

Math in Society

Edition 1.1

Revision 3

Contents

Voting Theory	1
<i>David Lippman</i>	
Weighted Voting	19
<i>David Lippman</i>	
Fair Division	31
<i>David Lippman</i>	
Graph Theory	49
<i>David Lippman</i>	
Scheduling	79
<i>David Lippman</i>	
Growth Models	95
<i>David Lippman</i>	
Finance	111
<i>David Lippman</i>	
Statistics	131
<i>David Lippman, Jeff Eldridge, onlinestatbook.com</i>	
Describing Data	143
<i>David Lippman, Jeff Eldridge, onlinestatbook.com</i>	
Historical Counting Systems	167
<i>Lawrence Morales</i>	
Solutions to Selected Exercises	201

David Lippman

Pierce College Ft Steilacoom

Copyright © 2010 David Lippman

This book was edited by David Lippman, Pierce College Ft Steilacoom
Development of this book was supported, in part, by the Transition Math Project

Statistics and Describing Data contain portions derived from works by:
Jeff Eldridge, Edmonds Community College (used under CC-BY-SA license)
www.onlinestatbook.com (used under public domain declaration)

Historical Counting Systems derived from work by:
Lawrence Morales, Seattle Central Community College (used under CC-BY-SA license)

Front cover photo:
Jan Tik, <http://www.flickr.com/photos/jantik/>, CC-BY 2.0

This text is licensed under a Creative Commons Attribution-Share Alike 3.0 United States License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

You are **free**:

- to Share** — to copy, distribute, display, and perform the work
- to Remix** — to make derivative works

Under the following conditions:

- Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license.

With the understanding that:

Waiver. Any of the above conditions can be waived if you get permission from the copyright holder.

Other Rights. In no way are any of the following rights affected by the license:

- Your fair dealing or fair use rights;
- Apart from the remix rights granted under this license, the author's moral rights;
- Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.
- Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page:
<http://creativecommons.org/licenses/by-sa/3.0/us/>

Voting Theory

In many decision making situations, it is necessary to gather the group consensus. This happens when a group of friends decides which movie to watch, when a company decides which design to produce, and when a democratic country elects its leaders.

While the basic idea of voting is fairly universal, the method by which those votes are used to determine a winner can vary. Amongst a group of friends, you may decide upon a movie by voting for all the movies you're willing to watch, with the winner being the one with the greatest approval. A company might eliminate unpopular designs then revote on the remaining. A country might look for the candidate with the most votes.

In deciding upon a winner, there is always one main goal: to reflect the preferences of the people in the most fair way possible.

Preference Schedules

To begin, we're going to need more information than a traditional ballot normally provides. A traditional ballot usually asks you to pick your favorite from a list of choices. This ballot fails to provide any information on how a voter would rank the alternatives if their first choice was unsuccessful.

A **preference ballot** is a ballot in which the voter ranks the choices in order of preference.

Example: A vacation club is trying to decide which destination to visit this year: Hawaii (H), Orlando (O), or Anaheim (A). Their votes are shown below:

	Bob	Ann	Marv	Alice	Eve	Omar	Lupe	Dave	Tish	Jim
1 st choice	A	A	O	H	A	O	H	O	H	A
2 nd choice	O	H	H	A	H	H	A	H	A	H
3 rd choice	H	O	A	O	O	A	O	A	O	O

These individual ballots are typically combined into one **preference schedule**, which shows the number of voters in the top row that voted for each option:

	1	3	3	3
1 st choice	A	A	O	H
2 nd choice	O	H	H	A
3 rd choice	H	O	A	O

Notice that by totaling the vote counts across the top of the preference schedule we can recover the total number of votes cast.

Plurality

The voting method we're most familiar with in the United States is the **plurality method**. In this method, the choice with the most first-preference votes is declared the winner. Ties are possible, and would have to be settled through some sort of run-off vote.

This method is sometimes mistakenly called the majority method, but it is not necessary for a choice to have gained a majority of votes to win. A majority is over 50%; it is possible for a winner to have a plurality without having a majority.

Example:

In our election from above, we had the preference table:

	1	3	3	3
1 st choice	A	A	O	H
2 nd choice	O	H	H	A
3 rd choice	H	O	A	O

For the plurality method, we only care about the first choice options. Totaling them up:

Anaheim: 4 votes

Orlando: 3 votes

Hawaii: 3 votes

Anaheim is the winner.

What's wrong with plurality?

The election above may seem totally clean, but there is a problem lurking that arises whenever there are three or more choices. Looking back at our preference table, how would our members vote if they only had two choices?

Anaheim vs Orlando: 7 out of the 10 would prefer Anaheim

	1	3	3	3
1 st choice	A	A	O	H
2 nd choice	O	H	H	A
3 rd choice	H	O	A	O

Anaheim vs Hawaii: 6 out of 10 would prefer Hawaii

	1	3	3	3
1 st choice	A	A	O	H
2 nd choice	O	H	H	A
3 rd choice	H	O	A	O

This doesn't seem right, does it? Anaheim just won the election, yet 6 out of 10 voters would prefer Hawaii! That hardly seems fair. Condorcet noticed how this could happen and for him we name our first **fairness criterion**. The fairness criteria are statements that seem like they should be true in a fair election.

Condorcet Criterion: If there is a choice that is preferred in every one-to-one comparison with the other choices, that choice should be the winner. We call this winner the Condorcet Winner.

In the example above, Hawaii is the Condorcet Winner. (Check for yourself that Hawaii is preferred over Orlando)

Example: Consider a city council election in a district that is 60% democratic voters and 40% republican voters. Even though city council is technically a nonpartisan office, people generally know the affiliations of the candidates. In this election there are three candidates: Don and Key, both democrats, and Elle, a republican. A preference schedule for the votes looks as follows:

	342	214	298
1 st choice	Elle	Don	Key
2 nd choice	Don	Key	Don
3 rd choice	Key	Elle	Elle

We can see a total of $342+214+298=854$ voters participated in this election. Computing percentage of first place votes:

Don: $214/854 = 25.1\%$

Key: $298/854 = 34.9\%$

Elle: $342/854 = 40.0\%$

So in this election, the democratic voters split their vote over the two democratic candidates, allowing the republican candidate Elle to win under the plurality method with 40% of the vote.

Analyzing this election closer, we see that it violates the Condorcet Criterion. Analyzing the one-to-one comparisons:

Elle vs Don: 342 prefer Elle; 512 prefer Don

Elle vs Key: 342 prefer Elle; 512 prefer Key

Don vs Key: 556 prefer Don; 298 prefer Key

So even though Don had the smallest number of first-place votes in the election, he is the Condorcet Winner, being preferred in every one-to-one comparison with the other candidates.

Insincere Voting

Situations like the one above, when there are more than one candidate that share somewhat similar points of view, can lead to **insincere voting**. Insincere voting is when a person casts a ballot counter to their actual preference for strategic purposes. In the case above, the democratic leadership might realize that Don and Key will split the vote, and encourage voters to vote for Key by officially endorsing him. Not wanting to see their party lose the election, as happened in the scenario above, Don's supporters might insincerely vote for Key, effectively voting against Elle.

Instant Runoff Voting

Instant Runoff Voting (IRV), also called Plurality with Elimination, is a modification of the plurality method that attempts to address the issue of insincere voting. In IRV, voting is done with preference ballots, and a preference schedule is generated. The choice with the *least* first-place votes is then eliminated from the election, and any votes for that candidate are redistributed to the voters' next choice. This continues until a choice has a majority (over 50%).

This is similar to the idea of holding runoff elections, but since every voter's order of preference is recorded on the ballot, the runoff can be computed without requiring a second costly election.

This voting method is used in several political elections around the world, including election of members of the Australian House of Representatives, and for county positions in Pierce County, Washington (until it was eliminated by voters in 2009). A version of it is used by the International Olympic Committee to select host nations.

Example: Consider the preference schedule below, in which a company's advertising team is voting on five different advertising slogans, called A, B, C, D, and E here for simplicity.

Round 1: Initial votes

	3	4	4	6	2	1
1 st choice	B	C	B	D	B	E
2 nd choice	C	A	D	C	E	A
3 rd choice	A	D	C	A	A	D
4 th choice	D	B	A	E	C	B
5 th choice	E	E	E	B	D	C

If this was a plurality election, note that B would be the winner with 9 votes, compared to 6 for D, 4 for C, and 1 for E.

There are total of $3+4+4+6+2+1 = 20$ votes. A majority would be 11 votes. No one yet has a majority, so we proceed to elimination rounds.

Round 2: We make our first elimination. Choice A has the fewest first-place votes, so we remove that choice, shifting everyone's options to fill the gaps.

	3	4	4	6	2	1
1 st choice	B	C	B	D	B	E
2 nd choice	C	D	D	C	E	D
3 rd choice	D	B	C	E	C	B
4 th choice	E	E	E	B	D	C

Still no choice has a majority, so we eliminate again.

Round 2: We make our second elimination. Choice E has the fewest first-place votes, so we remove that choice, shifting everyone's options to fill the gaps.

	3	4	4	6	2	1
1 st choice	B	C	B	D	B	D
2 nd choice	C	D	D	C	C	B
3 rd choice	D	B	C	B	D	C

Notice that the first and fifth columns have the same preferences now, we can condense those down to one column.

	5	4	4	6	1
1 st choice	B	C	B	D	D
2 nd choice	C	D	D	C	B
3 rd choice	D	B	C	B	C

Now B has 9 votes, C has 4 votes, and D has 7 votes. Still no majority, so we eliminate again.

Round 3: We make our third elimination. C has the fewest votes.

	5	4	4	6	1
1 st choice	B	D	B	D	D
2 nd choice	D	B	D	B	B

Condensing this down:

	9	11
1 st choice	B	D
2 nd choice	D	B

D has now gained a majority, and is declared the winner under IRV.

What's Wrong with IRV?

Example: Let's return to our City Council Election

	342	214	298
1 st choice	Elle	Don	Key
2 nd choice	Don	Key	Don
3 rd choice	Key	Elle	Elle

In this election, Don has the smallest number of first place votes, so Don is eliminated in the first round. The 214 people who voted for Don have their votes transferred to their second choice, Key.

	342	512
1 st choice	Elle	Key
2 nd choice	Key	Elle

So Key is the winner under the IRV method.

We can immediately notice that in this election, IRV violates the Condorcet Criterion. On the other hand, the temptation has been removed for Don's supporters to vote for Key; they now know their vote will be transferred to Key, not simply discarded.

Example: Consider the voting system below

	37	22	12	29
1 st choice	Adams	Brown	Brown	Carter
2 nd choice	Brown	Carter	Adams	Adams
3 rd choice	Carter	Adams	Carter	Brown

In this election, Carter would be eliminated in the first round, and Adams would be the winner with 66 votes to 34 for Brown.

Now suppose that the results were announced, but election official accidentally destroyed the ballots before they could be certified, and the votes had to be recast. Wanting to “jump on the bandwagon”, 10 of the voters who had originally voted in the order Brown, Adams, Carter change their vote to favor the presumed winner, changing those votes to Adams, Brown, Carter.

	47	22	2	29
1 st choice	Adams	Brown	Brown	Carter
2 nd choice	Brown	Carter	Adams	Adams
3 rd choice	Carter	Adams	Carter	Brown

In this re-vote, Brown will be eliminated in the first round, having the fewest first-place votes. After transferring votes, we find that Carter will win this election with 51 votes to Adams' 49 votes! Even though the only vote changes made *avored* Adams, the change ended up costing Adams the election. This doesn't seem right, and introduces our second fairness criterion:

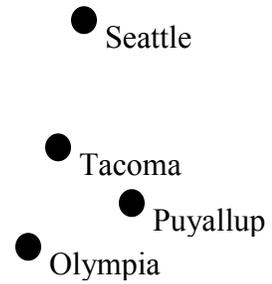
Monotonicity Criterion: If voters change their votes to increase the preference for a candidate, it should not harm that candidate's chances of winning.

This criterion is violated by this election. Note that even though the criterion is violated in this particular election, it does not mean that IRV always violates the criterion; just that IRV has the potential to violate the criterion in certain elections.

Borda Count

Borda Count is another voting method, named for Jean-Charles de Borda, who developed the system in 1770. In this method, points are assigned to candidates based on their ranking; 1 point for last choice, 2 points for second-to-last choice, and so on. The point values for all ballots are totaled, and the candidate with the largest point total is the winner.

Example: A group of mathematicians are getting together for a conference. The members are coming from four cities: Seattle, Tacoma, Puyallup, and Olympia. Their approximate relationship on a map is shown to the right. The votes for where to hold the conference were:



	51	25	10	14
1 st choice	Seattle	Tacoma	Puyallup	Olympia
2 nd choice	Tacoma	Puyallup	Tacoma	Tacoma
3 rd choice	Olympia	Olympia	Olympia	Puyallup
4 th choice	Puyallup	Seattle	Seattle	Seattle

In each of the 51 ballots ranking Seattle first, Puyallup will be given 1 point, Olympia 2 points, Tacoma 3 points, and Seattle 4 points. Multiplying the points per vote times the number of votes allows us to calculate points awarded:

	51	25	10	14
1 st choice	Seattle	Tacoma	Puyallup	Olympia
4 points	204	100	40	56
2 nd choice	Tacoma	Puyallup	Tacoma	Tacoma
3 points	153	75	30	42
3 rd choice	Olympia	Olympia	Olympia	Puyallup
2 points	102	50	20	28
4 th choice	Puyallup	Seattle	Seattle	Seattle
1 point	51	25	10	14

Adding up the points:

- Seattle: $204+25+10+14 = 253$ points
- Tacoma: $153+100+30+42 = 325$ points
- Puyallup: $51+75+40+28 = 194$ points
- Olympia: $102+50+20+56 = 228$ points

Under the Borda Count method, Tacoma is the winner of this election.

What's Wrong with Borda Count?

You might have already noticed one potential flaw of the Borda Count from the previous example. In that example, Seattle had a majority of first-choice votes, yet lost the election! This seems odd, and prompts our next fairness criterion:

Majority Criterion: If a choice has a majority of first-place votes, that choice should be the winner.

This example under the Borda Count violates the Majority Criterion. Notice also that this automatically means that the Condorcet Criterion will also be violated, as Seattle would have been preferred by 51% of voters in any head-to-head comparison.

Borda count is sometimes described as a consensus-based voting system, since it can sometimes choose a more broadly acceptable option over the one with majority support. In the example above, Tacoma is probably the best compromise location.

Because of this consensus behavior, Borda Count (or some variation) is commonly used in awarding sports awards. It is used to determine the Most Valuable Player in baseball, to rank teams in NCAA sports, and to award the Heisman trophy.

Copeland's Method (Pairwise Comparisons)

So far none of our voting methods have satisfied the Condorcet Criterion. The Copeland Method specifically attempts to satisfy the Condorcet Criterion by looking at pairwise (one-to-one) comparisons. In this method, each pair of candidates is compared, using all preferences to determine which of the two is more preferred. The more preferred candidate is awarded 1 point. If there is a tie, each candidate is awarded $\frac{1}{2}$ point. After all pairwise comparisons are made, the candidate with the most points, and hence the most pairwise wins, is declared the winner.

Variations of Copeland's Method are used in many professional organizations, including election of the Board of Trustees for the Wikimedia Foundation that runs Wikipedia.

Example: Consider our vacation group example from the beginning of the chapter.

	1	3	3	3
1 st choice	A	A	O	H
2 nd choice	O	H	H	A
3 rd choice	H	O	A	O

Making comparisons:

Anaheim vs Orlando: 7 votes to 3 votes: Anaheim gets 1 point

Anaheim vs Hawaii: 4 votes to 6 votes: Hawaii gets 1 point

Hawaii vs Orlando: 6 votes to 4 votes: Hawaii gets 1 point

Hawaii is the winner under Copeland's Method, having earned the most points.

Notice this is consistent with our determination of a Condorcet Winner.

Example: Consider the advertising group's vote we explored earlier:

	3	4	4	6	2	1
1 st choice	B	C	B	D	B	E
2 nd choice	C	A	D	C	E	A
3 rd choice	A	D	C	A	A	D
4 th choice	D	B	A	E	C	B
5 th choice	E	E	E	B	D	C

With 5 candidates, there are 10 comparisons to make:

A vs B: 11 votes to 9 votes A gets 1 point
 A vs C: 3 votes to 17 votes C gets 1 point
 A vs D: 10 votes to 10 votes A gets $\frac{1}{2}$ point, D gets $\frac{1}{2}$ point
 A vs E: 17 votes to 3 votes A gets 1 point
 B vs C: 10 votes to 10 votes B gets $\frac{1}{2}$ point, C gets $\frac{1}{2}$ point
 B vs D: 9 votes to 11 votes D gets 1 point
 B vs E: 13 votes to 7 votes B gets 1 point
 C vs D: 9 votes to 11 votes D gets 1 point
 C vs E: 17 votes to 3 votes C gets 1 point
 D vs E: 17 votes to 3 votes D gets 1 point

Totaling these up:

A gets $2\frac{1}{2}$ points
 B gets $1\frac{1}{2}$ points
 C gets $2\frac{1}{2}$ points
 D gets $3\frac{1}{2}$ points
 E gets 0 points

So Copeland's Method would choose D as the winner. Notice that in this case, D is not a Condorcet Candidate, but Copeland's Method still found a winner.

What's Wrong with Copeland's Method

As already noted, Copeland's Method does satisfy the Condorcet Criterion. It also satisfies the Majority Criterion and the Monotonicity Criterion. So is this the perfect method? Well, no.

Example: A committee is trying to award a scholarship to one of four students, Anna (A), Brian (B), Carlos (C), and Dimitry (D). The votes are shown below:

	5	5	6	4
1 st choice	D	A	C	B
2 nd choice	A	C	B	D
3 rd choice	C	B	D	A
4 th choice	B	D	A	C

Making the comparisons:

A vs B: 10 votes to 10 votes A gets $\frac{1}{2}$ point, B gets $\frac{1}{2}$ point
 A vs C: 14 votes to 6 votes: A gets 1 point
 A vs D: 5 votes to 15 votes: D gets 1 point
 B vs C: 4 votes to 16 votes: C gets 1 point
 B vs D: 15 votes to 5 votes: B gets 1 point
 C vs D: 11 votes to 9 votes: C gets 1 point

Totaling:

A has $1\frac{1}{2}$ points

B has $1\frac{1}{2}$ points

C has 2 points

D has 1 point

So Carlos is awarded the scholarship. However, the committee then discovers that Dimitry was not eligible for the scholarship (he failed his last math class). Even though this seems like it shouldn't affect the outcome, the committee decides to recount the vote, removing Dimitry from consideration:

	5	5	6	4
1 st choice	A	A	C	B
2 nd choice	C	C	B	A
3 rd choice	B	B	A	C

A vs B: 10 votes to 10 votes A gets $\frac{1}{2}$ point, B gets $\frac{1}{2}$ point

A vs C: 14 votes to 6 votes A gets 1 point

B vs C: 4 votes to 16 votes C gets 1 point

Totaling:

A has $1\frac{1}{2}$ points

B has $\frac{1}{2}$ point

C has 1 point

Suddenly Anna is the winner! This leads us to another fairness criterion.

The Independence of Irrelevant Alternatives (IIA) Criterion: If a non-winning choice is removed from the ballot, it should not change the winner of the election.

Equivalently, if choice A is preferred over choice B, introducing or removing a choice C should not make B preferred over A.

In this election, the IIA Criterion was violated.

This anecdote illustrating the IIA issue is attributed to Sidney Morgenbesser:

After finishing dinner, Sidney Morgenbesser decides to order dessert. The waitress tells him he has two choices: apple pie and blueberry pie. Sidney orders the apple pie. After a few minutes the waitress returns and says that they also have cherry pie at which point Morgenbesser says "In that case I'll have the blueberry pie."

Another disadvantage of Copeland's Method is that it is fairly easy for the election to end in a tie. There are a number of alternative methods based on satisfying the Condorcet Criterion that use more sophisticated methods for determining the winner when there is not a Condorcet Candidate.

So Where's the Fair Method?

At this point, you're probably asking why we keep looking at method after method just to point out that they are not fully fair. We must be holding out on the perfect method, right?

Unfortunately, no. A mathematical economist, Kenneth Arrow, was able to prove in 1949 that there is no voting method that will satisfy all the fairness criteria we have discussed. This is dubbed **Arrow's Impossibility Theorem**.

To see a very simple example of how difficult voting can be, consider the election below:

	5	5	5
1 st choice	A	C	B
2 nd choice	B	A	C
3 rd choice	C	B	A

Notice that in this election:

10 people prefer A to B

10 people prefer B to C

10 people prefer C to A

No matter whom we choose as the winner, 2/3 of voters would prefer someone else! This scenario is dubbed Condorcet's Voting Paradox, and demonstrates how voting preferences are not transitive (just because A is preferred over B, and B over C, does not mean A is preferred over C). In this election, there is no fair resolution.

It is because of this impossibility of a totally fair method that Plurality, IRV, Borda Count, and Copeland's Method are all still used. Usually the decision of which method to use is based on what seems most fair for the situation.

Approval Voting

Up until now, we've been considering voting methods that require ranking of candidates on a preference ballot. There is another method of voting that can be more appropriate in some decision making scenarios. With Approval Voting, the ballot asks you to mark all choices that you find acceptable. The results are tallied, and the option with the most approval is the winner.

Example: A group of friends is trying to decide upon a movie to watch. Three choices are provided, and each person is asked to mark with an "X" which movies they are willing to watch. The results are:

	Bob	Ann	Marv	Alice	Eve	Omar	Lupe	Dave	Tish	Jim
Titanic		X	X			X		X		X
Scream	X		X	X		X	X		X	
The Matrix	X	X	X	X	X		X			X

Totaling the results, we find
 Titanic received 5 approvals
 Scream received 6 approvals
 The Matrix received 7 approvals.

In this vote, The Matrix would be the winner.

What's Wrong with Approval Voting?

Approval voting can very easily violate the Majority Criterion. Consider the voting schedule:

	80	15	5
1 st choice	A	B	C
2 nd choice	B	C	B
3 rd choice	C	A	A

Clearly A is the majority winner. Now suppose that this election was held using Approval Voting, and every voter marked approval of their top two candidates.

A would receive approval from 80 voters
 B would receive approval from 100 voters
 C would receive approval from 20 voters

B would be the winner. Some argue that Approval Voting tends to vote the least disliked choice, rather than the most liked candidate.

Additionally, Approval Voting is susceptible to strategic insincere voting, in which a voter does not vote their true preference to try to increase the chances of their choice winning. For example, in the movie example above, suppose Bob and Alice would much rather watch Scream. They remove The Matrix from their approval list, resulting in a different result.

	Bob	Ann	Marv	Alice	Eve	Omar	Lupe	Dave	Tish	Jim
Titanic		X	X			X		X		X
Scream	X		X	X		X	X		X	
The Matrix		X	X		X		X			X

Totaling the results, we find Titanic received 5 approvals, Scream received 6 approvals, and The Matrix received 5 approvals. By voting insincerely, Bob and Alice were able to sway the result in favor of their preference.

Exercises

Skills

- To decide on a new website design, the designer asks people to rank three designs that have been created (labeled A, B, and C). The individual ballots are shown below. Create a preference table.

ABC, ABC, ACB, BAC, BCA, BCA, ACB, CAB, CAB, BCA, ACB, ABC

- To decide on a movie to watch, a group of friends all vote for one of the choices (labeled A, B, and C). The individual ballots are shown below. Create a preference table.

CAB, CBA, BAC, BCA, CBA, ABC, ABC, CBA, BCA, CAB, CAB, BAC

- The planning committee for a renewable energy trade show is trying to decide what city to hold their next show in. The votes are shown below.

Number of voters	9	19	11	8
1st choice	Buffalo	Atlanta	Chicago	Buffalo
2nd choice	Atlanta	Buffalo	Buffalo	Chicago
3rd choice	Chicago	Chicago	Atlanta	Atlanta

- How many voters voted in this election?
 - How many votes are needed for a majority? A plurality?
 - Find the winner under the plurality method.
 - Find the winner under the Borda Count Method.
 - Find the winner under the Instant Runoff Voting method.
 - Find the winner under Copeland's method.
- A non-profit agency is electing a new chair of the board. The votes are shown below.

Number of voters	11	5	10	3
1st choice	Atkins	Cortez	Burke	Atkins
2nd choice	Cortez	Burke	Cortez	Burke
3rd choice	Burke	Atkins	Atkins	Cortez

- How many voters voted in this election?
- How many votes are needed for a majority? A plurality?
- Find the winner under the plurality method.
- Find the winner under the Borda Count Method.
- Find the winner under the Instant Runoff Voting method.
- Find the winner under Copeland's method.

5. The student government is holding elections for president. There are four candidates (labeled A, B, C, and D for convenience). The preference schedule for the election is:

Number of voters	120	50	40	90	60	100
1st choice	C	B	D	A	A	D
2nd choice	D	C	A	C	D	B
3rd choice	B	A	B	B	C	A
4th choice	A	D	C	D	B	C

- How many voters voted in this election?
 - How many votes are needed for a majority? A plurality?
 - Find the winner under the plurality method.
 - Find the winner under the Borda Count Method.
 - Find the winner under the Instant Runoff Voting method.
 - Find the winner under Copeland's method.
6. The homeowners association is deciding a new set of neighborhood standards for architecture, yard maintenance, etc. Four options have been proposed. The votes are:

Number of voters	8	9	11	7	7	5
1st choice	B	A	D	A	B	C
2nd choice	C	D	B	B	A	D
3rd choice	A	C	C	D	C	A
4th choice	D	B	A	C	D	B

- How many voters voted in this election?
 - How many votes are needed for a majority? A plurality?
 - Find the winner under the plurality method.
 - Find the winner under the Borda Count Method.
 - Find the winner under the Instant Runoff Voting method.
 - Find the winner under Copeland's method.
7. Consider an election with 129 votes.
- If there are 4 candidates, what is the smallest number of votes that a plurality candidate could have?
 - If there are 8 candidates, what is the smallest number of votes that a plurality candidate could have?
8. Consider an election with 953 votes.
- If there are 7 candidates, what is the smallest number of votes that a plurality candidate could have?
 - If there are 8 candidates, what is the smallest number of votes that a plurality candidate could have?

9. Does this voting system having a Condorcet Candidate? If so, find it.

Number of voters	14	15	2
1st choice	A	C	B
2nd choice	B	B	C
3rd choice	C	A	A

10. Does this voting system having a Condorcet Candidate? If so, find it.

Number of voters	8	7	6
1st choice	A	C	B
2nd choice	B	B	C
3rd choice	C	A	A

11. The marketing committee at a company decides to vote on a new company logo. They decide to use approval voting. Their results are tallied below. Each column shows the number of voters with the particular approval vote. Which logo wins under approval voting?

Number of voters	8	7	6	3
A	X	X		
B	X		X	X
C		X	X	X

12. The downtown business association is electing a new chairperson, and decides to use approval voting. The tally is below, where each column shows the number of voters with the particular approval vote. Which candidate wins under approval voting?

Number of voters	8	7	6	3	4	2	5
A	X	X			X		X
B	X		X	X			X
C		X	X	X		X	
D	X		X		X	X	

Concepts

13. An election resulted in Candidate A winning, with Candidate B coming in a close second, and candidate C being a distant third. If for some reason the election had to be held again and C decided to drop out of the election, which caused B to become the winner, which is the primary fairness criterion violated in this election?
14. An election resulted in Candidate A winning, with Candidate B coming in a close second, and candidate C being a distant third. If for some reason the election had to be held again and many people who had voted for C switched their preferences to favor A, which caused B to become the winner, which is the primary fairness criterion violated in this election?

15. An election resulted in Candidate A winning, with Candidate B coming in a close second, and candidate C being a distant third. If in a head-to-head comparison a majority of people prefer B to A or C, which is the primary fairness criterion violated in this election?
16. An election resulted in Candidate A winning, with Candidate B coming in a close second, and candidate C being a distant third. If B had received a majority of first place votes, which is the primary fairness criterion violated in this election?

Exploration

17. In the election shown below under the Plurality method, explain why voters in the third column might be inclined to vote insincerely. How could it affect the outcome of the election?

Number of voters	96	90	10
1st choice	A	B	C
2nd choice	B	A	B
3rd choice	C	C	A

18. In the election shown below under the Borda Count method, explain why voters in the second column might be inclined to vote insincerely. How could it affect the outcome of the election?

Number of voters	20	18
1st choice	A	B
2nd choice	B	A
3rd choice	C	C

19. Compare and contrast the motives of the insincere voters in the two questions above.
20. Consider a two party election with preferences shown below. Suppose a third candidate, C, entered the race, and a segment of voters sincerely voted for that third candidate, producing the preference schedule from #17 above. Explain how other voters might perceive candidate C.

Number of voters	96	100
1st choice	A	B
2nd choice	B	A

21. In question 18, we showed that the outcome of Borda Count can be manipulated if a group of individuals change their vote.
- Show that it is possible for a single voter to change the outcome under Borda Count if there are four candidates.
 - Show that it is not possible for a single voter to change the outcome under Borda Count if there are three candidates.

22. Show that when there is a Condorcet Candidate, it is impossible to manipulate the vote under Pairwise Comparisons and help your preferred candidate. (possibly show that it would be possible to change the winner, just not to your preferred candidate)
23. The Pareto criterion is another fairness criterion that states: *If every voter prefers choice A to choice B, then B should not be the winner.* Explain why plurality, instant runoff, Borda count, and Copeland’s method all satisfy the Pareto condition.
24. Sequential Pairwise voting is a method not commonly used for political elections, but sometimes used for shopping and games of pool. In this method, the choices are assigned an order of comparison, called an agenda. The first two choices are compared. The winner of is then compared to the next choice on the agenda, and this continues until all choices have been compared against the winner of the previous comparison.

- a. Using the preference schedule below, apply Sequential Pairwise voting to determine the winner, using the agenda: A, B, C, D.

Number of voters	10	15	12
1st choice	C	A	B
2nd choice	A	B	D
3rd choice	B	D	C
4th choice	D	C	A

- b. Show that Sequential Pairwise voting can violate the Pareto criterion.
- c. Show that Sequential Pairwise voting can violate the Majority criterion.
25. The Coombs method is a variation of instant runoff voting. In Coombs method, the choice with the most last place votes is eliminated. Apply Coombs method to the preference schedules from questions 5 and 6.
26. Copeland’s Method is designed to identify a Condorcet Candidate if there is one, and is considered a Condorcet Method. There are many Condorcet Methods, which vary primarily in how they deal with ties, which are very common when a Condorcet winner does not exist. Copeland’s method does not have a tie-breaking procedure built-in. Research the Schulze method, another Condorcet method that is used by the Wikimedia foundation that runs Wikipedia, and give some examples of how it works.
27. The plurality method is used in most U.S. elections. Some people feel that Ross Perot in 1992 and Ralph Nader in 2000 changed what the outcome of the election would have been if they had not run. Research the outcomes of these elections and explain how each candidate could have affected the outcome of the elections (for the 2000 election, you may wish to focus on the count in Florida). Describe how an alternative voting method could have avoided this issue.

Weighted Voting

In a corporate shareholders meeting, each shareholders' vote counts proportional to the amount of shares they own. An individual with one share gets the equivalent of one vote, while someone with 100 shares gets the equivalent of 100 votes. This is called **weighted voting**, where each vote has some weight attached to it. Weighted voting is sometimes used to vote on candidates, but more commonly to decide “yes” or “no” on a proposal, sometimes called a motion. Weighted voting is applicable in corporate settings, as well as decision making in parliamentary governments and voting in the United Nations Security Council.

In weighted voting, we are most often interested in the power each voter has in influencing the outcome.

Beginnings

We'll begin with some basic vocabulary for weighted voting systems.

Each individual or entity casting a vote is called a **player** in the election. They're often notated as $P_1, P_2, P_3, \dots, P_N$, where N is the total number of voters.

Each player is given a **weight**, which usually represents how many votes they get.

The **quota** is the minimum weight needed for the votes or weight needed for the proposal to be approved.

A weighted voting system will often be represented in a shorthand form:

$$[q: w_1, w_2, w_3, \dots, w_n]$$

In this form, q is the quota, w_1 is the weight for player 1, and so on.

Example: In a small company, there are 4 shareholders. Mr. Smith has a 30% ownership stake in the company, Mr. Garcia has a 25% stake, Mrs. Hughes has a 25% stake, and Mrs. Lee has a 20% stake. They are trying to decide whether to open a new location. The company by-laws state that more than 50% of the ownership has to approve any decision like this. This could be represented by the weighted voting system:

$$[51: 30, 25, 25, 20]$$

Here we have treated the percentage ownership as votes, so Mr. Smith gets the equivalent of 30 votes, having a 30% ownership stake. Since more than 50% is required to approve the decision, the quota is 51, the smallest whole number over 50.

In order to have a meaningful weighted voting system, it is necessary to put some limits on the quota. While the quota does not have to be exactly $\frac{1}{2}$ the total number of votes, it must be at least $\frac{1}{2}$ the total number of votes. Likewise, the quota can't be larger than the total number of votes.

Why? Consider the voting system $[q: 3, 2, 1]$

Here there are 6 total votes. If the quota was set at only 3, then player 1 could vote yes, players 2 and 3 could vote no, and both would reach quota, which doesn't lead to a decision being made. If the quota was set to 7, then no group of voters could ever reach quota, and no decision can be made.

A Look at Power

Consider the voting system $[10: 11, 3, 2]$. Notice that in this system, player 1 can reach quota without the support of any other player. When this happens, we say that player 1 is a **dictator**. A player will be a dictator if their weight is equal to or greater than the quota. The dictator can also block any proposal from passing; the other players cannot reach quota without the dictator.

In the voting system $[8: 6, 3, 2]$, no player is a dictator. However, in this system, the quota can only be reached if player 1 is in support of the proposal; player 2 and 3 cannot reach quota without player 1's support. In this case, player 1 is said to have **veto power**. Notice that player 1 is not a dictator, since player 1 would still need player 2 or 3's support to reach quota.

With the system $[10: 7, 6, 2]$, player 3 is said to be a **dummy**, meaning they have no influence in the outcome. The only way the quota can be met is with the support of both players 1 and 2 (both of which would have veto power here); the vote of player 3 cannot affect the outcome.

Example: In the voting system $[16: 7, 6, 3, 3, 2]$, are any players dictators? Do any have veto power? Are any dummies?

No player can reach quota alone, so there are no dictators.

Without player 1, the rest of the players' weights add to 14, which doesn't reach quota, so player 1 has veto power. Likewise, without player 2, the rest of the players' weights add to 15, which doesn't reach quota, so player 2 also has veto power.

Since player 1 and 2 can reach quota with player 3 or player 4's support, neither player 3 or player 4 have veto power. However they cannot reach quota with player 5's support alone, so player 5 has no influence on the outcome and is a dummy.

To better define power, we need to introduce the idea of a **coalition**. A coalition is a group of players voting the same way. In the example above, $\{P_1, P_2, P_4\}$ would represent the coalition of players 1, 2 and 4. This coalition has a combined weight of $7+6+3 = 16$, which meets quota, so this would be a winning coalition.

A player is said to be **critical** in a coalition if their leaving the coalition would change it from a winning coalition to a losing coalition. In the coalition $\{P_1, P_2, P_4\}$, every player is critical. In the coalition $\{P_3, P_4, P_5\}$, no player is critical, since it wasn't a winning coalition to begin with. In the coalition $\{P_1, P_2, P_3, P_4, P_5\}$, only players 1 and 2 are critical; any other player could leave the coalition and it would still meet quota.

Example: In the Scottish Parliament, there were (as of 2009) 5 political parties: 47 representatives for the Scottish National Party, 46 for the Labour Party, 17 for the Conservative Party, 16 for the Liberal Democrats, and 2 for the Scottish Green Party. Typically all representatives from a party vote as a block, so the parliament can be treated like the weighted voting system:

[65: 47, 46, 17, 16, 2]

Consider the coalition $\{P_1, P_3, P_4\}$. No two players alone could meet the quota, so all three players are critical in this coalition.

In the coalition $\{P_1, P_3, P_4, P_5\}$, any player except P_1 could leave the coalition and it would still meet quota, so only P_1 is critical in this coalition.

Notice that a player with veto power will be critical in every winning coalition, since removing their support would prevent a proposal from passing.

Likewise, a dummy will never be critical, since their support will never change a losing coalition to a winning one.

Calculating Power: Banzhaf Power Index

The **Banzhaf power index** was originally created in 1946 by Lionel Penrose, but was reintroduced by John Banzhaf in 1965. The power index is a numerical way of looking at power in a weighted voting situation.

Banzhaf power index is calculated by:

- 1) List all winning coalitions
- 2) In each coalition, identify the players who are critical
- 3) Count up how many times each player is critical
- 4) Convert these counts to fractions or decimals by dividing by the total times any player is critical

Example: Consider the voting system [16: 7, 6, 3, 3, 2]. The winning coalitions are listed below, with the critical players underlined.

$\{\underline{P}_1, \underline{P}_2, \underline{P}_3\}$

$\{\underline{P}_1, \underline{P}_2, \underline{P}_4\}$

$\{\underline{P}_1, \underline{P}_2, P_3, P_4\}$

$\{\underline{P}_1, \underline{P}_2, \underline{P}_3, P_5\}$

$\{\underline{P}_1, \underline{P}_2, \underline{P}_4, P_5\}$

$\{\underline{P}_1, \underline{P}_2, P_3, P_4, P_5\}$

Counting up times that each player is critical:

$P_1 = 6$

$P_2 = 6$

$P_3 = 2$

$P_4 = 2$

$P_5 = 0$

Total of all: 16

Divide each player's count by 16 to convert to fractions or percents:

$$P_1 = 6/16 = 3/8 = 37.5\%$$

$$P_2 = 6/16 = 3/8 = 37.5\%$$

$$P_3 = 2/16 = 1/8 = 12.5\%$$

$$P_4 = 2/16 = 1/8 = 12.5\%$$

$$P_5 = 0/16 = 0 = 0\%$$

This power index measures a player's ability to influence the outcome of the vote. Notice that player 5 has a power index of 0, indicating that there is no coalition in which they would be critical and could influence the outcome. This means player 5 is a dummy, as we noted earlier.

Example: Revisiting the Scottish Parliament, with voting system [65: 47, 46, 17, 16, 2], the winning coalitions are listed, with the critical players underlined.

$$\begin{array}{l} \{\underline{P}_1, \underline{P}_2\} \\ \{\underline{P}_1, \underline{P}_2, P_3\} \\ \{\underline{P}_1, \underline{P}_2, P_4\} \\ \{\underline{P}_1, \underline{P}_2, P_5\} \\ \{P_1, P_2, P_3, P_4\} \\ \{P_1, P_2, P_3, P_5\} \\ \{\underline{P}_1, P_2, P_4, P_5\} \end{array} \qquad \begin{array}{l} \{\underline{P}_1, P_3, \underline{P}_4\} \\ \{\underline{P}_1, P_3, \underline{P}_5\} \\ \{\underline{P}_1, P_4, \underline{P}_5\} \\ \{\underline{P}_1, P_3, P_4, P_5\} \\ \{\underline{P}_2, \underline{P}_3, \underline{P}_4\} \\ \{\underline{P}_2, \underline{P}_3, \underline{P}_5\} \\ \{\underline{P}_2, \underline{P}_3, P_4, P_5\} \\ \{P_1, P_2, P_3, P_4, P_5\} \end{array}$$

Counting up times that each player is critical:

District	Times critical	Power index
P_1 (Scottish National Party)	9	$9/27 = 33.3\%$
P_2 (Labour Party)	7	$7/27 = 25.9\%$
P_3 (Conservative Party)	5	$5/27 = 18.5\%$
P_4 (Liberal Democrats Party)	3	$3/27 = 11.1\%$
P_5 (Scottish Green Party)	3	$3/27 = 11.1\%$

Interestingly, even though the Liberal Democrats party has only one less representative than the Conservative Party, and 14 more than the Scottish Green Party, their Banzhaf power index is the same as the Scottish Green Party's. In parliamentary governments, forming coalitions is an essential part of getting results, and a party's ability to help a coalition reach quota defines its influence.

Example: Banzhaf used this index to argue that the weighted voting system used in the Nassau County Board of Supervisors in New York was unfair. The county was divided up into 6 districts, each getting voting weight proportional to the population in the district, as shown below:

District	Weight
Hempstead #1	31
Hempstead #2	31
Oyster Bay	28
North Hempstead	21
Long Beach	2
Glen Cove	2

Translated into a weighted voting system, assuming a simple majority is needed for a proposal to pass:

$$[58: 31, 31, 28, 21, 2, 2]$$

Listing the winning coalitions:

- | | | |
|---------------------------------------|---------------------------------------|---------------------------------------|
| { <u>H1</u> , <u>H2</u> } | { <u>H1</u> , <u>OB</u> , NH} | { <u>H2</u> , <u>OB</u> , NH, LB} |
| { <u>H1</u> , <u>OB</u> } | { <u>H1</u> , <u>OB</u> , LB} | { <u>H2</u> , <u>OB</u> , NH, GC} |
| { <u>H2</u> , <u>OB</u> } | { <u>H1</u> , <u>OB</u> , GC} | { <u>H2</u> , <u>OB</u> , LB, GC} |
| { <u>H1</u> , <u>H2</u> , NH} | { <u>H1</u> , <u>OB</u> , NH, LB} | { <u>H2</u> , <u>OB</u> , NH, LB, GC} |
| { <u>H1</u> , <u>H2</u> , LB} | { <u>H1</u> , <u>OB</u> , NH, GC} | {H1, H2, OB} |
| { <u>H1</u> , <u>H2</u> , GC} | { <u>H1</u> , <u>OB</u> , LB, GC} | {H1, H2, OB, NH} |
| { <u>H1</u> , <u>H2</u> , NH, LB} | { <u>H1</u> , <u>OB</u> , NH, LB, GC} | {H1, H2, OB, LB} |
| { <u>H1</u> , <u>H2</u> , NH, GC} | { <u>H2</u> , <u>OB</u> , NH} | {H1, H2, OB, GC} |
| { <u>H1</u> , <u>H2</u> , LB, GC} | { <u>H2</u> , <u>OB</u> , LB} | {H1, H2, OB, NH, LB} |
| { <u>H1</u> , <u>H2</u> , NH, LB, GC} | { <u>H2</u> , <u>OB</u> , GC} | {H1, H2, OB, NH, GC} |
| | | {H1, H2, OB, NH, LB, GC} |

There's a lot of them! Counting up how many times each player is critical,

District	Times critical	Power index
Hempstead #1	16	$16/48 = 1/3 = 33\%$
Hempstead #2	16	$16/48 = 1/3 = 33\%$
Oyster Bay	16	$16/48 = 1/3 = 33\%$
North Hempstead	0	$0/48 = 0\%$
Long Beach	0	$0/48 = 0\%$
Glen Cove	0	$0/48 = 0\%$

It turns out that the three smaller districts are dummies. Any winning coalition requires two of the larger districts.

The weighted voting system that most Americans are most familiar with is the Electoral College system used to elect the President. In the Electoral College, states are given a number of votes equal to the number of their congress representatives (house + senate). Most states give all their electoral votes to the candidate that wins a majority in their state, turning the Electoral College into a weighted voting system, in which the states are the players. As I'm sure you can imagine, there are billions of possible winning coalitions, so the power index for the Electoral College has to be computed by a computer using approximation techniques.

Calculating Power: Shapley-Shubik Power Index

The **Shapley-Shubik** power index was introduced in 1954 by economists Lloyd Shapley and Martin Shubik, and provides a different approach for calculating power.

In situations like political alliances, the order in which players join an alliance could be considered the most important consideration. In particular, if a proposal is introduced, the player that joins the coalition and allows it to reach quota might be considered the most essential. The Shapley-Shubik power index counts how likely a player is to be **pivotal**. What does it mean for a player to be pivotal?

First, we need to change our approach to coalitions. Previously, the coalition $\{P_1, P_2\}$ and $\{P_2, P_1\}$ would be considered equivalent, since they contain the same players. We now need to consider the *order* in which players join the coalition. For that, we will consider **sequential coalitions** – coalitions in which the order players are listed reflect the order they joined the coalition. For example, the sequential coalition $\langle P_2, P_1, P_3 \rangle$ would mean that P_2 joined the coalition first, then P_1 , and finally P_3 . The angle brackets $\langle \rangle$ are used instead of curly brackets to distinguish sequential coalitions.

A **pivotal player** is the player in a sequential coalition that changes a coalition from a losing coalition to a winning one. Notice there can only be one pivotal player in any sequential coalition.

Example: In the weighted voting system $[8: 6, 4, 3, 2]$, which player is critical in the sequential coalition $\langle P_3, P_2, P_4, P_1 \rangle$?

The sequential coalition shows the order in which players joined the coalition. Consider the running totals as each player joins:

P_3	Total weight: 3	Not winning
P_3, P_2	Total weight: $3+4 = 7$	Not winning
P_3, P_2, P_4	Total weight: $3+4+2 = 9$	Winning
P_3, P_2, P_4, P_1	Total weight: $3+4+2+6 = 15$	Winning

Since the coalition *becomes* winning when P_4 joins, P_4 is the pivotal player in this coalition.

Shapley-Shubik Power Index is calculated by:

- 1) List all sequential coalitions
- 2) In each sequential coalition, determine the pivotal player
- 3) Count up how many times each player is pivotal
- 4) Convert these counts to fractions or decimals by dividing by the total number of sequential coalitions

How many sequential coalitions should we expect to have? If there are N players in the voting system, then there are N possibilities for the first player in the coalition, $N - 1$ possibilities for the second player in the coalition, and so on. Combining these possibilities, the total number of coalitions would be: $N(N - 1)(N - 2)(N - 3) \cdots (3)(2)(1)$. This calculation is called a **factorial**, and is notated $N!$. The number of sequential coalitions with N players is $N!$

Example: How many sequential coalitions will there be in a voting system with 7 players?

There will be 7! sequential coalitions. $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$

As you can see, computing the Shapley-Shubik power index by hand would be very difficult for voting systems that are not very small.

Example: Consider the weighted voting system [6: 4, 3, 2]. We will list all the sequential coalitions and identify the pivotal player. We will have $3! = 6$ sequential coalitions. The coalitions are listed, and the pivotal player is underlined.

$\langle P_1, \underline{P_2}, P_3 \rangle$ $\langle P_1, \underline{P_3}, P_2 \rangle$ $\langle P_2, \underline{P_1}, P_3 \rangle$
 $\langle P_2, P_3, \underline{P_1} \rangle$ $\langle P_3, P_2, \underline{P_1} \rangle$ $\langle P_3, \underline{P_1}, P_2 \rangle$

P_1 is pivotal 4 times, P_2 is pivotal 1 time, and P_3 is pivotal 1 time.

Player	Times pivotal	Power index
P_1	4	$4/6 = 66.7\%$
P_2	1	$1/6 = 16.7\%$
P_3	1	$1/6 = 16.7\%$

For comparison, the Banzhaf power index for the same weighted voting system would be P_1 : 60%, P_2 : 20%, P_3 : 20%. While the Banzhaf power index and Shapley-Shubik power index are usually not terribly different, the two different approaches usually produce somewhat different results.

Exercises

Skills

1. Consider the weighted voting system [47: 10,9,9,5,4,4,3,2,2]
 - a. How many players are there?
 - b. What is the total number (weight) of votes?
 - c. What is the quota in this system?

2. Consider the weighted voting system [31: 10,10,8,7,6,4,1,1]
 - a. How many players are there?
 - b. What is the total number (weight) of votes?
 - c. What is the quota in this system?

3. Consider the weighted voting system [q : 7,5,3,1,1]
 - a. What is the smallest value that the quota q can take?
 - b. What is the largest value that the quota q can take?
 - c. What is the value of the quota if at least two-thirds of the votes are required to pass a motion?

4. Consider the weighted voting system [q : 10,9,8,8,8,6]
 - a. What is the smallest value that the quota q can take?
 - b. What is the largest value that the quota q can take?
 - c. What is the value of the quota if at least two-thirds of the votes are required to pass a motion?

5. Consider the weighted voting system [13: 13, 6, 4, 2]
 - a. Identify the dictators, if any.
 - b. Identify players with veto power, if any
 - c. Identify dummies, if any.

6. Consider the weighted voting system [11: 9, 6, 3, 1]
 - a. Identify the dictators, if any.
 - b. Identify players with veto power, if any
 - c. Identify dummies, if any.

7. Consider the weighted voting system [19: 13, 6, 4, 2]
 - a. Identify the dictators, if any.
 - b. Identify players with veto power, if any
 - c. Identify dummies, if any.

8. Consider the weighted voting system [17: 9, 6, 3, 1]
 - a. Identify the dictators, if any.
 - b. Identify players with veto power, if any
 - c. Identify dummies, if any.

9. Consider the weighted voting system [15: 11, 7, 5, 2]
 - a. What is the weight of the coalition $\{P_1, P_2, P_4\}$
 - b. In the coalition $\{P_1, P_2, P_4\}$ which players are critical?
10. Consider the weighted voting system [17: 13, 9, 5, 2]
 - a. What is the weight of the coalition $\{P_1, P_2, P_3\}$
 - b. In the coalition $\{P_1, P_2, P_3\}$ which players are critical?
11. Find the Banzhaf power distribution of the weighted voting system [27: 16, 12, 11, 3]
12. Find the Banzhaf power distribution of the weighted voting system [33: 18, 16, 15, 2]
13. Consider the weighted voting system [q: 15, 8, 3, 1] Find the Banzhaf power distribution of this weighted voting system,
 - a. When the quota is 15
 - b. When the quota is 16
 - c. When the quota is 18
14. Consider the weighted voting system [q: 15, 8, 3, 1] Find the Banzhaf power distribution of this weighted voting system,
 - a. When the quota is 19
 - b. When the quota is 23
 - c. When the quota is 26
15. Consider the weighted voting system [17: 13, 9, 5, 2]. In the sequential coalition $\langle P_3, P_2, P_1, P_4 \rangle$ which player is pivotal?
16. Consider the weighted voting system [15: 13, 9, 5, 2]. In the sequential coalition $\langle P_1, P_4, P_2, P_3 \rangle$ which player is pivotal?
17. Find the Shapley-Shubik power distribution for the system [24: 17, 13, 11]
18. Find the Shapley-Shubik power distribution for the system [25: 17, 13, 11]

Concepts

19. Consider the weighted voting system [q: 7, 3, 1]
 - a. Which values of q result in a dictator (list all possible values)
 - b. What is the smallest value for q that results in exactly one player with veto power?
 - c. What is the smallest value for q that results in exactly two players with veto power?

20. Consider the weighted voting system $[q: 9, 4, 2]$
- Which values of q result in a dictator (list all possible values)
 - What is the smallest value for q that results in exactly one player with veto power?
 - What is the smallest value for q that results in exactly two players with veto power?
21. Using the Shapley-Shubik method, is it possible for a dummy to be pivotal?
22. If a specific weighted voting system requires a unanimous vote for a motion to pass:
- Which player will be pivotal in any sequential coalition?
 - How many winning coalitions will there be?
23. Consider a weighted voting system with three players. If Player 1 is the only player with veto power, there are no dictators, and there are no dummies:
- Find the Banzhof power distribution.
 - Find the Shapley-Shubik power distribution
24. Consider a weighted voting system with three players. If Players 1 and 2 have veto power but are not dictators, and Player 3 is a dummy:
- Find the Banzhof power distribution.
 - Find the Shapley-Shubik power distribution
25. An executive board consists of a president (P) and three vice-presidents (V_1, V_2, V_3). For a motion to pass it must have three yes votes, one of which must be the president's. Find a weighted voting system to represent this situation.
26. On a college's basketball team, the decision of whether a student is allowed to play is made by four people: the head coach and the three assistant coaches. To be allowed to play, the student needs approval from the head coach and at least one assistant coach. Find a weighted voting system to represent this situation.
27. In a corporation, the shareholders receive 1 vote for each share of stock they hold, which is usually based on the amount of money they invested in the company. Suppose a small corporation has two people who invested \$30,000 each, two people who invested \$20,000 each, and one person who invested \$10,000. If they receive one share of stock for each \$1000 invested, and any decisions require a majority vote, set up a weighted voting system to represent this corporation's shareholder votes.
28. A contract negotiations group consists of 4 workers and 3 managers. For a proposal to be accepted, a majority of workers and a majority of managers must approve of it. Calculate the Banzhof power distribution for this situation. Who has more power: a worker or a manager?

29. The United Nations Security Council consists of 15 members, 10 of which are elected, and 5 of which are permanent members. For a resolution to pass, 9 members must support it, which must include all 5 of the permanent members. Set up a weighted voting system to represent the UN Security Council and calculate the Banzhaf power distribution.

Exploration

30. In the U.S., the Electoral College is used in presidential elections. Each state is awarded a number of electors equal to the number of representatives (based on population) and senators (2 per state) they have in congress. Since most states award the winner of the popular vote in their state all their state's electoral votes, the Electoral College acts as a weighted voting system. To explore how the Electoral College works, we'll look at a mini-country with only 4 states. Here is the outcome of a hypothetical election:

State	Smalota	Medigan	Bigonia	Hugodo
Population	50,000	70,000	100,000	240,000
Votes for A	40,000	50,000	80,000	50,000
Votes for B	10,000	20,000	20,000	190,000

- If this country did not use an Electoral College, which candidate would win the election?
- Suppose that each state gets 1 electoral vote for every 10,000 people. Set up a weighted voting system for this scenario, calculate the Banzhaf power index for each state, then calculate the winner if each state awards all their electoral votes to the winner of the election in their state.
- Suppose that each state gets 1 electoral vote for every 10,000 people, plus an additional 2 votes. Set up a weighted voting system for this scenario, calculate the Banzhaf power index for each state, then calculate the winner if each state awards all their electoral votes to the winner of the election in their state.
- Suppose that each state gets 1 electoral vote for every 10,000 people, and awards them based on the number of people who voted for each candidate. Additionally, they get 2 votes that are awarded to the majority winner in the state. Calculate the winner under these conditions.
- Does it seem like an individual state has more power in the Electoral College under the vote distribution from part c or from part d?
- Research the history behind the Electoral College to explore why the system was introduced instead of using a popular vote. Based on your research and experiences, state and defend your opinion on whether the Electoral College system is or is not fair.

Fair Division

Whether it is two kids sharing a candy bar or a couple splitting assets during a divorce, there are times in life where items of value need to be divided between two or more parties. While some cases can be handled through mutual agreement or mediation, in others the parties are adversarial or cannot reach a decision all feel is fair. In these cases, fair division methods can be utilized.

A fair division method is a procedure that can be followed that will result in a division of items in a way so that each party feels they have received their fair share. For these methods to work, we have to make a few assumptions:

- 1) The parties are non-cooperative, so the method must operate without communication between the parties.
- 2) The parties act rationally, meaning they act in their best interest, and do not make emotional decisions.

With these methods, each party will be entitled to some **fair share**. When there are N parties equally dividing something, that fair share would be $1/N$. For example, if there were 4 parties, each would be entitled to a fair share of $1/4 = 25\%$ of the whole. More specifically, they are entitled to a share that *they* value as 25% of the whole.

It should be noted that a fair division method simply needs to guarantee that each party will receive a share they view as fair. A basic fair division does not need to be envy free; an **envy-free** division is one in which no party would prefer another party's share over their own. A basic fair division also does not need to be Pareto optimal; a **Pareto optimal** division is one in which no other division would make a participant better off without making someone else worse off. Nor does fair division have to be **equitable**; an equitable division is one in which the proportion of the whole each party receives, judged by their own valuation, is the same. Basically, a simple fair division doesn't have to be the best possible division – it just has to give each party their fair share.

Example: Suppose that 4 friends have pooled money to buy a \$12 pizza that is half pepperoni, half veggie. Since they all put in the same amount of money, each person's fair share is \$3, or a piece they value as 25% of the pizza.

It is important to keep in mind that each party might value portions of the whole differently. For example, a vegetarian would probably put zero value on the pepperoni half of the pizza. Suppose that Steve likes pepperoni twice as much as veggie. He would value the veggie half as being worth \$4 and the pepperoni half as \$8, twice as much. If the pizza was divided up into 4 pepperoni slices and 4 veggie slices, he would value a pepperoni slice as being worth \$2, and a veggie slice as being worth \$1. A fair share for Steve would be one pepperoni slice and one veggie slice, $1\frac{1}{2}$ pepperoni slices, 3 veggie slices, or a variety of more complicated possibilities.

You will find that many examples and exercises in this topic involve dividing food – dividing candy, cutting cakes, sharing pizza, etc. This may make this topic seem somewhat trivial, but instead of cutting a cake, we might be drawing borders dividing Germany after WWII. Instead of splitting a bag of candy, siblings might be dividing belongings from an inheritance. Mathematicians often characterize very important and contentious issues in terms of simple items like cake to separate any emotional influences from the mathematical method.

There are two broad classifications of fair division methods: those that apply to **continuously divisible** items, and those that apply to **discretely divisible** items. Continuously divisible items are things that can be divided into pieces of any size, like dividing a candy bar into two pieces or drawing borders to split a piece of land into smaller plots. Discretely divisible items are when you are dividing several items that cannot be broken apart easily, such as assets in a divorce (house, car, furniture, etc).

Divider-Chooser

The first method we will look at is a method for continuously divisible items. This method will be familiar to many parents - it is the “You cut, I choose” method. In this method, one party is designated the **divider** and the other the **chooser**, perhaps with a coin toss. The method works as follows:

- 1) The divider cuts the item into two pieces that are, in his eyes, equal in value.
- 2) The chooser selects either of the two pieces
- 3) The divider receives the remaining piece

Notice that the divider-chooser method is specific to a two-party division. Examine why this method guarantees a fair division; since the divider doesn’t know which piece he will receive, the rational action for him to take would be to divide the whole into two pieces he values equally. There is no incentive for the divider to attempt to “cheat” since he doesn’t know which piece he will receive. Since the chooser can pick either piece, she is guaranteed that one of them is worth at least 50% of the whole in her eyes. The chooser is guaranteed a piece she values as at least 50%, and the divider is guaranteed a piece he values at 50%.

Example: Two retirees, Fred and Martha, buy a vacation beach house in Florida together, with the agreement that they will split the year into two parts. Fred is chosen to be the divider, and splits the year into two pieces: November – February and March – October. Even though the first piece is 4 months and the second is 8 months, Fred places equal value on both pieces since he really likes to be in Florida during the winter. Martha gets to pick whichever piece she values more. Suppose she values all months equally. In this case, she would choose the March – October time, resulting in a piece that she values as $8/12 = 66.7\%$ of the whole. Fred is left with the November – February slot which he values as 50% of the whole.

Of course, in this example, Fred and Martha probably could have discussed their preferences and reached a mutually agreeable decision. The divider-chooser method is more necessary in cases where the parties are suspicious of each others motives, or are unable to communicate effectively, such as two countries drawing a border, or two children splitting a candy bar.

Things quickly become more complicated when we have more than two parties involved. We will look at three different approaches. But first, let us look at one that doesn't work.

How not to divide with 3 parties

When first approaching the question of 3-party fair division, it is very tempting to propose this method: Randomly designate one participant to be the divider, and designate the rest choosers. Proceed as follows:

- 1) Have the divider divide the item into 3 pieces
- 2) Have the first chooser select any of the three pieces they feel is worth a fair share
- 3) Have the second chooser select either of the remaining pieces
- 4) The divider gets the piece left.

Example: Suppose we have three people splitting a cake. We can immediately see that the divider will receive a fair share as long as they cut the cake fairly at the beginning. The first chooser certainly will also receive a fair share. What about the second chooser? Suppose each person values the three pieces like this:

	Piece 1	Piece 2	Piece 3
Chooser 1	40%	30%	30%
Chooser 2	45%	30%	25%
Divider	33.3%	33.3%	33.3%

Since the first chooser will clearly select Piece 1, the second chooser is left to select between Piece 2 and Piece 3, neither of which she values as a fair share ($1/3$ or about 33.3%). This example shows that this method does *not* guarantee a fair division.

To handle division with 3 or more parties, we'll have to take a more clever approach.

Lone Divider

The lone divider method works for any number of parties – we will use N for the number of parties. One participant is randomly designated the divider, and the rest of the participants are designated as choosers. The method proceeds as follows:

- 1) The divider divides the item into N pieces, which we'll label s_1, s_2, \dots, s_N .
- 2) Each of the choosers will separately list which pieces they consider to be a fair share. This is called their **declaration**, or **bid**.
- 3) The lists are examined by the referee. There are two possibilities:
 - a. If it is possible to give each party a piece they declared then do so, and the divider gets the remaining piece.
 - b. If two or more parties both want the same piece and none other, then give a non-contested piece to the divider. The rest of the pieces are combined and repeat the entire procedure with the remaining parties. If there are only two parties left, they can use divider-chooser.

Example: Consider the example from earlier, in which the pieces were valued as:

	Piece 1	Piece 2	Piece 3
Chooser 1	40%	30%	30%
Chooser 2	45%	30%	25%
Divider	33.3%	33.3%	33.3%

Each chooser makes a declaration of which pieces they value as a fair share. In this case,

Chooser 1 would make the declaration: Piece 1

Chooser 2 would make the declaration: Piece 1

Since both choosers want the same piece, we cannot immediately allocate the pieces. The lone divider method specifies that we give a non-contested piece to the divider. Both pieces 2 and 3 are uncontested, so we flip a coin and give Piece 2 to the divider. Piece 1 and 3 are then recombined to make a piece worth 70% to Chooser 1, and 70% to Chooser 2. Since there are only two players left, they can divide the recombined pieces using divider-chooser. Each is guaranteed a piece they value as at least 35%, which is a fair share.

Example: Suppose that Abby, Brian, Chris, and Dorian are dividing a plot of land. Dorian was selected to be the divider through a coin toss. Each person's valuation of each piece is shown below.

	Piece 1	Piece 2	Piece 3	Piece 4
Abby	15%	30%	20%	35%
Brian	30%	35%	10%	25%
Chris	20%	45%	20%	15%
Dorian	25%	25%	25%	25%

Based on this, their declarations should be:

Abby: Piece 2, Piece 4

Brian: Piece 1, Piece 2, Piece 4

Chris: Piece 2

This case can be settled simply – by awarding Piece 2 to Chris, Piece 4 to Abby, Piece 1 to Brian, and Piece 3 to Dorian. Each person receives a piece that they value as at least a fair share (25% value).

Example: Suppose the valuations in the previous problem were:

	Piece 1	Piece 2	Piece 3	Piece 4
Abby	15%	30%	20%	35%
Brian	20%	35%	10%	35%
Chris	20%	45%	20%	15%
Dorian	25%	25%	25%	25%

The declarations would be:

Abby: Piece 2, Piece 4

Brian: Piece 2, Piece 4

Chris: Piece 2

Notice in this case that there is no simple settlement. So in this case, the non-disputed piece, Piece 3, is awarded to the original divider Dorian, and the procedure is repeated with the remaining three players.

Suppose that on the second round of this method Brian is selected to be the divider, three new pieces are cut, and the valuations are as follows:

	Piece 1	Piece 2	Piece 3
Abby	40%	30%	30%
Brian	33.3%	33.3%	33.3%
Chris	50%	20%	30%

The declarations here would be:

Abby: Piece 1

Chris: Piece 1

Once again we have a standoff. Brian can be awarded either of Piece 2 or Piece 3, and the remaining pieces can be recombined. Since there are only two players left, they can divide the remaining land using the basic divider-chooser method.

Last Diminisher

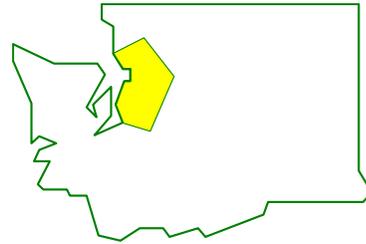
The Last Diminisher method is another approach to division among 3 or more parties. In this method, the parties are randomly assigned an order, perhaps by pulling names out of a hat.

The method then proceeds as follows:

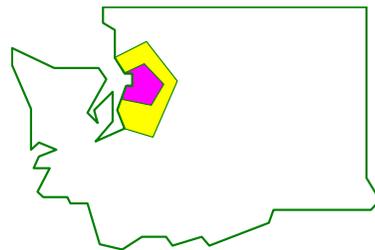
- 1) The first person cuts a slice they value as a fair share.
- 2) The second person examines the piece
 - a. If they think it is worth less than a fair share, they then pass on the piece unchanged.
 - b. If they think the piece is worth more than a fair share, they trim off the excess and lay claim to the piece. The trimmings are added back into the to-be-divided pile.
- 3) Each remaining person, in turn, can either pass or trim the piece
- 4) After the last person has made their decision, the last person to trim the slice receives it. If no one has modified the slice, then the person who cut it receives it.
- 5) Whoever receives the piece leaves with their piece and the process repeats with the remaining people. Continue until only 2 people remain; they can divide what is left with the divider-chooser method.

Example: Suppose that four salespeople are dividing up Washington State into sales regions; each will get one region to work in. They pull names from a hat to decide play order.

Round 1. The first salesman, Bob, draws a region around Seattle, the most populous area of the state. The piece Bob cuts and automatically lays claim to is shown in yellow.

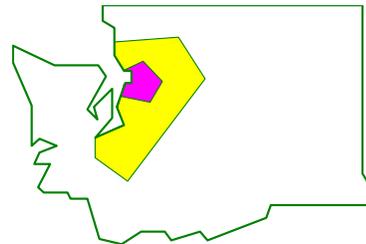


The second salesman, Henry, felt that this region was worth more than 25%, each player's fair share. Because of this, Henry opts to trim this piece. The new piece is shown in pink. The trimmings (in yellow) return to the to-be-divided portion of the state. Henry automatically lays claim to this smaller piece since he trimmed it.

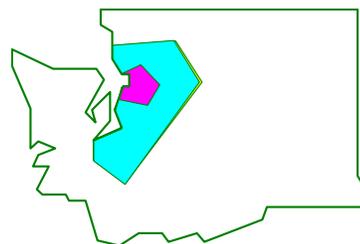


The third saleswoman, Marjo, feels this piece is worth less than 25% and passes, as does the fourth saleswoman, Beth. Since both pass, the last person to trim it, Henry, receives the piece.

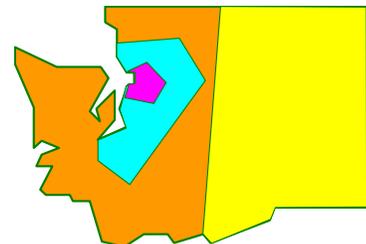
Round 2. The second round begins with Bob laying claim to a piece, shown again in yellow. Henry already has a piece, so is out of the process now. Marjo passes on this piece, feeling it is worth less than a fair share.



Beth, on the other hand, feels the piece as currently drawn is worth 35%. Beth is in an advantageous position, being the last to make a decision. Even though Beth values this piece at 35%, she can cut a very small amount and still lay claim to it. So Beth barely cuts the piece, resulting in a piece (blue) that is essentially worth 35% to her.



Round 3. At this point, Bob and Marjo are the only players without a piece. Since there are two of them, they can finish the division using the divider-chooser method. They flip a coin, and determine that Marjo will be the divider. Marjo draws a line dividing the remainder of the state into two pieces. Bob chooses the Eastern piece, leaving Marjo with the Western half.



Notice that in this division, Henry and Marjo ended up with pieces that they feel are worth exactly 25% - a fair share. Beth was able to receive a piece she values as more than a fair share, and Bob may feel the piece he received is worth more than 25%.

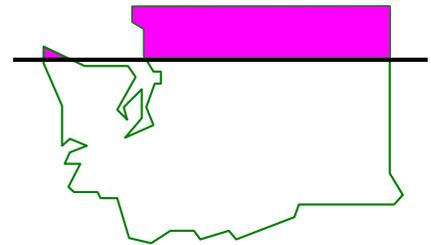
Moving Knife

A somewhat different approach to continuous fair division is called the moving knife procedure. In this method, applied to a cake,

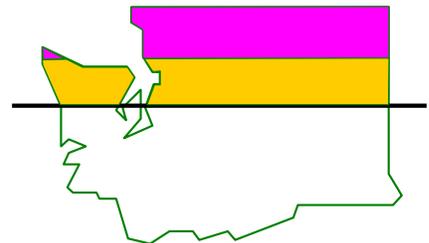
- 1) A referee starts moving a knife from left to right across a cake.
- 2) As soon as any player feels the piece to the left of the knife is worth a fair share, they shout "STOP." The referee then cuts the cake at the current knife position and the player who called stop gets the piece to the left of the knife.
- 3) This procedure continues until there is only one player left. The player left gets the remaining cake.

Example: Suppose that our four salespeople from above decided to use this approach to divide Washington. Rather than move the "knife" from left to right, they decide to move it from top to bottom.

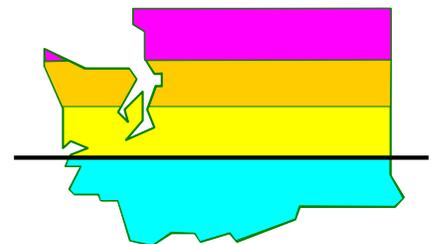
The referee starts moving a line down a map of the state. Henry is the first to call STOP when the knife is at the position shown, giving him the portion of the state above the line.



Marjo is the next to call STOP when the knife is at the position shown, giving her the second portion of the state.



Bob is the next to call STOP, leaving Beth with the southernmost portion of the state.



While this method guarantees a fair division, it clearly results in some potentially silly divisions in a case like this. The method is probably better suited to situations like dividing an actual cake.

Sealed Bids Method

The Sealed Bids method provides a method for *discrete* fair division, allowing for the division of items that cannot be split into smaller pieces, like a house or a car. Because of this, the method requires that all parties have a large amount of cash at their disposal.

The method begins by compiling a list of items to be divided. Then:

- 1) Each party involved lists, in secret, a dollar amount they value each item to be worth. This is their sealed bid.
- 2) An arbitrator collects the bids. For each party, they total the value of all the items, and divide it by the number of parties. This defines their fair share.
- 3) Each item is awarded to the highest bidder.
- 4) For each party, the value of all items received is totaled. If the value is more than that party's fair share, they pay the difference into a holding pile. If the value is less than that party's fair share, they receive the difference from the holding pile. This ends the initial allocation.
- 5) In most cases, there will be a surplus, or leftover money, in the holding pile. The surplus is divided evenly between all the players. This produces the final allocation.

Example: Sam and Omar have cohabitated for the last 3 years, during which time they shared the expense of purchasing several items for their home. Sam has accepted a job in another city, and now they find themselves needing to divide their shared assets.

Each records their value of each item, as shown below.

	Sam	Omar
Couch	\$150	\$100
TV	\$200	\$250
Video game system	\$250	\$150
Surround sound system	\$50	\$100

Sam's total valuation of the items is $\$150 + \$200 + \$250 + \$50 = \$650$, making a fair share for Sam $\$650/2 = \325 .

Omar's total valuation of the items is $\$100 + \$250 + \$150 + \$100 = \$600$, making a fair share for Omar $\$600/2 = \300 .

Each item is now awarded to the highest bidder. Sam will receive the couch and video game system, providing $\$150 + \$250 = \$400$ of value to Sam. Since this exceeds his fair share, he has to pay the difference, \$75, into a holding pile.

Omar will receive the TV and surround sound system, providing $\$250 + \$100 = \$350$ in value. This is more than his fair share, so he has to pay the difference, \$50, into the holding pile.

Thus, in the initial allocation, Sam receives the couch and video game system and pays \$75 into the holding pile. Omar receives the TV and surround sound system and pays \$50 into the holding pile. At this point, both players would feel they have received a fair share.

There is now \$125 remaining in the holding pile. This is the surplus from the division. This is now split evenly, and both Sam and Omar are given back \$62.50. Since Sam had paid in \$75, the net effect is that he paid \$12.50. Since Omar had originally paid in \$50, the net effect is that he receives \$12.50.

Thus, in the initial allocation, Sam receives the couch and video game system and pays \$12.50 to Omar. Omar receives the TV and surround sound system and receives \$12.50. At this point, both players feel they have received more than a fair share.

Example: Four small graphic design companies are merging operations to become one larger corporation. In this merger, there are a number of issues that need to be settled. Each company is asked to place a monetary value (in thousands of dollars) on each issue:

	Super Designs	DesignByMe	LayoutPros	Graphix
Company name	\$5	\$3	\$3	\$6
Company location	\$8	\$9	\$7	\$6
CEO	\$10	\$5	\$6	\$7
Chair of the board	\$7	\$6	\$6	\$8

We can then calculate for each company:

	Super Designs	DesignByMe	LayoutPros	Graphix
Total value of all issues	\$30	\$23	\$22	\$27
Fair share	\$7.50	\$5.75	\$5.50	\$6.75

The items would then be allocated to the company that bid the most for each.

Company name would be awarded to Graphix

Company location would be awarded to DesignByMe

CEO would be awarded to Super Designs

Chair of the board would be awarded to Graphix

For each company, we calculate the total value of the items they receive, and how much they get or pay in the initial allocation

	Super Designs	DesignByMe	LayoutPros	Graphix
Total value of issues awarded	\$10	\$9	\$0	\$14
Amount they pay/get	\$7.50-\$10 Pay \$2.50	\$5.75 - \$9 Pay \$3.25	\$5.50 - \$0 Get \$5.50	\$6.75 - \$14 Pay \$7.25

After the initial allocation, there is a total of $\$2.50 + \$3.25 - \$5.50 + \$7.25 = \$7.50$ in surplus. Dividing that evenly, each company gets \$1.875 (approximately)

	Super Designs	DesignByMe	LayoutPros	Graphix
Amount they pay/get	Pay \$2.50 Get \$1.875	Pay \$3.25 Get \$1.875	Get \$5.50 Get \$1.875	Pay \$7.25 Get \$1.875
After surplus	Pay \$0.625	Pay \$1.375	Get \$7.375	Pay \$5.375

So in the final allocation,

Super Designs wins the CEO, and pays \$625 (\$0.625 thousand)

DesignByMe wins the company location and pays \$1,375 (\$1.375 thousand)

LayoutPros wins no issues, but receives \$7,375 in compensation

Graphix wins the company name and chair of the board, and pays \$5,375.

Example: Fair division does not always have to be used for items of value. It can also be used to divide undesirable items. Suppose Chelsea and Mariah are sharing an apartment, and need to split the chores for the household. They list the chores, assigning a *negative* dollar value to each item; in other words, the amount they would pay for someone else to do the chore (a per week amount). We will assume, however, that they are committed to doing all the chores themselves and not hiring a maid.

	Chelsea	Mariah
Vacuuming	-\$10	-\$8
Cleaning bathroom	-\$14	-\$20
Doing dishes	-\$4	-\$6
Dusting	-\$6	-\$4

We can then calculate fair share:

	Chelsea	Mariah
Total value	-\$34	-\$38
Fair Share	-\$17	-\$19

We award to the person with the largest bid. For example, we award vacuuming to Mariah since she dislikes it less (remember $-8 > -10$).

Chelsea gets cleaning the bathroom and doing dishes. Value: -\$18

Mariah gets vacuuming and dusting. Value: -\$12

Notice that Chelsea's fair share is -\$17 but she is doing chores she values at -\$18. She should get \$1 to bring her to a fair share. Mariah is doing chores valued at -\$12, but her fair share is -\$19. She needs to pay \$7 to bring her to a fair share.

This creates a surplus of \$6, which will be divided between the two. In the final allocation:

Chelsea gets cleaning the bathroom and doing dishes, and receives $\$1 + \$3 = \$4/\text{week}$.

Mariah gets vacuuming and dusting, and pays $\$7 - \$3 = \$4/\text{week}$.

Exercises

Skills

1. Chance and Brianna buy a pizza for \$10 that is half pepperoni and half veggie. They cut the pizza into 8 slices.

If Chance likes veggie three times as much as pepperoni, what is the value of a slice that is half pepperoni, half veggie?

2. Ahmed and Tiana buy a cake for \$14 that is half chocolate and half vanilla. They cut the cake into 8 slices.

If Ahmed likes chocolate four times as much as vanilla, what is the value of a slice that is half chocolate, half vanilla?

3. Erin, Catherine, and Shannon are dividing a large bag of candy. They randomly split the bag into three bowls. The values of the entire bag and each of the three bowls in the eyes of each of the players are shown below. For each player, identify which bowls they value as a fair share.

	Whole Bag	Bowl 1	Bowl 2	Bowl 3
Erin	\$5	\$2.75	\$1.25	\$1.00
Catherine	\$4	\$0.75	\$2.50	\$0.75
Shannon	\$8	\$1.75	\$2.25	\$4.00

4. Jenna, Tatiana, and Nina are dividing a large bag of candy. They randomly split the bag into three bowls. The values of the entire bag and each of the three bowls in the eyes of each of the players are shown below. For each player, identify which bowls they value as a fair share.

	Whole Bag	Bowl 1	Bowl 2	Bowl 3
Jenna	\$8	\$4.50	\$1.75	\$1.75
Tatiana	\$4	\$1.00	\$1.00	\$2.00
Nina	\$6	\$1.75	\$2.50	\$1.75

5. Dustin and Kendra want to split a bag of fun-sized candy, and decide to use the divider-chooser method. The bag contains 100 Snickers, 100 Milky Ways, and 100 Reese's, which Dustin values at \$1, \$5, and \$2 respectively. (This means Dustin values the 100 Snickers together at \$1, or \$0.01 for 1 Snickers).

If Kendra is the divider, and in one half puts:

25 Snickers, 20 Milky Ways, and 60 Reese's

- a. What is the value of this half in Dustin's eyes?
- b. Does Dustin consider this a fair share?
- c. If Dustin was a divider, find a possible division that is consistent with his value system.

6. Dustin and Kendra want to split a bag of fun-sized candy, and decide to use the divider-chooser method. The bag contains 100 Snickers, 100 Milky Ways, and 100 Reese's, which Dustin values at \$1, \$3, and \$5 respectively. (This means Dustin values the 100 Snickers together at \$1, or \$0.01 for 1 Snickers).

If Kendra is the divider, and in one half puts:

30 Snickers, 40 Milky Ways, and 66 Reese's

- What is the value of this half in Dustin's eyes?
 - Does Dustin consider this a fair share?
 - If Dustin was a divider, find a possible division that is consistent with his value system.
7. Maggie, Meredith, Holly, and Zoe are dividing a piece of land using the lone-divider method. The values of the four pieces of land in the eyes of the each player are shown below.

	Piece 1	Piece 2	Piece 3	Piece 4
Maggie	21%	27%	32%	20%
Meredith	27%	29%	22%	22%
Holly	23%	14%	41%	22%
Zoe	25%	25%	25%	25%

- Who was the divider?
 - If playing honestly, what will each player's declaration be?
 - Find the final division.
8. Cody, Justin, Ahmed, and Mark are going to share a vacation property. The year will be divided into 4 time slots using the lone-divider method. The values of each time slot in the eyes of the each player are shown below.

	Time 1	Time 2	Time 3	Time 4
Cody	10%	35%	34%	21%
Justin	25%	25%	25%	25%
Ahmed	19%	24%	30%	27%
Mark	23%	31%	22%	24%

- Who was the divider?
- If playing honestly, what will player's declaration be?
- Find the final division.

9. A 6-foot sub valued at \$30 is divided among five players (P_1, P_2, P_3, P_4, P_5) using the last-diminisher method. The players play in a fixed order, with P_1 first, P_2 second, and so on. In round 1, P_1 makes the first cut and makes a claim on a piece. For each of the remaining players, the value of the *current* claimed piece at the time it is their turn is given in the following table:

	P_2	P_3	P_4	P_5
Value of the current claimed piece	\$6.00	\$8.00	\$7.00	\$6.50

- Which player gets his or her share at the end of round 1?
- What is the value of the share to the player receiving it?
- How would your answer change if the values were:

	P_2	P_3	P_4	P_5
Value of the current claimed piece	\$6.00	\$8.00	\$7.00	\$4.50

10. A huge collection of low-value baseball cards appraised at \$100 is being divided by 5 kids (P_1, P_2, P_3, P_4, P_5) using the last-diminisher method. The players play in a fixed order, with P_1 first, P_2 second, and so on. In round 1, P_1 makes the first selection and makes a claim on a pile of cards. For each of the remaining players, the value of the *current* pile of cards at the time it is their turn is given in the following table:

	P_2	P_3	P_4	P_5
Value of the current pile of cards	\$15.00	\$22.00	\$18.00	\$19.00

- Which player gets his or her share at the end of round 1?
- What is the value of the share to the player receiving it?
- How would your answer change if the values were:

	P_2	P_3	P_4	P_5
Value of the current pile of cards	\$15.00	\$22.00	\$18.00	\$21.00

11. Four heirs (A, B, C, and D) must fairly divide an estate consisting of two items - a desk and a vanity - using the method of sealed bids. The players' bids (in dollars) are:

	A	B	C	D
Desk	320	240	300	260
Vanity	220	140	200	180

- What is A's fair share?
- Find the initial allocation.
- Find the final allocation.

12. Three heirs (A, B, C) must fairly divide an estate consisting of three items - a house, a car, and a coin collection - using the method of sealed bids. The players' bids (in dollars) are:

	A	B	C
House	180,000	210,000	220,000
Car	12,000	10,000	8,000
Coins	3,000	6,000	2,000

- What is A's fair share?
 - Find the initial allocation.
 - Find the final allocation.
13. As part of an inheritance, four children, Abby, Ben and Carla, are dividing four vehicles using Sealed Bids. Their bids (in thousands of dollars) for each item is shown below. Find the final allocation.

	Abby	Ben	Carla
Motorcycle	10	9	8
Car	10	11	9
Tractor	4	1	2
Boat	7	6	4

14. As part of an inheritance, four children, Abby, Ben, Carla, and Dan, are dividing four vehicles using Sealed Bids. Their bids (in thousands of dollars) for each item is shown below. Find the final allocation.

	Abby	Ben	Carla	Dan
Motorcycle	6	7	11	8
Car	8	13	10	11
Tractor	3	1	5	4
Boat	7	6	3	8

15. After living together for a year, Sasha and Megan have decided to go their separate ways. They have several items they bought together to divide, as well as some moving-out chores. The values each place are shown below. Find the final allocation.

	Sasha	Megan
Couch	120	80
TV	200	250
Stereo	40	50
Detail cleaning	-40	-60
Cleaning carpets	-50	-40

16. After living together for a year, Emily, Kayla, and Kendra have decided to go their separate ways. They have several items they bought together to divide, as well as some moving-out chores. The values each place are shown below. Find the final allocation.

	Emily	Kayla	Kendra
Dishes	20	30	40
Vacuum cleaner	100	120	80
Dining table	100	80	130
Detail cleaning	-70	-40	-50
Cleaning carpets	-30	-60	-50

Concepts

Exploration

17. This question explores how bidding dishonestly can end up hurting the cheater. Four partners are dividing a million-dollar property using the lone-divider method. Using a map, Danny divides the property into four parcels s_1 , s_2 , s_3 , and s_4 . The following table shows the value of the four parcels in the eyes of each partner (in thousands of dollars):

	s_1	s_2	s_3	s_4
Danny	\$250	\$250	\$250	\$250
Brianna	\$460	\$180	\$200	\$160
Carlos	\$260	\$310	\$220	\$210
Greedy	\$330	\$300	\$270	\$100

- Assuming all players bid honestly, which piece will Greedy receive?
 - Assume Brianna and Carlos bid honestly, but Greedy decides to bid only for s_1 , figuring that doing so will get him s_1 . In this case there is a standoff between Brianna and Greedy. Since Danny and Carlos are not part of the standoff, they can receive their fair shares. Suppose Danny gets s_3 and Carlos gets s_2 , and the remaining pieces are put back together and Brianna and Greedy will split them using the basic divider-chooser method. If Greedy gets selected to be the divider, what will be the value of the piece he receives?
 - Extension: Create a Sealed Bids scenario that shows that sometimes a player can successfully cheat and increase the value they receive by increasing their bid on an item, but if they increase it too much, they could end up receiving less than their fair share.
18. Explain why divider-chooser method with two players will always result in an envy-free division.
19. Will the lone divider method always result in an envy-free division? If not, will it ever result in an envy-free division?

20. The Selfridge-Conway method is an envy-free division method for three players. Research how the method works and prepare a demonstration for the class.
21. Suppose that two people are dividing a \$12 pizza that is half pepperoni, half cheese. Steve likes both equally, but Maria likes cheese twice as much as pepperoni. As divider, Steve divides the pizza so that one piece is $\frac{1}{3}$ cheese and $\frac{2}{3}$ pepperoni, and the second piece is $\frac{1}{3}$ pepperoni and $\frac{2}{3}$ cheese.
- Describe the value of each piece to each player
 - Since the value to each player is not the same, this division is not equitable. Find a division that would be equitable. Is it still envy-free?
 - The original division is not Pareto optimal. To show this, find another division that would increase the value to one player without decreasing the value to the other player. Is this division still envy-free?
 - Would it be possible with this set of preferences to find a division that is both equitable and Pareto optimal? If so, find it. If not, explain why.
22. Is the Sealed Bids method Pareto optimal when used with two players? If not, can you adjust the method to be so?
23. Is the Sealed Bids method envy-free when used with two players? If not, can you adjust the method to be so?
24. Is the Sealed Bids method equitable when used with two players? If not, can you adjust the method to be so?
25. All the problems we have looked at in this chapter have assumed that all participants receive an equal share of what is being divided. Often, this does not occur in real life. Suppose Fred and Maria are going to divide a cake using the divider-chooser method. However, Fred is only entitled to 30% of the cake, and Maria is entitled to 70% of the cake (maybe it was a \$10 cake, and Fred put in \$3 and Maria put in \$7). Adapt the divider-choose method to allow them to divide the cake fairly.

Assume (as we have throughout this chapter) that different parts of the cake may have different values to Fred and Maria, and that they don't communicate their preferences/values with each other. Your goal is to come up with a method of fair division, meaning that although the participants may not receive equal shares, they should be *guaranteed* their fair share. Your method needs to be designed so that each person will always be guaranteed a share that they value as being worth at least as much as they're entitled to.

The last few questions will be based on the Adjusted Winner method, described here:

For discrete division between two players, there is a method called Adjusted Winner that produces an outcome that is always equitable, envy-free, and Pareto optimal. It does, however, require that items can be split or shared. The method works like this:

- 1) Each player gets 100 points that they assign to the items to be divided based on their relative worth.
- 2) In the initial allocation, each item is given to the party that assigned it more points. If there were any items with both parties assigned the same number of points, they'd go to the person with the fewest points so far.
- 3) If the assigned point values are not equal, then begin transferring items from the person with more points to the person with fewer points. Start with the items that have the smallest point ratio, calculated as (higher valuation/lower valuation).
- 4) If transferring an entire item would move too many points, only transfer a portion of the item.

Example: A couple is attempting to settle a contentious divorce¹. They assign their 100 points to the issues in contention:

	Mike	Carol
Custody of children	25	65
Alimony	60	25
House	15	10

In the initial allocation, Mike gets his way on alimony and house, and Carol gets custody of the children. In the initial allocation, Mike has 75 points and Carol has 65 points. To decide what to transfer, we calculate the point ratios.

	Mike	Carol	Point ratio
Custody of children	25	65	$65/25 = 2.6$
Alimony	60	25	$60/25 = 2.4$
House	15	10	$15/10 = 1.5$

Since the house has the smallest point ratio, the house will be the item we work with first. Since transferring the entire house would give Carol too many points, we instead need to transfer some fraction, p , of the house to that Carol and Mike will end up with the same point values. If Carol receives a fraction p of the house, then Mike will give up $(1-p)$ of the house. The value Carol will receive is $10p$: the fraction p of the 10 points Carol values the house at. The value Mike will get is $15(1-p)$. We set their point totals equal to solve for p :

$$65 + 10p = 60 + 15(1-p)$$

$$65 + 10p = 60 + 15 - 15p$$

$$25p = 10$$

$$p = 10/25 = 0.4 = 40\%. \text{ So Carol should receive 40\% of the house.}$$

¹ From *Negotiating to Settlement in Divorce*, 1987

26. Apply the Adjusted Winner method to settle a divorce where the couples have assigned the point values below

	Sandra	Kenny
Home	20	30
Summer home	15	10
Retirement account	50	40
Investments	10	10
Other	5	10

27. In 1974, the United States and Panama negotiated over US involvement and interests in the Panama Canal. Suppose that these were the issues and point values assigned by each side². Apply the Adjusted Winner method.

	United States	Panama
US defense rights	22	9
Use rights	22	15
Land and water	15	15
Expansion rights	14	3
Duration	11	15
Expansion routes	6	5
Jurisdiction	2	7
US military rights	2	7
Defense role of Panama	2	13

² Taken from *The Art and Science of Negotiation*, 1982

Graph Theory and Network Flows

In the modern world, planning efficient routes is essential for business and industry, with applications as varied as product distribution, laying new fiber optic lines for broadband internet, and suggesting new friends within social network websites like Facebook.

This field of mathematics started nearly 300 years ago as a look into a mathematical puzzle (we'll look at it in a bit). The field has exploded in importance in the last century, both because of the growing complexity of business in a global economy and because of the computational power that computers have provided us.

Graphs

Drawing Graphs

Here is a portion of a housing development from Missoula, Montana¹. As part of her job, the development's lawn inspector has to walk down every street in the development making sure homeowners' landscaping conforms to the community requirements.

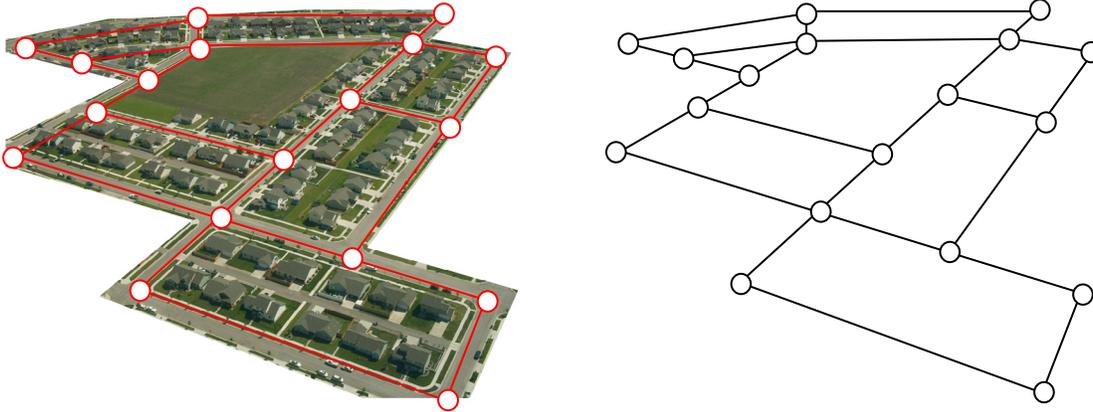


Naturally, she wants to minimize the amount of walking she has to do. Is it possible for her to walk down every street in this development without having to do any backtracking?

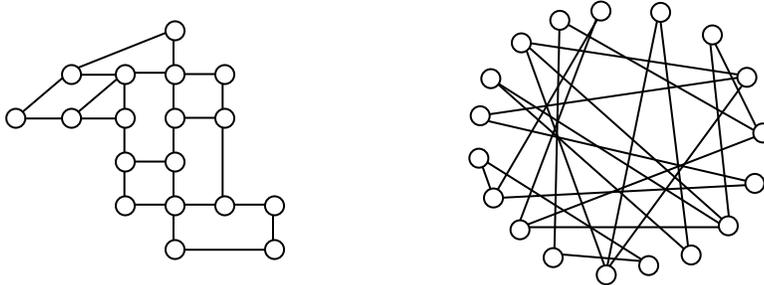
¹ Sam Beebe. <http://www.flickr.com/photos/sbeebe/2850476641/>

While you might be able to answer that question just by looking at the picture for a while, it would be ideal to be able to answer the question for any picture regardless of its complexity.

To do that, we first need to simplify the picture into a form that is easier to work with. We can do that by drawing a simple line for each street. Where streets intersect, we will place a dot.



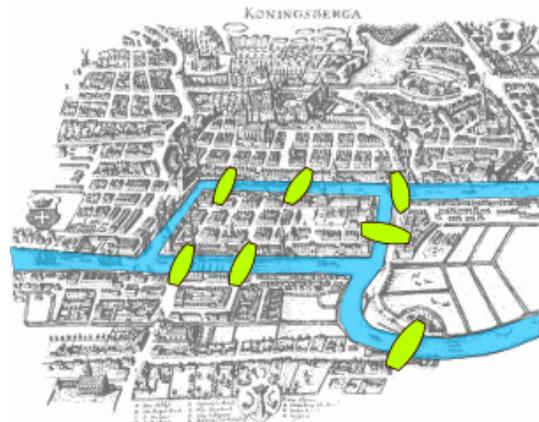
This type of simplified picture is called a **graph**. A graph consists of a set of dots, called **vertices**, and a set of **edges** connecting pairs of vertices. While we drew our original graph to correspond with the picture we had, there is nothing particularly important about the layout when we analyze a graph. Both of these graphs are equivalent to the one drawn above.



Example: Back in the 18th century in the Prussian city of Königsberg, a river ran through the city and seven bridges crossed the forks of the river. The river and the bridges are highlighted in the picture to the right².

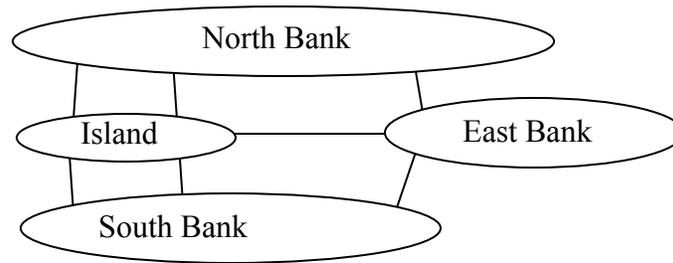
As a weekend amusement, townsfolk would see if they could find a route that would take them across every bridge once and return them to where they started.

Leonard Euler (pronounced OY-lur), one of the most prolific mathematicians ever, looked at this problem in 1735, laying the foundations for graph theory as a field in mathematics.

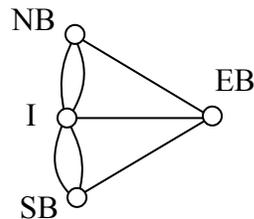


² Bogdan Giușcă. http://en.wikipedia.org/wiki/File:Königsberg_bridges.png

To analyze this problem, Euler introduced edges representing the bridges:



Since it is not relevant to the question of bridge crossing how large each land mass is, they can be shrunk down to single vertices representing each location:



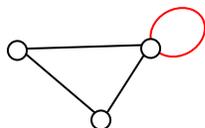
Notice that in this graph there are *two* edges connecting the north bank and island, corresponding to the two bridges in the original drawing. Depending upon the interpretation of edges and vertices appropriate to a scenario, it is entirely possible and reasonable to have more than one edge connecting two vertices.

Definitions

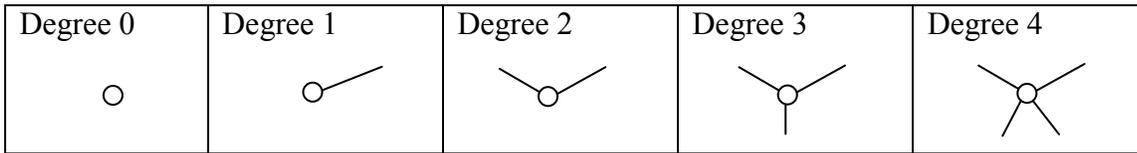
Vertex. A vertex is a dot in the graph where edges meet. A vertex could represent an intersection of streets, a land mass, or a general location, like “work” or “school”. Note that vertices only occur when a dot is explicitly placed, not whenever two edges cross. Imagine a freeway overpass – the freeway and side street cross, but it is not possible to change from the side street to the freeway at that point, so there is no intersection and no vertex would be placed.

Edges. Edges connect pairs of vertices. An edge can represent a physical connection between locations, like a street, or simply that a route connecting the two locations exists, like an airline flight.

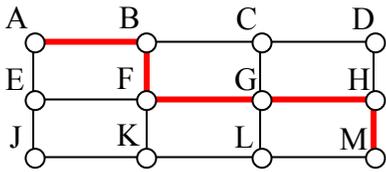
Loop. A loop is a special type of edge that connects a vertex to itself. Loops are not used much in street network graphs.



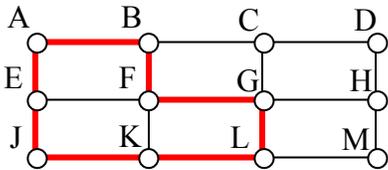
Degree of a vertex. The degree of a vertex is the number of edges meeting at that vertex. It is possible for a vertex to have a degree of zero or larger.



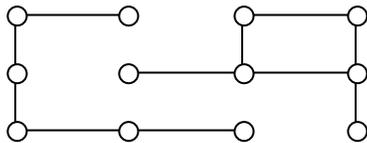
Path. A path is a sequence of vertices using the edges. Usually we are interested in a path between two vertices. For example, a path from vertex A to vertex M is shown below. It is one of many possible paths in this graph.



Circuit. A circuit is a path that begins and ends at the same vertex. A circuit starting and ending at vertex A is shown below.



Connected. A graph is connected if there is a path from any vertex to any other vertex. Every graph drawn so far has been connected. The graph below is **disconnected**; there is no way to get from the vertices on the left to the vertices on the right.



Weights. Depending upon the problem being solved, sometimes weights are assigned to the edges. The weights could represent the distance between two locations, the travel time, or the travel cost. It is important to note that the distance between vertices in a graph does not necessarily correspond to the weight of an edge.

Shortest Path

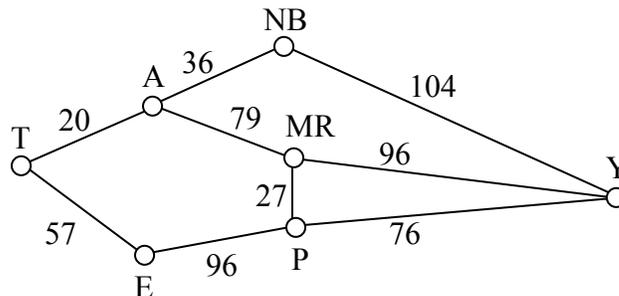
When you visit a website like Google Maps or MapQuest and ask for directions from home to your Aunt's house in Pasadena, you are usually looking for a shortest path between the two locations. These computer applications use representations of the street maps as graphs, with estimated driving times as edge weights.

While often it is possible to find a shortest path on a small graph by guess-and-check, our goal in this chapter is to develop methods to solve complex problems in a systematic way by following **algorithms**. An algorithm is a step-by-step procedure for solving a problem. Dijkstra's (pronounced dike-s-tra) algorithm will find the shortest path between two vertices.

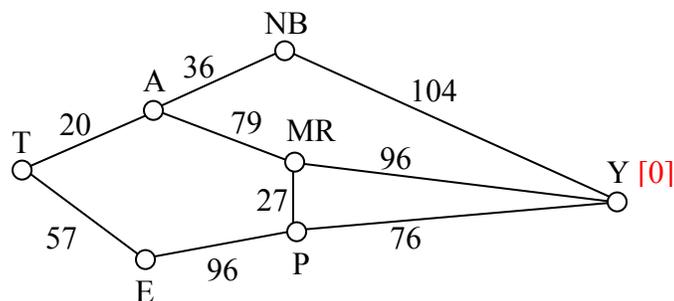
Dijkstra's Algorithm

- 1) Mark the ending vertex with a distance of zero. Designate this vertex as current.
- 2) Find all vertices leading to the current vertex. Calculate their distances to the end. Since we already know the distance the current vertex is from the end, this will just require adding the most recent edge. Don't record this distance if it is longer than a previously recorded distance.
- 3) Mark the current vertex as visited. We will never look at this vertex again.
- 4) Mark the vertex with the smallest distance as current, and repeat from step 2.

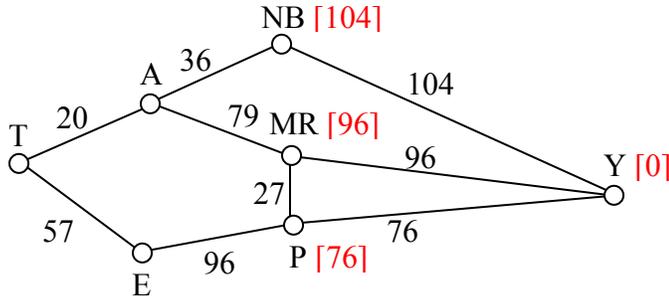
Example: Suppose you need to travel from Tacoma, WA (vertex T) to Yakima, WA (vertex Y). Looking at a map, it looks like driving through Auburn (A) then Mount Rainier (MR) might be shortest, but it's not totally clear since that road is probably slower than taking the major highway through North Bend (NB). A graph with travel times in minutes is shown below. An alternate route through Eatonville (E) and Packwood (P) is also shown.



Step 1: Mark the ending vertex with a distance of zero. The distances will be recorded in **[brackets]** after the vertex name.



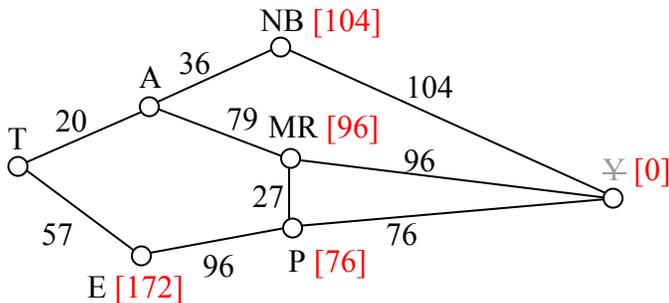
Step 2: For each vertex leading to Y, we calculate the distance to the end. For example, NB is a distance of 104 from the end, and MR is 96 from the end. Remember that distances in this case refer to the travel time in minutes.



Step 3 & 4: We mark Y as visited, and mark the vertex with the smallest recorded distance as current. At this point, P will be designated current. Back to step 2.

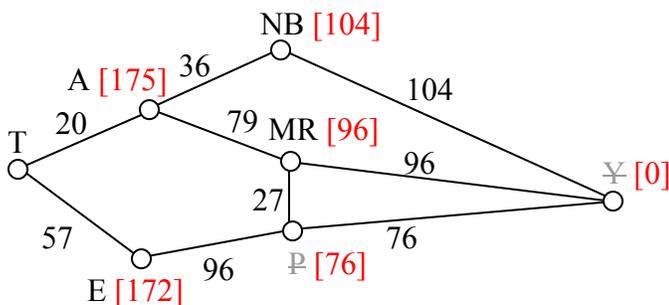
Step 2 (#2): For each vertex leading to P (and not leading to a visited vertex) we find the distance from the end. Since E is 96 minutes from P, and we've already calculated P is 76 minutes from Y, we can compute that E is $96+76 = 172$ minutes from Y.

If we make the same computation for MR, we'd calculate $76+27 = 103$. Since this is larger than the previously recorded distance from Y to MR, we will *not* replace it.



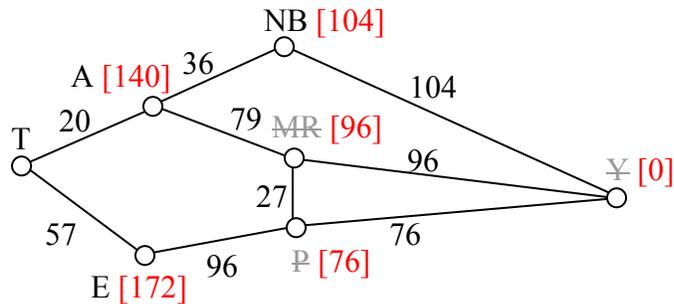
Step 3 & 4 (#2): We mark P as visited, and designate the vertex with the smallest recorded distance as current: MR. Back to step 2.

Step 2 (#3): For each vertex leading to MR (and not leading to a visited vertex) we find the distance to the end. The only vertex to be considered is A, since we've already visited Y and P. Adding MR's distance 96 to the length from A to MR gives the distance $96+79 = 175$ minutes from A to Y.



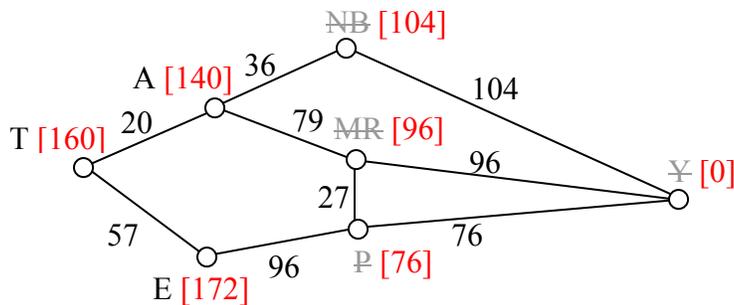
Step 3 & 4 (#3): We mark MR as visited, and designate the vertex with smallest recorded distance as current: NB. Back to step 2.

Step 2 (#4): For each vertex leading to NB, we find the distance to the end. We know the shortest distance from NB to Y is 104 and the distance from A to NB is 36, so the distance from A to Y through NB is $104+36 = 140$. Since this distance *is* shorter than the previously calculated distance from Y to A through MR, we replace it.



Step 3 & 4 (#4): We mark NB as visited, and designate A as current, since it now has the shortest distance.

Step 2 (#5): T is the only non-visited vertex leading to A, so we calculate the distance from T to Y through A: $20+140 = 160$ minutes.



Step 3 & 4 (#5): We mark A as visited, and designate E as current.

Step 2 (#6): The only non-visited vertex leading to E is T. Calculating the distance from T to Y through E, we compute $172+57 = 229$ minutes. Since this is longer than the existing marked time, we do not replace it.

Step 3 (#6): We mark E as visited. Since all vertices have been visited, we are done.

From this, we know that the shortest path from Tacoma to Yakima will take 160 minutes. Tracking which sequence of edges yielded 160 minutes, we see the shortest path is T-A-NB-Y.

Dijkstra's algorithm is an **optimal algorithm**, meaning that it always produces the actual shortest path, not just a path that is pretty short, provided one exists. This algorithm is also **efficient**, meaning that it can be implemented in a reasonable amount of time. Dijkstra's algorithm takes around V^2 calculations, where V is the number of vertices in a graph³. A graph with 100 vertices would take around 10,000 calculations. While that would be a lot to do by hand, it is not a lot for computer to handle. It is because of this efficiency that your car's GPS unit can compute driving directions in only a few seconds.

In contrast, an inefficient algorithm might try to list all possible paths then compute the length of each path. An inefficient algorithm could easily take 10^{25} calculations to compute the shortest path with only 25 vertices; that's a 1 with 25 zeros after it! To put that in perspective, the fastest computer in the world would still spend over 1000 years analyzing all those paths.

Example: A shipping company needs to route a package from Washington, DC to San Diego, CA. To minimize costs, the package will first be sent to their processing center in Baltimore, MD then sent as part of mass shipments between their various processing centers, ending up in their processing center in Bakersfield, CA. From there it will be delivered in a small truck to San Diego.

The travel times, in hours, between their processing centers are shown in the table below. Three hours has been added to each travel time for processing. Find the shortest path from Baltimore to Bakersfield.

	Baltimore	Denver	Dallas	Chicago	Atlanta	Bakersfield
Baltimore	*			15	14	
Denver		*		18	24	19
Dallas			*	18	15	25
Chicago	15	18	18	*	14	
Atlanta	14	24	15	14	*	
Bakersfield		19	25			*

While we could draw a graph, we can also work directly from the table.

Step 1: The ending vertex, Bakersfield, is marked as current.

Step 2: All cities connected to Bakersfield have their distances calculated which are Denver and Dallas; we'll mark those distances in the column headers.

Step 3 & 4: Mark Bakersfield as visited, and the Denver as current, since it is the vertex with the shortest distance.

³ It can be made to run faster through various optimizations to the implementation.

	Baltimore	Denver [19]	Dallas [25]	Chicago	Atlanta	Bakersfield [0]
Baltimore	*			15	14	
Denver		*		18	24	19
Dallas			*	18	15	25
Chicago	15	18	18	*	14	
Atlanta	14	24	15	14	*	
Bakersfield		19	25			*

Step 2 (#2): For cities connected to Denver, calculate distance to the end. For example, Chicago is 18 hours from Denver, and Denver is 19 hours from the end, the distance for Chicago to the end is $18+19 = 37$ (Chicago to Denver to Bakersfield)

Step 3 & 4 (#2): We mark Denver as visited and mark Dallas as current.

	Baltimore	Denver [19]	Dallas [25]	Chicago [37]	Atlanta	Bakersfield [0]
Baltimore	*			15	14	
Denver		*		18	24	19
Dallas			*	18	15	25
Chicago	15	18	18	*	14	
Atlanta	14	24	15	14	*	
Bakersfield		19	25			*

Step 2 (#3): For cities connected to Dallas, calculate the distance to the end. For Chicago, the distance from Chicago to Dallas is 18 and from Dallas to the end is 25, so the distance from Chicago to the end through Dallas would be $18+25 = 43$. Since this is longer than the currently marked distance for Chicago, we do not replace it. For Atlanta, we calculate $15+25 = 40$, and mark that distance.

Step 3 & 4 (#3): We mark Dallas as visited, and mark Chicago as current.

	Baltimore	Denver [19]	Dallas [25]	Chicago [37]	Atlanta [40]	Bakersfield [0]
Baltimore	*			15	14	
Denver		*		18	24	19
Dallas			*	18	15	25
Chicago	15	18	18	*	14	
Atlanta	14	24	15	14	*	
Bakersfield		19	25			*

Step 2 (#4): Baltimore and Atlanta are the only non-visited cities connected to Chicago. For Baltimore, we calculate $15+37 = 52$ and mark that distance. For Atlanta, we calculate $14+37 = 51$. Since this is longer than the existing distance of 40 for Atlanta, we do not replace that distance.

Step 3 & 4 (#4): Mark Chicago as visited and Atlanta as current.

	Baltimore [52]	Denver [19]	Dallas [25]	Chicago [37]	Atlanta [40]	Bakersfield [0]
Baltimore	*			15	14	
Denver		*		18	24	19
Dallas			*	18	15	25
Chicago	15	18	18	*	14	
Atlanta	14	24	15	14	*	
Bakersfield		19	25			*

Step 2 (#5): The distance from Atlanta to Baltimore is 14. Adding that to the distance already calculated for Atlanta gives a total distance of $14+40 = 54$ hours from Baltimore to Bakersfield through Atlanta. Since this is larger than the currently calculated distance, we do not replace the distance for Baltimore.

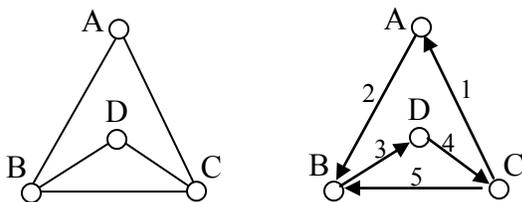
Step 3 & 4 (#5): We mark Atlanta as visited. All cities have been visited and we are done. The shortest route from Baltimore to Bakersfield will take 52 hours, and will route through Chicago and Denver.

Euler Circuits and the Chinese Postman Problem

In the first section, we created a graph of the Königsberg bridges and asked whether it was possible to walk across every bridge once. Because Euler first studied this question, they are named after him.

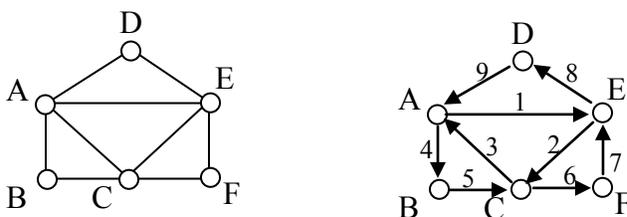
An **Euler path** is a path that uses every edge in a graph with no repeats. Being a path, it does not have to return to the starting vertex.

Example: In the graph shown below, there are several Euler paths. One such path is CABDCB. The path is shown in arrows to the right, with the order of edges numbered.



An **Euler circuit** is a circuit that uses every edge in a graph with no repeats. Being a circuit, it must start and end at the same vertex.

Example: The graph below has several possible Euler circuits. Here's a couple, starting and ending at vertex A: ADEACEFCBA and AECABCFEDA. The second is shown in arrows.



Look back at the example used for Euler paths – does that graph have an Euler circuit? A few tries will tell you no; that graph does not have an Euler circuit. When we were working with shortest paths, we were interested in the optimal path. With Euler paths and circuits, we're primarily interested in whether an Euler path or circuit *exists*.

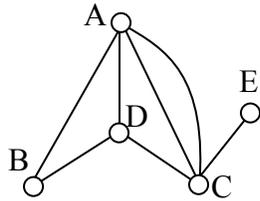
Why do we care if an Euler circuit exists? Think back to our housing development lawn inspector from the beginning of the chapter. The lawn inspector is interested in walking as little as possible. The ideal situation would be a circuit that covers every street with no repeats. That's an Euler circuit! Luckily, Euler solved the question of whether or not an Euler path or circuit will exist.

Euler's Path and Circuit Theorems

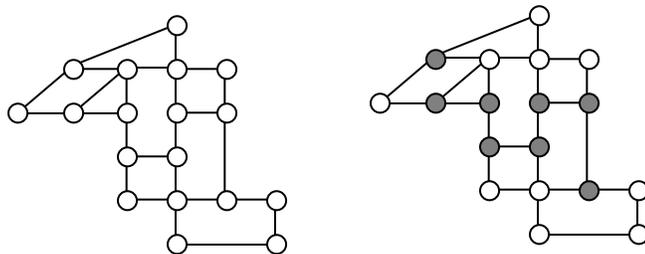
A graph will contain an Euler path if it contains at most two vertices of odd degree.

A graph will contain an Euler circuit if all vertices have even degree

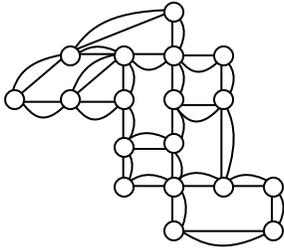
Example: In the graph below, vertices A and C have degree 4, since there are 4 edges leading into each vertex. B is degree 2, D is degree 3, and E is degree 1. This graph contains two vertices with odd degree (D and E) and three vertices with even degree (A, B, and C), so Euler's theorems tell us this graph has an Euler path, but not an Euler circuit.



Example: Is there an Euler circuit on the housing development lawn inspector graph we created earlier in the chapter? All the highlighted vertices have odd degree. Since there are more than two vertices with odd degree, there are no Euler paths or Euler circuits on this graph. Unfortunately our lawn inspector will need to do some backtracking.



Example: When it snows in the same housing development, the snowplow has to plow both sides of every street. For simplicity, we'll assume the plow is out early enough that it can ignore traffic laws and drive down either side of the street in either direction. This can be visualized in the graph by drawing two edges for each street, representing the two sides of the street.

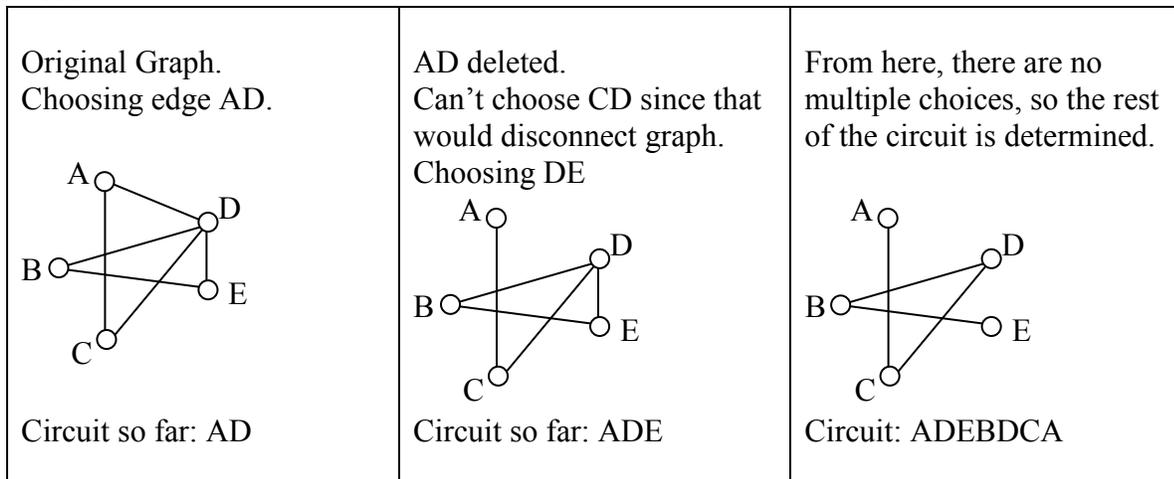


Notice that every vertex in this graph has even degree, so this graph does have an Euler circuit. Now how do we find one? While it usually is possible to find an Euler circuit just by pulling out your pencil and trying to find one, the more formal method is **Fleury's algorithm**.

Fleury's Algorithm

- 1) Start at any vertex if finding an Euler circuit. If finding an Euler path, start at one of the two vertices with odd degree.
- 2) Choose any edge leaving your current vertex, provided deleting that edge will not separate the graph into two disconnected sets of edges.
- 3) Add that edge to your circuit, and delete it from the graph.
- 4) Continue until you're done.

Example: Let's find an Euler Circuit on this graph using Fleury's algorithm, starting at vertex A.



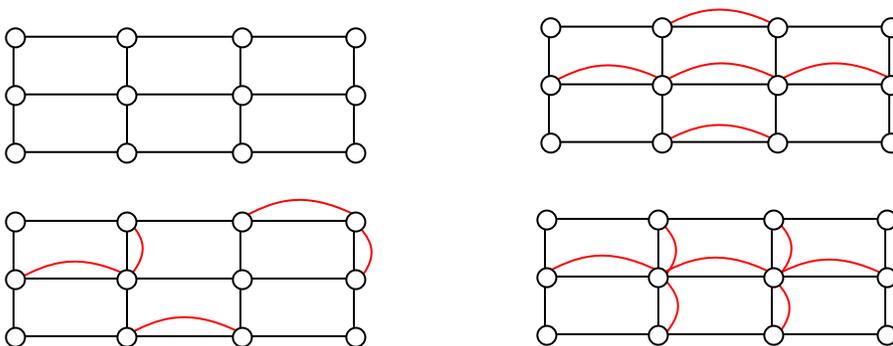
Eulerization and the Chinese Postman Problem

Not every graph has an Euler path or circuit, yet our lawn inspector still needs to do her inspections. Her goal is to minimize the amount of walking she has to do. In order to do that, she will have to duplicate some edges in the graph until an Euler circuit exists.

Eulerization is the process of adding edges to a graph to create an Euler circuit on a graph. To eulerize a graph, edges are duplicated to connect pairs of vertices with odd degree. Connecting two odd degree vertices increases the degree of each, giving them both even degree. When two odd degree vertices are not directly connected, we can duplicate all edges in a path connecting the two.

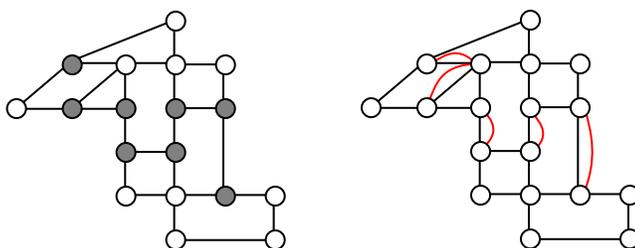
Note that we can only duplicate edges, not create edges where there wasn't one before; that's the difference between driving down a road twice and installing a new road!

Example: For the rectangular graph shown, three possible eulerizations are shown. Notice in each of these cases the vertices that started with odd degrees have even degrees after eulerization, allowing for an Euler circuit.



In the example above, you'll notice that the last eulerization required duplicating seven edges, while the first two only required duplicating five edges. If we were eulerizing the graph to find a walking path, we would want the eulerization with minimal duplications. If the edges had weights representing distances or costs, then we would want to select the eulerization with the minimal total added weight.

Example: For our lawn inspector route, the vertices with odd degree are shown highlighted. We can eulerize this graph by duplicating only four edges – the least possible with eight vertices to address. Without weights we can't be certain this is the eulerization that minimizes walking distance, but it looks pretty good.



The problem of finding the optimal eulerization is called the Chinese Postman Problem, a name given by an American in honor of the Chinese mathematician Mei-Ko Kwan who first studied the problem in 1962 while trying to find optimal delivery routes for postal carriers. This is important in determining efficient routes for garbage trucks, school buses, parking meter checkers, street sweepers, and more.

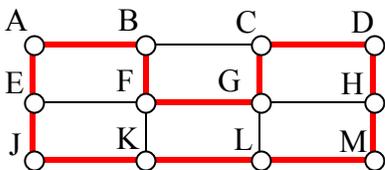
Unfortunately, algorithms to solve this problem are fairly complex. Some simpler cases are considered in the exercises.

Hamiltonian Circuits and the Traveling Salesman Problem

In the last section, we considered optimizing a walking route for a postal carrier. How is this different than the requirements of a package delivery driver? While the postal carrier needed to walk down every street (edge) to deliver the mail, the package delivery driver instead needs to visit every one of a set of delivery locations. Instead of looking for a circuit that covers every edge once, the package deliverer is interested in a circuit that visits every vertex once.

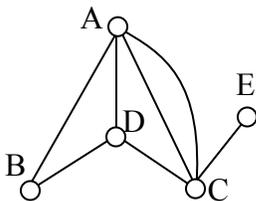
A **Hamiltonian circuit** is a circuit that visits every vertex once with no repeats. Being a circuit, it must start and end at the same vertex. A **Hamiltonian path** also visits every vertex once with no repeats, but does not have to start and end at the same vertex. Hamiltonian circuits are named for William Rowan Hamilton who studied them in the 1800's.

Example: One Hamiltonian circuit is shown on the graph below. There are several other Hamiltonian circuits possible on this graph. Notice that the circuit only has to visit every vertex once; it does not need to use every edge. This circuit could be notated by the sequence of vertices visited, starting and ending at the same vertex: ABFGCDHMLKJEA. Notice that the same circuit could be written in reverse order, or starting and ending at a different vertex.



Unlike with Euler circuits, there is no nice theorem that allows us to instantly determine whether or not a Hamiltonian circuit exists for all graphs.⁴

Example: Does a Hamiltonian path or circuit exist on the graph below?

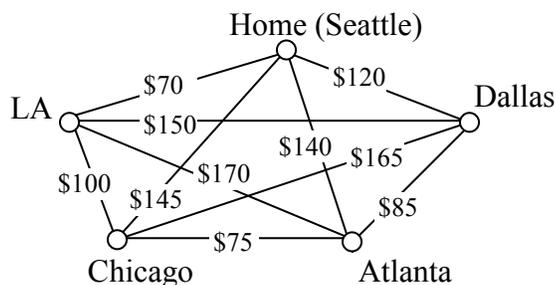


⁴ There are some theorems that can be used in specific circumstances, such as Dirac's theorem, which says that a Hamiltonian circuit must exist on a graph with n vertices if each vertex has degree $n/2$ or greater.

We can see that once we travel to vertex E there is no way to leave without returning to C, so there is no possibility of a Hamiltonian circuit. If we start at vertex E we can find several Hamiltonian paths, such as ECDAB and ECABD.

With Hamiltonian circuits, our focus will not be on existence, but on the question of optimization; given a graph where the edges have weights, can we find the optimal Hamiltonian circuit; the one with lowest total weight.

This problem is called the **Traveling salesman problem (TSP)** because the question can be framed like this: Suppose a salesman needs to give sales pitches in four cities. He looks up the airfares between each city⁵, and puts the costs in a graph. In what order should he travel to visit each city once then return home with the lowest cost?

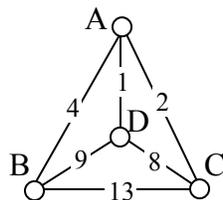


To answer this question, we will consider some possible approaches.

Brute Force Algorithm (a.k.a. exhaustive search)

- 1) List all possible Hamiltonian circuits
- 2) Find the length of each circuit by adding the edge weights
- 3) Select the circuit with minimal total weight.

Example: Apply the Brute force algorithm to find the minimum cost Hamiltonian circuit on the graph below.



To apply the Brute force algorithm, we list all possible Hamiltonian circuits and calculate their weight:

Circuit	Weight
ABCD	$4+13+8+1 = 26$
ABDC	$4+9+8+2 = 23$
ACBD	$2+13+9+1 = 25$

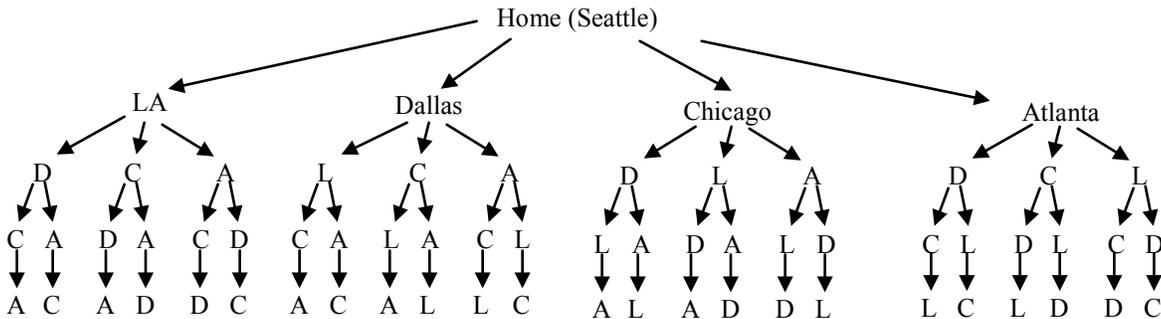
⁵ Travelocity.com. Retrieved 5/14/2009 for travel on 8/12/2009.

Note: These are the unique circuits on this graph. All other possible circuits are the reverse of the listed ones or start at a different vertex, but result in the same weights.

From this we can see that the second circuit, ABDCA, is the optimal circuit.

The Brute force algorithm is optimal; it will always produce the Hamiltonian circuit with minimum weight. Is it efficient? To answer that question, we need to consider how many Hamiltonian circuits a graph could have. For simplicity, let's look at the worst-case possibility, where every vertex is connected to every other vertex. This is called a **complete graph**.

Suppose we had a complete graph with five vertices like the air travel graph above. From Seattle there are four cities we can visit first. From each of those, there are three choices. From each of those cities, there are two possible cities to visit next. There is then only one choice for the last city before returning home. This can be shown visually:



Counting the number of routes, we can see there are $4 \cdot 3 \cdot 2 \cdot 1 = 24$ routes. For six cities there would be $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ routes. For N cities there would be $(n-1)! = (n-1)(n-2)(n-3)\cdots 3 \cdot 2 \cdot 1$ routes. The exclamation symbol is read “factorial” and is shorthand for the product shown.

Example: A complete graph with 8 vertices would have $(8-1)! = 7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$ possible Hamiltonian circuits. Half of the circuits are duplicates of other circuits but in reverse order, leaving 2520 unique routes. While this is a lot, it doesn't seem unreasonably huge. But consider what happens as the number of cities increase:

Cities	Unique Hamiltonian Circuits
9	$8!/2 = 20,160$
10	$9!/2 = 181,440$
11	$10!/2 = 1,814,400$
15	$14!/2 = 43,589,145,600$
20	$19!/2 = 60,822,550,204,416,000$

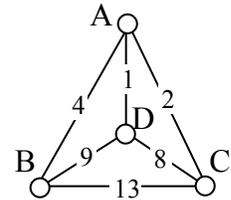
As you can see the number of circuits is growing extremely quickly. If a computer looked at one billion circuits a second, it would still take almost two years to examine all the possible circuits with only 20 cities! Certainly Brute Force is **not** an efficient algorithm.

Unfortunately, no one has yet found an efficient *and* optimal algorithm to solve the TSP. Since it is not practical to use brute force to solve the problem, we turn instead to heuristic algorithms; efficient algorithms that give approximate solutions. In other words, heuristic algorithms are fast, but may or may not produce the optimal circuit.

Nearest Neighbor Algorithm (NNA)

- 1) Select a starting point.
- 2) Move to the nearest unvisited vertex (the edge with smallest weight).
- 3) Repeat until the circuit is complete.

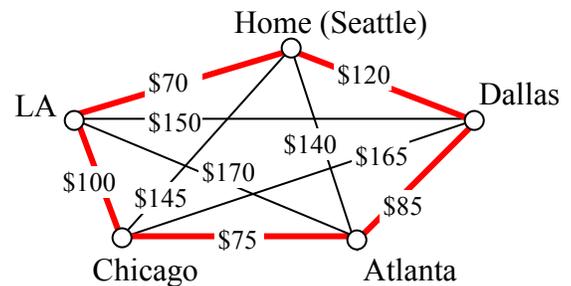
Example: Consider our earlier graph, shown to the right. Starting at vertex A, the nearest neighbor is vertex D with a weight of 1. From D, the nearest neighbor is C, with a weight of 8. From C, our only option is to move to vertex B, the only unvisited vertex, with a cost of 13. From B we return to A with a weight of 4. The resulting circuit is ADCBA with a total weight of $1+8+13+4 = 26$.



We ended up finding the worst circuit in the graph! What happened? Unfortunately, while it is very easy to implement, the NNA is a **greedy** algorithm, meaning it only looks at the immediate decision without considering the consequences in the future. In this case, following the edge AD forced us to use the very expensive edge BC later.

Example: Consider again our salesman. Starting in Seattle, the nearest neighbor (cheapest flight) is to LA, at a cost of \$70. From there:

LA to Chicago: \$100
 Chicago to Atlanta: \$75
 Atlanta to Dallas: \$85
 Dallas to Seattle: \$120
 Total cost: \$450



In this case, nearest neighbor did find the optimal circuit.

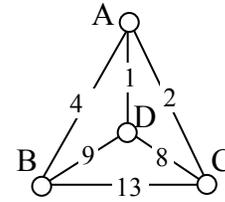
Going back to our first example, how could we improve the outcome? One option would be to redo the nearest neighbor algorithm with a different starting point to see if the result changed. Since nearest neighbor is so fast, doing it several times isn't a big deal.

Repeated Nearest Neighbor Algorithm (RNNA)

- 1) Do the Nearest Neighbor Algorithm starting at each vertex
- 2) Choose the circuit produced with minimal total weight

Example: Starting at vertex A resulted in a circuit with weight 26.

Starting at vertex B, the nearest neighbor circuit is BADCB with a weight of $4+1+8+13 = 26$. This is the same circuit we found starting at vertex A. No better.



Starting at vertex C, the nearest neighbor circuit is CADBC with a weight of $2+1+9+13 = 25$. Better!

Starting at vertex D, the nearest neighbor circuit is DACBA, the same circuit we found starting at C.

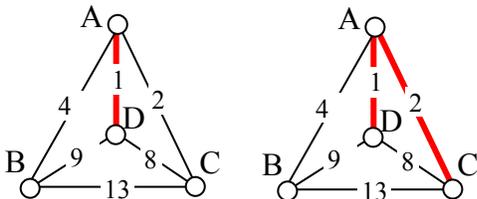
So the RNNA was able to produce a slightly better circuit with a weight of 25, but still not the optimal circuit. Notice that even though we found the circuit by starting at vertex C, we could still write the circuit starting at A: ADBCA or ACBDA.

Unfortunately, the RNNA is still greedy and will produce very bad results for some graphs. As an alternative, our next approach will step back and look at the “big picture” – it will select first the edges that are shortest, and then fill in the gaps.

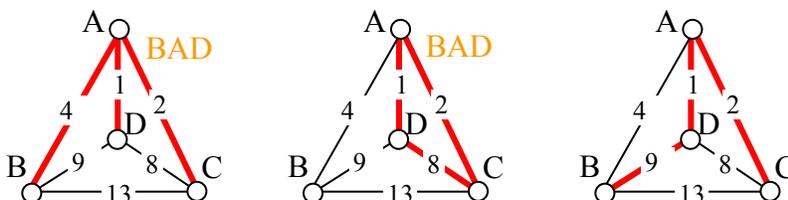
Sorted Edges Algorithm (a.k.a. Cheapest Link Algorithm)

- 1) Select the cheapest unused edge in the graph; highlight it.
- 2) Repeat step 1, adding the cheapest unused edge to the graph, unless:
 - a. adding the edge would create a circuit that doesn't contain all vertices, or
 - b. adding the edge would give a vertex degree 3.
- 3) Repeat until a circuit containing all vertices is formed.

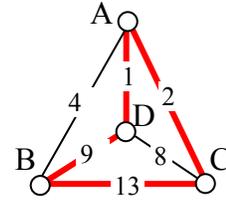
Example: Using the four vertex graph from above, the cheapest edge is AD, with a cost of 1. We highlight that edge. The next shortest edge is AC, with a weight of 2, so we highlight that edge.



For the third edge, we'd like to add AB, but that would give vertex A degree 3, which shouldn't happen in a Hamiltonian circuit. The next shortest edge is CD, but that edge would create a circuit ACDA that does not include vertex B, so we reject that edge. The next shortest edge is BD, so we add that edge to the graph.



We then add the last edge to complete the circuit: ACBDA with weight 25. Notice that the algorithm did not produce the optimal circuit in this case (circuit ACDBA with weight 23). While the Sorted Edge algorithm overcomes some of the shortcomings of NNA, it is still only a heuristic algorithm, and does not guarantee the optimal circuit.



Example: Your teacher's band, *Derivative Work*, is doing a bar tour in Oregon. The driving distances are shown below. Plan an efficient route for your teacher to visit all the cities and return to the starting location. Use NNA starting at Portland, and Sorted Edges.

	Ashland	Astoria	Bend	Corvallis	Crater Lake	Eugene	Newport	Portland	Salem	Seaside
Ashland	-	374	200	223	108	178	252	285	240	356
Astoria	374	-	255	166	433	199	135	95	136	17
Bend	200	255	-	128	277	128	180	160	131	247
Corvallis	223	166	128	-	430	47	52	84	40	155
Crater Lake	108	433	277	430	-	453	478	344	389	423
Eugene	178	199	128	47	453	-	91	110	64	181
Newport	252	135	180	52	478	91	-	114	83	117
Portland	285	95	160	84	344	110	114	-	47	78
Salem	240	136	131	40	389	64	83	47	-	118
Seaside	356	17	247	155	423	181	117	78	118	-

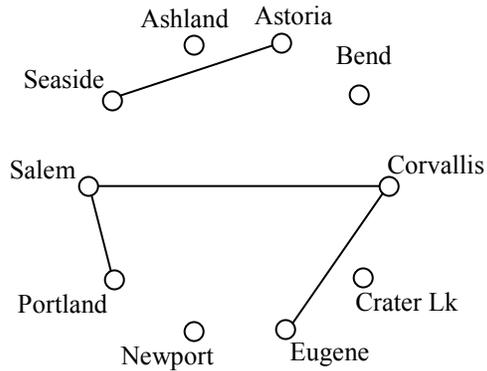
Using NNA with a large number of cities, you might find it helpful to mark off the cities as they're visited to keep from accidentally visiting them again. Looking in the row for Portland, the smallest distance is 47, to Salem. Following that idea, our circuit will be:

Portland to Salem	47
Salem to Corvallis	40
Corvallis to Eugene	47
Eugene to Newport	91
Newport to Seaside	117
Seaside to Astoria	17
Astoria to Bend	255
Bend to Ashland	200
Ashland to Crater Lake	108
Crater Lake to Portland	344
Total trip length:	1266 miles

Using Sorted Edges, you might find it helpful to draw an empty graph, perhaps by drawing vertices in a circular pattern. Adding edges to the graph as you select them will help you visualize any circuits or vertices with degree 3.

- Seaside to Astoria 17 miles
- Corvallis to Salem 40 miles
- Portland to Salem 47 miles
- Corvallis to Eugene 47 miles

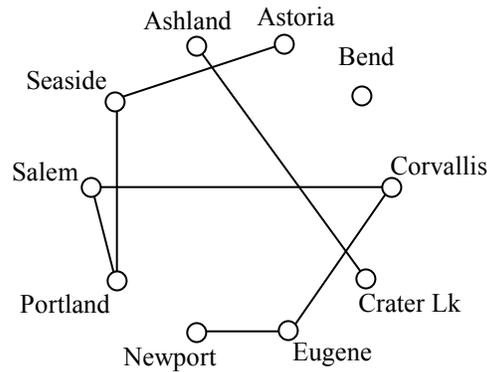
The graph after adding these edges is shown to the right. The next shortest edge is from Corvallis to Newport at 52 miles, but adding that edge would give Corvallis degree 3.



Continuing on, we can skip over any edge pair that contains Salem or Corvallis, since they both already have degree 2.

- Portland to Seaside 78 miles
- Eugene to Newport 91 miles
- Portland to Astoria (reject – closes circuit)
- Ashland to Crater Lk 108 miles

The graph after adding these edges is shown to the right. At this point, we can skip over any edge pair that contains Salem, Seaside, Eugene, Portland, or Corvallis since they already have degree 2.

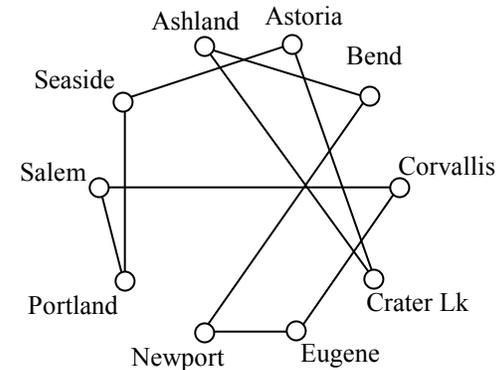


- Newport to Astoria (reject – closes circuit)
- Newport to Bend 180 miles
- Bend to Ashland 200 miles

At this point the only way to complete the circuit is to add:
Crater Lk to Astoria 433 miles

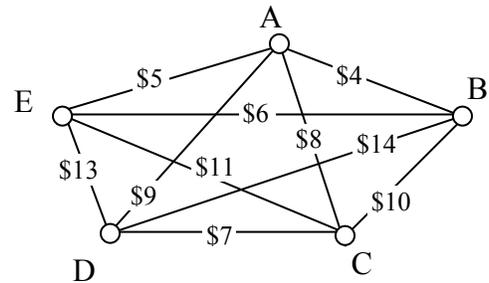
Total trip length: 1241 miles

Neither algorithm produced the optimal route. The following route can make the tour in 1069 miles:
Ashland, Crater Lake, Bend, Salem, Portland, Astoria, Seaside, Newport, Corvallis, Eugene, Ashland



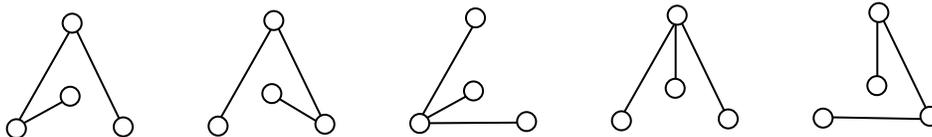
Spanning Trees

A company requires reliable internet and phone connectivity between their five offices (named A, B, C, D, and E for simplicity) in New York, so they decide to lease dedicated lines from the phone company. The phone company will charge for each link made. The costs, in thousands of dollars per year, are shown in the graph.



In this case, we don't need to find a circuit, or even a path; all we need to do is make sure there is a way to get from any vertex to any other vertex.

A **spanning tree** is a connected graph using all vertices in which there are no circuits. In other words, there is a path from any vertex to any other vertex, but no circuits. Some examples of spanning trees are shown below. Notice there are no circuits in the trees, and it is fine to have vertices with degree higher than two.



Usually we have a starting graph to work from, like in the phone example above. In this case, we form our spanning tree by finding a **subgraph** – a new graph formed using all the vertices and some of the edges from the original graph. No edges will be created where they didn't already exist.

Of course, any random spanning tree isn't really what we want. We want the **minimum cost spanning tree (MCST)**, the spanning tree with the smallest total edge weight. A nearest neighbor style approach doesn't make as much sense here since we don't need a circuit, so instead we will take an approach similar to sorted edges.

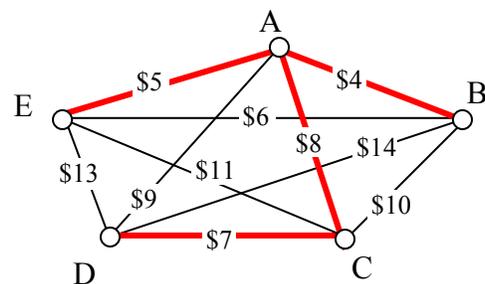
Kruskal's Algorithm

- 1) Select the cheapest unused edge in the graph; highlight it.
- 2) Repeat step 1, adding the cheapest unused edge to the graph, unless:
 - a. adding the edge would create a circuit
- 3) Repeat until a spanning tree is formed

Example: Using our phone line graph from above, begin adding edges:

AB	\$4	OK
AE	\$5	OK
BE	\$6	reject – closes circuit ABEA
DC	\$7	OK
AC	\$8	OK

At this point we stop – every vertex is now connected, so we have formed a spanning tree with cost \$24 thousand a year.



Remarkably, Kruskal’s algorithm is both optimal and efficient; we are guaranteed to produce the optimal MCST.

Example: The power company needs to lay updated distribution lines connecting the ten Oregon cities below to the power grid. How can they minimize the amount of new line to lay?

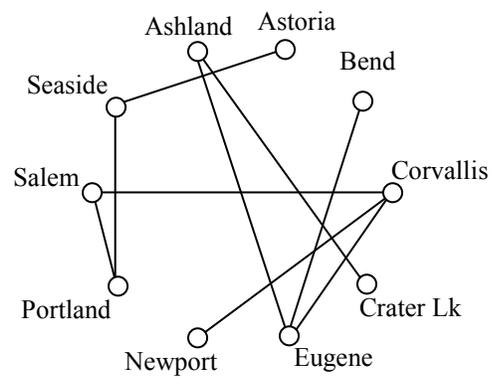
	Ashland	Astoria	Bend	Corvallis	Crater Lake	Eugene	Newport	Portland	Salem	Seaside
Ashland	-	374	200	223	108	178	252	285	240	356
Astoria	374	-	255	166	433	199	135	95	136	17
Bend	200	255	-	128	277	128	180	160	131	247
Corvallis	223	166	128	-	430	47	52	84	40	155
Crater Lake	108	433	277	430	-	453	478	344	389	423
Eugene	178	199	128	47	453	-	91	110	64	181
Newport	252	135	180	52	478	91	-	114	83	117
Portland	285	95	160	84	344	110	114	-	47	78
Salem	240	136	131	40	389	64	83	47	-	118
Seaside	356	17	247	155	423	181	117	78	118	-

Using Kruskal’s algorithm, we add edges from cheapest to most expensive:

- Seaside to Astoria 17 miles
- Corvallis to Salem 40 miles
- Portland to Salem 47 miles
- Corvallis to Eugene 47 miles
- Corvallis to Newport 52 miles
- Salem to Eugene reject – closes circuit
- Portland to Seaside 78 miles
- Newport to Salem reject – closes circuit
- Corvallis to Portland reject – closes circuit
- Eugene to Newport reject – closes circuit
- Portland to Astoria reject – closes circuit
- Ashland to Crater Lk 108 miles
- Eugene to Portland reject – closes circuit
- Newport to Portland reject – closes circuit
- Newport to Seaside reject – closes circuit
- Salem to Seaside reject – closes circuit
- Bend to Eugene 128 miles

- Bend to Salem reject – closes circuit
- Astoria to Newport reject – closes circuit
- Salem to Astoria reject – closes circuit
- Corvallis to Seaside reject – closes circuit
- Portland to Bend reject – closes circuit
- Astoria to Corvallis reject – closes circuit
- Eugene to Ashland 178 miles

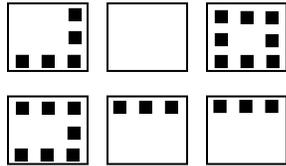
This connects the graph. The total length of cable to lay would be 695 miles.



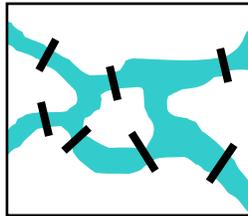
Exercises

Skills

- To deliver mail in a particular neighborhood, the postal carrier needs to walk to along each of the streets with houses (the dots). Create a graph with edges showing where the carrier must walk to deliver the mail.



- Suppose that a town has 7 bridges as pictured below. Create a graph that could be used to determine if there is a path that crosses all bridges once.



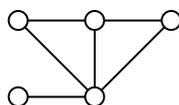
- The table below shows approximate driving times (in minutes, without traffic) between five cities in the Dallas area. Create a weighted graph representing this data.

	Plano	Mesquite	Arlington	Denton
Fort Worth	54	52	19	42
Plano		38	53	41
Mesquite			43	56
Arlington				50

- Shown in the table below are the one-way airfares between 5 cities⁶. Create a graph showing this data.

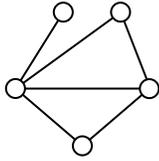
	Honolulu	London	Moscow	Cairo
Seattle	\$159	\$370	\$654	\$684
Honolulu		\$830	\$854	\$801
London			\$245	\$323
Moscow				\$329

- Find the degree of each vertex in the graph below.

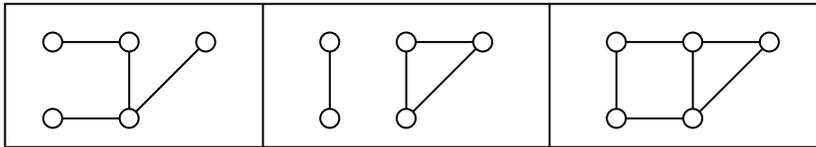


⁶ Cheapest fares found when retrieved Sept 1, 2009 for travel Sept 22, 2009

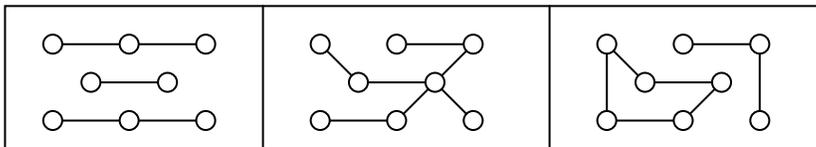
6. Find the degree of each vertex in the graph below.



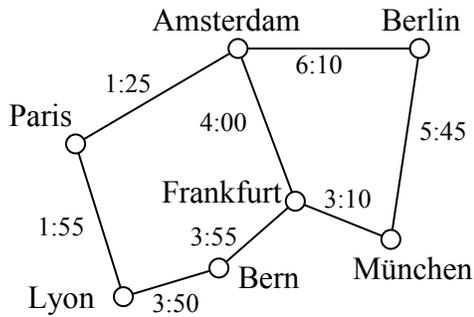
7. Which of these graphs are connected?



8. Which of these graphs are connected?

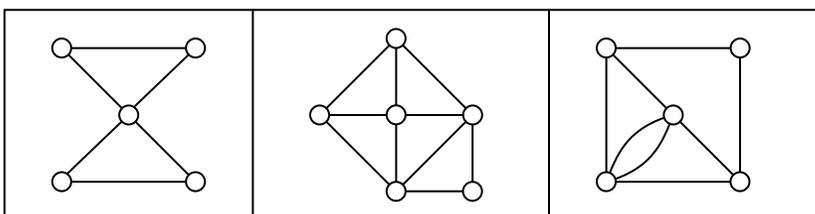


9. Travel times by rail for a segment of the Eurail system is shown below with travel times in hours and minutes⁷. Find path with shortest travel time from Bern to Berlin by applying Dijkstra's algorithm.



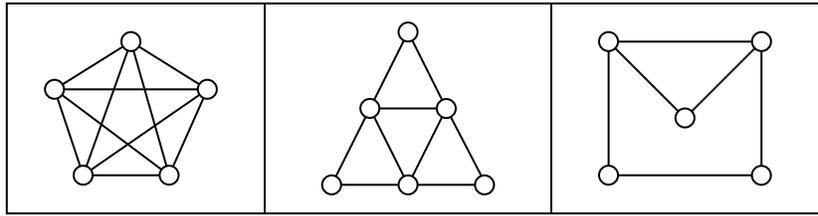
10. Using the graph from the previous problem, find the path with shortest travel time from Paris to München.

11. Does each of these graphs have an Euler circuit? If so, find it.

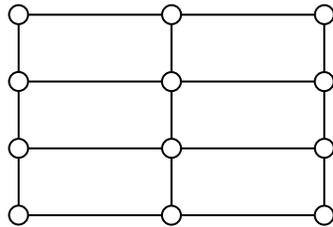


⁷ From <http://www.eurail.com/eurail-railway-map>

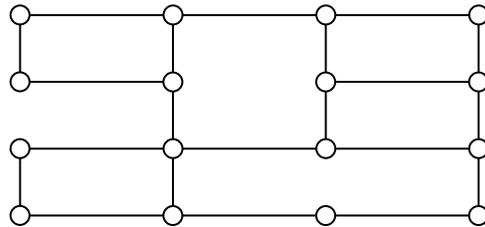
12. Does each of these graphs have an Euler circuit? If so, find it.



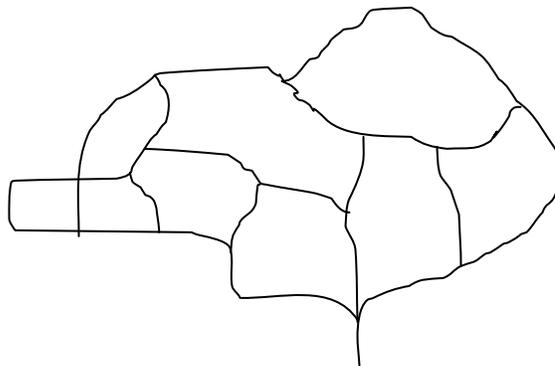
13. Eulerize this graph using as few edge duplications as possible. Then, find an Euler circuit.



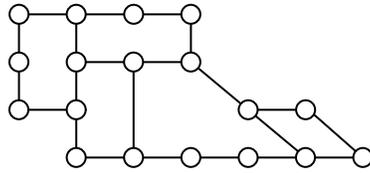
14. Eulerize this graph using as few edge duplications as possible. Then, find an Euler circuit.



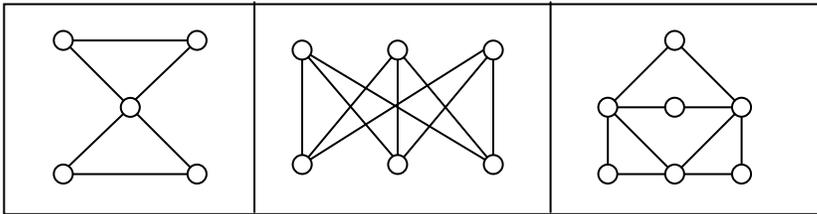
15. The maintenance staff at an amusement park need to patrol the major walkways, shown in the graph below, collecting litter. Find an efficient patrol route by finding an Euler circuit. If necessary, eulerize the graph in an efficient way.



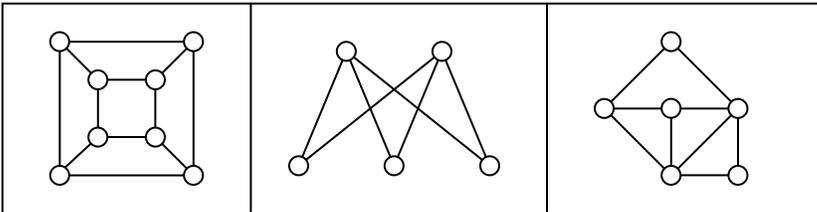
16. After a storm, the city crew inspects for trees or brush blocking the road. Find an efficient route for the neighborhood below by finding an Euler circuit. If necessary, eulerize the graph in an efficient way.



17. Does each of these graphs have at least one Hamiltonian circuit? If so, find one.



18. Does each of these graphs have at least one Hamiltonian circuit? If so, find one.

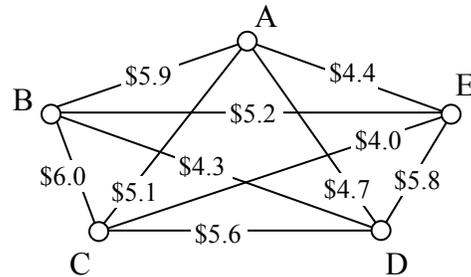


19. A company needs to deliver product to each of their 5 stores around the Dallas, TX area. Driving distances between the stores are shown below. Find a route for the driver to follow, returning to the distribution center in Fort Worth:
- Using Nearest Neighbor starting in Fort Worth
 - Using Repeated Nearest Neighbor
 - Using Sorted Edges

	Plano	Mesquite	Arlington	Denton
Fort Worth	54	52	19	42
Plano		38	53	41
Mesquite			43	56
Arlington				50

20. A salesperson needs to travel from Seattle to Honolulu, London, Moscow, and Cairo. Use the table of flight costs from problem #4 to find a route for this person to follow:
- Using Nearest Neighbor starting in Seattle
 - Using Repeated Nearest Neighbor
 - Using Sorted Edges

21. When installing fiber optics, some companies will install a sonet ring; a full loop of cable connecting multiple locations. This is used so that if any part of the cable is damaged it does not interrupt service, since there is a second connection to the hub. A company has 5 buildings. Costs (in thousands of dollars) to lay cables between pairs of buildings are shown below. Find the circuit that will minimize cost:
- Using Nearest Neighbor starting at building A
 - Using Repeated Nearest Neighbor
 - Using Sorted Edges



22. A tourist wants to visit 7 cities in Israel. Driving distances between the cities are shown below⁸. Find a route for the person to follow, returning to the starting city:
- Using Nearest Neighbor starting in Jerusalem
 - Using Repeated Nearest Neighbor
 - Using Sorted Edges

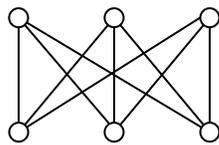
	Jerusalem	Tel Aviv	Haifa	Tiberias	Beer Sheba	Eilat
Jerusalem	--					
Tel Aviv	58	--				
Haifa	151	95	--			
Tiberias	152	134	69	--		
Beer Sheba	81	105	197	233	--	
Eilat	309	346	438	405	241	--
Nazareth	131	102	35	29	207	488

- Find a minimum cost spanning tree for the graph you created in problem #3
- Find a minimum cost spanning tree for the graph you created in problem #22
- Find a minimum cost spanning tree for the graph from problem #21

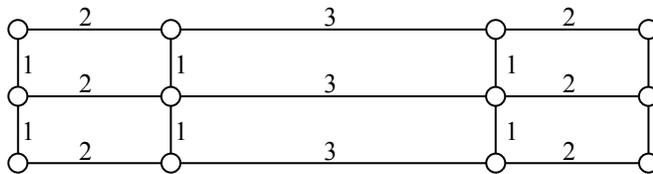
⁸ From <http://www.ddtravel-acc.com/Israel-cities-distance.htm>

Concepts

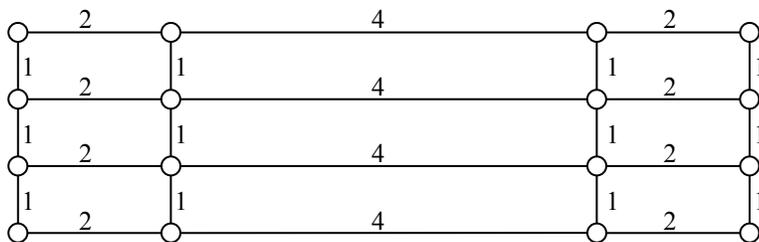
26. Can a graph have one vertex with odd degree? If not, are there other values that are not possible? Why?
27. A complete graph is one in which there is an edge connecting every vertex to every other vertex. For what values of n does complete graph with n vertices have an Euler circuit? A Hamiltonian circuit?
28. Create a graph by drawing n vertices in a row, then another n vertices below those. Draw an edge from each vertex in the top row to every vertex in the bottom row. An example when $n=3$ is shown below. For what values of n will a graph created this way have an Euler circuit? A Hamiltonian circuit?



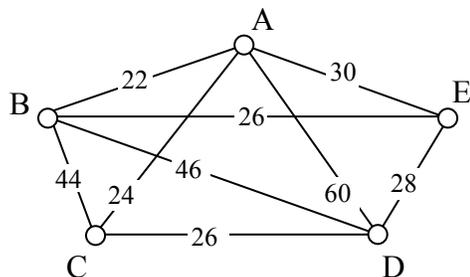
29. Eulerize this graph in the most efficient way possible, considering the weights of the edges.



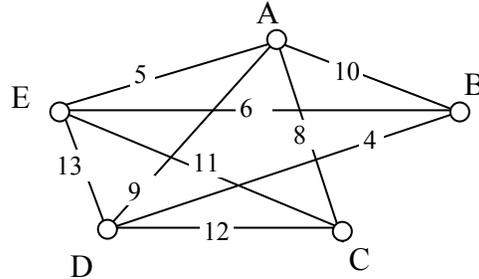
30. Eulerize this graph in the most efficient way possible, considering the weights of the edges.



31. Eulerize this graph in the most efficient way possible, considering the weights of the edges.



32. Eulerize this graph in the most efficient way possible, considering the weights of the edges.

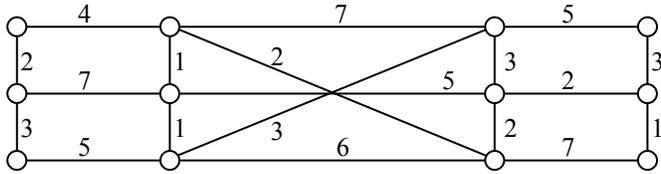


Explorations

33. Social networks such as Facebook and MySpace can be represented using graphs in which vertices represent people and edges are drawn between two vertices when those people are “friends.” The table below shows a friendship table, where an X shows that two people are friends.
- Create a graph of this friendship table
 - Find the shortest path from A to D. The length of this path is often called the “degrees of separation” of the two people.
 - Extension: Split into groups. Each group will pick 10 or more movies, and look up their major actors (www.imdb.com is a good source). Create a graph with each actor as a vertex, and edges connecting two actors in the same movie (note the movie name on the edge). Find interesting paths between actors, and quiz the other groups to see if they can guess the connections.
34. A spell checker in a word processing program makes suggestions when it finds a word not in the dictionary. To determine what words to suggest, it tries to find similar words. One measure of word similarity is the Levenshtein distance, which measures the number of substitutions, additions, or deletions that are required to change one word into another. For example, the words spit and spot are a distance of 1 apart; changing spit to spot requires one substitution (i for o). Likewise, spit is distance 1 from pit since the change requires one deletion (the s). The word spite is also distance 1 from spit since it requires one addition (the e). The word soot is distance 2 from spit since two substitutions would be required.
- Create a graph using words as vertices, and edges connecting words with a Levenshtein distance of 1. Use the misspelled word “moke” as the center, and try to find at least 10 connected dictionary words. How might a spell checker use this graph?
 - Improve the method from above by assigning a weight to each edge based on the likelihood of making the substitution, addition, or deletion. You can base the weights on any reasonable approach: proximity of keys on a keyboard, common language errors, etc. Use Dijkstra’s algorithm to find the length of the shortest path from each word to “moke”. How might a spell checker use these values?

35. The graph below contains two vertices of odd degree. To eulerize this graph, it is necessary to duplicate edges connecting those two vertices.

- a. Use Dijkstra's algorithm to find the shortest path between the two vertices with odd degree. Does this produce the most efficient eulerization and solve the Chinese Postman Problem for this graph?



- b. Suppose a graph has n odd vertices. Using the approach from part a, how many shortest paths would need to be considered? Is this approach going to be efficient?

Scheduling

An event planner has to juggle many workers completing different tasks, some of which must be completed before others can begin. For example, the banquet tables would need to be arranged in the room before the catering staff could begin setting out silverware. The event planner has to carefully schedule things so that everything gets done in a reasonable amount of time.

The problem of scheduling is fairly universal. Contractors need to schedule workers and subcontractors to build a house as quickly as possible. A magazine needs to schedule writers, editors, photographers, typesetters, and others so that an issue can be completed on time.

Getting Started

To begin thinking about scheduling, let us consider an auto shop that is converting a car from gas to electric. A number of steps are involved. A time estimate for each task is given.

Task 1: Remove engine and gas parts (2 days)

Task 2: Steam clean the inside of the car (0.5 day)

Task 3: Buy an electric motor and speed controller (2 days for travel)

Task 4: Construct the part that connects the motor to the car's transmission (1 day)

Task 5: Construct battery racks (2 days)

Task 6: Install the motor (0.5 day)

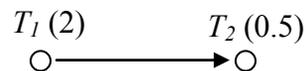
Task 7: Install the speed controller (0.5 day)

Task 8: Install the battery racks (0.5 day)

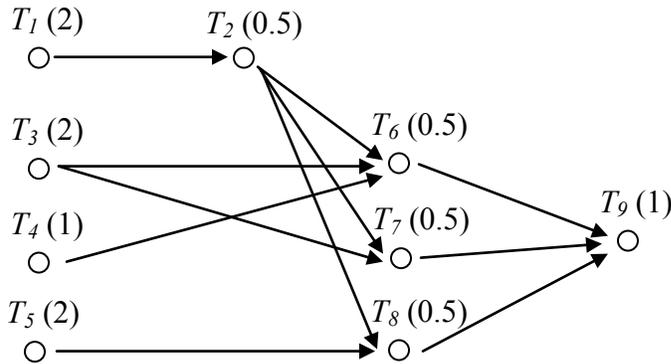
Task 9: Wire the electricity (1 day)

Some tasks have to be completed before others – we certainly can't install the new motor before removing the old engine! There are some tasks, however, that can be worked on simultaneously by two different people, like constructing the battery racks and installing the motor.

To help us visualize the ordering of tasks, we will create a **digraph**, a graphical representation in which tasks are represented with dots, called vertices, and arrows between vertices are used to show ordering. For example, this digraph shows that Task 1, notated T_1 for compactness, needs to be completed before Task 2. The number in parentheses after the task name is the time required for the task.

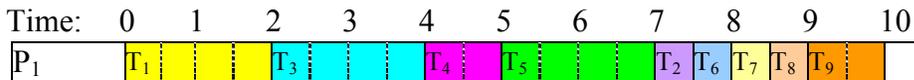


Example: A complete digraph for our car conversion would look like this:



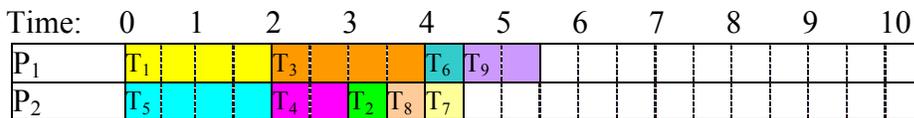
The time it takes to complete this job will partially depend upon how many people are working on the project. In scheduling jargon, the workers are called **processors**. While in this example the processors are humans, in some situations the processors are computers, robots, or other machines. For simplicity, we are going to make the very big assumptions that every processor can do every task, that they all would take the same time to complete it, and that only one processor can work on a task at a time.

If we had only one processor working on this task, it is easy to determine the **finishing time** - how long it will take to complete all the tasks; just add up the individual times. We assume one person can't work on two tasks at the same time, ignore things like drying times during which someone could work on another task. Scheduling with one processor, a possible schedule would look like this, with a finishing time of 10 days.



In this schedule, all the ordering requirements are met. This is certainly not the only possible schedule for one processor, but no other schedule could complete the job in less time. Because of this, this is an **optimal schedule** with **optimal finishing time** – there is nothing better.

For two processors, things become more interesting. For small digraphs like this, we probably could fiddle around and guess-and-check a pretty good schedule. Here would be a possibility:



With two processors, the finishing time was reduced to 5.5 days. What was processor 2 doing during the last day? Nothing, because there were no tasks for the processor to do. This is called **idle time**.

Is this schedule optimal? Could it have been completed in 5 days? Because every other task had to be completed before task 9 could start, there would be no way that both processors could be busy during task 9, so it is not possible to create a shorter schedule.

So how long will it take if we use three processors? About $10/3 = 3.33$ days? Again we will guess-and-check a schedule:

Time:	0	1	2	3	4	5	6	7	8	9	10
P ₁	T ₁		T ₂	T ₆	T ₈	T ₉					
P ₂	T ₃			T ₇							
P ₃	T ₄	T ₅									

With three processors, the job still took 4.5 days. It is a little harder to tell whether this schedule is optimal. However, it might be helpful to notice that since Task 1, 2, 6, and 9 have to be completed sequentially, there is no way that this job could be completed in less than $2+0.5+0.5+1 = 4$ days, regardless of the number of processors. Four days is, for this digraph, the absolute minimum time to complete the job, called the **critical time**.

Adding an algorithm

Up until now, we have been creating schedules by guess-and-check, which works well enough for small schedules, but would not work well with dozens or hundreds of tasks. To create a more procedural approach, we might begin by somehow creating a **priority list**. Once we have a priority list, we can begin scheduling using that list and the **list processing algorithm**.

List Processing Algorithm

- 1) On the digraph or priority list, circle all tasks that are **ready**, meaning that all pre-requisite tasks have been completed.
- 2) Assign to each available processor, in order, the first ready task. Mark the task as **in progress**, perhaps by putting a single line through the task.
- 3) Move forward in time until a task is completed. Mark the task as **completed**, perhaps by crossing out the task. If any new tasks become ready, mark them as such.
- 4) Repeat until all tasks have been scheduled.

Example: Using our digraph from above, schedule it using the priority list below:
T₁, T₃, T₄, T₅, T₆, T₇, T₈, T₂, T₉

Time 0: Mark ready tasks

Priority list: (T₁), (T₃), (T₄), (T₅), T₆, T₇, T₈, T₂, T₉

We assign the first task, T₁ to the first processor, P₁, and the second ready task, T₃, to the second processor. Making those assignments, we mark those tasks as in progress:

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, T₆, T₇, T₈, T₂, T₉

Schedule up to here:

Time:	0	1	2	3	4	5	6	7	8	9	10
P ₁	T ₁										
P ₂	T ₃										

Time 2: Both processors complete their tasks. We mark those tasks as complete. With Task 1 complete, Task 2 becomes ready:

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, T₆, T₇, T₈, ~~T₂~~, T₉

We assign the next ready task on the list, T₄ to P₁, and T₅ to P₂.

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, T₆, T₇, T₈, ~~T₂~~, T₉

Time:	0	1	2	3	4	5	6	7	8	9	10
P ₁	T ₁		T ₄								
P ₂	T ₃		T ₅								

Time 3: Processor 1 has completed T₄. Completing T₄ does not make any other tasks ready (note that all the rest require that T₂ be completed first).

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, T₆, T₇, T₈, ~~T₂~~, T₉

Since the next three tasks are not yet ready, we assign the next ready task, T₂ to P₁

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, T₆, T₇, T₈, ~~T₂~~, T₉

Time:	0	1	2	3	4	5	6	7	8	9	10
P ₁	T ₁		T ₄	T ₂							
P ₂	T ₃		T ₅								

Time 3.5: Processor 1 has completed T₂. Completing T₂ causes T₆ and T₇ to become ready. We assign T₆ to P₁

Priority list: ~~T₁~~, ~~T₃~~, ~~T₄~~, ~~T₅~~, ~~T₆~~, ~~T₇~~, T₈, ~~T₂~~, T₉

Time:	0	1	2	3	4	5	6	7	8	9	10
P ₁	T ₁		T ₄	T ₂	T ₆						
P ₂	T ₃		T ₅								

Time 4: Both processors complete their tasks. The completion of T_5 allows T_8 to become ready. We assign T_7 to P_1 and T_8 to P_2 .

Priority list: ~~T_1~~ , ~~T_3~~ , ~~T_4~~ , ~~T_5~~ , ~~T_6~~ , ~~T_7~~ , ~~T_8~~ , ~~T_2~~ , T_9

Time:	0	1	2	3	4	5	6	7	8	9	10
P_1	T_1			T_4	T_2	T_6	T_7				
P_2	T_3		T_5		T_8						

Time 4.5: Both processors complete their tasks. T_9 becomes ready, and is assigned to P_1 . There is no ready task for P_2 to work on, so P_2 idles.

Priority list: ~~T_1~~ , ~~T_3~~ , ~~T_4~~ , ~~T_5~~ , ~~T_6~~ , ~~T_7~~ , ~~T_8~~ , ~~T_2~~ , ~~T_9~~

Time:	0	1	2	3	4	5	6	7	8	9	10
P_1	T_1			T_4	T_2	T_6	T_7	T_9			
P_2	T_3		T_5		T_8						

With the last task completed, we have a schedule.

It is worth noting that the list processing algorithm itself does not influence the resulting schedule – the schedule is completely determined by the priority list followed. The list processing, while do-able by hand, could just as easily be computed by a computer. The interesting part of scheduling, then, is how to create the best priority list possible.

Choosing a priority list

We will explore two algorithms for selecting a priority list.

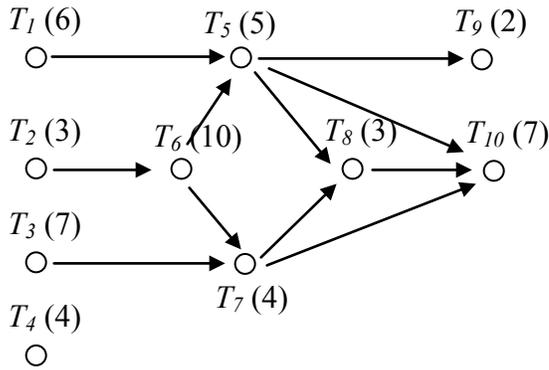
Decreasing time algorithm

The decreasing time algorithm takes the approach of trying to get the very long tasks out of the way as soon as possible by putting them first on the priority list.

Decreasing Time Algorithm

Create the priority list by listing the tasks in order from longest completion time to shortest completion time.

Example: Consider the scheduling problem represented by the digraph below.



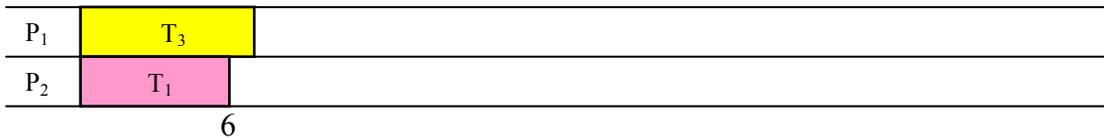
To use the decreasing time list algorithm, we create our priority list by listing the tasks in order from longest task time to shortest task time. If there is a tie, we will list the task with smaller task number first (not for any good reason, but just for consistency).

For this digraph, the decreasing time algorithm would create a priority list of:
 $T_6(10), T_3(7), T_{10}(7), T_1(6), T_5(5), T_4(4), T_7(4), T_2(3), T_8(3), T_9(2)$

Once we have the priority list, we can create the schedule using the list processing algorithm. With two processors, we'd get:

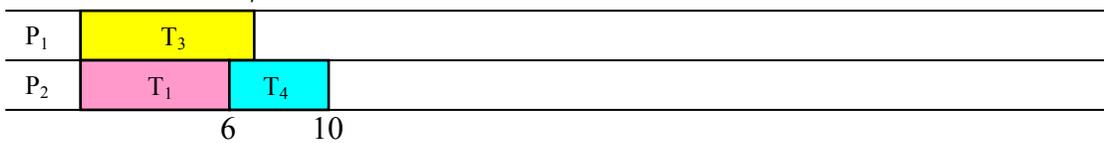
Time 0: We identify ready tasks, and assign T_3 to P_1 and T_1 to P_2

Priority list: $T_6, \cancel{T_3}, T_{10}, \cancel{T_1}, T_5, \cancel{T_4}, T_7, \cancel{T_2}, T_8, T_9$



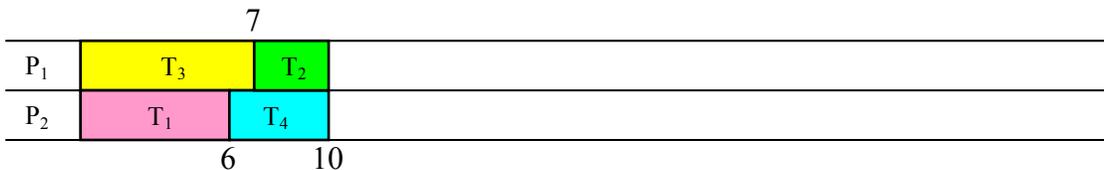
Time 6: P_2 completes T_1 . No new tasks become ready, so T_4 is assigned to P_2 .

Priority list: $T_6, \cancel{T_3}, T_{10}, \cancel{T_1}, T_5, \cancel{T_4}, T_7, \cancel{T_2}, T_8, T_9$



Time 7: P_1 completes T_3 . No new tasks become ready, so T_2 is assigned to P_1 .

Priority list: $T_6, \cancel{T_3}, T_{10}, \cancel{T_1}, T_5, \cancel{T_4}, T_7, \cancel{T_2}, T_8, T_9$



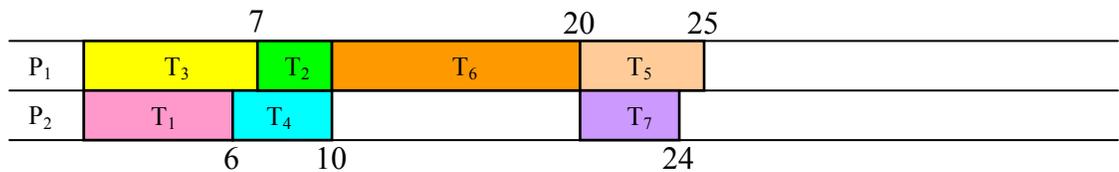
Time 10: Both processors complete their tasks. T_6 becomes ready, and is assigned to P_1 . No other tasks are ready, so P_2 idles.

Priority list: ~~T_6~~ , ~~T_3~~ , T_{10} , ~~T_1~~ , T_5 , ~~T_4~~ , T_7 , ~~T_2~~ , T_8 , T_9



Time 20: With T_6 complete, T_5 and T_7 become ready, and are assigned to P_1 and P_2 respectively.

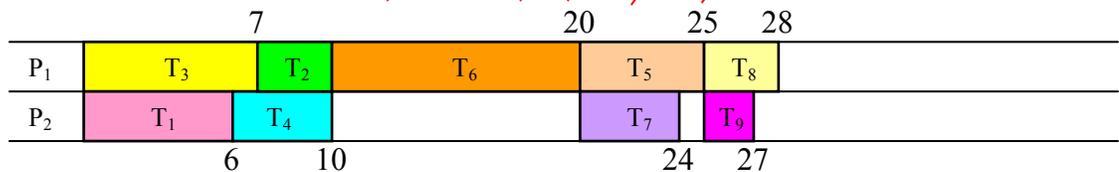
Priority list: ~~T_6~~ , ~~T_3~~ , T_{10} , ~~T_1~~ , ~~T_5~~ , ~~T_4~~ , ~~T_7~~ , ~~T_2~~ , T_8 , T_9



Time 24: P_2 completes T_7 . No new items become ready, so P_2 idles.

Time 25: P_1 completes T_5 . T_8 and T_9 become ready, and are assigned.

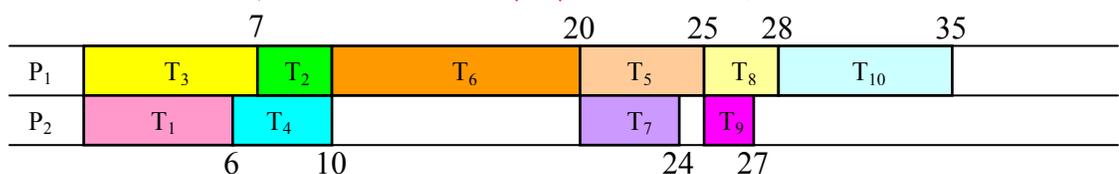
Priority list: ~~T_6~~ , ~~T_3~~ , T_{10} , ~~T_1~~ , ~~T_5~~ , ~~T_4~~ , ~~T_7~~ , ~~T_2~~ , ~~T_8~~ , ~~T_9~~



Time 27: T_9 is completed. No items ready, so P_2 idles.

Time 28: T_8 is completed. T_{10} becomes ready, and is assigned to P_1 .

Priority list: ~~T_6~~ , ~~T_3~~ , ~~T_{10}~~ , ~~T_1~~ , ~~T_5~~ , ~~T_4~~ , ~~T_7~~ , ~~T_2~~ , ~~T_8~~ , ~~T_9~~



Using the decreasing time algorithm, the priority list led to a schedule with a finishing time of 35. Is this good? It certainly looks like there was a lot of idle time in this schedule. To get some idea how good or bad this schedule is, we could compute the critical time, the minimum time to complete the job. To find this, we look for the sequence of tasks with the highest total completion time. For this digraph that sequence would appear to be: T_2 , T_6 , T_5 , T_8 , T_{10} , with total sequence time of 28. From this we can conclude that our schedule isn't horrible, but there is a possibility that a better schedule exists.

Critical path algorithm

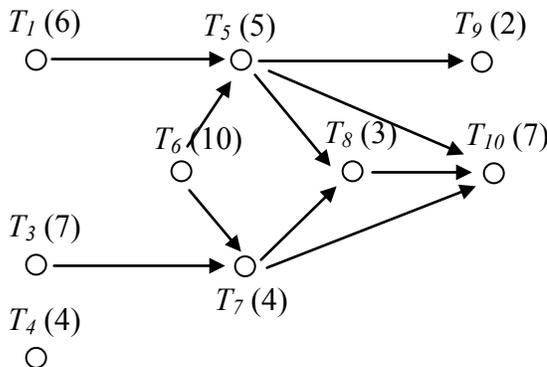
A sequence of tasks in the digraph is called a **path**. In the previous example, we saw that the critical path dictates the minimum completion time for a schedule. Perhaps, then, it would make sense to consider the critical path when creating our schedule. For example, in the last schedule, the processors began working on tasks 1 and 3 because they were longer tasks, but starting on task 2 earlier would have allowed work to begin on the long task 6 earlier.

The critical path algorithm allows you to create a priority list based on ideas of critical paths.

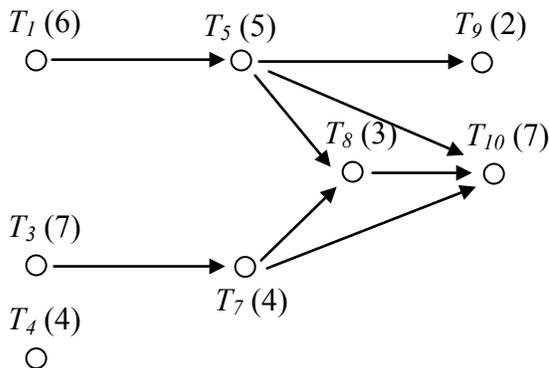
Critical Path Algorithm (version 1)

- 1) Find the critical path.
- 2) The first task in the critical path gets added to the priority list.
- 3) Remove that task from the digraph
- 4) Repeat, finding the new critical path with the revised digraph.

Example: The original digraph from above has critical path $T_2, T_6, T_5, T_8, T_{10}$, so T_2 gets added first to the priority list. Removing T_2 from the digraph, it now looks like:



The critical path is now T_6, T_5, T_8, T_{10} , so T_6 is added to the priority list and removed.



Now there are two paths with the same length: T_1, T_5, T_8, T_{10} and T_3, T_7, T_8, T_{10} . We can add T_1 to the priority list (or T_3 – we usually add the one with smaller item number) and remove it, and continue the process.

I'm sure you can imagine that searching for the critical path every time you remove a task from the digraph would get really tiring, especially for a large digraph. In practice, the critical path algorithm is implemented by first working from the end backwards. This is called the **backflow algorithm**.

Backflow Algorithm

- 1) Introduce an "end" vertex, and assign it a time of 0, shown in [brackets]
- 2) Move backwards to every vertex that has an arrow to the end and assign it a critical time
- 3) From each of those vertices, move backwards and assign those vertices critical times. Notice that the critical time for the earlier vertex will be that task's time plus the critical time for the later vertex.

Example:

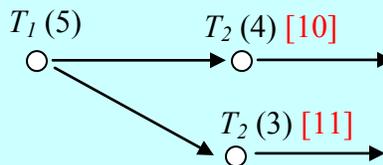


In this case, if T_2 has already been determined to have a critical time of 10, then T_1 will have a critical time of $5+10 = 15$



If you have already assigned a critical time to a vertex, replace it only if the new time is larger.

Example: In the digraph below, T_1 should be labeled with a critical time of 16, since it is the longer of $5+10$ and $5+11$.



- 4) Repeat until all vertices are labeled with their critical times

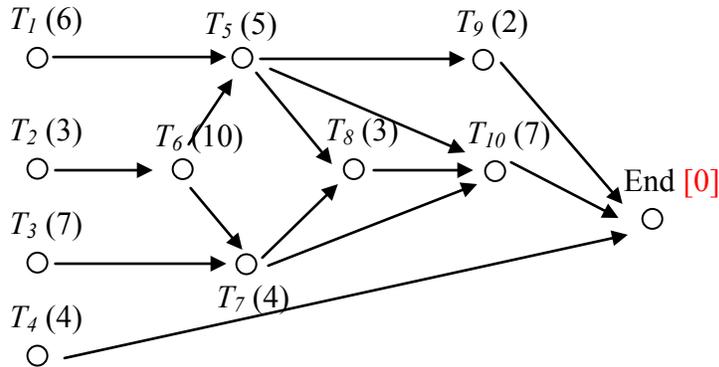
Once you have completed the backflow algorithm, you can easily create the critical path priority list by using the critical times you just found.

Critical Path Algorithm (version 2)

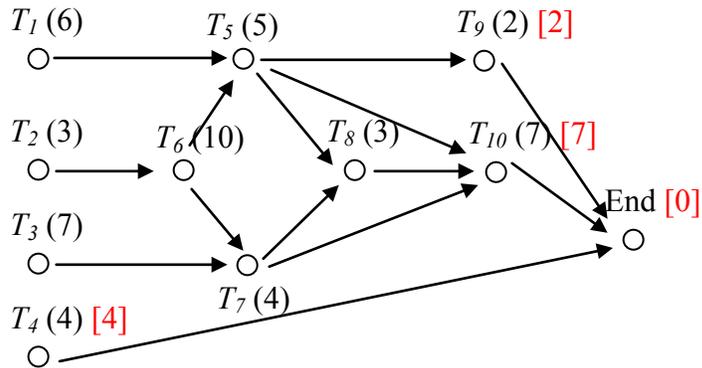
1. Apply the backflow algorithm to the digraph
2. Create the priority list by listing the tasks in order from longest critical time to shortest critical time

This version of the Critical Path Algorithm will usually be the easier to implement.

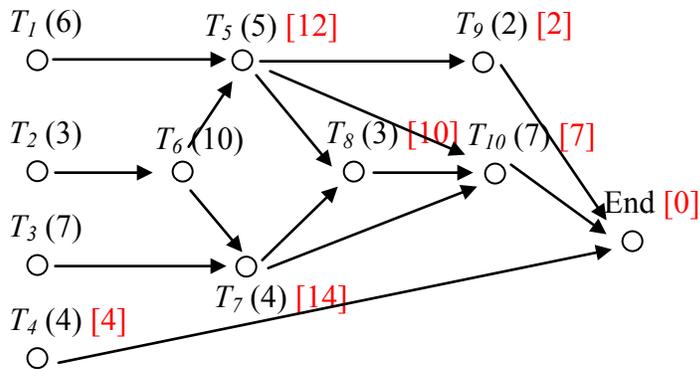
Example: Applying this to our digraph from earlier, we add an end vertex and give it a critical time of 0.



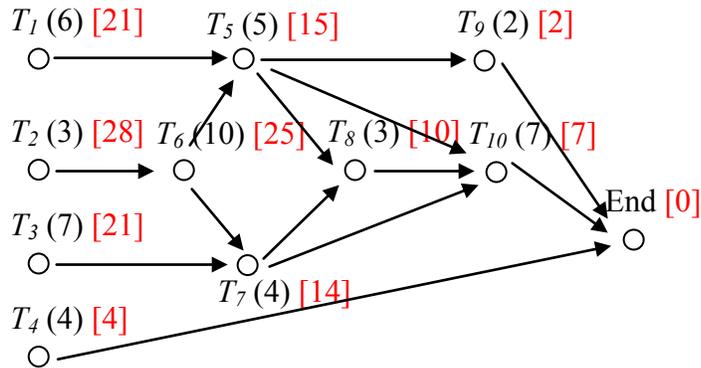
We then move back to T_4 , T_9 , and T_{10} , labeling them with their critical times



From each vertex marked with a critical time, we go back. T_7 , for example, will get labeled with a critical time 11 – the task time of 4 plus the critical time for T_{10} . For T_5 , there are two paths to the end. We use the longer, labeling T_5 with critical time $5+7 = 12$. (This value will end up getting replaced later with the even longer path through T_8).



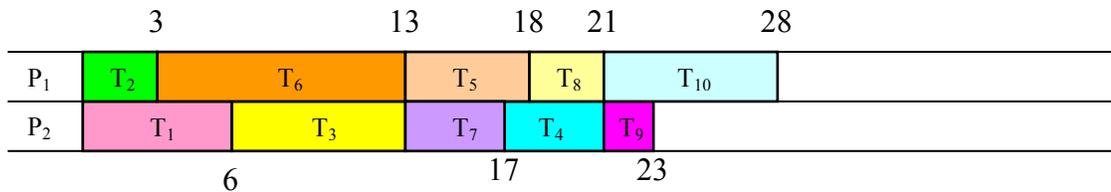
Continue the process until all vertices are labeled.



We can now quickly create the critical path priority list by listing the tasks in decreasing order of critical time:

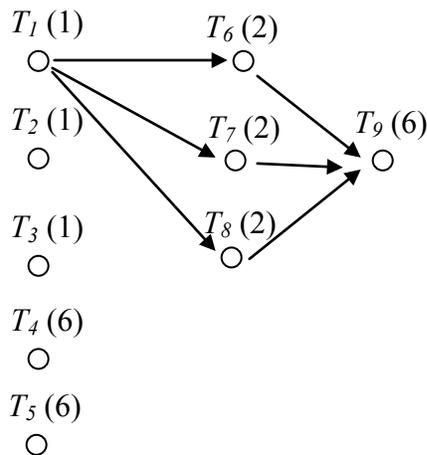
Priority list: T₂, T₆, T₁, T₃, T₅, T₇, T₈, T₁₀, T₄, T₉

Applying this priority list using the list processing algorithm, we get the schedule:

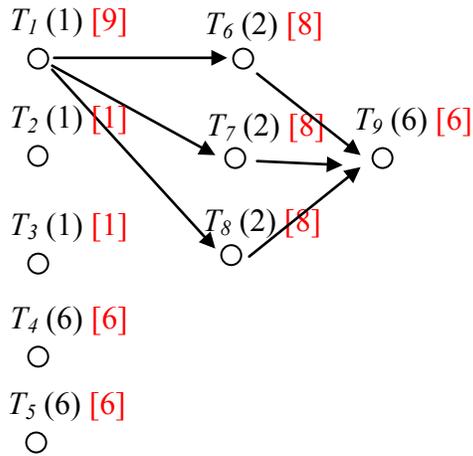


In this particular case, we were able to achieve the minimum possible completion time with this schedule, suggesting that this schedule is optimal. This is certainly not always the case.

Example: This example is designed to show that the critical path algorithm doesn't always work wonderfully.



To create a critical path priority list, we could first apply the backflow algorithm:



This yields the critical-path priority list: T₁, T₆, T₇, T₈, T₄, T₅, T₉, T₂, T₃.

Applying the list processing algorithm to this priority list leads to the schedule:

	1	3	5	7	13
P ₁	T ₁	T ₆	T ₇	T ₈	T ₉
P ₂	T ₄			T ₂	
P ₃	T ₅			T ₃	
	6				

By observation, we can see that a much better schedule exists:

	1	3	9
P ₁	T ₁	T ₆	T ₉
P ₂	T ₂	T ₇	T ₅
P ₃	T ₃	T ₈	T ₄

In most cases the critical path algorithm will lead to a very good schedule. There are cases, like this, where it will not. Unfortunately, there is no known algorithm to always produce the optimal schedule.

Exercises

Skills

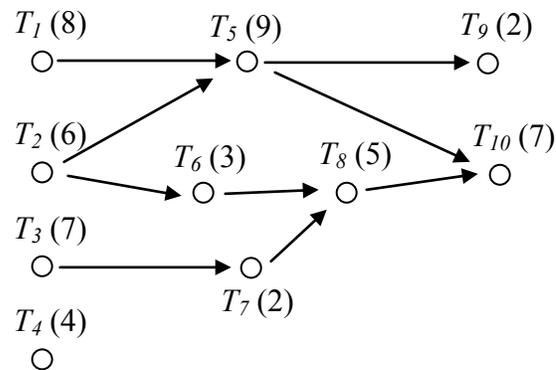
1. Create a digraph for the following set of tasks:

Task	Time required	Tasks that must be completed first
A	3	
B	4	
C	7	
D	6	A, B
E	5	B
F	5	D, E
G	4	E

2. Create a digraph for the following set of tasks:

Task	Time required	Tasks that must be completed first
A	3	
B	4	
C	7	
D	6	A
E	5	A
F	5	B
G	4	D, E

Use this digraph for the next 6 problems.

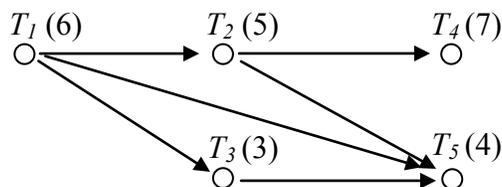


- Using the priority list $T_4, T_3, T_9, T_{10}, T_8, T_5, T_6, T_1, T_7, T_2$ schedule the project with two processors.
- Using the priority list $T_2, T_4, T_6, T_8, T_{10}, T_1, T_3, T_5, T_7, T_9$ schedule the project with two processors.
- Using the priority list $T_4, T_3, T_9, T_{10}, T_8, T_5, T_6, T_1, T_7, T_2$ schedule the project with three processors.

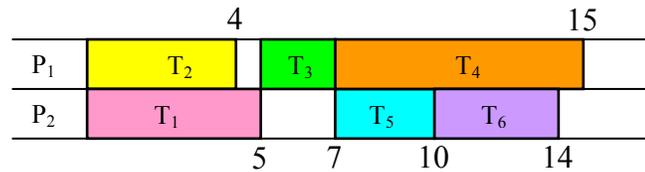
6. Using the priority list $T_2, T_4, T_6, T_8, T_{10}, T_1, T_3, T_5, T_7, T_9$ schedule the project with three processors.
7. Use the decreasing time algorithm to create a priority list for the digraph above, and schedule with two processors.
8. Use the decreasing time algorithm to create a priority list for the digraph above, and schedule with three processors.
9. Use the decreasing time algorithm to create a priority list for the problem from #1, and schedule with two processors.
10. Use the decreasing time algorithm to create a priority list for the problem from #2, and schedule with two processors.
11. With the digraph from above question 3:
 - a. Apply the backflow algorithm to find the critical time for each task
 - b. Find the critical path for the project and the minimum completion time
 - c. Use the critical path algorithm to create a priority list and schedule on two processors.
12. With the digraph from above question 3, use the critical path algorithm to schedule on three processors.
13. Use the critical path algorithm to schedule the problem from #1 on two processors.
14. Use the critical path algorithm to schedule the problem from #2 on two processors.

Concepts

15. If an additional order requirement is added to a digraph, can the optimal finishing time ever become longer? Can the optimal finishing time ever become shorter?
16. Will an optimal schedule always have no idle time?
17. Consider the digraph below.
 - a. How many priority lists could be created for these tasks?
 - b. How many unique schedules are created by those priority lists?



18. Create a digraph and priority list that would lead to the schedule below.



19. Is it possible to create a digraph with three tasks for which every possible priority list creates a different schedule? If so, create it.
20. Is it possible to create a digraph with four tasks for which every possible priority list creates a different schedule? If so, create it.

Exploration

21. Independent tasks are ones that have no order requirements; they can be completed in any order.
- Consider three tasks, with completion times 2, 2, and 4 hours respectively. Construct two different schedules on two processors with different completion times to show that the priority list still matters with independent tasks.
 - Choose a set of independent tasks with different completion times, and implement the decreasing time list algorithm and the critical path algorithm. What do you observe?
 - Will using the decreasing time list or critical path algorithms with independent tasks always produce an optimal schedule?
 - Will using the decreasing time list or critical path algorithms with independent tasks always produce an optimal schedule?
22. In a group, choose ten tasks necessary to throw a birthday party for a friend or child (for example, cleaning the house or buying a cake). Determine order requirements for the tasks, create a digraph, and schedule the tasks for two people.

Growth Models

Populations of people, animals, and items are growing all around us. By understanding how things grow, we can better understand what to expect in the future.

Linear (Algebraic) Growth

Example: Marco is a collector of antique soda bottles. His collection currently contains 437 bottles. Every year, he budgets enough money to buy 32 new bottles. How many bottles will he have in 5 years? How long will it take for his collection to reach 1000 bottles?

While both of these questions you could probably solve without an equation or formal mathematics, we are going to formalize our approach to this problem to provide a means to answer more complicated questions.

Suppose that P_n represents the number, or population, of bottles Marco has after n years. So P_0 would represent the number of bottles now, P_1 would represent the number of bottles after 1 year, P_2 would represent the number of bottles after 2 years, and so on. We could describe how Marco's bottle collection is changing using:

$$P_0 = 437$$

$$P_n = P_{n-1} + 32$$

This is called a **recursive relationship**. A recursive relationship is a formula which relates the next value in a sequence to the previous values. Here, the number of bottles in year n can be found by adding 32 to the number of bottles in the previous year, P_{n-1} . Using this relationship, we could calculate:

$$P_1 = P_0 + 32 = 437 + 32 = 469$$

$$P_2 = P_1 + 32 = 469 + 32 = 501$$

$$P_3 = P_2 + 32 = 501 + 32 = 533$$

$$P_4 = P_3 + 32 = 533 + 32 = 565$$

$$P_5 = P_4 + 32 = 565 + 32 = 597$$

We have answered the question of how many bottles Marco will have in 5 years. However, solving how long it will take for his collection to reach 1000 bottles would require a lot more calculations.

While recursive relationships are excellent for describing simply and cleanly *how* a quantity is changing, they are not convenient for making predictions or solving problems that stretch far into the future. For that, a closed or explicit form for the relationship is preferred. An **explicit equation** allows us to calculate P_n directly, without needing to know P_{n-1} . While you may already be able to guess the explicit equation, let us derive it from the recursive formula. We can do so by selectively not simplifying as we go:

$$P_1 = 437 + 32$$

$$P_2 = P_1 + 32 = 437 + 32 + 32 = 437 + 2(32)$$

$$P_3 = P_2 + 32 = (437 + 2(32)) + 32 = 437 + 3(32)$$

$$P_4 = P_3 + 32 = (437 + 3(32)) + 32 = 437 + 4(32)$$

You can probably see the pattern now, and generalize that

$$P_n = 437 + n(32) = 437 + 32n$$

From this we can calculate

$$P_5 = 437 + 32(5) = 437 + 160 = 597$$

We can now also solve for when the collection will reach 1000 bottles by substituting in 1000 for P_n and solving for n

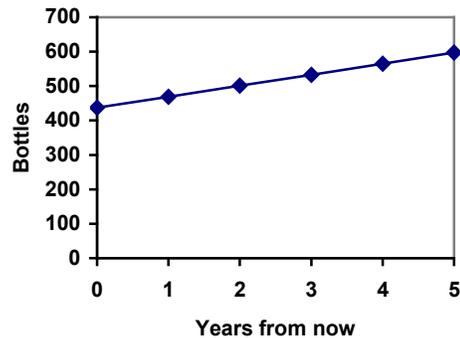
$$1000 = 437 + 32n$$

$$563 = 32n$$

$$n = 563/32 = 17.59$$

So Marco will reach 1000 bottles in 18 years.

In the previous example, Marco's collection grew by the *same number* of bottles every year. This constant change is the hallmark of linear growth. Plotting the values we calculated for Marco's collection, we can see the values form a straight line, the shape of linear growth.



Linear Growth

If a quantity starts at size P_0 and grows by d every time period, then the quantity after n time periods can be determined using either of these relations:

Recursive form:

$$P_n = P_{n-1} + d$$

Explicit form:

$$P_n = P_0 + d n$$

In this equation, d represents the **common difference** – the amount that the population changes each time n increases by 1. You may also recognize it as slope. In fact, the entire explicit equation should look familiar – it is the same linear equation you learned in algebra, probably stated as $y = mx + b$. In that equation recall m was the slope, or increase in y per x , and b was the y -intercept, or the y value when x was zero. Notice that the equations mean the same thing and can be used the same ways, we're just writing it somewhat differently.

Example: The population of elk in a national forest was measured to be 12,000 in 2003, and was measured again to be 15,000 in 2007. If the population continues to grow linearly at this rate, what will the elk population be in 2014?

To begin, we need to define how we're going to measure n . Remember that P_0 is the population when $n = 0$, so we probably don't want to literally use the year 0. Since we already know the population in 2003, let us define $n = 0$ to be the year 2003. Then $P_0 = 12,000$.

Next we need to find d . Remember d is the growth per time period, in this case growth per year. Between the two measurements, the population grew by $15,000 - 12,000 = 3,000$, but it took $2007 - 2003 = 4$ years to grow that much. To find the growth per year, we can divide: $3000 \text{ elk} / 4 \text{ years} = 750 \text{ elk in 1 year}$. We can now write our equation in whichever form is preferred.

Recursive form:

$$P_0 = 12,000$$

$$P_n = P_{n-1} + 750$$

Explicit form:

$$P_n = 12,000 + 750n$$

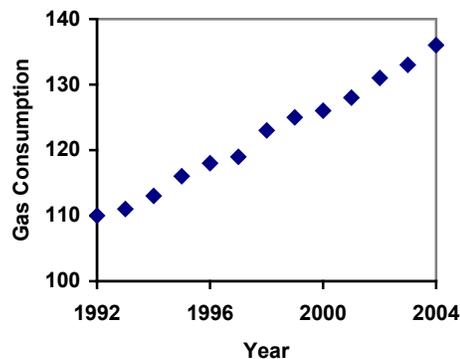
To answer the question, we need to first note that the year 2014 will be $n = 11$, since 2014 is 11 years after 2003. The explicit form will be easier to use for this calculation:

$$P_{11} = 12,000 + 750(11) = 20,250 \text{ elk}$$

Example: Gasoline consumption in the US has been increasing steadily. Consumption data from 1992 to 2004 is shown below¹

Year	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04
Consumption (billion of gallons)	110	111	113	116	118	119	123	125	126	128	131	133	136

Plotting this data, it appears to have an approximately linear relationship:



¹ http://www.bts.gov/publications/national_transportation_statistics/2005/html/table_04_10.html

While there are more advanced statistical techniques that can be used to find an equation to model the data, to get an idea of what is happening, we can find an equation by using two pieces of the data – perhaps the data from 1993 and 2003.

To find d , we need to know how much the gas consumption increased each year, on average. From 1993 to 2003 the gas consumption increased from 111 billion gallons to 133 billion gallons, a total change of $133 - 111 = 22$ billion gallons, over 10 years. This gives us an average change of $22 \text{ billion gallons} / 10 \text{ year} = 2.2 \text{ billion gallons/year}$. Letting $n = 0$ correspond with 1993 would give $P_0 = 111$ billion gallons.

We can now write our equation in whichever form is preferred.

Recursive form:

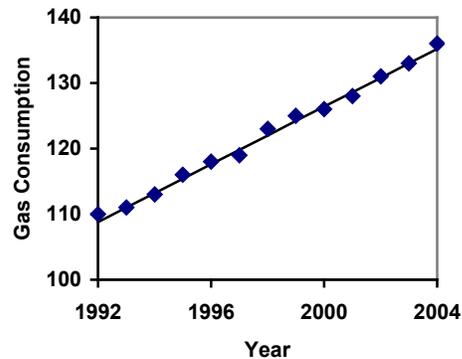
$$P_0 = 111$$

$$P_n = P_{n-1} + 2.2$$

Explicit form:

$$P_n = 111 + 2.2n$$

Calculating values using the explicit form and plotting them with the original data shows how well our model fits the data.



We could now use our model to make predictions about the future, assuming that the previous trend continues unchanged. For example, we could predict that the gasoline consumption in 2016 would be:

$$n = 23 \text{ (2016 - 1993 = 23 years later)}$$

$$P_{23} = 111 + 2.2(23) = 161.6$$

Our model predicts that the US will consume 161.6 billion gallons of gasoline in 2016 if the current trend continues.

When good models go bad

When using mathematical models to predict future behavior, it is important to keep in mind that very few trends will continue indefinitely.

Example: Suppose a four year old boy is currently 39 inches tall, and you are told to expect him to grow 2.5 inches a year. We can set up a growth model, with $n = 0$ corresponding to 4 years old.

Recursive form:

$$P_0 = 39$$

$$P_n = P_{n-1} + 2.5$$

Explicit form:

$$P_n = 39 + 2.5n$$

So at 6 years old, we would expect him to be
 $P_2 = 39 + 2.5(2) = 44$ inches tall

Any mathematical model will break down eventually. Certainly, we shouldn't expect this boy to continue to grow at the same rate all his life. If he did, at age 50 he would be
 $P_{46} = 39 + 2.5(46) = 154$ inches tall = 12.8 feet tall!

When using any mathematical model, we have to consider which inputs are reasonable to use. Whenever we **extrapolate**, or make predictions into the future, we are assuming the model will continue to be valid.

Exponential (Geometric) Growth

Suppose that every year, only 10% of the fish in a lake have surviving offspring. If there were 100 fish in the lake last year, there would now be 110 fish. If there were 1000 fish in the lake last year, there would now be 1100 fish. Absent any inhibiting factors, populations of people and animals tend to grow by a percent each year.

Suppose our lake began with 1000 fish, and 10% of the fish have surviving offspring each year. Since we start with 1000 fish, $P_0 = 1000$. How do we calculate P_1 ? The new population will be the old population, plus an additional 10%. Symbolically:

$$P_1 = P_0 + 0.10P_0$$

Notice this could be condensed to a shorter form by factoring:

$$P_1 = P_0 + 0.10P_0 = 1P_0 + 0.10P_0 = (1 + 0.10)P_0 = 1.10P_0$$

While 10% is the **growth rate**, 1.10 is the **growth multiplier**. Notice that 1.10 can be thought of as “the original 100% plus an additional 10%”

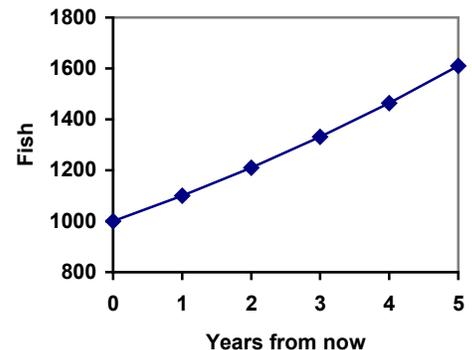
For our fish population,
 $P_1 = 1.10(1000) = 1100$

We could then calculate the population in later years:

$$P_2 = 1.10P_1 = 1.10(1100) = 1210$$

$$P_3 = 1.10P_2 = 1.10(1210) = 1331$$

Notice that in the first year, the population grew by 100 fish, in the second year, the population grew by 110 fish, and in the third year the population grew by 121 fish. While there is a constant *percentage* growth, the actual increase in number of fish is increasing each year. Graphing these values we see that this growth doesn't quite appear linear.



To get a better picture of how this percentage-based growth affects things, we need an explicit form, so we can quickly calculate values further out in the future.

Like we did for the linear model, we will start building from the recursive equation:

$$P_1 = 1.10P_0 = 1.10(1000)$$

$$P_2 = 1.10P_1 = 1.10(1.10(1000)) = 1.10^2(1000)$$

$$P_3 = 1.10P_2 = 1.10(1.10^2(1000)) = 1.10^3(1000)$$

$$P_4 = 1.10P_3 = 1.10(1.10^3(1000)) = 1.10^4(1000)$$

Observing a pattern, we can generate the explicit form to be:

$$P_n = 1.10^n(1000)$$

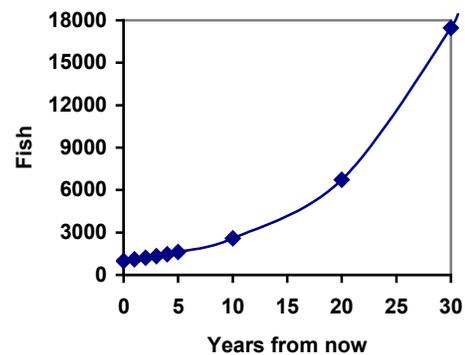
From this, we can calculate the number of fish in 10, 20, or 30 years:

$$P_{10} = 1.10^{10}(1000) = 2594$$

$$P_{20} = 1.10^{20}(1000) = 6727$$

$$P_{30} = 1.10^{30}(1000) = 17449$$

Adding these values to our graph reveals a shape that is definitely not linear. If our fish population had been growing linearly, by 100 fish each year, the population would have only reached 4000 in 30 years compared to almost 18000 with this percent-based growth, called **exponential growth**.



In exponential growth, the population grows proportional to the size of the population, so as the population gets larger, the same percent growth will yield a larger numeric growth.

Exponential Growth

If a quantity starts at size P_0 and grows by $R\%$ (written as a decimal, r) every time period, then the quantity after n time periods can be determined using either of these relations:

Recursive form:

$$P_n = (1+r) P_{n-1}$$

Explicit form:

$$P_n = (1+r)^n P_0$$

To evaluate expressions like 1.05^{20} , it will be easier to use a calculator than multiply 1.05 by itself twenty times. Most scientific calculators have a button for exponents. It is typically either labeled like:

$$\boxed{\wedge}, \boxed{y^x}, \text{ or } \boxed{x^y}.$$

To evaluate 1.05^{20} we'd type $1.05 \boxed{\wedge} 20$, or $1.05 \boxed{y^x} 20$. Try it out - you should get something around 2.6532977.

Example: Between 2007 and 2008, Olympia, WA grew almost 3% to a population of 245 thousand people. If this growth rate was to continue, what would the population of Olympia be in 2014?

As we did before, we first need to define what year will correspond to $n = 0$. Since we know the population in 2008, it would make sense to have 2008 correspond to $n = 0$, so $P_0 = 245,000$. The year 2014 would then be $n = 6$. We know the growth rate is 3%, giving $r = 0.03$.

Using the explicit form:

$$P_6 = (1+0.03)^6 (245,000) = 1.19405(245,000) = 292,542.25$$

So in 2014, Olympia would have a population of about 293 thousand people.

Example: In 1990, the residential energy use in the US was responsible for 962 million metric tons of carbon dioxide emissions. By the year 2000, that number had risen to 1182 million metric tons². If the emissions grow exponentially and continue at the same rate, what will the emissions grow to by 2050?

Similar to before, we will correspond $n = 0$ with 1990, as that is the year for the first piece of data we have. That will make $P_0 = 962$ (million metric tons of CO₂). In this problem, we are not given the growth rate, but instead are given that $P_{10} = 1182$.

$$\begin{aligned} P_{10} &= (1+r)^{10} P_0 \\ P_{10} &= (1+r)^{10} 962 \end{aligned}$$

We also know that $P_{10} = 1182$, so substituting that in, we get

$$1182 = (1+r)^{10} 962$$

We can now solve this equation for the growth rate, r .

$$\begin{aligned} \frac{1182}{962} &= (1+r)^{10} \\ \sqrt[10]{\frac{1182}{962}} &= 1+r \\ r &= \sqrt[10]{\frac{1182}{962}} - 1 = 0.0208 = 2.08\% \end{aligned}$$

So if the emissions are growing exponentially, they are growing by about 2.08% per year. We can now predict the emissions in 2050 by finding P_{60}

$$P_{60} = (1+0.0208)^{60} 962 = 3308.4 \text{ million metric tons of CO}_2 \text{ in 2050}$$

² <http://www.eia.doe.gov/oiaf/1605/ggrpt/carbon.html>

As a note on rounding, notice that if we had rounded the growth rate to 2.1%, our calculation for the emissions in 2050 would have been 3347. Rounding to 2% would have changed our result to 3156. A very small difference in the growth rates gets magnified greatly in exponential growth. For this reason, it is recommended to round the growth rate as little as possible. If you need to round, keep at least three significant digits - numbers after any leading zeros. So 0.4162 could be reasonably rounded to 0.416. A growth rate of 0.001027 could be reasonably rounded to 0.00103.

For the sake of comparison, what would the carbon emissions be in 2050 if emissions grow linearly at the same rate? Again we will get $n = 0$ correspond with 1990, giving $P_0 = 962$. Using an approach similar to that which we used to find the exponential equation, we'll use our knowledge that $P_{10} = 1182$.

$$P_{10} = P_0 + 10d$$

$$P_{10} = 962 + 10d$$

Since we know that $P_{10} = 1182$, substituting that in we get

$$1182 = 962 + 10d$$

We can now solve this equation for the common difference, d .

$$1182 - 962 = 10d$$

$$220 = 10d$$

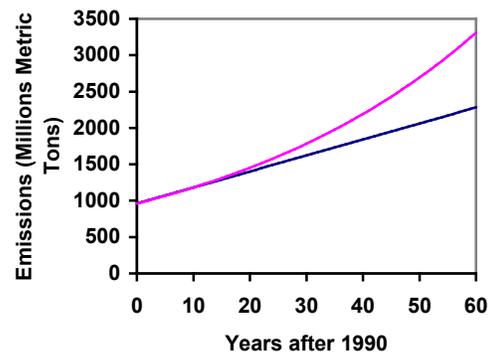
$$d = 22$$

So if the emissions are changing linearly, they are growing by 22 million metric tons each year. Predicting the emissions in 2050,

$$P_{60} = 962 + 22(60) = 2282 \text{ million metric tons.}$$

You will notice that this number is substantially smaller than the prediction from the exponential growth model. Calculating and plotting more values helps illustrate the differences.

So how do we know which growth model to use when working with data? There are two approaches which should be used together whenever possible:



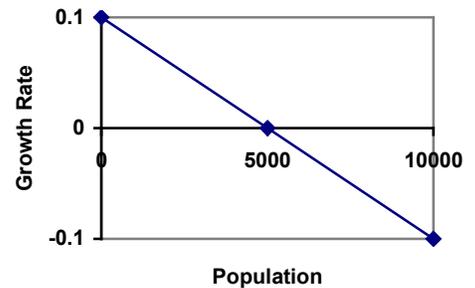
- 1) Find more than two pieces of data. Plot the values, and look for a trend. Does the data appear to be changing like a line, or do the values appear to be curving upwards?
- 2) Consider the factors contributing to the data. Are they things you would expect to change linearly or exponentially? For example, in the case of carbon emissions, we could expect that, absent other factors, they would be tied closely to population values, which tend to change exponentially.

Logistic Growth

In our basic exponential growth scenario, we had a recursive equation of the form

$$P_n = P_{n-1} + r P_{n-1}$$

In a confined environment, however, the growth rate may not remain constant. In a lake, for example, there is some *maximum sustainable population* of fish, also called a **carrying capacity**, which is the largest population that the resources in the lake can sustain. If the population in the lake is far below the carrying capacity, then the population will grow essentially exponentially, but as the population approaches the carrying capacity, the growth rate will decrease. If the population exceeds the carrying capacity, there won't be enough resources to sustain all the fish and there will be a negative growth rate. If the carrying capacity was 5000, the growth rate might vary something like that in the graph shown. Noting that this is a linear equation with intercept at 0.1 and slope $-0.1/5000$, we could write an equation for this adjusted growth rate as:



$$r_{adjusted} = 0.1 - 0.1/5000P = 0.1(1 - P/5000)$$

Substituting this in to our original exponential growth model gives

$$P_n = P_{n-1} + 0.1(1 - P_{n-1}/5000)P_{n-1}$$

Generalizing, given exponential growth rate r and carrying capacity K , we have a growth model for **Logistic Growth**.

Logistic Growth

$$P_n = P_{n-1} + r(1 - P_{n-1}/K)P_{n-1}$$

Unlike linear and exponential growth, logistic growth behaves differently if the populations grow steadily throughout the year or if they have one breeding time per year. The recursive formula provided above models generational growth, where there is one breeding time per year (or, at least a finite number); there is no explicit formula for this type of logistic growth.

Example: A forest is currently home to a population of 200 rabbits. The forest is estimated to be able to sustain a population of 2000 rabbits. Absent any restrictions, the rabbits would grow by 50% per year. Modeling this with a logistic growth model, $r = 0.50$, $K = 2000$, and $P_0 = 200$. Calculating the next few years:

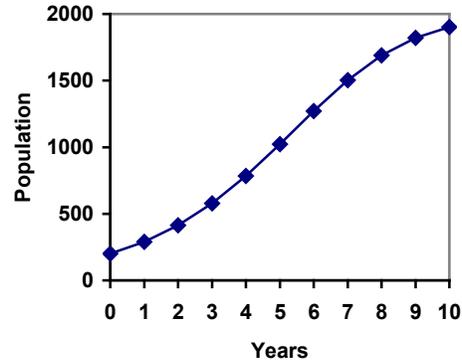
$$P_1 = P_0 + 0.50(1 - P_0/2000)P_0 = 200 + 0.50(1 - 200/2000)200 = 290$$

$$P_2 = P_1 + 0.50(1 - P_1/2000)P_1 = 290 + 0.50(1 - 290/2000)290 = 414 \text{ (rounded)}$$

A calculator was used to calculate more values:

n	0	1	2	3	4	5	6	7	8	9	10
P	200	290	414	578	784	1022	1272	1503	1690	1821	1902

Plotting these values, we can see that the population starts to increase faster and the graph curves upwards during the first few years, like exponential growth, but then the growth slows down as the population approaches the carrying capacity.



Example: On an island that can support a population of 1000 lizards, there is currently a population of 600. These lizards have a lot of offspring and not a lot of natural predators, so have very high growth rate, around 150%. Calculating out the next couple generations:

$$P_1 = P_0 + 1.50(1 - P_0/1000)P_0 = 600 + 1.50(1 - 600/1000)600 = 960$$

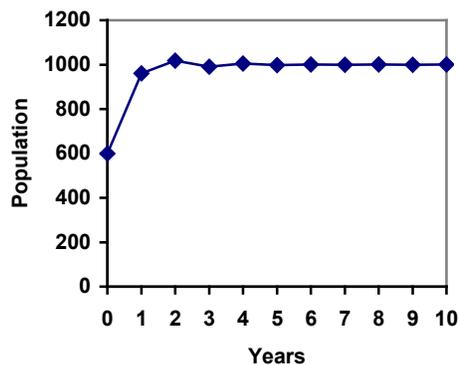
$$P_2 = P_1 + 1.50(1 - P_1/1000)P_1 = 960 + 1.50(1 - 960/1000)960 = 1018$$

Interestingly, even though the factor that limits the growth rate slowed the growth a lot, the population still overshoot the carrying capacity. We would expect the population to decline the next year.

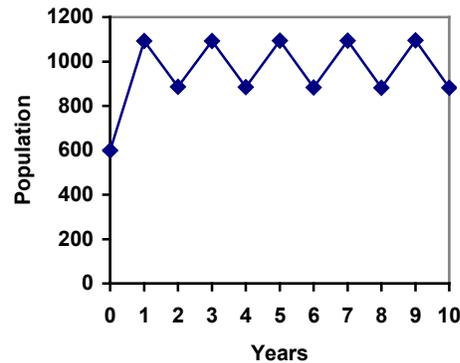
$$P_3 = P_2 + 1.50(1 - P_2/1000)$$

$$P_3 = 1018 + 1.50(1 - 1018/1000)1018 = 991$$

Calculating out a few more years and plotting the results, we see the population wavers above and below the carrying capacity, but eventually settles down, leaving a steady population near the carrying capacity.

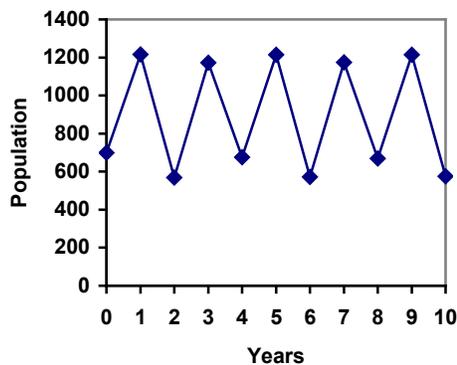


Example: On a neighboring island, the growth rate is even higher – about 205%. Calculating out several generations and plotting the results, we get a surprise: the population seems to be oscillating between two values, a pattern called a 2-cycle. While it would be tempting to treat this only as a strange side effect of mathematics, this has actually been observed in nature. Researchers from the University of California observed a stable 2-cycle in a lizard population in California³.

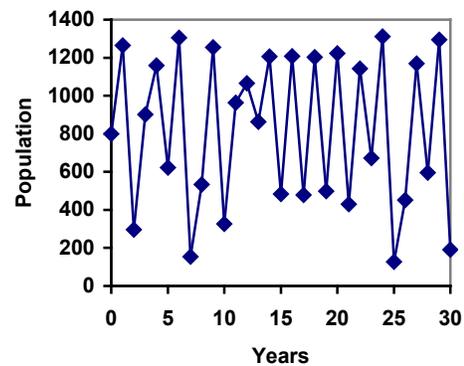


Taking this even further, we get more and more extreme behaviors as the growth rate increases higher. It is possible to get stable 4-cycles, 8-cycles, and higher. Quickly, though, the behavior approaches chaos (remember the movie Jurassic Park?).

r=2.46 A 4-cycle



r=2.9 Chaos!



³ <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/Populations2.html>

Exercises

Skills

1. Marko currently has 20 tulips in his yard. Each year he plants 5 more.
 - a. Write a recursive formula for the number of tulips Marko has
 - b. Write an explicit formula for the number of tulips Marko has
2. Pam is a Disc Jockey. Every week she buys 3 new albums to keep her collection current. She currently owns 450 albums.
 - a. Write a recursive formula for the number of albums Pam has
 - b. Write an explicit formula for the number of albums Pam has
3. A store's sales (in thousands of dollars) grow according to the recursive rule $P_n = P_{n-1} + 15$, with initial population $P_0 = 40$.
 - a. Calculate P_1 and P_2
 - b. Find an explicit formula for P_n
 - c. Use your formula to predict the store's sales in 10 years
4. The number of houses in a town has been growing according to the recursive rule $P_n = P_{n-1} + 30$, with initial population $P_0 = 200$.
 - a. Calculate P_1 and P_2
 - b. Find an explicit formula for P_n
 - c. Use your formula to predict the number of houses in 10 years
5. A population of beetles is growing according to a linear growth model. The initial population (week 0) was $P_0 = 3$, and the population after 8 weeks is $P_8 = 67$.
 - a. Find an explicit formula for the beetle population in week n
 - b. After how many weeks will the beetle population reach 187?
6. The number of streetlights in a town is growing linearly. Four months ago ($n = 0$) there were 130 lights. Now ($n = 4$) there are 146 lights. If this trend continues,
 - a. Find an explicit formula for the number of lights in month n
 - b. How many months will it take to reach 200 lights?
7. Tacoma's population in 2000 was about 200 thousand, and has been growing by about 9% each year.
 - a. Write a recursive formula for the population of Tacoma
 - b. Write an explicit formula for the population of Tacoma
 - c. If this trend continues, what will Tacoma's population be in 2016?
8. Portland's population in 2007 was about 568 thousand, and has been growing by about 1.1% each year.
 - a. Write a recursive formula for the population of Portland
 - b. Write an explicit formula for the population of Portland
 - c. If this trend continues, what will Portland's population be in 2016?

9. Diseases tend to spread according to the exponential growth model. In the early days of AIDS, the growth rate was around 190%. In 1983, about 1700 people in the U.S. died of AIDS. If the trend had continued unchecked, how many people would have died from AIDS in 2005?
10. The population of the world in 1987 was 5 billion and the annual growth rate was estimated at 2 percent per year. Assuming that the world population follows an exponential growth model, find the projected world population in 2015.
11. A bacteria culture is started with 300 bacteria. After 4 hours, the population has grown to 500 bacteria. If the population grows exponentially,
 - a. Write a recursive formula for the number of bacteria
 - b. Write an explicit formula for the number of bacteria
 - c. If this trend continues, how many bacteria will there be in 1 day?
12. A native wolf species has been reintroduced into a national forest. Originally 200 wolves were transplanted. After 3 years, the population had grown to 270 wolves. If the population grows exponentially,
 - a. Write a recursive formula for the number of wolves
 - b. Write an explicit formula for the number of wolves
 - c. If this trend continues, how many wolves will there be in 10 years?
13. One hundred trout are seeded into a lake. Absent constraint, their population will grow by 70% a year. The lake can sustain a maximum of 2000 trout. Using the logistic growth model,
 - a. Write a recursive formula for the number of trout
 - b. Calculate the number of trout after 1 year and after 2 years.
14. Ten blackberry plants started growing in my yard. Absent constraint, blackberries will spread by 200% a month. My yard can only sustain about 50 plants. Using the logistic growth model,
 - a. Write a recursive formula for the number of blackberry plants in my yard
 - b. Calculate the number of plants after 1, 2, and 3 months
15. In 1968, the U.S. minimum wage was \$1.60 per hour. In 1976, the minimum wage was \$2.30 per hour. Assume the minimum wage grows according to an exponential model where n represents the time in years after 1960.
 - a. Find an explicit formula for the minimum wage.
 - b. What does the model predict for the minimum wage in 1960?
 - c. If the minimum wage was \$5.15 in 1996, is this above, below or equal to what the model predicts?

Concepts

16. The population of a small town can be described by the equation $P_n = 4000 + 70n$, where n is the number of years after 2005. Explain in words what this equation tells us about how the population is changing.

17. The population of a small town can be described by the equation $P_n = 4000(1.04)^n$, where n is the number of years after 2005. Explain in words what this equation tells us about how the population is changing.

Exploration

All the examples in the text examined growing quantities, but linear and exponential equations can also describe decreasing quantities, as the next few problems will explore.

18. A new truck costs \$32,000. The car's value will depreciate over time, which means it will lose value. For tax purposes, depreciation is usually calculated linearly. If the truck is worth \$24,500 after three years, write an explicit formula for the value of the car after n years.
19. Inflation causes things to cost more, and for our money to buy less (hence your grandparents saying "In my day, you could buy a cup of coffee for a nickel"). Suppose inflation decreases the value of money by 5% each year. In other words, if you have \$1 this year, next year it will only buy you \$0.95 worth of stuff. How much will \$100 buy you in 20 years?
20. Suppose that you have a bowl of 500 M&M candies, and each day you eat $\frac{1}{4}$ of the candies you have. Is the number of candies left changing linearly or exponentially? Write an equation to model the number of candies left after n days.
21. A warm object in a cooler room will decrease in temperature exponentially, approaching the room temperature according to the formula $T_n = a(1-r)^n + T_r$, where T_n is the temperature after n minutes, r is the rate at which temperature is changing, a is a constant, and T_r is the temperature of the room. Forensic investigators can use this to predict the time of death of a homicide victim. Suppose that when a body was discovered ($n = 0$) it was 85 degrees. After 20 minutes, the temperature was measured again to be 80 degrees. The body was in a 70 degree room.
- Use the given information with the formula provided to find a formula for the temperature of the body.
 - When did the victim die, if the body started at 98.6 degrees? You will either need to guess-and-check answers or your teacher will need to teach logarithms.
22. Recursive equations can be very handy for modeling complicated situations for which explicit equations would be hard to interpret. As an example, consider a lake in which 2000 fish currently reside. The fish population grows by 10% each year, but every year 100 fish are harvested from the lake by people fishing.
- Write a recursive equation for the number of fish in the lake after n years.
 - Calculate the population after 1 and 2 years. Does the population appear to be increasing or decreasing?
 - What is the maximum number of fish that could be harvested each year without causing the fish population to decrease in the long run?

23. The number of Starbucks stores has been growing since they first opened. The number of stores, as reported on their corporate website⁴, is shown below.
- Carefully plot the data. Does it appear to be changing linearly or exponentially?
 - Try finding an equation to model the data by picking two points to work from. How well does the equation model the data?
 - Try using an equation of the form $P_n = P_0 n^k$, where k is a constant, to model the data. This type of model is called a Power model. Compare your results to the results from part b.

Year	Number of Starbucks stores
1990	84
1991	116
1992	165
1993	272
1994	425
1995	677
1996	1015
1997	1412
1998	1886
1999	2498
2000	3501
2001	4709
2002	5886
2003	7225
2004	8569
2005	10241
2006	12440
2007	15756

24. Thomas Malthus was an economist who put forth the principle that population grows based on an exponential growth model, while food and resources grow based on a linear growth model. Based on this, Malthus predicted that eventually demand for food and resources would outgrow supply, with doom-and-gloom consequences. Do some research about Malthus to answer these questions.
- What societal changes did Malthus propose to avoid the doom-and-gloom outcome he was predicting?
 - Why do you think his predictions did not occur?
 - What are the similarities and differences between Malthus's theory and the logistic growth model?

⁴ http://www.starbucks.com/aboutus/Company_Timeline.pdf retrieved May 2009

Finance

We have to work with money every day. While balancing your checkbook or calculating your monthly expenditures on espresso requires only arithmetic, when we start saving, planning for retirement, or need a loan, we need more mathematics.

Simple Interest

Discussing interest starts with the **principal**, or amount your account starts with. This could be a starting investment, or the starting amount of a loan. Interest, in its most simple form, is calculated as a percent of the principal. For example, if you borrowed \$100 from a friend and agree to repay it with 5% interest, then the amount of interest you would pay would just be 5% of 100: $\$100(0.05) = \5 . The total amount you would repay would be \$105, the original principal plus the interest.

Simple One-time Interest

$$I = P_0 r$$

$$A = P_0 + I = P_0 + P_0 r = P_0(1 + r)$$

I is the interest

A is the end amount: principal plus interest

P_0 is the principal (starting amount)

r is the interest rate (in decimal form. Example: 5% = 0.05)

Example: A friend asks to borrow \$300 and agrees to repay it in 30 days with 3% interest. How much interest will you earn?

$P_0 = \$300$ (the principal)

$r = 0.03$ (3% rate)

$I = \$300(0.03) = \9 . You will earn \$9 interest.

One-time simple interest is only common for extremely short-term loans. For longer term loans, it is common for interest to be paid on a daily, monthly, quarterly, or annual basis. In that case, interest would be earned regularly. For example, bonds are essentially a loan made to the bond issuer (a company or government) by you, the bond holder. In return for the loan, the issuer agrees to pay interest, often annually. Bonds have a maturity date, at which time the issuer pays back the original bond value.

Example: Suppose your city is building a new park, and issues bonds to raise the money to build it. You obtain a \$1,000 bond that pays 5% interest annually that matures in 5 years. Each year, you would earn 5% interest: $\$1000(0.05) = \50 in interest. So over the course of five years, you would earn a total of \$250 in interest. When the bond matures, you would receive back the \$1,000 you originally paid, leaving you with a total of \$1,250.

Simple Interest over Time

$$I = P_0rt$$

$$A = P_0 + I = P_0 + P_0rt = P_0(1 + rt)$$

I is the interest

A is the end amount: principal plus interest

P_0 is the principal (starting amount)

r is the interest rate (in decimal form. Example: 5% = 0.05)

t is time

Example: Treasury Notes (T-notes) are bonds issued by the federal government to cover its expenses. Suppose you obtain a \$1,000 T-note with a 4% annual rate, paid semi-annually, with a maturity in 4 years. How much interest will you earn?

First, it is important to know that interest rates are usually given as an **annual percentage rate (APR)** – the total interest that will be paid in the year. Since interest is being paid semi-annually (twice a year), the 4% interest will be divided into two 2% payments.

$P_0 = \$1000$ (the principal)

$r = 0.02$ (2% rate per half-year)

$t = 8$ (4 years = 8 half-years)

$I = \$1000(0.02)(8) = \160 . You will earn \$160 interest total over the four years.

Compound Interest

With simple interest, we were assuming that we pocketed the interest when we received it. In a standard bank account, any interest we earn is automatically added to our balance. This reinvestment of interest is called **compounding**.

Example: Suppose that we deposit \$1000 in a bank account offering 3% interest, compounded monthly. How will our money grow?

The 3% interest is an annual percentage rate (APR) – the total interest to be paid during the year. Since interest is being paid monthly, each month, we will earn $3\%/12 = 0.25\%$ per month.

So in the first month,

$P_0 = \$1000$

$r = 0.0025$ (0.25%)

$I = \$1000(0.0025) = \2.50

$A = \$1000 + \$2.50 = \$1002.50$

So in the first month, we will earn \$2.50 in interest, raising our account balance to \$1002.50. In the second month,

$$P_0 = \$1002.50$$

$$r = 0.0025 \text{ (0.25\%)}$$

$$I = \$1002.50 (0.0025) = \$2.51 \text{ (rounded)}$$

$$A = \$1002.50 + \$2.51 = \$1005.01$$

Notice that in the second month we earned more interest than we did in the first month. This is because we earned interest not only on the original \$1000 we deposited, but we also earned interest on the \$2.50 of interest we earned the first month. This is the key advantage that **compounding** of interest gives us.

Calculating out a few more months:

Month	Starting balance	Interest earned	Ending Balance
1	1000.00	2.50	1002.50
2	1002.50	2.51	1005.01
3	1005.01	2.51	1007.52
4	1007.52	2.52	1010.04
5	1010.04	2.53	1012.57
6	1012.57	2.53	1015.10
7	1015.10	2.54	1017.64
8	1017.64	2.54	1020.18
9	1020.18	2.55	1022.73
10	1022.73	2.56	1025.29
11	1025.29	2.56	1027.85
12	1027.85	2.57	1030.42

To find an equation to represent this, if P_m represents the amount of money after m months, then

$$P_0 = \$1000$$

$$P_m = (1+0.0025)P_{m-1}$$

To build an explicit equation for growth,

$$P_0 = \$1000$$

$$P_1 = 1.0025P_0 = 1.0025 (1000)$$

$$P_2 = 1.0025P_1 = 1.0025 (1.0025 (1000)) = 1.0025^2(1000)$$

$$P_3 = 1.0025P_2 = 1.0025 (1.0025^2(1000)) = 1.0025^3(1000)$$

$$P_4 = 1.0025P_3 = 1.0025 (1.0025^3(1000)) = 1.0025^4(1000)$$

Observing a pattern, we could conclude

$$P_m = (1.0025)^m(\$1000)$$

Notice that the \$1000 in the equation was P_0 , the starting amount. We found 1.0025 by adding one to the growth rate divided by 12, since we were compounding 12 times per year. Generalizing our result, we could write

$$P_m = P_0 \left(1 + \frac{r}{k} \right)^m$$

In this formula:

m is the number of compounding periods (months in our example)

r is the annual interest rate

k is the number of compounds per year.

While this formula works fine, it is more common to use a formula that involves the number of years, rather than the number of compounding periods. If N is the number of years, then $m = Nk$. Making this change gives us the standard formula for compound interest:

Compound Interest

$$P_N = P_0 \left(1 + \frac{r}{k} \right)^{Nk}$$

In this formula:

P_N is the balance in the account after N years.

P_0 is the starting balance of the account (also called initial deposit, or principal)

r is the annual interest rate (in decimal form. Example: 5% = 0.05)

k is the number of compounding periods in one year.

If the compounding is done annually (once a year), $k = 1$.

If the compounding is done quarterly, $k = 4$.

If the compounding is done monthly, $k = 12$.

If the compounding is done daily, $k = 365$.

The most important thing to remember about using this formula is that it assumes that we put money in the account once and let it sit there earning interest.

If you're unsure how to raise numbers to large powers, see the *Using Your Calculator* section at the end of the chapter.

Example 1. A certificate of deposit (CD) is savings instrument that many banks offer. It usually gives a higher interest rate, but you cannot access your investment for a specified length of time. Suppose you deposit \$3000 in a CD paying 6% interest, compounded monthly. How much will you have in the account after 20 years?

In this example,

$P_0 = \$3000$ (the initial deposit)

$r = 0.06$ (6%)

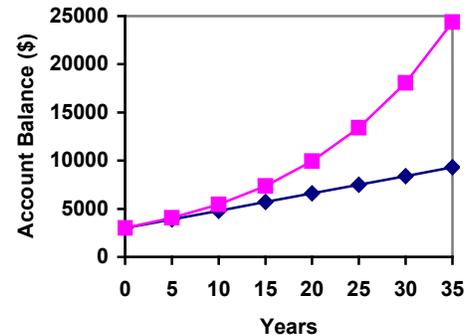
$k = 12$ (12 months in 1 year)

$N = 20$, since we're looking for P_{20}

$$\text{So } P_{20} = 3000 \left(1 + \frac{0.06}{12} \right)^{20 \times 12} = \$9930.61 \text{ (round your answer to the nearest penny)}$$

Let us compare the amount of money you will have from compounding against the amount you would have just from simple interest

Years	Simple Interest (\$15 per month)	0.5% interest compounded each month.
5	\$3900	\$4046.55
10	\$4800	\$5458.19
15	\$5700	\$7362.28
20	\$6600	\$9930.61
25	\$7500	\$13394.91
30	\$8400	\$18067.73
35	\$9300	\$24370.65



As you can see, over a long period of time, compounding makes a large difference in the account balance.

Example 2. You know that you will need \$40,000 for your child's education in 18 years. If your account earns 4% compounded quarterly, how much would you need to deposit now to reach your goal?

In this example,

We're looking for P_0 .

$r = 0.04$ (4%)

$k = 4$ (4 quarters in 1 year)

$N = 18$

$P_{18} = \$40,000$

In this case, we're going to have to set up the equation, and solve for P_0 .

$$40000 = P_0 \left(1 + \frac{0.04}{4} \right)^{18 \times 4}$$

$$40000 = P_0 (2.0471)$$

$$P_0 = \frac{40000}{2.0471} = \$19539.84$$

So you would need to deposit \$19,539.84 now to have \$40,000 in 18 years.

Note on rounding: You may have to round numbers sometimes while working these problems. Try to keep 3 significant digits after the decimal place (keep 3 non-zero numbers).

For example:

Round 1.00563663 to 1.00564

Round 2.06127 to 2.0613

For more information on rounding, see the *Note on Rounding* at the end of the chapter.

Annuities

For most of us, we aren't able to just put a large sum of money in the bank today. Instead, we save by depositing a smaller amount of money from each paycheck into the bank. This idea is called a **savings annuity**. This is what most retirement plans are.

An annuity can be described recursively in a fairly simple way. Recall that basic compound interest follows from the relationship

$$P_m = \left(1 + \frac{r}{k}\right) P_{m-1}$$

For a savings annuity, we simply need to add a deposit, d , to the account with each compounding period:

$$P_m = \left(1 + \frac{r}{k}\right) P_{m-1} + d$$

Taking this equation from recursive form to explicit form is a bit trickier than with compound interest. It will be easiest to see by working with an example rather than working in general.

Example: Suppose we will deposit \$100 each month into an account paying 6% interest. We assume that the account is compounded with the same frequency as we make deposits unless stated otherwise. So in this example:

$$r = 0.06 \text{ (6\%)}$$

$$k = 12 \text{ (12 compounds/deposits per year)}$$

$$d = \$100 \text{ (our deposit per month)}$$

So

$$P_m = \left(1 + \frac{0.06}{12}\right) P_{m-1} + 100 = (1.005) P_{m-1} + 100$$

Assuming we start with an empty account, we can begin using this relationship:

$$P_0 = 0$$

$$P_1 = (1.005)P_0 + 100 = 100$$

$$P_2 = (1.005)P_1 + 100 = (1.005)(100) + 100 = 100(1.005) + 100$$

$$P_3 = (1.005)P_2 + 100 = (1.005)(100(1.005) + 100) + 100 = 100(1.005)^2 + 100(1.005) + 100$$

Continuing this pattern,

$$P_m = 100(1.005)^{m-1} + 100(1.005)^{m-2} + \cdots + 100(1.005) + 100$$

In other words, after m months, the first deposit will have earned compound interest for $m-1$ months. The second deposit will have earned interest for $m-2$ months. Last months deposit would have earned only one month worth of interest. The most recent deposit will have earned no interest yet.

This equation leaves a lot to be desired, though – it doesn't make calculating the ending balance any easier! To simplify things, multiply both sides of the equation by 1.005:

$$1.005P_m = 1.005\left(100(1.005)^{m-1} + 100(1.005)^{m-2} + \dots + 100(1.005) + 100\right)$$

Distributing on the right side of the equation gives

$$1.005P_m = 100(1.005)^m + 100(1.005)^{m-1} + \dots + 100(1.005)^2 + 100(1.005)$$

Now we'll line this up with like terms from our original equation, and subtract each side

$$\begin{aligned} 1.005P_m &= 100(1.005)^m + 100(1.005)^{m-1} + \dots + 100(1.005) \\ P_m &= 100(1.005)^{m-1} + \dots + 100(1.005) + 100 \end{aligned}$$

Almost all the terms cancel on the right hand side, leaving

$$1.005P_m - P_m = 100(1.005)^m - 100$$

Solving for P_m

$$\begin{aligned} 0.005P_m &= 100\left((1.005)^m - 1\right) \\ P_m &= \frac{100\left((1.005)^m - 1\right)}{0.005} \end{aligned}$$

Replacing m months with $12N$, where N is years gives

$$P_N = \frac{100\left((1.005)^{12N} - 1\right)}{0.005}$$

Recall 0.005 was r/k and 100 was the deposit d . 12 was k , the number of deposit each year. Generalizing this, we get:

Annuity Formula

$$P_N = \frac{d \left(\left(1 + \frac{r}{k} \right)^{Nk} - 1 \right)}{\left(\frac{r}{k} \right)}$$

In this formula:

P_N is the balance in the account after N years.

d is the regular deposit (the amount you deposit each year, each month, etc.)

r is the annual interest rate (in decimal form. Example: 5% = 0.05)

k is the number of compounding periods in one year.

The compounding frequency is not always explicitly given. But:

If you make your deposits every month, use monthly compounding, $k = 12$.

If you make your deposits every year, use yearly compounding, $k = 1$.

If you make your deposits every quarter, use quarterly compounding, $k = 4$.

Etc.

The most important thing to remember about using this formula is that it assumes that you put money in the account on a regular schedule (every month, year, quarter, etc.) and let it sit there earning interest.

Compound interest: One deposit

Annuity: Many deposits.

Example 1. A traditional individual retirement account (IRA) is a special type of retirement account in which the money you invest is exempt from income taxes until you withdraw it. If you deposit \$100 each month into an IRA earning 6% interest. How much will you have in the account after 20 years?

In this example,

$d = \$100$ (the monthly deposit)

$r = 0.06$ (6%)

$k = 12$ (since we're doing monthly deposits, we'll compound monthly)

$N = 20$, since we're looking for P_{20}

Putting this into the equation:

$$P_{20} = \frac{100 \left(\left(1 + \frac{0.06}{12} \right)^{20(12)} - 1 \right)}{\left(\frac{0.06}{12} \right)}$$

$$P_{20} = \frac{100 \left((1.005)^{240} - 1 \right)}{(0.005)}$$

$$P_{20} = \frac{100(3.310 - 1)}{(0.005)}$$

$$P_{20} = \frac{100(2.310)}{(0.005)} = \$46,200$$

So you will have \$46,200 after 20 years. Notice that you deposited into the account a total of \$24,000 (\$100 a month for 240 months). The difference between what you end up with and how much you put in is the interest earned. In this case it is \$46,200 - \$24,000 = \$22,200.

Example 2. You want to have \$200,000 in your account when you retire in 30 years. Your retirement account earns 8% interest. How much do you need to deposit each month to meet your retirement goal?

In this example,

We're looking for d .

$r = 0.08$ (8%)

$k = 12$ (since we're depositing monthly)

$N = 30$ (30 years)

$P_{30} = \$200,000$

In this case, we're going to have to set up the equation, and solve for d .

$$200,000 = \frac{d \left(\left(1 + \frac{0.08}{12} \right)^{30(12)} - 1 \right)}{\left(\frac{0.08}{12} \right)}$$

$$200,000 = \frac{d \left((1.00667)^{360} - 1 \right)}{(0.00667)}$$

$$200,000 = d(1491.57)$$

$$d = \frac{200,000}{1491.57} = \$134.09$$

So you would need to deposit \$134.09 each month to have \$200,000 in 30 years.

Payout Annuities

In the last section you learned about annuities. In an annuity, you start with nothing, put money into an account on a regular basis, and end up with money in your account.

In this section, we will learn about a variation called a **Payout Annuity**. With a payout annuity, you start with money in the account, and pull money out of the account on a regular basis. Any remaining money in the account earns interest. After a fixed amount of time, the account will end up empty.

Payout annuities are typically used after retirement. Perhaps you have saved \$500,000 for retirement, and want to take money out of the account each month to live on. You want the money to last you 20 years. This is a payout annuity. The formula is derived in a similar way as we did for savings annuities. The details are omitted here.

Payout Annuity Formula

$$P_0 = \frac{d \left(1 - \left(1 + \frac{r}{k} \right)^{-Nk} \right)}{\left(\frac{r}{k} \right)}$$

P_0 is the balance in the account at the beginning (starting amount, or principal).
 d is the regular withdrawal (the amount you take out each year, each month, etc.)
 r is the annual interest rate (in decimal form. Example: 5% = 0.05)
 k is the number of compounding periods in one year.
 N is the number of years we plan to take withdrawals

Like with annuities, the compounding frequency is not always explicitly given, but is determined by how often you take the withdrawals.

The most important thing to remember about using this formula is that it assumes that you take money from the account on a regular schedule (every month, year, quarter, etc.) and let the rest sit there earning interest.

Compound interest: One deposit
 Annuity: Many deposits.
 Payout Annuity: Many withdrawals

Example 1. After retiring, you want to be able to take \$1000 every month for a total of 20 years from your retirement account. The account earns 6% interest. How much will you need in your account when you retire?

In this example,
 $d = \$1000$ (the monthly withdrawal)
 $r = 0.06$ (6%)
 $k = 12$ (since we're doing monthly withdrawals, we'll compound monthly)
 $N = 20$, since we were taking withdrawals for 20 years

We're looking for P_0 ; how much money needs to be in the account at the beginning. Putting this into the equation:

$$P_0 = \frac{1000 \left(1 - \left(1 + \frac{0.06}{12} \right)^{-20(12)} \right)}{\left(\frac{0.06}{12} \right)}$$

$$P_0 = \frac{1000 \times \left(1 - (1.005)^{-240} \right)}{(0.005)}$$

$$P_0 = \frac{1000 \times (1 - 0.302)}{(0.005)} = \$139,600$$

So you will need to have \$139,600 in your account when you retire. Notice that you withdrew a total of \$240,000 (\$1000 a month for 240 months). The difference between what you pulled out and what you started with is the interest earned. In this case it is $\$240,000 - \$139,600 = \$100,400$ in interest.

Example 2. You know you will have \$500,000 in your account when you retire. You want to be able to take monthly withdrawals from the account for a total of 30 years. Your retirement account earns 8% interest. How much will you be able to withdraw each month?

In this example,

We're looking for d .

$r = 0.08$ (8%)

$k = 12$ (since we're withdrawing monthly)

$N = 30$ (30 years)

$P_0 = \$500,000$

In this case, we're going to have to set up the equation, and solve for d .

$$500,000 = \frac{d \left(1 - \left(1 + \frac{0.08}{12} \right)^{-30(12)} \right)}{\left(\frac{0.08}{12} \right)}$$

$$500,000 = \frac{d \left(1 - (1.00667)^{-360} \right)}{(0.00667)}$$

$$500,000 = d(136.232)$$

$$d = \frac{500,000}{136.232} = \$3670.21$$

So you would be able to withdraw \$3,670.21 each month for 30 years.

Loans

In the last section, you learned about payout annuities.

In this section, you will learn about conventional loans (also called amortized loans or installment loans). Examples include auto loans and home mortgages. These techniques do not apply to payday loans, add-on loans, or other loan types where the interest is calculated up front.

One great thing about loans is that they use exactly the same formula as a payout annuity. To see why, imagine that you had \$10,000 invested at a bank, and started taking out payments while earning interest as part of a payout annuity, and after 5 years your balance was zero. Flip that around, and imagine that you are acting as the bank, and a car lender is acting as you. The car lender invests \$10,000 in you. Since you're acting as the bank, you pay interest. The car lender takes payments until the balance is zero.

Loans Formula

$$P_0 = \frac{d \left(1 - \left(1 + \frac{r}{k} \right)^{-Nk} \right)}{\left(\frac{r}{k} \right)}$$

In this formula:

P_0 is the balance in the account at the beginning (amount of the loan).

d is your loan payment (your monthly payment, annual payment, etc)

r is the annual interest rate (in decimal form. Example: 5% = 0.05)

k is the number of compounding periods in one year.

N is the length of the loan, in years

Like before, the compounding frequency is not always explicitly given, but is determined by how often you make payments.

The most important thing to remember about using this formula is that it assumes that you make loan payments on a regular schedule (every month, year, quarter, etc.) and are paying interest on the loan.

Compound interest: One deposit

Annuity: Many deposits.

Payout Annuity: Many withdrawals

Loans: Many payments

Example 1. You can afford \$200 per month as a car payment. If you can get an auto loan at 3% interest for 60 months (5 years), how expensive of a car can you afford (in other words, what amount loan can you pay off with \$200 per month)?

In this example,

$d = \$200$ (the monthly payment)

$r = 0.03$ (3%)

$k = 12$ (since we're doing monthly payments, we'll compound monthly)

$N = 5$, since we're making monthly payments for 5 years

We're looking for P_0 , the starting amount of the loan.

$$P_0 = \frac{200 \left(1 - \left(1 + \frac{0.03}{12} \right)^{-5(12)} \right)}{\left(\frac{0.03}{12} \right)}$$

$$P_0 = \frac{200 \left(1 - (1.0025)^{-60} \right)}{(0.0025)}$$

$$P_0 = \frac{200(1 - 0.861)}{(0.0025)} = \$11,120$$

So you can afford a \$11,120 loan. You will pay a total of \$12,000 (\$200 per month for 60 months) to the loan company. The difference between the amount you pay and the amount of the loan is the interest paid. In this case, you're paying \$12,000 - \$11,120 = \$880 interest total.

Example 2. You want to take out a \$140,000 mortgage (home loan). The interest rate on the loan is 6%, and the loan is for 30 years. How much will your monthly payments be?

In this example,

We're looking for d .

$r = 0.06$ (6%)

$k = 12$ (since we're paying monthly)

$N = 30$ (30 years)

$P_0 = \$140,000$ (the loan amount)

In this case, we're going to have to set up the equation, and solve for d .

$$140,000 = \frac{d \left(1 - \left(1 + \frac{0.06}{12} \right)^{-30(12)} \right)}{\left(\frac{0.06}{12} \right)}$$

$$140,000 = \frac{d \left(1 - (1.005)^{-360} \right)}{(0.005)}$$

$$140,000 = d(166.792)$$

$$d = \frac{140,000}{166.792} = \$839.37$$

So you will make payments of \$839.37 per month for 30 years. You're paying a total of \$302,173.20 to the loan company (\$839.37 per month for 360 months). You are paying a total of $\$302,173.20 - \$140,000 = \$162,173.20$ in interest over the life of the loan.

Which equation to use?

When presented with a finance problem (on an exam or in real life), you're usually not told what type of problem it is or which equation to use. Here are some hints on deciding which equation to use based on the wording of the problem.

The easiest types of problem to identify are loans. Loan problems almost always include words like: "loan", "amortize" (the fancy word for loans), "finance (a car)", or "mortgage" (a home loan). Look for these words. If they're there, you're probably looking at a loan problem. To make sure, see if you're given what your monthly (or annual) payment is, or if you're trying to find a monthly payment.

If the problem is not a loan, the next question you want to ask is: "Am I putting money in an account and letting it sit, or am I making regular (monthly/annually/quarterly) payments or withdrawals?" If you're letting the money sit in the account with nothing but interest changing the balance, then you're looking at a compound interest problem. The exception would be bonds and other investments where the interest is not reinvested; in those cases you're looking at simple interest.

If you're making regular payments or withdrawals, the next questions is: "Am I putting money into the account, or am I pulling money out?" If you're putting money into the account on a regular basis (monthly/annually/quarterly) then you're looking at a basic Annuity problem. Basic annuities are when you are saving money. Usually in an annuity problem, your account starts empty, and has money in the future. If you're pulling money out of the account on a regular basis, then you're looking at a Payout Annuity problem. Payout annuities are used for things like retirement income, where you start with money in your account, pull money out on a regular basis, and your account ends up empty in the future.

Remember, the most important part of answering any kind of question, money or otherwise, is first to correctly identify what the question is really asking, and to determine what approach will best allow you to solve the problem.

Using your Calculator

When we need to calculate something like 5^3 it is easy enough to just multiply $5 \cdot 5 \cdot 5 = 125$. But when we need to calculate something like 1.005^{360} , it would be very tedious to calculate this by multiplying 1.005 by itself 360 times! So to make things easier, we can harness the power of our scientific calculators.

Most scientific calculators have a button for exponents. It is typically either labeled like:

$$\boxed{\wedge}, \boxed{y^x}, \text{ or } \boxed{x^y}.$$

To evaluate 1.005^{360} we'd type $1.005 \boxed{\wedge} 360$, or $1.005 \boxed{y^x} 360$. Try it out - you should get something around 6.022575.

With some problems, you need to raise numbers to negative powers. Most calculators have a separate button for negating a number that is different than the subtraction button. Some calculators label this $\boxed{(-)}$, some with $\boxed{+/-}$. The button is often near the = key or the decimal point.

If your calculator displays operations on it (typically a calculator with multiline display), to calculate 1.03^{-10} you'd type something like: $1.03 \boxed{\wedge} \boxed{(-)} 10$

If your calculator only shows one value at a time, then usually you hit the (-) key after a number to negate it, so you'd hit: $1.03 \boxed{y^x} 10 \boxed{(-)} =$

Give it a try - you should get $1.03^{-10} = 0.744094$

A note on rounding

It is important to be very careful about rounding when calculating things with exponents. In general, you want to keep as many decimals during calculations as you can. Be sure to keep at least 3 significant digits (numbers after any leading zeros). So rounding 0.00012345 to 0.000123 will usually give you a “close enough” answer, but keeping more digits is always better.

To see why, suppose you were investing \$1000 at 5% interest compounded monthly for 30 years.

$P_0 = \$1000$ (the initial deposit)

$r = 0.05$ (5%)

$k = 12$ (12 months in 1 year)

$N = 30$, since we're looking for P_{30}

If we first compute r/k , we find $0.05/12 = 0.00416666666667$

Here is the effect of rounding this to different values:

r/k rounded to:	Gives P_{30} to be:	Error
0.004	\$4208.59	\$259.15
0.0042	\$4521.45	\$53.71
0.00417	\$4473.09	\$5.35
0.004167	\$4468.28	\$0.54
0.0041667	\$4467.80	\$0.06
no rounding	\$4467.74	

If you're working in a bank, of course you wouldn't round at all. But for our purposes, the answer we got by rounding to three significant digits is close enough - \$5 off of \$4500 isn't too bad. Certainly keeping that fourth decimal place wouldn't have hurt.

Exercises

Skills

1. A friend lends you \$200 for a week, which you agree to repay with 5% interest. How much will you have to repay?
2. Suppose you obtain a \$3,000 T-note with a 3% annual rate, paid quarterly, with maturity in 5 years. How much interest will you earn?
3. A T-bill is a type of bond that is sold at a discount over the face value. For example, suppose you buy a 13-week T-bill with a face value of \$10,000 for \$9,800. This means that in 13 weeks, the government will give you the face value, earning you \$200. What annual interest rate have you earned?
4. Suppose you are looking to buy a \$5000 face value 26-week T-bill. If you want to earn at least 1% interest, what is the most you should pay for the T-bill?
5. You deposit \$300 in an account earning 5% interest compounded annually. How much will you have in the account in 10 years?
6. How much will \$1000 deposited in an account earning 7% interest compounded annually be worth in 20 years?
7. You deposit \$2000 in an account earning 3% interest compounded monthly.
 - a. How much will you have in the account in 20 years?
 - b. How much interest will you earn?
8. You deposit \$10,000 in an account earning 4% interest compounded monthly.
 - a. How much will you have in the account in 25 years?
 - b. How much interest will you earn?
9. How much would you need to deposit in an account now in order to have \$6,000 in the account in 8 years? Assume the account earns 6% interest compounded monthly.
10. How much would you need to deposit in an account now in order to have \$20,000 in the account in 4 years? Assume the account earns 5% interest.
11. You deposit \$200 each month into an account earning 3% interest compounded monthly.
 - a. How much will you have in the account in 30 years?
 - b. How much total money will you put into the account?
 - c. How much total interest will you earn?
12. You deposit \$1000 each year into an account earning 8% compounded annually.
 - a. How much will you have in the account in 10 years?
 - b. How much total money will you put into the account?
 - c. How much total interest will you earn?

13. Jose has determined he needs to have \$800,000 for retirement in 30 years. His account earns 6% interest.
 - a. How much would you need to deposit in the account each month?
 - b. How much total money will you put into the account?
 - c. How much total interest will you earn?
14. You wish to have \$3000 in 2 years to buy a fancy new stereo system. How much should you deposit each quarter into an account paying 8% compounded quarterly?
15. You want to be able to withdraw \$30,000 each year for 25 years. Your account earns 8% interest.
 - a. How much do you need in your account at the beginning
 - b. How much total money will you pull out of the account?
 - c. How much of that money is interest?
16. How much money will I need to have at retirement so I can withdraw \$60,000 a year for 20 years from an account earning 8% compounded annually?
 - a. How much do you need in your account at the beginning
 - b. How much total money will you pull out of the account?
 - c. How much of that money is interest?
17. You have \$500,000 saved for retirement. Your account earns 6% interest. How much will you be able to pull out each month, if you want to be able to take withdrawals for 20 years?
18. Loren already knows that he will have \$500,000 when he retires. If he sets up a payout annuity for 30 years in an account paying 10% interest, how much could the annuity provide each month?
19. You can afford a \$700 per month mortgage payment. You've found a 30 year loan at 5% interest.
 - a. How big of a loan can you afford?
 - b. How much total money will you pay the loan company?
 - c. How much of that money is interest?
20. Marie can afford a \$250 per month car payment. She's found a 5 year loan at 7% interest.
 - a. How expensive of a car can she afford?
 - b. How much total money will she pay the loan company?
 - c. How much of that money is interest?
21. You want to buy a \$25,000 car. The company is offering a 2% interest rate for 48 months (4 years). What will your monthly payments be?
22. You decide finance a \$12,000 car at 3% compounded monthly for 4 years. What will your monthly payments be? How much interest will you pay over the life of the loan?

23. You want to buy a \$200,000 home. You plan to pay 10% as a down payment, and take out a 30 year loan for the rest.
- How much is the loan amount going to be?
 - What will your monthly payments be if the interest rate is 5%?
 - What will your monthly payments be if the interest rate is 6%?
24. Lynn bought a \$300,000 house, paying 10% down, and financing the rest at 6% interest for 30 years.
- Find her monthly payments.
 - How much interest will she pay over the life of the loan?

Concepts

25. Suppose you invest \$50 a month for 5 years into an account earning 8% compounded monthly. After 5 years, you leave the money, without making additional deposits, in the account for another 25 years. How much will you have in the end?
26. Suppose you put off making investments for the first 5 years, and instead made deposits of \$50 a month for 25 years into an account earning 8% compounded monthly. How much will you have in the end?
27. Mike plans to make contributions to his retirement account for 15 years. After the last contribution, he will start withdrawing \$10,000 a quarter for 10 years. Assuming Mike's account earns 8% compounded quarterly, how large must his quarterly contributions be during the first 15 years, in order to accomplish his goal?
28. Kendra wants to be able to make withdrawals of \$60,000 a year for 30 years after retiring in 35 years. How much will she have to save each year up until retirement if her account earns 7% interest?

Exploration

29. Pay day loans are short term loans that you take out against future paychecks: The company advances you money against a future paycheck. Either visit a pay day loan company, or look one up online. Be forewarned that many companies do not make their fees obvious, so you might need to do some digging or look at several companies.
- Explain the general method by which the loan works.
 - We will assume that we need to borrow \$500 and that we will pay back the loan in 14 days. Determine the total amount that you would need to pay back and the effective loan rate. The effective loan rate is the percentage of the original loan amount that you pay back. It is not the same as the APR (annual rate) that is probably published.
 - If you cannot pay back the loan after 14 days, you will need to get an extension for another 14 days. Determine the fees for an extension, determine the total amount you will be paying for the now 28 day loan, and compute the effective loan rate.

30. Suppose that 10 years ago you bought a home for \$110,000, paying 10% as a down payment, and financing the rest at 9% interest for 30 years.
- a. Let's consider your existing mortgage:
 - i. How much money did you pay as your down payment?
 - ii. How much money was your mortgage (loan) for?
 - iii. What is your current monthly payment?
 - iv. How much total interest will you pay over the life of the loan?
 - b. This year, you check your loan balance. Only part of your payments have been going to pay down the loan; the rest has been going towards interest. You see that you still have \$88,536 left to pay on your loan. Your house is now valued at \$150,000.
 - i. How much of the loan have you paid off? (i.e., how much have you reduced the loan balance by? Keep in mind that interest is charged each month - it's not part of the loan balance.)
 - ii. How much money have you paid to the loan company so far?
 - iii. How much interest have you paid so far?
 - iv. How much equity do you have in your home (equity is value minus remaining debt)
 - c. Since interest rates have dropped, you consider refinancing your mortgage at a lower 6% rate.
 - i. If you took out a new 30 year mortgage at 6% for your remaining loan balance, what would your new monthly payments be?
 - ii. How much interest will you pay over the life of the new loan?
 - d. Notice that if you refinance, you are going to be making payments on your home for another 30 years. In addition to the 10 years you've already been paying, that's 40 years total.
 - i. How much will you save each month because of the lower monthly payment?
 - ii. How much total interest will you be paying (you need to consider the amount from 2c and 3b)
 - iii. Does it make sense to refinance? (there isn't a correct answer to this question. Just give your opinion and your reason)

Statistics

Like most people, you probably feel that it is important to "take control of your life." But what does this mean? Partly it means being able to properly evaluate the data and claims that bombard you every day. If you cannot distinguish good from faulty reasoning, then you are vulnerable to manipulation and to decisions that are not in your best interest. Statistics provides tools that you need in order to react intelligently to information you hear or read. In this sense, Statistics is one of the most important things that you can study.

To be more specific, here are some claims that we have heard on several occasions. (We are *not* saying that each one of these claims is true!)

- 4 out of 5 dentists recommend Dentyne.
- Almost 85% of lung cancers in men and 45% in women are tobacco-related.
- Condoms are effective 94% of the time.
- Native Americans are significantly more likely to be hit crossing the streets than are people of other ethnicities.
- People tend to be more persuasive when they look others directly in the eye and speak loudly and quickly.
- Women make 75 cents to every dollar a man makes when they work the same job.
- A surprising new study shows that eating egg whites can increase one's life span.
- People predict that it is very unlikely there will ever be another baseball player with a batting average over 400.
- There is an 80% chance that in a room full of 30 people that at least two people will share the same birthday.
- 79.48% of all statistics are made up on the spot.

All of these claims are statistical in character. We suspect that some of them sound familiar; if not, we bet that you have heard other claims like them. Notice how diverse the examples are. They come from psychology, health, law, sports, business, etc. Indeed, data and data-interpretation show up in discourse from virtually every facet of contemporary life.

Statistics are often presented in an effort to add credibility to an argument or advice. You can see this by paying attention to television advertisements. Many of the numbers thrown about in this way do not represent careful statistical analysis. They can be misleading, and push you into decisions that you might find cause to regret. For these reasons, learning about statistics is a long step towards taking control of your life. (It is not, of course, the only step needed for this purpose.) These chapters will help you learn statistical essentials. It will make you into an intelligent consumer of statistical claims.

You can take the first step right away. To be an intelligent consumer of statistics, your first reflex must be to question the statistics that you encounter. The British Prime Minister Benjamin Disraeli famously said, "There are three kinds of lies -- lies, damned lies, and statistics." This quote reminds us why it is so important to understand statistics. So let us invite you to reform your statistical habits from now on. No longer will you blindly accept numbers or findings. Instead, you will begin to think about the numbers, their sources, and most importantly, the procedures used to generate them.

We have put the emphasis on defending ourselves against fraudulent claims wrapped up as statistics. Just as important as detecting the deceptive use of statistics is the appreciation of the proper use of statistics. You must also learn to recognize statistical evidence that supports a stated conclusion. When a research team is testing a new treatment for a disease, statistics allows them to conclude based on a relatively small trial that there is good evidence their drug is effective. Statistics allowed prosecutors in the 1950's and 60's to demonstrate racial bias existed in jury panels. Statistics are all around you, sometimes used well, sometimes not. We must learn how to distinguish the two cases.

Populations and samples

Before we begin gathering and analyzing data we need to characterize the **population** we are studying. If we want to study the amount of money spent on textbooks by a typical first-year college student, our population might be all first-year students at your college. Or it might be:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the entire United States.
- And so on.

Why is it important to specify the population? We might get different answers to our question as we vary the population we are studying. First-year students at the University of Washington might take slightly more diverse courses than those at your college, and some of these courses may require less popular textbooks that cost more; or, on the other hand, the University Bookstore might have a larger pool of used textbooks, reducing the cost of these books to the students. Whichever the case (and it is likely that some combination of these and other factors are in play), the data we gather from your college will probably not be the same as that from the University of Washington. Particularly when conveying our results to others, we want to be clear about the population we are describing with our data.

Example: A newspaper website contains a poll asking people their opinion on a recent news article. In this case, the population of the survey is all readers of the website.

If we were able to gather data on every member of our population, say the average (we will define "average" more carefully in a subsequent section) amount of money spent on textbooks by each first-year student at your college during the 2009-2010 academic year, the resulting number would be called a **parameter**. We seldom see parameters, however, since surveying an entire population is usually very time-consuming and expensive, unless the population is very small. A survey of an entire population is called a **census**.

You are probably familiar with two common censuses: the official government Census that attempts to count the population of the U.S. every ten years, and voting, which asks the opinion of all eligible voters in a district. The first of these demonstrates one additional problem with a census: the difficulty in finding and getting participation from everyone in a large population, which can bias, or skew, the results.

There are occasionally times when a census is appropriate, usually when the population is fairly small. For example, if the manager of Starbucks wanted to know the average number of hours her employees worked last week, she should be able to pull up payroll records or ask each employee directly.

Since surveying an entire population is often impractical, we usually select a **sample** to study; that is, a smaller subset of the entire population. We will discuss sampling methods in greater detail in a later section. For now, let us assume that samples are chosen in an appropriate manner. If we survey a sample, say 100 first-year students at your college, and find the average amount of money spent by these students on textbooks, the resulting number is called a **statistic**. The **sample size** was 100.

Example: A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. What is the sample and population?

The sample is the 500 shoppers questioned. The population is less clear. While the intended population of this survey was Tacoma citizens, the effective population was mall shoppers. There is no reason to assume that the 500 shoppers questioned would be representative of all Tacoma citizens.

Categorizing data

Once we have gathered data, we might wish to classify it. Roughly speaking, data can be classified as categorical data or quantitative data.

Also known as qualitative data, categorical data, not surprisingly, are pieces of information that allow us to classify the objects under investigation into various categories. For example, we might conduct a survey to determine the name of the favorite movie that each person in a math class saw in a movie theater. When we conduct such a survey, the responses would look like: *Finding Nemo*, *The Hulk*, or *Terminator 3: Rise of the Machines*. We might count the number of people who give each answer, but the answers themselves do not have any numerical values: we cannot perform computations with an answer like "*Finding Nemo*."

Quantitative data, on the other hand, are responses that are numerical in nature and with which we can perform meaningful arithmetic calculations. Examples of survey questions that would elicit quantitative answers are: the number of movies you have seen in a movie theater in the past 12 months (0, 1, 2, 3, 4, ...); the running time of the movie you saw most recently (104 minutes, 137 minutes, 104 minutes, ...); the amount of money you paid for a movie ticket the last time you went to a movie theater (\$5.50, \$7.75, \$9, ...).

Sometimes, determining whether or not data is categorical or quantitative can be a bit trickier. Suppose we gather respondents' ZIP codes in a survey to track their geographical location. ZIP codes are numbers, but we can't do any meaningful mathematical calculations with them (it doesn't make sense to say that 98036 is "twice" 49018 — that's like saying that Lynnwood, WA is "twice" Battle Creek, MI, which doesn't make sense at all), so ZIP codes are really categorical data.

Similarly, a survey about the movie you most recently attended might include a question such as "How would you rate the movie you just saw?" with these possible answers:

- 1 - it was awful
- 2 - it was just OK
- 3 - I liked it
- 4 - it was great
- 5 - best movie ever!

Again, there are numbers associated with the responses, but we can't really do any calculations with them: a movie that rates a 4 is not necessarily twice as good as a movie that rates a 2, whatever that means; if two people see the movie and one of them thinks it stinks and the other thinks it's the best ever it doesn't necessarily make sense to say that "on average they liked it."

As we study movie-going habits and preferences, we shouldn't forget to specify the population under consideration. If we survey 3-7 year-olds the runaway favorite might be *Finding Nemo*. 13-17 year-olds might prefer *Terminator 3*. And 33-37 year-olds might prefer...well, *Finding Nemo*.

Sampling methods

As we mentioned in a previous section, the first thing we should do before conducting a survey is to identify the population that we want to study. Suppose we are hired by a politician to determine the amount of support he has among the electorate should he decide to run for another term. What population should we study? Every person in the district? Not every person is eligible to vote, and regardless of how strongly someone likes or dislikes the candidate, they don't have much to do with him being re-elected if they are not able to vote.

What about eligible voters in the district? That might be better, but if someone is eligible to vote but does not register by the deadline, they won't have any say in the election either. What about registered voters? Many people are registered but choose not to vote. What about "likely voters?"

This is the criteria used in much political polling, but it is sometimes difficult to define a "likely voter." Is it someone who voted in the last election? In the last general election? In the last presidential election? Should we consider someone who just turned 18 a "likely voter?" They weren't eligible to vote in the past, so how do we judge the likelihood that they will vote in the next election?

In November 1998 former professional wrestler Jesse "The Body" Ventura was elected governor of Minnesota. Up until right before the election, most polls showed he had little chance of winning. There were several contributing factors to the polls not reflecting the actual intent of the electorate:

- Ventura was running on a third-party ticket and most polling methods are better suited to a two-candidate race.
- Many respondents to polls may have been embarrassed to tell pollsters that they were planning to vote for a professional wrestler.
- The mere fact that the polls showed Ventura had little chance of winning might have prompted some people to vote for him in protest to send a message to the major-party candidates.

But one of the major contributing factors was that Ventura recruited a substantial amount of support from young people, particularly college students, who had never voted before and who registered specifically to vote in the gubernatorial election. The polls did not deem these young people likely voters (since in most cases young people have a lower rate of voter registration and a turnout rate for elections) and so the polling samples were subject to **sampling bias**: they omitted a portion of the electorate that was weighted in favor of the winning candidate.

So even identifying the population can be a difficult job, but once we have identified the population, how do we choose an appropriate sample? Remember, although we would prefer to survey all members of the population, this is usually impractical unless the population is very small, so we choose a sample. There are many ways to sample a population, but there is one goal we need to keep in mind: we would like the sample to be *representative of the population*.

Returning to our hypothetical job as a political pollster, we would not anticipate very accurate results if we drew all of our samples from among the customers at a Starbucks, nor would we expect that a sample drawn entirely from the membership list of the local Elks club would provide a useful picture of district-wide support for our candidate.

One way to ensure that the sample has a reasonable chance of mirroring the population is to employ *randomness*. In a truly **simple random sample**, each member of the population has an equal probability of being chosen. If we could somehow identify all likely voters in the state, put each of their names on a piece of paper, toss the slips into a (very large) hat and draw 1000 slips out of the hat, we would have a simple random sample. In practice, computers are better suited for this sort of endeavor than millions of slips of paper and extremely large headgear.

It is always possible, however, that even a random sample might end up being biased or skewed. If we repeatedly take samples of 1000 people from among the population of likely voters in the state of Washington, some of these samples might tend to have a slightly higher percentage of Democrats (or Republicans) than does the general population; some samples might include more older people and some samples might include more younger people; etc. In most cases, this **sampling variability** is not significant.

To help account for this, pollsters might instead use a **stratified sample**. Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents. In a sample of 1000 people, they would then expect to get about 390 Democrats, 370 Republicans and 240 independents. To accomplish this, they could randomly select 390 people from among those voters known to

be Democrats, 370 from those known to be Republicans, and 240 from those with no party affiliation. Stratified sampling can also be used to select a sample with people in desired age groups, a specified mix ratio of males and females, etc.

A variation on this technique is called **quota sampling**, in which pollsters call people at random, but once they have met their quota of 390 Democrats, say, they only gather people who do not identify themselves as a Democrat. You may have had the experience of being called by a telephone pollster who started by asking you your age, income, etc. and then thanked you for your time and hung up before asking any "real" questions. Most likely, they already had contacted enough people in your demographic group and were looking for people who were older or younger, richer or poorer, etc. Quota sampling is usually a bit easier than stratified sampling, but also is not as random as a simple random sample.

Other sampling methods include **systematic sampling**, in which a sample is chosen in a systematic way, such as selecting every 100th name in the phone book. Systematic sampling is not as random as a simple random sample (if your name is Albert Aardvark and your sister Alexis Aardvark is right after you in the phone book, there is no way you could both end up in the sample) but it can yield acceptable samples.

Perhaps the worst types of sampling methods are **convenience samples** in which samples are chosen in a way that is convenient for the pollster, such as standing on a street corner and interviewing the first 100 people who agree to speak with you. One type of convenience sample is a **self-selected sample**, or **voluntary response sample**, in which respondents volunteer to participate. Usually such samples are skewed towards people who have a particularly strong opinion about the subject of the survey or who just have way too much time on their hands and enjoy taking surveys. A survey on a website is an example of a self-selected sample.

How to mess things up before you start

There are number of ways that a study can be ruined before you even start collecting data. The first we have already explored – **sampling** or **selection bias**, which is when the sample is not representative of the population. One example of this is **voluntary response bias**, which is bias introduced by only collecting data from those who volunteer to participate. This is not the only potential source of bias.

Consider a recent study which found that chewing gum may raise math grades in teenagers¹. This study was conducted by the Wrigley Science Institute, a branch of the Wrigley chewing gum company. This is an example of a **self-interest study**; one in which the researches have a vested interest in the outcome of the study. While this does not necessarily ensure that the study was biased, it certainly suggests that we should subject the study to extra scrutiny.

Another source of bias is **response bias**, in which something causes the responder to give inaccurate responses. This can be from sources as innocent as bad memory when answering a question like “when was the last time you visited your doctor”, or as intentional as

¹ Reuters. http://news.yahoo.com/s/nm/20090423/od_uk_nm/oukoe_uk_gum_learning. Retrieved 4/27/09

pressuring by the pollster. Consider, for example, how many voting initiative petitions people sign without even reading them.

Another source is **perceived lack of anonymity**, which can influence people. For example, in a poll about race relations, the respondent might not want to be perceived as racist even if they are, and give an untruthful answer. In other cases, answering truthfully might have consequences. For example, if an employer puts out a survey asking their employees if they have a drug abuse problem and need treatment help, the responses might not be accurate if the employees do not feel their responses are anonymous.

Another example of response bias is **loaded** or **leading questions**. For example, consider how people would be likely to respond to the question “Do you support the invasion and occupation of Iraq?” compared to “Do you support the liberation of Iraq from dictator rule?” These are leading questions – questions whose wording lead the respondent towards an answer - and can occur unintentionally. Also a concern is **question order**, where the order of questions changes the results. A psychology researcher provides an example²:

“My favorite finding is this: we did a study where we asked students, 'How satisfied are you with your life? How often do you have a date?' The two answers were not statistically related - you would conclude that there is no relationship between dating frequency and life satisfaction. But when we reversed the order and asked, 'How often do you have a date? How satisfied are you with your life?' the statistical relationship was a strong one. You would now conclude that there is nothing as important in a student's life as dating frequency.”

In addition to response bias, there is also **non-response bias**, introduced by people refusing to participate in a study or dropping out of an experiment. When people refuse to participate, we can no longer be so certain that our sample is representative of the population. Suppose, for example, we used a telephone poll to ask the question “Do you often have time to relax and read a book?”, and 50% of the people we called refused to answer the survey. It is unlikely that our results will be representative of the entire population.

Experiments

So far, we have primarily discussed **observational studies** – studies in which conclusions would be drawn from observations of a sample or the population. In some cases these observations might be unsolicited, such as studying the percentage of cars that turn right at a red light even when there is a “no turn on red” sign. In other cases the observations are solicited, like in a survey or a poll.

In contrast, it is common to use **experiments** when exploring how subjects react to an outside influence. In an experiment, some kind of **treatment** is applied to the subjects and the results are measured and recorded. Here are some examples of experiments:

² Swartz, Norbert. <http://www.umich.edu/~newsinfo/MT/01/Fal01/mt6f01.html>. Retrieved 3/31/2009

- 1) A pharmaceutical company tests a new medicine for treating Alzheimer's disease by administering the drug to 50 elderly patients with recent diagnoses. The treatment here is the new drug.
- 2) A gym tests out a new weight loss program by enlisting 30 volunteers to try out the program. The treatment here is the new program.
- 3) You test a new kitchen cleaner by buying a bottle and cleaning your kitchen. The new cleaner is the treatment.
- 4) A psychology researcher explores the effect of music on temperament by measuring people's temperament while listening to different types of music. The music is the treatment.

When conducting experiments, it is essential to isolate the treatment being tested. For example, suppose a middle school (junior high) finds that their students are not scoring well on the state's standardized math test. They decide to run an experiment to see if an alternate curriculum would improve scores. To run the test, they hire a math specialist to come in and teach a class using the new curriculum. To their delight, they see an improvement in test scores.

The difficulty with this scenario is that it is not clear whether the curriculum is responsible for the improvement, or whether the improvement is due to a math specialist teaching the class. This is called **confounding** – when it is not clear which factor or factors caused the observed effect. Confounding is the downfall of many experiments, though sometimes it is hidden.

For example, a drug company study about a weight loss pill might report that people lost an average of 8 pounds while using their new drug. However, in the fine print you find a statement saying that participants were encouraged to also diet and exercise. It is not clear in this case whether the weight loss is due to the pill, to diet and exercise, or a combination of both. In this case confounding has occurred.

There are a number of measures that can be introduced to help reduce the likelihood of confounding. The primary measure is to use a **control group**. When using a control group, the participants are divided into two or more groups, typically a control group and a treatment group. The treatment group receives the treatment being tested; the control group does not receive the treatment. Ideally, the groups are otherwise as similar as possible, isolating the treatment as the only potential source of difference between the groups. For this reason, the method of dividing groups is important. Some researchers attempt to ensure that the groups have similar characteristics (same number of females, same number of people over 50, etc.), but it is nearly impossible to control for every characteristic. Because of this, random assignment is very commonly used.

For example, to determine if a two day prep course would help high school students improve their scores on the SAT test, a group of students was randomly divided into two subgroups. The first group, the treatment group, was given a two day prep course. The second group, the control group, was not given the prep course. Afterwards, both groups were given the SAT. In another example a company testing a new plant food could grow two crops of plants, the treatment group receiving the new plant food and the control group not. The crop yield would then be compared.

Sometimes not giving the control group anything does not completely control for confounding variables. For example, suppose a medicine study is testing a new headache pill by giving the treatment group the pill and the control group nothing. If the treatment group showed improvement, we would not know whether it was due to the medicine in the pill, or a response to have taken any pill. This is called a **placebo effect** – when the effectiveness of a treatment is influenced by the patient’s perception of how effective they think the treatment will be. For example, a study found that when doing painful dental tooth extractions, patients told they were receiving a strong painkiller while actually receiving a saltwater injection found as much pain relief as patients receiving a dose of morphine.³

To control for the placebo effect, a **placebo**, or dummy treatment, is often given to the control group. This way, both groups are truly identical except for the specific treatment given. For the study of a medicine, a sugar pill could be used as a placebo. In a study on the effect of alcohol on memory, a non-alcoholic beer might be given to the control group as a placebo. In a study of a frozen meal diet plan, the treatment group would receive the diet food, and the control could be given standard frozen meals stripped of their original packaging. An experiment that gives the control group a placebo is called a **placebo controlled experiment**.

In some cases, it is more appropriate to compare to a conventional treatment than a placebo. For example, in a cancer research study, it would not be ethical to deny any treatment to the control group or to give a placebo treatment. In this case, the currently acceptable medicine would be given to the second group, called a **comparison group** in this case. In our SAT test example, the non-treatment group would most likely be encouraged to study on their own, rather than be asked to not study at all, to provide a meaningful comparison.

When using a placebo, it would defeat the purpose if the participant knew they were receiving the placebo. A **blind study** is one in which the participant does not know whether or not they are receiving the treatment or a placebo. In a study about anti-depression medicine, you would not want the psychological evaluator to know whether the patient is in the treatment or control group either, as it might influence their evaluation. A **double-blind study** is one in which those interacting with the participants don’t know who is in the treatment group and who is in the control group.

It should be noted that not every experiment needs a control group. For example, if a researcher is testing whether a new fabric can withstand fire, she simply needs to torch multiple samples of the fabric – there is no need for a control group.

³ Levine JD, Gordon NC, Smith R, Fields HL. (1981) Analgesic responses to morphine and placebo in individuals with postoperative pain. *Pain*. 10:379-89.

Exercises

Skills

1. A political scientist surveys 28 of the current 106 representatives in a state's congress. Of them, 14 said they were supporting a new education bill, 12 said there were not supporting the bill, and 2 were undecided.
 - a. What is the population of this survey?
 - b. What is the size of the population?
 - c. What is the size of the sample?
 - d. Give the sample statistic for the proportion of voters surveyed who said they were supporting the education bill.
 - e. Based on this sample, we might expect how many of the representatives to support the education bill?
2. The city of Raleigh has 9500 registered voters. There are two candidates for city council in an upcoming election: Brown and Feliz. The day before the election, a telephone poll of 350 randomly selected registered voters was conducted. 112 said they'd vote for Brown, 238 said they'd vote for Feliz, and 31 were undecided.
 - a. What is the population of this survey?
 - b. What is the size of the population?
 - c. What is the size of the sample?
 - d. Give the sample statistic for the proportion of voters surveyed who said they'd vote for Brown.
 - e. Based on this sample, we might expect how many of the 9500 voters to vote for Brown?
3. Identify the most relevant source of bias in this situation: A survey asks the following: Should the mall prohibit loud and annoying rock music in clothing stores catering to teenagers?
4. Identify the most relevant source of bias in this situation: To determine opinions on voter support for a downtown renovation project, a surveyor randomly questions people working in downtown businesses.
5. Identify the most relevant source of bias in this situation: A survey asks people to report their actual income and the income they reported on their IRS tax form.
6. Identify the most relevant source of bias in this situation: A survey randomly calls people from the phone book and asks them to answer a long series of questions.
7. Identify the most relevant source of bias in this situation: A survey asks the following: Should the death penalty be permitted if innocent people might die?
8. Identify the most relevant source of bias in this situation: A study seeks to investigate whether a new pain medication is safe to market to the public. They test by randomly selecting 300 men from a set of volunteers.

9. In a study, you ask the subjects their age in years. Is this data qualitative or quantitative?
10. In a study, you ask the subjects their gender. Is this data qualitative or quantitative?
11. Does this describe an observational study or an experiment: The temperature on randomly selected days throughout the year was measured.
12. Does this describe an observational study or an experiment? A group of students are told to listen to music while taking a test and their results are compared to a group not listening to music.
13. In a study, the sample is chosen by separating all cars by size, and selecting 10 of each size grouping. What is the sampling method?
14. In a study, the sample is chosen by writing everyone's name on a playing card, shuffling the deck, then choosing the top 20 cards. What is the sampling method?
15. A team of researchers is testing the effectiveness of a new HPV vaccine. They randomly divide the subjects into two groups. Group 1 receives new HPV vaccine, and Group 2 receives the existing HPV vaccine. The patients in the study do not know which group they are in.
 - a. Which is the treatment group?
 - b. Which is the control group (if there is one)?
 - c. Is this study blind, double-blind, or neither?
 - d. Is this best described as an experiment, a controlled experiment, or a placebo controlled experiment?
16. For the clinical trials of a weight loss drug containing *Garcinia cambogia* the subjects were randomly divided into two groups. The first received an inert pill along with an exercise and diet plan, while the second received the test medicine along with the same exercise and diet plan. The patients do not know which group they are in, nor do the fitness and nutrition advisors.
 - a. Which is the treatment group?
 - b. Which is the control group (if there is one)?
 - c. Is this study blind, double-blind, or neither?
 - d. Is this best described as an experiment, a controlled experiment, or a placebo controlled experiment?

Concepts

17. A teacher wishes to know whether the males in his/her class have more conservative attitudes than the females. A questionnaire is distributed assessing attitudes.
 - a. Is this a sampling or a census?
 - b. Is this an observational study or an experiment?
 - c. Are there any possible sources of bias in this study?

18. A study is conducted to determine whether people learn better with spaced or massed practice. Subjects volunteer from an introductory psychology class. At the beginning of the semester 12 subjects volunteer and are assigned to the massed-practice group. At the end of the semester 12 subjects volunteer and are assigned to the spaced-practice condition.
- Is this a sampling or a census?
 - Is this an observational study or an experiment?
 - This study involves two kinds of non-random sampling: (1) Subjects are not randomly sampled from some specified population and (2) Subjects are not randomly assigned to groups. Which problem is more serious? What affect on the results does each have?
19. A farmer believes that playing Barry Manilow songs to his peas will increase their yield. Describe a controlled experiment the farmer could use to test his theory.
20. A sports psychologist believes that people are more likely to be extroverted as adults if they played team sports as children. Describe two possible studies to test this theory. Design one as an observational study and the other as an experiment. Which is more practical?

Exploration

21. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a different set of 50 AIDS patients from the start of treatment until their deaths.
- What is the population of this study?
 - List two reasons why the data may differ.
 - Can you tell if one researcher is correct and the other one is incorrect? Why?
 - Would you expect the data to be identical? Why or why not?
 - If the first researcher collected her data by randomly selecting 40 states, then selecting 1 person from each of those states. What sampling method is that?
 - If the second researcher collected his data by choosing 40 patients he knew. What sampling method would that researcher have used? What concerns would you have about this data set, based upon the data collection method?
22. Find a newspaper or magazine article, or the online equivalent, describing the results of a recent study (the results of a poll are not sufficient). Give a summary of the study's findings, then analyze whether the article provided enough information to determine the validity of the conclusions. If not, produce a list of things that are missing from the article that would help you determine the validity of the study. Look for the things discussed in the text: population, sample, randomness, blind, control, placebos, etc.

Describing Data

Once we have collected data from surveys or experiments, we need to summarize and present the data in a way that will be meaningful to the reader. We will begin with graphical presentations of data then explore numerical summaries of data.

Presenting Categorical Data Graphically

Categorical, or qualitative, data are pieces of information that allow us to classify the objects under investigation into various categories. We usually begin working with categorical data by summarizing the data into a **frequency table** which is just a table with two columns, one for categories and another for the frequencies with which the items in the categories occur.

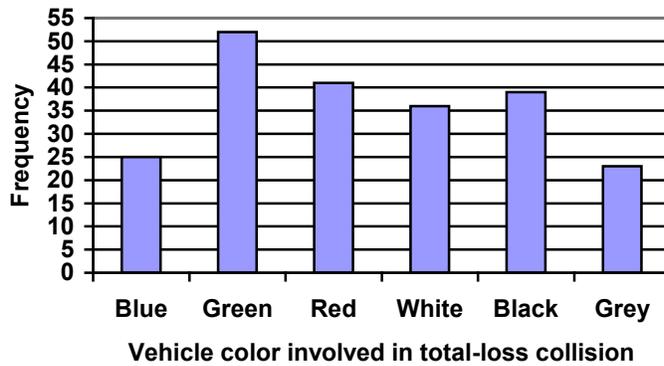
Example: An insurance company determines vehicle insurance premiums based on known risk factors. If a person is considered a higher risk, their premiums will be higher. One potential factor is the color of your car. The insurance company believes that people with some color cars are more likely to get in accidents. To research this, they examine police reports for recent total-loss collisions. The data is summarized in the frequency table below.

Color	Frequency
Blue	25
Green	52
Red	41
White	36
Black	39
Grey	23

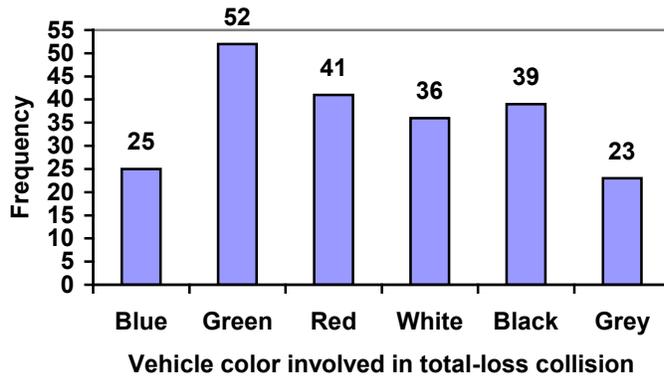
Sometimes we need an even more intuitive way of displaying data. This is where charts and graphs come in. There are many, many ways of displaying data graphically, but we will concentrate on one very useful type of graph called a bar graph. In this section we will work with bar graphs that display categorical data; the next section will be devoted to bar graphs that display quantitative data.

To construct a bar graph, we need to draw a vertical axis and a horizontal axis. The vertical direction will have a scale and measure the frequency of each category; the horizontal axis has no scale in this instance. The construction of a bar chart is most easily described by use of an example.

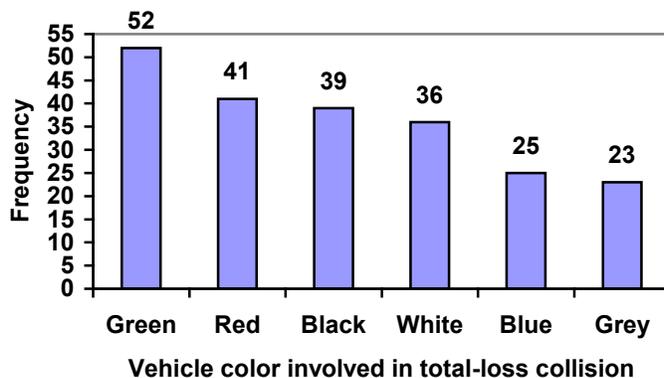
Using our car data from above, note the highest frequency is 52, so our vertical axis needs to go from 0 to 52, but we might as well use 0 to 55, so that we can put a hash mark every 5 units:



Notice that the height of each bar is determined by the frequency of the corresponding color. The horizontal gridlines are a nice touch, but not necessary. In practice, you will find it useful to draw bar graphs using graph paper, so the gridlines will already be in place. Instead of gridlines, we might also list the frequencies at the top of each bar, like this:



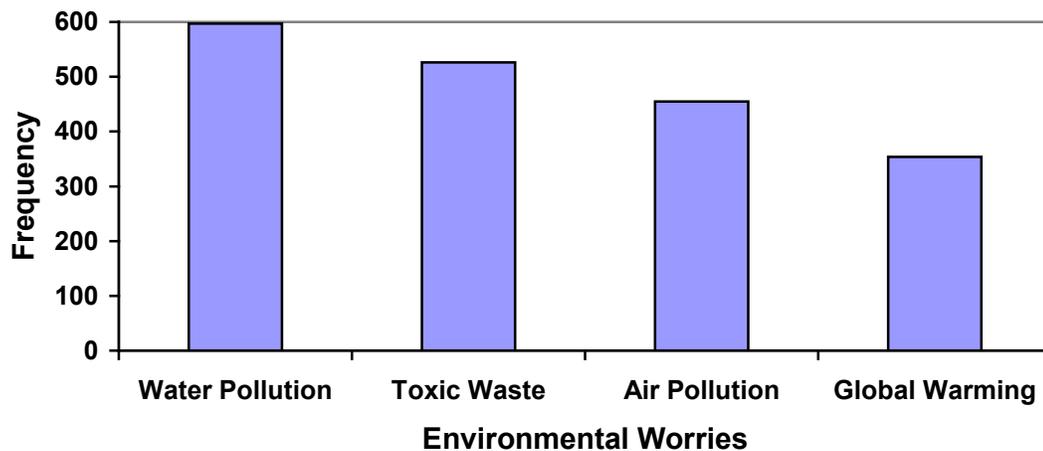
In this case, our chart might benefit from being reordered from largest to smallest frequency values. This arrangement can make it easier to compare similar values in the chart, even without gridlines. When we arrange the categories in decreasing frequency order like this, it is called a **Pareto chart**.



Example: In a recent survey¹, adults were asked whether they personally worried about a variety of environmental concerns. The number (out of 1012 surveyed) indicating that they worried “a great deal” about some selected concerns is summarized below.

Environmental Issue	Frequency
Pollution of drinking water	597
Contamination of soil and water by toxic waste	526
Air pollution	455
Global warming	354

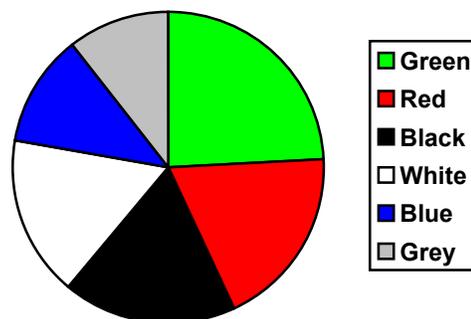
This data could be shown graphically in a bar graph:



Another type of chart in frequent use is a **pie chart**, which is basically a circle with wedges cut of varying sizes marked out like slices of pie or pizza. The relative sizes of the wedges correspond to the relative frequencies of the categories.

Example: For our vehicle color data, a pie chart might look like this:

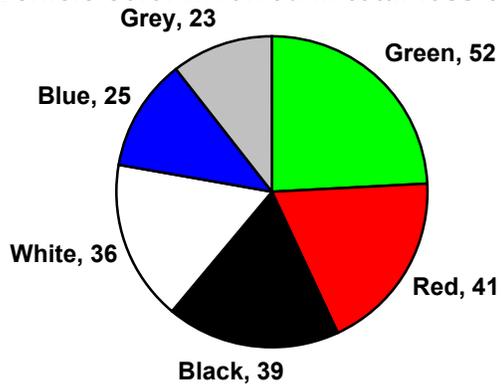
Vehicle color involved in total-loss collisions



Pie charts can often benefit from including frequencies or relative frequencies (percents) in the chart next to the pie slices. Often having the category names next to the pie slices also makes the chart clearer.

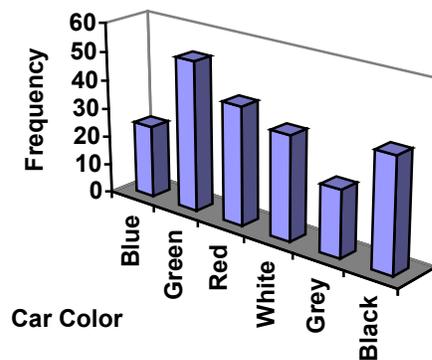
¹ Gallup Poll. March 5-8, 2009. <http://www.pollingreport.com/enviro.htm>

Vehicle color involved in total-loss collisions



Pie charts look nice, but are harder to draw by hand than bar charts since to draw them accurately we would need to compute the angle each wedge cuts out of the circle, then measure the angle with a protractor. Computers are much better suited to drawing pie charts. Common software programs like Microsoft Word or Excel, or OpenOffice.org Write or Calc are able to create bar graphs, pie charts, and other graph types. There are also numerous online tools that can create graphs².

Don't get fancy! People sometimes add features to graphs that don't help to convey their information. For example, 3-dimensional bar charts like the one shown below are usually not as effective as their two-dimensional counterparts.



Here is another way that fanciness can lead to trouble. Instead of plain bars, it is tempting to substitute meaningful images. This type of graph is called a **pictogram**.

Example: a labor union might produce the graph to the right to show the difference between the average manager salary and the average worker salary.

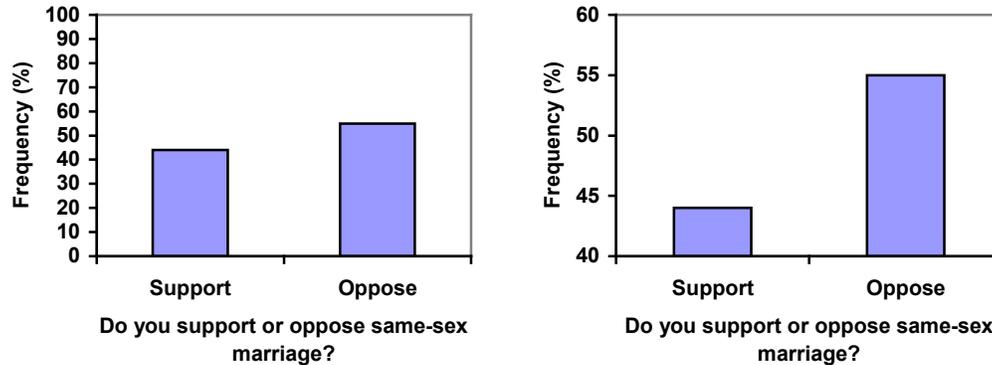


Looking at the picture, it would be reasonable to guess that the manager salaries is 4 times as large as the worker salaries – the area of the bag looks about 4 times as large. However, the manager salaries are in fact only twice as large as worker salaries, which were reflected in the picture by making the manager bag twice as tall.

² For example: <http://nces.ed.gov/nceskids/createAgraph/>

Another distortion in bar charts results from setting the baseline to a value other than zero. The baseline is the bottom of the vertical axis, representing the least number of cases that could have occurred in a category. Normally, this number should be zero.

Example: Compare the two graphs below showing support for same-sex marriage rights from a poll taken in December 2008³. Once again, the difference in areas suggests a different story than the true differences in percentages; the second graph makes it look like twice as many people oppose marriage rights as support it.



Presenting Quantitative Data Graphically

Quantitative, or numerical, data can also be summarized into frequency tables.

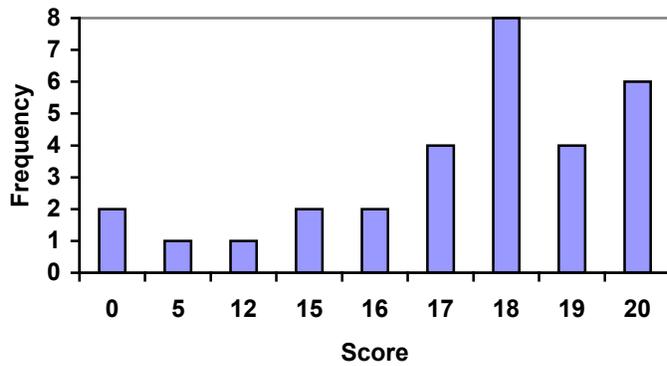
Example: A teacher records scores on a 20-point quiz for the 30 students in his class. The scores are:

19 20 18 18 17 18 19 17 20 18 20 16 20 15 17 12 18 19 18 19 17 20 18 16 15 18 20 5 0 0

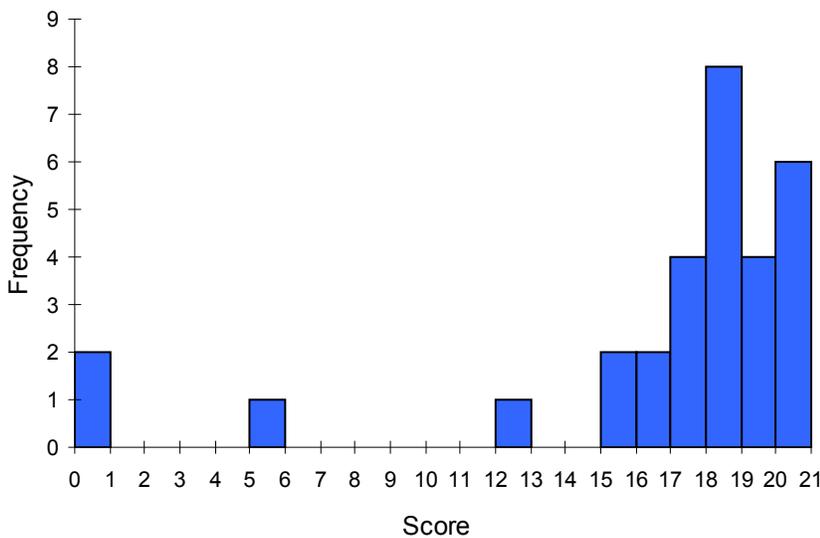
Score	Frequency
0	2
5	1
12	1
15	2
16	2
17	4
18	8
19	4
20	6

It would be possible to create a standard bar chart from this summary, like we did for categorical data:

³CNN/Opinion Research Corporation Poll. Dec 19-21, 2008, from <http://www.pollingreport.com/civil.htm>

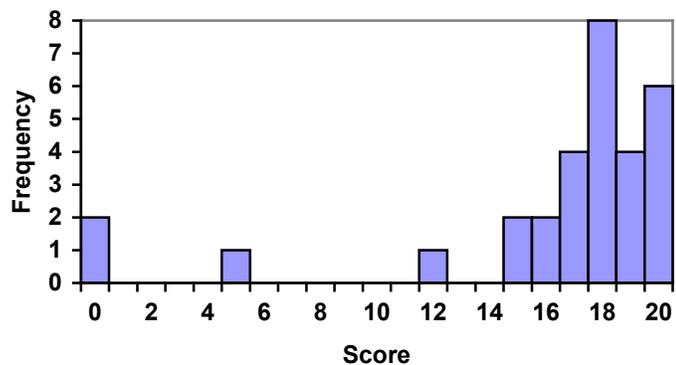


However, since the scores are numerical values, it would be more correct to treat the horizontal axis as a number line. This type of graph is called a **histogram**.



Notice that in the histogram, a bar represents values on the horizontal axis from that on the left hand-side of the bar up to, but not including, the value on the right hand side of the bar. Some people choose to have bars start at $\frac{1}{2}$ values to avoid this ambiguity.

Unfortunately, not a lot of common software packages can correctly graph a histogram. About the best you can do in Excel or Word is a bar graph with no gap between the bars and spacing added to simulate a numerical horizontal axis.



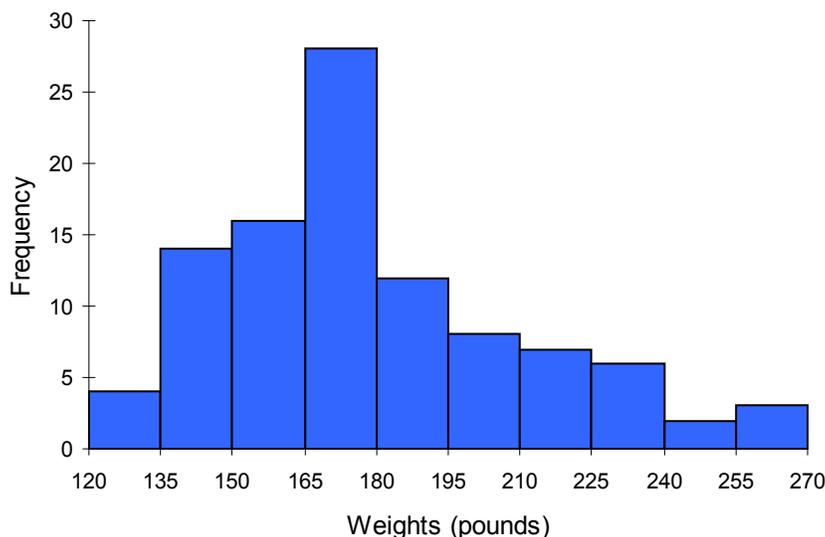
Now, suppose that we have collected weights from 100 male subjects as part of a nutrition study. Listing every possible weight with a frequency would create an exceptionally long frequency table, and probably would not reveal any patterns. For this reason, it is common with quantitative data to group data into **class intervals**. In general, we define class intervals so that:

- 1) Each interval is equal in size. For example, if the first class contains values from 120-129, the second class should include values from 130-139.
- 2) We have somewhere between 5 and 20 classes, typically, depending upon the number of data we're working with.

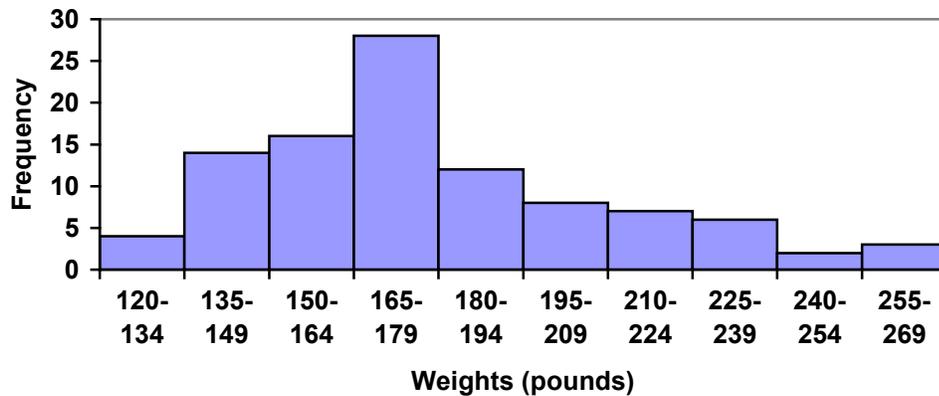
Example: For our weight data, we have values ranging from a low of 121 pounds to a high of 263 pounds, given a total span of $263 - 121 = 142$. We could create 7 intervals with a width of around 20, 14 intervals with a width of around 10, or somewhere in between. Often time we have to experiment with a few possibilities to find something that represents the data well. Let us try using an interval width of 15. We could start at 121, or at 120 since it is a nice round number.

Interval	Frequency
120 - 134	4
135 - 149	14
150 - 164	16
165 - 179	28
180 - 194	12
195 - 209	8
210 - 224	7
225 - 239	6
240 - 254	2
255 - 269	3

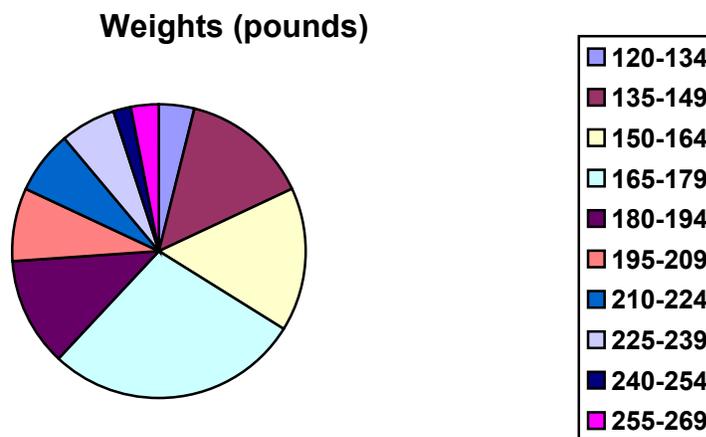
A histogram of this data would look like:



In many software packages, you can create a graph similar to a histogram by putting the class intervals as the labels on a bar chart.



Other graph types such as pie charts are possible for quantitative data. The usefulness of different graph types will vary depending upon the number of intervals and the type of data being represented. For example, a pie chart of our weight data is difficult to read because of the quantity of intervals we used.

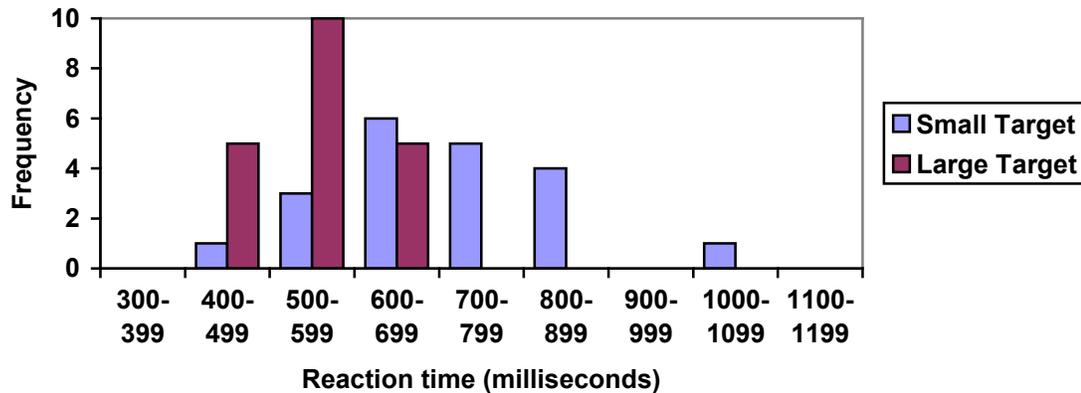


When collecting data to compare two groups, it is desirable to create a graph that compares quantities.

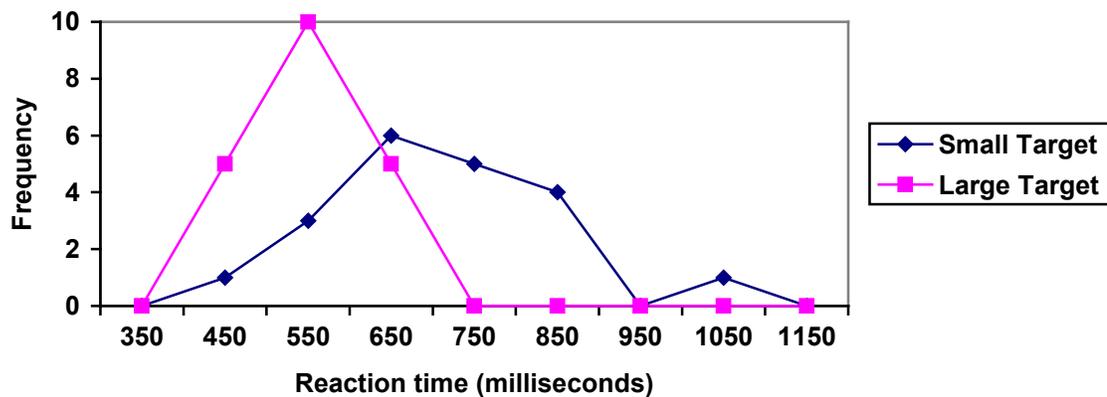
Example: The data below came from a task in which the goal is to move a computer mouse to a target on the screen as fast as possible. On 20 of the trials, the target was a small rectangle; on the other 20, the target was a large rectangle. Time to reach the target was recorded on each trial.

Interval (milliseconds)	Frequency small target	Frequency large target
300-399	0	0
400-499	1	5
500-599	3	10
600-699	6	5
700-799	5	0
800-899	4	0
900-999	0	0
1000-1099	1	0
1100-1199	0	0

One option to represent this data would be a comparative histogram or bar chart, in which bars for the small target group and large target group are placed next to each other.



An alternative representation is a **frequency polygon**. A frequency polygon starts out like a histogram, but instead of drawing a bar, a point is placed in the midpoint of each interval at height equal to the frequency. Typically the points are connected with straight lines to emphasize the distribution of the data. This graph makes it easier to see that reaction times were generally smaller for the larger target, and that the reaction times for the smaller target were more spread out.



Numerical Summaries of Data

It is often desirable to use a few numbers to summarize a distribution. One important aspect of a distribution is where its center is located. Measures of central tendency are discussed first. A second aspect of a distribution is how spread out it is. In other words, how much the data in the distribution vary from one another. The second section describes measures of variability.

Measures of Central Tendency

Let's begin by trying to find the most "typical" value of a data set.

Note that we just used the word "typical" although in many case you might think of using the word "average." We need to be careful with the word "average" as it means different things to different people in different contexts. One of the most common uses of the word "average" is what mathematicians and statisticians call the **arithmetic mean**, or just plain old **mean** for short. "Arithmetic mean" sounds rather fancy, but you have likely calculated a mean many times without realizing it; the mean is what most people think of when they use the word "average": the sum of the data values divided by the number of values.

Example: Marci's exam scores for her last math class were: 79, 86, 82, 94. The mean of these values would be:

$$\frac{79 + 86 + 82 + 94}{4} = 85.25$$

Typically we round means to one more decimal place than the original data had. In this case, that would 85.3.

Example: The number of touchdown (TD) passes thrown by each of the 31 teams in the National Football League in the 2000 season are shown below.

37 33 33 32 29 28 28 23 22 22 22 21 21 21 20
20 19 19 18 18 18 18 16 15 14 14 14 12 12 9 6

Adding these values, we get 634 total TDs. Dividing by 31, the number of data values, we get $634/31 = 20.4516$. It would be appropriate to round this to 20.5.

It would be most correct for us to report that "The mean number of touchdown passes thrown in the NFL in the 2000 season was 20.5 passes," but it is not uncommon to see the more casual word "average" used in place of "mean".

Example: The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars. The results are summarized in a frequency table below.

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

Calculating the mean by hand could get tricky if we try to type in all 100 values:

$$\frac{\overbrace{15+\cdots+15}^{6 \text{ terms}} + \overbrace{20+\cdots+20}^{8 \text{ terms}} + \overbrace{25+\cdots+25}^{11 \text{ terms}} + \cdots}{100}$$

We could calculate this more easily by noticing that adding 15 to itself six times is the same as $15 \cdot 6 = 90$. Using this simplification, we get

$$\frac{15 \cdot 6 + 20 \cdot 8 + 25 \cdot 11 + 30 \cdot 17 + 35 \cdot 19 + 40 \cdot 20 + 45 \cdot 12 + 50 \cdot 7}{100} = \frac{3390}{100} = 33.9$$

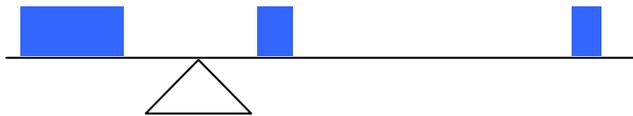
So the mean household income of our sample is 33.9 thousand dollars (\$33,900).

Now, suppose a new family moves into the neighborhood that has a household income of \$5 million (\$5000 thousand). Adding this to our sample, our mean is now

$$\frac{15 \cdot 6 + 20 \cdot 8 + 25 \cdot 11 + 30 \cdot 17 + 35 \cdot 19 + 40 \cdot 20 + 45 \cdot 12 + 50 \cdot 7 + 5000 \cdot 1}{100} = \frac{8390}{100} = 83.9$$

While 83.9 thousand dollars (\$83,900) is the correct mean household income, it no longer represents a “typical” value.

Imagine the data values on a see-saw or balance scale. The mean is the value that keeps the data in balance, like in the picture below.



If we graph our household data, the \$5 million data value is so far out to the right that the mean has to adjust up to keep things in balance



For this reason, when working with data that have **outliers** – values far outside the primary grouping – it is common to use a different measure of center, the **median**. Just like the median of a freeway is the strip that runs down the middle, the median of a set of numbers is the value in the middle.

To find the median, begin by listing the data in order from smallest to largest. If the number of data values, N , is odd, then the median is the middle data value. This value can be found by rounding $N/2$ up to the next whole number. If the number of data values is even, there is no one middle value, so we find the mean of the two middle values (values $N/2$ and $N/2 + 1$)

Example: Returning to the football touchdown data, we would start by listing the data in order. Luckily, it was already in decreasing order, so we can work with it.

37 33 33 32 29 28 28 23 22 22 22 21 21 21 20
20 19 19 18 18 18 18 16 15 14 14 14 12 12 9 6

Since there are 31 data values, an odd number, the median will be the middle number, the 16th data value ($31/2 = 15.5$, round up to 16, leaving 15 values below and 15 above). The 16th data value is 20, so the median number of touchdown passes in the 2000 season was 20 passes. Notice that for this data, the median is fairly close to the mean we calculated earlier, 20.5.

Example: Find the median of these quiz scores: 5 10 8 6 4 8 2 5 7 7

We start by listing the data in order: 2 4 5 5 6 7 7 8 8 10

Since there are 10 data values, an even number, there is no one middle number. So we find the mean of the two middle numbers, 6 and 7. Finding the mean: $(6+7)/2 = 6.5$. So the median quiz score was 6.5.

Example: Let us return now to our original household income data

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

Here we have 100 data values. If we didn't already know that, we could find it by adding the frequencies. Since 100 is an even number, we need to find the mean of the middle two data values - the 50th and 51st data values. To find these, we start counting up from the bottom:

Values 1 to 6 are \$15
 Values 7 to $(6+8)=14$ are \$20
 Values 15 to $(14+11)=25$ are \$25
 Values 26 to $(25+17)=42$ are \$30
 Values 43 to $(42+19)=61$ are \$35

From this we can tell that values 50 and 51 will be \$35 thousand, and the mean of these values is \$35 thousand. So the median income in this neighborhood is \$35 thousand.

Note that if we add in the new neighbor with a \$5 million household income, it will not affect the median in this case. The median is not swayed as much by outliers.

In addition to the mean and the median, there is one other common measurement of the "typical" value of a data set: the **mode**. The mode is defined as the element of the data set that occurs most frequently. The mode is fairly useless with data like weights or heights where there are a large number of possible values. The mode is most commonly used for categorical data, for which median and mean cannot be computed.

Example: In our vehicle color survey, we collected the data

Color	Frequency
Blue	3
Green	5
Red	4
White	3
Black	2
Grey	3

For this data, Green is the mode, since it is the data value that occurred the most frequently. It is possible for a data set to have more than one mode if several categories have the same frequency.

Measures of Variation

Consider these three sets of quiz scores:

Section A: 5 5 5 5 5 5 5 5 5

Section B: 0 0 0 0 0 10 10 10 10 10

Section C: 4 4 4 5 5 5 5 6 6 6

All three of these sets of data have a mean of 5 and median of 5, yet the sets of scores are clearly quite different. In section A, everyone had the same score; in section B half the class got no points and the other half got a perfect score, assuming this was a 10-point quiz. Section C was not as consistent as section A, but not as widely varied as section B.

In addition to the mean and median, which are measures of the "typical" or "middle" value, we also need a measure of how "spread out" or varied each data set is.

There are several ways to measure this "spread" of the data. The first is the simplest and is called the **range**. The range is simply the difference between the maximum value and the minimum value.

Example: Using the quiz scores from above,

For section A, the range is 0 since both maximum and minimum are 5 and $5 - 5 = 0$

For section B, the range is 10 since $10 - 0 = 10$

For section C, the range is 2 since $6 - 4 = 2$

So far range seems to be revealing how spread out the data is. However, if we add a fourth section:

Section D: 0 5 5 5 5 5 5 5 5 10

This section also has a mean and median of 5. The range is 10, yet this data set is quite different than Section B. To better illuminate the differences, we'll have to turn to more sophisticated measures of variation.

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean. A few important characteristics:

- Standard deviation is always positive. Standard deviation will be zero if all the data values are equal, and will get larger as the data spreads out.
- Standard deviation has the same units as the original data.
- Standard deviation, like the mean, can be highly influenced by outliers.

Using the data from section D, we could compute for each data value the difference between the data value and the mean:

data value	deviation: data value - mean
0	$0-5 = -5$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
5	$5-5 = 0$
10	$10-5 = 5$

We would like to get an idea of the "average" deviation from the mean, but if we find the average of the values in the second column the negative and positive values cancel each other out (this will always happen), so to prevent this we square every value in the second column:

data value	deviation: data value - mean	deviation squared
0	$0-5 = -5$	$(-5)^2 = 25$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
5	$5-5 = 0$	$0^2 = 0$
10	$10-5 = 5$	$(5)^2 = 25$

We then add the squared deviations up to get $25 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 25 = 50$. Ordinarily we would then divide by the number of scores, n , (in this case, 10) to find the mean of the deviations. But we only do this if the data set represents a population; if the data set represents a sample (as it almost always does), we instead divide by $n - 1$ (in this case, $10 - 1 = 9$)⁴

So in our example, we would have $50/10 = 5$ if section D represents a population and $50/9 =$ about 5.56 if section D represents a sample. These values (5 and 5.56) are called, respectively, the **population variance** and the **sample variance** for section D.

Variance can be a useful statistical concept, but note that the units of variance in this instance would be points-squared since we squared all of the deviations. What are points-squared? Good question. We would rather deal with the units we started with (points in this case), so to convert back we take the square root and get:

$$\text{population standard deviation} = \sqrt{\frac{50}{10}} = \sqrt{5} \approx 2.2$$

or

$$\text{sample standard deviation} = \sqrt{\frac{50}{9}} \approx 2.4$$

If we are unsure whether the data set is a sample or a population, we will usually assume it is a sample, and we will round answers to one more decimal place than the original data, as we have done above.

⁴ The reason we do this is highly technical, but we can see how it might be useful by considering the case of a small sample from a population that contains an outlier, which would increase the average deviation: the outlier very likely won't be included in the sample, so the mean deviation of the sample would underestimate the mean deviation of the population; thus we divide by a slightly smaller number to get a slightly bigger average deviation.

To compute standard deviation:

- 1) Find the deviation of each data from the mean. In other words, subtract the mean from the data value.
- 2) Square each deviation.
- 3) Add the squared deviations.
- 4) Divide by n , the number of data values, if the data represents a whole population; divide by $n-1$ if the data is from a sample.
- 5) Compute the square root of the result.

Example: Computing the standard deviation for Section B above, we first calculate that the mean is 5. Using a table can help keep track of your computations for the standard deviation:

data value	deviation: data value - mean	deviation squared
0	$0-5 = -5$	$(-5)^2 = 25$
0	$0-5 = -5$	$(-5)^2 = 25$
0	$0-5 = -5$	$(-5)^2 = 25$
0	$0-5 = -5$	$(-5)^2 = 25$
0	$0-5 = -5$	$(-5)^2 = 25$
10	$10-5 = 5$	$(5)^2 = 25$
10	$10-5 = 5$	$(5)^2 = 25$
10	$10-5 = 5$	$(5)^2 = 25$
10	$10-5 = 5$	$(5)^2 = 25$
10	$10-5 = 5$	$(5)^2 = 25$

Assuming this data represents a population, we will add the squared deviations, divide by 10, the number of data values, and compute the square root:

$$\sqrt{\frac{25 + 25 + 25 + 25 + 25 + 25 + 25 + 25 + 25 + 25}{10}} = \sqrt{\frac{250}{10}} = 5$$

Notice that the standard deviation of this data set is much larger than that of section D since the data in this set is more spread out.

For comparison, the standard deviations of all four sections are:

Section A: 5 5 5 5 5 5 5 5 5 5	Standard deviation: 0
Section B: 0 0 0 0 0 10 10 10 10 10	Standard deviation: 5
Section C: 4 4 4 5 5 5 5 6 6 6	Standard deviation: 0.8
Section D: 0 5 5 5 5 5 5 5 10	Standard deviation: 2.2

Where standard deviation is based on the mean, **quartiles** are based on the median. The median is the value so that 50% of the data values are below it. The first quartile (Q_1) is the value so that 25% of the data values are below it; the third quartile (Q_3) is the value so that 75% of the data values are below it. You may have guessed that the second quartile is the same as the median.

While quartiles are not a 1-number summary of variation like standard deviation, the quartiles are used with the median, minimum, and maximum values to form a **5 number summary** of the data looking like: Minimum, Q_1 , Median, Q_3 , Maximum

To find the first quartile, we need to find the data value so that 25% of the data is below it. If n is the number of data values, we compute a locator by finding 25% of n . If this locator is a decimal value, we round up, and find the data value in that position. If the locator is a whole number, we find the mean of the data value in that position and the next data value.

To find the first quartile, Q_1 :

Compute the locator: $L = 0.25n$

If L is a decimal value:

Round up to $L+$

Use the data value in the L^{th} position

If L is a whole number:

Find the mean of the data values in the L^{th} and $L+1^{\text{th}}$ positions.

To find the third quartile, Q_3 :

Use the same procedure, with locator: $L = 0.75n$

Examples should help make this clearer.

Example: Suppose we have measured 9 females and their heights (in inches), sorted from smallest to largest are:

59 60 62 64 66 67 69 70 72

To find the first quartile we first compute the locator: 25% of 9 is $0.25(9) = 2.25$. Since this value is not a whole number, we round up to 3. The first quartile will be the third data value: 62 inches.

Example: Suppose we had measured 8 females and their heights (in inches), sorted from smallest to largest are:

59 60 62 64 66 67 69 70

To find the first quartile we first compute the locator: 25% of 8 is $0.25(8) = 2$. Since this value is a whole number, we will find the mean of the 2nd and 3rd data values: $(60+62)/2 = 61$, so the first quartile is 61 inches.

The third quartile is computed similarly, using 75% instead of 25%. Note that the median could be computed the same way, using 50%.

Example: Using our 9 female sample of heights, to find the third quartile, we first compute the locator: 75% of 9 is $0.75(9) = 6.75$. Since this value is not a whole number, we round up to 7. The third quartile will be the seventh data value: 69 inches.

Example: Using our 8 female sample of heights, the locator would be: 75% of 8 is $0.75(8) = 6$. Since this value *is* a whole number, we will find the mean of the 6th and 7th data values: $(67+69)/2 = 68$, so the third quartile is 68 inches.

The 5-number summary combines the first and third quartile with the minimum, median, and maximum values.

Example: For the 9 female sample, the median is 66, the minimum is 59, and the maximum is 72. The 5 number summary is: 59, 62, 66, 69, 72.

For the 8 female sample, the median is 65, the minimum is 59, and the maximum is 70, so the 5 number summary would be: 59, 61, 65, 68, 70.

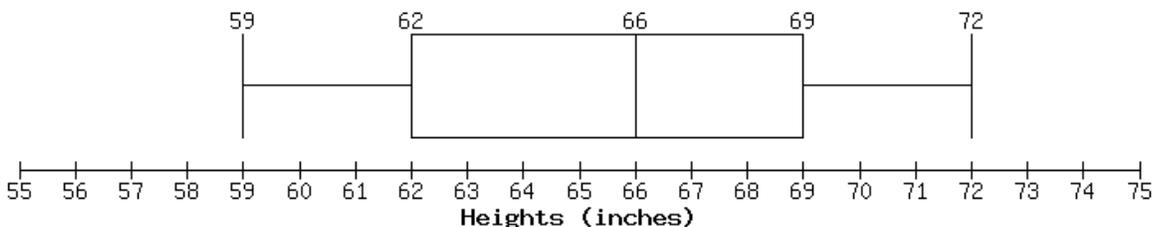
Example: Returning to our quiz score data. In each case, the first quartile locator is $0.25(10) = 2.5$, so the first quartile will be the 3rd data value, and the third quartile will be the 8th data value. Creating the 5-number summaries:

Section and data	5-number summary
Section A: 5 5 5 5 5 5 5 5 5	5, 5, 5, 5, 5
Section B: 0 0 0 0 0 10 10 10 10 10	0, 0, 5, 10, 10
Section C: 4 4 4 5 5 5 5 6 6 6	4, 4, 5, 6, 6
Section D: 0 5 5 5 5 5 5 5 10	0, 5, 5, 5, 10

Note that the 5 number summary divides the data into four intervals, each of which will contain about 25% of the data.

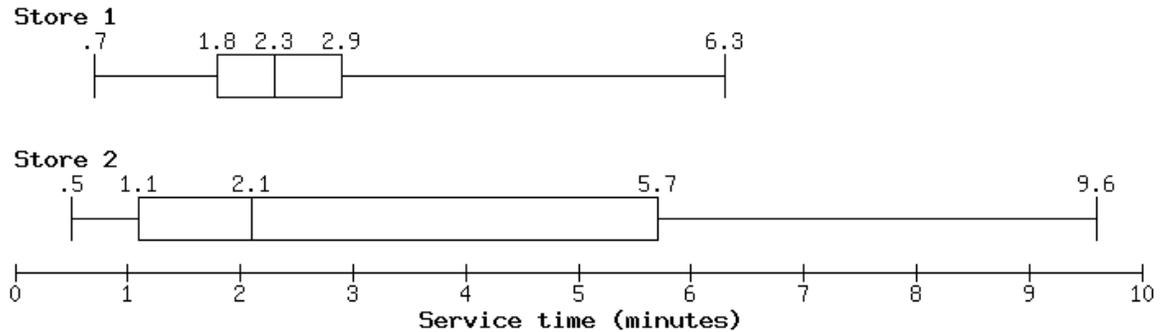
For visualizing data, there is a graphical representation of a 5-number summary called a **box plot**, or box and whisker graph. In a box plot, a number line is first drawn. A box is drawn from the first quartile to the third quartile, and a line is drawn through the box at the median. “Whiskers” are extended out to the minimum and maximum values.

Example: The box plot below is based on the 9 female height data with 5 number summary: 59, 62, 66, 69, 72.



Box plots are particularly useful for comparing data from two populations.

Example: The box plot of service times for two fast-food restaurants is shown below.



While store 2 had shorter median service time (2.1 minutes vs. 2.3 minutes), store 2 is less consistent, with wider spread of the data. At store 1, 75% of customers were served within 2.9 minutes, while at store 2, 75% of customers were served within 5.7 minutes. Which store should you go to in a hurry? That depends upon your opinions about luck – 25% of customers at store 2 had to wait between 5.7 and 9.6 minutes.

Exercises

Skills

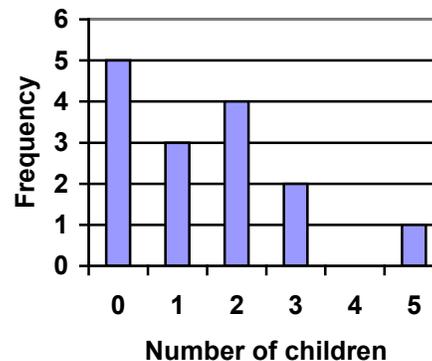
1. The table below shows scores on a Math test.
 - a. Complete the frequency table for the Math test scores
 - b. Construct a histogram of the data
 - c. Construct a pie chart of the data

80	50	50	90	70	70	100	60	70	80	70	50
90	100	80	70	30	80	80	70	100	60	60	50

2. A group of adults were asked how many cars they had in their household
 - a. Complete the frequency table for the car number data
 - b. Construct a histogram of the data
 - c. Construct a pie chart of the data

1	4	2	2	1	2	3	3	1	4	2	2
1	2	1	3	2	2	1	2	1	1	1	2

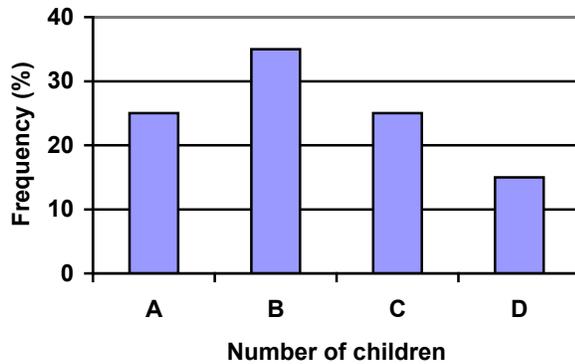
3. A group of adults were asked how many children they have in their families. The bar graph to the right shows the number of adults who indicated each number of children.
 - a. How many adults were questioned?
 - b. What percentage of the adults questioned had 0 children?



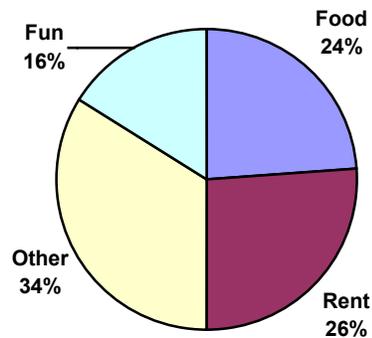
4. Jasmine was interested in how many days it would take an order from Netflix to arrive at her door. The graph below shows the data she collected.
 - a. How many movies did she order?
 - b. What percentage of the movies arrived in one day?



5. The bar graph below shows the *percentage* of students who received each letter grade on their last English paper. The class contains 20 students. What number of students earned an A on their paper?



6. Kori categorized her spending for this month into four categories: Rent, Food, Fun, and Other. The percents she spent in each category are pictured here. If she spent a total of \$2600 this month, how much did she spend on rent?



7. A group of diners were asked how much they would pay for a meal. Their responses were: \$7.50, \$8.25, \$9.00, \$8.00, \$7.25, \$7.50, \$8.00, \$7.00.
- Find the mean
 - Find the median
 - Write the 5-number summary for this data
8. You recorded the time in seconds it took for 8 participants to solve a puzzle. The times were: 15.2, 18.8, 19.3, 19.7, 20.2, 21.8, 22.1, 29.4.
- Find the mean
 - Find the median
 - Write the 5-number summary for this data
9. Refer back to the histogram from question #3.
- Compute the mean number of children for the group surveyed
 - Compute the median number of children for the group surveyed
 - Write the 5-number summary for this data.
 - Create box plot.

10. Refer back to the histogram from question #4.
- Compute the mean number of shipping days
 - Compute the median number of shipping days
 - Write the 5-number summary for this data.
 - Create box plot.

Concepts

11. The box plot below shows salaries for Actuaries and CPAs. Kendra makes the median salary for an Actuary. Kelsey makes the first quartile salary for a CPA. Who makes more money? How much more?



12. Referring to the boxplot above, what percentage of actuaries makes more than the median salary of a CPA?

Exploration

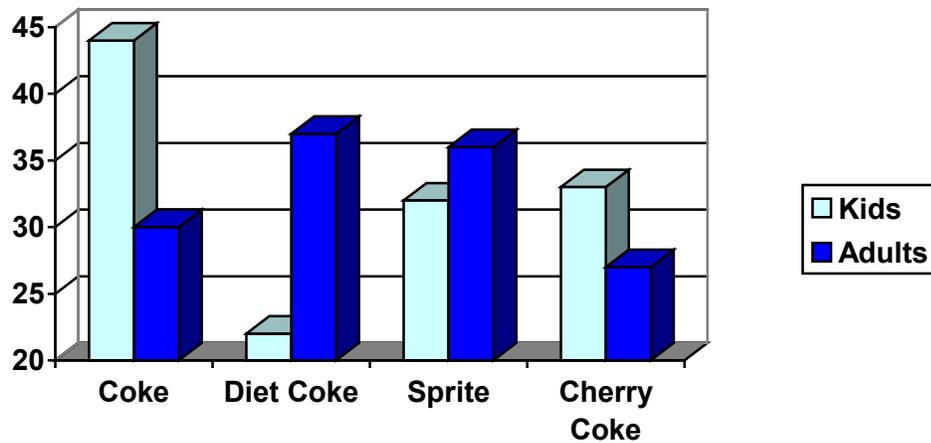
13. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a different set of 40 AIDS patients from the start of treatment until their deaths. The following data (in months) are collected.

Researcher 1: 3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher 2: 3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

- Create comparative histograms of the data
- Create comparative boxplots of the data

14. A graph appears below showing the number of adults and children who prefer each type of soda. There were 130 adults and kids surveyed. Discuss some ways in which the graph below could be improved



15. Make up three data sets with 5 numbers each that have:
- the same mean but different standard deviations.
 - the same mean but different medians.
 - the same median but different means.
16. A sample of 30 distance scores measured in yards has a mean of 7, a variance of 16, and a standard deviation of 4.
- You want to convert all your distances from yards to feet, so you multiply each score in the sample by 3. What are the new mean, median, variance, and standard deviation?
 - You then decide that you only want to look at the distance past a certain point. Thus, after multiplying the original scores by 3, you decide to subtract 4 feet from each of the scores. Now what are the new mean, median, variance, and standard deviation?

Historical Counting Systems

Introduction and Basic Number and Counting Systems

Introduction

As we begin our journey through the history of mathematics, one question to be asked is “Where do we start?” Depending on how you view mathematics or numbers, you could choose any of a number of launching points from which to begin. Howard Eves suggests the following list of possibilities.¹

Where to start the study of the history of mathematics...

- At the first logical geometric “proofs” traditionally credited to Thales of Miletus (600 BCE).
- With the formulation of methods of measurement made by the Egyptians and Mesopotamians/Babylonians.
- Where prehistoric peoples made efforts to organize the concepts of size, shape, and number.
- In pre-human times in the very simple number sense and pattern recognition that can be displayed by certain animals, birds, etc.
- Even before that in the amazing relationships of number and shapes found in plants.
- With the spiral nebulae, the natural course of planets, and other universe phenomena.

We can choose no starting point at all and instead agree that mathematics has *always* existed and has simply been waiting in the wings for humans to discover. Each of these positions can be defended to some degree and which one you adopt (if any) largely depends on your philosophical ideas about mathematics and numbers.

Nevertheless, we need a starting point. And without passing judgment on the validity of any of these particular possibilities, we will choose as our starting point the emergence of the idea of number and the process of counting as our launching pad. This is done primarily as a practical matter given the nature of this course. In the following chapter, we will try to focus on two main ideas. The first will be an examination of basic number and counting systems and the symbols that we use for numbers. We will look at our own modern (Western) number system as well those of a couple of selected civilizations to see the differences and diversity that is possible when humans start counting. The second idea we will look at will be base systems. By comparing our own base-ten (decimal) system with other bases, we will quickly become aware that the system that we are so used to, when slightly changed, will challenge our notions about numbers and what symbols for those numbers actually mean.

Recognition of More vs. Less

The idea of number and the process of counting goes back far beyond history began to be recorded. There is some archeological evidence that suggests that humans were counting as far back as 50,000 years ago.² However, we do not really know how this process started or developed over time. The best we can do is to make a good guess as to how things progressed. It is probably not hard to believe that even the earliest humans had some sense

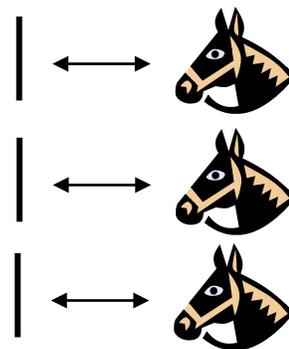
of *more* and *less*. Even some small animals have been shown to have such a sense. For example, one naturalist tells of how he would secretly remove one egg each day from a plover's nest. The mother was diligent in laying an extra egg every day to make up for the missing egg. Some research has shown that hens can be trained to distinguish between even and odd numbers of pieces of food.³ With these sorts of findings in mind, it is not hard to conceive that early humans had (at least) a similar sense of more and less. However, our conjectures about how and when these ideas emerged among humans are simply that; educated guesses based on our own assumptions of what might or could have been.

The Need for Simple Counting

As societies and humankind evolved, simply having a sense of more or less, even or odd, etc., would prove to be insufficient to meet the needs of everyday living. As tribes and groups formed, it became important to be able to know how many members were in the group, and perhaps how many were in the enemy's camp. And certainly it was important for them to know if the flock of sheep or other possessed animals were increasing or decreasing in size. "Just how many of them do we have, anyway?" is a question that we do not have a hard time imagining them asking themselves (or each other).

In order to count items such as animals, it is often conjectured that one of the earliest methods of doing so would be with "tally sticks." These are objects used to track the numbers of items to be counted. With this method, each "stick" (or pebble, or whatever counting device being used) represents one animal or object. This method uses the idea of *one to one correspondence*. In a one to one correspondence, items that are being counted are uniquely linked with some counting tool.

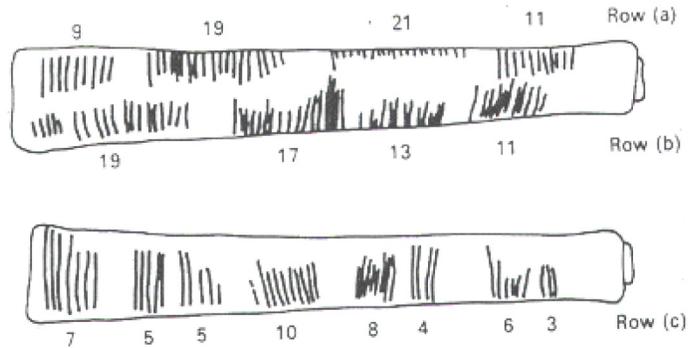
In the picture to the right, you see each stick corresponding to one horse. By examining the collection of sticks in hand one knows how many animals should be present. You can imagine the usefulness of such a system, at least for smaller numbers of items to keep track of. If a herder wanted to "count off" his animals to make sure they were all present, he could mentally (or methodically) assign each stick to one animal and continue to do so until he was satisfied that all were accounted for.



Of course, in our modern system, we have replaced the sticks with more abstract objects. In particular, the top stick is replaced with our symbol "1," the second stick gets replaced by a "2" and the third stick is represented by the symbol "3." But we are getting ahead of ourselves here. These modern symbols took many centuries to emerge.

Another possible way of employing the "tally stick" counting method is by making marks or cutting notches into pieces of wood, or even tying knots in string (as we shall see later). In 1937, Karl Absalom discovered a wolf bone that goes back possibly 30,000 years. It is believed to be a counting device.⁴ Another example of this kind of tool is the Ishango Bone, discovered in 1960 at Ishango, and shown below.⁵ It is reported to be between six and nine thousand years old and shows what appear to be markings used to do counting of some sort.

The markings on rows (a) and (b) each add up to 60. Row (b) contains the prime numbers between 10 and 20. Row (c) seems to illustrate for the method of doubling and multiplication used by the Egyptians (which we will study in the next topic). It is believed that this may also represent a lunar phase counter.



Spoken Words

As methods for counting developed, and as language progressed as well, it is natural to expect that spoken words for numbers would appear. Unfortunately, the development of these words, especially those for our numbers corresponding from one through ten, are not easy to trace. Past ten, however, we do see some patterns:

Eleven comes from “ein lifon,” meaning “one left over.”

Twelve comes from “twe lif,” meaning “two left over.”

Thirteen comes from “Three and ten” as do fourteen through nineteen.

Twenty appears to come from “twe–tig” which means “two tens.”

Hundred probably comes from a term meaning “ten times.”

Written Numbers

When we speak of “written” numbers, we have to be careful because this could mean a variety of things. It is important to keep in mind that modern paper is only a little more than 100 years old, so “writing” in times past often took on forms that might look quite unfamiliar to us today.

As we saw earlier, some might consider wooden sticks with notches carved in them as writing as these are means of recording information on a medium that can be “read” by others. Of course, the symbols used (simple notches) certainly did not leave a lot of flexibility for communicating a wide variety of ideas or information.

Other mediums on which “writing” may have taken place include carvings in stone or clay tablets, rag paper made by hand (12th century in Europe, but earlier in China), papyrus (invented by the Egyptians and used up until the Greeks), and parchments from animal skins. And these are just a few of the many possibilities.

These are just a few examples of early methods of counting and simple symbols for representing numbers. Extensive books, articles and research have been done on this topic and could provide enough information to fill this entire course if we allowed it to. The range and diversity of creative thought that has been used in the past to describe numbers and to count objects and people is staggering. Unfortunately, we don’t have time to examine them all, but it is fun and interesting to look at one system in more detail to see just how ingenious people have been.

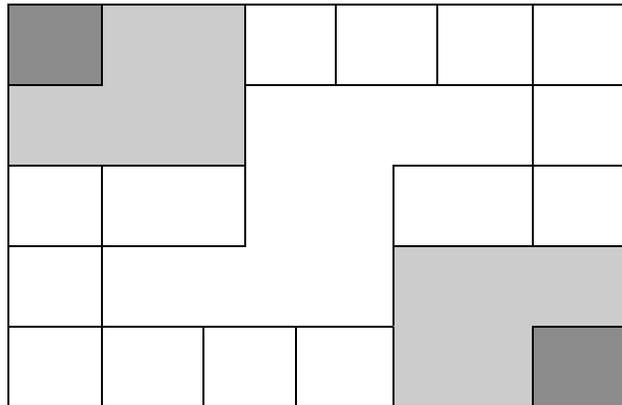
The Number and Counting System of the Inca Civilization

Background

There is generally a lack of books and research material concerning the historical foundations of the Americas. Most of the “important” information available concentrates on the eastern hemisphere, with Europe as the central focus. The reasons for this may be twofold: first, it is thought that there was a lack of specialized mathematics in the American regions; second, many of the secrets of ancient mathematics in the Americas have been closely guarded.⁶ The Peruvian system does not seem to be an exception here. Two researchers, Leland Locke and Erland Nordenskiold, have carried out research that has attempted to discover what mathematical knowledge was known by the Incas and how they used the Peruvian quipu, a counting system using cords and knots, in their mathematics. These researchers have come to certain beliefs about the quipu that we will summarize here.

Counting Boards

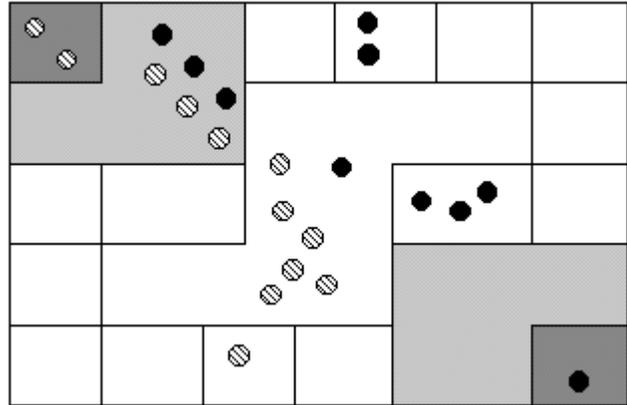
It should be noted that the Incas did not have a complicated system of computation. Where other peoples in the regions, such as the Mayans, were doing computations related to their rituals and calendars, the Incas seem to have been more concerned with the simpler task of record-keeping. To do this, they used what are called the “quipu” to record quantities of items. (We will describe them in more detail in a moment.) However, they first often needed to do computations whose results would be recorded on quipu. To do these computations, they would sometimes use a counting board constructed with a slab of stone. In the slab were cut rectangular and square compartments so that an octagonal (eight-sided) region was left in the middle. Two opposite corner rectangles were raised. Another two sections were mounted on the original surface of the slab so that there were actually three levels available. In the figure shown, the darkest shaded corner regions represent the highest, third level. The lighter shaded regions surrounding the corners are the second highest levels, while the clear white rectangles are the compartments cut into the stone slab.



Pebbles were used to keep accounts and their positions within the various levels and compartments gave totals. For example, a pebble in a smaller (white) compartment represented one unit. Note that there are 12 such squares around the outer edge of the figure. If a pebble was put into one of the two (white) larger, rectangular compartments, its value was doubled. When a pebble was put in the octagonal region in the middle of the slab, its value was tripled. If a pebble was placed on the second (shaded) level, its value was multiplied by six. And finally, if a pebble was found on one of the two highest corner levels, its value was multiplied by twelve. Different objects could be counted at the same time by representing different objects by different colored pebbles.

Example 1

Suppose you have the following counting board with two different kind of pebbles places as illustrated. Let the solid black pebble represent a dog and the striped pebble represent a cat. How many dogs are being represented?



Solution

There are two black pebbles in the outer square regions...these represent 2 dogs.

There are three black pebbles in the larger (white) rectangular compartments. These represent 6 dogs.

There is one black pebble in the middle region...this represents 3 dogs.

There are three black pebbles on the second level...these represent 18 dogs.

Finally, there is one black pebble on the highest corner level...this represents 12 dogs. We then have a total of $2+6+3+18+12 = 41$ dogs. ♦

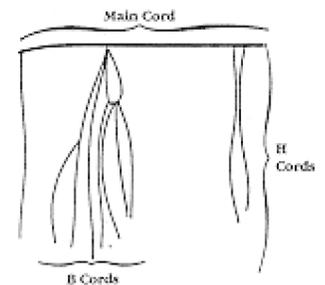
CheckPoint A

How many cats are represented on this board? See endnotes for the solution.⁷

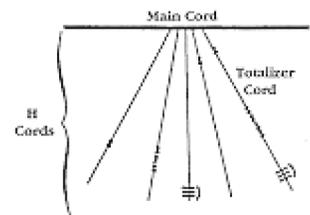
The Quipu

This kind of board was good for doing quick computations, but it did not provide a good way of keep a permanent recording of quantities or computations. For this purpose, they used the quipu. The quipu is a collection of cords with knots in them. These cords and knots are carefully arranged so that the position and type of cord or knot gives specific information on how to decipher the cord.

A quipu is made up of a main cord which has other cords (branches) tied to it. See pictures to the right.⁸

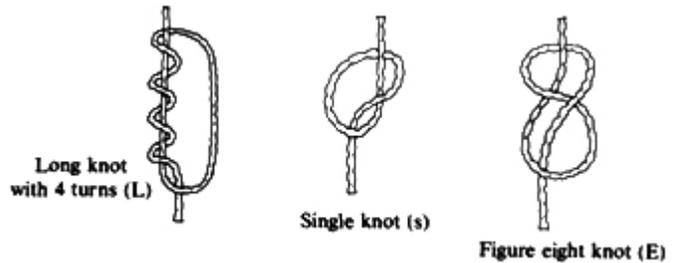


Locke called the branches H cords. They are attached to the main cord. B cords, in turn, were attached to the H cords. Most of these cords would have knots on them. Rarely are knots found on the main cord, however, and tend to be mainly on the H and B cords. A quipu might also have a “totalizer” cord that summarizes all of the information on the cord group in one place.



Locke points out that there are three types of knots, each representing a different value, depending on the kind of knot used and its position on the cord. The Incas, like us, had a

decimal (base-ten) system, so each kind of knot had a specific decimal value. The Single knot, pictured in the middle of the diagram⁹ was used to denote tens, hundreds, thousands, and ten thousands. They would be on the upper levels of the H cords. The figure-eight knot on the end was used to denote the integer “one.” Every other integer from 2 to 9 was represented with a long knot, shown on the left of the figure. (Sometimes long knots were used to represent tens and hundreds.) Note that the long knot has several turns in it...the number of turns indicates which integer is being represented. The units (ones) were placed closest to the bottom of the cord, then tens right above them, then the hundreds, and so on.



In order to make reading these pictures more easy, we will adopt a convention that is consistent. For the long knot with turns in it (representing the numbers 2 through 9), we will use the following notation:



The four horizontal bars represent four turns and the curved arc on the right links the four turns together. This would represent the number 4.

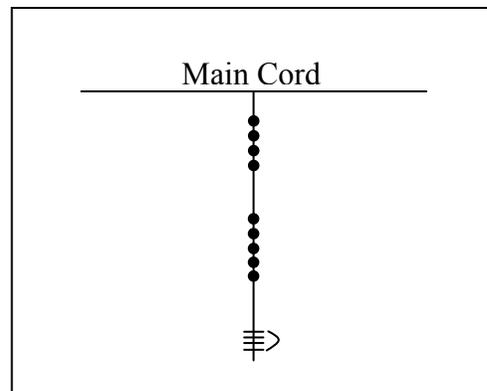
We will represent the single knot with a large dot (•) and we will represent the figure eight knot with the a sideways eight (∞).

Example 2

What number is represented on the cord shown?

Solution

On the cord, we see a long knot with four turns in it...this represents four in the ones place. Then 5 single knots appear in the tens position immediately above that, which represents 5 tens, or 50. Finally, 4 single knots are tied in the hundreds, representing four 4 hundreds, or 400. Thus, the total shown on this cord is 454. ♦

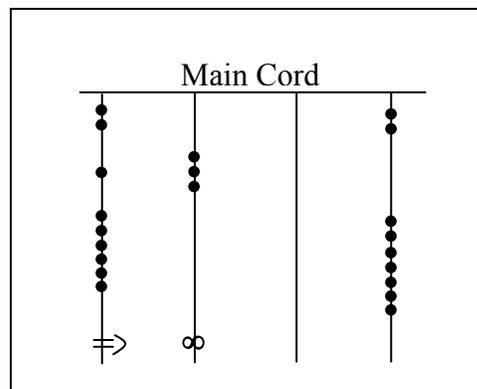


CheckPoint B

What numbers are represented on each of the four cords hanging from the main cord?

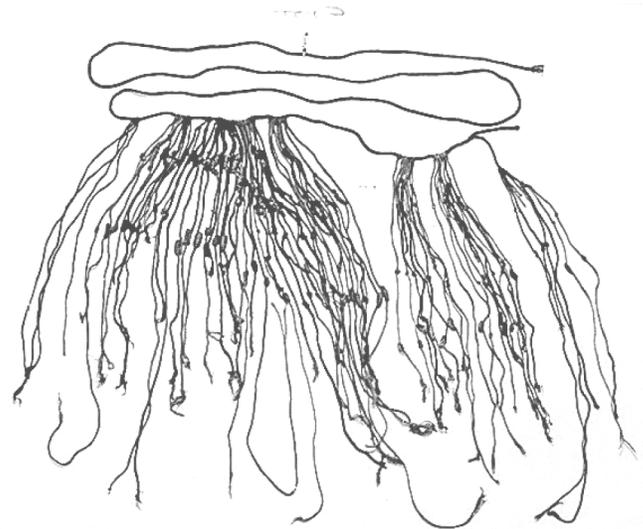
Answer:

- From left to right:
- Cord 1 = 2,162
- Cord 2 = 301
- Cord 3 = 0
- Cord 4 = 2,070 ♦



The colors of the cords had meaning and could distinguish one object from another. One color could represent llamas, while a different color might represent sheep, for example. When all the colors available were exhausted, they would have to be re-used. Because of this, the ability to read the quipu became a complicated task and specially trained individuals did this job. They were called Quipucamayoc, which means keeper of the quipus. They would build, guard, and decipher quipus.

As you can see from this photograph of an actual quipu, they could get quite complex.



There were various purposes for the quipu. Some believe that they were used to keep an account of their traditions and history, using knots to record history rather than some other formal system of writing. One writer has even suggested that the quipu replaced writing as it formed a role in the Incan postal system.¹⁰ Another proposed use of the quipu is as a translation tool. After the conquest of the Incas by the Spaniards and subsequent “conversion” to

Catholicism, an Inca supposedly could use the quipu to confess their sins to a priest. Yet another proposed use of the quipu was to record numbers related to magic and astronomy, although this is not a widely accepted interpretation.

The mysteries of the quipu have not been fully explored yet. Recently, Ascher and Ascher have published a book, *The Code of the Quipu: A Study in Media, Mathematics, and Culture*, which is “an extensive elaboration of the logical-numerical system of the quipu.”¹¹ For more information on the quipu, you may want to check out the following Internet link:

<http://wiscinfo.doit.wisc.edu/chaysimire/titulo2/khipus/what.htm>

We are so used to seeing the symbols 1, 2, 3, 4, etc. that it may be somewhat surprising to see such a creative and innovative way to compute and record numbers. Unfortunately, as we proceed through our mathematical education in grade and high school, we receive very little information about the wide range of number systems that have existed and which still exist all over the world. That’s not to say our own system is not important or efficient. The fact that it has survived for hundreds of years and shows no sign of going away any time soon suggests that we may have finally found a system that works well and may not need further improvement. But only time will tell that whether or not that conjecture is valid or not. We now turn to a brief historical look at how our current system developed over history.

The Hindu – Arabic Number System

The Evolution of a System

Our own number system, composed of the ten symbols {0,1,2,3,4,5,6,7,8,9} is called the *Hindu–Arabic system*. This is a base–ten (decimal) system since place values increase by powers of ten. Furthermore, this system is positional, which means that the position of a symbol has bearing on the value of that symbol within the number. For example, the position of the symbol 3 in the number 435,681 gives it a value much greater than the value of the symbol 8 in that same number. (We'll explore base systems more thoroughly later.) The development of these ten symbols and their use in a positional system comes to us primarily from India.¹²

It was not until the 15th century that the symbols that we are familiar with today first took form in Europe. However, the history of these numbers and their development goes back hundreds of years. One important source of information on this topic is the writer al–Biruni, whose picture is shown here.¹³ Al–Biruni, who was born in modern day Uzbekistan, had visited India on several occasions and made comments on the Indian number system. When we look at the origins of the numbers that al–Biruni encountered, we have to go back to the third century B.C.E. to explore their origins. It is then that the Brahmi numerals were being used.



The Brahmi numerals were more complicated than those used in our own modern system. They had separate symbols for the numbers 1 through 9, as well as distinct symbols for 10, 100, 1000, ..., also for 20, 30, 40, ..., and others for 200, 300, 400, ..., 900. The Brahmi symbols for 1, 2, and 3 are shown below.¹⁴

1	2	3
—	=	≡

Brahmi one, two, three

These numerals were used all the way up to the 4th century C.E., with variations through time and geographic location. For example, in the first century C.E., one particular set of Brahmi numerals took on the following form¹⁵:

1	2	3	4	5	6	7	8	9
—	=	≡	+	h	4	7	5	7

From the 4th century on, you can actually trace several different paths that the Brahmi numerals took to get to different points and incarnations. One of those paths led to our current numeral system, and went through what are called the Gupta numerals. The Gupta numerals were prominent during a time ruled by the Gupta dynasty and were spread throughout that empire as they conquered lands during the 4th through 6th centuries. They have the following form¹⁶:

1	2	3	4	5	6	7	8	9
—	=	≡	𑀓	𑀔	𑀕	𑀖	𑀗	𑀘

How the numbers got to their Gupta form is open to considerable debate. Many possible hypotheses have been offered, most of which boil down to two basic types¹⁷. The first type of hypotheses states that the numerals came from the initial letters of the names of the numbers. (This is not uncommon...the Greek numerals developed in this manner.) The second type of hypothesis states that they were derived from some earlier number system. However, there are other hypotheses that are offered, one of which is by the researcher Ifrah. His theory is that there were originally nine numerals, each represented by a corresponding number of vertical lines. One possibility is this:¹⁸

1	2	3	4	5	6	7	8	9
I	II	III	II II	III II	III III	III III	III III	III III

Because these symbols would have taken a lot of time to write, they eventually evolved into cursive symbols that could be written more quickly. If we compare these to the Gupta numerals above, we can try to see how that evolutionary process might have taken place, but our imagination would be just about all we would have to depend upon since we do not know exactly how the process unfolded.

The Gupta numerals eventually evolved into another form of numerals called the Nagari numerals, and these continued to evolve until the 11th century, at which time they looked like this:¹⁹

1	2	3	4	5	6	7	8	9	0
१	२	३	४	५	६	७	८	९	०

Note that by this time, the symbol for 0 has appeared! (The Mayans in the Americas had a symbol for zero long before this, however, as we shall see later in the chapter.)

These numerals were adopted by the Arabs, most likely in the eighth century during Islamic incursions into the northern part of India.²⁰ It is believed that the Arabs were instrumental in spreading them to other parts of the world, including Spain (see below).

Other examples of variations up to the eleventh century include:

Devangari, eighth century²¹:

१ २ ३ ४ ५ ६ ७ ८ ९

West Arab Gobar, tenth century²²:

۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹

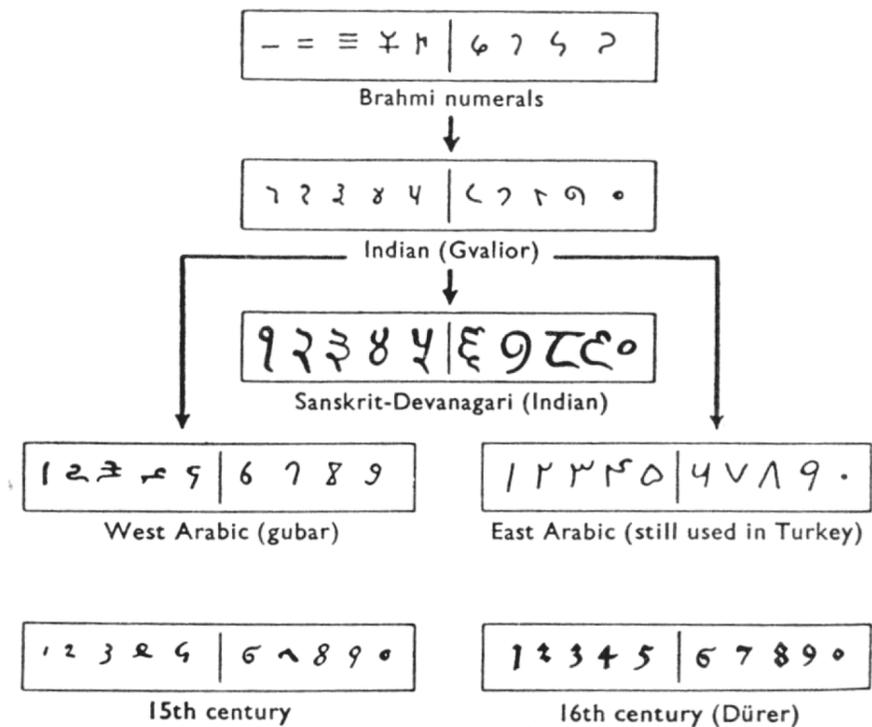
Spain, 976 B.C.E.²³:

۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹

Finally, one more graphic²⁴ shows various forms of these numerals as they developed and eventually converged to the 15th century in Europe.

The Positional System

More important than the form of the number symbols is the development of the place value system. Although it is in slight dispute, the earliest known document in which the Indian system displays a positional system dates back to 346 C.E. However, some evidence suggests that they may have actually developed a positional system as far back as the first century C.E.



The Indians were not the first to use a positional system. The Babylonians (as we will see in Chapter 3) used a positional system with 60 as their base. However, there is not much evidence that the Babylonian system had much impact on later numeral systems, except with the Greeks. Also, the Chinese had a base-10 system, probably derived from the use of a counting board²⁵. Some believe that the positional system used in India was derived from the Chinese system.

Wherever it may have originated, it appears that around 600 C.E., the Indians abandoned the use of symbols for numbers higher than nine and began to use our familiar system where the position of the symbol determines its overall value.²⁶ Numerous documents from the seventh century demonstrate the use of this positional system.

Interestingly, the earliest dated inscriptions using the system with a symbol for zero come from Cambodia. In 683, the 605th year of the Saka era is written with three digits and a dot in the middle. The 608th year uses three digits with a modern 0 in the middle.²⁷ The dot as a symbol for zero also appears in a Chinese work (*Chiu-chih li*). The author of this document gives a strikingly clear description of how the Indian system works:

*Using the [Indian] numerals, multiplication and division are carried out. Each numeral is written in one stroke. When a number is counted to ten, it is advanced into the higher place. In each vacant place a dot is always put. Thus the numeral is always denoted in each place. Accordingly there can be no error in determining the place. With the numerals, calculations is easy...*²⁸

Transmission to Europe

It is not completely known how the system got transmitted to Europe. Traders and travelers of the Mediterranean coast may have carried it there. It is found in a tenth-century Spanish manuscript and may have been introduced to Spain by the Arabs, who invaded the region in 711 C.E. and were there until 1492.

In many societies, a division formed between those who used numbers and calculation for practical, every day business and those who used them for ritualistic purposes or for state business.²⁹ The former might often use older systems while the latter were inclined to use the newer, more elite written numbers. Competition between the two groups arose and continued for quite some time. In a 14th century manuscript of Boethius' *The Consolations of Philosophy*, there appears a well-known drawing of two mathematicians. One is a merchant and is using an abacus (the "abacist"). The other is a Pythagorean philosopher (the "algorist") using his "sacred" numbers. They are in a competition that is being judged by the goddess of number. By 1500 C.E., however, the newer symbols and system had won out and has persevered until today. The Seattle Times recently reported that the Hindu-Arabic numeral system has been included in the book *The Greatest Inventions of the Past 2000 Years*.³⁰



One question to answer is *why* the Indians would develop such a positional notation. Unfortunately, an answer to that question is not currently known. Some suggest that the system has its origins with the Chinese counting boards. These boards were portable and it is thought that Chinese travelers who passed through India took their boards with them and

ignited an idea in Indian mathematics.³¹ Others, such as G. G. Joseph propose that it is the Indian fascination with very large numbers that drove them to develop a system whereby these kinds of big numbers could easily be written down. In this theory, the system developed entirely within the Indian mathematical framework without considerable influence from other civilizations.

The Development and Use of Different Number Bases

Introduction and Basics

During the previous discussions, we have been referring to positional base systems. In this section of the chapter, we will explore exactly what a base system is and what it means if a system is “positional.” We will do so by first looking at our own familiar, base-ten system and then deepen our exploration by looking at other possible base systems. In the next part of this section, we will journey back to Mayan civilization and look at their unique base system, which is based on the number 20 rather than the number 10.

A base system is a structure within which we count. The easiest way to describe a base system is to think about our own base-ten system. The base-ten system, which we call the “decimal” system, requires a total of ten different symbols/digits to write any number. They are, of course, 0, 1, 2, 9.

The decimal system is also an example of a *positional* base system, which simply means that the position of a digit gives its place value. Not all civilizations had a positional system even though they did have a base with which they worked.

In our base-ten system, a number like 5,783,216 has meaning to us because we are familiar with the system and its places. As we know, there are six ones, since there is a 6 in the ones place. Likewise, there are seven “hundred thousands,” since the 7 resides in that place. Each digit has a value that is explicitly determined by its position within the number. (We make a distinction between digit, which is just a symbol such as 5, and a number, which is made up of one or more digits.) We can take this number and assign each of its digits a value. One way to do this is with a table, which follows:

5,000,000	= $5 \times 1,000,000$	= 5×10^6	Five million
+700,000	= $7 \times 100,000$	= 7×10^5	Seven hundred thousand
+80,000	= $8 \times 10,000$	= 8×10^4	Eighty thousand
+3,000	= 3×1000	= 3×10^3	Three thousand
+200	= 2×100	= 2×10^2	Two hundred
+10	= 1×10	= 1×10^1	Ten
+6	= 6×1	= 6×10^0	Six
5,783,216	Five million, seven hundred eighty-three thousand, two hundred sixteen		

From the third column in the table we can see that each place is simply a multiple of ten. Of course, this makes sense given that our base is ten. The digits that are multiplying each place simply tell us how many of that place we have. We are restricted to having at most 9 in any

one place before we have to “carry” over to the next place. We cannot, for example, have 11 in the hundred–thousands place. Instead, we would carry 1 to the millions place and retain 1 in the hundred–thousands place. This comes as no surprise to us since we readily see that 11 hundred–thousands is the same as one million, one hundred thousand. Carrying is a pretty typical occurrence in a base system.

However, base-ten is not the only option we have. Practically any positive integer greater than or equal to 2 can be used as a base for a number system. Such systems can work just like the decimal system except the number of symbols will be different and each position will depend on the base itself.

Other Bases

For example, let’s suppose we adopt a base–five system. The only modern digits we would need for this system are 0,1,2,3 and 4. What are the place values in such a system? To answer that, we start with the ones place, as most base systems do. However, if we were to count in this system, we could only get to four (4) before we had to jump up to the next place. Our base is 5, after all! What is that next place that we would jump to? It would not be tens, since we are no longer in base–ten. We’re in a different numerical world. As the base–ten system progresses from 10^0 to 10^1 , so the base–five system moves from 5^0 to $5^1 = 5$. Thus, we move from the ones to the fives. After the fives, we would move to the 5^2 place, or the twenty fives. (Note that in base–ten, we would have gone from the tens to the hundreds, which is, of course, 10^2 .)

Let’s take an example and build a table. Consider the number 30412 in base five. We will write this as 30412_5 , where the subscript 5 is not part of the number but indicates the base we’re using. First off, note that this is NOT the number

	Base 5	This column converts to base–ten	In Base–Ten
	3×5^4	$= 3 \times 625$	$= 1875$
+	0×5^3	$= 0 \times 125$	$= 0$
+	4×5^2	$= 4 \times 25$	$= 100$
+	1×5^1	$= 1 \times 5$	$= 5$
+	2×5^0	$= 2 \times 1$	$= 2$
		Total	1982

“thirty thousand, four hundred twelve.” We must be careful not to impose the base–ten system on this number. Here’s what our table might look like. We will use it to convert this number to our more familiar base–ten system.

As you can see, the number 30412_5 is equivalent to 1,982 in base–ten. We will say $30412_5 = 1982_{10}$. All of this may seem strange to you, but that’s only because you are so used to the only system that you’ve ever seen.

Example 3

Convert 6234_7 to a base 10 number.

Solution

We first note that we are given a base-7 number that we are to convert. Thus, our places will start at the ones (7^0), and then move up to the 7's, 49's (7^2), etc. Here's the breakdown:

	Base 7	Convert	Base 10
	$= 6 \times 7^3$	$= 6 \times 343$	$= 2058$
+	$= 2 \times 7^2$	$= 2 \times 49$	$= 98$
+	$= 3 \times 7$	$= 3 \times 7$	$= 21$
+	$= 4 \times 1$	$= 4 \times 1$	$= 4$
		Total	2181

Thus $6234_7 = 2181_{10}$. ♦

CheckPoint C

Convert 41065_7 to a base 10 number. See endnotes for the answer.³²

Converting from Base 10 to Other Bases

Converting from an unfamiliar base to the familiar decimal system is not that difficult once you get the hang of it. It's only a matter of identifying each place and then multiplying each digit by the appropriate power. However, going the other direction can be a little trickier. Suppose you have a base-ten number and you want to convert to base-five. Let's start with some simple examples before we get to a more complicated one.

Example 4

Convert twelve to a base-five number.

Solution:

We can probably easily see that we can rewrite this number as follows:

$$12 = (2 \times 5) + (2 \times 1)$$

Hence, we have two fives and 2 ones. Hence, in base-five we would write twelve as 22_5 . Thus, $12_{10} = 22_5$ ♦

Example 5

Convert sixty-nine to a base-five number. We can see now that we have more than 25, so we rewrite sixty-nine as follows:

$$69 = (2 \times 25) + (3 \times 5) + (4 \times 1)$$

Solution

Here, we have two twenty-fives, 3 fives, and 4 ones. Hence, in base five we have 234_5 . Thus, $69_{10} = 234_5$. ♦

Example 6

Convert the base–seven number 3261_7 to base 10.

Solution

The powers of 7 are:

$$\begin{aligned}7^0 &= 1 \\7^1 &= 7 \\7^2 &= 49 \\7^3 &= 343 \\&\text{Etc...}\end{aligned}$$

$$3261_7 = (3 \times 343) + (2 \times 49) + (6 \times 7) + (1 \times 1) = 1170_{10}. \text{ Thus } 3261_7 = 1170_{10}. \blacklozenge$$

CheckPoint D

Convert 143 to base 5. See the footnotes for solution.³³

CheckPoint E

Convert the base–three number 21021_3 to base 10. See the endnotes for the solution.³⁴

In general, when converting from base–ten to some other base, it is often helpful to determine the highest power of the base that will divide into the given number at least once. In the last example, $5^2 = 25$ is the largest power of five that is present in 69, so that was our starting point. If we had moved to $5^3 = 125$, then 125 would not divide into 69 at least once.

Example 7

Convert the base–ten number 348 to base–five.

Solution

The powers of five are:

$$\begin{aligned}5^0 &= 1 \\5^1 &= 5 \\5^2 &= 25 \\5^3 &= 125 \\5^4 &= 625 \\&\text{Etc...}\end{aligned}$$

Since 348 is smaller than 625, but bigger than 125, we see that $5^3=125$ is the highest power of five present in 348. So we divide 125 into 348 to see how many of them there are:

$$348 \div 125 = 2 \text{ with remainder } 98$$

There are 98 left over, so we see how many 25's (the next smallest power of five) there are in the remainder:

$$98 \div 25 = 3 \text{ with remainder } 23$$

There are 23 left over, so we look at the next place, the 5's:

$$23 \div 5 = 4 \text{ with remainder } 3$$

This leaves us with 3 ones, and we are ready to assemble our base-five number:

$$348 = (2 \times 5^3) + (3 \times 5^2) + (4 \times 5^1) + (3 \times 1)$$

Hence, our base-five number is **2343**. We'll say that $348_{10} = 2343_5$. ♦

Example 8

Convert the base-ten number 4,509 to base-seven.

Solution

The powers of 7 are:

$$\begin{aligned} 7^0 &= 1 \\ 7^1 &= 7 \\ 7^2 &= 49 \\ 7^3 &= 343 \\ 7^4 &= 2401 \\ 7^5 &= 16807 \\ &\text{Etc...} \end{aligned}$$

The highest power of 7 that will divide evenly into 4,509 is $7^4 = 2401$. With division, we see that it will go in 1 time with a remainder of 2108. So we have 1 in the 7^4 place. The next power down is $7^3 = 343$, which goes into 2108 six times with a new remainder of 50. So we have 6 in the 7^3 place.

The next power down is $7^2 = 49$, which goes into 50 once with a new remainder of 1. So there is a 1 in the 7^2 place. The next power down is 7^1 but there was only a remainder of 1, so that means there is a 0 in the 7^1 's place and we still have 1 as a remainder. That, of course, means that we have 1 in the ones place. Putting all of this together means that $4,509_{10} = 16101_7$. ♦

$4,509 \div 7^4 =$	1 R 2108
$2108 \div 7^3 =$	6 R 50
$50 \div 7^2 =$	1 R 1
$1 \div 7^1 =$	0 R 1
$1 \div 7^0 =$	1
$4,509_{10} =$ 16101 ₇	

CheckPoint F

Convert 657_{10} to a base 4 number. See endnotes for the answer.³⁵

CheckPoint G

Convert 8377_{10} to a base 8 number. See endnotes for the answer.³⁶

A New Method For Converting From Base 10 to Other Bases

As you read the solution to this last example and attempted the “You Try It” problems, you may have had to repeatedly stop and think about what was going on. The fact that you are probably struggling to follow the explanation and reproduce the process yourself is mostly due to the fact that the non-decimal systems are so unfamiliar to you. In fact, the only system that you are probably comfortable with is the decimal system. As budding mathematicians, you should always be asking questions like “How could I simplify this process?” In general,

that is one of the main things that mathematicians do...they look for ways to take complicated situations and make them easier or more familiar. In this section we will attempt to do that.

To do so, we will start by looking at our own decimal system. What we do may seem obvious and maybe even intuitive but that's the point. We want to find a process that we readily recognize works and makes sense to us in a familiar system and then use it to extend our results to a different, unfamiliar system.

Let's start with the decimal number, 4863_{10} . We will convert this number to base 10. (Yeah, I know it's already in base 10, but if you carefully follow what we're doing, you'll see it makes things work out very nicely with other bases later on.) We first note that the highest power of 10 that will divide into 4863 at least once is $10^3 = 1000$. *In general, this is the first step in our new process; we find the highest power that a given base that will divide at least once into our given number.*

We now divide 1000 into 4863:

$$4863 \div 1000 = 4.863$$

This says that there are four thousands in 4863 (obviously). However, it also says that there are 0.863 thousands in 4863. This fractional part is our remainder and will be converted to lower powers of our base (10). If we take that decimal and multiply by 10 (since that's the base we're in) we get the following:

$$0.863 \times 10 = 8.63$$

Why multiply by 10 at this point? We need to recognize here that 0.863 thousands is the same as 8.63 hundreds. Think about that until it sinks in.

$$(0.863)(1000) = 863$$

$$(8.63)(100) = 863$$

These two statements are equivalent. So, what we are really doing here by multiplying by 10 is rephrasing or converting from one place (thousands) to the next place down (hundreds).

$$\begin{aligned} 0.863 \times 10 &\Rightarrow 8.63 \\ (\text{Parts of Thousands}) \times 10 &\Rightarrow \text{Hundreds} \end{aligned}$$

What we have now is 8 hundreds and a remainder of 0.63 hundreds, which is the same as 6.3 tens. We can do this again with the 0.63 that remains after this first step.

$$\begin{aligned} 0.63 \times 10 &\Rightarrow 6.3 \\ \text{Hundreds} \times 10 &\Rightarrow \text{Tens} \end{aligned}$$

So we have six tens and 0.3 tens, which is the same as 3 ones, our last place value.

Now here's the punch line. Let's put all of the together in one place:

$$\begin{array}{r}
 4863 \div 10 = \textcircled{4}.863 \\
 \downarrow \\
 0.863 \times 10 = \textcircled{8}.63 \\
 \downarrow \\
 0.63 \times 10 = \textcircled{6}.3 \\
 \downarrow \\
 0.3 \times 10 = \textcircled{3}.0
 \end{array}$$

Note that in each step, the remainder is carried down to the next step and multiplied by 10, the base. Also, at each step, the whole number part, which is circled, gives the digit that belongs in that particular place. What is amazing is that this works for any base! So, to convert from a base 10 number to some other base, b , we have the following steps we can follow:

Converting from Base 10 to Base b

1. Find the highest power of the base b that will divide into the given number at least once and then divide.
2. Keep the whole number part, and multiply the fractional part by the base b .
3. Repeat step two, keeping the whole number part (including 0), carrying the fractional part to the next step until only a whole number result is obtained.
4. Collect all your whole number parts to get your number in base b notation.

We will illustrate this procedure with some examples.

Example 9

Convert the base 10 number, 348_{10} , to base 5.

Solution

This is actually a conversion that we have done in a previous example. The powers of five are:

$$\begin{array}{l}
 5^0=1 \\
 5^1=5 \\
 5^2=25 \\
 5^3=125 \\
 5^4=625 \\
 \text{Etc...}
 \end{array}$$

The highest power of five that will go into 348 at least once is 5^3 . So we divide by 125 and then proceed.

$$\begin{array}{r}
 348 \div 5^3 = \textcircled{2}.784 \\
 \downarrow
 \end{array}$$

$$\begin{array}{l}
 0.784 \times 5 = \textcircled{3}.92 \\
 \downarrow \\
 0.92 \times 5 = \textcircled{4}0.6 \\
 \downarrow \\
 0.6 \times 5 = \textcircled{3}.0
 \end{array}$$

By keeping all the whole number parts, from top bottom, gives 2343 as our base 5 number. Thus, $2343_5 = 348_{10}$.

We can compare our result with what we saw earlier, or simply check with our calculator, and find that these two numbers really are equivalent to each other. ♦

Example 10

Convert the base 10 number, 3007_{10} , to base 5.

Solution

The highest power of 5 that divides at least once into 3007 is $5^4 = 625$. Thus, we have:

$$\begin{array}{l}
 3007 \div 625 = \textcircled{4}.8112 \\
 0.8112 \times 5 = \textcircled{4}.056 \\
 0.056 \times 5 = \textcircled{0}.28 \\
 0.28 \times 5 = \textcircled{1}0.4 \\
 0.4 \times 5 = \textcircled{2}0.0
 \end{array}$$

This gives us that $3007_{10} = 44012_5$. Notice that in the third line that multiplying by 5 gave us 0 for our whole number part. We don't discard that! The zero tells us that a zero in that place. That is, there are no 5^2 's in this number. ♦

This last example shows the importance of using a calculator in certain situations and taking care to avoid clearing the calculator's memory or display until you get to the very end of the process.

Example 11

Convert the base 10 number, 63201_{10} , to base 7.

Solution

The powers of 7 are:

$$\begin{array}{l}
 7^0 = 1 \\
 7^1 = 7
 \end{array}$$

$$\begin{aligned}
 7^2 &= 49 \\
 7^3 &= 343 \\
 7^4 &= 2401 \\
 7^5 &= 16807 \\
 &\text{etc...}
 \end{aligned}$$

The highest power of 7 that will divide at least once into 63201 is 7^5 . When we do the initial division on a calculator, we get the following:

$$63201 \div 7^5 = 3.760397453$$

The decimal part actually fills up the calculator's display and we don't know if it terminates at some point or perhaps even repeats down the road. (It must terminate or repeat since $63201 \div 7^5$ is a rational number.) So if we clear our calculator at this point, we will introduce error that is likely to keep this process from ever ending. To avoid this problem, we leave the result in the calculator and simply subtract 3 from this to get the fractional part all by itself. **DO NOT ROUND OFF!** Subtraction and then multiplication by seven gives:

$$\begin{aligned}
 63201 \div 7^5 &= \textcircled{3}.760397453 \\
 0.760397453 \times 7 &= \textcircled{5}.322782174 \\
 0.322782174 \times 7 &= \textcircled{2}.259475219 \\
 0.259475219 \times 7 &= \textcircled{1}.816326531 \\
 0.816326531 \times 7 &= \textcircled{5}.714285714 \\
 0.714285714 \times 7 &= \textcircled{5}.000000000
 \end{aligned}$$

Yes, believe it or not, that last product is exactly 5, *as long as you don't clear anything out on your calculator*. This gives us our final result: $63201_{10} = 352155_7$. If we round, even to two decimal places in each step, clearing our calculator out at each step along the way, we will get a series of numbers that do not terminate, but begin repeating themselves endlessly. (Try it!) We end up with something that doesn't make any sense, at least not in this context. So be careful to use your calculator cautiously on these conversion problems. ♦

Checkpoint H

Convert the base 10 number, 9352_{10} , to base 5. See endnotes for answer.³⁷

Checkpoint I

Convert the base 10 number, 1500, to base 3. See endnotes for answer.³⁸ Be careful not to clear your calculator on this one. Also, if you're not careful in each step, you may not get all of the digits you're looking for, so move slowly and with caution.

The Mayan Numeral System

Background

As you might imagine, the development of a base system is an important step in making the counting process more efficient. Our own base-ten system probably arose from the fact that we have 10 fingers (including thumbs) on two hands. This is a natural development. However, other civilizations have had a variety of bases other than ten. For example, the Natives of Queensland used a base-two system, counting as follows: “one, two, two and one, two two’s, much.” Some Modern South American Tribes have a base-five system counting in this way: “one, two, three, four, hand, hand and one, hand and two,” and so on. The Babylonians used a base-sixty (sexagesimal) system that we will study more in a later chapter. In this chapter, we wrap up with a specific example of a civilization that actually used a base system other than 10.

The Mayan civilization is generally dated from 1500 B.C.E to 1700 C.E. The Yucatan Peninsula (see map³⁹) in Mexico was the scene for the development of one of the most advanced civilizations of the ancient world. The Mayans had a sophisticated ritual system that was overseen by a priestly class. This class of priests developed a philosophy with time as divine and eternal.⁴⁰ The calendar, and calculations related to it, were thus very important to the ritual life of the priestly class, and hence the Mayan people. In fact, much of what we know about this culture comes from their calendar records and astronomy data. (Another important source of information on the Mayans is the writings of Father Diego de Landa, who went to Mexico as a missionary in 1549.)



- | | |
|---------------------------------|--|
| 1. Chichen Itza | 5. Bonampak, Yaxchilan |
| 2. Uxmal | 6. Tikal |
| 3. Tulum | 7. Altun Ha |
| 4. Palenque | 8. Copán |

There were two numeral systems developed by the Mayans – one for the common people and one for the priests. Not only did these two systems use



different symbols, they also used different base systems. For the priests, the number system was governed by ritual. The days of the year were thought to be gods, so the formal symbols for the days were decorated heads.⁴¹ (See sample left⁴²) Since the basic calendar was based on 360 days, the priestly numeral system used a mixed base system employing multiples of 20 and 360. This makes for a confusing system, the details of which we will skip in this particular course.

The Mayan Number System

Instead, we will focus on the numeration system of the “common” people, which used a more consistent base system. As we stated earlier, the Mayans used a base-20 system, called the “vigesimal” system. Like our system, it is positional, meaning that the position of a numeric symbol indicates its place value. In the following table you can see the place value in its vertical format.⁴³

Powers	Base-Ten Value	Place Name
20^7	12,800,000,000	Hablat
20^6	64,000,000	Alau
20^5	3,200,000	Kinchil
20^4	160,000	Cabal
20^3	8,000	Pic
20^2	400	<i>Bak</i>
20^1	20	Kal
20^0	1	Hun

In order to write numbers down, there were only three symbols needed in this system. A horizontal bar represented the quantity 5, a dot represented the quantity 1, and a special symbol (thought to be a shell) represented zero. The Mayan system may have been the first to make use of zero as a placeholder/number. The first 20 numbers are shown in the table that follows.⁴⁴

Unlike our system, where the ones place starts on the right and then moves to the left, the Mayan systems places the ones on the bottom of a vertical orientation and moves up as the place value increases.

When numbers are written in vertical form, there should never be more than four dots in a single place. When writing Mayan numbers, every group of five dots becomes one bar. Also, there should never be more than three bars in a single place...four bars would be converted to one dot in the next place up. (It’s the same as 10 getting converted to a 1 in the next place up when we carry during addition.)

Example 12

What is the value of this number, which is shown in vertical form?

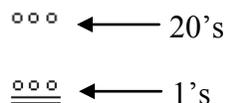
ooo

ooo

Number	Vertical Form	Number	Vertical Form
0		10	==
1	o	11	<u>o</u>
2	oo	12	<u>oo</u>
3	ooo	13	<u>ooo</u>
4	oooo	14	<u>oooo</u>
5	—	15	===
6	<u>o</u>	16	<u>oo</u>
7	<u>oo</u>	17	<u>ooo</u>
8	<u>ooo</u>	18	<u>oooo</u>
9	<u>oooo</u>	19	<u>oooo</u>

Solution

Starting from the bottom, we have the ones place. There are two bars and three dots in this place. Since each bar is worth 5, we have 13 ones when we count the three dots in the ones place. Looking to the place value above it (the twenties places), we see there are three dots so we have three twenties.



Hence we can write this number in base-ten as:

$$\begin{aligned} (3 \times 20^1) + (13 \times 20^0) &= (3 \times 20) + (13 \times 1) \\ &= 60 + 13 \\ &= 73 \end{aligned}$$



Example 13

What is the value of the following Mayan number?



Solution

This number has 11 in the ones place, zero in the 20's place, and 18 in the $20^2=400$'s place. Hence, the value of this number in base-ten is:

$$18 \times 400 + 0 \times 20 + 11 \times 1 = 7211. \quad \blacklozenge$$

CheckPoint J

Convert the Mayan number below to base 10. See the endnotes for the solution.⁴⁵



Example 14

Convert the base 10 number 3575_{10} to Mayan numerals.

Solution

This problem is done in two stages. First we need to convert to a base 20 number. We will do so using the method provided in the last section of the text. The second step is to convert that number to Mayan symbols.

The highest power of 20 that will divide into 3575 is $20^2 = 400$, so we start by dividing that and then proceed from there:

$$3575 \div 400 = 8.9375$$

$$0.9375 \times 20 = 18.75$$

$$0.75 \times 20 = 15.0$$

This means that $3575_{10} = 8,18,15_{20}$

The second step is to convert this to Mayan notation. This number indicates that we have 15 in the ones position. That's three bars at the bottom of the number. We also have 18 in the 20's place, so that's three bars and three dots in the second position. Finally, we have 8 in the 400's place, so that's one bar and three dots on the top. We get the following



◆

NOTE: We are using a new notation here. The commas between the three numbers 8, 18, and 15 are now separating place values for us so that we can keep them separate from each other. This use of the comma is slightly different than how they're used in the decimal system. When we write a number in base 10, such as 7,567,323, the commas are used primarily as an aide to read the number easily but they do not separate single place values from each other. We will need this notation whenever the base we use is larger than 10.

CheckPoint K

Convert the base 10 number 10553_{10} to Mayan numerals. See endnotes for answer.⁴⁶

CheckPoint L

Convert the base 10 number 5617_{10} to Mayan numerals. See endnotes for answer.⁴⁷

Adding Mayan Numbers

When adding Mayan numbers together, we'll adopt a scheme that the Mayans probably did not use but which will make life a little easier for us.

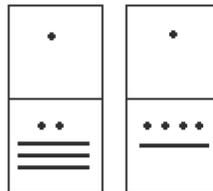
Example 15

Add, in Mayan, the numbers 37 and 29: ⁴⁸

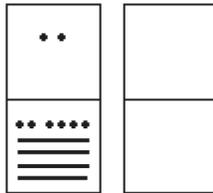


Solution

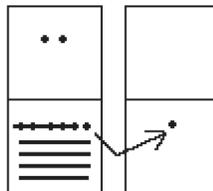
First draw a box around each of the vertical places. This will help keep the place values from being mixed up.



Next, put all of the symbols from both numbers into a single set of places (boxes), and to the right of this new number draw a set of empty boxes where you will place the final sum:

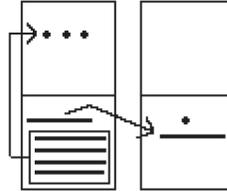


You are now ready to start carrying. Begin with the place that has the lowest value, just as you do with Arabic numbers. Start at the bottom place, where each dot is worth 1. There are six dots, but a maximum of four are allowed in any one place. (Once you get to five dots, you must convert to a bar.) Since five dots make one bar, we draw a bar through five of the dots, leaving us with one dot which is under the four-dot limit. Put this dot into the bottom place of the empty set of boxes you just drew:

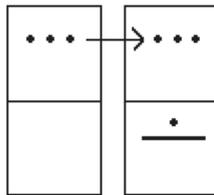


Now look at the bars in the bottom place. There are five, and the maximum number the place can hold is three. *Four bars are equal to one dot in the next highest place.* Whenever we have four bars in a single place we will automatically convert that to a *dot* in the next place up. So we draw a circle

around four of the bars and an arrow up to the dots' section of the higher place. At the end of that arrow, draw a new dot. That dot represents 20 just the same as the other dots in that place. Not counting the circled bars in the bottom place, there is one bar left. One bar is under the three-bar limit; put it under the dot in the set of empty places to the right.



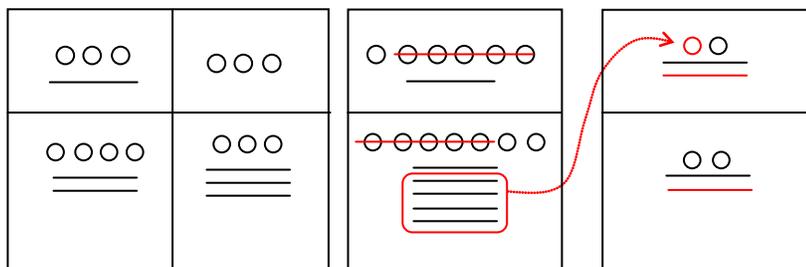
Now there are only three dots in the next highest place, so draw them in the corresponding empty box.



We can see here that we have 3 twenties (60), and 6 ones, for a total of 66. We check and note that $37 + 29 = 66$, so we have done this addition correctly. Is it easier to just do it in base-ten? Probably. But that's only because it's more familiar to you. Your task here is to try to learn a new base system and how addition can be done in slightly different ways than what you have seen in the past. Note, however, that the concept of carrying is still used, just as it is in our own addition algorithm. ♦

Checkpoint M

Try adding 174 and 78 in Mayan by first converting to Mayan numbers and then working entirely within that system. Do not add in base-ten (decimal) until the very end when you *check* your work. A sample solution is shown below, but you should try it on your own before looking at the one given.



Conclusion

In this first chapter, we have briefly sketched the development of numbers and our counting system, with the emphasis on the “brief” part. There are numerous sources of information and research that fill many volumes of books on this topic. Unfortunately, we cannot begin to come close to covering all of the information that is out there.

We have only scratched the surface of the wealth of research and information that exists on the development of numbers and counting throughout human history. What is important to note is that the system that we use every day is a product of thousands of years of progress and development. It represents contributions by many civilizations and cultures. It does not come down to us from the sky, a gift from the gods. It is not the creation of a textbook publisher. It is indeed as human as we are, as is the rest of mathematics. Behind every symbol, formula and rule there is a human face to be found, or at least sought.

Furthermore, I hope that you now have a basic appreciation for just how interesting and diverse number systems can get. Also, I’m pretty sure that you have also begun to recognize that we take our own number system for granted so much that when we try to adapt to other systems or bases, we find ourselves truly having to concentrate and think about what is going on. This is something that you are likely to experience even more as you study this chapter.

- 9) 100 in base 10 to base 2
 10) 2933 in base 10 to base 2
 11) Convert 653 in base 7 to base 10.
 12) Convert 653 in base 10 to base 7
 13) 3412 in base 5 to base 2
 14) 10011011 in base 2 to base 5
 (Hint: convert first to base 10 then to the final desired base)

The Caidoz System

Suppose you were to discover an ancient base-12 system made up twelve symbols. Let's call this base system the Caidoz system. Here are the symbols for each of the numbers 0 through 12:

0 = Υ	6 = Ω
1 = Υ	7 = \mathfrak{M}
2 = \mathbb{H}	8 = \times^{\wedge}
3 = \mathfrak{C}	9 = Υ_{\circ}
4 = \mathcal{O}	10 = \approx
5 = \mathfrak{M}	11 = \mathfrak{K}

Convert each of the following numbers in Caidoz to base 10

- 15) $\mathfrak{M}\Upsilon\approx$
 16) $\times^{\wedge}\mathfrak{C}\Upsilon\mathfrak{K}$
 17) $\Omega\mathcal{O}\mathbb{H}$
 18) $\mathfrak{M}\Upsilon\approx\Upsilon$

Convert the following base 10 numbers to Caidoz, using the symbols shown above.

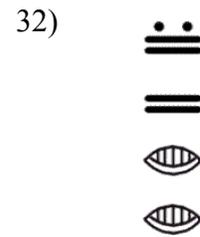
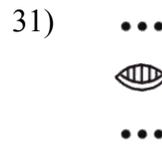
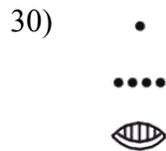
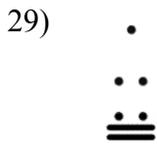
- 19) 175
 20) 3030
 21) 10,000
 22) 5507

Mayan Conversions

Convert the following numbers to Mayan notation. Show your calculations used to get your answers.

- 23) 135
 24) 234
 25) 360
 26) 1,215
 27) 10,500
 28) 1,100,000

Convert the following Mayan numbers to decimal (base-10) numbers. Show all calculations.



James Bidwell has suggested that Mayan addition was done by “simply combining bars and dots and carrying to the next higher place.” He goes on to say, “After the combining of dots and bars, the second step is to exchange every five dots for one bar in the same position.” After converting the following base 10 numbers into vertical Maya notation (in base 20, of course), perform the indicated addition:

33) $32 + 11$

34) $82 + 15$

35) $35 + 148$

36) $2412 + 5000$

37) $450 + 844$

38) $10,000 + 20,000$

39) $4,500 + 3,500$

40) $130,000 + 30,000$

41) Use the fact that the Mayans had a base-20 number system to complete the following multiplication table. The table entries should be in Mayan notation. Remember: Their zero looked like this... . Xerox and then cut out the table below, fill it in, and paste it onto your homework assignment if you do not want to duplicate the table with a ruler.

(To think about but not write up: Bidwell claims that only these entries are needed for “Mayan multiplication.” What does he mean?)

×	•	••	•••	••••	—	==	===
•							
••							
•••							
••••							
—							
==							
===							

Exploration

Write a short essay on the given topic. It should not be more than one page and if you can type it (double-spaced), I would appreciate it. If you cannot type it, your writing must be legible. Attention to grammar is important, although it does not have to be perfect grammatically...I just want to be able to understand it.

- 42) What are the advantages and disadvantages of bases other than ten.
- 43) Supposed you are charged with creating a base-15 number system. What symbols would you use for your system and why? Explain with at least two specific examples how you would convert between your base-15 system and the decimal system.
- 44) Describe an interesting aspect of Mayan civilization that we did not discuss in class. Your findings must come from some source such as an encyclopedia article, or internet site and you must provide reference(s) of the materials you used (either the publishing information or Internet address).
- 45) For a Papuan tribe in southeast New Guinea, it was necessary to translate the bible passage John 5:5 “And a certain man was there, which had an infirmity 30 and 8 years” into “A man lay ill one man, both hands, five and three years.” Based on your own understanding of bases systems (and some common sense), furnish an explanation of the translation. Please use complete sentences to do so. (Hint: To do this problem, I am asking you to think about how base systems work, where they come from, and how they are used. You won’t necessarily find an “answer” in readings or such...you’ll have to think it through and come up with a reasonable response. Just make sure that you clearly explain why the passage was translated the way that it was.)

Endnotes

- ¹ Eves, Howard; *An Introduction to the History of Mathematics*, p. 9.
- ² Eves, p. 9.
- ³ McLeish, John; *The Story of Numbers – How Mathematics Has Shaped Civilization*, p. 7.
- ⁴ Bunt, Lucas; Jones, Phillip; Bedient, Jack; *The Historical Roots of Elementary Mathematics*, p. 2.
- ⁵ http://www.math.buffalo.edu/mad/Ancient-Africa/mad_zaire-uganda.html
- ⁶ Diana, Lind Mae; The Peruvian Quipu in *Mathematics Teacher*, Issue 60 (Oct., 1967), p. 623–28.
- ⁷ Solution to **CheckPointA**: $1+6\times 3+3\times 6+2\times 12 = 61$ cats.
- ⁸ Diana, Lind Mae; The Peruvian Quipu in *Mathematics Teacher*, Issue 60 (Oct., 1967), p. 623–28.
- ⁹ <http://wiscinfo.doit.wisc.edu/chaysimire/titulo2/kipus/what.htm>
- ¹⁰ Diana, Lind Mae; The Peruvian Quipu in *Mathematics Teacher*, Issue 60 (Oct., 1967), p. 623–28.
- ¹¹ <http://www.cs.uidaho.edu/~casey931/seminar/quipu.html>
- ¹² http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Indian_numerals.html
- ¹³ <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Al-Biruni.html>
- ¹⁴ http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Indian_numerals.html
- ¹⁵ http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Indian_numerals.html
- ¹⁶ Ibid
- ¹⁷ Ibid
- ¹⁸ Ibid
- ¹⁹ Ibid
- ²⁰ Katz, page 230
- ²¹ Burton, David M., *History of Mathematics, An Introduction*, p. 254–255
- ²² Ibid
- ²³ Ibid
- ²⁴ Katz, page 231.
- ²⁵ Ibid, page 230
- ²⁶ Ibid, page 231.
- ²⁷ Ibid, page 232.
- ²⁸ Ibid, page 232.
- ²⁹ McLeish, p. 18
- ³⁰ http://seattletimes.nwsourc.com/news/health-science/html98/invs_20000201.html, Seattle Times, Feb. 1, 2000
- ³¹ Ibid, page 232.
- ³² Solution to **CheckPointC**: $41065_7 = 9994_{10}$
- ³³ Solution to **CheckPointD**: $143_{10} = 1033_5$
- ³⁴ Solution to **CheckPointE**: $21021_3 = 196_{10}$
- ³⁵ Solution to **CheckPointF**: $657_{10} = 22101_4$
- ³⁶ Solution to **CheckPointG**: $8377_{10} = 20271_8$
- ³⁷ Solution to **CheckPointH**: $9352_{10} = 244402_5$
- ³⁸ Solution to **CheckPointI**: $1500_{10} = 2001120_3$
- ³⁹ http://www.gorp.com/gorp/location/latamer/map_maya.htm
- ⁴⁰ Bidwell, James; Mayan Arithmetic in *Mathematics Teacher*, Issue 74 (Nov., 1967), p. 762–68.
- ⁴¹ <http://www.ukans.edu/~lctls/Mayan/numbers.html>
- ⁴² <http://www.ukans.edu/~lctls/Mayan/numbers.html>
- ⁴³ Bidwell
- ⁴⁴ http://www.vpds.wsu.edu/fair_95/gym/UM001.html
- ⁴⁵ Solution to **CheckPointJ**: 1562
- ⁴⁶ Solution to **CheckPointK**: $10553_{10} = 1,6,7,13_{20}$
- ⁴⁷ Solution to **CheckPointL**: $5617_{10} = 14,0,17_{20}$. Note that there is a zero in the 20's place, so you'll need to use the appropriate zero symbol in between the ones and 400's places.
- ⁴⁸ <http://forum.swarthmore.edu/k12/mayan.math/mayan2.html>

Solutions to Selected Exercises

Voting Theory

1.

Number of voters	3	3	1	3	2
1 st choice	A	A	B	B	C
2 nd choice	B	C	A	C	A
3 rd choice	C	B	C	A	B

3. a. $9+19+11+8 = 47$

b. 24 for majority; 17 for plurality

c. Atlanta, with 19 first-choice votes

d. Atlanta 94, Buffalo 111, Chicago 77. Winner: Buffalo

e. Chicago eliminated, 11 votes go to Buffalo. Winner: Buffalo

f. A vs B: B. A vs C: A. B vs C: B. B gets 2 pts, A 1 pt. Buffalo wins.

5. a. $120+50+40+90+60+100 = 460$

b. 231 for majority; 116 for plurality

c. A with 150 first choice votes

d. A 1140, B 1060, C 1160, D 1240. Winner: D

e. B eliminated, votes to C. D eliminated, votes to A. Winner: A

f. A vs B: B. A vs C: A. A vs D: D. B vs C: C. B vs D: D. C vs D: C

A 1pt, B 1pt, C 2pt, D 2pt. Tie between C and D.

Winner would probably be C since C was preferred over D

7. a. 33

b. 17

9. Yes, B

11. B, with 17 approvals

13. Independence of Irrelevant Alternatives Criterion

15. Condorcet Criterion

Weighted Voting

1. a. 9 players
b. $10+9+9+5+4+4+3+2+2 = 48$
c. 47
3. a. 9, a majority of votes
b. 17, the total number of votes
c. 12, which is $\frac{2}{3}$ of 17, rounded up
5. a. P1 is a dictator (can reach quota by themselves)
b. P1, since dictators also have veto power
c. P2, P3, P4
7. a. $11+7+2 = 20$
b. P1 and P2 are critical
9. Winning coalitions, with critical players underlined:
{P1,P2} {P1,P2,P3} {P1,P2,P4} {P1,P2,P3,P4} {P1,P3} {P1,P3,P4}
P1: 6 times, P2: 2 times, P3: 2 times, P4: 0 times. Total: 10 times
Power: P1: $\frac{6}{10} = 60\%$, P2: $\frac{2}{10} = 20\%$, P3: $\frac{2}{10} = 20\%$, P4: $\frac{0}{10} = 0\%$
11. a. {P1} {P1,P2} {P1,P3} {P1,P4} {P1,P2,P3} {P1,P2,P4} {P1,P3,P4} {P1,P2,P3,P4}
P1: 100%, P2: 0%, P3: 0%, P4: 0%
b. {P1,P2} {P1,P3} {P1,P4} {P1,P2,P3} {P1,P2,P4} {P1,P3,P4} {P1,P2,P3,P4}
P1: $\frac{7}{10} = 70\%$, P2: $\frac{1}{10} = 10\%$, P3: $\frac{1}{10} = 10\%$, P4: $\frac{1}{10} = 10\%$
c. {P1,P2} {P1,P3} {P1,P2,P3} {P1,P2,P4} {P1,P3,P4} {P1,P2,P3,P4}
P1: $\frac{6}{10} = 60\%$, P2: $\frac{2}{10} = 20\%$, P3: $\frac{2}{10} = 20\%$, P4: $\frac{0}{10} = 0\%$
13. $P3 = 5$. $P3+P2 = 14$. $P3+P2+P1 = 27$, reaching quota. P1 is critical.
15. Sequential coalitions with pivotal player underlined
<P1,P2,P3> <P1,P3,P2> <P2,P1,P3> <P2,P3,P1> <P3,P1,P2> <P3,P2,P1>
P1: $\frac{2}{6} = 33.3\%$, P2: $\frac{2}{6} = 33.3\%$, P3: $\frac{2}{6} = 33.3\%$
17. a. 6, 7
b. 8, given P1 veto power
c. 9, given P1 and P2 veto power
19. If adding a player to a coalition could cause it to reach quota, that player would also be critical in that coalition, which means they are not a dummy. So a dummy cannot be pivotal.

21. We know P_2+P_3 can't reach quota, or else P_1 wouldn't have veto power.
 P_1 can't reach quota alone.
 P_1+P_2 and P_1+P_3 must reach quota or else P_2/P_3 would be dummy.
- a. $\{P_1, P_2\}$ $\{P_1, P_3\}$ $\{P_1, P_2, P_3\}$. $P_1: 3/5, P_2: 1/5, P_3: 1/5$
- b. $\langle P_1, P_2, P_3 \rangle \langle P_1, P_3, P_2 \rangle \langle P_2, P_1, P_3 \rangle \langle P_2, P_3, P_1 \rangle \langle P_3, P_1, P_2 \rangle \langle P_3, P_2, P_1 \rangle$
 $P_1: 4/6, P_2: 1/6, P_3: 1/6$
23. [4: 2, 1, 1, 1] is one of many possibilities
25. [56: 30, 30, 20, 20, 10]
27. [54: 10, 10, 10, 10, 10, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1] is one of many possibilities

Fair Division

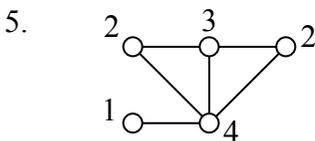
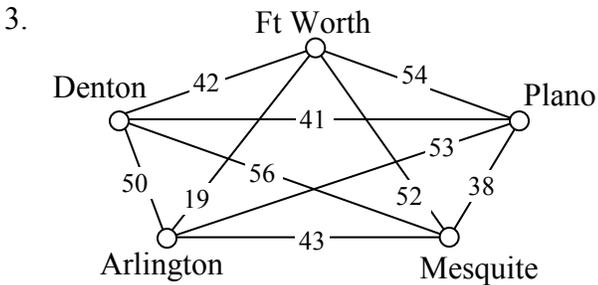
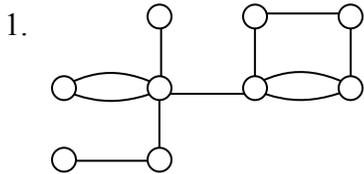
1. Chance values the veggie half at \$7.50 and pepperoni half at \$2.50.
 A full pepperoni slice is $\frac{1}{4}$ of the pepperoni half. Value $\$2.50/4 = \0.625
 A full veggie slice is $\frac{1}{4}$ of the veggie half. Value $\$7.50/4 = \1.875
 A slice that is $\frac{1}{2}$ pepperoni $\frac{1}{2}$ veggie is value $\$0.3125 + \$0.9375 = \$1.25$
3. Erin: Bowl 1, Catherine: Bowl 2, Shannon: Bowl 3
5. a. 25 Snickers @ \$0.01 each, 20 Milky Ways @ \$0.05 each, 60 Reese's @ \$0.02 each
 Value: $\$0.25 + \$1.00 + \$1.20 = \2.45
- b. No. Dustin values the whole bag at \$8, so a fair share would be \$4.
- c. Lots of possibilities. Here's a couple:
 80 Milky Ways, 0 Snickers, 0 Reese's
 50 Snickers, 50 Milky Ways, 50 Reese's
7. a. Zoe
 b. Maggie: s2, s3. Meredith: s1, s2. Holly: s3
 c. Maggie: s2, Meredith: s1, Holly: s3, Zoe: s4
9. a. P5
 b. \$6.50 (doesn't need to trim it much since they're last)
 c. P4 would receive it, with value \$6.00 (since P4 would trim it)
11. a. $(320+220)/4 = \$135$
 b. Desk and Vanity both go to A. A pays $\$320 + \$220 - \$135 = \405 to estate
 B gets \$95, C gets \$125, D gets \$110.
 c. Surplus of $\$405 - \$95 - \$125 - \$110 = \$75$ gets split, \$18.75 each.
 A gets desk and vanity, pays \$386.25 to estate
 B gets \$113.75, C gets \$143.75, D gets \$128.75

13. Fair shares: Abby: 10.333, Ben: 9, Carla: 7.667
 Motorcycle to Abby, Car to Ben, Tractor to Abby, Boat to Abby
 Initial: Abby pays \$10.667, Ben pays \$2, Carla gets \$7.667
 Surplus: \$5; \$1.667 each
 Final: Abby gets Motorcycle, Tractor and Boat, pays \$9
 Ben gets Car, pays \$0.333
 Carla gets \$9.334

15. Fair shares: Sasha: \$135, Megan: \$130
 Sasha gets: Couch, detail cleaning. Value \$80
 Megan gets: TV, Stereo, carpets. Value: \$260
 Initial: Sasha gets \$55, Megan pays \$130.
 Surplus: \$75; \$37.50 each
 Final: Sasha gets Couch and does detail cleaning, gets \$92.50
 Megan gets TV and stereo, and cleans carpets, pays \$92.50

17. a. s3, worth \$270
 b. s1 and s4 have combined value \$440 for Greedy, so piece would be worth \$220

Graph Theory

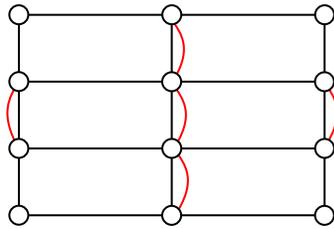


7. The first and the third graphs are connected

9. Bern to Frankfurt to Munchen to Berlin: 12hrs 50 min. (Though trip through Lyon, Paris and Amsterdam only adds 30 minutes)

11. The first two graphs have Euler circuits. The last graph has two vertices with odd degree

13. One of several possible eulerizations requiring 5 duplications:

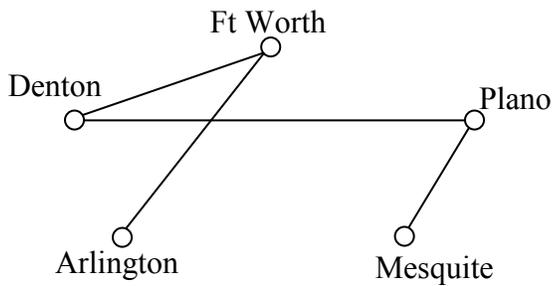


17. Only the middle graph has a Hamiltonian circuit.

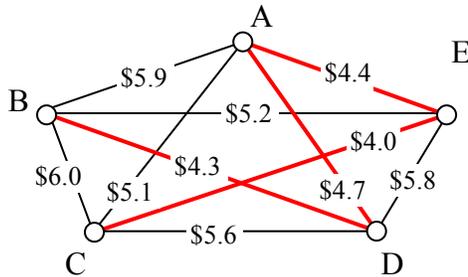
19. a. Ft Worth, Arlington, Mesquite, Plano, Denton, Ft Worth: 183 miles
 b. Same as part a
 c. Same as part a

21. a. ABDCEA
 b. ACEBDA
 c. ADBCEA

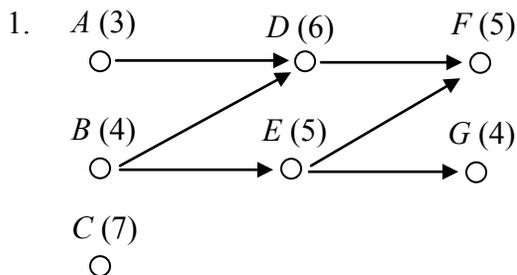
23.



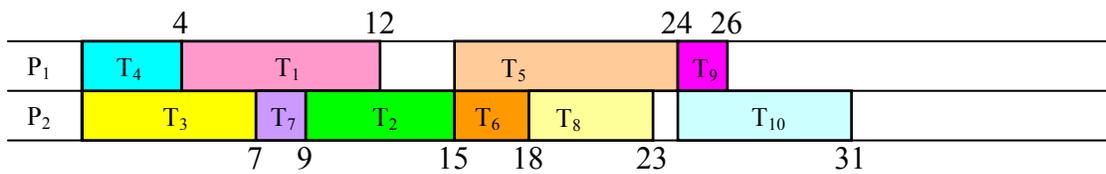
25.



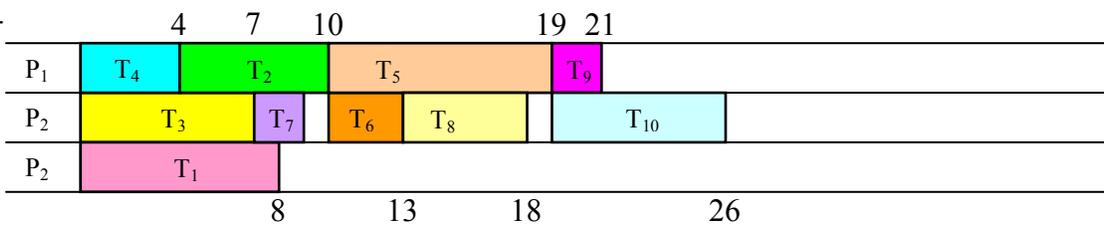
Scheduling



3.

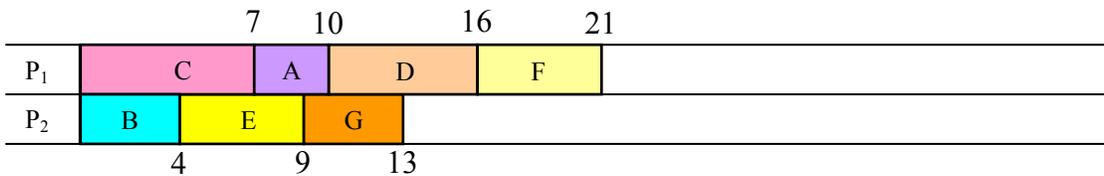


5.

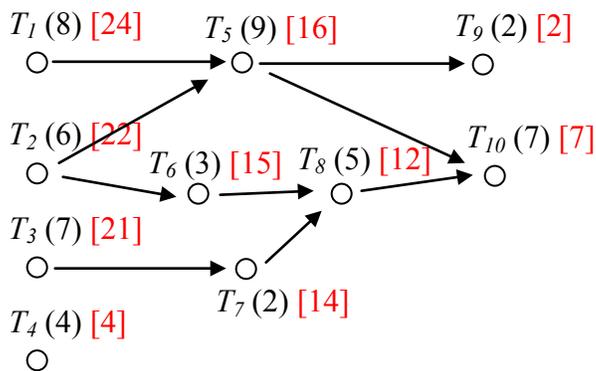


7. Priority List: T₅, T₁, T₃, T₁₀, T₂, T₈, T₄, T₆, T₇, T₉

9. Priority List: C, D, E, F, B, G, A



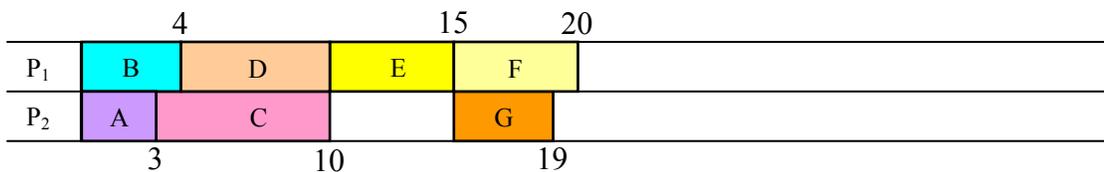
11. a.



b. Critical path: T₁, T₅, T₁₀. Minimum completion time: 24

c. Critical path priority list: T₁, T₂, T₃, T₅, T₆, T₇, T₈, T₁₀, T₄, T₉

13. Critical path priority list: B, A, D, E, C, F, G



Growth Models

1. a. $P_0 = 20$. $P_n = P_{n-1} + 5$
b. $P_n = 20 + 5n$
3. a. $P_1 = P_0 + 15 = 40 + 15 = 65$. $P_2 = 65 + 15 = 80$
b. $P_n = 40 + 15n$
c. $P_{10} = 40 + 15(10) = 190$ thousand dollars
5. Grew 64 in 8 weeks: 8 per week
a. $P_n = 3 + 8n$
b. $187 = 3 + 8n$. $n = 23$ weeks
7. a. $P_0 = 200$ (thousand), $P_n = 1.09n$, where n is years after 2000
b. $P_n = 200(1.09)^n$
c. $P_{16} = 200(1.09)^{16} = 794.061$ (thousand) = 794,061
9. Let $n=0$ be 1983. $P_n = 1700(2.9)^n$. 2005 is $n=22$. $P_{22} = 1700(2.9)^{22} = 25,304,914,552,324$ people. Clearly not realistic, but mathematically accurate.
11. If n is in hours, better to start with the explicit form. $P_0 = 300$. $P_4 = 500 = 300(1+r)^4$
 $500/300 = (1+r)^4$. $1+r = 1.136$. $r = 0.136$
a. $P_0 = 300$. $P_n = (1.136)P_{n-1}$
b. $P_n = 300(1.136)^n$
c. $P_{24} = 300(1.136)^{24} = 6400$ bacteria
13. a. $P_0 = 100$ $P_n = P_{n-1} + 0.70(1 - P_{n-1}/2000)P_{n-1}$
b. $P_1 = 100 + 0.70(1 - 100/2000)(100) = 166.5$
c. $P_2 = 166.5 + 0.70(1 - 166.5/2000)(166.5) = 273.3$
15. To find the growth rate, suppose $n=0$ was 1968. Then P_0 would be 1.60 and $P_8 = 2.30 = 1.60(1+r)^8$, $r = 0.0464$. Since we want $n=0$ to correspond to 1960, then we don't know P_0 , but P_8 would $1.60 = P_0(1.0464)^8$. $P_0 = 1.113$.
a. $P_n = 1.113(1.0464)^n$
b. $P_0 = \$1.113$, or about \$1.11
c. 1996 would be $n=36$. $P_{36} = 1.113(1.0464)^{36} = \5.697 . Actual is slightly lower.
17. The population in the town was 4000 in 2005, and is growing by 4% per year.

Finance

1. $A = 200 + .05(200) = \$210$
3. $I=200$. $t = 13/52$ (13 weeks out of 52 in a year). $P_0 = 9800$
 $200 = 9800(r)(13/52)$ $r = 0.0816 = 8.16\%$ annual rate

5. $P_{10} = 300(1 + .05/1)^{10(1)} = \488.67

7. a. $P_{20} = 2000(1 + .03/12)^{20(12)} = \3641.51 in 20 years

b. $3641.51 - 2000 = \$1641.51$ in interest

9. $P_8 = P_0(1 + .06/12)^{8(12)} = 6000$. $P_0 = \$3717.14$ would be needed

11. a. $P_{30} = \frac{200((1 + 0.03/12)^{30(12)} - 1)}{0.03/12} = \$116,547.38$

b. $200(12)(30) = \$72,000$

c. $\$116,547.40 - \$72,000 = \$44,547.38$ of interest

13. a. $P_{30} = 800,000 = \frac{d((1 + 0.06/12)^{30(12)} - 1)}{0.06/12}$ $d = \$796.40$ each month

b. $\$796.40(12)(30) = \$286,704$

c. $\$800,000 - \$286,704 = \$513,296$ in interest

15. a. $P_0 = \frac{30000(1 - (1 + 0.08/1)^{-25(1)})}{0.08/1} = \$320,253.29$

b. $30000(25) = \$750,000$

c. $\$750,000 - \$320,253.29 = \$429,756.71$

17. $P_0 = 500,000 = \frac{d(1 - (1 + 0.06/12)^{-20(12)})}{0.06/12}$ $d = \$3582.16$ each month

19. a. $P_0 = \frac{700(1 - (1 + 0.05/12)^{-30(12)})}{0.05/12} =$ a $\$130,397.13$ loan

b. $700(12)(30) = \$252,000$

c. $\$252,200 - \$130,397.13 = \$121,602.87$ in interest

21. $P_0 = 25,000 = \frac{d(1 - (1 + 0.02/12)^{-48})}{0.02/12} = \542.38 a month

23. a. Down payment of 10% is \$20,000, leaving \$180,000 as the loan amount

b. $P_0 = 180,000 = \frac{d(1 - (1 + 0.05/12)^{-30(12)})}{0.05/12}$ $d = \$966.28$ a month

c. $P_0 = 180,000 = \frac{d(1 - (1 + 0.06/12)^{-30(12)})}{0.06/12}$ $d = \$1079.19$ a month

25. First 5 years: $P_5 = \frac{50((1 + 0.08/12)^{5(12)} - 1)}{0.08/12} = \3673.84

Next 25 years: $3673.84(1 + .08/12)^{25(12)} = \$26,966.65$

27. Working backwards, $P_0 = \frac{10000(1 - (1 + 0.08/4)^{-10(4)})}{0.08/4} = \$273,554.79$ needed at retirement. To end up with that amount of money, $273,554.79 = \frac{d((1 + 0.08/4)^{15(4)} - 1)}{0.08/4}$. He'll need to contribute $d = \$2398.52$ a quarter.

Statistics

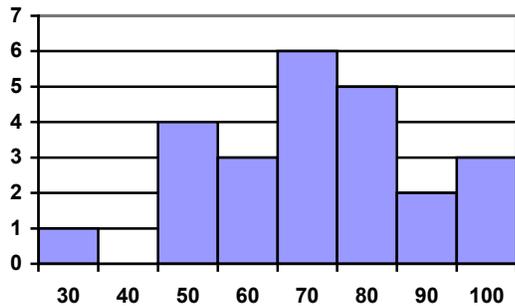
1. a. Population is the current representatives in the state's congress
 - b. 106
 - c. the 28 representatives surveyed
 - d. 14 out of 28 = $\frac{1}{2}$ = 50%
 - e. We might expect 50% of the 106 representatives = 53 representatives
3. This suffers from leading question bias
5. This question would likely suffer from a perceived lack of anonymity
7. This suffers from leading question bias
9. Quantitative
11. Observational study
13. Stratified sample
15. a. Group 1, receiving the vaccine
 - b. Group 2 is acting as a control group. They are not receiving the treatment (new vaccine).
 - c. The study is at least blind. We are not provided enough information to determine if it is double-blind.
 - d. This is a controlled experiment
17. a. Census
 - b. Observational study

Describing Data

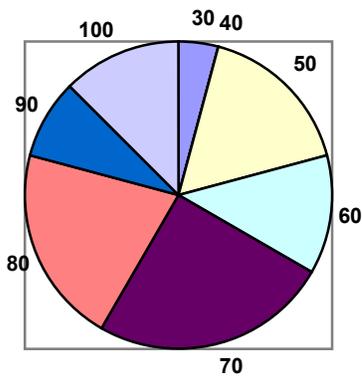
1. a. Different tables are possible

Score	Frequency
30	1
40	0
50	4
60	3
70	6
80	5
90	2
100	3

b. This is technically a bar graph, not a histogram:



c.



3. a. $5+3+4+2+1 = 15$

b. $5/15 = 0.3333 = 33.33\%$

5. Bar is at 25%. 25% of 20 = 5 students earned an A

7. a. $(7.25+8.25+9.00+8.00+7.25+7.50+8.00+7.00)/8 = \7.781

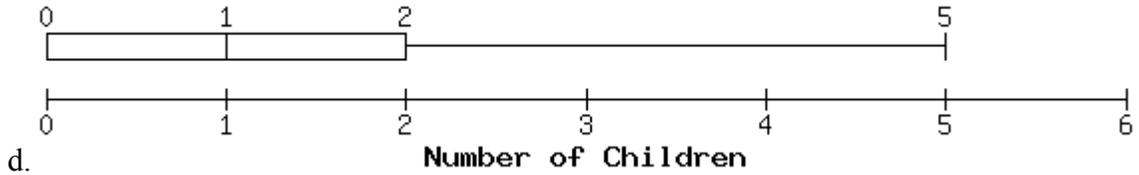
b. In order, 7.50 and 8.00 are in middle positions. Median = \$7.75

c. $0.25*8 = 2$. Q1 is average of 2nd and 3rd data values: \$7.375

$0.75*8 = 6$. Q3 is average of 6th and 7th data values: \$8.125

5-number summary: \$7.00, \$7.375, \$7.75, \$8.125, \$9.00

9. a. $(5*0 + 3*1 + 4*2 + 2*3 + 1*5)/15 = 1.4667$
 b. Median is 8th data value: 1 child
 c. $0.25*15 = 3.75$. Q1 is 4th data value: 0 children
 $0.75*15 = 11.25$. Q3 is 12th data value: 2 children
 5-number summary: 0, 0, 1, 2, 5



11. Kendra makes \$90,000. Kelsey makes \$40,000. Kendra makes \$50,000 more.

Historical Counting

1. Partial answer: Jars: 3 singles, 3 @ x2, 2 @ x6, 1 @ x12. $3+6+12+12 = 33$

3. 113

5. 3022

7. 53

9. 1100100

11. 332

13. 111100010

15. 7,1,10 base 12 = 1030 base 10

17. 6,4,2 base 12 = 914 base 10

19. 175 base 10 = 1,2,7 base 12 = $\text{I} \text{II} \text{M}$

21. 10000 base 10 = 5,9,5,4 base 12 = $\text{M} \text{V} \text{M} \text{Q}$

23. $135 = 6,15$ base 20 =

25. $360 = 18,0$ base 20 =

27. $10500 = 1,6,5,0$ base 20

$$\begin{array}{r} \bullet \\ \bullet \\ \hline \hline \end{array}$$


29. $1,2,12$ base 20 = 452 base 10

31. $3,0,3$ base 20 = 1203 base 10

33.

$$\begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{array} + \begin{array}{c} \bullet \\ \bullet \end{array} = \begin{array}{ccc} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{array}$$

35.

$$\begin{array}{ccc} & & \bullet \\ \bullet & \bullet & \bullet \\ \hline \end{array} + \begin{array}{ccc} & & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \hline \end{array} = \begin{array}{ccc} & & \bullet \\ \bullet & \bullet & \bullet \\ \hline \bullet & \bullet & \bullet \\ \hline \hline \end{array}$$