Math 89 Math of Voting - Spring 2020

§Course Info

Welcome to Math 89: Math of Voting

What do you hope to get out of this class?

Possible topics

- Gerrymandering
- Ranked voting systems
- Arrow's Impossibility Theorem
- Approval voting
- Strategic voting
- Weighted voting systems
- Power indices
- Electoral College
- Referendums
- Apportionment
- Polls
- Election Fraud

What to expect from this class.

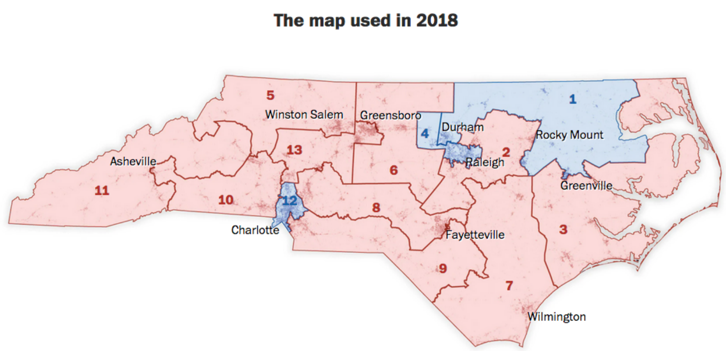
- Lots of questions
- Active participation
- Quizzes and homework assignments
  - Quizzes once every two weeks
  - Homework assigned on Sakai after every class, due once a week
- Final project

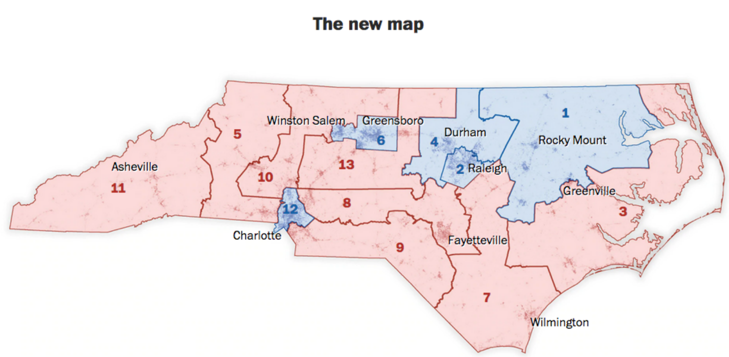
Resources for this class

- Materials posted on Sakai
  - Notes posted before and after class
  - Homework posted after class
  - Wiki to gather information about voting topics
- Textbook: *Math of Voting and Elections: A Hands-On Approach*
- Office hours: M 3:30 - 4:30, W 3:30 - 4:30, Th 4 - 5, and by appointment in Phillips 338
- Piazza
- Peers
- Math Help Center?

§12.1 Gerrymandering - Introduction

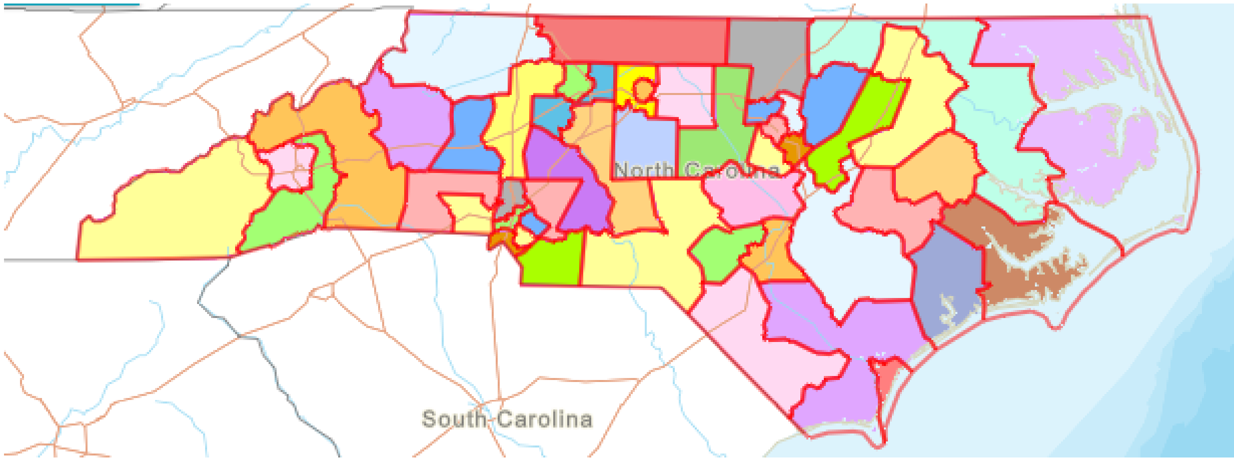
How are districts used to elect representatives to the the House of Representatives? to state legislative bodies?





Washington Post *North Carolina has a new congressional map for 2020*, Dec 3, 2019

State Legislature Districts

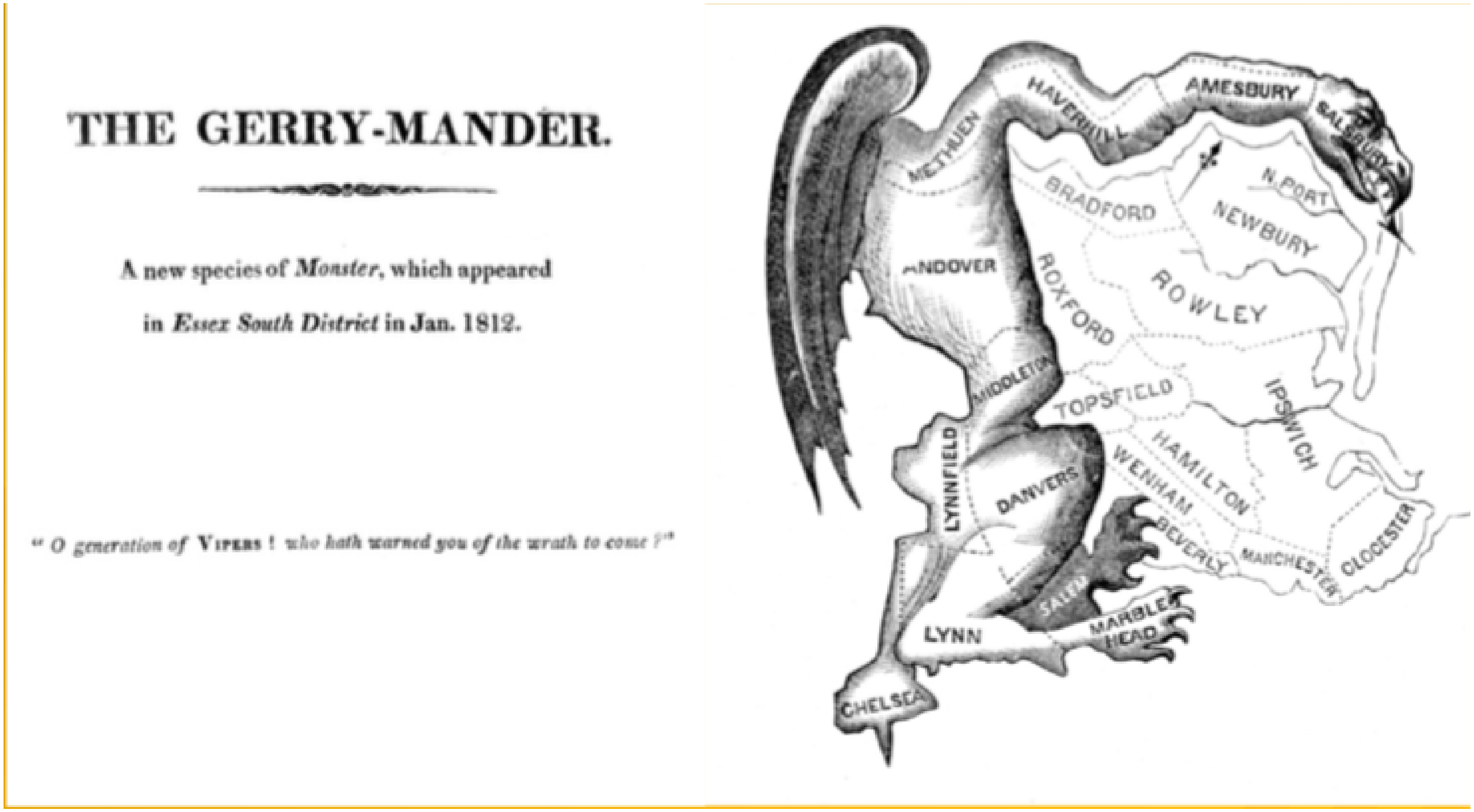


https://www.ncleg.gov/rnr/representation

Gerrymandering is ...

What is the difference between gerrymandering and redistricting?

Where does the term “gerrymandering” come from?

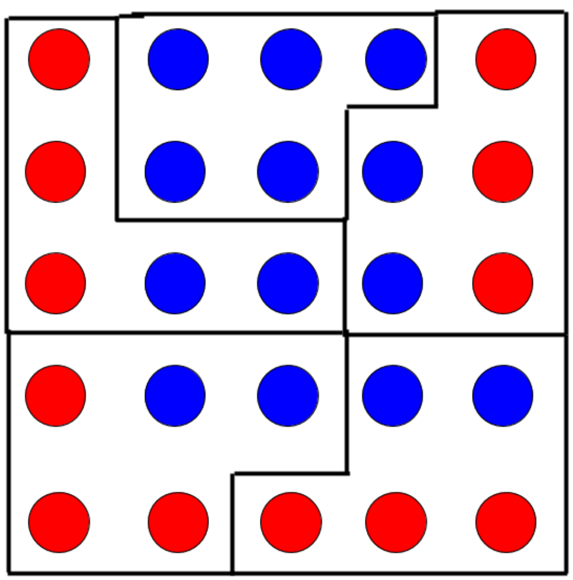


Boston Gazette March 1812

Try your hand at gerrymandering.

1. Draw a 6×6 grid of circles and color 19 of them red and 17 of them blue at random. These tokens represent two political parties, e.g. Republican and Democrat.
2. Draw districts of 3 voters per district so that that there are an equal number of red districts and blue districts. Note that a district's color depends on the majority of voters in that district.
3. Now, draw districts to get as many red districts as possible. How many could you achieve? What is your percent of red districts vs. your percent or red voters?
4. Next, draw districts so that as many blue districts as possible. How many could you achieve? What is your percent of blue districts vs. your percent or blue voters?
5. What methods did you use to gerrymander?

Two main techniques of gerrymandering are called “packing” and “cracking”. What do these terms mean?



North Carolina recent history of gerrymandering:

- 2011: districts for House of Representatives redrawn, giving 9-4 Republican edge
- 2016: legislature ordered to redraw districts for House of Representatives, because districts drawn in 2011 were determined to be racially gerrymandered. New districts gave Republicans 10-3 edge

  "I propose that we draw the maps to give a partisan advantage to 10 Republicans and three Democrats, because I do not believe it's possible to draw a map with 11 Republicans and two Democrats," David R. Lewis, a North Carolina state representative
- 2018: a panel of three federal judges again declared North Carolina's congressional district map to be unconstitutional, ruling that it was gerrymandered to unfairly favor Republican candidates; however, Supreme Court ordered a stay, and this map was used in 2018 elections
- June 2019: Supreme Court ruled that gerrymandering for political party advantage is not unconstitutional:

  "We conclude that partisan gerrymandering claims present political questions beyond the reach of the federal courts," Chief Justice John Roberts wrote.
- On Sep 3, 2019, a state court ruled that North Carolina's legislative districts were impermissibly gerrymandered and must be redrawn.

- October 28, 2019, a state court barred North Carolina from using its congressional district maps in the 2020 elections.

Topics to investigate on the history and laws around districting and gerrymandering

Questions about current rules:

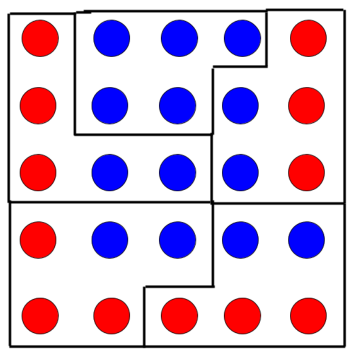
1. What does the US constitution say about districts for electing members of Congress?
2. What does the US constitution say about districts for electing representatives state legislatures?
3. What methods do states use to draw congressional districts?
4. How equal in size do congressional and state legislature districts have to be?
5. What are the current rules and standards for drawing districts, (i) universal to all states of the US, and (ii) specific to North Carolina?
6. What is Section 2 of the Voting Rights Act of 1965 and how has this been used in court cases to address gerrymandering?
7. What are majority-minority districts and what laws or court cases established that districting to create majority-minority districts is ok?
8. Does creating majority-minority districts help or hinder minorities from getting their voice heard? Do they tend to help Democrats or Republicans?
9. What is in the current news about gerrymandering?
10. For each of these Supreme Court cases, what was the case about and what precedent

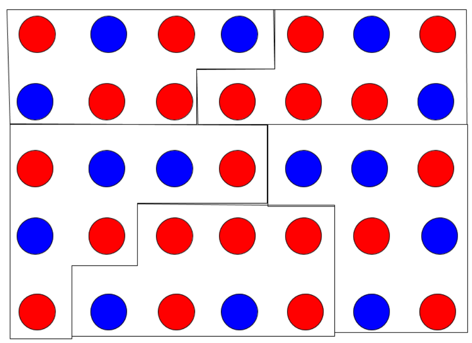
did it establish?

(a) Thornburg v. Gingles (1986)
(b) Larios v. Cox (2000)
(c) Vieth v. Jubeliere (2004)
(d) Harris v. Arizona Independent Redistricting (2016)
(e) Cooper v. Harris (2017)
(f) Rucho v. Common Cause and Lamone v. Benisek, (2019)
(g) Other cases?

§12.2 Gerrymandering and Efficiency Gap

How can you quantify the relative amount of gerrymandering in these two districting maps in terms of *wasted votes*?





Definition. The efficiency gap is ...

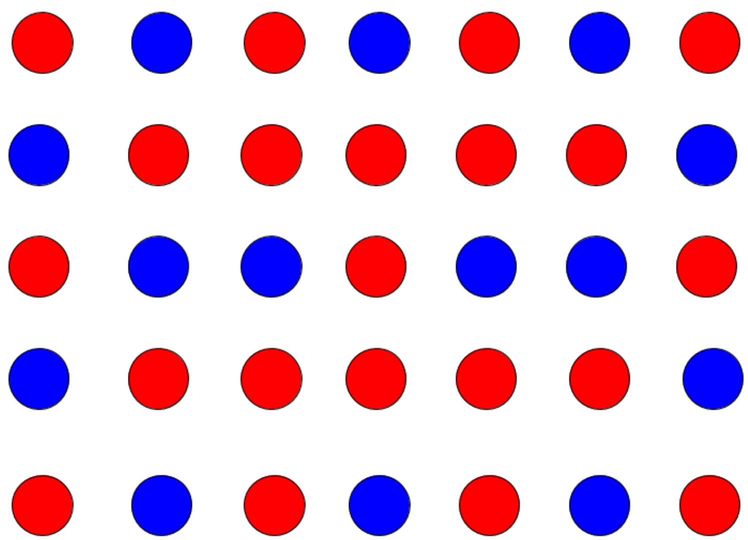
Question. Does a higher efficiency gap correspond to districts that are more or less gerrymandered?

Question. What is the maximum possible efficiency gap? Minimum possible?

Example. Calculate the efficiency gaps for the districts you created in the previous gerrymandering DIY exercise.

Extra Example. Try this gerrymandering puzzle.

- Create 5 districts, with as much advantage as possible for blue.
- Create 5 districts, with as much advantage as possible for red.
- What are the efficiency gaps?



Example. Calculate the efficiency gap for North Carolina districts for the 2016 election, and compare to the efficiency gaps in other years. See https://er.ncsbe.gov/ for data. There is a spreadsheet on google drive, linked to on Sakai, that can help you organize this data.

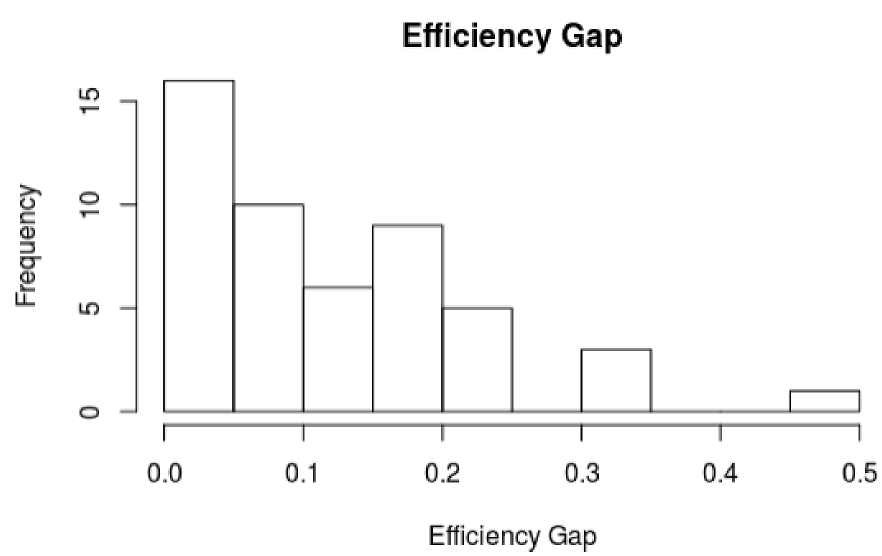
Example. Also look at Maryland for recent years.
See https://elections.maryland.gov/elections/2020/index.html for data.

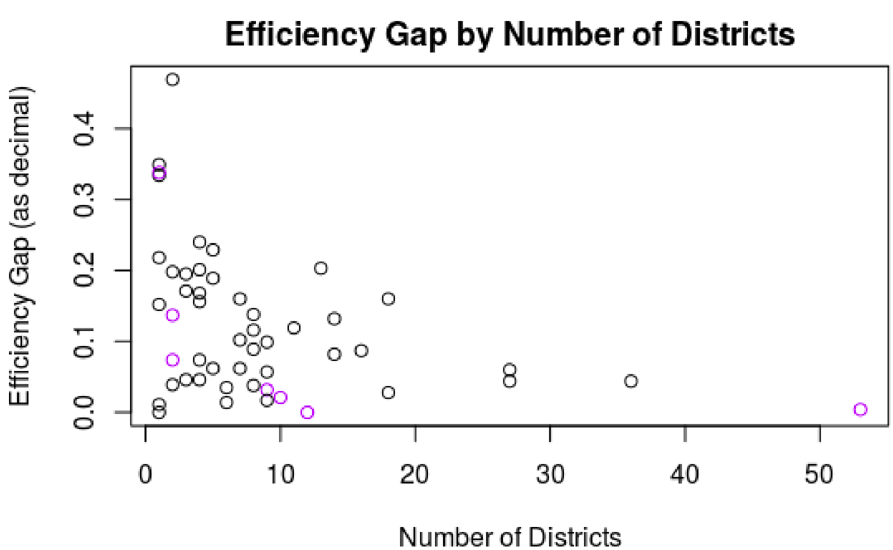
How do the efficiency gaps compare for various years (2010 - 2018) and between states (NC and Maryland)?

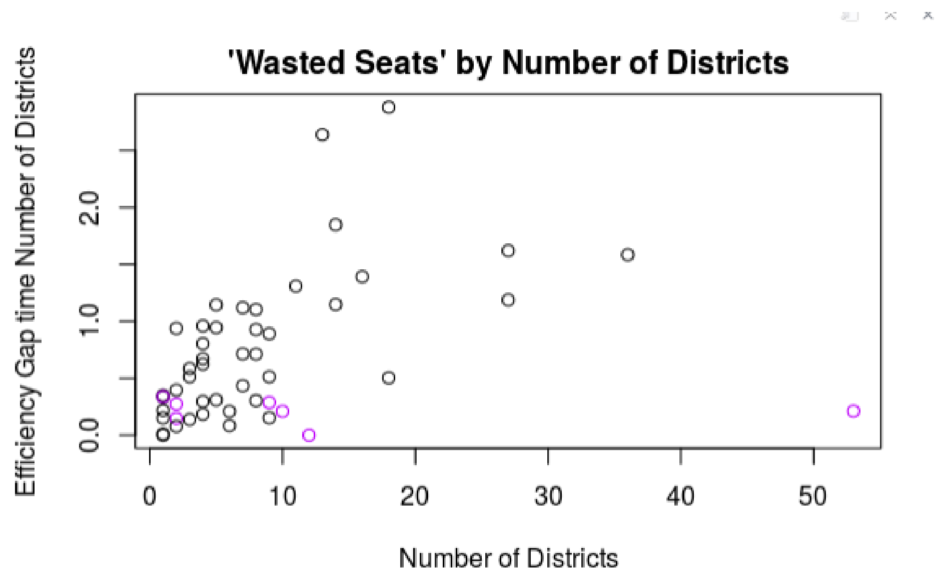
Is the efficiency gap a good measure of gerrymandering? What other information would be useful if we wanted to convince a court that a state does or does not have gerrymandered districts?

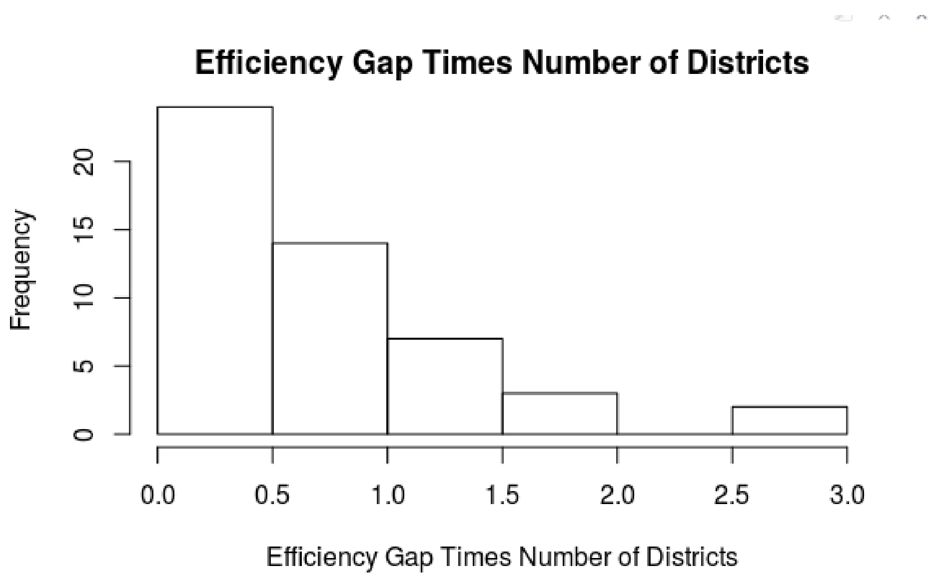
Look at the file of efficiency gaps for each state for the 2016 congressional elections. What do you observe? Set a criteria for how to tell if an efficiency gap is too high, and be ready to defend your choice.

Question. Look at the file of efficiency gaps for each state for the 2016 congressional elections. What do you observe? Set a criteria for how to tell if an efficiency gap is too high, and be ready to defend your choice.


Efficiency Gap
Frequency
0
5
10
15
0.0
0.1
0.2
0.3
0.4
0.5
Efficiency Gap

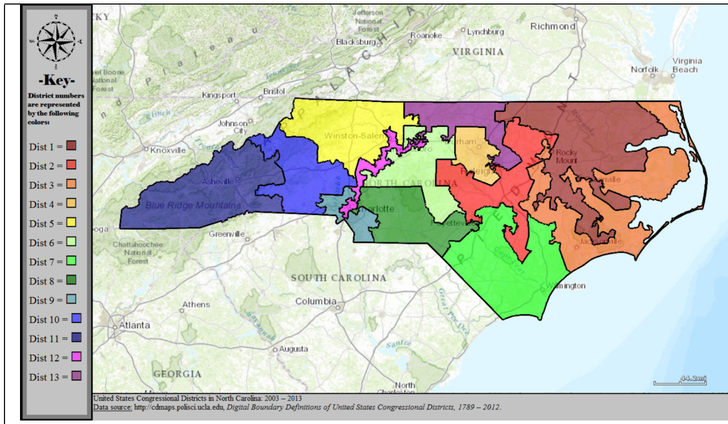

Efficiency Gap by Number of Districts
Efficiency Gap (as decimal)
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
Number of Districts


'Wasted Seats' by Number of Districts
Efficiency Gap time Number of Districts
0.0
1.0
2.0
0
10
20
30
40
50
Number of Districts

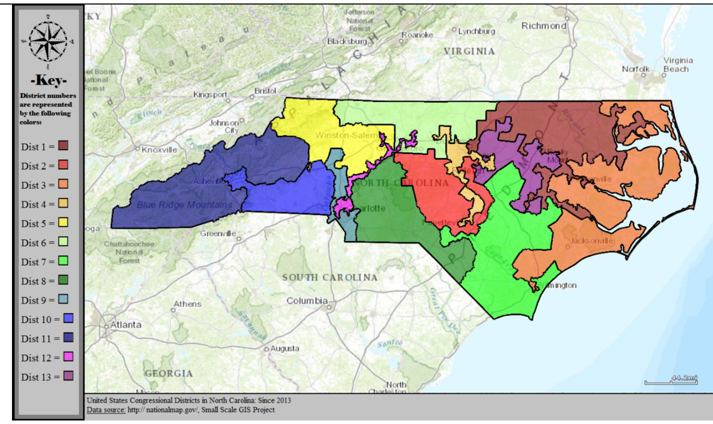

Efficiency Gap Times Number of Districts
Frequency
0
5
10
15
20
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Efficiency Gap Times Number of Districts

S12.3 Gerrymandering and Compactness

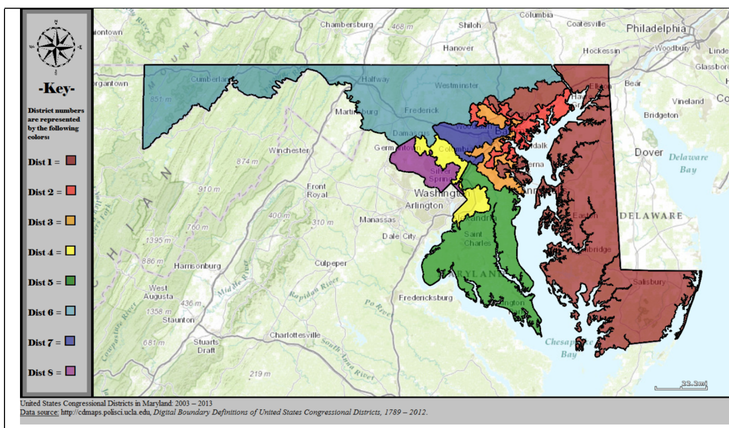
Look at the maps for NC and MD for some years.



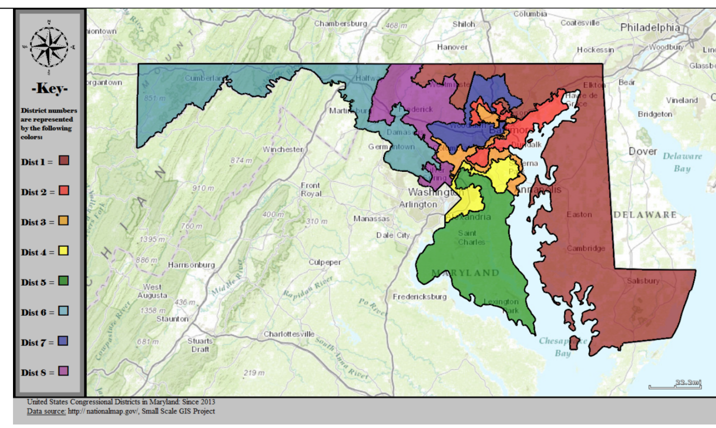
used to elect 2003 - 2013 House of Reps



drawn in 2011, used to elect 2013 - 2018 House of Reps



used to elect 2003 - 2013 House of Reps



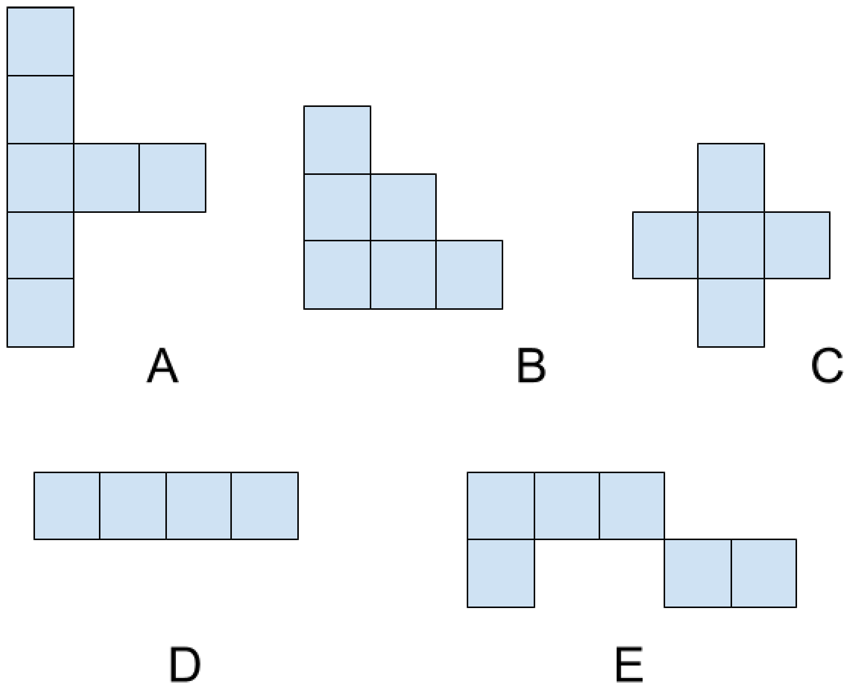
drawn in 2011, used to elect 2013 - 2020 House of Reps

What features of these district shapes suggest gerrymandering?

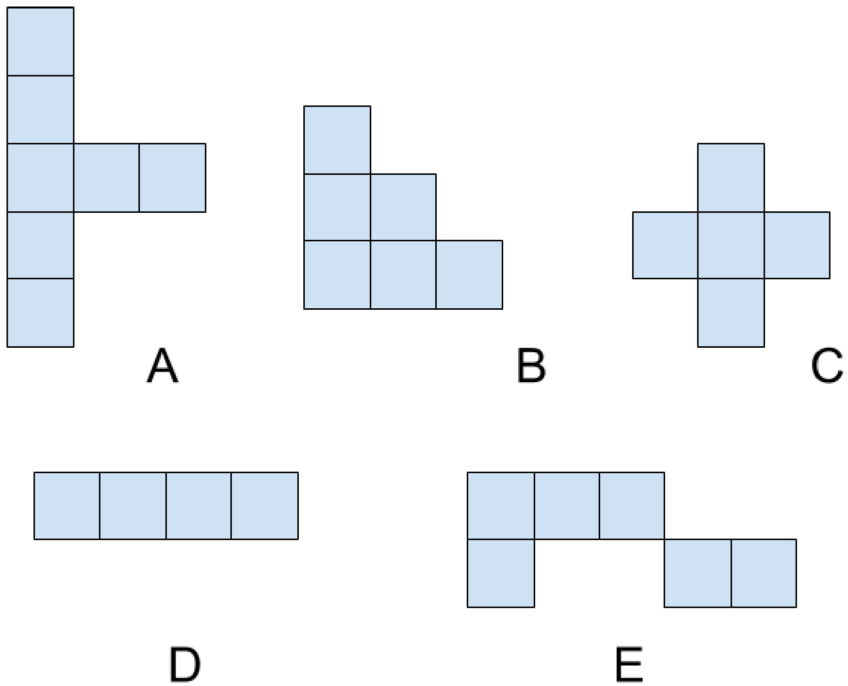
What features would you expect the shape of an un-gerrymandered district to have?

Goal: quantify gerrymandering looking only at the shape of the districts.

Based on intuition only, rank these shapes from most to least "compact" . .



Come up with one or two ways to measure / quantify “compactness”.



Use your measures to compute the compactness of one these simple hypothetical districts.

Description of measures of compactness:

1. Polsby-Popper method:

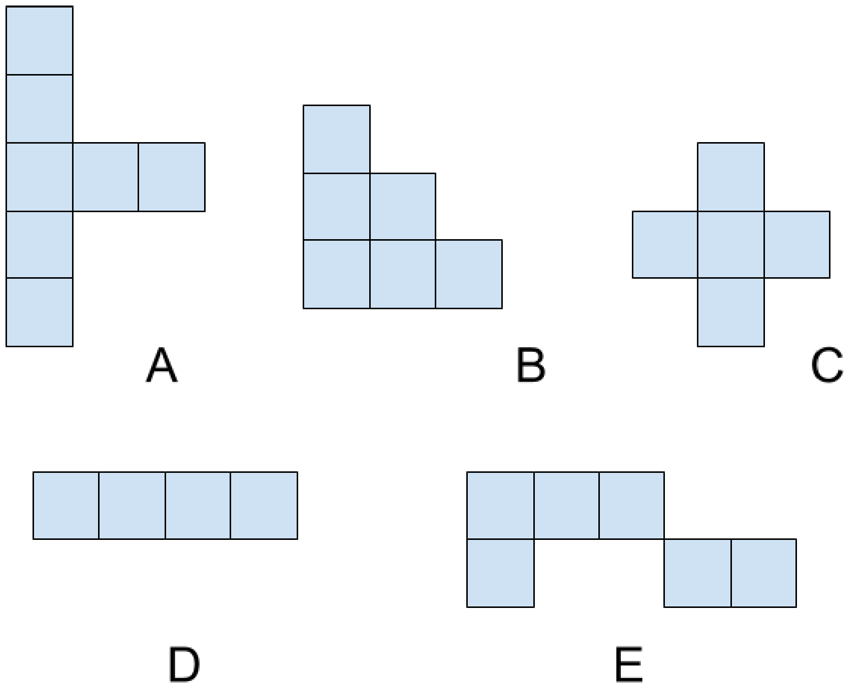
2. Schwarzberg method:

3. Reock method:

4. Convex Hull metnod:

5. Other methods:

Pick one shape and compute its compactness using each of the methods discussed.



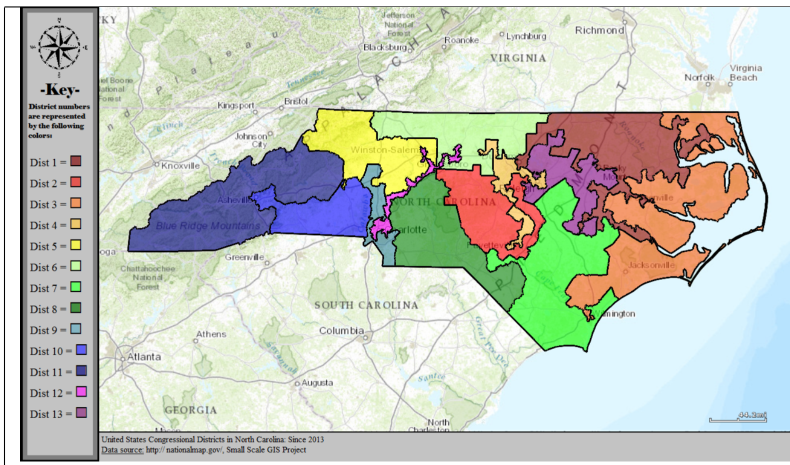
Do these measures all rank the shapes in the same order? Do they rank them in the same order as your intuition?

Does a high score or a low compactness score mean that the district is gerrymandered, or does it depend on the method?

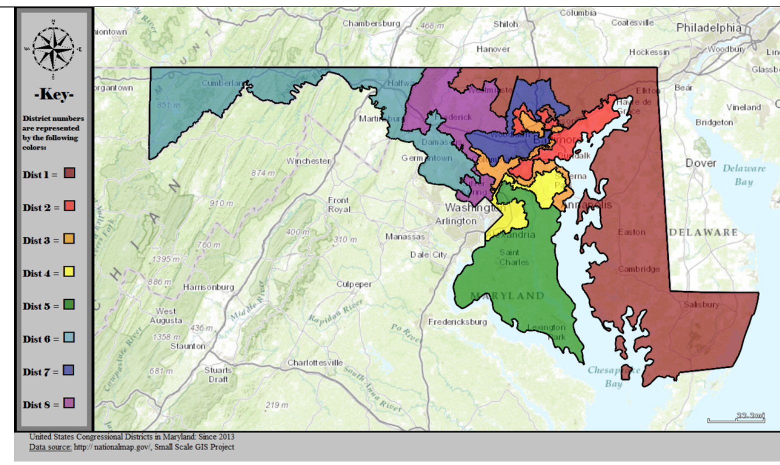
What is the maximum possible compactness score and what is the minimum?

How is the Polsby-Popper score related to the Schwarzberg score?

- Consider congressional districts for a chosen year. Each group should work with a different year.
- Census data on area and perimeter from https://www.census.gov/cgi-bin/geo/shapefiles/inde is available via a link from Sakai.
- Compute compactness for all districts in the state you chose, using at least two different measures.
- You can use the geogebra app at www.geogebra.com to estimate the area of the smallest circle that encloses a district.



North Carolina



Maryland

Question. What are advantages and disadvantages of these measures of compactness? Which of the measures do you think is the best? Why?

Question. If you were tasked to develop a mathematical criteria for detecting gerrymandering, what would you propose? Is there a compactness method and a threshold score beyond which you would argue a district is gerrymandered?

Question. Imagine that there is a court case involving the congressional districts for a certain state (e.g. NC or MD) for the current congress (116th congress, elected 2018). You are called as a subject matter expert to argue that the districts for that state are gerrymandered, or that they are not. Make your argument using data from class to support it. You can decide if you which state to choose and if you want to argue that the districts are gerrymandered or that they are not.

§12.4 Redistricting Algorithms and Redistricting with Real Data

What is meant by "an algorithm" and what are the advantages and disadvantages of an algorithmic approach to gerrymandering?

Here are a few algorithmic approaches to gerrymandering.

- Shortest split-line algorithm (see https://www.rangevoting.org/GerryExec.html)
- Brian Olson's least average distance algorithm. (see https://bdistricting.com/about.html "How does your solver work" for a brief description)
- Genetic algorithm (see http://autoredistrict.org/)
- K-means clustering (see https://www.brown.edu/news/2017-11-07/redistricting)
- B Districting
- Sweepcircle Redistricting

How do these work and what are potential advantages and disadvantages?

Try your hand at redistricting. Open Dave's Redistricting App at https://dra.indirect.cc/

Pick one challenge. Whichever challenge you pick, try to make population sizes as close to equal as possible. This means you are aiming for 733,499 people in each of 13 districts. Try to get within 1% if possible.

1. Gerrymander districts to favor Republicans
2. Gerrymander districts to favor Democrats.
3. Match the partisan breakdown of seats to the electorate. That is, if 53% of the state votes Democrat and 47% votes Republican (ignoring other parties), then 53% of the House of Reps seats should be Democrat and 47% Republican.
4. Promote highly competitive elections
5. Maximize the number of majority-minority districts
6. Draw a "fair map". In what way is it fair?
7. Draw a map in a systematic, algorithmic way, with regard only to geometry and not to distribution of Democrats and Republicans.

What are some voting reform measures that might give alternatives to districting and completely sidestep the gerrymandering question?

Divide into groups according to your favorite alternative.

- Is it better or worse than the current districting system?
- Why it is better than the other alternatives mentioned?
- What possible disadvantages does it have?

§1.1 Two Candidate Elections

Elections with two candidates are much simpler than elections with more than two candidates.

Example. Suppose we have a vote with only two alternatives, e.g. a ballot initiative on the legalization of marijuana, where you can only vote Yes or No.

What is wrong with each of the following election systems?

1. If Joe votes yes, then the initiative passes. If Joe votes no, the initiative fails.

Definition. This method is called ...

2. No matter what anyone votes, the ballot initiative fails.

Definition. This method is called ...

3. We count everyone's vote, and whichever choice gets the *least* number of votes wins.

Definition. This method is called ...

Question. What would you suggest as a better voting system for an election with two options?

Definition. A voting system is called **anonymous** if ...

Definition. A voting system is called **neutral** if. ...

Definition. A voting system is called **monotone** if ...

Example. Three friends get together for lunch every Thursday, and vote whether to to to Med Deli or Cosmic Cantina. Here is a history of some of the votes, and the outcome.

	Mo	Jo	Slo	Outcome
Aug 29	CC	MD	MD	CC
Sep 5	CC	CC	MD	MD
Sep 12	MD	MD	CC	MD

Assume they are always using some voting system (we don't know how it works), and the outcomes are not just random.

Is the voting system

1. anonymous?

2. neutral?

3. monotone?

Extra Example. Here is another group of four friends with more than two restaurant options: Med Deli, Cosmic Cantina, and Al's Burger Shop.

	Abe	Cabe	Dave	Mabe	Outcome
Aug 8	ABS	MD	ABS	CC	ABS
Aug 15	MD	CC	CC	ABS	CC
Aug 22	MD	ABS	MD	ABS	MD
Aug 29	ABS	MD	CC	CC	CC
Sep 5	CC	MD	ABS	CC	CC
Sep 12	MD	MD	ABS	ABS	ABS
Sep 19	MD	ABS	ABS	CC	CC

Is the voting system

1. anonymous?

2. neutral?

3. monotone?

Question. Which of the following voting systems – dictatorship, imposed rule, minority rule, majority rule – are anonymous? neutral? monotone?

	Anonymous	Neutral	Monotone
Dictatorship			
Imposed Rule			
Minority Rule			
Majority Rule			

Question. What voting systems for a two-option election are anonymous, neutral, and monotone?

Theorem. *May's Theorem: In a two-candidate election with an odd number of voters,*

Definition. A voting system is called a quota system if there is a number q, called a quota, such that ...

Question. Could there be a more than one winner in a quota system? More than one loser?

Question. Which of the following voting systems are quota systems?

- dictatorship
- imposed rule
- minority rule
- majority rule

Example. Suppose for an election with two candidates, Xena and Yolanda, you know the following about a voting system V.

1. If nobody votes for Xena, then V will not choose Xena as a winner.
2. If only Joe votes for Xena, then V will not choose Xena as a winner.
3. If Joe and Mary vote for Xena, then V will choose Xena as a winner.

Suppose you know that V is anonymous, neutral, and monotone. Using only this information, can you say that V is a quota system? Explain.

Theorem. *If a voting system for an election with two candidates is anonymous, neutral, and monotone, then it is a quota system.*

Proof of May's Theorem.

S2.1 Plurality

Example. In yesterday's Democratic primary in New Hampshire, candidates Sanders, Buttigieg, Klobuchar, Warren, and Biden got the following votes:

Candidate	Popular Votes
Bernie Sanders	71,759
Pete Buttigieg	68,141
Amy Klobuchar	55,164
Elizabeth Warren	25,899
Joe Biden	23,475
Steyer	10,035
Yang	7,880

1. Did any of the candidates receive a majority of the votes cast?

2. If Sanders and Buttigieg had been the only candidates in the 2016 U.S. presidential election, who do you think would have received more popular votes in New

Hampshire?

We say that Clinton received a **plurality** of the votes in the New Hampshire election. What does this mean?

Definition. A majority of the votes is ...

Definition. A plurality of the votes is ...

Consider an election with more than two candidates.
Definition. The **marjority rule** voting system is a system that elects a candidate ...

Definition. The **plurality** voting system is a system that elects a candidate ...

Is the plurality voting system

- anonymous?
- neutral?
- monotone?

Example. Twenty-one people competed for the 2016 Republican nomination for president. The total number of votes cast in 2016 Republican primaries was 34,183,841. Donald Trump received 14,015,993 of these votes.

1. Did Trump win a majority of the votes cast in the primaries?

2. If the nomination had been decided by plurailty from these 21 candidates, what is the smallest number of votes that Trump could have received and still have had a chance of winning the nomination?

3. What is the maximum number of voters who could have preferred Trump the LEAST among the 21 candidates in order for him to still have had a chance of winning the nomination?

   What message does this give about a plurality voting system?

To make the negatives of the plurality voting system more precise, we need some definitions.

Definition. A preference order (also called preference ballot, preference list) is

Note. Usually preference ballots do not allow ties. However, some voting systems do allow preference ballots with ties. We will assume that preference ballots do not allow ties unless otherwise specified.

Example. What would be a possible preference order for 2020 Democratic presidential candidates?

Example. Give a preference order for the following ice cream flavors, using the google form:

Chocolate, Vanilla, Cookies N'Cream, Mint Chocolate Chip

1. How many preference orders are possible?

2. Summarize the class data in a table that gives counts for each preference order. (This table is called a **preference schedule**.)

3. If we use the plurality voting system, which ice cream flavor will win? Which is second most popular? Third most popular?

Definition. For a voting system and an election, a **societal preference order** is ...

Question. Can a societal preference order have ties? Even if the individual preference ballots are not allowed to have ties?

4. What is the societal preference order for ice cream flavors for this class?

Definition. A **Condorcet winner** is a candidate in an election who ...

Definition. A **Condorcet loser** is a candidate in an election who ...

Example. Is there a Condorcet winner for ice cream flavors? A Condorcet loser?

Question. Was there a Condorcet winner in yesterday's New Hampshire primary?

Candidate	Popular Votes
Bernie Sanders	71,759
Pete Buttigieg	68,141
Amy Klobuchar	55,164
Elizabeth Warren	25,899
Joe Biden	23,475
Steyer	10,035
Yang	7,880

Question. For plurality voting, is there always a Condorcet winner in every election? Is there always a Condorcet loser?

Definition. A voting system is said to satisfy the **Condorcet winner criteria (CWC) if ...**

Definition. A voting system is said to satisfy the **Condorcet loser criteria (CLC) if ...**

Question. Does the plurality voting system satisfy CWC and CLC?

Summarize some of the advantages and disadvantages of plurality.

Note. There are two versions of a plurality voting system:

1. Voters only vote for one candidate
2. Voters submit ranked ballots

S2.2 Borda Count

Definition. The **Borda Count** voting system works like this for an election with n candidates:

- Each voter submits a preference ballot.

- For each ballot, points are awarded as follows:

- The winner is ...

- The societal preference order is determined by ...

NEXT TIME - need to insert example with a made up preference schedule!

Example. Going back to ice cream preferences, if we used the Borda Count system, what would the societal preference order be? Would this be the same or different from when we used plurality?

Question. Is the Borda count

A. anonymous?

B. neutral? What does this mean for a system with ranked ballots?

C. monotone? What does it mean for a voting system to be monotone when there are ranked ballots?

Example. The Borda count is used to rank college football teams. Here is a ranking that happened in 1971:

Rank	Team	Points	First-Place Votes
1	Notre Dame	885	15
2	Nebraska	870	26
3	Texas	662	5
4	Michigan	593	1
5	Southern California	525	1
6	Auburn	434	1
⋮	⋮	⋮	(all 0)
20	Northwestern	58	1
⋮	⋮	⋮	(all 0)

Do you notice anything surprising?

Definition. A voting system satisfies the **majority criteria** if ...

Question. Does the Borda count satisfy the majority criterion?

Question. Does the Borda count satisfy the Condorcet winner criteria?

Question. If a voting system satisfies the MC, does it satisfy the CWC?

Question. If a voting system satisfies the CWC, does it satisfy the MC?

Question. Which is a stronger condition? The majority criteria or the Condorcet winner criteria?

Example. Here is an example from the book: Filiz, Gerald, Helen, and Ivan are candidates for president of an organization.

Rank	12	7	5	3
1	F	G	H	I
2	G	H	I	H
3	H	I	F	G
4	I	F	G	F

A. Find the winner and societal preference order under 1) the plurality voting system, and 2) the Borda count

B. Which of these two systems best reflects the will of the voters in your opinion? Why?

Fill in the following chart:

System	Anonymous	Neutral	Monotone	Majority Criteria	CWC
Plurality					
Borda count					

Does this violate May's theorem?

Question. Is it possible for a candidate to win with the Borda count without getting any first place votes?

Summarize the pros and cons of the Borda count voting system.

S3.1 Sequential Pairwise Voting

We have seen that both plurality and Borda count share a major flaw:

Devise a voting system that does not have this flaw.

Definition. The sequential pairwise voting system works as follows:

Can you think of any voting systems that use this method?

Example. Return to the CVAAB election example from the book:

Rank	12	7	5	3
1	F	G	H	I
2	G	H	I	H
3	H	I	F	G
4	I	F	G	F

Recall that the plurality method produced the societal preference order:

and the Borda count produced the societal preference order:

What is the societal preference order that would result from a sequential pairwise voting system? Hint: it depends on the order of pairings. See if you can find orders of pairings that make different candidates win.

Suppose we pair G and H, then winner against F, then winner against I.

- What societal preference order results?

- Who would win in a head-to-head contest between G and I?

- Doest this seem strange?

Who would win in sequential pairwise for the ice cream poll?

Question. What pros and cons do you see for sequential pairwise voting? Consider anonymity, neutrality, monotonicity, CWC, majority criteria, manipulability, etc.

Question. is there a voting system that is similar to sequential pairwise voting, in that it eliminates candidates one at a time, but is not subject to the same pitfalls?

S3.2 Instant Runoff

Definition. The **instant runoff** voting system works in the following way:

Example. Find the societal preference order using instant runoff.

Rank	12	7	5	3
1	F	G	H	I
2	G	H	I	H
3	H	I	F	G
4	I	F	G	F

Example. Use instant runoff to determine a societal preference order for the ice cream poll.

Example. Use instant runoff to determine a societal preference order for this election for a math department chair:

Rank	7	6	5	3
1	A	B	C	D
2	B	A	B	C
3	C	C	A	B
4	D	D	D	A

What if the 3 voters on the far right switch their votes from $D > C > B > A$ to $A > D > C > B$?

Investigate the instant runoff voting method. What are its strengths and weaknesses?

Fill out this chart:

System	Anonymous	Neutral	Monotone	Majority Criteria	CWC
Plurality					
Borda count					
Sequential Pairwise Voting					
Instant Runoff					

What is your favorite voting system that we have studied so far? Why?

Can you devise a voting system that has ALL the desirable properties that we've discussed so far?

Independence of Irrelevant Alternatives

Example. For homework, you were asked to consider the F, G, H, I election using the plurality voting system, and to imagine that a small group of voters who were opposed to Filiz tried to manipulate the election by introducing another candidate.

Rank	12	7	5	3
1	F	G	H	I
2	G	H	I	H
3	H	I	F	G
4	I	F	G	F

Introduce a candidate Joe and write out a plausible preference schedule with Joe included, that makes Filiz lose (under pluraltiy) without making Joe win.

Joe is called a **spoiler candidate**. What does this mean?

What are some real-life elections whose outcomes may have been affected by spoiler candidates?

What property would you like for voting systems to have to avoid the possibility of spoiler candidates?

Definition. A voting system is said to have the **independence of irrelevant alternatives** criterion (IIA) if ...

Question. Investigate which voting systems satisfy IIA:

- Plurality?
- Borda count?
- Instant run-off?
- Sequential pairwise voting?
- Dictatorship?
- Imposed Rule?
- Majority Rule?

Hint: to try to break IIA, start with a preference schedule, such as the one below,

Rank	12	7	5	3
1	F	G	H	I
2	G	H	I	H
3	H	I	F	G
4	I	F	G	F

and EITHER:

(a) use the worst candidate to mess up the winning candidate, OR

(b) introduce a new candidate who is everyone's last choice, and then put that candidate where they need to go to mess up the winning candidate without winning themselves

Hint: for sequential pairwise: try the agenda of orders: F, I , G, H and change the preference ballots between two losing candidates in a way that derails the winning candidate.

Relationship between existence of spoiler candidates and IIA

Claim: Consider a voting system in which introducing a candidate who is everyone's last place choice has no effect on the societal preference order for the other candidates. Then: If there is an election with a spoiler candidate J, then the voting system cannot satisfy IIA.

Proof:

Question. Does there exist a voting system that is anonymous, neutral, monotone, and also satisfies IIA?

Arrow’s Theorem

Kenneth Arrow, 1951:

> I started out with some examples. I had already discovered that these led to some problems. The next thing that was reasonable was to write down a condition that I could outlaw. Then I constructed another example, and something else didn’t seem very right about it. Then I had to postulate that we have some other property. I found I was having difficulty satisfying all of these properties that I though were desireable ... After having formulated three or four conditions of this kind, I kept on experimenting. And lo and behold, no matter what I did, there was nothing that would satisfy these axioms.
>
> So after a few days of this, I began to get the idea that maybe ... there was no voting method that would satisfy al of the conditions that I regarded as rational and reasonable. It was at this point that I set out to prove it. And it actually turned out to be a matter of only a few days work.

Theorem. *Arrow's Theorem, loosely stated:*

To state Arrow's Theorem precisely, we need to specify ...

Voting Systems

Definition. A *ranked voting system* is a rule (or function) that assigns

The individual preference orders and the societal preference order are required to be *transitive*. This means:

Why is transitivity a reasonable requirement for a preference order? Can you think of a situation in which preference orders might not be transitive?

Unlike the book, we will assume unless otherwise specified that individual preference orders ARE NOT allowed to have ties. Like the book, we will assume that societal preference orders ARE allowed to have ties, as in $A \succ B \approx C \succ D$ or $A \approx B \approx C \succ D$

Arrow's Conditions for a Reasonable Ranked.Voting System

Condition 1: Universality

Condition 2: Positive Association of Social and Individual Values

Condition 3: Independence of Irrelevant Alternatives

Condition 4: Citizen Sovereignty

Condition 5: Non-dictatorship

Question. Which of Arrow's five conditions is most similar to the property of:

- Anonymity?
- Neutrality?
- Monotonicity?
- Which is new for us?

Theorem. *Arrow's Theorem: For an election with more than two candidates, ...*

Alternative versions of Arrow's Theorem

Theorem. *Arrow's Theorem (Weak Form): For an election with more than two candidates ...*

Theorem. *Arrow's Theorem (Strong Form): For an election with more than two candidates, it is impossible for a ranked voting system to satisfy all of the following:*

Condition 1: Universality

Condition 3: IIA

Condition 5: Not a dictatorship

and

Condition 6: Unanimity

What is unanimity?
Definition. Candidate A is *unanimously preferred* over candidate B in an an election if ...

Definition. A ranked voting system satisfies *unanimity* (also called the *Pareto Condition*) if, whenever one candidate is unanimously preferred over another in an election (say A unanimously preferred over B), then the voting system ...

Example. For each of the following elections, decide if the voting system satisfies unanimity. You can answer yes, no, or can't tell.

1.

Rank	7	6	3
1	A	B	C
2	B	A	D
3	C	D	B
4	D	C	A

Societal preference order: $D \succ C \succ B \succ A$.

2.

Rank	7	6	3
1	A	B	A
2	B	A	D
3	C	D	B
4	D	C	C

Societal preference order: $A \succ D \succ C \succ B$.

3.

Rank	7	6	3
1	A	B	A
2	C	A	D
3	B	D	B
4	D	C	C

Societal preference order: $A \succ D \succ C \succ B$.

Example. Which of the voting systems that we have encountered satisfy unanimity? Plurailty? Borda count? Instant runoff? Sequential pairwise voting? Dictatorship? Imposed rule?

Hint: for sequential pairwise, consider the preference schedule

Rank	1	1	1
1	A	B	C
2	B	C	D
3	C	D	A
4	D	A	B

Questions for thought:

- How can we prove that something is impossible?
- Are Arrow's conditions as reasonable as they seem at first glance?
- In no ranked voting systems are perfect, which one are best?
- Which of the ″fairness″ conditions would you be most willing to give up?
- Are their alternatives to ranked voting systems that would be more fair for elections involving more than two candidates?

Proving Arrow's Theorem

Last time we gave three versions of Arrow's Theorem:

Theorem. *Arrow's Theorem: For an election with more than two candidates, it is impossible for a ranked voting system to satisfy all of the following five conditions:*

Theorem. *Arrow's Theorem, Strong Form: For an election with more than two candidates, it is impossible for a ranked voting system to satisfy all of the following four conditions:*

Theorem. *Arrow's Theorem, Weak Form: For an election with more than two candidates, it is impossible for a ranked voting system to satisfy all of the following five conditions:*

How are these three forms related?

We will prove the Strong Form.

Then we will show that if the Strong Form is true, the regular theorem must be true.

Then we will show that if the regular theorem must be true, the weak version must be true.

Strategy for proving the strong version:

- We will assume we have a ranked voting system that satisfies the first three conditions: universality, IIA, and unanimity.
- We will show that this system must be a dictatorship.

Lemma. *Assume for an election with more than 2 candidates, a ranked voting system V satisfies universality, IIA, and unanimity. Suppose that B is some candidate in the election, and that every voter ranks B in either first place or last place (without ties) in their individual preference order. Then the societal preference order produced by V must ...*

Proof:

1. Suppose that B is not ranked either first or last in the societal preference order. Then there must be some candidate A with $A \geq B$ and some candidate C with $B \geq C$ in the societal preference order. Why?

2. It must be true that $A \geq C$ in the societal preference order. Why?

3. Now move C just above A in everyone's personal ballot. Will this affect any individual preference rankings between A and B or between B and C?

4. Will this affect the societal preference order between A and B? What about between B and C?

5. Based on the societal preference order between A and B and between B and C, what should the societal preference order between A and C be? How do you know?

6. What does unanimity tell us about the societal preference order between A and C?

7. How does this prove the Lemma?

Proof of Arrow's Theorem, Strong Form

We assume that we have a ranked voting system that satisfies universality, IIA, and unanimity.

We will look for a dictator. To do this, we'll name all the voters: $v_1, v_2, v_3, \cdots v_n$, where n is the number of voters.

1. Start with a particular preference schedule in which all the voters rank candidate B last (with no ties) in their individual preference orders. What will be the position of B in the societal preference order? What property allows you to conclude this?

2. Suppose ALL the voters move B from last place to first place (with no ties) in their individual preference orders. What will be the position B in the societal preference order now? How do you know?

3. Suppose that only SOME of the voters move B from last to first place in their individual preference orders? Where can B end up in the society preference order?

4. Starting with B last for all, suppose that, one by one, starting with v_1 and proceeding in order, each voter moves B from last place to first place in their individual preference order. Explain why there must be some voter, say v_j, for which this move first causes B to flip from last to first in the societal preference order?

5. In fact, for ANY preference schedule in which $v_1, v_2, \cdots, v_{j-1}$ rank B first and $v_j, v_{j+1}, v_{j+2}, \cdots, v_n$ rank B last, (not just the particular one we created) the voting system will rank B last. Why?

6. Also, for ANY preference schedule in which $v_1, v_2, \cdots, v_{j-1}, v_j$ rank B first and $v_{j+1}, v_{j+2}, \cdots, v_n$ rank B last, the voting system will rank B first. Why?

7. We want to show that v_j is a dictator.
   - First, we'll show that for any two candidates that do not include B, say A and C, if v_j prefers A over C, then the societal preference order ranks A over C.
   - Then we'll show the same thing for two candidates that do include B.

   How will this show that v_j is a dictator?

8. To show that v_j controls the societal preference order between any two candidates that are not B, say A and C, start with a preference schedule S in which v_j prefers A over C. We don't know anything about any other preferences. Our eventual goal is to show that the society preference order ranks A over C.

   Which of the following changes to S would change the societal preference between A and C? Explain your answer.

   (a) Voter v_j moves B between A and C in their individual preference order.

   (b) Each of the voters $v_1, v_2, \cdots, v_{j-1}$ (all of the voters listed before v_j) move B to first place in their individual preference orders.

(c) Each of the voters $v_{j+1}, v_{j+2}, \cdots v_n$ (all of the voters listed after v_j) move B to the last place in their individual preference order.

9. Let S' be the societal preference order in which all three of these things in the previous step have been done.

   (a) Explain why, in S', the societal preference order will rank B above C. Hint: will the societal ranking between B and C change if we move B up to the top in v_j's preference ballot?

   (b) Let S' be the societal preference order in which all three of these things have been done, as in the previous step. Explain why, in S', the societal preference order will rank A above B.

(c) Let S' be the societal preference order in which all three of these things have been done, as in the previous step. Explain why, in S', the societal preference order will rank A above C.

(d) Explain why, in the original preference schedule S, the societal preference order will rank A over C.

10. Is it now possible to conclude that v_j controls the societal preference order for any pair of candidates that does not include B? Why or why not?

11. To show that v_j also controls the societal preference order between B and any other candidate, say D, we have to be a little sneaky. First, let's repeat the above argument with a third candidate (say E), different from both B and D, in the place of B. Why can we use this to show that there is some voter v_i who controls the societal preference order between B and D?

12. We want to show $i = j$, that is, the two voters who control the societal preference order are the same person.

    (a) Suppose $i < j$. Find preference schedule that makes B lose even though v_i prefers B to D. Why is this a contradiction?

    (b) Suppose $i > j$. Find a preference schedule that makes B win even though v_i prefers D to B. Why is this a contradiction?

    (c) What do the above two steps tell you about v_i and v_j?

13. True or False: There is a single voter $v_\star$ that controls the societal preference order between any two candidates.

14. Explain why Arrow's Theorem, Strong Form, is proven.

To prove the regular version of Arrow's Theorem, we need a lemma:

Lemma. *If a ranked voting system satisfies universality, monotonicity, IIA, and citizen sovereignty, then it also satisfies ...*

Assume that we have a ranked voting system that satisfies universality, monotonicity, IIA, and citizen sovereignty. We need to prove that it satisfies ...

Proof of the Lemma:

1. Start with any preference schedule S in which every voter prefers A over B. We need to show ...

2. Now consider some other preference schedule S' in which $A \succ B$ in the societal preference order. How do we know such a preference schedule exists?

3. Now alter S' by moving A above B for every voter. Call this preference schedule S''. Do we know anything about the societal preference order for S'' between A and B? Why or why not?

4. How do S and S'' differ as far as individual preference orders between A and B?

5. What can we conclude about the societal preference order between A and B in S? What property allows us to conclude this?

6. How does this prove the Lemma?

Question. How does this prove the regular form of Arrow's Theorem?

To prove the weak form of Arrow's Theorem, we need the following Lemma:

Lemma. *If a ranked voting system satisfies anonymity, neutrality, monotonicity, IIA, and universality, then it must satisfy non-dictatorship, citizen sovereignity, monotonicity, IIA, and universality.*

Proof of Lemma:

1. Explain why a ranked voting system that satisfies anonymity must satisfy non-dictatorship.

2. Explain why a ranked voting system that satisfies neutrality must satisfy citizen sovereignity.

3. Why does this prove the lemma?

Explain how the regular version of Arrow's Theorem, plus this Lemma, proves the weak version of Arrow's Theorem.

Beyond Arrow Impossibility

Question. Since it is not possible to have all desirable properties, which are you willing to let go?

Intensity of Binary Independence

One option for sneaking past Arrow's Impossibility is to weaken IIA.

Maybe we don't want IIA:

- If some voters who used to rank A first and C last move A and C closer together so that they are nearly adjacent in rankings (but still prefer A to C) , maybe this change SHOULD catapult C to the lead.

Intensity of Binary Independence is a weaker condition than IIA, that lets a voting system take into account the intensity of voters preferences between candidates.

Definition. The **intensity** of a voter's preference between two candidates is ...

Example. For each of the following ice cream preference ballots, what is the intensity of the preference between chocolate and vanilla?

(a) Chocolate $\succ$ Mint Chococate Chip $\succ$ Strawberry $\succ$ Vanilla

(b) Mint Chocolate Chip $\succ$ Chocolate $\succ$ Vanilla $\succ$ Strawberry

(c) Chocolate $\succ$ Mint Chocolate Chip $\succ$ Vanilla $\succ$ Strawberry

Definition. A voting system is said to satisfy **Intensity of Binary Independence (IBI)** if ...

Example. NEXT TIME, GIVE AN EXAMPLE OR TWO OF A PAIR OF ELECTIONS AND ASK IF THE VOTING SYSTEM SATISFIES IBI (the start of an example is below but it has too many candidates and no SPO is given)

Rank	10	5
1	A	
2	C	
3	D	
4	B	
5	E	

Rank	10	5
1	C	
2	A	
3	D	
4	E	
5	B	

Does the intensity of preference change between A and B? What about between A and E? Does this voting system satisfy IBI?

Question. Do any voting systems that we are familiar with satisfy IBI? Hint: consider the example below.

Example. Suppose there is an election between Greg, Sharon, Dean, and Carolyn, but the preference ballots have been partial destroyed. All the info that you have is the following:

- Two ballots contain the partial ordering $G \succ S$.
- One ballot contains the partial ordering $S \succ C \succ G$.
- The remaining one ballot contains the partial ordering $G \succ D \succ C \succ S$.

You know these are the only ballots.

Is this enough information to determine the societal ranking of G and S under: purality? Borda count? instant runoff? sequential pairwise?

Question. True or False: If a ranked voting system satisfies IBI, then it satisfies IIA.

Question. True or False: If a ranked voting system satisfies IIA, then it satisfies IBI.

Question. Is there a ranked voting system that satisfies universality, IBI, unanimity, and is not a dictatorship?

Modify Unanimity

Unanimity (Pareto's condition) is a demanding condition. Even plurality and instant run-off don't have it, due only to ties.

Example. Give an example in which plurality fails unanimity. Also one for instant runoff.

Definition. Modified Pareto Condition If there is a pair of candidates, say A and B, such that every voter in the election prefers A over B, then ...

Question. Does plurality satisfy the Modified Pareto Condition? What about Borda Count? Instant runoff? Sequential pairwise? Explain your answers.

Question. True or False: If a ranked voting system satisfies Modified Pareto, then it satisfies unanimity (Pareto).

Question. True or False: If a ranked voting system satisfies unanimity (Pareto), then it satisfies Modified Pareto.

Question. Are there ranked voting systems that satisfy Arrow's conditions with the Modified Pareto condition instead of unanimity?

Theorem. *Corollary to Wilson's Theorem* A ranked voting system that satisfies universality, IIA, and the modified Pareto system must be either ....

Approval Voting

Question. How else might we design a voting system that has all the desirable properties that can sneak by Arrow Impossibility?

Definition. Approval Voting is a voting system in which ...

Example. Figure out who would win the vegetable vote if just our class voted, using approval voting.

Example. Where is approval voting used in the real world?

Question. Where should it be used?

Question. Is approval voting a ranked voting system?

Question. Does approval voting satisfy

- anonymity?
- neutrality?
- monotonicity?
- universality?
- unanimity (Pareto's condition)?
- IIA?
- citizen sovereinty?

Definition. A cardinal voting system is

Question. Is approval voting a cardinal voting system?

Question. What are other examples of cardinal voting systems?

S9.1 Electoral College

The U.S. presidential election is decided by a system called the Electoral College. How does it work?

1. Each state gets a certain number of electors.

2. When the people of a state vote for president, they are actually voting for electors.

3. The electors then cast their votes for president, and the winner is decided by majority rule among the elector votes.

Question. How many electors does each state get?

Question. How many electors are their total?

Question. How are the electors chosen?

Question. Do the electors have to vote for the candidate of their party?

Question. What happens if majority rule among the electors is a tie? Has this ever happened?

Question. What is the history of the electoral college system, and why was it created?

Question. Which of these properties does the electoral college system satisfy (from the perspective of the individual people being the voters, not the electors)? Ignore faithless electors.

- anonymity

- neutralty

- monotonicity

Question. Do you expect that the winner under the electoral college is usually the same as the winner would have been under a different voting system?

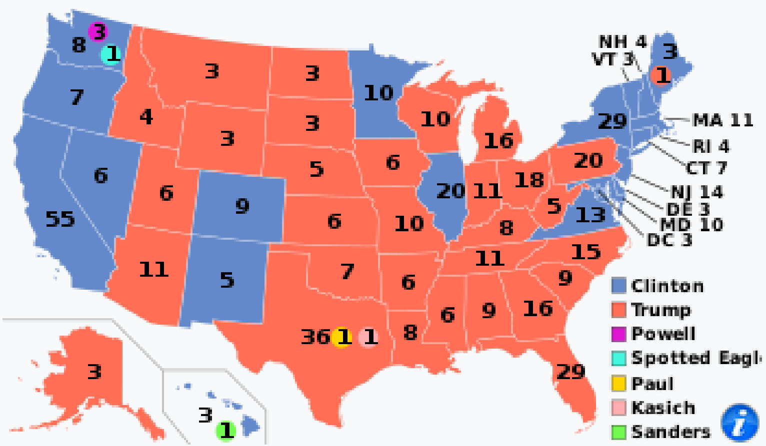
Example. In the 2016 presidential election, votes were cast as follows:

Candidate	Popular Votes
Hillary Clinton	65,853,516
Donald Trump	62,984,825
Gary Johnson	4,489,221
Jill Stein	1,457,216
Others	1,884,459

Question. Who do you think would have won the election under plurality? Instant runoff? Borda count? Approval voting?

Since ranked ballots weren't used, in order to apply a ranked voting method lie instant runoff or Borda count, you will have to make assumptions about how voters would have voted if ranked ballots had been used. You will also have to make guesses about how approval ballots would have been filled out.

Nominee	Donald Trump	Hillary Clinton
Party	Republican	Democratic
Home state	New York	New York
Running mate	Mike Pence	Tim Kaine
Electoral vote	304[a]	227[a]
States carried	30 + ME-02	20 + DC
Popular vote	62,984,828	65,853,514
Percentage	46.09%	48.18%



Presidential election results map. Red denotes states won by Trump and blue denotes those won by Clinton. Numbers indicate the number of electoral votes allocated to each state plus the District of Columbia, and show for whom they were cast. Altogether, Trump garnered 304 electoral votes and Clinton 227, as seven faithless electors, two pledged to Trump and five pledged to Clinton, voted for other persons.

Question. In which states do you think people have more power in the electoral college system? Bigger states? Smaller states? Other states, not necessarily bigger or smaller? Why?

Question. What are some advantages and disadvantages of the electoral college system?

Question. What are some alternatives to the electoral college system?

S11.1 Apportionment

In 1774, Congress had to decide how to apportion 105 representatives among 15 states. Here are the populations of the states, based on the 1790 Census.

State	Population
Connecticut	237,655
Delaware	59,096
Georgia	82,548
Kentucky	73,677
Maryland	319,728
Massachisetts	475,199
New Hampshire	141,899
New Jersey	184,139
New York	340,241
North Carolina	395,005
Pennsylvania	433,611
Rhode Island	69,112
South Carolina	249,073
Vermont	85,341
Virginia	747,550
TOTAL	3,893,874

How would you allocate them?

Actually Congress used slightly different population numbers due to fractional counting of slaves and native Americans

Definition. The **standard quota** is ...

Why can't we use the standard quota directly to allocate representatives?

Hamilton's method

Hamilton proposed the following method:

1. Find the standard quota for each state.

2. Give each state a number of seats equal to its standard quota rounded down.

3. See how many seats are left to be allocated (called surplus seats) and give them to ....

Use Hamilton's method to allocate seats for 1794.

What criticisms could you launch against Hamilton's method?

Hint 1: look at the standard quotas for Delaware and Maryland and look at the percent of the population corresponding to the decimal parts of the standard quotas.

Hint 2: look at the number of residents represented by each representative in each state. Which states get treated worst and which best by this measure?

Can you make an argument for Virginia getting the extra seat instead of Maryland?

Hamilton's method was approved by congress but vetoed by President George Washington! (Why?)

Jefferson's method
Definition. The **standard divisor** is ...

How can the standard divisor be used to get the standard quotients?

Jefferson's method is based on a **modified divisor** .

1. Find the standard quota for each state.

2. Give each state a number of seats equal to its standard quota rounded down. If these seats total to the number of seats to be apportioned, we are done. Otherwise, continue.

3. Use a modified divisor that is different from the standard divisor, and use this modified divisor to get each stat's modified quotas.

4. Round these modified quotas down. If these seats total to the number of seats to be apportioned, we are done. Otherwise, repeat the process with a different modified divisor.

Use Jefferson's method to allocate seats for 1794.

Question. Which states get treated worst and best under Jefferson's method?

Question. How does the outcome of Hamilton's method compare to the outcome of Jefferson's method in 1794? (What state loses a seat to which other state?)

Question. Does Jefferson's method tend to favor smaller states, larger states, or neither? Why?

Jefferson's method was actually used on 1794 (but with slightly different results due to slightly different population counts).

Example. The census of 1820 recorded a population of 1,368,775 for the state of New York and 8,969,878 for the entire U.S. The number of seats in the House of Representatives to be apportioned in 1822 was 213.

(a) What is the standard divisor and what is New York's standard quota?

(b) In the apportionment of 1822, Jefferson's method was used with a modified divisor of 39,900. Find New York's modified quota and the final number of seats the state was given. Does this seem fair?

Note. In the very next apportionment, in 1832, Jefferson's method gave New York 40 seats even though its standard quota was 38.59.

Definition. We say that Jefferson's method violated the **quota rule** in 1822 and 1832. The quota rule says that in an apportionment, each state should be given the number of seats that is equal to ...

After 1832, two new methods were proposed:

John Quincy Adam's method

Adam's method is identical to Jefferson's method except that ...

Try Adam's method on the 1790 data.

Question. Which states do better in the 1790 data and which do worse?

Question. Is Adam's method likely to favor small states or large states?

Daniel Webster's method

Webster's method is identical to Jefferson's method except that ...

Try Webster's method on the 1790 data.
Question. Which states do better in the 1790 data and which do worse?

Question. Is Webster's method likely to favor small states or large states?

Note. Webster's method was used in 1842.

Definition. Jefferson's method, Adam's method, and Webster's method are all called **divisor methods** because they work by ...

Note. It can be shown that any divisor method can violate quota (but violations using Webster's method are rare).

In 1852, Congress passed a law adopting (guess which method) ...

Population Paradoxes

Incident #1 - The Year 1882: The US Census Bureau supplied Congress with a table showing the apportionments under Hamilton's method for various sizes for the House of Representatives, from 275 - 350 seats. I have replicated this analysis using Census data from the 1880 Census, from wikipedia.

Do you notice anything unusual when the number of seats grows from 299 to 300?

This is called the ... paradox. It is very common with Hamilton's method.

In the end, Congress chose a House of Representatives size of 325.

incident # 2: The Year 1902: The US Census Bureau supplied Congress with a table showing the apportionments under Hamilton's method for all sizes of the House between 350 and 400.

A bill was presented in Congress to use Hamilton's method with 357 seats. Which state(s) do you think objected to this proposal?

In the end, Webster's method with 386 seats was used.

Incident # 3: The Year 1902 compared to 1892:

Compare the population growth of South Carolina to the population growth of New Hampshire from 1890 and 1900.

Compare the population growth of Washington to New Jersey.

For a House size of 357, what would happen to the number of seats for these four states under Hamilton's method?

In fact, under Hamilton's method, it is possible for one state to gain a seat and a second state to lose a seat even though the second state has a higher percent growth.

This is called the ... paradox.

Incident # 4: Oklahoma joined the Union: In 1907, Oklahoma joined the union as the 46th state.

Since it was not yet time for reapportionment, Congress opted to increase the size of the House from 386 to 391 and give Oklahoma 5 seats. .

What would have happened instead if Hamilton's method had been used to apportion the 391 seats to the 46 states in 1907? Can you find any states that might have objected?

This is called the ... paradox.

... in 1912 Webster's method was used again, but a new method by Joseph Hill was proposed.

... in 1922 no approtionment was done

... in 1932 Hill's method was signed into law and has been used since

Hill's method: Hill's method is just like Webster's method, except for the location of the cut-offs used for rounding up vs. rounding down.

Instead of using an average (also called arithmetic mean), Hill's method uses the geometric mean between two consecutive integers to decide whether to round up or down.

Definition. The **arithmetic mean** between two numbers x and y is the number a such that ...

Definition. The **geometric mean** between two numbers x and y is the number g such that ...

Equivalently, $g =$

Example. Would you round 5.482 up or down using Hill's method? What about 15.482?

Use Hill's method to divide up the 105 seats in 1794. Which states are treated worst and which are treated best?

Pros and cons:

- Hill's method and Webster's methods don't produce the paradoxes we saw with Hamilton's method, but they can violate quota.
- Hamilton's method will never violate quota. Why?

Question. Can we find an apportionment method that doesn't violate quota AND doesn't succumb to the three paradoxes?

Theorem. *Ballinski and Young's Theorem:*

Try these various methods for 2010 apportionment and compare results.

S13.1 Benford's Law and Election Fraud

Warm-up: Distributions

1. Everyone enter your birth month, height in inches (as in 65 inches), and the number of siblings you have on the google form birthHeightSiblings.
2. By hand, make a histogram (bar chart) for each of these variables, in which the x-axis give the values and the y axis give the number of people with those values.
3. Now use R to make bar charts.
   (a) Go to rstudio.cloud and make an account if needed
   (b) Click on New Project. You can name it if you like.
   (c) On the bottom right, click "upload" and upload a csv file with all the birth-height-sibling data in it.
   (d) Click on File > New > R Script.
   (e) Copy this code: myData = read.csv("myFile.csv", header = TRUE)
   (f) Change the "myFile.csv" to the name of the birth-height-sibling data file that you uploaded. What do you think "header = TRUE" means?
   (g) Click on this line of code to put the cursor there, then click on the "Run" button towards the top right. If you get red error messages below, something went

wrong, otherwise, your data file should be uploaded, and you can click on its name "myData" at the top left to see it.

(h) Click back to your script, and type: hist(myData$BirthMonth). If your header for the birth month was called something else, you will need to change "Birth-Month" to the correct header.

(i) Click on this line of code and click on "Run" to run it. You should see a histogram.

(j) You can modify your histogram, for example, by using parameters like hist(myData$Birth xlab= "Birth Month", breaks = c(0.5:12.5)). This specifies the label on the x axis and the breaks between consecutive bins. c(0.5:11.5) put the breaks at 0.5, 1.5, 2.5, etc., so that the bins will be centered at the integers 1, 2, 3, etc.

(k) Try making histograms for the other two variables, height and siblings.

4. What do you notice about the shapes of the distributions for birth month, height, and siblings?

Go to the wikipedia List of municipalities in North Carolina, down past the list of the top 50 cities and towns, and pick some cities and towns "at random" and record their populations on the spreadsheet.

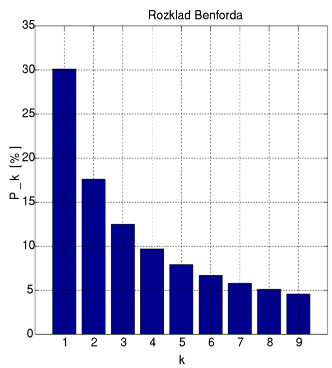
- What would you expect the distribution of populations to look like?
- Upload the file and plot a histogram using R, or use the NCCountyPopulations.csv file on Sakai instead. This file gives the populations of all 100 NC counties.
- Does the distribution look like what you would expect?
- What if you just took the first digit of each number in your file. What would you expect the distribution of these digits to look like?
- You can try it in R by using the substr function to grab the first digit. First, experiment with the following lines to see how this substr function works

  ```
  substr("A long time ago, in a galaxy far far away...", 21, 28)
  substr(387, 1, 2)
  as.numeric(substr(387, 1, 2))
  ```

- Now use as.numeric and substr to grab the first digits of all your population numbers.
- Now plot these first digits. What is the distribution like?

Benford's Law In many naturally occurring data sets, the first digit is more likely to be small than large.

- the number 1 appears as the first digit about ...
- the number 2 appears as the first digit about ...
- the number 9 appears as the first digit about



You can get the exact predicted frequencies of digits using this R code:

```
benfords_p <- data.frame(first_digit = 1:9,
                         ben_prop = log10(1 + 1/1:9))
```

According to Wikipedia, in 1881, "The discovery of Benford's law goes back to 1881, when the American astronomer Simon Newcomb noticed that in logarithm tables the earlier pages (that started with 1) were much more worn than the other pages."

Question. What sorts of data sets does Benford's Law apply to?

Question. How can Benford's Law be used to detect fraud?

On June 12 2009, the Republic of Iran held an election where President Mahmoud Ahmadinejad sought re-election against three challengers. When it was announced that Ahmadinejad had won with 62% of the vote, there were widespread allegations of election fraud.

The data file Iran2009PresidentialElection on Sakai gives vote totals for each of the four candidates: (Mahmoud Ahmadinejad, Mohsen Rezaee, Mehdi Karroubi, and Mir-Hossein Mousavi), as well as total votes, by province.

If there was election fraud, which columns of numbers might you expect to reflect these irregularities the most?

Using the data from http://irandataportal.syr.edu/2009-presidential-election that is posted on Sakai, see if the numbers of votes cast for the winner, Mahmoud Ahmadinejad, conform to Benford's Law. Be sure to feed your histogram command appropriate breaks: breaks = c(0.5:9.5), so that it'll center the bins around each integer and not lump some integers together.

Question. Do the first digits conform to Benford's Law? If not, what are some reasons why they might not?

We can try the same thing on other election results:

The OpenElections project is obtaining and standardizing precinct-level results from the 2016 US presidential election, among other US elections. To access the data, search OpenElections' GitHub page and pick an election that is marked "CSV" or "baked raw".

Pick a state and use this data to create a plot of that state's first digit distribution by precinct. Use the number of votes cast for Hillary Clinton in each precinct. Make sure you remove rows for "total" counts if your state has them.

Does the election you chose appear to fit the distribution better or worse than the Iran election?

Election Fingerprints

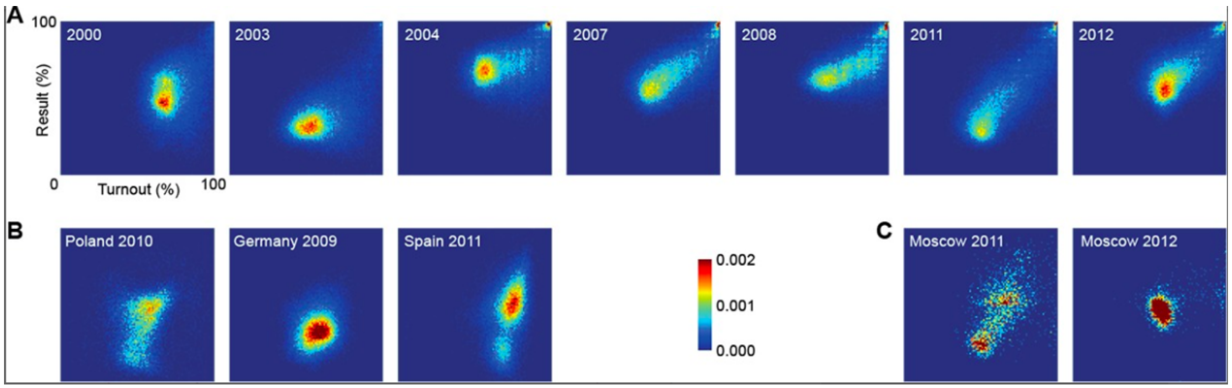
Suppose someone is committing election fraud by ”ballot stuffing”. That is ..

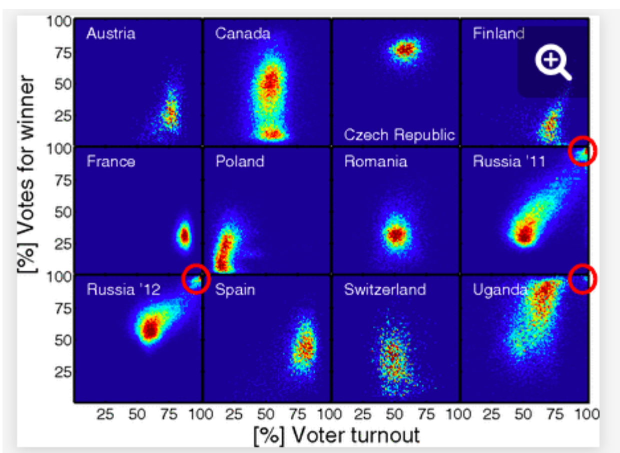
What effect would you expect this to have on reported voter turnout in the affected voting precincts?

What effect would you expect this to have on the percentage of the vote for the winning candidate in the affected voting precincts?

In these plots, % voter turnout is plotted on the x-axis and % vote for the winning candidate or party is plotted on the y-axis. The frequency (probability density) is shown by the color – red means lots of precincts had these vaues, and blue means few or none.

Which of these plots have "election fingerprints" that are suspicious for ballot stuffing?





What's going on with the two patches in Canada's election?

S6.1 Strategic Voting

Example. Consider this fictional election among three candidates, G from the Green Party, D from the Democratic Party, and R from the Republican Party.

Rank	25	15	15	45
1	G	D	D	R
2	D	G	R	D
3	R	R	G	G

(a) If instant run-off is used to decide the election, what will the resulting societal preference order be?

(b) Suppose that five of the $D \succ G \succ R$ voters decide that they actually like G best and change their preferences accordingly, to $G \succ D \succ R$. Would this change result in a better or worse outcome for these voters?

(c) Suppose that six of R's 45 voters change their preferences to $G \succ R \succ D$. How would this change impact the outcome of the election?

Definition. We say that a voter votes *sincerely* if ...

and insincerely otherwise.

Definition. *Strategic voting* means ...

Definition. We say that a voting system is *manipulable* if (informally) ...

Note. Here are some tactics of strategic voting. What do you think they mean?

1. Compromising.

2. Pushover or mischief voting.

3. Burying.

4. Bullet voting.

Example. Suppose that a form of approval voting was used in the G, D, R example above, and the ballots asked voters to pick their top two candidates. Show how bullet voting could be a smart strategic choice for some of the voters.

Rank	25	15	15	45
1	G	D	D	R
2	D	G	R	D
3	R	R	G	G

Example. Of the main voting systems we have considered: plurality rule, Borda count, sequential pairwise voting, instant runoff, and approval voting, which are most susceptible to::

(a) compromising?

(b) pushover voting?

(c) burying?

Question. Are there any voting systems that are immune to manipulation?

Precise definitions

Definition. A *choice function* is a function that receives as input ...

and gives as output ...

Definition. We will distiguish *ordinal* choice functions and *cardinal* choice functions as follows:

Definition. A choice function is *manipulable* if ...

Definition. For a given choice function, a *dictator* is a voter who can ...

Definition. A choice function is said to be *dictatorial* if ...

Definition. A choice function is *non-imposed* if ...

Definition. A choice function is *monotone* if the only way a single voter can cause the winner to change from A to B is for that voter to ...

Question. How is this definition of monotone similar and how is it different from our previous definition of monotone?

Note. If an ordinal choice function is not monotone, then it must be manipulable. Why?

Lemma: Every non-manipulable ordinal choice function is monotone.

Theorem. *(Gibbard-Satterthwaite Theorem) There does not exist an ordinal choice function for an election with more that two candidates that is ...*

The idea of the proof is to ...

.

Theorem. *(Gibbard-Satterthwaite Theorem) There does not exist a cardinal choice function for an election with more that two candidates that is ...*

The idea of the proof is ...

Choice Functions vs. Ranked Voting Systems?
Question. How are choice functions related to ranked voting systems?

Question. If you have a ranked voting system, how can you use it to create a choice function?

Question. If you have a choice function, how can you use it to created a ranked voting system?

The Gibbard-Satterthwaite Algorithm: To turn a choice function f into a ranked voting system F.

1. *To decide the societal ranking between any two pairs of candidates A and B ...*

2. *To get the entire societal preference ranking ...*

Question. Is there anything that you are worried about with this algorithm?

Example. Suppose you have a choice function f that takes the plurality winner of everyone's first place choices. If we use the Gibber-Satterthwaite algorithm to create a ranked voting system F, what is the societal ranking that results from the election below?

	10	6	5
1	A	B	C
2	B	C	B
3	C	A	A

Is this the same ranking that we would get using the usual plurality ranked voting system?

Extra Example. Assume that the voters in an election are arranged in some order, say $v_1, v_2, \cdots, v_n$ and let f be the choice function that works as follows:

- If v_1 has a unique top-ranked candidate, then that candidate is declared the winner. Otherwise, the candidates who are not among v_1's top choices are eliminated.
- If v_2 has a unique top-ranked candidate among those who remain, then that candidate is declared the winner. Otherwise, the remaining candidates who are not among v_2's top choices are eliminated.
- The process continues in this way, with each successive voter eliminating candidates until only one remains If multiple candidates remain after the last voter, the tie is broken by alphabetical order of the candidate's name.

Supposer there are three voters in the election with the following preference orders.

- $v_1 : A \approx B \succ C \approx C$
- $v_2 : B \succ C \approx D \succ A$
- $v_3 : D \succ C \succ A \succ B$

Given these ballots, what societal preference order would the ranked voting system produce?

Details to make sure that the algorithm produces a ranked voting system

Need to show that:

1. If A and B are moved to the top of everyone's preference ballot, then ...

2. If we get $A \succ B$ and $B \succ C$ in the societal preference order, then we get ...

Lemma. *If f is a non-imposed, non-manipulable choice function, and there is a set S of candidates at the top of every voter's preference order, then f must. ...*

Proof of the Lemma:

First, let's prove the Lemma when S has exactly two candidates in it: A and B.

Assume we have a non-imposed, non-manipulable choice function and every voter prefers A and B to any other candidate.

Call the ballots $b_1, b_2, b_3, \cdots, b_n$

1. Explain why there must be some other set of ballots, say $b_1^*, b_2^*, b_3^*, \cdots b_n^*$ such that with those ballots, f would choose A as the winner.

2. What would happen if we gave f the ballots $b_1, b_2^*, b_3^*, \cdots b_n^*$? Could f choose a candidate other than A or B as a winner? Hint: Only one voter's ballot is different from those in part (1). What would monotonicity require in this situation for the winner to change from A to some other candidate C?

3. What would happen if we gave f the ballots $b_1, b_2, b_3^*, \cdots b_n^*$? What possible winners could f choose?

4. If we continue this process, changing one ballot at a time, we will eventually get back to our original set of ballots: $b_1, b_2, b_3, \cdots, b_n$. Explain why this observation establishes that, given these ballots, f must choose A or B as the winner.

5. Could this argument be used to prove the Lemma if S has exactly one candidate? Walk through the steps in this case.

6. Could this argument be used to prove the Lemma if S has three, or more, candidate? Walk through the steps in this case.

Lemma. *If f is a non-imposed, non-manipulable choice function, and we use the Gibbard-Satterthwaite algorithm to produce a ranked voting system, then the ranked voting system must always produce societal preference orders that are ...*

Proof of the Lemma:

Suppose that for some set of ballots $b_1, b_2, b_3, \cdots, b_n$, the pairwise rankings produced combine to give the cycle $A > B$, $B > C$, and $C > A$.

1. Suppose that we move A, B, and C to the top of everyone's ballot (without changing anything else). Call these new ballots $b_1^{ABC}, b_2^{ABC}, b_3^{ABC}, \cdots, b_n^{ABC}$. What does the previous lemma tell us about what the choice function f chooses as a winner?

   Let's assume first that f chooses A as a winner. We will consider the other cases later.

2. Suppose that instead of moving A, B, and C to the top of the first ballot, we just move A and C to the top. Call this ballot b_1^{AC}. What winner will f choose if we give it the ballots $b_1^{AC}, b_2^{ABC}, b_3^{ABC}, \cdots, b_n^{ABC}$?

3. What winner will f choose if we give it $b_1^{AC}, b_2^{AC}, b_3^{ABC}, \cdots, b_n^{ABC}$?

4. Suppose we continue changing ballots one by one until we obtain $b_1^{AC}, b_2^{AC}, b_3^{AC}, \cdots, b_n^{AC}$? What winner will f choose for this set of ballots?

5. What does your answer to the previous question tell you about how the ranked voting system produced by f will rank A and C? Why is this a contradiction?

6. What needs to change in this argument so that we'll still reach a contradiction, if f chooses B as a winner instead of A for these ballots $b_1^{ABC}, b_2^{ABC}, b_3^{ABC}, \cdots, b_n^{ABC}$?

7. What needs to change in this argument if f chooses C as a winner instead of A for these ballots $b_1^{ABC}, b_2^{ABC}, b_3^{ABC}, \cdots, b_n^{ABC}$?

What have we proved about the ranked voting system F (created from f) so far? What do we still want to prove about it?

Lemma. *Prove that the ranked voting system F created from a non-imposed, non-manipulable choice function f satisfies ...*

Proof:

Assume that every voter ranks A above B in their ballot. What do we need to prove?

1. Suppose we move A and B to the top of every voter's ballot. Explain why A must then be the unique top-ranked candidate on everyone's ballot.

2. What does the Lemma about the set S of top-ranked candidates tell us about the candidate chosen by f?

3. Why does the ranked voting system F associated to f produce the societal ranking $A \succ B$? How does this prove the Lemma?

Lemma. *Prove that the ranked voting system F created from a non-imposed, non-manipulable choice function f satisfies ...*

Suppose we have two sets of ballots $b_1, b_2, b_3, \cdots, b_n$ and $b_1^*, b_2^*, b_3^*, \cdots, b_n^*$ that agree on their rankings of two candidates A and B. That is, for any i, the ranking of A and B on b_i is the same as on b_i^*.

1. Move A and B to the top of each ballot to obtain $b_1^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$ and $(b_1^*)^{AB}, (b_2^*)^{AB}, (b_3^*)^{AB}, \cdots, (b_n^*)^{AB}$. Explain why these new sets of ballots must still agree on their rankings of A and B. In other words, explain why the ranking of A and B on b_i^{AB} is always the same as on $(b_i^*)^{AB}$.

2. Explain why, given either set of ballots $b_1^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$ or $(b_1^*)^{AB}, (b_2^*)^{AB}, (b_3^*)^{AB}, \cdots, (b_n^*)^{AB}$ f must choose A or B as the winner.

3. We want to argue that f chooses the same candidate when given $b_1^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$ as when given $(b_1^*)^{AB}, (b_2^*)^{AB}, (b_3^*)^{AB}, \cdots, (b_n^*)^{AB}$. To do so, we will sequentially (one voter at a time) change the first set of ballots into the second. To start, explain why f will choose the same winner when given $(b_1^*)^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$ as when given $b_1^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$. Hint: only one ballot is different between the two sets, and the difference is limited by your answer to part (1).

4. Now continue changing ballots, one at a time, from b_i^{AB} to $(b_i^*)^{AB}$. Explain why each such change has no impact on the winner chosen by f.

5. Use the previous part to argue that f will choose the same winner given either $b_1^{AB}, b_2^{AB}, b_3^{AB}, \cdots, b_n^{AB}$ or $(b_1^*)^{AB}, (b_2^*)^{AB}, (b_3^*)^{AB}, \cdots, (b_n^*)^{AB}$. Explain why this implies the F ranks A and B the same way given either set of original ballots, and thus F satisfies IIA.

Question. What does Arrow's Theorem tell us about the ranked voting system F created from a non-imposed, non-manipulable choice function?

Lemma. *Suppose F is the ranked voting system created from a non-imposed, non-manipulable choice function f. If a voter v is a* *for F, then that voter is also a* *for f.*

Proof:

Suppose that v is a dictator for F, and suppose that A is the top-ranked candidate for v, with no ties. We need to show that ...

Let's suppose, instead that f chooses some candidate B. (What kind of proof are we setting up?)

1. Suppose we move A and B to the top of everyone's ballots. Explain why f will choose A in this situation.

2. Now, starting with the ballots in part (a), change one ballot at a time to put A and B back where they were before we moved them to the top. Explain why none of these changes can cause f to choose B rather than A.

3. How does this complete the proof?

Beyond the Gibbard-Satterthwaite Theorem

Is there a loophole for approval voting?

Are some kinds of voting systems less manipulable than others?