

Introduction to Computation for the Humanities and Social Sciences



CS 3
Chris Tanner

Lecture 2

“I Got 99 Computational Problems
and a Bit ain’t one”

— Jay Z

FIRST: SOME REMINDERS

1. Start bringing your laptops to class (for “Lab Time” session). If you don’t have a laptop, CIS can lend you one. Talk to us for details.
2. When e-mailing the course staff, use our emails (unless it’s a question for a specific individual, like a grading issue):
 1. cs0030tas@lists.brown.edu
 2. cs0030headtas@lists.brown.edu
3. Join Piazza for course announcements / questions (link available for our course website)

Lecture 2

- Problem Solving Workflow
- Our Data

Lecture 2

- Problem Solving Workflow
- Our Data

Worked Example

Where do your legislators exist on a political scale?



Sheldon Whitehouse



Jack Reed

?



Conservative

Liberal

Problem Solving Workflow

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Define Question

- Tradeoffs exist between data availability and questions you might want to answer.
- Ask yourself, what data is available to answer your question(s)?



?



Sheldon Whitehouse

Jack Reed



Problem Solving Workflow: Define Question

Conservative/Liberal scales

- **ontheissues.org:** Manually inspect politicians votes and statements to find their opinion on key issues
- **thatsmycongress.com:** Label bills and amendments liberal or conservative, then look at voting patterns
- **Esquire's Warren-Cruz scale:** Label issues liberal or conservative, then survey people

Problem Solving Workflow: Define Question

- **ontheissues.org**
statements
- **thatsmyco**
conservative
- **Esquire's**
conservative

Conservative/Liberal scales

On Social Issues	Hillary Clinton	Donald Trump
Abortion is a woman's unrestricted right	strongly favors	opposes
Legally require hiring women & minorities	strongly favors	opposes
Comfortable with same-sex marriage	strongly favors	opposes
Keep God in the public sphere	mixed opinion	strongly favors
Vouchers for school choice	strongly opposes	strongly favors
On Domestic Issues	Hillary Clinton	Donald Trump
Expand ObamaCare	strongly favors	opposes
EPA regulations are too restrictive	strongly opposes	strongly favors
Stricter punishment reduces crime	opposes	strongly favors
Absolute right to gun ownership	strongly opposes	strongly favors
Marijuana is a gateway drug	opposes	opposes
On Economic Issues	Hillary Clinton	Donald Trump
Privatize Social Security	strongly opposes	favors
Higher taxes on the wealthy	strongly favors	favors
Stricter limits on political campaign funds	strongly favors	favors
Stimulus better than market-led recovery	strongly favors	strongly opposes
Prioritize green energy	strongly favors	strongly opposes
On International Issues	Hillary Clinton	Donald Trump
Pathway to citizenship for illegal aliens	favors	strongly opposes
Support & expand free trade	mixed opinion	opposes
Support American Exceptionalism	opposes	favors
Expand the military	mixed opinion	strongly favors
Avoid foreign entanglements	opposes	favors

Problem Solving Workflow: Define Question

Conservative/Liberal scales

- **ontheissues.org:** Manually inspect politicians votes and statements to find their opinion on key issues
- **thatsmycongress.com:** Label bills and amendments liberal or conservative, then look at voting patterns
- **Esquire's Warren-Cruz scale:** Label issues liberal or conservative, then survey people

Problem Solving Workflow: Define Question

Detail the Specifics

- Choose a legislative body. We'll pick U.S. Senate
- What behavior should we evaluate? e.g.
 - Voting
 - Sponsorship/ co-sponsorship
 - Public speeches
 - Supporting donors
- Look for loaded words that are difficult to be precise about
 - Issues arise in defining how to measure conservative vs liberal
 - Labels are heavily context dependent; may hold diff stances on diff issues.

Problem Solving Workflow: Define Question

Where do your legislators exist on a political scale?



?



Sheldon Whitehouse

Jack Reed



Conservative

Liberal

Jim Inhofe



Elizabeth Warren

Problem Solving Workflow: Define Question

- Problem Solving Workflow
- Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Select Dataset

■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Select Dataset

- News articles
- Stated policies on the Senator's webpages
- Look at Congressional records:
 - voting
 - sponsorship
 - co-sponsorship
- Speeches during Senate sessions

Problem Solving Workflow: Select Dataset

News Articles

- Report information directly attained from elsewhere; there might be a better source
- Are heavily focused on the most popular politicians
- Sussing out political leanings requires a complex understanding of the issues
- Nunberg's analysis found that conservative/liberal labels close to names helped identify on which side they fell, but difficult to understand the degree

Problem Solving Workflow: Select Dataset

News Articles

- Report information directly attained from elsewhere; there

In the six years when **Ms. Abrams** led the **Democrats** in the State House, though, she proved to be a much more complex political character than the portraits now painted by either her most outspoken adversaries or her most fervent fans.

New York Times, August 19, 2018

- Nunberg's analysis found that conservative/liberal labels close to names helped identify on which side they fell, but difficult to understand the degree

Problem Solving Workflow: Select Dataset

Congressional Records

- Available at www.congress.gov

- Each bill, passed or not, is available with sponsors and cosponsors.

- **Voting:**

Issues put to a vote are limited by the issues considered by the majority party and the House

- **Sponsorship:**

Very few senators sponsor bills that get very far in the process. Difficult to compare alignment

- **Co-sponsorship:**

Like speeches, not all senators sponsor bills at the same frequency.

Problem Solving Workflow: Select Dataset

Congressional Records

Grouped By Vote Position

YEAs ---81

Barrasso (R-WY) Flake (R-AZ)
Bennet (D-CO) Franken (D-MN)
Blunt (R-MO) Gardner (R-CO)
Boozman (R-AR) Graham (R-SC)
Brown (D-OH) Grassley (R-IA)
Burr (R-NC) Harris (D-CA)
Cantwell (D-WA) Hassan (D-NH)
Capito (R-WV) Hatch (R-UT)
Cardin (D-MD) Heinrich (D-NM)
Carper (D-DE) Heitkamp (D-ND)
Casey (D-PA) Heller (R-NV)
Cassidy (R-LA) Hirono (D-HI)
Cochran (R-MS) Hoeven (R-ND)
Collins (R-ME) Inhofe (R-OK)
Coons (D-DE) Isakson (R-GA)
Corker (R-TN) Johnson (R-WI)
Cornyn (R-TX) Kaine (D-VA)
Cortez Masto (D-NV) Kennedy (R-LA)
Cotton (R-AR) King (I-ME)
Crapo (R-ID) Klobuchar (D-MN)
Cruz (R-TX) Lankford (R-OK)
Daines (R-MT) Lee (R-UT)
Donnelly (D-IN) Manchin (D-WV)
Enzi (R-WY) McCain (R-AZ)
Ernst (R-IA) McCaskill (D-MO)
Feinstein (D-CA) McConnell (R-KY)
Fischer (R-NE) Menendez (D-NJ)

NAYs ---17

Baldwin (D-WI) Leahy (D-VT)
Blumenthal (D-CT) Markey (D-MA)
Booker (D-NJ) Merkley (D-OR)
Duckworth (D-IL) Murphy (D-CT)
Durbin (D-IL) Murray (D-WA)
Gillibrand (D-NY) Sanders (I-VT)

Not Voting - 2

Alexander (R-TN) Moran (R-KS)

Problem Solving Workflow: Select Dataset

Speeches

- Speech transcripts are available at www.congress.gov/
- Not all senators get equal speech time on the floor, so there is a bias toward the more verbose
- We can analyze text using techniques we'll learn later in the semester
- However, text can be tricky even for humans, much less computers.

Problem Solving Workflow: Select Dataset

Speeches

- However, text can be tricky even for humans, much less computers.

**Can you discern if Senator Durbin is a
Republican or Democrat from the
following text?**

Problem Solving Workflow: Select Dataset

Speeches

- However, text can be tricky even for humans, much less computers.

Sen. DURBIN. "If we stick to the basic principles of the Affordable Care Act, or ObamaCare, we run into some problems in a hurry. The first basic principle accepted by President-Elect Trump is that we want to make sure that no health insurance company can ever discriminate against you or your family because of a preexisting condition—a baby born with cancer, a child with diabetes, a spouse who survives a cancer scare. In the old days before ObamaCare, that meant that you either were disqualified from insurance for your family or you couldn't afford it. So we said as part of the Affordable Care Act: No more—they cannot discriminate against those who are less than perfect when it comes to health because so many of us are less than perfect. OK, my friends in the Grand Old Party, how are you going to deal with that? How are we going to make sure that every family is protected with their health insurance plan? We haven't heard a word"

Problem Solving Workflow: Select Dataset

Redefined Question and Dataset

- Let's choose to look at the voting record
- Our question then becomes: which senators agree or disagree with Senator Warren, based on their voting record?

Problem Solving Workflow: Select Dataset

■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Design Method

■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Design Method

Redefined Question and Dataset

- Let's choose to look at the voting record
- Our question then becomes: which senators agree or disagree with Senator Warren, based on their voting record?

Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to



Elizabeth Warren



Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment



Elizabeth Warren



Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment



Sheldon Whitehouse

?



Elizabeth Warren

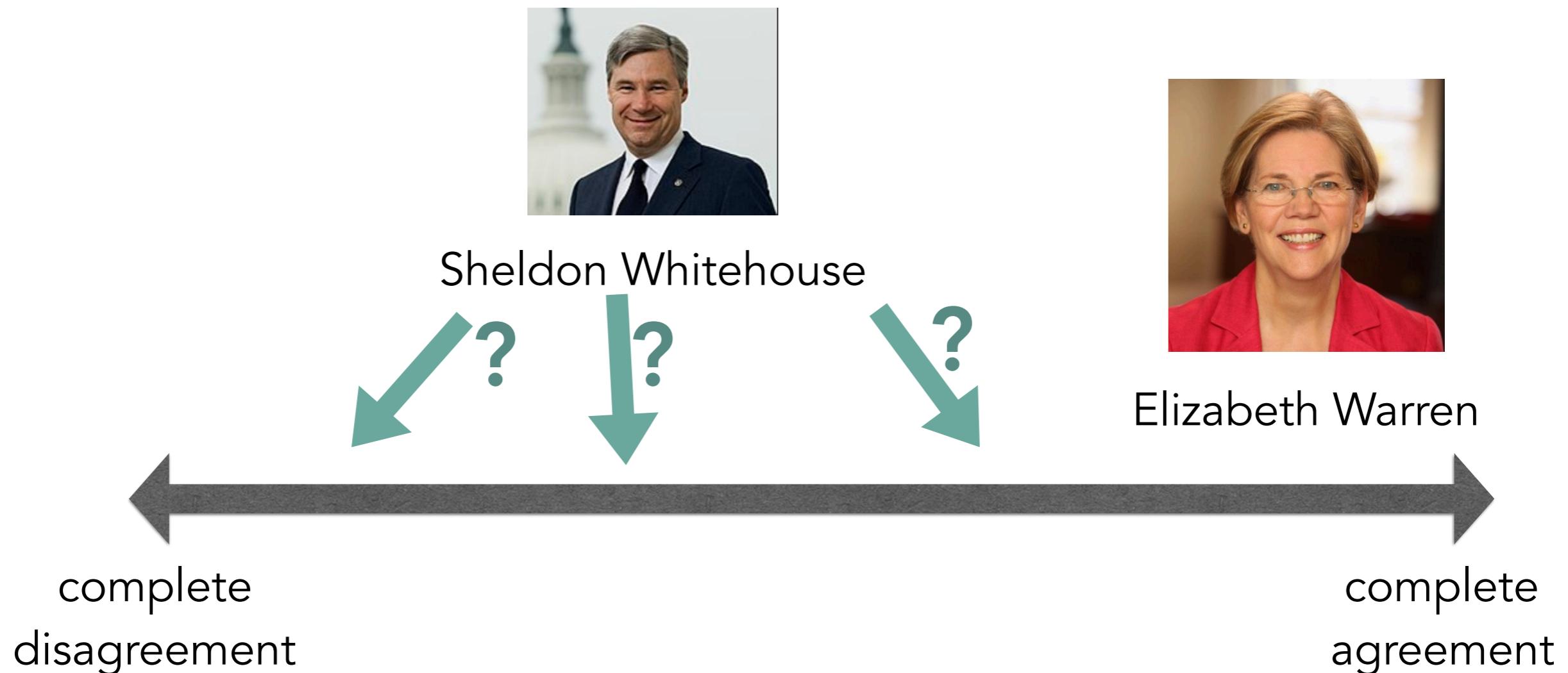


complete
disagreement

complete
agreement

Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment



Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment

Our Data

	Bill #1	Resolution #1	Bill #2
Warren	Yay	Nay	Yay
Whitehouse	Yay	Nay	Nay
Reed	Nay	Nay	Yay

Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment
- Computation always operates on inputs and produces an output

Input

Output

Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment
- Computation always operates on inputs and produces an output

Voting record
of two senators

Input

Alignment
Score

Output

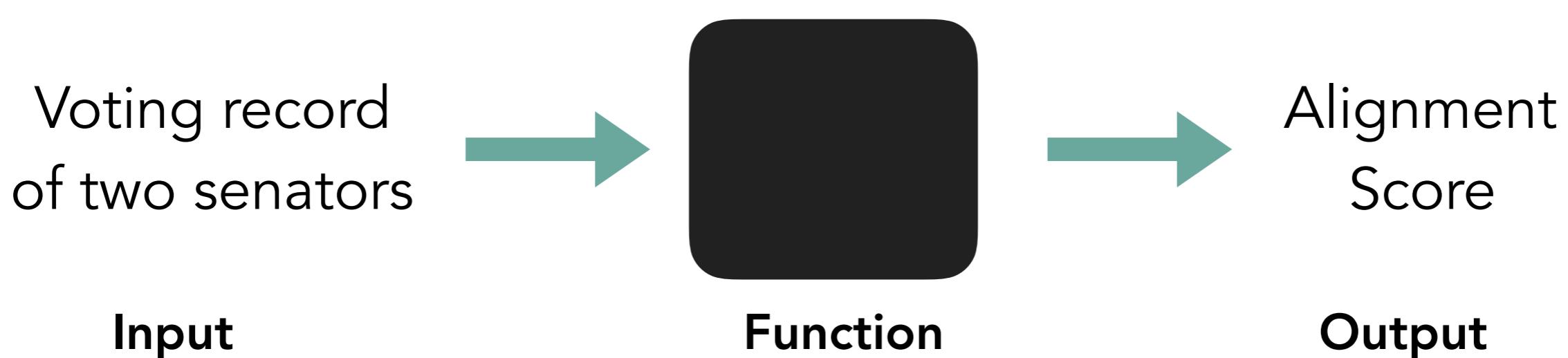
Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment
- Computation always operates on inputs and produces an output



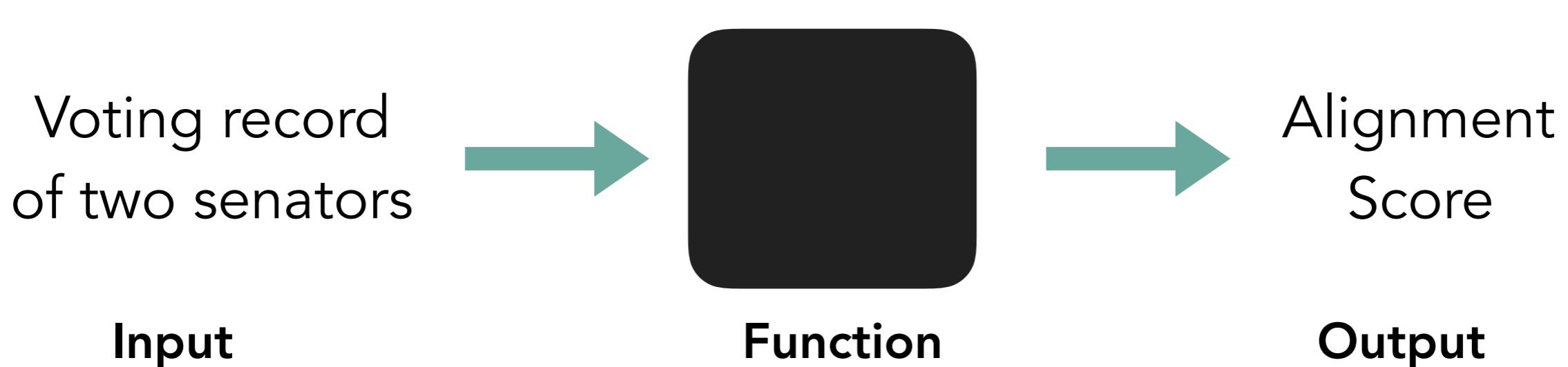
Problem Solving Workflow: Design Method

- To compute our spectrum, we choose one senator to align to
- For each other senator, we want to compute a score quantifying their voting alignment
- Computation always operates on inputs and produces an output



Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote
- Determine your output(s): aka the Range
 - What value should we use for complete disagreement?
Complete agreement? No overlapping voting info?



Problem Solving Workflow: Design Method

	Whitehouse	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

the Domain

aka the Range
for complete disagreement?
overlapping voting info?

Voting record
of two senators

Input

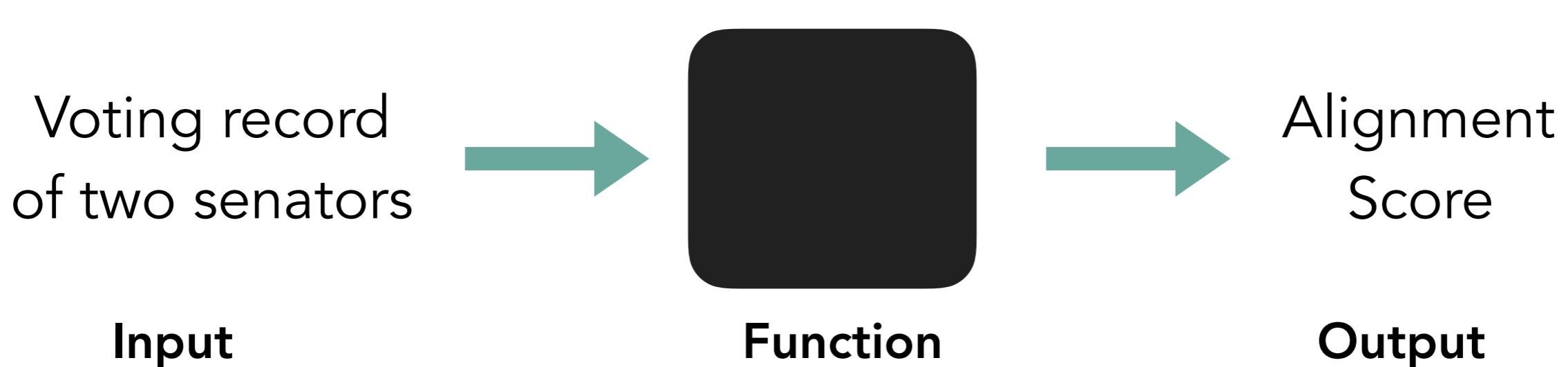
Function

Alignment
Score

Output

Problem Solving Workflow: Design Method

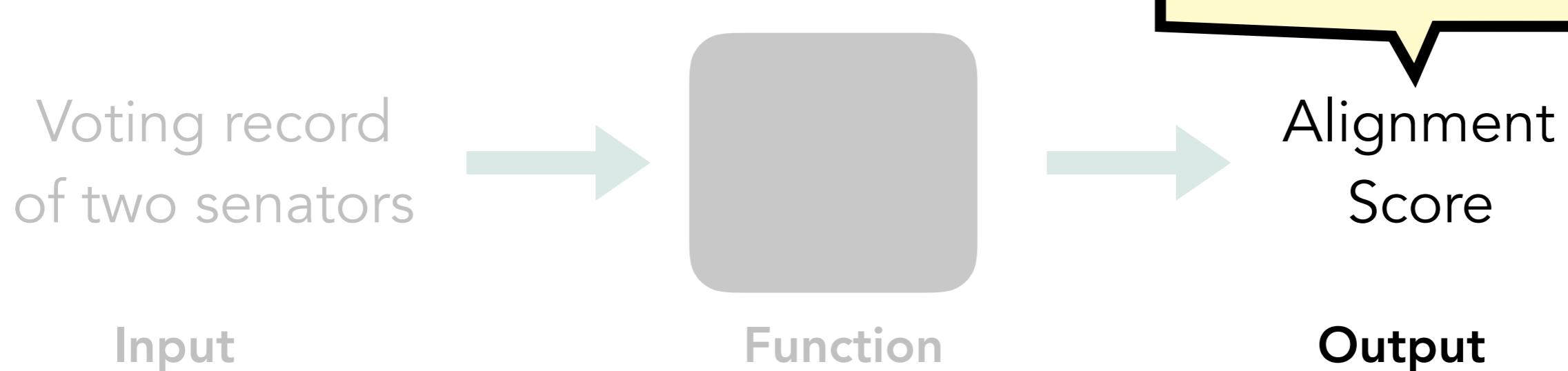
- Determine your input(s): aka the Domain
 - Yays or Nays for each vote
- Determine your output(s): aka the Range
 - What value should we use for complete disagreement?
 - Complete agreement? No overlapping voting info?



Problem Solving Workflow: Design Method

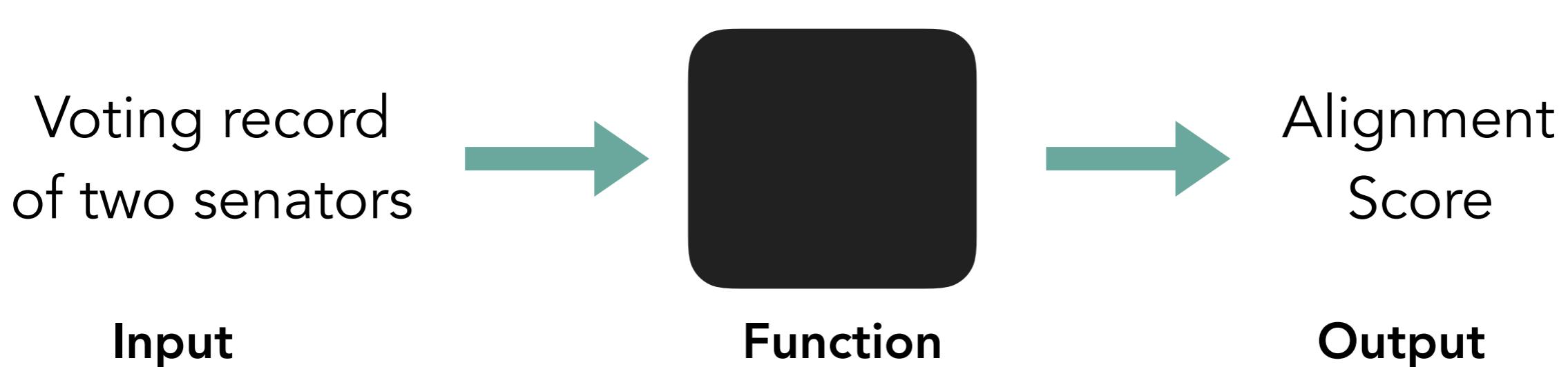
- Determine your input(s): aka the Domain
 - Yays or Nays for each vote
- Determine your output(s): aka the Range
 - What value should we use for complete disagreement? Complete agreement? No overlapping votes?

**value between
-1 and 1**



Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote
- Determine your output(s): aka the Range
 - What value should we use for complete disagreement?
 - Complete agreement? No overlapping voting info?

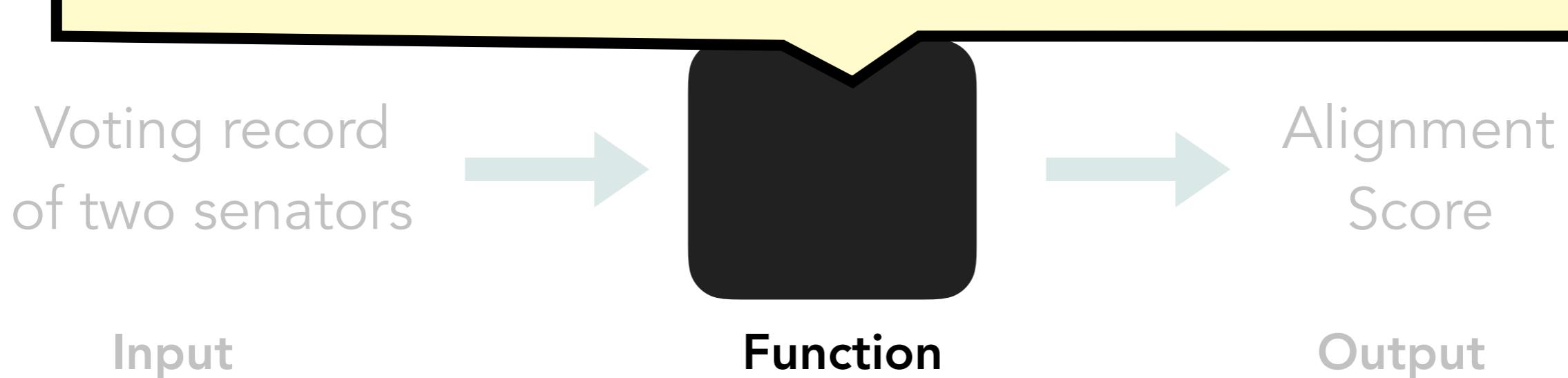


Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote

Just like in Algebra:

$$f(x) = 2x + 3$$



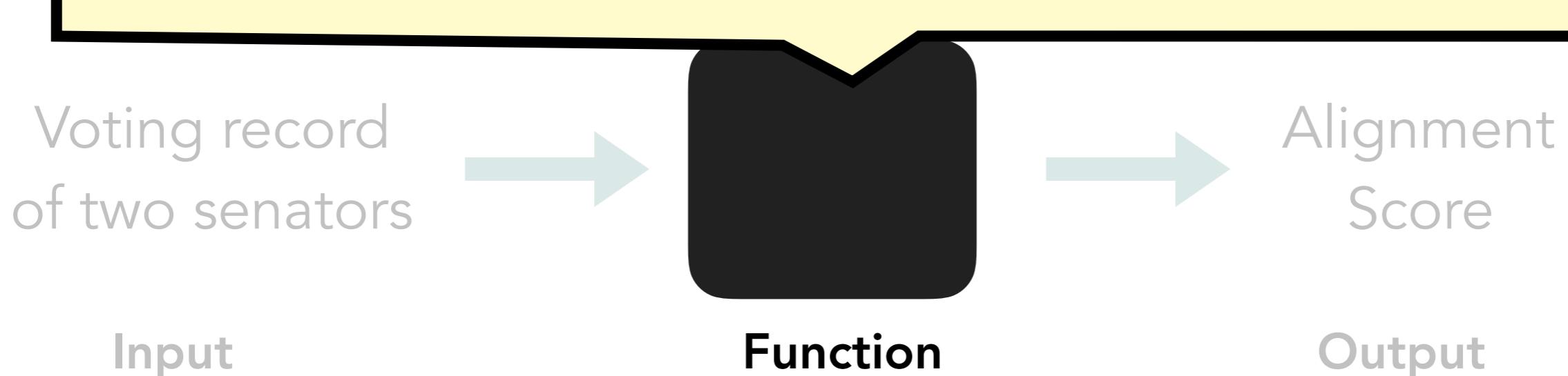
Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote

Just like in Algebra:

$$f(x) = 2x + 3$$

Given input, such as 1,
produce a value, such as 5



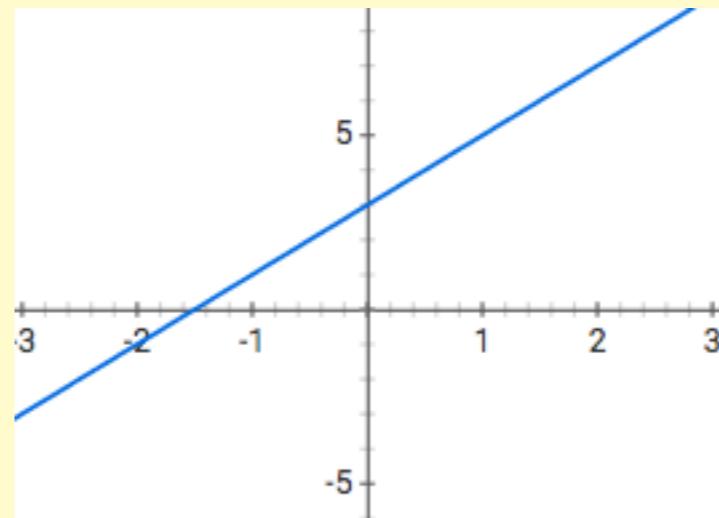
Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote

Just like in Algebra:

$$f(x) = 2x + 3$$

Given input, such as 1,
produce a value, such as 5



Voting record
of two senators

Input



Function



Alignment
Score

Output

Problem Solving Workflow: Design Method

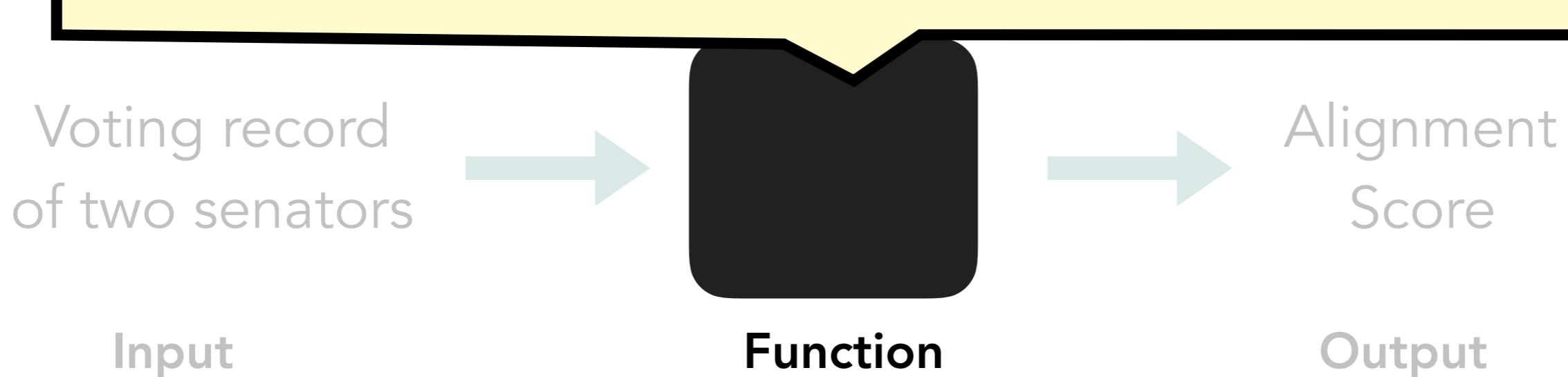
- Determine your input(s): aka the Domain
 - Yays or Nays for each vote

Just like in Algebra:

$$f(x) = 2x + 3$$

This is the crux of the entire class.

Given input, such as 1,
produce a value, such as 5



Problem Solving Workflow: Design Method

- Determine your input(s): aka the Domain
 - Yays or Nays for each vote

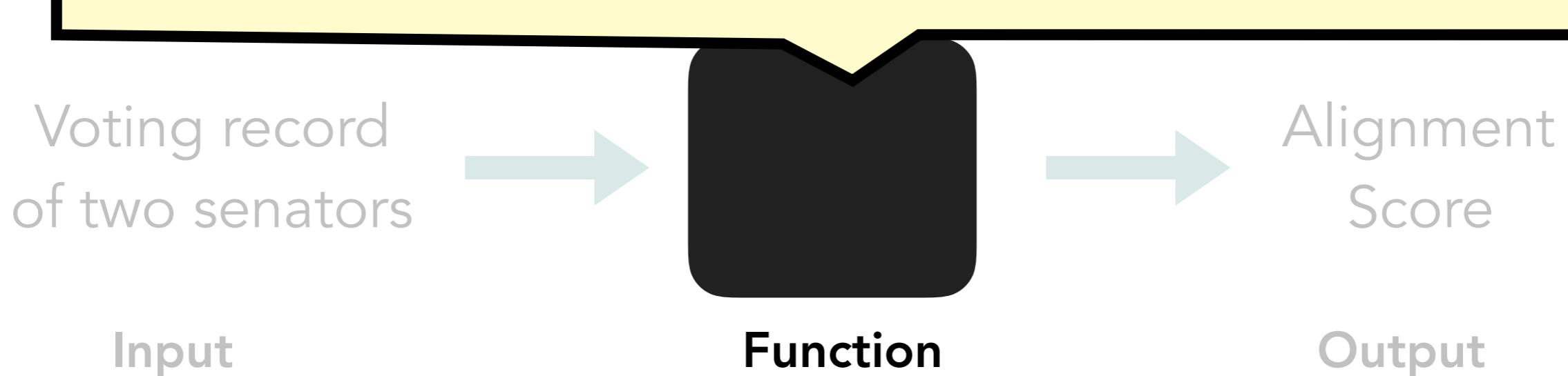
Just like in Algebra:

$$f(x) = 2x + 3$$

This is the crux of the entire class.

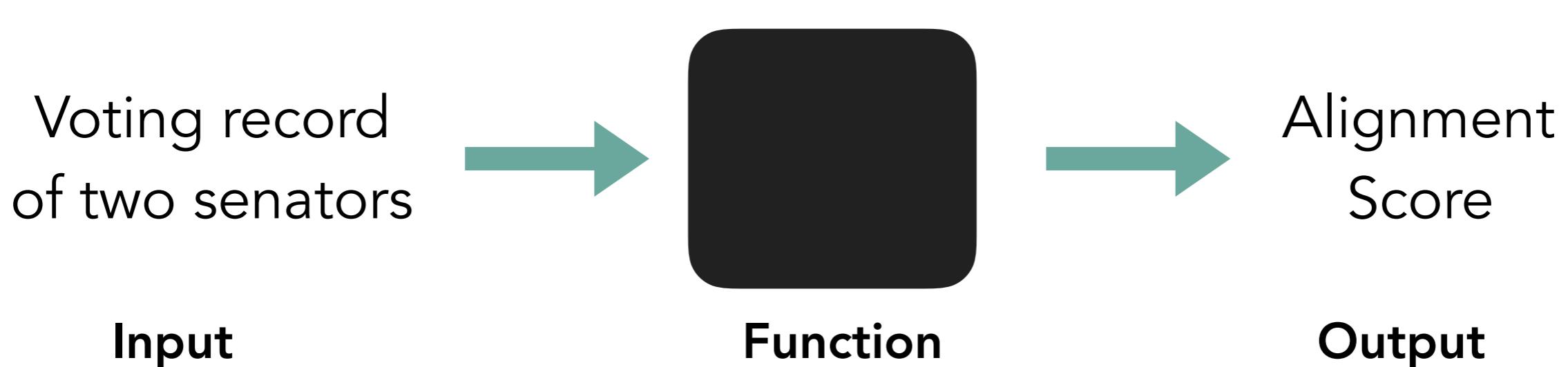
Given input, such as 1,
produce a value, such as 5

You get to make your own magic!



Problem Solving Workflow: Design Method

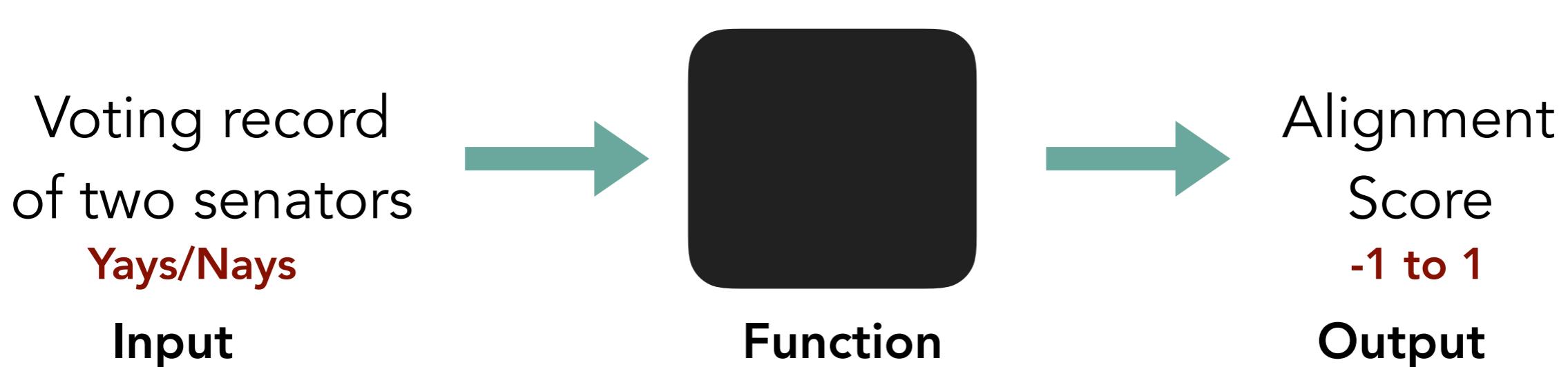
- Determine your input(s): aka the Domain
 - Yays or Nays for each vote
- Determine your output(s): aka the Range
 - What value should we use for complete disagreement?
 - Complete agreement? No overlapping voting info?



Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- Want to convert a non-numerical value into a number, what can we do?
- Our problem stresses agreement and disagreement, we need to convert yays/nays into a score based on agreement
- We can calculate the number of agreements and disagreements



Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?

$$f(A, D) \rightarrow [-1, 1]$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
- As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
- As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$
- Agreements should be **positive** in our function
$$f(A, D) = +A \dots$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
- As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$
 - Agreements should be **positive** in our function
$$f(A, D) = +A \dots$$
- As disagreements increase, the output value should decrease
$$f(0, 100) < f(50, 50)$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
- As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$
 - Agreements should be **positive** in our function
$$f(A, D) = +A \dots$$
- As disagreements increase, the output value should decrease
$$f(0, 100) < f(50, 50)$$
 - Disagreements should be **negative**
$$f(A, D) = -D \dots$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
 - As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$
 - Agreements should be **positive** in our function
$$f(A, D) = +A \dots$$
 - As disagreements increase, the output value should decrease
$$f(0, 100) < f(50, 50)$$
 - Disagreements should be **negative**
$$f(A, D) = -D \dots$$
 - Equal agreements/disagreements $\rightarrow 0.0$
$$f(50, 50) = 0.0$$

Problem Solving Workflow: Design Method

Converting Yays and Nays to Values

- How should the output value change with respect to changes of your inputs?
$$f(A, D) \rightarrow [-1, 1]$$
 - As agreements increase, the output value should increase
$$f(100, 0) > f(50, 50)$$
 - Agreements should be **positive** in our function
$$f(A, D) = +A \dots$$
 - As disagreements increase, the output value should decrease
$$f(0, 100) < f(50, 50)$$
 - Disagreements should be **negative**
$$f(A, D) = -D \dots$$
 - Equal agreements/disagreements $\rightarrow 0.0$
$$f(50, 50) = 0.0$$
 - Agreements and Disagreements should be equally weighted
$$f(A, D) = -f(D, A)$$

Problem Solving Workflow: Design Method

Scaling the Output

- Our function is bounded by 1 and -1
 - 100% agreements -> 1.0
 - Divide by positive agreements
 - 100% disagreements -> -1.0
 - Divide by positive disagreements
- Re-evaluate sample inputs and decide whether the function does what you want

$$f(A, D) = \frac{A - D}{\dots}$$

$$f(100, 0) = 1.0$$

$$f(A, D) = \frac{A - D}{A \dots}$$

$$f(0, 100) = -1.0$$

$$f(A, D) = \frac{A - D}{D \dots}$$

Problem Solving Workflow: Design Method

Scaling the Output

- Our function is bounded by 1 and -1

- 100% agreements -> 1.0

- Divide by positive agreements

- 100% disagreements -> -1.0

- Divide by positive disagreements

- Re-evaluate sample inputs and decide whether the function does what you want

$$f(A, D) = \frac{A - D}{\dots}$$

$$f(100, 0) = 1.0$$

$$f(A, D) = \frac{A - D}{A \dots}$$

$$f(0, 100) = -1.0$$

$$f(A, D) = \frac{A - D}{D \dots}$$

$$f(A, D) = \frac{A - D}{A + D}$$

Problem Solving Workflow: Design Method

How to Code our Function

- Generally, there are usually many ways to code the same calculation/function

Problem Solving Workflow: Design Method

How to Code our Function

- Generally, there are usually many ways to code the same calculation/function
- Coding (aka Programming) is simply writing a list of instructions to be executing

Problem Solving Workflow: Design Method

How to Code our Function

- Generally, there are usually many ways to code the same calculation/function
- Coding (aka Programming) is simply writing a list of instructions to be executing
- Akin to writing instructions for a food recipe, whereby a robot (aka a computer) will execute the recipe.

<ul style="list-style-type: none">• Step 1: add 1 tablespoon of salt• Step 2: add 1/2 tablespoon of cayenne	RECIPE #1	
<ul style="list-style-type: none">• Step 1: add 1.5 teaspoons of cayenne• Step 2: add 3 teaspoons of salt	RECIPE #2	
<ul style="list-style-type: none">• Step 1: add 1/2 lbs of butter• Step 2: add 3 lbs of cheese	RECIPE #3	

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

Problem Solving Workflow: Design Method

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Voting record
of two senators

Input

How to Code our Function

Output a specific pair of votes from 2 senators

Function

Output

value
between
-1 and 1

Alignment
Score

Problem Solving Workflow: Design Method

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Voting record
of two senators

Input

Now we can start to Code our Function

Let's start by putting a specific pair of votes from 2 senators

$$f(A, D) = \frac{A - D}{A + D}$$



Function

value
between
-1 and 1

Alignment
Score

Output

Problem Solving Workflow: Design Method

	White-house	Warren
Bill	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Voting record
of two senators

Input

Now we can start to Code our Function

Let's start by outputting a specific pair of votes from 2 senators

$$f(A, D) = \frac{A - D}{A + D}$$



Function

value
between
-1 and 1

Alignment
Score

Output

Problem Solving Workflow: Design Method

	White-house	Warren
Bill 1	Yay	Yay
Res	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Voting record
of two senators

Input

Now we have to Code our Function

Let's start by putting a specific pair of votes from 2 senators

$$f(A, D) = \frac{A - D}{A + D}$$



Function

value
between
-1 and 1

Alignment
Score

Output

Problem Solving Workflow: Design Method

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Voting record
of two senators

Input

Now we can start to Code our Function

Let's start by outputting a specific pair of votes from 2 senators

$$f(A, D) = \frac{A - D}{A + D}$$



Function

value
between
-1 and 1

Alignment
Score

Output

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

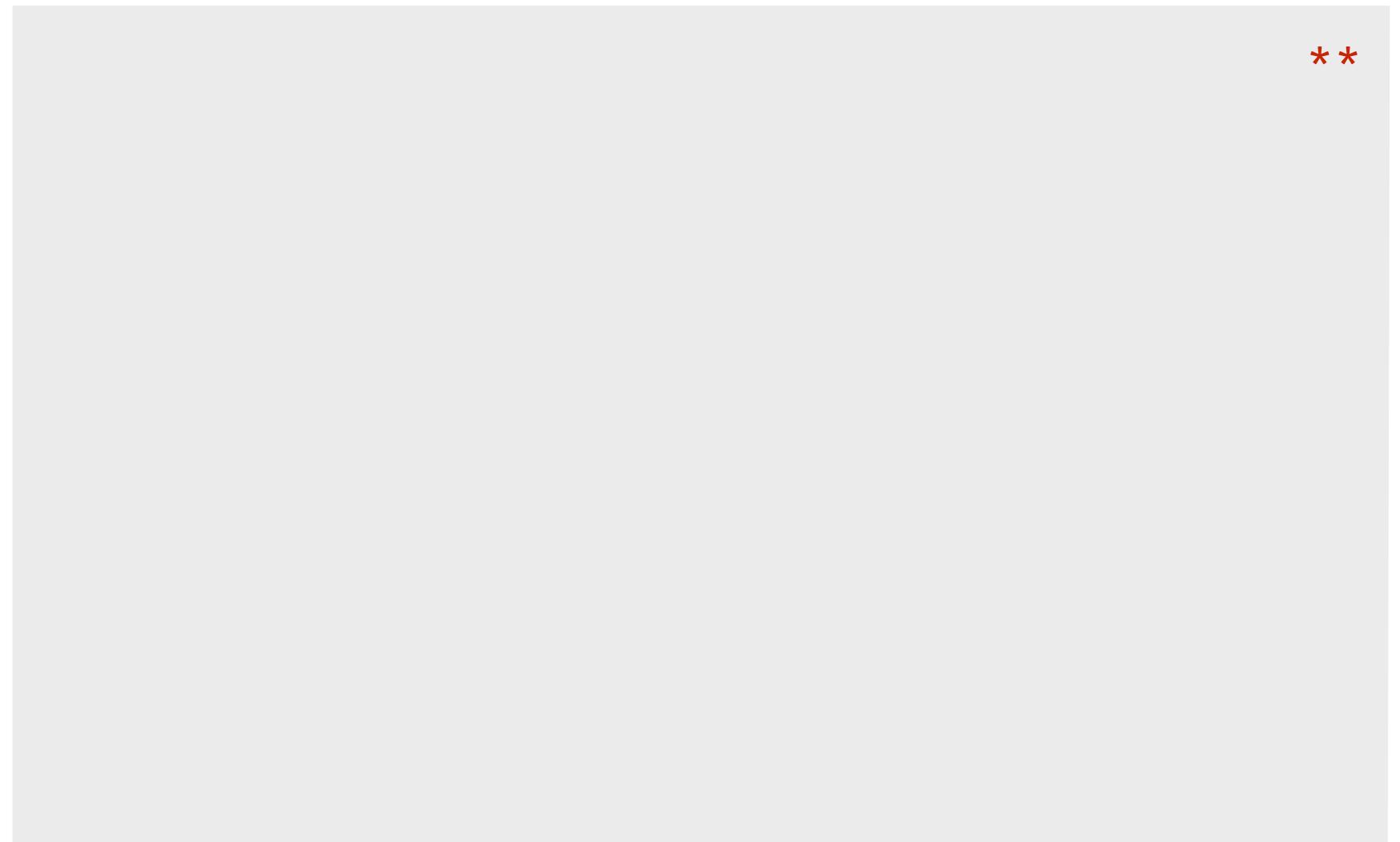
	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay



$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

**

Think, how would you instruct a person to keep track of the number of agreements and disagreements?

$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

**
if votel is Yay and vote2 is Yay:
increase agreement count by 1

or votel is Nay and vote2 is Nay:
increase agreement count by 1

$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

		White-house	Warren
		Bill 1	Bill 1
Bill 1	Res1	Yay	Yay
	Bill 2	Nay	Nay
Bill 3	Bill 3	Yay	Yay

**

```
if votel is Yay and vote2 is Yay:  
    increase agreement count by 1  
  
or votel is Nay and vote2 is Nay:  
    increase agreement count by 1  
  
or votel is Yay and vote2 is Nay:  
    increase disagreement count by 1  
  
or votel is Nay and vote2 is Yay:  
    increase disagreement count by 1
```

$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

	White-house	Warren
Bill 1	Yay	Yay
Res1	Nay	Nay
Bill 2	Nay	Yay
Bill 3	Yay	Yay

go through all pairs of votes:
**
if votel is Yay and vote2 is Yay:
 increase agreement count by 1

or votel is Nay and vote2 is Nay:
 increase agreement count by 1

or votel is Yay and vote2 is Nay:
 increase disagreement count by 1

or votel is Nay and vote2 is Yay:
 increase disagreement count by 1

$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

How to Code our Function

Let's make our input a specific pair of votes from 2 senators

		White-house	Warren
		Bill 1	Bill 1
Bill 1	Res1	Yay	Yay
	Bill 2	Nay	Nay
Bill 3	Bill 3	Yay	Yay

go through all pairs of votes:
**
if votel is Yay and vote2 is Yay:
 increase agreement count by 1

or votel is Nay and vote2 is Nay:
 increase agreement count by 1

or votel is Yay and vote2 is Nay:
 increase disagreement count by 1

or votel is Nay and vote2 is Yay:
 increase disagreement count by 1

return (agreement count - disagreement count) /
(agreement count + disagreement count)

$$f(A, D) = \frac{A - D}{A + D}$$

** not actually valid computer language; it's pseudocode, as we're just trying to convey the logic.

Problem Solving Workflow: Design Method

■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Analyze Results

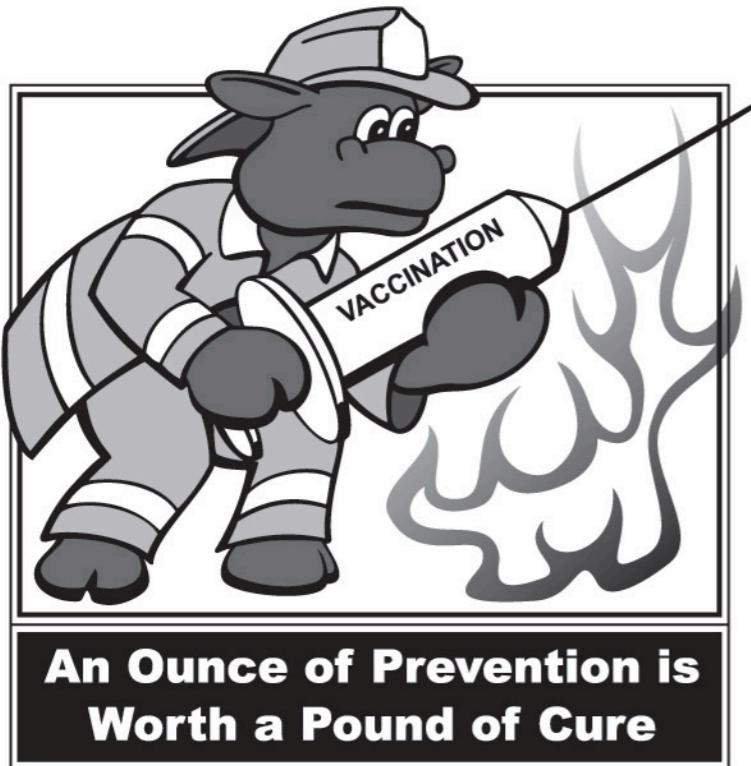
- Problem Solving Workflow
- Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Analyze Results

- Sanity check your results — start with simple, small examples
- Do you have errors?
- Are there values that do not fit into your desired range?
- Look at intermediate values you produce, like agreements and disagreements, do they agree with the data you have?
- Look for the extreme cases to ensure your function seems to work properly: do Sanders and Warren have a high score? Do Warren and Inhofe have a low score?

Problem Solving Workflow: Analyze Results



- Everyone makes bugs, quite often actually (even FB, gmail, etc have app updates regularly)
- “*An ounce of prevention is worth a pound of cure*”
— Benjamin Franklin
- Debugging code can sometimes take a long time
- Deliberately changing correct code, tinkering, and exploring to see what happens is one of the best ways to learn.

Problem Solving Workflow: Analyze Results

- Problem Solving Workflow
- Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Analyze Results

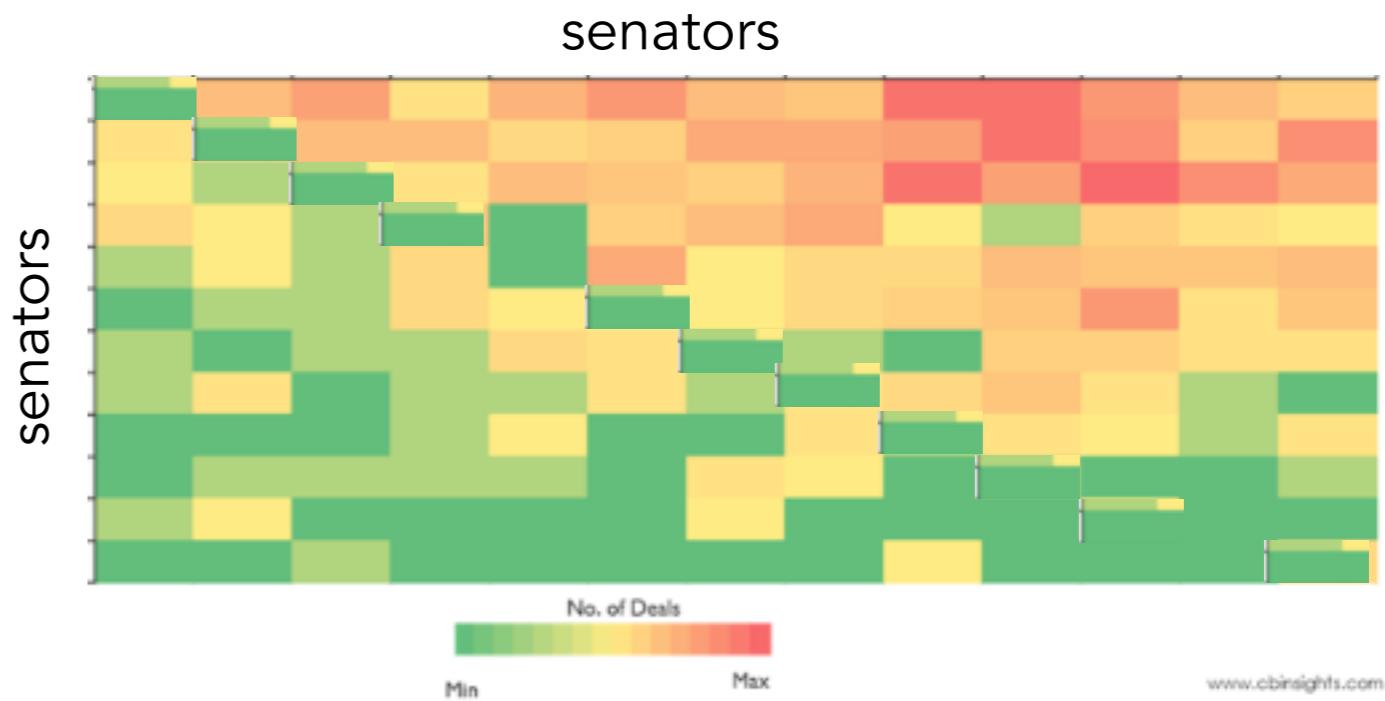
■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Problem Solving Workflow: Communicate Findings

- In this example, it's sufficient to just output a score for any two candidates
- One could imagine making a bar graph for a given candidate (Warren), where the X-axis is various other candidates (Whitehouse, Reed, Sanders, Inhofe, etc) and Y-axis is the score from -1 to 1.
- One could imagine making a heat map which displays similarity between *all* senators



Problem Solving Workflow: Analyze Results

■ Problem Solving Workflow

■ Our Data

1. Define question
2. Select appropriate dataset
3. Design a valid method
4. Analyze your results
5. Communicate your findings

Datasets

- Problem Solving Workflow
- Our Data

Overview of Data

- Remember from last lecture, data can come from a plethora of sources
- Two highly common sources, especially for this course:
 - downloadable pre-packaged, prepared data (in text file format)
 - webpages

Datasets

URL: Universal Resource Locator

Our Senate Voting data:

The screenshot shows the official website of the United States Senate at https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfms.cfm?congress=115&session=2&vote=00189. The page displays information about a roll call vote in the 115th Congress, 2nd Session. Key details include:

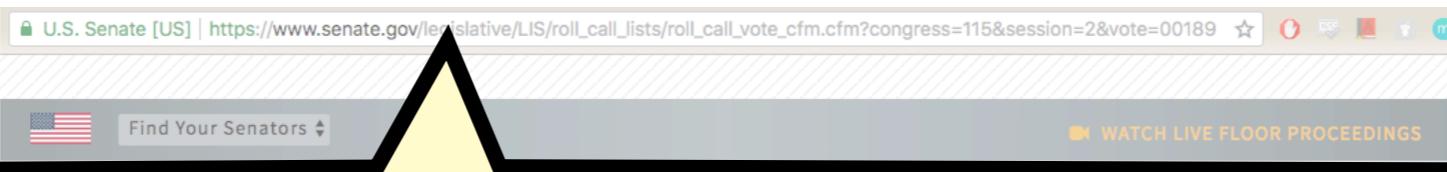
- Question:** On the Amendment (Kennedy Amdt. No. 3703)
- Vote Number:** 189 **Vote Date:** August 21, 2018, 12:53 PM
- Required For Majority:** 1/2 **Vote Result:** Amendment Agreed to
- Amendment Number:** S.Amdt. 3703 to S.Amdt. 3695 to H.R. 6157 (Department of Defense Appropriations Act, 2019)
- Statement of Purpose:** To increase funding for the National Suicide Prevention Lifeline.
- Vote Counts:** YEAs 95, NAYs 0, Not Voting 5

The page also includes links for "Vote Summary", "By Senator Name", "By Vote Position", and "By Home State". A sidebar on the left provides links to various legislative resources like "Votes", "Nominations", "Calendars & Schedules", etc., and sections for "Find Votes" and "Find Bills".

Datasets

URL: Universal Resource Locator

Our Senate Voting data:



URL: https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=115&session=2&vote=00189

text of recent bills and resolutions on the Web, or order them from the Senate or House Document Rooms, or you can find

Carper (D-DE), Yea
Casey (D-PA), Yea

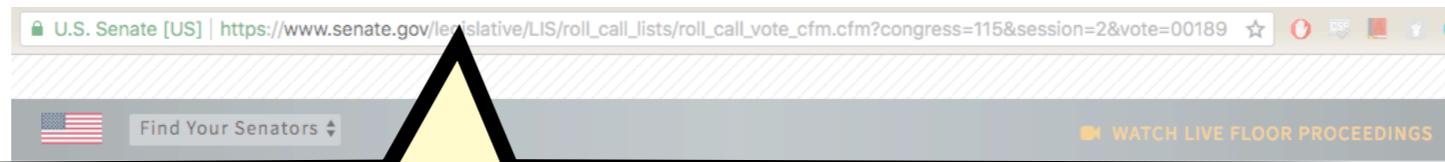
Isakson (R-GA), Yea
Johnson (R-WI), Yea

Schumer (D-NY), Yea
Scott (R-SC), Yea

Datasets

URL: Universal Resource Locator

Our Senate Voting data:



URL: https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=115&session=2&vote=00189

https://www.senate.gov: Homepage of the US Senate

/legislative: Information related to legislation (vs, people, history, etc)

/LIS: Stands for Legislative Information System

/roll_call_lists: list of roll votes for current session

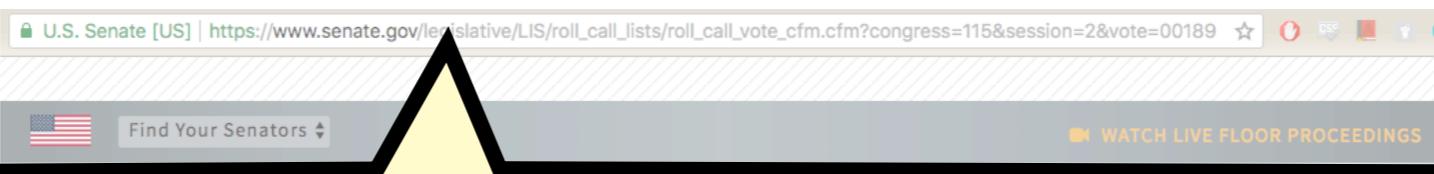
/roll_call_vote_cfm.cfm: A specific roll call vote page in ColdFusion Markup (cfm)

?congress=115&session=2&vote=00189: URL parameters. Server processes parameters to provide specific information

Datasets

URL: Universal Resource Locator

Our Senate Voting data:



The screenshot shows a browser window with the URL https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=115&session=2&vote=00189. A large black arrow points from the text in the slide to the parameter values in the URL.

?congress=115&session=2&vote=00189

- Parameters allow a single webpage to dynamically provide content based on their values
- Webpage URL is followed by ? then the parameters
- Parameters are paired with their values with =
- Additional parameters are separated by & symbols
- Public documentation may not always be available, but sometimes parameters and values can be guessed

text of recent bills and resolutions on the
Web, or order them from the Senate or
House Document Rooms. or you can find

Carper (D-DE), Yea
Casey (D-PA), Yea

Isakson (R-GA), Yea
Johnson (R-WI), Yea

Schumer (D-NY), Yea
Scott (R-SC), Yea

Datasets

URL: Universal Resource Locator

Our Senate Voting data:

The screenshot shows the official website of the United States Senate. At the top, there's a navigation bar with links for "SENATORS", "COMMITTEES", "LEGISLATION & RECORDS" (which is highlighted in gold), "ART & HISTORY", and "REFERENCE". Below the navigation, a sidebar on the left lists various legislative resources like "Votes", "Nominations", "Calendars & Schedules", etc. The main content area displays a "Roll Call Vote 115th Congress - 2nd Session" for "Amendment (Kennedy Amdt. No. 3703)". It provides details such as the "Vote Number: 189", "Vote Date: August 21, 2018, 12:53 PM", and the "Vote Result: Amendment Agreed to". A "Statement of Purpose" is also included. At the bottom of the main content, there's a section titled "Alphabetical by Senator Name" listing numerous senators with their party and voting status.

Senator	Party	Vote
Alexander (R-TN)	Yea	
Baldwin (D-WI)	Yea	
Barrasso (R-WY)	Yea	
Bennet (D-CO)	Yea	
Blumenthal (D-CT)	Yea	
Blunt (R-MO)	Yea	
Booker (D-NJ)	Yea	
Boozman (R-AR)	Yea	
Brown (D-OH)	Yea	
Burr (R-NC)	Yea	
Cantwell (D-WA)	Yea	
Capito (R-WV)	Yea	
Cardin (D-MD)	Yea	
Carper (D-DE)	Yea	
Casey (D-PA)	Yea	
Gillibrand (D-NY)	Yea	
Graham (R-SC)	Yea	
Grassley (R-IA)	Yea	
Harris (D-CA)	Yea	
Hassan (D-NH)	Yea	
Hatch (R-UT)	Yea	
Heinrich (D-NM)	Yea	
Heitkamp (D-ND)	Yea	
Heller (R-NV)	Yea	
Hirono (D-HI)	Yea	
Hoover (R-ND)	Yea	
Hyde-Smith (R-MS)	Yea	
Inhofe (R-OK)	Yea	
Isakson (R-GA)	Yea	
Johnson (R-WI)	Yea	
Nelson (D-FL)	Yea	
Paul (R-KY)	Yea	
Perdue (R-GA)	Yea	
Peters (D-MI)	Yea	
Portman (R-OH)	Yea	
Reed (D-RI)	Yea	
Risch (R-ID)	Yea	
Roberts (R-KS)	Yea	
Rounds (R-SD)	Yea	
Rubio (R-FL)	Yea	
Sanders (I-VT)	Yea	
Sasse (R-NE)	Yea	
Schatz (D-HI)	Not Voting	
Schumer (D-NY)	Yea	
Scott (R-SC)	Yea	

Datasets

URL: Universal Resource Locator

Our Senate Voting data:

Oooo, look, you can download the data (an XML file) which is being displayed here!

The screenshot shows the U.S. Senate website's "Senate Votes" page. A yellow callout box highlights the "XML" link next to the search bar. The URL in the browser is https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=115&session=2&vote=00189.

Alphabetical by Senator Name

Alexander (R-TN), Yea	Gillibrand (D-NY), Yea	Nelson (D-FL), Yea
Baldwin (D-WI), Yea	Graham (R-SC), Yea	Paul (R-KY), Yea
Barrasso (R-WY), Yea	Grassley (R-IA), Yea	Perdue (R-GA), Yea
Bennet (D-CO), Yea	Harris (D-CA), Yea	Peters (D-MI), Yea
Blumenthal (D-CT), Yea	Hassan (D-NH), Yea	Portman (R-OH), Yea
Blunt (R-MO), Yea	Hatch (R-UT), Yea	Reed (D-RI), Yea
Booker (D-NJ), Yea	Heinrich (D-NM), Yea	Risch (R-ID), Yea
Boozman (R-AR), Yea	Heitkamp (D-ND), Yea	Roberts (R-KS), Yea
Brown (D-OH), Yea	Heller (R-NV), Yea	Rounds (R-SD), Yea
Burr (R-NC), Yea	Hirono (D-HI), Yea	Rubio (R-FL), Yea
Cantwell (D-WA), Yea	Hoover (R-ND), Yea	Sanders (I-VT), Yea
Capito (R-WV), Yea	Hyde-Smith (R-MS), Yea	Sasse (R-NE), Yea
Cardin (D-MD), Yea	Inhofe (R-OK), Yea	Schatz (D-HI), Not Voting
Carper (D-DE), Yea	Isakson (R-GA), Yea	Schumer (D-NY), Yea
Casey (D-PA), Yea	Johnson (R-WI), Yea	Scott (R-SC), Yea

Datasets

- **XML: eXtensible Markup Language**
- A markup “language” provides a system for annotating a document
- Markup languages are not programming languages!
- **Extensible** implies that these documents can have a flexible architecture, unlike HTML which has a specific set of elements
- Different pieces of information can be given labels called **elements**
- Elements can also have **attributes**, which provide additional context

About XML

vote_115_2_00189.xml

```
<roll_call_vote>
  <congress>115</congress>
  <session>2</session>
  ...
  <members>
    <member>
      <member_full>Alexander (R-TN)</member_full>
      <last_name>Alexander</last_name>
      <first_name>Lamar</first_name>
      <party>R</party>
      <state>TN</state>
      <vote_cast>Yea</vote_cast>
      ...
    </member>
  </members>
</roll_call_vote>

<myelement class='blue'>
  ...
</myelement>
```

Datasets

- **XML: eXtensible Markup Language**
- A markup “language” provides a system for annotating a document
- Markup languages are not programming languages!
- **Extensible** implies that these documents can have a flexible architecture, unlike HTML which has a specific set of elements
- Different pieces of information can be given labels called **elements**
- Elements can also have **attributes**, which provide additional context

About XML

vote_115_2_00189.xml

TAGS



```
<roll_call_vote>
<congress>115</congress>
<session>2</session>
...
<members>
  <member>
    <member_full>Alexander (R-TN)</member_full>
    <last_name>Alexander</last_name>
    <first_name>Lamar</first_name>
    <party>R</party>
    <state>TN</state>
    <vote_cast>Yea</vote_cast>
  ...
  </member>
</members>
</roll_call_vote>

<myelement class='blue'>
  ...
</myelement>
```

Datasets

- **XML: eXtensible Markup Language**
- A markup “language” provides a system for annotating a document
- Markup languages are not programming languages!
- **Extensible** implies that these documents can have a flexible architecture, unlike HTML which has a specific set of elements
- Different pieces of information can be given labels called **elements**
- Elements can also have **attributes**, which provide additional context

About XML

vote_115_2_00189.xml

ELEMENT

```
<roll_call_vote>
  <congress>115</congress>
  <session>2</session>
  ...
  <members>
    <member>
      <member_full>Alexander (R-TN)</member_full>
      <last_name>Alexander</last_name>
      <first_name>Lamar</first_name>
      <party>R</party>
      <state>TN</state>
      <vote_cast>Yea</vote_cast>
      ...
    </member>
  </members>
</roll_call_vote>

<myelement class='blue'>
  ...
</myelement>
```

Datasets

- **XML: eXtensible Markup Language**
- A markup “language” provides a system for annotating a document
- Markup languages are not programming languages!
- **Extensible** implies that these documents can have a flexible architecture, unlike HTML which has a specific set of elements
- Different pieces of information can be given labels called **elements**
- Elements can also have **attributes**, which provide additional context

About XML

```
<roll_call_vote>
  <congress>115</congress>
  <session>2</session>
  ...
  <members>
    <member>
      <member_full>Alexander (R-TN)</member_full>
      <last_name>Alexander</last_name>
      <first_name>Lamar</first_name>
      <party>R</party>
      <state>TN</state>
      <vote_cast>Yea</vote_cast>
      ...
    </member>
  </members>
</roll_call_vote>

<myelement class='blue'>
  ...
</myelement>
```

vote_115_2_00189.xml
**ELEMENT'S
CONTENT**



Datasets

- **XML: eXtensible Markup Language**
- A markup “language” provides a system for annotating a document
- Markup languages are not programming languages!
- **Extensible** implies that these documents can have a flexible architecture, unlike HTML which has a specific set of elements
- Different pieces of information can be given labels called **elements**
- Elements can also have **attributes**, which provide additional context

About XML

vote_115_2_00189.xml

```
<roll_call_vote>
  <congress>115</congress>
  <session>2</session>
  ...
  <members>
    <member>
      <member_full>Alexander (R-TN)</member_full>
      <last_name>Alexander</last_name>
      <first_name>Lamar</first_name>
      <party>R</party>
      <state>TN</state>
      <vote_cast>Yea</vote_cast>
      ...
    </member>
  </members>
</roll_call_vote>

<myelement class='blue'>
  ...
</myelement>
```

ATTRIBUTE



Lecture 2: Homework

Install an editor, Python, and start playing around.

Lecture 2: Last Words

With practice, the problem solving workflow will not seem like a foreign, strict formula, but will flow naturally.