Forming a computational problem
Where do your legislators fit on a political scale?

Conservative  Sheldon Whitehouse

Liberal  Jack Reed
Problem Solving Workflow

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

- Tradeoffs exist between data availability and questions you might want to answer.
- Ask yourself, what data is available to answer these questions?
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
Detail the specifics

- Choose a legislative body. In our case, we’ll look at the US Senate
- What behavior should we evaluate?
  - Voting
  - Sponsorship/co-sponsorship of bills and resolutions
  - Public speeches
  - Supporting donors
- Look for loaded words that are difficult to be precise about
  - Issues arise when we try to decide how to measure conservative and liberal
  - Conservative/Liberal are heavily context dependent. People may hold conservative or liberal ideas for different issues
Where do your legislators fit on a political scale?

Jim Inhofe

Sheldon Whitehouse

Jack Reed

Elizabeth Warren
Finding a dataset

• News articles
• Stated policies on the senator’s webpages
• Look at congressional records
  • Voting
  • Sponsorship
  • Cosponsorship
• Speeches during senate session
Refining questions from data

Often there exists a tradeoff between the questions we want answer and the datasets available.

Remember our possible data sources

- News articles
- Stated policies on the senator’s webpages
- Look at congressional records
  - Voting
  - Sponsorship
  - Cosponsorship
- Speeches during senate session
WASHINGTON — At her confirmation hearing on Tuesday to be education secretary, Betsy DeVos vigorously defended her work steering taxpayer dollars from traditional public schools, arguing that it was time to move away from a “one size fits all” system and toward newer models for students from preschool to college.

News Articles

• News articles report information directly attained from elsewhere, there might be a better source

• They 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
Congressional Records

- Available at [http://www.congress.gov](http://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
- Cosponsorship
  - Like speeches, not all senators sponsor bills at the same frequency.
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”
Refined 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?*
Defining our 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.

<table>
<thead>
<tr>
<th></th>
<th>Bill 1</th>
<th>Res 1</th>
<th>Bill 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baldwin</td>
<td>Yay</td>
<td>Nay</td>
<td>Yay</td>
</tr>
<tr>
<td>Barrasso</td>
<td>Nay</td>
<td>Yay</td>
<td>Nay</td>
</tr>
<tr>
<td>Bennet</td>
<td>Nay</td>
<td>Nay</td>
<td>Yay</td>
</tr>
</tbody>
</table>
Defining our method

Voting record of two senators → Alignment score
Defining our method

- We want to derive a function that calculates an alignment score from the voting record of two senators.
Deriving a function

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

Voting record of two senators
Yays/Nays

Function

Alignment score [-1, 1]
Converting Yays/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

Voting record of two senators (Agreements, Disagreements) → Function → Alignment score [-1, 1]
Defining our function

• 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 
    \[ f(A, D) = +A \ldots \]

  • As disagreements increase, the output value should **decrease** 
    \[ f(0, 100) < f(50, 50) \]
  • Disagreements should be **negative** 
    \[ f(A, D) = -D \ldots \]

• Equal agreements/disagreements → 0.0 
  \[ f(50, 50) = 0.0 \]

• Agreements and Disagreements should be equally weighted
  \[ f(A, D) = (A - D) * w \]
Scaling the output

- Our function is bounded by 1 and -1
  \[ f(A, D) = \frac{A - D}{\ldots} \]

- 100% agreements -> 1.0
  \[ f(100, 0) = 1.0 \]

- Divide by positive agreements
  \[ f(A, D) = \frac{A - D}{A\ldots} \]

- 100% disagreements -> -1.0
  \[ f(0, 100) = -1.0 \]

- Divide by positive disagreements
  \[ f(A, D) = \frac{A - D}{D\ldots} \]

- Re-evaluate sample inputs and decide whether the function does what you want
  \[ f(A, D) = \frac{A - D}{A + D} \]
Choosing our coding scheme

• Converting Yay, Nay, Not Voting to agreements and disagreements could be done with nested conditional functions

\[ \text{IF}(A1=“Yea”, 1000, \text{IF}(A1=“Nay”, 0, 25)) \]

• Or we could define a function that takes in two votes and outputs agree, disagree, or someone abstained.
Coding the votes

- If the output is -1, 0, and 1 then let’s try an input of -1, 0, 1

<table>
<thead>
<tr>
<th></th>
<th>Yea</th>
<th>Not Voting</th>
<th>Nay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yea</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>Not Voting</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nay</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Coding the votes

- Nay = -1, Not Voting = 0, Yea = 1
Computation Summary

• Import voting records using **IMPORTXML**

• Convert Yeas, Not Voting and Nays to values 1,0, and -1.

  \[ \text{IF}(A1=\text{"Yea"}, 1, \text{IF}(A1=\text{Nay}, -1, 0)) \]

• Multiply coded votes to obtain agreement values

• Sum number of agreements and disagreements

• Compute alignment score from agreements and disagreements
Analyze your results

- Sanity check your results
- Do you have error cells?
- Are there values that do not fit into your desired range?
- Look for the extreme cases, do Sanders and Warren have high score? Do Warren and Inhofe have a low score?
- Look at intermediate values you produce, like agreements and disagreements, do they agree with the data you have?

<table>
<thead>
<tr>
<th></th>
<th>Baldwin</th>
<th>Barrasso</th>
<th>Bennet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warren</td>
<td>0.7</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
About our voting data
URL: Universal Resource Locator


- **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 Adobe ColdFusion Markup (cfm)
- **?congress=115&session=1&vote=00001**: URL parameters. Server processes parameters to provide specific information
URL Parameters

?congress=115&session=1&vote=00001

• 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
About XML

• 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

<roll_call_vote>
  <congress>115</congress>
  <session>1</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>
Spreadsheet tidbits for today
Conditional Formatting

- Format cells depending on their values
- Helpful to illustrate differences
Conditional Formatting

- Choose the cell range to apply it to
- Choose conditional formula
- Optional custom formula
- Top rule gets priority if there is a conflict
## Conditional Formatting

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hello</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IMPORTXML(url, xpath_query)

- Takes two inputs, a url to pull the data from and an xpath_query.
- Outputs data to the cells at or below and to the right of the current cell.
- XPath query is a search syntax for finding the elements’ data of interest to you.
XPath Query

- "//member" - find all <member> elements anywhere in the xml document

- "//member/member_full" find the <member_full> elements that exist in any <member> element

- "//member[@att1]" find <member> elements with the att1 attribute

- "//member[@att1='value']" find <member> elements with the att1 attribute equal to "value".

<roll_call_vote>
  <congress>115</congress>
  <session>1</session>
  ...
  <members type="senior">
    <member type="senior">
      <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>
Quotes within Quotes

- Frequently we want Google Sheets to interpret a value as text, so we will use ""

- What happens if we want the text to contain quotes themselves, like the xpath_query?

- We can use "" inside of "", but cannot use "" inside of ".

- We can escape the " inside of a text string by doubling it

Why doesn't this work?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>#ERROR!</td>
</tr>
</tbody>
</table>

However, these do work

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carl said, 'Hello'</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carl said, &quot;Hello&quot;</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Lab Activity

• You will receive credit for a ‘good effort completion’

• This means you have worked through all parts of the lab, but some errors may prevent you from getting the correct answer

• If you go to TA hours for help on the homework, your activity needs to be at the ‘good effort completed’ level.

• Working in pairs is allowed, but both people need to be engaged and working through the activity

• If two people are working at a single computer, switch who is at the keyboard after 15 minutes