# Project 1
# and
# Senator x Senator Ranking

Sep 29 2015

# Project 1

**Due Dates**
Proposal: 10/06
Project: 10/13

- Choose a problem related to the work we've done so far in the course
  - Choose a different dataset and perform a similar analysis
  - Perform a different analysis on the Senate dataset to answer a different political question
- You must have a **testable** hypothesis!
  - "Senators who vote similarly to Sanders are Democrats and senators who vote differently from Sanders are Republicans"

# Project 1

- Proposal
  - Background
  - Claim
  - Data
  - Analysis Steps
  - Potential Roadblocks

Meet with TAs/instructor!
Ask Questions!

# Project 1

- Proposal
  - Background
  - Claim
  - Data
  - Analysis Steps
  - Potential Roadblocks
- Project
  - All Google Spreadsheets and a **website**

Meet with TAs/instructor!
Ask Questions!

# Grading Rubric

- Proposal
  - Clarity
  - Forethought
- Design
- Execution
  - Did you do it right? Handle bad data?
- Website (Google Sites – easy!), Analysis, Discussion

# Proposal Rubric

- Testable hypothesis
- Problem context for hypothesis
- Data description (including source)
- Steps
  - *I'll import names and grades into two columns;*
  - *I'll compute the average number of cases per region;*
  - *I'll sort by the number of occurrences."*

  - Numbered
  - Specific
  - Manageable

# Proposal Rubric, cont.

- How will hypothesis be evaluated using the results?
- What would validate/invalidate the hypothesis?
- Description of a visual representation of results (or reason why no such thing is appropriate)
- Potential roadblocks
  - Example: “I don’t yet know how to measure variability in data”
  - Example: “Data is in form that may require tricky manipulation” (with details).

# Project 1

**Due Dates**
Proposal: 10/06
Project: 10/13

| Category | # Points Earned |
|---|---|
| Proposal | 25 |
| Design Elements | 25 |
| Execution | 25 |
| Code Quality | 10 |
| Website Presentation & Discussion | 15 |
| TOTAL | 100 |

~8h of focused work

**Proposal (25 points)**

_____ (4 points) A **hypothesis** is stated that can be tested using data and computation. It is specific enough that you can reasonably evaluate it within the time frame for the project. "I will rank all senators" is not a testable hypothesis – it is an activity that might result in some evidence that a hypothesis is true or false. Instead, a hypothesis is a statement one might suspect is true and can evaluate methodically. For instance, "Senators with democratic voting records are more likely to be called 'liberal' in the media than senators with more conservative voting records."

_____ (2 points) The hypothesis is placed in the **context** of a problem. Why is the hypothesis interesting to explore?

_____ (2 point) There is a brief **description of the data** to be used in the project, and the **data source** is specified, including a URL if the data is coming from the internet.

_____ (2 points) There is a brief description of the **format of the data** (e.g., file type and organization of file) and how it will be imported into a spreadsheet.

_____ (10 points) The **steps** of the analysis are numbered, specific, and manageable. "I will import the data and cluster according to votes" is not clear. "I will import all the data for all 2,012 congressional meetings" is not manageable. Be specific. Break your tasks into reasonable chunks.

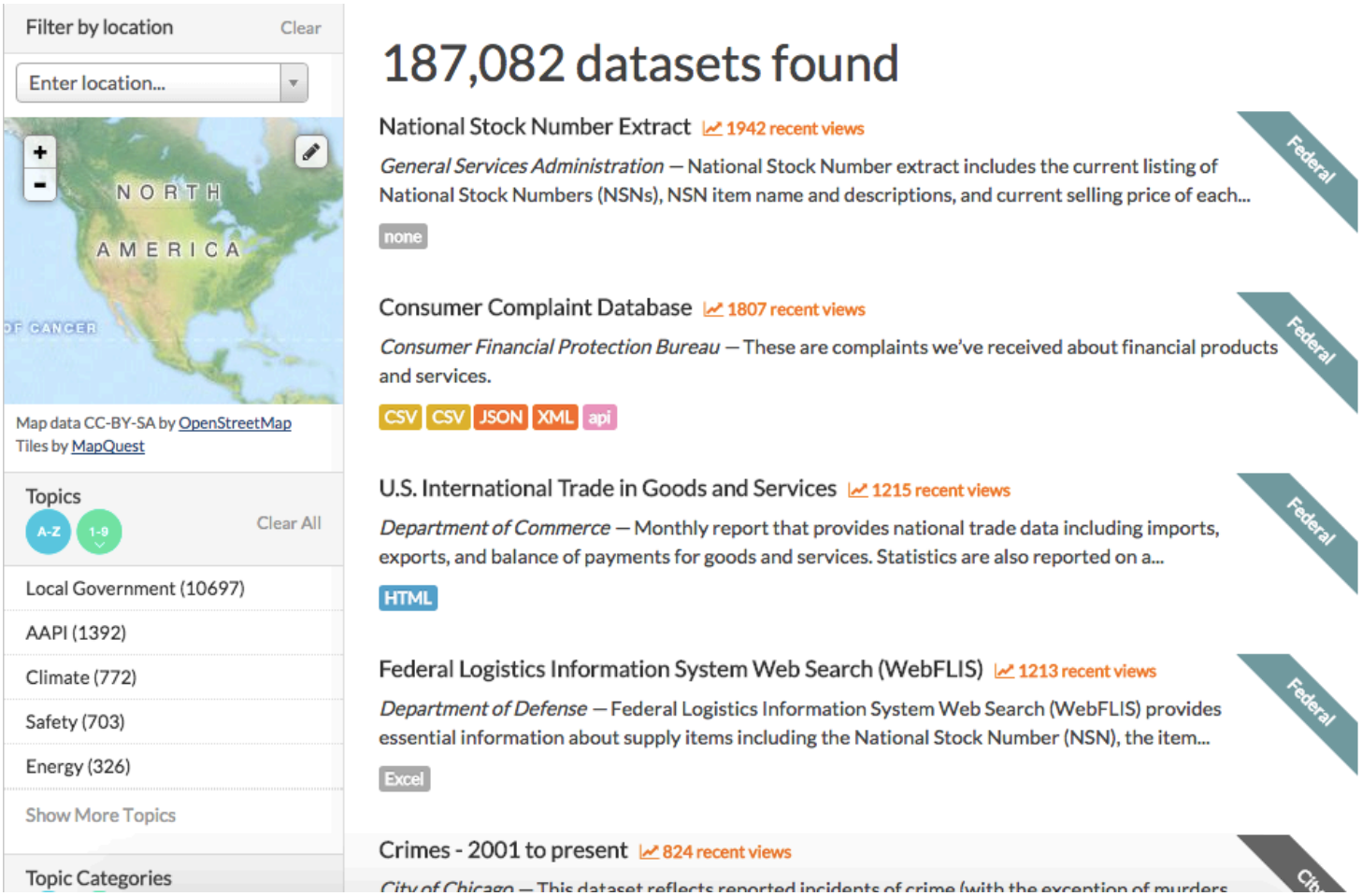
_____ (2 points) There is a description of how the hypothesis will be evaluated the **using the results**. What possible results would be confirming evidence for the hypothesis, and what would be disconfirming evidence for the hypothesis?

_____ (1 point) There is a description of a chart or visualization that will help **present the future results**. For example, "I will create a bar chart comparing these three averages." If no chart or visualization seems appropriate for presenting the results, there is an explanation why.

_____ (2 points) There are some **roadblocks** listed – what could go wrong with the steps you listed? For full credit, no obvious roadblocks are missed. Obvious roadblocks are things like "I want to perform a particular statistical test, but we haven't covered that formula in class and I don't know how to do it."

_____ Total

# http://catalog.data.gov/dataset



# OK, Last Class on Senator Ranking

- Learn some Google Spreadsheets tricks first

# Compare Every Pair in One Table

What We Have:

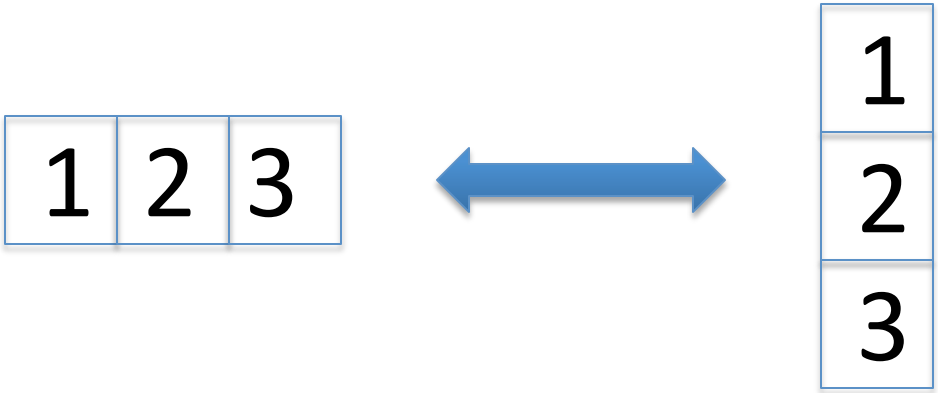
| | 1:1 | 1:2 | 1:3 | 1:4 |
|---|---|---|---|---|
| **Alexander** | Yea | Yea | Yea | Yea |
| **Ayotte** | Yea | Yea | Yea | Yea |
| **Baldwin** | Nay | Yea | Yea | Yea |
| **Barrasso** | Yea | Yea | Yea | Yea |

What We Want:

| | Alexander | Ayotte | Baldwin | Barraso |
|---|---|---|---|---|
| **Alexander** | 1 | 0.314285714 | 0.085714286 | 0.117647059 |
| **Ayotte** | 0.314285714 | 1 | 0.657142857 | 0.647058824 |
| **Baldwin** | 0.085714286 | 0.657142857 | 1 | 0.764705882 |
| **Barrasso** | 0.117647059 | 0.647058824 | 0.764705882 | 1 |

# TRANSPOSE

- Makes a “row” array into a “column” array or vice-versa



# SUMPRODUCT

Given two rows of the same length, multiply the corresponding cells and add up all these products.

# SUMPRODUCT

Given two rows of the same length, multiply the corresponding cells and add up all these products.

(doesn't only work with rows – the selections just have to be the same "shape")

# SUMPRODUCT example

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | 11 | 7 | 4 | | | |
| 2 | 0 | 2 | 1 | | | |

The formula =SUMPRODUCT(A1:C1, A2:C2) gives 18:

( 11 x 0 ) + ( 7 x 2 ) + ( 4 x 1 )

- Two ranges have to have same length...and both horizontal or both vertical
- SUMPRODUCT multiplies corresponding elements; sums up results

- Handy when one row contains prices, another contains quantities
  - SUMPRODUCT produces the total cost of buying things!

- You can use TRANSPOSE to get around the "both horizontal" restriction: see HW.

# Back to voting

We redefine

1 for a Yea
-1 for a Nay
0 for Not Voting

- The SUMPRODUCT of any 2 votes tells us how much they agree:
  +1 point (good ☺): both are 1 or -1
  -1 point (bad ☹): one is -1 and the other is 1
  0 points (neutral): at least one did not vote

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Alexander | 1 | 1 | 0 | -1 | … |
| 2 | Ayotte | -1 | 1 | 1 | 1 | … |

# Back to voting

- We've got this:

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Alexander | 1 | 1 | 0 | -1 | … |
| 2 | Ayotte | -1 | 1 | 1 | 1 | … |

- Formula to compute *agree - disagree* for Alexander and Ayotte?

```
=SUMPRODUCT(B1:AQ1, B2:AQ2)
```

- Formula to compute *agree + disagree?*

```
=SUMPRODUCT(ARRAYFORMULA(ABS(B1:AQ1)),
ARRAYFORMULA(ABS(B2:AQ2)))
```

# Our task:
# Do this for every pair of senators

- One strategy:
  - For each pair of senators
    - Form an agree/disagree list
    - Compute its sumproduct
    - Compute the sumproduct of its absolute values
    - Take the quotient

# Compare Every Pair in One Table

What We Have:

| | 1:1 | 1:2 | 1:3 | 1:4 |
|---|---|---|---|---|
| **Alexander** | Yea | Yea | Yea | Yea |
| **Ayotte** | Yea | Yea | Yea | Yea |
| **Baldwin** | Nay | Yea | Yea | Yea |
| **Barrasso** | Yea | Yea | Yea | Yea |

What We Want:

| | Alexander | Ayotte | Baldwin | Barraso |
|---|---|---|---|---|
| **Alexander** | 1 | 0.314285714 | 0.085714286 | 0.117647059 |
| **Ayotte** | 0.314285714 | 1 | 0.657142857 | 0.647058824 |
| **Baldwin** | 0.085714286 | 0.657142857 | 1 | 0.764705882 |
| **Barrasso** | 0.117647059 | 0.647058824 | 0.764705882 | 1 |

# Making labels

- Row labels are easy: copy from the re-coded vote tab
- Column labels????
  - Want to use the row labels
  - But we want to "fill" across instead of down

  - Let's open ACT 1-4

# Column Labels

- Solution 1:
  - In cell A3 of your table, put =TRANSPOSE(PivotTable!A3:A99)
  - Takes a block of cells and "flips it" over the NW-SE axis:

- Why is that built in?
  - Part of "matrix operations" that come up a lot

# Matrix Multiplication

**Grocery Store Prices (2 row and 4 cols)**

| | Bread | Milk | Doz. Eggs | Apples |
|---|---|---|---|---|
| **PriceRight** | $3.09 | $2.90 | $1.85 | $1.15 |
| **Stop'n'Shop** | $3.49 | $3.89 | $2.05 | $0.79 |

**Grocery Lists (4 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **Bread** | 2 | 2 | 1 |
| **Milk** | 1 | 0 | 2 |
| **Doz. Eggs** | 2 | 1 | 1 |
| **Apples** | 2 | 4 | 3 |

SUMPRODUCT(row 1, column 1)

**Food Bills (2 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **PriceRight** | 15,08 | | |
| **Stop'n'Shop** | | | |

# Matrix Multiplication

**Grocery Store Prices (2 row and 4 cols)**

| | Bread | Milk | Doz. Eggs | Apples |
|---|---|---|---|---|
| **PriceRight** | $3.09 | $2.90 | $1.85 | $1.15 |
| **Stop'n'Shop** | $3.49 | $3.89 | $2.05 | $0.79 |

**Grocery Lists (4 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **Bread** | 2 | 2 | 1 |
| **Milk** | 1 | 0 | 2 |
| **Doz. Eggs** | 2 | 1 | 1 |
| **Apples** | 2 | 4 | 3 |

SUMPRODUCT(row 1, column 2)

**Food Bills (2 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **PriceRight** | 15,08 | 12.63 | |
| **Stop'n'Shop** | | | |

# Matrix Multiplication

**Grocery Store Prices (2 row and 4 cols)**

| | Bread | Milk | Doz. Eggs | Apples |
|---|---|---|---|---|
| **PriceRight** | $3.09 | $2.90 | $1.85 | $1.15 |
| **Stop'n'Shop** | $3.49 | $3.89 | $2.05 | $0.79 |

**Grocery Lists (4 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **Bread** | 2 | 2 | 1 |
| **Milk** | 1 | 0 | 2 |
| **Doz. Eggs** | 2 | 1 | 1 |
| **Apples** | 2 | 4 | 3 |

SUMPRODUCT(row 2, column 1)

**Food Bills (2 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **PriceRight** | 15.08 | 12.63 | 14.19 |
| **Stop'n'Shop** | 16.55 | | |

# Matrix Multiplication

**Grocery Store Prices (2 row and 4 cols)**

| | Bread | Milk | Doz. Eggs | Apples |
|---|---|---|---|---|
| **PriceRight** | $3.09 | $2.90 | $1.85 | $1.15 |
| **Stop'n'Shop** | $3.49 | $3.89 | $2.05 | $0.79 |

**Grocery Lists (4 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **Bread** | 2 | 2 | 1 |
| **Milk** | 1 | 0 | 2 |
| **Doz. Eggs** | 2 | 1 | 1 |
| **Apples** | 2 | 4 | 3 |

MMULT(GREEN,ORANGE)

**Food Bills (2 rows and 3 cols):**

| | Bob | Carol | Dina |
|---|---|---|---|
| **PriceRight** | 15.08 | 12.63 | 14.19 |
| **Stop'n'Shop** | 16.55 | 12.19 | 15.69 |

# How can this help with our Senators?

MMULT

| | 1:1 | 1:2 | 1:3 | 1:4 |
|---|---|---|---|---|
| Alexander | 1 | 1 | 1 | 1 |
| Ayotte | 1 | -1 | -1 | -1 |
| Baldwin | 1 | 1 | 0 | 1 |
| Barrasso | 1 | 0 | 1 | -1 |

and

TRANSPOSE

| | Alexander | Ayotte | Baldwin | Barrasso |
|---|---|---|---|---|
| 1:1 | 1 | 1 | 1 | 1 |
| 1:2 | 1 | -1 | 1 | 0 |
| 1:3 | 1 | -1 | 0 | 1 |
| 1:4 | 1 | -1 | 1 | -1 |

RESULT: Agreements - Disagreements

| | Alexander | Ayotte | Baldwin | Barrasso |
|---|---|---|---|---|
| Alexander | 4 | -2 | 3 | 1 |
| Ayotte | -2 | 4 | -1 | 1 |
| Baldwin | 3 | -1 | 3 | 0 |
| Barrasso | 1 | 1 | 0 | 3 |

# How can this help with our Senators?

MMULT

ABS

| | 1:1 | 1:2 | 1:3 | 1:4 |
|---|---|---|---|---|
| Alexander | 1 | 1 | 1 | 1 |
| Ayotte | 1 | 1 | 1 | 1 |
| Baldwin | 1 | 1 | 0 | 1 |
| Barrasso | 1 | 0 | 1 | 1 |

and

TRANSPOSE of ABS

| | Alexander | Ayotte | Baldwin | Barrasso |
|---|---|---|---|---|
| 1:1 | 1 | 1 | 1 | 1 |
| 1:2 | 1 | 1 | 1 | 0 |
| 1:3 | 1 | 1 | 0 | 1 |
| 1:4 | 1 | 1 | 1 | 1 |

RESULT: Agreements + Disagreements

| | Alexander | Ayotte | Baldwin | Barrasso |
|---|---|---|---|---|
| Alexander | 4 | 4 | 3 | 3 |
| Ayotte | 4 | 4 | 3 | 3 |
| Baldwin | 3 | 3 | 3 | 2 |
| Barrasso | 3 | 3 | 2 | 3 |

# ACT 1-4

- Work with your neighbor
- Ask lots of questions
- Let us know when you're done with the 4-senator version

# What can we do with similarity table?

- Look for very similar people
  - Use color to help!
- Reorder the list to move similar people near each other
- If A is similar to B, and B similar to C...
  - Is A similar to C?
- Results

# Discoveries

- Two major blocks
  - Democrats
  - Republicans
- Some oddballs
  - Democrat In Name Only (DINO)
  - ... (RINO)
- Substructure
  - Blue Dog Democrats
  - Mountain-state republicans?

# Conclusion

- What matters isn't liberal vs conservative...
- It's the network structure in the senate!
  - Who is in what group, what subgroup, etc.
  - Who are connectors between groups?
- Computation enabled this insight.