Ranking Senators w/ Bernie Sanders’ Votes
Disclosure

• Brown requires that I tell you this:
  – I have investments in various companies whose primary business is related to the material in this course.
  – Brown thinks that this might lead me to present biased information to you...
    • as if what I say would ever make you buy a Google or Microsoft product, or the results of this course could influence those companies’ policies!
  – In particular: MSFT, GOOG, IBM, other companies in the office automation/computer industry
We’ve done this

Define Problem

Find Data

Write a set of instructions

Solution

XML Format

CSV Format

Make a HUGE spreadsheet table

Use Bernie Sanders’ votes to compare how liberal other senators are

Vote on bills only!
Last Class: Completed So Far

1. Open the spreadsheet from last class.
2. You should have 3,501 rows (Task 3.7)
3. You should have columns through E.
Next Task: enhancing data

• We want a unique identifier for the vote of each bill in this congress.
  – Which two columns together make a unique key?
• Add another column to the table by entering a \texttt{vote\_id} column in cell F1.
• Write a formula to output \texttt{session:vote\_number} values for this row.
• Use fill down or copy/paste, if necessary, to apply this formula to all the other rows.
Encoding data numerically: easier to process, esp. in spreadsheets

• Add a `numerical_vote` column in cell G1.
• Write a formula to output:
  – 0 if the senator did not vote or was absent
  – 1 if the senator voted Nay
  – 2 if the senator voted Yea
  – Error otherwise
vote_cast = “Not Voting”

0 if the senator did not vote
1 if the senator voted Nay
2 if the senator voted Yea
Error otherwise
vote_cast = “Absent”

0 if the senator did not vote or was absent
1 if the senator voted Nay
2 if the senator voted Yea
Error otherwise
Summarize data

• For each senator
  – For each vote
    • What was vote?
• Organize in a table whose rows are senators and columns are vote-names
  – Row-column entry is a senator’s vote
  – Row = all votes by senator
  – Column = all votes on this bill
• Do this with a pivot table
What is a pivot table good for?
Pivot table uses

- Restructure data (our use)
- Summarize data (more common use)
  - If our data was
    - salesman, customer, itemSold, itemCount
  - A “salesperson x itemSold” table could show the sum of all sales of this item, or the max sale of this item, etc.
We did these things!

- Find out what CSV is
- XML Format → CSV Format → Make HUGE Table
- Import the data we want into spreadsheet
- Format the table to get what we want
See anything odd in the pivot table?

• Use Home...Conditional Formatting to color all blank cells red
See anything odd in the pivot table?

• **Use** Home...Conditional Formatting to color all blank cells **red**

• What happened to Schatz and Ensign?
See anything odd in the pivot table?

- **Use** Home...Conditional Formatting to color all blank cells **red**

- **Name** csv_output RawData
- **Name** Sheet2 PivotTable
- **Make column widths nice:**
  - Home-> Cells-> Format->AutoFit Column Width
Activity 1-2

• Do Tasks 1 and 2
Are we done?

- We’ve got the similarity between any senator’s voting record and Bernie’s record...
- Isn’t that what we were hoping to produce?
- Who is the least Bernie-like?
- What if we compared to *that* person instead?
- Homework suggests that there could be surprises...
Project 1

• 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
Grading Rubric

• Proposal
  – Clarity
  – Forethought

• Design

• Execution
  – Did you do it right? Handle bad data?

• Website Presentation, 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 number of vowels in the name; I’ll sort by number of vowels…”
  – 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).
Note re data sources

- Lots of data now available in another form: JSON
- JSON data can also be imported into google spreadsheets
- Perhaps not quite as easy as you’d like...
Tonight’s HW

• A little longer than typical Tuesday HW – start early