Capstone

Project Proposal

Jonas Joehnk

Project 1 - Voting Behavior

Goal

Look into the relationship between major sources of funding from special interest groups and voting behaviors in Congress.

Datasets

- Federal Election Commission API (millions of possible rows)
- Voteview flat files on roll calls (largest of three - 586665 x 6)
- DIME PLUS flat file on legislation analysis (3182 x 33)

Approach

- Focus on data collection funnel of big data into summarized data for analysis
- Testing for variances among voting behavior
  - Do congresspeople funded by the same interest groups act the same? As expected by single interest group?

TOPIC:

Relationship between funding sources and voting patterns in Congress

API:
Federal Election Commission

FLAT DATA SOURCES:
- Voteview.com --> Roll Call votes
- Standard DIME PLUS repository
--> bill categorization

OTHER SOURCES:
Additional context, data may be added and planned for as research continues

.py

storage:
SQLite

Tableau --> Focus on one interactive dashboard

GitHub Repository

Limitations:

1. Complex topic that would not cover necessary additional variables (public opinion, party alignment/whips, personal ideology, complexity of the bills - just to name a few)

--> as such, goal is to build tool to view the relationship between these two elements for theoretical further analysis and NOT to come to any generalized conclusions.

2. Obfuscation of funding - with the restrictions placed on direct funding of politicians, Super PACs and other interest groups operate separately from the individuals they support.

--> will need to be careful in construction of pipeline to consider these relationships, and how the FEC tracks PAC spending and connect this to congress

3. Massive dataset: with the time constraint, only a small selection of a specific congress and certain special interests will be examined and analyzed.

--> goal is for backend to be documented in a way for anyone to pick up and apply for different funding/interest groups, in a different time period

Assumptions:

1. Complex legislation will need to be summarized down to digestible topics of focus and as 'yay' or 'nay' that aligns with funding sources' interests

2. More to come with data exploration

Goals and objs:

To create a tool that displays funding sources of politicians and how that relates to voting positions

(but cannot be used for sweeping generalizations of motive due to complexity! Only digestible summary of information and potential correlations)

Focused Goal:

Test of Variances among voting behavior by funding committees (and additional groupings)

Project 2 - US Wages and Unions

Goals

Look into the relationship between US wages by industry and occupation, and the influence of union membership and strike initiatives.

Datasets

- Bureau of Labor Statistics API (millions of possible rows)
- Strike actions (by year) flat file (35000 x 15)
- Union membership (by year) flat file (7,500 x 5)

Approach

- Time series analysis
  - Comparing industries with strong union presence or notable action to those without

- Additional focus on data collection of large dataset from API into summarizable data

Topic:

Influence of Union Membership and Collective Bargaining on Wages over Time

API:
Bureau of Labor Statistics on Wage by Industry

Flat Files:
- Strikes
- Union Membership

Additional Sources as needed

.py

storage:
SQLite

Tableau --> storyboard

GitHub Repository

Limitations:
1. Massive dataset --> loss of nuance and external variables that change wages by industry over time

2. Data from BLS already aggregated by industry and can't speak to differences within one occupation group

Assumptions:
1. BLS data collected from self-reported surveys, assuming that reports are accurate to actual wages

Goals and objs:

Determine potential correlations between union membership and collective bargaining on US wages

Focused Goal:

Time Series Analysis of Industry and Union Activity Levels on Dependent Variable

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution