Exploring Money's Influence on the US Congress

Matt Brock

SEA DAT1

Motivation and Objective

- Positions on campaign finance and money in politics tend to argue from strongly held beliefs, not data
  - "It's free speech!!!!!!!" vs. "it's corruption!!!!!"
- There is a ton of data out there – let's use it to make our institutions better, instead of arguing from dogma

Objective: Find measurable relationship(s) between lobbyist donations and behavior of US congressional representatives

What an answer might look like

- Ideally, a statistically valid modelable relationship between $ in and votes out, or between # of contacts and voting behavior, at a fine grained level
  - E.g., An X unit increase in donations to Representative Y makes that representative more likely to vote in favor of an issue
- Next best: any sort statistically valid of macro relationship between total $ (or total activity) and outcomes

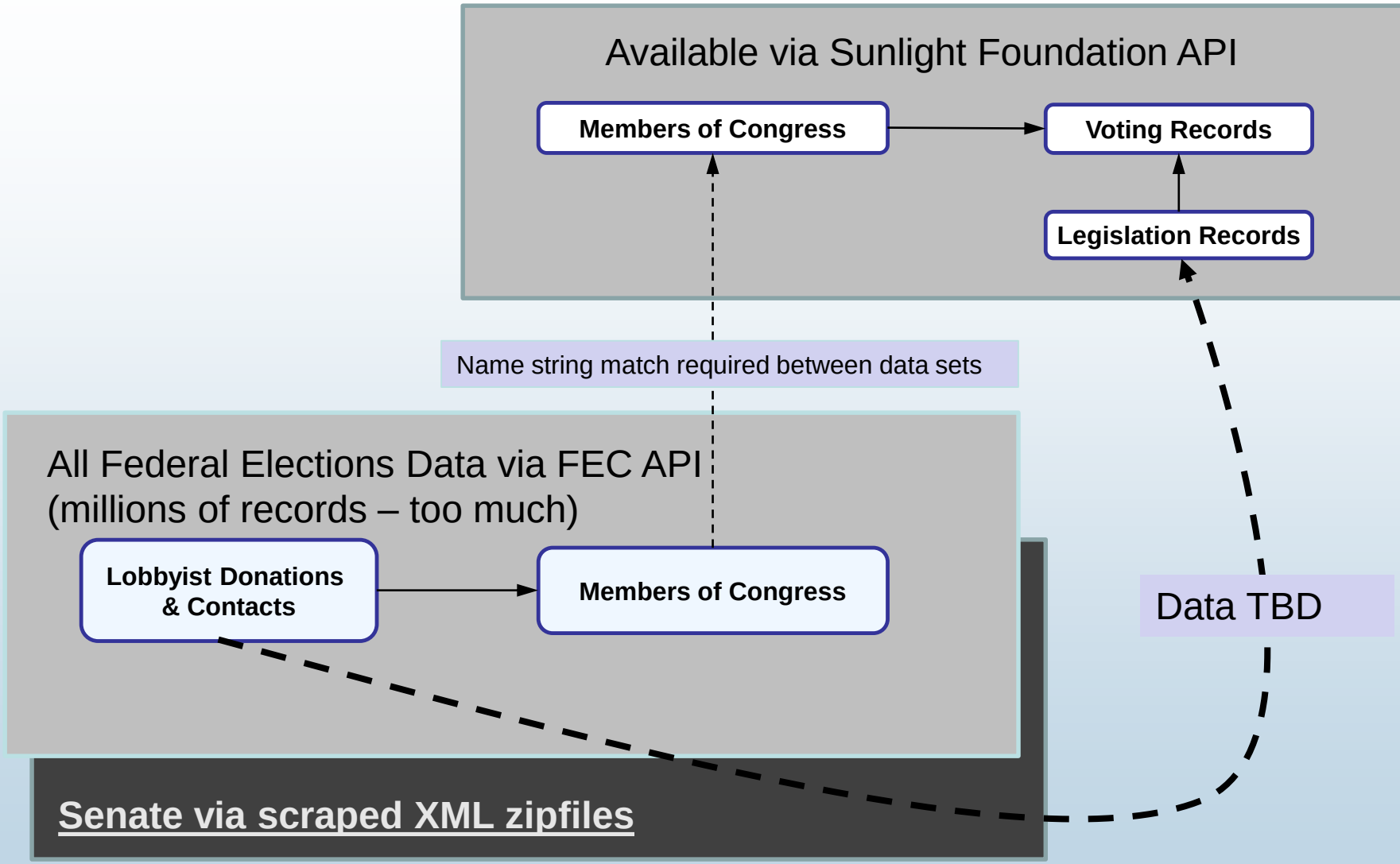
Progress so far

- Surveyed most of the data landscape
- Pulled in tables or large subsets of entity and donation data done basic index cleaning, and explored relationships using Excel
- Lost a lot of time on dead ends, capturing and pulling the wrong data, at scale

Data Sources

- Contributions and contacts
  - Federal Elections Commission API (beta.fec.gov)
    - Data on every federal campaign, going back decades. Tens of thousands of campaigns, millions of financial transaction records, across all donor categories. Heavily nested JSON with complicated pagination.
  - Senate records via website scraping
    - Data back to 2000, via several thousand XML files archived in zip fles
- Votes – congressional record back to 1800, available via Sunlight Foundation
- Still TBD – something machine-readable linking lobbyists to specific issues or bills

Conceptual Map of the Data and Sources



Next Steps

- Find the missing lobbyist-to-issue data and integrate it
- Move beyond data janitor mode into visualization/modeling
  - Unpack Senate XML files and build data frames
  - Pull in Congressional votes data, build data frames, match via name string to senate records
  - Will likely need a mix of supervised and unsupervised learning to discover and model relationships
- Backup plan: Change tack and simplify scope
  - Look at political behavior by political party subgroups using just the Congressional Votes data (Tea Party R's vs. Establishment R's; Progressive D's vs. Establishment D's)