Team Name: let teamName = “The League of Extraordinary Gentlemen” + “Danni”;
Members: Danni Li, Brian Sherrill, Drew Dickens, Joshua Veden

PM6: Conclusion

Value Proposition

The election process in America can often be convoluted and confusing, making it difficult for the average citizen to understand which candidate is the right choice. Many data points exist to compare candidates; however, campaign finance and lobby data points serve as tangible evidence of a candidate’s efficacy across the campaign trail, providing a logical data points which can be used to infer their potential efficacy in office.

Our tool makes this information accessible through a convenient web application. The value in this tool is clear; it provides useable data insights to be used when researching political candidates, leading up to an election.

Deliverables

Our initial deliverable set consisted of 2-3 in-app data visualization methods with various export formats for those data sets. Potential extensions for the app consisted of the following: map overlay functionality, reporting APIs, feedback sharing, and larger data sets. Over the course of the semester, we were able to develop a prototype that supports 8 different data visualizations; four for lobbying and four for campaign finance. Building out the entire development stack proved to be more time consuming than initially anticipated and the export formats were cut from the application.

UML Diagram

Campaign Finance

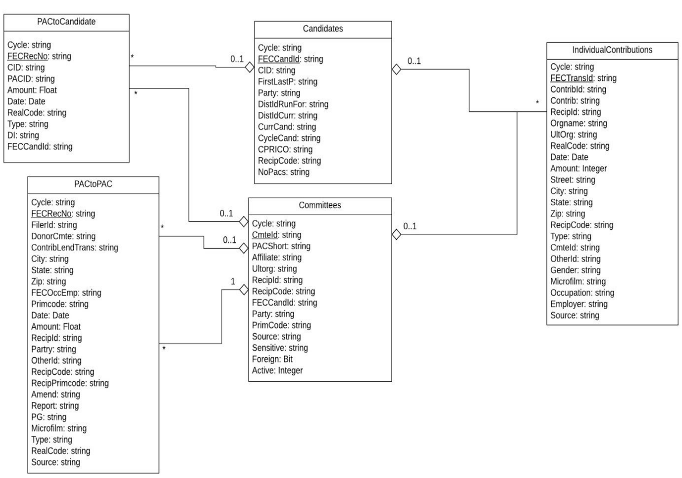

Diagram Changes

This is the final UML for our Campaign Finance model. Originally, when we exported the data from OpenSecrets.org website, we had 7 tables including PACtoCandidate, PACtoPAC, Candidates, Committees, IndividualContributions and Expenditures. But during Project Millstone 2, we decided to take out the Expenditure table due to the incompleteness of the Expenditure data, according to the data documentation on OpenSecrets.org.

Team Name: let teamName = “The League of Extraordinary Gentlemen” + “Danni”;
Members: Danni Li, Brian Sherrill, Drew Dickens, Joshua Veden

Lobbying

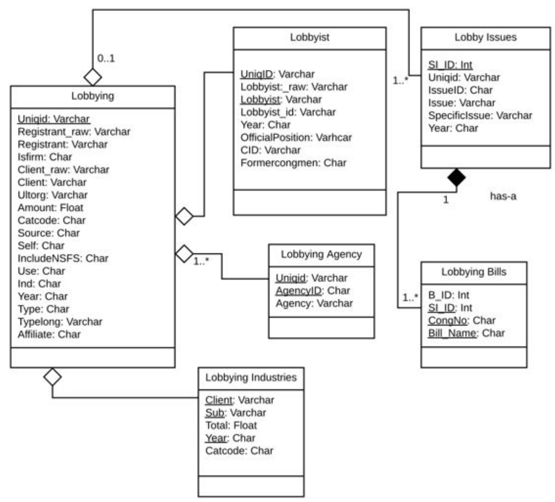

Diagram Changes

This is the final UML for our Lobbying model. We originally have 7 tables including Lobbying, Lobbyist, Lobbying Agency, Lobbying Industries, Lobby Issues, Lobbying Bills and Report Type. As we work with the data set and develop our product, we did not utilize the Report Type data, so we took out Report Types table in Project Millstone 6 and finalized our Lobbying model to include 6 tables shown above.

Reflection

Nearly any data oriented question one could have about campaign financing or bill lobbying, we could answer. For our website, we chose interesting questions and displayed them in a manner that is easy and appealing to the user. There are more questions we would have liked to answer, so If we were to do something different, we would have picked more modern technology to speed up development. While our quieres run relatively quick, they are not responsive enough for production website. We could benefit from a caching, e.g. running a caching worker and then retrieve all data from the cache.

The data itself also posed a problem. The data was made without good data practices. The data set did not make data definitions clear; we had to figure out much of the data connections. In fact, our UML gives a much better description that is available. Their implementation often used a single column as a foreign key, referencing different tables based on formatting. While importing data, often rows were improperly formatted causing us to lose data. Many rows were or unique identifiers were not unique. We lost a lot of data, and we can’t be sure if we lost important information. All these data difficulties taught us why good data practices are important and how to deal with bad data. The size of the dataset forced us to use proper indexing. If were were to do do this again, we would devote more time into preparing and fixing up our data. Alternatively, a better dataset could be used, but we did not find one.

Data warehousing proved Interesting for us. The most obvious questions to ask are how real world events effects our data, i.e. “How does the state of the economy change what gets lobbied”. These are hard to answer because there are so many confounding variables and not easily found data. Also, to answer interesting data warehousing questions involves a more complex analysis than we can do. If we were to do this again, we might try and either choose a more direct cause and effect question, or build a larger dataset and run analysis on the results.