

Leo Le, Elizabeth Willard, Lucy Herr

Github repository: <https://github.com/elizabethwillard/mids-200-project-2.git>

Project 2 Proposal

Summary:

The annual budget of the members of the House of Representatives for their Washington and district offices are public record. This data details all quarterly official spending, including lawmakers' offices, committee offices and administrative offices and through these insights, we can learn a lot about the spending habits of the government members, whether along partisan lines or in changes over time. We are particularly interested in the period between 2016 and 2018 when former president Donald J. Trump was elected to office as this was the time period when many changes occurred.

Using the publicly available "[House Office Expenditures](#)" dataset collected and published by Propublica, our analysis will address the following key questions about House expenditures between 2016 and 2018:

1. Are there any identifiable patterns in amounts spent on different expenditure categories and periods (quarter and/or year)?
 - a. Depending on the answers to the above questions, which categories yield the most interesting trends or patterns?
 - b. In particular, how does person spending vary between offices over time? Is there a relationship between spending on personal and office characteristics (such as turnover)?
2. What are the differences in overall spending breakdowns (by category and overall amounts) between Republican and Democratic house members between 2016 and 2018?
 - a. How do these trends vary on a quarterly and/or annual basis?
 - b. Depending on the answers to the above questions, which categories yield the most interesting trends or patterns in terms of partisan spending differences?

Primary Dataset:

The primary data source is "[House Office Expenditures](#)", which consists of summary and detailed expenditure datasets broken down by quarter and year. Examples of each type can be found here: [summary](#), [detail](#).

Some of the critical columns to examine in our analysis are:

- 'Bioguide_ID' – the official ID of members of the House (to be joined with supplementary dataset for member details)
- 'Description' - Expenditure categories
- 'Purpose' - Details of the categories expenditure
- 'Amount' - amount spent by office in that category and quarter

Supplementary Dataset: [Member of the House IDs and party affiliation](#)

This supplementary dataset will help us determine the Representatives' party affiliation in order to conduct comparison between spending at the party level.

Approach:

To prepare the data for analysis, we will first filter out any datasets from the comprehensive zip file that fall outside of the 3-year period of focus (2016-2018). Next, we will concatenate the quarterly "summary" datasets for the 3-year period (12 total). We will begin the preliminary analysis with the "summary" dataset types as opposed to the "detail" versions that contain itemized expenditure breakdowns (although we also plan to use these when applicable to explore interesting findings from the summary-level analysis).

Data cleaning will involve changing the data types of some columns (such as the "amount" column, which is currently of type object as it contains some string values). We will also examine and address any nulls, as early exploration of the dataset indicated some missing values in the "BIOGUIDE ID" (House member ID) column. That said, this may not in fact be due to error (e.g., it may result from observations that describe groups/committees rather than individuals). We are not currently expecting to drop any of the columns in the summary, although we will likely add columns containing additional information through joining with the supplementary dataset and/or detailed dataset versions.

One of the main relationships we are concentrating on in this proposal is between the categories and their corresponding spending amounts. We are approaching this analysis in a top-to-bottom method, where the first task is to identify categories that demonstrate big shifts and then dive deeper into these categories in the detailed spending report. Table 1 demonstrates the comparison between 2016 Q4 amount and 2017 Q1 amount spent. We can see there is a substantial change in the Personnel Compensation and Personal Benefits categories, which both attribute to the employee's salaries and their benefits.

2016 Q4 vs 2017 Q1 expenditure com

Category	2016 Q4 amount \$	2017 Q1 amount \$
Personnel Compensation	182,868,348.26	159,694,362.25
Personal Benefits	18,520,727.04	60,491,616.62
Rent Communication Utilities	13,963,332.27	13,728,321.33
Other Services	7,702,485.39	10,588,930.78
Equipment	-	9,372,522.37
Supplies and Materials	7,410,383.92	-

