2016-2017 House of Representatives Spending Analysis

Team: Leo Le, Elizabeth Willard, Lucy Herr

W200 Introduction to Data Science Programming: Project 2

Project Github repository: https://github.com/elizabethwillard/mids-200-project-2

Introduction

In late 2016, Donald Trump's election to the U.S. presidency spurred a sea change throughout the federal U.S. government, with impacts ranging from partisan norms to federal hiring. This election also translated to shifts in the U.S. House of Representatives, where Republicans lost several seats but retained their majority status. In addition, 42 incumbent Representatives declined to run during the 2016 election, ushering in new representatives with potentially different approaches to administration of their offices.

As one of two legislative bodies in the federal U.S. government, members of the House of Representatives receive an annual office budget to spend at their discretion (with restrictions on campaign or personal spending). Members of the House are required by law to report these expenditures, which Propublica has collected and published. This data details all quarterly official spending by individual Representative, committee, and administrative offices at both the summary and detailed level. By exploring the spending habits of House of Representatives members between 2016 and 2017, we can uncover insights into how House spending habits may or may not have varied over this period for major offices and spending categories, personnel practices, and partisan lines.

Research Questions

Using Propublica's 2016-2017 "House Office Expenditures" data, our analysis will address the following key questions about House expenditures during this period:

- Are there any identifiable patterns in amounts spent on different expenditure categories and periods (quarter and/or year)?
    - In particular, does personnel spending vary between offices over time, and if so, how? Is there a relationship between spending on personnel and office characteristics (such as turnover)?
- How do Republican and Democratic House members compare or differ in their spending during this time period?
    - Does total and/or average spending vary over time across the election period?
    - Which, if any, categories indicate partisan spending differences?

Data

1. Primary Dataset: 2016-1017 "House Office Expenditures"

This source includes summary and detailed office expenditure data broken down by quarter and year.

1.1: Summary Expenditures Data

In the summary expenditures data, we focused on the following variables in our analysis of Office and Party spending:

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

The 'Bioguide_ID' indicates which office or representative member filed for the expenses. This is key to find which political affiliation each of the member. Each expense is categorized in the 'Description' column with their corresponding amount value listed in the 'Amount' column.

1.2: Detailed Expenditures Data

In the detailed expenditures data, we focused on the following variables in our analysis of Personnel spending:

- 'Office' - House of Representatives offices
- 'Category' - Expenditure categories
- 'Purpose' - Details of the categories expenditure
- 'Payee' - Recipient of the payment
- 'Amount' - amount spent by office in that category

The 'Office' column listed the name of offices and Members. Their expenses are broken down and standardized in the 'Category' column with specific note details written in the 'Purpose' column. The 'Payee' column indicates a single person's name of which the expense is delegated to. Finally, the 'Amount' column indicates the dollar amount of the expense.

2. Supplementary Dataset.

In addition, we used data from the Biographical Directory of the United States Congresses (https://bioguide.congress.gov/) for U.S House Representatives between 2015 and 2019. This source consisted of 2000 unique Representative observations, with 14 original columns. While this source is described in more depth in the party analysis section, the variables of interest consisted of two original columns:

- 'Bioguide_id': the unique identifier for individual Representatives (aligns with column of same name in the expenditures datasets)
- 'congresses': json-type dictionary field including:
    - 'parties': Republican, Democrat, or independent/other
    - 'stateName': used for sanity checking and Representative categorization

Initial Data Exploration and Cleaning

The data was in a messy state, which was acknowledged in the ProPublica documentation: "The data has not been standardized, so simple aggregation on a recipient could result in multiple totals for the same individual or entity." With this note in mind, the data cleaning process had to begin by removing erroneous duplicate entries. Our process started with the summary files first.

For the “AMOUNTS” category within the summary files, they were typecast as strings, not values. These had to be cleaned of extraneous symbols like commas and then typecast into float64 values. Float64 was selected in order to assure that none of the values were rounded.

There was inconsistent naming within the “QUARTER” column as well, so this column was scrubbed to just the quarter value and then a new column was created that combined this quarterly value with the corresponding year value.

The offices with incomplete quarterly data (i.e., only 2016 Q3 and Q4) were dropped in the final aggregations. When computing the aggregations, there was another compensation category that listed the final total values, and these rows were dropped so as to not skew the total amounts of each office. This decision was also supported by the data documentation from Propublica, which acknowledged discrepancies between the listed totals and totals for the individual expenses.

Finally, before aggregating the data, assume that different House offices classify expenses in different ways. In order to accurately reflect all the unique categories and their corresponding expenses, we needed to account for overlapping words with the same meaning, and grammar errors such as incorrect spellings or punctuation. For example, within the “category” variable, there could be duplicated entries of separate expenses and then the combined expenses indicated by “expense totals:”. Rows containing the totaled amounts were filtered from the data set.

It is also important to note that aggregations of quarterly data could be understated because disbursements can be reported as late as three years later from when the disbursement was made, according to the guidance from ProPublica.

With the detailed data, ProPublica stated that “summary numbers may understate how much a member’s office spent in a quarter or a year… To look at specific expenditures–travel, salaries for individual congressional staffers and so on, the detailed files should be used.” This advice paved the data cleaning process for the detailed data files.

Analysis 1: Spending Categories Analysis

Some of the offices do not have enough data to define any trend patterns from as mentioned during the data cleaning process. The final amount of offices that a trend could be discerned from was 80 offices in total. For roughly a fifth of those offices, the highest amount of spending occurs during Quarter 4 of 2017.

From those 80 offices, the top ten highest spending offices were pulled and their largest spending category, which are tabulated below.

Table 1.1: Top Ten Highest Spending Offices and their Largest Spending Category

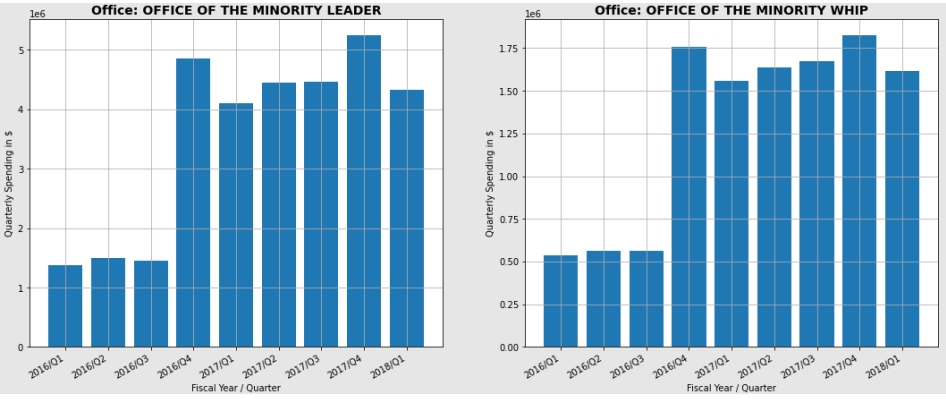
Office	Spending Category	Amount (in USD $)
Government Contributions	Personnel Benefits	315,838,400
Chief Admin Ofcr Of the House	Personnel Compensation	92,579,099.14

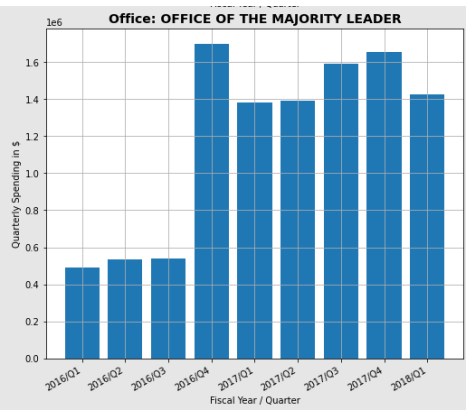
Office	Spending Category	Amount (in USD $)
Clerk of the House	Personnel Compensation	29,277,986.25
Committee on Appropriations	Personnel Compensation	28,341,849.89
Sergeant at Arms	Personnel Compensation	13,845,767.77
Committee on Energy & Commerce	Personnel Compensation	14,539,104.62
Joint Committee on Taxation	Personnel Compensation	12,700,998.37
Net Expenses Telecommunication	Office Expenditures (i.e., supplies, materials, rent, communication utilities)	7,691,841.23
Legislative Counsel	Personnel Compensation	12,411,011.83
Committee on Ways and Means	Personnel Compensation	11391,702.54

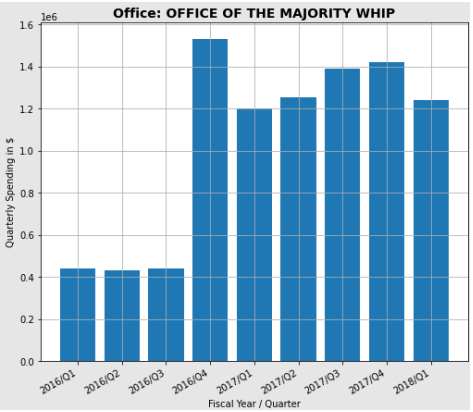
The Chief Administrative Officer of the House, Clerk of the House, Sergeant at Arms, Net Expenses Telecommunication, Legislative Counsel are all related to administrative work. The Joint Committee on Taxation and Committee on Ways and Means are both related to taxation. Therefore, it would not be unfair to say that the highest spending from offices is related to administrative and taxation-policy processes.

Another interesting comparison was between the Minority and Majority Offices, the spending of the Minority office surpassed that of the Majority Office. In the figures below, the minority leader reaches spending of five million dollars in one quarter, whereas the office of the majority leader exceeds 1.6 million dollars in one quarter. The office of majority and minority whip are closer, but the minority whip spending still exceeds that of the majority whip.

Figure 1.1: Spending amount of Minority vs Majority Offices







Analysis 2: Comparison of Expenditures by Party

The primary goal of this stage of the analysis was to evaluate whether spending trends differed for Republican and Democrat House Representatives based on time period and/or category. In order to compare Representative spending by party membership, I used supplementary data from the Biographical Directory of the United States to obtain party information for each Representative. This data source was limited to a search of House Representatives in the 2015-2017 and 2017-2019 terms and downloaded as a json file. My preliminary "sanity checks" of the data included confirming that each 'Bioguide_ID' was linked to a unique Representative and that the majority of these IDs were included in the cleaned summary dataset (with a few discrepancies to address later).

Without any missing values in the key columns ('Bioguide_ID' and 'congresses', which contained party and state information for each Representative), the main cleaning step consisted of parsing the "congresses" column into separate columns containing party and state information for each member. As the original format of the column values was a dictionary list (with a separate dictionary for each term served), I took the last entry in each dictionary and separated "party" and "state" into new columns. These columns were then merged with the summary expenditures data using Bioguide ID as the key to join them.

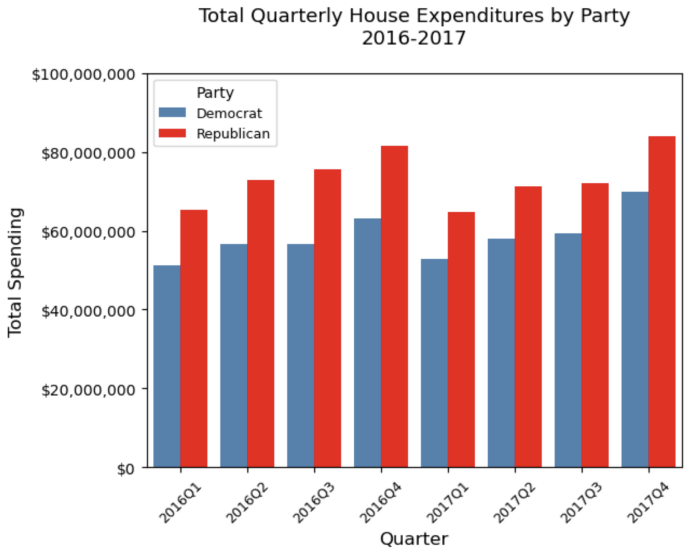
After joining, I examined the remaining observations with missing Bioguide IDs. Of these, the vast majority were for administrative offices or committees that could not be linked to specific party membership and were therefore excluded from this part of the analysis. I also made the decision to manually add party information for individual House leadership offices (e.g., Speaker of the House or Minority Whip) based on the assumption that, while these offices have higher budgets compared to non-leadership, excluding them would omit a major component of partisan spending. It's also important to note that I did not exclude expenditures based on incomplete quarterly reporting because I wanted to get as close as possible to the actual totals; however, this may introduce accuracy or bias into the results.

Ten remaining individual Reps had missing IDs to check. Based on my background research, three of these were individual Representatives during the correct terms who simply had missing information in this dataset, which I then found and imputed. Based on earlier rounds of data cleaning, one possibility for these discrepancies is the fact that a handful of Representatives

resigned mid-term due to sexual harrassment investigations, leading to discrepancies how Represenatives were listed in the Bioguide source. The remaining seven House members represented the U.S. Territories (e.g., Puerto Rico) or Districts at Large (e.g., American Samoa or D.C.). As a sanity check, I confirmed that six appointed "District at Large" House members were present in the data each year. I ultimately made the decision to remove these expenses from the data based on the fact that these House members have no voting rights and more limited roles and budgets in the House compared to state Representatives.

To understand possible differences in party spending over time, I first calculated quarterly spending totals by party over the two-year time period by summing all summary expenditures for Representatives for either party.

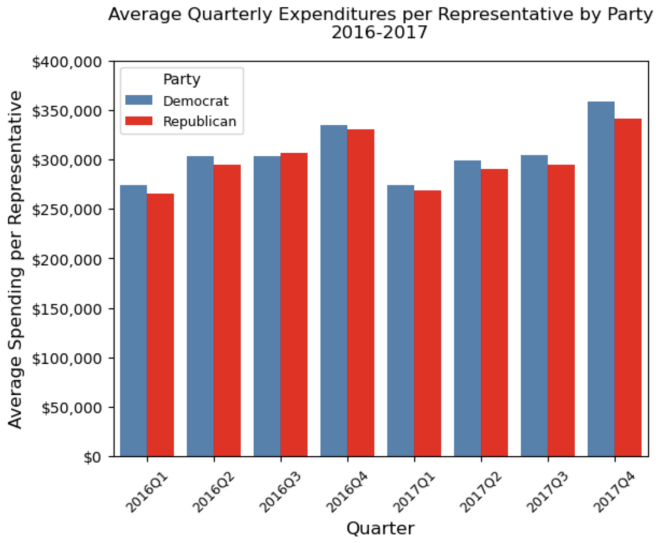
Figure 2.1: Total Quarterly Spending per Representative by Party



As Figure 2.1 demonstrates, the quarterly Republican totals consistently exceeded those for the Democrats by roughly $10 million or greater. However, it does not appear that the administrative shift after Trump's election translated to substantial differences in pre- and post-election spending totals by party. Determining the true explanations for this trend is beyond the scope of this project, but one possibility is that the actual budgets and party composition of the House stayed relatively consistent between the two years. More generally, the totals for either party ranged from approximately $50 million to $80 million per quarter, and quarterly spending for either party increased every quarter from the beginning to the end of each year.

In both years, Republicans held a House majority, with 250 Republicans to 188 Democrats in 2016 and 241 to 194, respectively, in 2017. To compensate for the potential bias in using totals for either party above, it was also important to compare average Representative expenditures by party. This was calculated by dividing the quarterly totals for each party by the number of representatives in the party during that period.

Figure 2.2: Average Quarterly Spending per Representative By Party



As Figure 2.2 demonstrates, the quarterly averages per Representative actually reverse the trend of higher Republican spending based on totals. The average Democrat spent approximately $5,000 to $10,000 more per quarter than did the average Republican. While it's tempting to make more in-depth inferences about this trend, there are a number of plausible explanations, with missing quarterly data at the forefront. As the overall budgets should theoretically average out in the same way regardless of party (the funding is impartially allocated by the Federal Government), it seems more likely that these discrepancies are based on incomplete quarterly reporting or otherwise missing data.

We also wanted to explore whether the parties reported any substantial differences in categorical spending given that Representatives allocate their budgets at their own discretion. Following a similar approach to the overall quarterly expenditures, total yearly expenses for either party were grouped and plotted by category.

Figure 2.3: Total Yearly House Expenditures by Category and Party

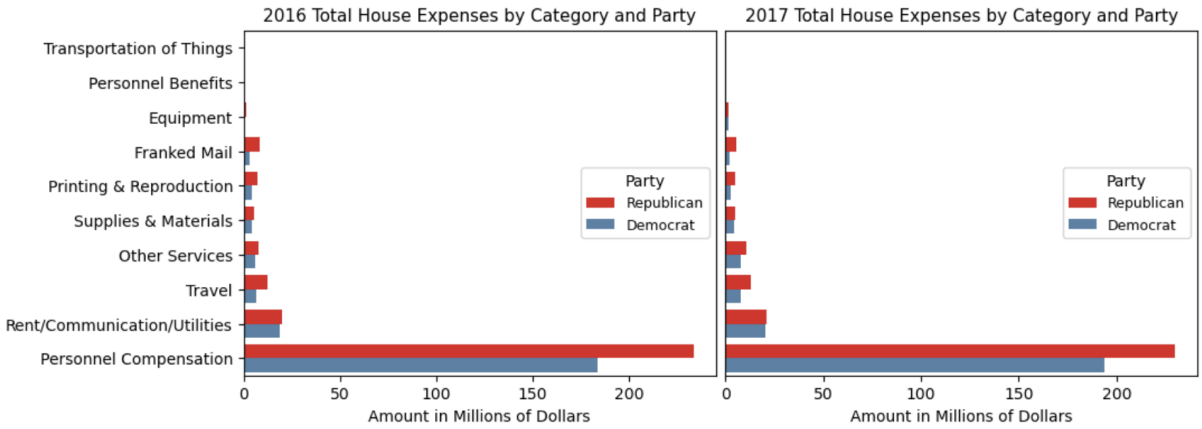
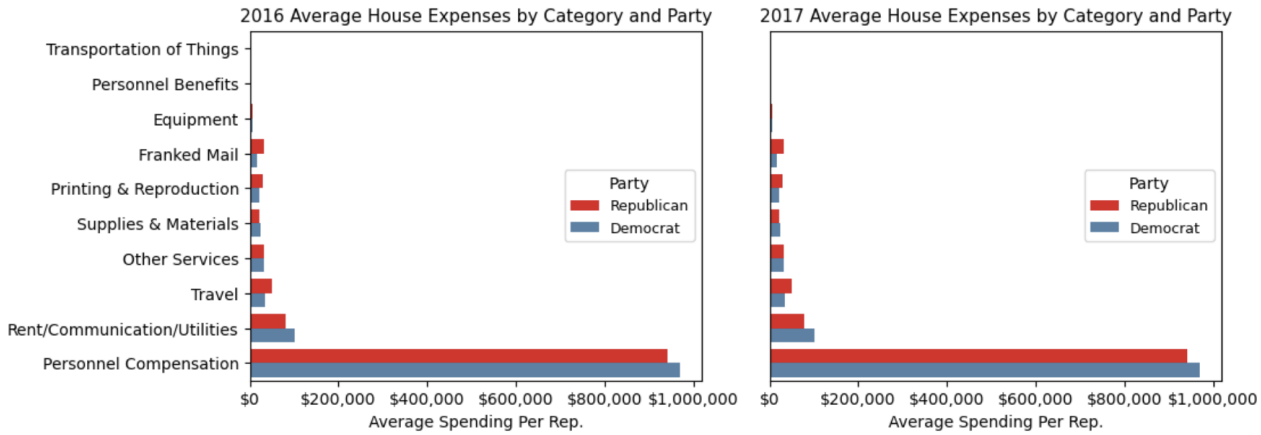


Figure 2.4: Average Yearly Representative Expenditures by Category and Party

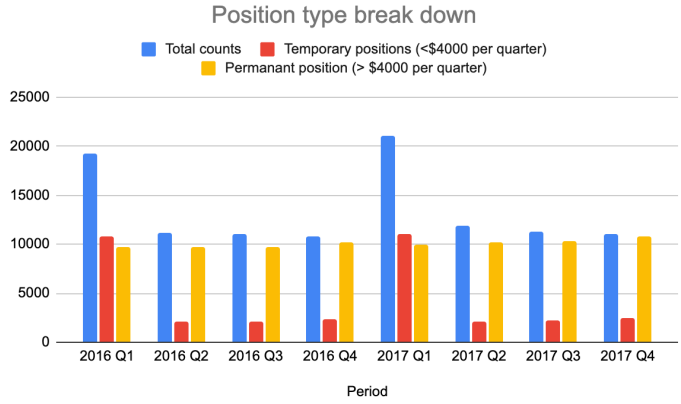


Similarly to the pattern seen in the quarterly totals by party (Figure 2.1), Republican Representatives consistently featured higher spending totals across each spending category in both 2016 and 2017 (Figure 2.3). The category with the greatest discrepancies along party lines was Personnel Compensation (with an approximate difference of $50 million - $75 million per year). This is unsurprising given how prominently the category features throughout our team's analysis as a major spending focus. According to our data, the Democrats collectively spent more on the top five categories in 2017 than they did in 2016, which most likely is due to gaining nine new Representatives at the loss of Republican seats. Figure 2.4, on the other hand, indicates that the partisan spending gap once again reverses slightly when party totals are averaged based on the number of Representatives in each party.

Analysis 3: Personnel Compensation and Retention Rate

The goal of this analysis is to better understand the Personnel Compensation category, which attributes to the labor distribution of the House Offices. From this category, by identifying unique positions and payee, the team can identify the head counts. The overall head counts are captured from 2016 Q1 to 2017 Q4 and broken down by temporary and permanent positions. A temporary position is defined as an individual who makes less than $4000 per quarter, and vice versa, a permanent position makes more than $4000 per quarter.

Figure 3.1: Position counts by temporary/permanent positions



One observation the team sees is the higher counts for Q1 of each year. Q2 to Q4 experiences the consistent same number of headcounts. We hypothesize the reason for this high headcount is the renewal of the House Office budget every year. Many of the offices decided to spend the first Q1 to prepare for the rest of the year. This spending habit plus the unique position aggregation can cause the high counts of every first Q1. Another observation that supports this is we can see the higher count in temporary positions in the first Q1 compared to the rest of the quarters. These temporary positions could be contractors, interns, or aides that went through a high turnover rate, which would explain the drop in head counts. The permanent position counts remain the same throughout the quarters.

One of the main analyses the team hopes to explore in this section is the staff retention rate of the House Offices. The reason why we chose the 2016-2017 time frame is because there were a lot of power transitions in this time period, such as the presidential election and midterm elections. Even though the datasets we are analyzing are for the House Office budgets, not Trump's administration, the team believes meaningful insights can still be deducted to understand the political climate overall. According to a report written by Tenpas for Brookings, a nonprofit public policy organization based in Washington D.C, Trump's A Team, which consisted of the executive officers of the president, had a turnover rate of 92%, significantly higher than previous presidency by approximately 10%. With this reason, the team hopes to find data that reflects this political event.

Table 3.1: Mean retention rate per period

Period	Mean retention rate
2016	47.38%
2017	80.69%
2016-2017	41.19%
2016 Q3 - 2016 Q4	50.00%

The team observes that the retention rate in 2016 was significantly lower than in 2017. Moreover, we wanted to zoom in on a specific quarter between 2016 Q3 to 2017 Q1, where many shifts of power and position occurred. The lowest retention is from 2016 Q3 - 2016 Q4, with 50%. This was when all the election results were released and power tensions may have been the highest. Contrary to 2016, 2017 retention was high, with the average of 80%. According to Propublica's research, the average retention rate between 2009 and 2012 (post-2008 recession) was 64.2%. The retention rate of 2016-2017 was 41.19%, which is substantially lower than in the 2008 recession time period.

Conclusion

In creating this project, most of our time and effort was spent on understanding, cleaning, and standardizing the data to fit the research's scope. We recognize that the existing raw data was messy because it was originally recorded by a variety of offices and individuals from the House Offices and later digitized to csv and json files for public record. We also recognize that there might be discrepancies between the data and actuality, due to the timing of when disbursements were entered. Moreover, the majority of our analysis focuses on the category names, so wording is an important factor to yield the most meaningful results. With these impacts, the team's focus was on the larger patterns rather than specific details of the budget.

Ultimately, although we anticipated that the intensely partisan nature of this period in American politics might translate to substantial differences in House spending by party, our data did not reflect this hypothesis. In regards to spending by offices, the bipartisan nature of the highest spending offices does not necessarily support or disprove our hypothesis. At the same time, the incomplete nature of the data as well as the assumptions involved in cleaning it limit these results, and we would need to use statistical methods beyond the scope of this project to draw meaningful conclusions about the trends we observe in our results.

In addition, the head counts and retention rate analysis supports our hypothesis that when former President Donald J. Trump was elected to office, there was a huge wave of decline to re-run from former Members of the House of Representatives. The 2016-207 retention rate was at 41.19%, lower than the 2008 recession rate which was at 64.2%. There was also a spending discrepancy between the Offices of the Minority and Majority leaders that, in the future, would be something to explore.

Appendix: References

1. https://www.politico.com/2016-election/results/map/house/
2. https://www.brookings.edu/research/tracking-turnover-in-the-trump-administration/
3. https://www.propublica.org/article/turnover-in-the-house-who-keeps-and-who-loses-the-most-staff
4. https://projects.propublica.org/represent/expenditures
5. https://www.propublica.org/article/house-operating-budget-cuts-paving-way-for-more-special-interest-influence
6. https://bioguide.congress.gov/
7. https://www.house.gov/the-house-explained/open-government/statement-of-disbursements
8. https://www.propublica.org/article/update-on-house-disbursements-a-few-notes-on-how-to-use-the-data
9. https://www.congress.gov/members
10. https://history.house.gov/People/