

Analysis Challenge 1 | Due: Monday 2023-02-06 by Midnight

The Challenge

In this first analysis challenge, imagine that you are an analyst for either the RNC or DNC (it's up to you which one you want to work for).

The party has collected data on the amount of money they've spent on various U.S. House elections from 2018, along with some details about those districts and the candidates that ran for office. They want you to **look at districts where their candidates lost to understand why**.

Write up a short research note (no more than 700 words) that:

1. Explains (politely!) to party leaders why the analysis they want you to perform is a bad idea.
2. Shows the results from a more appropriate analysis.

For part 1, a paragraph or two should suffice. For 2, look at how party spending in districts and candidate incumbency correlated with winning a greater share of votes in the election. For party spending, create a scatter plot with a regression line. For incumbency, create a column plot showing average vote share for their party's candidate by incumbency.

Before each figure provide a description of the figure (what kind of plot is it? and what are the variables?) and what the figure tells us (are the variables positively or negatively correlated? is there no correlation?).

Conclude with a paragraph summarizing your recommendations for party leaders.

Submit your write-up to Canvas as either a PDF or Word doc.

The Data

link: <https://raw.githubusercontent.com/milesdwilliams15/Teaching/main/DPR%20201/Data/HouseElectionsSpending2018.csv>

Details (Metadata):

This dataset shows electoral outcomes in U.S. House races for 2018. Observations in the data are at the district level. There are a total of 331 observations in the data. The data contains the following variables per district:

- state: Acronym for the state in which a district is located.
- dist: Numeric ID for a district within a state.
- incumbent: Indicator for whether the incumbent is a Democrat ("D") or Republican ("R").
- repvoteshare: The proportion (or share) of votes received by the Republican candidate (hint: the Democratic candidate's vote share is just 1 minus this).
- respending: Total spending by the Republican campaign in the district.
- demspending: Total spending by the Democratic campaign in the district.
- trumpvoteshare: The share of the vote that Donald Trump received in 2016.
- lagrepvoteshare: The share of the vote that the Republican candidate received in the last election (2016).