

# Skill Task 2

## Data Manipulation and Pipes

PS 811: Statistical Computing

Due February 14, 2020

This assignment revisits the CAFE data from the previous skills task. It describes how Senators voted for a certain fuel efficiency standards bill and their total campaign contributions-to-date from individuals who work for auto manufacturers.<sup>1</sup>

Begin a new session of R, and make sure you are working out of your PS 811 project.

Write an R file (R file! no R Markdown!) that fulfills the following tasks. Use the tidyverse verbs and the pipe operator (%>%).

1. Load the `here` and `tidyverse` packages, and import `CAFE.csv` into R (use `here()` to build the file path).
2. Investigate the dataset at a high level. Look at the variable names (`names()`), examine the number of rows (`nrow()`) and columns (`ncol()`), and get a rough summary of the variables (`summary()`).
3. Use the `count()` function to determine if the bill passed, based on the number of senators voting for or against.
4. Use `mutate()` and `case_when()` to create a more informative party affiliation variable<sup>2</sup> and an *indicator variable* of a Yea vote, taking the value of 1 if the Senator voted Yea or 0 if they voted Nay.
5. Which party was more supportive of the bill? Use your new variables along with the `group_by()` and `summarize()` functions to calculate the proportion of each party that supported the bill.<sup>3</sup>
6. **Bonus:** use the tools we've learned so far to determine which states had two Senators vote identically, and which states had two Senators vote differently. Your result should be a state-level data frame (50 rows) with a variable that indicates if Senators "co-voted."

When you are done, upload your R file to Canvas.

---

<sup>1</sup>I have removed/modified some of the original variables, but you can learn more about the data here: <https://vincentarelbundock.github.io/Rdatasets/doc/Stat2Data/CAFE.html>

<sup>2</sup>The party codes are the same as the data from lecture.

<sup>3</sup>Hint: if you calculate the mean of a binary indicator variable, it is equal to the proportion of 1s!