

## Problem Set 1

**Show all work.** For those questions that require you to write R code, turn in the scripts that you wrote and the output you obtained. Include all code in a file named `ps1_’your uni’.r` (e.g., I would turn in a file named `ps1_gjw10.r`) and embed comments in the file (i.e., using the `#` syntax) to answer questions posed. We should be able to run your R code without making any edits to the file. For the visualizations, please save these as separate PDF files. All files should be uploaded to the Assignments folder on CourseWorks. This assignment is due at 11:59 on 9/28.

From the files section on CourseWorks, download the file `fec22.txt`, which contains data for candidate political action committees for the 2022 elections in the U.S. Use the file `fec.codebook.txt` to see the values for the fields. Write R code to do the following.

1. Read the data into a data object called `fec22.df` using the appropriate command. Report the number of records/observations in the data. (2 points)
2. Report any variables that are missing values systematically. Is this what you expect? Why or why not? (2 points)
3. Subset the data to produce two different data objects—one for Senate candidates and one for House candidates (the variable `CAND_OFFICE_DISTRICT` equals 0 for Senate candidates, is greater than 0 for House candidates). Do a check that will give you a sense that the subsetting worked correctly. (2 points)
4. Calculate and report the mean, median, and standard deviation for total receipts (variable name `TTL_RECEIPTS`) for races for each chamber. Do this for the subsets produced in the previous step *without using dplyr*. Also do this on the original data that you read in (i.e., `fec22.df`) using `dplyr` and compare the results from the two approaches. (2 points)
5. For the data that includes only House candidates, produce density plots that shows two distributions—one for candidates who are incumbents and one for candidates who are challengers. The variable `CAND_ICI` equals “I” for incumbents, equals “C” for challengers, and equals “O” for candidates in open seat races. Write a sentence that summarizes what you see. (2 points)