Exercise 1: Presidential Polls in 2012

PS 270: Understanding Political Numbers

Due Wednesday, February 27

In this assignment, we will investigate polling data from three swing states in the 2012 presidential election: Florida, Ohio, and Virginia. Each case is a poll showing the support for Barack Obama (D) and Mitt Romney (R).

This exercise will test some important R skills for interacting with data. We will import data into R, examine and modify the data, and conduct simple analysis with a graphic.

Before beginning the assignment, do the following steps to prepare your computer.

1. On your computer, you should have a folder titled `PS-270` (or something similar) that contains all of your work for this class. *Inside* of your class folder, create two more folders (if you haven't already): one called `data` and another called `figures`.
2. Download and save the polling data from Canvas into your new `data` folder.
3. Open RStudio by double-clicking your `.Rproj` file inside your `PS-270` folder.
4. Run the following commands in the console to make sure that everything is set up to begin the assignment.
    a. If you enter the `getwd()` function, R should print a file pathway that ends with the name of your `PS-270` folder.
    b. The `list.files()` function prints the names of files and folders in your `PS-270` folder. You should see your `"data"` and `"figures"` folders among the output.
    c. Typing `list.files("data")` prints the names of files and folders inside the `data` folder. You should find the `.csv` data that you downloaded from Canvas.

Your Tasks

Write and save an R script that implements the following tasks in order.

1. Load the `tidyverse` and `here` packages.[1]
2. Import the polling data file into R. Do this with the `read_csv()` function as follows.

[1] Install any missing packages by typing `install.packages("package_name")` into the console. Don't save any package installation commands in your R script file.

```
polls <- read_csv(here("data", "FileNameGoesHere.csv"))
```

This code says "import a CSV (spreadsheet) file, which is located in the `data` folder, named `FileNameGoesHere.csv`." You will need to fix the file name so that R grabs the file you downloaded from Canvas.

3. Print the data object to the console to see what's inside.[2] Print the variable names.
4. How many polls were taken in each state? Find out by using the `count()` function to tabulate the state variable.
5. Obama's "margin" measures how much more support he had in the poll compared to Romney. It will be a negative number if Romney did better in a given poll. Create a histogram of Obama's vote margin. Plot states as separate panels (using one of the `facet_` functions). Use your judgment and taste to make your graphic look professional: polish up any sloppy axis titles, add a title, and so on. Save the graphic into your `figures` folder.

```
ggsave(here("figures", "ex-1_histogram.pdf"))
```

6. We often measure U.S. voting using a political party's *share of the two-party vote*: the vote percentage for a party when we set aside Independent and third-party voters. We would calculate the *Democratic share of the two-party vote* using the following equation.

$$\text{Dem. share of two-party vote} = \frac{\text{Dem. vote}}{(\text{Dem. vote} + \text{Rep. vote})}$$

Create this variable using `mutate()` and call it `dem_2party`.[3]

7. Calculate the mean of Obama's two-party vote share *in each state*. You can do this in one of two ways.
   - Use `filter()` to split the data into three separate objects, one for each state. Then use `summarize()` to calculate the mean of the two-party vote in each state.
   - Group the combined data (`polls`) by state, and find the mean in each group. If you `group_by()` the state, calculating the mean with `summarize()` would give you the mean in each group.

When you are finished, upload both your R script and your saved graphic to Canvas as your submission for the "Exercise 1" assignment.

[2] Tip: You can type `View(dataset_name)` into the console to see a spreadsheet view of the data.

[3] Tip: don't overwrite the original data object until you are confident that the code is working without any issues!