

R Studio Section Practice Activity to Prepare for PS 1
PLSC 373

1. For this question we're going to explore some (real!) survey data about which party respondents want to win in the 2024 general election. To do this we're going to look at what is called the congressional "generic ballot" question that asks:

What is your preference for the outcome of this year's congressional elections for the U.S. House of Representatives? The response options are the Democratic Party, The Republican Party, or No Preference/House Evenly Divided.

The website [fivethirtyeight.com](https://projects.fivethirtyeight.com/congress-generic-ballot-polls/) has aggregated generic ballot poll results from a number of survey firms, and the results can be seen here:

<https://projects.fivethirtyeight.com/congress-generic-ballot-polls/>.

1. Create a new R Script file called: ps373ps1yourname.R [Substitute your actual name into the file name.] All of the following steps should be a part of your R Script File. Be sure to annotate your code using the hashtag symbol # before any comments explaining your code.
2. Place the generic_ballot_polls.csv in your working directory.
3. Load the data into R Studio using the command `read.csv` and call the dataset `gen_poll`. Note, be sure to indicate that this dataset has variable names in the first row when you use the `read.csv` command.
4. Using the **`dim()`** command, how many rows are in the data (not including the variable names)? How many columns are in the data?
5. Look at the data using the **`View ()`** command. What are the observations in the data and what are the variables describing.

6. Using the `names()` command, what are the variable names in the dataset?
7. Averaging across all the polls in the dataset, what is the average (mean) "sample_size" of all points. Hint: `dataframe$variable` allows you to operate on variable in a dataframe. Does the `mean()` command work? If not, examine the data and figure out why.
8. Use the `prop.table(table())` command on the "methodology" variable to see the percentage of each survey methodology used. What is the most popular mode of polling?
9. Learning about functions: `?command`. For example, try `?summary()` to understand how to use the summary function.
10. Generate the summary statistics of the Democratic support variable ("dem") using the `summary()` command, and then generate summary statistics for the entire dataset. Understand the meanings of mean, median, mode, minimum, and maximum of a variable.
11. Generate two histograms to illustrate the distribution of Democratic support and another for Republican support using the `hist()` command.

12. Use the Democratic and Republican support variables (dem, rep) to create a new variable dem_leading if Democratic support is greater than Republican support. Then, calculate the average of this dem_leading variable and provide an interpretation using the ifelse() statement. Hint: Use ?ifelse() to figure out how.

13. Generate a plot(x,y) with the polling "start_date" as the horizontal (x) axis and support for the Republican candidate (rep) as the vertical (y) axis variable. Indicate the time trend with x = "start_date", y = "rep". and generate a second graph for democrats as well. Do you observe any trend?

Hint: convert start_date into a date variable first using as.Date(start_date, format="m/%d/%y")

14. Install and load packages using install.packages("packagename") and library(packagename). Install and load a package called "ggplot2". If you have done so correctly you should be able to execute this code:

```
# Assuming gen_poll is your dataframe with 'dem' and 'rep' columns
ggplot(gen_poll, aes(x = dem, fill = "Democrat")) +
  geom_histogram(data = gen_poll, aes(x = dem, y = ..density..), fill = "blue", alpha = 0.5, binwidth = 1) +
  geom_histogram(data = gen_poll, aes(x = rep, y = ..density..), fill = "red", alpha = 0.5, binwidth = 1) +
  labs(title = "Overlapping Histogram of Democratic and Republican Support",
       x = "Support",
       y = "Density") +
  scale_fill_manual(name = "Party", values = c("Democrat" = "blue", "Republican" = "red")) +
  theme_minimal()
```