Problem set 2

Your name here

Due 10/8/2021 at 5pm

NOTE: *Start with the file* `ps2_2021.Rmd` *(available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F21/tree/main/assignments). Modify that file to include your answers. Make sure you can "knit" the file (e.g. in RStudio by clicking on the* `Knit` *button). Submit both the Rmd file and the knitted PDF via Canvas*

Question 1: US presidential election results (again)

Download the file "tidy_county_pres_results.csv.zip" from the repository (https://github.com/UChicago-pol-methods/IntroQSS-F21/tree/main/data), unzip it, and put the CSV file in the same directory as your Rmd file.

Then load the data:

```
library(tidyverse)
df <- read_csv("tidy_county_pres_results.csv")
```

For each US county (uniquely identified by `FIPS` and labeled with `county` and `state`) in each presidential election `year`, we have the total number of votes cast (`total_vote`), number of votes for the Democratic candidate (`dem_vote`), and number of votes for the Republican candidate (`rep_vote`).

(1a) Add a variable called `other_vote_share`, which is the proportion of votes cast for candidates other than the Democratic and the Republican.

```
# your code here
```

(1b) Show a histogram of `other_vote_share` in 2000.

```
# histogram code here
```

(1c) Identify the counties with the highest `other_vote_share` in 2000. Output a table showing the county name, state, and `other_vote_share` for the six counties with the highest `other_vote_share` in 2000. (Don't worry about making the table look nice; just produce the raw `R` output.)

```
# table code here
```

(1d) Using `group_by()` and `summarize()`, produce and store a new `tibble` showing the two-party vote share for the Democrat in each election year. ("Two-party vote share for the Democrat" is the votes for the Democrat divided by the votes for either the Democrat or the Republican.) Use it to make a plot showing the Democrats' two-party vote share (vertical axis) across years (horizontal axis).

```
# your code here
```

(1e) Using `group_by()` and `summarize()`, produce and store a new `tibble` showing the proportion of counties in which the Democrat got more votes than the Republican in each election year.

Use it to make plot showing the share of counties won by the Democrat (vertical axis) across years (horizontal axis).

```
# your code here
```

(1f) Use `left_join()` to merge the two tibbles (one with county share, the other with vote share) and store the result. Use this new tibble to plot the Democratic county share (vertical axis) against the Democratic vote share (horizontal axis) over time, as in the last problem set.

```
# your code here
```

(1g) Use `pivot_longer()` to convert the tibble created in the last question to a format appropriate for plotting both the Democratic vote share and the Democratic county share (vertical axis) against the year (horizontal axis) as on the last problem set. Make that plot.

```
# your code here
```

Question 2: independent project data

Choose a dataset that you will use for your independent project. As noted last week, it should have many observations (but not too many) and many variables. And it should be interesting to you! Load the data and make a figure using tools we have learned in class.