Week 11 Programming Assignment

Instructor: Chien-Hao Fu

5/02/2019

Due: 5/08/2019

Preparation

1. Download the five csv files from the course website
2. Import the files and save them in the following names
    - Import `elec.csv` as data frame `elec`
    - Import `pop2012.csv` and `pop2016.csv` as data frames `pop12` and `pop16`
    - Import `state_info.csv` as data frame `state`
    - Import `edu.csv` as data frame `edu`
3. Take a moment to examine these data frames
4. Load the package `tidyverse`

Description of important variables

- `per.high`: share of population with high school degree or higher
- `per.bac`: share of population with bachelor degree or higher
- `R_share`: share of votes for Republican presidential candidate
- `D_share`: share of votes for Democratic presidential candidate
- `elderly`: share of population ages 65 or older
- `young_adult`: share of population ages 19-34

Programming Assignment

Goal: Draw two scatter plots and compare its patterns

1. A scatter plot of `elderly` (on the y-axis) against `R_share` (on the x-axis) with observations from both 2012 and 2016
2. A scatter plot of `per.bac` (on the y-axis) against `R_share` (on the x-axis) with observations from both 2012 and 2016

Key Steps

Step 1. Produce a single data frame for population data

Tips

- You can use `bind_rows` to combine both `pop12` and `pop16`
- You should add a variable `year` to indicate the year of data

Output

- A data frame `pop` that contains the following variables: `year`, `Location`, `elderly`, and `young_adult`
- The data frame should have 106 observations

Step 2. Process the data frame `elec` to be later merged to `pop.m`

- Examine the data frame `elec`. Can it be directly merged to `pop.m`?
- If not, what information does it need before it can joined with `pop.m`?

Tips

- You should use `xxx_join` to add the full state name from `state` to the main data frame `elec`

Output

- After joining the data tables, the data frame `elec.m` should have 104 observations
- The data frame should include the following variables: `year`, `state`, `R_share`, `D_share`, `State`, `Abbreviation`, `Capital City`

Step 3. Join the tables to create a single data frame for analysis

- First join the two tables: `elec.m` and `pop.m`
- Then join with the data frame `edu`
- The resulting data frame – (called `combined`, for example) – should have 104 observations and 12 variables

Step 4. Draw two scatter plots

Plot 1: Use `geom_point()` to plot the variable `elderly` on the y-axis and the variable `R_share` on the x-axis

Plot 2: Use `geom_point()` to plot the variable `per.bac` on the y-axis and the variable `R_share` on the x-axis

Step 5. Complete and submit "W11 Student Submission.docx" to the course website

Bonus Tasks

- Map `color` to `factor(year)`
- Use `scale_color_manual()` to adjust the values and labels of the color mapping. For example, add the following codes to your `ggplot` will adjust the color mapping and labeling:

```
+ scale_color_manual(
    limits=c(2012,2016),
    labels=c("2012 (D Won)", "2016 (R Won)"),
    values=c("navyblue", "maroon"),
    name = "Election Year")
```

- Adjust the label of your plots with the function `labs()`
- Use the piping operator `%>%` to simplfy your codes