Problem Set 4

(Due March 9, 1:00 PM)

Instructions

1. The following questions should each be answered within an R script. Be sure to provide many comments in the script to facilitate grading. Undocumented code will not be graded.

2. Work on git. Fork the repository found at `https://github.com/jeonghkim/PS4`. There is already code to get you started. As you add your code, commit and push frequently. Use meaningful commit messages – these may affect your grade.

3. You may work in teams, but each student should develop their own R script. To be clear, there should be no copy and paste. Each keystroke in the assignment should be your own.

4. If you have any questions regarding the Problem Set, contact the TA or use office hours.

5. For students new to programming, this may take a while. Get started.

Scraping Wikipedia

- Go to `https://en.wikipedia.org/wiki/List_of_United_States_presidential_elections_by_popular_vote_margin` and locate the first table. Your task will be to scrape the information from this table.

- Wikipedia uses https, which is not supported by the package `XML`. Instead you will need the package `rvest`. Code is provided (PS4.R). The data will need some minor cleaning to get numbers of interest.

- Creatively visualize two or three trends over time.

  - There is no strict requirement on what you choose to visualize, but it must be interesting in some way and each graph should demonstrate a specific trend in an intuitive and appealing way. Layout matters. Use `par()`, `layout()`, etc.

  - Write a brief explanation about each of the trends you find.
  - After creating plots, save all of your plots to a single pdf within your code. Upload your pdf to github in addition to your code.

- Scrape the electoral college votes won by the candidates at `https://en.wikipedia.org/wiki/United_States_presidential_election`. Merge this data with the data from the previous exercise. The final table should look like the first table, but with two additional columns: one for the winner's electoral college votes, and another for the runner up. Save this as an Rdata file, and upload it to github.