Problem Set 7

(Due April 26, 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/minheeseo/PS7` and add your code, committing and pushing 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 her office hours.

Your tasks

For this problem set, you will need to use `dplyr` and `ggplot2` R packages to summarize a given dataset and create an appropriate visualization. You will have to *only* use the functions within `dplyr` (ex. `filter()`,`select()`,`summarise()`, and `mutate()`) and pipe operator (`%>%`) to work with data. Please complete the following tasks in order:

1. Go to this link and download March2018 crime dataset: `http://www.slmpd.org/Crimereports.shtml`

2. Compute the number of crime per day by the type of crime (Hint: clean Description variable and use it). Which types of crime happened the most in March?

3. Compute the number of crime per day by neighborhood. Which neighborhood has the most number of crime?

4. Compute the proportion of crime related to robbery by district. Which district has the largest proportion of crime related to robbery?

5. Visualize changes of all types of crime over time using `ggplot2`. Write appropriate labels and titles.

6. Visualize changes of all types of crime over time by district using `ggplot2`. Choose different color to indicate each district. Write appropriate legend, labels and titles.