Application in Data Cleaning and Wrangling

YOUR NAME HERE PLEASE! :)

2024-01-31

Instructions: Please read carefully.

- Create the necessary code chunks and written answers to complete each task. Recall that all code in Markdown goes into "chunks", which can be created with Ctrl+Alt+I. Anything outside of a chunk will be rendered as text, not code.
- **Please keep all code chunks and text answers for each task within the section dedicated to that task.**
- Upon completion, delete the instructions, knit a pdf, change its filename to **lastname_firstinitial.pdf**, then upload to Canvas.
- You may collaborate with the person next to you, check documentation, and use google. Help from me will be limited to basic troubleshooting of errors.
- Grading is based on a combination of everything being well formatted, with accurate titles, axis labels, explanations, complete answers, and error-free code.

You are a data analyst consulting for a law firm. You have four tasks you must complete by 10:50am.

Task 1: The firm would like to know how the number of monthly cases they receive varies over time, grouped by zipcode, *but for only those zipcodes containing the string "292"* (after all, there are 600 zipcodes in here). Please provide at least one figure (plot) and a few sentences of explanation for the client around what the plot tells us.

Task 2: The firm would like to know how the number of monthly cases they receive varies over time, grouped by *source*. Please provide at least one figure (plot) and a few sentences of explanation for the client around what the plot tells us.

Task 3: The firm would like to know how the number of monthly cases they receive varies over time, grouped by *case type*. Please provide at least one figure (plot) and a few sentences of explanation for the client around what the plot tells us.

Task 4: The firm would like a map, divvied up by zipcode, colored by total number of cases for the *entire year of 2022.* Luckily, the "senior analyst" (me, lol) has already written the code for the loading in of map data and the merging of case data with the map data. He even wrote the basic map plot code for you :) . Technically, there are three lines missing that you must come up with to complete the chunk and get the map. *The map should also be clearly labelled with a title and a proper title for the legend.*

```
# library(tigris)
# zipcode_NC <- zctas(state="NC", year=2010)
# zipcode_SC <- zctas(state="SC", year=2010)
# zipcodes <- rbind(zipcode_NC, zipcode_SC)
```

```r
# line to rename ZCTA5CE10 to the same name zipcodes take on in your case data
# line to sum up cases by year and zipcode
# line to subset down to 2022


# zipcodes2 <- merge(zipcodes, cases_2022, by="Zipcode", all = TRUE)
# ggplot(zipcodes2) + geom_sf(aes(fill=cases))
```