Homework #1

Download `DSNY_Monthly_Tonnage_Data.csv`. This data set records the monthly collection tonnages that the Department of Sanitation collects from NYC residences and institutions. We will analyze this data set for this assignment. You can find further details about this data set, including descriptions for each column headings, at `https://data.cityofnewyork.us/City-Government/DSNY-Monthly-Tonnage-Data/ebb7-mvp5`.

Answer the following questions in a single R script called `hw01.R`. Answers must be given by R commands. You cannot simply look at the data set and answer the question via direct inspection. Use comments (#) to indicate which portion of your code answers which question. Be sure that you obtain the correct solutions to each question when you execute your script one line at a time from top to bottom.

Each question will be graded out of 4 points according to the following criteria:

- 0 points: No attempts is made to answer the question.
- 1 point: An attempt is made that, although unsuccessful, revealed some understanding of what the question was asking.
- 2 points: Solution is incorrect, but with some modifications, could be corrected.
- 3 points: Solution is incorrect, but easily resolved with minor modifications **OR** solution is correct, but obtained via convoluted reasoning or by avoiding standard approaches.
- 4 points: Solution is correct and uses standard approaches.

#1) What is the total amount of refuse collected (i.e. garabage) collected throughout NYC on 2018 / 7 (i.e. July 2018) according to this data set?

#2) NYC has five boroughs: Manhattan, Bronx Brooklyn, Queens and Staten Island. You can access boroughs by matching with the column `BOROUGH` or with `BOROUGH_ID`. What is the average amount of refuse collected per month from each borough? (Hint: Look up `na.rm` for the `mean` function.)

#3) NYC is divided up into 59 sanitation districts (indexed by `COMMUNITYDISTRICT`). A careful inspection of the data set will show that some districts do not record refuse collection data. Which sanitation districts have data recorded in this data set?

#4) Using `tapply`, compute the average refuse collected for each sanitation district.

#5) Produce a vector that explains why district 7A does not have an average refuse collected.

#6) The following code converts the output of `tapply` in #4 to a data frame and removes the row for district 7A.

```
dist_refuse <- data.frame(COMMUNITYDISTRICT=names(tapply_out),
    AVGREFUSETONSCOLLECTED=tapply_out)
dist_refuse <- dist_refuse[!(dist_df$COMMUNITYDISTRICT == "7A"),]
```

Using this approach, create a data frame that gives the average paper recycling and average metal, glass and plastic recycling for each sanitation district.

#7) Which sanitation districts have over 300 tons of paper recycling and less than 400 tons of metal, glass and plastic recycling?

#8) Which sanitation districts have less than 300 tons of paper recycling or less than 350 tons of metal, glass and plastic recycling?

#9) Identify the sanitation districts recycles more metals, plastics and glass than paper.

#10) Identify the sanitation district with the highest ratio of recycling tonnage (including paper and metal, glass and plastic) to refuse tonnage. (Hint: Do not add averages!)