MAT 128 Lab 20

Much of this lab was originally written by Katherine St. John (original lab)

1) Work through the DataCamp tutorial on dataframes in R:
   https://campus.datacamp.com/courses/free-introduction-to-r/chapter-5-data-frames?ex=1

2) We'll use the NYC historical population data that we explored with pandas. If you do not have a copy, download and save to the desktop.

3) To upload the file to R Studio, click on the Files tab in the bottom right window. Click on Upload in the menu that appears and upload the file.

4) nycHistPop.csv should now appear in your list of files (bottom right window). Click on the file and select "View File" to view the file and make sure it was uploaded correctly.

5) Look at the file. How many lines of comments and introductory material are there? As with pandas, we need to skip those row when we read in the program.
To find the option for skipping lines, bring up the help page for read.csv:

```
help(read.csv)
```

Read down the documentation until you find the command for skipping extra lines at the beginning of the file. Try it before looking at the answer below.

6) To read in the CSV file:

```
pop <- read.csv("nycHistPop.csv", skip=5)
```

7) Let's use the notation for R data frames to make a bar plot of the Bronx population over time:

```
barplot(names=pop$Year, pop$Bronx)
```

8) Next, let's add in the average population into the graph. First, we need to compute the mean:

```
mean(pop$Bronx)
```

When we do this in R, we end up with the value na ("not assigned").

9) Let's check why:

```
print(pop)
```

The early years in the table have no values for the Bronx population. There is an option to ignore unassigned values. See if you can find it using help:

```
help(mean)
```

What is the option to ignore the values? (The answer is below.)

```
mean(pop$Bronx, na.rm = TRUE)
```

10) Now we have the average population of the Bronx, over time. Let's assign it to a variable and add a line to the graph:

```
m <- mean(pop$Bronx, na.rm = TRUE)
abline(h = m)
```

You may have to click on the Plots tab in the bottom right window to see your plot.

Challenges

- Make a plot of the population of Manhattan. Include a line marking the mean on your plot.
- Using the shelter system dataset from Lab 2 (follow instructions in that lab to download again if necessary), make a plot of the total individuals in shelter over time.