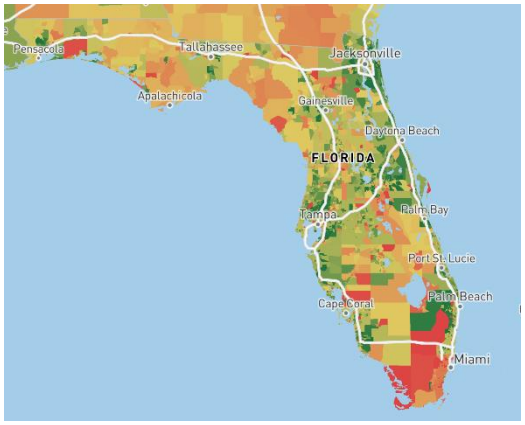
Module Assignment
Module 1
QMB-6304 Analytical Methods for Business

Write a simple R script to execute the following:



Preprocessing

1. Load the file "6304 Module 1 Assignment Data.xlsx" into R. This file contains information on crime in each of the 67 counties in Florida. This will be your master data frame.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 30 counties using the method presented in class. This will be your primary data frame.

```
#Carolina Aldana Yabur
#UID: U25124553

rm(list=ls())
install.packages("rio")
install.packages("moments")
library(rio)
library(moments)
counties_crime=import("6304 Module 1 Assignment
Data.xlsx",sheet="County Offense Data")
colnames(counties_crime)=tolower(make.names(colnames(counties_cr
ime)))
set.seed(25124553)
crime.sample=counties_crime[sample(1:nrow(counties_crime),30),]
attach(crime.sample)
```

Analysis

Using R calculate and report the following using your primary data frame:

1. The structure of the data frame using the str() command.

```
> str(crime.sample)
'data.frame':  30 obs. of  12 variables:
 $ county                          : chr  "Duval" "Citrus"
"Desoto" "Okaloosa" ...
 $ population                      : num  982080 149383 37082
203951 27443 ...
 $ total.crimes                    : num  34452 2333 736 3723 574
...
 $ murder                          : num  143 7 3 9 1 1 149 14 4
3 ...
 $ rape                            : num  477 25 22 94 6 3 592
193 26 8 ...
 $ robbery                         : num  961 38 13 43 8 ...
 $ aggravated.assault              : num  5074 335 148 514 79 ...
 $ burglary                        : num  4021 334 170 427 152
...
 $ larceny                         : num  20655 1403 336 2331 273
...
 $ vehicle.theft                   : num  3121 191 44 305 55 ...
 $ crime.rate.per.100k.popln       : num  3508 1562 1985 1825
2092 ...
 $ clearance.rate.per.100.offenses: num  18.6 34.9 43.9 31.5
54.4 54.6 18.9 22.8 36.8 44.3 ...
```

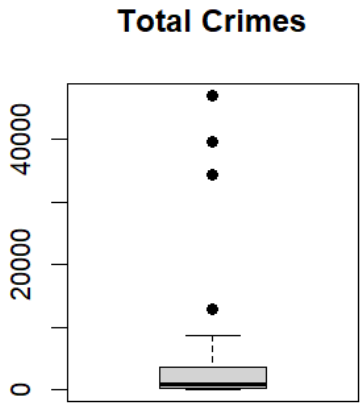
2. Mean, Median, Standard Deviation, Skewness, and Kurtosis of the Population variable. Based on these descriptive measurements how closely do you think this variable conforms to a theoretical normal distribution?

```
> mean(population)
[1] 251889.1
> median(population)
[1] 58602
> sd(population)
[1] 451678.7
> skewness(population)
[1] 2.505301
> kurtosis(population)
[1] 8.707224
```

It does not conform to a theoretical normal distribution. The mean and median are not equal, and since the mean is slightly higher than the median, it indicates a slight right skew in the data.

3. A boxplot of the Total Crimes variable. Based on this boxplot what can you say about the symmetry/skewness of this variable?

```
boxplot(total.crimes, main= "Total Crimes", pch=19)
```



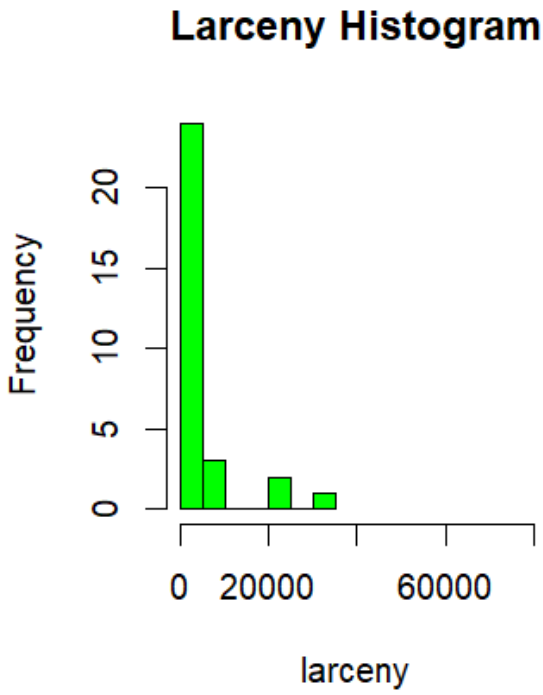
There are circles that are above the whiskers in the boxplot interquartile range, so it appears to be right-skewed.

4. Quartiles of the Aggravated Assault variable. Show your quartiles running from the minimum to maximum values for the variable, incrementing by .20.

```
> quantile(aggravated.assault,probs=seq(0,1,.2))
    0%    20%    40%    60%    80%   100%
   6.0   54.8  101.4  220.0 1024.4 5634.0
```

5. A simple histogram of the Larceny variable. Color your histogram green and give it an appropriate main title. Make sure the bottom axis of your histogram covers a range from 0 to 80,000. Based on this graphical tool would you say from this histogram the distribution of Larceny follows a symmetric distribution, or a skewed distribution?

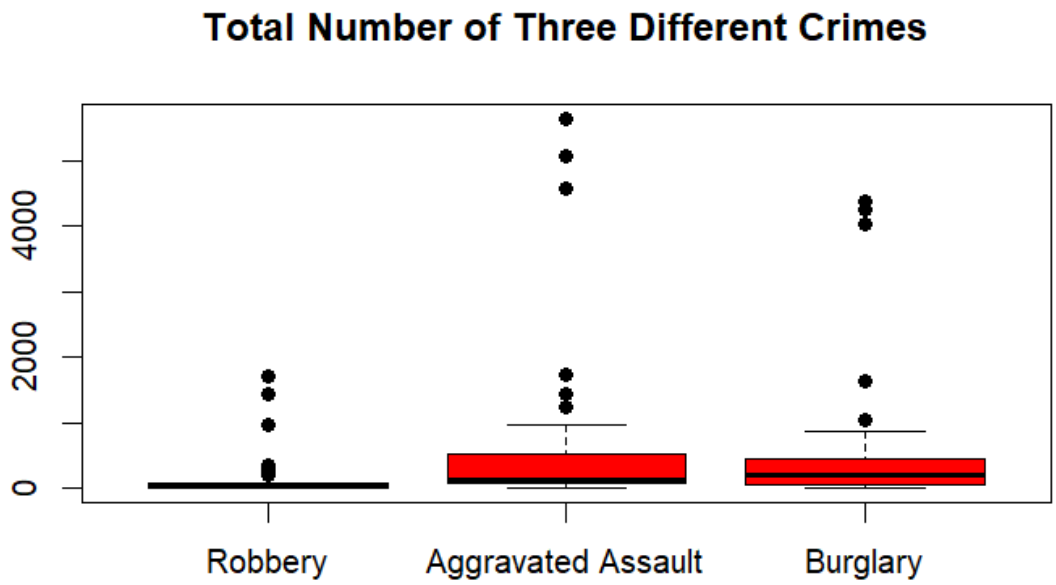
```
hist(larceny,col="green",xlim=c(0,80000), main="Larceny
Histogram")
```



The distribution of the Larceny variable appears to follow a skewed distribution since the tail on the right side of the histogram is longer than the left side, meaning a right skewed in the data.

6. Three comparative boxplots for the Robbery, Aggravated Assault, and Burglary variables. Your boxplots should be colored red and shown side by side in a single graphic with an appropriate main title and labels for the crime categories on the bottom axis. Based on these boxplots what can you say about the similarity in the number of crimes in these categories?

```
boxplot(robbery,aggravated.assault,burglary,
        main="Total Number of Three Different Crimes",
        col=c("red"),pch=19,
        names=c("Robbery","Aggravated Assault","Burglary"))
```



The number of burglary crimes is higher than the number of crimes for aggravated assault and robbery, and the number of robbery crimes is the lowest. It also seems that on average burglary crimes are more common.

7. Use R to determine and report the name of the county in primary data frame with the maximum number of Total Crimes.

```
> max(total.crimes)
[1] 47045
```

For inference we can say the name of the county is Broward.

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions and 2) the results of those instructions. The first line of your script file should be a "#" comment line showing your name as it appears in Canvas. Results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. **No collaboration of any sort is allowed on this assignment.**

The high standard deviation also indicates that the data is spread out. And lastly, the skewness and kurtosis also confirm that it appears to be right-skewed.