**Homework 4\_502**

**Name: Malik Hassan Qayyum**

**Data: FBI Crime Data, FBI Crime Data Geographically**

**Task:**

Using all three software (Excel, Tableau, R) create a times series plot for the data as well as a visual demonstrating the categorical nature of the data (ie comparison/distribution)

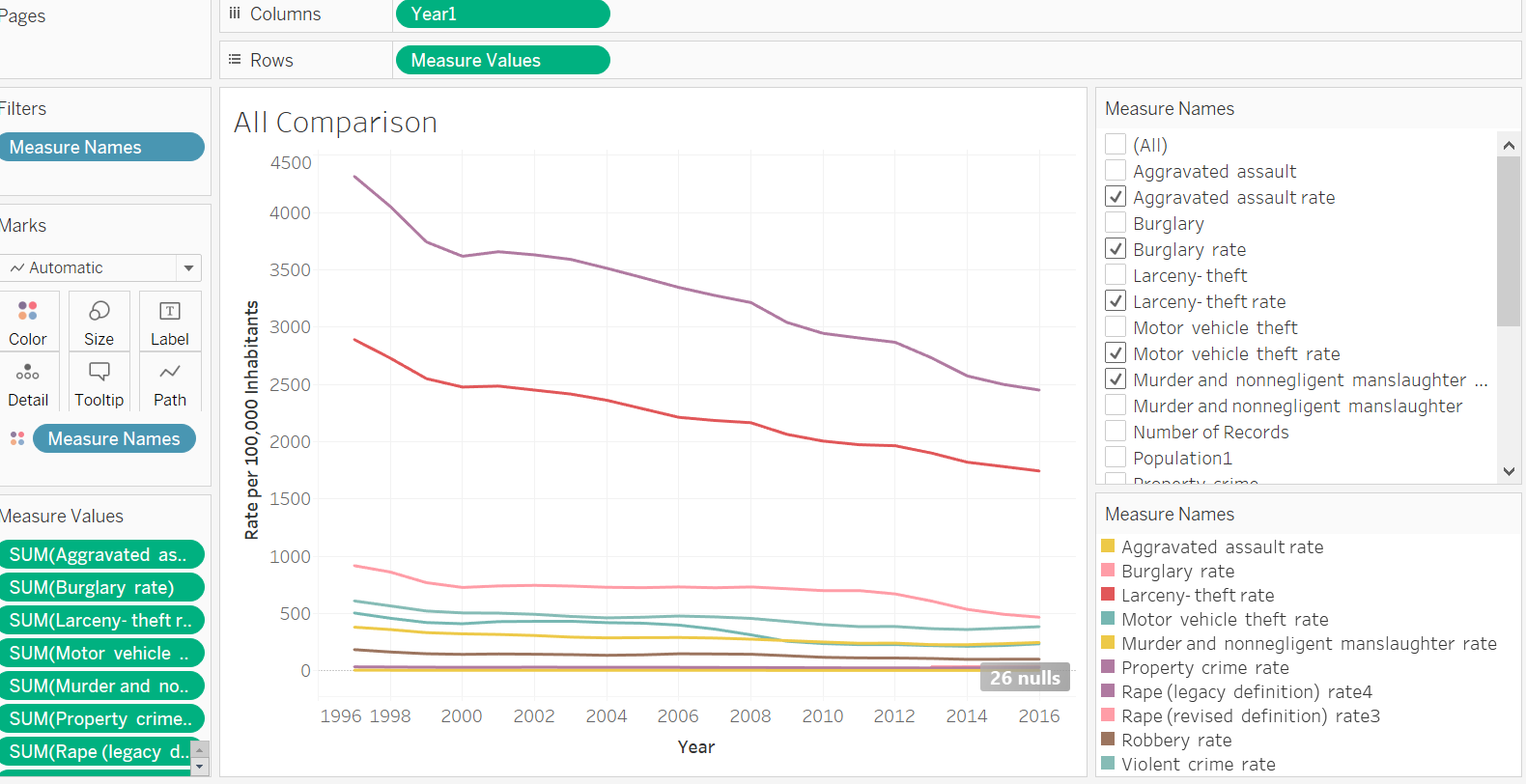
Label each of the x and y axis, along with a title for all plots

Briefly explain what you discover from your visuals as if you were presenting to an audience.

Create an additional plot in Tableau using the Geographical data file and justify your reasoning for the use of this plot.

**Update:** Use Rates

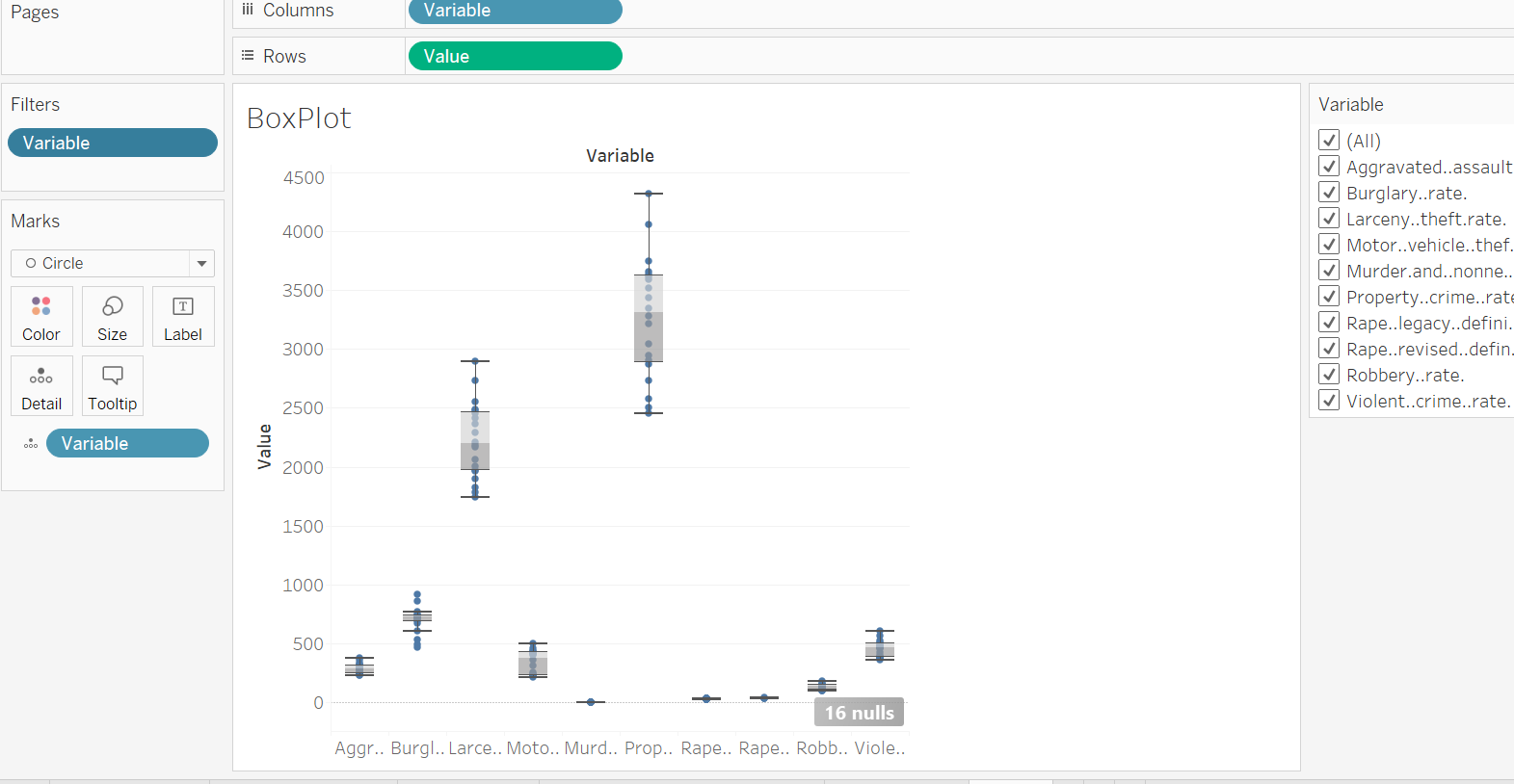
**Tableau:**

****

We can see that Property crime and larceny- left are the highest rates. Murders and Rapes are least rate as compared to other crimes throughout the period of 1997 to 2016. There is decrease in numbers of Property crime and larceny- left cases over the given period.

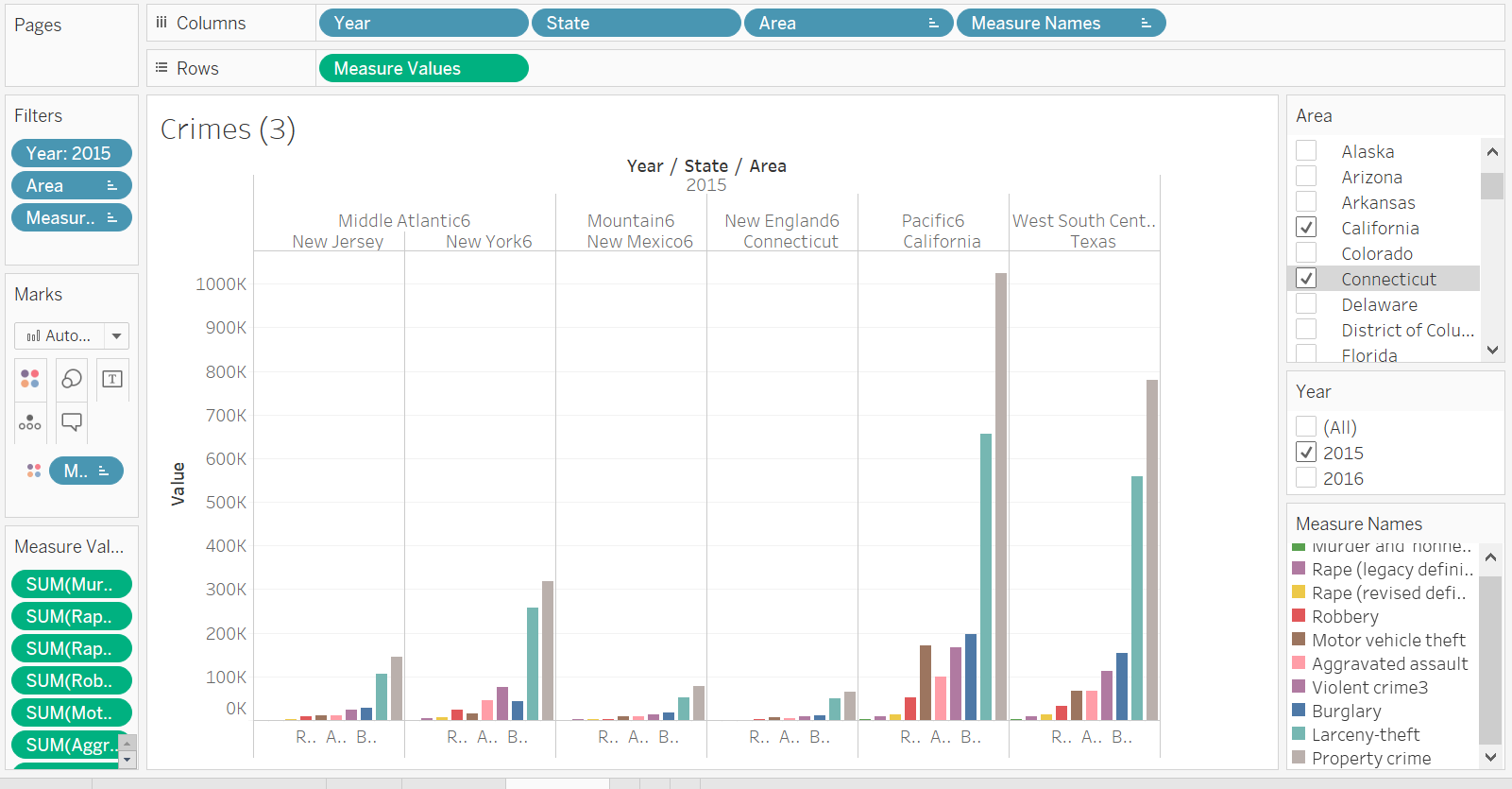


From 1997 to 2016 Population has increased from roughly 265M to 210M in the USA.



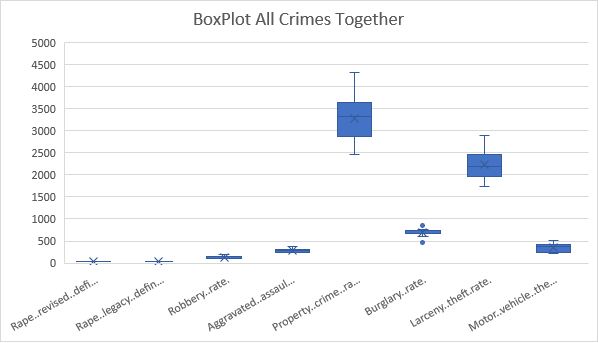
Shows same trends. Preoperty crimes have high distribution of rates

Geographical Data:



**Note:** I have selected only a few states and regions so the graph can fit on screen. States and year can be selected form the filters pane on the right side.

**Excel:**



**R - Studio**

Code:

# install.packages("reshape2")

library(ggplot2)

library(scales)

library(reshape2)

#################################################################################

crime\_df <-read.csv("C:/Users/Malik/Documents/GitHub/Data-Visualization-Data502/Dataset/DATA\_FBI Crime Rate.csv")

crime\_df\_cases = crime\_df[c(1:20),c(1,2,3,5,9,11,13,15,17,19,21)]

crime\_df\_rates = crime\_df[c(1:20),c(1,2,4,6,8,10,12,14,16,18,20,22)]

names(crime\_df\_rates)

#################################################################################

crime\_df\_rates\_melted = reshape2::melt(crime\_df\_rates,id.vars="Year", measure.vars=c(3:ncol(crime\_df\_rates)), value.name="value")

write.csv(crime\_df\_rates\_melted,"C:/Users/Malik/Documents/GitHub/Data-Visualization-Data502/Dataset/DATA\_FBI Crime Rate\_melted\_r.csv")

# Line Plot

ggplot(crime\_df\_rates\_melted, aes(Year,value,color=variable)) + geom\_line() + xlab("year") + ylab("value") + ggtitle("Crime Rate Comparison")

# Box Plot

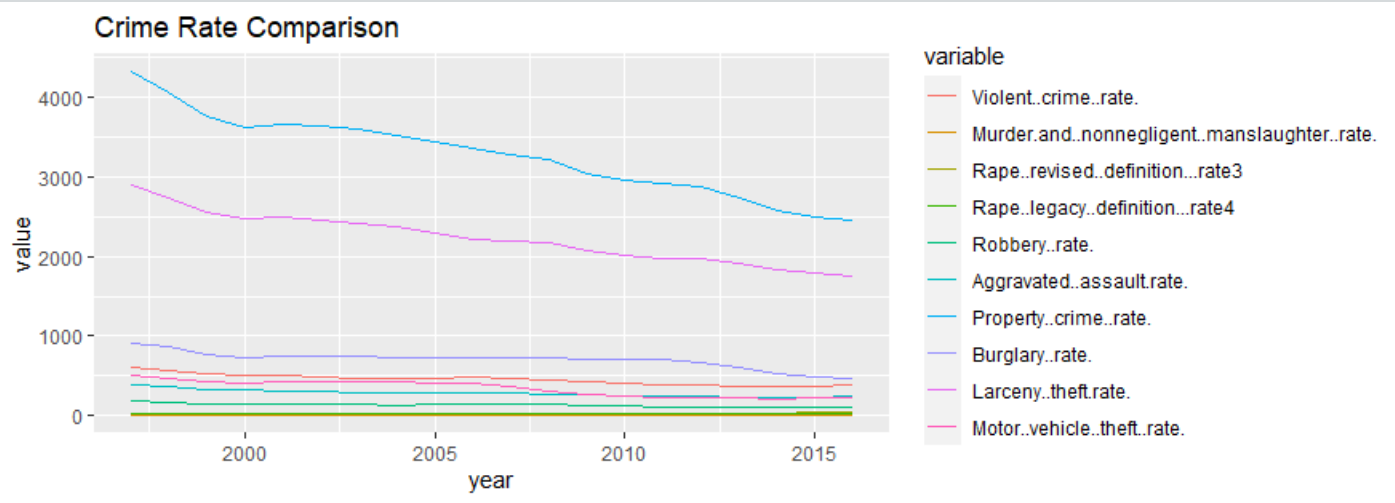
crime\_df\_rates\_violent = crime\_df\_rates[,3:8]

crime\_df\_rates\_property = crime\_df\_rates[,9:ncol(crime\_df\_rates)]

boxplot(crime\_df\_rates\_violent[,c(2:ncol(crime\_df\_rates\_violent))], main="violent crimes", xlab = "crimes", ylab= "values")

boxplot(crime\_df\_rates\_property[,c(2:ncol(crime\_df\_rates\_property))], main="property crines", xlab="crimes", ylab="values")

Plots:







----- END -----