HW1 Biz Analytics

Justin Markey

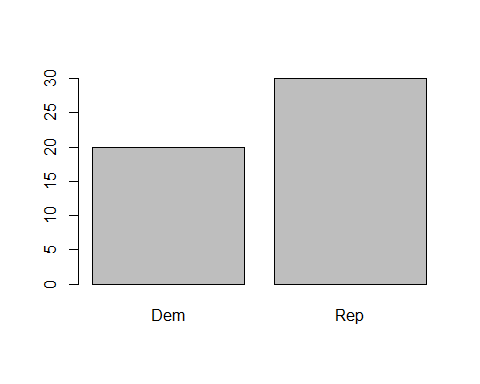
8/30/2021

# Import Data as statedata

statedata <- read.csv('C:/Projects/R/bizanalytics/data/HW1 Data.csv')

# Question 1

proportion <- table(statedata$Presidential.Election)  
  
barplot(proportion)

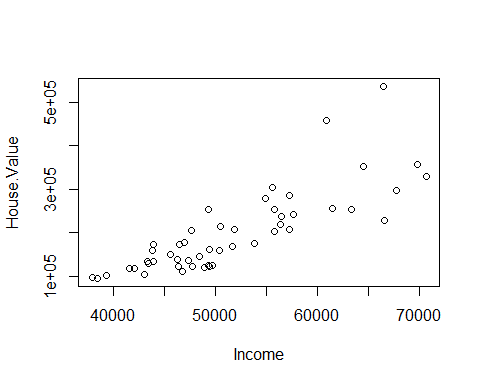


### Conclusion of barplot

Using a barplot to display the categorical data, we can clearly see a republican majority across all 50 states

# Question 2

with (statedata,  
 plot(x=Income, y=House.Value))



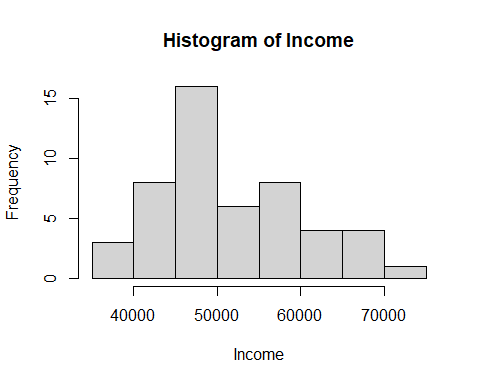
### Conclusion

There is a positive correlation between Income and House Value. This relationship is linear however it is dispersed more the greater the income is.

# Question 3

## Histogram representation of income

with (statedata,  
 hist(Income))

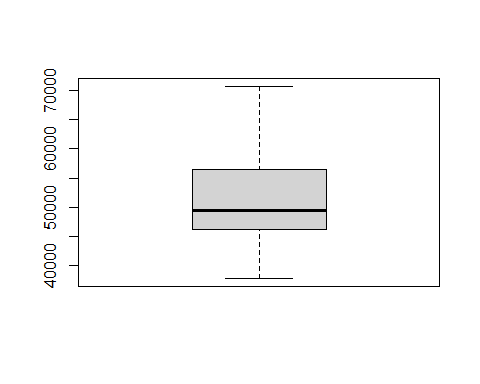


### Conclusion of histogram

The distribution is not symmetrical and is right skewed. Generally the income is crowded around the lower end of the income scale. There seems to be some outliers at the 70000 level.

## Boxplot Representation of income

with (statedata,  
 boxplot(Income))



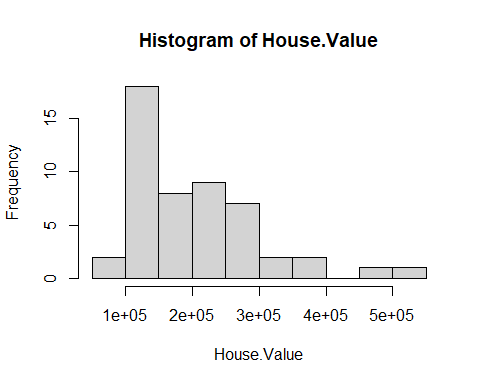
### conclusion of boxplot

The boxplot representation shows that the IQR being quite concentrated. The whiskers are quite long, indicating that some income levels are higher in comparison to the rest of the data. This matches our conclusion of the histogram. We can conclude that the upper range of income contains some outliers.

# Question 4

## Histogram representation of House Value

with (statedata,  
 hist(House.Value))

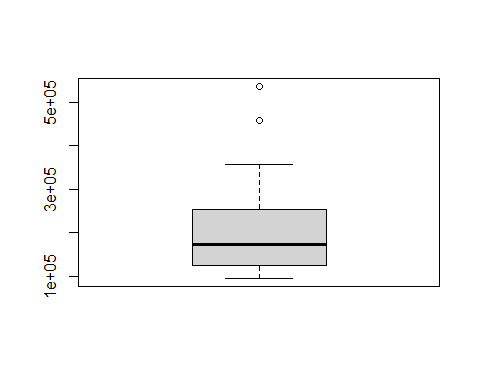


### Conclusion of histogram

The distribution of the house value histogram is not symmetrical and is right skewed. In addition, we can also see outliers in the 500,000 range.

## Boxplot representation of House Value

with (statedata,  
 boxplot(House.Value))



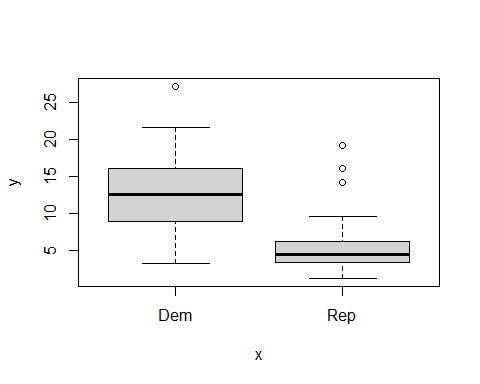
### Conclusion of boxplot

On the boxplot we can also see these outliers as represented by these dots. These are quite extreme given the IQR range and median.

# Question 5

## boxplot with foreign born vs political side

with(statedata,  
 plot(y=Foreign.Born, as.factor(x=Presidential.Election)))



### conclusion of foreign born vs political stance

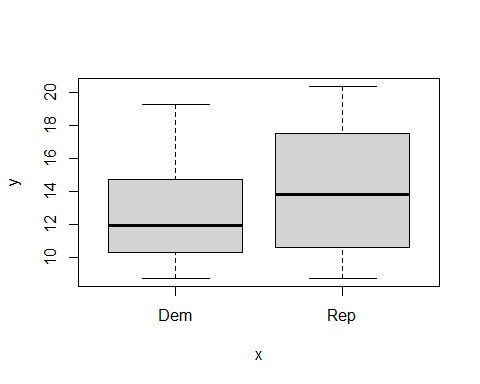
With the IQR and median being higher for dem, it is clear that foreign born people prefer a democratic stance.

## boxplot with HS diploma holders vs presidential side

y axis = percentage of HS diploma degree holders

x axis = political stance

with(statedata,  
 plot(y=HS.Diploma, as.factor(x=Presidential.Election)))



### conclusion of high school degree holder vs political stance

The median for rep is higher than dem however republican boasts an IQR with higher upper quantile. This means that in certain states, there is a higher concentration of republican high school degree holders compared to its democratic counterparts. The whiskers are longer for the rep boxplot indicating that some rep states have a very large percentage of high school degree holders in comparison to the rest of the rep data.

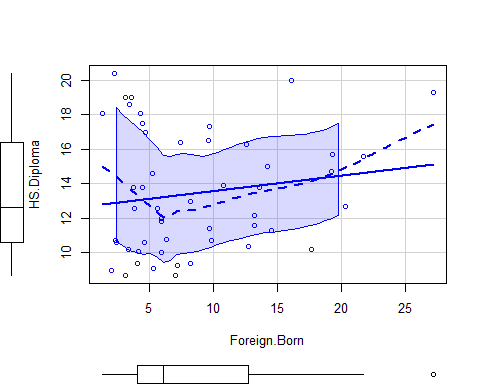
# Question 6

## conditional scatterplot

library(car)

## Loading required package: carData

scatterplot(data=statedata,  
 HS.Diploma~Foreign.Born)

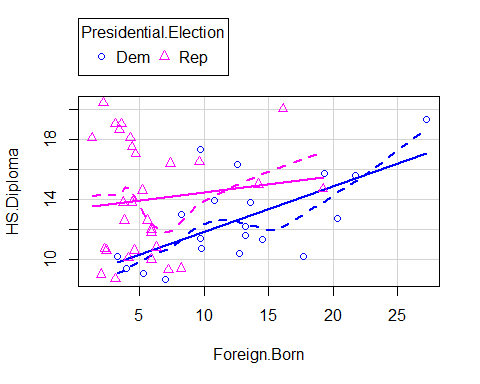


### conclusion of scatterplot comparing HS diploma holders and foreign born

There is a slight correlation between HS diploma holders and being foreign born. However the data is quite dispersed and there is a lot of outliers in the dataset.

## conditional scatterplot influenced by political stance? let’s take a look

scatterplot(data=statedata,  
 HS.Diploma~Foreign.Born | Presidential.Election)



### conclusion of scatterplot factoring in political stance.

There seems to be a weak positive correlation between more dem foreign born having a higher percentage of HS degree education. This is quite a fragile conclusion because of the small sample size of dem data and the amount of outliers in the data. As for rep foreign born, there seems to be no correlation in the data.