By: Nursyafikah Syahmi Bt Mohd Suhaimi (WIE180034)

This is an R markdown file that explains about data visualization

Question 1:

INTRODUCTION (DATA VISUALIZATION)

Data visualization is the representation of data in graphics forms including in maps, graphs and plots.

The uses of data visualizations: 1. Making massive collection of data more understandable. 2. Can use to identify certain patterns in given data. 3. Can identify emerging trends in data 4. To communicate results 5. Suggest modeling strategy

#Question 2:

#Declaring the dataset ChickWeight as chickdata  
chickdata <-datasets::ChickWeight

This dataset is regarding the effects of the chick diets towards its weght.

#Seeing the number of rows and columns in the dataset  
nrow(chickdata)

## [1] 578

ncol(chickdata)

## [1] 4

The number of rows and columns are 578 and 4 resepctively which means that we have tested 578 chicks

#Seeing the variables in datasets  
names(chickdata)

## [1] "weight" "Time" "Chick" "Diet"

The variables in the datasets are the weight, time, chick and diet.

#Seeing the summary of Titanic to explain regarding the dataset.  
summary(chickdata)

## weight Time Chick Diet   
## Min. : 35.0 Min. : 0.00 13 : 12 1:220   
## 1st Qu.: 63.0 1st Qu.: 4.00 9 : 12 2:120   
## Median :103.0 Median :10.00 20 : 12 3:120   
## Mean :121.8 Mean :10.72 10 : 12 4:118   
## 3rd Qu.:163.8 3rd Qu.:16.00 17 : 12   
## Max. :373.0 Max. :21.00 19 : 12   
## (Other):506

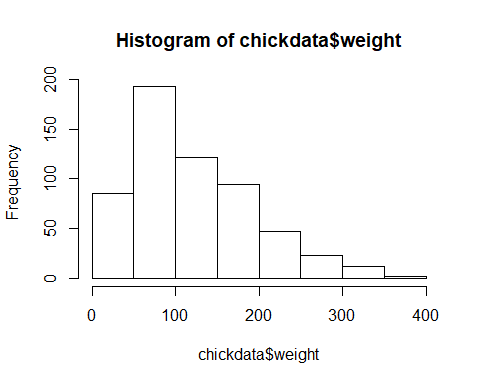
#Number of missing values  
sapply(chickdata, function(x)sum(is.na(x)))

## weight Time Chick Diet   
## 0 0 0 0

There are no missing values in the datasets as there is no na values during search.

#Question 3

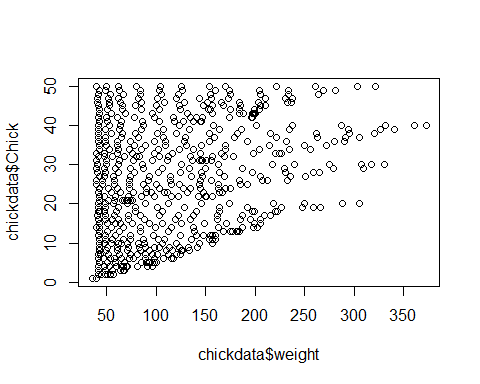
hist(chickdata$weight)

 This is a histogram

Histogram is used to identify distribution and frequency of data.

The histogram below identifies that there are almost 200 chicks that weight around 100 gram. There are less than 50 chicks that weight more than 300 gram.

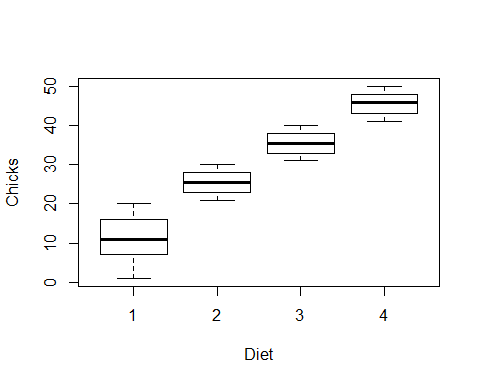
with(chickdata, plot(chickdata$weight,chickdata$Chick))

 This is a scatterplot

The boxplot is used for data inspection and identify the concentration interception of datas.

There is a large concentration of chicks between 50 gram to 100 gram. This indicates the majority of chicks are weight between 50 to 100 grams. There are lossely scattered dots after the weight enters 300 grams. This means that there are very few chicks weight more than 300

chicken <- transform(chickdata, Chick = factor(Diet))  
boxplot(Chick~Diet,chickdata,xlab = "Diet",ylab = "Chicks")

 This is a boxplot

The boxplot is used to visualize the spread of data

The boxplot below identifies that there are wider spread of chicks for the first diet compared to the 2nd and 3rd diet. The boxplot in diet 1 is incilning towards chicks in 0 - 10, making it indicate that the spread of data for the first diet is aldo incling towards 0 - 10 chicks.