ASSINGMENT1

## KUNAL SHARMA

#Load dataset packages

#import the excel file into the global environment

data <- read.csv('A:/DATA\_SETS/a1-cereals.csv')

#head and tail function

head(data)

## Cereal Manufacturer Type Calories Protein Fat Sodium  
## 1 Apple Cinnamon Cheerios G C 110 2 2 180  
## 2 Basic 4 G C 130 3 2 210  
## 3 Cheerios G C 110 6 2 290  
## 4 Cinnamon Toast Crunch G C 120 1 3 210  
## 5 Clusters G C 110 3 2 140  
## 6 Cocoa Puffs G C 110 1 1 180  
## Fiber Carbohydrates Sugars Shelf Potassium Vitamins Weight Cups  
## 1 1.5 10.5 10 1 70 25 1.00 0.75  
## 2 2.0 18.0 8 3 100 25 1.33 0.75  
## 3 2.0 17.0 1 1 105 25 1.00 1.25  
## 4 0.0 13.0 9 2 45 25 1.00 0.75  
## 5 2.0 13.0 7 3 105 25 1.00 0.50  
## 6 0.0 12.0 13 2 55 25 1.00 1.00

tail(data)

## Cereal Manufacturer Type Calories Protein  
## 72 Muesli Raisins, Peaches, & Pecans R C 150 4  
## 73 Rice Chex R C 110 1  
## 74 Wheat Chex R C 100 3  
## 75 Maypo A H 100 4  
## 76 Cream of Wheat (Quick) N H 100 3  
## 77 Quaker Oatmeal Q H 100 5  
## Fat Sodium Fiber Carbohydrates Sugars Shelf Potassium Vitamins Weight  
## 72 3 150 3.0 16 11 3 170 25 -1  
## 73 0 240 0.0 23 2 1 30 25 1  
## 74 1 230 3.0 17 3 1 115 25 1  
## 75 1 0 0.0 16 3 2 95 25 1  
## 76 0 80 1.0 21 0 2 -1 0 1  
## 77 2 0 2.7 -1 -1 1 110 0 1  
## Cups  
## 72 -1.00  
## 73 1.13  
## 74 0.67  
## 75 -1.00  
## 76 1.00  
## 77 0.67

#Working with functions for Stats

min(data$Fat)

## [1] 0

max(data$Fiber)

## [1] 14

range(data$Vitamins)

## [1] 0 100

mean(data$Weight)

## [1] 0.9776623

median(data$Protein)

## [1] 3

# function will give the summary of data

summary(data)

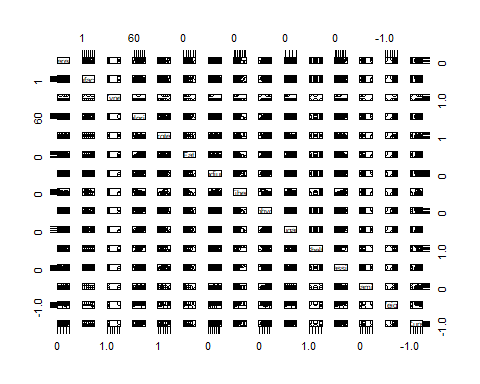
## Cereal Manufacturer Type Calories   
## 100% Bran : 1 A: 1 C:74 Min. : 50.0   
## 100% Natural Bran : 1 G:22 H: 3 1st Qu.:100.0   
## All-Bran : 1 K:23 Median :110.0   
## All-Bran with Extra Fiber: 1 N: 6 Mean :106.9   
## Almond Delight : 1 P: 9 3rd Qu.:110.0   
## Apple Cinnamon Cheerios : 1 Q: 8 Max. :160.0   
## (Other) :71 R: 8   
## Protein Fat Sodium Fiber   
## Min. :1.000 Min. :0.000 Min. : 0.0 Min. : 0.000   
## 1st Qu.:2.000 1st Qu.:0.000 1st Qu.:130.0 1st Qu.: 1.000   
## Median :3.000 Median :1.000 Median :180.0 Median : 2.000   
## Mean :2.545 Mean :1.013 Mean :159.7 Mean : 2.152   
## 3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:210.0 3rd Qu.: 3.000   
## Max. :6.000 Max. :5.000 Max. :320.0 Max. :14.000   
##   
## Carbohydrates Sugars Shelf Potassium   
## Min. :-1.0 Min. :-1.000 Min. :1.000 Min. : -1.00   
## 1st Qu.:12.0 1st Qu.: 3.000 1st Qu.:1.000 1st Qu.: 40.00   
## Median :14.0 Median : 7.000 Median :2.000 Median : 90.00   
## Mean :14.6 Mean : 6.922 Mean :2.208 Mean : 96.08   
## 3rd Qu.:17.0 3rd Qu.:11.000 3rd Qu.:3.000 3rd Qu.:120.00   
## Max. :23.0 Max. :15.000 Max. :3.000 Max. :330.00   
##   
## Vitamins Weight Cups   
## Min. : 0.00 Min. :-1.0000 Min. :-1.0000   
## 1st Qu.: 25.00 1st Qu.: 1.0000 1st Qu.: 0.5000   
## Median : 25.00 Median : 1.0000 Median : 0.7500   
## Mean : 28.25 Mean : 0.9777 Mean : 0.5873   
## 3rd Qu.: 25.00 3rd Qu.: 1.0000 3rd Qu.: 1.0000   
## Max. :100.00 Max. : 1.5000 Max. : 1.5000   
##

#str function will structure the data

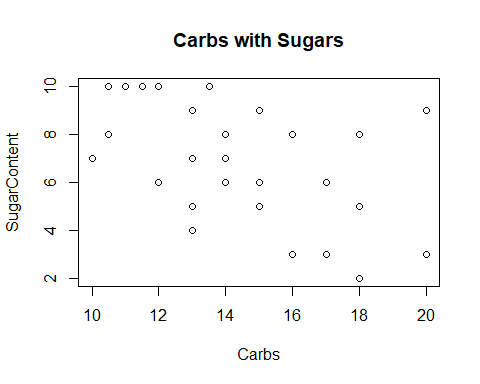
str(data)

## 'data.frame': 77 obs. of 15 variables:  
## $ Cereal : Factor w/ 77 levels "100% Bran","100% Natural Bran",..: 6 8 12 13 14 15 19 23 32 38 ...  
## $ Manufacturer : Factor w/ 7 levels "A","G","K","N",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Type : Factor w/ 2 levels "C","H": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Calories : int 110 130 110 120 110 110 110 100 110 110 ...  
## $ Protein : int 2 3 6 1 3 1 1 2 1 3 ...  
## $ Fat : int 2 2 2 3 2 1 1 1 1 1 ...  
## $ Sodium : int 180 210 290 210 140 180 180 140 280 250 ...  
## $ Fiber : num 1.5 2 2 0 2 0 0 2 0 1.5 ...  
## $ Carbohydrates: num 10.5 18 17 13 13 12 12 11 15 11.5 ...  
## $ Sugars : int 10 8 1 9 7 13 13 10 9 10 ...  
## $ Shelf : int 1 3 1 2 3 2 2 3 2 1 ...  
## $ Potassium : int 70 100 105 45 105 55 65 120 45 90 ...  
## $ Vitamins : int 25 25 25 25 25 25 25 25 25 25 ...  
## $ Weight : num 1 1.33 1 1 1 1 1 1 1 1 ...  
## $ Cups : num 0.75 0.75 1.25 0.75 0.5 1 1 0.75 0.75 0.75 ...

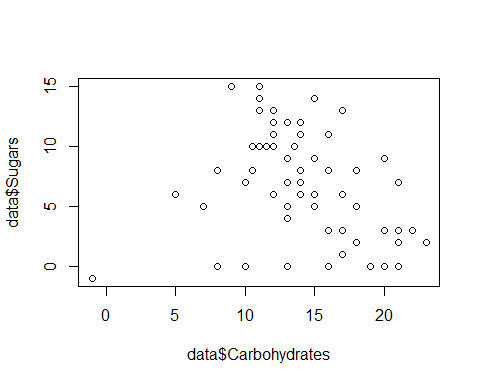
plot(data)

 #X-Y plot for two quantitative variables(Scatterplot)

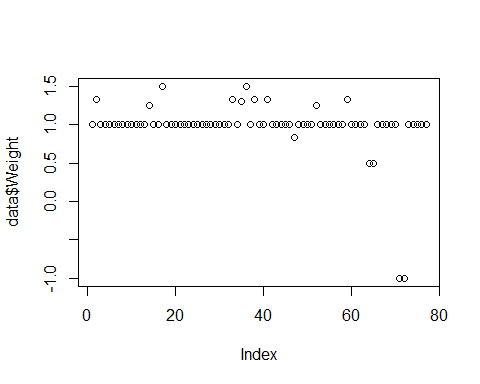
plot(x= data$Carbohydrates, y=data$Sugars,  
 xlab = "Carbs",  
 ylab = "SugarContent",  
 xlim = c(10,20),  
 ylim = c(2,10),  
 main = "Carbs with Sugars")

 #Add some Options

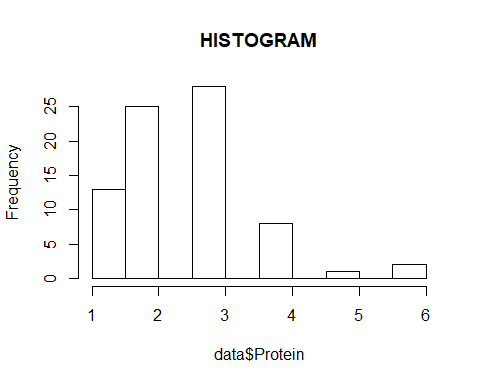
plot(data$Carbohydrates,data$Sugars)

 #Plotting a quantitative variable Audience\_Size

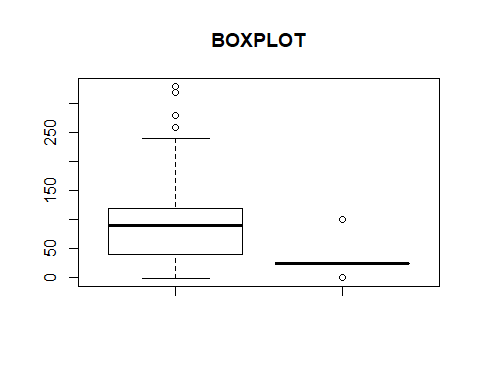
plot(data$Weight)

 #LineplotHistogram,Boxplot

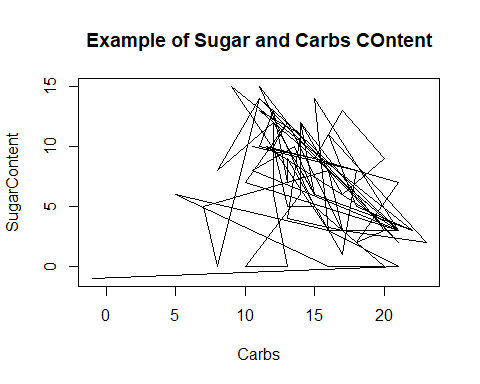
hist(data$Protein,main = "HISTOGRAM")



boxplot(data$Potassium,data$Vitamins,main="BOXPLOT")



plot(x= data$Carbohydrates, y=data$Sugars,  
  
 xlab = "Carbs",  
 ylab = "SugarContent",  
 type="l",main="Example of Sugar and Carbs COntent")



#The End