R\_markdown

Sayantan

April 28, 2018

#reading the cereal data into r  
library(readxl)

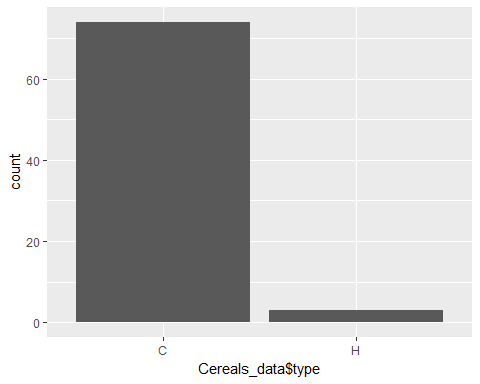
## Warning: package 'readxl' was built under R version 3.4.4

Cereals\_data <- read\_xlsx("cereals\_practice.xlsx",sheet = 1)

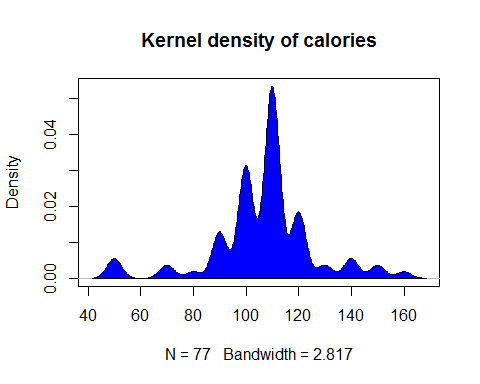
#lets see how many cereals are in each type  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.4

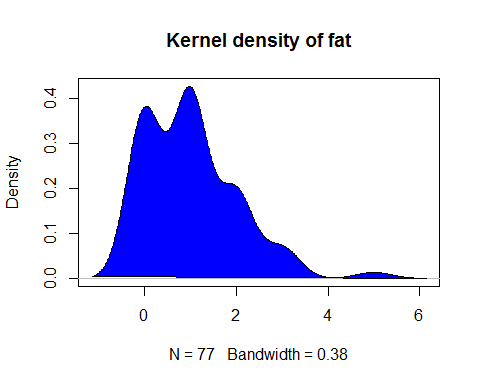
ggplot(data.frame(Cereals\_data), aes(x=Cereals\_data$type)) + geom\_bar()



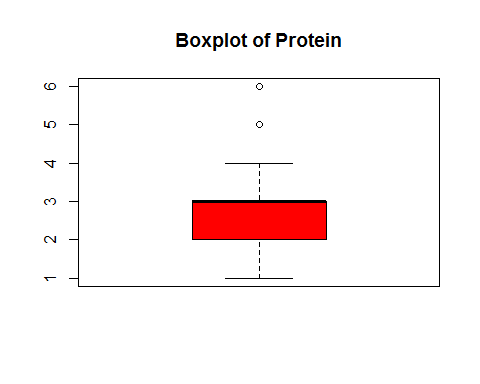
#kernel density of calories and fat  
d <- density(Cereals\_data$calories)  
plot(d,main = "Kernel density of calories")  
polygon(d,col = "blue",border = "black")



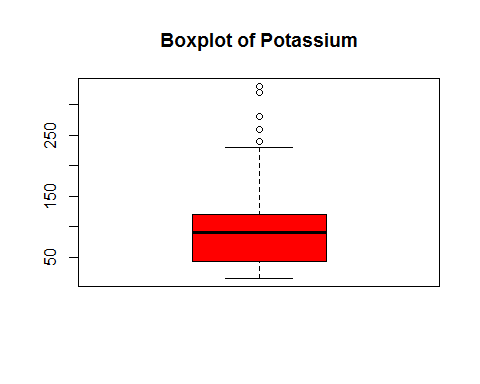
d1 <- density(Cereals\_data$fat)  
plot(d1,main = "Kernel density of fat")  
polygon(d1,col = "blue",border = "black")



#how calories and potassium are distributed  
boxplot(Cereals\_data$protein,col = "red",main = "Boxplot of Protein")



boxplot(Cereals\_data$potass,col = "red",main = "Boxplot of Potassium")



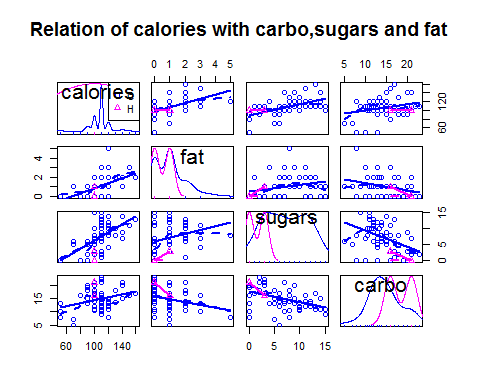
#how calories are related to fat sugars and carbohydrates  
library(car)

## Warning: package 'car' was built under R version 3.4.4

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.4.4

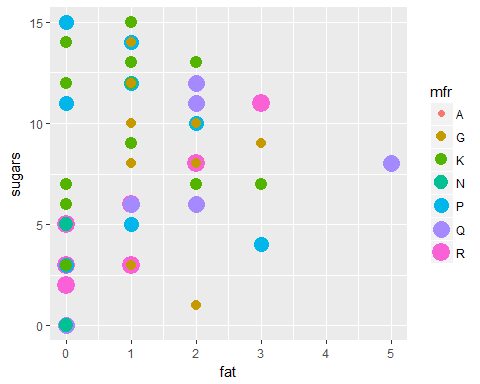
scatterplotMatrix(~calories+fat+sugars+carbo|type,data = Cereals\_data,main = "Relation of calories with carbo,sugars and fat")



#which mfr has the least amount of fat and sugars contents so that we can conclude which is the healthiest among them  
ggplot(data = Cereals\_data)+geom\_point(mapping = aes(x = fat,y = sugars,color = mfr,size = mfr))

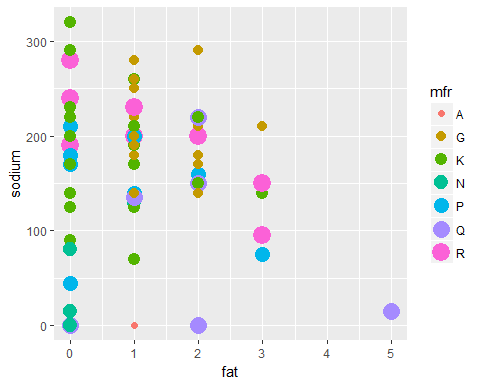
## Warning: Using size for a discrete variable is not advised.

## Warning: Removed 1 rows containing missing values (geom\_point).

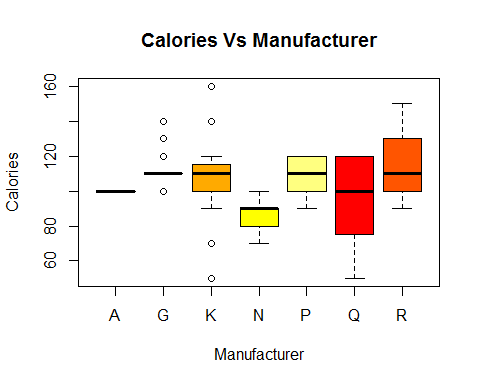


#now with fat and sodium  
ggplot(data = Cereals\_data)+geom\_point(mapping = aes(x = fat,y = sodium,color = mfr,size = mfr))

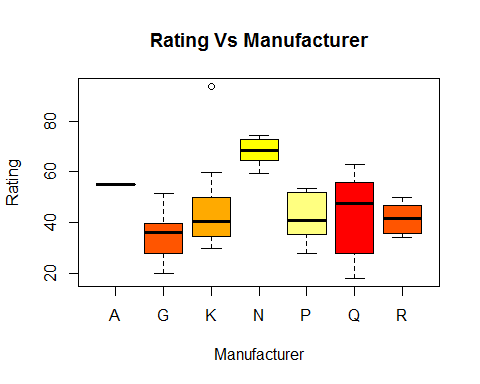
## Warning: Using size for a discrete variable is not advised.



#mfr which sells cereals with high calories  
Cereals\_data$mfr = as.factor(Cereals\_data$mfr)  
boxplot(calories~mfr,data = Cereals\_data,main = "Calories Vs Manufacturer",xlab = "Manufacturer",ylab = "Calories",col = heat.colors(5))



#now mfr and ratings  
boxplot(rating~mfr,data = Cereals\_data,main = "Rating Vs Manufacturer",xlab = "Manufacturer",ylab = "Rating",col = heat.colors(5))



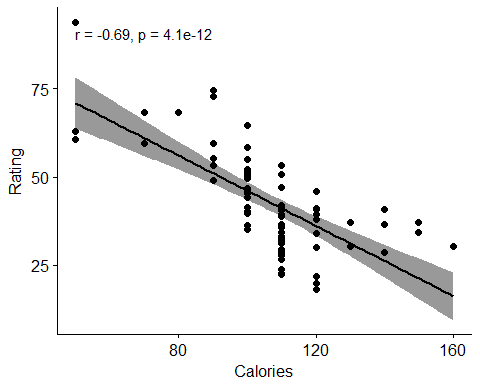
#now lets see what relation rating has with calories  
library('ggpubr')

## Warning: package 'ggpubr' was built under R version 3.4.4

## Loading required package: magrittr

## Warning: package 'magrittr' was built under R version 3.4.4

ggscatter(Cereals\_data,x = "calories",y = "rating",add = "reg.line",conf.int = TRUE,cor.coef = TRUE,cor.method = "pearson",xlab = "Calories",ylab = "Rating")



## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.