Assignment 1

Rob DiVincenzo 09-11-2023 BA 64060

This is the submission for Assignment 1.

#Tasks 1 and 2, download dataset and import into R. I'm using the Cereals.csv dataset located in our course CSV Files.  
DF=read.csv("./Cereals.csv")   
library(dplyr) # Install the dplyr library

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

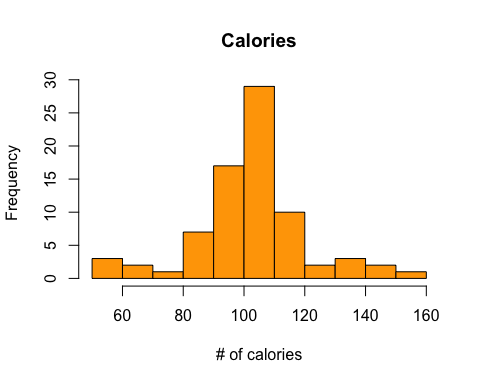
#Task 3, print out descriptive statistics  
summary(DF)

## name mfr type calories   
## Length:77 Length:77 Length:77 Min. : 50.0   
## Class :character Class :character Class :character 1st Qu.:100.0   
## Mode :character Mode :character Mode :character Median :110.0   
## Mean :106.9   
## 3rd Qu.:110.0   
## Max. :160.0   
##   
## 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   
##   
## carbo sugars potass vitamins   
## Min. : 5.0 Min. : 0.000 Min. : 15.00 Min. : 0.00   
## 1st Qu.:12.0 1st Qu.: 3.000 1st Qu.: 42.50 1st Qu.: 25.00   
## Median :14.5 Median : 7.000 Median : 90.00 Median : 25.00   
## Mean :14.8 Mean : 7.026 Mean : 98.67 Mean : 28.25   
## 3rd Qu.:17.0 3rd Qu.:11.000 3rd Qu.:120.00 3rd Qu.: 25.00   
## Max. :23.0 Max. :15.000 Max. :330.00 Max. :100.00   
## NA's :1 NA's :1 NA's :2   
## shelf weight cups rating   
## Min. :1.000 Min. :0.50 Min. :0.250 Min. :18.04   
## 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:0.670 1st Qu.:33.17   
## Median :2.000 Median :1.00 Median :0.750 Median :40.40   
## Mean :2.208 Mean :1.03 Mean :0.821 Mean :42.67   
## 3rd Qu.:3.000 3rd Qu.:1.00 3rd Qu.:1.000 3rd Qu.:50.83   
## Max. :3.000 Max. :1.50 Max. :1.500 Max. :93.70   
##

#Task 4 - data transformation, I'm going to summarize the number of different mfr in the dataset  
  
task4\_mfrs <- DF %>% group\_by(mfr) %>% summarise(manufacturerers\_count=n())  
  
print(task4\_mfrs)

## # A tibble: 7 × 2  
## mfr manufacturerers\_count  
## <chr> <int>  
## 1 A 1  
## 2 G 22  
## 3 K 23  
## 4 N 6  
## 5 P 9  
## 6 Q 8  
## 7 R 8

#Task 5 - Plot one qunatittative variable, and one scatterplot  
  
hist(DF$calories,  
 main="Calories",  
 xlab="# of calories",  
 col="orange",  
 freq=TRUE)



plot(DF$calories, DF$sugars)

