

FML Assignment 1

Jyothsna P - 811251679

1/28/2023

R Markdown

Install Readr Package #Install readr package to use read.csv file from the local machine #repos URL is required to use the install packages in the R markdown file

```
#install.packages("readr", repos = "http://cran.us.r-project.org")
#install.packages("tidyverse")
#library(tidyverse)
#library(readr)
#install.packages("pscl", repos = "https://cran.rstudio.com")
```

Load Cereals Data #read.csv function is useful to load the data available in the csv file to r environment #Cereals.csv file is downloaded from the FML course in Canvas

```
#install.packages("ggplot2")

cerealsData <- read.csv("C:/Users/peddi/Downloads/Cereals.csv", header = TRUE, sep = ",")
```

#Class of the Name column

```
class(cerealsData$name)
```

```
## [1] "character"
```

Descriptive Statistics

#Check if any null values are there in the cereals dataset

```
any(is.na(cerealsData))
```

```
## [1] TRUE
```

#Get Descriptive Statistics of all the columns in the dataset

```
summary(cerealsData)
```

```
##      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
##
```

Get Descriptive Statistics of the Calories column in the dataset

```
summary(cerealsData$calories)
```

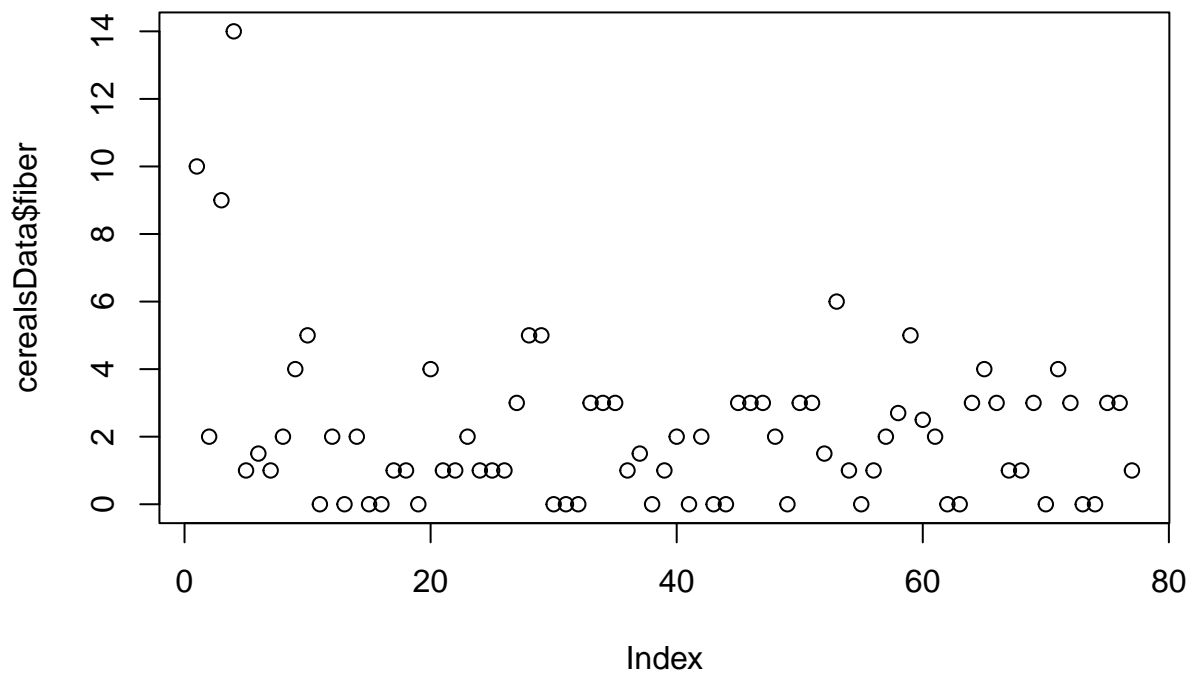
```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      50.0  100.0   110.0   106.9   110.0   160.0
```

Transformations

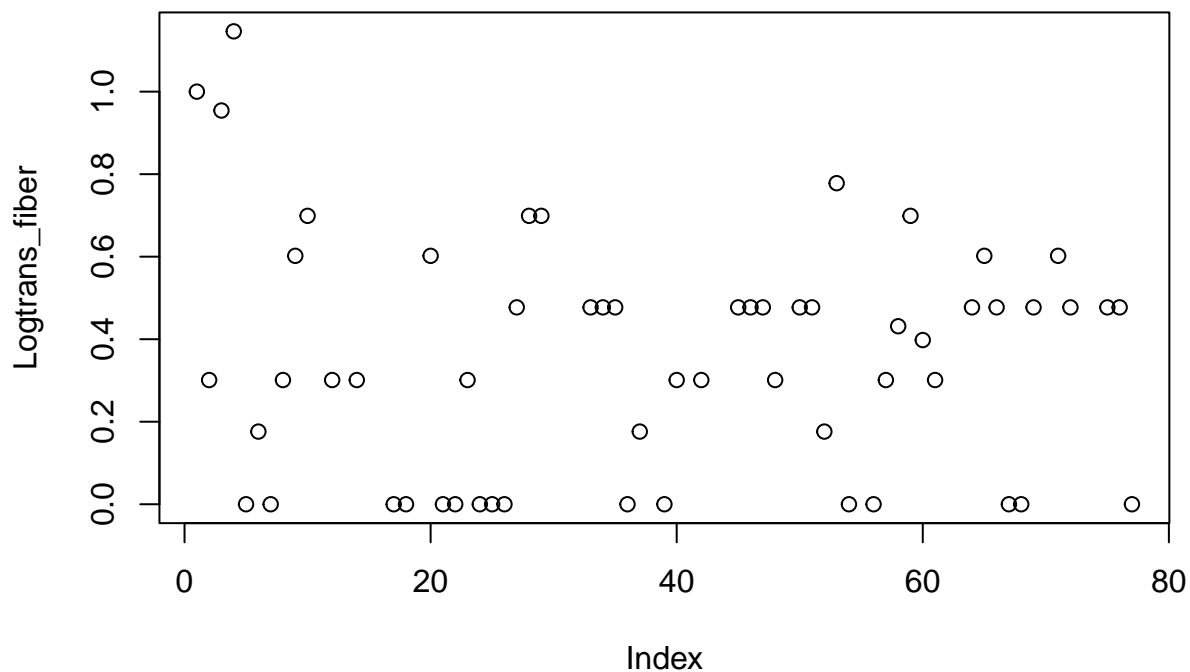
Transform column name type in Cereals dataset to factor

```
cerealsData$type <- factor(cerealsData$type)
```

```
plot(cerealsData$fiber)
```



```
Logtrans_fiber=log10(cerealsData$fiber)
plot(Logtrans_fiber)
```



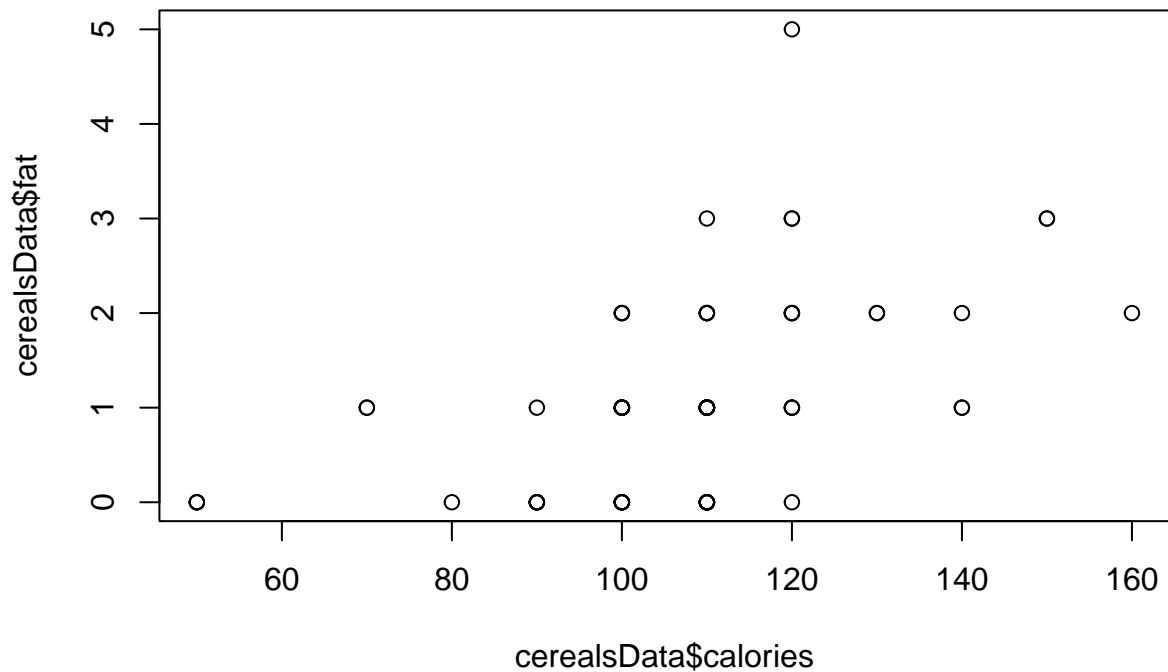
```
Logtrans_sodium=sqrt(cerealsData$sodium)
Logtrans_sodium
```

```
## [1] 11.401754  3.872983 16.124515 11.832160 14.142136 13.416408 11.180340
## [8] 14.491377 14.142136 14.491377 14.832397 17.029386 14.491377 11.832160
## [15] 13.416408 16.733201 17.029386  9.486833 13.416408 11.832160  8.944272
## [22] 14.832397 11.832160 13.784049 11.180340 14.142136  0.000000 12.649111
## [29] 15.491933 11.618950  6.708204 16.733201 11.832160 13.038405  8.660254
## [36] 14.832397 15.811388 13.416408 13.038405 13.038405 16.124515 12.247449
## [43] 13.416408  0.000000  9.746794 12.247449 12.247449 14.832397 13.784049
## [50] 14.832397 13.038405 13.038405 14.142136 17.888544  0.000000  0.000000
## [57] 11.618950  0.000000 14.491377 11.832160  0.000000 15.491933 17.029386
## [64]  0.000000  0.000000  0.000000  8.366600 15.165751  3.872983 14.142136
## [71] 13.784049 14.142136 15.811388 11.832160 15.165751 14.142136 14.142136
```

Including Plots

Scatter Plot for Calories and fat attributes

```
plot(cerealsData$calories,cerealsData$fat)
```



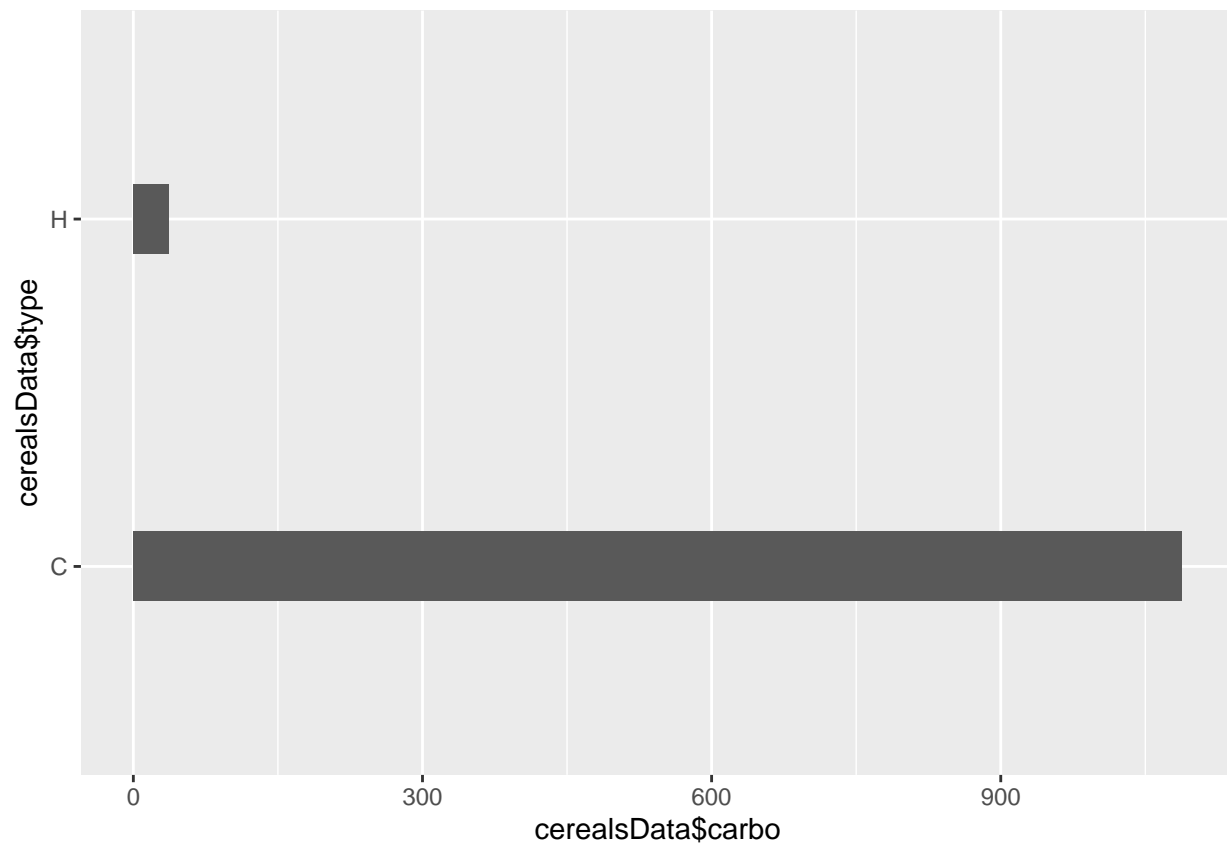
Plotting ggplot for categorical attribute type and carbo variables

```
library(ggplot2)
ggplot(cerealsData, aes(x=cerealsData$carbo, y=cerealsData$type)) +
  geom_bar(stat = "identity", width=0.2)
```

```
## Warning: Use of 'cerealsData$carbo' is discouraged.
## i Use 'carbo' instead.
```

```
## Warning: Use of 'cerealsData$type' is discouraged.
## i Use 'type' instead.
```

```
## Warning: Removed 1 rows containing missing values ('position_stack()').
```



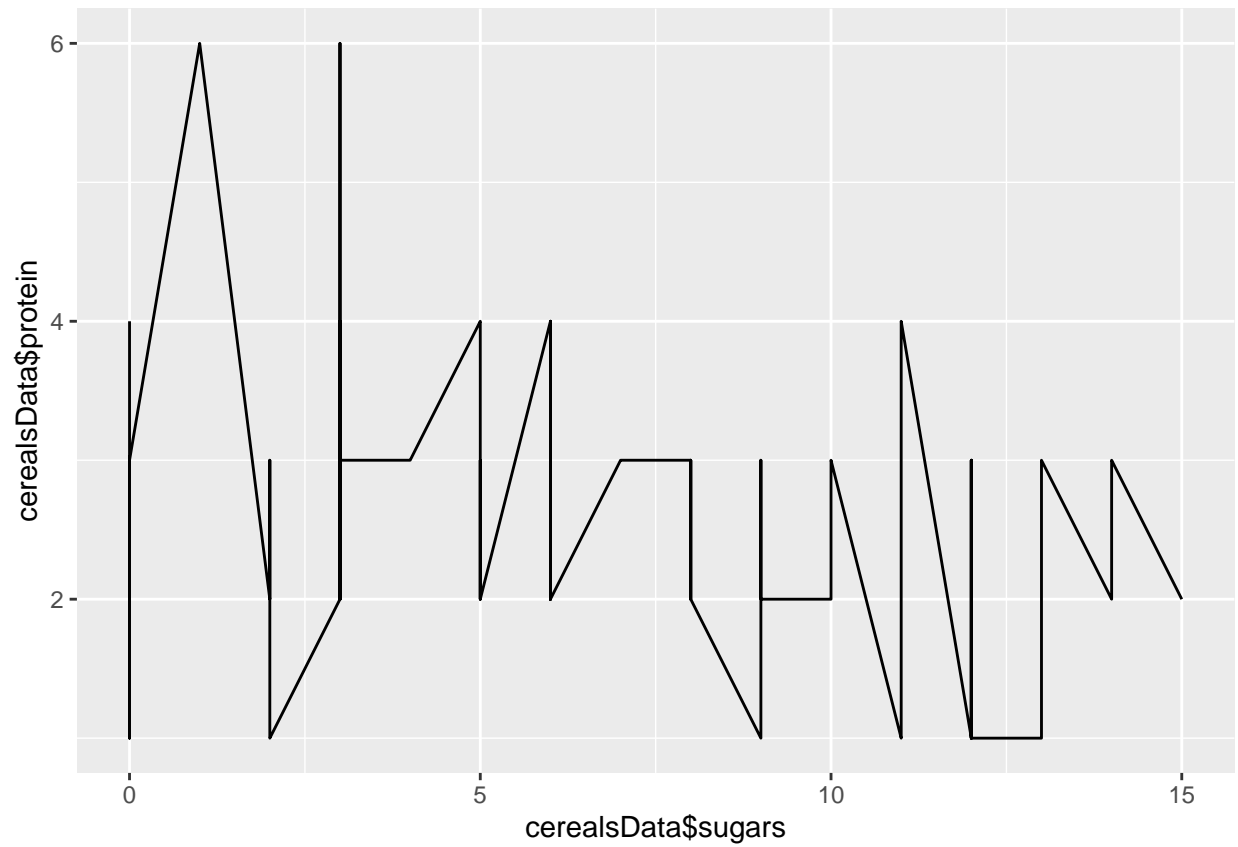
Plotting ggplot for sugars and protein variables

```
library(ggplot2)
ggplot(cerealsData, aes(x=cerealsData$sugars, y=cerealsData$protein)) +
  geom_line()
```

```
## Warning: Use of 'cerealsData$sugars' is discouraged.
## i Use 'sugars' instead.
```

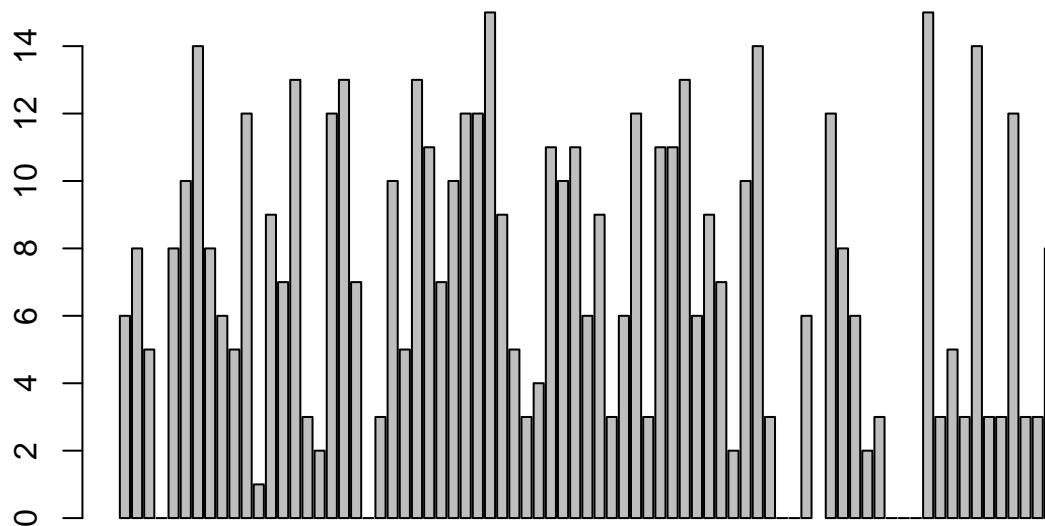
```
## Warning: Use of 'cerealsData$protein' is discouraged.
## i Use 'protein' instead.
```

```
## Warning: Removed 1 row containing missing values ('geom_line()').
```



Barplot for sugars variable

```
barplot(cerealsData$sugars)
```



plotting ggplot for mfr and type variables

```
ggplot(cerealsData, aes(mfr, ..count..)) + geom_bar(aes(fill = type), position = "dodge")
```

```
## Warning: The dot-dot notation ('..count..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(count)' instead.
```