Mcdonald Assignment

Statistical Methods for Decision Making

Table of Content

1. Project Objective

The objective of the report is to explore the Mcdonald data set ("Mcdonald.csv") in R and generate insights about the data set. This exploration report will consists of the following:
- Importing the dataset in R
- Understanding the structure of dataset
- Graphical exploration
- Descriptive statistics
- Insights from the dataset

2. Assumptions

Following assumption we made for this analysis
- The Data Provided to us was not tempered.
- Linearity - Linearity assumes a straight line relationship between each of the two variables.
- Homoscedasticity - Homoscedasticity assumes that data is equally distributed about the regression line.

3. Exploratory Data Analysis Step by step approach

A Typical Data exploration activity consists of the following steps:
1. Environment Set up and Data Import
2. Variable Identification
3. Univariate Analysis
4. Bi-Variate Analysis
5. Outlier Identification
6. Feature Creation & Exploration

We shall follow these steps in exploring the provided dataset.

3.1 Environment Set up and Data Import

3.1.1 Install necessary Packages and Invoke Libraries

Following are the Libraries are used in the analysis

Package	Library
dplyr	dplyr

ggplot2	ggplot2
corrplot	corrplot
DataExplorer	DataExplorer

Code for installing library

```
#Libraries Required for Analysis
library(dplyr)
library(ggplot2)
library(corrplot)
library(DataExplorer)
```

Please refer to Appendix A for Source Code.

3.1.2 Set up working Directory

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.

Command for installing library

```
#Setting the Working Directory
setwd("C:/Users/dignitas/Desktop/Learning/Mcdonald/Home 1")
getwd()
```

Please refer to Appendix A for Source Code.

3.1.3 Import and Read the Dataset

The given dataset is in .csv format. Hence, the command 'read.csv' is used for importing the file.

Command for Read the Dataset

```
#Importing the Data
myData = read.csv("Mcdonald.csv")
```

Please refer to Appendix A for Source Code.

3.2 Variable Identification

Functions is used for variable identifications with there functions:
- class(myData): To identify the class of Data
- str(myData): compactly display the (abbreviated) contents of lists.
- names(myData): Names of DataFrame variable
- dim(myData): Dimensions of Dataframe
- head(myData): Display top 6 elements of Variables
- tail(myData): Display last 6 elements of variables

3.2.1 Variable Identification – Inferences

Our Data contain 260 obs. of 24 variables with 3 variables as factors while others are numerical data.

Column name of our Data are:
- "Category"
- "Item"
- "Serving.Size"
- "Calories"
- "Calories.from.Fat"
- "Total.Fat"
- "Total.Fat.PDV"
- "Saturated.Fat"
- "Saturated.Fat.PDV"
- "Trans.Fat"
- "Cholesterol"
- "Cholesterol.PDV"
- "Sodium"
- "Sodium.PDV"
- "Carbohydrates"
- "Carbohydrates.PDV"
- "Dietary.Fiber"
- "Dietary.Fiber.PDV"
- "Sugars"
- "Protein"
- "Vitamin.A.PDV"
- "Vitamin.C.PDV"
- "Calcium.PDV"
- "Iron.PDV"

We also checked the top 6 and last 6 elements of each variable with command head and tail

Command for variable identifications and Output

```
> #General Data Analysis
> ##Check the Class of Data
> class(myData)
[1] "data.frame"
>
> ##First Inspection of Dataset using str
> str(myData)
'data.frame':   260 obs. of  24 variables:
 $ Category        : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Item            : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76 77 228 229 230 245 12 11 14 13
...
```

```
 $ Serving.Size     : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54 42 69 69 83 63 72 65 73 ...
 $ Calories         : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat: int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat        : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat.PDV    : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat    : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat.PDV: int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat        : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol      : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol.PDV  : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium           : int  750 770 780 860 880 960 1300 1410 1300 1420 ...
 $ Sodium.PDV       : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates    : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates.PDV: int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber    : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber.PDV: int  17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars           : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein          : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A.PDV    : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C.PDV    : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium.PDV      : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron.PDV         : int  15 8 10 15 10 20 15 20 10 15 ...
>
> ## Find the name of variable
> names(myData)
 [1] "Category"          "Item"              "Serving.Size"      "Calories"
 [5] "Calories.from.Fat" "Total.Fat"         "Total.Fat.PDV"     "Saturated.Fat"
 [9] "Saturated.Fat.PDV" "Trans.Fat"         "Cholesterol"       "Cholesterol.PDV"
[13] "Sodium"            "Sodium.PDV"        "Carbohydrates"     "Carbohydrates.PDV"
[17] "Dietary.Fiber"     "Dietary.Fiber.PDV" "Sugars"            "Protein"
[21] "Vitamin.A.PDV"     "Vitamin.C.PDV"     "Calcium.PDV"       "Iron.PDV"
>
> ##find the dimension of Data
> dim(myData)
[1] 260  24
>
> ##find first 6 elements of Data
> head(myData)
  Category                            Item   Serving.Size Calories Calories.from.Fat Total.Fat
1 Breakfast                   Egg McMuffin 4.8 oz (136 g)      300               120        13
2 Breakfast             Egg White Delight 4.8 oz (135 g)      250                70         8
3 Breakfast               Sausage McMuffin 3.9 oz (111 g)      370               200        23
4 Breakfast      Sausage McMuffin with Egg 5.7 oz (161 g)      450               250        28
5 Breakfast Sausage McMuffin with Egg Whites 5.7 oz (161 g)    400               210        23
6 Breakfast            Steak & Egg McMuffin 6.5 oz (185 g)      430               210        23
  Total.Fat.PDV Saturated.Fat Saturated.Fat.PDV Trans.Fat Cholesterol Cholesterol.PDV Sodium
1            20             5                25         0         260              87    750
2            12             3                15         0          25               8    770
3            35             8                42         0          45              15    780
4            43            10                52         0         285              95    860
5            35             8                42         0          50              16    880
6            36             9                46         1         300             100    960
  Sodium.PDV Carbohydrates Carbohydrates.PDV Dietary.Fiber Dietary.Fiber.PDV Sugars Protein
1         31            31                10             4                17      3      17
2         32            30                10             4                17      3      18
3         33            29                10             4                17      2      14
```

```
4         36           30             10             4             17      2    21
5         37           30             10             4             17      2    21
6         40           31             10             4             18      3    26
  Vitamin.A.PDV Vitamin.C.PDV Calcium.PDV Iron.PDV
1          10            0          25       15
2           6            0          25        8
3           8            0          25       10
4          15            0          30       15
5           6            0          25       10
6          15            2          30       20
>
> ##find last 5 elements of Data
> tail(myData)
            Category                                              Item    Serving.Size Calories
255 Smoothies & Shakes            McFlurry with M&Mâ€TMs Candies (Snack)  7.3 oz (207 g)     430
256 Smoothies & Shakes              McFlurry with Oreo Cookies (Small) 10.1 oz (285 g)     510
257 Smoothies & Shakes             McFlurry with Oreo Cookies (Medium) 13.4 oz (381 g)     690
258 Smoothies & Shakes              McFlurry with Oreo Cookies (Snack)  6.7 oz (190 g)     340
259 Smoothies & Shakes McFlurry with Reese's Peanut Butter Cups (Medium) 14.2 oz (403 g)     810
260 Smoothies & Shakes  McFlurry with Reese's Peanut Butter Cups (Snack)  7.1 oz (202 g)     410
    Calories.from.Fat Total.Fat Total.Fat.PDV Saturated.Fat Saturated.Fat.PDV Trans.Fat Cholesterol
255               140        15            24            10                48       0.0          35
256               150        17            26             9                44       0.5          45
257               200        23            35            12                58       1.0          55
258               100        11            17             6                29       0.0          30
259               290        32            50            15                76       1.0          60
260               150        16            25             8                38       0.0          30
    Cholesterol.PDV Sodium Sodium.PDV Carbohydrates Carbohydrates.PDV Dietary.Fiber
255              11    120          5            64                21             1
256              14    280         12            80                27             1
257              19    380         16           106                35             1
258               9    190          8            53                18             1
259              20    400         17           114                38             2
260              10    200          8            57                19             1
    Dietary.Fiber.PDV Sugars Protein Vitamin.A.PDV Vitamin.C.PDV Calcium.PDV Iron.PDV
255                 4     59       9            10             0          30        4
256                 4     64      12            15             0          40        8
257                 5     85      15            20             0          50       10
258                 2     43       8            10             0          25        6
259                 9    103      21            20             0          60        6
260                 5     51      10            10             0          30        4
```

3.3 Univariate Analysis

"summary" is used to find the minimum, 1st quartile, mean, median , 3rd quartile and max of the Data variables.
 "plot_missing" is used to plot to find the missing value if exist in the myData. In our data no value was missing.
"plot_histogram" is used to plot the histogram of numeric variables.

Command for Univariate analysis with output

```
> ##find summary of myData to get Min,median,Mean and Max with First and 3rd quartile.
> summary(myData)
           Category                        Item              Serving.Size
 Coffee & Tea        :95   1% Low Fat Milk Jug              :  1   16 fl oz cup: 45
 Breakfast           :42   Apple Slices                     :  1   12 fl oz cup: 38
 Smoothies & Shakes:28     Bacon Buffalo Ranch McChicken    :  1   22 fl oz cup: 20
 Beverages           :27   Bacon Cheddar McChicken          :  1   20 fl oz cup: 16
 Chicken & Fish      :27   Bacon Clubhouse Burger           :  1   21 fl oz cup:  7
 Beef & Pork         :15   Bacon Clubhouse Crispy Chicken Sandwich:  1   30 fl oz cup:  7
 (Other)             :26   (Other)                          :254   (Other)     :127
   Calories        Calories.from.Fat    Total.Fat        Total.Fat.PDV    Saturated.Fat
 Min.   :   0.0   Min.   :   0.0   Min.   :  0.000   Min.   :  0.00   Min.   : 0.000
 1st Qu.: 210.0   1st Qu.:  20.0   1st Qu.:  2.375   1st Qu.:  3.75   1st Qu.: 1.000
 Median : 340.0   Median : 100.0   Median : 11.000   Median : 17.00   Median : 5.000
 Mean   : 368.3   Mean   : 127.1   Mean   : 14.165   Mean   : 21.82   Mean   : 6.008
 3rd Qu.: 500.0   3rd Qu.: 200.0   3rd Qu.: 22.250   3rd Qu.: 35.00   3rd Qu.:10.000
 Max.   :1880.0   Max.   :1060.0   Max.   :118.000   Max.   :182.00   Max.   :20.000

 Saturated.Fat.PDV   Trans.Fat        Cholesterol      Cholesterol.PDV     Sodium
 Min.   :  0.00   Min.   :0.0000   Min.   :  0.00   Min.   :  0.00   Min.   :   0.0
 1st Qu.:  4.75   1st Qu.:0.0000   1st Qu.:  5.00   1st Qu.:  2.00   1st Qu.: 107.5
 Median : 24.00   Median :0.0000   Median : 35.00   Median : 11.00   Median : 190.0
 Mean   : 29.97   Mean   :0.2038   Mean   : 54.94   Mean   : 18.39   Mean   : 495.8
 3rd Qu.: 48.00   3rd Qu.:0.0000   3rd Qu.: 65.00   3rd Qu.: 21.25   3rd Qu.: 865.0
 Max.   :102.00   Max.   :2.5000   Max.   :575.00   Max.   :192.00   Max.   :3600.0

   Sodium.PDV       Carbohydrates    Carbohydrates.PDV Dietary.Fiber    Dietary.Fiber.PDV
 Min.   :  0.00   Min.   :  0.00   Min.   : 0.00   Min.   :0.000   Min.   : 0.000
 1st Qu.:  4.75   1st Qu.: 30.00   1st Qu.:10.00   1st Qu.:0.000   1st Qu.: 0.000
 Median :  8.00   Median : 44.00   Median :15.00   Median :1.000   Median : 5.000
 Mean   : 20.68   Mean   : 47.35   Mean   :15.78   Mean   :1.631   Mean   : 6.531
 3rd Qu.: 36.25   3rd Qu.: 60.00   3rd Qu.:20.00   3rd Qu.:3.000   3rd Qu.:10.000
 Max.   :150.00   Max.   :141.00   Max.   :47.00   Max.   :7.000   Max.   :28.000

    Sugars           Protein         Vitamin.A.PDV    Vitamin.C.PDV    Calcium.PDV
 Min.   :  0.00   Min.   : 0.00   Min.   :  0.00   Min.   :  0.000   Min.   : 0.00
 1st Qu.:  5.75   1st Qu.: 4.00   1st Qu.:  2.00   1st Qu.:  0.000   1st Qu.: 6.00
 Median : 17.50   Median :12.00   Median :  8.00   Median :  0.000   Median :20.00
 Mean   : 29.42   Mean   :13.34   Mean   : 13.43   Mean   :  8.535   Mean   :20.97
 3rd Qu.: 48.00   3rd Qu.:19.00   3rd Qu.: 15.00   3rd Qu.:  4.000   3rd Qu.:30.00
 Max.   :128.00   Max.   :87.00   Max.   :170.00   Max.   :240.000   Max.   :70.00

    Iron.PDV
 Min.   : 0.000
 1st Qu.: 0.000
 Median : 4.000
 Mean   : 7.735
 3rd Qu.:15.000
 Max.   :40.000

>
> ##plot the missing value
> plot_missing(myData)
> ##Plot histogram of the data
> plot_histogram(myData,nrow = 4,ncol = 4)
```

plot the missing value
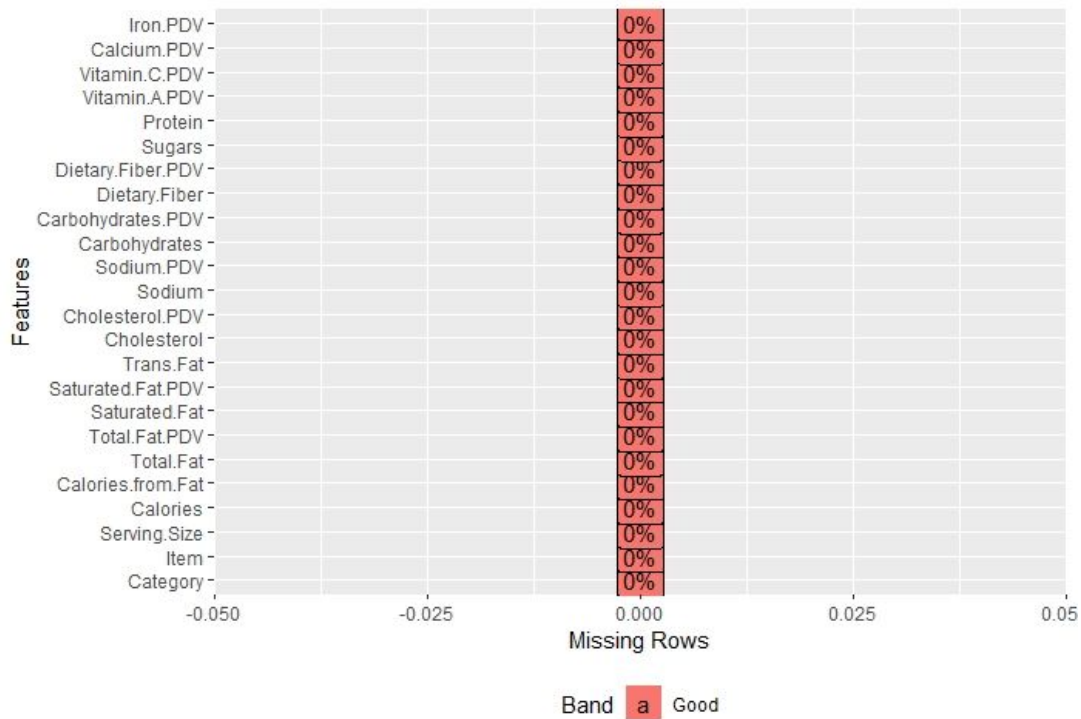
Histogram Plot
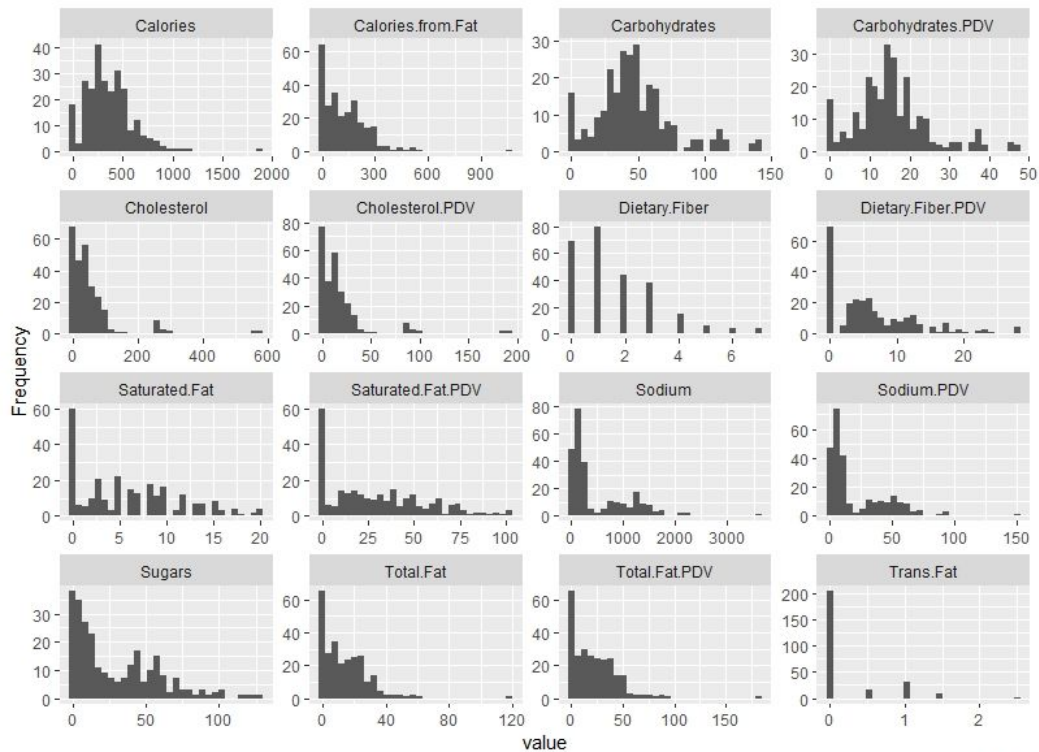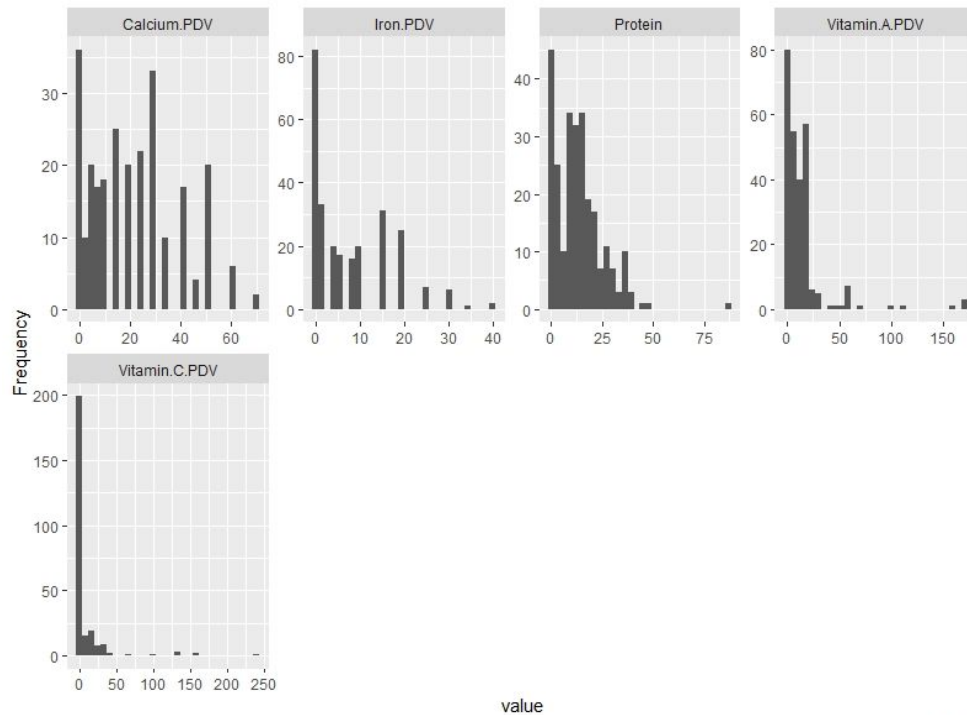
3.4 Bi-Variate Analysis

cor(CorData) is used to identify Correlations between the data.
Variable having correlations of more than 0.85 are as follows

1. "Calories" - Calories.from.Fat, Total.Fat, Total.Fat.PDV
2. "Calories.from.Fat"- Calories, Total.Fat, Total.Fat.PDV
3. "Total.Fat"- Calories, Calories.from.Fat, Total.Fat.PDV
4. "Total.Fat.PDV"- Calories, Calories.from.Fat, Total.Fat,
5. "Saturated.Fat"- Saturated.Fat.PDV
6. "Saturated.Fat.PDV" - Saturated.Fat
7. "Cholesterol"- Cholesterol.PDV
8. "Cholesterol.PDV"- Cholesterol
9. "Sodium"- Sodium.PDV, Protein, Iron.PDV
10 "Sodium.PDV"- Sodium, Protein, Iron.PDV
11 "Carbohydrates"- Carbohydrates.PDV
12 "Carbohydrates.PDV" - Carbohydrates
13 "Dietary.Fiber" - Dietary.Fiber.PDV
14 "Dietary.Fiber.PDV" - Dietary.Fiber
15 "Protein" - Sodium Sodium.PDV
16 "Iron.PDV" - Sodium Sodium.PDV

```
> Data_cor <- cor(CorData)
> Data_cor
                     Calories Calories.from.Fat     Total.Fat Total.Fat.PDV
Calories           1.00000000        0.90458780    0.90440916    0.90412255
Calories.from.Fat  0.90458780        1.00000000    0.99966350    0.99972526
Total.Fat          0.90440916        0.99966350    1.00000000    0.99976474
Total.Fat.PDV      0.90412255        0.99972526    0.99976474    1.00000000
Saturated.Fat      0.84556364        0.84700759    0.84670724    0.84737925
Saturated.Fat.PDV  0.84763077        0.84959196    0.84929279    0.84997301
Trans.Fat          0.52244092        0.43368614    0.43145290    0.43301574
Cholesterol        0.59639916        0.68216110    0.68054737    0.68093953
Cholesterol.PDV    0.59520769        0.68160700    0.67999970    0.68037809
Sodium             0.71230869        0.84662441    0.84615842    0.84672792
Sodium.PDV         0.71341497        0.84727635    0.84678022    0.84736817
Carbohydrates      0.78153946        0.46167225    0.46121347    0.46051625
Carbohydrates.PDV  0.78124203        0.46146307    0.46100479    0.46029841
Dietary.Fiber      0.53889351        0.58127379    0.58083725    0.58059179
Dietary.Fiber.PDV  0.54001419        0.57562058    0.57520633    0.57503300
Sugars             0.25959812       -0.11528469   -0.11544573   -0.11576122
Protein            0.78784745        0.80791326    0.80777296    0.80792221
Vitamin.A.PDV      0.10884404        0.05673072    0.05443396    0.05403819
Vitamin.C.PDV     -0.06874704       -0.08733073   -0.08935397   -0.08935258
Calcium.PDV        0.42842646        0.16103409    0.16285952    0.16203088
Iron.PDV           0.64355160        0.73589431    0.73468451    0.73547755
                  Saturated.Fat Saturated.Fat.PDV   Trans.Fat Cholesterol
Calories             0.84556364        0.84763077  0.52244092  0.59639916
Calories.from.Fat    0.84700759        0.84959196  0.43368614  0.68216110
Total.Fat            0.84670724        0.84929279  0.43145290  0.68054737
Total.Fat.PDV        0.84737925        0.84997301  0.43301574  0.68093953
Saturated.Fat        1.00000000        0.99927862  0.62061061  0.63121047
Saturated.Fat.PDV    0.99927862        1.00000000  0.62021025  0.63360307
Trans.Fat            0.62061061        0.62021025  1.00000000  0.25393452
Cholesterol          0.63121047        0.63360307  0.25393452  1.00000000
Cholesterol.PDV      0.63033406        0.63271160  0.25150204  0.99985535
Sodium               0.58407526        0.58869363  0.18758040  0.62436186
Sodium.PDV           0.58532342        0.58995842  0.18833862  0.62474286
Carbohydrates        0.59126063        0.59132201  0.46324990  0.27097747
Carbohydrates.PDV    0.59174306        0.59165462  0.46289148  0.27266167
Dietary.Fiber        0.35181784        0.35683067  0.05491839  0.43557480
Dietary.Fiber.PDV    0.34715171        0.35179740  0.05830128  0.44026646
Sugars               0.19773362        0.19592799  0.33475579 -0.13551832
Protein              0.60302754        0.60658146  0.38824850  0.56156141
Vitamin.A.PDV        0.06497234        0.06537608  0.07583347  0.08023904
Vitamin.C.PDV       -0.17967200       -0.17805875 -0.07661214 -0.08297821
Calcium.PDV          0.40331077        0.40113868  0.38533110  0.13207674
Iron.PDV             0.57806237        0.58048787  0.32547588  0.65499955
                  Cholesterol.PDV     Sodium  Sodium.PDV Carbohydrates
Calories               0.59520769 0.71230869  0.71341497    0.78153946
Calories.from.Fat      0.68160700 0.84662441  0.84727635    0.46167225
Total.Fat              0.67999970 0.84615842  0.84678022    0.46121347
Total.Fat.PDV          0.68037809 0.84672792  0.84736817    0.46051625
Saturated.Fat          0.63033406 0.58407526  0.58532342    0.59126063
Saturated.Fat.PDV      0.63271160 0.58869363  0.58995842    0.59132201
Trans.Fat              0.25150204 0.18758040  0.18833862    0.46324990
Cholesterol            0.99985535 0.62436186  0.62474286    0.27097747
Cholesterol.PDV        1.00000000 0.62331973  0.62372045    0.26929976
```

```
Sodium               0.62331973  1.00000000  0.99992860      0.20079562
Sodium.PDV           0.62372045  0.99992860  1.00000000      0.20242648
Carbohydrates        0.26929976  0.20079562  0.20242648      1.00000000
Carbohydrates.PDV    0.27099242  0.20103208  0.20266302      0.99962040
Dietary.Fiber        0.43494029  0.69438946  0.69391331      0.22457662
Dietary.Fiber.PDV    0.43981420  0.68999511  0.68946426      0.22825670
Sugars              -0.13645937 -0.42653553 -0.42494302      0.76236214
Protein              0.56095695  0.86980157  0.86986999      0.35212224
Vitamin.A.PDV        0.08005943  0.08306810  0.08325880      0.08380171
Vitamin.C.PDV       -0.08331470 -0.03076934 -0.03094523     -0.03472430
Calcium.PDV          0.13238169 -0.02407377 -0.02214497      0.58969874
Iron.PDV             0.65316750  0.87159268  0.87074199      0.21024093
                  Carbohydrates.PDV Dietary.Fiber Dietary.Fiber.PDV     Sugars
Calories                 0.78124203    0.53889351        0.54001419  0.25959812
Calories.from.Fat        0.46146307    0.58127379        0.57562058 -0.11528469
Total.Fat                0.46100479    0.58083725        0.57520633 -0.11544573
Total.Fat.PDV            0.46029841    0.58059179        0.57503300 -0.11576122
Saturated.Fat            0.59174306    0.35181784        0.34715171  0.19773362
Saturated.Fat.PDV        0.59165462    0.35683067        0.35179740  0.19592799
Trans.Fat                0.46289148    0.05491839        0.05830128  0.33475579
Cholesterol              0.27266167    0.43557480        0.44026646 -0.13551832
Cholesterol.PDV          0.27099242    0.43494029        0.43981420 -0.13645937
Sodium                   0.20103208    0.69438946        0.68999511 -0.42653553
Sodium.PDV               0.20266302    0.69391331        0.68946426 -0.42494302
Carbohydrates            0.99962040    0.22457662        0.22825670  0.76236214
Carbohydrates.PDV        1.00000000    0.22405781        0.22728497  0.76228199
Dietary.Fiber            0.22405781    1.00000000        0.98635029 -0.29517842
Dietary.Fiber.PDV        0.22728497    0.98635029        1.00000000 -0.28701389
Sugars                   0.76228199   -0.29517842       -0.28701389  1.00000000
Protein                  0.35217806    0.64134483        0.65664825 -0.17993965
Vitamin.A.PDV            0.08337579    0.34051837        0.36138045  0.04848771
Vitamin.C.PDV           -0.03545009    0.14193503        0.15001851 -0.06984698
Calcium.PDV              0.59026335    0.02871108        0.05235903  0.60009291
Iron.PDV                 0.21064338    0.74041104        0.73781398 -0.36476708
                     Protein Vitamin.A.PDV Vitamin.C.PDV Calcium.PDV
Calories           0.78784745    0.10884404  -0.068747036  0.42842646
Calories.from.Fat  0.80791326    0.05673072  -0.087330731  0.16103409
Total.Fat          0.80777296    0.05443396  -0.089353974  0.16285952
Total.Fat.PDV      0.80792221    0.05403819  -0.089352581  0.16203088
Saturated.Fat      0.60302754    0.06497234  -0.179672005  0.40331077
Saturated.Fat.PDV  0.60658146    0.06537608  -0.178058754  0.40113868
Trans.Fat          0.38824850    0.07583347  -0.076612139  0.38533110
Cholesterol        0.56156141    0.08023904  -0.082978213  0.13207674
Cholesterol.PDV    0.56095695    0.08005943  -0.083314698  0.13238169
Sodium             0.86980157    0.08306810  -0.030769341 -0.02407377
Sodium.PDV         0.86986999    0.08325880  -0.030945232 -0.02214497
Carbohydrates      0.35212224    0.08380171  -0.034724302  0.58969874
Carbohydrates.PDV  0.35217806    0.08337579  -0.035450092  0.59026335
Dietary.Fiber      0.64134483    0.34051837   0.141935030  0.02871108
Dietary.Fiber.PDV  0.65664825    0.36138045   0.150018508  0.05235903
Sugars            -0.17993965    0.04848771  -0.069846984  0.60009291
Protein            1.00000000    0.21409799  -0.045776805  0.32797109
Vitamin.A.PDV      0.21409799    1.00000000   0.069171110  0.17918994
Vitamin.C.PDV     -0.04577681    0.06917111   1.000000000 -0.21537981
Calcium.PDV        0.32797109    0.17918994  -0.215379809  1.00000000
Iron.PDV           0.79271946    0.13787902   0.001291741  0.03414874
```

```
                   Iron.PDV
Calories            0.643551604
Calories.from.Fat   0.735894311
Total.Fat           0.734684514
Total.Fat.PDV       0.735477546
Saturated.Fat       0.578062375
Saturated.Fat.PDV   0.580487865
Trans.Fat           0.325475885
Cholesterol         0.654999546
Cholesterol.PDV     0.653167496
Sodium              0.871592677
Sodium.PDV          0.870741991
Carbohydrates       0.210240935
Carbohydrates.PDV   0.210643376
Dietary.Fiber       0.740411035
Dietary.Fiber.PDV   0.737813978
Sugars             -0.364767078
Protein             0.792719455
Vitamin.A.PDV       0.137879022
Vitamin.C.PDV       0.001291741
Calcium.PDV         0.034148740
Iron.PDV            1.000000000
```

We Identify the variable having more than 0.85 correlation

```
> for (i in 1:nrow(Data_cor)){
+   correlations <-  which((Data_cor[i,] > 0.85) & (Data_cor[i,] != 1))
+   if(length(correlations)> 0){
+     print(colnames(CorData)[i])
+     print(correlations)
+   }
+ }
[1] "Calories"
Calories.from.Fat        Total.Fat     Total.Fat.PDV
                2                3                 4
[1] "Calories.from.Fat"
   Calories      Total.Fat Total.Fat.PDV
          1              3             4
[1] "Total.Fat"
        Calories Calories.from.Fat     Total.Fat.PDV
               1                 2                 4
[1] "Total.Fat.PDV"
        Calories Calories.from.Fat         Total.Fat
               1                 2                 3
[1] "Saturated.Fat"
Saturated.Fat.PDV
                6
[1] "Saturated.Fat.PDV"
Saturated.Fat
            5
[1] "Cholesterol"
Cholesterol.PDV
              9
[1] "Cholesterol.PDV"
Cholesterol
          8
```

```
[1] "Sodium"
Sodium.PDV    Protein    Iron.PDV
        11          17          21
[1] "Sodium.PDV"
  Sodium   Protein Iron.PDV
      10          17      21
[1] "Carbohydrates"
Carbohydrates.PDV
              13
[1] "Carbohydrates.PDV"
Carbohydrates
           12
[1] "Dietary.Fiber"
Dietary.Fiber.PDV
              15
[1] "Dietary.Fiber.PDV"
Dietary.Fiber
           14
[1] "Protein"
    Sodium Sodium.PDV
        10          11
[1] "Iron.PDV"
    Sodium Sodium.PDV
        10          11
```

3.5 Missing Value Identification

plot_missing(myData) is used to check the missing variable and our data has no missing value

```
> ##plot the missing value
> plot_missing(myData)
```
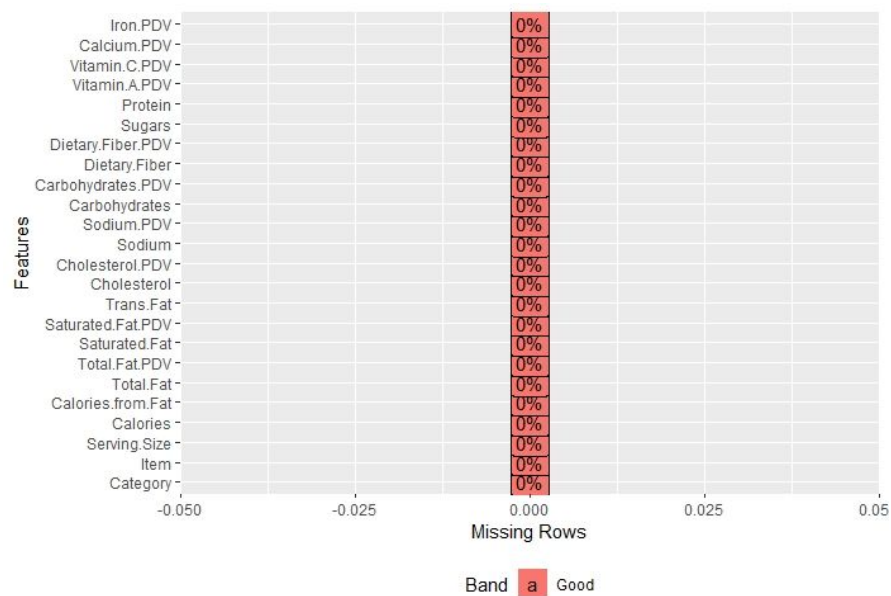
3.6 Outlier Identification

Outlier is identified using the Box Plot
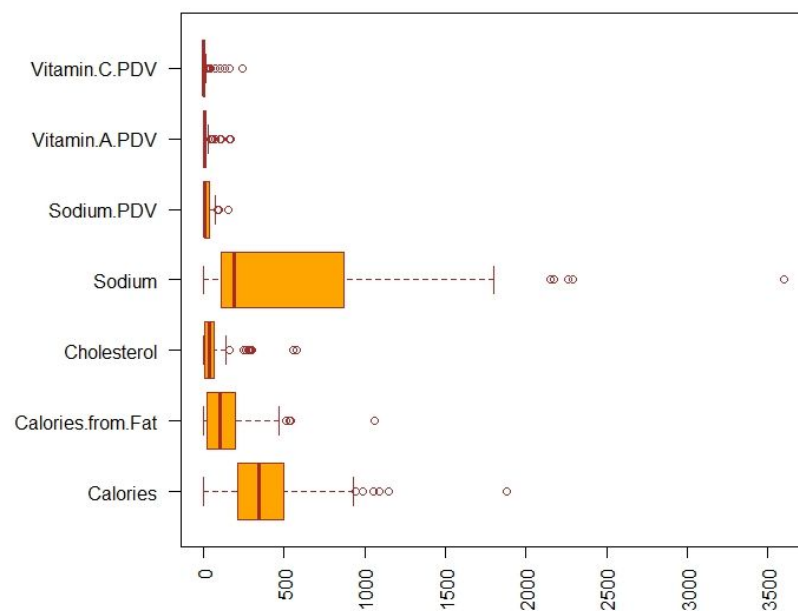
Box Plot of Variables - 1

Box Plot of Variables - 2

Variable having Outlier are
- "Calories"
- "Calories.from.Fat"
- "Total.Fat"
- "Total.Fat.PDV"
- "Trans.Fat"
- "Cholesterol"
- "Cholesterol.PDV"
- "Sodium"
- "Sodium.PDV"
- "Carbohydrates"
- "Carbohydrates.PDV"
- "Dietary.Fiber.PDV"
- "Sugars"
- "Protein"
- "Vitamin.A.PDV"
- "Vitamin.C.PDV"
- "Calcium.PDV"
- "Iron.PDV"

3.7 Feature Creation

We rename the column name of data frame to easy and understandable access the data.

Table_Cat_Var is new table contains the Category and there no.of items.

CorData is used to get all numeric variable in single DataFrame

Data_cor is used to store the correlation value between the variables

Table_Cat_Colestrol is new table contains Category and their total Cholesterol % value

4 Conclusion

6 Questions Asked and their solutions

Q1. Plot graphically which food categories have the highest and lowest varieties.
Ans. Coffee and Tea has maximum variety and Salads have minimum varieties.
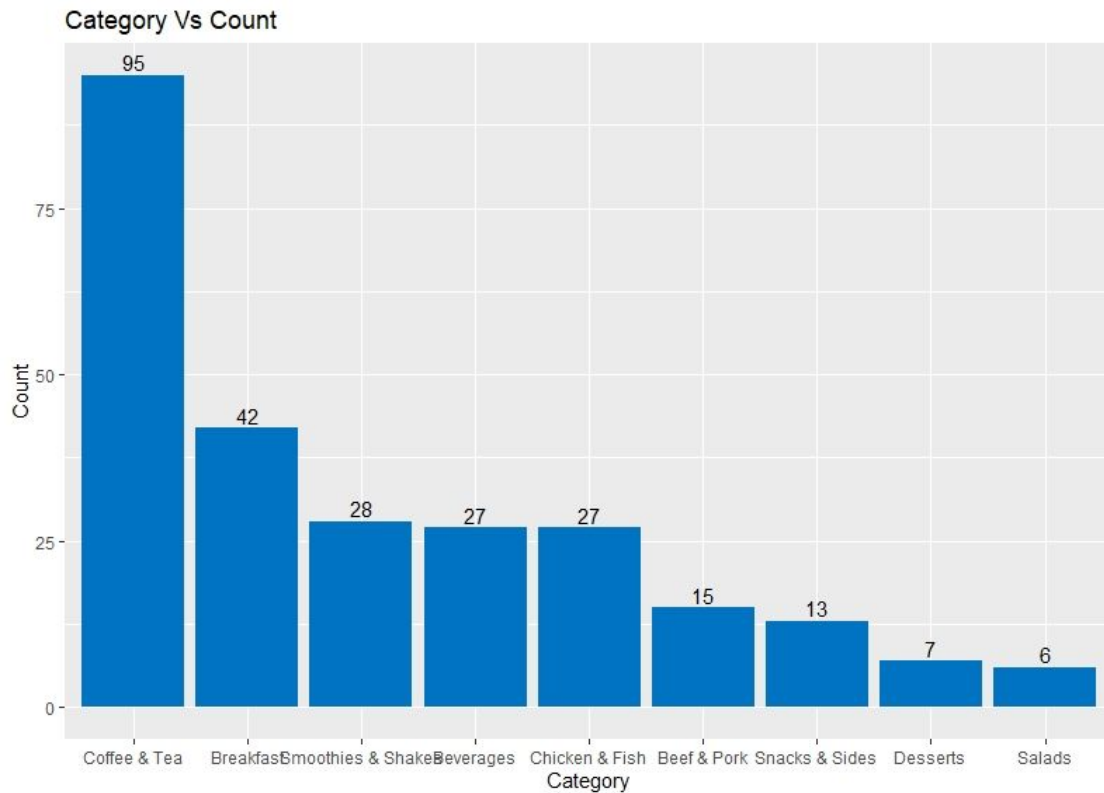
Q2. Which all variables have an outlier?
Ans. Variable having Outlier are
- "Calories"
- "Calories.from.Fat"
- "Total.Fat"
- "Total.Fat.PDV"
- "Trans.Fat"
- "Cholesterol"
- "Cholesterol.PDV"
- "Sodium"
- "Sodium.PDV"
- "Carbohydrates"
- "Carbohydrates.PDV"

- "Dietary.Fiber.PDV"
- "Sugars"
- "Protein"
- "Vitamin.A.PDV"
- "Vitamin.C.PDV"
- "Calcium.PDV"
- "Iron.PDV"

Outlier is identified using the Box Plot

Box Plot of Variables - 1
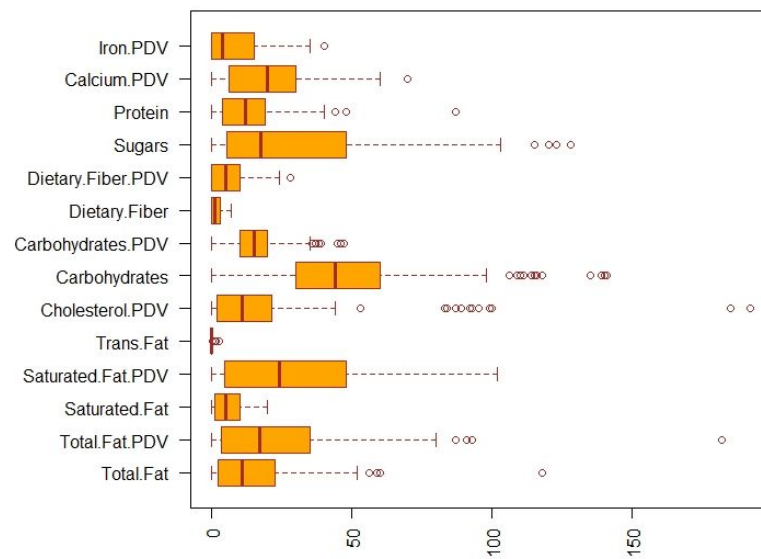
Box Plot of Variables - 2
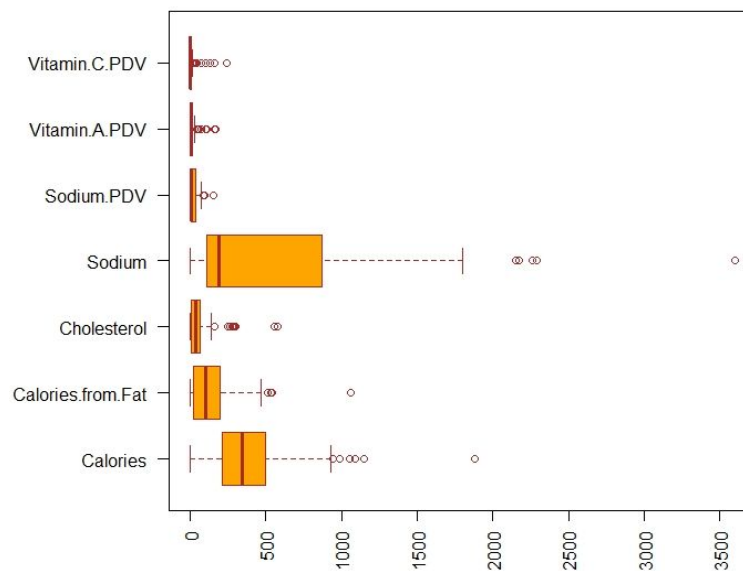
Q3. Which variables have the highest correlation? Plot them and find out the value?

Ans. Variable having correlations of more than 0.85 are as follows

1. "Calories" - Calories.from.Fat, Total.Fat, Total.Fat.PDV
2. "Calories.from.Fat"- Calories, Total.Fat, Total.Fat.PDV
3. "Total.Fat"- Calories, Calories.from.Fat, Total.Fat.PDV
4. "Total.Fat.PDV"- Calories, Calories.from.Fat, Total.Fat,
5. "Saturated.Fat"- Saturated.Fat.PDV
6. "Saturated.Fat.PDV" - Saturated.Fat
7. "Cholesterol"- Cholesterol.PDV
8. "Cholesterol.PDV"- Cholesterol
9. "Sodium"- Sodium.PDV, Protein, Iron.PDV
10 "Sodium.PDV"- Sodium, Protein, Iron.PDV
11 "Carbohydrates"- Carbohydrates.PDV
12 "Carbohydrates.PDV" - Carbohydrates
13 "Dietary.Fiber" - Dietary.Fiber.PDV
14 "Dietary.Fiber.PDV" - Dietary.Fiber
15 "Protein" - Sodium Sodium.PDV
16 "Iron.PDV" - Sodium Sodium.PDV

Q4. Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?
Ans.Breakfast contributes maximum % of Cholesterol in a diet

Category Vs Colestrol Percentage
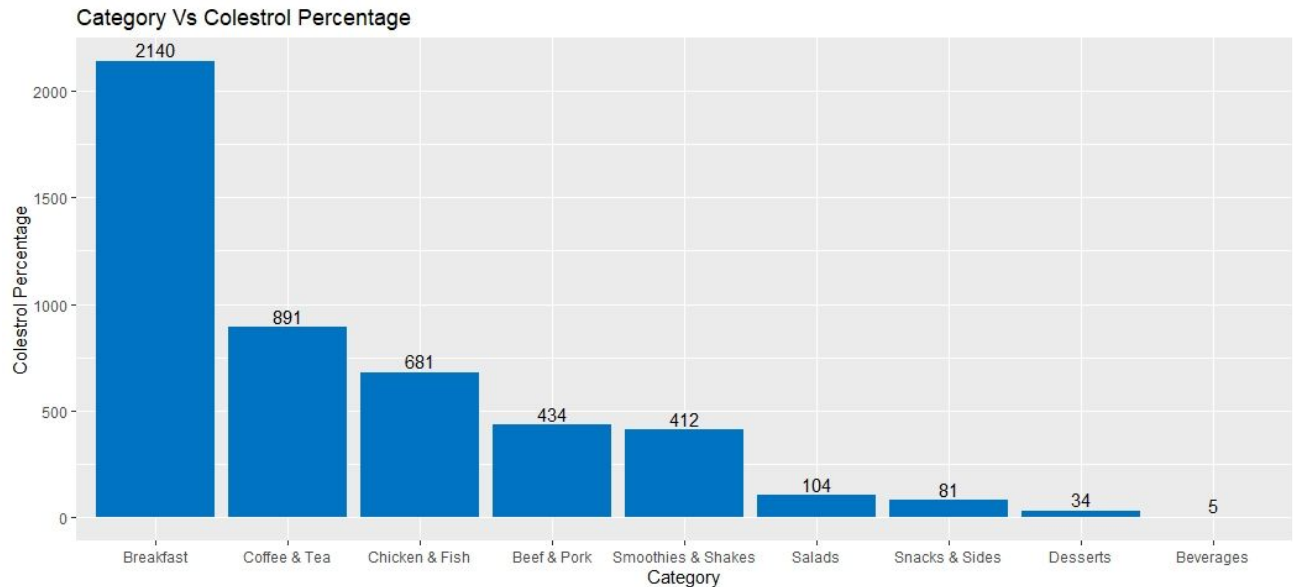
Q5. Which item contributes maximum to the Sodium intake?
Ans. "Chicken McNuggets (40 piece)"

Q6. Which 4 food items contain the most amount of Saturated Fat?
Ans.Item contribute max saturated fat are:
- Big Breakfast with Hotcakes (Large Biscuit)
- Chicken McNuggets (40 piece
- FrappÃ© Chocolate Chip (Large)
- McFlurry with M&Mâ€™s Candies (Medium)

5 Appendix A – Source Code

```
> #Setting the Working Directory
> setwd("C:/Users/dignitas/Desktop/Learning/Mcdonald/Home 1")
> getwd()
[1] "C:/Users/dignitas/Desktop/Learning/Mcdonald/Home 1"
>
> #Libraries Required for Analysis
> library(dplyr)
> library(ggplot2)
> library(corrplot)
> library(DataExplorer)
>
> #Importing the Data
> myData = read.csv("Mcdonald.csv")
>
> str(myData)
'data.frame':     260 obs. of  24 variables:
 $ Category                  : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Item                      : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76 77 228 229 230 245
12 11 14 13 ...
 $ Serving.Size              : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54 42 69 69 83 63 72
65 73 ...
 $ Calories                  : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat         : int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat                 : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat....Daily.Value. : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat             : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat....Daily.Value.: int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat                 : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol               : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol....Daily.Value. : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium                    : int  750 770 780 860 880 960 1300 1410 1300 1420 ...
 $ Sodium....Daily.Value.    : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates             : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates....Daily.Value.: int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber             : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber....Daily.Value.: int  17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars                    : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein                   : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A....Daily.Value. : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C....Daily.Value. : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium....Daily.Value.   : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron....Daily.Value.      : int  15 8 10 15 10 20 15 20 10 15 ...
> #Rnaming the Column name to more
> names(myData) <- c("Category"
+                    ,"Item"
+                    ,"Serving.Size"
+                    ,"Calories"
+                    ,"Calories.from.Fat"
+                    ,"Total.Fat"
+                    ,"Total.Fat.PDV"
+                    ,"Saturated.Fat"
+                    ,"Saturated.Fat.PDV"
```

```
+                         ,"Trans.Fat"
+                         ,"Cholesterol"
+                         ,"Cholesterol.PDV"
+                         ,"Sodium"
+                         ,"Sodium.PDV"
+                         ,"Carbohydrates"
+                         ,"Carbohydrates.PDV"
+                         ,"Dietary.Fiber"
+                         ,"Dietary.Fiber.PDV"
+                         ,"Sugars"
+                         ,"Protein"
+                         ,"Vitamin.A.PDV"
+                         ,"Vitamin.C.PDV"
+                         ,"Calcium.PDV"
+                         ,"Iron.PDV"
+                         )
> str(myData)
'data.frame':    260 obs. of  24 variables:
 $ Category          : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Item              : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76 77 228 229 230 245 12 11 14 13
...
 $ Serving.Size      : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54 42 69 69 83 63 72 65 73 ...
 $ Calories          : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat : int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat         : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat.PDV     : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat     : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat.PDV : int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat         : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol       : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol.PDV   : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium            : int  750 770 780 860 880 960 1300 1410 1300 1420 ...
 $ Sodium.PDV        : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates     : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates.PDV : int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber     : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber.PDV : int  17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars            : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein           : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A.PDV     : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C.PDV     : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium.PDV       : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron.PDV          : int  15 8 10 15 10 20 15 20 10 15 ...
>
> #Attach myData to directly access the variable
> attach(myData)
The following objects are masked from myData (pos = 3):

    Calcium.PDV, Calories, Calories.from.Fat, Carbohydrates,
    Carbohydrates.PDV, Category, Cholesterol, Cholesterol.PDV,
    Dietary.Fiber, Dietary.Fiber.PDV, Iron.PDV, Item, Protein,
    Saturated.Fat, Saturated.Fat.PDV, Serving.Size, Sodium,
    Sodium.PDV, Sugars, Total.Fat, Total.Fat.PDV, Trans.Fat,
    Vitamin.A.PDV, Vitamin.C.PDV

The following objects are masked from myData (pos = 8):
```

```
     Calcium.PDV, Calories, Calories.from.Fat, Carbohydrates,
     Carbohydrates.PDV, Category, Cholesterol, Cholesterol.PDV,
     Dietary.Fiber, Dietary.Fiber.PDV, Iron.PDV, Item, Protein,
     Saturated.Fat, Saturated.Fat.PDV, Serving.Size, Sodium,
     Sodium.PDV, Sugars, Total.Fat, Total.Fat.PDV, Trans.Fat,
     Vitamin.A.PDV, Vitamin.C.PDV

>
>
> #General Data Analysis
> ##Check the Class of Data
> class(myData)
[1] "data.frame"
>
> ##First Inspection of Dataset using str
> str(myData)
'data.frame':    260 obs. of  24 variables:
 $ Category        : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Item            : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76 77 228 229 230 245 12 11 14 13
...
 $ Serving.Size    : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54 42 69 69 83 63 72 65 73 ...
 $ Calories        : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat: int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat       : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat.PDV   : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat   : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat.PDV: int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat       : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol     : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol.PDV : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium          : int  750 770 780 860 880 960 1300 1410 1300 1420 ...
 $ Sodium.PDV      : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates   : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates.PDV: int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber   : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber.PDV: int  17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars          : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein         : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A.PDV   : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C.PDV   : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium.PDV     : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron.PDV        : int  15 8 10 15 10 20 15 20 10 15 ...
>
> ## Find the name of variable
> names(myData)
 [1] "Category"          "Item"              "Serving.Size"
 [4] "Calories"          "Calories.from.Fat" "Total.Fat"
 [7] "Total.Fat.PDV"     "Saturated.Fat"     "Saturated.Fat.PDV"
[10] "Trans.Fat"         "Cholesterol"       "Cholesterol.PDV"
[13] "Sodium"            "Sodium.PDV"        "Carbohydrates"
[16] "Carbohydrates.PDV" "Dietary.Fiber"     "Dietary.Fiber.PDV"
[19] "Sugars"            "Protein"           "Vitamin.A.PDV"
[22] "Vitamin.C.PDV"     "Calcium.PDV"       "Iron.PDV"
>
> ##find the dimension of Data
```

```
> dim(myData)
[1] 260  24
>
> ##find first 6 elements of Data
> head(myData)
   Category                           Item   Serving.Size Calories
1 Breakfast                 Egg McMuffin 4.8 oz (136 g)      300
2 Breakfast            Egg White Delight 4.8 oz (135 g)      250
3 Breakfast              Sausage McMuffin 3.9 oz (111 g)      370
4 Breakfast      Sausage McMuffin with Egg 5.7 oz (161 g)      450
5 Breakfast Sausage McMuffin with Egg Whites 5.7 oz (161 g)      400
6 Breakfast          Steak & Egg McMuffin 6.5 oz (185 g)      430
  Calories.from.Fat Total.Fat Total.Fat.PDV Saturated.Fat
1               120        13            20             5
2                70         8            12             3
3               200        23            35             8
4               250        28            43            10
5               210        23            35             8
6               210        23            36             9
  Saturated.Fat.PDV Trans.Fat Cholesterol Cholesterol.PDV Sodium
1                25         0         260              87    750
2                15         0          25               8    770
3                42         0          45              15    780
4                52         0         285              95    860
5                42         0          50              16    880
6                46         1         300             100    960
  Sodium.PDV Carbohydrates Carbohydrates.PDV Dietary.Fiber
1         31            31                10             4
2         32            30                10             4
3         33            29                10             4
4         36            30                10             4
5         37            30                10             4
6         40            31                10             4
  Dietary.Fiber.PDV Sugars Protein Vitamin.A.PDV Vitamin.C.PDV Calcium.PDV
1                17      3      17            10             0          25
2                17      3      18             6             0          25
3                17      2      14             8             0          25
4                17      2      21            15             0          30
5                17      2      21             6             0          25
6                18      3      26            15             2          30
  Iron.PDV
1       15
2        8
3       10
4       15
5       10
6       20
>
> ##find last 5 elements of Data
> tail(myData)
             Category                                          Item
255 Smoothies & Shakes         McFlurry with M&Mâ€TMs Candies (Snack)
256 Smoothies & Shakes               McFlurry with Oreo Cookies (Small)
257 Smoothies & Shakes              McFlurry with Oreo Cookies (Medium)
258 Smoothies & Shakes               McFlurry with Oreo Cookies (Snack)
259 Smoothies & Shakes McFlurry with Reese's Peanut Butter Cups (Medium)
```

```
260 Smoothies & Shakes  McFlurry with Reese's Peanut Butter Cups (Snack)
      Serving.Size Calories Calories.from.Fat Total.Fat Total.Fat.PDV
255  7.3 oz (207 g)      430             140        15           24
256 10.1 oz (285 g)      510             150        17           26
257 13.4 oz (381 g)      690             200        23           35
258  6.7 oz (190 g)      340             100        11           17
259 14.2 oz (403 g)      810             290        32           50
260  7.1 oz (202 g)      410             150        16           25
    Saturated.Fat Saturated.Fat.PDV Trans.Fat Cholesterol Cholesterol.PDV
255            10                48       0.0          35              11
256             9                44       0.5          45              14
257            12                58       1.0          55              19
258             6                29       0.0          30               9
259            15                76       1.0          60              20
260             8                38       0.0          30              10
    Sodium Sodium.PDV Carbohydrates Carbohydrates.PDV Dietary.Fiber
255    120          5            64                21             1
256    280         12            80                27             1
257    380         16           106                35             1
258    190          8            53                18             1
259    400         17           114                38             2
260    200          8            57                19             1
    Dietary.Fiber.PDV Sugars Protein Vitamin.A.PDV Vitamin.C.PDV
255                 4     59       9            10             0
256                 4     64      12            15             0
257                 5     85      15            20             0
258                 2     43       8            10             0
259                 9    103      21            20             0
260                 5     51      10            10             0
    Calcium.PDV Iron.PDV
255          30        4
256          40        8
257          50       10
258          25        6
259          60        6
260          30        4
>
> ##find summary of myData to get Min,median,Mean and Max with First and 3rd quartile.
> summary(myData)
              Category                                    Item
 Coffee & Tea        :95   1% Low Fat Milk Jug                 :  1
 Breakfast           :42   Apple Slices                        :  1
 Smoothies & Shakes:28     Bacon Buffalo Ranch McChicken       :  1
 Beverages           :27   Bacon Cheddar McChicken             :  1
 Chicken & Fish      :27   Bacon Clubhouse Burger              :  1
 Beef & Pork         :15   Bacon Clubhouse Crispy Chicken Sandwich:  1
 (Other)             :26   (Other)                             :254
      Serving.Size     Calories      Calories.from.Fat    Total.Fat
 16 fl oz cup: 45   Min.   :   0.0   Min.   :   0.0   Min.   :  0.000
 12 fl oz cup: 38   1st Qu.: 210.0   1st Qu.:  20.0   1st Qu.:  2.375
 22 fl oz cup: 20   Median : 340.0   Median : 100.0   Median : 11.000
 20 fl oz cup: 16   Mean   : 368.3   Mean   : 127.1   Mean   : 14.165
 21 fl oz cup:  7   3rd Qu.: 500.0   3rd Qu.: 200.0   3rd Qu.: 22.250
 30 fl oz cup:  7   Max.   :1880.0   Max.   :1060.0   Max.   :118.000
 (Other)     :127
 Total.Fat.PDV    Saturated.Fat    Saturated.Fat.PDV    Trans.Fat
```

```
  Min.   :  0.00   Min.   : 0.000   Min.   :  0.00   Min.   :0.0000
  1st Qu.:  3.75   1st Qu.: 1.000   1st Qu.:  4.75   1st Qu.:0.0000
  Median : 17.00   Median : 5.000   Median : 24.00   Median :0.0000
  Mean   : 21.82   Mean   : 6.008   Mean   : 29.97   Mean   :0.2038
  3rd Qu.: 35.00   3rd Qu.:10.000   3rd Qu.: 48.00   3rd Qu.:0.0000
  Max.   :182.00   Max.   :20.000   Max.   :102.00   Max.   :2.5000

  Cholesterol      Cholesterol.PDV      Sodium          Sodium.PDV
  Min.   :  0.00   Min.   :  0.00   Min.   :   0.0   Min.   :  0.00
  1st Qu.:  5.00   1st Qu.:  2.00   1st Qu.: 107.5   1st Qu.:  4.75
  Median : 35.00   Median : 11.00   Median : 190.0   Median :  8.00
  Mean   : 54.94   Mean   : 18.39   Mean   : 495.8   Mean   : 20.68
  3rd Qu.: 65.00   3rd Qu.: 21.25   3rd Qu.: 865.0   3rd Qu.: 36.25
  Max.   :575.00   Max.   :192.00   Max.   :3600.0   Max.   :150.00

  Carbohydrates    Carbohydrates.PDV Dietary.Fiber    Dietary.Fiber.PDV
  Min.   :  0.00   Min.   : 0.00    Min.   :0.000    Min.   : 0.000
  1st Qu.: 30.00   1st Qu.:10.00    1st Qu.:0.000    1st Qu.: 0.000
  Median : 44.00   Median :15.00    Median :1.000    Median : 5.000
  Mean   : 47.35   Mean   :15.78    Mean   :1.631    Mean   : 6.531
  3rd Qu.: 60.00   3rd Qu.:20.00    3rd Qu.:3.000    3rd Qu.:10.000
  Max.   :141.00   Max.   :47.00    Max.   :7.000    Max.   :28.000

     Sugars          Protein        Vitamin.A.PDV    Vitamin.C.PDV
  Min.   :  0.00   Min.   : 0.00    Min.   :  0.00   Min.   :  0.000
  1st Qu.:  5.75   1st Qu.: 4.00    1st Qu.:  2.00   1st Qu.:  0.000
  Median : 17.50   Median :12.00    Median :  8.00   Median :  0.000
  Mean   : 29.42   Mean   :13.34    Mean   : 13.43   Mean   :  8.535
  3rd Qu.: 48.00   3rd Qu.:19.00    3rd Qu.: 15.00   3rd Qu.:  4.000
  Max.   :128.00   Max.   :87.00    Max.   :170.00   Max.   :240.000

  Calcium.PDV       Iron.PDV
  Min.   : 0.00    Min.   : 0.000
  1st Qu.: 6.00    1st Qu.: 0.000
  Median :20.00    Median : 4.000
  Mean   :20.97    Mean   : 7.735
  3rd Qu.:30.00    3rd Qu.:15.000
  Max.   :70.00    Max.   :40.000

>
> ##plot the missing value
> plot_missing(myData)
>
> ##Plot histogram of the data
> plot_histogram(myData,nrow = 4,ncol = 4)
>
> #Problem Solutions
>
> ##Problem 1 : Plot graphically which food categories have the highest and lowest varieties.
> ###Create a Table having Category and its Count
> Table_Cat_Var = myData %>%   group_by(Category) %>% summarise(counts = n())
> Table_Cat_Var
# A tibble: 9 x 2
  Category         counts
  <fct>             <int>
1 Beef & Pork          15
```

```
2 Beverages             27
3 Breakfast             42
4 Chicken & Fish        27
5 Coffee & Tea          95
6 Desserts               7
7 Salads                 6
8 Smoothies & Shakes    28
9 Snacks & Sides        13
> ###Plotting that table Graphically
> Graph_Cat_Count = ggplot(Table_Cat_Var, aes(x = reorder(Category, -counts), y = counts)) +
+   geom_bar(fill = "#0073C2FF", stat = "identity") +
+   geom_text(aes(label =  counts), vjust = -0.3) +
+   ggtitle("Category Vs Count") +
+   xlab("Category") + ylab("Count")
> Graph_Cat_Count
> ####After Plotting Table We found "coffee and tea" has maximum and "Salad" has minimum variety
>
> ##Problem 2 : Which all variables have an outlier?
> ###For that we plotted boxplot of variables to find out the products having outlier
> ### Boxplot 1
> par(mar=c(4,10,4,4))
>
> Boxplot_Oulier_1 = boxplot(Total.Fat
+                            ,Total.Fat.PDV
+                            ,Saturated.Fat
+                            ,Saturated.Fat.PDV
+                            ,Trans.Fat
+                            ,Cholesterol.PDV
+                            ,Carbohydrates
+                            ,Carbohydrates.PDV
+                            ,Dietary.Fiber
+                            ,Dietary.Fiber.PDV
+                            ,Sugars
+                            ,Protein
+                            ,Calcium.PDV
+                            ,Iron.PDV
+                            ,names = c("Total.Fat"
+                                       ,"Total.Fat.PDV"
+                                       ,"Saturated.Fat"
+                                       ,"Saturated.Fat.PDV"
+                                       ,"Trans.Fat"
+                                       ,"Cholesterol.PDV"
+                                       ,"Carbohydrates"
+                                       ,"Carbohydrates.PDV"
+                                       ,"Dietary.Fiber"
+                                       ,"Dietary.Fiber.PDV"
+                                       ,"Sugars"
+                                       ,"Protein"
+                                       ,"Calcium.PDV"
+                                       ,"Iron.PDV")
+                            ,horizontal = TRUE
+                            ,las =2
+                            ,main = "Box Plot of Variables - 1"
+                            ,col = "orange"
+                            ,border = "brown"
+ )
```

```r
> 
> 
> Boxplot_Oulier_2 = boxplot(Calories
+                            ,Calories.from.Fat
+                            ,Cholesterol
+                            ,Sodium
+                            ,Sodium.PDV
+                            ,Vitamin.A.PDV
+                            ,Vitamin.C.PDV
+                            ,names = c("Calories"
+                                       ,"Calories.from.Fat"
+                                       ,"Cholesterol"
+                                       ,"Sodium"
+                                       ,"Sodium.PDV"
+                                       ,"Vitamin.A.PDV"
+                                       ,"Vitamin.C.PDV")
+                            ,horizontal = TRUE
+                            ,las =2
+                            ,main = "Box Plot of Variables - 2"
+                            ,col = "orange"
+                            ,border = "brown"
+ )
> Boxplot_Oulier_1
$stats
      [,1] [,2] [,3]  [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]  0.00  0.0    0   0.0    0  0.0    0    0    0     0   0.0     0
[2,]  2.25  3.5    1   4.5    0  2.0   30   10    0     0   5.5     4
[3,] 11.00 17.0    5  24.0    0 11.0   44   15    1     5  17.5    12
[4,] 22.50 35.0   10  48.0    0 21.5   60   20    3    10  48.0    19
[5,] 52.00 80.0   20 102.0    0 44.0   98   35    7    24 103.0    40
     [,13] [,14]
[1,]     0     0
[2,]     6     0
[3,]    20     4
[4,]    30    15
[5,]    60    35


$n
 [1] 260 260 260 260 260 260 260 260 260 260 260 260 260 260

$conf
          [,1]     [,2]      [,3]     [,4] [,5]      [,6]     [,7]     [,8]
[1,]  9.015754 13.9134 4.118113 19.73755    0  9.089245 41.06038 14.02013
[2,] 12.984246 20.0866 5.881887 28.26245    0 12.910755 46.93962 15.97987
           [,9]    [,10]    [,11]    [,12]   [,13]    [,14]
[1,] 0.7060377 4.020126 13.33553 10.53019 17.6483 2.530188
[2,] 1.2939623 5.979874 21.66447 13.46981 22.3517 5.469812


$out
  [1]  56.0  60.0  59.0 118.0  87.0  93.0  91.0 182.0   1.0   1.0   0.5
 [12]   0.5   1.5   1.0   1.5   1.5   1.5   1.5   2.5   0.5   1.0   1.5
 [23]   1.0   1.0   1.0   1.0   0.5   0.5   0.5   0.5   0.5   1.0   0.5
 [34]   0.5   0.5   0.5   0.5   0.5   0.5   1.0   1.0   1.0   1.0   1.0
 [45]   1.5   1.0   1.0   1.5   1.0   1.0   1.0   1.0   1.0   1.0   1.0
 [56]   1.0   1.0   1.0   1.0   0.5   1.0   0.5   1.0   1.0  87.0  95.0
 [67] 100.0  83.0  83.0  83.0  83.0  93.0  84.0  89.0  92.0  99.0 185.0
```

```
 [78] 185.0 192.0 192.0  53.0  89.0 111.0 116.0 110.0 115.0 118.0 111.0
 [89] 109.0 135.0 114.0 140.0 114.0 141.0 109.0 135.0 139.0 106.0 114.0
[100]  37.0  39.0  37.0  38.0  39.0  37.0  36.0  45.0  38.0  47.0  38.0
[111]  47.0  36.0  45.0  46.0  38.0  28.0  28.0  28.0  28.0 123.0 120.0
[122] 115.0 128.0  48.0  44.0  87.0  70.0  70.0  40.0  40.0

$group
  [1]  1  1  1  1  2  2  2  2  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
 [24]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
 [47]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  6  6  6  6  6
 [70]  6  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7  7  7  7  7  7  7
 [93]  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
[116] 10 10 10 10 11 11 11 11 12 12 12 13 13 14 14

$names
 [1] "Total.Fat"         "Total.Fat.PDV"     "Saturated.Fat"
 [4] "Saturated.Fat.PDV" "Trans.Fat"         "Cholesterol.PDV"
 [7] "Carbohydrates"     "Carbohydrates.PDV" "Dietary.Fiber"
[10] "Dietary.Fiber.PDV" "Sugars"            "Protein"
[13] "Calcium.PDV"       "Iron.PDV"

> Boxplot_Oulier_2
$stats
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    0    0    0  0.0    0    0
[2,]  210   20    5  105  4.5    2    0
[3,]  340  100   35  190  8.0    8    0
[4,]  500  200   65  870 36.5   15    4
[5,]  930  470  135 1800 75.0   30   10
attr(,"class")
 Calories
"integer"

$n
[1] 260 260 260 260 260 260 260

$conf
          [,1]      [,2]     [,3]     [,4]      [,5]     [,6]       [,7]
[1,] 311.5836  82.36226 29.12075 115.0396  4.864402 6.726163 -0.3919498
[2,] 368.4164 117.63774 40.87925 264.9604 11.135598 9.273837  0.3919498

$out
  [1] 1090 1150  990 1050  940 1880  510  540  530 1060  260  285  300  250
 [15]  250  250  250  280  250  265  275  295  555  555  575  575  160  265
 [29] 2150 2260 2170 2290 3600   90   94   91   95  150   60   60   60   60
 [43]   60   70   60   60  170  100  110  160  170  170   45   40   50   15
 [57]   15  130  130   15   20   25   20   20   20   20   25   30   20   25
 [71]   15   15   15   15   15   15   15   30   25   30   25   30   30   30
 [85]   45   70   15   25  160   15   25  100  130  160  240   30   35   45
 [99]   20   25   30

$group
  [1] 1 1 1 1 1 1 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 5 5
 [36] 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7
 [71] 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
```

```
$names
[1] "Calories"          "Calories.from.Fat" "Cholesterol"
[4] "Sodium"            "Sodium.PDV"        "Vitamin.A.PDV"
[7] "Vitamin.C.PDV"


>
> # ### Variable having Outlier
> # "Calories"
> # "Calories.from.Fat"
> # "Total.Fat"
> # "Total.Fat.PDV"
> # "Trans.Fat"
> # "Cholesterol"
> # "Cholesterol.PDV"
> # "Sodium"
> # "Sodium.PDV"
> # "Carbohydrates"
> # "Carbohydrates.PDV"
> # "Dietary.Fiber.PDV"
> # "Sugars"
> # "Protein"
> # "Vitamin.A.PDV"
> # "Vitamin.C.PDV"
> # "Calcium.PDV"
> # "Iron.PDV"

>
>
> ##Prblem 3: Which variables have the highest correlation? Plot them and find out the value?
> ###Segrigate Numeric Data
> CorData <- myData [, 4:24]
> str(CorData)
'data.frame':   260 obs. of  21 variables:
 $ Calories         : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat: int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat        : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat.PDV    : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat    : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat.PDV: int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat        : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol      : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol.PDV  : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium           : int  750 770 780 860 880 960 1300 1410 1300 1420 ...
 $ Sodium.PDV       : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates    : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates.PDV: int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber    : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber.PDV: int  17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars           : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein          : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A.PDV    : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C.PDV    : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium.PDV      : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron.PDV         : int  15 8 10 15 10 20 15 20 10 15 ...
>
> ###Find Corealation
> Data_cor <- cor(CorData)
```

```
> Data_cor
                      Calories Calories.from.Fat    Total.Fat Total.Fat.PDV
Calories            1.00000000       0.90458780   0.90440916    0.90412255
Calories.from.Fat   0.90458780       1.00000000   0.99966350    0.99972526
Total.Fat           0.90440916       0.99966350   1.00000000    0.99976474
Total.Fat.PDV       0.90412255       0.99972526   0.99976474    1.00000000
Saturated.Fat       0.84556364       0.84700759   0.84670724    0.84737925
Saturated.Fat.PDV   0.84763077       0.84959196   0.84929279    0.84997301
Trans.Fat           0.52244092       0.43368614   0.43145290    0.43301574
Cholesterol         0.59639916       0.68216110   0.68054737    0.68093953
Cholesterol.PDV     0.59520769       0.68160700   0.67999970    0.68037809
Sodium              0.71230869       0.84662441   0.84615842    0.84672792
Sodium.PDV          0.71341497       0.84727635   0.84678022    0.84736817
Carbohydrates       0.78153946       0.46167225   0.46121347    0.46051625
Carbohydrates.PDV   0.78124203       0.46146307   0.46100479    0.46029841
Dietary.Fiber       0.53889351       0.58127379   0.58083725    0.58059179
Dietary.Fiber.PDV   0.54001419       0.57562058   0.57520633    0.57503300
Sugars              0.25959812      -0.11528469  -0.11544573   -0.11576122
Protein             0.78784745       0.80791326   0.80777296    0.80792221
Vitamin.A.PDV       0.10884404       0.05673072   0.05443396    0.05403819
Vitamin.C.PDV      -0.06874704      -0.08733073  -0.08935397   -0.08935258
Calcium.PDV         0.42842646       0.16103409   0.16285952    0.16203088
Iron.PDV            0.64355160       0.73589431   0.73468451    0.73547755
                  Saturated.Fat Saturated.Fat.PDV    Trans.Fat Cholesterol
Calories             0.84556364        0.84763077   0.52244092  0.59639916
Calories.from.Fat    0.84700759        0.84959196   0.43368614  0.68216110
Total.Fat            0.84670724        0.84929279   0.43145290  0.68054737
Total.Fat.PDV        0.84737925        0.84997301   0.43301574  0.68093953
Saturated.Fat        1.00000000        0.99927862   0.62061061  0.63121047
Saturated.Fat.PDV    0.99927862        1.00000000   0.62021025  0.63360307
Trans.Fat            0.62061061        0.62021025   1.00000000  0.25393452
Cholesterol          0.63121047        0.63360307   0.25393452  1.00000000
Cholesterol.PDV      0.63033406        0.63271160   0.25150204  0.99985535
Sodium               0.58407526        0.58869363   0.18758040  0.62436186
Sodium.PDV           0.58532342        0.58995842   0.18833862  0.62474286
Carbohydrates        0.59126063        0.59132201   0.46324990  0.27097747
Carbohydrates.PDV    0.59174306        0.59165462   0.46289148  0.27266167
Dietary.Fiber        0.35181784        0.35683067   0.05491839  0.43557480
Dietary.Fiber.PDV    0.34715171        0.35179740   0.05830128  0.44026646
Sugars               0.19773362        0.19592799   0.33475579 -0.13551832
Protein              0.60302754        0.60658146   0.38824850  0.56156141
Vitamin.A.PDV        0.06497234        0.06537608   0.07583347  0.08023904
Vitamin.C.PDV       -0.17967200       -0.17805875  -0.07661214 -0.08297821
Calcium.PDV          0.40331077        0.40113868   0.38533110  0.13207674
Iron.PDV             0.57806237        0.58048787   0.32547588  0.65499955
                  Cholesterol.PDV     Sodium   Sodium.PDV Carbohydrates
Calories               0.59520769 0.71230869   0.71341497    0.78153946
Calories.from.Fat      0.68160700 0.84662441   0.84727635    0.46167225
Total.Fat              0.67999970 0.84615842   0.84678022    0.46121347
Total.Fat.PDV          0.68037809 0.84672792   0.84736817    0.46051625
Saturated.Fat          0.63033406 0.58407526   0.58532342    0.59126063
Saturated.Fat.PDV      0.63271160 0.58869363   0.58995842    0.59132201
Trans.Fat              0.25150204 0.18758040   0.18833862    0.46324990
Cholesterol            0.99985535 0.62436186   0.62474286    0.27097747
Cholesterol.PDV        1.00000000 0.62331973   0.62372045    0.26929976
Sodium                 0.62331973 1.00000000   0.99992860    0.20079562
```

```
Sodium.PDV               0.62372045  0.99992860  1.00000000   0.20242648
Carbohydrates            0.26929976  0.20079562  0.20242648   1.00000000
Carbohydrates.PDV        0.27099242  0.20103208  0.20266302   0.99962040
Dietary.Fiber            0.43494029  0.69438946  0.69391331   0.22457662
Dietary.Fiber.PDV        0.43981420  0.68999511  0.68946426   0.22825670
Sugars                  -0.13645937 -0.42653553 -0.42494302   0.76236214
Protein                  0.56095695  0.86980157  0.86986999   0.35212224
Vitamin.A.PDV            0.08005943  0.08306810  0.08325880   0.08380171
Vitamin.C.PDV           -0.08331470 -0.03076934 -0.03094523  -0.03472430
Calcium.PDV              0.13238169 -0.02407377 -0.02214497   0.58969874
Iron.PDV                 0.65316750  0.87159268  0.87074199   0.21024093
                   Carbohydrates.PDV Dietary.Fiber Dietary.Fiber.PDV
Calories                  0.78124203    0.53889351        0.54001419
Calories.from.Fat         0.46146307    0.58127379        0.57562058
Total.Fat                 0.46100479    0.58083725        0.57520633
Total.Fat.PDV             0.46029841    0.58059179        0.57503300
Saturated.Fat             0.59174306    0.35181784        0.34715171
Saturated.Fat.PDV         0.59165462    0.35683067        0.35179740
Trans.Fat                 0.46289148    0.05491839        0.05830128
Cholesterol               0.27266167    0.43557480        0.44026646
Cholesterol.PDV           0.27099242    0.43494029        0.43981420
Sodium                    0.20103208    0.69438946        0.68999511
Sodium.PDV                0.20266302    0.69391331        0.68946426
Carbohydrates             0.99962040    0.22457662        0.22825670
Carbohydrates.PDV         1.00000000    0.22405781        0.22728497
Dietary.Fiber             0.22405781    1.00000000        0.98635029
Dietary.Fiber.PDV         0.22728497    0.98635029        1.00000000
Sugars                    0.76228199   -0.29517842       -0.28701389
Protein                   0.35217806    0.64134483        0.65664825
Vitamin.A.PDV             0.08337579    0.34051837        0.36138045
Vitamin.C.PDV            -0.03545009    0.14193503        0.15001851
Calcium.PDV               0.59026335    0.02871108        0.05235903
Iron.PDV                  0.21064338    0.74041104        0.73781398
                       Sugars      Protein Vitamin.A.PDV Vitamin.C.PDV
Calories           0.25959812   0.78784745    0.10884404  -0.068747036
Calories.from.Fat -0.11528469   0.80791326    0.05673072  -0.087330731
Total.Fat         -0.11544573   0.80777296    0.05443396  -0.089353974
Total.Fat.PDV     -0.11576122   0.80792221    0.05403819  -0.089352581
Saturated.Fat      0.19773362   0.60302754    0.06497234  -0.179672005
Saturated.Fat.PDV  0.19592799   0.60658146    0.06537608  -0.178058754
Trans.Fat          0.33475579   0.38824850    0.07583347  -0.076612139
Cholesterol       -0.13551832   0.56156141    0.08023904  -0.082978213
Cholesterol.PDV   -0.13645937   0.56095695    0.08005943  -0.083314698
Sodium            -0.42653553   0.86980157    0.08306810  -0.030769341
Sodium.PDV        -0.42494302   0.86986999    0.08325880  -0.030945232
Carbohydrates      0.76236214   0.35212224    0.08380171  -0.034724302
Carbohydrates.PDV  0.76228199   0.35217806    0.08337579  -0.035450092
Dietary.Fiber     -0.29517842   0.64134483    0.34051837   0.141935030
Dietary.Fiber.PDV -0.28701389   0.65664825    0.36138045   0.150018508
Sugars             1.00000000  -0.17993965    0.04848771  -0.069846984
Protein           -0.17993965   1.00000000    0.21409799  -0.045776805
Vitamin.A.PDV      0.04848771   0.21409799    1.00000000   0.069171110
Vitamin.C.PDV     -0.06984698  -0.04577681    0.06917111   1.000000000
Calcium.PDV        0.60009291   0.32797109    0.17918994  -0.215379809
Iron.PDV          -0.36476708   0.79271946    0.13787902   0.001291741
                   Calcium.PDV     Iron.PDV
```

```
Calories            0.42842646   0.643551604
Calories.from.Fat   0.16103409   0.735894311
Total.Fat           0.16285952   0.734684514
Total.Fat.PDV       0.16203088   0.735477546
Saturated.Fat       0.40331077   0.578062375
Saturated.Fat.PDV   0.40113868   0.580487865
Trans.Fat           0.38533110   0.325475885
Cholesterol         0.13207674   0.654999546
Cholesterol.PDV     0.13238169   0.653167496
Sodium             -0.02407377   0.871592677
Sodium.PDV         -0.02214497   0.870741991
Carbohydrates       0.58969874   0.210240935
Carbohydrates.PDV   0.59026335   0.210643376
Dietary.Fiber       0.02871108   0.740411035
Dietary.Fiber.PDV   0.05235903   0.737813978
Sugars              0.60009291  -0.364767078
Protein             0.32797109   0.792719455
Vitamin.A.PDV       0.17918994   0.137879022
Vitamin.C.PDV      -0.21537981   0.001291741
Calcium.PDV         1.00000000   0.034148740
Iron.PDV            0.03414874   1.000000000
>
> ###Plot the graph
> cex.before <- par("cex")
> par(cex = 0.5)
> corrplot(Data_cor, method = "number", number.digits = 4)
> par(cex = cex.before)
>
> ###Finding the Variable with corelation of more than 0.85
> for (i in 1:nrow(Data_cor)){
+   correlations <-  which((Data_cor[i,] > 0.85) & (Data_cor[i,] != 1))
+   if(length(correlations)> 0){
+     print(colnames(CorData)[i])
+     print(correlations)
+   }
+ }
[1] "Calories"
Calories.from.Fat          Total.Fat      Total.Fat.PDV
              2                  3                  4
[1] "Calories.from.Fat"
    Calories       Total.Fat Total.Fat.PDV
          1               3             4
[1] "Total.Fat"
       Calories Calories.from.Fat      Total.Fat.PDV
             1                 2                  4
[1] "Total.Fat.PDV"
       Calories Calories.from.Fat         Total.Fat
             1                 2                 3
[1] "Saturated.Fat"
Saturated.Fat.PDV
              6
[1] "Saturated.Fat.PDV"
Saturated.Fat
          5
[1] "Cholesterol"
Cholesterol.PDV
```

```
                9
[1] "Cholesterol.PDV"
Cholesterol
          8
[1] "Sodium"
Sodium.PDV    Protein   Iron.PDV
        11         17         21
[1] "Sodium.PDV"
  Sodium   Protein  Iron.PDV
      10         17         21
[1] "Carbohydrates"
Carbohydrates.PDV
             13
[1] "Carbohydrates.PDV"
Carbohydrates
          12
[1] "Dietary.Fiber"
Dietary.Fiber.PDV
             15
[1] "Dietary.Fiber.PDV"
Dietary.Fiber
          14
[1] "Protein"
    Sodium Sodium.PDV
        10         11
[1] "Iron.PDV"
    Sodium Sodium.PDV
        10         11
>
>
> ##Problem 4 Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?
> ###Table of Category vs Colestrol Per
> Table_Cat_Colestrol = myData %>%   group_by(Category) %>% summarise(Cholesterol.Total.Per =
sum(Cholesterol.PDV))
> Table_Cat_Colestrol
# A tibble: 9 x 2
  Category          Cholesterol.Total.Per
  <fct>                             <int>
1 Beef & Pork                         434
2 Beverages                             5
3 Breakfast                          2140
4 Chicken & Fish                      681
5 Coffee & Tea                        891
6 Desserts                             34
7 Salads                              104
8 Smoothies & Shakes                  412
9 Snacks & Sides                       81
>
> ###Ploting the graph
> Graph_Cat_Colestrol = ggplot(Table_Cat_Colestrol, aes(x = reorder(Category, -Cholesterol.Total.Per), y
= Cholesterol.Total.Per)) +
+   geom_bar(fill = "#0073C2FF", stat = "identity") +
+   geom_text(aes(label = Cholesterol.Total.Per), vjust = -0.3) +
+   ggtitle("Category Vs Colestrol Percentage") +
+   xlab("Category") + ylab("Colestrol Percentage")
>
```

```
> Graph_Cat_Colestrol
>
> ####Breakfast is contributing maximum
>
>
> ##Problem 5 Which item contributes maximum to the Sodium intake?
> x= toString(myData[myData$Sodium == max(myData$Sodium),2])
> x
[1] "Chicken McNuggets (40 piece)"
>
> ####"Chicken McNuggets (40 piece)" has maximum sodium intake
>
>
> ##Problem 6
> myData %>%  select(Item,Saturated.Fat) %>%
+   arrange(desc(Saturated.Fat)) %>%
+   slice(1:4)
                                     Item Saturated.Fat
1 Big Breakfast with Hotcakes (Large Biscuit)           20
2               Chicken McNuggets (40 piece)           20
3            FrappÃ(c) Chocolate Chip (Large)             20
4     McFlurry with M&Mâ€TMs Candies (Medium)           20
> #### The 4 items having maximum Sodium intsake
> # Big Breakfast with Hotcakes (Large Biscuit)
> # Chicken McNuggets (40 piece)
> # FrappÃ(c) Chocolate Chip (Large)
> # McFlurry with M&Mâ???Ts Candies (Medium)
```

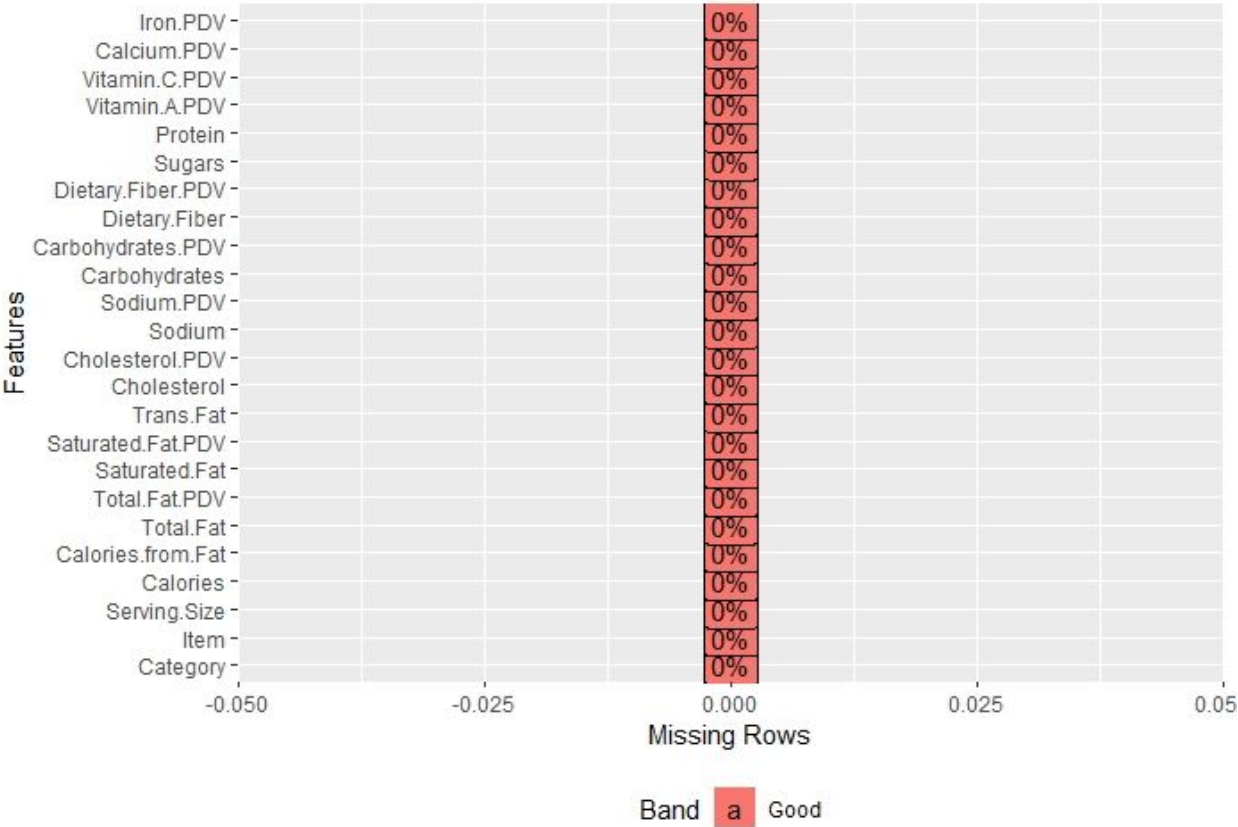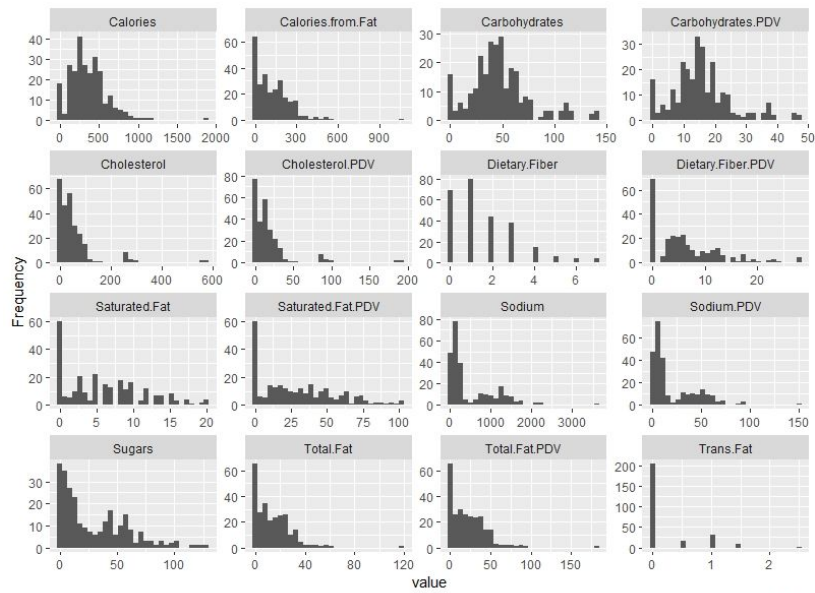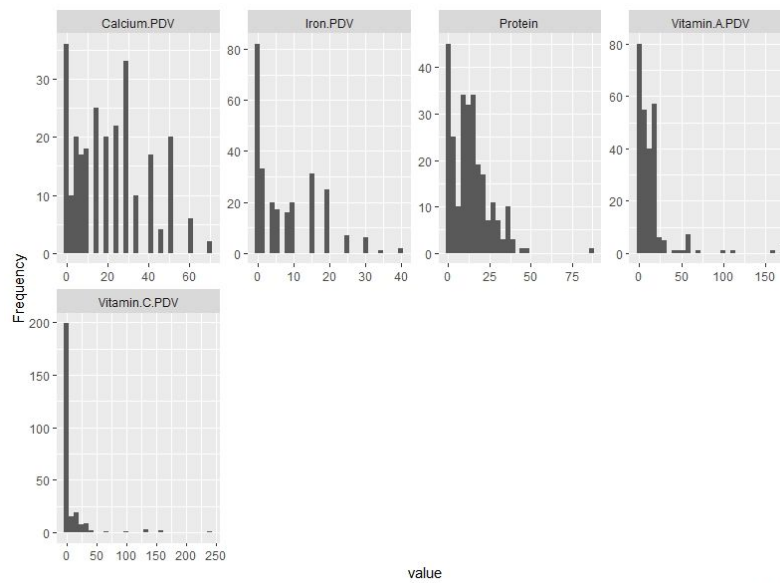# 6 Appendix B – Graphs and Plot

6.1 Missing Variable Plot
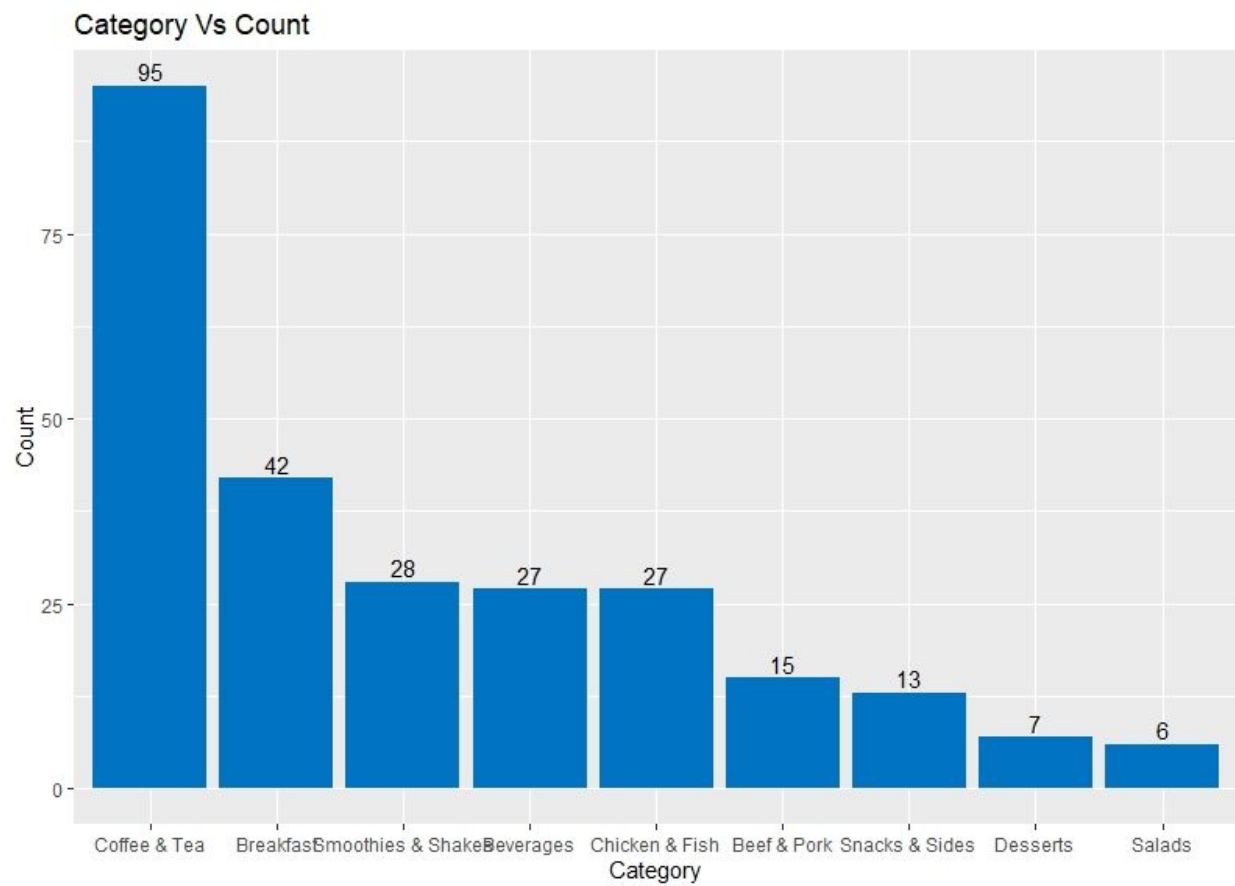
6.2 Histogram Plot of Numerical Variables
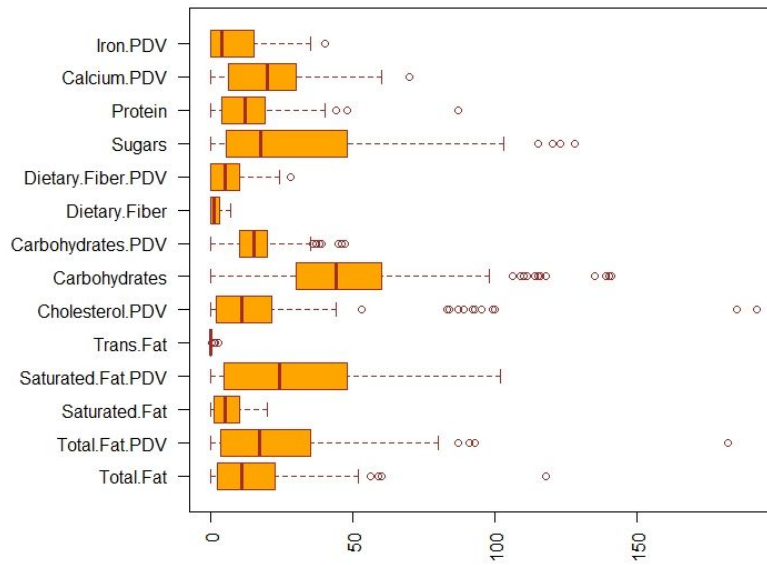
6.3 Category Vs Count

Category Vs Count

6.4 Box Plot of Variable

Box Plot of Variables - 1

Box Plot of Variables - 2

6.5 Correlations between Variables

	Calories	Calories.from.Fat	Total.Fat	Total.Fat.PDV	Saturated.Fat	Saturated.Fat.PDV	Trans.Fat	Cholesterol	Cholesterol.PDV	Sodium	Sodium.PDV	Carbohydrates	Carbohydrates.PDV	Dietary.Fiber	Dietary.Fiber.PDV	Sugars	Protein	Vitamin.A.PDV	Vitamin.C.PDV	Calcium.PDV	Iron.PDV
Calories	1	0.9046	0.9044	0.9041	0.8456	0.8476	0.5224	0.5964	0.5952	0.7123	0.7134	0.7815	0.7812	0.5389	0.54	0.2596	0.7878			0.4284	0.6436
Calories.from.Fat	0.9046	1	0.9997	0.9997	0.847	0.8496	0.4337	0.6822	0.6816	0.8466	0.8473	0.4617	0.4615	0.5813	0.5756	-0.1163	0.8079			0.161	0.7359
Total.Fat	0.9044	0.9997	1	0.9998	0.8467	0.8493	0.4315	0.6805	0.68	0.8462	0.8468	0.4612	0.461	0.5808	0.5752	-0.1154	0.8078			0.1629	0.7347
Total.Fat.PDV	0.9041	0.9997	0.9998	1	0.8474	0.85	0.433	0.6809	0.6804	0.8467	0.8474	0.4605	0.4603	0.5806	0.575	-0.1183	0.8079			0.162	0.7355
Saturated.Fat	0.8456	0.847	0.8467	0.8474	1	0.9993	0.6206	0.6312	0.6303	0.5841	0.5853	0.5913	0.5917	0.3518	0.3472	0.1977	0.603		-0.1737	0.4033	0.5781
Saturated.Fat.PDV	0.8476	0.8496	0.8493	0.85	0.9993	1	0.6202	0.6336	0.6327	0.5887	0.59	0.5913	0.5917	0.3568	0.3518	0.1958	0.6066		-0.1781	0.4011	0.5805
Trans.Fat	0.5224	0.4337	0.4315	0.433	0.6206	0.6202	1	0.2533	0.2515	0.1876	0.1883	0.4632	0.4629			0.3348	0.3882			0.3853	0.3255
Cholesterol	0.5964	0.6822	0.6805	0.6809	0.6312	0.6336	0.2533	1	0.9999	0.6244	0.6247	0.271	0.2727	0.4356	0.4403		0.5616			0.1121	0.655
Cholesterol.PDV	0.5952	0.6816	0.68	0.6804	0.6303	0.6327	0.2515	0.9999	1	0.6233	0.6237	0.2653	0.271	0.4345	0.4398		0.561			0.1324	0.6532
Sodium	0.7123	0.8466	0.8462	0.8467	0.5841	0.5887	0.1876	0.6244	0.6233	1	0.9999	0.2008	0.201	0.6944	0.69	-0.4265	0.8698				0.8716
Sodium.PDV	0.7134	0.8473	0.8468	0.8474	0.5853	0.59	0.1883	0.6247	0.6237	0.9999	1	0.2024	0.2027	0.6939	0.6895	-0.4249	0.8699				0.8707
Carbohydrates	0.7815	0.4617	0.4612	0.4605	0.5913	0.5913	0.4632	0.271	0.2653	0.2008	0.2024	1	0.9996	0.2248	0.2283	0.7624	0.3521			0.5897	0.2102
Carbohydrates.PDV	0.7812	0.4615	0.461	0.4603	0.5917	0.5917	0.4629	0.2727	0.271	0.201	0.2027	0.9996	1	0.2241	0.2273	0.7623	0.3522			0.5903	0.2106
Dietary.Fiber	0.5389	0.5813	0.5808	0.5806	0.3518	0.3568		0.4356	0.4345	0.6944	0.6939	0.2248	0.2241	1	0.9864	-0.2952	0.6413	0.3455	0.1415		0.7404
Dietary.Fiber.PDV	0.54	0.5756	0.5752	0.575	0.3472	0.3518		0.4403	0.4398	0.69	0.6895	0.2283	0.2273	0.9864	1	-0.287	0.6566	0.3614	0.15		0.7378
Sugars	0.2596	-0.1163	-0.1154	-0.1183	0.1977	0.1958	0.3348			-0.4265	-0.4249	0.7624	0.7623	-0.2952	-0.287	1	-0.1799			0.6001	-0.3648
Protein	0.7878	0.8079	0.8078	0.8079	0.603	0.6066	0.3882	0.5616	0.561	0.8698	0.8699	0.3521	0.3522	0.6413	0.6566	-0.1799	1	0.2141		0.328	0.7927
Vitamin.A.PDV														0.3455	0.3614		0.2141	1		0.1792	0.1379
Vitamin.C.PDV					-0.1737	-0.1781								0.1415	0.15				1	-0.2154	
Calcium.PDV	0.4284	0.161	0.1629	0.162	0.4033	0.4011	0.3853	0.1121	0.1324			0.5897	0.5903			0.6001	0.328	0.1792	-0.2154	1	
Iron.PDV	0.6436	0.7359	0.7347	0.7355	0.5781	0.5805	0.3255	0.655	0.6532	0.8716	0.8707	0.2102	0.2106	0.7404	0.7378	-0.3648	0.7927	0.1379			1

6.6 Category vs Cholesterol Percentage

Category Vs Colestrol Percentage