# MCDONALD'S ASSIGNMENT

-Report Submission

# Table of Contents

## 1. Objective of the assignment:

The objective of the assignment is to solve the questions asked in the assignment and performing a basic exploration of the dataset McDonald give a few insights.

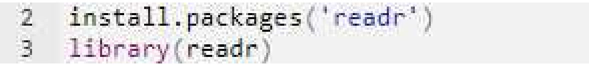
## 2. Exploration of the dataset:

### a. Setting up the environment and importing the dataset:

#### i. Calling out required libraries that will be used in the R code:

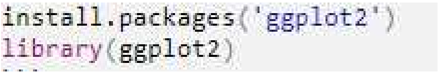
The libraries that are called for the assignment are as follows:

1. **Readr –** This library is required to import the dataset which is in CSV file into R environment as dataset



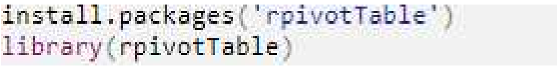
```
install.packages('readr')
library(readr)
```

2. **GGplot2 –** This library is required to create plots with colours and it provides a lot of customizations.



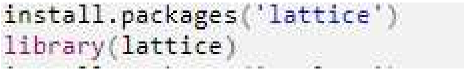
```
install.packages('ggplot2')
library(ggplot2)
```

3. **rpivotTable –** This library is useful for creating Pivot tables and Pivot graphs with ease.



```
install.packages('rpivotTable')
library(rpivotTable)
```

4. **lattice –** This library is used to easily display multivariate relationships.

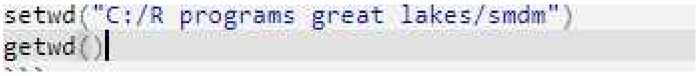


```
install.packages('lattice')
library(lattice)
```

### ii. Setting up the working directory:

The working directory is the location in the computer where the dataset is located. The dataset for this assignment is Mcdonad.

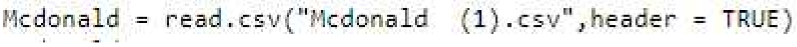
The working directory can be set using the function **setwd("Folder path")**. And the getwd() is used to get the current working directory.



```
setwd("C:/R programs great lakes/smdm")
getwd()
```

### iii. Importing the dataset and reading the dataset in R console:

The dataset in question is present in the form of a CSV(Comma Seperated Values) file in the working directory. The file name is "Mcdonald  (1).csv". We can use the **read.csv()** function with argument, **Header = TRUE** to import the dataset into R console as Data.Frame.



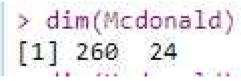
```
Mcdonald = read.csv("Mcdonald  (1).csv",header = TRUE)
```

The dataset is stored in the variable **Mcdonald.**

## b. Variable Identification:

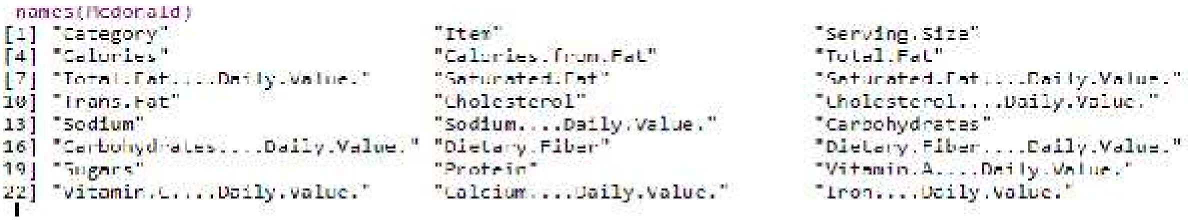
The functions that are used for Variable Identification of the datasets are as follows:

1. **dim() –** This function is used to get the dimensions of the dataset.



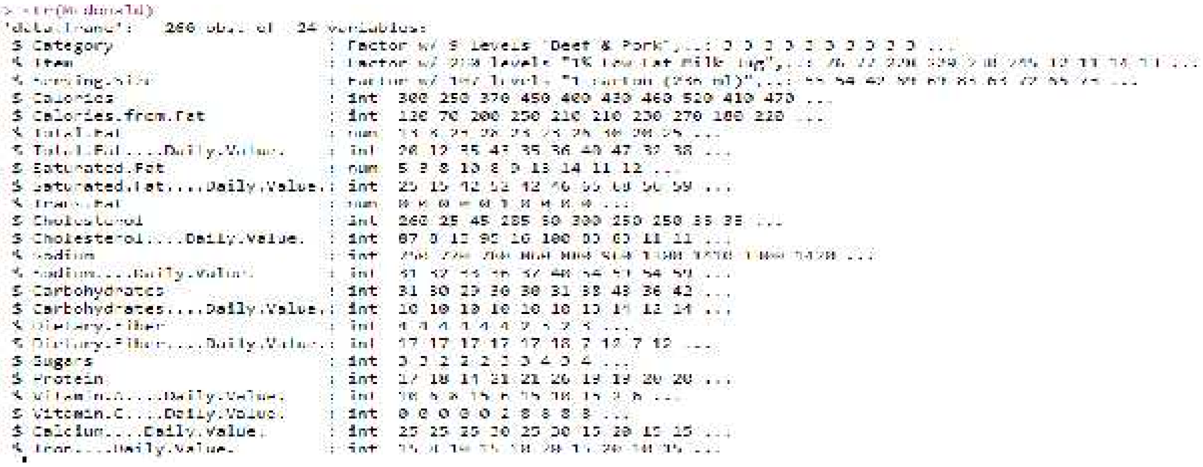
```
> dim(Mcdonald)
[1] 260  24
```

2. **names() –** This is used to get the names of the columns from the dataset.



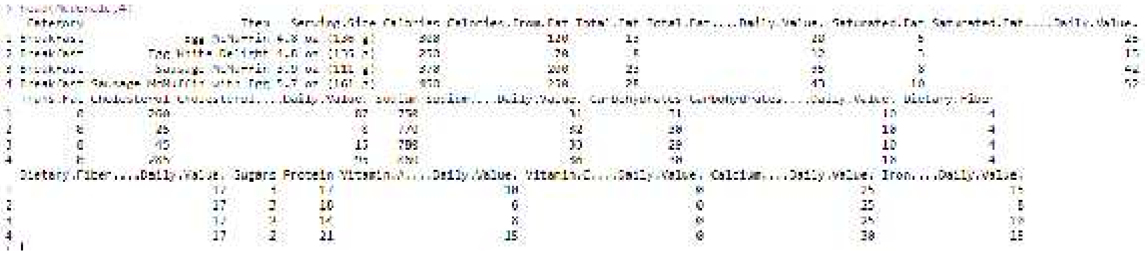
```
 names(Mcdonald)
[1] "Category"                        "Item"                           "Serving.Size"
[4] "Calories"                        "Calories.from.Fat"              "Total.Fat"
[7] "Total.Fat....Daily.Value."       "Saturated.Fat"                  "Saturated.Fat....Daily.Value."
10] "Trans.Fat"                       "Cholesterol"                    "Cholesterol....Daily.Value."
13] "Sodium"                          "Sodium....Daily.Value."         "Carbohydrates"
16] "Carbohydrates....Daily.Value."   "Dietary.Fiber"                  "Dietary.Fiber....Daily.Value."
19] "Sugars"                          "Protein"                        "Vitamin.A....Daily.Value."
22] "Vitamin.C....Daily.Value."       "Calcium....Daily.Value."        "Iron....Daily.Value."
```

3. **str() –** This function is used to the get basic information about each of the variables, like  type of variables, type of dataset, number of variables, first few values in the variable.



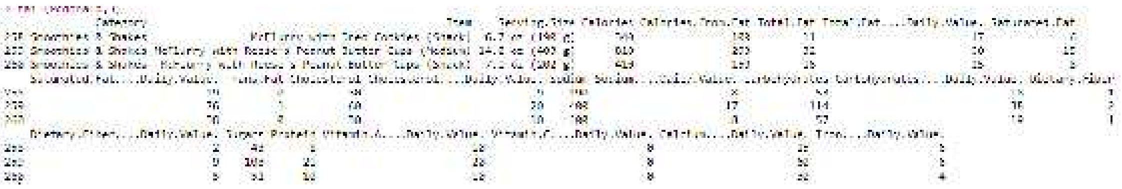
```
> str(McDonald)
'data.frame':   260 obs. of  24 variables:
 $ Category                     : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Item                         : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76 77 230 229 208 205 12 13 74 71 ...
 $ Serving.Size                 : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54 42 58 67 85 63 72 65 75 ...
 $ 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 40 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  750 770 780 860 880 810 1180 1340 1080 1290 ...
 $ Sodium....Daily.Value.       : int  31 32 33 36 37 34 49 56 45 54 ...
 $ 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 ...
```

4. **head() –** Give out the top value of the dataset. An additional argument will specify the number of top rows to print.



```
> head(McDonald,4)
   Category                            Item    Serving.Size Calories Calories.from.Fat Total.Fat Total.Fat....Daily.Value. Saturated.Fat Saturated.Fat....Daily.Value.
1 Breakfast                    Egg McMuffin 4.8 oz (136 g)      300               120        13                        20             5                            25
2 Breakfast                Egg White Delight 4.8 oz (135 g)     250                70         8                        12             3                            15
3 Breakfast                Sausage McMuffin 3.9 oz (111 g)      370               200        23                        35             8                            42
4 Breakfast Sausage McMuffin with Egg 5.7 oz (161 g)           450               250        28                        43            10                            52
  Trans.Fat Cholesterol Cholesterol....Daily.Value. Sodium Sodium....Daily.Value. Carbohydrates Carbohydrates....Daily.Value. Dietary.Fiber
1         0         260                          87    750                     31            31                            10             4
2         0          25                           8    770                     32            30                            10             4
3         0          45                          15    780                     33            29                            10             4
4         0         285                          95    860                     36            30                            10             4
  Dietary.Fiber....Daily.Value. Sugars Protein Vitamin.A....Daily.Value. Vitamin.C....Daily.Value. Calcium....Daily.Value. Iron....Daily.Value.
1                            17      3      17                        10                         0                      25                   15
2                            17      3      18                         6                         0                      25                    8
3                            17      2      14                         8                         0                      25                   10
4                            17      2      21                        15                         0                      30                   15
```

5. **tail() –** Give out the bottom value of the dataset. An additional argument will specify the number of bottom rows to print.



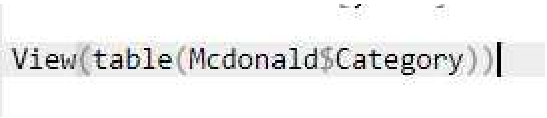
```
> tail(McDonald,3)
              Category                                               Item    Serving.Size Calories Calories.from.Fat Total.Fat Total.Fat....Daily.Value. Saturated.Fat
258 Smoothies & Shakes                   McFlurry with Oreo Cookies (Snack)  6.7 oz (190 g)      340               100        11                        17             7
259 Smoothies & Shakes McFlurry with Reese's Peanut Butter Cups (Medium) 14.2 oz (403 g)      810               290        32                        50            15
260 Smoothies & Shakes  McFlurry with Reese's Peanut Butter Cups (Snack)  7.3 oz (207 g)      410               150        16                        25             8
    Saturated.Fat....Daily.Value. Trans.Fat Cholesterol Cholesterol....Daily.Value. Sodium Sodium....Daily.Value. Carbohydrates Carbohydrates....Daily.Value. Dietary.Fiber
258                            34         0          30                           9    180                      8            54                            18             1
259                            76         1          60                          20    400                     17           114                            38             2
260                            38         0          30                          10    200                      8            57                            19             1
    Dietary.Fiber....Daily.Value. Sugars Protein Vitamin.A....Daily.Value. Vitamin.C....Daily.Value. Calcium....Daily.Value. Iron....Daily.Value.
258                             2     43       8                        10                         0                      25                    6
259                             9    103      21                        20                         0                      60                    6
260                             5     51      10                        10                         0                      30                    4
```

## The inferences made from the above outputs:

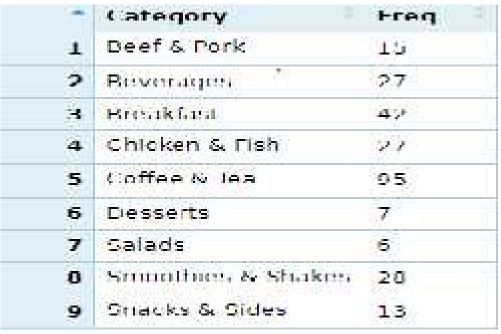
1) The dataset '**Mcdonald'** contains **260 rows** and **24 Columns.**
2) Some of the names of the columns are "**Category", "Item", "Calories", "Sugars", "Protein", etc.**
3) The first column **Category,** is a **factor** variable with 9 levels. This column contains **9** types of Food Categories which are available in Mcdonald's.
4) The column **Item,** has 260 character values indicating each food item that is present in the Mcdonald's. It means that this dataset has information on **260** different food items available at Mcdonald's.
5) The column **Serving Size**, has 107 levels indicating there are **107** food measurement factors for the **260** food items.
6) The columns starting from **Calories(4)** to **Iron….daily value(24)** are all either **numeric or integer** variables indicating that there are **no character values** in all of these variables.
7) Looking at the outputs of the **head()** and **tail()**, it can be inferred that the dataset is **not in an orderly manner,** the rows are placed randomly.

## c. Univariate Analysis:

The dataset contains two types of variables, **Categorical and Numerical.** The Categorical analysis can be carried out using the table () function.

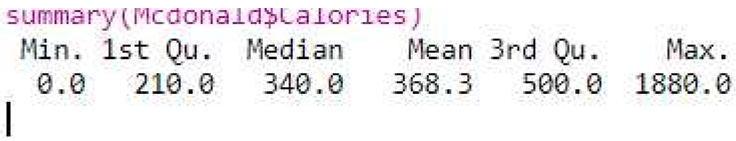


```
View(table(Mcdonald$Category))
```



|   | Category | Freq |
|---|---|---|
| 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 |

The Numerical Variable can be analysed by using **summary()** function.



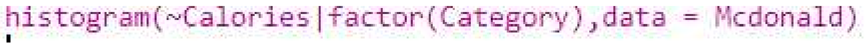
```
summary(Mcdonald$Calories)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.0   210.0   340.0   368.3   500.0  1880.0 
```

Inferences:

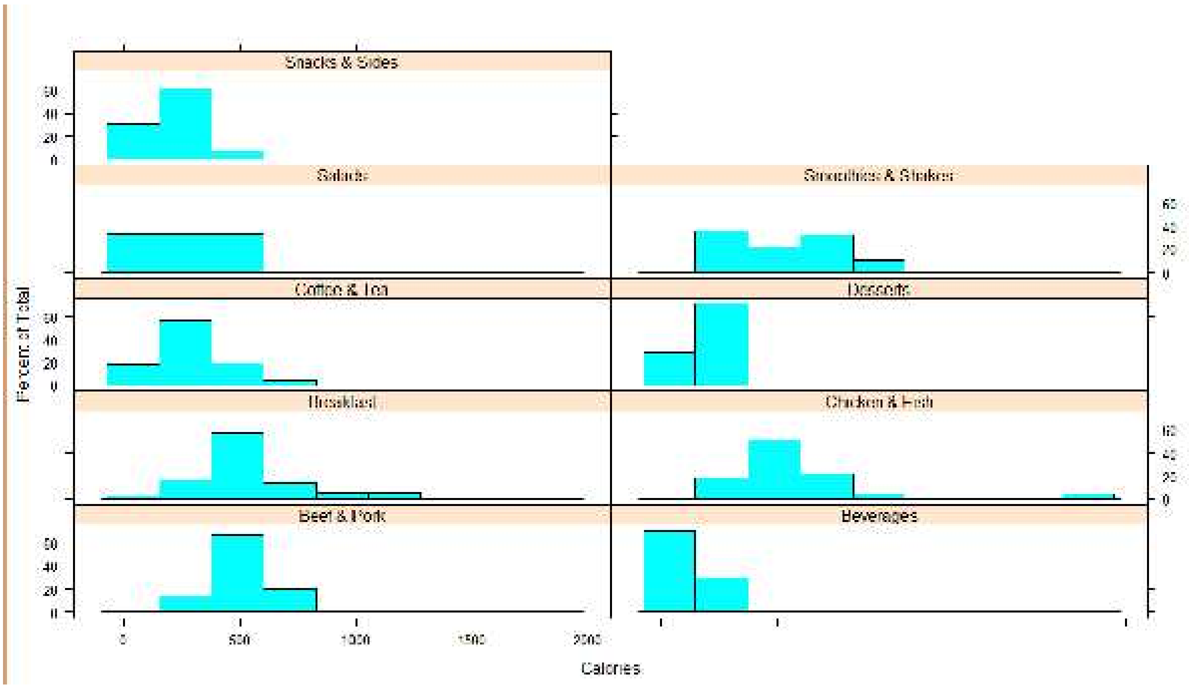
1) Coffee and Tea have the highest number of Items from all the categories.
2) There are least variety of items under Salads category in the Mcdonalds menu.
3) The highest calorific value from the Mcdonald's menu is **1880**.
4) The average calorific value of any item on the menu is around **368.3**

# d. Bi-Variate Analysis:

The Bi-Variate Analysis could be performed using the function **histogram()** from the **Lattice** library. We can create several histograms of a Numerical variable dependent on the categorical variable.



```
histogram(~Calories|factor(Category),data = Mcdonald)
```



<u>Inferences:</u>

1) When each of these categories are plotted in accordance with their calorific values, we can observe from the above graphs that all the graphs are **Left-Skewed**.
2) **Deserts** and **Beef and Pork** contribute to the highest amount of calories from the menu.
3) **Salads** can be judged as the **healthiest** of these categories because it amounts to the **lowest amount of calories** from all the categories in the menu.

## e. Outlier Identification:

The outliers can be described as the extreme most values away from the cluster of values that are situated around the mean. The Outlier Identification for this dataset has been done as part of solutions for one of the questions for this assignment. It can be found under *Solutions to the questions.*

## f. Variable transformation:

The variable transformation was done for this transformation to create a subset **Mcdonald2** from the dataset **Mcdonald**. This subset was created to remove the Categorical Variables which were present in the first four columns and to keep the rest of the numerical variables in the dataset.

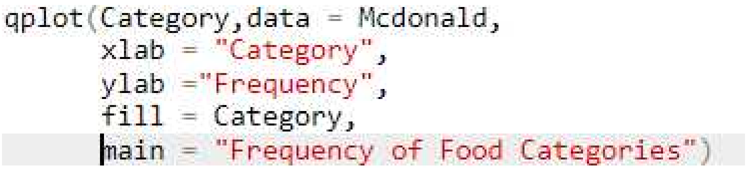
```
Mcdonald2 = Mcdonald[,4:24]
Mcdonald2
```

## g. Solutions to the questions asked:

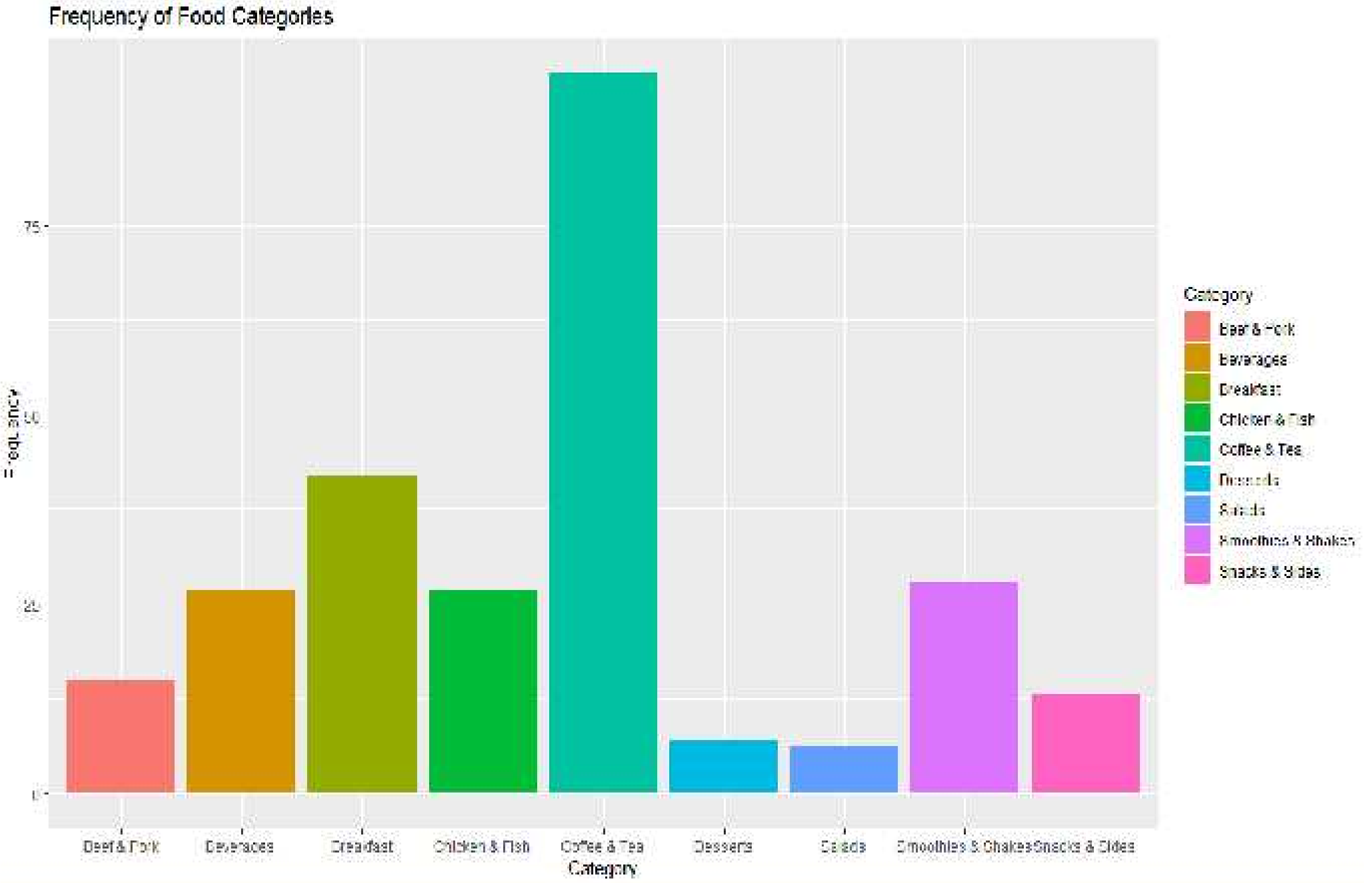
### 1. Plot graphically which food categories have the highest and lowest varieties.

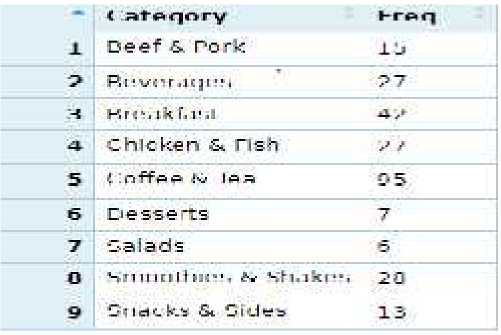
### Solution:

To find out which Categories have the highest and the lowest food items, we can plot a **bar graph** using the **qplot()** function from the **ggplot2** library. The below code prints a **Bar graph** of the Categories along with their **Frequencies.**



```
qplot(Category,data = Mcdonald,
      xlab = "Category",
      ylab ="Frequency",
      fill = Category,
      main = "Frequency of Food Categories")
```





| | Category | Freq |
|---|---|---|
| 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 |

## Inferences:

From the above graph and table, we can conclude that the Category **Coffee and Tea** have the **Highest** number of varieties and **Salads** have the **lowest** number of varieties.

*Therefore, Coffee and tea have the highest number of varieties and Salads have the lowest number of varieties.*

## 2. Which all variables have outliers?

### Solution:

There are total of 21 Numerical Variables. For the ease of usage, we use the subset of the dataset, **Mcdonald2.**

There are **3** steps involved for the solution of the problem. They are as follows:

i. The first and foremost step is identification of Outliers. For that, we create a **custom function** called outlier2 () with Column number as its argument. Upon executing, it gives out the number of Outliers in that Column.

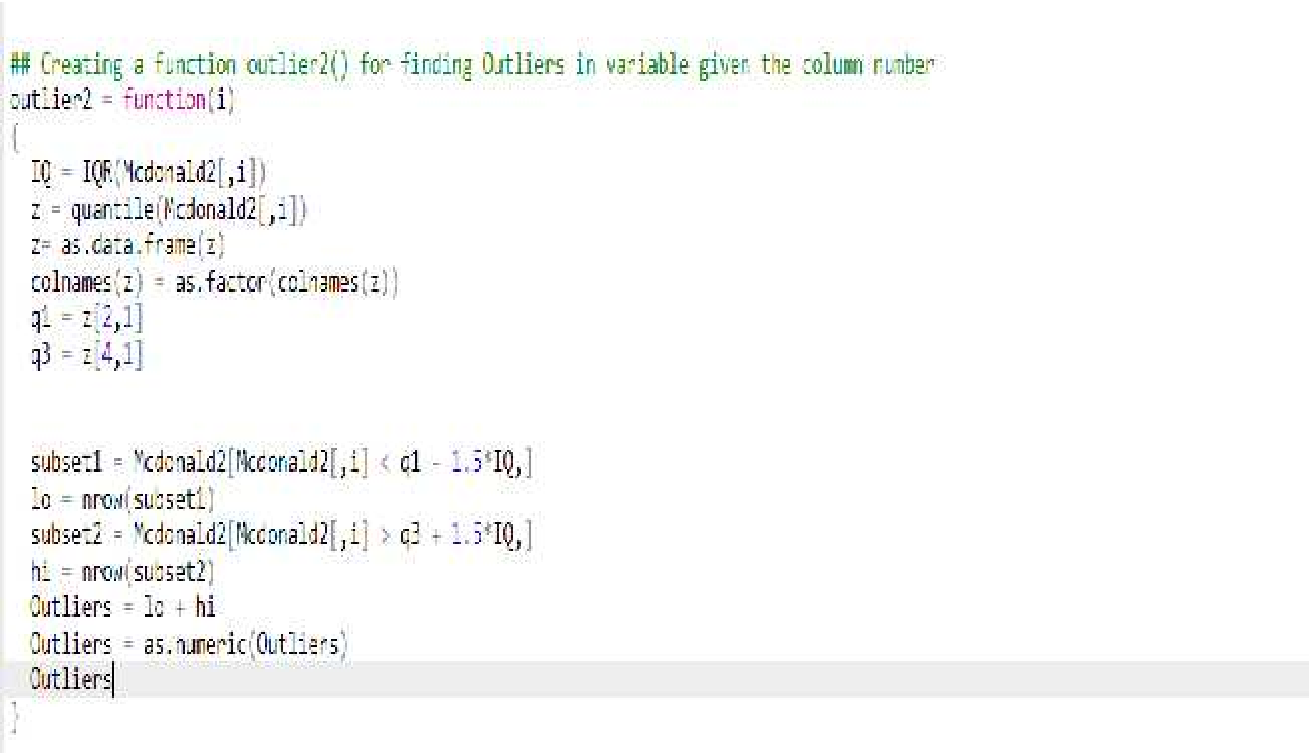
   The lower Outliers are all the values that are lower than (Q1 - 1.5*IQR), where Q1 = First Quartile,

   IQR = Inter Quartile Range

   The higher Outliers are all the values that are higher than (Q3 + 1.5*IQR), where Q3 = Third Quartile,

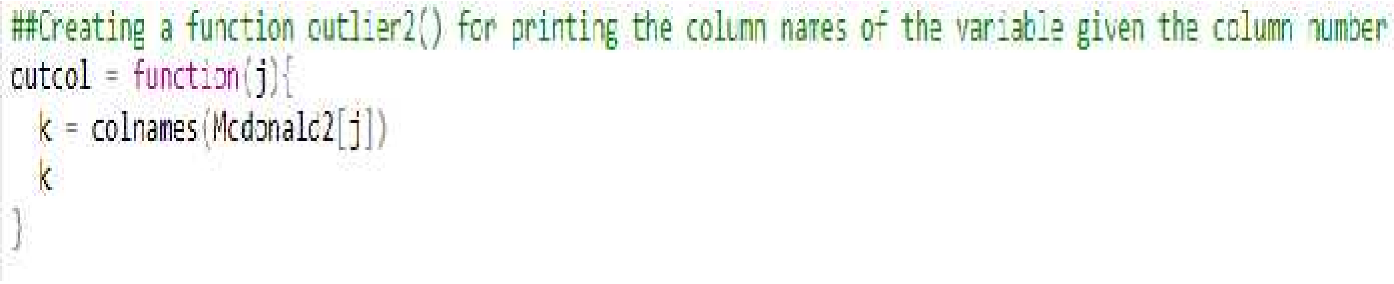
   IQR = Inter Quartile Range.

   The code for **outlier2()** is as follows:



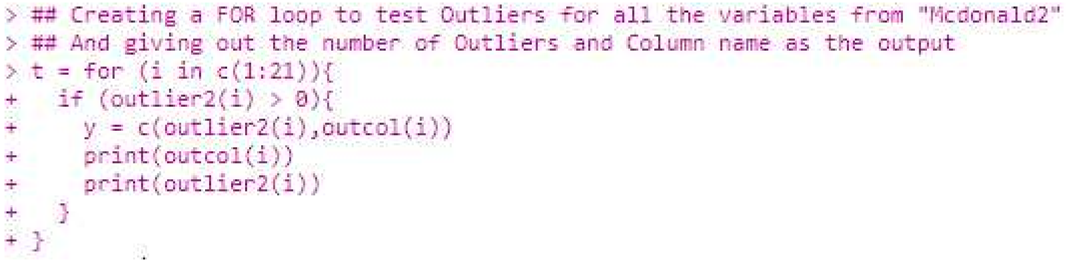
```
## Creating a function outlier2() for finding Outliers in variable given the column number
outlier2 = function(i)
{
  IQ = IQR(Mcdonald2[,i])
  z = quantile(Mcdonald2[,i])
  z= as.data.frame(z)
  colnames(z) = as.factor(colnames(z))
  q1 = z[2,1]
  q3 = z[4,1]


  subset1 = Mcdonald2[Mcdonald2[,i] < q1 - 1.5*IQ,]
  lo = nrow(subset1)
  subset2 = Mcdonald2[Mcdonald2[,i] > q3 + 1.5*IQ,]
  hi = nrow(subset2)
  Outliers = lo + hi
  Outliers = as.numeric(Outliers)
  Outliers
}
```

ii. The second step is creation of custom function **outcol2(),** which prints out the Column name when the Column number is given as the argument. The code is as below:



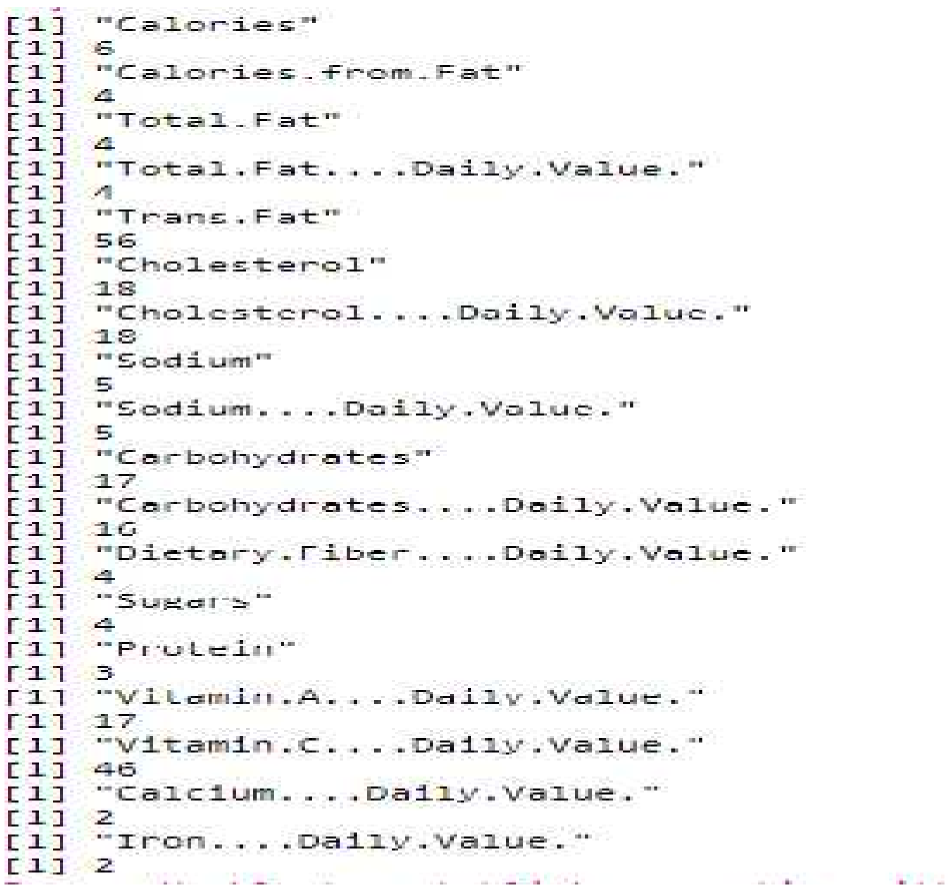
```
##Creating a function outlier2() for printing the column names of the variable given the column number
outcol = function(j){
  k = colnames(Mcdonald2[j])
  k
}
```

iii. The third and final step is creation of **FOR** loop and **IF** loop within the **FOR** loop for the printing out the Column names along with the number of outliers it contains. This is the most important part of the solution as it utilizes both the previously created functions. The code is as follows:



```
> ## Creating a FOR loop to test Outliers for all the variables from "Mcdonald2"
> ## And giving out the number of Outliers and Column name as the output
> t = for (i in c(1:21)){
+   if (outlier2(i) > 0){
+     y = c(outlier2(i),outcol(i))
+     print(outcol(i))
+     print(outlier2(i))
+   }
+ }
```

## Output:



```
[1] "Calories"
[1] 6
[1] "Calories.from.Fat"
[1] 4
[1] "Total.Fat"
[1] 4
[1] "Total.Fat....Daily.Value."
[1] 4
[1] "Trans.Fat"
[1] 56
[1] "Cholesterol"
[1] 18
[1] "Cholesterol....Daily.Value."
[1] 18
[1] "Sodium"
[1] 5
[1] "Sodium....Daily.Value."
[1] 5
[1] "Carbohydrates"
[1] 17
[1] "Carbohydrates....Daily.Value."
[1] 16
[1] "Dietary.Fiber....Daily.Value."
[1] 4
[1] "Sugars"
[1] 4
[1] "Protein"
[1] 3
[1] "Vitamin.A....Daily.Value."
[1] 17
[1] "Vitamin.C....Daily.Value."
[1] 46
[1] "Calcium....Daily.Value."
[1] 2
[1] "Iron....Daily.Value."
[1] 2
```

## Inferences:

1) The Variable **Vitamin.C....Daily.Value.** has **46** outliers
2) The Variable **Calcium…..Daily.Value and Iron…..Daily.Value** has **2** outliers.

***The Variables with outliers are***

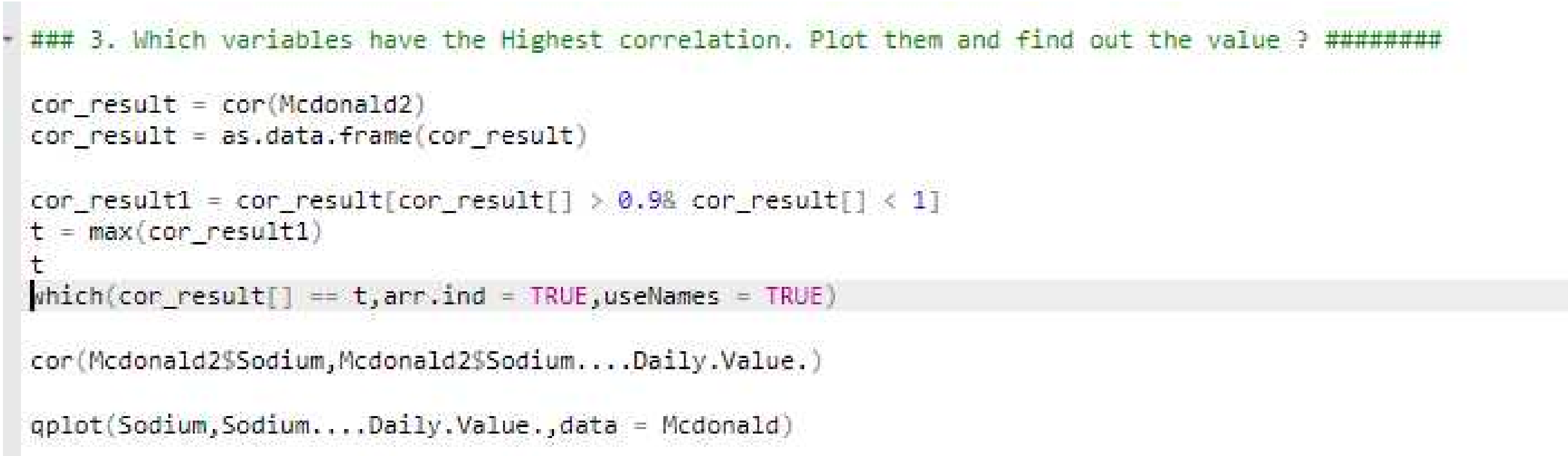
1) ***Calories,***
2) ***Calories.from.Fat,***
3) ***Total.Fat,***

***4) Total.Fat....Daily.Value.,***
***5) Trans.Fat,***
***6) Cholesterol,***
***7) Cholesterol....Daily.Value.,***
***8) Sodium,***
***9) Sodium....Daily.Value.,***
***10) Carbohydrates,***
***11) Carbohydrates....Daily.Value.,***
***12) Dietary.Fiber....Daily.Value.,***
***13) Sugars,***
***14) Protein,***
***15) Vitamin.A....Daily.Value.,***
***16) Vitamin.C....Daily.Value,***
***17) Calcium....Daily.Value.,***
***18) Iron....Daily.Value.***

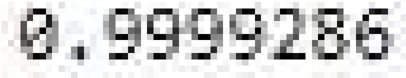
## 3. Which variables have the highest correlation? Plot them and find out the value?

### Solution:

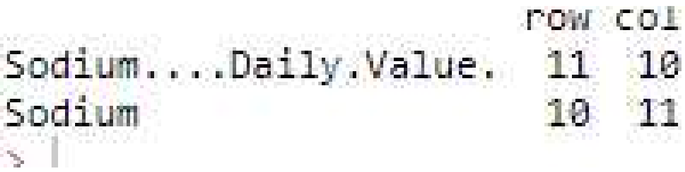
To find out which of the **21** variables have the highest correlation with each other, we apply the **cor()** function to the subset **Mcdonald2**. From the result of that function, we obtain a Data Frame which contains all the variables as rows and columns with their Correlation values as the elements. From that, we select the element with the highest Correlation value and find out the row name and column name of that element, giving us the two variables between which we have the highest correlation. A Scatter graph can be used to plot a graph between them. The code is as follows:



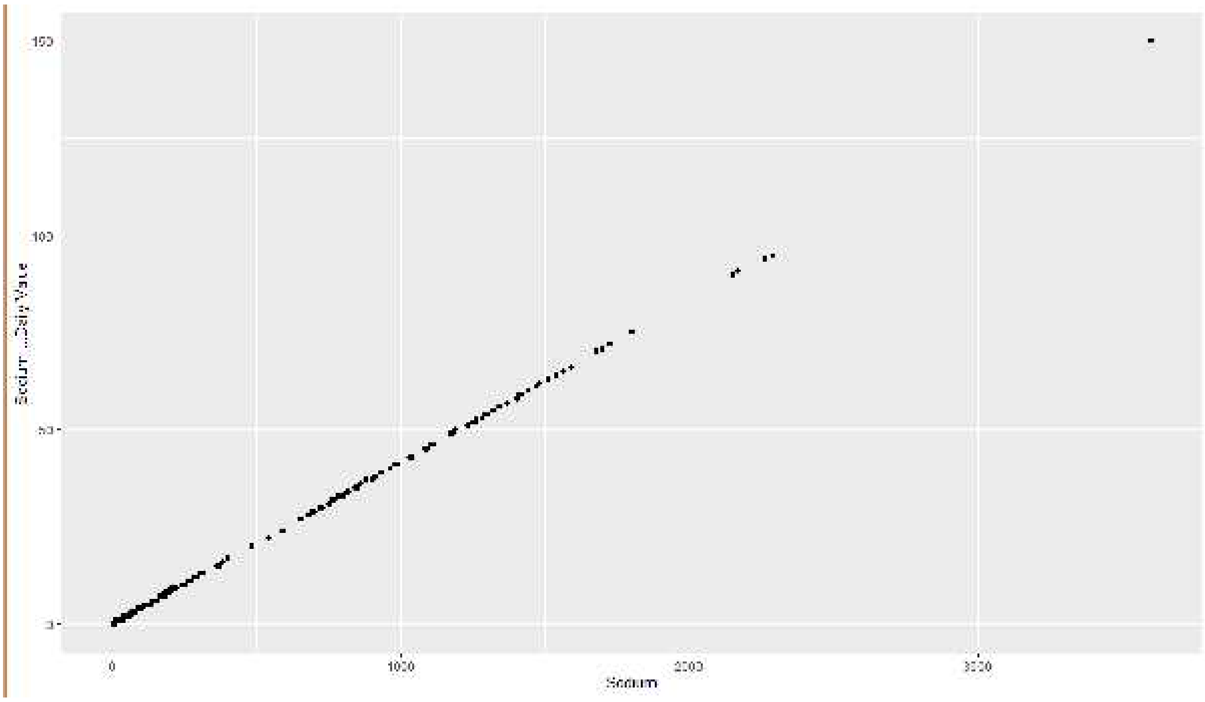
```
### 3. Which variables have the Highest correlation. Plot them and find out the value ? ########

cor_result = cor(Mcdonald2)
cor_result = as.data.frame(cor_result)

cor_result1 = cor_result[cor_result[] > 0.9& cor_result[] < 1]
t = max(cor_result1)
t
which(cor_result[] == t,arr.ind = TRUE,useNames = TRUE)

cor(Mcdonald2$Sodium,Mcdonald2$Sodium....Daily.Value.)

qplot(Sodium,Sodium....Daily.Value.,data = Mcdonald)
```



```
0.9999286
```



```
                        row col
Sodium....Daily.Value.   11  10
Sodium                   10  11
> 
```

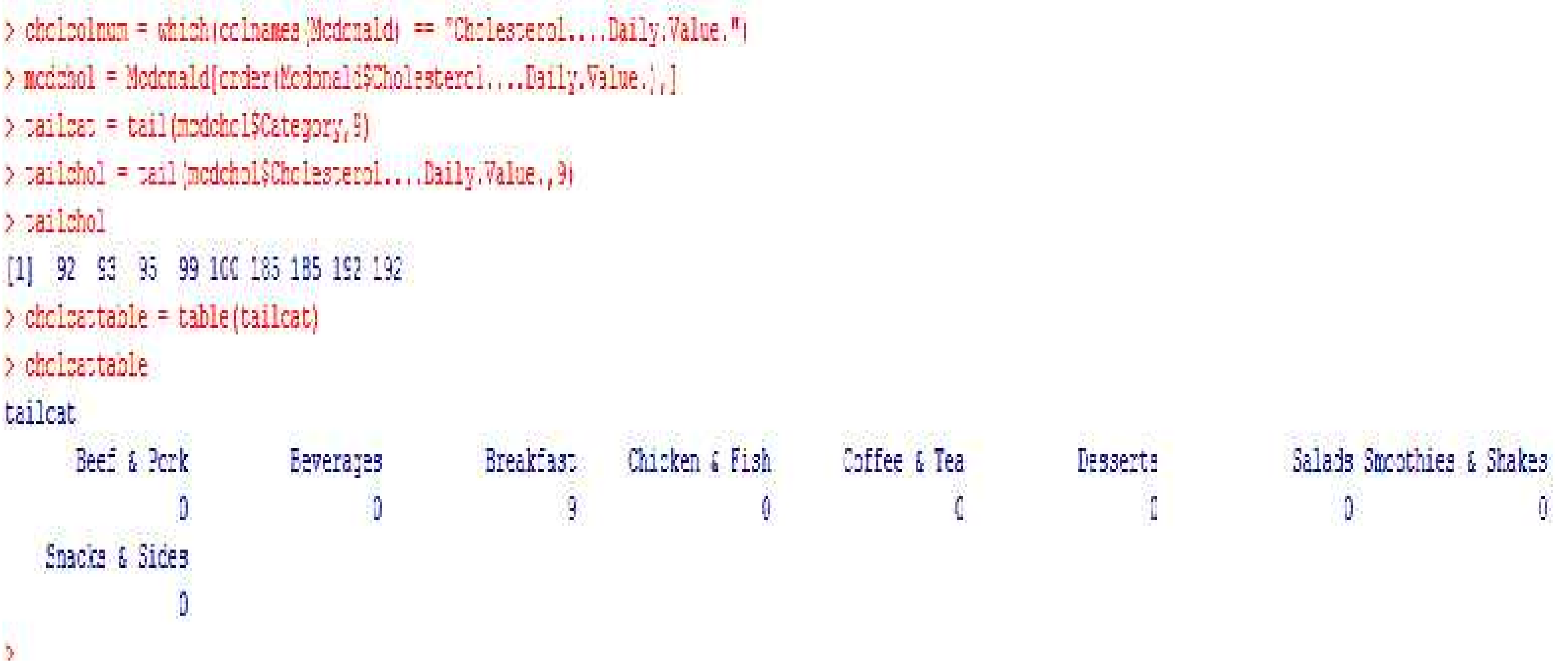


*The Variables **Sodium** and **Sodium.Daily.Value** are the two Variables with the highest correlation.*

## 4. Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?

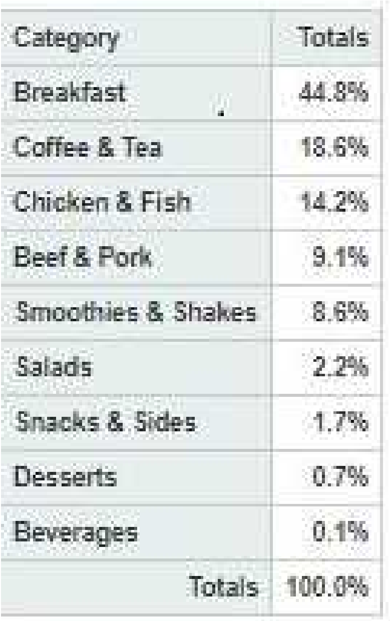
### Solution:

To find out which category contributes the maximum % of Cholesterol, we first arrange the whole dataset in the increasing order of the value of **Cholesterol (%daily value)** and assign to the dataset **mcdchol**. Then from that dataset, we use the **tail ()** function to find out the **highest Cholesterol (%daily value)** and correspondingly find the categories to which they correspond to. The category which **appear more number of times and correspond to the highest values** will be the category which **contributes the most to the Cholesterol (%daily value).** The code is as follows:



```
> cholcolnum = which(colnames(Mcdonald) == "Cholesterol....Daily.Value.")
> mcdchol = Mcdonald[order(Mcdonald$Cholesterol....Daily.Value.),]
> tailcat = tail(mcdchol$Category,9)
> tailchol = tail(mcdchol$Cholesterol....Daily.Value.,9)
> tailchol
[1]  92  93  95  99 100 185 185 192 192
> cholcattable = table(tailcat)
> cholcattable
tailcat
       Beef & Pork          Beverages          Breakfast     Chicken & Fish       Coffee & Tea           Desserts Salads Smoothies & Shakes
                 0                  0                  9                  0                  0                  0                  0                  0
    Snacks & Sides
                 0
>
```

```
rpivotTable(Mcdonald)
```



| Category | Totals |
|---|---|
| Breakfast | 44.8% |
| Coffee & Tea | 18.6% |
| Chicken & Fish | 14.2% |
| Beef & Pork | 9.1% |
| Smoothies & Shakes | 8.6% |
| Salads | 2.2% |
| Snacks & Sides | 1.7% |
| Desserts | 0.7% |
| Beverages | 0.1% |
| Totals | 100.0% |

## Inferences:

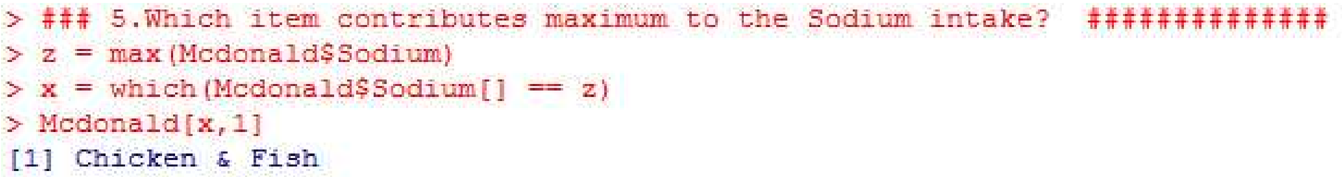
From the above outputs of the tail() function, we can infer that category **Breakfast**, contributes the most. This is also evident from the Pivot table made between **Breakfast and Sodium(%Daily.Value)** which shows its contribution is 44.8%.

*The Breakfast is the Category which contributes the most to the Sodium($Daily.Value)*

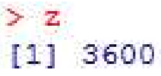
## 5. Which item contributes maximum to the Sodium intake?

## Solution:

We need to find the item which has the **highest value** of **Sodium** intake. To do this, we need to find the **highest value** of Sodium variable using the function **max().** Then we find the Corresponding **Item** name to that value using the **which()** function which gives out the Name of the Item. The code is as follows:



```
> ### 5.Which item contributes maximum to the Sodium intake?  ##############
> z = max(Mcdonald$Sodium)
> x = which(Mcdonald$Sodium[] == z)
> Mcdonald[x,1]
[1] Chicken & Fish
```



```
> z
[1] 3600
```

## Inferences:

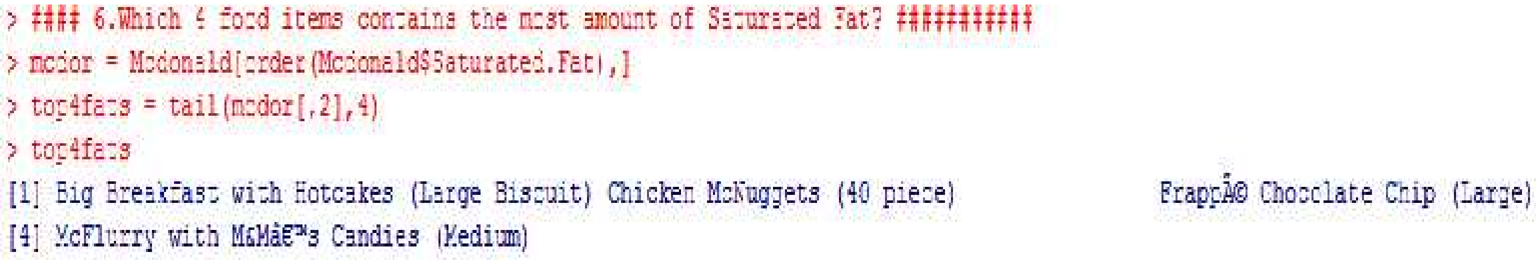
From the above outputs, we can see that the item **Chicken and Fish** has the highest amount of **Sodium** which is **3600.**

*The item Chicken and Fish contributes the more Sodium than the rest of the items.*

## 6. Which 4 food items contains the most amount of Saturated Fat?

## Solution:

First and foremost, we arrange the dataset into **ascending order** of the values of **Saturated Fat** and assign it to a new Data Frame called **mcdor.** Then we use the **tail()** function and get the **last four food items** which correspond **to the top 4 values** of Saturated Fat. The code is as follows:



```
> #### 6.Which 4 food items contains the most amount of Saturated Fat? ###########
> mcdor = Mcdonald[order(Mcdonald$Saturated.Fat),]
> top4fats = tail(mcdor[,2],4)
> top4fats
[1] Big Breakfast with Hotcakes (Large Biscuit) Chicken McNuggets (40 piece)            FrappÃ© Chocolate Chip (Large)
[4] McFlurry with M&Mâ€™s Candies (Medium)
```

## Inferences:

The above output shows that four variables which contribute the most **Saturated Fat**, all belong to different **Category**.

*The items which contribute the most Saturated Fat are,*

1) *Big Breakfast with Hotcakes (Large Biscuit)*
2) *Chicken McNuggets (40 piece)*
3) *FrappÃ© Chocolate Chip (Large)*
4) *McFlurry with M&Mâ€™s Candies (Medium)*

## i. Conclusion:

The dataset **Mcdonald** is a very useful information of ingredients of each every item found in the Mcdonald’s menu. If the right items are chosen off the menu, each and every customer will be leaving with a **Happy and Healthy meal.**

## Inferences:

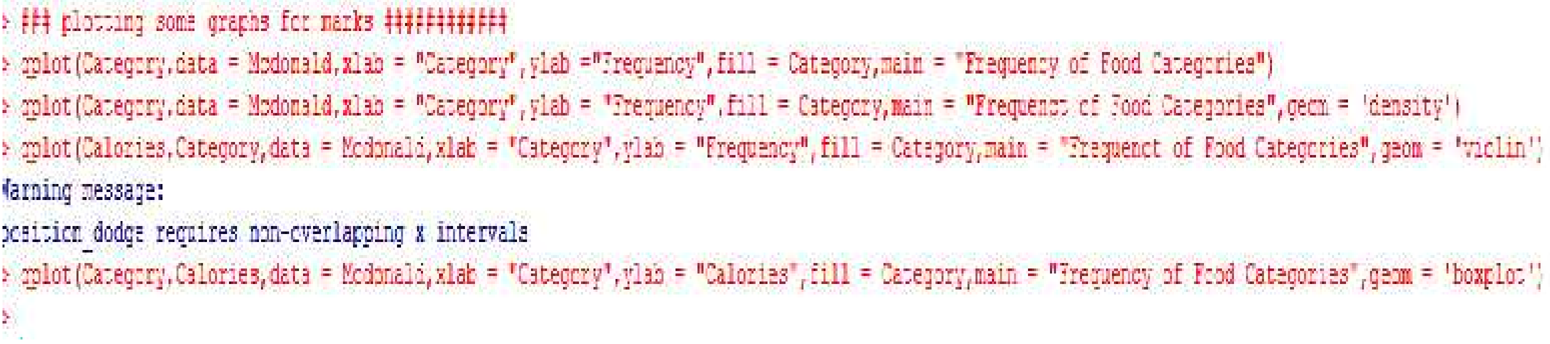
1) Even though they are less variety of items, **Salads** are the healthiest items due to their least calorific values.

2) Persons with High Blood Pressure must avoid **Chicken and Fish** as its high **Sodium** value increases the Blood Pressure even more.

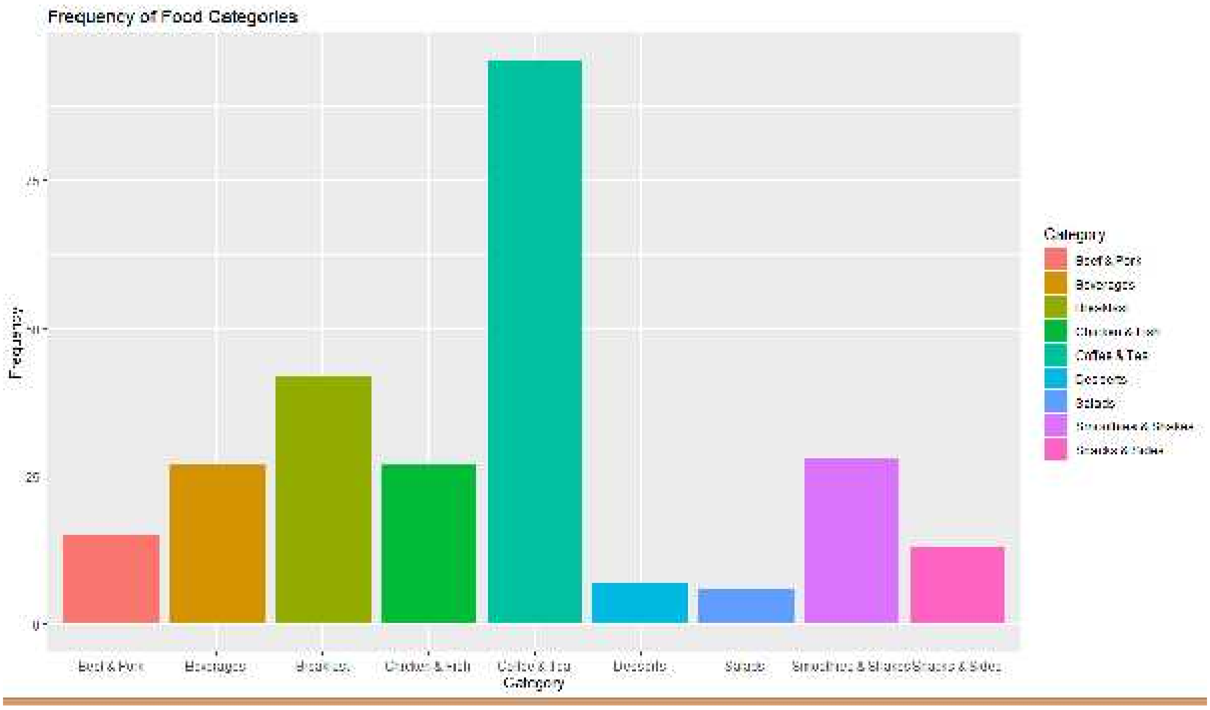
3) For the patients with heart diseases, all the items in the **Breakfast** category are rich in Cholesterol. Therefore, they can avoid the items in the **Breakfast** category and opt for other items.

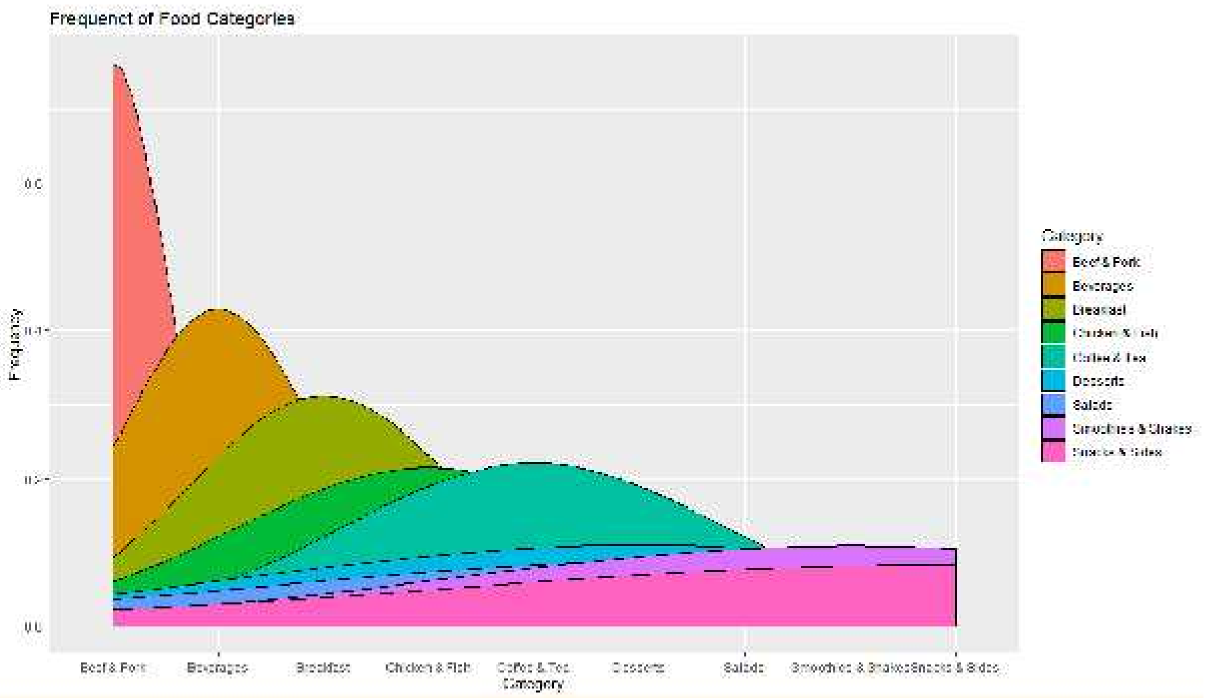
*5)* The four items**(**B*ig Breakfast with Hotcakes (Large Biscuit), Chicken McNuggets, FrappÃ© Chocolate Chip (Large), McFlurry with M&Mâ€™s Candies (Medium)***)** must be avoided by patients with heart problems at any cost as they are rich in **Saturated fats** and contribute a lot to increase in Cholestrol levels in the body.

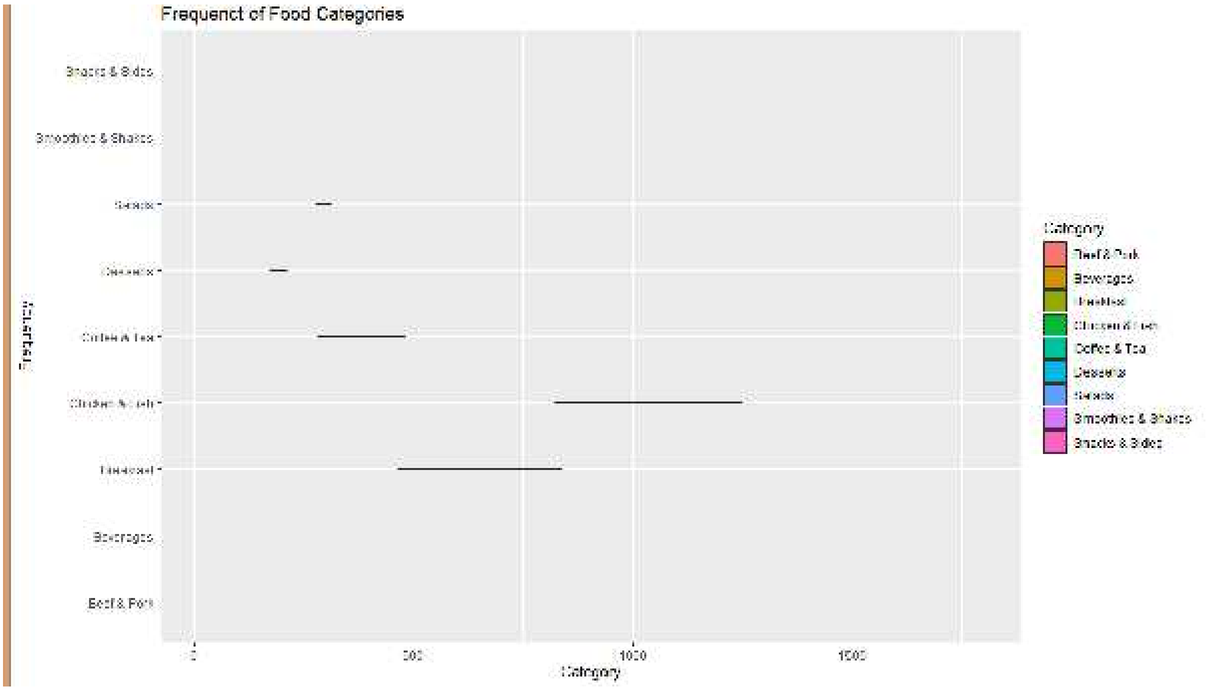
# Some other graphs related to Mcdonald dataset:

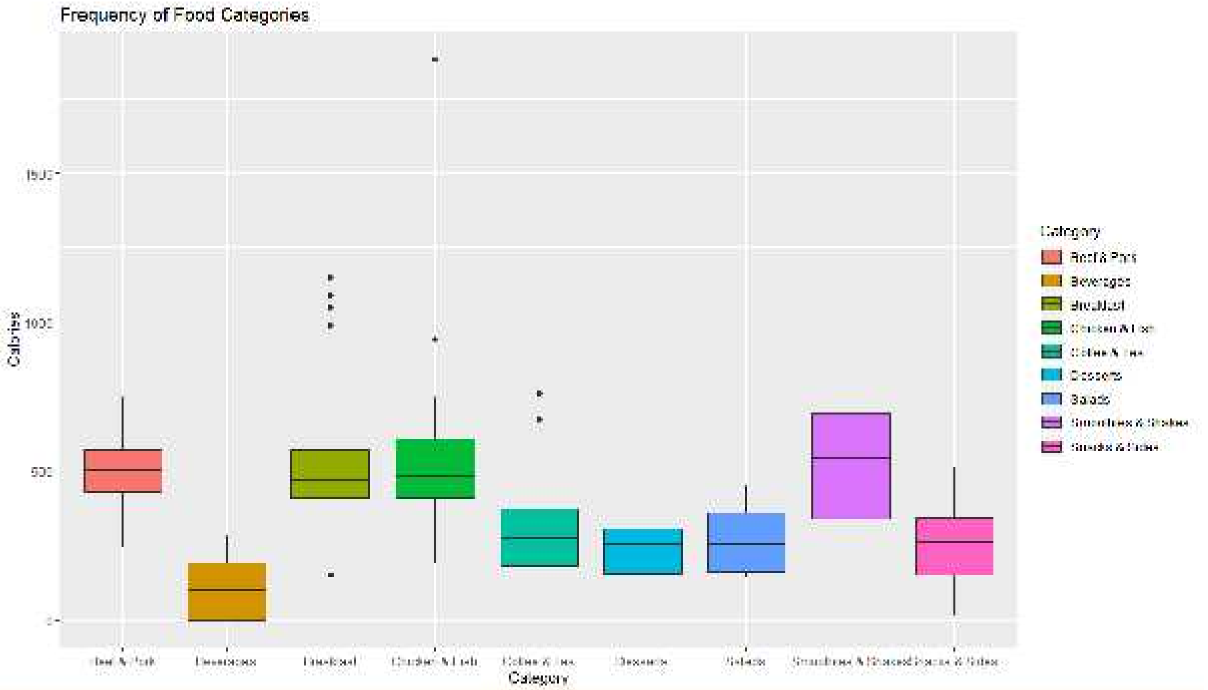


```
> ### plotting some graphs for marks ###########
> qplot(Category,data = Mcdonald,xlab = "Category",ylab ="Frequency",fill = Category,main = "Frequency of Food Categories")
> qplot(Category,data = Mcdonald,xlab = "Category",ylab = "Frequency",fill = Category,main = "Frequency of Food Categories",geom = 'density')
> qplot(Calories,Category,data = Mcdonald,xlab = "Category",ylab = "Frequency",fill = Category,main = "Frequenct of Food Categories",geom = 'violin'
Warning message:
position_dodge requires non-overlapping x intervals
> qplot(Category,Calories,data = Mcdonald,xlab = "Category",ylab = "Calories",fill = Category,main = "Frequency of Food Categories",geom = 'boxplot'
>
```




Frequenct of Food Categories
Frequency
0.6
0.4
0.2
0.0
Beef & Pork
Beverages
Breakfast
Chicken & Fish
Coffee & Tea
Desserts
Salads
Smoothies & Shakes
Snacks & Sides
Category
Category
Beef & Pork
Beverages
Breakfast
Chicken & Fish
Coffee & Tea
Desserts
Salads
Smoothies & Shakes
Snacks & Sides


Frequenct of Food Categories
Snacks & Sides
Smoothies & Shakes
Salads
Desserts
Coffee & Tea
Chicken & Fish
Breakfast
Beverages
Beef & Pork
Frequency
0
500
1000
1500
Category
Category
Beef & Pork
Beverages
Breakfast
Chicken & Fish
Coffee & Tea
Desserts
Salads
Smoothies & Shakes
Snacks & Sides



# j. Appendix - A(Source Code):

```
setwd("C:/R programs great lakes/smdm")
getwd()
"C:/R programs great lakes/smdm"
library(readr)
library(lattice)
library(rpivotTable)
library(ggplot2)

####Importing the dataset
> Mcdonald = read.csv("Mcdonald  (1).csv",header = TRUE)
Warning message:
In dontCheck(fnname) : reached elapsed time limit
> dim(Mcdonald)
[1] 260  24
> summary(Mcdonald)
             Category                                    Item           Serving.Size    Calories      Calories.from.Fat   Total.Fat
 Coffee & Tea       :95   1% Low Fat Milk Jug                      :  1   16 fl oz cup: 45   Min.   :    0.0    Min.   :    0.0    Min.   :  0.000
 Breakfast          :42   Apple Slices                          :  1   12 fl oz cup: 38   1st Qu.: 210.0   1st Qu.:  20.0    1st Qu.:  2.375
 Smoothies & Shakes:28   Bacon Buffalo Ranch McChicken          :  1   22 fl oz cup: 20   Median : 340.0   Median : 100.0    Median : 11.000
 Beverages          :27   Bacon Cheddar McChicken                 :  1   20 fl oz cup: 16   Mean   : 368.3   Mean   : 127.1    Mean   : 14.165
 Chicken & Fish     :27   Bacon Clubhouse Burger                  :  1   21 fl oz cup:  7   3rd Qu.: 500.0   3rd Qu.: 200.0    3rd Qu.: 22.250
 Beef & Pork        :15   Bacon Clubhouse Crispy Chicken Sandwich:  1   30 fl oz cup:  7   Max.   :1880.0   Max.   :1060.0    Max.   :118.000
 (Other)            :26   (Other)                                 :254   (Other)     :127
 Total.Fat....Daily.Value. Saturated.Fat    Saturated.Fat....Daily.Value.   Trans.Fat       Cholesterol     Cholesterol....Daily.Value.     Sodium
 Min.   :  0.00           Min.   : 0.000   Min.   :  0.00               Min.   :0.0000   Min.   :  0.00   Min.   :  0.00              Min.   :   0.0
 1st Qu.:  3.75           1st Qu.: 1.000   1st Qu.:  4.75               1st Qu.:0.0000   1st Qu.:  5.00   1st Qu.:  2.00              1st Qu.: 107.5
 Median : 17.00           Median : 5.000   Median : 24.00               Median :0.0000   Median : 35.00   Median : 11.00              Median : 190.0
 Mean   : 21.82           Mean   : 6.008   Mean   : 29.97               Mean   :0.2038   Mean   : 54.94   Mean   : 18.39              Mean   : 495.8
 3rd Qu.: 35.00           3rd Qu.:10.000   3rd Qu.: 48.00               3rd Qu.:0.0000   3rd Qu.: 65.00   3rd Qu.: 21.25              3rd Qu.: 865.0
 Max.   :182.00           Max.   :20.000   Max.   :102.00               Max.   :2.5000   Max.   :575.00   Max.   :192.00              Max.   :3600.0
 Sodium....Daily.Value. Carbohydrates    Carbohydrates....Daily.Value. Dietary.Fiber   Dietary.Fiber....Daily.Value.     Sugars          Protein
 Min.   :  0.00         Min.   :  0.00   Min.   : 0.00                 Min.   :0.000   Min.   : 0.000                Min.   :  0.00   Min.   : 0.00
 1st Qu.:  4.75         1st Qu.: 30.00   1st Qu.:10.00                 1st Qu.:0.000   1st Qu.: 0.000                1st Qu.:  5.75   1st Qu.: 4.00
 Median :  8.00         Median : 44.00   Median :15.00                 Median :1.000   Median : 5.000                Median : 17.50   Median :12.00
 Mean   : 20.68         Mean   : 47.35   Mean   :15.78                 Mean   :1.631   Mean   : 6.531                Mean   : 29.42   Mean   :13.34
 3rd Qu.: 36.25         3rd Qu.: 60.00   3rd Qu.:20.00                 3rd Qu.:3.000   3rd Qu.:10.000                3rd Qu.: 48.00   3rd Qu.:19.00
 Max.   :150.00         Max.   :141.00   Max.   :47.00                 Max.   :7.000   Max.   :28.000                Max.   :128.00   Max.   :87.00

 Vitamin.A....Daily.Value. Vitamin.C....Daily.Value. Calcium....Daily.Value. Iron....Daily.Value.
 Min.   :  0.00            Min.   :  0.000           Min.   : 0.00           Min.   : 0.000
 1st Qu.:  2.00            1st Qu.:  0.000           1st Qu.: 6.00           1st Qu.: 0.000
 Median :  8.00            Median :  0.000           Median :20.00           Median : 4.000
 Mean   : 13.43            Mean   :  8.535           Mean   :20.97           Mean   : 7.735
 3rd Qu.: 15.00            3rd Qu.:  4.000           3rd Qu.:30.00           3rd Qu.:15.000
 Max.   :170.00            Max.   :240.000           Max.   :70.00           Max.   :40.000

> str(Mcdonald)
'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 ...
> attach(Mcdonald)
The following objects are masked from Mcdonald (pos = 3):

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

The following objects are masked from Mcdonald (pos = 4):

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

The following objects are masked from Mcdonald (pos = 5):

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

The following objects are masked from Mcdonald (pos = 6):

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

The following objects are masked from Mcdonald (pos = 7):

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

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

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

> Mcdonald2 = Mcdonald[,4:24]
>
> histogram(~Mcdonald[,5]|factor(Mcdonald[,2]))
>
>
> ### 1. Plot graphically which food categories have the highest and lowest varieties. ##########
> catplottable = table(Category)
> catplottable
Category
      Beef & Pork          Beverages          Breakfast     Chicken & Fish        Coffee & Tea           Desserts          Salads Smoothies & Shakes
               15                 27                 42                 27                 95                  7                  6                 28
    Snacks & Sides
               13
> catplottable = as.data.frame(catplottable)
> qplot(Category,data = Mcdonald,xlab = "Category",ylab ="Frequency",fill = Category,main = "Frequency of Food Categories")
>
>
> ### 1. Plot graphically which food categories have the highest and lowest varieties. ##########
> catplottable = table(Category)
> catplottable
Category
      Beef & Pork          Beverages          Breakfast     Chicken & Fish        Coffee & Tea           Desserts          Salads Smoothies & Shakes
               15                 27                 42                 27                 95                  7                  6                 28
    Snacks & Sides
               13
> catplottable = as.data.frame(catplottable)
> qplot(Category,data = Mcdonald,xlab = "Category",ylab ="Frequency",fill = Category,main = "Frequency of Food Categories")
>
>
>
> #### 2. Which all variables have an outlier?  ################
>
> ## Creating a function outlier2() for finding Outliers in variable given the column number
> outlier2 = function(i)
+ {
+   IQ = IQR(Mcdonald2[,i])
+   z = quantile(Mcdonald2[,i])
+   z= as.data.frame(z)
+   colnames(z) = as.factor(colnames(z))
+   q1 = z[2,1]
+   q3 = z[4,1]
+
+
+   subset1 = Mcdonald2[Mcdonald2[,i] < q1 - 1.5*IQ,]
+   lo = nrow(subset1)
+   subset2 = Mcdonald2[Mcdonald2[,i] > q3 + 1.5*IQ,]
+   hi = nrow(subset2)
+   Outliers = lo + hi
+   Outliers = as.numeric(Outliers)
+   Outliers
+ }
>
> ##Creating a function outlier2() for printing the column names of the variable given the column number
> outcol = function(j){
+   k = colnames(Mcdonald2[j])
+   k
+ }
>
> ## Creating a FOR loop to test Outliers for all the variables from "Mcdonald2"
> ## And giving out the number of Outliers and Column name as the output
> t = for (i in c(1:21)){
+   if (outlier2(i) > 0){
+     y = c(outlier2(i),outcol(i))
+     print(outcol(i))
+     print(outlier2(i))
+   }
+ }
[1] "Calories"
[1] 6
[1] "Calories.from.Fat"
[1] 4
[1] "Total.Fat"
[1] 4
[1] "Total.Fat....Daily.Value."
[1] 4
[1] "Trans.Fat"
[1] 56
[1] "Cholesterol"
[1] 18
[1] "Cholesterol....Daily.Value."
[1] 18
[1] "Sodium"
[1] 5
[1] "Sodium....Daily.Value."
[1] 5
[1] "Carbohydrates"
[1] 17
[1] "Carbohydrates....Daily.Value."
[1] 16
[1] "Dietary.Fiber....Daily.Value."
[1] 4
[1] "Sugars"
[1] 4
[1] "Protein"
[1] 3
[1] "Vitamin.A....Daily.Value."
[1] 17
[1] "Vitamin.C....Daily.Value."
[1] 46
[1] "Calcium....Daily.Value."
[1] 2
[1] "Iron....Daily.Value."
[1] 2
>
>
> ### 3. Which variables have the Highest correlation. Plot them and find out the value ? ########
>
> cor_result = cor(Mcdonald2)
> cor_result = as.data.frame(cor_result)
>
> cor_result1 = cor_result[cor_result[] > 0.& cor_result[] < 1]
> t = max(cor_result1)
> t
[1] 0.9999286
> which(cor_result[] == t,arr.ind = TRUE,useNames = TRUE)
                       row col
Sodium....Daily.Value.  11  10
Sodium                  10  11
>
> cor(Mcdonald2$Sodium,Mcdonald2$Sodium....Daily.Value.)
[1] 0.9999286
>
> qplot(Sodium,Sodium....Daily.Value.,data = Mcdonald)
>
>
> ### 4. Which category contributes to the maximum % of Cholesterol in a diet (% daily value)? ##############
> t = table(Mcdonald$Cholesterol....Daily.Value.,Mcdonald$Category)
> Catchol = colSums(prop.table(t,2))
> Catchol
   Beef & Pork       Beverages       Breakfast  Chicken & Fish     Coffee & Tea        Desserts          Salads Smoothies & Shakes
             1               1               1               1                1               1               1                  1
  Snacks & Sides
             1
> qplot(Cholesterol....Daily.Value.,fill = Category,data = Mcdonald,geom = "density")
> rpivotTable(Mcdonald)
>
>
> cholcolnum = which(colnames(Mcdonald) == "Cholesterol....Daily.Value.")
> mcdchol = Mcdonald[order(Mcdonald$Cholesterol....Daily.Value.),]
> tailcat = tail(mcdchol$Category,9)
> tailchol = tail(mcdchol$Cholesterol....Daily.Value.,9)
> tailchol
[1]  92  93  95  99 100 185 185 192 192
> cholcattable = table(tailcat)
> cholcattable
tailcat
      Beef & Pork         Beverages         Breakfast    Chicken & Fish      Coffee & Tea          Desserts          Salads Smoothies & Shakes
                0                 0                 9                 0                 0                 0                 0                 0
    Snacks & Sides
                0
>
> ### 5.Which item contributes maximum to the Sodium intake?  ##############
> z = max(Mcdonald$Sodium)
> str(Mcdonald$Sodium)
 int [1:260] 750 770 780 860 880 960 1300 1410 1300 1420 ...
> x = which(Mcdonald$Sodium[] == z)
> Mcdonald[x,1]
[1] Chicken & Fish
Levels: Beef & Pork Beverages Breakfast Chicken & Fish Coffee & Tea Desserts Salads Smoothies & Shakes Snacks & Sides
>
>
>
> #### 6.Which 4 food items contains the most amount of Saturated Fat? ###########
> mcdor = Mcdonald[order(Mcdonald$Saturated.Fat),]
> top4fats = tail(mcdor[,2],4)
> top4fats
[1] Big Breakfast with Hotcakes (Large Biscuit) Chicken McNuggets (40 piece)              FrappÃ© Chocolate Chip (Large)
[4] McFlurry with M&Mâ€™s Candies (Medium)
260 Levels: 1% Low Fat Milk Jug Apple Slices Bacon Buffalo Ranch McChicken Bacon Cheddar McChicken ... Vanilla Shake (Small)
>
>
>
> ### plotting some graphs for marks ############
> qplot(Category,data = Mcdonald,xlab = "Category",ylab ="Frequency",fill = Category,main = "Frequency of Food Categories")
> qplot(Category,data = Mcdonald,xlab = "Category",ylab = "Frequency",fill = Category,main = "Frequenct of Food Categories",geom = 'density')
   qplot(Calories,Category,data = Mcdonald,xlab = "Category",ylab = "Frequency",fill = Category,main = "Frequenct of Food Categories",geom = 'violin')
  qplot(Category,Calories,data = Mcdonald,xlab = "Category",ylab = "Calories",fill = Category,main = "Frequency of Food Categories",geom = 'boxplot')
```