McDonald data Analysis

Read the data in R

menu = read.csv("menu.csv")

Basic Sanity check for data

Checkout the dimension for data

```
dim(menu)
[1] 260  24
```

Get character typr for data

```
str(menu)
'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 ...
```

Category, Serving Size and Item type is read as factor, rest all are numeric variables

View basic summary statistics

```
summary(menu)
```

```
               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....Daily.Value.
 Min.   :   0.0   Min.   :   0.0    Min.   :  0.000   Min.   :  0.00
 1st Qu.: 210.0   1st Qu.:  20.0    1st Qu.:  2.375   1st Qu.:  3.75
 Median : 340.0   Median : 100.0    Median : 11.000   Median : 17.00
 Mean   : 368.3   Mean   : 127.1    Mean   : 14.165   Mean   : 21.82
 3rd Qu.: 500.0   3rd Qu.: 200.0    3rd Qu.: 22.250   3rd Qu.: 35.00
 Max.   :1880.0   Max.   :1060.0    Max.   :118.000   Max.   :182.00

 Saturated.Fat    Saturated.Fat....Daily.Value.   Trans.Fat       Cholesterol
 Min.   : 0.000   Min.   :  0.00                 Min.   :0.0000   Min.   :
0.00
 1st Qu.: 1.000   1st Qu.:  4.75                 1st Qu.:0.0000   1st Qu.:
5.00
 Median : 5.000   Median : 24.00                 Median :0.0000   Median :
35.00
 Mean   : 6.008   Mean   : 29.97                 Mean   :0.2038   Mean   :
54.94
 3rd Qu.:10.000   3rd Qu.: 48.00                 3rd Qu.:0.0000   3rd Qu.:
65.00
 Max.   :20.000   Max.   :102.00                 Max.   :2.5000   Max.   :
575.00

 Cholesterol....Daily.Value.     Sodium       Sodium....Daily.Value.
Carbohydrates
 Min.   :  0.00              Min.   :   0.0   Min.   :  0.00         Min.   :
0.00
 1st Qu.:  2.00              1st Qu.: 107.5   1st Qu.:  4.75         1st Qu.:
30.00
 Median : 11.00              Median : 190.0   Median :  8.00         Median :
44.00
 Mean   : 18.39              Mean   : 495.8   Mean   : 20.68         Mean   :
47.35
 3rd Qu.: 21.25              3rd Qu.: 865.0   3rd Qu.: 36.25         3rd Qu.:
60.00
 Max.   :192.00              Max.   :3600.0   Max.   :150.00         Max.   :
141.00

 Carbohydrates....Daily.Value. Dietary.Fiber   Dietary.Fiber....Daily.Value.
Sugars
```

```
 Min.   : 0.00                 Min.   :0.000   Min.   : 0.000
Min.   :  0.00
 1st Qu.:10.00                 1st Qu.:0.000   1st Qu.: 0.000
1st Qu.:  5.75
 Median :15.00                 Median :1.000   Median : 5.000
Median : 17.50
 Mean   :15.78                 Mean   :1.631   Mean   : 6.531
Mean   : 29.42
 3rd Qu.:20.00                 3rd Qu.:3.000   3rd Qu.:10.000
3rd Qu.: 48.00
 Max.   :47.00                 Max.   :7.000   Max.   :28.000
Max.   :128.00

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

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

Observations:
No missing values seems to be there in data set
All factor in Item type are unique
Numerical variables might have outliers

Check first and last few records to ensure all variables are in proper format.

```
head(menu)
   Category                                  Item   Serving.Size Calories
Calories.from.Fat
1 Breakfast                      Egg McMuffin 4.8 oz (136 g)      300
120
2 Breakfast                Egg White Delight 4.8 oz (135 g)      250
70
3 Breakfast                 Sausage McMuffin 3.9 oz (111 g)      370
200
4 Breakfast        Sausage McMuffin with Egg 5.7 oz (161 g)      450
250
```

```
5 Breakfast Sausage McMuffin with Egg Whites 5.7 oz (161 g)      400
210
6 Breakfast             Steak & Egg McMuffin 6.5 oz (185 g)      430
210
  Total.Fat Total.Fat....Daily.Value. Saturated.Fat
Saturated.Fat....Daily.Value. Trans.Fat
1        13                          20             5
25         0
2         8                          12             3
15         0
3        23                          35             8
42         0
4        28                          43            10
52         0
5        23                          35             8
42         0
6        23                          36             9
46         1
  Cholesterol Cholesterol....Daily.Value. Sodium Sodium....Daily.Value.
Carbohydrates
1         260                           87    750                       31
31
2          25                            8    770                       32
30
3          45                           15    780                       33
29
4         285                           95    860                       36
30
5          50                           16    880                       37
30
6         300                          100    960                       40
31
  Carbohydrates....Daily.Value. Dietary.Fiber Dietary.Fiber....Daily.Value.
Sugars Protein
1                            10             4                            17
3      17
2                            10             4                            17
3      18
3                            10             4                            17
2      14
4                            10             4                            17
2      21
5                            10             4                            17
2      21
6                            10             4                            18
3      26
  Vitamin.A....Daily.Value. Vitamin.C....Daily.Value. Calcium....Daily.Value.
1                        10                         0                        25
2                         6                         0                        25
3                         8                         0                        25
4                        15                         0                        30
5                         6                         0                        25
6                        15                         2                        30
  Iron....Daily.Value.
1                   15
2                    8
3                   10
```

```
4                    15
5                    10
6                    20
```

```
tail(menu)
```

```
              Category                                                    Item
Serving.Size
255 Smoothies & Shakes             McFlurry with M&Mâ€™s Candies (Snack)  7.3
oz (207 g)
256 Smoothies & Shakes                McFlurry with Oreo Cookies (Small) 10.1
oz (285 g)
257 Smoothies & Shakes               McFlurry with Oreo Cookies (Medium) 13.4
oz (381 g)
258 Smoothies & Shakes                McFlurry with Oreo Cookies (Snack)  6.7
oz (190 g)
259 Smoothies & Shakes McFlurry with Reese's Peanut Butter Cups (Medium) 14.2
oz (403 g)
260 Smoothies & Shakes  McFlurry with Reese's Peanut Butter Cups (Snack)  7.1
oz (202 g)
    Calories Calories.from.Fat Total.Fat Total.Fat....Daily.Value.
Saturated.Fat
255      430               140        15                         24
10
256      510               150        17                         26
9
257      690               200        23                         35
12
258      340               100        11                         17
6
259      810               290        32                         50
15
260      410               150        16                         25
8
    Saturated.Fat....Daily.Value. Trans.Fat Cholesterol
Cholesterol....Daily.Value. Sodium
255                            48       0.0          35
11    120
256                            44       0.5          45
14    280
257                            58       1.0          55
19    380
258                            29       0.0          30
9    190
259                            76       1.0          60
20    400
260                            38       0.0          30
10    200
    Sodium....Daily.Value. Carbohydrates Carbohydrates....Daily.Value.
Dietary.Fiber
255                      5            64                           21
1
256                     12            80                           27
1
257                     16           106                           35
1
258                      8            53                           18
1
```

```
259                      17           114                             38
2
260                       8            57                             19
1
    Dietary.Fiber....Daily.Value. Sugars Protein Vitamin.A....Daily.Value.
255                             4     59       9                         10
256                             4     64      12                         15
257                             5     85      15                         20
258                             2     43       8                         10
259                             9    103      21                         20
260                             5     51      10                         10
    Vitamin.C....Daily.Value. Calcium....Daily.Value. Iron....Daily.Value.
255                         0                      30                    4
256                         0                      40                    8
257                         0                      50                   10
258                         0                      25                    6
259                         0                      60                    6
260                         0                      30                    4
```

Data looks in proper format with no custom headers or footers

Check for missing values

```
anyNA(menu)
[1] FALSE
> sapply(menu, function(x) sum(is.na(x)))
                    Category                         Item
Serving.Size
                           0                            0
0
                    Calories             Calories.from.Fat
Total.Fat
                           0                            0
0
    Total.Fat....Daily.Value.                 Saturated.Fat
Saturated.Fat....Daily.Value.
                           0                            0
0
                   Trans.Fat                   Cholesterol
Cholesterol....Daily.Value.
                           0                            0
0
                      Sodium        Sodium....Daily.Value.
Carbohydrates
                           0                            0
0
Carbohydrates....Daily.Value.                 Dietary.Fiber
Dietary.Fiber....Daily.Value.
                           0                            0
0
                      Sugars                       Protein
Vitamin.A....Daily.Value.
                           0                            0
0
```

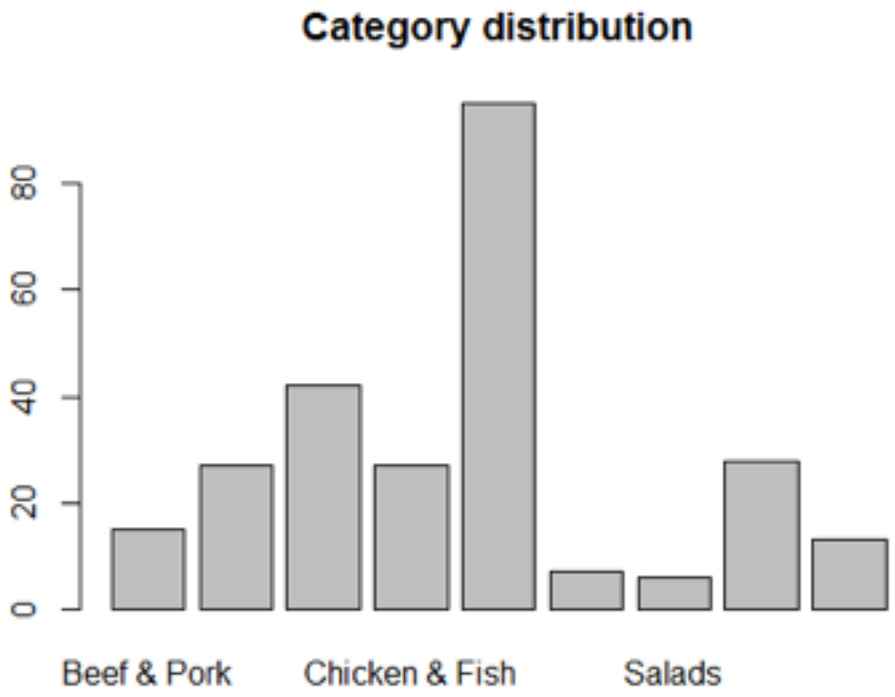
```
    Vitamin.C....Daily.Value.        Calcium....Daily.Value.
Iron....Daily.Value.
                            0                              0
0
```

This confirms that no missing values are present in data set

Exploratory Analysis

Category:

barplot(table(menu$Category), main = "Category distribution")

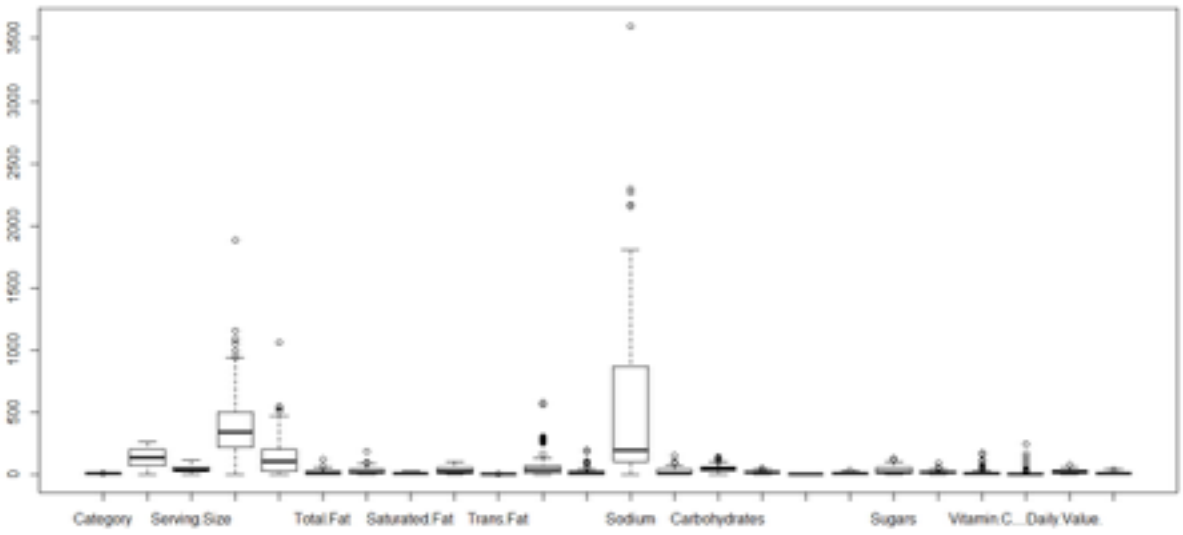



```
table(menu$Category)
```

```
       Beef & Pork          Beverages          Breakfast     Chicken & Fish
                15                 27                 42                 27
      Coffee & Tea           Desserts             Salads Smoothies & Shakes
                95                  7                  6                 28
    Snacks & Sides
                13
```

Coffee & Tea looks to be most popular while Salads seems to have least varieties
#Check for outliers in numeric variables

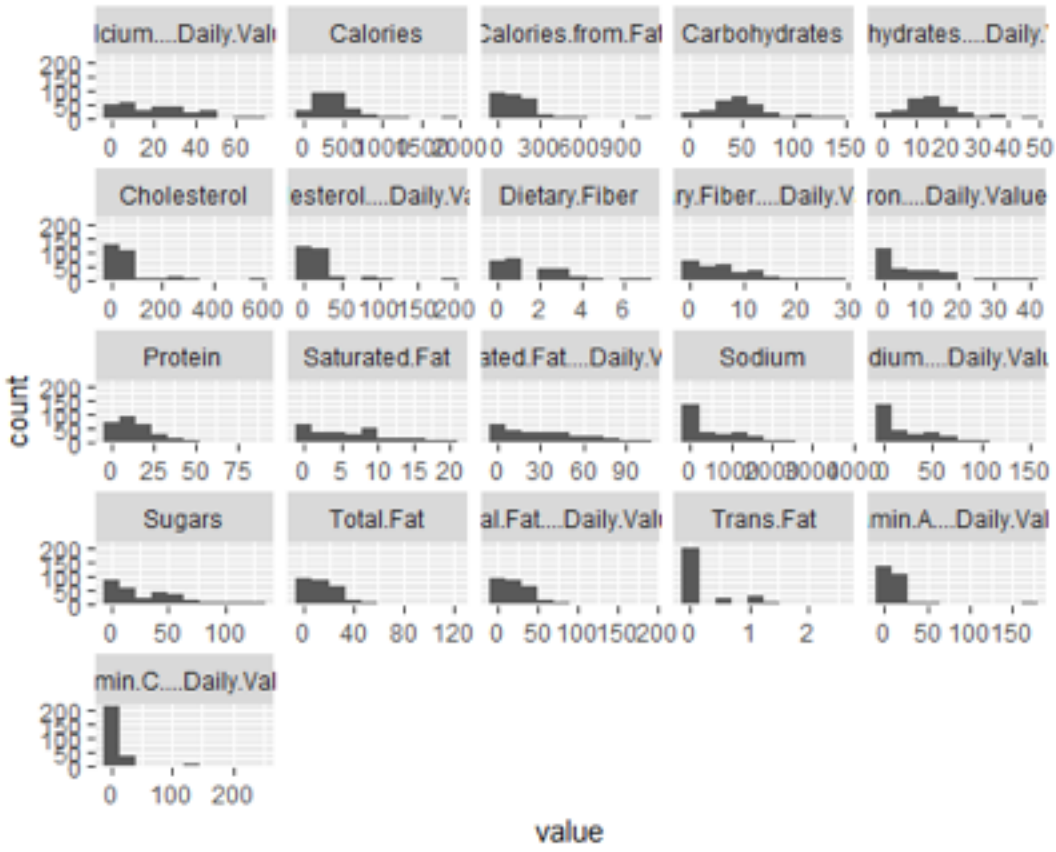
```
boxplot(menu)
```



We could see that outliers are there in most of variables.

Check for distribution of numerical variables

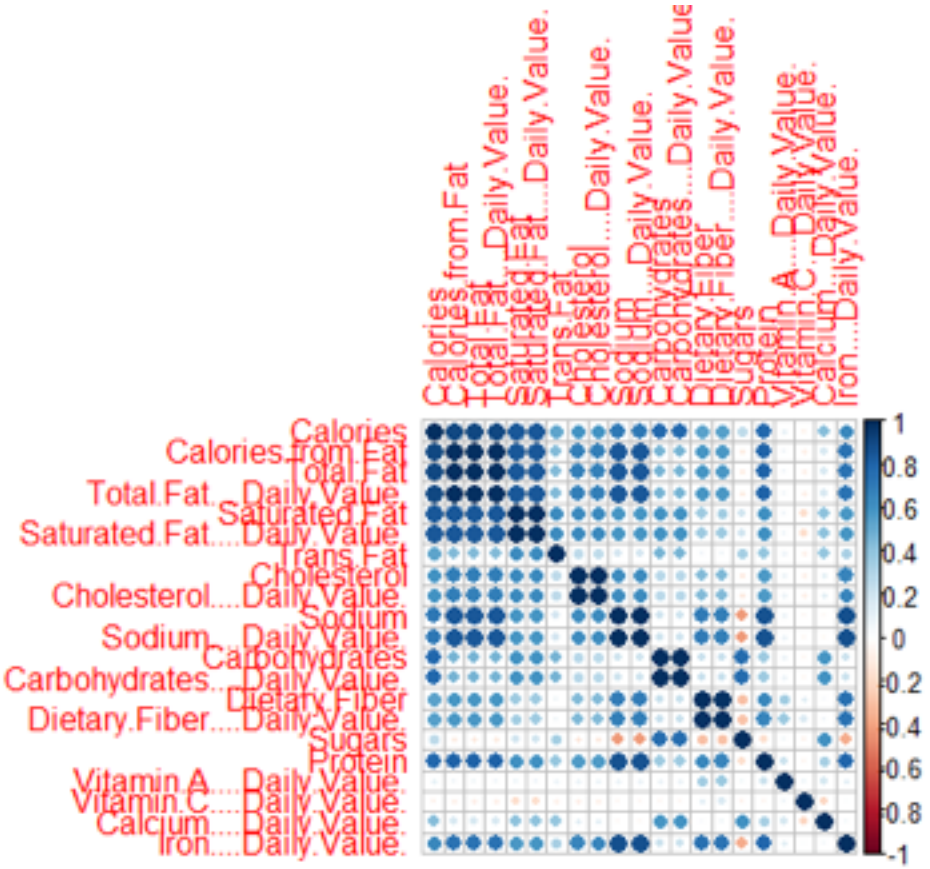
```
ggplot(gather(menu[,-1:-3]), aes(value)) +
+   geom_histogram(bins = 10) +
+   facet_wrap(~key, scales = 'free_x')
```



Carbohydrates variables looks normally distributed
Calcium, fiber, iron has good spread
Other variables show skewness

Check for correlation among numeric variables.

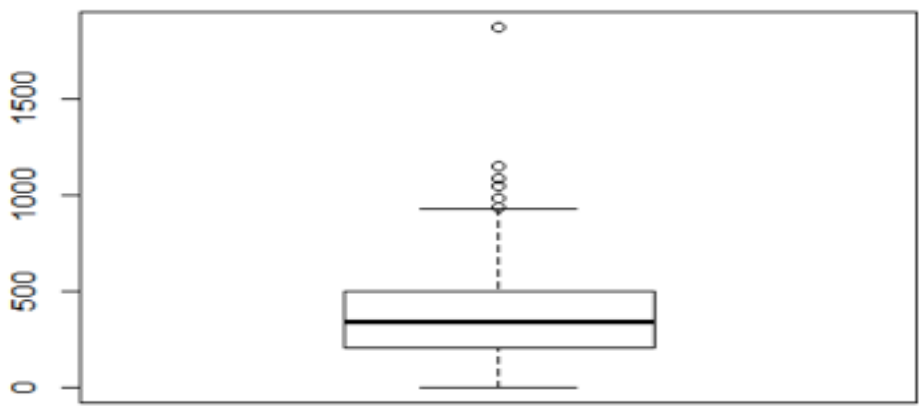
```
library(corrplot)
corrplot(cor(menu[,4:24]))
```



All fat variables show high correlation
Apart from variables of total values and daily value variables we can see strong correlation of proteins with Fat, sodium, Carbohydrates, fiber and iron
Similarly, iron shows strong correlation with above variables

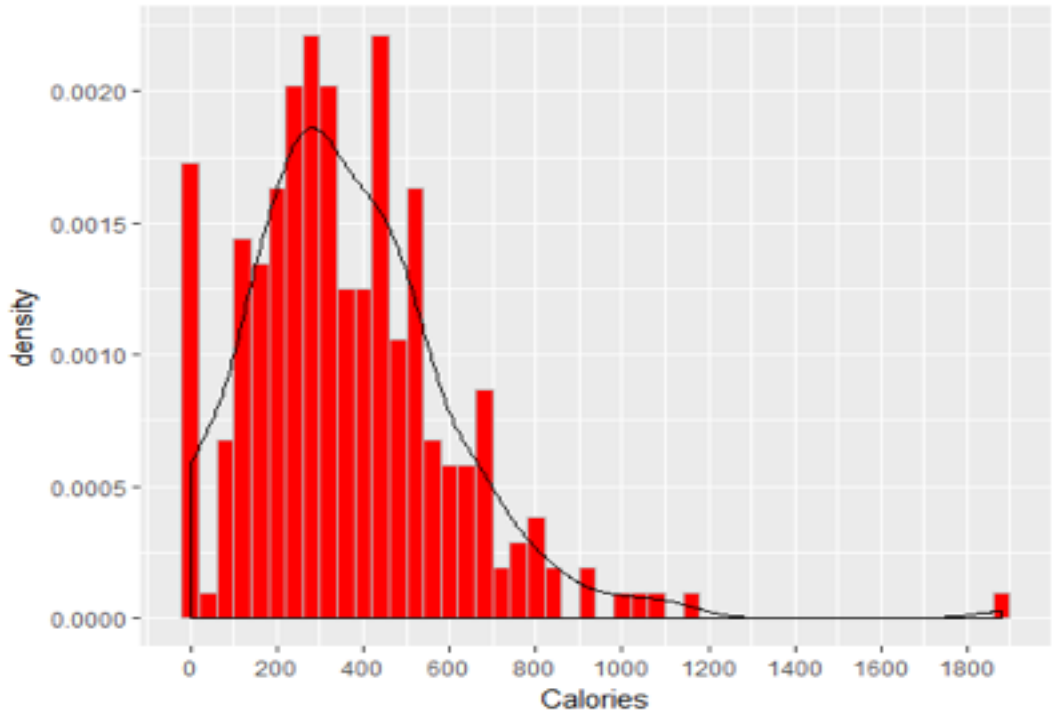
Calories

```
boxplot(menu$Calories)
```



Outliers are present

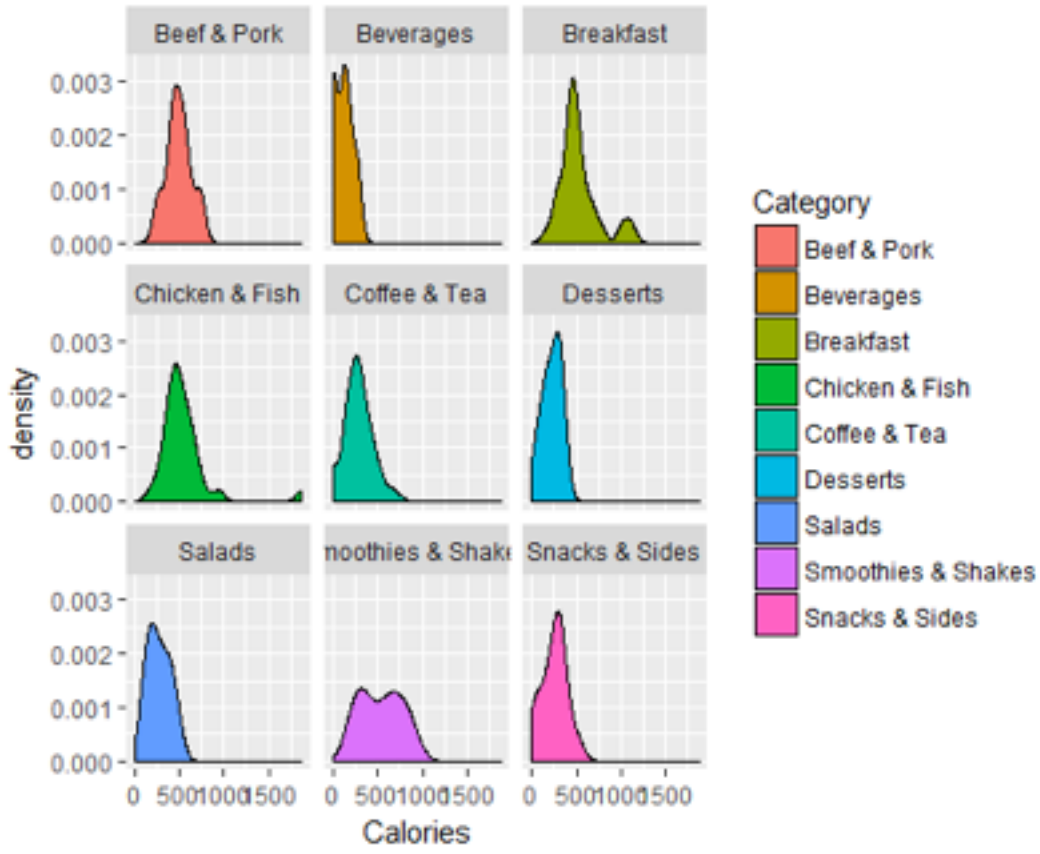
```
ggplot(menu, aes(x = Calories)) +
  geom_histogram(aes(y = ..density..), fill = "red", binwidth = 40, color="gray") +
  geom_density() +
  scale_x_continuous(breaks = seq(min(menu$Calories), max(menu$Calories), by = 200))
```



Most of items have calories of around 200-350
Outlier present in far end with calorie value of 1800

Let us check the calorie distribution by category

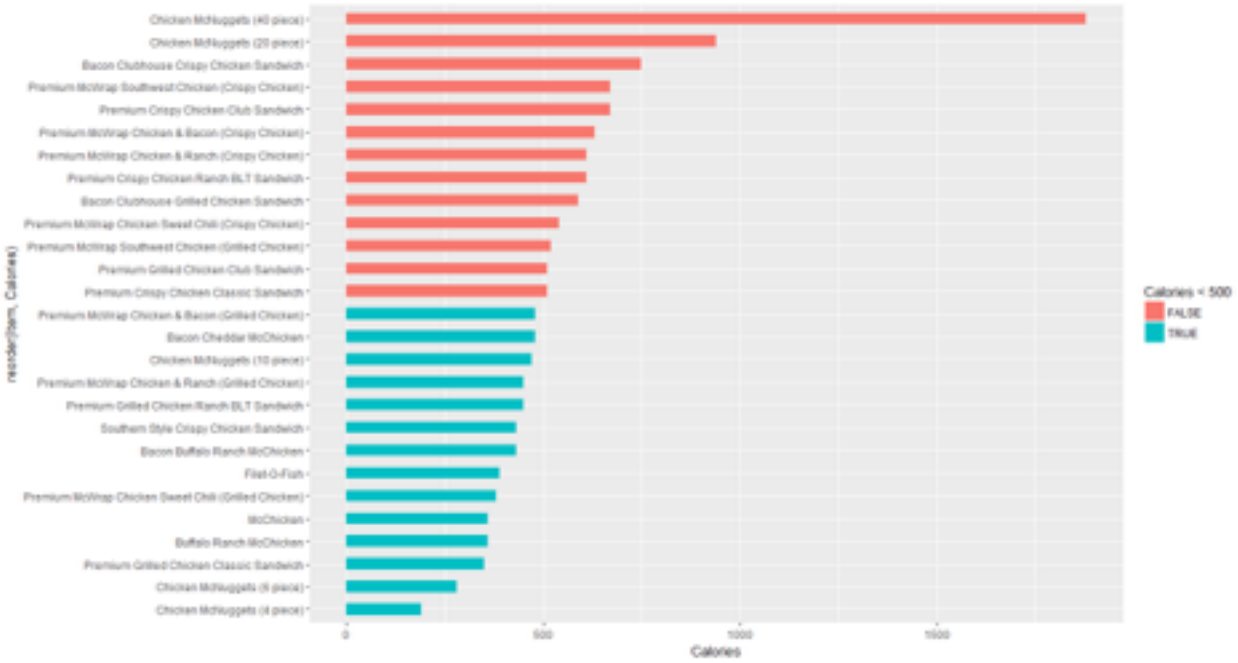
```
ggplot(menu, aes(x = Calories,fill=Category)) +
  geom_density() + facet_wrap( ~ Category)
```



Outlier seen in previous plot seems to have come from Chicken & Fish category
Apart from that Breakfast and Smoothies & Shakes have higher calorie on an average.

Check the distribution of calorie content in Chicken & Fish category

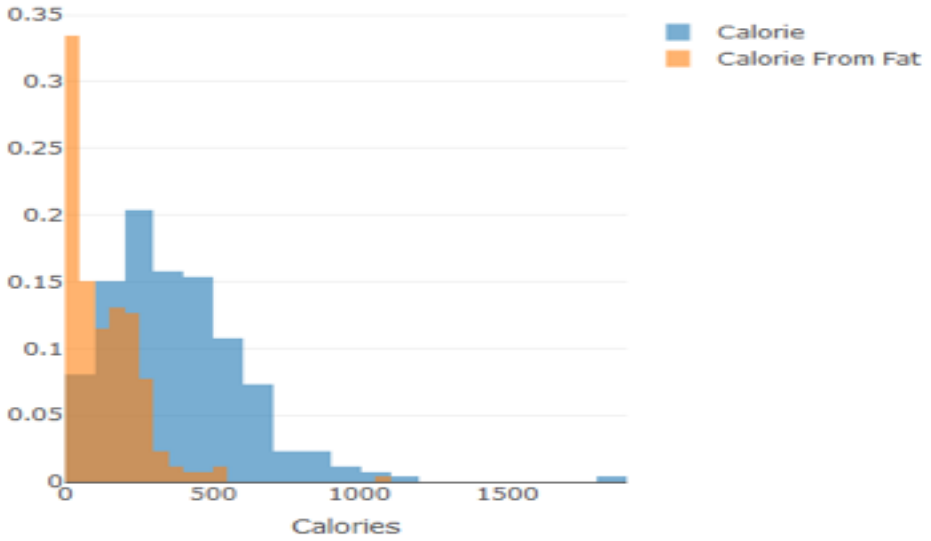
```
library(dplyr)
menu %>%
filter(.,Category=="Chicken & Fish") %>%
    ggplot(aes(x = reorder(Item, Calories), y = Calories)) +
    geom_bar(aes(fill=Calories<500), width=0.5, stat = "identity") + coord_flip()
```



1800 calorie value is from 40pcs of chicken, hence it is not an outlier as was concluded earlier

Let us now check the calories from fat as percentage of total calories

```
plot_ly(menu,x = ~Calories,
        type = "histogram",
        histnorm = "probability",name="Calorie",alpha=0.6)%>%
 add_histogram(x = ~Calories.from.Fat,name="Calorie From Fat",alpha=0.6) %>%
 layout(barmode = "overlay")
```



Check variables which have more than desired amount of nutrients value

```
menu[menu$Cholesterol....Daily.Value. > 100, cbind("Category", "Item",
"Cholesterol....Daily.Value.")]
    Category                                                Item
Cholesterol....Daily.Value.
28 Breakfast                Big Breakfast (Regular Biscuit)
185
29 Breakfast                  Big Breakfast (Large Biscuit)
185
32 Breakfast Big Breakfast with Hotcakes (Regular Biscuit)
192
33 Breakfast   Big Breakfast with Hotcakes (Large Biscuit)
192
```

We could see that above 4 items are not healthy as they contain almost double the amount of
cholesterol required daily. Expect it to be for single person

```
menu[menu$Total.Fat....Daily.Value. > 100, cbind("Category","Item",
"Total.Fat....Daily.Value.")]
         Category                          Item Total.Fat....Daily.Value.
83 Chicken & Fish Chicken McNuggets (40 piece)                        182
```

We will ignore this as it talks about 40 pieces

```
menu[menu$Saturated.Fat....Daily.Value. > 100, cbind("Category", "Item",
"Saturated.Fat....Daily.Value.")]
              Category                                      Item
Saturated.Fat....Daily.Value.
83      Chicken & Fish           Chicken McNuggets (40 piece)
101
232       Coffee & Tea         FrappÃ© Chocolate Chip (Large)
101
254 Smoothies & Shakes McFlurry with M&Mâ€™s Candies (Medium)
102

menu[menu$Vitamin.A....Daily.Value. > 100, cbind("Category", "Item",
"Vitamin.A....Daily.Value.")]
   Category                                             Item
Vitamin.A....Daily.Value.
85   Salads    Premium Bacon Ranch Salad (without Chicken)
170
87   Salads Premium Bacon Ranch Salad with Grilled Chicken
110
88   Salads      Premium Southwest Salad (without Chicken)
160
89   Salads    Premium Southwest Salad with Crispy Chicken
170
90   Salads   Premium Southwest Salad with Grilled Chicken
170

I would rather have this, provided it does not form a daily diet.
This can be classified as healthy food.


menu[menu$Vitamin.C....Daily.Value. > 100, cbind("Category", "Item",
"Vitamin.C....Daily.Value.")]
          Category                                        Item
Vitamin.C....Daily.Value.
41       Breakfast                     Fruit & Maple Oatmeal
130
42       Breakfast Fruit & Maple Oatmeal without Brown Sugar
130
102 Snacks & Sides                                 Apple Slices
160
134      Beverages          Minute Maid Orange Juice (Small)
130
135      Beverages         Minute Maid Orange Juice (Medium)
160
```