Eating Habits of People in America

Level: 6000

1. Abstract and Introduction

The traditional eating patterns that many in the United States are actually following do not agree wit h the Dietary Guidelines. A comparison is drawn in Figure1. About three-fourths of the population has a low vegetable, fruit, dairy and oils eating pattern. More than half of the population meets or exceeds the guidelines for total grain and protein foods, but within each of these food groups they do not follow the recommendations for the subgroups. Most Americans go beyond the recommendations for added sugars, saturated fats, and sodium.

Furthermore, many eating patterns are too high in calories. Compared to calorie needs, calorie intake over time is best assessed by measuring body weight status. The high percentage of the overweight or obese population means that many in the U.S. are over-consuming calories.

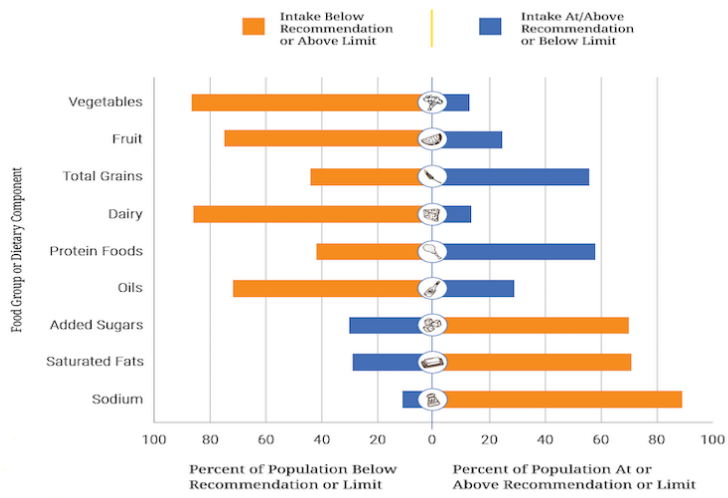

Figure1: Dietary Intakes Compared to Recommendations

It's no secret that the amount of calories people eat and drink has a direct impact on their weight: Consume the same number of calories that the body burns over time, and weight stays stable. Consume more than the body burns, weight goes up. Less, weight goes down. Obesity is becoming a worldwide problem affecting all levels of the society and is thus being globally described as an epidemic. The rapid increase in obesity among the world's population has become a major public health problem, affecting both developed and developing countries. The obesity epidemic results in a substantial decrease in the quality of life and life expectancy, and it accounts for heavy expenditure in provision of health care. Prevention of childhood obesity has been recognized as a public health priority due to difficulty in the treatment of obesity in adults and the many long-term adverse effects of childhood obesity. The development of obesity involves multiple factors, such as improper food consumption, sedentary behaviour, patterns of physical activity, social and environmental variables, and individual susceptibility; determined by unmodified factors such as genetic and biological factors.

In this project, my aim is to analyse the diets of people in America and compare them with those of the obese people. I will use amount of calories as a basis of comparison. I will also analyse the relation of calorie value with other nutritive values of diets of people in America.

2. Data Description

My analysis consist of two datasets, the first dataset contains the commonly eaten food items by people in America(Food_table). The second dataset is a weekly analysis of eating habits of Obese people(ObeseData).
The first dataset consists of commonly eaten food items with attributes such as solid fats, added sugars, calories, saturated fats, oils, alcohol content, meats, etc.

	Milk	Meats	Soy	Drybeans_Peas	Oils	Solid_Fats	Added_Sugars	Alcohol	Calories	Saturated_Fats
0	0	0	0	0	0	105.6485	1.57001	0	133.65	7.36898
0	0.29393	0	0	0	0	130.99968	95.20488	0	267.33	9.0307
0	0.2516	0.0962	0	0	0	213.06672	96.1034	0	368.52	15.2884
0	0.38233	0	0	0	0	170.39808	123.83793	0	347.73	11.7467

Figure2: Dataset of commonly eaten food items in America

The second dataset consist of categories of food eaten by Obese people on a weekly basis. Some of the food categories of this dataset include meat/fish, no color vegetables, fruits, grains, sweets etc.

food_type	times_per_week	number_of_males	number_of_females	Male_percentage	Female_percentage	Total_percentage	Total_number
Meat/fish	1	8	16	17.8	24.2	21.6	24
Meat/fish	4	25	36	55.6	54.5	55	61
Meat/fish	12	10	11	22.2	16.7	18.9	21
Meat/fish	21	2	3	4.4	4.5	4.5	5
beans/tofu	1	13	10	28.9	15.2	13.5	23

Figure3: Dataset of eating trends in Obese people

Source for the first dataset of food items:
https://data.world/us-usda-gov/27830bd2-53c4-4d7b-9686-eca1a695d92a
Source of the second dataset of eating habits of Obese people:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877769/

I selected the above to datasets so as to complement my analysis on the eating habits of people in America. The two detests provided me a basis for comparing the diets of non-obese people with those of the obese ones on the basis of calories consumed. The attribute of food_category provided me a basis of comparison between the two datasets. The first dataset was taken from the US government health data websites and the second dataset was manually prepared from a survey on Obese people(source for the details about the survey is given above).

My dataset was very clean apart from some missing values which I removed. I added a column of food category and categorised each of the food items in my first dataset so as to provide myself a basis of comparison between the two datasets.

3. Analysis

In analysing my datasets, I looked for various missing values and anomalies in my data and removed the data values with missing values and NA's. My dataset was relatively very clean and did not required much of cleaning. I performed Exploratory Data Analysis on my dataset and looked for distributions of various attributes in both of my datasets.

My dataset of eating habits of Obese people did not contain food items categorised according to their food category. I added a column of food category to my dataset so as to provide a basis of comparison among both the datasets.

a. Data Manipulation:
The dataset of food items were not categorised according to the food categories, so a column of food type was added to the dataset to provide a basis of comparison between the two datasets.

b. Data Cleaning:
The dataset of food items contained some missing values and NA's which were removed from the dataset.
> df.dropna()
> df.dropna(subset = ["colname"])
The data was normalised before performing computations so as to make different attributes lie in the same range.
> nor <-function(x) {(x-min(x))/(max(x)-min(x))}

c. Inspect data:
Several methods are used to inspect the dataset.
A. Get an overview of the dataframe (df):
head(Food_table)
tail(Food_table)
> summary(Food_table)

```
  Food_Code                    food_type                                       Display_Name
 Min.   :     7258   Snacks              :606   Cheese pizza, thick crust         :   7
 1st Qu.:27214100   Meat/fish           :496   French fries, deep-fried          :   7
 Median :54403090   Sweets              :220   Raw tomatoes                      :   7
 Mean   :52961704   milk/milk products:190   Chocolate-covered candy           :   6
 3rd Qu.:72901282   Fruits              :148   Dietetic chocolate-covered candy:   6
 Max.   :94210100   Grains              : 84   Ground beef (75% lean, regular) :   6
                    (Other)             :270   (Other)                           :1975
```

> dim(food_table)
[1] 2014   27

d. Import data:
R language (R)
A. Food Items dataset
Food_table<-read.csv(file.choose(),header = TRUE)
Food_table<-data.frame(food_table)

B. Obesity dataset

```
ObeseData<-read.csv(file.choose(),header = TRUE)
ObeseData<-data.frame(ObeseData)
```

I ordered my dataset of food items eaten by non-obese people in the increasing order of Calories and observed the variation of different nutritive quantities of food items.

```
food_table<-food_table[order(food_table$Saturated_Fats), ]
plot(food_table$Saturated_Fats)
plot(food_table$Solid_Fats)
```

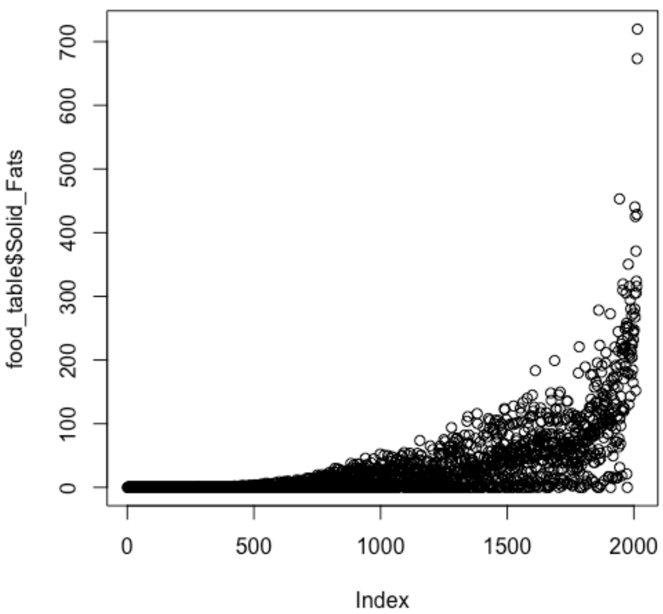

Figure4: Scatterplot for Saturated fats

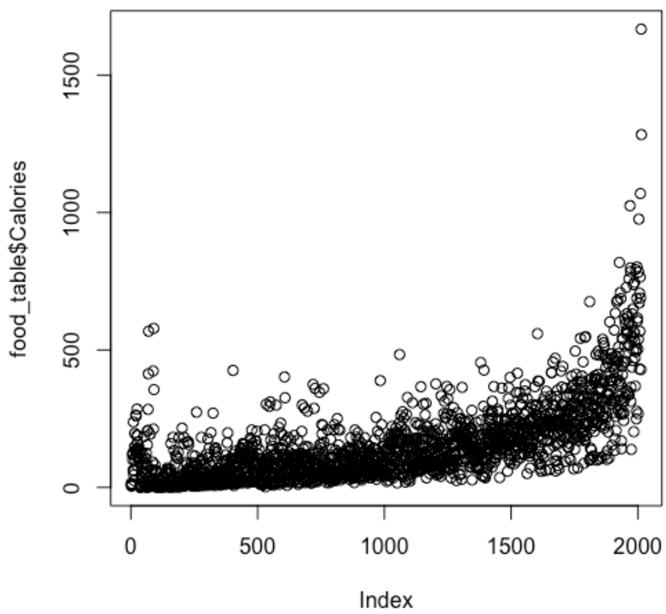

Figure5: Scatterplot for calories

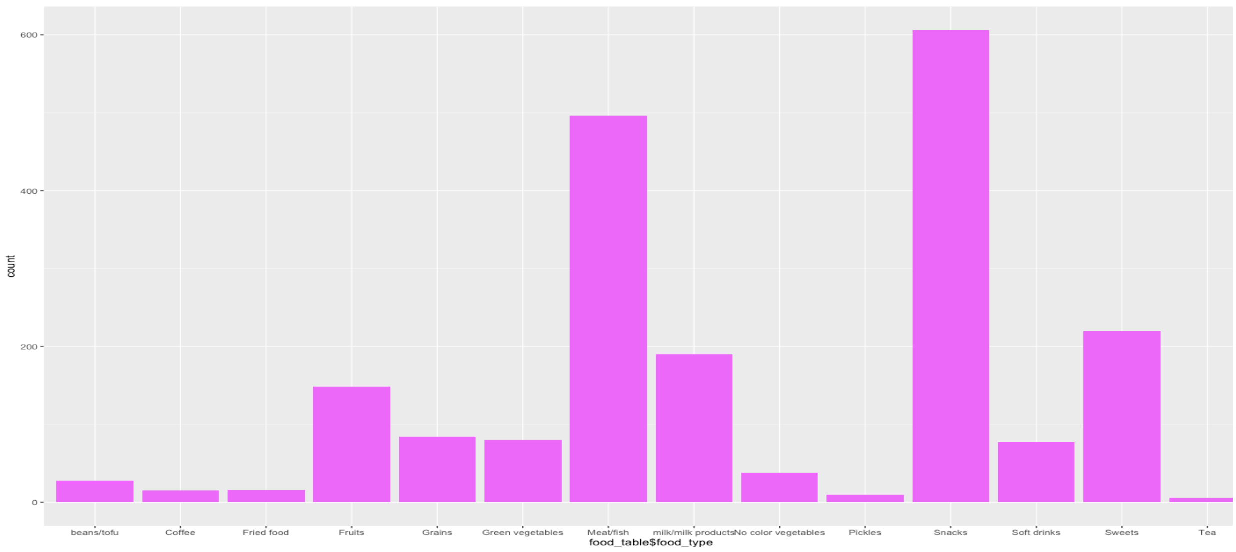

Figure6: Frequency of various categories of food items eaten

```
ggplot(food_table) +
  geom_bar(mapping = aes(x = food_table$food_type), fill = "magenta")

boxplot(food_table_norm$Saturated_Fats,food_table_norm$Alcohol,  food_table_norm$Solid_Fats,
food_table_norm$Added_Sugars,food_table_norm$Calories, col = "Yellow")
```

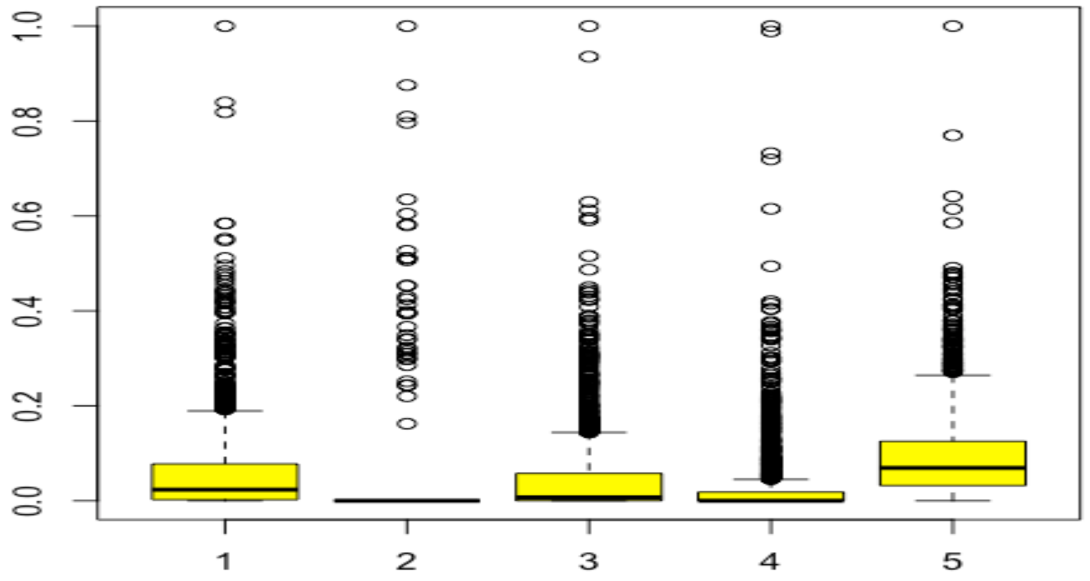

Figure7: Distribution of different attributes in the food items dataset

```
data <- as.matrix(food_table_norm)
heatmap(data, col = terrain.colors(256))
```

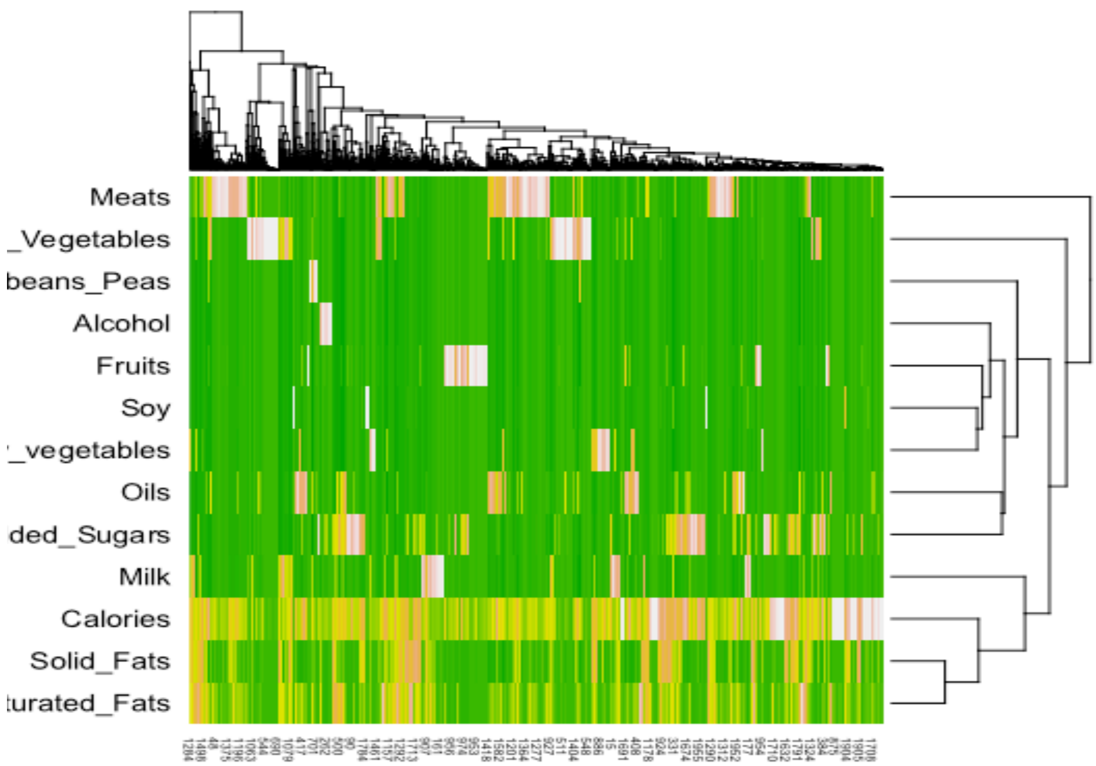

Figure8: Heatmap for different attributes in the food items dataset

I also performed correlation tests between the response variables and predictor variables to see the relation between different attributes in my data.

```
> cor(food_table$Milk, food_table$Calories)
[1] 0.3707003
> cor(food_table$Meats, food_table$Calories)
[1] 0.4310305
> cor(food_table$Soy, food_table$Calories)
[1] -0.0233141
> cor(food_table$Drybeans_Peas, food_table$Calories)
[1] 0.0986605
> cor(food_table$Oils, food_table$Calories)
[1] 0.2786845
> cor(food_table$Solid_Fats, food_table$Calories)
[1] 0.7456695
> cor(food_table$Added_Sugars, food_table$Calories)
[1] 0.2437717
> cor(food_table$Alcohol, food_table$Calories)
[1] 0.04432072
> cor(food_table$Saturated_Fats, food_table$Calories)
[1] 0.7974049
```

4. Model Development and Application of model(s)

The two motives of this project of mine are as follows:
(a) To compare the calorie intake of non-obese people with those of the obese ones.
(b) To predict calorie value of different food items based on other nutritive qualities of food.

I performed analytics on both of datasets and recorded the calorie intake by both the obese and the non-obese people. Secondly, I applied models of Linear Regression and KKNN to predict calorie value of different food items.

I chose the model of KKNN, which is a classification algorithm, so as to group my food items into categories of "Low calorie food item", "High Calorie food item" and "Optimum Calorie food item".

I applied the Linear Regression model on my dataset because of the high correlation between my response variables and predictor variables and predicted the calorie value of different food items with the model.

(a) To compare the calorie intake of non-obese people with those of the obese ones.

Code

Calorie intake by non-obese people

```
#calories consumed from different categories of food that are consumed by Americans in general
n<-nrow(food_table)
```

```
d<-ncol(food_table)

#average calories consumed from Meat/fish category
meat_fish = subset(food_table, food_table$food_type == "Meat/fish" )
avg_calories_meat_fish = sum(meat_fish$Calories)/nrow(meat_fish)
avg_calories_meat_fish

#average calories consumed from beans/tofu category
beans_tofu = subset(food_table, food_table$food_type == "beans/tofu" )
avg_calories_beans_tofu = sum(beans_tofu$Calories)/nrow(beans_tofu)
avg_calories_beans_tofu


#average calories consumed from milk/milk products category
milk_milkproducts = subset(food_table, food_table$food_type == "milk/milk products" )
avg_calories_milk_milkproducts = sum(milk_milkproducts$Calories)/nrow(milk_milkproducts)
avg_calories_milk_milkproducts

#average calories consumed from Green vegetables category
greenveg = subset(food_table, food_table$food_type == "Green vegetables" )
avg_calories_greenveg = sum(greenveg$Calories)/nrow(greenveg)
avg_calories_greenveg

#average calories consumed from No color vegetables category
nocolor_vegetables = subset(food_table, food_table$food_type == "No color vegetables" )
avg_calories_nocolor_vegetables = sum(nocolor_vegetables$Calories)/nrow(nocolor_vegetables)
avg_calories_nocolor_vegetables

#average calories consumed from Fruits category
Fruits = subset(food_table, food_table$food_type == "Fruits" )
avg_calories_Fruits = sum(Fruits$Calories)/nrow(Fruits)
avg_calories_Fruits

#average calories consumed from Grains category
Grains = subset(food_table, food_table$food_type == "Grains" )
avg_calories_Grains = sum(Grains$Calories)/nrow(Grains)
avg_calories_Grains

#average calories consumed from Sweets category
Sweets = subset(food_table, food_table$food_type == "Sweets" )
avg_calories_Sweets = sum(Sweets$Calories)/nrow(Sweets)
avg_calories_Sweets

#average calories consumed from Coffee category
Coffee = subset(food_table, food_table$food_type == "Coffee" )
avg_calories_Coffee = sum(Coffee$Calories)/nrow(Coffee)
avg_calories_Coffee
```

```
#average calories consumed from Tea category
Tea = subset(food_table, food_table$food_type == "Tea" )
avg_calories_Tea = sum(Tea$Calories)/nrow(Tea)
avg_calories_Tea

#average calories consumed from Snacks category
Snacks = subset(food_table, food_table$food_type == "Snacks" )
avg_calories_Snacks = sum(Snacks$Calories)/nrow(Snacks)
avg_calories_Snacks

#average calories consumed from Soft_drinks category
Soft_drinks = subset(food_table, food_table$food_type == "Soft drinks" )
avg_calories_Soft_drinks = sum(Soft_drinks$Calories)/nrow(Soft_drinks)
avg_calories_Soft_drinks

#average calories consumed from Pickles category
Pickles = subset(food_table, food_table$food_type == "Pickles" )
avg_calories_Pickles = sum(Pickles$Calories)/nrow(Pickles)
avg_calories_Pickles

#average calories consumed from Fried food category
Fried_food = subset(food_table, food_table$food_type == "Fried food" )
avg_calories_Fried_food = sum(Fried_food$Calories)/nrow(Fried_food)
avg_calories_Fried_food

#total amount of fats consumned from different food items
totalamt = food_table$Portion_Amount*food_table$Saturated_Fats
sumtotalamt = sum(totalamt)
fat_per_item = sumtotalamt/2000
fat_per_item

#total amount of calories consumned from different food items
totalcalories = food_table$Portion_Amount*food_table$Calories
sum_total_calories = sum(totalcalories)
calorie_per_item = sum_total_calories/2000
calorie_per_item

#total amount of sugars consumned from different food items
totalsugars = food_table$Portion_Amount*food_table$Added_Sugars
sum_total_sugars = sum(totalsugars)
sugar_per_item = sum_total_sugars/2000
sugar_per_item

#total amount vegetables consumed from different food items
totalveg = food_table$Portion_Amount*food_table$Vegetables
sum_total_vegetables = sum(totalveg)
veg_per_item = sum_total_vegetables/2000
veg_per_item
```

```
#total amount alcohol consumed from different food items
totalalcohol = food_table$Portion_Amount*food_table$Alcohol
sum_total_alcohol = sum(totalalcohol)
alc_per_item = sum_total_alcohol/2000
alc_per_item

plot(totalamt)
```

Food category	Average amount of calories per item in the food category
Meat/fish	212.8964
beans/tofu	150.7573
milk/milk products	125.291
Green vegetables	59.86665
No color vegetables	64.30158
Fruits	70.79822
Grains	109.4999
Sweets	136.2457
Coffee	77.55933
Tea	39.20167
Snacks	170.3584
Soft drinks	158.2419
Pickles	21.32025
Fried food	210.2632

Figure9: Calorie estimates of commonly eaten food items in America according to the food categories

Calorie intake by obese people

Code:

```
ObeseData<-read.csv(file.choose(), header = TRUE)
ObeseData<-data.frame(ObeseData)

average_foodtype_intake = ObeseData$times_per_week*ObeseData$Total_number
average_foodtype_intake

average_times_meat_fish_intake = sum(average_foodtype_intake[0:4])/111
average_times_meat_fish_intake
average_times_beans_tofu_intake = sum(average_foodtype_intake[4:8])/111
```

```
average_times_beans_tofu_intake
average_times_milk_intake = sum(average_foodtype_intake[8:12])/111
average_times_milk_intake
average_times_greenveg_intake  = sum(average_foodtype_intake[12:16])/111
average_times_greenveg_intake
average_times_nocolorveg_intake = sum(average_foodtype_intake[16:20])/111
average_times_nocolorveg_intake
average_times_fruits_intake = sum(average_foodtype_intake[20:24])/111
average_times_fruits_intake
average_times_grains_intake = sum(average_foodtype_intake[24:28])/111
average_times_grains_intake
average_times_sweets_intake = sum(average_foodtype_intake[28:30])/138
average_times_sweets_intake
average_times_coffee_intake = sum(average_foodtype_intake[30:32])/138
average_times_coffee_intake
average_times_tea_intake = sum(average_foodtype_intake[32:34])/138
average_times_tea_intake
average_times_snacks_intake = sum(average_foodtype_intake[34:36])/138
average_times_snacks_intake
average_times_softdrinks_intake = sum(average_foodtype_intake[36:38])/138
average_times_softdrinks_intake
average_times_pickle_intake = sum(average_foodtype_intake[38:40])/138
average_times_pickle_intake
average_times_fried_intake = sum(average_foodtype_intake[40:42])/138
average_times_fried_intake

#calorie intake by obesse people from different food categories
weekly_calories_meat_fish = avg_calories_meat_fish*average_times_meat_fish_intake #weekly calorie intake from meat/fish
weekly_calories_beans_tofu = avg_calories_beans_tofu*average_times_beans_tofu_intake
weekly_calories_milk = avg_calories_milk_milkproducts*average_times_milk_intake
weekly_calories_greenveg = avg_calories_greenveg*average_times_greenveg_intake
weekly_calories_nocolorveg = avg_calories_nocolor_vegetables*average_times_nocolorveg_intake
weekly_calories_fruits = avg_calories_Fruits*average_times_fruits_intake
weekly_calories_grains = avg_calories_Grains*average_times_grains_intake
weekly_calories_sweets = avg_calories_Sweets*average_times_sweets_intake
weekly_calories_coffee = avg_calories_Coffee*average_times_coffee_intake
weekly_calories_tea = avg_calories_Tea*average_times_tea_intake
weekly_calories_snacks = avg_calories_Snacks*average_times_snacks_intake
weekly_calories_softdrinks = avg_calories_Soft_drinks*average_times_softdrinks_intake
weekly_calories_pickle = avg_calories_Pickles*average_times_pickle_intake
weekly_calories_friedfood = avg_calories_Fried_food*average_times_fried_intake

weekly_calories_meat_fish
weekly_calories_beans_tofu
weekly_calories_milk
weekly_calories_greenveg
weekly_calories_nocolorveg
```

weekly_calories_fruits
weekly_calories_grains
weekly_calories_sweets
weekly_calories_coffee
weekly_calories_tea
weekly_calories_snacks
weekly_calories_softdrinks
weekly_calories_pickle
weekly_calories_friedfood

Food Category	Average no. of times eaten per week	Total calorie intake per week
Meat/ fish	5.630631	1198.741
Beans/tofu	4.54955	685.8777
Milk	4.225225	529.3827
Green Vegetables	8.621622	516.1476
Colorless Vegetables	13.32432	856.7751
Fruits	9.459459	669.7129
Grains	13.61261	1490.579
Sweets	8.905797	1213.377
Coffee	5.521739	428.2624
Tea	3.985507	156.2385
Snacks	5.362319	913.5161
Soft Drinks	4.101449	649.021
Pickle	4.231884	90.22483
Fried food	5.956522	1252.437

Figure10: Weekly calorie estimates of Obese people according to the food categories

(b) <u>To predict calorie value of different food items based on other nutritive qualities of food.</u>

<u>Model 1: Linear Regression</u>

mm <- lm(food_table_norm$Calories ~ food_table_norm$Saturated_Fats) # build linear regression model on full data
print(mm)

```
> print(mm)

Call:
lm(formula = food_table_norm$Calories ~ food_table_norm$Saturated_Fats)

Coefficients:
                   (Intercept)  food_table_norm$Saturated_Fats
                       0.04742                         0.77867
```

summary(mm) # model summary

```
> summary(mm)  # model summary

Call:
lm(formula = food_table_norm$Calories ~ food_table_norm$Saturated_Fats)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.42764 -0.03352 -0.00822  0.02396  0.30970 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    0.047418   0.001403   33.80   <2e-16 ***
food_table_norm$Saturated_Fats 0.778670   0.013137   59.27   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05287 on 2012 degrees of freedom
Multiple R-squared:  0.6359,	Adjusted R-squared:  0.6357 
F-statistic:  3513 on 1 and 2012 DF,  p-value: < 2.2e-16
```

```
# Create Training and Test data -
set.seed(100)  # setting seed to reproduce results of random sampling
trainingRowIndex <- sample(1:nrow(food_table_norm), 0.8*nrow(food_table_norm))
# row indices for training data
trainingData <- food_table_norm[trainingRowIndex, ] # model training data
testData  <- food_table_norm[-trainingRowIndex, ]

# Build the model on training data -
lmMod <- lm(food_table_norm$Calories ~ food_table_norm$Saturated_Fats  +
food_table_norm$Solid_Fats , data=trainingData)  # build the model
distPred <- predict(lmMod, testData)  # predict distance

summary (lmMod)
```

```
> summary (lmMod)

Call:
lm(formula = food_table_norm$Calories ~ food_table_norm$Saturated_Fats + 
    food_table_norm$Solid_Fats, data = trainingData)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.41431 -0.03323 -0.00826  0.02474  0.29924 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    0.047748   0.001379  34.637   <2e-16 ***
food_table_norm$Saturated_Fats 0.588381   0.025508  23.066   <2e-16 ***
food_table_norm$Solid_Fats     0.238525   0.027582   8.648   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05193 on 2011 degrees of freedom
Multiple R-squared:  0.6489,	Adjusted R-squared:  0.6486 
F-statistic:  1858 on 2 and 2011 DF,  p-value: < 2.2e-16
```

```
actuals_preds <- data.frame(cbind(actuals=testData$dist, predicteds=distPred)) # make actuals_predicteds dataframe.
correlation_accuracy <- cor(actuals_preds)
head(actuals_preds)
```

Model 2: KKNN

```
food_table<-read.csv(file.choose(), header = TRUE)
food_table<-data.frame(food_table)
#Calories.factor
food_table$Calories.factor <- factor(food_table$Calories)
food_table$Calories.cat <- NA
food_table$Calories.cat <- ifelse(food_table$Calories>=200, 'High Calorie food item',
food_table$Calories.cat)
food_table$Calories.cat <- ifelse((food_table$Calories<200 & food_table$Calories>100) , 'Medium Calorie Item', food_table$Calories.cat)
food_table$Calories.cat <- ifelse(food_table$Calories<=100, 'Low Calorie Item’,
food_table$Calories.cat)

food_table$Calories.cat <- factor(food_table$Calories.cat, levels = c("High Calorie food item", "Medium Calorie Item", "Low Calorie Item"))

##Generate a random number that is 90% of the total number of rows in dataset.
ran <- sample(1:nrow(food_table), 0.8 * nrow(food_table))

##the normalization function is created
nor <-function(x) {(x-min(x))/(max(x)-min(x))}

##Run nomalization on the selected coulumns of dataset because they are the predictors
food_table_norm <-
as.data.frame(lapply(food_table[,c(10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,27)], nor))

summary(food_table_norm)

##extract training set
food_table_train <- food_table_norm[ran,]
##extract testing set
food_table_test <- food_table_norm[-ran,]
##extract 12th column of train dataset because it will be used as 'cl' argument in knn function.
food_table_target_category <- food_table[ran,26]
##extract 12th column if test dataset to measure the accuracy
food_table_test_category <- food_table[-ran,26]
##load the package class
library(class)
##run knn function
pr <- knn(food_table_train, food_table_test, cl=food_table_target_category, k=3)
```

```
##create confusion matrix
tab <- table(pr,food_table_test_category)
```

##this function divides the correct predictions by total number of predictions that tell us how accurate teh model is.

```
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy(tab)
```

With the KKNN model, I achieved an Accuracy of 84.32% and with Linear Model, I achieved an accuracy of 64.89 %.

5. Conclusions and Discussion

In my project, I performed an analysis of the diets of non-obese people and the obese people and was able draw the following results:

Food category	Average amount of calories per item in the food category
Meat/fish	212.8964
beans/tofu	150.7573
milk/milk products	125.291
Green vegetables	59.86665
No color vegetables	64.30158
Fruits	70.79822
Grains	109.4999
Sweets	136.2457
Coffee	77.55933
Tea	39.20167
Snacks	170.3584
Soft drinks	158.2419
Pickles	21.32025
Fried food	210.2632

Figure11: Calorie estimates of commonly eaten food items in America according to the food categories

Combined with sedentary lifestyles, the easily available, inexpensive, high-calorie foods lead to a significant prevalence of overweight and obesity. Current eating pattern needs to be improved and they can be moved toward healthier eating patterns by making shifts in food choices over time. Making these shifts can help support a healthy body weight, meet nutrient needs, and lessen the risk for chronic disease.

Food Category	Average no. of times eaten per week	Total calorie intake per week
Meat/ fish	5.630631	1198.741
Beans/tofu	4.54955	685.8777
Milk	4.225225	529.3827
Green Vegetables	8.621622	516.1476
Colorless Vegetables	13.32432	856.7751
Fruits	9.459459	669.7129
Grains	13.61261	1490.579
Sweets	8.905797	1213.377
Coffee	5.521739	428.2624
Tea	3.985507	156.2385
Snacks	5.362319	913.5161
Soft Drinks	4.101449	649.021
Pickle	4.231884	90.22483
Fried food	5.956522	1252.437

Figure12: Weekly calorie estimates of Obese people according to the food categories

Obesity and overweigh is becoming an increasing problem presenting a risk to health. Improper dietary habits such as skipping breakfast, eating while watching TV, low fruit and vegetable consumption, and more sedentary living with no physical activity can be associated with the problem of overweight and obesity among people.

I applied the models of Linear Regression and KKNN in my project and was able to achieve accuracies of 64.89 % and 84.32% respectively. My overall analysis of Obese and non-obese people diets found that the categories of food with higher calorie content(and those that leads to obesity) are consumed on a larger scale than those that do not cause or leads to obesity/overweight.

Further in my project, I plan to take into consideration the other factors that leads to obesity in my project as well. This will give me a more reliable way of estimating the dietary habits of the people.

References

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877769/
2. https://health.gov/dietaryguidelines/2015/guidelines/chapter-2/current-eating-patterns-in-the-united-states/
3. https://www.hsph.harvard.edu/obesity-prevention-source/obesity-causes/diet-and-weight/
4. https://www.hsph.harvard.edu/obesity-prevention-source/diet-lifestyle-to-prevent-obesity/
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877769/