



The Fast Food Diet Trap: Analyzing Nutrition and Weight Watchers Points

Group 4

2024-10-11

Introduction

As graduate students, balancing academic responsibilities and maintaining a healthy lifestyle can become challenging, especially when convenience means grabbing a quick bite to eat for lunch at fast food restaurants like A&W, Diary Queen, McDonald's, Subway or Tim Hortons. They offer easily accessible meals but they also come with concerns about their nutritional content, high calorie and processed fat content.

Beyond graduate life, in today's world, we need to have an important discussion about the potential long-term impacts of these foods on our health. With this in mind, our project aims to address the issue of calorie and nutrient awareness among students, and make it easier for us to navigate the complexities of nutrition and fast food choices.

Dataset

Our dataset provides a breakdown of the nutrient and calorie value of items on the menu of selected fast-food restaurants. However, for the scope of this project, the focus is on three key nutrients: Fat (g) (including Trans Fat (g)), Protein (g), and Carbohydrate (g).

This dataset also contains Weight Watcher Points for each menu item. The Weight Watcher Points system is a program that enables people to easily track their diet and understand the nutrition implications of their fast food choices. People who join the Weight Watcher Program are given a maximum daily and weekly allowance of points. As you eat, you use up points, and once you hit zero, you should stop eating for the day. There are three types of Weight Watchers Points systems (WeightWatchers, n.d.):

1. *Plus Points*: assigns points based on calories and nutrition to help encourage healthier food choices.
2. *FreeStyle Points*: is the most flexible program that allows you to eat up to 200 foods without counting them towards daily points, while making activity a key part of your life.
3. *Smart Points*: uses calories, fats, and fiber to set points, with high-fiber foods costing fewer points, encouraging healthier, more filling choices. You can also rollover a set number of SmartPoints from previous days.

There are 745 rows and 28 columns in total, with 25 of those columns containing numeric data.

Below is a snippet of our dataset:

Company	Item	Calories	Calories.From.Fat	Total.Fat.g.	Saturated.Fat.g.	Trans.Fat.g.	FreeStyle.Points	SmartPoints	PointsPlus
1 A&W	Baby Burger without Cheese	240	72	8	3.0	0.2	8	8	6
2 A&W	Baby Burger with Cheese	298	117	13	6.0	0.4	10	10	8
3 A&W	Beyond Meat Burger	500	261	29	5.0	0.0	15	15	14
4 A&W	Buddy Burger without Cheese	310	117	13	4.5	0.2	11	11	8
5 A&W	Buddy Burger with Cheese	370	144	16	7.0	0.3	13	13	9
6 A&W	Cheddar Bacon Uncle Burger	760	432	48	19.0	1.0	25	25	20

Figure 1: Snippet of Dataset

We initially sourced our data from Kaggle. Within the Kaggle dataset, we noticed that the Weight Watcher Points in this data were more than the actual Weight Watcher Points obtainable. We also noticed that the Weight Watcher Points were oddly similar to the value of the calories for each menu item.

To solve this, we went to the original website where this information was hosted which is Fast Food Nutrition's website. We selected five Canadian fast food restaurants (A&W, Diary Queen, McDonald's, Subway, and Tim Hortons) from this site and webscraped their nutrition and weight watcher points information using Python. The webscrapping codes are available here. We then updated our code to reflect this new data.

Research Questions

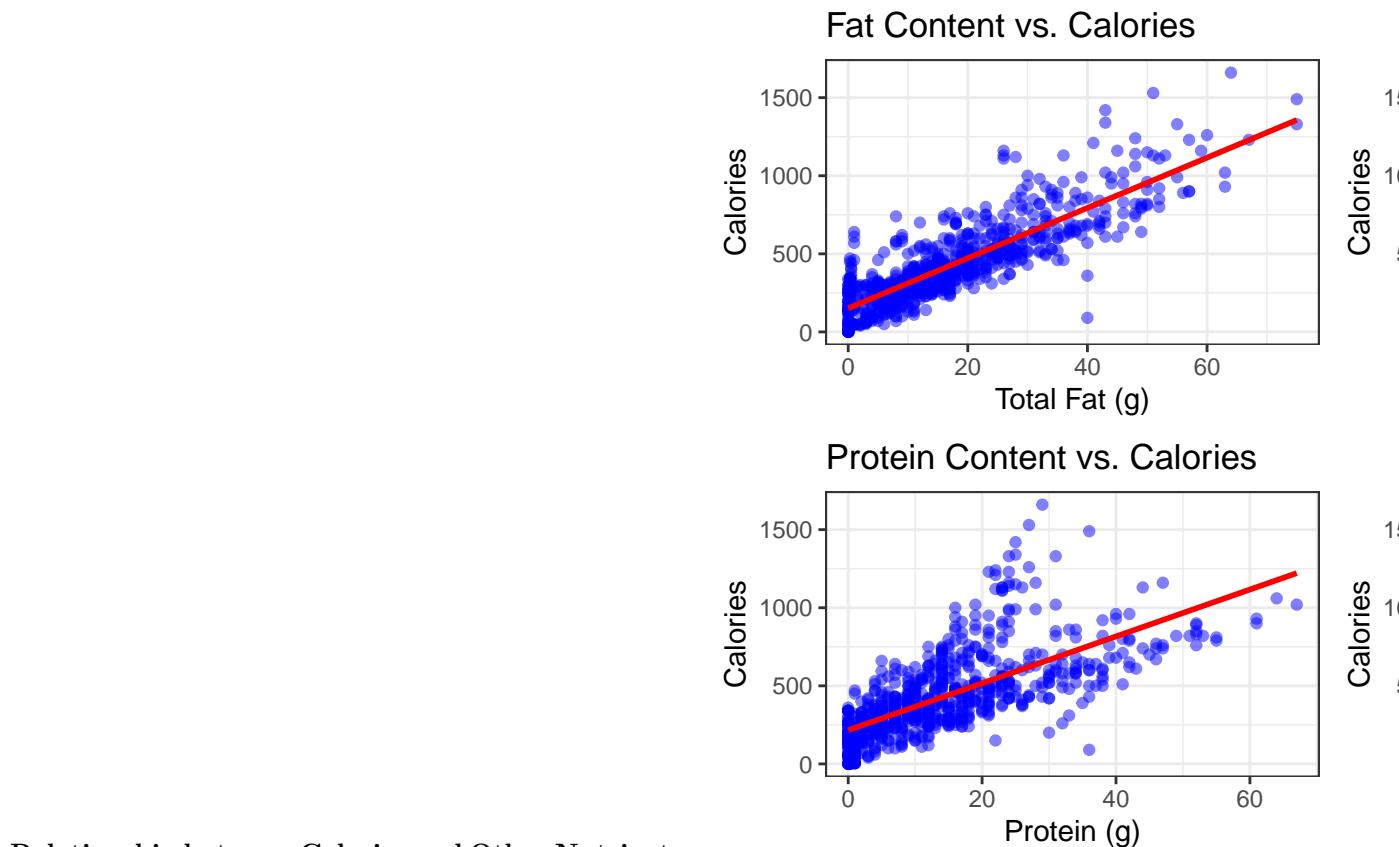
For this project, we will be exploring three research questions:

1. What is the relationship between calorie content and the three major nutrients: fat, carbohydrates, and protein, in various food items?
2. Is there a significant difference in the average trans fat content of menu items between A&W and Subway?
3. What is the relationship between calorie content and Plus Points across all listed companies?

Below, we will discuss the preliminary **exploratory data analysis (EDA)**, that informed the development of our research questions.

Question 1: What is the relationship between calorie content and the three major nutrients: fat, carbohydrates, and protein, in various food items?:

```
grid.arrange(cal_fat, cal_carb, cal_prot, cal_sodium, ncol=2)
```



Relationship between Calories and Other Nutrients

Calories are highly correlated with three major nutrients: fat, carbohydrates, and protein. The higher the content of these nutrients in food, the more calories it contains. Notably, some menu items have 0g of one of these nutrients but still possess a high caloric value. This suggests a multilinear relationship between calories and all three nutrients, rather than a simple linear relationship with any single nutrient, which we will address with a **multilinear regression analysis**.

Non-focal observations:

There seems to be two linear trends in the relationship between Calories and Sodium, which almost mimics the trend in Calories and Protein. Sodium is primarily contained in salt and spices. This could imply that the salted items on the menu are generally proteins.

Question 2 - Is there a significant difference in the average trans fat content of menu items between A&W and Subway?:

Processed (Bad) Fats Distributions in Each Restaurant

The types of fats found in food include:

Unsaturated Fat: These are the good fat which doctors say should be the fat people eat most. They come as liquid oils instead of solid fats. On food packages, they are listed as mono or poly unsaturated fats.

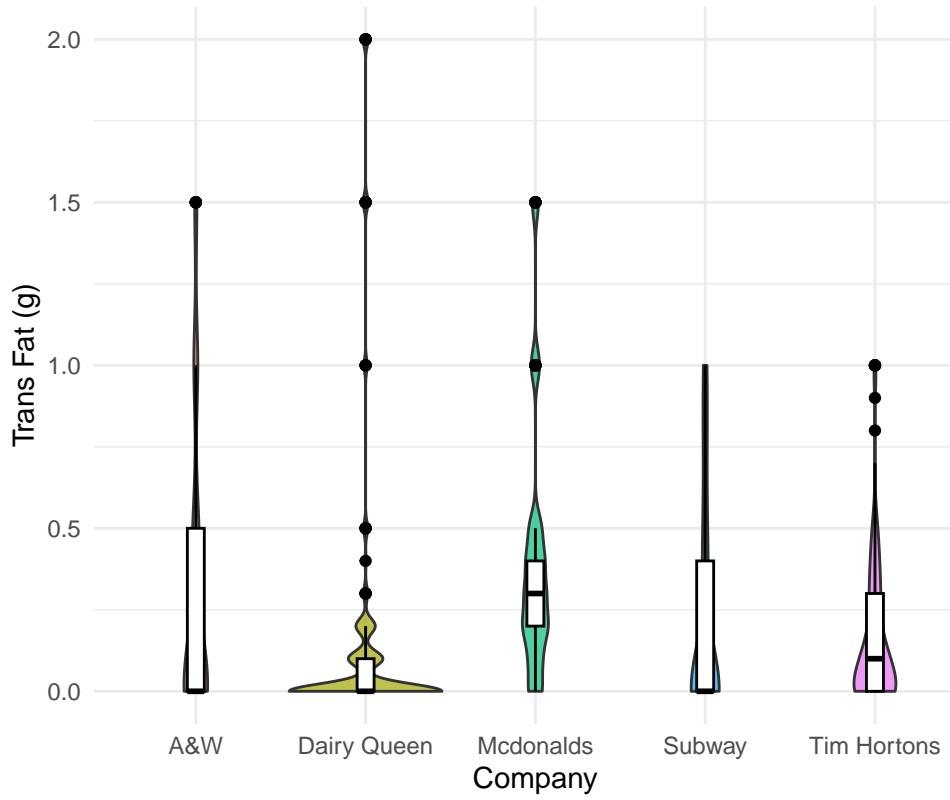
Saturated Fat: These fats come from animals and are generally solid. They contain “bad” cholesterol and can contribute to heart disease. It is recommended that they make up less than 10% of daily caloric intake. Examples include whole-fat milk, butter, high-fat cheeses etc.

Trans Fat: These are the worst fats, and the FDA is forcing food companies to phase them out. They form when hydrogen is added to vegetable oil. This is done to increase shelf life but are bad for our health. Examples of food more likely to contain trans fat include stick margarine, pie crusts, ready-to-use frosting, some microwave popcorn and frozen pizza, etc.

We aim to assess the content of these unhealthy Trans fats across various restaurants; therefore, we will examine the trans fat content of menu items from these different establishments.

```
print(trans_fat_violin_plot)
```

Violin Plot of Trans Fat Content by Restaurant



```

companies = unique(fast_food_df$Company)
for (company in companies) {
  cat("Favstats for", company, ":\n")
  stats = favstats(filter(fast_food_df, Company == company)$Trans.Fat...g., na.rm = TRUE)
  print(stats)
}

## Favstats for A&W :
##   min Q1 median Q3 max      mean        sd    n missing
##     0    0      0 0.5 1.5 0.2873874 0.4257864 111      0
## Favstats for Dairy Queen :
##   min Q1 median Q3 max      mean        sd    n missing
##     0    0      0 0.1  2 0.1391111 0.3809986 225      0
## Favstats for Mcdonalds :
##   min Q1 median Q3 max      mean        sd    n missing
##     0 0.2    0.3 0.4 1.5 0.3841727 0.37169 139      0
## Favstats for Subway :
##   min Q1 median Q3 max      mean        sd    n missing
##     0    0      0 0.4  1 0.2044444 0.3274943 45      0
## Favstats for Tim Hortons :
##   min Q1 median Q3 max      mean        sd    n missing
##     0    0      0.1 0.3  1 0.1564444 0.2122704 225      0

```

It looks like these companies are making efforts to phase out Trans Fat since their central tendencies are less than 0.5g. All of these companies have outliers with items that have more than 1g of trans fat.

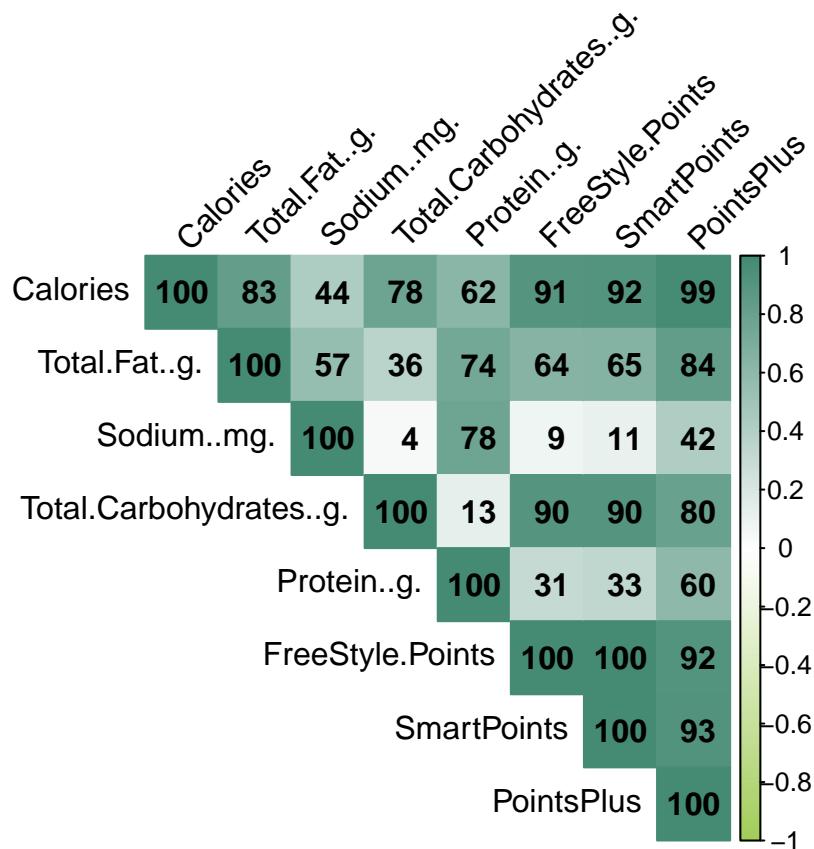
McDonald's has the highest median trans fat content compared to the others. The median trans fat content for A&W and Subway is closest to 0g, but which one has less trans fat content? We can test this hypothesis later on. Perhaps we can consider menu items from A&W and Subway as a "healthier" alternative to McDonald's.

Question 3 - What is the relationship between calorie content and Weight Watcher Plus Points across all listed companies?:

Correlations between Selected Nutrients, Calories, and Weight Watcher Points

There are three categories of Weight Watcher Points: Freestyle Points, Smart Points, and Plus Points.

```
# Heat map of the correlation matrix
corrplot(cor_matrix, method = "color", type = "upper", tl.col = "black",
         tl.srt = 45,
         col = colorRampPalette(c("darkolivegreen3", "white",
                                 "aquamarine4"))(200),
         addCoef.col = "black", addCoefasPercent = TRUE)
```



From the correlation plot, we see that Calories are highly correlated with Fat and Carbs, with medium correlation to Proteins.

For the Weight Watcher Points categories: 1. *Freestyle Points*: Highly correlated with calories and carbohydrates. 2. *SmartPoints*: Highly correlated with Calories and Carbohydrates. 3. *PointsPlus*: Most correlated with Calories. Good correlations with carbs and fat.

Since all the Weight Watchers point categories are correlated with calories, we will take a closer look at this relationship.

```

# calories vs freestyle
fsp = ggplot(fast_food_df,
             mapping = aes(x = Calories, y = FreeStyle.Points)) +
  geom_point() +
  geom_smooth(mapping = aes(group = Company, color = Company),
              show.legend = TRUE, method = 'loess', formula = 'y~x') +
  labs(title = "Scatter Plot of Calories vs. Weight Watchers Freestyle Points",
       x = "Calories", y = "Freestyle Points", color = "Company")

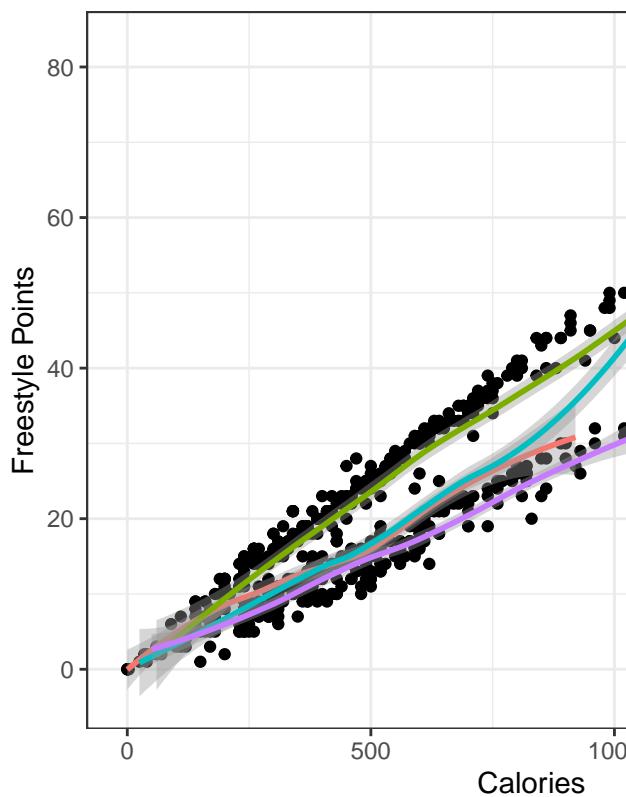
# calories vs points plus
pp = ggplot(fast_food_df,
            mapping = aes(x = Calories, y = PointsPlus)) +
  geom_point() +
  geom_smooth(mapping = aes(group = Company, color = Company),
              show.legend = TRUE, method = 'loess', formula = 'y~x') +
  labs(title = "Scatter Plot of Calories vs. Weight Watchers Plus Points",
       x = "Calories", y = "Plus Points", color = "Company")

# calories vs smart points
sp = ggplot(fast_food_df,
            mapping = aes(x = Calories, y = SmartPoints)) +
  geom_point() +
  geom_smooth(mapping = aes(group = Company, color = Company),
              show.legend = TRUE, method = 'loess', formula = 'y~x') +
  labs(title = "Scatter Plot of Calories vs. Weight Watchers Smart Points",
       x = "Calories", y = "Smart Points", color = "Company")

```

print(fsp)

Scatter Plot of Calories vs. Weight

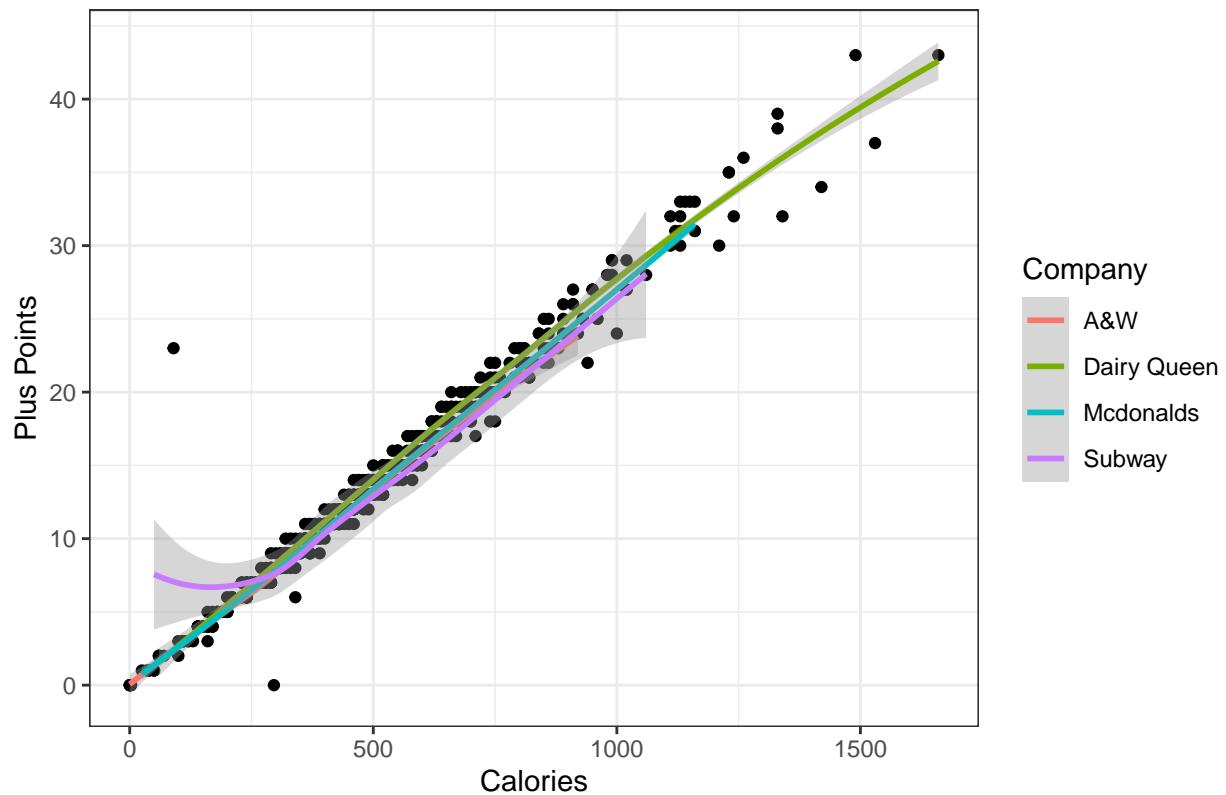


Relationships between Weight Watcher Points and Calories.

The calories of menu items in these companies are positively correlated with the Freestyle Points but McDonald's trend deviate slightly from that of A&W, Dairy Queen, and Subway.

```
print(pp)
```

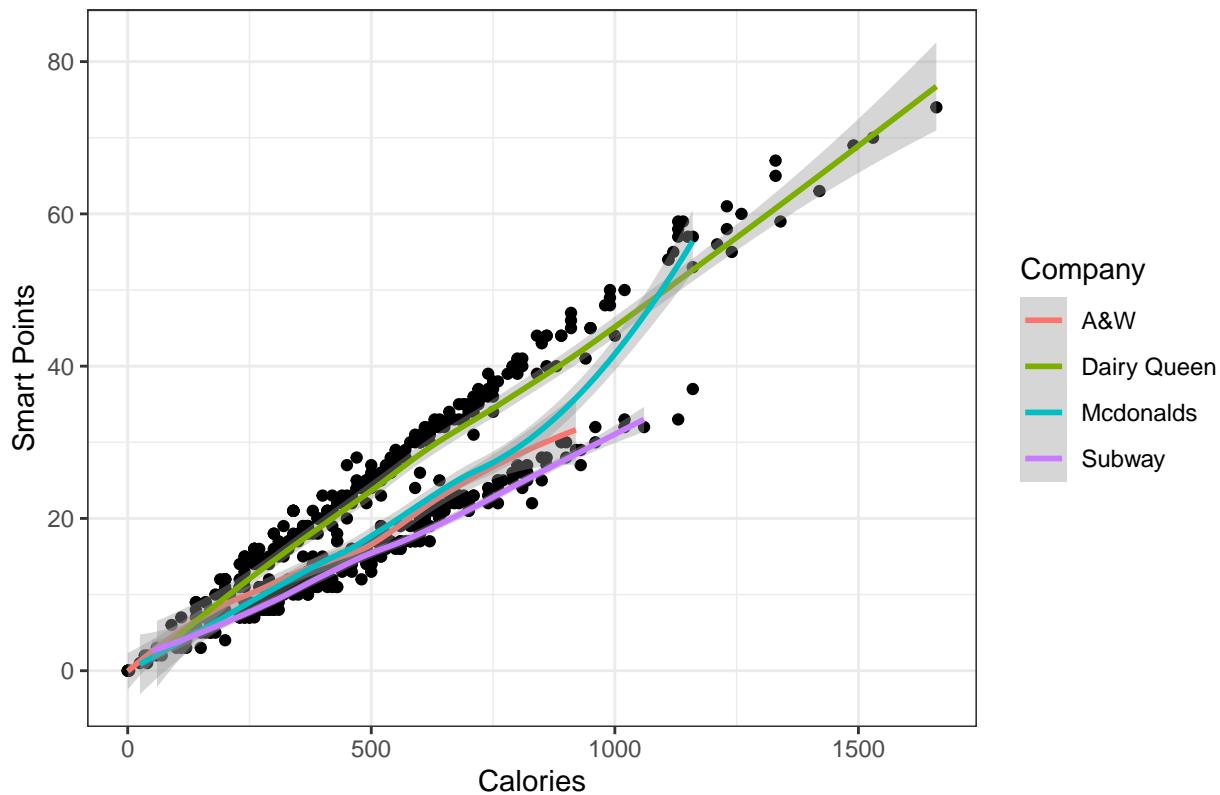
Scatter Plot of Calories vs. Weight Watchers Plus Points



The calories of menu items in these companies are positively correlated with the Plus Points and all listed restaurants follow the same trend.

```
print(sp)
```

Scatter Plot of Calories vs. Weight Watchers Smart Points



Similar to the Freestyle Points, the calories of menu items in these companies are positively correlated with the Freestyle Points but McDonald's trend deviate slightly from that of A&W, Dairy Queen, and Subway.

Application of Hypothesis Testing and Regression Analysis to Address Our Research Questions

Question 1: What is the relationship between calorie content and the three major nutrients: fat, carbohydrates, and protein, in various food items?

Multivariate Regression Analysis Carbohydrates, Fat, Protein and Calories

We identified that calories were highly correlated with fat, protein, and carbohydrates. We can create a multi-linear regression analysis between these nutrients and the calories.

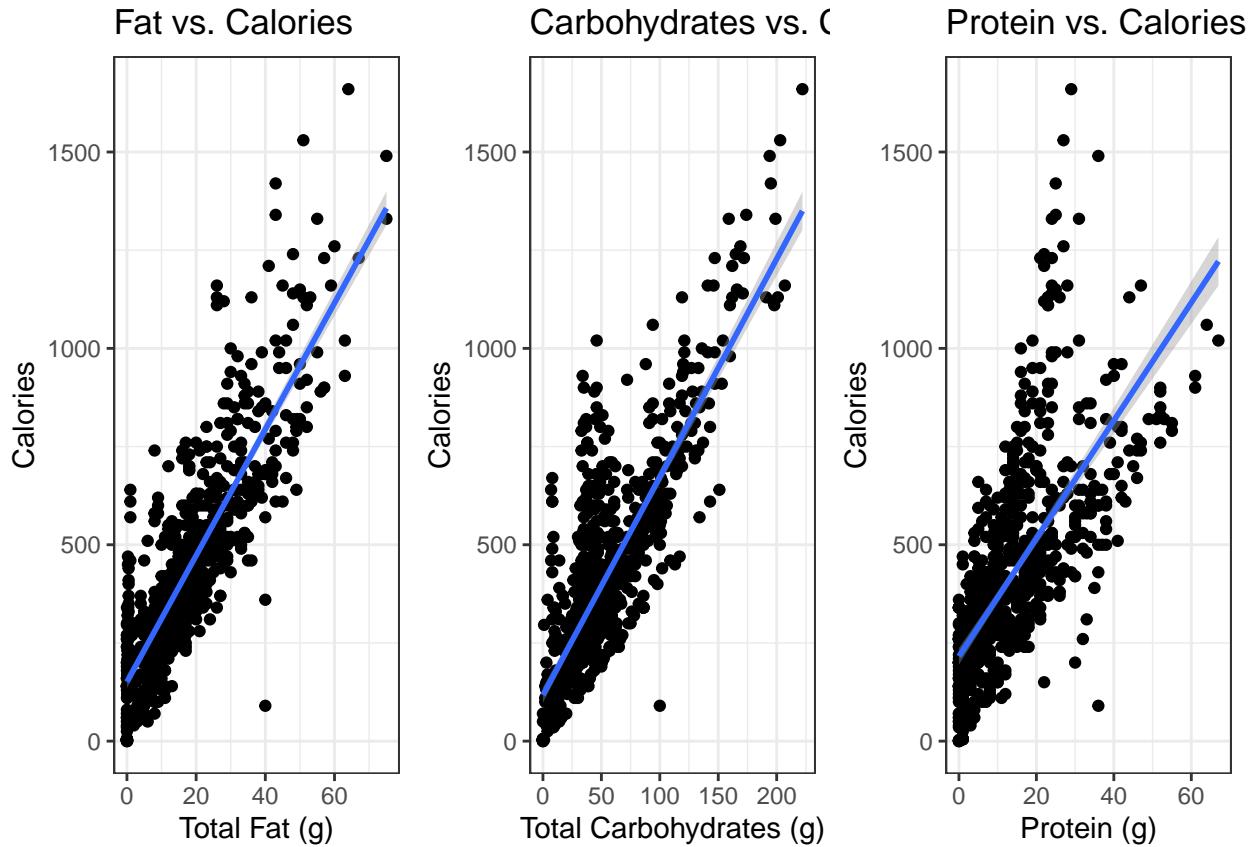
First we can recalculate the correlation coefficient between these nutrients and Calories.

```
cat("The linear correlation between calories and fat is", fat_cal_cor,
  "\nthat of calories and protein is", prot_cal_cor,
  "\nand that of calories and carbohydrates is", carb_cal_cor)
```

```
## The linear correlation between calories and fat is 0.8499438
## that of calories and protein is 0.6887784
## and that of calories and carbohydrates is 0.8133983
```

Next, we visualize and build a multi linear regression model.

```
grid.arrange(fat, carbs, protein, ncol=3)
```



```
print(ml_reg)
```

```

## Call:
## lm(formula = Calories ~ Total.Carbohydrates..g. + Protein..g. +
##      Total.Fat..g., data = fast_food_df)
##
## Coefficients:
##             (Intercept)  Total.Carbohydrates..g.          Protein..g.
##                   5.516                  3.847                  4.204
##      Total.Fat..g.
##                   8.643

```

Based on the multi linear regression model:

$\hat{y} = 5.516 + 3.847C + 4.204P + 8.643F$ where \hat{y} is the predicted Calories, C is carbohydrate content, P is the protein content, and F is the fat content.

Next, we check the significance of the model coefficients.

```
round(summary(ml_reg)$coefficients[, "Pr(>|t|)"], 4)
```

Protein...g.

```

##          0.0411
## Total.Fat..g.
##          0.0000

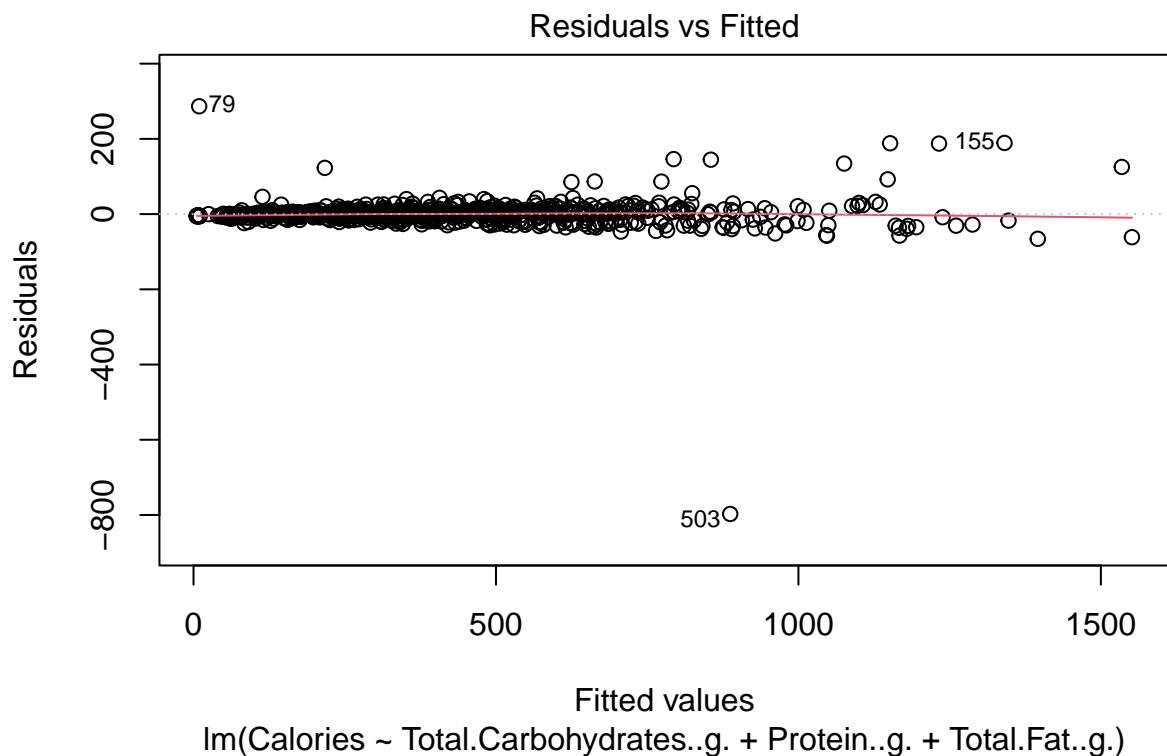
```

The p-values for the intercept (0.04) and all model (0) coefficients are less than 0.05, indicating that all are significant

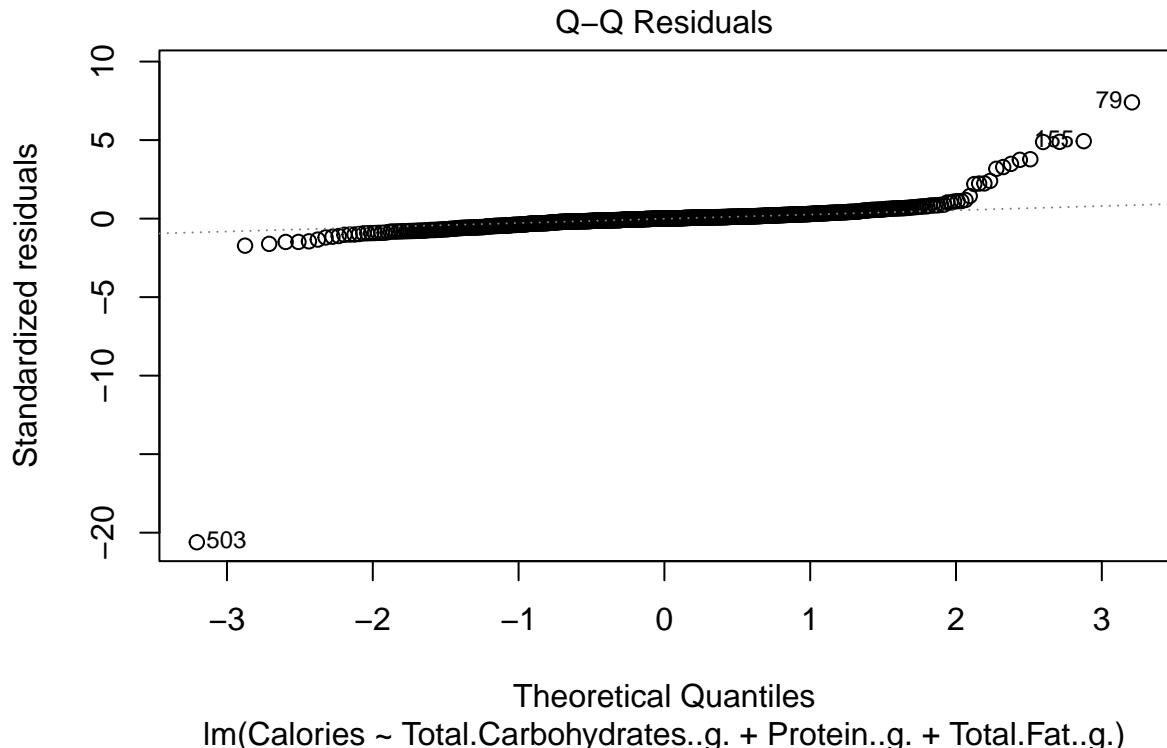
Essentially, for every 1g of carbohydrates, you can expect to add roughly 4 calories in a menu item, for every 1g of protein, you expect to add 4 calories, and for every 1g of Fat, you expect to add, you can expect to lose roughly 3 of your Weight Watchers Plus Points.

Finally, we visualize the residual plots.

```
plot(ml_reg, which=1)
```



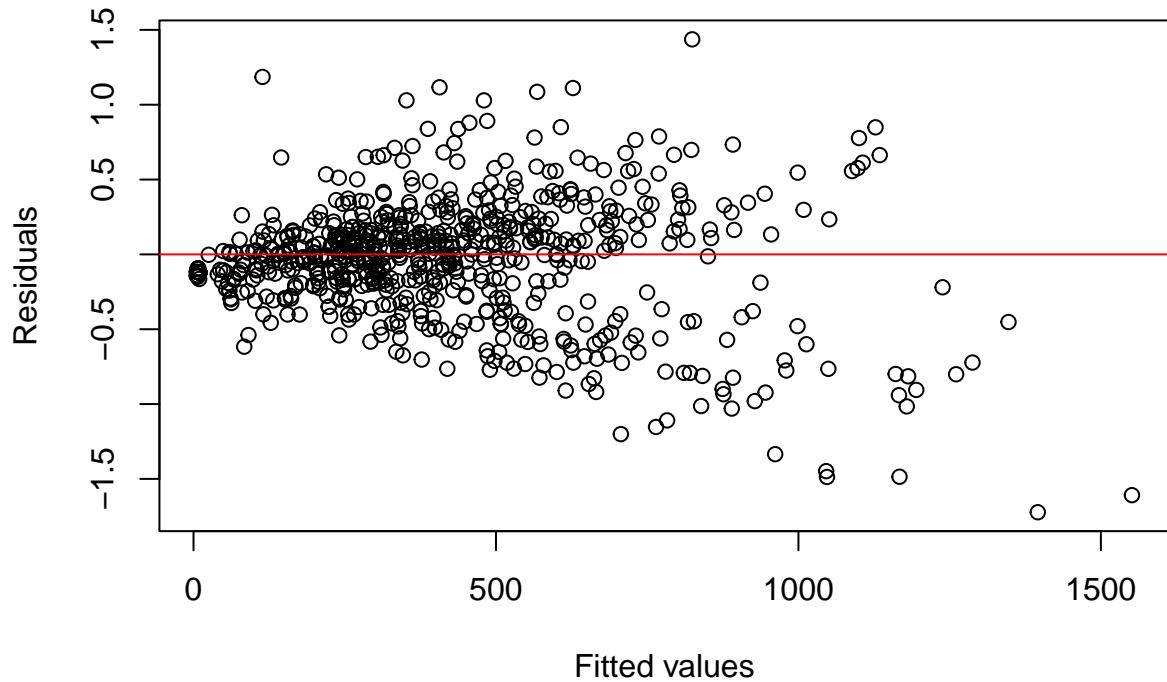
```
plot(ml_reg, which=2)
```



Since we have outliers in the plot, we will remove these outliers using a cut-off point of residuals greater than 2 and plot these residual graphs again without the outliers.

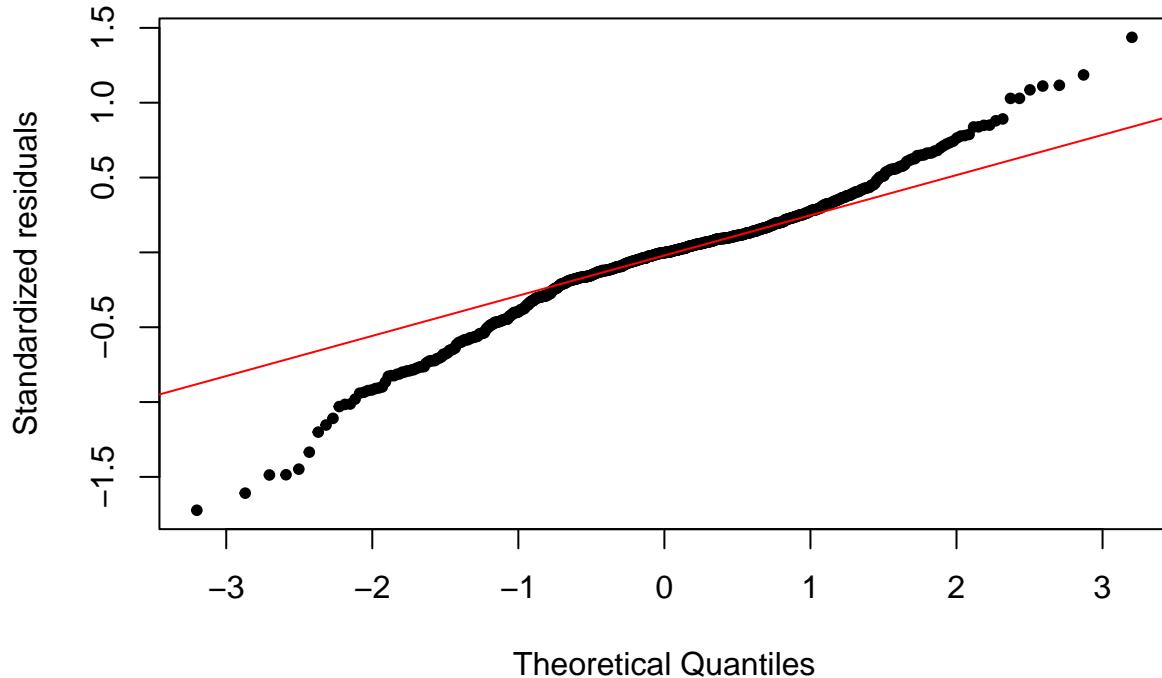
```
plot(fitted(ml_reg)[non_outlier_index] ,
      residuals[non_outlier_index] ,
      xlab = "Fitted values",
      ylab = "Residuals",
      main = "Residuals vs Fitted (Without Outliers)")
abline(h = 0, col = "red")
```

Residuals vs Fitted (Without Outliers)



```
qqnorm(residuals[non_outlier_index],  
       main = "Q-Q Plot (Without Outliers)",  
       xlab = "Theoretical Quantiles",  
       ylab = "Standardized residuals",  
       pch = 20)  
qqline(residuals[non_outlier_index], col = "red")
```

Q-Q Plot (Without Outliers)



The residuals show no pattern and pass the test of independence (Residuals vs Fitted). However, they fail the test of normality.

Question 2 :What is the relationship between calorie content and the three major nutrients: fat, carbohydrates, and protein, in various food items?

HYPOTHESIS TESTING

State the hypothesis

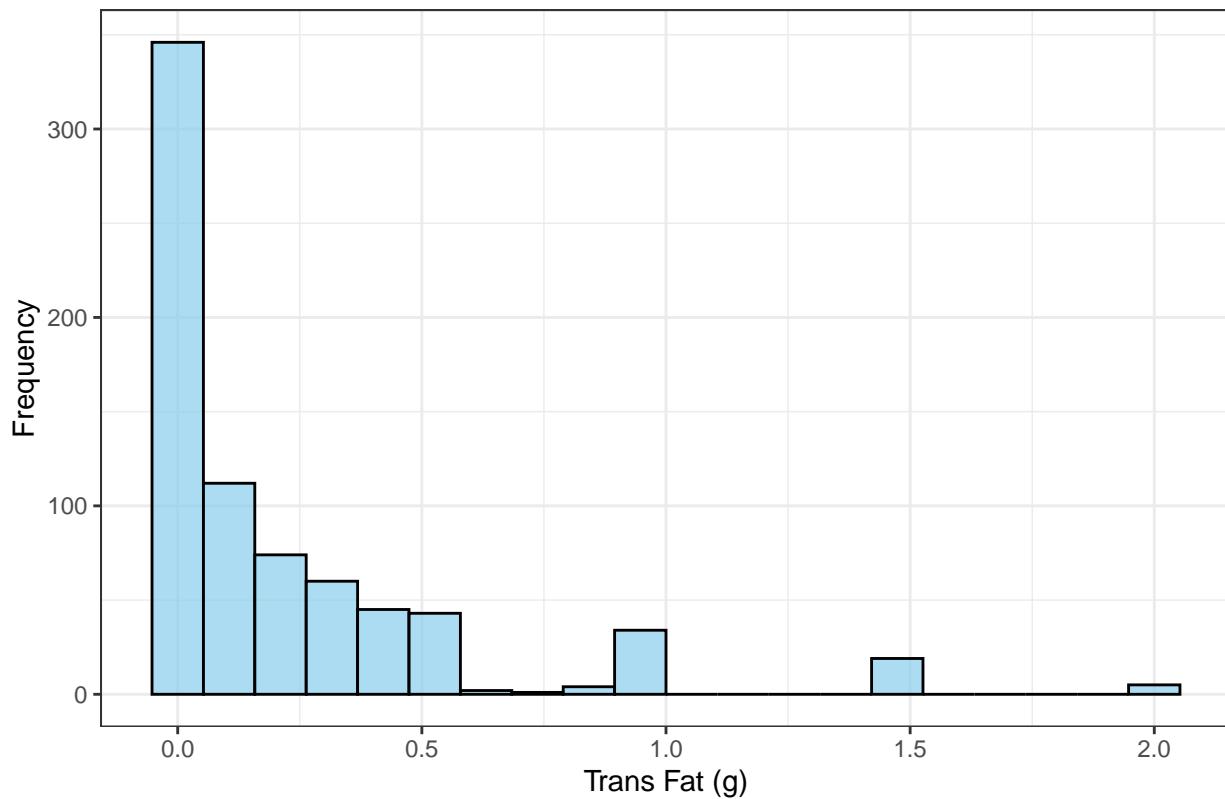
We identified in the EDA that the trans fat content of menu items in A&W and Subway are closer to 0g than other restaurants. We want to test and see whether the trans fat in menu items of both companies are equal, on average.

1. Null Hypothesis(H₀) - Mean Trans Fat of menu items in A&W and Subway are equal. i.e. $\mu_{aw} = \mu_{dq}$
2. Alternative Hypothesis(H₁) - Mean Trans fat of menu items in A&W and Subway are not equal i.e. $\mu_{aw} \neq \mu_{dq}$

First, check the distribution of Trans Fat.

```
print(hist_tf)
```

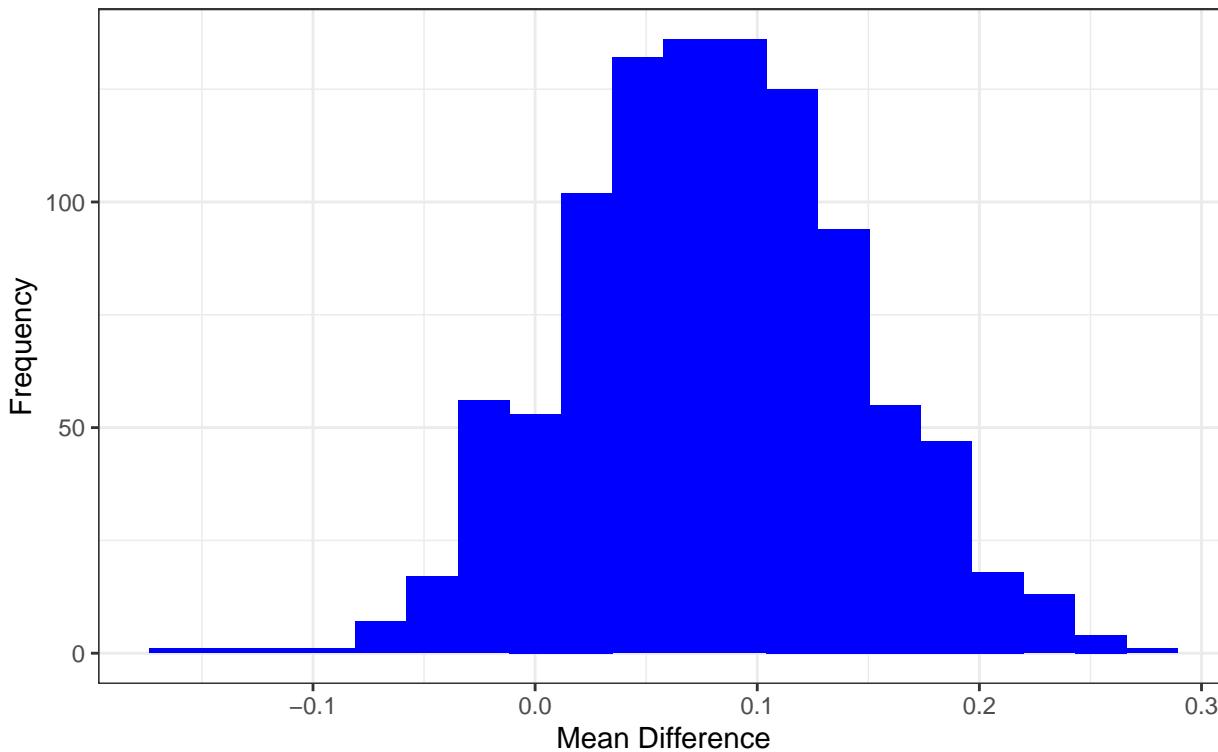
Histogram of Trans Fat Distribution



As the data is right skewed and not normally distributed, we will use bootstrapping as a non-parametric approach and permutation test as a parametric approach to test the hypothesis.

```
ggplot(b, aes(x=result)) + geom_histogram(fill='blue', bins=20) +
  labs(title = 'Mean difference in Trans Fat',
       subtitle = 'Between A&W and Subway',
       x = 'Mean Difference', y='Frequency')
```

Mean difference in Trans Fat Between A&W and Subway



Bootstrapping

```
# 95% confidence interval
quantile(b$result, c(0.025,0.975))
```

```
##          2.5%      97.5%
## -0.03935586  0.20992342
```

```
# Display the results from permutation test
cat("Permutation Test p-value:", p_value)
```

Permutation test

```
## Permutation Test p-value: 0.133
```

Result of the hypothesis testing:

Given that the p-value is 0.133, we do not have sufficient evidence to reject the null hypothesis at the 5% significance level. Since the confidence interval includes zero, this further supports our null hypothesis.

This means that we do not have evidence to conclude that the mean trans fat content of menu items in A&W and Subway is different.

Question 3: What is the relationship between calorie content and Plus Points across all listed companies?

Univariate Regression Analysis Calories and Weight Watchers Plus Points

From the EDA, since the Calories vs. Plus Points follow the same trend for all listed companies, we can create a regression analysis between the Calories and the Plus Points.

First we can recalculate the correlation coefficient between Calories and Plus Points.

```
cor(fast_food_df$Calories, fast_food_df$PointsPlus, use='complete.obs',
     method = 'pearson')
```

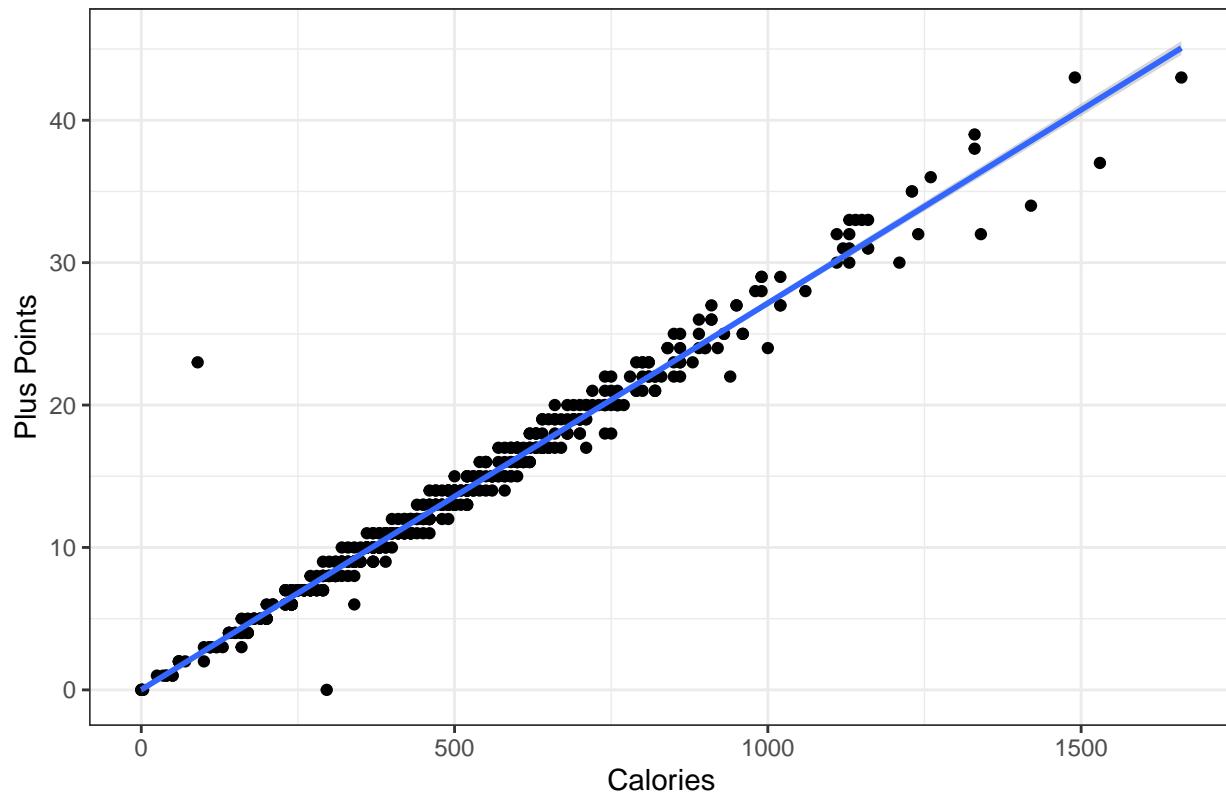
```
## [1] 0.9851189
```

There is a strong positive linear correlation between Calories and Plus Points.

Next, we visualize the regression line and build a simple linear regression model.

```
ggplot(fast_food_df, mapping = aes(x = Calories, y = PointsPlus)) +
  geom_point() +
  stat_smooth(geom='smooth', method = 'lm', formula = 'y~x') +
  labs(title = "Scatter Plot of Calories vs. Weight Watchers Plus Points",
       x = "Calories", y = "Plus Points")
```

Scatter Plot of Calories vs. Weight Watchers Plus Points



```
reg = lm (PointsPlus ~ Calories, data=fast_food_df)
print(reg)
```

```
##
## Call:
## lm(formula = PointsPlus ~ Calories, data = fast_food_df)
##
## Coefficients:
## (Intercept)    Calories
##      0.009975   0.027144
```

Based on the simple linear regression model:

$\hat{y} = 0.009975 + 0.027144X$ where \hat{y} is the predicted Weight Watchers Plus Points and X is the caloric value of the menu item.

Next, we check the significance of the model coefficients.

```
summary(reg)$coefficients[, "Pr(>|t|)"]
```

```
## (Intercept)    Calories
##      0.9333009   0.0000000
```

The p-value of the intercept is greater than 0.05, so the intercept model coefficient β_0 is insignificant.

The p-value of the calories (independent variable) is less than 0.05, so the independent variable model coefficient β_1 is significant.

Since, the model coefficient of the calories is significant, we can compute a 95% confidence interval for this model coefficient.

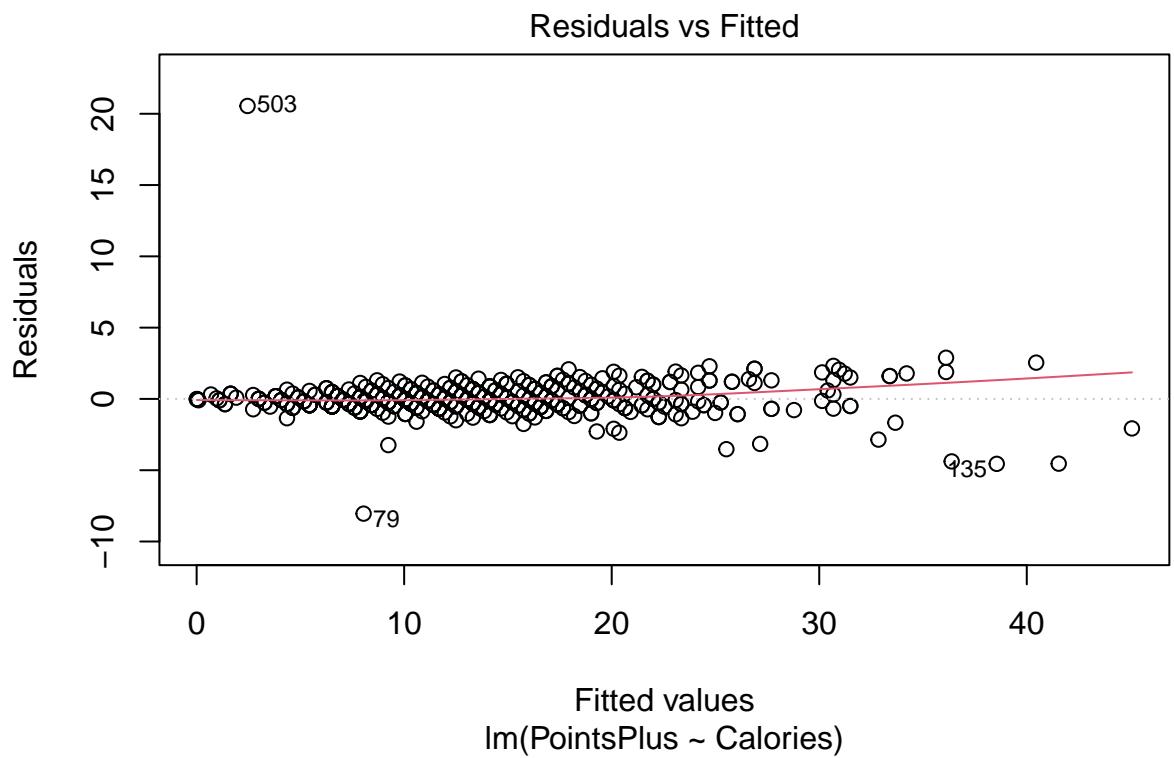
```
beta_1_CI = confint(reg, "Calories", level=0.95)
cat("We are 95% confident that the model coefficient of the Calories is
    in the range (",
    beta_1_CI[1], ",",
    beta_1_CI[2], ")")
```

```
## We are 95% confident that the model coefficient of the Calories is
##     in the range ( 0.02673505 , 0.02755262 )
```

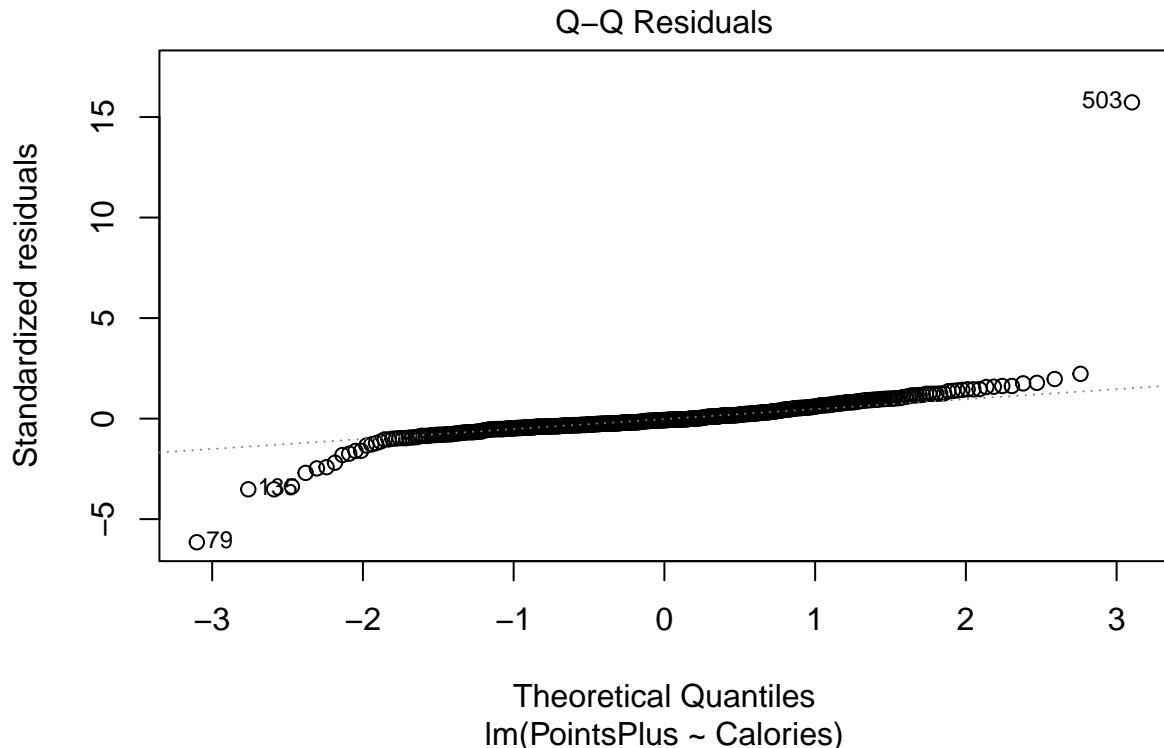
Essentially, for every 100 calories present in a menu item, you can expect to lose roughly 3 of your Weight Watchers Plus Points.

Finally, we visualize the residual plots.

```
plot(reg, which=1)
```



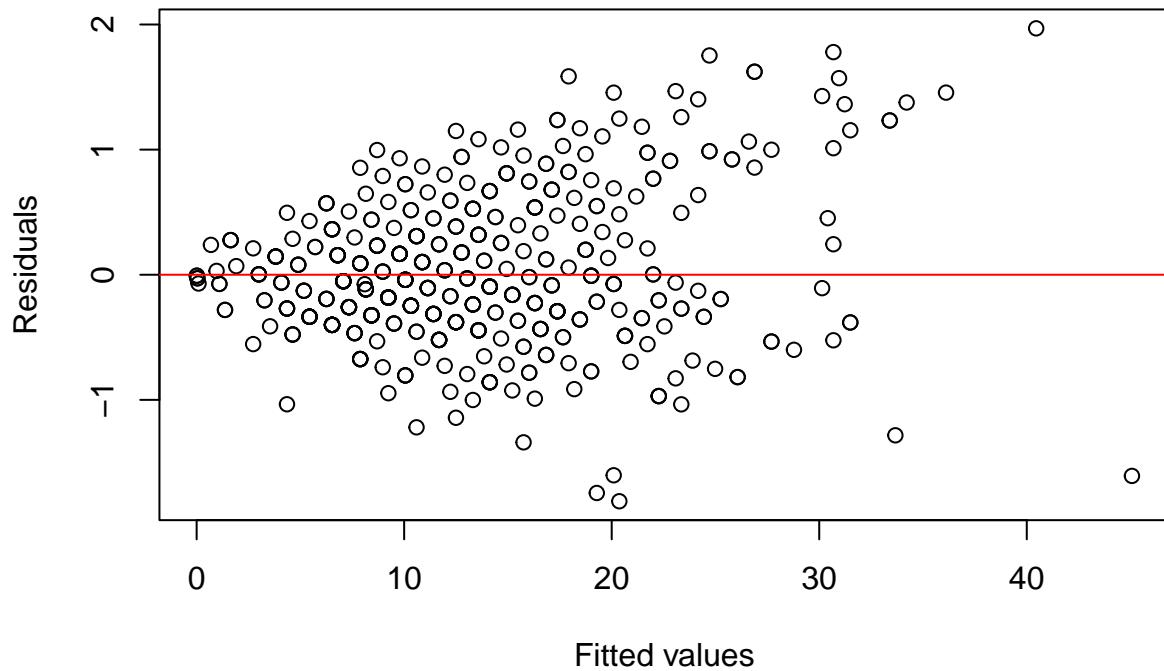
```
plot(reg, which=2)
```



Again, we remove the outliers and regenerate the plot.

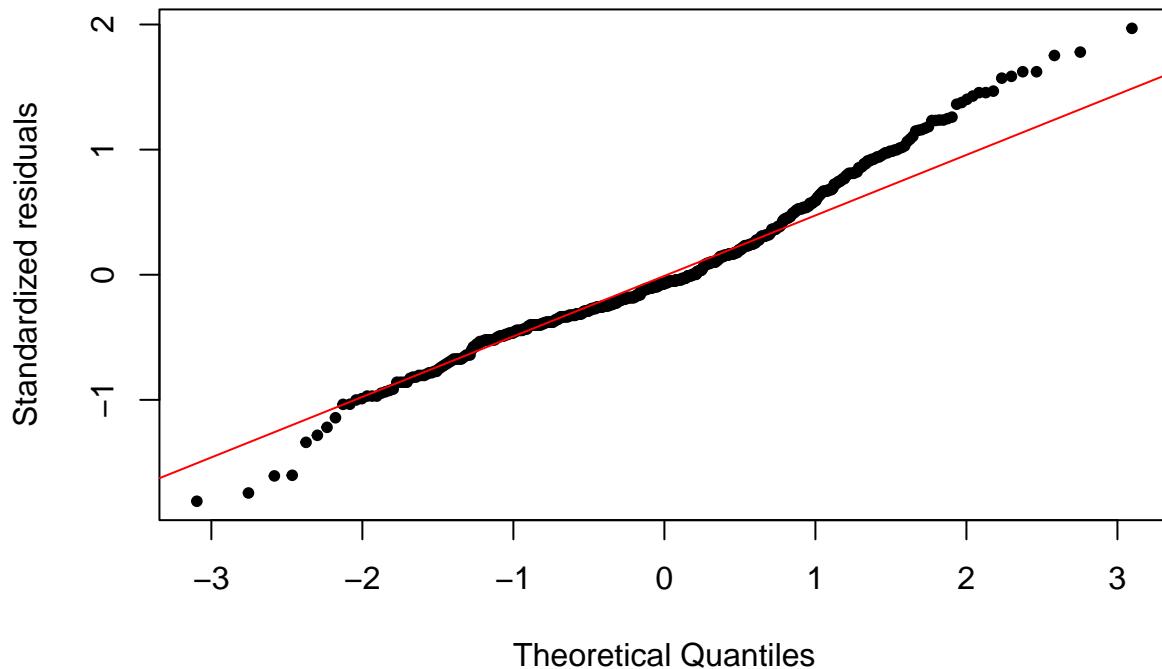
```
plot(fitted(reg)[non_outlier_index],
      residuals[non_outlier_index],
      xlab = "Fitted values",
      ylab = "Residuals",
      main = "Residuals vs Fitted (Without Outliers)")
abline(h = 0, col = "red")
```

Residuals vs Fitted (Without Outliers)



```
qqnorm(residuals[non_outlier_index],  
       main = "Q-Q Plot (Without Outliers)",  
       xlab = "Theoretical Quantiles",  
       ylab = "Standardized residuals",  
       pch = 20)  
qqline(residuals[non_outlier_index], col = "red")
```

Q-Q Plot (Without Outliers)



The residuals show no pattern and pass the test of independence (Residuals vs Fitted). From the Q-Q plot however, they fail the test of normality.

Applying the Weight Watchers Points System Calculator

```
print(plus_points_summary)
```

```
## # A tibble: 5 x 5
##   Company      minimum_pp median_pp  mean_pp max_pp
##   <chr>        <dbl>     <dbl>    <dbl>   <dbl>
## 1 A&W          0         9     9.57    24
## 2 Dairy Queen  2        14    16.1     43
## 3 Mcdonalds   1        12    12.9     31
## 4 Subway        1        13    13.6     28
## 5 Tim Hortons Inf       NA    NaN      -Inf
```

Since we do not have the Weight Watcher Points of Tim Hortons, we can use the regression analysis between calories and Weight Watcher Plus Points to predict the points value for items in Tim Hortons' menu provided that the calorie values of items in the Tim Hortons menu are within the same range as those from the other restaurants used in the prediction.

```
cat("The minimum calories from restaurants other than Tim Hortons is",
min(not_th$Calories), "\n and the maximum is", max(not_th$Calories))
```

```
## The minimum calories from restaurants other than Tim Hortons is 0
## and the maximum is 1660
```

```
cat("\nThe minimum calories from Tim Hortons is",
  min(th$Calories), "\n and the maximum is", max(th$Calories))
```

```
##  
## The minimum calories from Tim Hortons is 1  
## and the maximum is 680
```

Since the caloric value of items in Tim Hortons' menu is within the range of those used in the regression model, we can predict the Weight Watchers Plus Points for all items in Tim Hortons' menu.

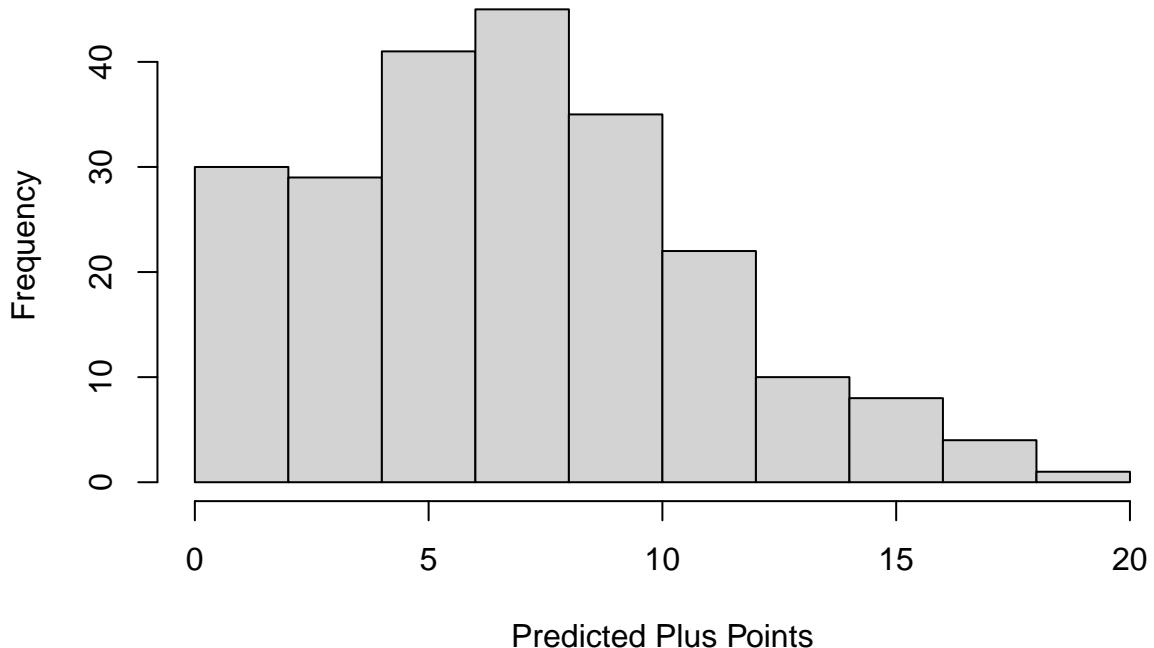
```
# predict the plus points
model_coeff = summary(reg)$coefficients[, "Estimate"]

# point estimate prediction
th$PointsPlus <- model_coeff[1] + model_coeff[2] * th$Calories

# predictions at 95% confidence interval
th_pp_pred = predict(reg, newdata = data.frame(Calories = th$Calories), interval = "confidence", level = 0.95)

# visualize the predicted Plus Points distribution
hist(th$PointsPlus, main='Distribution of Predicted Plus Points',
      xlab = "Predicted Plus Points")
```

Distribution of Predicted Plus Points



This distribution shows that most menu items are predicted to have low Plus Points (0-10), showing a majority of healthier options. As the Plus Points increase, the frequency of items significantly decreases, suggesting fewer high-calorie options.

References

1. <https://www.weightwatchers.com/ca/en/article/all-about-ww-freestyle>
2. <https://www.ctvnews.ca/health/the-difference-between-trans-fats-saturated-fats-and-unsaturated-fats-1.2426233?cache=piqndqvkh>
3. <https://www.statology.org/residual-plot-r/>
4. <https://www.geeksforgeeks.org/how-to-create-a-residual-plot-in-r/>