Data602GroupProject

Group

2024-10-11

Introduction to the Data

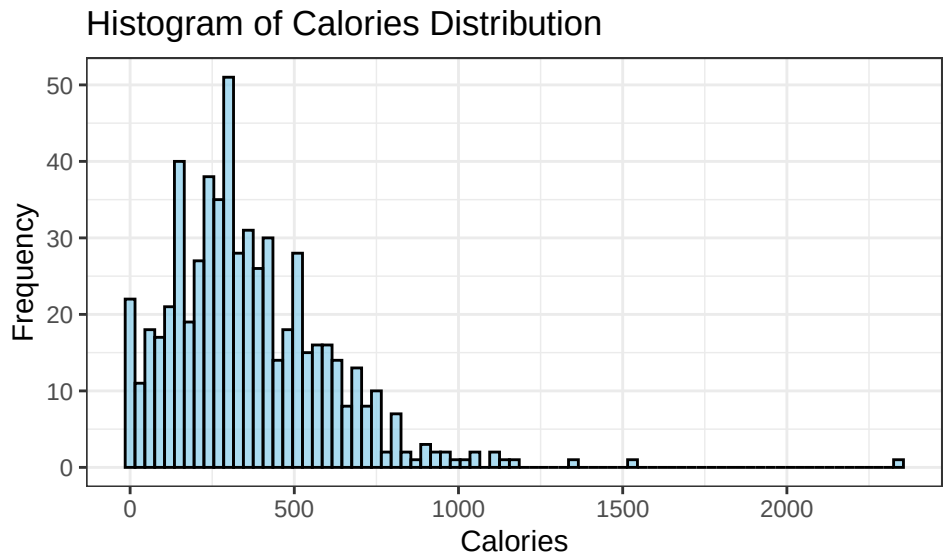
We intend to identify relationships between menu items of fast food restaurants, their nutritional information, and Weight Watchers Points. We initially got our data from Kaggle. However, we noticed that the Weight Watcher Points in this data were more than the actual Weight Watcher Points obtainable. We also noticed that the Weight Watcher Points were oddly similar to the value of the calories for each menu item.

To mitigate against 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, KFC, McDonald's, Subway, and Tim Hortons) from this site and webscraped their nutrition and weight watcher points information. We then updated our code to reflect this new data.

Exploratory Data Analysis

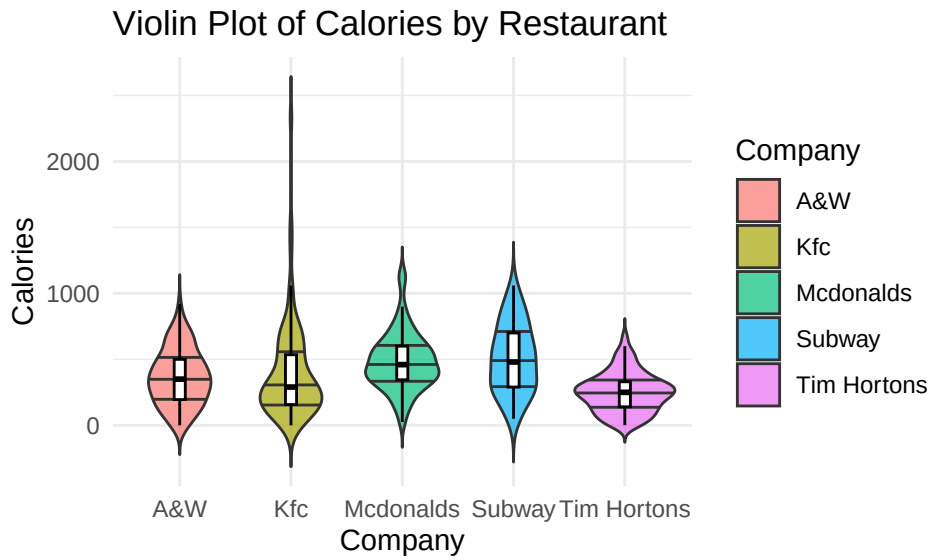
Graphics

```
fast_food_df = read.csv('FastFoodNutritionScraped20241011V2.csv')
hist_calories = ggplot(fast_food_df, aes(x = Calories)) +
geom_histogram(binwidth = 30, fill = "skyblue", color = "black", alpha = 0.7) +
labs(title = "Histogram of Calories Distribution",
x = "Calories",
y = "Frequency")
print(hist_calories)
```



The calories distribution is right-skewed.

```
violin_plot <- ggplot(fast_food_df, aes(x = Company, y = Calories, fill = Company)) +
geom_violin(trim = FALSE, draw_quantiles = c(0.25, 0.5, 0.75), alpha = 0.7) +
geom_boxplot(width = 0.1, fill = "white", color = "black", outlier.shape = NA) +
labs(title = "Violin Plot of Calories by Restaurant",
x = "Company",
y = "Calories",
fill = "Company") +
theme_minimal()
print(violin_plot)
```



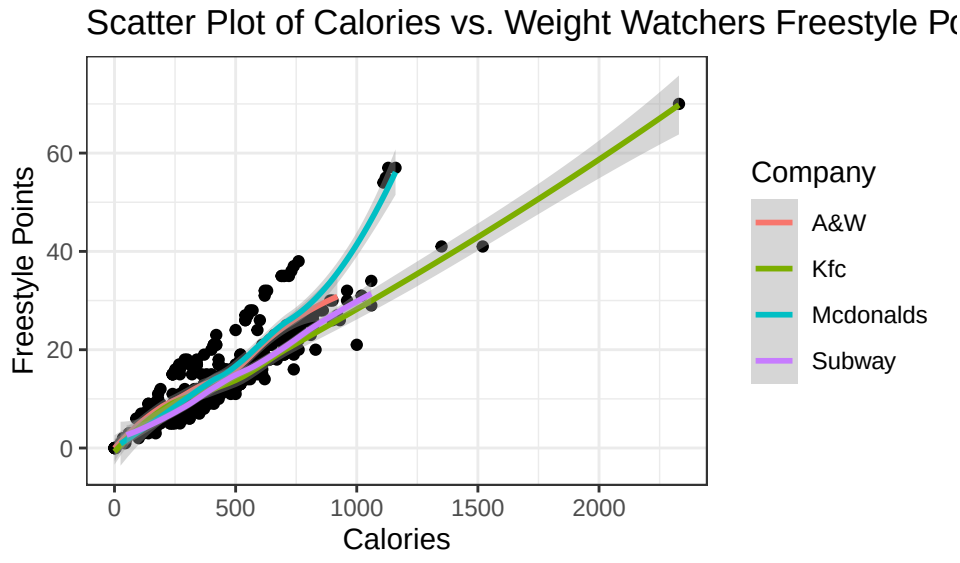
While the calories distribution for all listed restaurants are similar, McDonald's and Subway have slightly higher calories, on average, with Subway having more variability than McDonalds.

```
companies = unique(fast_food_df$Company)
for (company in companies) {
```

```
  cat("Favstats for", company, ":\n")
  stats = favstats(filter(fast_food_df, Company == company)$Calories, na.rm = TRUE)
  print(stats)
}
```

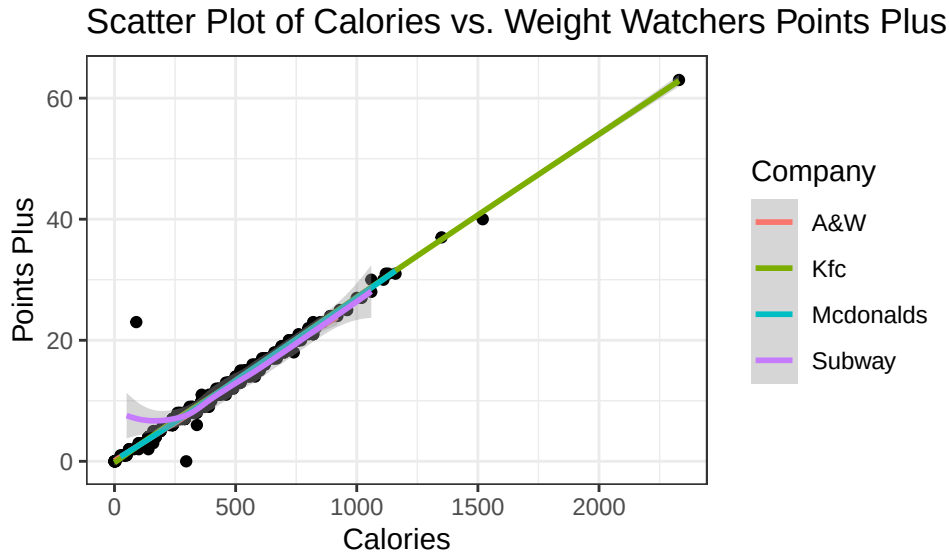
```
## Favstats for A&W :
##  min  Q1 median  Q3 max mean       sd   n missing
##    0 195    350 500 920  364 210.6246 111       0
## Favstats for Kfc :
##  min    Q1 median  Q3  max    mean       sd  n missing
##    0 157.5    290 535 2330 398.369 364.3555 84       0
## Favstats for Mcdonalds :
##  min  Q1 median  Q3  max     mean       sd   n missing
##   25 345    460 600 1160 482.6259 216.0493 139       0
## Favstats for Subway :
##  min  Q1 median  Q3  max     mean       sd  n missing
##   50 290    480 700 1060 508.2222 260.1511 45       0
## Favstats for Tim Hortons :
##  min  Q1 median  Q3 max  mean       sd   n missing
##    1 140    250 330 680 250.8 149.0066 225       0
```

```
scatter_plot = 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")
print(scatter_plot)
```



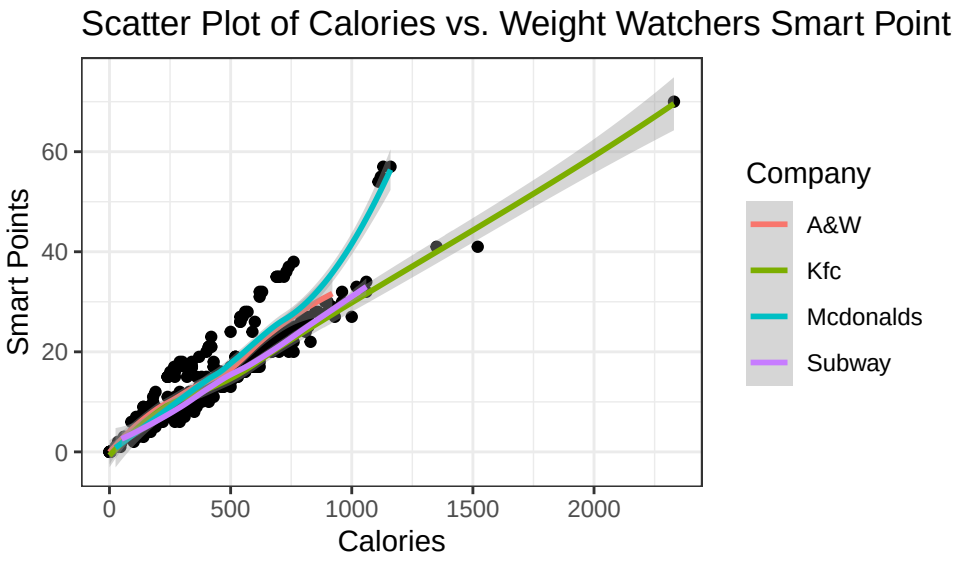
The calories are highly correlatable with the Freestyle Points but McDonald's trend deviate slightly from that of A&W, KFC, and Subway.

```
scatter_plot = 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 Points Plus",
x = "Calories",
y = "Points Plus",
color = "Company")
print(scatter_plot)
```



The calories are highly correlatable with the Plus Points and all listed rstaurants follow the same trend.

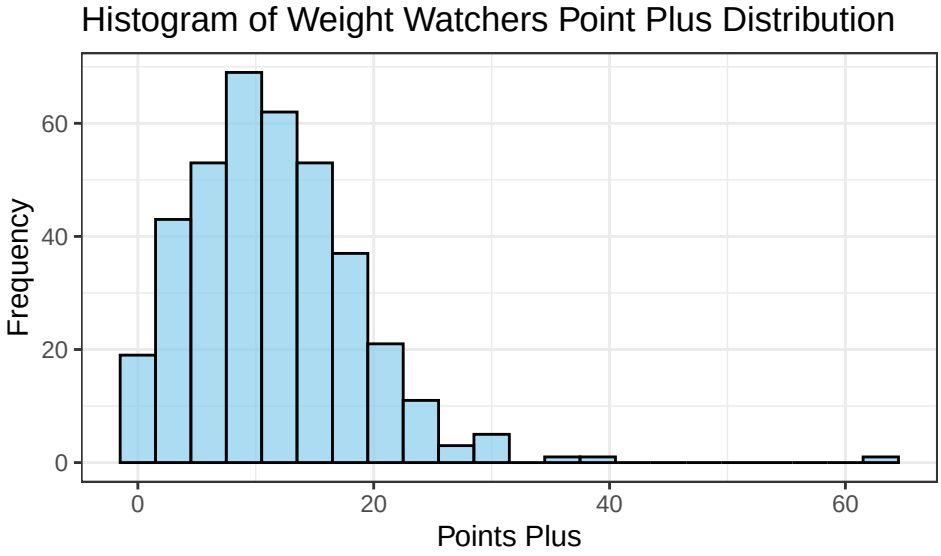
```
scatter_plot = ggplot(fast_food_df, mapping = aes(x = Calories, y = SmartPoints)) +
geom_point() +
  geom_smooth(mapping = aes(group = Company, color =
Company), show.legend = TRUE, method = 'loess', formula = 'y~x')  +
labs(title = "Scatter Plot of Calories vs. Weight Watchers Smart Points",
x =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



The calories are highly correlatable with the Smart Points but McDonald's trend deviate slightly from that of A&W, KFC, and Subway.

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

```
hist_pp = ggplot(fast_food_df, aes(x = PointsPlus)) +
geom_histogram(binwidth = 3, fill = "skyblue", color = "black", alpha = 0.7) +
labs(title = "Histogram of Weight Watchers Point Plus Distribution",
x = "Points Plus",
y = "Frequency")
print(hist_pp)
```



Since the Points Plus Distribution is left-skewed, we cannot assume normality. Rather, any confidence or prediction intervals will be created using bootstrapping methods.

```r
plus_points_summary = fast_food_df %>%
     group_by(Company) %>%
     summarize(
         minimum_pp  = min(`PointsPlus`, na.rm = TRUE),
         median_pp = median(`PointsPlus`, na.rm = TRUE),
         mean_pp = mean(`PointsPlus`, na.rm = TRUE),
         max_pp = max(`PointsPlus`, na.rm = TRUE)
     )
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 Kfc                  0         8   10.7      63
## 3 Mcdonalds            1        12   12.9      31
## 4 Subway               1        13   13.6      28
## 5 Tim Hortons        Inf        NA  NaN      -Inf
```

Research Question

Since we do not have the Weight Watcher Points of Tim Hortons, we can develop a regression analysis between calories and Weight Watcher Plus Points, and predict the points value for items in Tim Hortons' menu.