Wendy's ratings analysis

Danfei Li

2025-03-31

Import Data

```
setwd("/Volumes/PortableSSD/Programming/R/Statistical Programming")
data <- read.csv("data_wendys.csv")
str(data)
```

```
## 'data.frame':    1450 obs. of  17 variables:
##  $ responid            : chr  "3" "4" "5" "6" ...
##  $ age                 : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ gender              : int  2 1 2 2 2 2 2 2 2 2 ...
##  $ edu                 : int  3 3 3 4 2 3 3 3 3 3 ...
##  $ income              : int  1 1 1 8 2 1 8 2 6 3 ...
##  $ marital             : int  2 1 2 1 1 1 1 1 1 1 ...
##  $ consumption_rate    : int  1 4 10 15 3 20 28 25 2 5 ...
##  $ patronage_arbys     : int  3 2 1 4 2 1 3 1 3 3 ...
##  $ patronage_burgerking: int  3 1 1 1 2 1 1 1 3 3 ...
##  $ patronage_mcd       : int  3 2 3 1 2 1 2 1 3 1 ...
##  $ patronage_subway    : int  2 1 1 1 1 1 1 1 3 1 ...
##  $ patronage_wendys    : int  3 2 1 3 1 1 1 1 3 3 ...
##  $ rating_arbys        : int  NA 7 5 NA 8 8 NA 3 NA NA ...
##  $ rating_burgerking   : int  NA 1 5 6 7 7 8 8 NA NA ...
##  $ rating_mcd          : int  NA 3 NA 7 7 7 2 4 NA 6 ...
##  $ rating_subway       : int  7 5 6 4 9 9 9 7 NA 8 ...
##  $ rating_wendys       : int  NA 1 5 NA 8 8 8 5 NA NA ...
```

Data Cleaning (remove missing ratings)

```
data.clean <- data[which(!is.na(data$rating_wendys) &
                         !is.na(data$rating_mcd) &
                         !is.na(data$rating_subway) &
                         !is.na(data$rating_burgerking)), ]
str(data.clean)
```

```
## 'data.frame':    449 obs. of  17 variables:
##  $ responid            : chr  "4" "7" "8" "10" ...
##  $ age                 : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ gender              : int  1 2 2 2 2 2 2 2 2 2 ...
##  $ edu                 : int  3 2 3 3 3 3 3 3 4 3 ...
##  $ income              : int  1 2 1 8 2 2 1 1 3 2 ...
##  $ marital             : int  1 1 1 1 1 2 1 2 1 1 ...
##  $ consumption_rate    : int  4 3 20 28 25 10 30 5 10 6 ...
##  $ patronage_arbys     : int  2 2 1 3 1 3 2 2 2 1 ...
##  $ patronage_burgerking: int  1 2 1 1 1 2 1 1 2 1 ...
##  $ patronage_mcd       : int  2 2 1 2 1 1 1 1 1 1 ...
##  $ patronage_subway    : int  1 1 1 1 1 2 2 1 2 1 ...
##  $ patronage_wendys    : int  2 1 1 1 1 2 1 2 2 2 ...
##  $ rating_arbys        : int  7 8 8 NA 3 NA 10 10 8 8 ...
##  $ rating_burgerking   : int  1 7 7 8 8 9 2 5 7 7 ...
##  $ rating_mcd          : int  3 7 7 2 4 7 9 5 9 9 ...
##  $ rating_subway       : int  5 9 9 9 7 8 9 6 7 9 ...
##  $ rating_wendys       : int  1 8 8 8 5 8 10 8 9 8 ...
```

Comparative analysis of customer ratings

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

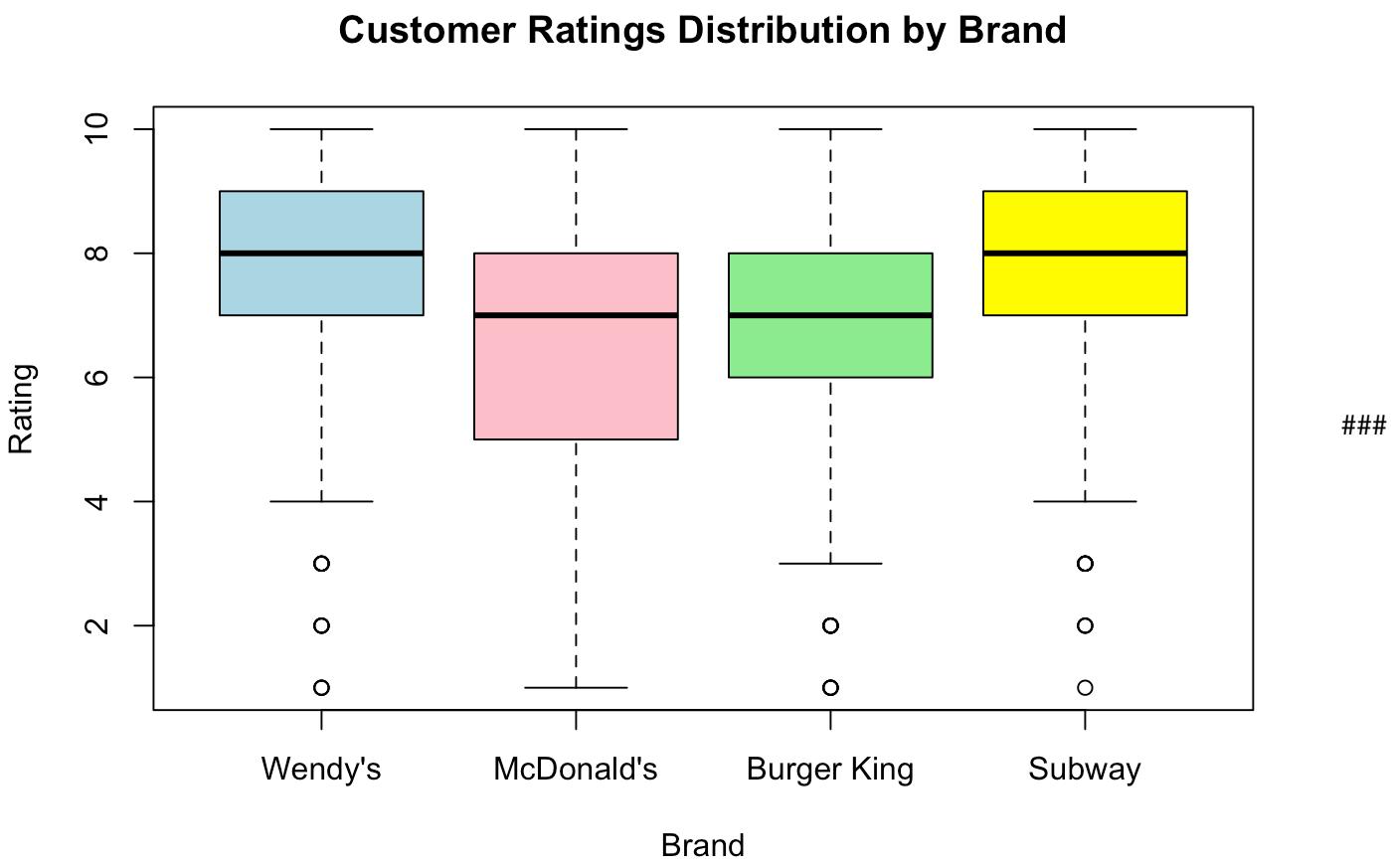
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
stats_table <- data.clean %>%
  summarise(
    Mean_Wendys = mean(rating_wendys),
    Median_Wendys = median(rating_wendys),
    SD_Wendys = sd(rating_wendys),
    IQR_Wendys = IQR(rating_wendys),

    Mean_McDonalds = mean(rating_mcd),
    Median_McDonalds = median(rating_mcd),
    SD_McDonalds = sd(rating_mcd),
    IQR_McDonalds = IQR(rating_mcd),

    Mean_BurgerKing = mean(rating_burgerking),
    Median_BurgerKing = median(rating_burgerking),
    SD_BurgerKing = sd(rating_burgerking),
    IQR_BurgerKing = IQR(rating_burgerking),

    Mean_Subway = mean(rating_subway),
    Median_Subway = median(rating_subway),
    SD_Subway = sd(rating_subway),
    IQR_Subway = IQR(rating_subway)
  ) %>%
  pivot_longer(everything(), names_to = c("Metric", "Brand"), names_sep = "_") %>%
  pivot_wider(names_from = "Brand", values_from = "value") %>%
  mutate(across(where(is.numeric), ~ round(., 2)))
print(stats_table)
```

```
## # A tibble: 4 × 5
##   Metric Wendys McDonalds BurgerKing Subway
##   <chr>   <dbl>     <dbl>      <dbl>  <dbl>
## 1 Mean     7.58      6.72       6.85   7.84
## 2 Median   8         7          7      8
## 3 SD       1.73      2.03       1.86   1.72
## 4 IQR      2         3          2      2
```

```
# Create boxplot visuals
boxplot(data.clean$rating_wendys, data.clean$rating_mcd, data.clean$rating_burgerking, d
ata.clean$rating_subway,
        names = c("Wendy's", "McDonald's", "Burger King", "Subway"),
        main = "Customer Ratings Distribution by Brand",
        xlab = "Brand",
        ylab = "Rating",
        col = c("lightblue", "pink", "lightgreen", "yellow"))
```



###

Key findings: 1. Consistency in scores: Wendy's has a SD of 1.73, indicating relatively consistent scores compared to McDonald's and Burger King. Subway has the lowest SD (1.72), suggesting that Wendy's is fairly consistent while slightly less than Subway. 2. Mean score: Wendy's performs well overall in customer satification with a mean score of 7.58, trailing only Subway with a mean score of 7.84. 3. Median score: the typical middle score of Wendy's is at the top level among its competitors, with a median score of 8 which is higher than McDonald's and Burger King, and on par with Subway. 4. From the boxplot, Wendy's and Subway both have smaller IQRs compared to McDonald's and Burger Kin, further supporting the consistent customer experience delivered by those two brands.

Correlation and brand positioning

```
rating_com <- data.frame(
    Wendys = data.clean$rating_wendys,
    McDonalds = data.clean$rating_mcd,
    BurgerKing = data.clean$rating_burgerking,
    Subway = data.clean$rating_subway
)
matrix <- cor(rating_com)
matrix <- round(matrix, 2)
matrix
```

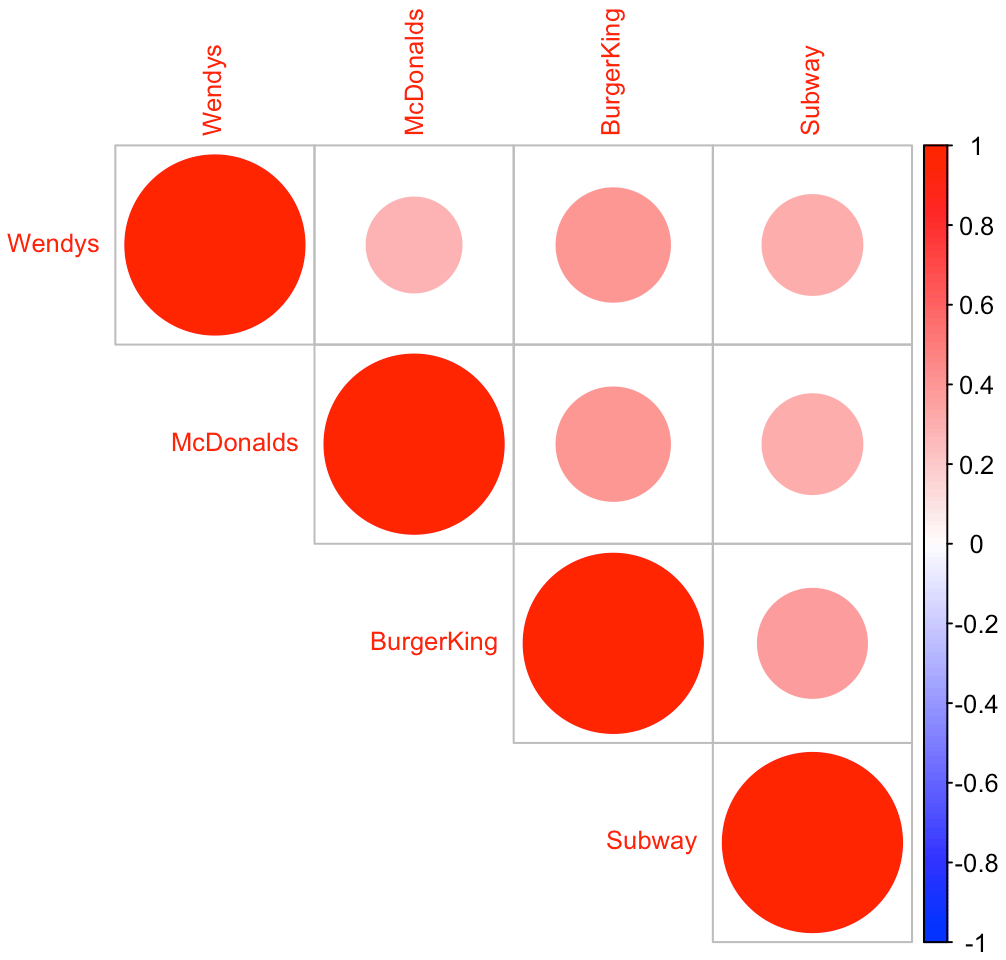
```
##            Wendys McDonalds BurgerKing Subway
## Wendys       1.00      0.28       0.40   0.31
## McDonalds    0.28      1.00       0.40   0.31
## BurgerKing   0.40      0.40       1.00   0.37
## Subway       0.31      0.31       0.37   1.00
```

```
# visualize the correlation
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.95 loaded
```

```
corrplot(matrix,
         method = "circle",        # Display numbers instead of colors/shapes
         type = "upper",           # Show only the upper triangle
         tl.cex = 0.8,             # Adjust text label size
         number.cex = 0.8,         # Adjust number size
         col = colorRampPalette(c("blue", "white", "red"))(200)) # Use a color gradient
```




```
cor.test(data.clean$rating_wendys, data.clean$rating_burgerking)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data.clean$rating_wendys and data.clean$rating_burgerking
## t = 9.2274, df = 447, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3192796 0.4749622
## sample estimates:
##       cor
## 0.4000023
```

```
cor.test(data.clean$rating_wendys, data.clean$rating_mcd)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data.clean$rating_wendys and data.clean$rating_mcd
## t = 6.2608, df = 447, p-value = 8.996e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1965599 0.3668389
## sample estimates:
##       cor
## 0.2839365
```

Key findings:

1. Among all brands, BurgerKing's ratings are most strongly correlated with that of Wendy's, and McDonalds' ratings are most weakly correlated with Wendy's, which suggests that Wendy's compete more directly with BurgerKing.
2. The p-value is extremely less than 0.5, suggesting that the positive correlation (0.4) between Wendy's and Burger King is significant in the population, which further supports the above finding.

Wendy's position relative to Burger King

```
diff <- data.clean$rating_wendys - data.clean$rating_burgerking
summary(diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -7.000   0.000   0.000   0.735   2.000   8.000
```

```
# calculate proportions of negative, positive, and zero values in the diff
prop_positive <- sum(diff > 0) / length(diff)
prop_negative <- sum(diff < 0) / length(diff)
prop_zero <- sum(diff == 0) / length(diff)
round(prop_positive, 2) # Wendy’s is rated higher than McDonald’s in 51% of cases (Most customers rate Wendy’s better than McDonald’s)
```

```
## [1] 0.48
```

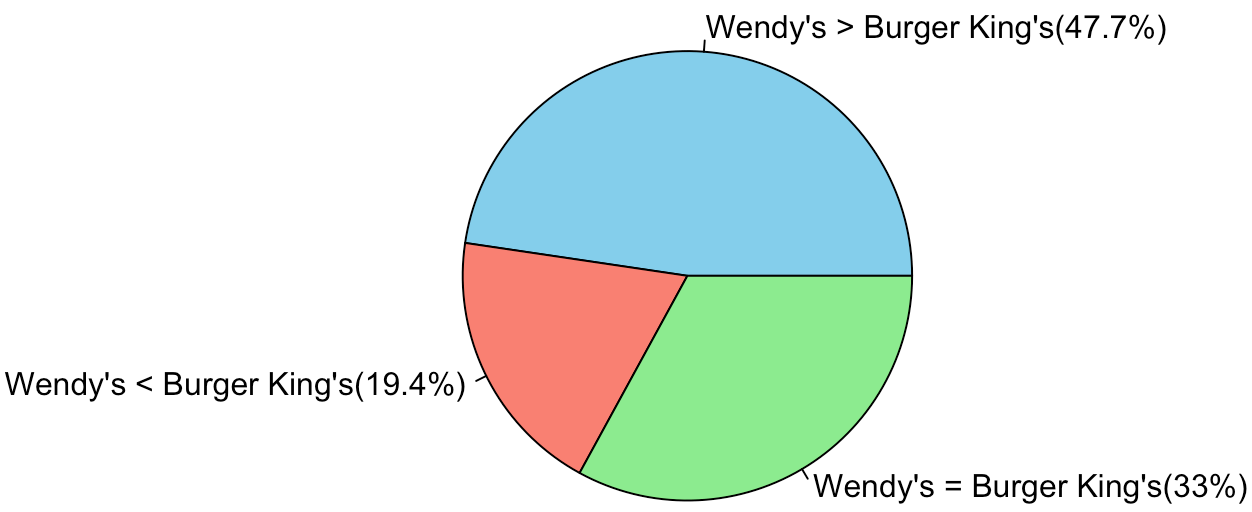
```
round(prop_negative, 2) # McDonald’s is rated higher than Wendy’s in 20% of cases
```

```
## [1] 0.19
```

```
round(prop_zero, 2)
```

```
## [1] 0.33
```

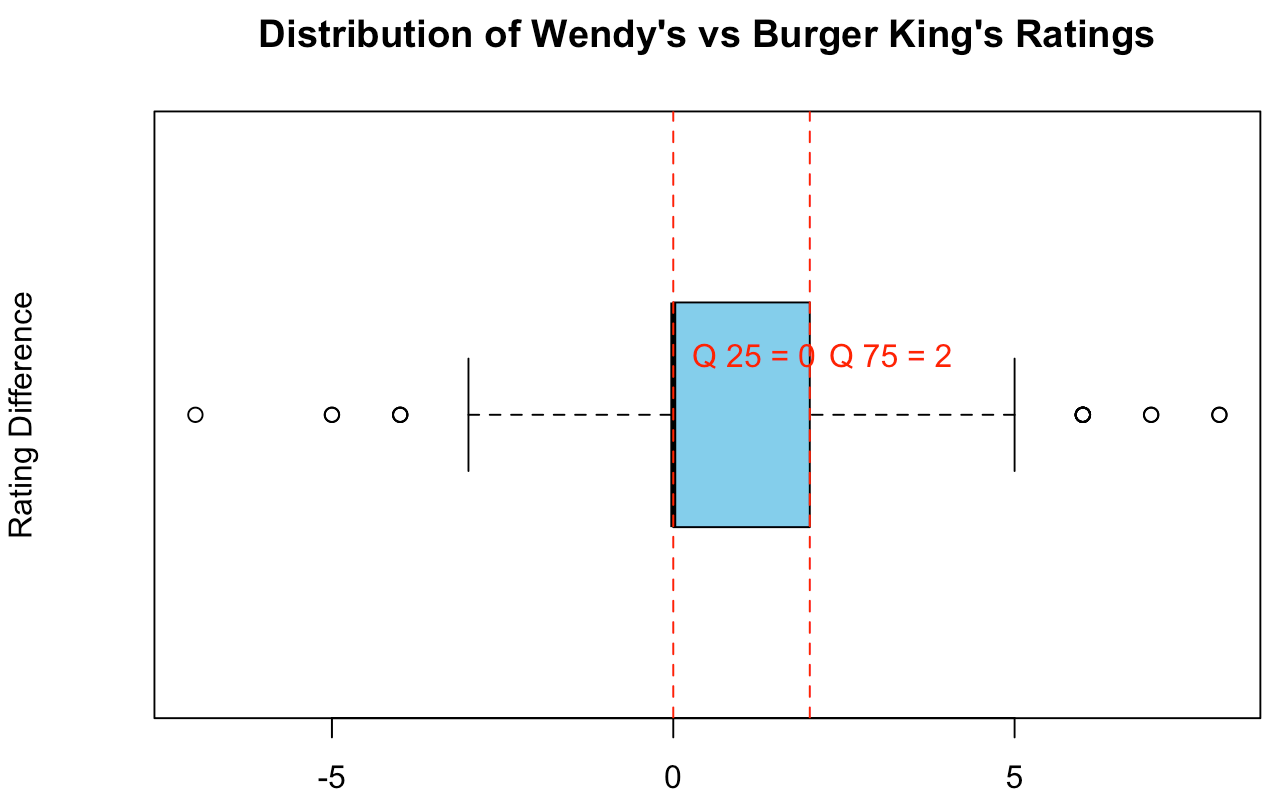
```
# Visualize above result using pie chart
proportions <- c(prop_positive, prop_negative, prop_zero)
labels <- c("Wendy's > Burger King's", "Wendy's < Burger King's", "Wendy's = Burger King's")
labels_with_values <- paste(labels, "(", round(proportions * 100, 1), "%)", sep = "")
pie(proportions, labels = labels_with_values, col = c("skyblue", "salmon", "lightgreen"), main = "Comparison of Wendy's and Burger King's Ratings")
```




```
# calculate the range of typical difference
quantile(diff, c(0.25, 0.75))
```

```
## 25% 75%
##   0   2
```

```
# visualize above result using boxplot
boxplot(diff, main = "Distribution of Wendy's vs Burger King's Ratings",
        ylab = "Rating Difference", col = "skyblue", horizontal = TRUE)
quantiles <- quantile(diff, c(0.25, 0.75))
abline(v = quantiles, col = "red", lty = 2)  # Add vertical dashed lines for the quantil
es
text(quantiles, rep(1.1, 2), labels = paste("Q", c("25", "75"), "=", round(quantiles,
2)), col = "red", pos = 4)
```




```
# conduct t-test
### H0: the mean rating of Wendy's is same as Burger King
### HA: the mean ratings of Wendy's and Burger King are different
t.test(data.clean$rating_wendys, data.clean$rating_burgerking, alternative = "two.side
d")
```

```
##
##  Welch Two Sample t-test
##
## data:  data.clean$rating_wendys and data.clean$rating_burgerking
## t = 6.1351, df = 890.8, p-value = 1.279e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4998484 0.9700848
## sample estimates:
## mean of x mean of y
##  7.583519  6.848552
```

```
#### the significant result (p-value < 0.05) means Wendy's is clearly favored in custome
r ratings compared to Burger King.
```