Fueling Your Workout: Fast-Food Chain Menu Analysis for Fitness Enthusiasts

Final Project for Stat 184

Honey Patel, Shailee Patel

Introduction

Maintaining adequate protein intake is critical for fitness goals, including muscle repair and growth. However, when pressed for time, many turn to fast-food options, which vary widely in protein content. This report explores how popular fast-food chains compare in terms of protein, fat, and calories—guiding fitness-minded consumers toward better choices.

Data Provenance

Our analysis draws on two publicly available datasets:

1. **fastfood.csv**:Nutritional data (calories, carbohydrates, fats, protein) for menu items from major U.S. fast-food chains, sourced from Fast Food Nutrition Dataset.
2. **Nutrition_Value_Dataset.csv**: Additional fast-food menu data, including newer items and broader restaurant coverage, sourced from Fast Food Joint Nutrition Values Dataset.

These datasets were merged after cleaning to create a comprehensive database for analysis. The cases in our data represent individual menu items across various restaurant chains.

FAIR and CARE Principles

Our project follows the FAIR principles by:

- **Findable**: We used publicly available datasets with clear metadata.
- **Accessible**: Files are hosted in open formats such as CSV without any access restrictions.
- **Interoperable**: All data columns follow standard nutritional labeling (i.e. Calories, Fat, Protein).
- **Reusable**: Our cleaning steps are documented and allow for others to reuse the combined, cleaned dataset.

We have also made sure that the CARE principles are ensured. The data used in our project is solely for educational purposes with attention given to responsibly interpreting its context. In presenting our findings, we have made a point to avoid misrepresenting the fast-food chains involved as we recognize the importance of maintaining ethical standards when working with publicly available data.

Focused and Derived Attributes

We concentrate on three key measures per menu item:

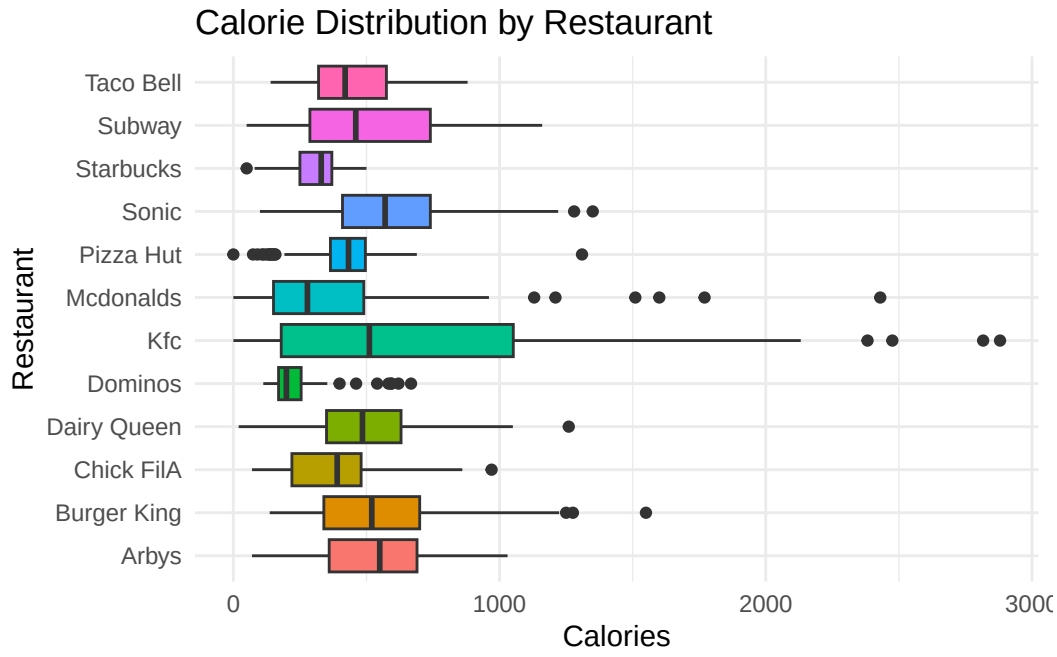
- **Calories (kCal)**: Total energy per item (original attribute from dataset).
- **Fat (g)**: Total fat content per item (original attribute).
- **Protein (g)**: Total protein content per item (original attribute).
- **Protein-to-Fat Ratio (PF Ratio)**: A newly derived attribute, calculated as Protein divided by Fat which is used to identify lean protein options.

These attributes allow us to assess the nutritional quality of menu items and make recommendations suited to fitness-focused consumers.

Analysis and Findings

1. Calorie Distribution by Restaurant

A box plot (Figure 1) shows the **spread, median, and outliers** of calorie counts across each chain. This highlights variation—for example, KFC has wider calorie ranges due to high-calorie menu items.



2. Average Calories, Fat, and Protein

Table 1 summarizes **per-chain averages**. Bar charts (Figures 2–4) compare each metric:

- Figure 2: Avg Calories
- Figure 3: Avg Fat (g)
- Figure 4: Avg Protein (g)

Fitness-focused readers can see that Chick Fil-A and KFC rank high in average protein, while Starbucks and Domino's offer lower calorie and fat options.

Figure 2. Average Calories by Restaurant

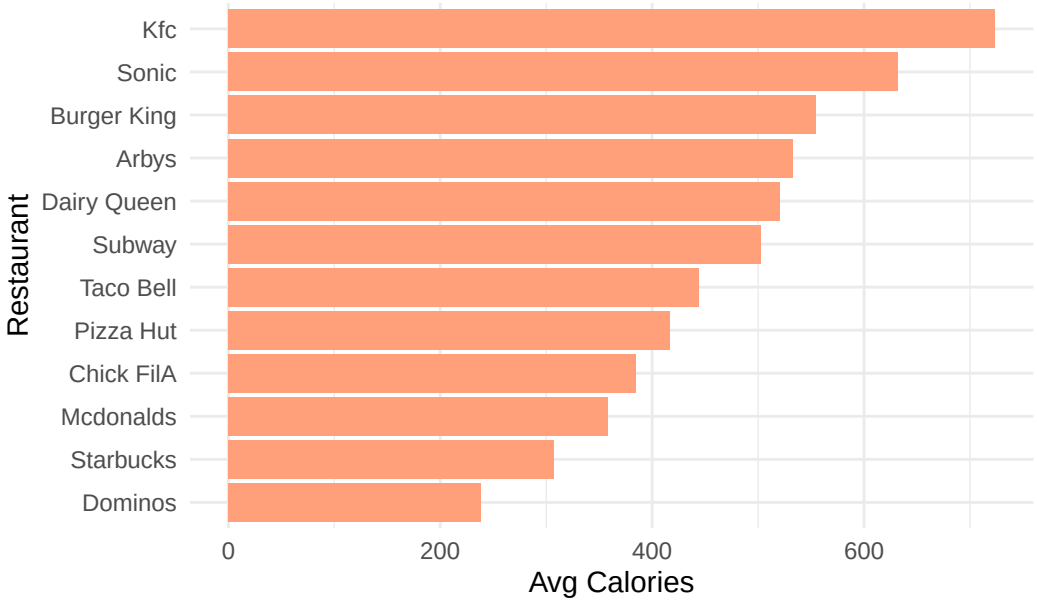


Figure 3. Average Fat (g) by Restaurant

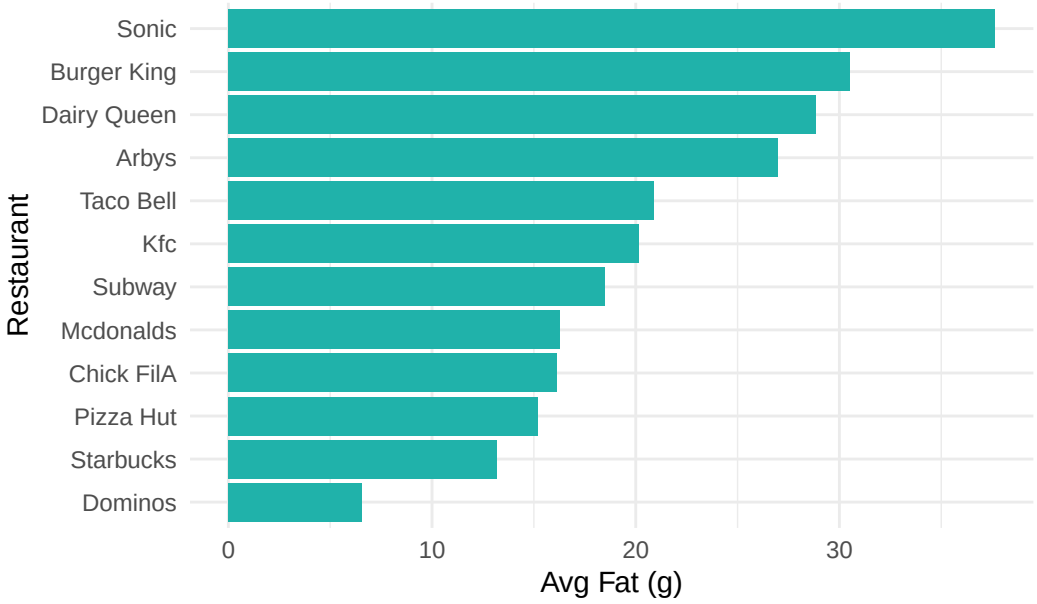


Figure 4. Average Protein (g) by Restaurant

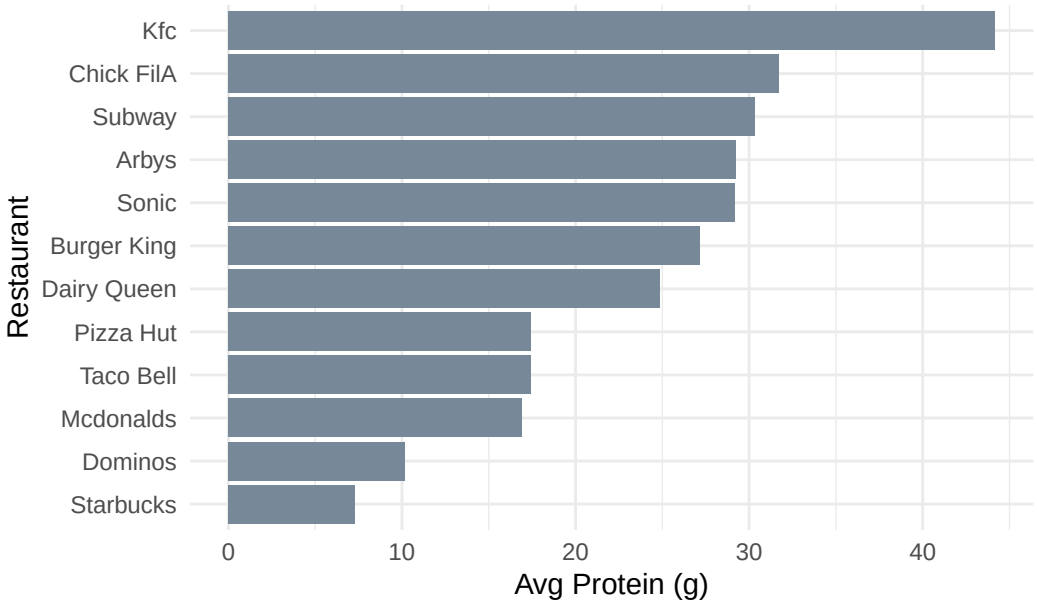


Table 1 shows each bars' **exact** values:

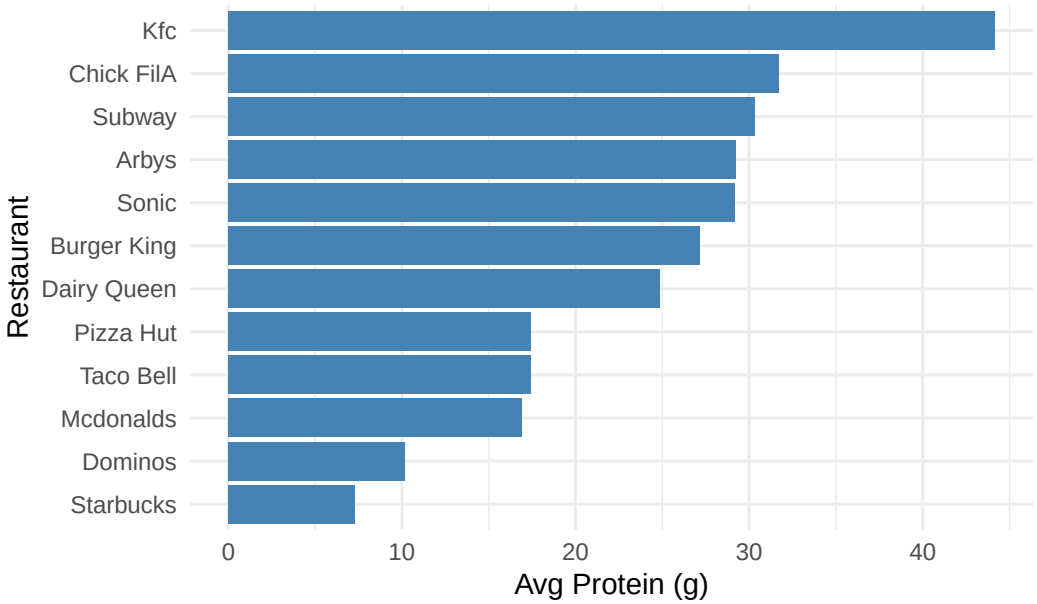
Table 1: Table 1: Average Calories, Fat, and Protein by Restaurant

Restaurant	Avg Calories	Avg Fat (g)	Avg Protein (g)
Arbys	532.7	27.0	29.3
Burger King	554.2	30.5	NA
Chick FilA	384.4	16.1	31.7
Dairy Queen	520.2	28.9	24.8
Dominos	238.5	6.5	10.1
Kfc	723.6	20.1	44.2
Mcdonalds	358.1	16.3	16.9
Pizza Hut	416.8	15.2	17.4
Sonic	631.7	37.6	29.2
Starbucks	307.4	13.2	7.3
Subway	503.0	18.5	30.3
Taco Bell	443.7	20.9	17.4

3. Average Protein per Menu Item by Restaurant

A bar chart (Figure 5) displays the **average protein content** per menu item at each restaurant. This helps fitness-focused consumers identify chains that consistently offer more protein-rich choices. For example, restaurants like KFC and Chick-fil-A rank among the highest for average protein per item, while Starbucks and Dunkin' offer significantly lower averages.

Figure 5. Average Protein per Menu Item by Restaurant



4. Ranking Protein-Fat Ratios Per Restaurants

Table 2 summarizes the **top three** menu items from each restaurant based on their protein-to-fat ratio.
This helps fitness-focused consumers quickly identify lean menu choices that offer the most protein with the least fat.

Table 2: Table 2: Top 3 Menu Items per Restaurant by Protein-to-Fat Ratio

Restaurant	Product	Calories	Protein	Fat	PF_Ratio
Arbys	Ham & Swiss Melt	300.00	18.00	9.00	2.00
Arbys	Turkey 'n Cheese Slider	200.00	14.00	7.00	2.00
Arbys	Arby-Q Sandwich	400.00	18.00	10.00	1.80
Burger King	Chicken wings Grilled (4Pcs)	335.00	27.70	8.80	3.15
Burger King	Chicken wings Grilled (15Pcs)	1275.50	105.40	33.60	3.14
Burger King	Chicken wings Grilled (2Pcs)	167.50	13.80	4.40	3.14
Chick FilA	12 Piece Grilled Chicken Nuggets	210.00	38.00	5.00	7.60
Chick FilA	4 Piece Grilled Chicken Nuggets	70.00	13.00	2.00	6.50
Chick FilA	6 Piece Grilled Chicken Nuggets	110.00	19.00	3.00	6.33
Dairy Queen	Grilled Chicken Garden Greens Salad	150.00	23.00	2.00	11.50
Dairy Queen	Grilled Chicken BLT Salad	380.00	42.00	19.00	2.21
Dairy Queen	Barbecue Pork Sandwich	310.00	17.00	9.00	1.89
Dominos	CHICKEN GOLDEN DELIGHT MEDIUM	185.20	8.20	1.50	5.47
Dominos	CHICKEN GOLDEN DELIGHT LARGE	225.10	9.90	1.90	5.21
Dominos	VEG EXTRAVAGANZA LARGE	242.60	12.30	3.30	3.73
Kfc	8 pc Bucket	1398.80	170.50	17.00	10.03

Kfc	5 PC	874.30	106.60	16.00	6.66
Kfc	6 Pc	1049.10	127.90	19.20	6.66
Mcdonalds	Mixed Fruit Beverage	72.25	0.65	0.02	32.50
Mcdonalds	Americano (R)	23.07	0.94	0.05	18.80
Mcdonalds	Americano (L)	26.71	1.09	0.06	18.17
Pizza Hut	Country Feast (Medium)	542.71	22.28	10.69	2.08
Pizza Hut	Country Feast (Personal)	407.60	16.73	8.03	2.08
Pizza Hut	Country Feast (Medium)	481.67	21.03	11.00	1.91
Sonic	3 Piece Crispy Chicken Tender Dinner	280.00	22.00	14.00	1.57
Sonic	5 Piece Crispy Chicken Tender Dinner	470.00	37.00	24.00	1.54
Sonic	Grilled Chicken Sandwich	430.00	28.00	20.00	1.40
Starbucks	Everything with Cheese Bagel	270.00	10.00	2.00	5.00
Starbucks	Multigrain Bagel	290.00	14.00	3.00	4.67
Starbucks	Dried Fruit	100.00	1.00	0.40	2.50
Subway	Double Chicken Salad	220.00	36.00	5.00	7.20
Subway	Sweet Onion Chicken Teriyaki Salad	200.00	20.00	3.00	6.67
Subway	Oven Roasted Chicken Salad	140.00	19.00	3.00	6.33
Taco Bell	Fresco Chicken Soft Taco	150.00	11.00	4.00	2.75
Taco Bell	Fresco Grilled Steak Soft Taco	150.00	9.00	4.00	2.25
Taco Bell	Fresco Burrito Supreme® – Chicken	340.00	17.00	8.00	2.12

Table 3 then highlights the **single best** menu item for each restaurant based on the highest protein-to-fat ratio and then ranks it compared to other restaurants. This provides clear recommendations for consumers seeking maximum protein with minimal fat intake.

Table 3: Table 3: Single-Best Menu Item per Restaurant by Protein-to-Fat Ratio Ranked

Restaurant	Product	Calories	Protein	Fat	PF_Ratio
Mcdonalds	Mixed Fruit Beverage	72.25	0.65	0.02	32.50
Dairy Queen	Grilled Chicken Garden Greens Salad	150.00	23.00	2.00	11.50
Kfc	8 pc Bucket	1398.80	170.50	17.00	10.03
Chick FilA	12 Piece Grilled Chicken Nuggets	210.00	38.00	5.00	7.60
Subway	Double Chicken Salad	220.00	36.00	5.00	7.20
Dominos	CHICKEN GOLDEN DELIGHT MEDIUM	185.20	8.20	1.50	5.47
Starbucks	Everything with Cheese Bagel	270.00	10.00	2.00	5.00
Burger King	Chicken wings Grilled (4Pcs)	335.00	27.70	8.80	3.15
Taco Bell	Fresco Chicken Soft Taco	150.00	11.00	4.00	2.75
Pizza Hut	Country Feast (Medium)	542.71	22.28	10.69	2.08
Arbys	Ham & Swiss Melt	300.00	18.00	9.00	2.00
Sonic	3 Piece Crispy Chicken Tender Dinner	280.00	22.00	14.00	1.57

Code Appendix

```
library(tidyverse)
library(kableExtra)

#reload data to prevent errors
data <- read_csv("https://raw.githubusercontent.com/Stat184-Spring2025/Stat184-Spring2025-Sec3_
library(tidyverse)

#1. Calorie Distribution by Restaurant
ggplot(data, aes(x = Restaurant, y = Calories, fill = Restaurant)) +
  geom_boxplot() +
  coord_flip() +                        #Horizontal boxes for readability
  theme_minimal() +
  theme(
    legend.position = "none",           #Drop the legend
    axis.title.x = element_text(),
    axis.title.y = element_text()
  ) +
  labs(
    x = "Restaurant",
    y = "Calories",
    title = "Calorie Distribution by Restaurant"
  )
library(tidyverse)

#2. Bar chart: Average Calories
restaurant_stats <- data %>%
  group_by(Restaurant) %>%
  summarise(
    avg_calories = mean(Calories, na.rm = TRUE),
    avg_fat      = mean(Fat, na.rm = TRUE),
    avg_protein  = mean(Protein, na.rm = TRUE)
  )

ggplot(restaurant_stats, aes(
    x = avg_calories,
    y = reorder(Restaurant, avg_calories)
  )) +
  geom_col(fill = "#FFA07A") +
  labs(
    title = "Figure 2. Average Calories by Restaurant",
    x     = "Avg Calories",
    y     = "Restaurant"
  ) +
  theme_minimal()
```

```r
#3. Bar chart: Average Fat
ggplot(restaurant_stats, aes(
    x = avg_fat,
    y = reorder(Restaurant, avg_fat)
  )) +
  geom_col(fill = "#20B2AA") +
  labs(
    title = "Figure 3. Average Fat (g) by Restaurant",
    x     = "Avg Fat (g)",
    y     = "Restaurant"
  ) +
  theme_minimal()

#4. Bar chart: Average Protein
restaurant_stats %>%
  filter(!is.na(avg_protein)) %>%
  ggplot(aes(
    x = avg_protein,
    y = reorder(Restaurant, avg_protein)
  )) +
  geom_col(fill = "#778899") +
  labs(
    title = "Figure 4. Average Protein (g) by Restaurant",
    x = "Avg Protein (g)",
    y = "Restaurant"
  ) +
  theme_minimal()
#Question 2 ------------------------------------------------------------
#Do different restaurants differ in how healthy their items are (based on calories, fat, and pr
#Goal: Compare central tendencies of Calories, Fat, and Protein across restaurants to assess re
#Selection of visualization: Summary table + grouped bar chart
#Rationale:
# - Table gives exact average values for each metric
# - Grouped bar chart lets you visually compare multiple metrics side-by-side

#1. Compute per-restaurant averages
restaurant_stats <- data %>%
  group_by(Restaurant) %>%
  summarise(
    avg_calories = mean(Calories),
    avg_fat      = mean(Fat),
    avg_protein  = mean(Protein)
  )
restaurant_stats %>%
  kable(
    digits    = 1,
    booktabs  = TRUE,
```

```
    col.names = c("Restaurant", "Avg Calories", "Avg Fat (g)", "Avg Protein (g)"),
    caption   = "Table 1: Average Calories, Fat, and Protein by Restaurant"
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),  # adds zebra stripes & fixes float
    full_width    = FALSE,
    position      = "center"
  )
library(tidyverse)

#3.Average Protein per Menu Item by Restaurant
protein_stats <- data %>%
  group_by(Restaurant) %>%
  summarize(Avg_Protein = mean(Protein, na.rm = TRUE)) %>% #group by restaurant and calulate av
  mutate(Avg_Protein = ifelse(is.na(Avg_Protein), 0, Avg_Protein)) #replace missing values with
ggplot(protein_stats, aes(x = reorder(Restaurant, Avg_Protein), y = Avg_Protein)) +
  geom_col(fill = "#4682B4") +
  coord_flip() +
  labs(
    title = "Figure 5. Average Protein per Menu Item by Restaurant",
    x = "Restaurant",
    y = "Avg Protein (g)"
  ) +
  theme_minimal()
library(tidyverse)
library(kableExtra)
library(dplyr)

#4. Ranking Protein-Fat Ratios Per Restaurants
data <- data %>% #Compute protein-fat ratio
  mutate(PF_Ratio = Protein / Fat)

#top three filter
invisible(
  top3_pf_items <- data %>%
    filter(Fat > 0, !is.na(Protein), !is.na(Fat)) %>%
    group_by(Restaurant) %>%
    slice_max(order_by = PF_Ratio, n = 3, with_ties = FALSE) %>%
    ungroup()
)

top3_pf_items %>% #display the results as a table
  select(Restaurant, Product, Calories, Protein, Fat, PF_Ratio) %>%
  arrange(Restaurant, desc(PF_Ratio)) %>%
  kable(
    digits = 2,
    booktabs = TRUE,
```

```
    caption = "Table 2: Top 3 Menu Items per Restaurant by Protein-to-Fat Ratio"
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE,
    position = "center"
  )
library(tidyverse)
library(kableExtra)
library(dplyr)

data <- data %>%
  mutate(PF_Ratio = Protein / Fat)

#single-best filter
invisible(
  best_pf_item <- data %>%
    filter(Fat > 0, !is.na(Protein), !is.na(Fat)) %>%
    group_by(Restaurant) %>%
    slice_max(order_by = PF_Ratio, n = 1, with_ties = FALSE) %>%
    ungroup()
)

best_pf_item %>% #display the results as a table
  select(Restaurant, Product, Calories, Protein, Fat, PF_Ratio) %>%
  arrange(desc(PF_Ratio)) %>%
  kable(
    digits = 2,
    booktabs = TRUE,
    caption = "Table 3: Single-Best Menu Item per Restaurant by Protein-to-Fat Ratio Ranked"
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE,
    position = "center"
  )
#Load required library
library(tidyverse)

#Read the original CSVs
fastfood <- read_csv("https://raw.githubusercontent.com/Stat184-Spring2025/Stat184-Spring2025-S
nutrition <- read_csv("https://raw.githubusercontent.com/Stat184-Spring2025/Stat184-Spring2025-

#Clean the fast food CSV
fastfood_clean <- fastfood %>%
  select(
    Restaurant = restaurant,
```

```
    Product = item,
    Calories = calories,
    Carbs = total_carb,
    Fat = total_fat,
    Cholesterol = cholesterol,
    Protein = protein
  ) %>%
  mutate(
    Restaurant = str_to_title(Restaurant),
    Restaurant = str_replace_all(Restaurant, "[^[:alnum:] ]", ""),
    Restaurant = str_squish(Restaurant)
  )

#Clean the second nutrition CSV
nutrition_clean <- nutrition %>%
  select(
    Restaurant = Company,
    Product = Product,
    Calories = `Energy (kCal)`,
    Carbs = `Carbohydrates (g)`,
    Fat = `Total Fat (g)`,
    Cholesterol = `Cholesterol (mg)`,
    Protein = `Protein (g)`
  ) %>%
  mutate(
    Restaurant = str_to_title(Restaurant),
    Restaurant = str_replace_all(Restaurant, "[^[:alnum:] ]", ""),
    Restaurant = str_squish(Restaurant)
  )

#Combine the cleaned datasets
combined_nutrition <- bind_rows(fastfood_clean, nutrition_clean)

#Save to a new CSV file
write_csv(combined_nutrition, "combined_nutrition_data.csv")

#Load require libraries
library(tidyverse)
library(RColorBrewer)
library(knitr)
library(kableExtra)

#Load the combined_nutrition_data.csv file
data <- read_csv("https://raw.githubusercontent.com/Stat184-Spring2025/Stat184-Spring2025-Sec3_

#Clean data further to prevent errors in visuali
data <- data %>%
```

```r
  mutate(Restaurant = str_to_title(Restaurant)) %>%
  mutate(Restaurant = str_replace_all(Restaurant, "[^[:alnum:] ]", "")) %>%
  mutate(Restaurant = str_squish(Restaurant)) %>%
  filter(!is.na(Protein))

#Question 1 ------------------------------------------------------------
#How does calorie count vary by restaurant?
#Goal: Reveal distribution of restaurants using calorie count (spread, outliers)
#Selection of visualization: Box Plot
#Rationale: Shows median, IQR, and outliers at once. We care about variation in data, e.g., whe

ggplot(data, aes(x = Restaurant, y = Calories, fill = Restaurant)) +
  geom_boxplot() +
  coord_flip() +                        #Horizontal boxes for readability
  theme_minimal() +
  theme(
    legend.position = "none",           #Drop the legend
    axis.title.x = element_text(),
    axis.title.y = element_text()
  ) +
  labs(
    x = "Restaurant",
    y = "Calories",
    title = "Calorie Distribution by Restaurant"
  )

#Question 2 ------------------------------------------------------------
#Do different restaurants differ in how healthy their items are (based on calories, fat, and pr
#Goal: Compare central tendencies of Calories, Fat, and Protein across restaurants to assess re
#Selection of visualization: Summary table + grouped bar chart
#Rationale:
# - Table gives exact average values for each metric
# - Grouped bar chart lets you visually compare multiple metrics side-by-side

#1. Compute per-restaurant averages
restaurant_stats <- data %>%
  group_by(Restaurant) %>%
  summarise(
    avg_calories = mean(Calories),
    avg_fat      = mean(Fat),
    avg_protein  = mean(Protein)
  )

restaurant_stats %>%
  filter(Restaurant=="Burger King") %>%
  pull(avg_protein)
```

```r
#2. Bar chart: Average Calories
ggplot(restaurant_stats, aes(
    x = avg_calories,
    y = reorder(Restaurant, avg_calories)
  )) +
  geom_col(fill = "#FFA07A") +
  labs(
    title = "Average Calories by Restaurant",
    x     = "Avg Calories",
    y     = "Restaurant"
  ) +
  theme_minimal()

#3. Bar chart: Average Fat
ggplot(restaurant_stats, aes(
    x = avg_fat,
    y = reorder(Restaurant, avg_fat)
  )) +
  geom_col(fill = "#20B2AA") +
  labs(
    title = "Average Fat (g) by Restaurant",
    x     = "Avg Fat (g)",
    y     = "Restaurant"
  ) +
  theme_minimal()

#4. Bar chart: Average Protein
restaurant_stats %>%
  filter(!is.na(avg_protein)) %>%
  ggplot(aes(
    x = avg_protein,
    y = reorder(Restaurant, avg_protein)
  )) +
  geom_col(fill = "#778899") +
  labs(
    title = "Average Protein (g) by Restaurant",
    x = "Avg Protein (g)",
    y = "Restaurant"
  ) +
  theme_minimal()

#Question 3 ------------------------------------------------------------
#Which restaurant offers the highest average protein per menu item?
#Goal: Compare protein content across brands to identify which offer the most protein-heavy mea
#Selection of visualization: Horizontal bar chart
#Rationale: Shows how different restaurants stack up in terms of lean/high-protein offerings.
```

```r
data %>%
  group_by(Restaurant) %>%
  summarize(Avg_Protein = mean(Protein, na.rm = TRUE)) %>% #group by restaurant and calulate av
  mutate(Avg_Protein = ifelse(is.na(Avg_Protein), 0, Avg_Protein)) %>% #replace missing values
  ggplot(aes(x = reorder(Restaurant, Avg_Protein), y = Avg_Protein)) + #create bar graph
  geom_col(fill = "#4682B4") +
  coord_flip() +
  labs(
    title = "Average Protein per Menu Item by Restaurant",
    x = "Restaurant",
    y = "Avg Protein (g)"
  ) +
  theme_minimal()

#Question 4 ------------------------------------------------------------
#Which menu items offer the best protein-to-fat ratio by restaurant?
#Goal: Help consumers identify lean, efficient options at each restaurant.
#Selection of visualization: Frequency table showing top 3 and single-best menu items per resta
#Rationale: Nutrient-dense items with low fat and high protein are often desired for health/per

library(tidyverse)
library(kableExtra)
library(dplyr)

data <- data %>% #Compute protein-fat ratio
  mutate(PF_Ratio = Protein / Fat)

#top three
invisible(
  top3_pf_items <- data %>%
    filter(Fat > 0, !is.na(Protein), !is.na(Fat)) %>%
    group_by(Restaurant) %>%
    slice_max(order_by = PF_Ratio, n = 3, with_ties = FALSE) %>%
    ungroup()
)

top3_pf_items %>% #display the results as a table
  select(Restaurant, Product, Calories, Protein, Fat, PF_Ratio) %>%
  arrange(Restaurant, desc(PF_Ratio)) %>%
  kable(
    digits = 2,
    booktabs = TRUE,
    caption = "Table 2: Top 3 Menu Items per Restaurant by Protein-to-Fat Ratio"
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE,
```

```
    position = "center"
  )

#single-best filter
invisible(
  best_pf_item <- data %>%
    filter(Fat > 0, !is.na(Protein), !is.na(Fat)) %>%
    group_by(Restaurant) %>%
    slice_max(order_by = PF_Ratio, n = 1, with_ties = FALSE) %>%
    ungroup()
)

best_pf_item %>% #display the results as a table
  select(Restaurant, Product, Calories, Protein, Fat, PF_Ratio) %>%
  arrange(desc(PF_Ratio)) %>%
  kable(
    digits = 2,
    booktabs = TRUE,
    caption = "Table 3: Best Single Menu Item per Restaurant by Protein-to-Fat Ratio Ranked"
  ) %>%
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE,
    position = "center"
  )
```