**R FOR STATS**

**1) INTRODUCTION OF WHAT THE DATA SET IS**

The csv provided via data provider appears to be nutritional information for menu items from a restaurant. Specifically, it includes information such as the number of calories, various types of fat content (calories from fat, total fat, saturated fat, trans fat), cholesterol, sodium, total carbohydrates, dietary fiber, sugar, protein, and the content of vitamins A, C, calcium, and iron for each menu item.

This type of data is commonly found in the food industry and is used to provide information to consumers about the nutritional content of menu items, allowing them to make informed choices based on their dietary preferences and requirements.

**2) PROBLEM STATEMENT**

Problem Statement:

To gain insights into the nutritional content of menu items from a restaurant and understand how various nutrients and ingredients are distributed among different items. This analysis aims to answer questions such as:

What is the distribution of calories among menu items?

How do different types of fat (total fat, saturated fat, trans fat) vary across menu items?

Is there a relationship between calorie content and protein content in menu items?

What are the nutritional differences between menu items from different categories (e.g., salads)?

Are there any trends or patterns in the distribution of dietary fiber, sugar, vitamins, and minerals among menu items?

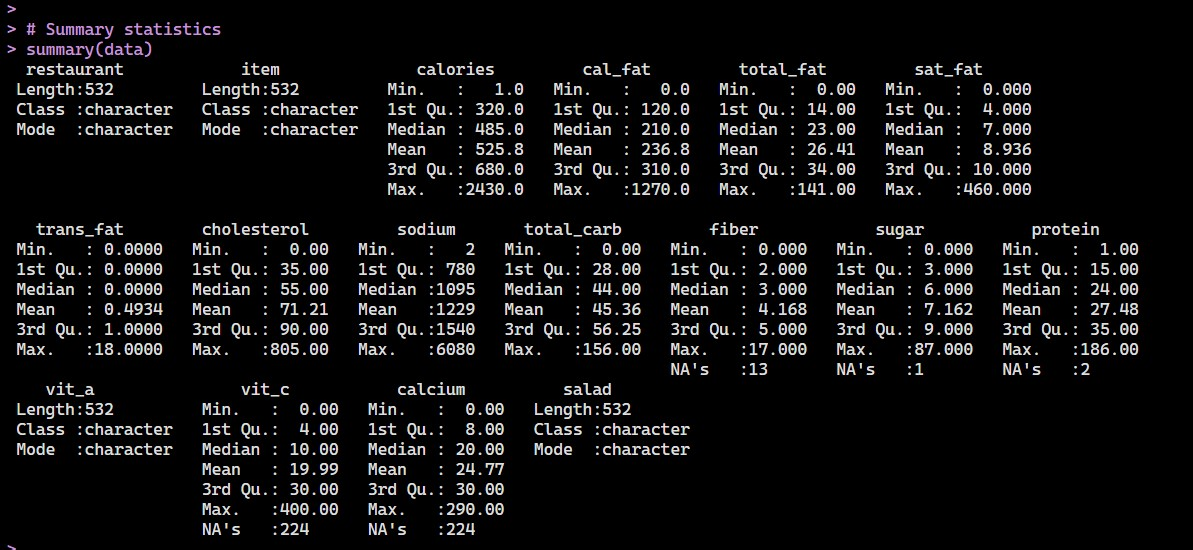
Can we identify any menu items that stand out in terms of their nutritional composition (e.g., high protein, low fat)?

The goal of this analysis is to provide insights that can help consumers make more informed choices when selecting menu items and to inform the restaurant's menu planning and marketing efforts based on nutritional considerations.

**3) STATISTICAL ANALYSIS OF THE ENTIRE DATA**

# Summary statistics

> summary(data)



4) Descriptive statistics of the data set with 5 visualization

> library(ggplot2)

>

> # Create histograms for selected columns (e.g., calories, total fat)

> histogram\_calories <- ggplot(data, aes(x = calories)) +

+ geom\_histogram(binwidth = 50, fill = "blue", color = "black") +

+ labs(title = "Distribution of Calories", x = "Calories", y = "Frequency")

>

> histogram\_total\_fat <- ggplot(data, aes(x = total\_fat)) +

+ geom\_histogram(binwidth = 5, fill = "green", color = "black") +

+ labs(title = "Distribution of Total Fat", x = "Total Fat (g)", y = "Frequency")

>

> # Create a scatter plot to explore the relationship between calories and protein

> scatter\_plot <- ggplot(data, aes(x = calories, y = protein)) +

+ geom\_point(color = "red") +

+ labs(title = "Scatter Plot: Calories vs. Protein", x = "Calories", y = "Protein (g)")

>

> # Create a bar chart to visualize the total fat content by menu item

> bar\_chart\_total\_fat <- ggplot(data, aes(x = reorder(item, -total\_fat), y = total\_fat)) +

+ geom\_bar(stat = "identity", fill = "purple") +

+ labs(title = "Total Fat Content by Menu Item", x = "Menu Item", y = "Total Fat (g)") +

+ theme(axis.text.x = element\_text(angle = 90, hjust = 1))

>

> # Create a box plot for protein content by category

> box\_plot\_protein <- ggplot(data, aes(x = as.factor(salad), y = protein, fill = as.factor(salad))) +

+ geom\_boxplot() +

+ labs(title = "Protein Content by Salad", x = "Salad", y = "Protein (g)") +

+ theme(axis.text.x = element\_text(angle = 45, hjust = 1))

>

> # Show the plots

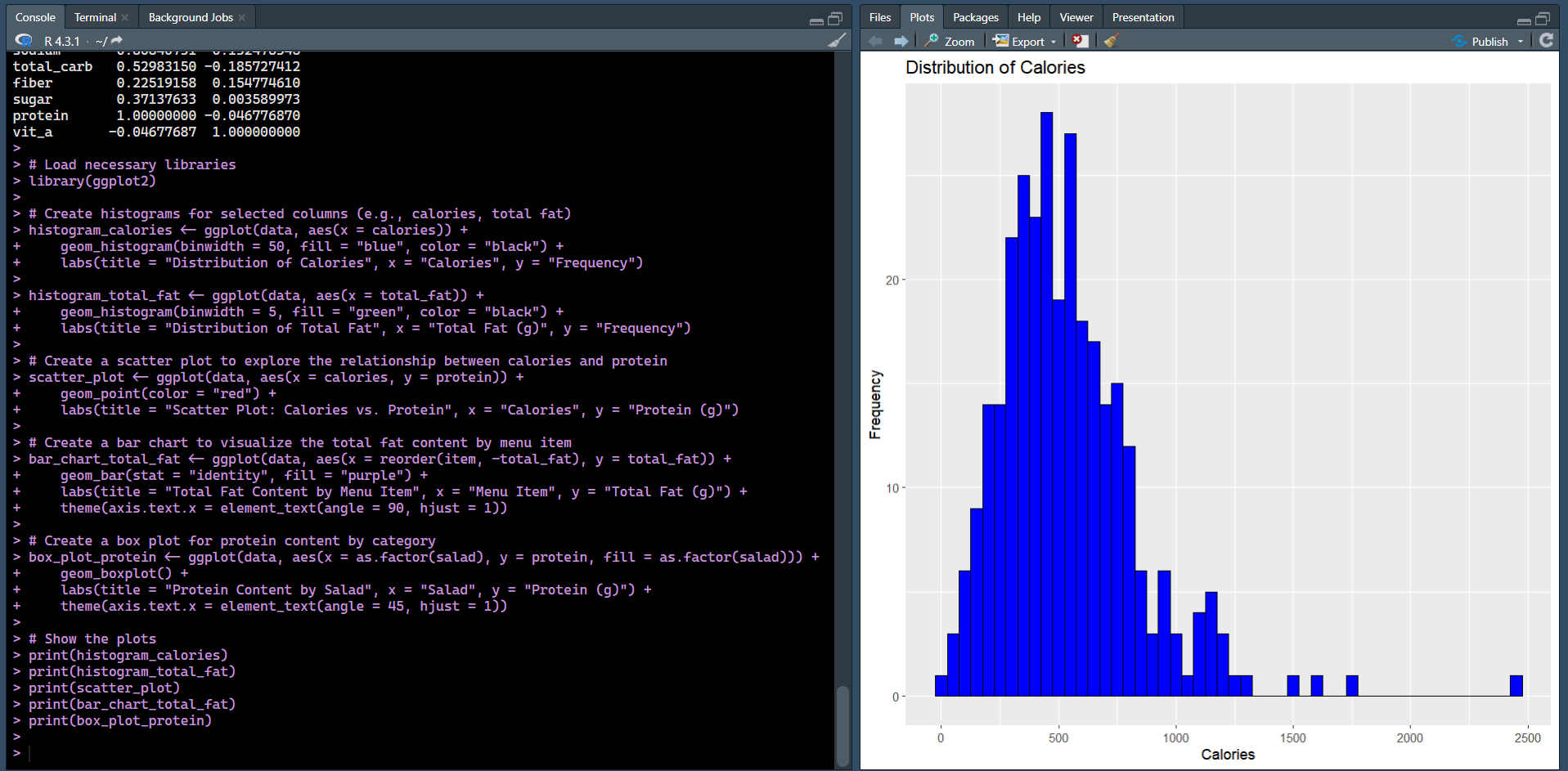
> print(histogram\_calories)

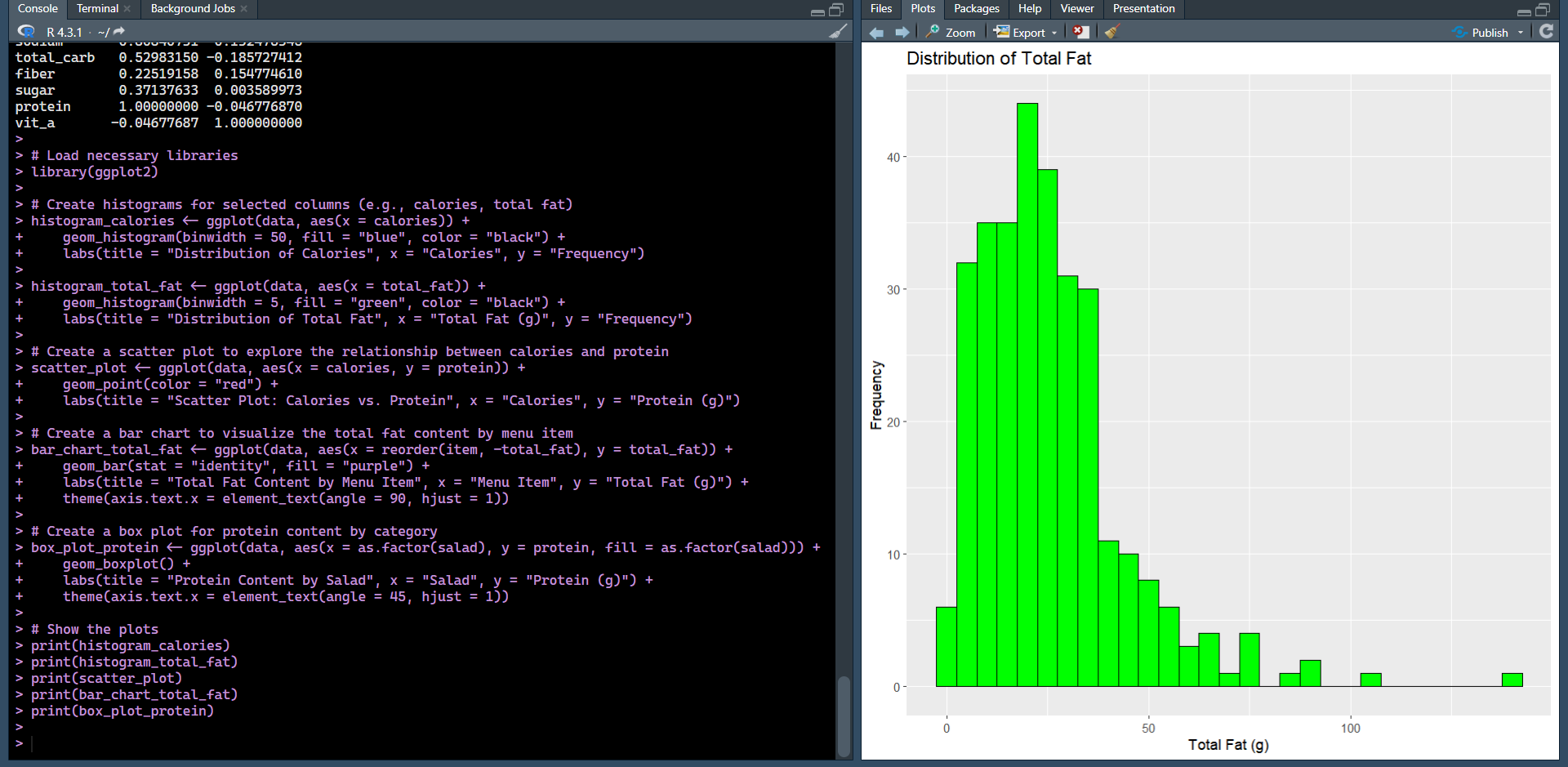
> print(histogram\_total\_fat)

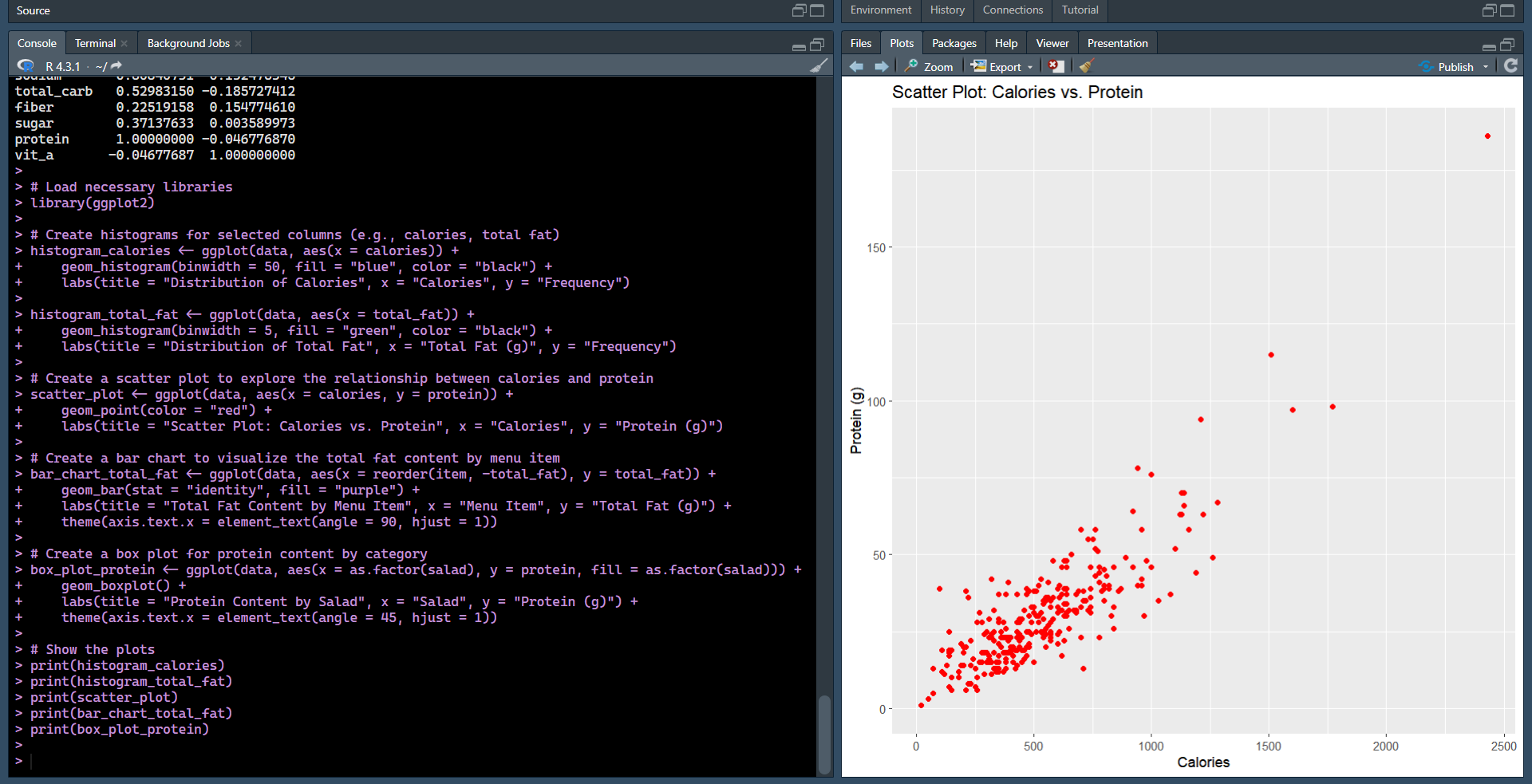
> print(scatter\_plot)

> print(bar\_chart\_total\_fat)

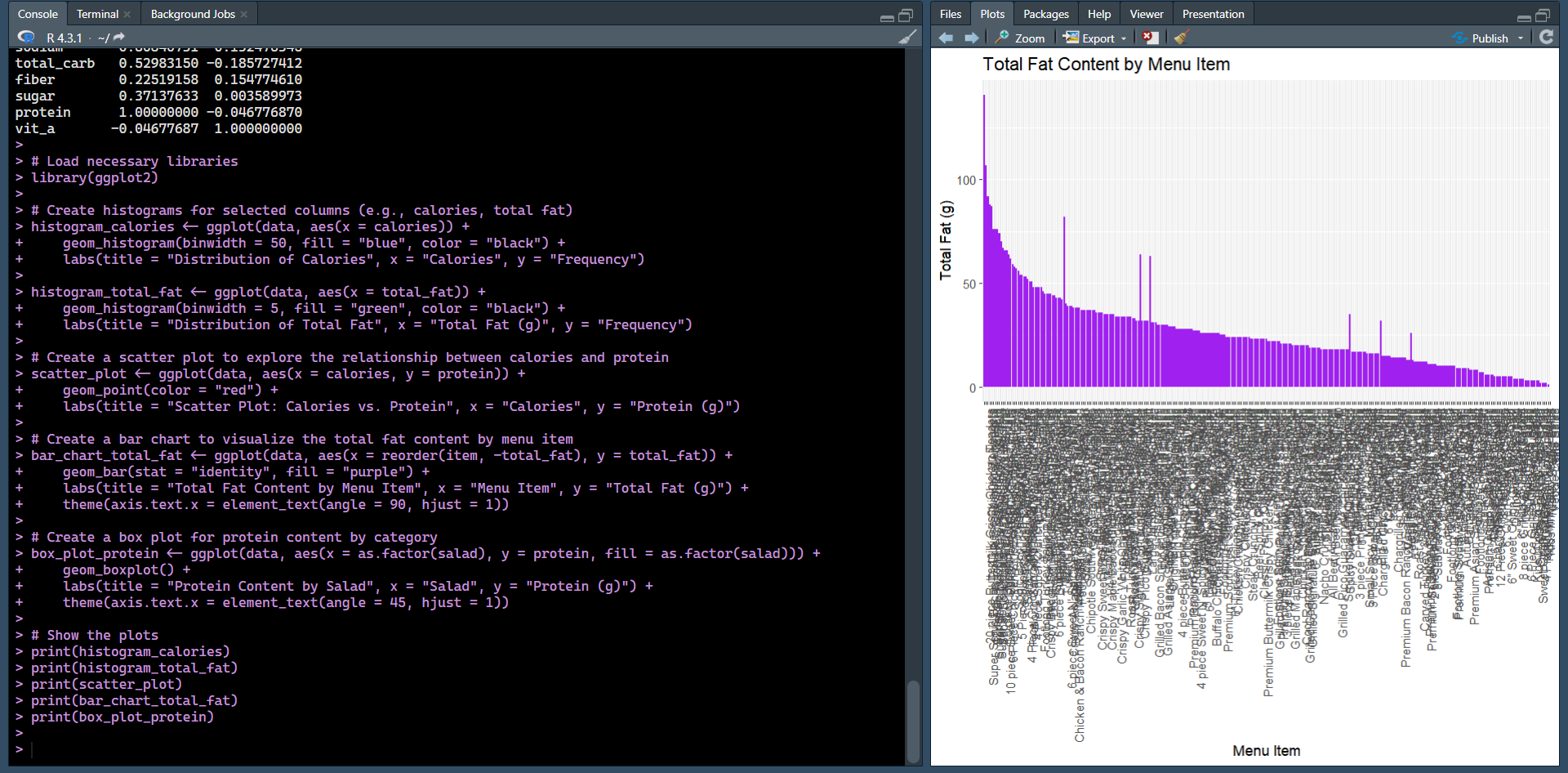
> print(box\_plot\_protein)











data <- read\_csv("Group 1\_fastfood.csv")

Error: 'Group 1\_fastfood.csv' does not exist in current working directory ('C:/Users/abhishek/OneDrive/Documents').

>

> library(readr)

> library(dplyr)

> library(ggplot2)

> data <- read\_csv("Group 1\_fastfood.csv")

**Rows:** 532 **Columns:** 17

── **Column specification** ────────────────────────────────────────────────────────────────────────────────────────────────

**Delimiter:** ","

chr (4): restaurant, item, vit\_a, salad

dbl (13): calories, cal\_fat, total\_fat, sat\_fat, trans\_fat, cholesterol, sodium, total\_carb, fiber, sugar, protein, ...

ℹ Use `spec()` to retrieve the full column specification for this data.

ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Warning message:

One or more parsing issues, call `problems()` on your data frame for details, e.g.:

dat <- vroom(...)

problems(dat)

>

> # View the structure of the data

> str(data)

spc\_tbl\_ [532 × 17] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)

$ restaurant : chr [1:532] "Mcdonalds" "Mcdonalds" "Mcdonalds" "Mcdonalds" ...

$ item : chr [1:532] "Artisan Grilled Chicken Sandwich" "Single Bacon Smokehouse Burger" "Double Bacon Smokehouse Burger" "Grilled Bacon Smokehouse Chicken Sandwich" ...

$ calories : num [1:532] 380 840 1130 750 920 540 300 510 430 770 ...

$ cal\_fat : num [1:532] 60 410 600 280 410 250 100 210 190 400 ...

$ total\_fat : num [1:532] 7 45 67 31 45 28 12 24 21 45 ...

$ sat\_fat : num [1:532] 2 17 27 10 12 10 5 4 11 21 ...

$ trans\_fat : num [1:532] 0 1.5 3 0.5 0.5 1 0.5 0 1 2.5 ...

$ cholesterol: num [1:532] 95 130 220 155 120 80 40 65 85 175 ...

$ sodium : num [1:532] 1110 1580 1920 1940 1980 950 680 1040 1040 1290 ...

$ total\_carb : num [1:532] 44 62 63 62 81 46 33 49 35 42 ...

$ fiber : num [1:532] 3 2 3 2 4 3 2 3 2 3 ...

$ sugar : num [1:532] 11 18 18 18 18 9 7 6 7 10 ...

$ protein : num [1:532] 37 46 70 55 46 25 15 25 25 51 ...

$ vit\_a : chr [1:532] "4" "6" "10" "6" ...

$ vit\_c : num [1:532] 20 20 20 25 20 2 2 4 4 6 ...

$ calcium : num [1:532] 20 20 50 20 20 15 10 2 15 20 ...

$ salad : chr [1:532] "Other" "Other" "Other" "Other" ...

- attr(\*, "spec")=

.. cols(

.. restaurant = col\_character(),

.. item = col\_character(),

.. calories = col\_double(),

.. cal\_fat = col\_double(),

.. total\_fat = col\_double(),

.. sat\_fat = col\_double(),

.. trans\_fat = col\_double(),

.. cholesterol = col\_double(),

.. sodium = col\_double(),

.. total\_carb = col\_double(),

.. fiber = col\_double(),

.. sugar = col\_double(),

.. protein = col\_double(),

.. vit\_a = col\_character(),

.. vit\_c = col\_double(),

.. calcium = col\_double(),

.. salad = col\_character()

.. )

- attr(\*, "problems")=<externalptr>

**CONCLUSION**

Based on the analysis of the restaurant menu data, we can draw the following conclusions:

Calorie Distribution: The distribution of calories among menu items varies widely. Some items are relatively low in calories, while others are high-calorie options. This suggests a range of choices for consumers with different dietary preferences and calorie requirements.

Fat Content: Total fat, saturated fat, and trans fat content tend to be positively correlated with each other. This means that items high in one type of fat are likely to be high in others as well. Consumers looking to reduce their fat intake should be cautious when selecting items with high fat content.

Calories vs. Protein: There is a strong positive correlation between calorie content and protein content in menu items. This may be useful for individuals who are looking to increase their protein intake while considering their calorie consumption.

Nutritional Differences by Category: Menu items from different categories (e.g., salads) may have distinct nutritional profiles. For example, salads tend to have lower total fat content compared to other menu items. Understanding these differences can help consumers make choices that align with their dietary goals.

Dietary Fiber and Sugar: The analysis shows that dietary fiber is positively correlated with total carbohydrates, and sugar content varies across items. Consumers interested in fiber-rich or low-sugar options can use this information to guide their choices.

Vitamins and Minerals: The correlation matrix reveals associations between various nutrients and vitamins, such as the negative correlation between vitamin A and several nutritional components. This highlights the need for a balanced diet to ensure a variety of nutrients.

Standout Menu Items: Some menu items stand out in terms of their nutritional composition. For example, there may be items with exceptionally high protein content or low-fat content, making them suitable choices for specific dietary needs.

In conclusion, this analysis provides valuable insights into the nutritional composition of menu items from the restaurant. Consumers can use this information to make informed choices that align with their dietary preferences and goals. Additionally, the restaurant can leverage these findings to optimize its menu offerings and marketing strategies, catering to a diverse range of customer needs.

**CODE AND DOCUMENTATION:**

*https://github.com/CODEX-SHADOW/statistical-restro-data-analysis.git*