GROUP 14

NUTRITIONAL DIETARY DATA

DATA SCIENECE FOR THE HEALTH SCIENCES

INTRODUCTION

A healthy diet is crucial for public and environmental health, with over 100 national dietary guidelines influencing health outcomes. However, people often opt for individual foods, posing challenges for researchers and policymakers in guiding public food choices based on these guidelines (Lawrence, 2024).

This study explores the dataset "**3_Nutritional_Dietary_data_Group_014.csv**", which contains data on patient nutritional diets such as body composition index (BMI). Muscle Mass, Body Fat Percent, Physical Activity, Daily Caloric Intake, Protein Intake, Fat Intake, Carbohydrate Intake, Vitamin C, Iron and Water Intake.

Analyzing this nutritional and dietary dataset, we will focus on understanding the relationship between dietary intake, physical activity, and body composition. By analyzing these variables, the goal is to uncover patterns and correlations that can inform us on how to make a strategy for these individuals dietary planning, fitness, and also public health strategies. We want to create effective strategies to better guidance for public food choices that align with national dietary guidelines.

METHODS

The dataset was first imported into the R environment and cleaned using the **tidyverse** and **skimr** packages.

Key steps included:

Data Cleaning: Using function glimpse() , summary (), and skim () to get an overview of the dataset's structure, summary statistics and initial insights into data.

Type Conversion: colSums(is.na()) is used for checking missing values. rename_() and str_replace_all () is to rename the category and simplifies data manipulation.

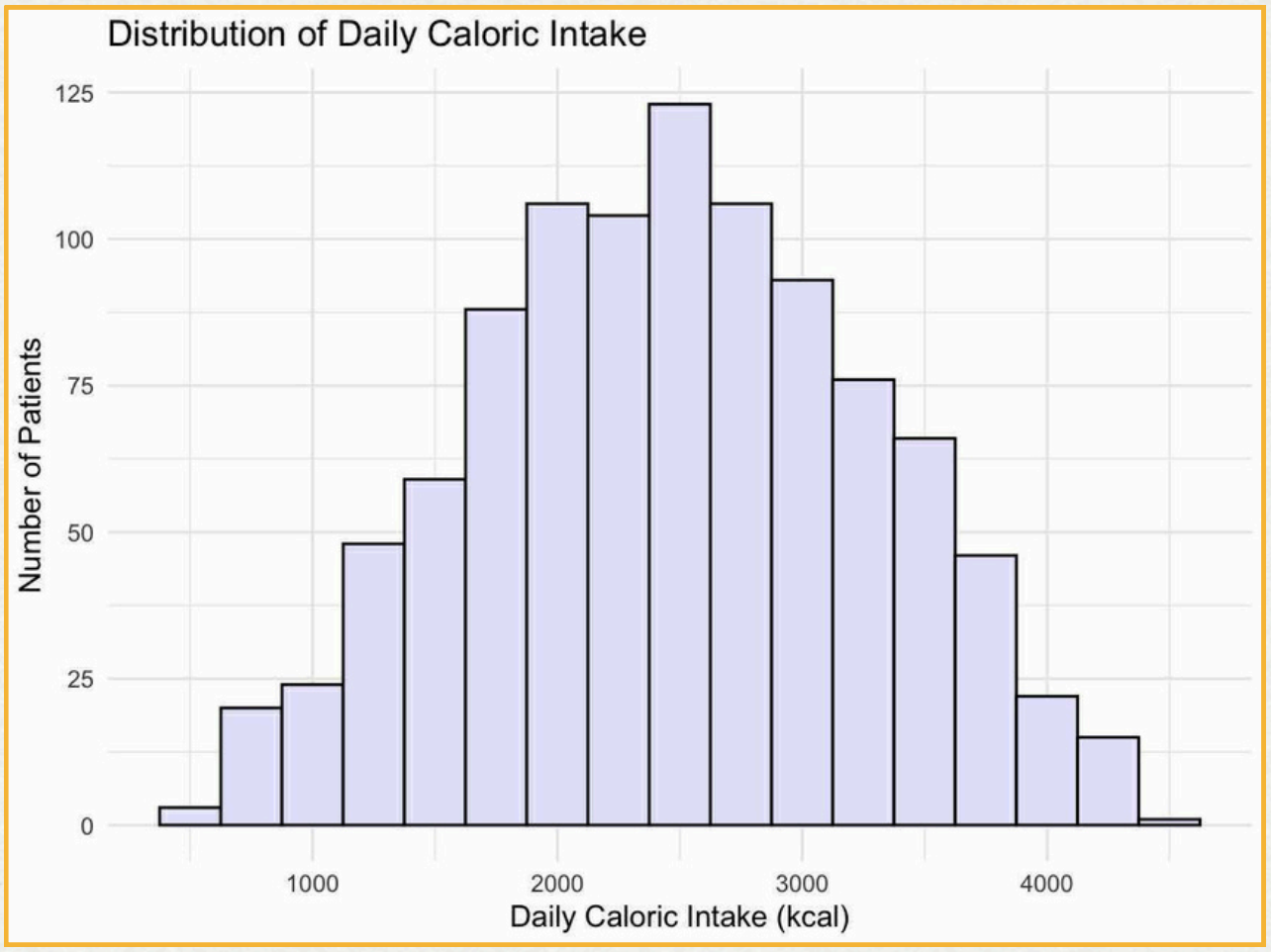
Formatting Data Types: All relevant variables are converted to numeric data types using mutate () and as.numeric(), so that conducting our analysis and visualization are performed correctly.

Descriptive Statistics: Summary metrics including mean, median, quartiles and standard deviation were computed for all numeric variables using summary () and skim().

Visual Analysis: A variety of plots were created using ggplot2 for visualizations of nutritional and body composition variables, including histograms for daily caloric intake and body fat percentage, scatter plots for the relationship between caloric intake and physical activity as well as protein intake and muscle mass, and a box plot depicting BMI distribution according to physical activity levels.

Linear Regression: Using lm() , a linear regression model was used to examine if daily caloric intake and physical activity hours can predict body fat percentage.

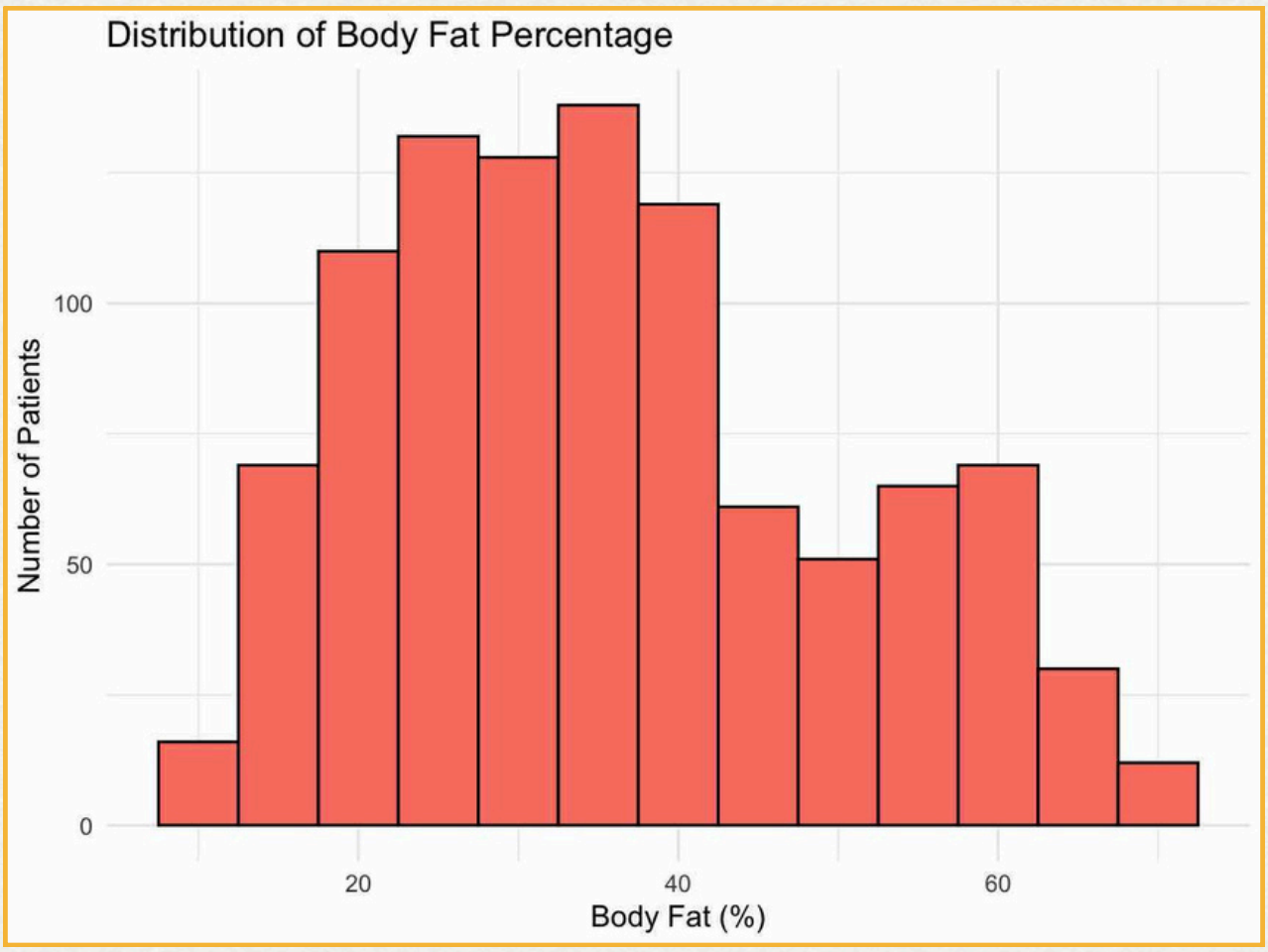
RESULTS



Distribution of Daily Caloric Intake

The histogram shows a bell-shaped distribution meaning that the majority of individuals have daily caloric intake clustering around the central average, peaking between approximately 2250 and 2500 kcal. Also there are fewer individuals consuming a very low or very high amount of calories.

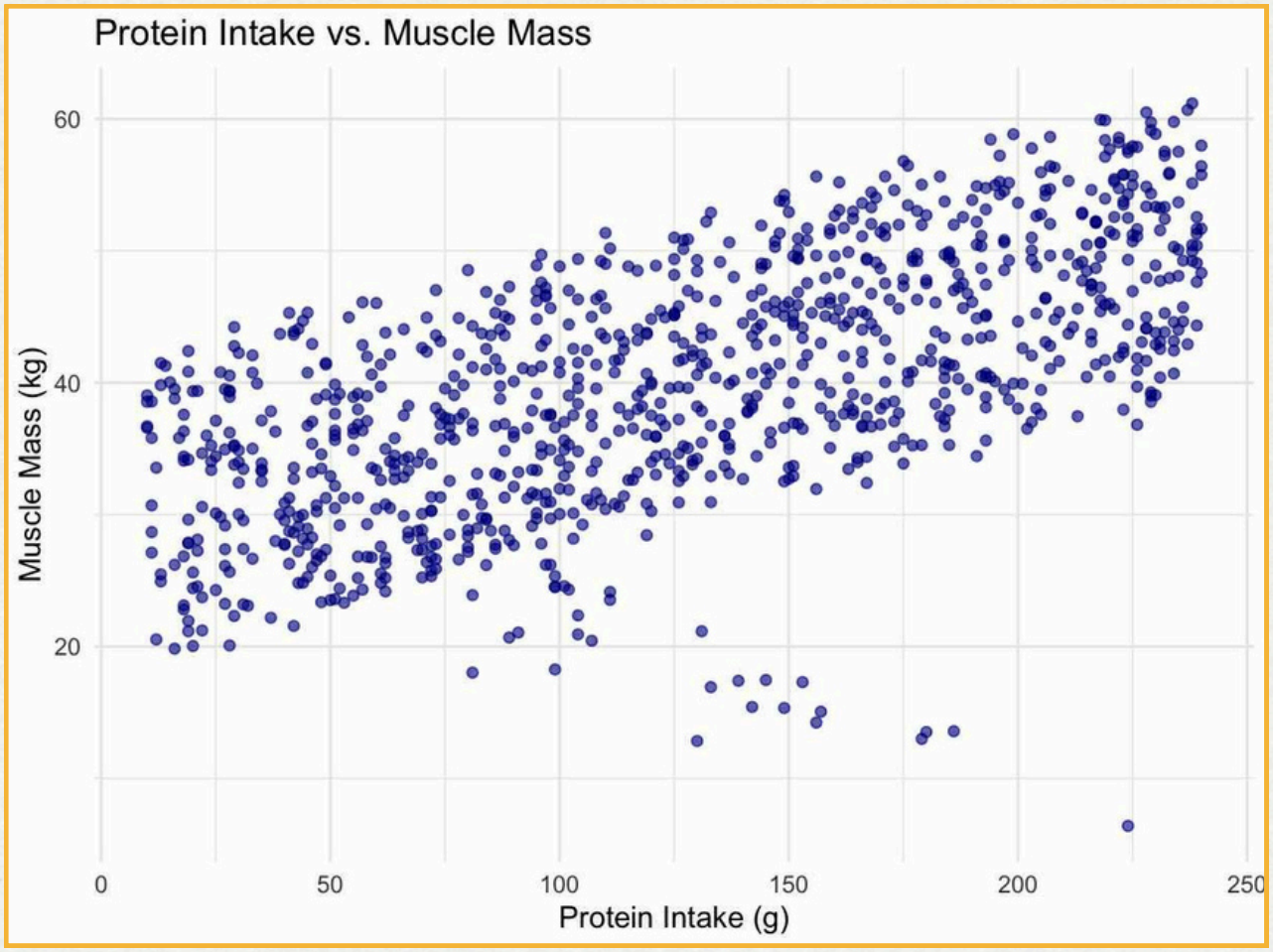
RESULTS



Distribution of Daily Caloric Intake

The histogram distribution shows a **right-skewed distribution**, indicating the tail of patients have higher body fat percentages extending towards to the right. There are two peaks, where one tall peak is 30% and 40% body fat and secondary peak is around 55% to 60% body fat. We could imply that there are different physiological or lifestyle characteristics with the patient groups.

RESULTS

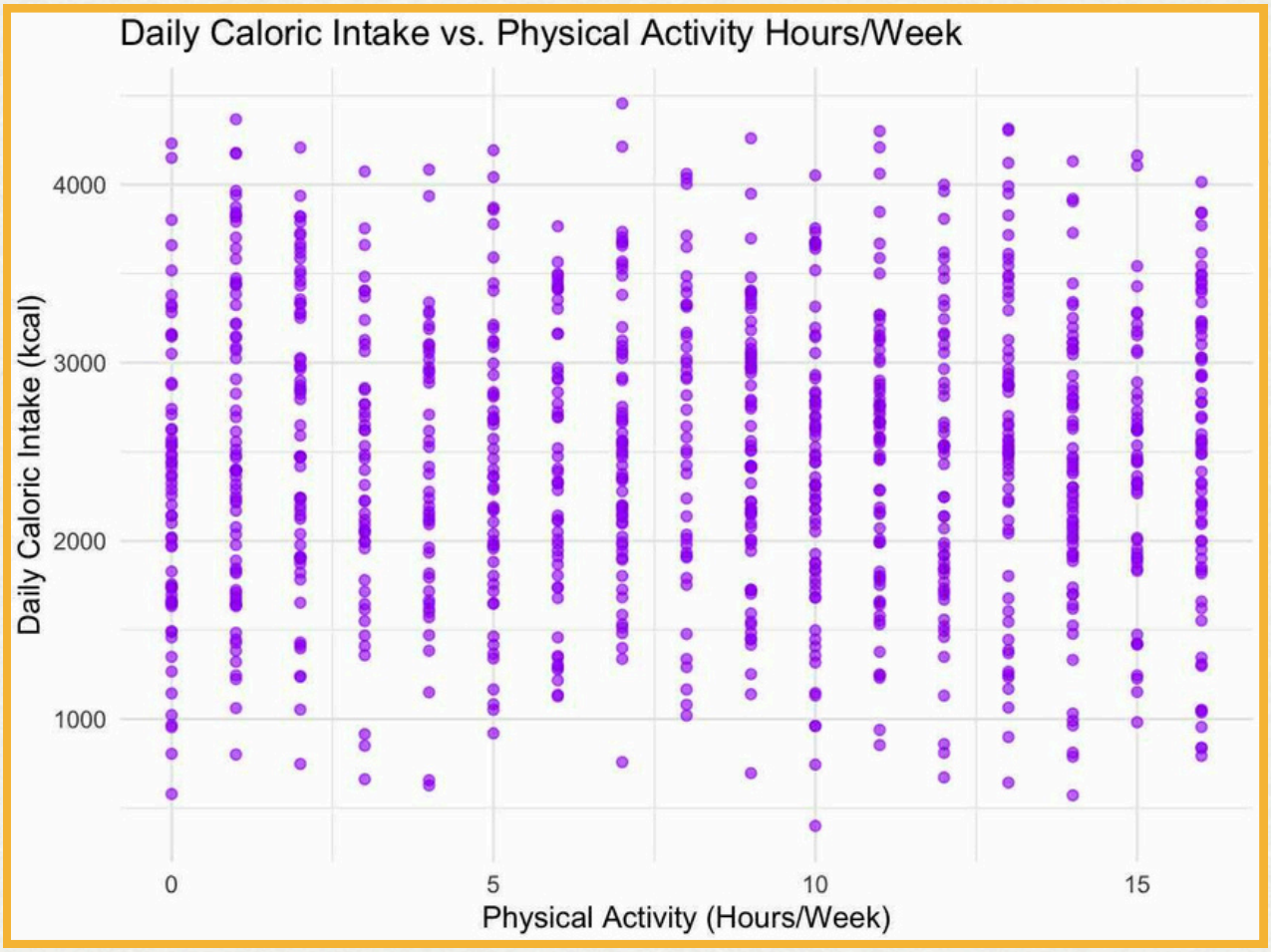


Protein Intake vs. Muscle Mass

The scatter plot shows a **strong positive correlation** between protein intake and muscle mass. As protein increases, muscle mass increases as well, having a clear slope that goes upward from left to right.

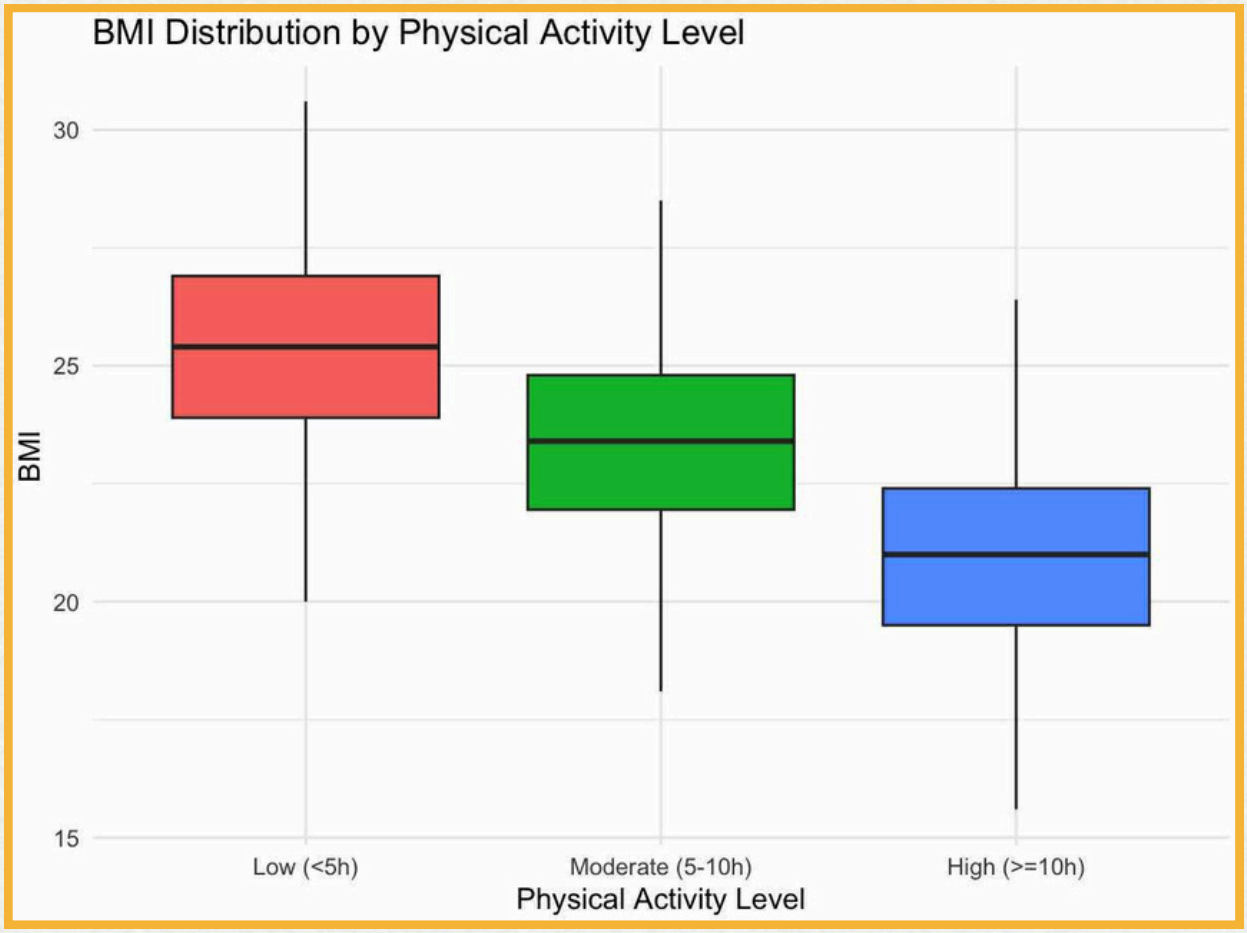
While the trend is strong, there is still some variability around the general trend. For any given protein intake, there's a range of muscle mass values, and vice versa. This indicates that while protein intake is a significant factor, other elements (like genetics, type of physical activity, age) also influence muscle mass.

RESULTS



The scatter plot shows **no strong linear correlation**—that physical activity hours are not a strong predictor of daily caloric intake and that the other factors likely influence dietary consumption habits.

RESULTS



BMI Distribution by Physical Activity Level

The boxplot shows a clear **inverse relationship** between physical level and BMI. As physical activity increases from low to high, the median decreases.

KEY FINDINGS

People that are less than 5 hours of activity level have the highest median BMI, 50% of these people fall in between 23.5 and 27.

While people that moderately have 5 to 10 hours of activity have a lower median BMI, these people falls between 19.5 and 24.5.

Lastly, people that have a high activity level that is more than or equal to 10 hours of movement have the lowest median BMI as these people fall between 19.5 and 22.

These findings show that the people have a low activity level indicating they are more likely overweight or obese. They are at risk of various health problems–that is why understanding the importance of physical activity in maintaining a healthy body mass.

CONCLUSION

This analysis provides insights into the relationship between dietary habits, physical activity, and body composition. Daily caloric intake usually follows a bell-shaped curve, with most people consuming between 2250 and 2500 kcal. However, body fat percentage has a right-skewed distribution, with two clear peaks around 30-40% and 55-60%. This suggests different physiological or lifestyle factors among patient groups.

Further examination shows no strong linear relationship between daily caloric intake and hours of physical activity per week. This indicates that other factors likely affect dietary habits. In contrast, there is a strong positive relationship between protein intake and muscle mass. Higher protein consumption is linked to greater muscle mass. Notably, BMI distribution by physical activity level shows an inverse relationship: as physical activity increases from low to high, the median BMI decreases. This highlights the important role of physical activity in maintaining a healthy body mass and reducing health risks related to higher BMI.

REFERENCE

Lawrence, M. Fundamentals of a healthy and sustainable diet. Nutr J 23, 150 (2024). https://doi.org/10.1186/s12937-024-01049-6