Health Data Visualization: Interactive Analysis of Obesity Factors GPH-Digital Health

Prof Dominik Böhler

Done By Chaden Khalaf

The number of words in the document: 2674 Words

Use Case

The obesity epidemic is a global public health concern characterized by a significant increase in the prevalence of obesity across populations worldwide. Obesity is a complex and multifactorial condition influenced by genetic, environmental, behavioral, and socioeconomic factors. The prevalence of obesity has risen dramatically, affecting individuals of all ages, genders, and socioeconomic backgrounds. According to the World Health Organization, global obesity rates have nearly tripled since 1975, with approximately 1.9 billion adults classified as overweight and over 650 million classified as obese in 2016. The obesity epidemic causes health challenges and consequences, including an increased risk of developing chronic diseases such as type 2 diabetes, cardiovascular disease, and musculoskeletal disorders. These conditions not only impair quality of life but also place a significant burden on healthcare systems and economies worldwide, due to the associated healthcare costs and productivity losses. Addressing the obesity epidemic requires a comprehensive and multifaceted approach involving multiple stakeholders at the local, national, and global levels.[1]

Stakeholder	Use of Visualization
Public Health Officials and Policy Makers	• Concerned with population health and disease prevention strategies. Can identify high-risk groups based on demographic and lifestyle factors for targeted interventions and public health campaigns.
Healthcare Providers	• Utilize visualization to understand relationship between lifestyle factors and obesity in patient population. • Can tailor personalized treatment plans and interventions for at-risk or affected patients.
Researchers	• Gain insights into complex interplay between factors contributing to obesity.
Individuals	• Gain understanding of how lifestyle choices impact risk of obesity. Empower individuals to make informed decisions about diet and exercise.
Educators and Health Advocates	• Use as teaching aid to raise awareness about healthy habits. • Motivate individuals to adopt healthier behaviors. • Promote positive changes in communities.
Insurance Companies	• Identify at-risk individuals and tailor wellness programs. • Encourage healthier behaviors among policyholders.
Fitness and Wellness Professionals	• Educate clients about importance of nutrition and physical activity. • Motivate clients to adopt healthy behaviors. • Achieve health goals.

Overall, this visualization has the potential to address a pressing public health issue by providing valuable insights into the factors contributing to obesity and facilitating targeted interventions to improve population health outcomes.

Using Histogram and Scatter Plot Visualizations to Interpret Factors Associated with Obesity Levels: Addressing Public Health Issues

The choice of histogram and scatter plot offer complementary visualizations that provide valuable insights into the factors influencing obesity levels.

Histogram: The histogram allows users to visualize the distribution of obesity levels across different categories of a selected variable, such as family history of overweight or frequency of physical activity. By observing the distribution of obesity levels within each category, users can identify trends, helping them understand which factors may contribute to higher obesity levels. This visualization is useful for identifying high-risk groups and informing targeted interventions aimed at reducing obesity among these group.

Scatter Plot: The scatter plot displays the relationship between age and BMI, with data points colored by gender. This allows users to observe how BMI varies with age and how this relationship differs between different gender. This enables users to gain deeper insights into demographic-specific trends in obesity levels. In addition, understanding how BMI changes with age and gender can help healthcare providers tailor personalized treatment plans and interventions for patients at risk.

This table highlights the challenges faced by the stakeholders in addressing the obesity epidemic before Data visualization, along with the impact of data visualization using the histogram and scatter plot.

Stakeholders	Challenges to Solve Epidemic Before Data Visualization	Impact of Data Visualization
Public Health Officials	Identifying high-risk groups without detailed data analysis.	Identifies high-risk groups based on demographic and lifestyle factors.
	Developing targeted interventions without understanding demographic disparities.	Enables targeted interventions and public health campaigns based on insights from visualization.
Healthcare Providers	Tailoring personalized treatment plans without comprehensive understanding of factors influencing obesity.	Provides insights into relationships between lifestyle factors and obesity, enabling better-informed treatment decisions.
	Addressing demographic disparities in obesity prevalence and treatment outcomes.	Illustrates demographic-specific trends in obesity, facilitating tailored interventions and patient education.

Researchers	Validating hypotheses without access to comprehensive data.	Provides visual confirmation of research findings, aiding in hypothesis validation and further investigation.
	Limited understanding of real-world implications of research findings.	Facilitates dissemination of research findings to broader scientific community, enhancing understanding of real-world implications.
Individuals	Limited access to information and resources for understanding personal risk factors.	Empowers individuals to understand how lifestyle choices impact obesity risk, motivating behavior change.
	Difficulty in interpreting complex health data.	Simplifies complex health data into easily understandable visualizations, promoting health literacy and informed decision-making.
Educators and Health Advocates	Inadequate tools for raising awareness about obesity and promoting healthy lifestyles.	Serves as educational tool for raising awareness about obesity risk factors and promoting healthy lifestyle choices.
Insurance Companies	Identifying at-risk individuals without access to detailed health data.	Identifies at-risk individuals based on insights from visualization, enabling targeted wellness programs and incentives.
	Developing effective preventive health measures without understanding underlying risk factors.	Informs development of preventive health measures by identifying underlying risk factors and trends in obesity prevalence.
Fitness and Wellness Professionals	Tailoring fitness and nutrition plans without comprehensive understanding of individual health profiles.	Provides insights into individual health profiles and factors influencing obesity, enabling personalized fitness and nutrition plans.
	Engaging clients in behavior change without clear understanding of motivations and barriers.	Facilitates client engagement through visual representation of health data, motivating behavior change and goal setting.

From Data Import to Shiny Web Page: A Journey in Data Cleaning and Visualization

1. Obtain the Dataset:
   The file "Dataset.csv" include data for the estimation of obesity levels in individuals from the countries of Mexico, Peru and Colombia, based on their eating habits and physical condition. The data contains 17 attributes and 2111 records, the records are labeled with the class variable NObesity (Obesity Level), that allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III. 77% of the data was generated synthetically using the Weka tool and the SMOTE filter, 23% of the data was collected directly from users through a web platform.

2. Importing and Cleaning the Dataset:

Step	Description	Changes Made
Package Installation and Loading	Check if required packages are installed, install them if not, and load necessary libraries.	Required packages: **rio**: Importing data, **here**: Relative file pathways, **janitor**: Data cleaning and tables, **matchmaker**: Dictionary-based cleaning, **epikit**: age_categories() function, **tidyverse**: Data management and visualization, **skimr**: Summary statistics and visualization
Data Import	Import the dataset from a CSV file using the **rio** package.	-
Summarizing the Dataset	Get a summary of the dataset using the **summary() function**.	-
Cleaning Column Names	Clean column names for consistency and ease of use using **janitor::clean_names()**.	-
Renaming Columns	Rename columns with spaces or non-standard names using **dplyr::rename()**.	"gender" to "Gender", "height" to "Height", "age" to "Age", "weight" to "Weight", "family_history_with_overweight" to "Family_History_With_Overweight", "favc" to "Frequent_High_Caloric_Food_Consumption", "fcvc" to "Frequency_Vegetables_Consumption", "ncp" to "Number_Of_Main_Meals", "caec" to "Consumption_Of_Food_Between_Meals", "smoke" to "Smoke", "ch2o" to "Consumption_Of_Water_Daily", "scc" to "Calories_Consumption_Monitoring", "faf" to "Frequency_Physical_Activity", "tue" to "Time_Using_Technology_Devices",

		"calc" to "Consumption_Alcohol", "mtrans" to "Transportation_Mode", "n_obeyesdad" to "Obesity_Level"
Handling Missing Values	Check for missing values and print the total count of missing values.	-
Handling Duplicates	Check for duplicates, remove them if found.	-
Data Summary and Inspection	Obtain summary statistics, check column names, and get a concise summary of the data structure.	-
Interactive Data Table	Create an interactive data table using the **DT::datatable() function**.	-
Checking Variable Structures and Classes	Check the structure of variables and classes using **str()** and **class().**	-
Converting Character Columns to Logical	Convert character columns to logical (TRUE/FALSE) using conditional statements.	"Frequent_High_Caloric_Food_Consumption", "Smoke", "Family_History_With_Overweight", "Calories_Consumption_Monitoring" converted to logical values
Converting Columns to Numeric	Convert character columns to numeric using **as.numeric() function** after converting them to character.	"Time_Using_Technology_Devices", "Frequency_Physical_Activity", "Consumption_Of_Water_Daily", "Frequency_Vegetables_Consumption", "Number_Of_Main_Meals" converted to numeric
Rounding Numeric Columns	Round numeric columns using the **round() function**.	-
Recode Categorical Variables	Recode categorical variables into meaningful categories using **conditional statements (ifelse())**.	"Time_Using_Technology_Devices", "Consumption_Of_Water_Daily", "Frequency_Physical_Activity", "Frequency_Vegetables_Consumption", "Number_Of_Main_Meals" recoded into descriptive categories
Saving Cleaned Data	Save the cleaned data to a CSV file named "linelist.csv" using **write.csv()**.	-
Checking Variable Classes	Check the class of specific variables Smoke and Transportation_Mode **using class()**.	This step was done just to make sure everything works.

The "linelist.csv" is the cleaned data file.

3. Exploratory Data Analysis
   After cleaning the dataset, an exploratory data analysis was made using Rstudio.

Step Description
Data Loading: Loaded the dataset "linelist.csv" into R using the **read.csv()** function.
Data Exploration: Utilized various visualization techniques to explore different aspects of the dataset, including age distribution, obesity level distribution, gender-wise obesity distribution, family history vs. obesity level relationship, and transportation mode preferences.
Interactive Table: Created an interactive table view of the dataset using the DT package, allowing for easy exploration of the data's details.
Visualization: Generated visualizations using the **ggplot2** package to effectively communicate insights about the dataset. These visualizations include density plots, bar plots, and mosaic plots.
Analysis Interpretation: Provided interpretations and insights for each visualization, highlighting key findings and observations related to obesity levels and associated factors.
Document Compilation: Compiled the analysis, including visualizations, interpretations, and insights, into an RMarkdown document for further processing.
Document Knitting: Knit the RMarkdown document to generate an HTML report containing the analysis results, visualizations, and interpretations.
Publication: Uploaded the HTML report to Posit and RPubs to publish and share the analysis with stakeholders. This step provides a convenient way for others to access and review the findings. https://rpubs.com/Chaden_khalaf/EDA

4. The Final Rscript for the Shiny WebPage:

Step Description	Code/Function
Load Required Libraries: Loaded necessary libraries including **shiny**, **dplyr**, and **plotly** for building the Shiny web application and performing data manipulation and interactive plotting.	library(shiny), library(dplyr), library(plotly)
Read Data: Imported the cleaned CSV data containing information on obesity factors from "linelist.csv" into the application environment.	data <- read.csv("linelist.csv") In my previous version, I used an absolute file path which lead to error when trying to run my application and deploy it. So in this version, I corrected the error using the relative file path, and to have all my files in the same working directory.
Define Variable Mapping: Created a mapping between user-friendly display names and original variable names to enhance the user experience when selecting variables for analysis.	variable_mapping <- list("Family History With Overweight" = "Family_History_With_Overweight", "Transportation Mode" = "Transportation_Mode", "Frequency Physical Activity" = "Frequency_Physical_Activity", "Consumption Of Food Between Meals" = "Consumption_Of_Food_Between_Meals", "Frequent High Caloric Food Consumption" = "Frequent_High_Caloric_Food_Consumption", "Number Of Main Meals" = "Number_Of_Main_Meals", "Consumption Of Water Daily" = "Consumption_Of_Water_Daily", "Calories Consumption

	Monitoring" = "Calories_Consumption_Monitoring", "Time Using Technology Devices" = "Time_Using_Technology_Devices")
Calculate BMI: Calculated Body Mass Index (BMI) for each individual.	data$bmi <- data$Weight / (data$Height)**2 In my previous version, the Y axis was incorrect giving numbers such as 500K. I was using this formula: BMI = Weight / (Height^2).
Create Tooltip: Generated tooltips for scatter plot data points to display additional contextual information, including age, BMI, and gender, upon hovering over data points.	data$tooltip <- paste("Age: ", data$Age, " ", "BMI: ", round(data$bmi, 2), " ", "Gender: ", data$Gender, sep = "") %>% sapply(htmltools::HTML)
Define UI (User Interface): Designed the layout and components of the Shiny web application user interface to optimize user interaction and visualization.	Defined in the ui object using fluidPage(), sidebarPanel(), mainPanel(), etc.
Create Server Logic: Developed server logic to handle user inputs, filter data accordingly, and generate interactive plots based on selected variables.	Defined in the server function using reactive expressions, render functions **(renderPlotly(), renderTable()),** and event observers **(observeEvent())**
Generate Histogram (Obesity Plot): Produced an interactive histogram plot illustrating the distribution of obesity levels across the selected variable.	output$obesityPlot <- renderPlotly({ p <- filtered_data() %>% plot_ly(x = ~get(variable_mapping[[input$x_var]]), color = ~gsub("_", " ", Obesity_Level)) %>% add_histogram() %>% layout(title = paste("Obesity Level by", input$x_var), xaxis = list(title = input$x_var), yaxis = list(title = "Count"), barmode = "stack") p })
Generate Scatter Plot: Created an interactive scatter plot depicting the relationship between age and BMI, with gender-coded data points and tooltips for enhanced interpretation.	output$scatterPlot <- renderPlotly({ p <- plot_ly(data, x = ~Age, y = ~bmi, color = ~Gender, colors = c("pink", "blue"), text = ~tooltip, hoverinfo = "text", type = 'scatter', mode = 'markers') %>% layout(title = 'Age vs BMI', xaxis = list(title = 'Age'), yaxis = list(title = 'Body mass index(BMI)'), legend = list(title = list(text = ' Gender '))) p })
Update BMI Classification Table: Dynamically updated the BMI classification table based on user-selected age range and gender. I used gsub function to remove the _ from legend. After so many errors I figured out how to code it.	output$bmi_classification_table <- renderTable({ classification <- data.frame(BMI_Classification = c("Underweight", "Normal", "Overweight", "Obesity I", "Obesity II", "Obesity III"), BMI_Range = c("Less than 18.5", "18.5 to 24.9", "25.0 to 29.9", "30.0 to 34.9", "35.0 to 39.9", "Higher than 40")) names(classification) <- gsub("_", " ", names(classification)) classification }, sanitize.text.function = function(x) x)
Collect User Feedback: Implemented functionality for users to provide feedback via text input and submit button, with feedback stored or processed accordingly (demonstration purpose only).	observeEvent(input$submit, { feedback <- isolate(input$feedback) cat("User feedback:", feedback, "\n") updateTextInput(session, "feedback", value = "") })
Display Project Context: Provided contextual information on the significance of understanding obesity factors and promoting healthy lifestyles.	Included in the ui object as text output components or explanatory text in the user interface elements.

5. Deployment Process

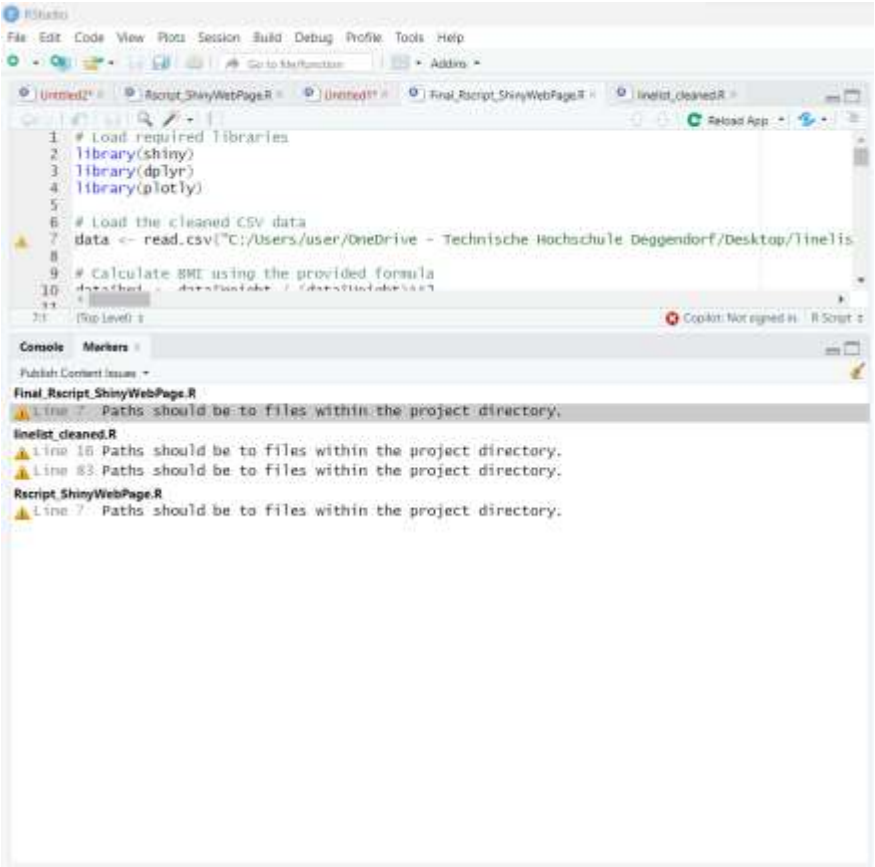
The Shiny web application was successfully deployed to the Shinyapps.io platform. The deployment process involved the following steps:

Preparation: Ensured that the Shiny application was fully developed and tested locally.

Installation and Authentication: Installed and loaded the rsconnect package in R. Authenticated the Shinyapps.io account using the rsconnect::setAccountInfo() function, providing the account name, token, and secret.

Deployment: Deployed the application using the rsconnect::deployApp() function. Progress messages were displayed in the R console indicating the status of the deployment process.

Here, I faced many errors due to the usage of absolute file path.



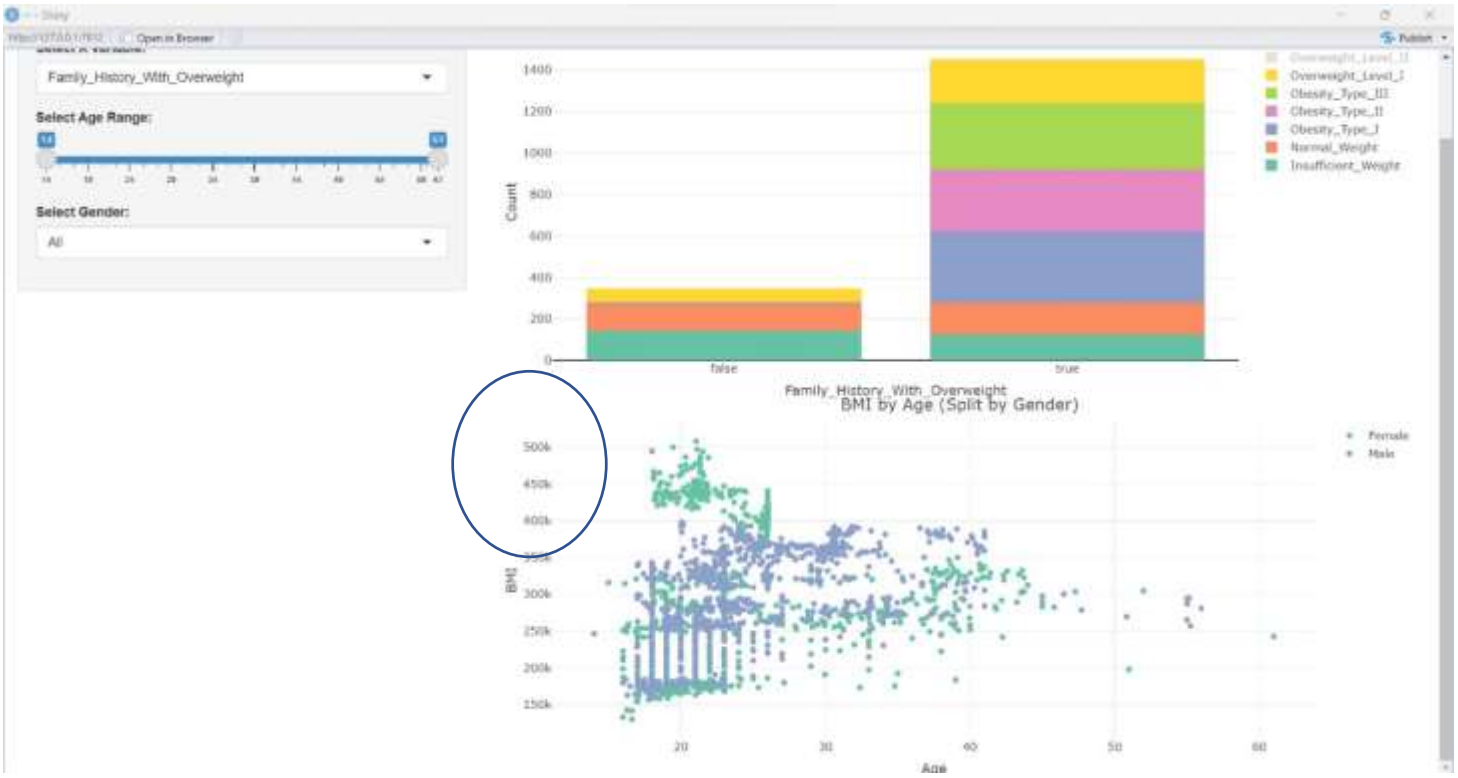
To correct this error, I used the relative file path instead of the absolute file path. In addition, all files should be in a same working directory.

Monitoring: Monitored the deployment process in the R console until completion. Once deployment was successful, received the URL to the deployed application. The deployed application can be accessed using the following URL: https://chadenkhalaf.shinyapps.io/Desktop/

Old Version of the Shiny Webpage

The BMI error:

I mentioned previously that I faced error with the BMI Y axis.
This was the previous version of the Shiny WebPage.



Before Converting Columns to Numeric, Rounding Numeric Columns, and Recode Categorical Variables:

Why to do these steps?

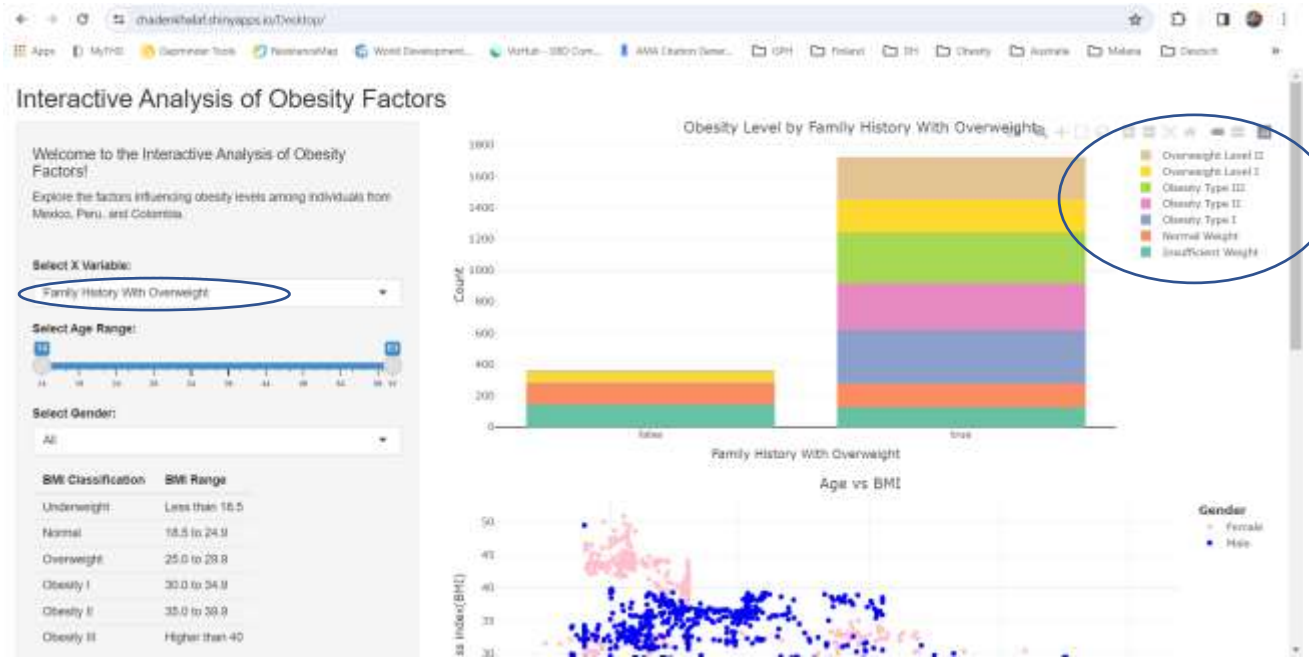
For Consistency and compatibility: Converting character variables to numeric ensures that they are in a format suitable for numerical operations and analysis. This ensures consistency in the data type across different variables and compatibility with statistical functions and models.

Data presentation: Rounding numeric variables can make them easier to interpret and present. It simplifies the values to a certain precision, making them more understandable for reporting and visualization purposes without losing important information.

After Converting Columns to Numeric, Rounding Numeric Columns, and Recode Categorical Variables:

The final touch: Variable Mapping

The variable mapping was established to streamline user interaction with the application interface. By providing clear and descriptive display names for each variable in the dataset, such as "Family History With Overweight" instead of "Family_History_With_Overweight," users can easily identify and select relevant factors for analysis.



Figures in the Exploratory Data Analysis

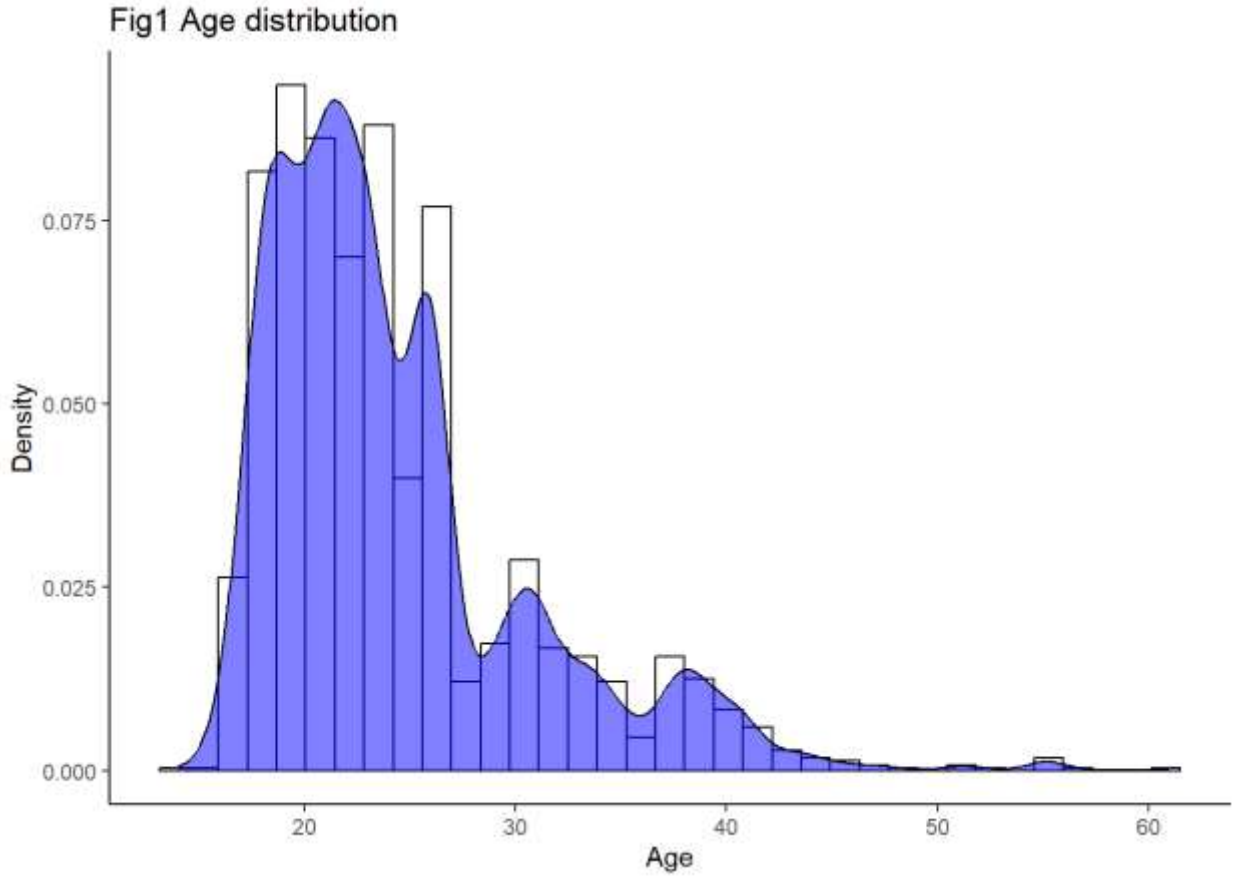


Fig2 Distribution across different obesity levels

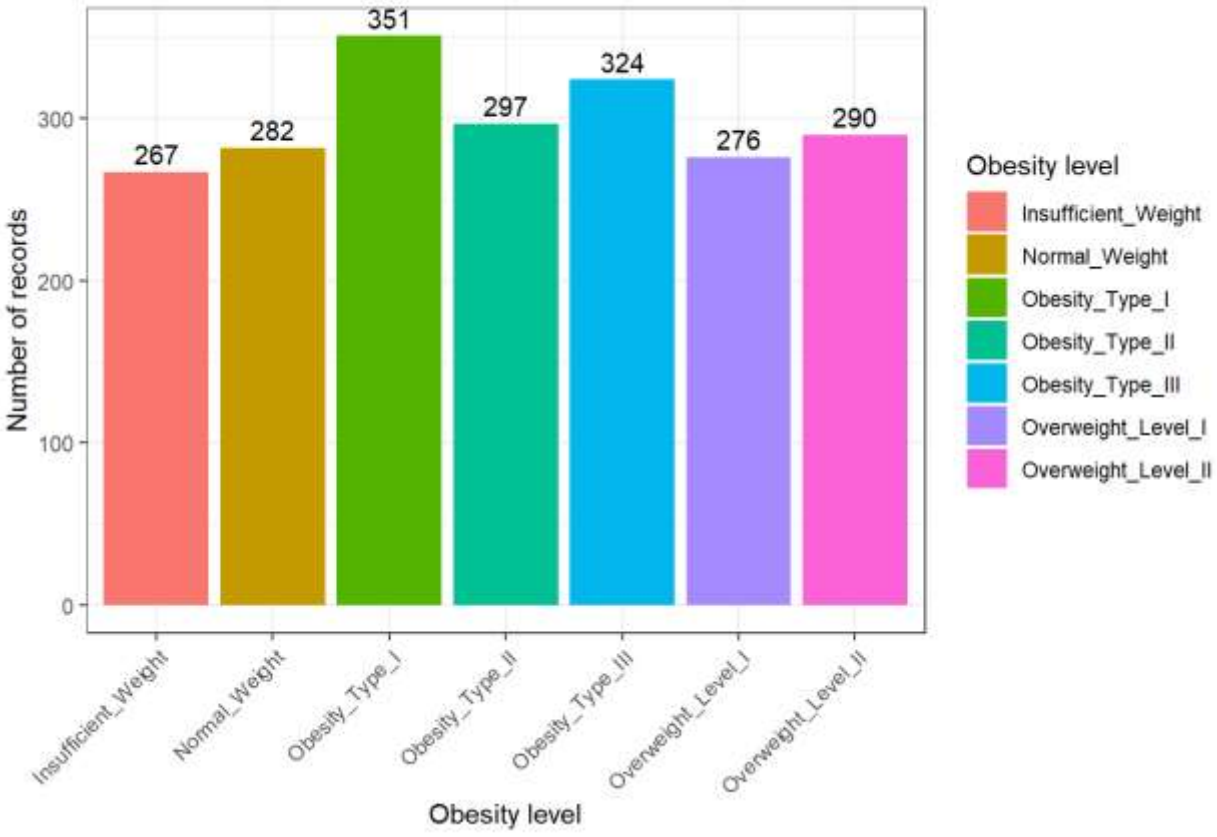


Fig3 Gender wise distribution

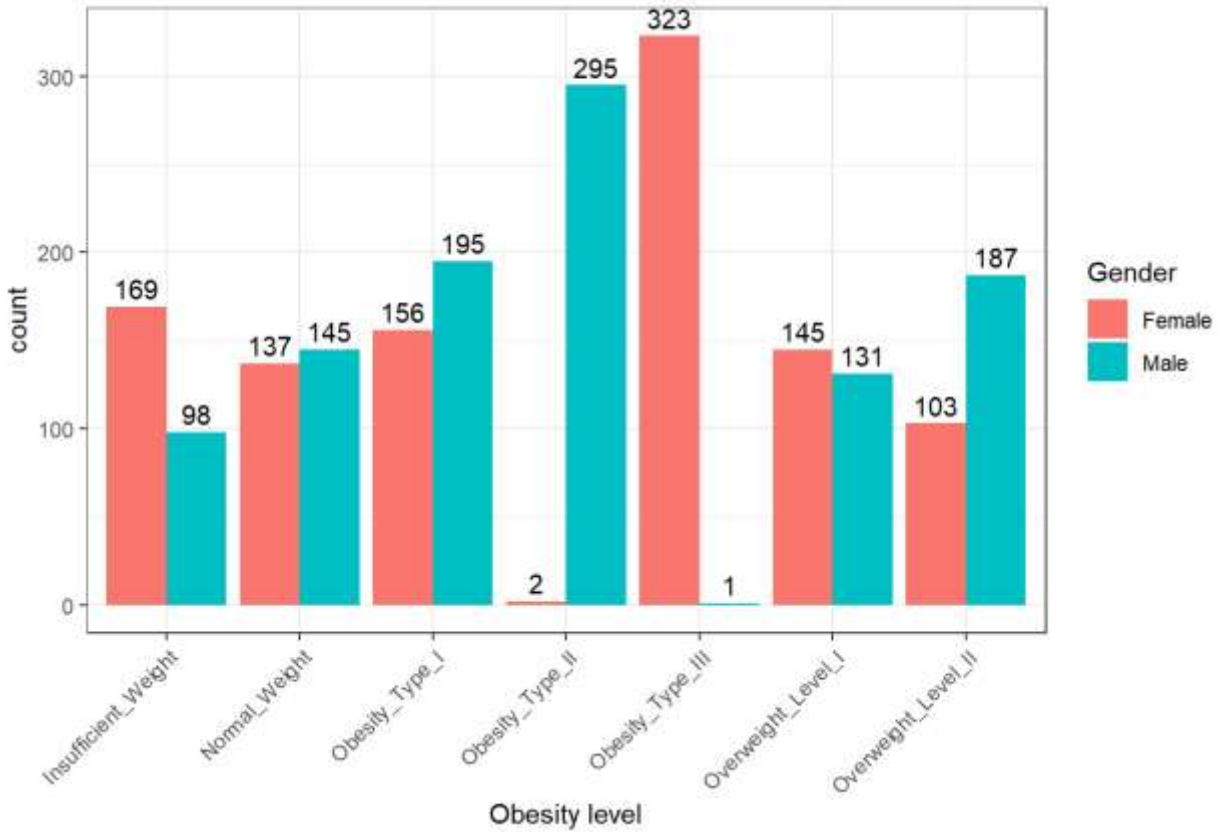
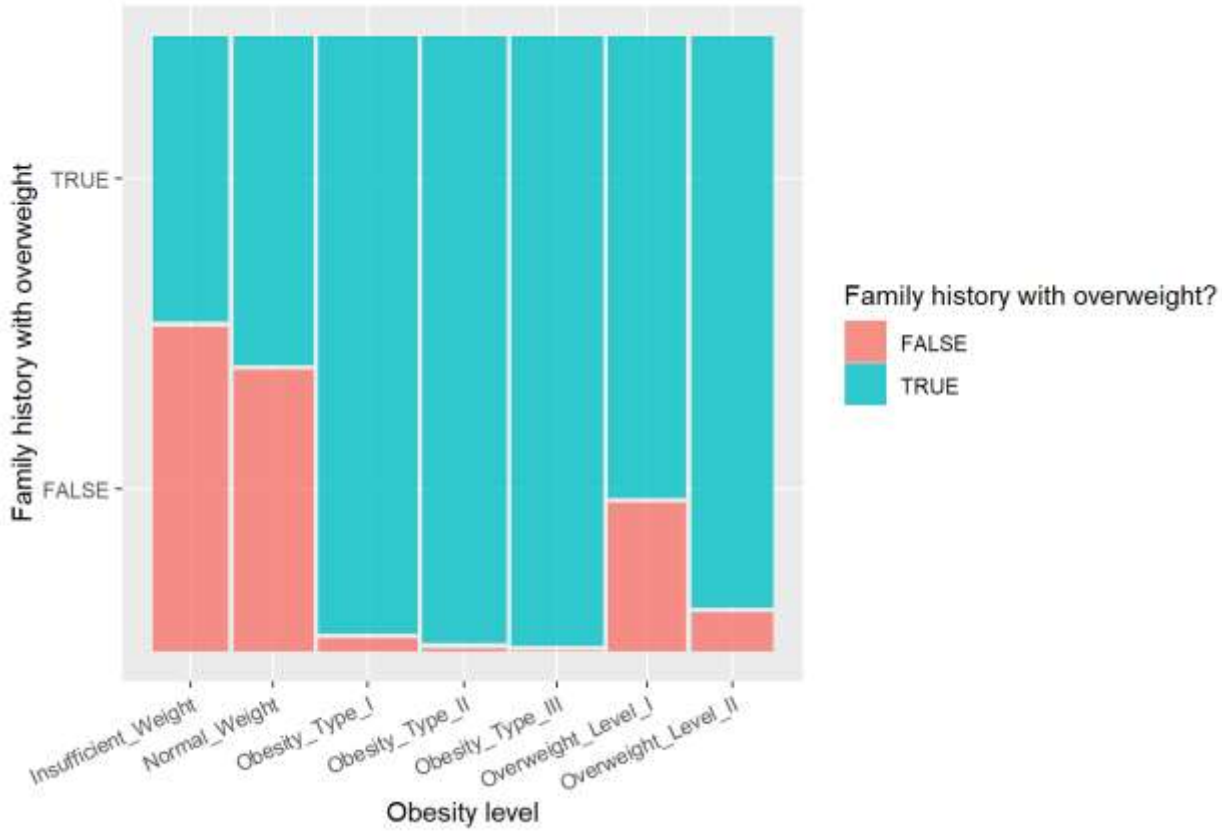


Fig4 Does family history with overweight leads to obesity level?



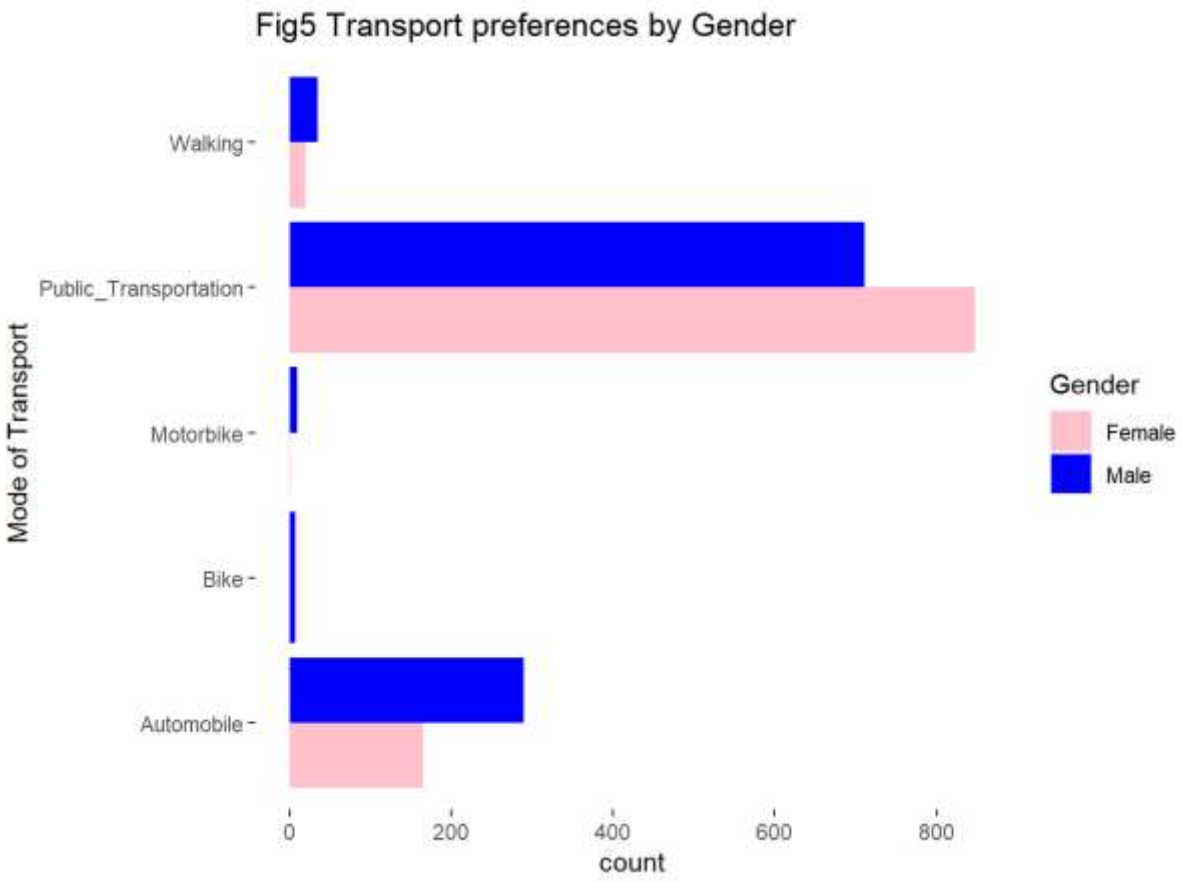


Fig5 Transport preferences by Gender
Walking
Public_Transportation
Motorbike
Bike
Automobile
Mode of Transport
0
200
400
600
800
count
Gender
Female
Male

The journey to develop this Shiny web application was both challenging and exciting. Throughout the process, I encountered numerous errors, each presenting an opportunity for learning. The experience has been valuable, equipping me with knowledge and skills in basic data visualization, application development, and problem-solving. You can access all the scripts, codes, and resources I utilized throughout the project journey on my GitHub repository at https://github.com/Chaden97/DH_Project_ChadenKhalaf .
Finally, I sincerely appreciate your contribution to this experience.

URL for Exploratory Data Analysis	https://rpubs.com/Chaden_khalaf/EDA
URL for shiny WebPage	https://chadenkhalaf.shinyapps.io/Desktop/

[1] World Health Organization. Obesity and overweight. World Health Organization. Published June 9, 2021. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight