**BMI Data Analysis**

**Data Wrangling Steps:**

The Data-Wrangling steps undertaken in this code are:

1. Importing the CSV files: The code imports the BMX D.csv and DEMO D.csv files from the provided file location.
2. Creating new Data Frames: The code generates two new data frames, "BMX new" and "DEMO new," by selecting specific columns from the original "BMX" and "DEMO" data frames.
3. Changing column names: The code modifies the column names of the "BMX new" and "DEMO new" data frames to make them more comprehensible.
4. Merging data frames: Using the "SEQN" column as the key, the code combines the two new data frames into a single data set.
5. The code categorizes the "AGE" column into several age categories.
6. The code changes the values 1 and 2 in the "GENDER" column to "Male" and "Female."
7. The code categorizes the "BMI" column as "Underweight," "Healthy weight," "Overweight," and "Obese."
8. The code classifies the "RACE/ETHNICITY" column into distinct racial and ethnic categories.
9. The code categorizes the "COUNTRY OF BIRTH" column into several categories depending on the country of birth.

**Exploration and Visualization Approach:**

Using the Shiny package, the code generates an interactive web application that enables users to explore the "BMX DEMO new" dataset by creating plots and summary statistics depending on user-specified variables. For example, the user may pick two variables from a list of dataset variables, one for the x-axis and one for the fill colour. The chosen variables are then supplied to the "plot\_func" function, which generates a bar plot of the specified variables using ggplot2. Finally, the "plotOutput" function shows the plot in the application.

In addition, the chosen variables are supplied to the "summary func" function, which utilises dplyr to group the data by the specified variables and compute the number of observations for each possible combination of the selected variables. Using the "tableOutput" function, the summary table is presented in the app.

The chosen variables are also supplied to the function "test func," which does a chi-squared test of independence on the specified variables. The programme displays the test result using the "verbatimTextOutput" method. The test is conducted only if there are sufficient observations to do the exam.

The application also has a "downloadButton" that enables the user to save the created plot as a PDF file. The filename of the PDF document is determined by the variables specified.

The code also loads many other libraries, including gtools, fitdistrplus, and shinyjs, in order to accomplish specific tasks, such as the chi-squared test, downloading the plot, and carrying out additional functionality, respectively.

The strategy for displaying the data in this code is to develop a Shiny application that enables the user to explore the connection between the variables in the dataset in an interactive manner. The application utilises the ggplot2 library to make bar graphs of the data and the dplyr library to provide summary statistics and conduct chi-squared tests of independence. For example, the user may pick two variables from the dataset, and the application will output a bar plot displaying the distribution of one variable by the other, a summary statistics table, and the results of a Cramer’s V- test of independence. This method facilitates the user's exploration of the connections between the variables in the dataset and the identification of potentially interesting patterns or trends. In addition, the user may download the data charts. Since the data is sorted and summarised by the two specified variables, this technique of data visualisation can also manage enormous data sets without becoming crowded or difficult to comprehend.

Overall, the Shiny visualisations make it simple to comprehend the data and draw conclusions from it while allowing the user to explore the correlations between various factors in the data.

**Interpretation:**

Chart, bar chart

Description automatically generated

The graph shows that the population within Age\_group 31-45 has been reported to have higher obesity than other ages.

**Chart, bar chart

Description automatically generated**

The graph shows that the female population has been reported to have higher obesity than males.

**Chart, bar chart

Description automatically generated**

The graph shows that Mexican Americans have reported the most obesity cases than the other.

**Chart, bar chart

Description automatically generated**

From the Graph, we can conclude that people born in Mexico are reported to have high Obesity than others.