# Exploratory Data Analysis of Heart Disease Dataset

**Introduction**: This project focuses on an exploratory data analysis (EDA) of the heart disease dataset. The objective is to understand the data structure, compute summary statistics, and identify key relationships between variables through correlation analysis. The insights gained from this analysis aim to enhance our understanding of heart disease and support further research.

**Dataset Overview**: The dataset comprises 303 observations and 14 variables related to heart disease. The variables include age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, and others.

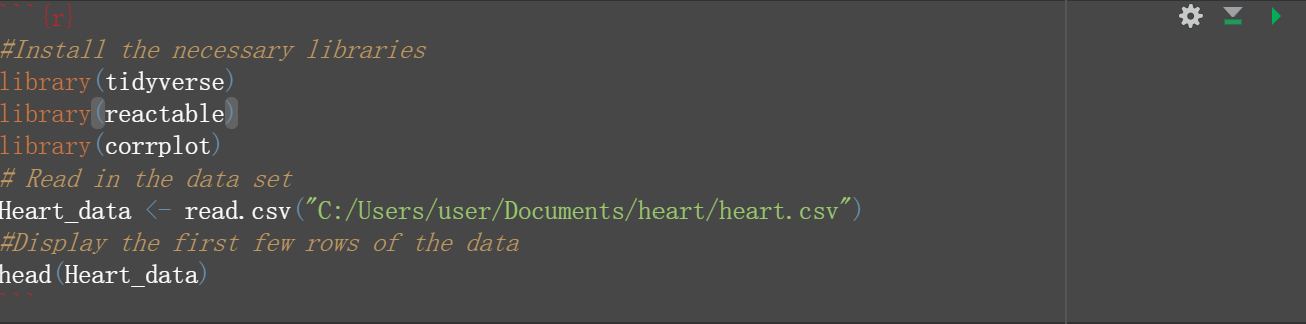


Fig 1 Screenshot of Data Overview

**Key Variables**:

* **age**: Age of the patient.
* **sex**: Gender (1 = male, 0 = female).
* **cp**: Chest pain type (1-4).
* **trtbps**: Resting blood pressure.
* **chol**: Serum cholesterol level.
* **fbs**: Fasting blood sugar (1 = > 120 mg/dl, 0 = ≤ 120 mg/dl).
* **restecg**: Resting electrocardiographic results (0-2).
* **thalachh**: Maximum heart rate achieved.
* **exng**: Exercise induced angina (1 = yes, 0 = no).
* **oldpeak**: Depression induced by exercise.
* **slp**: Slope of the peak exercise ST segment (0-2).
* **caa**: Number of major vessels colored by fluoroscopy (0-3).
* **thall**: Thalassemia (1 = normal, 2 = fixed defect, 3 = reversible defect).
* **output**: Presence or absence of heart disease (1 = presence, 0 = absence).

### **Exploratory Data Analysis**

**Summary Statistics**: I computed summary statistics to provide an overview of the dataset’s distribution.

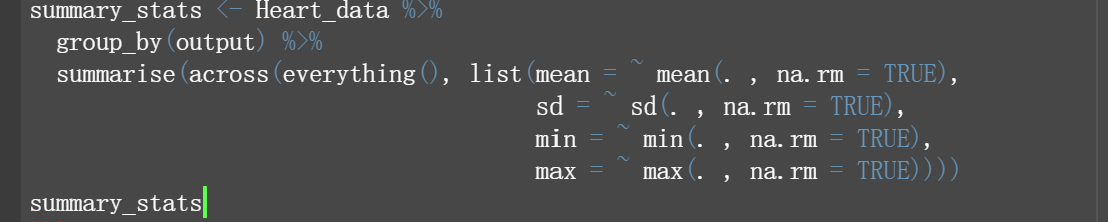


Figure 2

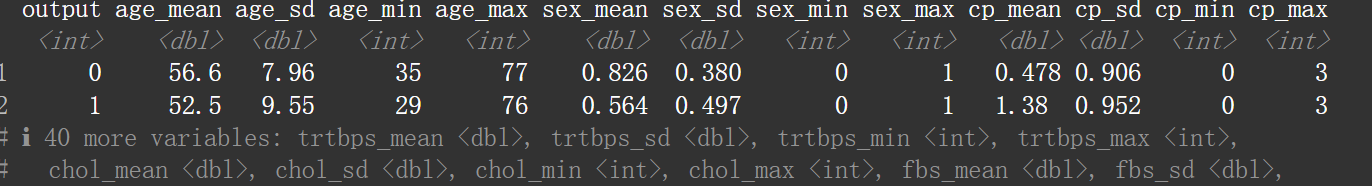


Figure 3 Screenshot of Summary Statistics

**Interactive Tables**: Interactive tables were created to present the summary statistics in a user-friendly format using reactable.

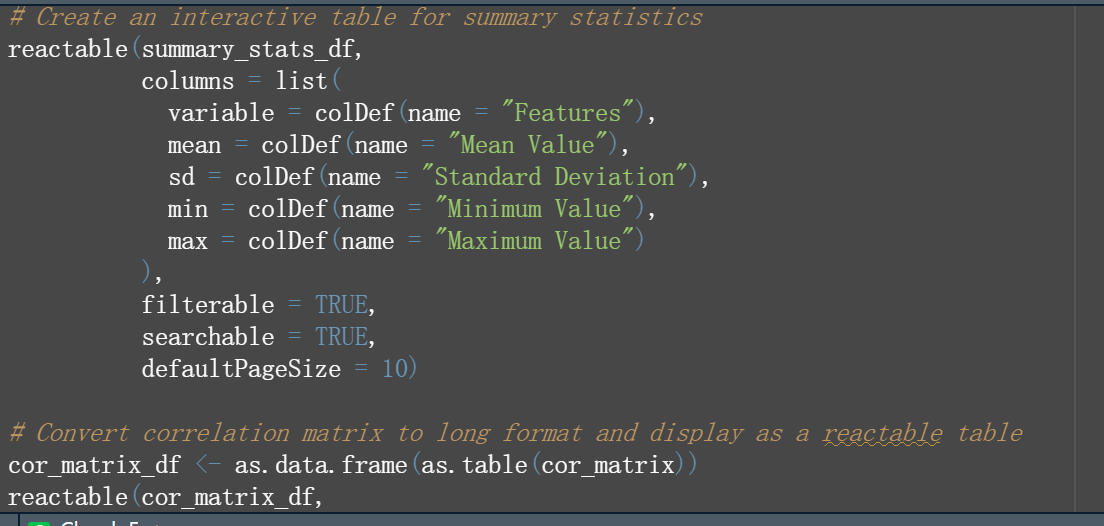
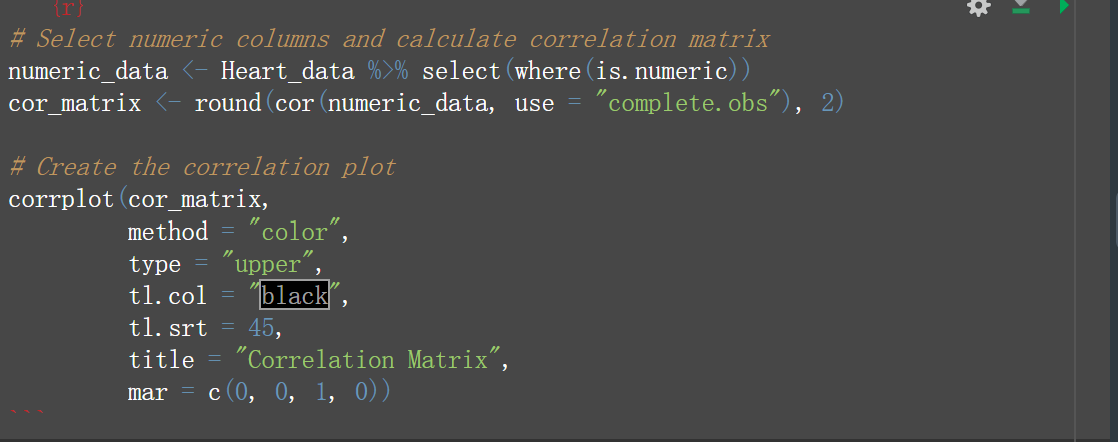


Figure 4 : A screenshot of interactive table

**Correlation Analysis**: Correlation analysis was performed to identify relationships between numeric variables.



**Interactive Correlation Table**: The correlation table provides a detailed look at the relationships between different numerical variables in your dataset. By calculating the correlation coefficients between pairs of variables, you can determine the strength and direction of their relationships.

**What the Correlation Table Represents**

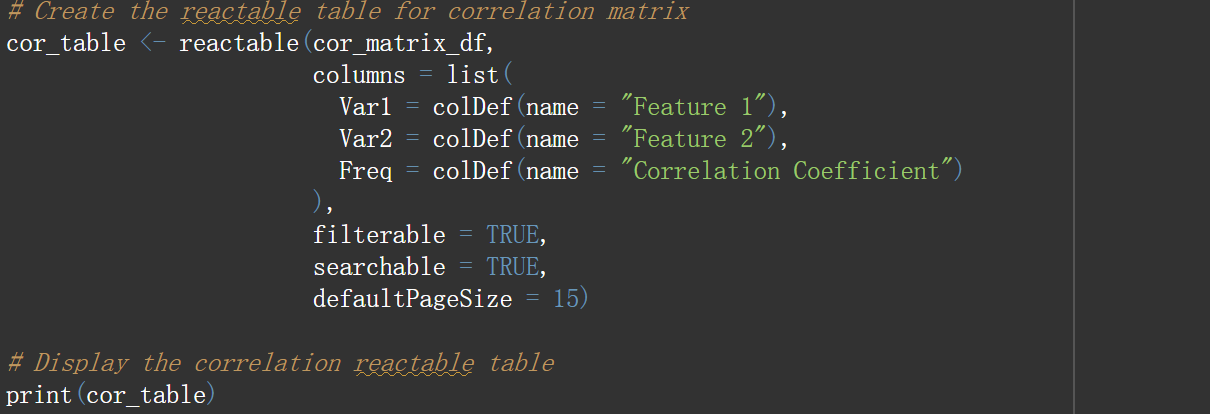
* **Correlation Coefficients**: These values range from -1 to 1. A coefficient close to 1 indicates a strong positive correlation (as one variable increases, so does the other), while a value close to -1 indicates a strong negative correlation (as one variable increases, the other decreases). A value close to 0 suggests no linear relationship between the variables.
* **Upper/Lower Triangle**: Since correlation matrices are symmetrical, typically only the upper or lower triangle is displayed to avoid redundancy. You've chosen to display the upper triangle of the matrix.

**Using the Reactable Table**

* The correlation table can be interactively displayed using reactable, which allows users to filter, search, and sort through the correlations between variables.
* This interactive table is useful when exploring specific relationships and identifying potential features for further analysis.

Top of Form

Bottom of Form



**Conclusion**

This project leveraged exploratory data analysis to dissect the heart disease dataset, revealing crucial connections among various health metrics. By generating comprehensive summary statistics and visualizing correlations, we highlighted key trends and relationships pertinent to heart disease. The interactive tables offered a user-friendly approach to navigate and interpret the data, enhancing our ability to uncover and understand underlying patterns. These insights not only contribute to a better grasp of cardiovascular health but also set the stage for future analysis and potential interventions

# References

UCI Machine Learning Repository. (2023). *Heart Disease Dataset*. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.

Wickham, H., & RStudio. (2023). *tidyverse: Easily install and load the 'tidyverse'*. CRAN.

K. Boettiger (2023). *reactable: Interactive Tables for R*. CRAN.

Wei, T., & Simko, V. (2023). *corrplot: Visualization of a Correlation Matrix*. CRAN.