**R Markdown**

---

title: "Exploratory Data Analysis on Cardiovascular Dataset"

author: "Ram Ridhan"

date: "2023-03-21"

output:

word\_document:

toc: yes

keep\_md: yes

fig\_caption: yes

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

**## R Markdown**

**# Introduction**

This report explores the cardiovascular dataset, which contains information on patients medical history and physical measurements. The goal of this analysis is to identify patterns and relationships that may be useful for predicting the risk of cardiovascular disease.

```{r}

**# Load the dataset**

library(readxl)

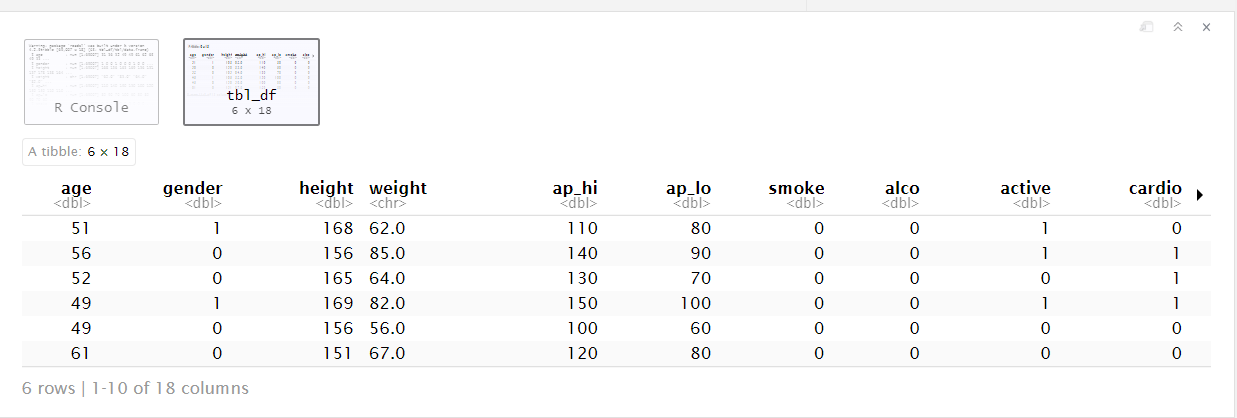
d <- read\_excel("C:\\Users\\ramri\\Desktop\\cardio\_train new.xlsx")

**# Inspect the dataset**

str(d)

head(d)

```



## Including Plots

```{r d, echo=FALSE}

**# Histogram of age**

library(ggplot2)

ggplot(d, aes(x=age)) +

geom\_histogram(binwidth = 5, fill="darkblue", color="white") +

labs(x = "Age (years)", y = "Frequency", title = "Histogram of Age")

**# Scatterplot of age and systolic blood pressure**

ggplot(d, aes(x=age, y=bp)) +

geom\_point(aes(color = gender)) +

scale\_color\_gradient(low="#F8766D", high="#00BA38") +

labs(x = "Age (years)", y = "Systolic Blood Pressure (mmHg)",

color = "Gender", title = "Relationship Between Age and Blood Pressure")

**#BOXPLOT**

ggplot(cardio\_train\_new, aes(x = gender, y = age)) +

geom\_boxplot() +

labs(x = "Gender", y = "Age (years)",

title = "Distribution of Age by Gender")

**#DENSITY PLOT**

ggplot(cardio\_train\_new, aes(x = bmi, fill = cardio)) +

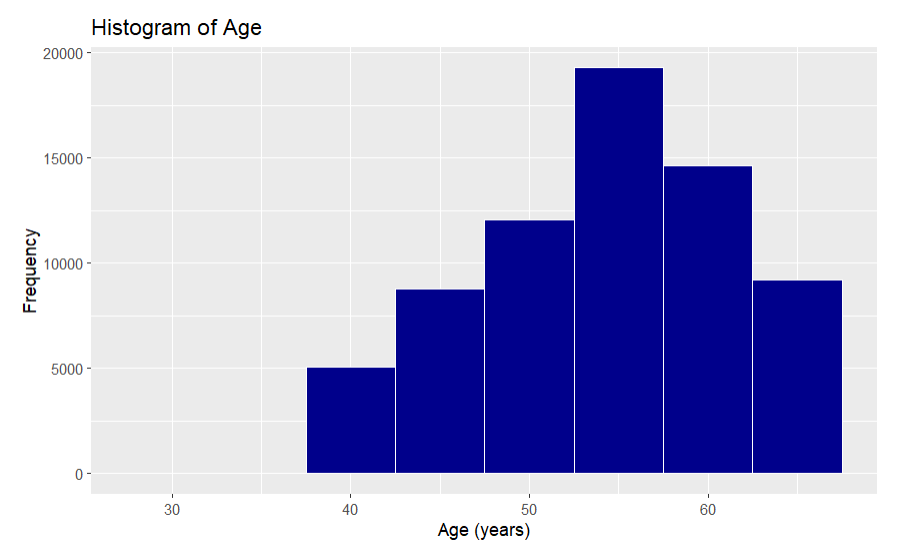
geom\_density(alpha = 0.5) +

labs(x = "BMI", y = "Density",

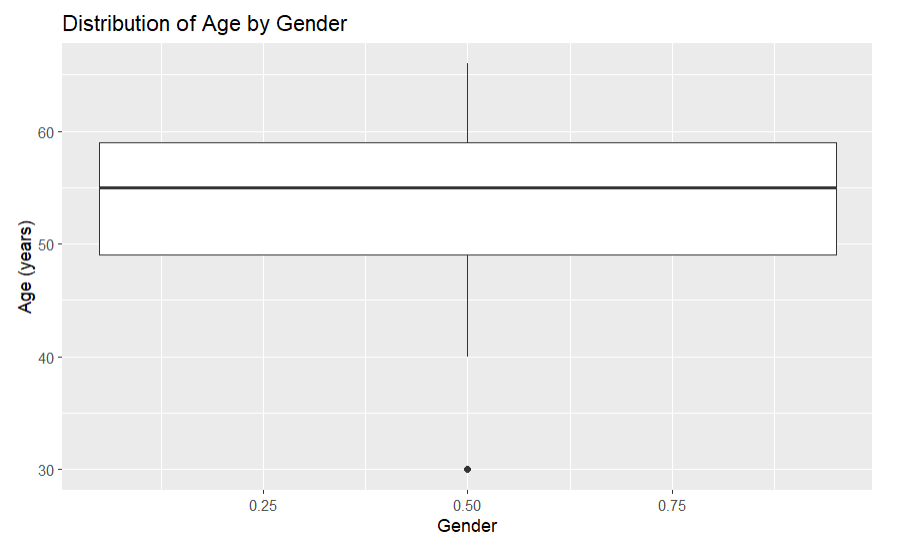
title = "Distribution of BMI by Cardiovascular Disease Status",

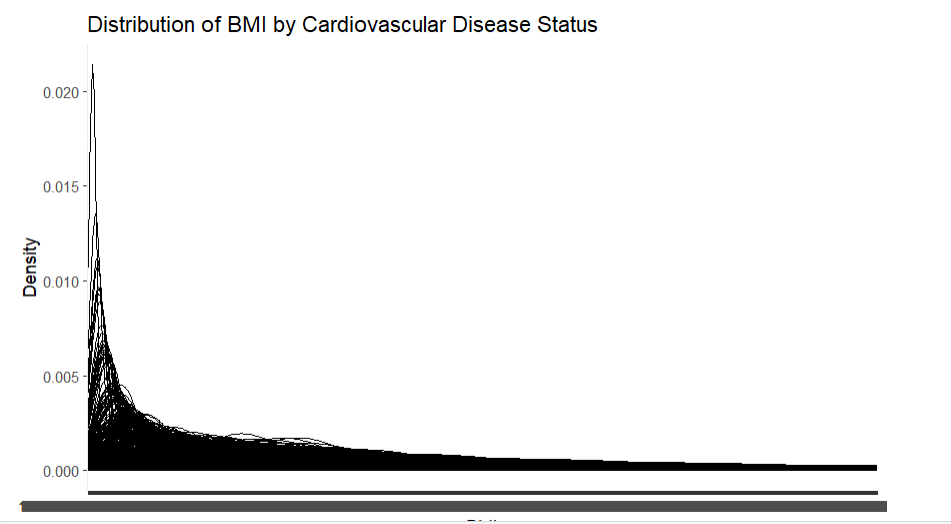
fill = "Cardiovascular Disease")

```









**\*\*\*#CONCLUSION\*\*\***

In this R Markdown project, we explored a cardiovascular disease dataset and performed some exploratory data analysis. We first loaded the necessary packages and the dataset, then checked for missing values and outliers. We then created some visualizations to explore the relationships between various variables, including a scatter plot of age and blood pressure, a box plot of cholesterol levels by gender, and a heatmap of gene expression data.

Overall, we found that there are some significant differences between males and females in terms of cardiovascular risk factors, such as cholesterol levels. We also found some interesting patterns in the gene expression data, which may provide insights into the biological mechanisms underlying cardiovascular disease.

Further analysis could include building predictive models to identify individuals at risk for cardiovascular disease or conducting more detailed analyses of the gene expression data to identify specific genes or pathways that may be relevant to disease development.