Heart Disease Analysis

A comprehensive statistical analysis of the UCI Heart Disease dataset to identify key risk factors and biomarkers associated with coronary artery disease.

Project Overview

This project analyzes cardiac catheterization data from 297 patients to predict the presence of significant heart disease (>50% vessel diameter narrowing). The analysis combines clinical symptoms, demographic data, and diagnostic test results to identify patterns that distinguish healthy individuals from those with coronary artery disease.

Medical Context

What is being predicted: Angiographic disease status based on cardiac catheterization

- No Significant Disease: < 50% vessel diameter narrowing (low risk)
- Significant Heart Disease: > 50% vessel diameter narrowing (high risk for heart attack)

Why this matters: Early identification of significant coronary blockages can prevent heart attacks through timely medical intervention (stents, bypass surgery, medication).

What This Analysis Shows

- Statistical comparison of heart disease patients vs healthy individuals
- Visualization of key biomarkers through interactive boxplots
- Identification of significant risk factors using t-tests
- Clinical relevance to cardiovascular health assessment

Key Variables Analyzed

- Age - Patient age in years
- Sex - Male/Female distribution
- Chest Pain Types - Typical angina, atypical angina, non-anginal pain, asymptomatic
- Cholesterol - Serum cholesterol levels (mg/dl)
- Maximum Heart Rate - Peak heart rate achieved during exercise testing
- Exercise-Induced Angina - Chest pain triggered by physical activity
- Fasting Blood Sugar - Diabetes indicator (>120 mg/dl)

Dataset Information

- Source: UCI Machine Learning Repository - Heart Disease Dataset
- Original size: 303 patients (Cleveland database)
- After cleaning: 297 patients (removed missing values)
- Features: 13 clinical and demographic variables
- Target: Angiographic disease status (binary classification)

Output Files

- heart_disease_data.csv - The cleaned dataset with interpretable labels
- analysis_results.csv - Summary table of all statistical tests

How to Run

1. Clone the repository:

   ```
   git clone https://github.com/yourusername/heart-disease-analysis.git
   cd heart-disease-analysis
   ```

2. Run the R script:

   ```
   source("heart_disease_analysis.R")
   ```