

**Using Exploratory Data Analysis and Visualization to Predict Heart Diseases**

**Prepared By:**

Hamza Halaweh 20190058

Noura Abdelhafez 20191005

Thabet Ghaleb

Tasneem Sanuri

Kholoud Qubbaj

Baraa Tahat

**Course Name:** Data Visualization 14364

**Tutor Name:** Osama Abdelhay

**Table of Contents**

[3. Discussion 2](#_Toc92994600)

[3.1. Main Statistical Features 2](#_Toc92994601)

[3.2. Central Statistical Relationships and Correlations 2](#_Toc92994602)

[3.4. Data Visuals 4](#_Toc92994603)

[3.4.1. Scatter and Point Plot 4](#_Toc92994604)

[3.4.2. Percentage Plots 5](#_Toc92994605)

[3.4.3. Tile Plots 8](#_Toc92994606)

# 3. Discussion

## 3.1. Main Statistical Features

Table 1: Main Statistical Features

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | age | trestbps | chol | thalach | exang | cp |
| Mean | 54.4 | 131.7 | 246.7 | 149.6 | 1 | 0.7 |
| Standard Deviation | 9 | 17.6 | 51.8 | 22.9 | 1.2 | 0.9 |
| Minimum | 29 | 94 | 126 | 71 | 0 | 0 |
| 25% | 48 | 120 | 211 | 133.5 | 0 | 0 |
| 50% | 56 | 130 | 241 | 153 | 0.8 | 0 |
| 75% | 61 | 140 | 275 | 166 | 1.6 | 1 |
| Maximum | 77 | 200 | 564 | 202 | 6.2 | 3 |

## 3.2. Central Statistical Relationships and Correlations

Table 2: Correlation Scale

|  |  |
| --- | --- |
| Size of Correlation | Interpretation |
| 0.9 to 1.00 (-0.9 to -0.9) | Very high positive (negative) correlation |
| 0.7 to 0.9 (-0.7 to -0.9) | High positive (negative) correlation |
| 0.5 to 0.7 (-0.5 to -0.7) | Moderate positive (negative) correlation |
| 0.3 to 0.5 (-0.3 to -0.5) | Low positive (negative) correlation |
| 0.0 to 0.3 (-0.0 to -0.3) | Negligible correlation |

Table 3: Pearson's Correlation

|  |  |  |  |
| --- | --- | --- | --- |
| Variable 1 | Variable 2 | Correlation | Identification |
| trestbps | age | 0.28 | Negligible correlation |
| age | chol | 0.21 | Negligible correlation |
| age | oldpeak | 0.20 | Negligible correlation |
| age | thalach | -0.39 | Low negative correlation |
| thalach | oldpeak | -0.34 | Low negative correlation |

Table 4: Spearman's Correlation

|  |  |  |  |
| --- | --- | --- | --- |
| Variable 1 | Variable 2 | Correlation | Identification |
| slope | oldpeak | 0.59 | Moderate positive correlation |
| thal | outcome | 0.53 | Moderate positive correlation |
| ca | outcome | 0.53 | Moderate positive correlation |
| cp | outcome | 0.49 | Low positive correlation |
| oldpeak | outcome | 0.46 | Low positive correlation |
| exang | outcome | 0.44 | Low positive correlation |
| slope | outcome | 0.40 | Low positive correlation |
| thalack | cp | -0.38 | Low negative correlation |
| thalack | age | -0.39 | Low negative correlation |
| thalack | exang | -0.40 | Low negative correlation |
| thalack | oldpeak | -0.43 | Low negative correlation |
| slope | thalack | -0.44 | Low negative correlation |
| thalack | outcome | -0.44 | Low negative correlation |

Table 5: Chi-squared Test

|  |  |  |  |
| --- | --- | --- | --- |
| x | y | p-value | result |
| outcome | cp | <2.2e-16 | dependent |
| outcome | thal | <2.2e-16 | dependent |
| outcome | ca | 1.17E-15 | dependent |
| cp | exang | 3.05E-14 | dependent |
| outcome | exang | 5.57E-14 | dependent |
| sex | thal | 1.11E-10 | dependent |
| outcome | slope | 1.14E-10 | dependent |
| exang | thal | 1.08E-07 | dependent |
| slope | thal | 4.54E-07 | dependent |
| cp | thal | 5.61E-07 | dependent |
| cp | thal | 5.61E-07 | dependent |
| outcome | sex | 1.45E-06 | dependent |
| exang | slope | 3.49E-06 | dependent |
| cp | slope | 0.000161 | dependent |
| cp | ca | 0.000284 | dependent |
| ca | thal | 0.000992 | dependent |
| outcome | restecg | 0.002214 | dependent |
| exang | ca | 0.005208 | dependent |
| sex | exang | 0.010930 | dependent |
| restecg | slope | 0.031590 | dependent |

## 3.4. Data Visuals

## 3.4.1. Scatter and Point Plot

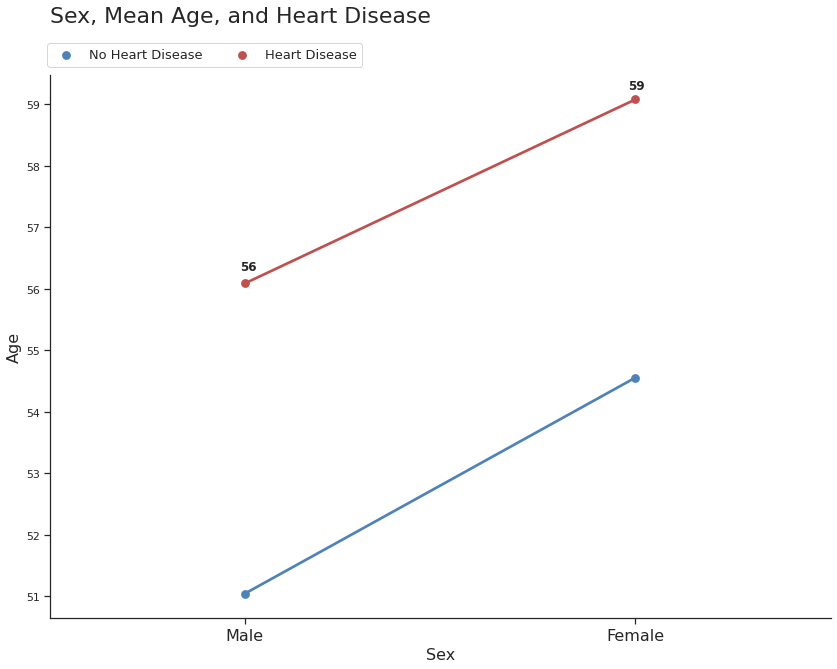


Figure 1: Sex, Mean Age, and Heart Disease

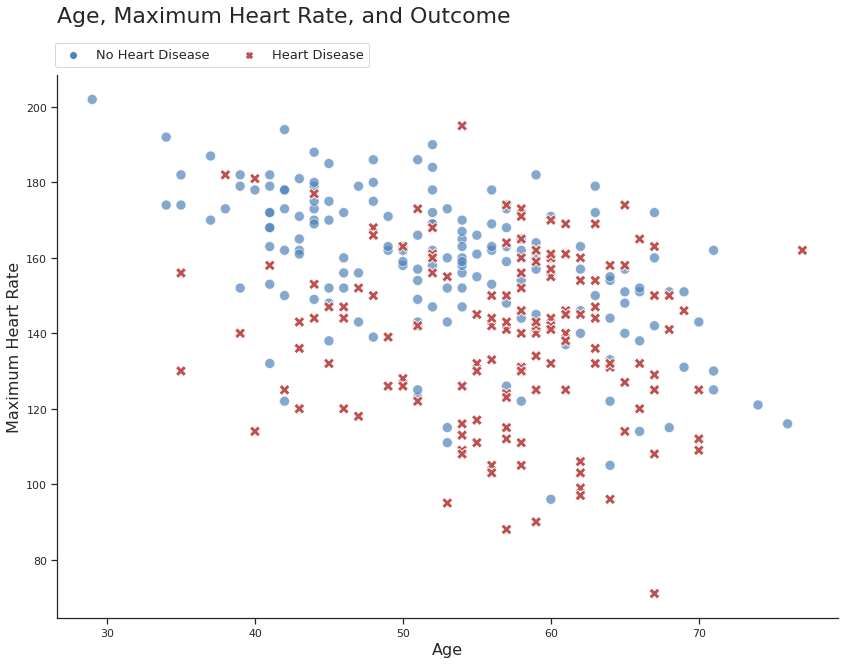


Figure 2: Age, Maximum Heart Rate, and Outcome

## 3.4.2. Percentage Plots

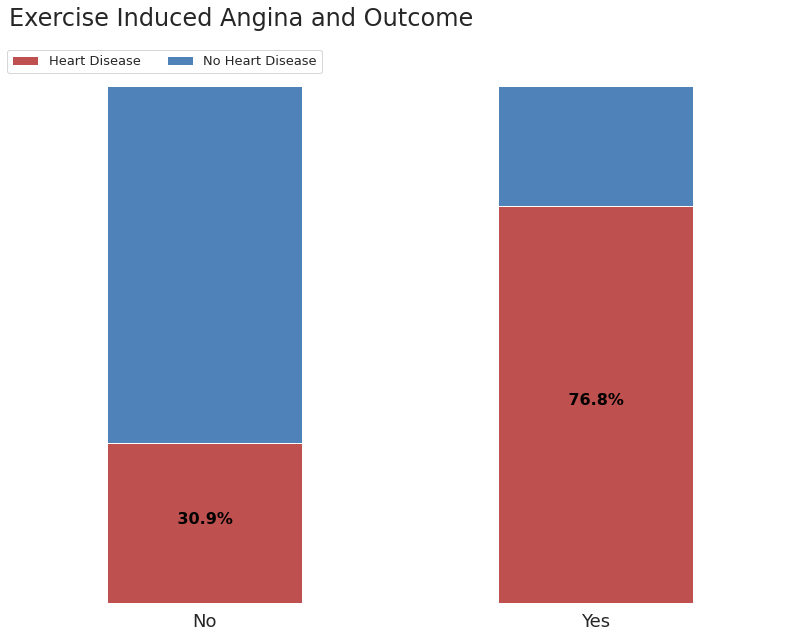


Figure 3: Exercise Induced Angina and Outcome

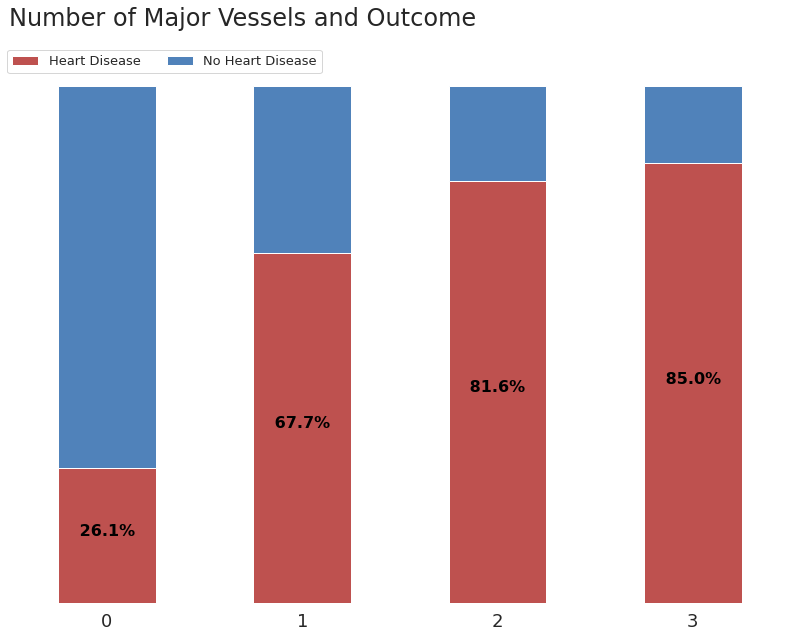


Figure 4: Number of Major Vessels and Outcome

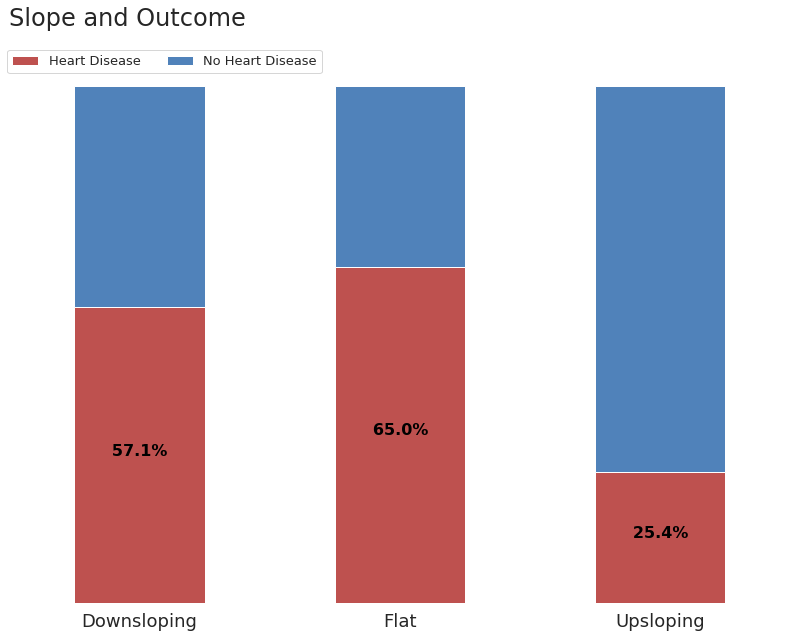


Figure 5: Slope and Outcome

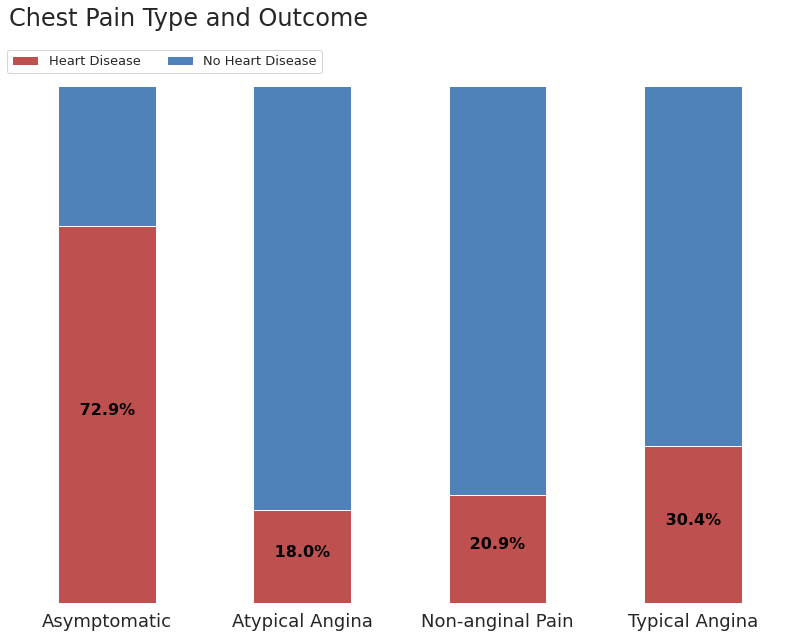


Figure 6: Chest Pain Type and Outcome

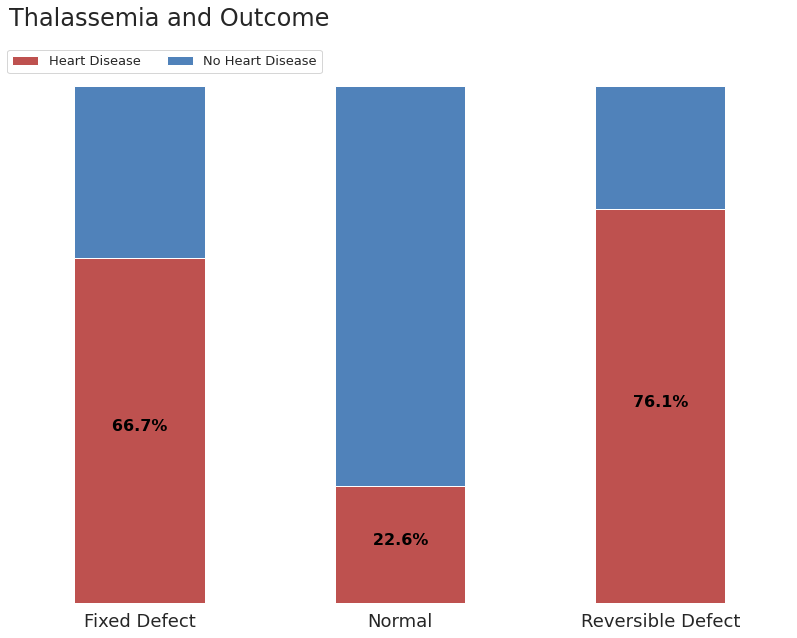


Figure 7: Thalassemia and Outcome

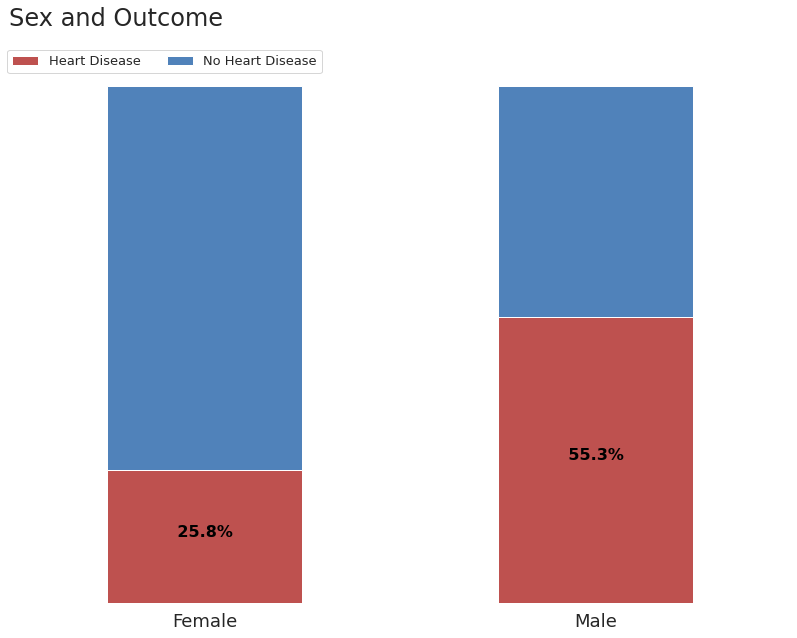


Figure 8: Sex and Outcome

## 3.4.3. Tile Plots

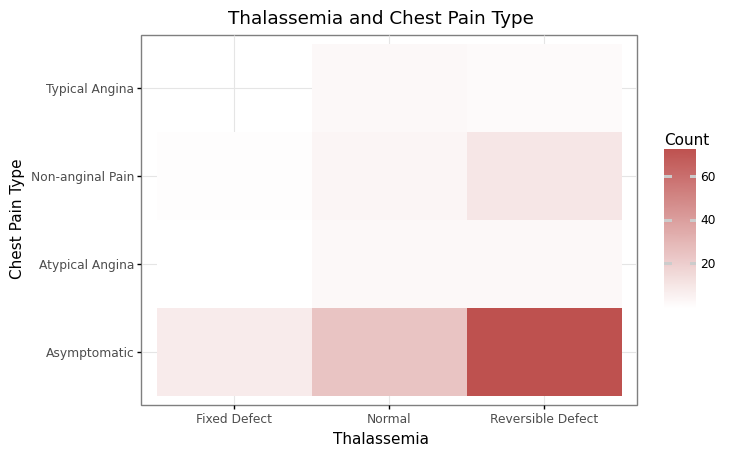


Figure 9: Thalassemia and Chest Pain Type

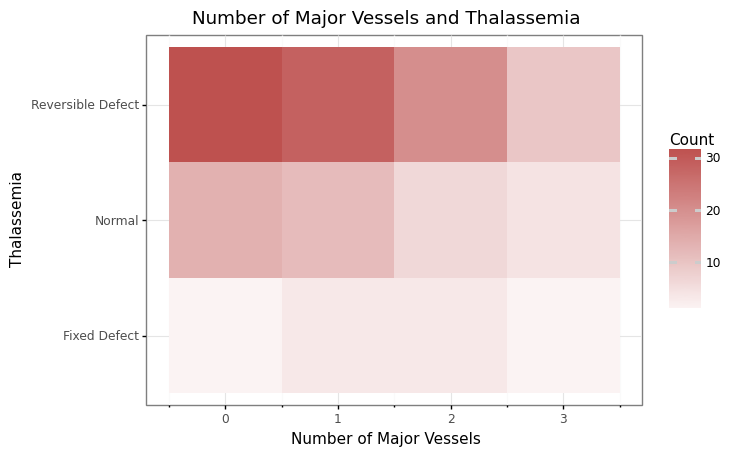


Figure 10: Number of Major Vessels and Thalassemia

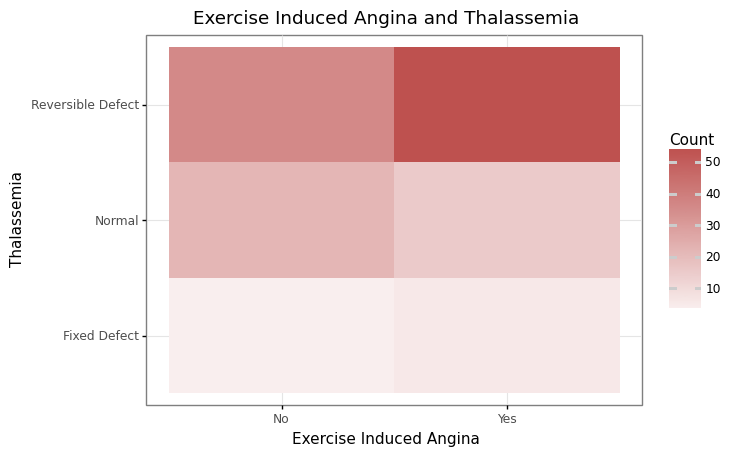


Figure 11: Exercise Induced Angina and Thalassemia

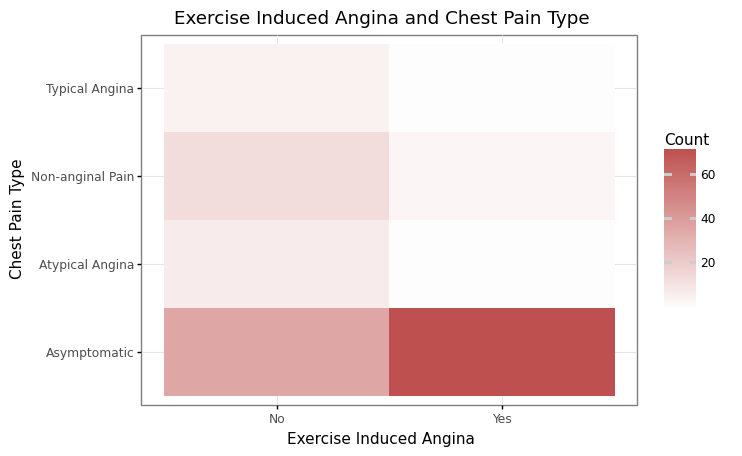


Figure 12: Exercise Induced Angina and Chest Pain Type