Zane Alderfer

Prof. Krudys

IST 719

4/23/24

Week 3 homework

Heart Disease Dataset

Description of Dataset

This dataset is a combination of 5 heart disease datasets that are combined over 11 common features relating to heart disease. The dataset has 1190 observations of individuals with heart disease to identify commonalities and differences within the variables among the patients.

str(hearts)

'data.frame': 1190 obs. of 12 variables:

$ age : int 40 49 37 48 54 39 45 54 37 48 ...

$ sex : int 1 0 1 0 1 1 0 1 1 0 ...

$ chest.pain.type : int 2 3 2 4 3 3 2 2 4 2 ...

$ resting.bp.s : int 140 160 130 138 150 120 130 110 140 120 ...

$ cholesterol : int 289 180 283 214 195 339 237 208 207 284 ...

$ fasting.blood.sugar: int 0 0 0 0 0 0 0 0 0 0 ...

$ resting.ecg : int 0 0 1 0 0 0 0 0 0 0 ...

$ max.heart.rate : int 172 156 98 108 122 170 170 142 130 120 ...

$ exercise.angina : int 0 0 0 1 0 0 0 0 1 0 ...

$ oldpeak : num 0 1 0 1.5 0 0 0 0 1.5 0 ...

$ ST.slope : int 1 2 1 2 1 1 1 1 2 1 ...

$ target : int 0 1 0 1 0 0 0 0 1 0 …

Plot 1: Distribution of Age

Plot 2: Distribution of Cholesterol Levels

Plot 3: Frequency of Chest Pain Types

Plot 4: Age vs. Max Heart Rate (Colored by Target)

Below is a description of numeric variables for better understanding of the graphs

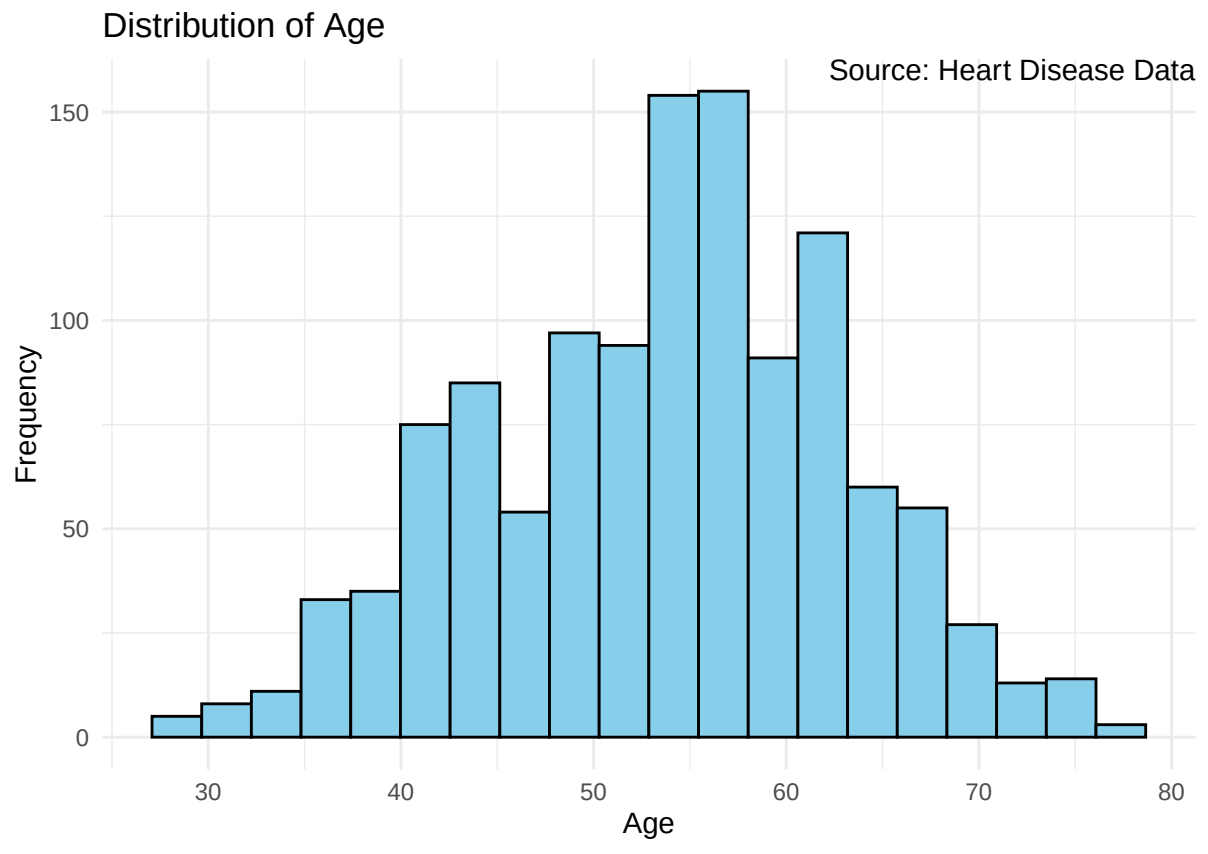
Description of Nominal Attributes

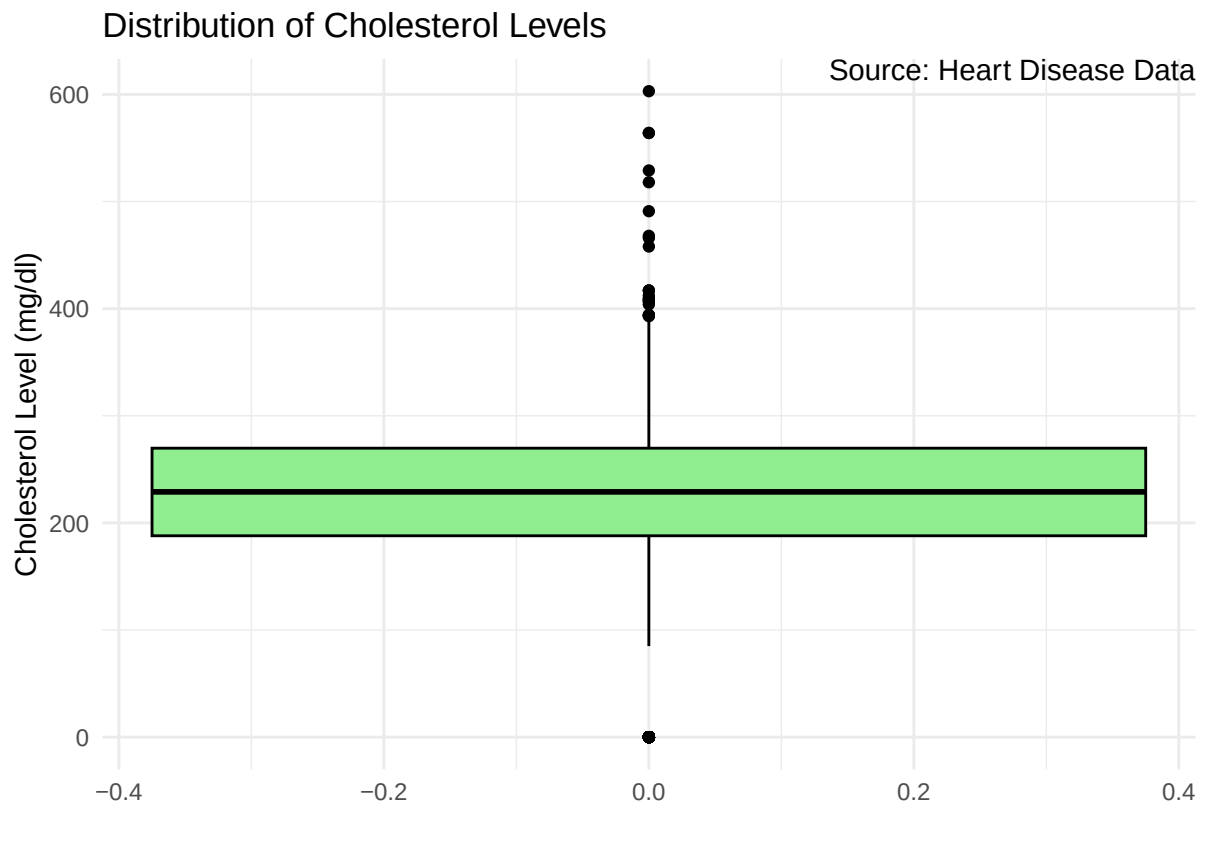
Attribute	Description
Sex	1 = male, 0= female;
Chest Pain Type	-- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
Fasting Blood sugar	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Resting electrocardiogram results	-- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
Exercise induced angina	1 = yes; 0 = no
the slope of the peak exercise ST segment	-- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
class	1 = heart disease, 0 = Normal

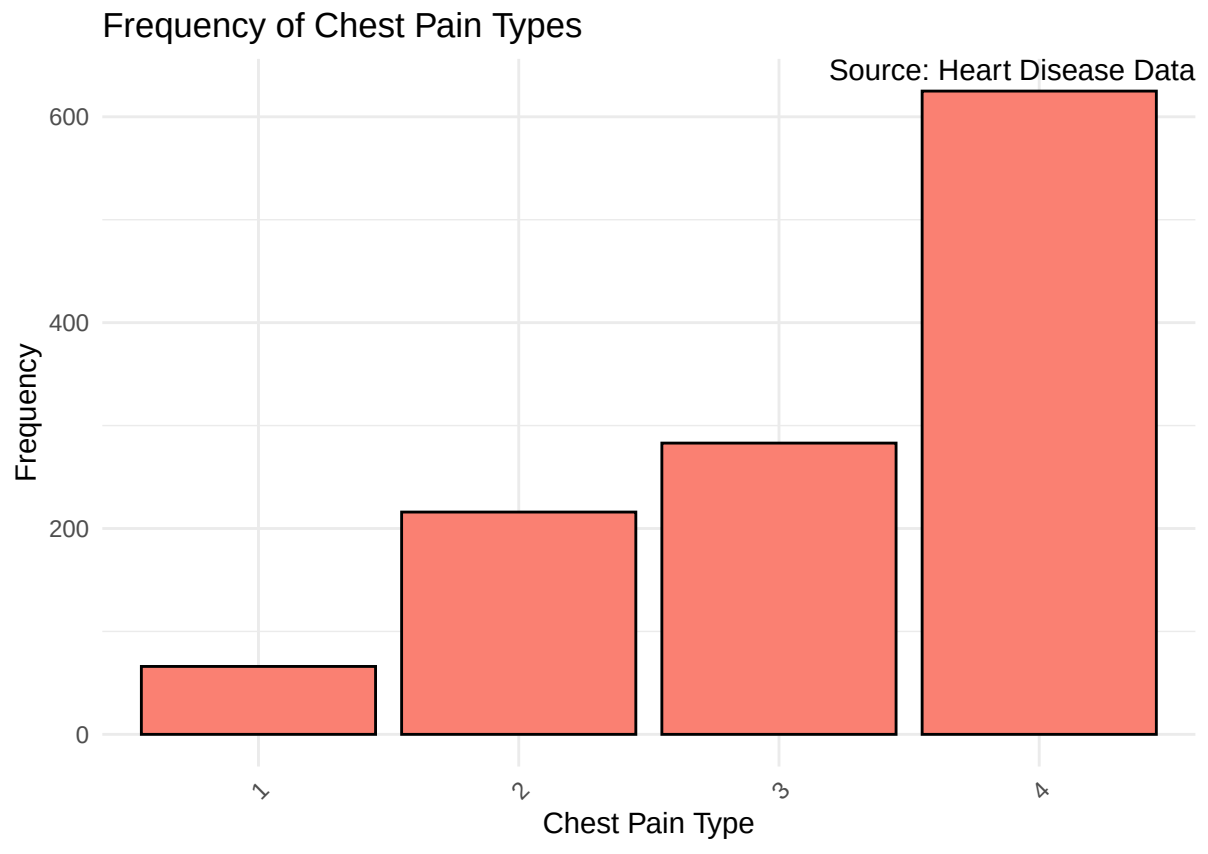
Source:
https://www.kaggle.com/datasets/mexwell/heart-disease-dataset?resource=download&select=documentation.pdf

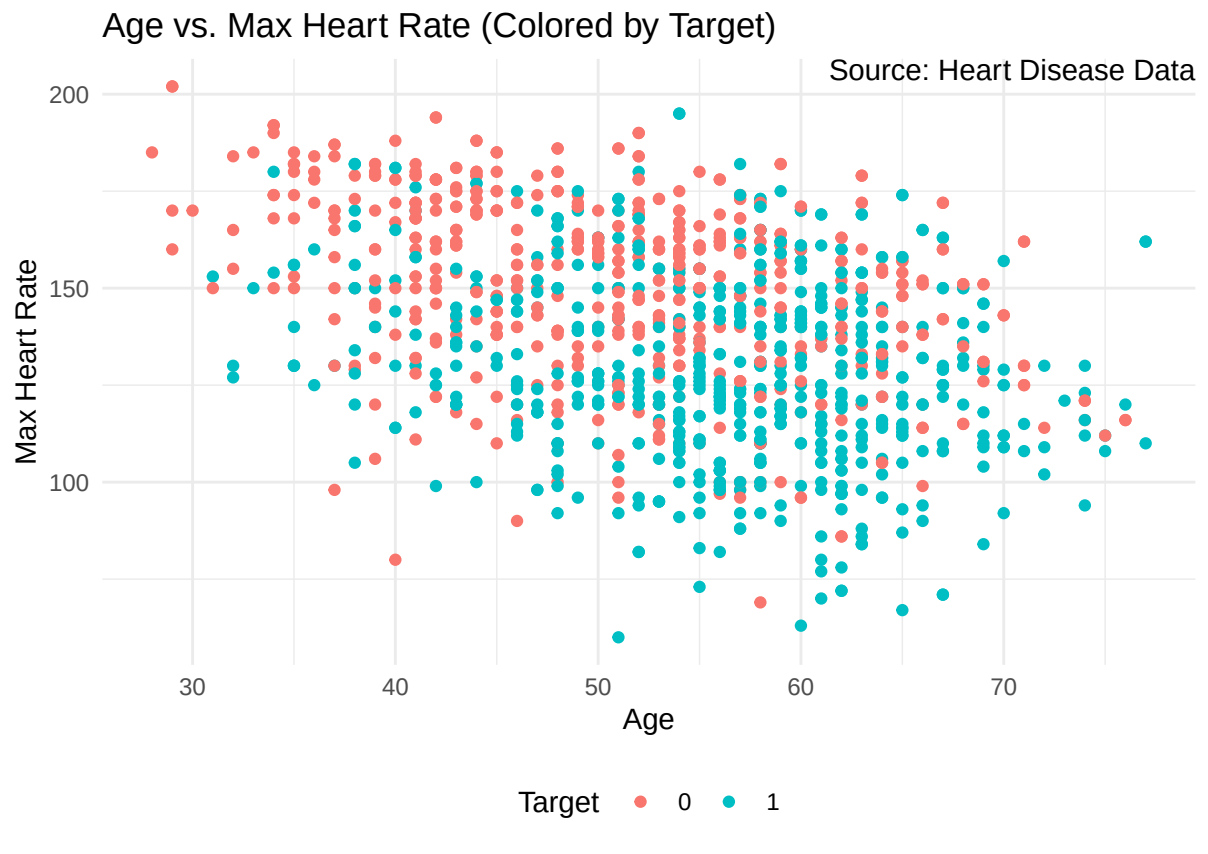
(NumberOfColumns * 4) * (NumberOfRows/100) >= 100

(11*4) * (1190/100) = 44 * 11.9 = 523.6


Distribution of Age
Source: Heart Disease Data
Frequency
150
100
50
0
30
40
50
60
70
80
Age


Distribution of Cholesterol Levels
Source: Heart Disease Data
Cholesterol Level (mg/dl)
600
400
200
0
−0.4
−0.2
0.0
0.2
0.4


Frequency of Chest Pain Types
Source: Heart Disease Data
600
400
200
0
Frequency
1
2
3
4
Chest Pain Type


Age vs. Max Heart Rate (Colored by Target)
Source: Heart Disease Data
Max Heart Rate
200
150
100
30
40
50
60
70
Age
Target
0
1