

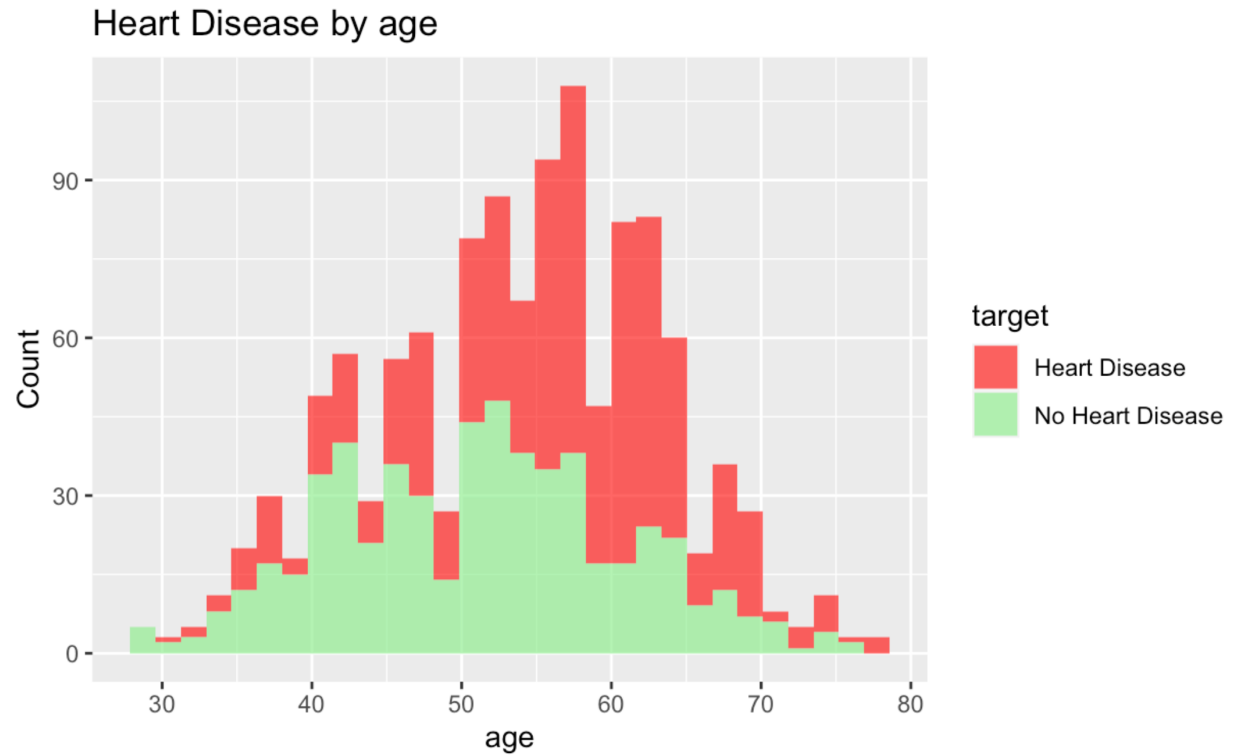
Data Visualization of Heart Disease dataset

The dataset is [Heart Disease UCI dataset from Kaggle](#)[Links to an external site.](#) The heart disease dataset contains information on patients who have been diagnosed with heart disease, as well as patients who have no evidence of heart disease. The dataset has a total of 12 columns, including both demographic and medical information.

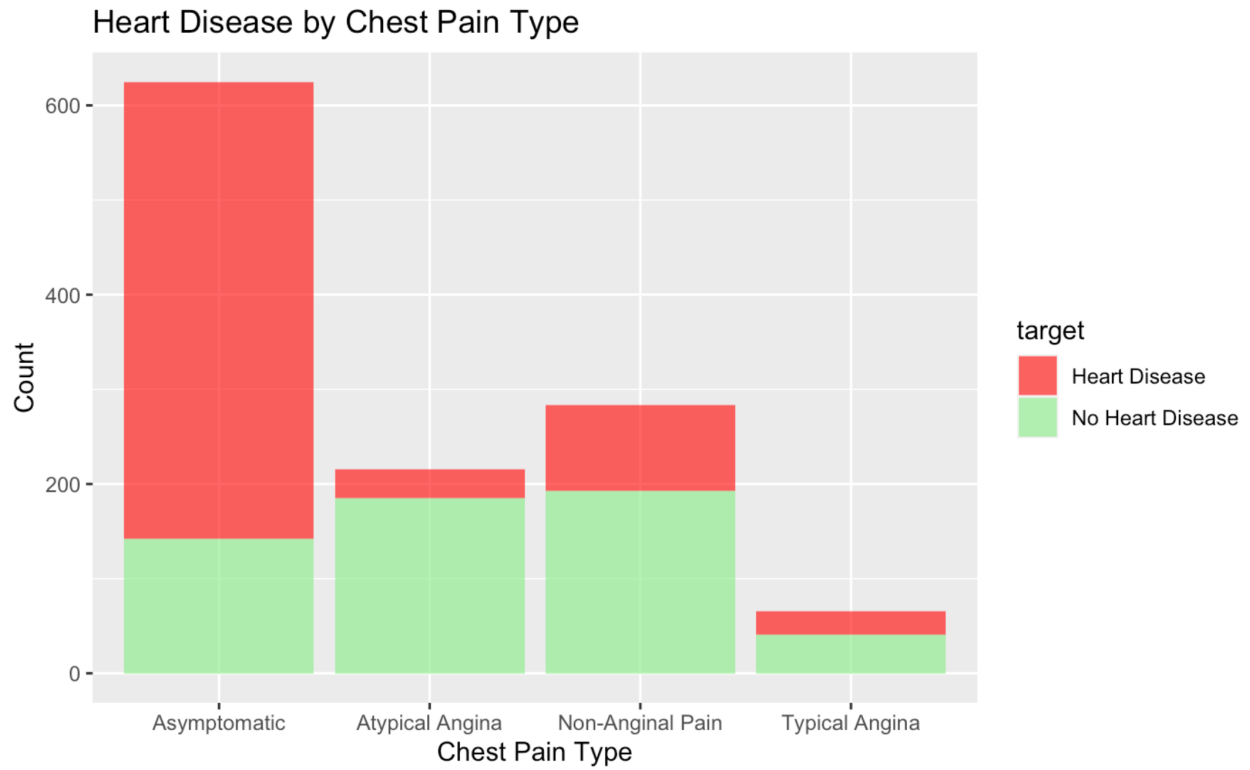
In this analysis, I was interested in exploring heart disease risk by age and chest pain type and heart disease prevalence by sex. To perform the analysis, I loaded the necessary libraries to read and clean the data in R. I performed some preprocessing steps to tidy and clean the data, including renaming the columns with suitable names, recoding the variables to make them easier to interpret, and imputing any missing values with meaningful values. The heart disease dataset had no missing values. After the cleaning process, the dataset contained 12 variables of interest. These variables included:

- `age`: age of the patient [years]
- `sex`: sex of the patient [1: Male, 0:Female]
- `chest_pain_type`: chest pain type [1: Typical Angina, 2:Atypical Angina, 3: Non-Anginal Pain, 4: Asymptomatic]
- `resting_bp`: resting blood pressure [mm/Hg]
- `cholesterol`: serum cholesterol [mm/dl]
- `fasting_bs`: fasting blood sugar [1: if FastingBS \geq 120 mg/dl, 0:otherwise]
- `resting_ecg`: resting electrocardiogram results [0:Normal, 1: ST-T wave abnormality, 2: Left ventricular hypertrophy]
- `max_heart_rate`: maximum heart rate achieved [Numeric value]
- `exercise_angina`: exercise-induced angina [1: Yes, 0: No]
- `oldpeak`: oldpeak = ST [Numeric value measured in depression]
- `ST_slope`: the slope of the peak exercise ST segment [0: Normal, 1: Upsloping, 2: Flat, 3: Downsloping]
- `target`: output class [1: heart disease, 0: Normal]

The exploratory data analysis revealed several key observations. First, the visualization of heart disease distribution by age showed that the majority of patients with heart disease tend to be older, with the highest risk appearing to be after the age of 50. The age group of 50-65 showed a higher proportion of individuals with heart disease, which indicates that such individuals are at higher risk of developing a heart disease.



Second, the visualization of the heart disease distribution by chest pain type showed that the majority of the patients with heart disease had asymptomatic chest pain followed by non-anginal pain, Atypical Angina and Typical Angina, which is interesting that most patients showing no symptoms of chest pain can develop heart disease, which is indicative of a need for routine screening for heart ailments.



Third, the visualization of heart disease distribution by sex revealed that men are at higher risk of developing heart diseases than women and a greater proportion of men tend to be diagnosed with heart diseases than women.

