Mini Project Report

On

**Heart Disease Detection Using R-Programming**

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Engineering

By

Gargi Virendra Kumbhare [ 38 ]

Rushikesh Sunildatt Mahadik [40]

Omkar Prabhakar Pawar [56]

Pradnya Janardan Pugaonkar[ 58 ]

Rajguru Shashikant Bhagyshree[59]

**Supervisor**

**Dr. S. N. Mahtre**



Department of Information Technology

Bharati Vidyapeeth College of Engineering, Navi Mumbai

Academic Year 2021-22



Bharati Vidyapeeth College of Engineering, Navi Mumbai

Department of Information Technology

**CERTIFICATE**

*This is to certify that,*

Gargi Virendra Kumbhare [ 38 ]

Rushikesh Sunildatt Mahadik [40]

Omkar Prabhakar Pawar [56]

Pradnya Janardan Pugaonkar[ 58 ]

Rajguru Shashikant Bhagyshree[59]

*Class- BEIT Semester-VIII have completed the Mini Project “****Heart Disease Detection Using R****-****Programming****”**Satisfactorily in the Department of Information Technology as prescribed by the Mumbai University in the academic year 2021-2022.*

Prof. S. N. Mhatre

(Mini Project Guide)

Prof. H. B. Sale Dr. S. D. Jadhav

Head Principal

**Table of Contents**

1. Introduction………………………………………………..3
   1. Abstract………………………………………………...3
2. Methodology………………………………………………3
3. Implementation…………………………………………….4

3.1 Diagnosis of Heart Disease…………………………….5

3.2 Data…………………………………………………….9

3.3 Data Transformation…………………………………..12

3.4 Data Visualization……………………………………..12

3.5 Presence and Absence of Heart Disease……………….13

3.6 Age Analysis

1. Result
2. Conclusion
3. Future Work
4. **Introduction:**

1.1 Abstract:

We are in a period of the “Information Age” where the traditional industry can pressure the rapid shift to the industrial revolution for industrialization, based on the economy of information technology Terabytes of data are produced and stored in day-to-day life because of fast growth in “Information Technology”.

Terabytes of data are produced and stored in day-to-day life because of the fast growth in “Information Technology”. The data which is collected is converted into knowledge by data analysis using various combinations of algorithms. For example, a huge amount of data regarding the patients is generated by the hospitals such as x-ray results, lungs results, heart paining results, chest pain results, personal health records (PHRs), etc. There is no effective use of the data which is generated from the hospitals. Some certain tools are used to extract the information from the database for the detection of heart diseases and other functions are not accepted.

The main aim of the project is the prediction of heart diseases using machine learning techniques by summarizing the few current types of research. In this project, the logistic regression algorithms are used and the health care data classifies the patients whether they are having heart diseases or not according to the information in the record. Also, we will try to use this data as a model which predicts the patient whether they are having heart disease or not.

1. **Methodology:**

**Logistic regression** is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as 1 (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.).

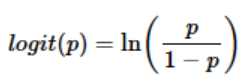
The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a *logit transformation* of the probability of presence of the characteristic of interest:

https://miro.medium.com/max/1048/1*5ZDCsRqMRFHVNQCXXfdj4A.png

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

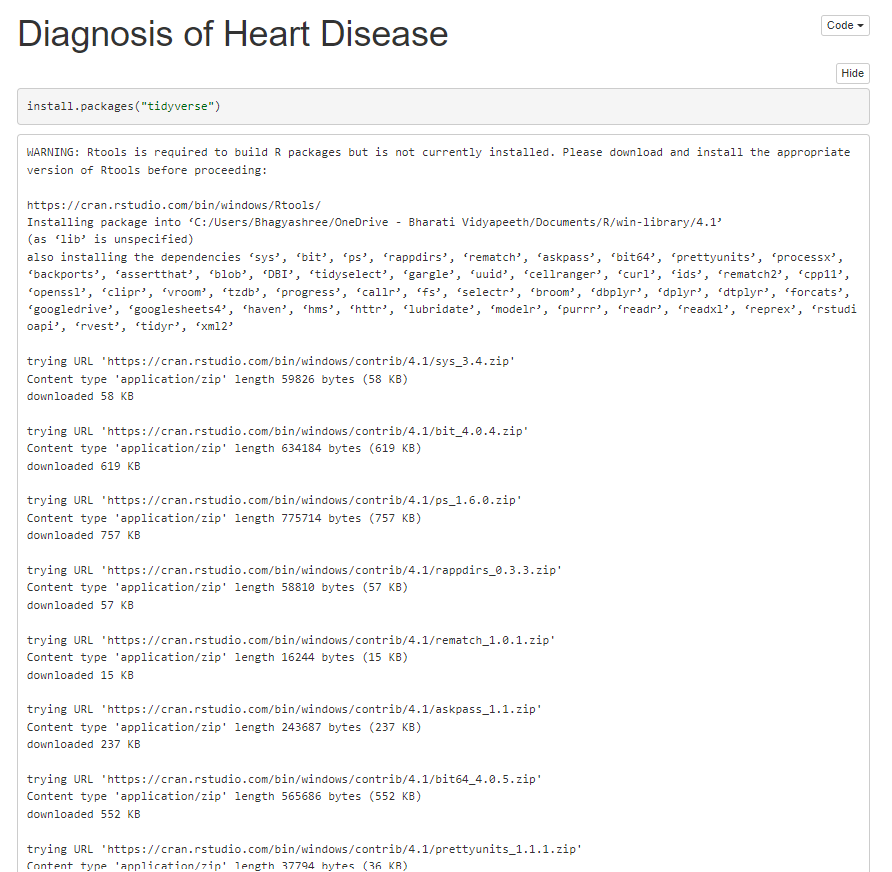
https://miro.medium.com/max/1310/1*7kictdVpel9Kp0pFj2UOPg.png

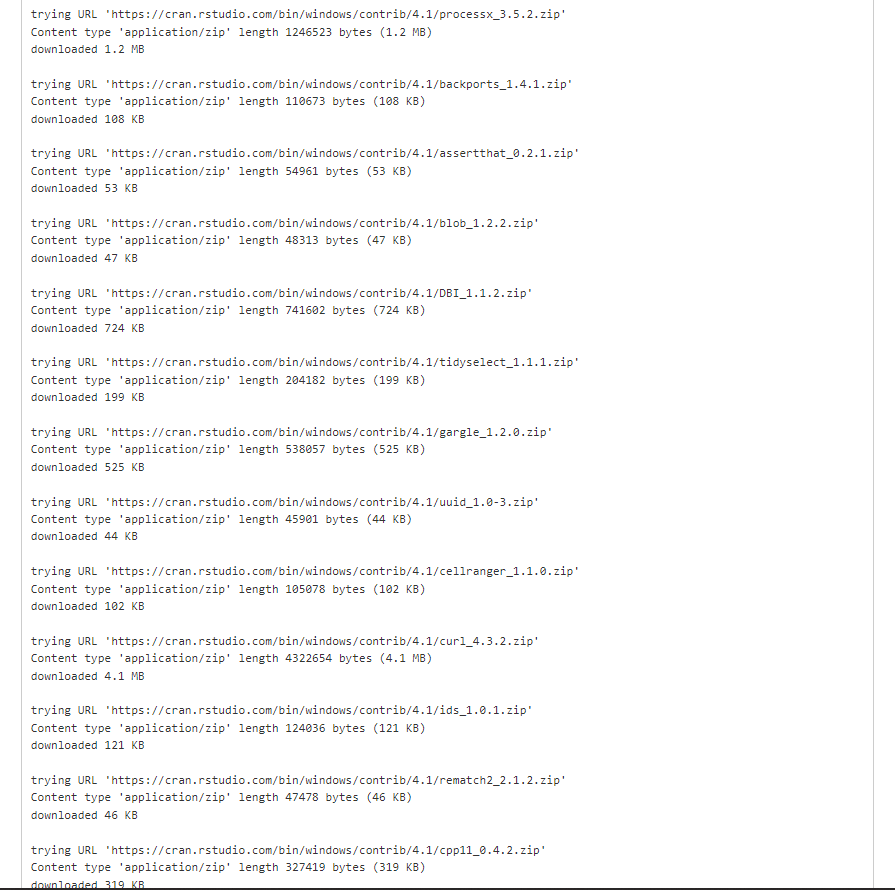
and

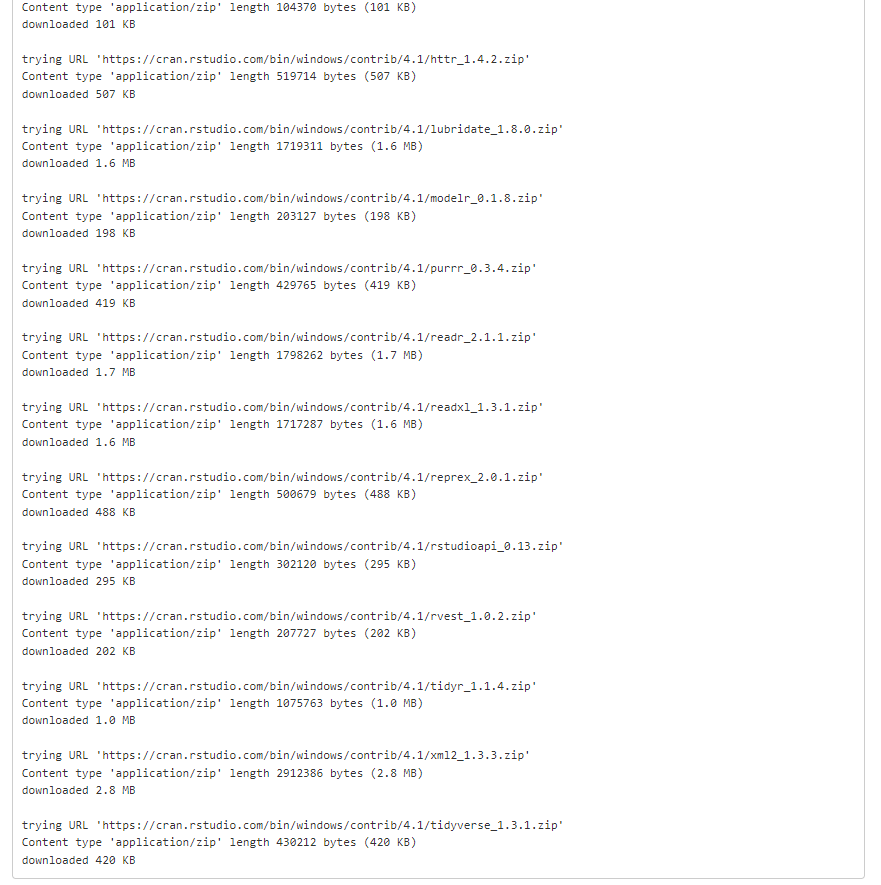


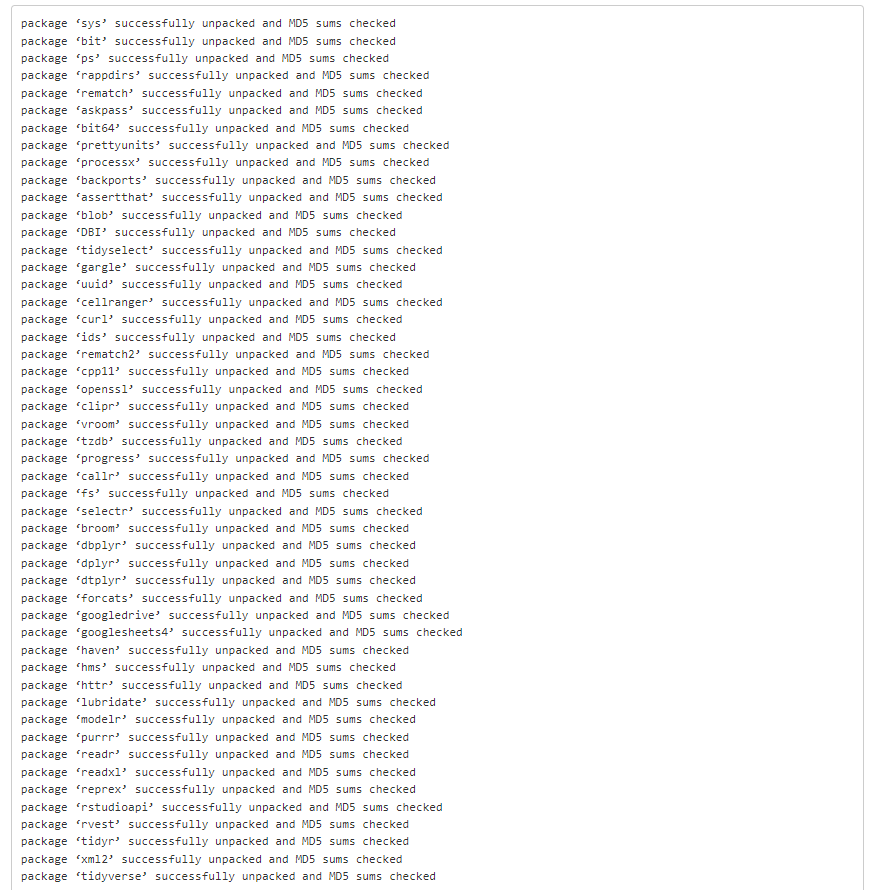
Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values.

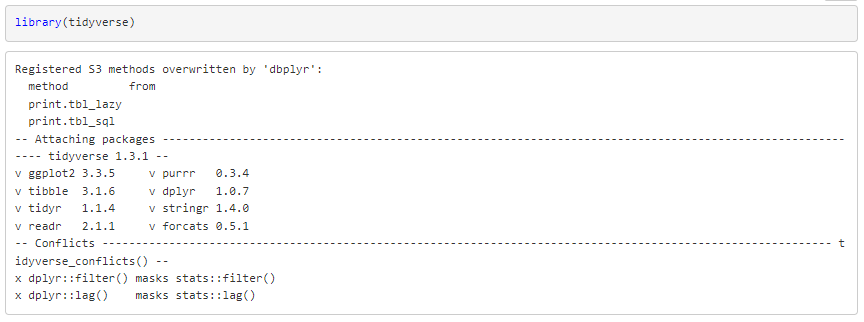
**3. Implementation:**

3.1 Diagnosis of Heart disease ****

****

****

****

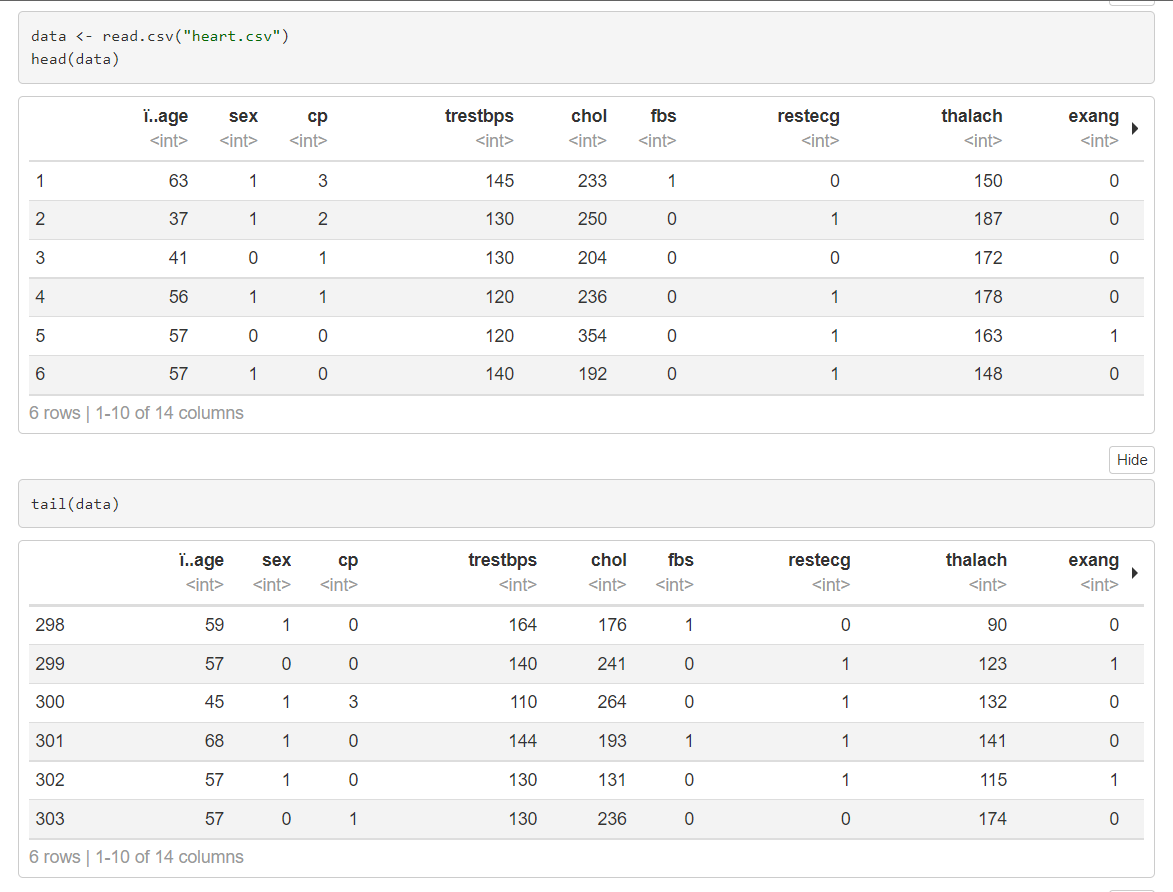
****

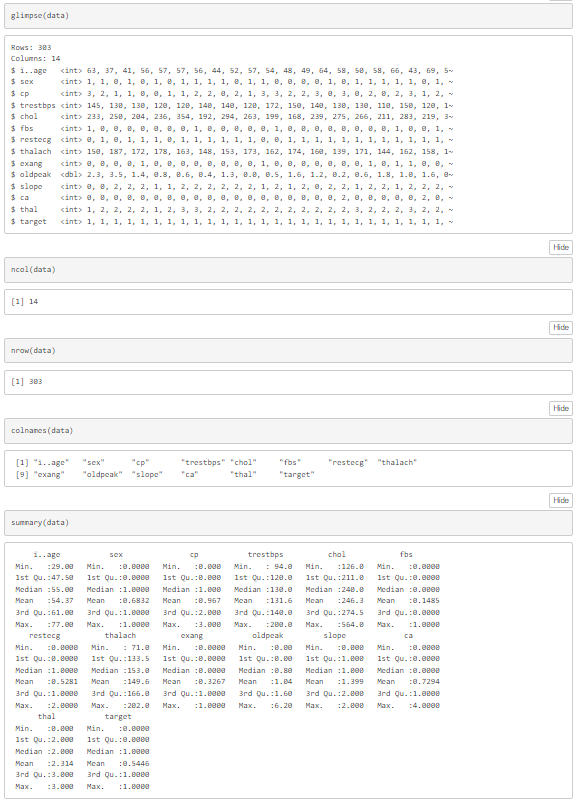
3.2 Data:

The dataset contains many medical indicators, This database contains 76 attributes. The dataset contains medical history of patients of Hungarian and Switzerland origin.

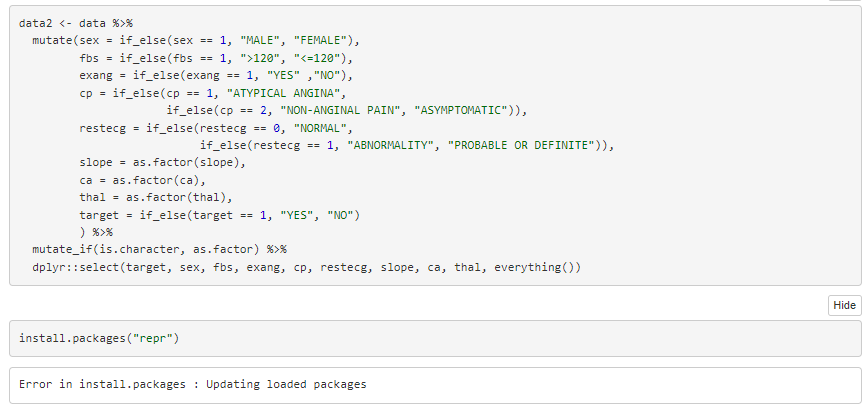
Attribute Information:

1. **age**: age in year
2. **sex:** (1 = male; 0 = female
3. **cp**: the chest pain experienced (value 1: typical angina, value 2: atypical angina, value 3: non-anginal pain, value 4: asymptomatic)
4. **trestbps**: resting blood pressure (in mm hg on admission to the hospital)
5. **chol**: serum cholesterol in mg/dl
6. **fbs**: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. [**restecg**](https://medium.com/@halima23121998/heart-disease-uci-logistic-regression-in-r-b95b821088e6?sk=def7a489c8ce9b249048e903d80f8591): resting electrocardiographic measurement (0 = normal, 1 = having st-t wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by estes’ criteria)
8. **thalach**: maximum heart rate achieved
9. **exang**: exercise-induced angina (1 = yes; 0 = no)
10. **oldpeak**: the slope of the peak exercise st segment (value 1: upsloping, value 2: flat, value 3: downsloping)
11. **slope**: the slope of the peak exercise st segment (value 1: upsloping, value 2: flat, value 3: downsloping)
12. **ca**: number of major vessels (0–3) colored by fluoroscopy
13. **thal**: a blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
14. **target**: heart disease (0 = no, 1 = yes)

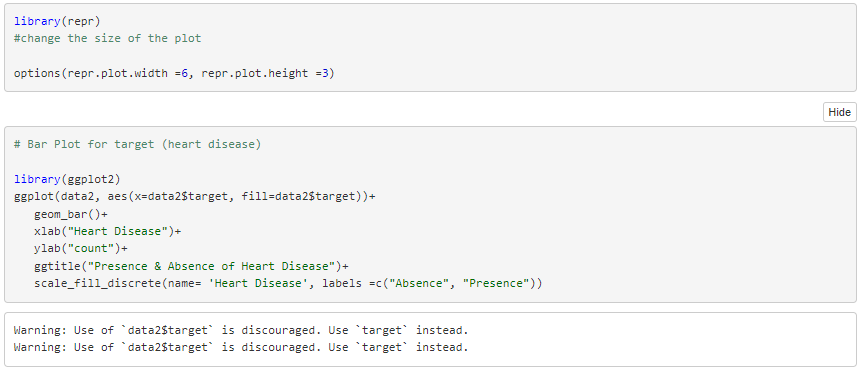


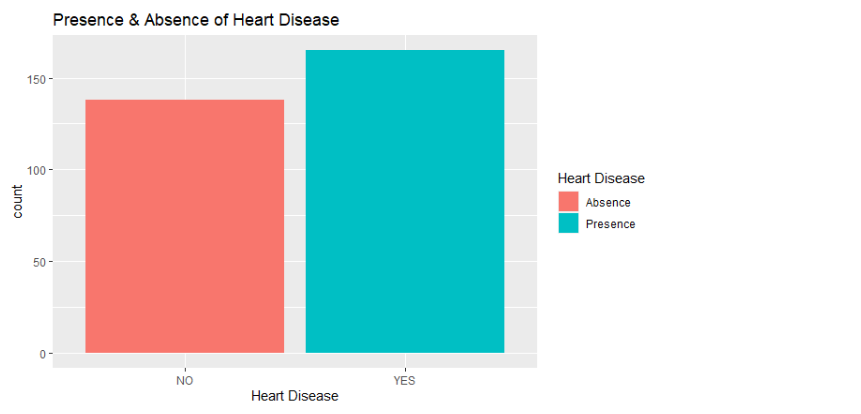
****

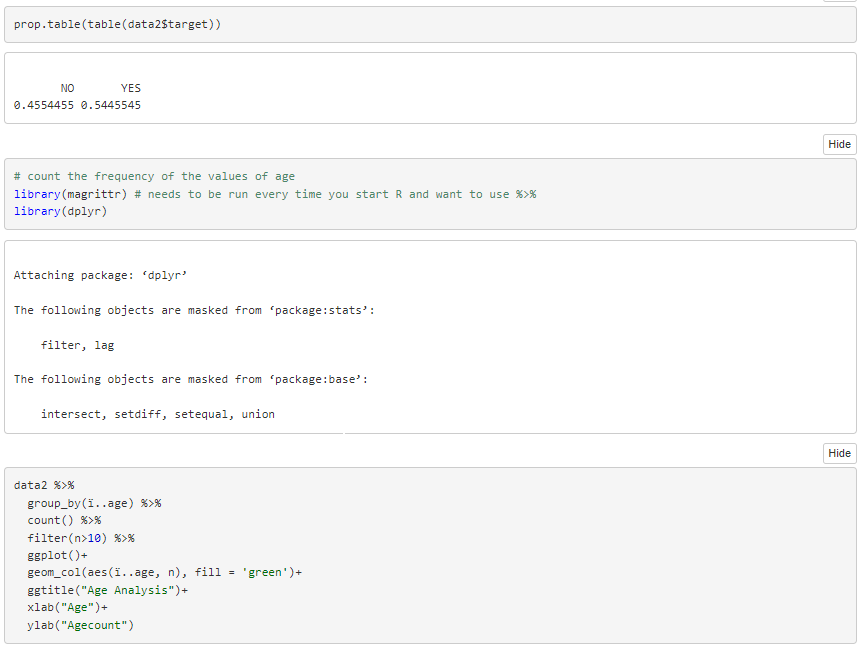
3.3 Data Transformation

****

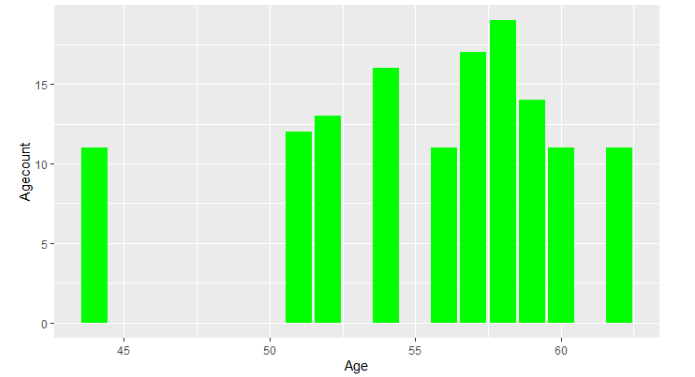
3.4 Data Visualization

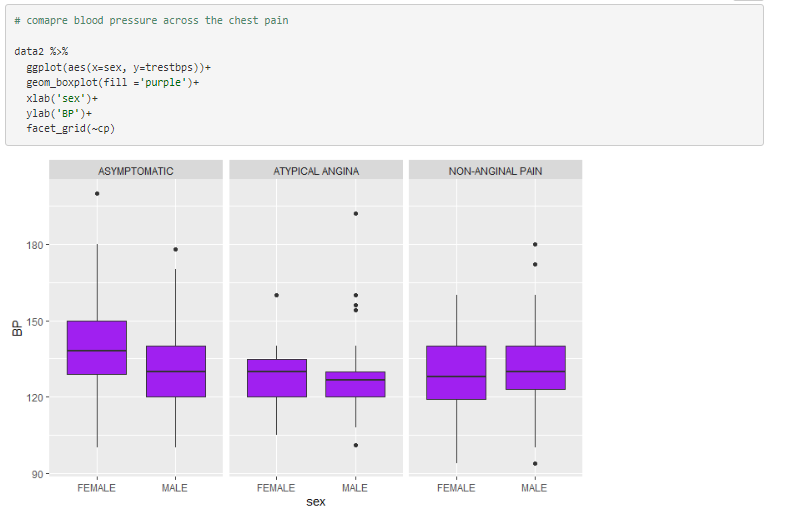
****

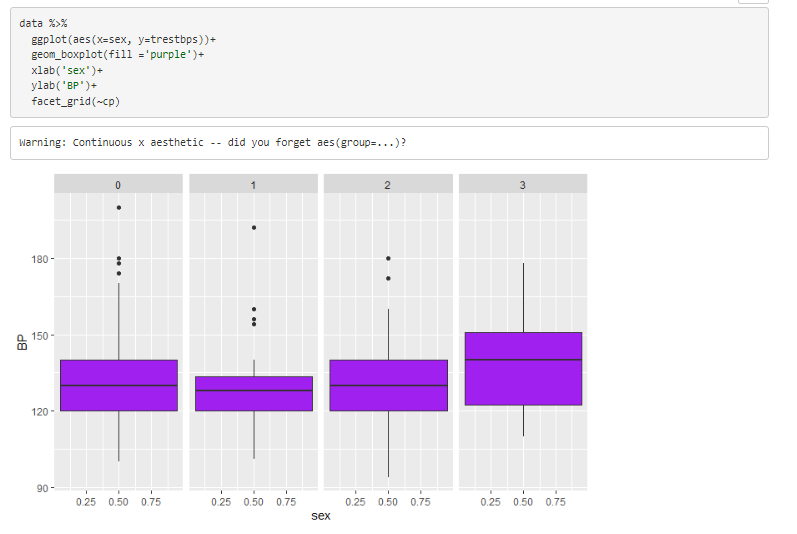
3.5 Presence and Absence of Heart Disease****

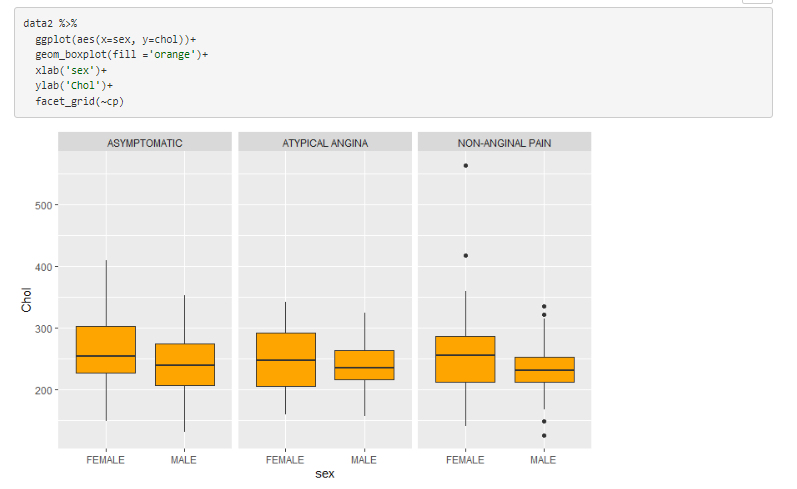
****

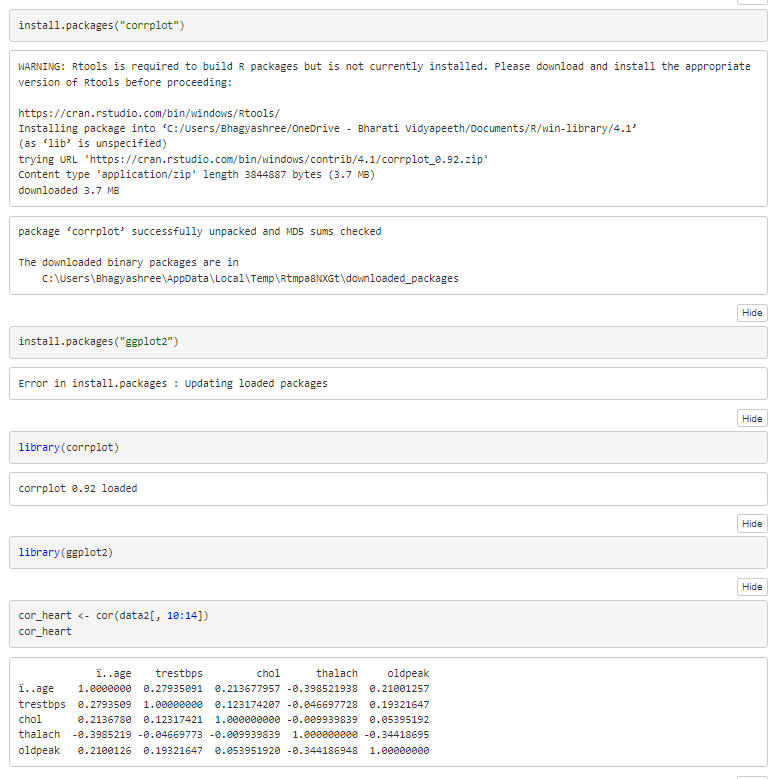
3.6 Age Analysis

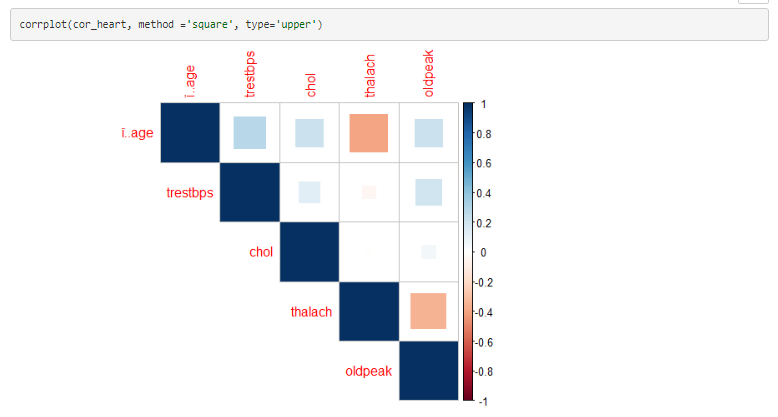
****

****

****

****

****

****

****

****

****

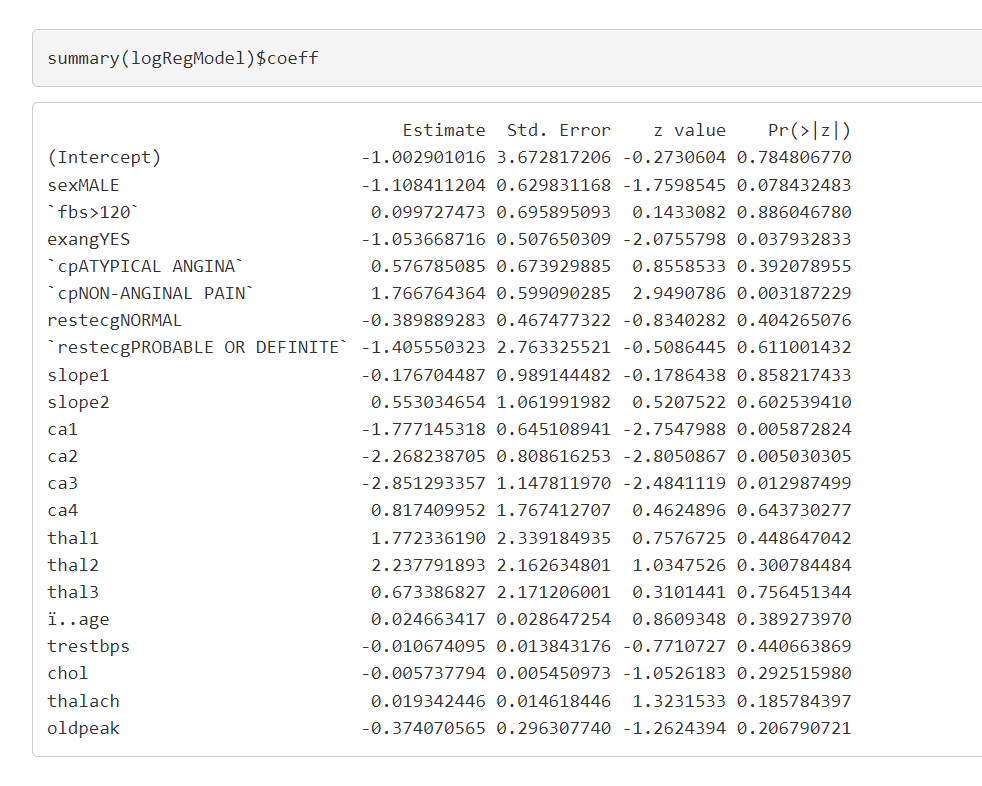
****

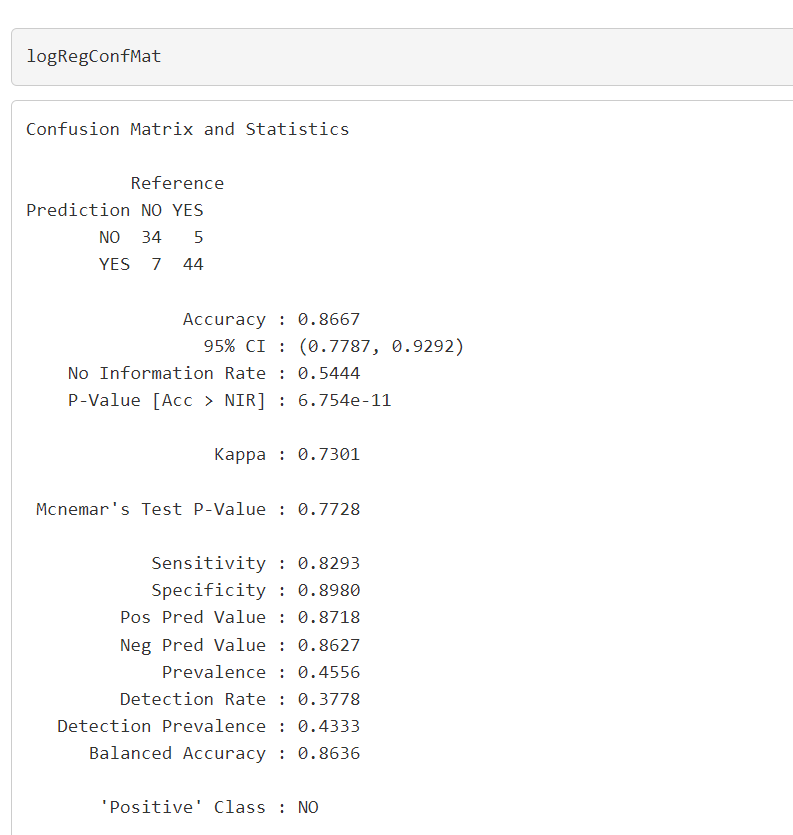
****

1. **Result:**

By using Logistic regression for heart disease detection we got accuracy about 87%.



****

****

1. **Conclusion:**

The number of heart diseases can exceed the control line and reach the maximum point. Heart disease is complicated and each and every year lots of people are dying with this disease. By using all systems one of the major drawbacks of these works mainly focuses only on the application of classifying techniques and algorithms for heart disease prediction, all these studying various data cleaning and mining techniques that prepare and build a dataset appropriate for data mining.

So that we can use this Machine Learning in that logistic regression algorithm by predicting if the patient has heart disease or not. Any nonmedical employee can use this software and predict heart disease and reduce the time complexity of the doctors.

1. **Future Work:**

In today’s, world most of the data is computerized, the data is distributed and it is not utilized properly. By analyzing the available data, we can also use it for unknown patterns. The primary motive of this research is the prediction of heart diseases with a high rate of accuracy. For predicting heart disease, we can use a logistic regression algorithm in machine learning. The future scope of the project is the prediction of heart diseases by using advanced techniques and algorithms in less time complexity.