TITLE: HEART FAILURE PREDICTION MODEL
INSIGHTS INTO HEART DIAGNOSIS

FINAL SEMESTER PROJECT

Name: ***BISHAL PAUL***
Registration No.: ***012-1111-0460-20***
Roll No.: ***203012-21-0007***
College Roll No.: ***89***
B.Sc. Honours in Statistics
Batch: ***2020-2023***
Semester: ***VI***

CREATING A PREDICTION MODEL BY STATISTICAL METHODS

CONTENTS

|ACKNOWLEDGEMENT

I, Bishal Paul, would like to thank my respected Supervisor Prof. Oindrila Bose for her continuous, patient & valuable guidance and support for execution of the project titled "HEART FAILURE PREDICTION MODEL INSIGHTS INTO HEART DIAGNOSIS ". I am grateful to her for her valuable suggestions and inputs enabling me to successfully complete my project with ease. Without her assistance the project could not be finished at all. In addition, I would also like to thank our Head of the Statistics Department, Asutosh College Dr. Dhiman Dutta and other faculty members Dr.Shirsendu Mukherjee, Dr. Parthasarathi Bera & Dr. Sankha Bhattacharya for their constant support. Without their valuable knowledge and assistance it was not possible to successfully complete the project

Bishal Paul

STUDENT

DEPARTMENT OF STATISTICS ,ASUTOSH COLLEGE

|**ABSTRACT**: Cardiovascular diseases (CVDs) are the number 1 cause of death globally,taking an estimated 17.9 million lives each year,which accounts for 31% of all deaths worldwide. Four out of 5 CVD deaths are due to heart attacks and strokes,and one-third of these deaths occur prematurely in people under 70 years of age.

Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension,diabetes,hyperlipidaemiaor already established disease)need early detection and management where in a machine learning model can be of great help.

<u>The main task is to develop a model to predict Heart failure from these variouscharacteristics</u>

In this data analysis project,Iwill:

➢ Find distribution of heart pain amongst various sex.

➢ At what resting blood pressure level do we start to see heart diseases.

➢ Find out cholesterol levels and heart failure levels for each age group.

➢ Creating a model for heart failure prediction.

➢ Accuracy of our model.

|**INTRODUCTION:**Cardiovascular diseases(CVDs)are a group of disorders of the heart and blood vessels. They include:

➢ Coronary heart disease–a disease of the blood vessels supplying the heart muscle;

➢ cerebrovasculardisease–adiseaseofthebloodvesselssupplyingthebrain;

➢ peripheralarterialdisease–adiseaseofbloodvesselssupplyingthearmsandlegs;

➢ rheumaticheartdisease–damagetotheheartmuscleandheartvalvesfromrheumaticfever, causedbystreptococcalbacteria;

➢ congenital heart disease – birth defects that affect the normal developmentand functioning of the heart caused by malformations of the heart structurefrombirth;and

➢ deepveinthrombosisandpulmonaryembolism–bloodclotsinthelegveins,whichcandislodgeandmovetotheheartandlungs.

Heart attacks and strokes are usually acute events and are mainly caused by ablockage that prevents blood from flowing to the heart or brain. The mostcommon reason for this is a build-up of fatty deposits on the inner walls of theblood vessels that supply the heart or brain. Strokes can be caused by bleedingfromabloodvesselinthebrain orfrombloodclots.

|**METHODOLOGY:**

- *Theoretical reference*: Thelinkofthedatasetis givenhere [CLICKHERE].
- *Software used*:RStudioversion1 .4.110

• ***OBJECTIVE:*** Many people around us face problems due to heart diseasesor other heart problems which in turn lead to bad heart health and also arisesconditionswhichour heartfailstocopeupwithandthuscollapsescausingheartfailures.

Mymainreasonbehindchoosingthistopicistoseehowbeautifullydatasciencewill let us draw more insights into data to develop metrices that will determinethisevent.

The other reason was to see how these different attributes chosen affects thehealth of our heart and which are the main attributes that causes our heart tofail.

• *Data Handling:* This dataset was created by combining different datasetsalready available independently but not combined before. In this dataset, 5heart datasets are combined over 11 common features which makes it thelargest heart disease dataset available so far for research purposes. The fivedatasets usedforitscurationare:

Cleveland: 303 observations.**Hungarian**: 294 observations.**Switzerland**: 123observations.**Long Beach VA**: 200 observations.**Stalog**(Heart) **Data Set**: 270observations.**Total**:1190observations.**Duplicated**: 272observations.

Finaldataset:918observations.

|GAININGINDUSTRY/BUISNESSKNOWLEDGE:

➢ **Typical angina (TA)**is defined as substernal chest pain precipitated byphysical exertionoremotionalstressandrelievedwithrestornitroglycerin.Findmoreinformation[HERE].

➢ **Atypicalangina(ATA)**chestpainapplieswhen2out of3criteriaofclassicangina arepresent.

➢ **Non-Angina**PainisalsoknownasNon-Cardiac ChestPain(NCCP).NCCPisatermusedtodescribechestpainthatresemblesheartpain(alsocalledangina)in patients who do not have heart disease. The pain typically is felt behind thebreast bone (sternum) and is described as oppressive, squeezing or pressure-like,symptomsareclassifiedasnon-specific.

➢ **Asymptomatic**isalsoknownassilentmyocardialischemia.Theyaresilentbecausewhentheyoccur,theirsymptomslacktheintensityofaclassicheartattack, such as extreme chest pain and pressure; stabbing pain in the arm,neck,orjaw;suddenshortnessofbreath; sweating,anddizziness.

➢ **FastingBloodSugar**isthebloodsugarafteranovernight fast.Zerobeingrecordedforthesemeanthepatientmustbedead.

➢ **Resting BP**That istheblooppressurewhenthebodyisinarestingposition.This is the normal blood pressure and when zero is recorded then the patientmustbedead.

|Codes:

• INSTALLINGNEEDEDPACKAGESFORANALYSIS:

install.packages('tidyverse')#fordatamanipulation

install.packages('dummies')#forcreatingdummyvariables

install.packages('ggpubr')#wouldbeusedtoarrangeplotsonasinglepagewith ggarange

install.packages('caTools')#Forrandomlysplittingdata

intotestandtraininstall.packages('rpart') #for creating a decision tree

modelinstall.packages('rpart.plot') #forplottingthedecisiontree

library(tidyverse)

library(dummies)

library(ggpubr)li

brary(caTools)lib

rary(rpart)library

(rpart.plot

• GETTINGOURDATAANDHAVINGAGLIMPSEOF IT:

> hdf<-read.csv('C:/Users/USER/Desktop/shouryarghyapatra/heart.csv')

> glimpse(hdf)#havingaglimpseofthedata

Rows:918

Columns:12

$Age <int>40,49,37,48,54,39,45,54,37,48,37,58,39,49,42,54,38,43,60,36,43,44,49, 44,40,36,53,52,53,51,53,56,54,41,43,32,65, 41,48,48,54,~

$Sex <chr> "M","F","M","F","M","M","F","M","M","F","F","M","M","M","F","F","M", "F","M","M","F","M","F","M","M","M","M","M","F","M","M","M","M"~

$ChestPainType<chr>"ATA","NAP","ATA","ASY","NAP","NAP","ATA","ATA","ASY","ATA","NAP", "ATA","ATA","ASY","NAP","ATA","ASY","ATA","ASY","ATA","TA","ATA","ATA","AT~

$RestingBP <int>140,160,130,138,150,120,130,110,140,120,130, 136,120,140, 115,120, 110,120,100,120,100,120,124,150,130,130,124,120,113,125,145,130,125~

$Cholesterol <int>289,180,283,214,195,339,237,208,207,284,211,164,204,234,211,273, 196,201,248,267,223,184,201,288,215,209,260,284,468,188,518,167,224~

$FastingBS <int>0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0~

$RestingECG

<chr>"Normal","Normal","ST","Normal","Normal","Normal","Normal","Normal","N ormal","Normal","Normal","ST","Normal","Normal","ST","Normal","Normal","Norm~

$MaxHR <int>172,156,98,108,122,170,170,142,130,120,142,99,145,140,137,150,166, 165,125,160,142,142,164,150,138,178,112,118,127,145,130,114,122,~

$ExerciseAngina<chr>"N","N","N","Y","N","N","N","N","Y","N","N","Y","N","Y","N","N","N", "N","N","N","N","N","N","Y","N","N","Y","N","N","N","N","N","N"~

$Oldpeak <dbl>0.0,1.0,0.0,1.5,0.0,0.0,0.0,0.0,1.5,0.0,0.0,2.0,0.0,1.0,0.0,1.5,0.0,0.0,1.0, 3.0,0.0,1.0,0.0,3.0,0.0,0.0,3.0,0.0,0.0,0.0,0.0,0.0, 2.0~

$ST_Slope <chr>"Up","Flat","Up","Flat","Up","Up","Up","Up","Flat","Up","Up","Flat","Up", "Flat","Up","Flat","Flat","Up","Flat","Flat","Up","Flat","Up","Flat~

$HeartDisease<int>0,1,0,1,0,0,0,0,1,0,0,1,0,1,0,0,1,0,1,1,0,0,0,1,0,0,0,0,0,0,1,0,1, 1,0,0,1,0,0,0,0,1,0,0,1,0,0,0,0,1,1,1,0,0,0~

> paste0('There are ', nrow(hdf),' observations in the dataset. This tallies withthe metadataofthedataset')#checkingthenumberof observations

> summary(hdf)#Let'sgetdescriptivestatisticsonthevariousvariables

"Thereare918observationsinthedataset.Thistallieswiththemetadataofthedataset"

Age Sex ChestPainType Cholesterol FastingBS RestingECG MaxHR RestingBPExerciseAngina Oldpeak

Min.:28.00Length:918Min.:60.0Length:918 Length:918 Min.:-2.6000 Min.:0.0 Min. :0.0 Min. :0.0000 Length:918

1stQu.:47.00Class:characterClass:character1stQu.:120.01stQu.:173.21stQu.:0.0000Class :character1stQu.:120.0Class:character1stQu.:0.0000

Median:54.00Mode:characterMode:characterMedian:130.0Median:223.0Median:0.0000Mode:characterMedian:138.0Mode:characterMedian:0.6000

Mean:53.51 Mean:132.4Mean:198.8Mean:0.2331 Mean :136.8 Mean:0.8874

3rd Qu.:60.00Qu. 3rdQu.:1.5000 3rdQu.:140.03rdQu.:267.03rdQu.:0.0000 3rd :156.0

Max.:77.00 Max. :200.0 Max. :603.0 Max. :1.0000 Max. :202.0 Max.:6.2000

ST_Slope

HeartDiseaseLengt

h:918

Min.:0.0000Class:

character1st

Qu.:0.0000Mode:characterMedia

n:1.0000

Mean:0.55343

rdQu.:1.0000

Max.:1.0000

ObservationsfromSummary:

➢ RestingBPandCholesterolhaszeroasminimumwhichisunusual.

➢ Theremaybeoutliers/MassingsinCholesterolandRestingBPbeingpresenteda szero.

➢ TheinterquartilerangeofMaxHRsuggeststhattheremaybeoutliersfromMinto 1stquartile.

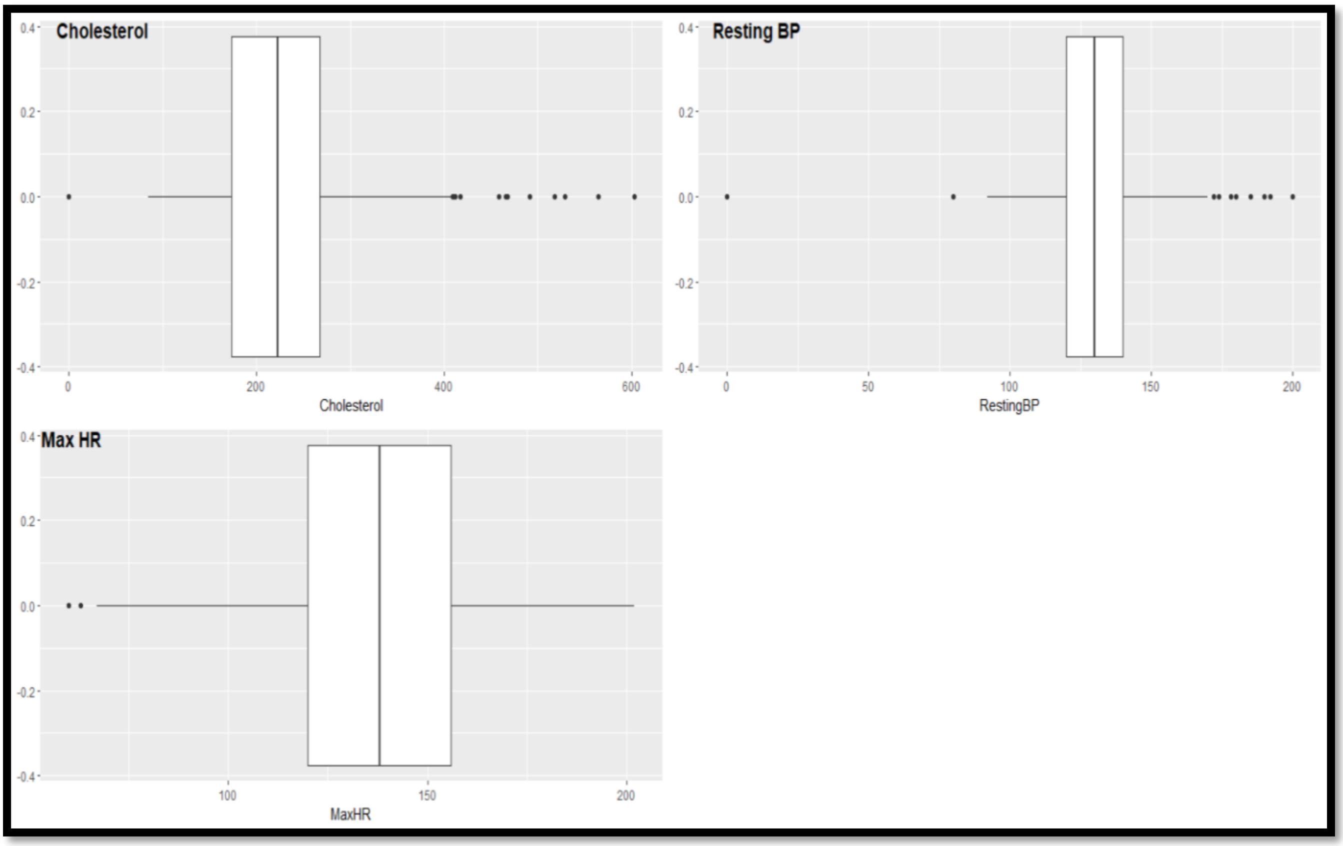
• **CHECKINGFOROUTLIERS:**

> chlout<-ggplot(data=hdf,aes(x=Cholesterol))+geom_boxplot()

> rbpout<-ggplot(data=hdf,aes(x=RestingBP))+geom_boxplot()

> maxhr<-ggplot(data=hdf,aes(x=MaxHR))+geom_boxplot()

> ggarrange(chlout,rbpout,maxhr,labels=c("Cholesterol","RestingBP","MaxHR")) #displaysalltheplotin onepage.



• DATACLEANING:

WewilltreatzerosinimpossiblefieldsasNA's.This includeszerosrecordedincholesterol and resting bp. The 0s will be replaced by the mean of theobservations.

➢ *FixingZeros/NAsinCholesterol:*

Ourbusinessknowledgetellsusthatitisimpossibletohavezerocholesterol.Henceitis highlypossibleitwasintroducedbyerror.

```
> hdf$Cholesterol[which(hdf$Cholesterol==0)]<-mean(hdf$Cholesterol)
```

• **CONCLUSION:** *Itisimpossibletohavezerocholesterol,sowecansayprobablyit was introduces due to some error. So to eradicate this error or to reduce wereplaceallthezero cholesterolwiththemeancholesterol.*

➢ *FixingZeros/NAsinRestingBP:*

Our business knowledge of the dataset suggests that zero resting bp is for thedead. Hence it might have been introduced into the data set by mistake. We'llfixthatbyinsertingthemeanofthatvariables.

```
> hdf$RestingBP[which(hdf$RestingBP==0)]<-mean(hdf$RestingBP)
```

```
> summary(hdf
```

Age Sex ChestPainType RestingBP Cholesterol FastingBS
RestingECGMaxHR ExerciseAngina Oldpeak

Min.:28.00Length:918 Length:918 Min.:80.0Min.:85.0Min.:0.0000Length:918
Min.:60.0Length:918 Min.:-2.6000

1stQu.:47.00Class:characterClass:character1stQu.:120.01stQu.:198.81stQu.:0.0000Class:chara
cter1stQu.:120.0Class:character1stQu.: 0.0000

Median:54.00Mode:characterMode:characterMedian:130.0Median:223.0Median
:0.0000Mode:characterMedian:138.0Mode:characterMedian:0.6000

Mean:53.51 Mean:132.5Mean:236.0Mean:0.2331 Mean
:136.8 Mean:0.8874

3rd 3rdQu.:140.03rdQu.:267.03rdQu.:0.0000 3rd
Qu.:60.00Qu. 3rdQu.:1.5000
:156.0

Max.:77.00 Max.:200.0Max.:603.0Max.:1.0000 Max.
:202.0 Max.:6.2000

ST_Slope

HeartDiseaseLeng

th:918 Min.

:0.0000Class:character

1stQu.:0.0000Mo

de:characterMedian:1.0000

Mean:0.55343

rdQu.:1.0000

Max.:1.0000

• **CONCLUSION:***Our knowledge leads us to the fact that Resting Bp zero is onlyfor deceased or dead people, so it is impossible for a living person to have zeroResting Bp. Thus, we come to conclusion that it was introduces into the dataset by mistake. So here we replace the zero Resting Bp with the mean RestingBp.*

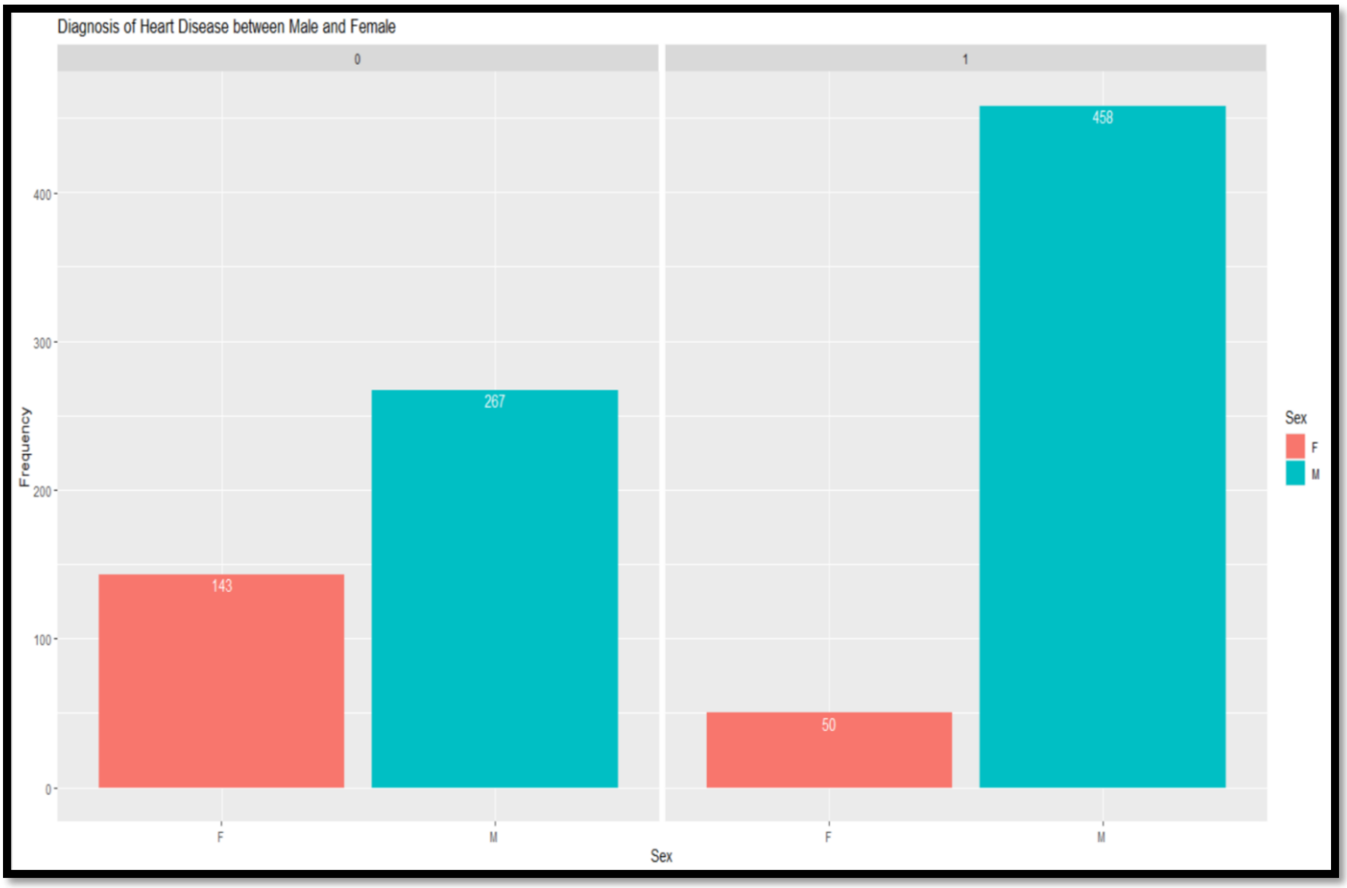
From the summary we also see that now the mean and median of theCholesterol isnowcloseasalltheoutliersorerrors havebeenremoved.

• VISUALIZATIONSANDDATAEXPLORATIONS:

Thissectioncontainsvisualizationsfromthecleaneddataset.

➢ *heartdiagnosisdistributionamongvarioussexes:*

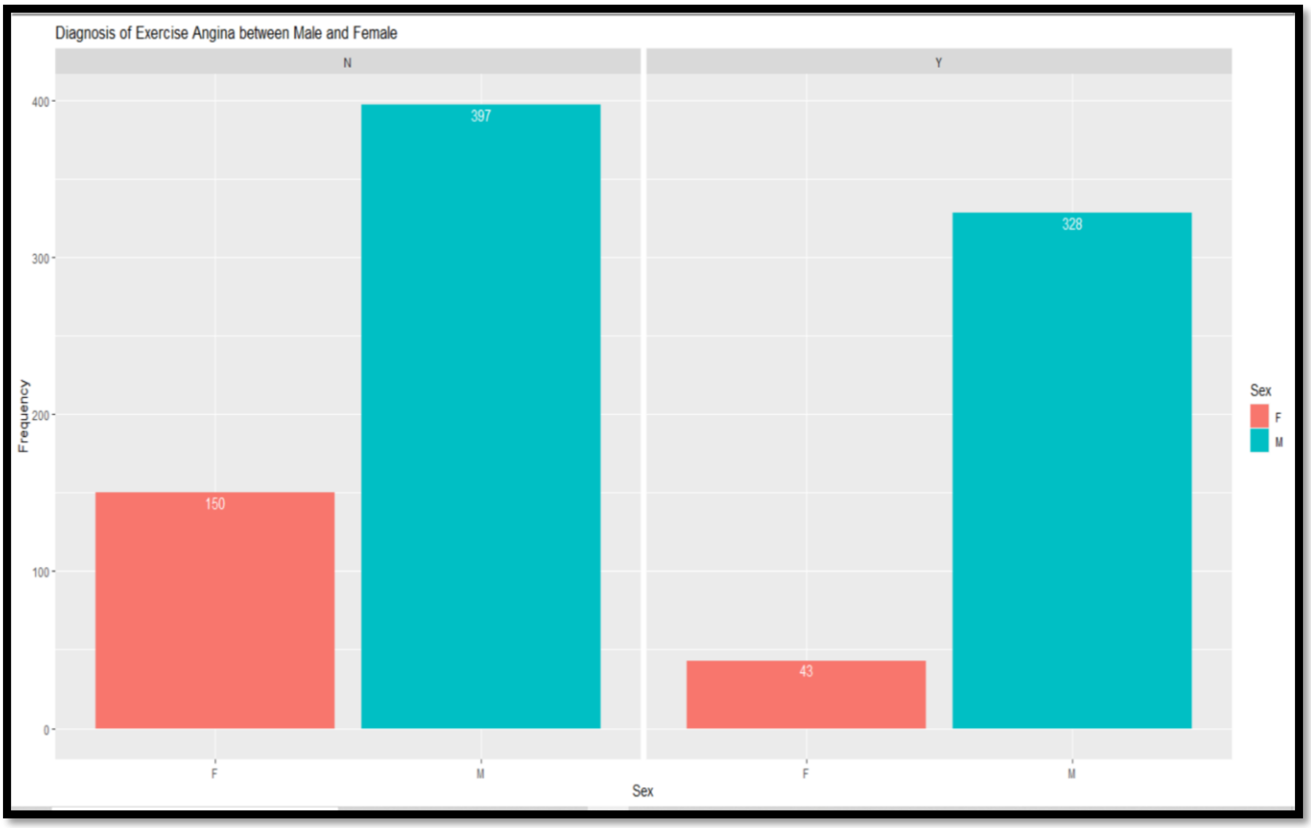
```
> ggplot(data = hdf, aes(x=Sex, fill =Sex))+geom_bar()+facet_grid(~HeartDisease)+geom_text(aes(label = ..count..),stat='count',vjust=1.2,color='white')+labs(title='DiagnosisofHeartDiseasebetweenMaleandFemale',y='Frequency')
```



• **CONCLUSION:***Here 0 signifies the total population from the data set with noheart disease and 1 signifies population suffering from heart disease. From 0column we find 267 male and 143 female with no heart disease. Total femalepopulation:193andmalepopulation:725.Therefore,percentageoffemalesuffering from heart disease is 26%(approx.) and male suffering from heartdisease is 63%(approx.). So we find that male population is high likely to sufferfromheartdiseaseascomparedtofemalepopulation.*

➢ ***ExerciseAnginadistributionamongvarioussexes:***

ggplot(data = hdf, aes(x=Sex, fill =Sex))+geom_bar()+facet_grid(~ExerciseAngina)+geom_text(aes(label=..count..),stat='count' , vjust = 1.2,color = 'white')+labs(title = 'Diagnosis of Exercise Angina between MaleandFemale',y= 'Frequency')

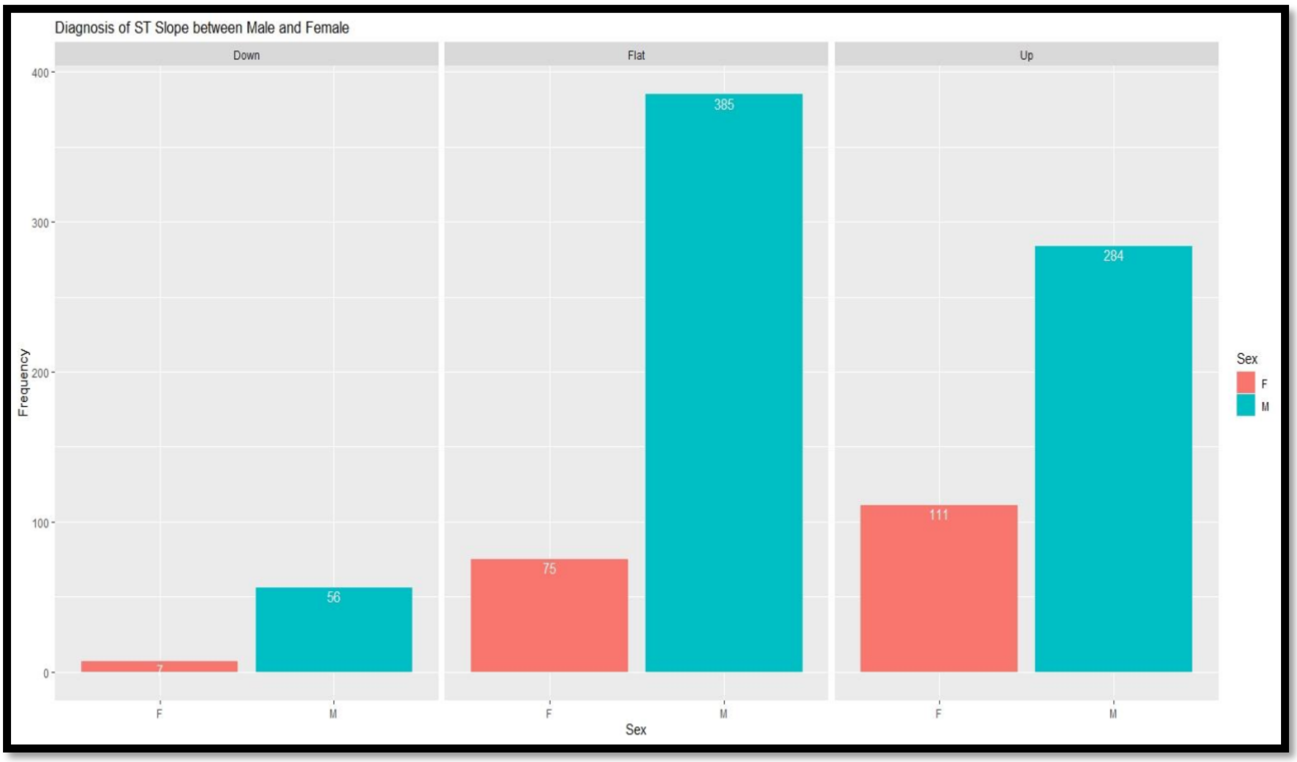


• **CONCLUSION:** *Angina refers to pain in the chest due to some physical activityor stress. Here N denotes a person is not suffering from Angina whereas Ydenotespersonfromthedatasetsufferingfrom Angina.Totalfemalepopulation:193 and male population:725. Females not suffering from Angina is150 and male not suffering is 397 whereas females suffering is 43 and malesufferingis328.Therefore,percentageoffemalessufferingfromAnginais22%(approx.) and percentage of males suffering is 45%(approx.). So we come toaconclusionthatmalepopulationismorelikelytosufferfromExerciseAngina.*

• **REMEDY:** *If chest pain last longer it is advisable to consult a doctor and take medications.Exerciselightlyoranyaerobicexerciseshouldbeimplementedindailylife.Warmingup isalsonecessaryasitallowsbloodvesselstowidentoreduceanginasymptoms.Thiswillalsoreducechancesof Heart attack.*

➢ ***STSlopedistributionamongvarioussexes:***

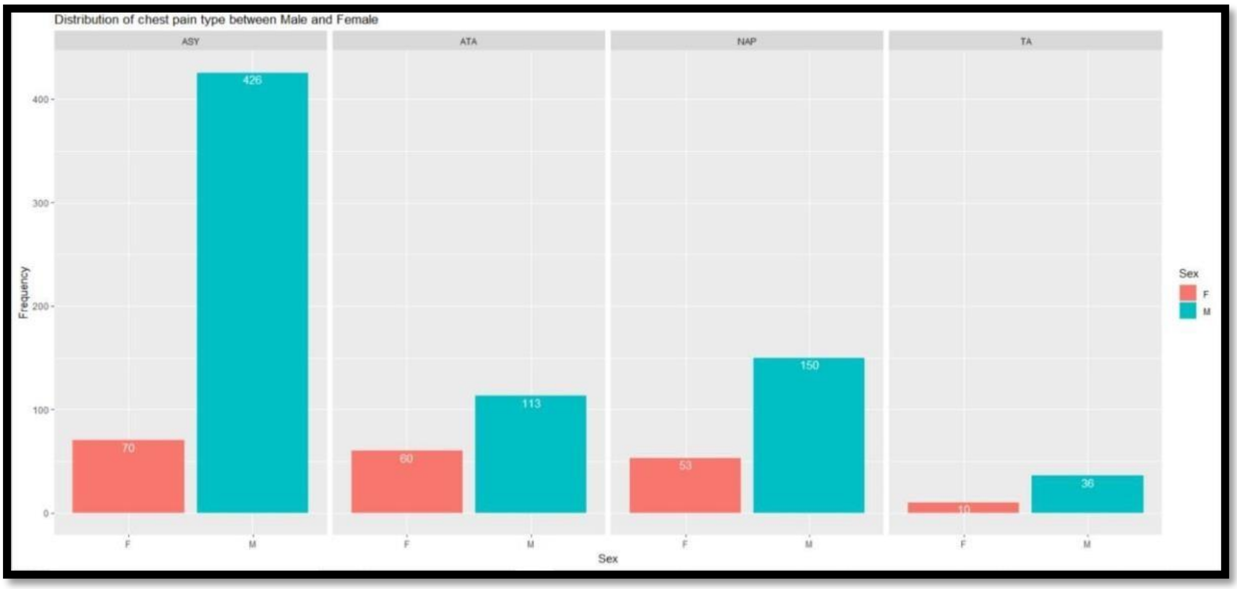
ggplot(data = hdf, aes(x=Sex, fill =Sex))+geom_bar()+facet_grid(~ST_Slope)+geom_text(aes(label=..count..),stat='count',vjust = 1.2,color = 'white')+labs(title = 'Diagnosis of ST Slope between Male and Female',y='Frequency')



- **CONCLUSION**:*TheSTsegmentisaportionoftheECGcycle.Here<u>Flat</u>representsthat theparticularpeoplehavenormalSTslopeorcanalsobetermedasisoelectric.Here<u>Down</u>refersmanifestationofmyocardialischemia.<u>Up</u>referstoanincreasechanceofCAD(coronaryarterydisease).Totalfemalepopulation:193andmalepopulation:725.Total numberoffemaleshavingdownSTSlopeis7andthatofmalesis56.Percentage offemaleslikelyto getaffectedwithmyocardialischemiais3%andthatofmalesis7%.TotalfemaleshavingflatSTSlopeis75andthatofmalesis385.PercentagehavingaNormalSTSlopewithnoriskis39%(approx.)andthatofmalesis53%(approx.).Totalfemaleswith UpST Slope is 111 and that of males is 284. Percentage of females likely to sufferfrom CAD is 57%(approx.) and that of males is 39%(approx.). So we come to a<u>conclusion</u>thattheentirepopulationhave<u>lesschancesofsufferingfromischemia.</u>Itisalsoseenthat<u>mentendtohaveNormalSTslopethanwomen andfemalesaremorelikelytosufferfromCAD.</u>*

➢ ***<u>Heartpaindistributionamongstvarioussex</u>***

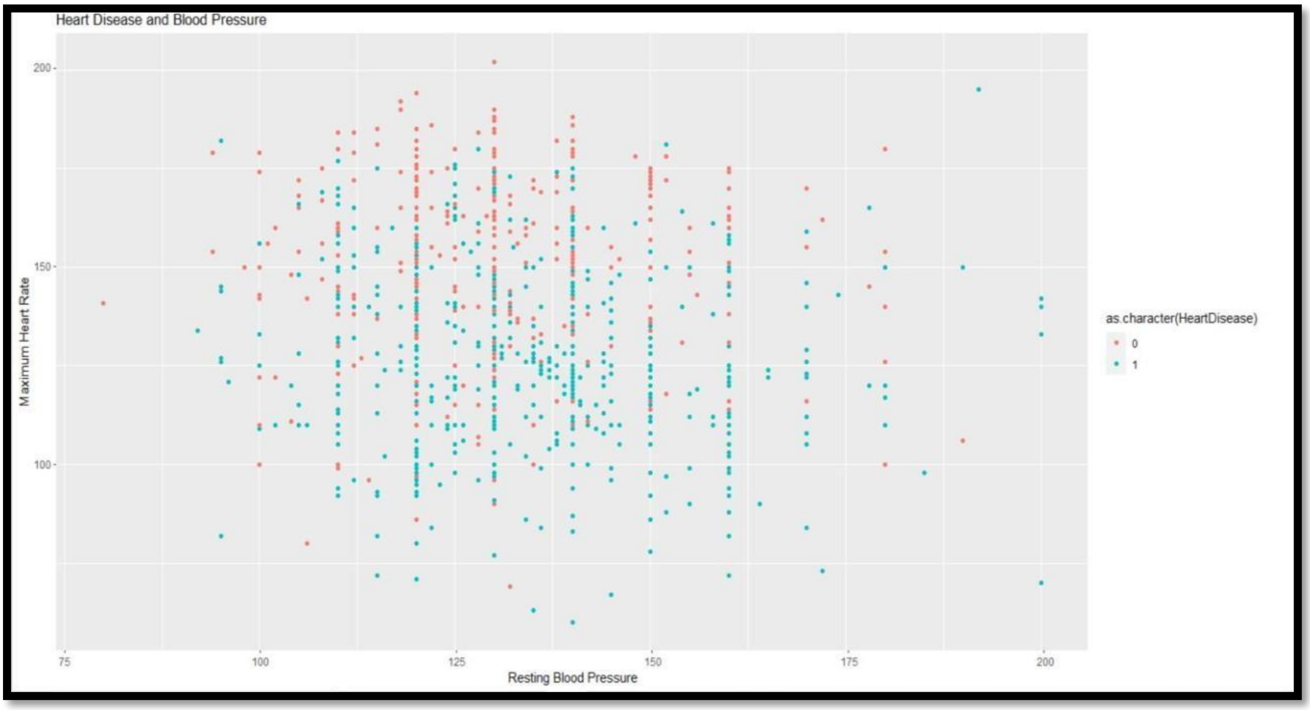
> ggplot(data=hdf,aes(x=Sex,fill=Sex))+geom_bar()+facet_grid(~ChestPainType) +geom_text(aes(label = ..count..), stat = 'count',vjust=1.2,colour='white')+labs(title='DistributionofchestpaintypebetweennMaleandFemale',y='Frequency')



• **<u>CONCLUSION</u>**: *Total female population:193 and male population:725. "ASY"refers to Asymptomatic (producing or showing no symptoms). Females showingnosymptomsare70andthepercentageis36%(approx.)andthatofmalesis426and the percentage is 59%(approx.). "ATA" refers to Atypical Angina a conditionwhen chest pain is observed but doesn't meet conditions of angina. Femalessuffering from ATA is 60 and the percentage is 31% and males suffering is 113and the percentage is 16%(approx.). "NAP" refers to Non Anginal Pain is acondition where chest pain is observed despite of none heart disease. FemaleswithNAPis53andthepercentageis27%andthatofmalesis150andpercentage is 21%. "TA" refers to Typical Angina a condition where chest painobservedduetophysicalexertionoremotionalstress,femalessufferingis10and percentage is 5% and males is 36 and the percentage is 5%(approx.). Thus, weconcludefromourreadingsisthatmenaremore likelytobe<u>asymptomatic</u>thanthat of women that is <u>chest pain will occur but show no symptoms or diseasesleadingtothatcondition</u>whichismoredangerous.Womenareevenmorelikelyto suffer from <u>ATA and even NAP</u>. Only in case of TA both men and women are<u>likelylesstosufferbut</u>both<u>get</u><u>diagnosedwiththesameprobability</u>.*

➢ ***Comparisonbetweenrestingbloodpressureandheartdiseases.***

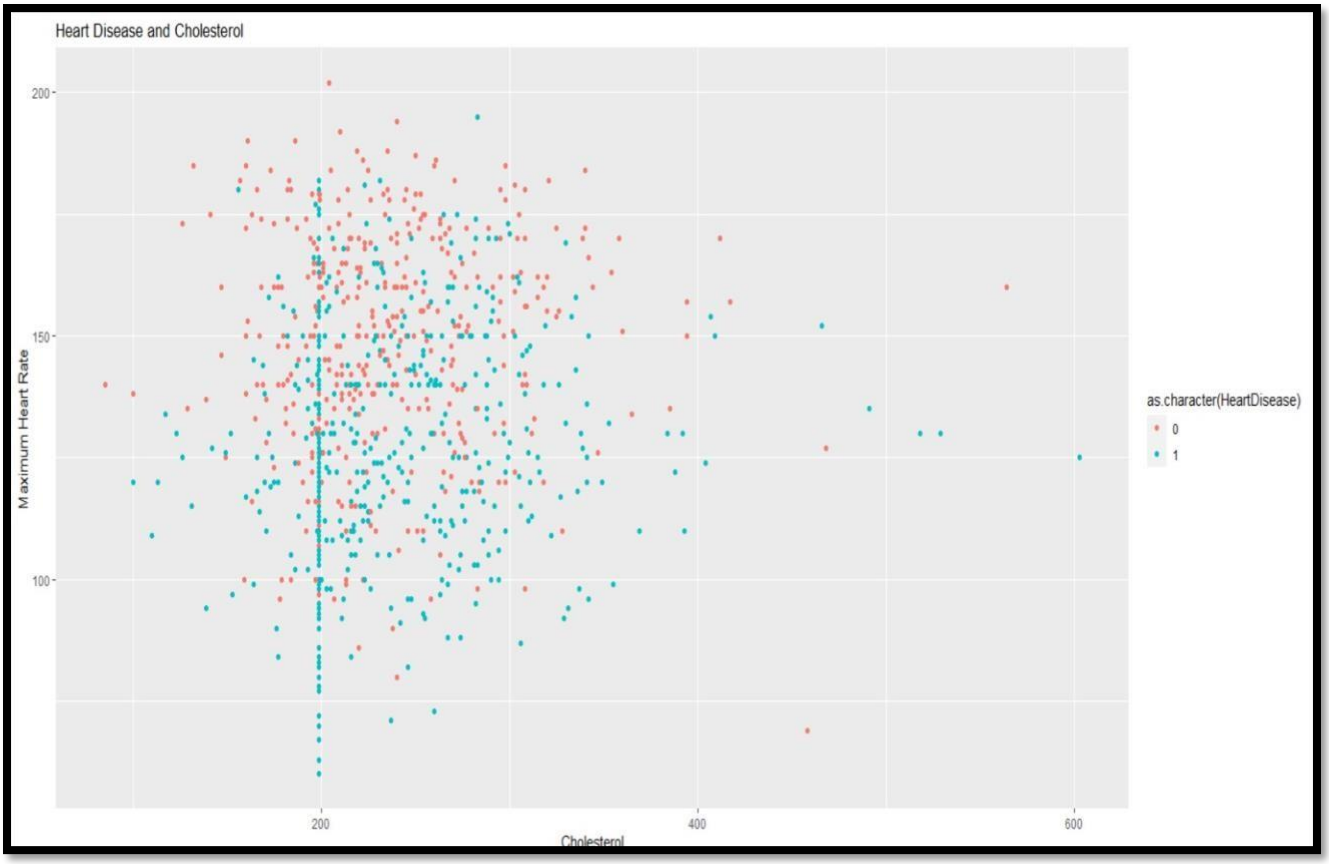
```
ggplot(data=hdf,aes(x=RestingBP, y=MaxHR, colour=as.character(HeartDisease)))+geom_point()+labs(title='Heart Disease and BloodPressure', x='RestingBloodPressure', y='MaximumHeartRate')
```



• **CONCLUSION**:*Here intheabove graph,weplotRestingBpintheX axisand Maximum Heart Rate in the Y axis and from the data set we find the bothtype of persons one suffering from heart disease (denoted by 1) and personswith no heart disease (denoted by 0). Then we plot people suffering from heartdisease(bluedot)bytakingintoaccount theirMaximumHeartRateandResting Bp and also the same for people with no heart disease (red dots). Thus,from the graph we get the idea that if the Max HR is less as compared to theincrease in Resting Bp a person is more likely to suffer from heart disease but aperson is less likely to have heart disease if his Max HR is within the range150-200 regardless of the Resting Bp. Thus, a person is advised to exercise toincrease heart rate thus decreasing chances of heart disease. We can alsoconclude that the Resting Bp of persons lying in the range of 100 to 145 is alsolesslikelytogetdiagnosedbyanysortofheartdisease.Althoughthebestrange is Max HR between 150 – 170 and Resting Bp between 105 - 135 (fromthegraph).*

ComparisonbetweenCholesterolandHeartdiseases.

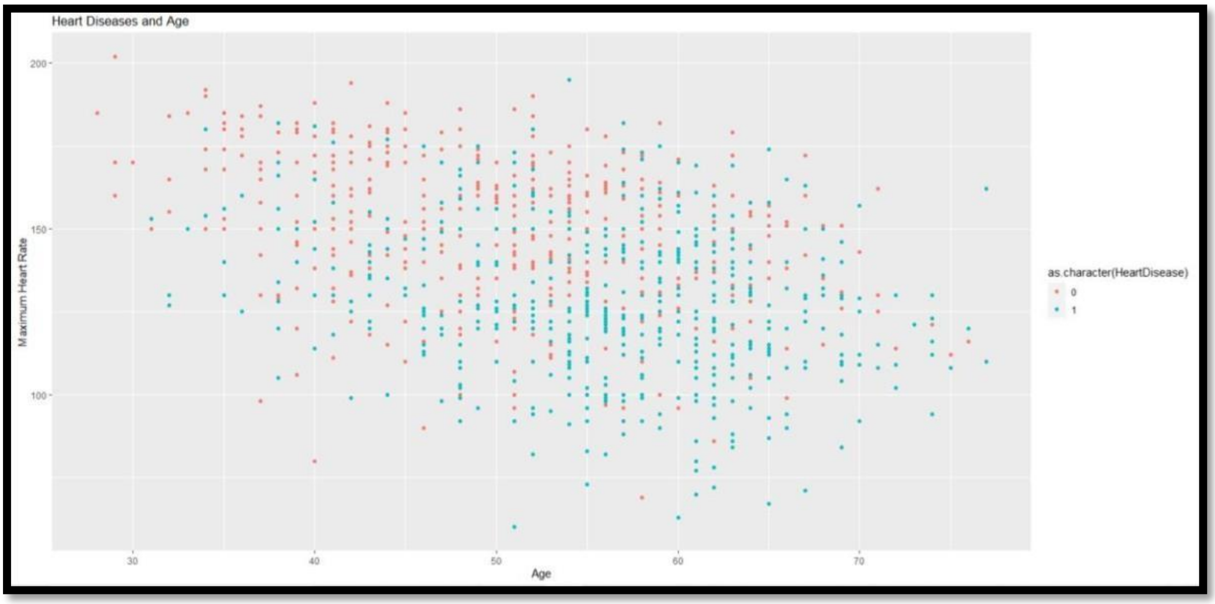
ggplot(data=hdf,aes(x=Cholesterol, y=MaxHR, colour=as.character(HeartDisease)))+ geom_point()+labs(title='HeartDiseaseandCholesterol',x='Cholesterol',y='MaximumHeartRate')



• **CONCLUSION:** *Here in the above graph, we plot Cholesterol in the X axisand Maximum Heart Rate in the Y axis and from the data set we find the bothtype of persons one suffering from heart disease (denoted by 1) and personswith no heart disease (denoted by 0). Then we plot people suffering from heartdisease (blue dot) by taking into account their Maximum Heart Rate andCholesterol also the same for people with no heart disease (red dots). Thepeople in the data set having cholesterol levels below 200 and Max HR between140 to 180 are less likelyto get affected byanyheartdisease. Eventhosewith cholesterol levelsbetween 200 to 300 but with Max HR between 150 to170 are also less probable to any heart disease. Any person with Max HR lessthan 100 are facing heart diseases regardless of any cholesterol levels. Any alsowith Max HR between 100to 150 but with cholesterol greater than 200can getconsumed by heart diseases. Best region or no heart diseaseare faced bypeoplewithMaxHRbetween150and 170andcholesterol less than 200.*

ComparisonbetweenAgeandHeartdiseases.

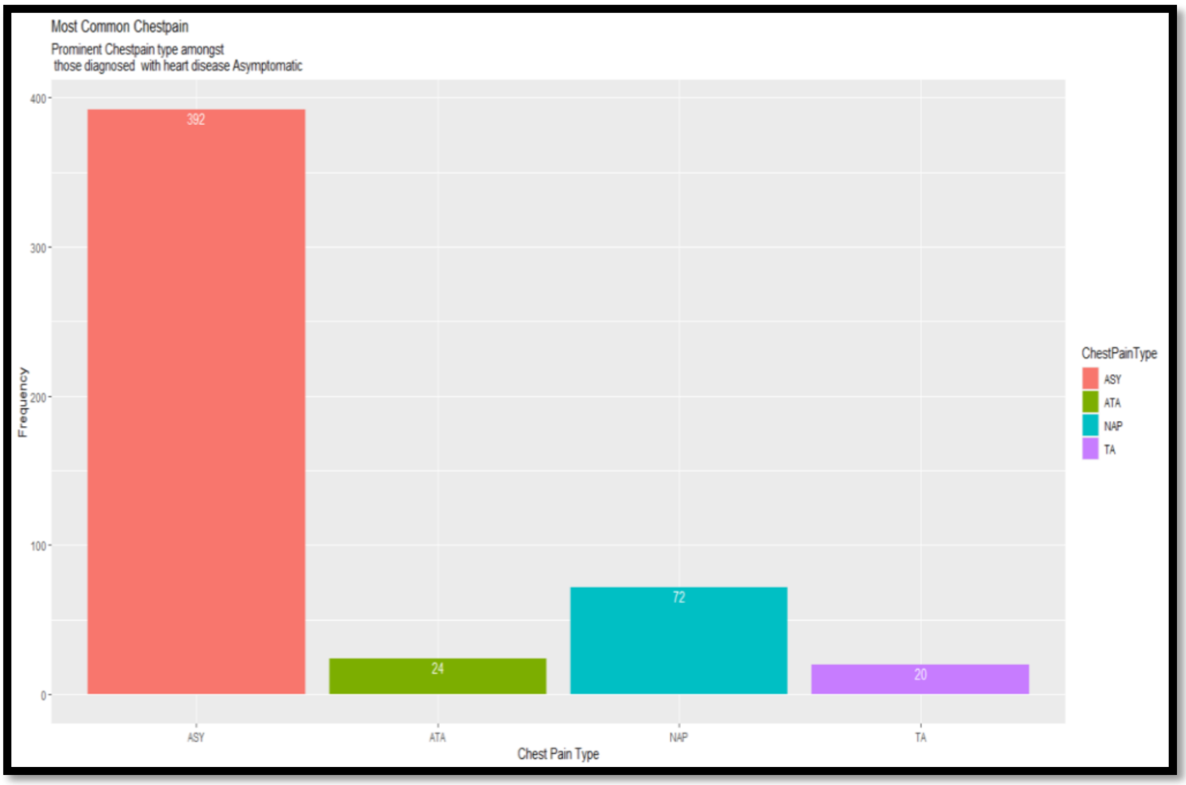
ggplot(data=hdf,aes(x=Age, y=MaxHR, colour=as.character(HeartDisease)))+geom_point()+labs(title='HeartDiseasesandAge',x='Age',y='MaximumHeart Rate')



• ***CONCLUSION***:*Here we plot Age in the X-axis which constitutes of thedifferent ages of the people in the study and the Y-axis contains the MaximumHeart Rate and after plotting the data we see that the peoplewith ages less than40 are at low chances or at low risk of suffering from any heart disease but as wecontinue our observation from 40+ age we see a considerable increase in theheartdiseaseanditincreaseorreachesitpeakaftertheageof 60.*

➢ ***WhatisthemostcommonChestPainamongstthosediagnosedwith HeartDisease?***

```
ggplot(data=filter(hdf,HeartDisease==1),aes(x=ChestPainType,fill=ChestPainTyp
e))+ geom_bar()+ geom_text(aes(label=..count..),
stat='count',vjust=1.2,colour='white')+                labs(title=
'MostCommonChestpain',
    subtitle='ProminentChestpaintypeamongst\nthosediagnosed
                              withheart                      disease
                              Asymptomatic',
    y                               =
                              'Frequency',x='ChestPain Type')
```



- **CONCLUSION:** *There are four types of Chest Painpresent:i.TypicalAngina(TA) ii.AtypicalAngina(ATA) iii.Non-Angina (NAP)iv.Asymptomatic (ASY)*

In the X-axis we plot the Chest Pain Type types and on the Y-axis we plot thefrequency and then after plotting the data we come to see that most people aresuffering from asymptomatic i.e. they are silent when they occur but theirsymptoms lack intensity for a classic heart attack such as extreme chest pain, paininthearm,suddenshortnessofbreath, sweating, dizziness.

➢ ***Whatarethecholesterollevelsandheartfailurecasesforeachageg roup***

hdf2<-hdf

>hdf2$AgeGroup[hdf2$Age >20&hdf2$Age<35]<-'20-34'

>hdf2$AgeGroup[hdf2$Age >34&hdf2$Age<45]<-'35-44'

>hdf2$AgeGroup[hdf2$Age >44&hdf2$Age<55]<-'45-54'

>hdf2$AgeGroup[hdf2$Age >54&hdf2$Age<65]<-'55-64'

>hdf2$AgeGroup[hdf2$Age >64&hdf2$Age<78]<-'65-77'

> head(hdf2)

```
  Age Sex ChestPainType RestingBP Cholesterol FastingBS RestingECG MaxHR ExerciseAngina Oldpeak ST_Slope HeartDisease AgeGroup
1  40   M           ATA       140         289         0     Normal   172              N     0.0       Up            0  35 - 44
2  49   F           NAP       160         180         0     Normal   156              N     1.0     Flat            1  45 - 54
3  37   M           ATA       130         283         0         ST    98              N     0.0       Up            0  35 - 44
4  48   F           ASY       138         214         0     Normal   108              Y     1.5     Flat            1  45 - 54
5  54   M           NAP       150         195         0     Normal   122              N     0.0       Up            0  45 - 54
6  39   M           NAP       120         339         0     Normal   170              N     0.0       Up            0  35 - 44
>
```

AgeSexChestPainTypeRestingBPCholesterolFastingBSRestingECGMaxHRExerciseAnginaOldpeakST_SlopeHeartDiseaseAgeGroup1.40

M	***ATA***	***140***	***289***	***0***	***Normal***	***172***	***N***	***0.0***	***Up***	***0***	***35-44***
2.49F	***NAP***	***160***	***180***	***0***	***Normal***	***156***	***N***	***1.0***	***Flat***	***1***	***45–54***
3.37M	***ATA***	***130***	***283***	***0***	***ST***	***98***	***N***	***0.0***	***Up***	***0***	***35–44***
4.48F	***ASY***	***138***	***214***	***0***	***Normal***	***108***	***Y***	***1.5***	***Flat***	***1***	***45-54***
5.54M	***NAP***	***150***	***195***	***0***	***Normal***	***122***	***N***	***0.0***	***Up***	***0***	***45–54***
6.39M	***NAP***	***120***	***339***	***0***	***Normal***	***170***	***N***	***0.0***	***Up***	***0***	***35-44***

•**CONCLUSION**: *After calculating the heart failure and cholesterol levels wecometotheconclusionthatmaleswithintheage35-44aremorepronetoheart failure. Then females within the age 45-54 are more prone to heart attackthan males an again males within the age range 35-44 with cholesterol levelsaround 283 are more prone to heart attack. Females with cholesterol levelsabove 210 and having ASY chest pain are prone to heart attacks within the agerange 45-54 are prone to heart attack and also involving a heart disease. Maleswith age around 54 with no heart diseases are also prone to heart attackbecause there are other significant factors which detoriates the heart health.Males with age 39 are prone to heart failure if they have cholesterol levels ashighas339.*

➢ ***CreatingSummarytableandvisualizationsfordisplay.***

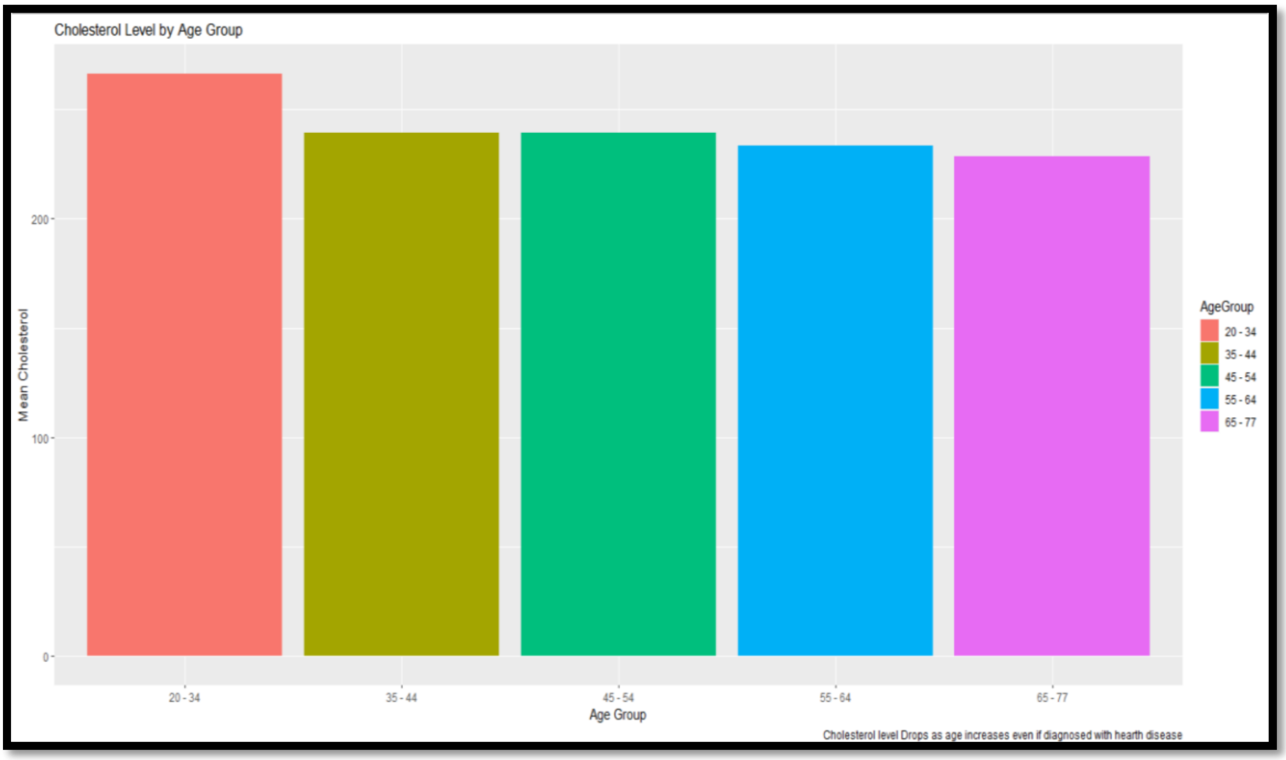
Overhere,let'shavealookatthesummarizeddata.

```
sum_hdf2<-hdf2%>%
  filter(HeartDisease==1)%>%#filtersforthosediagnosedwithheartdisease
  group_by(AgeGroup)%>%#groupsbyagegroup
  summarise(NumberofCases=n(),#getsmethenumberofparticipantsineach
group
  MeanCholesterol=mean(Cholesterol))#getsmethemeancholesterolhead(sum_hdf2)
```

```
#Atibble:Ag 5x3
   eGroup      NumberofCasesMeanCholesterol

             <chr>                    <int>              <dbl>1.20-
          34                            6                 266.4332
             2.35- 44                   52               240.5345
             3.45- 54                   139              239.4258
             4.55- 64                   239              234.5639
             5.65- 77                    72              228.3221
```

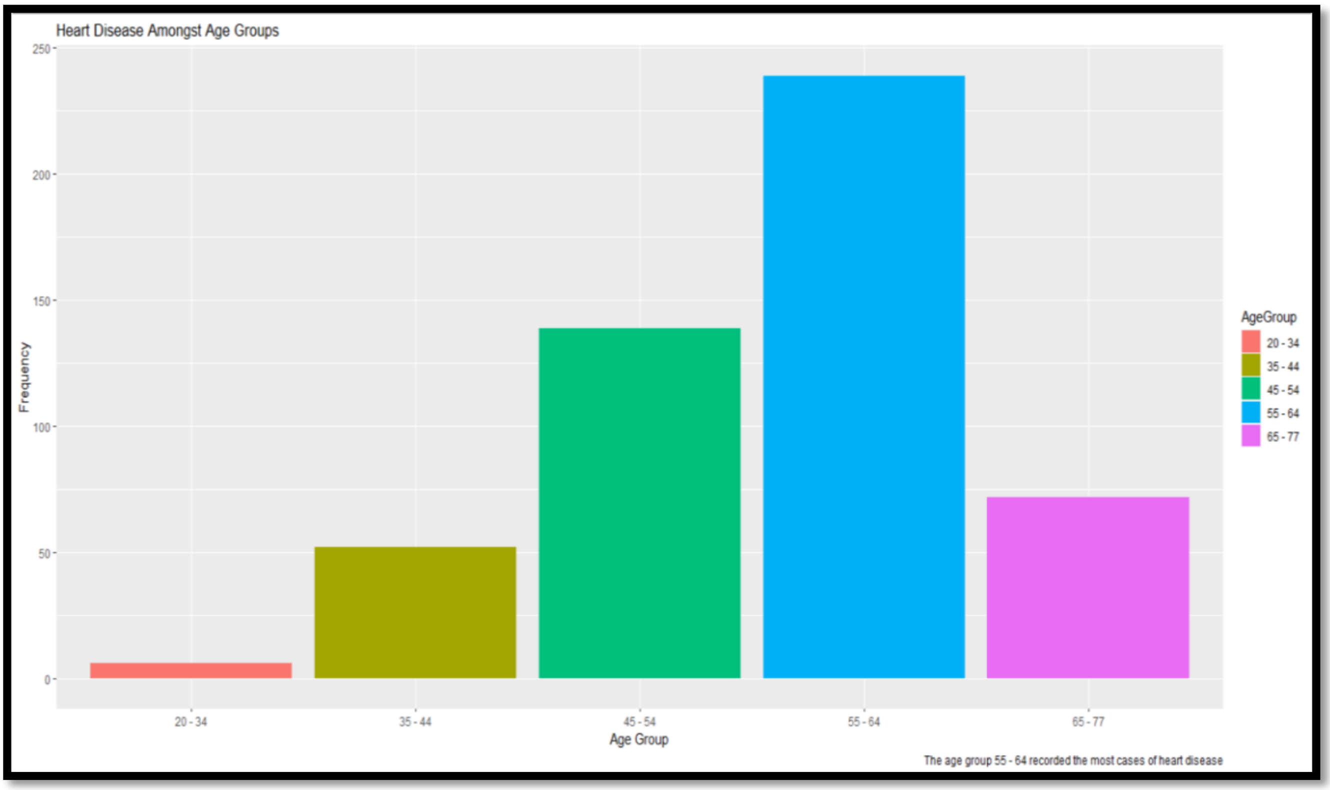
```
ggplot(data=sum_hdf2,aes(x=AgeGroup,y=MeanCholesterol,fill
 =AgeGroup))+
   geom_bar(stat='identity') +
   labs(title='CholesterolLevelbyAgeGroup',
        caption='CholesterollevelDropsasageincreasesevenifdiagnose
 dwith hearth disease',
        y='MeanCholesterol',x=
        'Age Group')
```



- <u>**CONCLUSION**</u>: *Here the X axis contains all the age groups from 20 to77brokenas20-34,35-44, 45-54,55-64, 65-77.*
  The Y axis contains all the Mean cholesterol and the data is represented here by abar diagram. We find that people with age rage 20-34 are most sufferers of HighCholesterol and at age 35-44 and 45-54 we find the people in the study has samelevel of cholesterol and it reduces just a little bit as we traverse through all the agegroups

DistributionofHeartDiseasesamongstsagegroups.

```
ggplot(data=sum_hdf2,aes(x=AgeGroup,y=NumberofCases,fill=AgeGroup))+
  geom_bar(stat='identity')+
  labs(title='HeartDiseaseAmongstAgeGroups',
       caption= 'Theagegroup55 -
64recordedthemostcasesofheartdisease',
y='Frequency',x='AgeGroup')
```

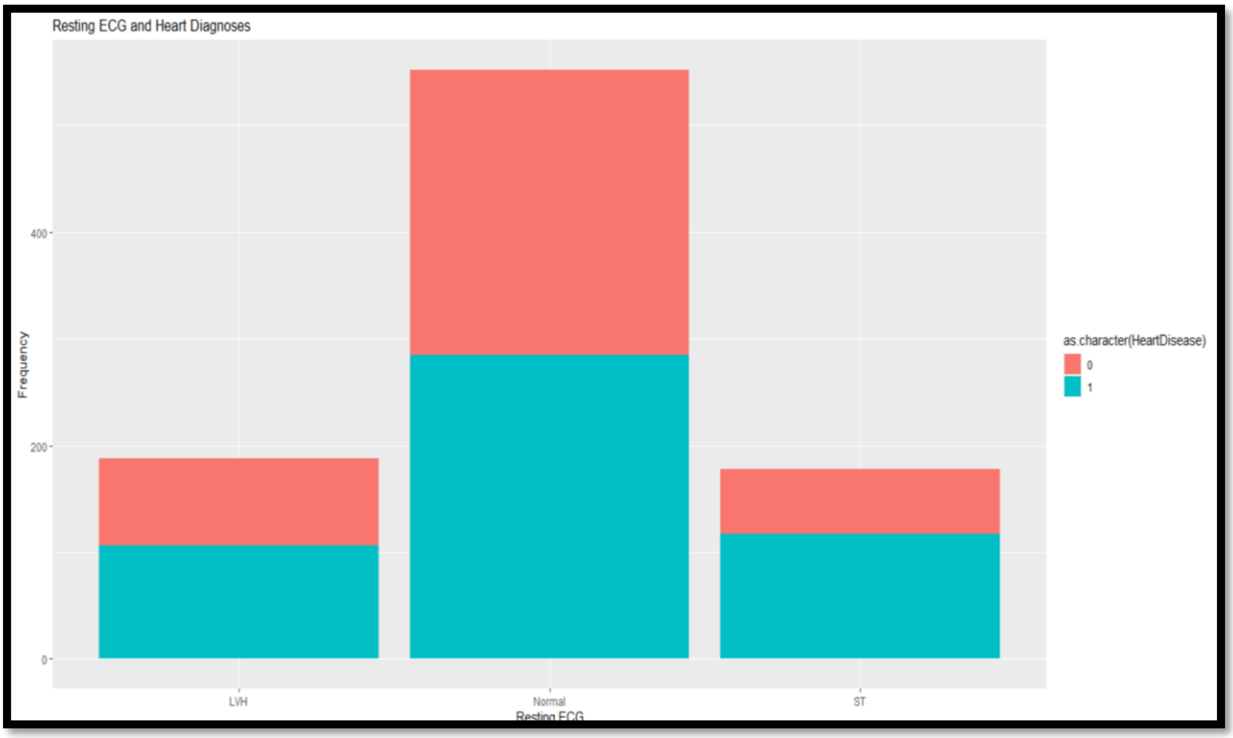


• CONCLUSION:

HeretheXaxiscontains allthe age groupsfrom20to 77brokenas20-34,35-44,45-54, 55-64, 65-77. The y axis contains all the frequencies and we plot out dataand represent the data by bar diagram and we find that the age 55-64 hasrecorded most number of heart diseases and the minimal heart diseases arepresent in the age group of 20-34 which is almost close to 5 persons from thedata given to us . Thus people in the age interval between 55-64 should takegoodcareandeatsfoodsthatinitiatemoregoodhearthealth

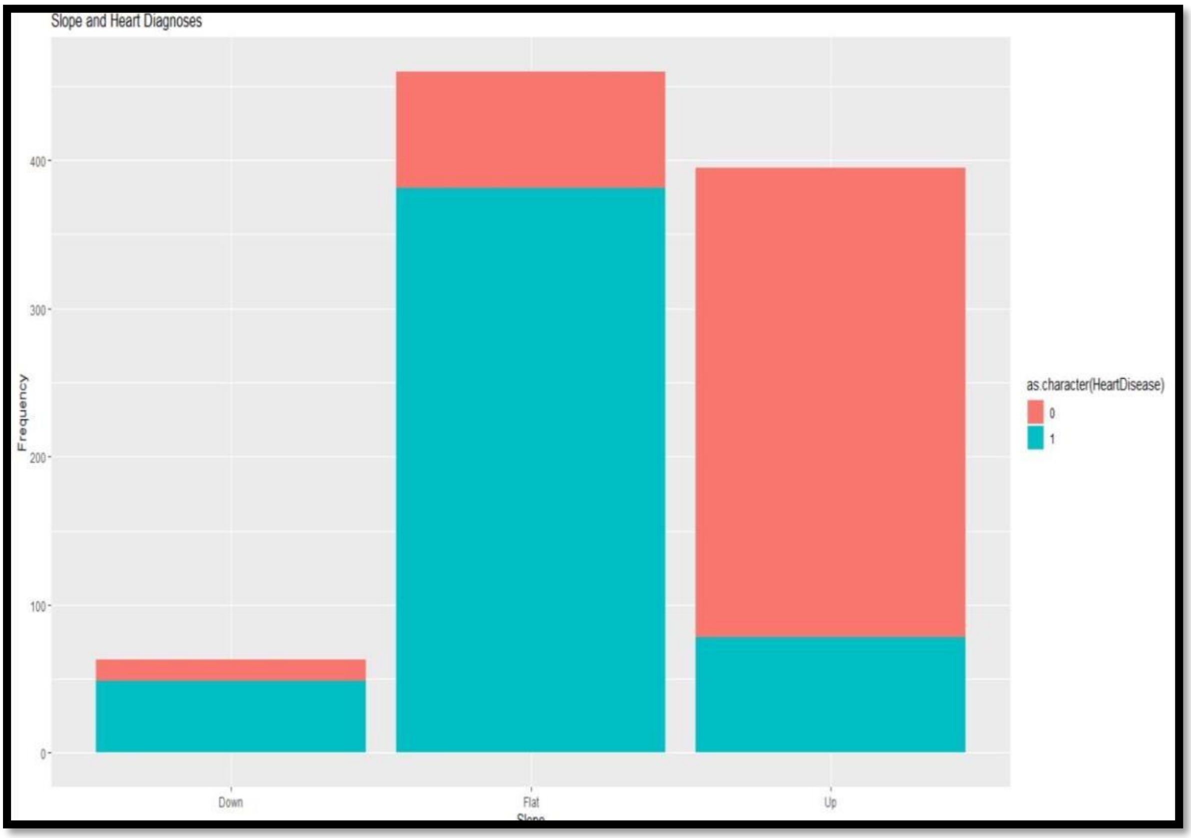
RestingECGdistributedamongstHeartDiagnosis

ggplot(data=hdf,aes(x=RestingECG,fill=as.character(HeartDisease)))+geom_bar()+labs(title='R estingECGandHeartDiagnoses',x='RestingECG',y ='Frequency')



STSlopedistributedamongHeartDiagnosis

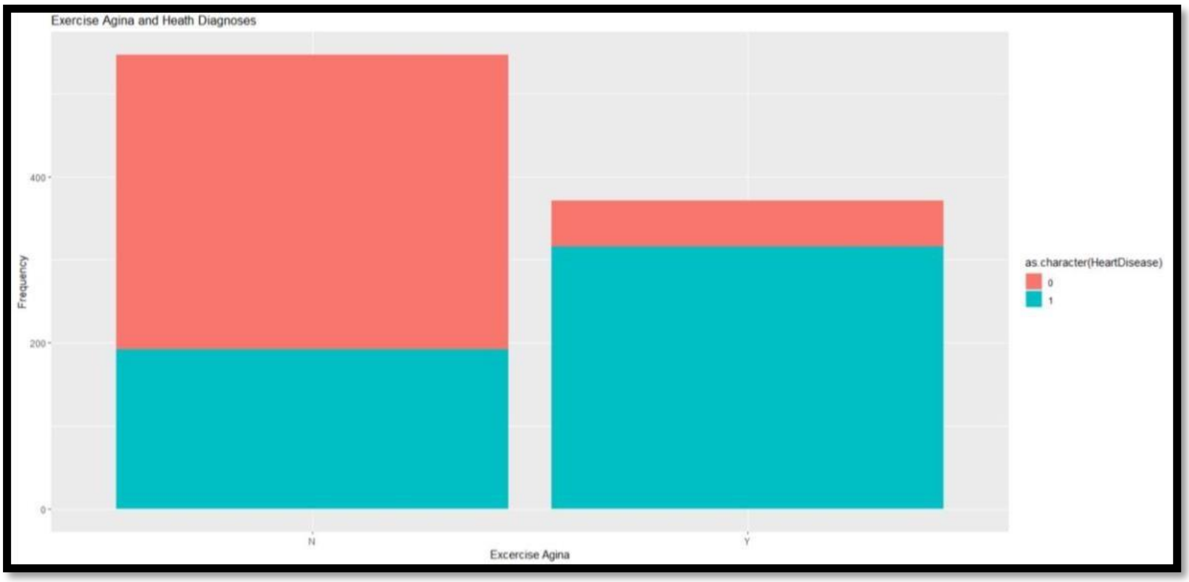
ggplot(data=hdf,aes(x=ST_Slope,fill=as.character(HeartDisease)))+geom_bar()+labs(title= 'SlopeandHeartDiagnoses',x ='Slope', y='Frequency')



• **CONCLUSION:** *The x axis contains the ST Slope which constitutes of UP,DOWN, FLAT and we find thatand the Y axis contains the frequencies and weconstruct a bar diagram to get a wider view on heart diagnosis and slope andwe find that down slope is too short and the flat slope is what most people haveand maximum people with heart disease have flat ST Slope and most peoplewithupslopearefreefromheartdisease.*

ExerciseAnginaDistributedAmongHeartdiagnosis

```
ggplot(data=hdf,aes(x=ExerciseAngina,fill=as.character(HeartDisease)))+geom_bar()+
  labs(title='ExerciseAginaandHeathDiagnoses',
     caption="'N'connotesnoExerciseAginaand'Y'connotesthepresenceofExerciseAgina",
     x='ExcerciseAgina',y
     = 'Frequency')
```



- **CONCLUSION:**

The X axis contains Exercise Angina which contains 0 and and the Y axis containsfrequencies. We plot bar diagram to get a wider view on Heart Diagnosis and wesee thatpeople positivetoexerciseangina are lesssufferersofheart diseases.

TESTING

1. Performingat-testbetweenHeartDiseaseandCholesterol:

t.test(hr$Cholesterol~hr$HeartDisease,mu=0,conf=0.95,var.eq=F,Paired=F)

data:hr$Cholesterolbyhr$HeartDiseaset=0.
37548,df=891.63,p-value =0.7074
alternativehypothesis: truedifferenceinmeansbetweengroup 0and group1isnot equalto0
95 percentconfidenceinterval:
-5.8974548.687877
sampleestimates: meaningroup 0 meaningroup1
236.8195 235.4243

FINDINGS : *Here we want to test if there is any significant difference between the Heart Disease forcholesterol and non-cholesterol Indivivuals is present or not. Our Null hypothesis (H0) is , there is nosignificant difference & our alternative hypothesis (H1) is , there exists true difference between HeartDisease for cholesterol and non-cholesterol Indivivuals. i.e. H0 : μ A – μ B= 0 VS. H1 : μ A – μ B ≠0.Where μ A= person with cholesterol having heart disease . μ B= person with no cholesterol havingheart disease. As the p-value (=0,7074) > 0.05 so we accept H0 at 5% level of significance , soweaccept the null hypothesis and conclude that the null hypothesis is true. So , there is a nosignificant difference between means of cholesterol and non-cholesterol Indivivuals The predictedvaluesof μ A& μ Bare236.8195and235.4243respectively.*

2. Performingat-testbetweenHeartDiseaseandRestingBP:

t.test(hr$RestingBP~hr$HeartDisease,mu=0,conf=0.95,var.eq=F,Paired=F)data

: hr$RestingBPbyhr$HeartDisease
t=-3.6467,df=910.43, p-value=0.0002807
alternativehypothesis: truedifferenceinmeansbetweengroup 0and group1isnot equalto0
95 percentconfidenceinterval:
-6.560571-1.969778
sampleestimates:
mean in group0mean ingroup1
130.1805 134.4457

FINDINGS : *Here we want to test if there is any significant difference between the Heart Diseaseand Resting Blood Pressure. Our Null hypothesis (H0) is , there is no significant difference & ouralternative hypothesis (H1) is , there exists true difference between Heart Disease andRestingBP i.e.H0 : μ A – μ B= 0 VS. H1 : μ A – μ B ≠0. Where μ A= person with high RBP having heart disease . μB= person with low RBP having heart disease. As the p-value (=0.0002) < 0.05 so we reject H0 at5% level of significance , so werejectthe null hypothesis and conclude that the null hypothesis isfalse. So , there is a significant difference between means of high RBP and low RBP.The predictedvaluesof μ A&μ B are130.1805and 133.4457respectively.*

3. Performinga t-testbetweenHeartDiseaseandMaxHR:

t.test(hr$MaxHR~hr$HeartDisease,mu=0,conf=0.95,var.eq=F,Paired=F)

data:hr$MaxHRbyhr$HeartDisease
t=13.231,df=877.04,p-value <2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 095percentconfidence interval:
17.4555123.53591
sampleestimates: mean ingroup0mean ingroup1
148.1512127.6555

FINDINGS *: Here we want to test if there is any significant difference between the Heart Diseaseand Max HR. Our Null hypothesis (H0) is , there is no significant difference & our alternativehypothesis(H1)is,thereexiststruedifferencebetweenHeartDiseaseforIndivivualswithhighbloodpressure. i.e. H0 : μ A – μ B= 0 VS. H1 : μ A – μ B ≠0. Where μ A= person with max Hr high havingheart disease . μ B= person with low or accurate resting BPhaving heart disease. As the p-value(=*2.2e-16*) < 0.05 so we rejectH0 at 5% level of significance , so we reject the null hypothesis andconclude that the null hypothesis is false. So , there is a a significant difference between means ofIndivivuals.Thepredicted valuesofμA&μBare141.5152and127.6555respectively.*

4. Performingachi-squaredtestbetweenHeartDiseaseandSex :

chisq.test(hr$HeartDisease,hr$Sex,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseandhr$Sex
X-squared=85.646,df=1,p-value <2.2e-16

FINDINGS *:Here we take null hypothesis (HO) to be as both the factors equal.i.e sex and heartdisease and alternative hypothesis that the factors are never equal. Here the p-value is (2.2e-16)<0.05 so we rejectH0 at 5% level of significance , so we reject the null hypothesis and concludethatthenullhypothesis isfalse.Thereforesignificantfactor.*

5. Performingchi-squaredbetweenHeartDiseaseandFastingBS:

chisq.test(hr$HeartDisease,hr$FastingBS,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseandhr$FastingBS
X-squared= 65.586, df= 1, p-value= 5.563e-16

FINDINGS *:Here we take null hypothesis (HO) to be as both the factors equal .i.eFastingBSandheart disease and alternative hypothesis that the factors are never equal. Here the p-value is(5.563e-16)<0.05 so we rejectH0 at 5% level of significance , so we reject the null hypothesis andconcludethatthenullhypothesisisfalse.Thereforesignificantfactor.*

6. Performingachi-squaredbetweenHeartDiseaseandChestPainType:

chisq.test(hr$HeartDisease,hr$ChestPainType,correct=FALSE)

Pearson's Chi-squared test data: hr$HeartDisease and hr$ChestPainType X-squared = 268.07, df =3,p-value <2.2e-16

FINDINGS *:Here we take null hypothesis (HO) to be as both the factors equal .i.e Chest Pain Typeand heart disease and alternative hypothesis that the factors are never equal. Here the p-value is(2.2e-16)<0.05 so we rejectH0 at 5% level of significance , so we reject the null hypothesis andconcludethatthenullhypothesisisfalse.Thereforesignificantfactor.*

7. Performingachi-squaredbetweenHeartDiseaseandResting ECG:

chisq.test(hr$HeartDisease,hr$RestingECG,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseandhr$RestingECG
X-squared=10.931,df=2,p-value=0.004229

FINDINGS *:Here we take null hypothesis (HO) to be as both the factor equal .i.eRestingECG andheart disease and alternative hypothesis that the factors are never equal. Here the p-value is(0,004229)<0.05 so we reject H0 at 5% level of significance , so we reject the null hypothesis andconcludethatthenullhypothesisisfalse.Thereforesignificantfactor.*

8. Performinga chi-squaredbetweenHeartDiseaseandExerciseAngina:

chisq.test(hr$HeartDisease,hr$ExerciseAngina,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseandhr$ExerciseAngina
X-squared=224.28,df=1,p-value<2.2e-16

FINDINGS *:Here we take null hypothesis (HO) to be as both the factors are equal .i.e ExerciseAgnina and heart disease and alternative hypothesis that the factors are never equal. Here the p-value is* (2.2e-16)*<0.05 so we reject H0 at 5% level of significance , so we reject the null hypothesisand concludethatthenullhypothesis isfalse.Thereforesignificanteffect.*

9. Performingachi-squaredbetweenHeartDiseaseandOldpeak:

chisq.test(hr$HeartDisease,hr$Oldpeak,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseand hr$Oldpeak
X-squared= 230.51, df =52, p-value< 2.2e-16

FINDINGS :*Here we take null hypothesis (HO) to be equal as both the factors .i.eOldpeak and heartdisease are equal under HO and alternative hypothesis that the factors are never equal. Here the p-value is* (2.2e-16*)<0.05 so we reject H0 at 5% level of significance , so we reject the null hypothesisand concludethatthenullhypothesisisfalse.Thereforesignificanteffect.*

10.Performingachi-squaredbetweenHeartDiseaseandSTSlope:

chisq.test(hr$HeartDisease,hr$ST_Slope,correct=FALSE)

Pearson'sChi-squaredtest
data:hr$HeartDiseaseandhr$ST_Slope
X-squared=355.92,df=2,p-value<2.2e-16

FINDINGS:*Here wetakenullhypothesis(HO)tobeequalasboththe factors.i.eST_Slope and heart disease are equal under HO and alternative hypothesis that the factors are never equal.Here the p-value is* 2.2e-16*)<0.05 so we reject H0 at 5% level of significance , so we reject the nullhypothesisand conclude thatthe nullhypothesisisfalse.Thereforesignificanteffect.*

SIMPLELINEARREGRESSION

We find the simple linear regression of the Sex, Chest Pain Type, Fasting BS, Resting ECG, ExerciseAngina, Old peak, ST_Slope to Heart Disease. Since Adjusted R-square gives a better insight thenChi square leadsto.

1. SimpleLinearRegressionbetweenHeartdiseaseandSex:

```
>rl.sex.hd<-lm(HeartDisease~Sex,data=hr)
>summary(rl.sex.hd)
```

Call:
lm(formula=HeartDisease~Sex,data=hr)

Residuals: Min 1Q Median3Q Max
-0.6317-0.6317 0.3683 0.36830.7409
Coefficients:
Estimate Std.Error tvalue Pr(>|t|)
(Intercept)0.25907 0.03411 7.595
7.61e-14***
SexM 0.372660.038389.708 <2e-16***

Signif.codes:0'***'0.001'**'0.01'*'0.05'.'0.1''
1 Residual standard error: 0.4739 on 916 degrees of freedom Multiple R-squared: 0.0933,AdjustedR-squared: 0.09231F-statistic: 94.25on 1and916DF,p-value:<2.2e-16

2. Simple Linear Regression between Chest Pain Type and HeartDiseases:

```
> rl.cpt.hd<-lm(HeartDisease~ChestPainType,data=hr)
> summary(rl.cpt.hd)
```

Call:
lm(formula=HeartDisease~ChestPainType,data=hr)Res
iduals:Min 1Q Median 3Q Max
-0.7903-0.3547 0.2097 0.2097 0.8613
Coefficients:
EstimateStd.Error tvalue
Pr(>|t|)(Intercept)0.79032 0.01882
41.986 <2e-16***
ChestPainTypeATA-0.651590.03702-17.603<2e-16***
ChestPainTypeNAP-0.435640.03493 -12.472 <2e-16***
ChestPainTypeTA-0.355540.06461 -5.503 4.86e-08***---

Signif.codes:0 '***'0.001'**' 0.01'*'0.05 '.'0.1' '1Residualstandarderror:0.4192on 914 degrees of freedom Multiple R-squared: 0.292, Adjusted R-squared: 0.2897 F-statistic:125.7 on 3 and 914 DF, p-value:< 2.2e-16

3. Simple Linear Regression between Old peak and HeartDiseases:

```
> rl.op.hd<-lm(HeartDisease~Oldpeak,data=hr)
> summary(rl.op.hd)
```

Call:
lm(formula = HeartDisease ~ Oldpeak, data =
hr)Residuals:

Min	1Q	Median	3Q	Max
-1.17744	-0.38621	0.04862	0.42540	1.10361

Coefficients:

	Estimat	Std.Error	tvalue	Pr(>|t|)
(Intercept)	0.38621	0.01955	19.75	<2e-16***
Oldpeak	0.18839	0.01410	13.37	<2e-16***

Signif.codes:0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1

Residual standard error: 0.4553 on 916 degrees of freedom Multiple R-squared: 0.1632,
AdjustedR-squared: 0.1623F-statistic:178.6on1and916DF,p-value:<2.2e-16

4. Simple Linear Regression between Exercise Angina and HeartDiseases:

```
> rl.a.hd<-lm(HeartDisease~ExerciseAngina,data=hr)
> summary(rl.a.hd)
```

Call:
lm(formula=HeartDisease~ExerciseAngina,data=hr)

Residuals:

Min	1Q	Median	3Q	Max
-0.8518	-0.3510	0.1482	0.1482	0.6490

Coefficients:

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	0.3510	0.0185	18.98	<2e-16***
ExerciseAnginaY	0.5008	0.0291	17.21	<2e-16***

Signif.codes:0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1
Residual standard error: 0.4326 on 916 degrees of freedom Multiple R-squared: 0.2443,
AdjustedR-squared: 0.2435F-statistic:296.1on1and916DF,p-value:<2.2e-16

CONCLUSION : *categorical data were present so I had to use chi-square test toget their p values but adjusted R square cannot be found from chi square so simplelinear regression is used to find out which variables increases the efficiency or alsogives us adjusted R square. The significant factors now can be used to constructmultipleregression.*

MultipleLinearRegression

We perform Multiple Linear Regression with the significant factors and HeartDiseaseso thatourmodelgetshigh accuracy.

```
> rlsn.multiple.hd<-
lm(HeartDisease~MaxHR+ChestPainType+ExerciseAngina+Oldpeak+ST_Slope,data=hr)
>summary(rlsn.multiple.hd)
```

Call:
lm(formula=HeartDisease~MaxHR+ChestPainType+ExerciseAngina+Oldpeak+ST_Sl
ope,data= hr)

```
Residuals:
      Min        1Q     Median     3Q      Max
-1.06949-0.11844 -0.00217 0.18087 1.04687
Coefficients:
                    EstimateStd.        Errort
valuePr(>|t|)(Intercept)
                        0.89689650.085256610.520<2e-16***
MaxHR
-0.00181610.0005111-3.553 0.000400 ***
ChestPainTypeATA -0.30650400.0350746 -8.739<2e-16***
ChestPainTypeNAP-0.24656330.0307143 -8.0283.06e-15***
ChestPainTypeTA-0.18563550.0550076-3.3750.000770***
ExerciseAnginaY 0.1358555 0.0288729 4.705 2.93e-06 ***
Oldpeak0.04365820.01292473.3780.000761 ***
ST_SlopeFlat0.10917220.04789842.2790.022884*
ST_SlopeUp -0.28311130.0532640-5.3151.34e-07***
---
Signif.codes:0 '***' 0.001 '**' 0.01 '*'0.05 '.' 0.1 ' ' 1
Residualstandarderror:0.3451 on909degrees offreedomMultipleR-squared:0.5228,AdjustedR-squared:0.5186 F-statistic:124.5on 8and909 DF, p-value:<2.2e-16
```

CONCLUSION:

After computing the multiple regression we can see that R-squared adjusted is0.5186 which roughly coverts to 51.86%. Thus we can say that the accuracy of myheartfailurepredictionmodelis51.86%thatmeansthesignificant factorspresentsum up to the fact that most of the significant factors are a reason for heart failure51.86% oftimes.

|Conclusion:

Visualizations give us an idea that men are more likely to suffer from heartdiseases than women and the percentage is around 63% for men and 26% for thatof women. So men aboveage 35 are advised to take good care of heart's healthandeatfoodsthatregulatemuch more flowofblood in heart.

It also gives us an idea that cholesterol should be kept under control as it might beanunderlyingcausefor seriousheart diseasesandsomemightevenbe fatalinlongrun.

Max Heart Rate should be between 150-170 as the visualizations have showedpeople with maximum heart rate in this range are less probable to get affetced byanyheartrelatedissuesorheartdiseases,butmaximumheartratemustbegreaterthan100as lessthan 100isa primecauseforany fatal heart relatedissues.

Eventestinggives usanideathatMaxHRisasignificant factor.

Chest pain type is also a significant factor because there are types or varities, anypersonsufferingfromasymptomaticchestpainwillfeeldizziness,stabbingkindofpainbutthe chancesofheart failure isless.

Exercise Angina is a pain in the chest that comes with exercise, stress, or otherthings that make the heart work harder as cholesterol gets clogged in coronaryarteries.Stillexercisingisadvisedsinceits best way forin andoutflowof oxygen.

The prediction model I created gives an accuracy of 51.86% that means it giveresultsthatcan betrusted more than 50%ofthe time.

I could have given much more effort to increase the accuracy but at this momentfrom all the knowledge I have gathered and the time given to me for completetionof the projectdidnotallowmeorhelptoperformany better.

|REFERENCES:

1. IntroductoryStatisticswithR,PeterDaalgard.
2. KAGGLE.COM
3. Fundamentals of statistics,Vol-I&Vol-II,Gun,Gupta,Dasgupta.
4. WIKIPEDIA.
5. GEEKSFORGEEKS.
6. TUTORIALSPOINT.

0

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina
40	M	ATA	140	289	0	Normal	172	N
49	F	NAP	160	180	0	Normal	156	N
37	M	ATA	130	283	0	ST	98	N
48	F	ASY	138	214	0	Normal	108	Y
54	M	NAP	150	195	0	Normal	122	N
39	M	NAP	120	339	0	Normal	170	N
45	F	ATA	130	237	0	Normal	170	N
54	M	ATA	110	208	0	Normal	142	N
37	M	ASY	140	207	0	Normal	130	Y
48	F	ATA	120	284	0	Normal	120	N
37	F	NAP	130	211	0	Normal	142	N
58	M	ATA	136	164	0	ST	99	Y
39	M	ATA	120	204	0	Normal	145	N
49	M	ASY	140	234	0	Normal	140	Y
42	F	NAP	115	211	0	ST	137	N
54	F	ATA	120	273	0	Normal	150	N
38	M	ASY	110	196	0	Normal	166	N
43	F	ATA	120	201	0	Normal	165	N
60	M	ASY	100	248	0	Normal	125	N
36	M	ATA	120	267	0	Normal	160	N
43	F	TA	100	223	0	Normal	142	N
44	M	ATA	120	184	0	Normal	142	N
49	F	ATA	124	201	0	Normal	164	N
44	M	ATA	150	288	0	Normal	150	Y
40	M	NAP	130	215	0	Normal	138	N
36	M	NAP	130	209	0	Normal	178	N
53	M	ASY	124	260	0	ST	112	Y
52	M	ATA	120	284	0	Normal	118	N
53	F	ATA	113	468	0	Normal	127	N
51	M	ATA	125	188	0	Normal	145	N
53	M	NAP	145	518	0	Normal	130	N
56	M	NAP	130	167	0	Normal	114	N
54	M	ASY	125	224	0	Normal	122	N
41	M	ASY	130	172	0	ST	130	N
43	F	ATA	150	186	0	Normal	154	N
32	M	ATA	125	254	0	Normal	155	N
65	M	ASY	140	306	1	Normal	87	Y
41	F	ATA	110	250	0	ST	142	N
48	F	ATA	120	177	1	ST	148	N
48	F	ASY	150	227	0	Normal	130	Y
54	F	ATA	150	230	0	Normal	130	N
54	F	NAP	130	294	0			
43	M	ASY	120	175	0	Normal	120	Y
59	M	NAP	130	318	0	Normal	120	Y
37	M	ASY	120	223	0	Normal	168	N
50	M	ATA	140	216	0	Normal	170	N

36	M	NAP	112	340	0	Normal	184	N
41	M	ASY	110	289	0	Normal	170	N
50	M	ASY	130	233	0	Normal	121	Y
47	F	ASY	120	205	0	Normal	98	Y
45	M	ATA	140	224	1	Normal	122	N
41	F	ATA	130	245	0	Normal	150	N
52	F	ASY	130	180	0	Normal	140	Y
51	F	ATA	160	194	0	Normal	170	N
31	M	ASY	120	270	0	Normal	153	Y
58	M	NAP	130	213	0	ST	140	N
54	M	ASY	150	365	0	ST	134	N
52	M	ASY	112	342	0	ST	96	Y
49	M	ATA	100	253	0	Normal	174	N
43	F	NAP	150	254	0	Normal	175	N
45	M	ASY	140	224	0	Normal	144	N
46	M	ASY	120	277	0	Normal	125	Y
50	F	ATA	110	202	0	Normal	145	N
37	F	ATA	120	260	0	Normal	130	N
45	F	ASY	132	297	0	Normal	144	N
32	M	ATA	110	225	0	Normal	184	N
52	M	ASY	160	246	0	ST	82	Y
44	M	ASY	150	412	0	Normal	170	N
57	M	ATA	140	265	0	ST	145	Y
44	M	ATA	130	215	0	Normal	135	N
52	M	ASY	120	182	0	Normal	150	N
44	F	ASY	120	218	0	ST	115	N
55	M	ASY	140	268	0	Normal	128	Y
46	M	NAP	150	163	0	Normal	116	N
32	M	ASY	118	529	0	Normal	130	N
35	F	ASY	140	167	0	Normal	150	N
52	M	ATA	140	100	0	Normal	138	Y
49	M	ASY	130	206	0	Normal	170	N
55	M	NAP	110	277	0	Normal	160	N
54	M	ATA	120	238	0	Normal	154	N
63	M	ASY	150	223	0	Normal	115	N
52	M	ATA	160	196	0	Normal	165	N
56	M	ASY	150	213	1	Normal	125	Y
66	M	ASY	140	139	0	Normal	94	Y
65	M	ASY	170	263	1	Normal	112	Y
53	F	ATA	140	216	0	Normal	142	Y
43	M	TA	120	291	0	ST	155	N
55	M	ASY	140	229	0	Normal	110	Y
49	F	ATA	110	208	0	Normal	160	N

39	F	NAP	110	182	0	ST	180	N
58	M	ASY	130	263	0	Normal	140	Y
43	M	ATA	142	207	0	Normal	138	N
39	M	NAP	160	147	1	Normal	160	N
56	M	ASY	120	85	0	Normal	140	N
41	M	ATA	125	269	0	Normal	144	N
65	M	ASY	130	275	0	ST	115	Y
51	M	ASY	130	179	0	Normal	100	N
40	F	ASY	150	392	0	Normal	130	N
40	M	ASY	120	466	1	Normal	152	Y
46	M	ASY	118	186	0	Normal	124	N
57	M	ATA	140	260	1	Normal	140	N
48	F	ASY	120	254	0	ST	110	N
34	M	ATA	150	214	0	ST	168	N
50	M	ASY	140	129	0	Normal	135	N
39	M	ATA	190	241	0	Normal	106	N
59	F	ATA	130	188	0	Normal	124	N
57	M	ASY	150	255	0	Normal	92	Y
47	M	ASY	140	276	1	Normal	125	Y
38	M	ATA	140	297	0	Normal	150	N
49	F	NAP	130	207	0	ST	135	N
33	F	ASY	100	246	0	Normal	150	Y
38	M	ASY	120	282	0	Normal	170	N
59	F	ASY	130	338	1	ST	130	Y
35	F	TA	120	160	0	ST	185	N
34	M	TA	140	156	0	Normal	180	N
47	F	NAP	135	248	1	Normal	170	N
52	F	NAP	125	272	0	Normal	139	N
46	M	ASY	110	240	0	ST	140	N
58	F	ATA	180	393	0	Normal	110	Y
58	M	ATA	130	230	0	Normal	150	N
54	M	ATA	120	246	0	Normal	110	N
34	F	ATA	130	161	0	Normal	190	N
48	F	ASY	108	163	0	Normal	175	N
54	F	ATA	120	230	1	Normal	140	N
42	M	NAP	120	228	0	Normal	152	Y
38	M	NAP	145	292	0	Normal	130	N
46	M	ASY	110	202	0	Normal	150	Y
56	M	ASY	170	388	0	ST	122	Y
56	M	ASY	150	230	0	ST	124	Y
61	F	ASY	130	294	0	ST	120	Y
49	M	NAP	115	265	0	Normal	175	N
43	F	ATA	120	215	0	ST	175	N
39	M	ATA	120	241	0	ST	146	N
54	M	ASY	140	166	0	Normal	118	Y

56	F	ATA	120	279	0	Normal	150	N
39	M	ASY	110	273	0	Normal	132	N
42	M	ATA	120	198	0	Normal	155	N
43	F	ATA	120	249	0	ST	176	N
50	M	ATA	120	168	0	Normal	160	N
54	M	ASY	130	603	1	Normal	125	Y
39	M	ATA	130	215	0	Normal	120	N
48	M	ATA	100	159	0	Normal	100	N
40	M	ATA	130	275	0	Normal	150	N
55	M	ASY	120	270	0	Normal	140	N
41	M	ATA	120	291	0	ST	160	N
56	M	ASY	155	342	1	Normal	150	Y
38	M	ASY	110	190	0	Normal	150	Y
49	M	ASY	140	185	0	Normal	130	N
44	M	ASY	130	290	0	Normal	100Y54	
M		ATA	160	195	0	ST	130N	
59	M	ASY	140	264	1	LVH	119	Y
49	M	ASY	128	212	0	Normal	96	Y
47	M	ATA	160	263	0	Normal	174	N
42	M	ATA	120	196	0	Normal	150	N
52	F	ATA	140	225	0	Normal	140	N
46	M	TA	140	272	1	Normal	175	N
50	M	ASY	140	231	0	ST	140	Y
48	M	ATA	140	238	0	Normal	118	N
58	M	ASY	135	222	0	Normal	100	N
58	M	NAP	140	179	0	Normal	160	N
29	M	ATA	120	243	0	Normal	160	N
40	M	NAP	140	235	0	Normal	188	N
53	M	ATA	140	320	0	Normal	162	N
49	M	NAP	140	187	0	Normal	172	N
52	M	ASY	140	266	0	Normal	134	Y
43	M	ASY	140	288	0	Normal	135	Y
54	M	ASY	140	216	0	Normal	105	N
59	M	ATA	140	287	0	Normal	150	N
37	M	NAP	130	194	0	Normal	150	N
46	F	ASY	130	238	0	Normal	90	N
52	M	ASY	130	225	0	Normal	120	Y
51	M	ATA	130	224	0	Normal	150	N
52	M	ASY	140	404	0	Normal	124	Y
46	M	ASY	110	238	0	ST	140	Y
54	F	ATA	160	312	0	Normal	130	N
58	M	NAP	160	211	1	ST	92	N
58	M	ATA	130	251	0	Normal	110	N
41	M	ASY	120	237	1	Normal	138	Y
50	F	ASY	120	328	0	Normal	110	Y
53	M	ASY	180	285	0	ST	120	Y

41	F	ATA	125	184	0	Normal	180	N
62	F	TA	160	193	0	Normal	116	N
49	M	ASY	120	297	0	Normal	132	N
42	M	ATA	150	268	0	Normal	136	N
53	M	ASY	120	246	0	Normal	116	Y
57	F	TA	130	308	0	Normal	98	N
47	M	TA	110	249	0	Normal	150	N
46	M	NAP	120	230	0	Normal	150	N
42	M	NAP	160	147	0	Normal	146	N
31	F	ATA	100	219	0	ST	150	N
56	M	ATA	130	184	0	Normal	100	N
50	M	ASY	150	215	0	Normal	140	Y
35	M	ATA	120	308	0	LVH	180	N
35	M	ATA	110	257	0	Normal	140	N
28	M	ATA	130	132	0	LVH	185	N
54	M	ASY	125	216	0	Normal	140	N
48	M	ASY	106	263	1	Normal	110	N
50	F	NAP	140	288	0	Normal	140	Y