Factors Influencing Deaths In Heart Failure Patients

Project Groups11
Parth Deshmukh(S3825055), Subbiah Soundarapandian(S3825012), Sudershan Ravi(S3829895)

• INTRODUCTION

Heart failure is the number one cause for death worldwide, which accounts for over 17 million lives i.e almost 31% of all people die due to heart failures. Most of the cardiovascular diseases can be prevented by addressing behavioural risk factors such as use of tobacco, unhealthy diet, hypertension , diabetes and many other physical inactivites.

In the first phase of this project we will analyze in total 12 features that can be used to predict the cause for the heart failure leading to death. So, we will analyze a dataset of 299 patients with heart failure which was collected in 2015. We will apply categorical data analysis to predict the patients who died and rank the features corresponding to the most important risk factors which lead to death.In this analysis we will be focussing majorly on age, platelets count, serium creatinine level in the blood and ejection fraction features to summarise our findings and find a comprehensive outcome based on this features leading to death due to heart failures.

In the next phase of this project we will be using multinominal regression technique to further extract the key features to make a conclusive statement on the visualized data from phase-1.

• DATA SET SOURCE AND DESCRIPTION

The dataset is referred from (“Heart Failure Prediction”, 2020) with compilation of various datasets from other sources. The dataset has 13 features and 299 observations which describes the reasons for health failures.The description for the dataset is provided below :

Age : Age of Patient.
Anaemia : Decrease of red blood cells or hemoglobin (0:Patient don’t have Anaemia , 1: Patient have Anaemia).
High Blood pressure : Does a patient has hypertension (0:Patient don’t have High Blood Pressure , 1: Patient have High Blood Pressure).
Creatinine phosphokinase : Level of CPK enzyme in the blood.
Diabetes : Does the patient has diabetes (0:Patient don’t have Diabetes, 1: Patient have Diabetes).
Ejection fraction : percentage of blood leaving the heart at each contraction.
Sex : Male or Female (0:female ,1:Male).
Platelets : Platelets in the blood.
Serum Ceratinine : Level of Ceratinine in the blood.
Serium sodium : Level of sodium in the blood.
Smoking : Does the patient smokes (0: Patient don’t smoke, 1: Patient smoke).
Time : Follow-up period.
Death Event : Did the patient died during the follow-up period (0: NO, 1: YES).

• DATA PREPARATION

Loading the packages required and reading the dataset.

```
library(readr)
assign_ds <- read.csv("Project Groups11_data.csv")
```

Handling the missing values and removing the irrelevant column (Time) as it is not necessary for our analysis.

```
assign_ds <- na.omit(assign_ds)
assign_ds <- assign_ds[,-12]
```

Statistical Quantitative description of the category features :-

```
#ANAEMIA
assign_ds_an <- as.data.frame(table(factor(assign_ds[, 2],labels = c("Non-Anaemic","Anaemic"
),levels=c(0,1))))
assign_ds_an
```

```
##          Var1 Freq
## 1 Non-Anaemic  170
## 2     Anaemic  129
```

```
#DIABETES
assign_ds_dia <- as.data.frame(table(factor(assign_ds[, 4],labels = c("Non-Diabetic","Diabeti
c"),levels=c(0,1))))
assign_ds_dia
```

```
##           Var1 Freq
## 1 Non-Diabetic  174
## 2     Diabetic  125
```

```
#SEX
assign_ds_sex <- as.data.frame(table(factor(assign_ds[, 10],labels = c("Female","Male"),level
s=c(0,1))))
assign_ds_sex
```

```
##     Var1 Freq
## 1 Female  105
## 2   Male  194
```

```
#SMOKING
assign_ds_smok <- as.data.frame(table(factor(assign_ds[, 11],labels = c("Non-Smoker","Smoker"
),levels=c(0,1))))
assign_ds_smok
```

```
##         Var1 Freq
## 1 Non-Smoker  203
## 2     Smoker   96
```

```
#DEATH_EVENT
assign_ds_death <- as.data.frame(table(factor(assign_ds[, 12],labels = c("Patient Survived",
"Patient Died"),levels=c(0,1))))
assign_ds_death
```

```
##              Var1 Freq
## 1 Patient Survived  203
## 2     Patient Died   96
```

Subsetting the dataset into people who died and people survived.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

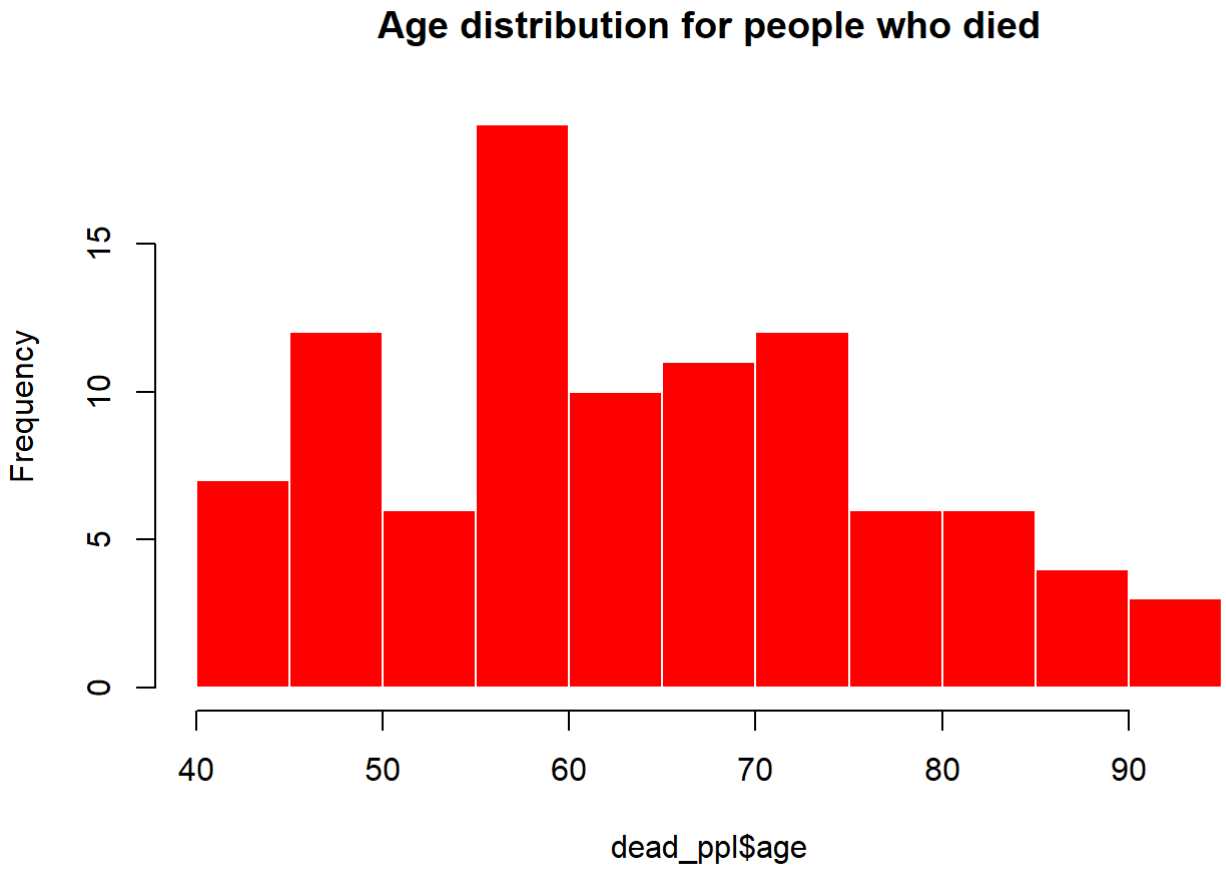
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

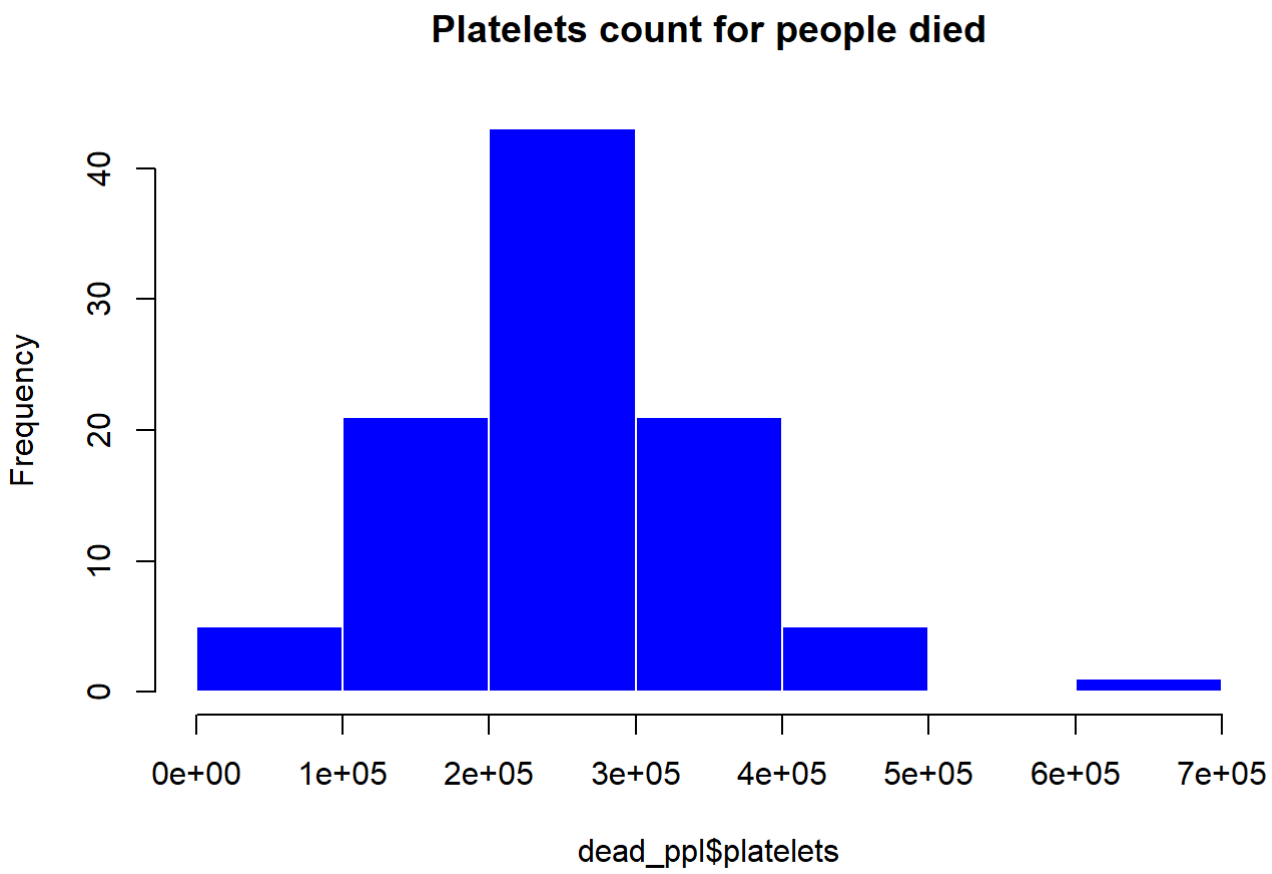
```
dead_ppl <- filter(assign_ds, DEATH_EVENT == 1)
alive_ppl <- filter(assign_ds, DEATH_EVENT == 0)
```

Histogram of multiple factors

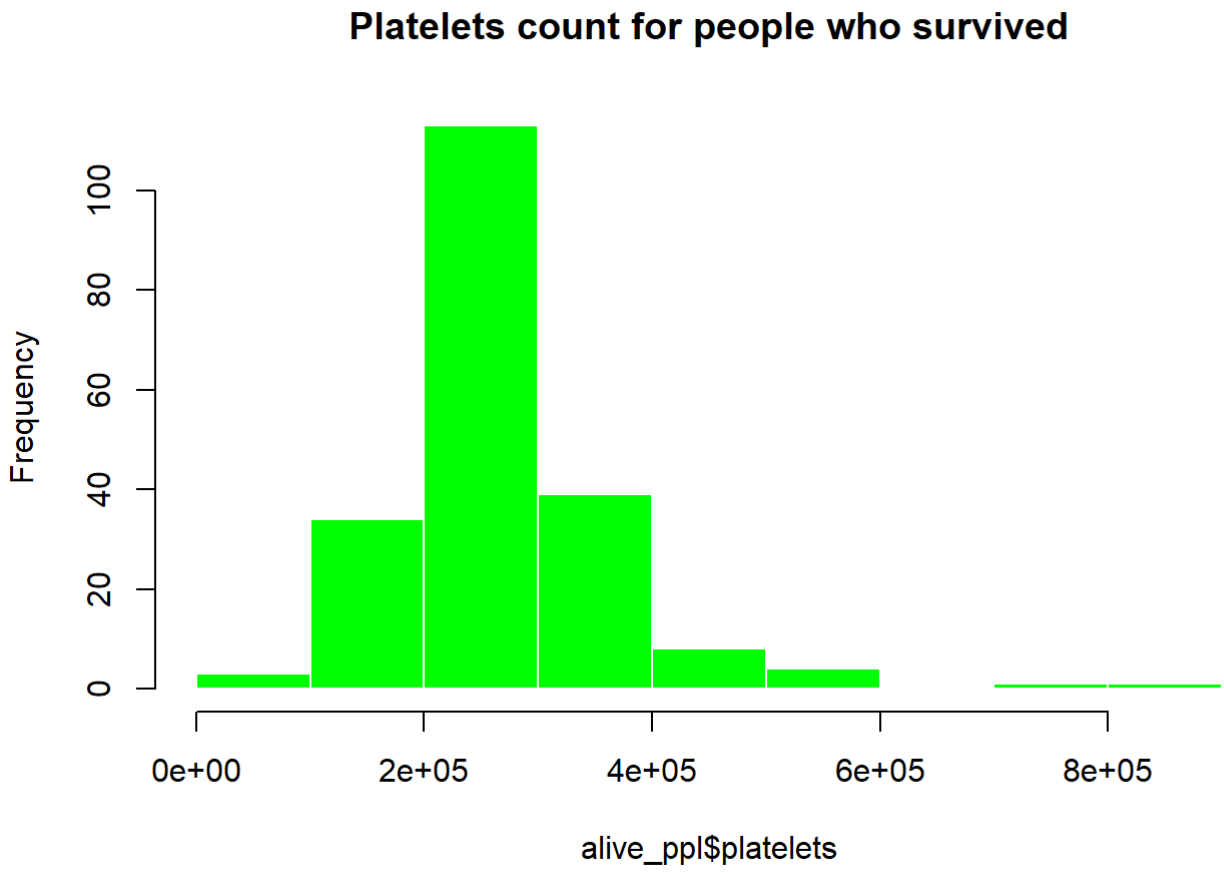
```
#AGE GROUP
hist(dead_ppl$age,border = "white", col = "red", main = "Age distribution for people who die
d")
```




```
#PLATELETS
hist(dead_ppl$platelets,border = "white", col = "blue",main = "Platelets count for people die
d")
```




```
hist(alive_ppl$platelets,border = "white", col = "green",main = "Platelets count for people w
ho survived")
```



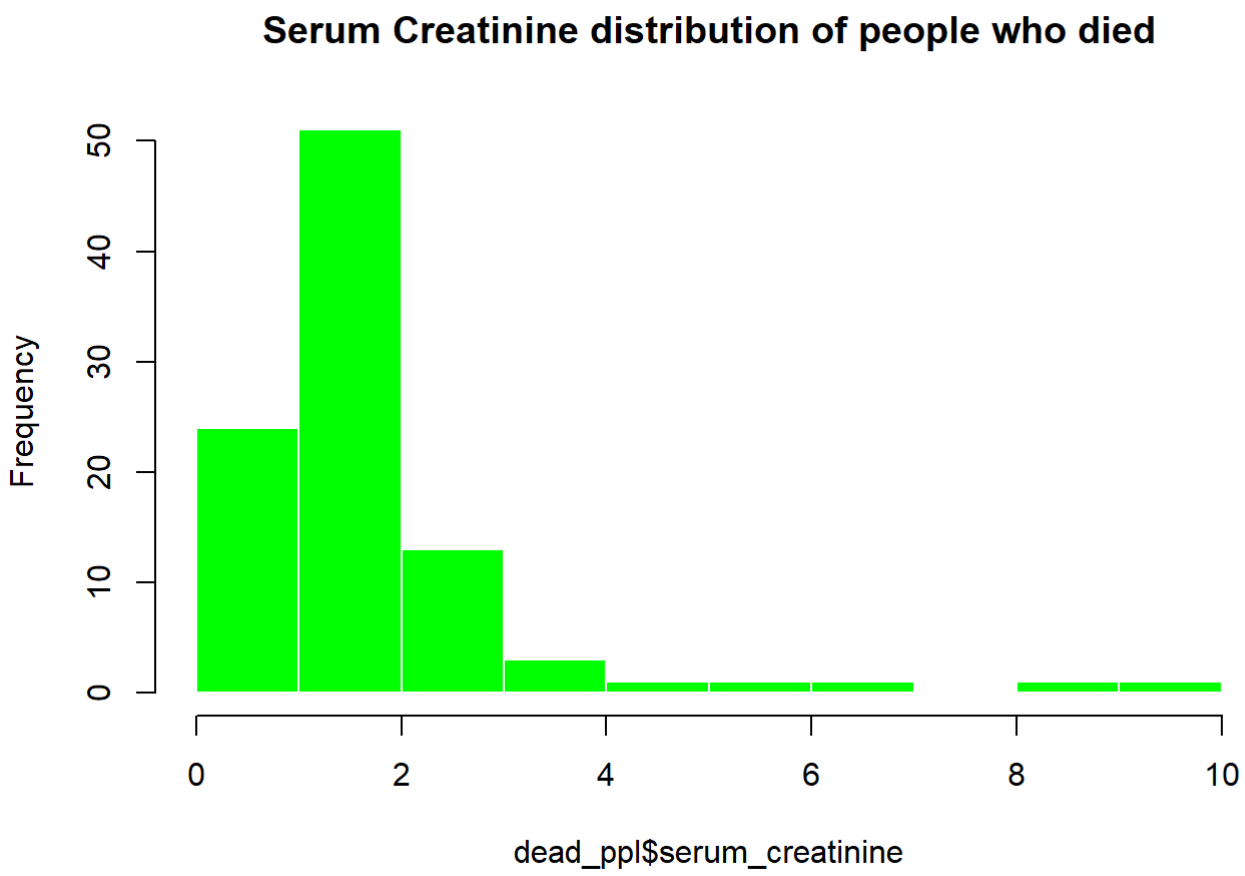
```
summary(dead_ppl$platelets)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   47000  197500  258500  256381  311000  621000
```

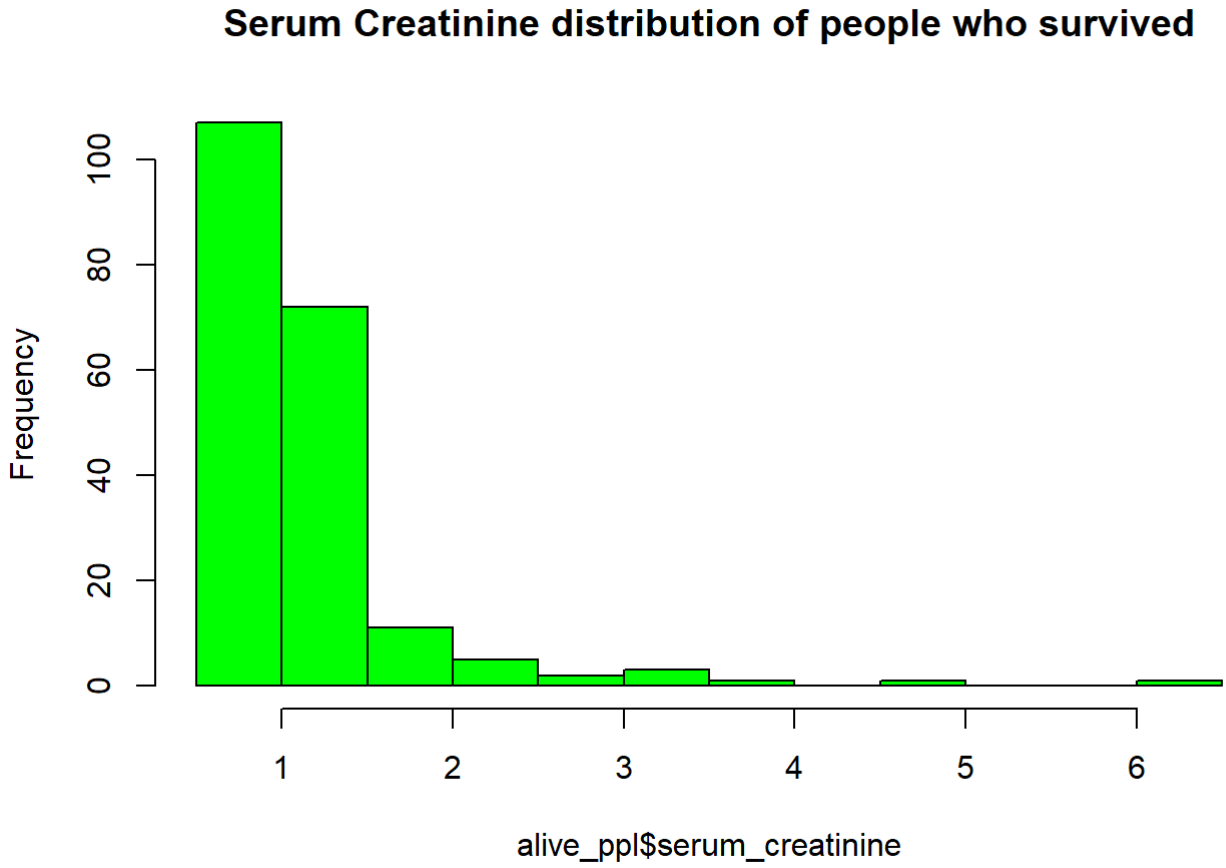
```
summary(alive_ppl$platelets)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25100  219500  263000  266658  302000  850000
```

```
#SERUM CERATININE
hist(dead_ppl$serum_creatinine,border = "white", col = "green", main = "Serum Creatinine dist
ribution of people who died")
```




```
hist(alive_ppl$serum_creatinine,border = "black", col = "green", main = "Serum Creatinine dis
tribution of people who survived")
```



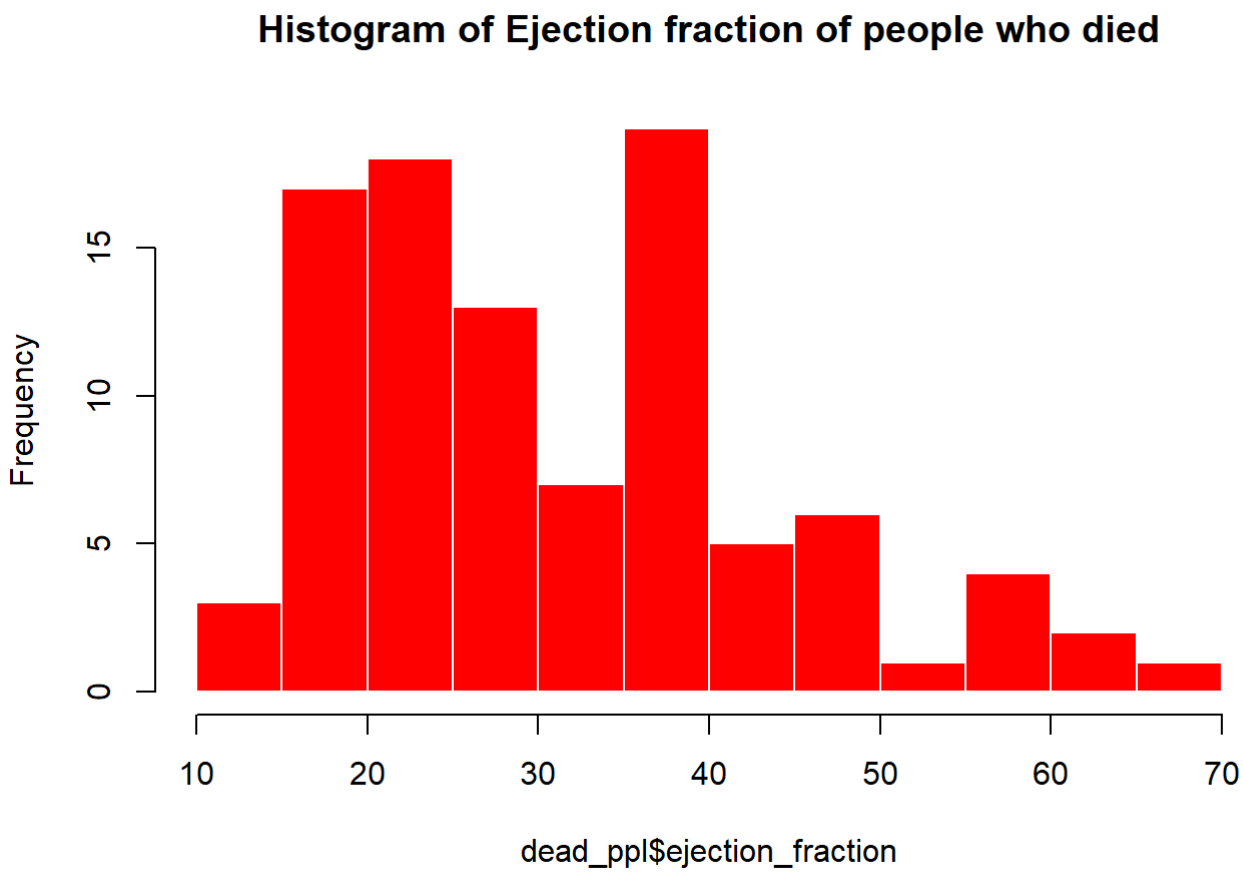

```
summary(dead_ppl$serum_creatinine)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.600   1.075   1.300   1.836   1.900   9.400
```

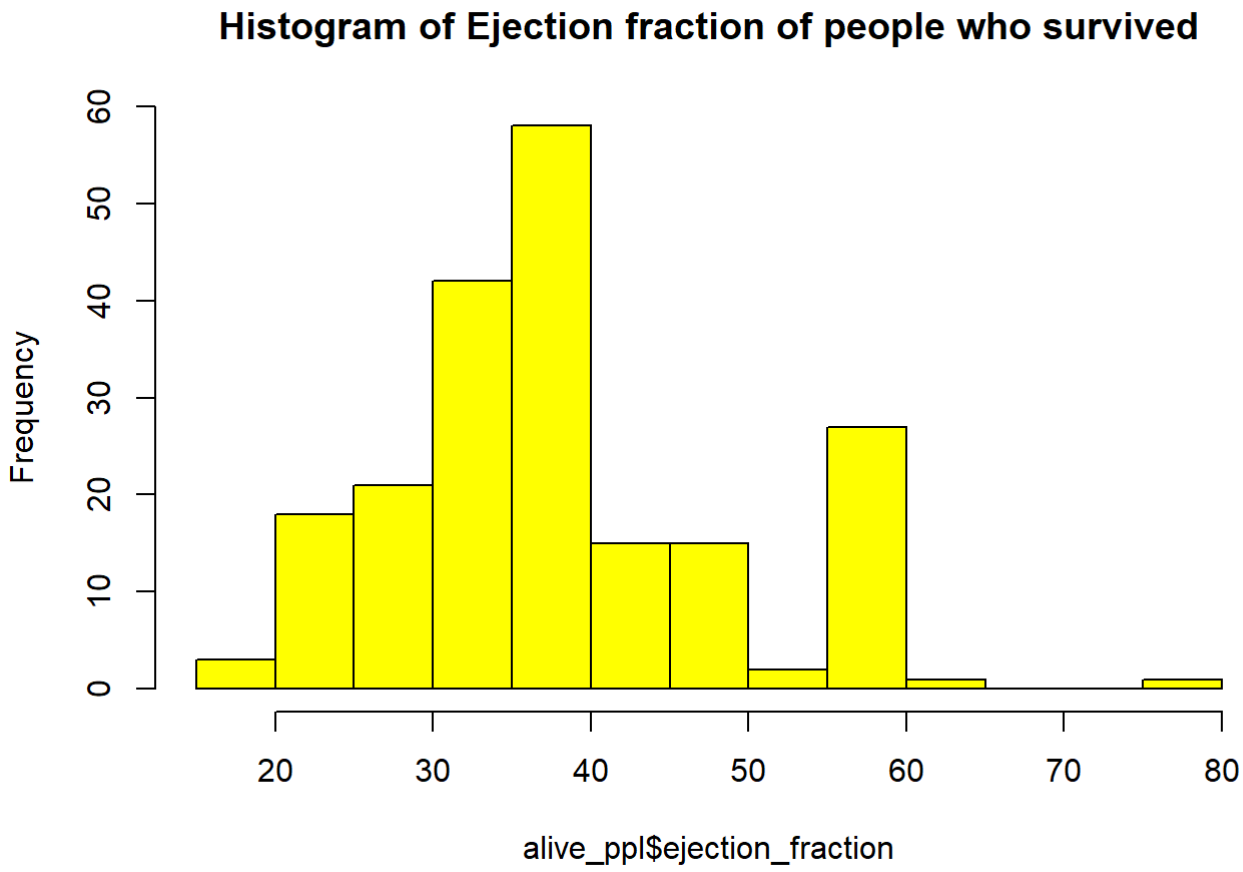
```
summary(alive_ppl$serum_creatinine)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.500   0.900   1.000   1.185   1.200   6.100
```

```
#EJECTION FRACTION
hist(dead_ppl$ejection_fraction,border = "white", col = "red", main = "Histogram of Ejection
 fraction of people who died")
```




```
hist(alive_ppl$ejection_fraction,border = "black", col = "yellow", main = "Histogram of Eject
ion fraction of people who survived")
```



```
summary(dead_ppl$ejection_fraction)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.00   25.00   30.00   33.47   38.00   70.00
```

```
summary(alive_ppl$ejection_fraction)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.00   35.00   38.00   40.27   45.00   80.00
```

Pie charts for different factors.

```
#ANAEMIA
anaemic <- table(assign_ds$DEATH_EVENT,assign_ds$anaemia,dnn = c("Death Event","Anaemic"))
rownames(anaemic) = c("alive","dead")
colnames(anaemic) = c("anaemic","not_anaemic")
anaemic
```

```
##            Anaemic
## Death Event anaemic not_anaemic
##       alive     120          83
##       dead       50          46
```

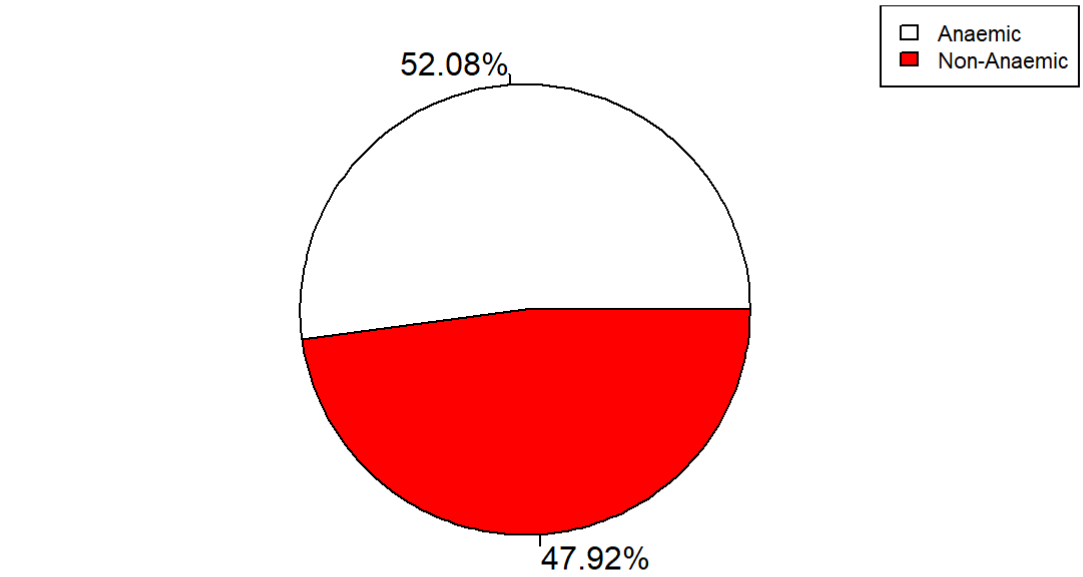
```
anaemic_pct <- anaemic[2,]
anaemic_pct <- round(anaemic_pct/sum(anaemic_pct)*100,2)
anaemic_pct
```

```
##     anaemic not_anaemic
##       52.08       47.92
```

```
labels_an = paste(anaemic_pct, "%", sep = "")
labels_an
```

```
## [1] "52.08%" "47.92%"
```

```
pie(anaemic[2,],labels = labels_an,main = "Pie Chart of Anaemic people who died of heart dise
ase"
,col = c("white","red"))
legend("topright",c("Anaemic","Non-Anaemic"),fill = c("white","red"), cex = 0.7)
```




```
#SMOKING
smoking <- table(assign_ds$DEATH_EVENT,assign_ds$smoking,dnn = c("Death Event","Smoking"))
rownames(smoking) = c("alive","dead")
colnames(smoking) = c("Smoker","Non-Smoker")
smoking
```

```
##            Smoking
## Death Event Smoker Non-Smoker
##       alive    137         66
##       dead      66         30
```

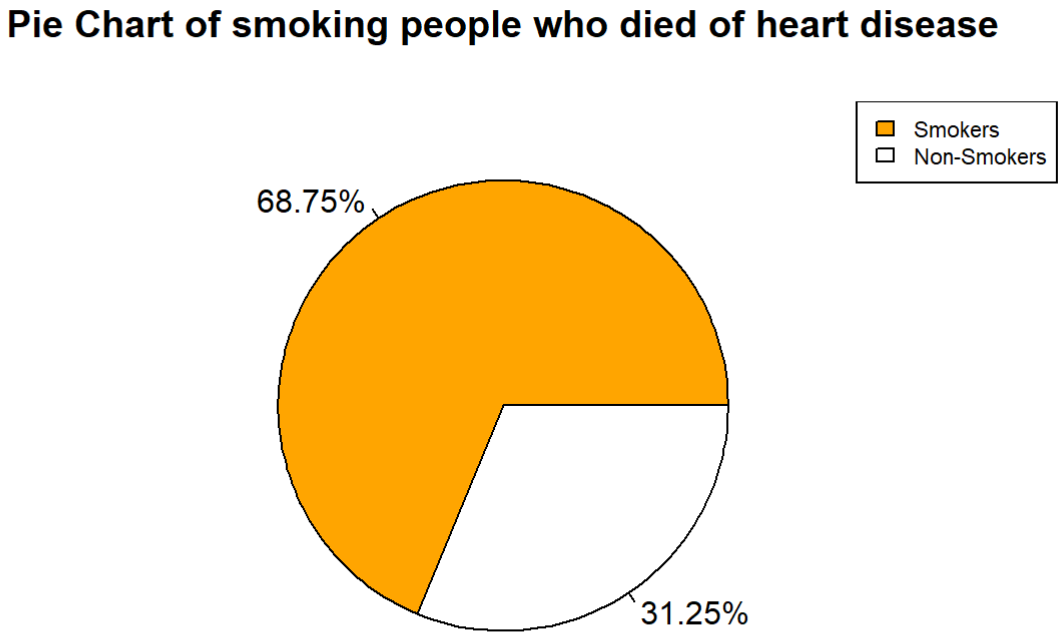
```
smoking_pct <- smoking[2,]
smoking_pct <- round(smoking_pct/sum(smoking_pct)*100,2)
smoking_pct
```

```
##     Smoker Non-Smoker
##      68.75      31.25
```

```
labels_sm = paste(smoking_pct, "%", sep = "")
labels_sm
```

```
## [1] "68.75%" "31.25%"
```

```
pie(smoking[2,],labels = labels_sm,main = "Pie Chart of smoking people who died of heart dise
ase"
    ,col = c("Orange","White"))
legend("topright",c("Smokers","Non-Smokers"),fill = c("Orange","white"),cex = 0.7)
```




```
#DIABETES
diabetes <- table(assign_ds$DEATH_EVENT,assign_ds$diabetes,dnn = c("Death Event","diabetes"))
rownames(diabetes) = c("alive","dead")
colnames(diabetes) = c("Diabetic","Non-Diabetic")
diabetes
```

```
##            diabetes
## Death Event Diabetic Non-Diabetic
##       alive      118           85
##       dead        56           40
```

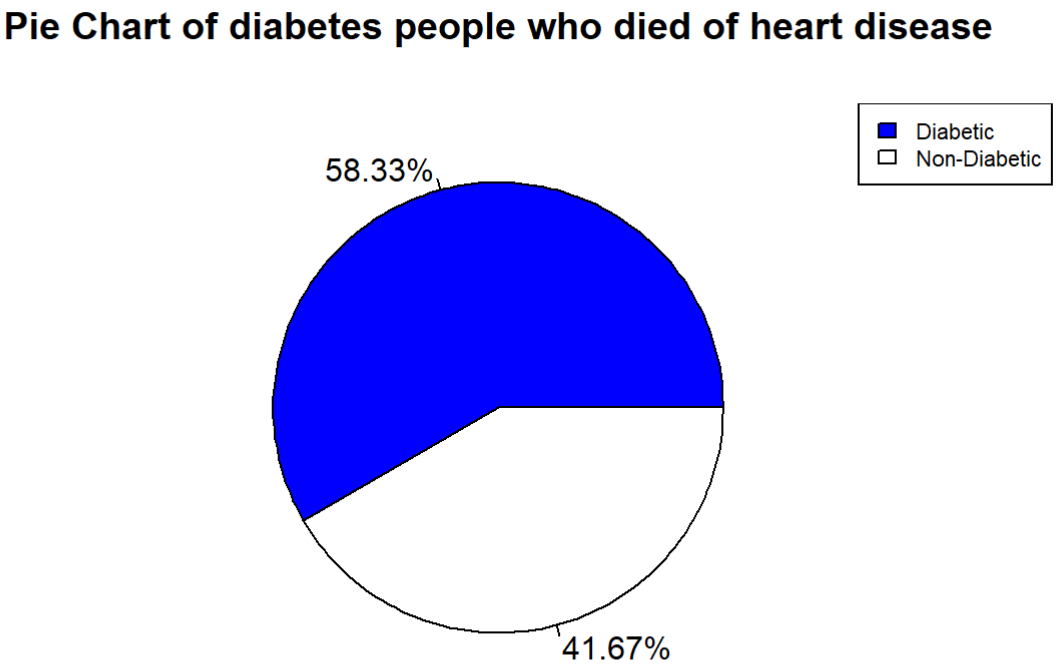
```
diabetes_pct <- diabetes[2,]
diabetes_pct <- round(diabetes_pct/sum(diabetes_pct)*100,2)
diabetes_pct
```

```
##     Diabetic Non-Diabetic
##        58.33        41.67
```

```
labels_db = paste(diabetes_pct, "%", sep = "")
labels_db
```

```
## [1] "58.33%" "41.67%"
```

```
pie(diabetes[2,],labels = labels_db,main = "Pie Chart of diabetes people who died of heart di
sease"
    ,col = c("Blue","White"))
legend("topright",c("Diabetic","Non-Diabetic"),fill = c("Blue","white"), cex = 0.7)
```




```
#BLOOD PRESSURE
high_blood_pressure <- table(assign_ds$DEATH_EVENT,assign_ds$high_blood_pressure,dnn = c("Dea
th Event","high_blood_pressure"))
rownames(high_blood_pressure) = c("alive","dead")
colnames(high_blood_pressure) = c("High_BP","Not High_BP")
high_blood_pressure
```

```
##            high_blood_pressure
## Death Event High_BP Not High_BP
##       alive     137          66
##       dead       57          39
```

```
prop.table(high_blood_pressure)
```

```
##            high_blood_pressure
## Death Event   High_BP Not High_BP
##       alive 0.4581940   0.2207358
##       dead  0.1906355   0.1304348
```

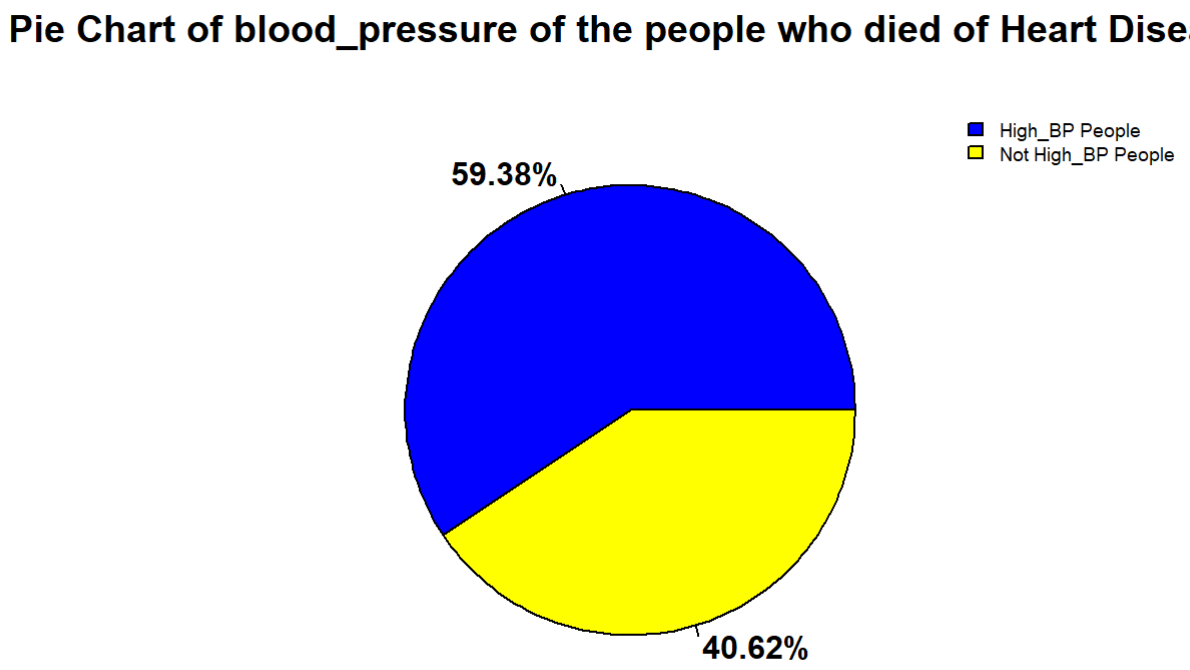
```
high_blood_pressure_pct <- high_blood_pressure[2,]
high_blood_pressure_pct <- round(high_blood_pressure_pct/sum(high_blood_pressure_pct)*100,2)
high_blood_pressure_pct
```

```
##     High_BP Not High_BP
##       59.38       40.62
```

```
labels_bp = paste(high_blood_pressure_pct, "%", sep = "")
labels_bp
```

```
## [1] "59.38%" "40.62%"
```

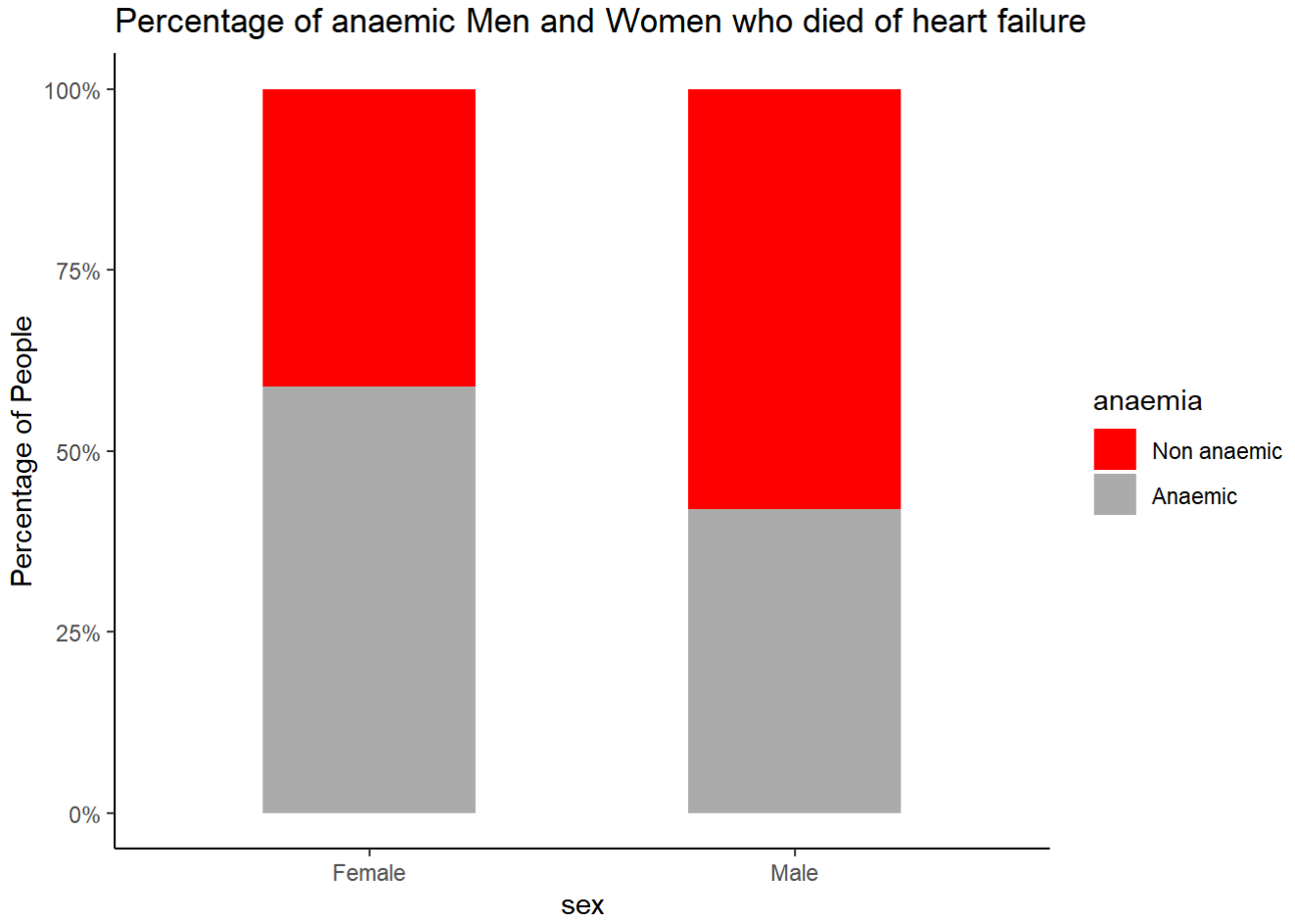
```
pie(high_blood_pressure[2,], labels = labels_bp ,font = 2,main = "Pie Chart of blood_pressure
of the people who died of Heart Disease"
    ,col = c("blue","yellow"))
legend("topright",c("High_BP People","Not High_BP People"),fill = c("blue","yellow"), bty =
"n",cex = 0.6)
```



Bar charts for various factors based on sex.

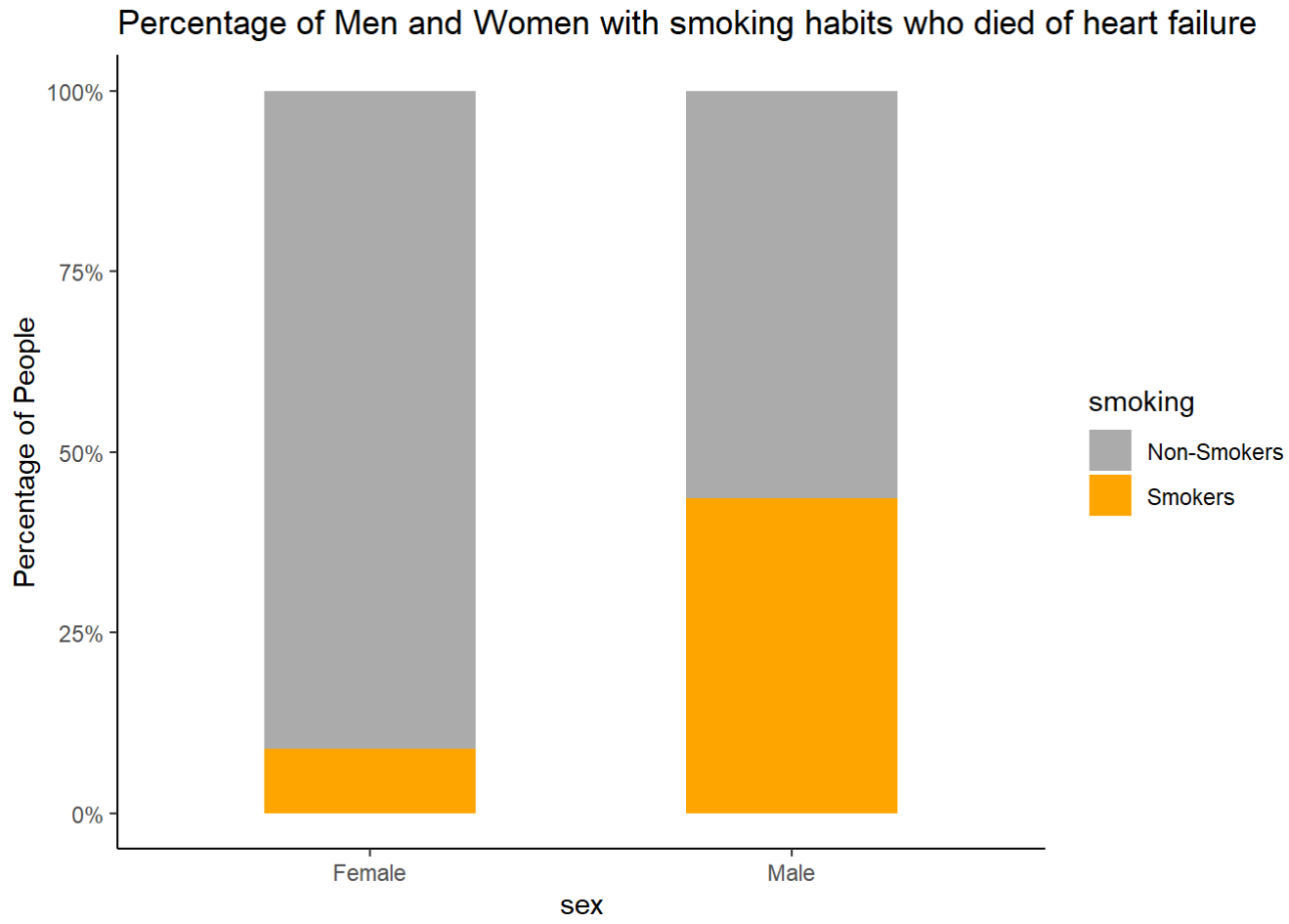
```
library(ggplot2)
dead_ppl$anaemia <- factor(dead_ppl$anaemia, labels = c("Non anaemic", "Anaemic"))
dead_ppl$sex <- factor(dead_ppl$sex, labels = c("Female","Male"))
dead_ppl$high_blood_pressure <- factor(dead_ppl$high_blood_pressure, labels = c("Not High BP"
,"High BP"))
dead_ppl$smoking <- factor(dead_ppl$smoking,labels = c("Non-Smokers","Smokers"))
dead_ppl$diabetes <- factor(dead_ppl$diabetes,labels = c("Non-Diabetic","Diabetic"))

anaemia_gender<-ggplot(data=dead_ppl, aes(x = sex, fill=anaemia)) +
  geom_bar(stat = "count",position = "fill",width = 0.5) + theme_classic() + labs(y = "Percen
tage of People") + ggtitle("Percentage of anaemic Men and Women who died of heart failure ")
anaemia_gender <- anaemia_gender+ scale_fill_manual(values = c("#FF0000","#ABABAB")) + scale_
y_continuous(labels = scales::percent)
anaemia_gender
```

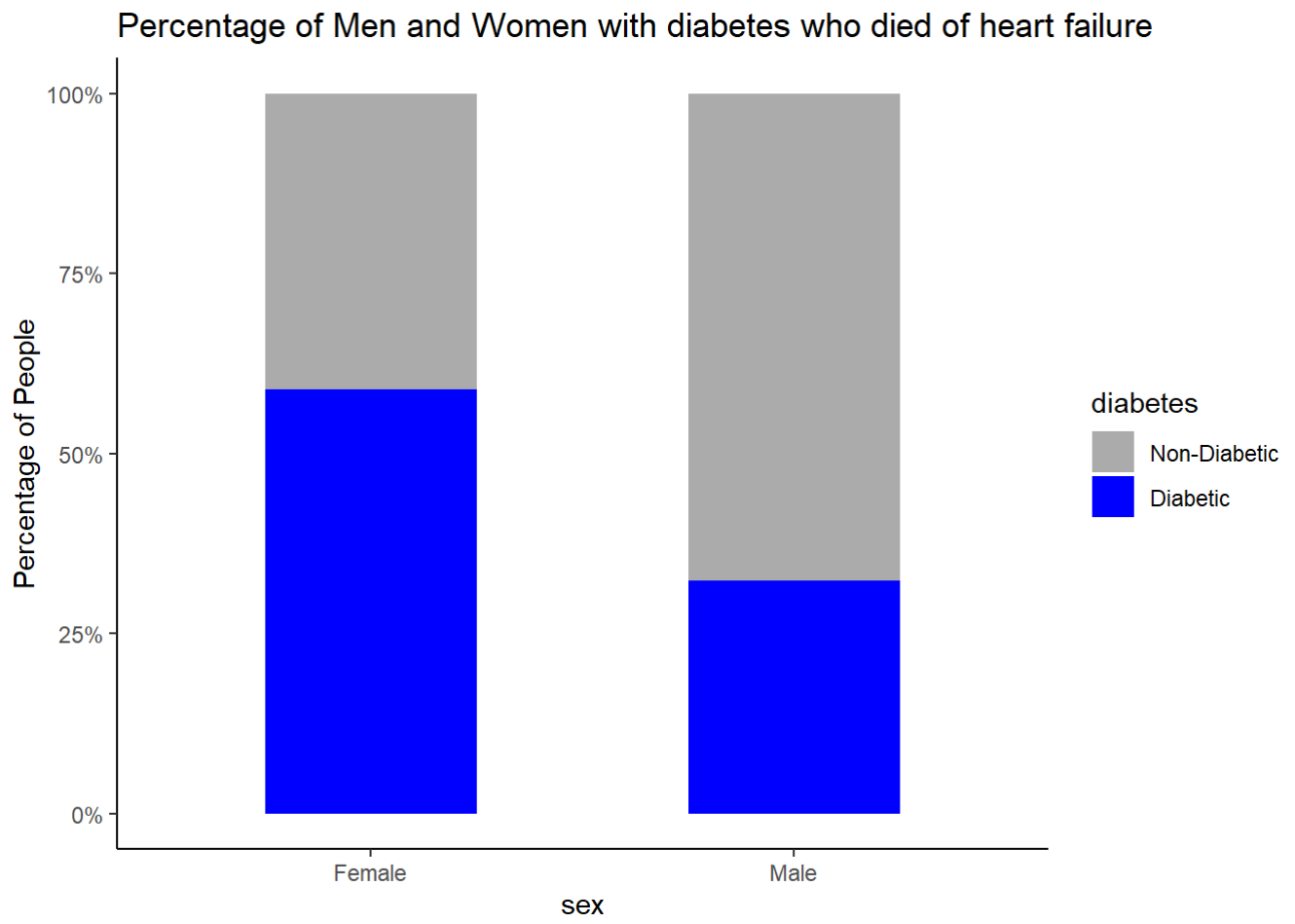



```
highbp_gender <- ggplot(data=dead_ppl, aes(x = sex, fill = high_blood_pressure)) +
    geom_bar(stat = "count",position = "fill",width = 0.5) + theme_classic() +  labs(y = "Per
centage of People") + ggtitle("Percentage of Men and Women with High blood pressure who died
 of heart failure ")
highbp_gender <- highbp_gender+ scale_fill_manual(values = c("#0000FF","#FFFF00")) + scale_y_
continuous(labels = scales::percent)
highbp_gender
```


```
smoking_gender <- ggplot(data=dead_ppl, aes(x = sex, fill = smoking)) +
  geom_bar(stat = "count",position = "fill",width = 0.5) + theme_classic() +  labs(y = "Perce
ntage of People") + ggtitle("Percentage of Men and Women with smoking habits who died of hear
t failure ")
smoking_gender <- smoking_gender+ scale_fill_manual(values = c("#ABABAB","#FFA500")) + scale_
y_continuous(labels = scales::percent)
smoking_gender
```




```
diabetes_gender <- ggplot(data=dead_ppl, aes(x = sex, fill = diabetes)) +
  geom_bar(stat = "count",position = "fill",width = 0.5) + theme_classic() +  labs(y = "Perce
ntage of People") + ggtitle("Percentage of Men and Women with diabetes who died of heart fail
ure ")
diabetes_gender <- diabetes_gender+ scale_fill_manual(values = c("#ABABAB","#0000FF")) + scal
e_y_continuous(labels = scales::percent)
diabetes_gender
```



RESULTS AND DISCUSSION

The initial analysis of the dataset taken, revealed that there were 99 mortalities out of the 299 patients considered. Diabetes,Blood Pressure level,Anaemia and the habit of smoking are found to be major factors impacting fatality rate in patients with an history of heart failure.Upon segregating the patients into various age groups, it was evident that the patients belonging to the age group of 55-60 years old were the most vulnerable to an event of death when compared to all other age groups. Further investigation suggested that increased Serum Creatinine level leads to increased mortality rate. To be more precise, patients with an average Serum Creatinine of 1.836 mg/dl faced an event of death while patients with an average Serum Creatinine of 1.185 mg/dl did not, while the normal range of average Serum Creatinine in the blood may be (0.84 – 1.21) mg/dl. Patients with an history of heart failure are more vulnerable to mortality if the average ejection fraction is 33.47% when compared to patients with average ejection fraction of 40.27%, while the normal range of ejection fraction is between 50% and 70%. Though the average blood platelets count of both terminal and non-terminal patients are 256,381 and 266,658 respectively, but it does not have statistically significant impact on the mortality rate as the average counts for both the type of patients fall under normal range, 150,000 to 350,000.

On further analysis of factors affecting mortality rate, it was discovered that an history of anaemia has a significant impact on the mortality rate of patients as the percentage of death is 52.08% in the case of anaemic patients while the percentage of death is 47.92% for non-anaemic patients. It is also discovered that smokers are 68.75% more likely to face mortality when compared to non-smokers with 31.25% mortality rate. Diabetes plays a major role in factors affecting the mortality rate in patients with heart failure as the percentage of diabetic patients facing mortality is 58.33% while the percentage of non-diabetic patients is 41.67%. On analysing the mortality rate distribution in genders, women are more vulnerable than men as the percentage of death in women is 64.58% whereas the percentage of death in men is 35.42%. Moreover, blood pressure plays a considerable role in the mortality rate as patients with higher blood pressure levels are 59.38% more likely to be terminal while the percentage value for patients with lower blood pressure is 40.62%.

On further examination of factors impacting mortality rate in the two common genders, it was discovered that there was significant impact of blood pressure levels in males with 64.52% likelihood to be terminal while the proportion is equally distributed in females. On factoring the habit of smoking, there is highly significant difference in proportion of patients facing mortality in female population as non-smokers have a higher mortality rate of 91.18% whereas the distribution is similar with non-smokers comprising 56.45% of the total male population. Diabetes tends to play a significant impact on mortality rate of both the genders as female population comprises of 58.82% of diabetic patients while male diabetic patients constitute only 32.25% of the total male population. Furthermore, it can be inferred that female anaemic patients with an history of heart failure constitute 58.82% of the female population while the population of male patients with heart failure comprises of 41.94% of anaemic patients.

CONCLUSION

On initial analysis, it is evident that there are various distinct factors impacting mortality in patients with history of heart failure. We can infer that age, platelet count, serum creatinine, ejection fraction, diabetes, anaemia, blood pressure levels and habit of smoking have significant relationships with number of heart failure patients being terminal.

REFERENCES

1] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics And Decision Making, 20(1). doi: 10.1186/s12911-020-1023-5

2] Heart Failure Prediction. (2020). Retrieved 27 September 2020, from https://www.kaggle.com/andrewmvd/heart-failure-clinical-data (https://www.kaggle.com/andrewmvd/heart-failure-clinical-data)