Assignment-3.R

nidhi

2020-10-01

```
#Installing and Loading Packages
#install.packages(c("tidyverse", "ggplot2", "ggthemes", "RColorBrewer",
"gridExtra", "kableExtra", "data.table", "dplyr", "corrplot"))
library(tidyverse)

## -- Attaching packages ------------------------------------ tidyverse
1.3.0 --

## √ ggplot2 3.3.2     √ purrr   0.3.4
## √ tibble  3.0.3     √ dplyr   1.0.2
## √ tidyr   1.1.2     √ stringr 1.4.0
## √ readr   1.3.1     √ forcats 0.5.0

## -- Conflicts --------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(ggplot2)
library(ggthemes)
library(RColorBrewer)
library(gridExtra)

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows

library(data.table)

##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose

library(dplyr)
library(corrplot)

## corrplot 0.84 loaded

#Loading dataset
rawdata <-
read.csv("C:/Users/nidhi/OneDrive/Desktop/MVA/heart_failure_clinical_records_
dataset.csv")
View(rawdata)

#Identifying different columns names
names(rawdata)

##  [1] "age"                       "anaemia"
##  [3] "creatinine_phosphokinase" "diabetes"
##  [5] "ejection_fraction"         "high_blood_pressure"
##  [7] "platelets"                 "serum_creatinine"
##  [9] "serum_sodium"              "sex"
## [11] "smoking"                   "time"
## [13] "DEATH_EVENT"

#Data Summary
str(rawdata)

## 'data.frame':    299 obs. of  13 variables:
##  $ age                      : num  75 55 65 50 65 90 75 60 65 80 ...
##  $ anaemia                  : int  0 0 0 1 1 1 1 1 0 1 ...
##  $ creatinine_phosphokinase: int  582 7861 146 111 160 47 246 315 157 123
...
##  $ diabetes                 : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ ejection_fraction        : int  20 38 20 20 20 40 15 60 65 35 ...
##  $ high_blood_pressure      : int  1 0 0 0 0 1 0 0 0 1 ...
##  $ platelets                : num  265000 263358 162000 210000 327000 ...
##  $ serum_creatinine         : num  1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4
...
##  $ serum_sodium             : int  130 136 129 137 116 132 137 131 138 133
...
##  $ sex                      : chr  "male" "male" "male" "male" ...
##  $ smoking                  : int  0 0 1 0 0 1 0 1 0 1 ...
##  $ time                     : int  4 6 7 7 8 8 10 10 10 10 ...
##  $ DEATH_EVENT              : chr  "No Death" "No Death" "No Death" "No
Death" ...
```

```
summary(rawdata)
```

```
##       age           anaemia       creatinine_phosphokinase    diabetes
##  Min.   :40.00   Min.   :0.0000   Min.   :  23.0           Min.   :0.0000
##  1st Qu.:51.00   1st Qu.:0.0000   1st Qu.: 116.5           1st Qu.:0.0000
##  Median :60.00   Median :0.0000   Median : 250.0           Median :0.0000
##  Mean   :60.83   Mean   :0.4314   Mean   : 581.8           Mean   :0.4181
##  3rd Qu.:70.00   3rd Qu.:1.0000   3rd Qu.: 582.0           3rd Qu.:1.0000
##  Max.   :95.00   Max.   :1.0000   Max.   :7861.0           Max.   :1.0000
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :14.00     Min.   :0.0000      Min.   : 25100   Min.   :0.500
##  1st Qu.:30.00     1st Qu.:0.0000      1st Qu.:212500   1st Qu.:0.900
##  Median :38.00     Median :0.0000      Median :262000   Median :1.100
##  Mean   :38.08     Mean   :0.3512      Mean   :263358   Mean   :1.394
##  3rd Qu.:45.00     3rd Qu.:1.0000      3rd Qu.:303500   3rd Qu.:1.400
##  Max.   :80.00     Max.   :1.0000      Max.   :850000   Max.   :9.400
##   serum_sodium       sex                smoking            time
##  Min.   :113.0   Length:299         Min.   :0.0000   Min.   :  4.0
##  1st Qu.:134.0   Class :character   1st Qu.:0.0000   1st Qu.: 73.0
##  Median :137.0   Mode  :character   Median :0.0000   Median :115.0
##  Mean   :136.6                      Mean   :0.3211   Mean   :130.3
##  3rd Qu.:140.0                      3rd Qu.:1.0000   3rd Qu.:203.0
##  Max.   :148.0                      Max.   :1.0000   Max.   :285.0
##  DEATH_EVENT
##  Length:299
##  Class :character
##  Mode  :character
##
##
##
```

```
head(rawdata)
```

```
##   age anaemia creatinine_phosphokinase diabetes ejection_fraction
## 1  75       0                     582        0                20
## 2  55       0                    7861        0                38
## 3  65       0                     146        0                20
## 4  50       1                     111        0                20
## 5  65       1                     160        1                20
## 6  90       1                      47        0                40
##   high_blood_pressure platelets serum_creatinine serum_sodium    sex
smoking
## 1                   1    265000             1.9          130   male
0
## 2                   0    263358             1.1          136   male
0
## 3                   0    162000             1.3          129   male
1
## 4                   0    210000             1.9          137   male
0
```

```
## 5                   0    327000              2.7          116 Female
0
## 6                   1    204000              2.1          132   male
1
##   time DEATH_EVENT
## 1    4    No Death
## 2    6    No Death
## 3    7    No Death
## 4    7    No Death
## 5    8    No Death
## 6    8    No Death
```

```
dim(rawdata)
```

```
## [1] 299  13
```

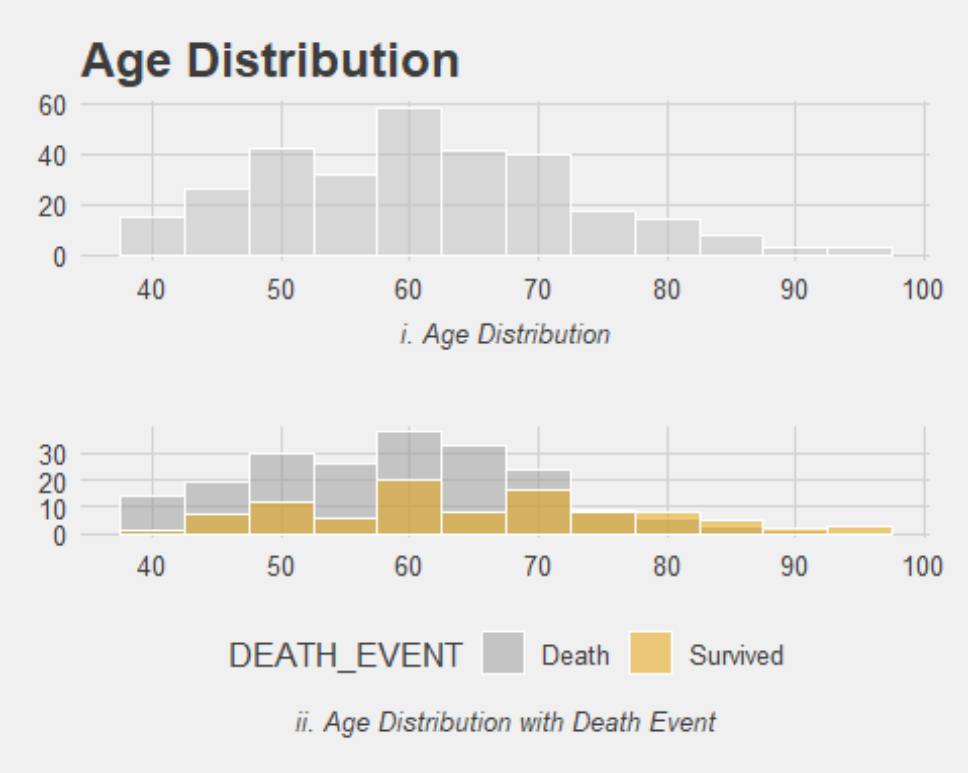
```
#Data Cleaning

#Checking for missing values
is.null(rawdata)
```

```
## [1] FALSE
```

```
##The "FALSE" output shows there is no missing data in the dataset.

#Transforming data (Converting 0,1's to meaningful form)

dataset <- rawdata %>%
  mutate(anaemia = ifelse(anaemia ==1, "Yes", "No"),
         high_blood_pressure = ifelse(high_blood_pressure ==1, "Yes", "No"),
         diabetes = ifelse(diabetes ==1, "Yes", "No"),
         smoking =ifelse(smoking ==1,"Yes","No"),
         DEATH_EVENT=ifelse(DEATH_EVENT=="No Death", "Survived", "Death")
  ) %>%
  mutate_if(is.character, as.factor) %>%
  dplyr::select(age, anaemia, creatinine_phosphokinase, diabetes,
ejection_fraction, high_blood_pressure, platelets,serum_creatinine,
serum_sodium, sex, smoking, time, DEATH_EVENT)

View(dataset)
summary(dataset)
```

```
##       age        anaemia   creatinine_phosphokinase diabetes
ejection_fraction
##  Min.   :40.00   No :170   Min.   :  23.0           No :174   Min.
:14.00
##  1st Qu.:51.00   Yes:129   1st Qu.: 116.5           Yes:125   1st
Qu.:30.00
##  Median :60.00             Median : 250.0                     Median
:38.00
##  Mean   :60.83             Mean   : 581.8                     Mean
```

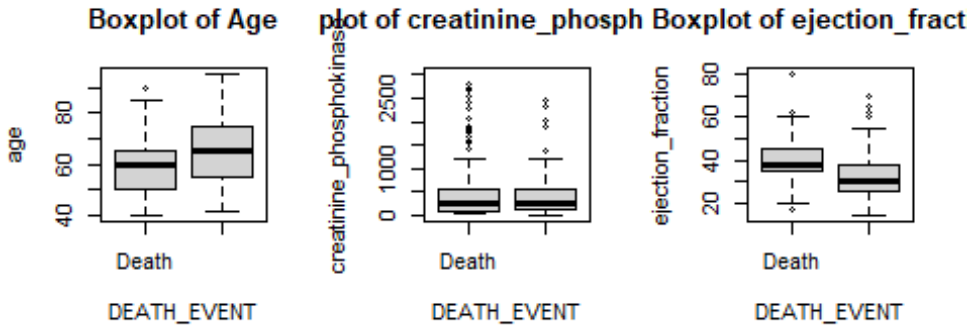
```
:38.08
##  3rd Qu.:70.00             3rd Qu.: 582.0                    3rd
Qu.:45.00
##  Max.   :95.00             Max.   :7861.0                    Max.
:80.00
##  high_blood_pressure   platelets      serum_creatinine  serum_sodium
##  No :194             Min.   : 25100   Min.   :0.500    Min.   :113.0
##  Yes:105             1st Qu.:212500   1st Qu.:0.900    1st Qu.:134.0
##                      Median :262000   Median :1.100    Median :137.0
##                      Mean   :263358   Mean   :1.394    Mean   :136.6
##                      3rd Qu.:303500   3rd Qu.:1.400    3rd Qu.:140.0
##                      Max.   :850000   Max.   :9.400    Max.   :148.0
##      sex      smoking        time         DEATH_EVENT
##  Female:105   No :203   Min.   :  4.0   Death   :203
##  male  :194   Yes: 96   1st Qu.: 73.0   Survived: 96
##                         Median :115.0
##                         Mean   :130.3
##                         3rd Qu.:203.0
##                         Max.   :285.0

#Understanding how Age affects the Death event
a<-ggplot(dataset,aes(x = age))+geom_histogram(binwidth = 5, color = "white",
fill = "grey",alpha = 0.5)+theme_fivethirtyeight()+labs(title = "Age
Distribution", caption = "i. Age Distribution")+
  theme(plot.caption = element_text(hjust = 0.5,face = "italic"))+
  scale_x_continuous(breaks = seq(40,100,10))

b<-ggplot(dataset,aes(x = age, fill = DEATH_EVENT))+geom_histogram(binwidth =
5, position = "identity",alpha = 0.5,color =
"white")+theme_fivethirtyeight()+scale_fill_manual(values = c("#999999",
"#E69F00"))+
  labs(caption = "ii. Age Distribution with Death Event")+
  theme(plot.caption = element_text(hjust = 0.5,face = "italic"))+
  scale_x_continuous(breaks = seq(40,100,10))

gridExtra::grid.arrange(a,b)
```

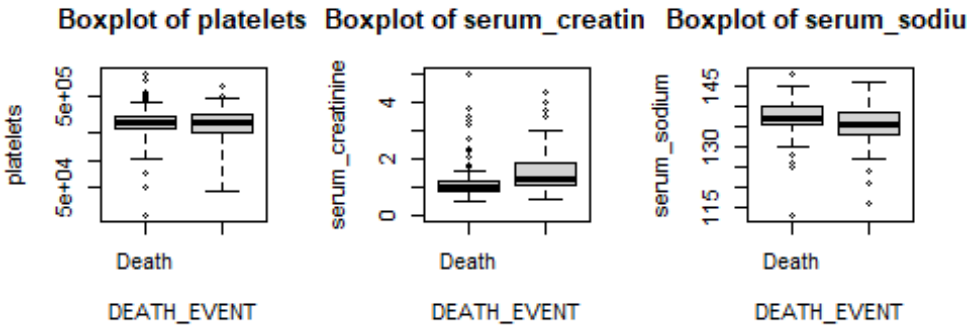


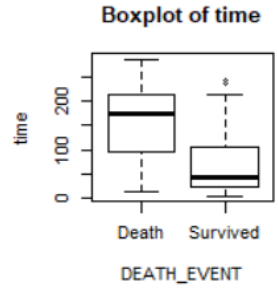
i. Age Distribution

ii. Age Distribution with Death Event

```
##From the age distributions we can see: (1) the age of patients is right-skewed; (2) there are more younger patients dead than survived; (3) there are more elder patients survived than dead.


#Plotting Boxplot to understand relationship of each variable with Death event
attach(dataset)
par(mfrow=c(2,3))
boxplot(age~DEATH_EVENT, main="Boxplot of Age")
boxplot(creatinine_phosphokinase~DEATH_EVENT, main="Boxplot of creatinine_phosphokinase",ylim=c(0,3000))
boxplot(ejection_fraction~DEATH_EVENT, main="Boxplot of ejection_fraction")
boxplot(platelets~DEATH_EVENT, main="Boxplot of platelets", log="y")
boxplot(serum_creatinine~DEATH_EVENT, main="Boxplot of serum_creatinine",ylim=c(0,5))
boxplot(serum_sodium~DEATH_EVENT, main="Boxplot of serum_sodium")
```








```r
boxplot(time~DEATH_EVENT, main="Boxplot of time")

##From the Box plots, we can see:
##Survived patients have a larger age range than dead patients;
##Creatinine Phosphokinase (CPK) has little difference between survived and
dead patients;
##Survived patients have lower Ejection Fraction than dead patients;
##Survived patients have a larger range (with small lower bound) of platelets
than dead patients;
##Survived patients have a larger range (with larger upper bound) of Serum
Creatinine than dead patients;
##Survived patients have a slightly larger range of Serum Sodium than dead
patients;
##Survived patients have shorter follow-up periods than dead patients.
```

```r
#Understanding the correlation between the variables
correlations <- cor(dataset[c(1,3,5,7,8,9,12)])
corrplot(correlations)
correlations
```

```
##                                 age creatinine_phosphokinase
ejection_fraction
## age                         1.00000000             -0.081583900
0.06009836
## creatinine_phosphokinase -0.08158390              1.000000000       -
0.04407955
## ejection_fraction         0.06009836             -0.044079554
1.00000000
## platelets                -0.05235437              0.024463389
0.07217747
## serum_creatinine          0.15918713             -0.016408480       -
0.01130247
## serum_sodium             -0.04596584              0.059550156
0.17590228
## time                     -0.22406842             -0.009345653
0.04172924
##                            platelets serum_creatinine serum_sodium
time
## age                      -0.05235437       0.15918713  -0.04596584 -
0.224068420
## creatinine_phosphokinase  0.02446339      -0.01640848   0.05955016 -
0.009345653
## ejection_fraction         0.07217747      -0.01130247   0.17590228
0.041729235
## platelets                 1.00000000      -0.04119808   0.06212462
0.010513909
## serum_creatinine         -0.04119808       1.00000000  -0.18909521 -
0.149315418
## serum_sodium              0.06212462      -0.18909521   1.00000000
0.087640000
## time                      0.01051391      -0.14931542   0.08764000
1.000000000
```

```r
##From the correlation plot and the table, we can say there exist little/weak
relationship between the numerical variables

#Understanding relationship of other variables (non-numerical) with Death
event
plot_1 <- ggplot(data = dataset, mapping = aes(x = sex, y = ..count.., fill =
DEATH_EVENT)) +
  geom_bar(stat = "count", position='dodge')+
  labs(title = "How gender affects death events?")
plot_1
##There are more male patients than females. The death:survival rate is about
the same (2:1) for male and female.
```

```r
plot_2 <- ggplot(data = dataset, mapping = aes(x = anaemia, y = ..count.., fill = DEATH_EVENT)) +
  geom_bar(stat = "count", position='dodge')+
  labs(title = "Barplot of anaemia")+
  theme_bw()
plot_2
##Patients with a decrease in red blood cell have a higher proportion of survival.

plot_3 <- ggplot(data = dataset, mapping = aes(x = diabetes, y = ..count.., fill = DEATH_EVENT)) +
  geom_bar(stat = "count", position='dodge')+
  labs(title = "Barplot of diabetes")+
  theme_bw()
plot_3
##There are fewer patients with diabetes. The death:survival rate is about the same (2:1) for diabeters and non-diabeters.

plot_4 <- ggplot(data = dataset, mapping = aes(x = high_blood_pressure, y = ..count.., fill = DEATH_EVENT)) +
  geom_bar(stat = "count", position='dodge')+
  labs(title = "Barplot of high_blood_pressure")+
  theme_bw()
plot_4
##There are fewer patients with high blood pressure. Patients with high blood pressure have a higher proportion of survival.

plot_5 <- ggplot(data = dataset, mapping = aes(x = smoking, y = ..count.., fill = DEATH_EVENT)) +
  geom_bar(stat = "count", position='dodge')+
  labs(title = "Barplot of smoking")+
  theme_bw()
plot_5
##There are fewer smoking patients than non-smoking patients. The death:survival rate is about the same (2:1) for smokers and non-smokers.



#T-Test

with(data=dataset,t.test(age[DEATH_EVENT=="Survived"],age[DEATH_EVENT=="Death"],var.equal=TRUE))

##
##  Two Sample t-test
##
## data:  age[DEATH_EVENT == "Survived"] and age[DEATH_EVENT == "Death"]
## t = 4.5206, df = 297, p-value = 8.917e-06
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.643992 9.262758
## sample estimates:
## mean of x mean of y
##  65.21528  58.76191

##p-value is smaller than alpha 0.05. There is a significant difference in
mean age between dead patients and survived patients.

with(data=dataset,t.test(creatinine_phosphokinase[DEATH_EVENT=="Survived"],cr
eatinine_phosphokinase[DEATH_EVENT=="Death"],var.equal=TRUE))

##
##  Two Sample t-test
##
## data:  creatinine_phosphokinase[DEATH_EVENT == "Survived"] and
creatinine_phosphokinase[DEATH_EVENT == "Death"]
## t = 1.0832, df = 297, p-value = 0.2796
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -106.3109  366.5984
## sample estimates:
## mean of x mean of y
##  670.1979  540.0542

##p-value is larger than alpha 0.05. There is no significant difference in
the mean level of CPK enzyme in blood between dead patients and survived
patients.

with(data=dataset,t.test(ejection_fraction[DEATH_EVENT=="Survived"],ejection_
fraction[DEATH_EVENT=="Death"],var.equal=TRUE))

##
##  Two Sample t-test
##
## data:  ejection_fraction[DEATH_EVENT == "Survived"] and
ejection_fraction[DEATH_EVENT == "Death"]
## t = -4.8056, df = 297, p-value = 2.453e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.580849 -4.013671
## sample estimates:
## mean of x mean of y
##  33.46875  40.26601

##p-value is smaller than alpha 0.05. There is a significant difference in
the mean ejection fraction between dead patients and survived patients.

with(data=dataset,t.test(platelets[DEATH_EVENT=="Survived"],platelets[DEATH_E
VENT=="Death"],var.equal=TRUE))
```

```
##
##  Two Sample t-test
##
## data:  platelets[DEATH_EVENT == "Survived"] and platelets[DEATH_EVENT ==
"Death"]
## t = -0.84787, df = 297, p-value = 0.3972
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -34129.06  13576.17
## sample estimates:
## mean of x mean of y
##  256381.0  266657.5
```

```
##p-value is larger than alpha 0.05. There is no significant difference in
mean platelets between dead patients and survived patients.
```

```
with(data=dataset,t.test(serum_creatinine[DEATH_EVENT=="Survived"],serum_crea
tinine[DEATH_EVENT=="Death"],var.equal=TRUE))
```

```
##
##  Two Sample t-test
##
## data:  serum_creatinine[DEATH_EVENT == "Survived"] and
serum_creatinine[DEATH_EVENT == "Death"]
## t = 5.3065, df = 297, p-value = 2.19e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.409539 0.892374
## sample estimates:
## mean of x mean of y
##  1.835833  1.184877
```

```
##p-value is smaller than alpha 0.05. There is a significant difference in
the mean level of Serum Creatinine between dead patients and survived
patients.
```

```
with(data=dataset,t.test(serum_sodium[DEATH_EVENT=="Survived"],serum_sodium[D
EATH_EVENT=="Death"],var.equal=TRUE))
```

```
##
##  Two Sample t-test
##
## data:  serum_sodium[DEATH_EVENT == "Survived"] and
serum_sodium[DEATH_EVENT == "Death"]
## t = -3.4301, df = 297, p-value = 0.0006889
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.8984440 -0.7850535
## sample estimates:
## mean of x mean of y
##  135.3750  137.2167
```

```r
##p-value is smaller than alpha 0.05. There is a significant difference in
the mean level of Serum Sodium between dead patients and survived patients.

with(data=dataset,t.test(time[DEATH_EVENT=="Survived"],time[DEATH_EVENT=="Dea
th"],var.equal=TRUE))
```

```
##
##  Two Sample t-test
##
## data:  time[DEATH_EVENT == "Survived"] and time[DEATH_EVENT == "Death"]
## t = -10.686, df = 297, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -103.5612  -71.3478
## sample estimates:
## mean of x mean of y
##  70.88542 158.33990
```

```r
##p-value is smaller than alpha 0.05. There is a significant difference in
the mean follow-up period between dead patients and survived patients.


#Hotelling's T2 test

#install.packages("Hotelling")
library(Hotelling)
```

```
## Loading required package: corpcor
```

```r
T2Test <- hotelling.test(age + creatinine_phosphokinase + ejection_fraction +
platelets + serum_creatinine + serum_sodium + time ~ DEATH_EVENT,
data=dataset)
T2Test
```

```
## Test stat:  29.086
## Numerator df:  7
## Denominator df:  291
## P-value:  0
```

```r
##p-value is smaller than alpha 0.05. The mean of at least one of the
numerical parameters (age, CPK, ejection fraction, serum creatinine, serum
sodium, time), or a combination of one or more parameters working together,
is significantly different between dead patients and survived patients.


#F-Test

var.test(age[DEATH_EVENT=="Survived"],age[DEATH_EVENT=="Death"])
```

```
##
##  F test to compare two variances
```

```
##
## data:  age[DEATH_EVENT == "Survived"] and age[DEATH_EVENT == "Death"]
## F = 1.5431, num df = 95, denom df = 202, p-value = 0.01112
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.103220 2.206112
## sample estimates:
## ratio of variances
##             1.5431
```

##p-value is smaller than alpha 0.05. There is a significant difference in variance of age between dead patients and survived patients.

```
var.test(creatinine_phosphokinase[DEATH_EVENT=="Survived"],creatinine_phospho
kinase[DEATH_EVENT=="Death"])
```

```
##
##  F test to compare two variances
##
## data:  creatinine_phosphokinase[DEATH_EVENT == "Survived"] and
creatinine_phosphokinase[DEATH_EVENT == "Death"]
## F = 3.0506, num df = 95, denom df = 202, p-value = 3.354e-11
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  2.180978 4.361306
## sample estimates:
## ratio of variances
##           3.050585
```

##p-value is smaller than alpha 0.05. There is a significant difference in variance of CPK level between dead patients and survived patients.

```
var.test(ejection_fraction[DEATH_EVENT=="Survived"],ejection_fraction[DEATH_E
VENT=="Death"])
```

```
##
##  F test to compare two variances
##
## data:  ejection_fraction[DEATH_EVENT == "Survived"] and
ejection_fraction[DEATH_EVENT == "Death"]
## F = 1.3302, num df = 95, denom df = 202, p-value = 0.09577
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.9510164 1.9017493
## sample estimates:
## ratio of variances
##           1.330209
```

##p-value is larger than alpha 0.05. There is no significant difference in variance of ejection fraction between dead patients and survived patients.

```r
var.test(platelets[DEATH_EVENT=="Survived"],platelets[DEATH_EVENT=="Death"])
```

```
##
##  F test to compare two variances
##
## data:  platelets[DEATH_EVENT == "Survived"] and platelets[DEATH_EVENT ==
"Death"]
## F = 1.0205, num df = 95, denom df = 202, p-value = 0.8915
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7295918 1.4589660
## sample estimates:
## ratio of variances
##           1.020497
```

```r
##p-value is larger than alpha 0.05. There is no significant difference in
variance of platelets between dead patients and survived patients.
```

```r
var.test(serum_creatinine[DEATH_EVENT=="Survived"],serum_creatinine[DEATH_EVE
NT=="Death"])
```

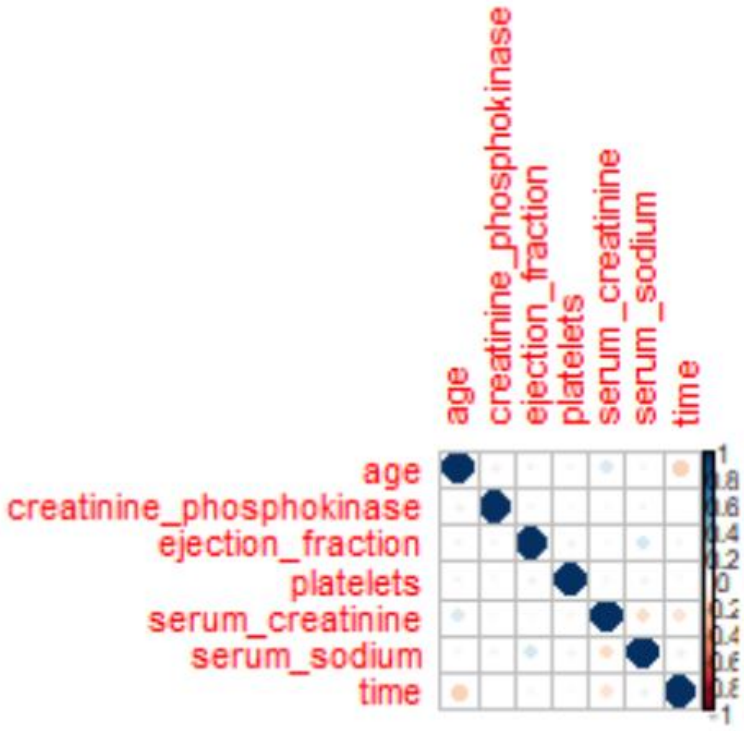
```
##
##  F test to compare two variances
##
## data:  serum_creatinine[DEATH_EVENT == "Survived"] and
serum_creatinine[DEATH_EVENT == "Death"]
## F = 5.041, num df = 95, denom df = 202, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  3.604020 7.206966
## sample estimates:
## ratio of variances
##           5.041027
```

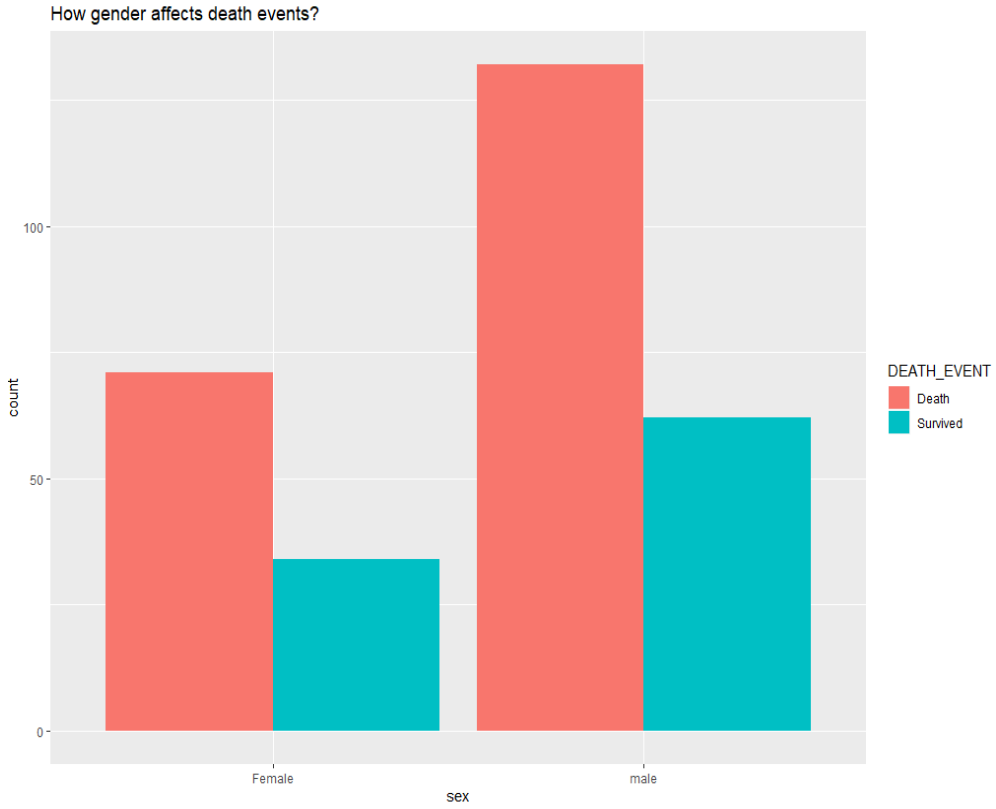
```r
##p-value is smaller than alpha 0.05. There is a significant difference in
variance of the level of Serum Creatinine between dead patients and survived
patients.
```

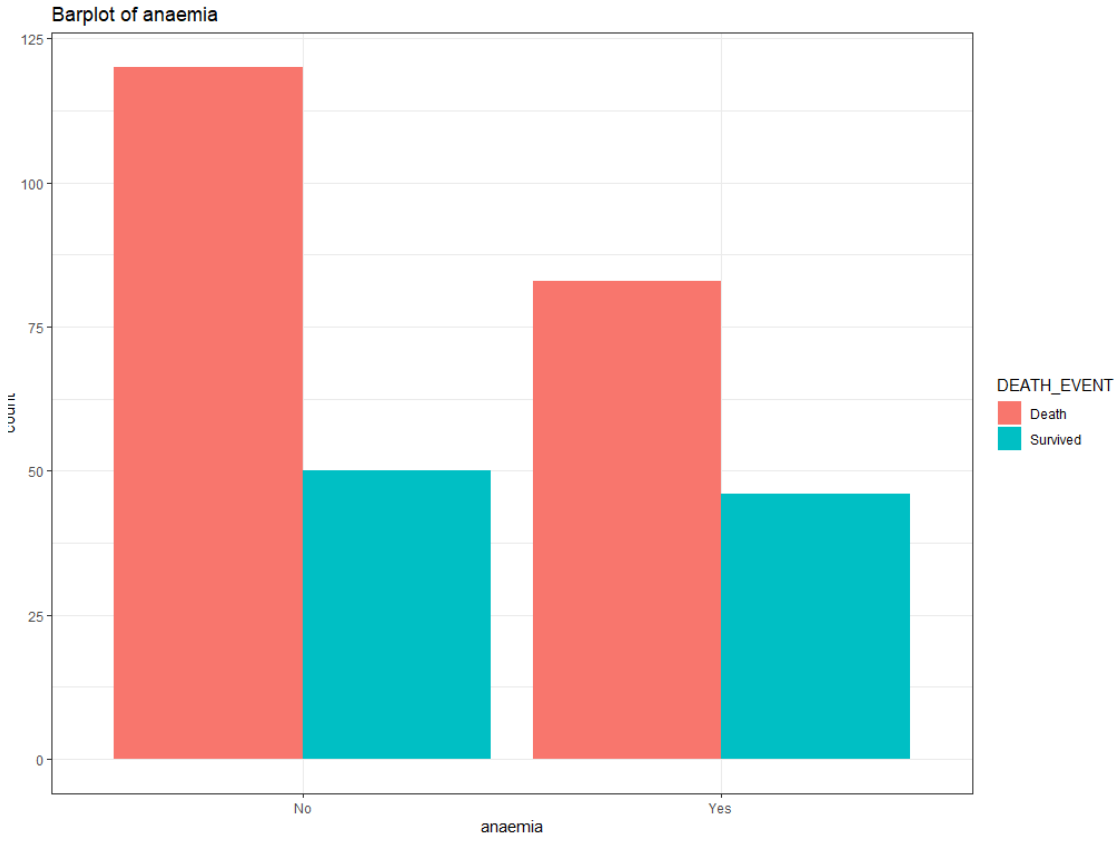
```r
var.test(serum_sodium[DEATH_EVENT=="Survived"],serum_sodium[DEATH_EVENT=="Dea
th"])
```

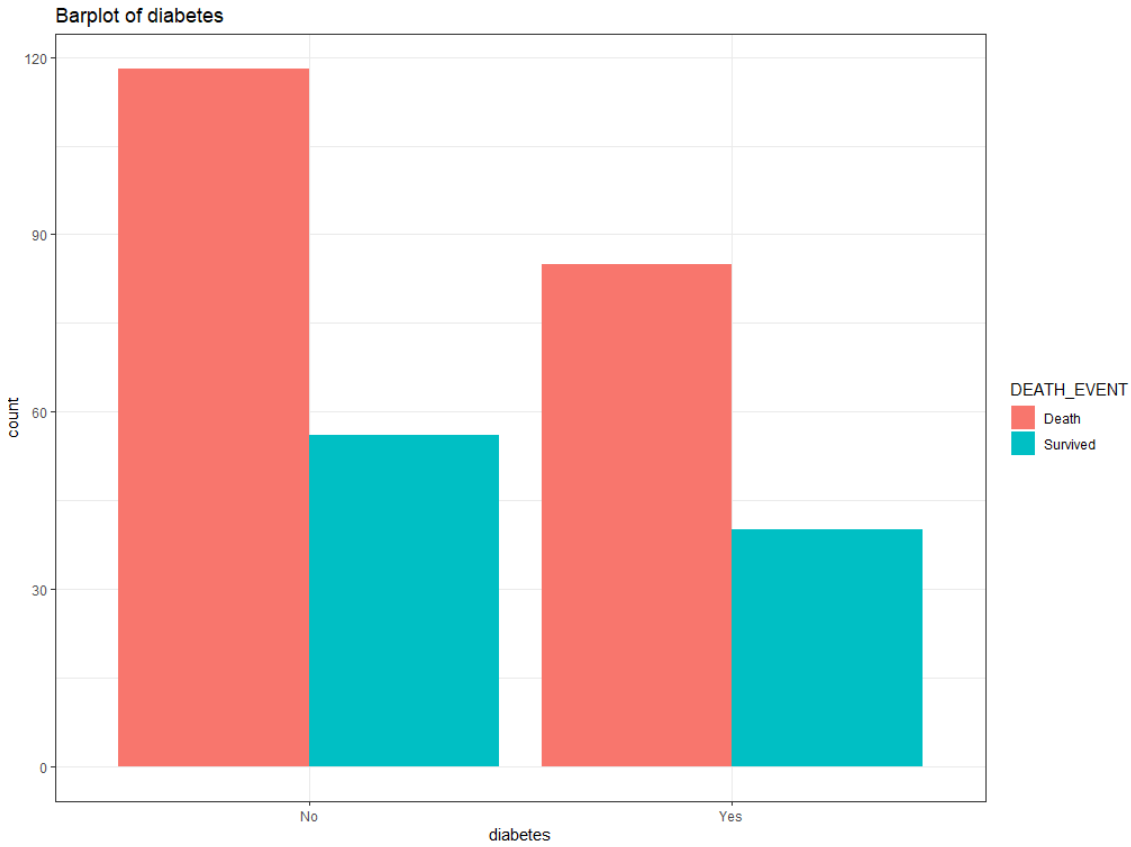
```
##
##  F test to compare two variances
##
## data:  serum_sodium[DEATH_EVENT == "Survived"] and
serum_sodium[DEATH_EVENT == "Death"]
## F = 1.5769, num df = 95, denom df = 202, p-value = 0.007646
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.127401 2.254466
```

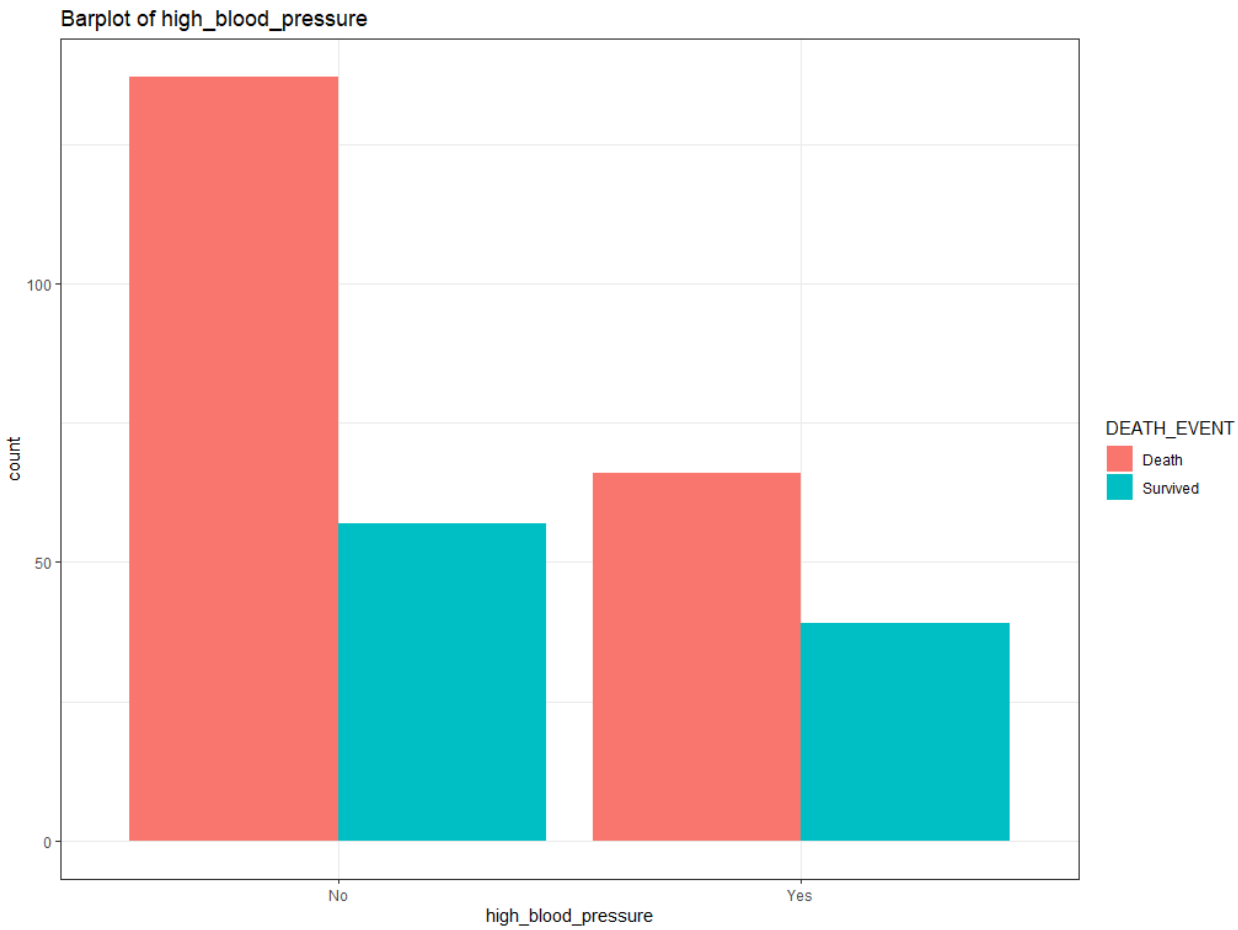
```
## sample estimates:
## ratio of variances
##           1.576922

##p-value is smaller than alpha 0.05. There is a significant difference in
variance of the level of Serum Sodium between dead patients and survived
patients.

var.test(time[DEATH_EVENT=="Survived"],time[DEATH_EVENT=="Death"])

##
##  F test to compare two variances
##
## data:  time[DEATH_EVENT == "Survived"] and time[DEATH_EVENT == "Death"]
## F = 0.84789, num df = 95, denom df = 202, p-value = 0.3652
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6061886 1.2121964
## sample estimates:
## ratio of variances
##          0.8478901

##p-value is larger than alpha 0.05. There is no significant difference in
variance of the follow-up period between dead patients and survived patients.
```

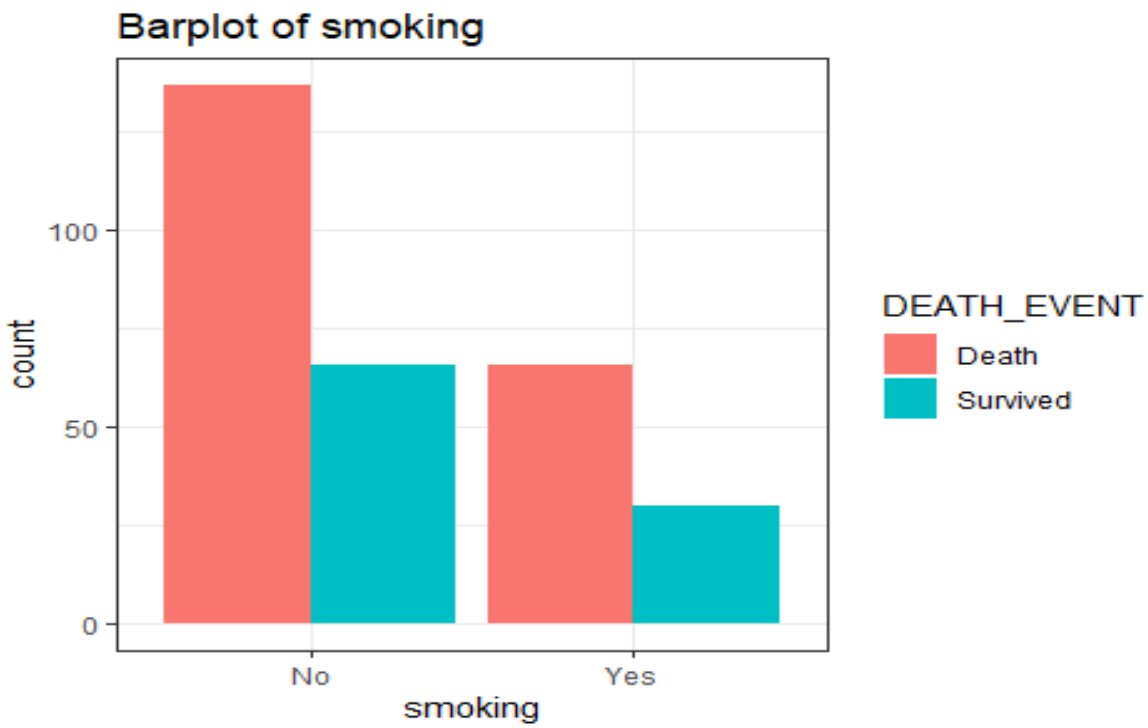



How gender affects death events?
count
100
50
0
Female
male
sex
DEATH_EVENT
Death
Survived


Barplot of anaemia
125
100
75
50
25
0
No
Yes
anaemia
DEATH_EVENT
Death
Survived


Barplot of diabetes
120
90
60
30
0
count
No
Yes
diabetes
DEATH_EVENT
Death
Survived


Barplot of high_blood_pressure
count
100
50
0
No
Yes
high_blood_pressure
DEATH_EVENT
Death
Survived


Barplot of smoking
count
100
50
0
No
Yes
smoking
DEATH_EVENT
Death
Survived