STAT515_FINAL

2023-05-06

```
library("ggplot2")
library(ISLR)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.2.1     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## v purrr   1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("reshape2")
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(class )
library(dplyr)
```

```
heart <- read.csv("/Users/mykola/Desktop/STAT515/stat515_final/heart_failure_clinical_records_dataset.cs
head(heart)
```

```
##   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 time
## 1                   1    265000             1.9          130   1       0    4
## 2                   0    263358             1.1          136   1       0    6
## 3                   0    162000             1.3          129   1       1    7
## 4                   0    210000             1.9          137   1       0    7
## 5                   0    327000             2.7          116   0       0    8
## 6                   1    204000             2.1          132   1       1    8
##   DEATH_EVENT
## 1           1
## 2           1
## 3           1
## 4           1
```

```
## 5           1
## 6           1
```

```
any(is.na(heart))
```

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

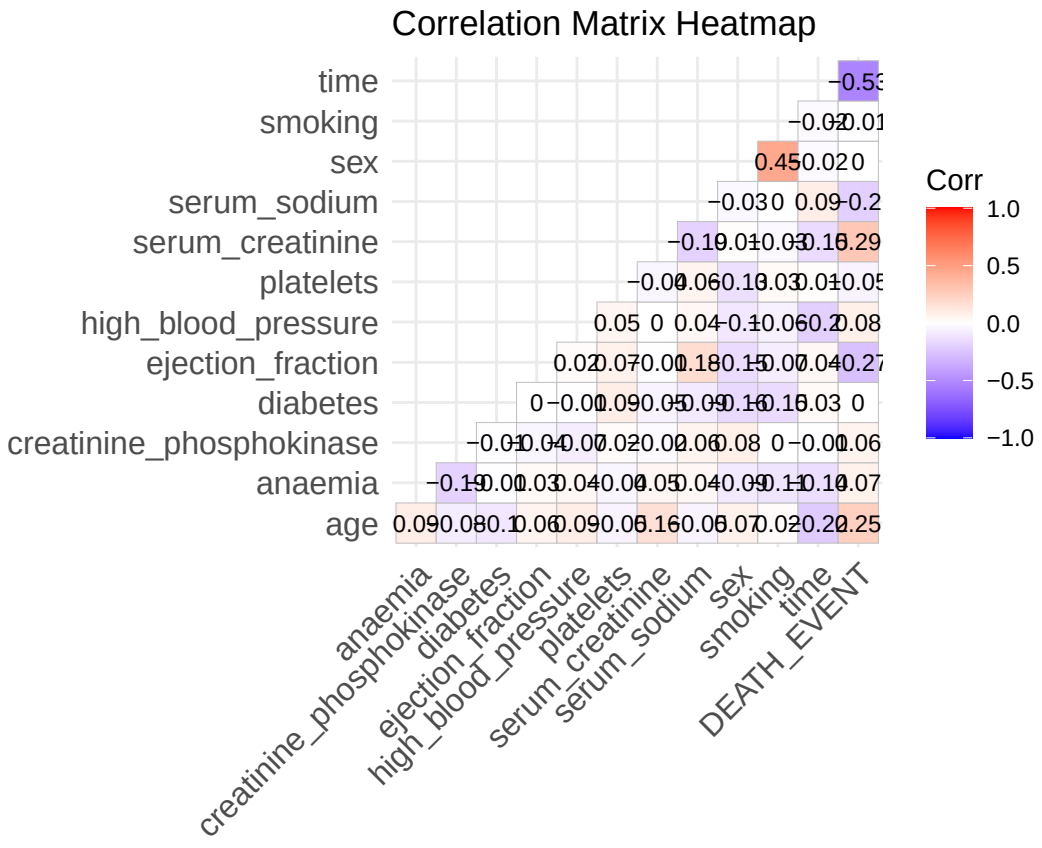
```
cor_mat <- cor(heart)
print(cor_mat)
```

```
##                                 age     anaemia creatinine_phosphokinase
## age                      1.00000000  0.08800644             -0.081583900
## anaemia                  0.08800644  1.00000000             -0.190741030
## creatinine_phosphokinase -0.08158390 -0.19074103              1.000000000
## diabetes                 -0.10101239 -0.01272905             -0.009638514
## ejection_fraction         0.06009836  0.03155697             -0.044079554
## high_blood_pressure       0.09328868  0.03818200             -0.070589980
## platelets                -0.05235437 -0.04378555              0.024463389
## serum_creatinine          0.15918713  0.05217360             -0.016408480
## serum_sodium             -0.04596584  0.04188161              0.059550156
## sex                       0.06542952 -0.09476896              0.079790629
## smoking                   0.01866787 -0.10728984              0.002421235
## time                     -0.22406842 -0.14141398             -0.009345653
## DEATH_EVENT               0.25372854  0.06627010              0.062728160
##                             diabetes ejection_fraction high_blood_pressure
## age                      -0.101012385        0.06009836         0.093288685
## anaemia                  -0.012729046        0.03155697         0.038182003
## creatinine_phosphokinase -0.009638514       -0.04407955        -0.070589980
## diabetes                  1.000000000       -0.00485031        -0.012732382
## ejection_fraction        -0.004850310        1.00000000         0.024444731
## high_blood_pressure      -0.012732382        0.02444473         1.000000000
## platelets                 0.092192828        0.07217747         0.049963481
## serum_creatinine         -0.046975315       -0.01130247        -0.004934525
## serum_sodium             -0.089550619        0.17590228         0.037109470
## sex                      -0.157729504       -0.14838597        -0.104614629
## smoking                  -0.147173413       -0.06731457        -0.055711369
## time                      0.033725509        0.04172924        -0.196439479
## DEATH_EVENT              -0.001942883       -0.26860331         0.079351058
##                            platelets serum_creatinine serum_sodium          sex
## age                      -0.05235437      0.159187133 -0.045965841  0.065429524
## anaemia                  -0.04378555      0.052173604  0.041881610 -0.094768961
## creatinine_phosphokinase  0.02446339     -0.016408480  0.059550156  0.079790629
## diabetes                  0.09219283     -0.046975315 -0.089550619 -0.157729504
## ejection_fraction         0.07217747     -0.011302475  0.175902282 -0.148385965
## high_blood_pressure       0.04996348     -0.004934525  0.037109470 -0.104614629
## platelets                 1.00000000     -0.041198077  0.062124619 -0.125120483
## serum_creatinine         -0.04119808      1.000000000 -0.189095210  0.006969778
## serum_sodium              0.06212462     -0.189095210  1.000000000 -0.027566123
## sex                      -0.12512048      0.006969778 -0.027566123  1.000000000
## smoking                   0.02823445     -0.027414135  0.004813195  0.445891712
## time                      0.01051391     -0.149315418  0.087640000 -0.015608220
## DEATH_EVENT              -0.04913887      0.294277561 -0.195203596 -0.004316376
##                              smoking         time  DEATH_EVENT
## age                       0.018667868 -0.224068420  0.253728543
## anaemia                  -0.107289838 -0.141413982  0.066270098
```

```
## creatinine_phosphokinase  0.002421235 -0.009345653  0.062728160
## diabetes                 -0.147173413  0.033725509 -0.001942883
## ejection_fraction        -0.067314567  0.041729235 -0.268603312
## high_blood_pressure      -0.055711369 -0.196439479  0.079351058
## platelets                 0.028234448  0.010513909 -0.049138868
## serum_creatinine         -0.027414135 -0.149315418  0.294277561
## serum_sodium              0.004813195  0.087640000 -0.195203596
## sex                       0.445891712 -0.015608220 -0.004316376
## smoking                   1.000000000 -0.022838942 -0.012623153
## time                     -0.022838942  1.000000000 -0.526963779
## DEATH_EVENT              -0.012623153 -0.526963779  1.000000000
```

```
library(ggcorrplot)

ggcorrplot(cor_mat,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap")
```




```
summary(heart)
```

```
##       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   Min.   :0.0000   Min.   :0.0000   Min.   :  4.0
##  1st Qu.:134.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 73.0
##  Median :137.0   Median :1.0000   Median :0.0000   Median :115.0
##  Mean   :136.6   Mean   :0.6488   Mean   :0.3211   Mean   :130.3
##  3rd Qu.:140.0   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:203.0
##  Max.   :148.0   Max.   :1.0000   Max.   :1.0000   Max.   :285.0
##   DEATH_EVENT
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3211
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
str(heart)
```

```
## '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                     : int  1 1 1 1 0 1 1 1 0 1 ...
##  $ 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             : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
# changing into factors

heart$DEATH_EVENT = as.factor(heart$DEATH_EVENT)
heart$smoking = as.factor(heart$smoking)
heart$diabetes = as.factor(heart$diabetes)
heart$high_blood_pressure = as.factor(heart$high_blood_pressure)
heart$sex = as.factor(heart$sex)
heart$anaemia = as.factor(heart$anaemia)
```

```
str(heart)
```

```
## 'data.frame':    299 obs. of  13 variables:
##  $ age                     : num  75 55 65 50 65 90 75 60 65 80 ...
##  $ anaemia                 : Factor w/ 2 levels "0","1": 1 1 1 2 2 2 2 2 1 2 ...
##  $ creatinine_phosphokinase: int  582 7861 146 111 160 47 246 315 157 123 ...
```

```
##  $ diabetes                : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
##  $ ejection_fraction       : int  20 38 20 20 20 40 15 60 65 35 ...
##  $ high_blood_pressure     : Factor w/ 2 levels "0","1": 2 1 1 1 1 2 1 1 1 2 ...
##  $ 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                     : Factor w/ 2 levels "0","1": 2 2 2 2 1 2 2 2 1 2 ...
##  $ smoking                 : Factor w/ 2 levels "0","1": 1 1 2 1 1 2 1 2 1 2 ...
##  $ time                    : int  4 6 7 7 8 8 10 10 10 10 ...
##  $ DEATH_EVENT             : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
set.seed(123)
```

```
dim(heart)
```

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

```
logit_model <- glm(DEATH_EVENT ~ ., data = heart, family = binomial)
```

```
summary(logit_model)
```

```
##
## Call:
## glm(formula = DEATH_EVENT ~ ., family = binomial, data = heart)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1848  -0.5706  -0.2401   0.4466   2.6668
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               1.018e+01  5.657e+00   1.801 0.071774 .
## age                       4.742e-02  1.580e-02   3.001 0.002690 **
## anaemia1                 -7.470e-03  3.605e-01  -0.021 0.983467
## creatinine_phosphokinase  2.222e-04  1.779e-04   1.249 0.211684
## diabetes1                 1.451e-01  3.512e-01   0.413 0.679380
## ejection_fraction        -7.666e-02  1.633e-02  -4.695 2.67e-06 ***
## high_blood_pressure1     -1.027e-01  3.587e-01  -0.286 0.774688
## platelets                -1.200e-06  1.889e-06  -0.635 0.525404
## serum_creatinine          6.661e-01  1.815e-01   3.670 0.000242 ***
## serum_sodium             -6.698e-02  3.974e-02  -1.686 0.091855 .
## sex1                     -5.337e-01  4.139e-01  -1.289 0.197299
## smoking1                 -1.349e-02  4.126e-01  -0.033 0.973915
## time                     -2.104e-02  3.014e-03  -6.981 2.92e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 219.55  on 286  degrees of freedom
## AIC: 245.55
##
## Number of Fisher Scoring iterations: 6
```

```
#confusion matrix
predicted_class <- ifelse(predict(logit_model, type = "response") > 0.5, 1, 0)
confusion_matrix <- table(predicted_class, heart$DEATH_EVENT)
accuracy <- sum(diag(confusion_matrix))/sum(confusion_matrix)
```

```
confusion_matrix
```

```
##
## predicted_class   0   1
##               0 187  27
##               1  16  69
```

```
accuracy
```

```
## [1] 0.8561873
```

```
set.seed(123)
train_index <- sample(nrow(heart), round(0.8 * nrow(heart)))
train <- heart[train_index, ]
test <- heart[-train_index, ]
```

```
model <- glm(DEATH_EVENT ~ ., family = binomial, data = train)
```

```
test$predicted <- ifelse(predict(model, test, type = "response") > 0.5, "Real", "Fake")
```

```
sum(diag(table(test$DEATH_EVENT, test$predicted))) / nrow(test)
```

```
## [1] 0.7666667
```

```
# Creating a list of feature subsets to use
feature_subsets <- list(c("age", "anaemia", "creatinine_phosphokinase", "diabetes", "ejection_fraction"
                        c("age", "anaemia", "creatinine_phosphokinase", "ejection_fraction", "platelets
                        c("age", "serum_creatinine", "serum_sodium"),
                        c("serum_creatinine", "serum_sodium"),
                        c("serum_creatinine", "serum_sodium", "ejection_fraction"),
                        c("serum_creatinine", "ejection_fraction"))
```

```
results <- lapply(feature_subsets, function(features){
  # only include selected features
  data_subset <- heart[, c(features, "DEATH_EVENT")]

  train_control <- trainControl(method = "cv", number = 10)

  # Train
  model <- train(DEATH_EVENT ~ ., method = "glm", family = "binomial", data = data_subset, trControl = t

  # Compute accuracy
  accuracy <- model$results$Accuracy[1]

  # the feature subset and accuracy
```

```
  list(features = features, accuracy = accuracy)
})
```

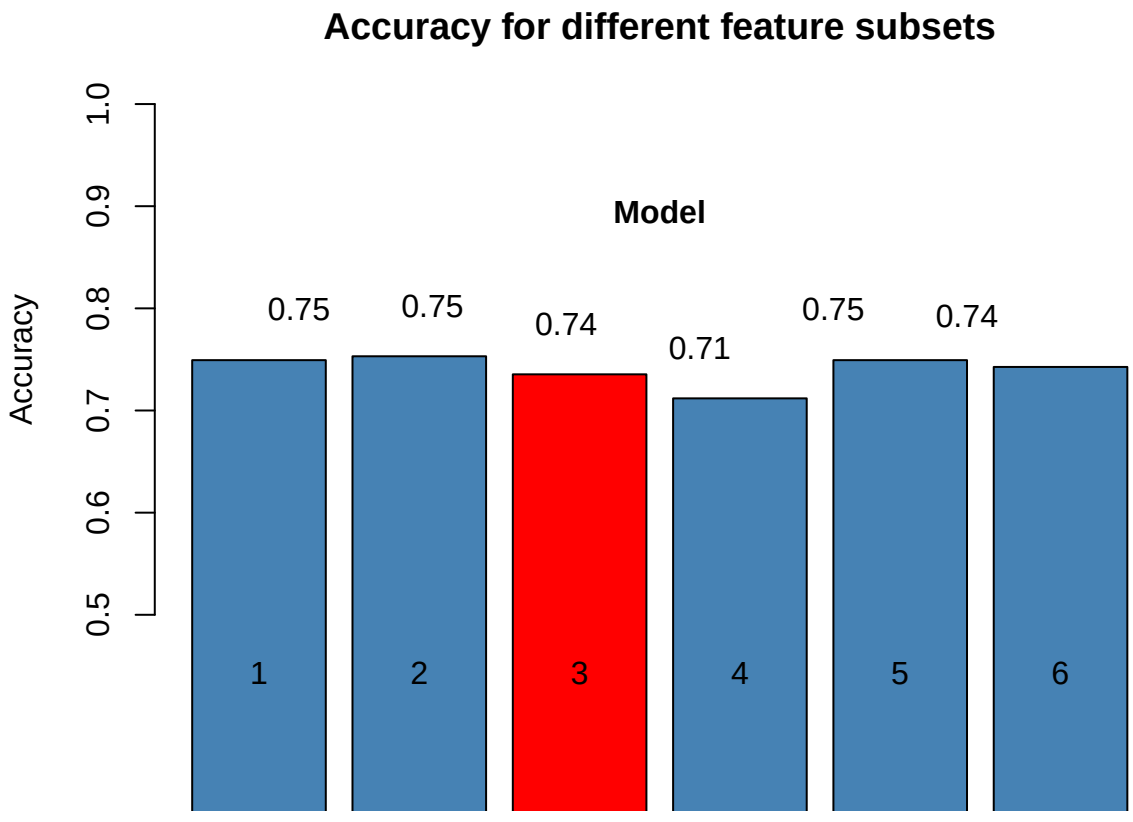
```
# printing accuracy
for (i in seq_along(results)) {
  cat(sprintf("Model %d: Features: %s, Accuracy: %.2f\n", i, paste(results[[i]]$features, collapse = ",
}
```

```
## Model 1: Features: age, anaemia, creatinine_phosphokinase, diabetes, ejection_fraction, high_blood_pr
## Model 2: Features: age, anaemia, creatinine_phosphokinase, ejection_fraction, platelets, serum_creati
## Model 3: Features: age, serum_creatinine, serum_sodium, Accuracy: 0.74
## Model 4: Features: serum_creatinine, serum_sodium, Accuracy: 0.71
## Model 5: Features: serum_creatinine, serum_sodium, ejection_fraction, Accuracy: 0.75
## Model 6: Features: serum_creatinine, ejection_fraction, Accuracy: 0.74
```

```
accuracy_plot <- sapply(results, function(x) x$accuracy)

# create a vector of colors with the same length as the number of models
colors <- rep("steelblue", length(results))
# set the color of the third bar to red
colors[3] <- "red"

barplot(accuracy_plot, names.arg = 1:length(results), ylim = c(0.50, 1), col = colors,
        xlab = "", ylab = "Accuracy", main = "Accuracy for different feature subsets") +
  text(x = 1:length(results), y = accuracy_plot + 0.02, labels = round(accuracy_plot, 2), pos = 3, col =
  title(xlab = expression(bold("Model")), line = -11)
```




```
## numeric(0)
```

```
head(heart)
```

```
##   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 time
## 1                   1    265000             1.9          130   1       0    4
## 2                   0    263358             1.1          136   1       0    6
## 3                   0    162000             1.3          129   1       1    7
## 4                   0    210000             1.9          137   1       0    7
## 5                   0    327000             2.7          116   0       0    8
## 6                   1    204000             2.1          132   1       1    8
##   DEATH_EVENT
## 1           1
## 2           1
## 3           1
## 4           1
## 5           1
## 6           1
```

```
heart_new <- read.csv("/Users/mykola/Desktop/STAT515/stat515_final/heart_failure_clinical_records_datas
sup <- subset(heart_new, DEATH_EVENT==1)
sup$death_interval <- ifelse(sup$DEATH_EVENT == 1 & sup$time <= 35, "Within 5 Week",
                      ifelse(sup$DEATH_EVENT == 1 & sup$time <= 70 & sup$time > 35, "Within 10 Week",
                      ifelse(sup$DEATH_EVENT == 1 & sup$time <= 105 & sup$time > 70, "Within 15 Week",
                      ifelse(sup$DEATH_EVENT == 1 & sup$time <= 140 & sup$time > 105, "Within 20 Week",
                      ifelse(sup$DEATH_EVENT == 1 & sup$time <= 175 & sup$time > 140, "Within 25 Week",
                      ifelse(sup$DEATH_EVENT == 1 & sup$time > 175, "more 25 Week", NA))))))
```

```
head(sup)
```

```
##   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 time
## 1                   1    265000             1.9          130   1       0    4
## 2                   0    263358             1.1          136   1       0    6
## 3                   0    162000             1.3          129   1       1    7
## 4                   0    210000             1.9          137   1       0    7
## 5                   0    327000             2.7          116   0       0    8
## 6                   1    204000             2.1          132   1       1    8
##   DEATH_EVENT death_interval
## 1           1  Within 5 Week
## 2           1  Within 5 Week
## 3           1  Within 5 Week
## 4           1  Within 5 Week
## 5           1  Within 5 Week
```

```
## 6           1  Within 5 Week
```

```
# Split the original data into separate datasets based on death_interval
df_list <- split(sup, sup$death_interval)

# Rename the list elements to match the death_interval values
names(df_list) <- paste0("death_", names(df_list))

# Create separate data frames from the list
for(i in seq_along(df_list)) {
  assign(names(df_list)[i], df_list[[i]])
}
```

```
`death_Within 5 Week` <- `death_Within 5 Week` %>% select(-death_interval)
`death_Within 10 Week` <- `death_Within 10 Week` %>% select(-death_interval)
`death_Within 15 Week` <- `death_Within 15 Week` %>% select(-death_interval)
`death_Within 20 Week` <- `death_Within 20 Week` %>% select(-death_interval)
`death_Within 25 Week` <- `death_Within 25 Week` %>% select(-death_interval)
```

```
cor_mat_2 <- cor(`death_Within 5 Week`)
```

```
## Warning in cor(`death_Within 5 Week`): the standard deviation is zero
```

```
cor_mat_3 <- cor(`death_Within 10 Week`)
```

```
## Warning in cor(`death_Within 10 Week`): the standard deviation is zero
```

```
cor_mat_4 <- cor(`death_Within 15 Week`)
```

```
## Warning in cor(`death_Within 15 Week`): the standard deviation is zero
```

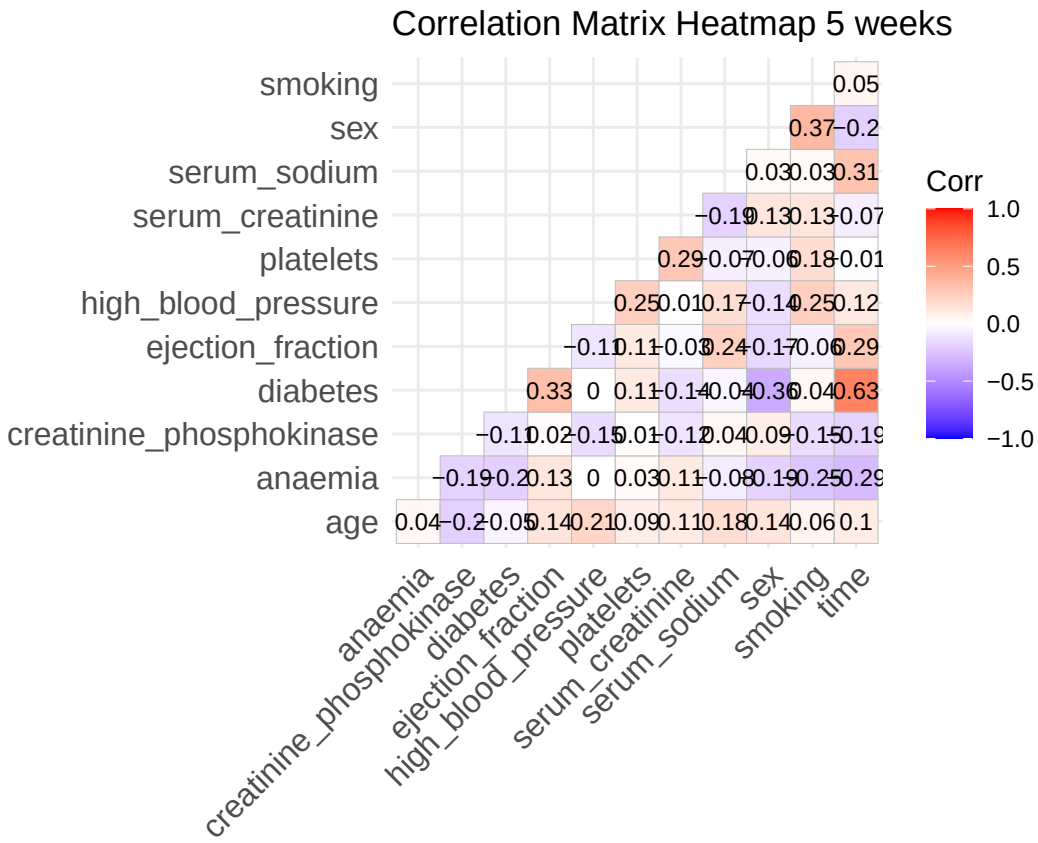
```
cor_mat_5 <- cor(`death_Within 20 Week`)
```

```
## Warning in cor(`death_Within 20 Week`): the standard deviation is zero
```

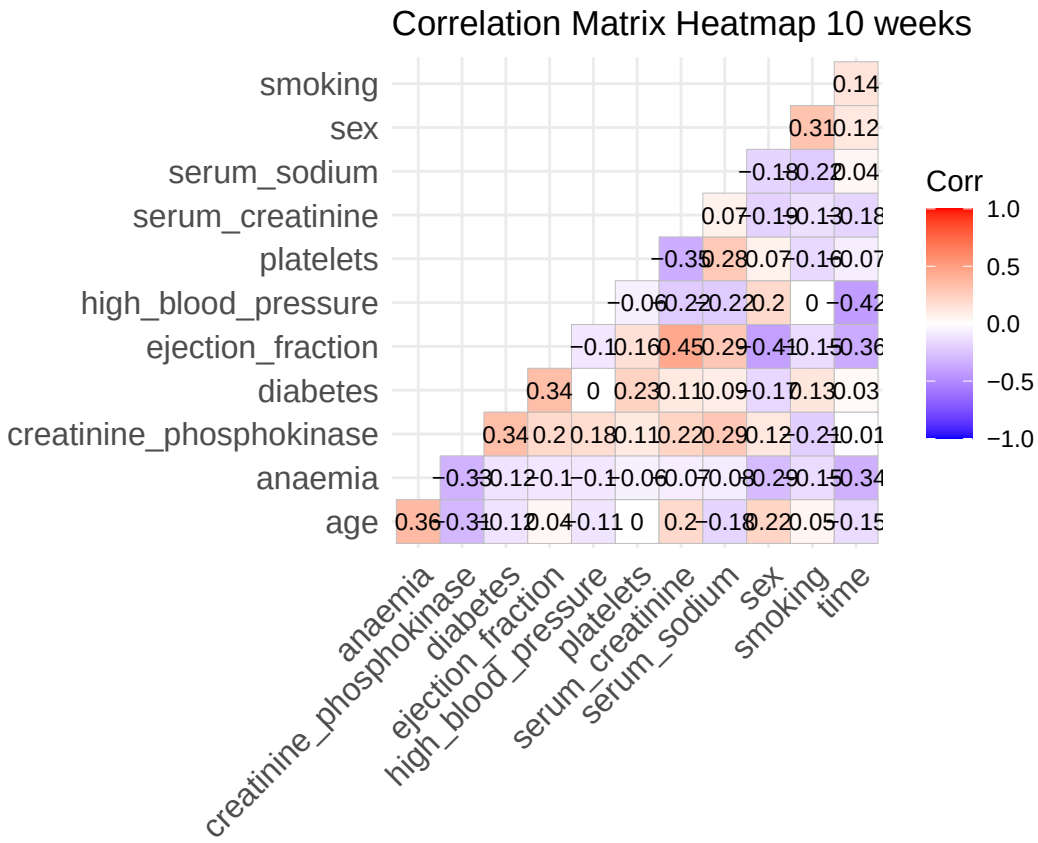
```
cor_mat_6 <- cor(`death_Within 25 Week`)
```

```
## Warning in cor(`death_Within 25 Week`): the standard deviation is zero
```

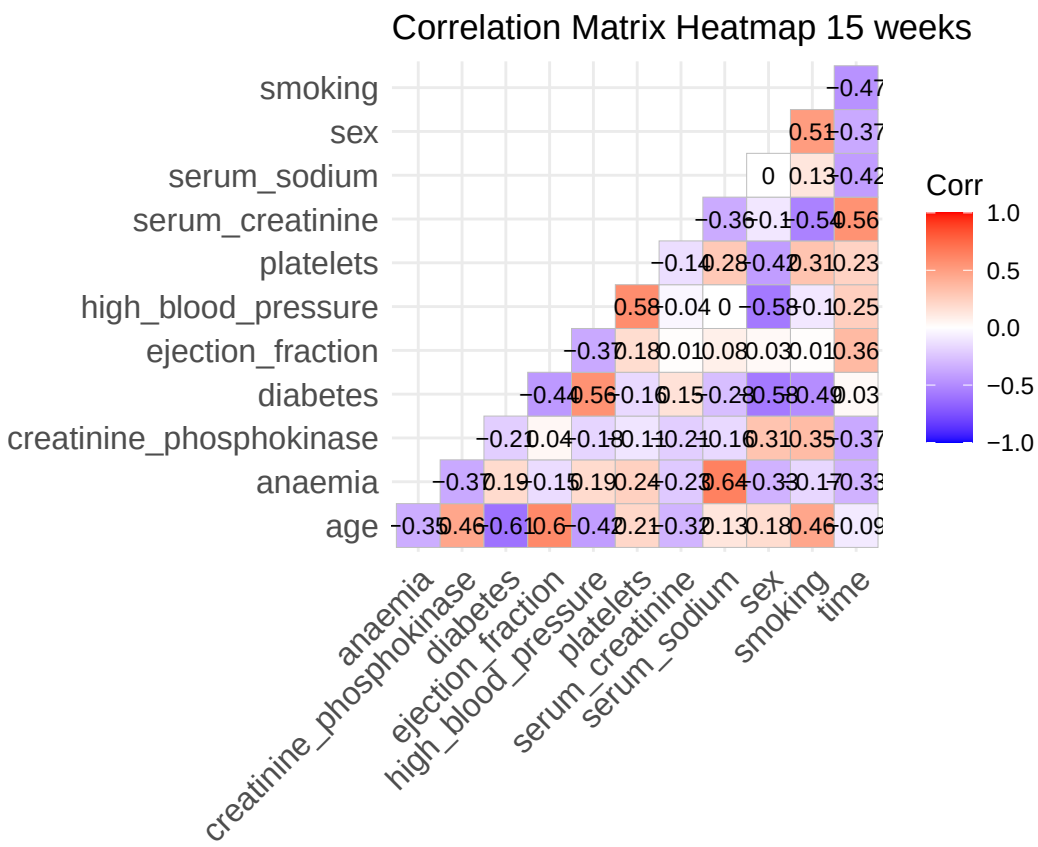
```
ggcorrplot(cor_mat_2,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap 5 weeks")
```



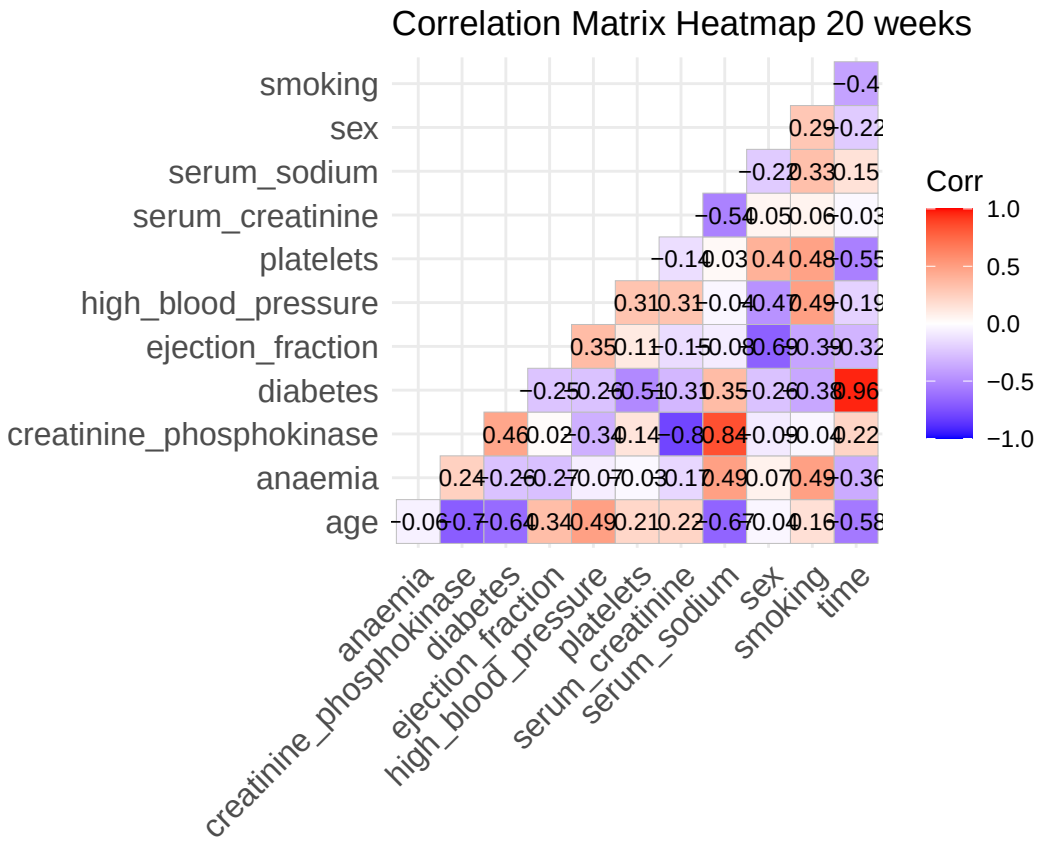

```
ggcorrplot(cor_mat_3,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap 10 weeks")
```



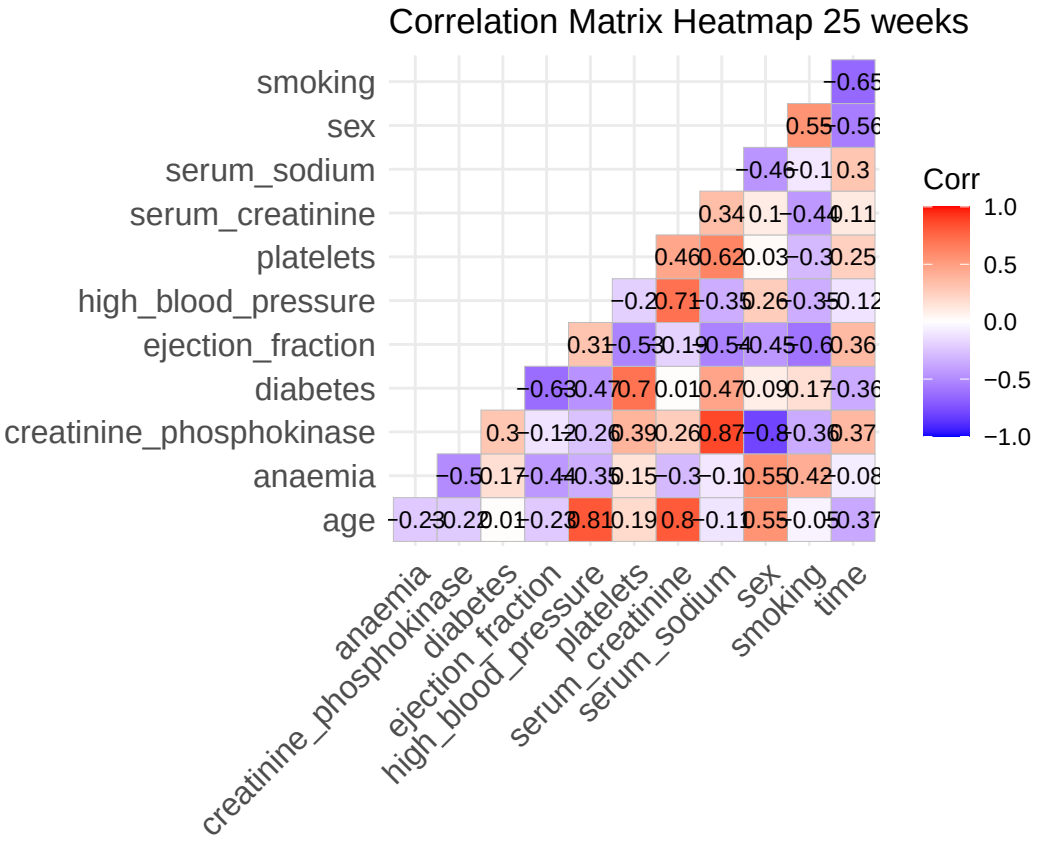

```
ggcorrplot(cor_mat_4,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap 15 weeks")
```




```
ggcorrplot(cor_mat_5,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap 20 weeks")
```




```
ggcorrplot(cor_mat_6,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           colors = c("blue", "white", "red"),
           title = "Correlation Matrix Heatmap 25 weeks")
```




```
summary(`death_Within 5 Week`)
```

```
##       age           anaemia      creatinine_phosphokinase    diabetes
##  Min.   :45.0   Min.   :0.000   Min.   :  23.0           Min.   :0.000
##  1st Qu.:59.5   1st Qu.:0.000   1st Qu.: 122.8           1st Qu.:0.000
##  Median :68.5   Median :1.000   Median : 194.0           Median :0.000
##  Mean   :68.3   Mean   :0.525   Mean   : 489.9           Mean   :0.475
##  3rd Qu.:80.0   3rd Qu.:1.000   3rd Qu.: 582.0           3rd Qu.:1.000
##  Max.   :95.0   Max.   :1.000   Max.   :7861.0           Max.   :1.000
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :14.00     Min.   :0.000       Min.   : 47000   Min.   :0.800
##  1st Qu.:30.00     1st Qu.:0.000       1st Qu.:187250   1st Qu.:1.075
##  Median :38.00     Median :0.000       Median :257500   Median :1.300
##  Mean   :36.65     Mean   :0.475       Mean   :241670   Mean   :1.882
##  3rd Qu.:45.00     3rd Qu.:1.000       3rd Qu.:291000   3rd Qu.:1.900
##  Max.   :65.00     Max.   :1.000       Max.   :454000   Max.   :9.400
##   serum_sodium        sex         smoking          time        DEATH_EVENT
##  Min.   :116.0   Min.   :0.0   Min.   :0.00   Min.   : 4.00   Min.   :1
##  1st Qu.:132.0   1st Qu.:0.0   1st Qu.:0.00   1st Qu.:10.00   1st Qu.:1
##  Median :136.0   Median :1.0   Median :0.00   Median :20.00   Median :1
##  Mean   :134.9   Mean   :0.7   Mean   :0.35   Mean   :19.43   Mean   :1
##  3rd Qu.:138.2   3rd Qu.:1.0   3rd Qu.:1.00   3rd Qu.:28.25   3rd Qu.:1
##  Max.   :145.0   Max.   :1.0   Max.   :1.00   Max.   :35.00   Max.   :1
```

```
summary(`death_Within 10 Week`)
```

```
##       age           anaemia      creatinine_phosphokinase    diabetes
##  Min.   :42.00   Min.   :0.00   Min.   :  68.0           Min.   :0.0
```

```
##  1st Qu.:50.75   1st Qu.:0.00   1st Qu.: 125.0           1st Qu.:0.0
##  Median :60.00   Median :0.00   Median : 341.0           Median :0.0
##  Mean   :61.20   Mean   :0.45   Mean   : 921.6           Mean   :0.4
##  3rd Qu.:70.00   3rd Qu.:1.00   3rd Qu.: 582.0           3rd Qu.:1.0
##  Max.   :95.00   Max.   :1.00   Max.   :7702.0           Max.   :1.0
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :15.00     Min.   :0.0         Min.   :119000   Min.   :0.600
##  1st Qu.:20.00     1st Qu.:0.0         1st Qu.:237000   1st Qu.:1.000
##  Median :25.00     Median :0.5         Median :272500   Median :1.250
##  Mean   :29.40     Mean   :0.5         Mean   :304118   Mean   :1.812
##  3rd Qu.:35.75     3rd Qu.:1.0         3rd Qu.:391250   3rd Qu.:2.050
##  Max.   :62.00     Max.   :1.0         Max.   :497000   Max.   :6.800
##   serum_sodium        sex          smoking          time        DEATH_EVENT
##  Min.   :127.0   Min.   :0.0   Min.   :0.0   Min.   :38.00   Min.   :1
##  1st Qu.:132.8   1st Qu.:0.0   1st Qu.:0.0   1st Qu.:43.00   1st Qu.:1
##  Median :136.0   Median :1.0   Median :0.0   Median :52.50   Median :1
##  Mean   :136.4   Mean   :0.6   Mean   :0.3   Mean   :52.55   Mean   :1
##  3rd Qu.:139.2   3rd Qu.:1.0   3rd Qu.:1.0   3rd Qu.:61.75   3rd Qu.:1
##  Max.   :146.0   Max.   :1.0   Max.   :1.0   Max.   :67.00   Max.   :1
```

```
summary(`death_Within 15 Week`)
```

```
##       age           anaemia    creatinine_phosphokinase    diabetes
##  Min.   :46.00   Min.   :0.0   Min.   :  47.0           Min.   :0.00
##  1st Qu.:60.00   1st Qu.:0.0   1st Qu.: 139.5           1st Qu.:0.00
##  Median :64.50   Median :0.5   Median : 224.0           Median :0.00
##  Mean   :67.67   Mean   :0.5   Mean   : 746.1           Mean   :0.25
##  3rd Qu.:75.25   3rd Qu.:1.0   3rd Qu.: 582.0           3rd Qu.:0.25
##  Max.   :86.00   Max.   :1.0   Max.   :5882.0           Max.   :1.00
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :17.00     Min.   :0.00        Min.   :196000   Min.   :0.700
##  1st Qu.:23.75     1st Qu.:0.00        1st Qu.:215250   1st Qu.:1.150
##  Median :27.50     Median :0.00        Median :264679   Median :1.500
##  Mean   :32.92     Mean   :0.25        Mean   :287530   Mean   :1.661
##  3rd Qu.:38.50     3rd Qu.:0.25        3rd Qu.:303000   3rd Qu.:1.897
##  Max.   :60.00     Max.   :1.00        Max.   :621000   Max.   :3.700
##   serum_sodium        sex          smoking            time          DEATH_EVENT
##  Min.   :124.0   Min.   :0.0   Min.   :0.0000   Min.   : 72.00   Min.   :1
##  1st Qu.:132.0   1st Qu.:0.0   1st Qu.:0.0000   1st Qu.: 76.00   1st Qu.:1
##  Median :134.0   Median :0.5   Median :0.0000   Median : 85.00   Median :1
##  Mean   :134.3   Mean   :0.5   Mean   :0.4167   Mean   : 84.50   Mean   :1
##  3rd Qu.:137.2   3rd Qu.:1.0   3rd Qu.:1.0000   3rd Qu.: 91.25   3rd Qu.:1
##  Max.   :141.0   Max.   :1.0   Max.   :1.0000   Max.   :100.00   Max.   :1
```

```
summary(`death_Within 20 Week`)
```

```
##       age           anaemia      creatinine_phosphokinase    diabetes
##  Min.   :45.00   Min.   :0.000   Min.   :  66.0           Min.   :0.0
##  1st Qu.:56.75   1st Qu.:0.000   1st Qu.: 494.8           1st Qu.:0.0
##  Median :72.00   Median :0.000   Median : 790.5           Median :0.5
##  Mean   :66.62   Mean   :0.375   Mean   :1022.6           Mean   :0.5
##  3rd Qu.:76.25   3rd Qu.:1.000   3rd Qu.:1290.8           3rd Qu.:1.0
##  Max.   :80.00   Max.   :1.000   Max.   :2442.0           Max.   :1.0
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :20.00     Min.   :0.000       Min.   : 70000   Min.   :0.900
```

```
##  1st Qu.:28.75     1st Qu.:0.000       1st Qu.:162750   1st Qu.:1.100
##  Median :32.50     Median :0.000       Median :230000   Median :1.500
##  Mean   :32.62     Mean   :0.375       Mean   :216545   Mean   :1.604
##  3rd Qu.:38.00     3rd Qu.:1.000       3rd Qu.:281018   3rd Qu.:1.972
##  Max.   :45.00     Max.   :1.000       Max.   :338000   Max.   :2.500
##   serum_sodium        sex           smoking           time        DEATH_EVENT
##  Min.   :134.0   Min.   :0.000   Min.   :0.000   Min.   :109.0   Min.   :1
##  1st Qu.:134.0   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:112.5   1st Qu.:1
##  Median :135.0   Median :1.000   Median :0.000   Median :120.5   Median :1
##  Mean   :136.5   Mean   :0.625   Mean   :0.125   Mean   :121.0   Mean   :1
##  3rd Qu.:139.0   3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:129.2   3rd Qu.:1
##  Max.   :142.0   Max.   :1.000   Max.   :1.000   Max.   :135.0   Max.   :1
```

```
summary(`death_Within 25 Week`)
```

```
##       age           anaemia       creatinine_phosphokinase    diabetes
##  Min.   :50.00   Min.   :0.0000   Min.   :  99.0           Min.   :0.0000
##  1st Qu.:58.50   1st Qu.:0.0000   1st Qu.: 124.5           1st Qu.:0.0000
##  Median :60.00   Median :0.0000   Median : 176.0           Median :1.0000
##  Mean   :61.10   Mean   :0.4286   Mean   : 485.3           Mean   :0.5714
##  3rd Qu.:62.83   3rd Qu.:1.0000   3rd Qu.: 488.5           3rd Qu.:1.0000
##  Max.   :75.00   Max.   :1.0000   Max.   :1896.0           Max.   :1.0000
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :25.00     Min.   :0.0000      Min.   :153000   Min.   :0.600
##  1st Qu.:25.00     1st Qu.:0.0000      1st Qu.:220000   1st Qu.:1.100
##  Median :25.00     Median :0.0000      Median :224000   Median :1.200
##  Mean   :31.14     Mean   :0.1429      Mean   :262286   Mean   :1.443
##  3rd Qu.:34.00     3rd Qu.:0.0000      3rd Qu.:315000   3rd Qu.:1.800
##  Max.   :50.00     Max.   :1.0000      Max.   :389000   Max.   :2.500
##   serum_sodium        sex             smoking            time        DEATH_EVENT
##  Min.   :134.0   Min.   :0.0000   Min.   :0.0000   Min.   :150.0   Min.   :1
##  1st Qu.:135.0   1st Qu.:0.5000   1st Qu.:0.0000   1st Qu.:158.0   1st Qu.:1
##  Median :136.0   Median :1.0000   Median :0.0000   Median :170.0   Median :1
##  Mean   :136.7   Mean   :0.7143   Mean   :0.4286   Mean   :164.4   Mean   :1
##  3rd Qu.:136.5   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:171.5   3rd Qu.:1
##  Max.   :144.0   Max.   :1.0000   Max.   :1.0000   Max.   :172.0   Max.   :1
```