

ProjectFirst

Teemu Sormunen, Abdullah Günay, Nicola Brazzale

11/18/2020

Contents

1	Introduction	1
2	Packages	1

1 Introduction

The dataset of 299 patients was produced as a result of study [1].

2 Packages

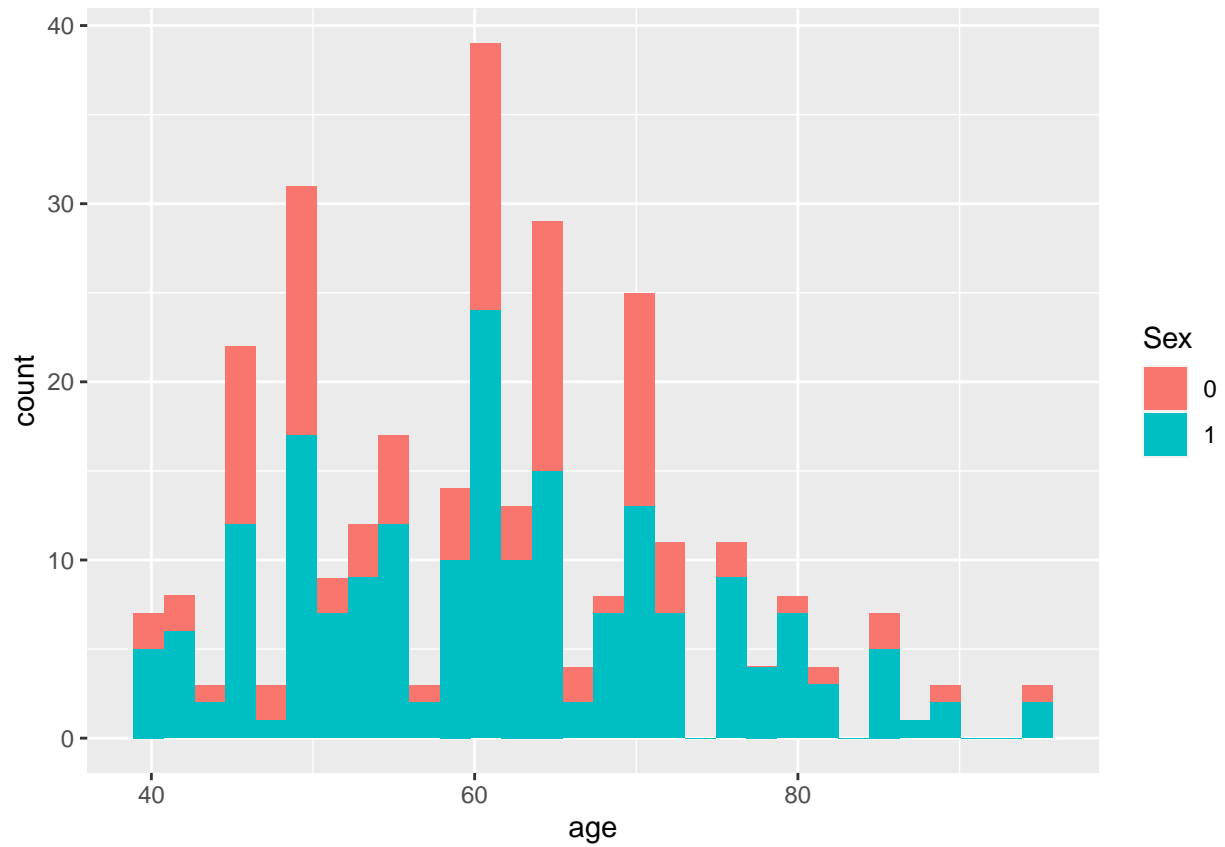
Load data

```
file.name <- './data/heart_failure_clinical_records_dataset.csv'  
heart <- read_csv(file.name)
```

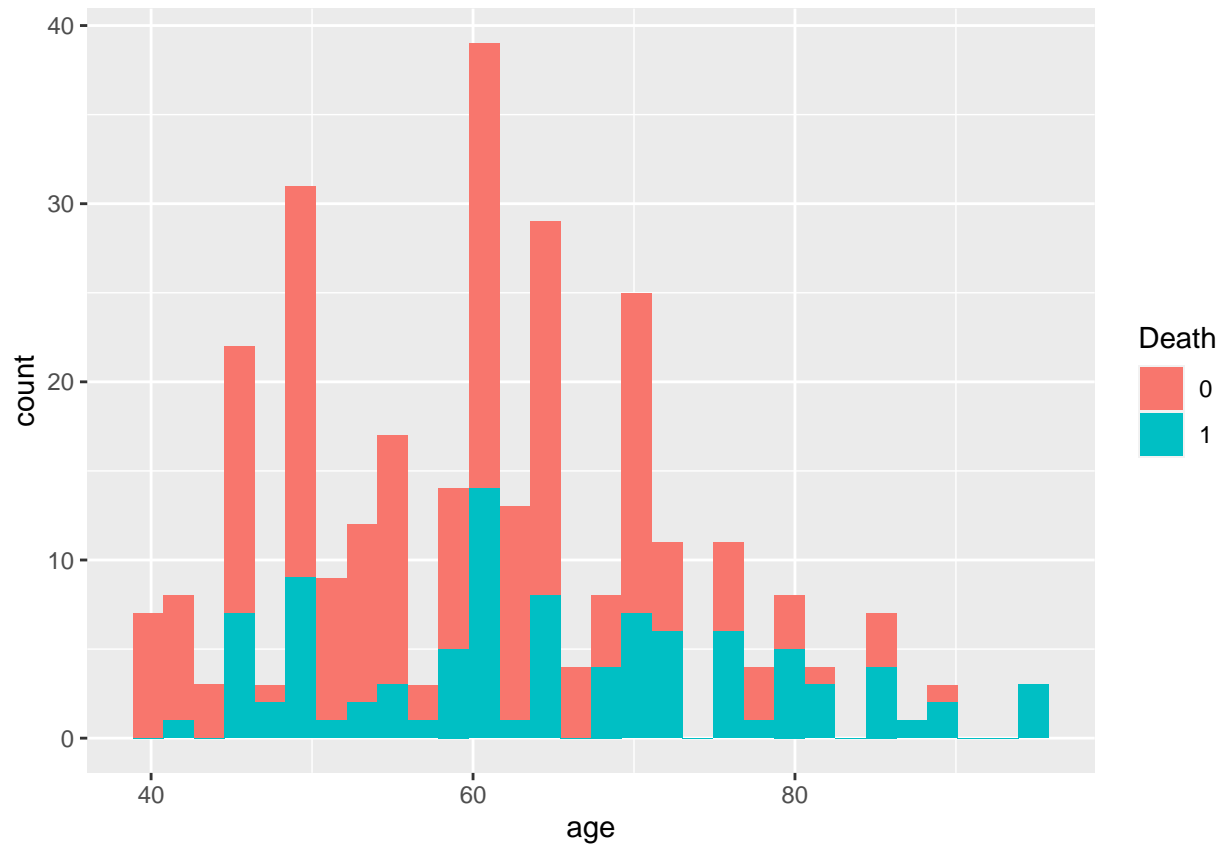
```
## Parsed with column specification:  
## cols(  
##   age = col_double(),  
##   anaemia = col_double(),  
##   creatinine_phosphokinase = col_double(),  
##   diabetes = col_double(),  
##   ejection_fraction = col_double(),  
##   high_blood_pressure = col_double(),  
##   platelets = col_double(),  
##   serum_creatinine = col_double(),  
##   serum_sodium = col_double(),  
##   sex = col_double(),  
##   smoking = col_double(),  
##   time = col_double(),  
##   DEATH_EVENT = col_double()  
## )
```

Plot histograms

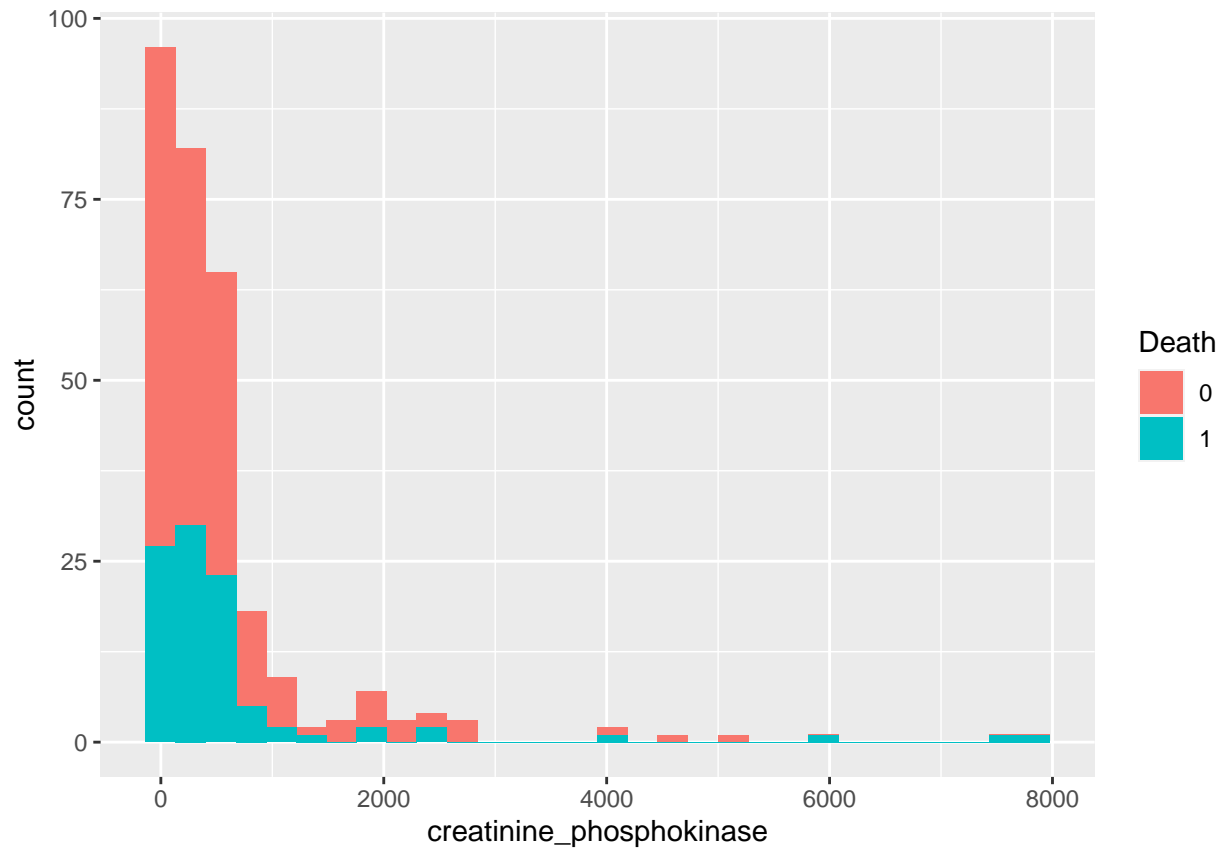
```
ggplot(heart, aes(x=age)) + geom_histogram(aes(fill=as.character(sex)), bins = 30) + labs(fill = "Sex")
```



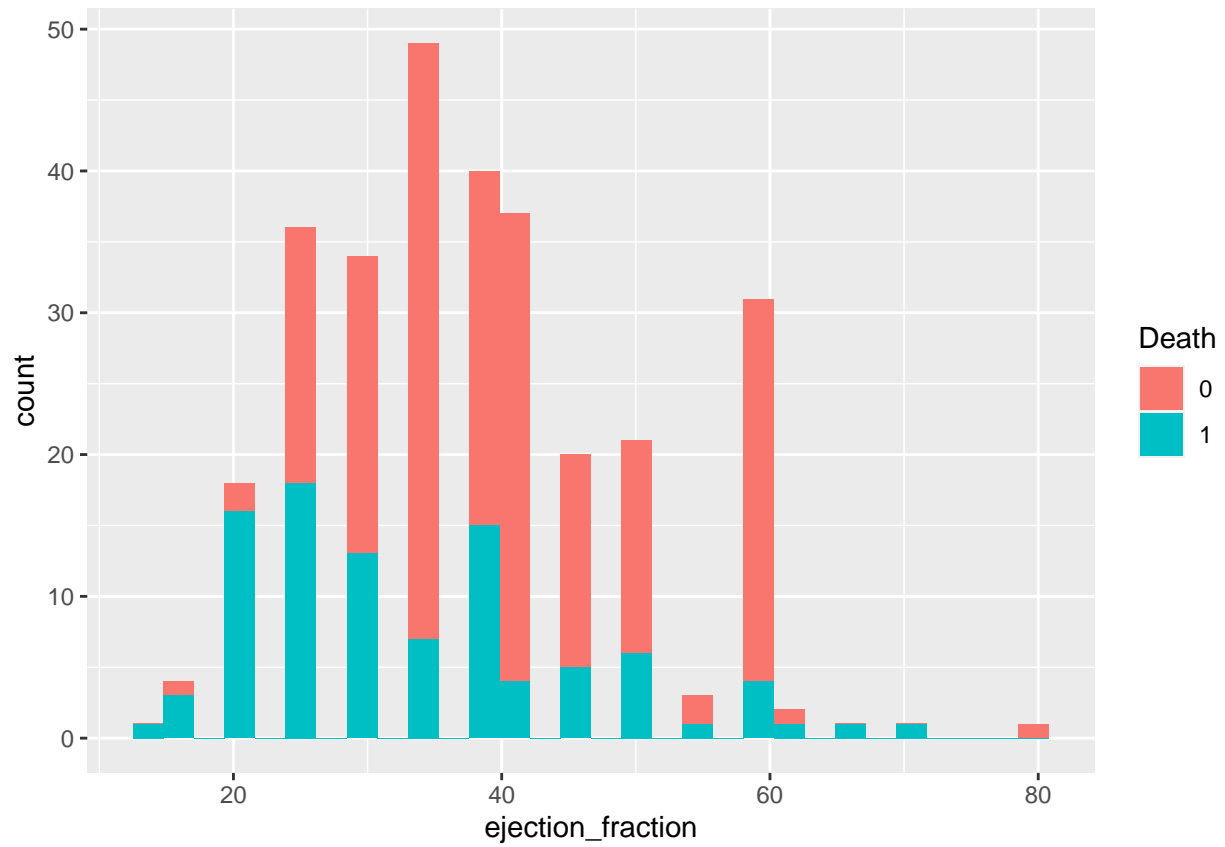
```
ggplot(heart, aes(x=age)) + geom_histogram(aes(fill=as.character(DEATH_EVENT)), bins = 30) + labs(fill = "Sex")
```



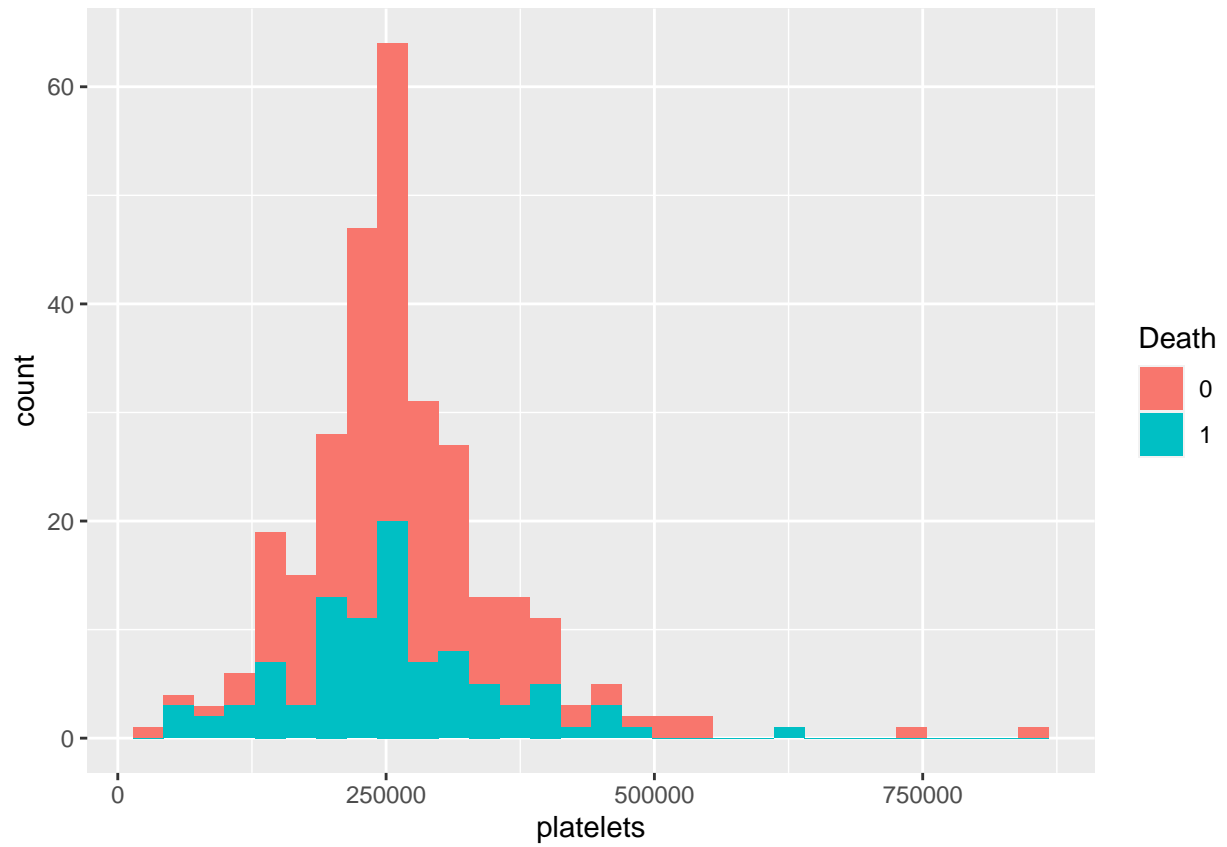
```
ggplot(heart, aes(x=creatinine_phosphokinase)) + geom_histogram(aes(fill=as.character(DEATH_EVENT)), bin
```



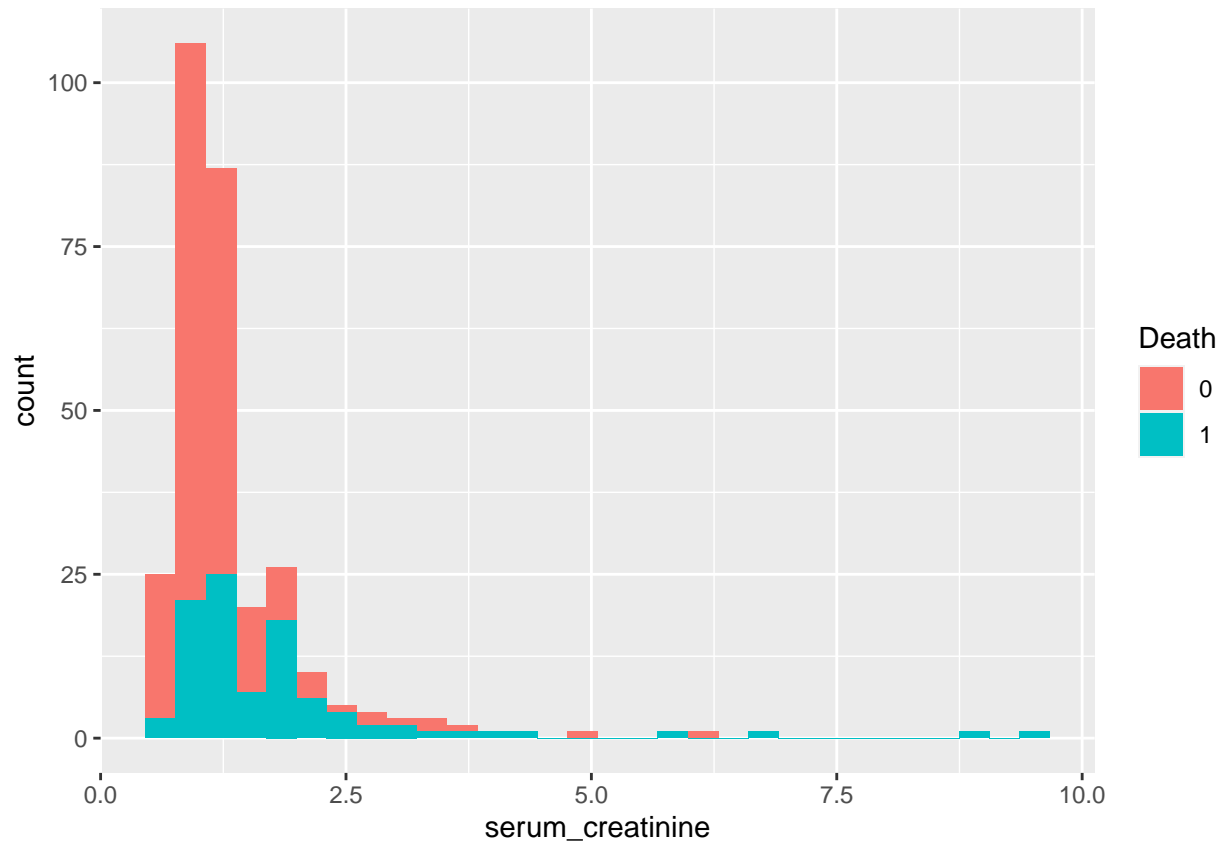
```
ggplot(heart, aes(x=ejection_fraction)) + geom_histogram(aes(fill=as.character(DEATH_EVENT)), bins = 30)
```



```
ggplot(heart, aes(x=platelets)) + geom_histogram(aes(fill=as.character(DEATH_EVENT)), bins = 30) + labs
```

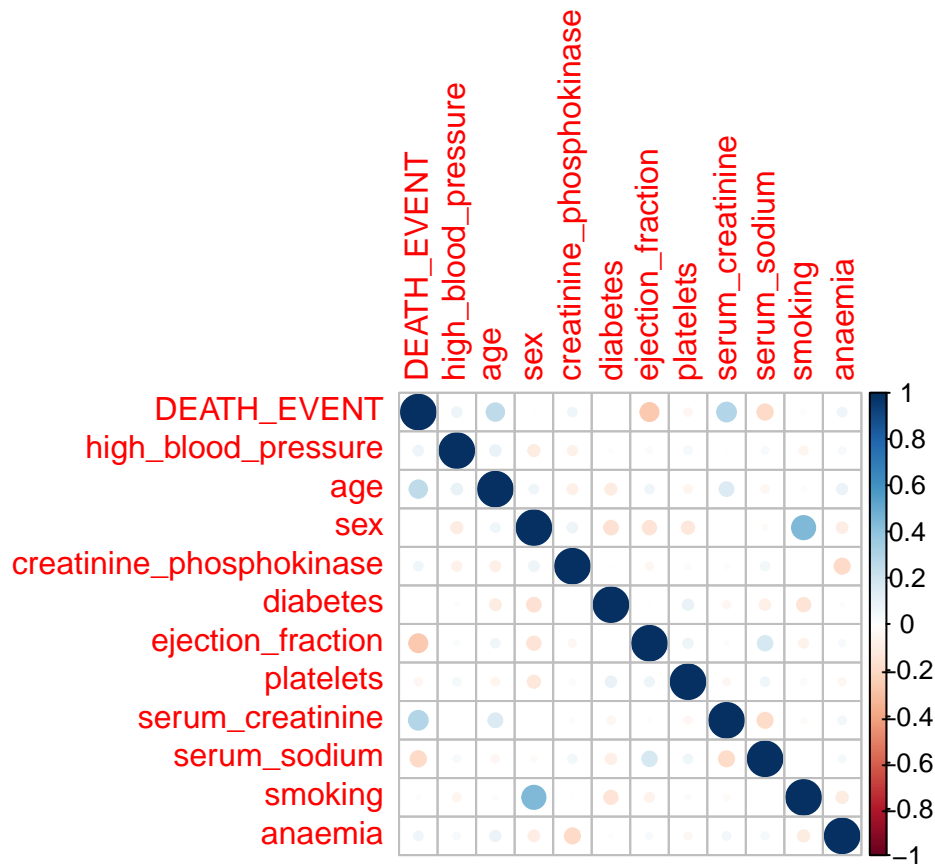


```
ggplot(heart, aes(x=serum_creatinine)) + geom_histogram(aes(fill=as.character(DEATH_EVENT)), bins = 30)
```



Correlation matrix

```
pred <- c("high_blood_pressure", "age", "sex", "creatinine_phosphokinase", "diabetes", "ejection_fractions")
target <- c("DEATH_EVENT")
#formula <- paste("DEATH_EVENT ~", paste(pred, collapse = "+"))
p <- length(pred)
n <- nrow(heart)
x = cor(heart[, c(target, pred)])
corrplot(x)
```



```
#PRIORS ?
# 1 LINEAR MODEL WITH VARIABLE SELECTION
# 2 LINEAR MODEL WITH ALL VARIABLES
# HIERARCHICAL - in the Titanic one a hier. model have been used so we can use that one for reference

# All models with bernoulli outcome (1-0, death or not)
```

References

- [1] Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA. Survival analysis of heart failure patients: A case study. 2017. doi: <https://doi.org/10.1371/journal.pone.0181001>.