Read the data

```
# install.packages("lme")
# install.packages("mgcv")
# install.packages("plotly")
# install.packages("ggplot2")
library('mgcv')
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-38. For overview type 'help("mgcv-package")'.
```

```
library(ggplot2)
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unable to identify current timezone 'C':
## please set environment variable 'TZ'
```

```
data  = read.csv('Emergency department 2012 to 17 datathon v1.1.xlsx - Data.csv', header=TRUE)
```

Recoding

Re-code DOW to only indicate weekday (=1) and weekend (=2)

```
data$DOW = as.factor(data$DOW)
data$Dischgto = as.factor(data$Dischgto)
```

Create new columns

The time in minutes from coming in and leaving

```
data$Time_in_Parse = as.POSIXct(data$Time_in, format="%H:%M:%S")
data$Time_out_Parse = as.POSIXct(data$Time_out, format="%H:%M:%S")
data$Hour_in = format(data$Time_in_Parse, format="%H")
data$Minutes_in = difftime(data$Time_out_Parse, data$Time_in_Parse, units="mins")
data$Minutes_in[data$Minutes_in < 0 & !is.na(data$Minutes_in)] = (24*60) + data$Minutes_in[data
$Minutes_in < 0 & !is.na(data$Minutes_in)]
data$DischICU[is.na(data$DischICU)] = 0
```

```
data$mrt = 0
data$mrt[data$X24hmrt == 1] = 1
data$mrt[data$mrt == 0 & data$X7dmrt ==  1] = 2
data$mrt[data$mrt == 0 & data$X30dmrt == 1] = 3
# assume that a Triagestm of NA is a new class == 0
data$Htriage[data$Htriage == 6] = 5
data$Triagestm[is.na(data$Triagestm)] = 0
```

Save a version

```
write.csv(data,file="triage_recoding.csv")
```

What column do we want to use?

```
cols = c("Age", "Kjonn", "Shift", "DOW", "EDLOS", "Hour_in", "mrt", "DischICU", "Htriage", "Triagestm")
data_filtered = data[,cols]
data_filtered$EDLOS = as.numeric(data_filtered$EDLOS)
```

```
## Warning: NAs introduced by coercion
```

```
data_filtered$Hour_in = as.numeric(data_filtered$Hour_in)
data_filtered$Htriage = as.factor(data$Htriage)
data_filtered$Triagestm = as.factor(data$Triagestm)
data_filtered$DischICU = as.factor(data$DischICU)
data_filtered$mrt = as.factor(data$mrt)
```

```
summary(complete.cases(data_filtered))
```

```
##    Mode   FALSE    TRUE
## logical     128  205359
```

```
library(tableone)
```

```
## Warning: package 'tableone' was built under R version 4.1.3
```

```
CreateTableOne(data=data_filtered)
```

```
## 
##                         Overall
##   n                     205487
##   Age (mean (SD))        56.20 (23.98)
##   Kjonn (%)
##                              7 ( 0.0)
##      Kvinne              97535 (47.5)
##      Mann               107945 (52.5)
##   Shift (%)
##                           9421 ( 4.6)
##      Dag                 98901 (48.1)
##      Kveld               72124 (35.1)
##      Natt                25041 (12.2)
##   DOW = Man-fre (%)   145952 (71.0)
##   EDLOS (mean (SD))   180.30 (1659.45)
##   Hour_in (mean (SD))  13.52 (5.60)
##   mrt (%)
##      0                  198518 (96.6)
##      1                    1072 ( 0.5)
##      2                    1946 ( 0.9)
##      3                    3951 ( 1.9)
##   DischICU = 1 (%)      9010 ( 4.4)
##   Htriage (%)
##      1                   19472 ( 9.5)
##      2                   30111 (14.7)
##      3                   73255 (35.6)
##      4                   80334 (39.1)
##      5                    2315 ( 1.1)
##   Triagestm (%)
##      0                   21888 (10.7)
##      1                   43453 (21.1)
##      2                   85867 (41.8)
##      3                   54279 (26.4)
```

```
write.csv(data_filtered,file="triage_recoding.csv", row.names=FALSE)
```

Plot the HTriage score over mortality

```
data2 = data_filtered[data_filtered$mrt == 1,]  # 24h
summary(data2$Htriage)
```

```
##   1   2   3   4   5 
## 819 123 103  25   2
```

```
data2 = data_filtered[data_filtered$mrt == 3,]  # 24h
summary(data2$Htriage)
```

```
##    1    2    3    4    5
##  844  946 1333  812   16
```

A linear model for feature importance

Now we can start a basic model to explain what we see:

```
data_filtered$mrt[data_filtered$mrt == 0] = 300
```

```
## Warning in `[<-.factor`(`*tmp*`, data_filtered$mrt == 0, value =
## structure(c(NA, : invalid factor level, NA generated
```

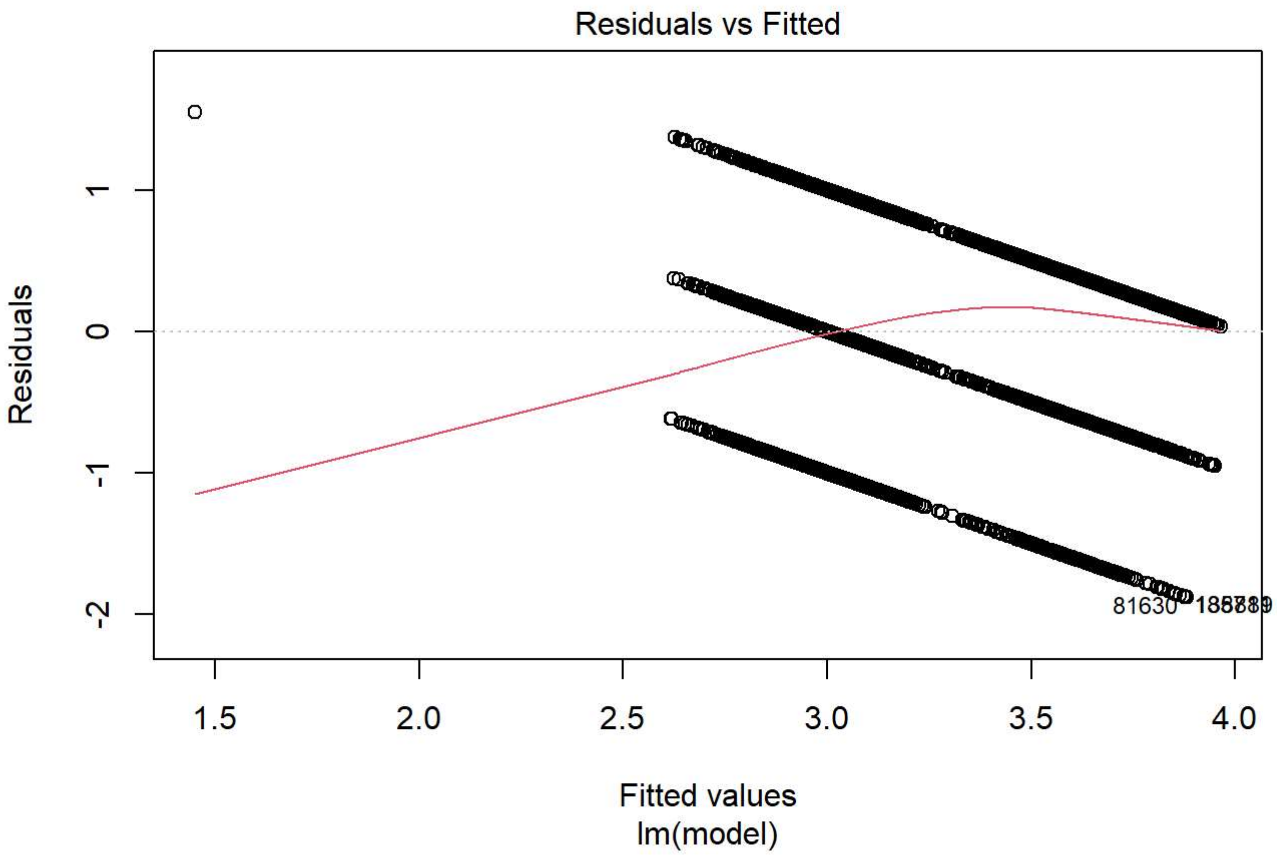
```
independent="mrt"
dependent=cols[-grep (independent, cols)]
dependent=dependent[-grep("Shift", dependent)]
#dependent=dependent[-grep("Hour_in", dependent)]
#dependent=dependent[-grep("EDLOS", dependent)]
data_filtered$mrt = as.numeric(data_filtered$mrt)
model = formula(paste("mrt~", paste(dependent, collapse="+")))
data_filtered = data_filtered[complete.cases(data_filtered),]
fit = lm(model, data=data_filtered)
```

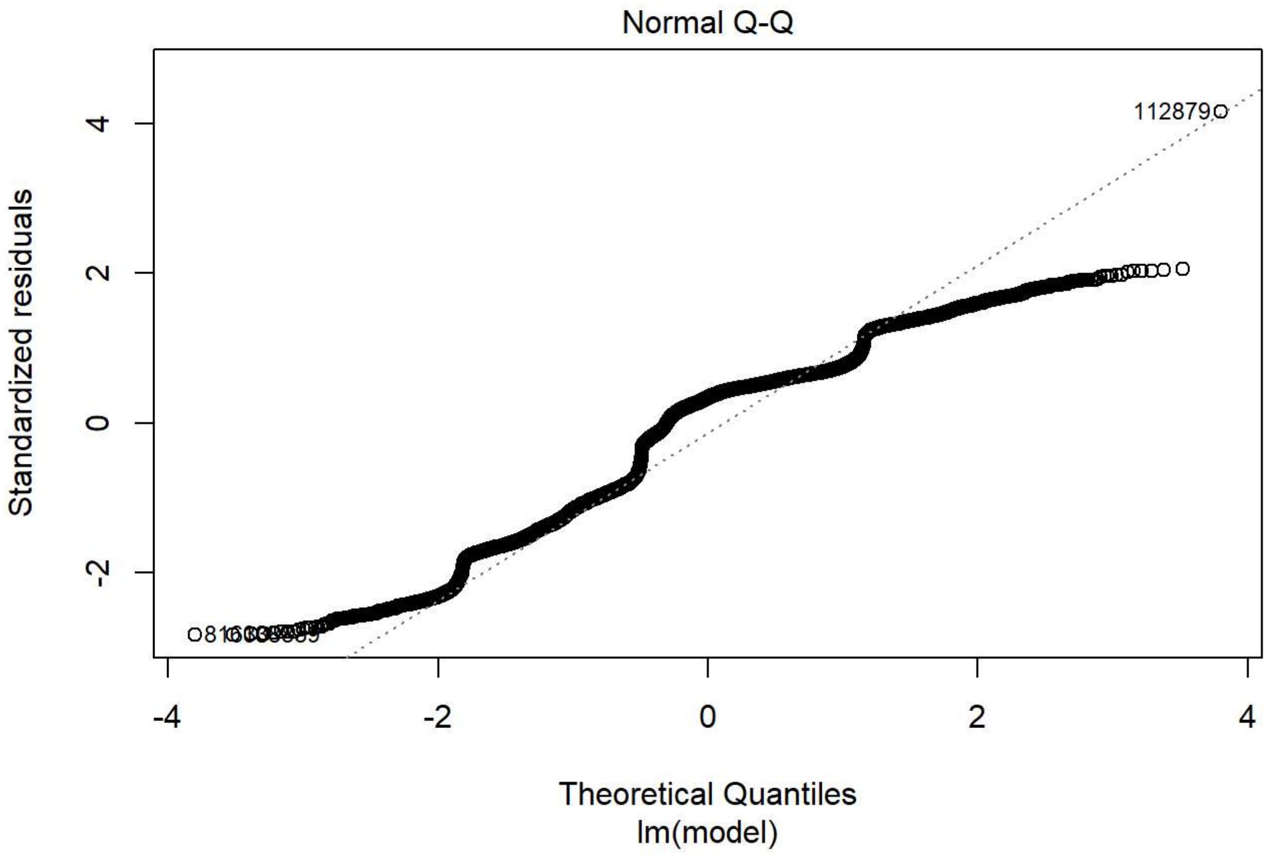
```
summary(fit)
```

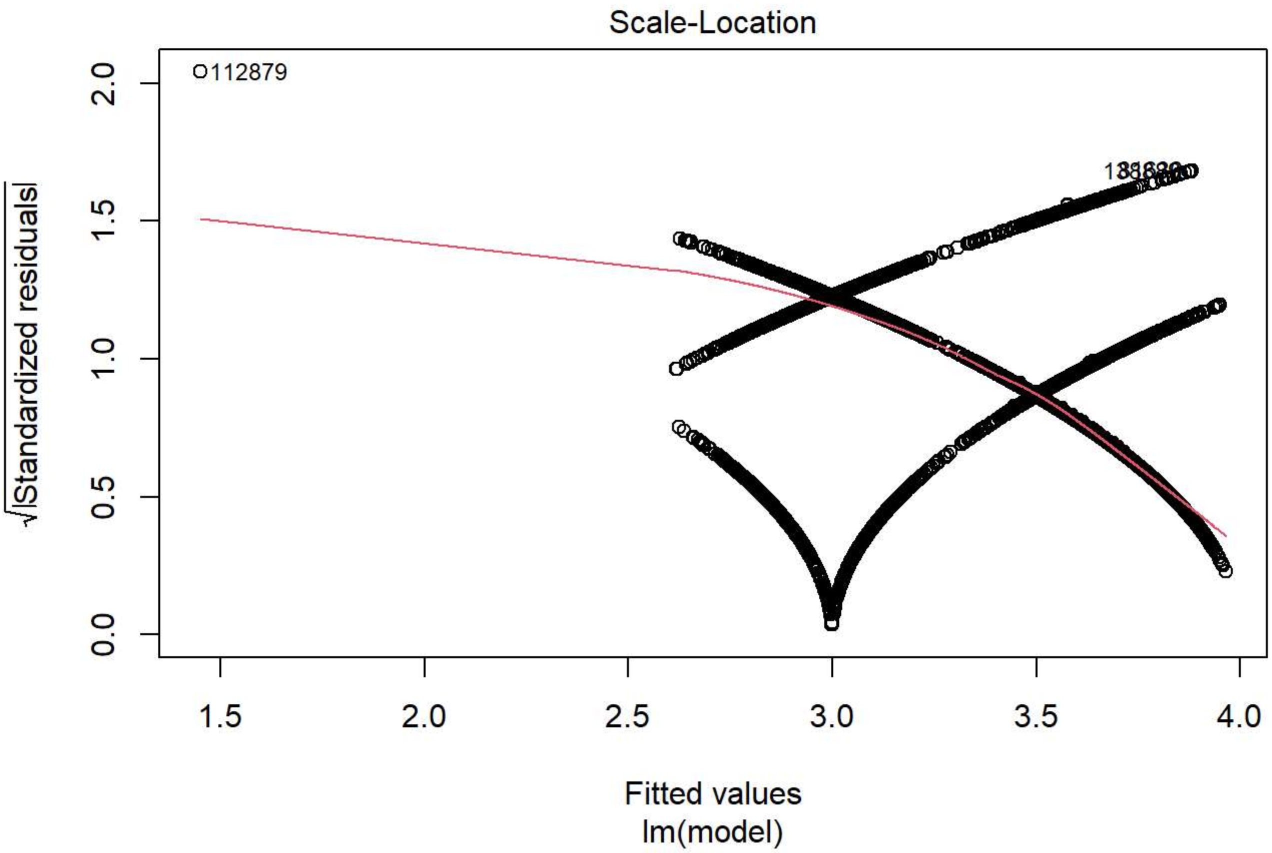
```
##
## Call:
## lm(formula = model, data = data_filtered)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8821 -0.5882  0.2257  0.4177  1.5492
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.421e+00  6.691e-01   3.618 0.000300 ***
## Age          1.348e-03  6.067e-04   2.222 0.026289 *
## KjonnKvinne  4.717e-01  6.667e-01   0.707 0.479292
## KjonnMann    4.585e-01  6.667e-01   0.688 0.491657
## DOWMan-fre   1.427e-02  1.742e-02   0.819 0.412893
## EDLOS       -1.919e-04  4.759e-05  -4.032 5.59e-05 ***
## Hour_in      8.500e-03  1.451e-03   5.860 4.84e-09 ***
## DischICU1   -2.171e-01  2.730e-02  -7.953 2.11e-15 ***
## Htriage2     4.695e-01  2.264e-02  20.735  < 2e-16 ***
## Htriage3     5.786e-01  2.185e-02  26.482  < 2e-16 ***
## Htriage4     7.468e-01  2.676e-02  27.902  < 2e-16 ***
## Htriage5     4.960e-01  1.401e-01   3.541 0.000401 ***
## Triagestm1  -1.126e-01  2.974e-02  -3.786 0.000154 ***
## Triagestm2   7.189e-03  2.753e-02   0.261 0.793982
## Triagestm3   3.028e-02  2.911e-02   1.040 0.298218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.666 on 6948 degrees of freedom
## Multiple R-squared:  0.1951, Adjusted R-squared:  0.1935
## F-statistic: 120.3 on 14 and 6948 DF,  p-value: < 2.2e-16
```

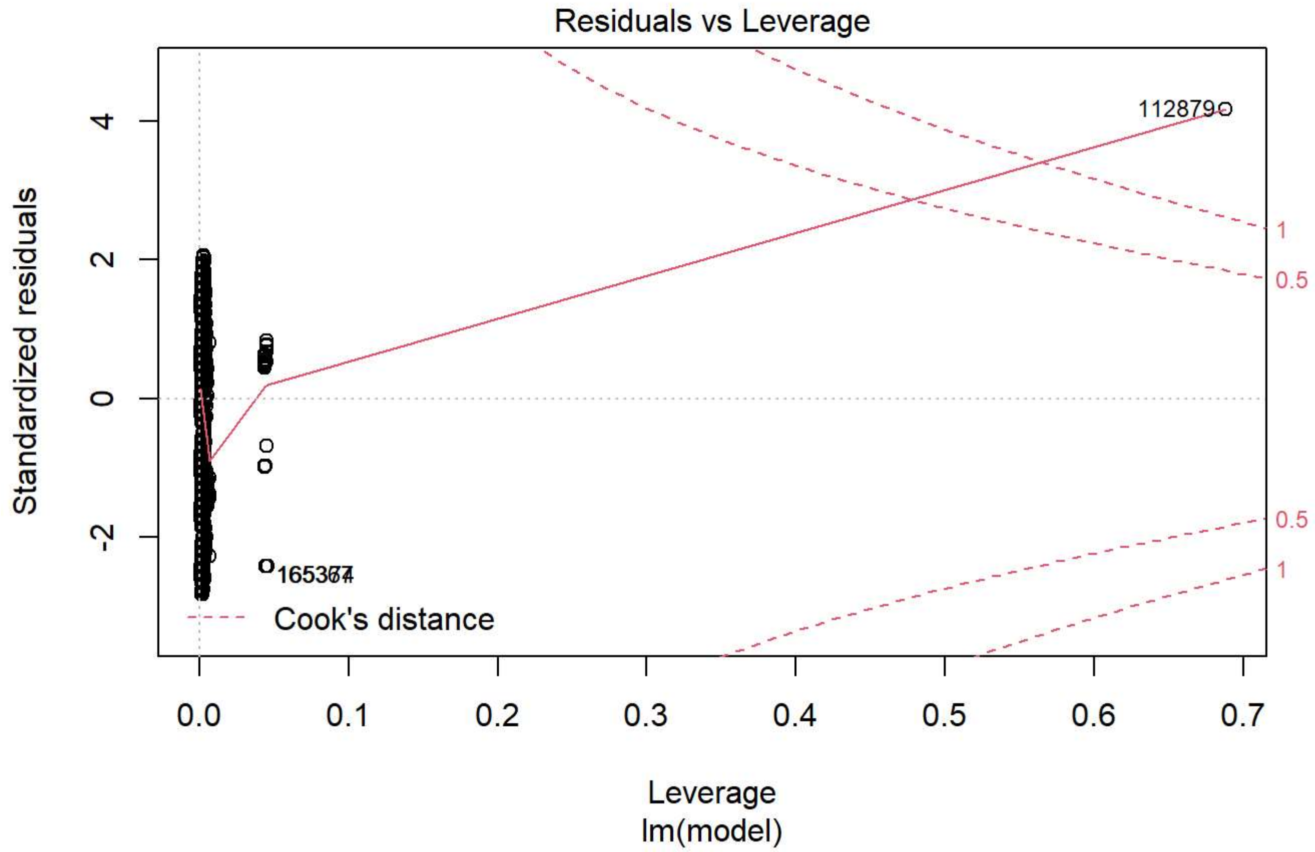
```
plot(fit)
```

```
## Warning: not plotting observations with leverage one:
##   1079
```


Residuals vs Fitted
Residuals
1
0
-1
-2
1.5
2.0
2.5
3.0
3.5
4.0
81630
185789
Fitted values
lm(model)


Normal Q-Q
Standardized residuals
4
2
0
-2
112879
81630
-4
-2
0
2
4
Theoretical Quantiles
lm(model)


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
112879
138839
1.5
2.0
2.5
3.0
3.5
4.0
Fitted values
lm(model)


Residuals vs Leverage
Standardized residuals
-2
0
2
4
112879
165374
Cook's distance
1
0.5
0.5
1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Leverage
lm(model)