

Latency and capacity estimation for a network connection from asymmetric measurements

P.1 - Analysis of the liglab2.log set

Read the data from the file and only take the date, size and time fields

```
dt = read.table('liglab2.log', sep=' ', na.strings = "", header=F, fill = TRUE) %>% select(V1, V2, V9)

# rename the columns to user friendly names
colnames(dt)[colnames(dt)=="V1"] <- "date"
colnames(dt)[colnames(dt)=="V2"] <- "size"
colnames(dt)[colnames(dt)=="V9"] <- "time"
```

Clean the data from any NA value

```
NA_rows = apply(dt, 1, function(x) any(is.na(x)))
dt = dt %>% drop_na()
```

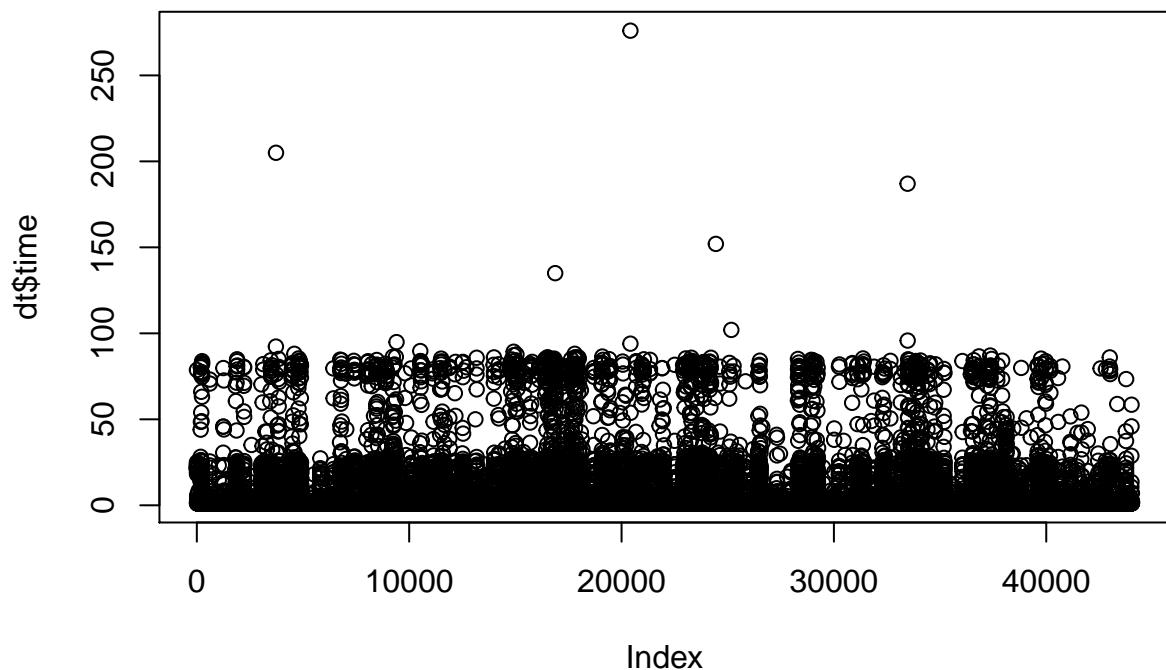
Convert times into readable values

```
# define the time conversion function
convertTime = function(time)
  gsub("[^0-9.]", "", time)

# apply the previously defined function to our data
dt$time = as.numeric(sapply(dt$time, convertTime))
```

Plot the time values

```
plot(dt$time)
```



Use `ggplot` to get a better view on the data. It clearly divides the data into 2 sets depending on the size of the message: `set(A) < 1500 & set(B) > 1500`.

```
library(ggplot2)

ggplot(data = dt, mapping = aes(x = size, y = time)) + geom_point() + theme_gray()
```

