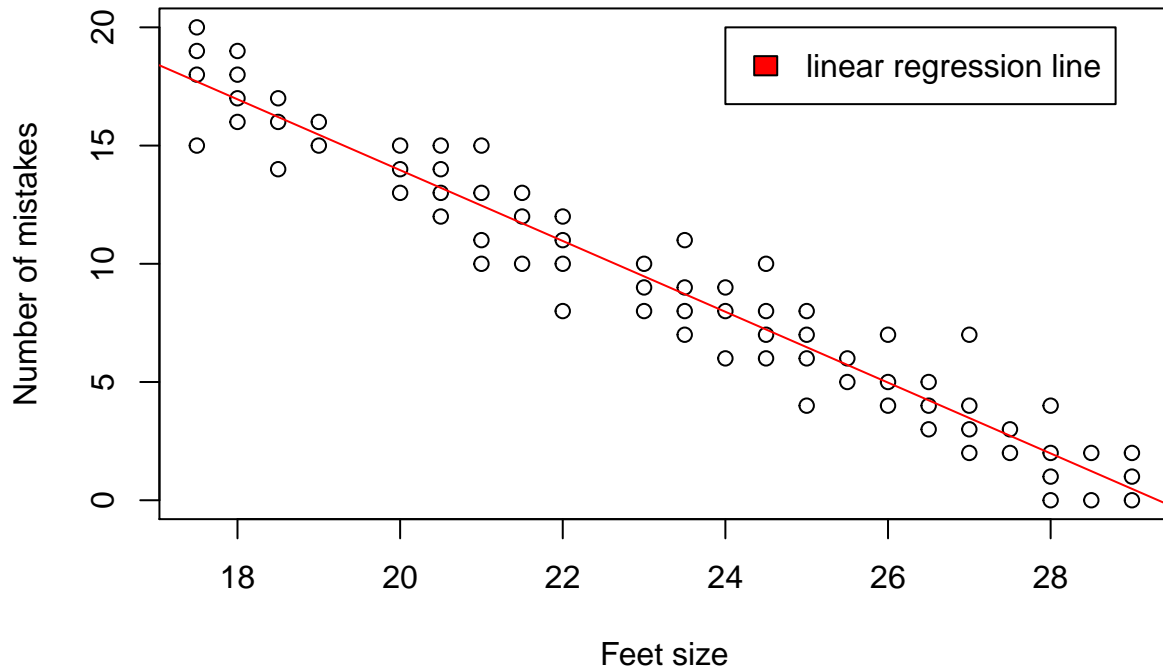# Analysis

Vincent RIBOT

11/23/2021

Experiment

The purpose of the experiment is to verify if there exist a correlation between the number of mistakes of a test made by kids of different ages in an elementary school and their feet size. Each of the 72 kids pass the same test and the number of mistake along with their feet size is collected.

Visualization

I choose to display the cloud of points (each one of them is a kid) so I would be able to see if this cloud is drawing some curve by its own form. In contrary, if all the points of the cloud are scattered all around the graph, that will probably mean that there is no correlation or that the experiment is not adequate to make an hypothesis on the correlation.

```r
data <- read.csv("~/Documents/Etude/M2bis/SM/SMPE-Perso/data.csv", dec=",")
plot(data,main="Number of mistakes mades by all kids in an elementary school\n related to their feet si
     xlab="Feet size", ylab="Number of mistakes")
reg<-lm(NB_OF_MISTAKES ~ FEET_SIZE, data = data)
abline(reg, col="red")
legend(24, 20, legend="linear regression line", fill="red")
```

Number of mistakes mades by all kids in an elementary school related to their feet size

As we can clearly see that the cloud is drawing a line, I deceived to print the regression linear curve. This will help visualize any correlation directly from the graph.

Analysis

By interpreting the data visualization made above, we see that there is and affine relation between the two data (number of mistakes and feet size). There is the correlation coefficient:

```
cor(data)
```

```
##                FEET_SIZE NB_OF_MISTAKES
## FEET_SIZE      1.0000000     -0.9665323
## NB_OF_MISTAKES -0.9665323     1.0000000
```

The result is really close to -1 so we can deduce that the two data are strongly correlated. However we must not take the linear regression line drew above as the function that connect the two data (negative values for feet size or number of mistakes have no sense fo instance).