

# Report Data Analysis with R

## 1<sup>st</sup> assignment

*Person in charge: Mr. Mai Long, Dr.Navet*

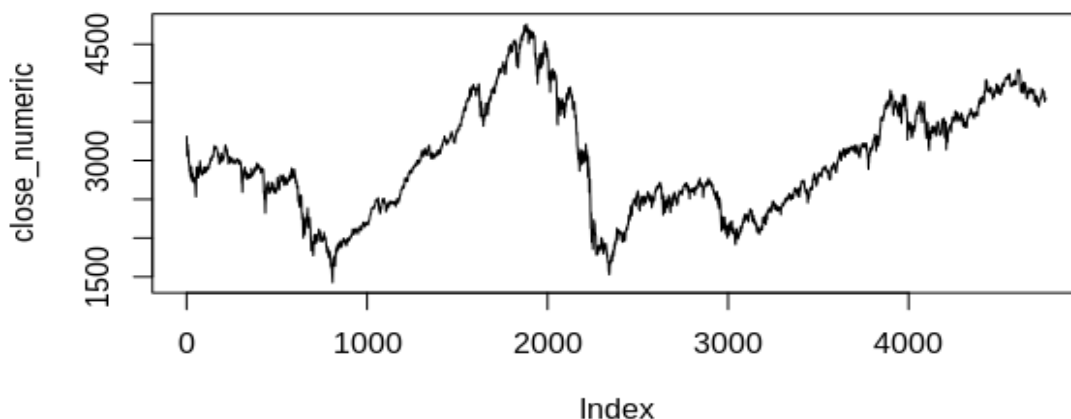
*Student: Pedro Gomes, 017066611B*

### Question 1.

It is worth noticing that after loading the file with the following command: `eof <- read.csv2("^BFX.txt", sep = ",")` the columns that seemed to be numerical such as the Close columns, was of the type "factor", a conversion to numeric was necessary. According to [0], just `as.numeric` is not enough, and I noticed that only when I was doing the other questions. Using the `levels()` functions ordered the numbers in Close, which was not ok for the plot since I got a continuous linear line and this was clearly not the case. Hence I used `as.numeric(as.character(eof$Close))`.

### Question 2.

a) After observing the following plot:



We can see that the close values change a lot from day to day. It does not seem to be a relation between the Close values.

b) There are a total of 4759 entries.

c) According to [1] the variance is :numerical measure of how the data values is dispersed around the mean. We are asked the max variance between the min and max, hence the whole dataset of Close. I simply calculated the normal variance for \$Close.

### Question 3.

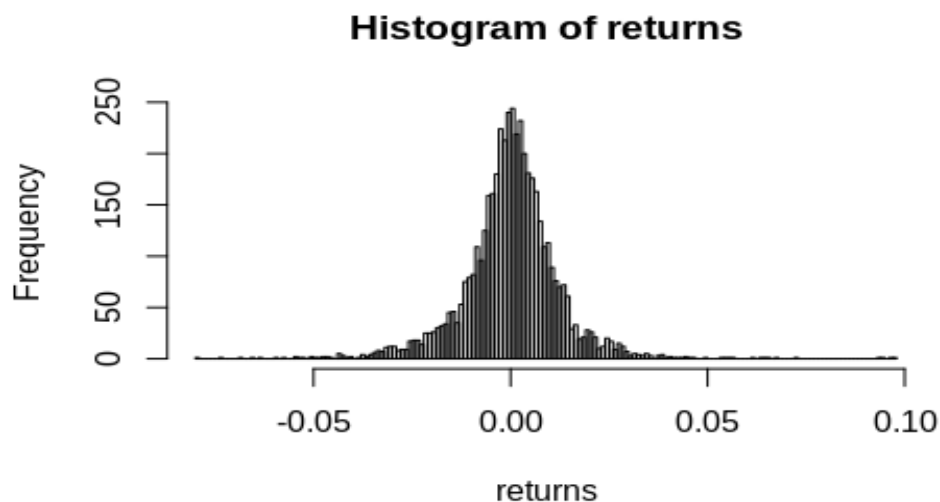
In there would be a loss of: -15.042056, and the highest value would be: 56.914272.

### Question 4.

- a) *See the code for more details.*
- b) The min is: -0.079826 ; the max is: 0.097835
- c) There are 24 records above 4%
- d) There are 4734 records below 4%

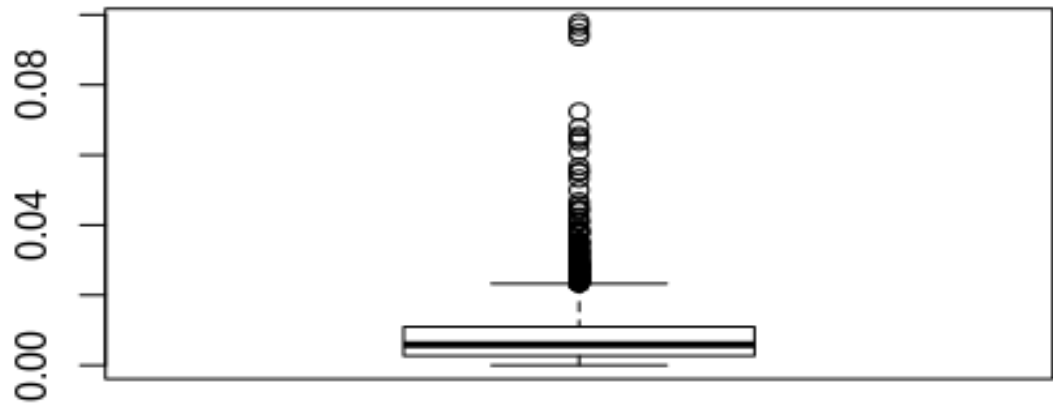
### Question 5.

From the first task of the empirical histogram, I got the following histogram:

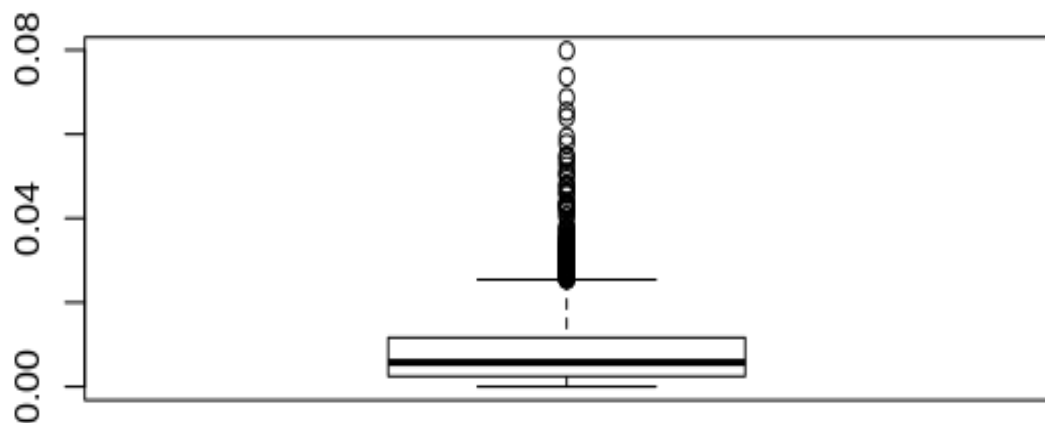


We can see that the great majority of returns, either negative or positive (non-profitable – profitable) are very close to 0.00. This means that the losses are not big however it also means that the gains are not great neither, at least in percentage.

From the task 1) I got the following boxplot:



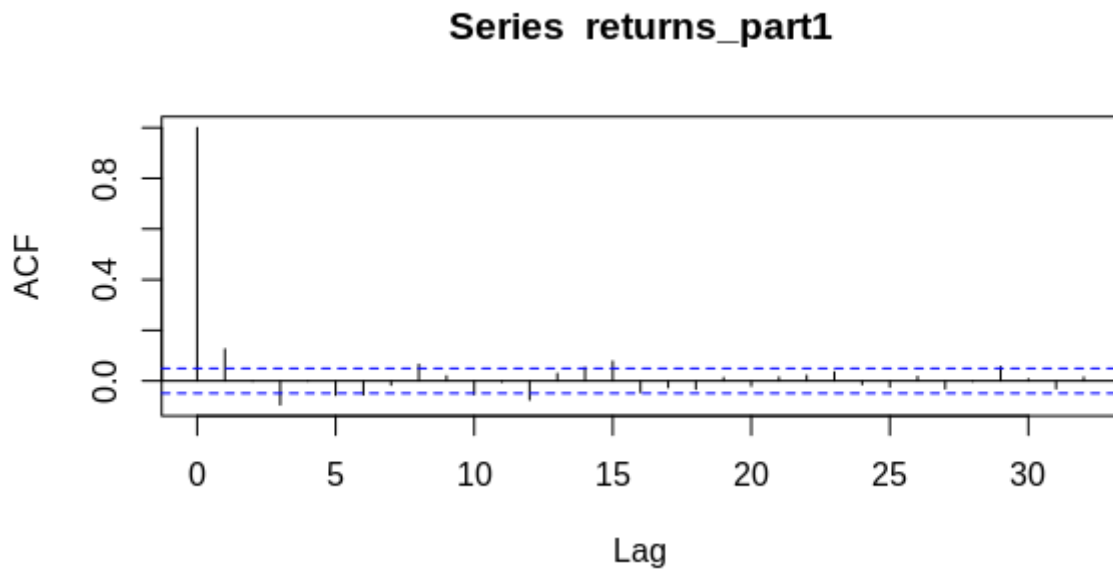
We can observe the Q1, Q2 and Q3 and the minimum value of 1)  
 From the task 2) I got the following boxplot:



The same observations can be taken from 1) . The outliers values of 2) are more close to the maximum( $Q3 + 1.5 \cdot \text{inter\_quantile range}$ ) than in 1) . I would say that the median is almost the same for both cases. Apart from the outliers I see almost no differences between 1) and 2). This may mean that the minimum money lost in the returns( $Q1 - 1.5 \cdot \text{IQR}$ ) is somehow proportional to the minimum money earned. Clearly, negative and positives values for each, since in 2) we use the absolute values. Same thoughts apply for the maximum.

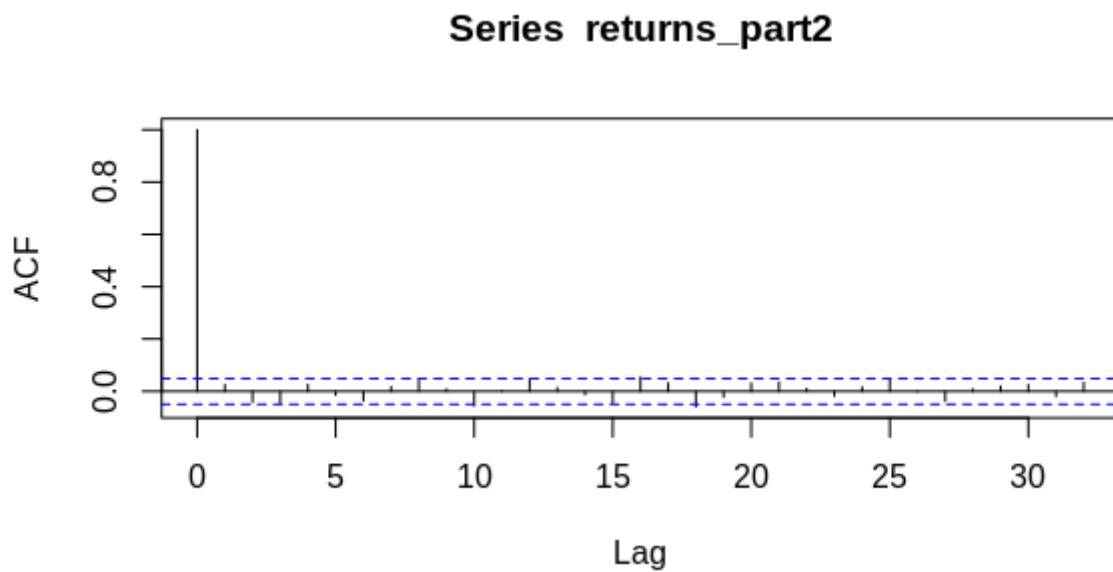
### Question 6.

For returns\_1 :



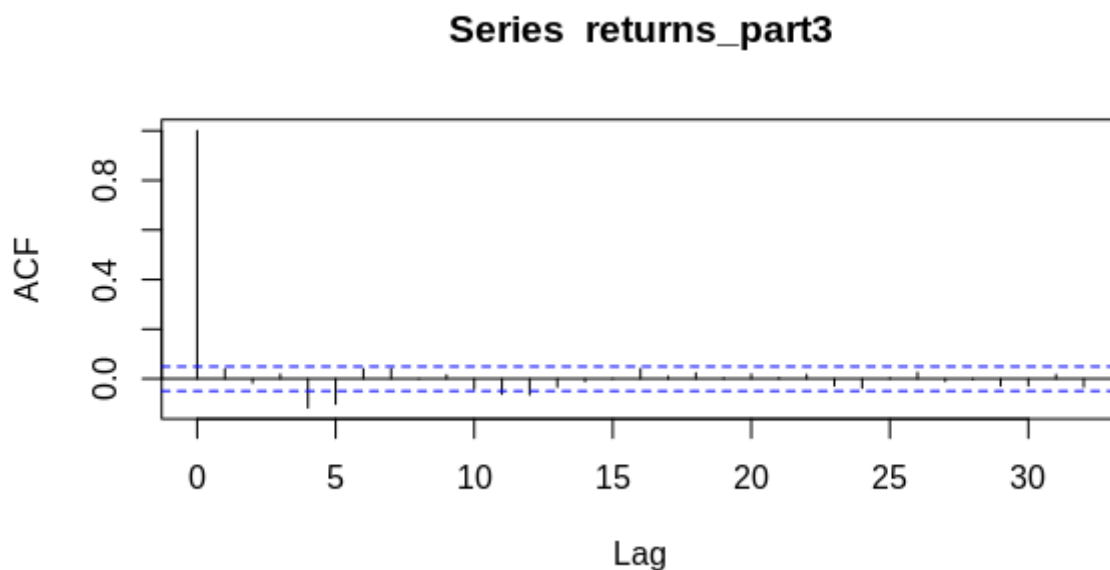
It seems that the only strong auto-correlation there is, is for 1 and 15.

For returns\_2:



There is no strong auto correlation

For returns\_3:



There is no strong auto-correlation as well.

For all the plots, I did not find an explanation why at lag 0 there is always a very strong correlation. This seems to be connected to R, or my code, rather than to the data. Returns\_2 and returns\_3 seem to be more close to each other, than returns\_1.

## Question 7.

```
[ reached getOption("max.print") -- omitted 3626 entries ]
> sprintf("Pearson result:%f", cor(result$dataset1, result$dataset1, use="everything", method="pearson"))
[1] "Pearson result:1.000000"
>
```

The result means we have a total positive correlation. If some values on NYA dataset increase, the values on my dataset(BFX) also tend to increase.

The time zone should play a roll, since the values of the stocks are always more advanced in time for my dataset(since it is in Belgium, and the other one is in New York). This means some value X in Belgium at some t1, will be the same at t1-4hours in New York. I cannot very well understand the relation between both sets BFX(my set) and NYA since they are different stocks. If they were the same, my explanations above would apply, and we would also have a total positive correlation.

## References

[0] : <https://stackoverflow.com/questions/3418128/how-to-convert-a-factor-to-integer-numeric-without-loss-of-information>