# Moneytalks - Report

Adriano Augusto - Grace Achenyo Okolo

## STEP1: Data pre-processing

First, we merge all the historical data of the different stocks into one unique dataset. This is meant to reduce the bias of the findings, since we want to analyse the entire NASDAQ100 and not single stocks.

```
library(data.table)
library(dplyr)
library(tidyr)
library(ggplot2)
library(pdist)
library(caret)
```

```
files <- list.files(pattern="*.csv")
data <- rbindlist(lapply(files, fread))
```

We rename some of the features, to ease their usage.

```
data <- rename(data, date = "Date", price = "Price", open = "Open", high = "High",
               low = "Low", volume =  "Vol.", d_change = "Change %", stock = "Stock",
               o_change = "O-Change", w_change = "W-Change", m_change = "M-Change",
               day = "Day", month = "Month", year = "Year")
```

We delete the rows containing missing values.

```
data <- data[!grepl("#DIV/0!", data$m_change),]
data <- data[!grepl("#REF!", data$m_change),]
data <- data[!grepl("#VALUE!", data$m_change),]
data <- data[,-1]
```

Finally, we export (only for the first execution) and (re)load the final dataset.

```
# data <-data %>% drop_na()
# ac <- write.csv(data, file = "dataset.csv")
# ac
data <- read.csv("dataset.csv")
```

## STEP2: Data exploration

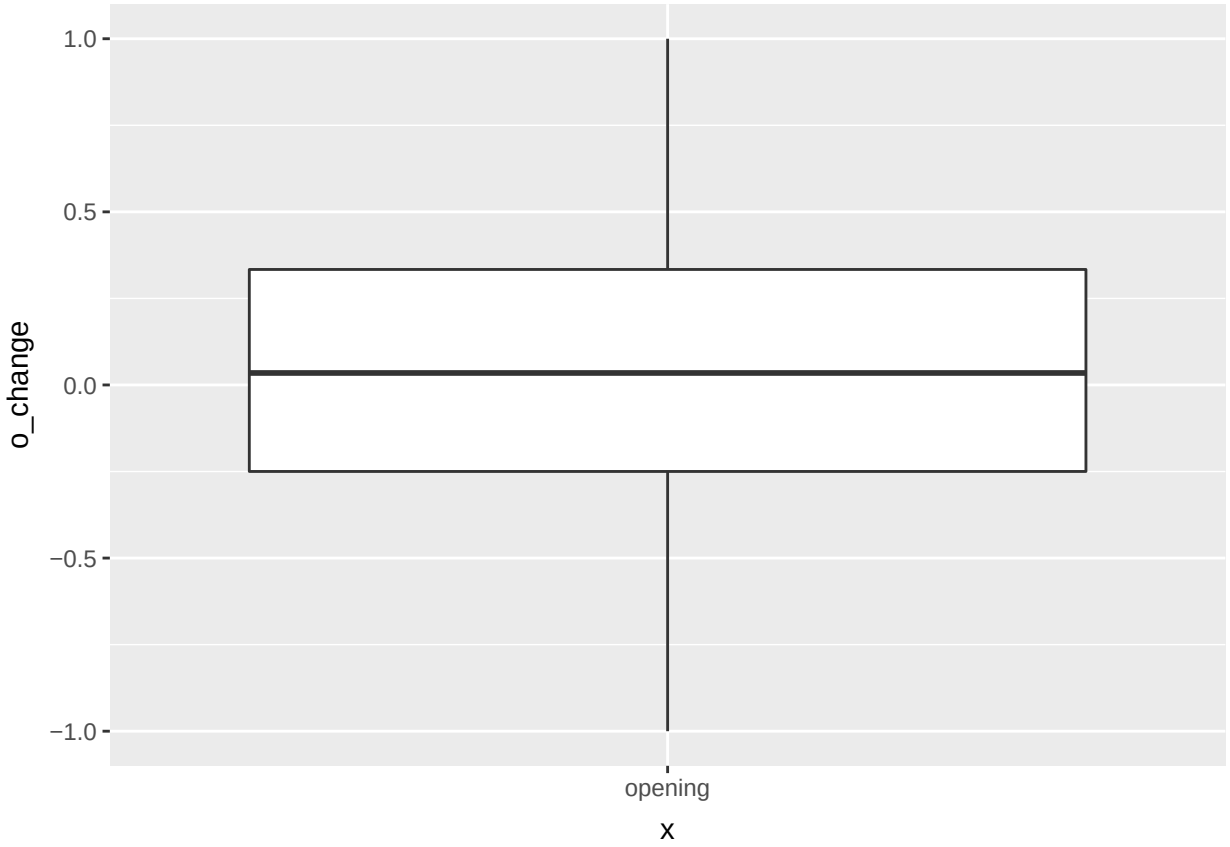
In this step, we focus on the most important features: the opening change, the daily change, the weekly change and the monthly change. For each of these features we report their ranges, box-plots and density plots. To note, that we defined some limits for the x-axis and y-axis in the plots. These limits were chosen taking into account the values of the 1st Qu. and 3rd Qu. of each value.

### Opening Change

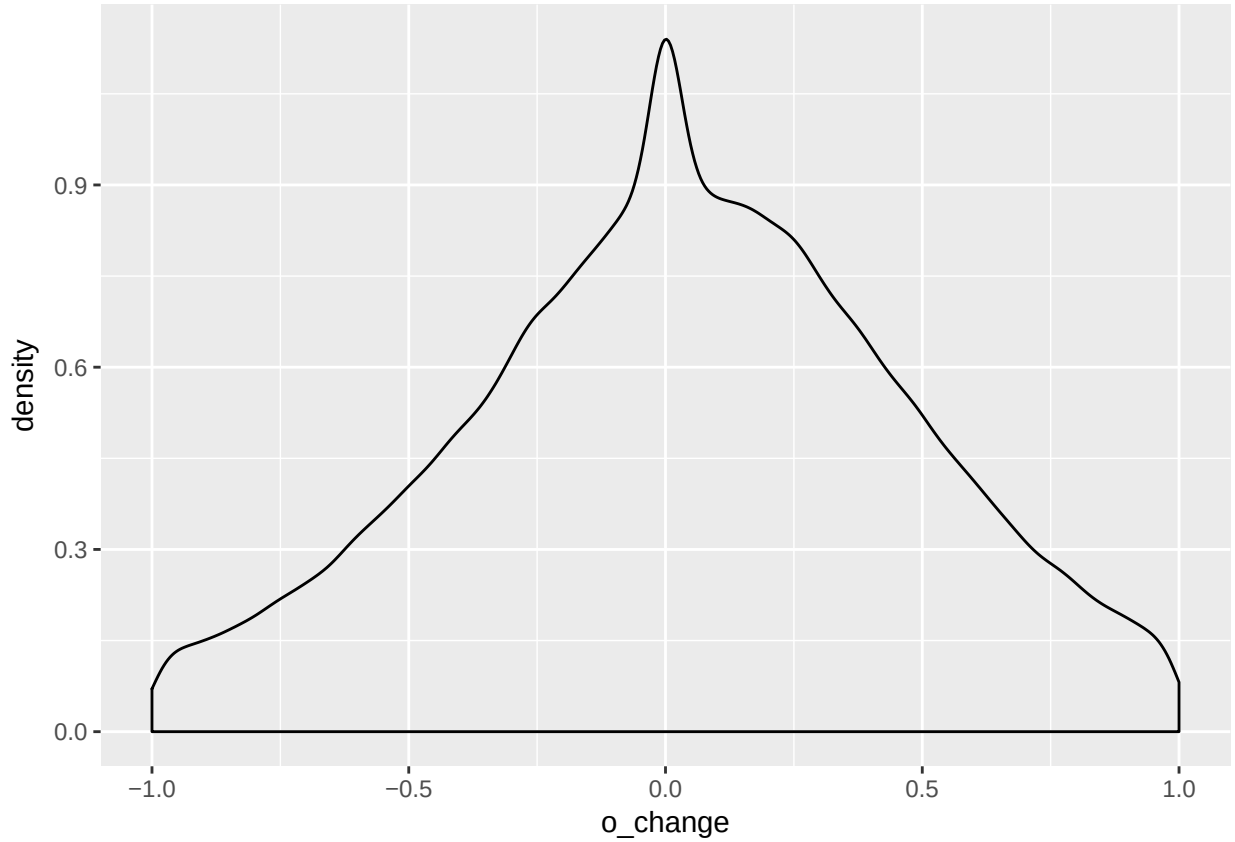
```
summary(data$o_change)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -66.71177  -0.30068   0.03946   0.04983   0.39901 199.55065
```

```
ggplot(data, aes(x = "opening", y = o_change)) + geom_boxplot() + ylim(-1, 1)
```




```
ggplot(data, aes(x = o_change)) + geom_density(alpha = 0.5) + xlim(-1, 1)
```




The summary highlights the fact that stock price change at the opening is used to fluctuate around the 0.00% change. I.e., a positive opening is eventually followed by an opening change with similar magnitude and opposite in sign. However, the fact that the mean is overall positive is a hint that it is slightly more probable to have positive openings than negative. Also, considering the MIN and the MAX change, we can notice that the MAX change is three times the MIN change. This shows that positive sentiment can be higher in magnitude than negative sentiment.

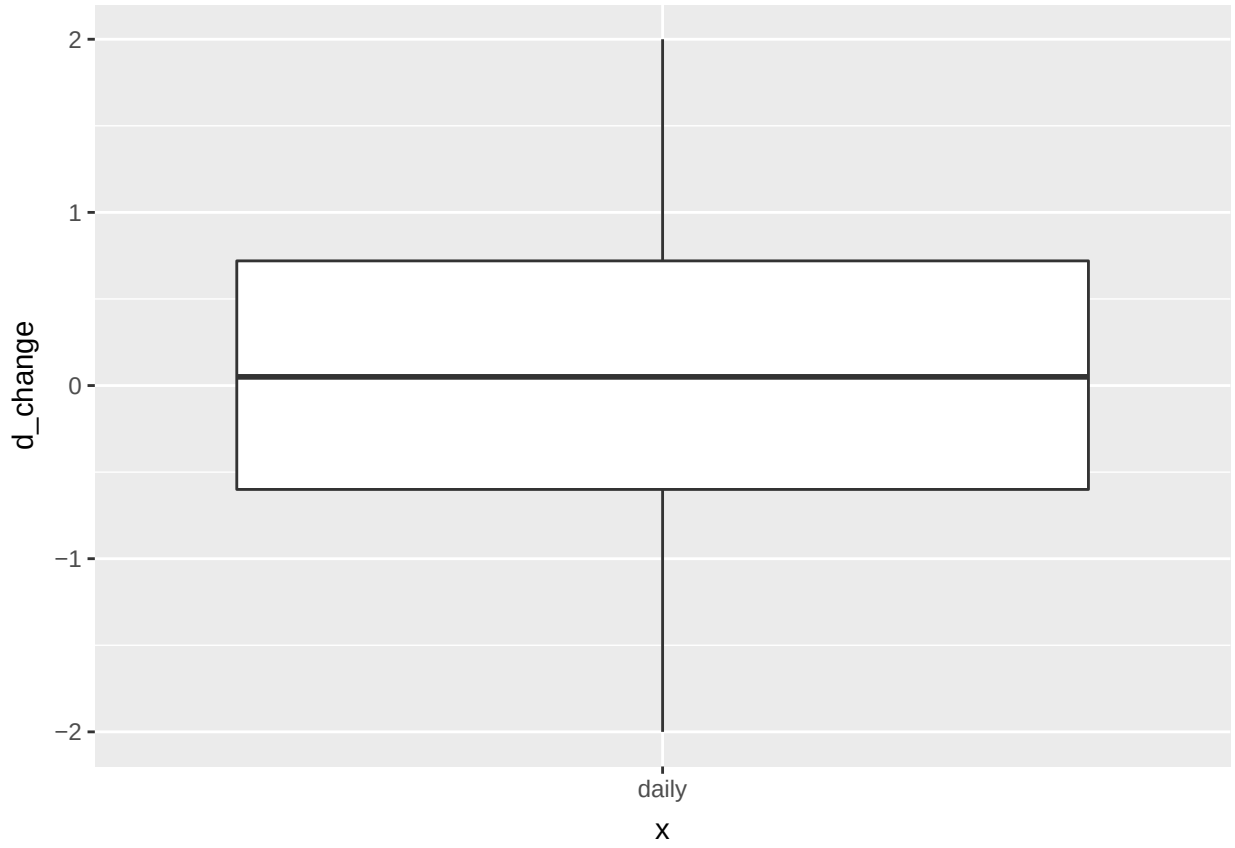
About the two plots, they show the distribution of the change of the stock price at the opening. We can see a slightly greater density for the positive values. This means that in the past five years stock prices have opened with a positive change more than the times they have opened with a negative change. This latter statements confirms the hint we had when looking at the summary of the feature.

**Daily Change**

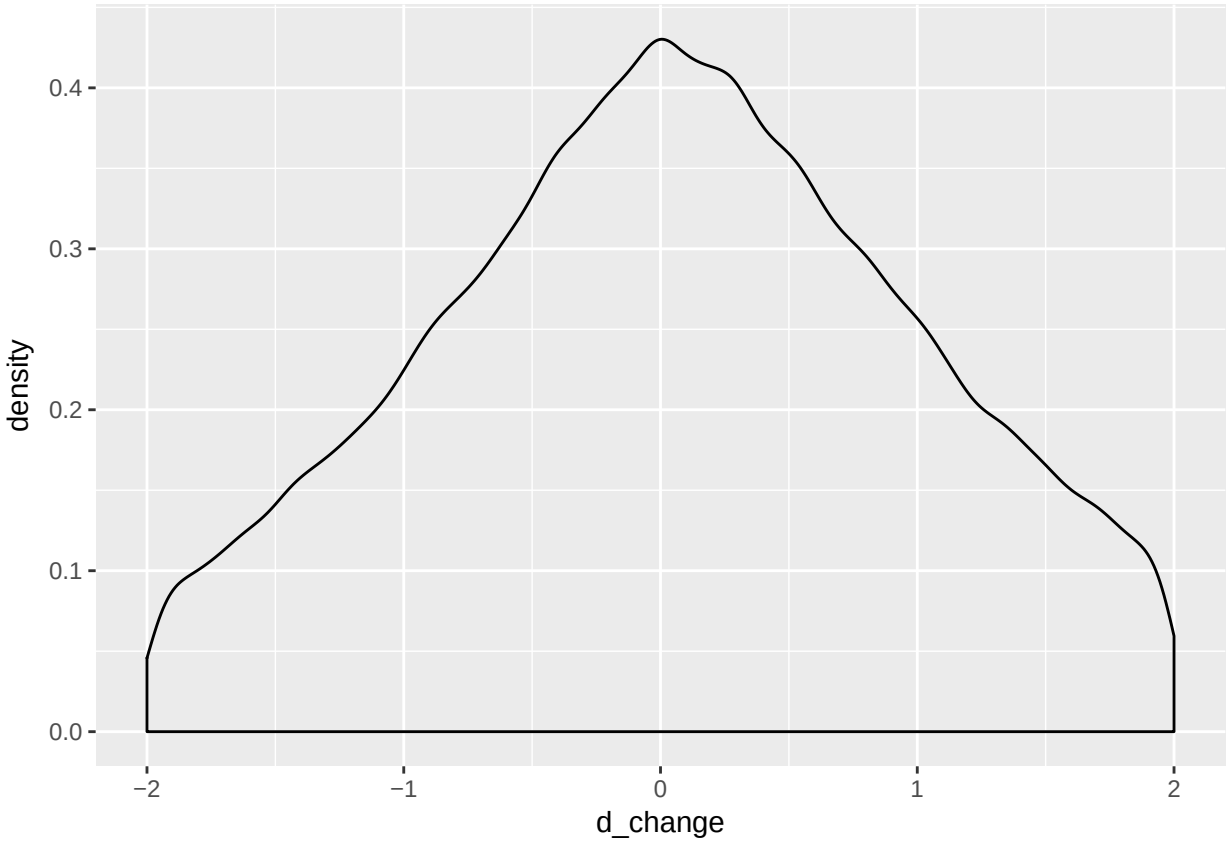
```
summary(data$d_change)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -66.88000  -0.76000   0.07000   0.08807   0.92000 199.55000
```

```
ggplot(data, aes(x = "daily", y = d_change)) + geom_boxplot() + ylim(-2, 2)
```




```
ggplot(data, aes(x = d_change)) + geom_density(alpha = 0.5) + xlim(-2, 2)
```




Similarly to the previous case, also the summary of the values for the daily change highlights the fact that stock price changes at the closing are used to fluctuate around the 0.00% change. However, we can notice that the values in this case have a wider range. Indeed, the value of the 1st Qu. and the 3rd Qu. are greater in absolute value w.r.t. the previous feature (the opening change). This may be a hint that the opening change may then vary widely across the whole day.

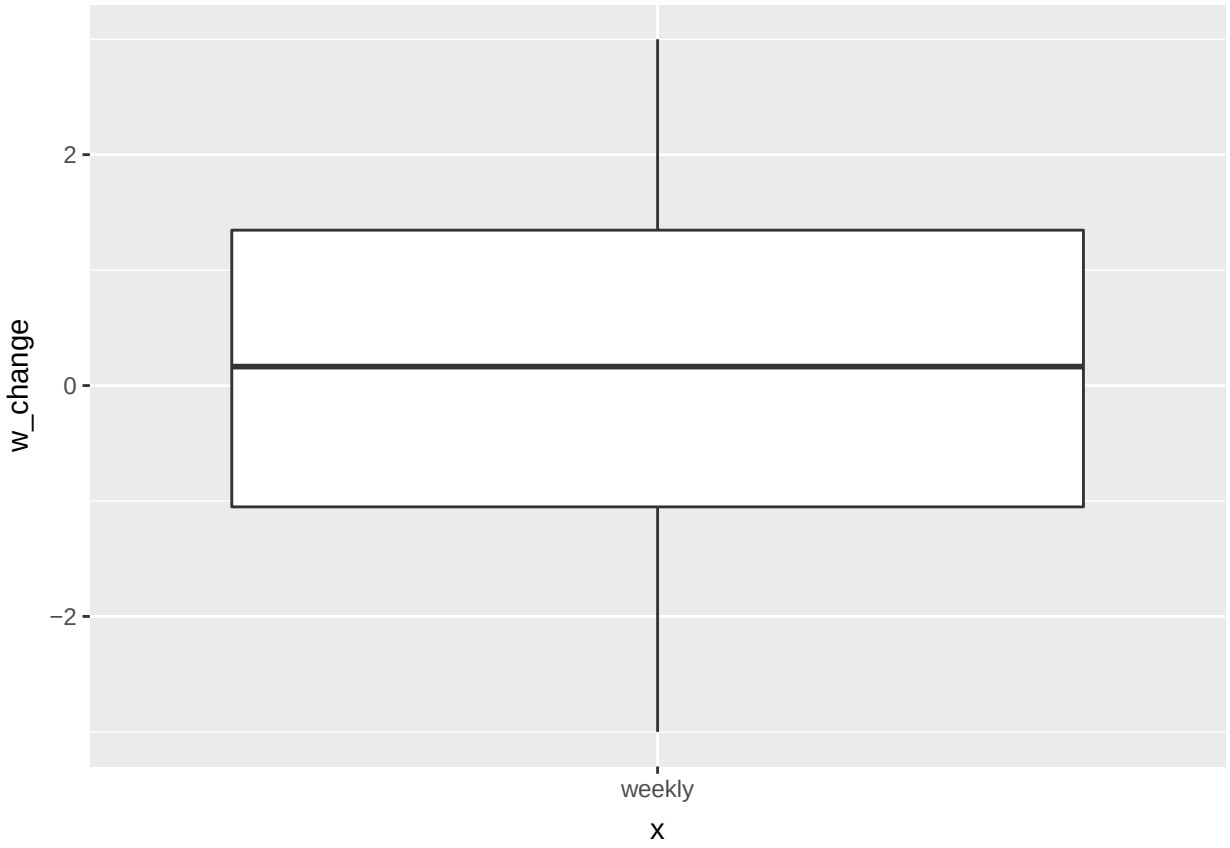
The two plots show the distribution of the change of the stock price at the closing. We noticed that the pattern observed for the opening change is present also in this second case. I.e. the positive values are slightly more than the negative ones.

**Weekly Change**

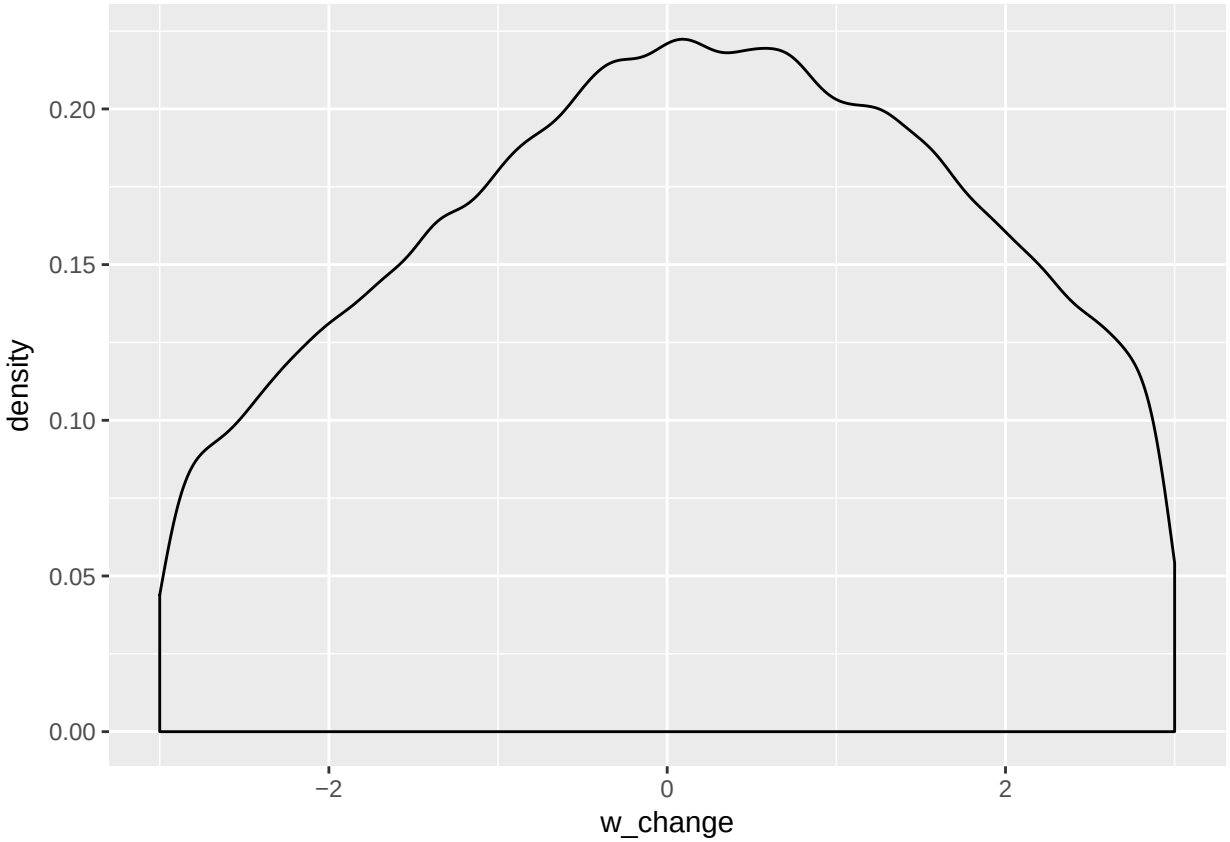
```
summary(data$w_change)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -69.7679  -1.4912   0.3336   0.3892   2.1922 208.2191
```

```
ggplot(data, aes(x = "weekly", y = w_change)) + geom_boxplot() + ylim(-3, 3)
```




```
ggplot(data, aes(x = w_change)) + geom_density(alpha = 0.5) + xlim(-3, 3)
```




The summary of the values of the weekly stock price changes clearly shows two important insights. First, over a longer timeframe, it is more probable that a stock price grows positively. This statement is supported by the following data: i) the mean keeps growing at the increasing of the time; ii) the number of positive values is greater than the number of negative values.

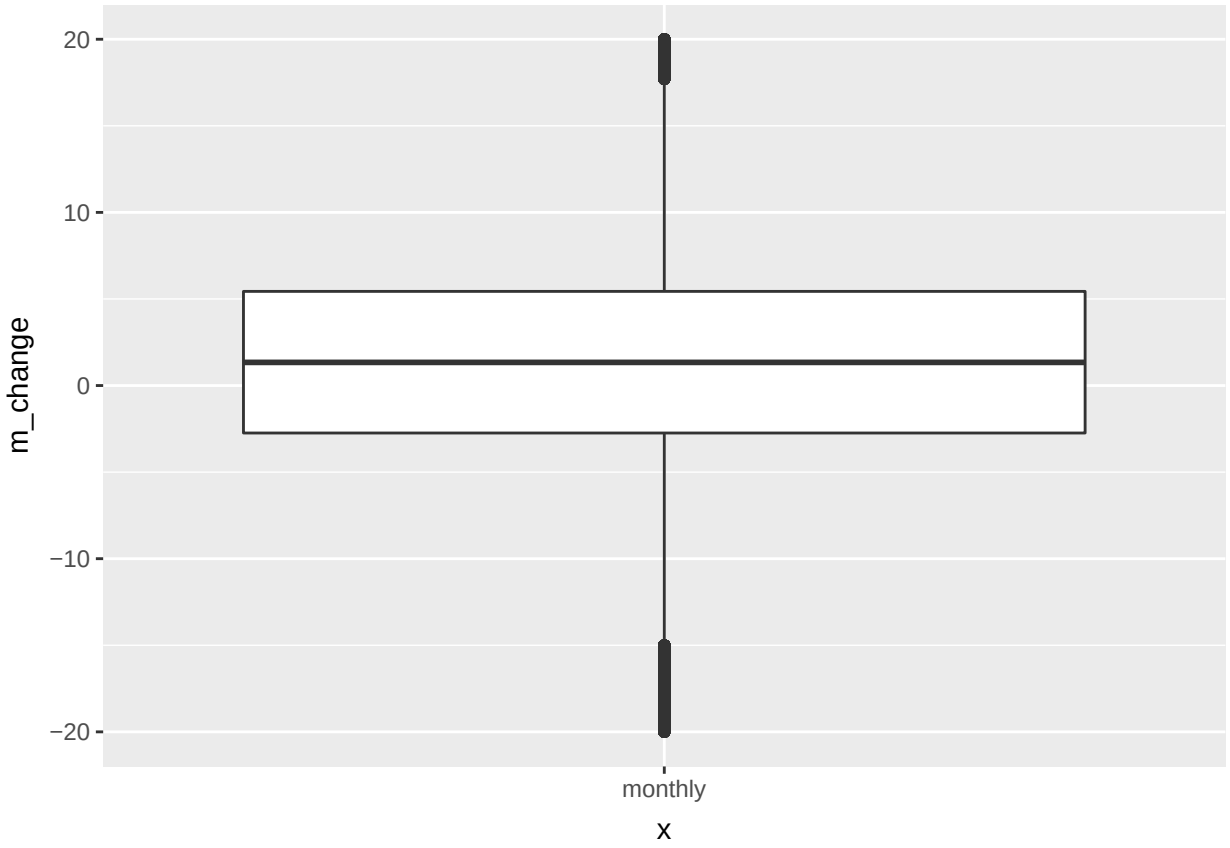
The two plots show the distribution of the weekly change of the stock price. We notice that the density for positive values is increasing.

**Monthly Change**

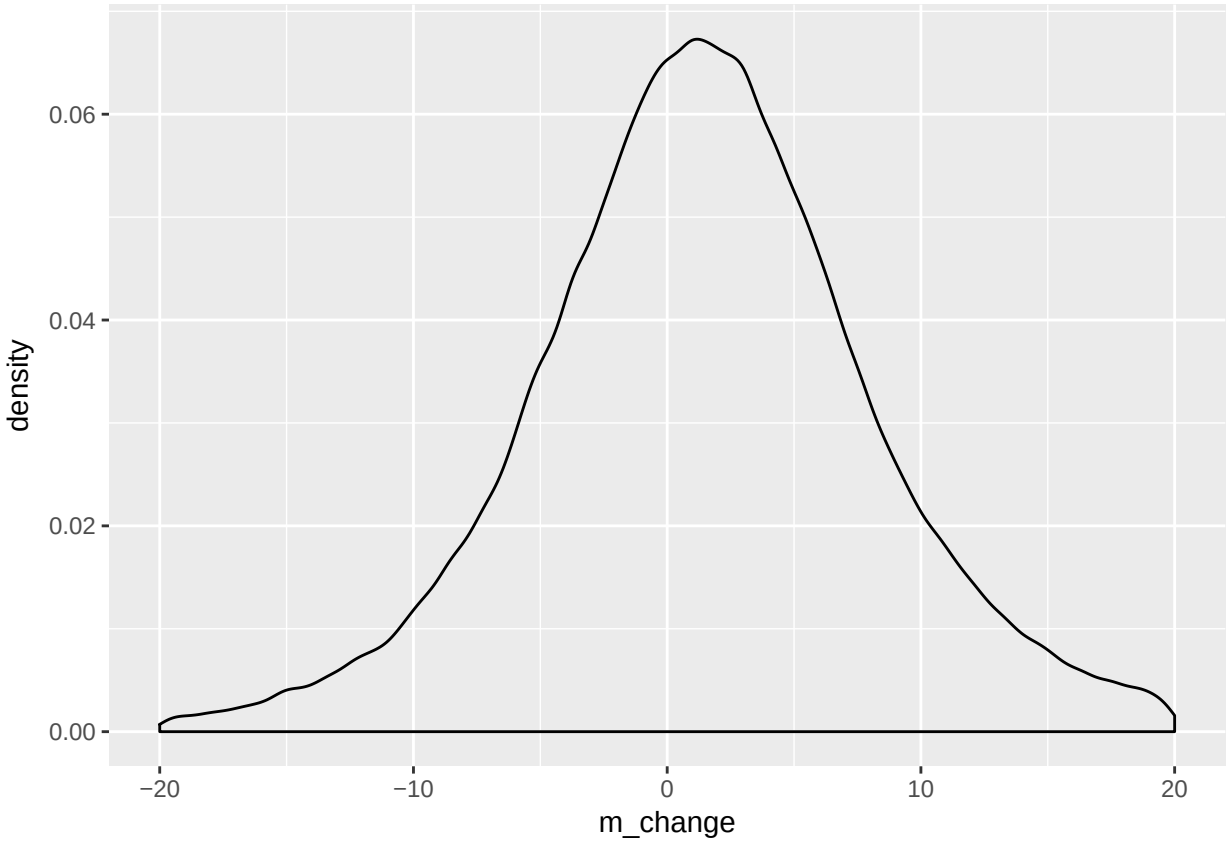
```
summary(data$m_change)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -72.269  -2.753   1.424   1.634   5.698 214.306
```

```
ggplot(data, aes(x = "monthly", y = m_change)) + geom_boxplot() + ylim(-20, 20)
```




```
ggplot(data, aes(x = m_change)) + geom_density(alpha = 0.5) + xlim(-20, 20)
```




Finally, increasing the timeframe to one month, and analysing the monthly price change of the stocks, we can confirm the previous statements. Overtime, the stock prices are more prone to increase, and therefore deliver a positive change. For the monthly change, we can see that the mean moved to the value of 1.6%, and the growth of the 3rd Qu. is much greater than the negative growth of the 1st Qu. This can be seen even better from the density and box plots. This time, the density on the positive side of the curve is much greater than the previous cases. Whilst, the boxplot moved upward centered on the value of 1.6% and clearly wider on the top part.