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)
```

```
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 =
```

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

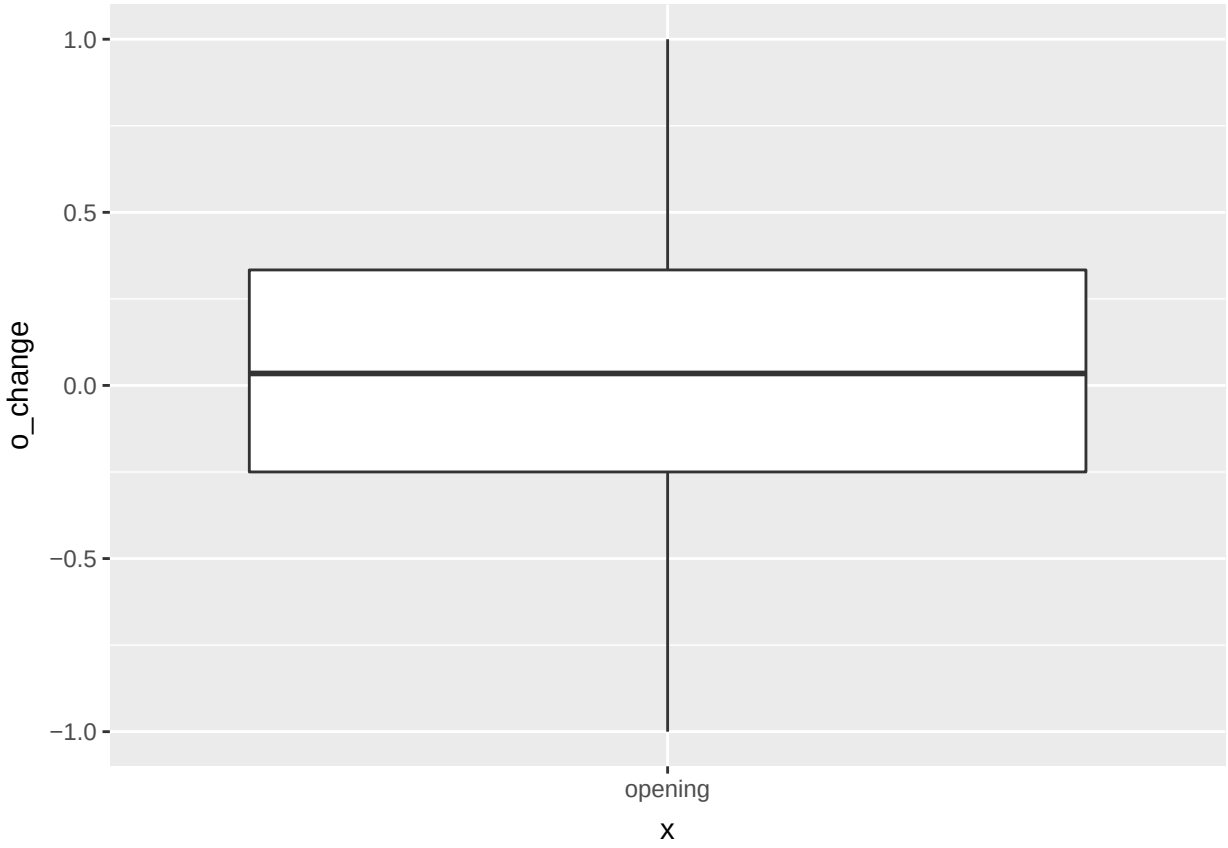
In this step, we focus on the most important features: the opening change, the daily change, the weekly change and the monthly change. For each of these features we report their ranges, box-plots and density plots. To note, that we defined some limits for the x-axis and y-axis in the plots. These limits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

Opening Change

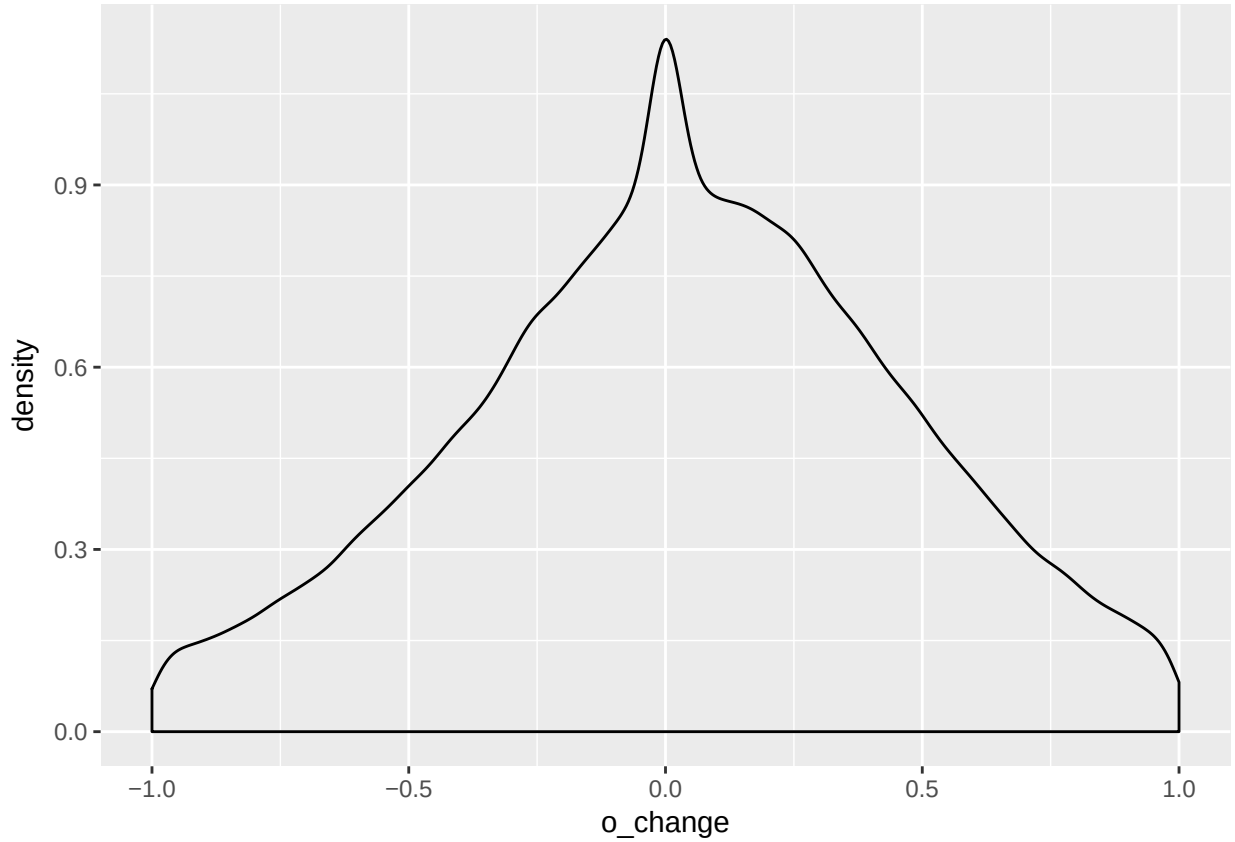
```
summary(data$o_change)
```

```
##       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.

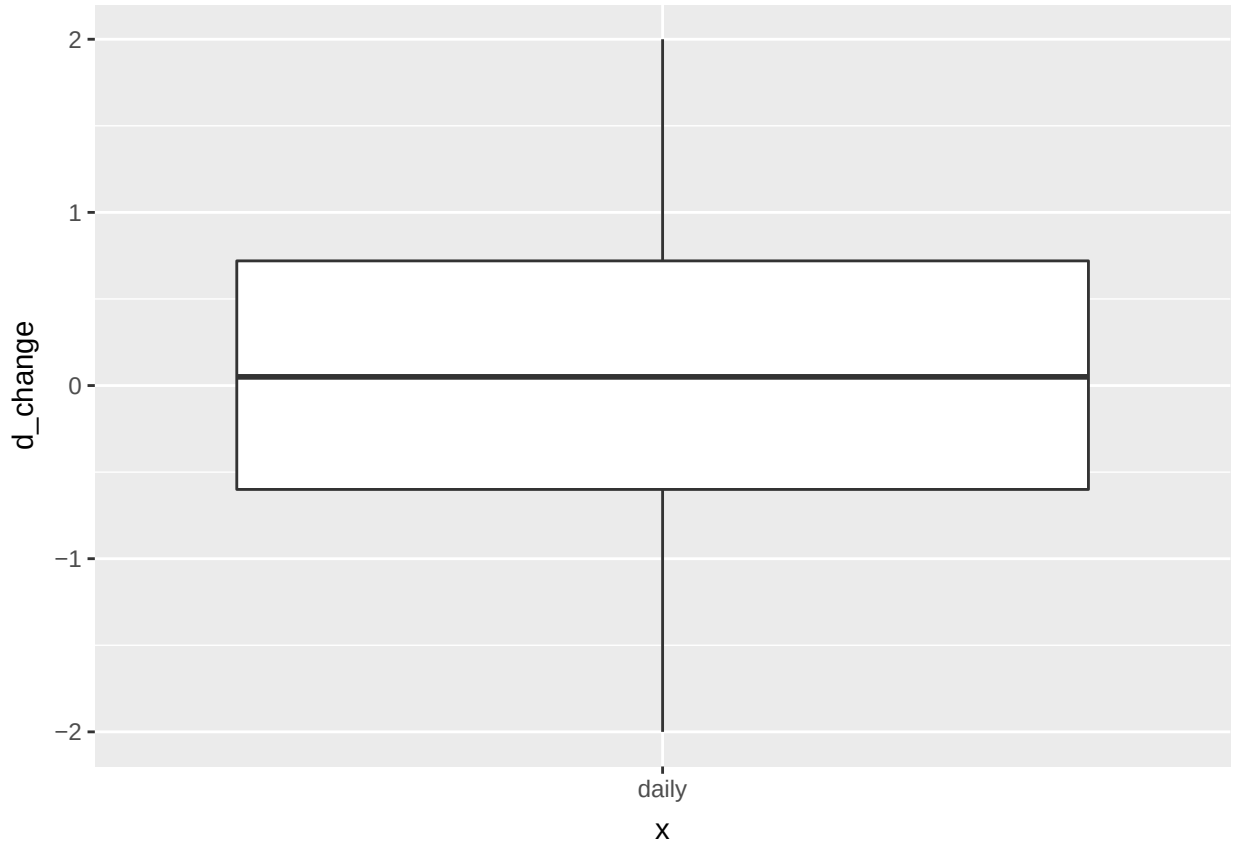
About the two plots, they show the distribution of the change of the stock price at the opening. We can see a slightly greater density for the positive values. This means that in the past five years stock prices have opened with a positive change more than the times they have opened with a negative change.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

Daily Change

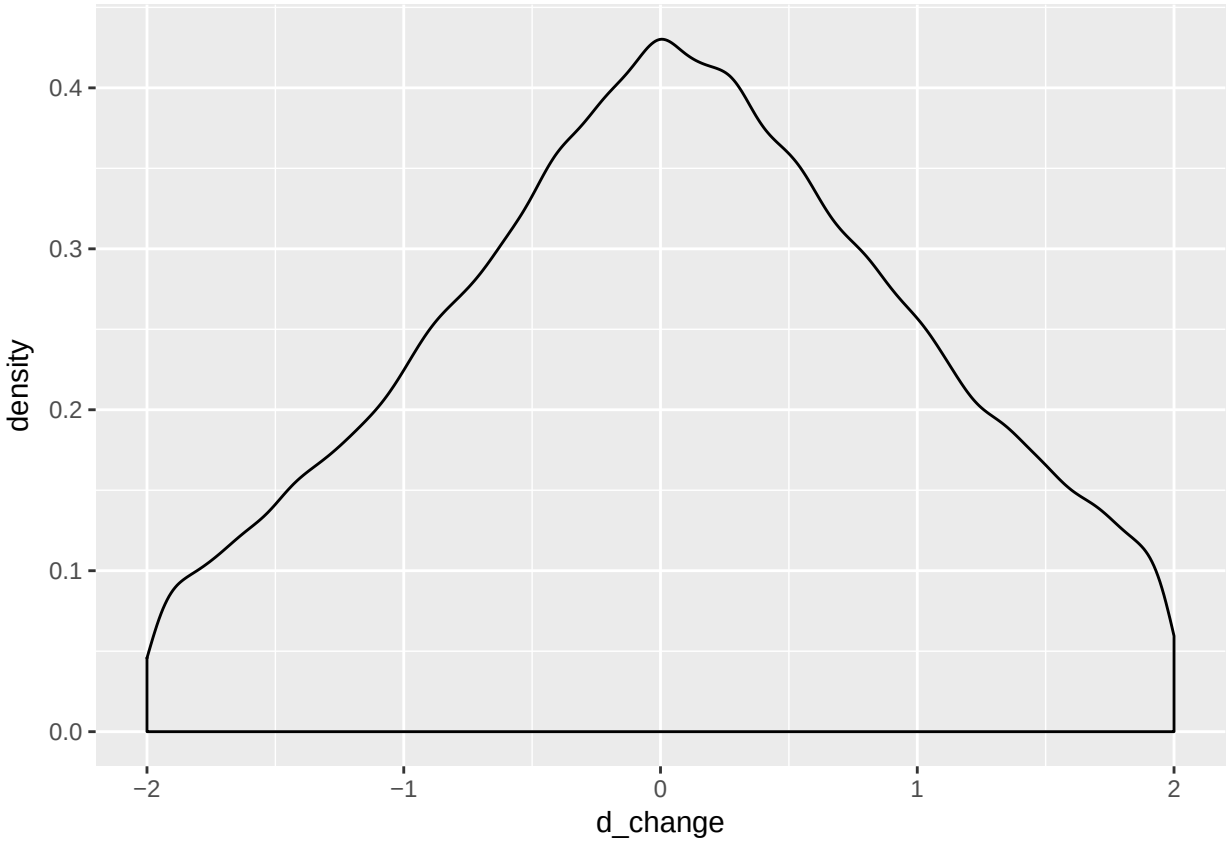
```
summary(data$d_change)
```

```
##      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.

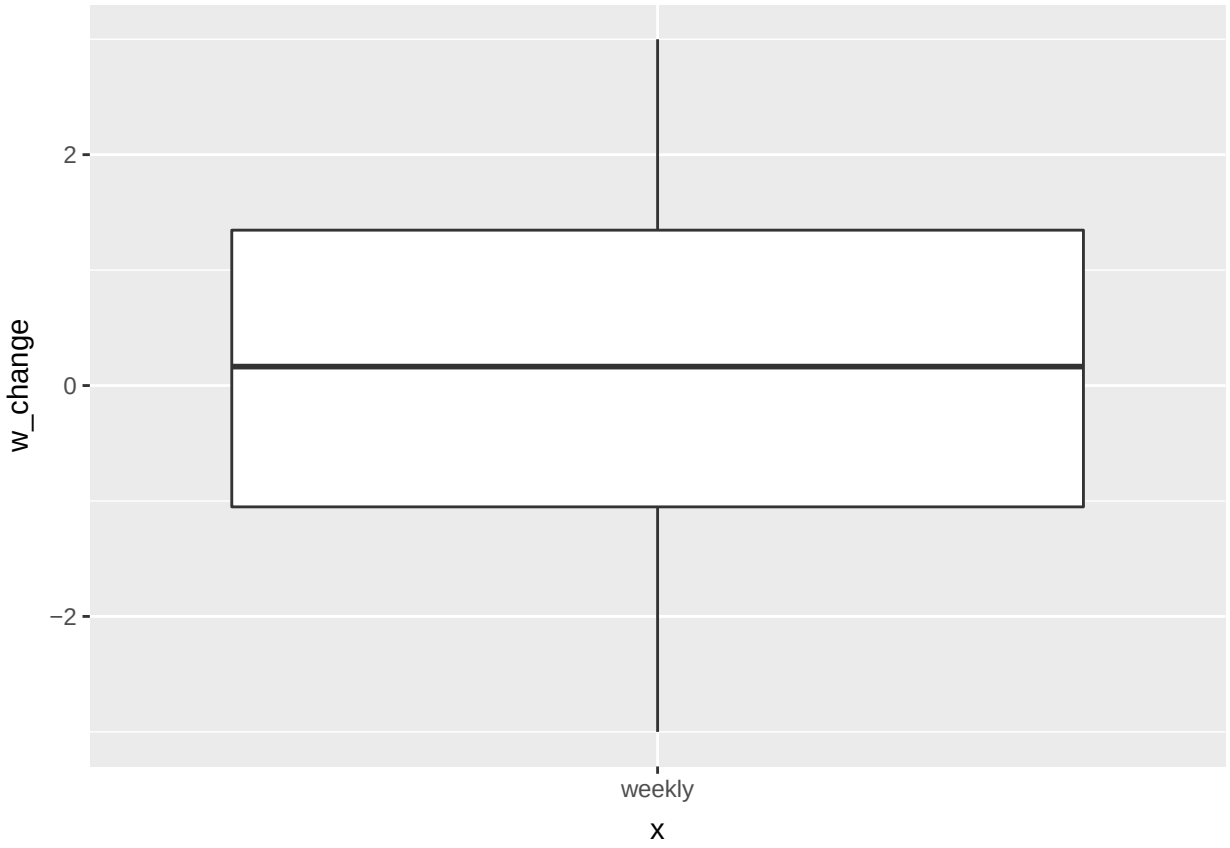
The two plots show the distribution of the change of the stock price at the closing. We noticed that the pattern observed for the opening change is present also in this second case. I.e. the positive values are slightly more than the negative ones.

Weekly Change

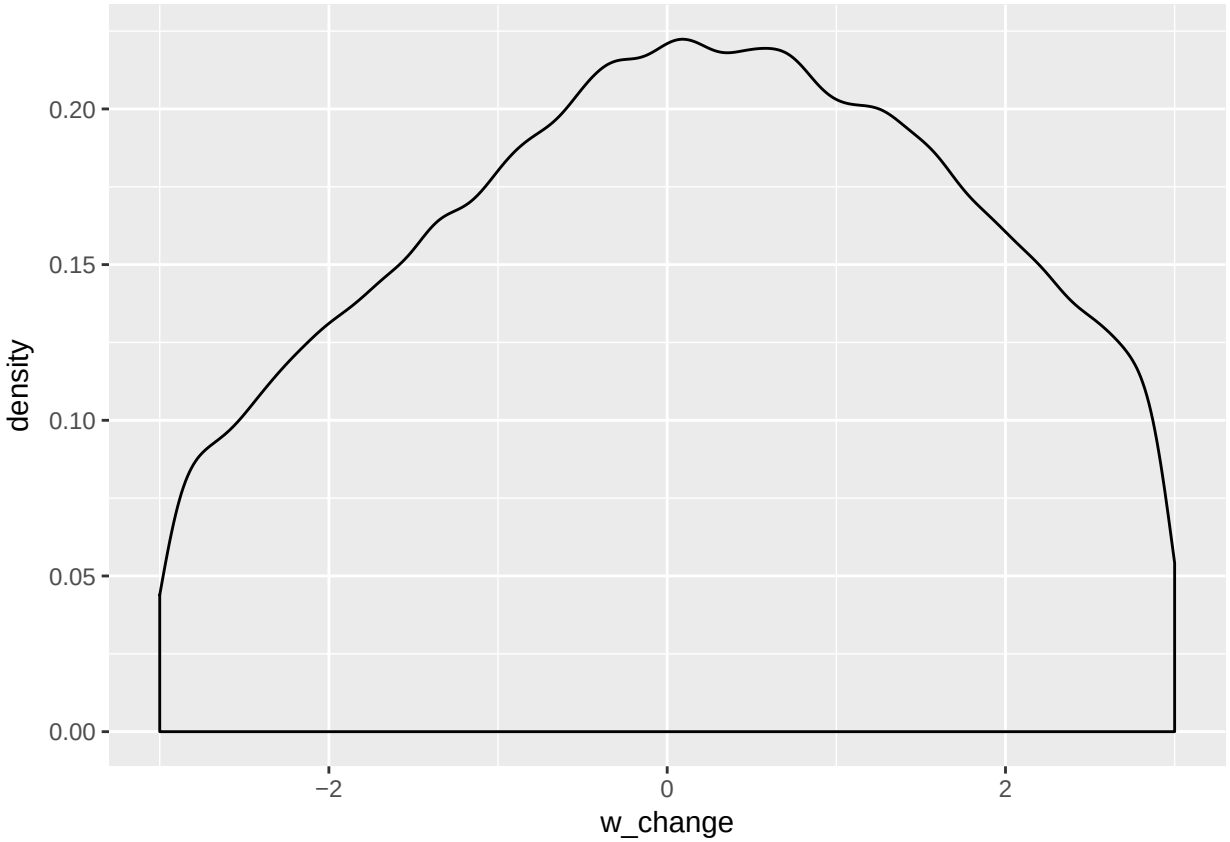
```
summary(data$w_change)
```

```
##     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.

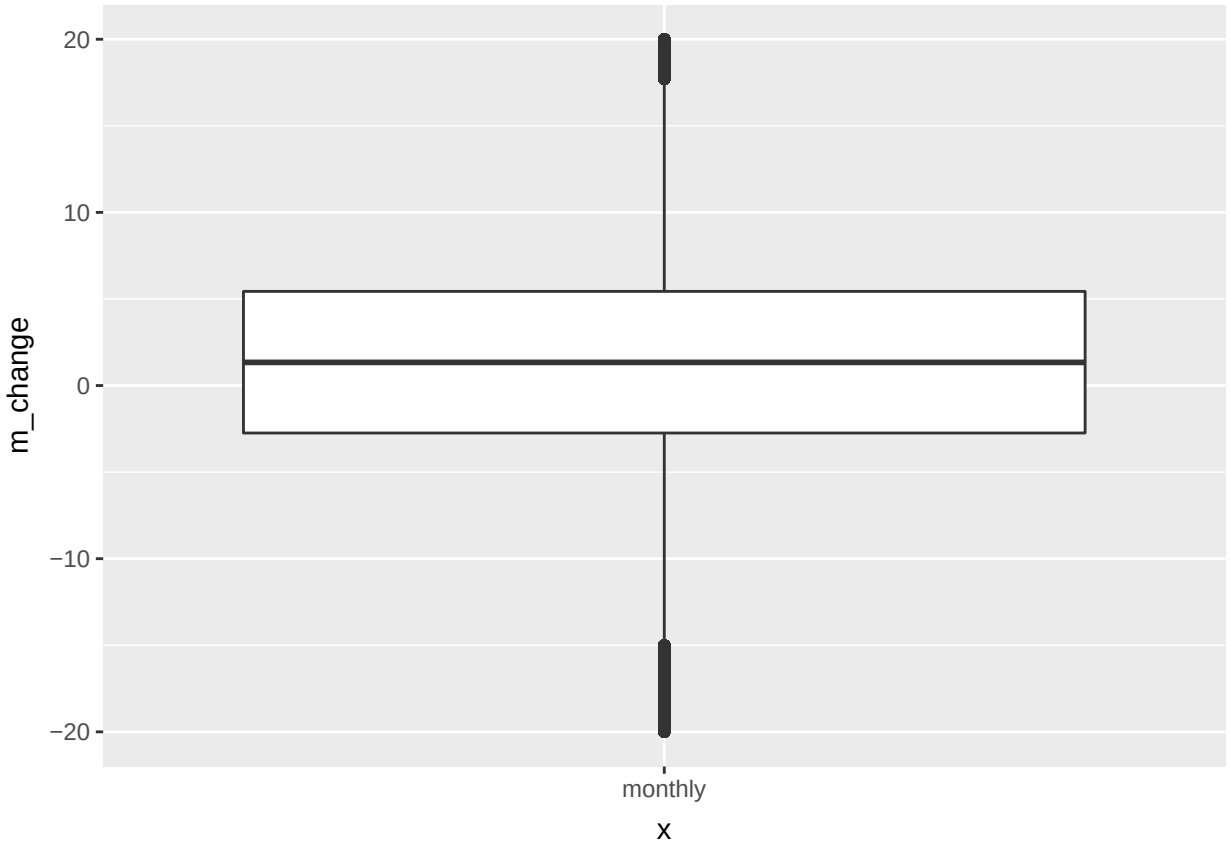
The two plots show the distribution of the weekly change of the stock price. We notice that the density for positive values is increasing.

Monthly Change

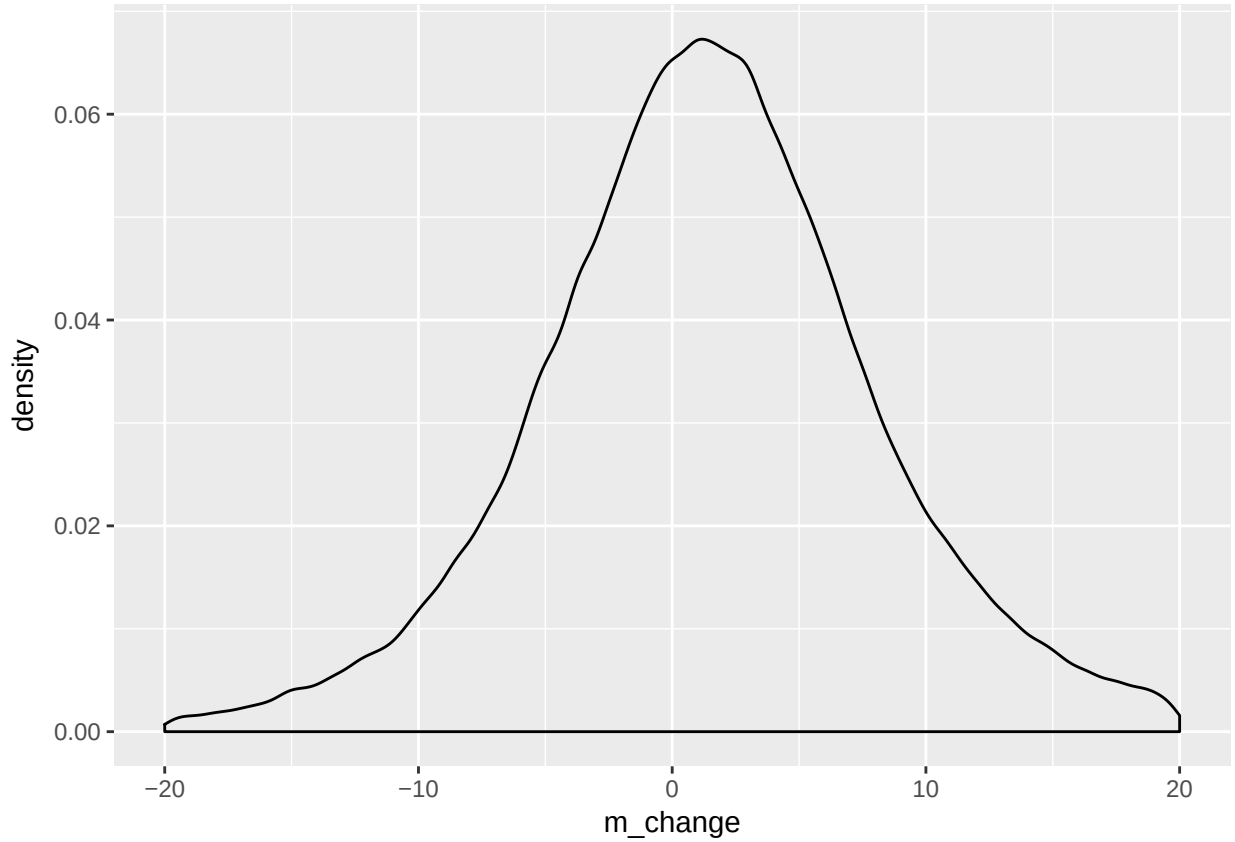
```
summary(data$m_change)
```

```
##    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.

STEP3 - Data analysis Vol.1

We decided to check whether there are months where most of the times a stock price closes in positive or negative. To do so, we counted the positive days (i.e. daily change > 0.5), and the negative days (i.e. daily change < -0.5), in each month. Then, we evaluated the difference, and we did group the counting by month and year and plot them. The values are normalized according to the number of stocks composing the NASDAQ100 (107).

Simultaneously, we did the same for the stock opening price. To do so, we counted the positive openings (i.e. opening change positive), and the negative openings (i.e. opening change negative), in each month. Then, we evaluated the difference, and we did group the counting by month and year and plot them. The values are normalized according to the number of stocks composing the NASDAQ100 (107).

```
dcount_pos <- function(df, value) {
  summarise(filter(df, d_change>value), n())
}
```

```r
dcount_neg <- function(df, value) {
  summarise(filter(df, d_change<value), n())
}

ocount_pos <- function(df, value) {
  summarise(filter(df, o_change>value), n())
}

ocount_neg <- function(df, value) {
  summarise(filter(df, o_change<value), n())
}
```

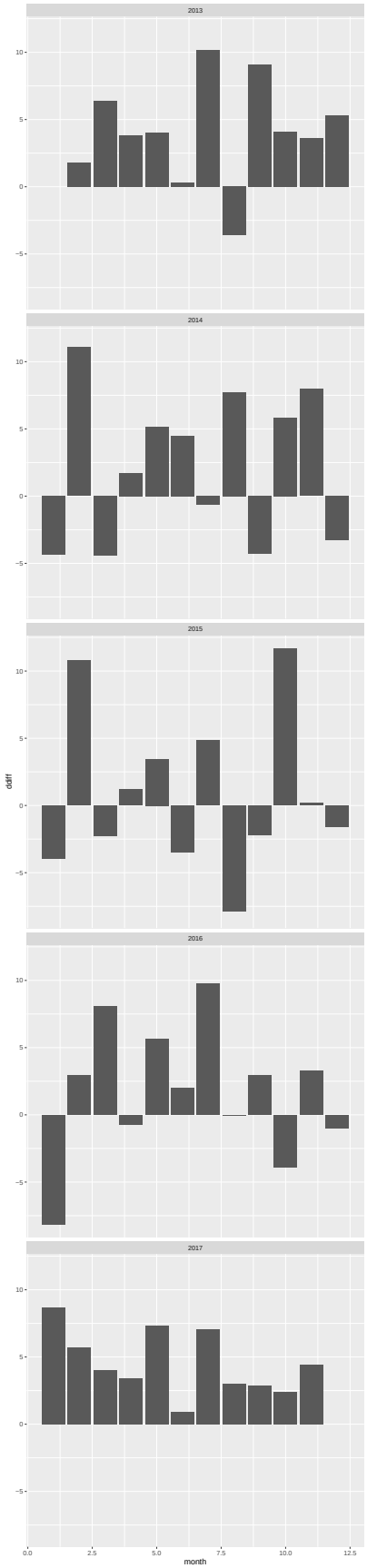
```r
years <- c(2013, 2014, 2015, 2016, 2017)
months <- c(1, 2, 3, 4, 5,6, 7, 8, 9, 10, 11, 12)
```

```r
iterations = 60
features = 6

changes <- matrix(ncol=features, nrow=iterations)

c <- 0
 for(y in years) {
  for(m in months) {

   monthly_data  <- filter(data, year == y & month == m)

   dmonthly_tot_pos <- dcount_pos(monthly_data, 0.5)[1,]
   dmonthly_tot_neg <- dcount_neg(monthly_data, -0.5)[1,]

   omonthly_tot_pos <- ocount_pos(monthly_data, 0.5)[1,]
   omonthly_tot_neg <- ocount_neg(monthly_data, -0.5)[1,]

   changes[c,1] <- y
   changes[c,2] <- m
   changes[c,3] <- dmonthly_tot_pos/107
   changes[c,4] <- dmonthly_tot_neg/107
   changes[c,5] <- omonthly_tot_pos/107
   changes[c,6] <- omonthly_tot_neg/107
   c = c+1
  }
 }

changes <- data.frame(changes)
changes <- rename(changes, year = X1, month = X2, dtot_pos = X3, dtot_neg = X4, otot_pos = X5, otot_neg
changes <- filter(changes, dtot_pos != 0 | dtot_neg != 0)
```

```r
changes <- transform(changes, ddiff= dtot_pos-dtot_neg)
changes <- transform(changes, odiff= otot_pos-otot_neg)
```

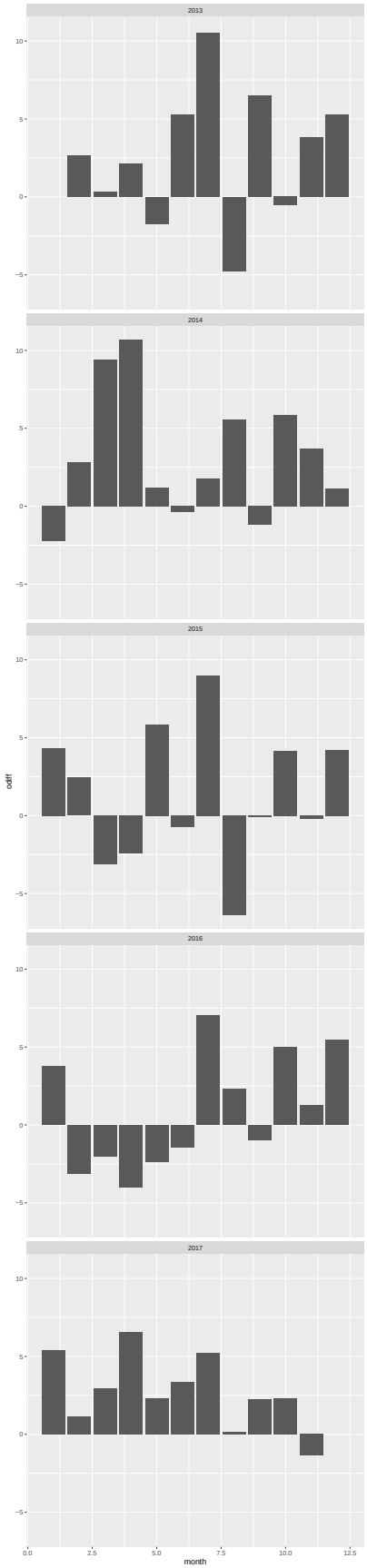
```r
#summary(changes$dtot_pos)
#summary(changes$dtot_neg)
#summary(changes$otot_pos)
#summary(changes$otot_neg)
```

```
ggplot(changes, aes(x=month, y=ddiff)) + geom_bar(stat = "identity") + facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
ddiff
month
10
5
0
−5
0.0
2.5
5.0
7.5
10.0
12.5

This first series of plots show how many days the stock price closed in positive (>0.5) for a specific month, w.r.t. the number of days the stock price closed in negative (<-0.5). A positive value X means that, for that month, there were X-days positive more than the negative days. What we can see from these plots is that, overall, there are more positive days than negative. Also, we can see that months having many positive days, are eventually followed by month with more negative days, and vice-versa. This entails that the stock prices movements follow a cycle that repeats itself. An example is the plot for year 2017, where first half and second half shows the same pattern. Also, the month of February was the only one able to have across the all 5 years an excess of positive days over the negative ones. Although this may be just casual, it may also be related to the fact that at the end of January, companies report the financial results of their first quarter financial year.

```
ggplot(changes, aes(x=month, y=odiff)) + geom_bar(stat = "identity") +  facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
odff
month
10
5
0
−5
0.0
2.5
5.0
7.5
10.0
12.5

This second series of plots report a similar analysis of the previous, however this time we focused on the opening price change instead of the closing price. The results show a very interesting fact, when compared with the results of the previous plots. Indeed, the number of positive days reduce overall. A clear example is the year 2016, first semester. Considering the closing price, there were more positive days than negative. Instead, when considering the opening price, there were more negative days than positive. This means that most of the times for those months a stock price opening with a negative change, turned the negative change into positive over the day. This is a first step toward the analysis of the relation between the opening price change and the closing price change. From this results, we can see that if there is a correlation it may be negative.

STEP4 - Data analysis Vol.2

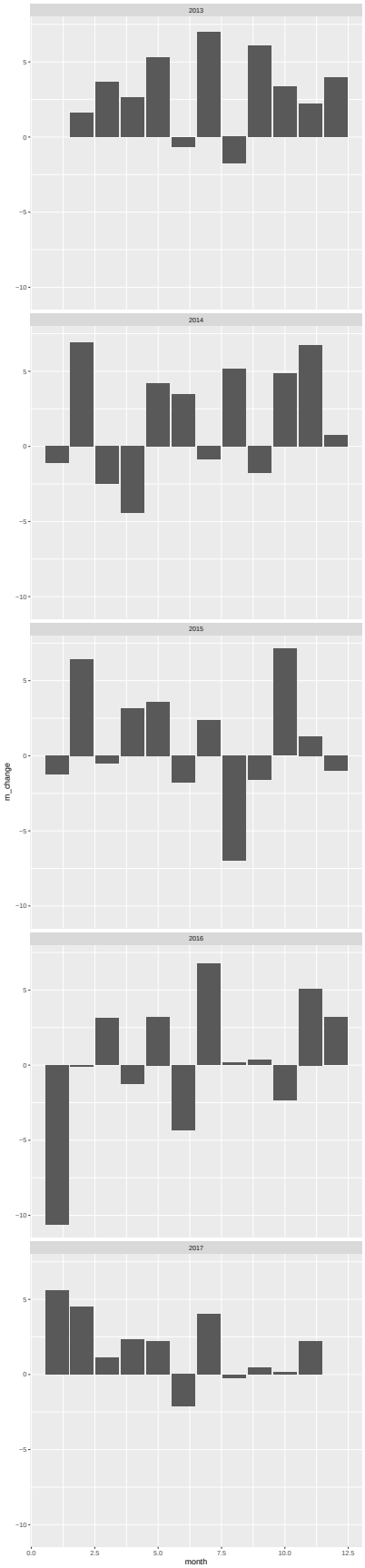
Successively, we decided to analyse what is the average monthly return for the NASDAQ100. First, we considered the monthly return of each month per year per each listed company in the NASDAQ100, and we evaluated the average. Then, we plot such average and we analysed the results, as well the summary of the values.

```
iterations = 60
features = 3

changes <- matrix(ncol=features, nrow=iterations)

c <- 0
 for(y in years) {
  for(m in months) {
   monthly_data  <- filter(data, year == y & month == m)

   i <- which.max(monthly_data$day)
   if(length(monthly_data[i,'m_change']) == 0) break

   d <- monthly_data[i,'day']

   monthly_data <- filter(monthly_data, day == d)

   changes[c,1] <- y
   changes[c,2] <- m
   changes[c,3] <- as.numeric(summarise(monthly_data, avg_mchange = mean(m_change)))
   c <-  c+1
  }
 }

changes <- data.frame(changes)
changes <- rename(changes, year = X1, month = X2, m_change = X3)
changes <- filter(changes, m_change != 0 )
```

```
summary(changes$m_change)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -10.5870  -0.9833   2.2114   1.5380   4.0028   7.1289
```

```
ggplot(changes, aes(x=month, y=m_change)) + geom_bar(stat = "identity") +  facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
m_change
month
5
0
−5
−10
0.0
2.5
5.0
7.5
10.0
12.5

The summary shows that the average return over the last 5 years has been 1.5% monthly. Despite some outliers (e.g. January 2016), we can say that across the last 5 years the NASDAQ100 had positive returns the majority of the month. Also, positive returns are often followed by negative returns.

STEP4 - Advices for investors

From the findings we got so far we designed four questions about stock market and stock prices and we did answer. The answers are backed by the data analysis did so far. The suggestions, despite simple, are fundamental for any investor.

1. Should I invest? Yes, everyone should invest in the stock market, i.e. everyone should be an investor. This statement is backed by the fact that the company listed in the NASDAQ100 showed to be solid and have positive returns across the 5 years, with an average of 1.5% monthly. Such return is way greater than any saving account offered by any bank in the world.
2. Short or long time investments? Buy and hold strategies are the most valuable, since the stocks' prices tend to grow over long time. The data exploration showed that the longer is the time-frame, the greater is the growth of the price stock. Despite there is no hint about how the prices change daily or weekly, it was clear that on a long run, the price tent to go up. This is of course strictly related to the fact that the company selected from the NASDAQ100 index are stable and solid, with strong financial setting.
3. When to buy? We saw in different graphs that despite it is not possible to forecast the magnitude of the price change, we can see patterns of the type "up-down-up-down". In other words, the best moment to buy is within negative periods. This does not mean that there is a warranty of the stock price to go up again quickly, however, the chances are higher that it could happen.
4. When to sell? The later the better. This question and its answer are strictly related to the type of investment strategy. Although the best rational strategy is to buy and hold a stock, it is clear that having knowledge of when a price could go up and when down, the returns would be much more profitable. However, the analysis of the data did so far showed that there is no way to forecast price movements, nor direction nor magnitude. Indeed, no specific patterns were found in the prices' movements.

Driving by the fact that having the change to forecast the near price movements would highly increase the profit of investments, we will try to analyse if any correlation holds between the price change at the opening, an the price change at the closure in terms of percetage.

STEP5 - Relation between Opening and Closing Price Change (%)

We first select the two column for opening change and closing change in terms of percentage, and we evaluate the correlation.

```
o <- data$o_change
d <- data$d_change

od_corr <- cor(o, d, use = "everything", method = c("pearson", "kendall", "spearman"))
od_corr
```

```
## [1] 0.6645693
```

The correlation between the two features is 0.66. This indicates there is crearly a positive correlation, that makes sense to investigate. Therefore, we evaluate the average distance between the opening and the closing change in %.

```
sum <- 0

for(i in 1:length(o)) {
  sum <- sum + sqrt((o[i] - d[i])^2)
}

avgd <- sum/length(o)
avgd
```

```
## [1] 1.056221
```

The average distance between the two value is 1.06. This is meaningful, since it shows that on average, the opening price gives and hint of the magnitude of the change at closing.

To further analyse this data, we check how many times the opening change was greater than the closing and viceversa.

```
growing <- 0
declining <- 0

for(i in 1:length(o)) {
  if( o[i] < d[i] ) growing <- growing + 1
  else declining <- declining + 1
}

growing
```

```
## [1] 200793
```

```
declining
```

```
## [1] 188718
```