RV

Collaborative project

4/9/2021

```
library(tidyquant)
```

```
## Loading required package: lubridate

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## Loading required package: PerformanceAnalytics

## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend

## Loading required package: quantmod

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## == Need to Learn tidyquant? ====================================================
## Business Science offers a 1-hour course - Learning Lab #9: Performance Analysis & Portfolio Optimiza
## </> Learn more at: https://university.business-science.io/p/learning-labs-pro </>
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
```

```
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x dplyr::first()           masks xts::first()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x dplyr::last()            masks xts::last()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
head(FANG)
```

```
## # A tibble: 6 x 8
##   symbol date        open  high   low close    volume adjusted
##   <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>    <dbl>
## 1 FB     2013-01-02  27.4  28.2  27.4  28    69846400     28
## 2 FB     2013-01-03  27.9  28.5  27.6  27.8  63140600     27.8
## 3 FB     2013-01-04  28.0  28.9  27.8  28.8  72715400     28.8
## 4 FB     2013-01-07  28.7  29.8  28.6  29.4  83781800     29.4
## 5 FB     2013-01-08  29.5  29.6  28.9  29.1  45871300     29.1
## 6 FB     2013-01-09  29.7  30.6  29.5  30.6 104787700     30.6
```

Comments by Martín in italics.

Natalia.

1. *I recommend using the pre-loaded FANG database as it is (with the corresponding length). Here, Natalia downloaded the data and took the first months of the 2021 year. She includes two plots. Are both plots the same? I would recommend adding some comments to understand the plots. Also, remember the objective is to link the volume with the returns as explained in the instructions (readme repo).*
2. *This is much better Natalia. I understand the value of the trend, it is useful, and it is a good idea to incorporate it. I agree with your contribution. However, the trend in loess reduces some detail which may be important later, when you link volume with stock returns. Ideally, this is where others start contributing to the project.*

Gonzalo.

1. *Interesting piece of evidence. Looks good and it was important to look at this at this moment. However, more questions arise. First, I recommend using ggplot to be consistent with the class and the previous analysis. The ggplot also will allow you to show differences by stocks. Gonzalo took all stocks together and there might be differences between them. There might also be differences by year. There is something else that could bring more contributions on board, this is to compare returns not only with respect to volume but changes in the volume (percentage changes). Just as you do a percentage change in price you can also calculate percentage change in volume. This will allow you to compare both variables under the same units.*

Diana.

1. *I have similar comments as in the case of Gonzalo. First, I recommend using ggplot to be consistent with the class and the previous analysis. The ggplot also will allow you to show differences by stocks. There is something else that could bring more contributions on board, this is to compare returns not only with respect to volume but changes in the volume (percentage changes). Just as you do a percentage*

change in price you can also calculate percentage change in volume. This will allow you to compare both variables under the same units.

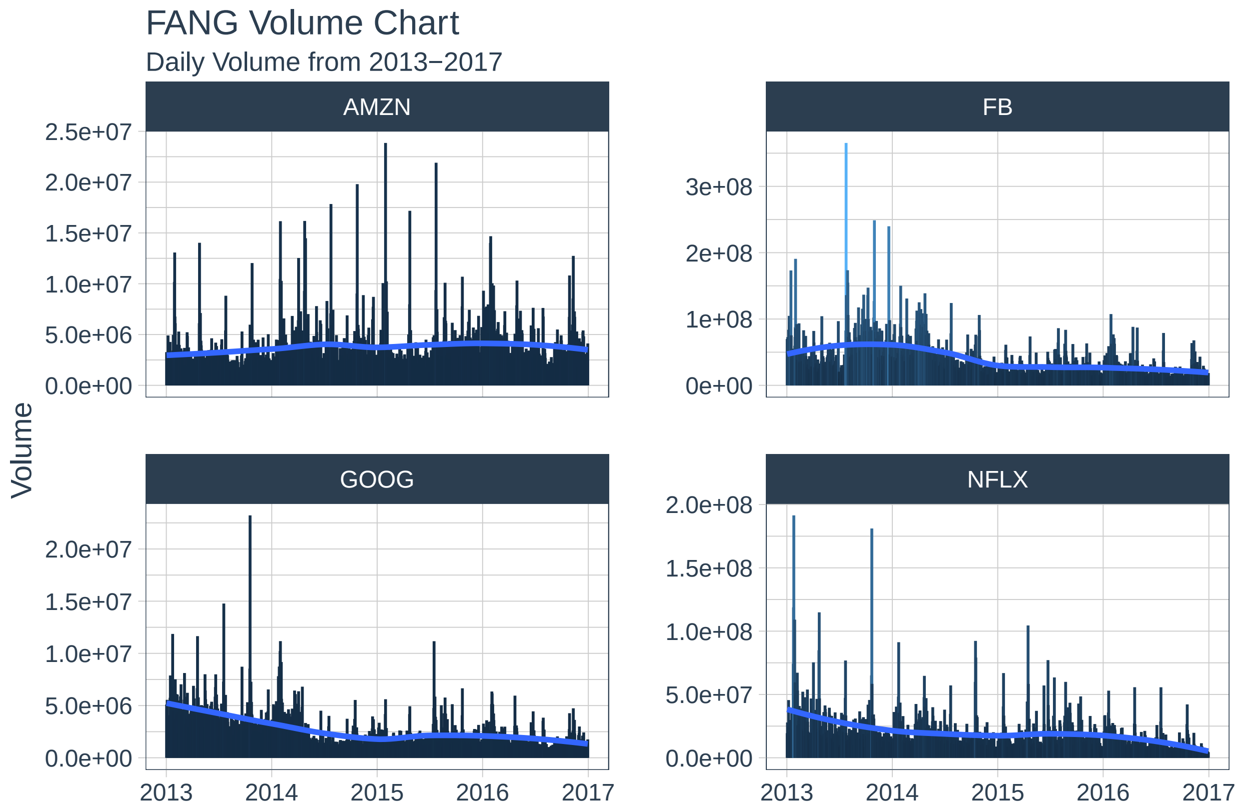
2. *This is much better as you listen to my comments. However, when you produce the PDF you end up with more than 160 pages because your code shows all data and this is not necessary, plots are fine. I already fixed that. You did good at producing ggplots and percentage changes of volume, apparently they are correct (unless someone sees a problem). I recommend adding a proper title so we can identify which plot corresponds to which stock. Also, I recommend colouring (add colours) the dots with respect to years so we can see whether specific years have some pattern in the plot.*

FANGs Volume in Recent Years.

Looking at volume patterns over time can help get a sense of the conviction behind rises and falls in specific stocks and entire markets, for that reason it is proposed to analyze the volume charts for each FANG stock, in order to understands its trend and then relate it with the stocks' returns.

```
FANG %>%
ggplot(aes(x = date, y = volume, group = symbol)) +
    geom_segment(aes(xend = date, yend = 0, color = volume)) +
    geom_smooth(method = "loess", se = FALSE) +
    labs(title = "FANG Volume Chart",
         subtitle = "Daily Volume from 2013-2017",
         y = "Volume", x = "") +
  facet_wrap(~ symbol, ncol = 2, scale = "free_y") +
  theme_tq() +
    theme(legend.position = "none")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The chart above shows the daily volumes and their trend line of FANGs stocks in the last few years. The bars represent the number of shares traded daily, and the line represents the trend that the volumes of each share have followed from 2013-2017.

It is clear that to analyze daily volumes, and see their highest and lowest points, it is necessary to have a graph focused on a shorter period of time, in order to obtain more clarity in the data and get more objective interpretations.

Still, trend lines are useful for analyzing data over long periods of time.In the graph, it is observed that FB, NFLX and GOOG stocks' volume has had a negative behavior, meaning that from 2013 to 2017 the average volume of shares decreased. In the case of FB, it is important to highlight that at the beginning it had a positive behavior, but at the end of 2014 it took a downward trend.In contrast, the trend line for AMZN looks stable throughout that period, so we infer that it has maintained a constant average in its volumes.

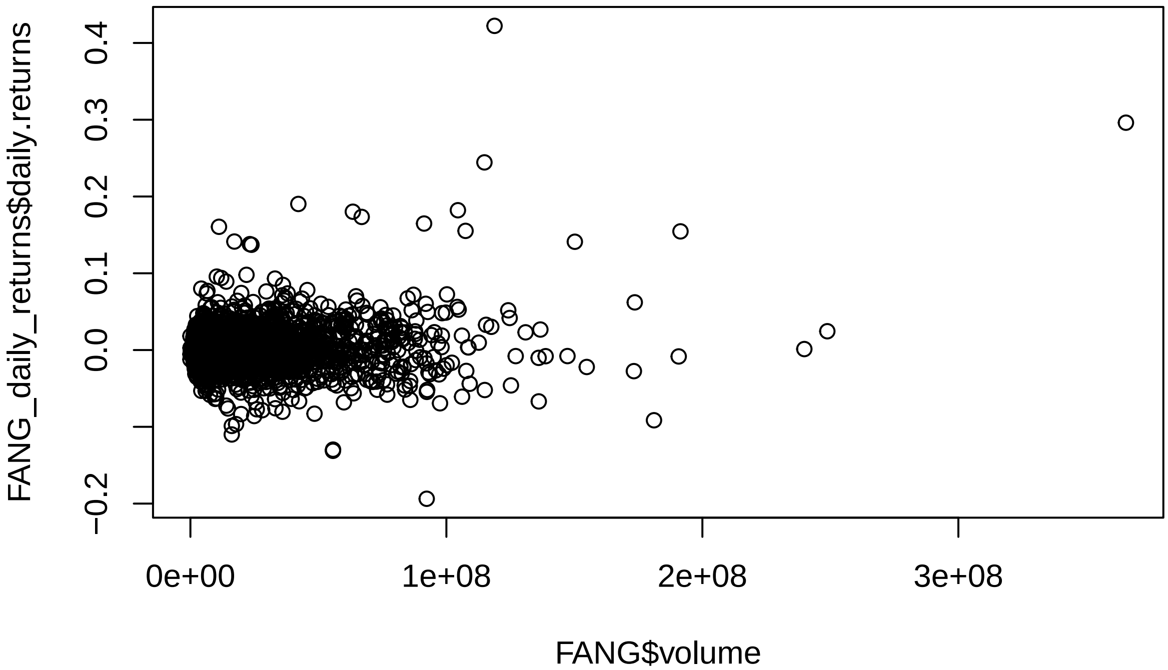
Now, it is proposed to analyze the a graph that shows the volumes by price

correlation between daily volume and daily returns

```
FANG_daily_returns <- FANG %>%
  group_by(symbol) %>%
  tq_transmute(select = adjusted,
               mutate_fun = periodReturn,
               period = "daily",
               type = "arithmetic")
```

```
## Registered S3 method overwritten by 'tune':
##   method                   from
##   required_pkgs.model_spec parsnip
```

```
plot(FANG_daily_returns$daily.returns~FANG$volume)
```



Thanks to the scatter diagram, we can see that the relationship between daily returns and the volume of FANG have an independent and non-linear relationship. We can interpret it this way because their correlation coefficient is very close to zero (0.1356115) and its diagram does not indicate any curves. This coefficient can be seen below:

```
cor.test(FANG_daily_returns$daily.returns, FANG$volume)
```

```
##
##  Pearson's product-moment correlation
##
## data:  FANG_daily_returns$daily.returns and FANG$volume
## t = 8.6892, df = 4030, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1051836 0.1657857
## sample estimates:
##       cor
## 0.1356115
```

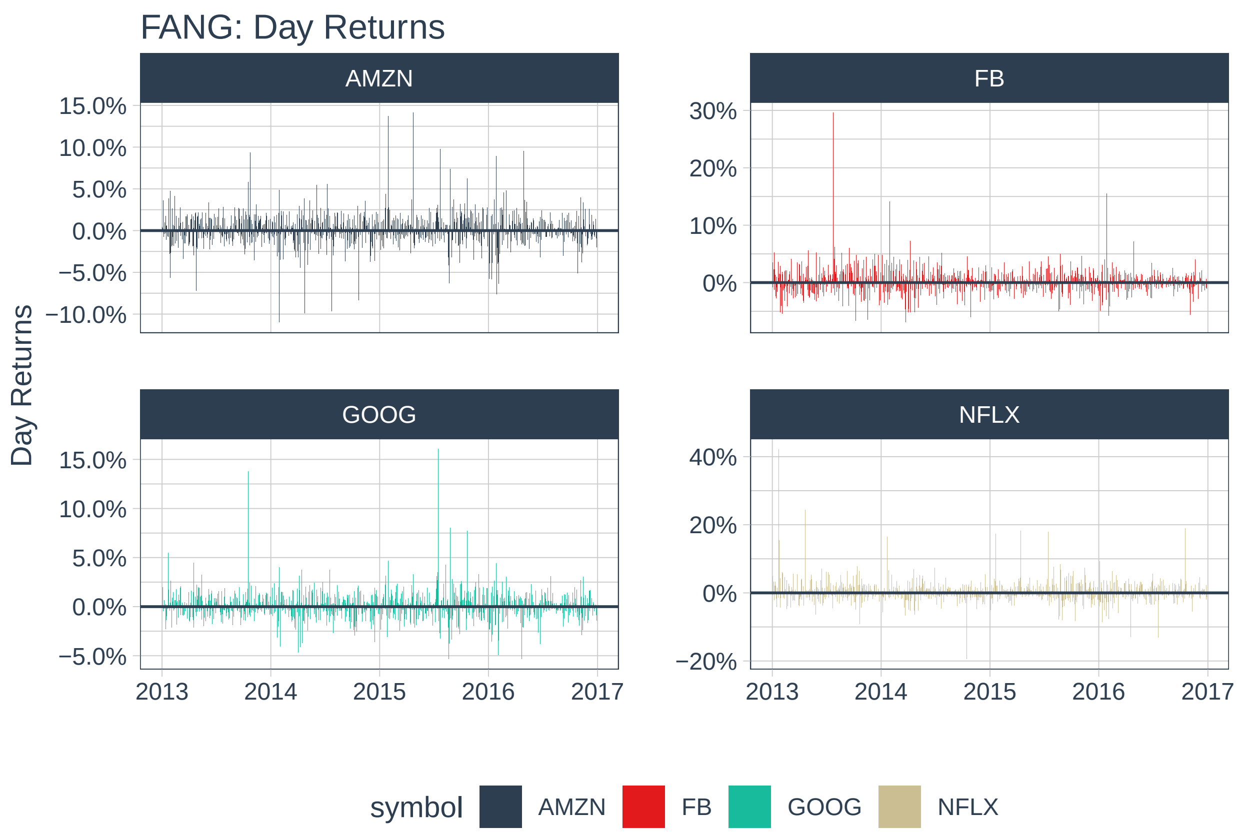
Thus, we could conclude that the return of a stock has not relationship with its volume in stock market; these variables are independent.

Does anyone disagree? or someone to improve my conclusion

```
FANG_daily_returns <- FANG %>%
  group_by(symbol) %>%
```

```
  tq_transmute(select = adjusted,
               mutate_fun = periodReturn,
               period = "daily",
               type = "arithmetic",
               date = "2015-12-30" )
```

```
FANG_daily_returns %>%
ggplot(aes(x = date-1, y = daily.returns, fill = symbol)) + geom_col() +
geom_hline(yintercept = 0, color = palette_light()[[1]]) + scale_y_continuous(labels = scales::percent)
labs(title = "FANG: Day Returns",
         y = "Day Returns", x = "") +
    facet_wrap(~ symbol, ncol = 2, scales = "free_y") +
    theme_tq() + scale_fill_tq()
```



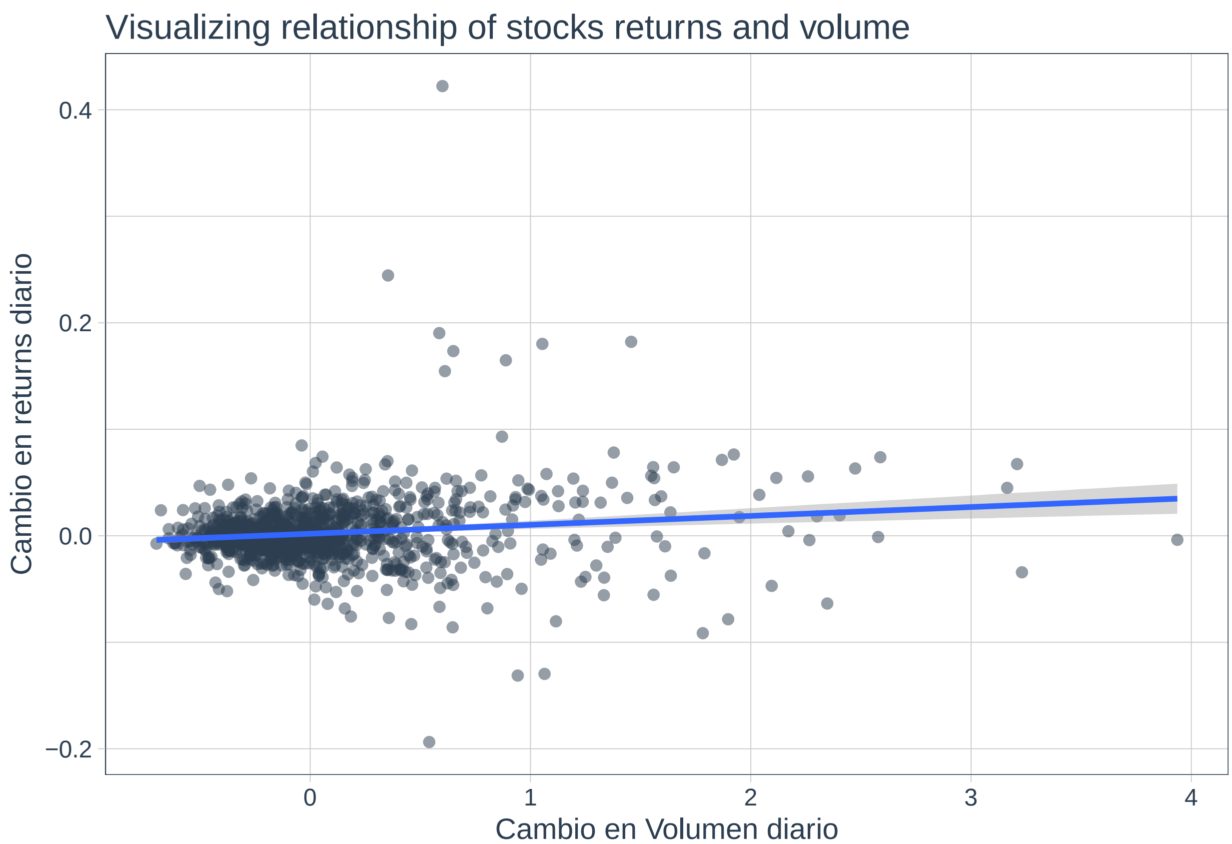

```
Netflix_daily_returns  <- FANG_daily_returns %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "NFLX")

Netflix <- FANG %>%
  filter(symbol == "NFLX")

Netflix_volume <- Netflix %>%
  tq_transmute(select = volume,
               mutate_fun = periodReturn,
               period = "daily",
               type = "arithmetic",
```

```
               date = "2015-12-30" )
Netflix1 <- merge(Netflix_daily_returns, Netflix_volume)
#Netflix1
```

```
FANG_daily_volume <- FANG %>%
  group_by(symbol) %>%
  tq_transmute(select = volume,
               mutate_fun = periodReturn,
               period = "daily",
               type = "arithmetic",
               date = "2015-12-30" )

Netflix_daily_returns  <- FANG_daily_returns %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "NFLX")

Netflix_daily_volume  <- FANG_daily_volume %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "NFLX")

Netflix1 <- data.frame(Netflix_daily_returns, Netflix_daily_volume)
#Netflix1
```

```
stock_ret_Netflix <- Netflix1 %>%
ggplot(aes(x = daily.returns.1, y = daily.returns)) +
geom_point(color = palette_light()[[1]], alpha = 0.5) + geom_smooth(method = "lm") +
labs(title = "Visualizing relationship of stocks returns and volume",  x = "Cambio en Volumen diario",y
stock_ret_Netflix
```

```
## `geom_smooth()` using formula 'y ~ x'
```



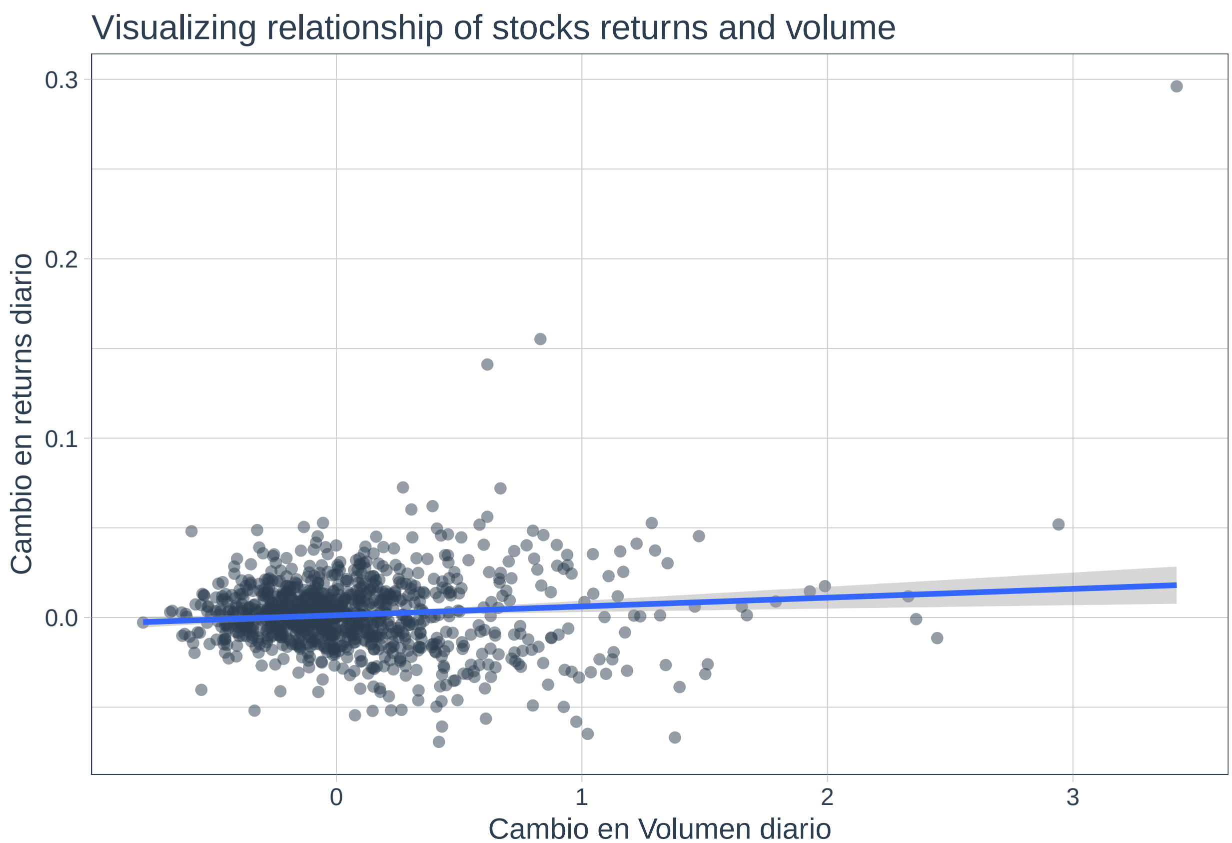

```
cor.test(Netflix1$daily.returns.1, Netflix1$daily.returns)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Netflix1$daily.returns.1 and Netflix1$daily.returns
## t = 4.494, df = 1006, p-value = 7.8e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07922777 0.20029939
## sample estimates:
##      cor
## 0.140288
```

```
FB_daily_returns  <- FANG_daily_returns %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "FB")


Fb_daily_volume  <- FANG_daily_volume %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "FB")

FB1 <- data.frame(FB_daily_returns, Fb_daily_volume)
#FB1
```

```
stock_ret_FB <- FB1 %>%
ggplot(aes(x = daily.returns.1, y = daily.returns)) +
geom_point(color = palette_light()[[1]], alpha = 0.5) + geom_smooth(method = "lm") +
labs(title = "Visualizing relationship of stocks returns and volume",  x = "Cambio en Volumen diario",y
stock_ret_FB
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
cor.test(FB1$daily.returns.1, FB1$daily.returns)
```

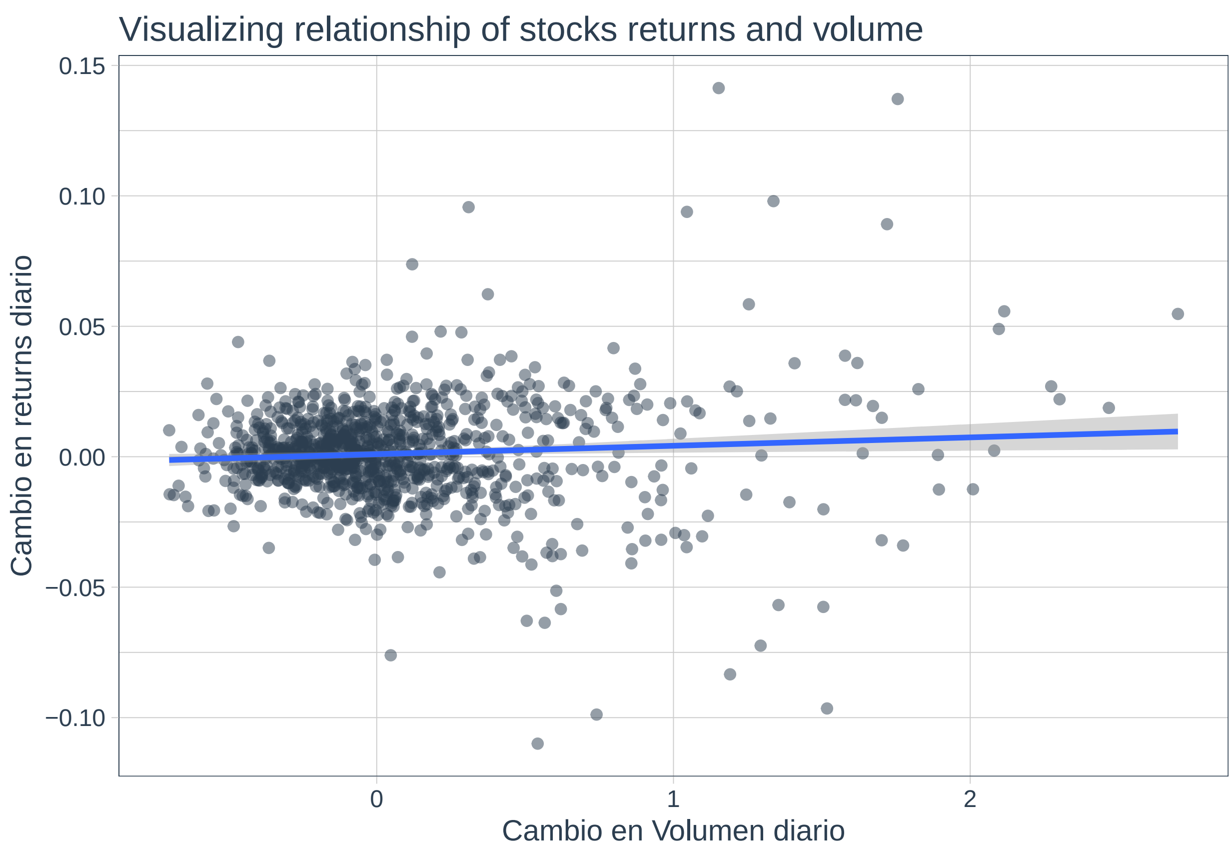
```
##
##  Pearson's product-moment correlation
##
## data:  FB1$daily.returns.1 and FB1$daily.returns
## t = 3.1306, df = 1006, p-value = 0.001795
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03670089 0.15900688
## sample estimates:
##        cor
## 0.09822478
```

```
AMZN_daily_returns  <- FANG_daily_returns %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "AMZN")
```

```
AMZN_daily_volume  <- FANG_daily_volume %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "AMZN")

AMZN1 <- data.frame(AMZN_daily_returns, AMZN_daily_volume)
#AMZN1
```

```
stock_ret_AMZN <- AMZN1 %>%
ggplot(aes(x = daily.returns.1, y = daily.returns)) +
geom_point(color = palette_light()[[1]], alpha = 0.5) + geom_smooth(method = "lm") +
labs(title = "Visualizing relationship of stocks returns and volume",  x = "Cambio en Volumen diario",y
stock_ret_AMZN
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
cor.test(AMZN1$daily.returns,AMZN1$daily.returns.)
```

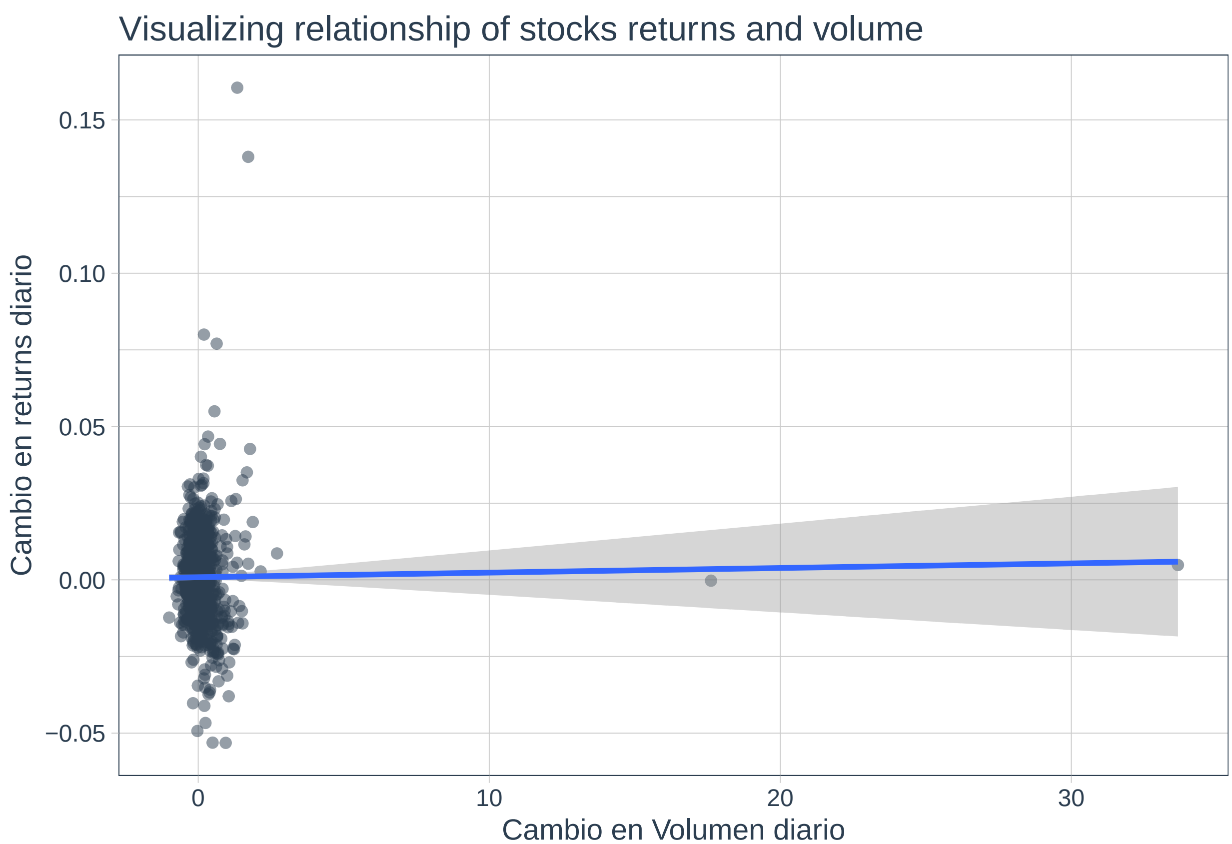
```
##
##  Pearson's product-moment correlation
##
## data:  AMZN1$daily.returns and AMZN1$daily.returns.
## t = 2.45, df = 1006, p-value = 0.01445
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.01534225 0.13810550
## sample estimates:
##        cor
```

```
## 0.07701577
```

```
GOOG_daily_returns  <- FANG_daily_returns %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "GOOG")



GOOG_daily_volume  <- FANG_daily_volume %>%
  mutate(day = as.factor(day(date))) %>%
  filter(symbol == "GOOG")

GOOG1 <- data.frame(GOOG_daily_returns, GOOG_daily_volume)
#GOOG1
```

```
stock_ret_GOOG <- GOOG1 %>%
ggplot(aes(x = daily.returns.1, y = daily.returns)) +
geom_point(color = palette_light()[[1]], alpha = 0.5) + geom_smooth(method = "lm") +
labs(title = "Visualizing relationship of stocks returns and volume",  x = "Cambio en Volumen diario",y
stock_ret_GOOG
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
cor.test(GOOG1$daily.returns,GOOG1$daily.returns.)
```

```
##
##  Pearson's product-moment correlation
##
```

```
## data:  GOOG1$daily.returns and GOOG1$daily.returns.
## t = 0.40708, df = 1006, p-value = 0.684
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.04895191  0.07452079
## sample estimates:
##        cor
## 0.01283336
```

As we can see, the graphs are very similar, except the Google graph, in which the values are concentrated in one area and the standard error are greater than the other graphs, with Google the standard error increases if the change in the daily volume increases. We can concluded that the volume changes are not representative for the returns. Continuing with this, we can see that the smallest correlation is with Google, which is 0.0128, we can conclude that the daily return of Google is not related to the change in daily volume. In contrast, Netflix's correlation is the biggest, which is 0.1402. The correlation between the daily return and change daily volume of Netflix is more than Google's correlation.