Do Technology Sector stocks outperform other sector of stocks ?

Hsiao-Han Chen, Zeyu Hua, Lianguan Zhou

12/10/2017

Abstract

After the financial crisis, technology stocks always attract extensive attention in the news and in stock market. We will like to analyze all the stock and sector data from Standard & Poor's 500 index since financial crisis. The data will be collected from Yahoo Finance and we will take a deep look into the distribution of return in each sector and test whether stocks from technology sector outperform stocks from other sectors.

1. Introduction

S&P 500 index is an American stock market index that includes largest 500 companies and highest market capitalization with common stocks listed on the exchange. The index is an important reference for the status of U.S. stock market as the components of the index possess a high market value. We can often hear the news reporting several biggest technology firms such as Apple Inc, Amazon, Intel and Microsoft Corporation are creating high income and drawing the attention from the public.

The Global Financial Crisis of 2007-2008 has been considered the worst financial crisis since the Great Depression of 1930s. U.S. stocks had dropped to the bottom during the crisis. Therefore, we are focusing on periods after the crisis and may provide us some thoughts about how fast the technology industry has grown. If the technology sector stocks do outperform other sectors, we can also discuss the reason why causing the technology stocks rally in this few years.

2. Methods

2.1 Participants

In this project, we will collect stock data from the Yahoo Finance and use R for data processing and analysis. There are three main components in this project

2.1.1. Daily Stocks Return

Daily Stock return from S&P 500 between the period from January 01, 2009 to September 30, 2017. The data includes 2202 daily return data for each stock and total 490 stocks are able to retrieve from the websites.

2.1.2. Sector and industries of the stocks

There are total 11 sectors in the S&P 500 index. Please see Table 1 for the stocks count in each sector.

Sectors	Stock counts
Consumer Staples	37
Energy	35
Financials	65
Health Care	60
Industrials	66
Information Technology	68
Materials	25
Real Estate	30
Telecommunications Services	5
Utilities	28

Table 1: Sector table

2.1.3. Market Capitalization

Market Capitalization means the market value of the stock and it is calculated by multiplying total outstanding shares of the company and the market price of the stock. In this project, we collected the market capitalization of all S&P 500 stock components as it is an important weighting factor in the process.

2.2 Measures and Procedures

2.2.1. Data Processing

First, we collect the 500 stock daily return data from YAHOO Finance. All the tickers of 500 stocks and market capitalization information is provided on the website. "Quantmod" package in r is installed for retrieving stock data with ticker. Next, from the market capitalization data, we will create weightings of stocks for each sector. The method is summing up the market capitalization by sector, and have the total market capitalization of each sector. Then we divide the individual stock's market capitalization by the market capitalization of each sector. The result represents the weighting factor of an individual stock in each sector. We multiply the weighting factor and the daily return of the stock and sum up to get the market cap weighted return of each sector. Here is the code for collecting the data.

```r
library(quantmod)
library(tidyverse)
library(tibbletime)

marketcap <- read_csv("../data/marketcap.csv")
```

```
## Parsed with column specification:
## cols(
##   Symbol = col_character(),
##   Name = col_character(),
##   Sector = col_character(),
##   Price = col_double(),
##   `Dividend Yield` = col_double(),
##   `Price/Earnings` = col_double(),
##   `Earnings/Share` = col_double(),
##   `Book Value` = col_double(),
##   `52 week low` = col_double(),
##   `52 week high` = col_double(),
##   `Market Cap` = col_double(),
##   EBITDA = col_double(),
##   `Price/Sales` = col_double(),
```

```
##   `Price/Book` = col_double(),
##   `SEC Filings` = col_character()
## )
```

```r
indsum <- marketcap %>%
  group_by(Sector) %>%
  summarise(sum_index = sum(`Market Cap`, na.rm = T))

# market cap sum by sector
datamerge <- marketcap %>%
  left_join(indsum, by = "Sector")

# create weight by marketcap for each sector
weight <-
  datamerge %>%
  mutate(weight = `Market Cap`/sum_index) %>%
  select(Sector, Symbol, weight)

sector <- levels(as.factor(weight$Sector))
```

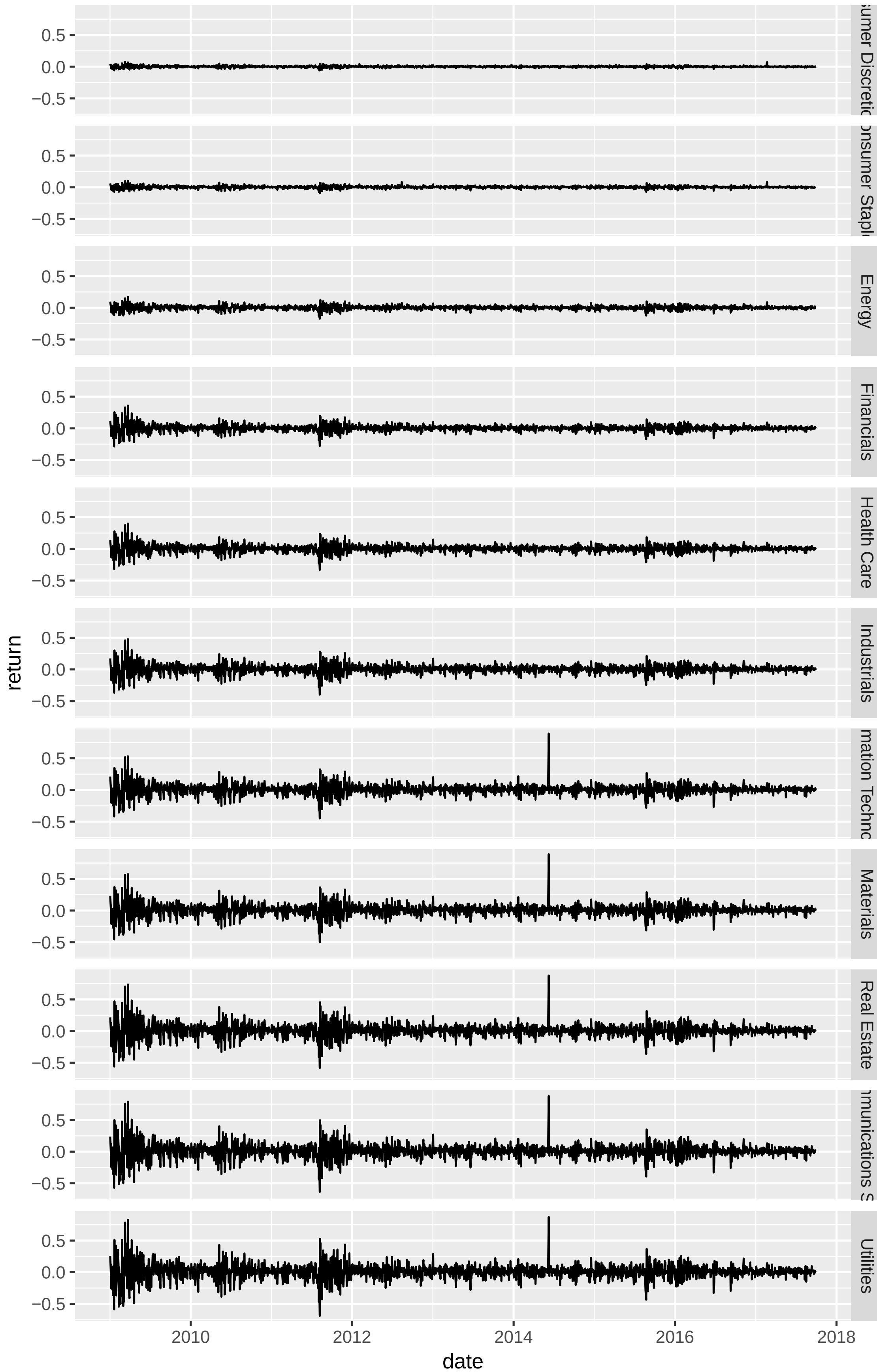
```r
# Total 11 sectors
# for each sector get all stocks daily information
for (j in seq_along(sector)) {
  # get symbols for each sector
  symbol <-
    weight %>%
    filter(Sector == sector[j]) %>%
    pull(Symbol)

  # wight for each symbol
  symbolweight <-
    weight %>%
    filter(Sector == sector[j]) %>%
    pull(weight)

  # quantmod package to get stock data from Yahoo
  for (i in seq_along(symbol)) {
    tryit <-
      try(getSymbols(
        symbol[i],
        from = "2009-01-01",
        to = "2017-09-30",
        src =  "yahoo",
        adjust =  TRUE
      ))
    if (inherits(tryit, "try-error")) {
      i <- i + 1
    } else {
      getSymbols(
        symbol[i],
        from = "2009-01-01",
        to = "2017-09-30",
        src =  "yahoo",
        adjust =  TRUE
      )
```

```
      return <- dailyReturn(get(symbol[i]))
      weightedreturn <- return * symbolweight[i]
      names(return) <- symbol[i]
      if ((i == 1) & (j == 1)) {
        dailyreturndata <- return
        weighteddata <- weightedreturn
      } else {
        dailyreturndata <- merge(dailyreturndata, return)
        weighteddata <- merge(weighteddata, weightedreturn)
      }
    }
  }
  sectorreturn <- rowSums(weighteddata, na.rm = TRUE)
  names(sectorreturn) <- sector[j]
  if (j == 1) {
    tempdata <- cbind(weighteddata, sectorreturn)
    sectordata <- tempdata[, ncol(tempdata)]
  } else {
    sectordata <- merge(sectordata, sectorreturn)
  }
}

# transform to time series data
names(sectordata) <- sector

# get daily return for each symbol and each sector
xtsdailyreturn<-as.xts(dailyreturndata)
# write.zoo(xtsdailyreturn,"data/dailyreturndata.csv", sep=",")

xtssector<-as.xts(sectordata)
# write.zoo(xtssector,"data/sectordata.csv", sep=",")

# data that didn't retrieve successfully
different.names <- weight$Symbol[!(weight$Symbol %in% names(dailyreturndata))]
# weight[different.names,2]
```

2.2.2. Descriptive Analysis

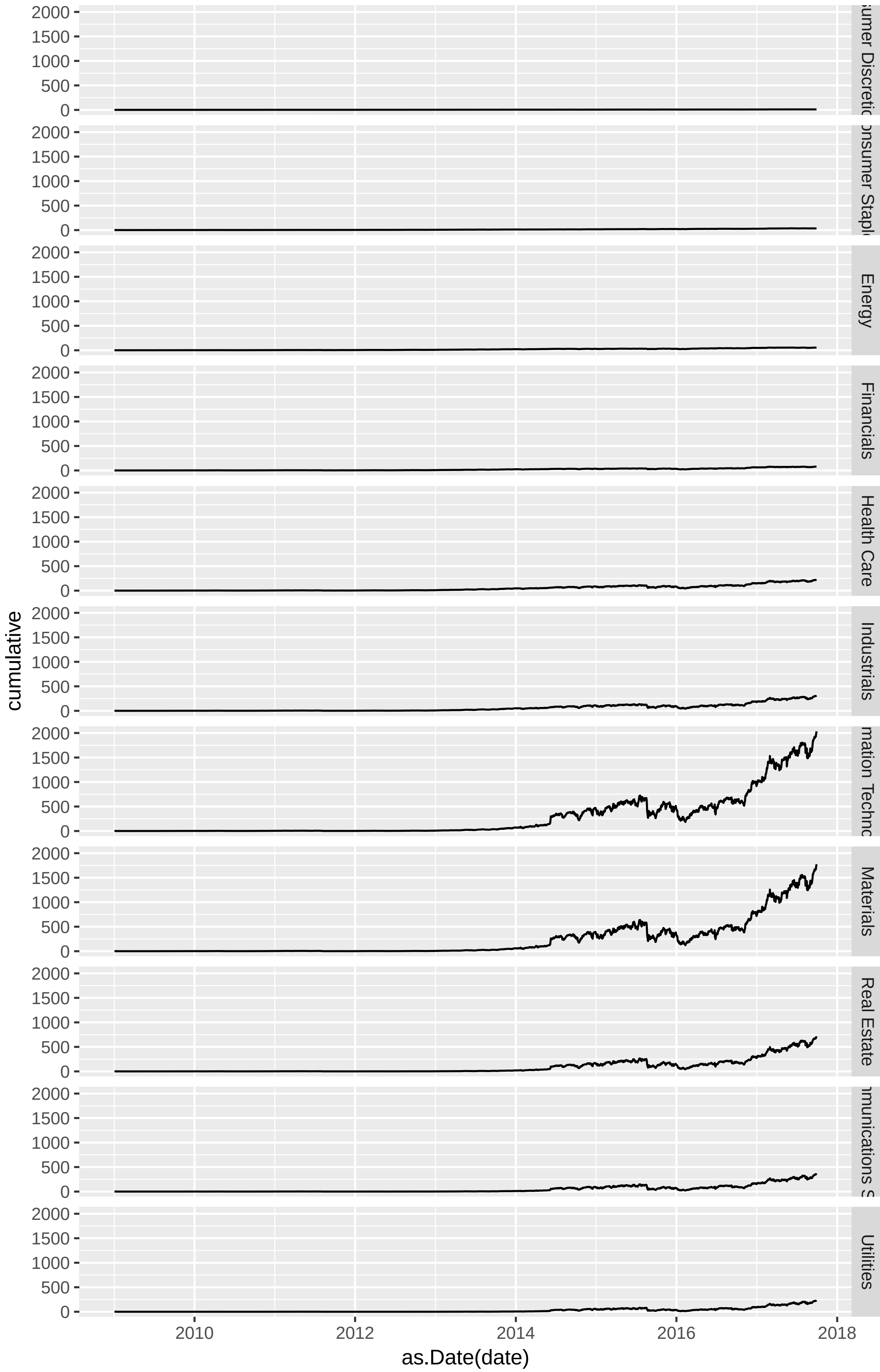
In this part, we want to know the return for each sector by day

```
# daily return for each sector
sectordata <- read_csv("../data/sectordata.csv")

sectordata <- as_tibble(sectordata) %>%
  rename(date = Index)

sectordata %>%
  gather(key = "sector", value = "return", -date) %>%
  ggplot(aes(x = date, y = return)) +
  geom_line() +
  # geom_smooth() +
  facet_grid(sector ~ .)
```



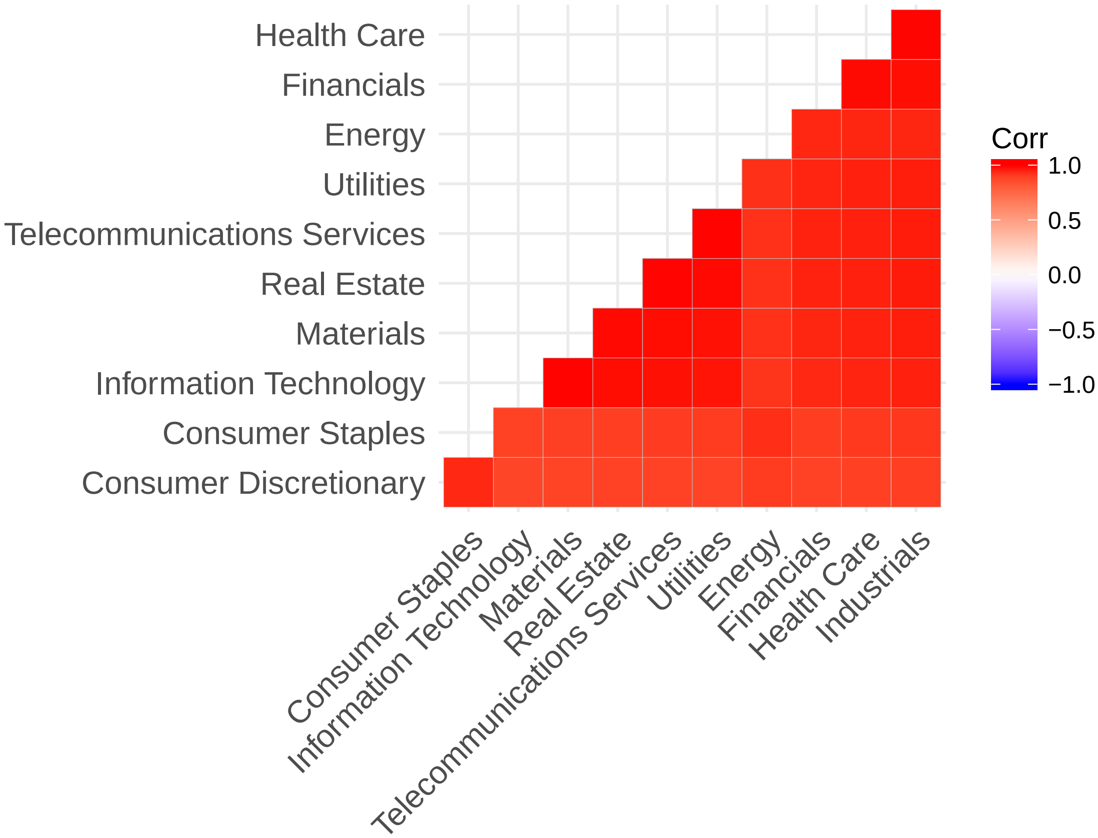
umer Discretio
onsumer Staple
Energy
Financials
Health Care
Industrials
mation Techno
Materials
Real Estate
munications S
Utilities
return
date
0.5
0.0
−0.5
2010
2012
2014
2016
2018

```r
# everyday's culumative change
sectordata %>%
  mutate_at(vars(`Consumer Discretionary`:Utilities),
            function(x) {x = x + 1}) %>%
  mutate_at(vars(`Consumer Discretionary`:Utilities), cumprod) %>%
  gather(key = "sector", value = "cumulative", -date) %>%
  ggplot(aes(x = as.Date(date), y = cumulative)) +
  geom_line() +
  # geom_smooth() +
  facet_grid(sector ~ .)
```


cumulative
as.Date(date)
2010
2012
2014
2016
2018
0
500
1000
1500
2000
umer Discretio
onsumer Staple
Energy
Financials
Health Care
Industrials
mation Techno
Materials
Real Estate
munications S
Utilities

And we can also explore the correlation between daily returns of differenct sectors:

```r
return_corr <-
  sectordata %>% dplyr::select(-date) %>%
  map_dfc(as.numeric) %>% as.matrix %>% cor

ggcorrplot(return_corr, hc.order = T, type = "lower")
```



We can see that the daily return of differenct sectors are highly correlated.

2.2.3. Hypothesis Testing

From the exploratory we can construct a hypothesis testing:

The average daily return of tech sector is larger than other sectors.

Since the sample size is relatively large, we can use z-test. For each sector i, it's sample variance s_i^2 and actual variance σ_i^2 are almost the same. So we can have a hypothesis testing for each sector i:

$$H_0 : \mu_{tech} \leq \mu_i H_a : \mu_{tech} > \mu_i$$

Use `z.test` function in `BSDA` package:

```r
tech <- sectordata$`Information Technology`
sectorwithouttech <- str_subset(sector, "^(?!Information Technology)")

for (sectorname in sectorwithouttech) {
  the_sector <- sectordata %>% pull(sectorname)
```

```
  print(paste("For sector:", sectorname))
  z.test(tech, the_sector, alternative = "greater",
         sigma.x = sd(tech)^2, sigma.y = sd(the_sector)^2) %>%
    print
}
```

```
## [1] "For sector: Consumer Discretionary"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 41.769, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.005105397          NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.001133657
##
## [1] "For sector: Consumer Staples"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 36.824, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.004480668          NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.001758176
##
## [1] "For sector: Energy"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 32.743, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.003994236          NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.002242842
##
## [1] "For sector: Financials"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 25.117, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.003167997          NA
```

```
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.003058342
##
## [1] "For sector: Health Care"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 17.799, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.002284922          NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.003930765
##
## [1] "For sector: Industrials"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = 10.659, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.001409112          NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.004782110
##
## [1] "For sector: Materials"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = -3.1625, p-value = 0.9992
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.0009650681            NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.007083212
##
## [1] "For sector: Real Estate"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = -6.3315, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.001945213           NA
## sample estimates:
##   mean of x   mean of y
```

```
## 0.006448345 0.007992425
##
## [1] "For sector: Telecommunications Services"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = -7.4022, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.002435243           NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.008440833
##
## [1] "For sector: Utilities"
##
##  Two-sample z-Test
##
## data:  tech and the_sector
## z = -8.5539, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.002993148           NA
## sample estimates:
##   mean of x   mean of y
## 0.006448345 0.008958757
```

2.3. Results

The hypothesis testing shows that the average daily return of tech sector is greater than most of other sectors, but it's daily average return is not greater than `Materials`, `Real Estat`, `Telecommunication Services` and `Utilities`.