Market Metrics

Eduardo Almeida

2020-03-27

Contents

Executive summary . . . 1
Introduction . . . 1
Geting Data . . . 2
Overview of data . . . 8
Portfolio Analysis . . . 32
KeyMetrics x Porfolio Mitigation . . . 39
Conclusion . . . 62

Executive summary

This capstone program is based on evaluating how financial metrics relates to each sector and how to explore which components tends to deliver the best overall risk mitigation portfolio decision and how to create a porftolio.

The data used for this project is on companies that are in the SP500 index. First there is an exploratory analysis of the data and how it relates to the metrics from industries and sector segments as well as its price and risk.

Sectors are then evaluated by daily prices in order to determine the best set of clusters that represents the groups of sectors presented in the SP500.

The first portfolio is created per sector to understand which sectors to chose in order to mitigate risks per ROI (Return on investment), then a second porfolio is created in each sector to evaluate the best and worst companies to mitigate risk.

All 57 metrics are then evaluated per sector and the best and worst companies of porftolio in order to understand how they relate with sectors and risk mitigation.

Lastly the price of the chosen companies of porftolio are predicted and the porfolio is rebalanced per period in order to evaluate how prices prediction changes the behavior of risk mitigation.

Introduction

This project aims to understand how financial metrics relates to each sector and how to explore which components tends to deliver the best overall risk mitigation portfolio decision and how to create a porftolio.

This is a personal attempt to merge a brief study on machine learning techniques and financial analysis from courses from michigan coursera and harvardx. From my perspectiva, most of the financial analysis is usually based on the ratio of risk versus return of an investment.

The issue is that studies teach us how to evaluate companies based on a set of benchmarked metrics and to evaluate data manually, my intent is to organize it in a broader perspective and gather as much data as possible in order to evaluate criticaly how financial metrics work per companies segments and how to use this information to evaluate risk and return of a portfolio investment.

This is a broad area of study and many metrics and analysis are beyond my understanding. This means that conceptual mistakes can happen, but I'm willing to evaluate this in order to achieve a better understanding of the market.

Geting Data

Libraries

Load the packages needed for this project:

- Handle API: `library(httr)` and `library(jsonlite)`

- Tidy metrics: `library(tidyverse)`
- List manipulation: `library(purrr)`
- Data preparation: `library(recipes)` and `library(janitor)`
- Data exploration visual: `library(patchwork)` and `library(ggstatsplot)`
- Machine learning models: `library(h2o)`
- Time based preparation: `library(anytime)`, `library(timetk)` and `library(tibbletime)`
- Markdown tables: `library(knitr)` and `library(kableExtra)`
- Portfolio Analytics: `library(PortfloioAnalytics)`

```r
require(httr) #Working with url
library(jsonlite) #Working with json data for API
library(tidyverse) #Tidy dataframe packages
library(purrr) #list manipulation
library(janitor) # Data cleansing and pivot
library(patchwork) #Easy grid arrange of ggplots
library(tidyquant) #Set of finance packages
library(anytime) #read any type of date format
library(readxl) #read/write excel data
library(stringr) #string manipulation
library(timetk) #tibble format for time based dataframe
library(tibbletime) #tibble format for time based dataframe
library(PortfolioAnalytics) #Porfolio analysis
library(ROI) #Optimization package
library(ROI.plugin.glpk) #Plugins needed
library(ROI.plugin.quadprog) #Plugins needed
library(knitr) #Tables in rmd
library(kableExtra) #Graphics for knitr tables
library(cowplot) #Grid plot for list plots
library(ggstatsplot) #Statistical testing in plot
library(h2o) #Machine learning models
library(lime) #Allow for black box models to be easily evaluated
library(lubridate) #Allow for changes in date format

library(gridExtra)
library(ggdendro)
library(zoo)
library(tsibble)
library(broom)
```

Table 1: API Structure

Category	Informaton	url
Company Valuation	Symbols List	https://financialmodelingprep.com/api/v3/company/stock/list
Company Valuation	Company Profile	https://financialmodelingprep.com/api/v3/company/profile/
Company Valuation	Income Statement	https://financialmodelingprep.com/api/v3/financials/income-statemen
Company Valuation	Balance Sheet Statement	https://financialmodelingprep.com/api/v3/financials/balance-sheet-sta
Company Valuation	Cash Flow Statement	https://financialmodelingprep.com/api/v3/financials/cash-flow-statem
Company Valuation	Company Financial Ratios	https://financialmodelingprep.com/api/v3/financial-ratios/
Company Valuation	Company Enterprise Value	https://financialmodelingprep.com/api/v3/enterprise-value/
Company Valuation	Company Key Metrics	https://financialmodelingprep.com/api/v3/company-key-metrics/
Company Valuation	Company Rating	https://financialmodelingprep.com/api/v3/company/rating/
Stock Price	Stock Real-time Price	https://financialmodelingprep.com/api/v3/stock/real-time-price/
Stock Price	Historical Daily Price	https://financialmodelingprep.com/api/v3/historical-price-full/

```r
options(scipen=999)
```

Data

The data obtained is from an API from financialmodelingprep.com which consists of:

```r
API_Structure <- tribble(
  ~Category, ~Informaton, ~url, ~Options, ~TimeUpdate,
  "Company Valuation","Symbols List", "https://financialmodelingprep.com/api/v3/company/stock/list", NUI
  "Company Valuation", "Company Profile","https://financialmodelingprep.com/api/v3/company/profile/","c
  "Company Valuation", "Income Statement","https://financialmodelingprep.com/api/v3/financials/income-s
  "Company Valuation", "Balance Sheet Statement", "https://financialmodelingprep.com/api/v3/financials/
  "Company Valuation", "Cash Flow Statement", "https://financialmodelingprep.com/api/v3/financials/cash
  "Company Valuation", "Company Financial Ratios", "https://financialmodelingprep.com/api/v3/financial-
  "Company Valuation", "Company Enterprise Value",  "https://financialmodelingprep.com/api/v3/enterpris
  "Company Valuation", "Company Key Metrics", "https://financialmodelingprep.com/api/v3/company-key-met
  "Company Valuation", "Company Rating", "https://financialmodelingprep.com/api/v3/company/rating/", "C
  "Stock Price", "Stock Real-time Price", "https://financialmodelingprep.com/api/v3/stock/real-time-pri
  "Stock Price", "Historical Daily Price", "https://financialmodelingprep.com/api/v3/historical-price-f
) %>%
  mutate(id = row_number()) %>%
  select(id, everything())


kable(API_Structure[,-1], caption = "API Structure") %>%
  kable_styling(full_width = F)
```

Brief overview of stock lists

In order to use the API structure two functions are created to help getting the data

```r
#Company informations
GetCompanyProfile <- function(url, company = NULL){

headers = c(
```

```r
  `Upgrade-Insecure-Requests`= '1'
)

params = list(
  `datatype` = 'json'
)

res <- httr::GET(url = paste0(url,"/",company),
                 httr::add_headers(.headers=headers), query = params)

data <- content(res, as = "text")

data <- fromJSON(data, flatten = T) %>%
  flatten_dfr()

return(data)

}

#Get data from API structure
GetData <- function(url, company = NULL, Period = NULL){


headers = c(
  `Upgrade-Insecure-Requests`= '1'
)

params = list(
  `datatype` = 'json'
)



if (is.null(company) & is.null(Period)) {
  res <- httr::GET(url = url,
                   httr::add_headers(.headers=headers), query = params)

} else if (is.null(Period)) {
  res <- httr::GET(url = paste0(url,"/",company),
                   httr::add_headers(.headers=headers), query = params)

} else {
  res <- httr::GET(url = paste0(url,"/",company, "?period=",Period),
                   httr::add_headers(.headers=headers), query = params)
}

data <- content(res, as = "text")

data <- fromJSON(data, flatten = T) %>%
        detect(is.data.frame) %>%
        as_tibble()

return(data)
```

```
}
```

Let's get all company symbols from the API

```
Stock_Lists <- GetData(url = "https://financialmodelingprep.com/api/v3/company/stock/list")

glimpse(Stock_Lists)
```

```
## Observations: 13,854
## Variables: 4
## $ symbol   <chr> "SPY", "CMCSA", "KMI", "INTC", "MU", "GDX", "GE", "BAC", "...
## $ name     <chr> "SPDR S&P 500", "Comcast Corporation Class A Common Stock"...
## $ price    <dbl> 254.19, 38.22, 12.64, 50.08, 37.38, 25.50, 7.08, 21.98, 33...
## $ exchange <chr> "NYSE Arca", "Nasdaq Global Select", "New York Stock Excha...
```

There is 13584 symbols, in order to explore the data we must choose a sample set from this dataset. In order to understand each sector, SP500 companies are a good choice since it is usually used to define how the US market is and represents a great variety of sectors and industries segments.

```
#SP500 Indexes
SP500 <- tq_index("SP500")
```

```
## Getting holdings for SP500
```

```
Stock_Lists <- GetData(url = "https://financialmodelingprep.com/api/v3/company/stock/list") %>%
               filter(symbol %in% SP500$symbol) %>% #Symbols of SP500
               filter(!symbol %in% c("J","AMCR")) #Companies that doesn't have data from API and causes


glimpse(Stock_Lists)
```

```
## Observations: 502
## Variables: 4
## $ symbol   <chr> "CMCSA", "KMI", "INTC", "MU", "GE", "BAC", "AAPL", "MSFT",...
## $ name     <chr> "Comcast Corporation Class A Common Stock", "Kinder Morgan...
## $ price    <dbl> 38.22, 12.64, 50.08, 37.38, 7.08, 21.98, 252.86, 146.57, 1...
## $ exchange <chr> "Nasdaq Global Select", "New York Stock Exchange", "Nasdaq...
```

Project Data

From the API structure the data required for this project is: 1. Segments: Data with information of sectors and industries segments of stocks 2. PriceSectors: Price of companies grouped by industries and sectors segments 3. KeyMetrics: Key financial metrics of stock market and companies 4. Historical prices: Stock market prices of companies

After exhaustive analysis, the capacity of memory for this project is at 400 stock market symbols and because of that the 502 stocks will be reduced to 400 on each data

- Segments:

```
segments <- Stock_Lists[1:400, ] %>% #Filter data for memory capacity
            mutate(Company_Profile = map(symbol, ~GetCompanyProfile(API_Structure[2,4], company = ..1)))
            select(Company_Profile) %>% #Select nested list
            unnest() %>% # Unnest it
            mutate(industry = case_when(industry == "" ~ "Funds", TRUE ~ industry), #Set sectors and in
                   sector = case_when(sector == "" ~ "Funds", TRUE ~ sector)) %>%
            select(symbol, companyName, industry, sector) #Select the data required for this dataframe
```

```
glimpse(segments)
```

```
## Observations: 400
## Variables: 4
## $ symbol      <chr> "CMCSA", "KMI", "INTC", "MU", "GE", "BAC", "AAPL", "MSF...
## $ companyName <chr> "Comcast Corporation Class A Common Stock", "Kinder Mor...
## $ industry    <chr> "Entertainment", "Oil & Gas - Midstream", "Semiconducto...
## $ sector      <chr> "Consumer Cyclical", "Energy", "Technology", "Technolog...
```

- PriceSectors:

```
PriceSectors <- Stock_Lists[1:400, ] %>% #Filter data for memory capacity
                mutate(Company_Profile = map(symbol, ~GetCompanyProfile(API_Structure[2,4],
                                                                         company = ..1))) %>% #Get Data f
                select(Company_Profile) %>% #Select nested list
                unnest() %>% # unnest it
                mutate(industry = case_when(industry == "" ~ "Funds", TRUE ~ industry),  #Set sectors a
                       sector = case_when(sector == "" ~ "Funds", TRUE ~ sector))

glimpse(PriceSectors)
```

```
## Observations: 400
## Variables: 17
## $ symbol            <chr> "CMCSA", "KMI", "INTC", "MU", "GE", "BAC", "AAPL"...
## $ price             <dbl> 38.22, 12.64, 50.08, 37.38, 7.08, 21.98, 252.86, ...
## $ beta              <chr> "1.061551", "0.75548", "0.90978", "1.951096", "1....
## $ volAvg            <chr> "25631996", "16375424", "27534578", "27161027", "...
## $ mktCap            <chr> "1.74016823E11", "2.86303601E10", "2.14192161E11"...
## $ lastDiv           <chr> "0.84", "0.8", "1.26", "0", "0.04", "0.6", "2.92"...
## $ range             <chr> "34.44-47.74", "12.32-22.58", "42.86-69.29", "32....
## $ changes           <dbl> 2.18, -0.09, 5.47, 2.91, 0.42, 1.54, 10.65, 11.15...
## $ changesPercentage <chr> "(+6.05%)", "(-0.71%)", "(+12.26%)", "(+8.44%)", ...
## $ companyName       <chr> "Comcast Corporation Class A Common Stock", "Kind...
## $ exchange          <chr> "Nasdaq Global Select", "New York Stock Exchange"...
## $ industry          <chr> "Entertainment", "Oil & Gas - Midstream", "Semico...
## $ website           <chr> "https://corporate.comcast.com", "http://www.kind...
## $ description       <chr> "Comcast Corp is a media and technology company. ...
## $ ceo               <chr> "Brian L. Roberts", "Steven J. Kean", "Brian M. K...
## $ sector            <chr> "Consumer Cyclical", "Energy", "Technology", "Tec...
## $ image             <chr> "https://financialmodelingprep.com/images-New-jpg...
```

- KeyMetrics:

Since there are 57 metrics in the API dataset, a description of each metric and measure formula was created

```
#metrics
path <- "Market KeyMetrics.xlsx"

Metrics_Info <- path %>%
                excel_sheets() %>%
                set_names() %>%
                map(read_excel, path = path)


kable(head(Metrics_Info$KeyMetrics), caption = "10 Metrics info") %>%
  kable_styling(full_width = F)
```

Table 2: 10 Metrics info

Segment	Metric	Explanation
Fundamental	Revenue per Share	Ratio that computes the total revenue earned per share over a desi
Income Statement	Net Income per Share	How much of a firm's net income was to each share of common sto
Fundamental	Operating Cash Flow per Share	Company's ability to generate cash
Fundamental	Free Cash Flow per Share	How much cash a business generates after accounting for capital ex
Fundamental	Cash per Share	Available cash to a business divided by the number of equity shares
Fundamental	Book Value per Share	Value of allshares divided by the number of shares issued.

```
KeyMetrics <- Stock_Lists[1:400, ] %>% #Filter data for memory capacity
              mutate(Company_Key_Metrics = map(symbol, ~GetData(API_Structure[8,4], company = ..1))) %>%
              select(symbol, name, Company_Key_Metrics) %>% #Select data and nested API data
              unnest(Company_Key_Metrics) %>% # Unnest it
              gather(key = "metric", value = "value", -symbol, -date, -name) %>% # Pivot the metrics pe
              inner_join(segments, by = "symbol") %>% #Get segments data to enrich the dataset
              inner_join(Metrics_Info$KeyMetrics, by = c("metric"="Metric")) %>% #Get the description a
              select(-companyName) %>% # Remove duplicate columns
              mutate(value = as.double(value), date = anydate(date)) %>% # Fix data structure
              group_by(metric, Explanation, Formula) %>% # Nest data per metric
              nest()


glimpse(KeyMetrics)
```

```
## Observations: 57
## Variables: 4
## Groups: metric, Explanation, Formula [57]
## $ metric      <chr> "Revenue per Share", "Net Income per Share", "Operating...
## $ Explanation <chr> "Ratio that computes the total revenue earned per share...
## $ Formula     <chr> "Total revenue / shares", "Net income / average outstan...
## $ data        <list<df[,7]>> CMCSA                                          ...
```

- Historical prices:

```
HistoricalPrices <- Stock_Lists[1:400, ] %>% #Filter data for memory capacity
                    #Get Data per symbo
                    mutate(Historical_Daily_Price = map(symbol, ~GetData(API_Structure[11,4],
                                                                         company = ..1) %>%
                                                                 mutate(date = anytime(date)))) %>%
                    #Adjust monthly price
                    mutate(Monthly_AdjPrice = map(Historical_Daily_Price,  ~..1 %>%
                                                    tq_transmute(select = close,
                                                                 mutate_fun = to.monthly,
                                                                 indexAt = "lastof"))) %>%
                    select(-price) %>% # Remove duplicated column
                    inner_join(segments, by = "symbol") %>% #Enrich dataframe with segments data
                    select(symbol:exchange, industry:sector, everything(), -companyName) #Select and or


glimpse(HistoricalPrices)
```

```
## Observations: 400
## Variables: 7
```

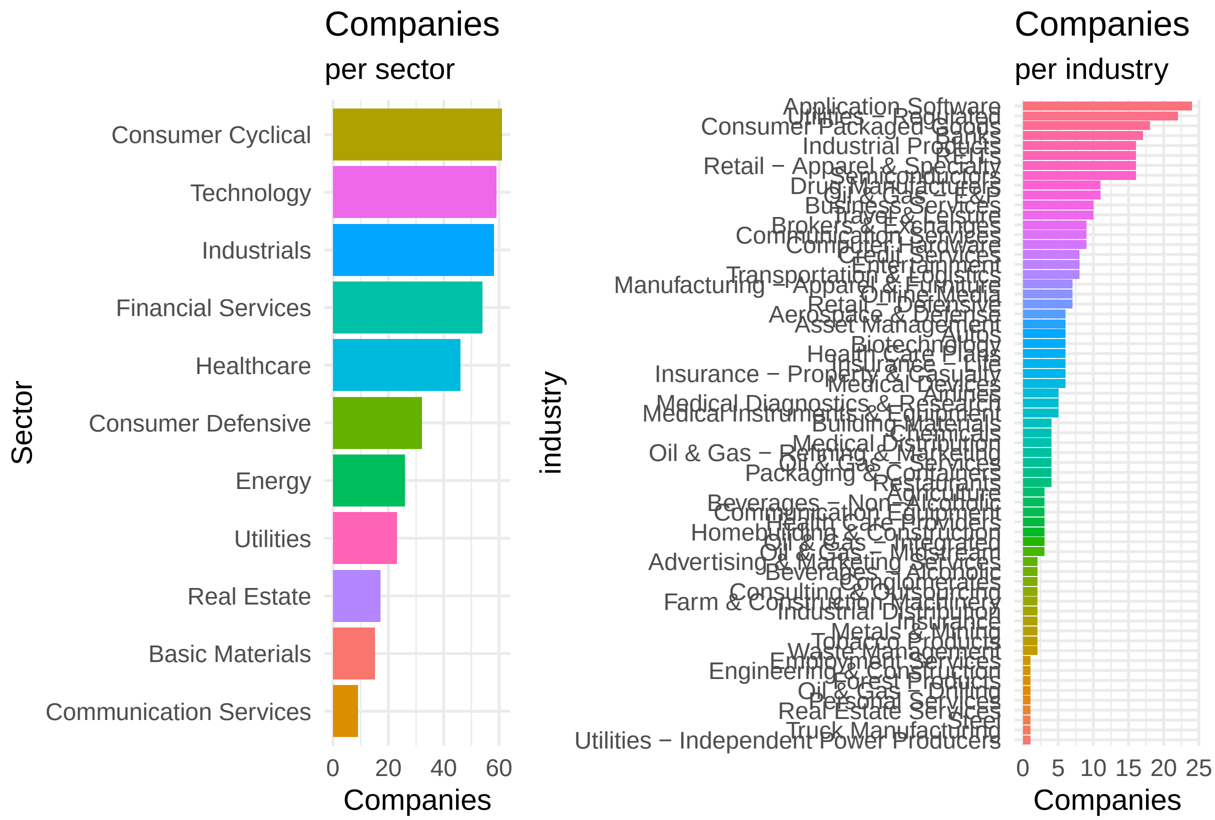
```
## $ symbol                <chr> "CMCSA", "KMI", "INTC", "MU", "GE", "BAC", "...
## $ name                  <chr> "Comcast Corporation Class A Common Stock", ...
## $ exchange              <chr> "Nasdaq Global Select", "New York Stock Exch...
## $ industry              <chr> "Entertainment", "Oil & Gas - Midstream", "S...
## $ sector                <chr> "Consumer Cyclical", "Energy", "Technology",...
## $ Historical_Daily_Price <list> [<tbl_df[1259 x 13]>, <tbl_df[1259 x 13]>, ...
## $ Monthly_AdjPrice      <list> [<tbl_df[61 x 2]>, <tbl_df[61 x 2]>, <tbl_d...
```

Overview of data

Industry & Sector

Let's check the amount of companies per sector and industry segments

```r
p1 <- segments %>%
      mutate(industry = fct_rev(fct_infreq(sector))) %>%
      ggplot() +
      aes(x = industry, fill = sector) +
      geom_bar() +
      coord_flip() +
      scale_fill_hue() +
      guides(fill = "none") +
      theme_minimal()+
      labs(title = "Companies", subtitle = "per sector", y = "Companies", x = "Sector")

p2 <- segments %>%
      mutate(industry = fct_rev(fct_infreq(industry))) %>%
      ggplot() +
      aes(x = industry, fill = industry) +
      geom_bar() +
      coord_flip() +
      scale_fill_hue() +
      guides(fill = "none") +
      theme_minimal() +
      labs(title = "Companies", subtitle = "per industry", y = "Companies")

p1 | p2
```



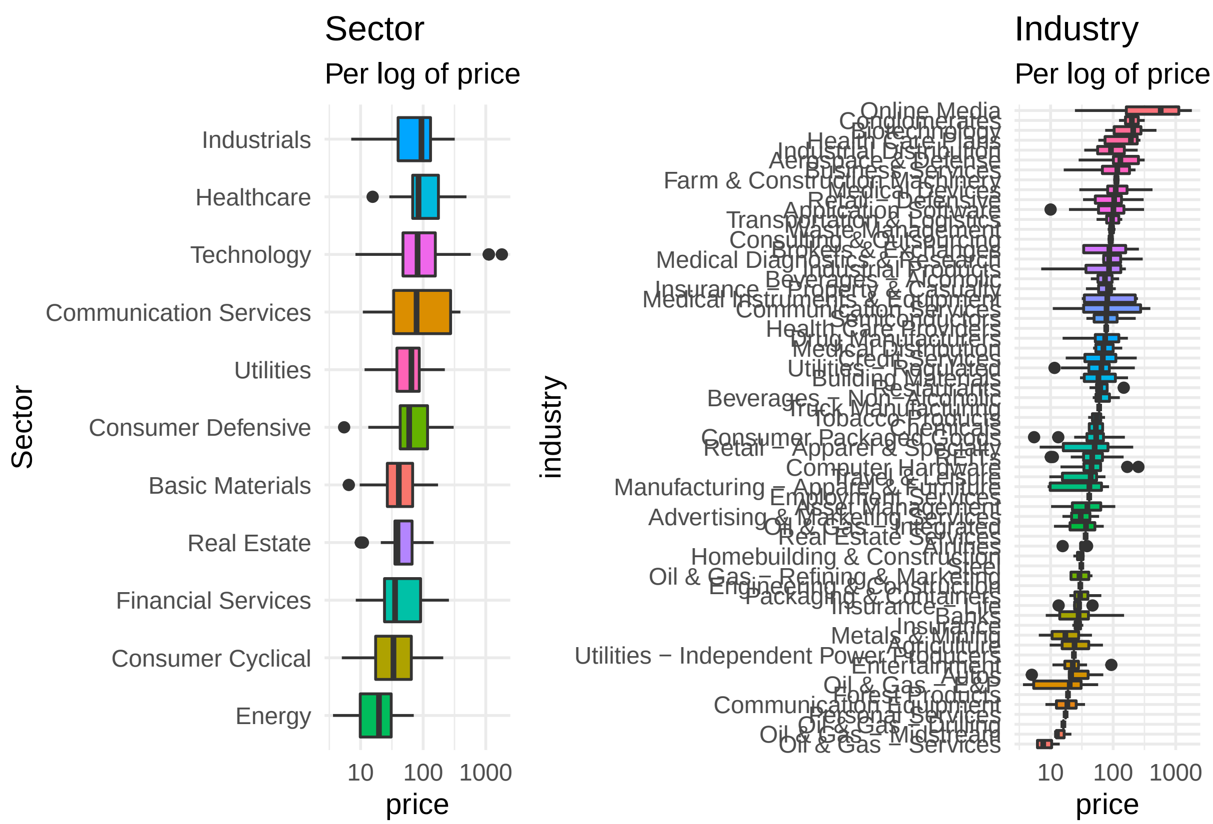
Some discoveries on SP500 companies segments:

1. The maximum amount of companies per sector is around 60 and it seems to be centered around 6 sectors and skewed to the right which could relate to some sectors having more companies in average or being more advantageous to the portfolio.
2. Industries segments shows that 6 sectors are mixed in the amount of companies, having software, consumer packaged goods and banks as top amount of companies in SP500

Now let's see how does price is distributed per segments:

```
p1 <- PriceSectors %>%
      mutate(industry = fct_reorder(sector, price)) %>%
      ggplot() +
      aes(x = industry, y = price, fill = sector) +
      geom_boxplot() +
      scale_y_log10() +
      coord_flip() +
      scale_fill_hue() +
      guides(fill = "none") +
      theme_minimal() +
      labs(title = "Sector", subtitle = "Per log of price", x = "Sector")

p2 <- PriceSectors %>%
      mutate(industry = fct_reorder(industry, price)) %>%
      ggplot() +
      aes(x = industry, y = price, fill = industry) +
      geom_boxplot() +
```

```
    scale_y_log10() +
    coord_flip() +
    scale_fill_hue() +
    guides(fill = "none") +
    theme_minimal()+
    labs(title = "Industry", subtitle = "Per log of price")

p1 | p2
```

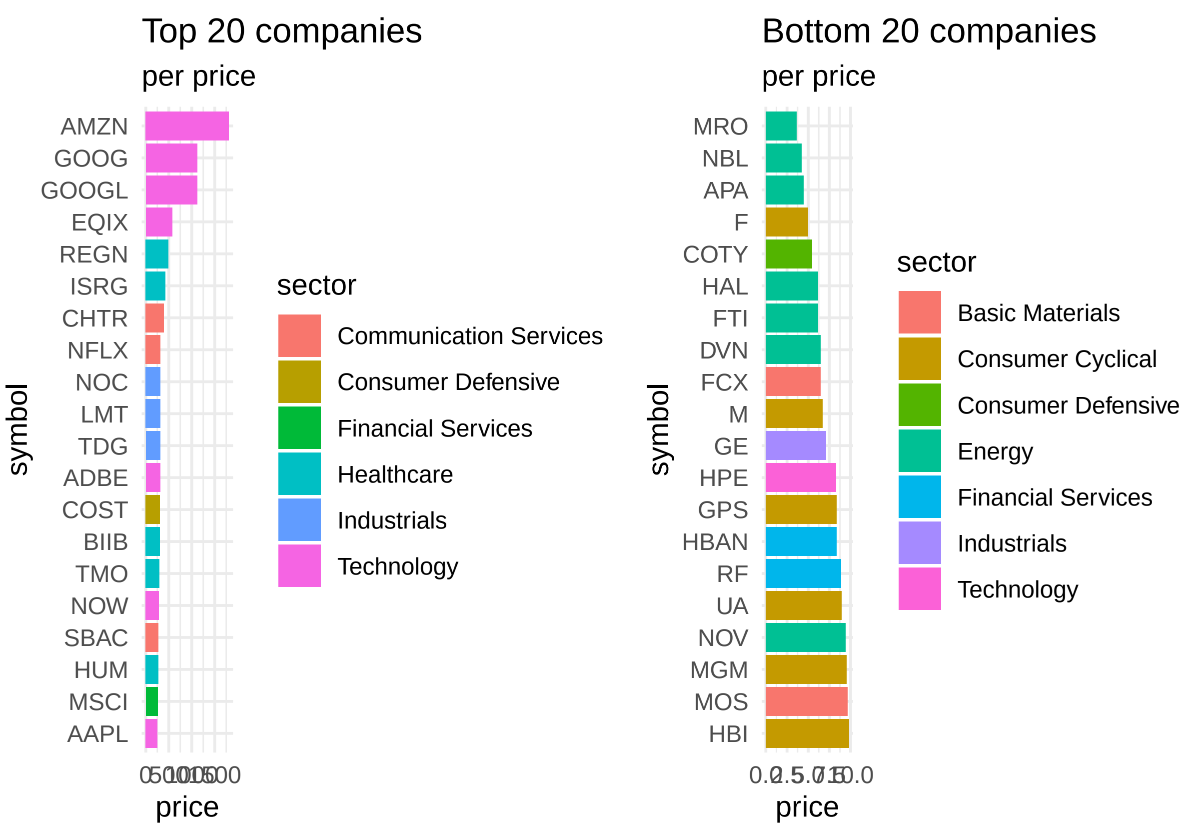


It seems that price per sector is usually around at the same mean with some variability, this variabiity is explained by the huge aount of difference on industry segment. This means that industry segment is a better metric to evaluate the dispersion of price rather than sectors.

Let's check the top/bottom 20 companies prices

```
p1 <- PriceSectors %>%
    mutate(price = as.double(price)) %>%
    arrange(-price) %>%
    head(20) %>%
    mutate(symbol = fct_reorder(symbol, price)) %>%
    ggplot() +
    aes(x = symbol, y = price, fill = sector) +
    geom_col() +
    coord_flip() +
    scale_fill_hue() +
    theme_minimal() +
    labs(title = "Top 20 companies", subtitle = "per price")
```

```r
p2 <- PriceSectors %>%
      mutate(price = as.double(price)) %>%
      filter(price > 0) %>%
      arrange(price) %>%
      head(20) %>%
      mutate(symbol = fct_rev(fct_reorder(symbol, price))) %>%
      ggplot() +
      aes(x = symbol, y = price, fill = sector) +
      geom_col() +
      coord_flip() +
      scale_fill_hue() +
      theme_minimal() +
      labs(title = "Bottom 20 companies", subtitle = "per price")

p1 | p2
```

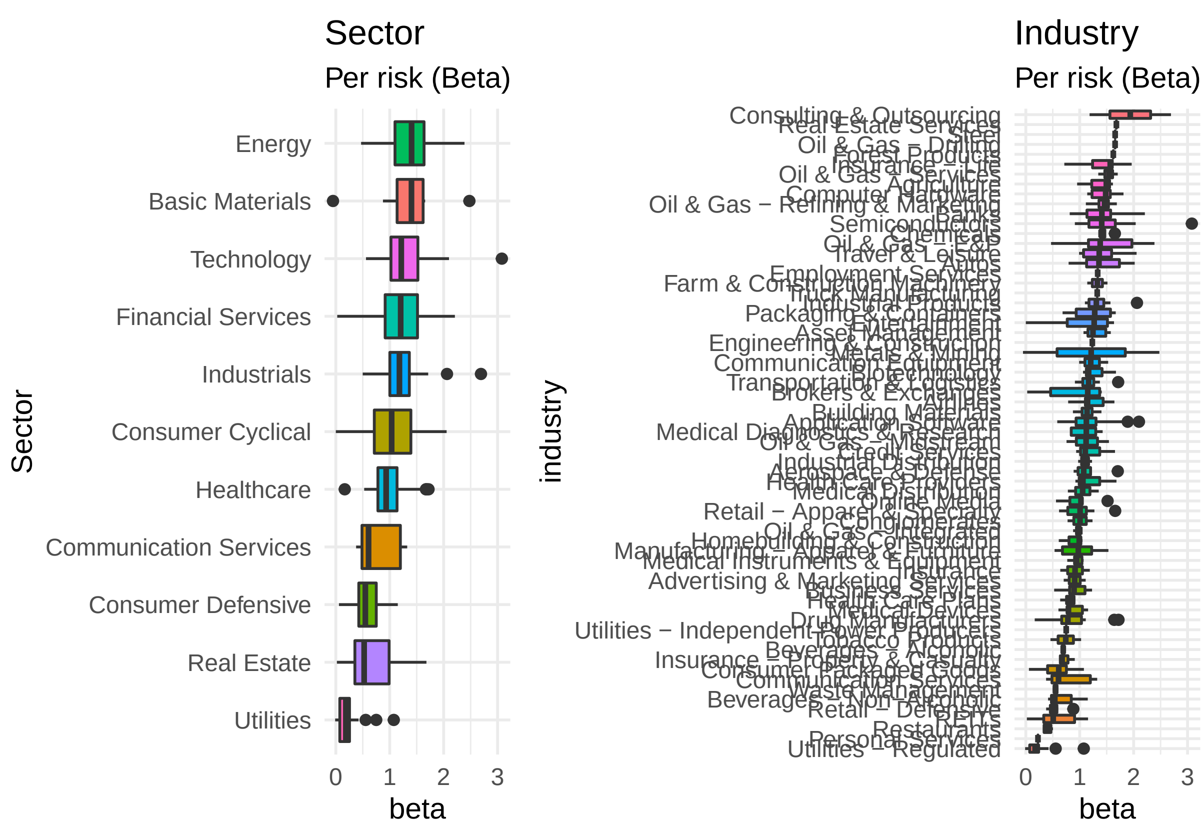


It does seem that those 6 sectors variability grants them in general the top 20 and bottom companies price

Now let's see how does risk is distributed per segments:

```r
p1 <- PriceSectors %>%
      mutate(beta = as.double(beta)) %>%
      mutate(industry = fct_reorder(sector, beta)) %>%
      ggplot() +
      aes(x = industry, y = beta, fill = sector) +
      geom_boxplot() +
```

```
      coord_flip() +
      scale_fill_hue() +
      guides(fill = "none") +
      theme_minimal() +
      labs(title = "Sector", subtitle = "Per risk (Beta)", x = "Sector")

p2 <- PriceSectors %>%
      mutate(beta = as.double(beta)) %>%
      mutate(industry = fct_reorder(industry, beta)) %>%
      ggplot() +
      aes(x = industry, y = beta, fill = industry) +
      geom_boxplot() +
      coord_flip() +
      scale_fill_hue() +
      guides(fill = "none") +
      theme_minimal() +
      labs(title = "Industry", subtitle = "Per risk (Beta)")


p1 | p2
```



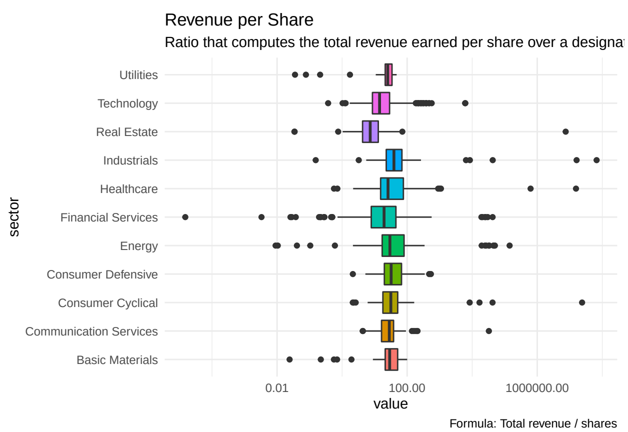
It seems that risk per sector changes slightly and it's quite impressive to see that energy and basic materials are on the top risk sector. On industry side, it seems some oindustries have a lot of variability on risk and consulting & outsourcing industry is on the top, it seems counterintuitive, I was expected to see financial services and technology being riskier.

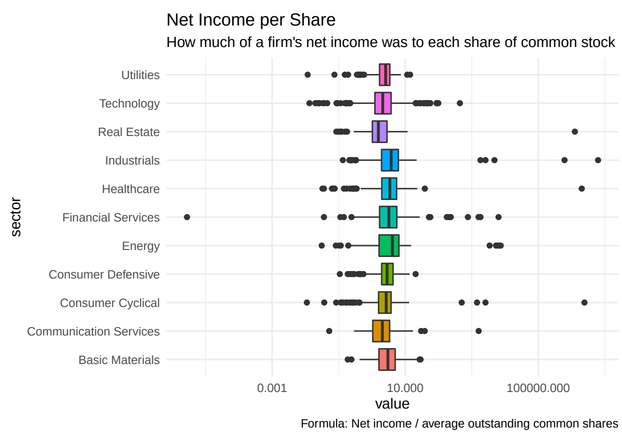
KeyMetrics

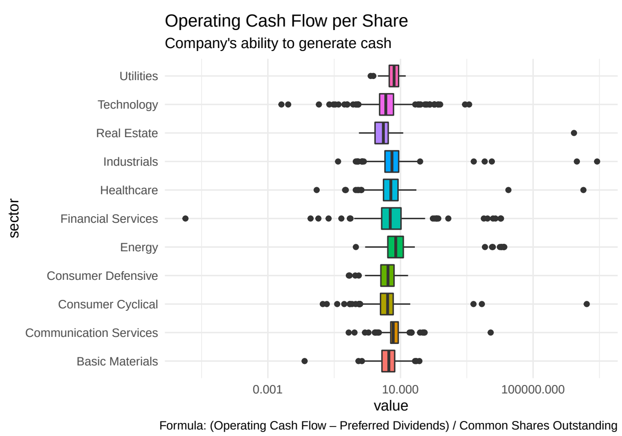
Since there are 57 metrics, it's important to create a ggplot function and use it in each sector, this function will include in the labs of plot the brief explanation and the formula in order to help understand each metric

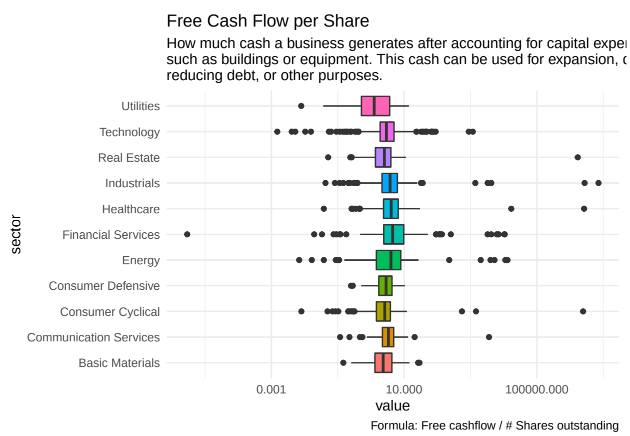
```
plots <- function(data, metric, Explanation, Formula){
    ggplot(data) +
      aes(x = sector, y = value, fill = sector) +
      geom_boxplot() +
      scale_fill_hue() +
      scale_y_continuous(trans = "log10") +
      theme_minimal() +
      coord_flip() +
      guides(fill = "none") +
      labs(title = metric, subtitle = Explanation, caption = paste0("Formula: ",Formula))
}
```

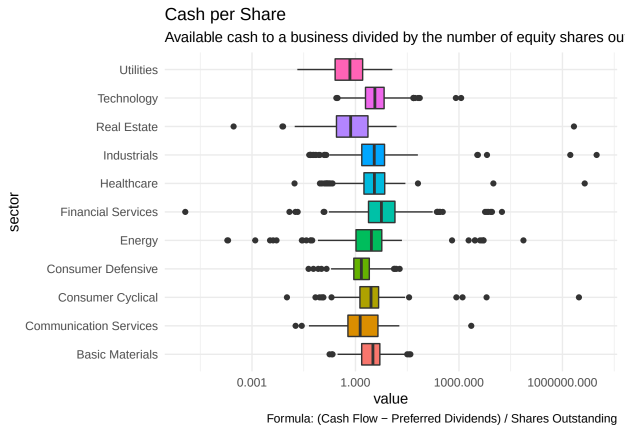
- Plots:

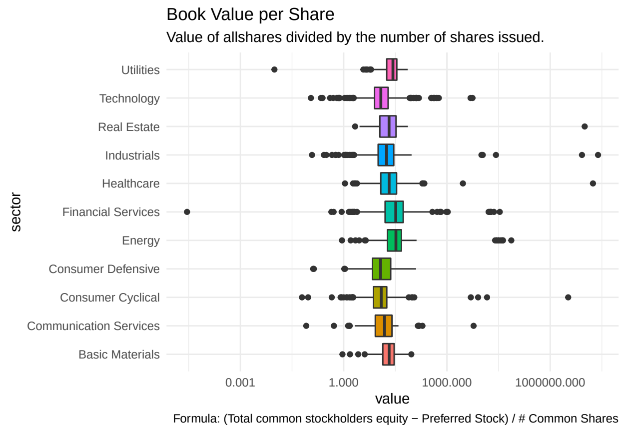


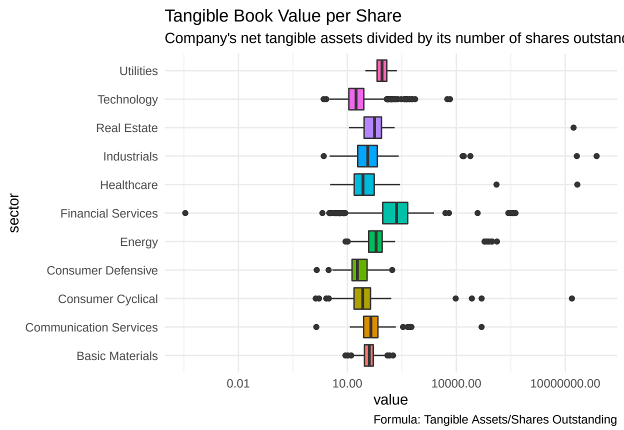


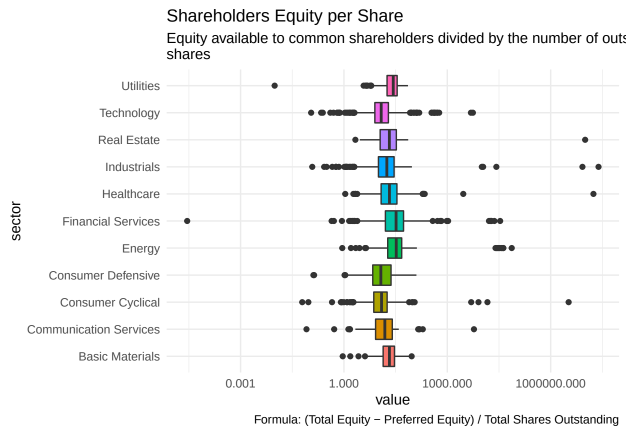


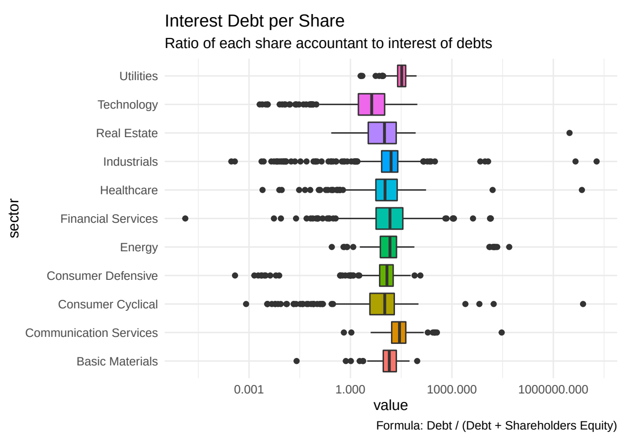


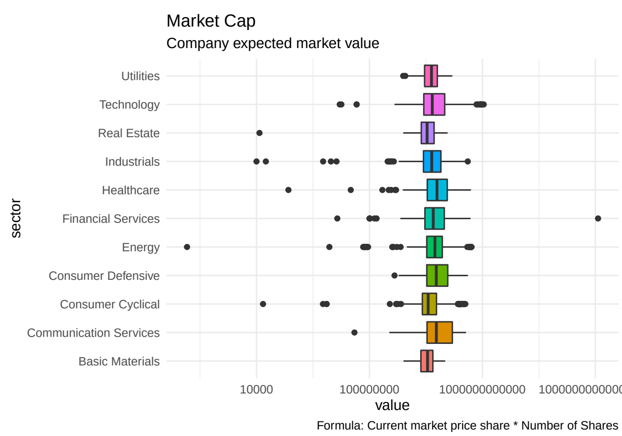

Cash per Share
Available cash to a business divided by the number of equity shares ou
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
0.001
1.000
1000.000
1000000.000
value
Formula: (Cash Flow − Preferred Dividends) / Shares Outstanding

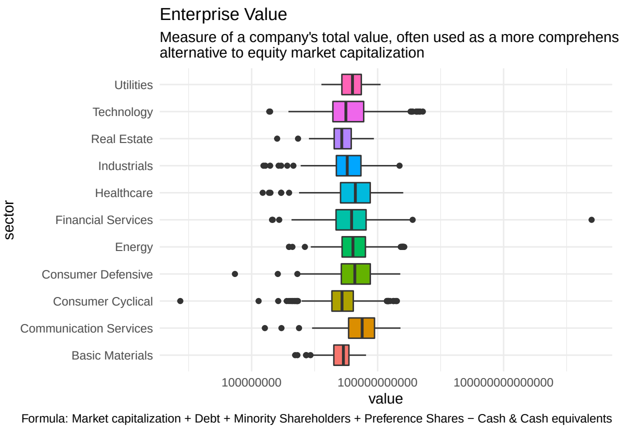

Book Value per Share
Value of allshares divided by the number of shares issued.
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
0.001
1.000
1000.000
1000000.000
value
Formula: (Total common stockholders equity − Preferred Stock) / # Common Shares

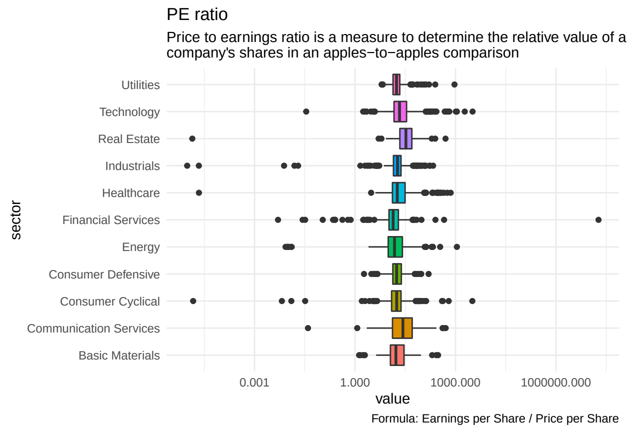

Tangible Book Value per Share
Company's net tangible assets divided by its number of shares outstanc
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
0.01
10.00
10000.00
10000000.00
value
Formula: Tangible Assets/Shares Outstanding

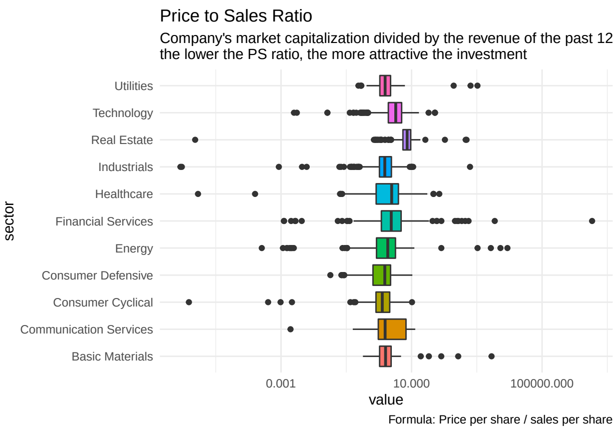

Shareholders Equity per Share
Equity available to common shareholders divided by the number of outs
shares
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
0.001
1.000
1000.000
1000000.000
value
Formula: (Total Equity − Preferred Equity) / Total Shares Outstanding

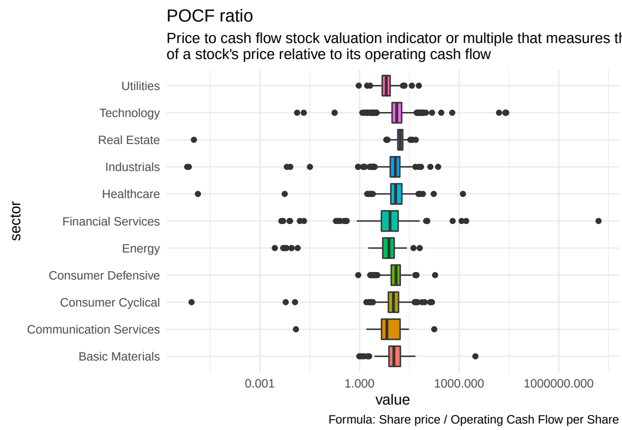

Interest Debt per Share
Ratio of each share accountant to interest of debts
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
0.001
1.000
1000.000
1000000.000
value
Formula: Debt / (Debt + Shareholders Equity)

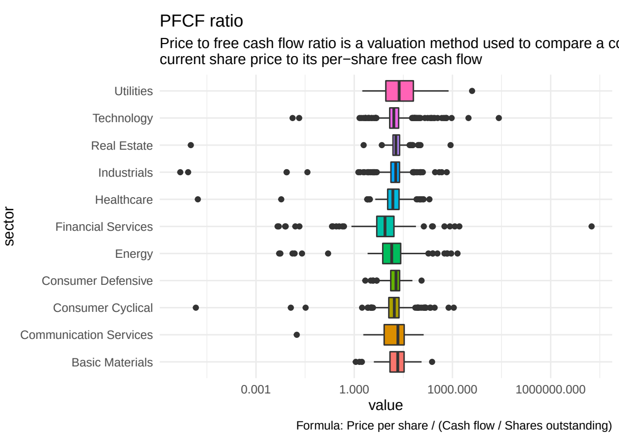

Market Cap
Company expected market value
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
10000
100000000
1000000000000
1000000000000
value
Formula: Current market price share * Number of Shares

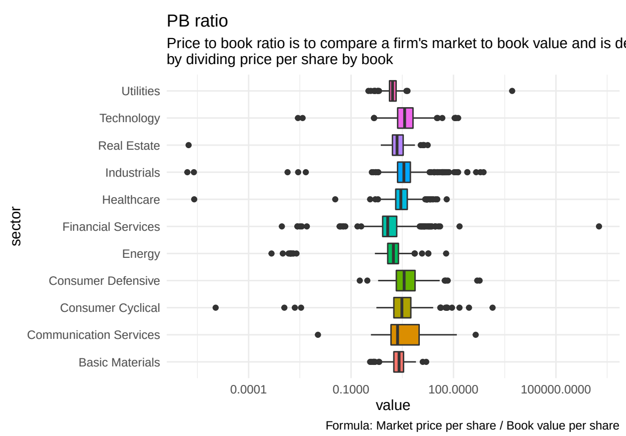

Enterprise Value
Measure of a company's total value, often used as a more comprehens
alternative to equity market capitalization
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
100000000
100000000000
100000000000000
value
Formula: Market capitalization + Debt + Minority Shareholders + Preference Shares − Cash & Cash equivalents

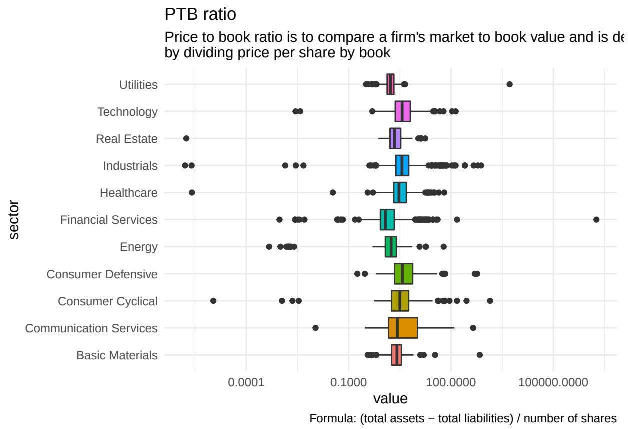

PE ratio
Price to earnings ratio is a measure to determine the relative value of a
company's shares in an apples−to−apples comparison
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
1000000.000
value
Formula: Earnings per Share / Price per Share

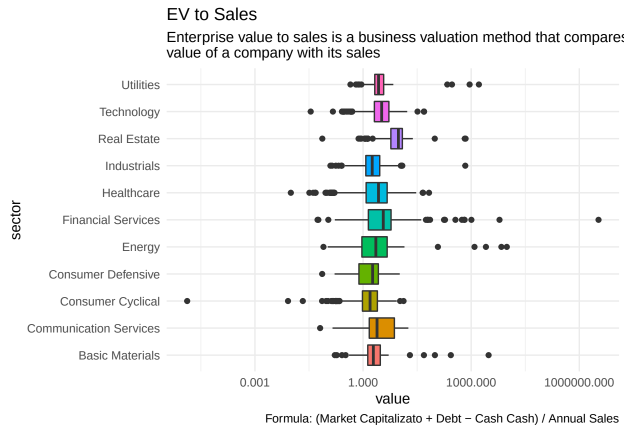

Price to Sales Ratio
Company's market capitalization divided by the revenue of the past 12 r
the lower the PS ratio, the more attractive the investment
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
10.000
100000.000
value
Formula: Price per share / sales per share

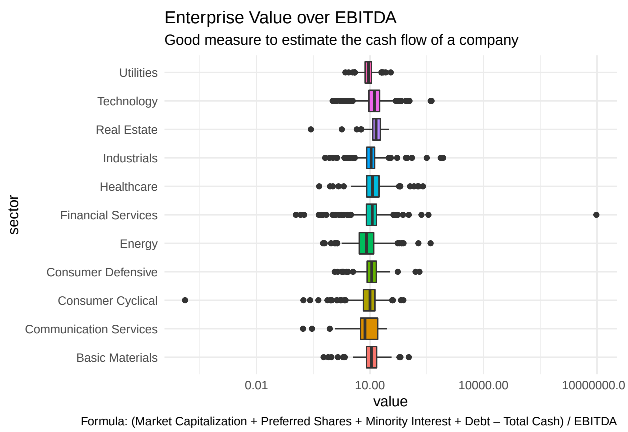

POCF ratio
Price to cash flow stock valuation indicator or multiple that measures th
of a stock's price relative to its operating cash flow
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
1000000.000
value
Formula: Share price / Operating Cash Flow per Share

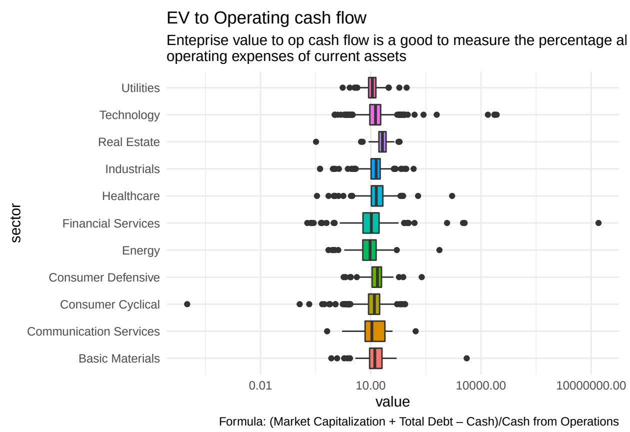

PFCF ratio
Price to free cash flow ratio is a valuation method used to compare a cc
current share price to its per−share free cash flow
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
1000000.000
value
Formula: Price per share / (Cash flow / Shares outstanding)

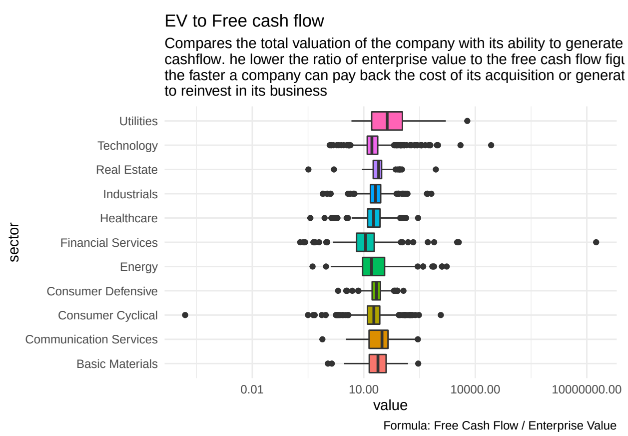

PB ratio
Price to book ratio is to compare a firm's market to book value and is de
by dividing price per share by book
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.0001
0.1000
100.0000
100000.0000
value
Formula: Market price per share / Book value per share

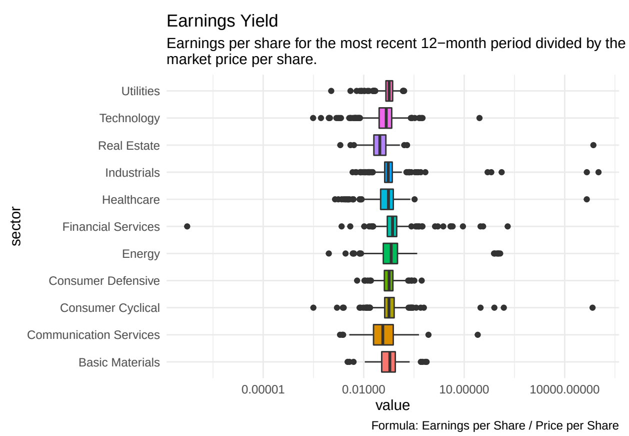

PTB ratio
Price to book ratio is to compare a firm's market to book value and is de
by dividing price per share by book
sector
0.0001
0.1000
100.0000
100000.0000
value
Formula: (total assets − total liabilities) / number of shares

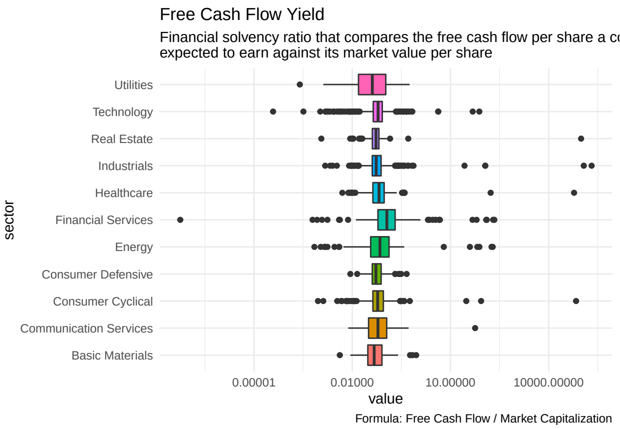

EV to Sales
Enterprise value to sales is a business valuation method that compares
value of a company with its sales
sector
0.001
1.000
1000.000
1000000.000
value
Formula: (Market Capitalizato + Debt − Cash Cash) / Annual Sales

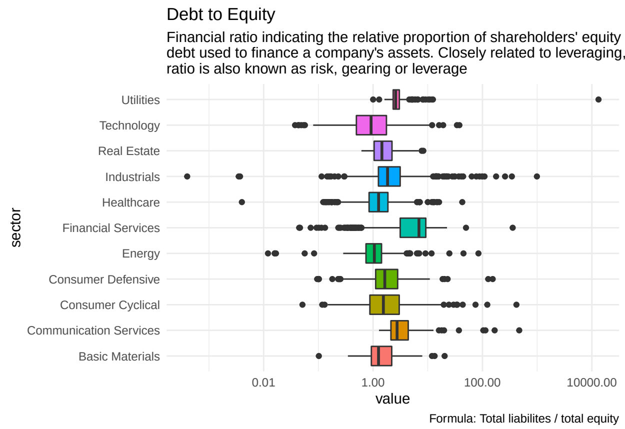

Enterprise Value over EBITDA
Good measure to estimate the cash flow of a company
sector
0.01
10.00
10000.00
10000000.0
value
Formula: (Market Capitalization + Preferred Shares + Minority Interest + Debt – Total Cash) / EBITDA

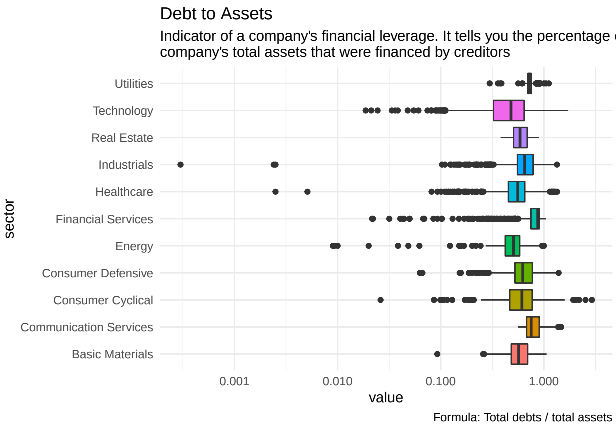

EV to Operating cash flow
Enteprise value to op cash flow is a good to measure the percentage al
operating expenses of current assets
sector
0.01
10.00
10000.00
10000000.00
value
Formula: (Market Capitalization + Total Debt – Cash)/Cash from Operations

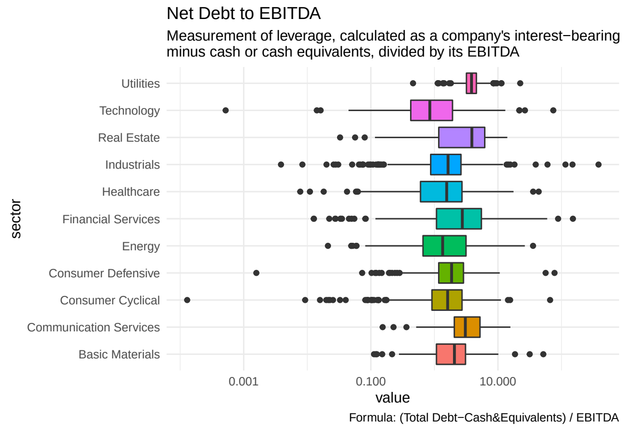

EV to Free cash flow
Compares the total valuation of the company with its ability to generate
cashflow. he lower the ratio of enterprise value to the free cash flow figu
the faster a company can pay back the cost of its acquisition or generat
to reinvest in its business
sector
0.01
10.00
10000.00
10000000.00
value
Formula: Free Cash Flow / Enterprise Value

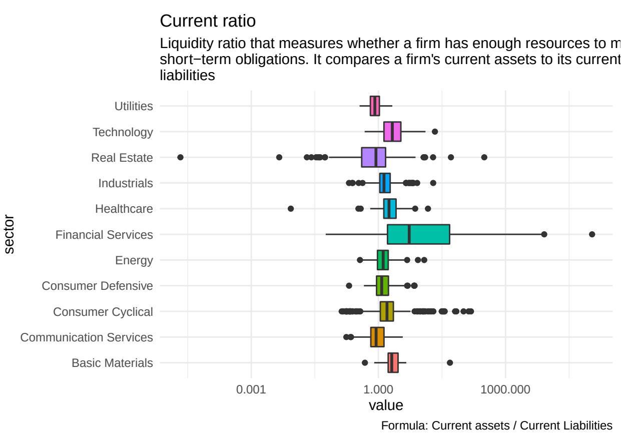

Earnings Yield
Earnings per share for the most recent 12−month period divided by the
market price per share.
sector
0.00001
0.01000
10.00000
10000.00000
value
Formula: Earnings per Share / Price per Share

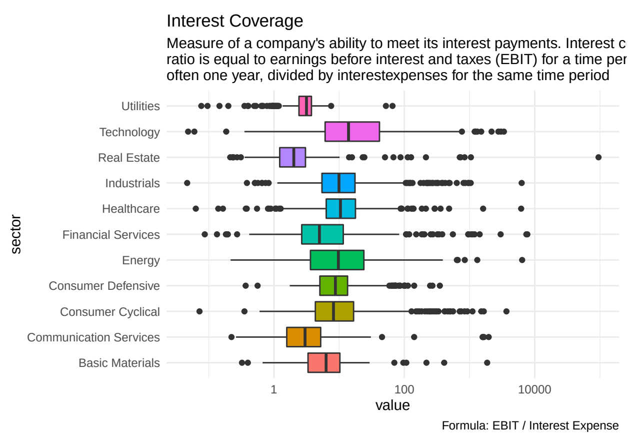

Free Cash Flow Yield
Financial solvency ratio that compares the free cash flow per share a co
expected to earn against its market value per share
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.00001
0.01000
10.00000
10000.00000
value
Formula: Free Cash Flow / Market Capitalization

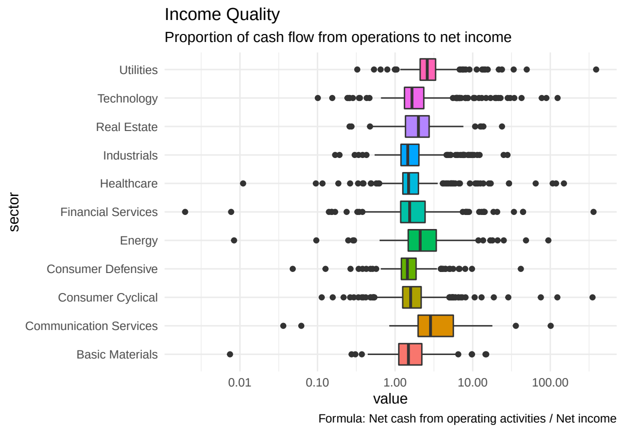

Debt to Equity
Financial ratio indicating the relative proportion of shareholders' equity a
debt used to finance a company's assets. Closely related to leveraging,
ratio is also known as risk, gearing or leverage
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
10000.00
value
Formula: Total liabilites / total equity

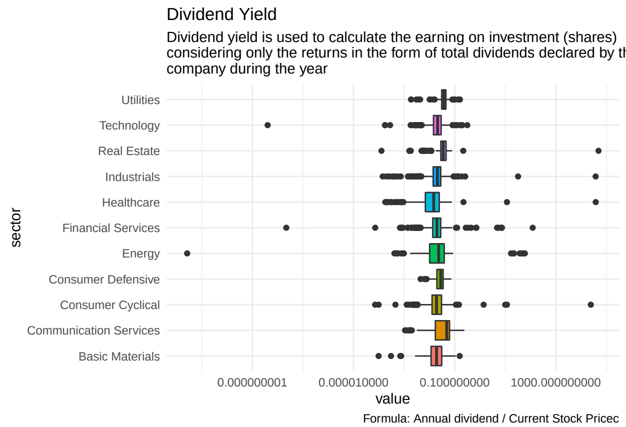

Debt to Assets
Indicator of a company's financial leverage. It tells you the percentage o
company's total assets that were financed by creditors
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
0.010
0.100
1.000
value
Formula: Total debts / total assets

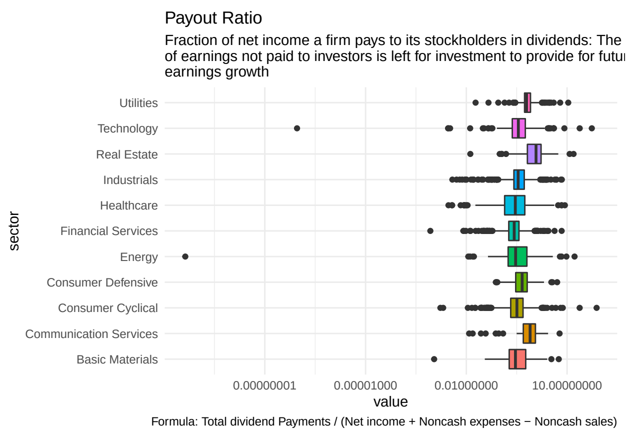

Net Debt to EBITDA
Measurement of leverage, calculated as a company's interest−bearing l
minus cash or cash equivalents, divided by its EBITDA
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
0.100
10.000
value
Formula: (Total Debt−Cash&Equivalents) / EBITDA

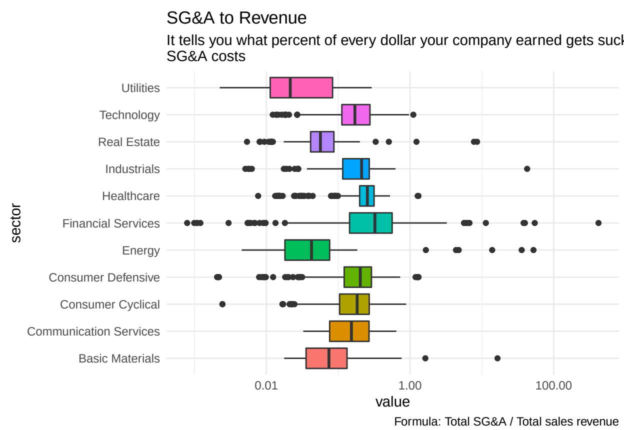

Current ratio
Liquidity ratio that measures whether a firm has enough resources to m
short−term obligations. It compares a firm's current assets to its current
liabilities
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
value
Formula: Current assets / Current Liabilities

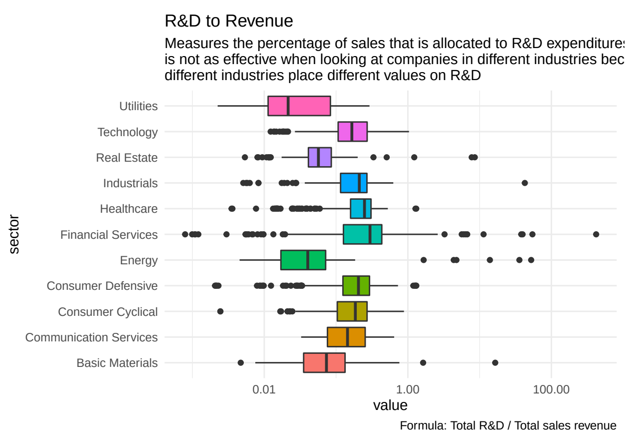

Interest Coverage
Measure of a company's ability to meet its interest payments. Interest c
ratio is equal to earnings before interest and taxes (EBIT) for a time per
often one year, divided by interestexpenses for the same time period
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
1
100
10000
value
Formula: EBIT / Interest Expense

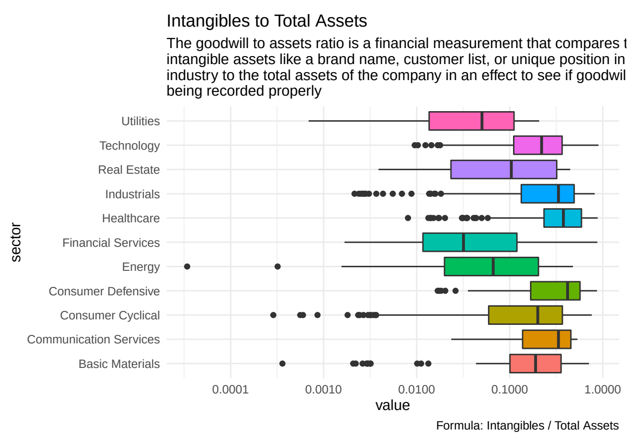

Income Quality
Proportion of cash flow from operations to net income
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
0.10
1.00
10.00
100.00
value
Formula: Net cash from operating activities / Net income

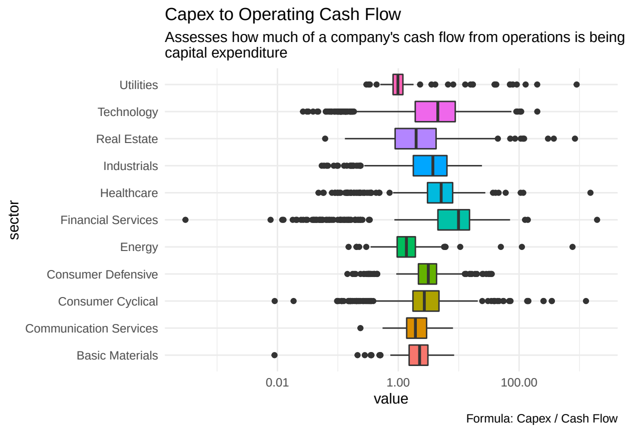

Dividend Yield
Dividend yield is used to calculate the earning on investment (shares) considering only the returns in the form of total dividends declared by th company during the year
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.000000001
0.000010000
0.100000000
1000.000000000
value
Formula: Annual dividend / Current Stock Pricec

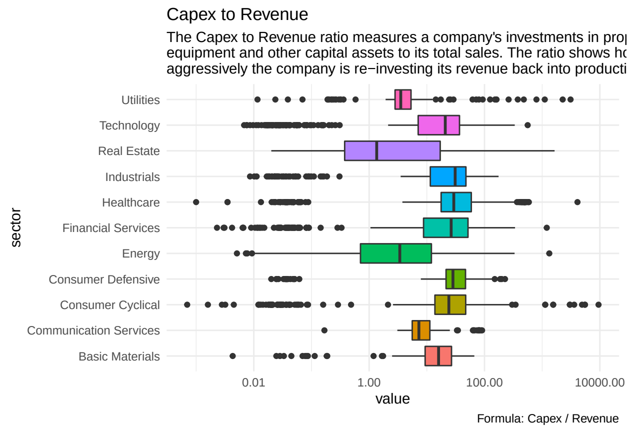

Payout Ratio
Fraction of net income a firm pays to its stockholders in dividends: The of earnings not paid to investors is left for investment to provide for futu earnings growth
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.00000001
0.00001000
0.01000000
10.00000000
value
Formula: Total dividend Payments / (Net income + Noncash expenses − Noncash sales)

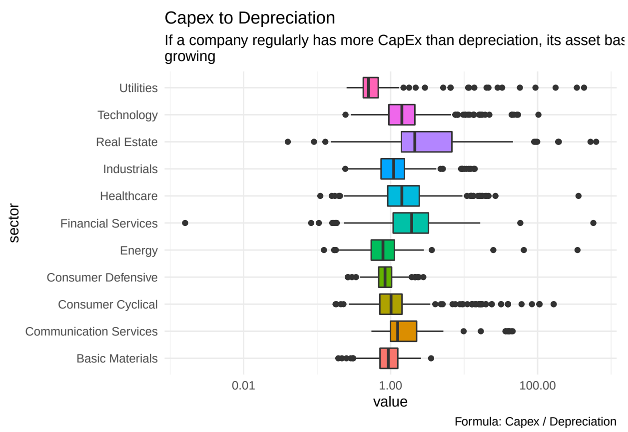

SG&A to Revenue
It tells you what percent of every dollar your company earned gets suck SG&A costs
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Total SG&A / Total sales revenue

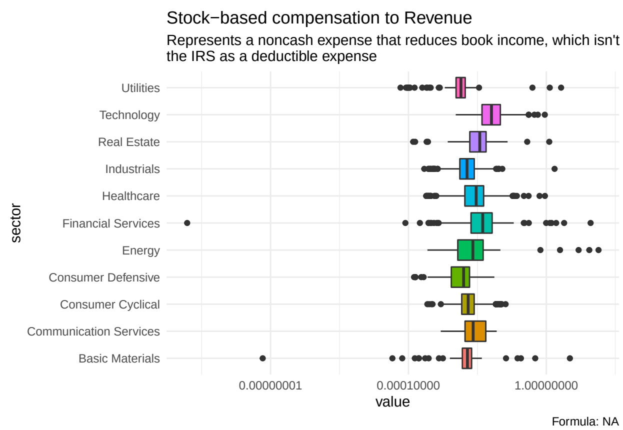

R&D to Revenue
Measures the percentage of sales that is allocated to R&D expenditure is not as effective when looking at companies in different industries bec different industries place different values on R&D
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Total R&D / Total sales revenue

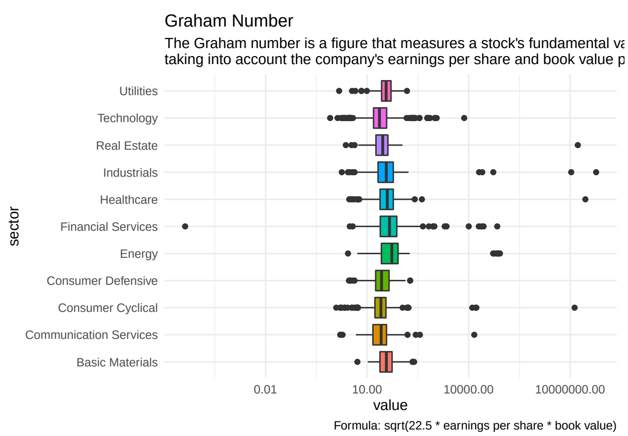

Intangibles to Total Assets
The goodwill to assets ratio is a financial measurement that compares t intangible assets like a brand name, customer list, or unique position in industry to the total assets of the company in an effect to see if goodwil being recorded properly
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.0001
0.0010
0.0100
0.1000
1.0000
value
Formula: Intangibles / Total Assets

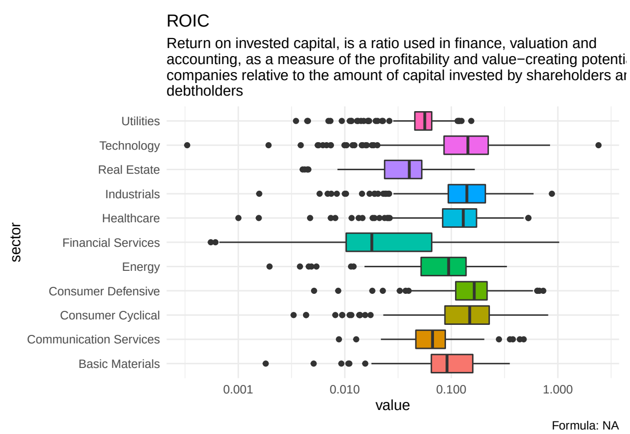

Capex to Operating Cash Flow
Assesses how much of a company's cash flow from operations is being
capital expenditure
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Capex / Cash Flow

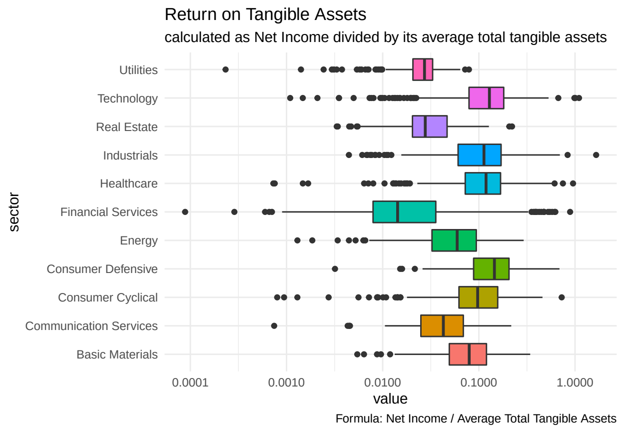

Capex to Revenue
The Capex to Revenue ratio measures a company's investments in prop
equipment and other capital assets to its total sales. The ratio shows ho
aggressively the company is re−investing its revenue back into producti
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
10000.00
value
Formula: Capex / Revenue

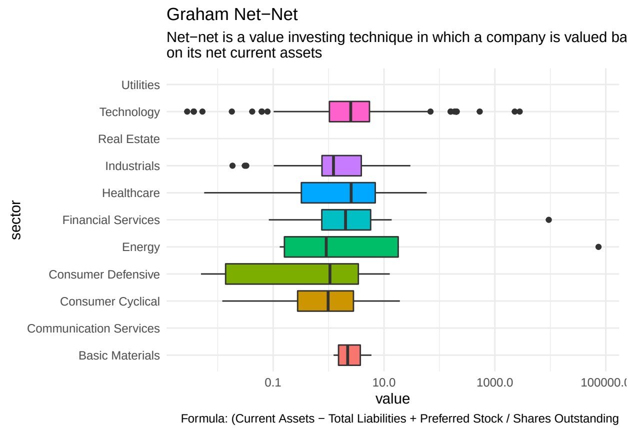

Capex to Depreciation
If a company regularly has more CapEx than depreciation, its asset bas
growing
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Capex / Depreciation

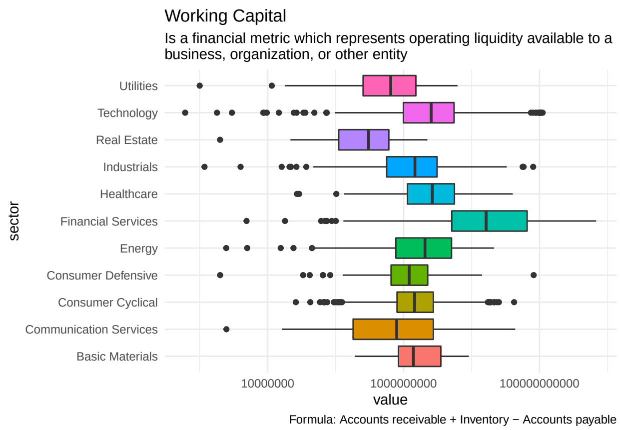

Stock−based compensation to Revenue
Represents a noncash expense that reduces book income, which isn't r
the IRS as a deductible expense
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.00000001
0.00010000
1.00000000
value
Formula: NA

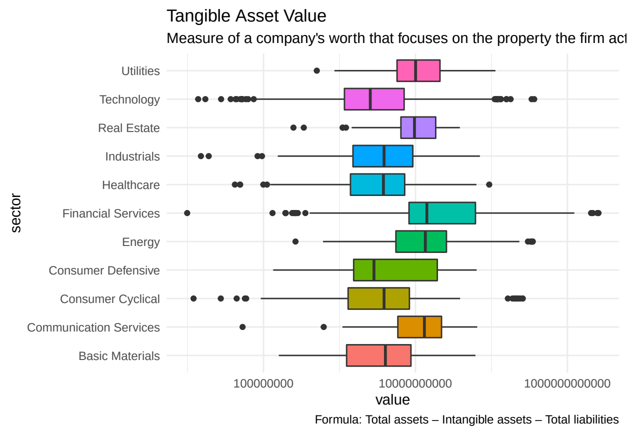

Graham Number
The Graham number is a figure that measures a stock's fundamental va
taking into account the company's earnings per share and book value p
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
10.00
10000.00
10000000.00
value
Formula: sqrt(22.5 * earnings per share * book value)

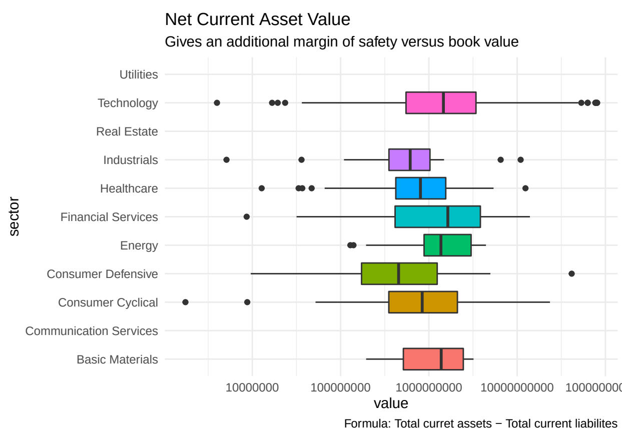

ROIC
Return on invested capital, is a ratio used in finance, valuation and
accounting, as a measure of the profitability and value−creating potenti
companies relative to the amount of capital invested by shareholders ar
debtholders
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
0.010
0.100
1.000
value
Formula: NA

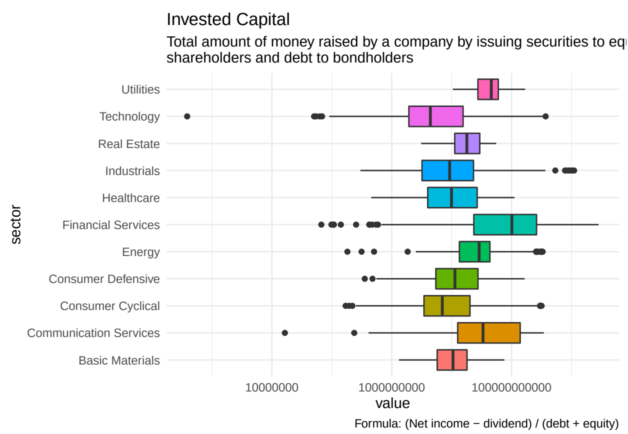

Return on Tangible Assets
calculated as Net Income divided by its average total tangible assets
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.0001
0.0010
0.0100
0.1000
1.0000
value
Formula: Net Income / Average Total Tangible Assets

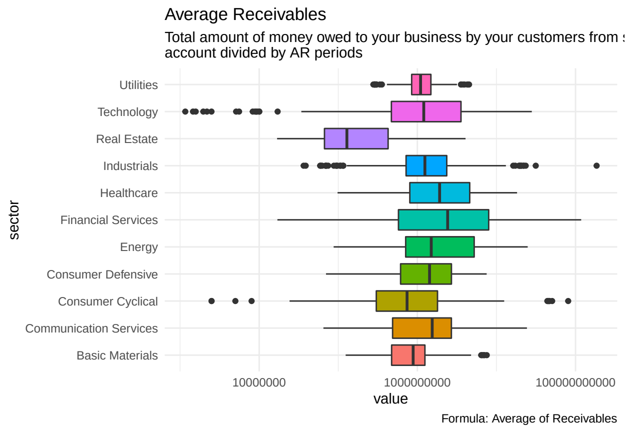

Graham Net−Net
Net−net is a value investing technique in which a company is valued ba
on its net current assets
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.1
10.0
1000.0
100000.0
value
Formula: (Current Assets − Total Liabilities + Preferred Stock / Shares Outstanding

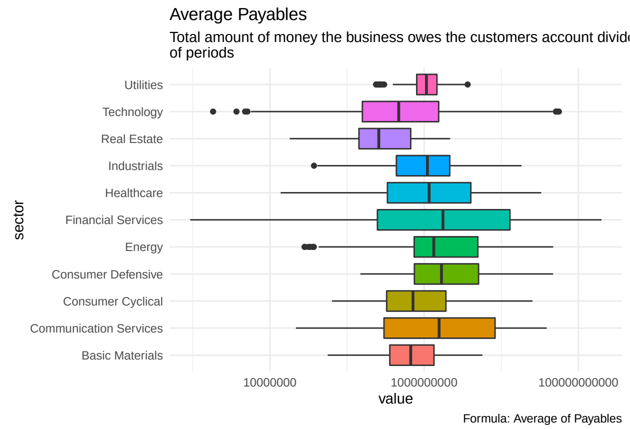

Working Capital
Is a financial metric which represents operating liquidity available to a
business, organization, or other entity
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
1000000000
100000000000
value
Formula: Accounts receivable + Inventory − Accounts payable

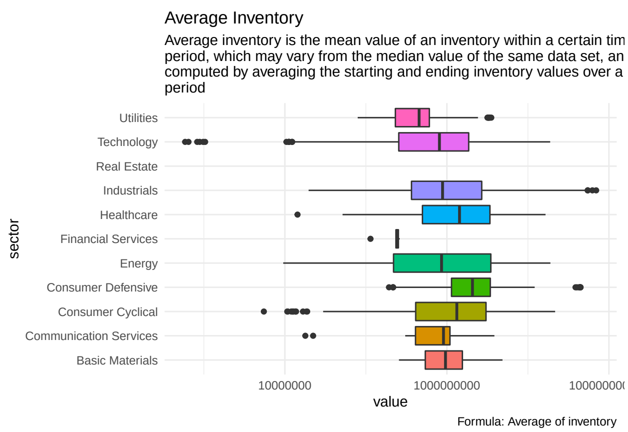

Tangible Asset Value
Measure of a company's worth that focuses on the property the firm act
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
100000000
10000000000
1000000000000
value
Formula: Total assets – Intangible assets – Total liabilities

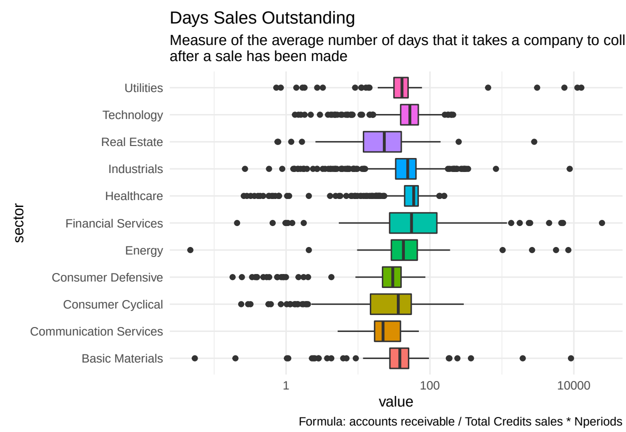

Net Current Asset Value
Gives an additional margin of safety versus book value
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
100000000
1000000000
10000000000
100000000
value
Formula: Total curret assets − Total current liabilites

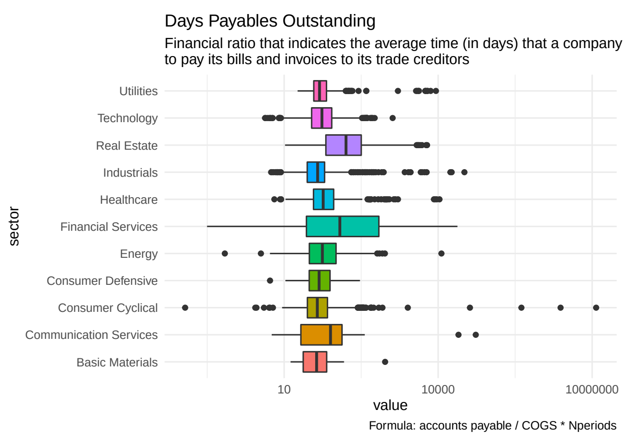

Invested Capital
Total amount of money raised by a company by issuing securities to eq
shareholders and debt to bondholders
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
1000000000
100000000000
value
Formula: (Net income − dividend) / (debt + equity)

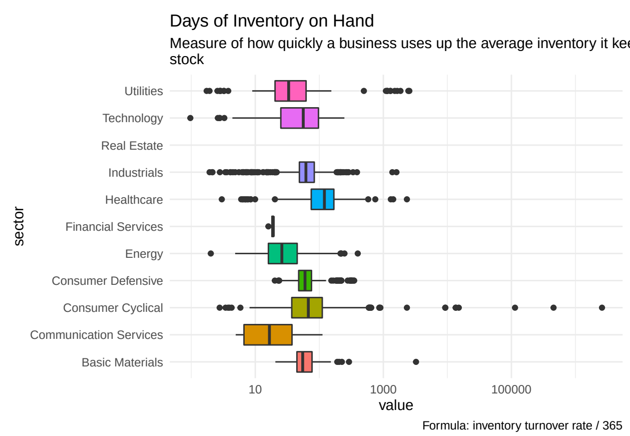

Average Receivables
Total amount of money owed to your business by your customers from s
account divided by AR periods
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
1000000000
100000000000
value
Formula: Average of Receivables

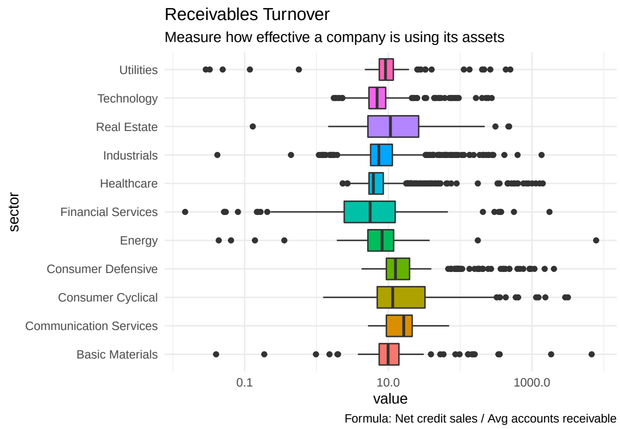

Average Payables
Total amount of money the business owes the customers account divide
of periods
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
1000000000
100000000000
value
Formula: Average of Payables

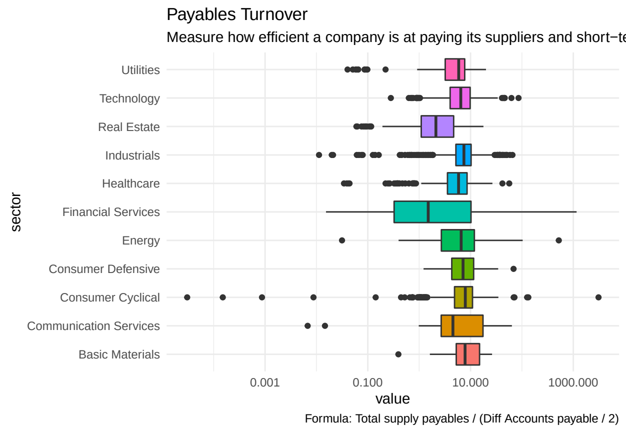

Average Inventory
Average inventory is the mean value of an inventory within a certain tim
period, which may vary from the median value of the same data set, an
computed by averaging the starting and ending inventory values over a
period
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10000000
1000000000
10000000
value
Formula: Average of inventory

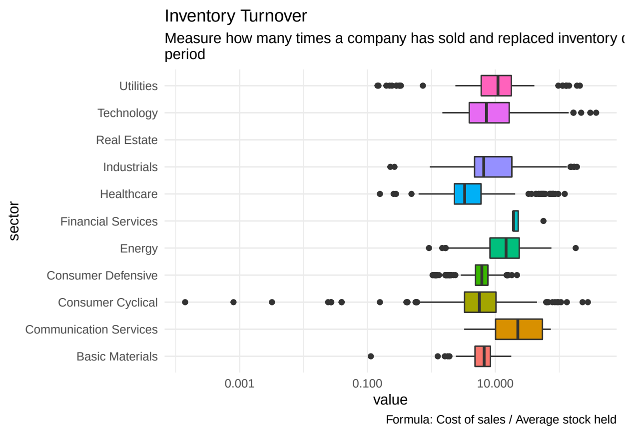

Days Sales Outstanding
Measure of the average number of days that it takes a company to colle
after a sale has been made
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
1
100
10000
value
Formula: accounts receivable / Total Credits sales * Nperiods

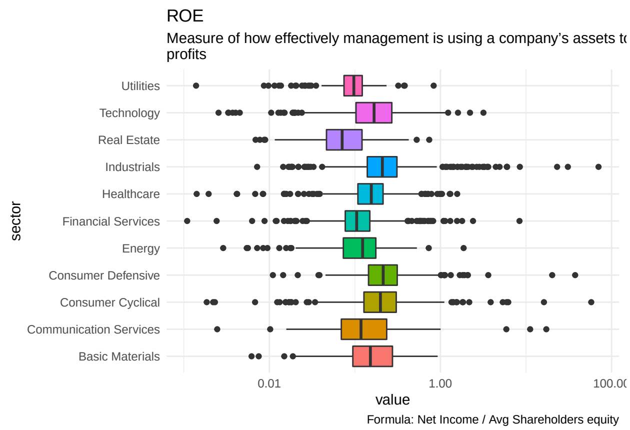

Days Payables Outstanding
Financial ratio that indicates the average time (in days) that a company
to pay its bills and invoices to its trade creditors
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10
10000
10000000
value
Formula: accounts payable / COGS * Nperiods

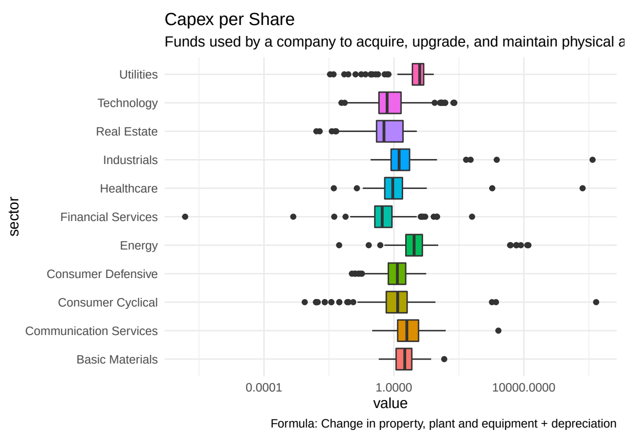

Days of Inventory on Hand
Measure of how quickly a business uses up the average inventory it kee
stock
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10
1000
100000
value
Formula: inventory turnover rate / 365











It seems that some metrics could be correlated and many of them tends to have the same pattern per sectors. That lead us to the question on how to evaluate them in risk mitigation assets, this will be done in the next chapter.

We can see that metrics that clearly show a difference in results per sectors are:

1. Debt to equity
2. Current ratio
3. Interest coverage
4. SG&A to revenue
5. R&D to revenue
6. Intangibles to assets
7. Capex to operatig cash flow

8. Capex to revenue
9. ROIC
10. Return on tangible assets
11. Working capital
12. Tangible Asset Value
13. Average Inventory
14. Days sales outstanding
15. Days payables outstanding
16. Days of inventory on hand
17. Receivables turnover
18. Inventory turnover
19. ROE

It is important to notice that x-axis is on log scale, that means that other metrics could be included as well

Let's check how these metrics are correlated

```
KeyMetrics %>%
    unnest(data) %>%
    ungroup() %>% #removes grouped data, otherwise select will bring grouped atributes as well
    select(sector, metric, symbol, date,  value) %>% # select variables needed to spread
    spread(key = metric, value = value) %>% # spread metrics to column that will be correlated
    select(-sector, -symbol, -date) %>% # remove columns not needed
    drop_na() %>% #Remove any na on metrics data, to fix correlation function return NA
    cor() %>% # Apply correlation function
    as.data.frame() %>% # Convert matrix class to data frame
    rownames_to_column("Metric") %>% # Include row names id from matrix to a column named data frame
    gather( "metric", "correlation", -Metric) %>% # gather all correlation into a single column
    filter(Metric != metric) %>% # Remove any metrics equal (That results in correlation 1)
    arrange(-correlation) %>% # Arrange correlation, this will be used in id creation later
    filter(correlation >= 0.8) %>% # Filter only correlations greater than 0.8
    mutate(id = case_when(Metric == lag(metric, 1) ~ 1, TRUE ~ 0)) %>% # Column created to remove dupli
    filter(id == 1) %>% # Removing duplicates
    select(-id) %>% # Removing aux column
    kable(caption = "Correlation of metrics greather than 80%") %>%
    kable_styling(full_width = F)
```

\begin{table}

\caption{Correlation of metrics greather than 80%}

Metric	metric	correlation
Book Value per Share	Shareholders Equity per Share	1.0000000
Enterprise Value	Market Cap	1.0000000
EV to Operating cash flow	POCF ratio	1.0000000
EV to Free cash flow	PFCF ratio	0.9999999
Enterprise Value over EBITDA	Market Cap	0.9999999
PB ratio	PTB ratio	0.9999999
Enterprise Value	Enterprise Value over EBITDA	0.9999999
Enterprise Value over EBITDA	PE ratio	0.9999999
PE ratio	Market Cap	0.9999998
Market Cap	PE ratio	0.9999998
PE ratio	Enterprise Value	0.9999998
Enterprise Value	PE ratio	0.9999998
EV to Sales	Price to Sales Ratio	0.9999993
Enterprise Value over EBITDA	PFCF ratio	0.9999987
PFCF ratio	Market Cap	0.9999986
Market Cap	PFCF ratio	0.9999986
PFCF ratio	Enterprise Value	0.9999986
Enterprise Value	PFCF ratio	0.9999986
PFCF ratio	PE ratio	0.9999986
PE ratio	PFCF ratio	0.9999986
PFCF ratio	EV to Operating cash flow	0.9999986
EV to Operating cash flow	PFCF ratio	0.9999986
PFCF ratio	POCF ratio	0.9999985
Enterprise Value over EBITDA	EV to Free cash flow	0.9999984
EV to Free cash flow	EV to Operating cash flow	0.9999984
EV to Free cash flow	Market Cap	0.9999984
Enterprise Value	EV to Free cash flow	0.9999984
EV to Free cash flow	PE ratio	0.9999983
EV to Free cash flow	POCF ratio	0.9999982
Enterprise Value over EBITDA	EV to Operating cash flow	0.9999974
Enterprise Value over EBITDA	POCF ratio	0.9999973
EV to Operating cash flow	Market Cap	0.9999973
Enterprise Value	EV to Operating cash flow	0.9999973
Market Cap	POCF ratio	0.9999973
POCF ratio	Enterprise Value	0.9999973
Enterprise Value	POCF ratio	0.9999973
EV to Operating cash flow	PE ratio	0.9999972
PE ratio	POCF ratio	0.9999972
Enterprise Value over EBITDA	Price to Sales Ratio	0.9999972
Price to Sales Ratio	Market Cap	0.9999971
Market Cap	Price to Sales Ratio	0.9999971
Price to Sales Ratio	Enterprise Value	0.9999971
Enterprise Value	Price to Sales Ratio	0.9999971
Price to Sales Ratio	PE ratio	0.9999971
PE ratio	Price to Sales Ratio	0.9999971
Price to Sales Ratio	PFCF ratio	0.9999959
PFCF ratio	Price to Sales Ratio	0.9999959
Price to Sales Ratio	EV to Free cash flow	0.9999956
EV to Free cash flow	Price to Sales Ratio	0.9999956
Enterprise Value over EBITDA	PB ratio	0.9999952
PB ratio	Market Cap	0.9999952
Market Cap	PB ratio	0.9999952
PB ratio	Enterprise Value	0.9999952
Enterprise Value	PB ratio	0.9999952
Enterprise Value over EBITDA	PTB ratio	0.9999952
PB ratio	PE ratio	0.9999951
Market Cap	PTB ratio	0.9999951

\end{table}

This is very interesting, it seems that 32 metrics have a correlation with one of these metrics by greather than 80% as we could see in the metrics plots.

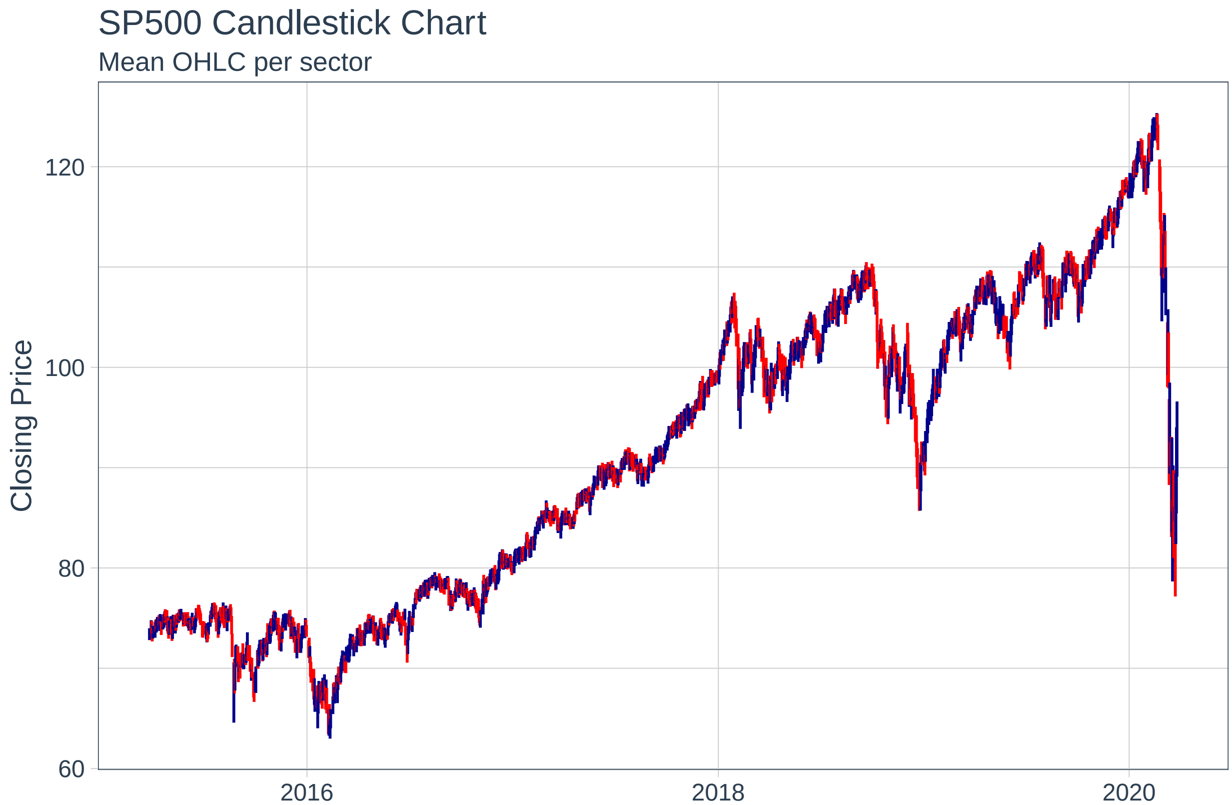
That actually makes sense because these metrics formulas are shared or have a common hierarchy formula variable.

That still lead us the question of risk mitigation on portfolio assets to these metrics, even though they're correlated that doesn't mean that these assets will follow the same pattern. In the next chapter this will be analyzed

Price

Let's take a look on market price in SP500 and per sector.

```
#Candlestick for SP500
HistoricalPrices %>%
  select(sector, Historical_Daily_Price) %>%
  unnest() %>%
  group_by(date) %>%
  summarise(close = mean(close), open = mean(open), low = mean(low), high = mean(high)) %>%
  ggplot(aes(x = date, y = close)) +
  geom_candlestick(aes(open = open, high = high, low = low, close = close)) +
  labs(title = "SP500 Candlestick Chart",
       subtitle = "Mean OHLC per sector",
       y = "Closing Price", x = "") +
  theme_tq()
```

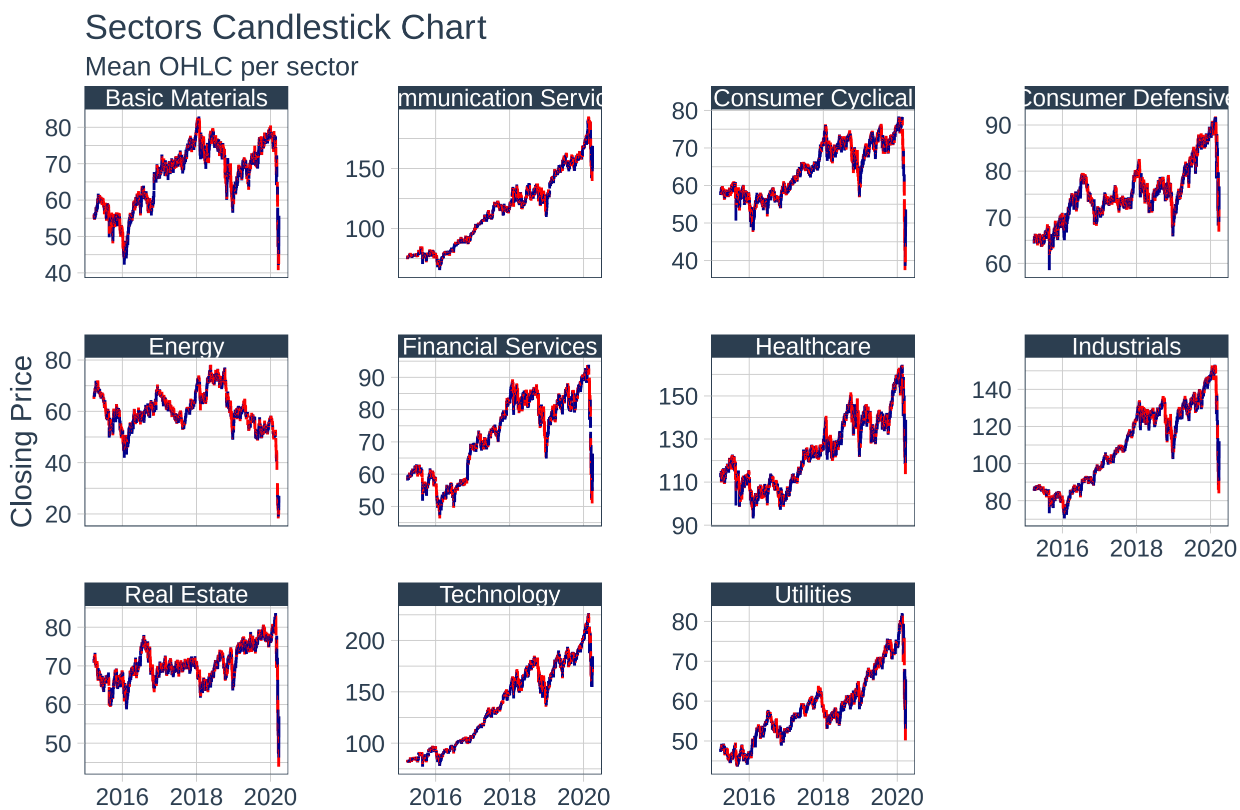


This candlestick time series plot shows some very interesting analysis:

1. Huge drop in the market due to coronavirus, it does seem to be one of the worst drops in the market for SP500.
2. Market has a lot of drops and ups, but usually shows some trends upwards stoped by some peaked collapses that we know are related to economy breakdowns.

Let's evaluate the same plot per sector

```
#Candlestick per sector
Sector_Daily_OHLC <- HistoricalPrices %>%
  select(sector, Historical_Daily_Price) %>%
  unnest() %>%
  group_by(sector,date) %>%
  summarise(close = mean(close), open = mean(open), low = mean(low), high = mean(high))

Sector_Daily_OHLC %>%
  ggplot(aes(x = date, y = close, group = sector)) +
  geom_candlestick(aes(open = open, high = high, low = low, close = close)) +
  labs(title = "Sectors Candlestick Chart",
       subtitle = "Mean OHLC per sector",
       y = "Closing Price", x = "") +
  facet_wrap(~ sector, ncol = 4, scale = "free_y") +
  theme_tq()
```



It does seem that all sectors were havily impacted, although technology and communication services were a bit less impacted in percent to its previous downfall price

Another intereseting thing is that some sectors seems to have a similar pattern, by looking at them it seems we have 3 groups of sectors.

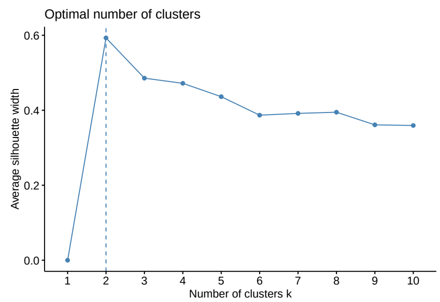
Let's take a look on these clusters, we'll use a silhouette method to define the optimal amount of clusters in sectors price

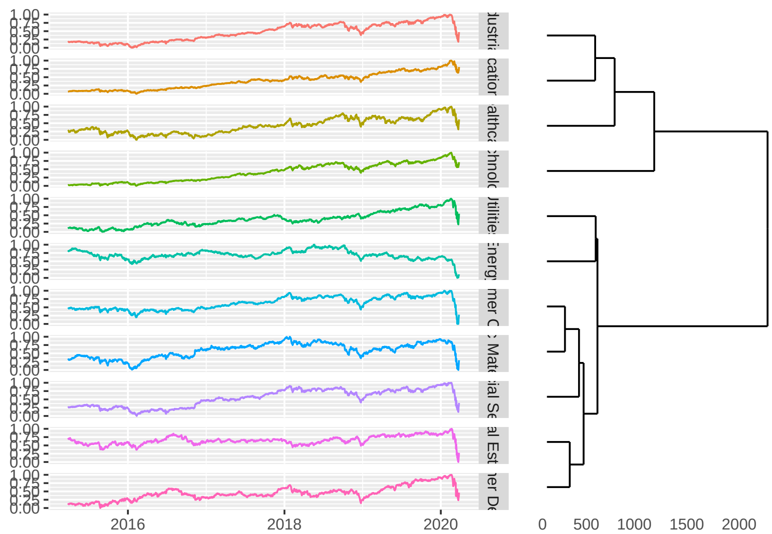
```
Clustering <- function(Cluster_DF, Df_aux){
  require(gridExtra)
  require(ggdendro)
  require(zoo)
  require(purrr)
  require(tsibble)
  require(broom)


  # Clustering
  hc <- hclust(dist(t(Df_aux[,-1])), "ave")

  # 8.1 DF clusters
  library(factoextra)
  NbClust <- fviz_nbclust(Df_aux[,-1], FUN = hcut, method = "silhouette")


  k <- which.max(NbClust$data$y)
```

```r
  cut_avg <- cutree(hc, k = k) %>%
    tidy() %>%
    rename("Data"="names", "cluster"="x")


  # Number of clusters plot
  NbClustersPlot <- plot(NbClust)

  ### Plot
  hcdata <- dendro_data(hc)
  names_order <- hcdata$labels$label

  # Use the folloing to remove labels from dendogram so not doubling up - but good for checking
  hcdata$labels$label <- ''
  p1 <- ggdendrogram(hcdata, rotate=TRUE, leaf_labels=FALSE)

  # Autoplot only accepts time series data type
  Zoo_DF <- read.zoo(Df_aux)

  # Scale the time series and plot
  maxs <- apply(Zoo_DF, 2, max)
  mins <- apply(Zoo_DF, 2, min)
  joined_ts_scales <- scale(Zoo_DF, center = mins, scale = maxs - mins)

  new_data <- joined_ts_scales[,rev(as.character(names_order))]

  p2 <- autoplot(new_data, facets = Series ~ . ) +
    xlab('') + ylab('') + theme(legend.position="none")

  gp1<-ggplotGrob(p1)
  gp2<-ggplotGrob(p2)

  grid <- grid.arrange(gp2, gp1, ncol=2, widths=c(4,2))


  aux <- data.frame(Model_Name = Cluster_DF) %>%
    mutate(Clustered = purrr::map(Model_Name, ~cut_avg),
           hc = purrr::map(Model_Name, ~hc),
           NbClust= purrr::map(Model_Name, ~NbClust),
           NbClustersPlot= purrr::map(Model_Name, ~NbClustersPlot),
           p1= purrr::map(Model_Name, ~p1),
           p2= purrr::map(Model_Name, ~p2),
           grid = purrr::map(Model_Name, ~grid)
    )


  return(aux)
}


Clust_DF <- Sector_Daily_OHLC %>%
            select(sector, date, close) %>%
            spread(sector, close) %>%
```

```
            filter_all(all_vars(!is.na(.)))

Clusters <- Clustering("Sectors", Clust_DF)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```





That's interesting! We have 2 groups of clusters, let's organize them:

```
Clusters_sectors <- map_dfr(Clusters$Clustered, ~..1) %>%
                    rename("sector"=Data) %>%
                    arrange(cluster)


kable(Clusters_sectors, caption = "Sector Clusters") %>%
  kable_styling(full_width = F)
```

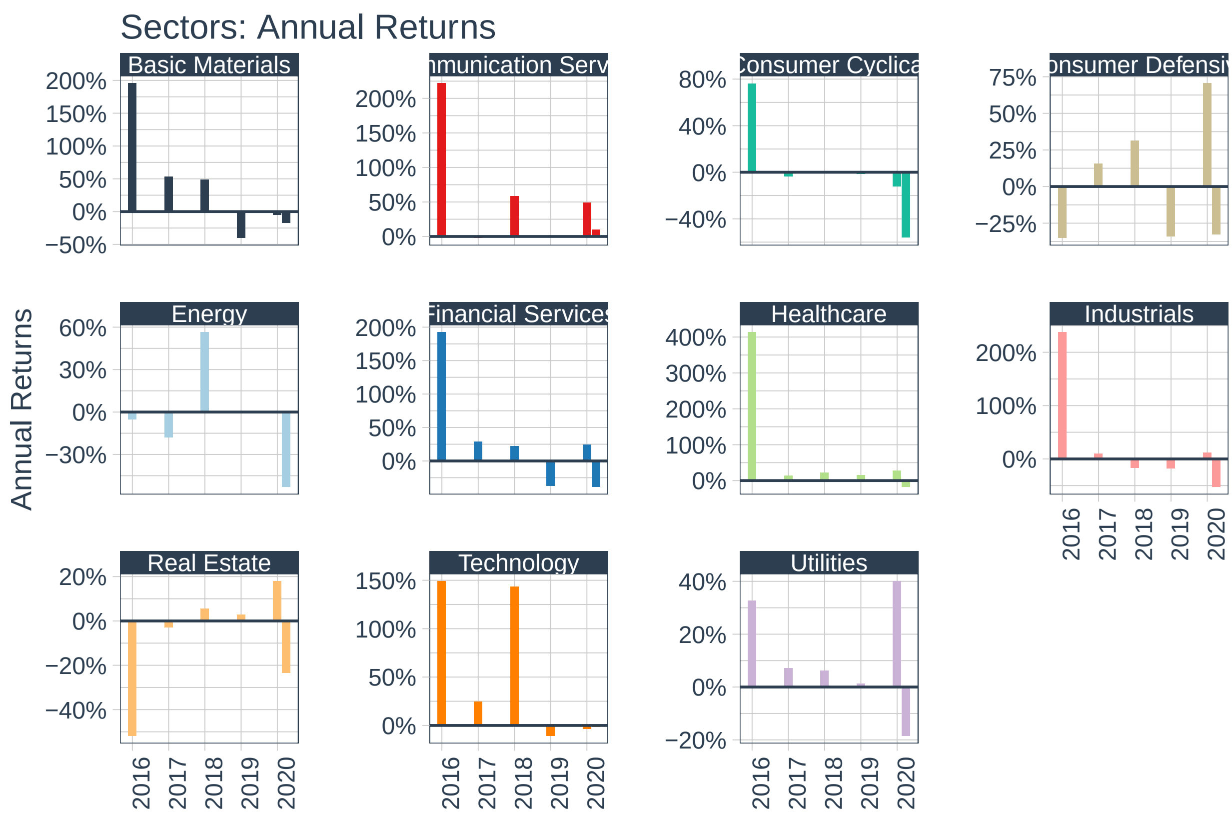
Now let's see how returns are ocurring annually per sector

```
#Annual returns per sectors
HistoricalPrices %>%
unnest(Monthly_AdjPrice) %>%
group_by(sector) %>%
tq_transmute(select = close, mutate_fun = periodReturn, period = "yearly", type = "arithmetic") %>%
ggplot(aes(x = date, y = yearly.returns, fill = sector)) +
  geom_col() +
  geom_hline(yintercept = 0, color = palette_light()[[1]]) +
  scale_y_continuous(labels = scales::percent) +
```

Table 3: Sector Clusters

sector	cluster
Basic Materials	1
Consumer Cyclical	1
Consumer Defensive	1
Energy	1
Financial Services	1
Real Estate	1
Utilities	1
Communication Services	2
Healthcare	2
Industrials	2
Technology	2

```
labs(title = "Sectors: Annual Returns",
     y = "Annual Returns", x = "") +
facet_wrap(~ sector, ncol = 4, scales = "free_y") +
theme_tq() +
theme(axis.text.x = element_text(angle = 90, hjust = 1),
      legend.position = "none") +
scale_fill_tq()
```

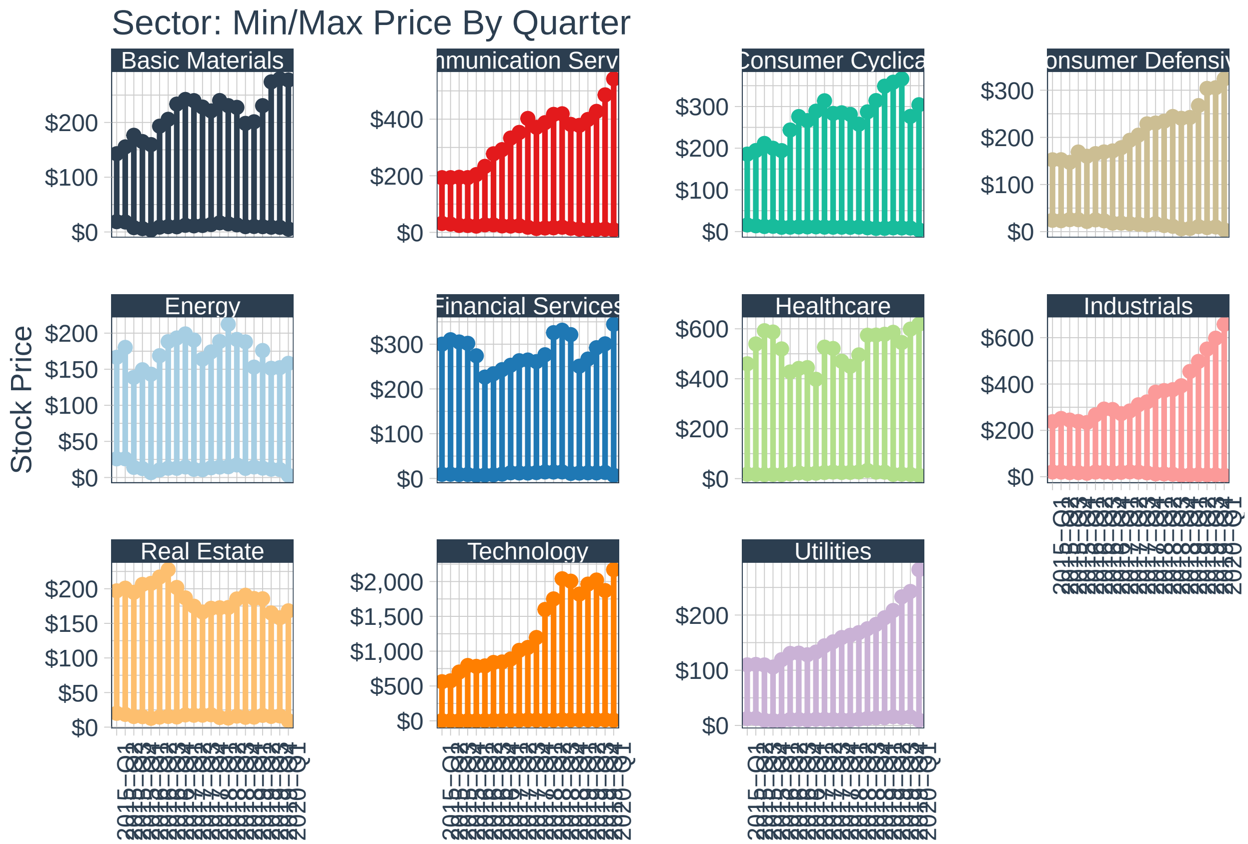


It seems that in general, the market has some downsides and upsides in annual return for all sectors, but it seems we have a huge drop on return untill this moment of 2020 due to coronavirus.

Another interesting result from this plot is that some sectors have some good percentual margin increase, but thats could be done due to low price being affected by any trend just as it seems to occur with basic materials.

The last analysis is to understand how prices moves quarterly, for that we'll get a min/max quarterly price per sector plot

```
#Quaterly max min per sector
Sector_max_by_qtr <- HistoricalPrices %>%
                     unnest(Historical_Daily_Price) %>%
                     group_by(sector) %>%
                     tq_transmute(select = close, mutate_fun = apply.quarterly, FUN= max,
                                  col_rename = "max.close") %>%
                     mutate(year.qtr = paste0(lubridate::year(date), "-Q",
                                              lubridate::quarter(date))) %>%
                     select(-date)



Sector_min_by_qtr <- HistoricalPrices %>%
        unnest(Historical_Daily_Price) %>%
        group_by(sector) %>%
        tq_transmute(select = close, mutate_fun = apply.quarterly,
                     FUN= min, col_rename = "min.close") %>%
        mutate(year.qtr = paste0(lubridate::year(date), "-Q",
                                 lubridate::quarter(date))) %>%
        select(-date)

Sector_by_qtr <- left_join(Sector_max_by_qtr, Sector_min_by_qtr,
                           by = c("sector" = "sector", "year.qtr" = "year.qtr"))


Sector_by_qtr %>%
  ggplot(aes(x = year.qtr, color = sector)) +
  geom_segment(aes(xend = year.qtr, y = min.close, yend = max.close),
               size = 1) +
  geom_point(aes(y = max.close), size = 2) +
  geom_point(aes(y = min.close), size = 2) +
  facet_wrap(~ sector, ncol =4, scale = "free_y") +
  labs(title = "Sector: Min/Max Price By Quarter",
       y = "Stock Price", color = "") +
  theme_tq() +
  scale_color_tq() +
  scale_y_continuous(labels = scales::dollar) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        axis.title.x = element_blank(),
        legend.position = "none")
```



This is interesting, all sectors have as low price as zero as well as higher prices. This could mean that companies could be integrated into SP500 quaterly and therefore have low prices due to start in the market, or that the market variability is across all sectors evenly.

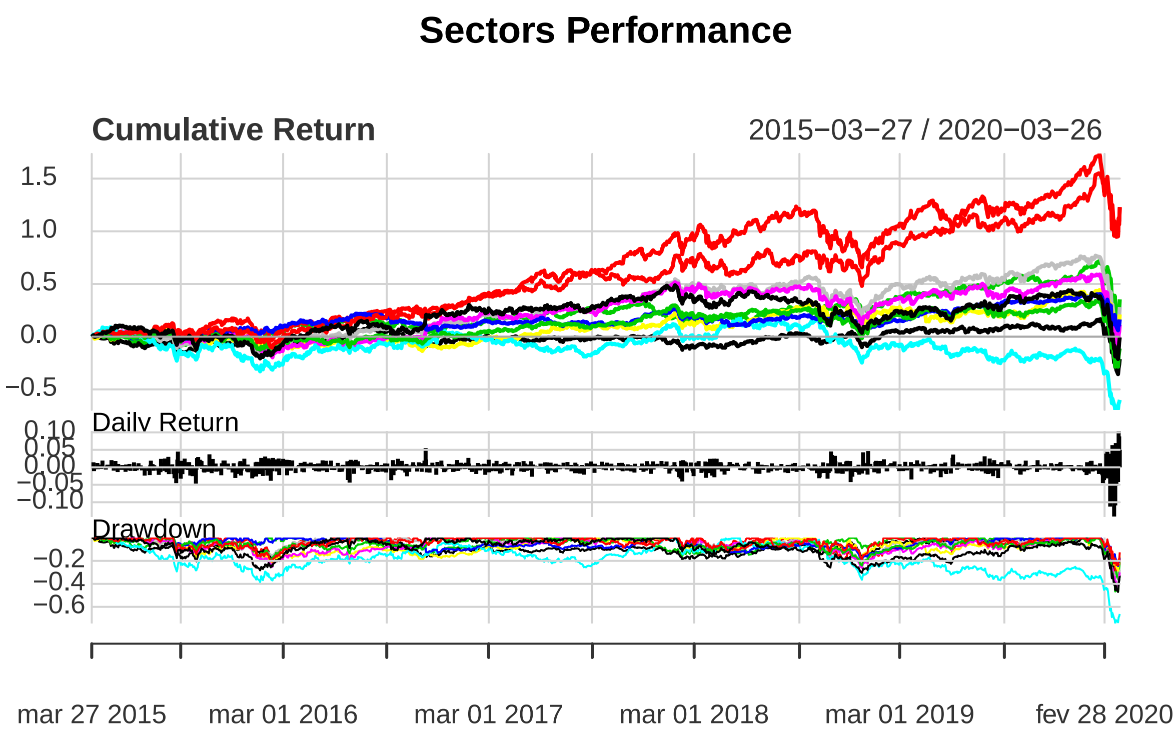
The major difference is that some sectors are more variable than others, and there are those that achieve higher prices in the long run such as technology sectors

Portfolio Analysis

Sectors

Let's make a portfolio with risk mitigation in order to evaluate how the optimization tries to deal with sector returns

```
Sector_Returns <- Sector_Daily_OHLC %>%
                  group_by(sector) %>%
                  tq_transmute(select     = close,
                               mutate_fun = periodReturn,
                               period     = "daily",
                               col_rename = "close") %>%
                  spread(sector, close) %>%
                  filter_all(all_vars(!is.na(.))) %>%
                  tk_xts(data = ., date_var = date, silent = TRUE)  #Its needed to run porfolio.spec

charts.PerformanceSummary(Sector_Returns,main = "Sectors Performance", legend.loc = NULL)
```



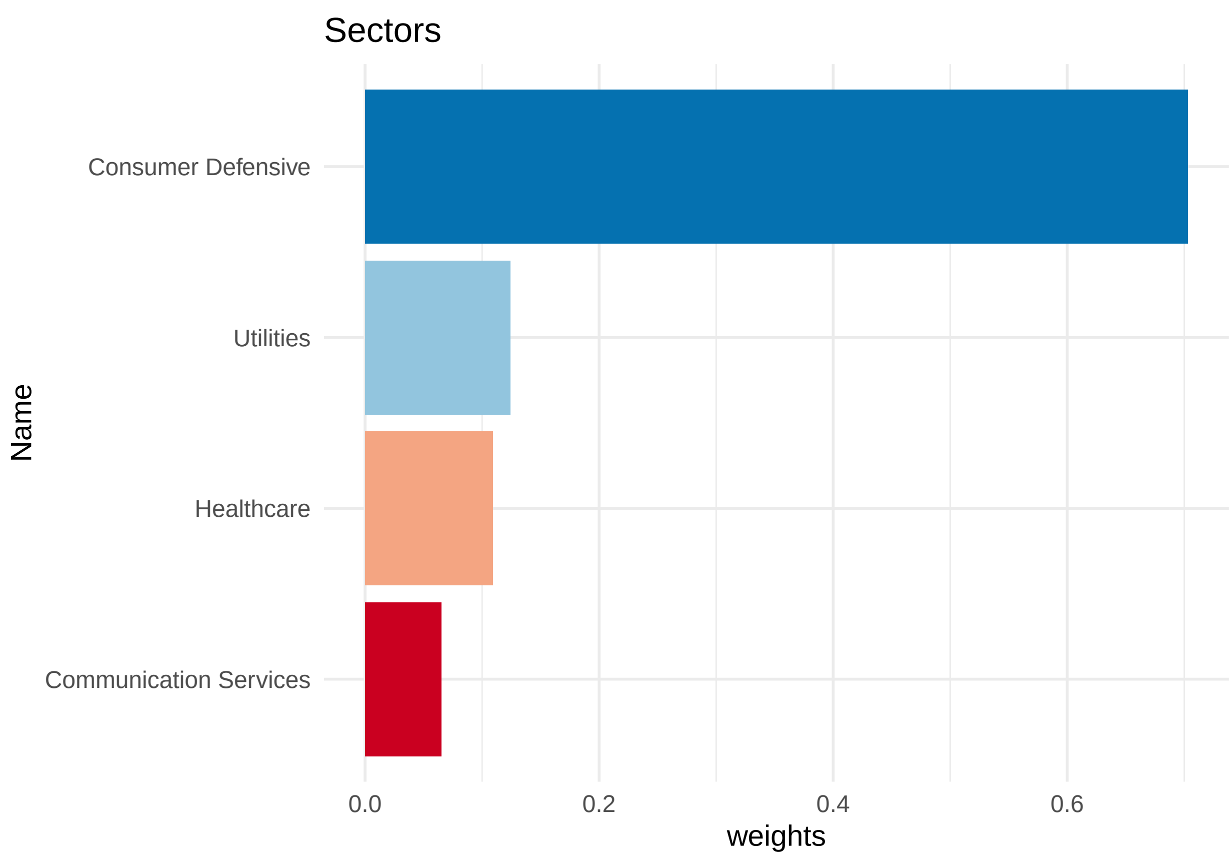
Let's run the portfolio in order to minimize risk and evaluate it's sector allocation called weight

```
Optimize <- function(Returns){

# Create the portfolio specification
port_spec <- portfolio.spec(colnames(Returns)) %>%

  # Add a full investment constraint such that the weights sum to 1
  add.constraint(portfolio = ., type = "full_investment") %>%

  # Add a long only constraint such that the weight of an asset is between 0 and 1
  add.constraint(portfolio = ., type = "long_only") %>%

  # Add an objective to minimize portfolio standard deviation
  add.objective(portfolio = ., type = "risk", name = "StdDev")

# Solve the optimization problem
opt <- optimize.portfolio(Returns, portfolio = port_spec,
                          optimize_method = "ROI", trace=TRUE)


return(opt)

}


SectorReturns <- Optimize(Sector_Returns) %>%
```

```
                 extractWeights() %>%
                 data.frame(Name = names(.), weights = round(.,3), row.names = NULL) %>%
                 select(-.)
```

Let's plot these weights

```
plots2 <- function(weights, sector=NULL){

  plot <- weights %>%
          mutate(Name = fct_reorder(Name, weights)) %>%
          filter(weights > 0.01) %>%
          ggplot(aes(x = Name, y=weights, fill = Name)) +
          geom_col() +
          scale_fill_brewer(palette = "RdBu") +
          theme_minimal() +
          coord_flip() +
          guides(fill = "none") +
          labs(title = paste0("Sectors ", sector))

  return(plot)

}


plots2(SectorReturns)
```

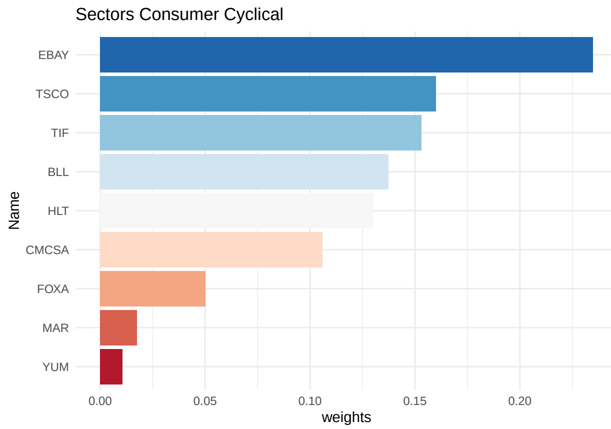


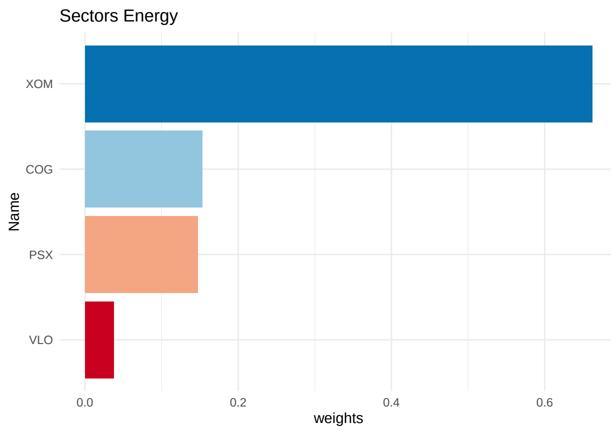
Let's check how each company per sector is structured in this portfolio analysis.

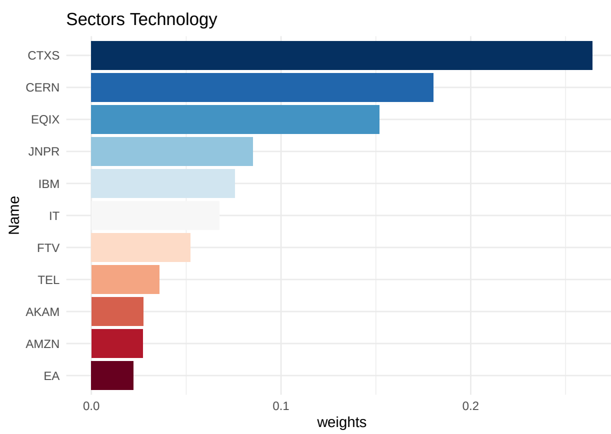
Companies per sector

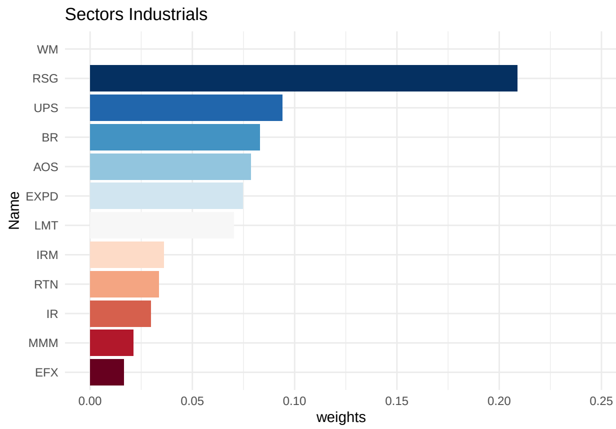
Since there are 400 countries and the main idea here is to understand how metrics relate to risk mitigation, we'll record only the final result of the best and worst weight allocation of portfolio companies per sector

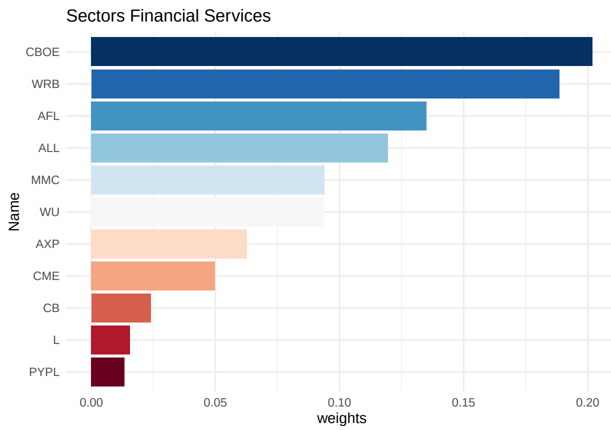
```
#Since we have to model this grouped per sector, all this piece of script is doing is merging the same
Symbol_Returns <- HistoricalPrices %>%
  select(symbol,sector, Historical_Daily_Price) %>%
  unnest() %>%
  group_by(sector) %>%
  nest() %>%
  mutate(data = map(data, ~..1 %>%
                        select(symbol, date, close) %>%
                        group_by(symbol) %>%
                        tq_transmute(select     = close,
                                     mutate_fun = periodReturn,
                                     period     = "daily",
                                     col_rename = "close") %>%
                        spread(symbol, close) %>%
                        filter_all(all_vars(!is.na(.))) %>%
                        tk_xts(data = ., date_var = date, silent = TRUE)))

#Optimizing per each purrr list of sectors
Symbol_Returns <- Symbol_Returns %>%
                  mutate(optimize = map(data, ~Optimize(..1)))

#Simple extract weights and organizing it to be able to plot
Symbol_Returns <- Symbol_Returns %>%
  mutate(weights = map(optimize, extractWeights),
         weights = map(weights, ~data.frame(Name = names(..1),
                                            weights = ..1, row.names = NULL)))


#Extracting worst and best symbols and ploting each sector weigths
Symbol_Returns <- Symbol_Returns %>%
        mutate(Best = map(weights, ~ filter(..1, weights == max(weights)) %>%
                           select(Name)),
               Worst = map(weights, ~ filter(..1, weights == min(weights)) %>%
                           select(Name)),
               plots = map(weights, ~plots2(..1, sector)))


walk(Symbol_Returns$plots, plot)
```

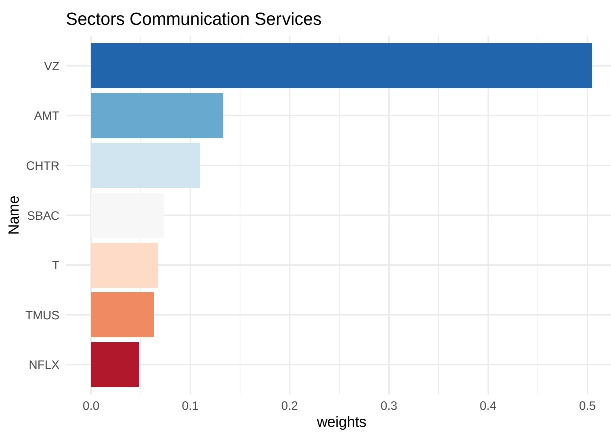

Sectors Consumer Cyclical
EBAY
TSCO
TIF
BLL
HLT
CMCSA
FOXA
MAR
YUM
Name
0.00
0.05
0.10
0.15
0.20
weights

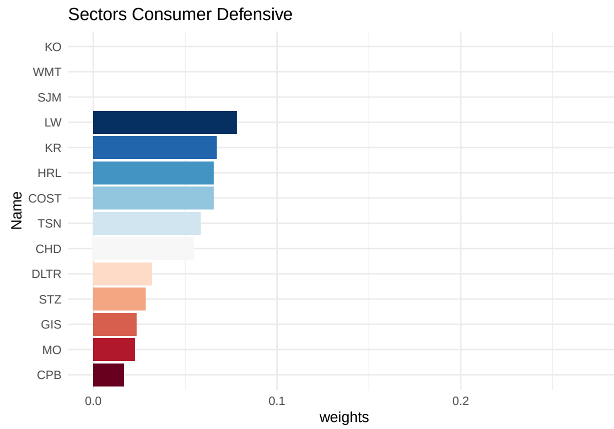

Sectors Energy
XOM
COG
PSX
VLO
Name
0.0
0.2
0.4
0.6
weights


Sectors Technology
CTXS
CERN
EQIX
JNPR
IBM
IT
FTV
TEL
AKAM
AMZN
EA
Name
0.0
0.1
0.2
weights

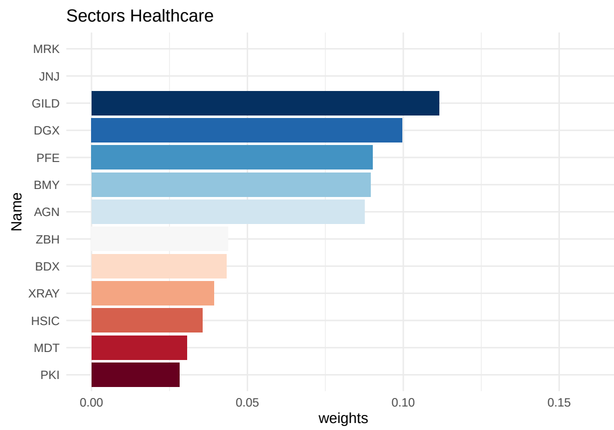

Sectors Industrials
WM
RSG
UPS
BR
AOS
EXPD
LMT
IRM
RTN
IR
MMM
EFX
Name
0.00
0.05
0.10
0.15
0.20
0.25
weights


Sectors Financial Services
CBOE
WRB
AFL
ALL
MMC
WU
AXP
CME
CB
L
PYPL
Name
0.00
0.05
0.10
0.15
0.20
weights

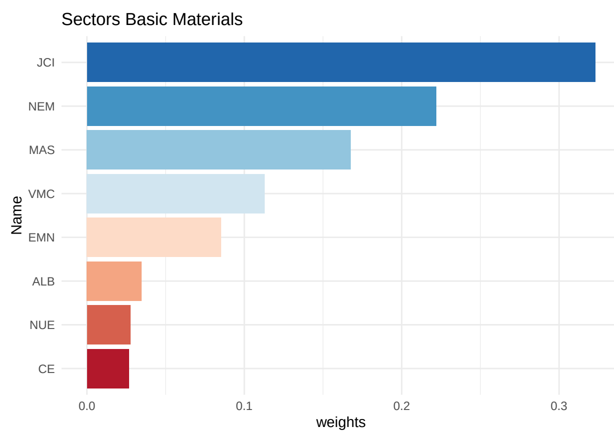

Sectors Communication Services
VZ
AMT
CHTR
SBAC
T
TMUS
NFLX
Name
0.0
0.1
0.2
0.3
0.4
0.5
weights



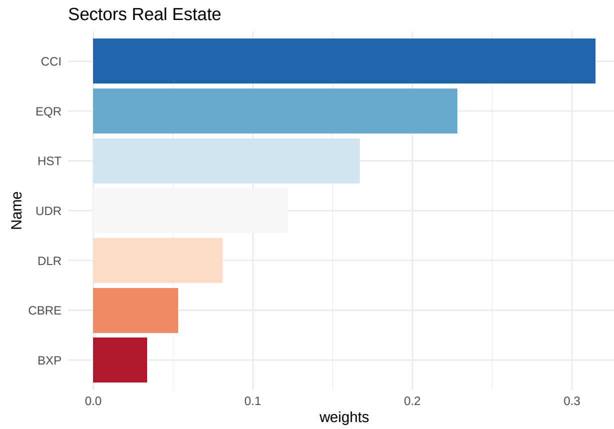
Sectors Consumer Defensive
KO
WMT
SJM
LW
KR
HRL
COST
TSN
CHD
DLTR
STZ
GIS
MO
CPB
Name
0.0
0.1
0.2
weights

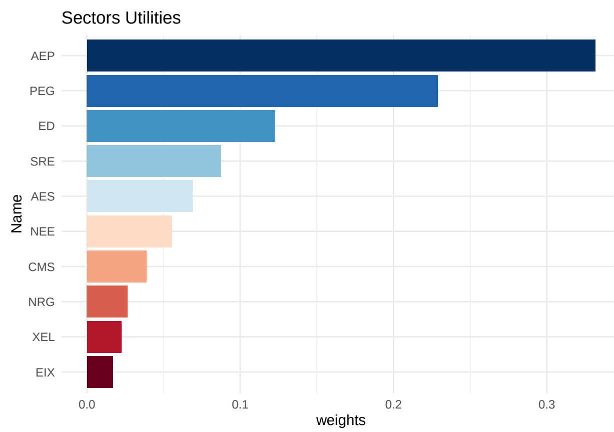


Sectors Healthcare
MRK
JNJ
GILD
DGX
PFE
BMY
AGN
ZBH
BDX
XRAY
HSIC
MDT
PKI
Name
0.00
0.05
0.10
0.15
weights



Sectors Basic Materials
JCI
NEM
MAS
VMC
EMN
ALB
NUE
CE
Name
0.0
0.1
0.2
0.3
weights





Now we can analyse each metric per best and worst companies allocated portfolio for risk mitigation.

KeyMetrics x Porfolio Mitigation

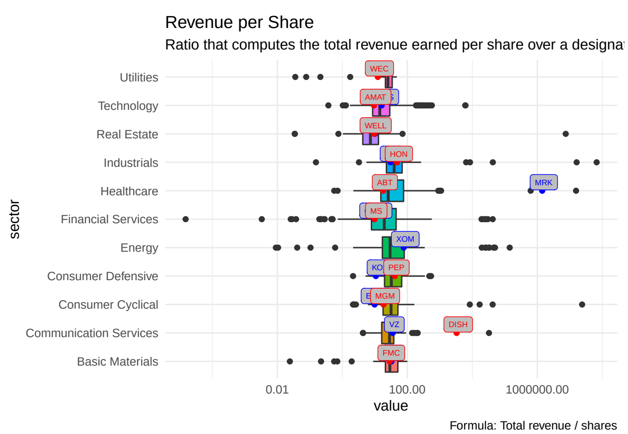
Best x Worst companies per sector

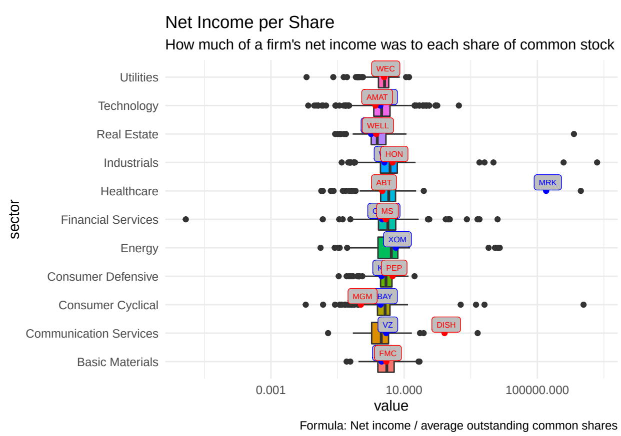
All we have to do know is to include the best and worst companies per sector in the keyMetrics dataset and plot each metric including the position of both best and worst companies in order to understand if these metrics relate to a decision on risk mitigation portfolio

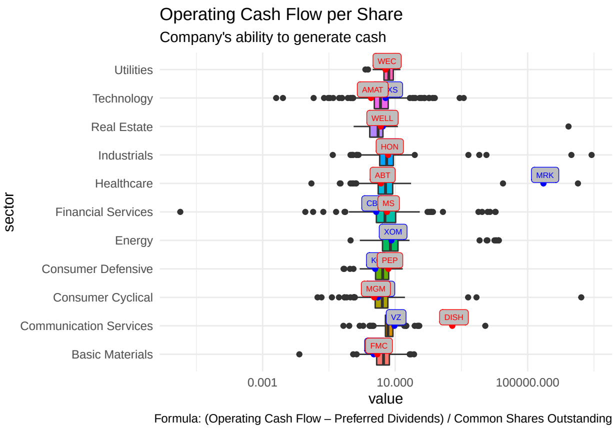
Table 4: Best and Worst company per Porfolio risk mitigation

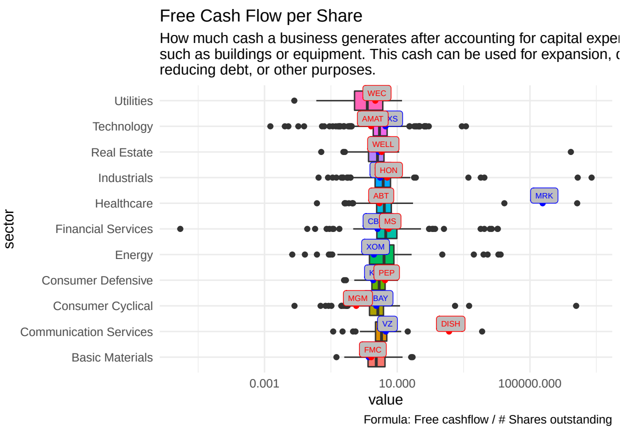
sector	Best	Worst
Consumer Cyclical	EBAY	MGM
Energy	XOM	FTI
Technology	CTXS	AMAT
Industrials	WM	HON
Financial Services	CBOE	MS
Communication Services	VZ	DISH
Consumer Defensive	KO	PEP
Healthcare	MRK	ABT
Basic Materials	JCI	FMC
Real Estate	CCI	WELL
Utilities	AEP	WEC

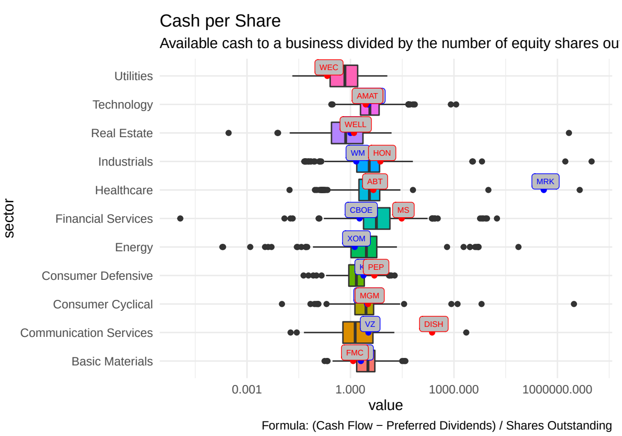
Plots

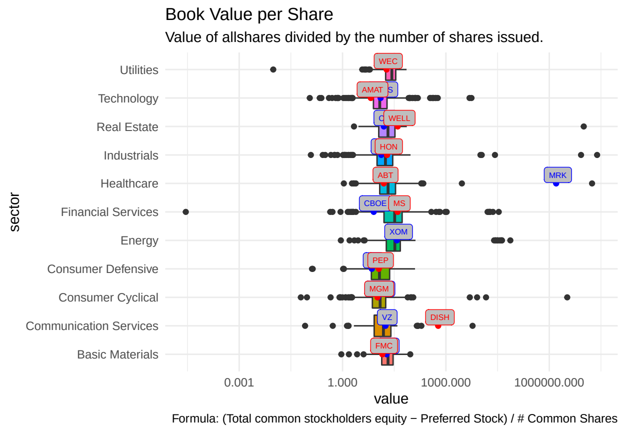


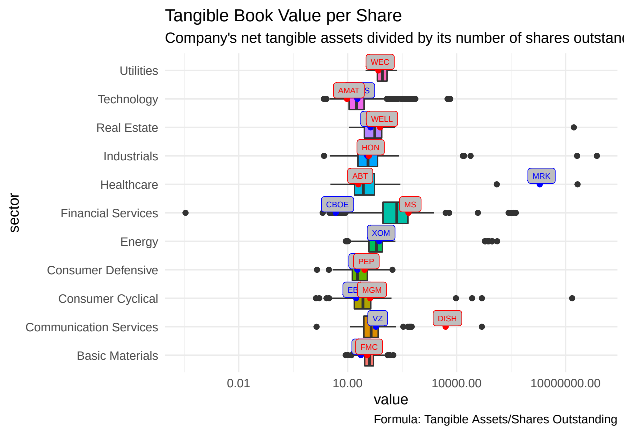


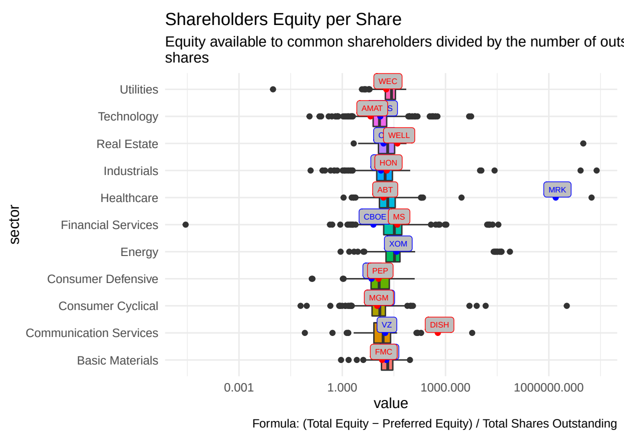

Operating Cash Flow per Share
Company's ability to generate cash
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
HON
ABT
MRK
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.001
10.000
100000.000
value
Formula: (Operating Cash Flow – Preferred Dividends) / Common Shares Outstanding


Free Cash Flow per Share
How much cash a business generates after accounting for capital expe
such as buildings or equipment. This cash can be used for expansion,
reducing debt, or other purposes.
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
HON
ABT
MRK
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.001
10.000
100000.000
value
Formula: Free cashflow / # Shares outstanding


Cash per Share
Available cash to a business divided by the number of equity shares ou
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
WM
HON
ABT
MRK
CBOE
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.001
1.000
1000.000
1000000.000
value
Formula: (Cash Flow − Preferred Dividends) / Shares Outstanding

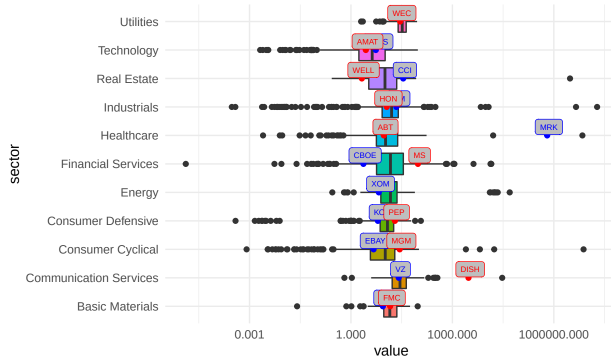

Book Value per Share
Value of allshares divided by the number of shares issued.
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
HON
ABT
MRK
CBOE
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.001
1.000
1000.000
1000000.000
value
Formula: (Total common stockholders equity − Preferred Stock) / # Common Shares


Tangible Book Value per Share
Company's net tangible assets divided by its number of shares outstanc
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
HON
ABT
MRK
CBOE
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.01
10.00
10000.00
10000000.00
value
Formula: Tangible Assets/Shares Outstanding


Shareholders Equity per Share
Equity available to common shareholders divided by the number of outs
shares
sector
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
WEC
AMAT
WELL
HON
ABT
MRK
CBOE
MS
XOM
PEP
MGM
VZ
DISH
FMC
0.001
1.000
1000.000
1000000.000
value
Formula: (Total Equity − Preferred Equity) / Total Shares Outstanding

Interest Debt per Share

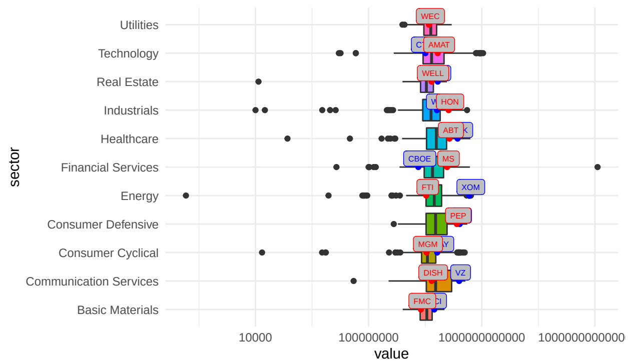
Ratio of each share accountant to interest of debts



Formula: Debt / (Debt + Shareholders Equity)

Market Cap

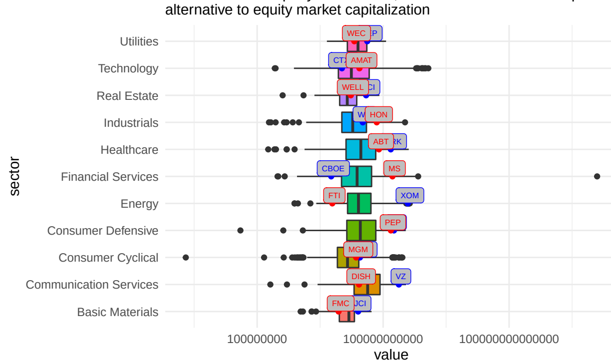
Company expected market value



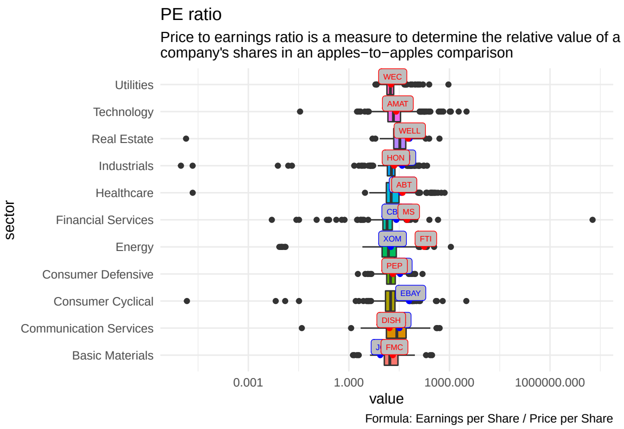
Formula: Current market price share * Number of Shares

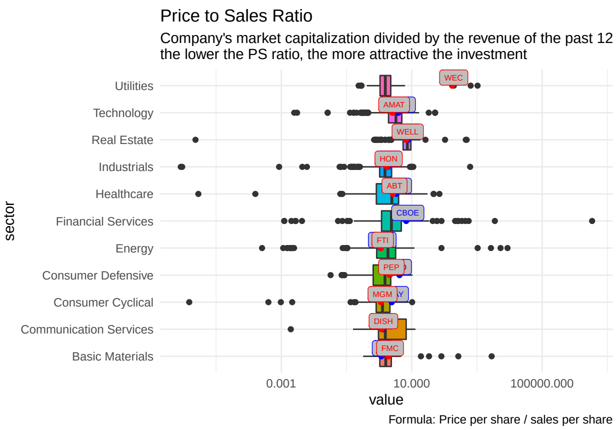
Enterprise Value

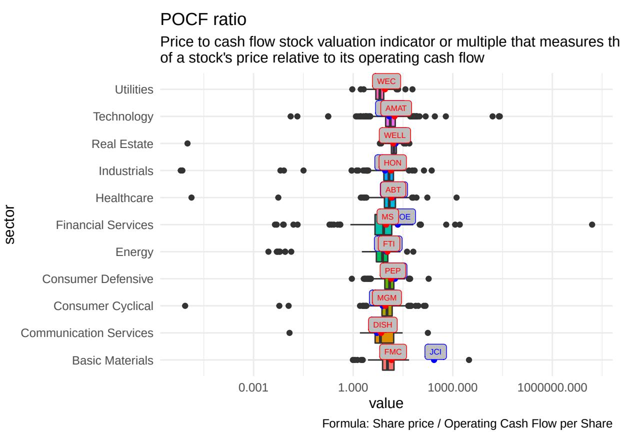
Measure of a company's total value, often used as a more comprehens alternative to equity market capitalization

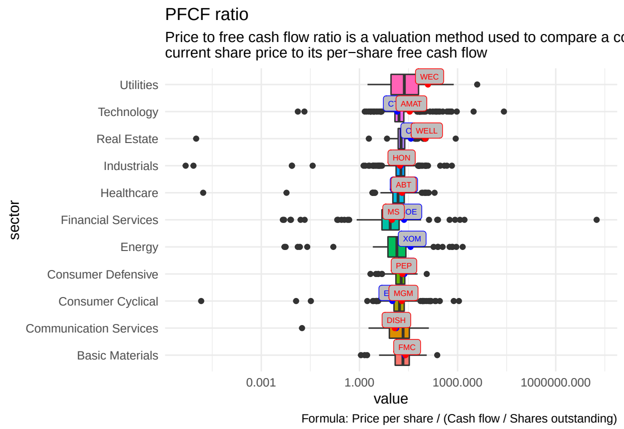


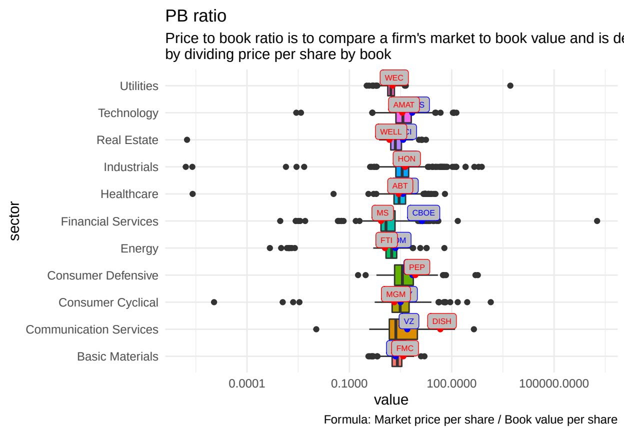
Formula: Market capitalization + Debt + Minority Shareholders + Preference Shares − Cash & Cash equivalents

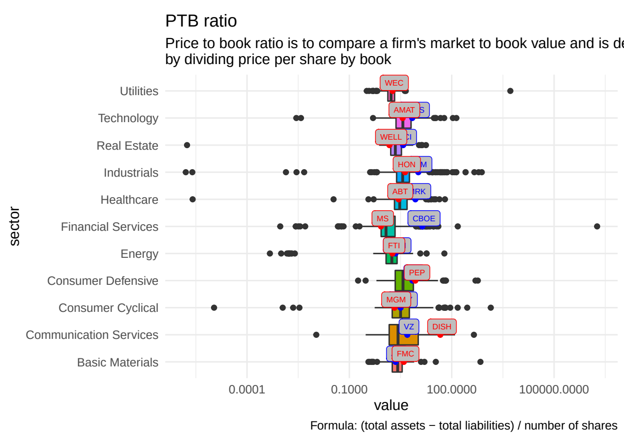

PE ratio
Price to earnings ratio is a measure to determine the relative value of a company's shares in an apples−to−apples comparison
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
1000000.000
value
Formula: Earnings per Share / Price per Share


Price to Sales Ratio
Company's market capitalization divided by the revenue of the past 12 r
the lower the PS ratio, the more attractive the investment
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
10.000
100000.000
value
Formula: Price per share / sales per share


POCF ratio
Price to cash flow stock valuation indicator or multiple that measures th
of a stock's price relative to its operating cash flow
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.001
1.000
1000.000
1000000.000
value
Formula: Share price / Operating Cash Flow per Share

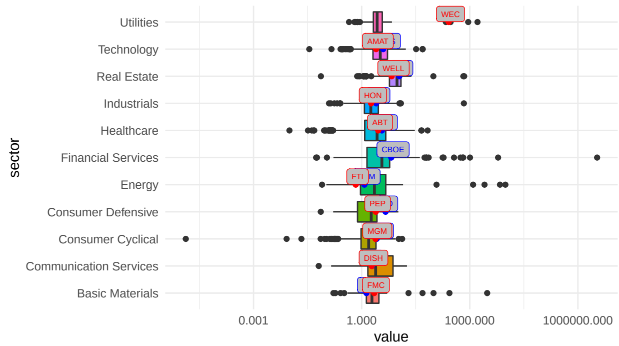

PFCF ratio
Price to free cash flow ratio is a valuation method used to compare a cc
current share price to its per−share free cash flow
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
WEC
AMAT
WELL
HON
ABT
MS
XOM
PEP
MGM
DISH
FMC
0.001
1.000
1000.000
1000000.000
value
Formula: Price per share / (Cash flow / Shares outstanding)


PB ratio
Price to book ratio is to compare a firm's market to book value and is de
by dividing price per share by book
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
WEC
AMAT
WELL
HON
ABT
MS
CBOE
FTI
PEP
MGM
VZ
DISH
FMC
0.0001
0.1000
100.0000
100000.0000
value
Formula: Market price per share / Book value per share


PTB ratio
Price to book ratio is to compare a firm's market to book value and is de
by dividing price per share by book
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
WEC
AMAT
WELL
HON
ABT
MS
CBOE
FTI
PEP
MGM
VZ
DISH
FMC
0.0001
0.1000
100.0000
100000.0000
value
Formula: (total assets − total liabilities) / number of shares

EV to Sales

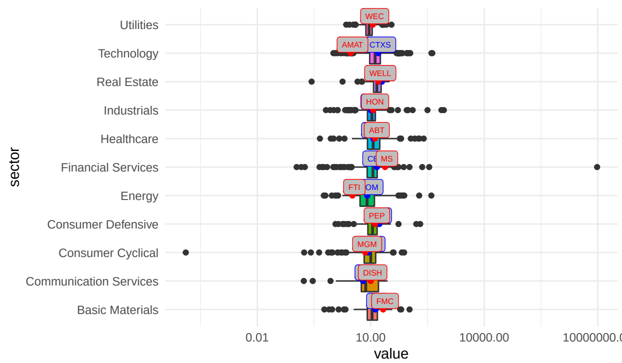
Enterprise value to sales is a business valuation method that compares
value of a company with its sales



Formula: (Market Capitalizato + Debt − Cash Cash) / Annual Sales

Enterprise Value over EBITDA

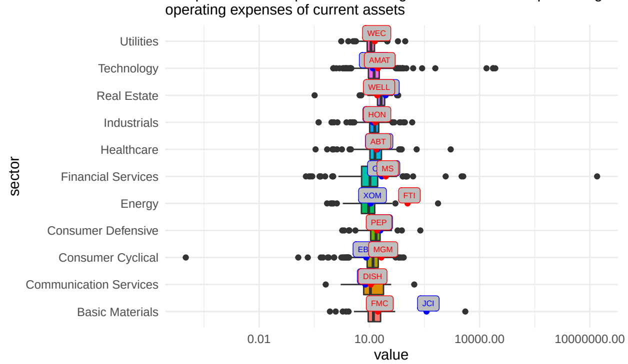
Good measure to estimate the cash flow of a company



Formula: (Market Capitalization + Preferred Shares + Minority Interest + Debt − Total Cash) / EBITDA

EV to Operating cash flow

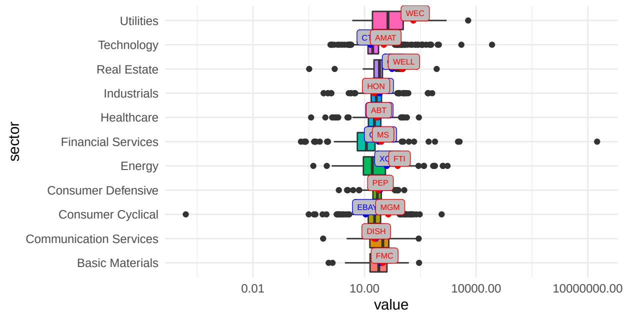
Enteprise value to op cash flow is a good to measure the percentage al
operating expenses of current assets



Formula: (Market Capitalization + Total Debt − Cash)/Cash from Operations

EV to Free cash flow

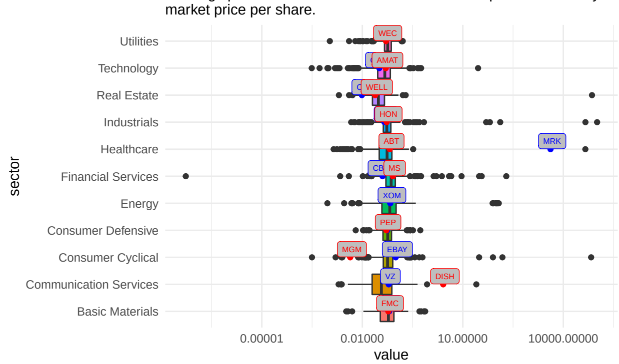
Compares the total valuation of the company with its ability to generate
cashflow. he lower the ratio of enterprise value to the free cash flow figu
the faster a company can pay back the cost of its acquisition or generat
to reinvest in its business



Formula: Free Cash Flow / Enterprise Value

Earnings Yield

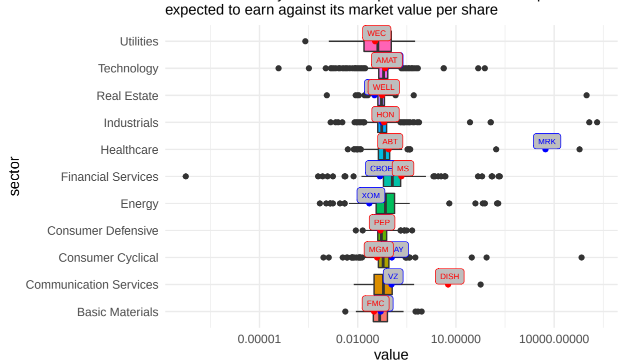
Earnings per share for the most recent 12−month period divided by the
market price per share.



Formula: Earnings per Share / Price per Share

Free Cash Flow Yield

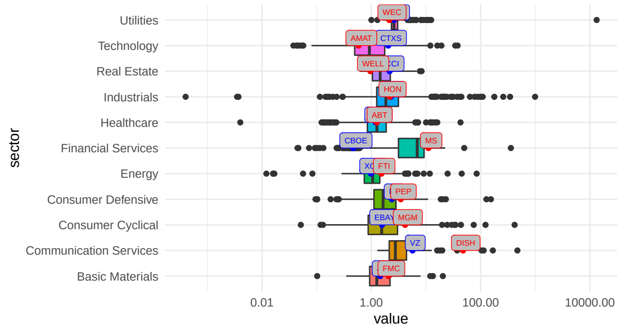
Financial solvency ratio that compares the free cash flow per share a co
expected to earn against its market value per share



Formula: Free Cash Flow / Market Capitalization

Debt to Equity

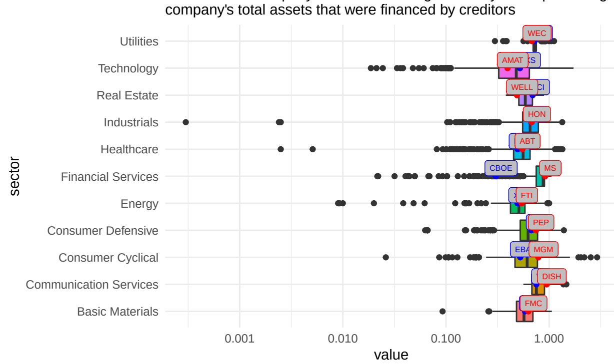
Financial ratio indicating the relative proportion of shareholders' equity a
debt used to finance a company's assets. Closely related to leveraging,
ratio is also known as risk, gearing or leverage



Formula: Total liabilites / total equity

Debt to Assets

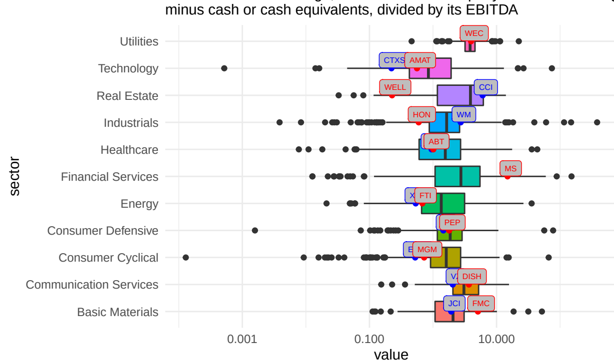
Indicator of a company's financial leverage. It tells you the percentage c
company's total assets that were financed by creditors



Formula: Total debts / total assets

Net Debt to EBITDA

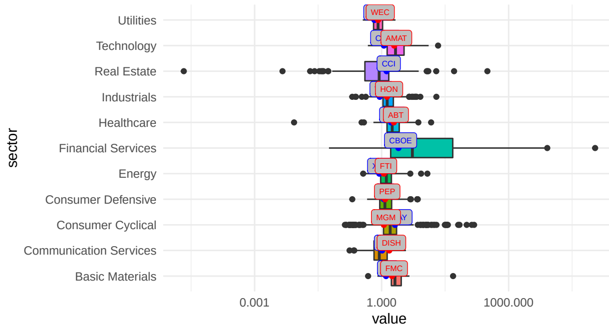
Measurement of leverage, calculated as a company's interest−bearing l
minus cash or cash equivalents, divided by its EBITDA



Formula: (Total Debt−Cash&Equivalents) / EBITDA

Current ratio

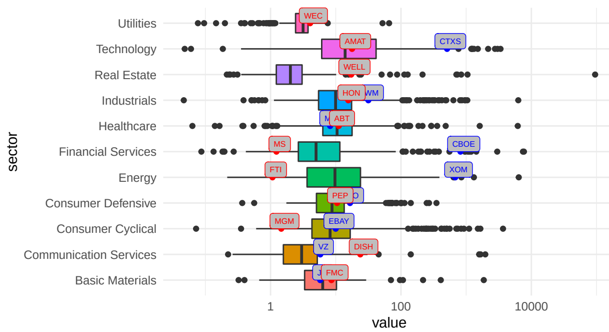
Liquidity ratio that measures whether a firm has enough resources to m
short−term obligations. It compares a firm's current assets to its current
liabilities



Formula: Current assets / Current Liabilities

Interest Coverage

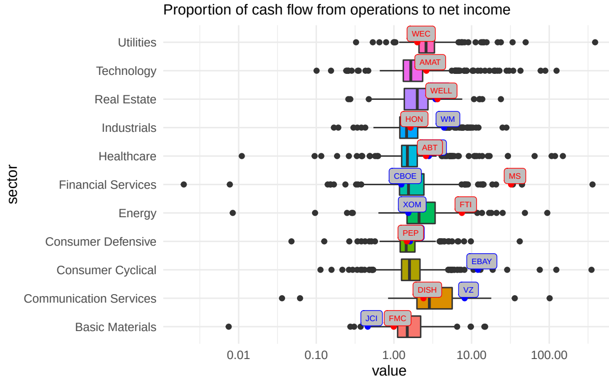
Measure of a company's ability to meet its interest payments. Interest c
ratio is equal to earnings before interest and taxes (EBIT) for a time per
often one year, divided by interestexpenses for the same time period



Formula: EBIT / Interest Expense

Income Quality

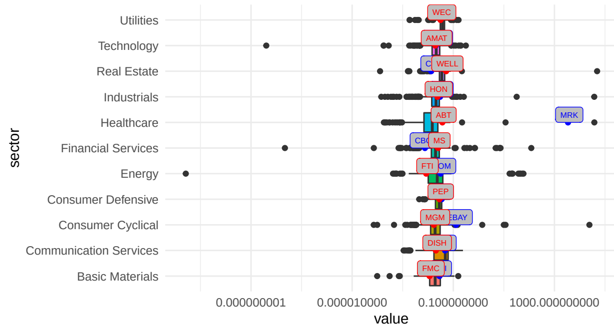
Proportion of cash flow from operations to net income



Formula: Net cash from operating activities / Net income

Dividend Yield

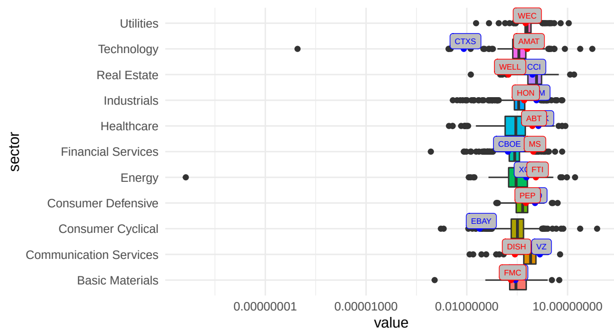
Dividend yield is used to calculate the earning on investment (shares) considering only the returns in the form of total dividends declared by tl company during the year



Formula: Annual dividend / Current Stock Pricec

Payout Ratio

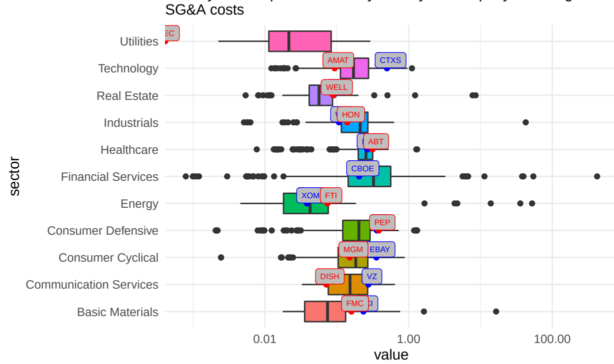
Fraction of net income a firm pays to its stockholders in dividends: The of earnings not paid to investors is left for investment to provide for futu earnings growth



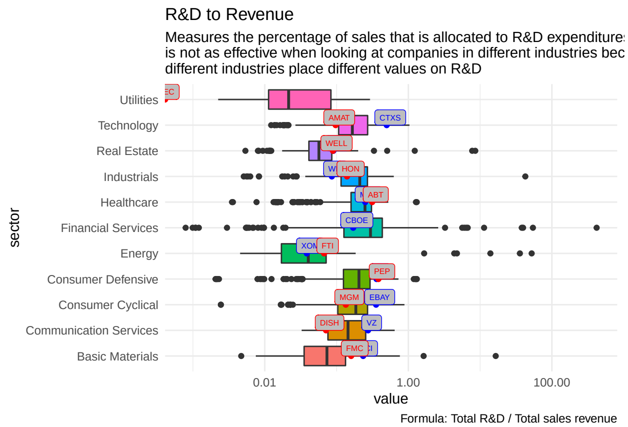
Formula: Total dividend Payments / (Net income + Noncash expenses − Noncash sales)

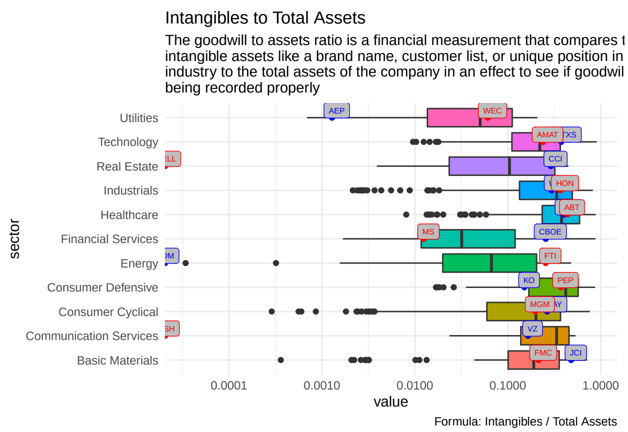
SG&A to Revenue

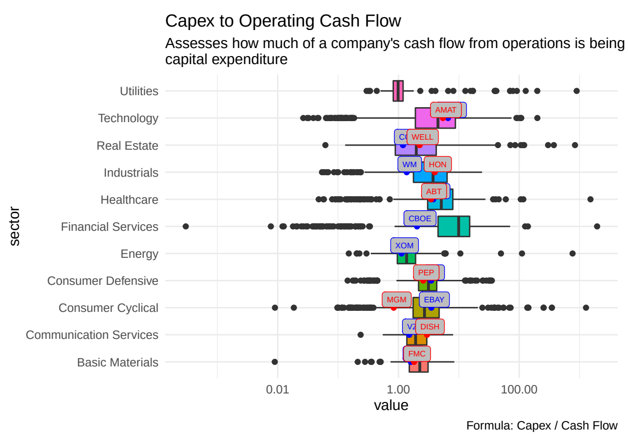
It tells you what percent of every dollar your company earned gets suck SG&A costs



Formula: Total SG&A / Total sales revenue

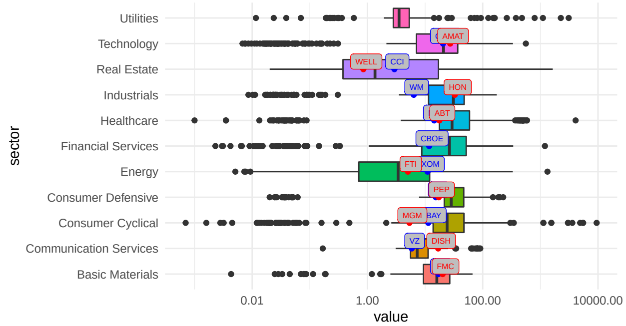

R&D to Revenue
Measures the percentage of sales that is allocated to R&D expenditure
is not as effective when looking at companies in different industries bec
different industries place different values on R&D
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Total R&D / Total sales revenue


Intangibles to Total Assets
The goodwill to assets ratio is a financial measurement that compares t
intangible assets like a brand name, customer list, or unique position in
industry to the total assets of the company in an effect to see if goodwil
being recorded properly
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.0001
0.0010
0.0100
0.1000
1.0000
value
Formula: Intangibles / Total Assets


Capex to Operating Cash Flow
Assesses how much of a company's cash flow from operations is being
capital expenditure
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.01
1.00
100.00
value
Formula: Capex / Cash Flow

Capex to Revenue

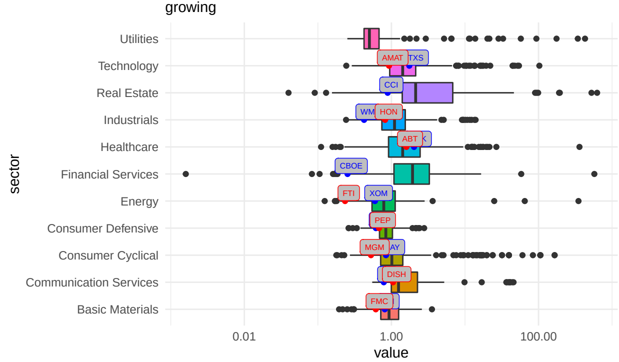
The Capex to Revenue ratio measures a company's investments in prop
equipment and other capital assets to its total sales. The ratio shows ho
aggressively the company is re−investing its revenue back into producti



Formula: Capex / Revenue

Capex to Depreciation

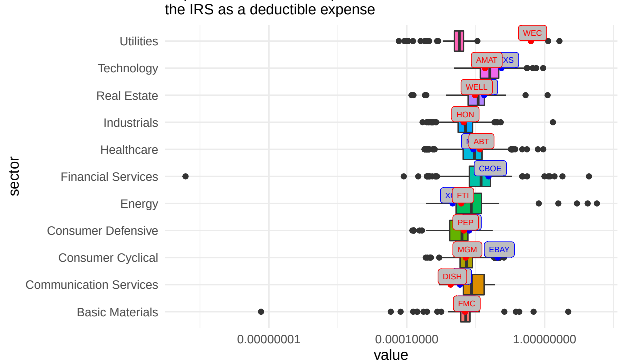
If a company regularly has more CapEx than depreciation, its asset bas
growing



Formula: Capex / Depreciation

Stock−based compensation to Revenue

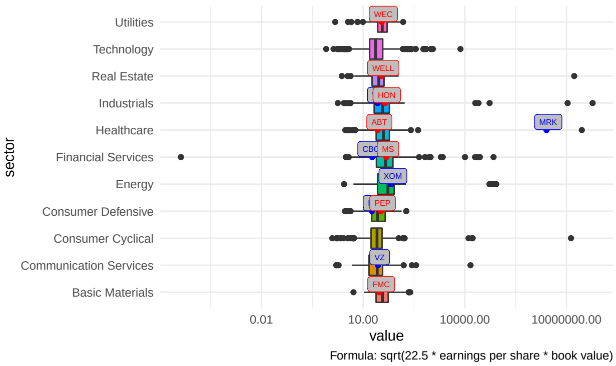
Represents a noncash expense that reduces book income, which isn't r
the IRS as a deductible expense



Formula: NA

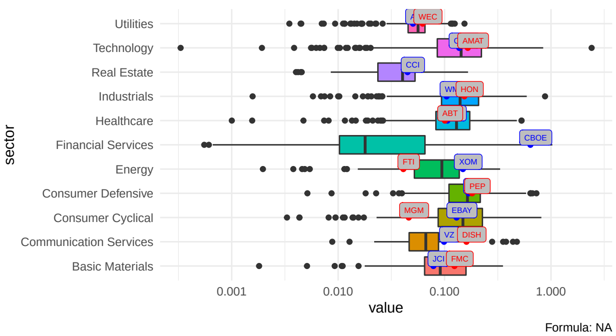
Graham Number

The Graham number is a figure that measures a stock's fundamental va
taking into account the company's earnings per share and book value p



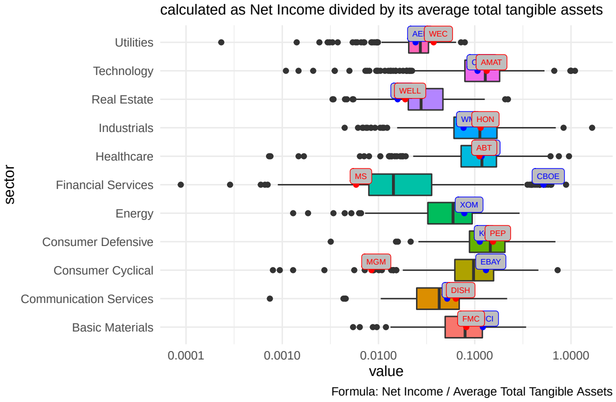
ROIC

Return on invested capital, is a ratio used in finance, valuation and
accounting, as a measure of the profitability and value−creating potenti
companies relative to the amount of capital invested by shareholders ar
debtholders



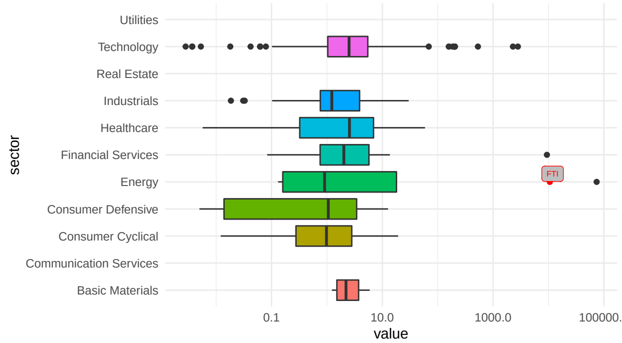
Return on Tangible Assets

calculated as Net Income divided by its average total tangible assets



Graham Net−Net

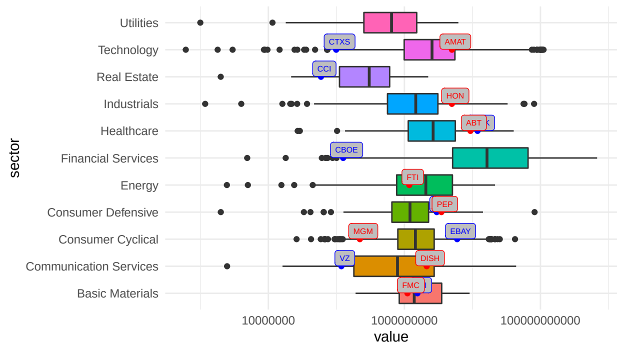
Net−net is a value investing technique in which a company is valued ba
on its net current assets



Formula: (Current Assets − Total Liabilities + Preferred Stock / Shares Outstanding

Working Capital

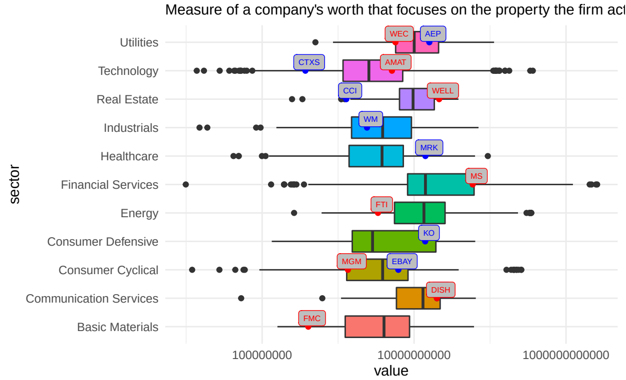
Is a financial metric which represents operating liquidity available to a
business, organization, or other entity



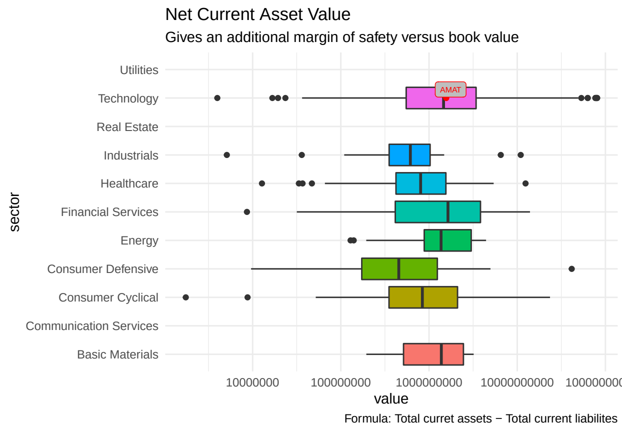
Formula: Accounts receivable + Inventory − Accounts payable

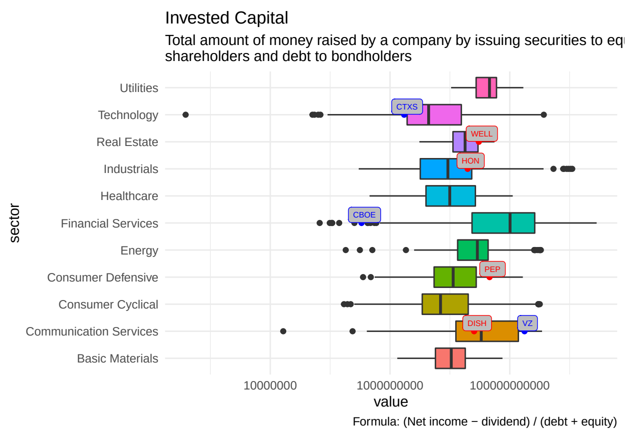
Tangible Asset Value

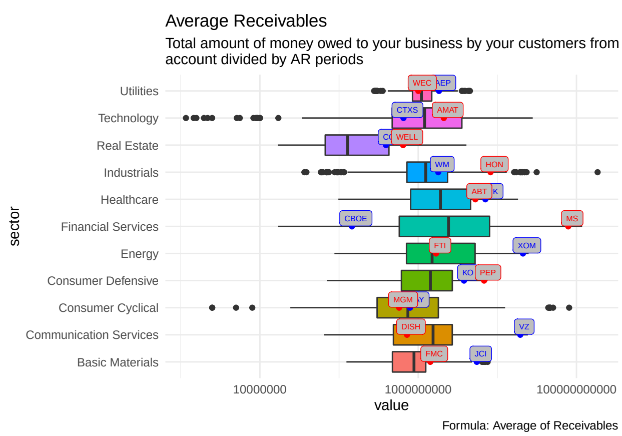
Measure of a company's worth that focuses on the property the firm act



Formula: Total assets – Intangible assets – Total liabilities

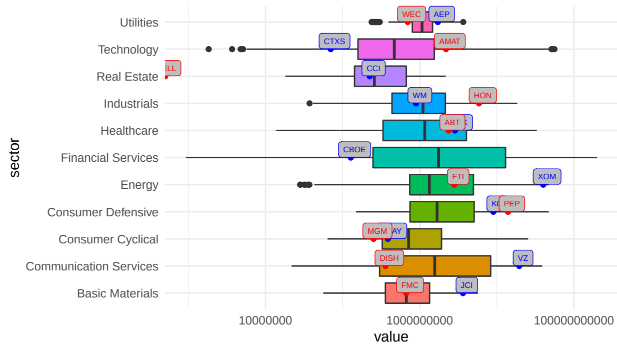

Net Current Asset Value
Gives an additional margin of safety versus book value
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
AMAT
10000000
100000000
1000000000
10000000000
100000000
value
Formula: Total curret assets − Total current liabilites


Invested Capital
Total amount of money raised by a company by issuing securities to eq
shareholders and debt to bondholders
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
CTXS
WELL
HON
CBOE
PEP
DISH
VZ
10000000
1000000000
100000000000
value
Formula: (Net income − dividend) / (debt + equity)


Average Receivables
Total amount of money owed to your business by your customers from
account divided by AR periods
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
WEC
CTXS
AMAT
WELL
WM
HON
ABT
CBOE
MS
FTI
XOM
KO
PEP
MGM
DISH
VZ
FMC
JCI
10000000
1000000000
100000000000
value
Formula: Average of Receivables

Average Payables

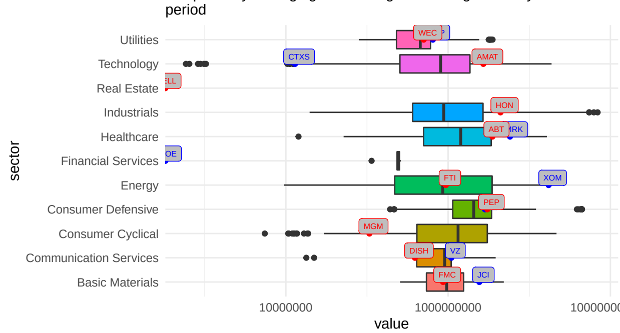
Total amount of money the business owes the customers account divid
of periods



Formula: Average of Payables

Average Inventory

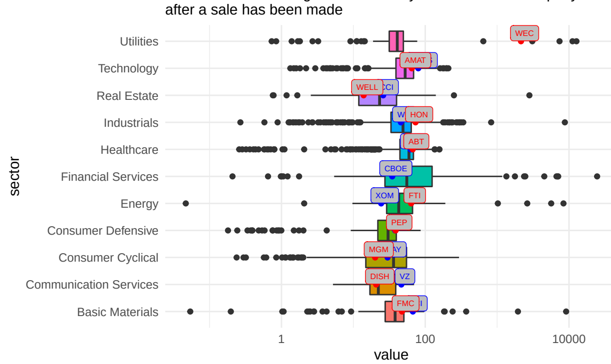
Average inventory is the mean value of an inventory within a certain tim
period, which may vary from the median value of the same data set, an
computed by averaging the starting and ending inventory values over a
period



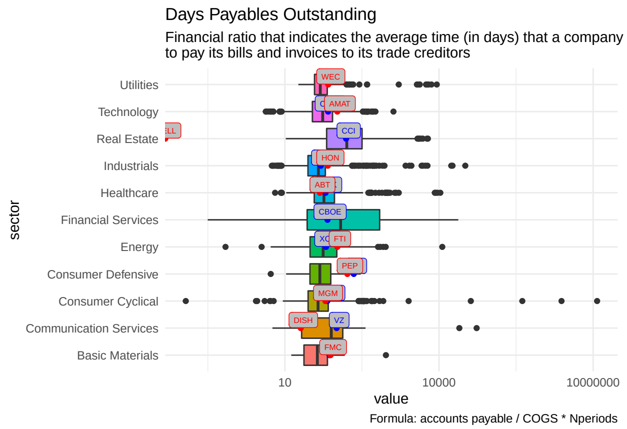
Formula: Average of inventory

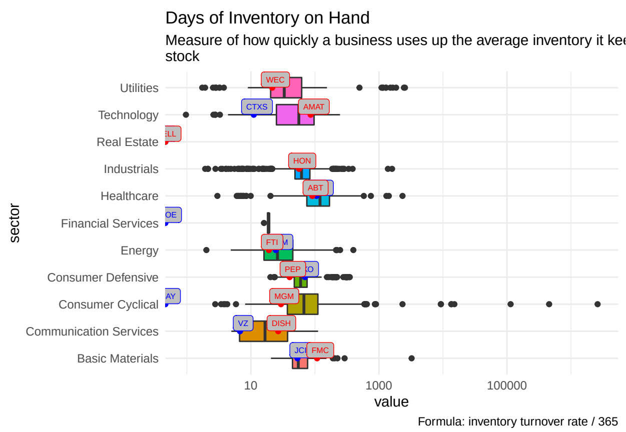
Days Sales Outstanding

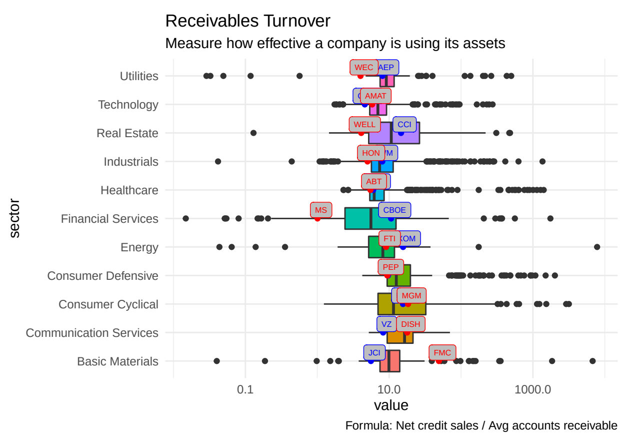
Measure of the average number of days that it takes a company to colle
after a sale has been made



Formula: accounts receivable / Total Credits sales * Nperiods

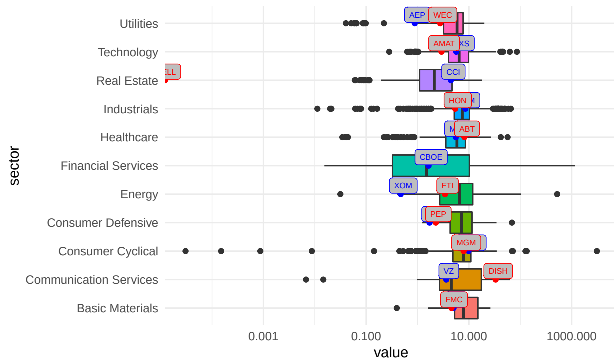

Days Payables Outstanding
Financial ratio that indicates the average time (in days) that a company
to pay its bills and invoices to its trade creditors
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10
10000
10000000
value
Formula: accounts payable / COGS * Nperiods


Days of Inventory on Hand
Measure of how quickly a business uses up the average inventory it kee
stock
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
10
1000
100000
value
Formula: inventory turnover rate / 365


Receivables Turnover
Measure how effective a company is using its assets
Utilities
Technology
Real Estate
Industrials
Healthcare
Financial Services
Energy
Consumer Defensive
Consumer Cyclical
Communication Services
Basic Materials
sector
0.1
10.0
1000.0
value
Formula: Net credit sales / Avg accounts receivable

Payables Turnover

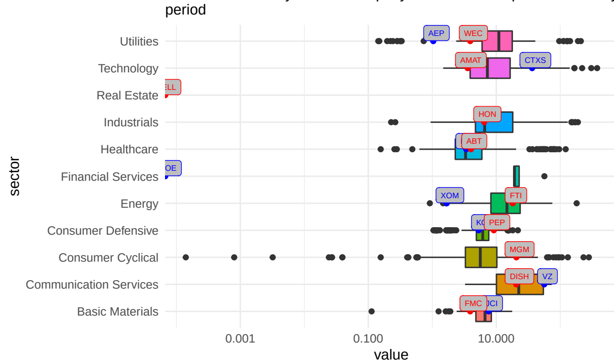
Measure how efficient a company is at paying its suppliers and short−te



Formula: Total supply payables / (Diff Accounts payable / 2)

Inventory Turnover

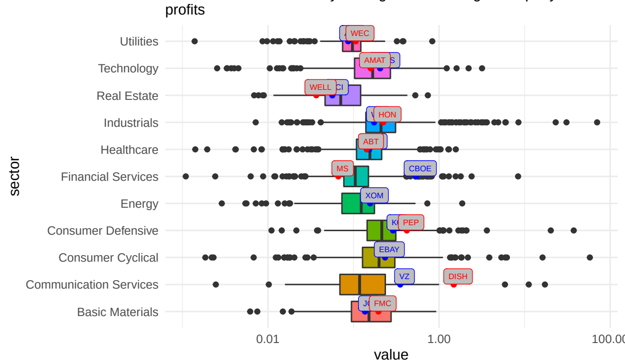
Measure how many times a company has sold and replaced inventory d
period



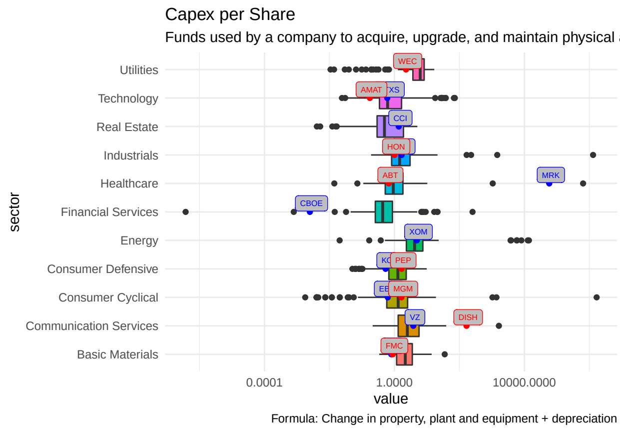
Formula: Cost of sales / Average stock held

ROE

Measure of how effectively management is using a company's assets to
profits



Formula: Net Income / Avg Shareholders equity



Thats awesome!! It does seem that some metrics and sector differ in relation to portfolio risk mitigation.

Let's just remember first that:

1. Sectors are correlated and we managed to find 2 clusters between them, this will be ignored in order to evaluate them per each sector
2. Metrics are correlated and we found that 32 of them are strongly correlated, but the main goal of this project is to evaluate per metric how they infer as a good result or bad result for risk mitigation. This is important though, if you're willing to analyze any subject including these metrics together, the strongest correlation will bias your model.
3. Prices have erratic movement that might injury any machine learning models and must be annomalized, this will be used for last chapter of this project when we model forecast for prices

Model

In order to model this, we'll follow the strategy:

1. List group sectors and get symbols weight the porftolio risk mitigation
2. Model how each metric value associates to the risk weight

In order to model it we'll use a GLM function from H2O wich consists of the formula:

$$\hat{y} = x^T \beta + \beta_0$$

Let's model it:

```
#Part 1
ListGroups <- Symbol_Returns %>%
              select(sector, weights) %>%
              unnest() %>%
              rename("symbol"="Name") #Rename to make it easier for innner join

# Join weight symbols, remove sector from nested data and nest it again...
KeyMetrics <- KeyMetrics %>%
               mutate(data2 = map(data, ~..1 %>%
                                    inner_join(ListGroups, by = c("sector","symbol"))))



# Part 2
```

```r
#Let's create a function with h2o to help us model per sector

h2o.init(max_mem_size = "5g")
```

```
##
## H2O is not running yet, starting it now...
##
## Note:  In case of errors look at the following log files:
##     C:\Users\EDUARD~1.ALM\AppData\Local\Temp\RtmpABkP1t/h2o_eduardo_almeida_started_from_r.out
##     C:\Users\EDUARD~1.ALM\AppData\Local\Temp\RtmpABkP1t/h2o_eduardo_almeida_started_from_r.err
##
##
## Starting H2O JVM and connecting:  Connection successful!
##
## R is connected to the H2O cluster:
##     H2O cluster uptime:         3 seconds 491 milliseconds
##     H2O cluster timezone:       America/Sao_Paulo
##     H2O data parsing timezone:  UTC
##     H2O cluster version:        3.26.0.2
##     H2O cluster version age:    8 months !!!
##     H2O cluster name:           H2O_started_from_R_eduardo.almeida_snc915
##     H2O cluster total nodes:    1
##     H2O cluster total memory:   4.44 GB
##     H2O cluster total cores:    8
##     H2O cluster allowed cores:  8
##     H2O cluster healthy:        TRUE
##     H2O Connection ip:          localhost
##     H2O Connection port:        54321
##     H2O Connection proxy:       NA
##     H2O Internal Security:      FALSE
##     H2O API Extensions:         Amazon S3, Algos, AutoML, Core V3, Core V4
##     R Version:                  R version 3.6.2 (2019-12-12)
```

```r
h2o.no_progress()

H2o_Model <- function(Data,x, y){
  require(h2o)
  require(tidyverse)
  require(purrr)

  Data$sector <- as.factor(Data$sector)

  Data_h2o <- as.h2o(Data)

  set.seed(123)

  automl_glm <- h2o.glm(
    x = x,
    y = y,
    training_frame = Data_h2o)

  Name_Model <- "H2O_GLM"

  coef <- h2o.coef(automl_glm) %>%
```

```r
          as.data.frame()

  #Model summary
  CV_Summary <- h2o.performance(automl_glm)

  #lime
  # explainer <- lime(x = Data[,x],
                     # model = automl_glm)

  # explanation <- explain(x = Data[,x], explainer = explainer, bin_continuous = TRUE,
                          # feature_select = "auto", n_features = 2)

  # Features_Plot <- plot_features(explanation,cases = 1)

  aux <- data.frame(Model_Name = Name_Model) %>%
    mutate(Model = map(Model_Name, ~automl_glm),
           coefs = map(Model_Name, ~coef),
           CV_Summary = map(Model_Name, ~CV_Summary)#,
           #explanation = map(Model_Name, ~explanation),
           #Features_Plot = map(Model_Name, ~Features_Plot)
    )


  return(aux)

}


# Let's model
KeyMetrics <- KeyMetrics %>%
               mutate(H2o_Model = map(data2, ~H2o_Model(..1, x = c("value","sector"),
                                                        y = "weights")))


h2o.shutdown(prompt = F)
```

```
## [1] TRUE
```

```r
#Organize coeficients for kable
Coefiecients <- KeyMetrics %>%
                mutate(coefs = map(H2o_Model, ~..1$coefs %>%
                              reduce(as.data.frame) %>%
                              rownames_to_column(var = "Parameter"))) %>%
                ungroup() %>%
                select(metric, coefs) %>%
                unnest(coefs) %>%
                rename("value" = ".") %>%
                mutate(Parameter = gsub(Parameter, pattern = "sector.", replacement = "")) %>%
                spread(key = Parameter, value = value) %>%
                select(metric, Intercept, value, everything()) %>%
                arrange(-Intercept, -value)
```

Let's check the coeficients

```r
kable(Coefieciients, caption = "Metrics coeficients per sectors", digits = 3) %>%
  kable_styling(bootstrap_options = "striped", full_width = F, font_size = 10)
```

This is great! As showed in the charts before, it does seem that metrics and sectors differs between each other in terms of risk mitigation.

In order to understand this table we must first understand how it works:

1. The value is a weight multiplication to the metric value, when it's negative it means that metrics value that are negatives will deliver higher risk mitigation.
2. When we get a difference between values and sector weights we can actually see that some sectors tend to increase or decrease the final risk mitigation by the metric

That means that we can actually measure theses metrics by comparing the highest ratio of difference from value to sector values in regards of negative x positive weights in order to understand:

1. Metrics that deliver the higher information gain for each sector
2. Metrics that deliver higher information gain in overall

This will not be done in this project.

Conclusion

This projects seeked to evaluate how each metric is related to a risk mitigation in a portfolio and how it differs per sectors.

For future projects we can measure:

1. Metrics that deliver the higher information gain for each sector
2. Metrics that deliver higher information gain in overall
3. Impact of other variables such as macroeconomic metrics and sentiment analysis of news
4. Predict the future price of models and re-model it for portfolio mitigation
5. Compare companies that are not in SP500 or have lower prices but have a good result on metrics and assess how it is used for risk mitigation with a lower price bond

```r
sessionInfo()
```

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18362)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Portuguese_Brazil.1252  LC_CTYPE=Portuguese_Brazil.1252
## [3] LC_MONETARY=Portuguese_Brazil.1252 LC_NUMERIC=C
## [5] LC_TIME=Portuguese_Brazil.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] factoextra_1.0.6           broom_0.5.4
##  [3] tsibble_0.8.6              ggdendro_0.1-20
##  [5] gridExtra_2.3              lime_0.5.1
##  [7] h2o_3.26.0.2               ggstatsplot_0.3.1
##  [9] cowplot_1.0.0              kableExtra_1.1.0
```

Table 5: Metrics coeficients per sectors

metric	Intercept	value	Basic Materials	Communication Services	Consumer Cycl
Debt to Assets	0.041	-0.013	0.035	0.077	-0.
Working Capital	0.033	0.000	0.035	0.074	-0.
Payables Turnover	0.033	0.000	0.035	0.075	-0.
Invested Capital	0.033	0.000	0.035	0.075	-0.
Net Current Asset Value	0.033	0.000	0.035	0.075	-0.
Stock-based compensation to Revenue	0.033	-0.001	0.035	0.075	-0.
Current ratio	0.033	0.000	0.035	0.075	-0.
Days Payables Outstanding	0.033	0.000	0.035	0.075	-0.
Days of Inventory on Hand	0.033	0.000	0.035	0.075	-0.
EV to Sales	0.033	0.000	0.035	0.075	-0.
Price to Sales Ratio	0.033	0.000	0.035	0.075	-0.
POCF ratio	0.033	0.000	0.035	0.075	-0.
EV to Operating cash flow	0.033	0.000	0.035	0.075	-0.
PB ratio	0.033	0.000	0.035	0.075	-0.
PTB ratio	0.033	0.000	0.035	0.075	-0.
Enterprise Value over EBITDA	0.033	0.000	0.035	0.075	-0.
PE ratio	0.033	0.000	0.035	0.075	-0.
Market Cap	0.033	0.000	0.035	0.075	-0.
PFCF ratio	0.033	0.000	0.035	0.075	-0.
EV to Free cash flow	0.033	0.000	0.035	0.075	-0.
Enterprise Value	0.033	0.000	0.035	0.075	-0.
Days Sales Outstanding	0.033	0.000	0.035	0.075	-0.
Net Debt to EBITDA	0.033	0.000	0.035	0.075	-0.
SG&A to Revenue	0.033	0.000	0.035	0.075	-0.
R&D to Revenue	0.033	0.000	0.035	0.075	-0.
Receivables Turnover	0.033	0.000	0.035	0.075	-0.
Debt to Equity	0.033	0.000	0.035	0.075	-0.
Revenue per Share	0.033	0.000	0.035	0.075	-0.
Capex per Share	0.033	0.000	0.035	0.075	-0.
Graham Net-Net	0.033	0.000	0.035	0.075	-0.
ROE	0.033	0.000	0.035	0.075	-0.
Tangible Book Value per Share	0.033	0.000	0.035	0.075	-0.
Interest Debt per Share	0.033	0.000	0.035	0.075	-0.
Net Income per Share	0.033	0.000	0.035	0.075	-0.
Operating Cash Flow per Share	0.033	0.000	0.035	0.075	-0.
Free Cash Flow per Share	0.033	0.000	0.035	0.075	-0.
Free Cash Flow Yield	0.033	0.000	0.035	0.075	-0.
Cash per Share	0.033	0.000	0.035	0.075	-0.
Graham Number	0.033	0.000	0.035	0.075	-0.
Earnings Yield	0.033	0.000	0.035	0.075	-0.
Shareholders Equity per Share	0.033	0.000	0.035	0.075	-0.
Book Value per Share	0.033	0.000	0.035	0.075	-0.
Dividend Yield	0.033	0.000	0.035	0.075	-0.
Payout Ratio	0.033	0.000	0.035	0.075	-0.
Income Quality	0.033	0.000	0.035	0.075	-0.
Capex to Revenue	0.033	0.000	0.035	0.075	-0.
Interest Coverage	0.033	0.000	0.035	0.075	-0.
Capex to Operating Cash Flow	0.033	0.000	0.035	0.075	-0.
Tangible Asset Value	0.032	0.000	0.036	0.075	-0.
Capex to Depreciation	0.032	0.000	0.035	0.074	-0.
Average Payables	0.032	0.000	0.036	0.074	-0.
Inventory Turnover	0.032	0.000	0.035	0.074	-0.
Return on Tangible Assets	0.032	0.015	0.035	0.076	-0.
Average Receivables	0.032	0.000	0.036	0.074	-0.
ROIC	0.031	0.019	0.035	0.076	-0.

```
## [11] knitr_1.26                 ROI.plugin.quadprog_0.2-5
## [13] ROI.plugin.glpk_0.3-0      ROI_0.3-3
## [15] PortfolioAnalytics_1.1.0   foreach_1.4.7
## [17] tibbletime_0.1.3           timetk_0.1.2
## [19] readxl_1.3.1               anytime_0.3.7
## [21] tidyquant_0.5.10           quantmod_0.4-15
## [23] TTR_0.23-6                 PerformanceAnalytics_2.0.4
## [25] xts_0.12-0                 zoo_1.8-7
## [27] lubridate_1.7.4            patchwork_1.0.0
## [29] janitor_1.2.1              forcats_0.4.0
## [31] stringr_1.4.0              dplyr_0.8.3
## [33] purrr_0.3.3                readr_1.3.1
## [35] tidyr_1.0.2                tibble_2.1.3
## [37] ggplot2_3.3.0              tidyverse_1.3.0
## [39] jsonlite_1.6               httr_1.4.1
##
## loaded via a namespace (and not attached):
##   [1] estimability_1.3          coda_0.19-3
##   [3] multcomp_1.4-12           data.table_1.12.8
##   [5] inline_0.3.15             RCurl_1.95-4.12
##   [7] generics_0.0.2            callr_3.4.0
##   [9] TH.data_1.0-10            webshot_0.5.2
##  [11] xml2_1.2.2                httpuv_1.5.2
##  [13] StanHeaders_2.21.0-1      assertthat_0.2.1
##  [15] gower_0.2.1               WRS2_1.0-0
##  [17] xfun_0.11                 hms_0.5.3
##  [19] evaluate_0.14             promises_1.1.0
##  [21] fansi_0.4.1               dbplyr_1.4.2
##  [23] htmlwidgets_1.5.1         DBI_1.1.0
##  [25] reshape_0.8.8             Quandl_2.10.0
##  [27] stats4_3.6.2              ellipsis_0.3.0
##  [29] paletteer_1.1.0           rcompanion_2.3.25
##  [31] backports_1.1.5           insight_0.8.2
##  [33] ggcorrplot_0.1.3          libcoin_1.0-5
##  [35] jmvcore_1.2.5             vctrs_0.2.1
##  [37] sjlabelled_1.1.3          abind_1.4-5
##  [39] withr_2.1.2               metaBMA_0.6.2
##  [41] bdsmatrix_1.3-4           emmeans_1.4.5
##  [43] prettyunits_1.0.2         fastGHQuad_1.0
##  [45] mnormt_1.5-6              cluster_2.1.0
##  [47] lazyeval_0.2.2            Rglpk_0.6-4
##  [49] crayon_1.3.4              labeling_0.3
##  [51] glmnet_3.0-2              pkgconfig_2.0.3
##  [53] slam_0.1-47               nlme_3.1-142
##  [55] statsExpressions_0.3.1    palr_0.2.0
##  [57] pals_1.6                  rlang_0.4.2
##  [59] lifecycle_0.1.0           miniUI_0.1.1.1
##  [61] groupedstats_0.2.1        skimr_2.1
##  [63] LaplacesDemon_16.1.4      MatrixModels_0.4-1
##  [65] sandwich_2.5-1            registry_0.5-1
##  [67] EMT_1.1                   modelr_0.1.5
##  [69] dichromat_2.0-0           cellranger_1.1.0
##  [71] matrixStats_0.55.0        broomExtra_2.5.0
##  [73] lmtest_0.9-37             Matrix_1.2-18
```

```
##  [75] loo_2.2.0                  mc2d_0.1-18
##  [77] carData_3.0-3              boot_1.3-23
##  [79] reprex_0.3.0               base64enc_0.1-3
##  [81] processx_3.4.1             viridisLite_0.3.0
##  [83] rjson_0.2.20               oompaBase_3.2.9
##  [85] bitops_1.0-6               parameters_0.6.0
##  [87] ggExtra_0.9                shape_1.4.4
##  [89] multcompView_0.1-8         coin_1.3-1
##  [91] ggsignif_0.6.0             scales_1.1.0
##  [93] magrittr_1.5               plyr_1.8.5
##  [95] compiler_3.6.2             rstantools_2.0.0
##  [97] bbmle_1.0.23.1             RColorBrewer_1.1-2
##  [99] lme4_1.1-21                cli_2.0.1
## [101] pbapply_1.4-2              ps_1.3.0
## [103] TMB_1.7.16                 Brobdingnag_1.2-6
## [105] MASS_7.3-51.4              mgcv_1.8-31
## [107] tidyselect_0.2.5           stringi_1.4.3
## [109] yaml_2.2.0                 ggrepel_0.8.1
## [111] bridgesampling_1.0-0       grid_3.6.2
## [113] tools_3.6.2                parallel_3.6.2
## [115] rio_0.5.16                 rstudioapi_0.10
## [117] foreign_0.8-72             ipmisc_1.2.0
## [119] pairwiseComparisons_0.2.5 farver_2.0.1
## [121] digest_0.6.23              shiny_1.4.0
## [123] nortest_1.0-4              quadprog_1.5-8
## [125] jmv_1.2.5                  Rcpp_1.0.3
## [127] car_3.0-6                  metafor_2.1-0
## [129] ez_4.4-0                   BayesFactor_0.9.12-4.2
## [131] performance_0.4.4          metaplus_0.7-11
## [133] later_1.0.0                psych_1.9.12.31
## [135] effectsize_0.2.0           sjstats_0.17.9
## [137] colorspace_1.4-1           rvest_0.3.5
## [139] fs_1.3.1                   splines_3.6.2
## [141] rematch2_2.1.0             expm_0.999-4
## [143] shinythemes_1.1.2          mapproj_1.2.7
## [145] jcolors_0.0.4              xtable_1.8-4
## [147] nloptr_1.2.1               rstan_2.19.2
## [149] zeallot_0.1.0              modeltools_0.2-23
## [151] scico_1.1.0                R6_2.4.1
## [153] broom.mixed_0.2.4          pillar_1.4.3
## [155] htmltools_0.4.0            mime_0.8
## [157] glue_1.3.1                 fastmap_1.0.1
## [159] minqa_1.2.4                codetools_0.2-16
## [161] maps_3.3.0                 pkgbuild_1.0.6
## [163] mvtnorm_1.0-12             lattice_0.20-38
## [165] numDeriv_2016.8-1.1        curl_4.3
## [167] DescTools_0.99.34          gtools_3.8.1
## [169] logspline_2.1.15           zip_2.0.4
## [171] openxlsx_4.1.4             survival_3.1-8
## [173] rmarkdown_2.0              repr_1.1.0
## [175] munsell_0.5.0              iterators_1.0.12
## [177] sjmisc_2.8.3               haven_2.2.0
## [179] reshape2_1.4.3             gtable_0.3.0
## [181] bayestestR_0.5.2
```