

Live Case: S&P500 (2 of 3)

Aug 10, 2023. --

Agenda: Understanding the S&P500 Index

S&P 500 Data - PRELIMINARY SETUP

1. We will continue our analysis of the S&P 500. Load the data, as described in the chapter Live Case: S&P500 (1 of 3)

```
# Read S&P500 stock data present in a Google Sheet.  
library(gsheet)  
prefix <- "https://docs.google.com/spreadsheets/d/"  
sheetID <- "11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM"  
url500 <- paste(prefix,sheetID) # Form the URL to connect to  
sp500 <- gsheets2tbl(url500) # Read it into a tibble called sp500
```

No encoding supplied: defaulting to UTF-8.

2. Rename columns, as described in the chapter Live Case: S&P500 (1 of 3).

```
suppressPackageStartupMessages(library(dplyr))  
  
# Define a mapping of new column names  
new_names <- c(  
  "Date", "Stock", "StockName", "Sector", "Industry",  
  "MarketCap", "Price", "Low52Wk", "High52Wk",  
  "ROE", "ROA", "ROIC", "GrossMargin",  
  "OperatingMargin", "NetMargin", "PE",  
  "PB", "EVEBITDA", "EBITDA", "EPS",  
  "EBITDA_YOY", "EBITDA_QYOY", "EPS_YOY",  
  "EPS_QYOY", "PFCF", "FCF",  
  "FCF_QYOY", "DebtToEquity", "CurrentRatio",  
  "QuickRatio", "DividendYield",
```

```

    "DividendsPerShare_YOY", "PS",
    "Revenue_YOY", "Revenue_QYOY", "Rating"
  )
# Rename the columns using the new_names vector
sp500 <- sp500 %>%
  rename_with(~ new_names, everything())

```

3. Remove Rows containing no data or Null values, as described in the chapter Live Case: S&P500 (1 of 3).

```

# Check for blank or null values in the "Stock" column
hasNull <- any(sp500$Stock == "" | is.null(sp500$Stock))
if (hasNull) {
  # Remove rows with null or blank values from the dataframe tibble
  sp500 <- sp500[!(is.null(sp500$Stock) | sp500$Stock == ""), ]
}

```

4. The S&P500 shares are divided into multiple Sectors. Thus, model Sector as a factor() variable, as described in the chapter Live Case: S&P500 (1 of 3).

```

sp500$Sector <- as.factor(sp500$Sector)

```

5. Stock Ratings: The S&P500 shares have Technical Ratings such as {Buy, Sell, ..}. Model the data column Rating as a factor() variable, as described in the chapter Live Case: S&P500 (1 of 3).

```

sp500$Rating <- as.factor(sp500$Rating)

```

6. Low52WkPerc: Create a new column to track Share Prices relative to their 52 Week Low, as described in the chapter Live Case: S&P500 (1 of 3).

```

sp500 <- sp500 %>% mutate(Low52WkPerc = round((Price - Low52Wk)*100 / Low52Wk,2))
colnames(sp500)

```

[1] "Date"	"Stock"	"StockName"
[4] "Sector"	"Industry"	"MarketCap"
[7] "Price"	"Low52Wk"	"High52Wk"
[10] "ROE"	"ROA"	"ROIC"

[13]	"GrossMargin"	"OperatingMargin"	"NetMargin"
[16]	"PE"	"PB"	"EVEBITDA"
[19]	"EBITDA"	"EPS"	"EBITDA_YOY"
[22]	"EBITDA_QYOY"	"EPS_YOY"	"EPS_QYOY"
[25]	"PFCF"	"FCF"	"FCF_QYOY"
[28]	"DebtToEquity"	"CurrentRatio"	"QuickRatio"
[31]	"DividendYield"	"DividendsPerShare_YOY"	"PS"
[34]	"Revenue_YOY"	"Revenue_QYOY"	"Rating"
[37]	"Low52WkPerc"		

Well done! Our data is now ready for analysis!!

7. Creating a new column MarketCapBillions = MarketCap/1000,000,000

```
sp500 <- sp500 %>% mutate(MarketCapBillions = MarketCap/ 1000000000)
colnames(sp500)
```

[1]	"Date"	"Stock"	"StockName"
[4]	"Sector"	"Industry"	"MarketCap"
[7]	"Price"	"Low52Wk"	"High52Wk"
[10]	"ROE"	"ROA"	"ROIC"
[13]	"GrossMargin"	"OperatingMargin"	"NetMargin"
[16]	"PE"	"PB"	"EVEBITDA"
[19]	"EBITDA"	"EPS"	"EBITDA_YOY"
[22]	"EBITDA_QYOY"	"EPS_YOY"	"EPS_QYOY"
[25]	"PFCF"	"FCF"	"FCF_QYOY"
[28]	"DebtToEquity"	"CurrentRatio"	"QuickRatio"
[31]	"DividendYield"	"DividendsPerShare_YOY"	"PS"
[34]	"Revenue_YOY"	"Revenue_QYOY"	"Rating"
[37]	"Low52WkPerc"	"MarketCapBillions"	

ANALYSIS OF S&P500 SECTORS

1. Number of shares in Each Sector

```
library(kableExtra)
```

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

group_rows

```
library(janitor) # This package helps us auto generate the total at the bottom of a table
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
tab1 <- table(Sector = sp500$Sector)
tab2 <- as.data.frame(tab1)
tab3 <- tab2 %>% arrange(desc(Freq))
tab4 <- tab3 %>%
  adorn_totals("row")
tab5 <- knitr::kable(tab4, "html") %>% kable_styling()
tab5
```

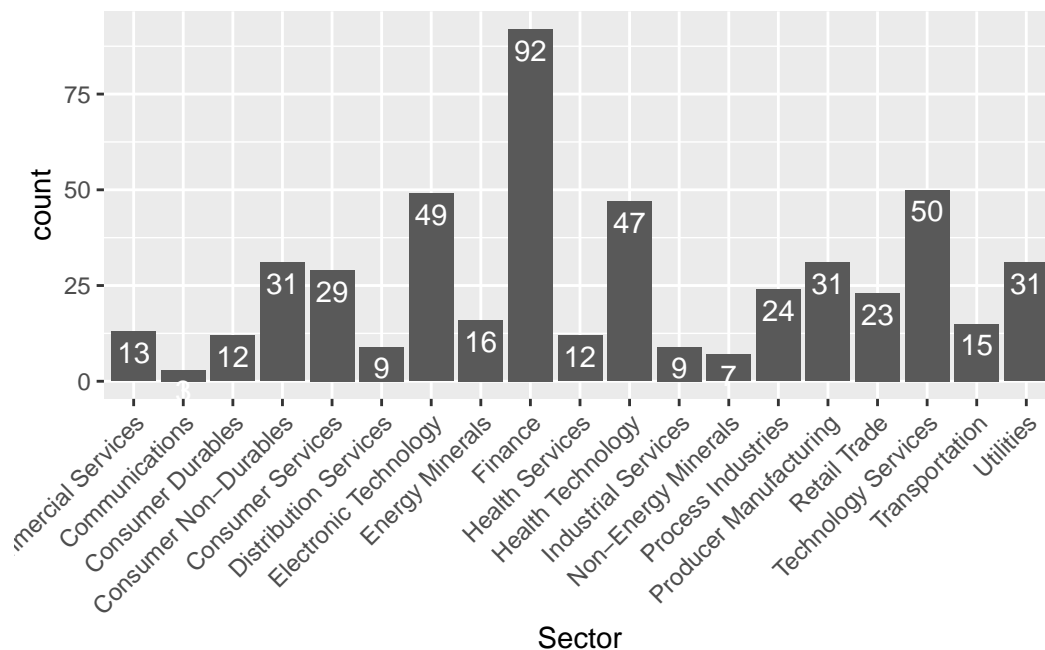
Sector	Freq
Finance	92
Technology Services	50
Electronic Technology	49
Health Technology	47
Consumer Non-Durables	31
Producer Manufacturing	31
Utilities	31
Consumer Services	29
Process Industries	24
Retail Trade	23
Energy Minerals	16
Transportation	15
Commercial Services	13
Consumer Durables	12
Health Services	12
Distribution Services	9
Industrial Services	9
Non-Energy Minerals	7

Sector	Freq
Communications	3
Total	503

2. Barplot for Showing Number of Stocks in Each Sector

```
library(ggplot2)

ggplot(sp500, aes(Sector)) + geom_bar() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  geom_text(aes(label = ..count..), stat = "count", vjust = 1.5, colour = "white")
```



3. Number of shares By Rating

```
library(janitor) # This package helps us auto generate the total at the bottom of a table
tab1 <- table(Rating = sp500$Rating)
tab2 <- as.data.frame(tab1)
tab3 <- tab2 %>% arrange(desc(Freq))
tab4 <- tab3 %>%
  adorn_totals("row")
tab5 <- knitr::kable(tab4, "html") %>% kable_styling()
tab5
```

Rating	Freq
Sell	203
Buy	130
Neutral	85
Strong Sell	72
Strong Buy	13
Total	503

4. Pie Chart Showing Number of shares By Rating

```
library(ggpubr)
# Compute counts of each Rating
Rating_count <- as.data.frame(table(sp500$Rating))
colnames(Rating_count) <- c("Rating", "n")

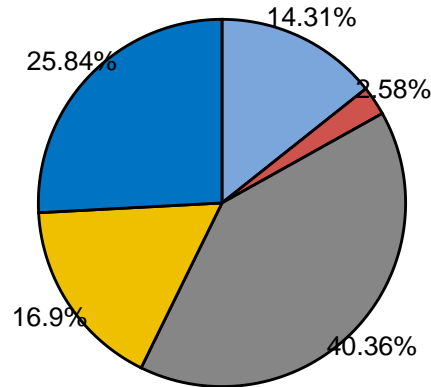
# Calculate proportions
Rating_count$prop <- Rating_count$n / sum(Rating_count$n)

# Create labels that display proportions as percentages
Rating_count$labels <- paste0(round(Rating_count$prop*100, 2), "%")

# Create the pie chart
ggpie(data = Rating_count,
      x = "n",
      fill = "Rating",
      label = "labels",
      palette = "jco",
      title = "Pie Chart of Rating")
```

Pie Chart of Rating

Rating ■ Buy ■ Neutral ■ Sell ■ Strong Buy ■ Strong Sell



5. Number of shares in Each Sector By Rating

```
library(janitor) # This package helps us auto generate the total at the bottom of a table
tab1 <- table(Rating = sp500$Rating, Sector = sp500$Sector)
tab2 <- as.data.frame(tab1)
tab3 <- tab2 %>% arrange(desc(Freq))
# excluding rows having 0 counts
tab3 <- subset(tab3 ,Freq > 0 )
tab4 <- tab3 %>%
  adorn_totals("row")
tab5 <- knitr::kable(tab4, "html") %>% kable_styling()
tab5
```

Rating	Sector	Freq
Buy	Finance	31
Sell	Electronic Technology	27
Sell	Finance	27
Sell	Technology Services	27
Neutral	Finance	22
Sell	Health Technology	22
Sell	Consumer Services	17
Sell	Utilities	15
Buy	Electronic Technology	12
Buy	Energy Minerals	12
Strong Sell	Health Technology	11

Rating	Sector	Freq
Buy	Consumer Non-Durables	10
Sell	Producer Manufacturing	10
Buy	Retail Trade	10
Strong Sell	Technology Services	10
Neutral	Utilities	10
Neutral	Consumer Non-Durables	9
Sell	Consumer Non-Durables	9
Strong Sell	Finance	9
Buy	Health Technology	9
Sell	Process Industries	9
Buy	Technology Services	9
Buy	Process Industries	8
Buy	Producer Manufacturing	8
Strong Sell	Producer Manufacturing	8
Sell	Retail Trade	7
Sell	Consumer Durables	6
Strong Sell	Electronic Technology	6
Buy	Transportation	6
Sell	Commercial Services	5
Sell	Distribution Services	5
Neutral	Producer Manufacturing	5
Neutral	Consumer Durables	4
Buy	Consumer Services	4
Neutral	Consumer Services	4
Neutral	Electronic Technology	4
Neutral	Health Services	4
Sell	Health Services	4
Strong Sell	Health Services	4
Neutral	Health Technology	4
Buy	Non-Energy Minerals	4
Strong Sell	Process Industries	4
Neutral	Technology Services	4
Sell	Transportation	4
Strong Sell	Utilities	4
Buy	Commercial Services	3
Strong Sell	Commercial Services	3
Sell	Communications	3
Strong Sell	Consumer Non-Durables	3
Strong Buy	Finance	3
Neutral	Industrial Services	3
Sell	Industrial Services	3

Rating	Sector	Freq
Neutral	Transportation	3
Neutral	Commercial Services	2
Strong Sell	Consumer Durables	2
Strong Buy	Consumer Services	2
Strong Sell	Consumer Services	2
Neutral	Distribution Services	2
Strong Buy	Energy Minerals	2
Strong Sell	Industrial Services	2
Sell	Non-Energy Minerals	2
Strong Buy	Process Industries	2
Neutral	Retail Trade	2
Strong Buy	Retail Trade	2
Strong Sell	Retail Trade	2
Strong Sell	Transportation	2
Buy	Utilities	2
Buy	Distribution Services	1
Strong Buy	Distribution Services	1
Neutral	Energy Minerals	1
Sell	Energy Minerals	1
Strong Buy	Health Technology	1
Buy	Industrial Services	1
Neutral	Non-Energy Minerals	1
Neutral	Process Industries	1
Total	-	503

MARKET CAP

1. Market Cap of all companies by Sector

```
library(janitor) # This package helps us auto generate the total at the bottom of a table
library(kableExtra)

# Market Cap by Sector
MCap <- sp500 %>%
  group_by(Sector) %>%
  summarise(
    MarketCapBi = sum(MarketCapBillions))

# Total Market Cap of the entire S&P 500 (in Millions)
```

```

SP500MarketCapBi <- sum(sp500$MarketCapBillions)

# calculating % market cap
PercentMarketCap <- round(MCap$MarketCapBi*100/SP500MarketCapBi,2)
MCapTab <- cbind(MCap,PercentMarketCap)

# sorting by PercentMarketCap
MCapTab <- MCapTab %>% arrange(desc(PercentMarketCap))

# Use package janitor to add sums at the bottom of the table
MCapTab <- MCapTab %>%
  adorn_totals("row")

# Use package knitr to format the appearance of the table
MCapTab <- knitr::kable(MCapTab, "html") %>% kable_styling()
MCapTab

```

Sector	MarketCapBi	PercentMarketCap
Technology Services	9389.67	22.63
Electronic Technology	6881.89	16.59
Finance	4817.87	11.61
Health Technology	3904.67	9.41
Retail Trade	3084.19	7.43
Consumer Non-Durables	2130.55	5.14
Producer Manufacturing	1406.97	3.39
Consumer Services	1403.52	3.38
Energy Minerals	1364.10	3.29
Commercial Services	1272.59	3.07
Consumer Durables	1086.98	2.62
Utilities	980.02	2.36
Health Services	943.70	2.27
Process Industries	791.45	1.91
Transportation	700.13	1.69
Industrial Services	405.70	0.98
Communications	398.00	0.96
Distribution Services	310.00	0.75
Non-Energy Minerals	217.47	0.52
Total	41489.47	100.00

2. Summary Statistics of Market Cap (in Billion) by each Sector of S&P500

```

SectorMC <- sp500 %>%
  group_by(Sector) %>%
  summarise(
    Total = sum(na.omit(MarketCapBillions)),
    Median= median(na.omit(MarketCapBillions)),
    Mean= mean(na.omit(MarketCapBillions)))

tab <- cbind(Sector = SectorMC$Sector, round(SectorMC[,2:4],2))

SMcap <- knitr::kable(tab, "html") %>% kable_styling()
SMcap

```

Sector	Total	Median	Mean
Commercial Services	1272.59	31.10	97.89
Communications	398.00	138.00	132.67
Consumer Durables	1086.98	19.15	90.58
Consumer Non-Durables	2130.55	43.70	68.73
Consumer Services	1403.52	21.50	48.40
Distribution Services	310.00	32.50	34.44
Electronic Technology	6881.89	43.70	140.45
Energy Minerals	1364.10	48.25	85.26
Finance	4817.87	25.60	52.37
Health Services	943.70	37.75	78.64
Health Technology	3904.67	37.00	83.08
Industrial Services	405.70	38.80	45.08
Non-Energy Minerals	217.47	29.70	31.07
Process Industries	791.45	18.15	32.98
Producer Manufacturing	1406.97	34.30	45.39
Retail Trade	3084.19	36.90	134.10
Technology Services	9389.67	34.15	187.79
Transportation	700.13	28.50	46.68
Utilities	980.02	23.90	31.61

3. Top 10 companies having highest Market Cap

```

Top10 <- sp500 %>% arrange(desc(MarketCapBillions)) %>% head(10)
Top10 <- `Top10`[,c(1:4,10:13,37)]

Top10 <- knitr::kable(Top10, "html") %>% kable_styling()
Top10

```

Date	Stock	StockName	Sector	ROE	ROA	ROIC
8/12/2023	AAPL	Apple Inc.	Electronic Technology	146.00	27.600	57.177930
8/12/2023	MSFT	Microsoft Corporation	Technology Services	38.80	18.600	28.062340
8/12/2023	GOOG	Alphabet Inc.	Technology Services	23.30	16.500	21.184760
8/12/2023	GOOGL	Alphabet Inc.	Technology Services	23.30	16.500	21.184760
8/12/2023	AMZN	Amazon.com, Inc.	Retail Trade	2.98	0.981	1.581103
8/12/2023	NVDA	NVIDIA Corporation	Electronic Technology	18.90	10.700	13.098850
8/12/2023	TSLA	Tesla, Inc.	Consumer Durables	28.00	15.400	25.481890
8/12/2023	META	Meta Platforms, Inc.	Technology Services	17.40	12.000	14.518990
8/12/2023	BRK.B	Berkshire Hathaway Inc. New	Finance	1.43	0.735	1.143973
8/12/2023	V	Visa Inc.	Commercial Services	43.60	18.600	28.104180

PRICE RELATIVE TO 52 WEEK LOW

1. Summary Statistics of Low52WkPerc by Sector

```
SM <- sp500 %>%
  group_by(Sector) %>%
  summarise(
    Mean = mean(na.omit(Low52WkPerc)),
    Median= sd(na.omit(Low52WkPerc)),
    Median= median(na.omit(Low52WkPerc)),
    Q1 = quantile(na.omit(Low52WkPerc), probs = 0.25, na.rm = TRUE),
    Q3 = quantile(na.omit(Low52WkPerc), probs = 0.75, na.rm = TRUE),
    Min = min(na.omit(Low52WkPerc)),
    Max = max(na.omit(Low52WkPerc))
  )

tab <- cbind(Sector = SM$Sector, round(SM[,2:7],2))

tab <- tab %>% arrange(Median)

SM <- knitr::kable(tab, "html") %>% kable_styling()
SM
```

Sector	Mean	Median	Q1	Q3	Min	Max
Communications	6.03	5.23	5.02	6.65	4.80	8.07
Utilities	16.13	13.69	6.99	17.31	1.06	58.39
Consumer Non-Durables	19.43	16.04	9.46	30.20	0.35	59.45
Health Services	27.11	18.76	13.33	35.51	8.30	64.45

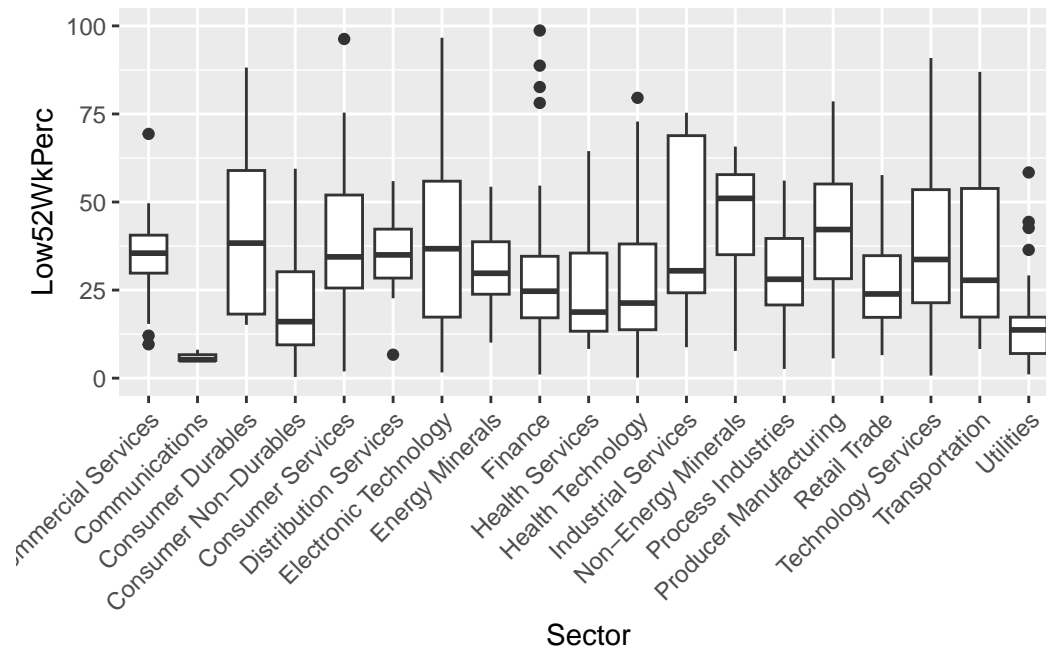
Sector	Mean	Median	Q1	Q3	Min	Max
Health Technology	28.40	22.86	13.80	39.69	0.17	111.43
Retail Trade	26.64	23.90	17.26	34.77	6.53	57.65
Finance	27.26	24.67	17.13	34.60	1.03	98.69
Transportation	36.79	27.79	17.34	53.84	8.28	86.95
Process Industries	28.55	28.07	20.78	39.65	2.62	56.08
Energy Minerals	31.46	29.77	23.82	38.72	10.09	54.33
Industrial Services	41.68	30.48	24.23	68.84	8.78	75.34
Distribution Services	33.82	34.99	28.41	42.30	6.66	55.92
Technology Services	45.16	35.26	21.77	55.13	0.73	255.51
Commercial Services	34.66	35.48	29.81	40.59	9.60	69.34
Consumer Services	50.27	35.97	25.94	58.61	1.91	190.10
Electronic Technology	48.08	38.84	17.85	60.86	1.62	311.70
Consumer Durables	58.17	42.84	18.51	79.84	15.13	150.04
Producer Manufacturing	50.00	43.22	29.05	62.26	5.62	132.17
Non-Energy Minerals	44.31	51.04	35.05	57.78	7.74	65.74

Sector Communications and Utilities are closest to its 52 week low.

2. Box Plot for Low52WkPerc by Sector

```
library(ggplot2)

ggplot(sp500, aes(Sector, Low52WkPerc)) + geom_boxplot() +
  scale_y_continuous(limits=c(0,100)) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



PROFITABILITY BY SECTOR

ROE

1. Summary Statistics of ROE by each Sector of S&P500

```
SectorROE <- sp500 %>%
  group_by(Sector) %>%
  summarise(
    Mean = mean(na.omit(ROE)),
    Median= sd(na.omit(ROE)),
    Median= median(na.omit(ROE)),
    Min = min(na.omit(ROE)),
    Max = max(na.omit(ROE))
  )

tab <- cbind(Sector = SectorROE$Sector, round(SectorROE[,2:5],2))

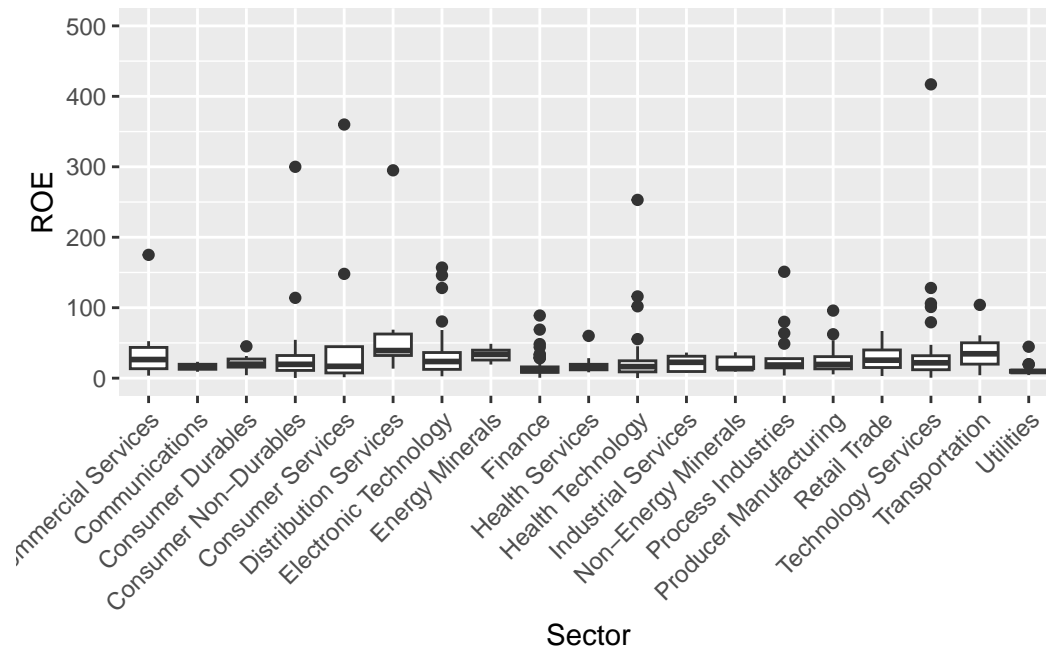
tab <- knitr::kable(tab, "html") %>% kable_styling()
tab
```

Sector	Mean	Median	Min	Max
Commercial Services	37.60	26.40	3.50	175.0
Communications	8.08	9.05	-8.01	23.2
Consumer Durables	13.56	17.75	-51.40	45.2
Consumer Non-Durables	132.99	18.15	-10.80	2880.0
Consumer Services	25.25	10.90	-186.00	360.0
Distribution Services	81.18	39.15	13.50	295.0
Electronic Technology	31.39	18.90	-14.20	157.0
Energy Minerals	90.99	33.80	19.30	954.0
Finance	22.29	10.80	-30.00	714.0
Health Services	20.26	16.10	8.31	60.0
Health Technology	23.04	13.30	-51.00	253.0
Industrial Services	20.71	22.50	7.67	36.3
Non-Energy Minerals	13.64	13.50	-3.83	36.8
Process Industries	27.85	18.00	-11.70	151.0
Producer Manufacturing	23.96	18.00	-13.60	95.9
Retail Trade	30.05	25.50	2.98	66.9
Technology Services	53.07	18.85	-70.60	844.0
Transportation	37.20	34.60	4.13	104.0
Utilities	7.96	8.70	-45.60	44.6

2. Box Plot for ROE by Sector

```
library(ggplot2)

ggplot(sp500, aes(Sector, ROE)) + geom_boxplot() +
  scale_y_continuous(limits=c(0,500)) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



ROA

1. Summary Statistics of ROA by each Sector of S&P500

```
SectorROA <- sp500 %>%
  group_by(Sector) %>%
  summarise(
    Mean = mean(na.omit(ROA)),
    Median= sd(na.omit(ROA)),
    Median= median(na.omit(ROA)),
    Min = min(na.omit(ROA)),
    max = max(na.omit(ROA))
  )

tab <- cbind(Sector = SectorROA$Sector, round(SectorROA[,2:5],2))

tab <- knitr::kable(tab, "html") %>% kable_styling()
tab
```

Sector	Mean	Median	Min	max
Commercial Services	10.10	6.60	1.74	27.20
Communications	2.14	2.93	-2.11	5.61

Sector	Mean	Median	Min	max
Consumer Durables	9.10	9.73	-8.81	27.90
Consumer Non-Durables	7.37	6.81	-5.57	18.90
Consumer Services	5.90	2.83	-10.80	29.00
Distribution Services	10.74	6.49	1.08	24.50
Electronic Technology	9.52	8.75	-6.73	27.60
Energy Minerals	15.90	15.80	7.39	23.70
Finance	3.21	2.38	-2.94	24.70
Health Services	5.75	5.52	2.96	11.00
Health Technology	7.46	7.07	-33.10	28.90
Industrial Services	6.48	5.47	3.30	10.60
Non-Energy Minerals	7.01	4.46	-2.18	21.20
Process Industries	7.24	6.92	-4.31	25.40
Producer Manufacturing	8.93	8.69	-3.10	25.10
Retail Trade	9.99	9.01	0.98	24.20
Technology Services	8.66	8.09	-35.40	41.50
Transportation	8.91	7.20	1.08	26.20
Utilities	2.32	2.55	-6.25	6.49

2. Box Plot for ROA by Sector

```
library(ggplot2)

ggplot(sp500, aes(Sector, ROA)) + geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```

