

Live Case: Overview of the S&P500

Chapter 17, *Last updated: Jan 05, 2024*

S&P 500

The S&P 500, also called the Standard & Poor's 500, is a stock market index that tracks the performance of 500 major publicly traded companies listed on U.S. stock exchanges. It serves as a widely accepted benchmark for assessing the overall health and performance of the U.S. stock market.

S&P Dow Jones Indices, a division of S&P Global, is responsible for maintaining the index. The selection of companies included in the S&P 500 is determined by a committee, considering factors such as market capitalization, liquidity, and industry representation.

The S&P is a float-weighted index, meaning the market capitalizations of the companies in the index are adjusted by the number of shares available for public trading. [1]

The performance of the S&P 500 is frequently used to gauge the broader stock market and is commonly referenced by investors, analysts, and financial media. It provides a snapshot of how large-cap U.S. stocks are faring and is considered a reliable indicator of overall market sentiment.

Typically, the S&P 500 index consists of 500 stocks. However, in reality, there are actually 503 stocks included. This discrepancy arises because three of the listed companies have multiple share classes, and each class is considered a separate stock that needs to be included in the index.

Strengths:

1. **Diverse Representation:** The S&P 500 isn't fixated on a single industry. From technology to healthcare, it offers a panoramic view of various economic sectors, making it an inclusive representation of the U.S. corporate sector.
2. **Benchmark for Investors:** For many fund managers, outperforming the S&P 500 stands as a golden standard. It's a yardstick, establishing it as a critical touchstone for gauging investment success.

3. **Liquidity and Visibility:** Constituent companies enjoy high liquidity and are subject to rigorous screening processes, ensuring that the index represents financially viable entities.

Critiques:

1. **Market Capitalization Weighting:** The index is weighted by market capitalization, meaning companies with higher market values have a more pronounced effect on its performance. Critics argue this approach can skew perceptions, especially during market bubbles when certain sectors are overvalued.
2. **Exclusivity:** Despite its broad purview, 500 companies cannot encapsulate the entire U.S. economy. Many sectors, especially emerging industries or smaller businesses, might not be adequately represented.
3. **Potential for Complacency:** The prominence of the S&P 500 has led many investors to adopt passive investment strategies, tracking the index rather than actively managing portfolios. Detractors argue this might lead to market inefficiencies and reduced capital allocation efficacy.

While the S&P 500 remains an influential and pivotal tool for investors, its dominance prompts a double-edged sword of advantages and critiques. In a constantly evolving economic landscape, understanding both its power and limitations is essential for informed financial decision-making. [2]

S&P 500 Data

Load some useful R packages

```
# Load the required libraries, suppressing annoying startup messages
library(dplyr, quietly = TRUE, warn.conflicts = FALSE) # For data manipulation
library(tibble, quietly = TRUE, warn.conflicts = FALSE) # For data manipulation
library(ggplot2, quietly = TRUE, warn.conflicts = FALSE) # For data visualization
library(ggpubr, quietly = TRUE, warn.conflicts = FALSE) # For data visualization

library(gsheet, quietly = TRUE, warn.conflicts = FALSE) # For Google Sheets
library(rmarkdown, quietly = TRUE, warn.conflicts = FALSE) # For writing
library(knitr, quietly = TRUE, warn.conflicts = FALSE) # For tables
library(kableExtra, quietly = TRUE, warn.conflicts = FALSE) # For tables
```

Read the S&P500 data from a Google Sheet into a tibble

1. We will analyze a real-world, recent dataset containing information about the S&P500 stocks, sourced from TradingView.com. [3]
2. The dataset is located in a Google Sheet and periodically updated.
3. The complete URL of the Google Sheet that has the data is
<https://docs.google.com/spreadsheets/d/14mU1NNpeuV2RouT9MKaAWKUpvjRijzQu40DdWJgyKPQ/>
4. Notice that the Google Sheet ID is: 14mU1NNpeuV2RouT9MKaAWKUpvjRijzQu40DdWJgyKPQ.

S&P Global Industry Classification Standard (GICS®)

In this case study, we will classify and analyze the S&P 500 stocks based on the GICS standard.

The Global Industry Classification Standard (GICS®) was developed in 1999 by S&P Dow Jones Indices and MSCI. The GICS methodology aims to enhance the investment research and asset management process for financial professionals worldwide. The GICS methodology has been widely accepted as an industry analysis framework for investment research, portfolio management and asset allocation. Its universal approach to industries worldwide has contributed to transparency and efficiency in the investment process, and the GICS methodology supports the trend towards sector-based investing. [4]

The GICS classification consists of **11** sectors, **25** industry groups. The 11 sectors are {Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Utilities}. The classification of each stock in the S&P 500 according to GICS is available at the following Google Sheet:

https://docs.google.com/spreadsheets/d/1WrVA8dPYvQsc_mXVctgTntRLS02qd7ubzcdAsw03Lgk/edit#gid=0

Again, we note that the Google Sheet ID is 1WrVA8dPYvQsc_mXVctgTntRLS02qd7ubzcdAsw03Lgk

Loading the data into R

We will connect to both these Google Sheets and read the data into a tibble. We can use the function `gsheet2tbl` in package `gsheet` to read the Google Sheet into a tibble, as demonstrated in the following code.

```
# Read S&P500 stock data present in a Google Sheet.
library(gsheet)
prefix <- "https://docs.google.com/spreadsheets/d/"
sheetID <- "14mU1NNpeuV2RouT9MKaAWKUpvjRijzQu40DdWJgyKPQ"
url500 <- paste(prefix,sheetID) # Form the URL to connect to
sp500Data <- gsheets2tbl(url500) # Read it into a tibble called sp500
```

4. **Reference:** This data has is current as of **5Jan2024**

Similarly, we can also read the GICS classification.

```
# Read GICS classificaiton of S&P 500 stocks from a Google Sheet.
library(gsheet)
prefix2 <- "https://docs.google.com/spreadsheets/d/"
sheetID2 <- "1WrVA8dPYvQsc_mXVctgTntRLS02qd7ubzcdAsw03Lgk"
urlgics <- paste(prefix2, sheetID2) # Form the URL to connect to
gics <- gsheets2tbl(urlgics) # Read it into a tibble called sp500
```

We join the two tibbles, using “Stock” as the key and name our joint tibble “sp500”. For this purpose, we can write the following code.

```
# Merging dataframes
sp500 <- merge(sp500Data,
               gics ,
               id = "Stock")
```

Review the S&P 500 data

1. The data corresponds to 503 companies that are part of the S&P500 and includes 39 data columns.

```
dim(sp500)
```

```
[1] 503  39
```

2. The first ten stocks in the S&P500 data, their GICS Sector and their recent prices are as follows:

```
# summary(sp500$`GICS Sector`)

sp500 %>%
  select(Stock, Description, GICSSector, Date, Price) %>%
  head(10) %>%
  kable("html", caption = "The first 10 companies in the S&P500 dataset") %>%
  kable_styling()
```

Table 0.1: The first 10 companies in the S&P500 dataset

Stock	Description	GICSSector	Date	Price
A	Agilent Technologies, Inc.	Health Care	5Jan2024	131.1500
AAL	American Airlines Group, Inc.	Industrials	5Jan2024	13.3297
AAPL	Apple Inc.	Information Technology	5Jan2024	182.0300
ABBV	AbbVie Inc.	Health Care	5Jan2024	162.1400
ABNB	Airbnb, Inc.	Consumer Discretionary	5Jan2024	135.1750
ABT	Abbott Laboratories	Health Care	5Jan2024	110.4500
ACGL	Arch Capital Group Ltd.	Financials	5Jan2024	77.7459
ACN	Accenture plc	Information Technology	5Jan2024	339.1400
ADBE	Adobe Inc.	Information Technology	5Jan2024	567.6200
ADI	Analog Devices, Inc.	Information Technology	5Jan2024	185.7200

3. Data Columns

- The data comprises of the following 39 columns:

```
colnames(sp500)
```

```
[1] "Stock"
[2] "Date"
[3] "Description"
[4] "Sector"
[5] "Industry"
[6] "Market Capitalization"
[7] "Price"
[8] "52 Week Low"
[9] "52 Week High"
[10] "Return on Equity (TTM)"
[11] "Return on Assets (TTM)"
[12] "Return on Invested Capital (TTM)"
```

```

[13] "Gross Margin (TTM)"
[14] "Operating Margin (TTM)"
[15] "Net Margin (TTM)"
[16] "Price to Earnings Ratio (TTM)"
[17] "Price to Book (FY)"
[18] "Enterprise Value/EBITDA (TTM)"
[19] "EBITDA (TTM)"
[20] "EPS Diluted (TTM)"
[21] "EBITDA (TTM YoY Growth)"
[22] "EBITDA (Quarterly YoY Growth)"
[23] "EPS Diluted (TTM YoY Growth)"
[24] "EPS Diluted (Quarterly YoY Growth)"
[25] "Price to Free Cash Flow (TTM)"
[26] "Free Cash Flow (TTM YoY Growth)"
[27] "Free Cash Flow (Quarterly YoY Growth)"
[28] "Debt to Equity Ratio (MRQ)"
[29] "Current Ratio (MRQ)"
[30] "Quick Ratio (MRQ)"
[31] "Dividend Yield Forward"
[32] "Dividends per share (Annual YoY Growth)"
[33] "Price to Sales (FY)"
[34] "Revenue (TTM YoY Growth)"
[35] "Revenue (Quarterly YoY Growth)"
[36] "Technical Rating"
[37] "Security"
[38] "GICSSector"
[39] "GICSSubIndustry"

```

- The names of the data columns are self-explanatory. The Financial terms are explained in depth on multiple external websites such as www.Investopedia.com

Rename Data Columns

4. The names of the data columns are lengthy and confusing. We will rename the data columns to make it easier to work with the data.

```

# Define a mapping of new column names
new_names <- c(
  "Stock", "Date", "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",
"Security", "GICSSector", "GICSSubIndustry"
)
# Rename the columns using the new_names vector
colnames(sp500)<-new_names

```

5. We review the column names again after renaming them, using the `colnames()` function can help.

```
colnames(sp500)
```

[1] "Stock"	"Date"	"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] "Security"	"GICSSector"	"GICSSubIndustry"

Understand the Data Columns

6. The complete data has 39 columns. Our next goal is to gain a deeper understanding of what the data columns mean. We reorganize the column names into eight tables, labeled Table 1a, 1b.. 1h.
 - a. The column names described in Table 1a. concern basic **Company Information** of each stock.

Table 1a: Data Columns giving basic Company Information	
ColumnName	Description
Stock	Stock Ticker (e.g. AAL)
Date	Date (e.g. "7/15/2023")
StockName	Name of the company (e.g "American Airlines Group, Inc.")
GICSSector	Sector, as per GICS Classification
GICSSubIndustry	Sub-Industry, as per GICS Classification
MarketCap	Market capitalization of the company
Price	Recent Stock Price

- b. The column names described in Table 1b. are related to **Technical Analysis** of each stock, including the 52-Week High and Low prices.

Table 1b: Data Columns related to Pricing and Technical Analysis	
ColumnName	Description
Low52Wk	52-Week Low Price
High52Wk	52-Week High Price
Rating	Technical Rating

- c. The column names described in Table 1c. are related to the **Profitability** of each stock.

Table 1c: Data Columns related to Profitability	
ColumnName	Description
ROE	Return on Equity
ROA	Return on Assets
ROIC	Return on Invested Capital
GrossMargin	Gross Profit Margin
OperatingMargin	Operating Profit Margin
NetMargin	Net Profit Margin

- d. The column names described in Table 1d are related to the **Earnings** of each stock.

Table 1d: Data Columns related to Earnings	
ColumnName	Description
PE	Price-to-Earnings Ratio
PB	Price-to-Book Ratio
EVEBITDA	Enterprise Value to EBITDA Ratio
EBITDA	EBITDA

Table 1d: Data Columns related to Earnings	
ColumnName	Description
EPS	Earnings per Share
EBITDA_YOY	EBITDA Year-over-Year Growth
EBITDA_QYOY	EBITDA Quarterly Year-over-Year Growth
EPS_YOY	EPS Year-over-Year Growth
EPS_QYOY	EPS Quarterly Year-over-Year Growth

- e. The column names described in Table 1e are related to the **Free Cash Flow** of each stock.

Table 1e: Data Columns related to Free Cash Flow	
ColumnName	Description
PFCF	Price-to-Free Cash Flow
FCF	Free Cash Flow
FCF_QYOY	Free Cash Flow Quarterly Year-over-Year Growth

- f. The column names described in Table 1f concern the **Liquidity** of each stock.

Table 1f: Data Columns related to Liquidity	
ColumnName	Description
DebtToEquity	Debt-to-Equity Ratio
CurrentRatio	Current Ratio
QuickRatio	Quick Ratio

- g. The column names described in Table 1g are related to the **Revenue** of each stock.

Table 1g: Data Columns related to Revenue	
ColumnName	Description
PS	Price-to-Sales Ratio
Revenue_YOY	Revenue Year-over-Year Growth
Revenue_QYOY	Revenue Quarterly Year-over-Year Growth

- h. The column names described in Table 1h are related to the **Dividends** of each stock.

Table 1h: Data Columns related to Dividends	
ColumnName	Description
DividendYield	Dividend Yield
DividendsPerShare__YOY	Annual Dividends per Share Year-over-Year Growth

Stock Ratings

1. In the data, the S&P500 shares have Technical Ratings such as {Strong Buy, Buy, Neutral, Sell, Strong Sell}. Since each Stock has a unique Technical Rating, it makes sense to model the data column Rating as a `factor()` variable.

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

2. We confirm that Rating is now modelled as a factor variable, by running the `str()` function. We can use the `levels()` function to review the different levels it can take.

```
str(sp500$Rating)
```

```
Factor w/ 5 levels "Buy","Neutral",...: 1 1 3 4 1 3 3 4 2 3 ...
```

```
levels(sp500$Rating)
```

```
[1] "Buy"          "Neutral"      "Sell"         "Strong Buy"   "Strong Sell"
```

4. The `table()` function allows us to count how many stocks have each Rating.

```
table(sp500$Rating)
```

```
Buy      Neutral      Sell Strong Buy Strong Sell
192       51       178       58       24
```

5. Thus, we can see how many stocks have ratings ranging from “Strong Sell” to “Strong Buy”. This completes our review of Rating.

Analysis of GICS Sectors in the S&P500

- The S&P 500 comprises a wide array of sectors, reflecting the diverse American corporate landscape.
- The data showcases the S&P500 divided across 11 Sectors. The S&P500 shares are divided into multiple Sectors. Each stock belongs to a unique sector. Thus, it makes sense to model Sector as a `factor()` variable.

```
sp500$GICSSector <- as.factor(sp500$GICSSector)
```

2. We confirm that `GICSSector` is now modelled as a factor variable, by running the `str()` function.

```
str(sp500$GICSSector)
```

Factor w/ 11 levels "Communication Services",...: 6 7 8 6 2 6 5 8 8 8 ...

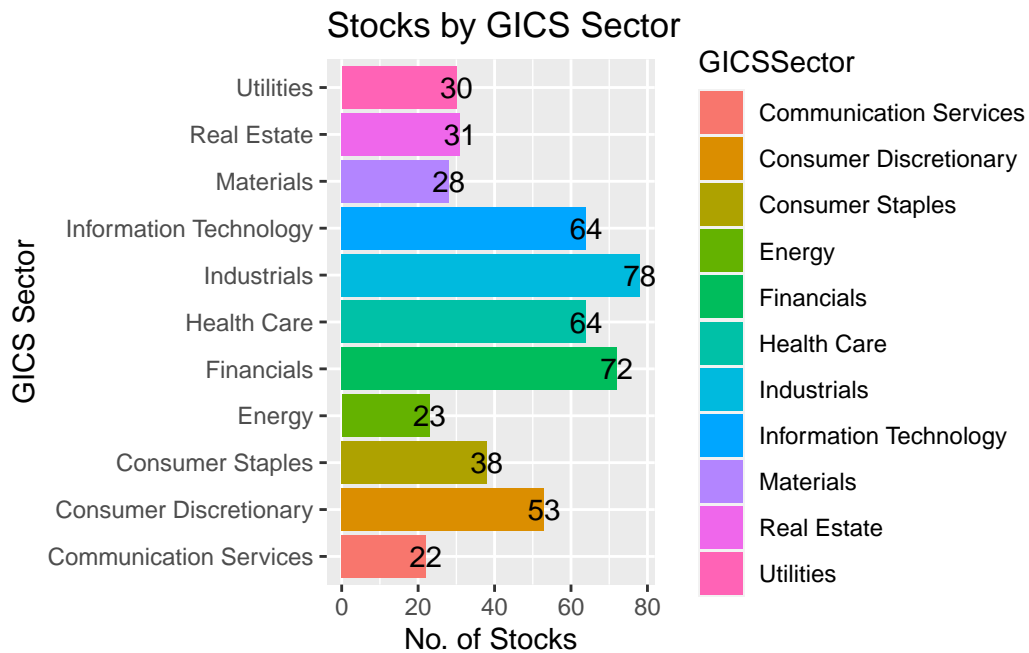
3. We can use the `levels()` function to review the different levels it can take.

```
levels(sp500$GICSSector)
```

```
[1] "Communication Services" "Consumer Discretionary" "Consumer Staples"
[4] "Energy"                "Financials"            "Health Care"
[7] "Industrials"            "Information Technology" "Materials"
[10] "Real Estate"           "Utilities"
```

- The S&P500 consists of 503 stocks, divided across 11 sectors.

```
ggplot(data = sp500,
       aes(y = GICSSector)) +
  geom_bar(aes(fill = GICSSector)) +
  geom_text(stat='count',
           aes(label=after_stat(count))) +
  labs(title = "Stocks by GICS Sector",
       x = "No. of Stocks",
       y = "GICS Sector")
```



- Thus, we can see how many stocks are part of each one of the 11 sectors. We can sum them to confirm that they add up to 503 stocks.

MarketCap by GICS Sector

1. We review the Market Cap of S&P500 stocks across GICS Sectors. We summarize the total Market Cap for each GICS Sector, using the following code.

```
MarketCapbySector <- sp500 %>%
  mutate(Market_Cap_Billions = round(MarketCap/1000000000, 2)) %>%
  group_by(GICSSector) %>%
  summarise(MarketCapBillions = sum(Market_Cap_Billions, na.rm = TRUE)) %>%
  arrange(-MarketCapBillions)
```

2. Next, we create a row, to display the total across all GICS Sectors.

```
# Create a row for the grand total
summary_row <- tibble(
  GICSSector = "Total",
  MarketCapBillions = sum(MarketCapbySector$MarketCapBillions)
)
summary_row
```

```
# A tibble: 1 x 2
  GICSSector MarketCapBillions
  <chr>         <dbl>
1 Total         43733.
```

3. Finally, we append this row to the bottom of the summary table and display it.

```
# Append the summary row to the result
MarketCapbySector <- bind_rows(MarketCapbySector, summary_row)

# Render the table
MarketCapbySector %>%
  kable("html", caption = "Market Capitalization (Billions of USD) of S&P500 by GICS Sector",
  kable_styling())
```

Table 0.10: Market Capitalization (Billions of USD) of S&P500 by GICS Sector

GICSSector	MarketCapBillions
Information Technology	11462.11
Financials	5734.92
Communication Services	5642.80
Health Care	5323.05
Consumer Discretionary	4596.22
Industrials	3598.75
Consumer Staples	2819.64
Energy	1622.53
Real Estate	1009.46
Materials	969.73
Utilities	953.59
Total	43732.80

4. This table gives us the breakdown of the Market Cap of the different stocks in the S&P500 across different GICS Sectors. We note that the Information Technology and Utilities sectors have the relatively highest and lowest Market Cap respectively. We also note that the entire S&P500 has a combined Market Cap of the S&P 500, as of 5Jan2024, is 43732.8 Billion USD.

Summary of Chapter – Exploring S&P500 Data

This chapter explores the S&P500, a significant stock market index encompassing 500 major publicly traded companies in the U.S. The chapter introduces the index's role as a benchmark for assessing the overall health and performance of the U.S. stock market, maintained by S&P Dow Jones Indices.

This chapter guides readers through the intricacies of exploring S&P500 data, employing practical examples and R code to foster a deeper understanding of the breakdown of the S&P 500 across different GICS Sectors. Further exploration is encouraged with a wealth of references for continued learning and analysis.

References

S&P 500

- [1] <https://www.investopedia.com/terms/s/sp500.asp>
- [2] S&P Global: S&P Global. (n.d.). S&P 500. Retrieved September 14, 2023, from <https://www.spglobal.com/spdji/en/indices/equity/sp-500/>
- MarketWatch: MarketWatch. (n.d.). S&P 500 Index. Retrieved September 14, 2023, from <https://www.marketwatch.com/investing/index/spx>
- Bloomberg: Bloomberg. (n.d.). S&P 500 Index (SPX:IND). Retrieved September 14, 2023, from <https://www.bloomberg.com/quote/SPX:IND>
- [3] TradingView.com <https://www.tradingview.com/screener/>
- [4] GICS: Global Industry Classification Standard: <https://www.spglobal.com/spdji/en/landing/topic/gics/>