

Live Case: S&P500 (1)

Sep 30, 2023

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. <https://www.investopedia.com/terms/s/sp500.asp>

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.

References:

1. **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/>
2. **MarketWatch:** MarketWatch. (n.d.). S&P 500 Index. Retrieved September 14, 2023, from <https://www.marketwatch.com/investing/index/spx>
3. **Bloomberg:** Bloomberg. (n.d.). S&P 500 Index (SPX:IND). Retrieved September 14, 2023, from <https://www.bloomberg.com/quote/SPX:IND>

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. The dataset is located in a Google Sheet and periodically updated.
2. The complete URL of the Google Sheet that has the data is
<https://docs.google.com/spreadsheets/d/11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM/>
3. The Google Sheet ID is: 11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM. 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 <- "11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM"
url500 <- paste(prefix,sheetID) # Form the URL to connect to
sp500 <- gsheet2tbl(url500) # Read it into a tibble called sp500
```

4. **Reference:** This data has been sourced from websites like Yahoo Finance and TradingView.com and it is current as of **9/30/2023**

Review the S&P 500 data

1. The data corresponds to 503 companies that are part of the S&P500 and includes 36 data columns.

```
dim(sp500)
```

```
[1] 503  36
```

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

```
sp500 %>%
  select(Stock, Description, Sector, 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	Sector	Date	Price
A	Agilent Technologies, Inc.	Health Technology	9/30/2023	111.8
AAL	American Airlines Group, Inc.	Transportation	9/30/2023	12.8
AAPL	Apple Inc.	Electronic Technology	9/30/2023	171.2
ABBV	AbbVie Inc.	Health Technology	9/30/2023	149.1
ABNB	Airbnb, Inc.	Consumer Services	9/30/2023	137.3
ABT	Abbott Laboratories	Health Technology	9/30/2023	96.8
ACGL	Arch Capital Group Ltd.	Finance	9/30/2023	79.7
ACN	Accenture plc	Technology Services	9/30/2023	307.3
ADBE	Adobe Inc.	Technology Services	9/30/2023	510.0
ADI	Analog Devices, Inc.	Electronic Technology	9/30/2023	175.1

3. Data Columns

- The data comprises of the following 36 columns:

```
colnames(sp500)
```

```
[1] "Date"
[2] "Stock"
[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"

```

- 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(
  "Date", "Stock", "StockName", "Sector", "Industry",
  "MarketCap", "Price", "Low52Wk", "High52Wk",
  "ROE", "ROA", "ROIC", "GrossMargin",
  "OperatingMargin", "NetMargin", "PE",
  "PB", "EBITDA", "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
colnames(sp500)<-new_names

```

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

```
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"

```

Understand the Data Columns

6. The complete data has 36 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
Date	Date (e.g. "7/15/2023")
Stock	Stock Ticker (e.g. AAL)

Table 1a: Data Columns giving basic Company Information	
ColumnName	Description
StockName	Name of the company (e.g "American Airlines Group, Inc.")
Sector	Sector the stock belongs to (e.g. "Transportation")
Industry	Industry the stock belongs to (e.g "Airlines")
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
EPS	Earnings per Share

Table 1d: Data Columns related to Earnings	
ColumnName	Description
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

Table 1h: Data Columns related to Dividends	
ColumnName	Description
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.

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

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

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

```
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
131         62      255         21         34
```

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

Sectors within 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 19 Sectors. Here's an overview of the sectors, along with some representative companies:
1. **Commercial Services (13 companies)**: This sector comprises firms offering services primarily to businesses. Example: Accenture, a management consulting and professional services firm.
 2. **Communications (3 companies)**: This sector covers companies involved in telecommunication and media. Example: AT&T, a telecommunications giant.
 3. **Consumer Durables (12 companies)**: Companies producing goods that aren't bought frequently and have a longer life, such as appliances and cars. Example: Whirlpool, an appliance manufacturer.
 4. **Consumer Non-Durables (32 companies)**: These are entities that produce goods that are consumed quickly or are non-reusable. Example: The Coca-Cola Company, known globally for its beverages.
 5. **Consumer Services (29 companies)**: This includes companies that offer services directly to consumers. Example: Walt Disney Company, known for its entertainment services.
 6. **Distribution Services (9 companies)**: Companies engaged in the distribution of products. Example: Sysco, a major food distributor.
 7. **Electronic Technology (49 companies)**: Represents companies engaged in electronics research and product development. Example: Apple, known for its iPhone and other electronic products.
 8. **Energy Minerals (16 companies)**: Companies that extract and produce energy resources. Example: ExxonMobil, one of the world's largest oil and gas companies.
 9. **Finance (92 companies)**: Covers banks, insurance companies, investment firms, RE-ITs and other financial institutions. Example: JPMorgan Chase, a leading global financial services firm.
 10. **Health Services (12 companies)**: Includes providers of health services like hospital management. Example: HCA Healthcare, a major hospital operator.
 11. **Health Technology (47 companies)**: Companies that produce medical equipment, pharmaceuticals, and other health-related tech. Example: Pfizer, a global pharmaceutical company.

12. **Industrial Services (9 companies)**: Firms providing services primarily to the industrial sector. Example: Caterpillar, a leading manufacturer of construction and mining equipment.
13. **Non-Energy Minerals (7 companies)**: Companies involved in the extraction of minerals other than energy resources. Example: Vulcan Materials, a producer of construction aggregates.
14. **Process Industries (24 companies)**: This sector comprises entities that transform raw materials into finished goods. Example: Dow Inc., a major chemical manufacturer.
15. **Producer Manufacturing (31 companies)**: Companies engaged in the manufacturing of products for industries and consumers. Example: 3M, known for its diversified manufacturing.
16. **Retail Trade (22 companies)**: Companies directly involved in selling products to consumers. Example: Walmart, the multinational retail corporation.
17. **Technology Services (50 companies)**: Companies that offer tech services, including IT consulting and software services. Example: Microsoft, a global leader in software and cloud services.
18. **Transportation (15 companies)**: Covers companies involved in the transportation of goods and people. Example: Delta Air Lines, a major U.S. airline.
19. **Utilities (31 companies)**: Companies providing essential services like electricity, gas, and water. Example: Duke Energy, a prominent utility company.

Analysis of S&P 500 Sectors

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

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

```
str(sp500$Sector)
```

```
Factor w/ 19 levels "Commercial Services",...: 11 18 7 11 5 11 9 17 17 7 ...
```

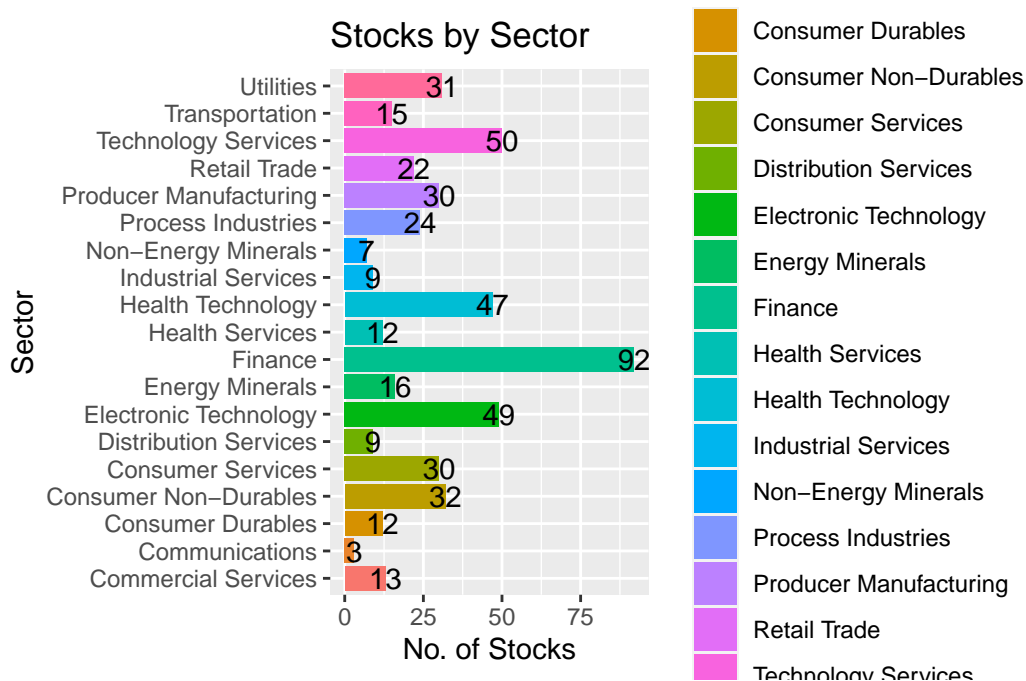
3. We can use the levels() function to review the different levels it can take.

```
levels(sp500$Sector)
```

```
[1] "Commercial Services"      "Communications"          "Consumer Durables"
[4] "Consumer Non-Durables"    "Consumer Services"        "Distribution Services"
[7] "Electronic Technology"     "Energy Minerals"          "Finance"
[10] "Health Services"          "Health Technology"        "Industrial Services"
[13] "Non-Energy Minerals"      "Process Industries"       "Producer Manufacturing"
[16] "Retail Trade"             "Technology Services"      "Transportation"
[19] "Utilities"
```

- The S&P500 consists of 503 stocks, divided across 19 sectors.

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



- Thus, we can see how many stocks are part of each one of the 19 sectors. We can sum them to confirm that they add up to 503 stocks.

MarketCap by Sector

1. 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.

We review the Market Cap of S&P500 stocks across Sectors.

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

# Create a summary row
summary_row <- tibble(
  Sector = "Total",
  MarketCapBillions = sum(MarketCapbySector$Market_Cap_Billions)
)
```

Warning: Unknown or uninitialised column: `Market_Cap_Billions`.

```
# 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) by S&P500 Sector") %>%
  kable_styling()
```

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

Sector	MarketCapBillions
Technology Services	9253.42
Electronic Technology	6354.28
Finance	4746.30
Health Technology	3752.43
Retail Trade	2981.37

Sector	MarketCapBillions
Consumer Non-Durables	1975.61
Energy Minerals	1455.78
Consumer Services	1411.99
Producer Manufacturing	1304.14
Commercial Services	1235.92
Consumer Durables	1049.33
Health Services	921.87
Utilities	890.46
Process Industries	746.42
Transportation	621.70
Communications	408.39
Industrial Services	403.77
Distribution Services	303.21
Non-Energy Minerals	197.15
Total	0.00

Summary of Chapter – Exploring S&P500 Data

This chapter embarks on an exploration of 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 skillfully guides readers through the intricacies of exploring S&P500 data, employing practical examples and R code to foster a deeper understanding of the dataset's structure and content. Further exploration is encouraged with a wealth of references for continued learning and analysis.