

Preliminary Data Visualizations

Last Updated: 2025-03-01

What is This?

The stocks used in this file were drawn from the [Huge Stock Market Dataset](#) we'll be using in this project. The purpose of this file is to examine and understand what the data sets we are using look like (i.e. variables available, time periods covered, data distributions).

We will be looking at both volatile (Solana) and nonvolatile (Apple, Microsoft) stocks.

To see the code used to compile this PDF, see the corresponding R Markdown file.

How to Run This File

For the R Markdown file, it is recommended to be opened using RStudio with the latest version of R.

This file uses the `tidyverse` package in R, which needs to be installed in the Console using the command `install.packages("tidyverse")` prior to running the following code that loads the package into this session.

```
library(tidyverse)
```

This file also uses the `quantmod` and `broom` package in R for plotting the S&P500 index. Again, this needs to be installed in the Console using the command `install.packages("quantmod")` prior to running the following code that loads the package into this session.

```
library(quantmod)
library(broom)
```

To load the data, make sure to have the following necessary `txt` files containing the data within the same folder you have this R Markdown file in:

- `aapl.us.txt`
- `msft.us.txt`
- `sol.us.txt`

To see any documentation about the built-in functions used, you can use the `help()` command in the Console in RStudio.

Notes About the Data

Each dataframe (i.e. each `txt` file) contains information about a different stock with the following 7 variables:

- **Date:** The day that row's data corresponds to. It is formatted as year-mm-dd.
- **Open:** The open price of the stock that day. This is the price at which the stock first trades at when the stock market opens for the day.
- **High:** The highest price the stock was sold for that day.
- **Low:** The lowest price that stock was sold for that day.
- **Close:** The closing price of the stock that day. This is the price that the stock was last recorded to be sold for before the stock market closes for the day.
- **Volume:** The total number of shares of the particular stock traded that day.
- **OpenInt:** ???

AAPL Stock Data Set

The following data is for Apple (APPL) stocks.

The following are the first 5 rows of the `apple` data set:

##	Date	Open	High	Low	Close	Volume	OpenInt
## 1	1984-09-07	0.42388	0.42902	0.41874	0.42388	23220030	0
## 2	1984-09-10	0.42388	0.42516	0.41366	0.42134	18022532	0
## 3	1984-09-11	0.42516	0.43668	0.42516	0.42902	42498199	0
## 4	1984-09-12	0.42902	0.43157	0.41618	0.41618	37125801	0
## 5	1984-09-13	0.43927	0.44052	0.43927	0.43927	57822062	0

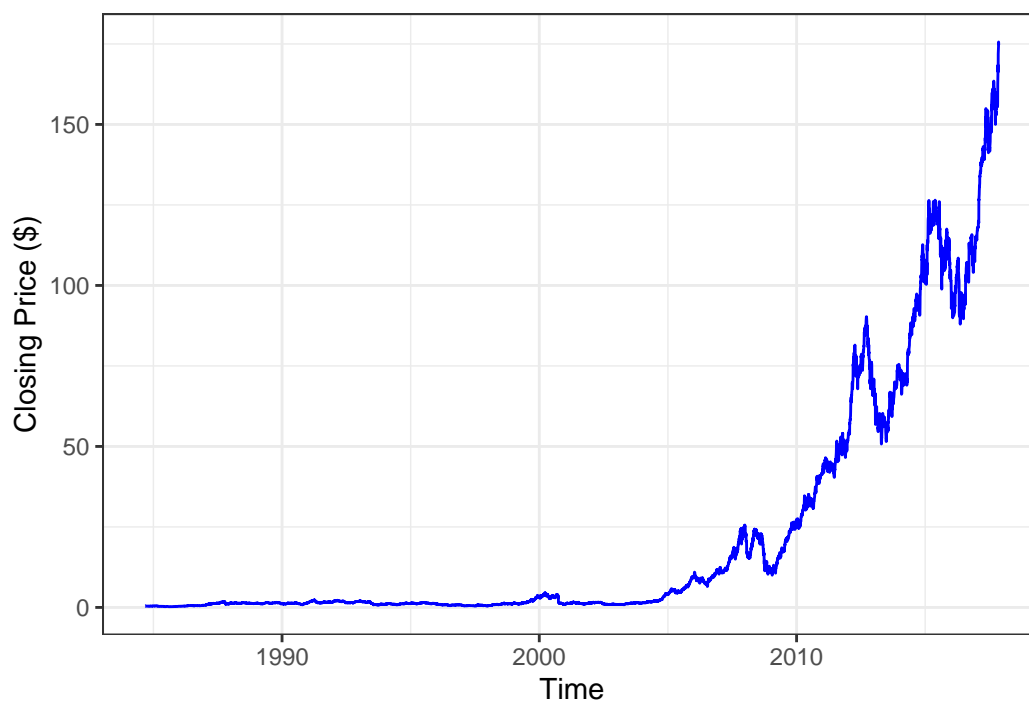
The following are the last 5 rows of the `apple` data set:

##	Date	Open	High	Low	Close	Volume	OpenInt
## 8360	2017-11-06	171.75	174.36	171.10	173.63	34901241	0
## 8361	2017-11-07	173.29	174.51	173.29	174.18	24424877	0
## 8362	2017-11-08	174.03	175.61	173.71	175.61	24451166	0
## 8363	2017-11-09	174.48	175.46	172.52	175.25	29533086	0
## 8364	2017-11-10	175.11	175.38	174.27	174.67	25130494	0

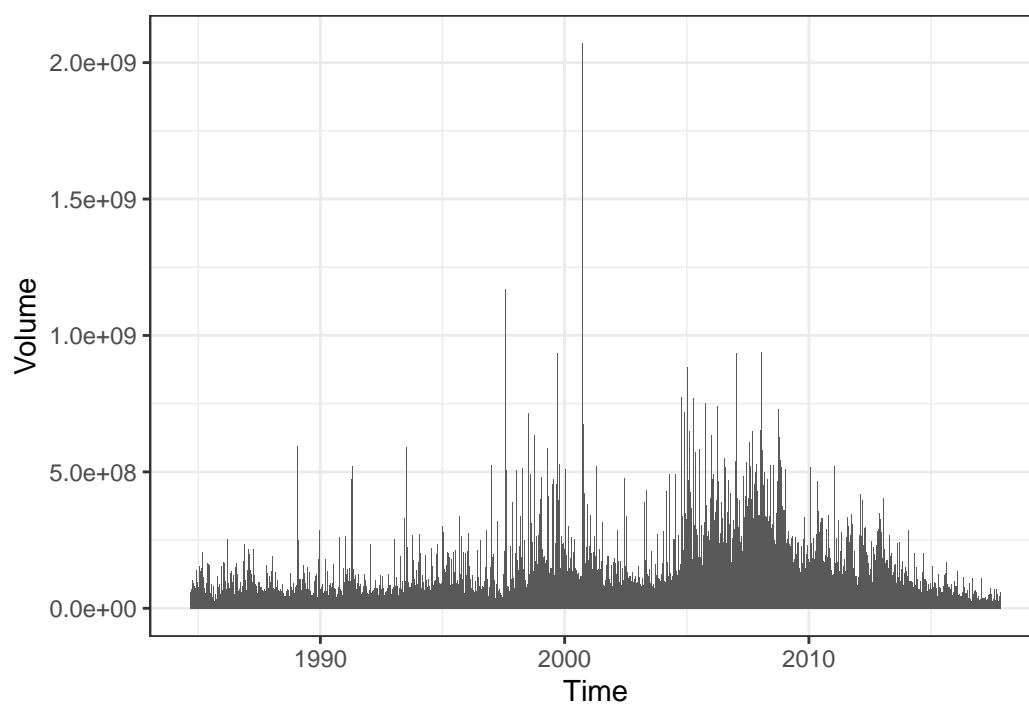
The following are some summary statistics of the `apple` data set:

##	Date	Open	High	Low
##	Length:8364	Min. : 0.2331	Min. : 0.2356	Min. : 0.2305
##	Class :character	1st Qu.: 1.1371	1st Qu.: 1.1642	1st Qu.: 1.1128
##	Mode :character	Median : 1.6328	Median : 1.6634	Median : 1.6006
##		Mean : 22.2843	Mean : 22.4959	Mean : 22.0542
##		3rd Qu.: 23.7390	3rd Qu.: 23.9305	3rd Qu.: 23.3358
##		Max. :175.1100	Max. :175.6100	Max. :174.2700
##	Close	Volume	OpenInt	
##	Min. : 0.2305	Min. :0.000e+00	Min. :0	
##	1st Qu.: 1.1371	1st Qu.:4.384e+07	1st Qu.:0	
##	Median : 1.6282	Median :7.481e+07	Median :0	
##	Mean : 22.2810	Mean :1.066e+08	Mean :0	
##	3rd Qu.: 23.6945	3rd Qu.:1.321e+08	3rd Qu.:0	
##	Max. :175.6100	Max. :2.070e+09	Max. :0	

Closing Prices for AAPL



Volume of AAPL Stocks



MSFT Stock Data Set

The following data is for Microsoft (MSFT) stocks.

The following are the first 5 rows of the `msft` data set:

##	Date	Open	High	Low	Close	Volume	OpenInt
## 1	1986-03-13	0.06720	0.07533	0.06720	0.07533	1371330506	0
## 2	1986-03-14	0.07533	0.07533	0.07533	0.07533	409569463	0
## 3	1986-03-17	0.07533	0.07533	0.07533	0.07533	176995245	0
## 4	1986-03-18	0.07533	0.07533	0.07533	0.07533	90067008	0
## 5	1986-03-19	0.07533	0.07533	0.07533	0.07533	63655515	0

The following are the last 5 rows of the `msft` data set:

##	Date	Open	High	Low	Close	Volume	OpenInt
## 8360	2017-11-06	171.75	174.36	171.10	173.63	34901241	0
## 8361	2017-11-07	173.29	174.51	173.29	174.18	24424877	0
## 8362	2017-11-08	174.03	175.61	173.71	175.61	24451166	0
## 8363	2017-11-09	174.48	175.46	172.52	175.25	29533086	0
## 8364	2017-11-10	175.11	175.38	174.27	174.67	25130494	0

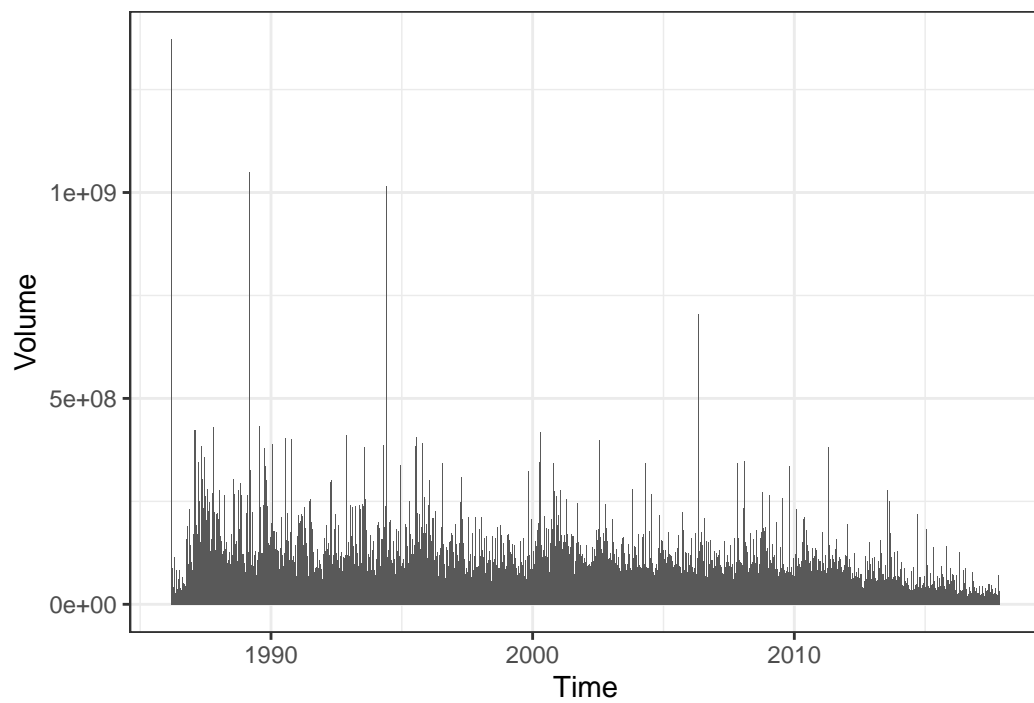
The following are some summary statistics of the `msft` data set:

##	Date	Open	High	Low
##	Length:7983	Min. : 0.0672	Min. : 0.0672	Min. : 0.0672
##	Class :character	1st Qu.: 2.1133	1st Qu.: 2.1551	1st Qu.: 2.0797
##	Mode :character	Median :20.4620	Median :20.6970	Median :20.2510
##		Mean :18.9779	Mean :19.1872	Mean :18.7736
##		3rd Qu.:25.4750	3rd Qu.:25.7540	3rd Qu.:25.2330
##		Max. :84.7700	Max. :86.2000	Max. :84.0825
##	Close	Volume	OpenInt	
##	Min. : 0.0672	Min. :0.000e+00	Min. :0	
##	1st Qu.: 2.1133	1st Qu.:4.911e+07	1st Qu.:0	
##	Median :20.4690	Median :7.059e+07	Median :0	
##	Mean :18.9847	Mean :7.946e+07	Mean :0	
##	3rd Qu.:25.4835	3rd Qu.:9.735e+07	3rd Qu.:0	
##	Max. :84.5600	Max. :1.371e+09	Max. :0	

Closing Prices for MSFT



Volume of MSFT Stocks



SOL Stock Data Set

The following data is for Solana (SOL) stocks. Solana is a cryptocurrency and blockchain platform.

The following are the first 5 rows of the sol data set:

```
##      Date  Open  High  Low Close  Volume  OpenInt
## 1 2008-01-29 65.40 68.65 62.0 64.95 1316110      0
## 2 2008-01-30 66.15 66.25 61.3 65.20  321948      0
## 3 2008-01-31 63.70 63.75 61.5 62.55   59761      0
## 4 2008-02-01 61.10 66.50 60.6 64.50  117039      0
## 5 2008-02-04 65.05 70.95 64.0 68.25  307036      0
```

The following are the last 5 rows of the sol data set:

```
##      Date  Open  High  Low  Close  Volume  OpenInt
## 2462 2017-11-06 2.72 2.7200 2.57 2.6000  19274      0
## 2463 2017-11-07 2.60 2.6501 2.59 2.5999  45012      0
## 2464 2017-11-08 2.67 2.8600 2.64 2.8400  89054      0
## 2465 2017-11-09 2.89 2.9000 2.80 2.8300  38238      0
## 2466 2017-11-10 2.82 2.9000 2.79 2.8800  44936      0
```

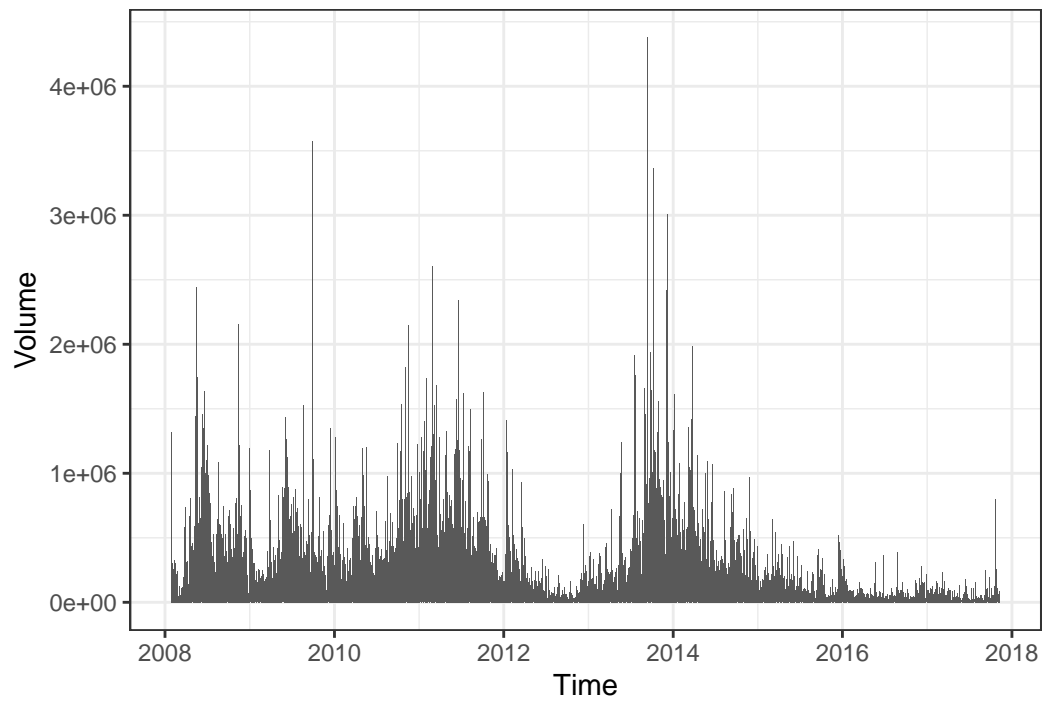
The following are some summary statistics of the sol data set:

```
##      Date      Open      High      Low
## Length:2466    Min.    :  2.18    Min.    :  2.23    Min.    :  2.12
## Class :character 1st Qu.:  6.90    1st Qu.:  7.05    1st Qu.:  6.65
## Mode  :character Median : 12.57    Median : 13.00    Median : 12.05
##              Mean  : 20.74    Mean  : 21.46    Mean  : 19.92
##              3rd Qu.: 25.84    3rd Qu.: 26.75    3rd Qu.: 25.19
##              Max.   :137.90    Max.   :147.40    Max.   :132.00
##      Close      Volume      OpenInt
## Min.    :  2.20    Min.    :  6601    Min.    :0
## 1st Qu.:  6.85    1st Qu.:  94148    1st Qu.:0
## Median : 12.50    Median : 236221    Median :0
## Mean  : 20.64    Mean  : 349144    Mean  :0
## 3rd Qu.: 25.94    3rd Qu.: 474693    3rd Qu.:0
## Max.   :139.00    Max.   :4377122    Max.   :0
```

Closing Prices for SOL



Volume of SOL Stocks

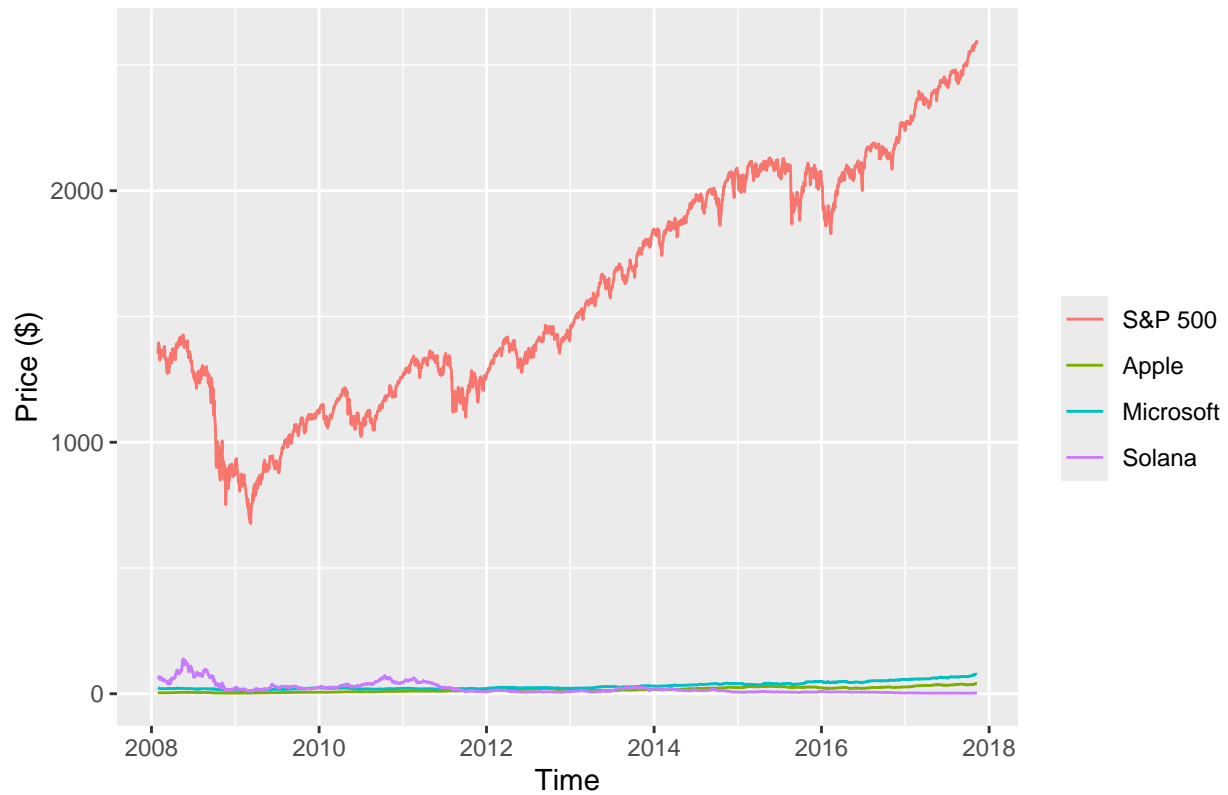


S&P 500 Index

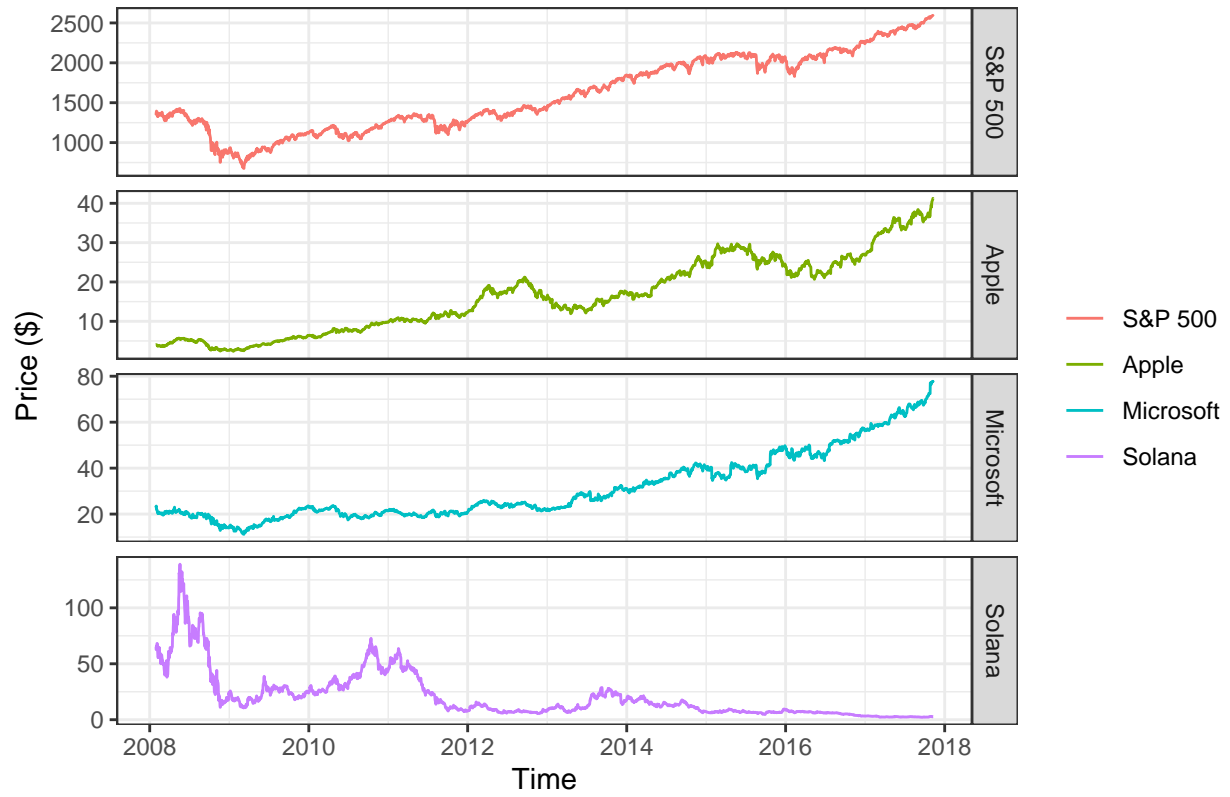
Note that in this section, we will be drawing data from Yahoo Finance instead of the `txt` files from before.

The time range covered here will be from 2008-01-29 to 2017-11-10 since this period is covered in all three of our data sets of interest from before.

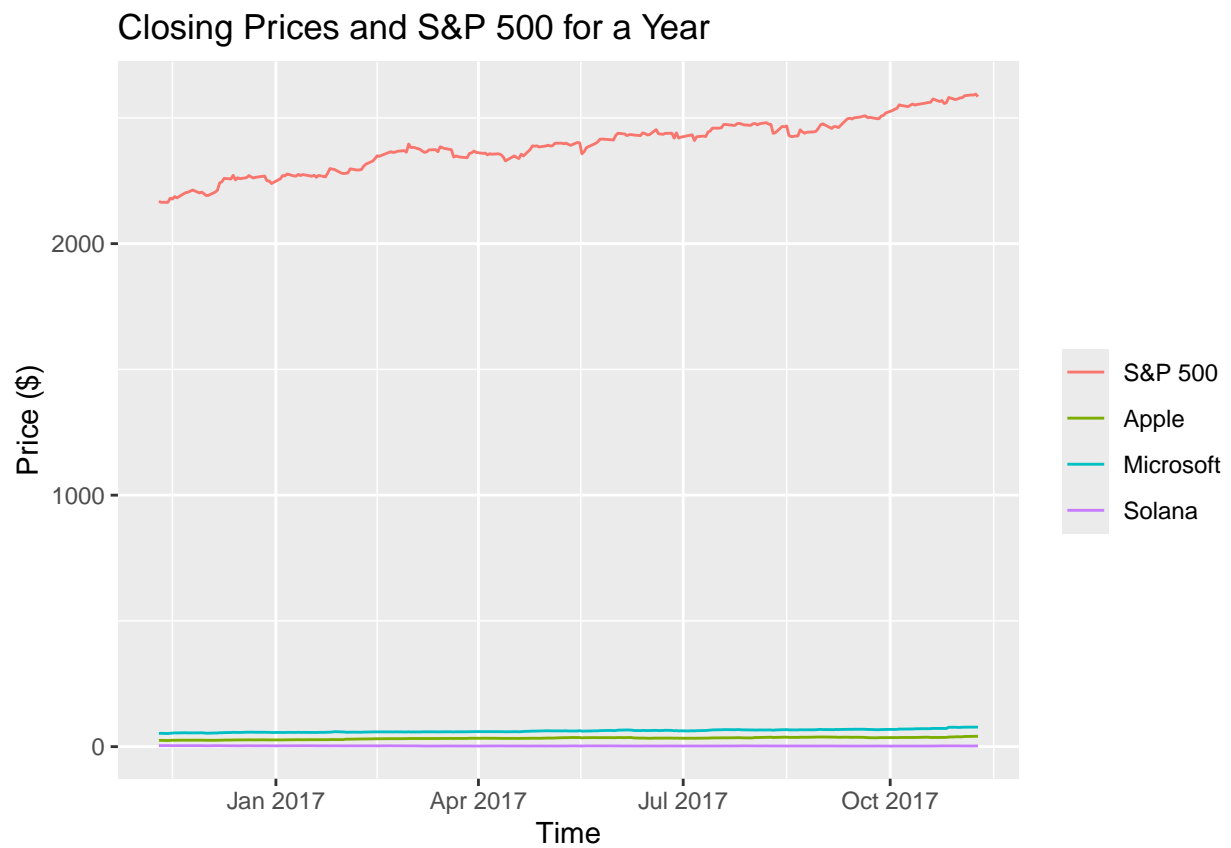
Closing Prices and S&P 500



Closing Prices and S&P 500



The following plots are observing only 2016-11-10 to 2017-11-10 (i.e. the last year of the previous plot).



Closing Prices and S&P 500 for a Year

