Lydia Teinfalt
02/17/2022

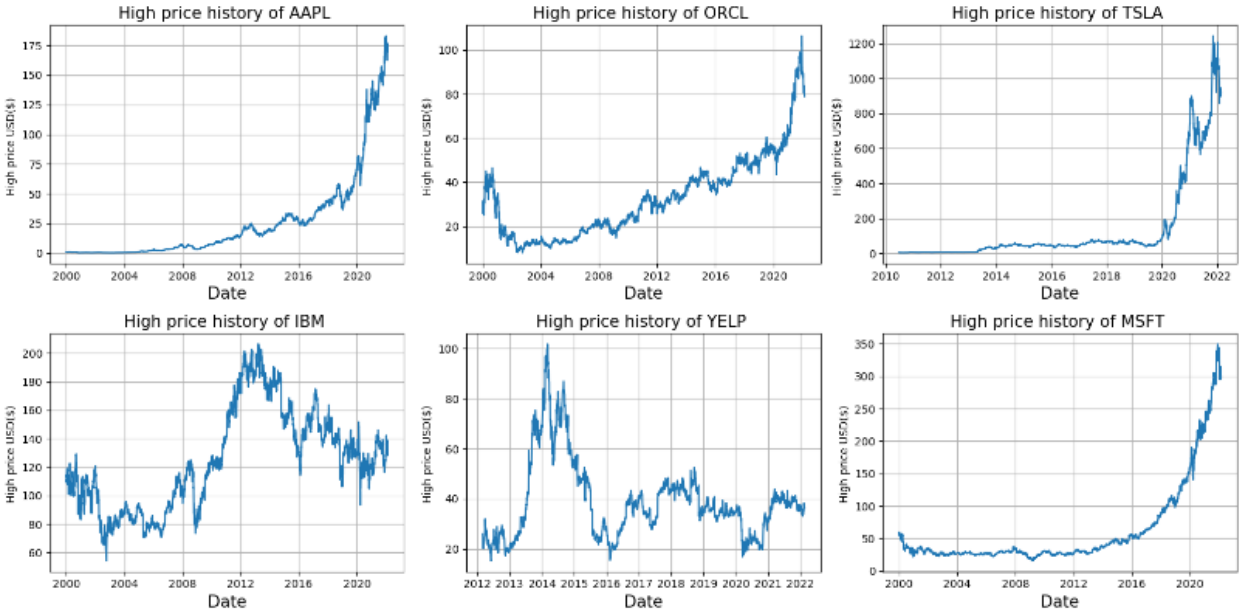
Homework 2

1. Using the pandas_datareader package connect to yahoo database and load the stock value for the following giant companies. Pick the start date as '2000-01-01' and the end date " today date" (2/17/2022)

```
stocks = ['AAPL','ORCL', 'TSLA', 'IBM','YELP', 'MSFT']
```

Date	High	Low	Open	...	Volume	Adj Close	Symbol
				...			
2000-01-03	1.004464	0.907924	0.936384	...	535796800.0	0.855796	AAPL
2000-01-04	0.987723	0.903460	0.966518	...	512377600.0	0.783644	AAPL
2000-01-05	0.987165	0.919643	0.926339	...	778321600.0	0.795112	AAPL
2000-01-06	0.955357	0.848214	0.947545	...	767972800.0	0.726304	AAPL
2000-01-07	0.901786	0.852679	0.861607	...	460734400.0	0.760708	AAPL

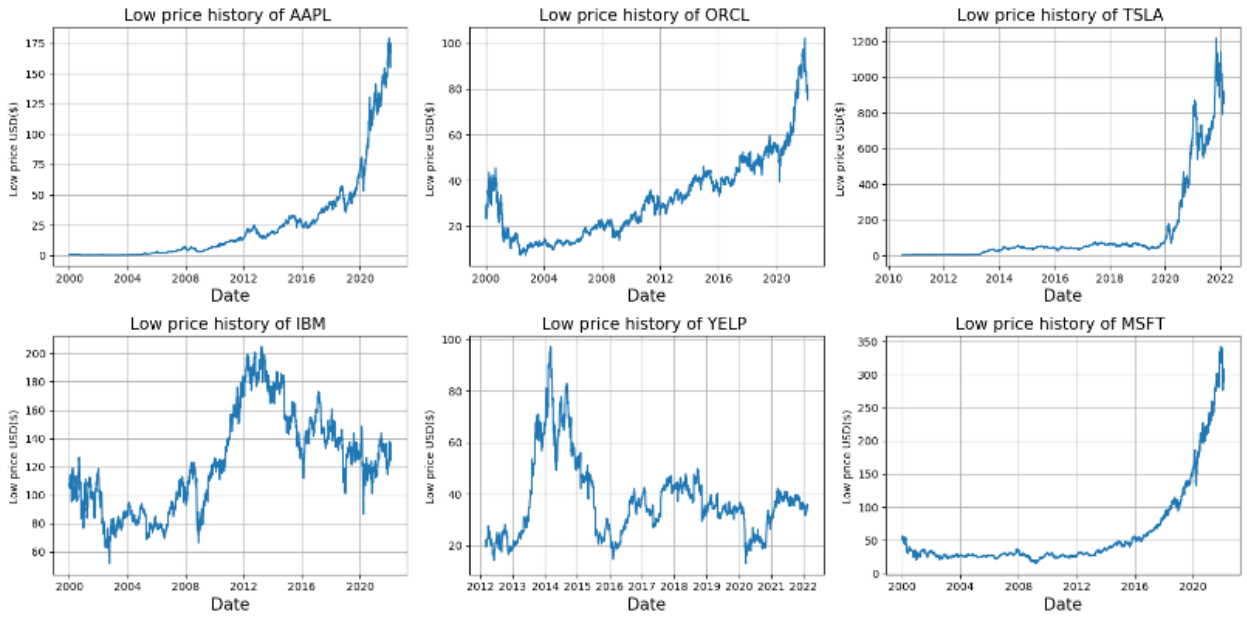
2. The database contains 6 features: "High", "Low", "Open", "Close", "Volume", "Adj Close" in USD($). Using the matplotlib.pyplot package and subplot command, plot the "High" columns for all companies in one figure with 3 rows and 2 columns graph. Make sure to add title, legend, x-label. y-label and grid to your plot. The plot should look like the following. Fig size = (16,8)



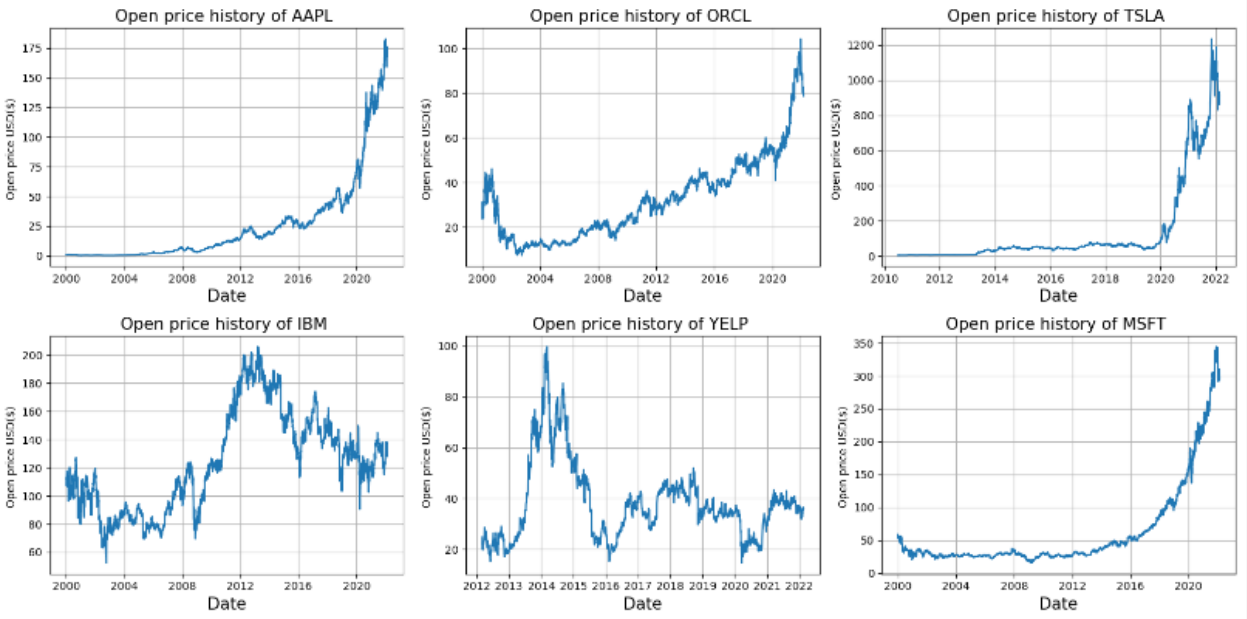
3. Repeat question 2 for, "Low", "Open", "Close", "Volume", "Adj Close".

Lydia Teinfalt
02/17/2022

Low Price ($USD)

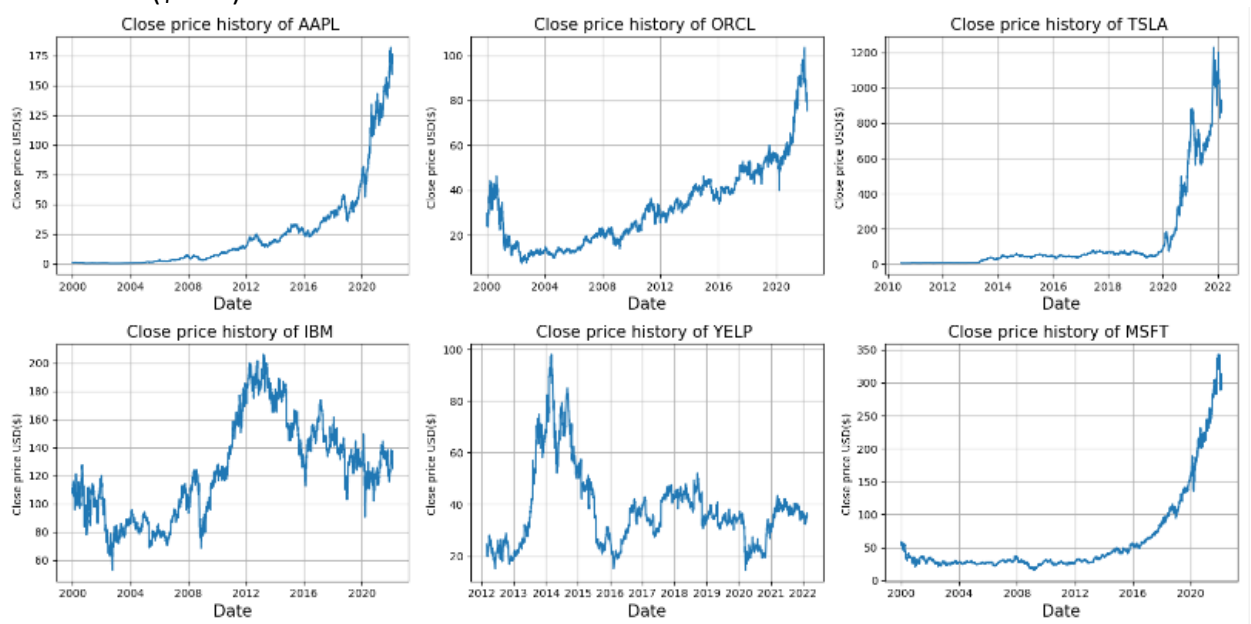


Open Price ($USD)

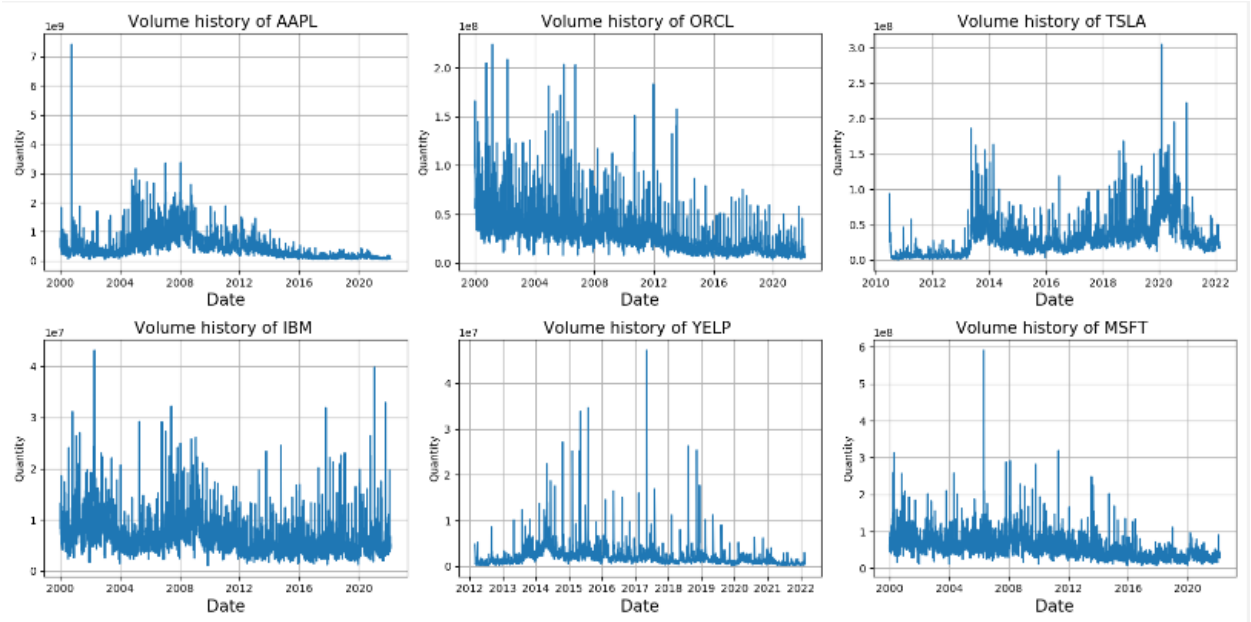


Lydia Teinfalt
02/17/2022

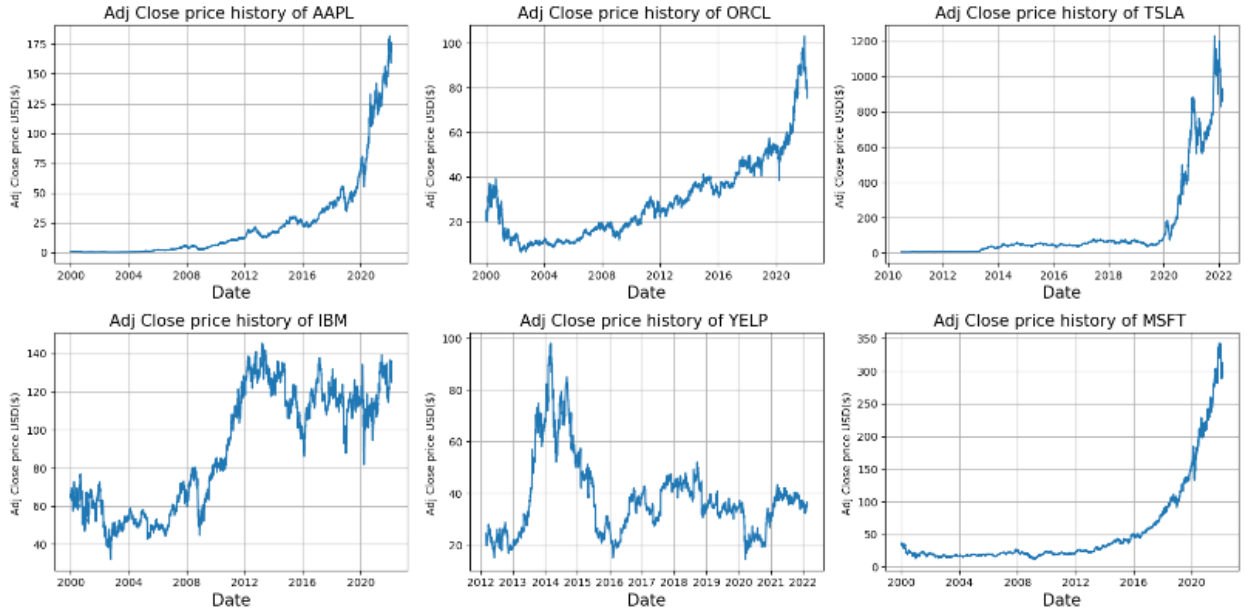
Close Price ($USD)



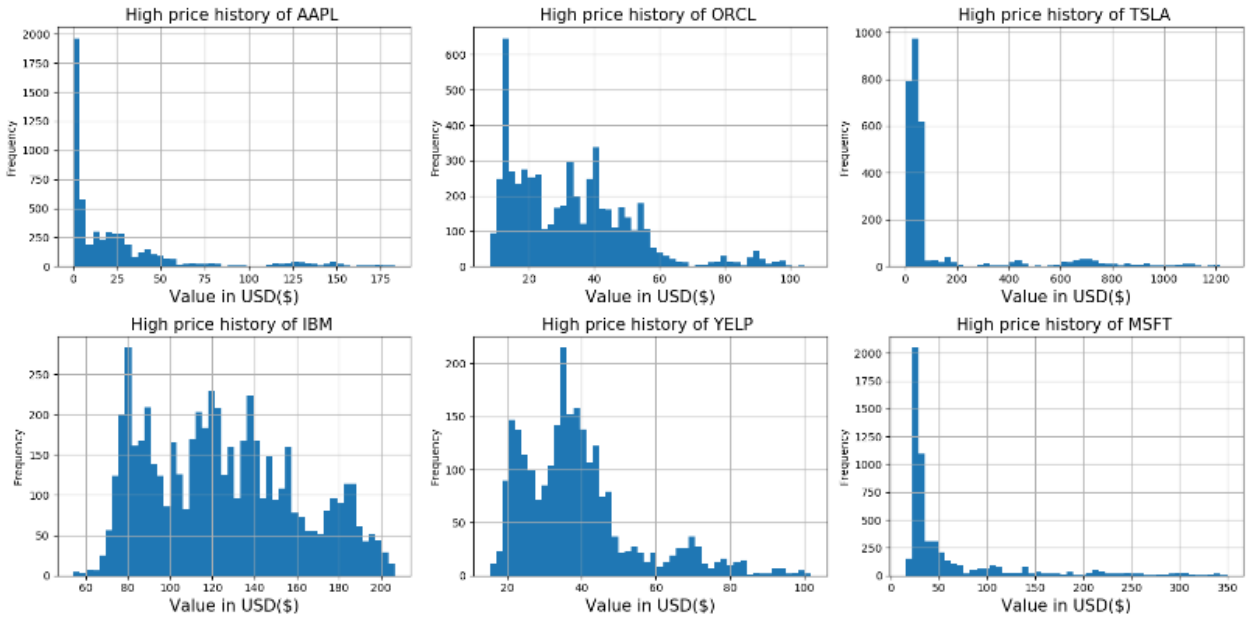
Volume



Adj Close Price ($USD)

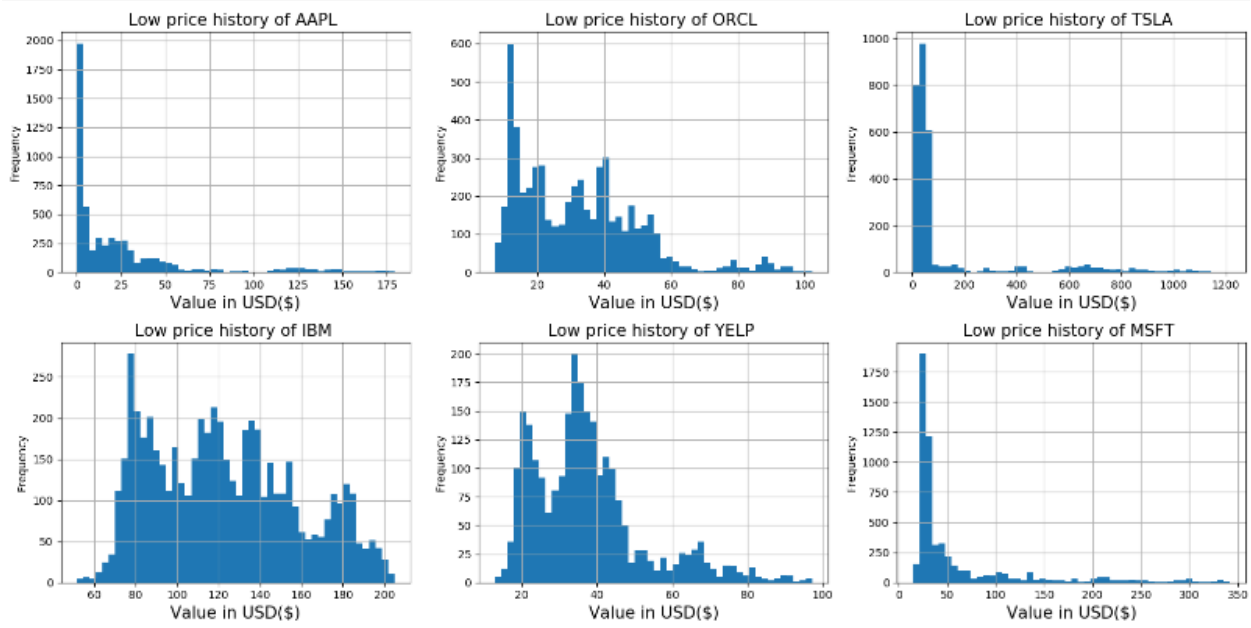


4. Using the matplotlib.pyplot package and hist command, plot the histogram plot of the “High” columns for all companies in a 3x2 graph. Make sure to add title, legend, x-label. y-label and grid to your plot. The final plot should look like the following. # of bins = 50.

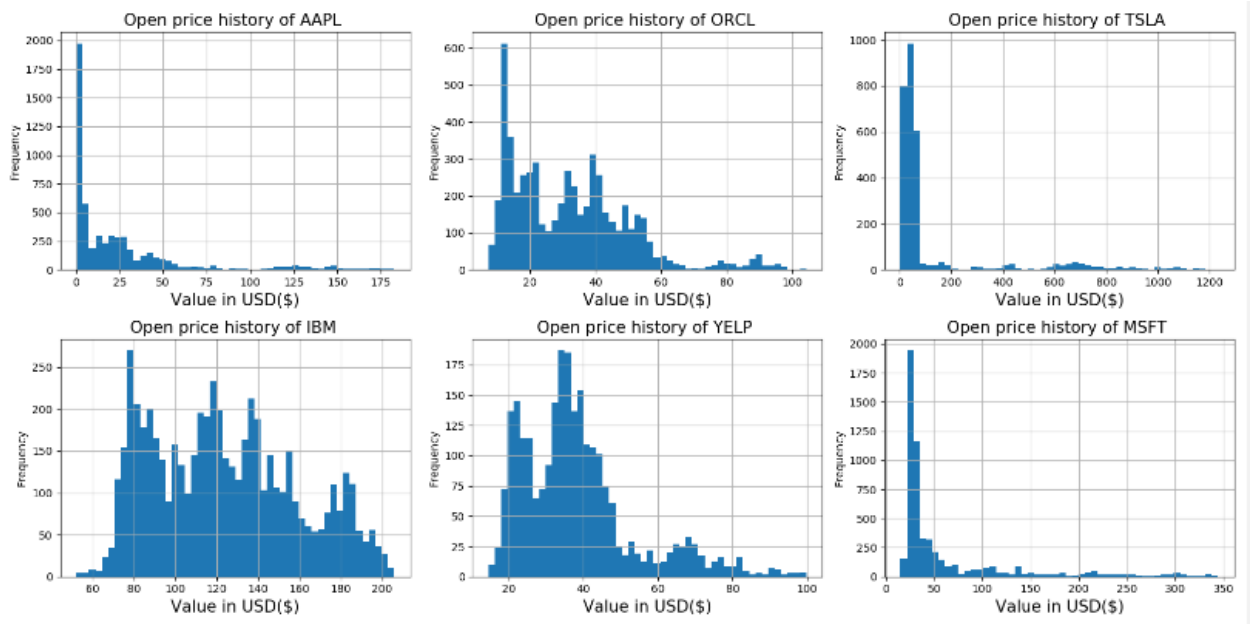


5. Repeat question 4 for, “Low”, “Open”, “Close”, “Volume”, “Adj Close”.

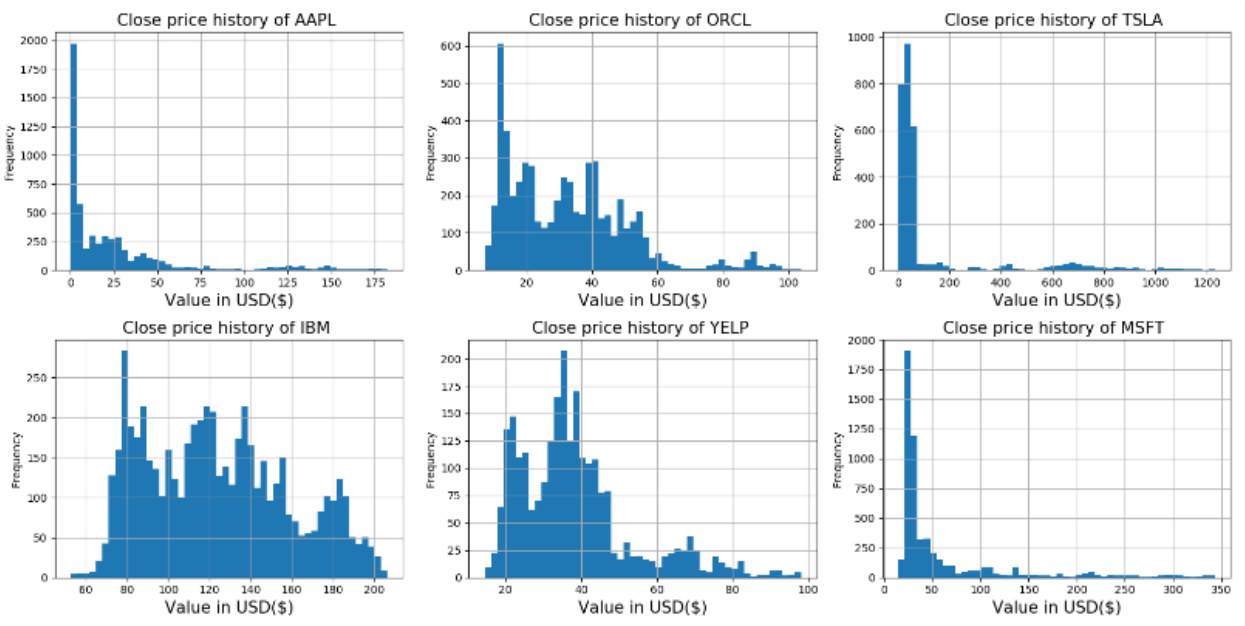
Low Price ($USD) Histogram



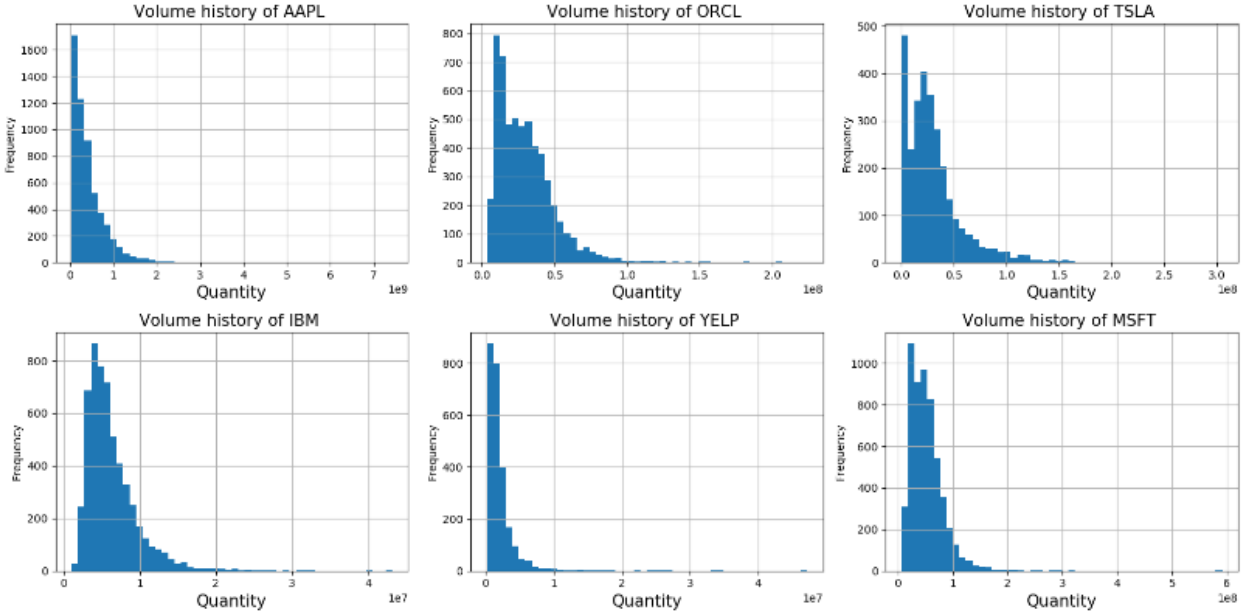
Open Price ($USD) Histogram



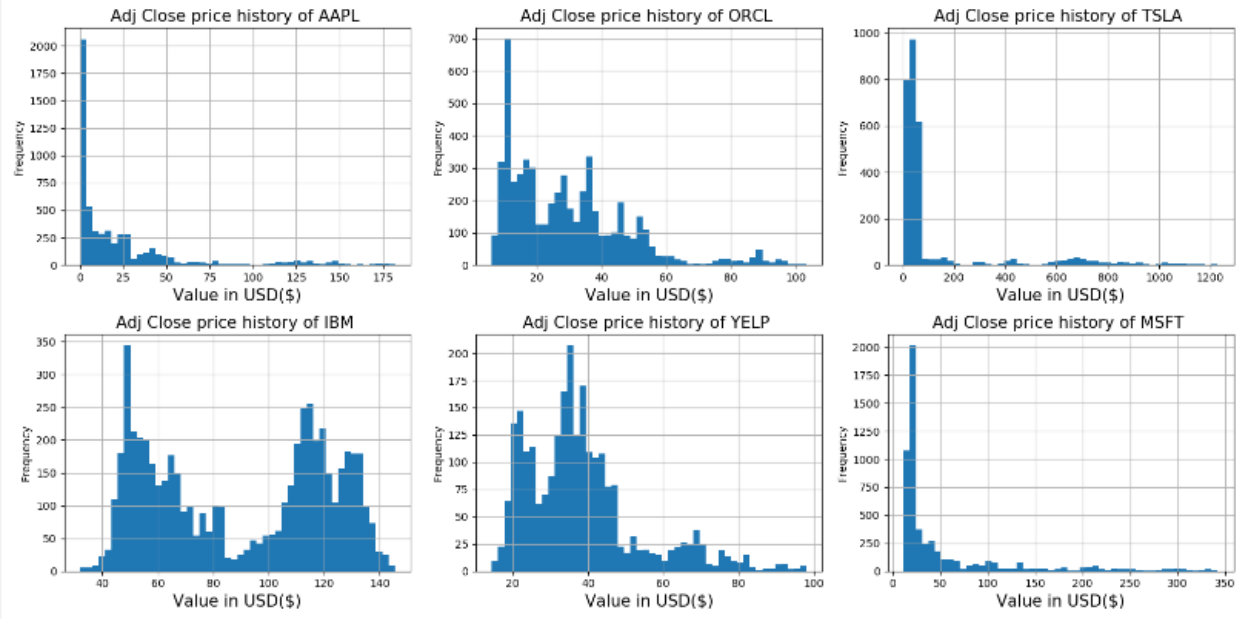
Close Price ($USD) Histogram



Volume Histogram



Adj Close Price ($USD) Histogram



6. Using the pandas package and. corr() function calculate the person correlation coefficients between all 6 features for the “AAPL” company. Display the correlation coefficient matrix through a table on the console. Which two feature has the highest correlation coefficient and which two features has the lowest correlation coefficient?

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.999890  0.999937  0.999917 -0.412495   0.999620
Low         0.999890  1.000000  0.999911  0.999918 -0.414764   0.999586
Open        0.999937  0.999911  1.000000  0.999840 -0.413401   0.999521
Close       0.999917  0.999918  0.999840  1.000000 -0.413642   0.999687
Volume     -0.412495 -0.414764 -0.413401 -0.413642  1.000000  -0.409563
Adj Close   0.999620  0.999586  0.999521  0.999687 -0.409563   1.000000
0.9999181271093175
AAPL ('Open', 'High') have highest correlation coefficient of 1.00
AAPL ('Volume', 'Low') lowest correlation coefficient of -0.41
```

7. Repeat question 6 for, “ORCL”, “TSLA”, “IBM”, “YELP” and “MSFT”.

Oracle

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.999469  0.999711  0.999700 -0.551196   0.997452
Low         0.999469  1.000000  0.999687  0.999697 -0.565413   0.997690
Open        0.999711  0.999687  1.000000  0.999467 -0.557553   0.997324
Close       0.999700  0.999697  0.999467  1.000000 -0.558491   0.997813
Volume     -0.551196 -0.565413 -0.557553 -0.558491  1.000000  -0.553658
Adj Close   0.997452  0.997690  0.997324  0.997813 -0.553658   1.000000
0.9997000682397463
ORCL ('Open', 'High') have highest correlation coefficient of 1.00
ORCL ('Volume', 'Low') lowest correlation coefficient of -0.57
```

Tesla

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.999595  0.999754  0.999671  0.100337   0.999671
Low         0.999595  1.000000  0.999617  0.999689  0.092262   0.999689
Open        0.999754  0.999617  1.000000  0.999297  0.096341   0.999297
Close       0.999671  0.999689  0.999297  1.000000  0.097150   1.000000
Volume      0.100337  0.092262  0.096341  0.097150  1.000000   0.097150
Adj Close   0.999671  0.999689  0.999297  1.000000  0.097150   1.000000
0.9996885414916639
TSLA ('High', 'Open') have highest correlation coefficient of 1.00
TSLA ('Volume', 'Low') lowest correlation coefficient of 0.09
```

IBM

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.999351  0.999556  0.999571 -0.383901   0.924120
Low         0.999351  1.000000  0.999544  0.999558 -0.402253   0.924675
Open        0.999556  0.999544  1.000000  0.999125 -0.392540   0.924341
Close       0.999571  0.999558  0.999125  1.000000 -0.393058   0.924585
Volume     -0.383901 -0.402253 -0.392540 -0.393058  1.000000  -0.426075
Adj Close   0.924120  0.924675  0.924341  0.924585 -0.426075   1.000000
0.9995582908527262
IBM ('High', 'Close') have highest correlation coefficient of 1.00
IBM ('Volume', 'Adj Close') lowest correlation coefficient of -0.43
```

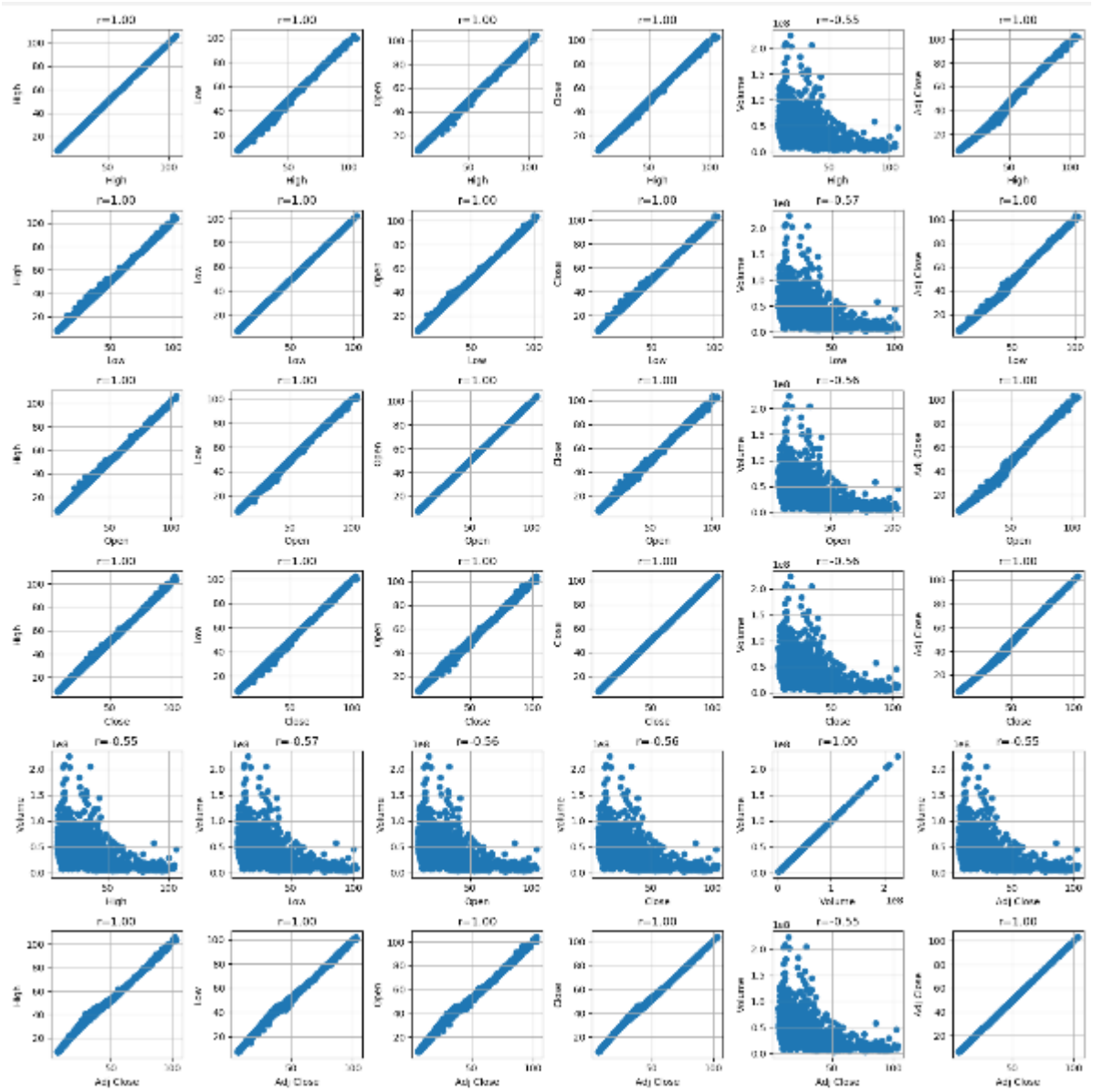
Yelp

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.998207  0.998754  0.998690  0.266151   0.998690
Low         0.998207  1.000000  0.998450  0.998843  0.237796   0.998843
Open        0.998754  0.998450  1.000000  0.997186  0.250192   0.997186
Close       0.998690  0.998843  0.997186  1.000000  0.253014   1.000000
Volume      0.266151  0.237796  0.250192  0.253014  1.000000   0.253014
Adj Close   0.998690  0.998843  0.997186  1.000000  0.253014   1.000000
0.9987544196302479
YELP ('Close', 'Low') have highest correlation coefficient of 1.00
YELP ('Volume', 'Low') lowest correlation coefficient of 0.24
```

Microsoft

```
                High       Low      Open     Close    Volume  Adj Close
High        1.000000  0.999871  0.999926  0.999906 -0.375817   0.999165
Low         0.999871  1.000000  0.999908  0.999917 -0.381045   0.999263
Open        0.999926  0.999908  1.000000  0.999835 -0.378182   0.999142
Close       0.999906  0.999917  0.999835  1.000000 -0.378754   0.999303
Volume     -0.375817 -0.381045 -0.378182 -0.378754  1.000000  -0.388660
Adj Close   0.999165  0.999263  0.999142  0.999303 -0.388660   1.000000
0.9999165616092668
MSFT ('High', 'Open') have highest correlation coefficient of 1.00
MSFT ('Volume', 'Adj Close') lowest correlation coefficient of -0.39
```

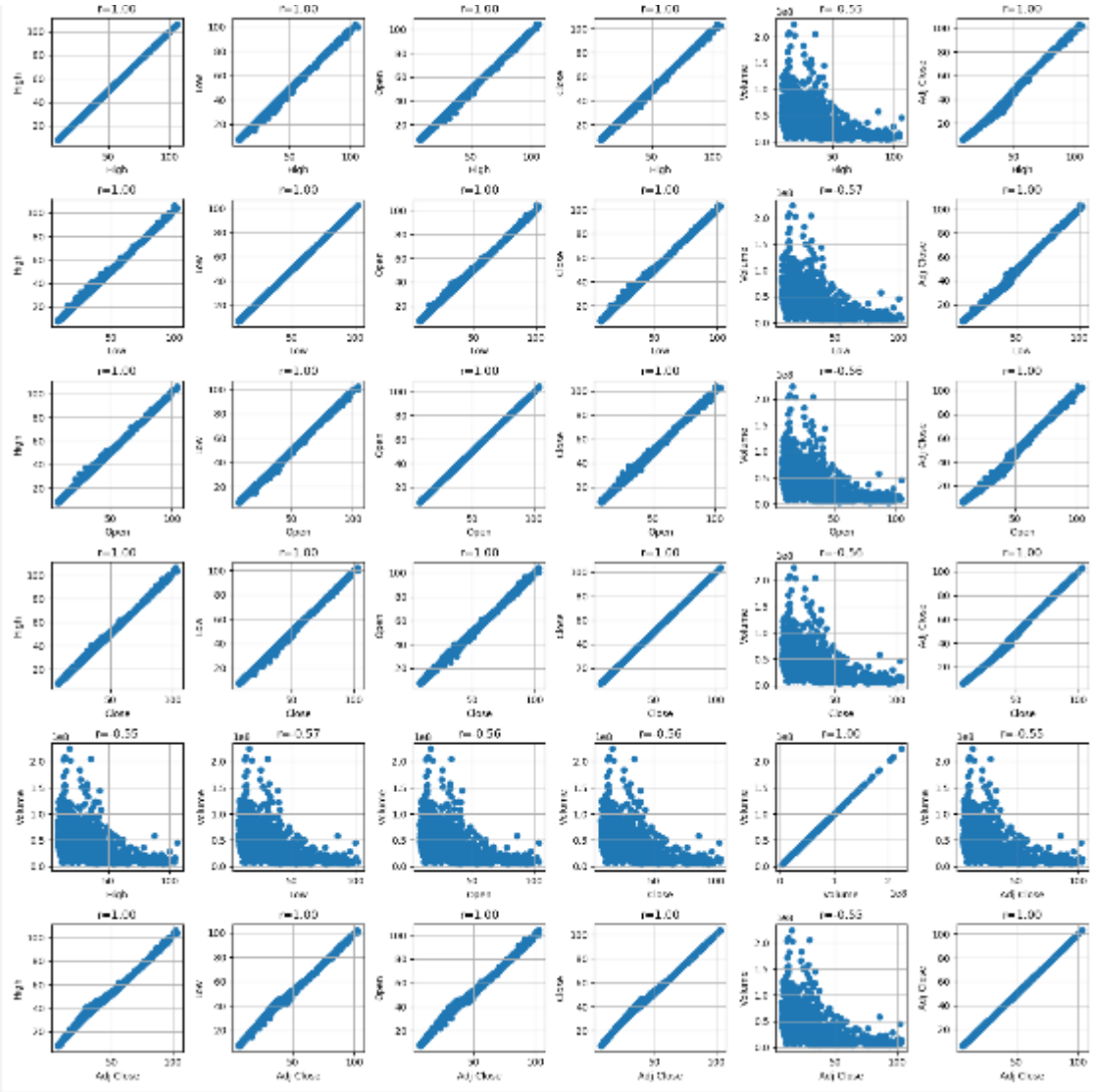
8. Using the matplotlib.pyplot package, subplot, and scatter() function plot the scatter plot for the “AAPL” company. You need to use the plt.subplots with 6x6 format to cover all the possible correlations between 6 feature. Add the calculated correlation coefficients in step 7 as a title to each subplot. Use two digits precision (.2f) for the correlation coefficients. Figure size = (16,16).
The final plot should look like the following.



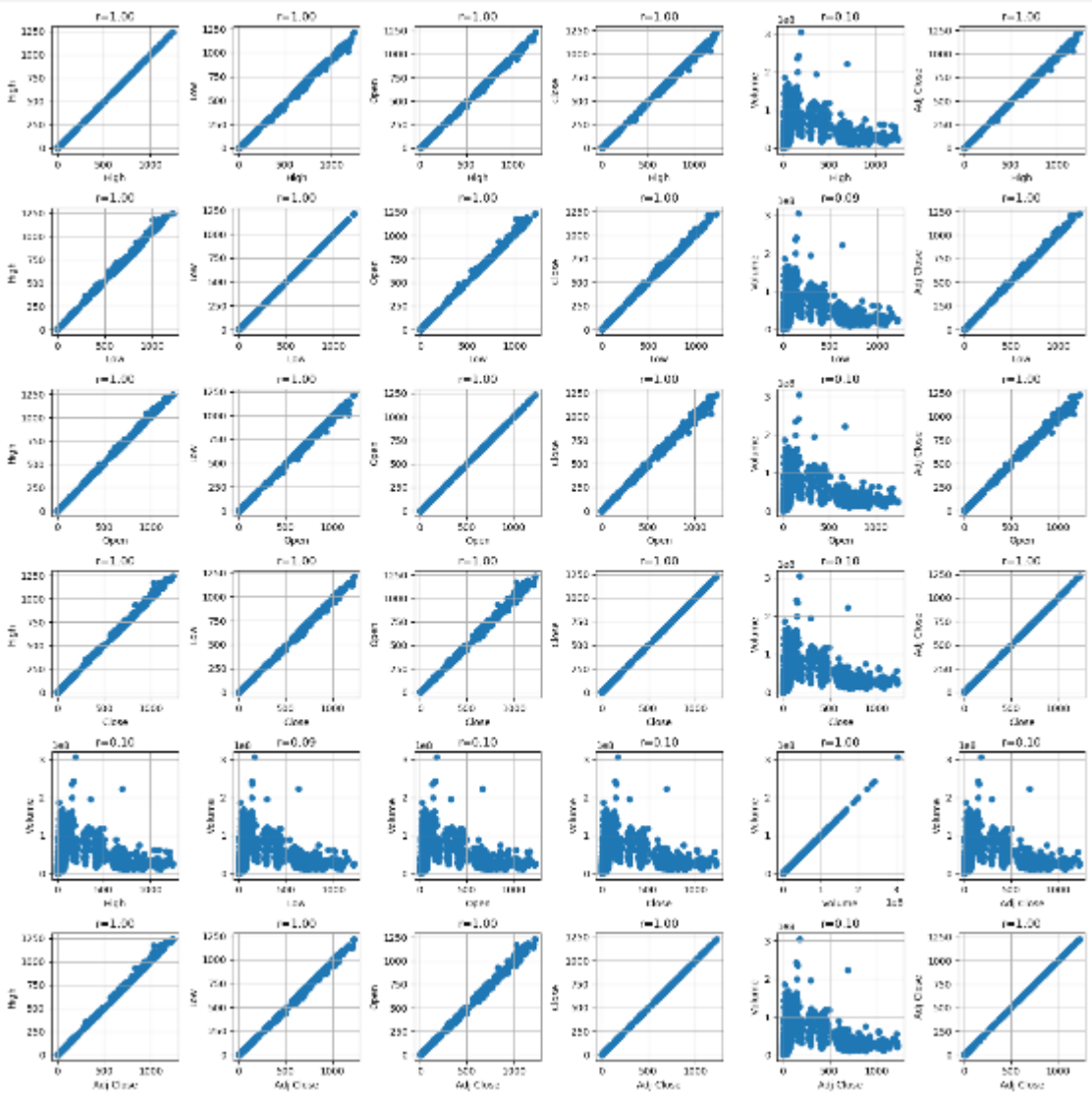
9. Repeat question 8 for, “ORCL”, “TSLA”, “IBM”, “YELP” and “MSFT”.

ORCL

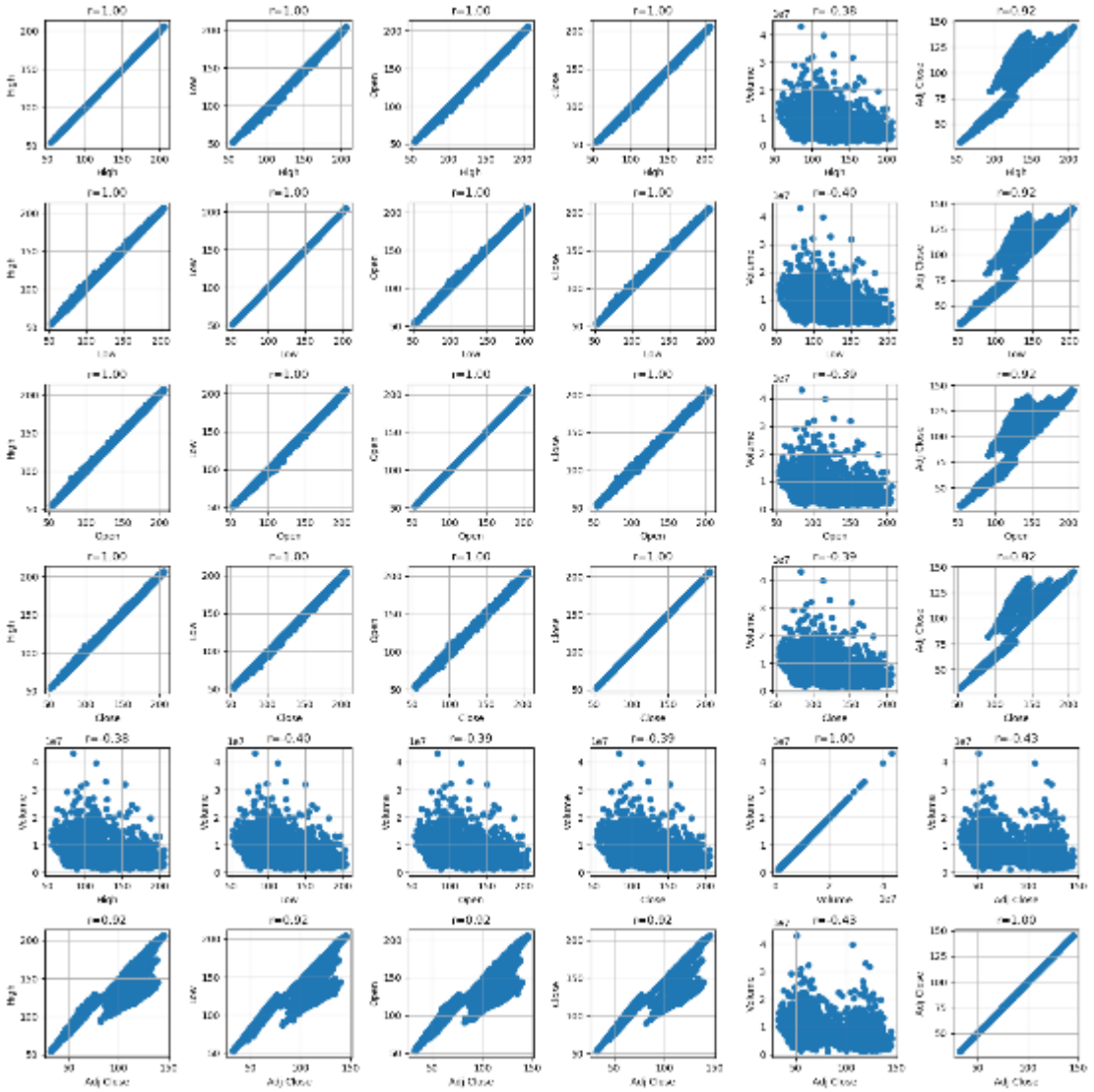
Lydia Teinfalt



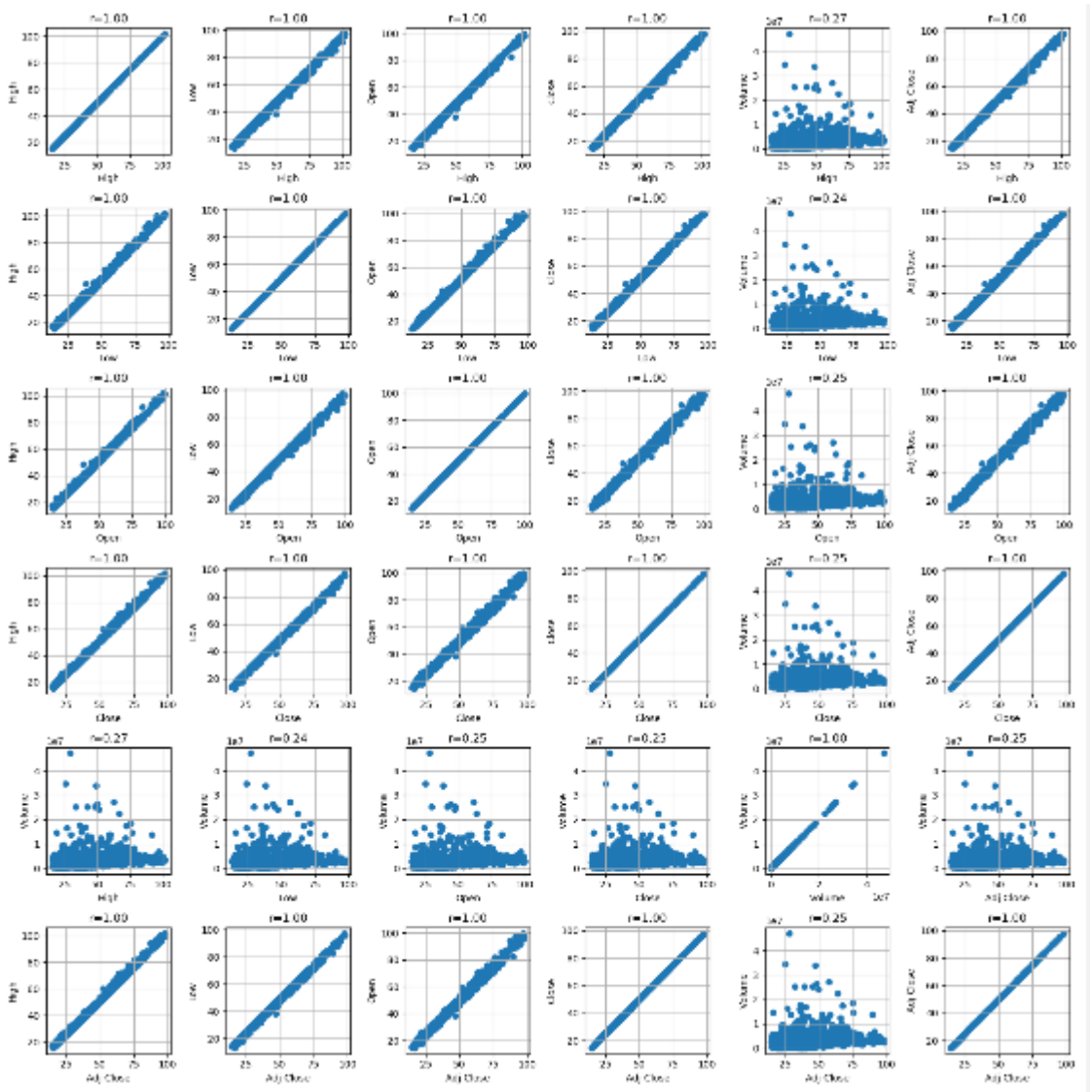
TSLA



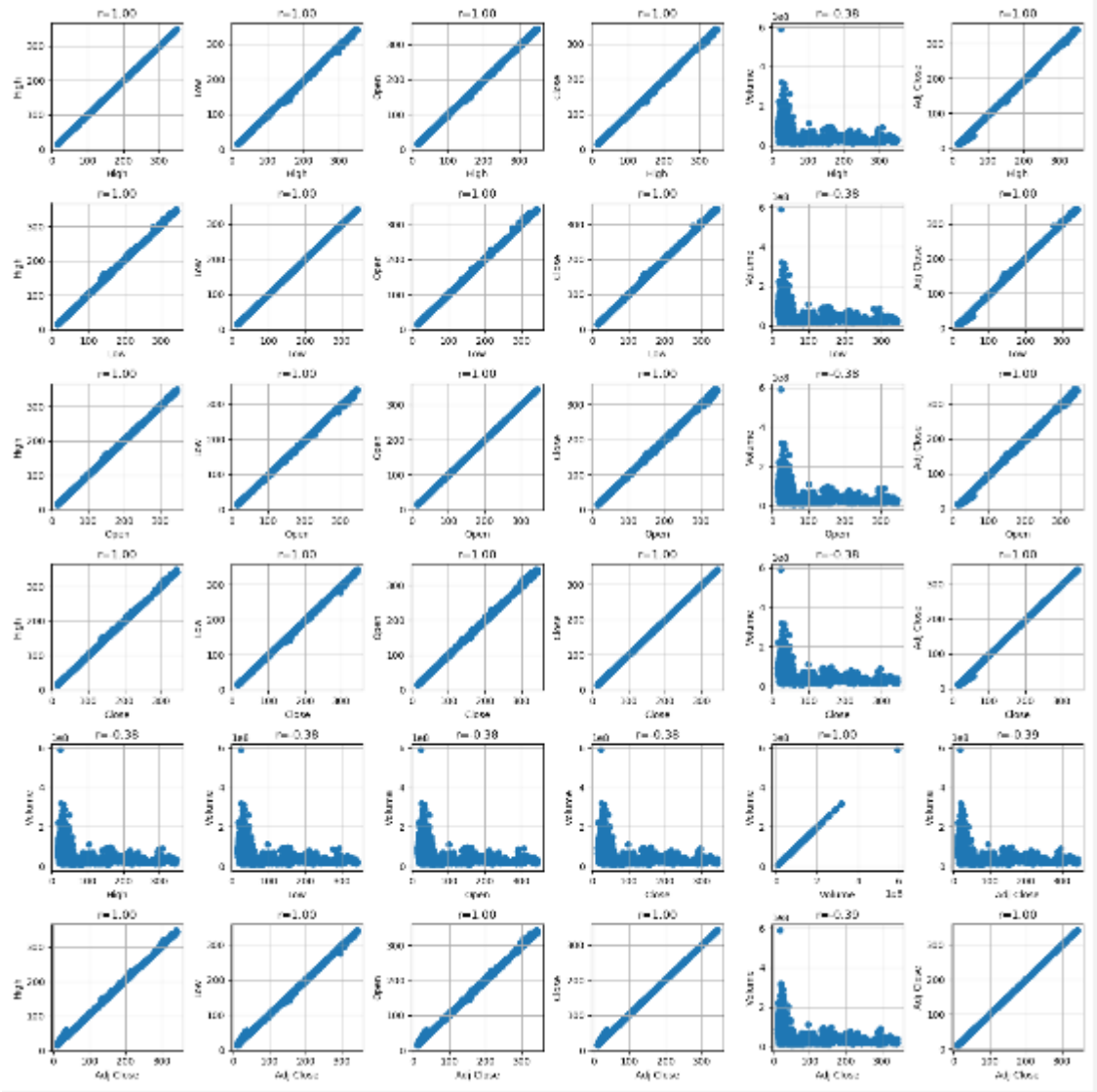
IBM



YELP



MSFT



10. Alternatively, one can use Pandas package to plot the scatter matrix. Using pandas package plot the scatter matrix plot of the “AAPL” company with the following parameters : hist_kwds= {‘bins’ : 50} , alpha = 0.5, s = 10, diagonal = ‘kde’. Hint: you can use the following command : pd.plotting.scatter_matrix()

Lydia Teinfalt
02/17/2022