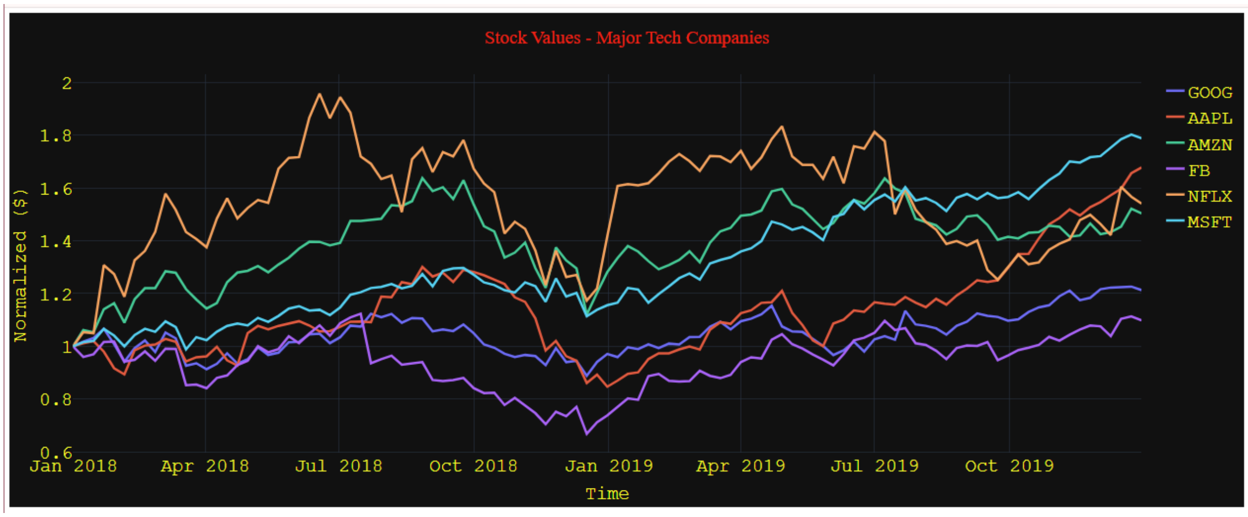
1.

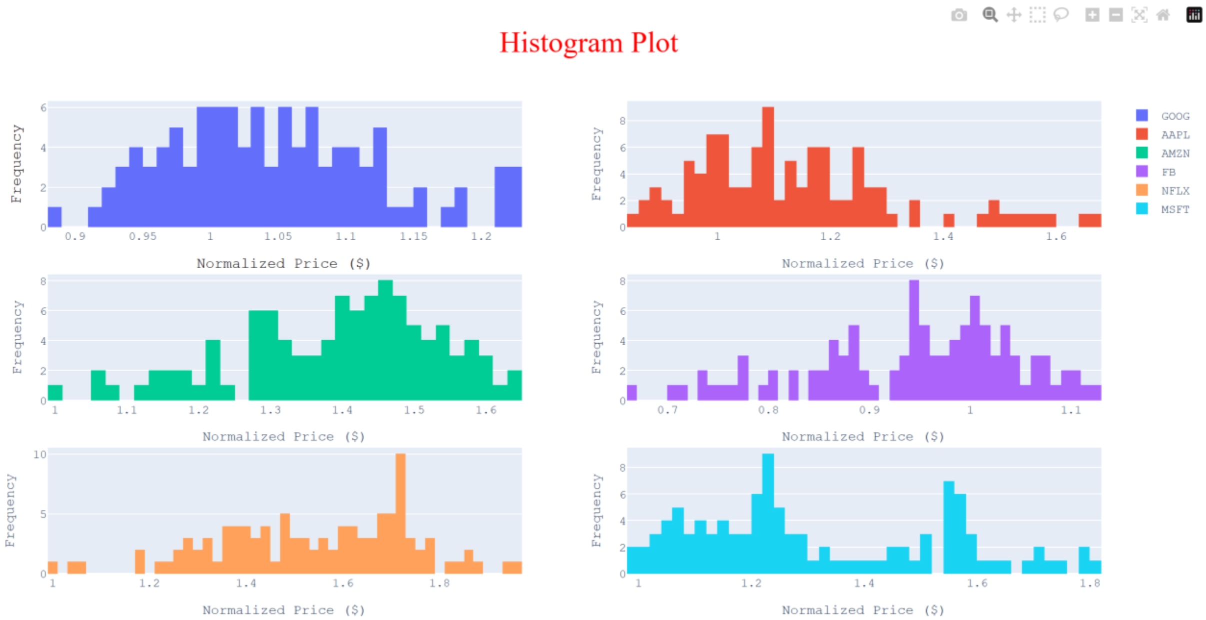
```
... print("Features:", df.columns.tolist())
... print(df.tail())
...
Features: ['date', 'GOOG', 'AAPL', 'AMZN', 'FB', 'NFLX', 'MSFT']
           date      GOOG      AAPL      AMZN        FB      NFLX      MSFT
100  2019-12-02  1.216280  1.546914  1.425061  1.075997  1.463641  1.720717
101  2019-12-09  1.222821  1.572286  1.432660  1.038855  1.421496  1.752239
102  2019-12-16  1.224418  1.596800  1.453455  1.104094  1.604362  1.784896
103  2019-12-23  1.226504  1.656000  1.521226  1.113728  1.567170  1.802472
104  2019-12-30  1.213014  1.678000  1.503360  1.098475  1.540883  1.788185
```

This is a time-series dataset. The feature in this dataset represents the date and 6 stocks (GOOG, AAPL, AMZN, FB, NFLX, MSFT) with their corresponding closing price in 2018/2019.

2.



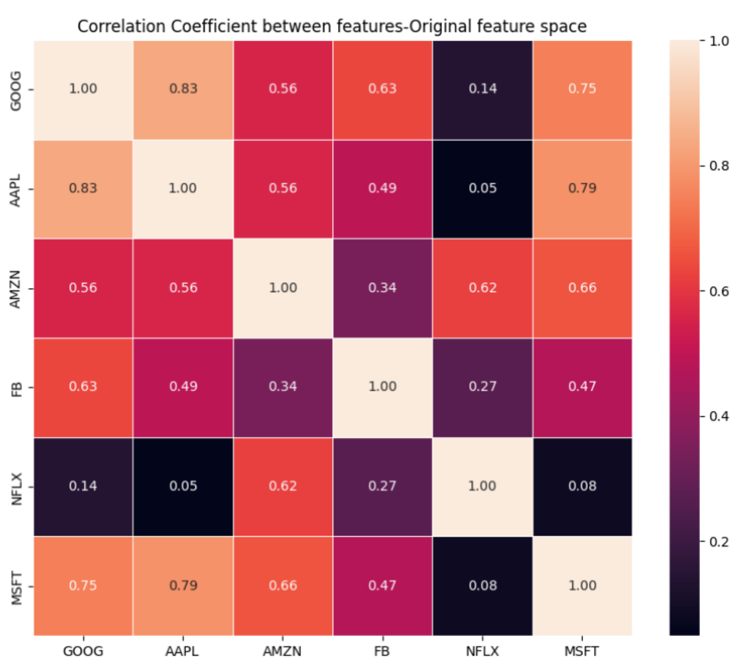
3.



4b.

```
Singular values:
 [19.34512086 11.38793518  8.49848778  5.20941662  3.8769269   3.4188233 ]
Condition Number: 5.658414955290657
```

4c.

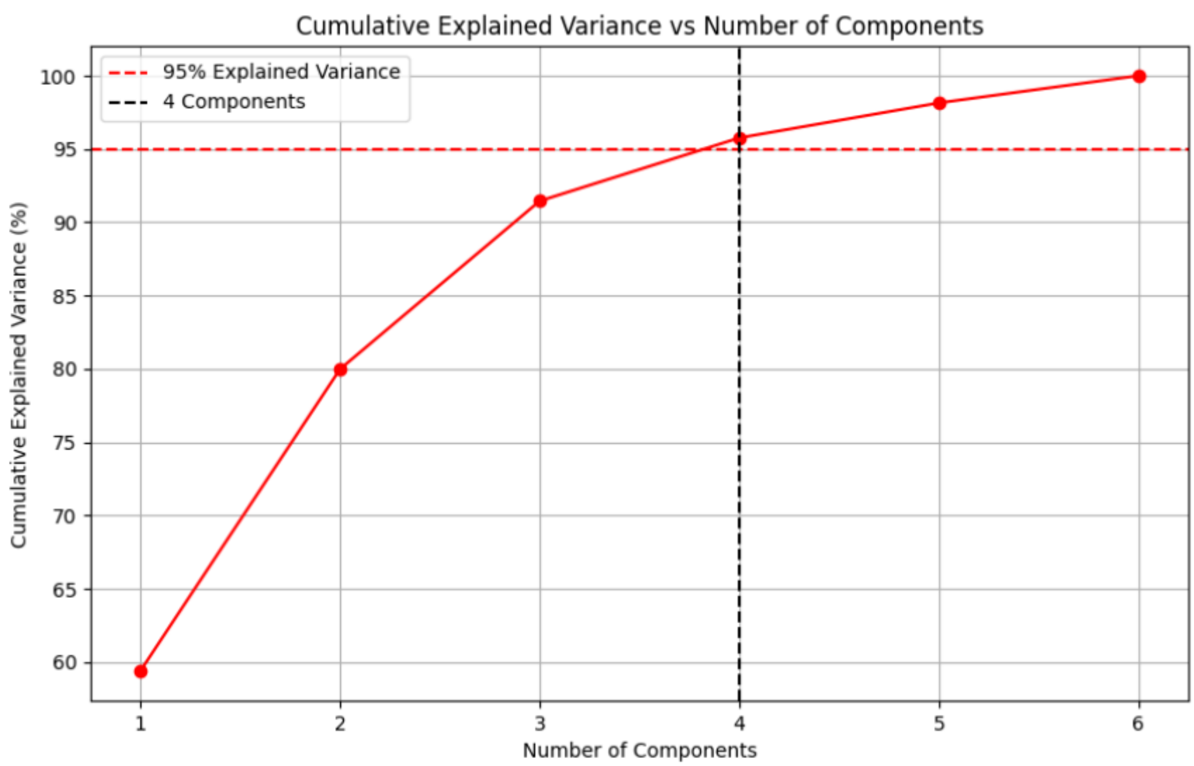


4d.

Two features should be removed. The reason is that when the threshold is 95%, there are two features that have cumulative explained variance ratio that higher than 95% (98.14% and 100%).

```
Explained variance ratio (original feature space):
 [0.59402175 0.20584931 0.11464174 0.04307622 0.02385804 0.01855294]
Explained variance ratio (reduced feature space): 0.957589023917707
```

4e.

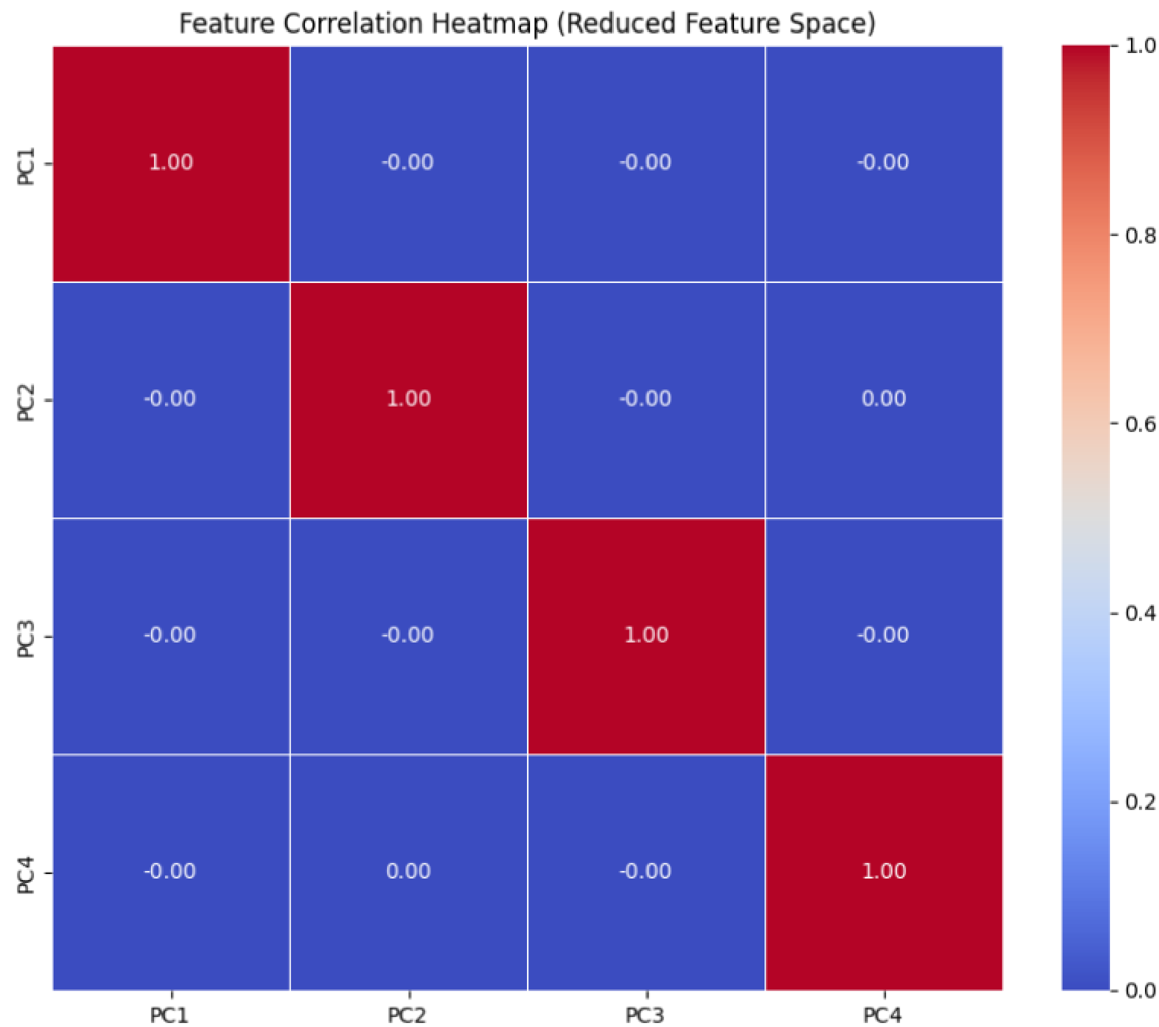


4f.

```
Singular values (Reduced Feature Space):
 [19.34512086 11.38793518  8.49848778  5.20941662]
Condition Number (Reduced Feature Space): 3.713490831529809
```

The singular values of the reduced feature space will be smaller compared to the original space because PCA removes low-variance components. The first few singular values remain relatively large, representing the retained variance in the dataset. Since PCA removes near-zero singular values (which contribute to multicollinearity), the condition number of the reduced dataset should be significantly lower, indicating a more numerically stable dataset. This confirms that PCA reduces redundancy and collinearity in the dataset.

4g.

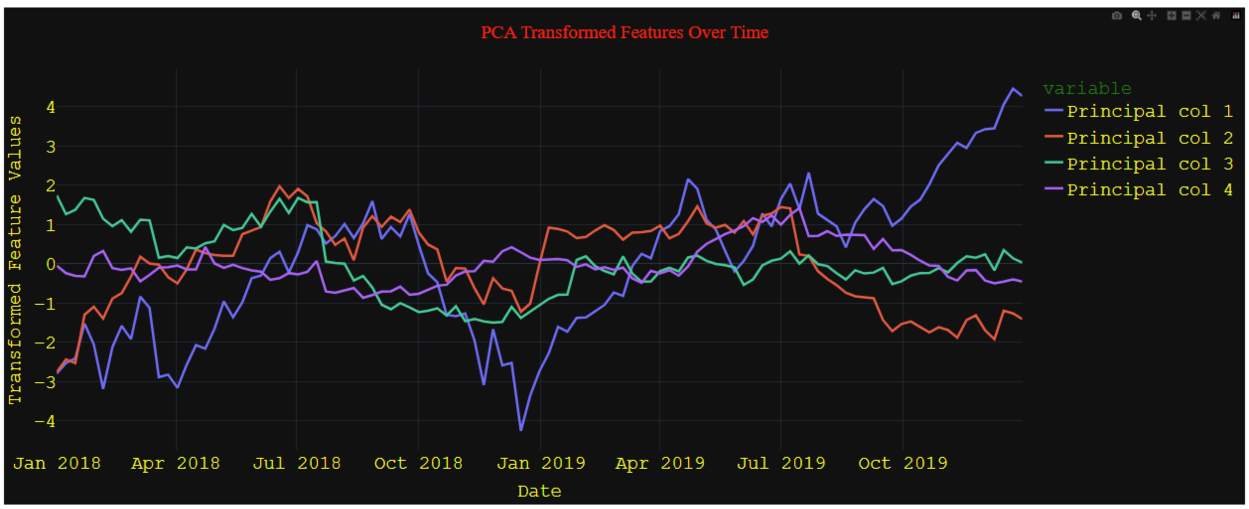


The correlation values appear to be around 0.00, so the principal components are uncorrelated among each other.

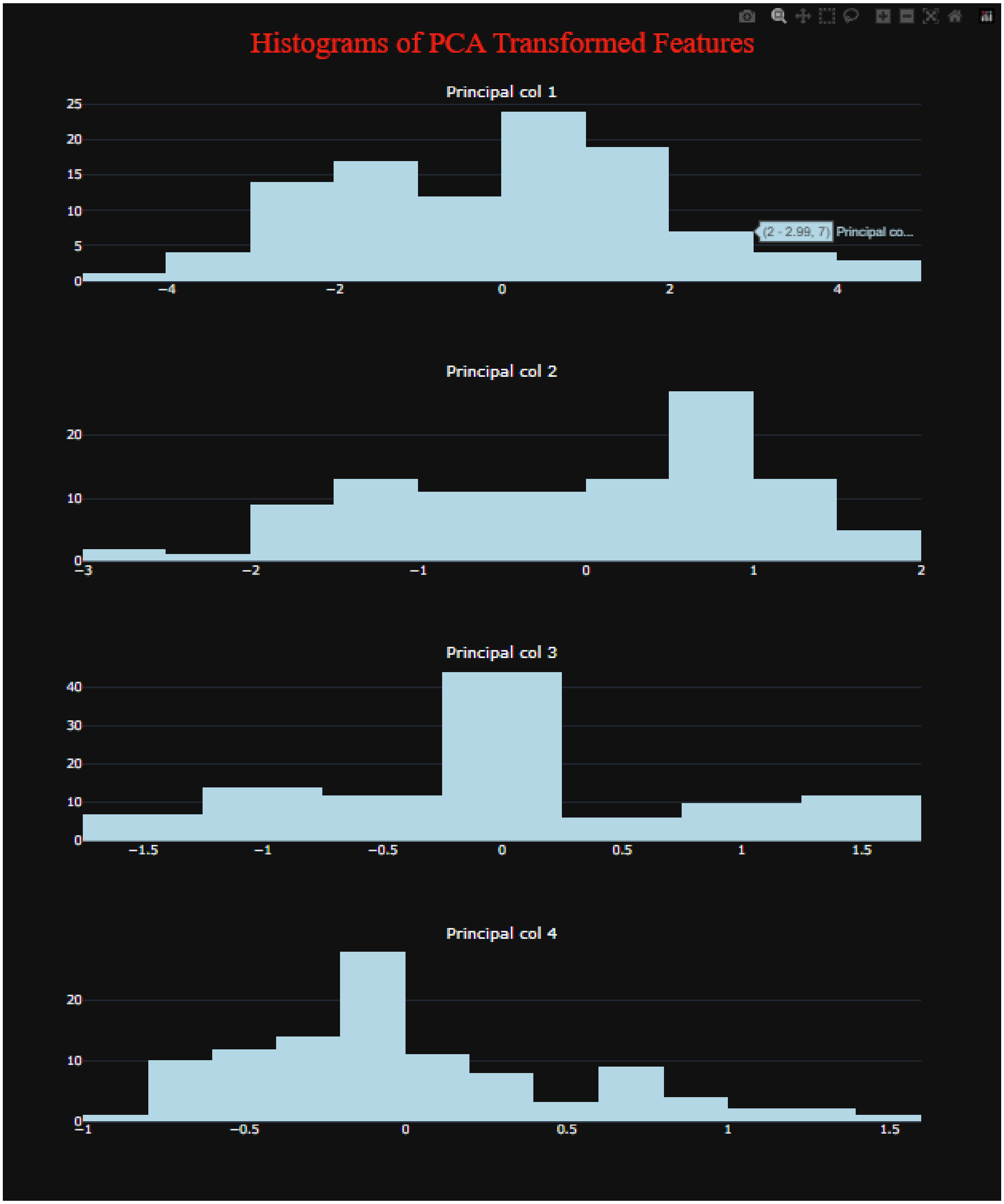
4h.

	Principal col 1	Principal col 2	Principal col 3	Principal col 4
0	-2.797664	-2.762156	1.738296	-0.051832
1	-2.529215	-2.434368	1.260494	-0.241111
2	-2.407053	-2.530062	1.371260	-0.311323
3	-1.526472	-1.298017	1.675201	-0.321514
4	-2.057901	-1.095378	1.622487	0.198700

4i.



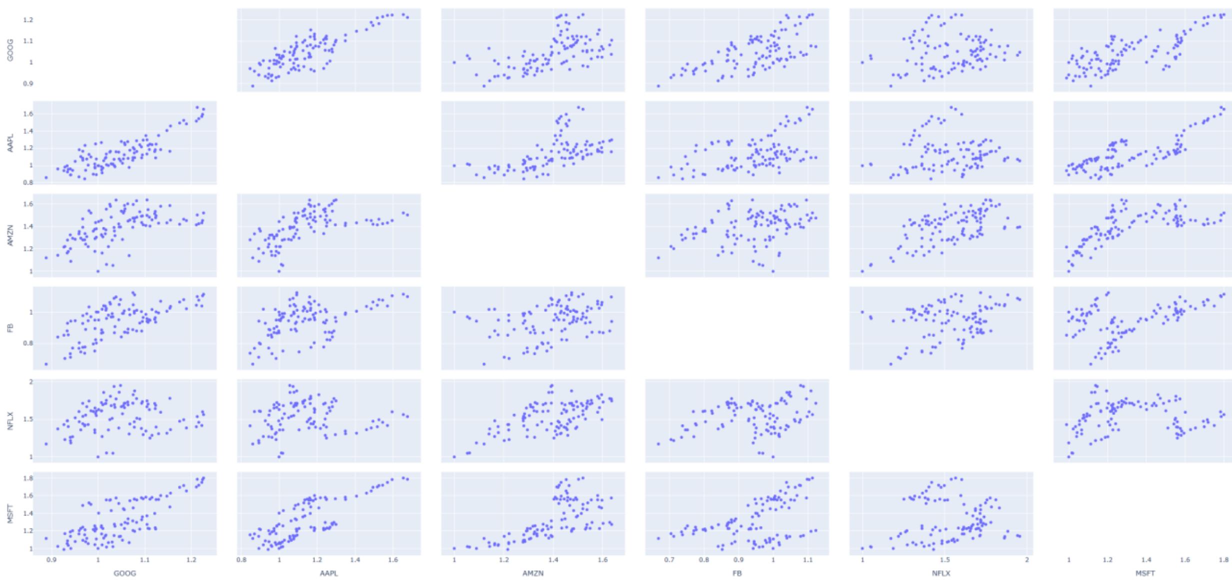
4j.



Histograms of PCA Transformed Features
Principal col 1
25
20
15
10
5
0
−4
−2
0
2
4
(2 - 2.99, 7) Principal co...
Principal col 2
20
10
0
−3
−2
−1
0
1
2
Principal col 3
40
30
20
10
0
−1.5
−1
−0.5
0
0.5
1
1.5
Principal col 4
20
10
0
−1
−0.5
0
0.5
1
1.5

4k.

Scatter Matrix of Original Feature Space



Scatter Matrix of Reduced Feature Space