Introduction

Stock markets are widely known as complex and multidimensional due to its complex and closely related variables (Lin, Shang & Zhou, 2014). The dataset reported the performance of 147 stocks in 5 securities markets under 9 financial variables and ESG (Environment, Social, and Governance) risk score. This report produces boxplots and biplots based on principal components to further analyze the relationship between those financial and non-financial variables. These analyses will help to identify the relationship between valuation index and the stock market.

Under the statistical summary, the variables of "forwarding price to earnings ratio (Forward P/E)" and "the five years expected price/earnings to growth ratio (PEG ratio)" have too many missing values (with 80 and 81 "NA" respectively), so these two variables will be ignored in the subsequent analysis. Moreover, the variable of "total ESG risk score" is the weight sum of the other three non-financial risk scores' variables. Thus, to increase the accuracy of the analysis results, this report also removes the variable of "total ESG risk score" during analysis.

Preliminary Analysis

Producing the boxplots of the selected variables and making comparisons with each one help to explore the features of variables before carrying out a Principal Components Analysis (PCA). The following is the faceted boxplot.

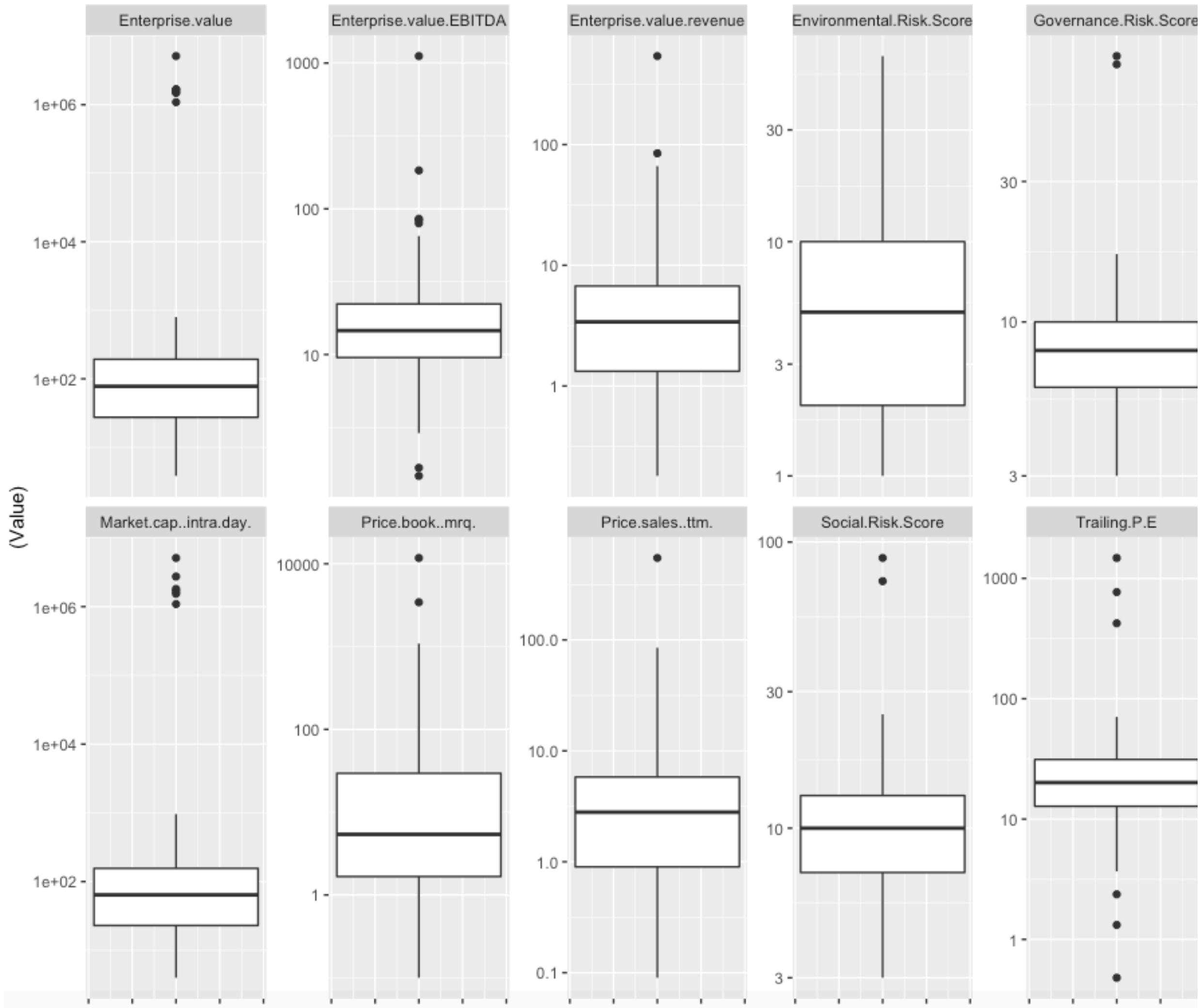


Figure 1: The Faceted Boxplot of Each Variables.

According to Figure 1, There evidently exist several outliers in the distribution for selected variables, except "Environmental Risk Score". Besides, "Enterprise Value" variable and "Market cap (intra-day)" variable have a very similar shape in the box plot, it may because the two variables have the same effect on stock price. The same relationship also exists between the variable of "Enterprise value revenue" and "Price sales ttm.". To verify this supposition, correlations between several variables are produced as the following table.

```
> round(cor(all.num.nona),5) #compute corr table
                         Market.cap..intra.day. Enterprise.value Trailing.P.E Price.sales..ttm. Price.book..mrq. Enterprise.value.revenue
Market.cap..intra.day.                  1.00000          0.99986      0.00098           0.06369         -0.03809                  0.05664
Enterprise.value                        0.99986          1.00000      0.00089           0.06291         -0.03803                  0.05601
Trailing.P.E                            0.00098          0.00089      1.00000           0.08837         -0.01559                  0.08568
Price.sales..ttm.                       0.06369          0.06291      0.08837           1.00000         -0.05630                  0.88921
Price.book..mrq.                       -0.03809         -0.03803     -0.01559          -0.05630          1.00000                 -0.06765
Enterprise.value.revenue                0.05664          0.05601      0.08568           0.88921         -0.06765                  1.00000
Enterprise.value.EBITDA                 0.02751          0.02742      0.31472           0.51142         -0.04097                  0.61254
Environmental.Risk.Score               -0.10664         -0.10651     -0.05026           0.03930         -0.02758                  0.07274
Social.Risk.Score                      -0.00533         -0.00486      0.04614          -0.04605         -0.01234                 -0.02632
Governance.Risk.Score                  -0.02682         -0.02618      0.00624          -0.06304         -0.06748                 -0.04933
                         Enterprise.value.EBITDA Environmental.Risk.Score Social.Risk.Score Governance.Risk.Score
Market.cap..intra.day.                   0.02751                 -0.10664          -0.00533              -0.02682
Enterprise.value                         0.02742                 -0.10651          -0.00486              -0.02618
Trailing.P.E                             0.31472                 -0.05026           0.04614               0.00624
Price.sales..ttm.                        0.51142                  0.03930          -0.04605              -0.06304
Price.book..mrq.                        -0.04097                 -0.02758          -0.01234              -0.06748
Enterprise.value.revenue                 0.61254                  0.07274          -0.02632              -0.04933
Enterprise.value.EBITDA                  1.00000                 -0.06292          -0.05064              -0.07466
Environmental.Risk.Score                -0.06292                  1.00000           0.74167               0.78630
Social.Risk.Score                       -0.05064                  0.74167           1.00000               0.92403
Governance.Risk.Score                   -0.07466                  0.78630           0.92403               1.00000
```

Table 1: Correlation Table

It is clear that there is a high correlation between "Enterprise Value" and "Market cap (intra-day)" which is valued 0.99986. Also, the correlation between "Enterprise value revenue" and "Price sales ttm." is 0.88921. Considering dropping and rebuilding with fewer variables in PCA may be more productive. However, further demonstration of this assumption will be discussed in the PCA part.

Moreover, “Price/book (mrq)” are positively skewed, which means the data constitutes higher frequency of high valued scores, the data in “Price/sales (ttm).” are negatively skewed and constitute higher frequency of low valued scores. It suggests that there may be a group of stocks that are particularly disadvantaged especially with respect to the two variables.

After detailed investigation, the outliers are found as listed："MSFT","AAPL", "RDSA.L", "LIN.DE", "0006.HK", "1299.HK", "TSLA", "1928.HK", "DB1.DE", "ILMN", "RTO.L", "VILMN", "1038.HK". These outliers bring significant errors and cannot be controlled affecting the further research. In order to figure out the influence of each variable, more filtered accumulations and dimensional reduction are needed in next steps.

Principal Components Analysis (PCA)

Nobi & Lee (2016) state that PCA is one of the common methods to identify the deriving correlation valuation index of stock markets and to analyse the related risk including environmental, social and governance risk.

After removing outliers, by using two principal components factors to produce the biplot (Figure 2 as follows).

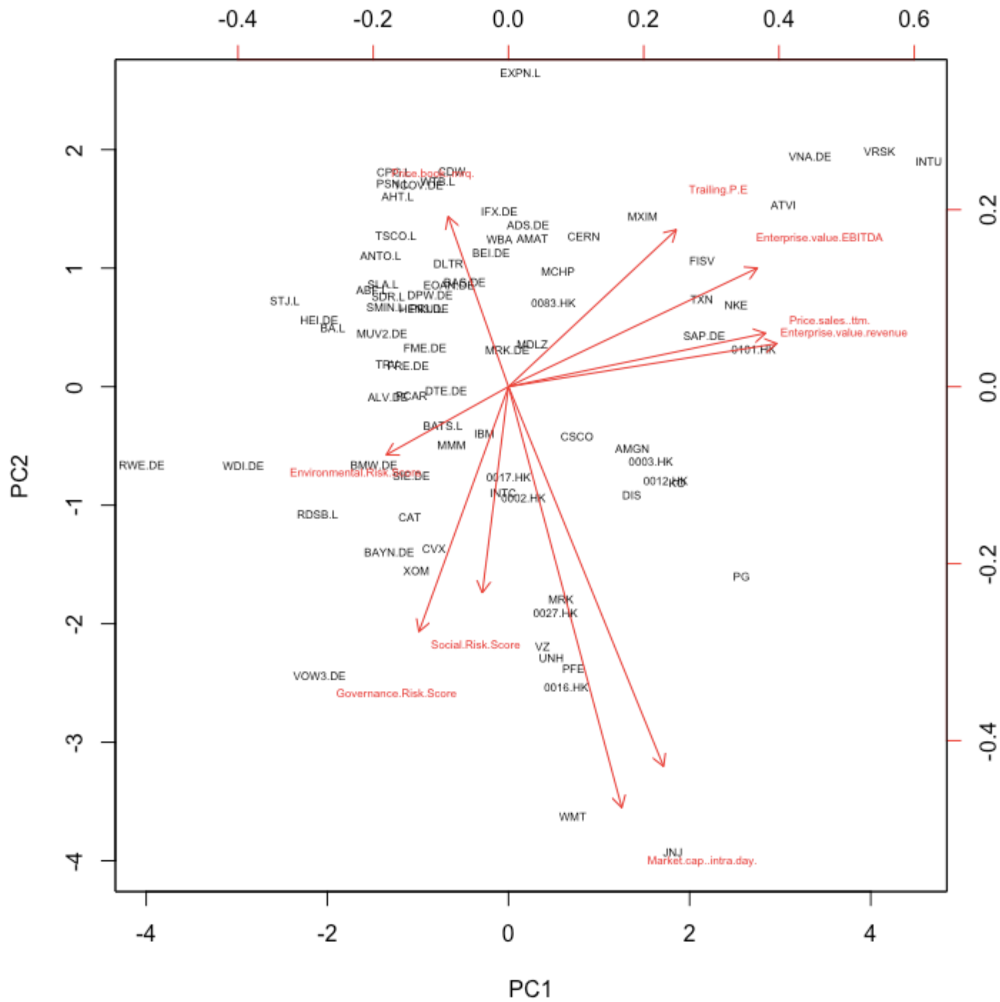


Figure 2: Two Principal Components of Biplot in Correlation.

```
> pca_rmoutliers$rotation
                                PC1         PC2          PC3         PC4         PC5
Market.cap..intra.day.    0.28671621 -0.53646484  0.043034481 -0.10016660 -0.30010170
Enterprise.value          0.20953673 -0.59474089  0.079382032 -0.08776315 -0.23845582
Trailing.P.E              0.31048334  0.22186787  0.054345460 -0.60961129  0.05240948
Price.sales..ttm.         0.47542557  0.07546671 -0.154892613  0.16186372  0.27141394
Price.book..mrq.         -0.11163706  0.24034313 -0.633013139 -0.17983977 -0.25481050
Enterprise.value.revenue  0.49643596  0.06081818 -0.065389878  0.15912742  0.29600932
Enterprise.value.EBITDA   0.46012395  0.16759505  0.003181648 -0.08178946  0.06609046
Environmental.Risk.Score -0.22599149 -0.09628825  0.352405099 -0.55770164  0.47000486
Social.Risk.Score        -0.04841086 -0.29133076 -0.619405806 -0.36455002  0.15688215
Governance.Risk.Score    -0.16558225 -0.34622444 -0.228159759  0.27596605  0.61174429
```

Table 2: The Principle Component Weights effect of remove outliers

The first principal component (PC1) is a method to measure the overall stock market. According to Table 2, “Price/book (mrq)”, “Environment risk”, “Social risk” and “Governance risk” values are negatively correlated with stock. It presents that these variables take a negative effect on stock price. Whereas, the rest of the variables show a positive correlated relationship, including “Market cap (intra-day)”, “Enterprise value”, “Trailing P/E”, “Price/sale”, “Enterprise value/EBITDA” and “Enterprise value/revenue”.

According to Table 2, the second principal component (PC2) indicates that “Market cap (intra-day)”, “Enterprise value”, “Environment risk”, “Social risk” and “Governance risk” are negatively correlated with stock. For example, Verizon Communications Inc (VZ), UnitedHealth Group Incorporated (UNH), Pfizer Inc (PFE) and Sun Hung Kai Properties Limited (0016.HK) present low values in the second principal component.

Moreover, as shown in Figure 2, the variables of “Trailing P/E”, “Enterprise value/EBITDA”, “Price/sales (ttm)” and “Enterprise value/revenue” are highly positively correlated relationships. Take “Enterprise value/EBITDA” as an example, it can be recognised as one of the most significant measures of a company's value for a future acquirer (Walkshäusl & Lobe, 2015). Based on Figure 2, “Enterprise value/EBITDA” is uncorrelated with “Price/book (mrq)”, “Market cap”. Moreover, “Trailing P/E”, “Enterprise value/EBITDA”, “Price/sales” and “Enterprise value/revenue” values are highly uncorrelated with “Environmental risk”, “Governance risk” and “Social risk”. Furthermore, some stocks including Fiserv, Inc (FISV), Texas Instruments

Incorporated (TXN) and NIKE, Inc (NKE) have highly positive value with “enterprise value/EBITDA” factor.

Further PCA Analysis

As discussed before, there is a strong correlation between “Market cap (intra day)” and “Enterprise value” and between “Enterprise value/revenue” and “Price/sales (ttm)”, with 0.99986 and 0.88921 respectively. ADAM (2020) argued that price sales factor takes an important role in stock prices. In this case, there is a new principal component analysis which removes variables of “Enterprise value/revenue” and “Price/sales (ttm)” for further analysis. After removing “enterprise value/revenue.” the correlation among variables does not change. The new biplot details showed in the following Figure 3.

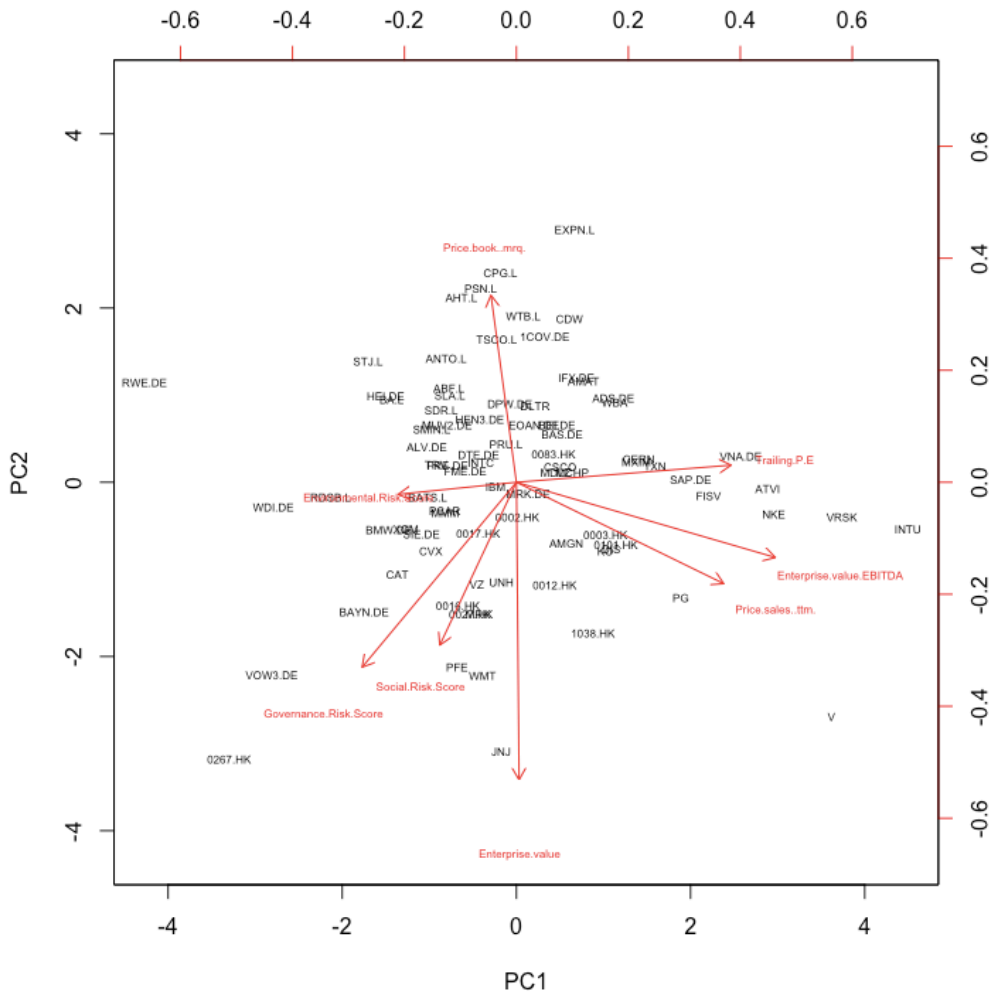


Figure 3: The Correlation Biplot after Remove "Enterprise value/revenue." & "Market cap (intra day)"

Obviously, Figure 3 represents that the data density decreased, compared with Figure 2. It makes the biplot easier to do analyses. Although the position of stocks was changed, the distance among each stock keeps constant. For example, the stock of "INTU" is still influenced by "Enterprise value/EBITDA" and "Trailing P/E" show a strong positive relationship. Moreover, there is an outlier that is named as "0267.HK" still shows a strong positive relationship with "governance risk score" and "social risk score", which are the same as Figure 3.

```
> pca_new$rotation
                                 PC1         PC2         PC3         PC4         PC5
Enterprise.value         0.005959497 -0.66418772  0.06887129 -0.07082741 -0.02485762
Trailing.P.E             0.479520233  0.03812592 -0.04356598  0.37262478 -0.56709020
Price.sales..ttm.        0.463078096 -0.22648065 -0.14560724  0.09114541  0.68737511
Price.book..mrq.        -0.056920123  0.41761540 -0.66615365  0.12177106  0.24639348
Enterprise.value.EBITDA  0.578522277 -0.16810534 -0.04456556  0.04456178  0.02054633
Environmental.Risk.Score -0.263961617 -0.02691092  0.32750681  0.85848811  0.23998283
Social.Risk.Score       -0.171082870 -0.36364657 -0.62753126  0.28771305 -0.28196772
Governance.Risk.Score   -0.344543900 -0.41289532 -0.15923037 -0.10598203  0.08412381
```

Table 3: The Principle Component Weights after Remove "Enterprise value/revenue." & "Market cap (intra.day)"

Table 3 provides the weights of each PC factor taken up in each variable. There are three variables that were affected obviously, including "Enterprise value", "Price/book (mrq)", and "Governance Risk Score". However, the relationship between the PC factors and variables does not change.

```
> summary(pca_new)
Importance of components:
                          PC1    PC2    PC3    PC4    PC5     PC6     PC7     PC8
Standard deviation     1.4894 1.2178 1.0833 0.9522 0.8390 0.82830 0.71310 0.56537
Proportion of Variance 0.2773 0.1854 0.1467 0.1133 0.0880 0.08576 0.06356 0.03996
Cumulative Proportion  0.2773 0.4627 0.6094 0.7227 0.8107 0.89648 0.96004 1.00000
```

Table 4: The Importance of Components after Remove "Enterprise value/revenue." & "Market cap (intra.day)"

After removing the variable of "Enterprise value/revenue." and "Market cap (intra day)", The first two PC factors can only explain 46.27% dataset, which decreased 6.21% compared to 52.48 %. However, the new PCA analysis uses less PC factor (8 PC factors) to explain the whole dataset.

Limitation

One limitation of analysis is principal components analysis is not a comprehensive representation of all observations. Only 64.49% of the overall variation was explained by the first three principal components.

According to the scree-plot, it needs three principal component factors to explain the whole dataset appropriately. However only two principal components were used to explain the data in this report, which reduced the accuracy in the analysis.

Finally, this report removed 13 outliers of the 147 observations during the whole analysis. It is likely to lead to data misleading.

Reference:

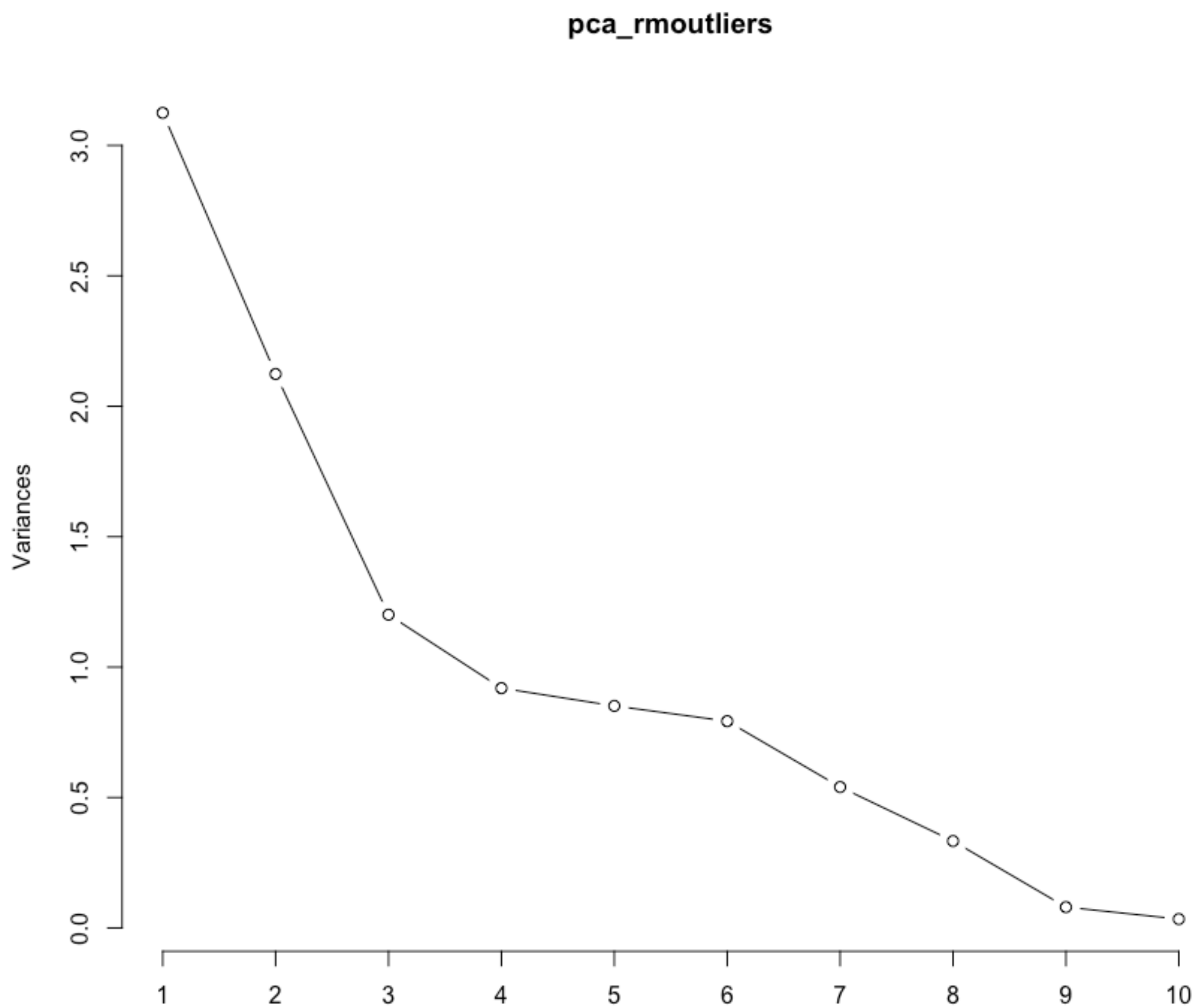
ADAM, H. (2020). Enterprise Value-to-Sales – EV/Sales Definition. Retrieved from https://www.investopedia.com/terms/e/enterprisevaluesales.asp

Lin, A., Shang, P., & Zhou, H. (2014). Cross-correlations and structures of stock markets based on multiscale MF-DXA and PCA. *Nonlinear Dynamics*, *78*(1), 485–494._ https://doi.org/10.1007/s11071-014-1455-5

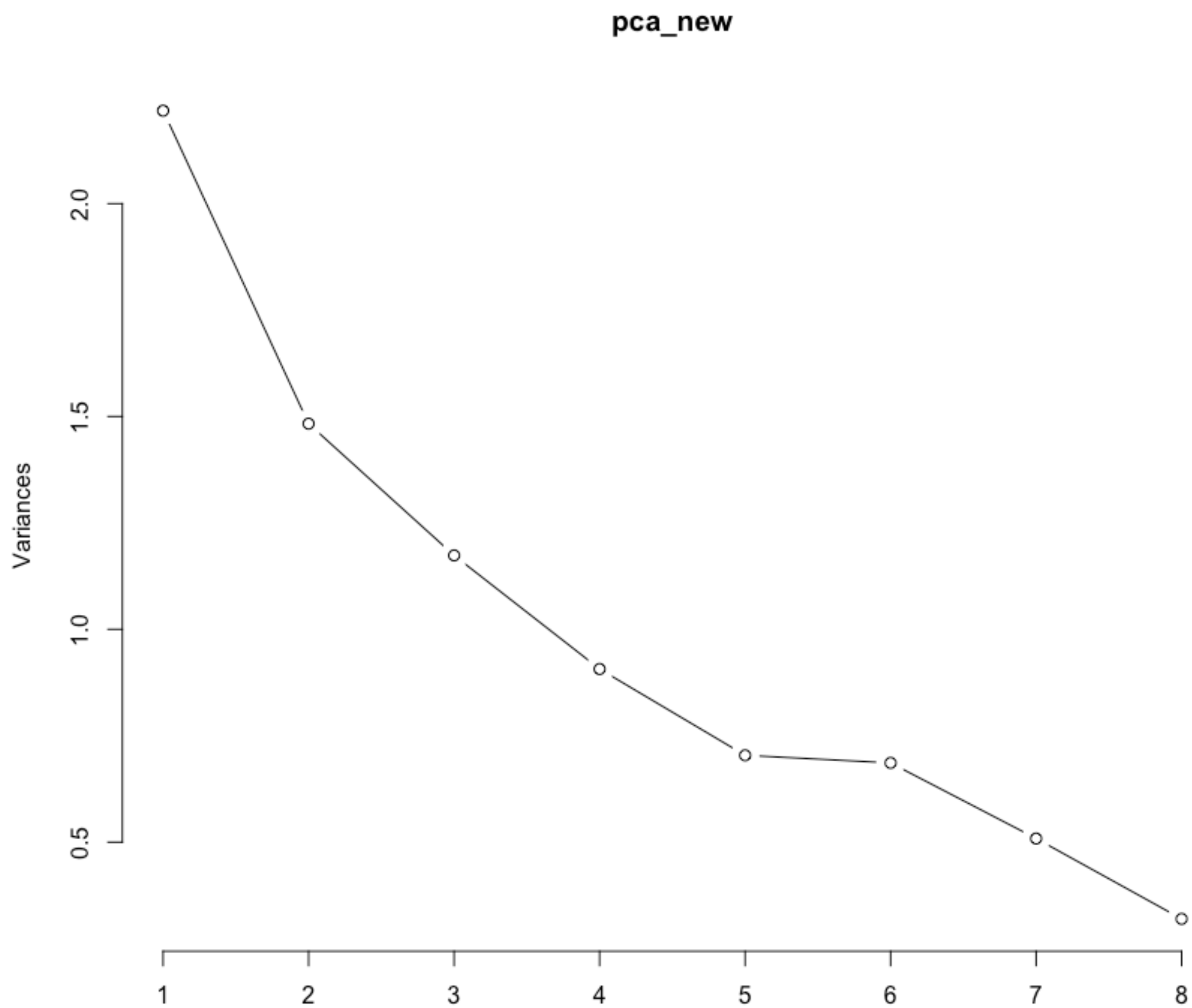
Nobi, A., & Lee, J. W. (2016). State and group dynamics of world stock market by principal component analysis. Physica A, 450(C), 85–94._ https://doi.org/10.1016/j.physa.2015.12.144

Walkshäusl, C., & Lobe, S. (2015). The Enterprise Multiple Investment Strategy: International Evidence. 50(4), 781–800. https://doi.org/10.1017/S002210901500023X

Appendix:



The Scree-plot after Removing Outliers



The Scree-plot after Removing "Enterprise.value.revenue."

The boxplots in Figure 1 were produced using the following code：

```
stock_rmNA %>%pivot_longer(cols = -
Symbol,names_to='Variable',values_to='Value')%>% ggplot(aes(y=(Value)))+
geom_boxplot()+
scale_y_log10()+
facet_wrap(~Variable, nrow = 2,scale='free_y') + theme(axis.text.x=element_blank())
```

The Correlation Table in Table 1 code:

```
all.num=stock_rmNA%>%select(-"Symbol") #remove non-numerical variables
all.num.nona=all.num[complete.cases(all.num),] #remove all na
round(cor(all.num.nona),5) #compute corr table
```

The principal components analysis and biplot were produced using the following code:

```
stock_rmNA%>%filter(!Symbol %in% c("MSFT","AAPL", "RDSA.L", "LIN.DE",
"0006.HK", "1299.HK", "TSLA",
"0669.HK", "V1928.HK",
"1928.HK", "DB1.DE", "ILMN", "RTO.L", "VILMN", "1038.HK"))%>%
column_to_rownames(var="Symbol")%>% na.omit()%>%
prcomp(scale = TRUE)->pca_rmoutliers
 summary(pca_rmoutliers)
pca_rmoutliers$rotation
screeplot(pca_rmoutliers,type="lines")
#Correlation
biplot(pca_rmoutliers,scale=0, xlim=c(-4,4.5), ylim=c(-4,2.5), cex=0.5)
#Distance
biplot(pca_rmoutliers,cex=0.5)
```

The principal components analysis and biplot after removing “Enterprise.value.revenue.” were produced using the following code:

```
## Remove Enterprise.value.revenue. same direction
stock_rmsame <- stock_rmNA %>% select(-"Enterprise.value.revenue")
summary(stock_rmsame)
## pca
```

```r
stock_rmsame%>%filter((Symbol %in% c("MSFT","AAPL", "RDSA.L", "LIN.DE",
"0006.HK", "1299.HK", "TSLA",
"0669.HK", "V1928.HK",
"1928.HK", "DB1.DE", "ILMN", "RTO.L"))==FALSE)%>%
column_to_rownames(var="Symbol")%>% na.omit()%>%
prcomp(scale = TRUE)->pca_new
pca_new$rotation
summary(pca_new)
screeplot(pca_new,type="lines")
#Correlation
biplot(pca_new,scale=0,cex=0.5)
#Distance
biplot(pca_new,cex=0.5)
```