Multivariate Statistical Analysis : A4

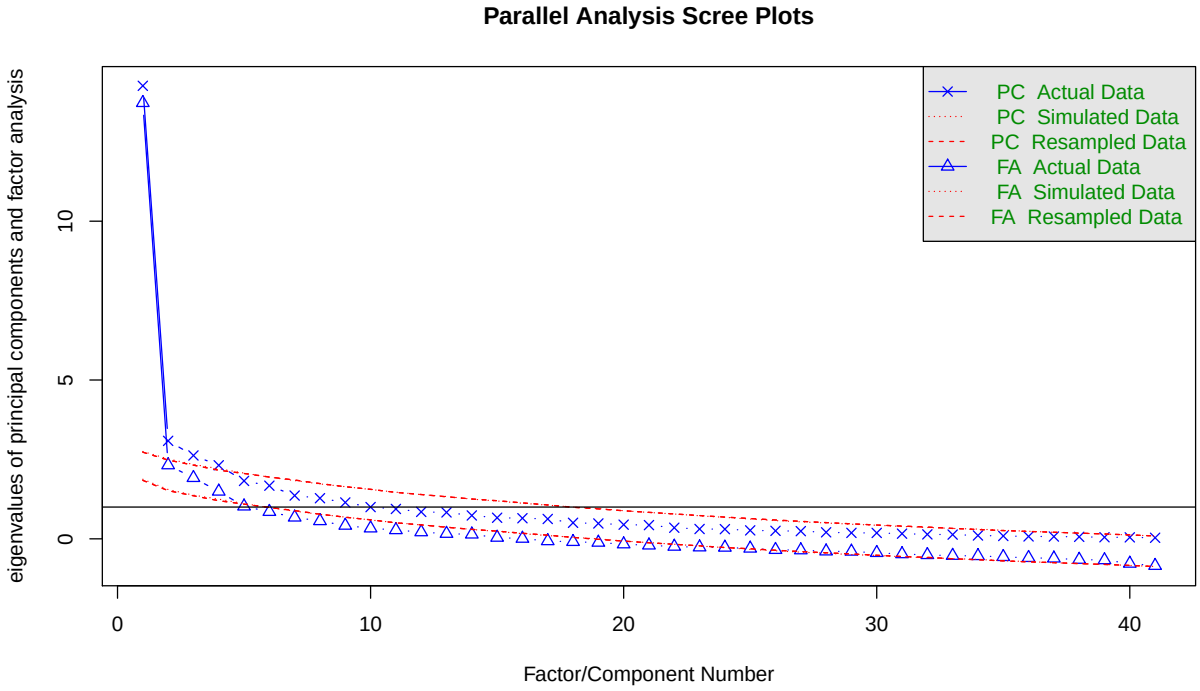
Chris Gervais (20042208)

May 3, 2017

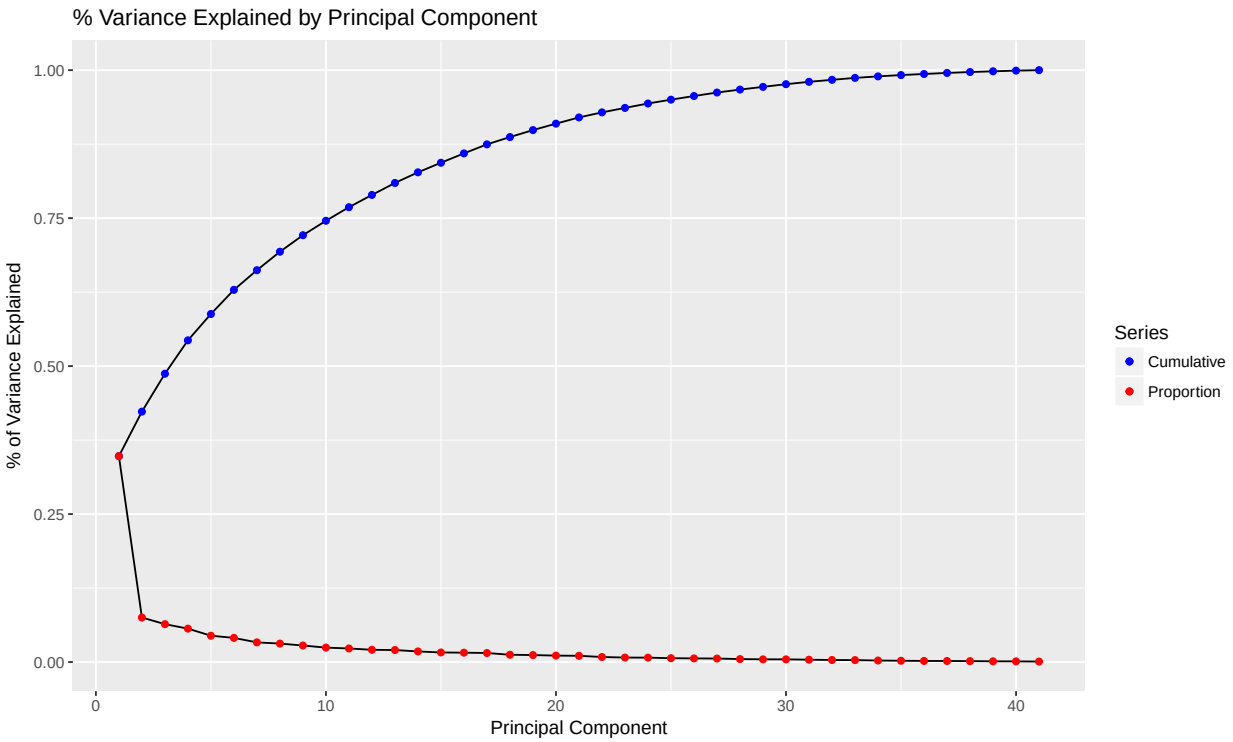
1. After completing a principal component analysis, it appears that *nine* dimensions are sufficient for capturing the most of the variation in the data. When the Eigen value for the corresponding principal component drop below 1, it means that the new principal component variable is doing a worse job explaining the variance than the original variables. The table of Eigen values, below, shows that the first *nine* of them are above zero. The corresponding scree plot supports this finding. Further, the variance plot shows that approximately 72% of the variance is accounted for in the first *nine* principal components.

Table 1: Table showing that the first nine Eigen values are above 1

Eigen Value	Value
1	14.262
2	3.084
3	2.624
4	2.315
5	1.824
6	1.673
7	1.362
8	1.279
9	1.145
10	0.999

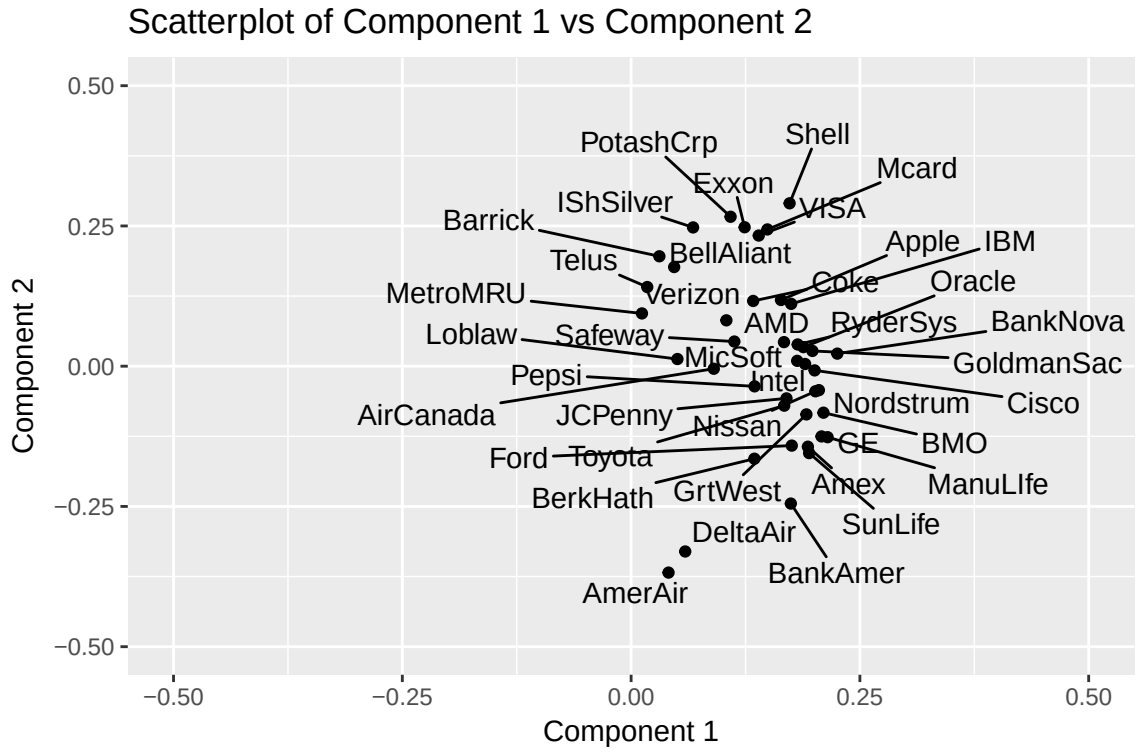


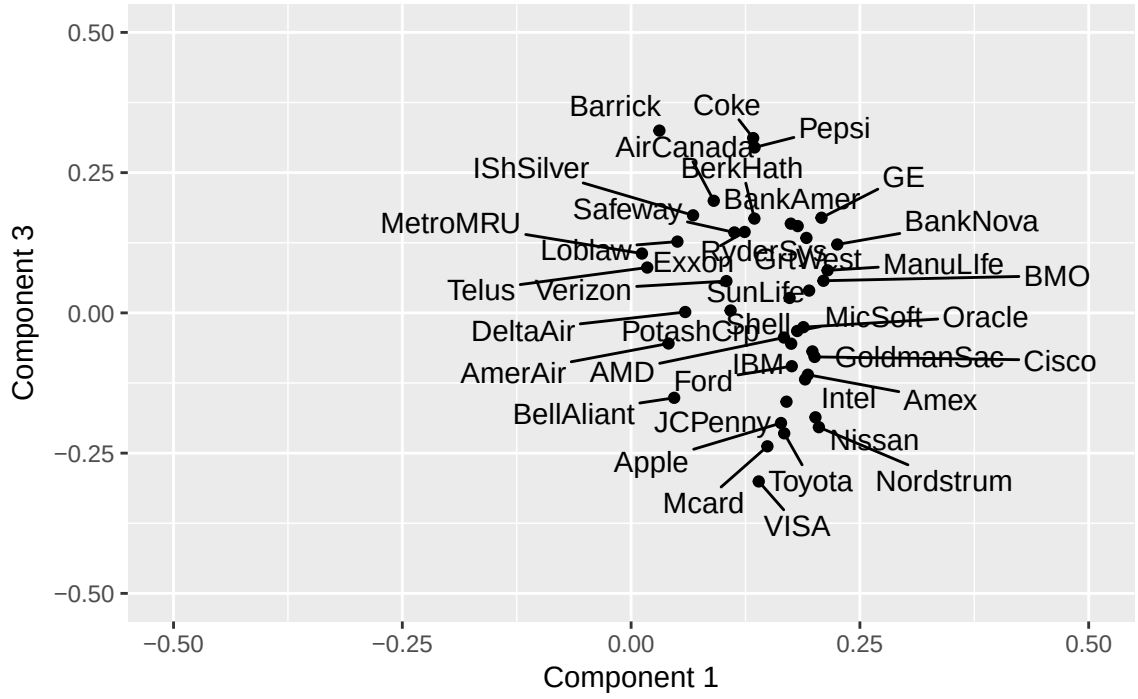
```
## Parallel analysis suggests that the number of factors =  4  and the number of components =  4
```

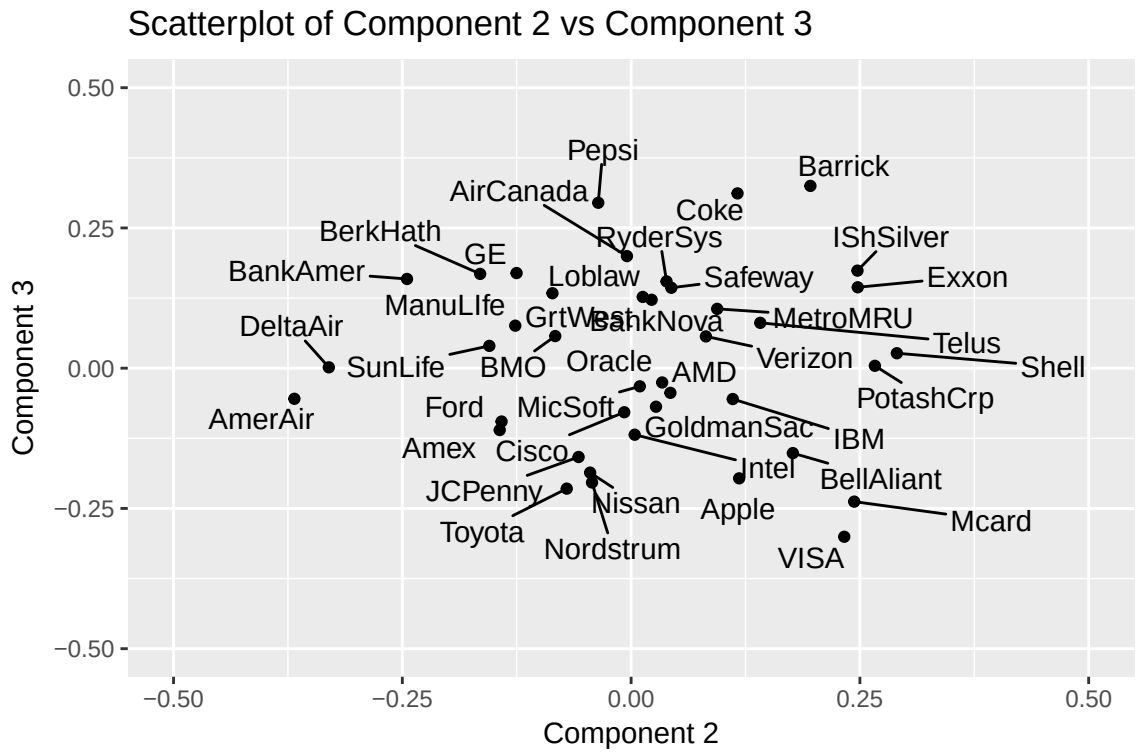


2. After creating the component pattern plots, the following groupings were observed:

- The plot of principal component 1 against component 2 appears to show a grouping of Shell and Exxon in the top right quadrant. These two may be grouped together because of their relationship to the fossil fuel industry. Also of interest is that PotashCorp is grouped in a similar area, which may suggest that the Exxon and Shell grouping generalizes to natural resource companies.
- The plot of principal component 1 against component 3 appears to show a grouping of Pepsi and Coke in the top right quadrant. This grouping may suggest that companies tied to the beverage or fast food industries move together in this dimension.
- Lastly, the plot of principal component 2 against 3 appears to show a grouping of MasterCard and VISA in the bottom left quadrant. This grouping may suggest that credit card companies move together in this dimension. Also present in this dimension is a grouping of technology companies, including IBM, Apple, and BellAliant.


Scatterplot of Component 1 vs Component 2
Component 2
Component 1
0.50
0.25
0.00
−0.25
−0.50
−0.50
−0.25
0.00
0.25
0.50
PotashCrp
Shell
Exxon
Mcard
VISA
IShSilver
Barrick
BellAliant
Apple
IBM
Telus
Coke
Oracle
MetroMRU
Verizon
AMD
RyderSys
BankNova
Loblaw
Safeway
MicSoft
GoldmanSac
Pepsi
Intel
Cisco
AirCanada
JCPenny
Nordstrum
Nissan
BMO
GE
Ford
Toyota
ManuLIfe
Amex
BerkHath
GrtWest
SunLife
DeltaAir
BankAmer
AmerAir


Scatterplot of Component 1 vs Component 3
Component 3
Component 1
0.50
0.25
0.00
−0.25
−0.50
−0.50
−0.25
0.00
0.25
0.50
Barrick
Coke
Pepsi
AirCanada
IShSilver
BerkHath
GE
BankAmer
Safeway
BankNova
MetroMRU
Loblaw
Exxon
GrtWest
ManuLIfe
BMO
Telus
Verizon
SunLife
MicSoft
Oracle
DeltaAir
PotashCrp
Shell
GoldmanSac
Cisco
AmerAir
AMD
IBM
Ford
Intel
Amex
BellAliant
JCPenny
Nissan
Apple
Toyota
Nordstrum
Mcard
VISA



3. After creating the following factor analyses (below), the Heywood condition appears to go away when *four* factors are used.

```
#Factor analysis with 6 factors:
fa_base <- fa(df_cor, fm = "mle", n.obs = 78, nfactors = 6)
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : A loading greater than abs(1) was detected. Examine the loadings
## carefully.
```

```
#Factor analysis with 5 factors:
fa_base <- fa(df_cor, fm = "mle", n.obs = 78, nfactors = 5)
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : A loading greater than abs(1) was detected. Examine the loadings
## carefully.
```

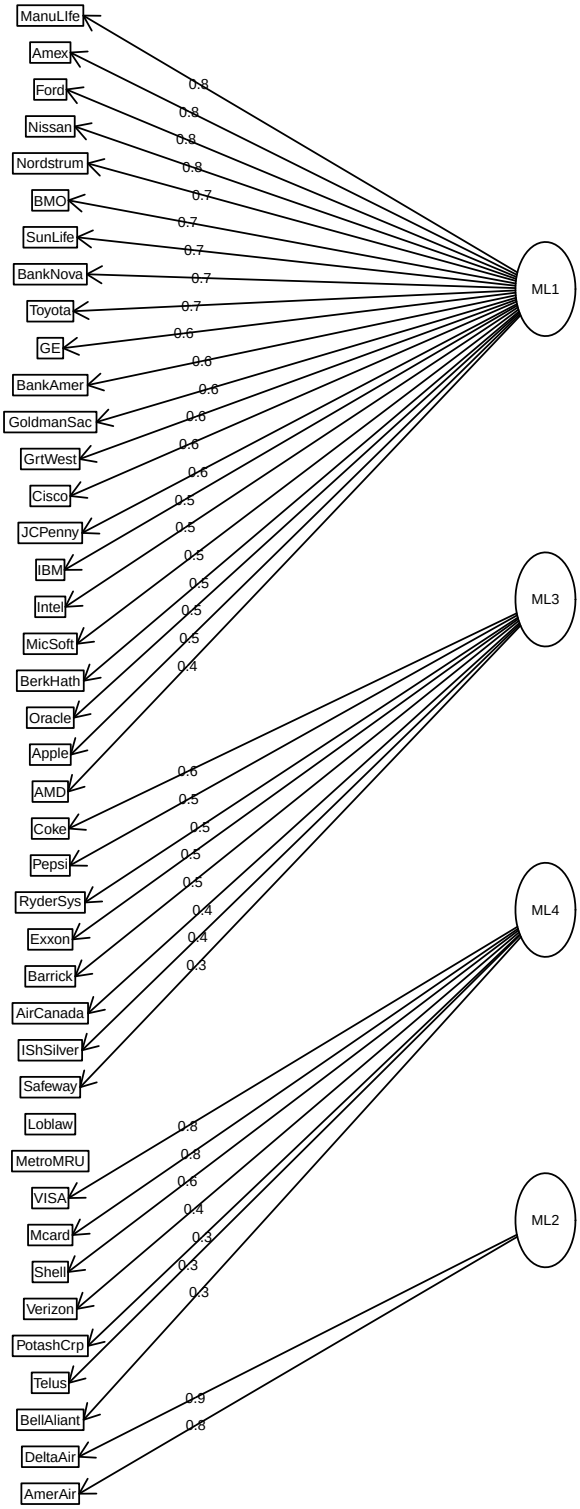
```
#Factor analysis with 4 factors:
fa_base <- fa(df_cor, fm = "mle", n.obs = 78, nfactors = 4)
```

Note that no warning was given for the case of *nfactors* = *4*, indicating that they Heywood condition is no longer present when 4 or less factors are used.

Table 2: Table of the factor loadings.

Stock	ML1	ML2	ML3	ML4
ManuLIfe	0.816	0.047	0.342	0.004
Amex	0.799	0.059	0.020	0.155
Ford	0.794	-0.035	-0.027	0.060
Nissan	0.773	0.009	0.030	0.276
Nordstrum	0.743	0.062	0.041	0.320
BMO	0.729	0.121	0.341	0.120
SunLife	0.703	0.163	0.276	0.044
BankNova	0.696	-0.001	0.510	0.197
Toyota	0.689	0.019	-0.070	0.218
GE	0.636	0.250	0.495	0.083
BankAmer	0.627	0.266	0.380	-0.128
GoldmanSac	0.621	0.108	0.233	0.301
GrtWest	0.619	0.074	0.394	0.096
Cisco	0.611	0.081	0.231	0.334
JCPenny	0.576	0.103	0.091	0.220
IBM	0.530	-0.097	0.258	0.301
Intel	0.518	0.227	0.200	0.407
MicSoft	0.511	0.046	0.291	0.271
BerkHath	0.480	0.048	0.305	-0.070
Oracle	0.470	0.193	0.332	0.379
Apple	0.470	0.004	0.114	0.450
RyderSys	0.449	0.090	0.501	0.235
AMD	0.439	0.156	0.243	0.321
VISA	0.362	-0.094	-0.144	0.808
Mcard	0.351	-0.111	-0.028	0.794
Pepsi	0.321	0.021	0.544	0.002
Shell	0.272	-0.040	0.471	0.638
Coke	0.248	-0.081	0.606	0.148
PotashCrp	0.226	-0.184	0.260	0.350
AmerAir	0.203	0.814	-0.137	-0.122
AirCanada	0.174	0.178	0.407	0.044
Safeway	0.167	0.246	0.327	0.297
Verizon	0.164	0.103	0.256	0.353
DeltaAir	0.162	0.911	-0.006	-0.027
Exxon	0.126	-0.094	0.495	0.432
IShSilver	0.087	-0.179	0.379	0.138
Loblaw	0.084	0.103	0.204	0.041
BellAliant	0.029	-0.017	0.021	0.331
Barrick	-0.013	-0.218	0.458	-0.084
MetroMRU	-0.106	-0.019	0.164	0.120
Telus	-0.235	0.220	0.240	0.343

Factor Analysis



4. Table 4 shows how each factor is loading on each stock, while the factor diagram shows the underlying cluster of stocks on each factor. From the factor diagram, we can clearly see that factor 2 is loading heavily on DeltaAir and AmerAir, suggesting that this factor may represent how airline industry stocks have performed. Factor 4 seems to have loaded heavily on VISA and Mastercard, suggesting that this factor may represent how credit card stocks have performed. Finally, factor 4 appears to be loading heavily on insurance stocks such as Manulife and SunLife, as well as car manfuacturers such as Ford, Nissan, and Toyota. However, this factor has also clustered companies from financial services and technology industries, suggesting that this factor may represent the general market trend for stocks over the time period.