fx_pca.R

brian

2021-09-21

```
# this data was downloaded from factset on 9/21/2021
# it is foreign exchange data, weekly average, over the last 5 years
# my goal is to do a PCA analysis

library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)

##### set up

# import data
fx <- read_xlsx('~/documents/data/factset_fx/fx_5yr_wkly_avg_transpose.xlsx')

# identify unwanted columns and delete
data.frame(colnames(fx))
```

```
##          colnames.fx.
## 1      Exchange Rates
## 2      Currency Pairs
## 3           Key Pairs
## 4             EUR/USD
## 5             GBP/USD
## 6             AUD/USD
## 7             USD/JPY
## 8             USD/CNY
## 9             USD/CAD
## 10            USD/CHF
## 11            EUR/GBP
## 12            EUR/CHF
## 13            EUR/JPY
## 14           Americas
## 15            USD/MXN
```

```
## 16               USD/BRL
## 17               USD/ARS
## 18               USD/CLP
## 19               USD/COP
## 20               USD/PEN
## 21          Asia-Pacific
## 22               AUD/JPY
## 23               AUD/CAD
## 24               EUR/AUD
## 25               EUR/NZD
## 26               NZD/USD
## 27               USD/HKD
## 28               USD/IDR
## 29               USD/INR
## 30               USD/KRW
## 31               USD/MYR
## 32               USD/PHP
## 33               USD/SGD
## 34               USD/THB
## 35               USD/TWD
## 36                Europe
## 37               EUR/CZK
## 38               EUR/DKK
## 39               EUR/HRK
## 40               EUR/HUF
## 41               EUR/NOK
## 42               EUR/PLN
## 43               EUR/RON
## 44               EUR/RSD
## 45               EUR/RUB
## 46               EUR/SEK
## 47               EUR/TRY
## 48               GBP/CHF
## 49               USD/RUB
## 50               USD/SEK
## 51               USD/TRY
## 52               USD/UAH
## 53 Middle East & Africa
## 54               USD/ILS
## 55               USD/ZAR
## 56               EUR/ZAR
```

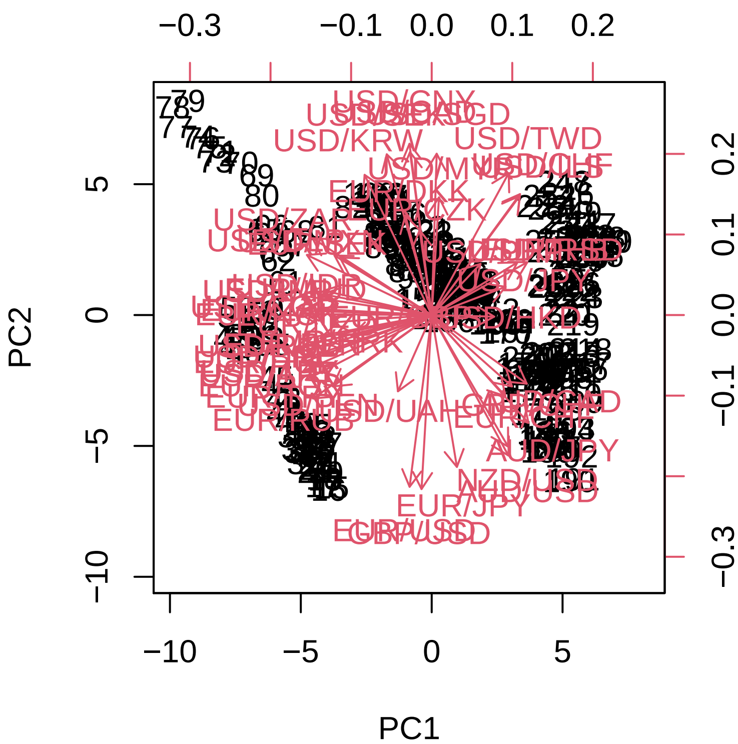
```
fx <- fx[,-c(1:3,14,21,36,53)]

# change data into numeric. This seems roundabout, but it works and every
# simpler solution I tried did not work

# create empty matrix
efex <- matrix(data = NA, nrow = dim(fx)[1], ncol = dim(fx)[2])
# add colnames
colnames(efex) <- colnames(fx)
# change each cell of fx into numeric and put in new matrix
for (i in 1:dim(fx)[2]) {
  efex[,i] <- c(as.numeric(fx[[i]]))
```

```
}

##### pca
pr.out <- prcomp(efex, scale = TRUE)
biplot(pr.out, scale=0)
```




```
pr.out$sdev
```

```
##  [1] 4.560322987 3.436203323 2.682905251 1.565076250 1.443134355 0.999352794
##  [7] 0.937077763 0.676391660 0.636878660 0.550080666 0.492198413 0.442694543
## [13] 0.416197347 0.370721763 0.361162850 0.345688884 0.321639114 0.280221198
## [19] 0.254478079 0.226265557 0.217970868 0.200178989 0.196984836 0.178197396
## [25] 0.160108261 0.152593637 0.150233434 0.136814778 0.124779068 0.109502683
## [31] 0.104332933 0.091858579 0.090456166 0.083255987 0.078728281 0.070832029
## [37] 0.058885534 0.022459179 0.019414305 0.015814503 0.011560132 0.010179267
## [43] 0.006952282 0.006720919 0.005802757 0.005241789 0.003974872 0.003404763
## [49] 0.002826015
```

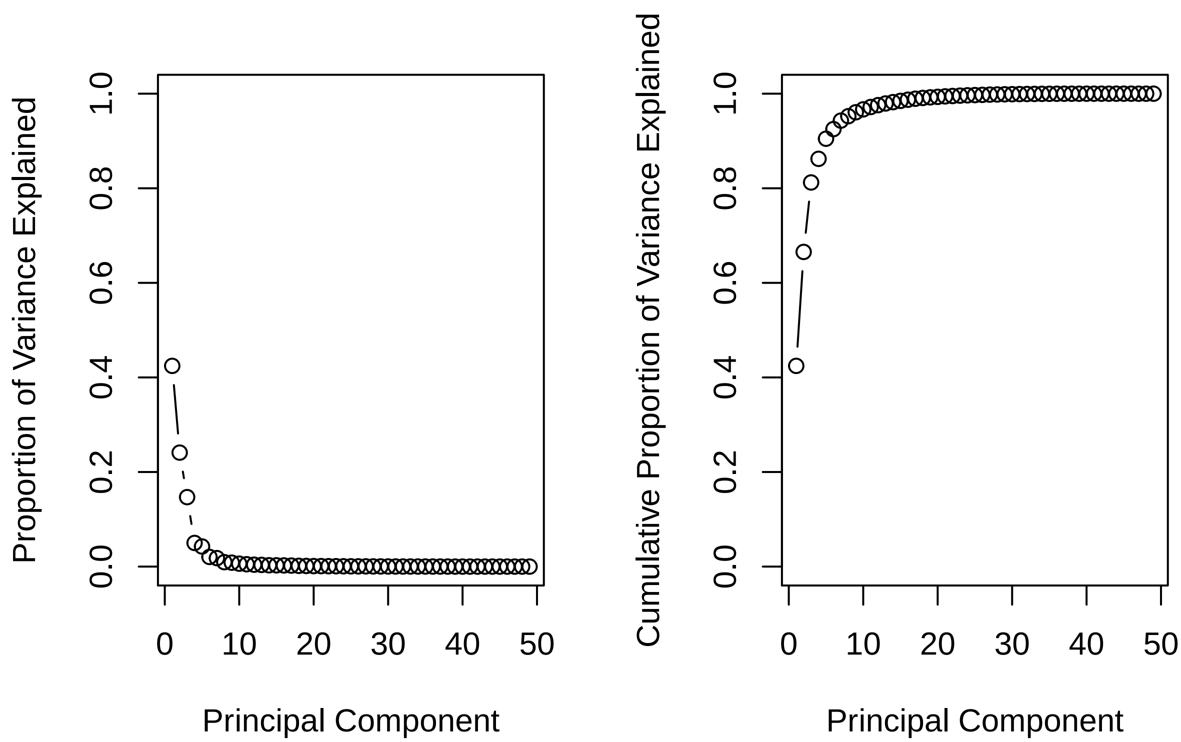
```
# proportion of variance explained
pr.var <- pr.out$sdev^2
pve <- pr.var/sum(pr.var)
pve
```

```
##  [1] 4.244193e-01 2.409693e-01 1.468976e-01 4.998905e-02 4.250279e-02
##  [6] 2.038176e-02 1.792071e-02 9.336851e-03 8.277845e-03 6.175280e-03
## [11] 4.944067e-03 3.999560e-03 3.535107e-03 2.804788e-03 2.662012e-03
## [16] 2.438792e-03 2.111260e-03 1.602529e-03 1.321614e-03 1.044818e-03
```

```
## [21] 9.696184e-04 8.177883e-04 7.918985e-04 6.480472e-04 5.231562e-04
## [26] 4.752004e-04 4.606140e-04 3.820058e-04 3.177513e-04 2.447110e-04
## [31] 2.221502e-04 1.722041e-04 1.669861e-04 1.414604e-04 1.264927e-04
## [36] 1.023914e-04 7.076543e-05 1.029418e-05 7.692147e-06 5.104051e-06
## [41] 2.727279e-06 2.114643e-06 9.864129e-07 9.218520e-07 6.871834e-07
## [46] 5.607420e-07 3.224410e-07 2.365799e-07 1.629870e-07
```

```r
# 42.4% variance explained by the first principal component

par(mfrow = c(1, 2))
plot(pve, xlab = "Principal Component",
    ylab = "Proportion of Variance Explained", ylim = c(0, 1),
    type = "b")
plot(cumsum(pve), xlab = "Principal Component",
    ylab = "Cumulative Proportion of Variance Explained",
    ylim = c(0, 1), type = "b")
```




```r
# you can explain over 81% of the variance in the forex market over the last
# 5 years with the first 3 principal components, and over 86% with 4

sum(pve[1:3])
```

```
## [1] 0.8122861
```

```r
sum(pve[1:4])
```

```
## [1] 0.8622752
```