20181108 PCA with FactoMineR and factoextra (for price change)

cheungngo

8 November 2018

library(FactoMineR)  
library(factoextra)

### Loading the data

library(readr)  
X20181108\_pricechange <- read\_csv("20181108-pricechange.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## Names = col\_character()  
## )

## See spec(...) for full column specifications.

View(X20181108\_pricechange)

### Tidying the dataset

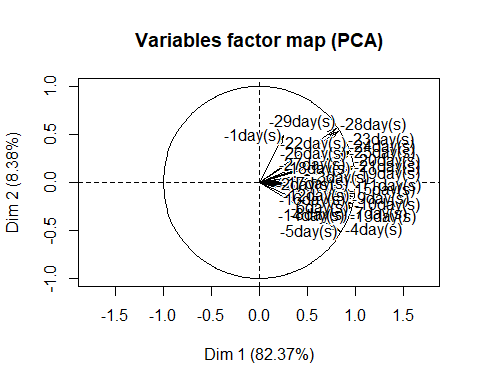
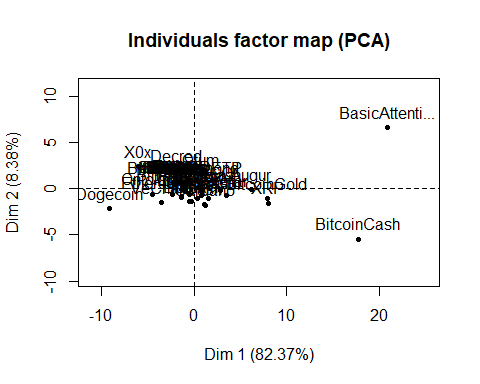
CMC = X20181108\_pricechange  
CMC = as.data.frame(CMC) # it became tibble when being read from csv  
rownames(CMC) = sapply(CMC[,1],as.character)  
CMC = CMC[,-1]  
head(CMC)

## -1day(s) -2day(s) -3day(s) -4day(s) -5day(s) -6day(s)  
## Bitcoin 1.0699566 1.720963 2.415415 2.654820 2.218069 2.388919  
## Ethereum -0.5813687 3.869147 4.670105 8.486937 8.249016 9.207020  
## XRP -2.4355285 7.698849 15.780726 17.565459 16.791916 17.551270  
## BitcoinCash -1.7915387 10.525185 10.187797 29.134501 33.601004 45.546936  
## EOS -1.3961606 3.290676 3.669725 6.403013 5.410448 7.007576  
## Stellar -3.3426708 4.132982 6.240740 8.914495 9.594269 15.148579  
## -7day(s) -8day(s) -9day(s) -10day(s) -11day(s) -12day(s)  
## Bitcoin 3.364089 3.092227 3.118925 0.6744892 0.7678562 0.855477  
## Ethereum 10.031411 9.931160 10.103929 5.7505965 6.3513050 6.811587  
## XRP 19.475581 20.448827 21.334091 15.8520867 17.3900181 16.638580  
## BitcoinCash 45.970297 46.292039 47.819527 39.9406003 41.0760405 40.686967  
## EOS 8.653846 9.922179 9.708738 4.0515654 5.0185874 4.823748  
## Stellar 14.877546 14.049748 15.453078 10.2404335 12.2215422 9.703390  
## -13day(s) -14day(s) -15day(s) -16day(s) -17day(s)  
## Bitcoin 0.8314946 0.5280302 0.8400584 0.6625395 0.7372326  
## Ethereum 7.1329913 6.5391219 6.2836449 6.4399137 5.8691625  
## XRP 16.9533153 17.1617782 15.9393409 18.2291678 17.3010812  
## BitcoinCash 40.2107989 39.5072890 39.4442562 38.0224056 37.8251647  
## EOS 5.2141527 5.0185874 4.2435424 5.0185874 4.2435424  
## Stellar 8.9422269 7.9735868 5.9876556 5.9238079 6.2029422  
## -18day(s) -19day(s) -20day(s) -21day(s) -22day(s) -23day(s)  
## Bitcoin 0.6310495 1.001174 0.8249559 -0.2183536 -1.000135 -1.006588  
## Ethereum 5.7197099 6.601875 6.8010819 4.8773421 3.359985 3.567000  
## XRP 16.6733820 17.950892 17.0368671 14.4191441 13.918890 19.760304  
## BitcoinCash 37.9514572 40.013610 40.7864425 37.5364870 34.767800 34.365884  
## EOS 5.2141527 6.003752 5.6074766 4.4362292 3.860294 3.290676  
## Stellar 5.1338756 5.709215 7.2106309 6.7865210 12.606480 15.119663  
## -24day(s) -25day(s) -26day(s) -27day(s) -28day(s) -29day(s)  
## Bitcoin 3.802458 3.884034 4.072942 4.378029 -0.8410864 -1.693604  
## Ethereum 10.964643 8.676942 10.394958 14.606860 -3.8047571 -4.737258  
## XRP 32.835506 27.890937 27.818618 39.823820 15.8568472 11.638884  
## BitcoinCash 40.530018 37.653041 39.261783 42.538795 20.3404040 19.208559  
## EOS 9.284333 7.619048 9.073359 11.220472 -3.9115646 -4.399323  
## Stellar 22.162898 19.509334 18.833321 23.996194 6.2605291 5.248960

## FactomineR

### PCA()

pca\_1108 = PCA(CMC,ncp=29)



### Variance and cummulative var in the PCA

pca\_1108$eig[,2] # variance for each component

## comp 1 comp 2 comp 3 comp 4 comp 5   
## 8.237369e+01 8.384570e+00 4.028471e+00 2.385890e+00 9.917854e-01   
## comp 6 comp 7 comp 8 comp 9 comp 10   
## 6.372392e-01 2.662147e-01 2.262284e-01 1.687502e-01 1.478132e-01   
## comp 11 comp 12 comp 13 comp 14 comp 15   
## 1.060866e-01 7.837573e-02 4.811112e-02 3.687294e-02 2.905720e-02   
## comp 16 comp 17 comp 18 comp 19 comp 20   
## 2.559424e-02 1.533794e-02 1.521996e-02 1.127710e-02 7.034065e-03   
## comp 21 comp 22 comp 23 comp 24 comp 25   
## 5.580934e-03 4.729846e-03 2.017318e-03 1.557455e-03 1.049307e-03   
## comp 26 comp 27 comp 28 comp 29   
## 8.911128e-04 4.036152e-04 1.343131e-04 1.270525e-05

pca\_1108$eig[,3] # cummulative variance for each component

## comp 1 comp 2 comp 3 comp 4 comp 5 comp 6 comp 7   
## 82.37369 90.75826 94.78673 97.17263 98.16441 98.80165 99.06786   
## comp 8 comp 9 comp 10 comp 11 comp 12 comp 13 comp 14   
## 99.29409 99.46284 99.61066 99.71674 99.79512 99.84323 99.88010   
## comp 15 comp 16 comp 17 comp 18 comp 19 comp 20 comp 21   
## 99.90916 99.93475 99.95009 99.96531 99.97659 99.98362 99.98920   
## comp 22 comp 23 comp 24 comp 25 comp 26 comp 27 comp 28   
## 99.99393 99.99595 99.99751 99.99856 99.99945 99.99985 99.99999   
## comp 29   
## 100.00000

### dimdesc()

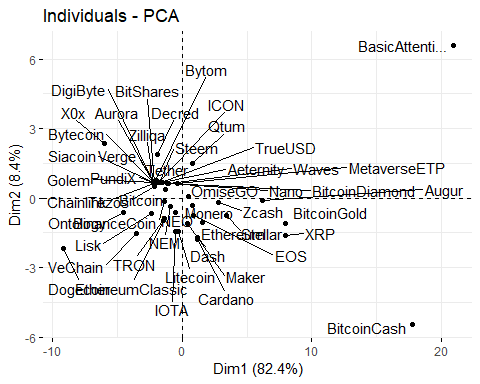
dimdesc(pca\_1108, axes = 1:2) # to get the most correlated variables

## $Dim.1  
## $Dim.1$quanti  
## correlation p.value  
## -11day(s) 0.9796779 3.776748e-35  
## -20day(s) 0.9774735 4.363902e-34  
## -19day(s) 0.9746592 7.132839e-33  
## -18day(s) 0.9721901 6.462180e-32  
## -10day(s) 0.9720897 7.037955e-32  
## -21day(s) 0.9692475 6.988952e-31  
## -12day(s) 0.9618705 1.121010e-28  
## -9day(s) 0.9546552 6.617803e-27  
## -6day(s) 0.9514830 3.235605e-26  
## -17day(s) 0.9474145 2.134632e-25  
## -7day(s) 0.9411487 2.963711e-24  
## -8day(s) 0.9392188 6.289788e-24  
## -13day(s) 0.9355185 2.490911e-23  
## -22day(s) 0.9288083 2.483172e-22  
## -26day(s) 0.9266479 4.964604e-22  
## -15day(s) 0.9258308 6.416299e-22  
## -25day(s) 0.9250103 8.277352e-22  
## -24day(s) 0.9222348 1.918220e-21  
## -27day(s) 0.9220353 2.035262e-21  
## -16day(s) 0.9177608 6.977102e-21  
## -23day(s) 0.9119609 3.350067e-20  
## -14day(s) 0.9045470 2.141247e-19  
## -4day(s) 0.8878427 8.461631e-18  
## -3day(s) 0.8617097 9.497812e-16  
## -5day(s) 0.8429247 1.616469e-14  
## -28day(s) 0.8309945 8.123741e-14  
## -29day(s) 0.8103389 1.008493e-12  
## -2day(s) 0.7932167 6.519873e-12  
##   
##   
## $Dim.2  
## $Dim.2$quanti  
## correlation p.value  
## -29day(s) 0.5610226 2.255571e-05  
## -28day(s) 0.5260101 8.740014e-05  
## -1day(s) 0.4863603 3.421779e-04  
## -23day(s) 0.3730289 7.629147e-03  
## -22day(s) 0.3248582 2.134566e-02  
## -25day(s) 0.3155712 2.559057e-02  
## -26day(s) 0.3010815 3.361212e-02  
## -24day(s) 0.2977815 3.570397e-02  
## -4day(s) -0.4076397 3.299769e-03  
## -5day(s) -0.5082559 1.644110e-04

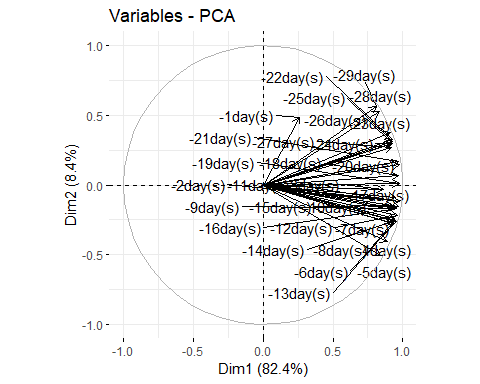
## factoextra

### fviz\_pca\_ind() and fviz\_pca\_var()

fviz\_pca\_ind(pca\_1108, repel = T)

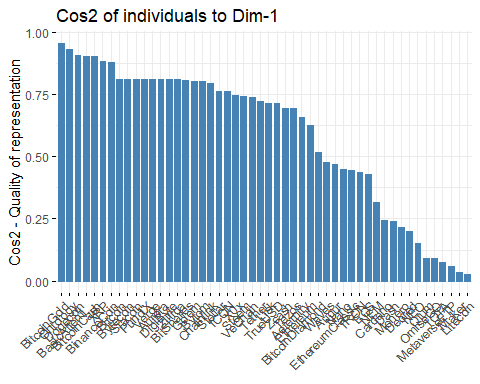


fviz\_pca\_var(pca\_1108, repel = T)

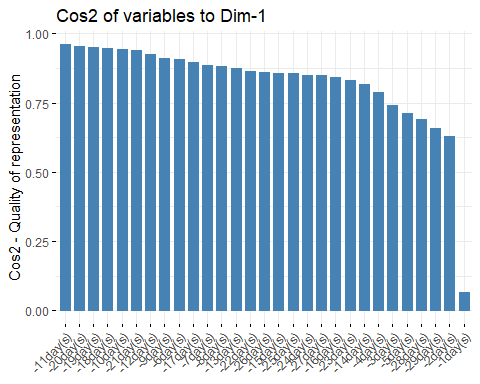


### Selecting the individuals and variables based on the quality of representation (cos2) and contributions(contrib)

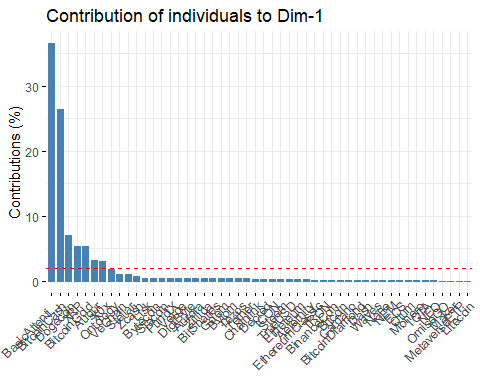
# Visualizing first  
fviz\_cos2(pca\_1108, choice = 'ind')



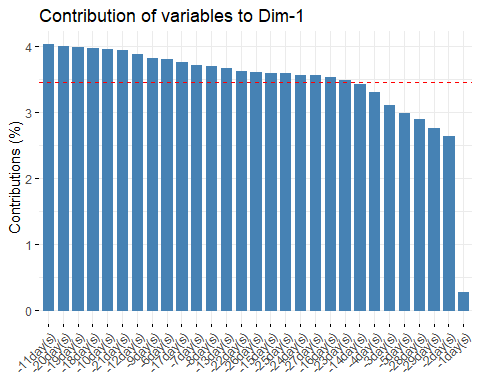
fviz\_cos2(pca\_1108, choice = 'var')



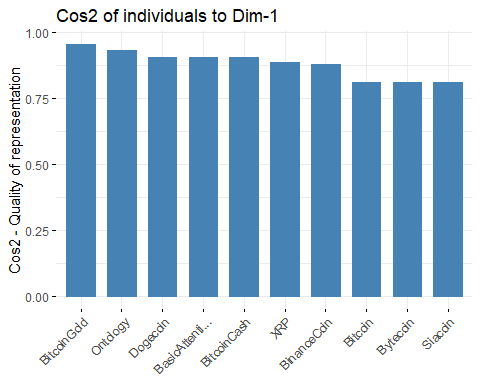
fviz\_contrib(pca\_1108, choice = 'ind')



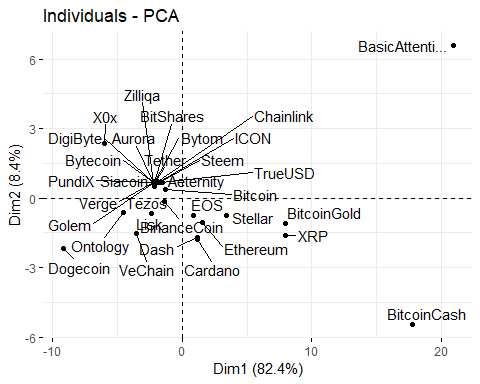
fviz\_contrib(pca\_1108, choice = 'var')



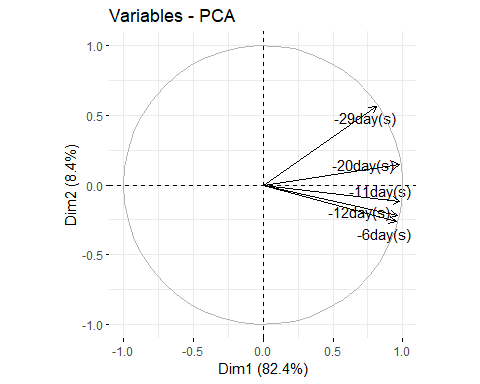
# We can also add the top =   
fviz\_cos2(pca\_1108, choice = 'ind', top = 10)



# Now selecting the ind and var based on cos2 and contrib  
fviz\_pca\_ind(pca\_1108, select.ind = list(cos2 = 0.7), repel = T)

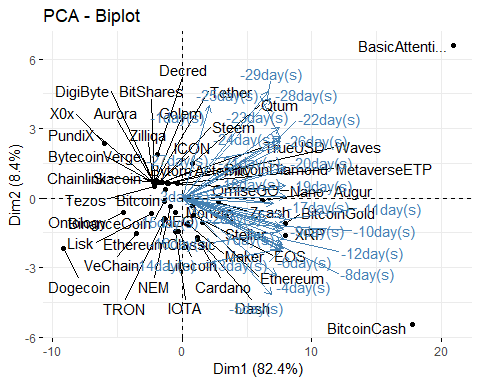


fviz\_pca\_var(pca\_1108, select.var = list(cos2 = 0.7, contrib = 5), repel = T) # plotting the top 5 contributing variables



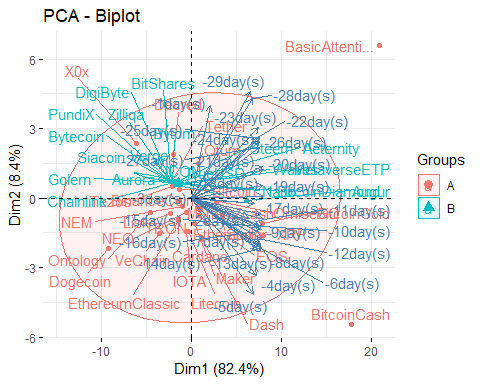
### biplots

fviz\_pca\_biplot(pca\_1108,  
 repel = T)



### biplots with epsilloids

n = length(CMC[,1])  
g = as.factor(c(rep('A',30),rep('B',20)))  
CMC$g = g  
fviz\_pca\_biplot(pca\_1108, axes = c(1,2),  
 habillage = CMC$g,  
 addEllipses = T,  
 repel = T)



fviz\_pca\_biplot(pca\_1108, axes = c(1,2),  
 habillage = CMC$g,  
 addEllipses = T,  
 repel = T,  
 select.ind = list(contrib=10))

## Too few points to calculate an ellipse

