Part1.R

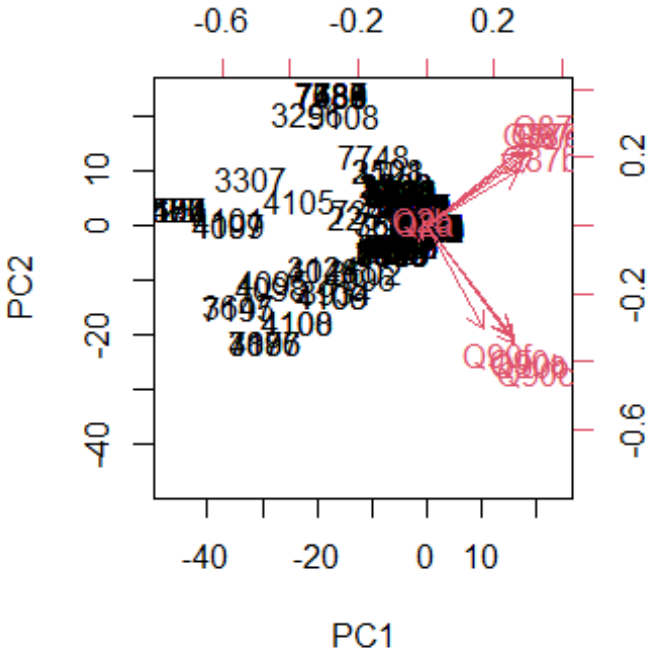
Cheng Jun

2021-03-26

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.4

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.0.4

## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

setwd("C:/Users/Cheng Jun/Desktop/SIM/Year 2/Machine Learning/Coursework")

data <- read.table("EWCS_2016.csv",header=1,sep = ',')
apply(data, 2,mean)

##        Q2a        Q2b       Q87a       Q87b       Q87c       Q87d
Q87e
##  1.3629848 41.7998208 -1.1526942 -0.5908102 -1.1640855 -0.6082171 -
3.4803533
##       Q90a       Q90b       Q90c       Q90f
## -1.9700499 -1.9015743 -1.1498784 -7.9462434

apply(data, 2,var)

##       Q2a       Q2b      Q87a      Q87b      Q87c      Q87d      Q87e
Q90a
##  128.3675 1681.0429 3582.8905 3201.7705 3582.8888 3330.2296 5871.7074
4089.4287
##      Q90b      Q90c      Q90f
## 4090.3005 3325.9265 9393.2895

data1 <- prcomp(data, center = TRUE, scale = TRUE)
data1$rotation

##                PC1          PC2           PC3          PC4          PC5
## Q2a   0.0006597898 -0.004501054 -0.7602485978  0.649595649 -0.004750643
## Q2b   0.0108053536 -0.019844962 -0.6492810942 -0.760092279 -0.008699139
## Q87a -0.3698905892 -0.268221531  0.0021831031  0.002119483  0.085932303
## Q87b -0.3404492936 -0.197076175  0.0016495101  0.001677391 -0.080105482
## Q87c -0.3684746170 -0.293726740  0.0011938530 -0.001304202 -0.057777415
## Q87d -0.3870423751 -0.268351596 -0.0011356811 -0.001233679  0.074497495
## Q87e -0.3454992520 -0.269425447  0.0051381984  0.001344620 -0.157493756
## Q90a -0.3047141808  0.395701669 -0.0078205459 -0.003165528  0.387724631
```

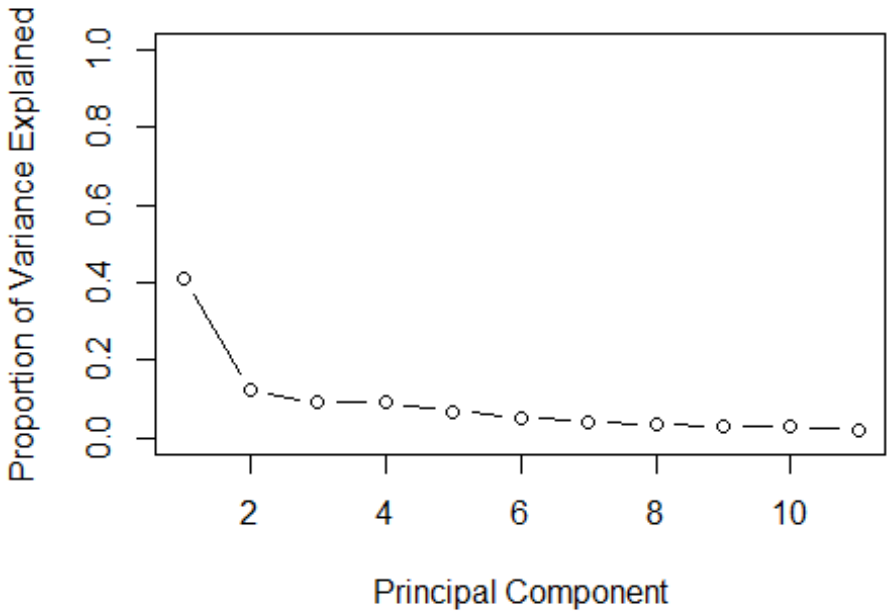
```
## Q90b -0.3128916187  0.409813689 -0.0109922309 -0.010242169  0.122374436
## Q90c -0.3260685016  0.434453678 -0.0153893547 -0.012401959  0.211288775
## Q90f -0.2124626982  0.380327480  0.0008396794 -0.002037385 -0.861660768
##                PC6           PC7           PC8           PC9          PC10
## Q2a   0.0001720072 -0.0014939129 -0.0003857357 -0.001052020  0.000123465
## Q2b  -0.0026682801  0.0007190446  0.0082321447 -0.002720565 -0.004294655
## Q87a -0.3448881456  0.2096985172 -0.1535163050  0.300316510 -0.524228007
## Q87b  0.6711584262  0.2602665045  0.4589842585  0.298622787 -0.081957732
## Q87c  0.0631005838  0.3108542910 -0.2294709436 -0.401063361  0.604532318
## Q87d -0.2558925656  0.0697267572 -0.2080490366 -0.112225524 -0.116057810
## Q87e -0.0580929708 -0.8404484970  0.2242317781 -0.022891136  0.090060853
## Q90a -0.2962232306  0.1242284200  0.6109788795 -0.348091877 -0.013108733
## Q90b  0.4766457623 -0.1954018163 -0.4405466554 -0.348364082 -0.364517171
## Q90c -0.0690894095 -0.0761802474 -0.2211571604  0.635148483  0.441165687
## Q90f -0.1950414597  0.1449870425  0.0564803974 -0.026766429 -0.061139189
##               PC11
## Q2a  -0.0008753406
## Q2b  -0.0025067417
## Q87a -0.4820279837
## Q87b  0.1186649160
## Q87c -0.3082686678
## Q87d  0.7955763452
## Q87e -0.1206829817
## Q90a -0.0484974083
## Q90b -0.0725646727
## Q90c  0.0486592456
## Q90f  0.0322425723

dim(data1$x)

## [1] 7813   11

data1$rotation = -data1$rotation
data1$rotation

##                PC1           PC2           PC3           PC4           PC5
## Q2a  -0.0006597898  0.004501054  0.7602485978 -0.649595649  0.004750643
## Q2b  -0.0108053536  0.019844962  0.6492810942  0.760092279  0.008699139
## Q87a  0.3698905892  0.268221531 -0.0021831031 -0.002119483 -0.085932303
## Q87b  0.3404492936  0.197076175 -0.0016495101 -0.001677391  0.080105482
## Q87c  0.3684746170  0.293726740 -0.0011938530  0.001304202  0.057777415
## Q87d  0.3870423751  0.268351596  0.0011356811  0.001233679 -0.074497495
## Q87e  0.3454992520  0.269425447 -0.0051381984 -0.001344620  0.157493756
## Q90a  0.3047141808 -0.395701669  0.0078205459  0.003165528 -0.387724631
## Q90b  0.3128916187 -0.409813689  0.0109922309  0.010242169 -0.122374436
## Q90c  0.3260685016 -0.434453678  0.0153893547  0.012401959 -0.211288775
## Q90f  0.2124626982 -0.380327480 -0.0008396794  0.002037385  0.861660768
##                PC6           PC7           PC8           PC9          PC10
## Q2a  -0.0001720072  0.0014939129  0.0003857357  0.001052020 -0.000123465
## Q2b   0.0026682801 -0.0007190446 -0.0082321447  0.002720565  0.004294655
## Q87a  0.3448881456 -0.2096985172  0.1535163050 -0.300316510  0.524228007
```

```
## Q87b -0.6711584262 -0.2602665045 -0.4589842585 -0.298622787  0.081957732
## Q87c -0.0631005838 -0.3108542910  0.2294709436  0.401063361 -0.604532318
## Q87d  0.2558925656 -0.0697267572  0.2080490366  0.112225524  0.116057810
## Q87e  0.0580929708  0.8404484970 -0.2242317781  0.022891136 -0.090060853
## Q90a  0.2962232306 -0.1242284200 -0.6109788795  0.348091877  0.013108733
## Q90b -0.4766457623  0.1954018163  0.4405466554  0.348364082  0.364517171
## Q90c  0.0690894095  0.0761802474  0.2211571604 -0.635148483 -0.441165687
## Q90f  0.1950414597 -0.1449870425 -0.0564803974  0.026766429  0.061139189
##               PC11
## Q2a   0.0008753406
## Q2b   0.0025067417
## Q87a  0.4820279837
## Q87b -0.1186649160
## Q87c  0.3082686678
## Q87d -0.7955763452
## Q87e  0.1206829817
## Q90a  0.0484974083
## Q90b  0.0725646727
## Q90c -0.0486592456
## Q90f -0.0322425723

data1$x =- data1$x
biplot(data1,scale=0)
```

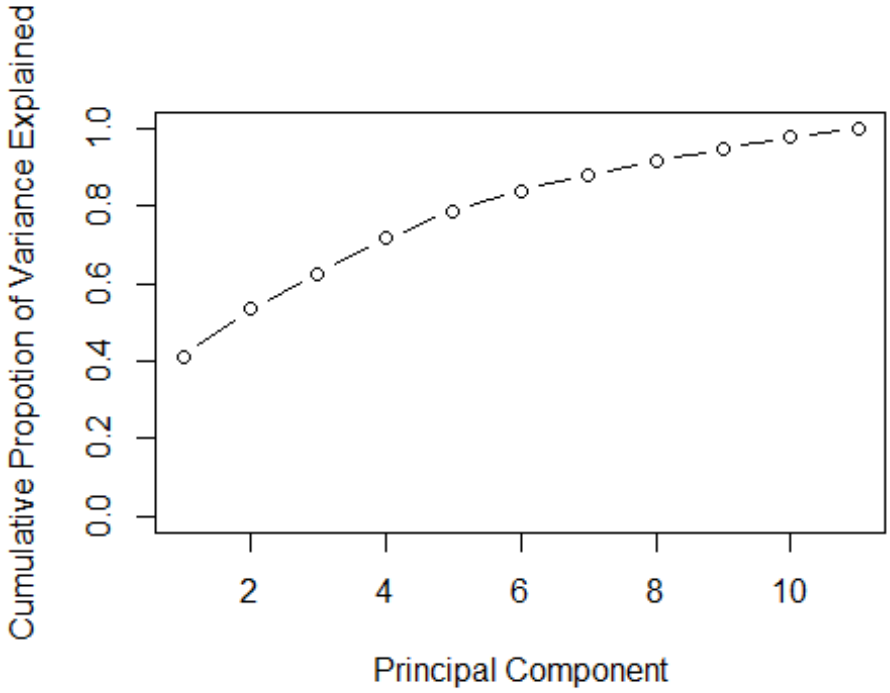



```
data1$sdev
```

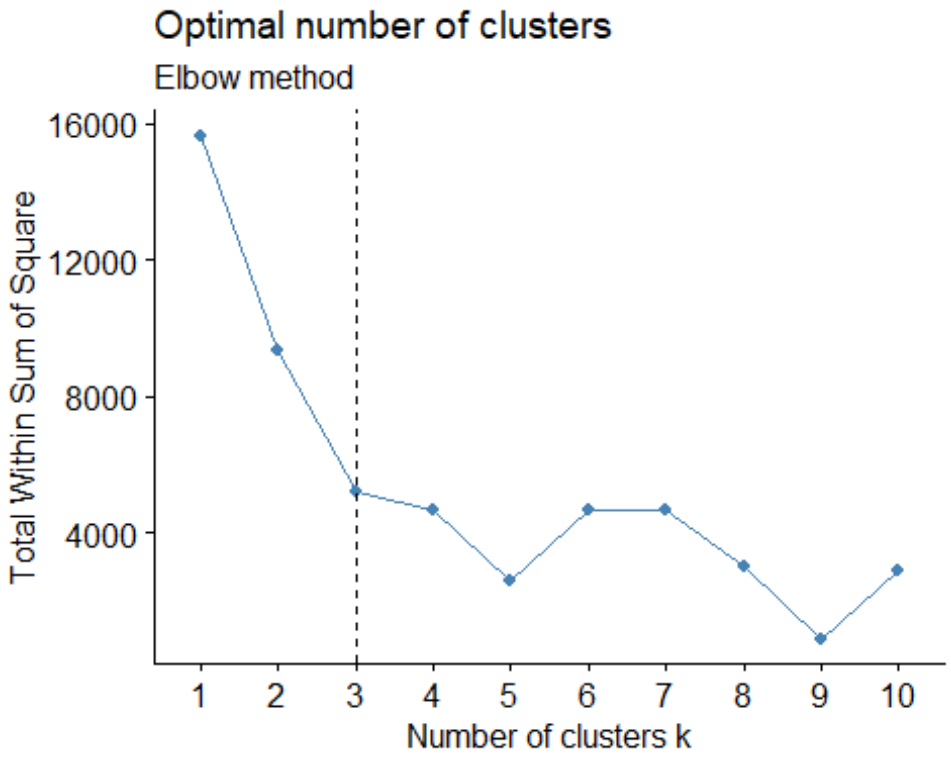
```
##  [1] 2.1284504 1.1669697 1.0006353 0.9991283 0.8713662 0.7680986 0.6653858
##  [8] 0.6373078 0.5821910 0.5754523 0.4900070

data1.var=data1$sdev^2
data1.var

##  [1] 4.5303010 1.3618182 1.0012710 0.9982574 0.7592790 0.5899755 0.4427382
##  [8] 0.4061613 0.3389464 0.3311453 0.2401068

data1.pve=data1.var/sum(data1.var)
data1.pve

##  [1] 0.41184555 0.12380165 0.09102463 0.09075067 0.06902536 0.05363413
##  [7] 0.04024893 0.03692375 0.03081331 0.03010412 0.02182789

plot(data1.pve,xlab="Principal Component",ylab="Proportion of Variance
Explained",ylim=c(0,1),type='b')
```



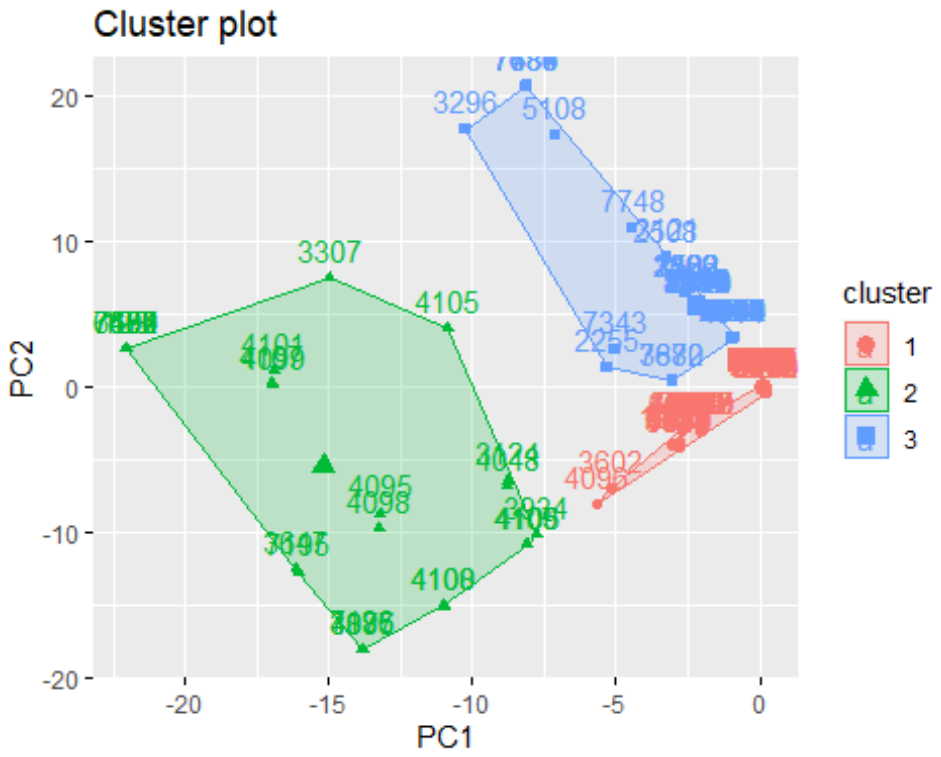

```
plot(cumsum(data1.pve),xlab="Principal Component",ylab="Cumulative Propotion
of Variance Explained",ylim=c(0,1),type='b')
```




```r
#clustering
set.seed(2)
x=matrix(rnorm(50*2),ncol=2)
x = data1$x[,c(1:2)]
x.scaled<- scale(x[,-11])
fviz_nbclust(x.scaled, kmeans, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2)+
  labs(subtitle = "Elbow method")
```




```r
#optimal number of cluster=3
km.out=kmeans(x,3)
km.out$cluster

#cluster plot
fviz_cluster(km.out,data=x)
```




```
km.out$centers

##           PC1         PC2
## 1   0.1631979 -0.05370104
## 2 -32.3041688 -6.42690927
## 3  -4.6247323  6.44929481
```