Homework 6

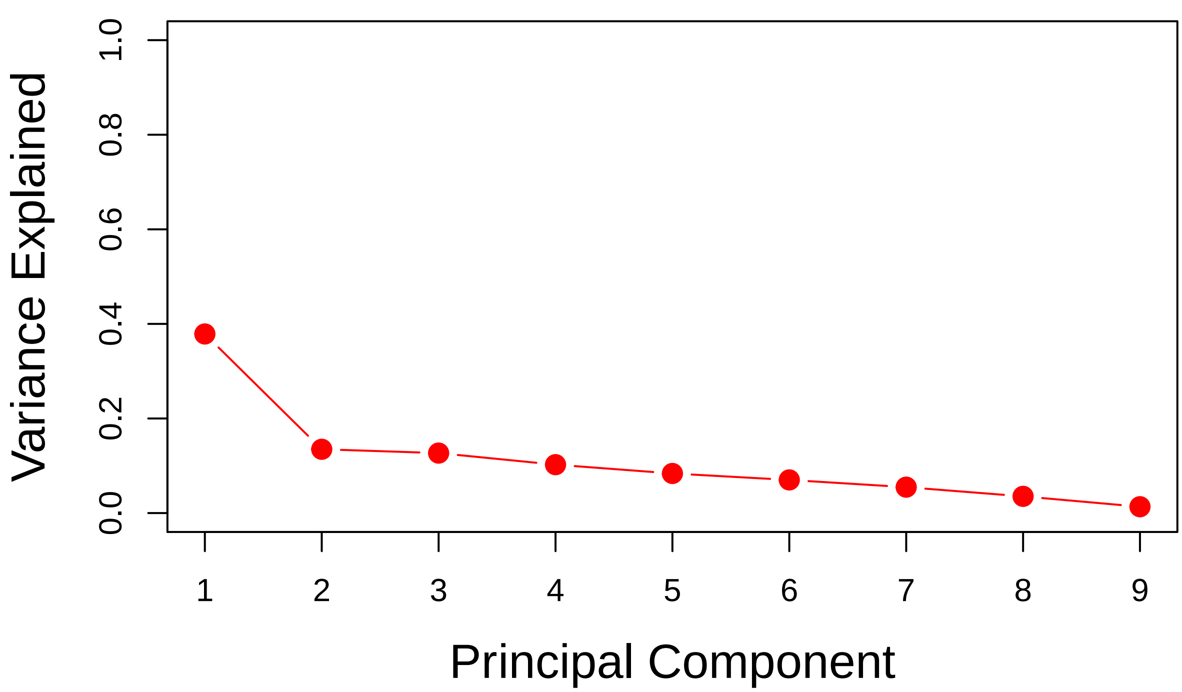
Joshua Oswari - A14751270

5/17/2019

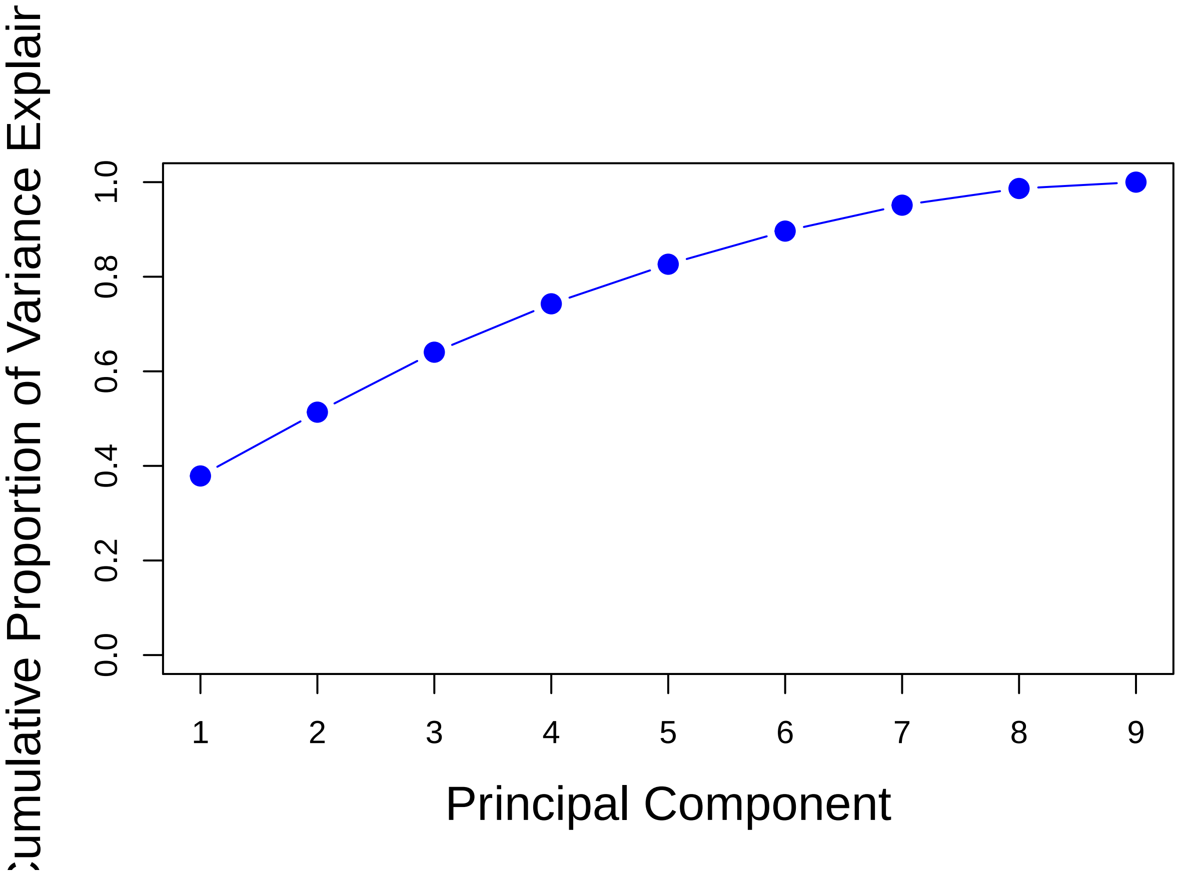
Problem 1

=============================== Part 1 ================================

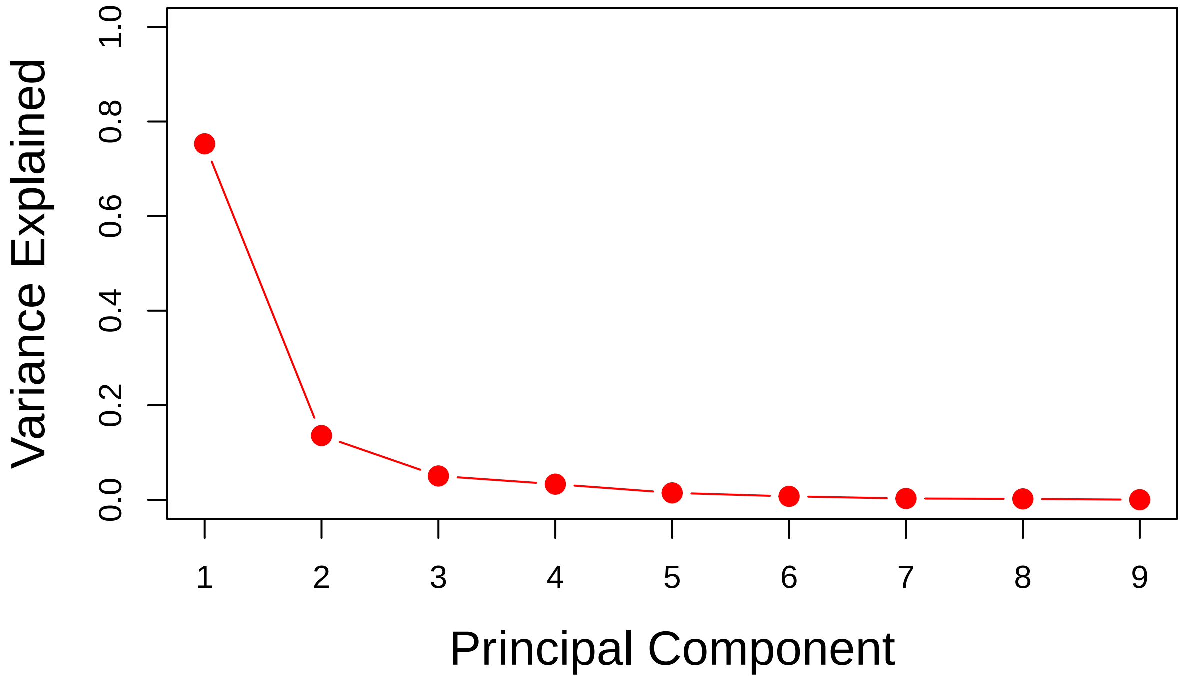
```
dat = read.table(file = "~/Documents/Math189/Places_rated.txt", fill = TRUE)
colnames(dat) <- c("Climate and Terrain","Housing","Health Care & the Environment",
                    "Crime", "Transportation","Education","The Arts","Recreation",
                    "Economics","Index of communities")
#scaled data
pca_result = prcomp(dat[,1:9], scale. = TRUE)
#eigen = eigen(var(pca_result$x), only.values = FALSE)

#Calculate variance explained by each PC
pca_var = pca_result$sdev ^2

#Calculate proportion of variance explained by each principal component
pve = pca_var/sum(pca_var)

#screeplot
plot(pve, xlab=" Principal Component ", ylab=" Proportion of
Variance Explained", ylim=c(0,1), xaxt="n" ,type='b', col="red", cex=2,
pch=20, cex.lab=1.5)
axis(1, at=c(1,2,3,4,5,6,7,8,9),labels=c(1,2,3,4,5,6,7,8,9))
```



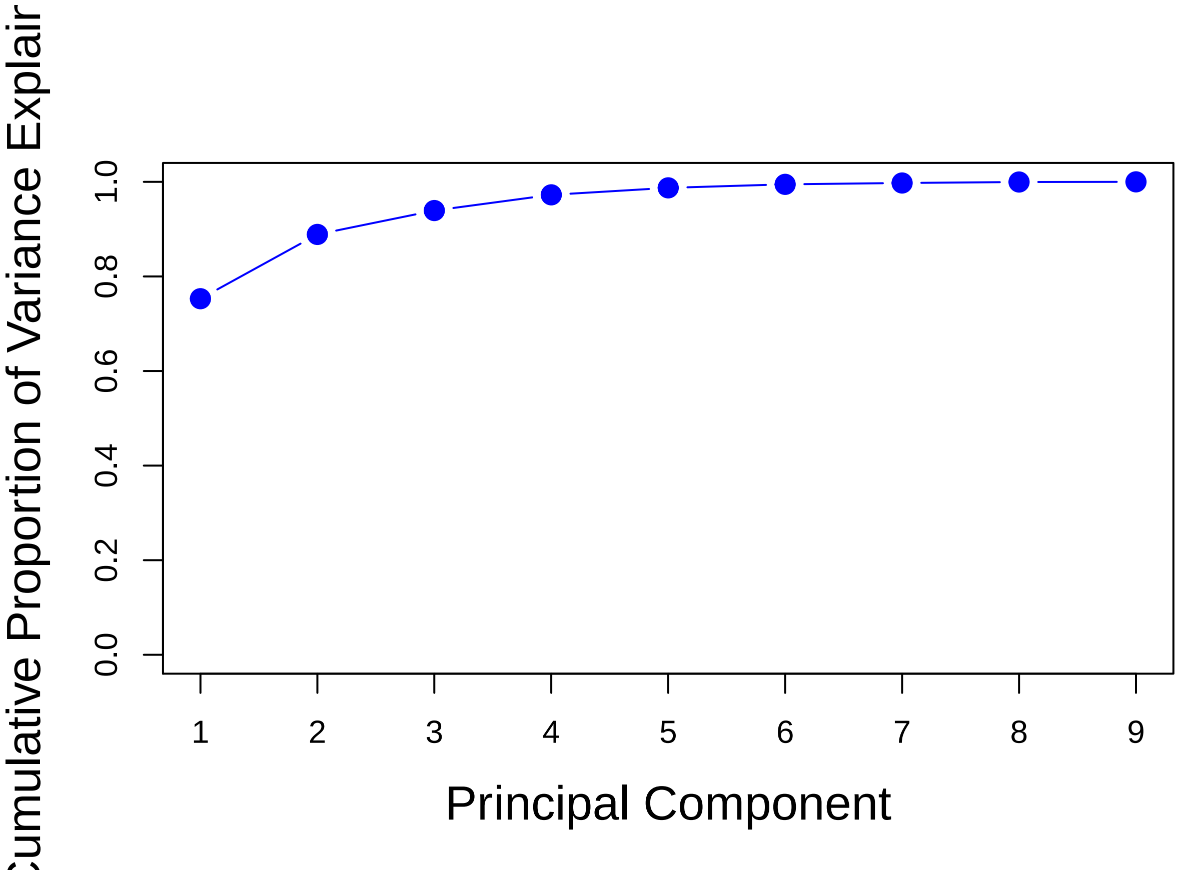
```r
#cumulative
plot(cumsum (pve), xlab=" Principal Component ", ylab ="
Cumulative Proportion of Variance Explained", ylim=c(0,1) , xaxt="n",
type='b', col="blue", cex=2, pch=20, cex.lab=1.5)
axis(1, at=c(1,2,3,4,5,6,7,8,9),labels=c(1,2,3,4,5,6,7,8,9))
```



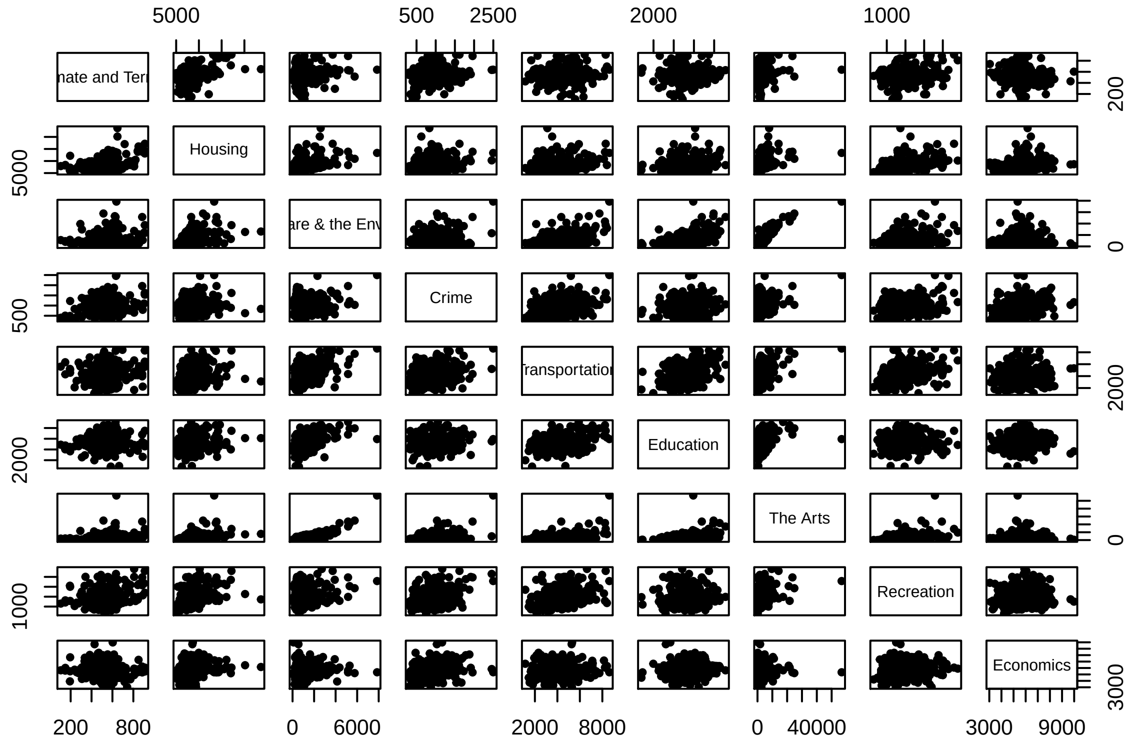

```r
#for unstrandardize
pca_unstd = prcomp(dat[,1:9], scale. = FALSE)

#Calculate proportion of variance explained by each principal component
pca_var_unstd = pca_unstd$sdev^2
pve_unstd = pca_var_unstd/sum(pca_var_unstd)

#scree plot
plot(pve_unstd , xlab=" Principal Component ", ylab=" Proportion of
Variance Explained", ylim=c(0,1), xaxt="n" ,type='b', col="red", cex=2,
pch=20, cex.lab=1.5)
axis(1, at=c(1,2,3,4,5,6,7,8,9),labels=c(1,2,3,4,5,6,7,8,9))
```




```
#cumulative plot
plot(cumsum (pve_unstd), xlab=" Principal Component ", ylab ="
Cumulative Proportion of Variance Explained", ylim=c(0,1) , xaxt="n",
type='b', col="blue", cex=2, pch=20, cex.lab=1.5)
axis(1, at=c(1,2,3,4,5,6,7,8,9),labels=c(1,2,3,4,5,6,7,8,9))
```




```
#When   we  perform PCA on  the unscaled    data,   the first   principal   component
#explains   more    than    95% of  total   variation.
#This   result  is  simply  a   consequence of  the scales  on  which   the variables   were
#measured   as  the total   variance    is  dominated   by  the variance    of  Assault
pairs(dat[,1:9], pch=20)
```



============================== Part 2 ==============================

```r
#scaled data
pca_result = prcomp(dat[,1:9], scale. = TRUE)

#adjust the negative direction
pca_result$rotation
```

```
##                                    PC1        PC2          PC3
## Climate and Terrain          0.2064140  0.2178353 -0.689955982
## Housing                      0.3565216  0.2506240 -0.208172230
## Health Care & the Environment 0.4602146 -0.2994653 -0.007324926
## Crime                        0.2812984  0.3553423  0.185104981
## Transportation               0.3511508 -0.1796045  0.146376283
## Education                    0.2752926 -0.4833821  0.229702548
## The Arts                     0.4630545 -0.1947899 -0.026484298
## Recreation                   0.3278879  0.3844746 -0.050852640
## Economics                    0.1354123  0.4712833  0.607314475
##                                      PC4        PC5         PC6
## Climate and Terrain           0.13732125 -0.3691499  0.37460469
## Housing                       0.51182871  0.2334878 -0.14163983
## Health Care & the Environment  0.01470183 -0.1032405 -0.37384804
## Crime                        -0.53905047 -0.5239397  0.08092329
## Transportation               -0.30290371  0.4043485  0.46759180
## Education                     0.33541103 -0.2088191  0.50216981
## The Arts                     -0.10108039 -0.1050976 -0.46188072
## Recreation                   -0.18980082  0.5295406  0.08991578
## Economics                     0.42176994 -0.1596201  0.03260813
##                                      PC7         PC8          PC9
## Climate and Terrain          -0.08470577 -0.36230833 0.0013913515
## Housing                      -0.23063862  0.61385513 0.0136003402
```

```
## Health Care & the Environment  0.01386761 -0.18567612 -0.7163548935
## Crime                          0.01860646  0.43002477 -0.0586084614
## Transportation                -0.58339097 -0.09359866  0.0036294527
## Education                      0.42618186  0.18866756  0.1108401911
## The Arts                      -0.02152515 -0.20398969  0.6857582127
## Recreation                     0.62787789 -0.15059597 -0.0255062915
## Economics                     -0.14974066 -0.40480926  0.0004377942
```

```r
pca_result$rotation <- -pca_result$rotation

#adjust to positive direction
pca_result$x <- - pca_result$x
#the the first 10
head(pca_result$x)
```

```
##             PC1         PC2         PC3         PC4        PC5         PC6
## [1,]  1.0401799 -0.89376897 -1.43665407 -0.50983413  0.5651365 -0.49785424
## [2,] -0.4398136 -0.07506618  1.15471654  1.11220718 -0.9968838  0.62854651
## [3,]  1.8755393 -0.06979169 -0.07334676  0.04623162  0.6795244  0.71409672
## [4,] -0.9107414  1.81758215 -1.09614720 -0.57358079 -0.7690800 -1.27123652
## [5,] -2.1492475 -0.32885808 -0.01973835  1.03150154  0.2385498 -1.28422401
## [6,]  1.7879611  0.78120167 -0.06083499 -0.46481865  0.8239929 -0.09922526
##              PC7        PC8         PC9
## [1,]  0.42375134  1.0017757 -0.34740439
## [2,]  0.01189488  0.4187458  0.12168548
## [3,]  0.23949403 -0.4418970  0.09420088
## [4,]  0.45871618  0.3714962 -0.31727660
## [5,]  0.15454404  0.1482641  0.30715341
## [6,] -0.56356212  0.1893559 -0.07334151
```

```r
biplot(pca_result, scale = 0)
pca_result$sdev
```

```
## [1] 1.8461560 1.1018059 1.0684003 0.9596446 0.8679199 0.7940793 0.7021736
## [8] 0.5639490 0.3469900
```

```r
(VE <- pca_result$sdev^2)
```

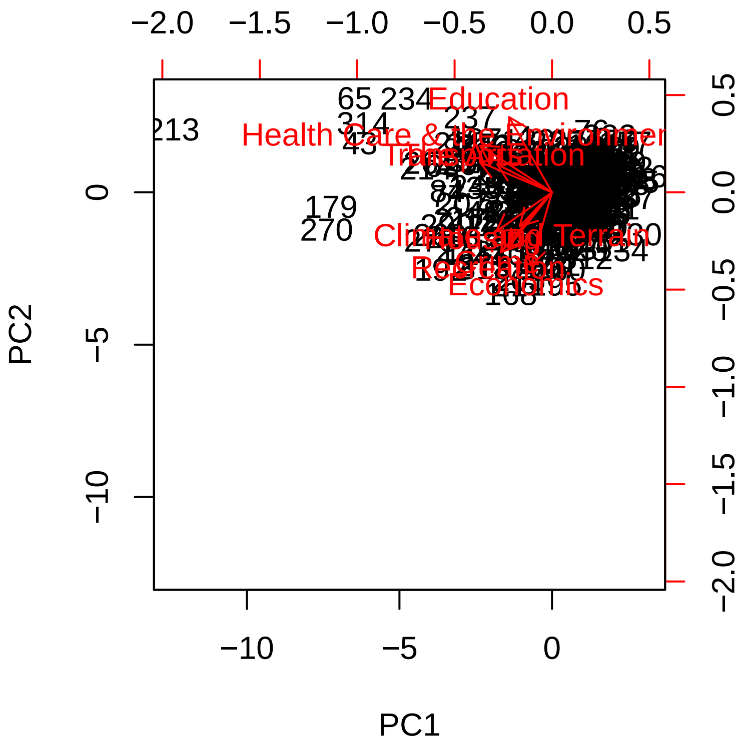
```
## [1] 3.4082918 1.2139762 1.1414791 0.9209178 0.7532849 0.6305619 0.4930477
## [8] 0.3180385 0.1204021
```

```r
PVE <- VE / sum(VE)
round(PVE, 2)
```

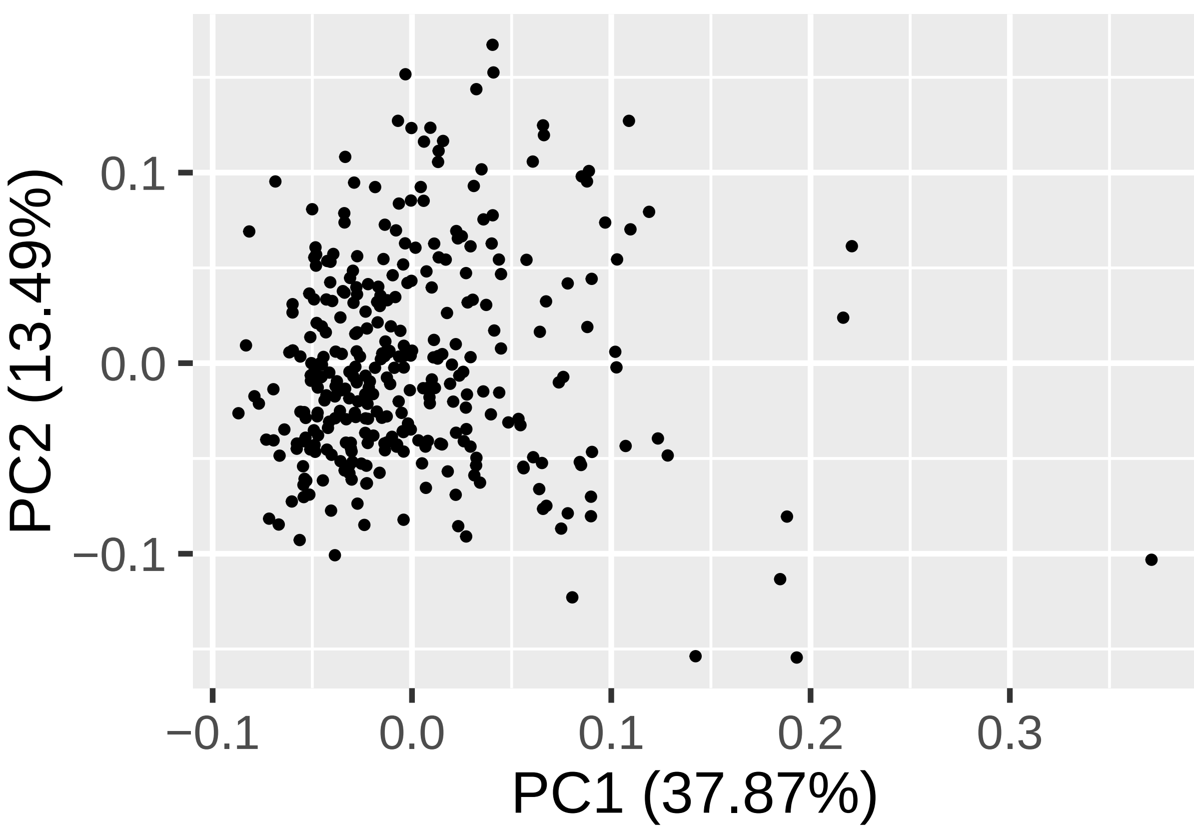
```
## [1] 0.38 0.13 0.13 0.10 0.08 0.07 0.05 0.04 0.01
```

```r
#Project data into first two principle components
biplot(pca_result , scale =0)

#Fancy plot
pca_result$rotation = -pca_result$rotation
pca_result$x = -pca_result$x


#{ggfortify} package let {ggplot2} know how to interpret PCA objects.
library(ggfortify)
```

```
## Loading required package: ggplot2
```




```
#Plot data projected to first two PCs
plot_0 = autoplot(pca_result, data = dat, colour = 'black')
plot_0+ theme_grey(base_size = 22)
```



```
plot_1=autoplot(pca_result, data = dat, color = 'steelblue', label=TRUE, shape=FALSE)
plot_1+ theme_grey(base_size = 22)
```