HW5

Sharon Zhao

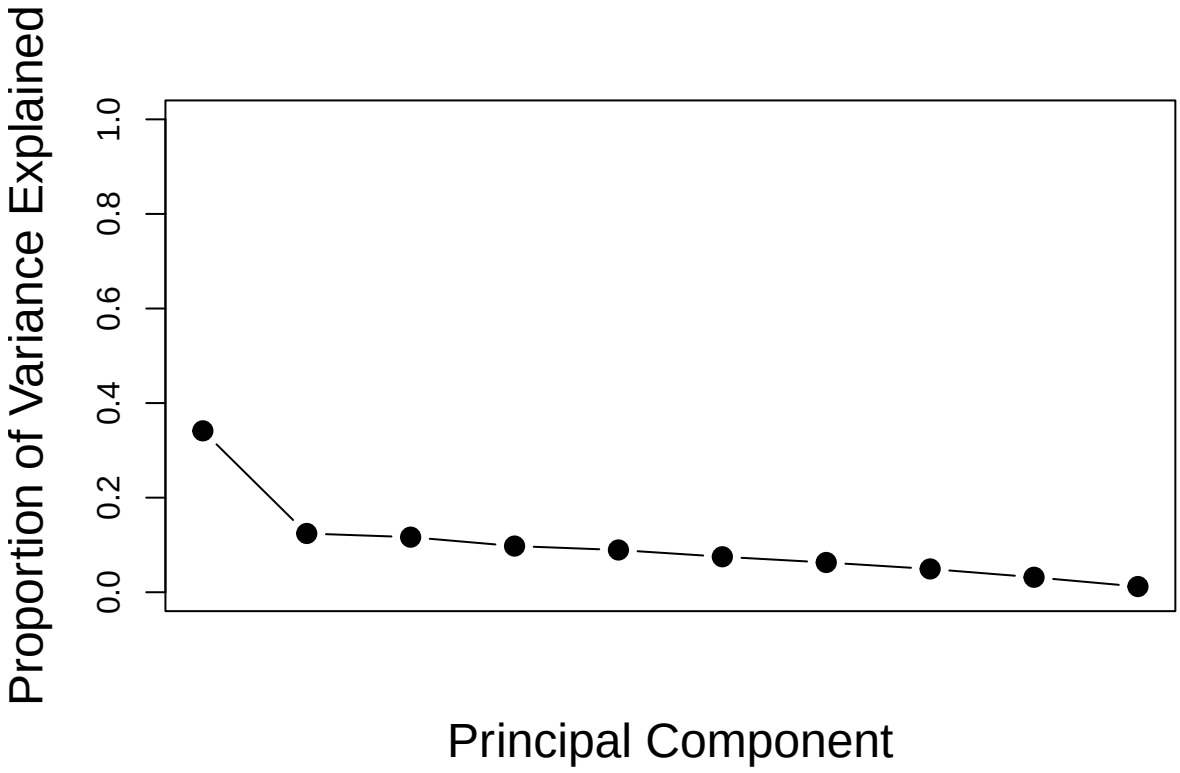
2/10/2022

Q1

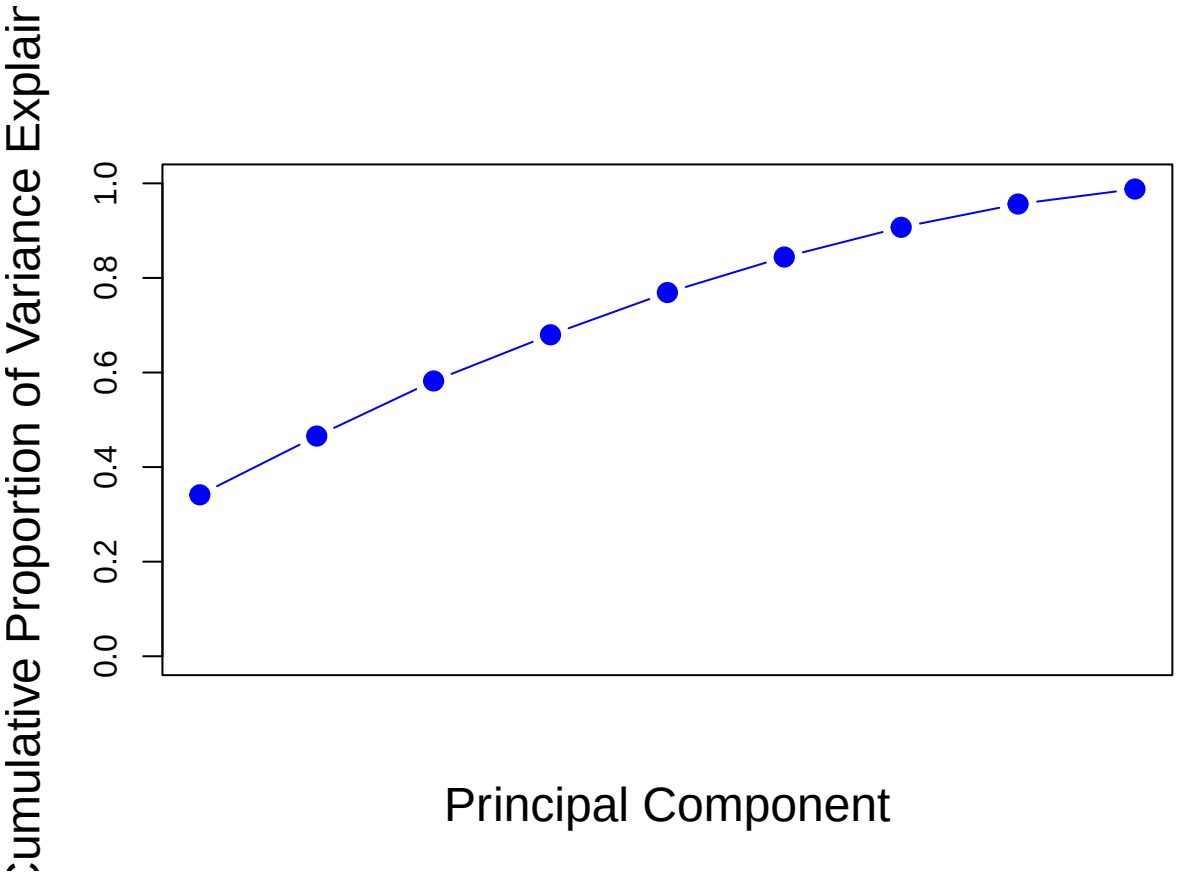
a

```
data <- read.delim("Places_Rated.txt", header =FALSE, sep = '')
df <- data.frame(data)
names(df) <- c("Climate and Terrain", "Housing",
               "Health Care & the Environment", "Crime",
               "Transportation", "Education", "The Arts",
               "Recreation", "Economics", "index")
standard_df <- scale(df)
standard_cov <- cov(standard_df)
standard_ev <- eigen(standard_cov)
standard_prop <- c()
for ( i in 1:10) {
  prop <- (standard_ev$values[i]) / (sum(standard_ev$values))
  standard_prop <- append(standard_prop, prop)
}
standard_cprop <- c()
cprop <- 0
for ( i in 1:9) {
  cprop <- cprop + standard_prop[i]
  standard_cprop <- append(standard_cprop, cprop)
}

plot(standard_prop, xlab="Principal Component",
     ylab="Proportion of Variance Explained",
     ylim=c(0,1), xaxt="n", type='b', cex=2, pch=20, cex.lab=1.5)
```

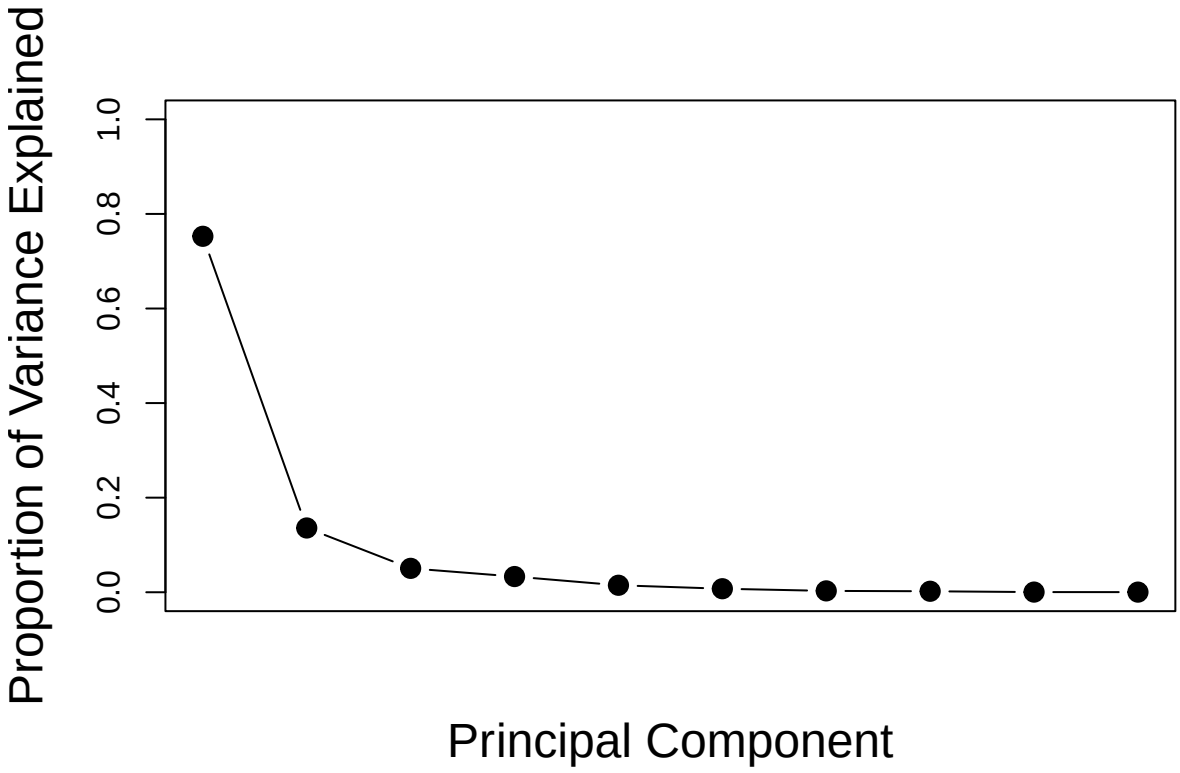



```
plot(standard_cprop, xlab="Principal Component",
     ylab="Cumulative Proportion of Variance Explained",
     ylim=c(0,1), xaxt="n", type="b", col="blue", cex=2,
     pch=20, cex.lab=1.5)
```

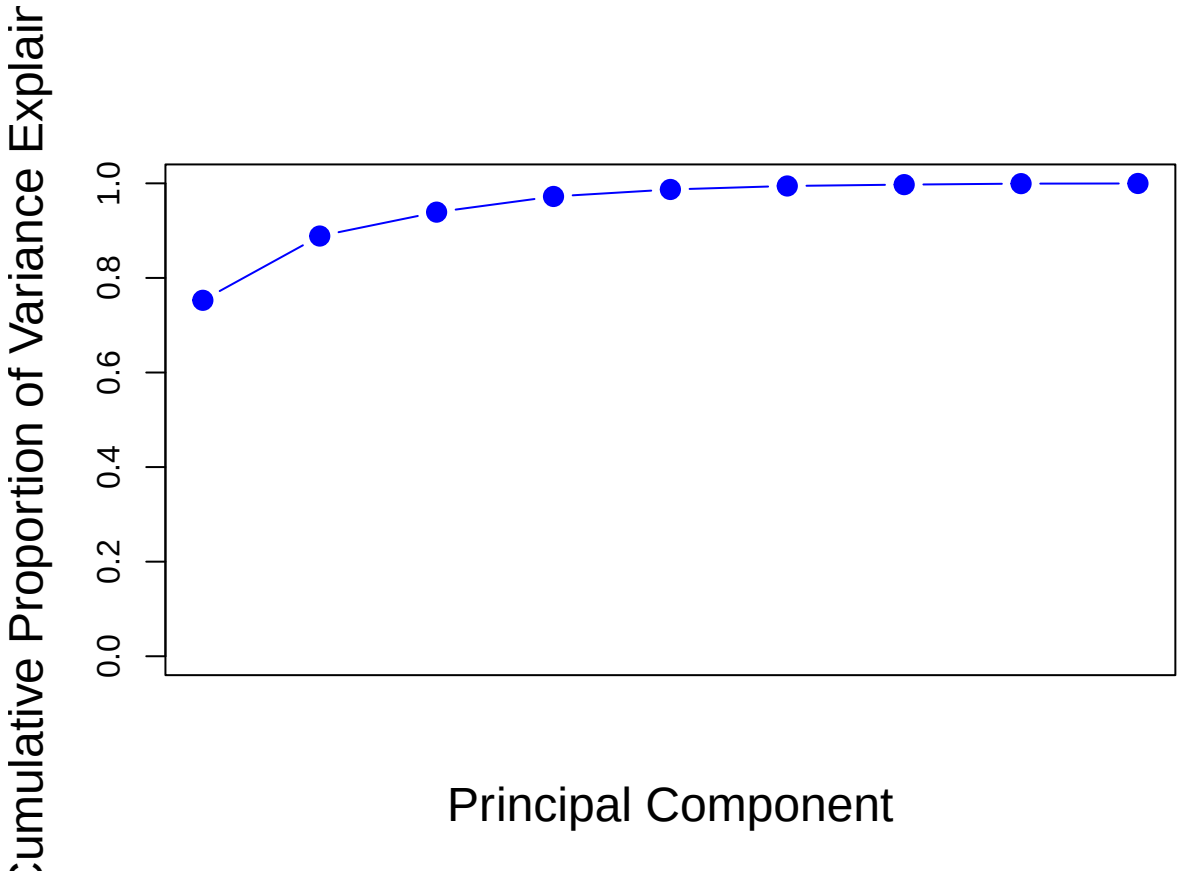



```
raw_cov <- cov(df)
raw_ev <- eigen(raw_cov)
raw_prop <- c()
for ( i in 1:10) {
  prop <- (raw_ev$values[i]) / (sum(raw_ev$values))
  raw_prop <- append(raw_prop, prop)
}
raw_cprop <- cumsum(raw_prop[1:9])

plot(raw_prop, xlab="Principal Component",
     ylab="Proportion of Variance Explained", ylim=c(0,1),
     xaxt="n", type='b', cex=2, pch=20, cex.lab=1.5)
```




```
plot(raw_cprop, 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)
```

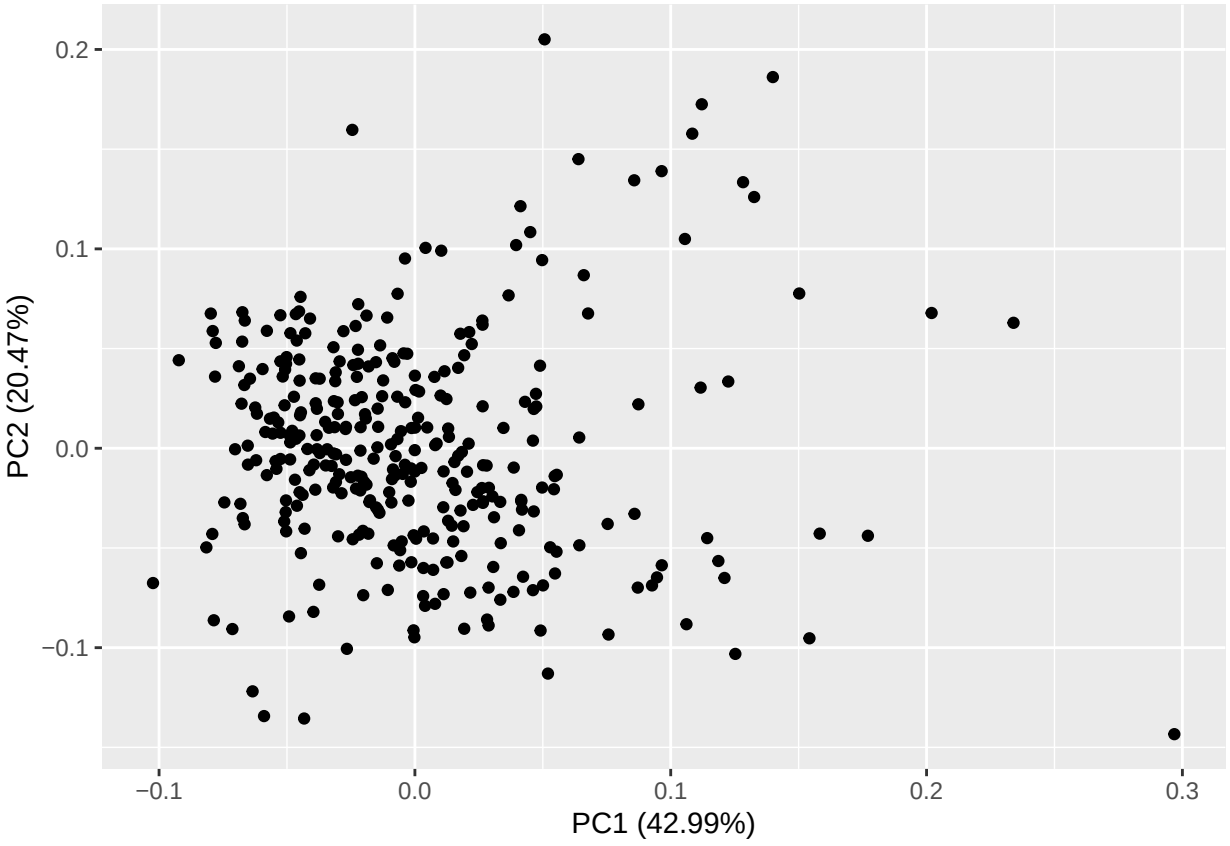


b

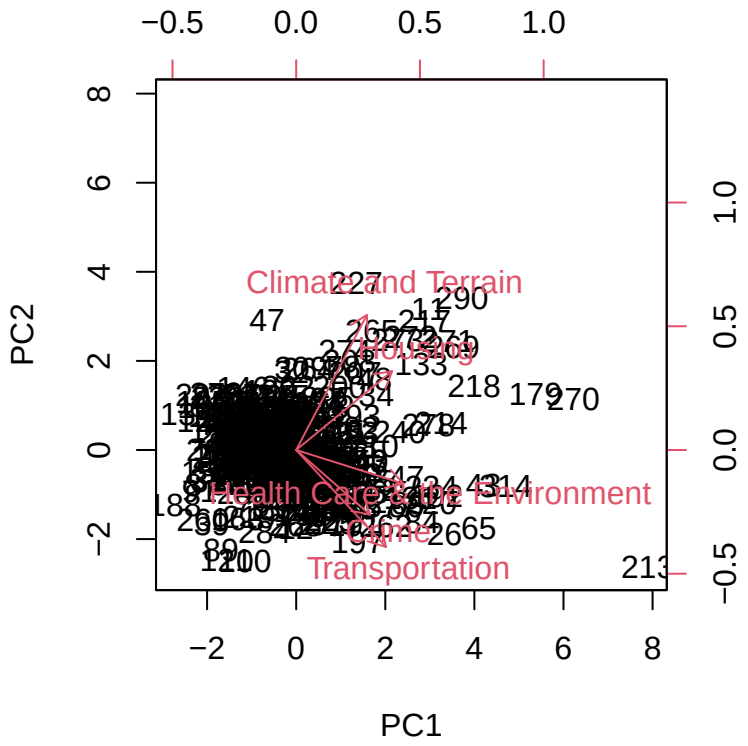
```
library("ggplot2")
library("ggfortify")
pca_result <- prcomp(df[,1:5], scale=TRUE)
pca_result$rotation
```

```
##                                     PC1        PC2        PC3        PC4
## Climate and Terrain           0.3568742  0.6823118 -0.3438239  0.4810650
## Housing                       0.4842331  0.3983105  0.3493474 -0.3856060
## Health Care & the Environment 0.5415405 -0.1704116  0.2556157 -0.3307141
## Crime                         0.3729212 -0.3265431 -0.7959444 -0.2794155
## Transportation                0.4536769 -0.4900294  0.2467278  0.6576015
##                                     PC5
## Climate and Terrain           0.2396936
## Housing                      -0.5797684
## Health Care & the Environment  0.7092127
## Crime                        -0.2066368
## Transportation               -0.2464430
```

```
autoplot(pca_result, data=df, color='black')
```




```
biplot(pca_result, scale=0)
```


−0.5
0.0
0.5
1.0
8
6
4
2
0
−2
PC2
1.0
0.5
0.0
−0.5
Climate and Terrain
Housing
Health Care
Crime
Transportation
Environment
227
290
47
133
218
179
270
314
65
213
−2
0
2
4
6
8
PC1