HW5_q4

Will Scheib

12/3/2021

4

```
swiss <- swiss
```

a

```
pr.out<-prcomp(swiss, scale=TRUE)

pr.out$rotation
```

```
##                          PC1        PC2         PC3         PC4         PC5
## Fertility        -0.4569876  0.3220284 -0.17376638  0.53555794 -0.38308893
## Agriculture      -0.4242141 -0.4115132  0.03834472 -0.64291822 -0.37495215
## Examination       0.5097327  0.1250167 -0.09123696 -0.05446158 -0.81429082
## Education         0.4543119  0.1790495  0.53239316 -0.09738818  0.07144564
## Catholic         -0.3501111  0.1458730  0.80680494  0.09947244 -0.18317236
## Infant.Mortality -0.1496668  0.8111645 -0.16010636 -0.52677184  0.10453530
##                          PC6
## Fertility         0.47295441
## Agriculture       0.30870058
## Examination      -0.22401686
## Education         0.68081610
## Catholic         -0.40219666
## Infant.Mortality -0.07457754
```

b

```
# change signs to make interpretation more intuitive,
# so that positive scores are above average values
pr.out$rotation<- -pr.out$rotation
pr.out$x<- -pr.out$x

# check first and second PCs
pr.out$rotation[,1:2]
```

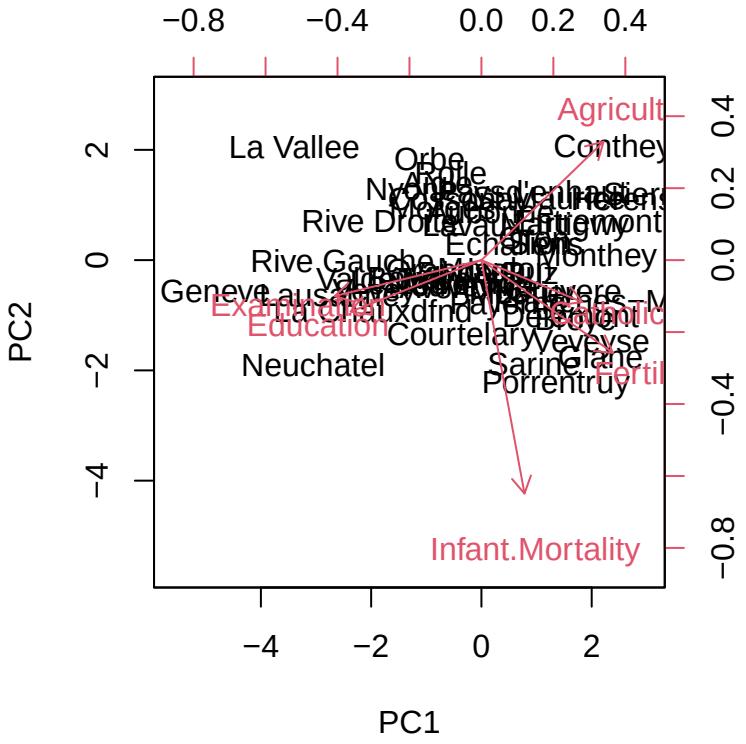
```
##                         PC1        PC2
## Fertility         0.4569876 -0.3220284
## Agriculture       0.4242141  0.4115132
## Examination      -0.5097327 -0.1250167
## Education        -0.4543119 -0.1790495
## Catholic          0.3501111 -0.1458730
## Infant.Mortality  0.1496668 -0.8111645
```

Higher Fertility, Agriculture, Catholic, and Infant Mortality all increase the first PC, but higher Examination and Education decrease it. Infant Mortality has the least impact, whereas the rest have similar impacts.

Higher Agriculture increases the second PC, but higher Fertility, Examination, Education, Catholic, and Infant Mortality decreases it. Infant Mortality has the highest impact, followed by Agriculture and Fertility.

c

```
# create biplot
biplot(pr.out, scale=0)
```



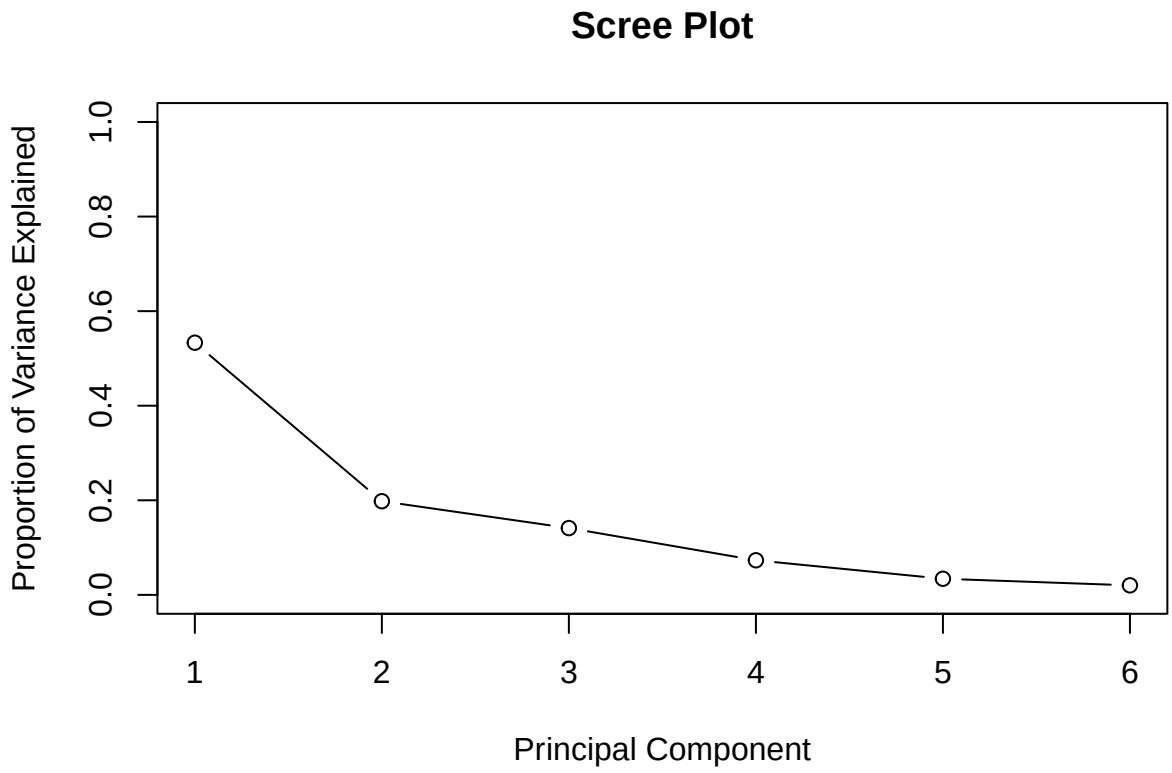
High examination and education. Low infant mortality, Catholic, and fertility. Middle agriculture.

d

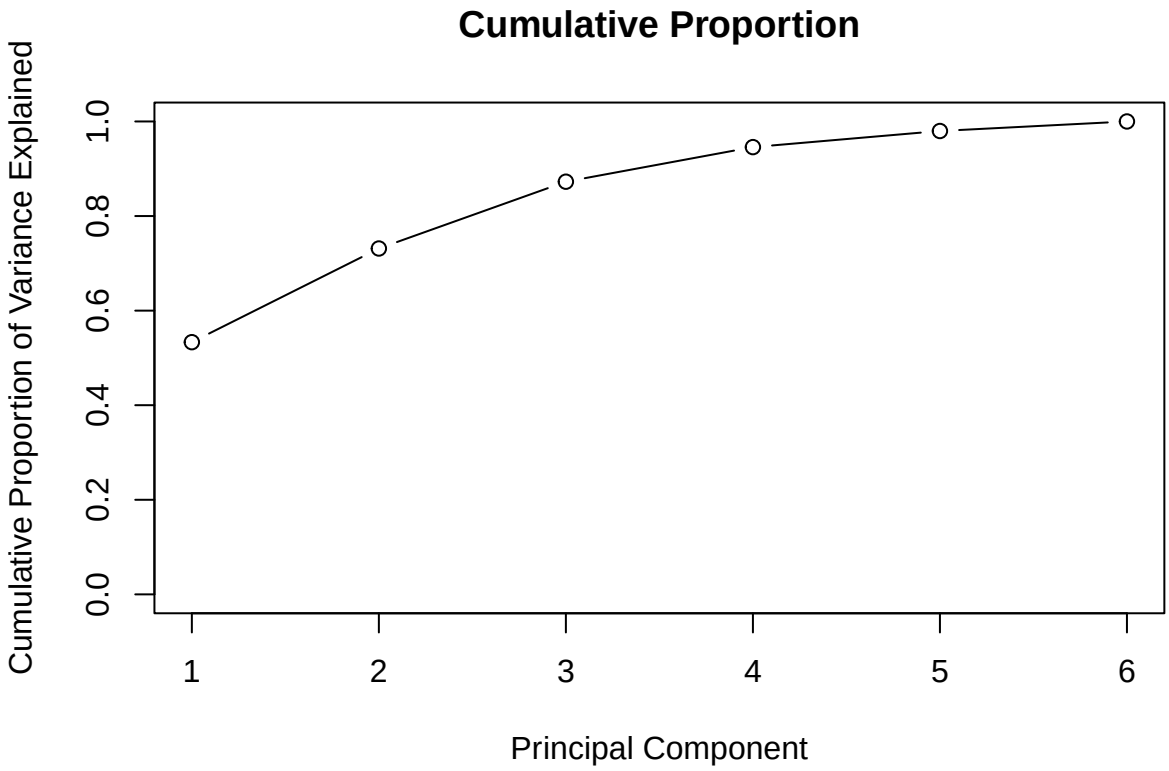
```
# variance of each PC
pr.var<-pr.out$sdev^2

# proportion of variance in features explained by each PC
pve<-pr.var/sum(pr.var)

# Scree plot
plot(pve, xlab="Principal Component", ylab="Proportion of Variance Explained", main="Scree Plot", ylim=
```



```r
# Cumulative plot
plot(cumsum(pve), xlab="Principal Component", ylab="Cumulative Proportion of Variance Explained", main="
```



Two, because after PC2, each principal component significantly decreases interpretability, but only slightly increases the amount of variance explained.