Happiest Countries

Rory Quinlan

Set Up

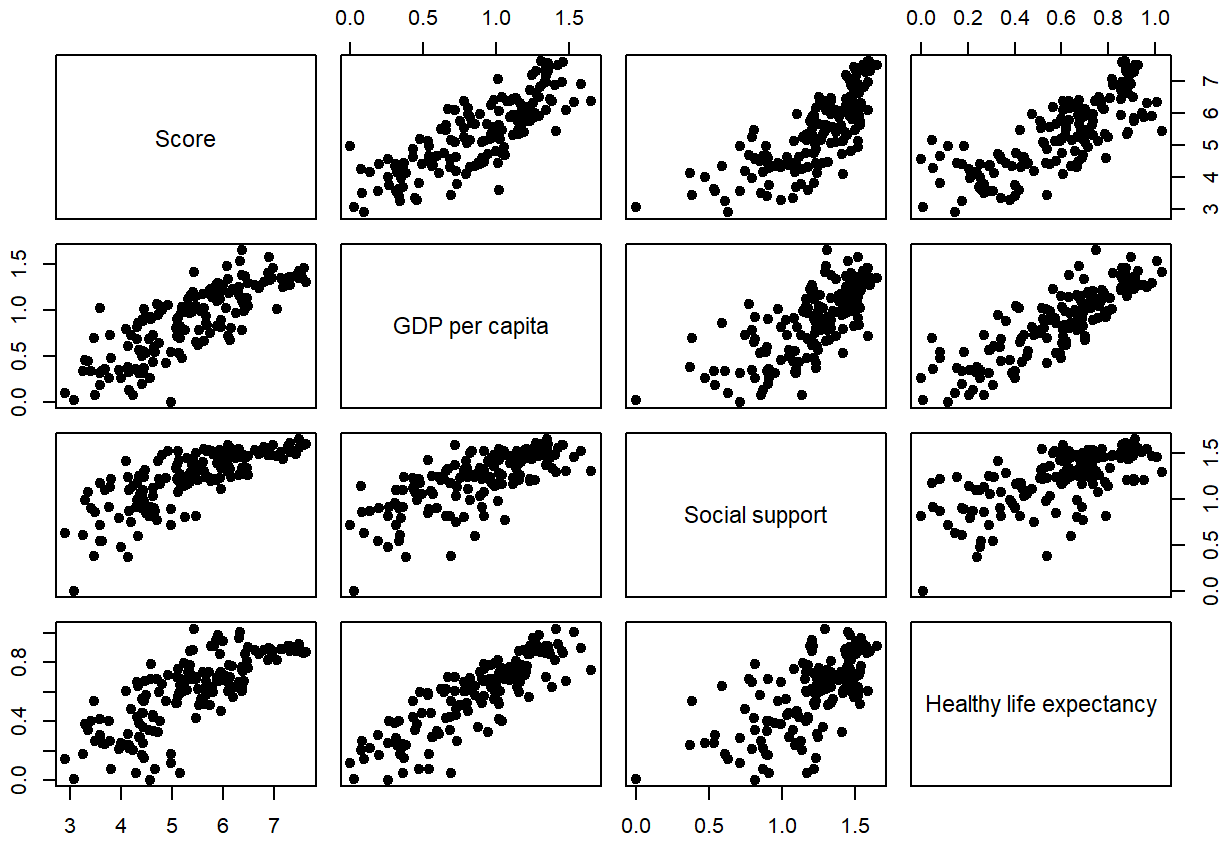
```
library(tidyverse)
library(dplyr)
library(factoextra)
library(cluster)
library(gridExtra)

# Load data
happiness_0 = read_csv("happiness_2018.csv", show_col_types = FALSE)

happiness_1 = happiness_0 %>%
 mutate(Corruption = as.numeric(`Perceptions of corruption`))

# Remove NA values
happiness_2 = happiness_1 %>% drop_na()
dim(happiness_1)
```

```
## [1] 156  10
```

```
happiness_1 = happiness_2 %>%
 select(-c(`Overall rank`, `Country or region`,`Perceptions of corruption`))
```

Explore Data

```
# Correlogram
pairs(happiness_1[,1:4], pch = 19)
```

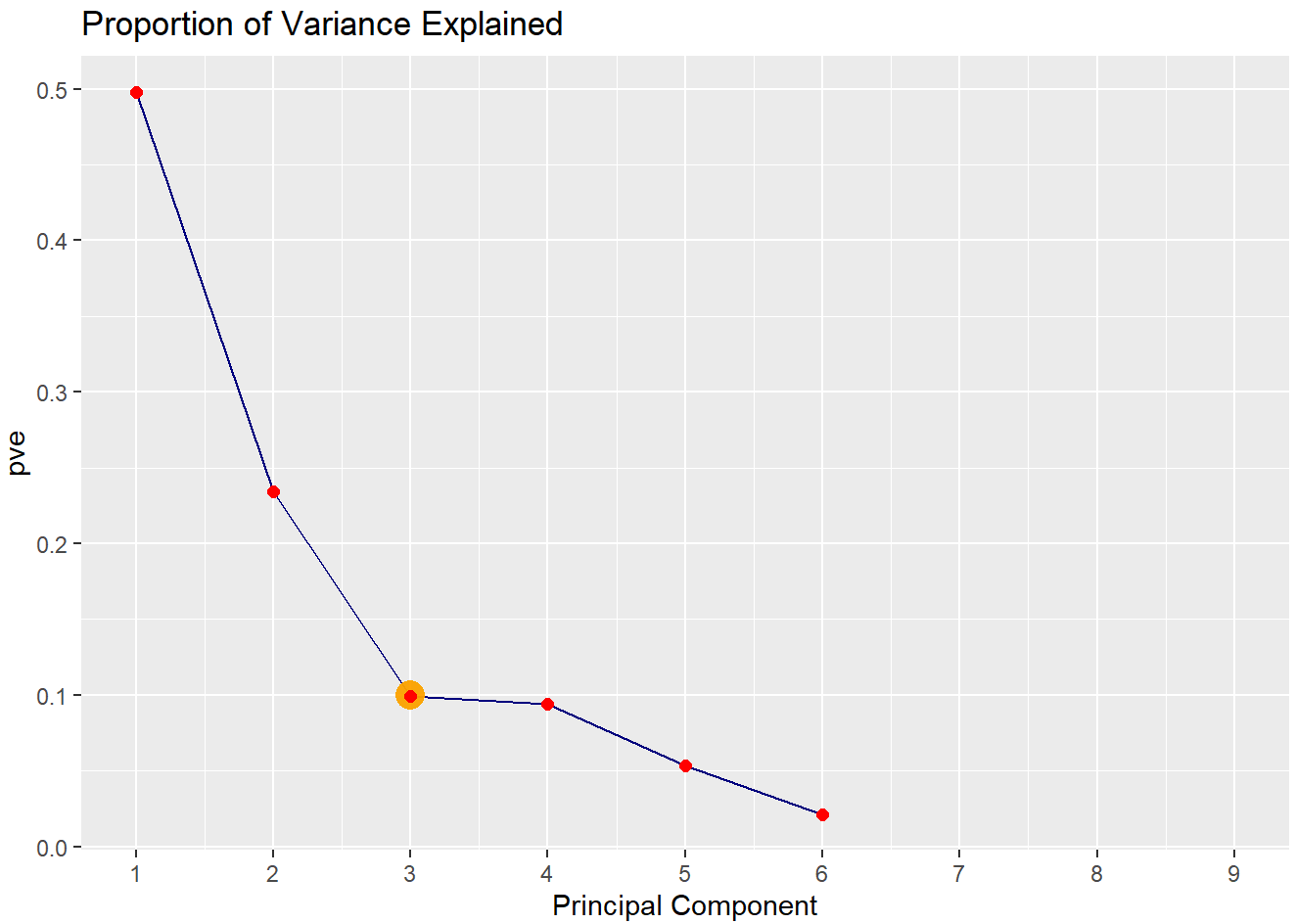


PCA

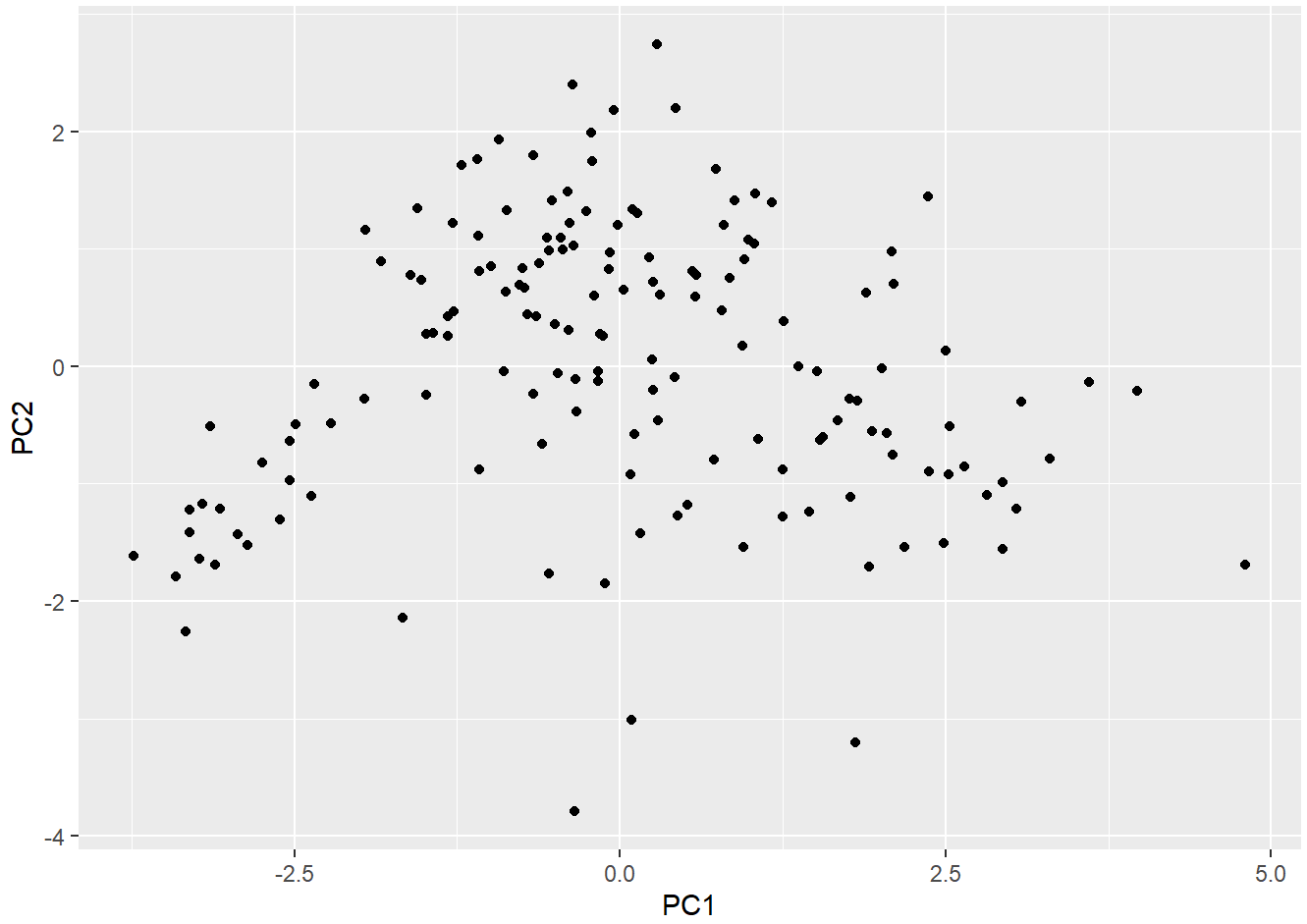
```
# Create A PCA
pc.happiness = happiness_1 %>%
 select(-Score) %>%
 prcomp(scale=TRUE)
pc.happiness
```

```
## Standard deviations (1, .., p=6):
## [1] 1.7286165 1.1856234 0.7717936 0.7501125 0.5653144 0.3581444
##
## Rotation (n x k) = (6 x 6):
##                                    PC1        PC2        PC3         PC4
## GDP per capita               -0.5126579  0.2666787 -0.1164253 -0.17098649
## Social support               -0.4720852  0.2376977 -0.1824560  0.27652670
## Healthy life expectancy      -0.5038326  0.2424268 -0.1581973 -0.20850817
## Freedom to make life choices -0.3752315 -0.3544097  0.4418517  0.68603484
## Generosity                   -0.1219418 -0.6758505 -0.7238170  0.03330243
## Corruption                   -0.3237308 -0.4808653  0.4571483 -0.61568506
##                                     PC5         PC6
## GDP per capita                0.23122805 -0.75485540
## Social support               -0.76997346  0.13423480
## Healthy life expectancy       0.45787741  0.63970950
## Freedom to make life choices  0.25983295 -0.01432444
## Generosity                    0.04285037 -0.03873052
## Corruption                   -0.27326438  0.03522586
```

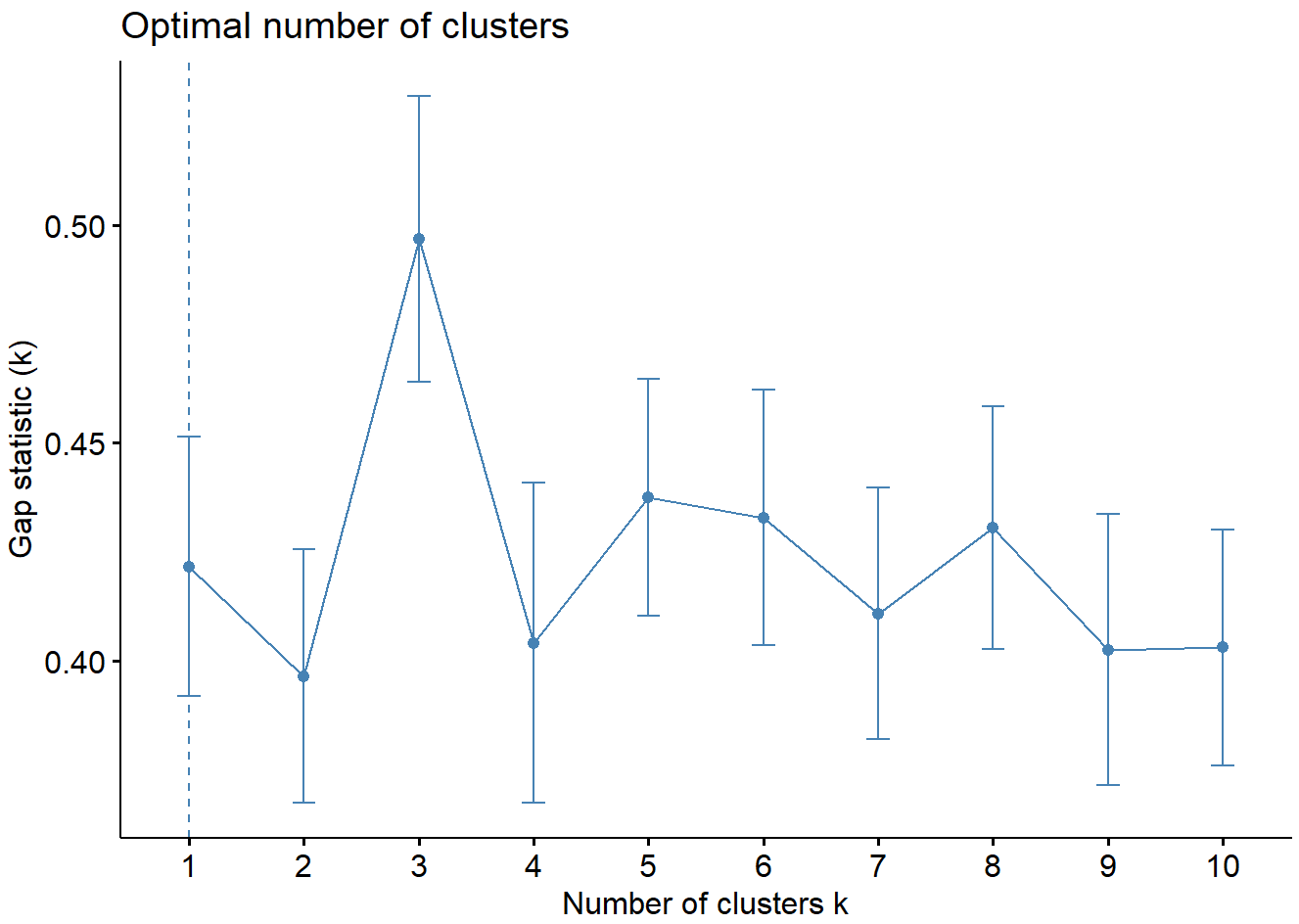
```
# Graph Explantory power of variables
PRVar<- pc.happiness$sdev^2
PVE<- PRVar[1:9]/sum(PRVar)

PC=1:9
data=data.frame(PC, PVE)
ggplot(data=data, aes(x=PC, y=PVE))+
 geom_line(color="navy")+
 geom_point(aes(x=3,y=0.1),cex=5,color="orange",alpha=0.3)+
 geom_point(color="red",cex=2)+
 labs(title="Proportion of Variance Explained", x="Principal Component",y="pve")+
 scale_x_continuous(breaks = 1:9)
```




```
# View scatterplot of greatest contributors from PCA
PC12 <- pc.happiness$x %>% as_tibble() %>% select(1:2)
pc.happiness$x %>% as_tibble() %>% select(PC1, PC2) %>%
 bind_cols(happiness_1) %>%
 ggplot(aes(x = `PC1`, y = `PC2`)) + geom_point()
```



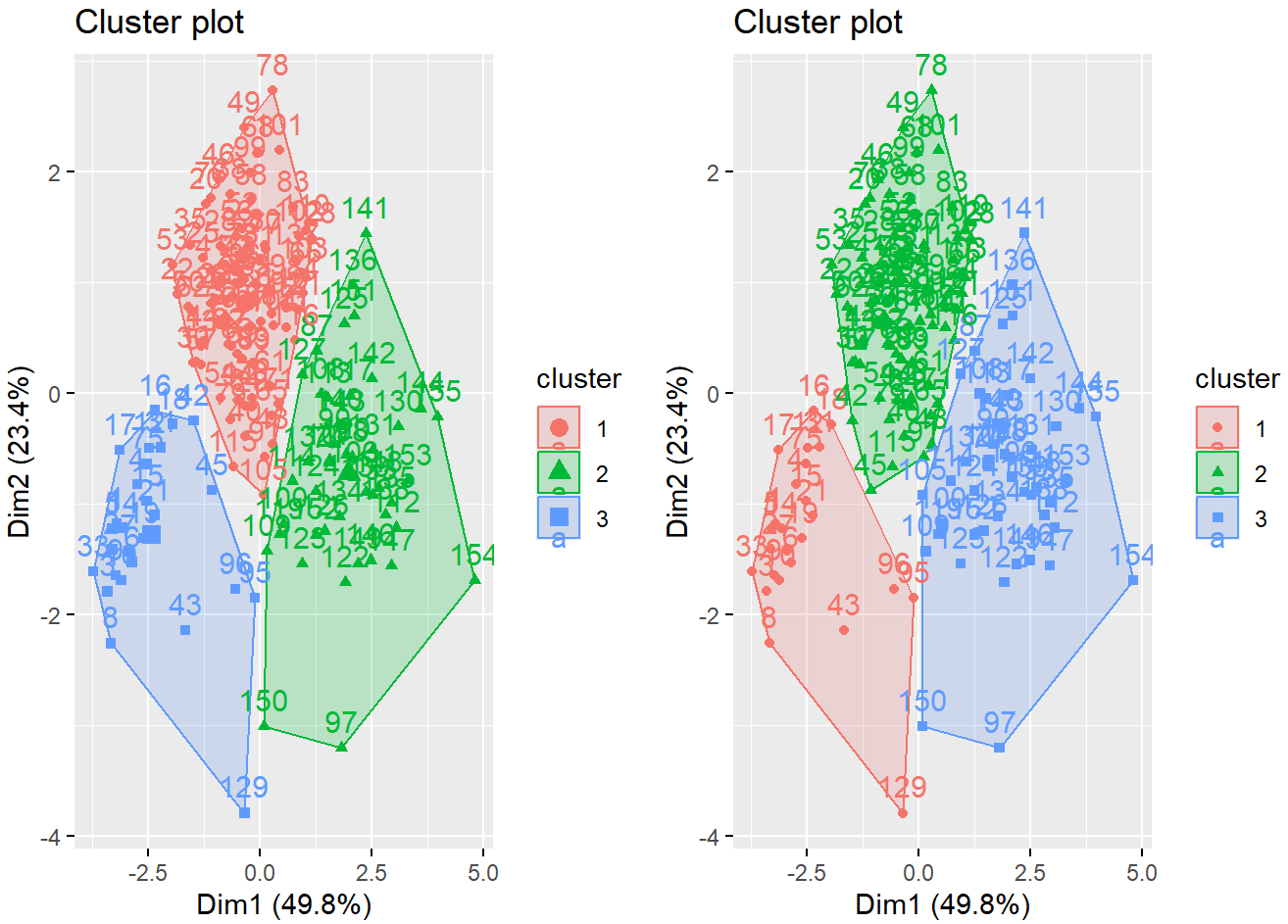

```
# Find optimal clusters
happiness_4 = scale(happiness_1[,-1])
fviz_nbclust(PC12, kmeans, method = "gap_stat")
```




```
km_mod = kmeans(happiness_4, centers=3)
pam_mod = pam(happiness_4, 3)
```

```
# Create variable for cluster number in df
PC12_cluster = happiness_1 %>% mutate(cluster=factor(pam_mod$cluster))
```

```
# Compare kmean and pam
p1<-fviz_cluster(km_mod, data = happiness_4)

p2<- fviz_cluster(pam_mod, data = happiness_4)

grid.arrange(p1, p2, ncol=2)
```



```r
# Create clusters with country name
PC12_cluster <- PC12 %>% mutate(cluster = factor(pam_mod$cluster), countryOrRegion = factor(happiness_2$`Country or region
`))

# Filter and display the happiest countries
as.data.frame(PC12_cluster %>% filter(cluster == 1) %>% select(countryOrRegion))
```

```
##    countryOrRegion
## 1          Finland
## 2           Norway
## 3          Denmark
## 4          Iceland
## 5      Switzerland
## 6      Netherlands
## 7           Canada
## 8      New Zealand
## 9           Sweden
## 10       Australia
## 11         Austria
## 12         Ireland
## 13         Germany
## 14         Belgium
## 15      Luxembourg
## 16   United States
## 17          Israel
## 18           Malta
## 19           Qatar
## 20       Singapore
## 21      Uzbekistan
## 22       Hong Kong
## 23       Indonesia
## 24          Bhutan
## 25         Myanmar
```