Lab 7 – MATH 243

Theodore Dounias

11/21/2017

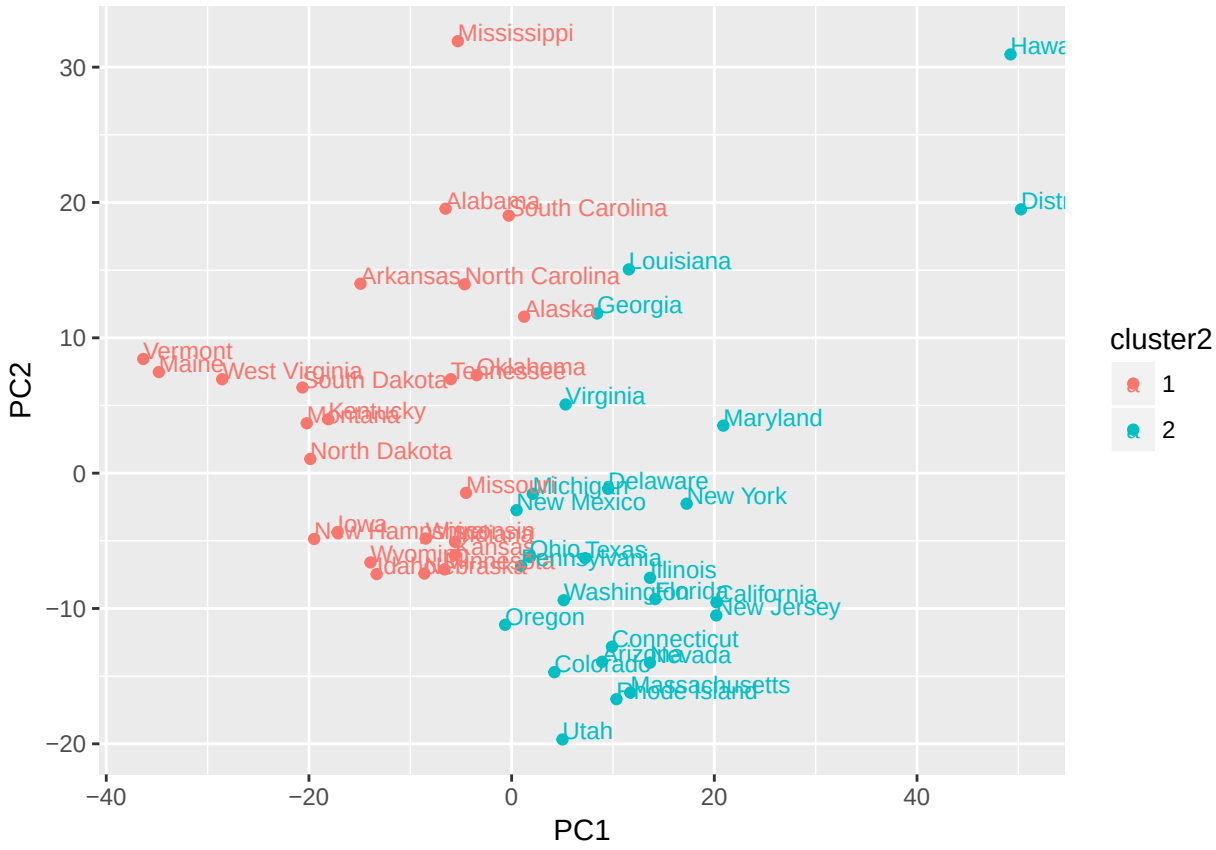
Part I

1

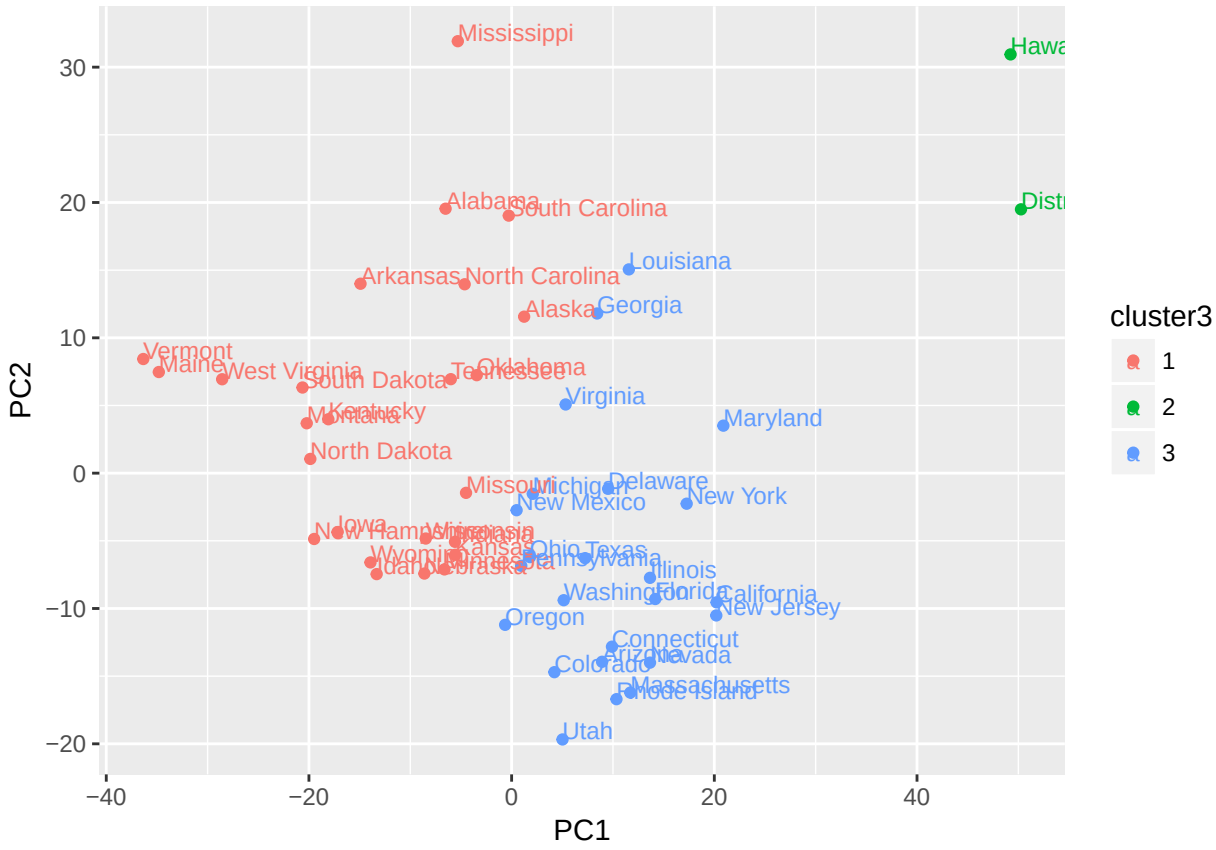
```
set.seed(40)

km2 <- kmeans(poverty[ ,-1], 2, nstart = 20)

km3 <- kmeans(poverty[ ,-1], 3, nstart = 20)

km4 <- kmeans(poverty[ ,-1], 4, nstart = 20)
```

2-3

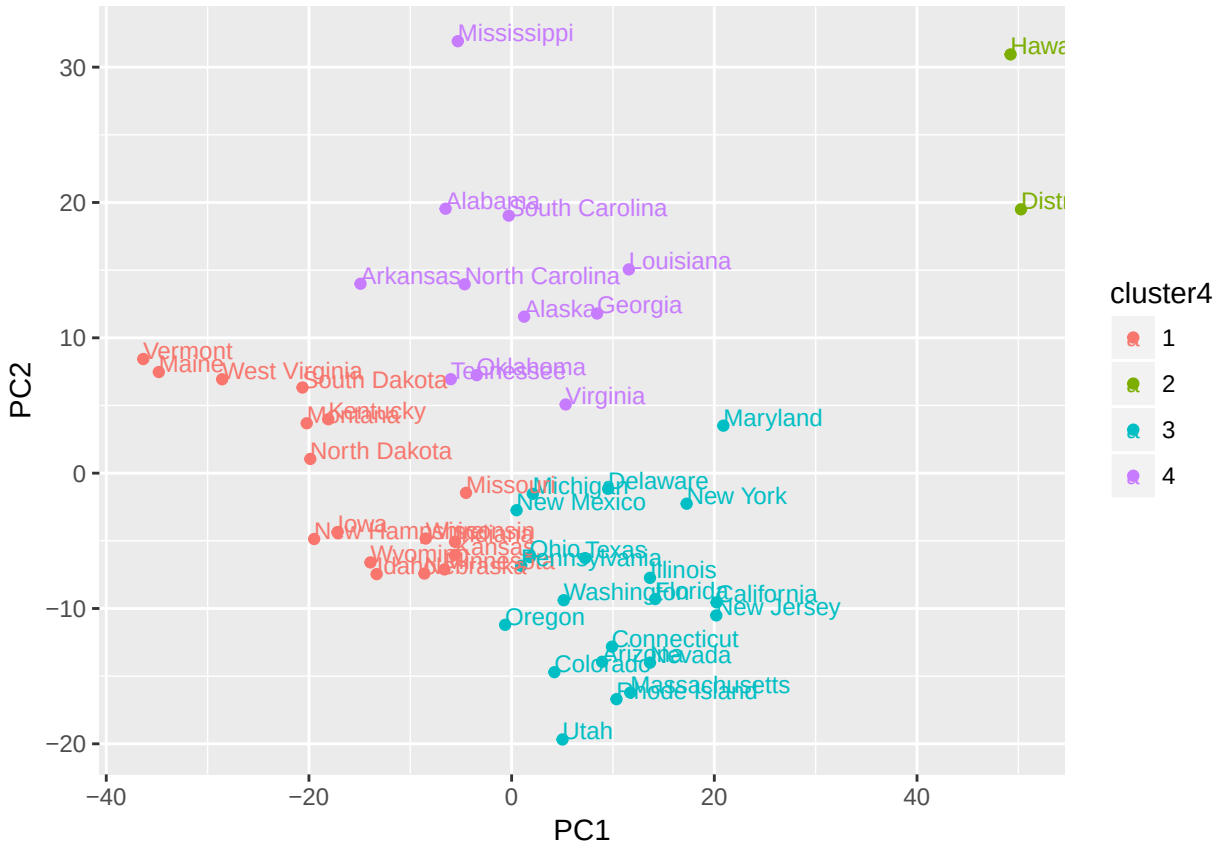
```
pca <- prcomp(poverty[, -1])

d <- data.frame(PC1 = pca$x[, 1],
                PC2 = pca$x[, 2],
                cluster2 = as.factor(km2$cluster),
                cluster3 = as.factor(km3$cluster),
                cluster4 = as.factor(km4$cluster),
                state = poverty$State)

ggplot(d, aes(x = PC1, y = PC2, col = cluster2)) +
  geom_point() +
  geom_text(aes(label = state), hjust = 0, vjust = 0, size = 3)
```




```
ggplot(d, aes(x = PC1, y = PC2, col = cluster3)) +
  geom_point() +
  geom_text(aes(label = state), hjust = 0, vjust = 0, size = 3)
```

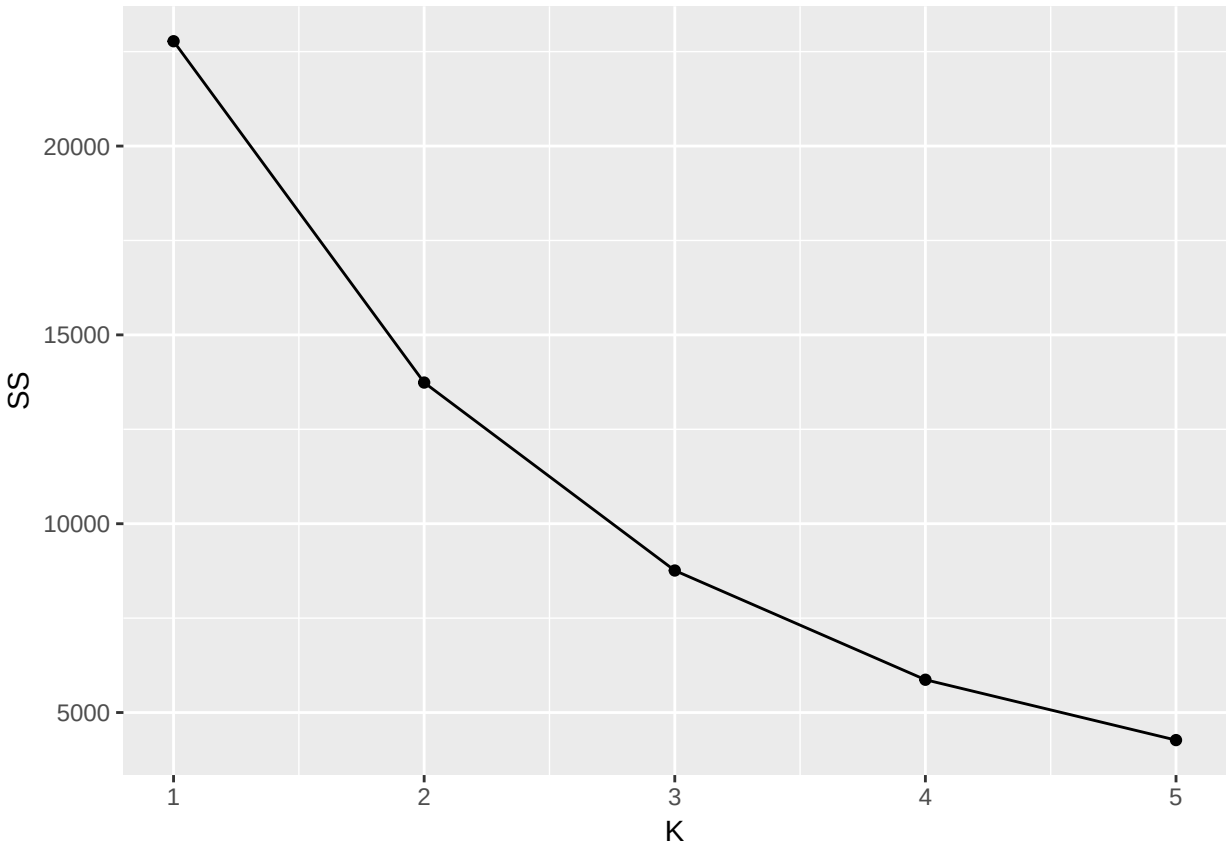



```
ggplot(d, aes(x = PC1, y = PC2, col = cluster4)) +
  geom_point() +
  geom_text(aes(label = state), hjust = 0, vjust = 0, size = 3)
```



4

```
km1 <- kmeans(poverty[ ,-1], 1, nstart = 20)

km5 <- kmeans(poverty[ ,-1], 5, nstart = 20)

df <- data.frame("K" = 1:5,
                 "SS" = c(km1$tot.withinss,
                          km2$tot.withinss,
                          km3$tot.withinss,
                          km4$tot.withinss,
                          km5$tot.withinss
                          ))
ggplot(df, aes(x = K, y = SS)) +
  geom_point() +
  geom_line()
```



Part II