

Data I use “Violent Crime Rates by US State dataset”, USArrests.

Load ggplot2 and fiftystater libraries and data USArrests.

```
library(ggplot2)
if(!require(fiftystater)){
  install.packages("fiftystater")
}
```

```
## Loading required package: fiftystater
```

```
library(fiftystater)
data("USArrests")
```

Analysis

PCA shows largely 2 separate groups of USArrests dataset, so I performed k-means clustering with k=2, and merge arrest data with geographic data from fifty_states library.

```
cluster=kmeans(USArrests,2)$cluster

usar=data.frame(id=tolower(rownames(USArrests)),cluster=cluster,
                TotalCrime=rowSums(USArrests[,c(1,2,4)]))

statell=data.frame(id=aggregate(fifty_states$lat,by=fifty_states["id"],mean)$id,
                   lat=aggregate(fifty_states$lat,by=fifty_states["id"],mean)$x,
                   long=aggregate(fifty_states$long,by=fifty_states["id"],mean)$x)

ur=merge(usar,statell,by="id")
```

Visualization

I then plotted crime data by state on USA map by ggplot, each marker colored by cluster membership, and sized by total crime number by state. The map shows clustering algorithm can separate groups.

```
ggplot(ur, aes(map_id = id))+geom_map(aes(fill = cluster), map = fifty_states)+
  expand_limits(x = fifty_states$long, y = fifty_states$lat)+
  geom_point(aes(x=long,y=lat,size=TotalCrime,col="red"))+
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  labs(x = "", y = "") +
  theme(legend.position = "bottom",panel.background = element_blank())+
  ggtitle("Violent Crime Rates by US State")
```

Violent Crime Rates by US State

