Knowledge Mining
(EPPS 6323)
Assignment 4

Submitted by

Samuel B. Adelusi

(BSA210004)

February 2023

School of Economic, Political and Policy Sciences

THE UNIVERSITY OF TEXAS AT DALLAS

Master in Social Data Analytics & Research

Question 2.

What problems do you encounter when working with the dataset? **?**

[1] "Murder"   "Assault" "UrbanPop" "Rape"

Question 2.

```
# Get means and variances of variables
> apply(USArrests, 2, mean)


Murder              Assault             UrbanPop                    Rape
7.788               170.760             65.540                      21.232

> apply(USArrests, 2, var)

Murder              Assault             UrbanPop                    Rape
18.97047            6945.16571          209.51878                   87.72916

> pr.out$center # the centering and scaling used (means)
  Murder            Assault             UrbanPop                    Rape
   7.788            170.760             65.540                      21.232
```

Question 2.

> pr.out$scale # the matrix of variable loadings (eigenvectors)

```
Murder         Assault         UrbanPop         Rape
 4.355510      83.337661       14.474763       9.366385
```

> pr.out$rotation

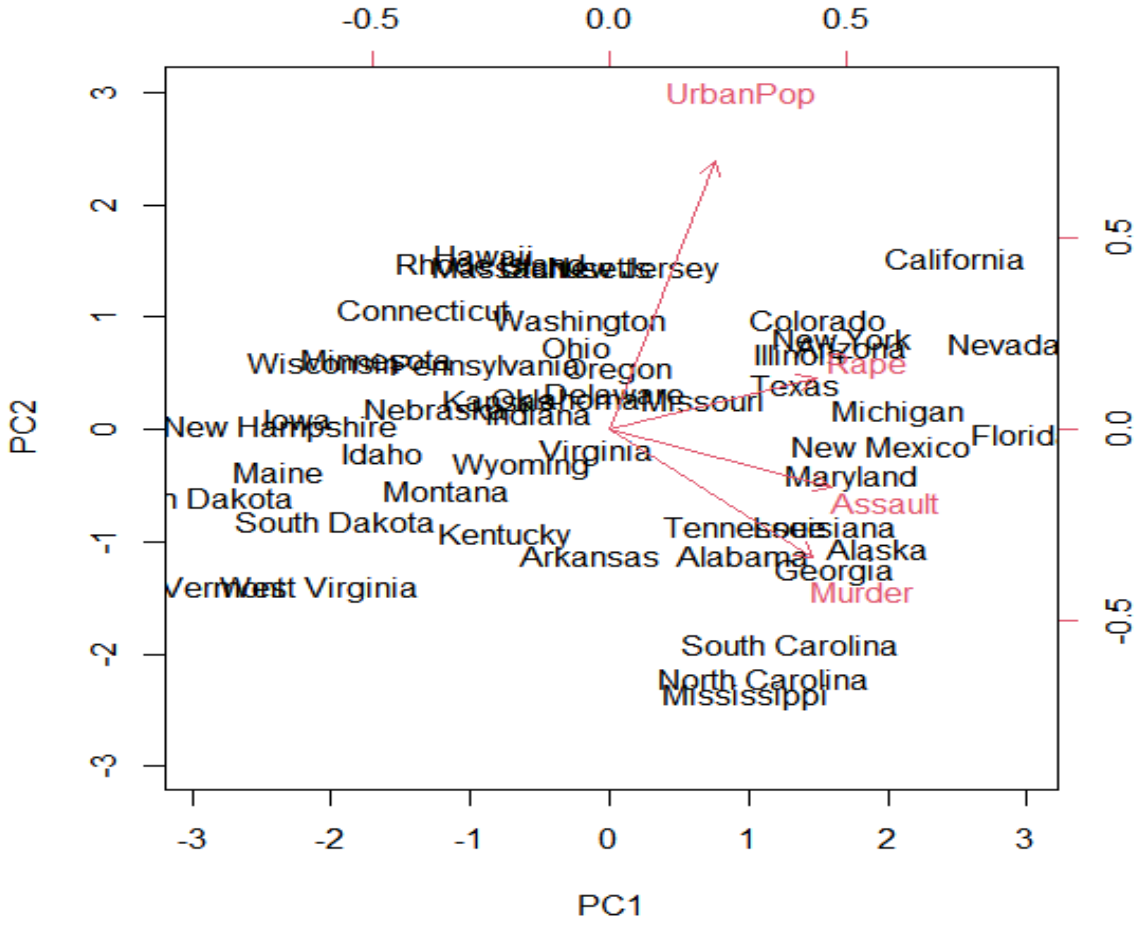
	PC1	PC2	PC3	PC4
Murder	-0.5358995	0.4181809	-0.3412327	0.64922780
Assault	-0.5831836	0.1879856	-0.2681484	-0.74340748
UrbanPop	-0.2781909	-0.8728062	-0.3780158	0.13387773
Rape	-0.5434321	-0.1673186	0.8177779	0.08902432

> dim(pr.out$x)

```
[1] 50  4
```

Question 2.

```
> pr.out$rotation=-pr.out$rotation
> pr.out$x=-pr.out$x
> biplot(pr.out, scale=0)
```



Question 2.

> pr.out$scale # the matrix of variable loadings (eigenvectors)

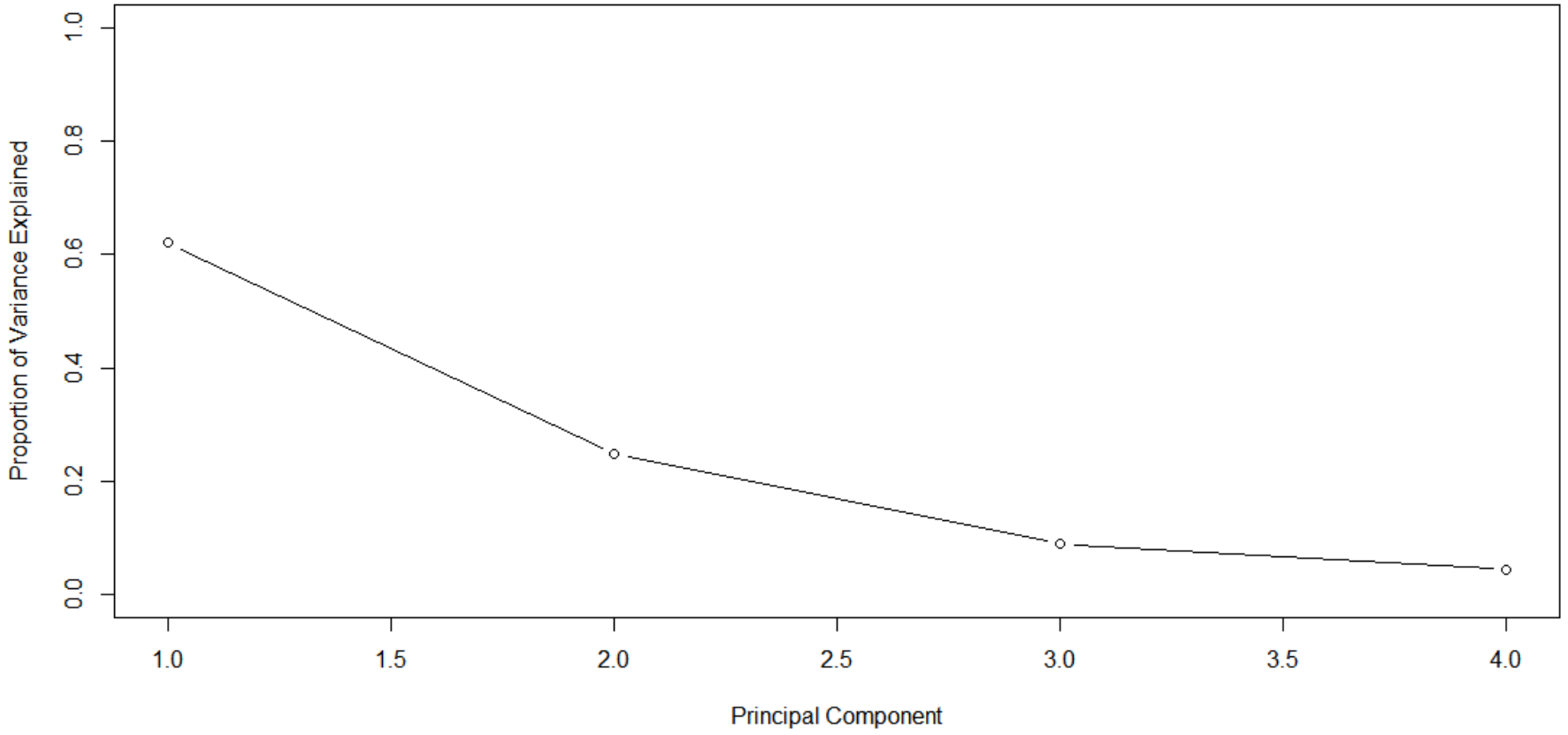
```
 Murder         Assault         UrbanPop         Rape
  4.355510      83.337661       14.474763       9.366385
```

```
> pr.out$sdev
  [1] 1.5748783 0.9948694 0.5971291 0.4164494
> pr.var=pr.out$sdev^2
> pr.var
  [1] 2.4802416 0.9897652 0.3565632 0.1734301

> pve=pr.var/sum(pr.var)
> pve
  [1] 0.62006039 0.24744129 0.08914080 0.04335752
```

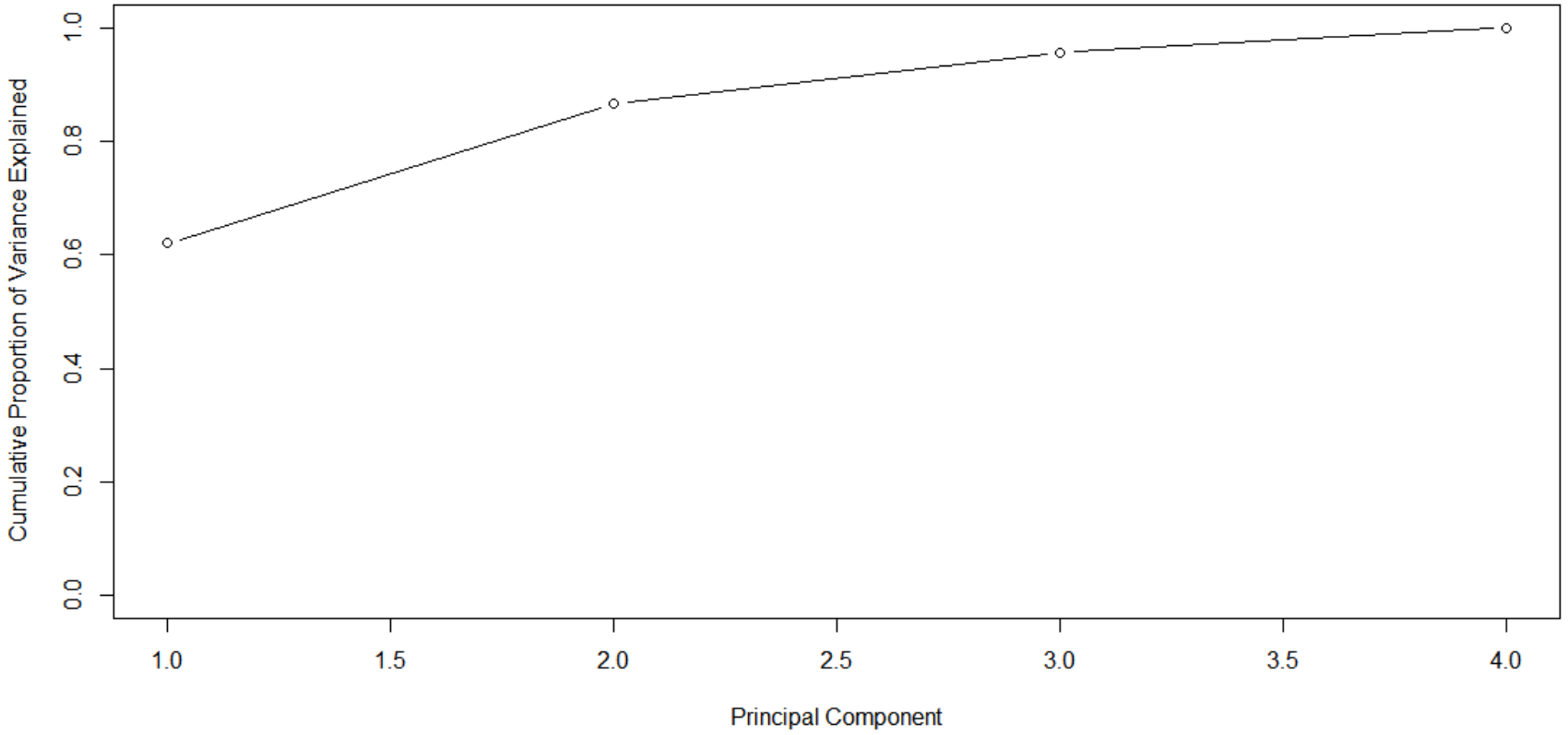
Question 2.

```
> plot(pve, xlab="Principal Component", ylab="Proportion of
Variance Explained", ylim=c(0,1),type='b')
```



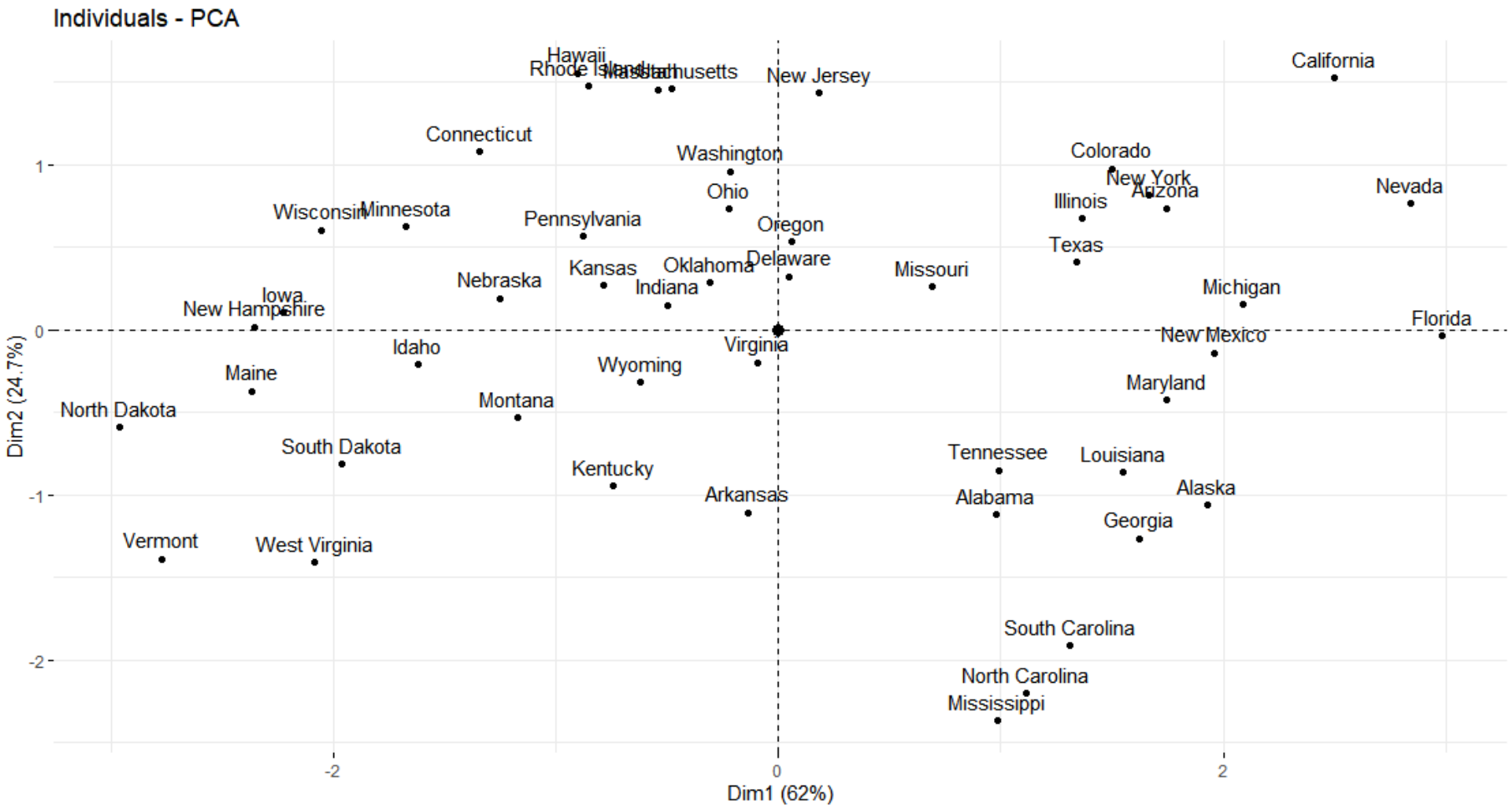
Question 2.

```
> plot(cumsum(pve), xlab="Principal Component", ylab="Cumulative
Proportion of Variance Explained", ylim=c(0,1),type='b')
```



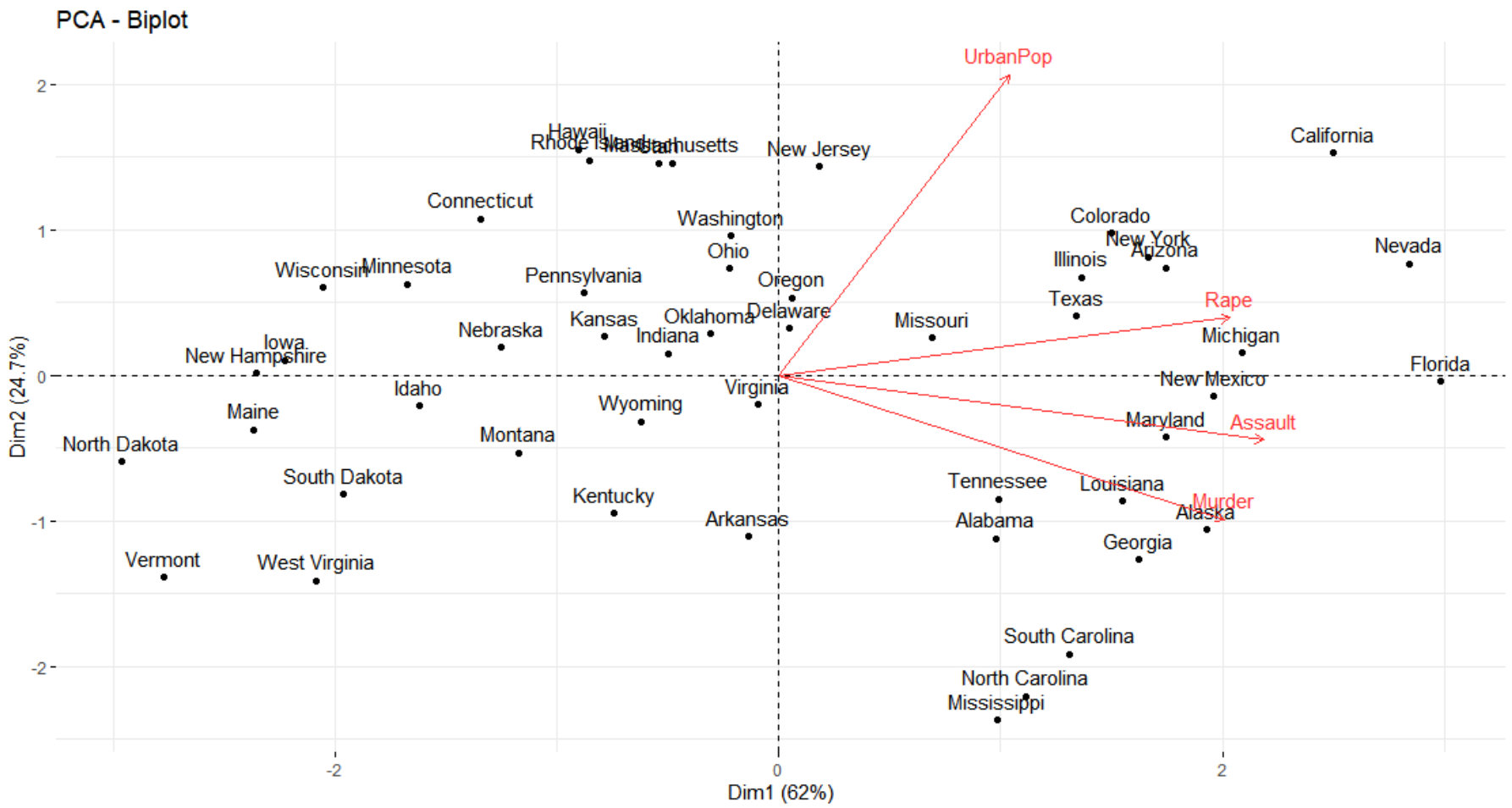
Question 2.

```
## Use factoextra package
library(factoextra)
fviz(pr.out, "ind", geom = "auto", mean.point = TRUE, font.family = "Georgia")
```



Question 2.

```
fviz_pca_biplot(pr.out, font.family = "Georgia", col.var="firebrick1")
```




Thank You