

27) Principal Components Analysis

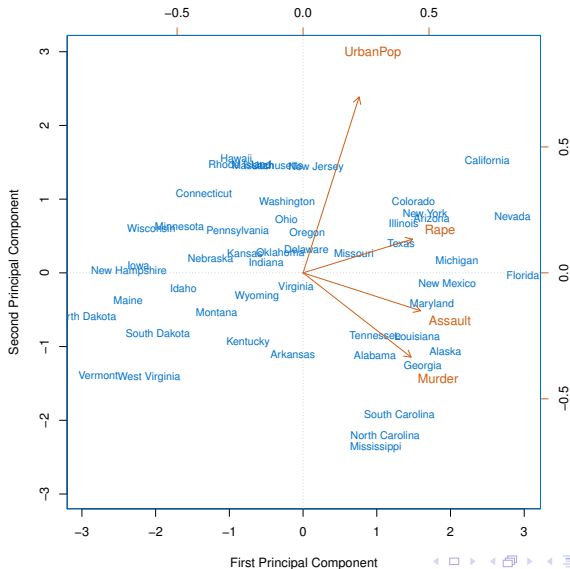
Vitor Kamada

February 2018

Tables, Graphics, and Figures from
An Introduction to Statistical Learning

James et al. (2017): Chapters: 10.2, and 10.4

USArrests Data



Principal Component Analysis (PCA)

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

$$z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + \dots + \phi_{p1}x_{ip}$$

$$\max_{\phi_{11}, \dots, \phi_{p1}} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{j1} x_{ij} \right)^2 \right\}$$

$$\text{subject to } \sum_{j=1}^p \phi_{j1}^2 = 1$$

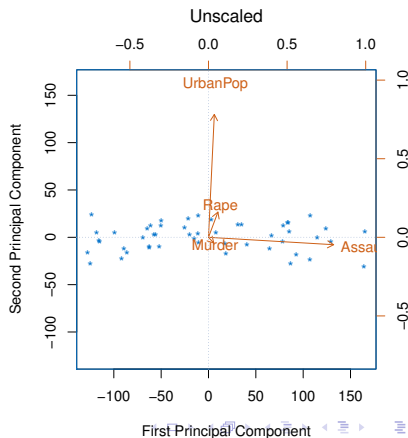
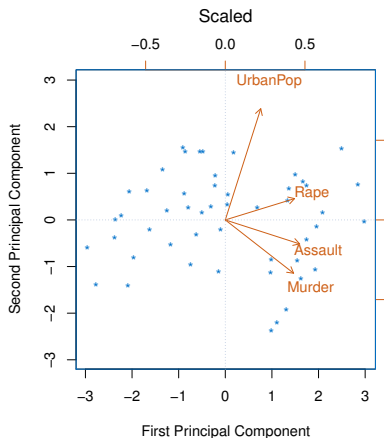
$$z_{i2} = \phi_{12}x_{i1} + \phi_{22}x_{i2} + \dots + \phi_{p2}x_{ip}$$

First and Second Principal Component

	PC1	PC2
Murder	0.5358995	-0.4181809
Assault	0.5831836	-0.1879856
UrbanPop	0.2781909	0.8728062
Rape	0.5434321	0.1673186

Scaling the Variables

Variance for Murder, Rape, Assault, and UrbanPop:
18.97, 87.73, 6945.16, and 209.5



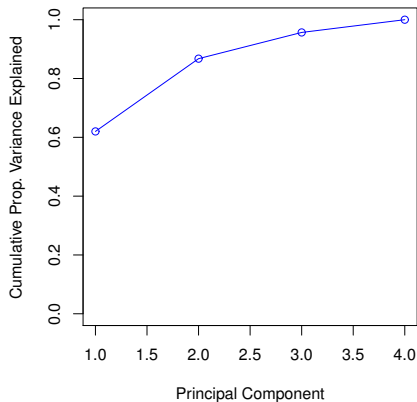
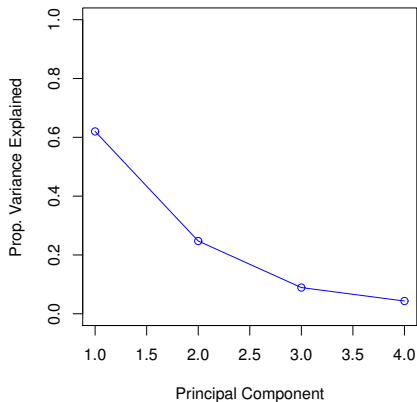
Proportion of Variance Explained (PVE)

$$PVE = \frac{\frac{1}{n} \sum_{i=1}^n z_{im}^2}{\sum_{j=1}^p \text{Var}(X_j)}$$

$$\sum_{j=1}^p \text{Var}(X_j) = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2$$

$$\frac{1}{n} \sum_{i=1}^n z_{im}^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{jm} x_{ij} \right)^2$$

Cumulative Proportion of Variance Explained




```
pr.out=prcomp(USArrests, scale=TRUE)
```

```
pr.out$center
```

Murder	Assault	UrbanPop	Rape
7.788	170.760	65.540	21.232

```
pr.out$scale
```

Murder	Assault	UrbanPop	Rape
4.355510	83.337661	14.474763	9.366385

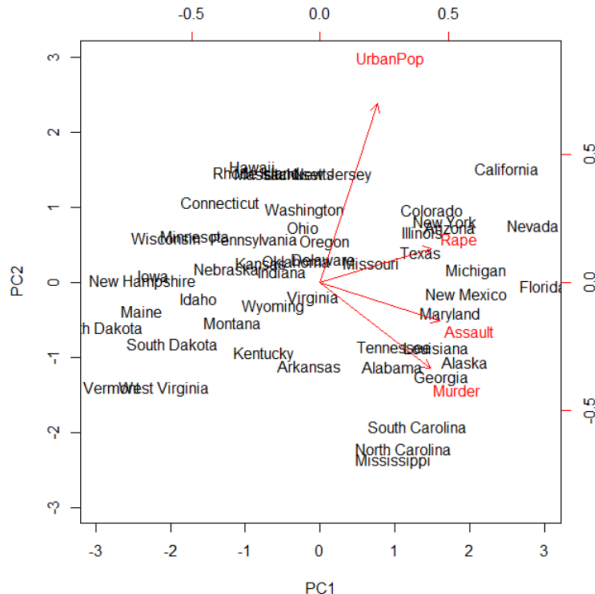
`pr.out$rotation=-pr.out$rotation`

`pr.out$x=-pr.out$x`

`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

biplot(pr.out, scale=0)



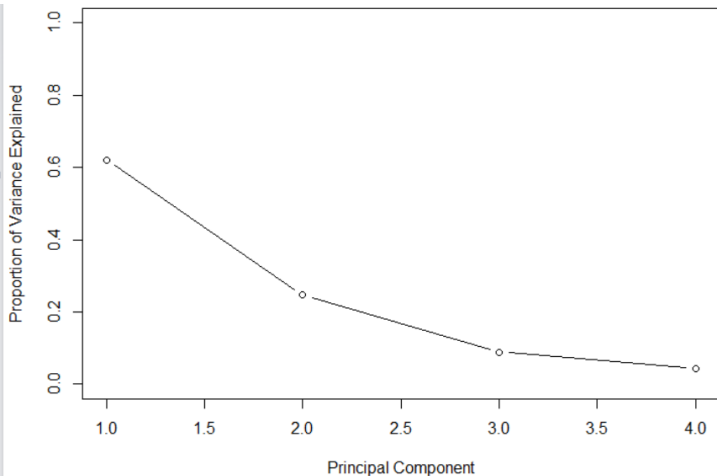
```
pr.var=pr.out$sdev^2; pr.var
```

```
2.4802416 0.9897652 0.3565632 0.1734301
```

```
pve=pr.var/sum(pr.var); pve
```

```
0.62006039 0.24744129 0.08914080 0.04335752
```

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



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

