Lab 1: Principal Components Analysis

Jonathan Bryan

August 31, 2018

```
states = row.names(USArrests)
states
```

```
##  [1] "Alabama"        "Alaska"         "Arizona"        "Arkansas"
##  [5] "California"     "Colorado"       "Connecticut"    "Delaware"
##  [9] "Florida"        "Georgia"        "Hawaii"         "Idaho"
## [13] "Illinois"       "Indiana"        "Iowa"           "Kansas"
## [17] "Kentucky"       "Louisiana"      "Maine"          "Maryland"
## [21] "Massachusetts"  "Michigan"       "Minnesota"      "Mississippi"
## [25] "Missouri"       "Montana"        "Nebraska"       "Nevada"
## [29] "New Hampshire"  "New Jersey"     "New Mexico"     "New York"
## [33] "North Carolina" "North Dakota"   "Ohio"           "Oklahoma"
## [37] "Oregon"         "Pennsylvania"   "Rhode Island"   "South Carolina"
## [41] "South Dakota"   "Tennessee"      "Texas"          "Utah"
## [45] "Vermont"        "Virginia"       "Washington"     "West Virginia"
## [49] "Wisconsin"      "Wyoming"
```

```
names(USArrests)
```

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

```
#examine mean and variance 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
```

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

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

```
#variable means
pr.out$center
```

```
##   Murder  Assault UrbanPop     Rape
##    7.788  170.760   65.540   21.232
```

```
#variable sd
pr.out$scale
```

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

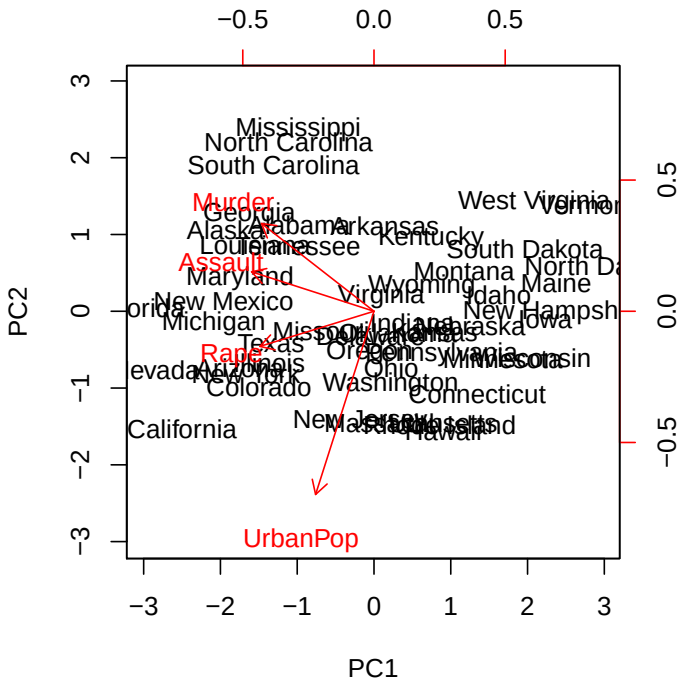
```
#variable loading vectors
pr.out$rotation
```

```
##                 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
```

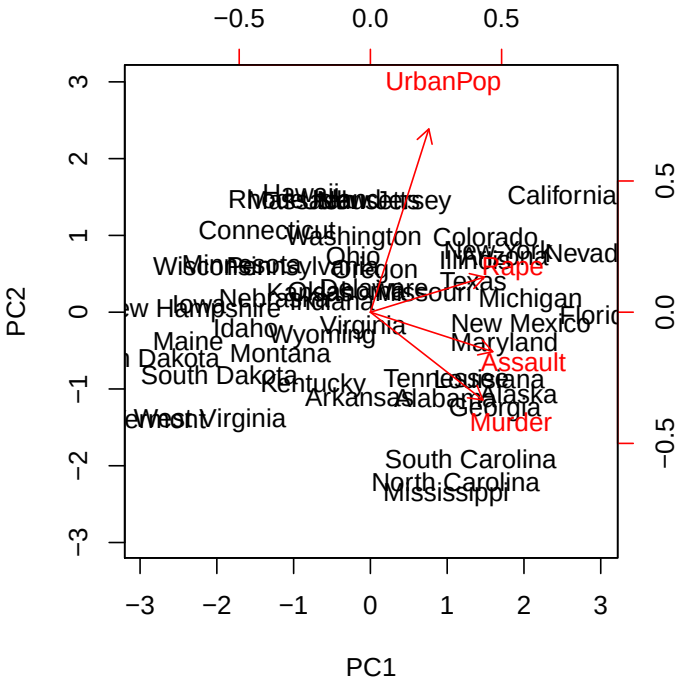
```
#dimensions of n x m score vectors
dim(pr.out$x)
```

```
## [1] 50  4
```

```
#plot of first two principal components
biplot(pr.out, scale = 0)
```




```
#flipping principal components
pr.out$rotation = -pr.out$rotation
pr.out$x = -pr.out$x
biplot(pr.out, scale = 0)
```




```
#principal component sd
round(pr.out$sdev,2)
```

```
## [1] 1.57 0.99 0.60 0.42
```

```
#principal component var
pr.var = pr.out$sdev^2
round(pr.var,2)
```

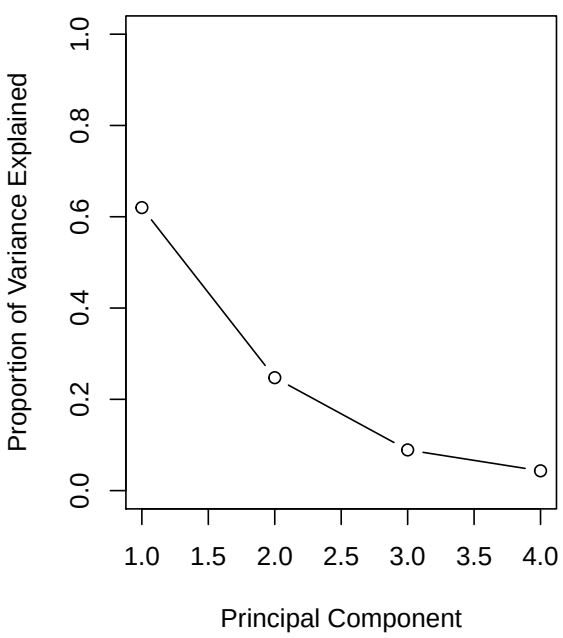
```
## [1] 2.48 0.99 0.36 0.17
```

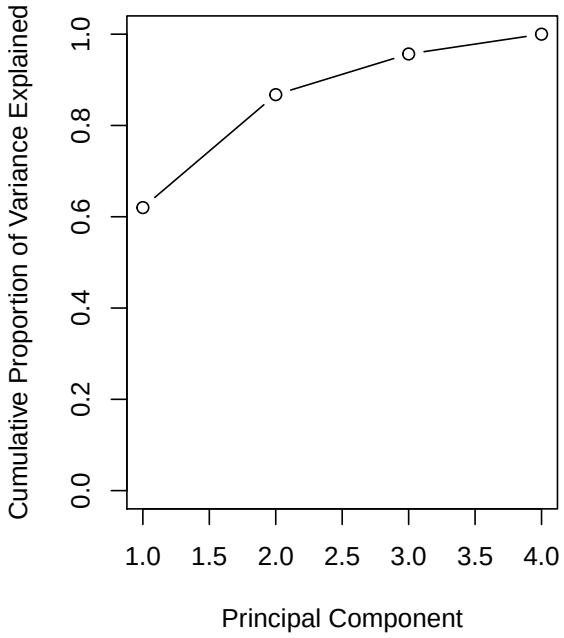
```
#proportion variance explained
pve = pr.var/sum(pr.var)
round(pve,2)
```

```
## [1] 0.62 0.25 0.09 0.04
```

```
par(mfrow = c(1,2))
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")
```






```
#cumulative sum function
a = c(1,2,8,-3)
cumsum(a)
```

```
## [1]  1  3 11  8
```