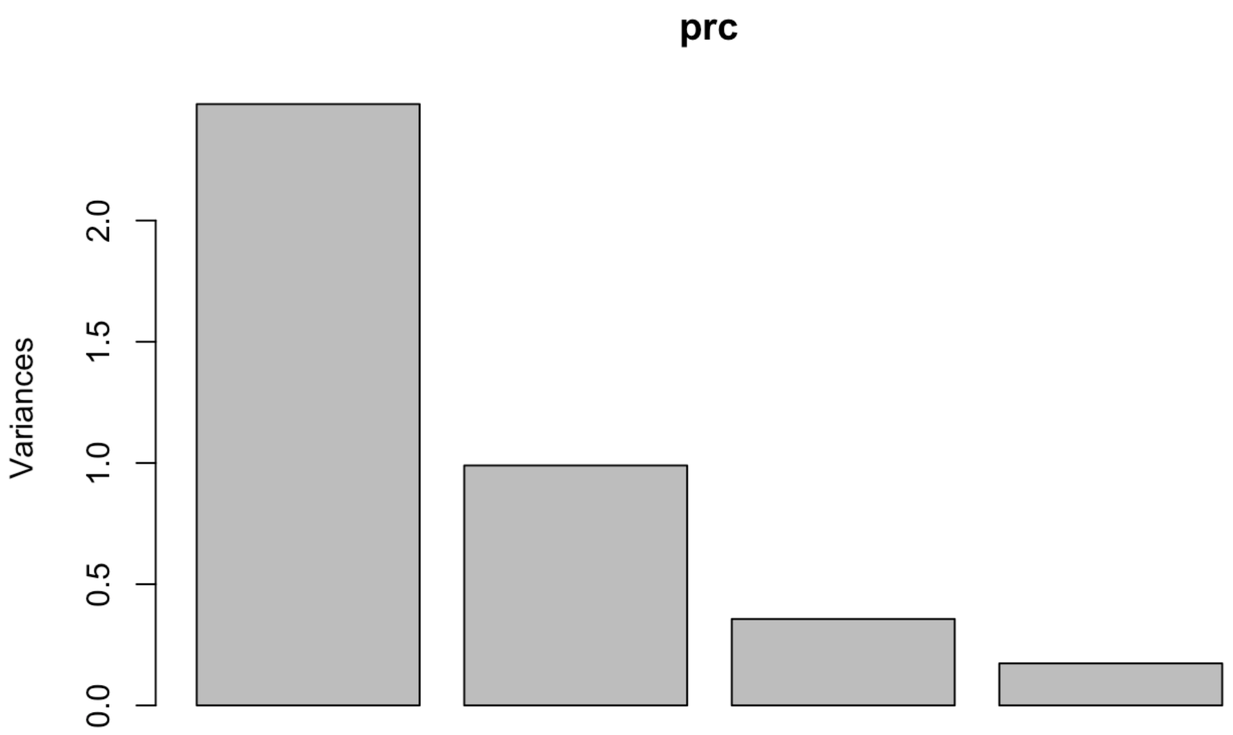
Practical Data Science: Reducing High Dimensional Data in R

Let’s start with prcomp (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html) and the example listed at the bottom of the page. (**Note**: the examples use the `USArrests` data set that is included in the stats package so you don’t have to download anything)

```
require(graphics)

# run prcomp on data set but scale all data first
prc <- prcomp(USArrests, scale = TRUE)
summary(prc)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4
## Standard deviation     1.5749 0.9949 0.59713 0.41645
## Proportion of Variance 0.6201 0.2474 0.08914 0.04336
## Cumulative Proportion  0.6201 0.8675 0.95664 1.00000
```

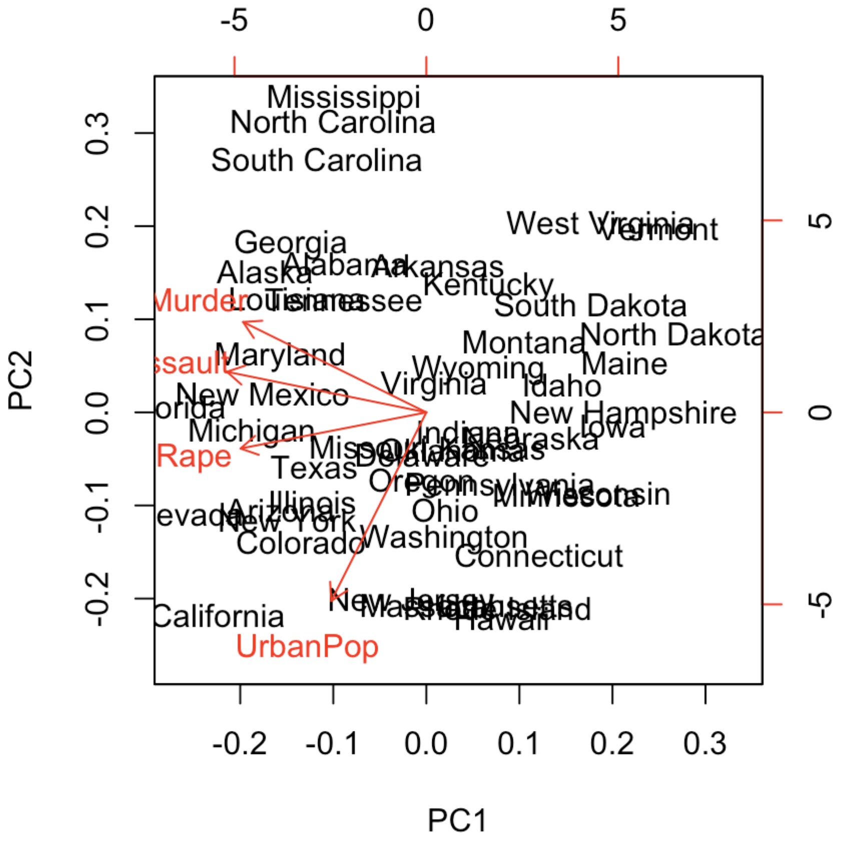
```
screeplot(prc)
```



```
# square the sdev to get the eigen value of each component
prc$sdev ^ 2 #
```

```
## [1] 2.4802416 0.9897652 0.3565632 0.1734301
```

```
# plot first two pcas along with feature correlations
biplot(prc)
```



```
# look at data
USArrests[order(USArrests$UrbanPop,decreasing=TRUE),]
```

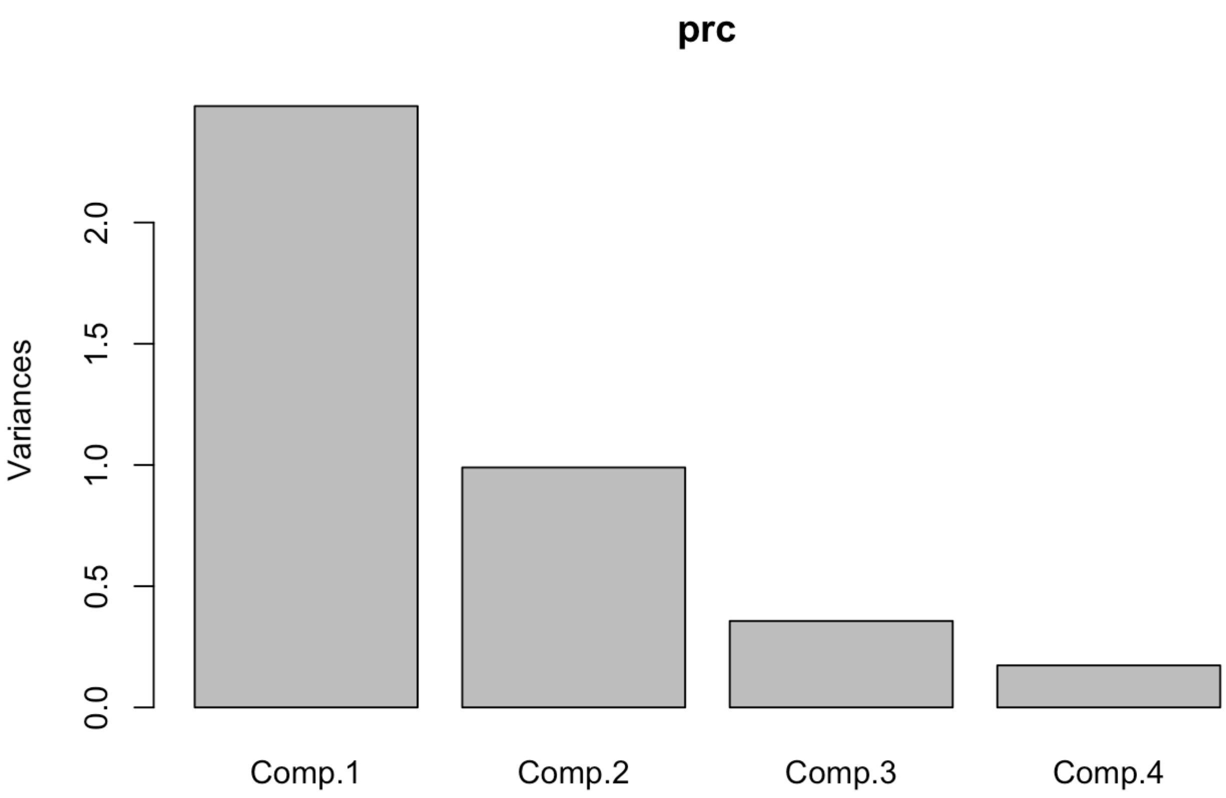
```
##                Murder Assault UrbanPop Rape
## California        9.0     276       91 40.6
## New Jersey        7.4     159       89 18.8
## Rhode Island      3.4     174       87  8.3
## New York         11.1     254       86 26.1
## Massachusetts     4.4     149       85 16.3
## Hawaii            5.3      46       83 20.2
## Illinois         10.4     249       83 24.0
## Nevada           12.2     252       81 46.0
## Arizona           8.1     294       80 31.0
## Florida          15.4     335       80 31.9
## Texas            12.7     201       80 25.5
## Utah              3.2     120       80 22.9
## Colorado          7.9     204       78 38.7
## Connecticut       3.3     110       77 11.1
## Ohio              7.3     120       75 21.4
## Michigan         12.1     255       74 35.1
## Washington        4.0     145       73 26.2
## Delaware          5.9     238       72 15.8
## Pennsylvania      6.3     106       72 14.9
## Missouri          9.0     178       70 28.2
## New Mexico       11.4     285       70 32.1
## Oklahoma          6.6     151       68 20.0
## Maryland         11.3     300       67 27.8
## Oregon            4.9     159       67 29.3
## Kansas            6.0     115       66 18.0
## Louisiana        15.4     249       66 22.2
## Minnesota         2.7      72       66 14.9
## Wisconsin         2.6      53       66 10.8
## Indiana           7.2     113       65 21.0
## Virginia          8.5     156       63 20.7
## Nebraska          4.3     102       62 16.5
## Georgia          17.4     211       60 25.8
## Wyoming           6.8     161       60 15.6
## Tennessee        13.2     188       59 26.9
## Alabama          13.2     236       58 21.2
## Iowa              2.2      56       57 11.3
## New Hampshire     2.1      57       56  9.5
## Idaho             2.6     120       54 14.2
## Montana           6.0     109       53 16.4
## Kentucky          9.7     109       52 16.3
## Maine             2.1      83       51  7.8
## Arkansas          8.8     190       50 19.5
## Alaska           10.0     263       48 44.5
## South Carolina   14.4     279       48 22.5
## North Carolina   13.0     337       45 16.1
## South Dakota      3.8      86       45 12.8
## Mississippi      16.1     259       44 17.1
## North Dakota      0.8      45       44  7.3
## West Virginia     5.7      81       39  9.3
## Vermont           2.2      48       32 11.2
```

Let’s take a look at example listed in princomp (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/princomp.html):

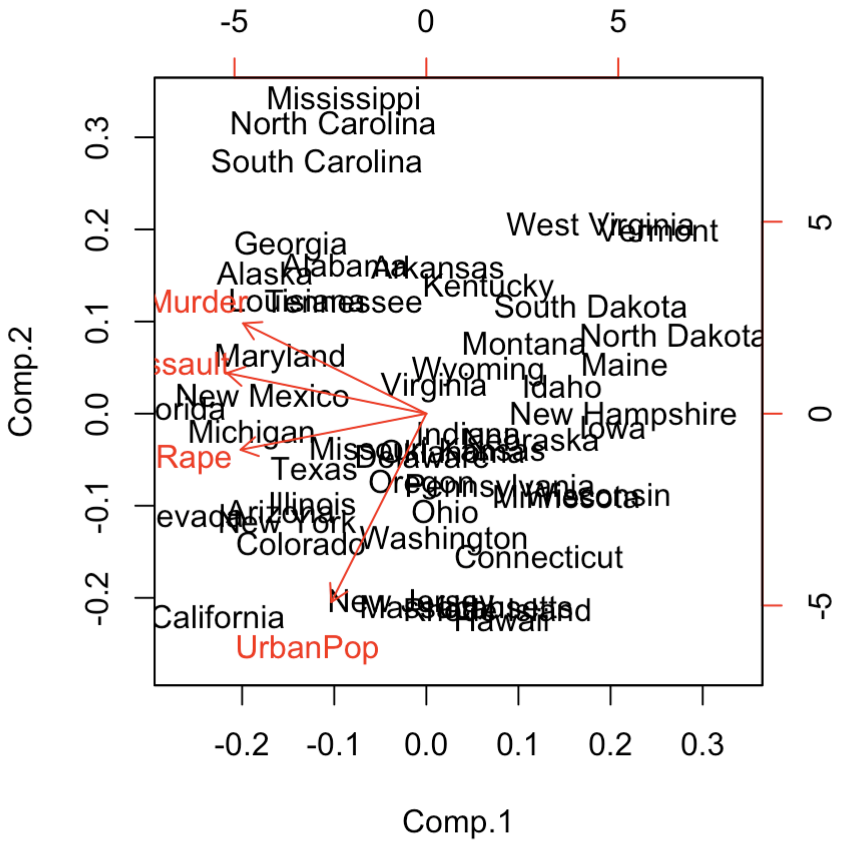
```
require(graphics)

prc <- princomp(USArrests, cor = TRUE, scale=TRUE)
```

```
## Warning: In princomp.default(USArrests, cor = TRUE, scale = TRUE) :
##  extra argument 'scale' will be disregarded
```

```
plot(prc) # shows a screeplot.
```



```
biplot(prc)
```



-5
0
5
0.3
0.2
0.1
0.0
-0.1
-0.2
5
0
-5
-0.2
-0.1
0.0
0.1
0.2
0.3
Comp.1
Comp.2
Mississippi
North Carolina
South Carolina
West Virginia
Vermont
Georgia
Alaska
Alabama
Arkansas
Kentucky
Murder
Louisiana
Tennessee
South Dakota
Montana
North Dakota
ssault
Maryland
Wyoming
Maine
Virginia
Idaho
New Mexico
orida
New Hampshire
Michigan
Indiana
Iowa
Nebraska
Rape
Missouri
Oklahoma
Delaware
Kansas
Texas
Oregon
Pennsylvania
Minnesota
Wisconsin
Illinois
Ohio
evada
Arizona
New York
Colorado
Washington
Connecticut
New Jersey
Massachusetts
Rhode Island
Hawaii
California
UrbanPop