PCA_510

Load USArrests data into usa

```
usa<-read.csv("c:/users/kun hu/desktop/USArrests.csv")
str(usa)
```

```
## 'data.frame':    50 obs. of  5 variables:
##  $ X       : Factor w/ 50 levels "Alabama","Alaska",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Murder  : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
##  $ Assault : int  236 263 294 190 276 204 110 238 335 211 ...
##  $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
##  $ Rape    : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

```
usa_pca<-princomp(~Murder+Assault+UrbanPop+Rape,data=usa)
attributes(usa_pca)
```

```
## $names
## [1] "sdev"     "loadings" "center"   "scale"    "n.obs"    "scores"
## [7] "call"
##
## $class
## [1] "princomp"
```

```
usa_pca$loadings
```

```
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4
## Murder                    0.995
## Assault  -0.995
## UrbanPop        -0.977 -0.201
## Rape            -0.201  0.974
##
##                Comp.1 Comp.2 Comp.3 Comp.4
## SS loadings      1.00   1.00   1.00   1.00
## Proportion Var   0.25   0.25   0.25   0.25
## Cumulative Var   0.25   0.50   0.75   1.00
```

```
usa_pca$center
```

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

```
usa_pca$scale
```

```
##   Murder  Assault UrbanPop     Rape
##        1        1        1        1
```

```
usa_pca$n.obs
```

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

```
usa_pca$scores
```

```
##         Comp.1      Comp.2       Comp.3     Comp.4
## 1   -64.802164  11.4480074  -2.49493284  2.4079009
## 2   -92.827450  17.9829427  20.12657487 -4.0940470
```

```
## 3  -124.068216  -8.8304030  -1.68744836 -4.3536852
## 4   -18.340035  16.7039114   0.21018936 -0.5209936
## 5  -107.422953 -22.5200698   6.74587299 -2.8118259
## 6   -34.975986 -13.7195840  12.27936280 -1.7214637
## 7    60.887282 -12.9325302  -8.42065719 -0.6999023
## 8   -66.731025  -1.3537978 -11.28095735 -3.7279812
## 9  -165.244370  -6.2746901  -2.99793315  1.2476807
## 10  -40.535177   7.2902396   3.60952946  7.3436728
## 11  123.536106 -24.2912079   3.72444284  3.4728494
## 12   51.797002   9.4691910  -1.52006356 -3.3478283
## 13  -78.992097 -12.8970605  -5.88326477  0.3676407
## 14   57.550961  -2.8462647   3.73816049  1.6494302
## 15  115.586790   3.3421305  -0.65402935 -0.8694960
## 16   55.789694  -3.1572339   0.38436416  0.6527917
## 17   62.383181  10.6732715   2.23708903  3.8762164
## 18  -78.277631   4.2949175  -3.82786965  4.4835590
## 19   89.261044  11.4878272  -4.69240562 -2.1161995
## 20 -129.330136   5.0070315  -2.34717282 -1.9283242
## 21   21.266283 -19.4501790  -7.50714835 -1.0348189
## 22  -85.451527  -5.9045567   6.46434210  0.4990479
## 23   98.954816  -5.2096006   0.00657376 -0.7318957
## 24  -86.856358  27.4284196  -5.00343624  3.8797577
## 25   -7.986289  -5.2756414   5.50057972  0.6794055
## 26   62.483635   9.5105021   1.83835536  0.2459426
## 27   69.096544   0.2111959   0.46802086 -0.6565664
## 28  -83.613578 -15.1021839  15.88869482  0.3341962
## 29  114.777355   4.7345584  -2.28238693 -0.9359106
## 30   10.815725 -23.1373389  -6.31015739  1.6124273
## 31 -114.868163   0.3364531   2.26126996 -1.3812478
## 32  -84.294231 -15.9239655  -4.72125960  0.8920194
## 33 -164.325514  31.0966153 -11.69616350 -2.1111927
## 34  127.495597  16.1350394  -1.31182982 -2.3009639
## 35   50.086822 -12.2793244   1.65733077  2.0291157
## 36   19.693723  -3.3701310  -0.45314329 -0.1803457
## 37   11.150240  -3.8660682   8.12998050 -2.9140109
## 38   64.689142  -8.9115466  -3.20646858  1.8749353
## 39   -3.063973 -18.3739704 -17.47001970 -2.3082597
## 40 -107.281069  23.5361159  -2.03279501  1.2517463
## 41   86.106720  16.5978586   1.31437998 -1.2522874
## 42  -17.506264   6.5065756   6.10012753  3.9228558
## 43  -31.291122 -12.9849566  -0.39340922  4.2420040
## 44   49.913397 -17.6484577   1.78816852 -1.8677052
## 45  124.714469  27.3135591   4.80277765 -2.0049857
## 46   14.817448   1.7526150   1.04538813  1.1738408
## 47   25.075839  -9.9679669   4.78112764 -2.6910819
## 48   91.544647  22.9528778  -0.40198344  0.7368781
## 49  118.176328  -5.5075792  -2.71132077  0.2049724
## 50   10.434539   5.9244529  -3.79444682 -0.5178674
```

```
usa_pca$call
```

```
## princomp(formula = ~Murder + Assault + UrbanPop + Rape, data = usa)
```

```r
summary(usa_pca)
```

```
## Importance of components:
##                             Comp.1      Comp.2      Comp.3       Comp.4
## Standard deviation     82.8908472 14.06956001 6.424204055 2.4578367034
## Proportion of Variance  0.9655342  0.02781734 0.005799535 0.0008489079
## Cumulative Proportion   0.9655342  0.99335156 0.999151092 1.0000000000
```

Alternative Method

```r
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
#only get numeric columns
usa_numeric<-usa[2:5]
usa_pca2<-preProcess(usa_numeric,method=c("BoxCox","center","scale","pca"),thresh=0.95)
attributes(usa_pca2)
```

```
## $names
##  [1] "dim"              "bc"               "yj"
##  [4] "et"               "invHyperbolicSine" "mean"
##  [7] "std"              "ranges"           "rotation"
## [10] "method"           "thresh"           "pcaComp"
## [13] "numComp"          "ica"              "wildcards"
## [16] "k"                "knnSummary"       "bagImp"
## [19] "median"           "data"
##
## $class
## [1] "preProcess"
```

```r
summary(usa_pca2)
```

```
##                   Length Class  Mode
## dim                2     -none- numeric
## bc                 4     -none- list
## yj                 0     -none- NULL
## et                 0     -none- NULL
## invHyperbolicSine  0     -none- NULL
## mean               4     -none- numeric
## std                4     -none- numeric
## ranges             0     -none- NULL
## rotation          12     -none- numeric
## method             5     -none- list
## thresh             1     -none- numeric
## pcaComp            0     -none- NULL
## numComp            1     -none- numeric
## ica                0     -none- NULL
## wildcards          2     -none- list
## k                  1     -none- numeric
## knnSummary         1     -none- function
## bagImp             0     -none- NULL
## median             0     -none- NULL
```

```
## data               0     -none- NULL
```

```
usa_num_pca<-predict(usa_pca2,usa_numeric)
head(usa_num_pca)
```

```
##           PC1        PC2         PC3
## 1 -1.03453112  1.0848115 -0.33595844
## 2 -1.67318523  1.3179300  1.37135111
## 3 -1.75849483 -0.6756439 -0.03729575
## 4 -0.06792624  1.1731575  0.05737966
## 5 -2.38210391 -1.4475095  0.30790267
## 6 -1.50855460 -0.7761294  0.83838786
```