12 - PCA Additional Example

Junvie Pailden

SIUE, F2017, Stat 589

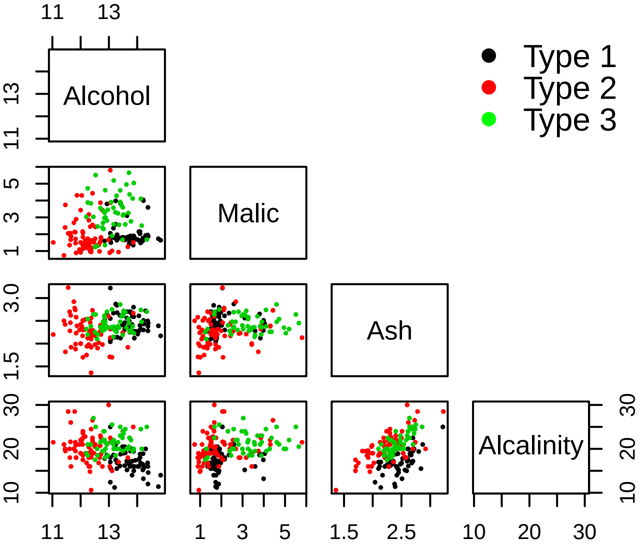
October 02, 2017

Wine Data

The wine data is the result of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

```r
data(wine, package = 'rattle.data')
names(wine)
```

```
#  [1] "Type"            "Alcohol"         "Malic"
#  [4] "Ash"             "Alcalinity"      "Magnesium"
#  [7] "Phenols"         "Flavanoids"      "Nonflavanoids"
# [10] "Proanthocyanins" "Color"           "Hue"
# [13] "Dilution"        "Proline"
```


Type 1
Type 2
Type 3
Alcohol
Malic
Ash
Alcalinity
11
13
1
3
5
1.5
2.5
3.0
10
20
30

PCA Wine Data using `prcomp`

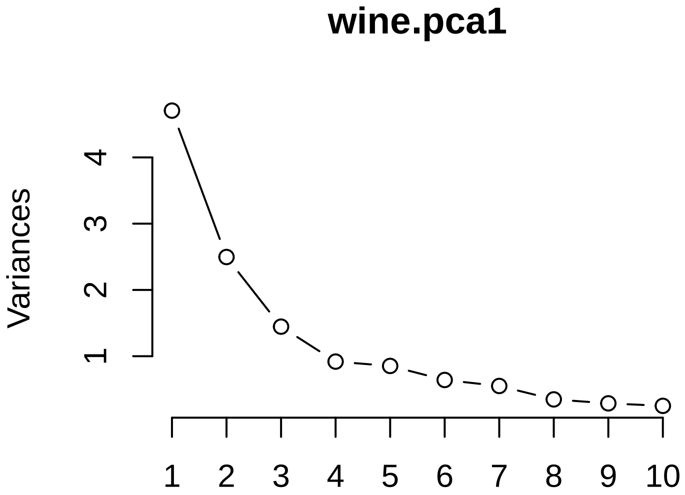
```
# need to center and scale the data
wine.pca1 <- prcomp(wine[,-1], scale. = T, center = T)
names(wine.pca1)
```

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

```
# Variances/Eigenvalues of PCA
(wine.pca1$sdev)^2
```

```
#  [1] 4.71 2.50 1.45 0.92 0.85 0.64 0.55 0.35 0.29 0.25 0.
```

```r
screeplot(wine.pca1, type="lines")
```



Summary of PCA

```r
summary(wine.pca1)
```

```
# Importance of components%s:
#                           PC1   PC2   PC3    PC4    PC5
# Standard deviation     2.169 1.580 1.203 0.9586 0.9237 0.
# Proportion of Variance 0.362 0.192 0.111 0.0707 0.0656 0.
# Cumulative Proportion  0.362 0.554 0.665 0.7360 0.8016 0.
#                           PC8    PC9   PC10   PC11  PC12
# Standard deviation     0.5903 0.5375 0.5009 0.4752 0.411
# Proportion of Variance 0.0268 0.0222 0.0193 0.0174 0.013
# Cumulative Proportion  0.9202 0.9424 0.9617 0.9791 0.992
```

First Four Eigenvectors (or Loadings) using `prcomp`

```
wine.pca1$rotation[,1:4]
```

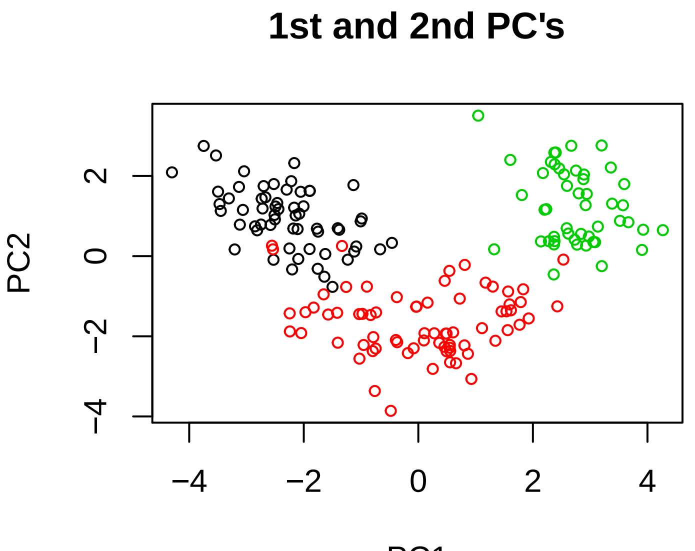
```
#                      PC1     PC2    PC3    PC4
# Alcohol          -0.1443  0.4837 -0.207  0.018
# Malic             0.2452  0.2249  0.089 -0.537
# Ash               0.0021  0.3161  0.626  0.214
# Alcalinity        0.2393 -0.0106  0.612 -0.061
# Magnesium        -0.1420  0.2996  0.131  0.352
# Phenols          -0.3947  0.0650  0.146 -0.198
# Flavanoids       -0.4229 -0.0034  0.151 -0.152
# Nonflavanoids     0.2985  0.0288  0.170  0.203
# Proanthocyanins -0.3134  0.0393  0.149 -0.399
# Color             0.0886  0.5300 -0.137 -0.066
# Hue              -0.2967 -0.2792  0.085  0.428
# Dilution         -0.3762 -0.1645  0.166 -0.184
# Proline          -0.2868  0.3649 -0.127  0.232
```

First Four Principal Components (or Scores) using prcomp

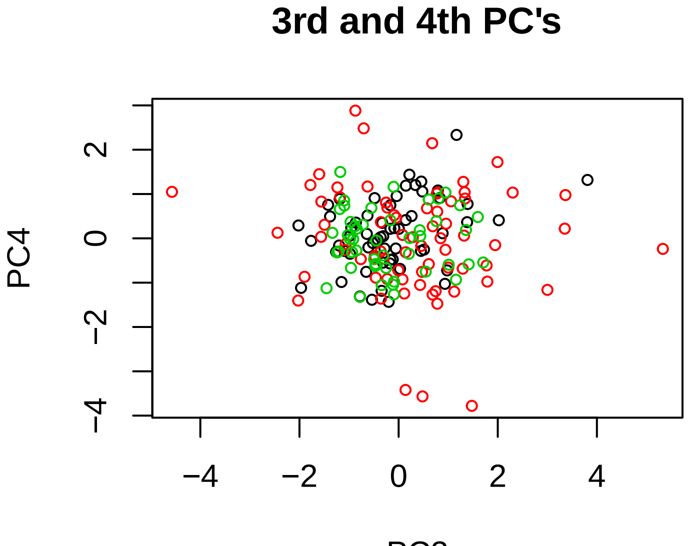
```
head(wine.pca1$x[,1:4], 10)
```

```
#         PC1   PC2   PC3    PC4
#  [1,] -3.3  1.44 -0.17  0.215
#  [2,] -2.2 -0.33 -2.02  0.291
#  [3,] -2.5  1.03  0.98 -0.723
#  [4,] -3.7  2.75 -0.18 -0.566
#  [5,] -1.0  0.87  2.02  0.409
#  [6,] -3.0  2.12 -0.63  0.514
#  [7,] -2.4  1.17 -0.97  0.066
#  [8,] -2.1  1.60  0.15  1.189
#  [9,] -2.5  0.92 -1.77 -0.056
# [10,] -2.7  0.79 -0.98 -0.348
```

```
plot(wine.pca1$x[,1:2], col = wine$Type, cex = 0.7, main =
```



```
plot(wine.pca1$x[,3:4], col = wine$Type, cex = 0.7, main =
```



PCA Wine Data using princomp

```r
# set cor = TRUE to use scaled data
wine.pca2 <- princomp(wine[,-1], cor = TRUE)
names(wine.pca2)
```

```
# [1] "sdev"     "loadings" "center"   "scale"    "n.obs"
# [7] "call"
```

```r
# Variances/Eigenvalues of PCA
(wine.pca2$sdev)^2
```

```
#  Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7
#    4.71    2.50    1.45    0.92    0.85    0.64    0.55
# Comp.10 Comp.11 Comp.12 Comp.13
#    0.25    0.23    0.17    0.10
```

First Four Eigenvectors (or Loadings) using `princomp`

```
wine.pca2$loadings[,1:4]
```

```
#                  Comp.1  Comp.2 Comp.3 Comp.4
# Alcohol         -0.1443 -0.4837 -0.207 -0.018
# Malic            0.2452 -0.2249  0.089  0.537
# Ash              0.0021 -0.3161  0.626 -0.214
# Alcalinity       0.2393  0.0106  0.612  0.061
# Magnesium       -0.1420 -0.2996  0.131 -0.352
# Phenols         -0.3947 -0.0650  0.146  0.198
# Flavanoids      -0.4229  0.0034  0.151  0.152
# Nonflavanoids    0.2985 -0.0288  0.170 -0.203
# Proanthocyanins -0.3134 -0.0393  0.149  0.399
# Color            0.0886 -0.5300 -0.137  0.066
# Hue             -0.2967  0.2792  0.085 -0.428
# Dilution        -0.3762  0.1645  0.166  0.184
# Proline         -0.2868 -0.3649 -0.127 -0.232
```

First Four Principal Components (or Scores) using `princomp`

```
head(wine.pca2$scores[,1:4], 10)
```

```
#       Comp.1 Comp.2 Comp.3 Comp.4
#  [1,]   -3.3  -1.44  -0.17 -0.216
#  [2,]   -2.2   0.33  -2.03 -0.291
#  [3,]   -2.5  -1.03   0.98  0.725
#  [4,]   -3.8  -2.76  -0.18  0.568
#  [5,]   -1.0  -0.87   2.03 -0.410
#  [6,]   -3.1  -2.12  -0.63 -0.516
#  [7,]   -2.4  -1.17  -0.98 -0.066
#  [8,]   -2.1  -1.61   0.15 -1.193
#  [9,]   -2.5  -0.92  -1.77  0.056
# [10,]   -2.8  -0.79  -0.98  0.349
```