

Principal Component Analysis in R

43141 (Sahil Naphade)

02/12/2020

Reading the Wine dataset and checking

```
setwd("E:/College/CL7/Assignment12")
winedf <- read.csv("../Datasets/wine.csv")
head(winedf)
```

```
##   Wine Alcohol Malic.acid  Ash  Acl  Mg Phenols Flavanoids Nonflavanoid.phenols
## 1    1   14.23      1.71 2.43 15.6 127   2.80      3.06                0.28
## 2    1   13.20      1.78 2.14 11.2 100   2.65      2.76                0.26
## 3    1   13.16      2.36 2.67 18.6 101   2.80      3.24                0.30
## 4    1   14.37      1.95 2.50 16.8 113   3.85      3.49                0.24
## 5    1   13.24      2.59 2.87 21.0 118   2.80      2.69                0.39
## 6    1   14.20      1.76 2.45 15.2 112   3.27      3.39                0.34
##   Proanth Color.int  Hue   OD Proline
## 1    2.29      5.64 1.04 3.92   1065
## 2    1.28      4.38 1.05 3.40   1050
## 3    2.81      5.68 1.03 3.17   1185
## 4    2.18      7.80 0.86 3.45   1480
## 5    1.82      4.32 1.04 2.93    735
## 6    1.97      6.75 1.05 2.85   1450
```

Calculate Correlation of all attributes

```
cor_mat <- cor(winedf[, -1])
cor_mat
```

```
##           Alcohol Malic.acid      Ash      Acl
## Alcohol      1.00000000  0.09439694  0.211544596 -0.31023514
## Malic.acid    0.09439694  1.00000000  0.164045470  0.28850040
## Ash           0.21154460  0.16404547  1.000000000  0.44336719
## Acl          -0.31023514  0.28850040  0.443367187  1.00000000
## Mg           0.27079823 -0.05457510  0.286586691 -0.08333309
## Phenols       0.28910112 -0.33516700  0.128979538 -0.32111332
## Flavanoids    0.23681493 -0.41100659  0.115077279 -0.35136986
## Nonflavanoid.phenols -0.15592947  0.29297713  0.186230446  0.36192172
## Proanth       0.13669791 -0.22074619  0.009651935 -0.19732684
## Color.int     0.54636420  0.24898534  0.258887259  0.01873198
## Hue          -0.07174720 -0.56129569 -0.074666889 -0.27395522
## OD            0.07234319 -0.36871043  0.003911231 -0.27676855
## Proline       0.64372004 -0.19201056  0.223626264 -0.44059693
```

```

##              Mg      Phenols Flavanoids Nonflavanoid.phenols
## Alcohol      0.27079823  0.28910112  0.2368149      -0.1559295
## Malic.acid   -0.05457510 -0.33516700 -0.4110066      0.2929771
## Ash          0.28658669  0.12897954  0.1150773      0.1862304
## Acl          -0.08333309 -0.32111332 -0.3513699      0.3619217
## Mg           1.00000000  0.21440123  0.1957838      -0.2562940
## Phenols      0.21440123  1.00000000  0.8645635      -0.4499353
## Flavanoids   0.19578377  0.86456350  1.0000000      -0.5378996
## Nonflavanoid.phenols -0.25629405 -0.44993530 -0.5378996      1.0000000
## Proanth      0.23644061  0.61241308  0.6526918      -0.3658451
## Color.int    0.19995001 -0.05513642 -0.1723794      0.1390570
## Hue          0.05539820  0.43368134  0.5434786      -0.2626396
## OD           0.06600394  0.69994936  0.7871939      -0.5032696
## Proline      0.39335085  0.49811488  0.4941931      -0.3113852
##              Proanth  Color.int      Hue      OD
## Alcohol      0.136697912  0.54636420 -0.07174720  0.072343187
## Malic.acid   -0.220746187  0.24898534 -0.56129569 -0.368710428
## Ash          0.009651935  0.25888726 -0.07466689  0.003911231
## Acl          -0.197326836  0.01873198 -0.27395522 -0.276768549
## Mg           0.236440610  0.19995001  0.05539820  0.066003936
## Phenols      0.612413084 -0.05513642  0.43368134  0.699949365
## Flavanoids   0.652691769 -0.17237940  0.54347857  0.787193902
## Nonflavanoid.phenols -0.365845099  0.13905701 -0.26263963 -0.503269596
## Proanth      1.000000000 -0.02524993  0.29554425  0.519067096
## Color.int    -0.025249931  1.00000000 -0.52181319 -0.428814942
## Hue          0.295544253 -0.52181319  1.00000000  0.565468293
## OD           0.519067096 -0.42881494  0.56546829  1.000000000
## Proline      0.330416700  0.31610011  0.23618345  0.312761075
##              Proline
## Alcohol      0.6437200
## Malic.acid   -0.1920106
## Ash          0.2236263
## Acl          -0.4405969
## Mg           0.3933508
## Phenols      0.4981149
## Flavanoids   0.4941931
## Nonflavanoid.phenols -0.3113852
## Proanth      0.3304167
## Color.int    0.3161001
## Hue          0.2361834
## OD           0.3127611
## Proline      1.0000000

```

We see that alcohol and Proline, alcohol and Color.int has fairly high correlation. Similarly, Phenols with Proanth, and Flavanoids with Proanth has high correlation. Flavanoids and nonflavanoids phenols has high negative correlation.

Checking the principal components of the dataset.

```

WPCA <- princomp(winedf[, -1], cor = T, scores = T, covmat = NULL)
summary(WPCA)

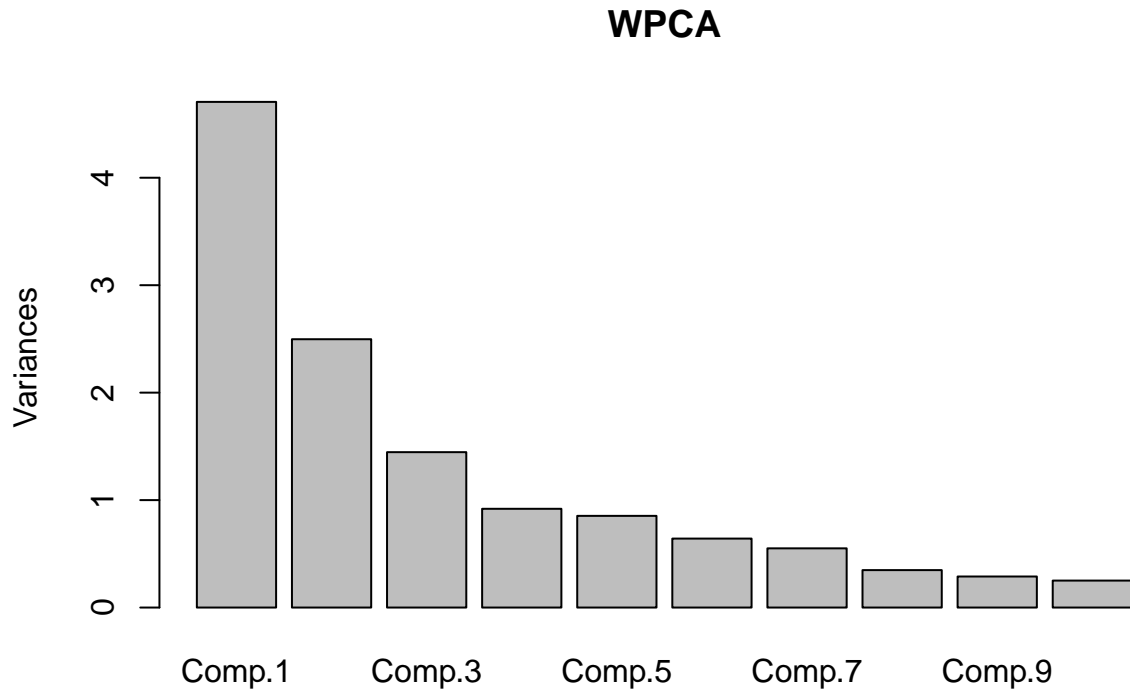
```

```
## Importance of components:
```

```
##               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation  2.1692972 1.5801816 1.2025273 0.9586313 0.92370351
## Proportion of Variance 0.3619885 0.1920749 0.1112363 0.0706903 0.06563294
## Cumulative Proportion 0.3619885 0.5540634 0.6652997 0.7359900 0.80162293
##               Comp.6   Comp.7   Comp.8   Comp.9   Comp.10
## Standard deviation  0.80103498 0.74231281 0.59033665 0.53747553 0.50090167
## Proportion of Variance 0.04935823 0.04238679 0.02680749 0.02222153 0.01930019
## Cumulative Proportion 0.85098116 0.89336795 0.92017544 0.94239698 0.96169717
##               Comp.11   Comp.12   Comp.13
## Standard deviation  0.47517222 0.41081655 0.321524394
## Proportion of Variance 0.01736836 0.01298233 0.007952149
## Cumulative Proportion 0.97906553 0.99204785 1.000000000
```

Plot the WPCA graph

```
plot(WPCA)
```



(component coefficients) correlation coefficients between variables(rows) and factors(columns)

```
loadings(WPCA)
```

```
##
## Loadings:
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
```

```

## Alcohol      0.144  0.484  0.207      0.266  0.214      0.396
## Malic.acid   -0.245  0.225      -0.537      0.537 -0.421
## Ash          0.316 -0.626  0.214  0.143  0.154  0.149 -0.170
## Acl          -0.239      -0.612      -0.101  0.287  0.428
## Mg           0.142  0.300 -0.131  0.352 -0.727      -0.323 -0.156
## Phenols      0.395      -0.146 -0.198  0.149      -0.406
## Flavanoids   0.423      -0.151 -0.152  0.109      -0.187
## Nonflavanoid.phenols -0.299      -0.170  0.203  0.501 -0.259 -0.595 -0.233
## Proanth      0.313      -0.149 -0.399 -0.137 -0.534 -0.372  0.368
## Color.int    0.530  0.137      -0.419  0.228
## Hue          0.297 -0.279      0.428  0.174  0.106 -0.232  0.437
## OD           0.376 -0.164 -0.166 -0.184  0.101  0.266
## Proline      0.287  0.365  0.127  0.232  0.158  0.120      0.120
##              Comp.9 Comp.10 Comp.11 Comp.12 Comp.13
## Alcohol      0.509  0.212  0.226  0.266
## Malic.acid   -0.309      -0.122
## Ash          -0.308      0.499      -0.141
## Acl          0.200      -0.479
## Mg           0.271
## Phenols      0.286 -0.320 -0.304  0.304 -0.464
## Flavanoids   -0.163      0.832
## Nonflavanoid.phenols 0.196  0.216 -0.117      0.114
## Proanth      -0.209  0.134  0.237      -0.117
## Color.int    -0.291      -0.604
## Hue          -0.522      -0.259
## OD           0.137  0.524      -0.601 -0.157
## Proline      -0.576  0.162 -0.539
##
##              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## SS loadings   1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var 0.077  0.077  0.077  0.077  0.077  0.077  0.077  0.077  0.077
## Cumulative Var 0.077  0.154  0.231  0.308  0.385  0.462  0.538  0.615  0.692
##              Comp.10 Comp.11 Comp.12 Comp.13
## SS loadings   1.000  1.000  1.000  1.000
## Proportion Var 0.077  0.077  0.077  0.077
## Cumulative Var 0.769  0.846  0.923  1.000

```

Checking the attributes of the Principal Component Analysis

```
attributes(WPCA)
```

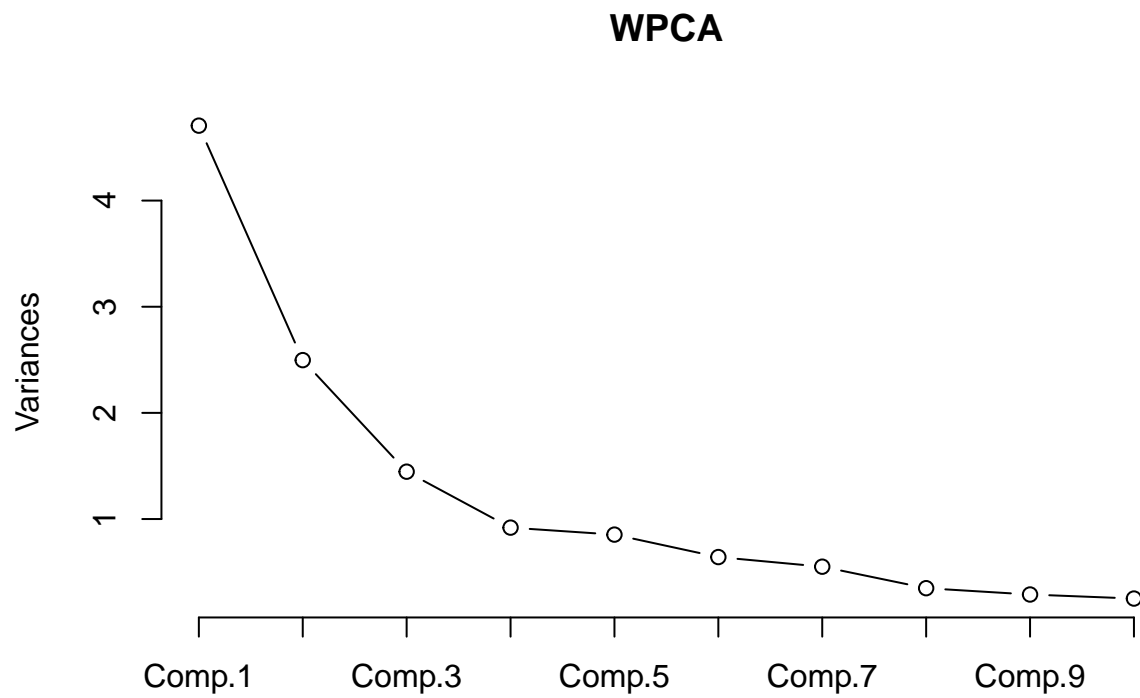
```

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

```

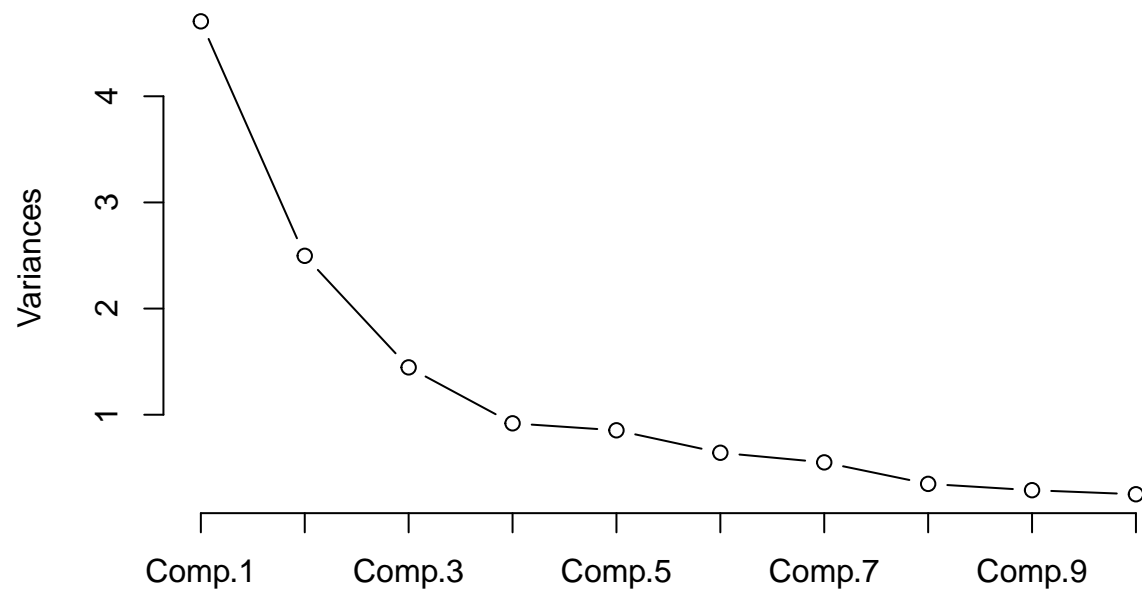
Plotting various graphs and scree graph to check the variation of information in the variables/components.

```
plot(WPCA, t='l')
```

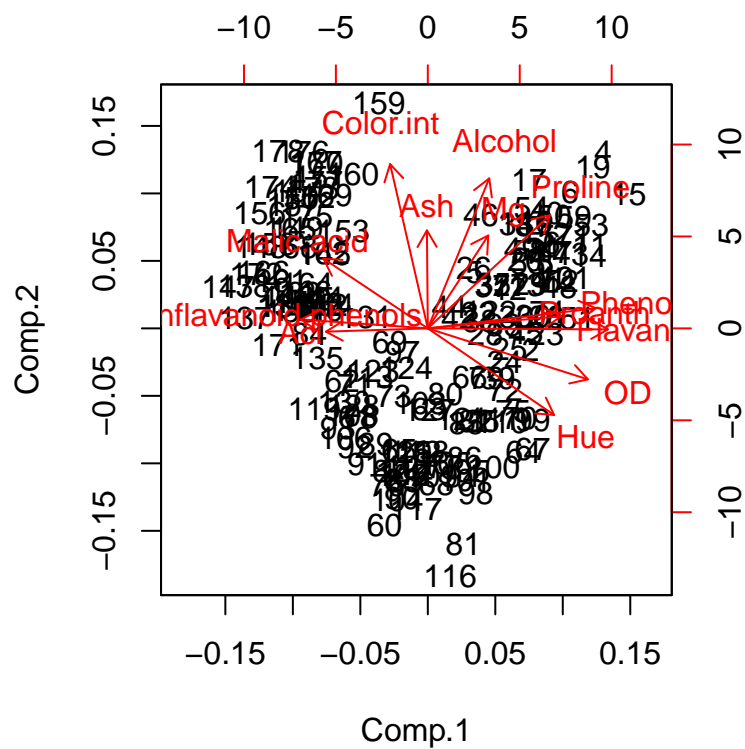


```
screeplot(WPCA, type="line", main="Scree plot")
```

Scree plot



```
biplot(WPCA)
```



```
## Check the score of PCA
```

```
WPCA$scores[1:10, 1]
```

```
## [1] 3.316751 2.209465 2.516740 3.757066 1.008908 3.050254 2.449090 2.059437
## [9] 2.510874 2.753628
```