R Notebook

Code

Dataset z https://rpubs.com/joelnudinas03/WQD

Obsahuje merenii ruznych parametru vin a jejich hodnoceni (kvalitu).

Hide

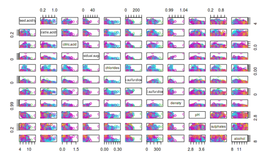
```
winequality <- read.csv("http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv", sep = ";")
summary(winequality)
```

```
 fixed.acidity    volatile.acidity  citric.acid     residual.sugar     chlorides       free.sulfur.dioxide total.
sulfur.dioxide    density             pH
 Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600   Min.   :0.00900   Min.   :  2.00    Min.
:  9.0    Min.   :0.9871   Min.   :2.720
 1st Qu.: 6.300   1st Qu.:0.2100   1st Qu.:0.2700   1st Qu.: 1.700   1st Qu.:0.03600   1st Qu.: 23.00    1st Q
u.:108.0        1st Qu.:0.9917   1st Qu.:3.090
 Median : 6.800   Median :0.2600   Median :0.3200   Median : 5.200   Median :0.04300   Median : 34.00    Median
:134.0          Median :0.9937   Median :3.180
 Mean   : 6.855   Mean   :0.2782   Mean   :0.3342   Mean   : 6.391   Mean   :0.04577   Mean   : 35.31    Mean
:138.4          Mean   :0.9940   Mean   :3.188
 3rd Qu.: 7.300   3rd Qu.:0.3200   3rd Qu.:0.3900   3rd Qu.: 9.900   3rd Qu.:0.05000   3rd Qu.: 46.00    3rd Q
u.:167.0        3rd Qu.:0.9961   3rd Qu.:3.280
 Max.   :14.200   Max.   :1.1000   Max.   :1.6600   Max.   :65.800   Max.   :0.34600   Max.   :289.00    Max.
:440.0          Max.   :1.0390   Max.   :3.820
   sulphates         alcohol        quality
 Min.   :0.2200   Min.   : 8.00   Min.   :3.000
 1st Qu.:0.4100   1st Qu.: 9.50   1st Qu.:5.000
 Median :0.4700   Median :10.40   Median :6.000
 Mean   :0.4898   Mean   :10.51   Mean   :5.878
 3rd Qu.:0.5500   3rd Qu.:11.40   3rd Qu.:6.000
 Max.   :1.0800   Max.   :14.20   Max.   :9.000
```

Zde vidime ruzne zavislosti mezi merenymi vlastnostmi.

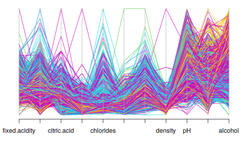
Hide

```
pairs(winequality[,1:11], col = winequality[,12])
```



Hide

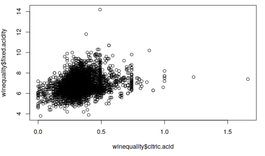
```
library(MASS)
parcoord(winequality[,1:11], col = winequality[,12])
```



Obsah kyseliny citronove vs urcite mereni kyselosti

Hide

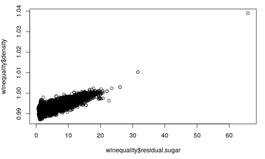
```
plot(winequality$citric.acid, winequality$fixed.acidity)
```



Obsah cukru vs. hustota vina

Hide

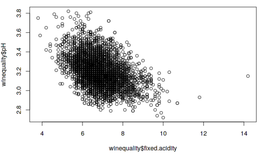
```
plot(winequality$residual.sugar, winequality$density)
```



Urcite mereni kyselosti vs. pH

Hide

```
plot(winequality$fixed.acidity, winequality$pH)
```



PCA na nestandardizovanych datech

Z dat vynechan sloupec s hodnocenim kvality vin.

Hide

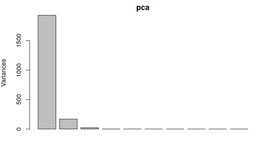
```
pca = prcomp(winequality[,1:11])
summary(pca)
```

```
Importance of components:
                            PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9     PC10      PC11
Standard deviation     43.9498 12.97894 4.64338 1.03654 0.82866 0.13613 0.11954 0.10699 0.09296 0.0199 0.0005629
Proportion Of Variance  0.9097  0.07933 0.01015 0.00051 0.00032 0.00001 0.00001 0.00000 0.00000 0.0000 0.0000000
Cumulative Proportion   0.9097  0.98899 0.99915 0.99965 0.99997 0.99998 0.99999 1.00000 1.00000 1.0000 1.0000000
```

Prvni dve komponenty zachyti 98 % rozptylu.

Hide

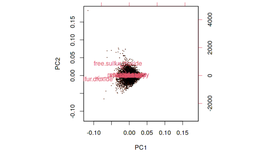
```
plot(pca)
```



Projekce dat na prvni dve hlavni komponenty.

Hide

```
biplot(pca, xlabs=rep(".", nrow(winequality)))
```



PCA pro standardizovana data

Hide

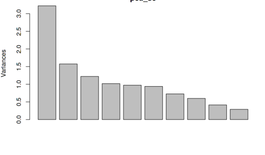
```
pca_sc = prcomp(winequality[,1:11], scale. = TRUE)
summary(pca_sc)
```

```
Importance of components:
                          PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9    PC10   PC11
Standard deviation     1.7951 1.2551 1.1053 1.00922 0.96658 0.96089 0.85241 0.77418 0.64354 0.53804 0.14370
Proportion Of Variance 0.2929 0.1432 0.1111 0.09259 0.08648 0.08534 0.06905 0.05449 0.03765 0.02632 0.00188
Cumulative Proportion  0.2929 0.4361 0.5472 0.63979 0.72827 0.81361 0.87967 0.93416 0.97181 0.99812 1.00000
```

Zde je situace velmi odlisna, pro 98 % je potreba 10 hlavnich komponent...

Hide

```
plot(pca_sc)
```



Hide

```
biplot(pca_sc,xlabs=rep(".", nrow(winequality)))
```