

# ISV51 – Programmation sous R

*Correction*

*devoir d'automne 2015 – 1h30*

## Exercice 1: absorption de dioxyde de carbone chez les graminées.

1. Combien y a t-il d'individus ? Quelles sont les variables qualitatives ? Quantitatives ?

```
data(CO2) # On charge les données
nrow(CO2) # on trouve 84 lignes ou individus
```

```
## [1] 84
```

```
head(CO2)
```

```
##   Plant   Type Treatment conc uptake
## 1   Qn1 Quebec nonchilled   95   16.0
## 2   Qn1 Quebec nonchilled  175   30.4
## 3   Qn1 Quebec nonchilled  250   34.8
## 4   Qn1 Quebec nonchilled  350   37.2
## 5   Qn1 Quebec nonchilled  500   35.3
## 6   Qn1 Quebec nonchilled  675   39.2
```

Les trois premières colonnes (`Plant`, `Type` et `Treatment`) sont des variables qualitatives. Les deux dernières sont des variables quantitatives (`conc` et `uptake`):

```
sapply(CO2, is.factor)
```

```
##      Plant      Type Treatment      conc      uptake
##      TRUE      TRUE      TRUE     FALSE     FALSE
```

2. a) Déterminer les effectifs de chaque modalité des variables `Plant`, `Type` et `Treatment`. Commenter.

Tout ceci se règle à l'aide de la commande `table`.

```
with(CO2, table(Plant))
```

```
## Plant
## Qn1 Qn2 Qn3 Qc1 Qc3 Qc2 Mn3 Mn2 Mn1 Mc2 Mc3 Mc1
##   7   7   7   7   7   7   7   7   7   7   7   7
```

```
with(CO2, table(Type))
```

```
## Type
##      Quebec Mississippi
##          42           42
```

```
with(CO2, table(Treatment))
```

```
## Treatment
## nonchilled    chilled
##           42         42
```

On remarque que les effectifs de chaque groupe sont équilibrés: il y a autant de plantes de chaque lignée (7 à chaque fois), de chaque origine géographique (42 de chaque), et les traitements sont appliqués de manière équilibrée à chaque lignée de plante.

b) Interpréter la sortie de la commande `table(CO2$Plant, CO2$Treatment)`

```
table(CO2$Plant, CO2$Treatment)
```

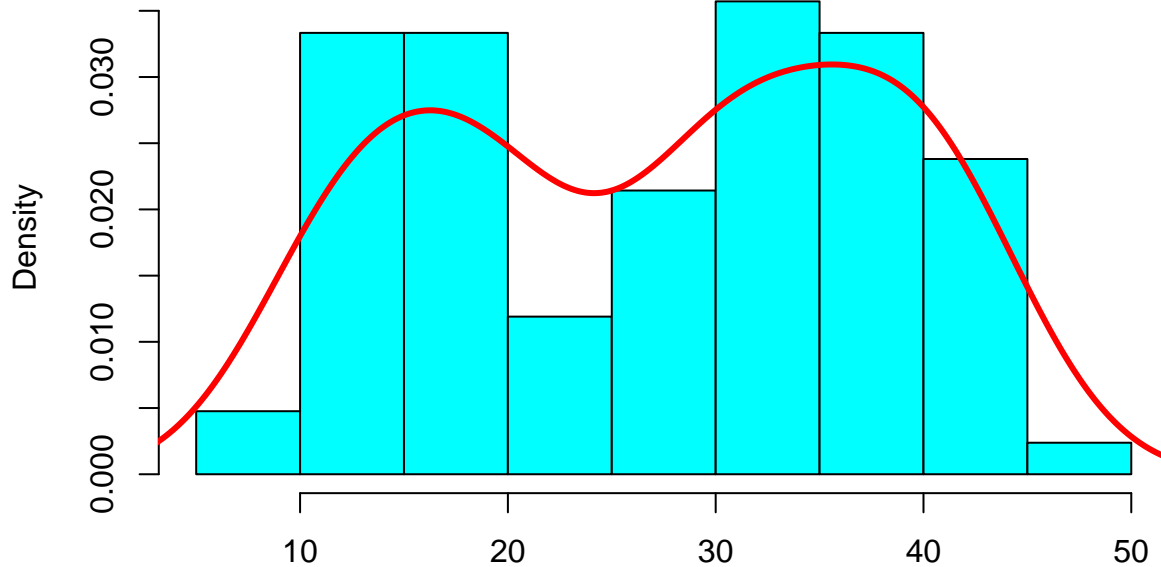
```
##
##      nonchilled chilled
## Qn1           7      0
## Qn2           7      0
## Qn3           7      0
## Qc1           0      7
## Qc3           0      7
## Qc2           0      7
## Mn3           7      0
## Mn2           7      0
## Mn1           7      0
## Mc2           0      7
## Mc3           0      7
## Mc1           0      7
```

On confirme le plan équilibré en terme de traitement sur chaque lignée de plante à l'aide de cette commande, qui réalise une table croisée des occurrences de chaque couple (`Treatment`, `Plant`) du jeu de données.

3. À l'aide d'un histogramme, représenter la distribution empirique de l'absorption en  $\text{CO}_2$ . Superposer l'estimation de la densité obtenue avec l'estimateur à noyau `density`. Qu'observez-vous ?

```
hist(CO2$uptake, col="cyan", prob=TRUE, xlab="", main="histogramme du taux d'absorption")
lines(density(CO2$uptake), col="red", lwd=3)
```

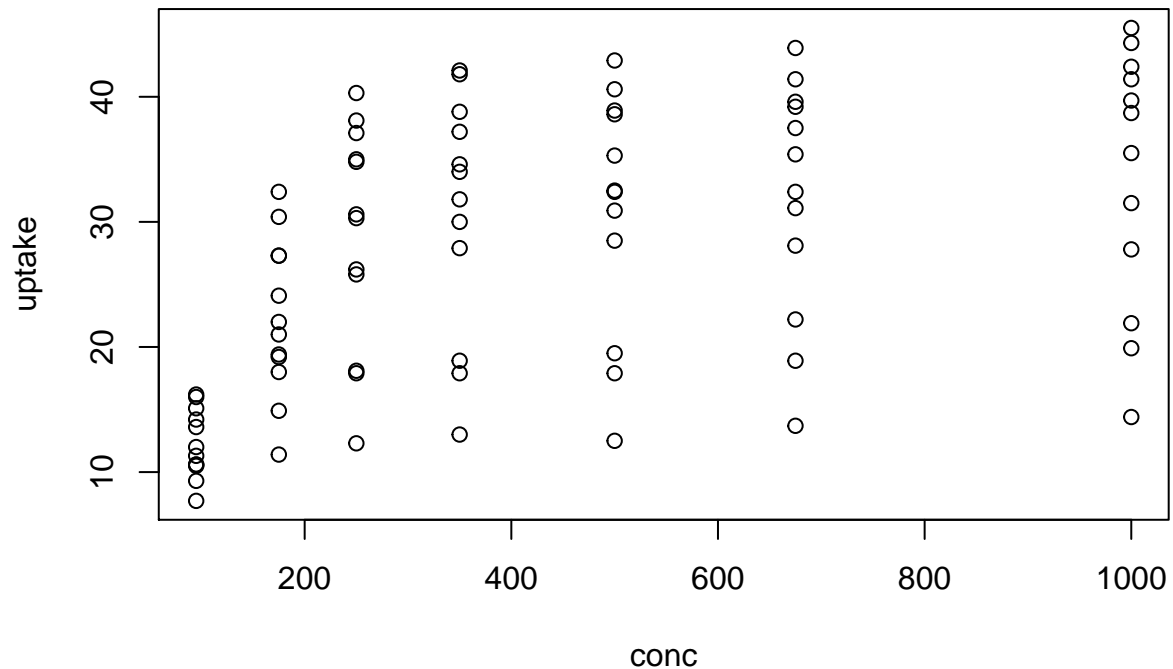
## histogramme du taux d'absorption



Deux modes de dégagent clairement dans cette distribution : sont-ils liés aux origines des plantes ? Au traitement ? À la concentration en dioxyde de carbone ? À un mélange de ces trois effets ?

4. a) Représenter l'évolution de l'absorption en  $\text{CO}_2$  en fonction de la concentration en  $\text{CO}_2$  à l'aide d'un nuage de points (ou diagramme de dispersion).

```
plot(uptake~conc, CO2)
```



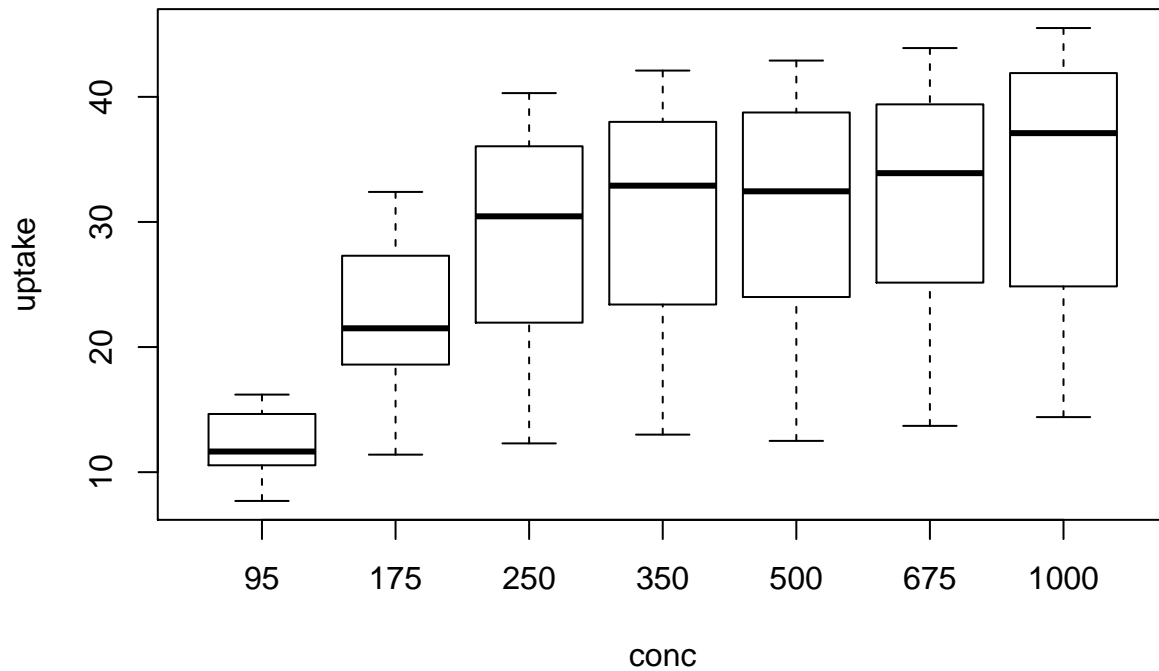
On remarque que la concentration est regroupée par niveau, d'où l'idée de transformer cette variable en facteur.

- b) Transformer la variable `conc` en facteur. Combien y a-t-il de niveaux ? Représenter à nouveau l'évolution de l'absorption en  $\text{CO}_2$  en fonction de la concentration en  $\text{CO}_2$  à l'aide de boxplot pour chaque niveau de concentration.

```
C02$conc <- as.factor(C02$conc)
nlevels(C02$conc) # on trouve 7 niveaux
```

```
## [1] 7
```

```
plot(uptake~conc, C02)
```

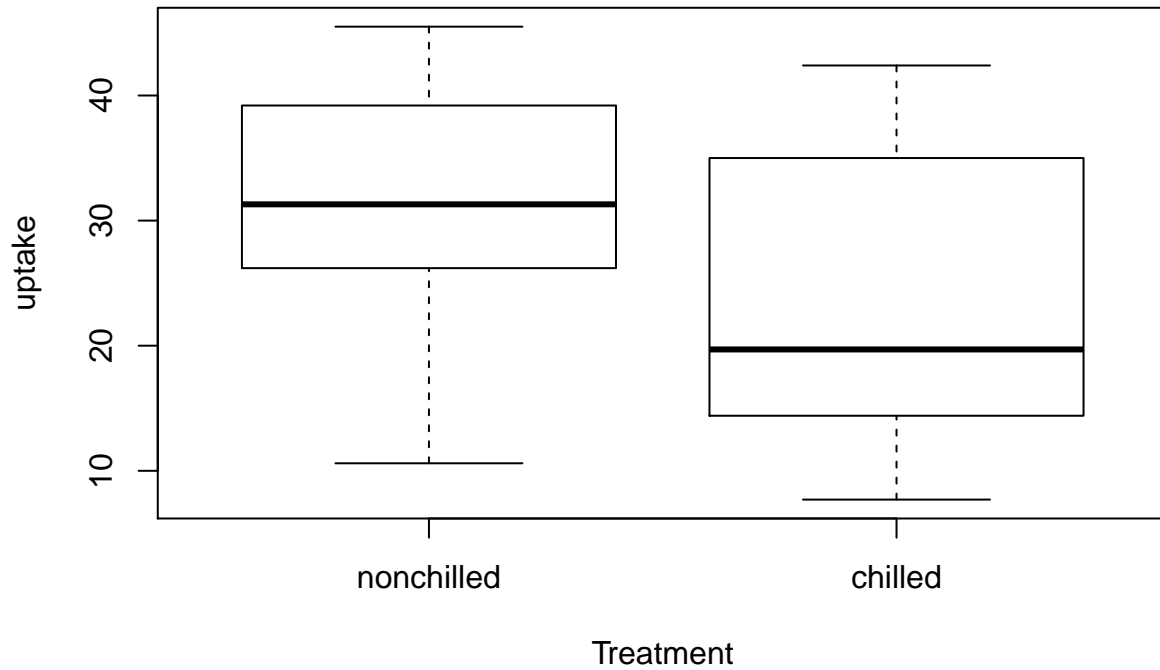


- c) Commenter ces deux graphes

Il y a clairement un effet dû à l'augmentation de la concentration en  $\text{CO}_2$  sur le degré d'absorption des plantes, mais cet effet sature assez rapidement. Les autres facteurs (origine et traitement) ont sans doute également un effet.

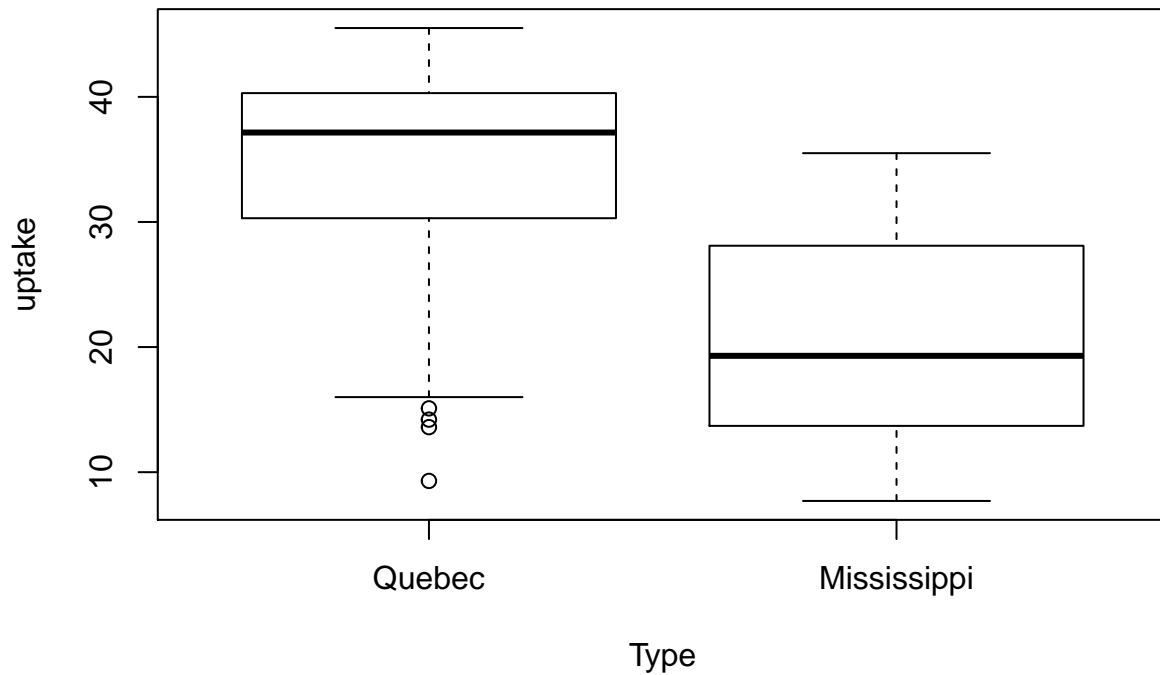
5. a) Représenter la distribution de l'absorption en  $\text{CO}_2$  pour les deux types de traitements.

```
plot(uptake~Treatment, C02)
```



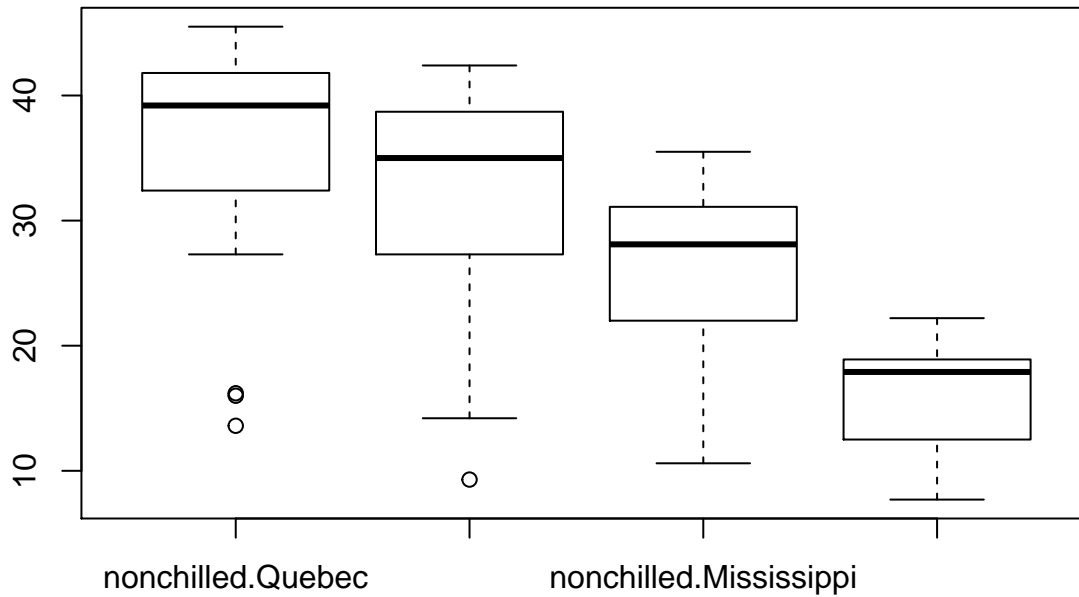
b) Représenter la distribution de l'absorption en CO<sub>2</sub> pour les deux origines géographiques.

```
plot(uptake~Type, C02)
```



c) Représenter la distribution de l'absorption en CO<sub>2</sub> par couple (traitement, origine).

```
boxplot(uptake~Treatment*Type, C02)
```



d) Afficher les individus dont le niveau d'absorption est supérieur à 1.5 fois la moyenne de tous les niveaux d'absorption.

```
subset(CO2, uptake > 1.5 * mean(uptake))
```

```
##   Plant  Type Treatment conc uptake
## 11  Qn2 Quebec nonchilled  350  41.8
## 13  Qn2 Quebec nonchilled  675  41.4
## 14  Qn2 Quebec nonchilled 1000  44.3
## 18  Qn3 Quebec nonchilled  350  42.1
## 19  Qn3 Quebec nonchilled  500  42.9
## 20  Qn3 Quebec nonchilled  675  43.9
## 21  Qn3 Quebec nonchilled 1000  45.5
## 35  Qc2 Quebec   chilled 1000  42.4
## 42  Qc3 Quebec   chilled 1000  41.4
```

e) À l'aide de la commande `tapply`, calculer le l'absorption moyenne pour chaque lignée de plantes.

```
with(CO2, tapply(uptake, Plant, mean))
```

```
##      Qn1      Qn2      Qn3      Qc1      Qc3      Qc2      Mn3      Mn2
## 33.22857 35.15714 37.61429 29.97143 32.58571 32.70000 24.11429 27.34286
##      Mn1      Mc2      Mc3      Mc1
## 26.40000 12.14286 17.30000 18.00000
```

f) Conclure sur l'origine des plantes les plus efficaces pour l'absorption de  $\text{CO}_2$ .

Clairement, les plantes originaires du Québec exhibent un plus haut degré d'absorption, quelque soit le traitement et la concentration en  $\text{CO}_2$ .

## Exercice 2: matrices élémentaires

1. Compléter la fonction suivante afin de construire la matrice élémentaire  $\mathbf{E}_\lambda^{(ij)}$  de taille  $n \times n$

```
E <- fonction(i, j, lambda, n=3) {
  matE <- diag(rep(1,n))
  matE[i,j] <- lambda
  return(matE)
}
```

2. Quelle matrice notée  $\mathbf{A}$  peut-elle se décomposer en les éléments suivants

$$\mathbf{E}_{-2}^{(21)} \mathbf{E}_{-1}^{(13)} \mathbf{E}_2^{(23)} \mathbf{E}_1^{(12)} \mathbf{E}_1^{(32)}$$

```
A <- E(2,1,-2) %*% E(1,3,-1) %*% E(2,3,2) %*% E(1,2,1) %*% E(3,2,1)
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    0   -1
## [2,]   -2    3    4
## [3,]    0    1    1
```

3. On rappelle que l'inverse de la matrice élémentaire  $\mathbf{E}_\lambda^{(ij)}$  est la matrice  $\mathbf{E}_{-\lambda}^{(ij)}$ . Écrire une fonction `invE(i, j, lambda, n)` qui calcule cette inverse.

```
invE <- fonction(i, j, lambda, n=3) {
  return(return(E(i,j,-lambda,n)))
}
```

4. À l'aide de `invE`, déterminer une décomposition en matrices élémentaires de  $\mathbf{A}^{-1}$ .

```
Am1 <- invE(3,2,1) %*% invE(1,2,1) %*% invE(2,3,2) %*% invE(1,3,-1) %*% invE(2,1,-2)
Am1
```

```
##      [,1] [,2] [,3]
## [1,]   -1   -1    3
## [2,]    2    1   -2
## [3,]   -2   -1    3
```

## Calcul des notes du DS

```
dispot <- c(1, 1, 0.75, 0.75, 1, 1, 0.25, 1, 1, 0, 1, 1, 0.75, 0, 0, 0, 0)
el.kurdi <- c(1, 1, 1, 0.75, 1, 1, 1, 1, 1, 0, 1, 1, 0.75, 0, 0, 0, 0)
gene <- c(1, 1, 1, 0.75, 1, 1, 1, 0.5, 0.5, 0, 0, 0, 0, 0, 0, 0, 0)
illiano <- c(0.5, 0, 0.5, 0.75, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)
le.hoang <- c(0.5, 0, 0.5, 0.75, 0.5, 1, 0.25, 0, 1, 0, 0.25, 0.75, 0, 0.25, 0, 0, 0)
manza <- c(1, 1, 0.75, 1, 1, 1, 0, 0, 0, 0, 0.75, 0, 0, 0, 0, 0, 0)
miagoux <- c(1, 1, 1, 0.75, 0.5, 0.75, 0, 0.5, 0.5, 0, 0.5, 0, 0, 0, 0, 0, 0)
ratovo <- c(1, 1, 0.5, 0.75, 1, 1, 0.5, 1, 1, 0.5, 0, 0, 0, 0, 0, 0, 0)
```

```

vilette <- c(0.5, 1, 0.75, 1, 1, 1, 1, 1, 1, 0, 0.75, 1, 1, 0.5, 1, 0.5, 0)

releves <- t(matrix(c(dispot, el.kurdi, gene, illiano, le.hoang, manza, miagoux, ratovo, vilette),
  nrow=9, byrow = TRUE,
  dimnames = list(c("dispot", "el.kurdi", "gene", "illiano",
    "le.hoang", "manza", "miagoux", "ratovo", "vilette"),
    c("e1.1", "e1.2a", "e1.2b", "e1.3", "e1.4a", "e1.4b", "e1.4c",
      "e1.5a", "e1.5b", "e1.5c", "e1.5d", "e1.5e", "e1.5f",
      "e2.1", "e2.2", "e2.3", "e2.4"))))

releves <- data.frame(releves)
bareme <- c(1, 1, 1, 2, 1, 2, 1, 1, 1, 1, 2, 2, 1, 1.5, 1.5, 1.5, 1.5)
notes <- sapply(releves, function(x) sum(x*bareme))
print(notes)

```

```

##   dispot el.kurdi   gene illiano le.hoang   manza miagoux   ratovo
##  14.250  15.250   9.500   5.500   8.625   9.250   8.500  10.000
##  vilette
##   17.750

```

```
barplot(sort(notes), las=3)
```

