

Introduction à R partie 2

L'organisation des données sous R

Les vecteurs

```
v <- 1:10
```

Les vecteurs

```
v1 <- seq(from=3, to=21, by=2)
```

Les vecteurs

```
v2 <- seq(from=2, by=3, length=10)
```

Les vecteurs

```
v3 <- seq(from=1, to=50, length=100)
```

Les vecteurs

```
v4 <- rep(1:2, times=3)
```

```
v5 <- rep(1:2, each=3)
```

Les vecteurs

```
length(v1)
```

```
## [1] 10
```

```
mean(v2)
```

```
## [1] 15.5
```

```
sd(v3)
```

```
## [1] 14.35922
```


Les vecteurs

```
exp(v1)
```

```
## [1] 2.008554e+01 1.484132e+02 1.096633e+03 8.103084e+03 5.987414e+04  
## [6] 4.424134e+05 3.269017e+06 2.415495e+07 1.784823e+08 1.318816e+09
```

```
log(v2)
```

```
## [1] 0.6931472 1.6094379 2.0794415 2.3978953 2.6390573 2.8332133 2.9957323  
## [8] 3.1354942 3.2580965 3.3672958
```

Les vecteurs

Quelques fonctions sur les vecteurs

```
v7 <- rep(v2, times = 3)
```

```
# classer  
sort(v7)
```

```
## [1] 2 2 2 5 5 5 8 8 8 11 11 11 14 14 14 17 17 17 20 20 20 23 23  
## [24] 23 26 26 26 29 29 29
```

```
rev(v7)
```

```
## [1] 29 26 23 20 17 14 11 8 5 2 29 26 23 20 17 14 11 8 5 2 29 26 23  
## [24] 20 17 14 11 8 5 2
```

Les vecteurs

Quelques fonctions sur les vecteurs

```
# compter le nombre d'occurrence de chaque valeur  
table(v7)
```

```
## v7  
##  2  5  8 11 14 17 20 23 26 29  
##  3  3  3  3  3  3  3  3  3  3
```

```
# extraire les valeurs uniques  
unique(v7)
```

```
## [1]  2  5  8 11 14 17 20 23 26 29
```

Les vecteurs

Un vecteur de chaîne de caractères

```
ch1 <- c("traitement_1")
```

```
ch2 <- c("traitement_2", "traitement_3")
```

```
ch3 <- c(ch1, ch2)
```

```
length(ch3)
```

```
## [1] 3
```

Les tableaux (dataframe)

```
df1 <- data.frame(v1, v2, v3)
```

```
df2 <- data.frame(ch1, ch2)
```

```
df3 <- data.frame(v=1:4, ch=c("a","b","b","c"), n=10)
```

Les tableaux (dataframe)

```
x <- c(1:10)
```

```
y <- c(2:11)
```

```
df4 <- cbind(x,y)
```

```
df5 <- rbind(x,y)
```

```
# ajouter les intitulés des colonnes  
colnames(df4) <- ch2
```

Les matrices (matrix)

```
m <- matrix(1:12, nrow=4, ncol=3)

m1 <- m+1

m2 <- m*2

m3 <- t(m2) # transposé d'une matrice

m4 <- m1*m2
```

Les matrices (matrix)

```
dim(m2)
```

```
## [1] 4 3
```

```
dim(m3)
```

```
## [1] 3 4
```

```
sum(m)
```

```
## [1] 78
```

```
mean(m)
```

```
## [1] 6.5
```



```
class(v)
```

```
## [1] "integer"
```

```
class(df1)
```

```
## [1] "data.frame"
```

```
class(m)
```

```
## [1] "matrix"
```

Indexation

Indexation des vecteurs

```
x<-1:20*pi
```

```
x[10] #le 10ème élément de l'objet x
```

```
## [1] 31.41593
```

```
x[-1] #tous les éléments de x sauf le premier
```

```
## [1] 6.283185 9.424778 12.566371 15.707963 18.849556 21.991149 25.132741
```

```
## [8] 28.274334 31.415927 34.557519 37.699112 40.840704 43.982297 47.123890
```

```
## [15] 50.265482 53.407075 56.548668 59.690260 62.831853
```

Indexation

Indexation des vecteurs

```
x[1:10]
```

```
## [1] 3.141593 6.283185 9.424778 12.566371 15.707963 18.849556 21.991148 25.132741 28.274334 31.415927  
## [8] 25.132741 28.274334 31.415927
```

```
x[c(1, 4, 2)]
```

```
## [1] 3.141593 12.566371 6.283185
```

```
x[c(5:10, 15:20)]
```

```
## [1] 15.70796 18.84956 21.99115 25.13274 28.27433 31.41593 34.55752 37.69911 40.84070 43.98229 47.12389 50.26548 53.40708 56.54867 59.69026 62.83185  
## [8] 50.26548 53.40708 56.54867 59.69026 62.83185
```

Indexation

Indexation des matrices

```
m[1,2] #l'élément de la ligne i, colonne j
```

```
## [1] 5
```

Indexation

Indexation des matrices

```
m[1,] #toute la ligne i
```

```
## [1] 1 5 9
```

```
m[,2] #toute la colonne j
```

```
## [1] 5 6 7 8
```

```
m[,c(1,3)] #les colonnes 1 et 3
```

```
##      [,1] [,2]  
## [1,]    1    9  
## [2,]    2   10  
## [3,]    3   11  
## [4,]    4   12
```

Indexation

Indexation des dataframes

Les matrices et les dataframes sont indexés de la même façon.

```
df3[["ch"]] # les éléments de la colonne nommée "ch"
```

```
## [1] a b b c  
## Levels: a b c
```

```
df3$ch
```

```
## [1] a b b c  
## Levels: a b c
```

Sélection des éléments

Dans un vecteur

- Sélection des éléments de x suivant des critères:

```
x[x>10]
```

```
## [1] 12.56637 15.70796 18.84956 21.99115 25.13274 28.274
## [8] 34.55752 37.69911 40.84070 43.98230 47.12389 50.265
## [15] 56.54867 59.69026 62.83185
```

```
x[x>10 & x<16]
```

```
## [1] 12.56637 15.70796
```

Sélection des éléments

Dans un tableau

- Sélection des éléments du tableau *data* suivant des critères:

```
data <- read.csv("data/data_poisson.csv")
```

```
data_demersal <- subset(data, position == "demersal")
```

```
newdata <- subset(data, groupe=="poisson" & life_span > 5)
```


Sélection des éléments

Dans un tableau

- Remplacer une valeur dans un tableau

```
data2 <- data
```

```
data2$life_span[data2$life_span==2] <- 1000
```

Statistiques conditionnelles dans un tableau

```
tapply(X=data$catch_2001, INDEX=data$groupe, FUN=mean)
```

```
## cephalopod  crustacea      fish  
##    1346606    2001715    1985685
```

- ▶ Calculez la somme des captures en 2005 en fonction de la position des espèces dans la colonne d'eau.
- ▶ Enregistrez les sorties dans l'objet : result

Enregistrer les sorties

```
write.csv(result, file="My_results.csv")
```

Fonctions mathématiques

```
x<-m[1,]*pi
```

```
y<-m[2,]/pi
```

```
max(x)
```

```
## [1] 28.27433
```

```
min(x)
```

```
## [1] 3.141593
```

```
range(x)
```

```
## [1] 3.141593 28.274334
```

Fonctions mathématiques

```
sum(x)
```

```
## [1] 47.12389
```

```
prod(x)
```

```
## [1] 1395.282
```

```
mean(x)
```

```
## [1] 15.70796
```

```
median(x)
```

```
## [1] 15.70796
```

```
quantile(x)
```

Fonctions mathématiques

```
rank(x)
```

```
## [1] 1 2 3
```

```
var(x)
```

```
## [1] 157.9137
```

```
cor(x, y)
```

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

```
round(x, 2)
```

```
## [1] 3.14 15.71 28.27
```

Exercice

Dessous les mesures d'échantillonnage de bivalve dans une lagune.

```
mussel.count <- c(20, 19, 13, 17, 21, NA, 19, 17, 24, 23,  
18, 24, 23, 20, 15)  
mussel.area<-c("area1", "area1", "area1", "area1", "area1",  
"area2", "area2", "area2", "area2", "area2",  
"area3", "area3", "area3", "area3", "area3")
```

- 1/ Entrer les données R en format vecteur
- 2/ Créer un tableau à partir des vecteurs
- 3/ Enregistrer les données dans votre répertoire de travail sous format csv
- 4/ Lire le fichier que vous avez enregistré et vérifier les données
- 5/ Calculer la moyenne, écart-type, somme et variance de l'abondance des moules par zone et les mettre dans un vecteur