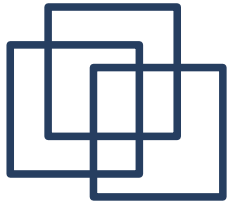


Module

AIDE A LA DECISION



Présentation

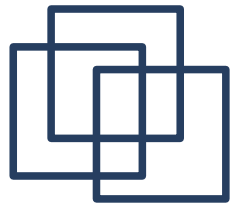
Deux intervenants :

- Guillaume Cleuziou
- Vincent NGuyen

7 semaines : Cours (1h) – TD (1h) – TP (1h30)

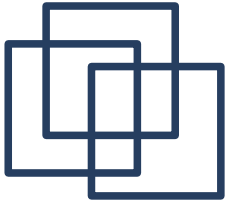
Un projet : sous forme de challenge

Notation : challenge+DS+QCMs+ TP noté (?)



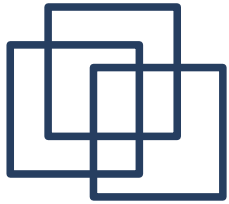
Premières semaines

- Introduction : Aide à la décision
- Initiation au langage R
- Les méthodes de classification
- Lancement du challenge



AIDE A LA DECISION

INTRODUCTION



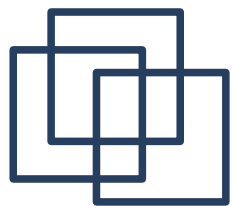
Définition

« En informatique, l'aide à la décision est un domaine qui vise à concevoir des outils informatiques pour aider un décideur à analyser un problème ou une situation, et à lui fournir des solutions. » (wikipédia)

En Anglais : *business intelligence* ou encore *decision making*

Domaines concernés :

- bases de données (entrepôts)
- analyse de données
- intelligence artificielle
- visualisation
- ...



Exemple 1 : ordonnancement

The screenshot shows a Google search results page for the query "iut informatique orléans". The browser window has a menu bar (File, Edit, View, History, Bookmarks, Tools, Help) and a search bar with the Google logo. The search results are displayed under the heading "Web". The first result is "Département Informatique - IUT d'Orléans" with a description of the department and a link to the website. The second result is "IUT d'Orléans - IUT d'Orléans" with a description of the department and a link to the website. The third result is "IUT INFORMATIQUE à ORLEANS - 89 anciens inscrits sur Copains D'avant !" with a description of the department and a link to the website. The fourth result is "IUT Orléans - IUT EN LIGNE" with a description of the department and a link to the website. The fifth result is "Iut Informatique" with a description of the department and a link to the website. The status bar at the bottom of the browser window shows "Done".

File Edit View History Bookmarks Tools Help

http://www.google.fr/#hl=fr&source=hp&q=iut+informatique+orléans&btnG=Recherche+Google&meta=&aq=

Most Visited Getting Started Latest Headlines Utils Perso Ressources R Enseignement Recherche mobilisation

Web Images Vidéos Maps Actualités Groupes Gmail plus

Connexion

Google iut informatique orléans Rechercher Recherche avancée Préférences

Rechercher dans : Web Pages francophones Pages : France

Web Résultats 1 - 10 sur un total d'environ 244 000 pour iut informatique orléans (0,28 secondes)

Département Informatique - IUT d'Orléans
268 étudiants sur les trois années de formation (IUT + Licence professionnelle) pour 2002/2003, encadrés par une équipe de 26 enseignants titulaires et une ...
www.univ-orleans.fr/iut-orleans/informatique/ - Pages similaires

IUT d'Orléans - IUT d'Orléans
16 mai 2009 ... Formations dans les filières Chimie, GMP (Génie Mécanique et Productique), **informatique**, OGP (Organisation et Génie de la Production), ...
www.univ-orleans.fr/iut-orleans/ - En cache - Pages similaires

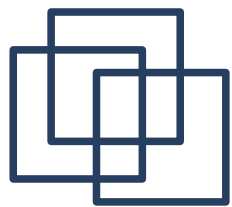
Plus de résultats de www.univ-orleans.fr

IUT INFORMATIQUE à ORLEANS - 89 anciens inscrits sur Copains D'avant !
IUT INFORMATIQUE à ORLEANS - 89 anciens photos de classe et trombinoscope de l'établissement IUT INFORMATIQUE. Retrouvez vos amis perdus de vues.
copainsdavant.linternaute.com/.../iut_informatique/ - En cache - Pages similaires

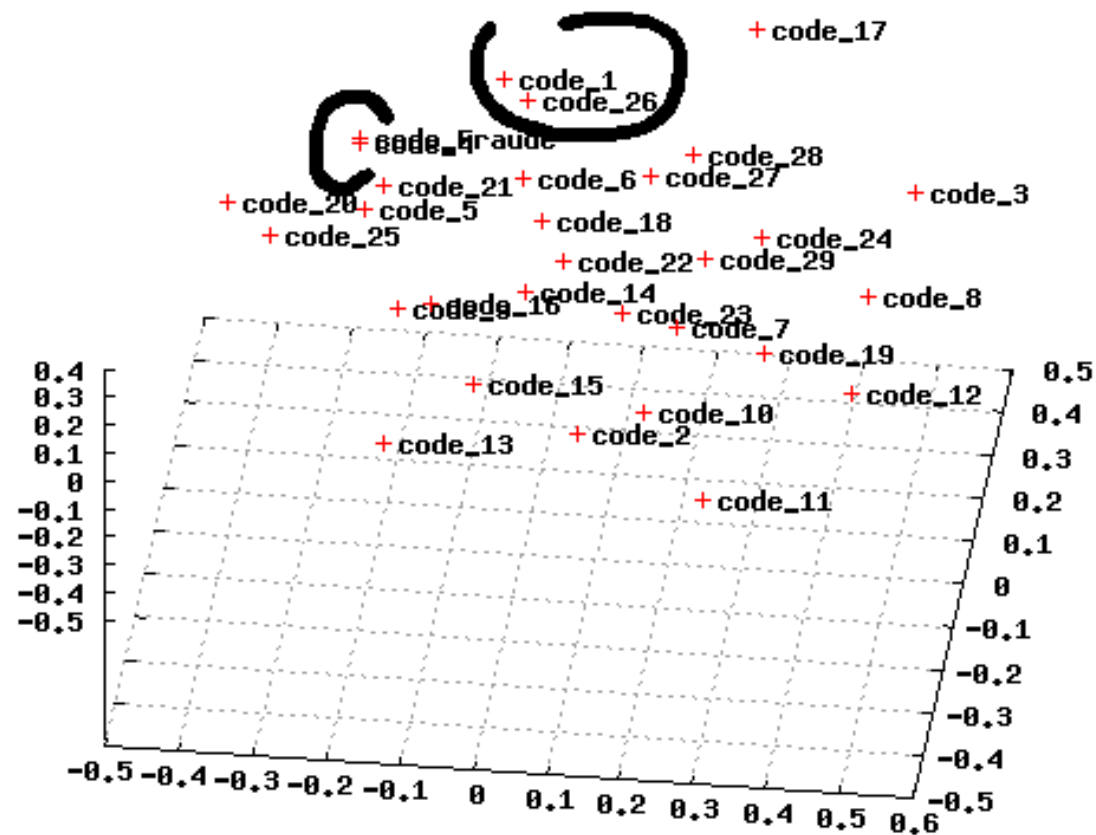
IUT Orléans - IUT EN LIGNE
Iut d'Orléans rue d'Issoudun 45000 Orléans Tél : 02 38 49 44 96 ... Former des professionnels de l'**informatique**, compétents au plan technologique, ...
www.iutenligne.net/IUT/Orl%E9ans - En cache - Pages similaires

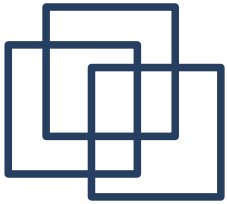
Iut Informatique
9 messages - 6 auteurs - Dernier message : 9 juin 2005
Iut Informatique : Salut tout le monde, voila je suis en term S et bon ... la quelle de ces iut est mieux: celle de nantes ou celle **orléans**? ...
www.infos-du-net.com/.../98998-28-informatique - En cache - Pages similaires

Done



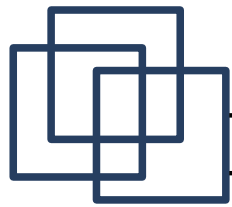
Exemple 2 : détection de fraude





Exemple 3 : classification





Exemple 4 : extraction de motifs





Exemple 5 : recommendation

The screenshot shows the Amazon.co.uk 'Your Recommendations' page. The browser window title is 'Amazon.co.uk: Your Recommendations - Microsoft Internet Explorer'. The address bar shows the URL: http://www.amazon.co.uk/exec/obidos/tg/stores/recs/instant-recs/-/books/0/batch/none/0/0/1/pc/ref=pd_jr_batch_list/026-0668331-0070807.

The page header includes the Amazon.co.uk logo, a 'MasterCard' logo, and navigation links: 'VIEW BASKET', 'WISH LIST', 'YOUR ACCOUNT', and 'HELP'. Below the header is a navigation bar with categories: 'WELCOME', 'J'S STORE', 'BOOKS', 'ELECTRONICS & PHOTO', 'MUSIC', 'DVD', 'VIDEO', 'SOFTWARE', 'PC & VIDEO GAMES', 'HOME & GARDEN', 'TOYS & KIDS!', and 'TRAVEL'. A sub-navigation bar highlights 'YOUR FAVOURITE STORES', 'YOUR RECOMMENDATIONS', 'THE PAGE YOU MADE', and 'NEW FOR YOU'.

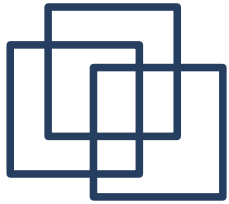
The main content area starts with a personalized greeting: 'Hello J.C. Dursteler Lopez, we have recommendations for you (if you're not J.C. Dursteler Lopez, [click here](#)). Here's your New For You™ recommendations'.

Under the heading 'Your Recommendations > Books', there is a section titled 'RECOMMENDATIONS'. It includes a list of categories on the left: 'Address Books, Journals & More', 'Art, Architecture & Photography', 'Audio CDs', 'Audio Cassettes', 'Biography', 'Business, Finance & Law', 'Children's Books', 'Comics & Graphic Novels', 'Computers & Internet', 'Crime, Thrillers & Mystery', 'Fiction', 'Food & Drink', 'Gay & Lesbian', 'Health, Family & Lifestyle', 'History', 'Home & Garden', 'Horror', 'Humour', and 'Mind, Body & Spirit'.

The main list of recommendations consists of five items, each with a book cover, title, author, and three radio button options: 'No Opinion', 'I own it', and 'Not interested'.

Item	Book Title	Author	No Opinion	I own it	Not interested
1.	Content Critical: Gaining Competitive Advantage Through High-Quality Web Content	by Gerry McGovern, Rob Norton	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2.	The Elements of User Experience	by Jesse James Garrett	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3.	The Design of Sites: Principles, Processes and Patterns for Crafting a Customer-centered Web Experience	by Douglas K. Van Duyne, et al	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.	Information Architecture: Blueprints for the Web	by Christina Wodtke	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.	Designing with Web Standards	by Jeffrey Zeldman	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

At the bottom of the recommendations list, there is a 'Save & Continue' button.



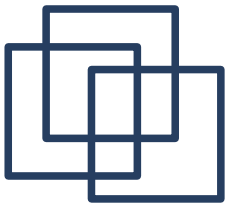
Les données

Parfois simples :

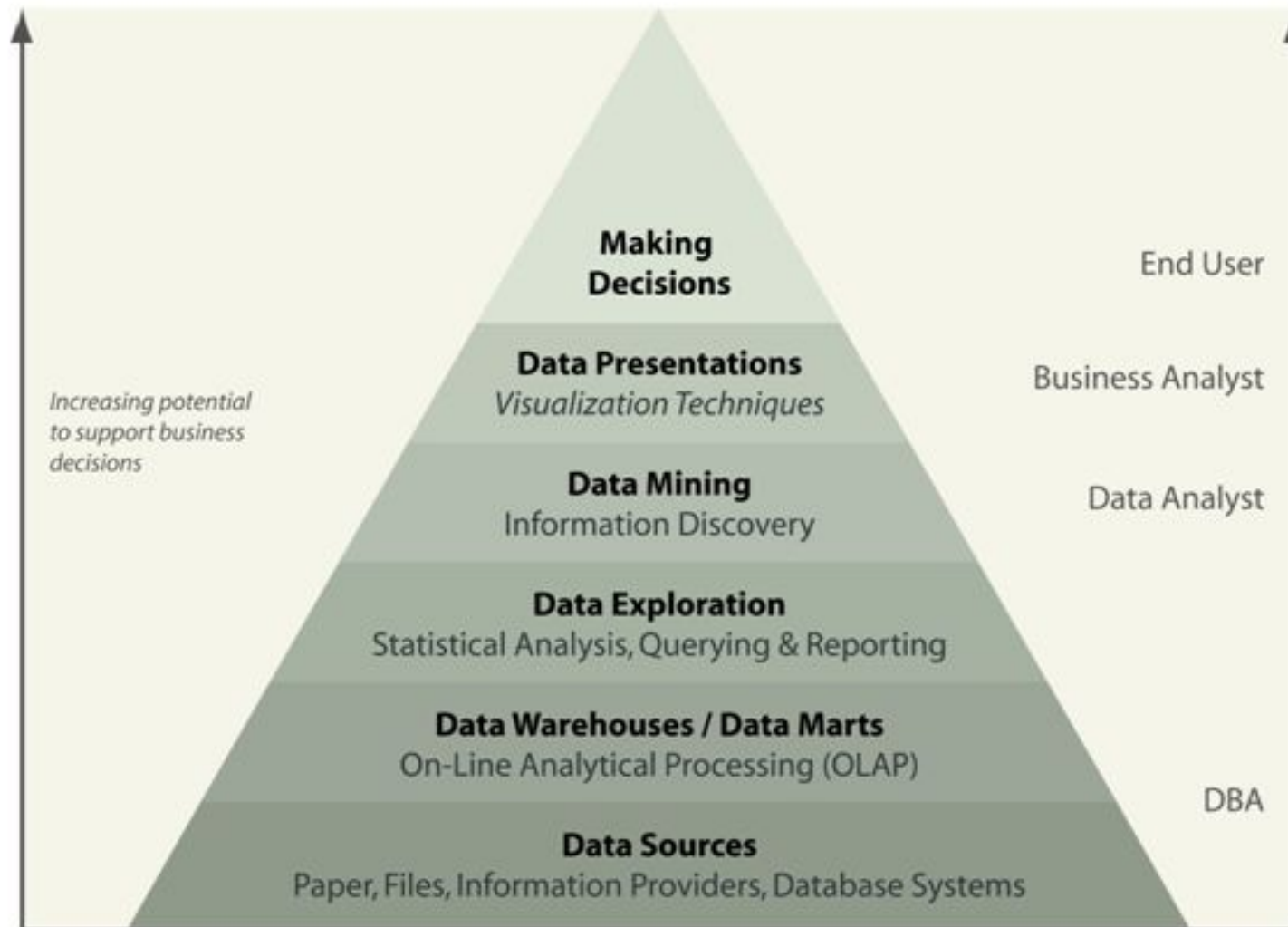
- numériques
- symboliques
- mixtes

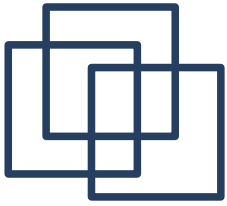
Souvent complexes :

- données temporelles,
- données séquentielles,
- données spatiales,
- données (hyper-)textuelles
- puces ADN
- réseaux (e.g. sociaux)
- flux de données



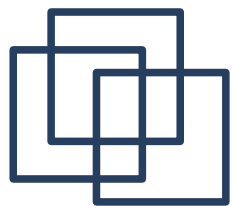
Le processus





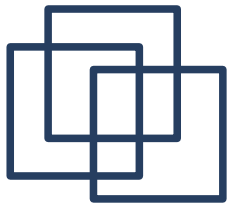
AIDE A LA DECISION

Le langage R : présentation

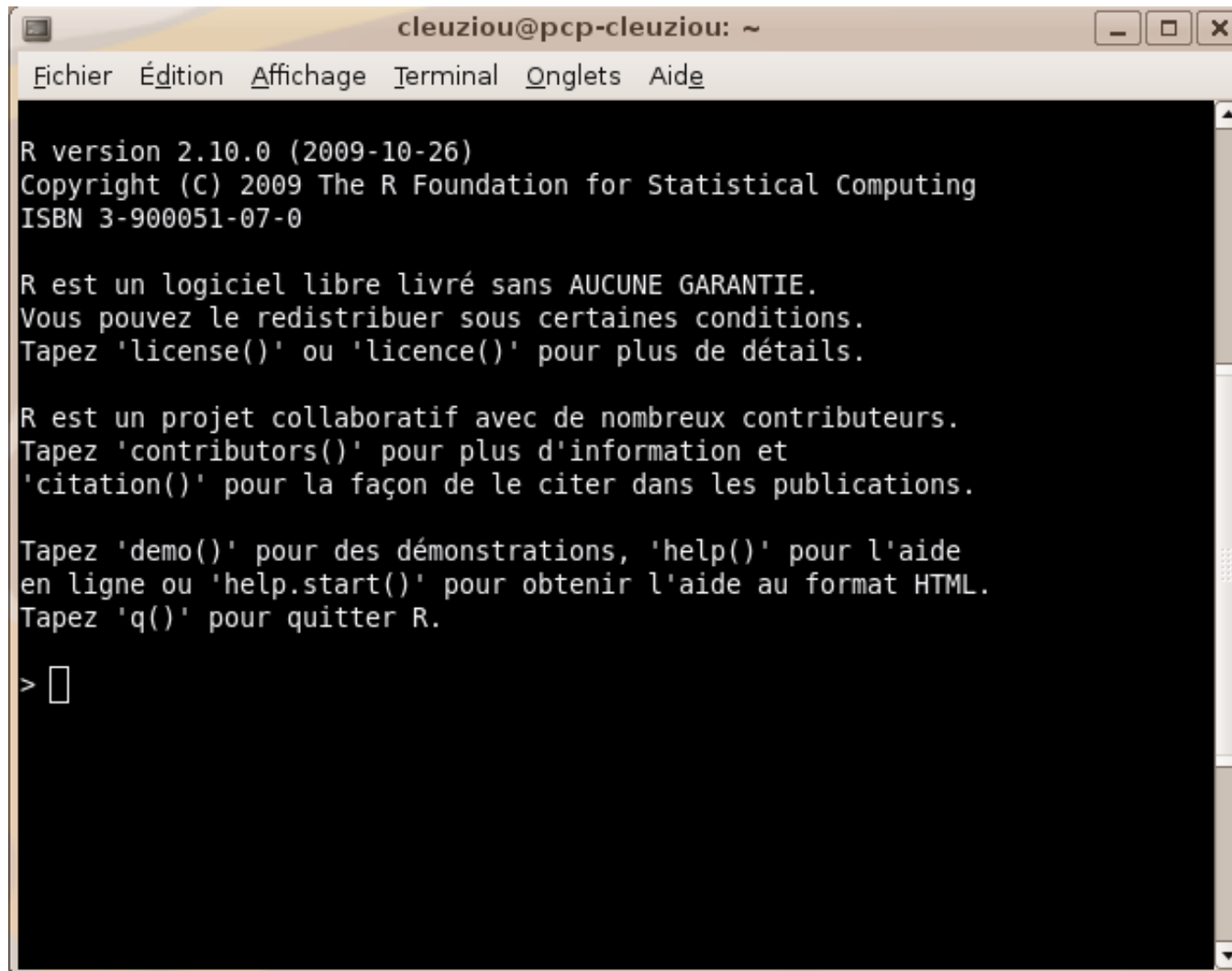


Présentation de R

- R est langage (logiciel) permettant de réaliser des analyses statistiques (manipulation de données, calcul de représentations, graphiques).
- R est sous licence GPL (Gnu Public Licence)
- Porté sur plusieurs plate-formes : Linux, Mac, Windows
- Dépourvu d'interface graphique (ligne de commande)

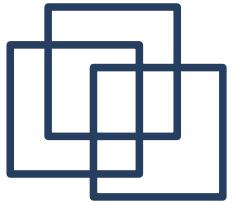


R sous Linux



The screenshot shows a terminal window titled 'cleuziou@pcp-cleuziou: ~'. The menu bar includes 'Fichier', 'Édition', 'Affichage', 'Terminal', 'Onglets', and 'Aide'. The terminal text displays the R version 2.10.0 (2009-10-26) copyright notice, followed by a disclaimer in French stating that R is free software without warranty. It provides instructions on how to view the license, contributors, and help. The prompt '>' is visible at the bottom.

```
cleuziou@pcp-cleuziou: ~  
Fichier  Édition  Affichage  Terminal  Onglets  Aide  
R version 2.10.0 (2009-10-26)  
Copyright (C) 2009 The R Foundation for Statistical Computing  
ISBN 3-900051-07-0  
  
R est un logiciel libre livré sans AUCUNE GARANTIE.  
Vous pouvez le redistribuer sous certaines conditions.  
Tapez 'license()' ou 'licence()' pour plus de détails.  
  
R est un projet collaboratif avec de nombreux contributeurs.  
Tapez 'contributors()' pour plus d'information et  
'citation()' pour la façon de le citer dans les publications.  
  
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide  
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.  
Tapez 'q()' pour quitter R.  
  
> 
```

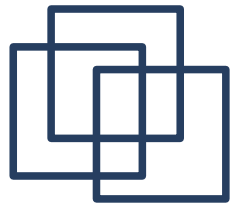


Obtenir R

Fichiers sources et documentation disponibles sur le site du CRAN (Comprehensive R Archive Network).

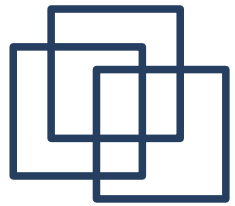
<http://cran.r-project.org/>





Notions élémentaires

- **Langage orienté objet** : le résultat d'une instructions peut être stocké dans un objet (réutilisation)
- **Erreur de syntaxe** : `Error : syntax error`
- **Instruction incomplète** : signe « + »
- **Opérateur d'affectation** : « <- » ou « = »
`X<-10` on affecte à l'objet X la valeur 10
- **Séparateur d'instructions** : « ; »
`X<-10 ; Y<-5`
- **Commentaires** : « # »
`Z<-X+5 # Z=X+5`



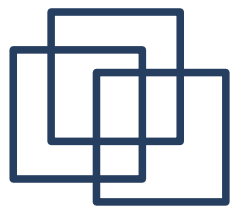
Notions élémentaires

- **Historique des commandes** : flèches ↑ et ↓
- **Gestion des objets** :
 - Liste des objets existants : `ls()`
 - Suppression d'un objet : `rm(nom_objet)`
 - Suppression de tous les objets : `rm(list=ls())`



Répertoire de travail

- Sauvegarde des objets :
`save.image("nomfic.RData")`
- Chargement des objets sauvegardés :
`load("nomfic.RData")`
- Modification du répertoire de travail :
`setwd("chemin d'accès")`



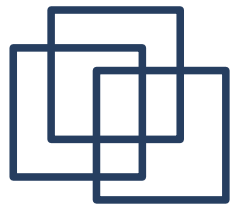
Répertoire de travail

Sorties-écran dans la console (par défaut).

Re-direction possible vers un fichier : `sink(nom_fic.txt)`

Exemple :

```
> X<-10  
> Y<-X+5 # Y=X+5  
> Z<-Y  
> sink("rep_R/test.txt")  
> X;Y;Z  
> sink()
```



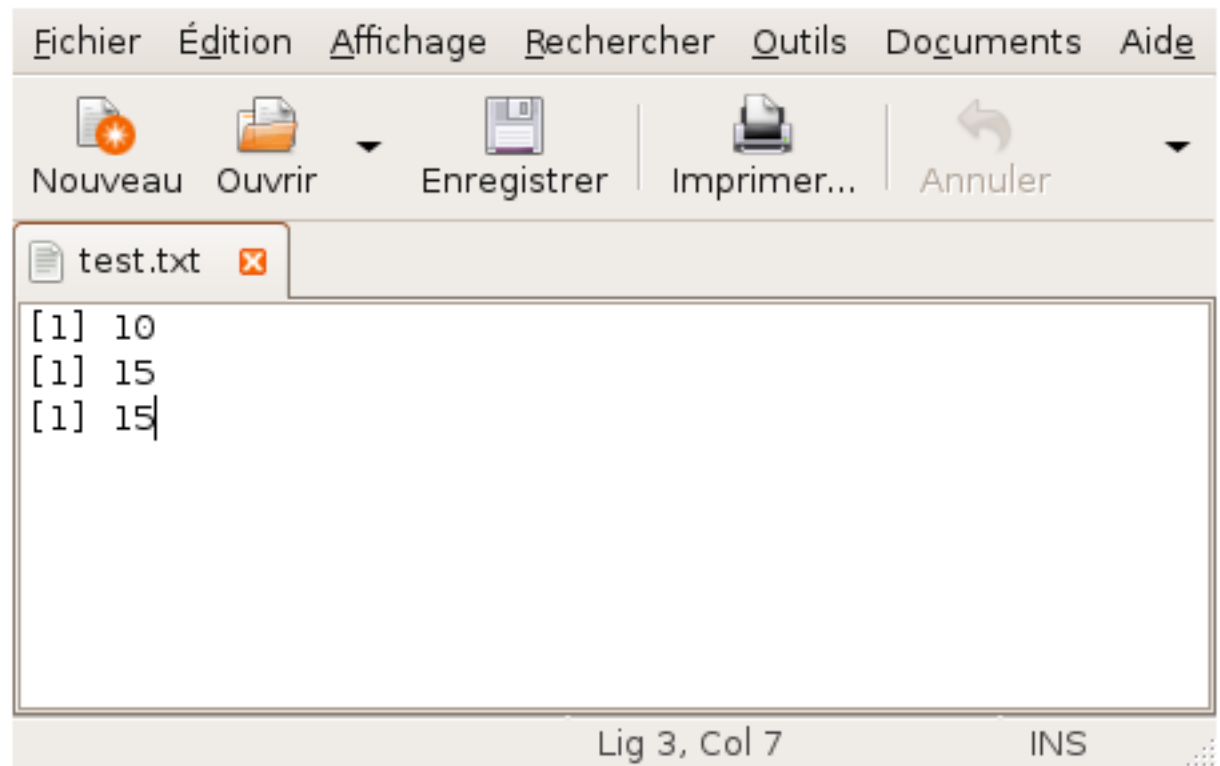
Répertoire de travail

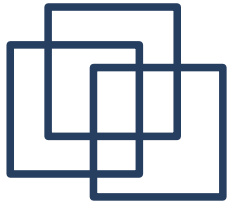
Sorties-écran dans la console (par défaut).

Re-direction possible vers un fichier : `sink(nom_fic.txt)`

Exemple :

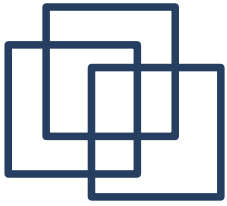
```
> X<-10  
> Y<-X+5 # Y=X+5  
> Z<-Y  
> sink("rep_R/test.txt")  
> X;Y;Z  
> sink()
```





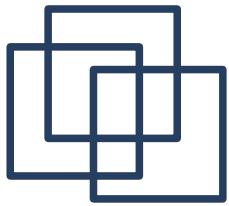
Aide en ligne

- R propose une aide en ligne : `help.start()`
- Pour une commande particulière : `?nom_commande`
- Recherche d'une commande : `apropos("mot_cle")`



AIDE A LA DECISION

Le langage R : les objets



Les objets

Les objets sont des éléments de structure définis :

- Un nombre est appelé **numeric**. Le séparateur de décimal est le point.

Exemple : 2, 0.3, 2.56E10, ...

- Un caractère ou une suite de caractères est appelée **character**.

Exemple : "a", 'abc', "boum boum", ...

Guillemets simples ou doubles

- Un booléen (vrai/faux) est appelé **logical**.

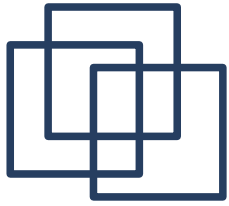
Exemple : TRUE, FALSE, T, F

Dans certains contextes, 1 est vrai, 0 est faux.

- Les facteurs sont les modalités d'une variable qualitative

Exemple : "bleu", "2,5", "5", "oui", ...

Les *factors* sont les variables, les *levels* sont les modalités possibles ordonnées ou non.



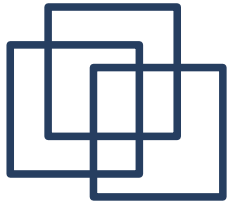
Les objets

On peut manipuler directement ces valeurs :

```
> 'plouf'
[1] "plouf"
> 4
[1] 4
> 5+6
[1] 11
> 4==5
[1] FALSE
> T
[1] TRUE
```

On peut également les stocker dans des variables :

```
> a<-T
> a
[1] TRUE
> class("a")
[1] "character"
> class(a)
[1] "logical"
```



Les variables

Sous R les noms de variables et fonctions sont sensibles à la casse.

```
> x<-5 ; X<-6 ; x ; X
```

```
[1] 5
```

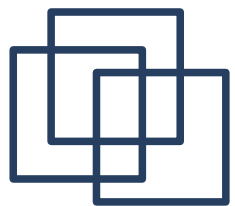
```
[1] 6
```

Les variables prennent le type du contenu au moment de l'affectation.
On peut également forcer le type avant d'affecter une valeur :

```
> b<-as.character(10)
```

```
> b
```

```
[1] "10"
```



Commandes génériques

R fournit un ensemble de fonctions génériques pour gérer ces types:
`as.<type>(quelquechose)`

```
> as.character(4)
```

```
[1] "4"
```

```
> as.numeric("e")
```

```
[1] NA
```

```
Warning message:
```

```
NAs introduced by coercion
```

```
> as.numeric("4")
```

```
[1] 4
```

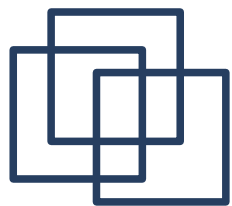
```
> as.logical(4)
```

```
[1] TRUE
```

```
> as.logical(0)
```

```
[1] FALSE
```

elle convertit, **si possible**, le contenu des parenthèses dans le type voulu.



Commandes génériques

`is.<type>(quelquechose)`

```
> is.character()
```

```
Error: 0 arguments passed to 'is.character' which requires 1
```

```
> is.character("R")
```

```
[1] TRUE
```

```
> is.character(4)
```

```
[1] FALSE
```

```
> is.logical(T)
```

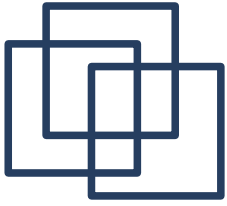
```
[1] TRUE
```

On récupère une valeur logique indiquant si la valeur ou la variable est du type indiqué.

C'est la fonction *is.na* qui permet de tester la présence ou l'absence de valeurs manquantes.

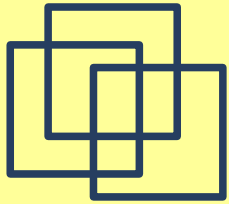
```
> is.na(as.numeric("e"))
```

```
[1] TRUE
```



AIDE A LA DECISION

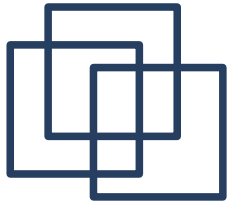
Le langage R : les objets « complexes »



Objets complexes

Des objets plus complexes peuvent être dérivés des types simples:

- les vecteurs, *vector* : une liste ordonnée de valeurs (types simples).
- les matrices, *matrix*: un tableau à 2 dimensions (valeurs)
- les *data.frame*: un ensemble de tableaux de même dim. (valeurs)
- les listes, *list* : un ensemble d'objets R de dimension quelconque



Les vecteurs

C'est une liste ordonnée de valeurs de même type (simple).
Peut se déclarer:

```
> c(5,6,9,6)
```

```
[1] 5 6 9 6
```

```
> c("a","b","c","d")
```

```
[1] "a" "b" "c" "d"
```

```
> rep(0,4)
```

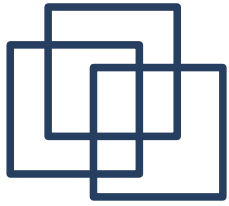
```
[1] 0 0 0 0
```

```
> seq(from=0,to=10,by=2)
```

```
[1] 0 2 4 6 8 10
```

```
> c(1:5) # raccourci pour seq(from=1,to=5,by=1)
```

```
[1] 1 2 3 4 5
```



Les matrices

Ce sont des tableaux à 2 dimensions. Les matrices se déclarent par la syntaxe suivante :

```
matrix( valeurs,nrow=nb_de_lignes, ncol=nb_colonnes,byrow=F )
```

ncol et *nrow* : les dimensions de la matrice.

Byrow : indique le mode de remplissage de la matrice

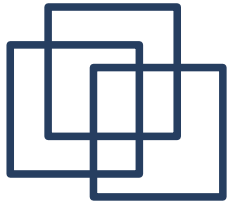
Valeurs : correspond aux valeurs à placer dans la matrice.

```
> matrix(c(1,2,3,4),2,2)
```

```
  [,1] [,2]  
[1,]  1  3  
[2,]  2  4
```

```
> matrix(rep(0,12),3,4)
```

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

Les matrices

Si le nombre de valeurs est insuffisant pour remplir la matrice, les valeurs sont recyclées:

```
> matrix(c(1,2),2,2)
```

```
  [,1] [,2]
```

```
[1,]  1  1
```

```
[2,]  2  2
```

```
> matrix(c(1,2),2,2,T)
```

```
  [,1] [,2]
```

```
[1,]  1  2
```

```
[2,]  1  2
```

```
> matrix(c(1,2,3),2,2)
```

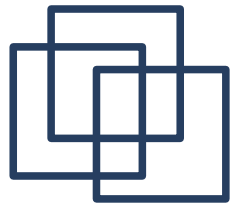
```
  [,1] [,2]
```

```
[1,]  1  3
```

```
[2,]  2  1
```

Message d'avis :

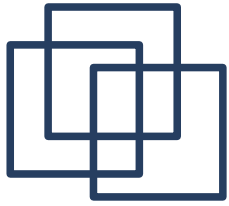
In matrix(c(1, 2, 3), 2, 2) : la longueur des données [3] n'est pas un diviseur ni un multiple du nombre de lignes [2]



Les tableaux à n-dimensions

Ils fonctionnent comme les matrices mais avec n dimensions.
Par exemple pour un cube 3x3 :

```
> array( data=c(1,2,3), dim=c(3,3,3) )  
,,1  
  [,1] [,2] [,3]  
[1,] 1 1 1  
[2,] 2 2 2  
[3,] 3 3 3  
,,2  
  [,1] [,2] [,3]  
[1,] 1 1 1  
[2,] 2 2 2  
[3,] 3 3 3  
...
```

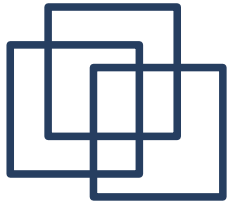


Les listes

Ce sont des assemblages (ordonnés) d'objets R de tout type (vecteur, matrice, liste, etc.)

```
> aa<-list( matrix(0,3,3) , c(1:4) , "abc" )
```

```
> aa
```



Les listes

Ce sont des assemblages (ordonnés) d'objets R de tout type (vecteur, matrice, liste, etc.)

```
> aa<-list( matrix(0,3,3) , c(1:4) , "abc" )
```

```
> aa
```

```
[[1]]
```

```
  [,1] [,2] [,3]
```

```
[1,]  0  0  0
```

```
[2,]  0  0  0
```

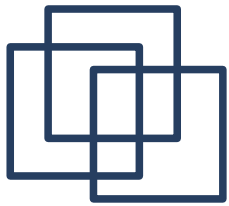
```
[3,]  0  0  0
```

```
[[2]]
```

```
[1] 1 2 3 4
```

```
[[3]]
```

```
[1] "abc"
```



Nommer les dimensions

Pour nommer les dimensions d'un objet, on utilise un vecteur *character*.

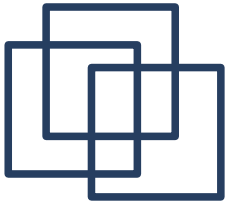
- Pour un vecteur :

```
> vv<-c(1,2,3,4,5,6)
> names(vv)<-c("d1","d2","d3","d4","d5","d6")
>vv
d1 d2 d3 d4 d5 d6
1  2  3  4  5  6
```

- Pour une matrice :

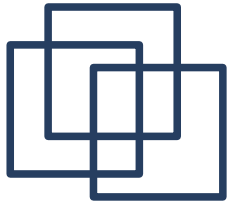
```
> aa<-matrix(c(1,2,3,4),3,4)
> colnames(aa)<-c("A","B","C","D")
> rownames(aa)<-c("X","Y","Z")
```

```
> aa
  A B C D
X 1 4 3 2
Y 2 1 4 3
Z 3 2 1 4
```



AIDE A LA DECISION

Le langage R : Calculs sous R



Calculs numériques

On peut réaliser des opérations de calcul sous R. On modifie la précedence en utilisant des parenthèses :

```
> 3*6
```

```
[1] 18
```

```
>(5+6)*5
```

```
[1] 55
```

```
> 5+6*5
```

```
[1] 35
```

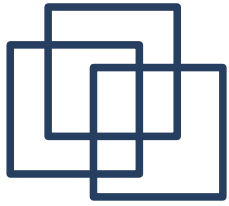
opérateurs simples : +,-,*,/ mais également :

%/% division entière

%% reste de la division entière

%*% produit matriciel

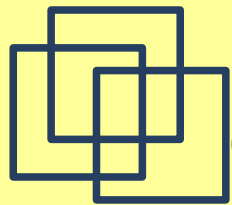
^n puissance n



Calculs booléens

Opérateurs logiques :

&	ET logique
	OU logique
== et !=	respectivement égalité et différent
>, <, >=, <=	inégalités
!	négation



Calculs sur les objets complexes

Par défaut, on réalise les opérations sur tous les éléments des objets :

```
> c(1,2,4)+1
```

```
[1] 2 3 5
```

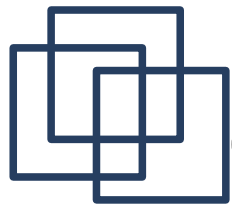
```
> aa<-matrix(1:9,3,3,T)
```

```
> aa
```

	[,1]	[,2]	[,3]
[1,]	1	2	3
[2,]	4	5	6
[3,]	7	8	9

```
> aa+1
```

	[,1]	[,2]	[,3]
[1,]	2	3	4
[2,]	5	6	7
[3,]	8	9	10



Calculs sur les objets complexes

Par défaut, on réalise les opérations sur tous les éléments des objets :

```
> c(1,2,4)+1
```

```
[1] 2 3 5
```

```
> aa<-matrix(1:9,3,3,T)
```

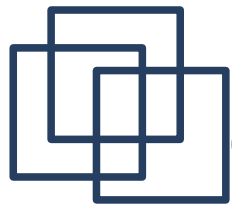
```
> aa
```

	[,1]	[,2]	[,3]
[1,]	1	2	3
[2,]	4	5	6
[3,]	7	8	9

```
> aa+c(1,2,3)
```

	[,1]	[,2]	[,3]
[1,]	2	3	4
[2,]	6	7	8
[3,]	10	11	12

R conforme l'objet pour que l'opération soit possible.



Calculs sur les objets complexes

Par défaut, on réalise les opérations sur tous les éléments des objets :

```
> c(1,2,4)+1
```

```
[1] 2 3 5
```

```
> aa<-matrix(1:9,3,3,T)
```

```
> aa
```

	[,1]	[,2]	[,3]
[1,]	1	2	3
[2,]	4	5	6
[3,]	7	8	9

```
> aa+c(1,2)
```

```
      [,1] [,2] [,3]
```

```
[1,]  2   4   4
```

```
[2,]  6   6   8
```

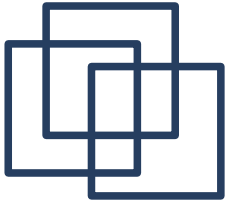
```
[3,]  8  10  10
```

Message d'avis :

In aa + c(1, 2) :

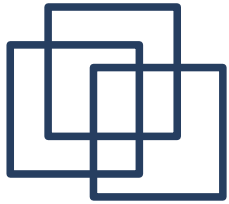
la taille d'un objet plus long n'est pas multiple
de la taille d'un objet plus court

Lorsque c'est possible !!!



AIDE A LA DECISION

Le langage R : Accès aux éléments des
objets complexes



Accès aux éléments d'un vecteur

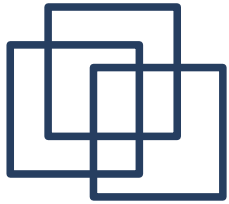
De plusieurs manières :

- Par leur indice (e.g. `x[2]`) **A partir de 1**
- Par leur nom si les dimensions sont nommées (e.g. `x["nom_dim"]`)
- Par des vecteurs logiques et/ou des conditions (e.g. `x[c(F,F,T)]`)

Exemple d'accès par l'indice :

```
> x<-c(4,6,8,5,2)
> x[3]
[1] 8
> x[-3]
[1] 4 6 5 2
```

Indice positif pour renvoyer l'élément correspondant. Indice négatif pour renvoyer tous les éléments sauf l'élément correspondant.



Accès aux éléments d'un vecteur

De plusieurs manières :

- Par leur indice (e.g. `x[2]`) **A partir de 1**
- Par leur nom si les dimensions sont nommées (e.g. `x["nom_dim"]`)
- Par des vecteurs logiques et/ou des conditions (e.g. `x[c(F,F,T)]`)

Exemple d'accès par l'indice :

```
> x<-c(4,6,8,5,2)
> x[3]
[1] 8
> x[-3]
[1] 4 6 5 2
```

Par un vecteur d'indices :

```
> x[c(1,3,5)]
[1] 4 8 2
> x[-c(1,3,5)]
[1] 6 5
```

Indice positif pour renvoyer l'élément correspondant. Indice négatif pour renvoyer tous les éléments sauf l'élément correspondant.