

Initiation R

E.Hirchaud

BiRD

3 septembre 2013



Sommaire

1 Notions informatiques

- L'ordinateur
- Système d'exploitation
- Les applications
- Chemins et fichiers
- Types de mémoires
- Synthèse

2 R

- Apperçu
- Objectif du TP
- Comparaison Excel R



L'ordinateur



Les système d'exploitation : OS (Operating System)



debian



FreeBSD



solaris



Les applications : logiciels, programmes,

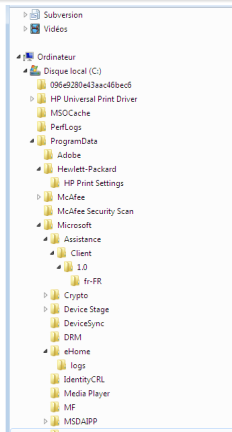


Arborescence

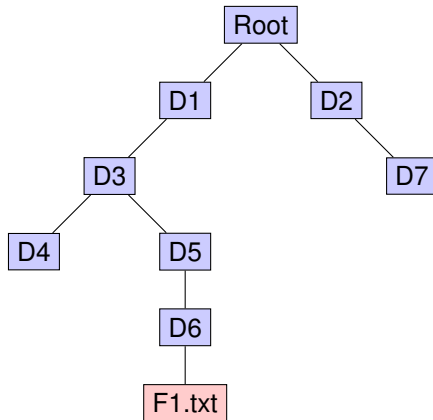
Apprélation

- Fichiers données numériques de type variés
- Directory = Répertoire = Dossier : contient des fichiers.

bin			- dossier
Documents			- dossier
Formation			- dossier
Formation_puce2012			- dossier
data			- dossier
15K_alea.atr	755 octets	inconnu	
15K_alea.cdt	3,0 Mo	inconnu	
15K_alea.gtr	716,4 ko	inconnu	
15K_alea.txt	3,8 Mo	document texte brut	
annotation_echantillons.txt	291 octets	document texte brut	
bg_matrix_lowess.txt	15,1 Mo	document texte brut	
matrix_lowess.txt	10,8 Mo	document texte brut	
Resultats			- dossier
filtrage_invariants			- dossier
filtrage_non_exprimes			- dossier
filtrage_non_exprimes_Excel			- dossier
matrix_totale.filtree.txt	5,7 Mo	document texte brut	
Scripts			- dossier
Feuille_route_TP_filtrage.doc	227,3 ko	Document Microsoft Word	
Formation_R_2013			- dossier
image			- dossier
logo.jpg	72,5 ko	image JPEG	
OS.png	66,4 ko	image PNG	



Shematisation



La navigation

Deux types de navigation

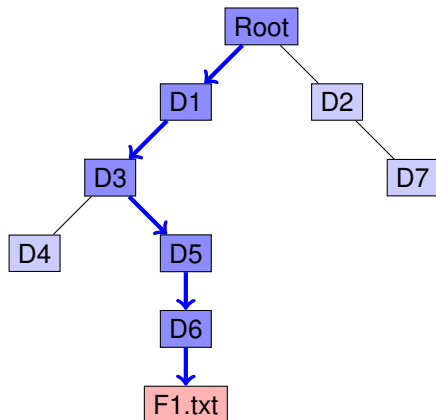
- Chemin absolu
- Chemin relatif

Symboles utilisés dans la navigation

- La racine : Windows une lettre, linux et mac, symbole /
- Séparateur de répertoire : Windows : \, linux et mac /
- Le Répertoire courant : . (point)
- Le Répertoire parent : .. (deux points)



Formalisation : Chemin absolu

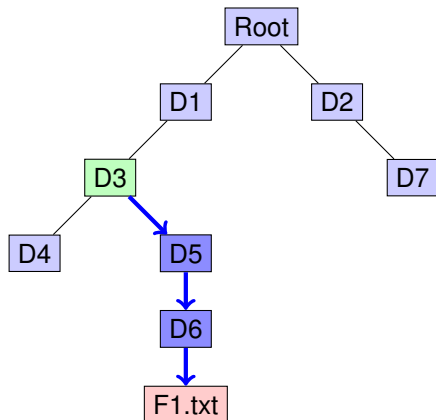


Chemin absolu

- A partir de la racine (C:)
- Root/D1/D3/D5/D6/F1.txt



Formalisation : Chemin relatif exemple 1

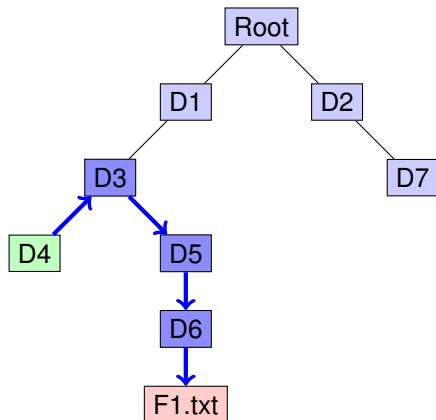


Chemin relatif

- D5/D6/F1.txt



Formalisation : Chemin relatif exemple 2



Chemin relatif

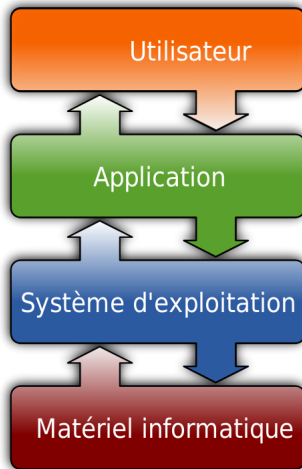
- ../D3/D5/D6/F1.txt



Les mémoires



Synthèse



R : Apperçu

- Créée par Ross Ihaka et Robert Gentleman (1996)
- C'est un logiciel libre et gratuit
- Il est basé sur le langage S qui est propriétaire
- Disponible sur les systèmes d'exploitation les plus utilisés
- Utilisé dans de nombreux domaines dont la bio analyse.



Objectif du TP

- Assimiler le vocabulaire
- Se servir de R comme d'une calculatrice
- Écrire et modifier des lignes de commande
- Utiliser un script déjà écrit
- Savoir où trouver de l'aide (documentation)
- Utiliser un éditeur convivial (RStudio)



Comparaison Excel R

Excel	R
Cellule	une valeur (Variable)
Plage de données	data.frame matrix, list, vector
Valeur	Valeur (value)
Format	Type
Fonction	Fonction
Macro	Script



Généralité sur les variables

- Un nom
- Valeur(s)
- Les valeurs ont un type :
 - ▶ numérique : 1,2, 3.14
 - ▶ chaîne de caractères : A,B genes
 - ▶ logique : TRUE/FALSE



Catégories de variables

- vector -> vecteur (type homogène)
- matrix -> matrice (type homogène)
- data.frame -> tableau de données (type hétérogène)
- factor -> classe de paramètres (type homogène)
- list -> liste(type hétérogène)



Les Fonctions

- Créent, modifient et informent sur les données
- Contiennent des arguments et des instructions



Règles de nomenclatures

- Importance de la casse (majuscule/minuscule)
 - ▶ : pizza \neq Pizza
- Informatique anglo-saxone
 - ▶ Ne pas nommer les noms des objets avec des accents
 - ▶ Le point sert de décimal, la virgule non !
- Ne JAMAIS mettre d'espace dans un nom
- Ne JAMAIS commencer un nom par un chiffre
- Eviter d'utiliser des symboles (+ - / ...)

