



RENCONTRE BIMENSUELLE AUTOUR DES LOGICIELS LIBRES



**Environnement Mathématique
et Langage de Programmation
Destiné à l'Analyse Statistique**

Logiciel de statistique : contrairement à d'autres logiciels de calcul numérique (Scilab, Octave).

Fonctions pour effectuer des traitements statistiques non triviaux :

- classiques (régression, analyse de la variance, arbres de décision, analyse en composantes principales, etc.) ou
- plus modernes (réseaux de neurones, bootstrap, etc.)

Langage de programmation : on n'est pas limité par l'imagination des concepteurs du logiciel. Implémentation libre de s-plus.

Langage interprété : programmation plus simple, mais exécution plus lente. Utilisable sans vraiment avoir de notions de programmation.

Un tableur : remplir un tableau avec du texte, des nombres et des formules

Quelques remarques à l'usage :

- + Rapide à mettre en œuvre et très répandu
- + Visualisation permanente des données traitées
- Sélection acrobatique des cellules non-contigües
- Un seul état : le résultat d'une formule
- Représentations graphiques limitées aux choix proposés

R : un « environnement » statistique

- + Un langage : grande liberté et possibilité d'automatiser (comme VBa, mais en plus compatible et portable)
- + Fonctions statistiques avec une aide détaillée et une communauté
- + Traçabilité des opérations effectuées sur les données
- Données « objets » : on ne les voit pas si on ne le demande pas
- Liberté importante : souvent noyé sous les arguments d'une fonction

Un logiciel libre pour les principaux systèmes d'exploitation :

- C.R.A.N : The Comprehensive R Archive Network : Site ressource
- Windows : un installateur, une très bonne interface graphique tierce
- Linux : un paquet pour la majorité des distributions, interface basique
- OSX : un paquet et une interface intégrée « luxueuse »

> Un confort d'utilisation différent mais le cœur est complètement identique.

Pour lancer R, taper « R » dans une console : une invite pour interface.

- C'est un langage interprété : R interprète une commande saisie
- Utilisé conjointement à un éditeur de texte (simple ou pensé pour R)
- Copie des commandes que l'on veut exécuter.

> La saisie de données est donc délicate, R peut être vu comme complémentaire à un tableur.

```
woinnie@hetre:~$ R

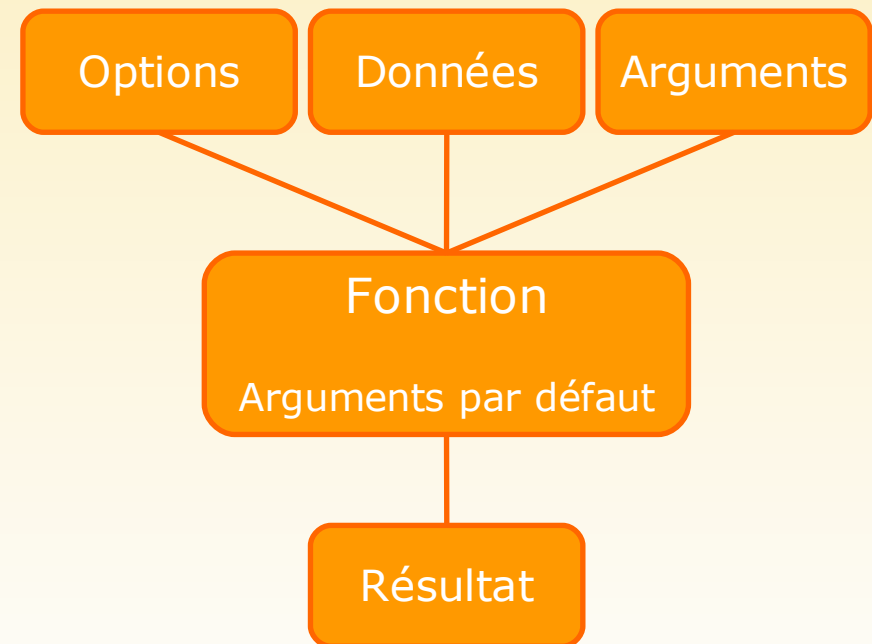
R : Copyright 2006, The R Foundation for Statistical Computing
Version 2.3.1 (2006-06-01)
ISBN 3-900051-07-0

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.

> a=3
> a
[1] 3
> b="Voilà donc cette interface"
> b
[1] "Voilà donc cette interface"
>
```



Depuis un fichier log de logiciel de synthèse d'image

- Création, import et export de données
- Graphiques
- Statistiques descriptives

Depuis des mesures architecturales de tournesol

- Régressions linéaires, non linéaires et multiples

Depuis des données agronomiques de tournesol

- Test paramétriques
- Analyse de variance

Création de données et types d'objets

- Assigner une valeur à un objet (numérique, chaîne de caractère, booléen)
- Quels types d'objets dans R ?
 - Vecteurs : une suite de valeurs numériques ou texte
 - Facteurs : la même chose mais pour des données qualitatives
 - Matrices : tableaux à 2 dimensions
 - Dataframe : tableaux avec des colonnes de type différents
- Exemples de génération de séquences

Importation de données

- Depuis un fichier texte : fonctions `scan()` ; `read.table()`
- Depuis une base de données : `SQLquery()` ...

Graphiques

- Gérés sous forme de « périphériques graphiques » : on choisit dans quel dispositif écrire : `dev.list()` ; `dev.cur()` ; `dev.set()`.
 - Une fenêtre graphique par défaut
 - Des fichiers images ou vectoriels
- L'apparence, légendes, axes, etc. sont personnalisables à travers des l'utilisation de fonctions spécifiques
- Il existe une autre bibliothèque (groupe de fonctions) graphique un peu plus sympathique avec l'utilisateur

Résumer ses données

- En les visualisant, pour repérer les grandes tendances (histogrammes, graphes bivariés)
- En utilisant des indicateurs statistiques classiques (moyenne, quantiles, écart-type) de manière numérique ou graphique

Prédire une variable depuis une autre : modèles

- Le cadre d'étude : l'architecture du tournesol



Longueur x largeur
Mesures



Prédire :

- La surface d'une feuille depuis ses dimensions : **régression linéaire**
- La surface de toutes les feuilles depuis certaines mesures : **régression non-linéaire**
- La surface de la plante de manière simple : **régression linéaire multiple**

Comparer ses données : est-ce « significativement » différent ?

- Réaliser des test paramétriques (test de moyennes) : toute une série de test est présente dans R
- Analyse de variance : test de moyenne généralisés : identifier statistiquement les effets de différents facteurs