

Introduction à R

Les types de données

Pascal Bessonneau

06/2015

Les types de données

Compléments sur les types de base

Typage

En informatique, toutes les informations sont stockées en binaire. 00111001 peut représenter aussi bien un nombre qu'un caractère. Les types sont la nature des données que l'on peut stocker.

Typage

En R, il n'existe pas de typage fort, c'est-à-dire qu'une nouvelle variable est défini comme caractère, numérique en fonction de la première valeur qu'on mets dedans (la plupart du temps). De plus une variable peut changer de type simplement en remplaçant le contenu par un autre. Toutefois il existe des exceptions.

Les types de base

Les différents types sont les suivants :

integer ou entier en français. Il permettent de sauvegarder des nombres non décimaux

numeric ce type sert à stocker des nombres décimaux

complex ce type permet de stocker des nombres avec partie entière et partie imaginaire des complexes

character permet de stocker des chaînes de caractères

logical permet de stocker des booléens (Vrai/Faux)

Les types de base

les dates , comme souvent les dates sont des objets délicats à manipuler. Les termes qui font référence à des type différents sont : *POSIXct*, *POSIXlt*, ...

factor les *factor* ce sont des modalités de variable ordinale ou qualitative

AsIs permet de stocker des données en utilisant l'opérateur *I()*. En pratique il se rencontre dans des fusions souvent c'est un signe que vous devriez vous inquiéter des résultats obtenus par la fusion

utilisateurs des packages (ou vous même) vous pouvez créer des types nouveaux

Les booléens

Les booléens sont des *TRUE* ou *FALSE* qui sont deux mots clefs de R. Ils peuvent être abrégés en *T* ou *F*.

Les booléens peuvent être convertis en numériques naturellement ou explicitement.

Dans ce cas leurs valeurs sont 0 pour FALSE et 1 pour TRUE. Ce qui autorise de faire des sommes de booléen.

Les *factor*

Les *factor* permettent de stocker des variables ayant peu de modalités. C'est par exemple le cas pour stocker : expérience A ou B, médicament Placebo ou Actif. L'étiquette est stockée sous forme de texte mais les facteurs sont aussi manipulables sous forme de nombres.

Cette dualité qui est très utile dans un laboratoire pour faire des ANOVA est un vrai problème quand on stocke des données complexes.

Si les facteurs sont indispensables pour faire certaines analyses comme les tests de Tukey certains auteurs font tout (ou presque) pour éviter leur utilisation comme H. Wickham.

La précision des nombres

Souvent, par exemple pour des entiers lors de sondages, il est nécessaire de s'inquiéter de la précision des types numériques. Ces limites peuvent dépendre de la machine utilisée.

Ces informations sont contenues dans l'objet *.Machine* que l'on peut taper directement ou on peut appeler seulement une partie. Par exemple pour trouver les entiers maximums que l'on peut stocker sur la machine qui a compilé ce document :

```
> .Machine$integer.max
```

```
## [1] 2147483647
```