

Préparer ses données avec R et le Tidyverse

Groupe des référents R

2018-05-15

Chapitre 1 Introduction

DATA



SORTED



ARRANGED



PRESENTED
VISUALLY



Ce module va vous permettre de découvrir un ensemble de méthodes sous R afin de préparer ses données. Préparer ses données sous R, cela veut dire :

- Savoir les importer dans un environnement R

- Mettre ses données dans de bons formats (date, catégorielle) et gérer les données manquantes
- Rajouter des variables en fonction de variables existantes
- Regrouper des modalités de variables
- Joindre des tables entre elles pour obtenir des informations de plusieurs sources
- Aggréger des données
- Bien définir notre table de travail en fonction des indicateurs à analyser et à leurs dimensions d'analyse ...

Bref, tout le travail technique préalable entre la collecte de la donnée et la valorisation proprement dite de ces sources. On estime qu'un scientifique de la donnée passe en général ***la moitié de son temps*** à cela.

Sous R, comme souvent, il y a plusieurs façons d'aborder cette question. Nous choisirons, lors de ce module de formation, d'explorer principalement les packages du framework *tidyverse*, qui ont l'avantage d'aborder ces différentes questions d'une façon intégrée et cohérente, d'une part entre elles, mais également avec d'autres.