

Data Science, Séance 5 : dplyr, tidyr

Etienne Côme

21 novembre 2019

dplyr, tidyr

2 libraries pour faciliter la manipulation de données (Cheatsheet) !
Introduction d'un nouvel opérateur ! Enchaîner des opérations
introduction de l'opérateur de pipe :

%>%

Faciliter la lecture du code produit

dplyr, tidyr

Sélection de lignes “filter”

dplyr, tidyr

Sélection de colonnes “select”

dplyr, tidyr

Transformation “mutate”

dplyr, tidyr

Résumé “summarise”

Avec des données groupées “group_by”

dplyr, tidyr

jointure "X_join"

dplyr, tidyr

Remise en forme

gather : format large -> format long

spread : format long -> format large

separate : split d'une colonne

unite : concatenation de colonnes

dplyr, tidyr, exercice

Prénoms

Faire une carte représentant le prénoms masculin le plus fréquemment donnés aux enfants nés en 2005 pour l'ensemble des départements français. Les données à utiliser sont disponibles dans le répertoire data :

- ▶ ./Data/exo5_dep.csv : centroid des départements
- ▶ ./Data/exo5_prenoms.csv : Fichier des prenomms

dplyr, tidyr, exercice