

# La science des données avec R

Fousseynou Bah

Faculté des Sciences Economiques et de Gestion (FSEG)  
Université des Sciences Sociales et de Gestion de Bamako (USSGB)

19-Mar-2019

- 1 Introduction générale
- 2 Les basiques dans R
- 3 Programmer dans R
- 4 Manipuler des données dans R
- 5 Visualiser des données dans R

# Introduction générale

# Science des données: *data science*

- La *data science*
- Le *data scientist*

# Les basiques dans R

# Introduction à R

- R
- Rstudio

# Objets dans R

- Notion d'objets
- Vecteurs
- Matrices
- *Data frames*
- Listes

# Programmer dans R



# S'exprimer dans R

- Les déclarations
- Les boucles
- Les fonctions

# Manipuler des données dans R

# Importer des données dans R

- Fichiers plats: format CSV
- Excel
- Autres formats: Stata, SPSS
- Données relationnelles
- Depuis Internet

# Transformer des données dans R avec **dplyr**

- Aperçu de **dplyr**
- Sélection et exclusion de variables: `select`
- Création et suppression de variables: `mutate`
- Sélection d'observations: `filter`
- Tri d'observations: `arrange`
- Vers l'agrégation: `group_by` et `summarize`

# Transformer des données dans R avec **tidyr**

- Du format large au format long: `gather`
- Du format long au format large: `spread`
- Séparateur d'attributs: `separate`
- Unificateur d'attributs: `unite`

# Visualiser des données dans R

# Visualiser des données dans R avec ggplot2

- La gamme des graphiques
- Points: à la recherche d'une relation
- Lignes: comprendre une évolution
- Barres: allier discret et continu (ou discret)
- Histogrammes et densités