
LNG-1100 : Méthodes expérimentales et analyse de données

Les variables binaires et le nettoyage des données

Guilherme D. Garcia

fr.gdgarcia.ca

10



Plan de la séance

1. Des variables catégorielles (binaires)
2. Pratique en R

Questionnaire 1

Collecte de données : forms.office.com/r/fcKjiCAx9p



Questionnaire 2

Collecte de données : forms.office.com/r/9yfdeMsw2W



Un nouveau script

Pratique : des données réalistes

1. Importez les données `questionnaire.csv` : résultats à partir de MS Forms
2. Le tableau est-il *tidy*?
3. Quels problèmes constatez-vous dans le tableau?

Un nouveau script

Nos données sont fréquemment chaotiques :

- format *wide* (pas *tidy*);
- plusieurs colonnes inutiles;
- noms longs pour les variables/colonnes;
- **NA**s.

☞ Heureusement, on peut régler tous ces problèmes dans notre script

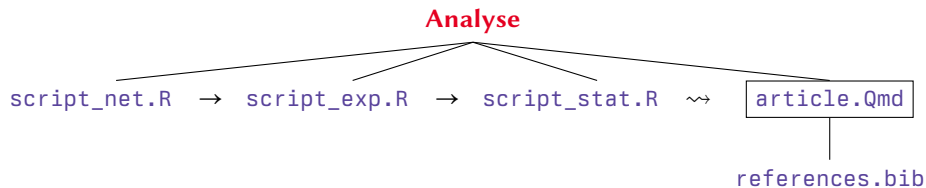
Une structure de travail organisée

L'analyse de données contient plusieurs étapes :

- importation et **nettoyage** des données;
- visualisation des patrons;
- analyse statistique.
- Communication des résultats (Quarto etc.)

☞ Jusqu'ici, on a utilisé un seul script pour toutes les étapes (pas idéal!)

Une structure de travail organisée



- Chaque **script** est responsable d'une étape : nettoyage, exploration, analyse statistique
- 👉 On utilise la fonction `source("script_net.R")` pour charger un script

Une structure de travail organisée

Aujourd'hui

- On travaille dans la **première étape** : `script_10_net.R`
- C'est un script normal : on charge les extensions et on importe les données

Nettoyage

- Importez et nettoyez le fichier `questionnaire2.csv`
- N'oubliez pas d'enlever la colonne qui indique le nom des participants

Bonus

- Quel est le score de chaque participant?
- Le goût pour la géographie affecte-t-il la précision des réponses?
- Créez une figure que vous considérez appropriée.