

---

# LNG-1100 : Méthodes expérimentales et analyse de données

Les variables binaires et le nettoyage des données

Guilherme D. Garcia

[fr.gdgarcia.ca](http://fr.gdgarcia.ca)

11



# Plan de la séance

1. Révision : nettoyage
2. Figures pour les données binaires
3. Combinaison des scripts à partir de `source()`



## Nettoyage

- Importez et nettoyez le fichier `questionnaire2.csv` (séance 10)
- N'oubliez pas d'enlever la colonne qui indique le nom des participants

👉 Réponses : `oui, non, oui, non, oui, non, oui, non, oui, non`

- Si vous avez déjà complété l'exercice :
  - Enchaînez les opérations dans une seule séquence de fonctions en utilisant le `>`



### Interprétation des patrons pertinents

- Quelle est la précision moyenne en fonction de la variable **langues**?
- Quelle est la précision moyenne en fonction de la variable **geo**?
- Quelle est la précision moyenne des participants 4, 7, 14, 15 et 17 (ensemble)?
- Créez une nouvelle colonne qui classe les participants qui aiment la géographie et qui parlent d'autres langues vs les autres. La précision de ces deux groupes est-elle affectée par la nouvelle classification?



### Interprétation des patrons pertinents

- Créez quelques figures pour explorer les patrons des données



# Pratique

## Partie IV

- On vient de créer des variables binaires/dichotomisées avec `if_else()`
- Cette fonction nous permet de recoder n'importe quelle variable catégorielle

### Pratique : le retour aux villes

1. Créer une nouvelle colonne, `langue`, qui dichotomise la langue officielle de la ville
2. Est-ce qu'il y a un effet de `langue` sur la `note`?



# Questionnaire 3

## Questions?

- Nettoyage de données + interprétation des patrons pertinents
- 5 questions; durée 2h50



# Prochaine séance

## Régression logistique

- Les séances 10 et 11 examinent une variable de réponse binaire : 0 ou 1
- ☞ Ce type de données ne peut pas être modélisé avec une régression linéaire (pas continu)

### La semaine prochaine : chapitre 11 du livre numérique du cours

- On explorera un nouveau type de modèle : régression logistique (Garcia 2021, chap. 7, Larmarange 2023, chap. 22)





# Références I

Garcia, G. D. (2021). *Data visualization and analysis in second language research*. Routledge, New York NY.

Larmarange, J. (2023). *Introduction à l'analyse d'enquêtes avec R et RStudio*. Available at <https://larmarange.github.io/analyse-R/analyse-R.pdf>.

