

Exercices et projets guidés

LA VISUALISATION DES DONNÉES ET LES TABLEAUX DE BORD

Entre les sessions

De la **session 1** à la **session 2**

- faites les exercices de la session 1
- téléchargez les ensembles de données depuis le site web

De la **session 2** à la **session 3**

- faites les exercices de la session 2

Après la **session 4**

- faites les exercices de la session 4
- essayez les projets guidés

De la **session 3** à la **session 4**

- faites les exercices de la session 3
- installez [R](#) / [RStudio](#) (Posit)
- lisez le [Programming Primer](#)(sections 1 - 4)
- lisez [Basic Visualizations with R](#)
- installez les paquets R suivants : tidyverse, ggplot2, dplyr

Projet guidé I

Identifiez un scénario pour lequel un tableau de bord pourrait s'avérer utile.

Déterminez les questions spécifiques auxquelles le tableau de bord pourrait aider à répondre ou les informations qu'il pourrait fournir. Identifiez les sources de données et les éléments de données qui pourraient être introduits dans votre tableau de bord.

Concevez un tableau (avec un stylo et du papier) avec des graphiques fictifs. Quels sont les points forts et les limites de votre tableau de bord ? Est-il fonctionnel ? Élégant ?

Utilisez certains des concepts de narration décrits dans le cours, au besoin.

Projet guidé II

Ce projet utilise l'[outil Gapminder](#) (il y a aussi une version [hors-ligne](#))

1. À quel moment du flux de travail de la science des données pensez-vous que des visualisations de cette nature pourraient être utiles ?
2. De quelle manière les observations peuvent-elles être anormales ? Avez-vous trouvé de telles anomalies ? Pouvez-vous offrir des explications ? Examinez en particulier le cas de l'Afrique du Sud en 2012, qui semble être une aberration manifeste. Suivez la trajectoire de la bulle sud-africaine de 1975 à 2020, par rapport au schéma général. L'explication de l'apartheid et de l'inégalité des revenus suggérée dans le texte ([DUDADS](#)) est-elle toujours valable ?
3. Choisissez 2+ visualisations "définitives" (méthodes, variables, etc.) autres que la configuration par défaut. Quelles sont les conclusions importantes ?
4. Comment décririez-vous les idées de l'étape 3 sans recourir à un vocabulaire visuel ?
5. Pouvez-vous imaginer comment les données qui vous intéressent dans vos activités quotidiennes pourraient bénéficier du même traitement ? Quelles situations pourriez-vous explorer dans un tel scénario ? Comment cela aiderait-il votre équipe à mieux comprendre le système à l'étude ?

Projet guidé III

Sélectionnez un ensemble de données dans la liste ci-dessous (ou tout autre ensemble) :

- [GlobalCitiesPBI.csv](#)
- [2016collisionsfinal.csv](#)
- [sondages_us_election_2016.csv](#)
- [HR_2016_Census_simple.xlsx](#)
- [custdata.tsv](#)

Pour votre/vos ensemble(s) de données :

1. Créez un dictionnaire de données. Établissez une liste de variables qui, selon vous, sont cruciales pour une bonne compréhension de l'ensemble de données. Justifiez vos choix.

3. Créez (au moins) 5 visualisations bivariées/ univariées qui peuvent vous aider à comprendre l'ensemble de données.
4. Produisez (au moins) 3 visualisations "définitives" pour l'ensemble de données. Utilisez les principes discutés en classe (notamment la documentation, les légendes, les annotations, etc.) L'accent doit être mis sur le contenu ET sur la présentation (suggestions : envisagez de créer un nombre raisonnablement élevé de graphiques en utilisant une sélection aléatoire d'un nombre aléatoire de variables afin de minimiser les chances de passer à côté d'informations utiles). Utilisez les concepts de narration lorsque cela est approprié.