

# Data-journalisme

MMI 2 – TP#1 S4



Danielo **JEAN-LOUIS**

# But du cours

- Découvrir les enjeux du journalisme du futur → data-journalisme
- Être capable d'utiliser les connaissances vues dans le cours de big data dans la rédaction d'articles
- Raconter des histoires avec des infographies et des cartes (interactives)

**Qu'est-ce que le data-journalisme ?**

- Évolution du journalisme "classique"
  - Appelé aussi journalisme de données
- "Nouvelle" discipline
- Émergence liée à la production et récolte intensive de données

- Domaine pluridisciplinaire regroupant :
  - Journalistes
  - Développeurs
  - Statisticiens
  - Graphistes
  - Experts thématiques
  - **La donnée**
- Nécessite des connaissances en programmation

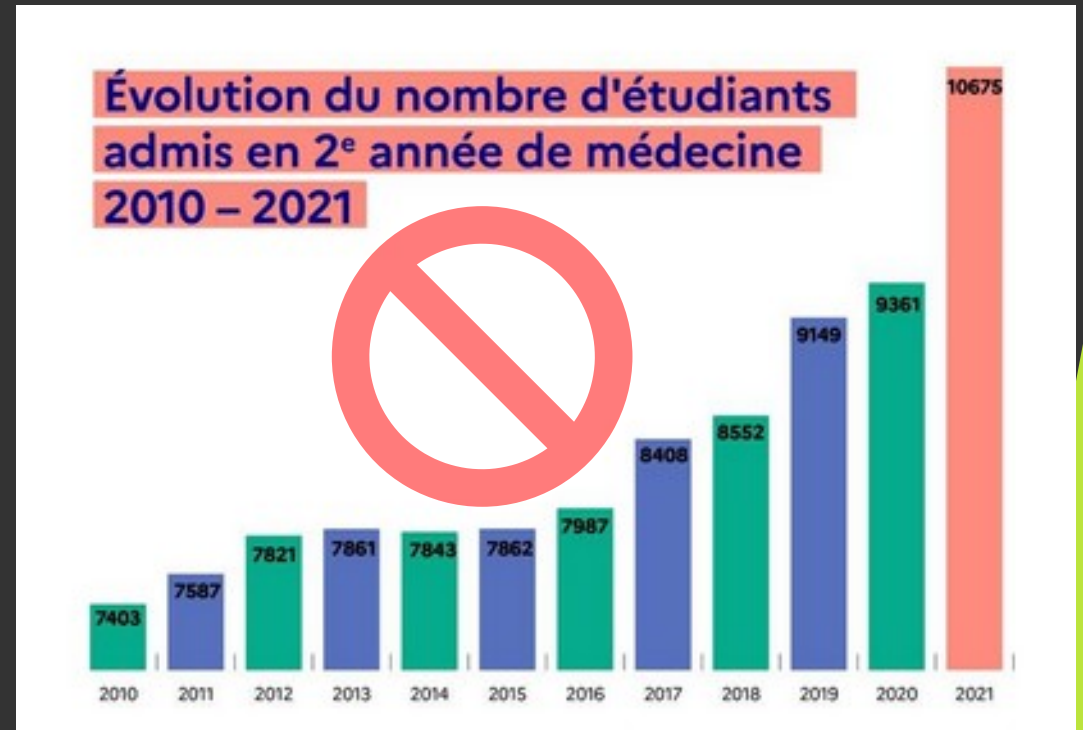
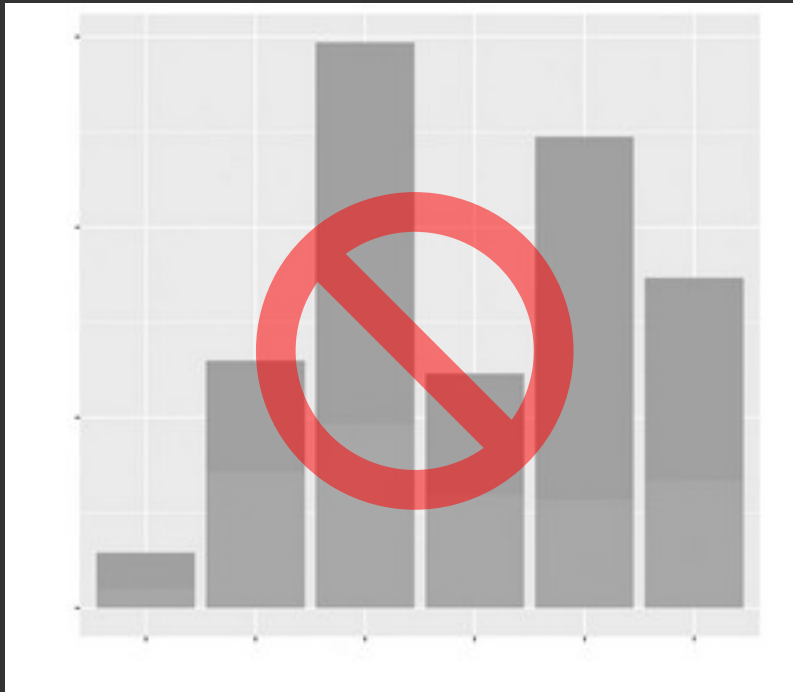
- Rédaction d'articles de presse à partir de données
  - Faiblement ou fortement structurées
  - Agrémentés de graphiques ou autres cartes (interactives ou non)
- Ces articles engagent plus le lecteur
  - Plus de profits pour l'organisme de presse

- Ces infographies servent les articles
  - **Elles ne sont pas une finalité**
  - Elles doivent être accompagnées d'analyses et d'explications
  - Elles doivent être claires et compréhensibles
    - Ethique journalistique



# Graphiques – Ethique journalistique

- Les graphiques sont souvent des outils vecteurs de mauvaise informations → fake news



**Le data-journalisme ce n'est pas  
qu'afficher des infographies**

**Un journalisme qui utilise la technologie  
pour accéder, analyser et découvrir des  
histoires via la donnée, donnée qui servira à  
raconter des histoires**

**Où trouve-t-on les données ?**

- Données ouvertes
  - insee.fr, data.gouv.fr, kaggle.com, données de gouvernements...
    - Florida man
- Fuite de données
  - Pandora paper, LuxLeaks...
- API
- Collecte manuelle...

Sources :

- <https://www.miamiherald.com/news/local/news-story/7608595>
- <https://www.kmcelwee.com/florida/>

# Collecte manuelle - Web-scrapping

- Technique d'extraction du contenu de sites web (tous les sites)
  - Récolte de données non conventionnelle
- Nécessite des connaissances en programmation
  - Connaissance du DOM
  - Python / Javascript et d'outils dédiés

# Web-scrapping – Outils

- Python
  - Selenium
    - Puissant fait fonctionner un navigateur en mode "headless"
  - BeautifulSoup
    - Léger mais limité, ne permet pas de lire les éléments HTMLS asynchrones

# Web-scrapping – Outils

- Javascript – Nécessite Nodejs
  - Selenium
  - puppeteer



# Web-scrapping – Précautions

- Scrappez de façon éthique
  - Eviter de le faire de façon intensive
    - Instaurez un délai entre chaque "scrap"
    - Limite les risques de se faire IP ban
- Scrappez de façon maligne
  - Changez de DNS entre chaque "scrap"
    - Limite les risques de se faire IP ban

# Les outils du data-journaliste

# Outils

- La donnée
  - open data, webscrapping, leaks...
- Les graphiques
  - datawrapper, seaborn, D3...
- Tableur
  - pandas, libre office calc, Excel...

## Sources :

- <https://datajournalism.tools/>

**On cherche toujours la vérité**

**La vérité doit toujours primer sur les  
sentiments**

*"Notre métier n'est pas de faire plaisir,  
non plus de faire du tort, il est de porter  
la plume dans la plaie."*

– Albert Londres, journaliste français

# Pratiquons ! - Graphiques

Pré-requis :

- Avoir la ressource `ressources/graphiques.jpg`
- A télécharger ici :  
<https://downgit.github.io/#/home?url=https://github.com/DanYellow/cours/tree/main/big-data-s4/travaux-pratiques/numero-7>

