

# 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)

## But du cours

- Traiter un sujet d'information complexe en récoltant les données et en les restituant dans un format clair et attractif

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

## Sources :

- <https://www.youtube.com/watch?v=9iGVzYIBelw>
- <https://www.youtube.com/watch?v=hJAR7UeMG8c>

- La donnée est au centre de l'investigation du data-journaliste
  - C'est sa **source principale**
- Évolution du journalisme "classique"
  - Le journaliste va se doter de nouvelles capacités :
    - programmation / graphisme / statistiques...

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

# "Nouvelle" discipline

- Color of money – 1988
  - Bill Dedman écrit un article avec des graphiques montrant que les banques dans la ville d'Atlanta discriminent les Noirs
    - Bill Dedman obtint le prix Pulitzer pour cet article

Source(s) :

- [http://powerreporting.com/color/color\\_of\\_money.pdf](http://powerreporting.com/color/color_of_money.pdf)



# "Nouvelle" discipline

- Spotlight – 2000
  - Le Boston Globe dévoile une affaire de pédophilie dans l'Église Catholique Romaine de Boston
    - Prix Pulitzer 2003
    - Film en 2015 oscarisé
    - Enquête à partir de courriers internes

## Source(s) :

- <https://www.franceculture.fr/emissions/mecaniques-du-journalisme-saison-7-laffaire-dupont-de-ligennes-lenquete-sans-fin/spotlight-des-pedophiles-dans-leglise-americaine-episode-1>

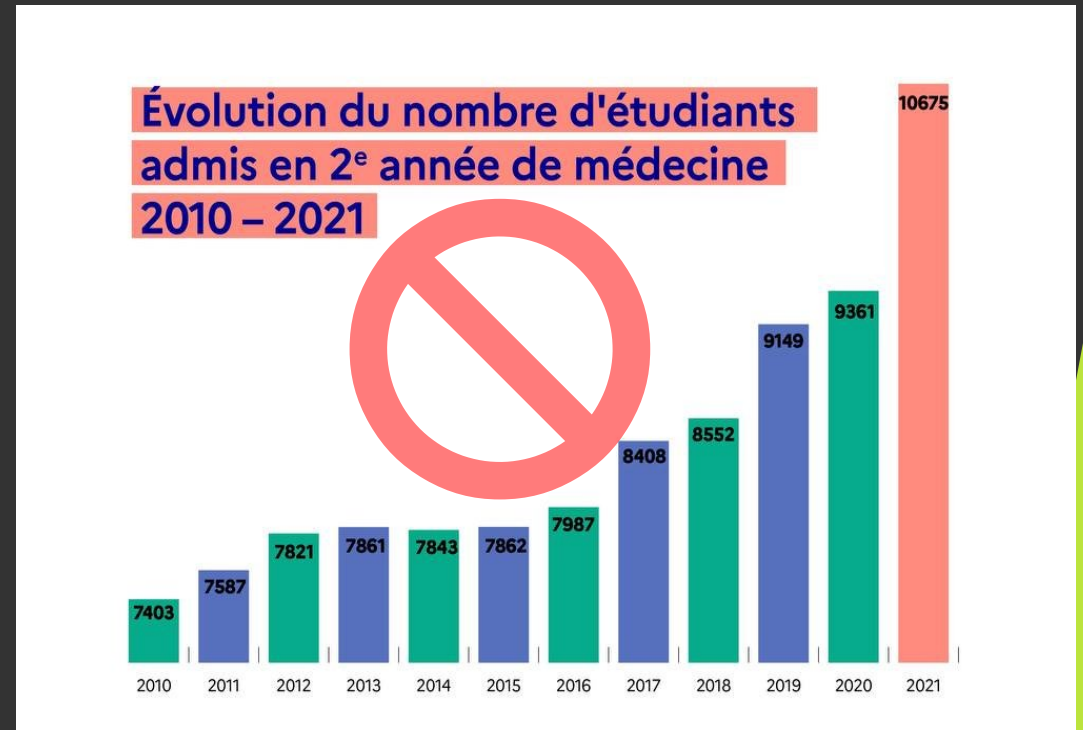
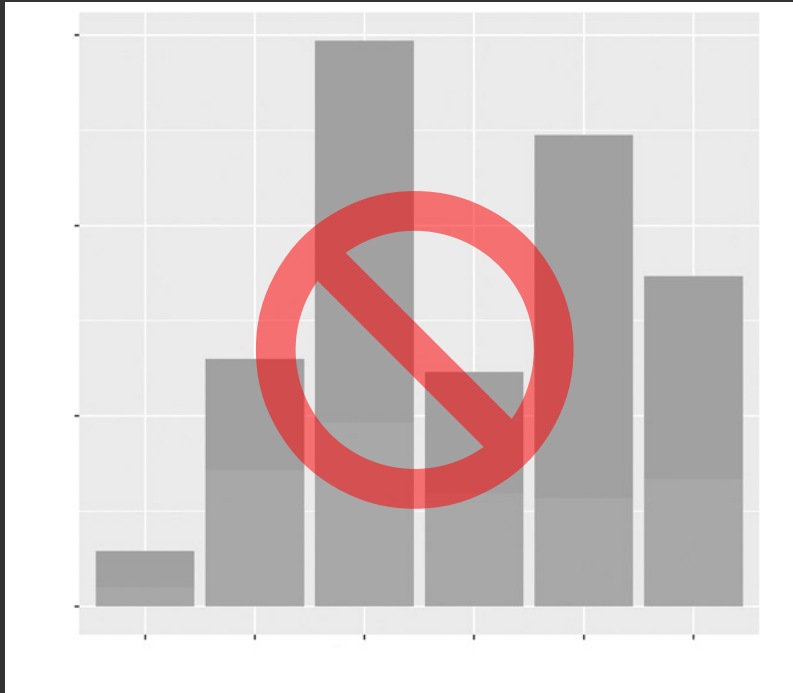
- Domaine pluridisciplinaire regroupant :
  - Journalistes
  - Développeurs
  - Statisticiens
  - Graphistes
  - Experts thématiques
  - **La donnée**

- 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

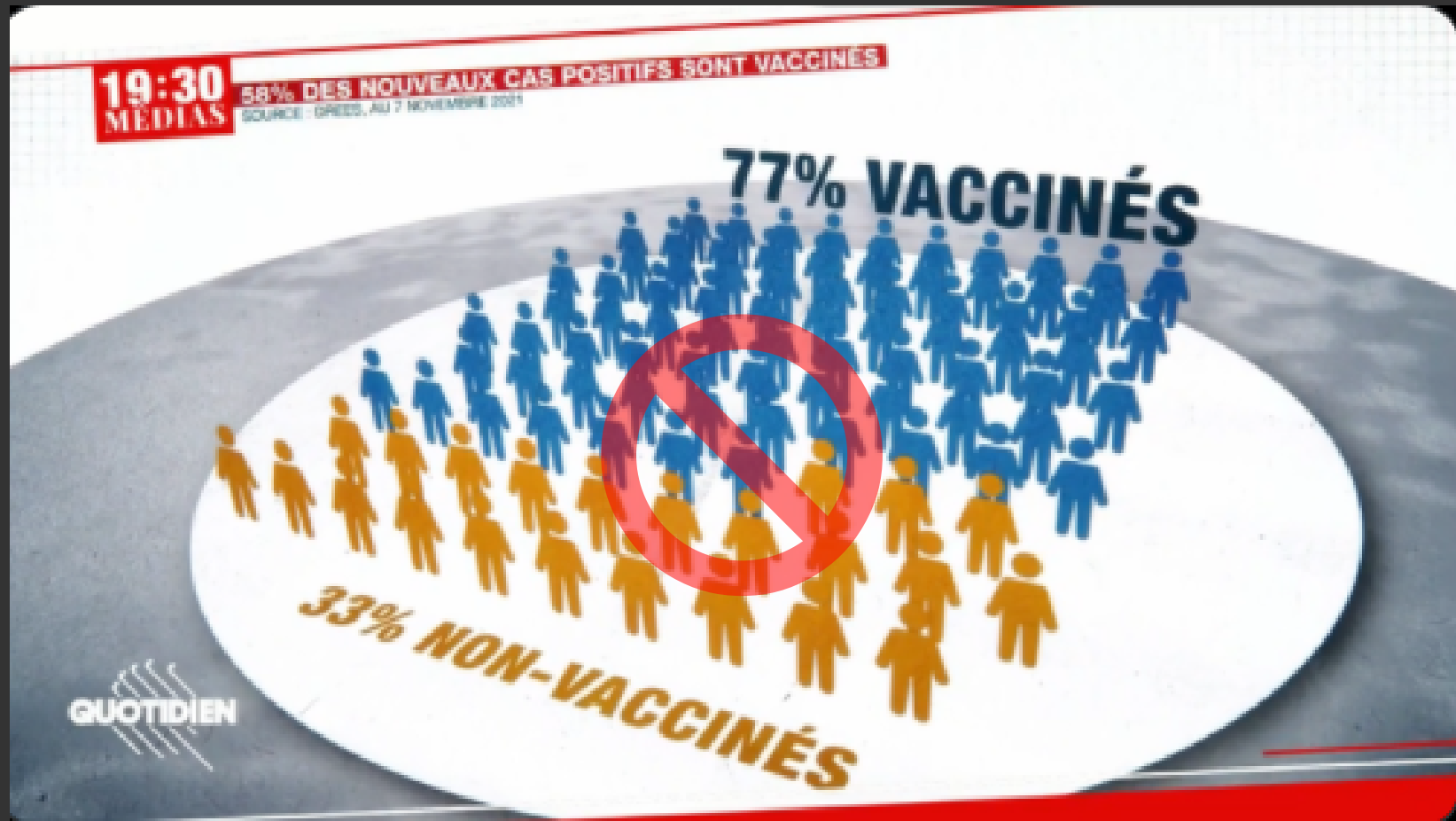
- Ces infographies servent les articles
  - **Elles ne sont pas une finalité**
  - Elles doivent être accompagnées d'analyses et d'explications
  - Elles doivent être claires et compréhensibles
    - Éthique journalistique

# Graphiques – Éthique journalistique

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



# Graphiques – Éthique journalistique



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

**Le data-journalisme c'est l'utilisation de la technologie pour accéder, analyser et découvrir des histoires via la donnée, donnée qui servira à raconter des histoires**



**Autrement dit, les données servent votre  
histoire, et non l'inverse.**

Source(s) :

- <https://blog.mondediplo.net/2012-08-23-L-infographie-dans-la-production-du-savoir>

Le journaliste interroge des personnes, le data-journaliste, lui, interroge des données.

Le data-journaliste **doit garder les méthodes d'investigation** du journaliste "classique" :

- Hiérarchisation
- Vérification des sources
- Tri des données

Trouver, récolter les données est la  
**première étape du travail** d'un data-  
journaliste

**Où trouve-t-on les données ?**

- Données ouvertes
  - insee.fr, data.gouv.fr, kaggle.com, données de gouvernements...
    - Florida man, exemple de données ouvertes
- Fuite de données
  - Pandora papers, LuxLeaks, Wikileaks...
- API / Bases données
- Moteurs de recherche
- ...

Sources :

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

# Moteurs de recherche

- Outils très puissants permettant de chercher à travers le web
  - Généraux : Google, ddg...
  - Spécialisés : linkedin, twitter...
- Permettent de faire des recherches plus poussées grâce à des opérateurs

# Moteurs de recherche web - Opérateurs

- Sur un site spécifique : "site:"
  - Ex: "site:https://cyiut.cyu.fr/"
- Version cache : "cache:"
  - Ex: "cache:https://cyiut.cyu.fr/"
- Expression exacte ""
  - Ex: "université de Cergy"
- ...

## Source(s) :

- <https://moz.com/learn/seo/search-operators> - anglais
- <https://www.journaldunet.com/solutions/seo-referencement/1194524-les-operateurs-de-recherche-google-la-recherche-avancee-pour-une-meilleure-analyse-seo/>

# Moteurs de recherche web - Opérateurs

- Après une certaine date : "after:"
  - Ex: "after:2021-01-01" – Format: YYYY-MM-DD
- Contenu de la balise <title> : "allintitle:"
  - Ex: "allintitle:MMI"
- Expression exacte ""
  - Ex: "université de Cergy"
- ...

## Source(s) :

- <https://moz.com/learn/seo/search-operators> - anglais
- <https://www.journaldunet.com/solutions/seo-referencement/1194524-les-operateurs-de-recherche-google-la-recherche-avancee-pour-une-meilleure-analyse-seo/>



# Moteurs de recherche

- <https://www.scinapse.io/>
  - Moteur de recherche académique
- <https://archive.org/>
  - Moteur de recherche des sites / pages du passé du web
- <https://osintframework.com/>
  - Liste de moteurs de recherche par thème
- <https://datasetsearch.research.google.com/>
  - Moteur Google dédié aux jeux de données
- <https://www.reddit.com/r/datasets/>
  - Subreddit dédié aux datasets

# Les outils du data-journaliste

# Outils

- La donnée : C'est sa source **principale** de travail
  - open data, leaks, API...
- Les graphiques / infographies
  - datawrapper.de, seaborn, D3, Illustrator...
- Tableur / Explorateur de données
  - pandas, libre office calc, Excel...
- Les moteurs de recherche

## Source(s) :

- <https://datajournalism.tools/> - anglais
- <http://infolabs.io/ressources>

**On cherche toujours la vérité**

**La vérité doit toujours primer sur vos  
sentiments/intérêts**

*"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 (1884 - 1932), journaliste français

