

Exploration des savoirs

L'enquête, ses données

Émilien Schultz - Nicolas Benvegna

Objectifs

- ▶ Lien enquête/données
- ▶ Réfléchir aux données possibles
- ▶ Discuter bonnes pratiques

La démarche d'enquête

1. Discussion 360°
2. Phase exploratoire / contexte de découverte
3. Consolidation de la question principale
4. Calendrier & objectifs & données pour l'argumentation
5. Démarche de collecte et de preuve
6. Analyse
7. Restitution

Parlons données

Les données dans l'enquête

Connaître la réalité quel accès ?

- ▶ **Sources**

- ▶ Résultats d'autres enquêtes
- ▶ Documenter directement les phénomènes
 - ▶ le langage des **données**

Mise à l'épreuve des affirmations

The (real) scientific method.



C'est quoi une donnée ?

Définition générique : *un ensemble d'informations issues par différentes transformations de traces laissées par un phénomène*¹

- ▶ définition a-disciplinaire
- ▶ importance du lien avec un **phénomène**
 - ▶ *la donnée n'est pas le phénomène*
- ▶ des étapes intermédiaires
- ▶ nature de la "trace" recueillie : comment ? quoi ?

¹Wikipédia : "Data, as a general concept, refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing."

La place des données dans la démarche

- ▶ *Theory-driven*
 - ▶ Les données arrivent après la formulation d'une hypothèse
- ▶ *Data-driven*
 - ▶ Les réponses émergent des données (notamment, le big data)

En fait

- ▶ Aucune donnée ne peut répondre à toutes les questions
- ▶ Aucune donnée ne répond en soit à une question

L'enquête : un mouvement plus incrémental

Deux principaux usages

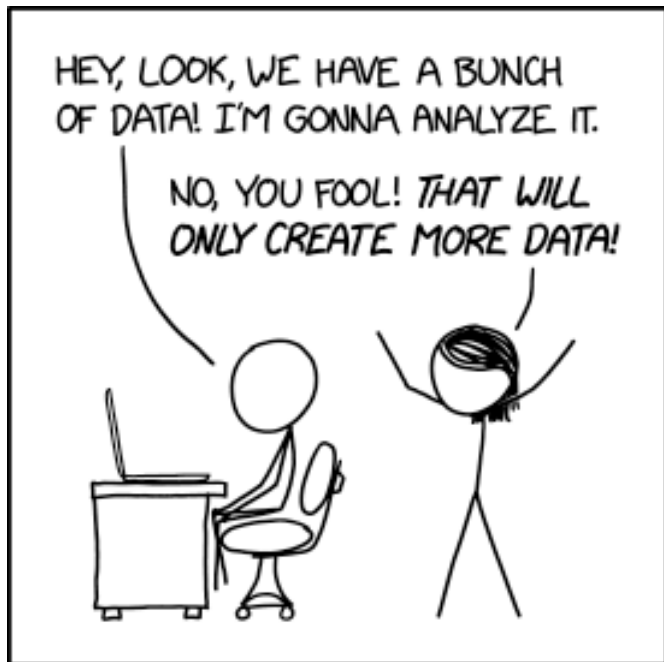
Les données interviennent à différents moments dans l'enquête

- ▶ Exploratoire/découverte : diversifier son rapport au monde
- ▶ Argumentation/justification : défendre un argument
- ▶ (Illustration)

Des données partout



Et sans fin



Mais des spécialités “régionales” de données

- ▶ Chaque domaine/secteur a ses données
 - ▶ Histoire/Physique/Biologie
 - ▶ Procès/Journalistes/...
- ▶ Pour l'enquête interdisciplinaire
 - ▶ Dépasser les habitudes
 - ▶ Relier des pratiques différentes
 - ▶ Se nourrir de la diversité

Importance d'un vocabulaire commun

Problématique fil rouge

i Cohabitation des mobilités en ville

Constat de transformations en cours et de conflits entre les différentes mobilités dans les zones urbaines.

- ▶ Quels sont les transformations en cours ?
- ▶ Quelle place pour les nouvelles mobilités ?
- ▶ Comment la cohabitation s'organise avec la croissance de l'usage du vélo ?

Une question interdisciplinaire et controversée

Enquêter pour aller au-delà des discours



La place des données

Le mythe du processus linéaire

DATA



SORTED



ARRANGED



PRESENTED
VISUALLY



EXPLAINED
WITH A STORY



Un processus itératif

- ▶ Le lien entre question & données n'est pas immédiat
- ▶ Processus itératif :
 - ▶ Certaines données amènent à des questions
 - ▶ Certaines questions amènent à chercher des données
- ▶ Mais
 - ▶ Certaines données n'existent pas

Un appariement pragmatique et toujours partiel ; passe souvent par expliciter le protocole

Quelles sont les données possibles ?

Pour enquêter sur les cohabitations des mobilités autour de l'université et la place du vélo.

Types de données

- ▶ Mon expérience de cycliste
- ▶ ...

Est-ce que toutes ces données se valent ?

- ▶ Format : brut ou transformé (étapes intermédiaires)
- ▶ Nature différentes (valeur, texte, image, ...)
- ▶ Disponibilité (déjà collectées, à collecter)
- ▶ Taille
- ▶ Qualité

Face à cette diversité

- ▶ Comment choisir les données les plus pertinentes ?
- ▶ Comment les manipuler et les stocker ?
- ▶ Comment passer des données à des résultats ?

Enjeu : prendre au sérieux les données pour des enquêtes plus riches et des affirmations plus solides.

Corpus : construction & représentativité

Définition : un ensemble de données rassemblée suivant une logique

Pourquoi ?

- ▶ Décrire la variation d'un phénomène
- ▶ Caractériser une population
- ▶ Réduire l'incertitude

i Pratiques du vélo

Une donnée : un entretien avec un cycliste

Un corpus : des entretiens avec tous les cyclistes qui ont accepté de répondre à mes questions un samedi soir rue Rivoli

Enjeu de la représentativité

A-t-on suffisamment de données ? Est-ce que l'on peut vraiment se dire que l'on peut caractériser l'entièreté du phénomène ?

Comment faire ? Entre méthodes & bricolage. Important de justifier/mettre en discussion.

Produire vs. réutiliser des données

- ▶ Données **primaire**
 - ▶ construire ses données/corpus
 - ▶ un gradient allant de la production complète à la mise en forme
- ▶ Données **secondaire**
 - ▶ données déjà constituées
 - ▶ multiplication des données disponibles open data
- ▶ Entre les deux : **l'explosion des traces numériques**
 - ▶ accès à des bases de données/API

Conséquences de la disponibilité

- ▶ Données primaires : possible de les adapter à sa question mais vulnérabilité du protocole
- ▶ Données secondaires : répondent souvent à une autre question que la votre

i Vélo & Justice

- ▶ Les données de comptage des vélos à Paris produits par la ville
- ▶ Compter moi-même les vélos dans une rue

3. La forme compte : quels formats ?

- ▶ Données “désagrégées” permettant les recombinaisons:
 - ▶ structurées
 - ▶ tabulaires
 - ▶ peu structurée
 - ▶ un mélange de formats
 - ▶ pas structuré
 - ▶ un contenu dont il faut extraire l'information
- ▶ Données déjà agrégées
 - ▶ Tableaux déjà calculés
 - ▶ Graphiques

Une pragmatique des données

Des arbitrages

- ▶ Balance coût / avantage de l'accès à des données
- ▶ Adéquation avec les objets de la question
- ▶ Adéquation avec la nature de la question
- ▶ Facilité de retransformer les données
- ▶ Disponibilités des méthodes d'analyse pour obtenir un résultat

Où mettre son énergie ?

Difficulté d'accès aux données

- ▶ Le coût des données :
 - ▶ ces données peuvent déjà exister
 - ▶ peuvent être collectées avec un effort raisonnable
 - ▶ sont coûteuses à récolter

i Des choix à faire

- ▶ Réaliser ma propre enquête par questionnaire, avec mes questions spécifiques
- ▶ Utiliser les données publiques disponibles à l'échelle des villes

Volume des données

- ▶ Le volume des données n'est pas l'élément le plus important
- ▶ Adéquation avec la question

Ne pas hésiter à restreindre la taille et privilégier l'adéquation

Qualité des données

Toutes les données ne se valent pas : une qualité multidimensionnelle.

- ▶ Tracabilité de la chaîne de production
- ▶ Fiabilités (producteurs douteux, ou au contraire légitime)
- ▶ Données manquantes (est-ce qu'il manque des éléments)
- ▶ Erreurs dans le corpus
- ▶ Ouverture versus fermeture de l'accès
- ▶ ...

L'analyse dans tout ça ?

- ▶ La valeur des données vaut par la capacité d'en faire découler des résultats (actionner les données)
 - ▶ Interdépendance collecte/données/résultats (par ex. l'expérimentation)
- ▶ Certaines analyses sont complexes et posent la question de la faisabilité (ex. statistiques)
- ▶ Des outils spécifiques...

Par quelles données débiter

Les données “accessibles”

- ▶ Toute enquête se pose la question:
 - ▶ De l'espace public
 - ▶ De la production des connaissances
- ▶ Avec quelques manières de les utiliser
 - ▶ Ce point sera traité dans une séance ultérieure plus en détail

Il n'y a pas vraiment de limites, n'hésitez pas à être créatifs

Documenter l'espace public : la presse

L'article de presse comme contenu complexe :

- ▶ date (évolution temporelle)
- ▶ contenu textuel (sujets)
- ▶ cadre : journal, journaliste, ...

Permet une diversité d'approches :

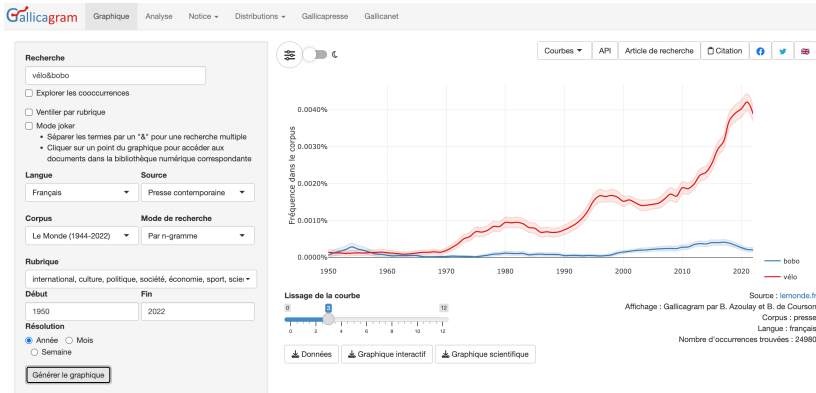
- ▶ l'article individuel comme source
- ▶ corpus d'articles sur un sujet/période
 - ▶ thèmes abordés
 - ▶ acteurs récurrents
 - ▶ évolutions temporelles

Accès captifs

Une presse sous droit d'auteur.

- ▶ Par des outils dédiés comme Gallicagram
- ▶ Par l'interface de portails comme Europresse
- ▶ Certains médias sont libres

Cadrage d'une tendance



Construire et extraire un corpus d'Europresse

Accès Europresse à des données payantes

Vélo & Justice

Comment aborde-t-on la question des inégalités quand on parle du vélo en région parisienne ?

Le contour du corpus :

- ▶ Définir la période
- ▶ Définir les mots-clés
- ▶ Définir les journaux concernés
 - ▶ Se concentrer sur un journal
 - ▶ Sur un bouquet?

Consulter vs. conserver

Possibilité de consulter sur Europresse. Mais comment conserver ?

- ▶ Collecter à la main vs. télécharger
- ▶ Des transformations nécessaires

Puis aller vers l'analyse :

- ▶ Descriptive : compter
- ▶ Plus avancée
- ▶ Visualisations



Statut des données

Ce sont des données sous droit d'auteur. Leur usage est très réglementé. Il y a une tolérance pour la recherche mais il est interdit de diffuser.

Recherche simple | Recherche avancée | Recherche express | Recherche de biographies

TEXT= vélo & (paris | parisien*)

Dans toutes les archives

Monde, Le



Presse

Télévision et radio

Médias sociaux

Études et rapports

Répertoires et références



50 sur 2 913



Pertinence



Le Monde

L'empêcheur de pédaler en rond

2023-03-29 - 1919 mots

Pascal Krémer - Regardez ! » Bicyclette stoppée net, Stein van Oosteren observe ce qui n'attire le regard de personne. Les tours de roue hésitants d'un petit, entre mère anxieuse et chaland du centre ...

Aussi paru dans

Le Monde

Vacances à vélo, la vérité à cru

2022-07-11 - 2071 mots

Guillemette Faure - Et si on essayait le «slow tourisme»? Voilà trois jours que je suis partie en vacances à vélo. Mon intention, avec ma fille et les amis qui nous rejoindront en route ...

Aussi paru dans

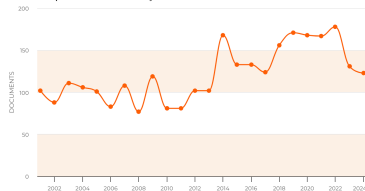
Le Monde

A Paris, la fin des trottinettes de location

0% - 0 documents
Répertoires et références
0% - 0 documentsPositif
70% - 1 337 documents
Neutre
7% - 14 documents
Négatif
23% - 566 documents

ÉVOLUTION

Pic médiatique : 11 documents le 27 juin 2016





Trends in local newspaper reporting of London cyclist fatalities 1992–2012: the role of the media in shaping the systems dynamics of cycling



Alex Macmillan^{a,*}, Alex Roberts^b, James Woodcock^c, Rachel Aldred^d, Anna Goodman^e

^a Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand

^b College of Medical and Dental Sciences: University of Birmingham, Birmingham, UK

^c UK CRC Centre for Diet and Activity Research (CEDAR), MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, University of Cambridge, Cambridge, UK

^d Department of Planning and Transport, Faculty of Architecture and the Built Environment, University of Westminster, London, UK

^e Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

ARTICLE INFO

Article history:

Received 20 August 2014

Received in revised form 27 May 2015

Accepted 18 October 2015

Available online 10 November 2015

Keywords:

Cycling

Fatality

Injury

Media

Systems dynamics

Trends

ABSTRACT

Background: Successfully increasing cycling across a broad range of the population would confer important health benefits, but many potential cyclists are deterred by fears about traffic danger. Media coverage of road traffic crashes may reinforce this perception. As part of a wider effort to model the system dynamics of urban cycling, in this paper we examined how media coverage of cyclist fatalities in London changed across a period when the prevalence of cycling doubled. We compared this with changes in the coverage of motorcyclist fatalities as a control group.

Methods: Police records of traffic crashes (STATS19) were used to identify all cyclist and motorcyclist fatalities in London between 1992 and 2012. We searched electronic archives of London's largest local newspaper to identify relevant articles (January 1992–April 2014), and sought to identify which police-reported fatalities received any media coverage. We repeated this in three smaller English cities.

Results: Across the period when cycling trips doubled in London, the proportion of fatalities covered in the local media increased from 6% in 1992–1994 to 75% in 2010–2012. By contrast, the coverage of motorcyclist fatalities remained low (4% in 1992–1994 versus 5% in 2010–2012; $p=0.007$ for interaction between mode and time period). Comparisons with other English cities suggested that the changes observed in

Mettre en forme les données

- ▶ Délimiter le corpus
- ▶ Téléchargement sous un format HTML
 - ▶ Gérer la limite
- ▶ Des fichiers HTML à dater + un document de métadonnées
- ▶ Transformer en un document tabulaire

Étendre le domaine de la presse

Europresse n'est pas le seul portail sur lequel collecter des données.

Suivant la question :

- ▶ se concentrer sur les articles parus sur un site spécifique spécialisé
- ▶ récupérer les newsletters d'une organisation

Statut des données

Récupérer les évènements de la FUB pour voir lesquels mentionnent la question de la justice :
<https://www.fub.fr/fub/actualites>

Parenthèse : la difficile question des médias sociaux

- ▶ Une source riche d'échanges
- ▶ Mais de plus en plus fermés (Twitter, Insta, etc.)
 - ▶ Difficulté de collecter massivement
 - ▶ Possibilités de constituer des petits corpus

Das, Subasish, et al. "Extracting patterns from Twitter to promote biking." IATSS research 43.1 (2019): 51-59.

L'état de la science & de la connaissances: les métadonnées des publications scientifiques

- ▶ La cumulativité des connaissances dans la recherche scientifique passe largement par les publications scientifiques
- ▶ L'étude de ces publications scientifiques permet :
 - ▶ De suivre la dynamique des connaissances
 - ▶ D'identifier les acteurs clés du monde scientifique
 - ▶ De reconstituer les communautés
- ▶ Depuis plus de 50 ans, des bases de méta-données
- ▶ Auxquelles se rajoutent des bases de données ouvertes

Des données multifactorielles

- ▶ Date
 - ▶ Evolutions
- ▶ Affiliations
 - ▶ Institutions/pays
- ▶ Mots-clés & résumés
 - ▶ Domaine de recherche
- ▶ Auteurs
 - ▶ Co-autorat / communautés
- ▶ Citation
 - ▶ Réputation
- ▶ Financements...

Le portail Scopus/Openalex

OpenAlex

Scopus par la bibliothèque de sciencespo

i Études sur les mobilités

Quels sont les principaux pays producteur d'une littérature sur les mobilités en vélo ?

Quelle place occupe le thème de la justice/inégalité, et quels sont les angles principaux traités dans la littérature ?

Encore une fois, la clé est le corpus

Comment délimiter un corpus pertinent :

- ▶ Approche biographique
 - ▶ Se concentrer sur un auteur
- ▶ Approche thématique
 - ▶ Identifier un thème
- ▶ Approche institutionnelle

Ne pas hésiter à viser des petits corpus & à lire quelques articles

Récupérer les données

Une fois le corpus délimité, enjeu de récupérer les données

- ▶ Select > Export
- ▶ Format CSV ou Plain text

Permet l'analyse ensuite

Autres sources

- ▶ Web Of Science
- ▶ Sources ouvertes
 - ▶ HAL
 - ▶ ArXiv
- ▶ ...

Documenter les répartitions nationales

L'INSEE rend disponible de nombreuses données

- ▶ Notamment Catalogue des données de l'INSEE
- ▶ Agrégées et à différents niveaux spatial (ville, IRIS)

(ça vaut le coup de creuser un peu)


Vélo & Justice

Mobilités professionnelles en 2020 : déplacements domicile - lieu de travail Recensement de la population - Base flux de mobilité

- ▶ voir la structure des données
- ▶ jeter un coup d'oeil sur les données

Que faire avec ces données

- ▶ Distributions nationales :
 - ▶ répartitions géographiques
 - ▶ recalculer des statistiques pour des échnatillons
 - ▶ faire des cartes
- ▶ Recours à des logiciels statistiques
 - ▶ Excel, Jamovi
 - ▶ R, Python

 Avant de faire de l'analyse

Prendre le temps de regarder les données et réfléchir en amont sur les traitements envisagés

Documenter les attitudes & pratiques : les sondages

Que pensent les individus ? Que disent-ils de leurs pratiques ?

- ▶ Les sondages sont devenus une source importante de connaissance des comportements sociaux
- ▶ Des producteurs très différents (Chercheurs, associations, ...)
- ▶ De plus en plus de données disponibles
 - ▶ Recherche : <https://data.progedo.fr/>
 - ▶ Autres acteurs : par exemple <https://opendata.parlons-velo.fr/>

Le contenu de ces données

- ▶ Données tabulaires : questions en colonnes, réponses en ligne
- ▶ Fichiers csv, spss, autre...
- ▶ Permet une diversité d'analyses:
 - ▶ Distribution
 - ▶ Croisement
 - ▶ Evolutions ...
- ▶ Limite de la formulation des questions

Entrepôts de données

https://data.gouv.fr/



data.gouv.fr

Se connecter S'enregistrer

Recherche



[Données](#) [Réutilisations](#) [Organisations](#) [Commencer sur data.gouv.fr](#) [Actualités](#) [Nous contacter](#)

[Accueil](#) > [Jeux de données](#)

Jeux de données

Rechercher parmi les 40005 jeux de données sur data.gouv.fr

[Rechercher dans les réutilisations](#)

vélo



Recherche

Filtres

Organisations

Toutes les organ...

Mots clés

Tous les mots-clés

Formats

Tous les formats

Licences

510 résultats

Trier par : **Pertinence**



Plan Vélo 2021-2026

[Ville de Paris](#) — Mis à jour le 3 mai 2022

Métadonnées : — 0 1 1

Jeu de données présentant le Plan Vélo 2021-2026 tel que présenté au conseil de Paris en octobre 2021. En savoir plus sur le Plan Vélo sur Paris.fr



Plan Vélo de la Métropole de Lyon

[Métropole de Lyon](#) — Mis à jour aujourd'hui

Métadonnées : — 0 0 0

Le Plan Vélo est un objet linéaire représentant un itinéraire du schéma directeur Plan Vélo 2009-2020, voté par le

S'adapter au corpus

- ▶ une structure/des contenus très variables
- ▶ une qualité variable
- ▶ ne jamais hésiter à explorer

i Vélo & Justice

Statistiques de subventions d'achat de vélo en Île-de-France

- ▶ Distribution par villes
- ▶ A rapporter aux données INSEE ?

Une diversité d'autres sources

- ▶ Google trends
- ▶ Banque mondiale
- ▶ OECD
- ▶ ...

Avoir une conception ouverte de ce qui peut constituer une donnée utile

Collecter vos propres données

Importance du protocole

- ▶ Identifier clairement le rôle des données dans votre question
 - ▶ Notamment les résultats que vous attendez
- ▶ Clarifier la population ciblée & les critères de représentativité
- ▶ Identifier les risques
 - ▶ Que se passe-t-il si vous n'arrivez pas à tout collecter ?

i Vélo & Justice

Je veux m'intéresser aux images du vélo dans les films. Comment je construis la population de films (uniquement ceux qui le mentionne dans le résumé ? un échantillon aléatoire ?) ? Comment ensuite je code pour chaque film (je me limite au résumé de manière automatique ? Je les regarde ?)

Importance du format de codage

- ▶ Quelle “trace” garder
 - ▶ Extensive : couteux
 - ▶ Condensé : perte d'information
- ▶ Construire des catégories en les testant
 - ▶ Ne jamais hésiter à faire des allers-retours

Données “quantitatives” et “qualitatives”

- ▶ Division un peu artificielle
- ▶ Passages possibles du qualitatif au quantitatif
 - ▶ Codage des entretiens
 - ▶ Analyse du texte
- ▶ Importance de la structure des jeux de données / leur meta-informations

Exemple des entretiens

Définir le dispositif de collecte

- ▶ Quel est le corpus (qui va être interrogé)
- ▶ Comment passer du discours à des données
 - ▶ Annotation : quoi ?
- ▶ Quels éléments doivent être présents dans l'entretien absoluments ?
- ▶ Que se passe-t-il s'il manque un entretien ?

Exemple des entretiens

Après : l'enregistrement/retranscription est une “trace” qu'il **faut mettre en forme**

- ▶ Mettre les entretiens dans un format unifié
 - ▶ .docx au mieux
- ▶ Nommer les fichiers de manière systématique
- ▶ Avoir un fichier avec les métadonnées (date, interlocuteur)
- ▶ Transformer les données :
 - ▶ Coder des éléments spécifiques des entretiens dans un fichier

Gestion matérielle des données

En pratique :

- ▶ Comment bien collecter et conserver les données ?
- ▶ Où les stocker ?
- ▶ Comment les décrire dans ses travaux ?
- ▶ Comment travailler collaborativement avec des données ?

Bonnes pratiques numériques

- ▶ Collecter et conserver les données brutes
- ▶ Documenter le contenu dans des **métadonnées**
- ▶ Documenter chacune des transformations
- ▶ Consolider un corpus nettoyé avant l'analyse

Dans l'idéal, aller vers la reproductibilité

Documenter documenter documenter

Journal collectif des traitements

UNFINISHED WORK

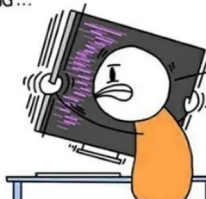
FRIDAY EVENING



PERFECT!
I'LL FINISH
THIS ON
MONDAY



MONDAY MORNING...



WHAT DOES
THIS MEAN!?!?



Format et stockage des données

- ▶ Simplifier le plus possible
 - ▶ .txt, .csv, ...
- ▶ Faire attention à la confidentialité
 - ▶ Anonymisation/pseudonymisation
- ▶ Noms clairs
- ▶ Dater les fichiers
- ▶ Système de dossier clair

Ranger ses données

- ▶ Conserver les données brutes
- ▶ Mettre ensuite en forme les données pour faciliter les analyses

country	year	cases	population
Afghanistan	1999	15	19997071
Afghanistan	2000	2666	200095360
Brazil	1999	30737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210766	1280008583

variables

country	year	cases	population
Afghanistan	1999	15	19997071
Afghanistan	2000	2666	200095360
Brazil	1999	30737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210766	1280008583

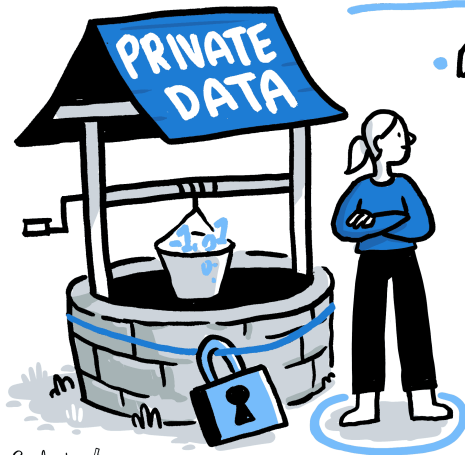
observations

country	year	cases	population
Afghanistan	1999	15	19997071
Afghanistan	2000	2666	200095360
Brazil	1999	30737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210766	1280008583

values

Sensibilité des données

- ▶ Limiter les données identifiantes
- ▶ Consentement et autorisation



PEOPLE DESERVE:

- DIGNITY
- AGENCY
- PRIVACY RIGHTS
- CONFIRMED CONSENT

Décrire les données dans ses travaux

Dans les travaux :

- ▶ Préciser la manière dont ont été constituées les données et.ou les sources
 - ▶ Pour chaque corpus
- ▶ Donner des informations descriptives sur le contenu
- ▶ Les transformations qui ont été faites
 - ▶ Ex. les filtrages
- ▶ Préciser les limitations s'il y en a
 - ▶ Données manquantes
 - ▶ Artefacts/problèmes

L'importance de décrire les données

In light of the significance of discursive construction in positioning environmental issues, I selected media presenting contrasting ideological views, since they shape public views towards key issues in different ways. *Le Figaro* and *Le Monde* with their respective rightist and leftist readerships, were selected for their wide spectrum of views (see Cohen 2012) and high circulation rates.

Articles were identified from the newspapers' online sites using the keywords *bicyclette* (bicycle) and *vélo* (bike) for the year 2017. This is indeed a pivotal year as it marks the implementation of Anne Hidalgo's 2015-2020 bicycle plan for Paris, because of the transition from JCDecaux's bike-sharing system to new operator Smovengo, the emergence of bike-sharing on the streets of French cities, and the boom in sales of electric bikes and subsidies (and their early withdrawal in September 2017). Only articles relevant to the study were selected. This means that articles dealing with sports events, cycling in France or bicycle touring were discarded, even though they may contribute to the view of cycling. The final corpus contains 68 articles from *Le Figaro* and 32 from *Le Monde*, totalling 100 articles and 56,868 words, as shown in Table 1. Although there are more articles in *Le Monde* as there are in *Le Figaro*, *Le Monde* articles are shorter on average vs. 474 words in *Le Figaro*, thus totalling 24,633 words for *Le Monde* and 32,235 words for *Le Figaro* dataset. The articles from the 2017 *Le Figaro* dataset

Aller vers l'analyse

- ▶ La valeur des données dépend des possibilités d'analyse
- ▶ Les analyses dépendent du type de données
- ▶ Trois approches :
 - ▶ “à la main” : manipulation (quasi) directe des données (Excel, codage manuel, ...)
 - ▶ “pipeline automatisé” : usage de logiciels spécialisés (Iramuteq, VosViewer, Jamovi, ...)
 - ▶ “programmation” : scripts de traitement de données (Python, R)

Les données : un ingrédient (central)

