



Sciences Po  
Bordeaux

# Données : Enjeux, collecte et analyse

S2 - Administration de la preuve (données quantitatives)

---

DECA 2 – 2023-2024

LÉO MIGNOT – INGÉNIEUR DE RECHERCHE – CENTRE ÉMILE DURKHEIM

# Éléments de contexte

## Multiplication des données

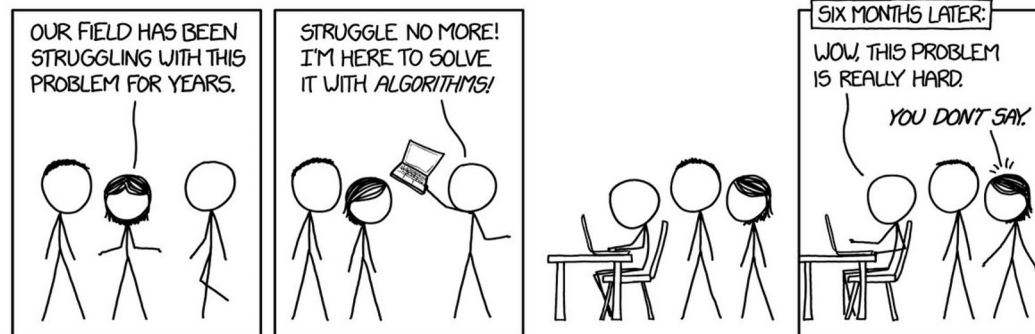
- données et traces numériques

## Progrès informatiques et puissance de calcul

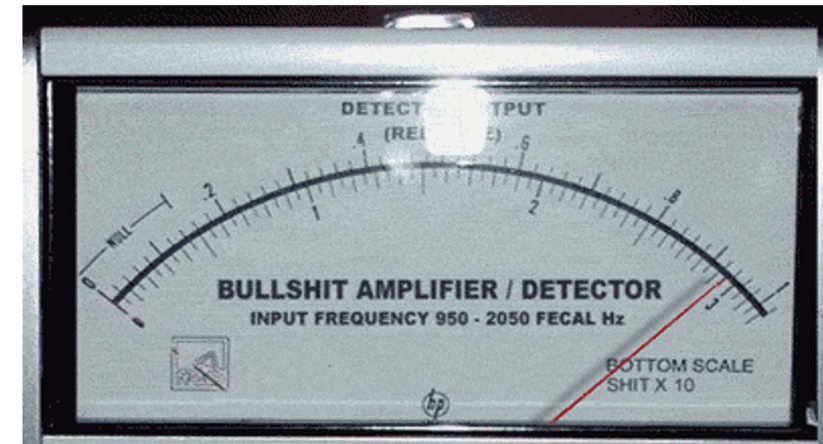
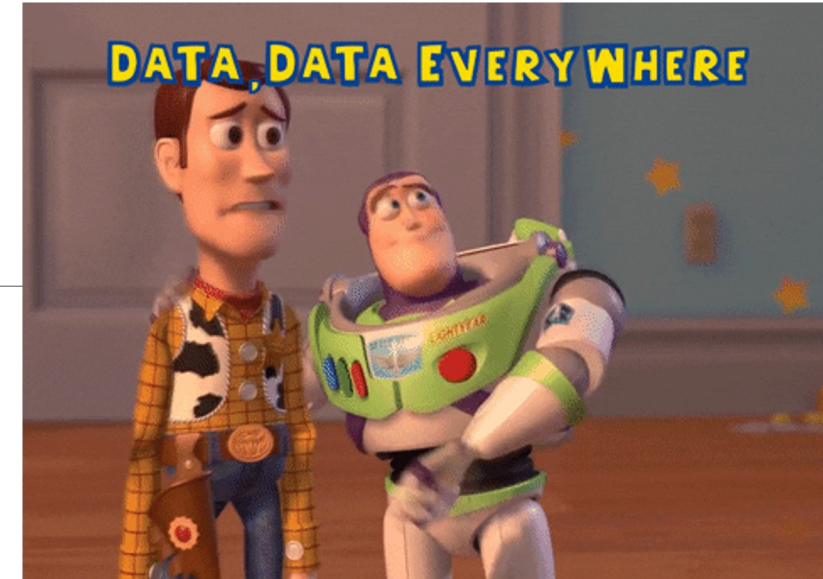
- nouvelles méthodes /outils

## Énorme "hype"

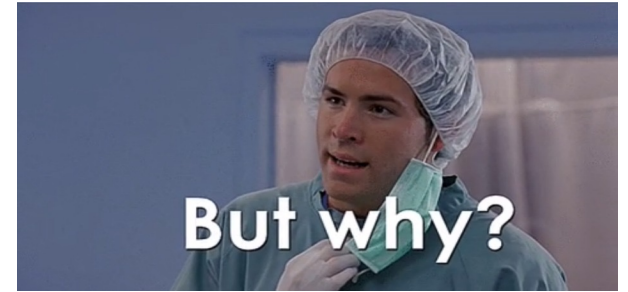
- Les data vont nous sauver...



[@xkcd](#)



# Pourquoi ?



## Être capable de naviguer dans un monde plein de données

- Familiariser à la notion de données
- Acquérir les compétences de lecture de tableaux et visualisation des données
- -> mieux comprendre les sujets, aide à la décision, etc.

## Plus long terme :

- effets évolutions numériques : renouvellement des objets d'étude et des outils
- multiplication des données pouvant faire l'objet d'investigations
- capacité des SHS à se saisir des nouveaux outils numériques et les mobiliser (évolution des outils statistiques, renouveau des méthodes quantitatives et qualitatives, etc.)

# Illustration : ma vie mon œuvre (?)

## Méthodes

stat, approche quanti+  
quali (des choses récentes,  
d'autres plus anciennes)  
ex : webscraping, ML, NLP,  
etc.



**Nouvelles  
données**  
ex : Twitter

#aventure #dynamique  
#c'estlematin

**Objet d'étude**  
IA en médecine  
(Radiologie surtout)



# Nota bene : je code comme une brouette



# Tour de table

WOOLAP

Aller sur le cours moodle (ça vous obligera à vous inscrire)

Dans l'onglet séances, aller sur le wooclap

Saisir 2-3 mots clés qui vous viennent à l'esprit quand on dit « quanti / stat ».

En discuter ensemble !



 [Copier le lien de participation](#)



1

Allez sur [wooclap.com](https://wooclap.com)

2

Entrez le code d'événement dans  
le bandeau supérieur

Code  
d'événement  
**GEQTXM**



1

Envoyez [@GEQTXM](#) au **06 44 60 96 62**

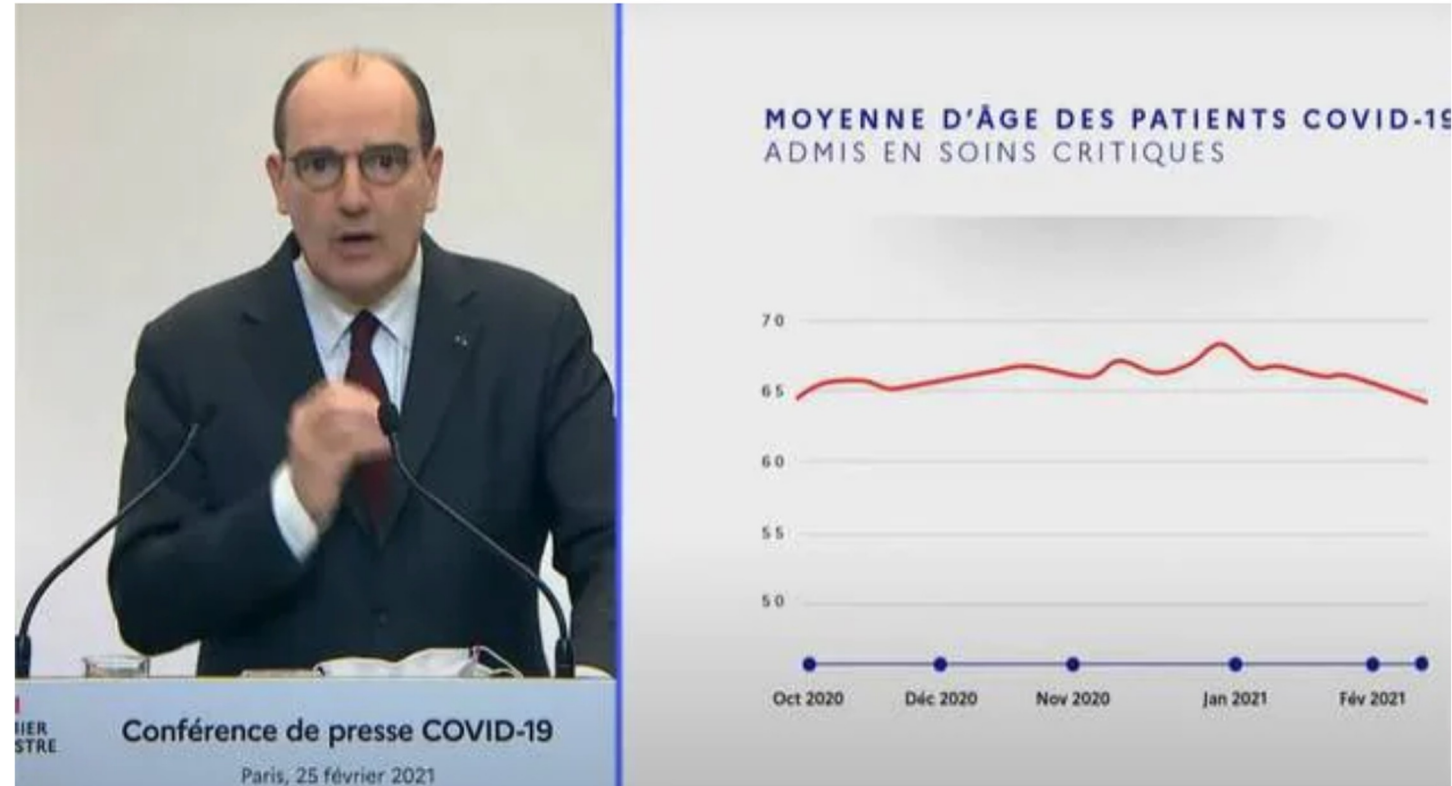
2

Envoyez votre réponse au même numéro

# J'ai du bon data



# Moins bien



Voir aussi @GraphCrimes

- <https://twitter.com/GraphCrimes>



# Séance 1. Séance introductive

---

# Objectifs de la séance 1

---

- Présentation du **semestre** (programme, thèmes, exercices & validation, le moodle)
- Découverte du logiciel **JAMOV**I (Installation, Ouvrir un jeu de données)
- “Jeux” de prise en main avec JAMOV
- Échanges pour **former les groupes** autour du ‘4 pages’
- Selon le temps : lecture d’un ‘4 pages’ pour comprendre l’exercice attendu

# Objectifs du cours

---

## **Cf. Être capable de naviguer dans un monde plein de données**

- Familiariser à la notion de données
- Acquérir compétences lecture de tableaux et visualisation des données
- -> mieux comprendre les sujets, aide à la décision, etc.

## **Se traduit par :**

1. Initier au vocabulaire et concepts de base de l'analyse de données (stat)
2. Comprendre leur place dans la démarche scientifique et argumentative
3. Construire et produire des analyses élémentaires (indicateurs, tableaux, graphiques, relations entre deux variables)
4. Développer la capacité d'analyse et d'interprétation des résultats

# Calendrier

Séance	Thème
1 - (17 janvier 2024)	Séance introductive
2 - (24 janvier 2024)	L'enquête par questionnaire 1/2
3 - (31 janvier 2024)	L'enquête par questionnaire 2/2
4 - (7 février 2024)	Statistique descriptive univariée
5 - (14 février 2024)	Statistique bivariée : relation entre 2 variables quantitatives
6 - (21 février 2024)	Statistique bivariée : relation entre 2 variables qualitatives
7 - (6 mars 2024)	Recodages et construction d'indicateurs
8 - (13 mars 2024)	Travail en groupe sur le 4 pages
9 - (20 mars 2024)	Visualisation graphique de résultats d'enquête
10 - (27 mars 2024)	De l'échantillon à l'inférence – Retour sur la notion de test statistique
11 - (3 avril 2024)	EXPOSE en groupe sur le 4 pages
12 - (10 mars 2024)	Travail en groupe sur le 4 pages

# Déroulement des séances

---

- Généralement 3 grands moments :
  1. Apprentissage des **notions** statistiques
  2. Illustration et **application** sur le jeu de données ;
  3. Application/**travail en groupe** pour avancer sur le '4 pages'.
- La mise en œuvre sur jeux de données, particulièrement celui de l'enquête
  - <https://ee-eu.kobotoolbox.org/single/cd663db88fdd535ceeeacca06fc991a3>
- Maîtriser les manipulations dans JAMOVI et notions stat d'une séance à l'autre
  - Révision des manip / test / entraide



# Évaluation : modalités et attendus

- Réalisation d'un '**4 pages**' par groupe de 3 étudiant.es
- Travaux **intermédiaires** en vue de la réalisation du '4 pages'
- Lectures préparatoires aux séances / validations individuelles des **manipulations** statistiques et informatiques abordées pendant les conférences
- **Participation** (+ assiduité, ponctualité et solidarité) / Plagiat

Exercice	Coefficient
Assiduité/participation/solidarité	1
'4 pages' rendu final	3
Rendus intermédiaires du '4 pages'	1
Interrogation sur les lectures préparatoires aux séances + validation individuelle des manipulations apprises en séances	1
Exposé	2
Total	8

# 4 pages ?

---

**Exemples :** dispo sur le (magnifique) Moodle du cours

- section méthodologie et bibliographie (doublon)

## Jeux de données

- disponibles dans la section méthodologie

## Attentes / consignes

- premiers éléments en ligne (section évaluation)
- précisions au fil du semestre

- Baromètre d'opinion de la Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (DREES)
- Bases de données pédagogiques du Centre de données socio-politiques (CDSP)
  - Styles de vie et Environnement // "Fécondité - Contraception - Dysfonctions sexuelles" en France Métropolitaine // Faire face au Covid-19 - vague 1 // Les conditions de vie professionnelle des femmes enceintes
- European Social Survey (ESS)
- Enquête "La vie en confinement" (Vico)
- World Values Survey (WVS)
- Enquête post-électorale Youngelect
- Enquête racisme de la Commission nationale consultative des droits de l'homme (CNCDH)
- et bien sûr l'enquête « ScPo2024 » !

# Guide : 1. Objectifs du "4 pages"

---

Le « 4-pages » est le rendu final du deuxième semestre de l'atelier DECA. Il s'inspire des briefs statistiques publiés par l'INSEE, type « INSEE Première » (voir les exemples proposés dans le dossier). Son format, court et très synthétique, doit s'appuyer sur l'utilisation et l'exploitation des bases de données mises à disposition sur Moodle). Ce 4 pages répond à plusieurs objectifs pédagogiques et d'évaluation:

- Formuler une question « quantitative » qui soit problématisée
- Sélectionner et manipuler une base de données quantitative (nettoyage des données, recodage des variables) permettant de répondre à votre question [ici contraint par les données mises à disposition]
- Décrire des variables d'intérêt (production d'indicateurs et analyses descriptives adaptées)
- Analyser les liens entre variables pour "confirmer ou infirmer" des hypothèses (analyses statistiques adaptées : tableaux croisés, graphiques)

Le "4 pages" doit être rédigé à partir de la formulation d'une problématique (question) claire, à laquelle vous répondez dans un plan organisé, mêlant éléments théoriques, empiriques et analytiques (interprétatifs). Il est recommandé d'intégrer un encadré « méthodes » dans lequel vous préciserez les variables choisies pour mesurer un phénomène, ainsi que les recodages appliqués (une solution alternative est de présenter à la fin du rapport - après la conclusion - une partie « Définitions », comme dans les « 4-pages » de l'INSEE, partie où vous présenterez les définitions des concepts utilisés, voire l'opérationnalisation de vos concepts dans les données).

# Guide : 2. Format et structure générale du "4 pages"

---

Comme son nom l'indique, la taille du rapport est de 4 pages. Comme pour les « 4-pages » de l'INSEE, il est préférable de présenter votre texte en plusieurs colonnes. Plusieurs exemples de l'INSEE (et d'autres) vous sont fournis sur le moodle. Les références bibliographiques doivent être harmonisées. La structure adoptée doit comporter les éléments suivants :

- Résumé (5 lignes : objet, périmètre, problématique, littérature, hypothèse, méthode, principaux résultats, conclusion)
- Introduction
- Parties et paragraphes
- Conclusion
- Bibliographie/webographie (respect des normes en notes de bas de page et dans la biblio finale)
- Définitions

Attention, les annexes ne sont pas autorisées. Les « 4-pages » seront soumis à une vérification anti-plagiat grâce au logiciel Compilatio. Si vous utilisez un outil d'IA (comme chat GPT) vous devez également le signaler et être transparent sur la manière dont vous l'avez utilisé.

# Rendus intermédiaires

---

1. **Projet de '4 pages'** (question posée - jeu de données mobilisé) et retroplanning (avec calage de créneaux de réunion de travail collectif + répartition du travail-qui a fait quoi).
  2. **Travail biblio** : identification de 3 travaux de référence sur la thématique choisie pour le '*4 pages*' (articles, ouvrages scientifiques) accompagnés pour chacune d'une fiche synthétique de lecture d'une page (problématique - jeu de données et variables utilisées - principaux résultats **en lien avec le projet du '4 pages'**) + répartition du travail (qui a fait quoi). NB : il est attendu que les groupes continuent d'enrichir ces premières lectures au cours du semestre, pour améliorer le travail final. La bibliographie finale se doit donc d'être davantage étoffée.
  3. **Schéma causal** (opérationnalisation) : formulation de 2 à 3 hypothèses de travail avec leur mode d'opérationnalisation (quelles variables ? quels recodages ? quelle(s) méthode(s) statistique(s) ? + Mise en œuvre pour une des hypothèses). + répartition du travail (qui a fait quoi).
- ➔ Ces 3 travaux intermédiaires feront l'objet d'explicitations sur les attendus et les critères d'évaluation au fur et à mesure du semestre. Pour chaque rendu intermédiaire, il est demandé de préciser qui a fait quoi dans le trio (y compris tâches d'organisation, etc.).



# Date des rendus

---

24 janvier (séance 2) : avoir créé les groupes avant la séance et envoyé le thème et jeux de données présentis.

31 janvier (séance 3) : **Rendu intermédiaire 1** (projet avec jeu de données choisi et rétroplanning)

14 février (séance 5) : **Rendu intermédiaire 2** (travail bibliographique)

6 mars (séance 7) : **Rendu intermédiaire 3** (schéma causal)

3 avril (séance 11) : **Exposé** en groupe sur le 4 pages

15 avril : Remise du **4 pages**

# Activité : lecture 4 pages - selon le temps

---

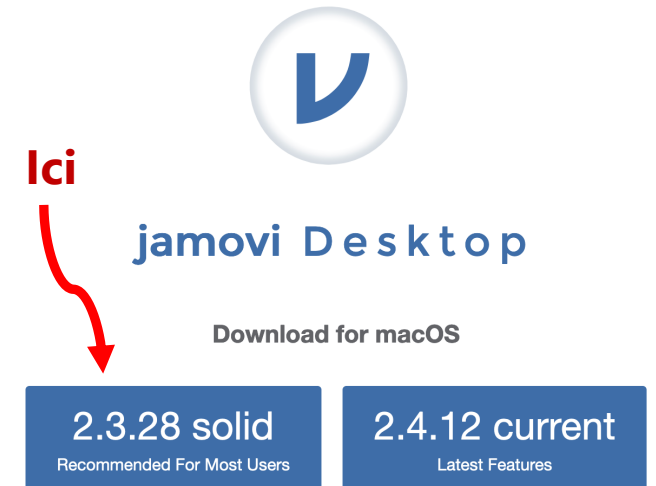
lecture d'un '*4 pages*' pour comprendre l'exercice attendu (identifier dans ce '*4 pages*' la question et la mise en œuvre empirique → passage des concepts à l'expérience empirique)

# Jamovi

---

Logiciel libre et gratuit

Téléchargement : <https://www.jamovi.org/download.html>



**Attention** : Les groupes doivent s'assurer qu'ils disposent d'au moins 1 ordinateur capable d'utiliser JAMОВI (ce qui n'est pas le cas des tablettes par exemple).

En *théorie* disponible en ligne par navigateur -> mais pas vraiment (cf. file attente, ipad, etc.)

# Activité : crash test Jamovi

---

- Installer Jamovi (si pas déjà fait)
- Crier (pas trop fort) s'il y a un problème
- Aller dans le moodle
- Récupérer le fichier contenu dans la séance 1
- Faire des premiers tests
  - Ouvrir le jeu de données
  - Prise en main
  - « essayer des machins » (terme scientifique de #dataanalyst)

# To do :

---

A FAIRE AVANT LA SEANCE 2 :

Les **groupes** '4 pages' sont **constitués** (cf. sur **moodle** dans séance 1) et ont envoyé par mail avant la séance 2 (à l'enseignant.e) les membres du groupe, leur idée de thème et le jeu de données a priori utilisé (tiré de moodle).

**Lire un 4 pages** (notamment si pas fait pendant la séance)