

Méthodes quantitatives pour les sciences sociales

Séance 2 : D'où viennent les données statistiques ?

Mattéo Lanoë

Printemps 2026

Objectifs de la séance

- ▶ **Identifier** les différentes sources de données statistiques (qui produit quoi?).
- ▶ **Comprendre** pourquoi différentes sources peuvent donner des mesures différentes d'un même phénomène (l'exemple de la délinquance).
- ▶ **Pratique** : Identifier des sources en Open Data et les importer dans RStudio.

Étude de cas : La mesure de la délinquance



Figure – Reportage France 2 : Lien

Étude de cas : La mesure de la délinquance

Support : Reportage sur les chiffres de la délinquance et des agressions.

Contexte politique : Pression pour modifier la mesure des agressions.

Étude de cas : La mesure de la délinquance

Support : Reportage sur les chiffres de la délinquance et des agressions.

Contexte politique : Pression pour modifier la mesure des agressions.

Problématique sociologique

Les statistiques sont un enjeu de contrôle et d'action publique.

Question : Pourquoi deux méthodes d'enquête mesurant le même phénomène produisent-elles des résultats différents ?

L'Acteur et la Mesure (Cesdip)

R. franç. sociol., 50-1, 2009, 31-62

Renée ZAUBERMAN
Philippe ROBERT
Sophie NÉVANEN
Emmanuel DIDIER

L'acteur et la mesure

Le comptage de la délinquance entre données administratives et enquêtes*

RÉSUMÉ

Depuis que la délinquance a émergé comme problème de société, il a fallu en estimer l'importance et l'évolution. On a d'abord recouru exclusivement aux mesures de l'activité des professionnels de l'action pénale. Au cours de la seconde partie du XX^e siècle, le doute sur la suffisance de ces données institutionnelles n'a cessé de monter – surtout avec l'émergence du débat sur la peur du crime et l'insécurité – amenant la création d'enquêtes en population générale. Dès lors, le problème de la mesure de la délinquance est devenu celui de la confrontation de différentes sources d'information, comptages d'activité institutionnelle d'un côté, enquêtes de l'autre. L'article montre comment comparer, dans le cas français, les données fournies par les enquêtes nationales de victimation disponibles et les statistiques policières éditées par le ministère de l'Intérieur, et quelles leçons tirer de cette confrontation. Finalement, il apparaît que la mesure ne doit pas servir à clore le débat sur la délinquance mais plutôt à l'ouvrir.

Dans les sociétés européennes d'ancien type, on pouvait bien s'alarmer de

Editeur: Oxydys / Tildesburg le 02/02/2024 sur <https://doi.org/10.1017/9781017000000>

Figure – Première page de (Zauberman et al. 2009)

L'Acteur et la Mesure (Cesdip)

Pour dépasser le débat médiatique, on s'appuie sur la sociologie de la quantification.

- ▶ **Référence** : Renée Zauberman, Philippe Robert, et al. (Cesdip).
- ▶ **Article** : « *L'acteur et la mesure. Le comptage de la délinquance entre données administratives et enquêtes* » (**Zauberman** et al. 2009).
- ▶ **Objectif** : Expliquer les écarts entre les données policières et les enquêtes de victimation.

Renée **Zauberman** et al. (avr. 2009). *L'acteur et la mesure : Le comptage de la délinquance entre données administratives et enquêtes*. In : *Revue française de sociologie* Vol. 50.1, p. 31-62

Typologie : Trois sources de données en sciences sociales

Avant d'analyser les biais, il faut distinguer trois grands modes de production des chiffres :

Typologie : Trois sources de données en sciences sociales

Avant d'analyser les biais, il faut distinguer trois grands modes de production des chiffres :

1. **Les enquêtes scientifiques** : Produites par des chercheurs pour répondre à une question de recherche.

Typologie : Trois sources de données en sciences sociales

Avant d'analyser les biais, il faut distinguer trois grands modes de production des chiffres :

1. **Les enquêtes scientifiques** : Produites par des chercheurs pour répondre à une question de recherche.
2. **Les données administratives** : Produites par l'État pour gérer et piloter l'action publique.

Typologie : Trois sources de données en sciences sociales

Avant d'analyser les biais, il faut distinguer trois grands modes de production des chiffres :

1. **Les enquêtes scientifiques** : Produites par des chercheurs pour répondre à une question de recherche.
2. **Les données administratives** : Produites par l'État pour gérer et piloter l'action publique.
3. **L'observation quantifiée** : Quantification de matériaux qualitatifs (observations, archives) par le chercheur.

Type 1 : Les enquêtes scientifiques

- ▶ **Producteurs** : Chercheurs (Ined, CNRS, Universités).

Type 1 : Les enquêtes scientifiques

- ▶ **Producteurs** : Chercheurs (Ined, CNRS, Universités).
- ▶ **Objectif** : Répondre à des questions sociologiques que l'administration ne pose pas.

Type 1 : Les enquêtes scientifiques

- ▶ **Producteurs** : Chercheurs (Ined, CNRS, Universités).
- ▶ **Objectif** : Répondre à des questions sociologiques que l'administration ne pose pas.
- ▶ **Caractéristiques** : Échantillons représentatifs, protocoles lourds (3 000 à 10 000 répondants).

Type 1 : Les enquêtes scientifiques

- ▶ **Producteurs** : Chercheurs (Ined, CNRS, Universités).
- ▶ **Objectif** : Répondre à des questions sociologiques que l'administration ne pose pas.
- ▶ **Caractéristiques** : Échantillons représentatifs, protocoles lourds (3 000 à 10 000 répondants).

Type 1 : Les enquêtes scientifiques

- ▶ **Producteurs** : Chercheurs (Ined, CNRS, Universités).
- ▶ **Objectif** : Répondre à des questions sociologiques que l'administration ne pose pas.
- ▶ **Caractéristiques** : Échantillons représentatifs, protocoles lourds (3 000 à 10 000 répondants).

Exemple 1 : La sexualité des jeunes

Question : La sexualité des 18-29 ans est-elle moins hétéronormée ?

→ Enquête **ENVIE** (Enquête sur la vie affective des jeunes adultes, Ined, 2022-2023).

Type 1 : Les enquêtes scientifiques



Figure – Page d'accueil de l'enquête ENVIE : <https://envie.site.ined.fr>

Type 1 : Les enquêtes scientifiques

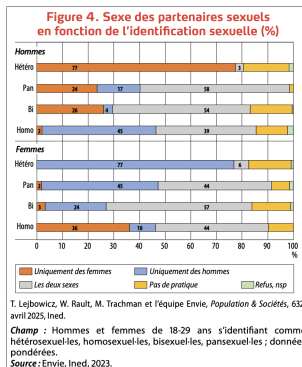


Figure – Résultats partiels issus de l'enquête ENVIE, dans (Lejbowicz et al. 2025)

Type 1 : Les enquêtes scientifiques (Suite)

Exemple 2 : La socialisation et le genre

Question : Quel est le rôle de la fratrie dans la construction du genre ?

- ▶ → Enquête **ELFE** (Enquête longitudinale française depuis l'enfance).
- ▶ Suivi de 18 000 enfants de la naissance à 20 ans.

Type 1 : Les enquêtes scientifiques (Suite)



Figure – Page d'accueil de l'enquête ELFE : <https://www.elfe-france.fr/fr/>

Type 1 : Les enquêtes scientifiques (Suite)

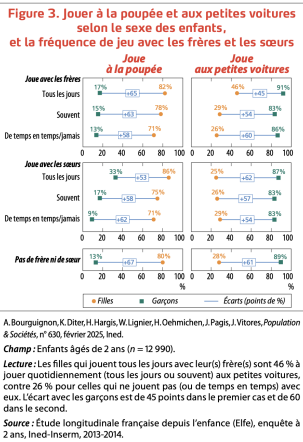


Figure – Résultats partiels issus de l'enquête ELFE, dans (Bourguignon et al. 2025)

Type 2 : Les données administratives

- ▶ **Producteurs** : Administrations, Ministères, Insee.

Type 2 : Les données administratives

- ▶ **Producteurs** : Administrations, Ministères, Insee.
- ▶ **Objectif** : Répondre aux besoins de l'État (combien de chômeurs ? Quelle efficacité économique ?).

Type 2 : Les données administratives

- ▶ **Producteurs** : Administrations, Ministères, Insee.
- ▶ **Objectif** : Répondre aux besoins de l'État (combien de chômeurs ? Quelle efficacité économique ?).
- ▶ **Exemple** : L'Enquête Emploi en Continu (EEC) de l'Insee.

Type 2 : Les données administratives

- ▶ **Producteurs** : Administrations, Ministères, Insee.
- ▶ **Objectif** : Répondre aux besoins de l'État (combien de chômeurs ? Quelle efficacité économique ?).
- ▶ **Exemple** : L'Enquête Emploi en Continu (EEC) de l'Insee.

Type 2 : Les données administratives

- ▶ **Producteurs** : Administrations, Ministères, Insee.
- ▶ **Objectif** : Répondre aux besoins de l'État (combien de chômeurs ? Quelle efficacité économique ?).
- ▶ **Exemple** : L'Enquête Emploi en Continu (EEC) de l'Insee.

Peut-on les utiliser pour la recherche ?

Oui, mais avec précaution. Les variables sont définies par des besoins de gestion, pas par des hypothèses sociologiques.

Type 3 : L'observation quantifiée (Méthodes mixtes)

Comment mesurer ce qui n'est pas enregistré (ni par l'administration, ni par questionnaire) ? *Exemple : Le contrôle au faciès.*

- ▶ **Étude de cas** : Fabien Jobard et al. (Cesdip) à la Gare du Nord (Paris).
- ▶ **Méthode** : "Population de référence" vs "Population contrôlée".
- ▶ On quantifie des observations directes pour objectiver une discrimination.

Fabien **Jobard** et al. (déc. 2012). *Mesurer les discriminations selon l'apparence : une analyse des contrôles d'identité à Paris* : in : *Population* Vol. 67.3, p. 423-451

Objectiver la discrimination (1)

Tableau 2 : La population de référence (Qui est dans la gare ?)

Tableau 2. Répartition des origines apparentes selon les lieux d'observation dans la population de référence (% et effectifs)

Lieu d'observation	Répartition des origines apparentes (%)						Effectifs
	Blancs	Noirs	Maghrébins	Indo-Pakistanaïs	Asiatiques	Autres	
GDN Station	70,6	15,3	8,7	1,9	3,3	0,3	8008
GDN RER	42,7	32,1	14,8	6,1	3,7	0,7	8496
GDN Thalys	86,4	7,0	2,8	0,7	3,0	0,1	3726
Châtelet Station	52,1	27,9	11,1	3,4	4,9	0,5	9409
Châtelet Innocents	54,8	22,7	14,3	1,8	5,8	0,6	7687
Total							37326
Source : Enquête CI (Cesdip-OSI, 2007-2008).							

Fabien Jobard et al. (déc. 2012). *Mesurer les discriminations selon l'apparence : une analyse des contrôles d'identité à Paris* : in : *Population* Vol. 67.3, p. 423-451

Objectiver la discrimination (2)

Tableau 3 : La population contrôlée (Qui est contrôlé par la police ?)

Tableau 3. Effectifs et répartition (%) de la population de référence et de la population contrôlée selon l'origine apparente

Population de référence					
Lieux d'observation	Effectifs	Répartition par origine apparente (%)			
		Blancs	Noirs	Maghrébins	Autres
GDN Station	8008	70,6	15,3	8,7	5,5
GDN RER	8496	42,7	32,1	14,8	10,5
GDN Thalys	3726	86,4	7,0	2,8	3,8
Châtelet Station	9409	52,1	27,9	11,1	8,8
Châtelet Innocents	7687	54,8	22,7	14,3	8,2

Population contrôlée					
Lieux d'observation	Effectifs	Répartition par origine apparente (%)			
		Blancs	Noirs	Maghrébins	Autres
GDN Station	123	17,9	26,0	29,3	26,8
GDN RER	129	20,2	50,4	12,4	17,1
GDN Thalys	121	52,1	23,1	9,9	14,9
Châtelet Station	68	8,8	54,4	27,9	8,8
Châtelet Innocents	83	28,9	47,0	22,9	1,2

Lecture : Parmi les 8008 personnes faisant partie de la population de référence à GDN-Station, 5 654 (soit 70,6 %) sont blanches. Au sein des 123 personnes contrôlées dans ce même lieu, c'est le cas de 22 d'entre elles (soit 17,9 %).

Source : Enquête CI (Cesdip-OSI, 2007-2008).

Fabien **Jobard** et al. (déc. 2012). *Mesurer les discriminations selon l'apparence : une analyse des contrôles d'identité à Paris* : in : *Population* Vol. 67.3, p. 423-451

Retour à la délinquance : Comprendre les écarts

Pourquoi les chiffres de la police diffèrent-ils des enquêtes ?

Statistiques de Police

- ▶ Mesurent l'**activité de l'institution** (plaintes, arrestations).
- ▶ Dépendent du "taux de plainte" (propension des victimes à se déplacer).
- ▶ Dépendent de l'initiative policière (proactivité).

Enquêtes de Victimation

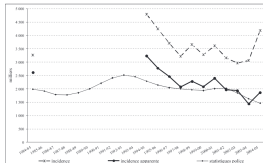
- ▶ Interrogent directement la population ("Avez-vous été victime... ?").
- ▶ Mesurent la **masse noire** (ce qui n'est jamais signalé).
- ▶ Donnent une image plus proche de la "réalité subie".

"Le comptage dépend de l'acteur qui compte."

L'écart entre police et victimation (Zauberman et al.)

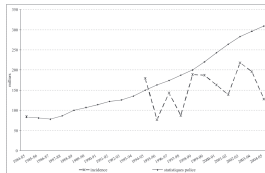
Comparaison des tendances sur longue période (1985-2005)

FIGURE II. – Vols de et dans les véhicules (1984-2005) (en milliers)



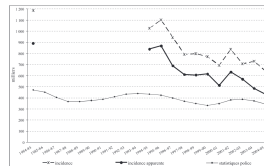
Source : Cerdip, Insee, ministère de l'Intérieur.
Champ : France hexagonale.

FIGURE VII. – Violences physiques caractérisées (1984-2005) (en milliers)



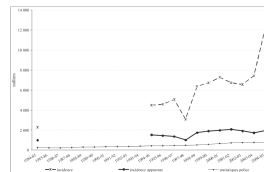
Source : Cerdip, Insee, ministère de l'Intérieur.
Champ : France hexagonale.

FIGURE I. – Cambriolages (1984-2005) (en milliers)



Source : Cerdip, Insee, ministère de l'Intérieur.
Champ : France hexagonale.

FIGURE VI. – Ensemble des agressions (1984-2005) (en milliers)



Source : Cerdip, Insee, ministère de l'Intérieur.
Champ : France hexagonale.

Source : Zauberman et al., 2009.

Rappel : L'environnement de travail

1. Créer un Projet

- ▶ File > New Project.
- ▶ Cela définit votre "Espace de travail" (le dossier où tout se passe).

2. Le format CSV (*Comma Separated Values*)

- ▶ C'est un fichier texte brut, léger et universel.
- ▶ **Attention au séparateur !**
- ▶ Anglais : virgule (,) → `read.csv()`
- ▶ Français : point-virgule (;) → `read.csv(..., sep=";")`

La Démonstration (Titanic)

Le code à exécuter

1. Je télécharge le fichier.
 2. Je le déplace dans mon dossier Projet.
 3. J'importe :
- ```
titanic <- read.csv("titanic.csv")
View(titanic)
```



## Exercice : Chercher et utiliser des données

**Objectif** : Par groupe de 3, identifier une source fiable, importer les données et répondre à une question factuelle.

### La mission

1. **Chercher** le jeu de données sur le web (Indices fournis).
2. **Télécharger** le fichier (format CSV ou Excel/XLS).
3. **Déplacer** le fichier dans votre dossier de projet R.
4. **Importer** dans RStudio (Attention au séparateur!).
5. **Trier/Calculer** pour trouver la réponse.
6. **Remplir** la fiche d'identité du jeu de données.

# Groupe 1 : La Représentation Politique

**Mission :** Identifier le doyen ou le benjamin de l'Assemblée nationale.

## Questions de recherche

1. Qui est le ou la député(e) le plus jeune de l'actuelle Assemblée ?
2. Quel est l'âge moyen des députés ?

## Vos indices :

- ▶ Cherchez sur des sites de dépôts officiels de données de l'État (*data.gouv...*).
- ▶ **Appelez-moi** une fois le fichier identifié avant de l'importer.
- ▶ Identifiez les variables (colonnes) pertinentes (Date de naissance ? Âge ?).

## Groupe 2 : La Géographie Culturelle

**Mission :** Identifier les hauts lieux de la culture en province.

### Question de recherche

Quel est le nombre minimum de visites dans un musée en France, quel est le nombre maximum ? Quelle moyenne ?

### Vos indices :

- ▶ Source institutionnelle : Ministère de la Culture.
- ▶ Cherchez sur la plateforme Open Data du ministère.
- ▶ Mots clés : "Fréquentation" et "Musées".
- ▶ **Appelez-moi** une fois le fichier identifié.

## Groupe 3 : La Démographie Locale

**Mission :** Identifier les écarts de population sur le territoire.

### Questions de recherche

1. Quelle est la commune la moins peuplée ?
2. Quelle est la commune qui en compte le plus ?
3. Quelle est la population moyenne des communes en France ?

### Vos indices :

- ▶ Cherchez des informations en lien avec les "Communes de France" sur les dépôts officiels.
- ▶ Choisissez le jeu de données le plus récent (2023 ou 2024).
- ▶ **Appelez-moi** une fois le fichier identifié.

# Fiche d'identité du jeu de données

*À remplir pour valider l'exercice (une feuille par groupe).*

**1. Nom du fichier téléchargé**

(Indiquez le nom exact, ex : donnees\_brutes.csv)

**2. Qui est le producteur de la donnée ?**

(Quelle administration, institut ou association ? Source officielle ?)

**3. Nature de la donnée**

(Données administratives de gestion OU Enquête par questionnaire ?)

**4. Format du fichier**

(CSV virgule, CSV point-virgule, Excel ?)

**5. La réponse à la question posée**

(Nom + Chiffre associé)