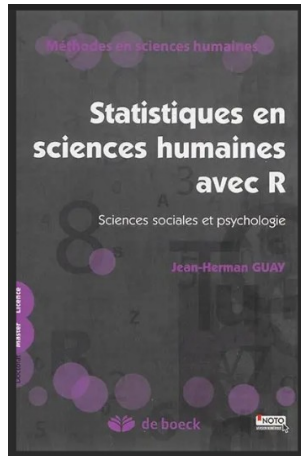
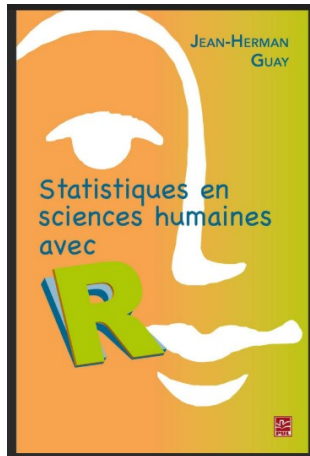


# Explorer les données

Jean-Herman Guay, professeur émérite, Université de Sherbrooke



# Explorer les données

## « explorer », définition dans le dictionnaire Littré

### explorer

(èk-splo-ré) v. a.

1 Parcourir en examinant, en cherchant à découvrir. Il voulut explorer ces contrées.

• Parcourir du regard.

*Et quand sur cette mer, las de chercher sa route, Du firmament splendide il explore la route, Des astres inconnus s'y lèvent à ses yeux, LAMARTINE, Harm. IV, 13.*

2 Fig. Explorer une bibliothèque, les vieux documents.

3 Examiner attentivement les symptômes d'une maladie ; sonder une plaie, une fistule.

### HISTORIQUE

XVI<sup>e</sup> s.

*Apportez-moi les œuvres de Virgile, et, par trois fois avec l'ongle les ouvrant, explorerons, par les vers du nombre entre nous convenu, le sort futur de votre mariage, RABELAIS, Pant. III, 10.*

### ÉTYMOLOGIE

Lat. *explorare*, explorer. D'après Pott, *plorare* serait ici le même que *plorare*, pleurer, viendrait du radical sanscrit *plu*, couler, et aurait pris le sens de aller, et, avec *ex*, aller au loin.

[PLORARE] (latin : crier, appeler, se lamenter)

[EX-PLORARE]

explorer

= examiner,

aller voir

[IM-PLORARE]

implorer

= supplier

en suppliant

[DE-PLORARE]

déplorer

= regretter,

se lamenter

[PLEURER]

(héritage direct)

= verser des larmes

# Explorer les données

« donnée », définition dans le dictionnaire Littré

## donnée [1]

(do-née) **s. f.**

**1** Le point sur lequel on fonde un raisonnement, et qui est reconnu. Il faut partir de cette donnée. Les données de l'observation.

*Si un homme raisonne mal, c'est qu'il n'a pas les données pour raisonner mieux, DIDEROT, Sur le livre de l'Esprit.*

## Définition

Représentation d'une [information](#) sous une forme conventionnelle destinée à en permettre le stockage, la transmission ou le traitement par des moyens automatiques.

## Notes

Les données peuvent être stockées dans des [fichiers](#) ou des [bases de données](#), et être rassemblées en [enregistrements](#).



# L'esprit du cours

Imaginez avoir accès aux données utilisées chaque mois pour calculer le taux de chômage au Canada. Imaginez travailler avec les variables de l'Enquête mondiale sur les valeurs (World Values Survey), couvrant une centaine de pays depuis 1980. Imaginez obtenir en temps réel des données boursières ou climatiques, ou encore explorer les vastes ressources de la Banque mondiale, cet organisme des Nations Unies qui compile des informations en matière d'éducation, de santé, de violence ou de démographie. C'est l'invitation de ce cours!

Si au terme de nos quatre semaines, vous avez moins peur des données, si vous êtes heureux de raconter que vous avez fait un peu de programmation ou si vous pouvez questionner des chiffres que vous lisez dans les médias... c'est que nous aurons atteint nos objectifs. Si vous décidez ensuite d'aller plus loin, d'explorer d'autres données, de développer une curiosité pour le traitement des données avec Python, nous allons lever nos verres! Contribuer à développer une citoyenneté numérique plus forte, résiliente, capable de vérifier à la source les données... c'est peut-être l'objectif ultime du cours!

# Le plan du cours

## Première semaine

Présentation de la démarche du cours et des rudiments

Visionnement de la vidéo d'introduction

Premier exemple: **les incidents de sécurité à Sherbrooke**. Source : Données Québec

Deuxième exemple: **la fréquentation du Marché de Racine**. Source : Sondage de Céline Garant

Consultations personnelles

## Deuxième semaine

Retour sur les difficultés rencontrées et partage

Troisième exemple: les **données climatiques** au Québec. Source : Environnement Canada

Quatrième exemple: le **boycott des produits américains**. Source : Census Bureau

Consultations personnelles

## Troisième semaine

Retour sur les difficultés rencontrées et partage

Cinquième exemple: Les **valeurs des Américains et des Canadiens**. Source : World Values Survey

Sixième exemple: Le **taux de chômage** au Canada. Source : Statistique Canada, Enquête sur la population active

Consultations personnelles

## Quatrième semaine

Retour sur les difficultés rencontrées et partage

Septième exemple: **L'espérance de vie en Chine et en Inde** : deux modèles. Source : la Banque mondiale

Huitième exemple: **Les données boursières**, Bombardier et le Dow Jones. Source : Yahoo Finance

## Conclusion du cours

Ce que vous avez retenu

Ce qui a été le plus difficile: les frustrations

Ce qui a été le plus facile: les eurékas

# Les objectifs du cours

- **TROUVER.** Découvrir des sources de données québécoises, canadiennes et internationales, reconnues et utilisées par les médias, les administrations publiques, les associations de la société civile et la recherche universitaire. Comprendre comment les télécharger (limites, formats, spécificités)
- **TRAITER.** Comprendre les outils de base pour traiter une variable à la fois, pour combiner deux variables et pour réorganiser les données (python, pandas, etc).
- **INTERPRÉTER.** Savoir comment lire les résultats obtenus, ce qu'on peut en dire et ce qu'on peut encore aller chercher pour aller plus loin.

# Les rudiments

Des mots : variable, valeur, enregistrement, quantitatif, qualitatif

Un format : des colonnes et des rangées

Des exemples et encore des exemples



# Est-ce que toutes les femmes de ma famille jouent de la musique ?

- enregistrements : membres de ma famille
- Première variable : le **genre** (valeurs : homme, femme)
- Deuxième variable : joue un **instrument** de musique (valeurs)

1

## Variables

nom	Adèle
genre	Femme
age	55
ville	Montréal
joue_musique	Oui
connais_python	Non
nb ordi	2

Enregistrements\* :  
membres de ma  
famille

nom	genre	age	ville	joue_musique	connais_python	nb_ordi
Adèle	Femme	55	Montréal	Oui	Non	2
Alice	Femme	53	Magog	Oui	Non	2
Adrienne	Femme	75	magog	Oui	Non	1
Alma	Femme	71	Laval	Oui	Non	1
Pierre	Homme	58	Carignan	Oui	Non	2
Joseph	Homme	38	Laval	Non	Non	1
Julie	Femme	50	Laval	Non	Non	2
Jean-Herman	Homme	66	Sherbrooke	Non	Oui	2
Vincent	Homme	32	montréal	Non	Oui	3
Émilie	Femme	40	Montréal	Oui		3

Qualitative

Qualitative

Quantitative

Qualitative

Qualitative

Qualitative

Quantitative

Équivalents à enregistrements :  
unités d'analyse, observations,  
cas, individus, etc.  
Ensemble des  
enregistrements : population,  
échantillon probabiliste, non  
probabiliste



# Combien y a-t-il de d'entreprises enregistrées comme coopérative au Québec? Combien d'entreprises ont de 750 à 999 employés?

- enregistrements : entreprises
- Première variable : la forme juridique (valeurs : coopérative, etc.)
- Deuxième variable : nombre d'employés (valeurs : 1... ou catégories)

# 2

## Variables

Enregistrements :  
entreprises

NEQ	IND_FAIL	DAT_IMMAT	COD_REGIM_JURI	COD_INTVAL_EMPLO_QUE	DAT_CESS_PREVU	COD_STAT_IMMAT	COD_FORME_JURI
1140030355		1994-01-18	24N			RO	CIE
1140030363		1994-01-18	24A			RO	CIE
1140031379		1994-01-19	23N			RO	APE
1140031486		1994-01-19	24A			RO	CIE
1140031510	N	1994-01-19	24O			IM	CIE
1140031536		1994-01-19	24O			RO	CIE
1140031551		1994-01-19	24A			RO	CIE
1140031908		1994-01-19	24A			RO	CIE
1140031916		1994-01-19	24N			RO	CIE
1140031924	N	1994-01-19	24O			RO	CIE
1140031957		1994-01-19	24A			RO	CIE
1140031965	N	1994-01-19	24B			IM	CIE
1140031973		1994-01-19	24A			RO	CIE
1140031999	O	1994-01-19	24O			IM	CIE
1140032336		1994-01-20	23O			RO	APE
1140032419		1994-01-20	24O			RO	CIE
1140032435	N	1994-01-20	24C			IM	CIE
1140032468		1994-01-20	400A			RO	CIE
1140032500		1994-01-20	24O			RO	CIE
1140032526		1994-01-20	24O			RO	CIE
1140032567		1994-01-20	400O			RO	CIE
1140032625		1994-01-20	24A			RO	CIE
1140033045		1994-01-20	24O			RO	CIE
1140033052		1994-01-20	24A			RO	CIE

NEQ	1140030355
IND_FAIL	NaN
DAT_IMMAT	1994-01-18
COD_REGIM_JURI	24.0
COD_INTVAL_EMPLO_QUE	N
DAT_CESS_PREVU	NaN
COD_STAT_IMMAT	RO
COD_FORME_JURI	CIE
DAT_STAT_IMMAT	1997-07-17
COD_REGIM_JURI_CONSTI	400.0
DAT_DEPO_DECLR	NaN
AN_DECL	NaN
AN_PROD	2025.0
DAT_LIMIT_PROD	2025-11-15
AN_PROD_FRE	2024.0
DAT_LIMIT_PROD_PRE	2024-11-15
DAT_MAJ_INDEX_NOM	1997-07-17
COD_ACT_ECON_CAE	2.0
NO_ACT_ECON_ASSUJ	1.0
DESC_ACT_ECON_ASSUJ	NON DÉCLARÉ
COD_ACT_ECON_CAE2	NaN
NO_ACT_ECON_ASSUJ2	NaN
DESC_ACT_ECON_ASSUJ2	NaN
NOM_LOCLT_CONSTI	MONTREAL
DAT_CONSTI	1994-01-05
IND_CONVEN_UNMN_ACTNR	N
IND_RET_TOUT_POUVR	NaN
IND_LIMIT_RESP	NaN
DAT_DEB_RESP	NaN
DAT_FIN_RESP	NaN
OBJET_SOC	NaN
NO_MTR_VOLONT	NaN
ADR_DOMCL_ADR_DISP	N
ADR_DOMCL_LIGN1_ADR	2707, CAZENEUVE
ADR_DOMCL_LIGN2_ADR	SAINT-LAURENT (QUEBEC)
ADR_DOMCL_LIGN3_ADR	NaN
ADR_DOMCL_LIGN4_ADR	H4R1Z7





# Combien y a-t-il de d'entreprises enregistrées comme coopérative au Québec? Combien d'entreprises ont de 750 à 999 employés?

- enregistrements : entreprises
- Première variable : la forme juridique (valeurs : coopérative, etc.)
- Deuxième variable : nombre d'employés (valeurs : 1... ou catégories)

## Valeurs

# 2

## 4 SPÉCIFICATIONS FONCTIONNELLES

Cette section traite de certains éléments fonctionnels qui aident à la compréhension et à l'interprétation du jeu de données.

### 4.1 Bloc de données : Entreprise

## Variables

Position	Libellé d'en-tête	Domaine de valeur <sup>1</sup>	Interprétation, particularité ou précision
1	NEQ	S. O.	Numéro d'entreprise du Québec
2	IND_FAIL	S. O.	Indicateur de faillite
3	DAT_IMMAT	S. O.	Date d'immatriculation
4	COD_REGIM_JURI	REGIM_JURI	Code associé au régime juridique sous lequel a été constituée l'entreprise
5	COD_INTVAL_EMPLO_QUE	INTVAL_EMPLO_QUE	Ordre de grandeur du nombre d'employés au Québec
6	DAT_CESS_PREVU	S. O.	Date de cessation prévue de l'existence de l'entreprise
7	COD_STAT_IMMAT	STAT_IMMAT	Statut d'immatriculation
8	COD_FORME_JURI	FORM_JURI	Forme juridique
9	DAT_STAT_IMMAT	S. O.	Date de la dernière mise à jour du statut liée à l'immatriculation de l'entreprise au registre
10	COD_REGIM_JURI_CONSTI	REGIM_JURI	Code du régime juridique lors de la constitution
11	DAT_DEPO_DECLR	S. O.	Date à laquelle la dernière déclaration de mise à jour annuelle a été déposée
12	AN_DECL	S. O.	Année visée par la dernière déclaration de mise à jour annuelle déposée
13	AN_PROD	S. O.	Année de production de la déclaration de mise à jour annuelle
14	DAT_LIMIT_PROD	S. O.	Date limite de production de la déclaration de mise à jour annuelle
15	AN_PROD_PRE	S. O.	Année de production de la déclaration de mise à jour annuelle précédente
16	DAT_LIMIT_PROD_PRE	S. O.	Date limite de production de la déclaration de mise à jour annuelle précédente
17	DAT_MAJ_INDEX_NOM	S. O.	Date de mise à jour de l'index des noms
18	COD_ACT_ECON_CAE	ACT_ECON	Code de l'activité économique  La classification des activités économiques utilisée par le Registraire des entreprises est disponible sur son site Internet.
19	NO_ACT_ECON_ASSUJ	S. O.	Numéro de rang de l'activité économique de l'assujetti
20	DESC_ACT_ECON_ASSUJ	S. O.	Description de l'activité économique de l'assujetti

### 6.6 INTVAL\_EMPLO\_QUE

Type de domaine de valeur	Code du domaine de valeur	Domaine de valeur
INTVAL_EMPLO_QUE	A	De 1 à 5
INTVAL_EMPLO_QUE	B	De 6 à 10
INTVAL_EMPLO_QUE	C	De 11 à 25
INTVAL_EMPLO_QUE	D	De 26 à 49
INTVAL_EMPLO_QUE	E	De 50 à 99
INTVAL_EMPLO_QUE	F	De 100 à 249
INTVAL_EMPLO_QUE	G	De 250 à 499
INTVAL_EMPLO_QUE	H	De 500 à 749
INTVAL_EMPLO_QUE	I	De 750 à 999
INTVAL_EMPLO_QUE	J	De 1000 à 2499
INTVAL_EMPLO_QUE	K	De 2500 à 4999
INTVAL_EMPLO_QUE	L	Plus de 5000
INTVAL_EMPLO_QUE	N	Non déclaré
INTVAL_EMPLO_QUE	O	Aucun

### 6.7 FORM\_JURI

Type de domaine de valeur	Code du domaine de valeur	Domaine de valeur en français
FORM_JURI	APE	Personne morale sans but lucratif
FORM_JURI	ASS	Association
FORM_JURI	AU	Autre
FORM_JURI	AUG	Groupement de personnes
FORM_JURI	CIE	Société par actions ou compagnie
FORM_JURI	COP	Coopérative
FORM_JURI	IND	Entreprise individuelle
FORM_JURI	MUT	Mutuelle d'assurances
FORM_JURI	SEC	Société en commandite
FORM_JURI	SENC	Société en nom collectif
FORM_JURI	SENCRL	Société en nom collectif à responsabilité limitée
FORM_JURI	SEP	Société en participation
FORM_JURI	SMU	Société de secours mutuel
FORM_JURI	SOC	Société non constituée au Québec
FORM_JURI	SYC	Syndicat de copropriété
FORM_JURI	SYP	Syndicat professionnel



REGISTRAIRE DES ENTREPRISES

DONNÉES PUBLIQUES SUR LES ENTREPRISES AU QUÉBEC

Guide d'utilisation

OU  
Codebook



# Combien y a-t-il de d'entreprises enregistrées comme coopérative au Québec? Combien d'entreprises ont de 750 à 999 employés?

- enregistrements : entreprises
- Première variable : la forme juridique (valeurs : coopérative, etc.)
- Deuxième variable : nombre d'employés (valeurs : 1... ou catégories)

## Valeurs

2

### 4 SPÉCIFICATIONS FONCTIONNELLES

Cette section traite de certains éléments fonctionnels qui aident à la compréhension et à l'interprétation du jeu de données.

#### 4.1 Bloc de données : Entreprise

## Variables

Position	Libellé d'en-tête	Domaine de valeur <sup>1</sup>	Interprétation, particularité ou précision
1	NEQ	S. O.	Numéro d'entreprise du Québec
2	IND_FAIL	S. O.	Indicateur de faillite
3	DAT_IMMAT	S. O.	Date d'immatriculation
4	COD_REGIM_JURI	REGIM_JURI	Code associé au régime juridique sous lequel a été constituée l'entreprise
5	COD_INTVAL_EMPLO_QUE	INTVAL_EMPLO_QUE	Ordre de grandeur du nombre d'employés au Québec
6	DAT_CESS_PREVU	S. O.	Date de cessation prévue de l'existence de l'entreprise
7	COD_STAT_IMMAT	STAT_IMMAT	Statut d'immatriculation
8	COD_FORME_JURI	FORM_JURI	Forme juridique
9	DAT_STAT_IMMAT	S. O.	Date de la dernière mise à jour du statut liée à l'immatriculation de l'entreprise au registre
10	COD_REGIM_JURI_CONSTI	REGIM_JURI	Code du régime juridique lors de la constitution
11	DAT_DEPO_DECLR	S. O.	Date à laquelle la dernière déclaration de mise à jour annuelle a été déposée
12	AN_DECL	S. O.	Année visée par la dernière déclaration de mise à jour annuelle déposée
13	AN_PROD	S. O.	Année de production de la déclaration de mise à jour annuelle
14	DAT_LIMIT_PROD	S. O.	Date limite de production de la déclaration de mise à jour annuelle
15	AN_PROD_PRE	S. O.	Année de production de la déclaration de mise à jour annuelle précédente
16	DAT_LIMIT_PROD_PRE	S. O.	Date limite de production de la déclaration de mise à jour annuelle précédente
17	DAT_MAJ_INDEX_NOM	S. O.	Date de mise à jour de l'index des noms
18	COD_ACT_ECON_CAE	ACT_ECON	Code de l'activité économique La classification des activités économiques utilisée par le Registraire des entreprises est disponible sur son site Internet.
19	NO_ACT_ECON_ASSUJ	S. O.	Numéro de rang de l'activité économique de l'assujetti
20	DESC_ACT_ECON_ASSUJ	S. O.	Description de l'activité économique de l'assujetti

#### 6.6 INTVAL\_EMPLO\_QUE

Type de domaine de valeur	Code du domaine de valeur	Domaine de valeur
COD_INTVAL_EMPLO_QUE	A	De 1 à 5
O 1773213	B	De 6 à 10
A 840475	C	De 11 à 25
B 81905	D	De 26 à 49
N 72692	E	De 50 à 99
C 52243	F	De 100 à 249
D 20159	G	De 250 à 499
E 11523	H	De 500 à 749
F 5622	I	De 750 à 999
G 1699	J	De 1000 à 2499
H 483	K	De 2500 à 4999
J 428	L	Plus de 5000
I 340	N	Non déclaré
L 132	O	Aucun
K 112		

#### 6.7 FORM\_JURI

Type de domaine de valeur	Code du domaine de valeur	Domaine de valeur en français
COD_FORME_JURI	APE	Personne morale sans but lucratif
CIE 1447992	ASS	Association
IND 979342	AU	Autre
SENC 211490	AUG	Groupement de personnes
APE 118732	CIE	Société par actions ou compagnie
SYC 44381	COP	Coopérative
ASS 16671	IND	Entreprise individuelle
SEC 13416	MUT	Mutuelle d'assurances
AUG 9640	SEC	Société en commandite
COP 7929	SENC	Société en nom collectif
SEP 6927	SENCRL	Société en nom collectif à responsabilité limitée
SOC 5139	SEP	Société en participation
AU 145	SMU	Société de secours mutuel
MUT 107	SOC	Société non constituée au Québec
	SYC	Syndicat de copropriété
	SYP	Syndicat professionnel



REGISTRAIRE DES ENTREPRISES

DONNÉES PUBLIQUES SUR LES ENTREPRISES AU QUÉBEC

Guide d'utilisation

OU  
Codebook



# Parmi vous, combien ont déjà travaillé avec Excel? Combien ont téléchargé des fichiers?

- enregistrement : répondants
- Première variable : travail\_excel (valeurs : souvent, quelques fois, etc.)
- Deuxième variable : télécharger\_fichiers (valeurs : souvent, quelques fois, etc)

Variables

3

Explorer les données. Petit questionnaire pour mieux vous connaître!

B I U ↺ ↻

Les données sont confidentielles. Votre courriel ne sera pas retenu comme information. Merci.

1. Avez-vous déjà suivi un cours à l'UTA? \*

☐ Oui

☐ Non

4. Avez-vous déjà eu l'occasion de travailler avec Excel ou un logiciel semblable? \*

☐ Souvent

☐ Quelques fois

☐ Rarement

☐ Jamais

Enregistrements :  
répondants

5. Avez-vous déjà eu l'occasion d'importer ou de télécharger des fichiers sur votre ordinateur? \*

☐ Souvent

☐ Quelques fois

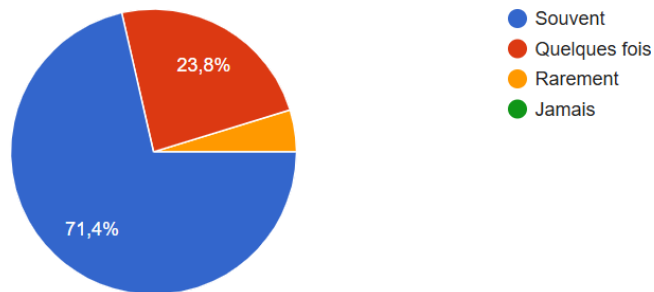
☐ Rarement

☐ Jamais

	A	B	C	D
1	Horodateur	1. Avez-vous déjà suivi un cours à l'UTA?	4. Avez-vous déjà eu l'occasion de travailler avec Excel ou un logiciel semblable?	5. Avez-vous déjà eu l'occasion d'importer ou de télécharger des fichiers sur votre ordinateur?
2	9-11-2025 7:45:12	Oui	Souvent	Souvent
3	9-11-2025 7:46:25	Oui	Quelques fois	Souvent
4	9-11-2025 11:49:46	Non	Souvent	Souvent
5	9-29-2025 11:47:13	Oui	Souvent	Souvent
6	9-29-2025 14:51:13	Oui	Souvent	Souvent
7	9-30-2025 14:03:57	Oui	Souvent	Souvent
8	9-30-2025 14:05:35	Oui	Souvent	Souvent
9	9-30-2025 14:10:39	Non	Quelques fois	Quelques fois
10	9-30-2025 14:12:06	Oui	Souvent	Souvent
11	9-30-2025 14:12:18	Non	Quelques fois	Souvent
12	9-30-2025 14:13:34	Oui	Souvent	Souvent
13	9-30-2025 14:19:01	Non	Souvent	Souvent
14	9-30-2025 14:25:29	Oui	Souvent	Souvent
15	9-30-2025 14:26:43	Oui	Rarement	Souvent
16	9-30-2025 14:27:08	Non	Souvent	Souvent
17	9-30-2025 14:29:18	Oui	Souvent	Quelques fois
18	9-30-2025 14:29:40	Oui	Quelques fois	Souvent
19	9-30-2025 14:29:43	Oui	Souvent	Souvent
20	9-30-2025 14:53:20	Oui	Quelques fois	Souvent
21	9-30-2025 19:15:17	Oui	Souvent	Souvent
22	10-1-2025 10:31:16	Oui	Quelques fois	Souvent

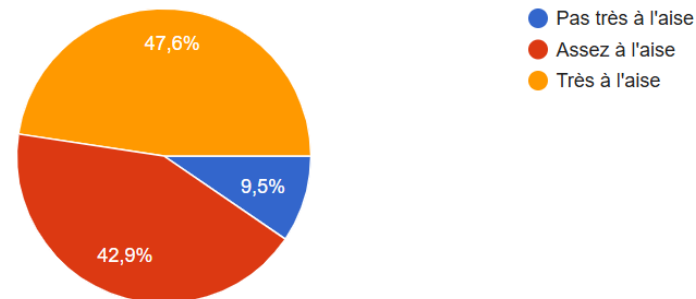
6. Dans les médias, quand on vous montre un tableau de chiffres, y portez-vous attention?

21 réponses



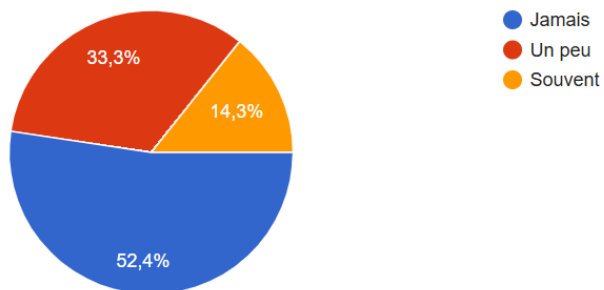
7. Diriez-vous que vous êtes à l'aise pour comprendre des informations statistiques comme des moyennes ou des médianes?

21 réponses



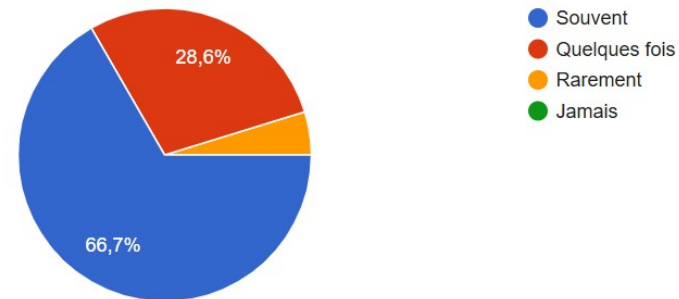
8. Avez-vous déjà fait un peu de programmation (html, javascript, java, c++)?

21 réponses














4. Avez-vous déjà eu l'occasion de travailler avec Excel ou un logiciel semblable?

21 réponses



# Excel vs Python (pandas)

Fonctionnalité / Usage	Excel	Python (pandas)
 Taille des données	✗ Limité à ~1M lignes	✓ Millions de lignes (selon RAM)
 Automatisation	✗ Peu reproductible (clics manuels)	✓ Code réutilisable et automatisé
 Valeurs manquantes	✗ Pas de gestion native	✓ <code>.fillna()</code> , <code>.dropna()</code> , interpolation
 Jointures / fusion de tables	✗ Complexe	✓ <code>merge()</code> , <code>join()</code> comme SQL
 Analyse temporelle	✗ Basique	✓ Décalage, fenêtres glissantes, resampling
 Statistiques avancées	⚠ Très limité	✓ Régressions, stats, IA, ML (via scikit-learn, statsmodels)
 Formats supportés	⚠ CSV, XLSX	✓ CSV, Excel, JSON, SQL, Parquet, HDF5, etc.
 Vitesse	✗ Lent sur gros fichiers	✓ Optimisé (NumPy en arrière-plan)
 Visualisation	✓ Graphiques intégrés	✓ Graphiques avancés (Matplotlib, Seaborn)
 Collaboration	✓ Facile (fichier partagé)	⚠ Nécessite Git/Jupyter pour collaborer efficacement
 Programmabilité	⚠ VBA mais limité	✓ Python complet + écosystème scientifique

# Essayons ensemble



## Un questionnement

- Enregistrement
- Première variable, les valeurs, type de variable
- Deuxième variable, les valeurs, type de variable



# Est-ce que les femmes, plus que les hommes, fréquentent régulièrement le marché public ?

- Enregistrements : des répondants à un sondage
- Première variable : le **genre** (valeurs : homme, femme)
- Deuxième variable : la **fréquentation** du marché public (valeurs : régulièrement, rarement, etc.)

4

Variables		
Index	<i>genre</i>	<i>fréquentation</i>
0	femme	régulièrement
1	homme	rarement
2	femme	régulièrement
3	femme	régulièrement
4	homme	souvent
Qualitative		Qualitative

Enregistrements : des répondants



# Les Américains, plus que les Canadiens, croient qu'il faut inculquer aux enfants la valeur d'obéissance?

- Enregistrements : des répondants à un sondage mené dans les deux pays
- Première variable : le pays du répondant (valeurs : Canada, États-Unis)
- Deuxième variable : l'obéissance est-elle choisie comme valeur à inculquer ? (valeurs : oui, non)

5

Variables		
Index	<i>pays</i>	<i>obéissance</i>
0	États-Unis	présente
1	États-Unis	absente
2	Canada	absente
3	Canada	présente
4	Canada	absente
Qualitative		Qualitative

Enregistrements : des répondants





# Parmi les incidents de sécurité publique traités par un service de police, quel type se produit le plus souvent ?

- Enregistrements : les incidents traités
- Première variable : le type d'incident (valeurs : accident, vol, introduction par effraction, etc.)

6

Variables	
Index	type
0	Voie de fait
1	Voie de fait
2	Méfait
3	Accident mortel
4	Méfait
Qualitative	

Enregistrements : des incidents



# Depuis 1960, l'espérance de vie a-t-elle connu une croissance plus forte en Chine ou en Inde ?

- Enregistrements : les années
- Première variable : l'espérance de vie en Chine (valeurs : 50 ans, 60 ans, etc.)
- Deuxième variable : l'espérance de vie en Inde (valeurs : 50 ans, 60 ans, etc.)

7

## Variables

Index	<i>année</i>	<i>Esp_Chine</i>	<i>Esp_Inde</i>
0	1960	50	56
1	1961	51.5	56.7
2	1962	52	57
3	1963	53	57.1
4	1964	53	57.3

Enregistrements :  
des  
dates

Quantitative

Quantitative

Quantitative



# Dans le fleuve Saint-Laurent, l'eau en profondeur s'est-elle réchauffée depuis 30 ans ?

- Enregistrements : des enregistrements de capteurs
- Première variable : le moment de l'enregistrement (valeurs : 6 juin 2002, 3 janvier 2022, etc.)
- Deuxième variable : la température de l'eau (valeurs : en °C)

8

Variables		
Index	<i>dates</i>	<i>température</i>
0	6 juin 2002	7
1	7 juin 2002	7.51
2	8 juin 2002	7.50
3	9 juin 2002	7.49
4	...	7.34
Date		Quantitative

Enregistrements : des résultats des capteurs



Certains disent que le taux de chômage a longtemps été plus élevé au Québec qu'en Ontario, et que maintenant c'est l'inverse, notamment depuis l'arrivée de François Legault à la tête du gouvernement québécois. Est-ce vrai ?

- Enregistrements : des données mensuelles de Statistique Canada
- Première variable : le mois et l'année (valeurs : août 2025, janvier 2023, etc.)
- Deuxième variable : le taux de chômage au Québec (valeurs : 8%, 3%, etc)
- Troisième variable : le taux de chômage en Ontario (valeurs : 8%, 3%, etc)

### Variables

Enregistrements :  
des  
dates

Index	<i>mois</i>	<i>cho_quebec</i>	<i>cho_ontario</i>
0	Septembre 1976	11.4	8.6
1	Octobre 1976	11.5	8.7
2	Novembre 1976	11.4	8.7
3	Décembre 1976	11.5	8.8
4	Janvier 1977	11.6	8.8

Date

Quantitative

Quantitative

# Avantages d'une approche fondée sur les données

- Connaître les phénomènes de façon plus systématique et moins anecdotique
- Appréhender les phénomènes de façon plus objective et moins subjective
- Structurer la recherche d'informations
- Confirmer, contredire ou nuancer les théories ou les paradigmes
- Vérifier les affirmations présentes dans l'espace public

# Ajoutons une certaine profondeur historique à notre démarche

**La démarche scientifique moderne**

**Trois exemples remarquables :**

- **Florence Nightingale**
- **Émile Durkheim**
- **Evarts Ambrose Graham**

## Approche classique, jusqu'au XVIIe siècle

### Phénomène

#### Aristote

Le Mouvement des animaux  
La Locomotion des animaux

Traduction et présentation  
par Pierre-Marie Morel

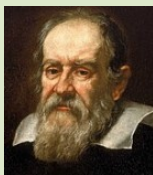


**Pourquoi ce phénomène existe? Le but, la finalité :**  
« la nature ne fait rien en vain, mais toujours en visant pour chaque être le meilleur dans la limite des possibilités, préservant la substance propre, c'est-à-dire l'essence même de chacun » Aristote

## Approche moderne, à partir XVIIe siècle

### Phénomène

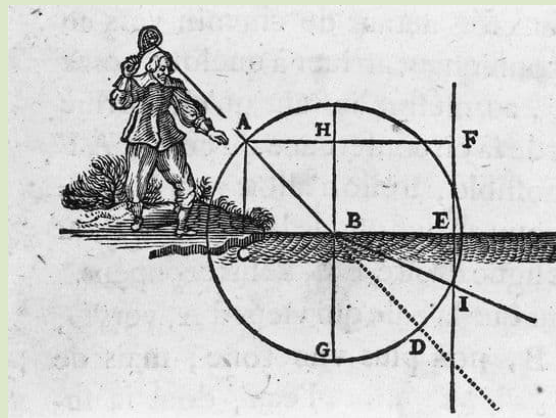
**Comment ce phénomène varie? Quelle en est la cause antérieure?** « Nous ne devons admettre aucune cause des choses naturelles que celles qui sont à la fois véritables et suffisantes pour expliquer leurs phénomènes. » Isaac Newton, Principia (1726)



« La philosophie est écrite dans cet immense livre qui se tient toujours ouvert devant nos yeux, je veux dire l'Univers, mais on ne peut le comprendre si l'on ne s'applique d'abord à en comprendre la langue et à connaître les caractères avec lesquels il est écrit. Il est écrit dans la **langue mathématique** et ses caractères [...] sans le moyen desquels il est humainement impossible d'en comprendre un mot. Sans eux, c'est une errance vaine dans un labyrinthe obscur. » **Galileo Galilei**, Il Saggiatore (1623)



« Il faut examiner chaque chose selon les proportions et les mesures qu'on peut en établir, afin de la ramener à une connaissance exacte. [...] Il n'y a rien dans la nature qui ne puisse être conduit par l'esprit humain vers ce que nous savons déjà, pour autant qu'on en étudie les **rapports** et qu'on les soumette à un raisonnement certain. » **René Descartes**, Règles pour la direction de l'esprit (1628)



## Approche classique, jusqu'au XVIIe siècle

### Phénomène

#### Aristote

Le Mouvement des animaux  
La Locomotion des animaux

Traduction et présentation  
par Pierre-Marie Morel



**Pourquoi ce phénomène existe? Le but, la finalité :**  
« la nature ne fait rien en vain, mais toujours en visant pour chaque être le meilleur dans la limite des possibilités, préservant la substance propre, c'est-à-dire l'essence même de chacun » Aristote

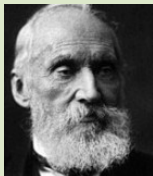
## Approche moderne, à partir XVIIe siècle

### Phénomène

**Comment ce phénomène varie? Quelle en est la cause antérieure?** « Nous ne devons admettre aucune cause des choses naturelles que celles qui sont à la fois véritables et suffisantes pour expliquer leurs phénomènes. » Isaac Newton, Principia (1726)



« L'essence des choses devant nous rester toujours ignorée, nous ne pouvons connaître que les **relations** de ces choses, et les phénomènes ne sont que des résultats de ces relations. » **Claude Bernard**, Introduction à l'étude de la médecine expérimentale (1865)



« Je dis souvent que lorsque l'on peut mesurer ce dont on parle, et l'exprimer en nombres, on sait quelque chose à son sujet ; mais lorsque l'on ne peut pas l'exprimer en nombres, notre connaissance est maigre et insatisfaisante ; cela peut être le début de la connaissance, mais on n'a guère, dans nos pensées, atteint le stade de la science, quel que soit le sujet considéré. » **Lord Kelvin**, (1889)



« Maintenant qu'est-ce que la science ? Je l'ai expliqué ... c'est avant tout une classification, une façon de rapprocher des faits que les apparences séparaient, bien qu'ils fussent liés par quelque parenté naturelle et cachée. **La science, en d'autres termes, est un système de relations.** Or nous venons de le dire, c'est dans les relations seulement que l'objectivité doit être cherchée ; il serait vain de la chercher dans les êtres considérés comme isolés les uns des autres. » **Henri Poincaré**, La valeur de la science (1905)



« La science classique cherchait l'éternel, l'invariable ; la science contemporaine découvre le primat de l'instable, du fluctuant et du variable. » **Ilya Prigogine**, La fin des certitudes (1996)



# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



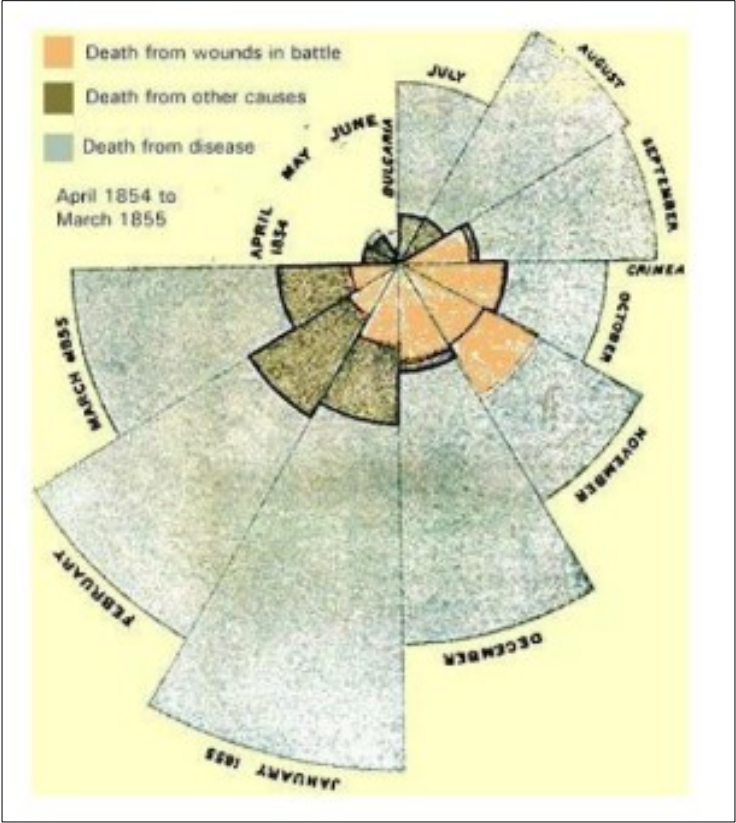
# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



TABLE SHOWING the ESTIMATED AVERAGE MONTHLY STRENGTH of the ARMY ; and the Deaths and Annual Rate of Mortality per 1,000, in each Month, from April, 1854, to March, 1856, (inclusive), in the Hospitals of the Army in the East.

Months	Estimated Average Monthly Strength of the Army.	DEATHS.			ANNUAL RATE OF MORTALITY PER 1,000.		
		Zymotic Diseases.	Wounds and Injuries.	All other Causes.	Zymotic Diseases.	Wounds and Injuries.	All other Causes.
1854 April .. ..	8,571	1	..	5	1.4	..	7.0
May .. ..	23,333	12	..	9	6.2	..	4.6
June .. ..	28,333	11	..	6	4.7	..	2.5
July .. ..	28,722	359	..	23	150.0	..	9.6
August .. ..	30,246	828	1	30	328.5	.4	11.9
September .. ..	30,290	788	81	70	312.2	32.1	27.7
October .. ..	30,643	503	132	128	197.0	51.7	50.1
November .. ..	29,736	844	287	106	340.6	115.8	42.8
December .. ..	32,779	1,725	114	131	631.5	41.7	48.0
1855 January .. ..	32,393	2,761	83	324	1022.8	30.7	120.0
February .. ..	30,919	2,120	42	361	822.8	16.3	140.1
March .. ..	30,107	1,205	32	172	480.3	12.8	68.6
April .. ..	32,252	477	48	57	177.5	17.9	21.2
May .. ..	35,473	508	49	37	171.8	16.6	12.5
June .. ..	38,863	802	209	31	247.6	64.5	9.6
July .. ..	42,647	382	134	33	107.5	37.7	9.3
August .. ..	44,614	483	164	25	129.9	44.1	6.7
September .. ..	47,751	189	276	20	47.5	69.4	5.0
October .. ..	46,852	128	53	18	32.8	13.6	4.6
November .. ..	37,853	178	33	32	56.4	10.5	10.1
December .. ..	43,217	91	18	28	25.3	5.0	7.8
1856 January .. ..	44,212	42	2	48	11.4	.5	13.0
February .. ..	43,485	24	..	19	6.6	..	5.2
March .. ..	46,140	15	..	35	3.9	..	9.1

The Deaths under the head of "Wounds and Injuries," comprise the following causes:—Luxatio, Sub-Luxatio, Vulnus Scelopitorum, Vulnus Incisum, Contusio, Fractura, Ambustio, and Concussio Cerebri.





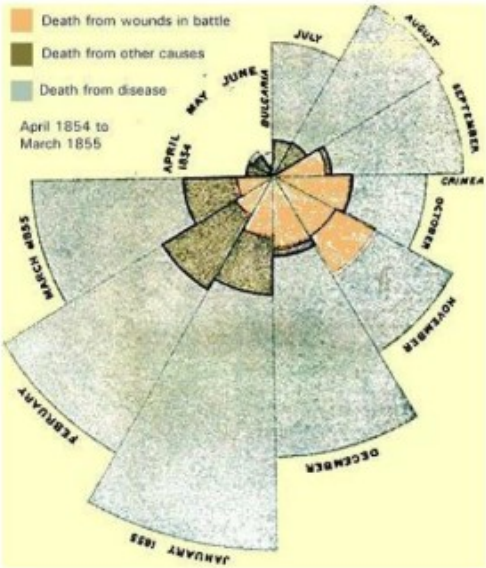
# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



TABLE SHOWING the ESTIMATED AVERAGE MONTHLY STRENGTH of the ARMY; and the Deaths and Annual Rate of Mortality per 1,000, in each Month, from April, 1854, to March, 1856, (inclusive), in the Hospitals of the Army in the East.

Months	Estimated Average Monthly Strength of the Army.	DEATHS.			ANNUAL RATE OF MORTALITY PER 1,000.		
		Zymotic Diseases.	Wounds and Injuries.	All other Causes.	Zymotic Diseases.	Wounds and Injuries.	All other Causes.
1854 April .. ..	8,571	1	..	5	1.4	..	7.0
May .. ..	23,333	12	..	9	6.2	..	4.6
June .. ..	28,333	11	..	6	4.7	..	2.5
July .. ..	28,722	359	..	23	150.0	..	9.6
August .. ..	30,246	828	1	30	328.5	..	11.9
September .. ..	30,290	788	81	70	312.2	32.1	27.7
October .. ..	30,645	503	132	128	197.0	61.7	50.1
November .. ..	29,736	844	287	106	340.6	115.8	42.8
December .. ..	32,779	1,725	114	131	631.5	41.7	48.0
1855 January .. ..	32,393	2,761	83	324	1022.8	30.7	120.0
February .. ..	30,919	2,120	42	361	822.8	16.3	140.1
March .. ..	30,107	1,205	32	172	480.3	12.8	68.6
April .. ..	32,262	477	48	57	177.5	17.9	21.2
May .. ..	35,473	508	49	37	171.8	16.6	12.5
June .. ..	38,863	802	209	31	247.6	64.5	9.6
July .. ..	42,647	382	134	33	107.5	37.7	9.3
August .. ..	44,614	483	164	25	129.9	44.1	6.7
September .. ..	47,751	189	276	20	47.5	69.4	5.0
October .. ..	46,852	128	53	18	32.8	13.6	4.6
November .. ..	37,853	178	33	32	56.4	10.5	10.1
December .. ..	43,217	91	18	28	25.3	5.0	7.8
1856 January .. ..	44,212	42	2	48	11.4	..	13.0
February .. ..	43,485	24	..	19	6.6	..	5.2
March .. ..	46,140	15	..	35	3.9	..	9.1

The Deaths under the head of "Wounds and Injuries," comprise the following causes:—Luxatio, Sub-Luxatio, Vulus Scloporum, Vulus Incisum, Contusio, Fractura, Ambustio, and Concussio Cerebri.



## Septembre 1854

Maladies infectieuses

$$\frac{788 \times 12}{30\,290} \times 1000 = \frac{9\,456}{30\,290} \times 1000 \approx 312,2$$

Tableau de Florence Nightingale (1854-1856) .

Mois	Effectif moyen de l'armée	Décès – Maladies infectieuses	Décès – Blessures et traumatismes	Décès – Autres causes	Taux annuel pour 1 000 – Maladies infectieuses	Taux annuel pour 1 000 – Blessures et traumatismes	Taux annuel pour 1 000 – Autres causes
1854 Avril	8 571	1	..	5	1,4	..	7,0
Mai	23 333	12	..	9	6,2	..	4,6
Juin	28 333	11	..	6	4,7	..	2,5
Juillet	28 722	359	2	23	150,0	0,8	9,6
Août	30 246	828	1	30	328,5	..	11,9
Septembre	30 290	788	81	70	312,2	32,1	27,7

# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



Table 4. Crimean War death rates by hospital †

		No. of Deaths	Hosp Pop. ‡	Cases treated §	Rate/yr of cases treated
Scutari	Je54-Je56	4923	41,325		11.9%
Koulali	Feb-Je55	509	1,963		25.9%
Varna	Je54-Ja55	374	2,846		13.1%
Balaclava	Oc54-Je56	438	5,686		7.7%
Castle	Mar55-Je56	96	2,554		3.8%
Camp	Ap55-Je56	204	1,083		18.8%
Monastery	Jy55-Je56	28	911		3.1%
Abydos	Se54-Se55	82	814		10.1%
Smyrna	Fe55-No56	154	1,887		8.2%
Renkioi	Oc55-Je56	50	1,330		3.8%

† Nightingale (1859b) 25.  
‡ mean of weekly numbers remaining in hospital  
§ mean of admissions and discharges, including deaths

Tableau de mortalité comparée (Florence Nightingale, Guerre de Crimée)

Rang	Hôpital / Lieu	Taux de mortalité (%)
1	Koulali (févr. 1855 – juin 1855)	25,9 %
2	Camp (avr. 1855 – juin 1856)	18,8 %
3	Varna (juin 1854 – janv. 1855)	13,1 %
4	Scutari (juin 1854 – juin 1856)	11,9 %
5	Abydos (sept. 1854 – sept. 1855)	10,1 %
6	Smyrna (févr. 1855 – nov. 1856)	8,2 %
7	Balaclava (oct. 1854 – juin 1856)	7,7 %
8	Renkioi (oct. 1855 – juin 1856)	3,8 %
9	Castle (mars 1855 – juin 1856)	3,8 %
10	Monastery (juil. 1855 – juin 1856)	3,1 %

- Problèmes d'hygiène :
  - Manque de propreté
  - Insalubrité des locaux
  - Mauvaise gestion des déchets
- Problèmes de ventilation :
  - Air stagnant favorisant la propagation des maladies
- Problèmes de lumière :
  - Manque de lumière naturelle pour les patients
- Problèmes nutritionnels :
  - Alimentation inadéquate pour les patients
- Problèmes d'infrastructures :
  - Hôpitaux surpeuplés et mal équipés
- Problèmes de soins :
  - Manque de formation des infirmières
  - Soins dispensés de manière désorganisée

# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



Table 4. Crimean War death rates by hospital †

		No. of Deaths	Hosp Pop.	‡Cases treated	§Rate/yr of cases treated
Scutari	Je54-Je56	4923		41,325	11.9%
Koulali	Feb-Je55	509		1,963	25.9%
Varna	Je54-Ja55	374		2,846	13.1%
Balaclava	Oc54-Je56	438		5,686	7.7%
Castle	Mar55-Je56	96		2,554	3.8%
Camp	Ap55-Je56	204		1,083	18.8%
Monastery	Jy55-Je56	28		911	3.1%
Abydos	Se54-Se55	82		814	10.1%
Smyrna	Fe55-No56	154		1,887	8.2%
Renkioi	Oc55-Je56	50		1,330	3.8%

† Nightingale (1859b) 25.  
‡ mean of weekly numbers remaining in hospital  
§ mean of admissions and discharges, including deaths

Tableau de mortalité comparée (Florence Nightingale, Guerre de Crimée)

Rang	Hôpital / Lieu	Taux de mortalité (%)	Remarques sur les conditions
1	Koulali (févr. 1855 – juin 1855)	25,9 %	Très mauvaises conditions sanitaires
2	Camp (avr. 1855 – juin 1856)	18,8 %	Huttes insalubres, semi-enterrées
3	Varna (juin 1854 – janv. 1855)	13,1 %	Épidémies (choléra), pas d’infirmières femmes
4	Scutari (juin 1854 – juin 1856)	11,9 %	Principal hôpital militaire, amélioré après réformes
5	Abydos (sept. 1854 – sept. 1855)	10,1 %	Hôpital civil
6	Smyrna (févr. 1855 – nov. 1856)	8,2 %	Hôpital civil
7	Balaclava (oct. 1854 – juin 1856)	7,7 %	Bon bâtiment en pierre, bien situé
8	Renkioi (oct. 1855 – juin 1856)	3,8 %	Hôpital préfabriqué, très bonnes conditions
9	Castle (mars 1855 – juin 1856)	3,8 %	Conditions excellentes, altitude favorable
10	Monastery (juil. 1855 – juin 1856)	3,1 %	Hôpital de convalescence, conditions optimales

- **Problèmes d'hygiène :**
  - Manque de propreté
  - Insalubrité des locaux
  - Mauvaise gestion des déchets
- **Problèmes de ventilation :**
  - Air stagnant favorisant la propagation des maladies
- **Problèmes de lumière :**
  - Manque de lumière naturelle pour les patients
- **Problèmes nutritionnels :**
  - Alimentation inadéquate pour les patients
- **Problèmes d'infrastructures :**
  - Hôpitaux surpeuplés et mal équipés
- **Problèmes de soins :**
  - Manque de formation des infirmières
  - Soins dispensés de manière désorganisée

# Florence Nightingale 1820-1910 infirmière et statisticienne anglaise



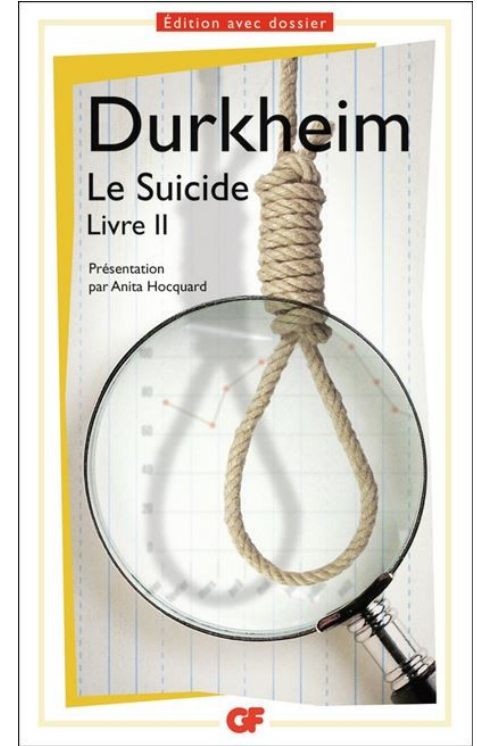
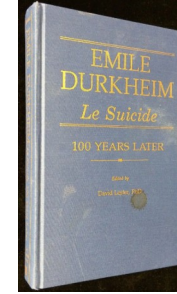
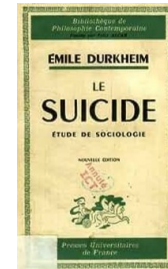
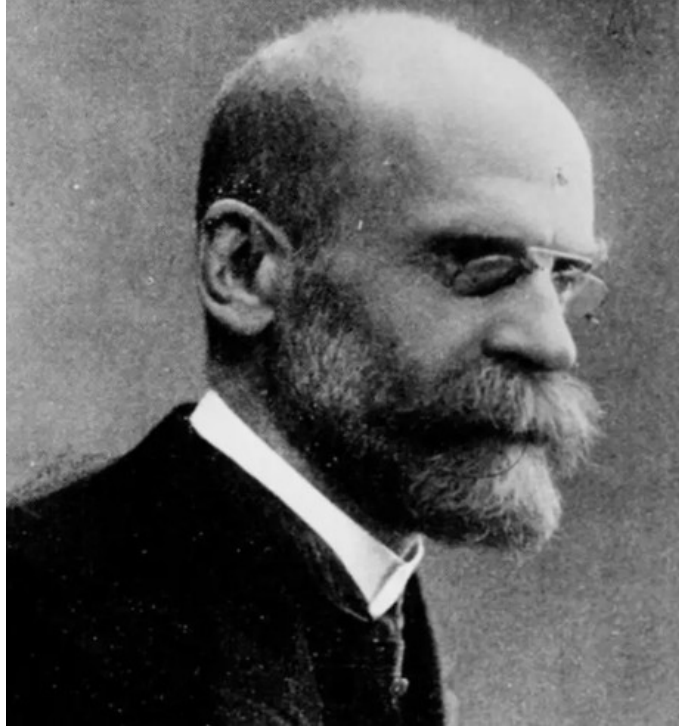
**Conditions  
d'hospitalisation**



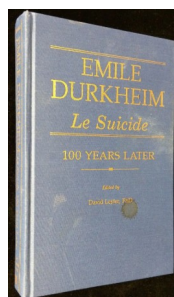
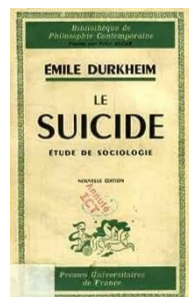
**Mortalité**



# Émile Durkheim 1858-1917 sociologue français



# Émile Durkheim 1858-1917 sociologue français



TABEAU XXIV

Suicides pour 1 million de sujets de chaque profession.

	COMMERCE.	TRANS- PORTS.	INDUSTRIE.	AGRI- CULTURE.	CARRIÈRES libérales (1).
France (1878-87) (2).....	440		340	240	300
Suisse (1876).....	664	1514	577	304	558
Italie (1866-76)...	277	152,6	80,4	26,7	618 (3)
Prusse (1883-90)...	754		456	315	832
Bavière (1884-91)...	465		369	153	454
Belgique (1886-90).....	421		160	160	100
Wurtemberg (1873-78) .....	273		190	206	
Saxe (1878).....		341,59		71,17	

Les fonctions industrielles et commerciales sont, en effet, parmi les professions qui fournissent le plus au suicide (V. Tableau XXIV, p. 286). Elles sont presque au niveau des carrières libérales, parfois même elles le dépassent; surtout, elles sont sensiblement plus éprouvées que l'agriculture. C'est que l'industrie agricole est celle où les anciens pouvoirs régulateurs font encore le mieux sentir leur influence et où la fièvre des affaires a le moins pénétré. C'est elle qui rappelle le mieux ce qu'était autrefois la constitution générale de l'ordre économique. Et encore l'écart serait-il plus marqué si, parmi les suicides de l'industrie, on distinguait les patrons des ouvriers, car ce sont probablement les premiers qui sont le plus atteints par l'état d'*anomie*. Le taux énorme de la population rentière (720 pour un million) montre assez que ce sont les plus fortunés qui souffrent le plus. C'est que tout ce qui oblige à la subordination atténue les effets de cet état. Les classes inférieures ont du moins leur horizon limité par celles qui leur sont superposées et, par cela même, leurs désirs sont plus définis. Mais ceux qui n'ont plus que le vide au-dessus d'eux, sont presque nécessités à s'y perdre, s'il n'est pas de force qui les retienne en arrière.

## 1. Recueillir les données de base

- Nombre de suicides annuels enregistrés (dans les statistiques criminelles ou de police)
- Répartition par profession (commerce, agriculture, industrie, transports, professions libérales, etc.)
- Effectifs de population dans chaque profession (issus des recensements).

## 2. Calculer un taux brut

Pour chaque catégorie professionnelle dans chaque pays :

$$\text{Taux de suicides} = \frac{\text{Nombre de suicides dans la profession}}{\text{Population totale de cette profession}}$$

## 3. Standardiser pour comparer

Comme les effectifs varient d'une profession ou d'un pays à l'autre, Durkheim exprime le résultat pour 1 million de personnes (un taux standardisé).

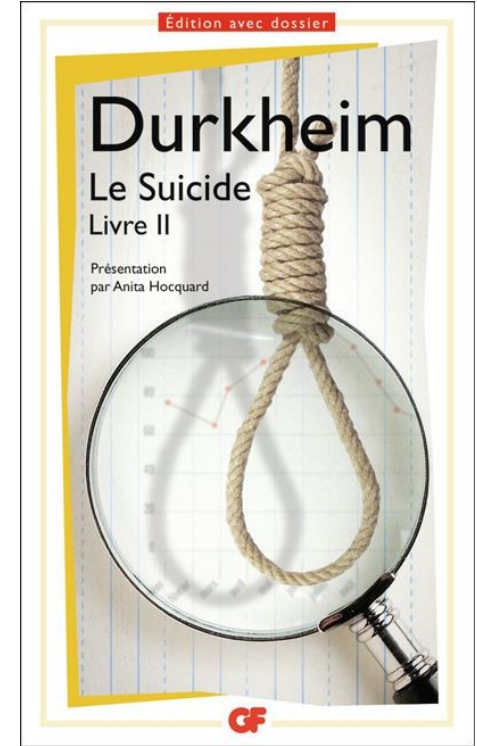
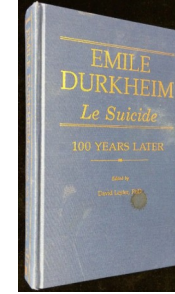
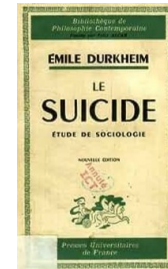
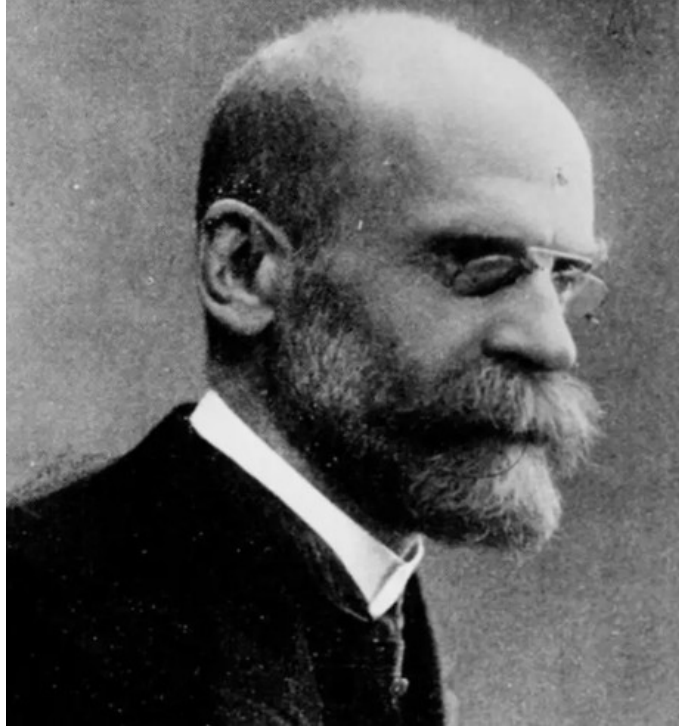
$$\text{Taux pour 1 million} = \frac{\text{Nombre de suicides}}{\text{Population de la profession}} \times 1\,000\,000$$

Exemple (chiffres fictifs) :

- 150 suicides parmi 500 000 commerçants
- $\frac{150}{500\,000} \times 1\,000\,000 = 300$  suicides par million.



# Émile Durkheim 1858-1917 sociologue français



Professions



Suicide

Evarts Ambrose Graham 1883-1957 médecin et chercheur

...ent men in town. While his  
dose may say *Office Hours*  
2 to 4, he's actually on call  
24 hours a day.

The doctor is a scientist,  
a diplomat, and a friendly  
sympathetic human being  
all in one, no matter how  
long and hard his schedule.

According to a recent Nationwide survey:

**MORE DOCTORS SMOKE CAMELS**  
**THAN ANY OTHER CIGARETTE**

DOCTORS in every branch of medicine—113,997  
in all—were queried in this nationwide study  
of cigarette preference. Three leading research or-  
ganizations made the survey. The gist of the query  
was—What cigarette do you smoke, Doctor?  
*The brand named most was Camels!*

The rich, full flavor and cool mildness of Camel's  
especial blend of costlier tobaccos seem to have the  
same appeal to the smoking tastes of doctors as to  
millions of other smokers. If you are a Camel  
smoker, this preference among doctors will hardly  
surprise you. If you're not—well, try Camels now.

Your "T-Zone" Will Tell You...

T for Taste...  
T for Throat...  
that's your  
proving ground  
for any cigarette.  
See if Camels  
don't suit your  
"T-Zone" to a "T."

**CAMELS** Costlier Tobaccos



The Journal of the  
American Medical Association

Published Under the Auspices of the Board of Trustees

VOL. 143, No. 4 CHICAGO, ILLINOIS MAY 27, 1950  
COPYRIGHT, 1950, BY AMERICAN MEDICAL ASSOCIATION

**TOBACCO SMOKING AS A POSSIBLE ETIOLOGIC  
FACTOR IN BRONCHIOGENIC CARCINOMA**

A Study of Six Hundred and Eighty-Four Proved Cases

ERNEST L. WYNDER  
and  
EVARTS A. GRAHAM, M.D.  
St. Louis

*General Increase.*—There is rather general agreement  
that the incidence of bronchiogenic carcinoma has  
greatly increased in the last half-century. Statistical  
studies at the Charity Hospital of New Orleans  
(Ochsner and DeBakey),<sup>1</sup> the St. Louis City Hospital  
(Wheeler)<sup>2</sup> and the Veterans Administration Hospital  
of Hines, Ill. (Avery)<sup>3</sup> have revealed that at these  
hospitals cancer of the lung is now the most frequent  
visceral cancer in men.

a few of the workers who thought that there was some  
evidence that tobacco was an important factor in the  
increase of cancer of the lungs. Müller<sup>14</sup> in 1939,  
from a careful but limited clinical statistical study,  
offered good evidence that heavy smoking is an impor-  
tant etiologic factor. In 1941 Ochsner and DeBakey<sup>15</sup>  
called attention to the similarity of the curve of increased  
sales of cigarets in this country to the greater prevalence  
of primary cancer of the lung. They emphasized the  
possible etiologic relationship of cigaret smoking to  
this condition. In a recent paper Schrek<sup>16</sup> conclude<sup>d</sup>  
that there is strong circumstantial evidence that cigar  
smoking is an etiologic factor in cancer of the respira-  
tory tract and finds that his data are in agreement with  
the results of a preliminary report presented by Wynder  
and Graham at the National Cancer Conference in  
February 1949.<sup>17</sup>

*Purpose of Study.*—The purpose of the present study  
was to attempt to determine, so far as possible by clin-

# The Journal of the American Medical Association

Published Under the Auspices of the Board of Trustees

VOL. 143, No. 4

CHICAGO, ILLINOIS  
COPYRIGHT, 1950, BY AMERICAN MEDICAL ASSOCIATION

MAY 27, 1950

## TOBACCO SMOKING AS A POSSIBLE ETIOLOGIC FACTOR IN BRONCHIOGENIC CARCINOMA

A Study of Six Hundred and Eighty-Four Proved Cases

ERNEST L. WYNDER  
and  
EVARTS A. GRAHAM, M.D.  
St. Louis

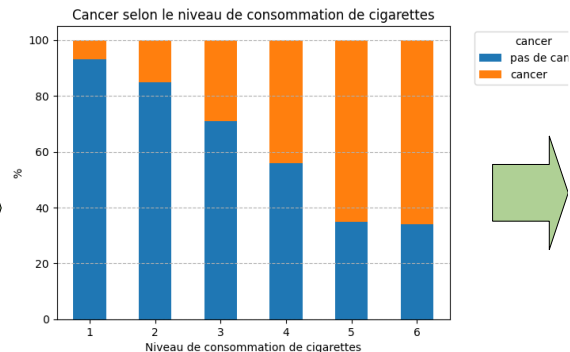
*General Increase.*—There is rather general agreement that the incidence of bronchiogenic carcinoma has greatly increased in the last half-century. Statistical studies at the Charity Hospital of New Orleans (Ochsner and DeBakey),<sup>1</sup> the St. Louis City Hospital (Wheeler)<sup>2</sup> and the Veterans Administration Hospital of Hines, Ill. (Avery)<sup>3</sup> have revealed that at these hospitals cancer of the lung is now the most frequent visceral cancer in men.

a few of the workers who thought that there was some evidence that tobacco was an important factor in the increase of cancer of the lungs. Müller<sup>14</sup> in 1939, from a careful but limited clinical statistical study, offered good evidence that heavy smoking is an important etiologic factor. In 1941 Ochsner and DeBakey<sup>15</sup> called attention to the similarity of the curve of increased sales of cigarets in this country to the greater prevalence of primary cancer of the lung. They emphasized the possible etiologic relationship of cigaret smoking to this condition. In a recent paper Schrek<sup>16</sup> concluded that there is strong circumstantial evidence that cigaret smoking is an etiologic factor in cancer of the respiratory tract and finds that his data are in agreement with the results of a preliminary report presented by Wynder and Graham at the National Cancer Conference in February 1949.<sup>17</sup>

*Purpose of Study.*—The purpose of the present study was to attempt to determine, so far as possible by clinical

cancer	cigarettes	cancer	0	1	Total
0	1	4			
1	0	4			
2	0	3			
3	0	4			
4	0	6			
5	0	6			
6	1	4			
7	0	4			
8	1	4			
9	1	5			
Total		780	605	1385	

cancer	0	1
1	93.0	7.0
2	85.0	15.0
3	71.0	29.0
4	56.0	44.0
5	35.0	65.0
6	34.0	66.0



**TABLE 18.33** Data for Exercise 18.30. Case-control data on long-term smoking habit and lung cancer.

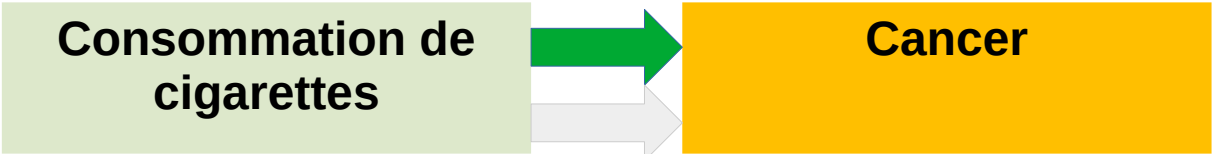
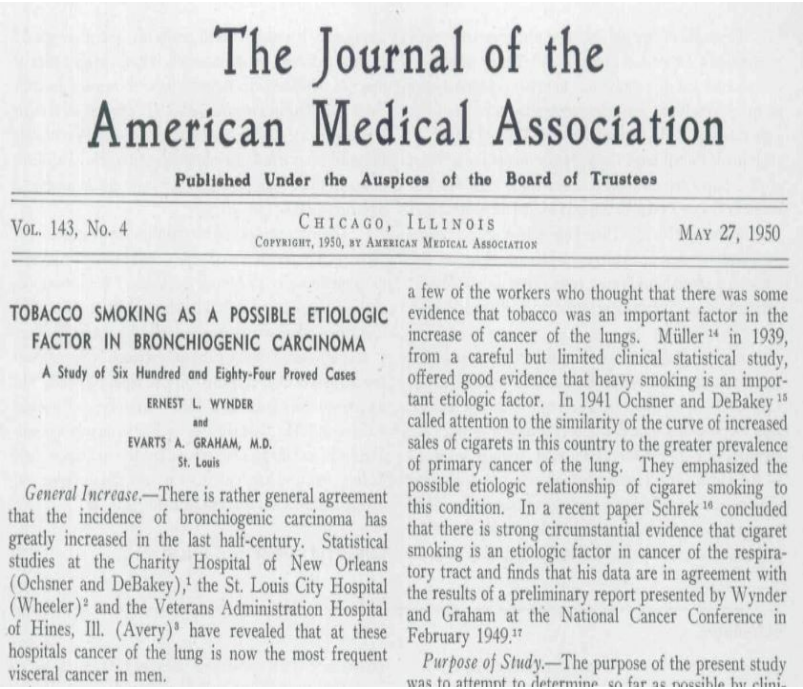
Smoking level*	Cases	Noncases
1 Nonsmoker (<1 cigarette per day)	8	115
2 Light smoker (1–9 cigs per day)	14	82
3 Moderate (10–15 cigs per day)	61	147
4 Heavy (16–20 cigs per day)	213	274
5 Excessive (21–34 cigs per day)	186	98
6 Chain (35 or more cigs per day)	123	64
Total	605	780

\*If subject smoked for less than 20 years, the amount of smoking was reduced in proportion to duration.

Data from Wynder, E. L., & Graham, E. A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *JAMA*, 143 (1)(4), 329–336.



**Evarts Ambrose Graham 1883-1957** médecin et chercheur





Les données ouvertes: pourquoi et comment avec Python, Pandas et Google Colab



Jean-H...  
1 abonné

Données analytiques

Modifier la vidéo



Partager

Promouvoir

