

Le vrai du “Fake”

21 novembre 2024

Les bonnes
pratiques
de la
Statistique





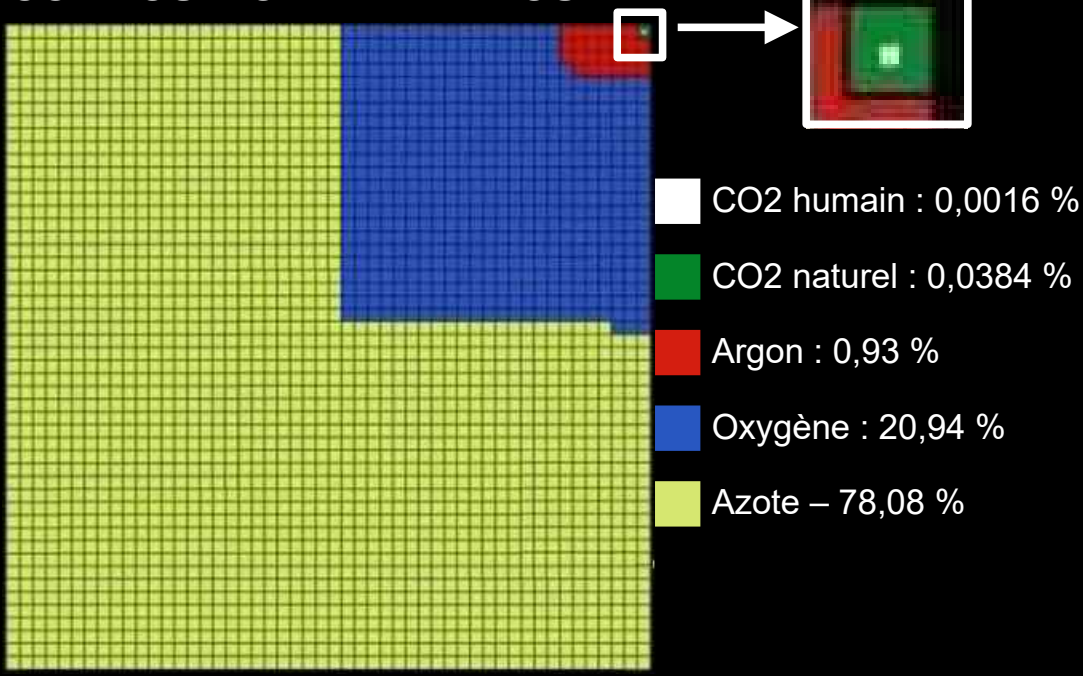
Un exemple de mésinformation*

***mauvaise information, qualitativement imparfaite en raison d'erreurs** de différentes natures (précipitation et absence de vérification, superficialité de traitement et incomplétude, non-actualisation de contenus et obsolescence, faux pour faire rire).

 Le CO2 émis par les activités humaines représente 0,0016 % de l'atmosphère

 Et on veut nous faire croire que c'est ce qui fait augmenter la température sur Terre de 1 °C

COMPOSITION DE L'ATMOSPHERE



Information juste

Provenant du Groupe d'experts inter-gouvernemental sur l'évolution du climat



L'interprétation est fausse

Didier Hauglustaine, climatologue au CNRS
Cathy Clerbaux, physicienne

"Il est erroné de comparer le CO2 à des gaz qui sont inertes."

"Sans gaz à effet de serre, il ferait -18°C sur Terre au lieu de +15°C en moyenne."



Un exemple de désinformation*

* processus de communication qui consiste à utiliser les médias de masse, dont les réseaux sociaux, pour **transmettre des informations erronées dans le but de tromper ou d'influencer** l'opinion publique et de l'amener à agir dans une certaine direction (motivations idéologiques, financières ou ludiques).

**Idée très répandue
chez les Français :**

**Le passage à l'Euro
a fait bondir les prix.**

Notamment, parce qu'à sa mise en place, les commerçants en auraient profité pour augmenter les prix, en particulier les arrondir à la hausse.

Exemple :
2,5 Francs = 0,38 €
arrondi à 0,40 €
soit une augmentation de 5 %

Information fausse



Pour prévenir une potentielle augmentation des prix :

Beaucoup de pièces de 1 et 2 centimes avaient par exemple été mises en circulation pour éviter que les commerçants arrondissent les prix à l'euro supérieur.



AFP Factuel : la cellule de fact-checking de l'Agence France Presse (AFP).

Décodeurs du Monde : les journalistes du journal Le Monde traitent des rumeurs et des intox qui circulent sur la Toile pour y distinguer le vrai du faux.

Citizen Evidence : Outil d'Amesty International qui permet de vérifier les différentes utilisations d'une vidéo sur Youtube.

HoaxBuster : plateforme collaborative contre la désinformation : son site internet permet d'identifier les hoax sur la Toile c'est-à-dire les canulars.

Fake Off de 20 minutes : la rédaction de 20 minutes met à votre disposition sa rubrique de fact-checking "fake off" pour lutter contre les fake news.

Matinale ROSS

21 novembre 2024

Sommaire



Partie 1 : Comment collecter les données ?

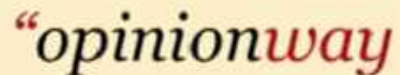
Partie 2 : Comment bien transcrire la réalité ?

Partie 1 : Comment collecter les données ?





1.1 Insee vs instituts de sondage : Il y a sondage et sondage...



Fondés sur **un ensemble de méthodes, basées sur les mathématiques**
= un ensemble de méthodes, prouvées par les mathématiques

...pour tirer un échantillon, traiter les non-réponses, faire des redressements,
calculer la précision des résultats (intervalle de confiance)

**Tout individu de la population étudiée doit avoir la même probabilité
d'être tiré au sort**

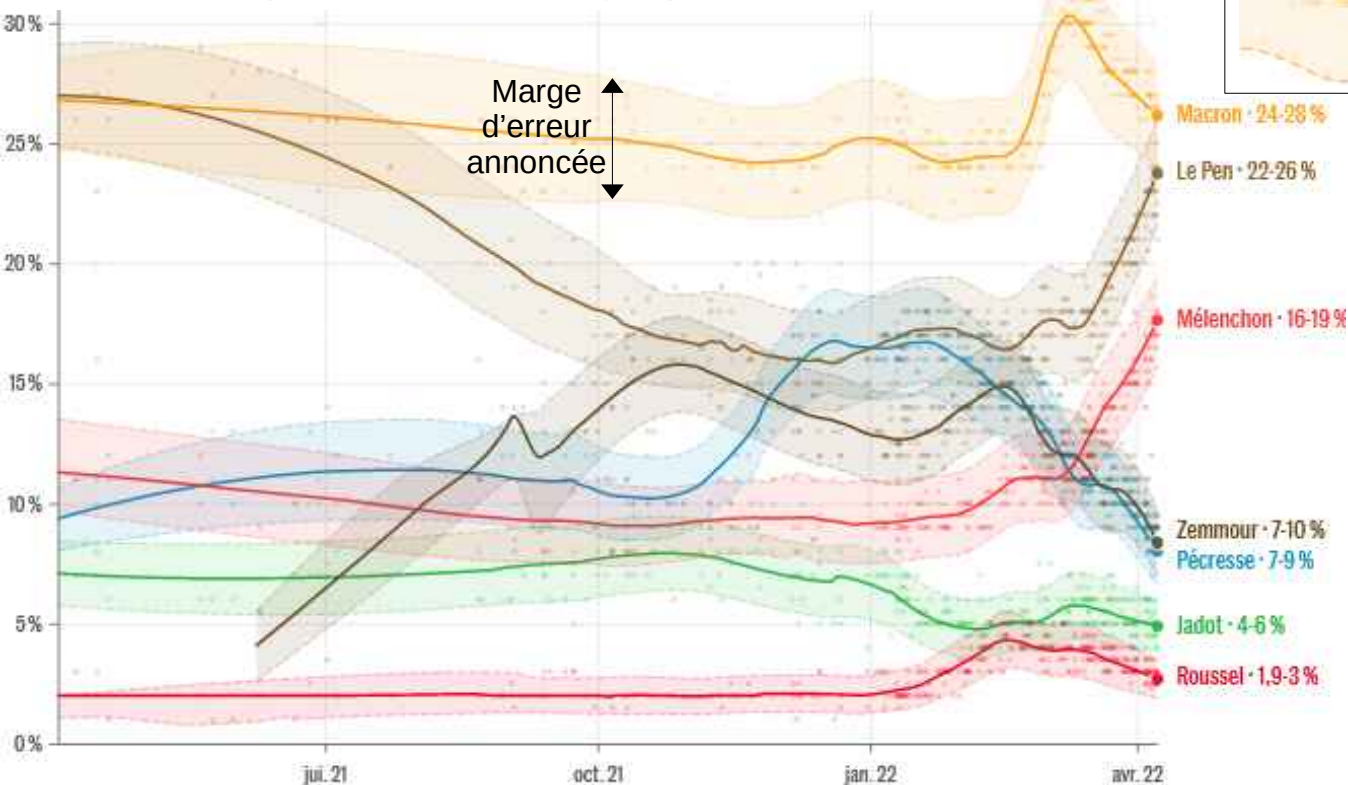
Or, seul l'Insee :

- dispose d'une base de toute la population française
- utilise des tailles d'échantillon conséquentes :
 - 80 000 logements pour enquête emploi
 - 16 000 pour enquête condition de vie
- à ses enquêtes rendues obligatoires par la législation

- Ne décrivent pas totalement leurs méthodes (secret des affaires)
 - Des biais qui ne peuvent pas être totalement redressés
 - Souvent les bases de sondages sont des volontaires :
chaque Français n'a donc pas les mêmes chances d'être enquêté
 - Enquêtes souvent en ligne : et les personnes qui n'ont pas internet ?
 - Pas de cadre probabiliste ou de méthode permettant de mesurer la qualité de réponse (ou d'erreur)
- => Des méthodes “empiriques”,
basées sur leurs expériences de l'exercice.**

Le Monde

Synthèse des enquêtes d'opinion
du 1er tour des présidentielles 2022 jusqu'au 8 avril



Résultats du
1er Tour

27,84 %

23,15 %

21,95 %

7,07 %

4,78 %

4,63 %

2,28 %

Alphabet



1.2 Insee vs Google : deux modèles de collecte et d'usage des données opposés

Insee vs. Google

Un traitement de l'information et des objectifs différents...



exploitation **commerciale** de l'information au **niveau individuel**

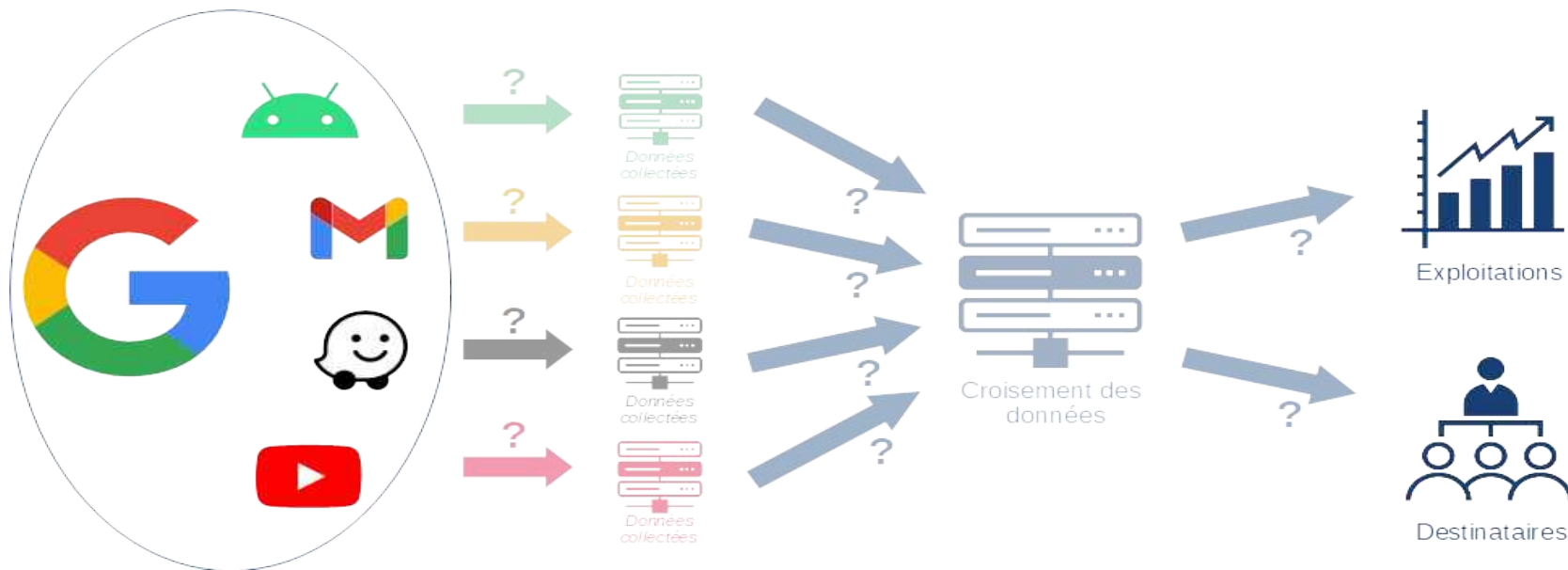


notre objet social : recueil de données individuelles
anonymisées par la suite pour produire des analyses sur des
données agrégées

...auquel s'ajoute un positionnement opposé en
matière de **transparence**.

Que sait-on de ce que Google sait sur les Français ? **Pas grand-chose...**

Une masse impressionnante de données numériques collectées



Que sait-on de ce que l'Insee sait sur les Français ? Une transparence encadré par le législateur...

Collecte de données
si utilité publique



Conseil national
de l'information statistique

Dossier respect des lois
informatiques et libertés



COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Protocole d'enquête
documenté



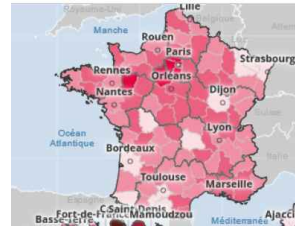
Communication
préalable



Organisme de
contrôle



Résultats et analyses diffusés



Les avantages des “traces numériques”



Avantages



Disponibilité instantanée



Homogénéité



Coût marginal



Charge de collecte

Les **limites** des “traces numériques”

=> Un problème de représentativité

⇒ 1 Français de plus de 15 ans sur 6 n'utilise pas Internet

⇒ Parmi les personnes âgées, plus d'1 sur 2 manquent de connaissances informatiques (illectronisme)

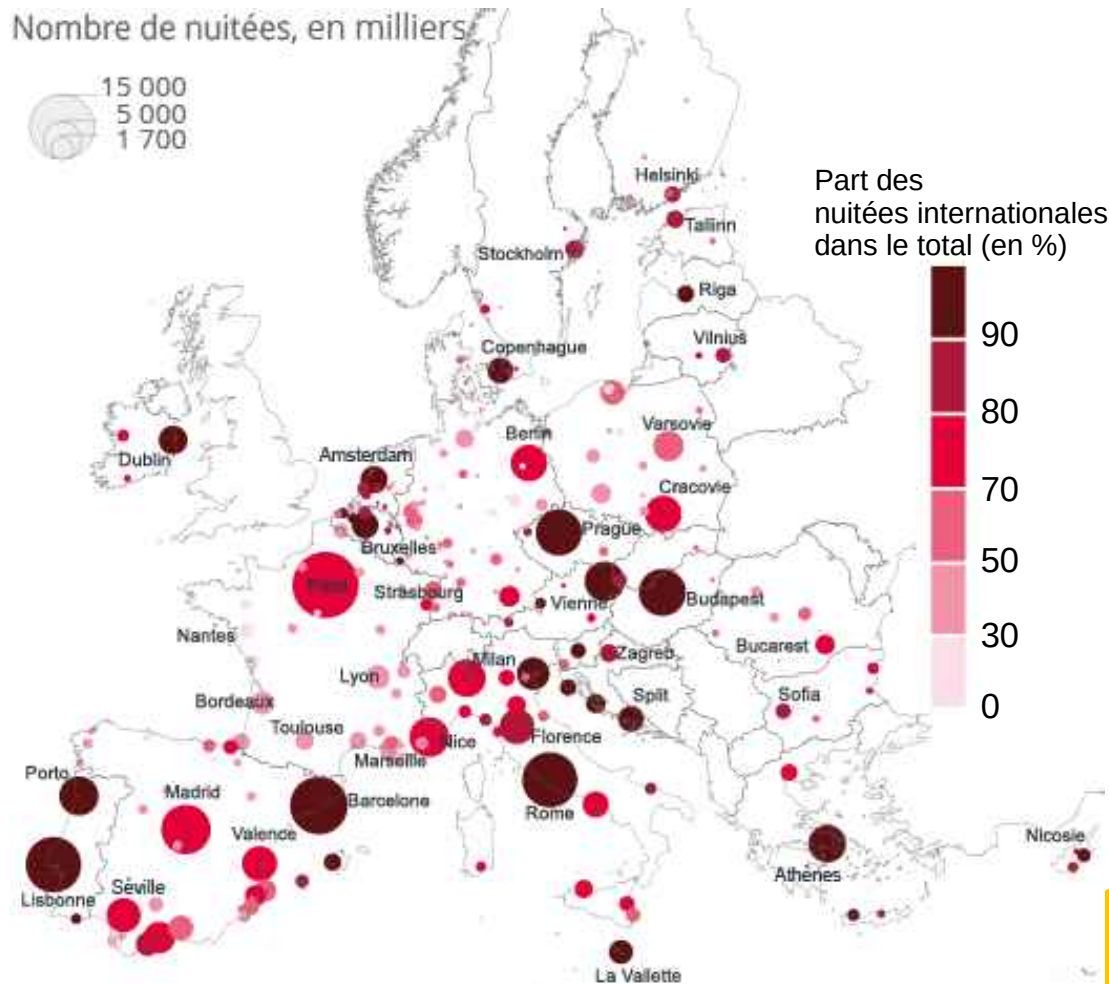
⇒ et les utilisateurs de Google ne sont pas ceux de Qwant, de Bing, etc

Néanmoins,
l'Insee utilise des “traces
numériques” pour compléter ses
statistiques sur le Tourisme

=> L'exemple des réservations
touristiques en ligne
(Airbnb, Booking, Expedia Group et Tripadvisor)

Lecture : à Nice, en 2019, la fréquentation des hébergements touristiques proposés par des particuliers via les principales plateformes internationales représente 5 350 000 nuitées. Parmi celles-ci, 73 % sont réalisées par des non-résidents.

Source : IGN – Eurostat 2021.



Partie 2 : Comment bien transcrire la réalité ?



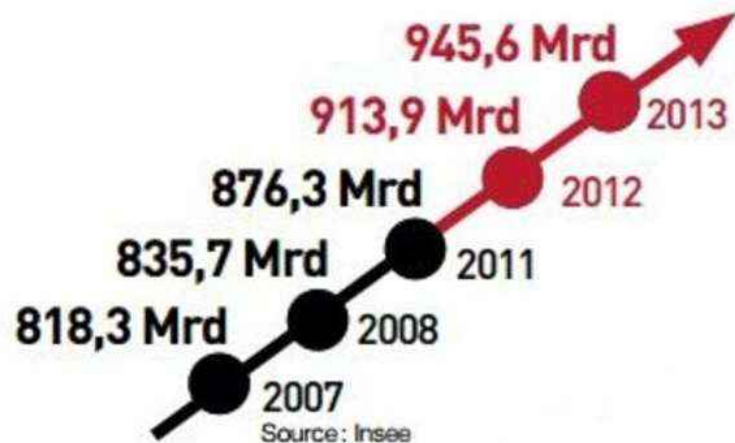
Attention aux manipulations



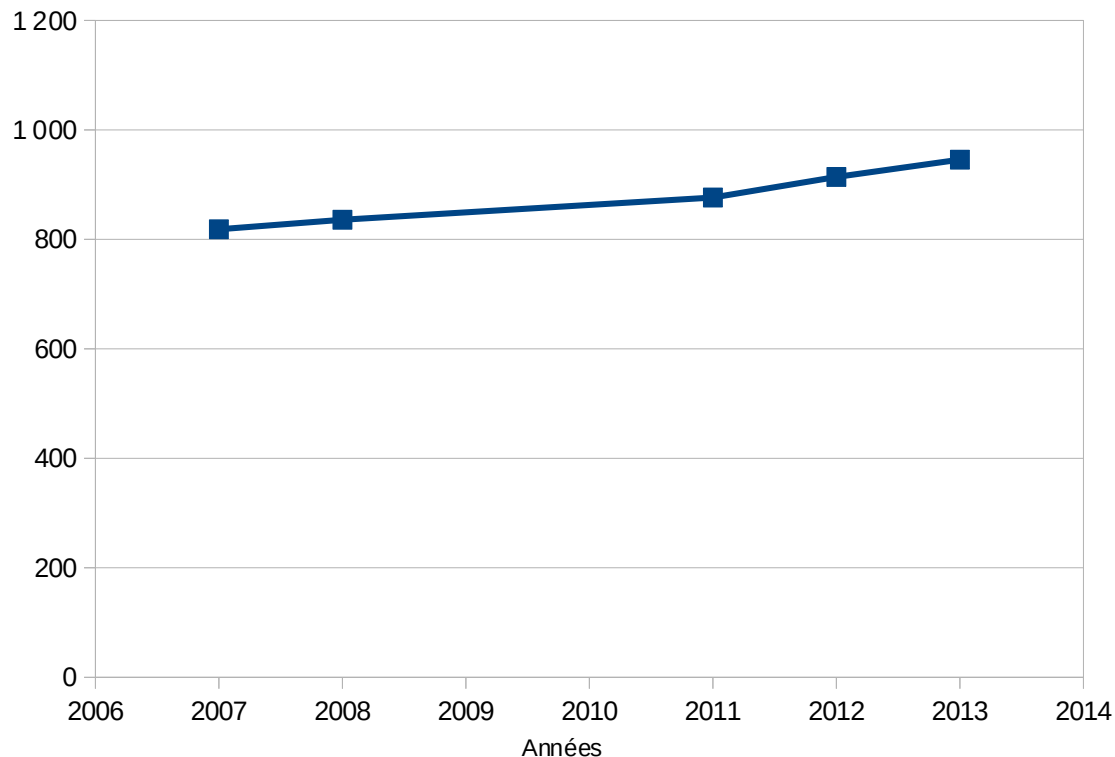
[TAXES ET IMPÔTS]

84 nouveaux impôts
en 2 ans !

(2011-2013)



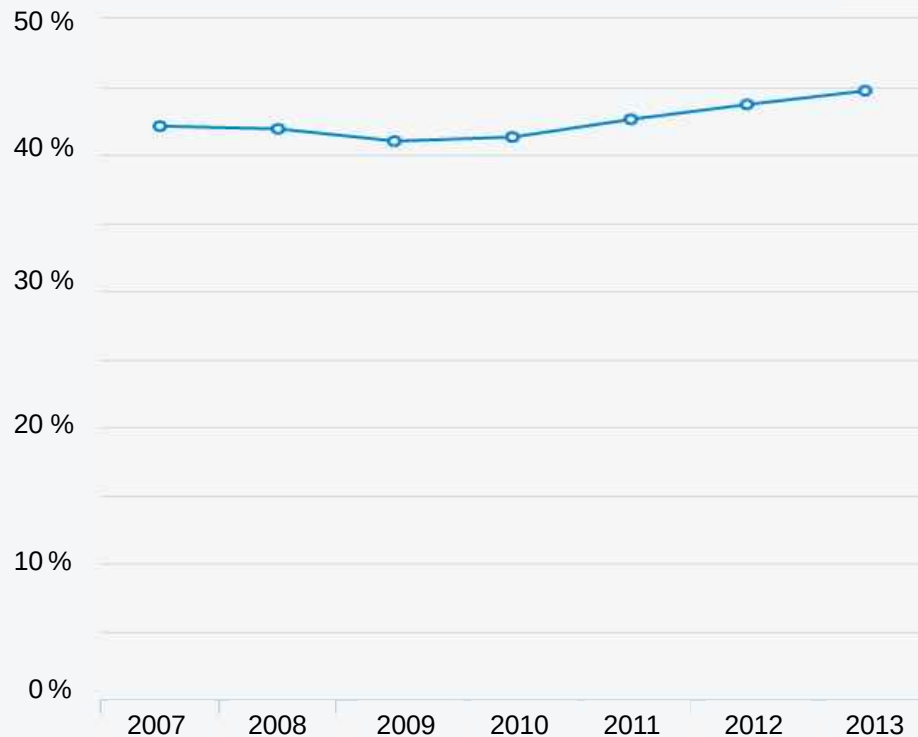
Milliards d'euros





Les prélèvements obligatoires à un niveau record en 2013

Evolution des prélèvements obligatoires depuis 2007
en part du produit intérieur brut (PIB)



Source : Insee

Poids des prélèvements obligatoires

En % du pib



Note : hors cotisations sociales imputées et crédit d'impôts.

Lecture : en 2021, le taux de prélèvements obligatoires s'établit à 44,3 %.

Champ : France.

Source : Insee, comptes nationaux - base 2014.



Transformez une baisse en hausse...

Ventes cumulées d'iPhone

Nombre d'iPhone vendus

Cumulative iPhone sales

Cumulative iPhone sales

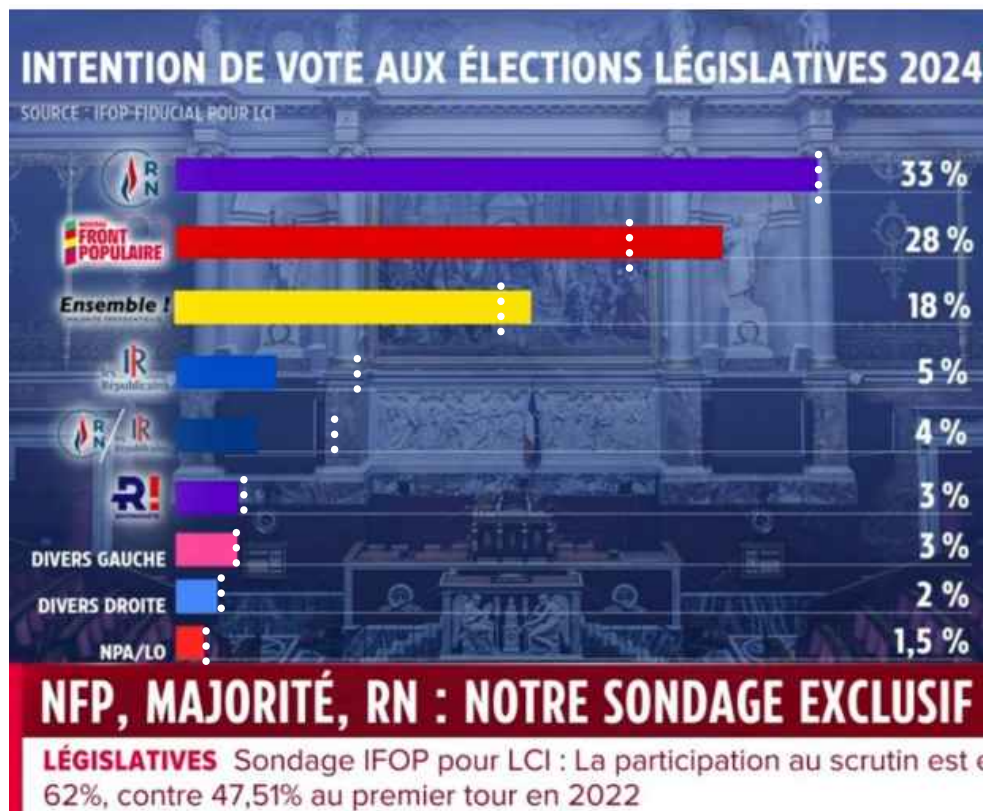
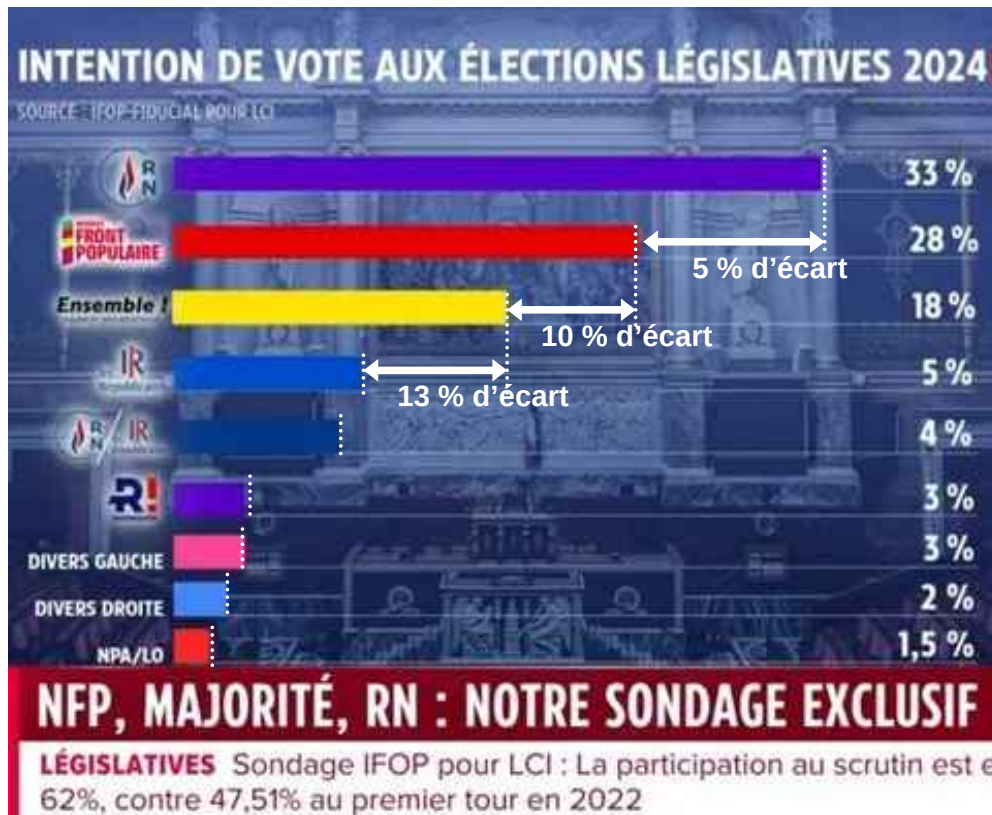
Ventes trimestrielles

Attention aux représentations

Des exemples de graphiques trompeurs



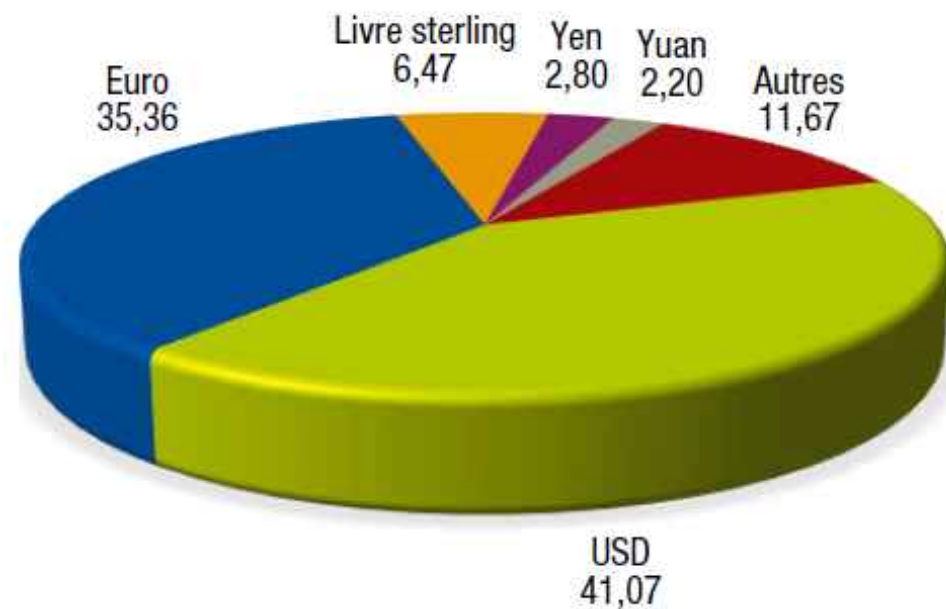
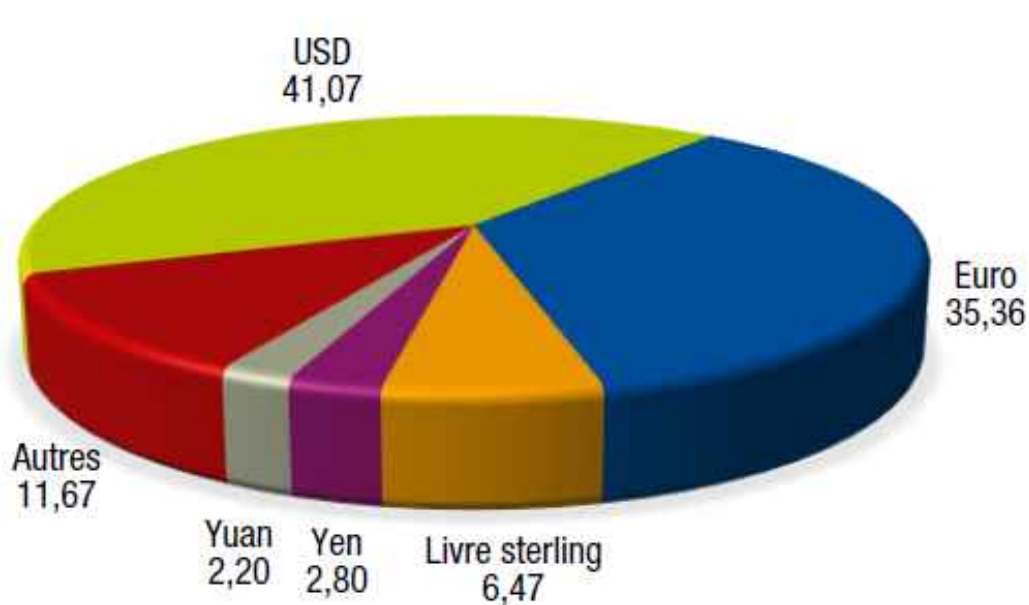
Graphique corrigé :



Des exemples de graphiques trompeurs



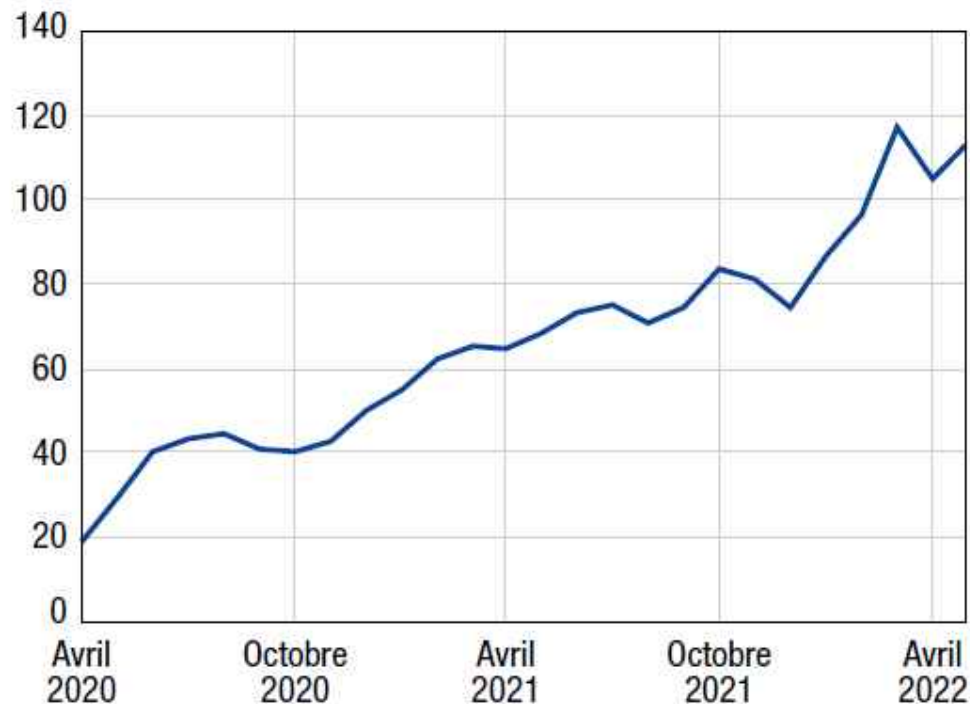
Part des différentes devises dans les paiements internationaux (en %)



Source : SWIFT

Attention à la simplification

Evolution du prix du pétrole
(en dollars américains par baril)



Limites des outils statistiques : les “pouvoirs limités” de la moyenne



Les iris (1889)
Van Gogh



Sa moyenne
colorimétrique

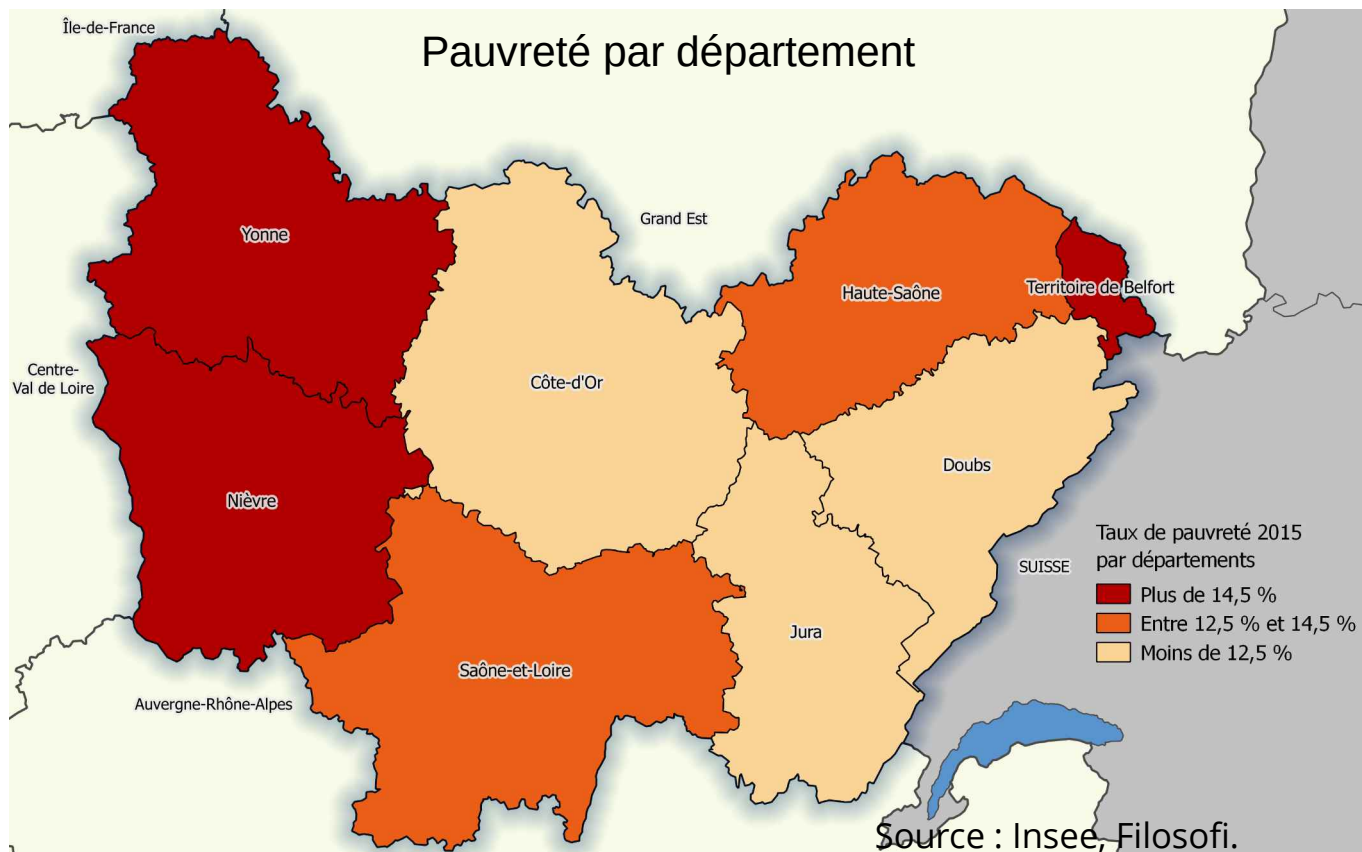
Comment la moyenne peut
nous dire que tout va bien :



Limites des outils statistiques :



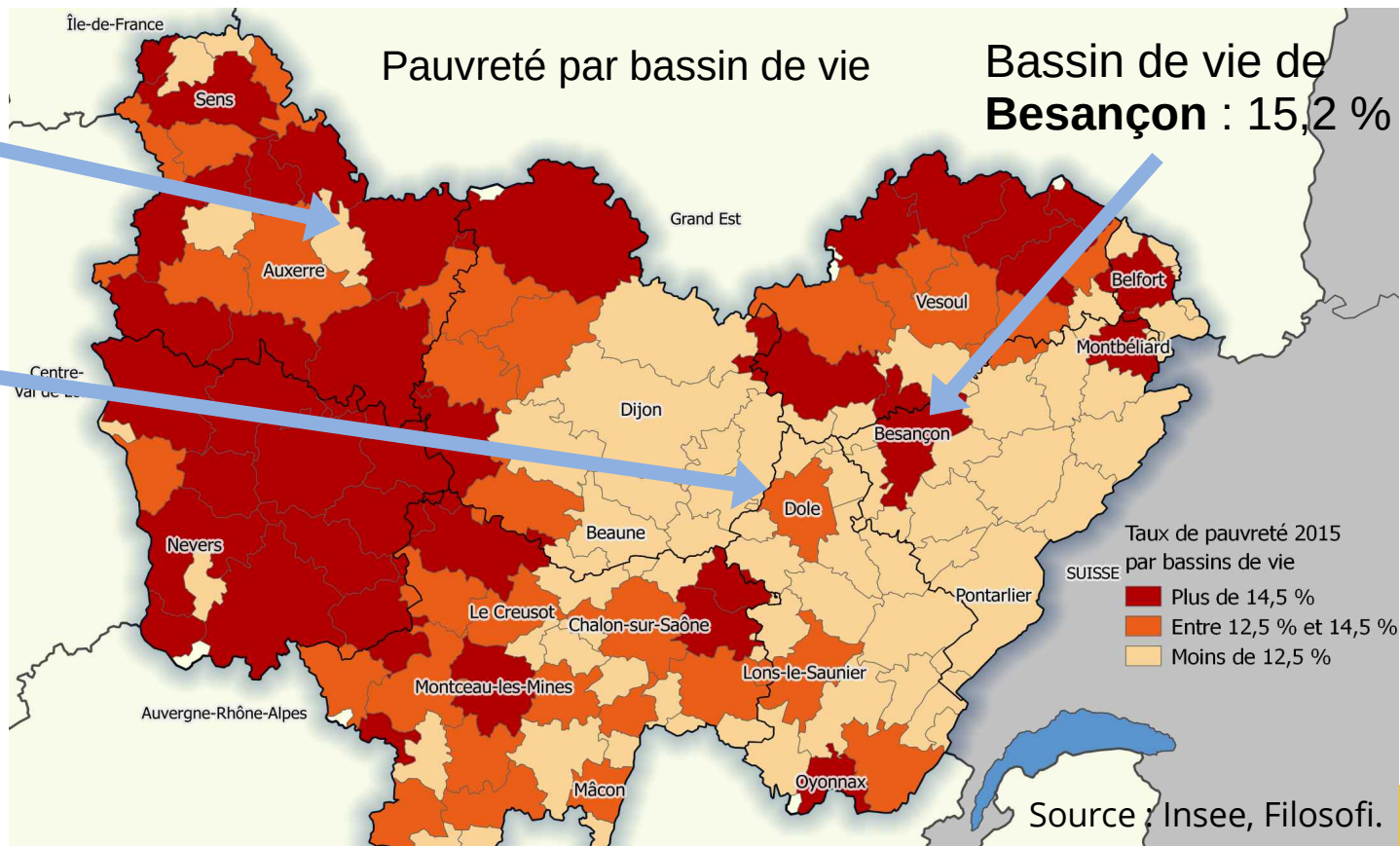
Quelles informations peut-on tirer de cette carte ?



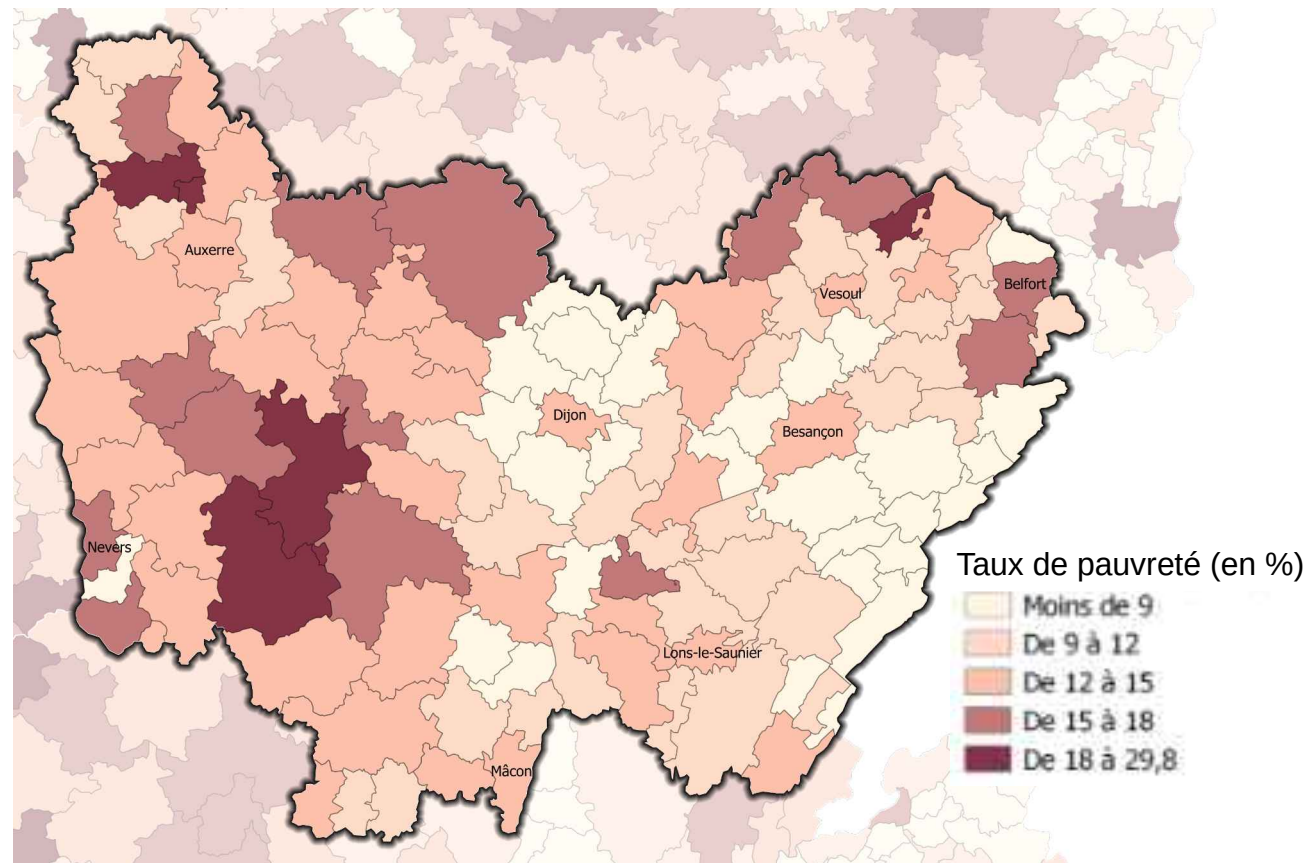
Limites des outils statistiques : la maille d'observation

Bassin de vie de
Chablis : 11,3 %

Bassin de vie de
Dole : 14,5 %



Limites des outils statistiques :

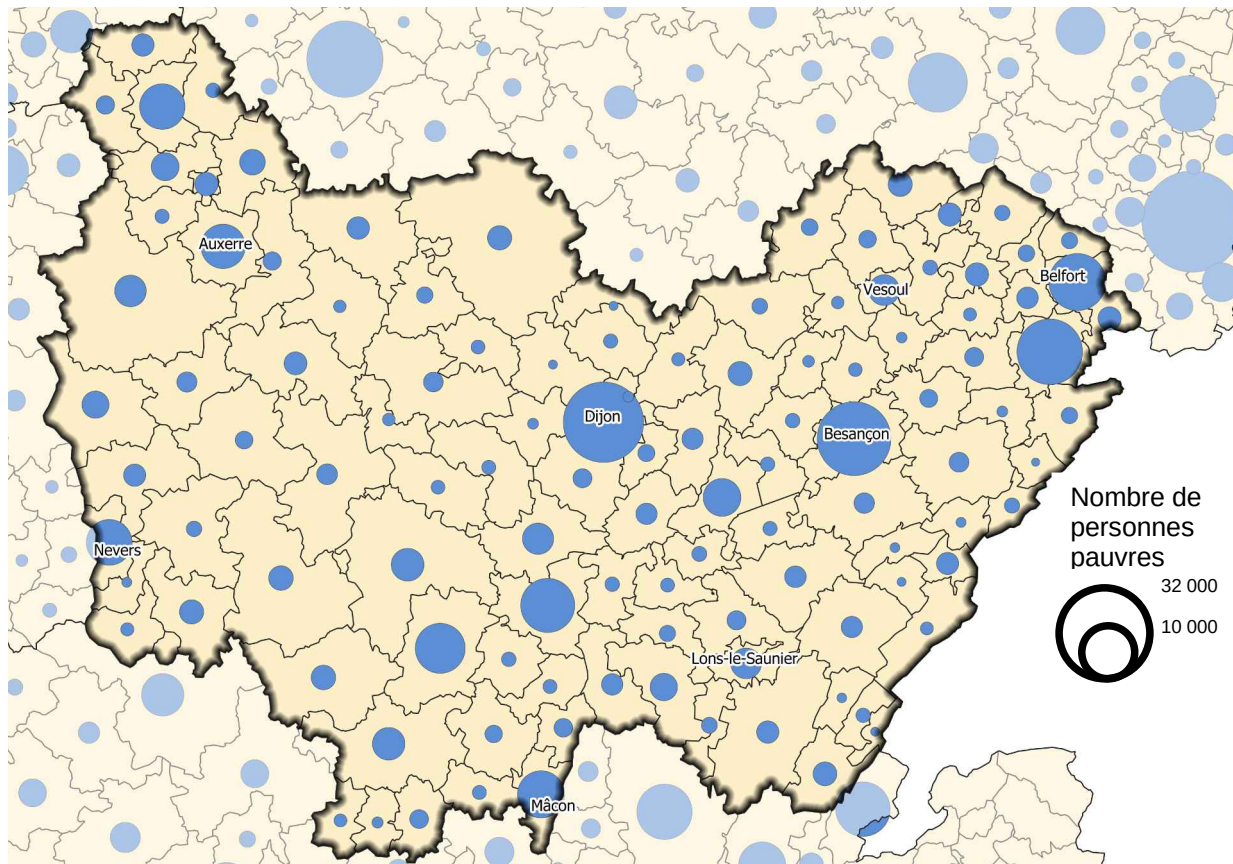


Quelles informations peut-on tirer de cette carte ?

Limites des outils statistiques :



Quelles informations peut-on tirer de cette carte ?

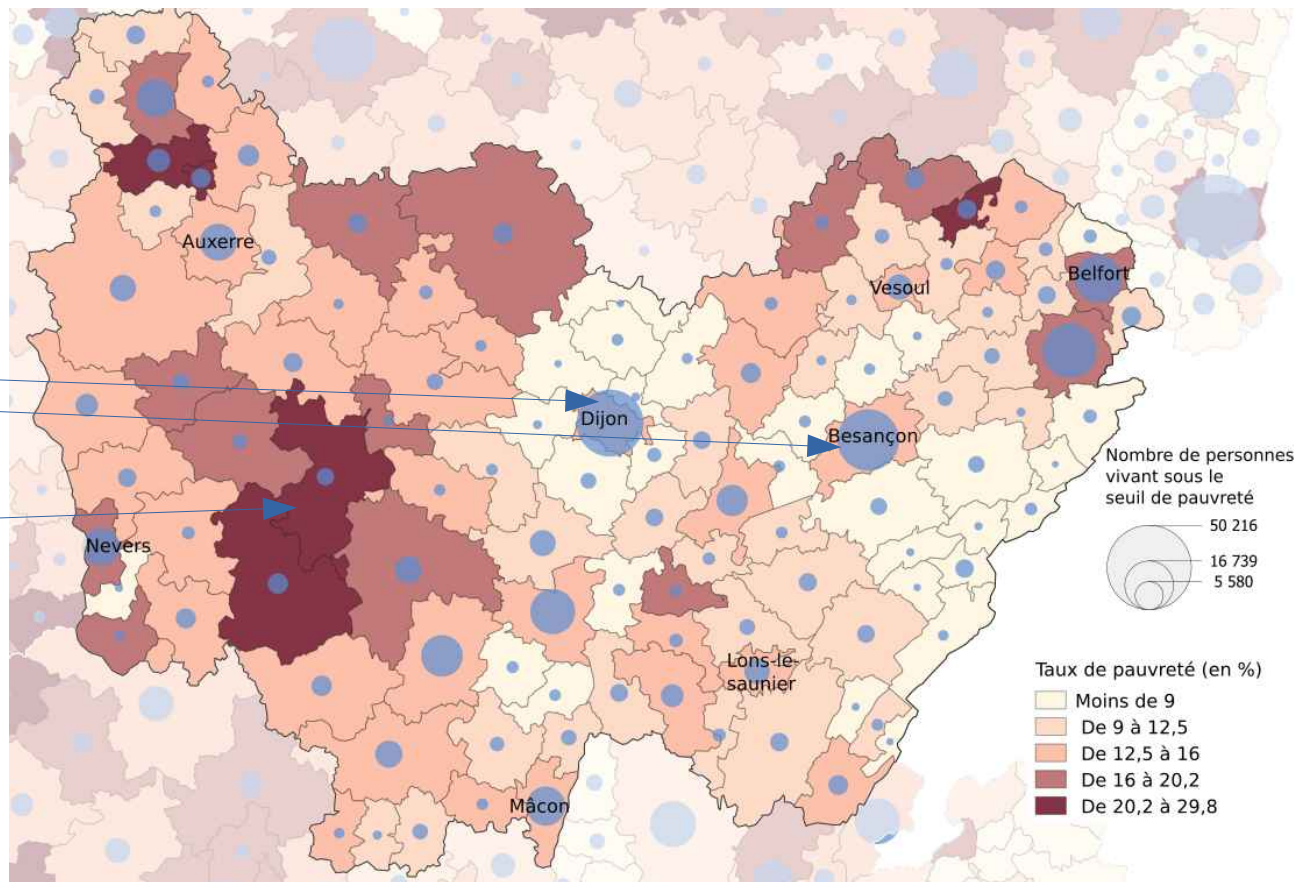


Limites des outils statistiques : un seul indicateur est réducteur

Avec deux indicateurs représentés,
les personnes pauvres

- résident surtout
dans les grandes villes

- représentent une part
importante de la population
dans de nombreuses
zones rurales



Attention à l'interprétation



Regroupement des résultats

Exemple d'un sondage sur le travail dominical



« Travailler le dimanche et être payé davantage qu'en semaine. Si votre employeur vous proposait de travailler le dimanche, accepteriez-vous ? »

A : non jamais

33 %

B : peut-être de temps en temps

50 %

C : oui toujours

17 %



**Le Journal
du Dimanche**

du 12 octobre 2008

Résultat du sondage :

67 % des Français sont favorables au travail le dimanche

«Les Français veulent travailler le dimanche »

**OUVERT
LE DIMANCHE**



Exemple :

plus **Nicolas Cage** apparaît dans des films,
plus le nombre de **morts noyés dans une piscine** augmente

Nombre d'apparitions de Nicolas Cage dans un film

corrélé au

Nombre de morts noyés dans une piscine
(corrélation : 66,6 %)



→ La corrélation
n'implique pas la
causalité !



80 % des accidents de voiture survenus entre 9h et 11h sont le fait de conducteur ayant bu du café au petit-déjeuner

"Le mariage est la cause principale de divorce" *Oscar Wilde*

Les ventes de crème glacée augmentent avec le nombre de morts par noyade



Une nouvelle mesure de la mobilité intergénérationnelle des revenus en France

Insee analyse n°73 Mai 2022

Les enfants de familles aisées ont trois fois plus de chances
d'être parmi les 20 % les plus aisés que ceux issus de familles modestes [...].

Le Monde

En France, selon l'Insee, l'ascenseur social n'est pas en panne

Selon une étude novatrice, 70 % des enfants de parents modestes grimpent dans l'échelle des revenus.

Les Echos

L'ascenseur social fonctionne encore en France



Déterminisme

**L'ascenseur social en panne ? Surtout
si vous êtes une femme élevée par une
mère seule dans le Nord**



***L'ASCENSEUR SOCIAL FRANÇAIS EST-
IL RÉELLEMENT EN PANNE?***

LE FIGARO

En France, l'ascenseur social est bien grippé

Conclusion

LES DONNÉES NE SONT PAS DONNÉES ELLES SONT CONSTRUITES



Une réflexion importante est souvent nécessaire pour **choisir le (ou les) indicateur(s) pertinent(s)** pour éclairer et mesurer un phénomène

Merci de votre attention

insee.fr



PARTIE 3 : EXEMPLES DE MESSAGES/D'ANALYSES