

Que peut-on faire dire aux chiffres, aux représentations ?

« Je ne crois aux statistiques que lorsque je les ai moi-même falsifiées », disait Winston Churchill.

Que peut-on ou que veut-on faire dire à certaines données (pourcentages, statistiques) ?

Que pourraient induire certaines représentations (diagrammes, graphiques) ?

Peut-on améliorer les choix des représentations ?

1. Interprétation différente d'un pourcentage :

Les deux groupes comparés ne sont **pas de même taille**.

Source :

[CLEMI \(Centre pour L'Éducation aux Médias et à l'Information\)](#)



ISRAËL : 40% DES NOUVEAUX CAS SONT VACCINÉS

Par Clémence Barbier - Mis à jour le 27/06/2021 à 09:50
Publié le 27/06/2021 à 09:48

Malgré une couverture vaccinale importante en Israël, 40% des nouvelles contaminations sont des personnes vaccinées selon Gabi Barbash, l'ancien directeur général du ministère de la Santé, qui craint la flambée du variant Delta.

Le masque est de retour dans les rues de Jérusalem après une recrudescence des cas de contamination à la Covid-19, en partie liée au variant Delta. Israël enregistre une centaine de nouveaux cas par jour. Selon l'ancien directeur général de la santé, 40% d'entre eux sont des personnes vaccinées.

2. Interprétation de 2 valeurs choisies :

On donne l'impression d'une baisse en **prenant la valeur journalière la plus forte et en la comparant à une autre date aléatoire**.

La notion de « **moyenne glissante** » permet de montrer la fausseté de l'explication : cette moyenne permet de lisser une série de valeurs exprimées en fonction du temps, en éliminant les fluctuations les moins significatives.

Par exemple, comparer une moyenne sur une semaine entière est plus significatif que de prendre deux dates au hasard car les variations journalières peuvent être liées à des questions purement techniques (moins de remontées statistiques effectuées le week-end par exemple).



Source :

[CLEMI \(Centre pour L'Éducation aux Médias et à l'Information\)](#)

Autres activités :

Licences FFF

Il fait froid et on parle de réchauffement climatique !

3. L'utilisation des pourcentages :

Les calculs d'inégalités de salaires ont des **résultats différents selon que l'on parle en termes de temps complets, de temps partiels, ou encore en équivalent temps plein.**

Si on prend les salaires mensuels nets en équivalent temps plein, les femmes gagnent 2 113 € contre 2 533 € pour les hommes. L'écart mensuel est de 420 €.

Les inégalités sont donc évidentes mais comment les calculer ?

Le chiffre ne sera pas le même si on calcule ce que les femmes perçoivent en moins ou ce que les hommes gagnent en plus.

S'agissant du salaire net équivalent temps plein on peut dire que les femmes gagnent un salaire équivalent à 83,4% de celui des hommes donc inférieur de 16,6% à celui des hommes ou que les hommes gagnent 19,9% de plus que les femmes.

L'observatoire des inégalités fait à ce propos le commentaire suivant : Comment présenter l'écart de salaires hommes / femmes : la vision des hommes et celle des femmes ? Aucune des deux méthodes n'est plus « juste » ou meilleure, indique l'Observatoire des inégalités.

Il est toutefois impressionnant de constater que le chiffre le plus faible s'est imposé dans tous les documents et dans le débat public. La présentation la plus fréquente par le chiffre le plus faible peut engendrer de mauvais réflexes lorsqu'il s'agit de savoir de combien il faudrait augmenter en moyenne les salaires des femmes pour que ceux-ci soient égaux à ceux des hommes. Pour connaître l'augmentation qui mettrait le salaire moyen féminin au niveau de celui des hommes, il faut accroître le salaire des femmes de 19,9%, c'est-à-dire de ce que les hommes gagnent en plus.

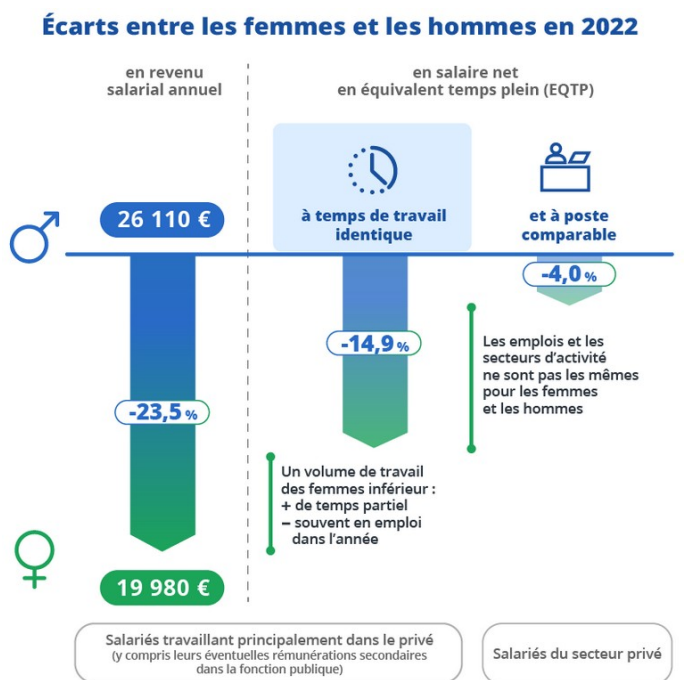
Prolongement :

Inégalités entre les femmes et les hommes

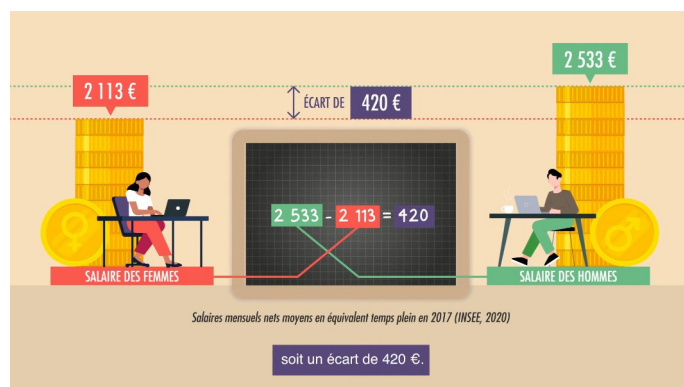
4. Moyenne et médiane :

La moyenne est l'un des outils les plus fréquemment utilisés pour décrire une série statistique. Elle est, pourtant, souvent peu représentative de la distribution effective des données étudiées. Il est alors utile d'inclure la médiane dans l'analyse.

Les exemples de distribution des salaires et des patrimoines, détaillés ci-dessous, montrent qu'une analyse fondée uniquement sur la moyenne, produirait une vision déformée de la réalité.



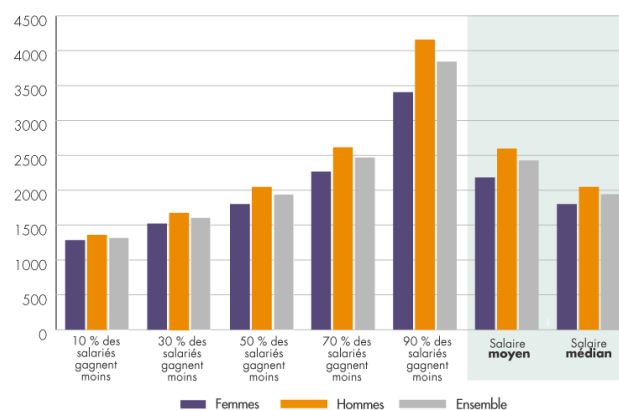
Source : [INSEE](#)



Source : [La finance pour tous](#)

- En 2019, le salaire moyen était de 2 424 € et le salaire médian était de 1 940 €. Le salaire médian est donc nettement inférieur au salaire moyen. Autrement dit, plus de la moitié de la population gagne moins que la moyenne et moins de la moitié gagne plus ; cela est dû au fait qu'une petite partie de la population perçoit un salaire très élevé, ce qui tire la moyenne vers le haut.

DISTRIBUTION DES SALAIRES MENSUELS NETS
(EN 2019, EN EUROS, SALARIÉS EN ÉQUIVALENT TEMPS PLEIN)



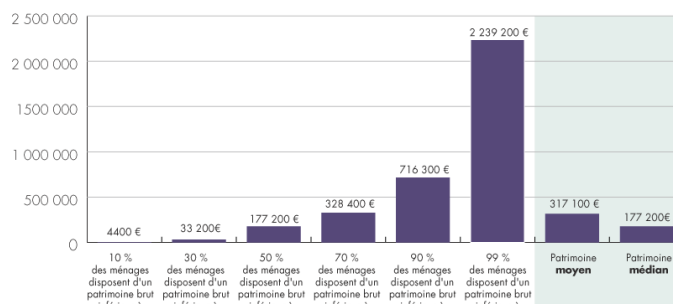
Source : lafinancepourtous.com d'après INSEE



- Le patrimoine brut des ménages présente une distribution particulièrement inégalitaire. Le patrimoine brut moyen est de 317 100 € par ménage en France en 2021. Il s'agit d'un montant largement supérieur à la médiane qui est de 177 200 €. Cela signifie que 50% des ménages disposent d'un patrimoine brut inférieur à 177 200 € et les 50% restants ont un patrimoine d'une valeur supérieure.

PATRIMOINE BRUT DES MÉNAGES

EN 2021, EN EUROS



Source : lafinancepourtous.com d'après INSEE



La différence de plus de 40% entre la moyenne et la médiane s'explique par le fait que **les hauts patrimoines « tirent la distribution vers le haut »** :

1% des ménages disposent d'un patrimoine brut supérieur à 2 239 200 €.

La moyenne est un indicateur suffisant dans le cas où les données sont réparties également. Mais il est insuffisant lorsque ce n'est pas le cas et il masque alors la réalité. Pour savoir si une répartition est égalitaire ou inégalitaire, il faut calculer la médiane et observer si la médiane est à peu près égale à la moyenne (répartition « égalitaire ») ou si la médiane s'écarte de la moyenne (répartition « inégalitaire »). Si la médiane est inférieure à la moyenne, les inégalités se situent en haut de l'échelle. Dans le cas inverse, les inégalités se situent en bas de l'échelle.

Source : La finance pour tous

Autres activités : Elon Musk et ses amis
Déchets collectés
Inégalités salariales

5. L'effet de migration des stades, ou effet Will Rogers :

en hommage à l'acteur du même nom qui aurait déclaré « *quand les Okies quittèrent l'Oklahoma et vinrent en Californie, ils élevèrent l'intelligence moyenne des deux côtés* ».

En déplaçant un élément d'un groupe à l'autre, on peut paradoxalement faire monter la moyenne dans... les deux groupes ! Représentons-nous par exemple un ensemble de patients qui indiquent sur une échelle de 1 à 10 la douleur moyenne qu'ils ressentent. On crée un groupe A à douleur faible, $A = \{ 1, 2, 3, 4 \}$ et un groupe B de douleur forte, à partir de 5, soit $B = \{ 5, 6, 7, 8, 9 \}$ Si l'on fait la moyenne de chaque groupe, on obtiendra 2,5 pour A et 7 pour B.

Mais imaginons que les normes d'inclusion changent, et que par exemple, après une recommandation de l'Organisation mondiale de la **Santé**, 5 est finalement considéré comme faible. Les groupes deviennent alors $A = \{ 1, 2, 3, 4, 5 \}$ et $B = \{ 6, 7, 8, 9 \}$. Or de ce fait, la moyenne (notée ci-dessous μ) de A est montée à 3, et celle de B, à 7,5. Les deux groupes ont vu leur moyenne augmenter.

Situation 1 : $A = \{ 1, 2, 3, 4 \} \rightarrow \mu = 2,5$ et $B = \{ 5, 6, 7, 8, 9 \} \rightarrow \mu = 7$

Situation 2 : $A = \{ 1, 2, 3, 4, 5 \} \rightarrow \mu = 3$ et $B = \{ 6, 7, 8, 9 \} \rightarrow \mu = 7,5$

Quelles en sont les conséquences ? Si le système de détection d'une maladie par exemple s'améliore et permet du dépistage précoce, certains individus passeront du groupe des sujets en bonne santé vers le groupe des sujets malades. À travers ce changement, la moyenne de la durée de vie augmentera paradoxalement dans les deux groupes, et cela quel que soit le traitement que l'on fera. On aura ainsi tendance à conclure à l'efficacité du traitement, alors que c'est un problème d'un critère d'inclusion dans les groupes qui a changé.

Source : [CORTECS](https://www.cortece.fr/)

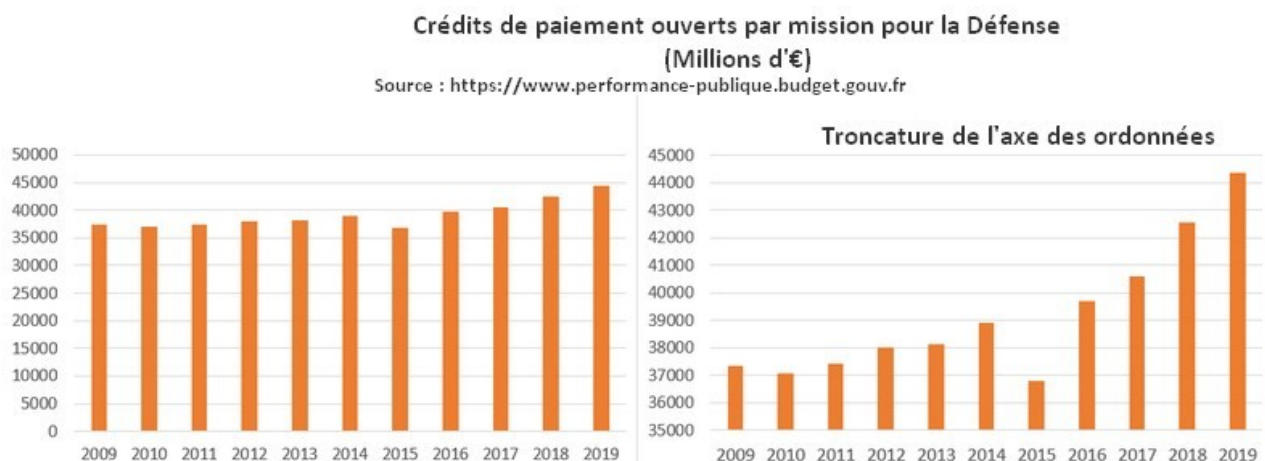
Autre activité : Evolution des moyennes

6. Graphiques trompeurs :

Un **graphique** est un outil pour visualiser une grande quantité de données. Mais il peut aussi induire en erreur, parfois intentionnellement.

Voici quelques stratégies utilisées pour rendre un graphique trompeur :

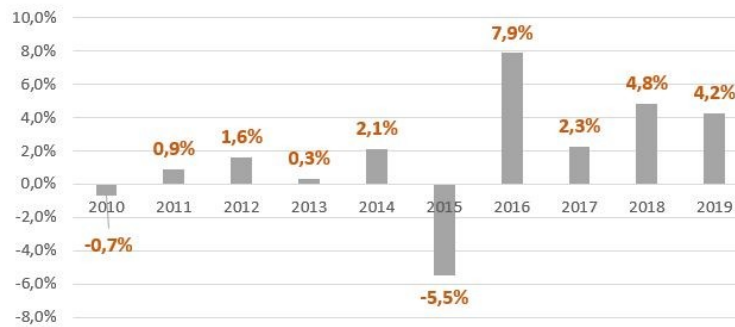
a. Modifier l'échelle :



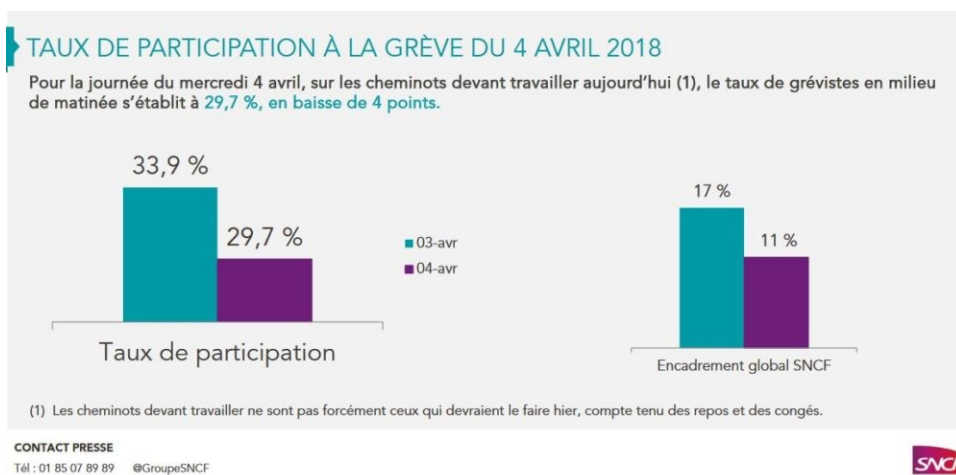
- A gauche, on a un graphique avec un axe classique : il montre un budget stable de 2009 à 2016, avec une légère augmentation en fin de période.
- A droite, on a un graphique avec un **axe tronqué** : en réduisant l'échelle de l'axe des ordonnées, la même évolution donne l'impression d'une augmentation marquée et d'un investissement croissant.

- Dessous, un autre moyen d'étudier l'évolution d'un paramètre serait de **calculer les variations**.

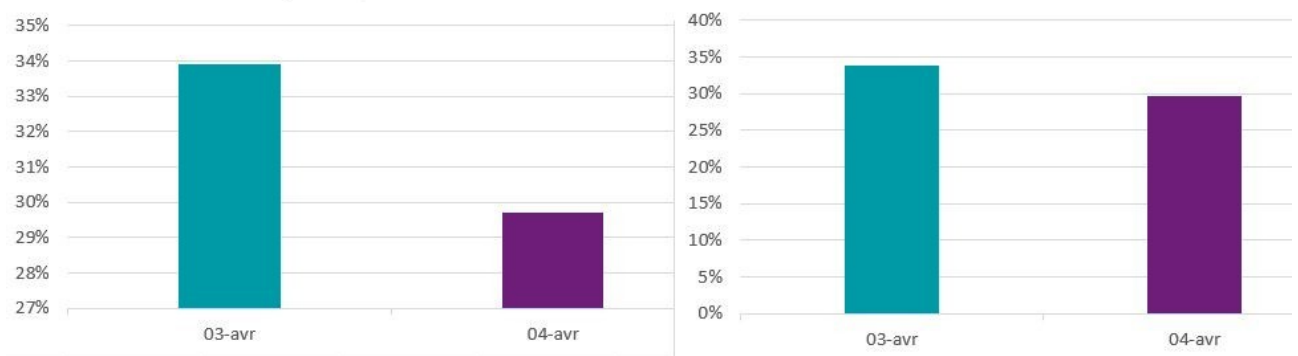
Evolution par rapport à l'année précédente des crédits de paiement alloués à la Défense



b. Omettre certaines données :



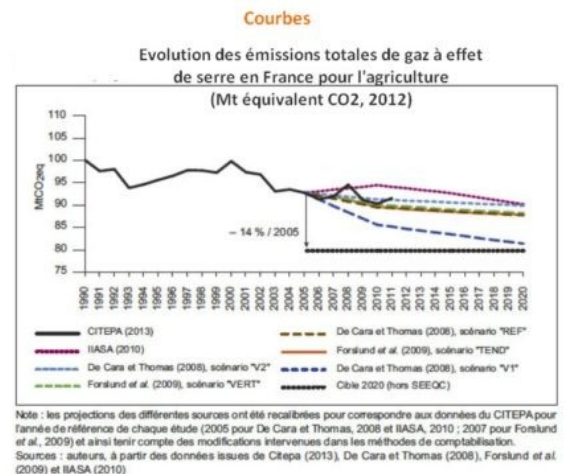
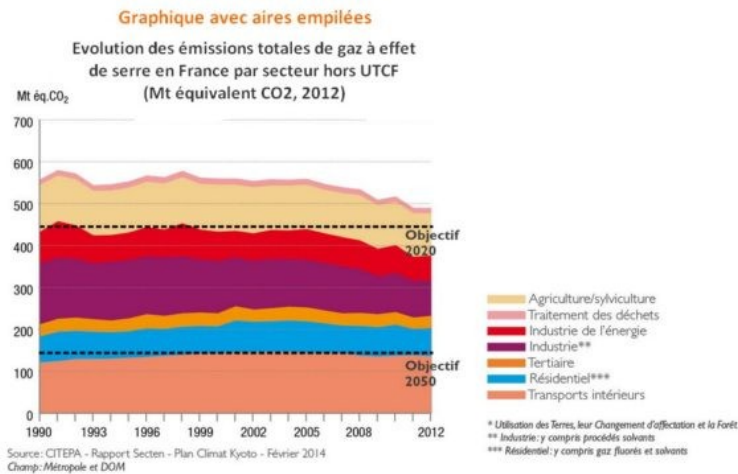
Taux de participation à la grève SNCF



- Dans un communiqué de presse publié pendant une grève, la SNCF a présenté un graphique concernant le taux de grévistes, mais a omis l'**axe des ordonnées**. Cette suppression rend la lecture et l'interprétation des données beaucoup plus compliquées. De plus, l'**axe** semble avoir été **tronqué** en ne commençant pas à zéro, réduisant l'apparence du succès de la grève pour démotiver les cheminots.
- Mais, en traçant les graphiques avec un axe des ordonnées puis une échelle partant de zéro, la baisse semble moins significative.

c. Les graphiques à aires empilées :

Ils compliquent souvent l'interprétation des données. Ces graphiques demandent de comparer des aires pour évaluer les tendances, ce qui n'est pas une tâche intuitive.

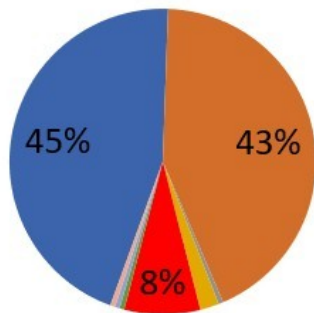
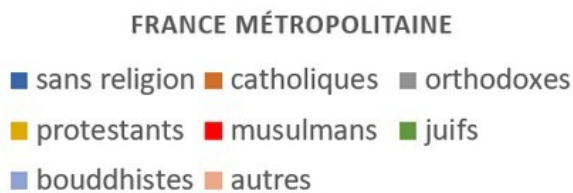


- Dans le graphique de gauche, les émissions agricoles sont représentées en jaune - saumon. Pour comprendre l'évolution de ce secteur, il faut comparer la taille des aires entre les différentes catégories, ce qui complique la lecture et peut masquer des tendances.
- A droite, en isolant uniquement les émissions du secteur agricole, le graphique devient beaucoup plus clair. On distingue une tendance nette à la diminution des gaz à effet de serre, avec des projections représentées par plusieurs courbes.

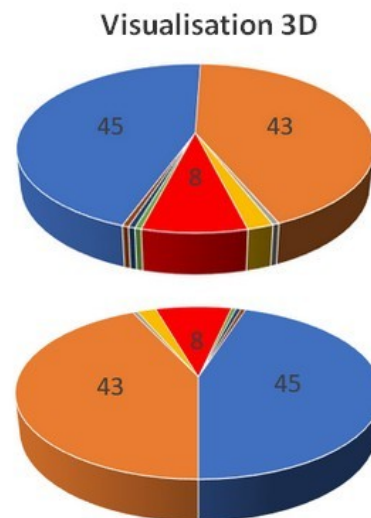
d. L'illusion de la 3D :

Enquête de l'INED en 2008 sur la répartition religieuse en France

- 45% des Français déclaraient ne pas avoir de religion,
- 43% étaient catholiques, 8% musulmans et 2% protestants.



Source : INED



Trois diagrammes circulaires peuvent représenter ces données de manière correcte. Cependant, en ajoutant une perspective 3D au graphique, une manipulation visuelle peut émerger. Par exemple, placer la part des musulmans (8 %) au premier plan, en accentuant l'épaisseur 3D, donne l'impression qu'elle est bien plus importante qu'en réalité.

Source : [Quoi dans mon assiette](#)

Autres activités : [News maths & fakes - Production académie Rennes](#)

Qu'en penses-tu ?
Importation de charbon

7. Les corrélations trompeuses :

Activités : [News maths & fakes - Production académie Rennes](#)

8. Autres activités intéressantes :

[Esprit critique, es-tu là ? TraAM académie Rennes](#)

[Paradoxe de Simpson - TraAM académie Rennes](#)

[Étude de performances paradoxales au baseball - TraAM académie Rennes](#)