

Chiffres et statistiques

G. Haesbroeck

DÉPARTEMENT DE MATHÉMATIQUE – UNIVERSITÉ DE LIÈGE

Université des champs 2023

Statistiques dans la presse

De nombreux articles de la presse présentent des données brutes ou des résultats d'analyses statistiques. Il n'est pas rare de constater que

- la présentation n'est pas optimale (voire est biaisée)
- l'interprétation n'est pas correcte/complète/assez prudente

Deux constats:

- De nombreux citoyens acceptent ces informations sans esprit critique.
- Les effets d'annonce (parfois/souvent divergents) renforcent l'idée que "la statistique est la première des sciences inexactes" (citation parue dans le journal des frères Goncourt).

Inspiration de cet exposé: la formation de la Société Royale Belge de Statistique

La Société Royale Belge de Statistique a été créée en 1937 et compte aujourd'hui de l'ordre de 250 membres.

Les objectifs de la Société sont de

- contribuer au progrès scientifique en incitant à la collaboration entre les statisticiens belges
- aider le public au sens large à mieux comprendre la place de la statistique dans le monde moderne.

Un projet concret était envisagé par le Conseil d'Administration : créer une formation en statistique pour les journalistes.

Impulsion finale: présentation des *statistiques Covid* dans la presse à partir de mars 2020...

Comment bien rapporter et interpréter des données

FORMATION À DESTINATION DES
JOURNALISTES

ORGANISÉE PAR LA SOCIÉTÉ ROYALE BELGE
DE STATISTIQUE (SRBS)



Supports de cette formation

La première version du contenu et des supports de cette formation a été développée par la plateforme technologique de Support en Méthodologie et Calcul Statistique ([SMCS](#)) de l'Université catholique de Louvain sous la coordination de la Société Royale Belge de Statistique (SRBS).



Contenu de l'exposé de ce 18 avril

- Présentation et comparaison de chiffres
- Calcul et présentation de pourcentages
- Corrélation/association
- Graphiques
- Conclusion

L'objectif est de passer d'un extrait de journal à un autre pour attirer l'attention sur certains aspects et faire réfléchir à l'interprétation des chiffres et à leur portée.

Note importante "personnelle" :

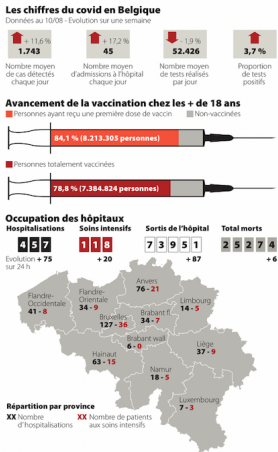
- Ce regain d'intérêt pour les articles de presse est lié aux nombreuses statistiques publiées pendant la crise sanitaire; de nombreux exemples proviennent de cette période.
- Je suis abonnée au journal Le Soir.

Présentation et comparaison de chiffres

- Comparaison à l'aide de chiffres bruts
- Et en termes relatifs?
- “Raccourci” de présentation/effet d'annonce

Utilisation de chiffres bruts dans la presse

Il n'est pas rare de "montrer" des chiffres bruts dans la presse:
Par exemple: *les chiffres de la pandémie* du journal Le Soir du 11 août 2021:



Utilisation de chiffres bruts dans la presse

Comptage des cas Covid dans le monde (article sur le site info 7 sur 7, 25 décembre 2020):



The screenshot shows a news article from the website '7 SUR 7'. The header includes the site logo and navigation tabs for ACTU, SHOW, SPORT, LIFESTYLE, and VIDEO. Below the header, there is a sub-header with categories: Belgique, Monde, Faits divers, Insolite, Ecologie, Tech, Santé, Economie, Sciences. The main text of the article discusses the global impact of the COVID-19 pandemic, mentioning 1.743.187 deaths and 25 million cases as of December 25, 2019. It also highlights the USA as the most affected country in terms of both deaths and cases. Social media sharing buttons for Facebook and Twitter are visible. At the bottom of the article, there is a list of other countries with their respective case counts.

7 SUR 7  **ACTU** **SHOW** **SPORT** **LIFESTYLE** **VIDEO** 

Belgique **Monde** Faits divers Insolite Ecologie Tech Santé Economie Sciences

La pandémie du nouveau coronavirus a fait au moins 1.743.187 morts dans le monde depuis que le bureau de l'OMS en Chine a fait état de l'apparition de la maladie fin décembre 2019, alors que l'Europe a franchi en ce jour de Noël le seuil des 25 millions de cas, selon un bilan établi par l'AFP à partir de sources officielles vendredi à 12h00 (heure belge). La

Rédaction 25-12-20, 14:37 Dernière mise à jour: 25-12-20, 14:42 Source: AFP

Plus de 79.313.380 cas d'infection ont été officiellement diagnostiqués depuis le début de l'épidémie, dont au moins 50.154.100 sont aujourd'hui considérés comme guéris.

Les USA reste le pays le plus touché tant en nombre de morts que de cas

Les États-Unis sont le pays le plus touché tant en nombre de morts que de cas, avec 329.106 décès pour 18.655.012 cas recensés, selon le comptage de l'université Johns Hopkins. Au moins 6.298.082 personnes ont été déclarées guéries.

Après les États-Unis, les pays les plus touchés sont le Brésil avec 189.982 morts et 7.423.945 cas, l'Inde avec 147.092 morts (10.146.845 cas), le Mexique avec 121.172 morts (1.362.564 cas), et l'Italie avec 70.900 morts (2.009.317 cas).

Comparaison à l'aide de valeurs relatives

L'utilisation de chiffres bruts pour transmettre de l'information n'est pas problématique en soi mais les exploiter pour effectuer des comparaisons/ des rankings est trompeur.

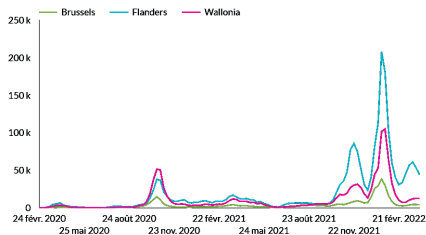


The screenshot shows a news article header from 'Le Vif'. The navigation bar includes 'Rubriques', 'Le magazine', 'Avantages pour abonnés', and 'Abonnez-vous'. The article is dated '16/12/21 à 11:47' and 'Mise à jour à 12:58'. The title is 'Covid: la Belgique se classe-t-elle toujours parmi les pays les plus touchés au monde? (infographies)'. The author is 'Maily Chavagne', a 'Journaliste Web'. The main text states: 'Nombre total de décès: Belgique, 26^e du classement. Bref, tout semble indiquer que le pire est derrière nous. Néanmoins, tous ces chiffres ne font pas de la Belgique un bon élève en matière de lutte contre le covid. En chiffres bruts, la Belgique figure à la 26^e place du classement des pays qui présentent le nombre le plus élevé de décès, derrière l'Iraq et la Thaïlande.'

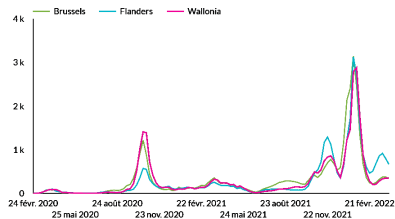
Pour comparer des chiffres obtenus sur des entités de masse différente, il convient de travailler en pourcentages ou taux.

Le “ranking” des régions varie si on parle en termes de chiffres bruts ou en termes de taux:

Weekly New Cases per Region



Weekly New Cases per Region per 100k Inhabitants



La Belgique première du "classement" en termes de taux de mortalité (DH, 4 juillet 2020; Le Vif, décembre 2021)

DH Actus SPORTS Régions Coseo Médias Lifestyle & people Buzz Nos podcasts

Coronavirus: la Belgique a le taux de mortalité le plus élevé au monde

Belgique

AFP

Publié le 04-07-20 à 21h45 - Mis à jour le 05-07-20 à 11h14



Selon un bilan établi par l'AFP à partir de sources officielles, la Belgique est le pays qui déplore le plus grand nombre de morts par rapport à sa population, avec 84 décès pour 100.000 habitants.

La Belgique est suivie par le Royaume-Uni, qui compte 65 morts pour 100.000 habitants, puis par l'Espagne (61), l'Italie (58), et la Suède (54).

Selon ce même bilan, la pandémie du coronavirus a fait au moins 527.241 morts dans le monde depuis que le bureau de l'OMS en Chine a fait état de l'apparition de la maladie fin décembre.

Mais... la difficulté d'opérer une comparaison honnête dépasse le simple fait d'imposer une base commune.

LE VIF Rubriques Le magazine Avantages pour abonnés Abonnez-vous

16/12/21 à 11:47
Mise à jour à 12:58

Covid: la Belgique se classe-t-elle toujours parmi les pays les plus touchés au monde? (infographies)

 **Maily Chavagne**
Journaliste Web

Taux de mortalité: la Belgique, 22^e du classement

Par rapport à son nombre d'habitants, la Belgique reste aujourd'hui encore l'un des pays les plus durement frappés par la pandémie. Avec un taux de 242,77 décès pour 100.000 habitants, notre plat pays figure à la 22^e place du classement des États présentant le plus haut taux de mortalité.

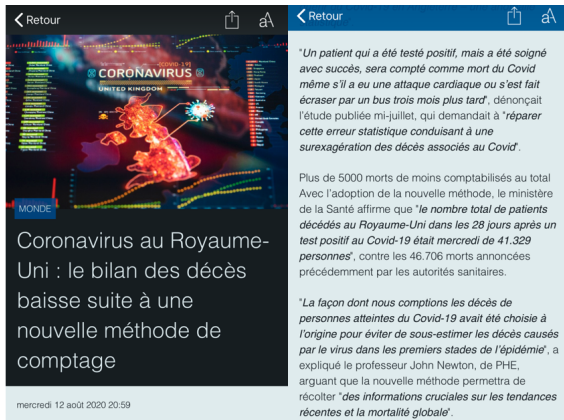
Elle se classe juste derrière les États-Unis (245,29 morts pour 100.000 habitants), dont le bilan est pourtant de loin le plus élevé au monde; le nombre de morts aux États-Unis de Covid-19 a en effet dépassé les 800.000, un chiffre autrefois inimaginable.

Il y a tellement de facteurs qui entrent en jeu...

Dans le cas des taux de mortalité Covid:

- La densité de population est différente d'un pays à l'autre.
- La méthode de comptage et la fiabilité des chiffres rapportés peuvent varier d'un pays à l'autre. Certains pays ne comptabilisent que les décès dans les hôpitaux (et non en maisons de retraite), d'autres comptent tous les décès de personnes testées positives (même si le décès est dû à une autre pathologie).
- Les données ne correspondent pas nécessairement au même stade de l'épidémie (même variant du virus en activité).

Illustration concrète de la difficulté des comparaisons “spatiales”



The image shows a screenshot of a news article from a mobile device. The article title is "Coronavirus au Royaume-Uni : le bilan des décès baisse suite à une nouvelle méthode de comptage". The main text discusses a change in the counting method for COVID-19 deaths, leading to a decrease in the reported toll. It mentions that more than 5,000 deaths were previously uncounted. The article also includes a quote from Professor John Newton of PHE, explaining that the new method aims to avoid underestimating deaths by including cases that die from other causes shortly after a positive COVID-19 test.

Retour

CORONAVIRUS
UNITED KINGDOM

MONDE

Coronavirus au Royaume-Uni : le bilan des décès baisse suite à une nouvelle méthode de comptage

mercredi 12 août 2020 20:59

"Un patient qui a été testé positif, mais a été soigné avec succès, sera compté comme mort du Covid même s'il a eu une attaque cardiaque ou s'est fait écraser par un bus trois mois plus tard", dénonçait l'étude publiée mi-juillet, qui demandait à "réparer cette erreur statistique conduisant à une surexagération des décès associés au Covid".

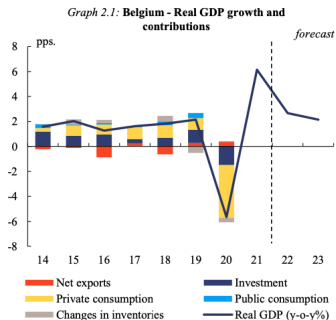
Plus de 5000 morts de moins comptabilisés au total Avec l'adoption de la nouvelle méthode, le ministère de la Santé affirme que *"le nombre total de patients décédés au Royaume-Uni dans les 28 jours après un test positif au Covid-19 était mercredi de 41.329 personnes"*, contre les 46.706 morts annoncées précédemment par les autorités sanitaires.

"La façon dont nous comptons les décès de personnes atteintes du Covid-19 avait été choisie à l'origine pour éviter de sous-estimer les décès causés par le virus dans les premiers stades de l'épidémie", a expliqué le professeur John Newton, de PHE, arguant que la nouvelle méthode permettra de récolter "des informations cruciales sur les tendances récentes et la mortalité globale".

A la décharge des journalistes, des appels à la prudence apparaissent de temps à autre dans les articles... mais est-ce suffisant?

Février 2022: le taux de croissance du PIB des pays de l'UE

Graphique décrivant l'évolution du PIB de la Belgique publié dans le rapport de la Commission Européenne de février 2022:



Estimation pour 2022: 2,7%, ce qui fait de la Belgique le pays en dernière position!!!

Beaucoup de réactions sur les réseaux sociaux en suivi de cette publication

Extrait du fil info de la RTBF du 14 février 2022



C'est un tableau qui suscite pas mal de commentaires sur les réseaux sociaux : il s'agit du tableau comparé des taux de croissance des pays de l'Union Européenne pour 2022. La Belgique y apparaît en toute dernière position des 27 pays membres de l'Union, avec 2,7% de croissance annoncée. Comment expliquer cette lanterne rouge ? Déclic a voulu y voir clair.

Il désigne quoi ce chiffre exactement ?

Ce taux de croissance décrit l'évolution du produit intérieur brut (PIB), un indicateur qui quantifie, en gros, les richesses produites par un pays en une année. Ici, les prévisions de la commission imaginent que si on prend pour base le PIB belge au 1er janvier 2022 et qu'on regarde son évolution jusqu'au 31 décembre, il aura augmenté de 2,7%. Ce qui ne dit encore rien de la hauteur du PIB au final, ni de son niveau de départ au début de



En cause : un régime fiscal particulièrement accommodant pour toute une série de multinationales, notamment les géants du numérique qui viennent installer leurs quartiers généraux européens en Irlande et y payer le moins d'impôts possibles. Cela signifie que l'Irlande grignote de la sorte une petite partie de la croissance des autres Etats membres... Mais surtout leurs recettes fiscales.

Cela démontre aussi que comparer des taux de croissances annuels entre des pays qui ont des politiques économiques et fiscales aussi différentes au sein de l'UE n'a pas forcément beaucoup de sens.

Auteur: Arnaud Ruysen

“Raccourci” de présentation/effet d’annonce

Des “raccourcis” (titre accrocheur, calculs trop simplistes...) peuvent amener des fausses impressions aux lecteurs.

Économie

Selon Test Achats, les ménages belges ont perdu en moyenne 1626 euros durant la crise du coronavirus

RTBF

🕒 Publié le vendredi 22 mai 2020 à 05h05

527 

Plus de 60% des ménages belges déclarent **avoir perdu de l'argent**, pour un montant moyen de 1626 euros, lors de **la crise sanitaire du coronavirus**. C'est ce qui ressort d'un sondage réalisé par Test Achats les 14 au 15 mai auprès d'un échantillon de 1000 Belges, âgés de 18 à 74 ans. Au total, cette perte financière représente 4,9 milliards d'euros.

Le titre “généralise” le contenu de l'article!

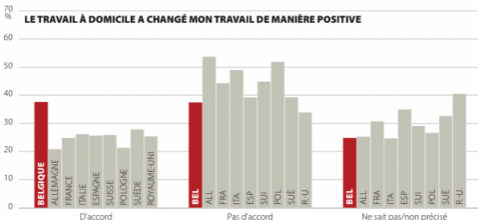
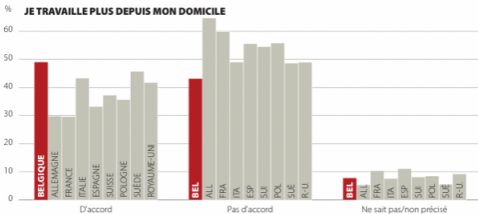
Le Belge télétravaille plus que ses voisins européens et s'en dit plutôt satisfait

Le Soir Samedi 19 et dimanche 20 décembre 2020

16 à la une

YouGov
a interrogé
10.000
personnes
dans neuf pays
européens
dont la
Belgique.

En quoi la crise du covid a-t-elle changé votre vie professionnelle en 2020 ?



The screenshot shows the Futura Santé website interface. At the top, there is a navigation bar with links for 'Accueil', 'EN CE MOMENT', '#Futura20ans', 'Pluie d'étoiles filantes de l'été', 'L'effondrement en Nouvelle-Zélande', and 'Événement extrême réchauffement'. Below this is the 'FUTURA SANTÉ' logo and navigation options: 'Explorer', 'Médias', 'Experts', 'Forum', and 'Bons Plans'. A social media bar follows with icons for Facebook, Twitter, Instagram, LinkedIn, YouTube, RSS, and others. The main article header features a red banner with 'Accueil / Santé / Actualités' and a large image of red handprints and a vaccine vial. The title reads: '— SANTÉ — Pourquoi plus de personnes vaccinées que de non vaccinées meurent de la Covid-19 en Angleterre'. Below the title, it says 'ACTUALITÉ' and 'Classé sous : VACCIN ANTI-COVID, CORONAVIRUS, EFFICACITÉ DES VACCINS'. The author is identified as Céline Deluzarche, Journaliste, with a small profile picture. The publication date is 'Publié le 24/07/2021'.

C'est une statistique assez perturbante dont se nourrissent les antivax : entre le 1^{er} février et le 21 juin 2021 en Angleterre, il y a eu davantage de morts de la Covid-19 parmi les personnes vaccinées et testées positives au variant Delta que parmi les personnes non vaccinées, atteste un rapport du gouvernement. Plus exactement, sur les 257 décès, 45 personnes avaient reçu une dose, 118 avaient reçu deux doses, et 92 n'avaient reçu aucune injection. La conclusion hâtive qu'il pourrait être tiré de ces statistiques est que le vaccin n'est pas efficace, voire dangereux.

Données du rapport du gouvernement

Table 5. Attendance to emergency care and deaths by vaccination status among Delta confirmed cases (sequencing and genotyping) including all confirmed Delta cases in England, 1 February 2021 to 21 June 2021

	Age group (years)**	Total	Cases with specimen date in past 28 days	Vaccination status unknown	<21 days post dose 1	≥21 days post dose 1	Received 2 doses	Unvaccinated
Delta cases	All cases	123,620	63,707	14,359	8,562	17,933	10,834	71,932
	<50	111,008	57,673	12,900	8,453	13,391	5,600	70,664
	≥50	12,404	5,957	1,252	109	4,542	5,234	1,267
Deaths within 28 days of positive specimen date	All cases	257	N/A	2	1	44	118	92
	<50	26	N/A	-	-	3	2	21
	≥50	231	N/A	2	1	41	116	71

Simplifions l'analyse en considérant uniquement les catégories Vaccinés 2 doses et Non Vaccinés.

Globalement, risque plus élevé parmi les vaccinés

	Vaccinés 2 doses	Non Vaccinés
Nombre de cas	10834	71932
Nombre de décès	118	92
Risque	1%	0,1%

Mais un *facteur confondant* important n'est pas pris en compte dans ce calcul de risque: l'âge...

Les groupes des personnes vaccinées ou non vaccinées ne sont pas homogènes selon cette caractéristique.

En tenant compte de l'âge

	Vaccinés 2 doses		Non Vaccinés	
	< 50 ans	≥ 50 ans	< 50 ans	≥ 50 ans
Nombre de cas	5600	5234	70644	1267
Nombre de décès	2	116	21	71
Risque	0,03%	2,2%	0,03%	5,6%

Dans chaque “tranche d'âge”, les risques de décès sont égaux ou plus importants pour les non vaccinés par rapport aux vaccinés. Cependant, globalement, la catégorie plus âgée est “sous-représentée” dans la catégorie des non vaccinés, ce qui change l'effet (paradoxe de Simpson).

Utilisation et calcul de pourcentages (de variation)

- Choix de la base
- Interprétation/portée de l'information
- Erreur (notamment de copier-coller) ou imprécision

Rappelez-vous en juillet/août 2020: évolution des cas Covid

Les pourcentages d'augmentation annoncés quotidiennement dans la presse sont "vertigineux".



Les calculs sont corrects

Le pourcentage de variation caractérisant l'évolution (augmentation ou diminution) d'une variable (à valeur positive) au cours du temps est donné par

$$\% \text{ de variation} = \frac{\text{valeur finale} - \text{valeur initiale}}{\text{valeur initiale}} \cdot 100\%,$$

et les résultats annoncés dans la presse et calculés à partir des chiffres bruts étaient corrects.

Mais

- Il faut “relativiser” ou interpréter avec prudence les pourcentages (de variation) lorsque ceux-ci sont calculés sur des chiffres très réduits.
- Il faut s'assurer du caractère “invariable” des conditions de calcul.

Réaction de B. Rentier (ex-Recteur de l'ULiège)

7M7 ACTU SHOW SPORT LIFESTYLE VIDÉO Q

Belgique Monde Faits divers Insolite Ecologie Tech Santé Economie Sciences



▲ Bernard Rentier © belga

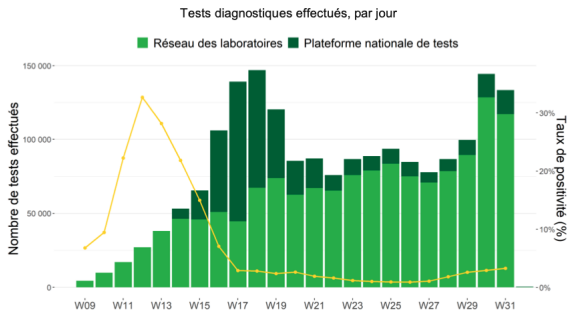
L'ex-recteur de l'ULiège et virologue Bernard Rentier veut présenter les chiffres du Covid-19 autrement: "Ce n'est pas +60%, mais +1% de cas confirmés"

The image shows Bernard Rentier, a man with a white beard, wearing a dark academic gown with a white fur stole and several medals on his chest. He is standing in front of a wooden door. The background is slightly blurred, showing an outdoor setting with a building.

Définition du taux de positivité

Selon B. Rentier, cela n'avait pas de sens de suivre l'évolution des cas (notamment à cause de la variation de la stratégie de testing). Il faut relativiser les nombres de cas détectés en fonction du nombre de tests effectués → il a proposé de calculer le "taux de positivité" et de suivre l'évolution temporelle de celui-ci.

Ce taux a ensuite été ajouté dans les rapports (de la RTBF notamment).



Proximus 09:42 99 %

Cas détectés¹ : entre le 01/08 et le 07/08, 1755 nouvelles infections au coronavirus ont été détectées en moyenne chaque jour. C'est une hausse de 12% par rapport à la semaine précédente.

Tests : entre le 01/08 et le 07/08, une moyenne de 50.973 tests ont été effectués quotidiennement, un total en baisse de 6% par rapport à la semaine précédente.

Admissions : elles s'élevaient en moyenne à 47,4 entre le 04/08 et le 10/08. C'est une hausse de 22% par rapport à la semaine précédente.

Personnes hospitalisées² : 467 patients sont actuellement hospitalisés en lien avec le Covid-19, dont 121 patients traités en soins intensifs.

Taux de positivité³ : sur la base des résultats des tests obtenus entre le 01/08 et le 07/08, il est de 3,8%, en hausse de 0,6% par rapport à la semaine dernière.

Féminisation des noms de rues à Bruxelles, RTBF, 3 mars 2020

INFO Rechercher sur l

À la une Fil Info Coronavirus Belgique Régions Monde Économie Société Médias

Régions Bruxelles Brabant Wallon Hainaut Liège Namur Luxembourg Flandre

A peine 6 % des rues bruxelloises portent des noms de femmes



03 mars 2020 à 11:39 - mise à jour 03 mars 2020 à 11:39 · 2 min

Par Elisabeth Groutars

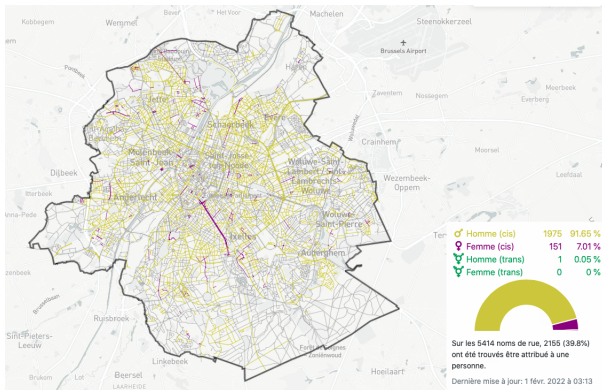
Info Regions Regions Bruxelles Info

PARTAGER f t w i n e

A Bruxelles, seules 6 % des rues portent des noms de femmes. Pour souligner cette inégalité, une association faisant la promotion des données ouvertes "Open Knowledge Belgium" et l'association féministe "Noms peut-être" ainsi qu'une soixantaine de citoyens ont créé une carte en ligne. Cette carte met en lumière les personnages féminins qui ont donné leur nom aux rues bruxelloises. Le site s'appelle EqualStreetNames.Brussels.

Seules 6 % des rues bruxelloises portent des noms de femmes © Tous droits réservés

Informations complémentaires dans l'article (information mise à jour “en continu”)



2,8% des rues portent un nom de femme et 37% des rues portent un nom d'homme;

Parmi les rues qui portent un nom de personne, 7% portent un nom de femme et 93% portent un nom d'homme.

Interprétation/portée de l'information : utilisation d'un "risque relatif"

SANTÉ & BIEN-ÊTRE

Vaccin contre le coronavirus : la 4ème dose est-elle utile et pour qui ?

hier à 17:12 - mise à jour il y a 3 heures · 3 min

Par Aline Delvoye

Santé & Bien-être Info Société

Coronavirus Info Accueil

CORONAVIRUS VACCIN

Accueil Vidéo Audio Mon choix Menu

11:14

98 %

U ne [étude israélienne publiée le 5 avril](#) confirme l'intérêt de la quatrième dose de vaccin contre le COVID chez les plus de 60 ans. Le risque de mourir du Covid est réduit de moitié grâce à la 4ème dose dans ce groupe d'âge.

Publicité

Il s'agit d'une étude menée par l'hôpital universitaire Assuta Ashdod, sous la direction du Dr Tal Brosh, chef du département des maladies infectieuses. L'étude a comparé un groupe de patients ayant reçu trois doses et un autre groupe ayant reçu quatre doses. Les résultats ont été collectés entre le 3 janvier et le 18 février 2022.

Réduction du nombre de décès

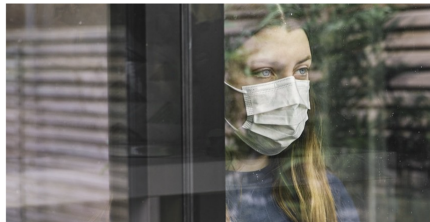
Accueil Vidéo Audio Mon choix Menu

Pour être parfaitement pertinente pour le lecteur, l'information donnée dans l'article devrait être complétée par le risque "initial" ... Si ce risque est de 60%, alors le réduire de moitié "vaut le coup". Si le risque est de 1%, l'intérêt de la réduction n'est pas le même.

Interprétation/portée de l'information : utilisation d'un pourcentage de variation "isolé"



Home / News / COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide



COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide

Wake-up call to all countries to step up mental health services and support

2 March 2022 | News release | Reading time: 3 min (927 words)

In the first year of the COVID-19 pandemic, global prevalence of anxiety and depression increased by a massive 25%, according to a scientific brief released by the World Health Organization (WHO) today. The brief also highlights who has been most affected and summarizes the effect of the pandemic on the availability of mental health services and how this has changed during the pandemic.

Pour être parfaitement pertinente pour le lecteur, un "point de départ" ou un "point d'arrivée" devrait être fourni afin de pouvoir interpréter cette croissance majeure de 25%...



Pourcentage de variation ou différence en points de pourcentage?



Coronavirus: malgré les craintes, le taux de réussite dans l'enseignement supérieur est à la hausse cette année

BELGA, publié le 07 juillet 2020 à 09h24



Ainsi, dans l'ensemble des six universités francophones, le taux de réussite lors de cette session de juin progresse de pas moins de 9% en moyenne. Selon des chiffres livrés parallèlement par le Conseil des recteurs francophones (Cref), le taux d'examens réussis les universités en juin est de 77%, alors qu'il tournait autour de 68% les trois années précédentes.

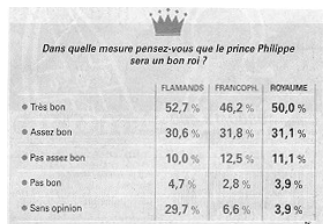
Calculs “incorrects”

Afin de relativiser les chiffres bruts ou calculer/interpréter des pourcentages, encore faut-il

- se méfier des arrondis ou des copiers-collers;
- choisir la bonne base ou du moins la préciser;
- effectuer le calcul “dans le bon sens” ou simplement correctement.

Sondage sur la monarchie, Le Soir du **27 nov. 1999**: “Dans quelle mesure pensez-vous que le prince Philippe sera un bon roi?”

Extrait:



Dans quelle mesure pensez-vous que le prince Philippe sera un bon roi ?

	FLAMANDS	FRANCOPH.	ROYAUME
• Très bon	52,7 %	46,2 %	50,0 %
• Assez bon	30,6 %	31,8 %	31,1 %
• Pas assez bon	10,0 %	12,5 %	11,1 %
• Pas bon	4,7 %	2,8 %	3,9 %
• Sans opinion	29,7 %	6,6 %	3,9 %

Plus clairement:

	Flamands	Francoph	Royaume
Très bon	52,7%	46,2%	50,0%
Assez bon	30,6%	31,8%	31,1%
Pas assez bon	10,0%	12,5%	11,1%
Pas bon	4,7%	2,8%	3,9%
Sans opinion	29,7%	6,6%	3,9%
Total	127,7%	99,9%	100%

Erreur de “sens”

JT de la RTBF (12 janvier 2018)



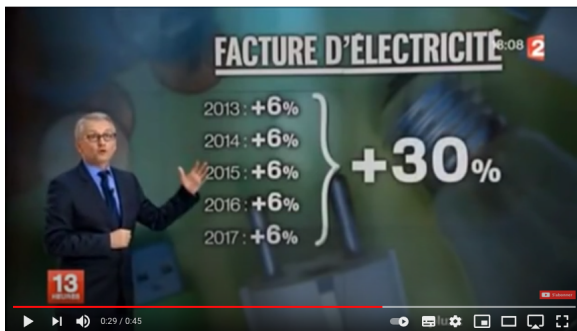
Annnonce d'une augmentation de 30% d'utilisateurs du système de leasing en un an sur base des éléments suivants:

- Janvier 2016: 7000 utilisateurs
- Janvier 2017: 10000 utilisateurs

D'autres erreurs/raccourcis sont possibles

JT de FR2 (janvier 2013,

<https://www.youtube.com/watch?v=z6OyOpDqhwk>)



Corrélation et causalité

Exemple “réputé” :

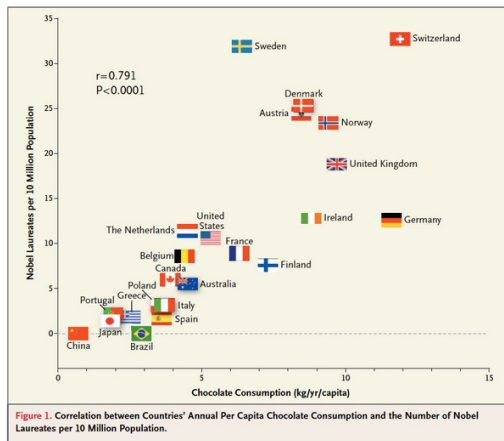


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

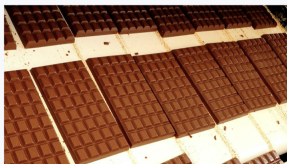
Correlation of chocolate consumption with Nobel Laureates (Image credit: New England Journal of Medicine)

Facteur confondant



ACCUEIL VIDÉO AUDIO MON CHOIX CHÂNES THÉMATIQUES PLUS

Consommation de chocolat et Prix Nobel: "aucun lien", selon des chercheurs



13 juin 2018 à 11:03 - mise à jour 13 juin 2018 à 13:18 - 1 min
Par Sofya Neum

Ils ont également établi une forte corrélation entre les magasins IKEA d'un pays et son nombre de Prix Nobel, pour démontrer par l'absurde que "le rapport entre deux variables ne signifie pas qu'un lien réel existe entre elles". Il serait en effet "farfelu de supposer que la nécessité de comprendre les instructions de montage des meubles IKEA augmente le niveau d'intelligence d'un pays", expliquent les scientifiques, pour qui le lien entre deux variables "dépend surtout d'une troisième variable".

Dans le cas du chocolat et du Prix Nobel, la troisième variable serait le PIB du pays. Ce qu'a observé le docteur Franz Messerli, auteur de l'essai paru en 2012, "pourrait donc être en partie expliqué par le fait que le chocolat est un produit de luxe et serait davantage consommé dans les pays où le développement économique est optimal" et donc propice à l'obtention de Prix Nobel, concluent les chercheurs.

L'étude parue l'an dernier avançait que les flavonoïdes, de puissants antioxydants contenus en grande quantité dans le cacao, le thé et le vin rouge amélioreraient les fonctions mentales.

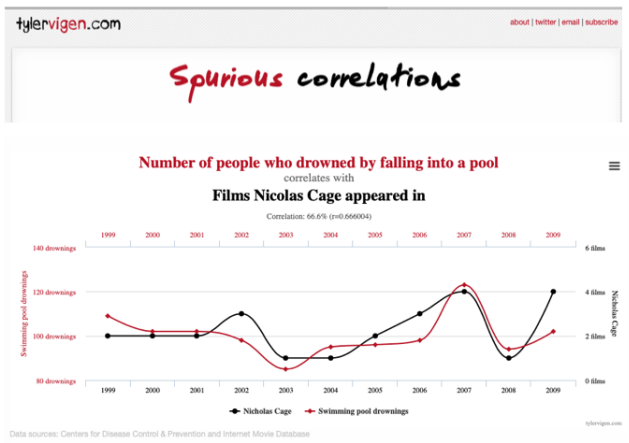


Nicolas Cage movies linked to drownings

► An analysis of data by a Harvard criminology student shows a clear correlation between the number of movies Nicolas Cage appears in each year and the number of people who drown in their swimming pools.



Analyse en question



Oups... un nouveau film sort en mai 2023

Renfield : Vidéos & bande annonce | Bande-annonce

FILMS / Cinémas / Tous les films / Horreur / Vidéos & bande annonce



Mise en garde dans l'article

And it doesn't seem to be a fluke, either: detailed analysis shows the correlation is statistically significant. Many scientists would take that as pretty good evidence that we're not dealing with random chance.

Before all this gets traction on social media (or Cage calls his lawyers), it should be made clear that it's all true – and all baloney.

That may sound paradoxical, and it highlights a problem in the use of statistics. There really is a clear, strong and statistically significant correlation between Mr Cage's annual output and deaths by falling into swimming pools (at least, in the US). And there is also no reason whatever for believing it to be true.

There is every reason for Mr Vigen winning an award for the most entertaining demonstration of one of the most important lessons in statistical science: correlation is not causation.

Hardly a week goes by without headlines proclaiming some “correlation” between one thing and another. Some of these make sense. There is, for example, a correlation between lung-cancer risk and number of cigarettes smoked.

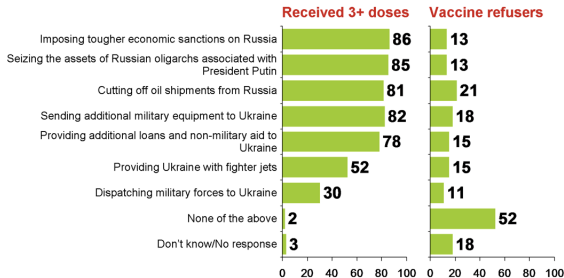
But statisticians routinely warn against mistaking mere correlation for a genuine, causative connection.

Lien entre l'opinion et le statut vaccinal des Canadiens

Le site *Ekos Politics* a mené une enquête le 18 mars 2022 auprès de 1035 Canadiens âgés de 18 ans ou plus.

Preferred response by vaccine acceptance

Q. How should Canada respond to the Russian invasion of Ukraine? Select all that apply.



BASE: Canadians; March 9-13, 2022. n=1,035, MOE +/- 3.1%, 19 times out of 20

Copyright 2022
No reproduction without permission

La presse locale s'est empressée de résumer les informations...

The screenshot shows the top section of the Toronto Star website. At the top left is the "TORONTO STAR" logo in a blue banner. To the right are buttons for "Subscribe Now" (Less than \$1/week) and "Sign In". Below these are links for "Newsletters", "Today's paper", and "Sports Bet". A navigation bar contains links for "HOME", "MY LOCAL", "GTA", "CANADA", "POLITICS", "WORLD", "OPINION", "LIFE", "SPORTS", "ENTERTAINMENT", "BUSINESS", "INVESTIGATIONS", and "PODCASTS".

The main article is titled "How vaccination status might predict views on the Russian invasion of Ukraine" and is marked as a "STAR EXCLUSIVE". The sub-headline reads: "New poll indicates that 'vaccine refusers are much more sympathetic to Russia.'" The author is identified as Grant LaFleche, a Standard Reporter, with the date "Sat., March 19, 2022" and a "5 min. read" indicator.

“Corrélation” entre le prénom et le résultat au BAC

Journal

LE FIGARO.fr
étudiant

Parcoursup | Études sup | Bac | Collège - Lycée | JPO | L'annuaire du sup | Vie étudiante

Bac 2018 : Garance, Juliette ou Diane ont plus de chance d'avoir une mention

Par Naïla Khelifi • Publié le 18/07/2018 à 13:06 • Mis à jour le 18/07/2018 à 16:11



Cette année, 25% des Garance ont obtenu la mention très bien. Crédits photo: CHARLY TRIBALLEAU/AFP

LE FIGARO L'abonnement numérique 4,90€ / mois

37 f t in

Depuis 2011, Baptiste Coulmont, sociologue, publie une étude sur la corrélation entre les mentions au bac et les prénoms. En fonction des prénoms des candidats et des résultats, il tire des conclusions surprenantes...

Quel est le lien entre le prénom et le résultat d'un candidat au bac? Passionné par ce que peuvent cacher les prénoms, Baptiste Coulmont, sociologue, a réalisé sur son [blog](#), une étude afin d'en tirer d'éventuelles corrélations entre le prénom des lycéens et la mention au baccalauréat. Comme l'écrit Baptiste Coulmont, dans son [article](#), «en tant qu'individu, elles sont toutes différentes. En tant que groupe (du simple fait de partager un prénom) elles sont semblables». Comme chaque année depuis 1991, ce sociologue a recensé les candidats ayant eu plus de 8/20 (qui permet d'aller aux rattrapages) au baccalauréat général et technologique en 2018. Cette année, 25% des Garance ont obtenu une mention très bien.

Les prénoms qui obtiennent le plus de mention « très bien » au bac

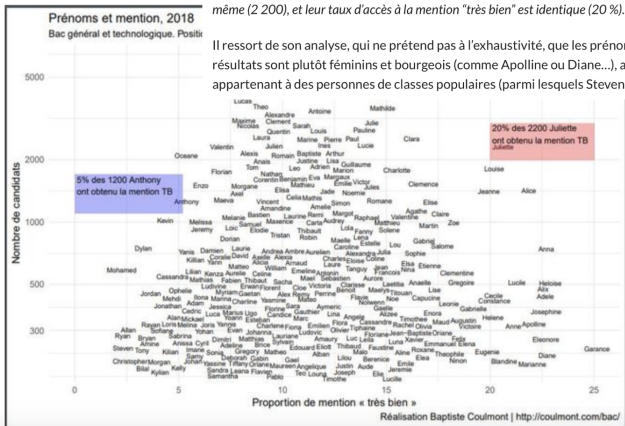
Vingt-cinq pour cent des Garance et 18 % des Augustin ont décroché la mention la plus prestigieuse au bac 2018, selon les travaux du sociologue Baptiste Coulmont, qui permettent de retrouver le taux d'accès associé à 335 prénoms.

Par Adrien de Tricornot

Publié le 19 juillet 2018 à 15h50 - Mis à jour

Ce qui est étonnant est que si les personnes ne sont pas les mêmes d'une année sur l'autre, les résultats sont similaires. Ce sociologue explique ainsi sur son blog: *"les Juliette qui ont passé le bac en 2017 ne sont pas celles qui ont passé le bac en 2018. Et même plus : les Juliette de 2017 n'ont pas les mêmes parents que les Juliette de 2018. Et pourtant leur nombre est presque le même (2 200), et leur taux d'accès à la mention "très bien" est identique (20 %).* ».

Il ressort de son analyse, qui ne prétend pas à l'exhaustivité, que les prénoms des personnes ayant obtenues les meilleurs résultats sont plutôt féminins et bourgeois (comme Apolline ou Diane...), alors que les moins bons sont masculins et appartenant à des personnes de classes populaires (parmi lesquels Steven, Ryan, Christopher ou Allan).



Etude longitudinale

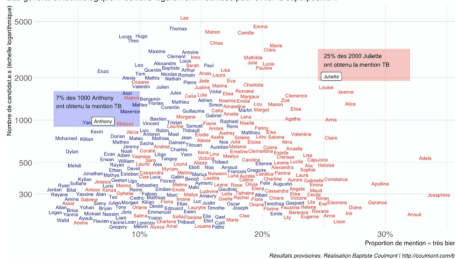
France : les filles prénommées Adèle, Joséphine et Anouk ont obtenu de beaux résultats au bac cette année



08 juil. 2020 à 11:57 - mise à jour 08 juil. 2020 à 11:57 - 2 min

Prénoms et mention, 2020

Bac général et technologique. Positions légèrement modifiées pour éviter la superposition.



Facteurs confondants:

Comme il le précise dans son article publié dès la fin des résultats du baccalauréat, réussite scolaire et choix des prénoms sont corrélés à l'origine sociale, le diplôme des parents ou encore le sexe de l'enfant, les filles réussissant mieux que les garçons.

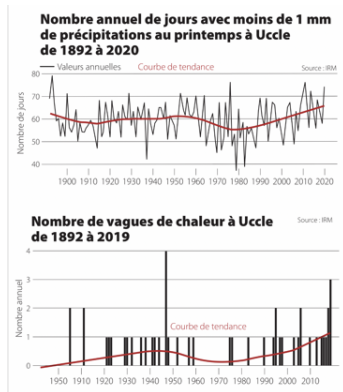
Graphiques

Des graphiques de “plus en plus sophistiqués” apparaissent dans les médias, avec, dans certains cas des notions sous-jacentes non triviales.

Ce sont les diagrammes en barres et en secteurs qui sont les plus utilisés mais des cartogrammes, des courbes temporelles, des nuages de mots... se retrouvent fréquemment dans les journaux.

Séries temporelles “continues” et “discrètes”

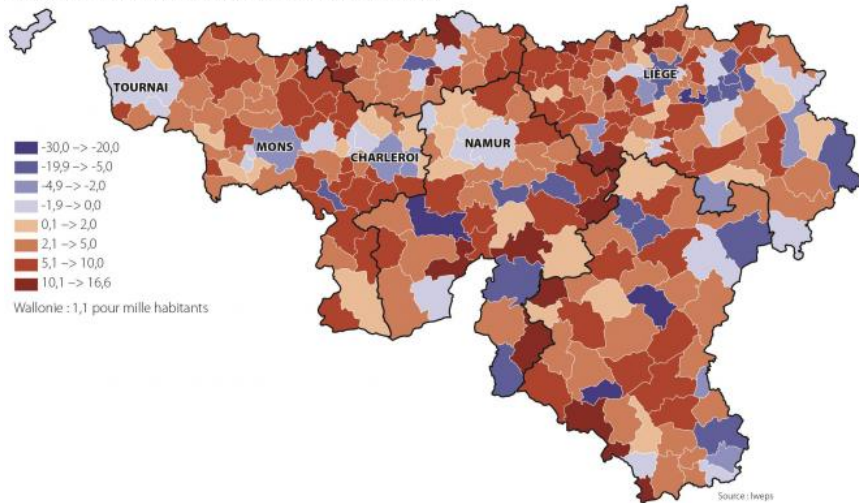
Le journal Le Soir du 9 août 2021



Cartogramme: Le Soir du 30 octobre 2022

Solde migratoire interne par an pour mille habitants

Moyenne annuelle sur 5 ans (du 01/01/2017 au 01/01/2022)

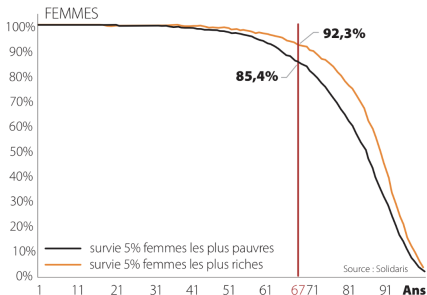
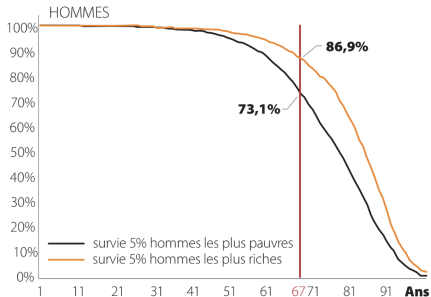


Courbes de survie

Le Soir du 9 février 2023

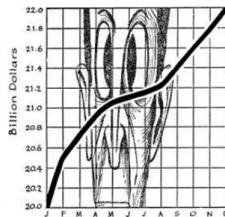
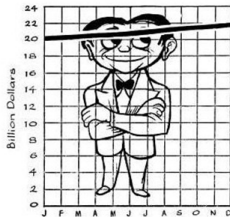
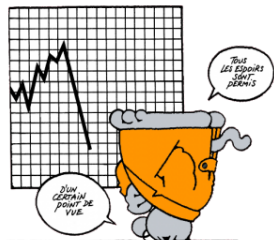
Courbes de survie

Courbes de survie (années) selon le sexe et le niveau socio-économique, 2014-2018

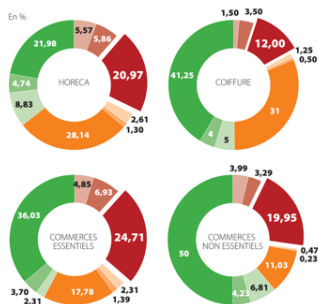


Graphiques “trompeurs”

Il est bien connu que le “point de vue” peut être trompeur:



Le Soir du 21 janvier 2021



Vingt types de commerces dans six villes

Le Soir a demandé à Graydon d'appliquer son modèle à une vingtaine de commerces (restaurants, bars, snacks, épicerie, poissonneries, boucheries, boulangeries, fromageries, cavistes, magasins de bricolage, jardinerie, librairies, marchands de journaux, magasins de vêtements, de chaussures, d'électro-ménager, de télécoms, parfumeries et salons de coiffure) en Région bruxelloise et dans cinq villes wallonnes (Charleroi, Liège, Louvain-la-Neuve, Mons et Namur). Nous avons ensuite regroupé ces commerces en quatre types (horeca, commerces essentiels et non essentiels, coiffure) avant de regrouper, en Wallonie, les cinq villes en une seule, comparable à Bruxelles. B.J.Y ET V.L.A.

Le Soir Lundi 26 juillet 2021

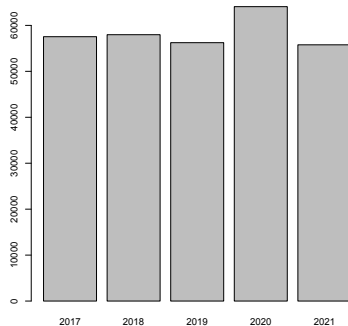
à la une 11

DÉCÈS

La mortalité à son niveau le plus bas

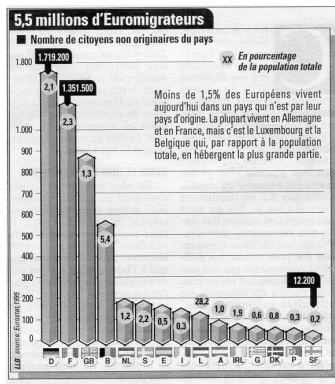
Les statistiques pour le premier semestre 2021 sont sorties. Elles montrent un recul des décès après le pic de 2020 provoqué par le covid. Leur nombre est le plus bas enregistré au cours des cinq dernières années.

L'évolution des décès sur les six premiers mois de l'année



La Libre Entreprise du 4 septembre 1999

Les effets de la libre circulation en Europe sont discutés à partir du graphique ci-dessous:



Le Soir Mercredi 8 février 2017

4 LABELGIQUE

Le niveau de vie baisse pour les plus défavorisés

BIEN-ÊTRE Niveau de vie, travail, santé : les plus pauvres sont toujours plus mal lotis

- Le rapport du Bureau fédéral du Plan a été présenté mardi.
- Il n'est pas dénué de bonnes nouvelles, sauf pour les plus pauvres.

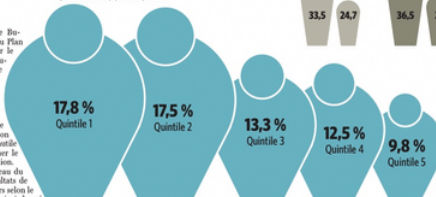
Depuis 2014, le Bureau fédéral du Plan est chargé par le gouvernement de mesurer la qualité de la vie en Belgique. Pour ce faire, il lui a été demandé d'établir une série d'indicateurs complémentaires au PIB qui, s'il s'intéresse au niveau de production d'un pays, reste assez inutile quand il s'agit d'évaluer le bien-être de sa population. Cette année, le Bureau du Plan a ventilé les résultats de treize de ces indicateurs selon le revenu. Et le tableau ainsi dressé est peu réjouissant.

Si l'écart de bonheur entre les 20 % de Belges les plus pauvres et les 20 % les plus riches (notamment dit entre les quintiles 1 et 5, ainsi que le rapport de richesse) est

Population obèse selon la catégorie de revenu

Pourcentage de la population de 18 ans et plus

Chaque quintile représente 20 % de la population globale, ils sont classés ici, sur base de leurs revenus. Ainsi, le premier quintile représente les 20 % de Belges avec le plus faible revenu. Le second quintile représente la moyenne basale. Le troisième groupe dispose d'un revenu moyen. Le quatrième quintile représente les 20 % de population disposant d'un revenu plus élevé, le cinquième quintile représente les 20 % de population les plus riches.



disposer d'un poste de télévision (alors qu'elle le souhaite), ne pas avoir de téléphone (alors qu'elle le souhaite), ne pas avoir de machine à laver (alors qu'elle le souhaite) et

plus roses. L'obésité, si elle concerne 13,7 % de la population, touche environ 17 % du premier quintile et un peu moins de 10 % des 20 % de Belges les plus riches.

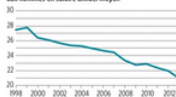
Espérance de vie à 50 ans selon le sexe et le niveau d'éducation

En années de vie après 50 ans



Ecart salarial entre les hommes et les femmes

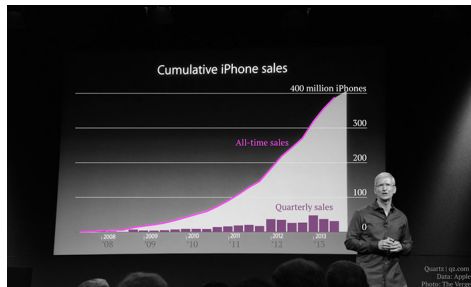
Pourcentage de différence des femmes par rapport aux hommes en salaire annuel moyen



Les inégalités hommes-femmes se réduisent

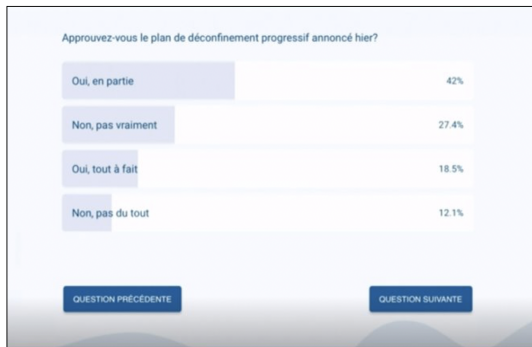
Utilisation de courbes cumulatives

En septembre 2013, Tim Cook a présenté l'évolution des ventes d'iphones à l'aide d'un graphique cumulé:

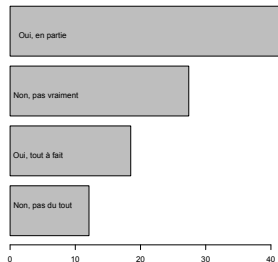


... pour ne pas (selon ses détracteurs) montrer un graphique trimestriel illustrant une baisse des ventes.

Sondage présenté au JT de la RTBF du 25 avril 2020: Approuvez-vous le plan de déconfinement progressif annoncé hier?

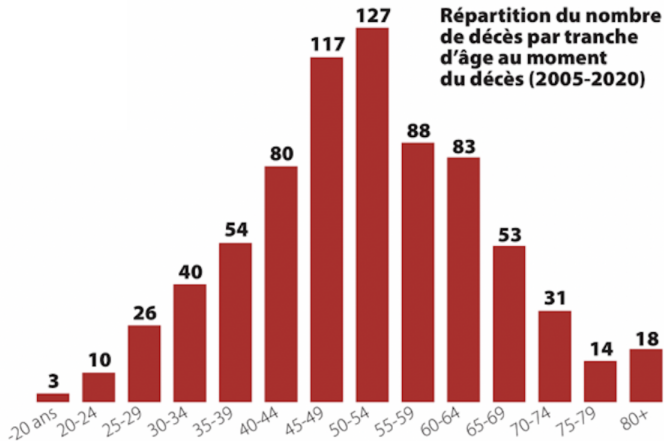


“Zoom”



Distribution des âges (Le Soir le 3 novembre 2021)

L'article s'intéressait à l'âge des 744 personnes sans-abris décédées dans la rue à Bruxelles entre 2005 et 2020.



Des erreurs "flagrantes" sont aussi possibles

Reportage sur BFMTV (25 avril 2018)



Et tant d'autres choses à dire concernant les statistiques dans la presse

- A propos des sondages (même si, globalement, la présentation des résultats s'est nettement améliorée ces dernières années, avec, assez souvent, la prise en compte de la marge d'erreur dans l'interprétation des résultats)
- A propos du calcul d'une moyenne ou d'une médiane
- ...

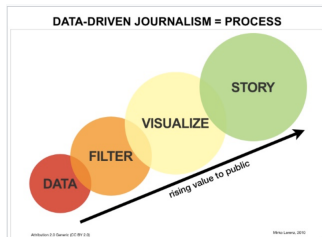
La plupart des statistiques présentées dans la presse sont “triviales” mais le citoyen “moyen” est-il suffisamment formé pour les comprendre et les interpréter?

Il est clair qu'il faut améliorer la *littératie statistique* dès l'école...

Le “data journalism” arrive...

“Data journalism” = mouvement visant à renouveler le journalisme en se focalisant sur le traitement et la présentation de données (idéalement publiques). L’objectif est de “raconter” l’information différemment.

Ce mouvement se développe (notamment et/ou plus rapidement dans les pays anglosaxons). Cela implique le besoin de compétences additionnelles pour les journalistes, mais également pour les citoyens...



Et plus globalement dans nos vies quotidiennes...

