

Graphique mon beau graphique, dis-moi ?

Benoît Jadin, Camille Kroonen, Juliette Petitjean

Mots clés : Graphiques, échelles, indices, affinités, perspective, invariants, boîtes à moustaches, valeurs centrales : sens et positions relatives.

Résumé. *Comment faire des graphiques en bâtonnets, circulaires et cartésiens pour mieux tromper ? Ou, à rebours, comment déjouer les pièges et comment interpréter des graphiques pour y voir bien plus que ce qu'ils ne semblent montrer ?*

Les graphiques c'est comme les pesticides, on s'en méfie mais on ne peut s'en passer. On s'en méfie parce qu'ils ne montrent qu'une partie de la réalité qu'ils illustrent, ou qu'ils n'en font voir que les belles facettes, ou qu'ils sont tout simplement trompeurs. On peut essayer de faire sans eux, commenter la réalité et donner des tableaux, mais c'est tellement plus rapide, plus explicite et plus facile à comprendre avec un graphique. On peut penser à des « graphiques écologiques » qui respecteraient certaines règles éthiques mais ce sont surtout les capacités critiques citoyennes des lecteurs enfants, adolescents et adultes qu'il faut développer.

1. Marine Le Pen et le Frexit

Le 20 mars 2017, lors du débat entre cinq candidats à la présidentielle diffusé par TF1, « au moment où ses adversaires lui ont fait remarquer que le Brexit n'avait pas encore produit ses effets, la candidate du Front National a montré aux caméras un graphique (figure 1) censé prouver que la mise en place de l'euro en 2002 a provoqué une dégringolade de la production industrielle en France. [...] L'image aurait été diffusée pour la première fois le 12 février par un membre de l'équipe de campagne de Marine Le Pen, sur Twitter. Elle a ensuite été partagée le 14 février par Louis Aliot, vice-président du FN, puis plus largement relayée. » ⁽¹⁾

Pour savoir si Marine a menti, il faut d'abord savoir de quoi parle ce graphique. Selon le titre et les drapeaux, il s'agit de la production industrielle dans quatre pays européens et la source des données serait l'OCDE. Un petit détour par cette dernière et nous apprenons que « la production industrielle désigne la production des entités industrielles et recouvre des secteurs tels que l'extraction minière, les activités manufacturières et les services publics collectifs (électricité, gaz et eau). Cet indicateur prend la forme d'un indice qui exprime les variations de volume de production par rapport à une période de référence. » ⁽²⁾ S'il s'agit d'un indice, c'est qu'on s'intéresse à la croissance de la production au fil du temps et qu'on considère donc des valeurs relatives par rapport à une année de référence (2001 pour ce graphique). Si pour un pays donné, on note $p(t)$ la production annuelle pour une année t , alors $i_n(t) = \frac{p(t)}{p(n)}$.

1. Le parisien en ligne à l'adresse <http://www.leparisien.fr/elections/presidentielle/debat-que-dit-le-graphique-brandi-par-marine-le-pen-21-03-2017-6781617.php>. Consulté le 28 février 2018

2. Site de l'OCDE à l'adresse <https://data.oecd.org/fr/industry/production-industrielle.htm>. Consulté le 28 février 2018

Graphiques

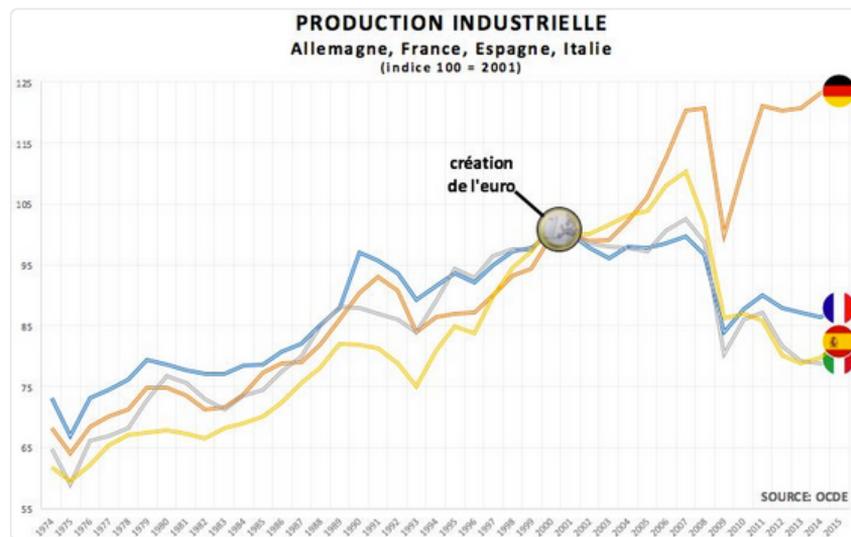


Fig. 1 : Le graphique de Marine

Est-il vrai qu'entre 2001 et 2015, la production industrielle a chuté en France ? Est-ce la faute de l'euro ? On constate assez rapidement qu'une chute importante a eu lieu en 2008 pour les quatre pays considérés. On devine qu'il s'agit des effets de la crise des subprimes...

1.1. Piège graphique

Intéressons nous d'abord aux échelles. À partir de données de la banque mondiale ⁽³⁾, nous avons reconstruit un graphique en considérant pour l'axe des ordonnées une échelle d'origine 0 et pour l'axe des abscisses une échelle d'origine 1995 (figure 2). En effet, quel sens ont des données avant cette date quand on connaît le recul actuel de l'industrie classique au profit des multinationales de la communication, des réseaux et des données comme les GAFAM ⁽⁴⁾, par exemple ?

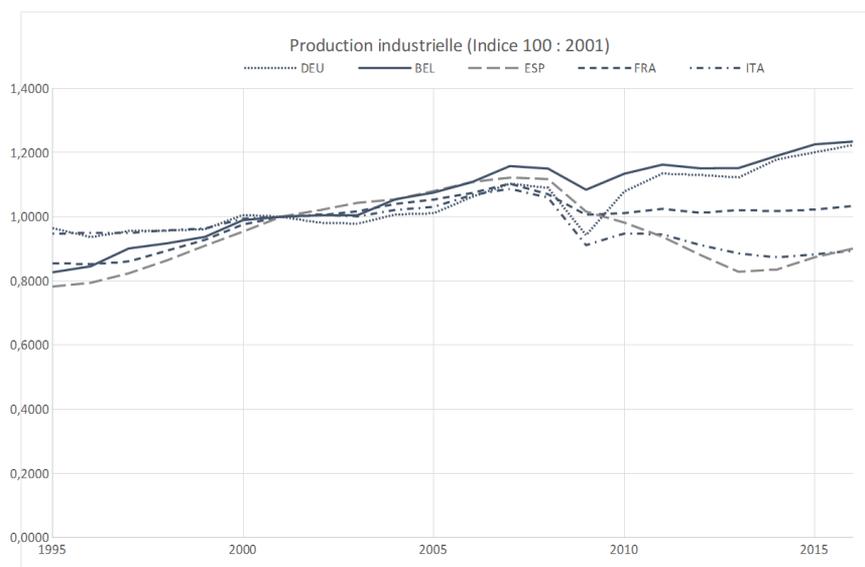


Fig. 2 : Changement d'échelles

La situation de la France paraît déjà nettement moins désespérée.

3. Site de la Banque Mondiale à l'adresse <http://databank.banquemondiale.org/data/reports.aspx?source=2&series> consulté le 28 février 2018

4. Google, Apple, Facebook, Amazon, Microsoft

Graphiques

Un effet trompeur pourrait aussi provenir du choix de l'année de référence pour le calcul de l'indice ? Nous en avons donc choisi une autre, 2011, postérieure à la crise des subprimes (figure 3). On a surtout l'impression que la catastrophe est espagnole et non française. La production industrielle a progressé en France entre 2001 et 2007 puis a chuté en 2008 et en 2009, pour stagner ensuite à partir de 2010.

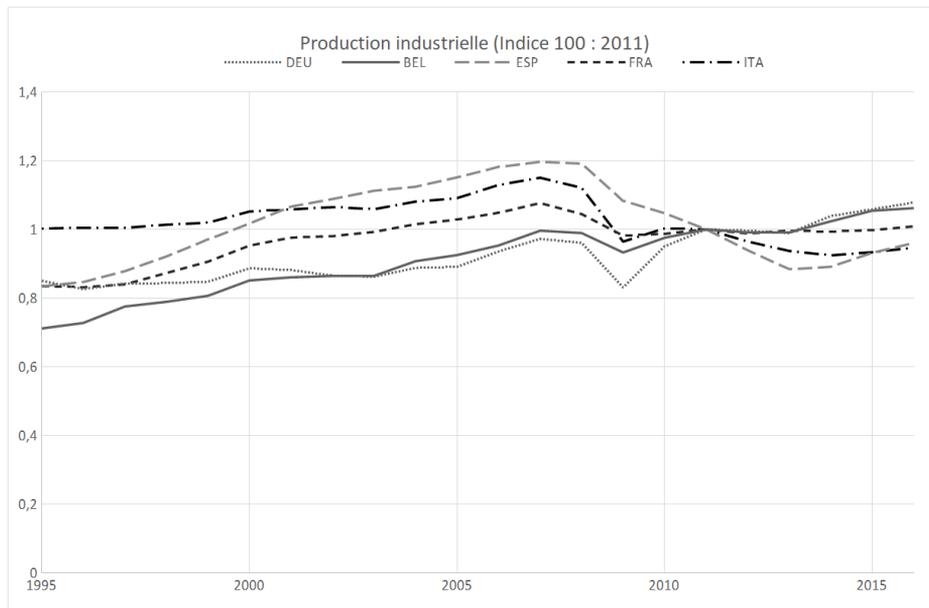


Fig. 3 : Changement de l'année de référence

1.2. Piège de l'indice

Et si le problème venait du fait qu'on considère des indices ? Intéressons-nous (figure 4) donc aux productions absolues (et non relatives à une année de référence). L'Allemagne caracole en tête tandis que la Belgique est bien bas. Entre les deux, la France maintient une certaine constance.

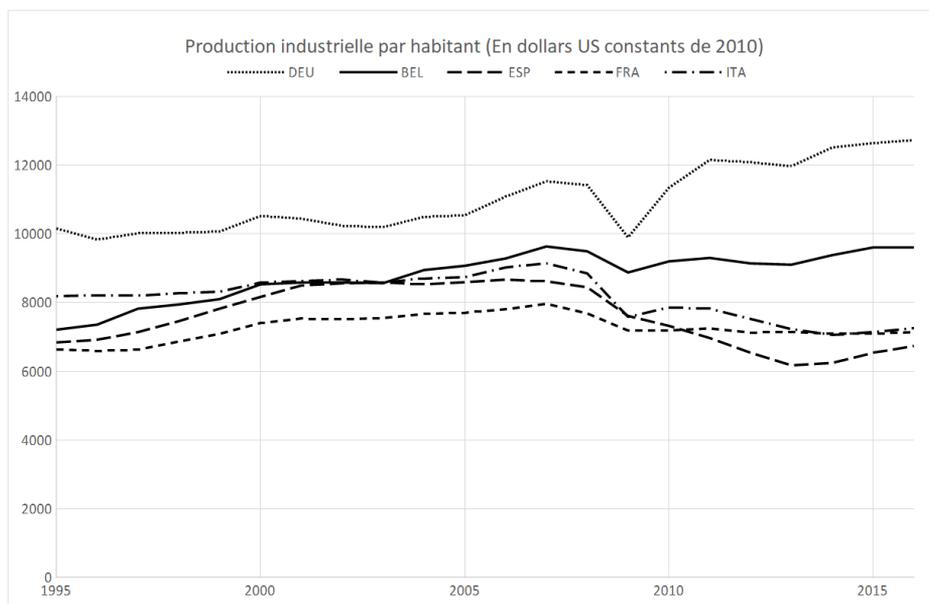


Fig. 4 : Productions absolues

Graphiques

Les valeurs sont exprimées en dollars constants, ce qui élimine toute variation due à l'inflation ou à la déflation. En dollars courants⁽⁵⁾, on pourrait donc s'attendre à des fluctuations bien plus marquées sachant que celles-ci seraient les mêmes pour tous les pays considérés puisqu'ils ont une même monnaie, l'euro.

Entre 2007, les cinq pays considérés vont connaître une chute qui s'accroît encore l'année suivante avant de repartir, à l'exception de l'Espagne, à la croissance en 2010... Pour connaître de nouveaux problèmes en 2011, 2012. On peut se demander si toute croissance de la production se traduit par une croissance de l'indice référent à une année donnée ou quelle que soit l'année de référence choisie. Nous y reviendrons plus loin.

Une autre question se pose... Peut-on comparer des pays de tailles aussi différentes que l'Allemagne et la Belgique? Il faut donc considérer la production par habitant (figure 5). On constate la stagnation de la France qui ne s'est pas vraiment relevée après la crise des subprimes et revient en 2016 à un niveau comparable à celui de 1996.

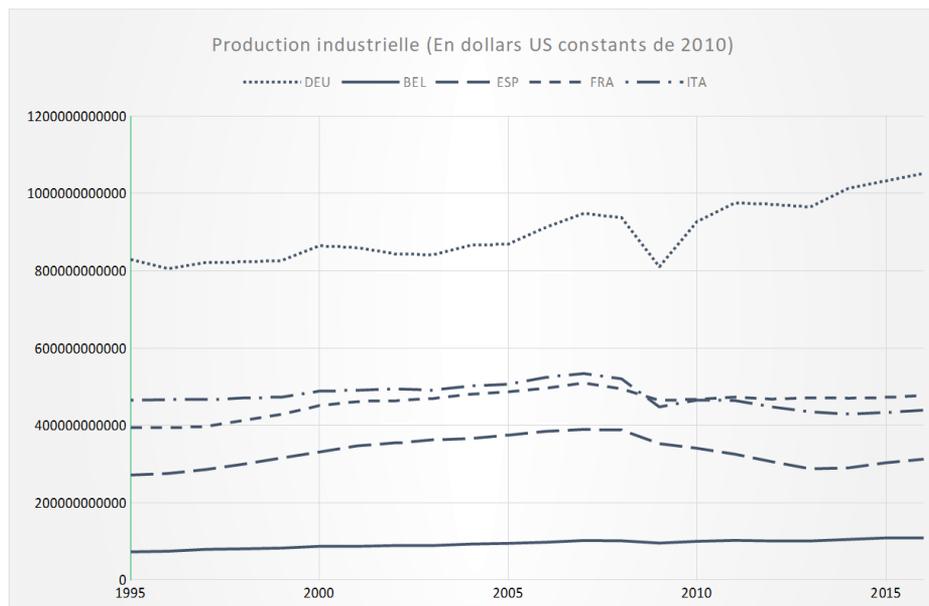


Fig. 5 : Production par habitant

1.3. Et plus si affinité

Que peut-on dire, pour un pays donné, du graphique de variation de l'indice de production par rapport au graphique de variation des productions données en valeurs absolues? De l'équation $i_n(t) = \frac{p(t)}{p(n)}$, on peut conclure que, par rapport au graphique de $p(t)$, celui de $i_n(t)$ subit une affinité d'axe Ot (ou axe des abscisses) parallèlement à l'axe Op (ou axe des ordonnées) et de rapport $\frac{1}{p(n)}$.

Considérons la Belgique. Pour passer du graphique de $p(t)$ à celui de $i_{2011}(t)$, il faut une affinité de rapport $\frac{1}{p(2011)}$. Pour passer du graphique de $p(t)$ à celui de $i_{2001}(t)$, il faut une affinité de rapport $\frac{1}{p(2001)}$. Pour passer du graphique de $i_{2001}(t)$ à celui de $i_{2011}(t)$, il faut une affinité de rapport $\frac{p(2001)}{p(2011)}$. Ce qui peut être perturbant pour la lecture du graphique, c'est que ce dernier rapport est propre à chaque pays. Quand on passe d'une année de référence à une autre, toutes les courbes ne vont

5. Pour passer des dollars courants aux dollars constants à un moment donné, on divise la valeur courante par l'indice des prix à ce moment.

Graphiques

donc pas subir la même affinité mais chaque graphique va être transformé différemment par un rapport d'affinité qui lui est propre, comme le montre clairement la figure 6. Pour la clarté, nous avons seulement considéré deux pays, la Belgique et l'Espagne.

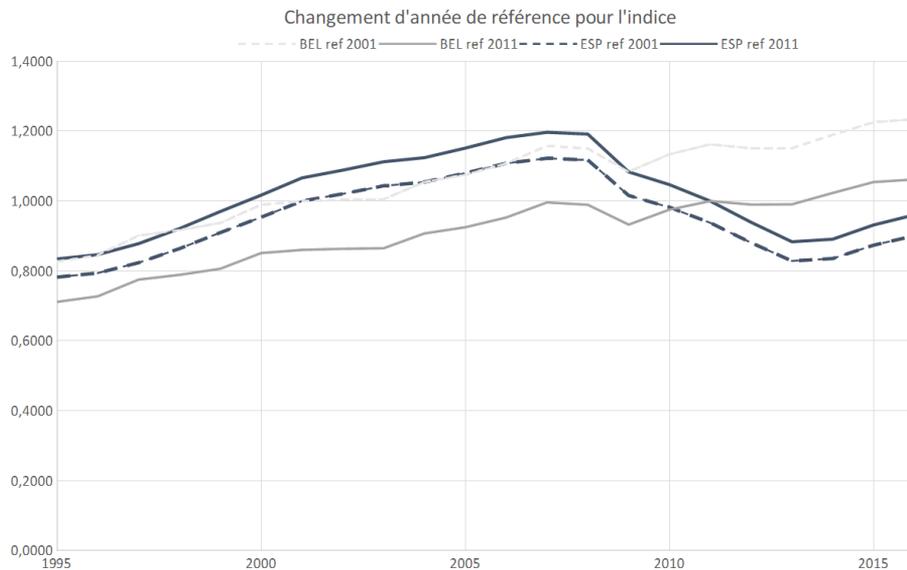


Fig. 6 : Production par habitant

Par ailleurs, entre les graphiques de production et d'indice (réfèrent à une année donnée), l'affinité qui permet le passage de l'un à l'autre à l'autre, n'affecte ni les zones de croissances et de décroissance ni les extréma locaux, ces deux fonctions étant multiples l'une de l'autre et leur dérivées égales à une constante près.

2. Benoît Lutgen et le PSexit

Le 19 juin 2017, le président du Cdh décidait de débrancher la prise du gouvernement wallon : « Benoît Lutgen demande au MR, Ecolo et Défi de former une majorité sans le PS aux gouvernements wallon, bruxellois et de la Communauté française. Le parti humaniste ne veut plus gouverner avec le parti socialiste. » ⁽⁶⁾

À l'annonce de cette décision, les journalistes et autres analystes politiques ont rappelé l'état des forces en présence. Depuis les élections de 2014 et jusqu'aux élections suivantes en 2019, le Cdh dispose de 17 sièges sur 75 au Parlement Wallon, comme le montrent les figures 7 et 8 au travers de graphiques circulaires.

2.1. Et plus si affinités

Le graphique de la figure 8, construit à partir des mêmes données que celui de la figure 7, est une perspective axonométrique, non d'un disque, mais d'un cylindre. Cela a visiblement pour conséquence de donner plus d'importance au Cdh que ce que lui donne le graphique de la figure 7 ? Mais n'est-ce qu'une impression ? Quelle transformation permet le passage de l'un à l'autre ? Avec quels effets ?

6. Eric Deffet, Le Soir en ligne du 19/06/2018 à l'adresse <http://www.lesoir.be> consulté le 10 avril 2018.

Répartition des sièges au Parlement Wallon 2014-2019

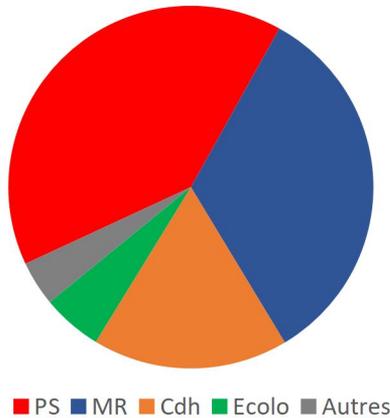


Fig. 7 : Le poids du Cdh

Répartition des sièges au Parlement Wallon 2014-2019

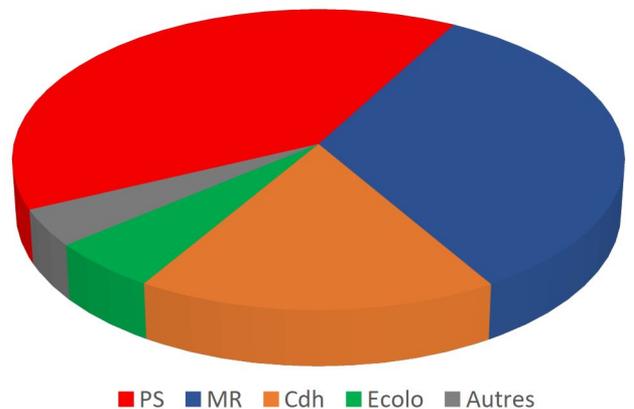


Fig. 8 : Le poids du Cdh

Considérons tout d'abord le disque. Il est transformé en ellipse, image affine d'un cercle. On constate que les amplitudes d'angles ne sont pas conservées (figure 9), certaines subissent une augmentation et d'autres une diminution. Par contre, comme il s'agit d'une affinité, les aires des secteurs angulaires du cercle et de l'ellipse sont proportionnelles entre elles dans un rapport qui correspond au rapport d'affinité. On peut s'en convaincre intuitivement en quadrillant finement le disque avec une famille de lignes parallèles à l'axe de l'affinité et l'autre de lignes perpendiculaires. Les carrés du quadrillage initial sont tous transformés en rectangles isométriques dont une dimension n'a pas changé tandis que l'autre a subi l'affinité. En passant du disque à l'ellipse, les rapports d'aires des secteurs angulaires n'ont donc pas changé.

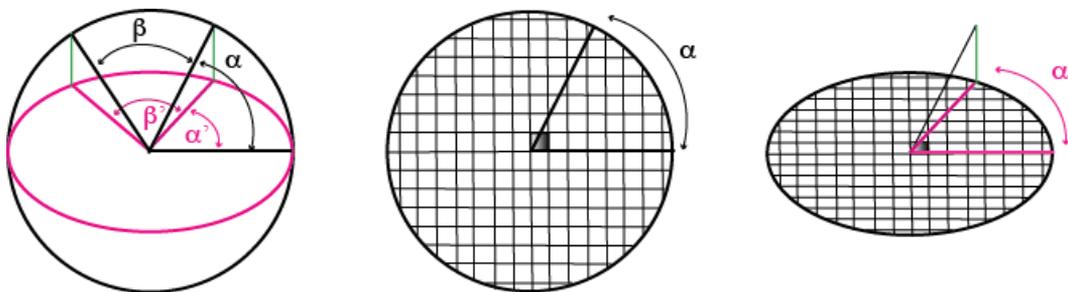


Fig. 9 : Conservation des aires et non des angles

Par contre, si on prend en compte l'épaisseur du cylindre représenté, on constate que la différence de traitement est très nette entre les secteurs situés en arrière et ceux situés en avant : pour ceux-ci l'impact visuel est renforcé par la tranche visible du cylindre comme on le perçoit au travers des deux graphiques des figures 8 et 10 qui correspondent à la même réalité : le premier met en évidence le Cdh et le second, le PS.

2.2. 3D déformante

Plutôt qu'une perspective axonométrique du cylindre (figure 10), on peut considérer une perspective classique (figure 11). Dans ce dernier cas, le centre du disque n'est pas conservé. C'est-à-dire qu'il n'est plus centre de symétrie de l'ellipse. Ce qui a pour effet, dans le cas présent d'augmenter l'aire relative associée au PS...

Graphiques

Répartition des sièges au Parlement Wallon 2014-2019

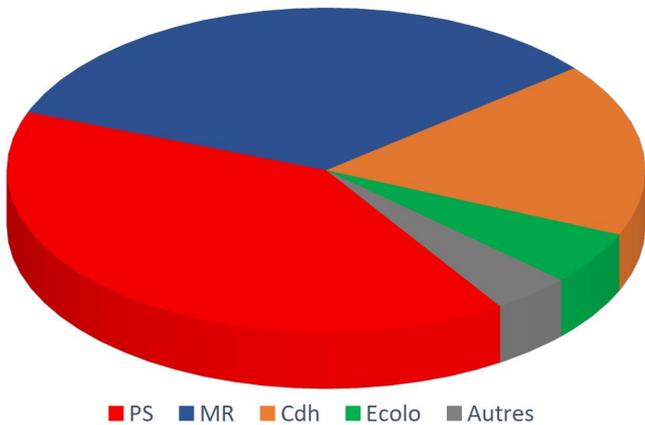


Fig. 10 : Le poids du PS

Répartition des sièges au Parlement Wallon 2014-2019

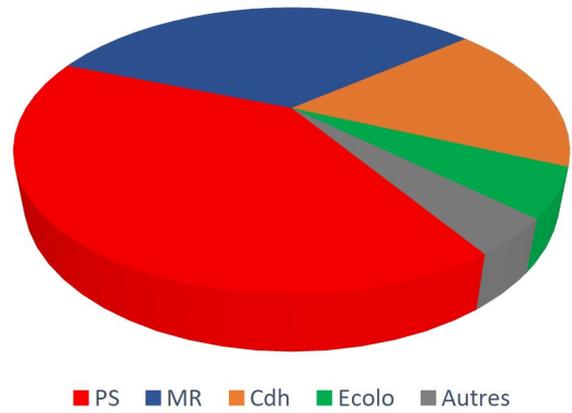


Fig. 11 : Le cauchemar du Cdh

3. Nous et notre exit

« En 2016, l'espérance de vie à la naissance en Belgique s'établissait à 83,8 ans pour les femmes et à 79,0 ans pour les hommes. Pour l'ensemble de la population, elle s'établissait à 81,4 ans. La hausse de l'espérance de vie à la naissance par rapport à 2015 pour les deux sexes réunis est de 0,5 an. » (7)

Pour y voir clair nous avons consulté les tables de mortalité établies par l'institut national de statistique belge, à savoir Statbel et nous avons superposé les graphiques des femmes et des hommes (figure 12).

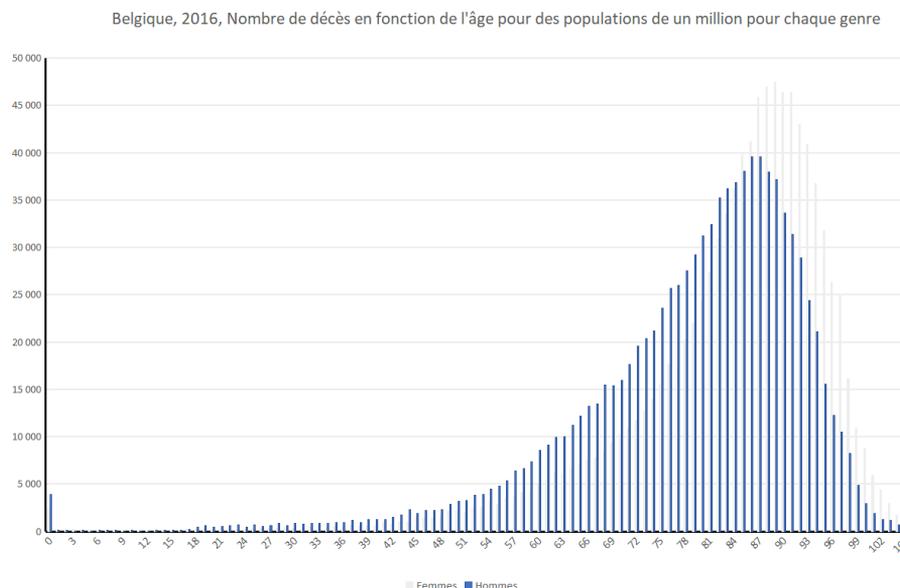


Fig. 12 : La nature est injuste

Bien sûr, on constate le décalage entre les deux genres. Cela ne fait aucun doute, la nature paraît injuste avec les hommes qui meurent plus jeunes. Mais d'autres questions se posent. Pourquoi la distribution de décès des femmes (exprimée en « pour millions » et non en pourcents) est-elle

7. Tables de mortalité et espérance de vie, à l'adresse <https://statbel.fgov.be/fr/themes/population/mortalite-et-espérance-de-vie/tables-de-mortalite-et-espérance-de-vie> consulté le 7 août 2018.

plus resserrée que celle des hommes ? Pourquoi sont-ce des allures de courbes en cloche ? Pourquoi ces « cloches » s'étirent-elles vers la gauche du graphique ? Nous ne répondrons pas à toutes ces questions, certaines demandent un travail assez long tandis que d'autres se verront éclairées au cours de notre cheminement.

3.1. Décédés ou survivants

Le graphique de la figure 13 fait apparaître les fréquences cumulées de mortalité toujours exprimées en « pour millions ». L'écart entre les deux courbes considéré du point de vue des abscisses (ordonnées fixes) est variable mais est proche de cinq années, en faveur des femmes, sur un intervalle assez grand (entre 60 ans et 90 ans). Les deux fonctions représentées sont forcément croissantes puisqu'on cumule des fréquences par nature positives. On constate dans les deux cas, une concavité vers le haut puis une concavité vers le bas quand on va dans le sens des âges croissants. Le changement de concavité n'est pourtant pas facile à déterminer avec précision, quoi correspond-il ? Une concavité vers le haut correspond à une croissance de plus en plus forte ; une concavité vers le bas, à une croissance de plus en plus faible. Le passage de l'une à l'autre est donc associé à un maximum, à savoir le mode de la distribution en termes statistiques. À partir de la figure 12, on peut être sûr que le changement de concavité est en 89 ans pour la courbe des femmes et en 87 ans pour la courbe des hommes.

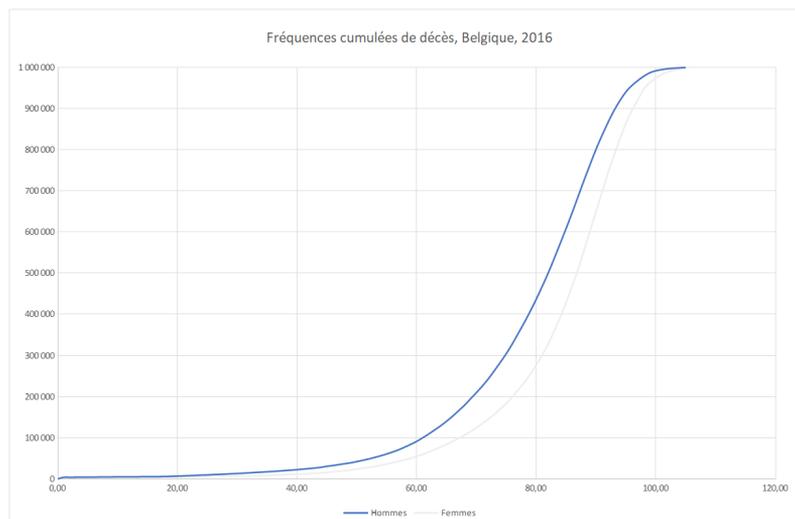


Fig. 13 : Du côté des décès

Le lecteur attentif aura remarqué que le traitement statistique a été fait en données non groupées sur deux populations belges⁽⁸⁾, les femmes et les hommes, la variable considérée étant l'âge exact⁽⁹⁾ de décès. Les fréquences cumulées sont donc des fonctions en escalier. Nous les avons lissées et transformées en fonctions continues...

À partir des courbes de fréquence cumulées, on détermine facilement les quartiles dont la médiane. On construit ensuite la boîte à moustache comme nous le montre la figure 14. On perçoit rapidement au travers de ces boîtes, l'écart entre les sexes ainsi que la dispersion moins grande chez les femmes.

8. Il s'agit, comme nous le verrons plus loin, de populations fictives.

9. On considère pour ce faire, l'écart entre l'année de décès et l'année de naissance.

Graphiques

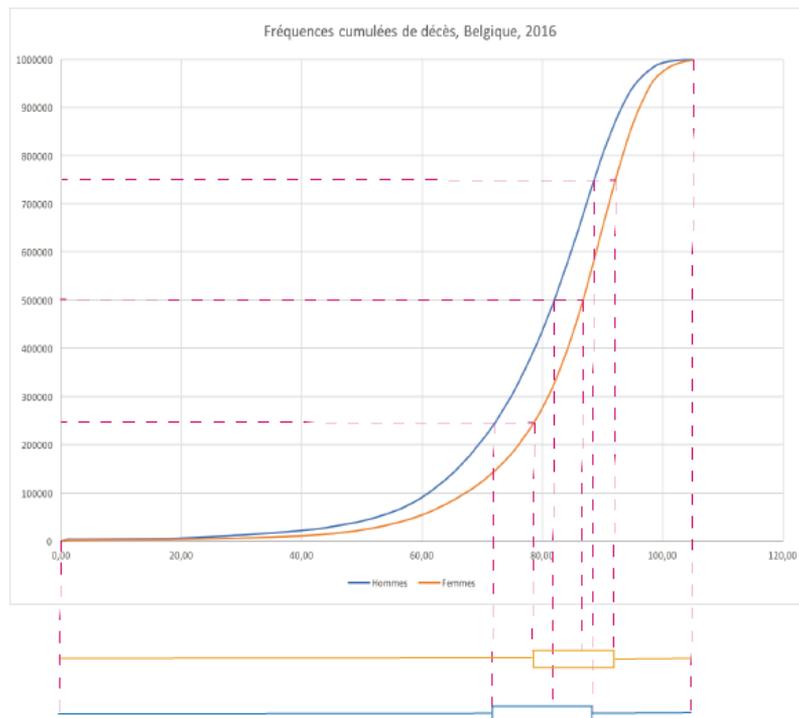


Fig. 14 : La boîte à moustaches

Plutôt que de considérer les fréquences cumulées de décès, on peut considérer les fréquences de survivants. C'est ce que nous montre le graphique de la figure 15, toujours avec des populations de référence d'un million d'individus. C'est ce qu'a fait Christian Huygens dans de longs échanges qu'il a entretenus avec son frère Louis à propos des tables de mortalité de John Graunt établies pour Londres au début du dix septième siècle. Louis pose la question « *jusqu'à quel aage doit vivre naturellement un enfant aussi tost qu'il est conçu ?* »⁽¹⁰⁾ Alors que Louis fait un calcul d'espérance basé sur la moyenne, Christian considère la courbe des survivants et se demande à partir de quel âge, la moitié de la population est encore en vie. Il fait naturellement le pari d'une chance sur deux pour qu'une personne vive encore à cet âge-là. Il passe aux probabilités et fait un choix de médiane comme espérance.

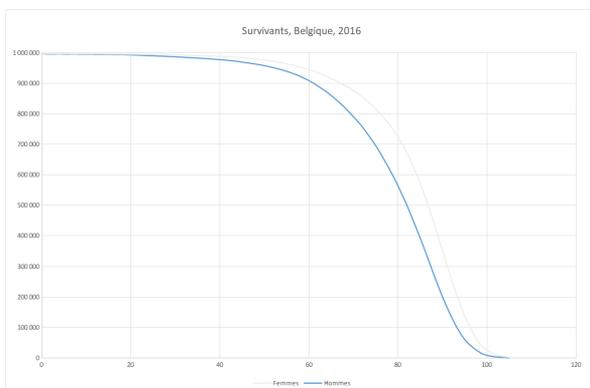


Fig. 15 : Du côté des survivants

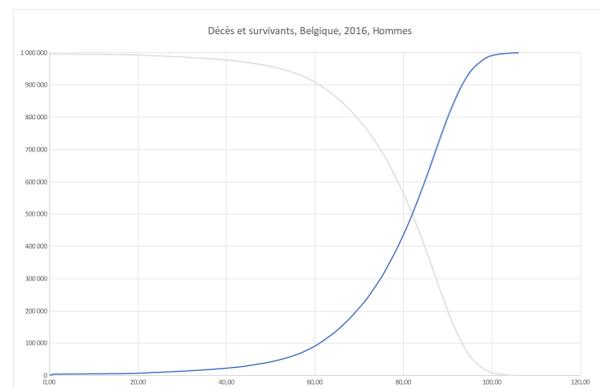


Fig. 16 : En miroir

Si on note $d(t)$ la fonction de fréquence cumulée de décès en fonction de l'âge, et $s(t)$ la fréquence de survivants, on peut écrire que $s(t) = 1 - d(t)$ ou encore $s(t) = 1000000 - d(t)$ si on les exprime

10. Oeuvres complètes de Ch. Huygens, Tome 6 sur Gallica à l'adresse <https://gallica.bnf.fr/ark:/12148/bpt6k778547.texteImage> consulté le 10 août 2018.

en « pour millions ». Les deux courbes (figure 16) sont images l'une de l'autre par une symétrie orthogonale d'axe parallèle à l'axe Ot et d'ordonnée $1/2$ (ou 500000).

3.2. Valeurs centrales

À la figure 17, on a considéré la population des hommes en données groupées par classe de 5 ans. Comme les classes sont de même largeur, le graphique en bâtonnets (perception basée sur les hauteurs des bâtonnets) est aussi un histogramme (perception basée sur les aires des bâtonnets). La classe modale apparaît directement, elle correspond au maximum des fréquences, il s'agit de la classe $[85,90[$. Par ailleurs, le graphique de la figure 12, nous montrait un mode en 87 ans.

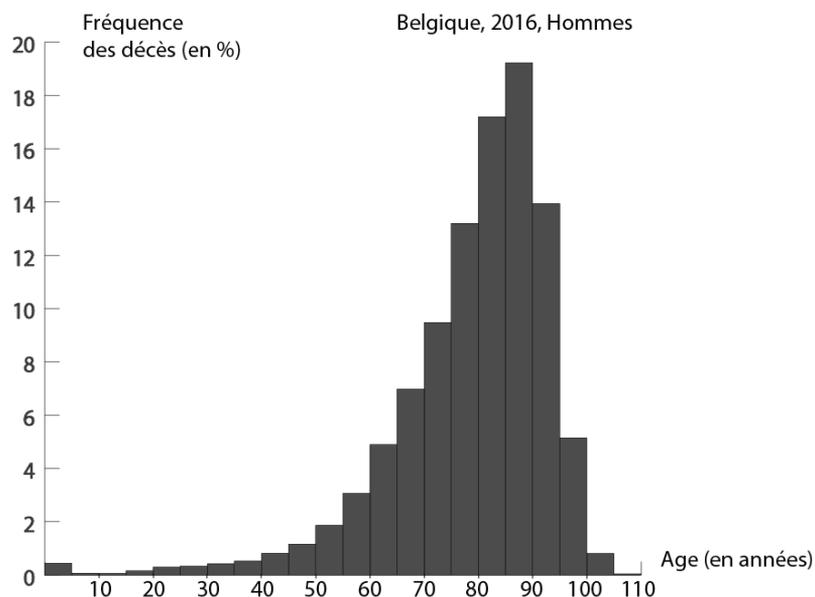


Fig. 17 : Histogramme des décès

Sur le graphique, la médiane partage l'aire totale en deux parties égales. Si la distribution des fréquences était symétrique, la médiane correspondrait au mode. Mais comme dans le cas présent, l'aire en dessous du mode est visiblement plus grande que l'aire au dessus, la médiane est inférieure au mode. La figure 14 permet de la déterminer avec plus de précisions, elle vaut 82 ans.

Sur l'histogramme, si on considère l'axe des abscisses comme une tige horizontale et chaque bâtonnet comme un poids exercé sur cette tige horizontale en un point donné, la moyenne correspond au point d'appui qui assure l'équilibre de la tige. Celui-ci est fonction des poids et des « bras de levier ». Quand la distribution est symétrique, le point d'équilibre est au centre de la distribution. La moyenne correspond alors au mode et la médiane. Par contre, d'une part, si les valeurs restent les mêmes mais que les poids changent, les modifications entraînent la moyenne là où les poids se font plus importants. D'autre part, si les poids restent les mêmes mais que ce sont les valeurs qui changent, la moyenne suit ce déplacement des valeurs. Dans le cas présent, on se rend compte que par rapport à la médiane qui divise les poids en deux parties égales, la moyenne est tirée vers le bas par ses plus petites valeurs. Un calcul plus précis à partir des tables elle-mêmes permet de calculer une moyenne de 78,29 ans.

Sachant que le graphique de distribution des fréquences de mortalité en fonction de l'âge n'est pas symétrique mais « s'étire vers la gauche », cela conduit à une situation où la moyenne est inférieure

Graphiques

à la médiane, elle-même inférieure au mode (figure 18). Dans une situation de revenus⁽¹¹⁾ comme celle de la figure 19, c'est tout le contraire : le graphique de distribution « s'étire vers la droite » et la la moyenne est supérieure à la médiane, elle-même supérieure au mode.

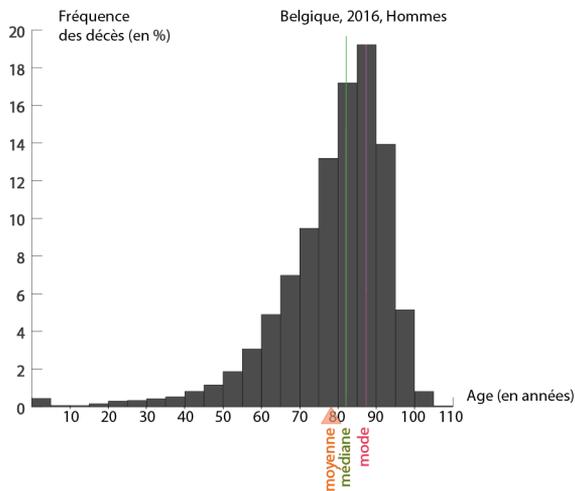


Fig. 18 : Moyenne < Médiane < Mode

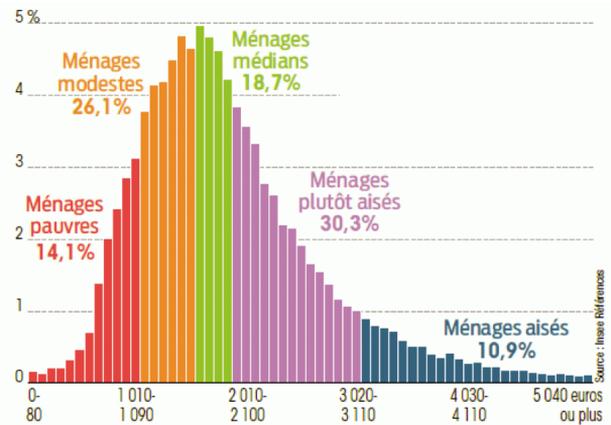


Fig. 19 : Mode < Médiane < Moyenne

3.3. Tables et réalité

Nous terminerons cette section par quelques remarques relatives aux tables de mortalité qui ont alimenté nos questions, constats et réflexions, par rapport aux graphiques de fréquence et de fréquences cumulées, par rapport aux valeurs centrales ainsi qu'à la dispersion au travers des boîtes à moustaches et des positions relatives des valeurs centrales. Pour établir les probabilités de décès puis de survie à chaque âge, Statbel (office belge de statistique) part des fréquences de décès sur un territoire donné (le pays, les régions...) et d'une population ciblée (femmes, hommes, deux sexes réunis). Considérant ensuite une population hypothétique d'un million de personnes, ils appliquent alors les probabilités successives de décès. La probabilité de décès à 0 an (il s'agit de ce qu'on appelle l'âge exact, ce sont donc les personnes qui meurent durant leur première année de vie qui sont considérées) est de 0,003927. Ce qui conduit à 3927 morts. Il y a donc 966073 survivants. La probabilité de décès à un an est de 0,000142. Il y a donc 141 morts et 995931 survivants. Et ainsi de suite, jusqu'à 105 où la probabilité de décès est de 1. Les derniers survivants meurent donc à cet âge. On constate donc qu'une table de mortalité n'est pas établie sur une population donnée dont on enregistrerait les décès d'année en année. Il faudrait attendre plus de cent ans pour connaître les résultats et la table ne serait plus d'actualité. La table repose sur une population hypothétique mais en utilisant des fréquences de décès bien réelles.

Dans le paragraphe d'introduction, vous aurez remarqué qu'on parle d'espérance de vie de 79 ans, alors que nous avons calculé 78,3 ans. La différence vient du fait qu'on considère soit les âges (0,1, 2...) ou les classes d'âge ([0,1[, [1,2[, [2,3[...). Dans ce dernier cas, on considère les milieux de classe pour le calcul de la moyenne. Ce qui donne une moyenne d'un demi supérieure à celle que nous avons calculé. On obtient alors une moyenne de 78,79. Ce n'est pas encore les 79 ans annoncés dans le texte d'introduction. La différence peut venir de ce qu'ils ont peut-être considérés

11. Clarisse Josselin, L'Insee dresse le portrait de la France du milieu, posté le 25 novembre 2017 sur le site de Force ouvrière, à l'adresse <https://www.force-ouvriere.fr/l-insee-dresse-le-portrait-de-la-france-du-milieu> et consulté le 10 août 2018.

les âges révolus⁽¹²⁾ et non les âges exacts. Mais, exprimée en âges révolus, l'espérance de vie est (forcément) légèrement plus courte et vaut 78,28...

4. Dernière sortie

Nous avons abordé trois sujets sociaux et/ou politiques qui illustrent la richesse du questionnement tant citoyen que mathématique, de pareilles situations. Pourquoi dès lors, parmi les onze pages relatives à l'approche graphique des fonctions dans un manuel de troisième année, ne trouve-t-on pas un seul exemple de fonction issu d'une réalité politique, sociale, citoyenne, quotidienne ? Pourquoi observe-t-on dans les classes du primaire et du secondaire, tant de calculs de valeurs centrales (surtout la moyenne) et si peu de lien avec les graphiques, si peu de comparaison entre situations (2 populations ou 2 échantillons) ? Pourquoi aux cours de statistique ou de traitement de données, rencontre-t-on tant de situations relatives aux loisirs des jeunes, aux tirs au but de joueurs de football ? Et presque jamais des questions de cotes, de revenus, de mortalité... Pourquoi en troisième année secondaire et après, y a-t-il tant d'études de domaines et de signes qui par ailleurs semblent moins fréquents dans les « vrais » problèmes ? Pourquoi reporte-t-on l'étude de la dispersion à la quatrième secondaire alors qu'elle est accessible, notamment au travers des graphiques, des boîtes à moustaches et de la comparaison des valeurs centrales, pour des plus jeunes et qu'elle joue un rôle important dans l'approche de la fluctuation d'échantillonnage.⁽¹³⁾

Ce ne sont pourtant pas les situations qui manquent... Les graphiques de croissance de taille, de poids, de périmètre crânien, d'indice de masse corporelle, et l'angoisse des parents dont les enfants ne sont pas dans les courbes. La démographie, ses cartes, ses pyramides droites ou renversées, ses disques de taille de population, et les inquiétudes ou la peur liées aux migrations, à l'immigration, au vieillissement de la population. L'économie des ménages avec l'indexation, les salaires, les revenus, les taux de pauvreté, les indices d'inégalité, et les interprétations politiques disparates.

Bien sûr, en classe, ce n'est pas toujours simple à gérer parce qu'il y a débat, prises de position, voire polémique. Mais c'est riche et incontournable pour développer les capacités critiques citoyennes des enfants et des adolescents qui sont dans nos classes.

Benoît Jadin est membre de GEM (Groupe d'Enseignement Mathématique), membre de CGé (Changements pour l'égalité) et enseignant à HELMO Liège. ✉ benjadin@gmail.com

Camille Kroonen et Juliette Petitjean sont étudiantes à HELMO Liège dans l'option « Devenir enseignant en mathématiques dans l'enseignement secondaire supérieur ».

12. L'âge révolu est l'âge au dernier anniversaire. Si vous êtes nés le 3 juillet 1940 et que vous êtes mort le 28 février 2018, votre âge exact de décès est de 78 ans tandis que votre âge révolu est de 77 ans.

13. Dans certains pays comme la France, les Pays-Bas ou l'Angleterre, on aborde cette question bien plus tôt dans le cursus.