

DANS NOS CLASSES

LE FACTEUR DE MENSONGE

Gilles Waehren

Le programme de Statistiques en Seconde (2009) indique, dans les capacités attendues, « Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées). » avec comme commentaire « [...]faire réfléchir les élèves sur des données réelles, riches et variées [...] proposer des représentations pertinentes. » - à noter que les représentations graphiques statistiques ont disparu des nouveaux programmes (2019). Après les avoir traitées de façon cavalière pendant plusieurs années, il m'a semblé utile de donner aux élèves des bases plus solides pour les construire et ce pour deux raisons. D'une part, quand les TPE, dans mon lycée, intégraient encore des mathématiques, la confusion entre histogramme et diagramme en barres était récurrente. D'autre part, un bon dessin valant mieux qu'un long discours, nos médias préférés regorgent d'infographies diverses et variées, mais pas toujours très mathématiques (voir la rubrique Maths et Médias du Petit Vert). C'est la page : [Comment mentir avec un graphique](#) du site flamand [Smals Research](#) qui m'a fourni le sujet du Devoir Maison que j'ai donné à mes deux classes de Seconde (2°10 et 2°11) sur ce thème.

Dans le sujet (voir en annexe), je fournissais un premier lien vers un article du site [Et que faire ?](#), qui propose des pages assez généralistes et hétéroclites, consacré à la lecture des graphiques statistiques. On trouvera la source de cet article dans un e-book : [Statistiques appliquées à la psychologie](#). Cette lecture avait pour but de leur montrer les diverses manières de mal construire un diagramme en barres ou un diagramme circulaire. Il s'agissait ensuite de se donner un outil pour mesurer le niveau de distorsion d'un graphique : le « lie factor ».

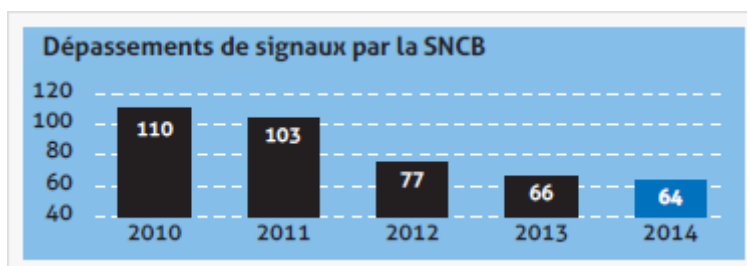
Le « lie factor » est une invention d'Edward Tufte, statisticien et concepteur d'infographies, défini par le rapport : $\text{lie factor} = \frac{\text{taille de l'effet dans le graphique}}{\text{taille de l'effet dans les données}}$.

La taille de l'effet est, en quelque sorte, un taux d'évolution :

$$\text{taille de l'effet} = \frac{|\text{seconde valeur} - \text{première valeur}|}{\text{première valeur}}.$$

La valeur absolue n'étant pas au programme 2009, il a fallu indiquer deux variantes de cette formule. Le critère de validité d'un graphique décidé par Tufte est que l'infographie est correcte si le facteur de mensonge est compris en 0,95 et 1,05.

Dans un premier exercice, le calcul du « lie factor » était exposé puis mis en pratique sur un graphique fourni par Smals Research :



Ce cas, relativement classique, de mise à l'ordonnée incorrecte est souvent utilisé, de manière intentionnelle, pour accentuer les écarts entre les valeurs. Le calcul du lie factor a souvent été bien réussi, comme chez cette élève (très bonne, cela dit, dans une classe de 2°10 plutôt faible) :

taille de l'effet dans le graphique = 0,66
 ↳ On passe sur le graphique de $110 - 40 = 70$ à $64 - 40 = 24$
 donc la taille de l'effet dans le graphique est $\frac{70 - 24}{70} \approx 0,66$.

ici, la 1^{ère} valeur (70) est supérieure à la 2^{ème} (24), donc on utilise la formule suivante :

$$\frac{\text{première valeur} - \text{seconde valeur}}{\text{première valeur}}$$

taille de l'effet dans les données $\approx 0,42$
 ↳ ici, première valeur (110) > 2^{ème} valeur (64) donc on utilise la formule :

$$\frac{\text{première valeur} - 2^{\text{ème}} \text{ valeur}}{\text{première valeur}} = \frac{110 - 64}{110} \approx 0,42$$

le factor = $\frac{\text{taille de l'effet dans le graphique}}{\text{taille de l'effet dans les données}} = \frac{0,66}{0,42} \approx 1,57$

↳ 1,57 n'est pas compris entre 0,95 et 1,05, donc il y a une distorsion de la réalité et l'intégrité de la visualisation n'est pas assurée

Toutefois, la mise en œuvre d'une formule, voire d'un petit algorithme (calculer les tailles des effets puis évaluer le « lie factor ») ne s'est pas toujours faite sans mal. S'approprier de nouveaux concepts n'est pas chose aisée pour tous les élèves.

Ici, l'élève (2°10) a interprété la question concernant la période de 2010 à 2014 en ajoutant les tailles de tous les effets entre ces deux années :

$\frac{110 - 103}{110} \approx 0,06$	$\frac{70 - 63}{70} \approx 0,1$	
$\frac{103 - 77}{103} \approx 0,25$	$\frac{63 - 37}{63} \approx 0,70$	$110 - 40 = 70$
$\frac{77 - 66}{77} \approx 0,15$	$\frac{37 - 26}{37} \approx 0,2$	$103 - 40 = 63$
$\frac{66 - 64}{66} \approx 0,03$	$\frac{26 - 24}{26} \approx 0,07$	$77 - 40 = 37$
		$66 - 40 = 26$
		$64 - 40 = 24$
$0,06 + 0,25 + 0,15 + 0,03 = 0,49$		
$0,1 + 0,70 + 0,2 + 0,07 = 1,07$		
$\frac{1,07}{0,49} \approx 2,2$		

Dans le cas suivant, sa camarade (2°10) utilise également toutes les valeurs pour la période.

$$\frac{110 - 103 - 77 - 66 - 64}{110} = -\frac{20}{11} \approx -1,8$$

On prend l'effet de taille $\rightarrow$ dans les données première valeur > seconde valeur.
On obtient le calcul ci dessus. On peut donc en déduire

exemple

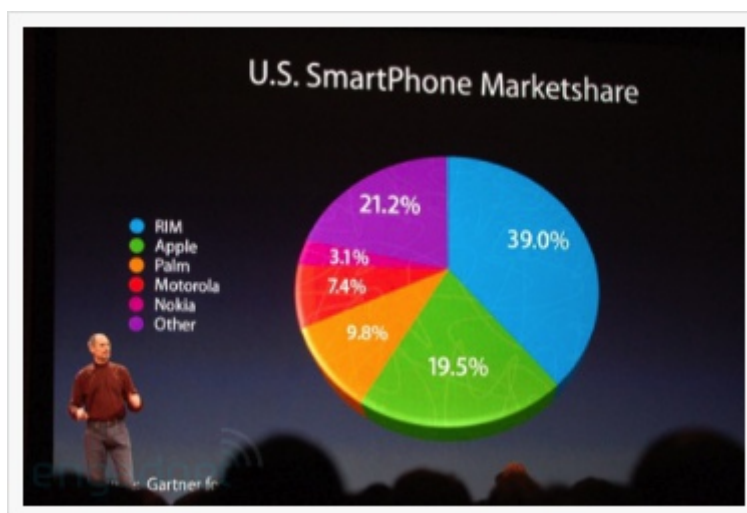
$$\frac{70 - 63 - 37 - 26 - 24}{70} = -\frac{8}{7} \approx -1,1$$

donc le $\Leftarrow$ lie factor $\Rightarrow$ est $\frac{-1,1}{-1,8} = \frac{11}{18} \approx 0,6$

Je pense que la formulation de la question pouvait prêter à confusion, mais l'exemple de mise en pratique ne devait pas laisser de place à ce genre d'interprétation ; d'autant plus que le sujet du devoir avait été éclairci en classe.

Je profitais de cet exercice pour insister sur deux aspects d'un graphique qui sont souvent négligés lors de la lecture et qui ne sont pourtant pas toujours évidents à énoncer : la population et le caractère considérés.

Le deuxième exercice s'appuyait sur un autre exemple fourni par la même page de Smals Research



Ce diagramme circulaire pouvait être analysé en mesurant les angles, selon le principe de construction défini en cours, mais aussi, en calculant les aires des secteurs. Dans les deux cas, le travail se faisait en traçant sur le polycopié du sujet. La tâche n'étant pas très aisée, j'avais déposé sur l'ENT une version couleur et A3 du document. Cette année, j'ai reconstruit sur tableur le diagramme, mais je n'ai pas su retrouver l'angle d'inclinaison du cylindre par rapport au plan horizontal. Les élèves devaient comparer les secteurs d'Apple (19,5%) et d'"Other" (21,2%).

Le travail sur les angles a donné ce genre de production (2°11, assez forte)

1) Pour le « lie factor » entre « 21,2% » et « 19,5 » :

- on a, sur le graphique, de 78° à 79° donc la taille de l'effet sur le graphique est : $\frac{79-78}{79} \approx 0,013$.
- on a, avec les données, de 19,5% à 21,2%, donc la taille de l'effet dans les données est : $\frac{21,2-19,5}{21,2} \approx 0,08$.

donc le « lie factor » est : $\frac{0,013}{0,08} \approx 0,1625$. Le graphique est distordu car il n'est pas compris $[0,99 ; 1,05]$, donc marque utilise cette technique pour faire croire qu'elle a un plus grande importance donc d'être 2^e alors qu'elle 3^e.

Cette élève (2°10) prétend mesurer les angles, mais elle fournit en fait les angles que l'on calcule pour construire le diagramme. Elle ne devrait pas trouver de distorsion, et pourtant.

1) Pour construire un diagramme circulaire il suffit de faire correspondre à chaque effectif le pourcentage ou fréquence un angle de mesure proportionnelle.

2) Je mesure les angles de :

$$21,2 = 360 \times 21,2 \div 100 = 76,32^\circ$$

$$19,5 = 360 \times 19,5 \div 100 = 70,20^\circ$$

donc : $\frac{76,32 - 70,20}{76,32} \approx 75,40$

La taille de l'effet sur le graphique est égale à environ 75,40.

3) donné : $\frac{19,5 - 21,2}{19,5} \approx 18,41$ graphique : $\frac{76,32 - 70,20}{76,32} \approx 75,40$

les valeurs ne correspondent pas

donc le lie factor est : $\frac{75,40}{18,41} \approx 4$

Les valeurs (première et seconde) n'ont pas été clairement identifiées, donc le calcul donne une distorsion importante. Peut-être que, comme Tufte utilise les mêmes mots pour des nombres différents, il serait bon de choisir d'autres termes plus évocateurs. Quoiqu'il en soit, la mesure au rapporteur ne s'est pas imposée comme une évidence. Le premier trimestre, en Seconde, est encore une période où les initiatives ne vont pas de soi. Par ailleurs, la demande d'explications sur les DM n'était pas toujours la même dans les deux classes : les élèves les plus forts sont aussi souvent ceux qui ont le plus de questions...

Les données étant relativement proches (8 % d'écart relatif de la plus petite à la plus grande), le calcul du « lie factor » montrait une distorsion relativement importante. Les mesures au rapporteur ne pouvant prétendre à une grande précision, on pouvait se demander si cette distorsion n'était pas due à cette imprécision. C'est pour cela que j'avais suggéré à certains élèves de recourir à un calcul d'aire, vu que les deux secteurs n'ont pas le même rayon.

Pour mesurer la taille de l'effet sur le graphique

effet données : $\frac{21,2 - 19,5}{19,5} = 0,087$

effet graphique : $\frac{2,96 - 0,80}{2,96} = 0,72$ $19,5\% \rightarrow 1,95$
 $21,2 \rightarrow 2,80$

$77 \times \pi \times 7,1^2 \div 360 = 2,96$
 $81 \times \pi \times 1,15^2 \div 360 = 0,80$

le lie factor : $\frac{0,72}{0,087} = 8,27$ le lie factor du graphique est 8,27

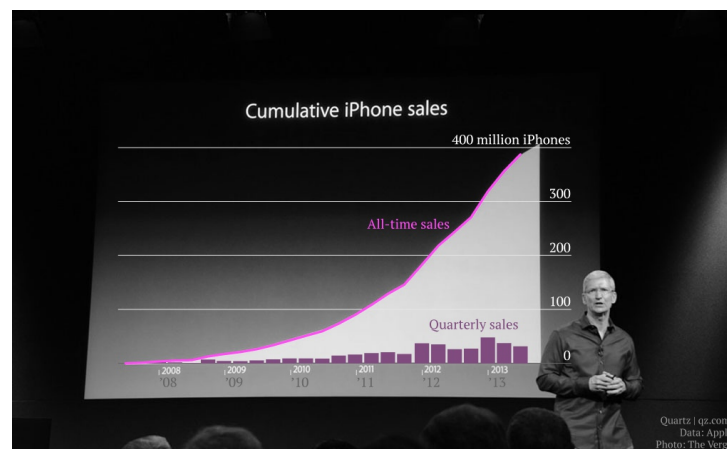
(Le travail a ici été rafistolé après l'indication sur les aires)

L'aide du professeur n'a donc pas été la même pour tous, mais comme j'évalue les DM par compétences, cela n'a pas impacté la note finale. Il faudrait peut-être demander dans l'énoncé de calculer le « lie factor » sur les angles et sur les aires et comparer les résultats. C'est le genre d'attitude que l'on peut légitimement attendre d'un lycéen : porter un regard critique sur son travail pour éviter des conclusions hâtives. Une tentation très forte chez nos élèves (voire le troisième exercice) est de surinterpréter. C'est aussi le reproche que l'on peut parfois faire aux médias dans leur utilisation des résultats statistiques. Soyons prudents avec les données numériques !

Le troisième exercice fournissait un diagramme moins courant dans les manuels de Seconde. Et pour cause, la firme à la pomme avait produit une courbe donnant en fonction de l'année (entre 2008 et 2012), le total des ventes cumulées depuis 2008 : on obtient nécessairement la courbe d'une fonction croissante sur toute la période.



Ce dernier exemple était là pour insister sur la pertinence des représentations. J'avais donc complété cette figure par le deuxième graphique qui l'accompagne sur Smals Research (ci-dessous).



Il était demandé aux élèves d'estimer les augmentations entre 2011 et 2012 puis entre 2012 et 2013 et de constater que cette courbe, qui semble montrer une forte progression des ventes cache une réalité plus terne, comme l'indiquent les petites barres. La question de la plus forte augmentation était quelque peu insidieuse : un certain nombre d'élèves s'en est tenu à la variation absolue. Mais la variation relative sur les données cumulées n'est pas forcément pertinente non plus.

2) Contrairement à la fausse impression que pourrait donner la courbe représentant les ventes d'iphones cumulées (c'est-à-dire les nouvelles ventes ajoutées aux précédentes), le diagramme en bâtons révèle que l'augmentation la plus forte n'a pas été celle de 2012 à 2013 mais celle de 2011 à 2012. En effet, il semble que les ventes aient doublé de 2011 à 2012, tandis qu'elles ont légèrement augmenté de 2012 à 2013.

Ca t'emb?

2) De 2011 à 2012, le nombre d'iphones vendus passe de 100 millions à 200 millions.

Pourcentage d'augmentation = $\frac{(\text{valeur d'arrivée} - \text{valeur de départ})}{\text{valeur de départ}} \times 100$

$$= \frac{(200\,000\,000 - 100\,000\,000)}{100\,000\,000} \times 100$$

$$= 100\% \text{ d'augmentation.}$$

De 2012 à 2013, le nombre d'iphones vendus passe de 200 millions à 320 millions.

Pourcentage d'augmentation = $\frac{(\text{valeur d'arrivée} - \text{valeur de départ})}{\text{valeur de départ}} \times 100$

$$= \frac{(320\,000\,000 - 200\,000\,000)}{200\,000\,000} \times 100$$

$$= 60\% \text{ d'augmentation.}$$

Le pourcentage d'augmentation est légèrement supérieur de 2011 à 2012 que de 2012 à 2013 (100% > 60%) / celui.

En tout cas, pour donner une conclusion appropriée, les langues se sont déliées :

Sur le 1er graphique, Apple souhaite dupliquer le public en indiquant une forte hausse de ventes. En réalité, elle représente toutes les ventes d'iphones depuis la création de ce produit, une courbe qui ne peut qu'être croissante. Le graphique rajoute l'évolution réelle des ventes par période, une progression qui stagne en indiquant de réelles ventes peu impressionnantes.

Finalement, j'ai trouvé ce troisième exercice trop compliqué pour des élèves de Seconde et j'ai choisi de le supprimer cette année. Il aurait pu trouver sa place dans le chapitre de ES ou de STMG consacré aux taux d'évolution.

La lecture de graphiques correctement réalisés n'est jamais chose aisée : les informations de premier niveau (lecture directe) sont souvent bien perçues, celles de second niveau (interprétation des données, mode de construction) donnent parfois lieu à des contresens. Si, en plus, on fournit aux citoyens des graphiques mal construits, la lecture est une vraie gageure. Mon objectif était que les élèves prennent conscience qu'une information aux apparences mathématiques n'est pas forcément plus sérieuse qu'une autre et que l'école fournit des outils pour la décrypter. Ce type de travail dans lequel les élèves ont plus de latitude (beaucoup de contenu informationnel, quelques questions pas trop directives) leur donne souvent des ailes et devient, pour certains élèves, l'occasion de s'exprimer de façon plus personnelle, au travers d'opinions parfois tranchées ou d'une recherche mathématique plus ou moins aboutie.

Annexe : Le sujet du DM

Infos

Info 1 : Consulter la page « Apprendre à lire les graphiques en statistique et les erreurs ou méthodes de manipulation » (à saisir dans un moteur de recherche)

(<http://www.etquefaire.fr/Articles/apprendre-a-lire-les-graphiques-en-statistique-et-les-erreurs-ou-methodes-de-manipulation.php>)

Info 2 : le « lie factor »

Pour pouvoir mesurer à quel point un graphique est une distorsion de la réalité, Edward Tufte, auteur majeur dans le domaine de la visualisation de l'information, a défini le concept de "lie factor" (que l'on pourrait traduire par facteur de mensonge) de la façon suivante :

$$\text{lie factor} = \frac{\text{taille de l'effet dans le graphique}}{\text{taille de l'effet dans les données}}$$

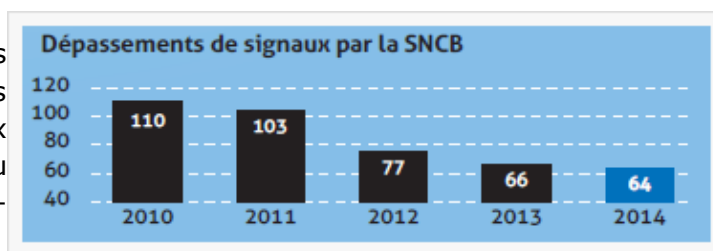
avec $\text{taille de l'effet} = \frac{\text{seconde valeur} - \text{première valeur}}{\text{première valeur}}$ si $\text{seconde valeur} > \text{première valeur}$

et $\text{taille de l'effet} = \frac{\text{première valeur} - \text{seconde valeur}}{\text{première valeur}}$ si $\text{première valeur} > \text{seconde valeur}$

Un "lie factor" de 1 indique donc qu'il n'y a pas de distorsion. Tufte estime que ce facteur doit rester entre 0,95 et 1,05 pour assurer l'intégrité de la visualisation.

Exercice 1

La SNCB (Société Nationale des Chemins de fer Belges) a mesuré sur cinq années le nombre de dépassements de signaux (pour simplifier : le train passe au feu rouge) et a produit le graphique ci-contre.



Exemple de calcul du « lie factor » pour ce graphique :

Pour l'évolution entre 2011 et 2012 :

- on passe, avec les données, de 103 à 77 donc la taille de l'effet dans les données est :

$$\frac{103 - 77}{103} \approx 0,25$$

- on passe, sur le graphique, de $103 - 40 = 63$ à $77 - 40 = 37$ donc la taille de l'effet dans le

graphique est : $\frac{63 - 37}{63} \approx 0,70$

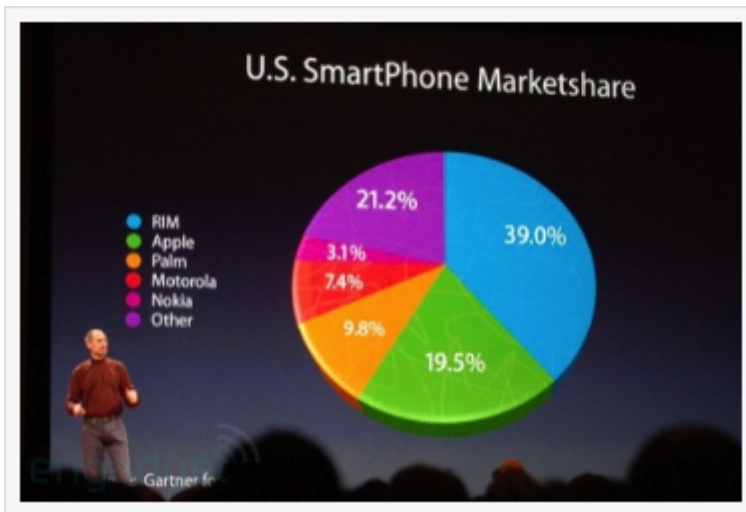
donc le « lie factor » est : $\frac{0,7}{0,25} = 2,8$

Questions :

1. Quels sont la population et le caractère considérés ?
2. Calculer le « lie factor » pour la période de 2010 à 2014. Commenter.

Exercice 2

Lors d'une de ses célèbres « keynotes », la marque « Apple » a diffusé ce document pour présenter la répartition du marché des smartphone aux Etats-Unis.



1. Rappeler le principe de construction d'un tel diagramme.

2. En mesurant sur la figure, comparer l'importance des secteurs « 21,2 % » et « 19,5 % » en calculant la « taille de l'effet sur le graphique ».

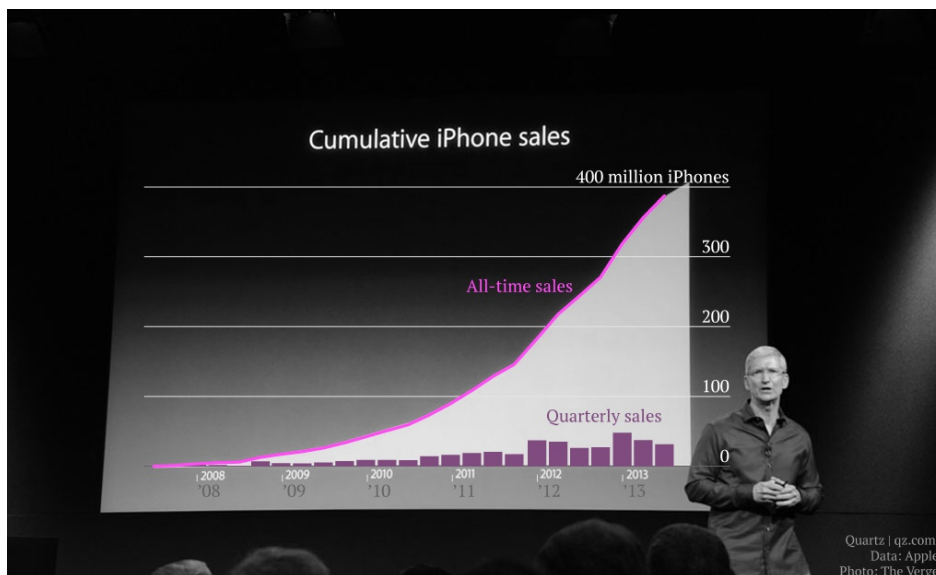
3. Calculer le « lie factor ».

Exercice 3

La même marque a voulu impressionner ses actionnaires avec des résultats de ventes en affichant le graphique ci-contre.



Un décodeur d'information américain a repris et complété cette image pour améliorer sa lisibilité :



1. Quel élément utile manquait sur le premier graphique ?

2. Quelle est l'augmentation la plus forte : celle de 2011 à 2012 ou celle de 2012 à 2013 ? Commenter.