

L'ART DU DESSIN

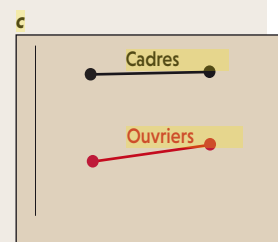
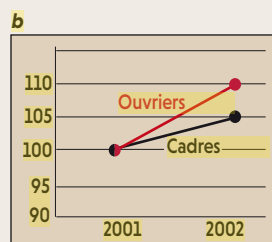
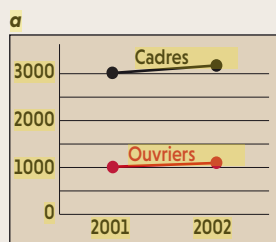
Il y a mille façons de projeter graphiquement des données, chacune permettant de mettre en avant un fait, d'en masquer un autre, ou de donner à croire autre chose que ce que les données indiquent. L'exemple suivant a d'abord été proposé dans la revue *The Economist*. Entre 2001 et 2002, les salaires des cadres et des ouvriers ont évolué ainsi :

	SALAIRES MENSUELS MOYENS	
	2001	2002
OUVRIERS	1000 €	1100 €
CADRES	3000 €	3150 €

En moyenne, les cadres et les ouvriers gagnent plus en 2002 qu'en 2001. En valeurs brutes, l'augmentation des cadres (150 euros) est supérieure à celle des ouvriers (100 euros). Cependant, la situation est inversée en valeur relative :

5% pour les cadres contre 10 % pour les ouvriers. Ces points de vue différents ont leurs équivalents graphiques, respectivement en *a* et en *b*. Le commentaire naturel de la représentation *a* est : « Les ouvriers et les cadres ont été augmentés. Les cadres recevaient plus, et reçoivent encore plus que les ouvriers. L'écart entre les salaires empire. » La représentation *b* s'attache aux augmentations relatives. Elle montre que les cadres ont été moins augmentés en pourcentage que les ouvriers, mais elle ne renseigne pas sur les éventuelles modifications de rémunération en valeur absolue. Cette courbe donne

l'impression que les cadres sont moins gâtés que les ouvriers... Peut-on représenter les évolutions relatives et les différences entre les professions ? Oui, en ayant recours à une échelle logarithmique en ordonnée (*c*). Avec cette graduation, une même différence de hauteur correspond à un même coefficient multiplicateur. La lecture de ce graphique suggère cette fois : « Les cadres recevaient et reçoivent encore plus que les ouvriers, mais les inégalités se réduisent. » Ces trois représentations étant toutes parfaitement honnêtes, on peut commenter l'évolution des salaires selon sa préférence.



Le revenu médian en Coccagne passe alors illico à 1400 sols, et le demi-revenu médian à 700 sols, et par un miracle statistique à couper le souffle, l'indice de pauvreté passe aussitôt, en mai 2068, à 0, le minimum possible. »

Créé suite à une polémique sur l'indice de pauvreté, le BIP 40 ou baromètre des inégalités et de la pauvreté est un indicateur synthétique des inégalités et de la pauvreté. Cet indice a été proposé par le Réseau d'alerte sur les inégalités en 2002 et il élimine certains des problèmes mentionnés ci-dessus. Toutefois, il est assez complexe, si bien qu'il est délicat d'en comprendre le sens et qu'on ne peut garantir qu'il évitera toutes les absurdités des indices plus simples.

UNE AFFAIRE DE FAMILLE

Parmi tous les pièges que détaille le livre de Nicolas Gauvrit, l'un d'eux est assez subtil et mérite une attention particulière, nous le nommerons le paradoxe du nombre moyen d'enfants. Une enquête exhaustive menée dans une ville lointaine indique que les familles ayant des enfants de moins de 18 ans se répartissent de la manière suivante : 10% de familles à 1 enfant, 50% à 2 enfants, 30% à 3 enfants, 10% à 4 enfants. Le nombre moyen d'enfants par famille (parmi celles qui ont des enfants) est donc de $(10 + 100 + 90 + 40)/100 = 2,4$.

Pour contrôler cette statistique, les autorités administratives procèdent à un sondage. On interroge 1000 enfants de moins de 18 ans soigneusement pris au hasard et on leur demande combien il y a d'enfants dans leur famille, eux compris. En faisant la moyenne des

réponses, on obtient... 2,68 ! Cela semble absurde. On recommence donc le sondage en interrogeant cette fois 10000 enfants, on trouve maintenant 2,67. Un troisième sondage sur 100000 enfants donne 2,668 à nouveau. Pourquoi cet écart si important avec les 2,4 de la statistique qui prenait en compte toutes les familles ayant des enfants ?

La réponse tient dans le fait qu'en interrogeant des enfants au hasard, vous interrogerez 4 fois plus d'enfants des familles à 4 enfants que vous n'en interrogerez dans les familles à 1 enfant, ce qui fausse la moyenne. S'il y a 1000 familles, il y aura 100 enfants uniques, 1000 enfants appartenant à une famille de 2 enfants, 900 enfants appartenant à une famille de 3 enfants, 400 enfants appartenant à une famille de 4 enfants. Au total, les réponses données par ces 2400 enfants conduiront au résultat de 2,666... enfants par famille.

Les sondages opérés n'évaluent pas le nombre moyen d'enfants d'une famille prise au hasard, mais le nombre moyen d'enfants qu'on trouve dans la famille d'un enfant pris au hasard. « Prendre une famille au hasard » et « Prendre un enfant au hasard » n'est pas la même chose.

Nous le savons depuis Condorcet pour les votes, il est bien difficile de synthétiser un ensemble de nombres en un seul. Nous espérons que les pièges de la statistique, des représentations graphiques, des indices synthétiques, des sondages présentés ici – et ceux que vous trouverez dans le livre de Nicolas Gauvrit – vous aideront à mieux comprendre les vérités cachées derrière l'inquiet et fluctuant monde des nombres. ■

BIBLIOGRAPHIE

N. GAUVRIT, *Statistiques. Méfiez-vous*, Éditions Ellipses, 2014.

I. EKELAND, *Statistiques incroyables, Pour la Science* n° 334, p. 6, août 2005.

P. BICKEL ET AL., *Sex Bias in Graduate Admissions: Data from Berkeley*, *Science*, vol. 187, n° 4175, pp. 398-404, 1975.

G. YULE, *Notes on the Theory of Association of Attributes in Statistics*, *Biometrika*, vol. 2, n° 2, pp. 121-134, 1903.