# Statistiques et fantaisies journalistiques

Frédéric Schütz

schutz@mathgen.ch


PME Magazine, octobre 2008. Un exemple-type dans lequel le design (effet 3D et reflets divers) se fait au détriment de l'information.

Made with Scribus

Article publié dans le *Bulletin de la Société Statistique Suisse* (numéro 70), octobre 2011.
(version corrigée du 17 octobre 2011)

# Statistiques et fantaisies journalistiques

Frédéric Schütz, SIB Institut suisse de bioinformatique, Lausanne,
schutz@mathgen.ch

**Résumé**

Pourquoi les statisticiens grimacent-ils souvent devant les chiffres et graphiques publiés dans les médias ? Probablement parce qu'ils vivent dans un autre monde que celui des journalistes et les infographistes: leur formation est différente, ils utilisent d'autres outils, et leurs priorités ne sont pas les mêmes. Il n'est dès lors pas surprenant qu'ils travaillent autrement et n'obtiennent pas les mêmes résultats. En se basant sur une gallerie d'exemples collectés au fil du temps, nous allons explorer et expliquer ces différences, en essayant d'aller au-delà d'une simple critique du contenu de ces graphiques et autres données statistiques.

## 1. Introduction

Les statistiques portées à la connaissance du grand public le sont en grande majorité par les médias et les journalistes plutôt que par les professionnels du domaine. Cette situation rappelle le livre «*How to lie with statistics*», vendu à plus d'exemplaires que tout autre livre de statistiques, alors que son auteur n'est pas un statisticien. Mais si le livre de Darrell Huff est généralement apprécié des gens du domaine, ces derniers sont souvent très critiques sur le traitement des chiffres par les médias. La plupart du temps, à juste titre: preuve en sont les exemples de bourdes statistiques découvertes dans les journaux que de nombreux statisticiens s'amusent à collectionner.

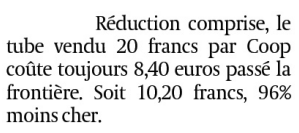
Le but de cet article, en partant d'une telle collection, est d'aller un peu plus loin, en essayant de décrire d'où proviennent ces erreurs, et en particulier, comment la façon de travailler du journaliste ou de l'infographiste diffère de celle du statisticien. Sans surprise, l'accent sera porté sur les graphiques: outils favoris de toutes ces professions, ils sont abondants dans les médias, et présentent mille opportunités pour déformer l'information. Du côté statistique, ils forment l'un des sujets les plus connus du domaine, et sont associés à des noms célèbres tels que Edward Tufte, William Cleveland ou Howard Wainer.

Nous parlerons aussi de quelques curiosités qui ne sont pas dues aux journalistes, mais qui sont néanmoins intéressantes ou instructives.

## 2. Quand les chiffres sont simplement faux

Une catégorie de problèmes concerne les simples erreurs sur les chiffres, la plupart du temps involontaires. Certains chiffres erronés restent plausibles, et nécessitent de recourir à des sources pour être découverts et corrigés (par exemple, en vérifiant avec l'Office fédéral de la statistique, qui répond efficacement à ce genre de requêtes). D'autres erreurs sont plus simples, ainsi que le montrent les exemples récents suivants:

- Un tube de crème vendu 20 francs en Suisse et 10.20 francs (8.40 euros) en France est décrit par *Le Temps* du 9 juillet 2011, une fois comme étant «*96% plus cher en Suisse*», et une autre fois, de façon symétrique, mais incorrecte, «*96% moins cher en France*».


Réduction comprise, le tube vendu 20 francs par Coop coûte toujours 8,40 euros passé la frontière. Soit 10,20 francs, 96% moins cher.

- La probabilité pour un couple de se séparer, sachant que certains de leurs proches se sont séparés, serait de 147% selon une brève parue sur le site internet de la *Télévision Suisse Romande*. Dans la source (un article sur le site du magazine *L'Hebdo*), cette probabilité était *augmentée* de 147% par rapport aux autres couples.
- Un calcul réalisé par *L'Hebdo* dans le cadre de la votation sur le 2ème pilier en mars 2010 expliquait qu'il manque 600 millions de francs d'argent chaque année pour payer les rentes, parce que 300'000 nouveaux rentiers par année coûtent chacun 20'000 CHF de trop. Le journaliste notait en particulier ce chiffre élevé de 300'000 nouvelles rentes. Il y avait de quoi s'étonner en effet: le produit des deux chiffres est de 6 milliards de francs, et le chiffre de 300'000 est incorrect.

Le but de l'exercice n'est pas de jeter la pierre aux responsables de ces erreurs, dont les compétences ne sont pas forcément en cause: des coquilles se cachent dans les textes beaucoup plus souvent qu'on ne le voudrait (comme tout le monde l'a expérimenté personnellement, y compris l'auteur de ces lignes). Il y a cependant une différence entre les coquilles et ces erreurs statistiques: les rédactions emploient généralement des correcteurs pour l'orthographe et la grammaire, alors que personne ne vérifie les chiffres.

Les trois problèmes mentionnés ci-dessus peuvent être découverts sans vérifier les sources originales, uniquement avec du bon sens ou des calculs simples, mais une telle vérification ne fait pas partie des habitudes.
Si ces erreurs sont généralement involontaires chez les journalistes, la situation est différente chez certains politiciens dont les statistiques peu nettes finissent quelques fois dans les pages de la presse. Certains lecteurs se souviendront de l'annonce qu'un comité de droite avait fait paraître en 2004 et qui faisait état d'une augmentation exponentielle (et surréaliste) du nombre de musulmans en Suisse. Depuis, certains partis diffusent régulièrement des statistiques douteuses. Le tout-ménage que l'UDC a distribué dans tous le pays en été 2010, et dont certaines parties ont fait l'objet de publicité dans les médias, est à cet égard exemplaire, et les commentaires statistiques à son sujet pourraient faire l'objet d'un article entier.

**3. Les différences entre statisticiens et journalistes**

Quelle est la différence principale entre un statisticien et un journaliste ou un infographiste ? La réponse est contenue dans la question: ils font des métiers différents, et ont des formations différentes. Le statisticien n'est généralement pas réputé pour ses talents de graphiste, et à l'inverse, l'infographiste a la plupart du temps une formation dans le domaine du graphisme, mais aucune dans la représentation des chiffres. Cette approche différente des sujets statistiques explique en grande partie les problèmes relevés par les statisticiens dans les médias. Mais d'autres raisons entrent également en ligne de compte, en particulier des priorités différentes, ainsi que l'accès à d'autres outils.

**La formation**

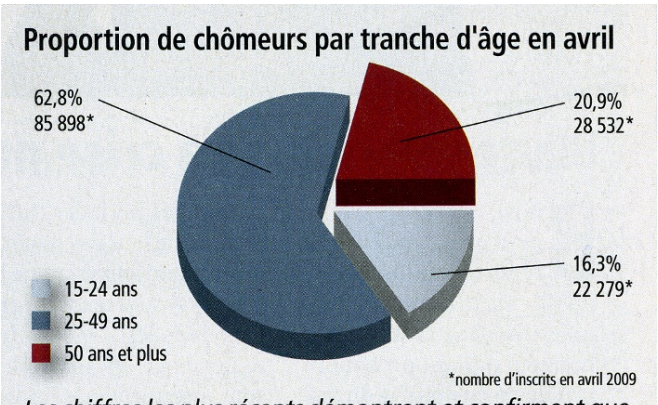
Le problème de la connaissance statistique n'est pas spécifique aux métiers des médias: il est lié au manque général d'habitude du grand public de travailler avec les chiffres, une incommodité parfois nommée «*innumérisme*». Mais il est particulièrement handicapant quand il touche les journalistes, souvent amenés à traiter des données chiffrées. Un exemple-type est la notion de médiane: souvent utilisée (à raison) par l'Office fédéral de la statistique en lieu et place de la moyenne, son interprétation est plus simple que celle de cette dernière. Et pourtant, elle reste quasiment inconnue en dehors des milieux scientifiques: d'après un sondage informel réalisé lors de cours au Centre Romand de Formation des Journalistes (CRFJ), seuls 10% des journalistes stagiaires connaissent l'existence ou la définition de cette mesure.

C'est ainsi que *L'Hebdo* a publié il y a quelques temps une brève qui commençait par l'affirmation suivante:

> «*5845 francs: tel est le salaire médian des Suisses (à ne pas confondre avec le salaire moyen, il indique le salaire perçu par le plus grand nombre d'employés d'un secteur donné).*»

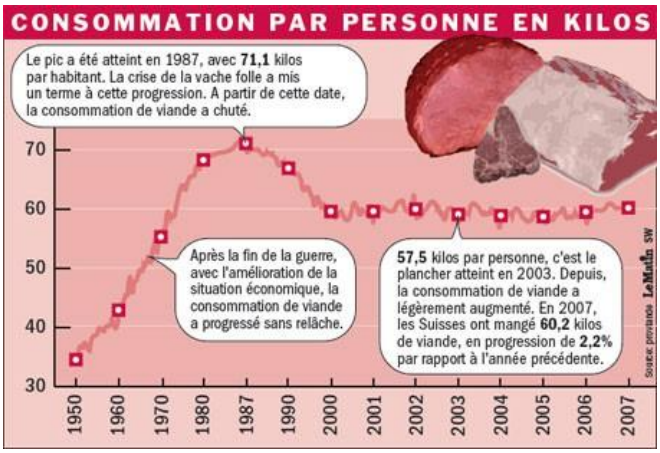
Si le lecteur peut probablement deviner, sans savoir ce qu'est une médiane, que ces 5845 francs représentent une certaine mesure centrale des revenus des suisses, la définition donnée entre parenthèse, qui confond médiane et mode, est certaine de le laisser confus.

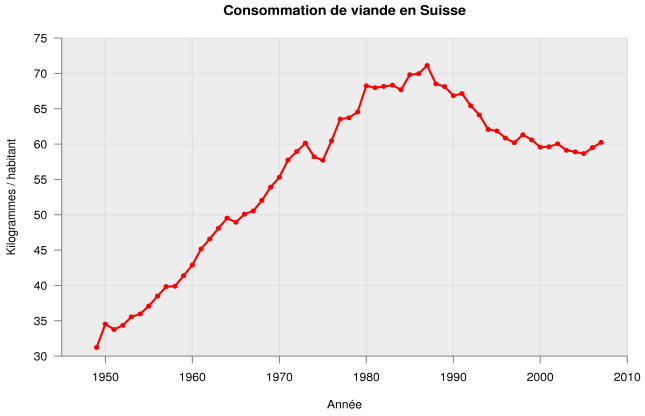
Sur un sujet différent, le mensuel *PME Magazine* a publié en juin 2009 le diagramme suivant, indiquant la répartition par classe d'âge des chômeurs en Suisse en avril 2009. Le commentaire interprète ces données en concluant que «*la classe d'âge la plus touchée reste celle des 25-49 ans*». Il oublie ainsi de répondre à la question fondamentale en statistique: «*comparé à quoi ?*» en tenant compte des tailles différentes des



*Les chiffres les plus récents démontrent et confirment que la classe d'âge la plus touchée reste celle des 25-49 ans.*

classes d'âge. L'utilisation de pourcentage par classe aurait montré que ce sont en fait les jeunes qui ont le *taux* de chômage le plus élevé (mais pas le nombre absolu le plus élevé).



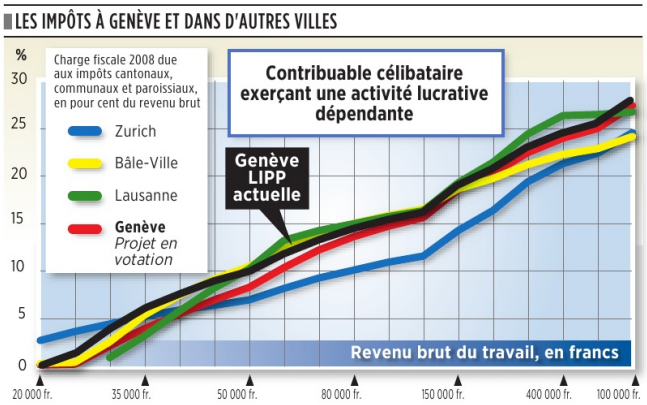

**Les priorités**

Là où le statisticien cherche généralement à montrer les données le plus simplement et le plus efficacement possible (on se souvient en particulier du «*data-ink ratio*» d'Edward Tufte, qui mesure la quantité d'encre effectivement utilisée pour montrer les données), le journaliste et l'infographiste doivent arbitrer entre ce même but, la charte graphique de leur média, l'angle qu'ils ont choisi pour leur sujet et d'autres contraintes.

Un exemple parmi les plus marquants est le graphique (ci-dessus, à gauche) publié dans le quotidien *Le Matin* du 15 août 2008 et qui montre la consommation de viande en Suisse entre 1950 et 2007. L'oeil entraîné repère immédiatement que l'axe horizontal n'est pas linéaire: les années 1950 à 2000 (50 ans) occupent le même espace que les années 2000 à 2007 (8 ans). Le résultat est différent sur un graphique à la bonne échelle (en haut à droite); en particulier, l'apparence de stabilité de la consommation ces dernières années, notée par le journaliste, n'est plus aussi évidente. Cette non-linéarité n'était pas une erreur du journaliste, mais une décision consciente, pour des «*raisons de place*» peu convaincantes...

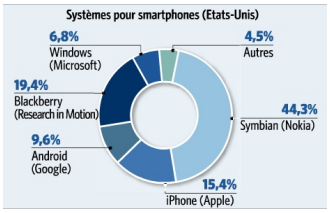
Mais une autre différence est observable: le graphique du *Matin* montre des variations entre les années, alors que l'autre version n'en contient pas. Il s'agit à nouveau d'une décision consciente du journaliste et de l'infographiste: les variations entre les points ont été ajoutées, aléatoirement, uniquement pour que le graphique ne soit pas trop «*lisse*» – une priorité qui est clairement opposée à celles du statisticien!

Si cet exemple porte sur un sujet relativement léger, les mêmes erreurs apparaissent aussi dans d'autres contextes. Le graphique ci-dessous, publié dans la *Tribune de Genève* du 9 septembre 2009, prend également des libertés avec la linéarité de l'axe horizontal, sur un sujet un peu plus important. Difficile de se rendre compte de l'évolution de la charge fiscale par rapport au revenu sur un graphique qui donne une impression aussi faussée ! (on notera également la faute de frappe dans la dernière graduation de l'axe horizontal).

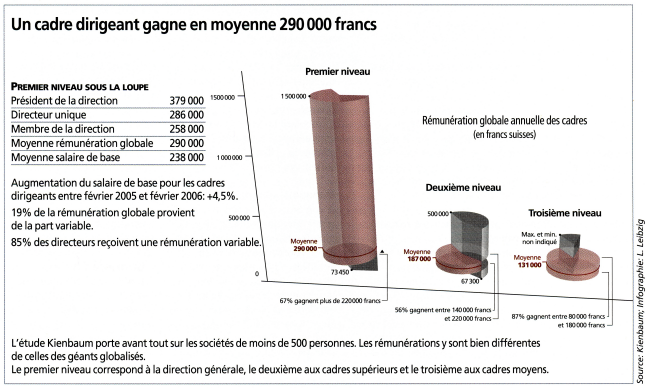



Les différences de priorités sont également apparentes dans le choix des types de graphiques, qui se fait trop souvent pour des raisons de design plutot que d'efficacité. Les diagrammes en camembert sont ainsi très utilisés par les infographistes, comme on le voit dans ces pages, alors qu'ils sont loin d'être les préférés des statisticiens. De plus, ils sont souvent accompagnés d'effets en trois dimensions, qui compliquent la lecture sans apporter de réel avantage – allant au contraire jusqu'à déformer complétement les données.

Depuis quelques années, ces camemberts sont généralement representés sous forme de «doughnuts» (avec l'intérieur évidé, comme dans l'exemple ci-contre publié dans *Le Temps*). Si cette modification permet d'alléger un peu le graphisme, cela se fait au prix d'une importante perte de lisibilité: l'information quantitative principale (l'angle entre les différentes tranches) n'est plus apparente à l'oeil.

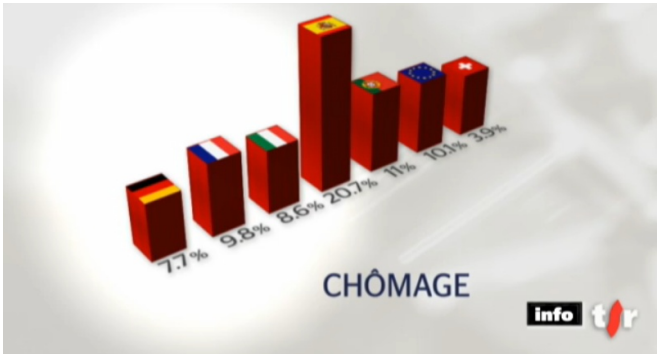



Il n'y a pas vraiment de limites à l'ingéniosité déployée pour créer des graphiques qui ne montrent plus les données originales; l'exemple ci-dessous, tiré de *PME Magazine*, a ainsi résisté jusqu'à présent à tous les efforts de compréhension de l'auteur.

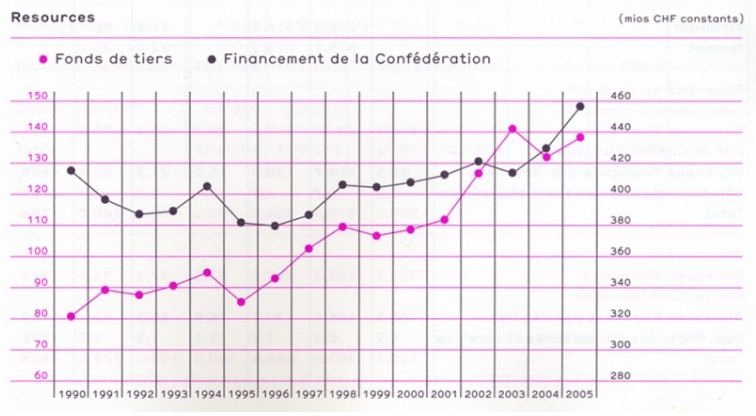


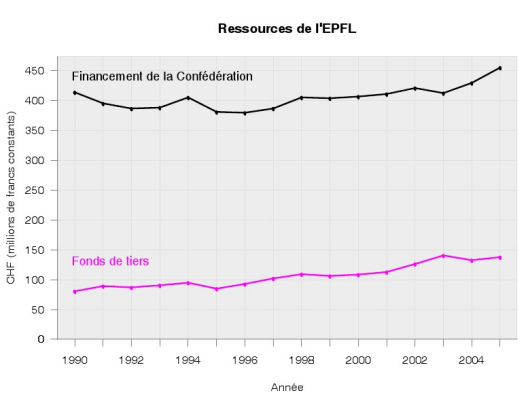

L'exemple de priorités inattendues donné ci-dessous ne provient pas des médias, mais de l'Ecole Polytechnique Fédérale de Lausanne, hôte de l'un des plus importants départements de statistiques en Suisse – mais qui semble cependant très bien séparé du service de la communication ! Sur le graphique de gauche, deux courbes, dessinées sur deux échelles différentes, semblent se croiser dans le passé proche, donnant l'impression que le financement privé de l'institution est arrivé au niveau du financement par la Confédération. Le graphique correct (en bas à droite) montre cependant une image très différente de la situation.

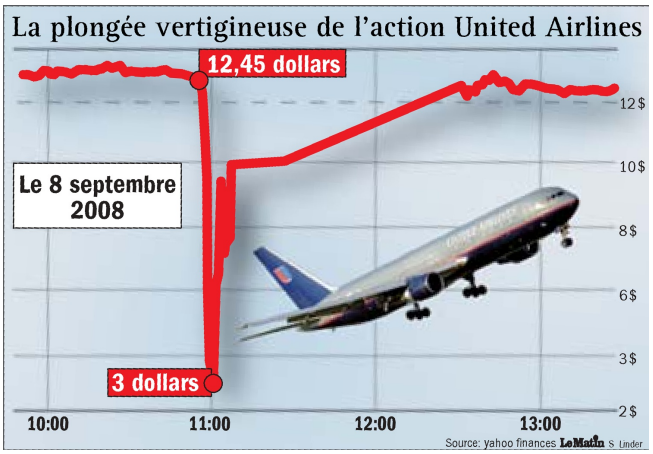
Revenons vers les médias avec un dernier exemple, tiré cette fois du monde de la télévision. L'image ci-dessous a été diffusée lors du journal de 19:30 de la *Télévision Suisse Romande*, le 13 janvier 2011.




Combinant l'absence d'axe à des données en trois dimensions, ce graphique est particulièrement difficile à lire. Un problème d'autant plus important que l'image n'apparaît que quelques secondes à l'antenne, rendant impossible la lecture des pourcentages indiqués (en écriture déformée par la perspective) au bas de chaque barre. Et pourtant, un peu de temps serait bienvenu pour permettre au téléspectateur de remarquer que la dernière barre, qui indique le taux de chômage pour la Suisse, ne peut pas représenter un taux de 3.9% en comparaison des autres données. Dans un tel cas où la lisibilité serait particulièrement importante, elle est entièrement sacrifiée sur l'autel du design.






**Les outils**

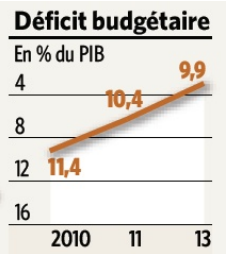
Un certain nombre d'erreurs commises par les médias peuvent sembler étonnantes au premier abord pour des statisticiens, comme le montrent les trois exemples suivants.

Sur l'infographie ci-dessus, publiée dans *Le Matin Dimanche* du 14 septembre 2008, l'axe vertical apparaît irrégulier: difficile de réconcilier le point indiqué en rouge par «3 dollars» avec les valeurs de l'axe. Il faut un moment d'attention pour se convaincre que l'on a affaire à une erreur d'annotation de l'axe (la ligne marquée 3$ correspond en fait à 4$). Si l'erreur est mineure, elle pose un problème de compréhension, et empêchera certains des lecteurs de saisir immédiatement le contenu du graphique.

Deuxième exemple, le diagramme en camembert ci-dessous, publié dans *Le Matin*. En plus de l'effet en trois dimensions déjà discuté, on voit que ce sont 105.4% des participants qui ont donné leur avis !


Une inversion de chiffres est probablement à l'origine de l'erreur: si la tranche de 8.2% était en fait de 2.8%, la somme ferait bien 100%. Ce chiffre corrigé est plus cohérent avec la méthode de sondage utilisée (réalisé sur Internet, avec auto-sélection des participants), dont on attend probablement un faible taux de réponses sans avis.

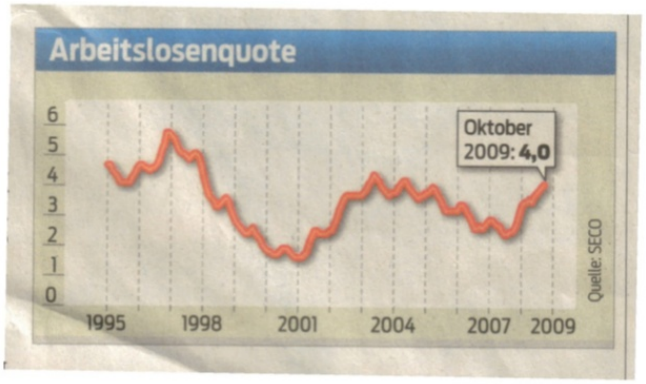
Finalement, le graphique ci-contre, publié par *Le Temps* du 31 août 2011, montre le déficit budgétaire du Royaume-Uni entre 2010 et 2013. Tout d'abord, la conception du graphique est inhabituelle (axe vertical allant de haut en bas), donnant une impression faussée d'augmentation du déficit budgétaire, et l'axe horizontal n'est pas linéaire. Mais c'est surtout la cohérence entre les points et les axes qui pose problème: les valeurs indiquées ne correspondent pas du tout aux graduations de l'axe. Il est impossible de savoir si ce sont les chiffres ou la droite qui sont corrects.




Ces trois exemples ont un point commun: ils décrivent des erreurs que des statisticiens ne commettraient pas. Pas parce que ces derniers sont au dessus de ce genre de problèmes, mais parce qu'ils ne dessinent jamais un axe ou un diagramme en camembert à la main, et qu'ils ne tracent généralement pas une courbe séparément de son système d'axes. Dans les trois cas, leur logiciel préféré (par exemple R) s'occupe de ces tâches, et fait généralement bien son travail. Même Excel, souvent décrié, produit généralement des résultats corrects de ce point de vue.

Les infographistes, de leur côté n'utilisent généralement pas de logiciel de statistiques, mais des logiciels de graphisme, avec en tête Adobe Illustrator. Ceux-ci leur donnent tout contrôle sur le contenu des graphiques, leur permettant de modifier tous les détails à leur convenance. Mais avec ce grand pouvoir vient une grande responsabilité: les graphiques sont considérés comme une série de courbes, et non pas comme un tout, et il est facile d'en modifier une partie et de l'altérer sans s'en rendre compte, rendant le résultat incohérent.

Le graphique ci-dessous, publié dans le Blick du 7 novembre 2009, est un autre exemple de technique souvent utilisée dans le monde de l'infographie. L'étiquette attachée à la fin de la courbe ne semble pas correspondre à l'échelle horizontale (qui est elle-même assez peu claire). Une comparaison avec les données originales (disponibles sur le site internet du Secrétariat d'Etat à l'économie) montre que le premier point, début 1995, est correct, mais que la courbe avance plus vite que l'axe, le point correspondant à octobre 2009 apparaissant *avant* le début de l'année 2009. Vraisemblablement, l'infographiste n'avait pas à disposition les données chiffrées, mais uniquement une version graphique. Il a ensuite redessiné la courbe dans le logiciel graphique, en la superposant à la




courbe originale et en ajoutant manuellement des axes, pour obtenir le résultat cohérent avec la charte graphique de son journal. Cette technique est rapide et fréquemment utilisée, mais comme on le voit, elle peut facilement conduire à des erreurs.

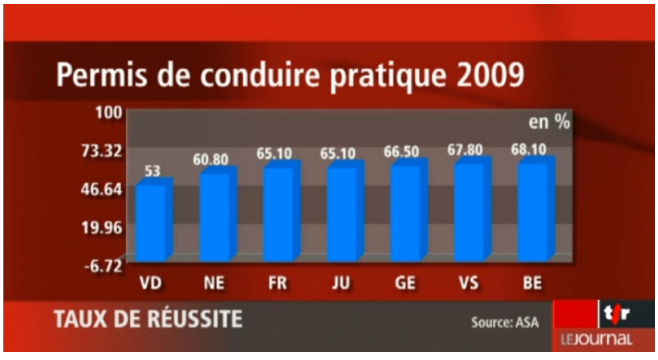
Si de telles opérations manuelles peuvent être dangereuses, l'automatisation n'est pas forcément la meilleure solution, surtout quand on utilise de mauvais outils. Ainsi, le graphique ci-contre, présenté lors du journal de 19:30 de la *Télévision Suisse Romande* du 1er avril 2010, indique le taux de réussite au permis de conduire en 2009 dans différents cantons. Il montre des données qui semblent correctes (même si elles sont affublées d'un effet tri-dimensionnel inutile), mais avec un axe fantaisiste qui démarre à -6.72. Les graduations calculées automatiquement sont probablement correctes également, mais elles sont peu intuitives et apportent de la confusion au lieu de la lisibilité.

Les occurences de ce type de problèmes sont nombreuses, en particulier avec les proportions, qui sont souvent représentées sur des graphiques dont les axes sortent du domaine des valeurs possibles (exemple fréquemment rencontré: un axe vertical qui se prolonge jusqu'à 120%).

**4. Conclusions: quelle place pour le statisticien ?**

Les statistiques sont loin d'être la discipline qui fait l'objet du plus de vulgarisation auprès du grand public et dans les médias, même si les tentatives ne sont pas inexistantes (trois bons exemples sont donnés en référence à la fin de l'article). Il y a donc probablement une place à prendre dans ce domaine, surtout les quelques fois où elles «font» l'actualité (on peut citer par exemple le procès de la Banque Cantonale de Genève, où les statistiques ont eu une place importante pour montrer que le tirage au sort des jurés n'était pas aléatoire).

A titre d'exemple, l'auteur de ces lignes a le plaisir d'écrire une chronique régulière dans la page «Sciences&Environnement» du *Temps*; l'un de ces textes a d'ailleurs été republié dans le précédent bulletin de la SSS, et montrait comment juger de l'efficacité des prévisions astrologiques (en introduisant le concept de test de permutation pour le grand public). En parallèle, il donne une heure de cours dans le cadre de la formation des journalistes au CRFJ; cet enseignement, basé principalement sur des exemples similaires à ceux présentés ici, accompagnés de quelques principes généraux («corrélation n'implique pas causalité», et, bien entendu, une présentation de la médiane !) est très court, mais devrait permettre, espérons-le, d'éviter quelques unes des erreurs les



plus grossières. De son côté, l'Office fédéral de la statistique donne des cours similaires dans des programmes de formation de journalistes en Suisse allemande.

Il vaut probablement la peine de transmettre aux médias les problèmes découverts par des statisticiens; la plupart des journalistes et des infographistes contactés au fil du temps ont répondu très positivement aux questions et commentaires qui ont été faits sur leur travail (même s'ils étaient quelques fois peu convaincus au premier abord). Comme l'ont montré les exemples décrits plus hauts, tous les médias sont concernés (à des degrés différents) par ces problèmes. Il ne sera pas possible d'éliminer toutes les erreurs (de la même façon que certaines coquilles passent entre les gouttes des correcteurs), mais il y a certainement des opportunités pour des améliorations.

**5. Post scriptum: la genèse**

Le contenu de cet article est né d'un cours «de service» (de statistiques, mais destiné à des non-statisticiens) enseigné à l'Université de Genève. Les étudiants doivent, chaque semaine, collecter au moins trois exemples d'utilisation des statistiques dans les médias, des publications scientifiques ou internet, et expliquer pourquoi ces exemples sont particulièrement bons ou mauvais, ces commentaires faisant ensuite partie de leur évaluation. Les étudiants jouent généralement très bien le jeu et à la fin du semestre, ils ne regardent plus une information statistique de la même manière. Cet esprit critique est probablement ce qu'ils peuvent retirer de plus utile d'un tel cours, loin devant la théorie et autres formules mathématiques.

**Références**


- Olivier Dessibourg. «Quand les chiffres travestissent la réalité». *Le Temps*, 19 octobre 2010.
- Daniel Saraga. «Prédire le futur: probablement possible». *Le Temps*, 26 février 2011.
- Lucia Sillig. «Sondages, loterie et Internet». *Le Temps*, 16 mars 2011.


**Remerciements**


Merci en particulier à Olivier Dessibourg, Samuel Rouge, Joël Sutter et Gilles Laplace, qui m'ont permis de découvrir le monde du journalisme et de l'infographisme.