

Scruter les sociétés, après l'"homme moyen'

Claire Lemercier

▶ **To cite this version:**

Claire Lemercier. Scruter les sociétés, après l"homme moyen'. Mokrane Bouzeghoub; Rémy Mosseri. Les Big Data à découvert, CNRS Éditions, pp.268 - 269, 2017, 9782271114648. hal-01520605

HAL Id: hal-01520605
https://sciencespo.hal.science/hal-01520605

Submitted on 22 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Version auteure de Claire Lemercier, « Scruter les sociétés, après l''homme moyen' », in Mokrane Bouzeghoub et Rémy Mosseri (éd.), *Les Big Data à découvert*, Paris, CNRS Éditions, 2017, p. 268-269.

Scruter les sociétés, après l'« homme moyen »
Claire Lemercier

Les Big Data contribuent à l'étude des sociétés en mobilisant des quantités inédites de données, souvent produites, au départ, pour des raisons n'ayant rien à voir avec la recherche ; et en développant de nouveaux algorithmes d'analyse, adaptés à leur masse et à leur caractère souvent peu structuré (cf. VII.13). Elles marquent ainsi une nouvelle étape dans la longue histoire des statistiques sociales – une histoire de science, mais aussi d'États, comme le terme « statistique » l'indique, et d'entreprises.

Les données sur la société : des traces de pratiques

On peut faire commencer cette histoire avec les premières tables de mortalité, qui paraissent dans les années 1660 et qui recensent les décès à Londres selon l'âge (figure 1). En livrant des probabilités conditionnelles de survie, elles introduisent de nouvelles manières de calculer remarquées dans les milieux savants. Un de leurs objectifs est de mieux comprendre les épidémies, notamment la peste ; et l'assurance sur la vie s'empare rapidement de l'outil. Par ailleurs, même si le traitement statistique est inédit, les données, elles, n'ont pas été produites pour la recherche : elles proviennent des registres d'enterrement de l'Église.

Au cours des siècles suivants, des données en plus grande quantité ont été collectées, des méthodes différentes ont été développées pour les traiter, mais les statistiques sociales ont conservé la même démarche, au croisement de la science (mathématiques et physique, puis histoire, archéologie, sciences sociales et informatique), de l'administration et des entreprises. Les chercheurs ont parfois eu les moyens de produire des données (par questionnaire) spécifiquement dans un but de connaissance et les recensements ont inclus certaines de leurs questions à côté de celles des gouvernements. Mais le plus souvent, la recherche sur les sociétés a travaillé sur des traces de pratiques constituées et préservées pour des raisons non scientifiques, par des entreprises ou des administrations qui avaient la possibilité d'inciter, voire d'obliger les individus à fournir telle ou telle information.

Ces traces sont très variées : les squelettes permettent de dire quelque chose de l'épidémiologie du passé, les tickets de caisse de travailler sur la consommation, les statistiques policières disent quelque chose des illégalités, etc. Mais l'accès à ces traces n'est jamais facile et elles présentent de nombreux biais. Ainsi, toute la population n'est pas enterrée dans des cimetières, la consommation porte aussi sur des produits non achetés au supermarché et la police ne connaît et ne documente pas toutes les illégalités. De plus, les questions des chercheurs ne correspondent jamais exactement à ce que les sources permettent de dire : toutes les maladies ne laissent pas des traces qui se conservent, les tickets de caisse ne disent rien de la position sociale de l'acheteur et la police nomme ce qu'elle observe dans son propre vocabulaire.

L'apport des sciences sociales

Le cœur de métier des sciences sociales est de réfléchir sur ces biais. Pour cela, elles ont développé une capacité à « inventer » des sources parfois inattendues pour répondre à des questions scientifiques (c'est-à-dire à identifier des traces qui pourraient avoir été collectées et préservées dans d'autres buts) et à « critiquer » ces sources, c'est-à-dire à réfléchir sur leurs conditions de production, les biais qu'elles impliquent et les conséquences à en tirer pour l'interprétation.

Aux XIX^e^ et XX^e^ siècles, ces savoir-faire ont été développés dans un contexte où les données, même si elles devenaient de plus en plus abondantes, ne pouvaient pas, la plupart du temps, concerner l'ensemble des populations (cf. VII.15). Cela incitait à laisser de côté les comportements minoritaires au profit des plus fréquents, et à faire émerger un « homme moyen » (figure 2), expression introduite par le statisticien Adolphe Quetelet au XIX^e^ siècle. Les entreprises acceptaient de se concentrer, par exemple, sur la « ménagère de moins de 50 ans ». Les sciences sociales des années 1950 et 1960, notamment dans leurs applications au service des politiques publiques, envisageaient une « société de masse » et des solutions uniformes (cf. VII.9). Lorsqu'on tentait de contacter des extra-terrestes, on leur envoyait, à côté d'une représentation des nombres, de quelques constantes physiques et formules chimiques et d'une représentation de l'ADN, une silhouette d'être humain générique, accompagnée de l'effectif de la population mondiale à la date du message et d'une indication de la stature moyenne.

Depuis les années 1980, sciences sociales, politiques publiques, marketing ou encore médecine ont revalorisé les exceptions, les configurations plus « micro » (inter-individuelles) et la notion de complexité.

D'où leur intérêt, aujourd'hui, pour les grandes masses de données qui semblent tendre à l'exhaustivité. Est-ce pour autant la fin du travail d'invention et de critique des sources ? Le croire serait aussi dangereux pour la recherche que pour les sociétés. Le coût et les bénéfices potentiels liés à l'utilisation de ces informations font que, comme les bases de données plus réduites du passé, elles ne sont pas toujours accessibles aux chercheurs. Surtout, la masse ne fait pas disparaître la question des biais. Ainsi, la population qui tweete à propos de santé ou d'alimentation n'est pas une portion représentative de la société. En effet, dans ces domaines, les écarts entre classes sociales restent abyssaux et il est probable que la partie de la société qui préoccupe le plus les politiques publiques, c'est-à-dire celle qui mange « mal » et meurt jeune, est à peu près absente des données.
Évitons le fantasme de la carte à l'échelle 1/1, décrit par Jorge Luis Borges dans une nouvelle de 1946 (et avant lui par Lewis Carroll) : les données ne sont pas la société. Pour approcher malgré tout cette dernière (comme du reste pour comprendre les fantasmes de connaissance parfaite et immédiate que font naître les Big Data), les savoir-faire critiques accumulés par les sciences humaines et sociales restent notre meilleur guide.

Références bibliographiques

Alain Desrosières, *La Politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte, 1993.
Étienne Ollion et Julien Boelaert, « Au-delà des *big data*. Les sciences sociales et la multiplication des données numériques », *Sociologie*, 2015.

Illustrations :
https://commons.wikimedia.org/wiki/File:Houghton_HEW_11.11.1_-_Graunt_1.jpg
Couverture d'une des premières tables de mortalité, publiée en 1665, attribuée à John Graunt

https://commons.wikimedia.org/wiki/File:Arecibo_message.svg
Extrait du message radio d'Arecibo, envoyé en direction de l'amas globulaire M13