

Une équipe de chercheurs du LIEPP analyse une base de donnée de 1 million deux cent mille personnages de l'histoire depuis 3000 avant Jésus Christ

L'article de Morgane Laouenan, Olivier Gergaud et Etienne Wasmer est le fruit d'un travail de deux ans réalisé au LIEPP (Laboratoire Interdisciplinaire d'Evaluation des Politiques Publiques) à Sciences Po. Les auteurs ont collecté et analysé des données de 1 243 776 personnages notables (*notable people*) et 7 184 575 lieux auxquels ils sont liés et localisés par leurs longitude et latitude (*Geolinks*) à travers l'histoire (3000 avant JC jusqu'en 2015).

L'article décrit en détail la méthodologie et les procédures d'extraction de l'information pertinente à partir de la base communautaire *Freebase* recensant des individus et en appariant de façon automatique ces informations à des pages *Wikipedia* et ensuite analyse la base. Dix faits principaux émergent.

1. Il y a une croissance exponentielle de la base, avec plus de 60% de personnalités encore en vie en 2015, à l'exception d'un déclin relatif des cohortes nées au XVIIIe siècle avec un minimum entre 1645 et 1655.
2. La durée de vie de ces personnages a augmenté de 20 ans environ, de 60 jusqu'à un peu moins de 80 ans, entre les cohortes nées en 1400 après JC et celle née en 1900.
3. La part des femmes dans la base présente une forme en U, avec un minimum aux alentours du XVIIIe siècle et atteint 25% pour les cohortes les plus récentes.
4. La part des personnages dans la catégorie Gouvernance (politique, religieux, militaire, noblesse) a diminué au cours du temps et la part dans les activités artistiques, littéraires, media et sport a augmenté au cours des siècles; le sport a rattrapé la catégorie arts/littérature/media pour les cohortes nées en 1870 et ensuite restée au même niveau jusqu'aux cohortes nées en 1950; puis la catégorie « sport » domine la base de données après 1950.
5. Les 10 personnalités les plus visibles nées avant 1890 sont toutes non-américaines; toutes sont de nationalités différentes. En revanche, six des 10 personnalités nées après 1890 sont des citoyens américains. Depuis 1800, la part des personnages nés en Europe et aux Etats-Unis diminue et le nombre de personnes d'Asie et de l'hémisphère Sud augmente pour atteindre 20% de la base en 2000. Le barycentre (centre de gravité) de la base se situait entre la Grèce et la Turquie entre 2000 avant JC et 500 après JC, puis se déplace vers l'Ouest. Ce même centre se retrouve dans le Nord-Est de la France entre 1500 et 1700. Avec le peuplement de l'Amérique du Nord, ce dernier se déplace dans l'Atlantique, au large des côtes françaises et espagnoles.
6. La distance moyenne entre le lieu de naissance et celui de décès suit une courbe en U également: la distance moyenne était de 316km avant l'année 500 après JC, de 100km entre 500 and 1500, et a augmenté continuellement depuis. La plus grande part de cette distance est franchie entre l'âge de 15 ans et de 25 ans.

7. Les individus les plus visibles sont les plus distants de leur lieu de naissance, avec une distance médiane de 785km pour le 1% des individus de l'histoire les plus visibles, à comparer avec 389km pour les 10% les plus visibles et 176km pour l'ensemble de ces personnages.

8. Dans toutes les activités, on observe une augmentation de la mobilité internationale depuis 1960. La part des lieux visités situés dans un pays différent du pays de naissance est passée de 15% en 1955 à 35% après 2000.

9. Il n'y a pas de corrélation positive entre la taille des villes et la visibilité des individus mesurée en 2015. Cette corrélation serait plutôt négative.

10. Enfin, sur la période 1800-1939 analysée dans cette version 1.0 du texte, il existe une corrélation positive et stable entre le nombre d'entrepreneurs à une date donnée et la croissance urbaine de la ville où ils sont localisés au cours de la décennie suivante ; de façon peut-être moins attendue, c'est également vrai du nombre d'artistes à une date donnée qui affecte la croissance les années suivantes; en revanche, on trouve une corrélation nulle ou négative entre le nombre de « gouvernants » (politiques, militaires, religieux et noblesse) et la croissance des décennies suivantes.