

→ Aperçu

LES STATISTIQUES

EILEEN MAGNELLO & BORIN VAN LOON

edp sciences

Dans la même collection:

- L'infini en images*, 2016, ISBN: 978-2-7598-1771-9
Stephen Hawking en images, 2016, ISBN: 978-2-7598-1966-9
L'intelligence artificielle en images, 2015, ISBN: 978-2-7598-1772-6
Les mathématiques en images, 2015, ISBN: 978-2-7598-1737-5
La génétique en images, 2015, ISBN: 978-2-7598-1767-2
La logique en images, 2015, ISBN: 978-2-7598-1748-1
La relativité en images, 2015, ISBN: 978-2-7598-1728-3
Le temps en images, 2014, ISBN: 978-2-7598-1228-8
La théorie quantique en images, 2014, ISBN: 978-2-7598-1229-5
La physique des particules en images, 2014, ISBN: 978-2-7598-1230-1
La psychologie en images, 2014, ISBN: 978-2-7598-1231-8

Édition originale: Statistics, © Icon Books Lts, London, 2013.

Traduction: Alan Rodney - Relecture: Gaëlle Courty

Imprimé en France par Présence Graphique, 37260 Monts

Mise en page de l'édition française: studiowakeup.com

ISBN: 978-2-7598-1770-2

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays. La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (alinéa 1^{er} de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

© EDP Sciences, 2016

La noyade par les nombres

Nous sommes noyés aujourd'hui sous les statistiques. Et il ne s'agit pas uniquement de nombres. Pour les médias, les statistiques sont systématiquement « damnées », « horribles », « létales », « troublantes » – ou aussi, de temps à autre, « encourageantes ». La presse suggère constamment que les données statistiques qui traitent de crimes, de maladies, de pauvreté et de retards dans les transports sont non seulement une source de maux, mais représentent aussi des entités réelles ou des personnes réelles au lieu de se limiter à n'être que des points sur un graphique.

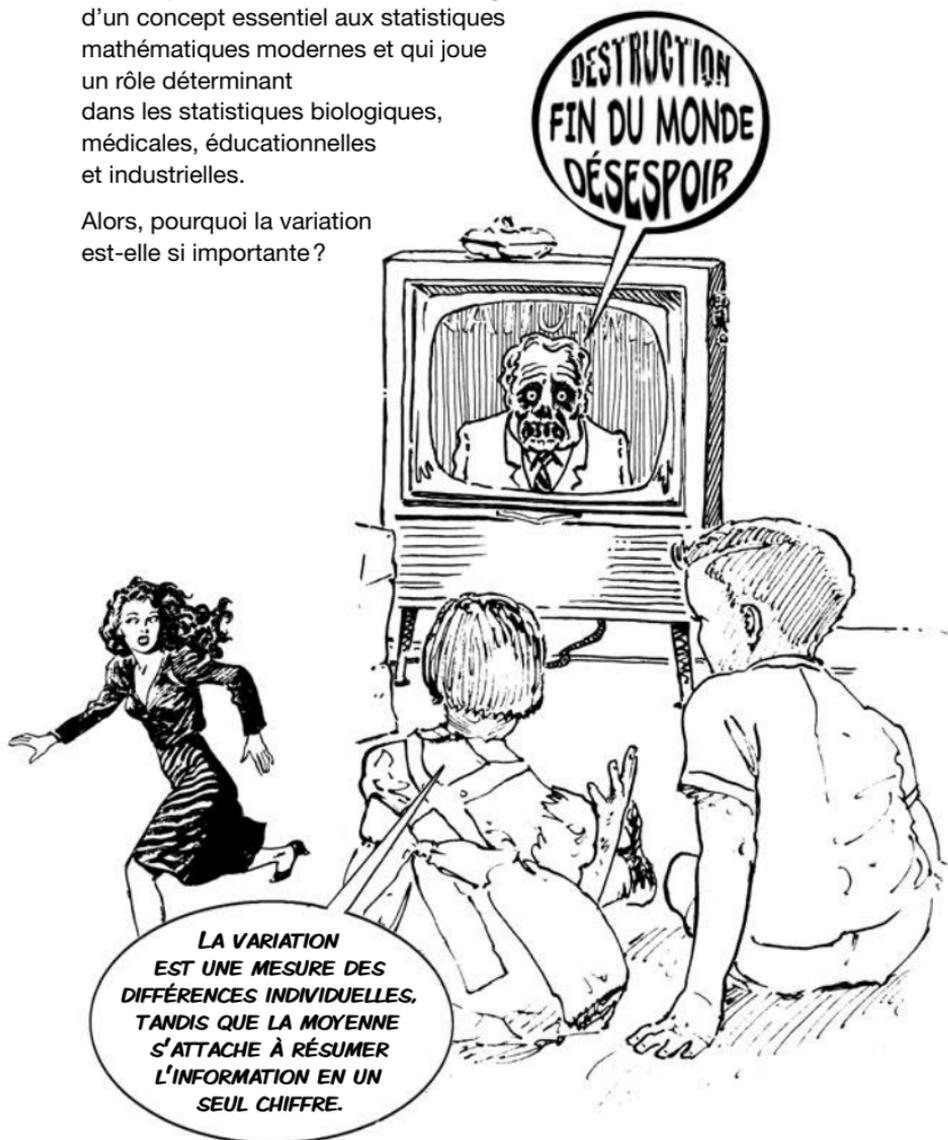
CETTE TENDANCE À AFFECTER UN SENS À UNE SEULE ENTITÉ, PAR EXEMPLE EN SE FOCALISANT SUR UN POINT D'UNE DISTRIBUTION STATISTIQUE, CRÉE BEAUCOUP DE CONFUSION ET DE CRAINTES INUTILES.



Moyennes ou variations ?

Une grande partie des informations statistiques qui véhiculent des histoires choc et d'horreur, utilisées par les médias, sont basées sur des moyennes **statistiques**. Et, en dépit de cette obsession trompeuse pour les moyennes, le concept statistique le plus important, mais négligé par les journalistes et les reporters, est la **variation**. Il s'agit d'un concept essentiel aux statistiques mathématiques modernes et qui joue un rôle déterminant dans les statistiques biologiques, médicales, éducationnelles et industrielles.

Alors, pourquoi la variation est-elle si importante ?



On peut facilement reconnaître des variations dans un Royaume-Uni aujourd'hui multiculturel, surtout à Londres qui regroupe plus de 300 sous-cultures avec autant de langues parlées (de l'acholi au zoulou) et treize religions différentes. Pour certains, le multiculturalisme doit mettre en valeur tout un chacun et ne pas faire en sorte que chaque individu soit identique (ou ne pas réduire des groupes ethniquement différents à une seule personne représentative).

**IL EXISTE TANT DE
DIFFÉRENCES INDIVIDUELLES AU SEIN
DE LA POPULATION BRITANNIQUE QU'IL EST
DÉSORMAIS PRESQUE ILLUSOIRE DE PARLER
D'UN CITOYEN BRITANNIQUE «MOYEN»,
COMME ON POUVAIT PEUT-ÊTRE
LE FAIRE AVANT 1950.**



Ces différences multiformes individuelles incarnent des variations statistiques, qui sont au cœur des statistiques mathématiques modernes.

Pourquoi étudier les statistiques ?

Les scientifiques, les économistes, les agents gouvernementaux, l'industrie et les fabricants, tous utilisent des statistiques. Des décisions sur des bases statistiques sont prises tous les jours et affectent nos vies – des médicaments qui nous sont prescrits, les traitements que nous recevons, les tests d'aptitude et psychométriques pratiqués régulièrement par les employeurs, les voitures que nous conduisons, les vêtements que nous portons (les fabricants de laine se servent de tests statistiques pour choisir le tissage du fil idéal pour notre confort), jusqu'à la nourriture que nous mangeons voire même la bière que nous buvons.



Une connaissance des bases de statistiques peut même sauver ou prolonger nos vies – comme ce fut le cas pour un certain Stephen Jay Gould, dont nous reparlerons plus tard.

Les statistiques, que sont-elles au juste ?

Tout en étant conscients qu'elles sont omniprésentes, nous ne savons pas quoi penser précisément des statistiques. Pour citer un journaliste de la presse écrite, « ce sont les cigarettes qui constituent la plus grande origine des statistiques ». Les gens expriment le souhait d'écarter le mauvais sort en disant : « je n'ai pas envie d'être une simple statistique ». Est-ce que les statisticiens pensent que l'humanité peut être réduite à quelques chiffres ?

Bien que certains estiment que les résultats statistiques sont irréfutables, d'autres pensent en revanche que toute information statistique est sujette à caution.



Bien que Mark Twain ait fait l'erreur d'attribuer cet aphorisme au Premier ministre britannique de l'époque, Benjamin Disraeli, en 1904, en réalité, le premier à le prononcer fut un autre Britannique, le baron Leonard Henry Courtney, professeur d'économie à l'université de Londres et, par ailleurs, président de la *Royal Statistical Society* (société savante), lors d'une intervention à Saratoga Springs, dans l'État de New York en 1895, qui traitait de la représentation proportionnelle des 44 États américains.

Certains représentants du gouvernement voient même dans les statistiques la source de problèmes économiques. Quand l'attaché de presse de la Maison-Blanche Scott McClellan a essayé d'expliquer en février 2004 la raison pour laquelle le gouvernement Bush avait fait marche arrière, révisant à la baisse une prévision qui aurait vu la création de davantage de postes aux États-Unis, sa ligne de défense était des plus simples.



Au Royaume-Uni, la *Statistics Commission* (Comité national des statistiques, équivalent de l'INSEE, Institut national de la statistique et des études économiques) a même recommandé « que les ministres du Gouvernement britannique n'aient pas accès aux données statistiques avant leur publication officielle, et ce, afin d'éviter toute interprétation ou utilisation politique ». Néanmoins, les statistiques mises à disposition du public peuvent (dé)former les opinions, influencer les politiques gouvernementales et même (dés)informer les citoyens, s'agissant parfois de découvertes scientifiques ou d'innovations technologiques remarquables.

Que signifie le terme « statistique » ?

Le mot « statistique » vient du latin *status*, qui a donné le mot italien *statista*, dont la première utilisation remonte au *xvi^e* siècle, se référant à un « statiste » ou homme d'État, c'est-à-dire quelqu'un qui s'occupe des affaires de l'État. Les Allemands, en 1750, employaient le terme « *Statistik* », les Français ont introduit « statistique » en 1785 et les Hollandais ont adopté « *statistiek* » en 1807.



LES PREMIÈRES
STATISTIQUES ÉTAIENT
UN SYSTÈME QUANTITATIF
POUR DÉCRIRE DES QUESTIONS
ÉTATIQUES - UNE FORME
D'« ARITHMÉTIQUE
POLITIQUE ».

Le système statistique fut utilisé pour la première fois en Angleterre au *xvii^e* siècle par le marchand londonien **John Graunt** (1620–1674) et le physicien irlandais **William Petty** (1623–1687).



À noter qu'au XVIII^e siècle, bon nombre de statistes étaient des juristes formés souvent en droit public (la spécialité qui traite directement des affaires d'État).

Mais il revient à **Sir John Sinclair** (1754–1834), propriétaire terrien écossais et premier président du Conseil national de l'agriculture, d'avoir introduit le mot « statistics » en anglais en 1798 dans son *Statistical Account of Scotland* (Compte statistique sur l'Écosse).

UN QUANTUM DE BONHEUR



Sinclair s'est servi des statistiques pour étudier des phénomènes sociaux plutôt qu'à des fins politiques. Quelque temps plus tard, à la moitié du XIX^e siècle, ses travaux débouchèrent sur des statistiques vitales.

Statistiques vitales versus statistiques mathématiques

Les statistiques ne sont pas toutes de même nature. En réalité, il y en a deux sortes : les statistiques vitales et les statistiques mathématiques.

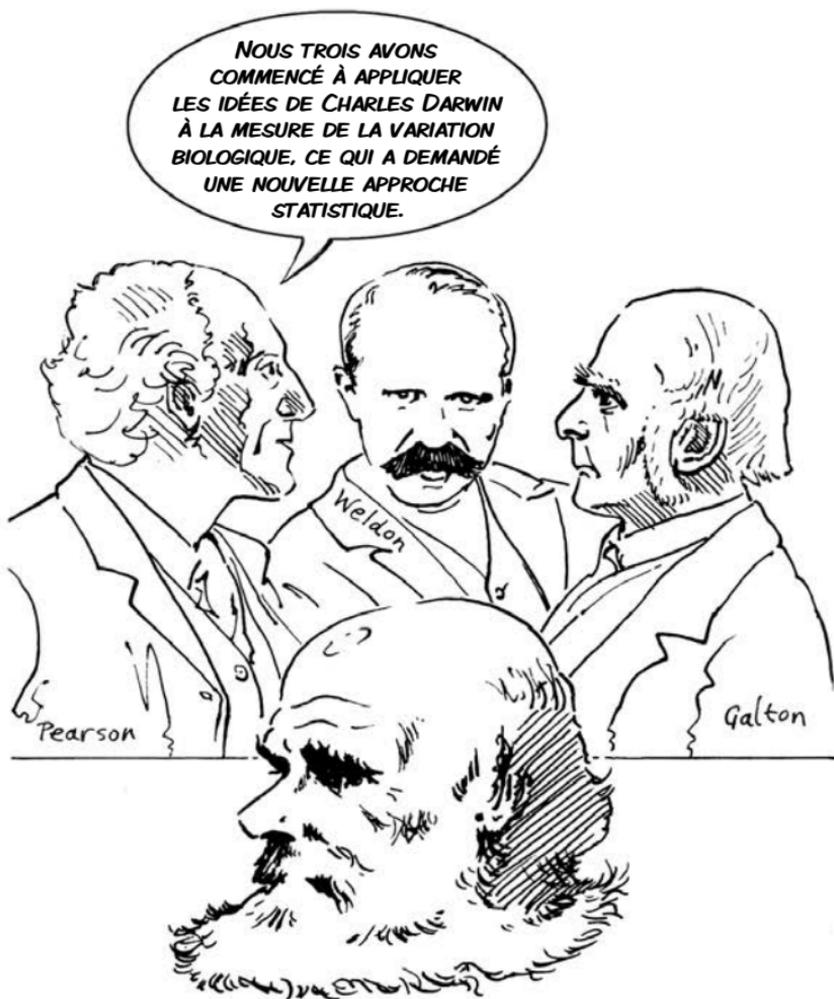
Les statistiques vitales correspondent à ce que la plupart des personnes assimilent aux statistiques. Employées au pluriel, elles désignent un ensemble agrégé de données.



La procédure employée concerne principalement les valeurs moyennes et met en œuvre des tables de données sur la vie, les pourcentages, les proportions et les ratios : les calculs de probabilité sont le plus souvent utilisés à des fins d'actes notariés (assurance-vie). Ce n'est qu'au ^{xx}e siècle qu'est apparu le terme « statistique » au singulier, impliquant ainsi des faits individuels.

Les mathématiques statistiques ont pris leur essor à partir de la théorie mathématique de la probabilité, à la fin ^{xviii} siècle, et des travaux du Suisse Jacob Bernoulli, des Français Abraham de Moivre et Pierre-Simon de Laplace, et de l'Allemand Johann Carl Friedrich Gauss.

À la fin du ^{xix} siècle, les statistiques mathématiques ont commencé à prendre la forme d'une discipline académique à part entière, notamment dans les travaux de **Francis Ysidro Edgeworth** (1845–1926), **John Venn** (1834–1923), **Francis Galton** (1822–1911), **W. F. R. Weldon** (1860–1906) et **Karl Pearson** (1857–1936).



Les statistiques mathématiques englobent une discipline scientifique qui analyse les variations et basée le plus souvent sur l'algèbre matricielle. Elles traitent de la collecte, l'ordonnancement, la description et l'interprétation de données issues d'enquêtes sociales, d'expérimentations scientifiques et d'essais cliniques. La probabilité est utilisée pour des tests statistiques de significativité.



LES STATISTIQUES MATHÉMATIQUES SONT DE NATURE ANALYTIQUE ET PEUVENT SERVIR À ÉLABORER DES PRÉVISIONS STATISTIQUES OU DES INFÉRENCES RELATIVES À UNE POPULATION DONNÉE.

DE PLUS, ELLES CAPITALISENT LES DIFFÉRENCES INDIVIDUELLES DES MEMBRES D'UN GROUPE EN EXAMINANT L'ÉTENDUE DE CETTE VARIATION STATISTIQUE AU TRAVERS DE MÉTHODES CLASSIQUES, TELLES QUE LES MESURES DE PORTÉE OU D'ÉCART-TYPE, QUE NOUS ALLONS DÉCRIRE EN DÉTAIL PLUS LOIN.

LES STATISTIQUES VITALES CALCULENT DES MOYENNES, TANDIS QUE LES STATISTIQUES MATHÉMATIQUES ANALYSENT DES VARIATIONS.

Prises dans ce sens, les statistiques représentent une discipline technique et, dans la mesure où elles sont mathématiques, il est important de bien comprendre les concepts statistiques qui sous-tendent ces procédures mathématiques.

La philosophie des statistiques

La décision d'examiner des moyennes ou de mesurer des variations trouve ses racines dans des idéologies philosophiques qui gouvernaient la pensée des statisticiens, des physiciens et des scientifiques tout au long du XIX^e siècle. Le fait de se focaliser sur des moyennes statistiques était soutenu par les défenseurs du **déterminisme**

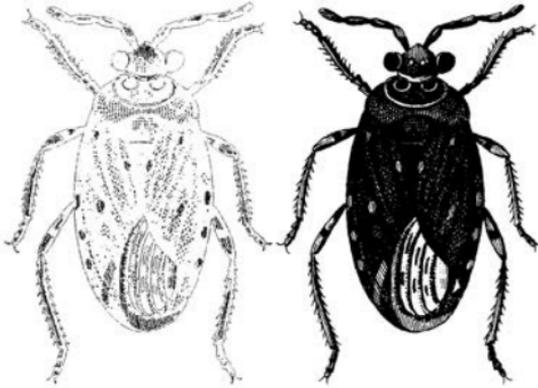
et des idées **typologiques** des espèces biologiques, ce qui a permis de perpétuer l'idée d'une moyenne idéalisée.

Le déterminisme implique l'existence d'ordre et de perfection dans l'Univers...

AINSI, LES VARIATIONS SONT IMPARFAITES ET PERÇUES COMME DES ERREURS QUI DOIVENT ÊTRE ÉRADIQUÉES, DANS LA MESURE OÙ ELLES INTERFÈRENT AVEC LE GRAND DESSEIN DE DIEU ET SON OBJECTIF POUR SON MONDE.



Le concept typologique des espèces, qui constituait la pensée dominante des taxonomistes*, des typologistes et des morphologistes jusqu'à la fin du XIX^e siècle, a donné lieu au concept morphologique des espèces. Ces dernières sont censées représenter un type idéal.



La présence d'un type idéal a été déduite d'une forme de similitude morphologique, qui devient le critère de l'espèce pour les typologistes. Cela aurait pu avoir pour conséquence de créer une prolifération des espèces, puisque toute déviation d'un type donné aurait conduit à la détermination d'une nouvelle espèce.

Les véritables changements, selon le concept morphologique des espèces, n'étaient possibles que par **saut ou bond** vers une nouvelle espèce d'une génération à la suivante. De ce fait, la théorie de Darwin, qui était basée sur des changements « graduels », était incompatible avec l'essentialisme.



* Les **taxonomistes** classent les organismes vivants par groupes.
 Les **typologistes** classent les organismes par types d'ordre général.
 Les **morphologistes** étudient les formes physiques des organismes.

Darwin et les populations statistiques

La transition vers des mesures de variations statistiques représentait une remise en cause idéologique, apparue au milieu du XIX^e siècle quand **Charles Darwin** (1809–1882) a commencé ses observations sur de minuscules variations biologiques chez des animaux et des plantes.



Chaque idée de Darwin, depuis les variations, la sélection naturelle, l'hérédité de traits, en passant par la réversion, semblait exiger des analyses statistiques.

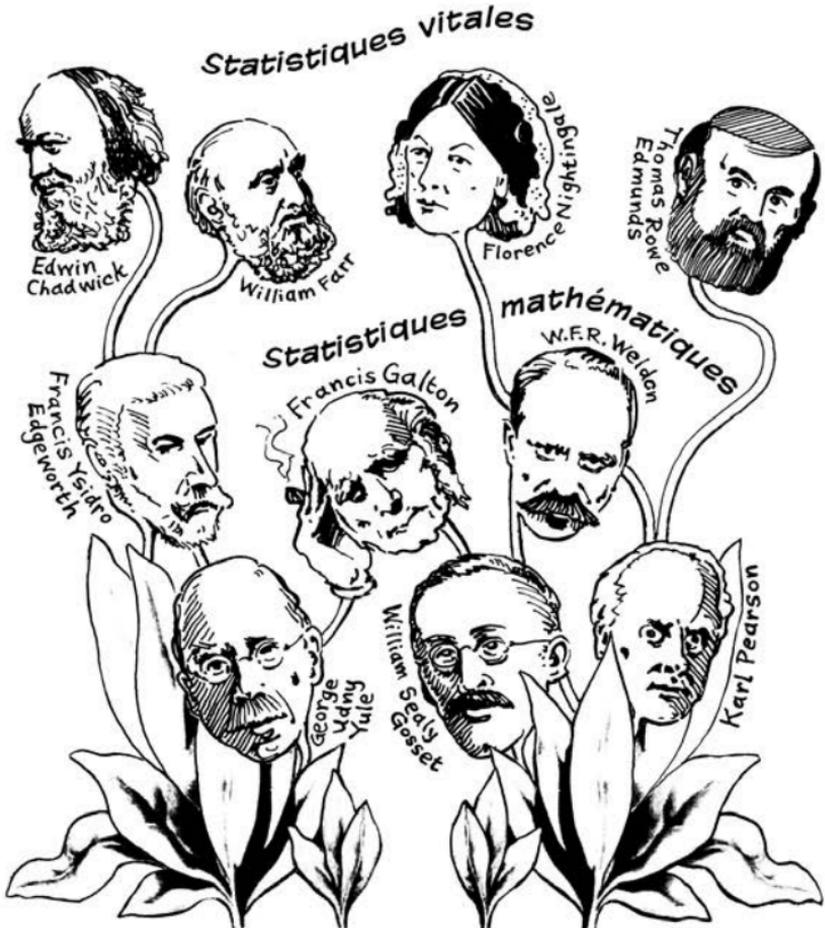
Darwin n'a pas seulement démontré que la variation biologique était mesurable et prenait du sens en soulignant l'existence de populations statistiques plutôt que se focaliser sur un type d'animal ou une essence de plante, mais il a analysé, de plus, plusieurs formes de corrélations qui pouvaient expliquer la sélection naturelle. Comme l'a remarqué en 1931 le biologiste évolutionniste **Sewall Wright** (1899–1988):

DARWIN ÉTAIT LE PREMIER À CONSIDÉRER VÉRITABLEMENT L'ÉVOLUTION PRINCIPALEMENT COMME UN PROCESSUS STATISTIQUE.

A black and white illustration of Sewall Wright, a man with curly hair wearing a suit and tie. He is shown from the chest up, looking slightly to the left. A speech bubble is positioned in front of him, containing text. His signature 'S. Wright' is visible at the bottom right of the illustration.

Les valeurs victorienne

Bien que plusieurs développements aient vu le jour en Europe continentale dans les domaines des statistiques vitales et mathématiques, nous sommes très redevables, concernant les rapides dissémination et application des statistiques vitales au milieu du XIX^e siècle et des statistiques mathématiques à la fin du XIX^e et au début du XX^e siècle, aux Victoriens suivants :



Le développement des deux types de statistiques a émergé de la culture de la mesure des Victoriens. Ces derniers appréciaient la précision et la justesse apportées par des instruments scientifiques, qui leur fournissaient des informations plus fiables. Dans le contexte d'une économie industrielle en plein essor, il était primordial de vérifier que les résultats obtenus étaient reproductibles en vue du marché international.

Les ingénieurs et les physiciens ont passé de longues heures dans les laboratoires à enregistrer et mesurer avec précision des constantes électriques, mécaniques et physiques pour des machines, des appareils et autres objets. Les biologistes et les géologues ont collecté autant de données que possible lors de leurs expéditions afin d'établir des cartes géographiques, relevant la longitude et la latitude, et classifiant de nouvelles espèces de plantes et d'animaux.



DR JOHN SNOW

LES
STATISTIQUES
OFFRENT UNE MANIÈRE
DE QUANTIFIER LES
MESURES HUMAINES,
NOTAMMENT POUR DES
SUJETS QUI TRAITENT DE
LA SANTÉ ET L'HYGIÈNE
PUBLIQUES, L'HÉRÉDITÉ
ET LA MÉDECINE.

