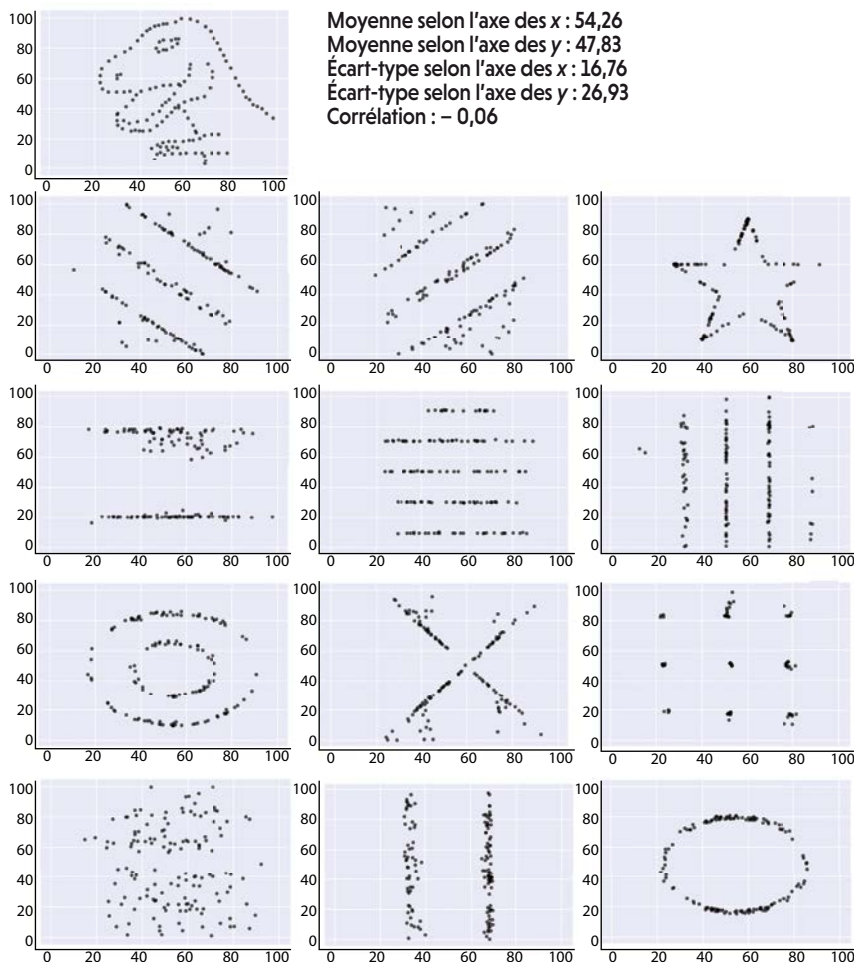


4

MÊMES PARAMÈTRES...
ET POURTANT SI DIFFÉRENTS

Nous pensons en général que si des données ont des paramètres statistiques proches (moyennes, écarts-types, etc.), elles se ressemblent nécessairement. Les paramètres statistiques, pensons-nous, sont comme des versions compressées des séries de données. Justin Matejka et George Fitzmaurice prouvent graphiquement que c'est faux. À l'aide d'une méthode qui leur permet de créer facilement des séries ayant exactement les mêmes paramètres statistiques, ils ont notamment produit ces 13 schémas, utiles à tous ceux qui souhaitent prendre conscience de notre fragile compréhension intuitive du hasard et des statistiques.

Il s'agit de graphiques de séries de couples de nombres (x, y) , qui apparaissent très différents. Pourtant, d'une série à l'autre, les 5 paramètres suivants sont les mêmes : moyenne des x , écart-type des x , moyenne des y , écart-type des y , corrélation entre les x et les y . Lorsque la corrélation entre les x et les y est parfaite, ce type de schéma est une droite. (Voir J. Matejka et G. Fitzmaurice, « Same stats, different graphs : Generating datasets with varied appearance and identical statistics », *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1290-1294, ACM, 2017.)



> toujours en escadrille!... Certaines analyses tentent d'en identifier l'origine en parlant d'une « attente excessive d'étalement » (voir mon article « Notre vision du hasard est bien hasardeuse », *Pour la Science* n° 293, mars 2002) : notre intuition nous souffle à tort que, par exemple, les dates des accidents d'avion doivent être régulièrement espacées, alors que les statistiques nous montrent autre chose.

Cette attente excessive d'étalement a été clairement analysée dans les cas précis des accidents d'avions par Élise Janvresse et Thierry de la Rue, de l'université de Rouen (voir l'encadré 3 page 111).

DES SÉRIES TRÈS DIFFÉRENTES,
MAIS AYANT LA MÊME STATISTIQUE

Notre bon sens est défaillant pour traiter les probabilités et nous sommes surpris dans des cas où il n'y a pas lieu de l'être. Utiliser le mot « coïncidence » n'explique rien ou alors cela introduit des idées étranges comme la synchronicité de Carl Gustav Jung ou les champs morphiques de Rupert Sheldrake, dont les scientifiques ont vainement cherché à prouver l'existence (voir par exemple www.sceptiques.qc.ca/dictionnaire/).

Cette faiblesse de notre esprit à percevoir correctement les probabilités a encore récemment été mise en évidence. On est tenté de penser que si des données ont des paramètres statistiques proches, elles doivent nécessairement se ressembler; or Justin Matejka et George Fitzmaurice, de la société Autodesk Research à Toronto, ont créé délibérément des exemples magnifiques montrant que c'est faux.

Pour engendrer leurs séries, ils ont mis au point une technique qui leur permet de jouer facilement avec les formes des graphes créés. Ainsi, la figure de l'encadré 4 ci-contre visualise une collection de 13 séries de couples de nombres (x, y) qui, malgré leurs graphes très différents (chaque couple (x, y) représentant un point du plan), ont les mêmes 5 paramètres suivants : moyenne des x , écart-type des x , moyenne des y , écart-type des y , corrélation entre les x et les y .

SIMPLE ET INATTENDU

Toujours afin de comprendre et de classer les situations où notre esprit s'étonne alors qu'il ne devrait pas, détaillons à présent un cas moins connu, car lié à une théorie assez récente. La mauvaise compréhension de ce qui est probable, car simple, conduit à percevoir certains événements comme étonnants alors qu'ils ne le sont pas, et donc à croire être en présence de coïncidences miraculeuses sans bonne explication, alors que la situation est banale.

La notion la plus générale de simplicité est celle qui provient de la théorie algorithmique de l'information. Comme nous n'en avons pas