

## 0 Des amateurs scrupuleux

Au dix-neuvième siècle, une véritable avalanche de chiffres s'est abattue sur tous les secteurs de la connaissance. Chacun avait à cœur de prouver ses opinions avec des séries de nombres plus ou moins justifiés, plus ou moins farfelus. Rares étaient ceux qui se posaient vraiment la question : « que démontrent les chiffres ? » Et parmi eux, tous n'étaient pas des représentants de la science officielle.

histoires de statistique

### Des amateurs scrupuleux

*intuitions non paramétriques*



hist-math.fr

Bernard YCART

## 1 Charles Dupin (1784–1873)

À l'époque dont nous parlons, celui qui incarnait la statistique en France était Charles Dupin.

Charles Dupin (1784–1873)



## 2 Carte choroplèthe (1826)

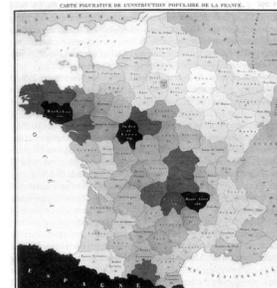
En 1826, il avait réalisé un coup médiatique, en publiant sa carte choroplèthe. Cette carte était censée représenter le niveau d'instruction des départements français. Ceux qui envoyaient le moins d'enfants dans les écoles étaient les plus sombres, et ceux où l'instruction était la plus répandue les plus clairs. Il opposait ainsi 54 départements au sud de la ligne Saint-Malo – Genève, aux 32 départements du nord. La méthodologie statistique laissait beaucoup à désirer, mais personne ne la remettait en cause.

La carte de Dupin avait fait beaucoup parler d'elle. Elle avait lancé une mode de la statistique, et pour le grand public, Charles Dupin en était l'incarnation.

Ce n'est pas pour autant qu'il était un statisticien au sens moderne du terme, ni même d'ailleurs un scientifique rigoureux. Voici par exemple un extrait d'un mémoire qu'il présente à l'Académie des sciences en 1842. Le sujet est le rapport des sexes à la naissance, qui est la pierre de touche des méthodes statistiques depuis Arbuthnot et Laplace.

Carte choroplèthe (1826)

Charles Dupin (1784–1873)



### 3 variations du rapport entre le nombre de naissances. . .

« Les résultats que je viens d'offrir démontrent évidemment qu'il est impossible de considérer comme une quantité *immuable* le rapport entre le nombre des naissances des deux sexes.

Les quinze dernières années, quoique offrant le rapport moyen le plus élevé, présentent une tendance marquée à la diminution successive du rapport des naissances. »

Dupin offre surtout une démonstration évidente. . . de ses immuables arguments d'autorité. Y avait-il un test pour démontrer ses affirmations ? Non, aucun. Cela faisait pourtant plus d'un demi-siècle que Laplace avait donné les bases de la statistique inférentielle, mais quasiment personne n'appliquait ses méthodes.

Certains, dans la presse en particulier, se posaient parfois des questions. Voici un extrait de la Revue des deux Mondes en 1843.

### 4 mieux se contenir

« Si la statistique ne sait pas mieux se contenir, elle se fera, auprès des esprits sérieux, un tort irréparable. C'est une science qui renferme des calculs et des arguments pour toutes les causes, fussent-elles diamétralement opposées. »

À la tribune de l'Assemblée, Lamartine ne se prive pas d'ironiser.

### 5 prestidigitateurs

« J'ai appris dans cette étude ce que valent les statistiques, j'en demande bien pardon aux partisans de la statistique qui se trouvent ici, excusez cette comparaison vulgaire : les prestidigitateurs font leurs tours avec des gobelets ; les économistes font leurs théories avec des statistiques. »

Ou encore, dans la Revue de Paris :

#### variations du rapport entre le nombre de naissances. . .

C. Dupin, CRAS (1842)

Les résultats que je viens d'offrir démontrent évidemment qu'il est impossible de considérer comme une quantité *immuable* le rapport entre le nombre des naissances des deux sexes.

Les quinze dernières années, quoique offrant le rapport moyen le plus élevé, présentent une tendance marquée à la diminution successive du rapport des naissances.

#### mieux se contenir

Le Revue des deux mondes (1843)

Si la statistique ne sait pas mieux se contenir, elle se fera, auprès des esprits sérieux, un tort irréparable. C'est une science qui renferme des calculs et des arguments pour toutes les causes, fussent-elles diamétralement opposées.

#### prestidigitateurs

Alphonse de Lamartine, 22 avril 1846

J'ai appris dans cette étude ce que valent les statistiques, j'en demande bien pardon aux partisans de la statistique qui se trouvent ici, excusez cette comparaison vulgaire : les prestidigitateurs font leurs tours avec des gobelets ; les économistes font leurs théories avec des statistiques.

## 6 catalogue de momies

« Sans de tels résultats, la statistique serait une science morte, un catalogue de momies. Évoquer des chiffres seulement pour qu'ils se placent dans l'addition définitive, c'est ne rien faire ; il faut leur demander ce qu'ils prouvent, et les forcer ainsi à être éloquents. »

### catalogue de momies

La revue de Paris (1829)

Sans de tels résultats, la Statistique serait une science morte, un catalogue de momies. Évoquer des chiffres seulement pour qu'ils se placent dans l'addition définitive, c'est ne rien faire ; il faut leur demander ce qu'ils prouvent, et les forcer ainsi à être éloquents.

## 7 Antoine Augustin Cournot (1801–1877)

Antoine Cournot était un des rares qui savait forcer les chiffres à être éloquents, parce qu'il avait compris la théorie de Laplace.

En 1828 il écrit un article pour l'enseignement de la théorie des probabilités.

### Antoine Augustin Cournot (1801–1877)



## 8 De la théorie des probabilités (1828)

« Où il n'y a pas de données statistiques les formules du calcul des probabilités sont illusoires ; et partout où ces données mettent en évidence les faits dont nous parlons, les formules trouvent incontestablement leur application. »

On ne fait pas de statistique sans probabilités, on n'affirme pas sans appliquer un test, on ne donne pas un résultat moyen sans préciser un intervalle de confiance. Voilà ce que veut dire Cournot.

### De la théorie des probabilités (1828)

Antoine Augustin Cournot (1801–1877)

Où il n'y a pas de données statistiques les formules du calcul des probabilités sont illusoires ; et partout où ces données mettent en évidence les faits dont nous parlons, les formules trouvent incontestablement leur application.

## 9 Exposition de la théorie des chances (1843)

« De nos jours, la statistique a pris un développement en quelque sorte exhubérant ; et l'on n'a plus qu'à se mettre en garde contre les applications prématurées et abusives qui pourraient la décréditer pour un temps, et retarder l'époque si désirable où les données de l'expérience serviront de bases certaines à toutes les théories qui ont pour objet les diverses parties de l'organisation sociale. »

Dans cette Exposition de la théorie des chances, qui est un des premiers manuels de probabilité, Cournot donne des applications. Et en passant, l'air de ne pas y toucher, il reprend les données de l'article de Dupin sur le rapport des sexes.

### Exposition de la théorie des chances (1843)

Antoine Augustin Cournot (1801–1877)

De nos jours, [...] la statistique a pris un développement en quelque sorte exhubérant ; et l'on n'a plus qu'à se mettre en garde contre les applications prématurées et abusives qui pourraient la décréditer pour un temps, et retarder l'époque si désirable où les données de l'expérience serviront de bases certaines à toutes les théories qui ont pour objet les diverses parties de l'organisation sociale.

## 10 Exposition de la théorie des chances (1843)

« Nous aurons  $P = 0.1834$ ,  $\Pi = 0.5917$ , pour la probabilité que l'écart n'est pas imputable aux anomalies du hasard. Cette probabilité est sans signification possible, quoique l'année 1840 soit, après 1830 et 1828, celle des années de la période où le rapport a atteint sa plus petite valeur. »

En clair, l'affirmation péremptoire de Dupin n'est pas justifiée par le calcul de p-valeur. Cournot se garde bien d'insister. Simplement, il en est conscient, et il est l'un des rares.

Il y a eu pourtant des amateurs pour oser contredire Dupin. Dont un jeune homme, à peu près inconnu. Je n'en ai pas de portrait...

### Exposition de la théorie des chances (1843)

Antoine Augustin Cournot (1801-1877)

Nous aurons  $P = 0.1834$ ,  $\Pi = 0.5917$ , pour la probabilité que l'écart n'est pas imputable aux anomalies du hasard. Cette probabilité est [...] sans signification possible, quoique l'année 1840 soit, après 1830 et 1828, celle des années de la période où le rapport a atteint sa plus petite valeur.

## 11 Acte de naissance (7 juin 1803)

Seulement son acte de naissance en 1803 à Condrieu dans le Rhône...

### Acte de naissance (7 juin 1803)

Jules Honoré Modeste Bigeon (1803-1831)



## 12 Acte de décès (26 mai 1831)

Et son acte de décès en 1831 à Ollioules dans le Var. Il n'avait pas tout à fait 28 ans. Que s'est-il passé? On le sait par un entrefilet dans le Journal des Débats.

### Acte de décès (26 mai 1831)

Jules Honoré Modeste Bigeon (1803-1831)



## 13 Journal des Débats, 3 juin 1831

« M. Bigeon fils, rédacteur du journal l'Aviso, a été foulé aux pieds des chevaux d'une diligence, qui traversait Ollioules au grand trot, hier 26 mai, à deux heures après-midi : il a succombé une demi-heure après. »

Le destin de Bigeon, mort aussi jeune en 1831, évoque ceux de Galois, mort à 20 ans en 1832, ou d'Abel, mort à 27 ans en 1829. Les travaux de Bigeon n'ont pas eu le retentissement de ceux de Galois et Abel. Néanmoins, il avait à 28 ans une liste de publications tout à fait respectable.

### Journal des Débats, 3 juin 1831

Jules Honoré Modeste Bigeon (1803-1831)

M. Bigeon fils, rédacteur du journal l'Aviso, a été foulé aux pieds des chevaux d'une diligence, qui traversait Ollioules au grand trot, hier 26 mai, à deux heures après-midi : il a succombé une demi-heure après.

## 14 liste de publications

Il a écrit sur la théorie de l'électricité, et a correspondu avec Ampère. Il a écrit un mémoire sur les développées des courbes planes en collaboration avec Dubois-Aymé. Il a traduit un poème de Byron, et a écrit une Ode sur la guerre d'Orient. La guerre d'Orient, c'était la guerre d'indépendance des Grecs, à laquelle Byron a participé. C'était la cause qui enflammait la jeunesse romantique du temps.

Euuh c'est pas pour dire, mais l'Ode à la guerre d'Orient, il aurait peut-être mieux fait de ne pas.

## 15 Ode sur la guerre d'Orient (1829)

« Et ce n'est pas en vain que la Grèce t'implore ;  
France ! tu vas chasser de ses derniers remparts,  
Le dernier musulman dont l'aspect souille encore  
La terre des beaux arts.

[...]

Marche, peuple Français ! la gloire te convie  
À ses banquets de mort et d'immortalité :  
Vas ! qu'une fois encore à ton nom se marie  
Un cri de liberté. »

Bref, n'insistons pas.

## 16 Aperçu Statistique (1829)

Ce qui nous intéresse, c'est cet « Aperçu statistique sur la durée de la vie, dans les différentes parties de la France, et dans ses rapports avec l'instruction », etc.

Et en plus petit juste au-dessus du nom de l'auteur : « où l'on rectifie quelques fausses allégations de M. Charles Dupin ». C'est vous dire que Bigeon prenait moins de gants que Cournot.

## 17 je n'oserais conclure hardiment

« En considérant, comme l'a fait M. Charles Dupin dans ses calculs statistiques, la France du nord d'une part et celle du sud de l'autre, celle-ci comprenant la Bretagne, on trouverait que la vie moyenne est plus longue dans la première partie que dans la seconde, mais je n'oserais en conclure hardiment que c'est l'effet de l'instruction populaire plus répandue au septentrion qu'au midi attendu les exceptions trop formelles que présentent sous le rapport de la durée de la vie les départements de la Haute-Loire, du Cantal etc. marqués sur sa carte d'une teinte si noire... »

C'est vrai, Dupin ne tenait aucun compte des valeurs extrêmes.

### liste de publications

Jules Honoré Modeste Bigeon (1803-1831)

- 1 Note sur la Théorie de l'électricité (1826)
- 2 Sur les formules Barométriques et les Réfractions (1827)
- 3 Le Giaour (traduction de Byron) (1828)
- 4 Mémoire sur les développées des courbes planes (1828)
- 5 Note sur la dispersion de la lumière (1828)
- 6 Ode sur la guerre d'Orient (1829)
- 7 Aperçu statistique sur la durée de la vie (1829)
- 8 Note sur quelques expériences galvanométriques (1830)
- 9 Itinéraire du royaume d'Alger (1830)
- 10 De l'influence que peuvent avoir les Guerres... (1830)
- 11 L'influence du luxe sur la population (1831 ?)
- 12 Beppo histoire vénitienne (traduction de Byron) (1831 ?)

### Ode sur la guerre d'Orient (1829)

Jules Honoré Modeste Bigeon (1803-1831)

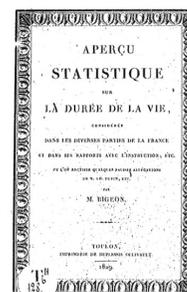
Et ce n'est pas en vain que la Grèce t'implore ;  
France ! tu vas chasser de ses derniers remparts,  
Le dernier musulman dont l'aspect souille encore  
La terre des beaux arts.

[...]

Marche, peuple Français ! la gloire te convie  
A ses banquets de mort et d'immortalité :  
Vas ! qu'une fois encore à ton nom se marie  
Un cri de liberté.

### Aperçu Statistique (1829)

Jules Honoré Modeste Bigeon (1803-1831)



### je n'oserais conclure hardiment

Bigeon, Aperçu Statistique (1829)

En considérant, comme l'a fait M. Ch. Dupin dans ses calculs statistiques, la France du nord d'une part et celle du sud de l'autre, celle-ci comprenant la Bretagne, on trouverait que la vie moyenne est plus longue dans la 1.<sup>re</sup> partie que dans la seconde, mais je n'oserais en conclure hardiment que c'est l'effet de l'instruction populaire plus répandue au septentrion qu'au midi attendu les exceptions trop formelles que présentent sous le rapport de la durée de la vie les départements de la Haute-Loire, du Cantal etc. marqués sur sa carte d'une teinte si noire...

## 18 je serais embarrassé

« ... et j'ajouterai, quoique ce ne soit pas ici le lieu de traiter cette question, au sujet de la liaison admise entre l'instruction et la richesse relatives, que le département de la Seine, malgré sa nuance obscure, est d'un si grand poids dans l'industrie de la France du nord, que je serais, je l'avoue embarrassé de rien décider sur ce point là même, si M. Dupin n'avait pas tranché la question. »

Remarquez l'ironie de la dernière phrase. Devant ce problème des valeurs extrêmes, qui rend difficile le calcul des moyennes, Bigeon a une solution originale. Il divise la série statistique en deux classes, au-dessus, et au-dessous de la moyenne.

## 19 définitions

« J'appellerai riches ou pauvres, instruits ou ignorants, etc. les départements où les revenus industriels sont plus ou moins considérables, l'instruction populaire plus ou moins répandue que dans la France moyenne. »

Ces définitions sont apparemment abruptes, mais de fait rigoureuses. Elles lui permettent de faire des constatations d'association entre variables. Parler de corrélation serait anachronique. Bigeon se contente d'observer que certaines intersections ont une taille remarquable.

## 20 intersections

« Sur les 49 départements où la vie est longue il en est : 23 instruits, 26 ignorants. Sur les 37 où elle est courte : 12 instruits, 25 ignorants. »

Comprenez que les 49 départements où la durée de vie est au-dessus de la moyenne, se répartissent à peu près également selon le niveau d'instruction. Par contre sur les 37 départements ayant une durée de vie au-dessous de la moyenne, il y en a deux fois plus où le niveau d'instruction est faible. C'est un moyen de relier durée de vie et niveau d'instruction par une sorte de méthode non-paramétrique avant la lettre.

Un autre amateur scrupuleux va pousser la même logique encore plus loin.

## 21 Château de Lompnès

Il s'agit d'Adolphe d'Angeville. Je n'ai trouvé aucun portrait de lui. Voici le château de Lompnès dans l'Ain, propriété de la famille d'Angeville. La sœur d'Adolphe, Henriette d'Angeville est beaucoup plus célèbre, et il n'y a aucun problème pour en trouver des portraits.

### je serais embarrassé

Bigeon, Aperçu Statistique (1829)

... et j'ajouterai, quoique ce ne soit pas ici le lieu de traiter cette question, au sujet de la liaison admise entre l'instruction et la richesse relatives, que le département de la Seine, malgré sa nuance obscure, est d'un si grand poids dans l'industrie de la France du nord, que je serais, je l'avoue embarrassé de rien décider sur ce point là même, si M. Dupin n'avait pas tranché la question.

### définitions

Bigeon, Aperçu Statistique (1829)

J'appellerai riches ou pauvres, instruits ou ignorants, etc. les départements où les revenus industriels sont plus ou moins considérables, l'instruction populaire plus ou moins répandue que dans la France moyenne.

### intersections

Bigeon, Aperçu Statistique (1829)

Sur les 49 départements où la vie est longue il en est : 23 instruits, 26 ignorants.

Sur les 37 où elle est courte : 12 instruits, 25 ignorants.

### Château de Lompnès

Adolphe d'Angeville (1796-1856)



## 22 ascension au Mont-Blanc (1838)

Elle est la deuxième femme à avoir atteint le sommet du Mont-Blanc, et la première à l'avoir atteint sur ses propres jambes.

## 23 ascension au Mont-Blanc (1838)

« Pendant les trois jours que j'ai encore passé à Chamonix, il n'aurait tenu qu'à moi de me croire reine ; je croyais *rêver toute éveillée* en me trouvant tout à coup célèbre pour avoir eu deux bonnes grosses jambes de montagnarde et la forte volonté de m'en servir, pour aller à quinze mille pieds de hauteur. Le rêve de Chamonix continue à Genève, *on se m'arrache* et comme je ne suis point insensible à la gloriole, tout cela me chatouille agréablement, je l'avoue, ce côté un peu faible qu'on nomme l'amour-propre. »

Une fresque sur un immeuble d'Annecy commémore son exploit. Henriette d'Angeville a fait de la montagne jusqu'à la fin de sa vie. C'était quelqu'un d'original et une forte personnalité. Tout comme son frère Adolphe.

## 24 Biographie des hommes du jour (1840)

« M. d'Angeville a ses travers, c'est un esprit fort disparate et de la famille de ceux auxquels on donne le titre *d'original*. C'est un mélange du noble et du roturier, de l'homme léger et de l'esprit appliqué ; du royaliste et du démocrate. »

Il était capable de penser sans se laisser influencer par les idées dominantes. Par exemple, en pleine période coloniale, seulement deux ans après la conquête de l'Algérie, voici ce qu'il écrit.

## 25 Aperçu sur nos colonies (1832)

« *Madagascar* n'a jamais figuré comme colonie française que sur l'*Almanach Royal* : la fièvre et les nègres ont toujours mis bon ordre à nos absurdes projets de colonisation sur ce pays. »

### ascension au Mont-Blanc (1838)

Henriette d'Angeville (1794-1871)



### ascension au Mont-Blanc (1838)

Henriette d'Angeville (1794-1871)

Pendant les trois jours que j'ai encore passé à Chamonix, il n'aurait tenu qu'à moi de me croire reine ; je croyais *rêver toute éveillée* en me trouvant tout à coup célèbre pour avoir eu deux bonnes grosses jambes de montagnarde et la forte volonté de m'en servir, pour aller à quinze mille pieds de hauteur. Le rêve de Chamonix continue à Genève, *on se m'arrache* et comme je ne suis point insensible à la gloriole, tout cela me chatouille agréablement, je l'avoue, ce côté un peu faible qu'on nomme l'amour-propre.

### Biographie des hommes du jour (1840)

Adolphe d'Angeville (1796-1856)

M. d'Angeville a ses travers, c'est un esprit fort disparate et de la famille de ceux auxquels on donne le titre *d'Original*. C'est un mélange du noble et du roturier, de l'homme léger et de l'esprit appliqué ; du royaliste et du démocrate.

### Aperçu sur nos colonies (1832)

Adolphe d'Angeville (1796-1856)

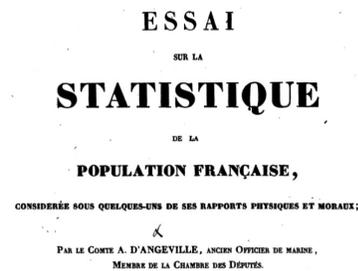
*Madagascar* n'a jamais figuré comme colonie française que sur l'*Almanach Royal* : la fièvre et les nègres ont toujours mis bon ordre à nos absurdes projets de colonisation sur ce pays.

## 26 Essai sur la Statistique (1837)

Ce qui nous intéresse ici, c'est son Essai sur la Statistique de la population française. C'est un travail énorme, précis, méticuleux. Un recueil et une compilation de données qui représentent probablement des milliers d'heures de travail. D'Angeville a étudié 127 variables sur les 86 départements. Il est probable qu'il attendait beaucoup du jugement de l'Académie. Il espérait peut-être même être nommé membre correspondant. Il ne s'attendait sûrement pas à ce qui s'est passé le lundi 5 mars 1838.

### Essai sur la Statistique (1837)

Adolphe d'Angeville (1796-1856)



## 27 Académie des sciences, lundi 5 mars 1838

« Enfin, une des principales conclusions de ses immenses recherches a été qu'un pays est d'autant plus éclairé, et que son intelligence est d'autant plus cultivée, qu'il y a plus de fenêtres aux maisons. Messieurs, qui ne s'attendaient pas à ce brusque résultat, n'ont pu réprimer les éclats d'un rire homérique; ce qui a fort scandalisé le président, M. Becquerel, car il est défendu de rire à l'Académie des sciences. »

Et si on écoutait d'Angeville, plutôt que son rapporteur ?

### Académie des sciences, lundi 5 mars 1838

Revue du Dauphiné, avril 1838

Enfin, une des principales conclusions de ses immenses recherches a été qu'un pays est d'autant plus éclairé, et que son intelligence est d'autant plus cultivée, qu'il y a plus de fenêtres aux maisons. Messieurs, qui ne s'attendaient pas à ce brusque résultat, n'ont pu réprimer les éclats d'un rire homérique; ce qui a fort scandalisé le président, M. Becquerel, car il est défendu de rire à l'académie des sciences.

## 28 première phrase

« Les études statistiques ont contre elles, outre l'aridité qui leur est propre, les préjugés soulevés par l'abus qu'on a fait des chiffres; il faut le dire, on repousse une science qui paraît prêter un égal appui aux systèmes les plus opposés. »

### première phrase

d'Angeville, Essai sur la Statistique (1838)

Les études statistiques ont contr'elles, outre l'aridité qui leur est propre, les préjugés soulevés par l'abus qu'on a fait des chiffres; il faut le dire, on repousse une science qui paraît prêter un égal appui aux systèmes les plus opposés.

## 29 mais...

« La statistique est le meilleur flambeau de la raison, lorsqu'elle est employée avec bonne foi, sans préoccupation d'aucun système, et avec des bases assez larges de *nombre* et de *durée*, pour que l'on ne coure pas la chance de présenter pour des résultats généraux des résultats partiels, incomplets, ou choisis sans discernement. »

Comment s'y est-il pris? Il invente une méthode originale, qui consiste à classer d'abord les valeurs dans l'ordre. Ce que nous appelons maintenant les statistiques d'ordre.

### mais...

d'Angeville, Essai sur la Statistique (1838)

La statistique est le meilleur flambeau de la raison, lorsqu'elle est employée avec bonne foi, sans préoccupation d'aucun système, et avec des bases assez larges de *nombre* et de *durée*, pour que l'on ne coure pas la chance de présenter pour des résultats généraux des résultats partiels, incomplets, ou choisis sans discernement.

## 30 numéro d'ordre

« Pour les calculs qui ont quelque importance, nous avons donné à tous les départements un numéro d'ordre relatif ; de cette manière, on peut voir du premier coup d'œil l'ordre des divers départements entre eux, pour les faits qui s'appliquent aux recherches les plus utiles. »

Ensuite, il divise la population en cinq.

### numéro d'ordre

d'Angeville, *Essai sur la Statistique* (1838)

Pour les calculs qui ont quelque importance, nous avons donné à tous les départements un numéro d'ordre relatif [...]; de cette manière, on peut voir du premier coup d'œil l'ordre des divers départements entr'eux, pour les faits qui s'appliquent aux recherches les plus utiles.

## 31 série de départements

« C'est ici le lieu d'expliquer ce que nous entendons par *série de départements*. Il y a 86 départements ; si on laisse en dehors des séries la *Corse*, il n'en reste que 85, dont le cinquième est 17. C'est ce dernier nombre qui compose une série. Ainsi lorsque nous parlons de la série où la population est la plus dense, nous voulons parler des 17 départements où il y a le plus de population par myriamètre carré. Cette explication était nécessaire pour l'intelligence de la suite du travail ; elle s'applique à tous nos calculs. »

Et voilà : les statistiques d'ordre de chaque variable divisent la population des 85 départements en 5 séries de 17. D'Angeville reprend le procédé de Dupin des cartes choroplèthes, mais il le rend rigoureux. Il utilise cinq couleurs, et colorie chaque département en fonction de la série à laquelle il appartient.

Surtout, il utilise les intersections de ses séries de 17 comme indicateurs d'associations entre variables. Prenons l'exemple des fenêtres qui a tant fait rire l'Académie des sciences.

### série de départements

d'Angeville, *Essai sur la Statistique* (1838)

C'est ici le lieu d'expliquer ce que nous entendons par *série de départements*. Il y a 86 départements ; si on laisse en dehors des séries la *Corse*, il n'en reste que 85, dont le cinquième est 17. C'est ce dernier nombre qui compose une série. Ainsi lorsque nous parlons de la série où la population est la plus dense, nous voulons parler des 17 départements où il y a le plus de population par myriamètre carré. Cette explication était nécessaire pour l'intelligence de la suite du travail ; elle s'applique à tous nos calculs.

## 32 fenêtres

« Sur les 17 départements de la série où il y a le plus d'ignorance, 12 sont aussi de celle où il y a le moins d'ouvertures imposables aux maisons. »

Que veut-il dire ? Dans la variable « instruction publique », il prend les 17 départements qui ont les valeurs les plus basses. Dans la variable « ouvertures » (les fenêtres étant utilisées comme base de calcul pour l'impôt foncier), il prend encore les 17 départements de plus faible valeur. Ces deux séries de 17 ont une intersection de 12. D'Angeville a l'intuition qu'une intersection de 12 entre deux sous-ensembles de 17 parmi 85, c'est significativement élevé.

Effectivement, cela s'appelle un test de Fisher exact, et on calcule la p-valeur avec la loi hypergéométrique. Sauf que le test de Fisher exact n'a été formalisé qu'un siècle après d'Angeville.

### fenêtres

d'Angeville, *Essai sur la Statistique* (1838)

Sur les 17 départements de la série où il y a le plus d'ignorance, 12 sont aussi de celle où il y a le moins d'ouvertures imposables aux maisons.

### 33 fenêtres

R dit que la p-valeur est de 2 dix moins sept, donc d'Angeville a raison : son observation est bien significative.

Et il conclut : « c'est-à-dire que toutes les fois qu'en traversant un pays on voit des maisons bien aérées, ayant beaucoup de portes et de fenêtres, on peut en conclure que l'instruction est répandue et que la civilisation est avancée dans ce pays. »

Il n'a pas dit qu'un département est d'autant plus éclairé qu'il y a plus de fenêtres. Il sait simplement que les ouvertures des maisons, en tant que base d'imposition sont un bon indicateur économique. Donc plus un département est riche, plus l'instruction publique est répandue. C'est autrement plus rigoureux que la ligne Saint-Malo – Genève de Dupin.

Voici une autre application de la méthode d'Angeville.

### 34 goîtres

« Si l'on examine tous les départements baignés par la mer, on voit que pas un n'est de la série des 17 départements où il y a le plus de goîtres. »

### 35 goîtres

Effectivement, la p-valeur du test de Fisher exact est de  $4 \cdot 10^{-3}$ .

« Y aurait-il dans le voisinage de la mer quelque chose qui s'oppose à ce mal ? »

Ça, c'est quand même impressionnant, parce que la découverte de l'effet bénéfique de l'iode par Coindet en Suisse, était toute récente, et d'Angeville l'ignorait.

### 36 étonnement

« Quel n'a pas été notre étonnement, lorsque nous avons vu que les 32 départements de la France du Nord, qui sont si éclairés, contiennent 13 des 17 départements de la série qui présente le plus d'accusés de crimes, tandis que le midi, c'est-à-dire 53 départements, n'en renferme que 4. »

#### fenêtres

d'Angeville, Essai sur la Statistique (1838)

Sur les 17 départements de la série où il y a le plus d'ignorance, 12 sont aussi de celle où il y a le moins d'ouvertures imposables aux maisons.

```
> phyper(11,17,85-17,17,lower.tail=FALSE)
[1] 2.081593e-07
```

c'est-à-dire que toutes les fois qu'en traversant un pays on voit des maisons bien aérées, ayant beaucoup de portes et de fenêtres, on peut en conclure que l'instruction est répandue et que la civilisation est avancée dans ce pays.

#### goîtres

d'Angeville, Essai sur la Statistique (1838)

Si l'on examine tous les départements baignés par la mer, on voit que pas un n'est de la série des 17 départements où il y a le plus de goîtres.

#### goîtres

d'Angeville, Essai sur la Statistique (1838)

Si l'on examine tous les départements baignés par la mer, on voit que pas un n'est de la série des 17 départements où il y a le plus de goîtres.

```
> phyper(0,21,85-21,17)
[1] 0.00431921
```

Y aurait-il dans le voisinage de la mer quelque chose qui s'oppose à ce mal[...]?

#### étonnement

d'Angeville, Essai sur la Statistique (1838)

Quel n'a pas été notre étonnement, lorsque nous avons vu que les 32 départements de la France du Nord, qui sont si éclairés, contiennent 13 des 17 départements de la série qui présente le plus d'accusés de crimes, tandis que le midi, c'est-à-dire 53 départements, n'en renferme que 4.

## 37 étonnement

Oui, la p-valeur est faible. Il a raison de s'étonner.

C'est vrai, les raccourcis auxquels d'Angeville se laisse aller sont parfois un peu loufoques.

### étonnement

d'Angeville, *Essai sur la Statistique* (1838)

Quel n'a pas été notre étonnement, lorsque nous avons vu que les 32 départements de la France du Nord, qui sont si éclairés, contiennent 13 des 17 départements de la série qui présente le plus d'accusés de crimes, tandis que le midi, c'est-à-dire 53 départements, n'en renferme que 4.

```
> phyper(12,17,85-17,32,lower.tail=FALSE)
[1] 0.0003556381
```

## 38 châtaignes

« Les départements de la *Haute-Vienne*, de la *Corrèze*, de la *Charente*, de la *Dordogne* et du *Lot*, qui sont contigus, semblent être l'agglomération où les tailles sont le plus généralement faibles ; c'est aussi la partie de la France où l'on consomme le plus de *châtaignes*. »

D'autres conclusions, qui sont pourtant bien dans l'air du temps, nous paraissent beaucoup moins drôles.

### châtaignes

d'Angeville, *Essai sur la Statistique* (1838)

Les départements de la *Haute-Vienne*, de la *Corrèze*, de la *Charente*, de la *Dordogne* et du *Lot*, qui sont contigus, semblent être l'agglomération où les tailles sont le plus généralement faibles ; c'est aussi la partie de la France où l'on consomme le plus de *châtaignes*.

## 39 bâtards

« Dès lors, nous avons dû considérer, non comme une base absolue de moralité, mais comme l'un de ses éléments les plus directs, le nombre relatif des bâtards de chaque département. Les calculs de la 19<sup>e</sup> colonne prouvent que les départements les plus industriels sont, en général, ceux où les mœurs sont le plus relâchées. »

### bâtards

d'Angeville, *Essai sur la Statistique* (1838)

Dès lors, nous avons dû considérer, non comme une base absolue de moralité, mais comme l'un de ses éléments les plus directs, le nombre relatif des bâtards de chaque département. Les calculs de la 19<sup>e</sup> colonne prouvent que les départements les plus industriels sont, en général, ceux où les mœurs sont le plus relâchées.

## 40 hospices

« On doit comprendre, en effet, que les 12 millions que l'état ou les communes payent annuellement aux hospices d'enfants trouvés, agissent comme une prime pour l'encouragement de la production de ces infortunés ; et c'est ainsi que l'administration et les chambres poussent au vice les classes pauvres, et se rendent complices de la plus grande cause de démoralisation qui existe dans la nation. »

### hospices

d'Angeville, *Essai sur la Statistique* (1838)

On doit comprendre, en effet, que les 12 millions que l'état ou les communes payent annuellement aux hospices d'enfants trouvés, agissent comme une prime pour l'encouragement de la production de ces infortunés ; et c'est ainsi que l'administration et les chambres poussent au vice les classes pauvres, et se rendent complices de la plus grande cause de démoralisation qui existe dans la nation.

Écoutez, je ne suis pas bien sûr, mais j'ai l'impression que plus il y a de vanes dans une histoire, plus elle est drôle. Vous croyez qu'on peut essayer de le démontrer ?

### références

- P. Beyls (2007) *Dubois Aymé, égyptologue, directeur des douanes, député*, Montbonnot : l'auteur et communication privée
- J. H. M. Bigeon (1829) *Aperçu statistique sur la durée de la vie*, Toulon : Duplessis Ollivault
- A. A. Cournot (1843) *Exposition de la théorie des chances et des probabilités*, Paris : Hachette
- A. d'Angeville (1837) *Essai sur la Statistique de la population française*, Bourg : Dufour
- O. Rey (2016) *Quand le monde s'est fait nombre*, Paris : Stock
- B. Y. (2016) 1827 : la mode de la statistique en France, *Histoire et Mesure*, 31(1) 161–194