

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

LUCIEN MARCH

Les représentations graphiques et la statistique comparative

Journal de la société statistique de Paris, tome 45 (1904), p. 407-420

http://www.numdam.org/item?id=JSFS_1904__45__407_0

© Société de statistique de Paris, 1904, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

III.

LES REPRÉSENTATIONS GRAPHIQUES ET LA STATISTIQUE COMPARATIVE.

On a souvent recours aux représentations graphiques pour guider l'esprit, soit dans l'exercice des arts, soit dans les études scientifiques (1). Elles prêtent à la statistique un appui dont l'utilité n'a point échappé aux premiers chercheurs : l'atlas de Playfair date de 1786.

De nos jours, lorsque les ressources financières le permettent, on accompagne les tableaux arides de chiffres, de cartogrammes et de diagrammes qui éclairent l'exposé ou l'analyse des faits numériques, comme l'image colore le récit ou appuie la théorie, et que le lecteur accueille toujours avec satisfaction, comme l'on fête le rayon de soleil qui, à travers une brume légère, fait saillir les contours des choses.

Parmi les publications de statistique graphique parues en France depuis une cinquantaine d'années, il me suffira de rappeler les *Atlas de statistique agricole*, les *Albums* du Ministère des travaux publics, ceux qu'ont dressés l'administration des contributions directes, le bureau de la Statistique générale de France, l'Office du travail, la ville de Paris, une place à part devant être réservée aux œuvres privées si méritantes de Guerry (2), d'Adolphe Bertillon et d'autres auteurs.

Au sein de notre Société, MM. Levasseur, Cheysson, de Foville, Jacques Bertillon, Loua, Turquan, ont exposé des procédés propres à enrichir les ressources de la statistique graphique ou signalé des précautions utiles et, parmi ceux que la mort nous a élevés, M. Vauthier a tous les droits à ne point être oublié.

Mais personne ne sera surpris — et M. Levasseur permettra certainement — qu'entre ces noms je reprenne celui de l'auteur éminent qui, dans la préparation des plus importantes publications officielles citées plus haut, a montré que l'ingé-

(1) « Il faut employer toutes les ressources de l'intelligence, de l'imagination, *des sens* et de la mémoire, soit pour avoir une intuition distincte des propositions simples, soit pour comparer convenablement ce qu'on cherche avec ce qu'on connaît, afin de le découvrir par ce moyen. » (DESCARTES, *Règles pour la direction de l'esprit*.)

(2) Dans son atlas relatif à la statistique morale de l'Angleterre, publiée à Paris en 1863, Guerry a appliqué, sans les formuler, une partie des règles dont il sera question plus loin.

niosité dans le choix des moyens pouvait s'allier à une parfaite sûreté de méthode, et qui nous a fait ensuite profiter de son expérience dans quelques-unes de ces communications que nous accueillons aussi comme le rayon de soleil dont je parlais tout à l'heure.

C'est sous l'autorité de ce maître que je placerai les réflexions qui vont suivre ; ce sont en partie ses travaux qui me les ont inspirées, et je rappellerai tout à l'heure ce que nous devons à M. Cheysson sur un point essentiel de la statistique graphique comparative.

Le rôle des représentations graphiques en statistique peut être envisagé sous deux aspects.

Du point de vue de la statistique descriptive, elles présentent à l'œil, sous une forme saisissante, le tableau schématique des faits susceptibles de mesure. Du point de vue de la statistique comparative, ou analytique, elles mettent en évidence les relations réciproques des faits et elles peuvent aider à découvrir ce qu'il y a de constant dans ces relations. Elles facilitent aussi la comparaison des différentes phases d'un même phénomène, décèlent les irrégularités et les anomalies et fournissent des moyens, soit de rectifier les données, soit de suppléer à leur absence.

Je laisserai de côté les applications à la statistique descriptive : je ne pourrais que répéter ce que d'autres ont dit mieux que je ne saurais le dire. Un mot seulement. On a insisté sur la qualité maîtresse qu'il convient d'exiger des représentations graphiques descriptives : la clarté ; mon sentiment est que l'on doit en effet tout sacrifier à cette exigence. Ne pas chercher à représenter trop de choses à la fois, user des moyens les plus simples, subordonner le choix de ces moyens de représentation plutôt aux facultés de l'œil qu'aux facultés de l'esprit : telles me semblent les règles acceptées le plus généralement et sans le respect desquelles les représentations descriptives peuvent perdre toute leur utilité.

Je m'occuperai exclusivement de l'application des représentations graphiques à la statistique comparative.

Ici encore la clarté est de rigueur. Il ne s'agit plus, il est vrai, de vulgarisation ou d'enseignement, mais de démonstrations et d'investigations. Cependant il ne faut point oublier que d'autres moyens s'offrent à nous pour la recherche ou la mise en évidence des rapports et que, si nous avons recours à la méthode graphique, c'est pour voir plus clair. La simplicité des moyens de représentation s'impose donc encore sur le terrain de la statistique analytique, et l'on doit la recommander, sans d'ailleurs en faire une obligation, chacun, dans le domaine de la recherche scientifique, restant libre de son inspiration.

Mais il est une partie du terrain où des barrières s'imposent, qu'un règlement inflexible doit empêcher de franchir : c'est celui des fausses impressions que peut donner un graphique mal combiné, des relations erronées qu'il laisse supposer.

Comment se limite le territoire interdit, quel est le domaine exploitable et comment l'exploiter ? Je voudrais vous soumettre à ce propos quelques réflexions et quelques applications.

Les représentations graphiques les plus aisément utilisables, et peut-être même les seules vraiment utiles, dans le domaine de la statistique comparative, sont les représentations dans un plan, cartogrammes ou diagrammes.

Les cartogrammes fournissent des représentations topographiques des faits, c'est-à-dire qu'ils signalent les relations de faits mesurables avec les différents points ou avec les différentes portions de surface du plan. Comme, dans le plan, la connaissance de deux éléments est nécessaire pour situer un point, on peut dire que le cartogramme est un moyen de représenter la relation d'un élément observé avec deux autres éléments. Les diagrammes fournissent une représentation des relations de deux éléments seulement.

I. — LES CARTOGRAMMES.

En général, les parties du plan d'un cartogramme correspondent à des divisions territoriales et l'on représente la grandeur de l'élément observé dans chaque compartiment, soit par une surface dont l'étendue varie avec la grandeur de l'élément, soit par une teinte unique d'intensité variable ou, comme l'on dit, plus ou moins dégradée, soit par des teintes diverses qui changent de couleur suivant la grandeur de l'élément.

Ce dernier procédé semble devoir être rejeté du domaine de la statistique comparative, à cause de l'effort de mémoire qu'exige l'emploi de plusieurs couleurs et aussi parce que la juxtaposition de couleurs différentes ne donne pas l'impression de plus et de moins qui convient aux mesures comparatives; au contraire, elle laisse une impression de différenciation d'espèces qui n'est généralement pas dans la nature des choses.

Les cartogrammes à teintes dégradées n'offrent pas ces inconvénients, mais, pour qu'ils puissent être correctement appliqués à la statistique comparative, il convient de les établir avec méthode. Je rappellerai brièvement en quoi consiste la méthode exposée en 1887, à notre Société, par M. Cheysson (1) et j'en montrerai quelques applications.

Cartogrammes à teintes dégradées.

Proposons-nous par exemple de représenter la distribution en France du personnel des industries de l'alimentation, d'après les résultats du recensement de 1896, et admettons que nous disposions de 8 nuances différentes de la même couleur. On peut s'y prendre de plusieurs façons.

On peut se servir des nombres absolus des personnes engagées dans les industries considérées, et partager les 87 départements en 8 groupes suivant le nombre de personnes que chacun comprend. La carte ainsi dressée permettra d'apprécier d'un coup d'œil la répartition géographique des industries étudiées; on se rendra compte par exemple que le département de la Seine comprend à peu près le dixième du personnel total, tandis que le Territoire de Belfort n'en comprend guère plus de 1 millième. Ce sera une bonne carte industrielle décrivant la situation géographique des industries considérées, mais sans intérêt comparatif. En effet, nous n'avons pas

(1) *Journal de la Société de statistique de Paris*, numéro d'avril 1887.

besoin de cette carte pour savoir que dans les départements très peuplés, les bouches à nourrir étant plus nombreuses et la main-d'œuvre plus abondante que dans les autres, les industries de l'alimentation prennent un grand développement.

Il est plus intéressant de choisir comme élément à comparer, non pas le nombre absolu des personnes actives de chaque département, mais des nombres proportionnels, par exemple le nombre des personnes actives pour 10 000 habitants. Alors, les départements fortement teintés sont ceux qui, à population égale, comprennent le plus grand nombre d'individus occupés par les industries visées. Ainsi, le département où le personnel des industries de l'alimentation est relativement le plus nombreux est celui de la Marne et non celui de la Seine.

Je me reprocherais d'insister ici sur des recommandations aussi banales; cependant, du point de vue où nous sommes placés, il n'est pas inutile de formuler d'une manière précise la raison principale de préférer la seconde carte à la première. Sur la première carte, consacrée aux nombres absolus, nous observions un phénomène complexe : les variations d'un département à l'autre dépendaient à la fois de la consistance des industries considérées et du nombre des habitants. La seconde présente un phénomène plus simple : les changements ne dépendent plus du nombre des habitants; cette influence a été, comme l'on dit, éliminée, et l'étude comparative peut porter exclusivement sur les autres influences. Rien n'empêche, d'ailleurs, pour que le cartogramme fasse connaître aussi les nombres absolus, d'inscrire dans le cercle de chaque département le nombre absolu correspondant au lieu du nombre relatif, tout en graduant les teintes dans l'ordre des nombres relatifs.

Étant admis que les départements sont classés dans l'ordre des nombres relatifs, c'est-à-dire d'après des coefficients, on peut encore hésiter quant au mode de graduation des teintes et au mode de distribution des départements entre les diverses teintes.

On peut ordonner les départements suivant la valeur du coefficient, partager l'échelle du coefficient en 8 parties, si l'on dispose de 8 teintes, et attribuer une nuance à chacune de ces parties. Parfois, on corrige ce qu'une telle division a de trop rigoureux en tenant compte aussi de la discontinuité des coefficients.

Admettons qu'on applique rigoureusement la méthode. Dans le cas que nous avons pris comme exemple, où il s'agit de représenter la distribution des industries de l'alimentation, la série des coefficients est comprise entre la proportion de 60 personnes actives pour 10 000 habitants et celle de 220 pour 10 000; pour 8 divisions de l'intervalle, chacune correspondra à 20 unités; la teinte la plus faible sera appliquée aux départements pour lesquels le coefficient est compris entre 60 et 80, la suivante de 80 à 100 et ainsi de suite.

On établirait de même des cartes relatives à d'autres groupes industriels. Par exemple, dressons la carte du personnel des industries textiles proprement dites. Dans ce cas, l'échelle des coefficients-nombres de personnes actives pour 10 000 habitants, est beaucoup plus étendue. Tandis que pour les industries de l'alimentation, les coefficients varient de 60 à 220, dans le cas des industries textiles ils oscillent de 0 à 560 personnes actives par 10 000 habitants. En divisant cet intervalle en 8 parties, chaque division correspondra à 70 personnes pour 10 000 habitants, tandis que chaque division de l'échelle des industries de l'alimentation ne correspondait qu'à 20 personnes.

M. Cheysson a fait remarquer que l'emploi de la même gamme de teintes, dans toute son étendue, pour représenter des variations aussi différentes, présentait une image déformée de la réalité. Examinée isolément, chacune des deux cartes fournit un aperçu satisfaisant de la répartition dans l'un ou l'autre groupe d'industries. Mais, lorsqu'on les juxtapose, on est porté à attribuer aux mêmes teintes les mêmes valeurs, ce qui n'est point conforme à la réalité.

On pourrait, il est vrai, augmenter le nombre des teintes lorsqu'il s'agit des industries textiles, en conservant aux teintes employées pour les industries de l'alimentation la valeur qui leur avait été assignée primitivement. Le procédé serait impraticable. Si l'on emploie 8 teintes pour les industries de l'alimentation, il en faudrait 28 pour les industries textiles; se bornât-on à 4 dans le premier cas, qu'il en faudrait encore 14 dans le second. L'œil ne percevrait pas les différences imperceptibles des nuances.

Si pourtant on avait recours à une gamme très adoucie et si l'on supposait un œil très sensible, on se heurterait par ailleurs à une difficulté plus grave.

Dans l'exemple précédent, il s'agit de comparer des coefficients de même espèce; il peut arriver que les faits à comparer soient de nature différente. Alors une même teinte représentera sur les deux cartes des nombres qui pourront être égaux, mais qui ne comprendront pas des unités de même nature; par suite, la comparaison supposera une convention implicite quant au rapport des deux unités et elle n'aura aucune signification si la convention est arbitraire.

Supposons, par exemple, que l'on se propose de mettre en présence une carte de la natalité par département en France et une carte représentant les variations des salaires d'un département à l'autre, afin d'apprécier s'il semble exister un certain rapport entre le taux de la natalité et le taux des salaires.

A la suite des enquêtes de l'Office du travail de 1891 à 1896, on a pu classer les départements d'après l'évaluation du taux ordinaire des salaires dans les villes chefs-lieux de département. Les nombres varient de 2 fr. 50 c. dans le département de la Lozère à 6 fr. 50 c. dans le département de la Seine, pour une journée de travail.

D'autre part, durant la période décennale 1888-1897, le nombre des naissances pour 1 000 habitants dans les divers départements est compris entre 14 dans le Gers et 32 dans le Finistère.

Si l'on dispose de 9 teintes et que l'on prenne pour unités le franc, en ce qui concerne les salaires, une naissance par mille habitants en ce qui concerne les naissances, chaque teinte pourra correspondre à deux unités, c'est-à-dire à deux naissances pour 1 000 habitants, de façon à utiliser toute la gamme lorsqu'on établit la carte des naissances. Chaque teinte correspondant à deux unités sur la carte des salaires, celle-ci n'utilisera que trois teintes. De la comparaison des deux cartes résultera donc cette impression que le salaire varie fort peu d'un département à l'autre, tandis que la natalité varie beaucoup. Il suffirait de prendre pour unité des salaires le centime au lieu du franc pour que la comparaison des deux cartes laissât exactement l'impression contraire. De même si, conservant le franc comme unité du salaire, on adoptait comme unité de natalité une naissance pour 100 habitants, au lieu d'une naissance pour 1 000 habitants.

Ainsi, lorsque nous interrogeons les cartogrammes pour nous faire une idée des

rapports des faits représentés, la réponse dépend avant tout du choix des unités. Par conséquent, pour que les cartogrammes aient une valeur comparative, ce choix ne doit point être arbitraire.

Le principal intérêt de la méthode exposée par M. Cheysson est, semble-t-il, que le choix des unités se trouve réglé d'après une convention uniforme, indépendante de la nature des faits.

Pour qu'il en soit ainsi, quelle que soit la série de faits à représenter, l'unité doit être une même fonction des nombres de la série. Cette fonction sera de préférence très simple : ce pourrait être la valeur médiane des termes de la série (valeur qui laisse de chaque côté un nombre égal d'observations), ou la valeur normale (point de concentration des observations) ou l'une des valeurs moyennes.

M. Cheysson a proposé de prendre pour unité la moyenne arithmétique, calculée en ayant égard au poids des observations, c'est-à-dire ce que l'on pourrait appeler la moyenne arithmétique quantitative ou *barique* (1). Les nombres de la série sont rapportés à cette moyenne et l'échelle des rapports ainsi calculés est mise en correspondance avec l'échelle des teintes.

Cette méthode, dont l'application est indispensable lorsque les faits comparés sont d'espèce différente, est aussi la plus rationnelle lorsqu'il s'agit de faits de même espèce parce que, suivant l'expression de M. Cheysson, « elle conserve l'importance respective des faits, leur hiérarchie ».

Elle a été appliquée par l'Office du travail pour la préparation des cartes qui, à l'Exposition de 1900, représentaient la distribution en France du personnel des différents groupes industriels.

En comparant ces cartes, on se rend compte que, pour certains groupes tels que celui des industries de l'alimentation, celui des industries du bâtiment, etc., la distribution est presque uniforme; un petit nombre de teintes suffisent pour marquer les différences d'un département à l'autre. Au contraire, l'emploi d'un plus grand nombre de teintes pour les industries textiles, pour le groupe de la céramique et de la verrerie, etc, indique que le personnel de ces industries est concentré dans certains départements, très rare dans d'autres, et ces appréciations comparatives ont ici une signification qui fait entièrement défaut lorsque le choix des unités est arbitraire.

Je n'insiste pas sur le mode d'opérer : les détails nécessaires ont été donnés par M. Cheysson dans sa communication précitée; j'ai voulu surtout appeler de nouveau l'attention sur une méthode féconde et justifier son extension à d'autres modes de représentation.

On peut la modifier. Par exemple, au lieu de graduer en nombres absolus les écarts des nombres observés à partir de leur moyenne, on pourrait graduer ces écarts en nombres relatifs par rapport à leur propre moyenne. Quels que soient les procédés, il suffit que dans la construction des cartogrammes comparatifs des précautions soient prises pour éviter toute fausse impression quant à la répartition relative des faits comparés.

(1) En anglais : *weighted average*.

Cartogrammes à éléments représentatifs échelonnés.

Au lieu d'employer comme mode de représentation une teinte dégradée, on emploie aussi des lignes ou des surfaces plus ou moins grandes, suivant que les faits qu'elles représentent sont plus ou moins nombreux. La représentation peut d'ailleurs être rendue plus frappante si les surfaces-indices sont de plus distinguées par des teintes dégradées. Pour rendre comparables les cartogrammes de ce système il suffit, comme pour les précédents, de mesurer les faits de chaque série par rapport à la moyenne. Par exemple, il pourrait être intéressant de dresser par ce procédé un cartogramme des salaires moyens dans les villes où fonctionne une bourse de travail et de le comparer à un autre cartogramme représentant pour chacune de ces villes l'importance relative du nombre des ouvriers affiliés à la bourse de travail.

Cartogrammes tabulaires.

En général, les cartogrammes servent à représenter la répartition d'éléments numériques entre des circonscriptions géographiques. Si l'on regarde comme cartogramme tout moyen de représentation qui permet de figurer dans un plan la relation d'un fait principal avec deux éléments variables, les tables à double entrée, telles que la table des mariages suivant l'âge des époux, peuvent être traduites en cartogrammes, soit à l'aide de teintes dégradées, soit à l'aide de courbes de niveau, soit par la combinaison de ces deux systèmes, soit enfin à l'aide de lignes ou de surfaces teintées uniformément ou non et d'étendue variable suivant la grandeur des nombres à représenter.

Les trois premiers systèmes fournissent des figures intéressantes de statistique descriptive; on se rappelle celles qu'ont dressées MM. Vauthier, Perozzo, Turquan. Il ne semble pas que ces figures soient bien commodes pour les recherches de statistique comparative, à cause de leur complication et de la nécessité de leur donner du relief à l'aide d'un lavis très soigné et d'effets de perspective. De même que les solides, ou représentations à trois dimensions dont elles sont l'image dans le plan, elles n'offrent pas assez de simplicité et elles ne soulagent guère le travail de l'esprit.

Il n'en est pas de même des représentations par lignes ou surfaces d'étendue variable. Dans ce genre, je citerai comme un excellent instrument de démonstration et de recherche le système de représentation employé par M. Lexis pour l'étude de la mortalité et de la natalité, et dont ce savant professeur nous a exposé une application, lors du dernier congrès de l'Institut international de statistique.

II. — LES DIAGRAMMES.

Les diagrammes constituent un moyen d'expression graphique plus simple que les cartogrammes puisqu'ils servent à représenter sur le plan les relations de deux éléments seulement au lieu de trois; ils offrent cependant plus de ressources parce que le champ de la représentation n'est limité ni par un nombre de teintes ni par la juxtaposition de divisions topographiques.

Pour la commodité du langage, l'un des deux éléments comparés est désigné sous le nom de variable, l'autre sous le nom de variante. D'autre part, afin d'éviter toute complication, on se borne aux cas où l'un au moins des deux éléments ne prend qu'une seule valeur pour chaque valeur de l'autre et c'est ce dernier que l'on choisit comme variable.

Du point de vue de la statistique comparative, il n'est pas utile de traiter séparément les différentes formes graphiques que l'on distingue d'ordinaire : diagrammes à coordonnées rectilignes, diagrammes polaires, diagrammes à ordonnées discontinues, diagrammes à surfaces continues. Quelle que soit la forme adoptée, le tracé s'effectue ainsi : on compte dans un certain sens des grandeurs proportionnelles aux valeurs successives de la variable, puis dans un autre sens déterminé des grandeurs proportionnelles aux valeurs successives de la variante. On détermine de la sorte des points que l'on relie entre eux au moyen d'une ligne brisée à laquelle, pour simplifier le langage, on donne le nom de courbe, quoique ce ne soit jamais une véritable courbe tant que l'on représente des faits observés, l'observation humaine étant inévitablement discontinue. Dans la suite nous supposerons les courbes tracées en coordonnées rectilignes.

Pour l'étude comparative de ces courbes, il est avantageux de les diviser en deux catégories, suivant que l'on a égard à la succession des faits dans le temps, à leur chronologie, ou suivant qu'on les examine simultanément, indépendamment de l'ordre dans lequel ils ont pu se produire, comme dans le cas où l'on observe la distribution des conscrits classés suivant la taille, des ouvriers classés suivant le taux du salaire (1), etc.

Il ne semble pas qu'il y ait jamais intérêt à rapprocher des courbes de l'une et des courbes de l'autre catégorie : on peut se borner à rendre comparables les courbes de chaque catégorie séparément.

Courbes de succession.

La première catégorie de courbes statistiques se rapporte à des faits dont on étudie la succession ou les changements dans le temps. Les dispositions qui me semblent devoir être conseillées pour leur construction offrent la plus grande analogie avec celles que M. Cheysson recommandait pour les cartogrammes.

Lorsqu'il s'agit d'une courbe isolée il suffit d'éviter que le lecteur ne soit tenté, en comparant les ordonnées aux abscisses, de juger non seulement du sens des changements subis par les ordonnées, ce qui est légitime, mais encore de l'allure de ces changements, ce qui ne l'est plus.

A cet effet, M. Bertillon (2) conseille de choisir les unités de manière que l'allure générale de la courbe puisse être caractérisée par une droite inclinée à 45° sur l'axe des abscisses. Pour obtenir ce résultat, il suffit à la rigueur que l'ordonnée maximum, ou mieux la moyenne des ordonnées les plus hautes, soit égale à l'abscisse maximum.

C'est surtout lorsqu'il s'agit de comparer dans le cours du temps les mouvements

(1) Division adoptée par Marey. (*La Méthode graphique*, 2^e edit., Paris, 1885.)

(2) Rapport à la VIII^e session de l'Institut international de statistique.

de faits statistiques, soit de même nature, soit de nature différente, que les représentations graphiques sous forme de diagrammes offrent le plus d'utilité, mais aussi c'est alors que des précautions sont le plus nécessaires.

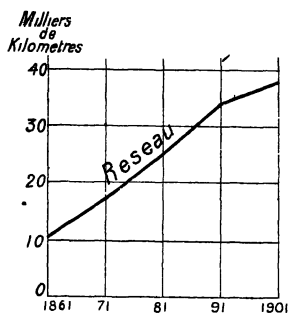


Fig. 1.

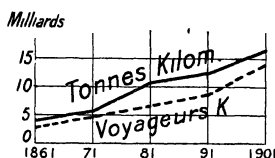


Fig. 2.

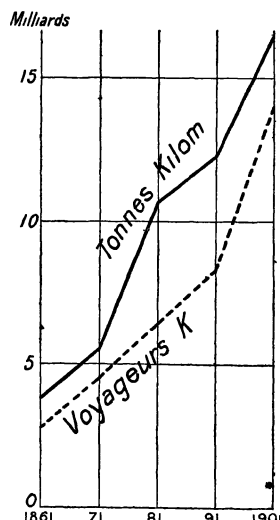


Fig. 3.

Traçons, par exemple, la courbe du développement du réseau des chemins de fer d'intérêt général depuis 1850, celle du tonnage kilométrique et celle du nombre des voyageurs kilométriques, en vue de comparer les trois phénomènes économiques.

Le schéma comparatif sera très différent suivant le choix des unités. Ainsi, sur la figure (2) le tonnage kilométrique et le nombre de voyageurs kilométriques semblent avoir crû beaucoup plus rapidement que le réseau dont le développement est représenté figure (1); sur la figure (3), au contraire, on a l'impression inverse, et ce résultat est obtenu simplement par le changement de l'unité de longueur des ordonnées : sur la figure (3) l'unité adoptée est cinq fois plus grande que sur la figure (2).

Or, la comparaison des diagrammes est un des moyens auxquels on recourt le plus souvent pour justifier les relations que l'on suppose exister entre des phénomènes soumis aux mesures statistiques. Cette comparaison est viciée si le choix des unités est arbitraire et toute conclusion est interdite.

Le moyen d'échapper à cet arbitraire est encore celui que Guerry et M. Cheysson ont appliqué aux cartogrammes, mais ici il y a quelque intérêt à modifier légèrement la formule.

Au lieu de rapporter chaque ordonnée à l'ordonnée moyenne ou à la somme des ordonnées, il est suffisant et plus rapide de les rapporter à l'une d'entre elles, par exemple à celle qui correspond à la dernière année de l'observation. Soit Ω la dernière ordonnée et M l'ordonnée moyenne, une ordonnée O quelconque mesurée par rapport à cette moyenne sera égale à $\frac{O}{M}$, fraction que l'on peut écrire $\frac{O}{\Omega} \times \frac{\Omega}{M}$. La longueur de chaque ordonnée de la courbe rapportée à Ω sera égale à l'ordonnée correspondante de la courbe rapportée à M , à un coefficient constant près.

Pratiquement, on remplacera par 100 le chiffre correspondant à l'année choisie comme année de base, les autres seront remplacés par des nombres proportionnels.

Dans un but d'uniformité, la longueur qui représente 100 aura un rapport fixe avec l'unité de temps. De cette façon, les diagrammes seront toujours comparables, quel que soit le nombre d'années auquel s'étendent les observations.

En fait, il est très rare que l'on construise des diagrammes s'étendant sur plus de cent années, car il n'est guère de matières statistiques qui prêtent à d'aussi longues comparaisons. D'autre part, des nombres annuels ne peuvent guère être utilement comparés si la période est inférieure à dix années. 10 ans et 100 ans peuvent donc être regardés comme les limites de l'étendue des diagrammes chronologiques représentant des séries de nombres annuels; la moyenne des inclinaisons des courbes qui partent de ces limites est à peu près de 45° lorsque le coefficient 100 est représenté par une longueur égale à celle qui, sur l'axe des abscisses, mesure trente années. Pour représenter 100 en ordonnée, il serait donc convenable de choisir une longueur égale à celle qui représente 30 ans sur l'axe des temps.

Avec ces dispositions les trois courbes relatives à l'exploitation des chemins de fer ont les formes indiquées sur les figures (4, 5, 6), et leur examen comparatif permet cette fois d'affirmer que le tonnage kilométrique ainsi que le nombre des voyageurs se sont accrus un peu plus vite que le réseau, de noter les périodes où l'accroissement a été plus ou moins rapide. Nous n'avons figuré que les résultats de cinq années échelonnées à 10 ans l'une de l'autre; il serait aussi facile de suivre d'après la même méthode ces résultats année par année.

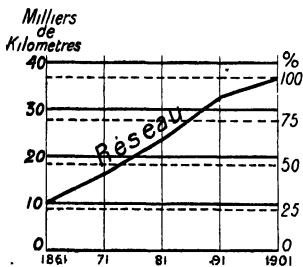


Fig. 4.

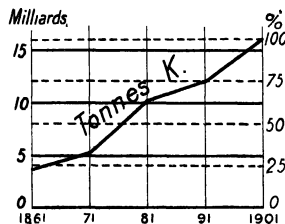


Fig. 5.

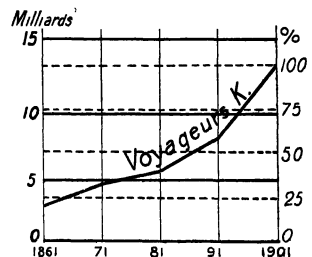


Fig. 6.

Au lieu de prendre comme terme de comparaison la dernière année de l'observation, on pourrait aussi bien prendre une année quelconque, ou mieux un groupe d'années, afin d'éviter les anomalies possibles. Si l'on ne veut faire porter le calcul sur toutes les années, il vaut mieux choisir des années récentes, d'abord parce que, pour beaucoup de statistiques, les chiffres récents sont plus sûrs et plus complets que les plus anciens, et en outre parce que dans la suite il sera facile de prolonger la courbe en conservant la même base et par conséquent la même allure.

Graduation en nombres absolus.

Telles qu'elles viennent d'être construites, les courbes statistiques ne font connaître que des nombres proportionnels; la graduation des ordonnées est établie en proportions pour cent de l'ordonnée de base, ce qui est gênant lorsque l'on veut se rendre compte des variations réelles en nombres absolus. Des calculs sont néces-

saies qu'il convient d'éviter au lecteur en ajoutant une seconde graduation en nombres absolus.

Prenons à titre d'exemple la courbe du réseau des chemins de fer d'intérêt général. La longueur de ce réseau était en 1901 de 38 000 kilomètres, en 1891 de 34 400, en 1881 de 25 200, etc. Pour graduer l'échelle en nombres proportionnels on a élevé, au point 1901, sur l'axe des abscisses, une longueur égale à celle qui mesure trois années et l'on a divisé cette longueur en quatre parties égales.

Les points de la courbe sont déterminés par les rapports $\frac{34,4}{38} \times 100$, $\frac{25,2}{38} \times 100$, etc., que l'on mesure sur l'échelle précédente; on obtient ainsi des points dont les hauteurs sont : 100 en 1901; 90,5 en 1891; 66,3 en 1881; etc. Pour construire l'échelle des nombres absolus, on remarque que la longueur 38 milliers de kilomètres correspondant à 100 unités de l'échelle proportionnelle, 10 unités de l'échelle absolue correspondent à $\frac{1\ 000}{38} = 26,4$ unités de l'échelle proportionnelle, par suite 20 à 52,7; 30 à 158, etc.

Les deux échelles sont figurées l'une à droite, l'autre à gauche de chacun des diagrammes.

On peut demander encore un peu plus à la représentation. Lorsque l'on compare des diagrammes de même espèce, il peut être intéressant de donner au lecteur l'impression de la consistance différente des phénomènes représentés.

Si, d'après le système précédemment exposé, on trace trois diagrammes représentant respectivement le mouvement des syndicats patronaux, ouvriers et mixtes depuis 1884 jusqu'en 1900, on obtient des courbes dont l'allure différente met en évidence le fait que le nombre des syndicats patronaux a augmenté assez régulièrement au cours de la période, tandis que le nombre des syndicats ouvriers et celui des syndicats mixtes, qui s'était accru très vite jusqu'en 1894, a augmenté ensuite plus lentement ou a diminué.

Si, de plus, on fait subir une réduction proportionnelle convenable aux échelles adoptées tant en ordonnées qu'en abscisses pour les syndicats patronaux et pour les syndicats mixtes, par rapport aux échelles adoptées pour les syndicats ouvriers, le diagramme des syndicats mixtes couvre alors une étendue superficielle beaucoup moindre que les deux autres et le diagramme des syndicats patronaux une superficie moindre que celui des syndicats ouvriers, ce qui rend sensible aux yeux les différences des nombres des trois catégories de syndicats, sans que l'allure respective des phénomènes se trouve modifiée.

Grâce à de semblables dispositions, les diagrammes de succession sont rendus comparables et donnent au lecteur une juste impression, tant de la grandeur relative que de la grandeur absolue des éléments en présence. A défaut de règles de construction uniformes, toute comparaison des courbes de succession est dépourvue de sens.

Courbes de distribution.

La seconde catégorie de courbes statistiques est celle des courbes de distribution. Trois cas peuvent être distingués. Si l'on trace une courbe isolée, on évitera de donner une fausse impression de la distribution des faits. Comme les seules lon-

guez que l'œil soit porté à comparer sont les abscisses et les ordonnées et comme d'autre part leur grandeur est pratiquement limitée, il suffit encore de choisir l'unité d'abscisse et l'unité d'ordonnée de façon que l'étendue de la figure en largeur (abscisse maxima) soit égale à son étendue en hauteur (ordonnée maxima). Ainsi le lecteur ne sera pas tenté d'apprécier la dispersion plus ou moins grande des faits d'après le rapport des ordonnées aux abscisses, ce qui constitue un jugement *instinctif* mais *faux*.

Le second cas est celui où l'on trace une série de courbes de distribution ayant rapport à des faits statistiques de même nature. Prenons par exemple des courbes de distribution de salaires, en vue de nous rendre compte si, à telle époque, ou dans tel pays, les ouvriers les moins payés sont relativement moins nombreux qu'à telle autre époque ou dans tel autre pays.

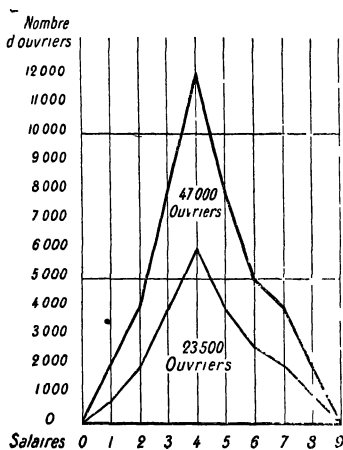


Fig. 7.

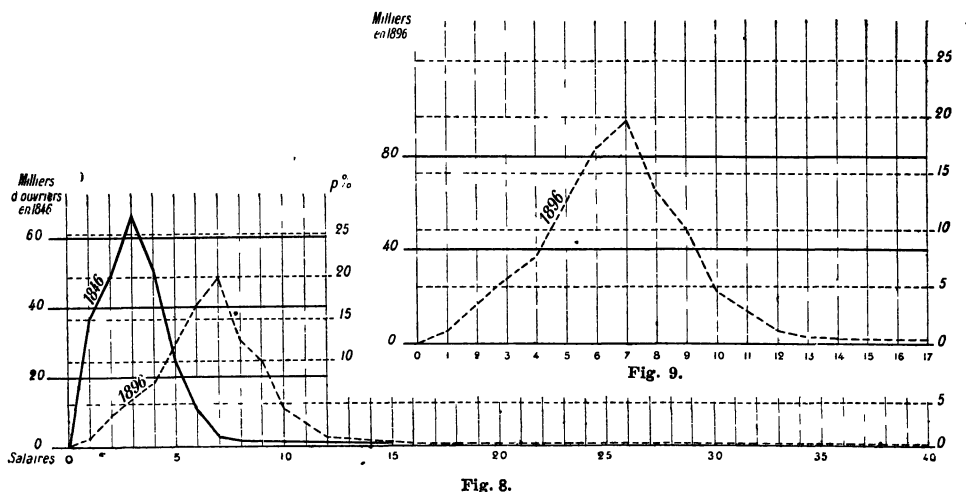
Supposons deux groupes ouvriers dont les salaires soient répartis de la même façon, mais dont l'un comprend deux fois plus d'ouvriers que l'autre. Si l'on trace pour chacun de ces groupes le diagramme représentatif de la distribution des salaires, en portant sur un axe horizontal des longueurs proportionnelles aux différents taux de salaires, puis en élevant en chaque point ainsi déterminé une perpendiculaire à l'axe horizontal dont la longueur soit proportionnelle au nombre des ouvriers payés à ce taux, les deux courbes donneront l'impression que les ouvriers du groupe le plus nombreux sont plus fortement groupés autour du salaire normal que les ouvriers du groupe le moins nombreux alors que, par hypothèse, la répartition est la même dans les deux groupes.

Pour éviter cette fausse apparence, il convient de porter en ordonnées non plus les nombres absolus d'ouvriers payés à chaque taux de salaire, mais des nombres proportionnels obtenus en rapportant chaque nombre partiel absolu au nombre total des ouvriers.

Voici par exemple (voir p. 419) tracées par ce procédé, les courbes de distribution des salaires des ouvriers industriels de Belgique en 1846 et en 1896. On voit combien en cinquante ans la courbe s'est aplatie, montrant une distribution toute différente en 1896 de ce qu'elle était en 1846, tous les ouvriers n'ayant pas bénéficié au même degré de l'amélioration générale, résultat qui n'aurait pas été mis en évidence si l'on avait comparé les courbes tracées à l'aide des nombres absolus.

On objectera sans doute que si le procédé est avantageux pour les comparaisons, il offre l'inconvénient de ne pas fournir une représentation directe des faits observés. Les ordonnées sont graduées en rapports et non plus en nombres absolus; pour retrouver les nombres absolus il faut se livrer à des calculs, au lieu de se borner à une simple lecture. Il est facile d'éviter cette objection; à cet effet, il suffit de munir le diagramme d'une double graduation des ordonnées, l'une en nombres proportionnels, l'autre en nombres absolus, ainsi qu'on l'a figuré sur le croquis ci-contre.

Si même on voulait que la simple inspection du graphique rendît compte de l'accroissement de la population ouvrière, on y parviendrait en modifiant l'échelle des abscisses et l'échelle des ordonnées de l'une des deux courbes.



NOTA. — Sur l'axe des salaires, une division représente 0 fr. 50.

Soient, pour l'année 1846, a l'unité d'abscisse correspondant à 1 fr. de salaire et b l'unité d'ordonnée correspondant à une variation égale à 10 pour cent du nombre total des ouvriers. Sur la figure (8) les deux unités a et b ont été conservées les mêmes pour le tracé de la courbe de 1896, en sorte que le nombre des ouvriers est supposé le même aux deux époques. Si le graphique doit laisser l'impression que le nombre des ouvriers est deux fois et quart plus élevé en 1896 qu'en 1846, il suffit de prendre en 1896 pour unité des abscisses une longueur égale à $a \times \sqrt{2,25}$ et pour unité des ordonnées une longueur égale à $b \times \sqrt{2,25}$ (fig. 9). Avec ces unités, on obtient une représentation graphique qui, d'une part, donne une impression exacte du changement survenu dans la répartition relative des faits et qui, en même temps, laisse aussi une impression exacte des variations absolues qu'ils ont subies. Elle permet aussi, grâce à la double échelle, de mesurer directement aussi bien les changements absolus que les changements relatifs.

Passons au troisième cas; les faits à comparer se rapportent à des faits de nature différente. Supposons par exemple que l'on juxtapose deux courbes dont l'une représente la distribution d'un groupe d'ouvriers colineurs suivant le taux du salaire, et l'autre la distribution de ces ouvriers suivant la taille, en vue d'examiner si la distribution de ces ouvriers suivant le taux du salaire est ou non semblable à la distribution suivant la taille et de se rendre compte si le travail des ouvriers les plus grands est estimé à un plus haut prix que celui des autres.

On mesurera encore les ordonnées des deux courbes à l'aide d'une même unité supposée représenter le nombre total des ouvriers ou une même fraction de ce total. Il s'agit de savoir maintenant comment s'échelonnent sur les axes des abscisses les taux successifs de salaires : 1 fr., 2 fr., 3 fr., etc., ou les tailles successives : 1^m,60, 1^m,65, 1^m,70, etc. Entre ces deux séries il n'existe pas de commune mesure et

pourtant il faut s'efforcer d'éviter que le choix des ordonnées soit arbitraire. On adoptera donc une convention, semblable à celle qui a été adoptée pour les cartogrammes. Les abscisses, soit de la courbe des salaires, soit de la courbe des tailles, seront mesurées à l'aide d'une même unité et cette unité représentera une même fonction, soit des salaires, soit des tailles. Cette fonction sera la plus simple possible : ce sera par exemple la moyenne arithmétique tenant compte du poids des observations, que nous avons désignée sous le nom de moyenne barique. L'unité d'abscisse de la courbe des salaires sera le salaire moyen ; pour la courbe des tailles, ce sera la taille moyenne.

Cette convention est d'une application très commode ; aussi commode est une autre convention dont on a fait souvent usage, depuis que M. F. Galton l'a proposée, et qui consiste à prendre, pour origine des abscisses des courbes de fréquence considérées, la valeur médiane, c'est-à-dire une grandeur telle qu'à l'ensemble des abscisses plus petites correspond une moitié des observations, tandis que l'autre moitié correspond à l'ensemble des abscisses plus grandes. Puis l'axe des abscisses est gradué de chaque côté de la nouvelle origine en divisions telles qu'à chaque division correspond une même fraction du nombre total des faits observés. Dans cette méthode les divisions de l'axe des abscisses sont inégales, mais l'on comprend que ces divisions étant déterminées uniquement par les nombres partiels d'observations, peu importe la nature des objets observés : les courbes sont toujours comparables.

Il en est de même si l'on prend pour nouvelle origine la valeur normale, c'est-à-dire celle qui correspond au maximum de densité des observations, et que l'on gradue l'axe des abscisses en prenant pour unité, soit la moyenne des valeurs absolues des écarts observés à partir de la nouvelle origine, soit la racine carrée de la somme des carrés de ces écarts, soit toute autre fonction.

Ces moyens de rendre comparables les courbes de fréquence ont principalement de l'intérêt quand l'une des courbes est celle qui caractérise la distribution de faits purement accidentels, parce que la comparaison permet de sérier les autres faits observés suivant qu'ils offrent ou non un caractère accidentel.

* * *

En résumé, les représentations graphiques des phénomènes dont s'occupe la statistique peuvent être rendues comparables, à la fois quant aux grandeurs relatives et quant aux grandeurs absolues des manifestations de ces phénomènes.

Lorsque les précautions indispensables pour assurer cette comparabilité ont été prises, mais seulement dans ce cas, il est permis de se servir des représentations graphiques soit pour découvrir les relations des phénomènes entre eux ou les liens qui peuvent exister entre les modalités successives d'un même phénomène, soit pour vérifier les rapports dont on suppose *a priori* l'existence.

Bien rarement la comparaison des graphiques démontre des rapports évidents et inflexibles ; presque toujours elle indique des liaisons soumises à des influences diverses. Il y a des degrés dans l'étroitesse de ces liaisons, dans la dépendance réciproque des phénomènes, et ces degrés, il serait utile de les mesurer : nous examinerons ultérieurement les moyens d'y parvenir.

LUCIEN MARCH.