Chapter 1

Some Prehistory of CARME: Visual Language and Visual Thinking

Michael Friendly and Matthew Sigal

> *If statistical graphics, although born just yesterday, extends its reach every day, it is because it replaces long tables of numbers and it allows one not only to embrace at glance the series of phenomena, but also to signal the correspondences or anomalies, to find the causes, to identify the laws.*
> —Émile Cheysson, c. 1877

Correspondence Analysis and Related Methods (CARME), as described in the preface, includes simple and multiple correspondence analysis (CA and MCA), biplots, singular value decomposition (SVD) and principal components analysis (PCA), canonical correlation analysis (CCA), multidimensional scaling (MDS) and so forth. The commonalities shared by these methods can be grouped in relation to the features of hypothesized lateralized brain functions. The left brain elements are more logical, formal, and mathematical: matrix expression, eigenvalue formulations, dimension reduction, while the right brain features are more visual: (point) clouds, spatial data maps, geometric vectors, and a geometric approach to data analysis. This lateralization of brain function is often exaggerated in popular culture, but it resembles a conjecture I have long held regarding data analysis (see Friendly and Kwan, 2011):

Conjecture (Bicameral minds)**.** *There are two kinds of people in this world– graph people and table people.*

The term *bicameral mind* comes from Julian Jaynes' (1978) book on the "origin of consciousness," in which he argued that ancient peoples before roughly 1000BC lacked self-reflection or meta-consciousness. For bicameral humans, direct sensory neural activity in the dominant left hemisphere operated largely by means of automatic, nonconscious habit-schemas, and was separated from input of the right hemisphere, interpreted as a vision or the voice of a chieftain or deity.

We don't fully believe the strong, two-point, discrete distributional form of the above conjecture; rather, a weaker claim for bimodality or clearly separated latent classes in the general population. That being said, we also believe that the CARME community is largely composed of 'graph people', who, despite their interest in formal mathematical expression, can still hear the voice of a deity proclaiming the importance of data visualization for understanding.

With these distinctions in mind, this chapter aims to sketch some of the historical antecedents of the topics that form the basis of this book. As self-confessed graph people, we confine ourselves to the right-brain, deity side, and consider developments and events in the history of data visualization that have contributed to two major revolutions: the rise of visual language for data graphics and successes in visual thinking.

To further align this chapter with the themes of this book, we focus largely on French contributions and developments to this history. We rely heavily on the resources publicly available via the Milestones Project (Friendly and Denis, 2001; Friendly, 2005; Friendly et al., 2013b). A graphical overview appears in Figure 1.1, showing birth places of 204 authors who are important contributors to this history. Of these, 36 were born in France, second only to the UK. The Google map on the `http://datavis.ca/milestone` site is global, zoomable and interactive, with each geographic marker linked to a query giving details about that individual.

[Figure 1 about here.]

1.1 Visual language

Data and information visualization, particularly for the descriptive and exploratory aims of CARME methods, are fundamentally about showing quantitative and qualitative information so that a viewer can see patterns, trends or anomalies, in ways that other forms— text and tables— do not allow (e.g., Tufte, 1997; Few, 2009; Katz, 2012; Yau, 2013). It is also important to realize that data displays are also communication tools— a visual message from a producer to a consumer— designed to serve some communication goal: *exploration & analysis* (to help see patterns and relations); *presentation* (to attract attention, illustrate a conclusion); and for *rhetoric* (to persuade). As such, effective data displays rely upon commonly understood elements and shared rules of visual language.

The rise of visual language

Such rules were developed through use and experimentation. In data-based maps, as well as statistical graphs and diagrams, this modern visual language arose over a period of time, largely the 18^{th} and 19^{th} centuries. This era also witnessed the rise of quantification in general, with many aspects of social, political and economic life measured and recorded (Porter, 1995); visual summaries were necessary to take stock of and gain insight about the growing body of data at hand.

With this increase in data, the graphical vocabulary for thematic maps surged. New features were introduced to show quantitative information, such as: *contour lines* that revealed the level curves of a surface (de Nautonier, 1604; Halley, 1701; von Humboldt, 1817), *dot symbols* that could be used to represent intensities such as population density (Frère de Montizon, 1830), *shading*, as in choropleth and dasymetric maps, to show the distribution of data variables such as education or crime (Dupin, 1826; Guerry, 1832), and *flow maps*, to show movement or change on a geographic background, such as those developed by Minard (1863).

Likewise, while the modern lexicon of statistical graphs stems largely from the work of William Playfair (1801) with the line graph, bar chart and pie chart, other methods soon followed, such as the scatterplot (Herschel, 1833), area charts (Minard, 1845) and other precursors to modern mosaic displays (Friendly, 1994), polar area diagrams (Guerry, 1829; Lalanne, 1845) or "rose diagrams" (Nightingale, 1858) and so forth.

In the second half of the 19^{th} century, a period we call the "Golden Age of Statistical Graphics" (Friendly, 2008), the International Statistical Congress began (in the third session, Vienna, 1857) to devote considerable attention to standardization of this graphical language. This work aimed to unify disparate national practices, avoid "babelisation" and codify rules governing conventions for data display (see Palsky, 1999).

However, absent of any over-arching theory of data graphics (what works, for what communication goals?) these debates faltered over the inability to resolve the differences between the artistic freedom of the graph designer to use the tools that worked, and the more rigid, bureaucratic view that statistical data must be communicated unequivocally, even if this meant using the lowest common denominator. For example, many of Minard's elegant inventions and combinations of distinct graphical elements (e.g., pie charts and flow lines on maps, subtended line graphs) would have been considered outside the pale of a standardized graphical language.

It is no accident that the next major step in the development of graphical language occurred in France (extending the tradition of Émile Cheysson, and Émile Levasseur) with Jacques Bertin's (1967; 1983) monumental *Sémiologie graphique.* Bertin codified (a) the "retinal variables" (shape, size, texture, color, orientation, position, etc.), and related these in combination with (b) the levels of variables to be represented ($Q :=$ quantitative, $O :=$ ordered, $\neq :=$ selective (categorical), $\equiv :=$ associative (similar)); (c) types of "impositions" on a planar display (arrangement, rectilinear, circular, orthogonal axes); and (d) common graphic forms (graphs, maps, networks, visual symbols).

Moreover, Bertin provided extensive visual examples to illustrate the the graphical effect of these combinations and considered their syntax and semantics. Most importantly, he considered these all from the perceptual and cognitive points of view of readability (elementary, intermediate, overall), efficiency (mental cost to answer a question), meaningfulness and memorability.

The most recent stage in this development of graphical language is best typified by Lee Wilkinson's (2005) *Grammar of Graphics.* It considers the entire corpus of data graphics from the perspectives of syntax (coordinates, graphical elements, scales, statistical summaries, aesthetics, ...) and semantics (representations of space, time, uncertainty). More importantly, it incorporates these features within a computational and expressive language for graphics, now implemented in the GPL language for SPSS (IBM Corporation, 2008) and the `ggplot2` (Wickham, 2009) package for R (R Development Core Team, 2012).

This is no small feat. Now consumers of statistical graphics can learn to *speak* (or write) in this graphical language; moreover, contributors to these methods, as in the present volume, can present their methods in computational form, making them more easily accessible to applied researchers. A leading example is the *Understanding Biplots* book (Gower, Lubbe, and Roux, 2011) that provides R packages to do all of the elaborate graphical displays in 2D and 3D that comprise a general biplot methodology related to PCA, CA, MCA, CCA and more.

The historical roots of these developments of visual language are firmly intertwined with those of data-based maps and statistical graphics. In the remainder of this section we highlight a few important contributions, largely from a French perspective.

Maps

In this subsection, there are many important French contributions we could emphasize. For example, amongst the earliest uses of isolines on a map was the world map by Guillaume de Nautonier de Castelfranc (1604) showing isogons of geomagnetism. This considerably predated Halley (1701) who is widely credited as the inventor of this graphic form.

Among many others, Phillipe Buache (1752) deserves mention for an early contour map of the topography of France that would later lead to the first systematic recording of elevations throughout the country by Charles Lallemand, mentioned later in this chapter. Moreover, although Playfair is widely credited as the inventor of the bar chart, the first known (to me) exemplar of this graphic form occurred in a graphic by Buache (1770), charting the ebb and flow of the waters in the Seine around Paris over time.

However, there is only one contribution of sufficient importance to describe and illustrate in any detail here, and that must be the work of André-Michel Guerry (1801–1864) on "moral statistics," which became the launching pad for criminology and sociology and much of modern social science. Guerry's work is especially relevant for this volume because it considers multivariate data in a spatial context. Beyond Guerry's own work, his data has proved remarkably useful for modern applications and demonstrations. For example, in Friendly (2007a) biplots, canonical discriminant plots, HE plots (Friendly, 2007b) and other CARME-related methods were used to provide a modern reassessment of his contributions and suggest other challenges for data analysis.

The choropleth map, showing the distribution of instruction in the French regional zones, called départements (departments), was invented by Charles Dupin (1826). Shortly after, Guerry, a young lawyer working for the Ministry of Justice, began the systematic study of relations between such variables as rates of crime, suicide, literacy, illegitimate births and so forth, using centralized, national data collected by state agencies. Guerry's life-long goal was to establish that constancies in such data provided the basis for social laws, analogous to those in the physical world and open discussion of social policy to empirical research.

In 1829, together with Adriano Balbi, he published the first *comparative* moral maps (Balbi and Guerry, 1829) showing the distribution of crimes against persons and against property in relation to the level of instruction in the départements of France, allowing direct comparison of these in a "small multiples" view (see Friendly, 2007a, Fig. 2). Surprisingly, they seemed to show an inverse relation between crimes against persons and property, yet neither seemed strongly related to levels of instruction.

[Figure 2 about here.]

Guerry followed this line in two major works (Guerry, 1833, 1864), both of which were awarded the Montyon prize in statistics from the Académie Française des Sciences. The 1833 volume, titled *Essai sur la Statistique Morale de la France*, established the methodology for standardized comparisons of rates of moral variables over time and space, and the rationale for drawing conclusions concerning social laws. In addition to tables, bar graphs and an innovative proto-parallel coordinates plot (showing relative ranking of crimes at different ages (Friendly, 2007a, Fig. 9)), he included six shaded maps of his main moral variables. A modern reproduction of these is shown in Figure 1.2.

[Figure 3 about here.]

Guerry wished to reason about the relationships among these variables, and, ultimately (in his final work, Guerry (1864)) about causative or explanatory social factors such as wealth, population density, gender, age, religious affiliation, etc. This is all the more remarkable because even the concept of correlation had not yet been invented.

We can give Guerry a bit of help here with the biplot of his data shown in Figure 1.3. This two-dimensional version accounts for only 56.2% of total variation, yet contains some interesting features. The first dimension aligns positively with property crime and illegitimate births (*infants naturelles*) and suicides, and negatively with literacy. The second dimension weights strongly on personal crime and donations to the poor. Using this and other dimension reduction techniques (e.g., CDA), Guerry could have seen more clearly how the regions of France and individual départements relate to his moral variables and underlying dimensions.

Graphs and diagrams

Aside from the standard, and now familiar methods to display quantitative data, French civil and military engineers made another important contribution to graphic language: nomograms and computational diagrams. These arose from the need to perform complex calculations (calibrate the range of field artillery, determine the amount of earth to be moved in building a railway or fortification) with little more than a straight-edge and a pencil (Hankins, 1999).

Toward the end of the 19^{th} century these developments, begun by Léon Lalanne (1844), gave rise to a full-fledged theory of projective geometry codified by Maurice d'Ocagne (1899). These ideas provide the basis for nonlinear scales used in nonlinear PCA (De Leeuw, this book), linear and nonlinear biplot calibrations (Gower, this book), contribution biplots (Greenacre, 2013) and the modern parallel coordinates plot, whose theoretical basis was also established by d'Ocagne (1885). This includes the principles of duality, by which points in Cartesian coordinates map into lines in alignment diagrams with parallel or oblique axes and vice versa, polar transformations of curves and surfaces, and so forth.

Among the most comprehensive of these is Lalanne's "Universal calculator," which allowed graphic calculation of over 60 functions of arithmetic (log, square root), trigonometry (sine, cosine), geometry (area, surface, volume) and so forth (see `http://datavis.ca/gallery/Lalanne.jpg` for a high-resolution image). Lalanne combined the use of parallel, nonlinear scales (as on a slide-rule) with a log-log grid on which any three-variable multiplicative relation could be represented.

[Figure 4 about here.]

Charles Lallemand, a French engineer, produced what might be considered the most impressive illustration of this work with the multi-graphic nomogram (Lallemand, 1885) shown in Figure 1.4. This tour-de-force graphic was designed to calculate the magnetic deviation of the compass at sea, which depends on seven variable through complex trigonometric formulas given at the top of the figure. It incorporates three-dimensional surfaces, an anamorphic map with nonlinear grids, projection through the central cone and an assortment of linear and nonlinear scales. Using this device, the captain could follow simple steps to determine magnetic deviation without direct calculation, and hence advise the crew when they might arrive at some destination.

Lallemand was also responsible for another grand project: the *Nivellement général de la France*, which mapped the altitudes of *all* of continental France. Today, you can still find small brass medallions embedded in walls in many small towns and villages throughout the country, indicating the elevation at that spot.

1.2 Visual thinking

The development of graphic language through the end of the 19^{th} century and the widespread adoption of graphic methods by state agencies did much more than make data graphics commonly available, in both popular expositions and official publications. For example, the *Album de Statistique Graphique*, published under the direction of Émile Cheysson by the Ministère des Traveaux Publiques from 1879–1897 represents a high point in the use of diverse graphic forms to chart the development of the modern French state.

It also presented a concrete means to plan for economic and social progress (where to build railroads and canals, how to bolster international trade) to reason and perhaps draw conclusions about important social issues (e.g., the discussion above of Guerry) and make some scientific discoveries that arguably could not have been arrived at otherwise.

We focus here on two aspects of this rise in visual thinking that characterize the Golden Age of statistical graphics: visual explanation, as represented by the work of Charles Joseph Minard and visual discovery, typified by the work of Francis Galton.

The graphic vision of Charles Joseph Minard

Minard, of course, is best known for his compelling and now iconic depiction of the terrible losses sustained by Napoleon's Grande Armée in the disastrous 1812 Russian campaign (Minard, 1869). However the totality of Minard's graphic work, comprising 63 *cartes figuratives* (thematic maps) and *tableaux graphiques* (statistical diagrams) is arguably more impressive as an illustrations of visual thinking and visual explanation.

[Figure 5 about here.]

Minard began his career as a civil engineer for the École Nationale des Ponts et Chausées (ENPC) in Paris. In 1840, he was charged to report on the collapse of a suspension bridge across the Rhone at Bourg-Saint-Andèol. The (probably apocryphal) story is that his report consisted essentially of a self-explaining before-after diagram (Friendly, 2008, Fig. 4) showing that the bridge collapsed because the river bed beneath one support column had eroded.

Minard's later work at the ENPC was that of a visual engineer for planning. His many graphics were concerned with matters of trade, commerce and transportation. We illustrate this here with another before-after diagram (Figure 1.5), designed to explain what happened to the trade in cotton and wool as a consequence of the U.S. Civil War. The conclusion from this pair of *cartes figuratives* is immediate and interoccular: Before the war, the vast majority of imports came from the southern U.S. states. By 1862, the Union naval blockade of the Confederacy reduced this to a tiny fraction; demand for these raw materials in Europe was only partially met by greater imports from Brazil and Egypt, but principally from India.

Francis Galton's visual discoveries

De Leeuw (this volume) points out that the early origin of PCA stems from the idea of principal axes of the "correlation ellipsoid," discussed by Galton (1889), and later developed mathematically by Pearson (1901). It actually goes back a bit further to Galton (1886) where he presented the first fully formed diagram of a bivariate normal frequency surface together with regression lines of $\mathcal{E}(y \mid x)$ and $\mathcal{E}(x \mid y)$, and also with the principal axes of the bivariate ellipse. This diagram and the correlation ellipsoid can arguably be considered the birth of modern multivariate statistical methods (Friendly et al., 2013a).

What is remarkable about this development is that Galton's statistical insight stemmed from a largely geometrical and visual approach using the smoothed and interpolated isopleth lines for 3D surfaces developed earlier by Halley, Lalanne and others. When he smoothed the semi-graphic table of heights of parents and their children and found that isolines of approximately equal frequency formed a series of concentric ellipses, Galton's imagination could complete the picture, and also offer the first true explanation of "regression toward mediocrity." Pearson (1920, p. 37) would later call this "one of the most noteworthy scientific discoveries arising from pure analysis of observations."

[Figure 6 about here.]

However, Galton achieved an even more notable scientific, visual discovery 25 years earlier in 1863— the anticyclonic relation between barometric pressure and wind direction that now forms the basis of modern weather maps and prediction. This story is described and illustrated in detail in Friendly (2008, §3.2) and will not be replayed here. In the book *Meteorographica* (Galton, 1863), he describes the many iterations of numerical and graphical summaries of the complex multivariate and spatial data he had elicited from over 300 weather stations throughout Europe at precise times (9am, 3pm, 9pm) for an entire month (December, 1861).

The result was a collection of micromaps (Figure 1.6) in a 3 × 3 grid of schematic contour maps showing barometric pressure, wind direction rain and temperature by time of day, using color, shape, texture and arrows. From this he observed something totally unexpected: whereas in areas of low barometric pressure, winds spiraled inwards rotating counterclockwise (as do cyclones), high pressure areas had winds rotating clockwise in outward spirals, which he termed "anti-cyclones." This surely must be among the best exemplars of scientific discovery achieved almost entirely through high-dimensional graphs.

1.3 Conclusion

This chapter demonstrates how the underlying attitudes of CARME— data exploration and analysis (largely model-free), reduction of complex, high-dimensional data to comprehensible low-dimensional views, and an emphasis on visualization— are rooted in a long, primarily European history that gave rise to the elements of visual language and visual thinking. Along with the rise of quantification and novel methods for visualization, came new ways to think about data and mathematical relationships, and to express them graphically.

Many of these innovations came from France, and were popularized and taught through works like *La méthod graphique* (Marey, 1885). The spirit of CARME, embodied in this volume, gives due attention to these historical developments we consider commonplace today.

References

Balbi, A. and Guerry, A.-M. (1829). Statistique comparée de l'état de l'instruction et du nombre des crimes dans les divers arrondissements des académies et des cours royales de France. Jules Renouard, Paris.

Bertin, J. (1967). *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes.* Paris: Gauthier-Villars.

Bertin, J. (1983). *Semiology of Graphics.* Madison, WI: University of Wisconsin Press. (trans. W. Berg).

Buache, P. (1752). Essai de géographie physique. *Mémoires de L'Académie Royale des Sciences*, (pp. 399–416).

Buache, P. (1770). Profils représentants la cruë et la diminution des eaux de la Seine et des rivieres qu'elle recoit dans le Paris-haut au dessus de Paris. Paris: G. de L'Isle et P. Buache.

de Nautonier, G. (1602–1604). *Mecometrie de l'eymant, c'est a dire la maniere de mesurer les longitudes par le moyen de l'eymant.* Paris: R. Colomies.

Dupin, C. (1826). *Carte figurative de l'instruction populaire de la France.* Jobard.

Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis.* Oakland, California: Analytics Press.

Frère de Montizon, A. J. (1830). Carte philosophique figurant la population de la France.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89, 190–200.

Friendly, M. (2005). Milestones in the history of data visualization: A case study in statistical historiography. In C. Weihs and W. Gaul, eds., *Classification: The Ubiquitous Challenge*, (pp. 34–52). New York: Springer.

Friendly, M. (2007a). A.-M. Guerry's Moral Statistics of France: Challenges for multivariable spatial analysis. *Statistical Science*, 22(3), 368–399.

Friendly, M. (2007b). HE plots for multivariate general linear models. *Journal of Computational and Graphical Statistics*, 16(2), 421–444.

Friendly, M. (2008). The Golden Age of statistical graphics. *Statistical Science*, 23(4), 502–535.

Friendly, M. and Denis, D. (2001). Milestones in the history of thematic cartography, statistical graphics, and data visualization. Web document. `http://www.math.yorku.ca/SCS/Gallery/milestone/`.

Friendly, M. and Kwan, E. (2011). Comment (graph people versus table people). *Journal of Computational and Graphical Statistics*, 20(1), 18–27.

Friendly, M., Monette, G., and Fox, J. (2013a). Elliptical insights: Understanding statistical methods through elliptical geometry. *Statistical Science*, 28(1), 1–39.

Friendly, M., Sigal, M., and Harnanansingh, D. (2013b). The Milestones Project: A database for the history of data visualization. In M. Kimball and C. Kostelnick, eds., *Visible Numbers*. London, UK: Ashgate Press. In press.

Galton, F. (1863). *Meteorographica, or Methods of Mapping the Weather*. London: Macmillan.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, 15, 246–263.

Galton, F. (1889). *Natural Inheritance*. London: Macmillan.

Gower, J., Lubbe, S., and Roux, N. (2011). *Understanding Biplots*. Chchester, UK: Wiley.

Greenacre, M. (2013). Contribution biplots. *Journal of Computational and Graphical Statistics*, 22(1), 107–122.

Guerry, A.-M. (1829). Tableau des variations météorologique comparées aux phénomènes physiologiques, d'aprés les observations faites à l'obervatoire royal, et les recherches statistique les plus récentes. *Annales d'Hygiène Publique et de Médecine Légale*, 1, 228–237.

Guerry, A.-M. (1832). *Statistique comparée de l'état de l'instruction et du nombre des crimes*. Paris: Everat.

Guerry, A.-M. (1833). *Essai sur la statistique morale de la France*. Paris: Crochard. English translation: Hugh P. Whitt and Victor W. Reinking, Lewiston, N.Y. : Edwin Mellen Press, 2002.

Guerry, A.-M. (1864). *Statistique morale de l'Angleterre comparée avec la statistique morale de la France, d'après les comptes de l'administration de la justice criminelle en Angleterre et en France, etc.* Paris: J.-B. Baillière et fils.

Halley, E. (1701). The description and uses of a new, and correct sea-chart of the whole world, shewing variations of the compass. London.

Hankins, T. L. (1999). Blood, dirt, and nomograms: A particular history of graphs. *Isis*, 90, 50–80.

Herschel, J. F. W. (1833). On the investigation of the orbits of revolving double stars. *Memoirs of the Royal Astronomical Society*, 5, 171–222.

IBM Corporation (2008). *GPL Reference Guide for IBM SPSS Visualization Designer.*

Jaynes, J. (1978). *The Origin of Consciousness in the Breakdown of the Bicameral Mind.* London: Houghton Mifflin.

Katz, J. (2012). *Designing Information.* Hoboken, New Jersey: John Wiley and Sons.

Lalanne, L. (1844). *Abaque, ou Compteur univsersel, donnant á vue á moins de 1/200 près les résultats de tous les calculs d'arithmétique, de geometrie et de mécanique practique.* Paris: Carilan-Goery et Dalmont.

Lalanne, L. (1845). Appendice sur la representation graphique des tableaux météorologiques et des lois naturelles en général. In L. F. Kaemtz, ed., *Cours Complet de Météorologie*, (pp. 1–35). Paulin. Translated and annotated by C. Martins.

Lallemand, C. (1885). *Les abaques hexagonaux: Nouvelle méthode générale de calcul graphique, avec de nombreux exemples d'application.* Paris: Ministère des travaux publics, Comité du nivellement général de la France.

Marey, E.-J. (1885). *La méthode graphique dans les sciences expérimentales.* Paris: Masson.

Minard, C. J. (1845). Tableau figuratif du mouvement commercial du canal du Centre en 1844 dressé d'apré les renseignements de M. Comoy. lith. (n.s.).

Minard, C. J. (1862). Carte figurative et approximative des quantités de coton en laine importées en Europe en 1858 et en 1861. lith. (868 x 535). ENPC: Fol 10975.

Minard, C. J. (1863). Carte figurative et approximative des grands ports du globe, 2 ed. corrigée et augmentée de 26 ports. lith. (765 x 540).

Minard, C. J. (1869). Carte figurative des pertes successives en hommes de l'armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe). Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie, 1812–1813. lith. (624 x 207, 624 x 245).

Nightingale, F. (1858). *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army*. London: Harrison and Sons.

d'Ocagne, M. (1885). *Coordonnées Parallèles et Axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèlles*. Paris: Gauthier-Villars.

d'Ocagne, M. (1899). *Traité de nomographie: Théorie des Abaques, Applications Pratiques*. Paris: Gauthier-Villars.

Palsky, G. (1999). The debate on the standardization of statistical maps and diagrams (1857-1901). *Cybergeo*, (65). Retrieved from `http://cybergeo.revues.org/148`.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6(2), 559–572.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13(1), 25–45.

Playfair, W. (1801). *Statistical Breviary; Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*. London: Wallis. Re-published in Wainer, H. and Spence, I. (eds.), *The Commercial and Political Atlas and Statistical Breviary*, 2005, CAmbridge, UK: Cambridge University Press, ISBN 0-521-85554-3.

Porter, T. M. (1995). *Trust in Numbers*. Princeton, New Jersey: Princeton University Press.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Tufte, E. R. (1997). *Visual explanations: Images and quantities evidence and narrative*. Cheshire, Connecticut: Graphics Press.

von Humboldt, A. (1817). Sur les lignes isothermes. *Annales de Chimie et de Physique*, 5, 102–112.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

Wilkinson, L. (2005). *The Grammar of Graphics.* New York: Springer, 2nd edn.

Yau, N. (2013). *Data Points.* Indianapolis, Indiana: John Wiley and Sons.

List of Figures

1.1 Birth places of milestones authors. Left: Portion of an interactive Google map, centered on France. The highlighted point is that for André-Michel Guerry, born in Tours, Dec. 24, 1802. Right: Frequencies, by country of birth. 16

1.2 Reproduction of Guerry's (1833) maps of moral statistics of France. Shading, as in Guerry's originals, is such that darker shading signifies *worse* on each moral variable, ranked across departments (shown by numbers in the map). 17

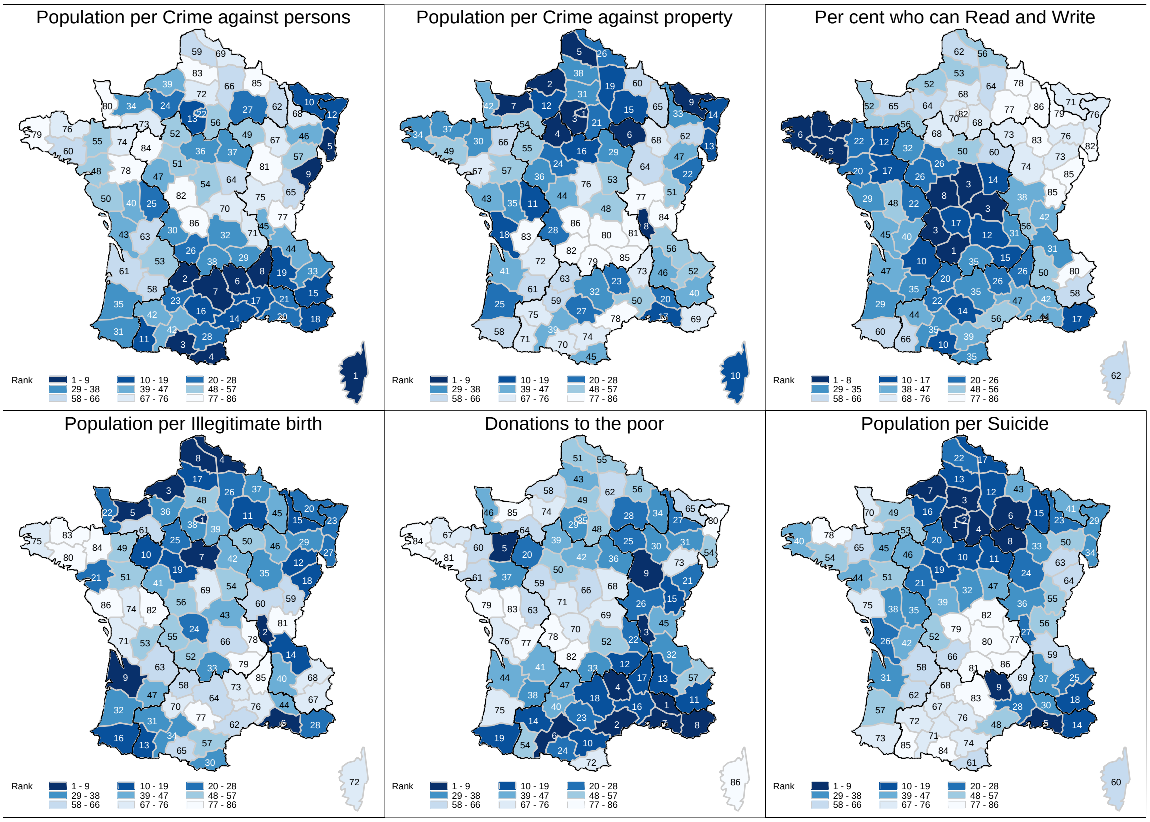
1.3 A symmetric 2D biplot of Guerry's six moral variables shown in maps in Figure 1.2. The points for the départements of France are summarized by region (N, S, E, W, C) with 68% data ellipses and points outside their ellipse are labeled by department name. 18

1.4 Nomograms: computational diagrams and axis calibration. This tour-de-force nomogram by Charles Lellemand combines diverse graphic forms (anamorphic maps, parallel coordinates, 3D surfaces) to calculate magnetic deviation at sea. *Source*: École des Mines, Paris, reproduced by permission. 19

1.5 Visual explanation: What happened to the trade in cotton and wool from Europe in the U.S. civil war? Left: In 1858, most imports to Europe came from the southern U.S. states. Right: by 1862, U.S. imports had been reduced to a trickle, only partially compensated by increased imports from India, Brazil and Egypt. *Source*: Minard (1862), image from École Nationale des Ponts et Chaussées, Paris, reproduced by permission. 20

1.6 Visual discovery: Top portion of Galton's (1863) multivariate schematic micromaps. Each 3×3 grid shows barometric pressure, wind, rain and temperature (rows) by time of day (columns). *Source*: Galton (1863), Appendix, p. 3, image from a private collection. 21

Figure 1.1: Birth places of milestones authors. Left: Portion of an interactive Google map, centered on France. The highlighted point is that for André-Michel Guerry, born in Tours, Dec. 24, 1802. Right: Frequencies, by country of birth.



Population per Crime against persons

Population per Crime against property

Per cent who can Read and Write

Population per Illegitimate birth

Donations to the poor

Population per Suicide

Figure 1.2: Reproduction of Guerry's (1833) maps of moral statistics of France. Shading, as in Guerry's originals, is such that darker shading signifies *worse* on each moral variable, ranked across departments (shown by numbers in the map).

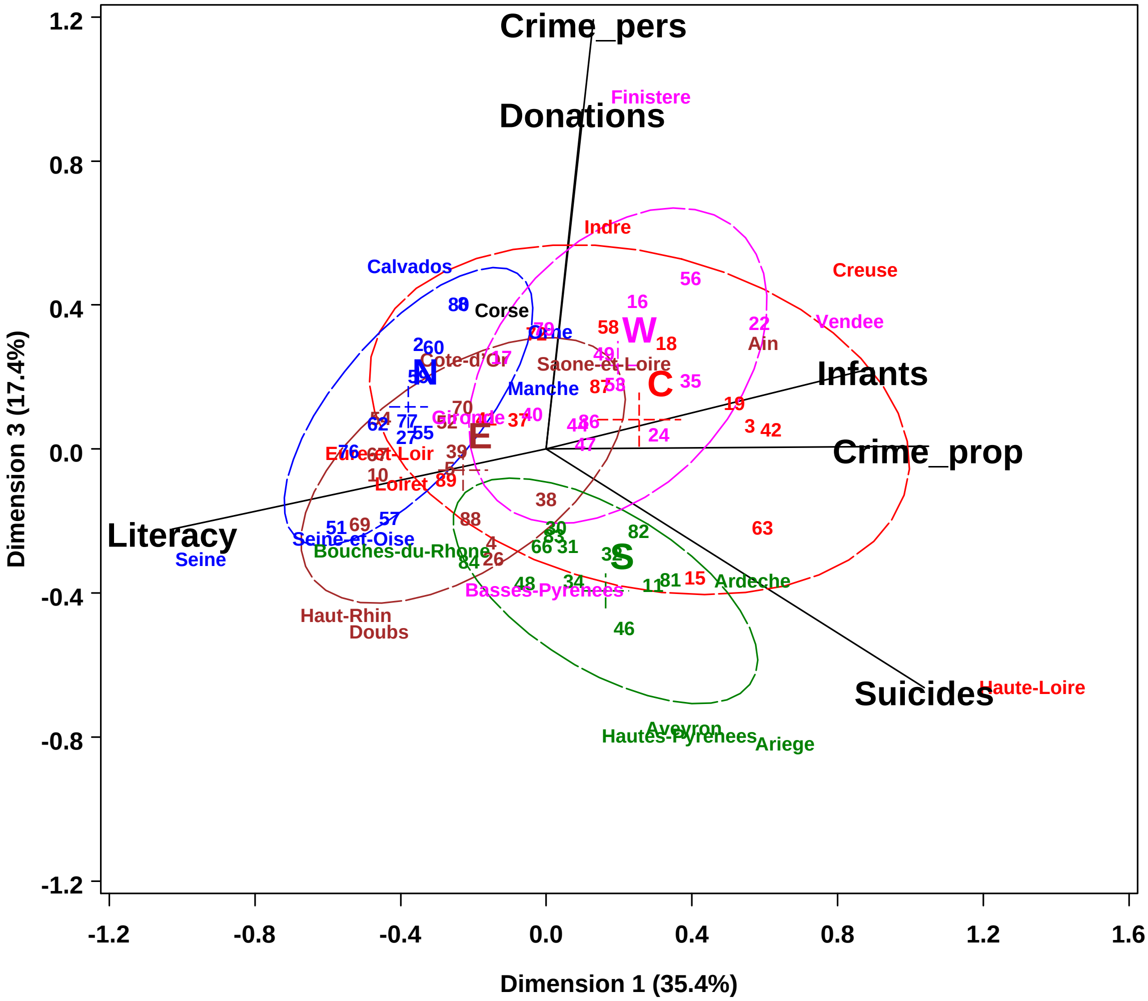
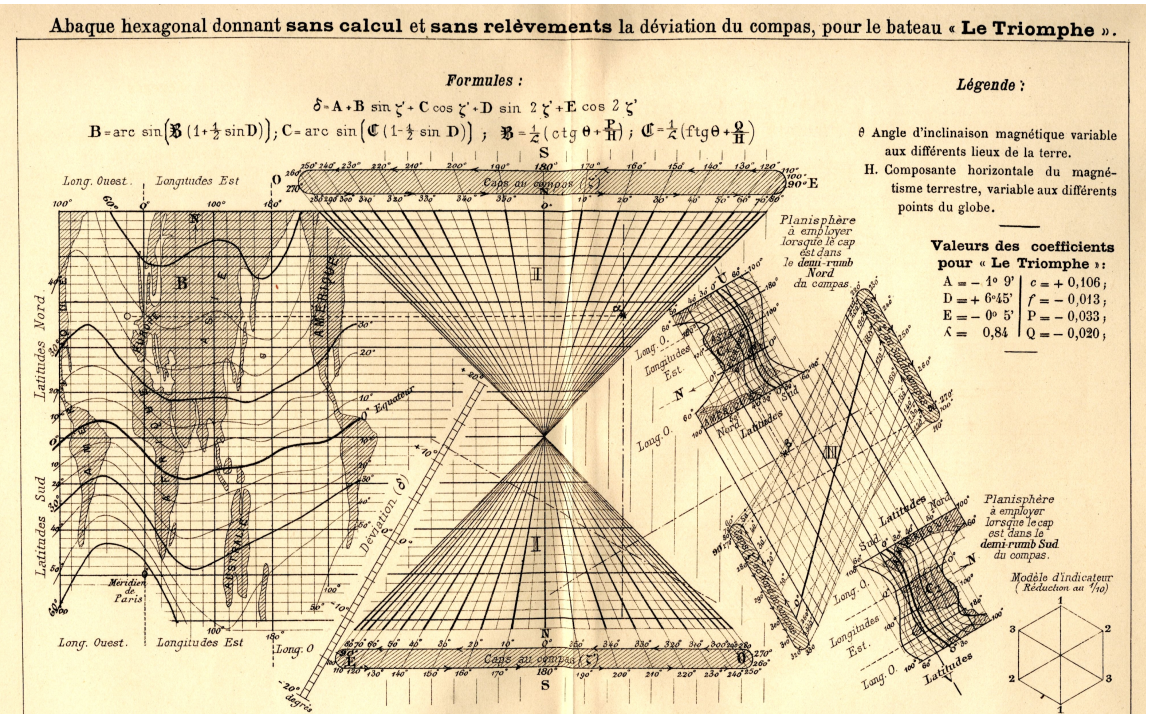


Figure 1.3: A symmetric 2D biplot of Guerry's six moral variables shown in maps in Figure 1.2. The points for the départements of France are summarized by region (N, S, E, W, C) with 68% data ellipses and points outside their ellipse are labeled by department name.



Abaque hexagonal donnant **sans calcul** et **sans relèvements** la déviation du compas, pour le bateau « **Le Triomphe** ».

Formules :

$$\delta = \mathbf{A} + \mathbf{B} \sin \zeta' + \mathbf{C} \cos \zeta' + \mathbf{D} \sin 2\zeta' + \mathbf{E} \cos 2\zeta'$$

$$\mathbf{B} = \text{arc} \sin\left(\mathfrak{B}\left(1 + \tfrac{1}{2} \sin \mathbf{D}\right)\right); \ \mathbf{C} = \text{arc} \sin\left(\mathfrak{C}\left(1 - \tfrac{1}{2} \sin \mathbf{D}\right)\right) \ ; \ \mathfrak{B} = \tfrac{1}{2}\left(\text{ctg}\, \theta + \tfrac{\mathbf{P}}{\mathbf{H}}\right); \ \mathfrak{C} = \tfrac{1}{2}\left(\text{ftg}\, \theta + \tfrac{\mathbf{Q}}{\mathbf{H}}\right)$$

Légende :

θ Angle d'inclinaison magnétique variable aux différents lieux de la terre.

H. Composante horizontale du magnétisme terrestre, variable aux différents points du globe.

Valeurs des coefficients pour « Le Triomphe » :

A = − 1° 9' | c = + 0,106 ;
D = + 6°45' | f = − 0,013 ;
E = − 0° 5' | P = − 0,033 ;
λ = 0,84 | Q = − 0,020 ;

Figure 1.4: Nomograms: computational diagrams and axis calibration. This tour-de-force nomogram by Charles Lellemand combines diverse graphic forms (anamorphic maps, parallel coordinates, 3D surfaces) to calculate magnetic deviation at sea. *Source*: École des Mines, Paris, reproduced by permission.

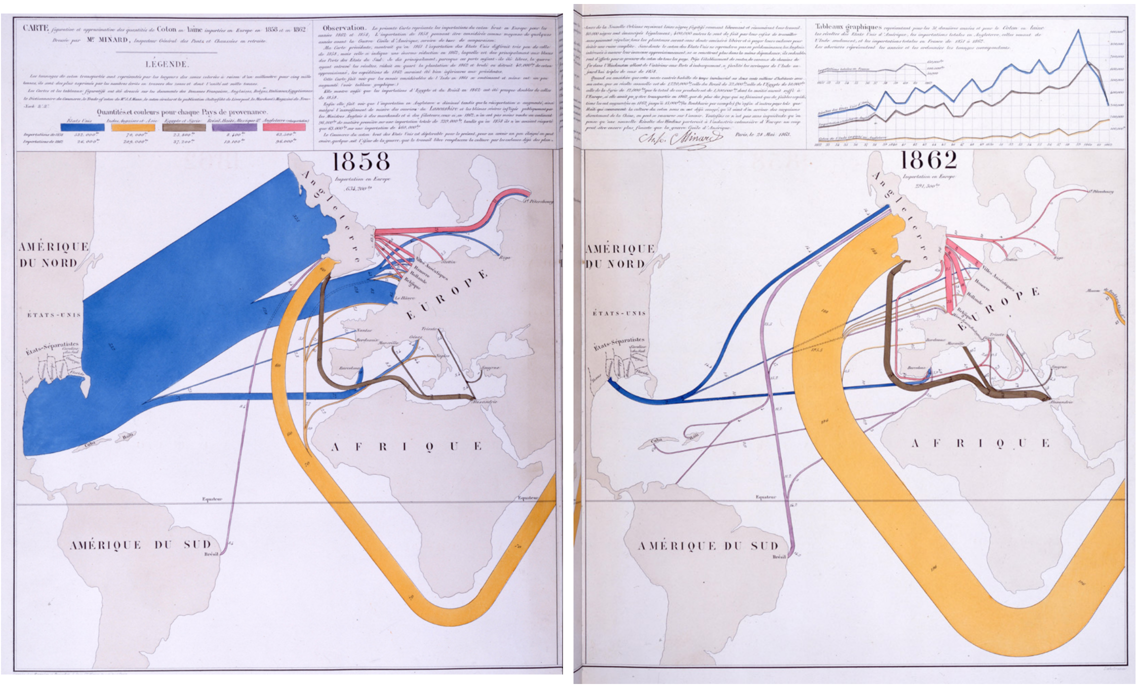


Figure 1.5: Visual explanation: What happened to the trade in cotton and wool from Europe in the U.S. civil war? Left: In 1858, most imports to Europe came from the southern U.S. states. Right: by 1862, U.S. imports had been reduced to a trickle, only partially compensated by increased imports from India, Brazil and Egypt. *Source*: Minard (1862), image from École Nationale des Ponts et Chaussées, Paris, reproduced by permission.

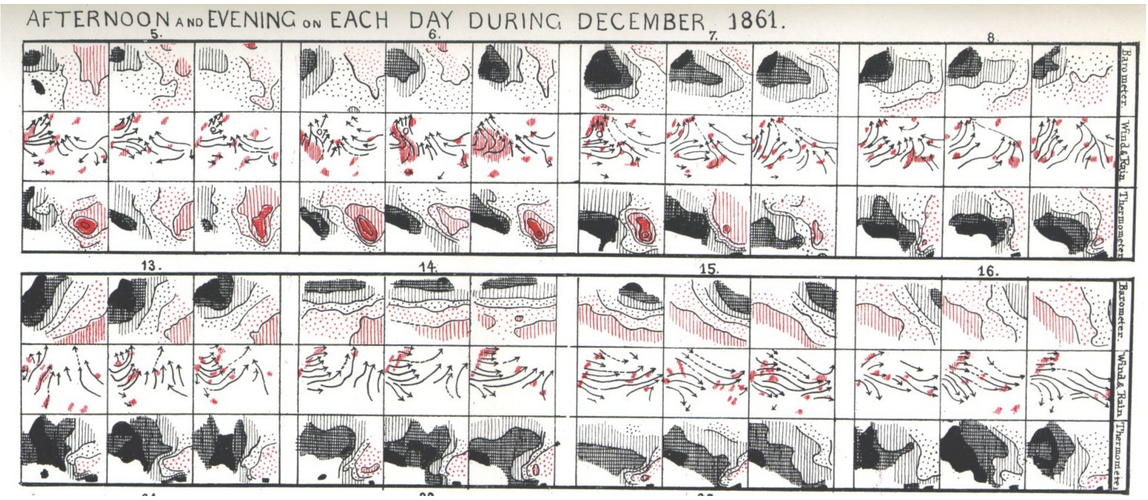


Figure 1.6: Visual discovery: Top portion of Galton's (1863) multivariate schematic micromaps. Each 3×3 grid shows barometric pressure, wind, rain and temperature (rows) by time of day (columns). *Source*: Galton (1863), Appendix, p. 3, image from a private collection.