DAVID SEPKOSKI*

Data in Time: Statistics, Natural History, and the Visualization of Temporal Data

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

ABSTRACT

One of the best arguments for approaching the history of information processing and handling in the human and natural sciences as a "history of data" is that it focuses our attention on relationships, convergences, and contingent historical developments that can be obscured following more traditional areas of focus on individual disciplines or technologies. This essay explores one such case of convergence in nineteenth-century data history between empirical natural history (paleontology and botany), bureaucratic statistics (cameralism), and contemporary historiography, arguing that the establishment of visual conventions around the presentation of temporal patterns in data involved interactions between ostensibly distinct knowledge traditions. This essay is part of a special issue entitled *Histories of Data and the Database* edited by Soraya de Chadarevian and Theodore M. Porter.

KEY WORDS: statistics, natural history, visualization, cameralism, data

In recent years, historians of science have been drawn to a burgeoning subfield, the "history of data."[1] Although there have been many good arguments for adopting data as a historical category, there are legitimate questions about how this recent scholarly interest relates to established histories of quantification, statistics, mathematics, computing, and of the various disciplines in the natural and social sciences. My own position is that histories of data do not replace or subordinate these other research programs, but rather bring together scholarly perspectives that have not otherwise necessarily been in conversation

*Department of History, University of Illinois, Urbana-Champaign, 301 Gregory Hall, 810 S. Wright St., Urbana, IL 61801, sepkoski@illinois.edu

1. See, for example, Elena Aronova, Christine von Oertzen, and David Sepkoski, eds., "Data Histories," *Osiris* 32 (2017).

Historical Studies in the Natural Sciences, Vol. 48, Number 5, pps. 581–593. ISSN 1939-1811, electronic ISSN 1939-182X. © 2018 by the Regents of the University of California. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Reprints and Permissions web page, http://www.ucpress.edu/journals.php?p=reprints. DOI: https://doi.org/10.1525/hsns.2018.48.5.581.

to shed light on practices, epistemologies, technologies, and cultures around the collection and analysis of data in the *longue durée*. This has been my approach in recent work linking nineteenth-century natural history, state administrative sciences, and humanistic historical conventions, where I have found that it has been illuminating, for example, to link the work of Ted Porter and others on the history of statistics with studies of natural history and geology in the tradition of Martin Rudwick.[2]

This essay does not attempt to summarize those more detailed studies, but rather offers some broad reflections on a particular theme in the history of data: the turn toward visual depictions of temporal patterns in data in the nineteenth-century natural and economic sciences. It has long been established that the nineteenth century was a period when statistical arguments became a central component of scientific epistemology in Europe and beyond. This was a time during which, to use Porter's memorable phrase, social and natural scientists established "trust in numbers" as the basis for statistical decision-making in a wide variety of disciplines.[3] This was a process that was political as well as methodological and epistemological: statistical reasoning, as Porter, Alain Desrosiers, Ian Hacking, and others have demonstrated, requires the tacit social approval that numbers and statistical objects support valid representations of reality, and that their analysis can provide a sound basis for economic and social interventions.[4] The statistical turn introduced the broad acceptance of a new kind of object—the statistical aggregate—that was neither wholly abstract, being derived from individual data points, nor reducible to any specific individual or moment in time, but able to stand for a general class. Statistical entities such as "the average man," "unemployment," or "crime rate"

2. David Sepkoski, "Towards 'a Natural History of Data': Evolving Practices and Epistemologies of Data in Paleontology, 1800–2000," *Journal of the History of Biology* 46, no. 3 (2013): 401–44. Sepkoski, "The Earth as Archive," in *Archiving Sciences: Pasts, Presents, Futures*, ed. Lorraine Daston (Chicago: University of Chicago Press, 2017), 53–83; David Sepkoski and Marco Tamborini, "'An Image of Science': Cameralism, Natural History, and the Visual Language of Statistics in the Nineteenth Century," *Historical Studies in the Natural Sciences* 48, no. 1 (2018): 56–109. Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton. NJ: Princeton University Press, 1995); Porter, *The Rise of Statistical Thinking, 1820–1900* (Princeton, NJ: Princeton University Press, 1986). Martin J. S. Rudwick, *Bursting the Limits of Time: The Reconstruction of Geohistory in the Age of Revolution* (Chicago: University of Chicago Press, 2005).

3. Porter, *Trust in Numbers* (ref. 2).

4. Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA: Harvard University Press, 2002); Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990).

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

became categories that shaped new perceptions of social, political, economic, and scientific reality.

Two implicit corollaries of this turn toward aggregate statistical reasoning stand out. The first is that the construction of broad statistical objects—whether in economics or state administration or natural history—required large amounts of data as grist for the statistical mill. The nineteenth century was, as Hacking has put it, the era of an "avalanche of printed numbers," and armies of bureaucratically organized collectors funneled data on census figures, weather observations, botanical and zoological specimens, and the like toward centralized experts tasked with organizing and interpreting them.[5] These projects were supported by new institutions (e.g., large statistical bureaus), new technologies (e.g., index cards, paper slips, and eventually punched cards), new social arrangements (a hierarchical system in which many relatively untrained "invisible workers" supplied raw material to fewer expert analysts), and new disciplines (the state sciences, economics, the modern professionalized natural sciences). As much as this process required trust in numbers, it also required trust in *data* and in the patterns extracted from large data collections. We speak now of the twenty-first century being the age of "Big Data," but much of the machinery—to say nothing of the underlying totalizing ambitions of data-driven statistical projects—was established long ago. The nineteenth century was an era of big data, too.[6]

The second is that in order for the results of statistical reasoning to become legible and translatable to a relatively broad audience—and for statistical methods to be standardized to the degree that they could be adapted across many disparate disciplines—new methods of representing statistical arguments were developed. Many of the characteristic elements of today's statistical language—particularly pictorial charts and graphs of all descriptions—were invented or popularized during the later nineteenth century, an era one historian has labeled "the Golden Age of statistical graphics."[7] The statistical turn required a new visual culture that enabled people to "see" statistical objects: graphical representations made visible entities that were composed of abstracted data points or mathematical relationships. We take such

5. Ian Hacking, "Biopower and the Avalanche of Printed Numbers," *Humanities in Society* 5 (1982): 279–95.

6. See, particularly, Christine von Oertzen, "Datafication and Visualization in Nineteenth-Century Census Statistics," in this Issue.

7. Michael Friendly, "The Golden Age of Statistical Graphics," *Statistical Science* 23, no. 4 (2008): 502–35.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

visualizations to be intuitive today, but it took many decades over the nineteenth century for these conventions to be accepted, and their emergence was tied to highly contingent convergences between practices in often disparate fields of inquiry. How did this transformation come about?

For insight into this question, I will point to contemporary developments in two seemingly disparate fields of knowledge: early nineteenth-century state science (cameralism), and paleontology. As my central case study, I will highlight one particular convergence: the emergence of visual approaches to patterns in temporal data. My argument is that the temporalizing of data in graphical form drew on developments in both traditions, but that the knowledge exchange involved depended on actors whose interests were highly idiosyncratic and divergent. This case study, then, offers both an insight into the contingent circumstances around the establishment of a particular visual idiom, and an advertisement for the usefulness of "history of data" as a category.

As Daniel Rosenberg and Anthony Grafton have shown in their innovative study of the history of the "timeline," the now-familiar convention that time can be visually represented as a line passing through measured graphical space did not fully emerge until the late eighteenth century.[8] One of the most influential early presentations of this visual idiom was Joseph Priestley's 1764 *A Description of a Chart of Biography*, which depicted the lives of some 2,000 notable historical figures as horizontal lines measured against absolute graphic space. As Priestly put it, "Thus the abstract idea of TIME . . . admits of a natural and easy representation in our minds by the idea of a measurable space, and particularly that of a line; which like time, may be extended in length, without giving any idea of breadth or thickness."[9]

This visual device caught on quickly and was often imitated, but in its initial form it was somewhat limited: being two-dimensional, the only meaningful information is conveyed on the horizontal (temporal) axis. However, others quickly realized that the vertical axis could also be employed, enabling a graph to demonstrate some *change* in a variable over time. Most notably, the Scottish engineer and political economist William Playfair adapted Priestley's convention to now-familiar line graphs, representing (for example) fluctuating imports and exports in England. The importance of this innovation was that it merged the geometric tradition of the Cartesian coordinate system

8. Daniel Rosenberg and Anthony Grafton, *Cartographies of Time: A History of the Timeline* (Princeton, NJ: Princeton Architectural Press, 2010), 113.

9. Joseph Priestley, *A Description of a Chart of Biography* (London: J. Johnson, 1764), 5.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

(developed in the seventeenth century) with Priestley's projection of time as graphical space.

Another important innovation of Playfair's was that his graphs arranged economic data in temporal series; that is to say, they presented a historical account of a phenomenon composed of data points. But in his own day, this epistemic convention did not catch on—at least in the economic sciences. At the turn of the nineteenth century there was a well-established tradition in Northern Europe and Britain of quantitative analysis of state resources, labeled variously "cameralism" or "political arithmetic," that attempted to quantify or forecast state resources and population statistics using large quantities of data. However, these approaches almost never utilized graphs, being instead focused on the tabular presentation of data.[10] Indeed, despite the apparently obvious fact that "tables of statistics in general, and comparative political data in particular, suggests the need for graphical comparison," prior to the 1860s graphical visualizations of statistics were not warmly welcomed into economic arguments.[11] When, for example, the English mathematical economist William Stanley Jevons approached the noted statistician and publisher William Newmarch in the early 1860s about publishing a "Statistical Atlas" he was composing, he was taken aback by Newmarch's profound disinterest in his statistical visualizations. As economic historian Haro Maas puts it, "Apparently, it was not at all obvious for someone like Newmarch to present a table of numbers in a graph."[12]

However, one place where narrative graphical summaries of data were fairly well-established prior to the mid-nineteenth century was taxonomic natural history. For example, Alexander von Humboldt devised a number of innovative visual methods for summarizing biogeographical and climactic information, such as the image depicting the geographical distribution of plants on the island of Tenerife in Figure 1. Although somewhat stylized, this image is, effectively, a line graph: as von Humboldt himself acknowledged, "This graphical method is analogous to what M. Playfair first employed," noting that "the map which M. Playfair gives of the national debt of England brings to mind the section of the Pic of Teneriffe." Graphical techniques such as these were essential to Humboldt's communication of complex relationships in data

10. James R. Beniger and Dorothy L. Robyn, "Quantitative Graphics in Statistics: A Brief History," *The American Statistician* 32 (1978): 2.

11. Ibid., 3.

12. Harro Maas, *William Stanley Jevons and the Making of Modern Economics* (Cambridge: Cambridge University Press, 2005), 218.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

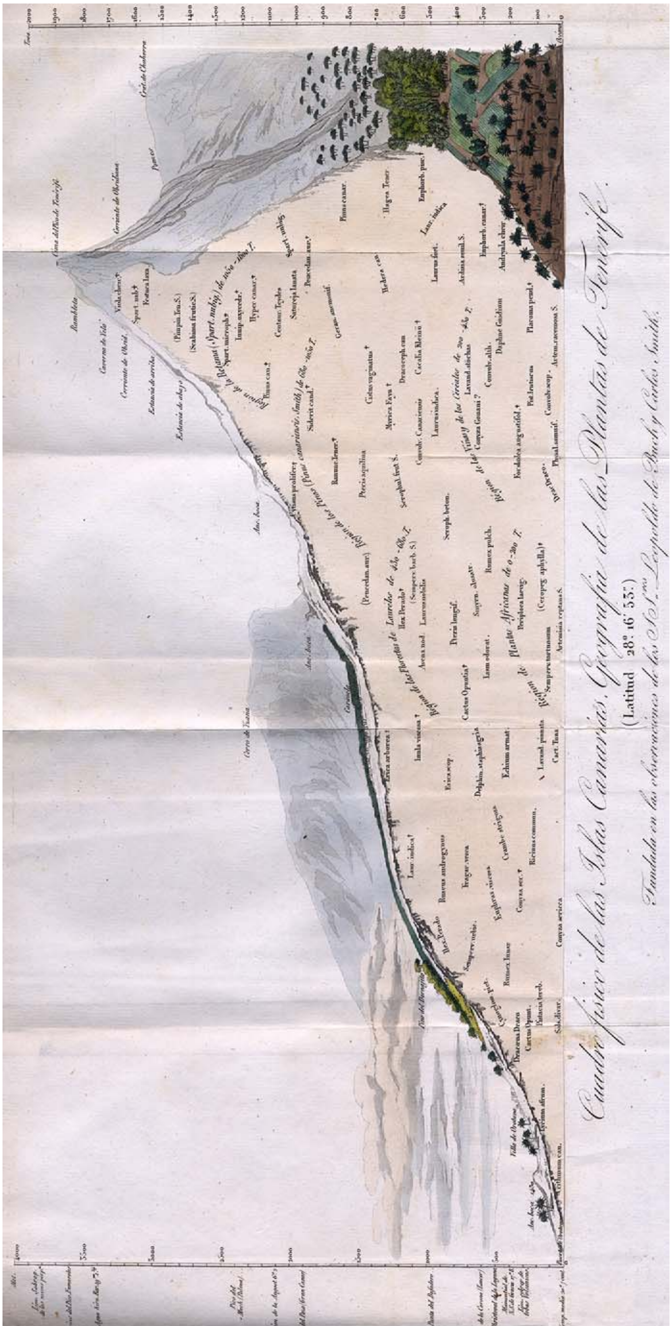

FIGURE 1: Alexander von Humboldt's "line graph" showing distribution of plants on the island of Tenerife. *Source:* Alexander von Humboldt, *Viage á las regiones equinocciales del nuevo continente: Hecho en 1799 hasta 1804, por Al. de Humboldt y A. Bonpland* (Paris: En Casa de Rosa, 1826), vol. 1; reproduced in Humboldt, *Political Essay on the Kingdom of New Spain* (London: Longman, 1814), vol. 1.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

since, as he put it, “statistical projections that speak to the senses without fatiguing the mind, possess the advantage of fixing the attention on a great number of important facts.”[13]

Geologists, too, had a fairly established visual culture by the early decades of the nineteenth century. Whereas some of the most famous of these images were spatial representations of the stratigraphic layers of the earth, arranged either as topographical maps (William Smith) or idealized stratigraphic columns (Georges Cuvier and Alexandre Brongniart), others followed the tradition of Priestley and Playfair by representing temporal data in graphical space.[14] One of the first true line graphs in paleontology was published in 1860 by the English geologist John Phillips (who, interestingly, was the nephew and collaborator of the stratigraphical pioneer William Smith), depicting changes in the taxonomic diversity of fossils over the history of life (Fig. 2). Phillips’s image abstracted a pattern of the waxing and waning of life’s diversity from a compendium of taxonomic data in a manner precisely analogous to the method that had been employed by Playfair and by contemporaries like William Jevons—it revealed a complex temporal phenomenon in data that would otherwise have been invisible.[15]

Whereas Humboldt was quite open about the source of his graphical inspiration, the story in geology and paleontology is a bit more complex. Phillips’s graph was, in 1860, published at a moment when similar graphs were becoming more popular in political economy, demography, and the physical sciences, and Phillips might have had any number of models to draw on. But similar kinds of narrative visual summaries had already emerged significantly earlier, in the late 1830s and 1840s, well before such visualizations of economic statistics had become broadly accepted. One particularly significant example was the work of the German naturalist Georg Heinrich Bronn, a paleontologist who occupied a chair at Heidelberg and was considered in his day to be one of the leading scientists in Europe. Bronn advocated a distinctive approach to the study of life that was explicitly statistical in nature: basically, he spent his career tabulating data on all known taxonomic groups of fossils, which he compiled in huge compendia, like his 1849 *Index palaeontologicus: Oder Übersicht der bis*

13. Alexander von Humboldt, *Political Essay on the Kingdom of New Spain*, vol. 1 (London: Longman, 1814), cxxxii–iii.

14. Georges Cuvier and Alexandre Brongniart, *Description géologique des environs de Paris* (Paris, 1812); William Smith, *Strata Identified by Organized Fossils* (London,: W. Arding, 1816).

15. John Phillips, *Life on the Earth* (London: John Murray, 1860).

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

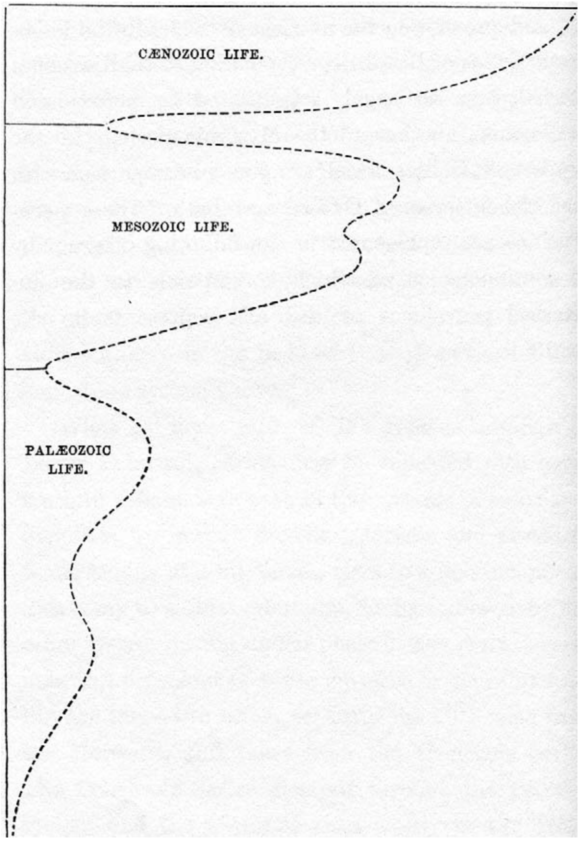

FIGURE 2: John Phillips' depiction of the history of life as a line graph tracing three successive curves. *Source:* John Phillips, *Life on the Earth* (1860), 66.

jetzt bekannten fossilen Organismen (Index of Paleontology: or an Overview of the Hitherto Known Fossil Organisms).[16]

Bronn's method was to convert standard taxonomic entries found in paleontological catalogs and monographs into numerical data, which he then subjected to a variety of statistical analyses (e.g., calculating average durations of species, ratios of species to genera at different periods, relative composition and distribution of these groups in time). These relationships were expressed, in the first instance, as numerical tables, much like the tables favored by the cameralists and political arithmeticians of the late eighteenth century. There

16. H. G. Bronn, *Index Paleontolgicus, oder Übersicht der bist jetzt bekannten Fossilen Organismen* (Stuttgart: 1849).

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

was a very good reason for this approach: Bronn, though ultimately appointed to the first Chair in Zoology at his university, was in fact trained not as a naturalist, but as a practitioner of the bureaucratic science of *Kameralwissenschaft* (cameralism). Indeed, Bronn received his habilitation in that subject from the University of Heidelberg in 1821, and immediately joined the faculty in the school of cameral studies. He did not publish on geological topics until the early 1830s, although his particular focus in state administration was forestry and agriculture, which led him to travel widely, observing natural environments in Germany, Swizerland, and Italy that ultimately seems to have inspired his interest in fossils.

However, it is noteworthy not just that Bronn was a cameralist, but especially that he was trained and taught at Heidelberg during a period of transition in economic state sciences. Although it has been fairly extensively documented that most cameralist practice—and, indeed, statistical practice at the time more generally—viewed the accumulation of numerical data as a purely descriptive enterprise that explicitly avoided seeking explanatory patterns and laws, as Marco Tamborini and I have shown elsewhere, the Heidelberg school was an exception.[17] Specifically, Heidelberg cameralists followed the German agronomist Albrech Daniel Thaer and his student Carl von Wulffen in distinguishing between "statistics" and "statics," the latter of which they defined as an approach explicitly contrasted to what was perceived as the dry collection of data for data's sake. As practiced at Heidelberg—for example, by the forestry expert Johann Christian Hundeshagen—cameralism involved a two-part process: first the data was collected and tabulated, which was statistics, and next mathematical relationships within the tables were calculated, which was statics.

These lessons were absorbed deeply by Bronn. His compendia of fossil data were assembled not just to document the empirical record of fossils, but to determine relationships among the data that could reveal patterns in the history of life, and Bronn seems to have been aware of the novelty of his approach. In his first paleontological study—an analysis of the succession of strata in the Italian Tertiary formations—he concluded that "these studies are sufficient not only to settle a dispute concerning the Italian Tertiary structure, but also to demonstrate the application of a numerical approach to characteristics of the fossil deposits in rock strata, that has so far not been considered."[18]

17. Sepkoski and Tamborini, "'An Image of Science'" (ref. 2).

18. Heinrich Georg Bronn, *Italiens Tertiär-Gebilde Und Deren Organische Einschlüsse: Vier Abhandlungen* (Heidelberg: Groos, 1831), 175 and 74.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

This was, for Bronn, the beginning of a new discipline, which he appropriately termed "paleontological statics." The need for better data prompted him to devote his career to assembling his massive taxonomic data compilation, which he believed would provide a firm basis for analyzing the patterns and regularities in the history of life. Whereas a number of other naturalists of Bronn's day—in disciplines ranging from botany to zoology to paleontology—had compiled numerical tabular accounts of species and genera, Bronn's approach was distinctive because it applied statistical analysis to a *causal* explanation of the data he collected—and this was the influence of his cameralist training in statics.

But Heidelberg cameralists generally did not employ pictorial summaries of data in the style of Playfair or von Humboldt—their visual culture was still firmly rooted in the tabular idiom. In his more mature paleontological studies—the 1849 *Index*, for example—Bronn introduced visualizations of his data patterns, like the one reproduced in Figure 3. This image is one of the earliest examples of what is now known as a "spindle" diagram, one of the most common visual devices in paleontology for depicting the duration and diversity of a higher taxon as a line of varying thickness (according to the number of constituent taxa at a given time). This particular diagram depicts the history of major groups of organisms throughout geological history, and effectively takes the form of a visual historical summary. We see, at a glance, the rise and diversification of the ammonites before their abrupt departure alongside the waxing and waning of the brachiopods, vascular plants, crustaceans, mammals, etc. The picture, then, tells a story about the history of life, and has become a standard tool in evolutionary biology—and especially paleontology—to this day.

What is significant about this image is not just that it reflects the incorporation of statistical visual practices to natural history, but that it also signals an important epistemic shift in thinking about—and visualizing—history itself. In describing the somewhat later innovations of Jevons in economic statistics, Maas observes that "before graphs could reveal economic phenomena and feature in economic explanations. . . . [h]istorical *events* had to be repackaged as *data*," something Maas labels "the timing of history."[19] This is precisely what Bronn, drawing on his cameralist training, saw was necessary for narrating broad patterns in the history of life. In the first place, historical "events" (i.e., the individual life histories of organisms) were aggregated and converted

19. Maas, *William Stanley Jevons* (ref. 12), 220.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

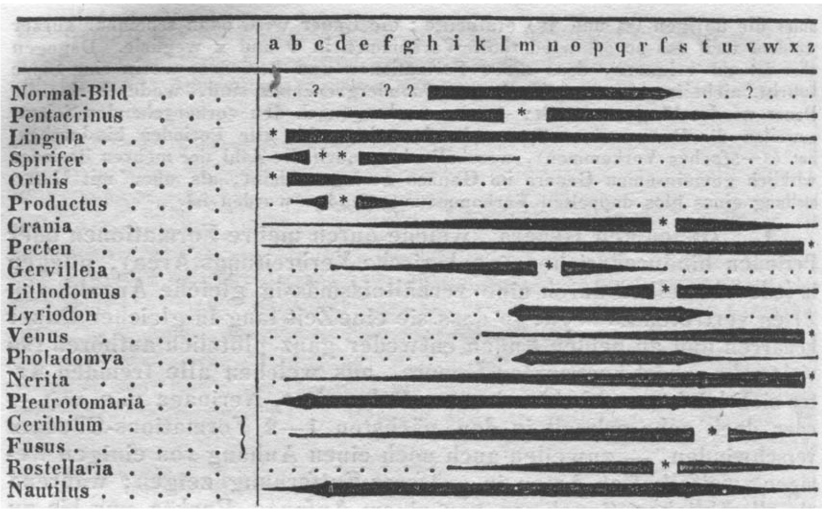

FIGURE 3: One of Bronn's spindle diagrams depicting the relative changes in diversity among different groups of animals throughout geological time. *Source:* H. G. Bronn, *Index Paleontolgicus, oder Übersicht der bist jetzt bekannten Fossilen Organismen* (Stuttgart: 1849), 778.

to data points representing taxonomic units that could be counted and treated statistically. Secondly, those data were projected as a temporal progression in visual space, "stabilized" not only along the horizontal (temporal) axis, but also on the vertical, which represented change in quantity. Indeed, it is only through the abstraction of individual events to commensurable data points, and the aggregation of those data, that it is possible to identify the regularities or "empirical laws" that Bronn hoped to discover—in this case, what Bronn would later describe as the tendency for major taxonomic groups to progressively diversify, reach a maximum point, and then slowly decline, only to be succeeded by similar cycles in other groups.[20]

20. H. G. Bronn, *Untersuchungen Über Die Entwickelungs-Gesetze der organischen Welt Während der Bildungs-Zeit unserer Erd-Oberfläche* (Stuttgart: Schweizerbart'sche Verlagsbuchhandlung, 1858), 489. Interestingly, the historiographic conventions implicit Bronn's statistical visualizations also resonate suggestively with contemporary philosophies of human history, such as works by G. F. Hegel and the English historian Henry Buckle, the latter of whom promoted a "statistical" approach to history in his *History of Civilization in England* (London: John Parker, 1857). Notably, Buckle's work received a great deal of attention in Germany, where it was read appreciatively by statisticians but often criticized for, as Prussian Statistical Bureau director Ernst Engel asserted, confusing statistical regularity with *law*. Engel's view—which is discussed in detail by Christine von Oertzen in this Issue—reflected an important contemporary debate about the

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

My purpose in highlighting Bronn's statistical visualization of natural history in the context of contemporary economic statistics is to give an example of what I believe "history of data" is most useful for: that is, illuminating the complex web of practices, influences, and contexts associated with the development of modern information cultures. I have not attempted to tell a heroic story of discovery here, but rather to suggest that behind a humble graph—Bronn's spindle diagram—is a complex story involving a series of important transformations in the way Western science has approached issues as disparate as quantification, history, visual language, and statistical abstraction. The establishment of conventions around the legitimacy of statistical inferences and the visual representation of aggregative temporal patterns has had a profound effect on the shaping of modern politics and culture, as well as science. Though few historians today would attempt a statistical explanation for major trends in human history, in the historical sciences—including the economic and social as well as the natural sciences—statistical language is pervasive. Although relatively little attention has been paid to the implicit historiography of those disciplines, it strikes me that a catholic approach to the history of data is an ideal means of addressing it more closely.

Another important lesson is that visualizations of quantitative data invoke the power of numbers without requiring advanced numeracy in the viewer. As Porter and others have pointed out, trust in numbers became a kind of *implicit* trust, and in visualizations the numerical data and their analysis are effectively black-boxed. Statistical visualizations make patterns in data visible, but they also obscure the phenomena that stand behind the data themselves. Bronn's spindle diagram may have made it easier for his reader to grasp the statistical narrative he hoped to convey, but unlike his data tables or fossil catalogs, the story the images tell is unequivocal. In the sciences, this has recently become a source of tension, where in a variety of disciplines stark divisions—often associated with status, credit, and access to funding—have emerged between

viability of statistics for historical reconstruction in natural history and other disciplines as well. In the decades following Bronn's death (in 1862), German-speaking paleontologists debated whether an aggregative statistical approach was valid for identifying historical patterns in the fossil record. Like Engel, paleontologists Melchior Neumayr and Karl Alfred von Zittel argued that statistical analysis could not identify causal relationships that are the proper foundation for establishing laws of nature, and urged instead a focus on reconstructing individual "snapshots" of life's history through careful study of individual fossil groups. See Hacking, *Taming of Chance* (ref. 4), 125–29; Marco Tamborini, "Paleontology and Darwin's Theory of Evolution: The Subversive Role of Statistics at the End of the 19th Century," *Journal of the History of Biology* 48 (2015): 575–612.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021

data "producers" (i.e., bench scientists) with deep understanding of phenomena and "analysts" (or "data jockeys") who are sometimes accused of glib disinterest in anything other than correlations and models.[21] In political and economic contexts the disconnect between data and visualization or aggregate summary is potentially more striking and perilous: in some examples of truly huge data (such as Google analytics of web activity) the data are so vast and so widely distributed that no person can claim to really know it. Trust in this kind of data amounts to trusting the algorithms that handle it for us. A genealogical approach to data history can help us better understand, if not escape, the implicit choices and compromises that have been made in seeing the world as data.

ACKNOWLEDGEMENTS

I would like to thank Soraya de Chadarevian and Ted Porter for the invitation to contribute to this special issue and the workshop that preceded it. Thanks also to the members of the initial workshop for thoughtful comments about my oral presentation, and to Soraya, Ted, Hallam Stevens and Christine von Oertzen for helpful comments on the written draft. Portions of this article have drawn on research conducted in collaboration with Marco Tamborini; I would like to acknowledge his important role in the larger project on which this essay is based.

21. See, for example, Bruno J. Strasser, "Data-Driven Sciences: From Wonder Cabinets to Electronic Databases," *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (2012): 85–87.

Downloaded from http://online.ucpress.edu/hsns/article-pdf/48/5/581/377932/hsns_2018_48_5_581.pdf by Max Planck Institute for the History of Science user on 15 June 2021