

Graph-based inductive reasoning



Marcel Boumans

Utrecht University School of Economics, Kriekenpitplein 21–22, 3584 EC Utrecht, Netherlands

ARTICLE INFO

Article history:

Received 3 February 2016

Received in revised form

18 April 2016

Keywords:

Expert judgment
Graphical method
Inductive reasoning
Measurement
Statistical method
Visual cognition

ABSTRACT

This article discusses methods of inductive inferences that are methods of visualizations designed in such a way that the “eye” can be employed as a reliable tool for judgment. The term “eye” is used as a stand-in for visual cognition and perceptual processing. In this paper “meaningfulness” has a particular meaning, namely accuracy, which is closeness to truth. Accuracy consists of precision and unbiasedness. Precision is dealt with by statistical methods, but for unbiasedness one needs expert judgment. The common view at the beginning of the twentieth century was to make the most efficient use of this kind of judgment by representing the data in shapes and forms in such a way that the “eye” can function as a reliable judge to reduce bias. The need for judgment of the “eye” is even more necessary when the background conditions of the observations are heterogeneous. Statistical procedures require a certain minimal level of homogeneity, but the “eye” does not. The “eye” is an adequate tool for assessing topological similarities when, due to heterogeneity of the data, metric assessment is not possible. In fact, graphical assessments precedes measurement, or to put it more forcefully, the graphic method is a necessary prerequisite for measurement.

© 2016 Elsevier Ltd. All rights reserved.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

1. Introduction

The twentieth century witnessed an exponential growth of social statistics. At its beginning, the magnitude of data was in the order of a few kilobytes; half way through that century it was in the order of megabytes, and at the end of the twentieth century it was in the order of gigabytes, thereby earning the label of Big Data (see Diebold, 2013); in other words, each 50 years the magnitude of social statistics increased by a factor 1000.ⁱ Despite this enormous expansion of data, the nature of the core problem of inductive inference remained the same: How to infer meaningful patterns from these masses of data when there is little or no theory to guide the inferences and there are not yet any standardized objective procedures to follow? Part of the answer is that one needs additional expert judgments.ⁱⁱ But then the subsequent question

is—when and in what manner are these expert judgments instrumental?

This article will discuss methods of inductive inference that are methods of visualizations designed in such a way that the “eye” can be employed as a reliable tool for judgment. I use the term “eye” as a stand-in for visual cognition and perceptual processing. The reason is that I will discuss literature from the beginning of the twentieth century in which the term “eye” was commonly used instead of the currently used concept of visual-cognitive system. The method of visualizations that will be discussed is the method of graphs, also called the method of curves, that was developed around 1900.ⁱⁱⁱ

In her position piece, Annamaria Carusi (2012) advocates that more attention should be paid in philosophy of science to the epistemological role of visualizations: “[I]t is necessary to understand how vision works embedded in *epistemic* contexts, as playing

E-mail address: m.j.boumans@uu.nl.

ⁱ Although these numbers are not so impressive as they are in natural science, where they are in the order of petabytes, the additional “big” problem is that social data are much more heterogeneous.

ⁱⁱ See Boumans 2015 for an extensive substantiation of this argument.

ⁱⁱⁱ Visual displays have of course various different roles in science, for example to make theories comprehensible, to present data sets, or to analyze data. This article focuses on the latter.

a crucial role in the formation of evidence for claims” (p. 107).^{iv} Notwithstanding the bulk of studies in cognitive sciences which has accumulated in the past few decades, attention to the visual as such is still scarce in philosophy of science today. According to Carusi, these studies challenge philosophy of science to re-think its position on three key distinctions: the qualitative/quantitative, the subjective/objective, and the causal/non-causal distinction. This article hopes to contribute to re-thinking the first two distinctions.

A good overview of studies in cognitive science on the use of visualization in scientific thinking and reasoning is provided by Hegarty (2011).^v Maria Hegarty summarizes some of the main advantages that visual displays afford for cognitive tasks: 1. External storage of information; 2. Organization of information; 3. Offloading of cognition on perception, and 4. Offloading cognition on action. This article will focus on the second and third tasks of visual displays. It will show that in the organization of information—here called visualization—the “eye” also plays a crucial role. The organization of information requires not only “vision to think,” that is, the offloading of cognition on perception, but also “vision to visualize.”

Halfway through the twentieth century, visualizations came to imply the involvement of computers, and were often called simulations.^{vi} Although in that period visualizations were only possible with analog computers,^{vii} in economics the digital computer was considered as an instrument of observation, supplying “a viewing equipment to the economist in a manner analogous to the microscope for biologists” (Shubik, 1960). According to Mary Morgan (2004, p. 363), who discusses Martin Shubik’s account of simulation, Shubik may have inherited this metaphor of the microscope from Oskar Morgenstern (1954). Morgenstern made this comparison with the microscope for cases in which one has to deal with an enormous amount of data for which there is no theory that could attach any meaning to data. Without theory, one would be “just looking” or “merely looking.” In an earlier stage of a science this may lead to data of a new kind. When the telescope and microscope were invented, “all that mattered was to take these wonderful new instruments and to look, to look practically anywhere. Some phenomena would turn up. Totally unsuspected, be they the moons of Jupiter or some tiny amoebae in a drop of water” (1954, p. 540).

While Morgenstern spoke of “just looking” when making observations with a microscope, Shubik noted rightly that, as every user of a microscope would admit, the metaphor of a microscope implies a “specimen” to be observed with the microscope, and that the setting up of such a specimen would require “a great amount of work” (Shubik, 1960, p. 908). In social science, however, these specimens are not material objects but

models of those things. Thus, where the biologist prepares slides so that they can “see” certain things with the microscope, economists prepare models so that the relevant parts of the world they specify can be “read” by the computer. And if the models prepared for the economists’ computer-as-microscope are not natural, they are, of course, artificial constructions, man-made. (Morgan, 2004, p. 364)

Both Shubik and Morgan observe that inductive inference from an enormous data set is not “just looking” at the “raw data” but it

requires a specific kind of preparation—a visualization—to have something to look at. Morgan calls this kind of preparation—rightfully in my view—modeling, in the sense of making a representation.^{viii}

The analogy between specimen and model also implies an analogy between the epistemic problems on both sides. As much as a specimen can lead to artefacts caused by the method of preparation, a model built to make a phenomenon visible may also create artefacts about it. The problem however is that the epistemic strategies to distinguish between valid facts and artefacts, such as control of possible confounding effects and systematic error, replicability, data reduction and calibration do not travel very well from the material world of specimens to the virtual world of models. Although the problem of artefacts is a relevant issue, it will not be further discussed in this paper because it will go beyond the topic of this paper and is already extensively discussed elsewhere, see for example Boumans, 2002 and Franklin, 1997.

To explore the preparation of “specimens” which enables the observation of social phenomena, it is useful to move to the period when these kinds of modeling strategies were discussed most explicitly. Firstly, I will discuss the general shared epistemology of the method of graphs around 1900. Second, I will discuss a specific and rich case in which the method of graphs has been used to infer meaningful patterns from the enormous amount of available data. This case study concerns how Warren M. Persons designed a “barometer” to inform businessmen and politicians about the “general business conditions.” The last section will discuss—as an illustration—a current case of pattern recognition that applies a similar strategy based on the epistemological advantages of the “eye” to organize information.

2. The method of graphs

Histories of the graphic method (Funkhouser, 1937; Hankins, 1999; Klein, 1995; Maas and Morgan, 2002) point to William Playfair as the originator of this method for social data. In the introduction of his (1796) *A Real Statement of the Finances and Resources of Great Britain*, Playfair mentions explicitly the advantage of using graphs and charts, which he called “lineal arithmetic”:

[I]t is to give a more simple and permanent idea of the gradual progress and comparative amounts, at different periods, by presenting to the eye a figure, the proportions of which correspond with the amount of the sums intended to be expressed. As the eye is the best judge of proportion, being more accurate and quicker than any other of our organs, it follows, that where-ever *relative quantities*, a gradual increase or decrease of any revenue, receipt or expenditure of money, or other value, are to be stated, this mode of representing it is peculiarly applicable, as it gives a simple, accurate, and permanent idea; it produces form and shape to a number of separate ideas, which are otherwise abstract and unconnected; for in a numerical table there are as many distinct ideas given, and to be remembered, as there are sums. The order and progression, therefore, of those sums, are also to be recollected by another effort of memory, while this unites proportion, progression, and amount, all under one simple impression of vision, and consequently one act of memory. (Playfair, 1796, pp. v–vi)

Although a century earlier René Descartes, one of the originators of analytic geometry, had stated that “imagination or visualization

^{iv} See Mößner, 2015 for a similar position.

^v I thank one of the anonymous referees for pointing me at the relevant literature.

^{vi} See Morgan, 2004 for an insightful discussion of simulations as a new “technology.”

^{vii} See Boumans 2012 for a treatment of analog computers as tools of intelligibility.

^{viii} This is more explicitly stated and explored in Section 5.iii of Morgan 2012, which carries the title “Specimens = Models.”

and in particular the use of diagrams has a crucial part to play in scientific investigation” (quoted in Beniger and Robyn, 1978, p. 2), it would take until the nineteenth century before the method of graphs came to be more generally accepted as a method of investigation in the social sciences. As a matter of fact, this was mainly due to the rise of statistics and the emerging questions about their analysis, that is to say, questions about the proper methods of inductive inference.

It was William Whewell’s (1858) *Philosophy of the Inductive Sciences*, which discusses the Method of Curves as one of the inductive methods that was influential in the acceptance of the use of graphs as tools of investigation. In aphorism 44, Whewell states that its “efficacy”

depends upon the faculty which the eye possesses of readily detecting regularity and irregularity in forms. The Method may be used to detect the Laws which the observed quantities follow; and also, when the Observations are inexact, it may be used to correct these observations, so as to obtain data more true than the observed facts themselves. (Whewell, 1858, p. 202)

Or in other words, “that order and regularity are more readily and clearly recognized, when thus exhibited to the eye in a picture, than they are when presented to the mind in any other manner” (p. 204).

Relevant to the discussion of the epistemological power of the Method of Curves is the second part of the aphorism. Graphs—visualizations—are needed to correct for observational errors.^{ix} The second efficacy of the Method of Curves implies that this method leads to more accurate results; the Method “may be made to give us a formula with great accuracy. The Method enables us to perceive, among our observations, an order, which without the method, is concealed in obscurity and perplexity” (p. 205). Observations are “imperfect”, that is to say they contain observational errors. If these observations were plotted in a graph, it would show not a smooth and flowing curve, but a broken and irregular line. These irregularities are assumed to indicate these observational errors; smoothness is considered to be an indication of the working of the underlying laws. “The regular curve which we thus obtain, thus freed from the casual errors of observation, is that in which we endeavor to discover the laws of change and succession” (p. 207).

By this method, thus getting rid at once, in a great measure, of errors of observation, we obtain data which are *more true than the individual facts themselves*. The philosopher’s business is to compare his hypotheses with facts, as we have often said. But if we make the comparison with separate special facts, we are liable to be perplexed or misled, to an unknown amount, by the errors of observation; which may cause the hypothetical and the observed result to agree, or to disagree, when otherwise they would not do so. If, however, we thus take the *whole mass of the facts*, and remove the errors of actual observation, by making the curve which expresses the supposed observation regular and smooth, we have separate facts corrected by their general tendency. We are put in possession, as we have said, of something more true than any fact by itself is. (Whewell, 1858, p. 207)

To analyze exactly what is the epistemological power of the Method of Curves, it is helpful to compare this method with the

method of graduation developed in the same period.^x Assume that one wishes to determine a variable x , and the value of x has to be inferred from a set of available observations y_i ($i = 1, \dots, n$), which inevitably involve observational errors ε_i . One can then represent this relation between these observations, y_i , the target variable, x , and error, ε_i , by the following equation:

$$y_i = x + \varepsilon_i. \quad (1)$$

The method of graduation is a numerical combination of the values of the observations, $M[y_i]$, such that the estimate, \hat{x} , is as accurate as possible. Hence, the estimate $\hat{x} = M[y_i]$ of a target variable x is accurate when \hat{x} is very close to x .

As a result, to determine x a model, denoted by M , has to be specified, for which the observations y_i function as input and \hat{x} , the estimation of x , functions as output:

$$\hat{x} = M[y_i]. \quad (2)$$

Substitution of the first equation (Eq. 1) into the second equation (Eq. 2) shows what should be modeled (assuming that M is a linear operator, which is usually the case):

$$\hat{x} = M[x + \varepsilon_i; \alpha] = M_x[x; \alpha] + M_\varepsilon[\varepsilon_i; \alpha].$$

From this algebraic exercise one can see that a necessary condition for \hat{x} to be an estimate of x is that model M must be a representation of how the observations are related to the true value of x . Therefore we need both a representation of the target variable, M_x , and a specification of the observational errors, M_ε .

To show what the problem of accuracy entails, we first split the estimation error $\hat{\varepsilon}$ in two parts:

$$\hat{\varepsilon} = \hat{x} - x = M_\varepsilon + (M_x - x).$$

To explore how this estimation error, that is, the inaccuracy, is dealt with in the first instance, it is helpful to compare this measurement error with the mean-squared error of an estimator as defined in statistics:

$$E[\hat{\varepsilon}^2] = E[(\hat{x} - x)^2] = \text{Var}\hat{\varepsilon} + (E\hat{x} - x)^2.$$

The first term of the right-side of this expression of the mean-squared error, the variance of the measurement error, is a measure of precision and the second term is called the bias of the estimator. Comparing the mean-squared error with the estimation error, one can see that the reduction of the error term M_ε is aimed at by attempts to obtain precision. Hence, one part of (but not equal to) accuracy is precision. To deal with precision, procedures have been developed, like the Method of Mean, which contribute to the mechanical objectivity of the estimate. The other part of accuracy, however, is much more problematic. The reduction of the bias ($M_x - x$) requires theory which may inform about the underlying laws. But for the cases we are dealing with here, such theory does not exist (yet); moreover the inductive inferences we are discussing here are meant as a first step to arrive at such theories.

To investigate how to reduce the bias term, it will prove instructive to look at a field where this problem is most prominently dealt with, namely actuarial science. The general approach to dealing with observational errors in actuarial science is graduation or smoothing, which Morton D. Miller (1946, p. 4) defined as “the process of securing from an irregular series of observed values of a continuous variable a smooth regular series of values consistent in a general way with the observed series of values.”

^{ix} This is not about the problem of detecting outliers. Because these data points lie out far enough, trained researchers can detect them by looking at the data alone, and so do not need a visual representation as an aid. Whewell is referring to the determination of the size of the error in *each* data point.

^x The following brief exposition is based upon chapter 3 of (Boumans, 2015).

Graduation is based on the view that there is an underlying law that produces a smooth, regular and continuous sequence of values, but that all kinds of disturbances have turned this sequence into an irregular one.^{xi} The irregularity represents deviations from the true values, and thus the revised, “graduated,” sequence should be taken as a representation of the underlying law. However, the only empirical knowledge that we have about these laws are these observations.

Since no law of mortality, in the sense of a physical law, is known to us, nor is one likely to be discovered, we have no way of knowing *a priori* what the basic pattern of mortality is. We must therefore rely on the information supplied by observations of the rates of mortality actually being experienced. (Miller, 1946, pp. 1–2)

However, for the analysis of these observations there are too many possible methods of graduation available to choose from; the choice of the most appropriate graduation method is under-determined by the observations. To come to a definite choice additional assumptions are needed, such as one about the assumed underlying law, but, as Miller noted: “The theoretical reasons justifying graduations are those upon which this assumption rests” (p. 5). Therefore it is too much to expect that the errors in the observations will be completely eliminated. There will always be a “residual error.” For that reason Miller suggests that we think of the graduated series as a “representation of the underlying law rather than as the law itself” (p. 5), so we are aware that in principle various different representations are possible.

But graduation is not only characterized by smoothness. The other “essential quality,” according to Miller (1946, p. 5), is “fit, or consistency with the observed data.” But these two different qualities of smoothness and fit are “basically inconsistent.” Improving one is at the cost of the other. Therefore, “any graduated series must of necessity follow a middle course between optimum fit and optimum smoothness; it must represent the result of a compromise between the two” (p. 5). There exists, however, no standard for this “compromise,” and therefore must be left to the judgment of the graduator: “a graduation method must allow the graduator some latitude in choosing the relative emphasis to place on smoothness and fit in the graduated series” (p. 5).

It was Edmund T. Whittaker (1922) who developed a “new method of graduation” that captured both qualities. To do this he formulated the problem in terms of probability, so his graduation method was designed such to obtain the most probable values of x . He defined the measure of smoothness as $S = \sum_i (\Delta^2 \hat{x}_i)^2$ and fit, which he called the “fidelity of the graduated to the ungraduated values” (p. 65) as $F = \sum_i (\hat{x}_i - y_i)^2$. Using Bayesian reasoning, he showed that the most probable value of \hat{x} is that for which $\lambda S + F$ is a minimum. According to Whittaker, one of the advantages of this method is its “elasticity” due to the freedom of the choice of λ :

A satisfactory method of graduation *ought* to possess such elasticity, because the degree to which we are justified in sacrificing fidelity in order to obtain smoothness varies greatly from one problem to another. (Whittaker, 1922, p. 73)

There is of course another choice that has to be made, and that is the choice of the degree of smoothness z . As Miller (1946)

emphasized in his textbook, “graduation does not have a single solution” (p. 7). It depends upon the choice of the method, upon a choice of how much fit and how much smoothness, the field of application, but also “upon the skill and experience of the graduator” (p. 7). Moreover, the choice of the method will rest upon:

- (i) The purpose to which the graduated table is to be put;
- (ii) The form in which the data are given and their general characteristics;
- (iii) The extent of the data; and
- (iv) The experience, technical knowledge and preferences of the graduator.

These factors, of course, are not fully separable but rather interdependent. (Miller, 1946, p. 54)

In other words, there is always an experienced graduator needed to make decisions wherever there are neither standardized procedures nor theory available.

How are these decisions made? This paper argues that these decisions could be based on the judgment of the “eye.” More recent evidence for this claim is how Whittaker’s method of graduation is used in modern macroeconomics, where the minimum of $\lambda S + F$ is known as the Hodrick–Prescott filter to separate the cyclical component of a time series from the unfiltered data in order to arrive at the growth component (Hodrick and Prescott, 1997). In macroeconomics a given time series y_t is considered to be the sum of a growth component g_t and a cyclical component c_t : $y_t = g_t + c_t$, where the growth is considered to be the smooth component and the cycles the irregular ones. Without any explication, it is assumed that the “measure of the smoothness” z is two. The problem then is the choice of the value of parameter λ .

We found that if the time-series are quarterly, a value of $\lambda = 1600$ is reasonable. With this value, the implied trend path for the logarithm of real GNP is close to the one that students of business cycles and growth would draw through a time plot of this series, as shown in Chart 2. (Kydland and Prescott, 1990, p. 9)

The current general consensus is that expert judgment is the cause of “bias.” Theodore Porter (1995) and Lorraine Daston and Peter Galison (2007) have shown that the rise in dominance of the epistemological virtue of mechanical objectivity can be understood as an insistent drive to reduce this subjective bias. This article, however, argues that in addition to mechanical procedures, expert judgments are needed to reduce bias, as they are part of the solution of bias.^{xii} Whewell showed how this kind of judgment could be involved by using the Method of Curves, that is, by defining unbiasedness as a specific feature of a visual display of the data, namely smoothness, for which the “eye” is an effective tool for assessment.

Inductive reasoning appears to be a combination of mechanical methods and expert judgments and finding the most appropriate balance between these two aspects. It is a balance between bias and precision, where precision is dealt with by an objective method, for example the Method of Mean, and bias is reduced by expert judgment. More specifically for graduation, this kind of inductive reasoning is based on a balance between fit and smoothness, where fit is dealt with by, for example, least squares, and smoothness

^{xi} Note that “smoothness” is a feature of a visual display. Moreover, whatever method of graduation one designs, one can only “test” this method for its aimed level of smoothness by looking at a graph that shows the graduated series.

^{xii} See Boumans 2015 for an extensive discussion of this claim.

involves expert judgment. The example of smoothness shows very clearly that this kind of judgment involves visualization (“preparation of a specimen”) and vision (“looking”).

While Whewell in a more abstract way discussed the usefulness of the Method of Curves, it was Arthur L. Bowley who offered the most professional treatment of the graphic method (see [Morgan, 1997](#), p. 50). His *Elements of Statistics* (1907), first published in 1901, was a textbook based on his lectures given at the London School of Economics, and his *An Elementary Manual of Statistics*, first published in 1910, was designed as a practical manual on how to do statistics, including the “use of diagrams.” [Bowley \(1907, p. 143\)](#) describes the graphic method as one of the “two main methods of elementary statistics,” the other being the method of averages. Both methods are appropriate as “when we deal with large and complex masses of figures we are unable to grasp them in their entirety, however clearly they may be tabulated” (p. 143). According to Bowley, averages “afford the best summary of the whole group in question that the mind can grasp”, and “the main use of diagrams is also to present large groups of figures so that they shall be intelligible in their entirety, and the test for all diagrams is that the diagram as drawn should afford the best view of the series or group of figures that the eye can appreciate” (p. 143). In contrast with Whewell, however, Bowley saw a diagram as “less essential” than averages, because the latter represent “true types of the quantities which are being measured; and by their use alone are further comparisons of complex groups made possible” (pp. 143–4). In my view, Bowley was mistaken herein. As I will show in the next section, when discussing a specific case of business cycle analysis, the method of average alone is insufficient for the comparison of time series.

According to [Judy Klein’s \(1997, p. 17\)](#) history of time series analysis, “the golden stage for graphs” was that of empirical investigation in the late nineteenth and early twentieth centuries. But this “geometrical reasoning of statistical method” involved a “search for form” and was initially seen in opposition to the method of mean; see for example Bowley’s remarks on the difference between the graphic method and the method of averages. But Klein shows that the smoothing process became a “means to reconstructing form consistent with the science of means” (p. 101). In other words, smoothing techniques were used to “prepare” the specimen such that measurements could be undertaken, as will be shown in the next section.

3. Construction of a business barometer

To explore in more detail a distinct practice of visualization, I will now discuss Warren Persons’ work of developing a graphical display of an index of business conditions, which he called a “business barometer.” This business barometer is a prime example of the preparation of data for a particular application of the science of means-correlation analysis. As with effective use of the tool of graduation, the “eye” was fundamental to the process of constructing a business barometer; in other words, this practice is a clear illustration of “vision to visualize.” Besides it being a practice of visualization, I have also chosen to discuss this particular practice because it is so well documented.

As [Walter Friedman \(2014, p. 136\)](#) recounts, Persons published a lengthy series of articles on the methodology of designing and making the barometer in the *Review of Economic Statistics*, which was founded in 1917 with the purpose of publishing this index. He devoted the first three regular issues and four of the monthly supplements to an extensive and detailed discussion of the methods. He produced a total of two hundred pages of methodological articles in 1919 alone. Thanks to these efforts, we now have a very good insight of this very specific practice of building graphs for

the purpose of economic analysis, that is, of the “great amount of work” of building of a “specimen.” The method that Persons applied for developing an index of business conditions was truly a method of graphs. The two first issues contained 70 (37 and 33 respectively) charts in which he analyzes and compares time series.

Persons (1878–1937) was born in West De Pere, Wisconsin.^{xiii} He received a bachelor of science degree from the University of Wisconsin and worked as a mathematics instructor there from 1901 to 1906. He then became an assistant professor of economics and finance at Dartmouth College, where he taught for about five years. In 1913, he became a professor of economics at Colorado College, in Colorado Springs, and began writing a PhD dissertation on the distribution of wealth and income. He completed his doctorate in economics three years later at the University of Wisconsin. His work attracted attention and eventually led to an appointment at Harvard in 1918.

In 1917, Harvard University had organized the Committee on Economic Research to study economic statistics and improve the scientific quality of economic investigation. The committee hired Persons to lead its statistical work. Persons’ primary responsibility was to edit the committee’s quarterly journal with monthly supplements, the *Review of Economic Statistics*, which first appeared in January 1919. As editor of the *Review*, Persons was also in charge of the formation of the index of business conditions, a “business barometer” on which ideas Persons had published an *American Economic Review* article in 1916.

In his presidential address read at the Annual Meeting of the American Statistical Association in 1923, [Persons \(1924, p. 2\)](#) emphasized “necessity of the accumulation of statistics of the complex world of affairs in which we are immersed and the equal necessity of the development of special methods, different from those of the exact sciences, for summarizing these data.” The summary of the data he developed was an index, a “barometer,” which was built by a combined application of the graphic method and the correlation coefficient.

To develop an “index of general business conditions,” Persons first formulated two principles, one of composition and the other of decomposition of a time series. A time series is a sequence of data, called “items,” observed at successive points in time spaced at uniform time intervals. The principle of composition is that “isolated items of statistical series [...] can have no significance by themselves. Only by a comparison of items over a period of time can we ascertain their meaning” ([Persons, 1920, p. 39](#)); in other words, the time series as a whole and not the individual items are the objects that will be compared with each other. The principle of decomposition is that “items pertaining to widely separated periods cannot, however, be used in their crude form” (p. 39). Each monthly item is considered to be a “composite” of four elements: a secular trend, seasonal variation, cyclical fluctuations, and a residual factor.

The *secular trend* is the regular increase, “according to some principle,” over the whole period under consideration. According to Persons, the secular trend is a growth element, “a normal change,” dependent upon population growth and the development of industry, “just as there is a normal change in the physical or mental status of a growing child” (p. 39).^{xiv} This component is determined by least squares regression analysis. The “chief” problem was the

^{xiii} This brief biography of Persons and history of the Harvard Economic Service is distilled from [Friedman 2009](#) and [Friedman 2014](#).

^{xiv} In the late nineteenth century both Karl Pearson and Udney Yule used the term “secular variation” to refer to non-cyclical, trend-like variation, which they probably took from astronomy, geology and meteorology where it meant non-cyclical or persisting for a long time, e.g. a century. (Judy Klein in a personal communication).

determination of the period for which the trend has to be calculated. For such a period the conditions have to be “homogeneous.”

The question is, how long a time is necessary to show significant trend; how long a time can safely be taken consistent with the requirement of homogeneity of data? To this question no absolute answer can be given. (Persons, 1919a, p. 9)

Persons gave two criteria to determine this period: First, the minimum length of this period should be at least as long as a “full business cycle of prosperity and depression. This indefinite period is the minimum for which it may be determined” (p. 10). This minimum period was indefinite because at that time it was not yet determined how long the period of a business cycle was. Second, the maximum length of the period is determined by the homogeneity of the data, but “since there are no criteria for deciding precisely what degree of homogeneity is requisite for all series, this maximum also is indefinite” (p. 10). To determine this period, one has to look at the charts of the time series and to see where changes of trends are clearly visible.

The *seasonal variation* is the movement of the items within a year, attributed to the change of the seasons. “There is a seasonal change in various lines of business activity just as there is a seasonal change in temperature or rainfall” (1920, p. 39). The seasonal variation of an item for any month is given by an index which expresses the “normal” for that month as a percentage of the monthly average of the year. An obvious method to determine this “normal”—as suggested by calling it a “normal”—would be to take the average for a specific month during the ten preceding years. But this method will not be satisfactory because this average will not isolate the seasonal component and therefore will capture the cycle components as well as the residual factor. To find a more appropriate way to isolate the seasonal variation, the graphic method had to be used. Although “none of the graphs for the series examined showed a characteristic shape indicative of seasonal variation” (1919a, p. 22), the graphs did show that the month-to-month movements were systematic and revealed seasonal fluctuation. And so, Persons concluded that the determination of seasonal variation should be based upon percentage change from one month to the next. Because one has to deal with extreme values and only a small number of observations, Persons chose to take the median instead of the mean as the most appropriate estimate of the seasonal component.

The *cycles* are the components secured by removing from the actual items the secular trend and the seasonal variations, and expressing the result in terms of “comparable units.” These units are the standard deviations of the respective series. As a result, the cycles are the percentage deviations of actual items from secular trend corrected for seasonal variation, divided by the standard deviation.

The residual element includes

all sporadic developments which affect individual series, or widespread changes due to momentous occurrences, such as wars or national catastrophes, which affect a number of series simultaneously. Thus pig-iron production may take a sudden slump if a strike occurs, railway gross earnings may drop because of unusual storms or floods; trading on the stock exchange may be greatly affected by a court decision. (Persons, 1920, p. 39)

There was no general rule possible to specify this highly idiosyncratic component. The specification depends on the knowledge of the background history of the data. For example, a marking event

was the outbreak of the “Great War” (World War I). Because of its unsystematic character, the “residual element” was not eliminated from the cycle component. Hence, the resulting statistics used in constructing the index consisted of both the cycle component and this residual element.

Some fifty time series were considered; of these, twenty series were “corrected” according to the above procedures and then compared, because these series had “well marked cyclical fluctuations or wave movements, connected with the ebb and flow of business activity” (Persons, 1919b, p. 111). The twenty series showed similarities and differences. The elapsed periods “from crest to crest or trough to trough” were approximately the same, but the times at which crests or troughs of the waves of the several series occurred were generally not the same. For the index to be informative about the general business conditions, the series had to be clustered in groups for which the wave movements were somehow similar and simultaneous, so that an order of the sequence of these grouped series of fluctuations became visible. The grouping of these series was not determined by comparing the dates of the maximum and minimum items (the dates of the crests and the troughs), but by comparing the series in their entirety. Next, the “synchronous items of the series in each group” were averaged so that three “series of synthetic group indices” were secured. These three series of indices “epitomize the business situation” (p. 111) and were presented graphically by three curves in a chart, see Fig. 1.

The criteria for grouping the series were, thus, similarity and simultaneity of the cyclical fluctuations. The method of “measuring the correspondence or lack of correspondence between the fluctuations” that was adopted, was the “obvious procedure” to plot the graphs and to examine visually these graphs for similarity and dissimilarity (p. 120). To “facilitate” this visual examination, the charts were drawn on “translucent tracing cloth” and a box was constructed “with a glazed top illuminated by electric lights placed within the box” (see Fig. 2). One cycle chart was placed upon another and shifted till a best fit was reached of one curve to the other.

The twenty cycle charts were compared in this way, making 190 comparisons by each of three “independent observers.” Each observer recorded the results as follows:

- a) Degree and sign of correlation: high, moderate, low; + or –.
- b) Direction and extent of lag in months for the maximum correlation during the entire period.
- c) Characterization of the consistency of lag for maximum correlation: excellent, good, fair, poor. (Persons, 1919, p. 121)

The judgments of the three observers were in the main consistent, but, as Persons reported, there were also differences of opinion. The reason for disagreement is that the observer’s “personal equation, preconceived notions, or theoretical bias influence his conclusions” (p. 121). The observer’s judgments, therefore, were considered to be “first approximations”, which “need to be checked up by a more objective method of measuring correlation [...] of a coefficient of correlation, which will enable us to decide” (pp. 121–2). The three-term classification (high, moderate, low) could only indicate the degree of correlations very roughly (see Fig. 3); therefore the correlation coefficients had to be computed before a final decision could be made, “decisions concerning the nature and degree of correlation between pairs of series were actually based upon graphic comparison as a first approximation and the coefficients of correlation as a second approximation” (p. 128).

Comparisons of the series over the illuminated box was an essential part of the procedure. It showed that the cycles of the several series are related but the wave movements are neither

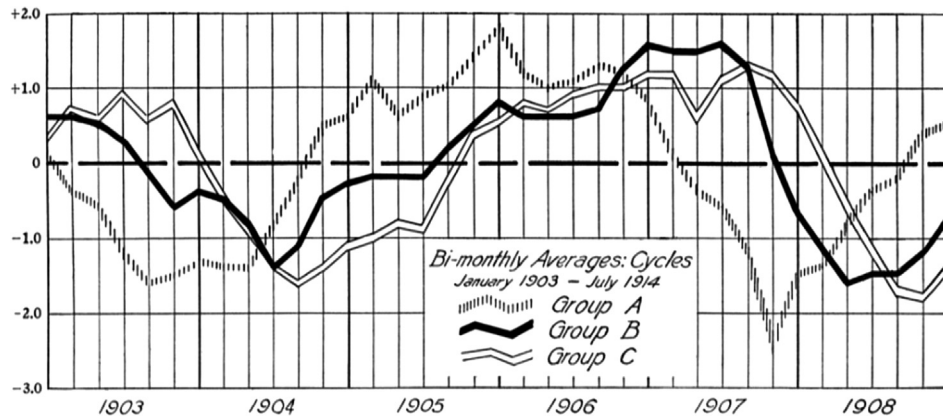


Fig. 1. A–B–C barometer.

(Source: Persons, 1919b, p. 112. Courtesy of the President and Fellows of Harvard University and the MIT Press.)

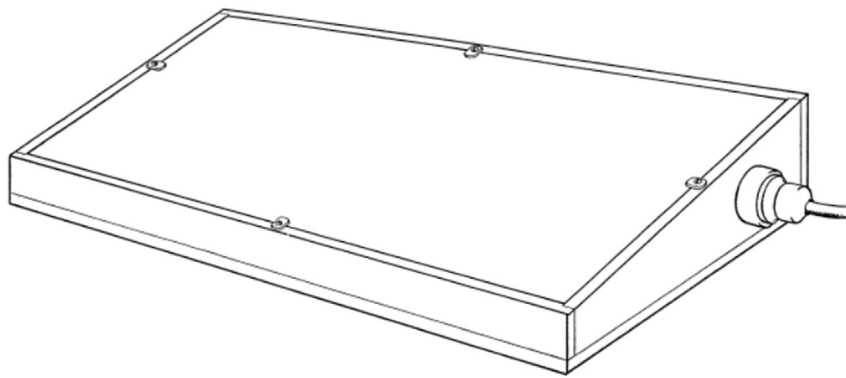


Fig. 2. Light box for comparing cycle charts.

(Source: Persons, 1919b, p. 121. Courtesy of the President and Fellows of Harvard University and the MIT Press.)

synchronous nor always positively correlated; that there are various degrees of correlation and lag; that the waves in the charts do not occur simultaneously but consecutively; that for some pairs the lag is approximately constant, for other pairs variable. (Persons, 1919b, p. 128)

4. Graphs versus correlation

According to Persons, the deployment of three observers was needed “in order to decrease the number of coefficients of correlations computed” (Persons, 1919b, p. 129). In a footnote, Persons explained how many calculations would be needed if the comparisons were made on the basis of comparing correlation coefficients alone: Because 20 series are compared which each other, there are $\binom{20}{2} = 190$ pairs for which correlation coefficients have to be calculated. If three coefficients are computed for each pair, 570 coefficients would be found. Since there are 138 items in

each series a total of 78,660 products would be required—“a considerable task to compute, record, and add” (p. 129).

But, even if Persons had had the availability of current computer technology, he still would have needed the three observers for deciding which times series is similar to which other one. The graphic method is more than simply “a necessary preliminary to taking correlation measures; it helped to decide what correlations to calculate. It was an early form of specification searching, appropriate to the pre-computer age, when calculation was very time-consuming” (Morgan, 1997, p. 73). To show this I will build on Klein (1997), who discusses extensively the distinction between “statistical visions” in “logical time” and those in “historical time,” and an article by Morgan (1997) which discusses the graphic method in the context of searching for causal relations in economic statistics. Although the latter article was written on a more restricted topic of comparing the graphic method with the correlation coefficient in the context of causal inference, it also sheds some light on my more general question of inductive inference to meaningful patterns, and whether this inference can be done by standardized mechanical procedures alone.

STANDARD 1. MONTHLY BANK CLEARINGS OF NEW YORK CITY COMPARED WITH										
New York Clearings	Production of Pig Iron	Outside Clearings	Bradstreet's Prices	Imports	Building Permits	R.R. Gross Earnings	Shares Traded	Unfilled Orders U. S. S. C.	Bus. Failures (Bradstreet)	
I	2	3	4	5	6	7	8	9	10	11
a.	a. High+	a. Moderate+	a. High+	a. Moderate+	a. Moderate+	a. Moderate+	a. High+	a. High+	a. Moderate-	
b.	b. 2 mos. lag	b. 2 mos. lag	b. 6 mos. lag	b. 3 mos. lag	b. Concurrent	b. 6 mos. lag	b. Concurrent	b. 2 mos. lag	b. 4 mos. lag	
c.	c. Good	c. Fair	c. Fair	c. Fair	c. Fair	c. Poor	c. Good	c. Good	c. Fair	

Fig. 3. Result of comparison of charts by three observers.

(Source: Persons, 1919b, p. 184. Courtesy of the President and Fellows of Harvard University and the MIT Press.)

The correlation coefficient originated in biometrics and was designed for the treatment of biometric data, which means particularly a historical data. The question was whether this tool could also be used to analyze time-series, the “time-correlation problem” as it was called (Klein, 1997, p. 222). Klein (1997, p. 224) lists three characteristics that the early biometric studies have in common but which were not shared by the studies of time series:

1. The co-relationship was between the same organ of different generations or different organs of the same organism. Thus the x and y variables were measured in comparable units and although they were samples drawn from different populations, they were organically related.
2. The observations for each variable were taken from a cross-section of the population at one point of time. Although the goal was to study heredity and evolution, the observations comprised a sample from a static, single population.
3. The observations for each variable usually displayed a “normal,” bell-curve frequency distribution similar to the law of error. Similarly, the correlation table of paired variables yielded a near-normal bell-shaped surface.

This time-correlation problem was extensively discussed by Persons (1919b) in a nine-page-long section particularly devoted to it. The correlation coefficient was originally developed as an aid to study relationships between features, like size of biological objects such as organs. But clearly sizes of organs are different in nature from items of time-series. Persons listed the following “essential differences”:

1. The items of time series must be defined for a selected time unit; they measure one form of activity of industrial society, whereas biological measurements are individual observations of objects whose measurements are independent of time.
2. Because they are defined for units of time, the items of time series are *ordered* in time; each item has a definite position with respect to the other items. The way to compare the items of two such series is to set up a relationship in time; that is, to pair concurrent items, or those definitely related in time. [...] There are various options possible in pairing items. On the other hand, when we are considering the correlation between the measurements of organs [...] there is a *unique* pairing of items determined by the problem. [...] the concept of lag is thus peculiar to time series.
3. Because they are ordered in time, the adjacent items in time series may be determined by overlapping or persistent causes; a succession of items of similar sizes is the rule rather than the exception. Adjacent measurements of organs, on the other hand, are not so connected (Persons, 1919b, p. 132).

The main problem was the problem of pairing. The organs may be in the same or in different individuals (e.g. parent and child). The pairing is determined by a “complex,” that is, “constituted by a natural or artificial tie of any kind, but the tie is to remain the same for every complex, whether it be the result of mating or parentage, or from any physiological or social relation” (Pearson quoted in Persons, 1919b, p. 131). Or, as Persons emphasized, “correlation is defined mathematically by any constant, or series of constants, which determine the above function [complex]” (p. 131). The problem of the time series which represent historical processes is that one cannot assume that these series represent “constants” in a similar way. This may be possible if the historical conditions across a period were constant, but this is never the case.

As Klein (p. 229) argues, “correlation analysis begged the question of which of the several components of the time series were to be correlated.” Persons saw the time-correlation problem to be solved by decomposition, that is isolating the different components and correlating only similar components, here the cycle components. “If our problem is to ascertain the relationship between two series ordered in time it is of little avail (or actually misleading) to compute the coefficient of correlation from pairs of the actual items” (Persons quoted in Klein, 1997, p. 229).

As the conditions change constantly, the graphic method is “a necessary preliminary to taking correlation measures; it helped to decide what correlations to calculate” (Morgan, 1997, p. 73). To be more precise, the correlation coefficient is a “reliable method to *measure* correlation” (Persons, 1910, p. 294; emphasis added), whereas the “charts do not answer the questions [of measurement]” (p. 294). “The graphic method of comparing fluctuations is well enough as a preliminary” for these measurements (p. 294).^{xv}

The correlation coefficient never superseded the graphic method. According to Morgan (1997, p. 74), the reason is fairly simple:

A correlation is easy, often useful, but many times a senseless thing. In many cases, it must be confessed, the correlation tells very little. [...] The correlation coefficient is informative only about the overall relationship; it does not allow you to retrieve any knowledge about what has happened to the two individual series over extended periods of time of the sort that can easily be gained from visual inspection of the time-series graph. Correlation does not tell you much. It does not tell you the history of the variable or help you to explain what has happened, nor even help you predict a time series; it is a complementary, not a replacement, technology.

The graphic method and the correlation coefficient differ from each other in the kind of information they provide: the graphic method provides *topological* information and the correlation coefficient provides *geometric* information. Topology and geometry are both mathematical theories of space; the essential difference is that geometry—as the name indicates—implies a metric to describe a space, whereas topology does not.^{xvi} One could also say that geometry describes a space in terms of quantities and topology does so in terms of qualities. A metric is a nonnegative function of two elements of a set, describing the “distance” between these two elements.^{xvii} As such, the correlation coefficient squared can be seen as the “distance” between two time series. In this sense, a metric is “one-dimensional”—it reduces the multi-dimensional relationship between two objects to only one dimension, namely its distance expressed as a number; it actually destroys the information about the other non-metric differences. A topological property of an object is a property that is preserved through deformations, twistings, and stretchings (but not tearing) of that object. (Topology therefore is sometimes called rubber-sheet geometry.) So two different time series can be similar with respect to certain topological properties, even though their shapes have different sizes for different time periods. So, even though two time

^{xv} Morgan (1997, p. 70) seems to suggest that for the purpose of measurement Persons (1910) was arguing here that “the graphic method had to give way to correlation,” but in my view Persons has never considered the graphic method to be a method of measurement.

^{xvi} I try to develop here a language to indicate more precisely the difference between the graphic method and the correlation coefficient. As such, geometry is considered here to be a “metric topology.”

^{xvii} This function has only to satisfy the triangle inequality ($g(x, y) + g(y, z) \geq g(x, z)$) and is symmetric ($g(x, y) = g(y, x)$).

series may be stretched differently, their topological properties can be the same.

The “eye” is a very good tool for assessing topological properties, but not so good as a tool of measurement. I define measurement in a very general way as the assignment of numbers to a property of a phenomenon according to a specific rule. Therefore, this rule should include a definition of a metric. The eye is a reliable tool for assessing differences, but not for assigning numbers to these differences in a reliable way, or in other words for assessing consistently what the distances are of these differences.^{xviii} Bowley (1907) spoke in a similar way about the ability of the “eye” to assess topological properties, which he called the “optical power”: “The eye can judge—(1) Distances; (2) ratios; (3) angles. [...] the eye is a fairly safe judge of distances; [...] The eye can also judge differences quickly” (p. 148). Similarly, he mentioned that “the accuracy with which the eye can make such measurements is not great” (p. 148).

The difference between the eye and a ruler can be clarified by taking, for example, a caricature of a famous figure such as a current president or prime minister. To give its intended message (e.g., whether the person is a weak or strong leader), the caricature will deform, twist, and stretch the original picture in such a way that it shows these features (e.g. a small or wide chin respectively). Irrespective of the deformations, the “eye” will have no problems recognizing the intended person, but a method based on “distances” alone will find it very hard to recognize an existing, real person.

5. Pattern recognition and machine learning

A century later, matters have not changed in principal. Notwithstanding the invention of the computer and the development of computer programs to analyze big data—called machine learning—one can still state that “computers do not perform as well as humans in visual recognition tasks” (von Ahn et al., 2008; see also; Puthala and Agarwal, 2011).^{xix} A core concept and tool in machine pattern recognition is the “kernel” which is nothing more than a distance function, a metric, that measures the similarity of two graphs (or “any two vectors in input space,” see Bishop, 2006). But distance functions are not sufficient for pattern recognition when the patterns are really messy.

An evidential illustration for this claim are reCAPTCHAs. reCAPTCHAs are machines that build upon the concept of CAPTCHAs. A CAPTCHA is a Turing test on the World Wide Web to determine whether a user is a human or computer. The acronym stands for Completely Automated Public Turing test to tell Computers and Humans Apart. A typical CAPTCHA is an image containing several distorted characters that appears at the bottom of Web registration forms. Users are asked to type the wavy characters to show they are human. Current computer programs cannot read distorted text as well as humans can, so CAPTCHAs act as sentries against automated programs that attempt to abuse online services.

But CAPTCHA texts are machine-generated and therefore can and will be decoded at some stage by another machine. The reason for this is that the artificial distortions of characters in CAPTCHA come from a limited (and usually simple) distribution of possible transformations that remain readable to humans. Therefore, it is feasible to build machine-learning algorithms that, after some “training,” can recognize the distorted characters. When decoded,

one could shift to another type of machine that produces texts that cannot be decoded yet, until again it is decoded, and so on. As such one will have an endless machine race between CAPTCHA and AI (see von Ahn et al., 2003).

This problem, however, does not apply to reCAPTCHA because its main distortion is not artificial but natural. Whereas standard CAPTCHAs display images of random characters rendered by a computer, reCAPTCHA displays words taken from scanned texts which an optical character recognition (OCR) software has difficulty deciphering because it involves older prints with faded ink and yellowed pages. Of those texts, OCR cannot recognize 16.5% of the words. The percentage of words on which two OCR systems make a mistake is 7.3%. By contrast, humans are more accurate at deciphering such print. For example, two humans using the “key and verify” technique, where each types the text independently and then any discrepancies are identified, can achieve an accuracy of 99.1%.

Physical books and other texts written before the computer age are currently being digitized en masse. reCAPTCHA is being developed to help digitize these texts by enlisting humans to decipher the words that computers cannot recognize. reCAPTCHA displays words taken from scanned texts that two automated OCR programs fail to recognize. These words have three types of distortions. First, and most importantly, there are natural distortions that result from the underlying texts having faded through time. Second, the scanning process introduces noise. Third, artificial transformations are introduced similar to those used by standard CAPTCHAs. However, to meet the goal of CAPTCHA (differentiating between humans and computers), the system needs to be able to verify the user's answer. To do this, reCAPTCHA gives the user two words, one for which the answer is not known and a second “control” word for which the answer is known. If users correctly type the control word, the system assumes they are human and gains confidence that they also typed the other word correctly. And as in Persons' case, each word that needs recognition is identified by two or three humans, and their judgments are verified on consensus.

6. Conclusions

This paper discussed inductive inference to meaningful patterns, when there is little or no theory to assist, by exploring the case of the construction of an index for business conditions in the 1920s by Persons. The index was a graph, the A–B–C Barometer, showing three lines, A, B, and C, going up and down, one after the other, indicating what had happened and was happening in the US economy. The construction of this meaningful pattern was based on a combination of the employment of the “eye” and statistical tools such as correlation.

In this paper “meaningfulness” has a particular meaning, namely accuracy, which is closeness to the truth. Accuracy consists of several components, such as precision and unbiasedness. Precision is dealt with by statistical methods, like the method of least squares, but for unbiasedness one needs expert judgment. Irrespective of how much one tries to cover with objective tools and procedures to find patterns in data, there will always be a need for a residual expert judgment to attain accuracy. The question is—how to make the most efficient use of this kind of judgment? The common view at the beginning of the twentieth century was that if the data can be represented in shapes and forms, the “eye” can function as a reliable judge to reduce bias. Data contain observational errors, but these can be removed by plotting them in a graph where these errors reveal themselves by the irregularities of the graph. Smoothness, that is to say the required degree of

^{xviii} Consistency is defined such that the assignment of numbers apply rigorously the rules of triangle inequality and symmetry.

^{xix} The classic paper arguing that diagrammatic and sentential representations that are informationally equivalent are not necessarily computationally equivalent is Larkin and Simon 1987.

smoothness, is then evaluated by the “eye.” Although it sounds like a paradox, the “eye” is actually used to reduce observational errors.

The need for judgment of the “eye” is even more necessary when the background conditions of the observations are heterogeneous. Statistical procedures require a certain minimal level of homogeneity, but the “eye” does not. The “eye” is an adequate tool for assessing topological similarities when, due to heterogeneity of the data, metric assessment is not possible. In contrast with metric assessments, that is to say measurements, the “eye” is much better in assessing topological features. In other words, the “eye” is able to assess similarities between shapes and forms when their sizes differ in various dimensions. In fact, graphical assessment precedes measurement, or to put it more forcefully, the graphic method is a necessary prerequisite for measurement.

Acknowledgments

I would like to thank Judy Klein, Tarja Knuuttila, Mary Morgan and Nicola Mößner, and the anonymous referees for their constructive comments. I also would like to thank Sabina Leonelli and the participants of the ERC Exploratory Workshop ‘What is Data-Intensive Science?’ at the University of Exeter, December 2014, where I presented this paper for the first time, for their support in developing this paper.

References

- von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003). CAPTCHA: Using hard AI problems for security. In E. Biham (Ed.), *Advances in cryptography—EUROCRYPT 2003*, LNCS 2656 (pp. 294–311). Berlin: Springer.
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321 (12 September), 1465–1468.
- Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. *The American Statistician*, 32(1), 1–11.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Boumans, M. (2002). Calibration of models in experiments. In L. Magnani, & N. J. Nersessian (Eds.), *Model-based reasoning. Science, technology, values* (pp. 75–93). New York: Kluwer Academic.
- Boumans, M. (2012). Visualisations for understanding complex economic systems. In C. Bissell, & C. Dillon (Eds.), *Ways of thinking, ways of seeing: Mathematical and other modelling in engineering and technology* (pp. 145–166). Berlin: Springer.
- Boumans, M. (2015). *Science outside the laboratory. Measurement in field science and economics*. New York: Oxford University Press.
- Bowley, A. L. (1907). *Elements of statistics* (3rd ed.). London: P.S. King and Son.
- Bowley, A. L. (1910). *An elementary manual of statistics*. London: Macdonald and Evans. <https://archive.org/details/elementarymanual00bowluoft> Accessed 04.04.2016.
- Carusi, A. (2012). Making the visual visible in philosophy of science. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 6(1), 106–114.
- Daston, L., & Galison, P. (2007). *Objectivity*. Brooklyn, NY: Zone Books.
- Diebold, F. X. (2013). A personal perspective on the origin(s) and development of “Big Data”: The phenomenon, the term, and the discipline, second version. *PIER Working Paper*, 13–003.
- Franklin, A. (1997). Calibration. *Perspectives on Science*, 5, 31–80.
- Friedman, W. A. (2009). The Harvard economic service and the problems of forecasting. *History of Political Economy*, 41(1), 57–88.
- Friedman, W. A. (2014). *Fortune tellers. The story of America's first economic forecasters*. Princeton and Oxford: Princeton University Press.
- Funkhouser, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris*, 3, 269–404.
- Hankins, T. L. (1999). Blood, dirt, and nomograms: A particular history of graphs. *Isis*, 90(1), 50–80.
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3, 446–474.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.
- Klein, J. L. (1995). The method of diagrams and the black arts of inductive economics. In I. H. Rima (Ed.), *Measurement, quantification and economic analysis* (pp. 98–139). London: Routledge.
- Klein, J. L. (1997). *Statistical visions in time. A history of time series analysis 1662–1938*. Cambridge: Cambridge University Press.
- Kydland, F. E., & Prescott, E. C. (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 14(2), 3–18.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.
- Maas, H., & Morgan, M. S. (2002). Timing history: The introduction of graphical analysis in 19th century British economics. *Revue d'histoire des Sciences Humaines*, 7, 97–127.
- Miller, M. D. (1946). *Elements of graduation*. Philadelphia: The Actuarial Society of America and American Institute of Actuaries.
- Morgan, M. S. (1997). Searching for causal relations in economic statistics. Reflections from history. In V. R. McKim, & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 47–80). Notre Dame, IN: University of Notre Dame Press.
- Morgan, M. S. (2004). Simulation: The birth of a technology to create “evidence” in economics. *Revue d'Histoire des Sciences*, 57(2), 339–375.
- Morgan, M. S. (2012). *The world in the model. How economists work and think*. Cambridge: Cambridge University Press.
- Morgenstern, O. (1954). Experiment and large scale computation in economics. In O. Morgenstern (Ed.), *Economic activity analysis* (pp. 484–549). New York: Wiley.
- Mößner, N. (2015). Visual information and scientific understanding. *Axiomathes*, 25, 167–179.
- Persons, W. M. (1910). The correlation of economic statistics. *Publications of the American Statistical Association*, 12(92), 287–322.
- Persons, W. M. (1916). Construction of a business barometer based upon annual data. *The American Economic Review*, 6(4), 739–769.
- Persons, W. M. (1919a). Indices of business conditions. I. A study of statistical method. *The Review of Economic Statistics*, 1(1), 5–37.
- Persons, W. M. (1919b). An index of general business conditions. *The Review of Economic Statistics*, 1(2), 111–205.
- Persons, W. M. (1920). A non-technical explanation of the index of general business conditions. *The Review of Economics and Statistics*, 2(2), 39–48.
- Persons, W. M. (1924). Some fundamental concepts of statistics. *Journal of the American Statistical Association*, 19(145), 1–8.
- Playfair, W. (1796). *A real statement of the finances and resources of Great Britain*. London: Whittingham. <https://archive.org/details/foruseofenemieso05play> Accessed 04.04.2016.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and objective life*. Princeton, NJ: Princeton University Press.
- Putchala, S., & Agarwal, N. (2011). Machine vision: An aid in reverse Turing test. *AI & Society*, 26, 95–101.
- Shubik, M. (1960). Simulation of the industry and the firm. *The American Economic Review*, 50(5), 908–919.
- Whewell, W. (1858). *Novum organon renovatum*. In (3rd ed.) *Philosophy of inductive sciences* (3rd ed.), (Vol. 2). London: Parker <https://archive.org/details/novumorganonren01whewgoog> Accessed 04.04.2016.
- Whittaker, E. T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75.