Estadística Española • Volumen 59, número 192 / 2017, pp. 29-35

On the origin of Karl Pearson's term "histogram"

Daniel Riaño Rufilanchas
Departamento de Filología Clásica. Facultad de Filosofía y Letras. Universidad Autónoma de Madrid

Abstract

Many modern scholars think that the term "histogram" is related to the word "history". Recent work in the field of the history of statistics has only increased this misunderstanding. The etymology is incorrect for several reasons: first, the word "histogram" does not share a stem with "history"; second, we show that Karl Pearson, who coined the term, used a clearly defined method to designate recently devised graphs. His method, inspired by Levasseur's work, excludes any relation to the word "history" and confirms the link to the Greek *ἱστός* (mast). Some clarifications of Pearson's use of the term are given in the paper.

Keywords: history of statistics, mathematical terminology, histogram, Karl Pearson, graphics, etymology of mathematical terms.

AMS Classification: 01A55.

Acerca del origen del término 'histograma' acuñado por Karl Pearson

Resumen

Muchos especialistas contemporáneos, dentro y fuera del campo de la estadística, piensan que el término "histograma" está relacionado con el sustantivo "historia", una confusión que agravan algunos trabajos recientes en el campo de la historia de la estadística. Esta etimología es necesariamente errónea por varias razones: en primer lugar, desde un punto de vista etimológico "histograma" no puede compartir la raíz de "historia". Pero lo más relevante es que sabemos que Karl Pearson, que fue quien acuñó el término y contaba con una buena preparación filológica, había diseñado un método muy preciso para la designación de los gráficos que se habían diseñado recientemente para visualizar datos estadísticos. Su método, inspirado en el trabajo de Levasseur, excluye cualquier relación con "historia" y confirma su relación con el sustantivo griego "ἱστός" ("mástil"). Se hacen algunas clarificaciones sobre el uso que Pearson daba al término estudiado.

Palabras Clave: historia de la estadística, terminología matemática, histograma, Karl Pearson, gráficos, etimología de términos matemáticos.

Clasificación AMS: 01A55.

1. Introduction

The first aim of this short paper is to clarify the etymology of the well-known, yet often misunderstood, term "histogram"[1] For example, Adrian Rice, in his introduction to the otherwise excellent *Mathematics in Victorian Britain*, notes that:

"Pearson introduced several now-commonplace statistical tools. One of these was the *histogram*, a diagram similar to a bar chart, but which represents a set of continuous, rather than discrete, data. For this reason, Pearson explained that it could be employed as a tool in the study of history, for example to chart historical time periods, and coined the name 'histogram' in 1891 to convey its use as a 'historical diagram'" (Flood, Rice and Wilson 2011: 12; original italics).[2]

Of course, there is nothing in the histogram as we know it (and as Pearson used it) that makes it better suited to represent historical data graphically than any other kind of continuous quantitative data. And naturally, if it were etymologically related to "history", we would expect a word like "historiogram", which does, in fact, exists[3]. Yet the above quote illustrates a common misunderstanding for many, including some of the leading scholars in the field. Schwartzman (1994, s.v. "histogram"), whose work remains the most complete account of the etymology of mathematical terms, goes to great lengths to make sense of the presumed link between "histogram" and the Greek root of the modern word "history" in the second part of the entry for "histogram":

"histogram (noun): […] (2) *Histo-* may be a contraction of *history*, from Latin *historia*, in turn from Greek *istoria* "inquiry, observation." Greek *histor* "a learned man," represents a presumed *wid-tor*, from the Indo-European root *weid-* "to see." Compare the native English cognate *wise*. Etymologically speaking, history is "what has been seen (and presumably also understood)." According to this explanation, a histogram is a "picture history" of a statistical distribution. (3) Whoever coined the term *histogram* may have had both of the above associations in mind, since each is plausible".

Although Schwartzman, unlike many authors, is aware of the incongruity of relating "histogram" to the modern sense of "history" (the study and record of past events), his argument becomes muddy when he attempts to make sense of the supposed link to the Indo-European root **weid*, as it appears in Ancient Greek *ἱστορία* (cf. Chantraine 1968-1983: s.v. *οἶδα*).

[1] Research for this article was financed through the Spanish Ministerio de Economía y Competitividad and the Subprograma Ramón y Cajal (Projects FFI2013-44797-P and RYC-2011-09425).

[2] The examples of this misunderstanding are too many to quote. To illustrate this common mistake made by academic and learned online publications, take this extract from an online entry for "Karl Pearson: "Two days later he [i.e., Pearson] introduced the histogram – a term he coined to designate a 'time-diagram' to be used for historical purposes" *(StabProb,* entry "Karl Pearson").

[3] The Ancient Greek term ἱστορία (*historía*, "inquiry" > "knowledge" > "history" i.e. the account of one's inquiries is derived from the noun ἵστωρ (*hístor*, "one who knows"), itself derived from οἶδα (*oída* "know"), a verbal perfect tense built from *ϝid "to see". ἵστωρ, then, and its derivate ἱστορία is a formation of this Indo-European root *ϝιδ- followed by the suffix -τωρ. The second part of the word, the suffix "-tor", must be present in all their derivates, as in "historic", "historicity", etc., cf. Benveniste 1950: 29-33, Chantraine 1968-83: s.v. οἶδα.

On the other hand, and in spite of Schwartzman's final remarks about who coined the term, we know who it was: the British mathematician Karl Pearson (1957-1936), the founder of mathematical statistics (Porter 2004: 236). Pearson's coining of the term is commonly recognised, but the date and publication thereof are often incorrectly reported. Thus, Dodge (2009: s.v. "histogram") notes, "[t]he term histogram was used for the first time by Pearson, Karl in 1895" and this year (or 1894) is commonly given as the datum for the first appearance of the term (Beniger and Robyn 1978: 4, followed by so many, see Snee and Pfeiffer 2006, Scott 2015: 4, Ross 2010: 59, etc.). But we know that Pearson was using the term "histogram" at least since his 1891 Gresham Lectures (Pearson, 1938: 144; Magnello 2005: note 38).

The confusion surrounding the original meaning of the term grew after M. Eileen Magnello, who is one of the main living specialists on Karl Pearson, gave us an account of Pearson's use of the term, which is misleading in its wording, if not downright incorrect.

"Pearson had already introduced the 'histogram' on 18 November 1891, a term he coined to designate a 'time-diagram' in his lecture on 'Maps and Chartograms'. He explained that the *histogram* could be used for *historical* purposes to create blocks of time of 'charts about reigns or sovereigns or periods of different prime ministers'." (Magnello 2005, cf. note 38)

Magnello takes this piece of information from Pearson's personal notes for his Gresham Lecture (20 November 1891) on 'Maps and Chartograms', archived with a wealth of his personal papers in the Karl Pearson collection in the UCL Library.[4] Although the wording of the above quote strongly suggests that there was some relation between the word "history" and the term "histogram" in Pearson's mind, we can be sure that this was not the case.

Karl Pearson's son Egon published (Pearson, 1938: 142-154) the syllabuses of his father's *Gresham Lectures* 1891-1894. As befit Pearson's character and his methodological approach to science, he considered how he introduced the different "geometrical methods" to represent statistics to the audience. Pearson (who knew Ancient Greek well and could equally be presented as a "philologist turned statistician")[5] started naming the different methods of graphical display, some of them previously unnamed. To do so, he generalised a method already employed by continental statisticians for some of this vocabulary. He constructed a one-word term for each display, the composite being a two-part noun with a Greek (or Latin, for want of a better alternative) first element describing *the main graphical feature of the figure used in the chart.* The second component is "-gram", an English suffix adapted from a Greek substantive *γράμμα* "draw", a noun with the same root as the verb *γράφω*, '*grapho*', "to write".

At least two existing terms showed the pattern: *διάγραμμα,* "diagram", an Ancient Greek term (already used by Plato) had passed into Latin by Vitrubius's times (1st century BC).

[4] Signature KP: UCL/49/21.

[5] Pearson had started a career in German philology before turning his attention to statistics; see Magnello 2005.

This same technical author also used the word "*euthygramma*" to describe a rule for drawing lines. The word, derived from Greek, passed into the technical vocabulary of some romance languages, such as French.[6]

The main inspiration for Pearson's systematisation of graph shapes, and partially for the systematisation of their names, was Levasseur 1885, a paper by the French economist and geographer Émile Levasseur (1828-1911). In addition to "*diagramme*" Levasseur used the existing terms "*cartogramme*" and "*stéréogramme*", which follow the described pattern.[7] The rest of the terms are likely to be Pearson's creations or adaptations.

During his 1891 lectures, Karl Pearson presented the ten following terms to describe as many different kinds of graphical displays for statistical data (Pearson, 1938: 143-144). I give his ten terms here, preceded by Pearson's own words to describe them, and followed by the explanation of the etymology (which is always transparent) and the page in Levasseur 1885 where the French statistician dealt with a similar graphical device. Levasseur's distinctions are not as meticulous as Pearson's, but he adds several illustrations that were probably used either as a source for Pearson's descriptions or as an example during the lecture.

1. By points: stigmograms. From Greek *στιγμός*, '*stigmós'*, "small dot", a rare derivative of *στίγος,* '*stigos*', "point", cf. Levasseur 1885: 220)

2. By lines: euthygrams. From *εὐθύς*, '*euthús,*' "straight"; Levasseur (1885: 221). As explained above, "*euthygramme*" designated a "rule". The meaning "diagram built with straight lines" was Pearson's suggestion. Levasseur had no special term for this kind of graph.

3. By areas: epipedograms. From *ἐπίπεδος*, '*epípedos*', "plane", cf. Levasseur 1885: 223, 225.

4. By columns: histograms. From *ἱστός,* '*histós*', "mast", cf. Levasseur 1885: 222, 228, 228 bis.

5. By curves: diagrams. (see above, and cf. Levasseur 1885: 226).

6. By rays from a point: radiograms. From Latin *radius* "ray", related to Greek *ῥάδις 'rádis'*, cf. "diagramme polaire" in Levasseur 1885: 232.

7. By straight line Linkages: hormograms. From *ὅρμος,* '*hórmos*', "chain, necklace", cf. Levasseur 1885: 230 bis, tris.

8. By maps: chartograms. From Latin *charta,* "map", itself coming from Greek *χάρτης,* '*chártes*', "papyrus leaf", cf. Levasseur 1885: 235.

9. By surfaces drawn on the plane: topograms. From *τόπος*, '*tópos,*' "place", cf. Levasseur 1885: 245, who calls them "*cartes topographiques*" or "carte en relief".

[6] "Euthygramma" derives from εὐθύγραμμος, 'euthúgrammos', an actual Greek technical term used by Aristotle with the meaning "rectilinear figure" (cf. Arist.Cael.286b13).

[7] The term "stereogramma" appears in an 1880 chart prepared by the Italian Direzione Generale della statistica used by Levasseur in his paper (246 ss).

10. By solid surfaces: stereograms. From *στερεός*, '*stereós*', "firm, solid", cf. Levasseur 1885: 246 bis.

Today, only *histogram*, *diagram* and *cartogram* (with a spelling that identifies the link to the French form of the Latin word) are used in statistics-related terminology (cf. Porter 2004: p. 236). The term *euthygram* to designate a "ruler" is an archaism in French, if used at all. *Topogram* has a different, if somehow related meaning, while the current use in English of *radiogram* (or *radiograph*) and *stereogram* is altogether different, and their genesis seem to be unrelated to Pearson's term. *Hormograms, stigmograms* and *epipedograms* seem to have died with their creator.

2 Clarifying Pearson's use of "histogram"

Most of us use the term histogram in the sense defined by Ross: "[a] bar graph plot of the data, with the bars placed adjacent to each other" (Ross 2010: 33, and see Triola 2010: 55, etc.), typically used in frequency distributions. This is not the way in which Pearson intended the word be used.

As we saw in his systematic account of graphical displays, Pearson was following Levasseur's "*Figures de statistique*" (1885; see Porter 2004). Levasseur had left some figures unnamed or termed in a way that Pearson found too vague (for example, Pearson's "hormograms" were mere "*courbes*"). To describe the different figures, Levasseur proceeded, in a very French way, from the simplest form (the point) towards the most complicated figures (lines, surfaces, and "*diagrammes proprement dits ou courbes*"), although things became complicated halfway through.

The graph that both men were trying to define in their seminal papers is not always what comes to mind for the modern statistician. The "point figure", corresponding to Pearson's stigmogram was nothing like our scatterplot, but rather a simple figure drawn using points representing unities, as one can represent the square of 4 drawing 16 points in a four-point sided square. The euthygram is similar to our simple horizontal bar chart, but drawn using only horizontal one-dimensional lines (continuous or dotted) instead of surfaces. Epipedograms are statistical graphs where the different values are represented as areas with the same ratio scale. Today, we call Pearson's hormograms "frequency lines/ polygons" (with continuous lines joining the dots indicating a measure) and his radiograms "radar charts".

For Pearson, histograms are bar-like charts, using either horizontal or vertical lines. It seems as if he was indifferent to the bars being placed continuously or with an interval, but he seems to have intended them to be used mainly for frequency distributions where data values are sorted and placed in a class interval: in a footnote to his first use of the word "histogram" in Pearson (1895: 399), it is specified that the word was "[i]ntroduced by the writer in his lectures on statistics as a term for a common form of graphical representation, *i.e.,* by columns marking as areas the frequency corresponding to the range

of their base".[8] Pearson, who seemed to be mostly interested in the psychological effect of the difference in orientation of the bars, found some "optical advantage of vertical over horizontal columns" (Pearson, 1938: 144), hence the choice for a word specifically meaning a vertical structure like a mast as the element for the root of the new word "histogram".

William Playfair (1759-1823), the Scottish economist, inventor and polymath who is usually considered the first to use histograms, if not the inventor of the term, favoured a display using a separation between bars throughout his 1786 epochal work *The Commercial and Political Atlas*.[9] Levasseur (from whom Person takes some of his examples for the lecture) referred the reader to figures originally drawn using contiguous bars (222, 228, 228 bis), but neither author showed a specific concern for the significant use of this separation.

The mentions of the word "history" that can be found in Pearson's syllabuses are just an indication of the "origin of Diagrams in time Histograms" [Pearson 1938: 143, with Karl Pearson's' capitalisations. They refer to a particular set of diagrams, and therefore they have no relation whatsoever with the constituent parts of the word "histogram" (see also Lo Bello 2013: s.v.).

References

BENIGER, JAMES R. AND DOROTHY L. ROBYN 1978 «Quantitative graphics in Statistics: A brief History» in *The American Statistician*, 32:1, 1-11

BENVENISTE, ÉMILE 1948, «Noms d'agent et noms d'action en indo-européen», Paris: Adrien-Maisonneuve

CHANTRAINE, PIERRE 1968-1983, «Dictionnaire étymologique de la Langue Greque», Paris: Klinksieck.

DODGE, YADOLAH 2009 «The Concise Encyclopaedia of Statistics» New York: Springer Verlag.

FLOOD, RAYMOND; ADRIAN RICE AND ROBIN WILSON (eds.) 2011 «Mathematics in Victorian Britain», Oxford: Oxford University Press.

LEVASSEUR, ÉMILE 1885 «La Statistique graphique» in *Journal of the Statistical Society of London, Jubilee Volume*. pp. 218-250.

LO BELLO, ANTHONY. 2013 «Origins of mathematical words: A comprehensive dictionary of Latin, Greek, and Arabic roots» Baltimore: Johns Hopkins University Press.

8 This note, made by the editors of Pearson 1895, and likely suggested by the author himself, seems to be the source of the error in the date for the introduction of the term: The "lectures" referred to in this paper are the 1891 Gresham Lectures, not the present ones.

9 He does use a crude form of histogram in his latter work *Letter on our agricultural distresses, their causes and remedies*, printed in London in 1822, without a terminological distinction between both.

MAGNELLO, M. EILEEN. 2005 «Karl Pearson and the origins of modern Statistics: An elastician becomes a statistician» en *The Rutherford Journal,* 1 [accessible at < http://www.rutherfordjournal.org/article010107.html>]

PEARSON, EGON SHARPE. 1938 «Karl Pearson: An Appreciation of Some Aspects of His Life and Work*»* Cambridge: CUP.

PEARSON, KARL. 1895 «Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material» in *Philosophical Transactions,* 186. 1,343-414

PORTER, THEODORE M. 2004 «Karl Pearson. The scientific life in a statistical age» Princeton: Princeton University Press.

ROSS, SHELDON M. 2010 (3ª ed.) «Introductory Statistics*»* Burlington, MA: Academic Press.

SCHWARTZMAN, STEVEN. 1994 «The Words of Mathematics. An Etymological Dictionary of Mathematical Terms used in English» Washington: The Mathematical Association of America.

SCOTT, DAVID W. 2015 «Multivariate Density Estimation*»* New Jersey: Wiley.

SNEE, RONALD D. AND CHARLES G. PFEIFER. 2006 «Histograms» in *Encyclopedia of Statistical Sciences,* Vol. 5. London: Wiley.

STABPROB*,* Springer and the PlanetMath.org [on-line resource, http://statprob.com/encyclopedia/KarlPEARSON.html, accessed 2015.09.10]

TRIOLA, MARIO 2010 (11th ed.) «*Elementary Statistics»* Boston, MA: Addison-Wesley.