

**Technological Institute of Tijuana****Academic Subdirectorates****Systems and Computing Department****SEMESTER:** August - December 2021**CAREER:** Computer Systems Engineer**MATTER:** Data Mining**JOB NAME:** Unit 2 - Investigation 1 - Grammar of Graphs**STUDENT NAME AND CONTROL NUMBER:**

Castro Cebreros Alejandro - 16211341

Márquez Millán Seashell Vanessa - 17212153

TEACHER NAME: Jose Christian Romero Hernandez**DATE OF DELIVERY:** October 29, 2020**1. What is the Origin of the graphs?**

It is not easy to pinpoint the exact moment when statistical charts were born. Graphical representation techniques were used since ancient times, in the form of terrestrial and celestial maps, but it could be said that it was not until the 18th century when modern statistical graphics emerged, mainly due to the advancement of the different scientific and social branches.

The word statistics derives from the modern Latin *statisticum collegium* ("council of state"), from the old Latin *status* (position, "form of government"), from the modern Italian word *statista* ("statesman", "politician") and from the old Italian *stato* ("State"). In 1749, the German, Gottfried Achenwall (1719-1792) uses the term *Statistik* in his book entitled "*Staatswissenschaft der vornehmen Europäischen Reiche und Republiken*", who originally designated the statistical word for the analysis of government data, defining it as the "Science of the State". Gottfried Achenwall is known as the "Father of Statistics."

The first person to introduce the term statistics in England was Sir John Sinclair (1754-1835) with his work "Statistical Account of Scotland" (1791-, 1799) compiled in 21 volumes. The author explains in his book that the word statistics was adopted thanks to the study of research carried out in Germany, as a new word that would attract the attention of the English; In contrast, in Germany statistics is used as an instrument to measure the strength of a state, while Sinclair would use it as a generator of internal information to find shortcomings and propose improvements in the country. This work was followed by two publications: the second edition produced between 1834 and 1845; the third edition begins after the second world war covering the periods between 1951 and 1992. At the beginning of the 19th century, the word statistics adopted a more generalized meaning towards the collection and classification of any type of quantitative data.

Scottish engineer and economist William Playfair (1759–1823) is considered the father of statistical graphics. He pioneered the use of the line chart to represent time series and was the creator of the bar chart and the pie chart. Playfair's life was marked by multiple vicissitudes as befits the times in which he lived. He carried out numerous activities (engraver, engineer, real estate speculator, journalist, newspaper editor, economist, author of treatises) until in the final part of his life, he began to write. The result of this activity is The Commercial and Political Atlas (1786) considered the first work of graphical representation of data. This statistical atlas uses 43 line graphs and a bar graph. Later he would publish The Statistical Breviary (1801) where pie charts and pie charts would appear for the first time.

2. What is the grammar of graphs?

In the same way that to construct a sentence in our language, we make use of combinations of words (which can be nouns, verbs, adjectives, etc.) following a set of rules known as grammar, the grammar of graphics provides us with the rules to construct statistical graphs.

In short, the grammar of graphs means that: A statistical graph is the mapping of variables in our data to the aesthetic attributes of geometric figures.

We can build a graph from three essential components (Taking into account that the language used is R): data: the data set that contains the variables of interest geom: the geometric figures in question. We refer to the type of object that we can observe in a graph. For example: points, lines or bars. aes: the aesthetic attributes of the geometric figure. For example, its position on the x / y axes, color, shape, and size. The aesthetic attributes are mapped to the variables contained in our data set. An example of the grammar in this case will use the data from the gapminder package. Based on life expectancy data in Peru over the years. For this we will create df_example.

In this example, the ggplot2 package, elaborated by Hadley Wickham, was used, which is the R implementation of the grammar of graphs, a theory developed by Leland Wilkinson. How does the grammar work in this example? Ggplot requires three basic components to create graphs: the data, the mapping of variables, and the geometric figures that will represent our data. In the above code:

*We tell the ggplot () function that our data will be df_example. *With mapping we ask you to map the year variable to the X axis and the lifeExp variable to the Y axis. *In addition to doing this, it is necessary to tell it which geometric figure ggplot should use to represent our data. We indicate this instruction with geom_line (), asking it to do so with a line.

Bibliography

Dos siglos de graficos estadisticos: 1750 - 1950. Ine.es. (2021). Retrieved 30 October 2021, from https://www.ine.es/expo_graficos2010/expogra_inicio.htm#presentacion.

Quesada, V., & Vergara, J. (2021). 1.1 UN POCO DE HISTORIA. Eumed.net. Retrieved 30 October 2021, from <https://www.eumed.net/libros-gratis/2007a/239/1a.htm>.

William Playfair, el inventor de los gráficos y la visualización de datos - Dataly. Dataly. (2020). Retrieved 30 October 2021, from <https://dataly.es/william-playfair-inventor-los-graficos-la-visualizacion-datos/>.

Calderon, S. (2021). RPubS - La gramática de los gráficos. Rpubs.com. Retrieved 30 October 2021, from <https://rpubs.com/scalderon/taller-I-sesion-06>.