Advances in Social Science, Education and Humanities Research, volume 670
Proceedings of the 2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022)

Historical Overview of Data Visualization and its Attempts and Reflections in LIS

Zhi-Wei Liu[1,2,*], Ming-Yueh Tsay[1]

[1] *National Chengchi University, Taiwan*
[2] *Zenith Steel Group Corporation, China*
[*]*Corresponding author. Email: ziwee.leo@gmail.com*

ABSTRACT

Starting from the papyrus map of Turin in ancient Egypt, this study reviews the complete historical process of data visualization development. By taking "The Papyrus map of Turin to before the First Industrial Revolution", "The First Industrial Revolution to the eve of the birth of computers" and "After the emergence of computers" as the narrative dimension, this study lists the characters, backgrounds, design concepts and influences of different data visualization graphics development in detail, integrates the development process of data visualization tools, and summarizes the two thought development stages of data visualization and its tools in the development process. At the same time, this study make a dissicusion of the data visualization attempts in China's LIS and found problems such as single application form, lack of design, poor readability, and unfriendly publishing style when using data visualization.

Keywords:*Review, Data visualization, LIS*

1. INTRODUCTION

The development of data visualization in the modern sense can be traced back to computer graphics after the computer's invention in the early 1950s. The National Science Foundation report "Visualization in Scientific Computing", written in 1987 by McCormick et al., has been a tremendous boost and excitement for this area [1]. With computer computing power improvement, the subject's scope became wider and wider, and the data scale and complexity increased. Wehrend developed the concept of "Scientific Visualization" in 1984 for Visualization purposes [2]. In 1999, Card introduced the concept of "Information Visualization" to support more domains, with a review of relationships using Visualization thinking and related papers such as interactive graphics Visualization extension thinking. In 2002, Post introduced a new term for Visualization, "Data Visualization" to unify the above concepts [3]. The emergence of new concepts greatly enriched the range of technical approaches in the field, allowing Data to be expressed, modeled using Visualization approaches such as graphics, image processing, computer vision, and user interfaces, and has continued to this day. However, the history of human visual representation of quantified information goes far beyond that. Tracing back to its origin, data visualization benefits from the development of thematic mapping and statistical graphics and is influenced by the ancient design ideas of statistics and data.

At present, the widespread application of data visualization technology has greatly accelerated this field's development and has achieved remarkable results. However, while a large number of researchers are pursuing practicality, it seems that few researchers are willing to pay attention to its budding state and development context. Few researchers reflect on the shortcomings of data visualization methods in the subject application, which will not be conducive to data visualization's long-term development.

2. RESEARCH METHODS AND PROCESSES

This research adopts the systematic literature research method, firstly combing lots of historical documents that visually portray quantitative information. From the three stages ("The Papyrus map of Turin to before the First Industrial Revolution", "The First Industrial Revolution to the eve of the birth of computers" and "After the emergence of computers"), systematically explain the technological evolution in the development of data visualization Status, focusing on the characters, background, design concept and influence during the development of visual graphics. Secondly, review the development process of data visualization tools, and summarize the characteristics. Thirdly, comprehensively discuss the current status of China's LIS data visualization attempts, reflect on the existing shortcomings, and make suggestions.

3. FINDINGS

3.1. The foundation for data visualization (The Papyrus map of Turin to Before the First Industrial Revolution)

Most of the recorded visualization developments originally came from ancient geometric diagrams, astrological diagrams, celestial diagrams, and maps, mainly used to navigate unknown areas. Limited by the amount of knowledge, scope of activities, and technological level at the time, the form of early data visualization was relatively simple. The earliest surviving ancient visualized topographic map globally is the Turin papyrus map drawn by the ancient Egyptians (Figure1). The map is the first to record mineral distribution and mining data in the Nile's middle reaches using text signs and a graphical interface [4]. In 950 AD, European scientists used grid coordinates to plot celestial bodies' trajectories over time (Figure 2). This "multi-line graph" with horizontal axis, vertical axis, and polyline has begun to have some characteristics of modern statistical charts [5].

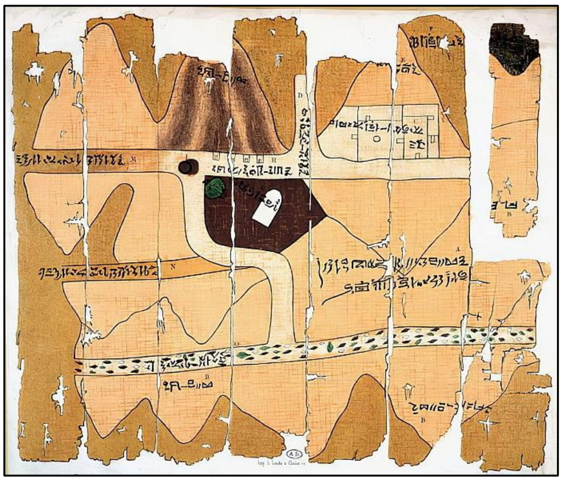

Figure 1 Papyrus Map of Turin (1150 BC)

In 1305, Palma mathematician Ramon Llull gave full play to rationality and romance. He created an unprecedented visualization method "the tree of knowledge" (Figure 3), which used the growth path and hierarchical relationship of the trunk and fruit to explain the knowledge in the scientific field Growth, which vividly shows the structure, connection and evolution process of knowledge in different fields, is considered to be the earliest knowledge graph masterpiece [6].

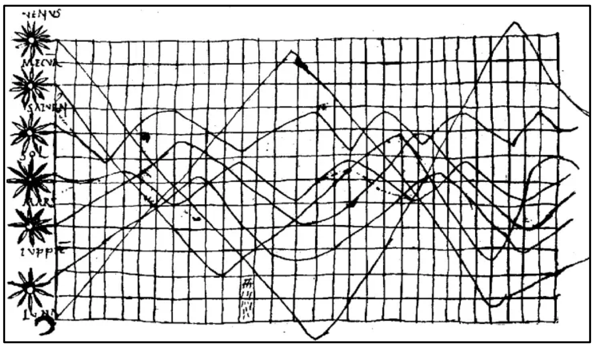

Figure 2 Celestial motion map (AD 950)

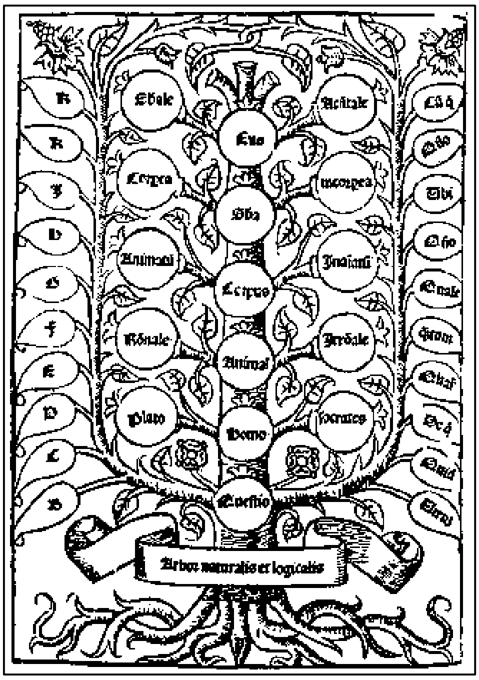

Figure 3 The Tree of Knowledge (1305)

In 1350, the French philosopher Nicole Oresme Oresme invented the original, theoretically functional bar chart for Oresme to plot the variable (Figure4). In 1637, the philosopher Descartes put forward the discourse of analytical geometry and coordinate system. During the same period, with the rise of John Graunt's demographics at the time, and the rise of Fermat and Laplace's probability theory, data visualization was formally moved towards the systematic development of data collection, processing, and rendering. In 1644, Langren drew a chart showing the difference in longitude between Toledo and Rome, which became a typical visual case of deviation in statistical estimation. Starting from practice, he explained that charts are the most effective way to highlight information features (Figure 5) [7]. Langren's idea largely explains the root cause of data visualization. Since then, people began to improve the quality of expression while seeking the accuracy of visualization data.

3.2. The modern process of data visualization (The first industrial revolution to The eve of the birth of computers)

In the 18th century, mathematics, physics, and chemistry began to develop vigorously, and statistics began to sprout, which promoted the development of data visualization in the direction of precision and quantification. Data has begun to be valued by people, and empirical data such as population and business have been systematically collected, sorted and recorded, and various data graphics have also begun to emerge.

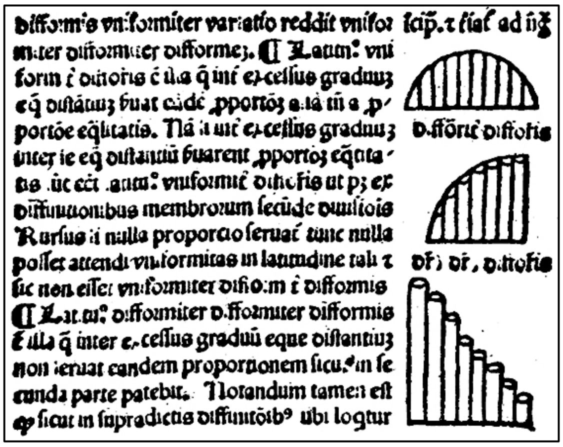

Figure 4 Original bar chart (1350)

In 1753, French physician Dubourg designed a method of visualizing histograms "Carte chronographique". He used the scroll structure to show the world history spanning 140 years, which also indicated the official birth of the idea of "interaction" in the development of data visualization (Figure6) [8][9]. In 1779, the French mathematician Lambert pioneered the first dual format chart that combined tables and graphics to record the periodic changes of soil temperature, and simultaneously presented multidimensional information such as numerical values, trends, and fluctuations (Figure7) [10]. In 1782, the French mathematician Fourcroy introduced the "geometric scale" drawing method in his book "Comparison of City Sizes at the Same Scale", abstracting the city area as a rectangular figure of equal proportions and used it for comparison of different cities (Figure8) [11], which pushed data visualization from an abstract perspective to an expression method constructed with fictional feature graphics. It has been influencing it ever since. In 1786, Playfair, a Scottish engineer, invented bar chart, line chart and area chart in his work "Atlas of Business and Politics" [12], and proposed pie chart again in statistical Abstract published in 1801 [13], which laid the basic paradigm of statistical graph and was later regarded as the founder of graphical statistical method. In 1826, the French mathematician Dupin pioneered a new model of presenting statistics by color proportion and drew the first choropleth map (Figure 9) [14]. In 1837, British soldier Harness used residents' travel data in different regions to create a flow map between Dublin and Ireland and other regions (Figure 10). In this map, the independent values are presented in a coherent visual manner, and the thickness of the connection is used to reflect the size of the flow, which intuitively shows the density of traffic between cities at that time and plays a role in identifying traffic hubs [15]. In 1855, British physician John Snow used a dot map to describe London's cholera cases (Figure11). The medical community later recognized this event as the founding event of epidemiology, and dot plots were also widely used in demographic statistics. In 1884, the Irish statistician Mulhall developed a pictogram that used pictograms to refer to the main objects of statistics and used the pattern's size to intuitively reflect the numerical value of the statistics, which further improved the efficiency of information transmission (Figure12). In 1888, French engineer Cheysson designed an Anamorphic Map that used the radial space's size change to display variables (Figure13). He proposed a data visualization processing mode that does not need to strictly follow the map specifications, but uses graphics to illustrate social issues. In 1920, when the American geneticist Sewall Wright analyzed genetic paths, he used arrows to connecting genetic factors, drawing a causal genetic path diagram (Figure14), which was used to reveal the influence law and transmission characteristics of multiple factors in the genetic process, that also indirectly promoted the generation of social network diagram. In 1933, British cartographer Henry C. Beck broke the cartographic tradition of navigation maps and proposed the concept of non-geographic linear graphs, creating a complete color system route map for the London Underground (Figure15), which is still the world's subway common design standards for the roadmap today [16]. In general, from the middle of the 18th century to the first half of the 20th century, the form of data visualization has been greatly enriched, and the field of application has been continuously expanded. It has experienced the evolution from single-dimensional to multi-dimensional; from single-color to multi-color; from single information to complex information; from statistical graphics to theme drawing; from strict drawing to non-strict drawing; from science and technology to economic management. In this process, people's ability to use data is constantly improving, and their recognition of the value of data is also constantly improving. All this seems to be preparing for the computer's arrival, and the latter pushes the data visualization to the next climax.

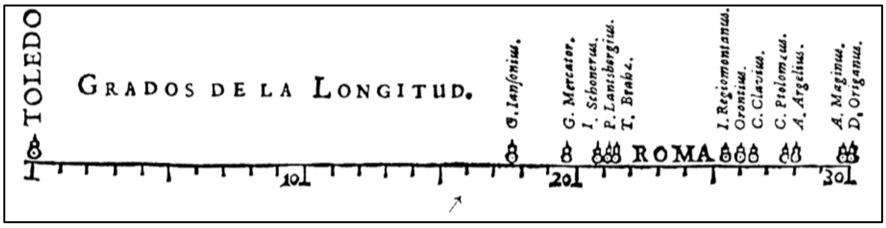

Figure 5 Longitude distance between Toledo and Roma (1644)

3.3. Contemporary development of data visualization (After the emergence of computers)

In 1946, the world's first general-purpose computer "ENIAC" was born at the University of Pennsylvania. With the advancement of computer graphics capabilities, the way of hand-drawing diagrams was gradually replaced by calculation programs, and the presentation and logic of graphics were further liberated.

Figure 6 Scroll of Time (1753)

In 1981, the American semantic analyst Furnas invented the Fisheye View, which created a new information focus and guidance method by providing focus and more detail in a large information area while weakening the surrounding environment (Figure16) [17]. In 1989, American statistician Haslett et al. used computer programs to create the first computer interactive statistical graphics "Interactive maps". Users can click on the map's links to view statistical graphics information of different locations (Figure 17) [18]. The emergence of interactive graphics has completely broken the limitation that graphics can only present static data and opened a precedent to develop dynamic, multidimensional, and interactive data visualization.

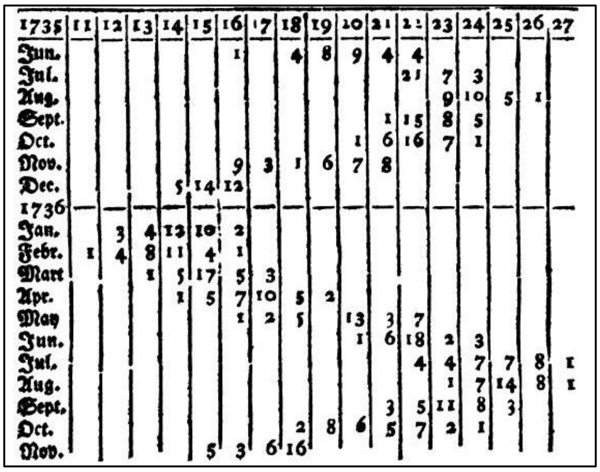

Figure 7 Periodic diagram of soil temperature (1779)

In 2002, interface interaction designer Butterfield created a Tag cloud tool (Figure18) for statistics of user-generated tag content when organizing website system logs [19]. It extended the perspective of humans in data visualization from focusing on numerical value to the content, and thoroughly promoted content analysis methods. In 2009, geneticist Krzywinski invented the Chord diagram to interpret the relationship between multiple genetic objects, using a circular form to provide a new idea for later association analysis of complex data (Figure19) [20]. Since 2009, human society has entered a data-driven development stage, and its dependence on data visualization has continued to deepen. Although the wave of graphic innovation has passed, in the era of big data, data visualization began to focus on the optimization of massive data management methods, the improvement of the ability to correctly communicate the value of data, the innovation of interactive methods, and the broadening of application fields. Based on the above reasons, various data visualization tools have also begun to appear one after another, and have made considerable progress and development, which has greatly promoted the popularization and maturity of data visualization in various fields.

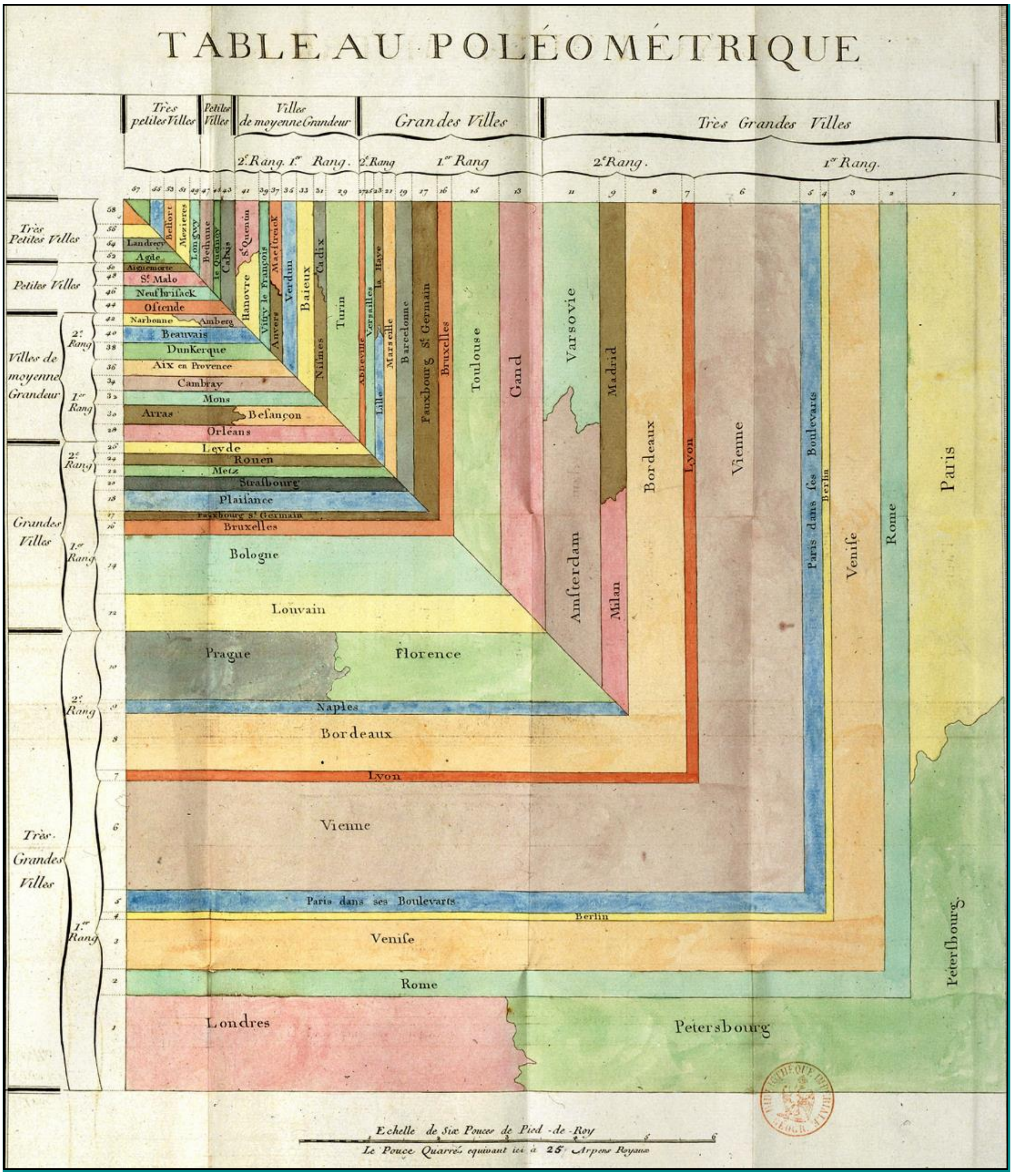

Figure 8 City Area Scale (1782)

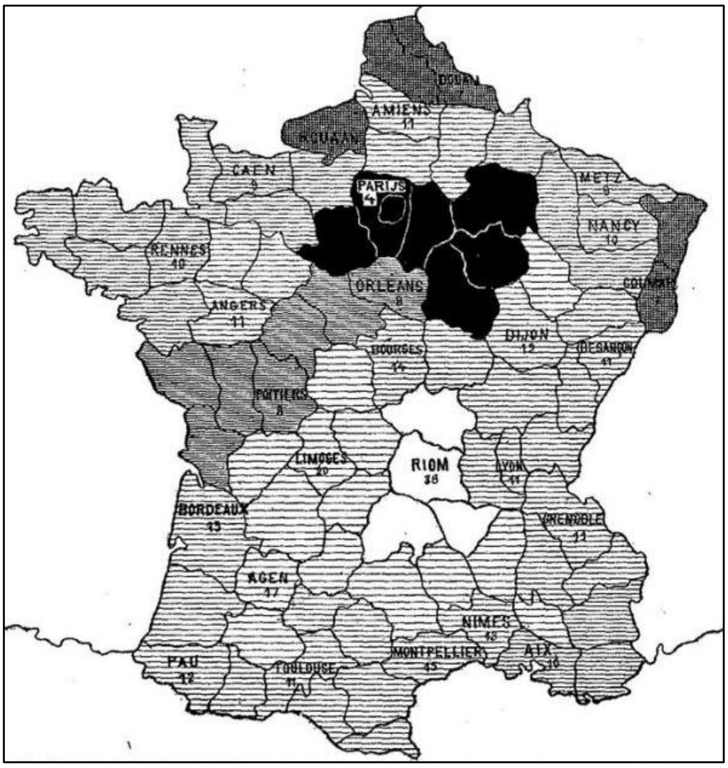

Figure 9 Distribution of the number of illiteracies in France (1826)

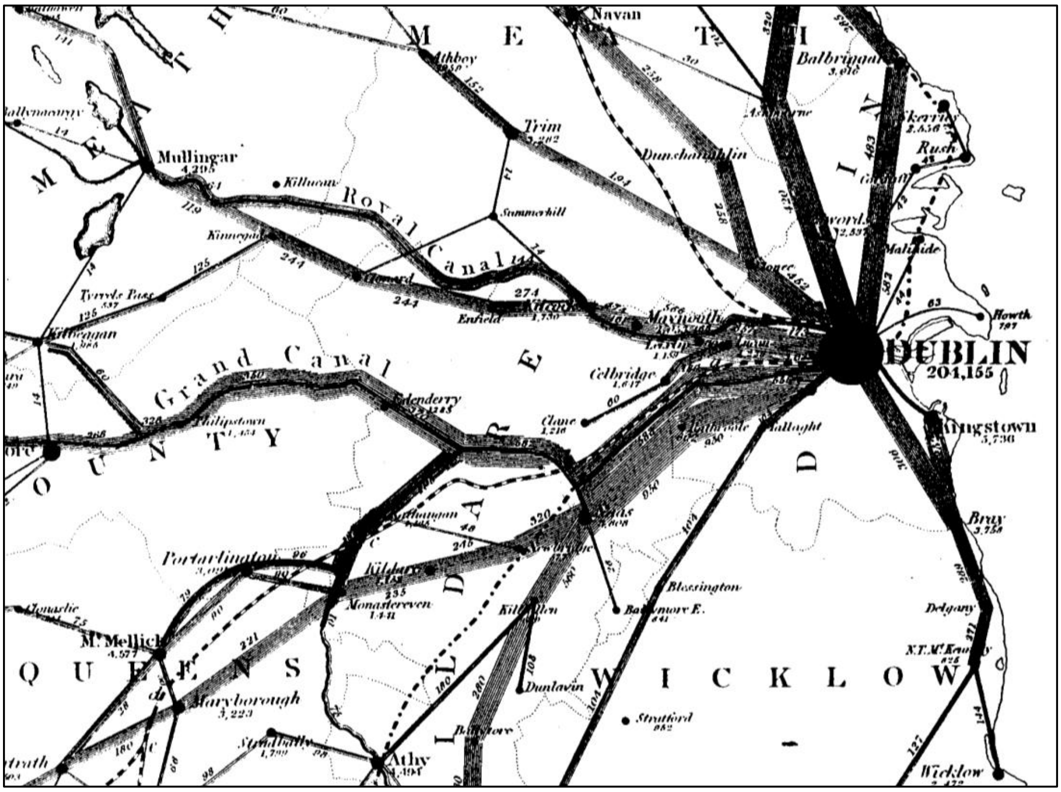

Figure 10 Traffic Flow Map (1837)

Figure 11 Distribution of Cholera Cases (1855)

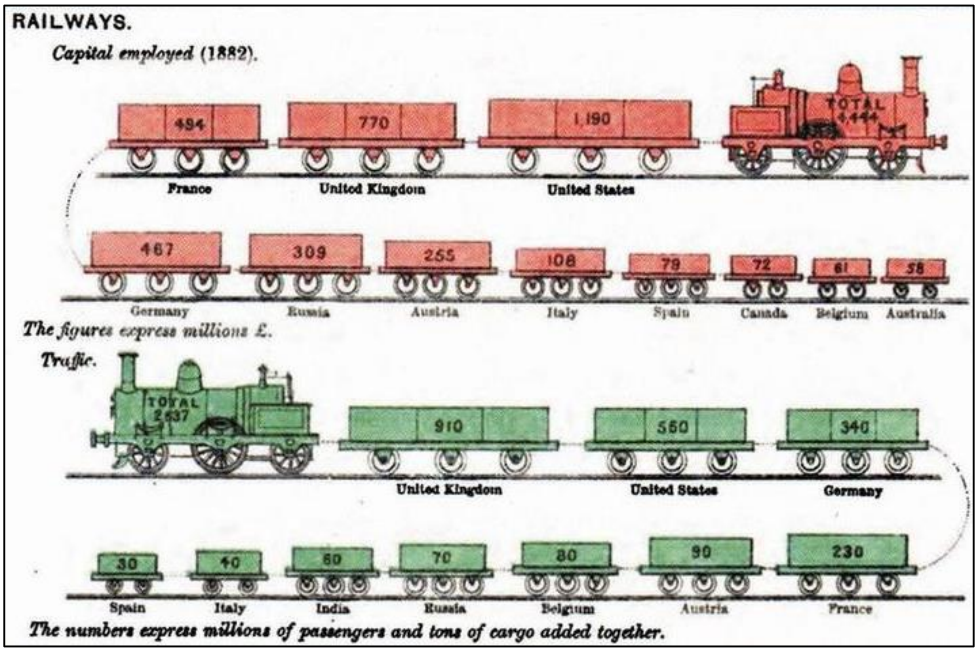

Figure 12 Map of Import and Export Volume of Countries (1884)

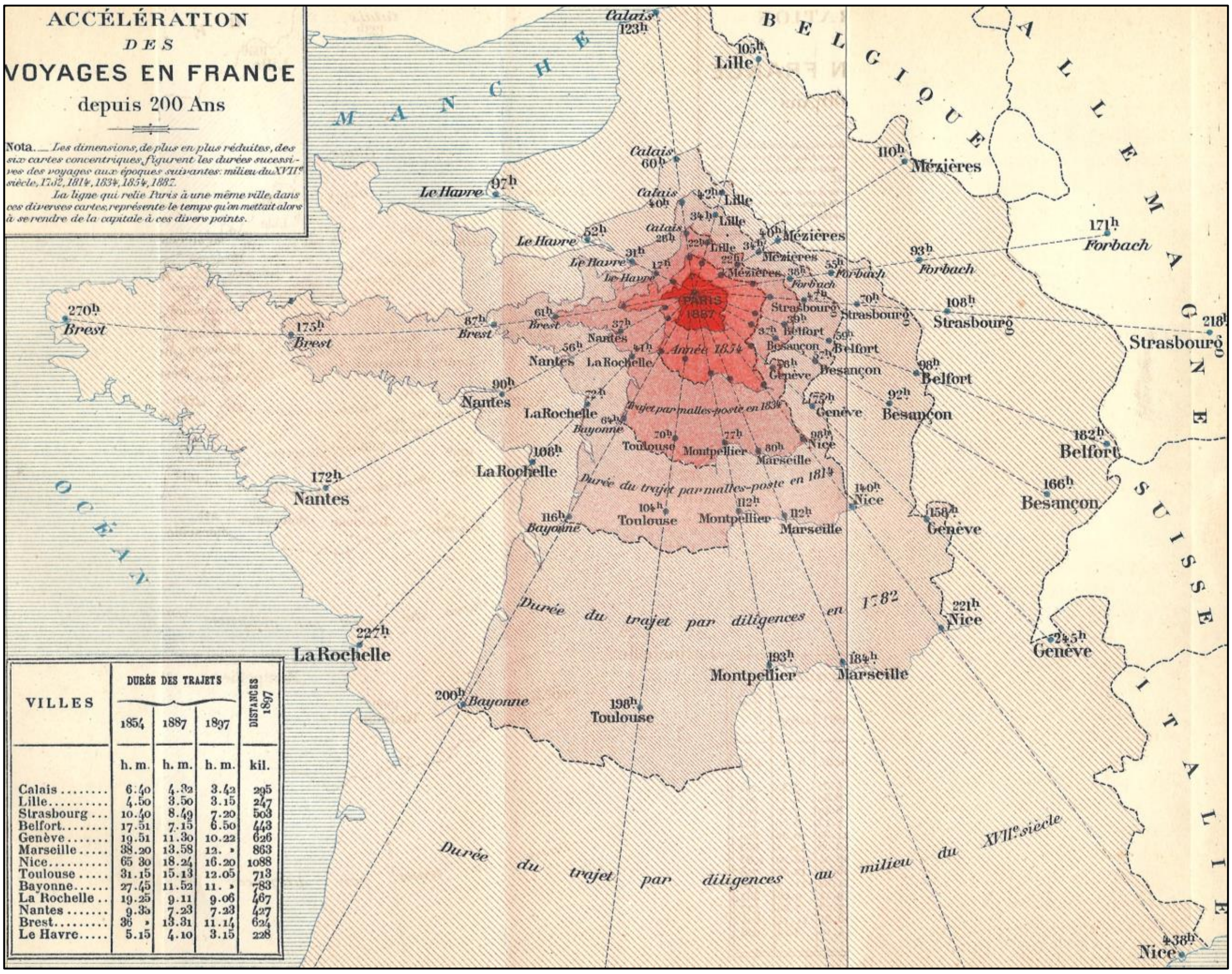

Figure 13 Travel time chart (1888)

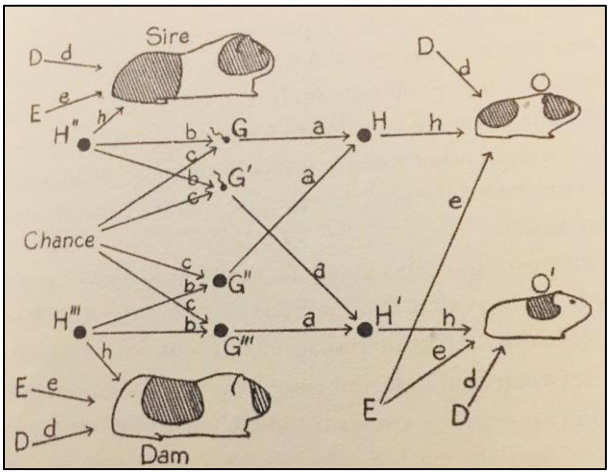

Figure 14 Genetic Path Map (1920)

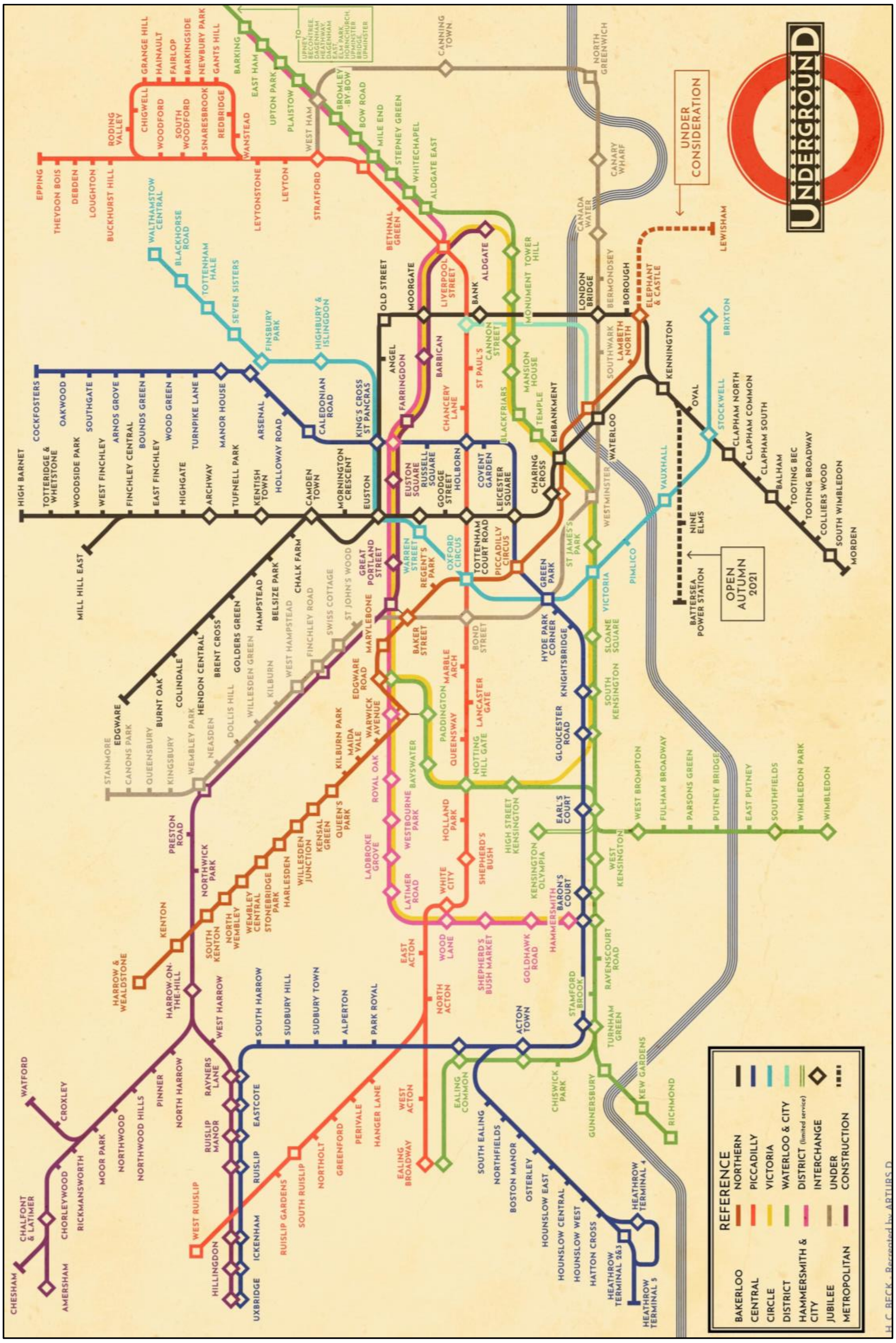

Figure 15 London Underground Route Map (1933)

3.4. Birth and development of data visualization tools

The germination of data visualization tools can be traced back to Excel 2.0, developed by Microsoft in 1988. Although it was still in the most primitive state at that time, only simple charts (such as pie charts and line charts, etc.) could be created, but its appearance created a precedent for the use of computer-aided data graphics presentation, which ultimately inspired people to use graphic tools to express data The desire for information. In 1992, the US Environmental Systems Research Institute (ESRI) developed ArcView, a data visualization mapping tool based on the PC platform, which made it possible to draw geospatial graphics on the desktop. In 2003, the first network relationship visualization tool Cytoscape was born in bioinformatics, and it was used in the research of tracing biological, genetic network relationships [21].

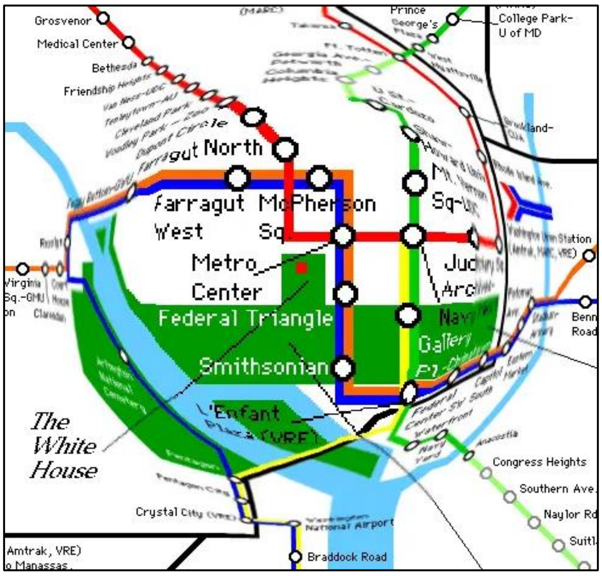

Figure 16 Fisheye views of subway lines (1981)

In the same year, Tableau Software Company was established. The company is committed to the visualization application of business data analysis. Drawing static data visualization graphics adds support for relational databases and multidimensional data sets and realizes data storage and retrieval [22]. After 2005, programming language-based visualization graphics libraries represented by Prefuse, ggplot, and D3.js became the mainstream. These general graphic libraries provide more abundant graphic presentation styles and better strengthen the expressive power of interactive information graphics. With its high compatibility, it has wholly broadened the application scenarios of data visualization tools. Simultaneously, the early visualization tools with single functions have been gradually updated and iterated to meet the visualization needs of various industries, while the later development of visualization tools are mostly updates or functional integrations based on the above concepts.

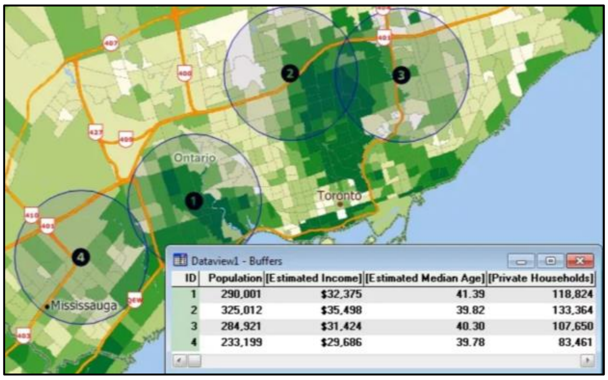

Figure 17 Interactive map Statistical Graphics (1989)

4. RESULTS

This research reviews the development of both data visualization graphics and tools in detail. Although the history of data visualization is much longer than the development of its tools, comprehensively speaking, both the data visualization ideas and tool development ideas are both Following a similar pattern, it can be summarized in the following two stages. The first stage is mainly characterized by the use of graphics/tools to restore data information. Using simple charts and superimposing simple colors, labels, notes, and other elements, the visual graphics can be directly read as much as possible. The second stage is about developing visual graphics systems and specifications, focusing on the cartographic principles of assigning data attributes to graphics. The visual graphics/tools developed abandoned the previous cartographic concepts that required text annotations and color matching. They emphasized the diversity of data content through the geometric characteristics of different graphics and the construction of the logical relationship between graphics and data. The advancement of thinking has triggered a new data research paradigm and has also promoted the reform of research methods in various professional fields. In this process, LIS, which has always used data, materials, and documents as its important research objects, has significantly benefited.

5. DISCUSSION

In recent years, with the popularity of data science research in LIS, the application of data visualization methods and tools is favored. In recent years, with the popularity of data science research in LIS, the application of methods and tools for data visualization has been favored. The research on data visualization in LIS first sprouted in the late 1990s in China, and started from the early discussion on the visualization of citation network systems in the field of bibliometrics [23], and then gradually affected the field of library systems, mainly involving information retrieval results' visualization processing and the visualized transformation of library business data [24][25][26].

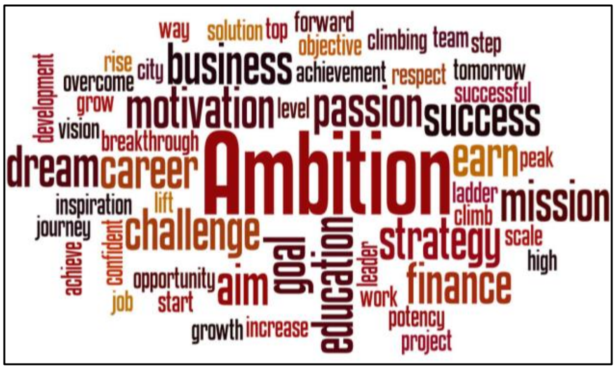

Figure 18 Tag Cloud (2002)

In 2003, Jing Peidong et al. formally introduced the concept of "information visualization" into the field of information science [27], which promoted LIS's further exploration of visualization research and application. Especially in the field of bibliometrics, some scholars have gradually realized the graphical interpretation of citation relation [28], cooperative relation [29], textual relation [30], and other metrological problems based on the visualization idea. Meanwhile, in the ten years from 2001 to 2010, foreign data visualization analysis tools (such as Histcite, Ucinet, CiteSpace, Gephi, VOSviewer, Bibexcel, CitNetExplorer, etc. [31]) were successively developed and introduced into Chinese academics almost at the same time. The emergence of tools has dramatically reduced the threshold of data visualization applications and has completely set off a research boom in knowledge graphs and social network analysis in LIS and has further enriched the scope and form of scientific research evaluation. Today, the data visualization application of LIS in China involves many aspects, such as hotspot analysis, trend analysis, frontier analysis, citation analysis, social network analysis, co-word analysis, cluster analysis, public sentiment analysis, knowledge mapping, scientific research cooperation, scientific research evaluation, knowledge moving, humanities history, etc.

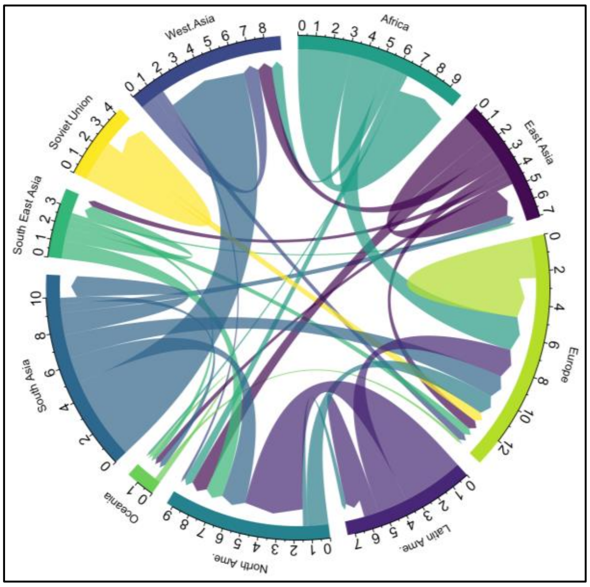

Figure 19 Chord diagram (2009)

Data visualization has played an essential role in the rapid development of domestic graphic science in recent years in LIS, but some shortcomings can also be seen after a long period of development. First of all, it is shown that the current visualized graphics presentation form is relatively single and convergent. A large number of studies have used network relationship maps to present various data relationships. On the one hand, it is limited by the functional limitations of visualization tools; on the other hand, it also shows that visual graphics thinking in the discipline needs to be broadened. The good logical mapping relationship between data and graphics still needs attention and optimization. Secondly, the resulting data visualization results are less readable. Although the existing tools can meet data visualization analysis needs, the visualization logic and specifications of different tools are different from each other. Researchers still need to use much text to explain the graphics, and the graphics themselves are insufficient. Thirdly, the overall data visualization result lacks design and poor visual experience. That is reflected in the limited graphics design capabilities of existing visualization tools, but also, the researchers themselves generally lack the ability and thinking for data design and interactive visual experience. Fourthly, the current way of publishing scientific research of LIS cannot entirely support data visualization results. Traditional publishing is limited by the publishing medium, unable to express rich graphic information; electronic publishing also lacks the dynamic presentation of visual graphics and the support for graphic interaction for complex multi-dimensional data.

6. CONCLUSION

In summary, this study proposes the following recommendations and prospects for the application and development of data visualization in LIS in China in the future:

(1) Broaden the idea of data visualization graphic expression. LIS researchers need to create new data presentation methods and incorporate more graphic styles to conform to different data characteristics. Simultaneously, different visualization tool functions are integrated to shift from single type of data graphics to "research problem" centered comprehensive data graphics expression and construction.

(2) Pay attention to data design and information aesthetics to improve the readability of graphics. The current data analysis teaching of LIS should include data design courses, strengthen the ability to use visual graphics to interpret data information, cultivate the quality of information aesthetics, pay attention to the reading experience, and reduce the obstacles of visual graphics in the reading process.

(3) Turn to dynamic and interactive data presentation mode. LIS should actively optimize how to publish scientific research results, integrate data visualization tools in the scientific research results publishing platform, and support researchers to upload dynamic graphics and visualization data during the submission process. This method can effectively enhance the expression of scientific research results and promote the interaction opportunities between readers and scientific research results, and differentiation meets the demand of readers.

(4) Restrained and effective visual expression. Summarizing the history of data visualization development, the wave of new visualizations in each period has never been late. However, it should be noted that data visualization is only a helper on the road of scientific research, not a prop to please users or decorate the facade. Visual skills, such as design, graphics, color, dynamics, and interaction, must ultimately serve rigorous scientific research. The only restrained and effective expression can produce the best data visualization practice effect.

AUTHORS' CONTRIBUTIONS

Zhi-Wei Liu: Data curation, Writing Original draft preparation. Ming-Yueh Tsay: Conceptualization, Supervision, Writing Reviewing and Editing.

REFERENCES

[1] McCormick, B. H., Visualization in scientific computing, Acm Sigbio Newsletter, 1988, pp. 15-21. DOI: https://doi.org/10.1145/43965.43966

[2] Wehrend, Stephen Charles, A categorization of scientific visualization techniques, University of Colorado, Colorado, 1984.

[3] CARD, Mackinlay. Readings in information visualization: using vision to think, Morgan Kaufmann, 1999.

[4] Notton, J. H. F., Ancient Egyptian gold refining, Gold Bulletin, 1974, pp. 50-56. DOI: https://doi.org/10.1007/bf03215038

[5] Funkhouser, H. G., A note on a tenth century graph, Osiris, 1936, pp. 260-262. DOI: https://doi.org/10.1086/368425

[6] Lulle, Raymond, Arbor scientiae, Google Book, 2000.

[7] Friendly M, Valero-Mora P, Ulargui J I., The First (Known) Statistical Graph: Michael Florent van Langren and the "Secret" of Longitude, American Statian, 2010, pp.174-184. DOI: https://doi.org/10.1198/tast.2010.09154.

[8] List of physical visualizations and related artifacts, http://dataphys.org/list/barbeu-dubourgs-machine-chronologique/, last accessed 2020/07/18.

[9] FERGUSON, The 1753 "Carte chronographique" of Jacques Barbeu-Dubourg, Princeton University Library Chronicle, 1991, pp. 190. DOI: https://doi.org/10.2307/26404421

[10] HULLIGER, Beat, Johann Heinrich Lambert, an admirable applied statistician, Bullet, Swiss Stat,2013, pp. 4-10.

[11] DUPAIN-TRIEL, Jean-Louis, Essai d'une table poléométrique, ou Amusement d'un amateur de plans sur les grandeurs de quelques villes, Dupain-Triel père, 1782.

[12] Robyn B D L., Quantitative Graphics in Statistics: A Brief History, American Statistician, 1978, pp. 1-11. DOI: https://doi.org/10.1080/00031305.1978.10479235

[13] Tufte E R, The Visual Display of Quantitative Information, American Journal of Physics, 2001.DOI: https://doi.org/10.1080/00401706.1987.10488194

[14] Dupin, Charles, Carte figurative de l'instruction populaire de la France, Bachelier, 1827.

[15] Robinson, Arthur H., The 1837 maps of henry drury harness, The Geographical Journal, 1995, pp.440-450. DOI: https://doi.org/10.2307/1791753

[16] GIEßMANN, Sebastian, Synchronisation im Diagramm Henry C. Beck und die Londoner Tube Map von 1933, Wilhelm Fink, 2013, pp.339-364. DOI: https://doi.org/10.30965/9783846748084_015

[17] FIRMAS, G., The FISHEYE view: A new look at structured files, Bell Laboratories Technical Memorandum, 1981.DOI: https://doi.org/10.5555/300679.300769

[18] WILLS, G., et al., Dynamic interactive graphics for spatially referenced data, Fortschritte der statistik-software 2, 1990, pp.278-287.

[19] BAUSCH, Paul, BUMGARDNER, Jim., Make a Flickr-style tag cloud, Flickr hacks, 2006, pp. 82-86.

[20] KRZYWINSKI, Martin, et al., Circos: an information aesthetic for comparative genomics, Genome research,2009, pp. 1639-1645. DOI: https://doi.org/10.1101/gr.092759.109.

[21] The 3 waves of data visualization: A brief history and predictions for the future, https://www.tableau.com/about/blog/2019/2/three-waves-data-visualization-brief-history-and-predictions-future-100830, last accessed 2020/08/25.

[22] Create Tableau Maps from Spatial Files, https://help.tableau.com/current/pro/desktop/en-us/maps_shapefiles.htm, last accessed 2020/08/25.

[23] Dang Ya ru, Structural modeling of citation network systems, Library and Information Service,1996, pp. 58-61. DOI: https://doi.org/CNKI:SUN:TSQB.0.1996-04-016

[24] Zhou Ning, Wen Yanping, Visual study of retrieval results, Journal of Library Science in China, 2002, pp.47-49.DOI: https://doi.org/10.3969/j.issn.1001-8867.2002.06.011

[25] Zhou Ning, Visual and Intelligent development direction of information organization and retrieval in the 21st century, Modern Library and Information Technology,2002, pp. 4-6 (2002). DOI: https://doi.org/10.11925/infotech.1003-3513.2002.04.02

[26] Luo Longyan, Preliminary Research on Information Retrieval based on Visualization Technology, Modern Library and Information Technology, 2002, pp.36-38. DOI: https://doi.org/10.3969/j.issn.1003-3513.2002.04.014

[27] Jing Peidong, Information Visualization - a New Field of Information Science research, Information Science, 2003, pp. 685-687. DOI: https://doi.org/10.3969/j.issn.1007-7634.2003.07.003

[28] Hu Liyong, Chen Dingquan. Visualization research of Citation Analysis. Journal of Intelligence, 2004, pp.78-79.

[29] Ou Yangxia, EAMOLA members' scientific research cooperation behaviors, Documentation, Information & Knowledge, 2006. pp. 33-37. DOI: https://doi.org/10.3969/j.issn.1003-2797.2006.06.008

[30] Wang Zixi, Ma Lei, Visualization of the Relationship between Words in Chinese Thesaurus. Modern Library and Information Technology, 2006, pp.86-88. DOI: https://doi.org/10.3969/j.issn.1003-7136.2006.02.009

[31] Hu zewen, Sun Jianjun, Wu Yishan, A Review on the Application of Knowledge Mapping in China, Library and Information Service, 2013, pp.131-137. DOI: https://doi.org/CNKI:SUN:TSQB.0.2013-03-025