

Seven clusters of data visualization articles in Scopus using social network analysis

Riki Satia Muharam^{a*}, Budiman Rusli^a, Heru Nurasa^a and Entang Adhy Muhtar^a

^aPublic Administration Department, Faculty of Social and Political Sciences, Universitas Padjadjaran, Indonesia

CHRONICLE

Article history:

Received: December 25, 2022
Received in revised format: March 2, 2023
Accepted: April 8, 2023
Available online: April 8, 2023

Keywords:

Data Visualization
Scopus
VOSviewer

ABSTRACT

The aim of this study was to analyse the bibliographic characteristics and content of articles on Data Visualization published in journals indexed by Scopus written by researchers from throughout the world. We conducted a bibliometric and content analysis of publication in the Scopus database. We only retrieved articles written in English. We conducted content analysis using the VOSviewer software and visualized the co-occurrence of keywords and bibliographic coupling of sources and countries. Following the study protocol, we found 862 articles on Data Visualization over the past 30 years. The most productive journal that published these articles was Lecture Notes In Computer Science (n=32). The most productive country was the United States (n=305). Based on citations, the most influential authors, and journals were Thorvaldsdóttir et al., (2013) [Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192.] (n=4699), and IEE Transactions on Visualization and Computer Graphics (n=656). The keywords of research on Data Visualization formed 7 clusters (e.g. Data Visualization, Visualization, and Human). From a global perspective, Data Visualization research in the past 30 years has increased significantly. There were European published journals nominated publications. Thus, Asian countries need to conduct more active research on this topic.

© 2023 by the authors; licensee Growing Science, Canada.

1. Introduction

This article discusses the discourse Data Visualization. The discourse understanding is inseparable from bibliometric analysis (Lee, 2020; Mifrah, 2020; Omoregbe et al., 2020; Saravanan & Dominic, 2014), referring to the incorporation of various frameworks and methods to analyze citations from scientific publications. Such attempts lead to the development of different metrics to gain insight into the intellectual structure of a broad academic discipline and to evaluate the impact of a particular field of study (Akhavan et al., 2016; Putera et al., 2020). Although some researchers have produced Data Visualization articles, we have not found research in bibliometric articles on Data Visualization utilized social network analysis. The aim of this article was to provide useful data for understanding global publication trends regarding Data Visualization. This study aimed to analyze the bibliographic characteristics and trends of articles on Data Visualization published in journals indexed in Scopus written by researchers from throughout the world and to conduct an analysis of keyword co-occurrence using VOSviewer.

2. Methods

This study did not involve human subjects; therefore, neither institutional review board approval nor informed consent was needed. This study was a descriptive and bibliometric analysis based on a literature database. The data in this study were retrieved from the Scopus database. To obtain the necessary data, this study used the keyword “Data Visualization” in the

* Corresponding author.

E-mail address: riki19001@mail.unpad.ac.id (R. S. Muharam)

title, abstracts, and author’s keywords. In this step, we found 862 articles. In the next step, we downloaded the articles from the Scopus database and analyzed the 862 articles that had been sorted by relevance. In this study, the metadata and refined Scopus result values were retrieved in the Csv dataset format. However, before the bibliometric analysis, the consistency and reliability of the data were checked to address issues such as a lack of consistency in country names and keywords. The data were also standardized to ensure consistency regarding key words that sometimes appeared in singular or plural, abbreviations, or other forms. The data obtained from the Scopus database were analyzed using VOSviewer software, and simple statistics were calculated using Microsoft Excel.

3. Results

The development of ‘Data Visualization’ publications indicates an increase, from 2019 to 2020. The trend analysis (see Fig. 1) indicates that before 1991, there were no ‘Data Visualization’ publications on the Scopus database.

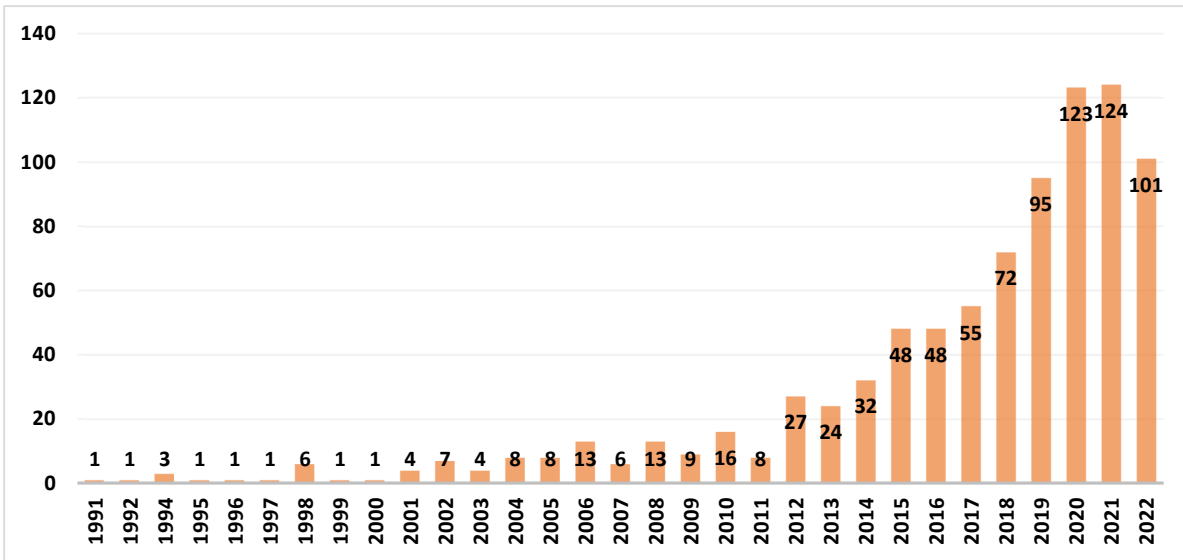


Fig. 1. Publication Trend “Data Visualization” (Source : Processed by Author)

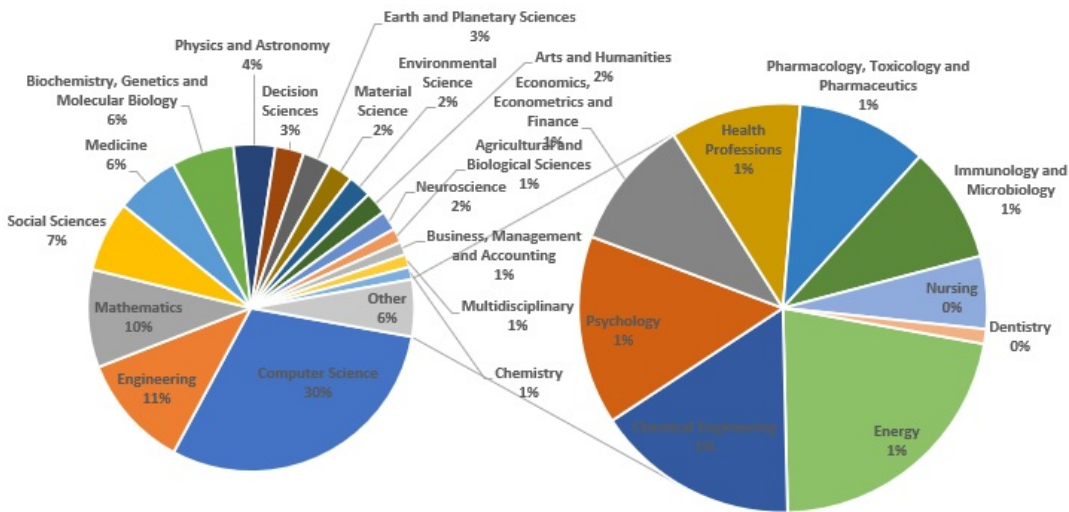


Fig. 2. Publication Trend “Data Visualization” based on Subject Area (Source : Processed by Author)

Based on a search with the keyword “Data Visualization”, the result showed approximately 862 documents. Most articles were listed under Computer Science (n=466, 30%), Engineering (n=175, 11%), Mathematics (n=150, 10%), Social Sciences (n=108, 7%), and Medicine (n=98, 6%). The full distribution of Data Visualization articles across subject areas is shown in Fig. 2. According to VOSViewer, the articles were published in 160 different journals. The highest number of articles were published in Lecture Notes In Computer Science, with 32 publications, followed by IEEE Transactions on Visualization and Computer Graphics (n=29), Journal of Physics Conference Series (n=22), Bioinformatics (n=17), and Computer Graphics Forum (n=12). The other most productive journals with the most publications are shown in Table 1.

Table 1
The most production journals based on the number of publications

Rank	Source	No. Of Publication	Country
1 st	Lecture Notes In Computer Science	32	Germany
2 nd	IEEE Transactions on Visualization and Computer Graphics	29	United States
3 rd	Journal of Physics Conference Series	22	United Kingdom
4 th	Bioinformatics	17	United Kingdom
5 th	Computer Graphics Forum	12	United Kingdom
6 th	BMC Bioinformatics	11	United Kingdom
7 th	Conference On Human Factors in Computing Systems Proceedings	11	United States
8 th	IEEE Access	11	United States
9 th	IEEE Computer Graphics and Applications	10	United States
10 th	ACM International Conference Proceeding Series	8	United States
11 th	Proceeding of The International Conference on Information Visualization	8	Portugal
12 th	Frontiers In Neuroinformatics	7	Switzerland
13 th	Iop Conference Series Earth and Environmental Science	6	United Kingdom
14 th	Iop Conference Series Materials Science and Engineering	6	United Kingdom
15 th	Nucleic Acids Research	6	United Kingdom
16 th	Plos One	6	United States
17 th	Procedia Computer Science	6	Netherlands
18 th	Applied Sciences Switzerland	5	Switzerland
19 th	IFIP Advances In Information And Communication Technology	5	United States
20 th	International Archives of The Photogrammetry Remote Sensing and Spatial Information Sciences ISPRS Archives	5	Germany

Table 2 lists the most influential authors based on citations recorded by the Scopus database. The most influential author was Thorvaldsdóttir et al., (2013), with 4699 citations, followed Li, (1992) (n=445), Allen et al., (2021) (n=397), Francisco et al., (2012) (n=375), and Van Wijk, (1991) (n=292). Table 3 presents the influential source (i.e. journals) based on citations. IEE Transactions on Visualization and Computer Graphics (n=656) was the most influential journal, followed by Bioinformatics (n=540), BMC Bioinformatics (n=533), Lecture Notes in Computer Science (n=387), and Computer Graphics Forum (n=328).

Table 2
The most influential authors and source based on citation analysis

Rank	Authors	Title	Source title	Cited by
1 st	Thorvaldsdóttir et al., (2013)	Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration	Briefings in Bioinformatics	4699
2 nd	Li, (1992)	On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma	Journal of the American Statistical Association	445
3 rd	Allen et al., (2021)	Raincloud plots: A multi-platform tool for robust data visualization [version 1; peer review: 2 approved]	Welcome Open Research	397
4 th	Francisco et al., (2012)	PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods	BMC Bioinformatics	375
5 th	Van Wijk, (1991)	Spot noise texture synthesis for data visualization	Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1991	292
6 th	Aguado et al., (2019)	The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stel-	Astrophysical Journal, Supplement Series	248
7 th	Peng et al., (2004)	Clutter reduction in multi-dimensional data visualization using dimension reordering	Proceedings - IEEE Symposium on Information Visualization, INFO VIS	245
8 th	Williamson, (2016)	Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments	Journal of Education Policy	232
9 th	Djorgovski et al., (2018)	Immersive and collaborative data visualization using virtual reality platforms	Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014	210
10 th	Marcus et al., (2013)	Human Connectome Project informatics: Quality control, database services, and data visualization	NeuroImage	210
11 th	Buja et al., (2008)	Data visualization with multidimensional scaling	Journal of Computational and Graphical Statistics	200
12 th	Bhatia et al., (2009)	Software tool for researching annotations of proteins: Open-source protein annotation software with data visualization	Analytical Chemistry	188
13 th	Y. Q. Wang, (2014)	MeteoInfo: GIS software for meteorological data visualization and analysis	Meteorological Applications	170
14 th	Gatto & Lilley, (2012)	Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation	Bioinformatics	166
15 th	Swayne et al., (2003)	GGobi: Evolving from XGobi into an extensible framework for interactive data visualization	Computational Statistics and Data Analysis	153
16 th	Bishop & Tipping, (1998)	A hierarchical latent variable model for data visualization	IEEE Transactions on Pattern Analysis and Machine Intelligence	148
17 th	C. Wang et al., (2008)	Importance-driven time-varying data visualization	IEEE Transactions on Visualization and Computer Graphics	136
18 th	Childs et al., (2005)	A contract based system for large data visualization	Proceedings of the IEEE Visualization Conference	129
19 th	Nolte et al., (2018)	Instant Clue: A Software Suite for Interactive Data Visualization and Analysis	Scientific Reports	111
20 th	Claessen & Van Wijk, (2011)	Flexible linked axes for multivariate data visualization	IEEE Transactions on Visualization and Computer Graphics	107

Source : Processed by Author

Table 3 The most influential countries and source based on citation analysis.

Rank	Source	Cited by	Country	Quartile
1 st	IEE Transactions on Visualization and Computer Graphics	656	United States	Q1
2 nd	Bioinformatics	540	United Kingdom	Q1
3 rd	BMC Bioinformatics	533	United Kingdom	Q1
4 th	Lecture Notes in Computer Science	387	Germany	Q2
5 th	Computer Graphics Forum	328	United Kingdom	Q1
6 th	IEE Computer Graphics and Applications	268	United States	Q2
7 th	Frontier in Neuroinformatic	147	Switzerland	Q1
8 th	Conference on Human Factors in Computing Systems	126	United States	-
9 th	Proceedings of the International Conference on Information Visualization	123	Portugal	-
10 th	Nucleic Acids Research	110	United Kingdom	Q1

Source : Processed by Author

There were four journals from the United Kingdom, three journals from the United States, and one journal each from Switzerland, Portugal, and Germany. In the period 2020 to 2022, the United States was the country with the most publications on Data Visualization, with 305 articles, followed by China with 98 articles.

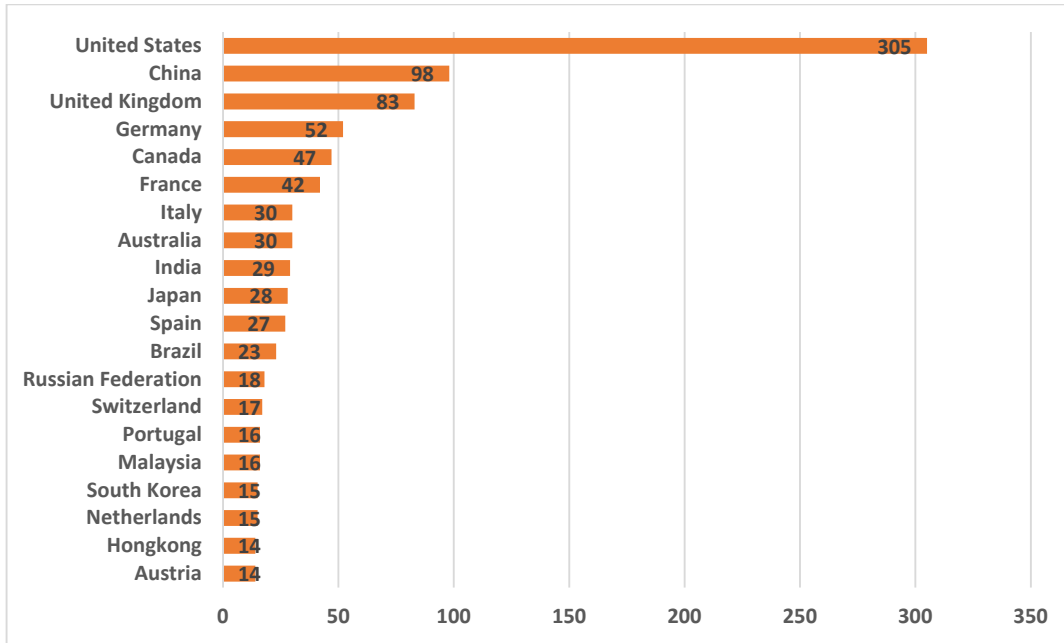


Fig. 3. Top 20 countries with publication of Data Visualization

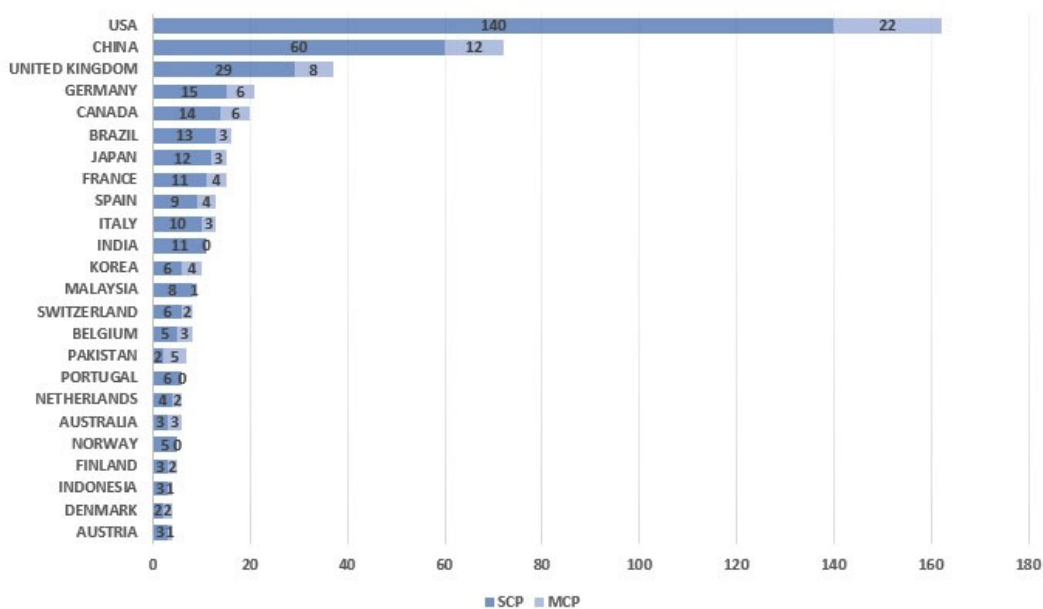


Fig. 4. Corresponding Authors

China, India, Japan, Malaysia, South Korea and Hongkong were the Asian countries ranked in the top 20 countries in terms of the most Data Visualization publications. These Six Asian countries ranked two, nine, ten, sixteen, eighteen and nineteen. The top 20 countries can be seen in Fig. 3. Fig. 4 shows Corresponding Authors of Pandemic Covid-19. The highest number of articles were in the United States, Followed by China, and the United Kingdom. A content analysis was performed of the 862 publications sorted by relevance. Next, we performed a co-occurrence analysis with VOSviewer, using the “all keyword” analysis unit and the “full counting” method. We limited the frequency of keyword occurrence to 5 times; out of 6179 keywords VOSviewer found 318 keywords that met the threshold. The results of this analysis are presented in Fig. 5. Data Visualization (593), Visualization (459), and Human (136) were the top three keywords that appeared most frequently. Moreover, we found seven clusters in this analysis. Fig. 5 shows these keywords divided into 7 clusters (each with a different number of keywords), which are represented by colors. The first cluster (red, 107 keywords) focused on Data Visualization, Visualization, Information Systems, Big Data and Decision Making. The second cluster (green, 55 keywords) centered on Software, Computer Graphics, Procedures, Priority Journal, and Data Processing. The third cluster (blue, 55 keywords) related to Data Mining, Algorithms, Data Reduction, Data Structures, and Mapping.

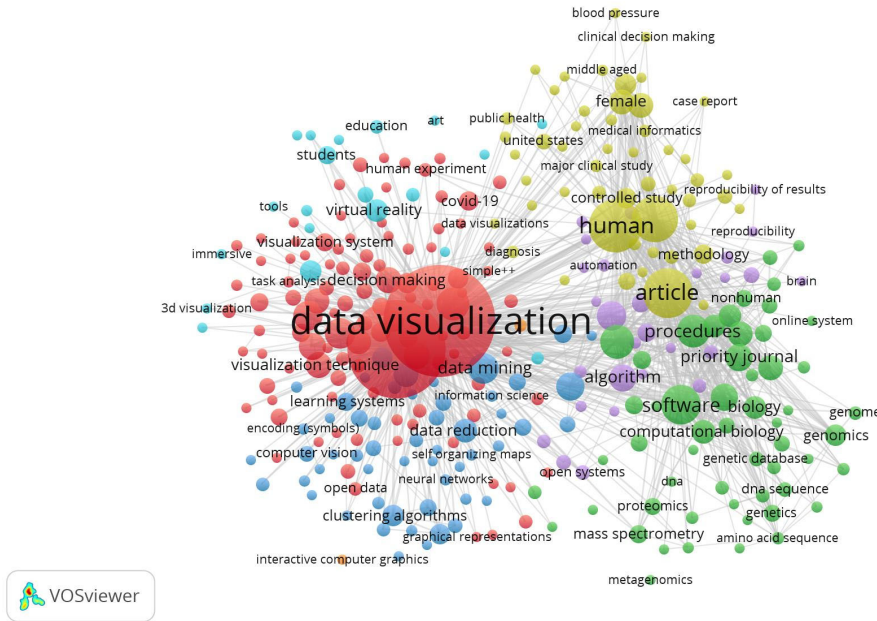


Fig. 5. Network visualization of data visualization articles published globally in the Scopus database (Source: Processed by Author)

The fourth cluster (yellow, 51 keywords) related to Human, Article, Methodology, Factual Database, and Female. The five clusters (purple, 28 keywords) focused on Data Analysis, Quality Control, Quantitative Analysis, Data Sharing, and Automation. The sixth cluster (light blue, 20 keywords) dealt with Three Dimensional, Virtual Reality, Students, Education, and Tools. The seventh cluster (orange, 2 keywords) focused on Interactive Computer Graphics and Research.

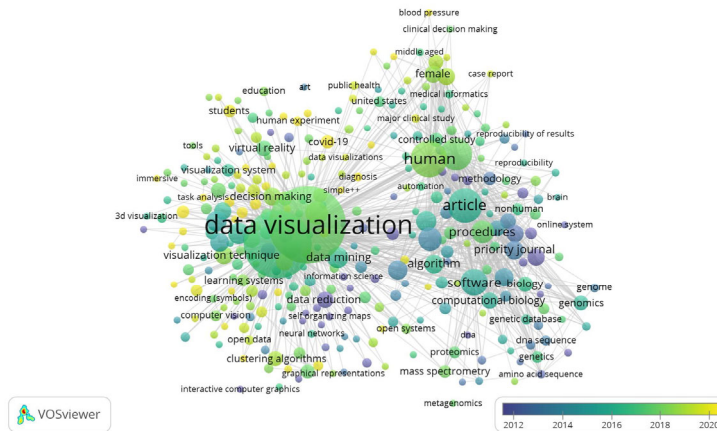


Fig. 6. Overlay visualization of global data visualization articles

Fig. 6 shows an overlay visualization of Data Visualization literature with the average number of publications from 2012 to 2020. There was a shift in topics; around 2014, the literature on Data Visualization contained extensive discussions of the terms “Information Visualization”, “Information Systems” and “Data Mining”, and then the last 3 years discussed “Data Analytics”, “Open Source”, and “Covid-19”.

4. Discussion

Based on data from Scopus, the publication trends, journal performance, content analysis, and bibliographic coupling of countries and sources were analysed for research on Data Visualization issues throughout the world. The current study focused on articles published in Data Visualization. This study aimed to provide information on the status of publications in these fields. A total of 862 studies published were recorded in the Scopus database. The data showed the rapidity of article publications and the responsiveness of researchers in analyzing Data Visualization around the world. However, limited research from a global perspective on Data Visualization in the past 3 years has discussed “Data Analytics”, “Open Source”, and “Covid-19” and its relationship with governance within the scope of social science. Based on Fig. 3, the most productive and influential country was the United States followed by China. Although countries from Europe dominated the top 20 countries with the most publications by affiliated researchers, China, India, Japan, Malaysia, South Korea and Hongkong is the country from Asia in the top 20. The current study has limitations, we only retrieved studies from Scopus and did not use other source such as Web of Science, Crossref, or PubMed Central. Finally, we did not use other analyses in VOSviewer, such as co-citation or co-authorship. Thus, we hope that bibliometric research on this topic will expand in terms of the databases used, the subject areas, and the analyses conducted in order to provide a broader overview of the issue.

5. Conclusion

In the past 30 years, global research on Data Visualization has increased significantly. The theme of research on Data Visualization related to local government could be interesting for future discussions. There are also opportunities to foster discussion about Data Visualization in social science journals related to public administration. Finally, Europe dominated this field in terms of publications, while research from Asia on this topic remains limited, and further research is therefore necessary.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

Author delivered many thanks to Prof. Dr. Rina Indriastuti, S.E., M.SIE as Rector of Universitas Padjadjaran, Indonesia, Dr. R. Widya Setiabudi Sumadinata, M.T., M.Si. as Dean of Faculty of Social and Political Sciences and the Direktorat Riset Pengabdian Masyarakat (DRPM).

References

- Aguado, D. S., Ahumada, R., Almeida, A., Anderson, S. F., Andrews, B. H., Anguiano, B., Ortíz, E. A., Aragón-Salamanca, A., Argudo-Fernández, M., Aubert, M., Avila-Reese, V., Badenes, C., Barboza Rembold, S., Barger, K., Barrera-Ballesteros, J., Bates, D., Bautista, J., Beaton, R. L., Beers, T. C., ... Zou, H. (2019). The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library. *The Astrophysical Journal Supplement Series*, 240(2), 23. <https://doi.org/10.3847/1538-4365/aaf651>
- Akhavan, P., Ebrahim, N. A., Fetrati, M. A., & Pezeshkan, A. (2016). Major trends in knowledge management research: a bibliometric study. *Scientometrics*, 107(3), 1249–1264. <https://doi.org/10.1007/s11192-016-1938-x>
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4(January), 63. <https://doi.org/10.12688/wellcomeopenres.15191.2>
- Bhatia, V. N., Perlman, D. H., Costello, C. E., & McComb, M. E. (2009). Software tool for researching annotations of proteins: Open-source protein annotation software with data visualization. *Analytical Chemistry*, 81(23), 9819–9823. <https://doi.org/10.1021/ac901335x>
- Bishop, C. M., & Tipping, M. E. (1998). A hierarchical latent variable model for data visualization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 281–293. <https://doi.org/10.1109/34.667885>
- Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, 17(2), 444–472. <https://doi.org/10.1198/106186008X318440>
- Childs, H., Brugger, E., Bonnell, K., Meredith, J., Miller, M., Whitlock, B., & Max, N. (2005). A contract based system for large data visualization. *Proceedings of the IEEE Visualization Conference*, 25. <https://doi.org/10.1109/VIS.2005.3>
- Claessen, J. H. T., & Van Wijk, J. J. (2011). Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2310–2316. <https://doi.org/10.1109/TVCG.2011.201>
- Djorgovski, S. G., Donalek, C., Lombeyda, S., Davidoff, S., & Amori, M. (2018). Immersive and Collaborative Data

- Visualization and Analytics Using Virtual Reality. *American Geophysical Union, Fall Meeting 2018, Abstract #IN53B-01*, 609–614. <http://adsabs.harvard.edu/abs/2018AGUFMIN53B.01D>
- Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., & Carriço, J. A. (2012). PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13(1). <https://doi.org/10.1186/1471-2105-13-87>
- Gatto, L., & Lilley, K. S. (2012). Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*, 28(2), 288–289. <https://doi.org/10.1093/bioinformatics/btr645>
- Lee, D. (2020). Bibliometric Analysis of Korean Journals in Arts and Kinesiology – from the Perspective of Authorship. *Journal of Information Science Theory and Practice*, 8(3), 15–29. <https://doi.org/10.1633/JISTaP.2020.8.3.2>
- Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *Journal of the American Statistical Association*, 87(420), 1025–1039. <https://doi.org/10.1080/01621459.1992.10476258>
- Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., Barch, D. M., Archie, K. A., Burgess, G. C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., ... Van Essen, D. C. (2013). Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage*, 80, 202–219. <https://doi.org/10.1016/j.neuroimage.2013.05.077>
- Mifrah, S. (2020). Toward a Semantic Graph of Scientific Publications: A Bibliometric Study. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 3323–3330. <https://doi.org/10.30534/ijatse/2020/129932020>
- Nolte, H., MacVicar, T. D., Tellkamp, F., & Krüger, M. (2018). Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Scientific Reports*, 8(1), 6–13. <https://doi.org/10.1038/s41598-018-31154-6>
- Omogrebe, O., Mustapha, A. N., Steinberger-Wilckens, R., El-Kharouf, A., & Onyeaka, H. (2020). Carbon capture technologies for climate change mitigation: A bibliometric analysis of the scientific discourse during 1998–2018. *Energy Reports*, 6, 1200–1212. <https://doi.org/10.1016/j.egy.2020.05.003>
- Peng, W., Ward, M. O., & Rundensteiner, E. A. (2004). Clutter reduction in multi-dimensional data visualization using dimension reordering. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, 89–96. <https://doi.org/10.1109/INFVIS.2004.15>
- Putera, P. B., Suryanto, S., Ningrum, S., & Widianingsih, I. (2020). A bibliometric analysis of articles on innovation systems in Scopus journals written by authors from Indonesia, Singapore, and Malaysia. *Science Editing*, 7(2), 177–183. <https://doi.org/10.6087/KCSE.214>
- Saravanan, G., & Dominic, J. (2014). A Ten-year Bibliometric Analysis of Research Trends in Three Leading Ecology Journals during 2003-2012. *Journal of Information Science Theory and Practice*, 2(3), 40–54. <https://doi.org/10.1633/jistap.2014.2.3.4>
- Swayne, D. F., Temple Lang, D., Buja, A., & Cook, D. (2003). GGobi: Evolving from XGobi into an extensible framework for interactive data visualization. *Computational Statistics and Data Analysis*, 43(4), 423–444. [https://doi.org/10.1016/S0167-9473\(02\)00286-4](https://doi.org/10.1016/S0167-9473(02)00286-4)
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Van Wijk, J. J. (1991). Spot noise texture synthesis for data visualization. *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1991*, 25, 309–318. <https://doi.org/10.1145/122718.122751>
- Wang, C., Yu, H., & Ma, K. L. (2008). Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1547–1554. <https://doi.org/10.1109/TVCG.2008.140>
- Wang, Y. Q. (2014). MeteoInfo: GIS software for meteorological data visualization and analysis. *Meteorological Applications*, 21(2), 360–368. <https://doi.org/10.1002/met.1345>
- Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, 31(2), 123–141. <https://doi.org/10.1080/02680939.2015.1035758>



© 2023 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).