Data Visualization Techniques for Research Publications and Scientific Computational Methods using Python: A Review

Thirumalesh[1*], Ashok A S[1], Meenakshi H[1], Aruna R[1], Shashidhar S N[1], Madhu N R[1], Reshma S R[1]

R L Jalappa Institute of Technology, Doddaballapur, Bangalore, India[1]

Corresponding author:thirumalesh@rljit.in[1]

Abstract: *Data visualization plays a crucial role in research publications, enabling researchers to communicate their findings effectively and enhance the understanding of complex data. Python, with its extensive libraries and versatile capabilities, has emerged as a popular choice for data visualization and scientific computation. This article explores various data visualization techniques for research publications and highlights the scientific computation methods facilitated by Python. We discuss the importance of data visualization in research publications and the advantages offered by Python for visualizing scientific data. Additionally, we explore Python's role in scientific computation, encompassing areas such as physical, chemical and biological science computations, numerical computing, data analysis, statistical modeling, and machine learning. Also in this review article we tried to emphasize the significance of reproducibility and open science practices facilitated by Python's code-centric nature, enabling researchers to share their work and foster collaboration. We conclude by highlighting the vibrant Python community and the availability of resources for researchers to learn, contribute, and stay updated with the latest developments in data visualization and scientific computation. Overall, this article demonstrates the value of data visualization techniques and Python's contributions to research publications and scientific computation methods, empowering researchers to gain insights, communicate results, and advance knowledge in their respective fields*

Keywords: Python, Data Visualization, Python Computation tools, Python Scientific Libraries

REFERENCES

[1]. Pine, David J. Introduction to Python for science and engineering. CRC press, 2019.

[2]. Millman, K. Jarrod, and Michael Aivazis. "Python for scientists and engineers." Computing in science & engineering 13, no. 2 (2011): 9-12.

[3]. Embarak, Dr Ossama, Embarak, and Karkal. Data analysis and visualization using python. Berkeley, CA, USA: Apress, 2018.

[4]. Cao, Shengjia, Yunhan Zeng, Shangru Yang, and Songlin Cao. "Research on Python data visualization technology." In Journal of Physics: Conference Series, vol. 1757, no. 1, p. 012122. IOP Publishing, 2021.

[5]. Jhamb, Shreya, Richa Gupta, Vinod Kumar Shukla, Insha Mearaj, and Princy Agarwal. "Understanding Complexity in Language Learning Through Data Visualization Using Python." In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), pp. 268-274. IEEE, 2020.

[6]. Sahoo, Kabita, Abhaya Kumar Samal, Jitendra Pramanik, and Subhendu Kumar Pani. "Exploratory data analysis using Python." International Journal of Innovative Technology and Exploring Engineering 8, no. 12 (2019): 4727-4735.

[7]. Barrett, Paul, John Hunter, J. Todd Miller, J-C. Hsu, and Perry Greenfield. "matplotlib--A Portable Python Plotting Package." In Astronomical data analysis software and systems XIV, vol. 347, p. 91. 2005.

[8]. Yim, Aldrin, Claire Chung, and Allen Yu. Matplotlib for Python Developers: Effective techniques for data visualization with Python. Packt Publishing Ltd, 2018.

[9]. Sial, Ali Hassan, Syed Yahya Shah Rashdi, and Abdul Hafeez Khan. "Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python." International Journal 10, no. 1 (2021).

[10]. Van Der Donckt, Jonas, Jeroen Van der Donckt, Emiel Deprost, and Sofie Van Hoecke. "Plotly-resampler: Effective visual analytics for large time series." In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 21-25. IEEE, 2022.

[11]. Harper, Charlie. "Visualizing Data with Bokeh and Pandas." *The Programming Historian* (2018).

[12]. Python Documentation: https://docs.python.org/3/

[13]. Python community: https://www.python.org/community/

[14]. Van Der Walt, Stefan, S. Chris Colbert, and Gael Varoquaux. "The NumPy array: a structure for efficient numerical computation." *Computing in science & engineering* 13, no. 2 (2011): 22-30.

[15]. McKinney, Wes, and P. D. Team. "Pandas-Powerful python data analysis toolkit." *Pandas—Powerful Python Data Analysis Toolkit* 1625 (2015).

[16]. Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature methods* 17, no. 3 (2020): 261-272.

[17]. GitHub: https://github.com/

[18]. Cosentino, Valerio, Javier Luis, and Jordi Cabot. "Findings from GitHub: methods, datasets and limitations." In *Proceedings of the 13th International Conference on Mining Software Repositories*, pp. 137-141. 2016.

[19]. Erdogan Taskesen: Creating beautiful stand-alone interactive D3 charts with Python, https://towardsdatascience.com/creating-beautiful-stand-alone-interactive-d3-charts-with-python-804117cb95a7

[20]. Sullivan, C., and Alexander Kaszynski. "PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK)." Journal of Open Source Software 4, no. 37 (2019): 1450.

[21]. Schreiber, Andreas, Lynn von Kurnatowski, Annika Meinecke, and Claas de Boer. "An interactive dashboard for visualizing the provenance of software development processes." In *2021 Working Conference on Software Visualization (VISSOFT)*, pp. 100-104. IEEE, 2021.

[22]. Lemenkova, Polina. "Python libraries matplotlib, seaborn and pandas for visualization geo-spatial datasets generated by QGIS." *Analele stiintifice ale Universitatii" Alexandru Ioan Cuza" din Iasi-seriaGeografie* 64, no. 1 (2020): 13-32.

[23]. Waskom, Michael L. "Seaborn: statistical data visualization." *Journal of Open Source Software* 6, no. 60 (2021): 3021.

[24]. Zhang, Tianyu, and Long Mei. "Analysis and research on computer visualization in data science with bokeh and JavaScript." In *Journal of Physics: Conference Series*, vol. 2033, no. 1, p. 012154. IOP Publishing, 2021.

[25]. Ding, Wubin, David Goldberg, and Wanding Zhou. "PyComplexHeatmap: A Python package to visualize multimodal genomics data." *iMeta*: e115.

[26]. Sunitha, Gurram, A. V. Sriharsha, OlimjonYalgashev, and Islom Mamatov. "Interactive Visualization WithPlotly Express." In *Advanced Applications of Python Data Structures and Algorithms*, pp. 182-206. IGI Global, 2023.

[27]. Khorasani, Mohammad, Mohamed Abdou, and Javier Hernández Fernández. "Streamlit Use Cases." In *Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework*, pp. 309-361. Berkeley, CA: Apress, 2022.

[28]. Kanazawa, Naoki, Daniel J. Egger, Yael Ben-Haim, Helena Zhang, William E. Shanks, Gadi Aleksandrowicz, and Christopher J. Wood. "Qiskit Experiments: A Python package to characterize and calibrate quantum computers." Journal of Open Source Software 8, no. 84 (2023): 5329.

[29]. Anis, M. S., Mitchell, A., Abraham, H., Offei, A., Agarwal, R., Agliardi, G., Aharoni, M., Akhalwaya, I. Y., Aleksandrowicz, G., & et al. (2021). Qiskit: An open-source framework for quantum computing. https://doi.org/10.5281/zenodo.2573505.

[30]. Kaiser, Sarah C., and Christopher Granade. Learn quantum computing with python and Q#: A hands-on approach. Simon and Schuster, 2021.

[31]. Koch, Daniel, Laura Wessing, and Paul M. Alsing. "Introduction to coding quantum algorithms: A tutorial series using Pyquil." arXiv preprint arXiv:1903.05195 (2019).

[32]. Hancock, Andrew, Austin Garcia, Jacob Shedenhelm, Jordan Cowen, and Calista Carey. "Cirq: A Python Framework for Creating, Editing, and Invoking Quantum Circuits."

[33]. IBM: https://www.ibm.com/quantum

[34]. Steffen, Matthias, David P. DiVincenzo, Jerry M. Chow, Thomas N. Theis, and Mark B. Ketchen. "Quantum computing: An IBM perspective." IBM Journal of Research and Development 55, no. 5 (2011): 13-1.

[35]. Google Quantum AI: https://quantumai.google/

[36]. Courtland, Rachel. "Google aims for quantum computing supremacy [news]." IEEE Spectrum 54, no. 6 (2017): 9-10.

[37]. Sun, Qiming, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu et al. "PySCF: the Python-based simulations of chemistry framework." Wiley Interdisciplinary Reviews: Computational Molecular Science 8, no. 1 (2018): e1340.

[38]. Smith, Daniel GA, Lori A. Burns, Andrew C. Simmonett, Robert M. Parrish, Matthew C. Schieber, Raimondas Galvelis, Peter Kraus et al. "PSI4 1.4: Open-source software for high-throughput quantum chemistry." The Journal of chemical physics 152, no. 18 (2020).

[39]. Muller, R. "Pyquante: Python quantum chemistry." URL http://pyquante. sourceforge. net (2017).

[40]. Landrum, Greg. "Rdkit documentation." *Release* 1, no. 1-79 (2013): 4.

[41]. Kasim, Muhammad F., Susi Lehtola, and Sam M. Vinko. "DQC: A Python program package for differentiable quantum chemistry." The Journal of chemical physics 156, no. 8 (2022).

[42]. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.

[43]. Singh, Pramod, Avinash Manure, Pramod Singh, and Avinash Manure. "Introduction to tensorflow 2.0." *Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python* (2020): 1-24.

[44]. Goldsborough, Peter. "A tour of tensorflow." *arXiv preprint arXiv:1610.01178* (2016).

[45]. Manaswi, Navin Kumar, and Navin Kumar Manaswi. "Understanding and working with Keras." *Deep learning with applications using Python: Chatbots and face, object, and speech recognition with TensorFlow and Keras* (2018): 31-43.

[46]. Grattarola, Daniele, and Cesare Alippi. "Graph neural networks in tensorflow and keras with spektral [application notes]." *IEEE Computational Intelligence Magazine* 16, no. 1 (2021): 99-106.

[47]. Robitaille, Thomas P., Erik J. Tollerud, Perry Greenfield, Michael Droettboom, Erik Bray, Tom Aldcroft, Matt Davis et al. "Astropy: A community Python package for astronomy." *Astronomy & Astrophysics* 558 (2013): A33.

[48]. Robitaille, Thomas, and Eli Bressert. "APLpy: astronomical plotting library in Python." Astrophysics Source Code Library (2012): ascl-1208.

[49]. Chapman, Brad, and Jeffrey Chang. "Biopython: Python tools for computational biology." *ACM Sigbio Newsletter* 20, no. 2 (2000): 15-19.

[50]. Pulli, Kari, Anatoly Baksheev, Kirill Kornyakov, and Victor Eruhimov. "Real-time computer vision with OpenCV." *Communications of the ACM* 55, no. 6 (2012): 61-69.

[51]. Hagberg, Aric, and Drew Conway. "Networkx: Network analysis with python." *URL: https://networkx. github. io* (2020).

[52]. Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck. "Probabilistic programming in Python using PyMC3." *PeerJ Computer Science* 2 (2016): e55.

[53]. Michaud-Agrawal, Naveen, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations." *Journal of computational chemistry* 32, no. 10 (2011): 2319-2327.

[54]. Clewley, Robert. "Hybrid models and biological model reduction with PyDSTool." (2012): e1002628.

[55]. VanderPlas, Jacob, Brian Granger, Jeffrey Heer, Dominik Moritz, KanitWongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. "Altair: interactive statistical visualizations for Python." Journal of open source software 3, no. 32 (2018): 1057.

[56]. Helmus, Jonathan J., and Christopher P. Jaroniec. "Nmrglue: an open source Python package for the analysis of multidimensional NMR data." Journal of biomolecular NMR 55 (2013): 355-367.

[57]. Carter, Faustin Wirkus, Trupti S. Khaire, Valentyn Novosad, and Clarence L. Chang. "scraps: An open-source python-based analysis package for analyzing and plotting superconducting resonator data." IEEE Transactions on Applied Superconductivity 27, no. 4 (2016): 1-5.

[58]. Roberts, Wade, Gustavious P. Williams, Elise Jackson, E. James Nelson, and Daniel P. Ames. "Hydrostats: A Python package for characterizing errors between observed and predicted time series." Hydrology 5, no. 4 (2018): 66.

[59]. Herath, Uthpala, Pedram Tavadze, Xu He, Eric Bousquet, Sobhit Singh, Francisco Muñoz, and Aldo H. Romero. "PyProcar: A Python library for electronic structure pre/post-processing." Computer Physics Communications 251 (2020): 107080.

[60]. Wasser, Leah, Maxwell B. Joseph, Joe McGlinchy, Jenny Palomino, Nathan Korinek, Chris Holdgraf, and Tim Head. "EarthPy: A Python package that makes it easier to explore and plot raster and vector data using open source Python tools." Journal of Open Source Software 4, no. 43 (2019): 1886.

[61]. Millman, K. Jarrod, and Matthew Brett. "Analysis of functional magnetic resonance imaging in Python." Computing in Science & Engineering 9, no. 3 (2007): 52-55.

[62]. Habowski, Amber N., T. J. Habowski, and M. L. Waterman. "GECO: gene expression clustering optimization app for non-linear data visualization of patterns." BMC bioinformatics 22 (2021): 1-13.

[63]. Meunier, David, Annalisa Pascarella, Dmitrii Altukhov, Mainak Jas, Etienne Combrisson, Tarek Lajnef, Daphné Bertrand-Dubois et al. "NeuroPycon: An open-source python toolbox for fast multi-modal and reproducible brain connectivity pipelines." NeuroImage 219 (2020): 117020.

[64]. Gouws, Andre D., Will Woods, Rebecca E. Millman, Antony B. Morland, and Gary GR Green. "DataViewer3D: an open-source, cross-platform multi-modal neuroimaging data visualization tool." Frontiers in Neuroinformatics 3 (2009): 344.

[65]. Franklin, Janet, Josep M. Serra-Diaz, Alexandra D. Syphard, and Helen M. Regan. "Big data for forecasting the impacts of global change on plant communities." Global Ecology and Biogeography 26, no. 1 (2017): 6-17.

[66]. Winkler, Karina, Richard Fuchs, Mark Rounsevell, and Martin Herold. "Global land use changes are four times greater than previously estimated." Nature communications 12, no. 1 (2021): 2501.

[67]. Lock, Oliver, Tomasz Bednarz, and Christopher Pettit. "The visual analytics of big, open public transport data–a framework and pipeline for monitoring system performance in Greater Sydney." Big Earth Data 5, no. 1 (2021): 134-159.

[68]. Klein, Karsten, Sabrina Jaeger, Jörg Melzheimer, Bettina Wachter, Heribert Hofer, Artur Baltabayev, and Falk Schreiber. "Visual analytics of sensor movement data for cheetah behaviour analysis." Journal of Visualization 24 (2021): 807-825.

[69]. Sibolla, Bolelang H., Serena Coetzee, and Terence L. Van Zyl. "A framework for visual analytics of spatio-temporal sensor observations from data streams." ISPRS International Journal of Geo-Information 7, no. 12 (2018): 475.

[70]. Klein, Karsten, Sabrina Jaeger, Jörg Melzheimer, Bettina Wachter, Heribert Hofer, Artur Baltabayev, and Falk Schreiber. "Visual analytics for cheetah behaviour analysis." In Proceedings of the 12th International Symposium on Visual Information Communication and Interaction, pp. 1-8. 2019.

[71]. Quiroz, Dorys, Byron Guanochanga, Walter Fuertes, Diego Benítez, Jenny Torres, Freddy Tapia, and Theofilos Toulkkeridis. "Visual analytics for the reduction of air pollution on real-time data derived from WSN." In Developments and Advances in Defense and Security: Proceedings of MICRADS 2019, pp. 109-119. Springer Singapore, 2020.

[72]. Kehl, Christian, Erik van Sebille, and Angus Gibson. "Speeding up python-based Lagrangian fluid-flow particle simulations via dynamic collection data structures." arXiv preprint arXiv:2105.00057 (2021).

[73]. Wellmann, J. Florian, Adrian Croucher, and Klaus Regenauer-Lieb. "Python scripting libraries for subsurface fluid and heat flow simulations with TOUGH2 and SHEMAT." Computers & Geosciences 43 (2012): 197-206.