

Eaglescope: an interactive visualization and cohort selection tool for biomedical data exploration.

Ryan Birmingham^{1*}, Nan Li^{1*}, Tony Pan¹, and Yahia Zakaria²

¹ Emory University, United States of America ² Independent Researcher, Egypt

* These authors contributed equally.

DOI: [10.21105/joss.06837](https://doi.org/10.21105/joss.06837)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Charlotte Soneson](#)

Reviewers:

- [@flekschas](#)
- [@sebastian-raubach](#)

Submitted: 29 April 2024

Published: 05 November 2024

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Eaglescope is a configurable code-free interactive visualization and cohort definition/selection tool designed for biomedical data exploration. It is designed to be hosted flexibly without the need for a dedicated server, and creates an interactive dashboard based upon a configuration file and either an API or data file. It uses visualizations of sets of features to describe and enable contextual filtering of the data. This allows for users to understand deeper patterns or anomalies within the data, and to create datasets specifically tuned to their requirements effortlessly.

Eaglescope is typically utilized either as a tool to create refined datasets tailored for training and validating machine learning AI models, or as a central hub for further exploration, allowing users to seamlessly navigate to biomedical viewers such as DICOM or whole slide imaging (WSI) platforms.

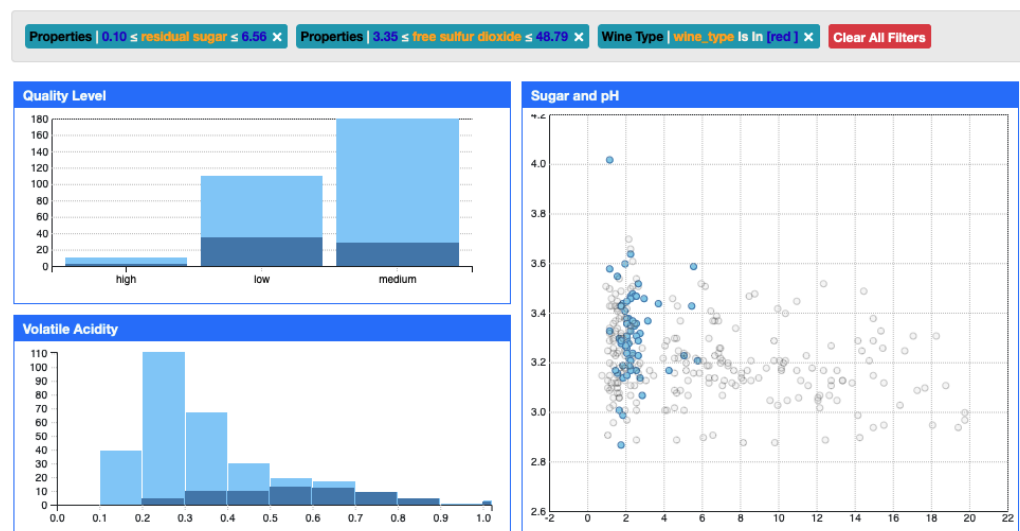


Figure 1: Interactive Contextual Visualizations

To create a dashboard, users simply need to create a file specifying the data source, configurations for each visualization, and any further desired customizations to the platform. Hosting is as straightforward as copying the static files, along with the configuration and data files if applicable, to any location capable of hosting static files. This streamlined process was intentionally designed to support the visualization of multiple datasets without added

complexity or specialized requirements. Additionally, the flexibility of hosting allows for seamless scalability with demand, eliminating the need for modifications to Eaglescope itself.

State of the Field

Cohort definition is an essential aspect of both retrospective and prospective studies. Different workflows are used to define cohorts depending on the nature and needs of a study. For example, in cases where the criteria are known, retrospective data for the criteria can be downloaded using user-focused natural language tools such as Criteria2Query (Yuan et al., 2019).

In addition to statistical and code methods, visual tools can be and are used in cohort studies. A tool called Composer (Rogers et al., 2019) allows for surgeons and patients to visually interpret results of cohort studies. Some tools, such as CAVA (Cohort Analysis via Visual Analytics) (Zhang et al., 2015) can be used for both batch analytics and interactive analytics for exploratory cohort analysis for data stored in its own relational database.

Cohort definition or dataset selection can be done at the dataset level. For an example, the Gene Expression Omnibus Database (Clough & Barrett, 2016) enables interactive exploratory analysis on multiple of its component datasets. Many other dataset providers such as the Cancer Imaging Archive (TCIA) (Clark et al., 2013), cBioPortal (Gao et al., 2013), and NBIA (Nguyen et al., 2020) host similar kinds of filtering and exploratory analysis. Ultimately, users may also download data and use tools such as Bokeh (Bokeh Contributors, 2024) to perform analytics by creating and defining their own interactive visualizations.

Statement of Need

Eaglescope was developed as a visual human-centered cohort definition and analysis tool. For cohort studies, this fits in a workflow niche for initial patient/record and feature selection. To support this dynamically and with minimal computational requirements, Eaglescope supports multiple datasets, creates dashboards which are reconfigurable without coding, and can be deployed in a serverless environment. It offers hierarchical functionality, allowing dashboards to represent and interlink with other dashboards.

In many datasets, filtering on a variable can significantly alter other properties of the data. To address this, Eaglescope employs visually contextualized filtering, which displays filtered data within its broader context. This approach helps users identify patterns and insights that might otherwise be overlooked, thereby enhancing decision-making in biomedical research. By incorporating this feature, Eaglescope supports a broader effort to mitigate systematic bias in cohort definition (Nohr & Liew, 2018). Additionally, Eaglescope can serve as a de facto home page for datasets, clarifying its features and limitations through its dashboards.

Eaglescope is a direct successor to the Datascope project (Iyer et al., 2017). It is used by the Cancer Imaging Archive (TCIA) (Clark et al., 2013) and the National Cancer Institute for exploring and exporting extensive data across collections and modalities. Additionally, the PRISM (Sharma et al., 2020) project incorporates Eaglescope for dataset creation and visualization.

Acknowledgements

We acknowledge all contributors to the Eaglescope project, as well as grant support sub-awarded by the University of Arkansas Medical School and both financial and logistical support from the Emory University Department of Biomedical Informatics.

References

- Bokeh Contributors. (2024). *Bokeh: An interactive visualization library for modern web browsers*. <https://github.com/bokeh/bokeh>
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Moore, S., Phillips, S., Maffit, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- Clough, E., & Barrett, T. (2016). The gene expression omnibus database. In E. Mathé & S. Davis (Eds.), *Statistical genomics: Methods and protocols* (pp. 93–110). Springer New York. https://doi.org/10.1007/978-1-4939-3578-9_5
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, 6(269). <https://doi.org/10.1126/scisignal.2004088>
- Iyer, G., DuttaDuwarah, S., & Sharma, A. (2017). DataScope: Interactive visual exploratory dashboards for large multidimensional data. *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, 17–23. <https://doi.org/10.1109/VAHC.2017.8387496>
- Nguyen, T., Shafi, A., Nguyen, T.-M., Schissler, A. G., & Draghici, S. (2020). NBIA: A network-based integrative analysis framework - applied to pathway analysis. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-60981-9>
- Nohr, E. A., & Liew, Z. (2018). How to investigate and adjust for selection bias in cohort studies. *Acta Obstetricia Et Gynecologica Scandinavica*, 97(4), 407–416. <https://doi.org/10.1111/aogs.13319>
- Rogers, J., Spina, N., Neese, A., Hess, R., Brodke, D., & Lex, A. (2019). Composer: Visual cohort analysis of patient outcomes. *Applied Clinical Informatics*, 10(02), 278–285. <https://doi.org/10.1055/s-0039-1687862>
- Sharma, A., Tarbox, L., Kurc, T., Bona, J., Smith, K., Kathiravelu, P., Bremer, E., Saltz, J. H., & Prior, F. (2020). PRISM: A platform for imaging in precision medicine. *JCO Clinical Cancer Informatics*, 4. <https://doi.org/10.1200/cci.20.00001>
- Yuan, C., Ryan, P. B., Ta, C., Guo, Y., Li, Z., Hardin, J., Makadia, R., Jin, P., Shang, N., Kang, T., & Weng, C. (2019). Criteria2Query: A natural language interface to clinical databases for cohort definition. *Journal of the American Medical Informatics Association*, 26(4), 294–305. <https://doi.org/10.1093/jamia/ocy178>
- Zhang, Z., Gotz, D., & Perer, A. (2015). Iterative cohort analysis and exploration. *Information Visualization*, 14(4), 289–307. <https://doi.org/10.1177/1473871614526077>