

Alex Mutebe
MSc. Data Science
A Reflective Piece
23-July-2023

This is my reflection on data management and statistical competence in clinical research as a student at the University of Essex pursuing a master's degree in data science

I'm Alex Mutebe, and I'm presently studying for a master's degree in data science at Essex University. I work with the Medical Research Council Uganda Unit at this point in my academic career, where I am involved in clinical research and data management.

I commit much effort into making sure that the data management procedure conforms with accepted clinical standards in my role as a clinical data manager with the Medical Research Council Uganda section. I've been producing data reports for the principal investigator and project leads, who are in charge of overseeing and reporting the data collection progress during the research study period (Ellenberg et al., 2002). Additionally, I manage data analysis for a range of datasets, from large to small basic to difficult. I believe that proper data interpretation and insightful conclusions depend on having a strong foundation in statistical ideas.

It takes a lot of practice to accurately clean and report data results. I've occasionally had trouble managing large databases with incomplete data. These issues occasionally lead to my analysis being somewhat off without appropriate exploratory data analysis techniques. The dimensionality reduction lectures gave me a variety of imputation techniques and data reduction tactics to assist me get past them as an Essex University student earning a master's degree in data science (James et al., 2021).

I intend to continue exploring clustering and classification models in multivariate analysis and machine learning in order to further develop my machine learning abilities. I hope to undertake this in order to improve my ability to manage complicated datasets and develop a better understanding of clinical research data (Self, 2018).

Among the various ways, visualization is one of the most crucial exploratory data analysis strategies. Before I took a course on visualizing data, I used to have trouble choosing the most suitable models to represent data visually (Gour, 2019). For research findings to be successfully communicated, data visualization was essential. It enabled me to make difficult information accessible to both technical and non-technical audiences by presenting it in a visually beautiful and simple-to-understand way (Bradford, 2011).

In employing data visualization tools to make various charts, graphs, and interactive dashboards, I've grown quite confident. I was able to find significant connections and support what was learned by visualizing trends and patterns in the data (Nolan & Perrett, 2015). Returning to my place of employment, I recently led an introduction to R for data science workshop for my coworkers. I focused mainly on ggplot2 data visualization knowledge sharing (Khaled, 2020).

I intend to investigate cutting-edge visualization libraries and technologies to further develop my data visualization abilities. Additionally, in order to improve my visualizations and effectively communicate the research findings, I will ask my colleagues and mentors for their input.

During this class, I had the opportunity to study and use data visualization technologies like Tableau. Similar to other tools, tableau is a data analytics program that has shown to be useful in expediting the data visualization process, allowing me to produce dynamic and aesthetically pleasing representations of the data (Tableau, 2022).

I can now efficiently combine data from many sources and create perceptive representations using tableau. My exploratory data analysis is strengthened by the interactive features' ability to

drill down into particular data points and analyze the data from many perspectives (Komorowski, 2016).

I want to expand my knowledge of data visualization in the future by investigating additional platforms and tools. The ability to adapt to changing project requirements and work more productively with interdisciplinary teams will come from learning new tools.

I had the chance to work extensively with programming languages like R and Python for data analysis and visualization during this course. These languages offer a vast array of statistical libraries and visualization capabilities, and as a result, my analytical abilities have significantly improved (Golemund & Wickham, 2017)

R, in my opinion, excels in statistical functionality, whereas Python, because of its adaptability and interoperability with other libraries, is a strong tool for data manipulation and display.

I'll keep working on personal projects that require data analysis and visualization using both R and Python to further develop my skills. This practical experience will aid in my development as a proficient user of these programming languages for a range of research scenarios (Huddar & Kulkarni, 2018).

Conclusion

As a Masters in Data Science student, my experience with clinical research and data management has been tremendously enlightening. I've learned a lot about statistical analysis and data visualization techniques, which has helped me communicate my research findings clearly and come to insightful conclusions. I am enthusiastic to advance the application of data science approaches in the future while continuing to build on this foundation, contributing to the field of clinical research.

References

- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021) An Introduction to Statistical Learning with Applications in R. 2nd ed. New York: Springer.
- Self, G. (2018) The 5 Clustering Algorithms Data Scientists Need to Know. Towards Data Science. Available from <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68> [Accessed on 02 July 2023]
- Gour, R. (2019) Top 5 BI Tools Widely used for Data Visualization. Towards Data Science. Available from <https://towardsdatascience.com/top-5-bi-tools-that-you-must-use-for-data-visualization-7ccc2a852bd3> [Accessed on 20 June 2023]
- Bradford, W.C. (2004) Reaching the Visual Learner: Teaching Property Through Art. Researchgate. Available from https://www.researchgate.net/publication/228146305_Reaching_the_Visual_Learner_Teaching_Property_Through_Art [Accessed on 03 July 2023]
- Nolan, D., Perrett, J. (2015). Teaching and Learning Data Visualization: Ideas and Assignments. *The American Statistician*. 70. 10.1080/00031305.2015.1123651.
- Tableau Software (2023). Tableau Desktop. Available from <https://www.tableau.com/why-tableau/what-is-tableau>
- Komorowski, M., Marshall, D., Salciccioli, J., Crutain, Y. (2016) Exploratory Data Analysis. SPRINGER LINK. Available from https://link.springer.com/chapter/10.1007/978-3-319-43742-2_15 [Accessed on 15 July 2020]
- Huddar, R., Kulkarni, R. (2018). Role of R and Python in Data Science. RESEARCH JOURNEY International Multidisciplinary E- Research Journal. Available from www.researchjourney.net
[Accessed on 28 July 2023]
- haled, A. (2020). Data Visualization with ggplot2. Towards Data Science. Available from <https://towardsdatascience.com/data-visualization-with-ggplot2-db04c4956236> [Accessed on 20 June 2023]