

FOSTERING DATA SCIENCE AND STATISTICS EDUCATION IN AFRICA VIA ONLINE TEAM-BASED LEARNING

O. Olawale Awe, Kim Love and Eric A. Vance

LISA 2020 Global Network USA/ Global Humanistic University, Curacao

K.R. Love Quantitative Consulting and Collaboration, Athens, Georgia, USA

Laboratory for Interdisciplinary Statistical Analysis (LISA), University of Colorado Boulder, USA

olawaleawe@gmail.com

Many African students are not usually exposed to the analytical experience with data and computing skills they need to be successful in the workplace after graduation. Also, students often have limited exposure to team-based data science and the principles and tools that are encountered outside of school. In this paper, we describe the ADA Global Academy-Laboratory for Interdisciplinary Statistical Analysis (AGA-LISA) program, a LISA 2020 Global Network data science development project in which teams of graduate students are mentored online by a local non-profit organization on various collaborative data-focused projects. To help the students develop and improve confidence in their technical and non-technical data science skills, the project promoted a team-based approach to data science. Evidence from the project evaluation survey is presented to document the degree to which the project was successful in engaging students in team-based data science, and how the project impacted their technical and non-technical skills.

INTRODUCTION

In recent times, the need for evidence-based statistical reasoning and capacity building in developing countries has been growing steadily (Love et al, 2022). ADA Global Academy (AGA) is an online capacity-building organization of learning for holistic data science education and continental development in the application of computational data science methods across disciplines. Its mission is to impart the necessary knowledge and practical skills that would enable data scientists to become self-reliant industrial practitioners, continental data science leaders, and researchers with global engagement and local relevance. It is an affiliate member of the LISA 2020 Global Network (www.lisa2020.org).

The objectives of AGA are to (1) provide basic online data science training and education to professionals aspiring to acquire or upgrade knowledge in statistics and data science; (2) to train individuals and corporate organizations on how to properly design experiments or surveys, collect data and perform data analysis using statistical software packages such as R and Python; and (3) to train enrollees on data science skills that can be used for analyzing data in business, environment, health, science, government, and technology fields. It has qualified professionals who can guide young professionals on how to manage, understand, and analyze complex data so that they will be able to communicate results with comprehensible data visualizations to influence key business decision-making in society. Approved by the Government of the Federal Republic of Nigeria under the auspices of the Corporate Affairs Commission (CAC, Reg. No. 3390293), AGA is a not-for-profit organization that focuses on capacity building, training and research development, facilitation of workshops and seminars, and project management. It is a group of professionals who are committed to ensuring that data science is popularly known and freely practiced among students (undergraduates and graduates) of higher institutions, researchers, professionals in corporate organizations, scholars, and academic (faculty) staff members in developing countries and beyond. We are linking African data scientists with business organizations and companies abroad to provide exposure to practical data and practice their data science skills.

According to Vance and Pruitt (2022), a stat lab is not a big room filled with computers but a team of statistical collaborators serving as research infrastructure to solve real-world problems through statistics and data science. This is the main motivation behind this exercise. Team-based learning has been used to successfully mentor data science students in the past (see, for instance, Vance, 2021; Zgheib et al., 2010; Clair and Chihara, 2012). However, we used online team-based learning in this experiment because online classes have been found to have a comparable impact with physical classes (Lino-Neto et al., 2022). It allows students from geographically diverse locations to participate in classes and offers the benefit of allowing learning to continue during a pandemic. According to Vance (2021), the field of data science is a collaborative field, and its students should learn teamwork and collaboration. Team-

Based Learning (TBL) is a pedagogical strategy that can help educators teach data science better by flipping the classroom to employ small-group collaborative learning to actively engage students in doing data science (Charalambous et al., 2021).

MATERIALS AND METHODS

In this experiment, we offered a free (online) six-week mentorship data science short course to thirty (30) aspiring data scientists and researchers of African origin. Selected candidates worked under the guidance and mentorship of the LISA 2020 Ambassador to Africa and other mentors (including the co-authors of this paper) selected from North and South America to build predictive models for applications in many fields to solve societal problems. The course was designed for anyone whose work interfaces with data analysis, and who wants to learn the key concepts, formulations, algorithms, and practical examples of what is possible in machine learning and data science. The main aim of this course was to bridge the skill gap in machine learning and data science among African scientists and ensure the sustainability of these skills. Trainees were selected with a good gender mix (about a 60:40 ratio). Participants underwent three weeks of intensive online classes and spent three weeks working on a supervised project of their choice, which would later be developed into manuscripts. To sustain their acquired skills, we will keep in touch with the participants for continuous collaboration to help master their skills and solve real-world problems. The aim is for them to also publish manuscripts through the skills they have learned. Most of them will also become trainers for our future programs to sustain our organization and other labs in the LISA Global Network. In our experiment, we utilized the essential elements and established guidelines of team-based learning (TBL) to support students collaborating within permanent teams on well-designed application exercises to undertake several data science projects in the spirit of Vance (2021) and Burgess et al. (2014). We also utilized tutorial ideas in literature for teaching the students (Awe et al., 2022). Some of the basic data science topics taught during this online experiment include:

- Descriptive statistics, data exploration, and wrangling with R and Python.
- Statistical modeling and machine learning techniques.
- Inbuilt Packages-contributed packages.
- Data visualization: base plots, ggplot, gganimate, ggmap, etc.
- Using Tidyverse for Data Science.
- Developing simple and effective programming skills.
- Basic research and collaboration skills.

After some weeks of practical technical sessions, the students were divided into groups to enable them to try new codes and apply them to new group projects of their choice. The participants were grouped according to their areas of interest and expertise. Each group had at least three members.

RESULTS

The results of our experiment using TBL to teach a modern, introductory data science course indicate that the course effectively taught reproducible data science workflows, beginning R and Python programming, communication, and collaboration skills. Students also reported how much they have improved in their learning of statistical thinking and basic R concepts. We have been able to improve the technical know-how and statistical literacy of researchers and students in Africa through various online courses, consultations, and collaborations. The overall result of the experiment was adjudged successful, as indicated by the feedback of the participants. There will be a continuous feedback process as well as continuous monitoring and evaluation of the effect of this activity or experiment. A longitudinal survey would be administered to the participants at regular intervals for up to two years after this first experiment. We will also measure success via the number of manuscripts developed into publications as a result of this experiment at the end of the year. Some manuscripts developed by the students are currently under review in high-impact journals. Published papers from this experiment would be archived as part of the LISA 2020 publications. Most of the trainees will become trainers for our future programs to sustain our organization and other labs in the LISA 2020 Global Network. Participants were asked to fill out a questionnaire, and the graphical representations of their responses

were recorded. In Fig. 1, participants were asked if they wanted to participate in more programs like this in the future, and 97.4% agreed, while 5.3% were unsure. When asked how satisfied they were with the program, the participants responded as shown in Fig. 2, with 52.6% indicating high satisfaction and 47.4% indicating satisfaction. In Fig. 3, participants were asked to rate the program's usefulness to their career or future studies on a scale of 1 to 5. 84.2% rated it a 5, and 15.8% rated it a 4. When asked if their expectations were met, 89.5% of participants in Fig. 4 responded positively, compared to 10.5% who did not. These outcomes demonstrate how beneficial the program was, and with more funding, it will advance the careers of young data scientists in Africa.

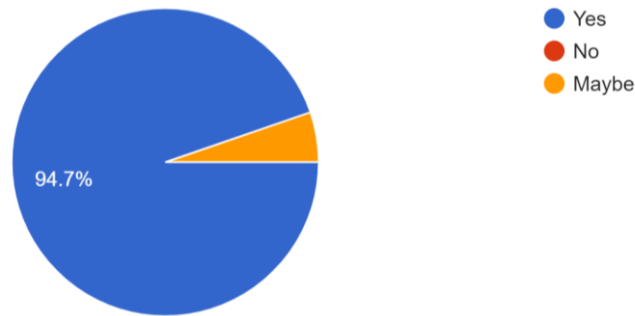


Figure 1. Would you love to attend more courses/internships like this in the future?

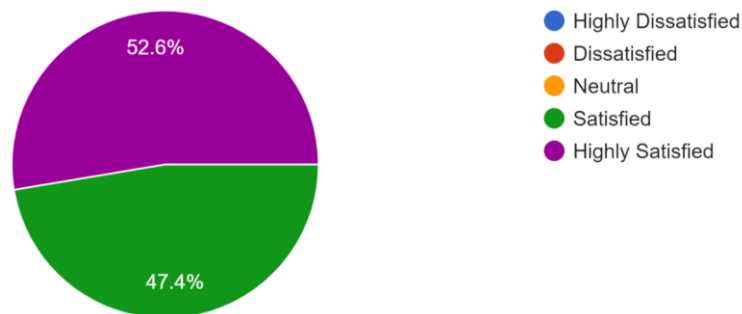


Figure 2. How satisfied are you with this course/Internship?

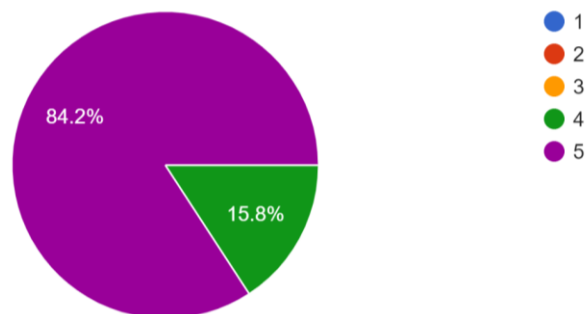


Figure 3. On a scale of 1-5 (5 being the highest), how useful will what you have learned in this course/Internship be to your career or future studies?

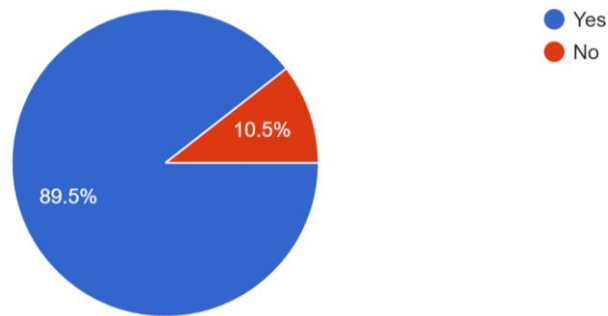


Figure 4. Were your expectations met?

FEEDBACK

After the course, a feedback survey was administered to the participants. Here are some comments quoted from the course feedback survey concerning the success of our experiment:

- “I like ADA Global Academy because it provides training on different statistical packages which is key to statisticians and help in building our career.” “It is really informative learning data science at ADA Global Academy.”, “The quality of the expertise of the tutors is second to none!”
- “The ADA Global Academy is an excellent learning platform for young scholars to diversify their quantitative skills in modern statistical techniques, while gaining an added opportunity to produce a publishable research paper”. “ADA GLOBAL Academy is one of the best places to be, either as a Data Scientist or prospective Data Scientist.” ‘ADA GLOBAL Academy is a place I recommend for anyone interested in building career or skills in Data Science and Machine Learning!’
- “The trainers were down to earth, they have a full knowledge of the courses thought and the hands on sessions were really impactful.” “The mentoring sessions were even more than what I experienced during my postgraduate days. My mentor was always there and was ready to pour out the knowledge until I understood and perfected my work.” “Being able to publish a paper at the end of the course was another milestone in my career. I really appreciate ADA Global Academy for this selfless service to humanity.”
- “Learning from great minds is a privilege I will never take for granted.” ‘ADA Global Academy is investing in lives. I love the organization, and there is precision and promptness in answering questions generated as a result of each lecture throughout the program.’
- ‘My experience as an intern at ADA Global Academy was exceptional. The curriculum was integrated with values, impact and motivation for growth with emphasis on the importance of teamwork, communication skills and leadership building. I recommend ADA global academy for anyone anticipating excellence.’
- ‘ADA academy opened up an opportunity to collaborate with other researchers across different fields. It was an avenue to develop skills in data analysis and machine learning. ADA is definitely a part of my success story.’
- ‘Professor Olawale Awe is my role model; his impactful training is top-notch. I sincerely want to appreciate ADA Global Academy for this golden opportunity to learn and apply what I learnt, during the cause of our project.’

CONCLUSION

Although we hope to improve on the methodology mentioned earlier in this study in the future, it was a successful and satisfying experiment. To help the data science education community grow, we suggest that other organizations adopt this appealing online pedagogical strategy that we have adopted for teaching data science. A consequence of this teaching method would be that it would help students achieve the workforce-relevant data science learning skills of effective communication, teamwork, and collaboration (Vance, 2021). Using TBL to teach data science is a relatively new method. There is still a lot of room for improvement in building data science capacity in Africa (Awe et al, 2015). We

encourage various organizations to embark on similar projects in order to boost statistics and data science capacity in Africa.

REFERENCES

- Awe, O. O., Crandell, I., & Vance, E. A. (2015). Building statistics capacity in Nigeria through the LISA 2020 program. In *Proceedings of the International Statistical Institute's 60th World Statistics Congress*. Rio de Janeiro.
- Awe, O. O., Jegede, P. O., & Cochran, J. A. (2022). Comprehensive Tutorial on Factor Analysis with R: Empirical Insights from an Educational Perspective. *Promoting Statistical Practice and Collaboration in Developing Countries*, 265. Chapman and Hall/CRC.
- Burgess, A. W., McGregor, D. M., and Mellis, C. M. (2014), "Applying Established Guidelines to Team-Based Learning Programs in Medical Schools: A Systematic Review," *Academic Medicine*, 89, 678–688.
- Charalambous, M., Hodge, J. A., & Ippolito, K. (2021). Statistically significant learning experiences: Towards building self-efficacy of undergraduate statistics learners through team-based learning. *Educational Action Research*, 29(2), 226-244.
- Clair, K. S., & Chihara, L. (2012). Team-based learning in a statistical literacy class. *Journal of Statistics Education*, 20(1).
- Lino-Neto, T., Ribeiro, E., Rocha, M., & Costa, M. J. (2022). Going virtual and going wide: comparing Team-Based Learning in-class versus online and across disciplines. *Education and Information Technologies*, 27(2), 2311-2329.
- Love, K., Awe, O. O., Gunderman, D. J., Druckenmiller, M., & Vance, E. A. (2022). LISA 2020 Network Survey on Challenges and Opportunities for Statistical Practice and Collaboration in Developing Countries. *Promoting Statistical Practice and Collaboration in Developing Countries*, 47-59. Chapman and Hall/CRC.
- Vance, E. A. (2021). Using Team-Based Learning to Teach Data Science. *Journal of Statistics and Data Science Education*, 29(3), 277-296.
- Vance, E. A., & Pruitt, T. R. (2022). Statistics and data science collaboration laboratories: Engines for development. In *Promoting Statistical Practice and Collaboration in Developing Countries* (pp. 3-26). Chapman and Hall/CRC.
- Zgheib, N. K., Simaan, J. A., & Sabra, R. (2010). Using team-based learning to teach pharmacology to second-year medical students improves student performance. *Medical teacher*, 32(2), 130-135.