INTELLIGENCE BY THE NUMBERS: DATA ANALYTICS TRAINING FOR UNDERGRADUATE INTELLIGENCE STUDENTS

COBB, MATTHEW, Coastal Carolina University
mcobb1@coastal.edu

ROBERTS, JORDAN, Coastal Carolina University
jrobert2@coastal.edu

ABSTRACT

This article explores an undergraduate course, Data Analytics for Intelligence and Security, at Coastal Carolina University. The course, a requirement for the new Data Analytics for Intelligence and Security minor, focuses on theory and practice. It equips students with practical skills highly applicable to analytical careers in national security. To enhance their practical skills, students utilize the *R programming language*, a free and open-source software widely used for statistical computations. They gain extensive hands-on experience managing and visualizing data and analyzing relationships between variables using regression analysis. This practical experience develops students' self-efficacy.

Keywords: *data analytics, quantitative social science, R programming language*

The technological capacity to collect and store near-infinite amounts of information is rife with challenges and opportunities. There is a growing need for intelligence and security professionals to possess basic data literacy skills (Coulthart et al., 2024; Lim, 2016; Ramsay & Macpherson, 2024; Van Puyvelde et al., 2017). Borrowing from the social sciences, intelligence professionals have long used estimative language to describe the findings of their analysis, a process formalized in government policy such as the Office of the Director of National Intelligence's (2015) *Intelligence Community Directive 203, Analytic Standards*. However, in some situations, descriptors such as *likelihood* are insufficient to communicate information meaningfully to intelligence consumers. To avoid unintentionally obscuring important details, some intelligence scholars recommend presenting information with numerical probabilities (Barnes, 2016; Dhami, 2018).

While intelligence scholars have been slower to adopt quantitative methods than scholars from comparable fields (Macpherson, 2020), there is an increasing amount of quantitative research in intelligence on topics related to intelligence capacity (Stottlemyre, 2024), covert action (Roberts, 2023), and counterintelligence (Levin, 2023). In response to the need for greater data literacy and quantitative research skills, Coastal Carolina University's Intelligence and Security Studies department created a data analytics course, Data Analytics for Intelligence and Security (INTEL

309). The course is the cornerstone of the University's new Data Analytics for Intelligence and Security minor. This article analyzes the course, focusing on its role in educating future intelligence professionals.

COASTAL CAROLINA UNIVERSITY

Coastal Carolina University is a public university in Conway, South Carolina, accredited by the Southern Association of Colleges and Schools Commission on Colleges. As of 2024, the University has approximately 10,800 students, 10,300 of whom are undergraduates. It is home to the Department of Intelligence and Security Studies, which hosts approximately 350 undergraduate intelligence majors.

The Data Analytics for Intelligence and Security minor includes three required classes from the Department of Intelligence and Security Studies and two elective courses in statistics, computer science, and geographic information systems from other departments. The first required course is Intelligence Research and Communication (INTEL 301), which introduces students to the scientific method and prepares them to conduct their capstone research projects. The second is Social Networks (INTEL 316), a course that uses intelligence analysis and social science research to teach social network analysis. This article focuses on the final required course, Data Analytics for Intelligence and Security, which focuses on statistical analysis, data wrangling and management, and data visualization.

DATA ANALYTICS FOR INTELLIGENCE AND SECURITY

The Data Analytics for Intelligence and Security course helps students build *hard skills*, the technical capabilities employers increasingly value and often find difficult to teach through on-the-job training. Student learning outcomes for this course include articulating the application of data science to intelligence and security-related problems, mastering basic data management skills, using software to develop and run statistical models, creating graphs and charts to communicate information to consumers, and demonstrating basic proficiency in computer programming needed for data science tasks. The skills covered in class enrich students' abilities to complete intelligence analytic tasks utilizing data science and research methodologies derived from the social sciences. This aligns with what Stoian (2023) characterizes as *research-in-support* of intelligence missions, blending intelligence and the social sciences. The class does not replicate the experience of supporting any single employer or producing any specific intelligence product; instead, the class develops students' technical skills.

Carefully building students' technical skills over a semester is essential, considering how many students enter the course feeling intimidated by programming and statistics. The instructors have found that a slow and steady approach emphasizing small, concrete accomplishments is best for building students' *self-efficacy*. Self-efficacy is the conviction that one can successfully execute the behavior required to produce the outcomes (Bandura, 1977). It has long been theorized that individuals with higher self-efficacy are capable of higher performance on specified tasks

(Bandura, 1982) and that self-efficacy is a primary predictor of learning (Chiaburu & Lindsay, 2008). Training in specific skills to raise self-efficacy can apply to technical skills, such as in the data analytics course, or to other skills, such as interpersonal interaction, public speaking, or critical reading and writing. This approach to developing skills and self-efficacy is highly relevant to individuals interested in security or public safety careers. For example, law enforcement officers who receive crisis intervention training demonstrate enhanced self-efficacy for interacting with individuals struggling with mental health or addiction issues (Bahora et al., 2008).

Data Analytics for Intelligence and Security has been taught three times by two faculty members in person in a computer lab. Initially, a pilot version of the class was offered in the spring 2022 semester before the official launch of the Data Analytics for Intelligence and Security minor. Students must complete an introductory statistics course as a prerequisite or receive permission from the instructor to register. Students in this course are expected to learn the *R programming language*, a computer programming language closely related to Python optimized for statistical analysis, including a wide variety of tools for processing large volumes of quantitative data, running statistical models, and generating high-quality graphs. Using R, the students complete various data analytic activities, such as formatting data appropriately for real-world tasks, generating descriptive statistics for key variables, and using regression models to evaluate the relationships between variables.

This hands-on class requires students to engage in data management and analytics, including merging datasets, reformatting data, and overcoming difficulties caused by incomplete observations. The course is facilitated with a textbook that covers basic statistical concepts and focuses on the practical application of statistical tools and techniques rather than theoretical concepts. Additionally, the course emphasizes data visualization, and examples of figures produced by students can be seen in Figure 1. The first graph visualizes the outcomes of international crises according to levels of involvement by the U.S. government based on data from the International Crisis Behavior dataset (Brecher & Wilkenfeld, 1997; Brecher et al., 2023). The second graph depicts the number of fatal terrorist attacks occurring in four African countries from 2000 to 2020, based on data from the Global Terrorism Database (National Consortium for the Study of Terrorism and Responses to Terrorism [START], 2022).

Different course instructors assign students to work with various datasets, including the Global Terrorism Database (START, 2022), the International Crisis Behavior dataset (Brecher & Wilkenfeld, 1997; Brecher et al., 2023), and the World Bank Global Development Indicators (World Bank, 2023). Students construct datasets with variables they code manually, predict the likelihood of international crises breaking out between countries, and generate graphs showing changing levels of poverty and terrorism in various countries over time. Students wishing to work on more advanced topics have also continued their learning with independent studies that build on the foundation provided by the Data Analytics for Intelligence and Security course.

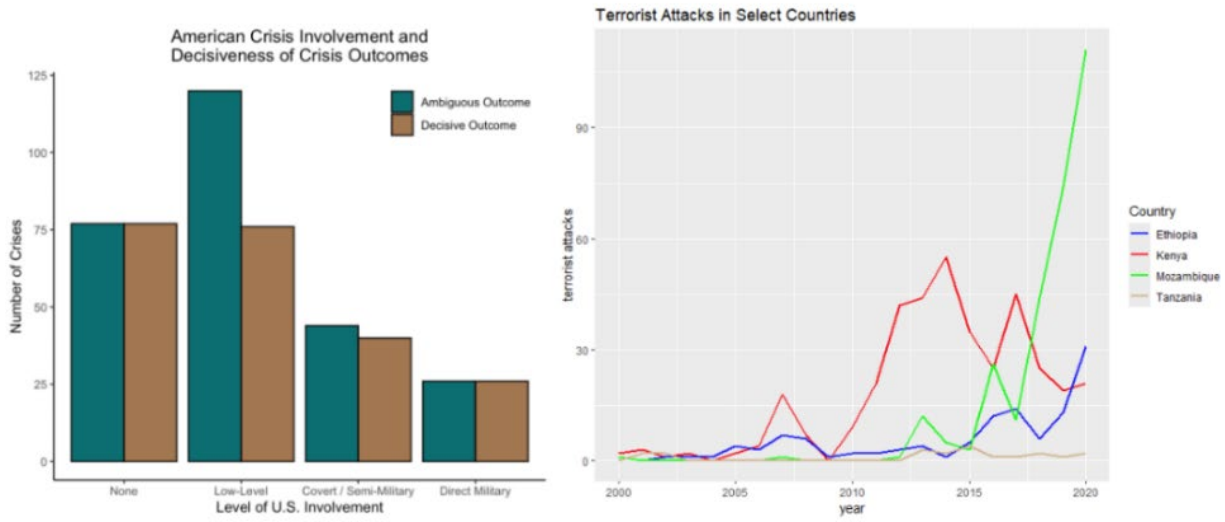

Figure 1. Examples of Graphs Produced by Students

Some sections of Data Analytics for Intelligence and Security offered have included several weeks of instruction in Microsoft Excel, including constructing datasets, generating graphs, and summarizing data with pivot tables. Students learn to load and merge datasets, summarize variables, create and recode variables, and generate graphs in the R programming language. Students do these activities with R's base functions and a family of R software packages known as the *tidyverse*. The specific tests conducted with the R programming language vary from section to section. However, they have included analysis of variance, linear regression, logistic regression, ordered logistic regression, and negative binomial regression.

COURSE ASSESSMENT

Most course assignments are practical exercises through which students demonstrate proficiency in completing tasks practiced in class. Students also complete a research design paper in which they plan a scientific study of a security-relevant topic using quantitative methods. This includes reviewing the topic, proposing a theory about the relationship between two or more variables, identifying relevant datasets to obtain data for the study's independent, dependent, and control variables, and specifying a statistical model appropriate for the analysis.

In the early offerings of Data Analytics for Intelligence and Security, the instructors interviewed students about their learning experiences in the course. In this way, faculty assessed student performance and the course's effectiveness in teaching them essential data science skills. The faculty has used information from these interviews to adjust course content and assessments to maximize educational impact. The faculty hopes that graduate feedback will help gauge the effectiveness of the course. The feedback may assist the faculty in improving learning outcomes for current students and keeping the repertoire of skills taught in class relevant to the field.

THE FUTURE OF DATA ANALYTICS FOR INTELLIGENCE AND SECURITY

Based on student feedback and discussions between the faculty members teaching Data Analytics for Intelligence and Security, there are several changes that the instructors will implement in future iterations of the course. One idea is to add a presentation in which students formally present their work, complete with data visualizations and statistical results, in such a way as to accurately communicate data-driven findings to an audience that is unfamiliar with statistical methods.

In teaching this course, the authors have found that students' prior knowledge of programming and engagement with the material varies immensely. The range of students' pre-course programming knowledge presents a pedagogical dilemma. Instructors must meet the needs of the strongest and weakest students. In the most recent iteration of the course, the department hired a student as a teaching assistant (TA) to work with students struggling with course content, particularly coding-related topics. The department plans on employing TAs to assist students in future iterations of this course.

Finally, the emergence of readily accessible artificial intelligence (AI), specifically large language models such as ChatGPT, offers a major opportunity and a significant challenge for future iterations of the data analytics course. It is increasingly clear that artificial intelligence will assist in *R programming* and other data analysis tasks in the future, and integrating training on how best to utilize this new technology will be a necessary addition for the course to remain cutting-edge. However, AI also poses a significant challenge. AI models can program tasks, allowing students to become so dependent on the technology that they fail to learn the fundamentals necessary to engage in advanced work. Additionally, communicating the bright line between using AI to complete menial programming tasks and "cheating" by having a large language model "do the homework" for the student may prove difficult.

CONCLUSION

This article examined a new data analytics course, Data Analytics for Intelligence and Security, to help students build quantitative analytical skills not covered in more traditional intelligence studies courses. The course provides students with practical experience with data analysis tasks and helps them develop a foundational proficiency in statistical programming. The course content gives students hands-on experience managing and visualizing data on relevant security-related topics. Students also learn and use a variety of statistical models and complete a research paper that plans a scientific study using quantitative variables and statistical methods. The skills covered in this course build on research methods from the social sciences and prepare students for intelligence careers. The course iterations offered at Coastal Carolina University have been successful, as measured based on interviews with students completing the course and instructors' observations of students' performance. The instructors will continue to assess and update the course to optimize their pedagogy and maximize student learning.

REFERENCES

Bahora, M., Hanafi, S., Chien, V. H., & Compton, M. T. (2008). Preliminary evidence of effects of crisis intervention team training on self-efficacy and social distance. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*, 159–167. https://doi.org/10.1007/s10488-007-0153-8

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Barnes, A. (2016). Making intelligence analysis more intelligent: Using numeric probabilities. *Intelligence and National Security*. *31*(3), 327–344. https://doi.org/10.1080/02684527.2014.994955

Brecher, M., & Wilkenfeld, J. (1997). *A study of crisis*. University of Michigan Press. https://doi.org/10.3998/mpub.14982

Brecher, M., Wilkenfeld, J., Beardsley, K., James, P., & Quinn, D. (2023). *International crisis behavior data codebook* (Version 15) [Data set]. http://sites.duke.edu/icbdata/data-collections/

Chiaburu, D. S., & Lindsay, D. R. (2008). Can do or will do? The importance of self-efficacy and instrumentality for training transfer. *Human Resource Development International*, *11*(2), 199–206. https://doi.org/10.1080/13678860801933004

Coulthart, S., Shahriar Hossain, M., Sumrall, J., Kampe, C., & Vogel, K. M. (2024). Data-science literacy for future security and intelligence professionals. *Journal of Policing, Intelligence and Counter Terrorism*. *19*(1), 40–60. https://doi.org/10.1080/18335330.2023.2187705

Dhami, M. K. (2018). Towards an evidence-based approach to communicating uncertainty in intelligence analysis. *Intelligence and National Security*, *33*(2), 257–272. https://doi.org/10.1080/02684527.2017.1394252

Levin, D. H. (2023). Is sunlight the best counterintelligence technique? The effectiveness of covert operation exposure in blunting the Russian intervention in the 2020 U.S. election. *Intelligence and National Security*. *38*(5), 816–834. https://doi.org/10.1080/02684527.2023.2181923

Lim, K. (2016). Big data and strategic intelligence. *Intelligence and National Security*. *31*(4), 619–635. https://doi.org/10.1080/02684527.2015.1062321

Macpherson, A. (2020). Quantitative intelligence literature: Are intelligence scholars behind the curve? *Journal of Security, Intelligence & Resilience Education*, *9*(1). https://jsire.org/wp-content/uploads/sites/661/2020/05/2019-4-macpherson-F-5.19.20.pdf

National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2022). *Global terrorism database 1970–2020* [Data set]. https://www.start.umd.edu/gtd

Office of the Director of National Intelligence. (2015). *Analytical standards*. Intelligence Community directive 2003. https://www.dni.gov/files/documents/ICD/ICD-203.pdf

Ramsay, J., & Macpherson, A. (2024). The integration of statistical learning in intelligence education: Is the academy equipping tomorrow's intelligence professionals to analyze data-centric threats? *Journal of Policing, Intelligence and Counter Terrorism*, *19*(1), 3–21. https://doi.org/10.1080/18335330.2022.2048965

Roberts, J. (2023). State preferences, viable alternatives, and American covert action, 1946–1989. *Intelligence and National Security*, *38*(4), 501–524. https://doi.org/10.1080/02684527.2022.2123937

Stoian, V. (2023). The best of both worlds? A hybrid approach to intelligence and research. *Intelligence and National Security*. *38*(4), 611–627. https://doi.org/10.1080/02684527.2022.2137152

Stottlemyre, S. (2024). Intelligence for human security: Measuring outcomes quantitatively. *Intelligence and National Security*, *39*(1), 93–118. https://doi.org/10.1080/02684527.2023.2250478

Van Puyvelde, D., Coulthart, S., & Hossain, M. S. (2017). Beyond the buzzword: Big data and national security decision-making. *International Affairs*, *93*(6), 1397–1416. https://doi.org/10.1093/ia/iix184