2021 Mini-Grant Awardee Abstracts

Atlanta University Center Data Science Initiative

The AUC Data Science Initiative's Mini-Grants program aims to stimulate data science-related research and curriculum development productivity while strengthening research and inform teaching. For data science curriculum and education-focused proposals, the project may adopt, adapt, create, or improve curricular materials, curriculum design, or practices in ways that improve the learning and learning environments of data science. For data science research-focused proposals, the project should address how the research will expand advances in data science and the potential for discovery across different disciplinary areas.

The Mini-Grants Program has four tracks:

- Track 1 Curriculum Development (New Module or New Course),
- Track 2 Research Development,
- Track 3 Summer Program Development, and
- Track 4 Academic Year Development.

Below is a summary of awardees and followed by a complete summary with corresponding abstracts. *We are happy to support our 2021 cohort!*

Summary

Track 1 – Summer Curriculum Development: New Module

The award amount is $1,000.

Data Science Module for Database Systems Courses, Alfred Watkins, Academic Program Director & Senior Assistant Professor, Computer Science, Morehouse College; alfred.watkins@morehouse.edu

How to Manage Data: Creation of an Interdisciplinary Research Data Management Module for AUC Student Researchers, Bryan Briones, Reference Librarian/Facilitator for Research Data Management, Research, Learning and Technology Services Department, Data Science and Technology Services Unit, Robert W. Woodruff AUC Library; bbriones@auctr.edu

Say Her Name: Quantifying Black Women's Experience with Police Violence, Celeste Lee, Assistant Professor, Sociology and Anthropology, Spelman College; celestenleephd@gmail.com

A Data Science Module for the Community Health Care Course, Kenya Jones, Associate Professor, Whitney M. Young Jr. School of Social Work, Clark Atlanta University; kjones1@cau.edu

Enhancing Computation in a Physics Capstone Course, Michael Burns-Kaurin, Associate Professor, Physics/Academic Affairs, Spelman College; mburns-k@spelman.edu

Track 1 – Summer Curriculum Development: New Course

The award amount is $3,000.

A Statistics for Data Science Course, Chuang Peng, Professor, Mathematics and Computational Sciences, Morehouse College; chuang.peng@morehouse.edu

Integrating Data Science and Programming into a New Discrete Structures Course, Mohammad Ali Sazegarnejad, Computer Science Lecturer, Cyber-Physical Systems, Clark Atlanta University;

asazegar@cau.edu

A New Course on Data Visualization at Morehouse, Perry Sweeper, Professor of Practice, Experiential Learning and Interdisciplinary Studies, Morehouse College; perry.sweeper@morehouse.edu

Developing a Data Science Summer Program for K-12 Students, Shonda Lawrence, Associate Professor/PhD Chair, Social Work, Clark Atlanta University, with J. Fidel Turner, Jr., Dean, School of Education, Clark Atlanta University; slawrence@cau.edu

Designing the Data and the African Diaspora Course at Morehouse College, Sonya Dennis, Assistant Professor, Computer Science, Morehouse College; sonyamd@yahoo.com

Track 2 – Summer Research Development

The award amount is $3,000.

A Multivariate Statistical Analysis of the Effectiveness of ALEKS in Improving the Educational Outcomes of STEM Majors, Daniel Teodorescu, Professor, Educational Leadership, Clark Atlanta University; dteodorescu@cau.edu

Data Visualization Platform for the War on Drugs with a Focus on Atlanta, Georgia, Fernando Esquivel-Suarez, Assistant Professor, English, Spelman College; fesquive@spelman.edu

Creating a Database to Categorize Structural Trends and Chemical Functions of Benzalkonium Salts, Issifu Harruna, Professor, Chemistry, Clark Atlanta University; iharruna@cau.edu

Predicting the Disease Threat of Microbial Pathogens Using Data Science, Jeticia Sistrunk, Assistant Professor, Biology, Spelman College; jsistru2@spelman.edu

Using Data Science and Experimentation to Determine Properties of Novel Materials, Natarajan Ravi, Professor, Physics, Spelman College; nravi@spelman.edu

Using High Performance Computing and Machine Learning Tools to Develop Fullerene-Free Acceptors for Organic Solar Cells, Seyhan Salman, Assistant Professor, Chemistry, Clark Atlanta University; ssalman@cau.edu

COVID-19 and the 2020 General Election Participation Among African Americans, Teri Platt, Associate Professor, Public Administration, Clark Atlanta University; tplatt@cau.edu

Using Survey Data to Understand the Election of Women Leaders in South Asia, Tinaz Pavri, Professor, Political Science, Spelman College; tpavri@spelman.edu

Disparities in Cardiology Research and Patients of African Descent: A Data Scientific Approach, Unislawa Williams, Associate Professor, Political Science, Spelman College; uwilliams@spelman.edu

Track 3 – Summer Program Development

The award amount is $25,000.

Provost's Summer Virtual Data Science Training, Research, and Education Initiative, Eric Mintz, Professor, Chemistry, Clark Atlanta University; emintz@cau.edu

A Health Data Science Summer Bridge Program in Biotechnology and Bioentrepreneurship, James Lillard, Associate Dean for Research Affairs, OSRA, with Angelita Howard, Assistant Dean, Online Education and Expanded Programs, Morehouse School of Medicine; jlillard@msm.edu

Abstracts

Track 1 – Summer Curriculum Development: New Module

The award amount is $1,000.

Data Science Module for Database Systems Courses, Alfred Watkins, Academic Program Director & Senior Assistant Professor, Computer Science, Morehouse College; alfred.watkins@morehouse.edu

Database systems courses from all the undergraduate institutions in the Atlanta University Center (AUC) have been included in the lists of data science electives in their respective data science minor proposals. This proposal seeks to enhance traditional database systems courses with a data science module in support of the proposed AUC-wide data science minor. The proposed module can be implemented as a single assignment or as a final project.

How to Manage Data: Creation of an Interdisciplinary Research Data Management Module for AUC Student Researchers, Bryan Briones, Reference Librarian/Facilitator for Research Data Management, Research, Learning and Technology Services Department, Data Science and Technology Services Unit, Robert W. Woodruff AUCLibrary; bbriones@auctr.edu

Training students on how to appropriately conduct research is one of the hallmarks of higher education. However, little has been done to increase students' awareness about research data management. The Atlanta University Center (AUC) Robert W. Woodruff Library proposes the development of a module to introduce the fundamentals of Research Data Management (RDM) to AUC student researchers. Entitled "How to Manage Data," this module will provide an overview of the fundamental concepts associated with RDM and the research project lifecycle. The proposed interdisciplinary module will be comprised of interactive information sessions, hands-on activities to build data literacy skills, and presentations on best practices that should be implemented when conducting research. Social justice-related case studies and other practical activities will afford students the opportunity to apply the presented concepts and to explore how and where data science is incorporated across a research project's life cycle. Upon completion of the full module, student participants will gain a clear understanding about how research and research data management overlap, and the student's role in ensuring that best practices are utilized across the span of future research experiences. The "How to Manage Data" module has the potential to contribute positively to the process of developing within students an ethos of conducting research both effectively, ethically and with confidence.

Say Her Name: Quantifying Black Women's Experience with Police Violence, Celeste Lee, Assistant Professor, Sociology and Anthropology, Spelman College; celestenleephd@gmail.com

Dominant discussions around police brutality and state-sanctioned violence center the experiences of Black men. Meanwhile, the experiences of Black women are often rendered invisible. The African American Policy Forum (APPF) started the #SayHerName campaign to bring awareness to all too often forgotten names of Black women who have been targets of law enforcement. This project embraces the #SayHerName campaign mission by using statistics to further document the intersections of race and gender with regards to police brutality and state-sanctioned violence. In a previous course, students produced a virtual memorial that honors the lives and memory of 25 Black women/girls who lost their lives as a result of police violence. In this module, students will draw upon the qualitative data from the virtual memorial (and other narratives) to construct a dataset that quantitatively documents Black

women and girls' experience with police violence. In doing so, students will learn tools for data organization in spreadsheets, engage in descriptive statistics, and utilize data visualization techniques. The value of data science lies in its power to take raw data and make sense of it, process it, extract value from it, visualize it, and communicate it to broader audiences. The proposed module will allow students to hone their skills in each of these areas. This module is significant to the field of data science because it provides an opportunity to train future data scientists while also centering the importance of data science in making "invisible" groups and experiences more visible.

A Data Science Module for the Community Health Care Course, Kenya Jones, Associate Professor, Whitney M. Young Jr. School of Social Work, Clark Atlanta University; kjones1@cau.edu

The Community Health Care course is an advanced direct practice elective that provides students with the opportunity to analyze health problems in a selected community and is offered in the Whitney M. Young, Jr. School of Social Work - Master of Social Work program. This project will enhance this course by adding a data science module to encourage the emphasis placed onaccess and barriers at health services providers. The data science module will expand the students' understanding of mezzo and macro learn methods of assessing community health issues through data analysis. The roles and functions of multi-disciplinary health teams can be further understood through community needs assessment and interventions via research to identify interventions and determine, through evaluations, which teams are successful. This module will last for two weeks and consist of both individual and group assignments. It also offers students visual, reading, and auditory opportunities to engage and fulfill all learner preferences. Lastly, this course will add to the AUC's dedication to teach all students in various disciplines about data science. By introducing data science within an elective platform, this module will add to the capstone signature assignment as well as provide students with an opportunity to use their voice in identifying issues that they see as prevalent and provide the skills for them to initiate change.

Enhancing Computation in a Physics Capstone Course, Michael Burns-Kaurin, Associate Professor, Physics/Academic Affairs, Spelman College; mburns-k@spelman.edu

The capstone course in Physics at Spelman College, "Physics 462 Advanced Experiments, Theory, and Modeling", requires students to work on three projects. One of those projects involves doing an experimental study of the temperature change of some object in contact with something hotter or colder by, for instance, boiling a potato. Students also use the physics of thermal energy transfer to build a computational model of the temperature change of the object as time passes and adjust the thermal properties of the object in the model to match the experimental data. Currently, the model is one-dimensional and is usually done in a spreadsheet. The proposed module will improve the data science in the project by having a sequence of activities that leads to students using the Python programming language and related libraries of functions to create a three-dimensional computational model. The module also includes finding the numerical parameters in the model that make the model best match the experimental data of temperature versus time.

Track 1 – Summer Curriculum Development: New Course

A Statistics for Data Science Course, Chuang Peng, Faculty, Mathematics and Computational Sciences, Morehouse College; chuang.peng@morehouse.edu

Data Science has emerged out as one of the most popular fields of the 21st century. It uses Mathematics, Computer Science, and in particular, Statistics to perform technical analysis of data and gain deeper and

more fine-grained insights into how exactly our data is structured. Based on that structure, we can optimally apply other data science techniques to get even more information. The proposed new course will start with basic concepts of statistics and probability, such as bias, mean, median, range, percentile, mode, variance, and standard deviation. It will include the key concepts of probability distribution, sampling, statistical significance, hypothesis testing and regression, for example, uniform distribution, normal distribution, and Poisson distribution. It will also introduce Bayesian statistics and machine learning, conditional probability, priors and posteriors, and maximum likelihood. The work in this project includes developing a rationale for the new proposed course, a description of how this new proposed course will be aligned with the mission of AUC Data Science Initiative, setting up learning goals and outcomes for the course. The project aims to develop a complete course syllabus, sample tests and quizzes, as well as sample course projects, along with grading rubrics and assessment plans, both formative and summative, for student performance.

Integrating Data Science and Programming into a New Discrete Structures Course, Mohammad Ali Sazegarnejad, Computer Science Lecturer, Cyber-Physical Systems, Clark Atlanta University; asazegar@cau.edu

Most students start their college with a cloudy vision about big data, data science, and artificial intelligence. Many of our students at the Atlanta University Center (AUC) might have the basic knowledge and skills used in data sciencesuch as mathematics, statistics without knowing how their skills and talent can be used to solve big data problems and apply them in data science projects. The aim of this proposal is to help such students with basic math skills to explore their potential in data science while they pursue their degree in the AUC. This proposal aims to help with creating a holistic learning experience for students in the AUC by introducing applied mathematics fundamentals in data science for instructional use as well as interested students with no prior knowledge of data science. Some of the topics that this course targets are at the intersection of data science and set theory, number theory, intro to statistics, functions (fast and slow growth), derivatives, basics of probabilities, and Bayes' network. These topics are often covered by several courses that the instructor might be unaware of the fact that students can apply what they learn to solve problems related to data science. This course will bring the core fundamentals of mathematical skills used in data science into a new Discrete Structures Course that will enhance students' learning experiences and inform teaching. In this course, students will learn how to implement data science techniques using the Python programming language.

A New Course on Data Visualization at Morehouse, Perry Sweeper, Professor of Practice, Experiential Learning and Interdisciplinary Studies, Morehouse College; perry.sweeper@morehouse.edu

Data Visualization is a new course that will enhance the learning environment within the Atlanta University Center (AUC) by offering content that teaches students the foundational principles, methods, and best practices of data visualization. In the course, students will learn how to analyze data and dynamically visualize it in a human-centered manner that effectively communicates vital information to an audience. The course will culminate with a project centered on digital activism through data journalism based around social justice issues that directly impact the Black Diaspora. The curricula created will be offered within the Applied Software Engineering program currently offered in the Division of Experiential Learning and Interdisciplinary Studies at Morehouse College.

Developing a Data Science Summer Program for K-12 Students, Shonda Lawrence, Associate Professor/PhD Chair, Social Work, Clark Atlanta University, with J. Fidel Turner, Jr., Dean, School of Education, Clark Atlanta University; slawrence@cau.edu

In general, K-12 education curriculum needs more teaching of data literacy (Levitt, 2019). In addition, African American and Latino children lack access to high-level science and math (Loewus, 2016). This proposal addresses the gap for children of color through the development of a data science curriculum. These efforts will lead to a data science program for K-12 students in metropolitan Atlanta, Georgia. A review of current literature and examination of best practices and frameworks for data science curriculum development for K-12 will be conducted. The curriculum will be implemented in a summer program in 2022.

Designing the Data and the African Diaspora Course at Morehouse College, Sonya Dennis, Assistant Professor, Computer Science, Morehouse College; sonyamd@yahoo.com

Data Science is one of the leading careers in technology with an estimated job growth of 2.7 million starting in 2020 (source IBM). This proposal will design a new course in Data Science at Morehouse College in the Department of Computer Science to incorporate into the new data science minor. The proposed course entitled "Data and the African Diaspora" will introduce Morehouse students to the field of data science via coursework using large datasets on topics impacting Black males. Our team will extend the work of the "Data and the African Diaspora" course that was piloted at Clark Atlanta University and Spelman College. To develop this new course, we will use the backward course design that will provide a student-centered approach for data science instruction. Using the backward course design, we will determine student learning outcomes and develop rubrics and assessments to demonstrate evidence of successful learning while providing an experiential learning experience on the Black male experience. The "Data and the African Diaspora" course is proposed to be offered in the Spring of 2022 at Morehouse as an elective.

Track 2 – Summer Research Development

The award amount is $3,000.

A Multivariate Statistical Analysis of the Effectiveness of ALEKS in Improving the Educational Outcomes of STEM Majors, Daniel Teodorescu, Professor, Educational Leadership, Clark Atlanta University; dteodorescu@cau.edu

This project will develop and use data science methods to examine the effectiveness of an adaptive learningsystem (ALEKS) for African American students in mathematics gateway courses, while controlling for pre-college achievement and demographic characteristics. For the past three academic years, Clark Atlanta University has implemented ALEKS to support classroom instruction in pre-calculus and calculus courses. The impetus for this intervention was an analysis of pass/fail data across all STEM gateway courses which revealed that many of these classes had a high failure rate. The causes for this failure are rooted in the inequities experienced in the K-12 education system by many African American students. The U.S. Department of Education Office of Civil Rights (2014) categorized the deep racial disparities in high schools in three key areas that are critical for college readiness: the level of coursework available, the experience level of the teachers, and access to guidance counselors. This project seeks to use a multivariate statistical analysis to establish whether the implementation of ALEKS improves student mastery of learning outcomes in Calculus and Pre-Calculus and retention rates in STEM majors. The findings of this project can inform other institutions on how to better support African American students pursuing a STEM major.

Data Visualization Platform for the War on Drugs with a Focus on Atlanta, Georgia, Fernando Esquivel-Suarez, Assistant Professor, English, Spelman College; fesquive@spelman.edu

The project aims to carry out initial collection and visualization of data related to the War on Drugs (WoD) in the city of Atlanta, Georgia, from the 2000s until the present. This is an initial step in a larger data visualization project that seeks to produce a transnational comparative perspective of the drug war. The platform will initially visualize and compare data on drug-related incarceration, violence, inequality, police brutality, and racial and gender disparities from Atlanta and the city of Cali in Colombia. This project is a response to the need for data-driven support to the recent scholarship that has demonstrated the disproportionate impact of the WoD in communities of African descent. By presenting and analyzing data on the populations affected by the WoD, this platform will inform and support the efforts by advocates, organizers, researchers, policymakers, students, and general audiences aimed at ending the drug war.

Creating a Database to Categorize Structural Trends and Chemical Functions of Benzalkonium Salts, Issifu Harruna, Professor, Chemistry, Clark Atlanta University; iharruna@cau.edu

Since their invention in the 1930s, some benzalkonium salts have been very effective against some microorganisms. Benzalkonium salts are widely used as disinfectants, biocides and detergents, among a variety of other applications. However, with continuous usage over time, microorganisms develop resistance which can render the benzalkonium salt ineffective. Therefore, it is imperative that the structural and functional groups of the benzalkonium salt be modified to develop a more effective and efficacious agent. Researchers have generally modified a collection of benzalkonium salts, characterized and then tested them against microorganisms. This “hit or miss” approach is time consuming and expensive. Hence, we will develop a database of benzalkonium salts and study their structure activity relationships (SAR) against different microorganisms. The database will be used to inform the development of future antimicrobial agents. Furthermore, the database could be enhanced by artificial intelligence (AI) to predict drug – microorganism interactions. The results of this work will serve as a basis for a proposal to be submitted to the NSF or NIH.

Predicting the Disease Threat of Microbial Pathogens Using Data Science, Jeticia Sistrunk, Assistant Professor, Biology, Spelman College; jsistru2@spelman.edu

Predicting and understanding the threat of microorganisms to cause human disease by using data science techniques is important because it allows for the analysis of large datasets that were previously difficult to study collectively. Antibiotic resistance is an increasing and ongoing concern for treating many bacterial infections as a result of incomplete or overuse in patients and agriculture. Salmonella enterica is a bacterial pathogen that is a major source of foodborne infection around the world. To understand how bacteria acquire resistance to antibiotics, how prevalent antibiotic resistance genes are in soil and water in the state of Georgia, and how we can use biological datasets to predict the threat of antibiotic resistant infections, we will analyze DNA from 50 Salmonella enterica bacteria from both environmental and clinical sites within the state of Georgia. Data science tools will be designed to compare large sets of biological data will be used to identify the genes and associated with antibiotic resistance and clinical disease severity. Algorithms and mathematical models will be trained on this dataset to predict the threat of human disease based on features in the DNA. Lastly, functional analyses will test the results of these predictions by identifying the production of antibiotic resistant and virulence proteins produced by these bacterial strains in the laboratory. This work aims to explore how environmental sources may lead to clinical infection and will highlight the importance of investing research and resources on the surveillance of dangerous pathogens.

Using Data Science and Experimentation to Determine Properties of Novel Materials, Natarajan Ravi,

Professor, Physics, Spelman College, nravi@spelman.edu

Nanoscale materials (a billionth of a meter) are widely used in electronic devices due to their promising electrical properties, especially high conductivity. This project intends to incorporate optimization process and numerical analysis techniques to accomplish data analysis of the complex spectroscopic data by writing codes with Numpy, a library for the Python programming language. Two classes of nanomaterials a) metal encapsulated carbon based nanotubes/graphene and b) perovskites based interface oxides are of interest and work is currently underway. Experimental techniques such as X-ray diffraction and Mössbauer spectroscopy are employed to unravel structural, electronic, and magnetic characteristics of such systems. This project will build on our preliminary results on these systems and this project will provide valuable information on spin-canting mechanism of the oxide system.

Using High Performance Computing and Machine Learning Tools to Develop Fullerene-Free Acceptors for Organic Solar Cells, Seyhan Salman, Assistant Professor, Chemistry, Clark Atlanta University; ssalman@cau.edu

There is no more significant challenge facing our planet than the climate crisis. Among the renewable energy sources, solar energy has the greatest potential to sustain the lifestyle of more than 7 billion people. Non-fullerene materials have attracted attention as high-performance molecular acceptors in organic solar cells (OSCs). In order to address current materials challenges and support new product development, this project will develop and apply a multiscale theoretical approach combining high-performance computing (HPC) and machine learning (ML) approaches. These approaches will determine the key electronic properties (e.g., highest occupied molecular orbitals levels, lowest unoccupied molecular orbitals levels, and bandgap) and develop a correlation between the molecular structure and physical property for the non-fullerene organic solar cells efficiency prediction. The project will impact the development of non-fullerene OSCs in two ways: (1) it will help to extract complex correlation between various descriptors and device performance, and (2) will improve device performance by utilizing data to expand material design and device architecture. Moreover, the project has components in innovation and workforce talent that aligns with the Atlanta University Center (AUC) Data Science Initiative goals and will contribute to infusing data science-related activities at the AUC.

COVID-19 and the 2020 General Election Participation Among African Americans, Teri Platt, Associate Professor, Public Administration, Clark Atlanta University; tplatt@cau.edu

Using 2020 American National Election Study (ANES 2020) data, noting that only the preliminary pre-and post-election data are currently available, this project seeks to understand the correlation relationships and multiple variable connections between attitudes and behaviors about the 2020 general election and COVID-19. Additionally, variable relationships between (1) election candidate preference, (2) attitudes about civic/political engagement, (3) COVID-19 vaccine hesitancy and access, (4) COVID-19 vaccine participation, and (5) self-health assessments will be explored. This project adds value to data science-driven inquiry because it transforms data into valuable information about political and health-related choices. The project will advance understanding of the relationship between individual-level political attitudes and behavior and attitudes about personal health efficacy. The work is an extensionof a “Pilot Study of the Relationship between Health and Political Engagement (HPE) in African-American Men” which is supported by the Center for Cancer Research and Therapeutic Development (CCRTD) and the National Institutes of Health (NIH) through the Research Centers at Minority Institutions (RCMI). The results of this research project and the pilot study will support a grant proposal to the National Institute for Minority Health and Health Disparities (NIMHD) of the NIH.

Using Survey Data to Understand the Election of Women Leaders in South Asia, Tinaz Pavri, Professor, Political Science, Spelman College, tpavri@spelman.edu

South Asia has historically been a space where women have struggled in claiming their rights. After independence from the British, the region was left impoverished and infrastructure poor, and women bore the greatest brunt. However, in five countries, India, Pakistan, Sri Lanka, Myanmar, and Bangladesh, women leaders have been elected to the highest office of Prime Minister (or, in the case of Myanmar, Counsellor). This presents a puzzle. Why have women been widely discriminated against, but yet voters have elected women to the highest office of the land? We turn to survey data for answers tothis question. Utilizing the World Values Survey (WVS), this project will track voters' values and beliefs as capturedin the WVS data during specific time frames to see how they responded to questions of gender, equality, family, and diversity. The data will provide a context for analyzing the election of these women leaders. It will provide valuable insight into the mindset of voters about women, gender, and diversity issues in these countries and shed light on why they might have elected women leaders. This research will be of interest to those following questions of gender, leadership, the developing South, and those interested in data science around opinion and values surveys.

Disparities in Cardiology Research and Patients of African Descent: A Data Scientific Approach, Unislawa Williams, Associate Professor, Political Science, Spelman College, uwilliams@spelman.edu

Despite significant advances, articles about cardiac health of populations of African descent continue to be disproportionately rare. Quantifying not only the number of articles published but also the interest in relevant research can highlight the benefits to authors, especially junior scientists, of specializing in the cardiac health of populations of African descent. The study's approach borrows several techniques from data science. The study utilizes culturomic analyses, common in the social sciences, to examine interest in cardiac health disparities. The study combines these with a bibliometric approach, which evaluates original research published in the high-impact cardiology journals in the past forty years. Of the 56,005 publications analyzed so far, only 408 research articles or less than 1% reference populations of African descent. However, those that do have a significantly higher impact (Mood's median test, p-value < 0.001) with a median citation count of 30.5 compared to others (median citation count of 10). Finally, widespread use of phrases related to cardiac health disparities shows hunger for relevant research insights among the public. The percentage of publications about the health of populations of African descent has grown steadily though it remains low. However, high citation counts of the same articles indicate strong demand for relevant research. Utilizing both the data science and social science tools reveals a challenge and an opportunity to diminish disparities in medical research.

Track 3 – Summer Program Development

The award amount is $25,000.

Provost's Summer Virtual Data Science Training, Research, and Education Initiative, Eric Mintz, Professor, Chemistry, Clark Atlanta University; emintz@cau.edu

This project will allow us to expand two data science education and research training programs the 1) CAU Provost's Summer Virtual Faculty Training in Data Science Initiative and the 2) CAU Provost's Summer Virtual Undergraduate Research Participation in Data Science Initiative. These programs will run June 1 – July 30, 2021, and will be 100% online, and will utilize a course from HarvardX's Professional Certificate in Data Science. Faculty Participants will complete R Basic, Data Visualization, Data Wrangling, and will be encouraged to complete one or more courses from Linear Regression, Probability,

Productivity Tools, Data Set ad Research Questions, and Inferencing and Modeling depending on their discipline and intended applicationof data science in their teaching and/or research. The selected courses will provide faculty with the essential foundation to develop and implement data science modules into existing courses during the 2021-2022 academic year and mentor students conducting data science projects and prepare a draft proposal to seek external funding to support their education or research agenda. Undergraduate participants will complete R Basic, Data Visualization, Probability, Data Wrangling, Linear Regression, Productivity Tools, Data Set and Research Questions, and Inferencing and Modeling. These courses will provide students the foundation to complete a research project in data science with a faculty mentor and enhance their research training and professional development. Students will maintain and submit a daily activity log outlining daily goals, completion of material, and any challenges that occur. Faculty and students will present their summer projects at the end of the summer.

A Health Data Science Summer Bridge Program in Biotechnology and Bioentrepreneurship, James Lillard, Associate Dean for Research Affairs, OSRA, with Angelita Howard, Assistant Dean, Online Education and Expanded Programs, Morehouse School of Medicine; jlillard@msm.edu

This project develops an innovative pathway that facilitates the transition into the biotechnology industry through a 5-week summer program entitled “Bridges to Biotechnology & Bio-entrepreneurship” (B2BB). Despite representing 13-15% of the US population, African Americans represent less than 3% of the biotechnology workforce. The B2BB program has a focus on engaging African American participants who are rising college senior and postbaccalaureate professionals and supports the pathway to the Master of Science in Biotechnology (MSBT) at Morehouse School of Medicine. The project will expand health data science education in the B2BB program to support the transition into the biotechnology industry as well as the MSBT, which is now ranked #1 out of 139 similar programs by Intelligent.com. The B2BB program’s training in health data science includes a module on public health informatics that includes statistics and its uses and visualization; databases and an introduction to MySQL, manipulating data sets with HTML, XML, and JSON; and an introduction to data analysis and python coding. The module on biomedical data science includes principles and databases, an introduction to SQL, cloud computing, and data mining. The module on making medicines will integrate the health and data science fields to investigate drug discovery, development, and testing. Finally, the B2BB program will have a module on bio-entrepreneurship built on the I-Corps model. This project will address the workforce diversity needs of biotechnology and will provide training to rising seniors and postbaccalaureate professionals in health data science, biotechnology, and bio-entrepreneurship, including how biotechnologies and health data science is used to bring these products to the market.