2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #50302

Data Science 101 for Faculty: Designing and Assessing a Foundational Workshop

Dr. Karl D. Schubert FIET, University of Arkansas

Dr. Karl D. Schubert is a Professor of Practice and Director for the Data Science Program at the University of Arkansas College of Engineering, the Sam M. Walton College of Business, and the Fulbright College of Arts & Sciences and led the team that developed the B.S. Data Science degree.

After earning his Ph.D., Dr. Schubert spent 35 years in industry in various roles with IBM, Dell, Silicon Valley, and other start-ups. He returned to his alma mater to give back by helping develop the next generation of STEM workforce. He has patents in various technology areas and is the author and co-author of several books.

Dr. Schubert is a Senior Member of the IEEE, Senior Member of ACM, and Senior Member of IISE. He is also Vice Chair of the Ozark Section of the IEEE Computer Society and is the ASEE Data Science & Artificial Intelligence (DSAI) Constituent Delegate to the Commission on P-12 Engineering Education (CP12) and the DSAI Delegate to the Interdivisional Town Hall.

Stephen R. Addison, University of Central Arkansas

Dr. Stephen R. Addison is a Professor of Physics and Dean of the College of Science and Engineering at the University of Central Arkansas. Dr. Addison joined the faculty of the University of Central Arkansas in 1984, and has previously served as Dean and Associate Dean of the College of Natural Sciences and Mathematics, Chair of the Department of Physics and Astronomy, Interim Chair of the Department of Computer Science, and Interim Director of the Center for Mathematics and Science Education. Dr. Addison was one of the founders of the Arkansas Space Grant Consortium. He has been active in research throughout his career and has been active in innovative educational efforts across Arkansas in computer science, data science, mathematics and physics.

Dr. Addison is a Senior Member of the IEEE, a member of the Arkansas Academy of Computing, a member of Sigma Xi, and is a Life Member and Secretary of the Arkansas Academy of Science. Dr. Addison has previously served as the President of three professional organizations: the Arkansas Academy of Science, the Arkansas Deans' Association, and the Arkansas-Oklahoma-Kansas section of the American Association of Physics Teachers. He is also a member of the Acoustical Society of America, and the American Association of Physics Teachers.

Dr. Carol S Gattis, University of Arkansas

Dr. Carol Gattis is Associate Dean Emeritus of the Honors College and an Adjunct Associate Professor of Industrial Engineering at the University of Arkansas. With over 30 years of experience in STEM education, she has led major efforts to build innovative programs, expand student pathways, and align academic initiatives with workforce needs. A frequent PI or Co-PI on multi-million-dollar NSF and state grants, Dr. Gattis focuses on student success, transfer pathways, and industry collaboration in engineering and data science.

Jessica Hilton, University of Arkansas

Jessica Hilton is a Data Science Program Coordinator at the University of Arkansas, where she plays a key role in coordinating curriculum, faculty learning workshops, and certificate program development, as well as fostering statewide partnerships and supporting student success initiatives. With over 17 years of experience in adult and postsecondary education, her work centers on expanding equitable access to data science and technical education across Arkansas and strengthening pathways for adult learners to succeed in a rapidly evolving workforce. She holds an M.Ed. in Adult Education and Lifelong Learning from the University of Arkansas.

Devanshu Priyam, University of Arkansas

Devanshu Priyam is a graduate researcher and engineer with both academic and industry experience in the field of computer science. He is pursuing an M.S. in Operations Management (College of Engineering) at the University of Arkansas, where he serves as a Graduate Research Assistant in the Data Science Department under the Arkansas HIRED State Grant. In this role, which he has held for over a year, he supports departmental research and training initiatives aimed at advancing and expanding standardized data science practices in higher education across Arkansas. He also brings over 4.5 years of industry experience as a Senior Software Engineer at a global IT company. He holds a Bachelor of Engineering in Computer Science from Visvesvaraya Technological University, India (2020). His technical expertise spans various programming languages, machine learning, cloud platforms, and enterprise software architecture and solution development.

Recently, Devanshu was awarded the 2026 Graduate Academic Award by the University of Arkansas (UARK), recognizing his sustained academic performance, his research contributions within the College of Engineering, and his collaborative research with the University of Arkansas for Medical Sciences (UAMS) and two additional centers under the UARK Division of Research and Innovation. He remains committed to advancing engineering education through applied research and service to the broader engineering community.

Data Science 101 for Faculty: Designing and Assessing a Foundational Workshop

Abstract

To expand data science education capacity in Arkansas, the HIRED workforce development grant [1] team launched a faculty development initiative aimed at equipping faculty from 2-year colleges and 4-year higher education universities to teach foundational data science. This is particularly important as work and daily life become increasingly dependent on data and more jobs require basic data skills [2]. The first offering in this effort, "Data Science 101 Workshop," was a 1.5-day in-person workshop held in September 2025. Nine faculty members participated, representing a range of disciplines and institutions across the state, many of which serve working adult learners and place-bound students who seek workforce-relevant credentials.

This workshop was intentionally designed for beginner-level faculty, including those with little or no prior experience in teaching or applying data science. Pre-survey data confirmed this need: most participants reported limited confidence and minimal experience in areas such as applying the data science process, using Python for data analysis, teaching exploratory data techniques, or integrating AI tools. Though nearly all had some experience with spreadsheets, few felt skilled enough to teach data science tools effectively. These gaps limit the capacity of programs to offer data science and analytics experiences that help students and incumbent workers upskill or reskill for data-intensive roles [2].

The workshop agenda included hands-on practice with Python in Google Colab, spreadsheet-based data analysis, an introduction to the data science process, and practical use of AI tools for teaching and research. Content was scaffolded to support entry-level faculty, with opportunities for interaction, troubleshooting, and guided examples.

Post-survey results revealed self-reported gains related to feeling more skilled in the topical areas presented in the workshop and self-reported confidence in teaching those areas. For example, participants who self-reported being somewhat confident or very confident in teaching the data science process rose from 33% to 78%, and the confidence in teaching basic data analysis rose from 44% to 100%. Open-ended feedback highlighted the value of hands-on exercises, shared instructional materials, and instructor insights. Participants expressed a desire for more time to apply the tools and more opportunities to collaborate.

Additionally, the feedback from the participants through open-ended questions in the survey and through live "retrospective" discussions were used to improve the subsequent workshops, with the most significant changes being slowing down the pace at which the materials were covered, adding a section on the use of Generative AI, and including time for the workshop participants to partner, use what they learned, and present it to the instructors and other participants.

Introduction: Vision & Strategy

The vision for the state of Arkansas is to have a workforce that is data literate. Through the HIRED grant [1] initiative, the educational ecosystem described in previous papers by the authors [3] – [5] has been expanded to include workforce development [6]. Workforce development, for purposes of the HIRED grant, includes upskilling, reskilling, and career development for those who already have jobs or are upskilling and/or reskilling in both cases to better position themselves in the workforce. This could, for example, be an employee who has recently been placed into a role (or is interested in being in a role) that requires data analyses and needs to upskill or reskill to be able to meet the requirements of the new position.

The strategy to achieve the vision is twofold, building on the previous works:

1. Create workforce-focused state-wide training and education focused on data literacy and analysis.
2. Create Data Science Foundational Workshops for faculty and instructors that is workforce focused.

The first goal is addressed by developing certificate-based modules for the state-wide ecosystem, available in branded and unbranded formats, enabling institutions to deliver certificate courses and workforce training locally through the institutions closest to those needing them.

The second goal is addressed by providing introductory level, hands-on, in-person, subject matter training for faculty and instructors throughout the state. The second goal is the focus of this paper.

This paper contributes a replicable, workforce-aligned faculty development model that integrates pedagogical design principles, tool accessibility constraints, and data-informed iterative refinement, to expand data science instructional capacity.

Overall Workshop Goals and Design

The overall workshop goals were to introduce the concepts of data science and the fundamentals of performing data science to faculty and instructors who would be teaching data science courses or supporting the students who pursue the data science workforce development certificates

The workshop was designed for faculty with limited or no prior experience with or applying data science. Participants are expected to have some experience with Excel and a basic understanding of graphs, but programming and statistics experience is not required. This is a challenge and does cause the workshop to move at a slower pace, but it is critical to meet the faculty at the level they bring. The future more advanced workshops will at least have participants all with a known common minimum set of skills. During the workshop, knowledgeable grant team members move around the room to help participants as they get stuck or fall behind to ensure each participant successfully progresses.

The design of the workshops was based on industry experience and best practices in workforce readiness, pedagogical best practices, and academic experience in designing, implementing, and teaching data science at the post-secondary and dual credit high school environments. As such, it was determined that the best way to host the workshops was to go to the regions where the faculty and instructors were geographically based. This meant the Central, South, East, Northeast, and Northwest geographical areas of Arkansas.

The learning objectives for the workshops are:

- The participants will be able to demonstrate an understanding and hands-on use of data science and the data science and analytics process.
- The participants will be able to demonstrate an understanding and hands-on use of Exploratory Data Analysis (EDA) and apply critical thinking to evaluate the results.
- The participants will be able to draw and communicate inferences and implications from their analyses and present and communicate the results visually and verbally to the level of the listener (technical and/or non-technical).
- The participants will experience working on a team-based EDA and explain the importance of team-based, real-world, real-data projects for data science and analytics workforce readiness.

Experience from previous data science pedagogical workshops informed the design of the workshops that were based on informal feedback from the participants. For example:

1. There is a wide range of post-secondary institutions participating, a variety of level of experience with the subject matter and the hardware and software, and a variety of the "core tools" used by their home institution (i.e., a Microsoft shop or a Google shop).
2. There is a variety of the participants' hardware (i.e., Windows laptops or macOS laptops) and software (i.e., Windows 11, macOS, or Linux).
3. There is a wide range of the participants' experience in downloading software and the use of files systems and folder hierarchies.

These key factors created significant challenges in being able to get the workshops started and to making sure no one was left behind as topics progressed. As a result, "loaner" laptops were provided for the workshop that were pre-configured to be the same and with folder structures to be set up ahead of time and logins of a sequential nature (e.g., DASC_01, DASC_02, ….). A Google file system structure was created on each laptop with a shared Google drive so software, examples, and data files used in the workshops could be pre-loaded and shared real-time during the workshop.

Detailed directions were also provided for the participants – as a pre-workshop activity – to create a Google Account that was not connected to their academic institution's account to ensure there were no restrictions placed on it as those restrictions vary greatly by institution.

The laptops chosen were MacBook Pro systems of the most recently available models. These were chosen because most of the tools used in data science are developed on Unix, Linux, and/or macOS and the Windows Subsystem for Linux (WSL 2) is not fully supportive of the libraries typically used in data science (as was discovered very early on in the B.S. Data Science degree program).

The business applications were chosen based on workforce readiness: the Microsoft suite for business applications: the online free versions of Microsoft Word, Excel, and PowerPoint. The programming / integrated development environment (IDE) chosen was Google Colab (pronounced Co' – Lab). This environment was chosen because:

1. It does not require any downloading of software, and
2. It has excellent debugging capabilities for beginners.

In addition, the tie-in with Gemini was useful for the portions of the workshop on using AI for data science.

These design choices reflect an intentional shift from content transmission toward capacity building, prioritizing faculty confidence, tool transferability, and instructional feasibility in heterogeneous institutional contexts.

The agenda for the most recent workshop is shown in Table 1, below.

		HIRED Data Science 101 Foundational Workshops	
Day 1		**Topic**	**Speaker**
8:00 AM	9:00 AM	Laptop / Google Account Verification. Presurvey, & Workshop Overview - Warm Breakfast Provided	Program Coordinator (PC) Graduate Assistant (GRA), & Co-PI Gattis
9:00 AM	10:00 AM	What is Data Science and How is it Used?	Principal Investigator Schubert
10:00 AM	10:15 AM	Break - snacks provided	
10:15 AM	Noon	Hands-On Exploratory Data Analysis for Data Science with Excel	GRA
Noon	12:45 PM	Lunch - Provided	
1:00 PM	3:15 PM	Hands-On Exploratory Data Analysis for Data Science with Excel (continued)	GRA
3:15 PM	3:30 PM	Break - snacks provided	
3:30 PM	5:00 PM	Hands-On Data Visualization for Data Science with Excel & GenAI	GRA
5:00 PM		Dinner on your own (not provided)	
Day 2			
8:00 AM	8:30 AM	Warm Breakfast Provided	
8:30 AM	10:00 AM	Hands-On Python Programming for Data Science (including GenAI)	Co-PI Addison
10:00 AM	10:15 AM	Break - snacks provided	
10:15 AM	Noon	Hands-On Python Programming for Data Science (continued)	Co-PI Addison
Noon	12:45 PM	Lunch - Provided	

12:45 PM	2:30 PM	Hands-On Python Programming for Data Science (continued)	Co-PI Addison
2:30 PM	2:45 PM	Break - snacks provided	
2:45 PM	4:30 PM	Hands-On Python Programming for Data Science (continued)	Co-PI Gattis
4:30 PM	5:00 PM	Wrap-up, Post-survey, & Feedback	Co-PI Gattis

Table 1. The HIRED Grant Data Science 101 Foundational Workshop Agenda

Workshop Planning

Preparation for the data science workshop began by identifying a location in the targeted state region that was easy to access and equipped to support interactive computing activities. The team looked for venues, and successfully located them, that could be reserved without rental costs, such as non-profits, education centers, or community rooms. The venue was reserved, and internet connectivity and audio-visual equipment were checked well in advance.

Workshop attendance was limited to faculty nominated by invited institutions. Outreach focused on partner and potential partner schools in the region, including community colleges, universities, and technical institutions. Invitations were sent to department heads, program directors, and other leaders in data science education, encouraging them to share the workshop details with faculty at their institutions. Participant schools were asked to nominate two faculty to attend, with additional seats added for schools requesting more than two participants, as space allowed. The email invitation for the most recent workshop is shown below.

> Good morning, HIRED Southern State Academic Partners!
>
> We are excited to invite your institution to participate in the upcoming Data Science 101 Workshop, hosted by the HIRED Data Science team! This hands-on, two-day learning experience is designed to empower faculty and administrators to bring new data science skills and strategies back to their campuses. This workshop is open to individuals who teach (or will teach) Data Science or administer Data Science programs/curriculum.

Each institution is invited to send **two participants** to the event. A **$1,500 stipend** will be provided to each attendee who participates fully in both days of the workshop.

☞ **RSVP here:** [active link URL]

Workshop Details

- **What:** Data Science 101 Workshop
- **When:**
    - **Thursday, November 20** | 8:00 AM – 5:00 PM
    - **Friday, November 21** | 8:00 AM – 5:00 PM
- **Where:**
    - [Location address]
- **Who Should Attend:**
    - Faculty who currently teach or plan to teach data science
    - Program administrators involved in data science curriculum or initiatives
- **Stipend:** $1,500 per participant (contingent on full participation in both days)
- **Meals:** Breakfast and lunch provided both days
- **Technology:** Laptops will be provided and are required to be used (personal laptops may not be used)

Workshop Experience

- Sessions and skills build on each other
- Highly interactive format: learn by doing
- Participants will receive ongoing support from the HIRED initiative Data Science team
- Opportunities to collaborate with peers from across the region
- Access to learning materials from workshop

While we have planned for two people per institution, if you would like to request more than 2 attendees, please feel free to contact me.

We look forward to welcoming your team to this energizing and practical workshop, and to continuing our work together to expand access to high-quality data science education across state. **Please RSVP ASAP, space is limited!**

Best regards,
Signature

Participant registration was collected through a Google Form RSVP page. This allowed the team to provide logistical details while gathering information from participants, including name, institution, role/title, dietary restrictions, expected learning outcomes, familiarity with basic data science software (Google Drive, Excel, Python), and Gmail account details. Gmail accounts were required to access a shared folder during and after the workshop. During registration, participants were asked whether their Gmail accounts were active and had been accessed recently. Once registration was complete, confirmation emails were sent to review key workshop details and check Gmail accounts, helping ensure all participants were ready before the event.

As part of the grant, participants received a $1,500 stipend to offset costs such as travel, lost work time, and childcare. To receive the stipend, participants had to attend and actively participate for the entire workshop. The team preloaded participants into the Workday financial management system which helped make stipend disbursement faster once workshop completion was confirmed.

Workshop Logistics

The workshop team included three data science educators, one graduate research assistant (GRA), and one program coordinator (PC). Two educators led the instructional sessions, while the third focused on research surveys and tracking participant learning outcomes using Qualtrics. The GRA presented content sections, provided technical support, and assisted participants as needed, while the PC oversaw logistics, communications, and the overall participant experience.

The team prepared materials ahead of time such as presentation materials, surveys, Apple laptops, learning aids, drinks, snacks, and sign-in sheets to ensure workshop flowed smoothly. Laptops were preconfigured for participants with a guest account and all required programs (e.g., Google Chrome, Excel). These laptops remain the property of the team for future workshops.

For the second iteration, a printed and digital learning guide was developed. The guide included the Wi-Fi login, commonly used programs and tools, key data science definitions, Apple laptop shortcuts, and instructions for accessing shared folders and workshop activities. This allowed participants to use the guide throughout the workshop, which reduced confusion and let them focus more fully on the activities.

Problem-Solving and Preparedness

Several challenges came up during the first workshop that required creative problem-solving. Catering, previously handled by the venue, failed to deliver lunch on time and did not meet specific dietary requests. For later workshops, the program coordinator took over all catering responsibilities. Meals were delivered as planned, and all dietary needs were met, so the same issues did not happen again.

Technical issues also occurred. On the second day of the first workshop, the projector stopped working unexpectedly, causing a brief scramble while the team tried to restart it. Internet connectivity was inconsistent during some sessions as well. In response, the team purchased a backup projector and stand, verified strong internet access at the venue before the second workshop, and identified an on-site technical contact. With these changes, the second workshop ran smoothly, with minimal interruptions.

Lessons Learned, Best Practices, and Refinements

Across workshop iterations, several transferable design principles emerged that informed subsequent refinements and may guide similar faculty development efforts. These included prioritizing proactive logistical planning and participant support, such as checking technical requirements ahead of time, providing backup equipment, offering pre-workshop guidance, and directly managing catering.

The learning guide introduced in the second workshop was especially helpful. Participants could refer to it whenever needed, which kept them engaged and reduced confusion. Moving forward, planning and communicating with participants before the workshop will continue to guide both workshop design and delivery.

Workshop Implementation

The agenda for the workshops was published beforehand. As shown in Table 1 above, the agenda for the workshops includes:

1. Laptop / Google Account Verification & Workshop Overview
2. Pre-survey of participants
3. Introduction to Data Science
4. Hands-on Exploratory Data Analysis (EDA) with Excel
5. Hands-on Introduction to Python Programming and GenAI for Data Science
6. Post survey of participants

Each topic is described, below.

Laptop / Google Account Verification & Workshop Overview

Although explicit instructions were provided for creating a non-academic, institution-based Google account, it was important to assist participants who may not have completed this step or who encountered challenges for a variety of reasons. Each participant was provided with a laptop workstation upon arrival, and successful login and access to their Google account was confirmed prior to beginning the session. For participants who had not completed the pre work, staff provided assistance during the initial preparation, verification, and breakfast period to ensure they were fully ready to begin. Participants used the same laptop for both days. After the workshop ended, laptops were wiped clean in preparation for the next workshop.

After participants had completed preparation and breakfast, the workshop commenced with the pre-survey described below in the section entitled: *Survey Results from the "Data Science 101 Foundational Workshops."*

Introduction to Data Science

The presentation for the introduction is *What is Data Science & How is it Used*. Topics include:

1. Data science as data-focused in contrast to computing or programming focused.
2. Data Science roles and how they interact with data engineers and business stakeholders.
3. Data Science as a Venn Diagram of
    a. Domain Expertise (Concentrations)

b. Mathematics & Statistics (Probability & Optimization, Machine Learning, Cloud Computing, AI)
c. Computer Science (MIS, Business Fundamentals)
4. The definition of a data scientist by the National Academy of Sciences [2].
5. Learning outcomes for a Data Science Program, based on the ABET accreditation draft criteria [7].
6. Other professions that use or rely on data and what differentiates them.
7. The impact and growth of Data Science jobs in the United States based on data from the U.S. Bureau of Labor Statistics (BLS) [8].
8. The data science and analytics process overview.
    a. Understanding the need and what came before.
    b. The data.
    c. Cleaning the data.
    d. Wrangling the data.
    e. Exploratory Data Analysis (EDA).
    f. Modeling.
    g. Understanding the results (applying critical thinking, concluding, interpreting, implicating).
    h. Communicating the results (answer first, tell the story, verbally and visually).
    i. Iterative improvement (retrospective).

The process is visualized in Figure 1, below.

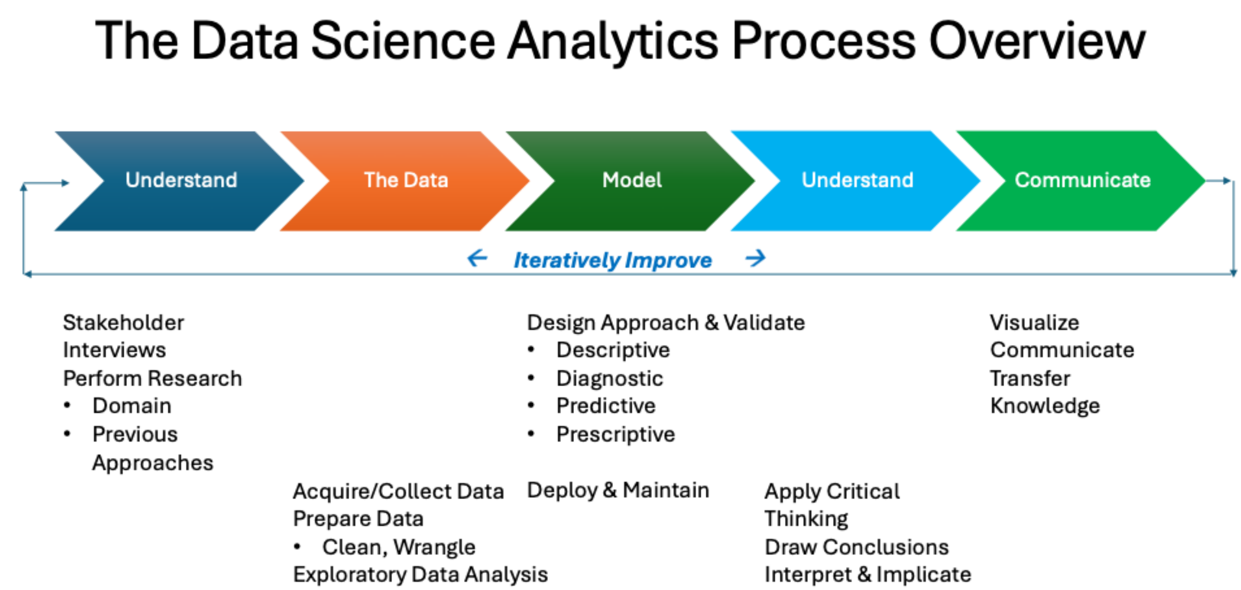

Figure 1. The Data Science Analytics Process Overview (developed by the authors)

The purpose of the presentation is to establish a data-based foundation and way of thinking, to put data science in perspective to other data-related professions and differentiation from information systems or computer science, and to provide a foundational process for understanding the question or problem, understanding the data, preparing the data for analysis, analyzing the data, drawing conclusions and recommendations from the data, and communication the results of the analyses.

The nomenclature, vocabulary, and processes of data science are fully integrated and continually used and reinforced throughout the remainder of the workshop.

This leads us to the next workshop topic, which uses the process and introduces basic data analysis using the free version of Excel on the Web.

Hands-on Introduction to Exploratory Data Analysis (EDA) with Excel

This session was designed with an objective of capacity building for the participants, who were teaching faculty and instructors from diverse disciplines that included non-data science, non-computer science, and non-STEM disciplines. The goal was for participants to leave with a practical, transferable understanding of exploratory data analysis (EDA) from basic to intermediate level that they could apply in their own domains and begin integrating into classroom instruction. Excel was intentionally positioned as a conceptual scaffold rather than as an endpoint tool, allowing participants to internalize the EDA mindset before transitioning to programming-based workflows. In addition, Excel is one of the most widely used analytical tools in industry. Microsoft Excel on the Web was chosen as the primary tool because it is widely accessible, familiar, and cost-free. The session was structured in such a way that participants could succeed with no prerequisites. Table 2 below shows topics presented during this EDA session.

Schedule of Topic: Introduction to EDA with Excel
Welcome, Learning Outcomes & Expectations
EDA Mindset & Workflow (questions, iteration, limits)
Understand the data: structure, variable types & measurement scales
Data quality essentials: missingness, outliers, validity, & consistency
Descriptive summaries: Center, Spread, & Distributions
Visualization for insight: choosing the charts for the question
Hands on setup: dataset, data dictionary, import, & Excel tables
Profile the dataset: first impressions (types, missing values, anomalies)
Clean & Prepare: standardize text, handle errors, remove/flag issues
Summarize with PivotTables (group, compare, and slice)
Visualize & Communicate: charts, key takeaways, & EDA checklist wrap-up

Table 2. Agenda for Session: Hands-on Introduction to EDA with Excel

The initial priority revolved around first developing a theoretical understanding, so that participants could understand what EDA is before learning how to apply it. EDA was introduced as a mindset of open-ended inquiry rather than a fixed procedure, and some basic guiding questions were used to anchor that mindset, such as:

- What is the structure of a dataset?
- Where are the missing values or anomalies?
- What patterns or relationships might exist across variables?

This framing was paired with defining limits. EDA supports discovery and hypothesis generation, but it does not establish causal claims, and it should not be treated as conclusive proof. Understanding these distinctions was especially important for this audience, who would later translate the ideas into instructional activities and help them design assessments.

Similarly, to strengthen data literacy, shared vocabulary in data science and analytics and some basic data evaluation standards that apply across disciplines were introduced. Participants learned the levels of measurement (nominal, ordinal, interval, ratio); core data-quality dimensions such as completeness, consistency, validity, timeliness, and accuracy; and basic statistical analyses, such as mean, median, range, and quartiles. This common language positioned datasets as raw material that must be examined and prepared before analysis can be trusted.

Furthermore, data visualization was focused on as a data understanding and decision-making tool rather than a decorative presentation artifact. Participants learned about the chart selection based on their analytical purposes, such as distributions, comparisons, compositions, or relationships, and the structure of their data. Some common EDA interpretation pitfalls, particularly the difference between correlation and causation, were reinforced. This segment concluded with ethical and professional responsibilities and by emphasizing that biased data, inappropriate visual choices, or careless interpretation can produce misleading conclusions. (cf. Figure 1)

After establishing the theory aspect, the session transitioned to hands-on implementation in Excel to mirror the same workflow in practice using some real-world datasets. For the hands-on experiences, the following open public datasets were used: the Titanic dataset [9] and The Metropolitan Museum of Art Open Access dataset [10].

The participants began by importing the data, performing first impression checks, and consulting the data dictionary as needed. They converted the data into Excel tables and performed first impressions checks such as row/column counts, variable types, and quick scans for missing values or unusual entries. Next, they conducted data-quality checks and cleaning using basic Excel tools. Sorting and conditional formatting were used to flag missing values, duplicates, outliers, and suspicious and invalid ranges. The next steps focused on text standardization and error handling. This stage reinforced a key message that many quality issues are visible through systematic checks before any data analysis is attempted.

Participants then moved into descriptive statistics and structured summarization. They computed measures of central tendency and spread using built-in functions and PivotTables to summarize outcomes by category. Throughout, interpretability was emphasized so that statistics were generated to answer a specific question, not computed simply because the tool was available.

For visualization, participants created different types of charts aligned with the earlier selection principles, such as histograms for distributions, box-and-whisker plots for spread and outliers, scatter plots for relationships, and category comparisons using bar or stacked bar charts (with pie charts used only when justified). Every chart had to earn its place by answering a defined question, and decorative chart elements that do not support interpretation were discouraged.

Important concepts such as data synthesis and communication were addressed. Participants produced a brief, structured summary of findings to practice translating exploration into stakeholder-ready language. This reinforced a core EDA competency that an insight is not complete until it can be communicated clearly, accurately, and responsibly.

The overall teaching design followed a consistent pattern: short concept, guided task, immediate interpretation and reflection, allowing participants to build confidence through visible progress. This practical implementation helped the audience reframe how Excel can be used as a tool for systematic data exploration. Participants also showed a stronger appreciation for EDA as iterative, where each summary statistic or visualization naturally prompts the next question.

This session was concluded by contrasting effective EDA with common anti-patterns, such as creating visuals without a guiding question or selectively hunting for patterns that support a preferred narrative. Participants were able to create a practical checklist for approaching any dataset, along with an ability to justify technique choices based on the question and data type, an ethical lens for interpretation, and an end-to-end EDA workflow executed in Excel on the Web.

Hands-on EDA with Excel and GenAI

Building on the EDA foundation established in the previous session, this session introduced Generative AI (GenAI) as a practical collaborator in the analytical workflow without replacing the conceptual discipline or excel skills participants had just developed. For an audience new to data science, the core message of this session was that participants should be responsible for the question, context, and judgment, while GenAI can be used as a tool for speed, breadth, and drafting, mainly as a complementary instrument**.**

Schedule of topics: EDA with Excel & GenAI
A Quick Recap of EDA Workflow & Session Goals
GenAI in the EDA Workflow
AI Tool Access and its Safe, Responsible Use (best-practices checklist)
Basics of Writing EDA-oriented Prompts
AI-assisted planning for EDA
AI-suggested cleaning steps and Excel functions
Validate results in Excel
AI-assisted Summarization Plan
AI-assisted visualization

Compare workflow: Excel-only vs. AI-Augmented
Communicate findings and wrap-up

Table 3. Hands-On Exploratory Data Analysis (EDA) with Excel and Generative AI Topics

Because the workshop was aimed to help faculty and instructors from diverse domains (e.g., non-data science, non-computer science, and non-STEM domains) adopt data science practices in their own teaching, transferability was emphasized so that participants would be able to reuse this approach in discipline-specific classroom examples where time and tool access may be constrained.

To ensure accessibility, well-known GenAI tools were used, including free versions of ChatGPT and Microsoft Copilot**.** Using low-barrier tools was a key workshop objective so that participants would be able to replicate the workflow at their home institutions with minimal setup.

This session reused the same dataset from the previous session, enabling participants to directly compare an "Excel-only" approach with an "AI-augmented" approach. Rather than presenting GenAI as a one-step solution, the work was structured in a way that every AI output was verified in Excel.

Participants learned to craft prompts that were specific, contextual, and aligned to EDA steps they already understood, such as data preparation and cleaning formulas**,** analysis scaffolding**,** and **s**tatistical interpretation support. The main idea was to learn how to ask good analytical questions and supply enough context so that AI suggestions could be evaluated responsibly.

Participants were explicitly guided to test, debug, and validate AI-generated formulas and steps. This requirement reinforced the value of the previous session, where the foundational understanding gained by the participants enabled them to detect incorrect results. Participants also learned turning exploration into insights through various visualization tables and charts based on their based on the analytical question.

Therefore, this demonstrated to the participants the means to bring data science into their teaching domains, the session demonstrated a practical pathway that starts with a disciplined EDA foundation, then uses GenAI to accelerate execution and expand perspective. This session framed GenAI not as a replacement for skill development, but as a force multiplier and accelerator once foundational competence is established.

Hands-on Python Programming and GenAI for Data Science

As anticipated, participants in the Python sessions brought a wide range of abilities and prior experiences to the sessions. Some had little or no prior experience in programming, others were unfamiliar with Python, and a small number had experience in building machine learning models but no real exposure to data science. This range of abilities is challenging, but the sessions were designed so that all participants would be able to increase their knowledge of using Python in data science and would gain enough confidence so that they would continue to engage with the learning materials after the workshop had ended.

Schedule of Topics in Python Sessions
Session 1
Introduction to the Instructor
The History of Python
Python Distributions and Interactive Development Environments (IDEs)
Thoughts on Programming
Introduction to Google Colab
Entering Text and Code in Colab
Using Python/Colab as a Calculator
Variables in Python
1st Program Total cost of items including sales tax
Python Modules
Naming Conventions for Imported Python Modules
2nd Program – The Fibonacci Sequence
Basic Mathematical Operations in Python
Python.org – Documentation and Downloads
Active State Python
Anaconda Python
Python References for Further Study
Session 2
Python Packages and How to Use Them
Generating Data: Range and Linspace
Introduction to Matplotlib
Introduction to Numpy
Making Plots
Introduction to Pandas
Plotting in Pandas
Introduction to Data frames
Importing .csv files and other File Types
EDA with Pandas

Modifying Data Frames including Generating Data Series
Session 3
Additional Methods of Getting Data into Colab
Mounting Google Drives in Colab
More EDA
Overview of Pandas commands for EDA
Modifying and Exploring Dataframes
Correlation
Heatmaps and Seaborn
Introduction to Polars for large Datasets
Introduction to AI
Google Gemini
You.com
Ethical uses of AI in Data Science

Table 4. Schedule of Topics for Python Sessions.

It is important to use effective pedagogies even in a workshop environment. Empirical research in learning has been conducted for more than 135 years. Brown *et al.* [11] and Agarwal and Bain [12] summarize the results of modern cognitive learning theory that underlie the design of the Python sessions under discussion and those that will be offered in the future. In short, the workshops are based on active learning, interleaving, and spaced repetition, and they build on the knowledge that participants already have by linking new concepts to those with which they are already familiar.

The sessions were built around the Google suite; coding was performed in Google Colab [13] and session slides were created with Google Slides [14]. The sessions incorporated both coding and instruction using Google Colab. Some of the earlier sessions were captured in Google Slides so that participants could use them as step-by-step instructions if they were unable to remember how to use the Colab notebooks after the sessions.

The sessions were designed to teach content and to reinforce effective pedagogy, with instructors both teaching the material and discussing why the content was presented in a particular way. First, the history of Python was reviewed; this included an introduction to available resources and a discussion of why Google Colab was chosen. It was explained that an advantage of a tool like Colab is that it enables participants to get started without installing applications or packages on their own computers and that Colab can be used on any operating system. While participants at the HIRED workshops were using issued computers, this point was emphasized because participants are expected to teach the same content themselves in the future, often to students using their own devices.

While the workshop focus was Python for data science, sessions began with a general introduction to Python using the official Python website [15]. At the Python website, participants were introduced to the available documentation and directed to the beginner's guide. They were also shown how to download standalone versions of Python, and several interactive development environments (IDEs) commonly encountered in practice - IDLE, Eric, Visual Studio Code, and Spyder - were discussed. Jupyter Notebooks and JupyterLab [16] were also introduced, and it was explained that Google Colab is an online version of Jupyter that includes the Python libraries that must otherwise be downloaded and installed for local use. A brief introduction to the Anaconda Python distribution [17] was provided; participants were told that it is widely used in the data science industry and that many books and resources describe its use, such as Vaughan's book on Python tools [18]. Additional general books on using Python in data science were also recommended [19] - [22].

The session began with an introduction to programming. It was noted that although Python is an object-oriented programming (OOP) language, it is multi-paradigm and does not require the use of OOP. Introductory data science investigations more often use procedural programming. Vaughan [18] provides an accessible, example-based introduction to OOP that helps clarify the differences between object-oriented and procedural paradigms. To begin using Google Colab, participants navigated to the welcome page and opened a new notebook. They learned the difference between code cells and text cells, were introduced to Markdown for writing explanatory text, and were shown how to incorporate mathematical expressions into text cells.

Python/Colab instruction began with learning basic arithmetic operations by using Python code cells as a calculator. Through this process, participants learned how to execute code cells and progressed to assigning values to variables, understanding Python's basic mathematical operations, and expanding Python's capabilities through modules. Over the course of the workshop, participants were introduced to several commonly used modules, including NumPy [23], SciPy [24], Matplotlib [25], pandas [26], Seaborn [27], and Polars [28]. Basic programming concepts were reinforced through exercises such as computing total cost with tax and exploring loops and output formatting using the Fibonacci sequence. These programs were analyzed step by step to support participants with limited prior programming experience. The remainder of the workshop focused on using Python and Colab to perform introductory data science tasks.

Participants were introduced to data visualization and produced a range of plots using Matplotlib. To support this, instruction was provided on generating data using Python's range function and NumPy's linspace function. Plotting directly with pandas was then introduced, during which participants learned how to import .csv files from online sources into Colab and explore additional exploratory data analysis (EDA) capabilities in pandas. Most of the major EDA functions in pandas were demonstrated. Artificial intelligence tools, including Google's Gemini [29], were also introduced. Gemini is integrated into Colab and suggests corrections when errors occur; when a Pandas DataFrame is loaded, Gemini can also generate code and suggest visualizations. Errors were intentionally introduced into code cells to demonstrate Gemini's error-correction capabilities. The resulting plots were used to explore datasets, and a basic introduction to categorical data was provided.

After this introduction, additional methods for importing data into Colab were explored, including transferring files to and from local drives and Google Drive. Approaches for working with other data formats, including Excel and SPSS files, were discussed. Instruction then returned to exploratory data analysis and to manipulating and modifying pandas DataFrames. Seaborn was used to demonstrate heatmaps, and Polars was introduced to illustrate the availability of tools designed to handle larger datasets more efficiently.

Following the Python sessions, participants were introduced to AI tools using Google's Gemini and the multi-agent platform You.com [30]. This session demonstrated how AI agents can be used to accelerate research and learning and how to perform basic EDA with Python and pandas within AI tools. Participants also learned about data-privacy considerations associated with AI usage and were cautioned against exposing data or other confidential information when using public AI systems. The sessions concluded with a discussion of the ethical use of AI in learning, teaching, and research, including the obligation to disclose any use of AI tools.

Throughout the sessions, Generative AI was framed as a force multiplier or accelerator rather than a substitute for understanding, with explicit emphasis on validation, judgment, and ethical responsibility.

Survey Results from the "Data Science 101 Foundational Workshops"

To assess and improve the Data Science 101 Workshop, IRB approved pre- and post-surveys were developed. The pre-survey was deployed at the very beginning of the workshop before any workshop content was presented. The post-survey was deployed after all workshop content was delivered. The surveys were administered anonymously to encourage candid self-assessment of skills and confidence. Anonymity also allowed participants to provide open and critical feedback on workshop content and delivery. All twenty-one of the workshop participants (thirteen from 4-year colleges, seven from 2-year colleges, and one from the ADE Career4and Technical Education program) completed both the pre-survey and the post-survey. The data from the two workshops were pooled by combining pre-survey responses across both workshop iterations and post-survey responses across both workshop iterations to provide the information below.

Workshop participants typically brought years of general teaching experience; however, most had limited exposure to teaching data science topics prior to the workshop, as shown in Figures 2 and 3, below. The percentage shown at the top of each bar represents the integer percentage of participants in that experience range; percentages are rounded to whole numbers and therefore may not sum exactly to 100%.

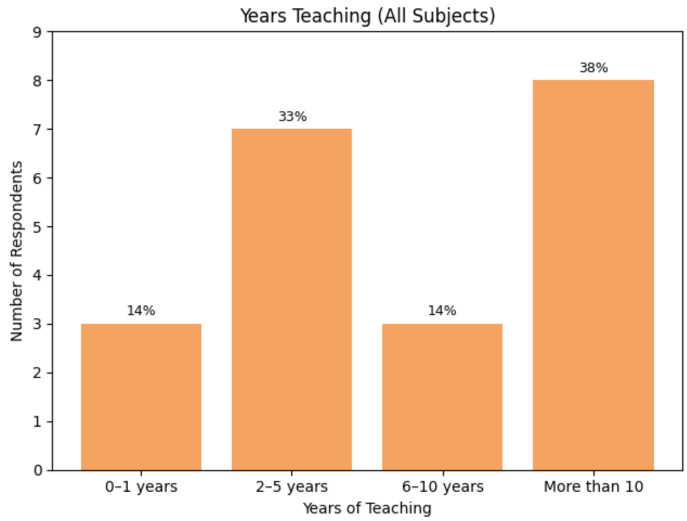

Figure 2. Years of General Teaching Experience.

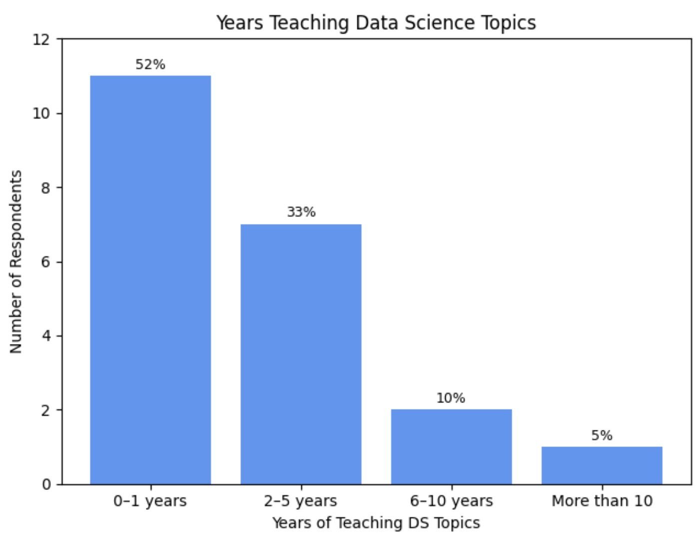

Figure 3. Prior Data Science Teaching Experience.

Despite limited experience at the start, nearly all participants reported feeling skilled in the workshop topical areas after the workshop. Figure 4 shows the number of participants who reported either no or limited experience in each of the main workshop topical areas in the pre-survey. Figure 5 shows the number of participants who reported feeling either somewhat or very skilled in those same topical areas in the post-survey.

While there is a limit to what can be learned in two workshop days, the results indicate that participants developed familiarity with the topics and confidence in their ability to teach the material. This confidence and familiarity with foundational content represent an important first step toward viewing data science instruction as both accessible and engaging.

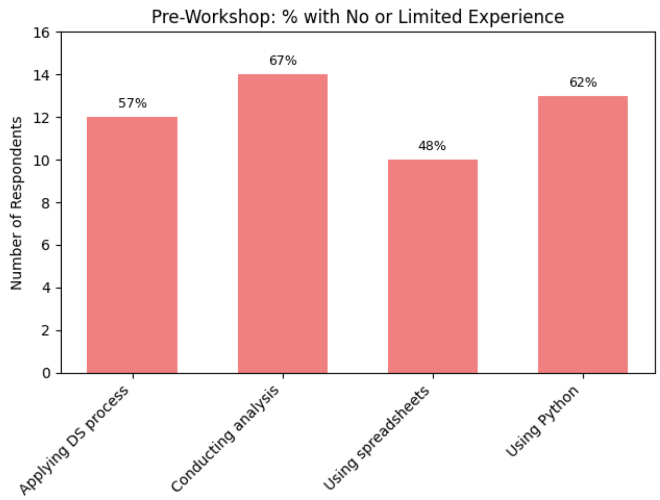

Figure 4. Pre-Survey Experience Levels.

Figure 5. Post-Survey Skill Levels.

Confidence in teaching key data science topics rose sharply, with most participants reporting that they felt somewhat or very confident across all areas by the end of the workshop (see Figure 6). Note that the data reflects responses in the top two categories of "Somewhat Confident" and "Very Confident." The blue bar on the left of each pair shows results from the pre-survey, while the green bar on the right of each pair shows the results of the same question from the post-survey. Because the workshops were held recently, it is too early to determine how increased participant confidence will translate into development of new data science courses, integration of data science content into existing courses, or adoption of the certificate offerings. Additional study will be required over time to assess this impact.

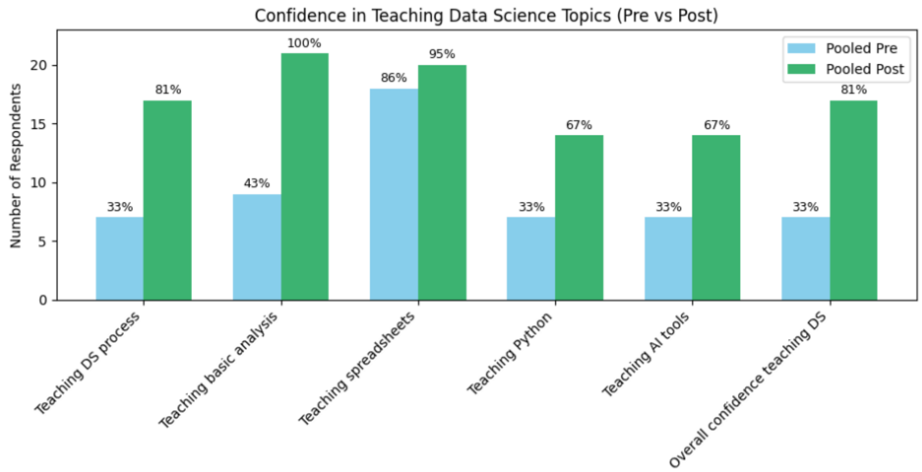

Figure 6. Teaching Confidence Gains.

Relative to overall impressions of the workshop, 100% of participants rated the workshop positively, with 83% rating it as "very valuable." Many noted that the workshop exceeded their expectations in both content and relevance. A majority plan to use the workshop materials provided in their existing courses or use it to begin creating data science content for new courses. Several participants noted plans to collaborate with colleagues or lead curriculum updates.

Participants in the first workshop, which was 1.5 days in length, were asked in the post-survey about the pace of delivery. Most participants indicated that the workshop should be expanded to two full days. In response, the second workshop was extended to two days, the pace was slowed, and additional time was provided to ensure participants could keep up with the hands-on learning. Following the expanded second workshop, almost all participants reported that the pace was just right. While self-reported confidence does not equate to instructional mastery, prior faculty development research indicates that confidence is a necessary precursor to curricular adoption and sustained instructional change. In this context, confidence gains represent an enabling condition for integration rather than a terminal learning outcome.

Data-informed Iterative Improvements to the Workshops

Over the past five years or so, different data science pedagogical workshops have been designed, implemented, and taught [5], [31] with each successive offering being informed and improved by the feedback from the prior workshops. While there were many improvements over the years, these are the most significant points:

1. Teach what data science is, how to use the data science and analytics process, and how to use spreadsheets (still the most common tool used in the workplace), programming, and Generative AI responsibly and ethically – *before* teaching how to teach them!
2. Design the workshop to meet the participants where they are – and even those who are more advanced appreciate the refresher.
3. Provide the tools (pre-workshop directions, laptops, software, libraries, files, IDE, etc.) pre-configured as much as possible, so the participants can focus on the workshop. This

also improves “leaving no one behind” to ensure everyone gets everything they can from the workshop.

4. Bring everything needed when the workshops are regionally on-site. This includes projectors, screens, multi-plugs, extension cords, and coffee maker. Frequently items confirmed as available on-site turn out to be unavailable, not working, or break down in the middle of the workshop.
5. It is reasonable to assign homework between the two days when it is feasible and appropriately scoped. Nonetheless, the demands of the workshop may leave participants with limited energy to complete it.
6. It is important to recognize that what may seem obvious to the facilitator is not necessarily obvious to the participants if they are new to the topic, approach, etc. While the material may be very familiar to the instructor, participants may be encountering it for the first time.
7. Participants truly appreciate the intentional focus on ensuring they leave the workshop more knowledgeable and having met the objectives.

Summary and Conclusions

The development and expansion of the state-wide Data Science Educational Ecosystem represent a significant step toward meeting the state’s workforce development needs. By integrating 2-year colleges into the Ecosystem, creating stackable certificates, and aligning curricula with industry demands, this initiative addresses the growing demand for skilled data science professionals. The “start anywhere, finish anywhere” model where learners can start at any Ecosystem educational institution and finish at any Ecosystem institution without losing credits ensures flexibility for learners, while the emphasis on upskilling and credit for prior learning (CPL) make data science education accessible to a broader audience, including current workers and non-traditional students.

Through the first year of the grant, significant input was provided by the HIRED Advisory Council and the Workshop participants. Workshops have been and will continue to be adapted accordingly. Based on the first “Workshop 101”, significant changes – well received – were implemented for the second one. Most of the changes related to the flow and pacing of the content, ensuring no one was left behind.

The HIRED grant pays for the workshop to be held and provides the $1,500 stipend for workshop participants. Obviously, this cannot be sustained once grant funds are exhausted. Hopefully, after faculty attend workshops, they can provide training of other faculty at their institution and help lead data science curriculum development. All workshop materials are provided to the participants, so it is possible that attendees can use these materials to hold local workshops.

Participants clearly gained confidence in teaching the presented data science fundamentals by the end of the workshop. Confidence is only a first step. Follow-on workshops to be held in 2026 are intended to provide deeper content as well as methods for teaching data science effectively. In these two-day workshops presented in this paper, no assessment was made to determine the participant skill level beyond that of self-assessment. In the follow-on workshop, participants will work in teams at the end of the workshop to perform an EDA and presentation to demonstrate capability.

As the initiative continues, the approach will continue to be refined and assessed. Arkansas is a rural state, and the HIRED grant has limited funding to support a finite number of workshop participants, resulting in a small sample size to date. As the workshops are rolled out to the other areas of the state in 2026 and the follow-on workshops are also held, the number of participants will increase, and more information can be learned about the effectiveness of the workshops. These workshops were just recently held, so it will take time to determine the ultimate impact of the workshop as it relates to the goal of increasing the capacity for quality data science instruction in the state.

Although developed within a specific state ecosystem, the approach described here is adaptable to other states and institutions seeking to expand data science capacity through workforce-aligned faculty development, particularly in environments characterized by diverse faculty backgrounds and constrained technical resources. These successes and lessons learned can serve as a model for other states seeking to build similar education ecosystems and bypass the challenges that were encountered and addressed to accelerate their own implementation.

AI Use Disclosure: Generative AI tools were used only as instructional examples within the workshop activities; all data analysis, interpretation, and conclusions presented in this paper were performed by the authors.

References

[1] Arkansas Department of Higher Education, “HIRED workforce development grants.” [Online]. Available: https://adhe.edu/institutions/workforce-grants. Accessed: Jan. 20, 2026.

[2] National Academies of Sciences, Engineering, and Medicine, Data Science for Undergraduates: Opportunities and Options. Washington, DC, USA: National Academies Press, 2018. doi: 10.17226/25104.

[3] K. D. Schubert and M. D. Rossetti, “Creating a multi-college interdisciplinary B.S. data science program with concentrations,” in Proc. ASEE Virtual Annual Conf., Virtual Conference, Jul. 2021.

[4] K. D. Schubert, L. Shoultz, and S. Romer, “Expanding, improving, and completing a multi-college interdisciplinary B.S. data science program with concentrations,” in Proc. 2023 ASEE Annual Conf. & Exposition, Baltimore, MD, USA, Jun. 2023.

[5] K. D. Schubert, S. Romer, S. R. Addison, T. D. Moore, L. J. Berry, J. M. Fowler, L. Shoultz, and C. C. Davis, “Envisioning and realizing a statewide data science ecosystem,” in Proc. 2024 ASEE Annual Conf. & Exposition, Portland, OR, USA, Jun. 2024.

[6] K. D. Schubert, C. S. Gattis, S. R. Addison, T. J. Dryer, and A. Musto, “Expanding a state-wide data science educational ecosystem to meet workforce development needs,” in Proc. 2025 ASEE Annual Conf. & Exposition, Montreal, QC, Canada, Jun. 2025, doi: 10.18260/1-2--56474.

[7] ABET, “Draft accreditation criteria for Data Science programs.” [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/draft-criteria-for-data-science-programs/. Accessed: Jan. 20, 2026.

[8] U.S. Bureau of Labor Statistics, “Data scientists,” Occupational Outlook Handbook, Washington, DC, USA, 2024. [Online]. Available: https://www.bls.gov/ooh/math/data-scientists.htm. Accessed: Jan. 20, 2026.

[9] Data Science Dojo, “Titanic dataset.” [Online]. Available: https://github.com/datasciencedojo/datasets/blob/master/titanic.csv. Accessed: Jan. 15, 2026.

[10] The Metropolitan Museum of Art, “The Metropolitan Museum of Art Open Access dataset.” [Online]. Available: https://github.com/metmuseum/openaccess. Accessed: Jan. 15, 2026.

[11] P. C. Brown, H. L. Roediger III, and M. A. McDaniel, Make It Stick: The Science of Successful Learning. Cambridge, MA, USA: Belknap Press of Harvard University Press, 2014.

[12] P. K. Agarwal and P. M. Bain, Powerful Teaching. San Francisco, CA, USA: Jossey-Bass, 2019.

[13] Google, "Google Colab." [Online]. Available: https://colab.research.google.com/. Accessed: Jan. 15, 2026.

[14] Google, "Google Slides." [Online]. Available: https://workspace.google.com/products/slides/. Accessed: Jan. 15, 2026.

[15] Python Software Foundation, "Python." [Online]. Available: https://www.python.org/. Accessed: Jan. 15, 2026.

[16] Project Jupyter, "Project Jupyter." [Online]. Available: https://jupyter.org/. Accessed: Jan. 15, 2026.

[17] Anaconda, Inc., "Anaconda Python." [Online]. Available: https://www.anaconda.com/. Accessed: Jan. 15, 2026.

[18] L. Vaughan, Python Tools for Scientists. San Francisco, CA, USA: No Starch Press, 2023.

[19] F. Nelli, Python Data Analytics, 3rd ed. Rome, Italy: Apress, 2023.

[20] N. George, Practical Data Science with Python. Birmingham, U.K.: Packt, 2023.

[21] M. Walker, Python Data Cleaning. Birmingham, U.K.: Packt, 2020.

[22] W. McKinney, Python for Data Analysis, 3rd ed. Boston, MA, USA: O'Reilly Media, 2022.

[23] NumPy Developers, "NumPy." [Online]. Available: https://numpy.org/. Accessed: Jan. 15, 2026.

[24] SciPy Developers, "SciPy." [Online]. Available: https://scipy.org/. Accessed: Jan. 15, 2026.

[25] Matplotlib Development Team, "Matplotlib." [Online]. Available: https://matplotlib.org/. Accessed: Jan. 15, 2026.

[26] Pandas Development Team, "pandas." [Online]. Available: https://pandas.pydata.org/. Accessed: Jan. 15, 2026.

[27] Seaborn Development Team, "Seaborn." [Online]. Available: https://seaborn.pydata.org/. Accessed: Jan. 15, 2026.

[28] Polars Developers, "Polars." [Online]. Available: https://pola.rs/. Accessed: Jan. 15, 2026.

[29] Google, "Google Gemini." [Online]. Available: https://gemini.google.com/app. Accessed: Jan. 15, 2026.

[30] You.com, “You.com.” [Online]. Available: https://you.com/. Accessed: Jan. 15, 2026.

[31] DART Project, “Education – Data Analytics that are Robust and Trusted (DART).” [Online]. Available: https://dartproject.org/education/. Accessed: Jan. 20, 2026.