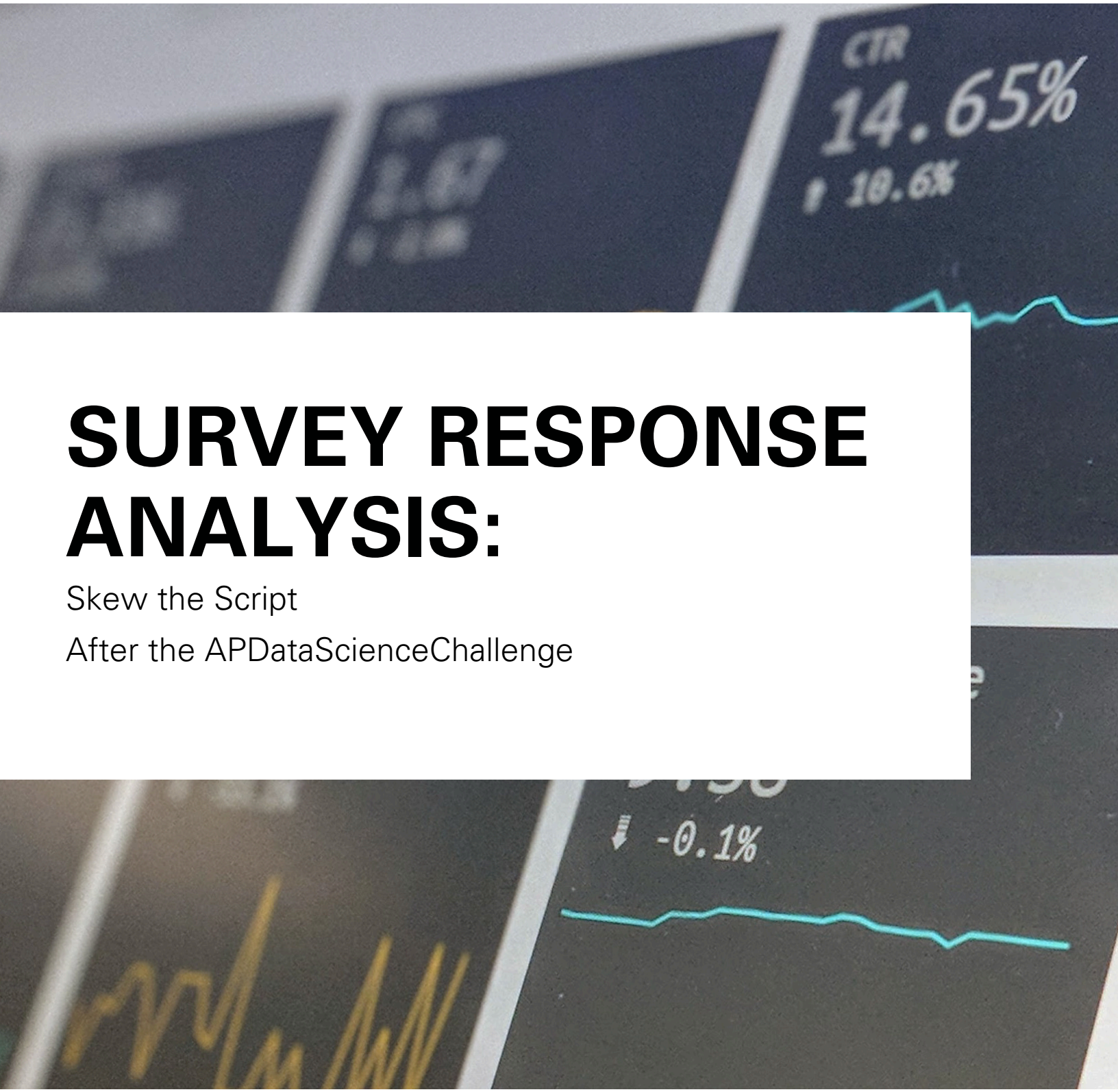
NC STATE UNIVERSITY

Data Science and AI Academy



SURVEY RESPONSE ANALYSIS:

Skew the Script

After the APDataScienceChallenge

Taryn Shelton, K-12 Programs Coordinator Data Science and AI Academy and The Science House

At NC State, Data Science is for Everyone.

Learn more at datascienceacademy.ncsu.edu

Background

Skewthe Script(skewthescript.org), is an educational platform that provides free and socially relevant math lessons designed to engage students in mathematics in contextually authentic ways. In an effort to extend the reach of their mathematics content into the area of data science, Skew the Script has partnered with Data Science for Everyone, a coalition focused on advancing K-12 data science education, and CourseKata, which provides interactive online textbooks for teaching introductory statistics and data science. Through this collaboration, the **After the AP Data Science Challenge** was created and launched to present students with raw data and guidance in building their own models while learning some low-level coding skills.

The **After the AP Data Science Challenge** is a four-module data science challenge, intended for students who have completed AP Statistics. In Spring 2023, classrooms across the US piloted the challenge with their students, particularly those who took the AP Stats exam and had time to explore data science concepts. These modules were delivered through Canvas and consisted of Jupyter notebooks as provided by the CourseKata Jupyter hub. The notebooks walked students through an introduction to data science concepts and the coding language R in order to conduct data explorations and visualizations. The data presented were about college debt sourced from the U.S. Department of Education College Scorecard (https://collegescorecard.ed.gov/data/). The target areas of the four notebooks were

- **Module 1:** Learn basic commands in the R coding language; Calculate summary statistics from a large data frame; Create various types of data visualizations.
- **Module 2:** Create scatterplots; Fit simple linear regression models; Compare the strength of predictor variables.
- **Module 3:** Fit multiple regression models; Interpret the coefficients of multiple regression models; Compare the strength of different multiple regression models.
- **Module 4:** Fit polynomial models on a subset of the data; Tune models based on the principles of underfitting and overfitting; Use machine learning principles to compare the prediction strength of different models.

Serving as an external evaluator, the Data Science Academy at NC State University created a post-survey to be completed by educators who piloted the **After the AP Data Science Challenge** with their students in April and May 2023 and then analyzed their responses. This report summarizes the results from that analysis and presents recommendations for future improvements to the lesson. The goal of this report is to guide the revision of the modules based on teacher feedback and does not include direct feedback from students or an assessment of student learning outcomes.

It is important to note that since the post-pilot survey was voluntary, some educators with the least desirable experiences may not have been motivated to complete the challenge or to participate in the survey. In addition, we do not have feedback from teachers who opted not to participate. That said, the response from teachers was overwhelmingly positive.

Findings

SkewtheScript reported that over 2,600 high school students participated in the **After the AP Data Science Challenge**, completing at least one Notebook. We received feedback from 53 of the 89 educators who registered to pilot the project with students. This feedback encompasses the experiences of those educators and their 1,500 students. The majority of these students were 11th and 12th graders in public schools throughout the U.S., with 21% of the public schools having Title I status. (Table 1, Figure 1) Of the classes that piloted the lesson, most were AP Statistics plus several Computer Science classes. It is interesting to note that while this challenge was specifically marketed toward AP Statistics and Computer Science classes, there were a couple of Calculus classes that also participated. (Table 1)

School Type

Private	Charter	Public	Public with Title I status
10	2	41	11
19%	4%	77%	21%

Grades

9th	10th	11th	12th
3	19	48	42
3%	17%	43%	38%

Classes

AP Stats	Stats / Probability	Computer Science	Calculus
44	5	10	2

Table 1. School Type (Private, Charter, Public) of the 53 post-pilot survey respondents, 21% of all the participating schools were public schools with Title I status. Educators at each school piloted the challenge with multiple classes and grade levels. Note that some teachers used the materials in more than one course.

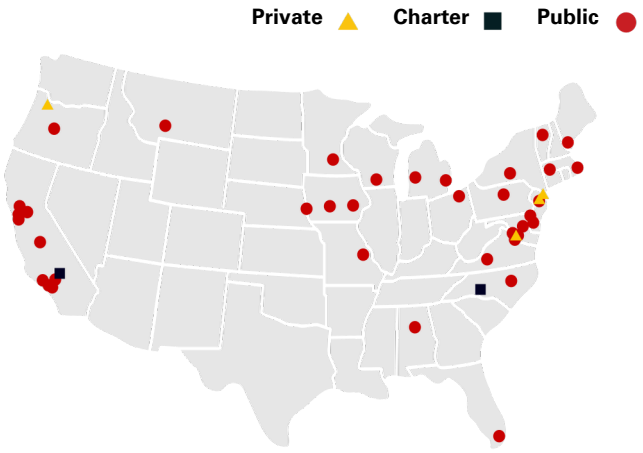


Figure 1. Map showing the city and state locations of the 53 post-pilot respondents and the types of educational institution.

When asked as an open response question why they were interested in participating in the **After the AP Data Science Challenge**, most respondents described wanting to provide their students with exposure to real-world data and data science skills while experiencing a practical application of their statistical skills. Some also expressed trust in the Skew the Script platform and felt that it was a good use of time after the AP Statistics exam with a topic that was relevant to their students. Several of the respondents also indicated that their schools may be considering or have plans to implement a data science course in the near future.

> **"I wanted students to experience a more authentic project-based investigation and apply their knowledge of statistics in a real setting."**

As noted above, the **After the AP Data Science Challenge** consisted of four total Jupyter notebooks, which were provided by the CourseKata Jupyter hub. Among all students who took part in the challenge (not just the students of survey respondents), 82% completed at least three of the four notebooks, and 76% completed all four Notebooks (Figure 2). Survey respondents reported giving their students an average of 8-10 class periods (50-minutes each) to complete the challenge, while also reporting that they needed an estimated 6 to 8 class periods to complete all four notebooks (although some expressed needing as few as 3 classes or as many as 20). Each notebook took students an average of 2 hours to complete, though most respondents reported students needing one hour to complete one notebook.

Notebooks Completed
% of all student participants (not just survey respondents)

Figure 2. Among all students who took part in the challenge, 87% completed Notebooks 1-2. Notebooks 3 and 4 covered completely new academic material for all students (multiple regression and machine learning). 82% of students completed Notebooks 1-3 and 76% completed Notebooks 1-4.

Educator Experience

The **After the AP Data Science Challenge**was designed to be accessible to both students and educators with little to no prior coding or computer programming experience. Of those surveyed, 51% expressed that they had little to no experience with R or computer programming languages and 84% had little to no prior experience with Jupyter notebooks. (Figure 3, Table 2) Yet despite their degree of experience, most educators (83%) found the level of coding in the lesson to be of low difficulty. (Figure 3)

Instructor's Prior Coding Experience

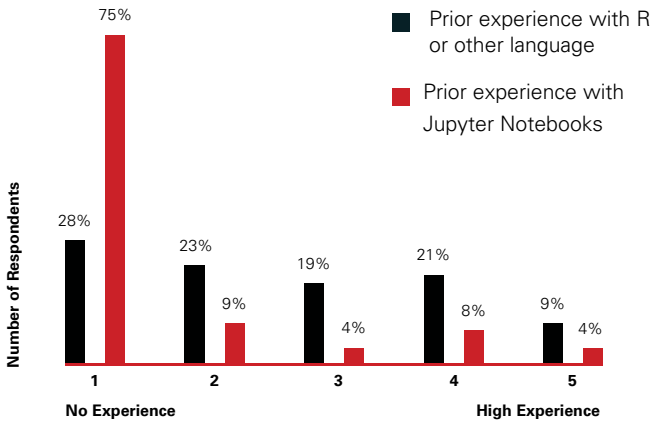


Figure 3. While there was a range of instructor experience with R or other programming languages among respondents, most (75%) reported no prior experience with Jupyter Notebooks.

Level of Coding Difficulty for Instructors

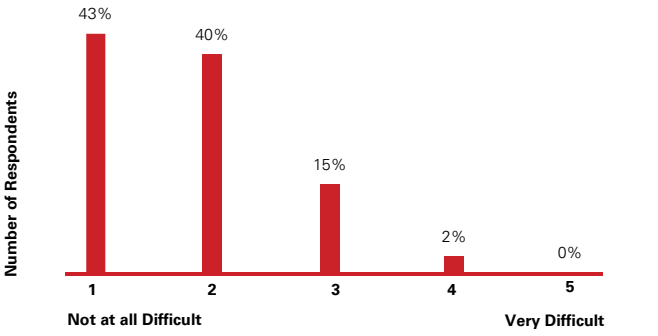


Figure 4. Most respondents (83%) reported low difficulty in coding within the **After the AP Data Science Challenge**.

Student Experience

TheAfter the AP Data Science Challenge allowed students to apply their statistical skills to a real dataset in a context relevant to them in order to increase engagement. The goal was to design content that feels authentic and meaningful with tasks that are just difficult enough to keep students interested, without being so difficult that they disengage. Survey responses indicated that students were engaged in varying degrees with the notebooks. When given a scale between 1 and 5, where 1 is very disengaged and 5 is very engaged, 68% of respondents rated student engagement between 4 and 5. (Figure 5)

Likewise, when asked about the level of coding difficulty for students on a scale between 1 and 5, where 1 is too easy and 5 is too difficult, 98% responded that the level of difficulty for their students was between too difficult and too easy, which is ideal for instructional material. Only one teacher reported that the level of coding was too easy, while none of the respondents considered the coding to be too difficult for their students. (Figure 6)

Students' Level of Engagement while completing the Notebooks

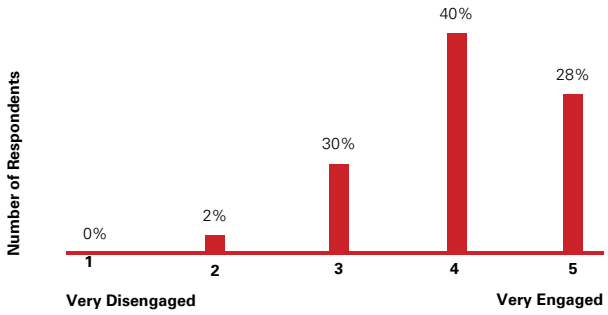


Figure 5. Most respondents (68%) reported that their students were to some degree engaged with the **After the AP Data Science Challenge**.

Level of Coding Difficulty for Students

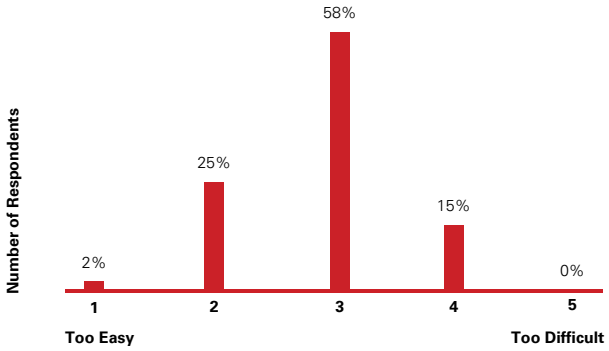


Figure 6. The large majority of respondents (98%) indicated that the level of coding difficulty for their students was between too difficult and too easy. No respondents reported the coding as too difficult for their students.

The level of coding difficulty is not the only factor in maintaining student engagement. Providing students with data in a context that is relevant to them is also critical. In the **After the AP Data Science Challenge**, students were presented with a dataset about college debt. Overwhelmingly, educators responded that this context was relevant to their students. (Figure 7)

Context (College Debt) Relevant to Students

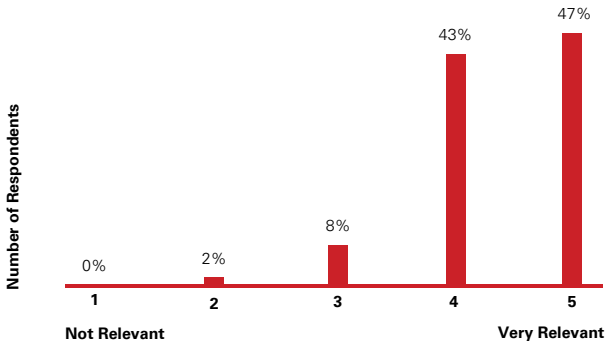
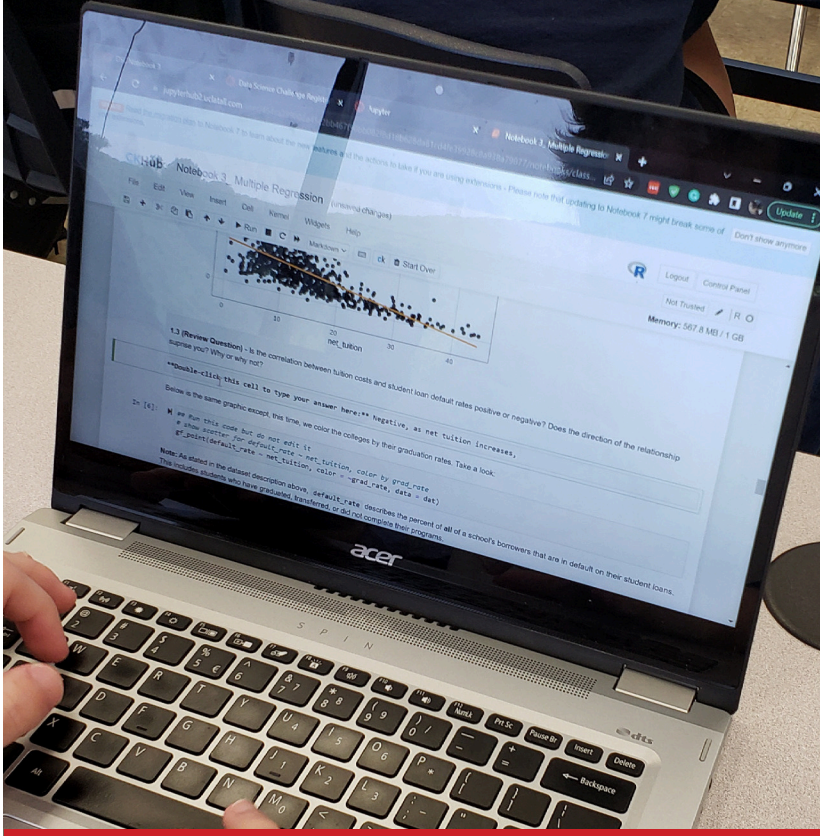


Figure 7. The majority of respondents indicated that the context of College Debt was relevant to their students.

Learning Supports and Ease of Use

Educators were asked how easy the lesson platform was to navigate with their students, and most agreed that both the Canvas platform and the Jupyter Notebook were clear and comfortable to use. The majority of respondents also agreed that the instructional videos were helpful and that the level of learning support provided was sufficient. (Figure 8) There was consistent feedback that indicated a need for the program to instruct users that reloading the CSV file and re-running the code may be necessary. Several respondents also commented on the Canvas platform, describing issues with logging in, as well as a desire to integrate the program into their current school account. Another common suggestion was to provide some additional resources, such as a list of commands or codes, and explanations that would help students self-diagnose and fix errors.



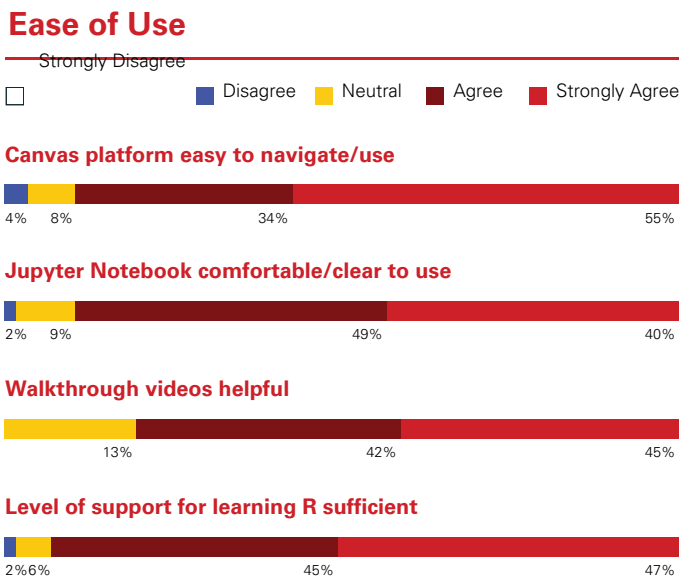


Figure 8. An overwhelming majority of respondents agreed that the Data Science Challenge platform was easy to use and navigate with sufficient support and helpful videos. Only 2-4% of respondents disagreed on the platform's ease of use.

Final Thoughts and Recommendations

Overall, feedback from educators was positive. Generally speaking, most found the **After the AP Data Science Challenge** to be both engaging and relevant to their students, with just enough difficulty to keep them invested in the tasks. The platform was mostly easy to use whether or not educators or students had any prior coding experience, and the supporting material was helpful. However, the sentiments among students were much more varied. While many educators reported that their students enjoyed the challenge and found the content interesting, others relayed that students were overwhelmed, confused, and frustrated. For some, their frustrations subsided once they overcame the learning curve and got the hang of coding. Educators appreciated most that the content was relevant and that their students were able to apply their statistical skills to real data in an authentic way.

"They enjoyed the context and were invigorated by the "realness" of the data and the question they were considering."

It is common to have a wide range in student perceptions of classroom work, especially at the end of the year when student apathy is at its highest. Some suggestions that may help with this divergence are to better scaffold the lessons by offering coding and troubleshooting guides for those new to programming language, and more opportunities for deeper investigations for those with prior coding experience. One teacher even suggested having an "open data dump" once students learn some R skills, and then the option for an additional open-ended challenge (like "pick your own outcome variable and build a model to predict it.") Additionally, for those who struggled more, offering more frequent and shorter video clips may be helpful, as some students were not interested in the longer videos and skipped them. Many educators also requested that the lessons be integrated with their own classroom Canvas accounts and that they provided a more efficient means of grading student work, as well as a better way to monitor student progress in real time. The **After the AP Data Science Challenge** pilot was a success and educators are enthusiastically awaiting the opportunity to offer it in their classes again. 85% of respondents said that they would use this project again, while only 15% were undecided. Zero respondents indicated they wouldn't consider using the challenge in future years. We recommend revisions include better accommodation for variance in coding abilities, the addition of guidance sheets, and options for deeper open- ended data dives. With these adjustments, we anticipate that this program will continue to be well-received and implemented in Statistics, Computer Science, and Calculus to increase high school students' exposure to the world of data science.

"I found that the vast majority were engaged by the context and enjoyed learning both technology skills and data science concepts. There was a wide range of prior experience and comfort levels in my classes, but all the students seemed interested and found value in the project."

Instructor Experience

	No Experience			High Experience	
	1	**2**	**3**	**4**	**5**
Prior experience with R or other languages	15	12	10	11	5
	28%	23%	19%	21%	9%
Prior experience with Jupyter Notebooks	40	5	2	4	2
	75%	9%	4%	8%	4%

	Not at all Difficult			Very Difficult	
	1	**2**	**3**	**4**	**5**
Level of Coding Difficulty	23	21	8	1	0
	43%	40%	15%	2%	0%

Student Experience

	Very Disengaged			Very Engaged	
	1	**2**	**3**	**4**	**5**
Students' level of engagement while completing the Notebooks	0	1	16	21	15
	0%	2%	30%	40%	28%

	Much too Easy			Much too Difficult	
	1	**2**	**3**	**4**	**5**
Students' level of difficulty coding	1	13	31	8	0
	2%	25%	58%	15%	0%

	Not useful			Very Useful	
	1	**2**	**3**	**4**	**5**
Data science skills covered are useful for students	0	0	3	12	38
	0%	0%	6%	23%	72%
					94%

	Not Relevant			Very Relevant	
	1	**2**	**3**	**4**	**5**
Context (college debt) relevant to students	0	1	4	23	25
	0%	2%	8%	43%	47%
					91%

NC STATE UNIVERSITY

Data Science and AI Academy

Data Science 4 everyone

Skew The Script

CourseKata