Learning Python and Data Science In Berkeley Lab Director's Apprenticeship Program

Brandon Abrego[1], Alisa Bettale[2], Laurel Hales[2], Faith Dukes[2]

[1]UC Berkeley, [2]Lawrence Berkeley National Laboratory

ABSTRACT

This summer, I worked with the K-12 Team at Lawrence Berkeley National Lab where I assisted in helping teach Python to local high school students. I worked with Alisa Bettale and Laurel Hales as instructors and Faith Dukes as head of the team. The students took a survey both before the program and after the program to determine what coding skills they had coming into the program, and what skills they had leaving the program. I analyzed this data using Google sheets to determine what concepts students understood and what concepts students were still struggling with.

BACKGROUND INFO

BLDAP (Berkeley Lab Director's Apprenticeship Program) is a STEM integration program designed for local high school students (students from Alameda, San Francisco, and West Contra Costa counties) from underrepresented backgrounds in STEM to get exposure and hands-on experience with different science fields. All of the students applied for the program and completed a round of interviews, and are paid a stipend as a part of the program. Students participated in an introductory Python/data science course, as well as worked on various projects from an energy design challenge to Arduinos. Students applied their Python/data science skills to research sets provided from Berkeley Lab scientists.

I primarily focused on the Python/data science aspect of the program, creating content to present and teach via direct instruction to students and developing assessments. 23 students are coming from various local high schools and 6 have no prior experience, 13 have some experience, and 2 identified as intermediate with coding. Their final Python lesson was applying data science to the field of astrophysics. Students learned throughout the program that virtually all fields of science require a significant amount of coding, making understanding coding essential for a large variety of STEM fields. For educators is is important to know which concepts students find more difficult to understand than others.

RESEARCH QUESTION

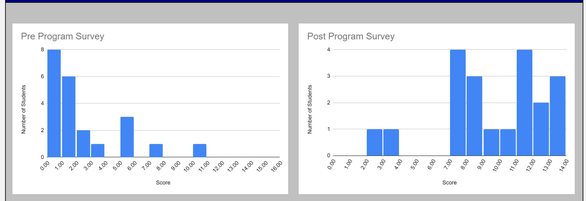
What concepts do students find difficult and what concepts are students understanding with Python/data science?

RESEARCH METHODS

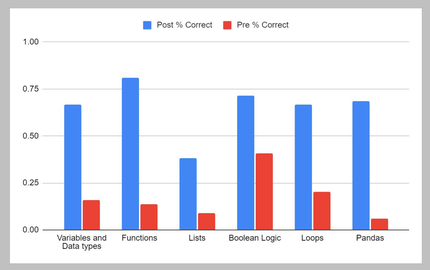
Data comes from surveys assigned to students at the beginning of the program and at the end of the program. Surveys consisted of multiple choice programing questions, with students having the option to say "I am not familiar with this concept" which will count as incorrect. This is done to remove correct responses that result from guessing. I will be comparing the results of the Pre Program survey and the Post Program Survey to analyze overall learning and effectiveness of the program. I will then divide the survey into question categories, with each category representing a different concept. Note that the Pre Program survey has 16 questions while the Post Program Survey has 14 questions. 2 questions were removed from the post survey because they were on dictionaries, which we did not get to cover by the end of the program. For the purposes of analysis I will refer to the questions by the number they appeared in the Pre Program Survey. Students also anonymously indicated how they felt about the lesson for the day on a poster board, Blue feeling most confident and red feeling completely lost. Week 3 students were online and not doing Python work.

Catagory	Question Numbers
Variables and Data types	1-2
Functions	3
Lists	5
Boolean Logic	7
Loops	8-9
Pandas	10-16

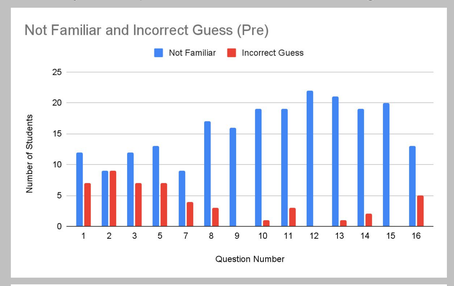
RESULTS

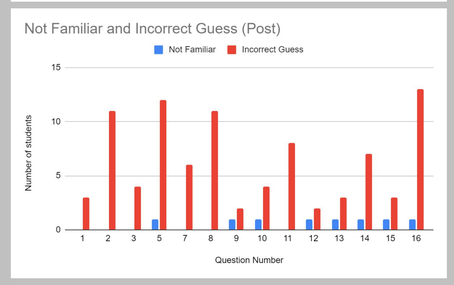


There is a clear increase in the number of correct responses before and after the program, indicating that BLDAP is generally increasing student's competency in coding. The average score increased from **2.05** to **9.38** out of 14.



Every question category showed some level of improvement. With the greatest improvement in Pandas and lists showing substantially less improvement than the other categories.



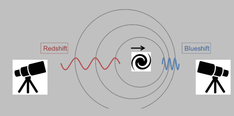


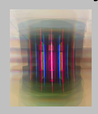
Students are much less likely to say they are not familiar with a concept after the program, which does show improvement, even in the case of lists.

Students felt the least confident at the beginning of week 1 when they first started coding with a majority feeling they need more time. Lists took place on Wednesday and Thursday of Week 2, showing no noticeable dip in confidence. Pandas took place during week 4, which had some select students showing less confidence than normal.

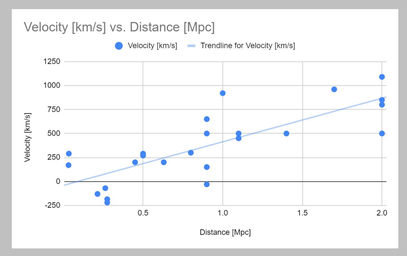
LESSON PLAN

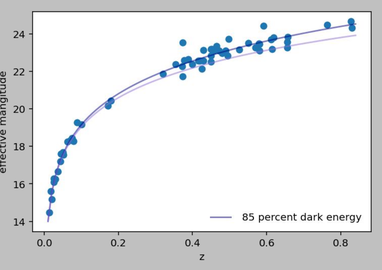
The students utilized programming and data science concepts they learned throughout the program to engage with the astrophysics lesson based on Jupyter Notebooks from DESI High. This lesson was taught over two days and covers spectroscopy and redshift, relating it to how astronomers determine velocity.





The lesson culminated with students discovering two important facts; the universe is expanding and the expansion is accelerating due to dark energy. Students created graphs from real data and used these visualizations to discover and prove these phenomenon.





Students utilized lists, functions and pandas to generate their plots. A final challenge was given for them to generate the lowest Chi squared value to have the most accurate fit for the amount of dark energy in the universe.

References

Hubble E. 1929 *Proc. Natl. Acad. Sci.* **15** 168

S. Perlmutter *et al.* 1999 *ApJ* **517** 565

Lesson based off DESI High