

# 2022–23 Data4All Bridge Workshops

## The Impact in Students' Own Words

*Made possible with funding from the Hymen Milgrom Supporting Organization and the University of Chicago's Office of Civic Engagement and Data Science Institute*

Prepared by Julia Koschinsky, PhD  
Center for Spatial Data Science  
[spatial@uchicago.edu](mailto:spatial@uchicago.edu)



THE UNIVERSITY OF CHICAGO  
DATA SCIENCE  
INSTITUTE

Argonne  
NATIONAL LABORATORY

THE UNIVERSITY OF  
CHICAGO | THE CENTER FOR  
SPATIAL  
DATA  
SCIENCE

THE UNIVERSITY OF CHICAGO  
Civic Engagement

# Foreword

by Damaris Hernandez,  
MEd, Peer Reviewer of  
Data4All Teaching  
Materials

## Why We're Advancing Diversity, Equity & Inclusion When Teaching Data Science in a Scientific Reasoning Framework

As a first generation, low-income, woman of color, I have experienced the educational disparities that exist in the public education system firsthand. Throughout my K12 education I oftentimes found it challenging to engage in my educational spaces and learning content. Preoccupied with what was going on outside of the classroom and aware that space for these concerns were not being made inside the classroom, I felt disconnected from my education. Space for meaningful learning did not exist, and math and science were often too fixated on rote learning, in contrast to focusing on the learning process and scientific reasoning.

It was not until my senior year of high school when my physics and calculus teacher, Mr. Eastvedt, attempted to reimagine the way we had been taught for the past 11 years and make learning student centered and exploratory. Similar to Data4All's *Data Science Reasoning Framework*, Mr. Eastvedt presented us with the opportunity to explore and get our minds thinking about a specific real world problem or phenomena that we would take into our own hands and try to explain, test, and analyze findings for.

This type of learning initially brought along many mixed feelings and emotions, one of them being as if I was not capable of understanding because I had never been in spaces such as these and did not see myself, a brown girl, being part of learning in this manner. I felt like I did not belong. I continued to see this type of learning and teaching when I began my undergraduate mathematics education at UC Berkeley and immediately felt scared again because this way of thinking and learning was still new to me. Again, I felt like I did not belong.



# Foreword

by Damaris Hernandez,  
MEd, Peer Reviewer of  
Data4All Teaching  
Materials

As I gained more experience in the public primary education system through my student teaching at UC Berkeley, I realized that my sentiments were paralleled by that of many first-generation students under my tutelage at low-income K12 Bay Area schools. While volunteering at Richmond High School, I worked primarily with students who had recently migrated to the U.S. from Latin American countries. With a language barrier and a curriculum that was not tailored to the unique cultural, social, and economic challenges these students were facing, I witnessed the pivotal impact of redesigning learning environments and creating equitable access to curriculum that exposes students from nondominant communities to programming, statistics, and mapping in the context of scientific reasoning and real-world problem-solving with data.

Other experiences I participated in proved how empowering and valuable it is for students to explore hands-on learning and know how data is in everything we do and how we can work with it to do better and make an impact. Data4All's vision of teaching students how to solve real world health problems with a scientific reasoning framework targets this educational problem, attempts to bridge students' learning and helps them build a problem solving mindset that they are capable of.

As a math educator now, my goal is to rehumanize math education, so students can view education as more than just black and white spaces, and empower them through math and data to make a difference in their communities. Shifting teaching and learning early in the classroom to be taught with the capacity to critically solve math and science problems would allow students from nondominant communities to recognize that they are scientists and mathematicians and have a sense of belonging in spaces that impacts them and they can make an impact in. The integration of data and scientific reasoning have the power of making that happen for students from nondominant communities.



# Table of Contents



<u>About this Report</u> .....	5
<u>Data4All Impacts</u> .....	7
<u>Best Practices</u> .....	8
<u>Challenges</u> .....	9
<b><u>ABOUT DATA4ALL</u></b> .....	10
○ <u>Data4All in Numbers</u> .....	11
○ <u>Why this Workshop</u> .....	12
○ <u>Data4All's Approach</u> .....	13
○ <u>Reasoning Framework</u> .....	14
○ <u>Workshop Activities</u> .....	15
○ <u>Learning Objectives</u> .....	17
<b><u>EVALUATION</u></b> .....	18
1. <u>Target Audience</u> .....	19
2. <u>Learning Outcomes</u> .....	26
3. <u>Program Evaluation</u> .....	36
○ <u>Small Groups &amp; Mentors</u> .....	37
○ <u>Stats &amp; Coding Exercises</u> .....	39
○ <u>Real Cases &amp; Mapping</u> .....	42
○ <u>Spark Activities</u> .....	45
○ <u>Guest Speakers</u> .....	48
<b><u>BACKGROUND</u></b> .....	51
○ <u>About the Hosts</u> .....	52
○ <u>About the Surveys</u> .....	53
○ <u>Open Teaching Materials</u> .....	54
○ <u>Learning Objective Details</u> .....	55
○ <u>In the News</u> .....	57
○ <u>Acknowledgements</u> .....	59

# About this Report

**The goal of the Data4All Bridge workshops** (Data4All) has been to teach high school students how to use computation, statistics, and mapping to address real-world data problems and puzzles. This focus on reasoning with data seeks to broaden students' understanding of what data science (DS) is, beyond a more narrow technical focus on programming and statistics. It is also intended to highlight the relevance of DS to a broad variety of science, technology, engineering, and mathematics (STEM) and other fields, including college and career options students did not associate with DS before.

**The target audience** for the workshops are sophomores, juniors, and seniors from Chicago Public Schools who are underrepresented in data science in terms of race/ethnicity or gender and who are less exposed to computer science (CS) curricula at their schools. Data4All considered trauma-informed education frameworks and provided technological and financial support to students. For instance, mentors, instructors, and curriculum developers received several hours of sensitivity training regarding trauma-informed education (co-led by the Office of Civic Engagement and Tyler Skluzacek). The technical environment was geared to students without their own laptops (Jupyter Hub, Chromebook loaners, no data or software downloads) and scholarships for all students (\$500) offset some of the costs for travel and potential lost earned income.

So far, **we held Data4All three times:** The first workshop was a 1-week virtual training during COVID in spring of 2021, followed by in-person workshops on eight Saturdays for 4 hours in fall 2022 and spring 2023. Data4All was hosted at the University of Chicago (UChicago)'s Data Science Institute (DSI), in collaboration with Argonne National Labs (ANL), the Center for Spatial Data Science (CSDS), and the Office of Civic Engagement (OCE). OCE and ANL recruited students, ANL instructors taught the workshop, and DSI recruited mentors and managed Data4All. The curriculum was co-developed by ANL, CSDS, and Tyler Skluzacek (Oak Ridge National Labs).



# About this Report

This is **an internal evaluation** of the last two workshops held in fall 2022 (with 24 students) and in spring 2023 (with 26 students).

It highlights Data4All's impact in the students' own words. We completed this internal evaluation for several reasons: To provide an overview of our efforts so far, summarize how the workshop works, and assess if its learning objectives were reached or not. The audience we had in mind for this report include instructors and mentors who will teach the workshop in the future (at UChicago and elsewhere), funders to whom we are accountable, and other stakeholders interested in teaching data science to high schoolers.

The **evaluation is based on student surveys** that we conducted before, in the middle, and at the end of the workshops in 2022 and 2023 (more details [here](#)). It also draws on feedback from mentors and instructors. Since not all students completed the mid-point and post surveys, we often report 2022 mid-point results (which had more respondents) and cannot always directly compare survey responses between 2022 and 2023 since the 2023 survey was updated to reflect the [logic model](#) that we developed after the fall 2022 workshop.

**Three workshop aspects are evaluated** in this report: 1) [Target audience](#), 2) short-term [learning objectives](#), and 3) [workshop activities and staff](#). The report starts with an [overview of Data4All's rationale and approach](#), followed by the above [evaluation](#) based on student surveys, and ends with a [background](#) section with further details about the hosts, surveys, and logic model.

All workshop materials will be [publicly released](#) after peer review in fall 2023. Data4All will continue to be taught in fall 2023 and spring 2024 at the UChicago's Data Science Institute.

This workshop was made possible by **Milgrom's Computer Coding Fellowship Program Grant**, which provided stipends to students and mentors, and **OCE's** and **DSI's** operating support. We are also grateful for the invaluable contributions of everyone who worked on Data4All, [acknowledged at the end of the report](#).



# Data4All Impacts at a Glance

The results of this Data4All evaluation at a glance:

## 1. Student Diversity & DS Exposure

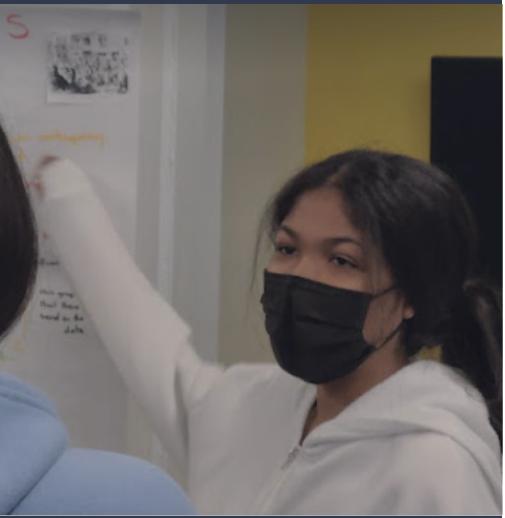
- 90-96% of Data4All Students were underrepresented in STEM ([NSF definition](#)) or girls (2022 & 2023)
- Students came from 18 schools across Chicagoland in 2022 and 2023, with a majority being public schools and half Southside high schools
- Prior to Data4All (2022), almost all students (86%) were exposed to scientific concepts, and most had some programming experience (69%) but fewer (32%) had data analysis & visualization skills
- Prior to Data4All (2023), almost all students (85%) were comfortable with evidence-based arguments, but two thirds did not feel comfortable programming (65%) and over a third (38%) were not comfortable with statistics

## 2. Learning Outcomes

- Before Data4All, students saw data science (DS) more narrowly as programming and statistics. Afterwards, they saw it in a larger context of scientific reasoning.
- Integration of scientific reasoning broadened appeal of CS & stats for students who might not major in CS
- Through the reasoning framework and speakers, students realized broader connection of CS & statistics to STEM
- Programmers added skills in reasoning with data while debaters added programming skills
- Students without prior experience in data analysis or data visualization reported increased confidence in both in 2022
- In 2023, Data4All increased 81% of students' confidence in programming and 57% in statistics
- Students found Data4All useful for clarifying data science-related college & career goals

## 3. Program Evaluation

- Students LOVED their small groups & mentors
- Students really liked the programming exercises ... even though they also found the work with Jupyter notebooks the most challenging
- The historical cholera and modern covid cases & mapping were interesting, with slight preference for the modern covid case
- A frequent workshop highlight was tracing why & how cholera spread
- Students found lectures & spark activities (very) useful
- Guest speakers connected data science to STEM, college & career options, which students found (very) useful



# Best Practices from 2022-23 Workshops



Based on this evaluation of the Data4All 2022 and 2023 workshops, the following best practices emerged:

## **Scientific reasoning framework appealed to debaters and programmers**

The workshop's framework of how to reason and solve real problems with statistical evidence made data science appealing for two groups that are usually hard to teach in the same class: Students who preferred debate (they learned to program) and students who preferred to program (who learned to reason with data).

## **Problem-driven approach to programming made it more accessible to non-programmers**

Starting with a real-world puzzle to solve made data science more accessible and relevant for students who did not intend to major in computer science – compared to traditional data science bootcamps that only code.

## **Small groups and near-peer mentors were key to Data4All's success**

Working in small groups of 4-6 students with near-peer mentors was essential for grasping content at students' own pace and debugging their code. Jupyter notebooks worked well for integrating computation, statistics, and mapping with scientific reasoning.

## **Mix of activities kept students engaged for 4 hours on Saturdays**

The mix of activities – exercises to practice programming, statistics, and mapping, spark activities, lectures, ice breakers, and lunch speakers – kept students engaged over 4 hours on a weekend and appealed to different types of learners.

## **Guest Speakers helped students see the broader relevance of data science to different domains & careers**

Guest speakers were effective at introducing students to a larger world of data science in colleges and careers. Students walked away with a greater sense of the large range of domains that data science is relevant to and that match their own broader interests.

# Challenges

Beyond best practices, the 2022 and 2023 workshops also revealed several implementation challenges:

## Instructors & Mentors

- Data4All requires more prerequisite knowledge for instructors and mentors than when programming, statistics and science are taught separately since it combines these three domains. If instructors do not have all three skills, the key skill to be comfortable with is critical reasoning with data and statistics. If they cannot program, mentors can cover the programming in the small groups.
- Since students are often not used to open, inquiry-based learning, the experience has to be scaffolded

## Students

- The main student-related challenges were about student recruitment, attendance and student-to-student interactions.
- A key concern for recruitment was to obtain sufficient applications from diverse students in Chicago Public Schools with feasible commutes.
- Ensuring full attendance to the end has been an issue, which we tried to improve by tying student stipends to attendance.
- Students seemed more reluctant to interact with each other post COVID, which hampered engagement. We added more icebreakers to facilitate such interactions.

## Technology

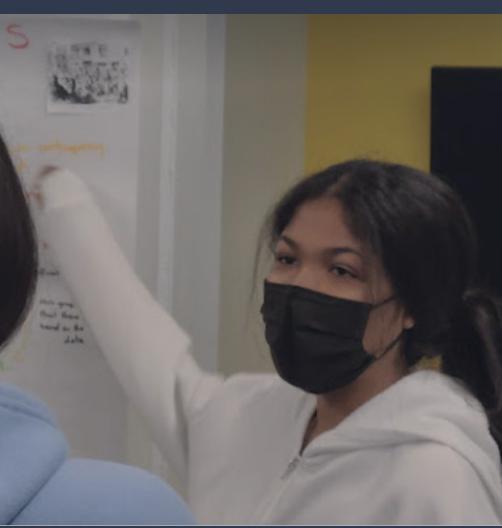
- The main technological challenges were getting the JupyterHub set up and maintained (in our case on AWS), having a Google Colab backup option, organizing a dozen loaner Chromebooks, and managing wifi guest permission problems at UChicago.

## New Cases

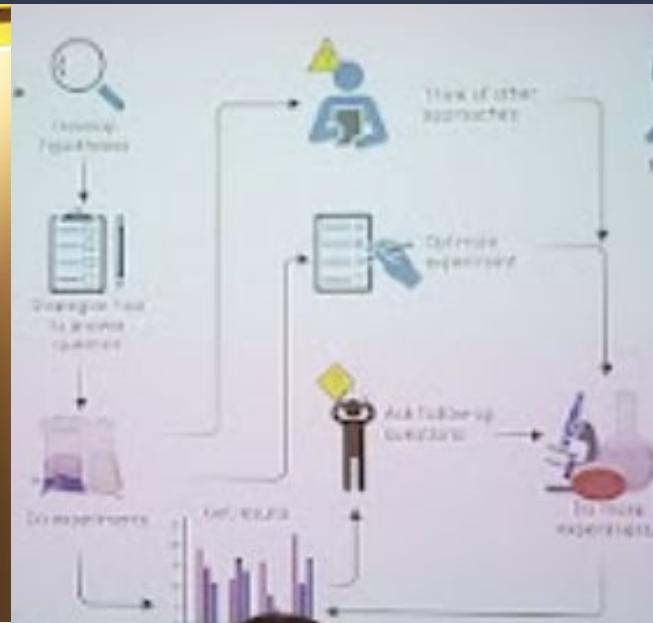
- The development of the initial cholera case across three iterations was time-intensive but additional cases that follow a similar structure (like the COVID case) take less time to develop. For instance, students were interested in community development or violence prevention as additional topics, which could be added following the initial template.

## Replicability

- Key challenges with replicating Data4All are multi-domain knowledge of instructors and mentors (see above) and the larger number (in our case six) of near-peer mentors.



# ABOUT DATA4ALL



# Data4All in Numbers

32h Workshop Content

65 Student Applications

Student Cohort: 24 (2022) | 26 (2023)

110 Pages Instructor Guide

10 Jupyter Notebooks

10 Game Activities

2 Project Managers

4 Curriculum Developers

4 High School Teachers

6 Near-Peer Mentors

3 Postdoc Mentors

# Why this Workshop

## Data4All addresses 3 gaps in data science education:

1

While data science is one of the fastest growing employment sectors, data science courses are often not yet taught in high schools. For instance, of the students who participated in Data4All in 2022, almost two thirds had taken a computer science (CS) course and three quarters had some programming experience. Most students (86%) have been introduced to scientific concepts. But only a third of students had been exposed to data analysis or data visualization – and if data analysis is introduced, it is taught separately from scientific reasoning. Data4All seeks to close these gaps.

2

At the college level, many data science students get proficient at the technical skills in computation and statistics but still struggle to critically solve problems with data. To address this challenge early on, Data4All students learn foundational programming, statistical methods, and mapping within the context of a data science reasoning framework: For eight weeks, students are working through puzzles of two cases (related to cholera and covid) while analyzing data in a browser-based programming environment (using the Python language in [Jupyter notebooks](#)). They learn to apply technical skills to assess the plausibility of evidence in relation to competing explanations and to communicate their findings to non-technical audiences.

3

There is a need for more data scientists in general, and for more data scientists from underrepresented groups in particular. Almost all of the Data4All students (90%) in 2022 and 96% in 2023 were underrepresented minorities or girls. The goal of the workshop is to introduce these students to data science-related college and career opportunities – not only in computer science or data science but also to data science in other STEM fields. This is done through the case studies, which apply data science to public health problems, through guest speakers who demonstrate the use of scientific reasoning and data analysis in modern microbiology, and through talks about college and career preparation.



# Data4All's Approach: Reasoning & Problem First

## Traditional Approach: Method First

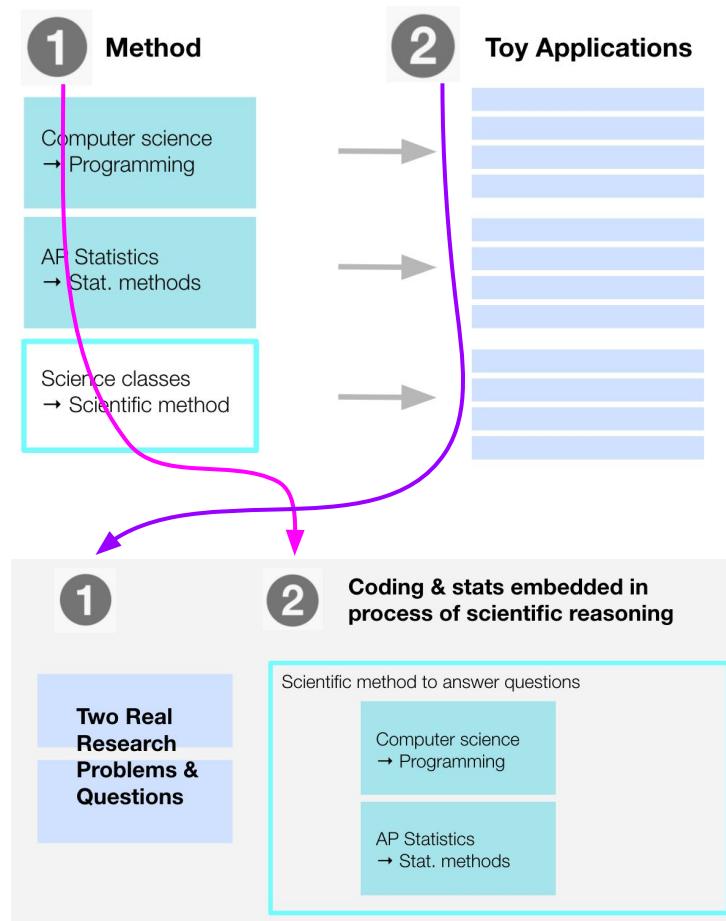
## Data4All Approach: Problem & Reasoning First

As illustrated below, in traditional high school classes, computer science, statistics, and science are typically taught separately, with little overlap. Each class often starts with a method (such as programming or statistics), followed by toy or textbook applications to illustrate the method.

With this approach, students often feel: “If I don’t want to become a computer scientist, data science isn’t for me.” Also, the relevance of why students are learning methods often is not clear to them when instructors say: “trust me, learn the methods and then you’ll see why later.” It remains unclear that one needs to learn methods in order to solve interesting problems.

To avoid both problems, Data4All starts with a [reasoning framework](#), a real research problem and a question. Then students learn how to address this question using scientific reasoning, programming, statistics, and mapping.

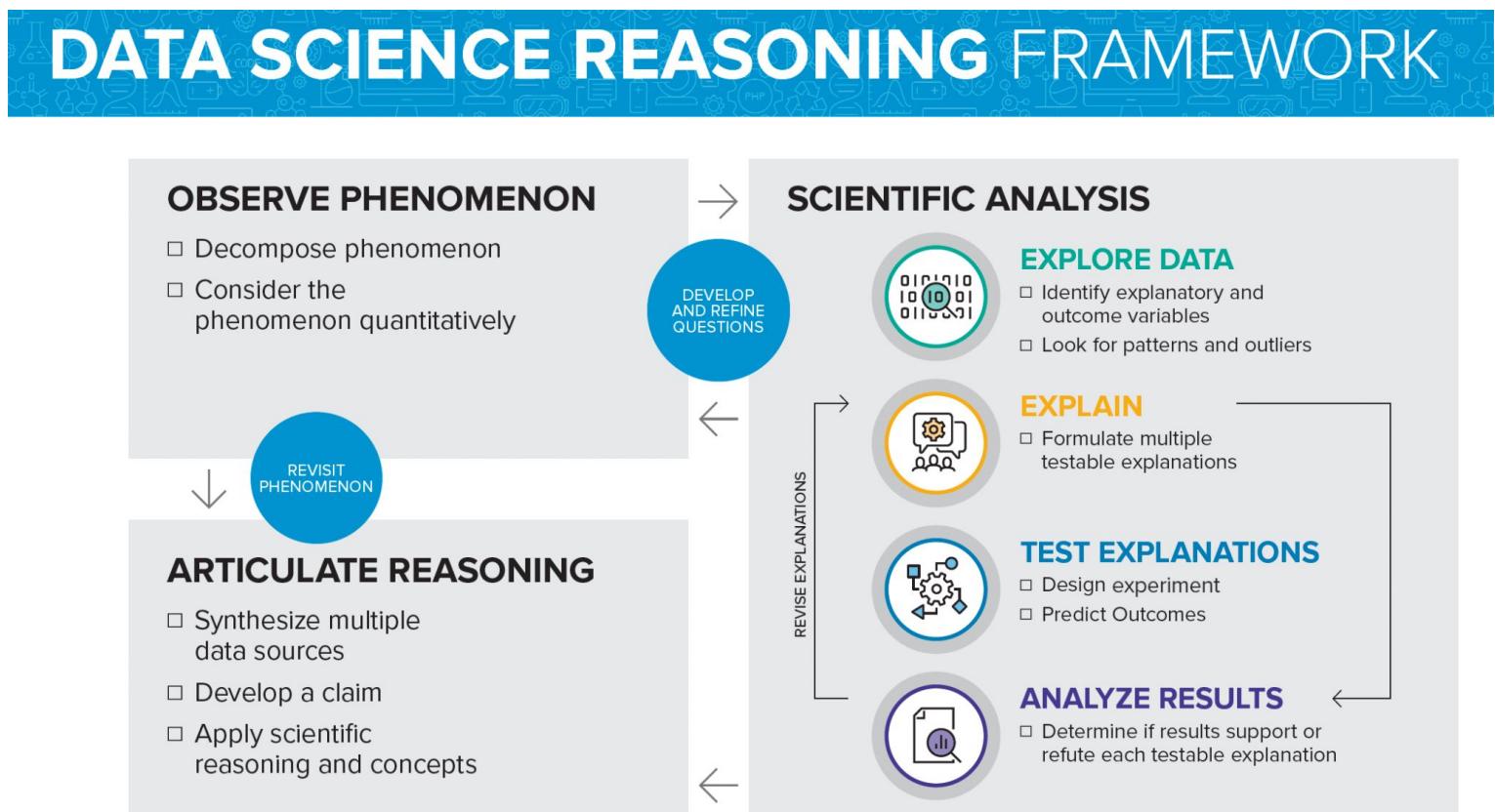
The reasoning framework on the next page outlines a logic for analyzing data.



# Data4All's 'Special Sauce': The Data Science Reasoning Framework

The Data Science Reasoning Framework (DSRF) drives the design of the curriculum. It embodies the overall progression of using data to answer questions and solve problems through describing a phenomenon, identifying interesting patterns or outliers, proposing explanations for these patterns and testing these explanations against each other as one moves toward a claim that is supported by evidence and reasoning.

The DSRF is meant to provide structure to doing data science in a way that encourages the discovery of unexpected patterns and developing explanations for these patterns as well as addressing some of its pitfalls that plague data science such as confirmation bias and focus on expected patterns.



# Workshop Activities

Data4All offers a variety of activities for learning: They range from near-peer mentors in small groups to experiential & social learning through games (spark activities) to traditional classroom formats (lectures), and guest speakers to demonstrate data science in other real-world applications and highlight related college and career opportunities. These activities are summarized here.

## Small groups with near-peer mentors

give students the opportunity to ask questions and work through coding roadblocks with their mentors and peers, so all students can get personal attention tailored to their level of understanding (groups are designed with similar previous knowledge).

## Jupyter Notebooks & Kepler

Jupyter notebooks are web-based programming platforms that allow students to learn and work with the reasoning framework while learning to analyze real data in Python with statistical methods. Kepler is a web-based spatial data visualization platform where students can learn to analyze data in map formats.

## Spark Activities

give students an intuitive sense of the reasoning framework and statistical concepts through collaborative games, e.g. illustrating the structure of a difference-in-difference research design through two teams competing to hit can pyramids where one team's results are manipulated through a can magnet intervention.

## Lectures

are presentations by high school teachers to the whole class to introduce and explain key concepts used in a given class, such as the difference between correlation and causation or between demographic characteristics and explanatory mechanisms.

## Lunch Speakers

give students an overview of how the workshop content is related to college and career options. E.g., in 2022, two PhD biologists walked students through innovative case studies of how they used data science to solve problems that save lives and improve health. Other speakers explained their own career trajectories to make the steps for choosing college majors and careers more transparent, and they overviewed where and how to access college and career information.



# Mix of Activities in each Week

To keep students engaged, the workshop activities described on the previous page were alternated throughout the 4-hour class. Below is the color-coded schedule for 2022 for all activities in 15-minute blocks. It also gives an overview of which concepts in reasoning, coding, statistics and mapping were covered each week for the cholera and covid cases.

## CASE 1: CHOLERA (learning the reasoning logic from a historical case that has been resolved)

	<b>Reasoning</b>	<b>Coding</b>	<b>Stats</b>	<b>Mapping</b>						
Week 1	Beriberi Seeing the Problem as Data	Meet the Londoners Brainstorm questions	Normalization on (Lebron vs. Jordan) Seeing the problem in the data (outcome variable)	How to ask scientific vs other questions Represent real data in Python store, group, organize data manipulate variables & lists	Normalization outcome var.					
Week 2	Is it random? Is it spatial? Dataframe manipulations	Finalizing the Question Patterns & mechanisms	Finding patterns in the variable of interest	Summarize patterns	Data exploration (filter & group) in Pandas scatter plots Patterns in randomness explanatory mechanisms outcome & explan. var.s correlation vs causation					
Week 3	Develop proposed explanations	Annotate base map: predictions	Explore spatial patterns of cholera/qualitative analysis of predictions	Correlation, causation and contingency tables Dr. Campbell: DS in microbiology	Distance analysis for pumps and sewers	Developing alternative explanations Predictions Distance analysis for pumps and sewers	Correlation, causation quasi-exp. design contingency tables statistical significance	Explore spatial patterns of cholera/qualitative analysis of predictions		
Week 4	Analyze results, refine theories	The pump handle & adding time	Spurious correlations, experimental design	Correlation and Causation South London Experiment (part 1)	Difference in Difference Amy Kirby - CDC	Predict DID results for each theory	South London Experiment (part 2)	Predictions, testing explanations I, analyzing results I, refine explanations, testing explanations II	contingency tables, statistical significance, and data visualization applied to the cholera case	temporal analysis spurious correlations correlation vs causation difference-in-difference
Week 5	Analyze DID Results	Evidence-Based Reasoning and Confirmation Bias	Changing Minds: Arguing from Evidence	Construct argument poster	Persuade the Londoners	Research lifecycle meta activity		Analyzing results II, evidence-based reasoning, articulate reasoning confirmation bias Construct argument poster Convince others		

## CASE 2: COVID (applying the reasoning logic to a modern case with open questions)

	<b>Reasoning</b>	<b>Coding</b>	<b>Stats</b>	<b>Mapping</b>				
Week 6	The early days of covid Seeing the problem as data Covid impact on counties	Measuring covid's impact	Dom McCoy (To & Through)	Spatial patterns of positivity Revise research question, How can a community be described with data?	Asking the right questions, exploring data, revising research questions in light of new results creating variables data exploration	outcome var.s	Spatial patterns of covid	
Week 7	Categorize vars from spark activity	Uncover correlations between variables and positivity rates	Proxy variables Proposed explanations	MADD Center with Ashlyn Sparrow	Test explanations I Multiple causes Revise theory: more vars	Exploring data, proposing explanations, analyze results I, testing explanations I, revise theory	correlations btw outcome and explanatory variables contingency table	diff btw demographic characteristics & explanatory mechanisms, multiple causes, p-values
Week 8	Groups present initial findings	Comparing 2 Explanatory Variables	Test explanations II	Analyze results and finalize theory; Construct Argument UChicago Promise Future opportunities at DSI + ANL	Guide policy decisions	Testing competing explanations II, analyze results II, articulate reasoning, finalize theory, construct argument, present findings, guide policy	loading data & required libraries, scatter plots and best-fit lines, histograms, contingency tables, line graphs, analysis	

15-min blocks  
(without intro, admin & breaks)

- spark / large group activity
- notebook
- small group
- lecture
- lunch speaker

2022

# Activities & Learning Objectives

The graph below shows how we expect the workshop activities to relate to short-term learning objectives. It also summarizes what we invested to produce the activities.

These objectives include 1) a broader understanding and appeal of data science, 2) improved confidence in programming, statistics, math and reasoning skills, and 3) greater college and career preparedness. The [background section spells out in more detail](#) how each activity is expected to impact each outcome in the short and [long term](#).

[Students were surveyed](#) before and after the workshop to assess in how far these short-term objectives were reached or not after 8 weeks.

The learning objectives are referenced on each page of the respective evaluation section.

## What We Invest

**Curriculum** developed by 4 team members (2 Argonne, Tyler, CSDS), including: [Data science reasoning framework](#) and scoping of 2 real projects with data

### Staffing

1 project manager (DSI)  
4 high school teachers (Argonne)  
6 near-peer mentors (CS/DSI)  
3 postdoc mentors (DSI)  
5 lunch speakers

### Partners

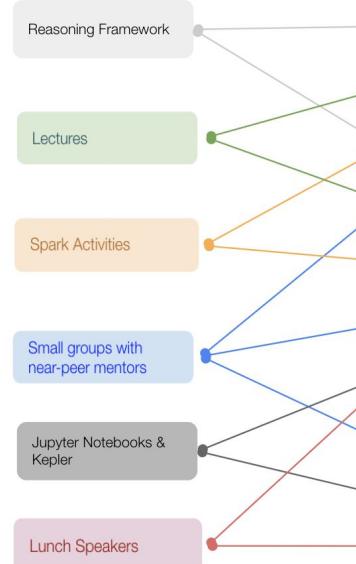
Data Science Institute  
Argonne National Labs  
Center for Spatial Data Science  
Office for Civic Engagement

**Funding** from Milgrom (student scholarships) & DSI (lunch & project mgmt)

**Tech infrastructure:** Jupyter Hub setup, Chromebook loaners

**Meeting space** at UChicago's Crerar Library (DSI and CS Dept)

## Activities



## Learning Objectives

### 1. Understanding: What is Data Science?

- 1.1 Students see CS & data science in a larger context of scientific reasoning & problem solving with data
- 1.2 Students who might not major in CS find new relevance and broader appeal in CS and data science
- 1.3 Students realize broader connection of CS & statistics to STEM fields like microbiology & env. science

### 2. Skills: Programming, Statistics, Mapping & Reasoning

- 2.1 Programmers add skills in reasoning with data while others (like debaters) add programming skills
- 2.2 Students' confidence increases in:
  - analyzing data with Python & Kepler (mapping)
  - understanding foundational statistical concepts
  - using both to critically explore causes of problems

### 3. College & Career Preparedness

- 3.1 Students realize that you can use data science in lots of different fields that match their interests
- 3.2 Students find Data4All useful for clarifying data science-related college & career goals

# EVALUATION

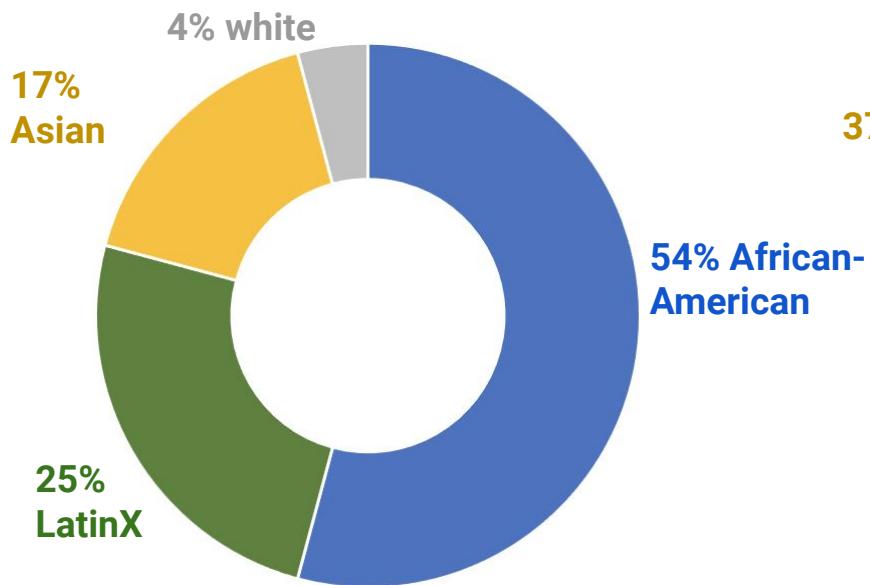


# 1. DID DATA4ALL REACH ITS TARGET AUDIENCE? YES

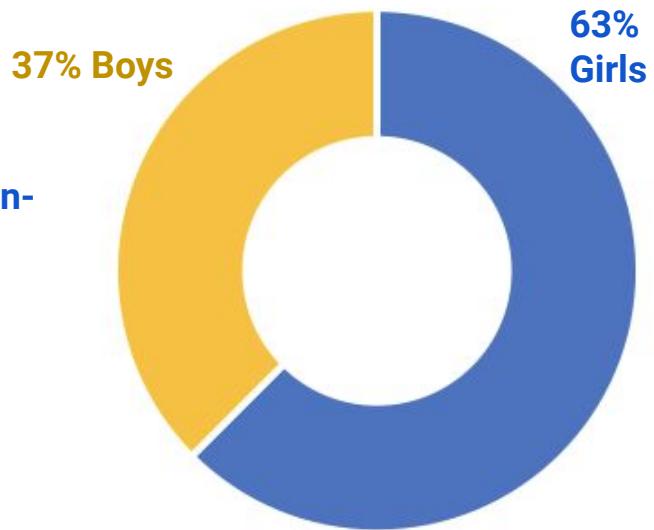


# 92% of the 24 Data4All Students were underrepresented in data science (2022)

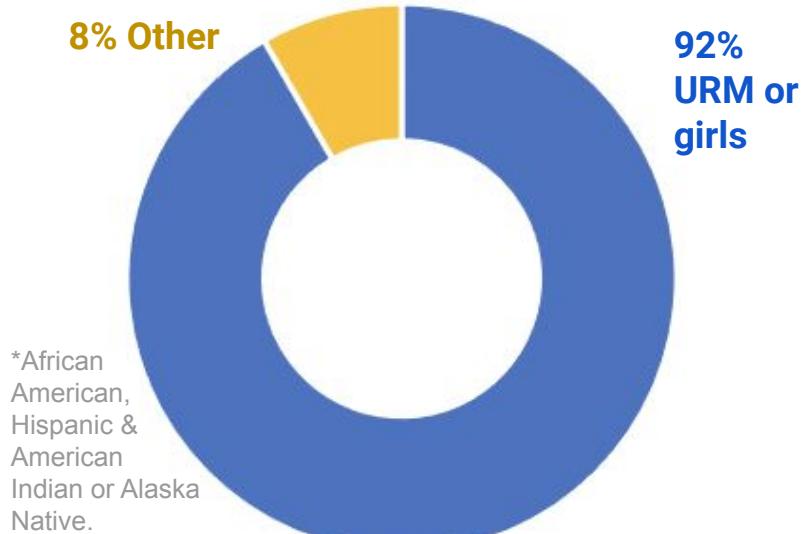
**96% of students** were racially or ethnically diverse (not white)



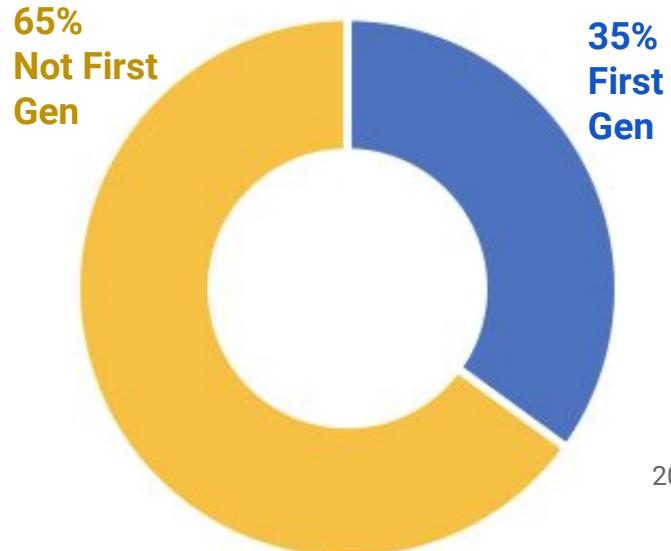
**Almost two thirds of students** were girls



**92% of students** met NSF's criteria of underrepresented minorities in STEM (URM\*) or were girls. 79% were URM.

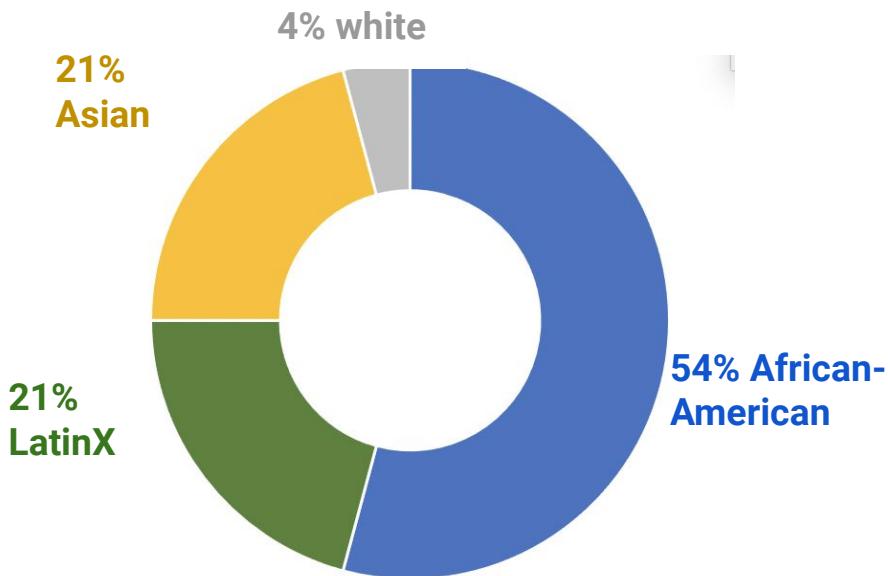


**About a third of students** would be the first generation to go to college

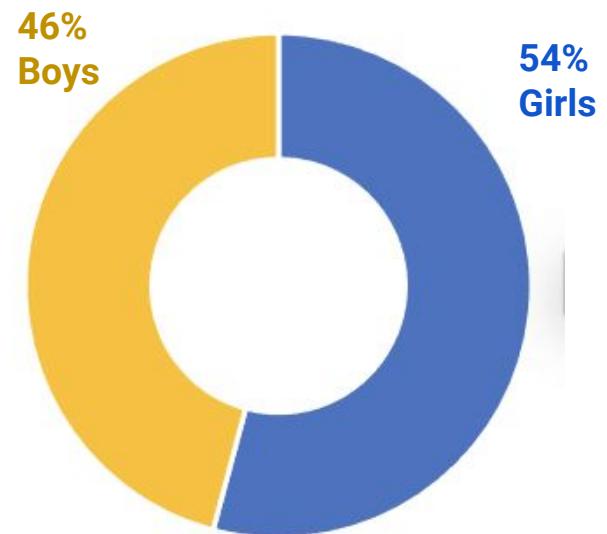


# 96% of the 26 Data4All Students were underrepresented in data science (2023)

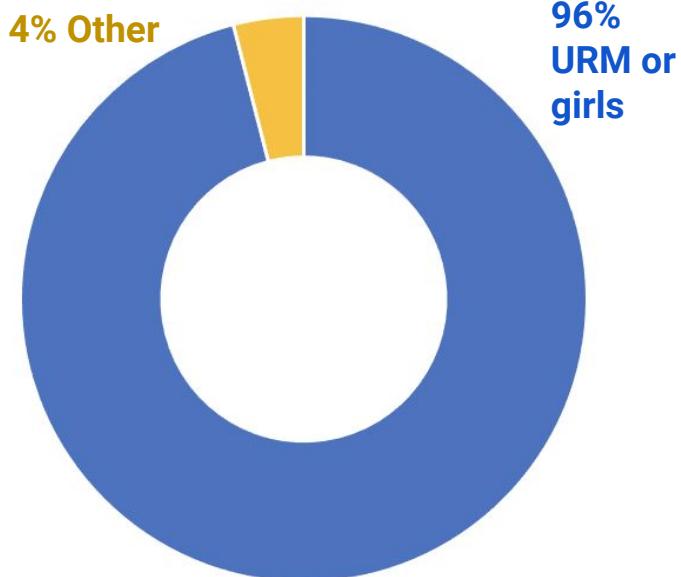
**96% of students** were racially or ethnically diverse (not white)



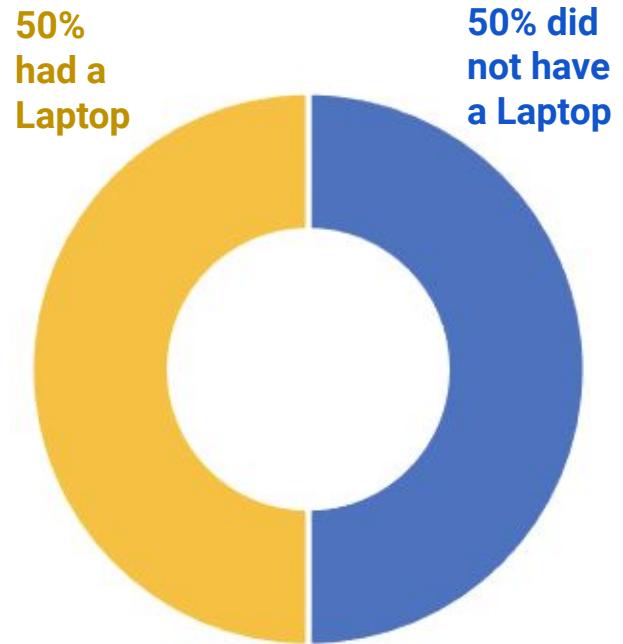
**More than half of students** were girls



**96% of students** met NSF's criteria of underrepresented minorities (URM) in STEM or were girls. **75% of students were URM.**



**Half of all students** needed a loaner laptop

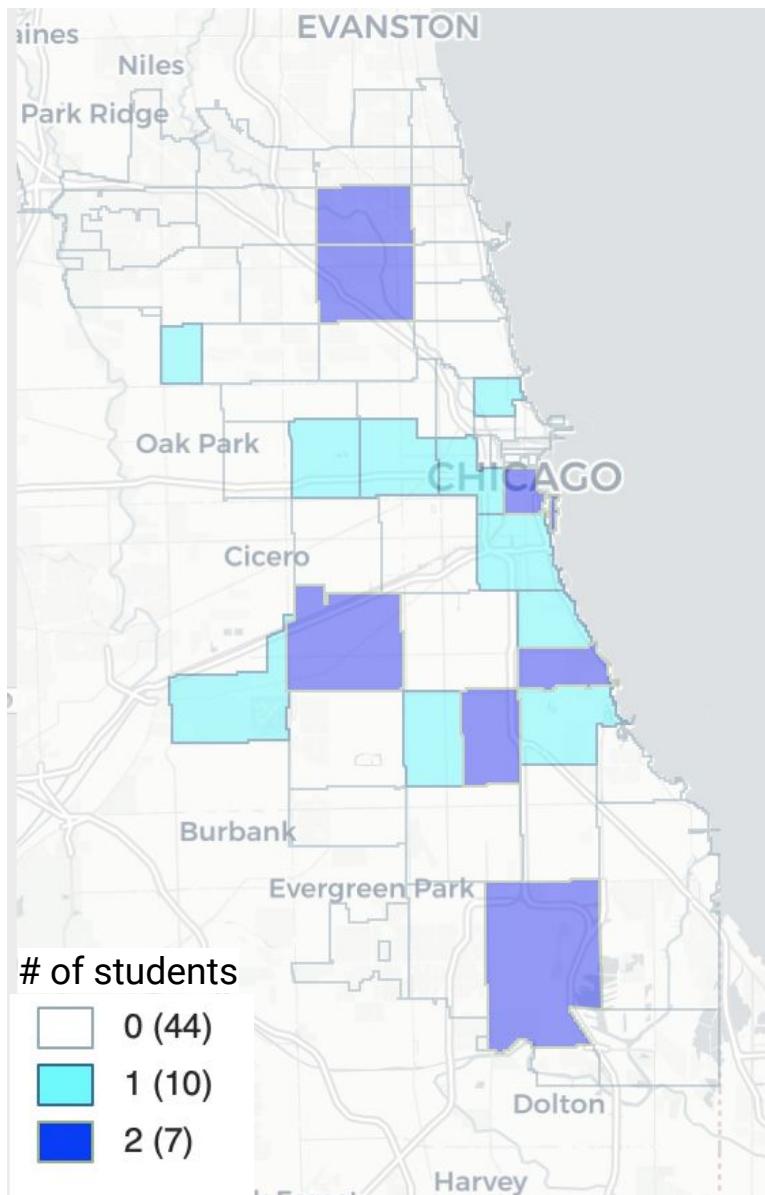


Students came from 18 schools across Chicagoland in 2022 and 2023, majority public schools and half Southside high schools

## Locations of Students' High Schools (HS)

2022

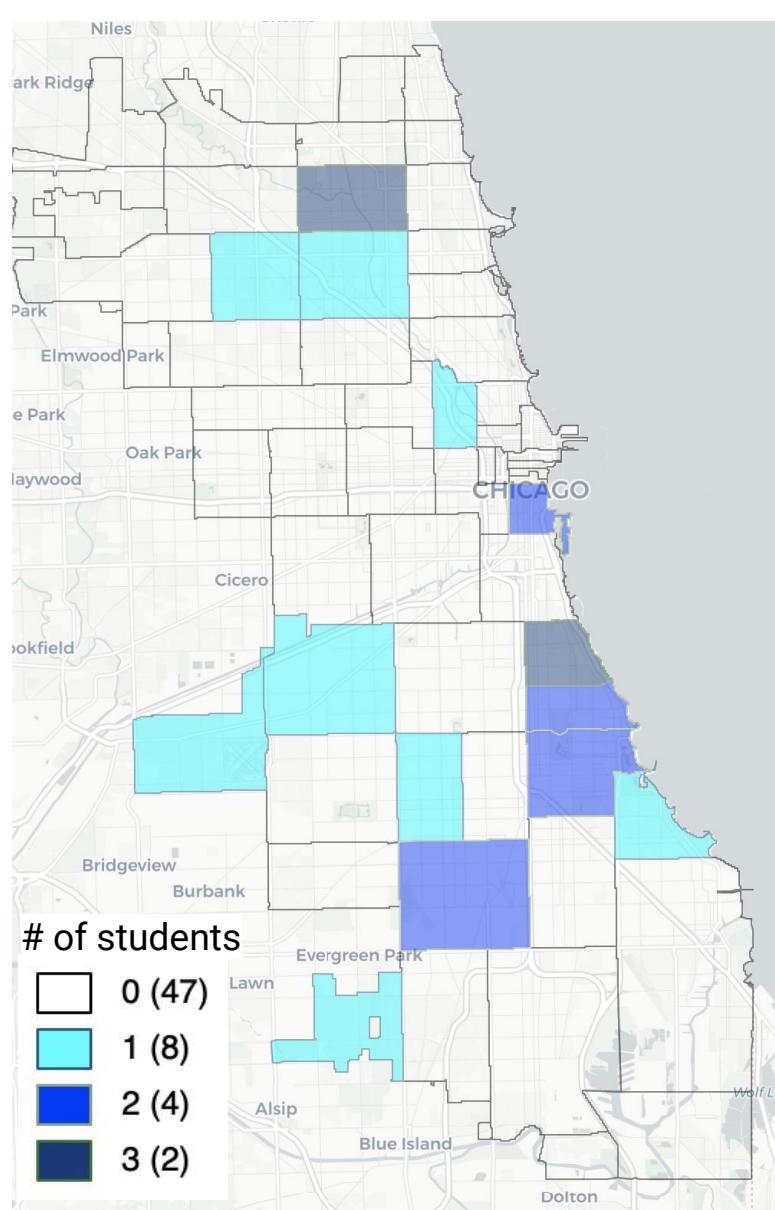
50% from Southside HS (12 out of 24 students)



Homewood HS not included (outside of Chicago)

2023

54% from Southside HS (14 out of 26 students)



4 students from outside of Chicago not included on map: Willows Academy, Homewood-Flossmoor HS and Thornwood HS

Students came from 18 schools across Chicagoland in 2022 and 2023, majority public schools and half Southside high schools

## **Names of Students' Schools 2022**

### **CPS or Public (21 students or 88%)**

- 1 Englewood STEM High School
- 2 George Westinghouse College Prep
- 3 Gwendolyn Brooks College Prep
- 4 Homewood-Flossmoor High School
- 5 Hyde Park Academy High School
- 6 John Hancock College Prep
- 7 Jones College Prep
- 8 Kenwood Academy
- 9 Lane Tech High School
- 10 Lindblom Math and Science Academy
- 11 Marie Skłodowska Curie Metropolitan High School
- 12 Northside College Prep
- 13 Richard T. Crane Medical Preparatory High School
- 14 Von Steuben Metropolitan High School
- 15 Walter Payton College Prep
- 16 Whitney Young

### **Charter (3 students or 12%)**

- 17 Mansueto High School
- 18 Urban Prep Academy

## **Names of Students' Schools 2023**

### **CPS or Public (20 students or 77%)**

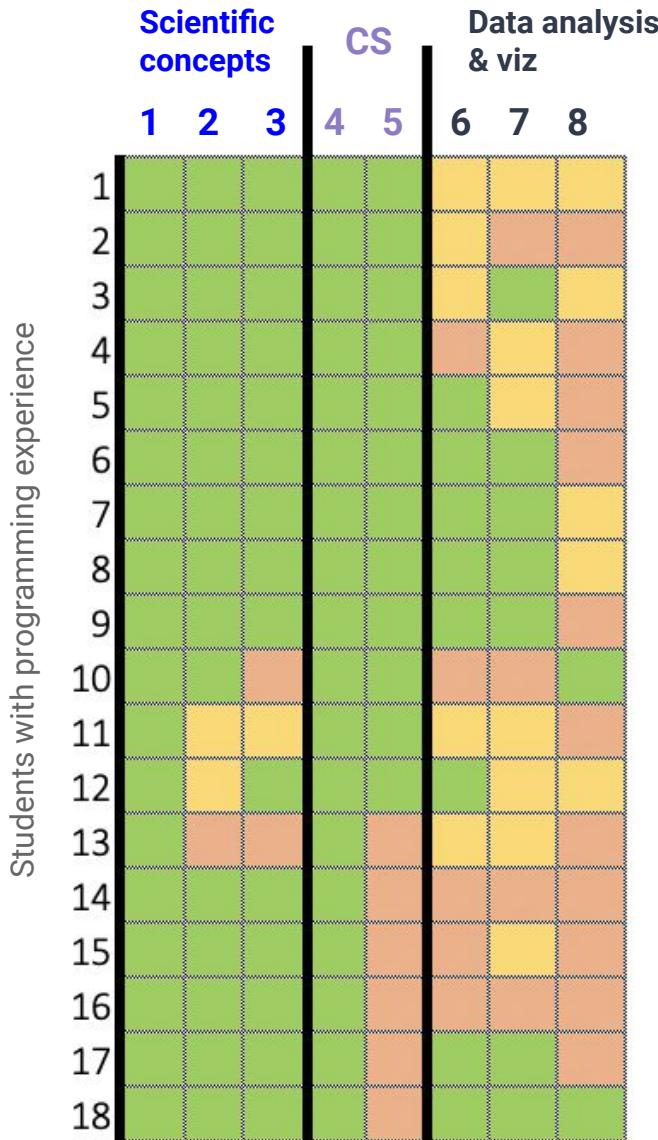
- 1 Eric Solorio Academy High School
- 2 Homewood-Flossmoor High School
- 3 Hyde Park Academy
- 4 Jones College Prep
- 5 Kenwood Academy
- 6 King College Prep
- 7 Lane Tech
- 8 Lindblom Math and Science
- 9 Mather HS
- 10 Northside College Prep
- 11 Robert Lindblom Math and Science Academy
- 12 Schurz High School
- 13 South Shore International College Prep

### **Private or Homeschooled (6 students or 23%)**

- 14 Leo Catholic High School
- 15 Marist High School
- 16 Thornwood High School
- 17 Willows Academy
- 18 Homeschooled



Prior to Data4All (2022), almost all students (86%) were exposed to scientific concepts, and most had some programming experience (69%) but fewer (32%) had data analysis & visualization skills



2022

#### COLUMN HEADERS

##### Scientific concepts

- 1 Problem-solving
- 2 Developing a hypothesis
- 3 Testing a hypothesis

##### Computer science

- 4 Programming languages/coding
- 5 CS Course Experience

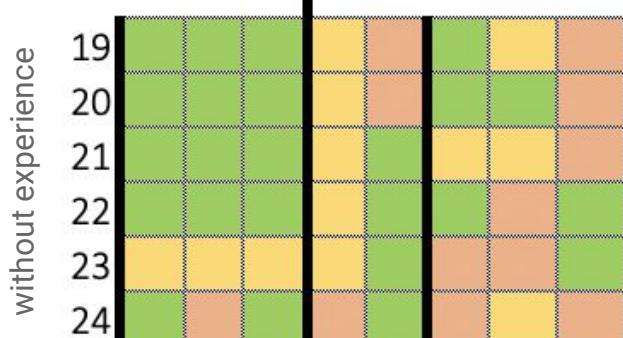
##### Data analysis & visualization

- 6 Data analysis
- 7 Data visualization
- 8 Jupyter Notebooks

#### COLOR CODING

- I've worked with or used this before.
- I might have heard of this before, but I haven't used or worked with it before.
- I'm not familiar with this.

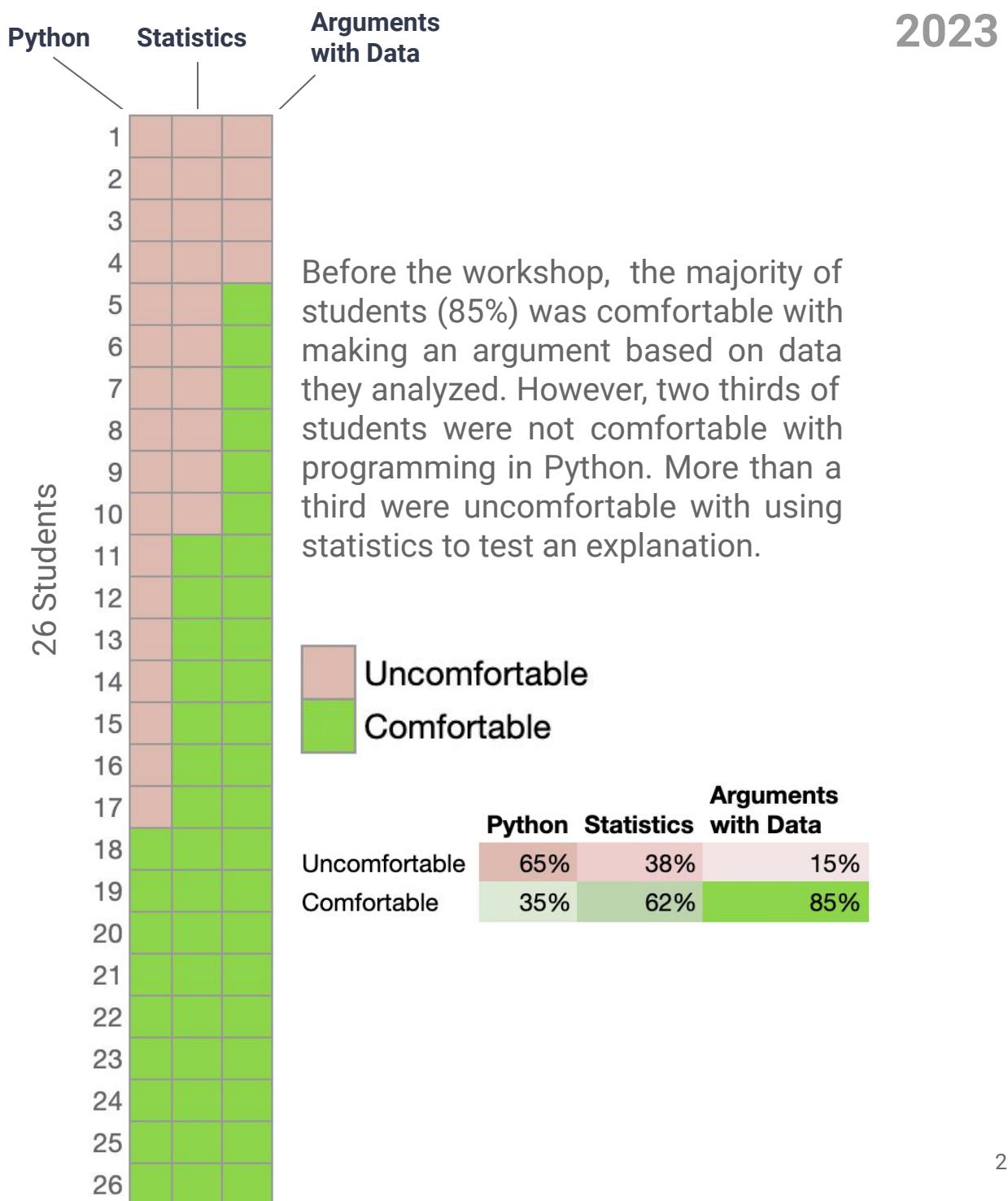
#### SHARE OF STUDENTS WITH EXPERIENCE IN:



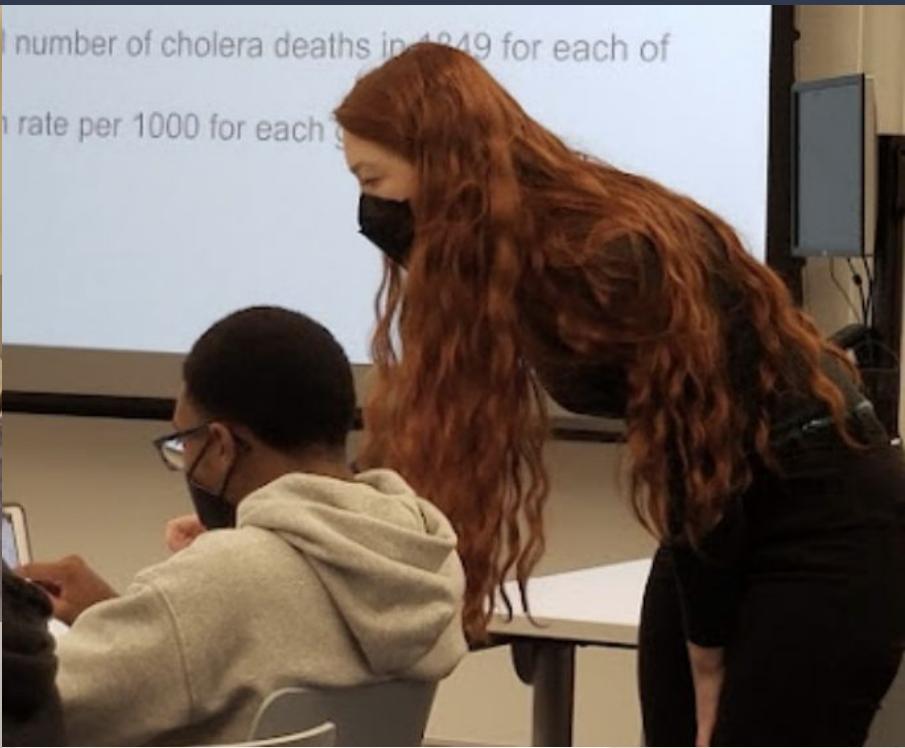
has experience

scientific reasoning	Problem-solving	96%
86%	Testing a hypothesis	83%
	Developing a hypothesis	79%
computer science	Programming languages/coding	75%
69%	CS Course Experience	63%
data analysis	Data analysis	46%
32%	Data visualization	33%
	Jupyter Notebooks	17%

Prior to Data4All (2023), almost all students (85%) were comfortable with evidence-based arguments, but two thirds did not feel comfortable programming (65%) and over a third (38%) were not comfortable with statistics



## 2. DID DATA4ALL REACH ITS LEARNING OBJECTIVES? YES



# After Data4All, students did see data science in a larger context of scientific reasoning

## Learning Objective 1.1



**Walter Payton senior who took AP CS:** [Before Data4All] I thought data science was about analyzing large data sets. [Afterwards:] In part, I still think data science is about analyzing large data sets. But, it's also breaking down the data sets to look for evidence that can support a hypothesis through correlation, p-value, etc.

**Mansueto HS senior with little programming experience:** [Before Data4All] I didn't think anything about it, like it never crossed my mind. And for the very little times it did I always thought it was too difficult for me. [Afterwards:] Data science is using data to answer difficult questions that without it you probably wouldn't be able to answer.

**Von Steuben HS junior who took a CS class:** "It has influenced me to think in many ways, so to be open minded basically."

**Crane Medical Prep HS sophomore who took a CS class:** "It has shown me a new side of data science by seeing that it can come with many different factors that i haven't thought about."

**Kenwood Academy sophomore with programming experience** [highlight of the workshop overall]: "explaining why i believe stuff"

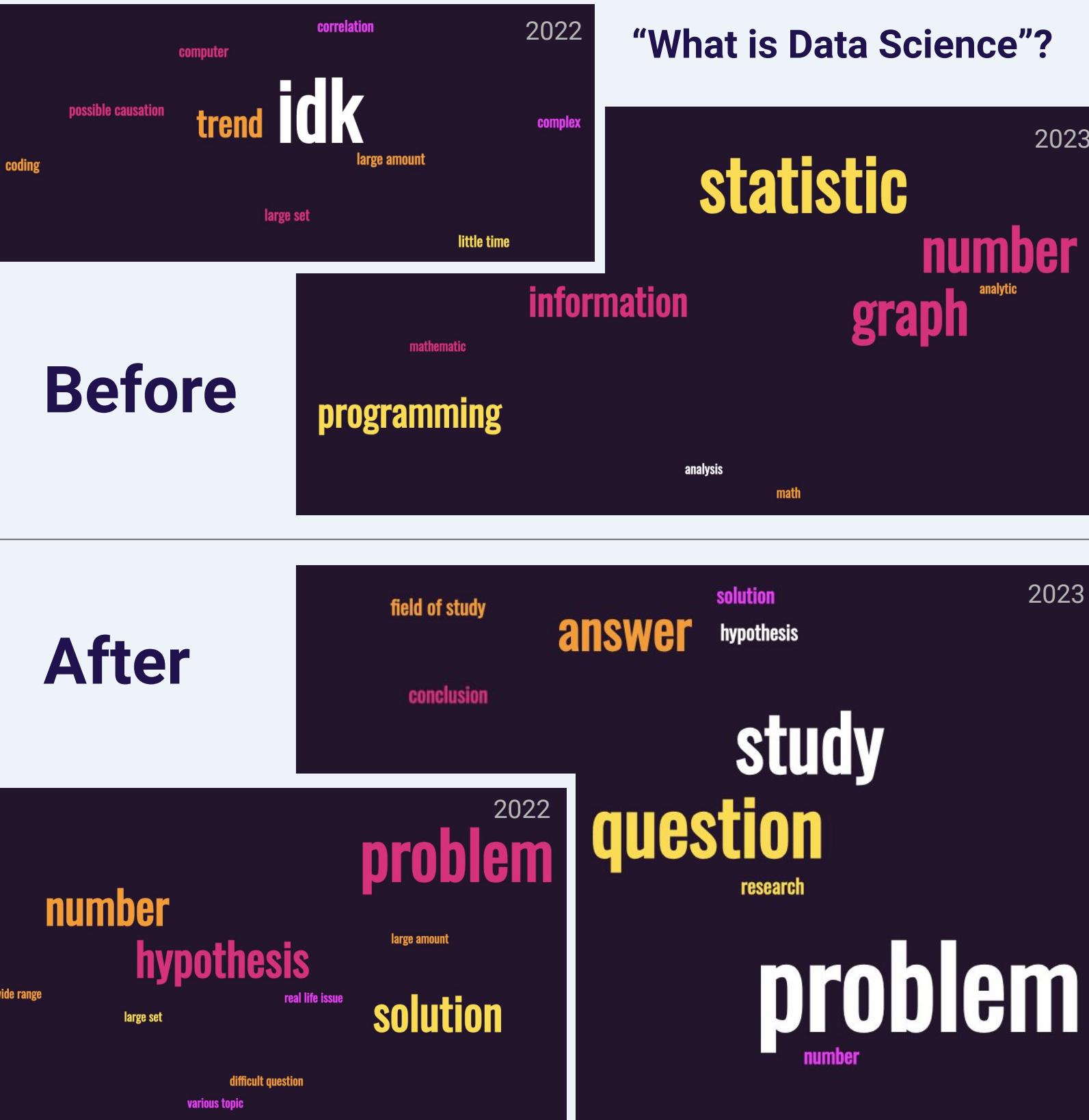
**Jones College Prep sophomore with programming experience:** "A highlight of the program is finding the reason that cholera was spreading. And the collaboration to create hypotheses in the final week."

**Anonymous student:** "I liked the different problem solving techniques we learned, and how a lot of the activities required us to work together and piece together the different information we're given to find the solution to a problem."

Before Data4All, students saw data science more narrowly as programming and statistics. Afterwards, they saw it in a larger context of scientific reasoning.

The responses of students before and after the 2022 and 2023 workshops to the question **When you hear the phrase "data science", what comes to mind?**

*The words "data" and "science" have been removed from the answers before running the word clouds.*



# Integration of scientific reasoning broadened appeal of CS & stats for students who might not major in CS

## Learning Objective 1.2

Midpoint feedback from students: I'm interested in exploring fields beyond computational or data science.

**2022**

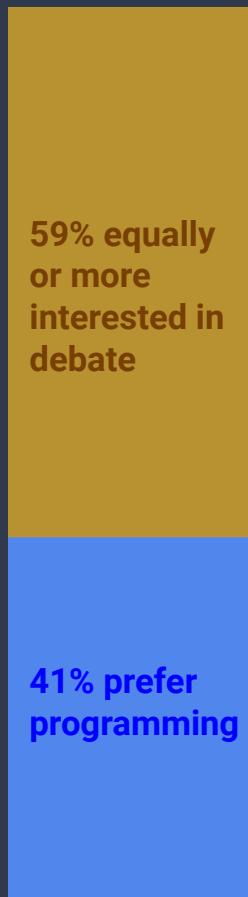
20 students



Pre-feedback from students: Are you more or equally interested in debate than programming or more interested in programming than debate?

**2023**

26 students



**Two seniors from Englewood STEM and Homewood-Flossmoor who are interested in exploring fields beyond CS and data science:**

"My experience in data science workshops have been very influential, i love science, debate, and mysteries, and data science is a great way to bring all of my interests together into one very influential field of study."

"It's expanded my view of data science so much, which is so helpful because I feel like I can use this in my future."

**Northside Prep College junior without prior programming experience:**

"I enjoyed the notebooks and the Cholera problems had me more interested. I also enjoyed learning more about UChicago from all the mentors, they helped a lot with telling us about the school and showing us around during lunches."

Through the reasoning framework and speakers, students realized broader connection of CS & statistics to STEM

Learning Objective 1.3

**Jones College Prep senior who is interested in exploring fields beyond computation and data science:** “My experience in the workshop has made me even more inclined to pursue computer science more than before. Mainly because I was in between pursuing public health and computer science but now I am sure that I can use both my skills in data and computer science to also pursue my passion for solving health and environmental issues.”

**Lane Tech senior with AP CS experience:** “It made me realize of new career possibilities I was not aware of before. For example, combining data science with medicine (Microbiology).”

**Crane Medical Prep sophomore with programming experience:** “It has shown me a new side of data science by seeing that it can come with many different factors that i haven't thought about. ... “The highlight was this group activity of showing and proving our hypothesis/claim.”

**Walter Payton senior with AP CS experience:** “I really enjoyed being able to talk with my teammates about our theories for our case-studies it really feels like we are scientists conducting and testing real-life experiments.”

Programmers added skills in reasoning with data while others also added programming skills

#### Learning Objective 2.1

The data science reasoning framework addressed skills gaps for students with and without programming experience



#### **Students with extensive prior programming experience:**

Biggest workshop challenge at midpoint for **Lane Tech senior with AP CS experience**: “Arguing against the opponent's beliefs and persuading them.”

Workshop highlight at midpoint for **Walter Payton senior with AP CS experience**: “A highlight of the program is being able to verbalize my ideas and findings.”

[biggest challenge of workshop overall]: “The biggest challenge was communicating my ideas, especially presenting data.”

#### **Students with little prior programming experience:**

[Highlights of workshop overall]:

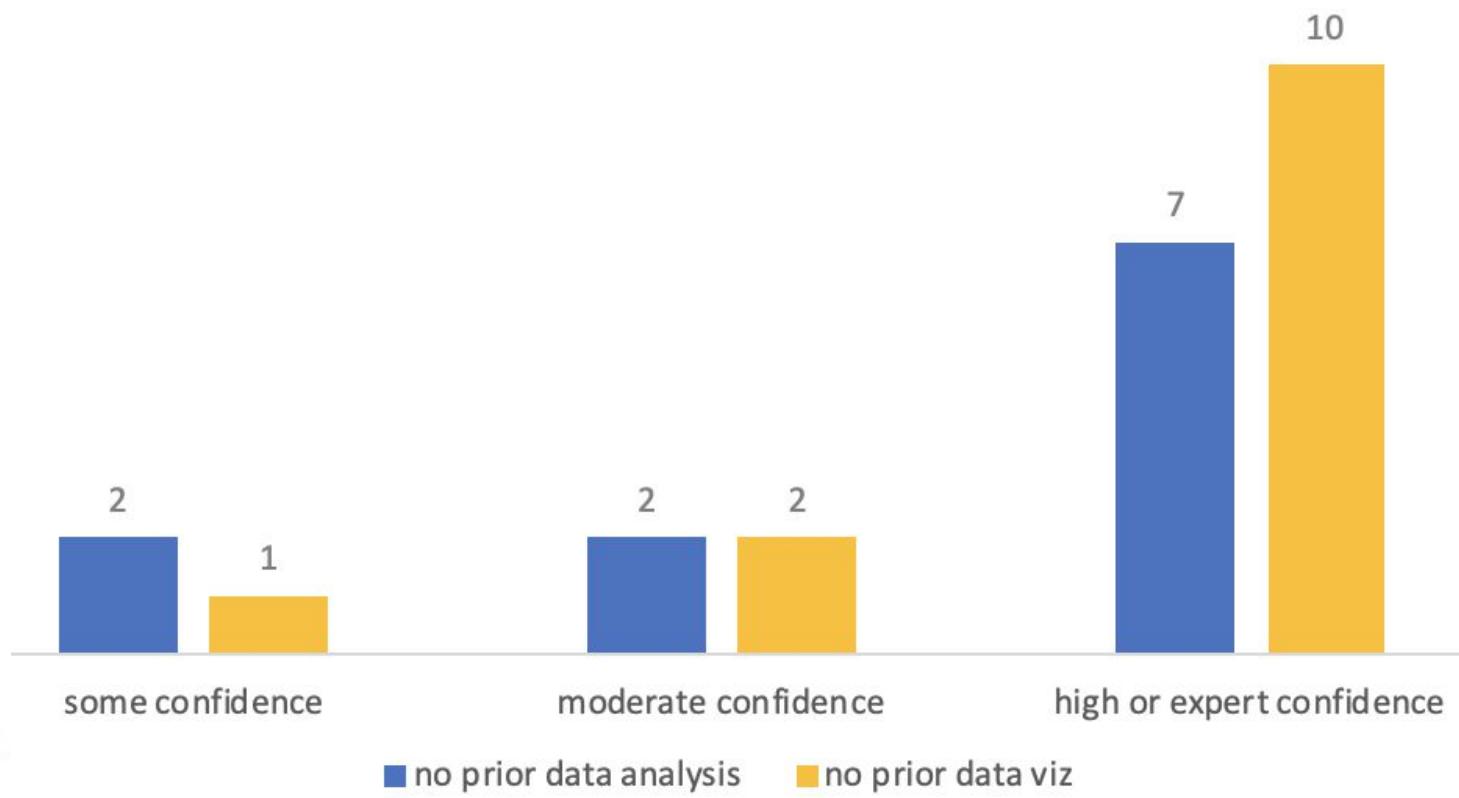
**Whitney Young junior:** “Learning more about coding.”

**Marie Curie HS sophomore:** “I made the code more effective.”

# Students without prior experience in data analysis or data visualization reported increased confidence in both as a result of Data4All (2022)

## Learning Objective 2.2

***Self-reported confidence levels by students without prior experience in data analysis or data visualization at workshop midpoint or end (2022)***



# Data4All increased 81% of students' confidence in programming and 57% in statistics (others stayed comfortable) (2023)

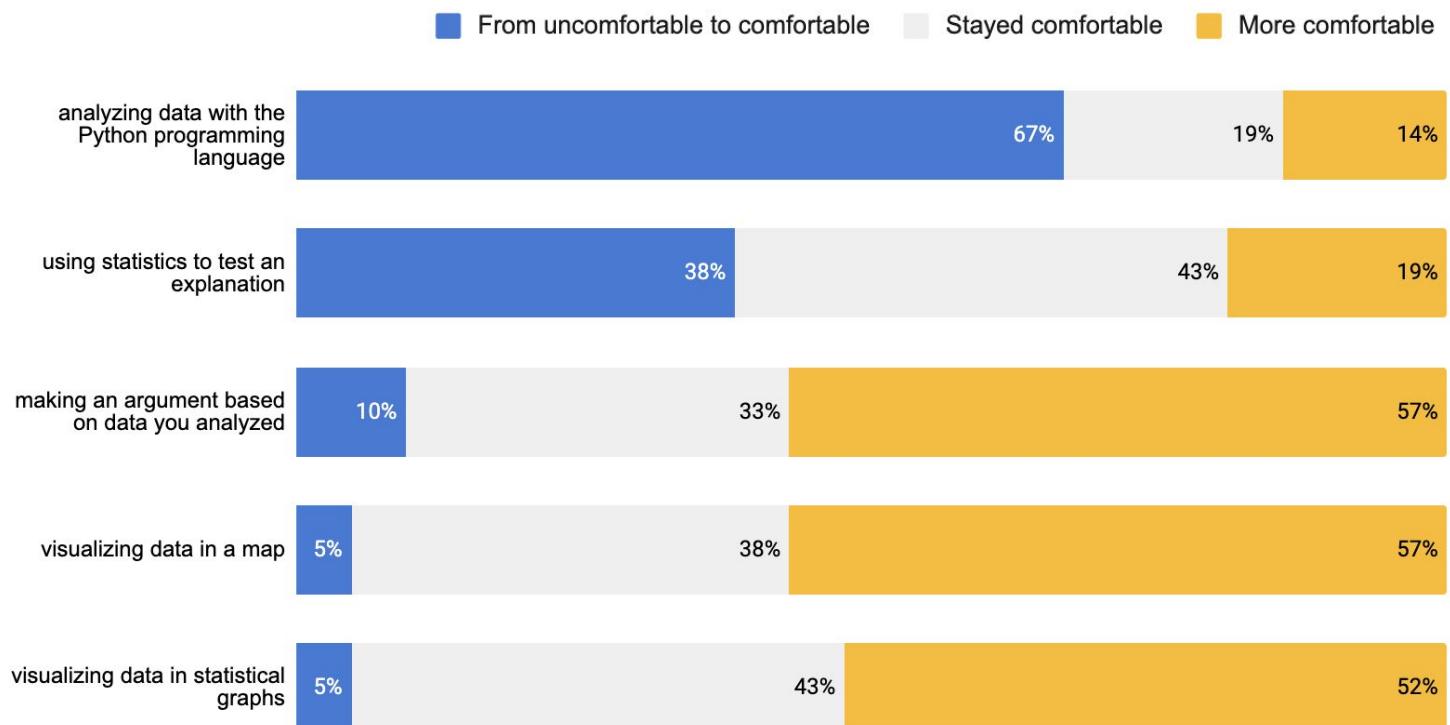
## Learning Objective 2.2

The second learning objective of Data4All was to increase students' skills in Python programming, statistics, mapping & reasoning with data. To assess progress in these areas, we asked students about their level of confidence before and after the 2023 workshop. Before the workshop, students were least comfortable with Python, followed by statistical tests. They had greater confidence in making an argument based on data and visualizing data on a map or in a statistical graph.

Given these baseline results, the greatest increase in confidence occurred in programming, where Data4All increased 81% of students' confidence (67% of all students went from feeling uncomfortable to comfortable after the workshop). In statistics the workshop increased confidence for 57% of students. Detailed results are summarized below.

Increasing students' confidence in these areas is important in light of the 2022 National Assessment of Educational Progress, which showed that students' achievements in data science declined more than in other areas, especially for underrepresented groups (see *DataScience4Everyone's* review [here](#)).

## Students' Comfort Level Before and After the 2023 Workshop



*Blue on left:* Students who started uncomfortable.

*Yellow on right:* Students whose confidence improved for other comfort levels.

# Students realize broad relevance of data science through guest speakers

## Learning Objective 3.1

### **Weekly feedback from anonymous students (2023):**

"I enjoyed learning about the different game labs and careers within them. I think Ashley did a really good job explaining the different aspects of the lab and the connections within them, like how data scientists team up with others to help them create visuals with their data."

"Today's highlight for me was the tour we took of the computer science labs downstairs. It was really exciting to learn more about specific careers and majors in computer science, and what you could do in them."

"A highlight today was seeing how much terminology was used from biology and U.S History. It was very cool to see."

# Students found Data4All useful for clarifying data science-related college & career goals

## Learning Objective 3.2

Midpoint feedback from 20 students in 2022: This experience has helped me evaluate whether I would like to pursue data science & computer science in the future.



No opinion

**Northside College Prep sophomore with little programming experience:** "This workshop experience has definitely influenced me and gotten me interested in data science. It was my first time working with coding, or taking a computer science class, and it has encouraged me to look for more computer science classes to take in the future."

**Lane Tech sophomore with some programming experience:** "I came into this loving computer science and not knowing much about data science but I am now leaving loving computer science even more and know that I can pursue dat science as a career of choice."

**Mansueto HS senior with little programming experience:** "Before i never thought about using data science and computer science for my future career but now I feel like no matter what I end up going into I could definitely be using what I learned."

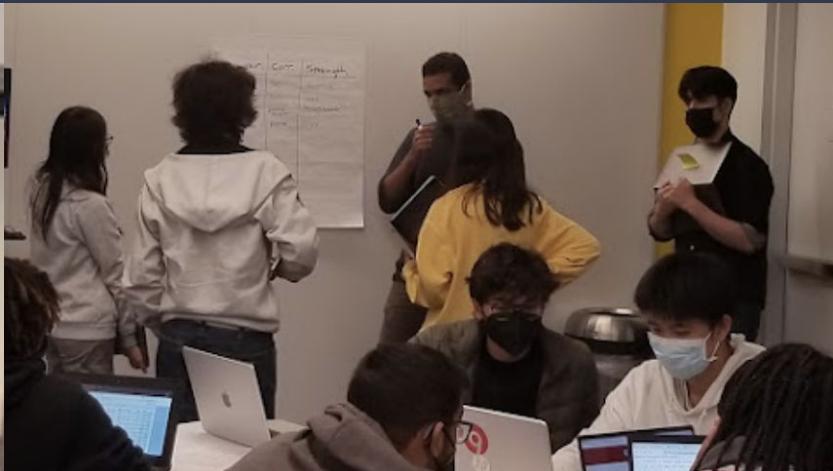
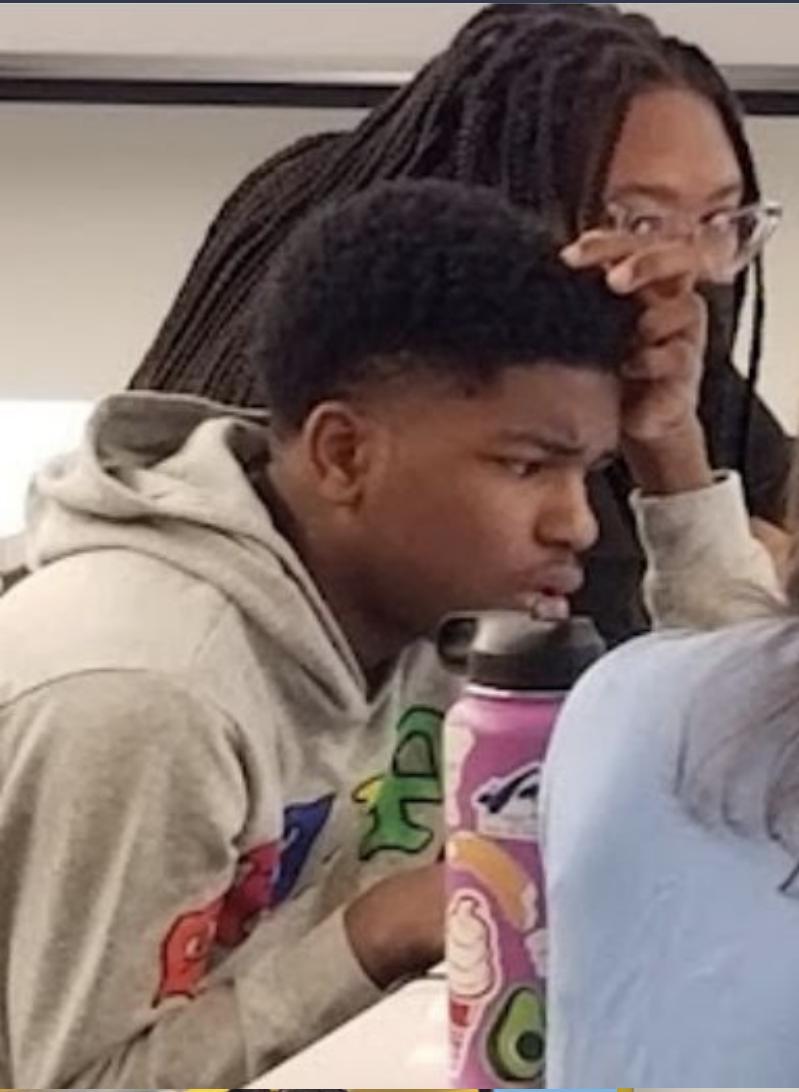
**Lane Tech sophomore with little programming experience:** "I have really enjoyed working on the Jupyter notebooks because while I have experience with programming, I have never used it for data science visualizations as we have done in this program. Working with my mentor has also been great because not only has she helped me learn more about data science, but she has also given me advice for school and college applications."

**South Shore International College Prep sophomore without programming experience:** [the highlight was] "getting to meet guest speakers and learning about jobs related to data science."

### 3. EVALUATION OF DATA4ALL ACTIVITIES AND STAFF



# SMALL GROUPS & MENTORS



# Students LOVED their small groups & mentors

Feedback from students:  
Rate your satisfaction with  
your mentor:

**2022** Mid  
20 students



**2023** Mid  
24 students



Students worked in small groups of 4-6 with one near-peer mentor each (a total of six mentors). The near-peer mentors were UChicago undergraduate students with expertise in data science and scientific reasoning. For some of the workshop weeks the near-peer mentors were joined or substituted by postdoc preceptors from the Data Science Institute who focused on data science pedagogy.

Mentors helped students work through the reasoning, programming, mapping, and statistical concepts and addressed their questions about college. As the feedback illustrates, students found their mentors and small groups especially helpful for learning how to program and debug their code.

**Englewood STEM senior with extensive Python skills:** “I adore my group, we always have a fun time while getting our work done. We share thoughts, ideas, and just feed off of each other's creativity.”

**Lane Tech senior who had taken AP CS:** “Small groups worked more efficiently and at a deeper level than whole class discussions. Having TAs who are studying data and programming was a major privilege and help.”

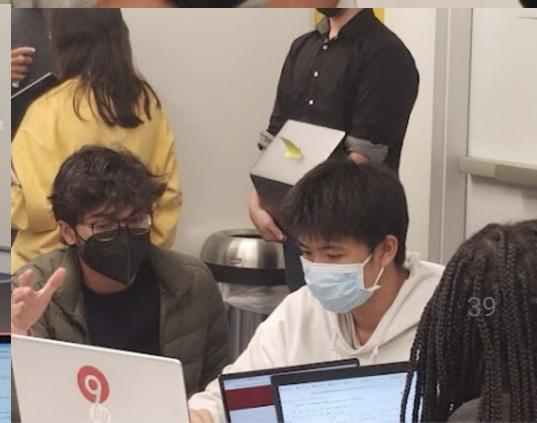
**Mansueto HS senior with less programming experience:** “Sometimes when we would be working on the jupyter hub I would get stuck on some parts but then I would ask for help and I would have my mentor or one of the people in my group help me.”

**Jones College Prep senior with some Python skills:** “My biggest challenge has been learning how to connect the pieces of data so far. My teammates really help me see the connections through our discussions but I hope to also develop those skills on my own as well.”

What was a **highlight** of the program (2022)?



# STATS & PROGRAMMING EXERCISES



# Students really liked the stats & programming exercises ...

... even though they also found the work with Jupyter notebooks the most challenging

Feedback from students: How useful are the Jupyter notebooks?

**2022** Mid

20 students



**2023** Mid

24 students



**Walter Payton senior who took AP CS:** “I usually don't enjoy coding, but the way things were broken down in the notebook and how extensive things were explained (either in the notebook or by the mentor) really made it more enjoyable and manageable. It's also nice to know that I can now analyze both tables and graphs to draw conclusions.”

**Brooks College Prep junior who loves CS:** “the biggest highlight would be using Jupiter notebooks , i like the way i can code in the notebooks”

**Northside College Prep sophomore with less programming experience:** “I have really enjoyed using Jupyter to not only perform calculations using python, but also to visualize the data with histograms and line plots.”

**Englewood STEM senior with less programming experience:** “I liked the coding parts, they were challenging but fun, especially when I tried to fix the issue”

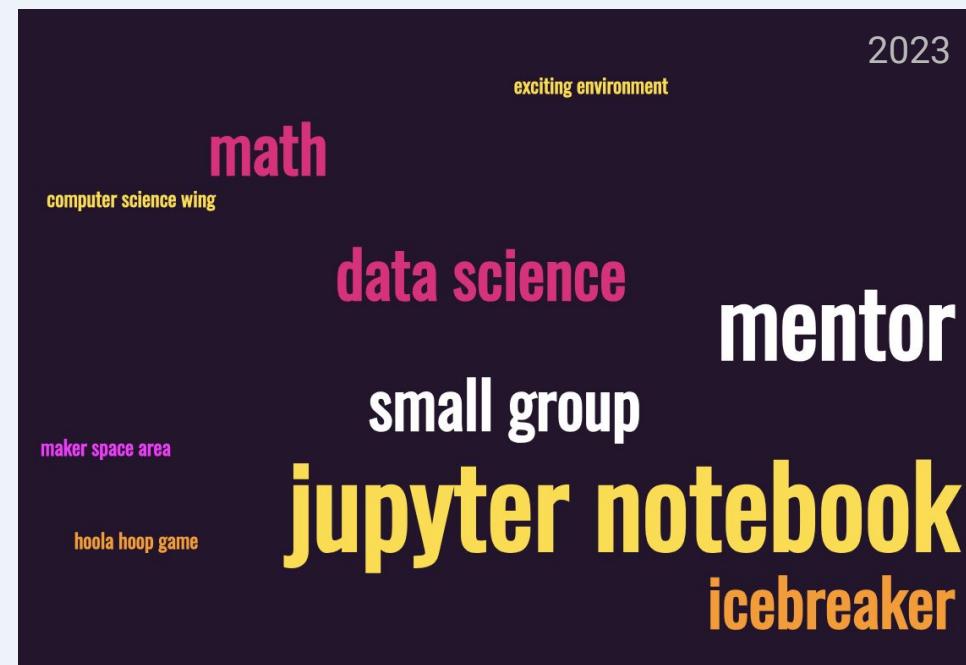
**Kenwood Academy sophomore with programming experience [the biggest challenge of the workshop overall]:** “the errors.”

**Lane Tech sophomore with little programming experience:** “I have always been interested in computer science, but I feel like I am even more so interested now as I found the Jupyter Hub notebooks to be my favorite part of this experience. Using python to model the data was fun, engaging, and also significantly helpful in proving our theories about cholera and COVID-19. I am definitely more interested in pursuing a career related to data science or CS after high school.”

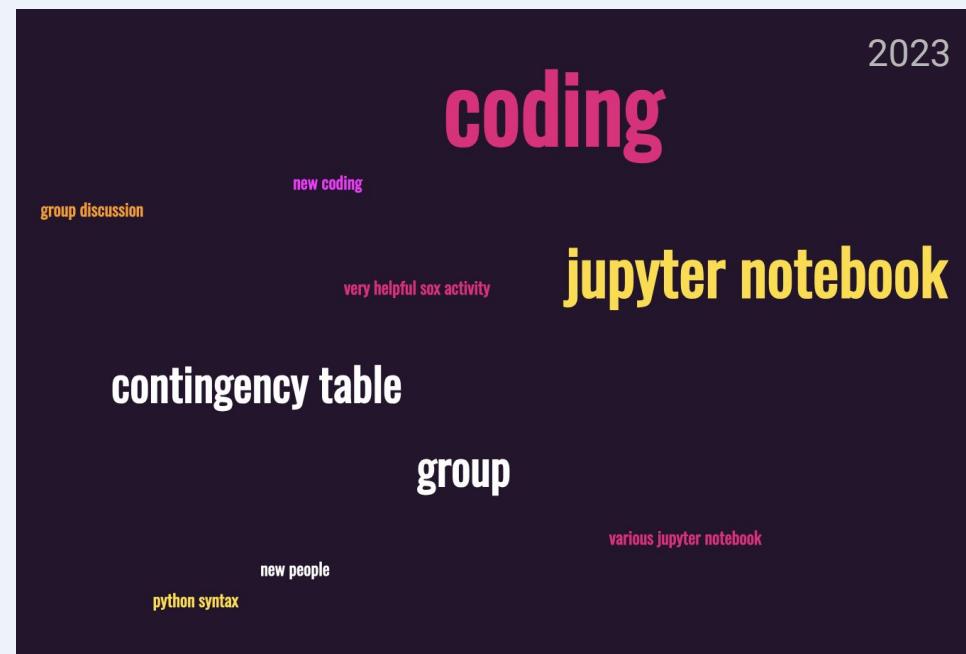
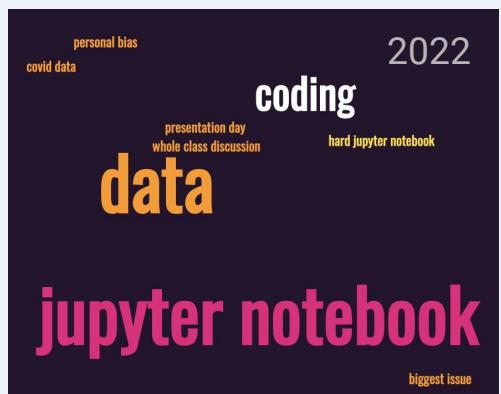
Students really liked the stats & programming exercises ...

... even though they also found the work with Jupyter notebooks the most challenging

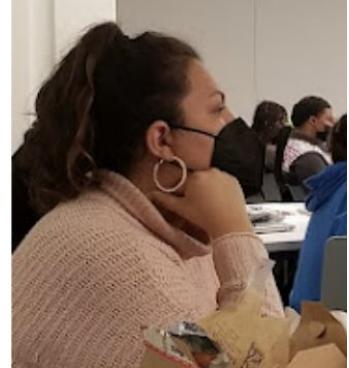
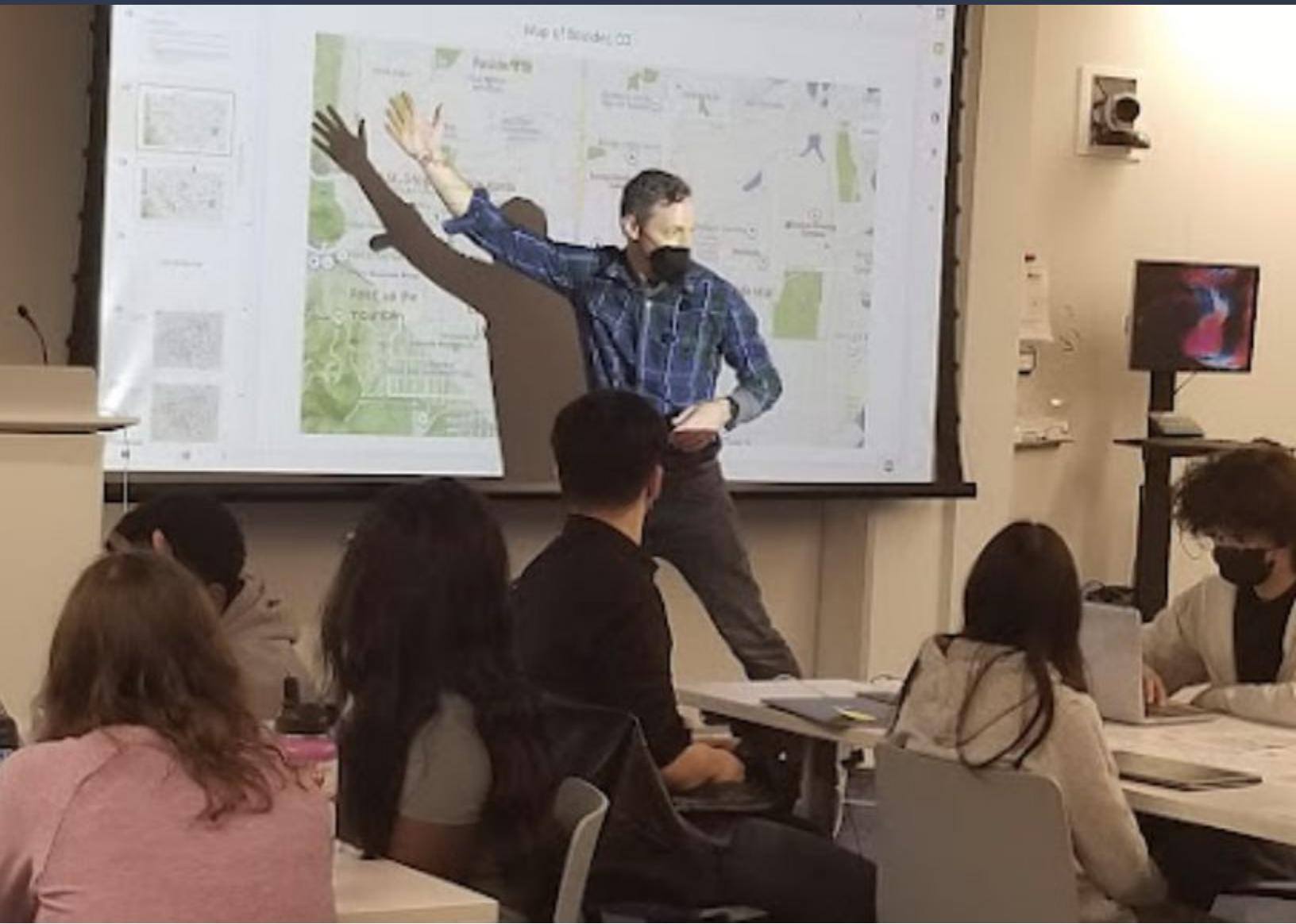
What was a highlight of your time in the program?



What was the biggest challenge in the program?



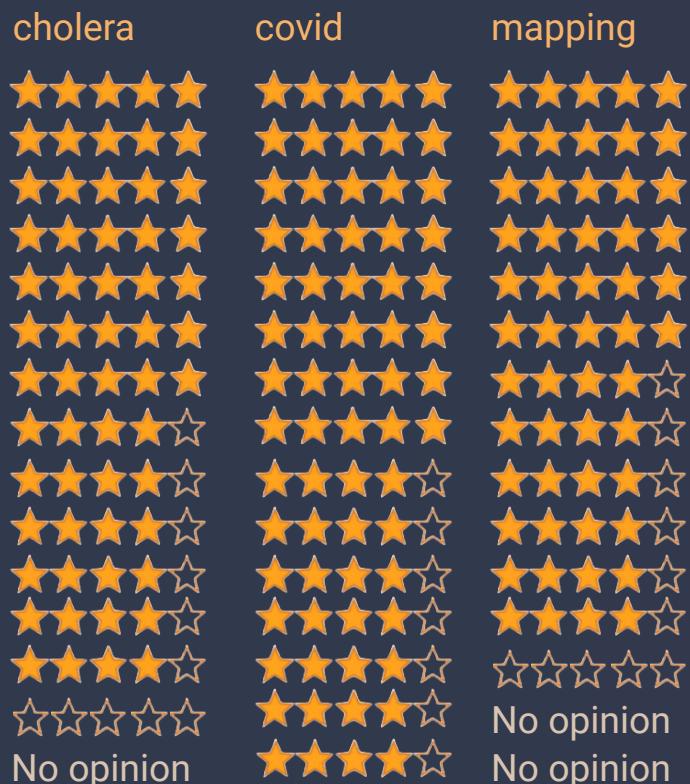
# REAL-WORLD CASES & MAPPING



The cases & mapping were interesting, with slight preference for covid case

2022

Final feedback from 15 students: How interesting and engaging were the cases and mapping ...



Data4All applies its reasoning framework to two real-world problems that students address with programming, mapping, and statistical tools: 1) **Why cholera spread** in 19th century London and 2) **how COVID-19 impacted Chicago neighborhoods differently** in the early 2020s. Since the first case is historical, it is more scaffolded, while the 2nd case allows students to apply the same reasoning to a more open-ended problem.

2023

Final feedback from 21 students: How interesting and engaging were the cases and mapping ...



# A frequent workshop highlight was tracing why & how cholera spread



**Englewood STEM senior with extensive Python skills:** “Today we had to defend our theories with all of the data and I had so much fun doing this. It was great using data to change peoples' minds about the way cholera spread.”

**Von Steuben Metro HS Asian junior with less programming experience:** [Highlight of the workshop at midpoint]: “How the first day we made a bunch of questions on the list of the reasons we thought why and how cholera was spreading.”

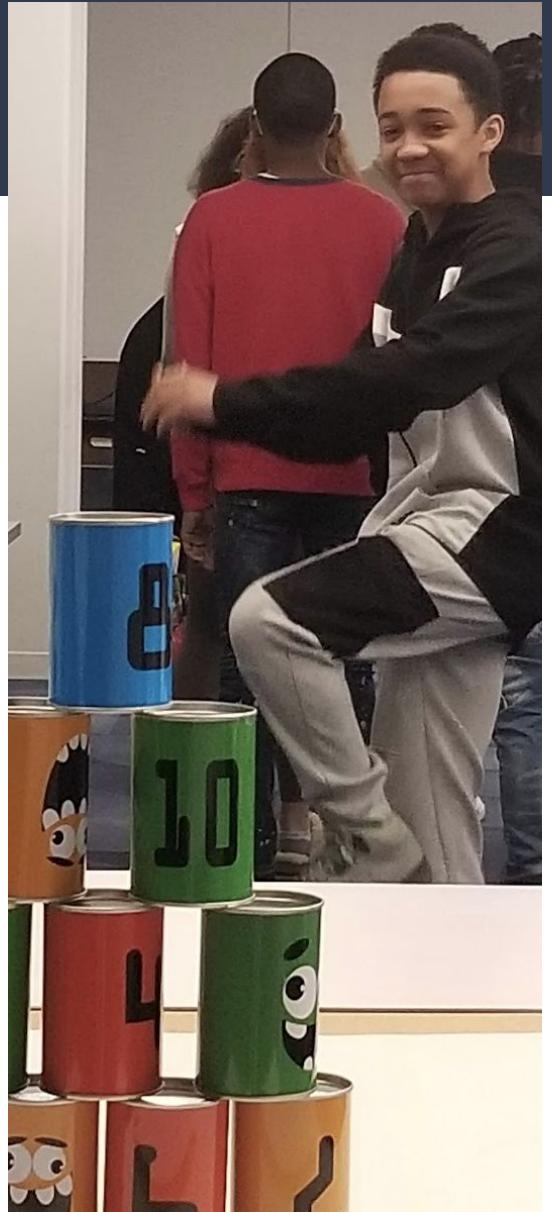
[highlight of the workshop overall]: “When we created arguments for Cholera”

**Lane Tech sophomore with some programming experience:** [Highlight of the workshop overall]: “I liked when we were presenting our cholera information to the speakers and making the poster.”

**Willows Academy junior with little prior programming experience:** “The highlight of my experience so far has probably been learning about the real life data science problem, the cholera outbreak, and how John Snow was able to navigate the challenges the disease presented with victorian age technology and medicine.”

**Jones College Prep sophomore with programming experience:** “A highlight was finding the cause of Cholera.”

# THE SPARK ACTIVITIES



# Students found lectures & spark activities (very) useful

2022

Midpoint feedback from 20 students: How useful are ...

the lectures? the activities?



Most students found the lectures and spark activities useful or very useful in 2022 and 2023. In 2023, students' feedback about the lectures was that they were too foundational. We are revising the lectures before publicly releasing the Data4All teaching materials to address this concern.

2023

Final feedback from 21 students: How useful are ...

the lectures? the activities?



# Students found the spark activities fun

**Hyde Park Academy sophomore with little prior programming experience who was curious what data science is:** "This was very fun and showed that doing data science wasn't as "boring" as I had imaged and support my interest in computer science. ..."

[highlight of the whole workshop]: I like when we did the cans activities and ordering how to do research."

*[In the can activity, we graphed the number of cans still standing for two teams of students who threw bean bags at a pyramid of cans. In round two, magnets were secretly added to the second team's cans, which decreased their scores. This was used to illustrate the quasi-experimental research design with an intervention for the cholera case.]*

"I liked when we did activities that were more hands on."

**Whitney Young junior with basic prior programming experience:** [about the spark activities]: "They were enjoyable."

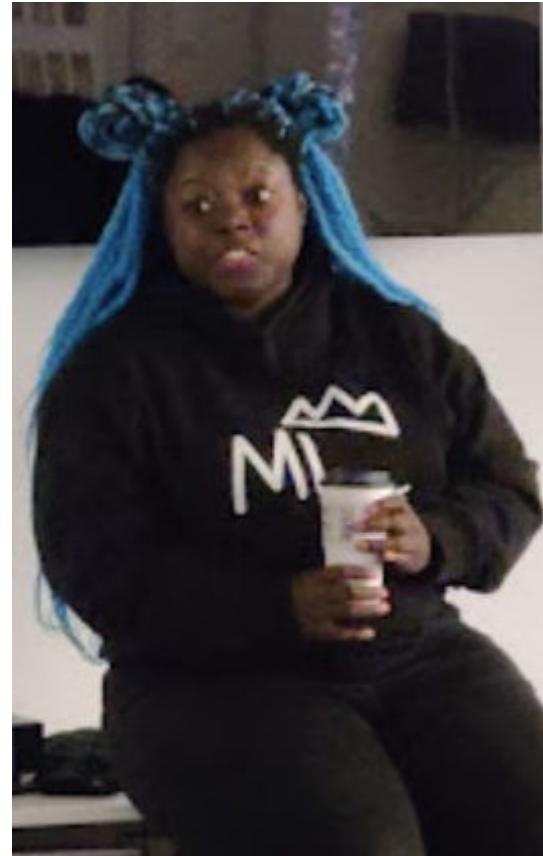
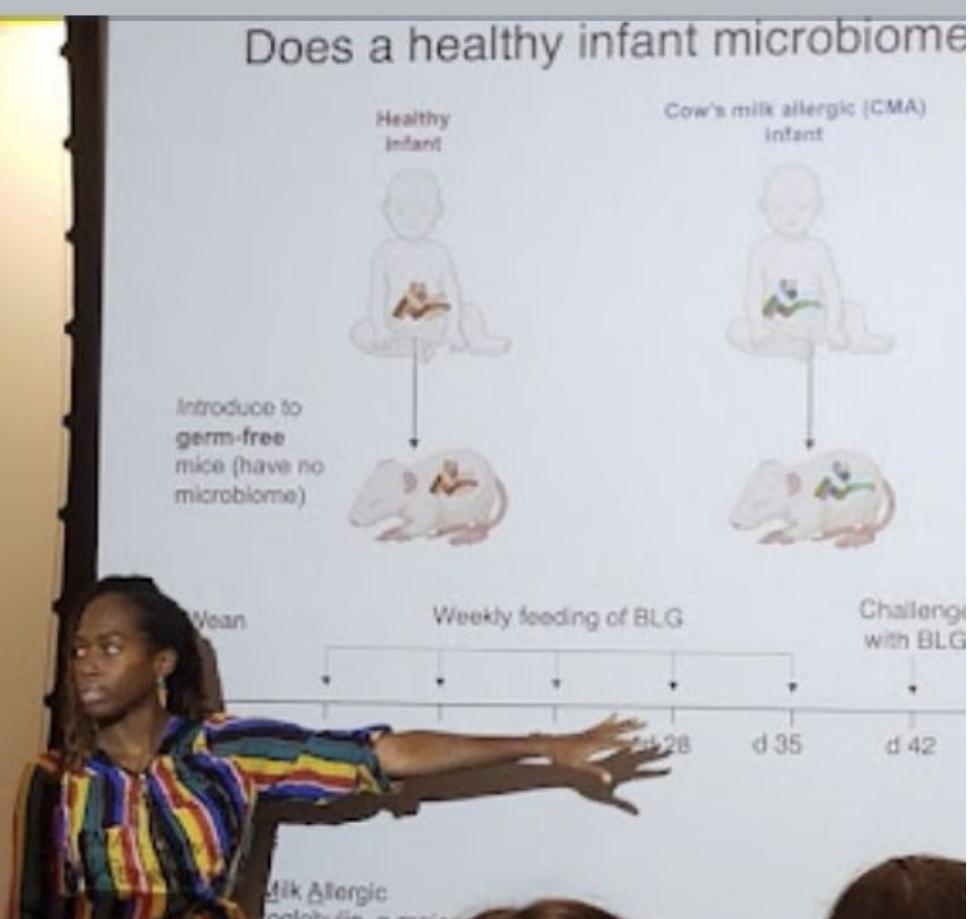
**Northside College Prep Junior with prior programming experience:** "The highlight of my time in the program was all the fun games we played and the friends made both within my table and with the rest of the group. I really enjoyed how we were able to learn in an exciting environment and how we were able to move around a lot instead of sitting in a chair all day."



# THE GUEST SPEAKERS

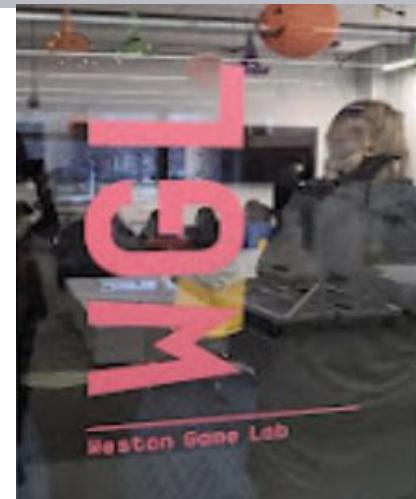
Guest speakers discussed data science in the context of college and career options

*Dr. Evelyn Campbell explains how she used data science in her microbiology dissertation*



*Ashlyn Sparrow: From video game playing to designing games*

*Dr. Amy Kirby (CDC) about data analysis with bacteria to find contaminated water that is causing diseases*



# Guest speakers connected data science to STEM, college & career options

## 2022

**Data science and microbiology – DNA sequencing for food allergies** | Evelyn Campbell, PhD (Preceptor/Postdoc at Data Science Institute)

**Using scientific reasoning to trace the cause of a PAM brain infection** | Amy Kirby, PhD (Microbiologist at CDC)

**Making it to and through college** | Dom McKoy (Deputy Director, UChicago To & Through)

**From gaming to game dev: A field trip to the Weston Game Lab** | Ashlyn Sparrow, Assistant Director, UChicago Weston Game Lab

**Summer internship and college prep overview** | UChicago Promise and future opportunities at DSI + Argonne

We invited the following guest speakers in 2022 and 2023 to help broaden students' understanding of what they can do with data science and how it can help them prepare for careers or undergraduate studies in data science or related fields:

## 2023

**Playing the Epi game & a field trip to the Weston Game Lab** | Ashlyn Sparrow, Assistant Director, UChicago Weston Game Lab. *She developed the epi game where one student plays a contagious disease and other students collaborate to optimally place health clinics across Chicago to prevent the spread of the disease.*

**Data science and microbiology – DNA sequencing for food allergies** | Evelyn Campbell, PhD (Preceptor/Postdoc at Data Science Institute). *Included her personal journey from high school to completing her dissertation.*

**Internships at Argonne** | Azucena Rodriguez

**Summer internship and college prep overview** | UChicago Promise: Brandon McCallister



# Students found guest speakers' connection of data science to college & career (very) useful

Feedback from students:  
How useful are the guest speakers?

**2022** Mid  
20 students



**2023** Final  
21 students



**Northside College Prep junior with prior programming experience:** "I think a highlight for me was the day that we played the infection board game and toured the game lab on the first floor."

**Robert Lindblom Math and Science Academy junior with little Python experience:** "My highlight was when the Doctor came in, I forgot her name but she was a huge inspiration for me because she said how she struggled in math and I also struggle in math and she talked about how she almost wanted to give up but kept going."

**Lane Tech senior with AP CS experience:** [the highlight of the program:] "The microbiologist speaker. It was the most insightful."

# BACKGROUND



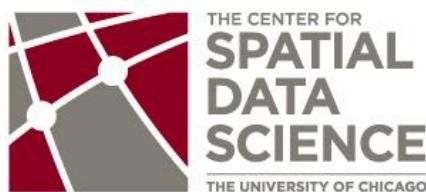
# About the Hosts



The Data Science Institute (DSI) executes the University of Chicago's bold, innovative vision of Data Science as a new discipline. The DSI seeds research on the interdisciplinary frontiers of this emerging field, forms partnerships with industry, government, and social impact organizations, and supports holistic data science education. It partners with the UChicago Department of Computer Science, where it is located, and the UChicago Department of Statistics.



Argonne is a multidisciplinary science and engineering research center, where talented scientists and engineers work together to answer the biggest questions facing humanity, from how to obtain affordable clean energy to protecting ourselves and our environment. Through its middle school to graduate programming, Argonne Education connects with over 30,000 youth and families at Outreach events, hosts over 4,400 middle and high school students at the Learning Center, and employs over 900 undergraduate & graduate students each year.



The team at the Center for Spatial Data Science (CSDS) thinks spatially about research problems. CSDS develops state-of-the-art methods for geospatial analysis; and applies them to policy-relevant research in the social sciences. It developed the open spatial software GeoDa, which has been downloaded over half a million times, and hosts spatial analytics lectures with over 670,000 views on its YouTube channel. It hosts a larger research project on integrating scientific reasoning with spatial data science: <https://puttingscienceintodatascience.org/>



The Office of Civic Engagement (OCE) connects the University to the city and the South Side. OCE stewards the University's commitment to the city, supporting other academic and administrative units as they develop and advance their distinct civic priorities. It also leads areas of strategic work that extend the University's reach and impact to the city of Chicago and the South Side.

# About the Surveys



This report is based on the results of 3 student surveys conducted with Google forms in each of the 2022 and 2023 workshops before, in the middle, and after the workshop. It also draws on feedback from near-peer mentors and instructors obtained in three meetings after the workshop. Because we revised the survey questions for 2023 to take the logic model into account that we developed after the 2022 workshop, in some cases there is no direct comparison between the 2022 and 2023 results.

The stars correspond to Likert scale ratings such as: Strongly agree (5 stars), slightly agree (4), agree (3), disagree (2), slightly disagree (1), strongly disagree (0). Degrees of usefulness or confidence are alternatives.

## **1) Baseline survey before workshop (N=24 in 2022 and N=26 in 2023)**

All students completed this survey before the start of the workshop (Sept. 2022 and March 2023). The survey asked questions about their demographics, schools, and experience with and confidence in computing, data analysis, and scientific reasoning.

## **2) Midpoint survey (N=20 in 2022 and N = 24 in 2023)**

A slightly smaller number of students completed the survey in the fourth week of the 8-week workshop due to absences. It contained questions about Data4All's mentors, teachers, activities and levels of confidence in workshop-related skills. Since a larger number of students completed the midpoint than final survey in 2022, many of the 5-point scale results in this report are based on midpoint results for 2022 (which tended to be consistent with the final results).

## **3) Survey at workshop end (N=15 in 2022 and N=21 in 2023)**

The smallest number of students answered the final survey's questions in the last workshop (week 8). We saw some attrition on the last workshop day in both workshops (an issue we plan to address in future workshops). Most of the questions were the same as in the midpoint survey, with a few additions, e.g. comparing students' understanding of data science before and after the workshop.

# Open Teaching Materials

All of the teaching materials are being peer reviewed and revised by Damaris Hernandez in summer 2023. She earned a BS and MEd from UC Berkeley, has extensive expertise in math and data science, and works as a math middle school teacher at Sayre Language Academy in Chicago.

The revised materials will be accessible openly and for free in fall 2023 here:

<https://github.com/uchicago-dsi/2023-data4all>

Please contact us at [spatial@uchicago.edu](mailto:spatial@uchicago.edu) with any questions.



# Expected Short and Long-Term Outcomes & Impacts of Data4All

What results we hoped for **in the short term** ...

## Understanding: What is data science?

- Students see CS & data science in a larger context of scientific reasoning & problem solving with data
- Students who might not major in CS find new relevance and broader appeal in CS and data science
- Students realize broader connection of CS & statistics to STEM fields like microbiology & env. science

## Skills: Programming, statistics, mapping & reasoning

- Programmers add skills in reasoning with data while others (like debaters) add programming skills

Students' confidence increases in:

- analyzing data with Python & Kepler (mapping)
- understanding foundational statistical concepts
- using both to critically explore causes of problems

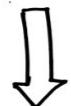
## College & career preparedness

- Students realize that you can use data science in lots of different fields that match their interests
- Students find Data4All useful for clarifying data science-related college & career goals

... and in the **longer term**

Data4All students:

- apply for and are admitted to DSI's Summer Lab
- are aware of and utilize OCE's college and career preparation resources to increase their chances of successful transition to college and jobs
- are inspired by the workshop to advance their data science skills, e.g. through additional courses or online learning
- Enroll in advanced coursework in high school that is related to data science (CS, stats, AP science)
- continue to practice their critical reasoning skills with data, statistics, and computing in their fields of interest and in everyday life



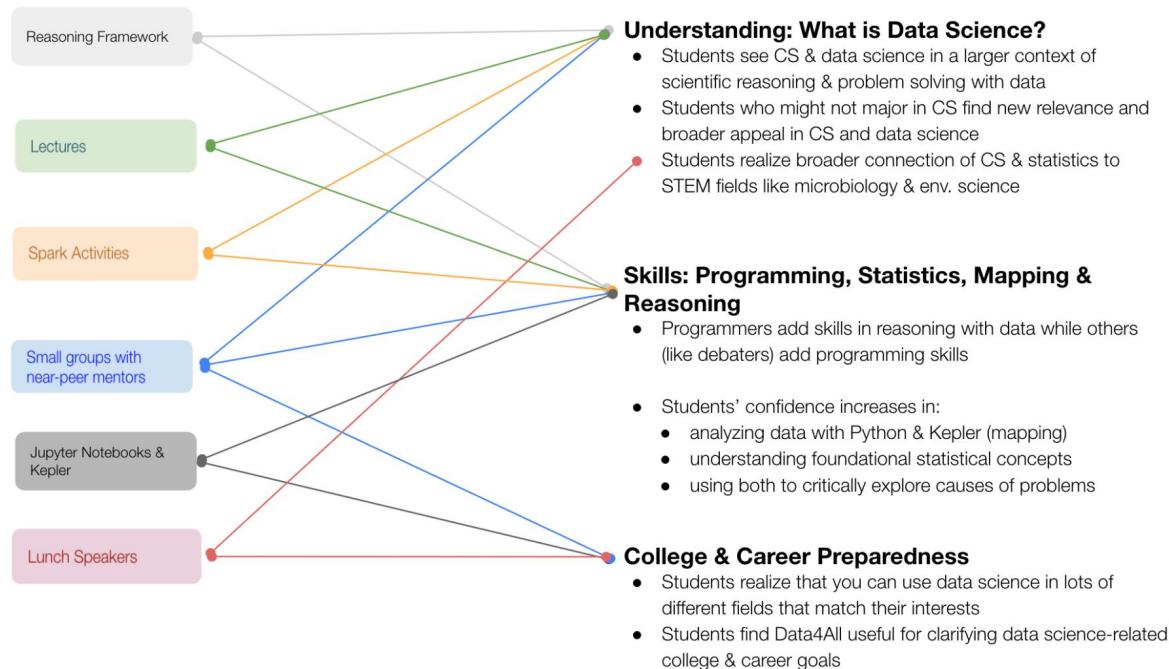
## Contribution to Long-Term Impacts

- More students currently underrepresented in data science choose data science as their college major and/or career
- A more informed citizenry that can reason with data and is less vulnerable to disinformation
- More educational offerings that integrate scientific reasoning with data science to unlock data science potential in students who had not considered data science



# Logic Model: Activities and their Expected Outcomes & Impacts

Logic Model Overview: Relation between  
**Activities** and **Expected Outcomes**



## Logic Model Details

ACTIVITIES	EXPECTED OUTCOMES		
	Understanding: What is data science?	Skills: Programming, statistics, mapping & reasoning	College & career preparedness
Reasoning Framework	Through the reasoning framework, students realize the connections between programming/statistics and other areas like debate, scientific inquiry and evidence-based decisions.	The reasoning framework centers the programming, stats and mapping lessons on real cases, so students see the relevance of the methods and learn how to reason with them (as opposed to applying methods mechanically).	
Lectures	Lectures help students see DS and stats in larger context by explaining key concepts in the context of case studies.	Lectures help students understand how the statistical and reasoning concepts work before they practice them in the small groups.	
Spark Activities	Spark activities help students realize the connections between statistics and real life, e.g. by answering a favorite question of who's the GOAT with statistical measures.	Spark activities improve students' intuitive understanding of statistical and scientific reasoning concepts so they have the big picture before they practice these concepts in more detail in the Jupyter notebooks.	
Small groups with near-peer mentors	Mentors address specific questions about how everything ties together for each student, e.g. how DS can be used for forensic analysis.	Mentors improve students' technical skills by addressing the specific programming, statistics, and mapping questions students have and show them how to debug coding errors.	Students are better prepared for college and jobs since mentors (who just started College) address students' questions about what College life and classes are like, how they wrote their admissions essays, accessed internships, etc. during breaks.
Jupyter Notebooks & Kepler	Jupyter notebooks help students see programming & stats in context since the coding is taught within a logic of problem solving with real data.	Jupyter notebooks & Kepler are the widely used technical practice grounds to help students improve their skills in programming, statistics and mapping while reasoning with real data.	
Lunch Speakers	Diverse women PhD speakers' case studies of DS in biology help students see how CS & statistics are used in STEM fields like microbiology & env. Science – and in game development.		Speakers from To & Through and UChicago Promise, ANL and DSI help students access college & career prep info & summer internships through concrete info about what opportunities are available where.

# In the News

2 Data4All students were interviewed by the local paper

Excerpt from:

## HYDE PARK HERALD

### KENWOOD ACADEMY'S PARTNERSHIP WITH ARGONNE PROMPTS VISIT WITH SECRETARY OF ENERGY

Aaron Gettinger, staff writer Nov 14, 2022

Skylar Spratt does Argonne's Data4All program, a partnership with the University of Chicago's Data Science Institute and Office of Civic Engagement. They do computer science data collecting during Saturday workshops. One of their first projects was an adaptation of epidemiologist John Snow's investigation of the 1854 cholera outbreak in London, which established that contaminated water was infecting people by mapping clusters of people suffering the disease near certain pumps.

In Data4All, Spratt and other students gathered information and made the hypothesis that cholera is waterborne, not airborne.

"We basically, starting from nothing — not having knowledge about cholera, not knowing about the situation itself — found and collected data, and then analyzed that data," she said. They looked at who was being exposed to the water sources, what areas of the city people who were being infected lived in and found commonalities between the infected people.

As Snow did, they plotted cases on a map; unlike him, they made the map a 3D representation on a computer to show how many people were getting infected.

Surrey Jones, another Kenwood student in Data4All, talked about a newer project about the way COVID-19 has affected different parts of Chicago and how whether or not people have health insurance affected them during the pandemic.

"In different communities, they have different resources that others don't," Jones said. "The West Side might not have the same amount of resources that the South Side has, so there would be a lot more cases on the West Side because they don't have those resources."

Jones said Data4All has strongly affected her career interests, in terms of exposure to different kinds of engineering and the way new technologies are changing the world.

"I really didn't think that data was a big deal. I did not think it was as big as it is, so this has definitely helped me," she said. "And it's actually pretty fun exploring. It gives you a lot more knowledge as to what's happening in the world."

Spratt likes engineering's problem-solving aspect. She compared it to debate, which she has also done, in terms of coming up with different points to support a hypothesis. She's done other engineering enrichment programs before, but she liked Data4All because it gave a sense of what she could be doing for a career.

At one point interested in civil engineering, Spratt thinks she's more into computer science now, given how much she enjoyed working with data.

Source: <https://tinyurl.com/3ntdk6tr>



# In the News

2023 Workshop  
featured in article  
by UChicago's  
Social Sciences  
Division

THE UNIVERSITY OF CHICAGO | Division of the Social Sciences

INFO FOR Current Students Faculty Staff Alumni

Search

PROGRAMS OF STUDY THE FACULTY ADMISSIONS ACADEMIC DEVELOPMENT PROFESSIONAL DEVELOPMENT RESEARCH DEVELOPMENT ABOUT

Home > News > Data4All Workshop Introduces High School Students Data Science Research

## Data4All workshop introduces high school students to data science research

May 3, 2023

Sociology | Faculty | Student | Announcements

f o in t G+ e

The collaborative effort between the Center for Spatial Data Science (CSDS), the Data Science Institute (DSI), Argonne National Laboratory, and the Office of Civic Engagement opens high schoolers up to the "thrill of discovery" of using data science to solve research problems.

By Sarah Steiner

Right now, 24 individuals are conducting inquiry-driven research. They're learning Python, a popular programming language for data science; training on how to investigate data problems and present findings; and developing critical collaboration, problem-solving, and communication skills — all under the leadership of UChicago's Data Science Institute, Argonne National Laboratory, the Center for Spatial Data Science, and the Office of Civic Engagement. They're also all high schoolers.



High school students learn the workshop content in 6 small groups of 4 students each.

The program, called Data4All, is a bridge workshop to introduce Chicago high school students to data science research and inspire them to consider careers in data science and AI. It addresses three main gaps: only a third of students have been exposed to data analysis or data visualization in their schools; many data science students struggle to critically solve problems with data; and there is a need for more data scientists in general, and especially from underrepresented groups.

"Data4All is critically important to DSI's efforts to broaden participation and increase equity and access to cutting edge data science education," says David Umnisky, executive director of DSI. "We are fortunate to have such outstanding experts to partner with in the CSDS, ANL, and OCE."

Currently in its third round, Data4All helps high school students see how data science is relevant to their interests and to critical problem-solving. Not only does the eight-week program improve their technical DS skills, but it also connects them with college and career preparation and internships at UChicago and Argonne.

"We know that data-centered skills and perspectives are becoming increasingly valuable in college and the workforce," says John Dominyancic, Manager of STEM Education Learning at Argonne National Laboratory, who leads the curriculum development and instruction of the program. "Data4All has allowed us to envision what data science education could look like at the high school level. We hope to take what we learn from Data4All and expand on its successes to empower more students, especially ones from underserved communities."

The program was developed to include gamified statistical concepts, small group activities, and field trips. For example, students have visited Ashlyn Sparrow, Assistant Director of the Weston Game Lab. Sparrow developed a spatial optimization game called Infection City that lets students experiment with where to put clinics to minimize STDs. Students played the game prior to their visit, where Sparrow explained how technology and reasoning are used to design these games.

"Students can really get an intuitive sense of how these concepts work and get some practice," says Julia Koschinsky, who is on Data4All's curriculum development team and operationally directs the Center for Spatial Data Science, with faculty direction by Luc Anselin. "Learning how to analyze data is more like skiing. As opposed to just memorizing something, you really need to get the practice down."



Dr. Evelyn Campbell explains how she used data science to tackle the big data problems in her microbiology dissertation.



Districted participants discuss how to address one of the challenges in their programming activity.

For CSDS, Koschinsky says the workshop is part of a larger project (<https://puttingscienceintodatascience.org/>) to determine whether data science or spatial data science could be taught to high schoolers and undergraduates within a scientific reasoning framework. Does that engage students? Does it lead them to think more critically about solving problems with data? Does it broaden their understanding of what data science is from programming to quantitative reasoning and scientific analysis? And does it get students away from a mechanical application of statistical tools and methods, toward a thrill of discovery reminiscent of solving a mystery?

"Can we get to those outcomes if we don't just teach programming, but teach programming in the context of scientific reasoning?" Koschinsky asks.

The answer appears to be yes. From an internal evaluation report of last fall's program, which included quotes from students, it's clear that the connections between data and social science are clicking. Not only did students speak — unprompted — of how they were grateful for a greater understanding of data science, but they appreciated seeing how it's applicable in their own lives or potential careers.

The high schoolers are also mentored by UChicago students, such as Doug Williams — who himself entered the university as "not a math kid," he says. But a work project after his freshman year helped him realize both his passion and capacity to pursue data science, and he saw Data4All as an experience that could invoke a similar passion for high schoolers. He sees his role as a mentor in the program as helping to tackle any technical skills needed for students to pursue their ideas.



Doug Williams, who double-majors in data science and public policy at the College, mentors high schoolers in a small group.

"As the program progressed, students realized that good data science is driven by critical thinking and formulating strong questions," Williams says. "Once they came to see the intimidating math and code as nothing but tools to aid their exploration of a problem, they became super proactive about trying new computational or statistical methods in pursuit of their idea."

Data4All is also hitting its goals of engaging underrepresented groups: 90% of the students in last year's cohort were minorities or girls, and students hailed from 18 public schools across the Chicago region — students without prior experience in data analysis or data visualization who, after the program wrapped, reported increased confidence in these areas.

"It opened up students to how they approach data science in a way that also incorporates social, scientific, and humanities thinking," Koschinsky says. "And that was what kept me hooked and why I'm still engaged, because I feel like we're onto something here: This special sauce of integrating these fields is working. It's exciting."

Source:<https://socialsciences.uchicago.edu/news/data4all-workshop-introduces-high-school-students-data-science-research>

# Acknowledgements

**Funding:** Hymen Milgrom Supporting Organization, UChicago Office of Civic Engagement, Data Science Institute

**In-kind contributions:** Data Science Institute (DSI), Argonne National Lab (ANL), Center for Spatial Data Science (CSDS)

**Collaborators:** UChicago Office of Civic Engagement, Peter Vinten-Johansen (historian | cholera case) and Kevin Credit (data and research | covid case), Jeanne Century (program & logic model)

**Curriculum Developers:** John Domyancich (ANL), Tyler Skluzacek, Julia Koschinsky (CSDS), Bethany Frank (ANL)

**Instructors:** John Domyancich, Kelly Sturner, Jacqueline Otmanski, Azucena Rodriguez (all ANL)

**Mentors (2022):** Doug Williams, Eliza Migdal, Anna Poon, Erin Ku, Pratham Gandhi, Arnav BrahmaSandra (all UChicago), Evelyn Campbell, Susanna Lange, Amanda Kube (all DSI)

**Mentors (2023):** Doug Williams, Arnav BrahmaSandra, Disha Mohta, Ziyu Ren, Victoria Kielb, Yurou Li (all UChicago), Evelyn Campbell, Susanna Lange (both DSI)

**Logistics:** Jessica Sweeney and Katherine Rosengarten (2022), Maria Victoria Fernandez and Shwetha Srinivasan (2023) (all DSI)

**Speakers:** Evelyn Campbell (DSI), Amy Kirby (CDC), Ashlyn Sparrow (UChicago Weston Game Lab), Dom McKoy (To & Through), UChicago Promise, Brandon McCallister (UChicago Admissions)

**Photography:** Joe Sterbenc: all high-resolution photos (buzzpictures.tv) and Julia Koschinsky: all others (CSDS)

**Teaching Material Revisions:** Damaris Hernandez (peer review) and ANL's Creative Services Team (graphic design)

Dr. Evelyn Campbell explains how she used data science in her microbiology dissertation

