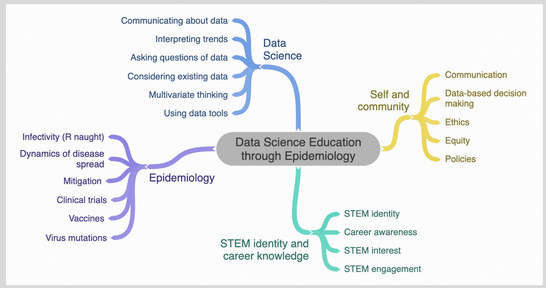
Data Detectives Clubs: COVID-Inspired Data Science Education through Epidemiology

Jan Mokros (jmokros@scieds.com), *Science Education Solutions*;
Penny Noyce (PendredNoyce@gmail.com) & **Laura Martin** (lmartin39@cox.net), *Tumblehome Learning*

Objectives

A collaboration of data science educators, data tool developers, a children's book publisher, and a science lab has created a 20-hour "Data Detectives Club" for middle school youth, focusing on historical and current pandemics. Goals are:

- To develop youths' skills in using digital data tools to track the spread of infectious disease and to ask and address their own questions of epidemiological datasets.
- To foster a deeper interest in and awareness of epidemiology and related STEM careers.
- To have youth become more fluent in using data (rather than opinion) to make sense of the COVID pandemic.



Framework

The program is anchored by a science-based adventure book, *The Case of the COVID Crisis*, and incorporates data activities, podcasts, animations, and digital data tools, as well as visits by epidemiologists.

- At each session, youth discuss one of the book's chapters, focusing on ethics, equity, and policy issues related to how infectious diseases spread and are treated.
- Youth engage in data activities using CODAP (Common Online Data Analysis Platform–Concord Consortium, 2022) and NetLogo (2020) to understand growth rate of infections, variability, rate of vaccination, and predictive models.
- Youth produce graphs, maps, and diagrams as well as data-based recommendations.

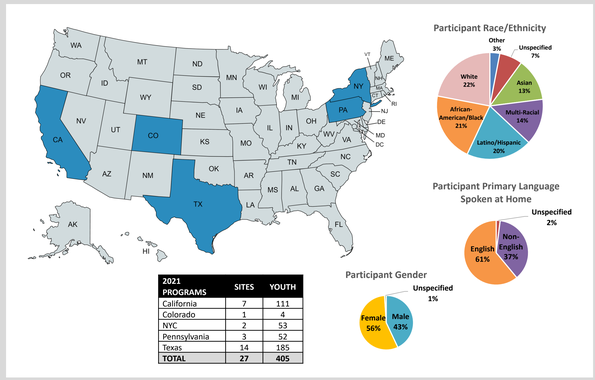
This project is funded by the National Science Foundation, grant no. DRL-2048463. Any opinions, findings and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Perspective

Epidemiological datasets are socio-technical tools entering everyday usage (Pea & Cole, 2019). Our research is based on cultural-historical activity theory (CHAT) (LCHC, 2010) which links the influence of direct interactions to broader contexts and intuitive concepts to scientific ones. The work is also based on emerging research in data science education, with a focus on asking and addressing questions about secondary datasets.

Participants

Participants are 400+ middle school youth (ages 9–14) from diverse, low-income communities in several states. They are enrolled in summer and afterschool programs offered through the YMCA, Boys and Girls Clubs, Girls, Inc., or 4-H that are part of the Imagine Science network, which supports STEM programming at these agencies. These youth are underserved with respect to STEM.

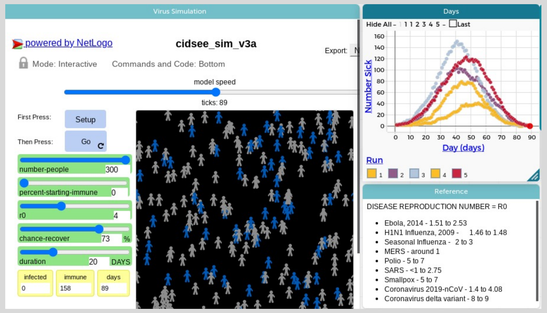


2021 PROGRAMS	SITES	YOUTH
California	7	111
Colorado	1	4
NYC	2	53
Pennsylvania	3	52
Texas	14	185
TOTAL	27	405

Sources of information

Three major types of data are collected from the youth:

- Surveys developed by project partner PEAR that measure changes in STEM interest, career awareness, and identity.
- Focus groups interviews.
- Youth-generated artifacts (CODAP graphs, data visualizations, and recorded final project presentations). *Images below show youth artifact screenshot examples of: CODAP/NetLogo disease spread simulation (left); New COVID cases (right, top) and fully vaccinated people in Orleans Parish, LA through December 2021 (right, bottom).*



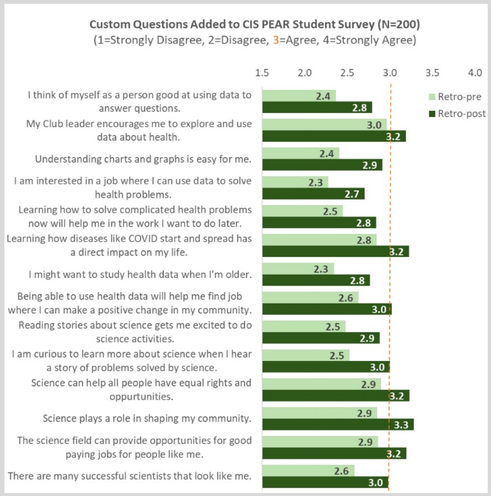
Results

- Retrospective pre/post surveys from summer and fall 2021 showed significant increases in STEM engagement, career awareness, and identity. *Image below shows analysis of custom survey item results.*
- Focus groups showed that youth felt that information and activities on virus mitigation and the associated fields/jobs in science and technology resonated the most with their own set of personal interests. They gained new knowledge about the types of careers used to examine questions in epidemiology and other fields related to STEM and disease management.
- Analysis of the graphs and presentations showed that youth learned how to use CODAP to construct, read, and compare data-based graphs. They also made many policy recommendations, but these were not always based on data.

Significance

Our project, lying at the intersection of data science education and epidemiology, promotes engagement with critical yet largely unfamiliar aspects of STEM. Currently, neither epidemiology nor data science play more than a small role in national science or math curricula. Specifically:

- We demonstrate the potential of afterschool and summer programs for immersing underserved youth in contemporary, relevant, and cutting-edge STEM.
- We show the power of narrative to convey scientific ideas and motivate youth to learn more about epidemiology.
- We examine how youth master the use of data tools when used in a compelling context.



References:

Concord Consortium. (2022). Common Online Data Analysis Platform (CODAP). http://codap.concord.org

Laboratory for Comparative Human Cognition (LCHC). (2010). Cultural-Historical Activity Theory. http://lchc.ucsd.edu/Pubs/lchc.encyc.pdf

NetLogo. (2020). https://ccl.northwestern.edu/netlogo

Pea, R. & Cole, M. (2019). The living hand of the past: The role of technology in development. *Human Development, 62* (no. 1-2), pp. 1-3.

imagine science | INSPIRING THE NEXT GENERATION OF STEM LEADERS
Boys & Girls Clubs of America • Girls Inc. • National 4-H Council • Y-USA

SLP4i STRATEGIC LEARNING PARTNERS FOR INNOVATION

pear

DATA DETECTIVES CLUB

The Concord Consortium

The Jackson Laboratory