Data Detectives Clubs in the time of COVID-19

Paderborner Kolloquium
27ten Oktober 2021

Jan Mokros (jmokros@scieds.com)
Bill Finzer (wfinzer@concord.org)

Link to these slides: https://bit.ly/3C5Svyf

Our approach: Data Detectives Clubs

Introduce youth to data science & epidemiology through 20-hour out-of-school experience centered on an adventure book. The two anchors are the book and the data activities.

- Most chapters are accompanied by data activities
- Digital data activities are done in CODAP/NetLogo
- Social-emotional learning and public-health decision-making are incorporated
- Youth do project using epidemiological data to make help policy-makers make a decision
- Youth learn about epidemiology and data-based careers through the book and virtual visits

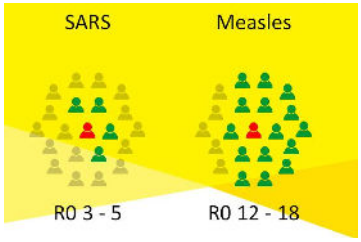
The adventure story as one anchor

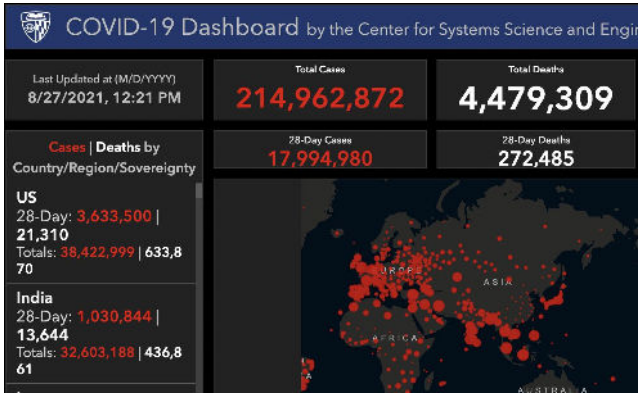
- Why a story? – Curiosity, connection, empathy
- What the story is: a journey through space and time
- The story is evolving
- Chance to address children's fears and concerns

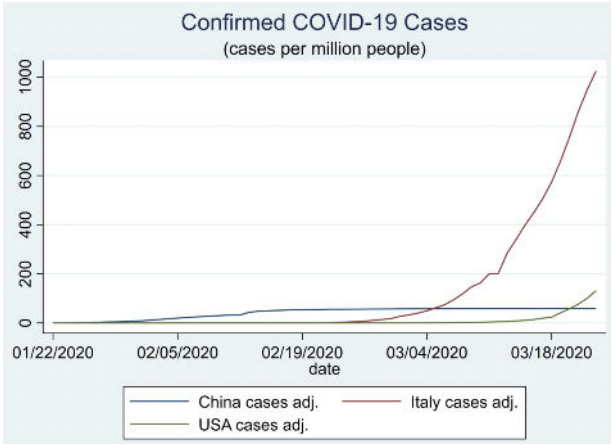
Data activities as the second anchor

- Data activities using CODAP accompany most of the chapters
- Data are about different pandemics (Ebola, smallpox, 1918 flu), but most of the data are about COVID.
- We start by posing questions
- Youth ask their own questions about patterns, relationships, and change over time.

Background



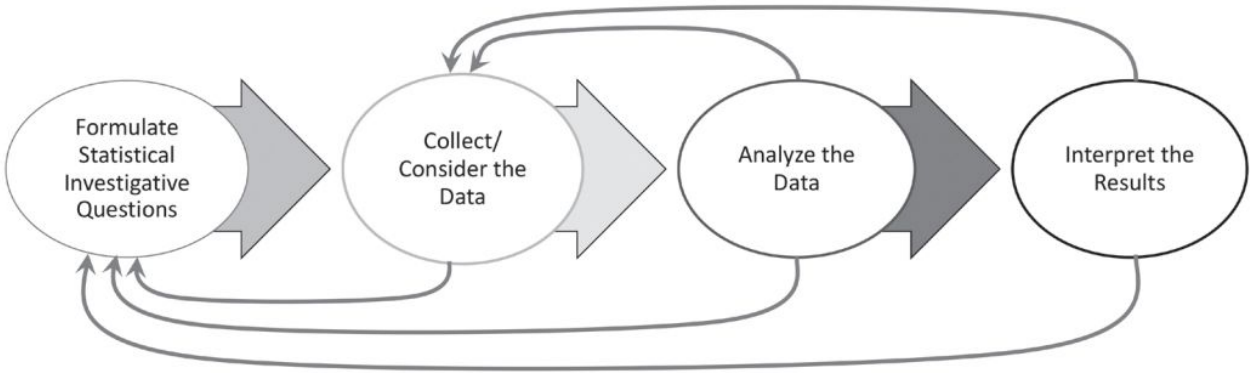




- Epidemiology is rooted in data science
- Epidemiology uses time series data, multivariate data, rates, probability, and R0
- American youth typically do not learn about either epidemiology (Bracken, 2014) or using large datasets (Finzer, 2013) in science classes

Data science framework

- Investigative process from Guidelines for Assessment and Instruction in Statistics Education II (GAISE II, 2021, p. 13)



GAISE II increased areas of emphasis (pp. 8–11)

- The role of technology in statistics
- Questioning throughout the statistical problem-solving process
- Different data and variable types
- Multivariate thinking
- Probabilistic thinking

Humanistic data science framework

(Lee, Wilkerson, & Lanouette, 2021)

- **Personal layer:** Direct experiences with designing and generating data, and with the phenomena and contexts.
- **Cultural layer:** Norms relating to how datasets are constructed and used, including the use of technology & cultural practices of different disciplines (in our case, epidemiology)
- **Sociopolitical layer:** Political and social forces that affect how data are constructed and interpreted.

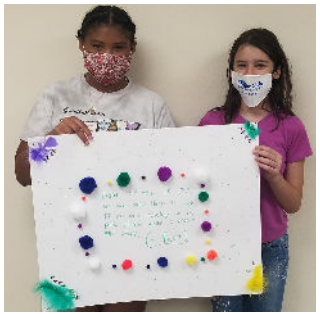
Mae and Clinton visit Dr. Anthony Fauci in Washington, DC

Implementation

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

Imagine Science is implementing the 20-hour Data Detectives Clubs program *(70% minority youth, 60% low income, 49% girls)*

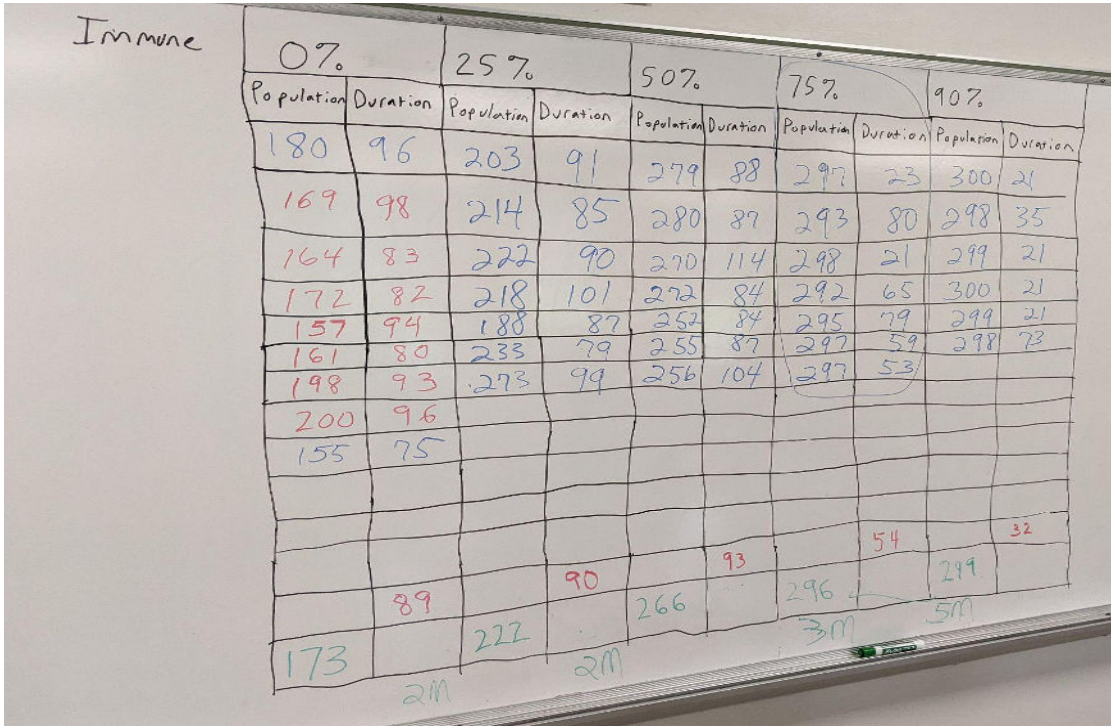
- To date: 30 clubs with 450 youth (ages 10 to 14) involved in 7 communities (6 US states)
- Leaders of clubs get 6 hours of professional development
- Youth work is collected, especially data-rich artifacts
- Post-club survey data about STEM and data science attitudes, dispositions, and identity is also collected
- Post-club focus groups are conducted with youth and with leaders



A Simulation of an Epidemic:
bit.ly/CCCODAPNetLogo

- Using NetLogo Web, a multi-agent programmable modeling environment from Northwestern University, US
- Embedded in the Common Online Data Analysis Platform (CODAP)
- Free, browser-based with no installation or login

Variation in simulations



Parameters used
Population at start: 300
Survival rate: 50%
Duration of illness: 21 days
R_0 (infectivity rate): 5

Learning Experiences Emerging from Working with a Stochastic Simulation with Data Output

- Noise!
- Search for a signal
- Play ⇒ Experimentation
- Variation within groups versus variation between groups
- Decisions about what to measure
- Transfer to real world

The Beauty of Data Portals: bit.ly/DDCODAPportal

- Some portals can be made to be always up to date by leveraging work of institutions like the US CDC
- Learners can choose from a very large dataset the portion of most interest to them
- Portals can be created for particular learning contexts (like the Data Detective Clubs)
- Simplicity can be a goal
- The richness of the data makes a given portal useful for multiple learning contexts

Research questions

1) What changes do youth experience with respect to STEM identity, STEM career interests and STEM engagement as a result of their participation? Special emphasis on the data science part of STEM.
2) How do youth use datasets/data tools to study infectious disease spread? How do they ask questions, examine patterns, use models, make predictions?

Our research on Question 2: “The Mission”

- How do youth use datasets/data tools to study infectious disease spread?
- Youth are given a mission directly from one of the characters in the book (Selectra Volt)
- Youth use data from the portals
- Data-based arguments for community health decisions

Challenges and questions for discussion

- A challenge is to make Club materials more data-rich while simultaneously engaging youth in the narrative and activities.
- Leaders need more time to learn about data and CODAP
- COVID data change as we are developing materials
- Will the project continue be relevant as COVID evolves?
- Does the project have cross-cultural relevance?

Danke!

This work was funded by the National Science Foundation under grant DRL-2048463. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Project partner logos:

Selected references

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education. National Council of Teachers of Mathematics (NCTM). American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Bracken, M. B. (2014). Epidemiology as a liberal art: From graduate school to middle school, an unfulfilled agenda. *Annals of Epidemiology*, *24*, 171–173. https://doi.org/10.1016/j.annepidem.2013.11.010

Concord Consortium. (2021). Common Online Data Analysis Platform (CODAP) [Computer software]. https://codap.concord.org/

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press. https://data-feminism.mitpress.mit.edu/

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, *7*(2). https://escholarship.org/uc/item/7gv0q9dc

Lee, V. R., Wilkerson, M. H., &, K. (2021, September 23). A call for a humanistic stance toward K–12 data science education. Educational Researcher. https://doi.org/10.3102/0013189X211048810

Noyce, P. (2021). The case of the COVID crisis. Tumblehome Books. https://tumblehomebooks.org/book/the-case-of-the-covid-crisis/