Designing Data Science Experiences for Middle School Youth

Andee Rubin, TERC February 14, 2019

TORC Because math and science build future

Overview

• Some History (Is there anything new under the sun?)

- O Data Clubs project
 - O Design Criteria
 - How it played out in one or two cases
 - O Instrument development
 - O Wild hair idea



History or "How my career became sexy after 30 years."

DATA ANALYSIS STATISTICS STATISTICAL REASONING DATA LITERACY DATA SCIENCE



X

Back in the Good Old Days (1989)



• "We introduce students to good literature in their early years. We do not reserve great literature until they are older – on the contrary, we encourage them to read it or we read it to them. Similarly, we can give young students experience with real mathematical processes rather than save the good mathematics for later. Through collecting and analyzing real data, students encounter the uncertainty and intrigue of real mathematics." (p. 1)



And in 2005...

ANDEE RUBIN, M.S., E.E. • • Senior Scientist, TERC





MATH THAT MATTERS

The case for replacing the algebra/calculus track with data literacy–a critical skill for modern life.

Have we made any progress?? (Now Data Science is part of Computational Thinking?)

TORC | Because math and science build futures

Data Clubs Overview

- Data Clubs is a STEM+C grant (thank you, NSF) with partners TERC, Science Education Solution, PERG, Gulf of Maine Research Institute, Oxford Hills School District, Malden YMCA, Girls Inc.
 - Data Science = Mathematics + Computation (for this project)
 - After school/summer camp materials, 10-12 hours/module
 - Participants entering 7/8 grade (12-13 years old)
 - Three modules, focusing on 3 different topics and large publicly available data sets.



Data Clubs Conceptual Foci

- Data are everywhere; turning "life" into data requires "measurement" decisions. Lehrer and Schauble include this in the process of "data modeling"
- O Data are organized into cases, each of which has multiple attributes.
- Any data set can be used to answer some questions and <u>not</u> others.



Design Criteria

O Topics
O Datasets
O Tools

O Activities



Criteria - Topics

- Working out of school means we have freedom to choose any topic and have to keep participants interested. The relevant statistical concepts emerge from the data.
- Both a "window" and a "mirror" plus potential for social action
- Topics chosen with help of Youth Advisory Group:
 - Teens and technology
 - Tick-borne diseases such as Lyme
 - Sports injuries or animals in animal shelters??



Criteria – Data Sets

- "Goldilocks and the Three Bears" story
- Looking for "just right" datasets that are small enough to be manageable, but have rich relationships





Criteria - Tools

- Accessible for middle school students
- Can deal with largish data sets
- Makes the difference between cases and attributes clear
- O Dynamically links multiple representations of the same data
- Makes filtering and looking at subsets of the data (data moves) accessible



Criteria - Activities

- Tim Erickson's criteria for "smelling like data science": "awash in data" and using "data moves"
- Participants should both analyze a large data set and have some experience collecting data.
- Participants should have experience representing data with computer tools and by hand.



Teens and Technology topic

- "Hook" the lead article in the Boston newspaper last year!
 - Forty-five percent of teens say they are online "almost constantly," according to a new Pew Research Center study on teens and social media use. That percentage has nearly doubled in just a few years: In a 2014-15 Pew survey, just 24 percent of teens said the same.

Pew shares findings on teen social media use





Teens and Technology dataset

- Pew Research Center Technology and Internet Use 2014 survey
- 37 page interview administered online to a parent/child combination
- O About 1200 cases
- 300+ variables
- Introductory dataset: 50 cases, 10 variables
 - Three thematic subsets: internet use, friends in the digital age, gaming, 200 cases each



Some of the Pew Questions

- Overall, how often do you use the Internet?
 - Almost constantly
 - Several times a day
 - About once a day
 - Several times a week
 - Once a week
 - Less often



More Pew Questions

- O On an average day, about how many text messages do you send and receive on your cell phone?
- Which of the following social media do you use MOST often?
 - Facebook
 - O Twitter
 - O Instagram
 - O Google+
 - Snapchat
 - O Vine
 - O Tumblr



And a few more...

• When you play games online with others, do you ever

- Feel more connected to friends you play games with who you already know?
- Feel more relaxed and happy?
- Feel more angry and frustrated?
- Feel connected to people you play games with, but aren't friends with yet?

Possible answers: (Yes, a lot; Yes, a little; No)



Tool: CODAP (Common Online Data Analysis Platform)

- CODAP, designed and implemented by Concord Consortium with many user-partners.
- Cases have physical presence and are linked across representations.
- O Show CODAP!



What participants explored this summer

O <u>Nadia's project</u>

O Josh's project

O <u>Arnie's project</u>



Instrument Development

- An interview, using CODAP, to see if a student can tell which questions can be answered from a data set and which can not.
- A survey or interview measure of "data dispositions."



x

Hair-brained Idea (which I love)

- Dear Data: postcards between two graphic artists who love data.
- Where else could we use this form of communication?











Dear Data

The challenge of creating a legend requires sophisticated thinking about mapping between attributes and representation.



week of phone addiction	Stefanie
ANN ANA ANA ANA ANA ANA ANA ANA ANA ANA	MAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
DEAR DATA-WEEK 08 A WEEK OF PHONE ADDICTION THE DATA: I TRACKED EVERSI TIME I PICKED UPDAME IN THE PICST INSTANCE (WITTHE HOW TO READ IT: THE WITTHE STATES) HOW TO READ IT: THE WITTHE STATES PARENCE INSTANCE (WITTHE THE WITTHE STATES) REASONS FOR PICKING UP PHONE CHILDING WEEK THE WITTHE STATES ON CALL BRANK CHILDING WEEK THE STATES ON CALL BRANK CHILDING THE STATES ON CALL BRANK CH	REAME S. FROME LABOON LA TO: GIORGIA LUPI BROOKLYN, NY USA BY AIR MAIL Dryal Mair

On Tuesday and Saturday Stefanie was mostly airborne and listening to music, as can be seen in her data drawing

image part with relationship ID rId2 was not found in the file.

Lessons Learned

- Because we are driven by topic, different datasets have different structures and present different challenges.
 - Teens and technology has many ordinal and categorical variables; tick-borne diseases involves changes over time and geographical trends
- It's sometimes tricky to balance "lies, damned lies, and statistics" and "data tell the truth"
 - Appropriate skepticism often requires a deep dive into the details



Thank You!

<u>Andee_rubin@terc.edu</u>

Let's talk more over lunch