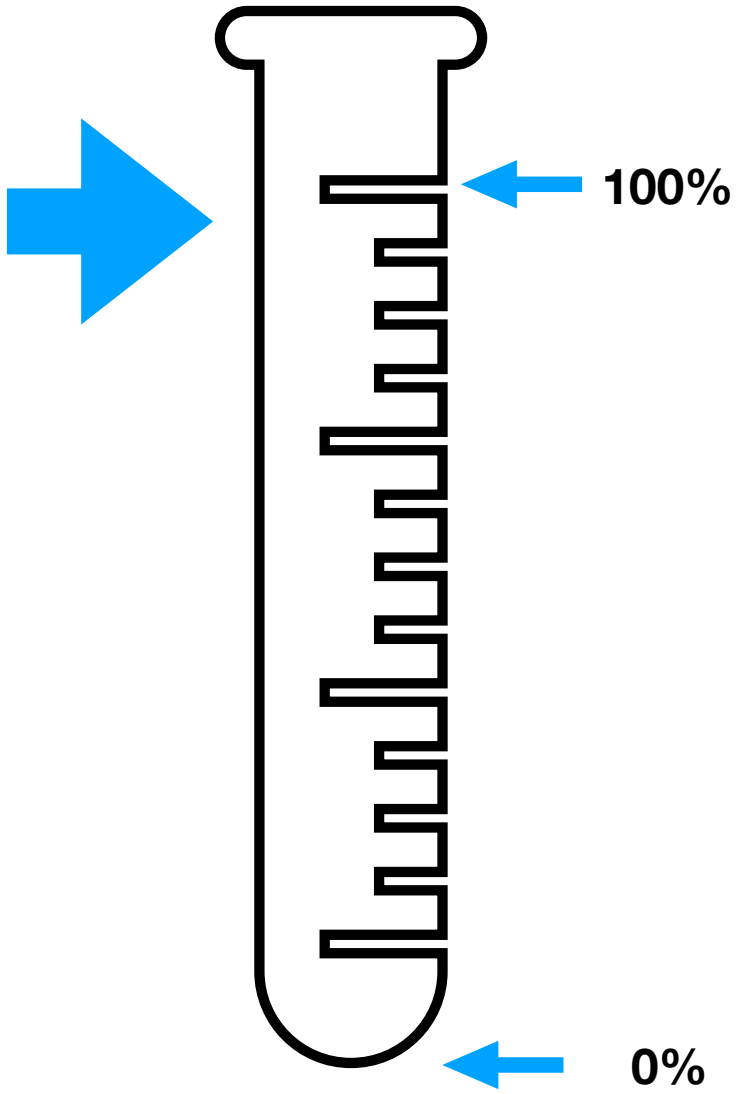
Citizen Statisticians Then and Now: Looking Back Over the Last 8 Years

Rob Gould
rgould@stat.ucla.edu

data challenge

- A student should be able to bring to intro stats any collection of data and by the end of class learn the tools need to discover knowledge otherwise unavailable.

Progress o' meter



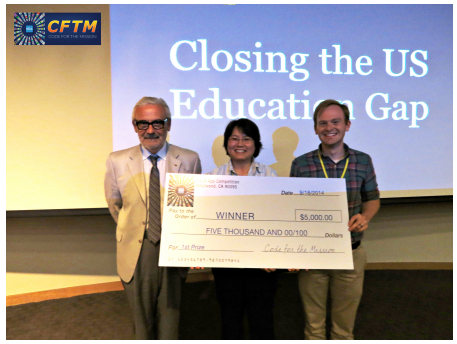
START

https://www.introdatascience.org/

mobilize

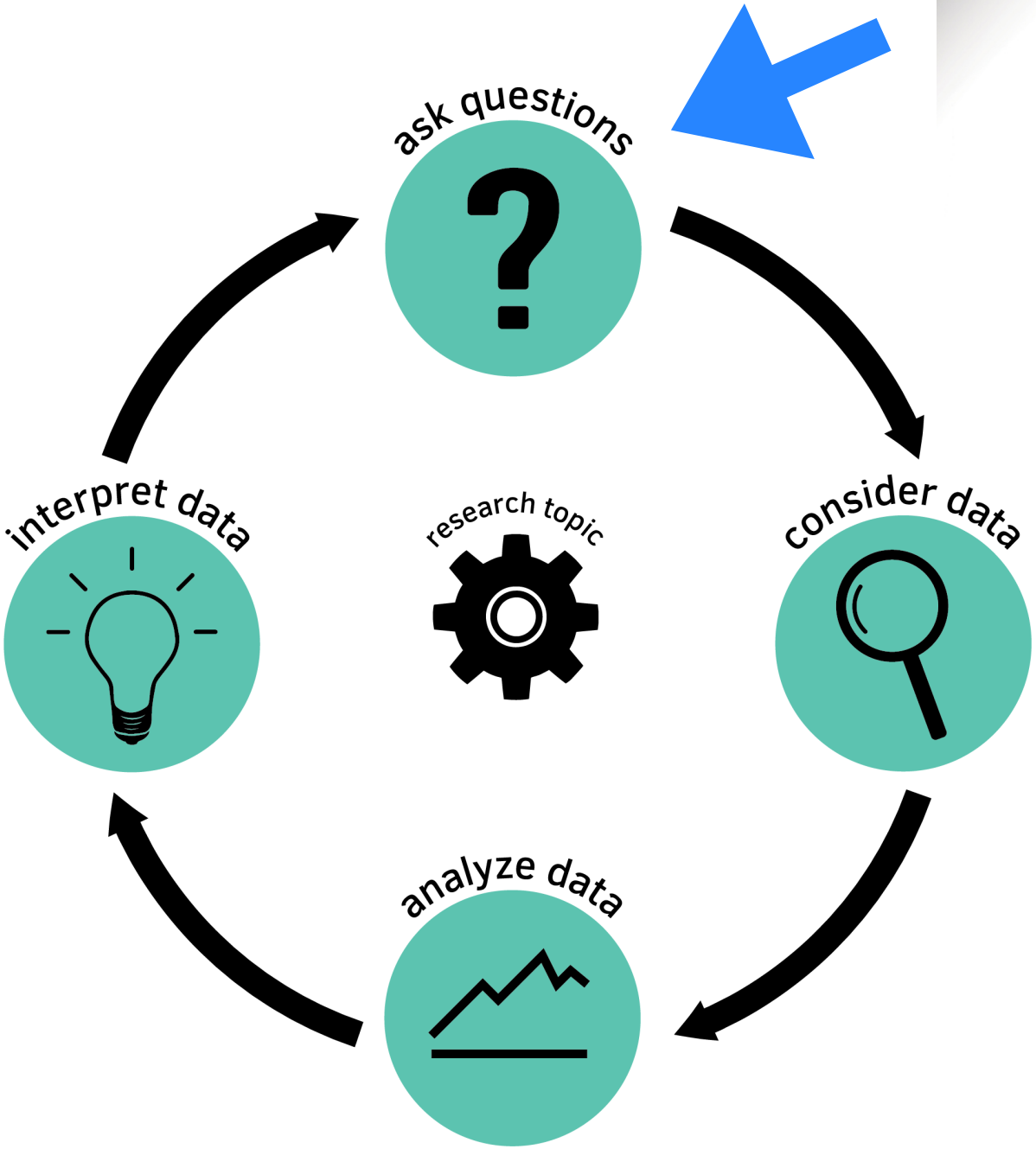
IDS

Introduction to Data Science



Suyen Moncada-Machado, James Molyneux, Amelia McNamara, Terri Johnson, LeeAnn Trusela, Hongsuda Tangmunarunkit, Steve Nolen

The Data Cycle



ASA DataFest

Intro Stats Project

- Phase I: Using a list of **vetted data repositories**, select a data set of interest to you. Down load it, and upload it into StatCrunch.

- Phase II: Join a team of 3 students, and choose one of your datasets. For this dataset, write a paper that describes the data: who what where when why how etc. Pose three interesting questions that you believe can be answered by these data.

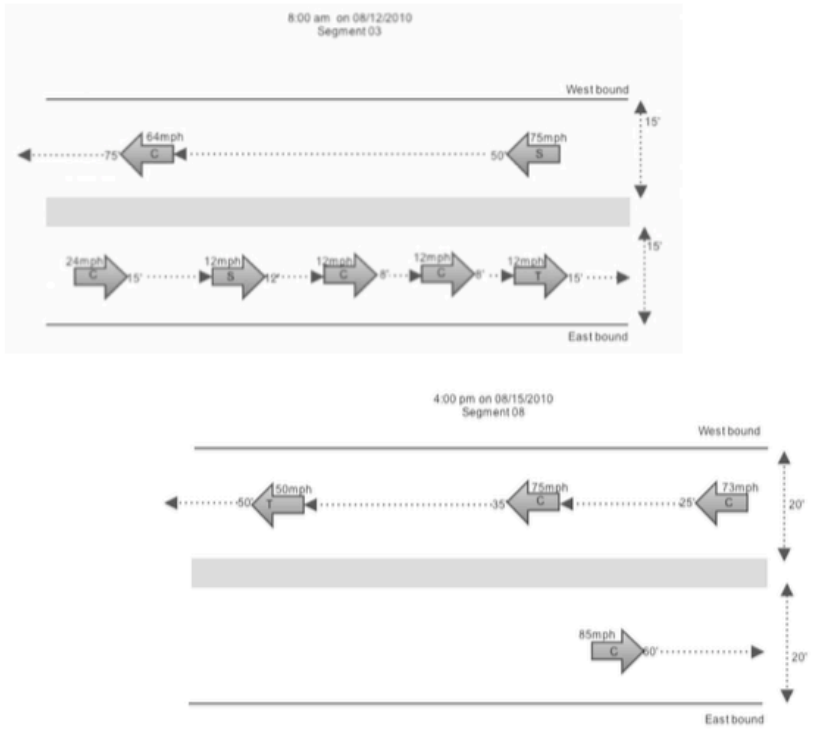
- Phase III: Answer the questions you posed or explain why a question cannot be answered.

Students' Issues

- I uploaded a date, but a strange "random" number was uploaded instead. (data formats: dates, strings, characters, floating)
- Are observations people, or records? (hierarchical structures)
- I got several files zipped together. Which one is the data? (file extensions, managing files)
- Which document do I download? The documentation says it provides only SAS/SPSS/Stata/ASCII, but not 'csv". (file extensions, managing files)
- What to do with fixed format? (data storage)
- I click "download all files" but nothing opens up. (file management)
- I can't upload a .tsv file. (file extensions)
- I need step-by-step instructions. (?)

Four Suggestions For Moving Forward

1. Understand what "data" means.



"Modeling as a Core Component of Structuring Data", Konold, Finzer, Kreetong, 2017, SERJ.

What do our students think?

- Bowler et al (2017*): What do students think the term “data” means? 2 camps: Data as a product of scholarly enterprise; Data in terms of the networked, digital world

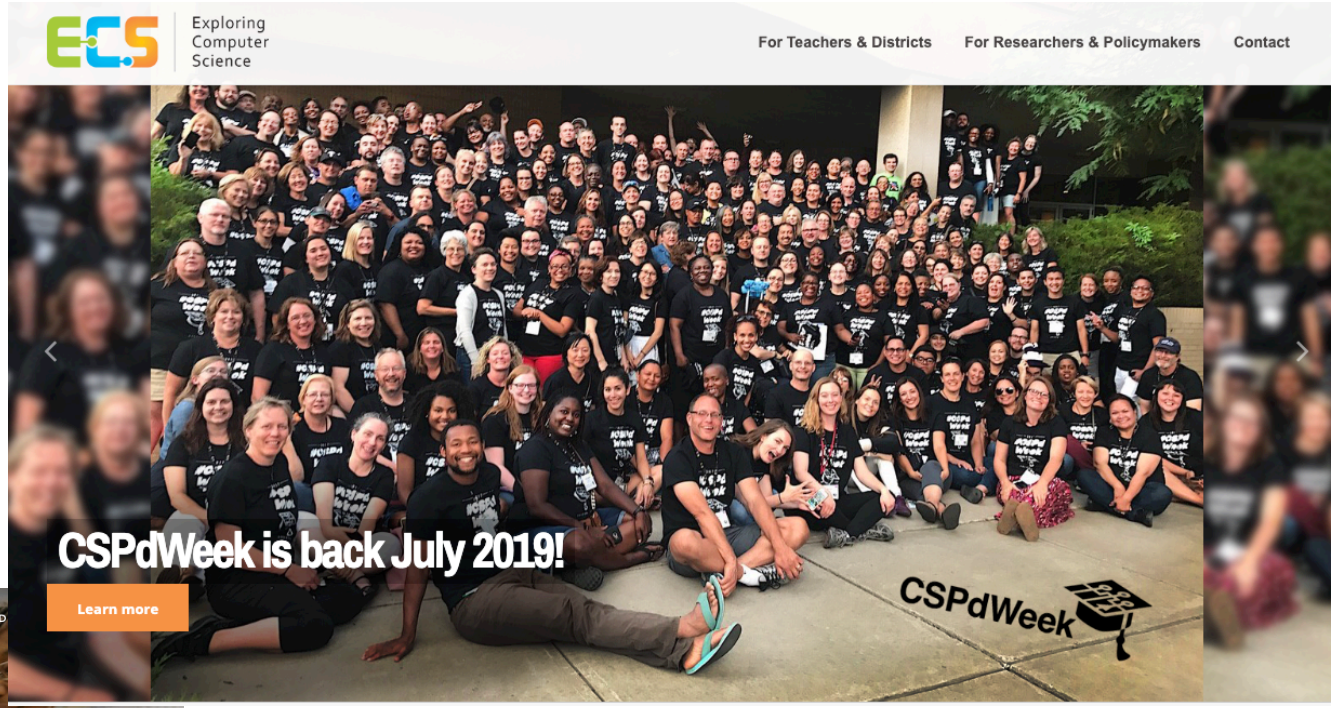
- "...the teens...had varying interpretation of the nature of data...but most found it difficult to connect with data at a concrete and personal level."

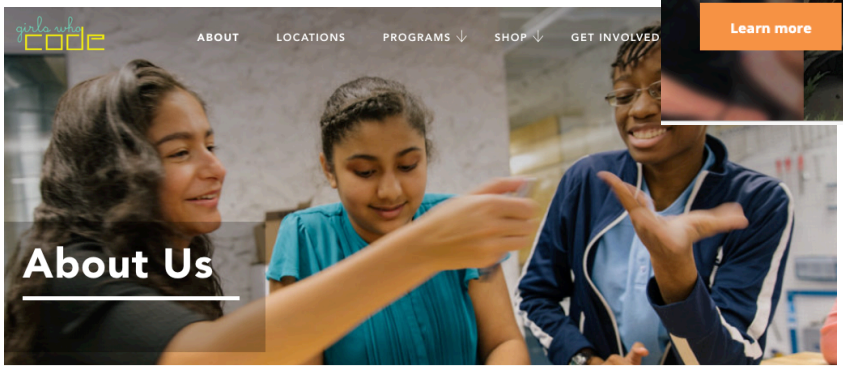
*** Proceedings from 80th annual meeting of association for information science & technology, Washington, dC, Oct 27-Nov 1, 2017**

Four Suggestions For Moving Forward

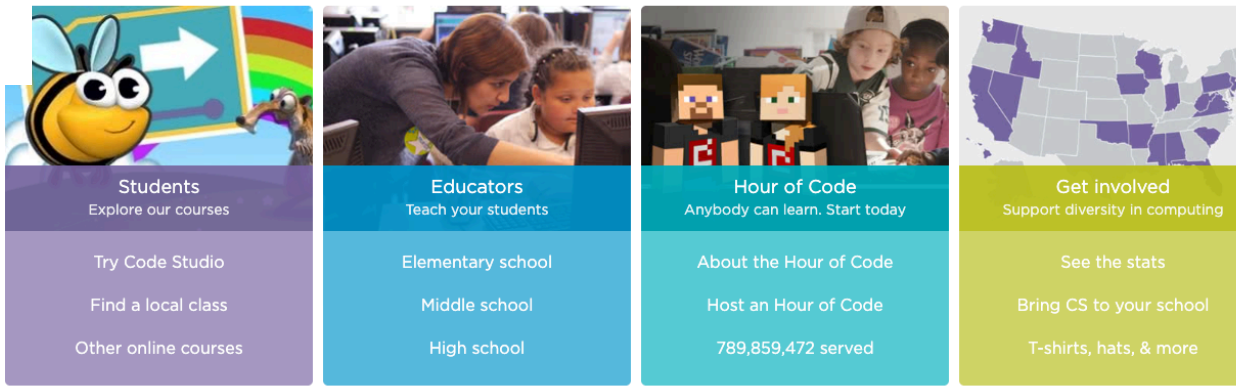
1. Understand what "data" means.

2. Become a Nuisance to Computer Science Educators.





Girls Who Code is on a mission to close the gender gap in technology and to change the image of what a programmer looks likes and does



CSTA Standards

1A-DA-05	K-2	Store, copy, search, retrieve, modify, and delete information using a computing device and define the information stored as data.	Data & Analysis	Abstraction
1A-DA-06	K-2	Collect and present the same data in various visual formats.	Data & Analysis	Communica Abstraction
1A-DA-07	K-2	Identify and describe patterns in data visualizations, such as charts or graphs, to make predictions.	Data & Analysis	Abstraction
1B-DA-06	3-5	Organize and present collected data visually to highlight relationships and support a claim.	Data & Analysis	Communica
1B-DA-07	3-5	Use data to highlight or propose cause-and-effect relationships, predict outcomes, or communicate an idea.	Data & Analysis	Communica
2-DA-07	6-8	Represent data using multiple encoding schemes.	Data & Analysis	Abstraction
2-DA-08	6-8	Collect data using computational tools and transform the data to make it more useful and reliable.	Data & Analysis	Testing
2-DA-09	6-8	Refine computational models based on the data they have generated.	Data & Analysis	Creating, Abstraction
3A-DA-09	9-10	Translate between different bit representations of real-world phenomena, such as characters, numbers, and images.	Data & Analysis	Abstraction
			Data &	Computatic

Four Suggestions For Moving Forward

1. Understand what "data" means.

2. Become a Nuisance to Computer Science Educators.

3. Develop assessment tools to better understand the role of computation in understanding statistical concepts and performing data analysis.

Some course goals

- Learn to think with and about variability
- Understand how to represent data in different structures for analysis
- Ask probing and productive questions
- Recognize when questions can be and should be answered with data, and whether the data are appropriate
- Understand the role of chance variation in measurements.
- Understand the scope of inference

Getting Started

FAQs
Frequently asked questions and some answers offered by the Research Advisory Board

Assessment Tools (for Research use)
ARTIST (Assessment Resource Tools for Improving Statistical Thinking):
Tests and measurement tools to use in research studies in statistics education.

Graduate Programs
Graduate program emphasizing statistical education:
University of Minnesota

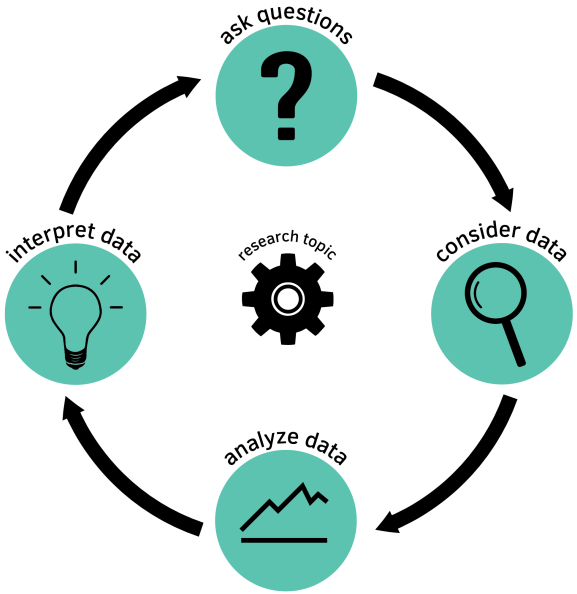
Statistics Education Theses
Graduate programs where students have completed a dissertation on teaching or learning statistics.

LOCUS

Levels of Conceptual Understanding in Statistics

https://locus.statisticseducation.org/

The Data Cycle



Browse questions by component

- Formulate Questions
- Collect Data
- Analyze Data
- Interpret Results

Four Suggestions For Moving Forward

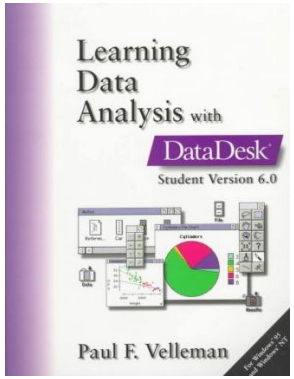
1. Understand what "data" means.

2. Become a Nuisance to Computer Science Educators.

3. Develop assessment tools to better understand the role of computation in understanding statistical concepts and performing data analysis.

4. Bridge the gap between technology for learning and technology for doing

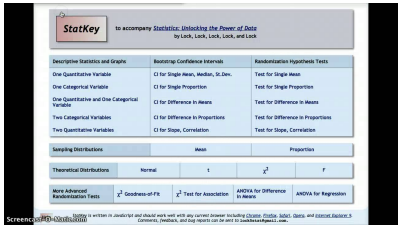
Learning



Fathom
Dynamic Data Software

TinkerPlots™
DYNAMIC DATA EXPLORATION

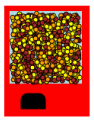
CODAP
Common Online Data Analysis Platform

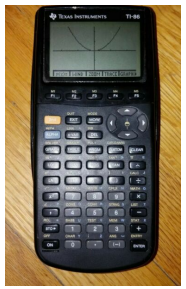


Rossman/Chance Applet Collection

Reeses Pieces

Probability of orange 0.5
Number of candies 25
Number of samples 1
Animate
Draw Samples
Total = 0





jmp
Statistical Discovery.™ From SAS.

güroo

Doing

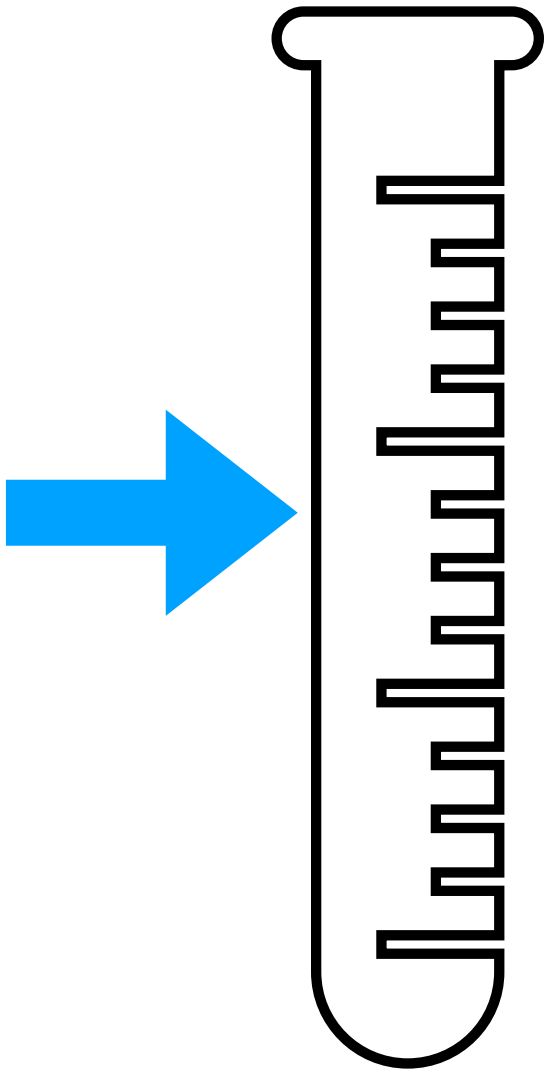
RStudio®

python™

SPSS®

STATA

sas



What I've learned:

- Ask Questions!
- Use the Data Cycle
- Give hard problems, but with a safety net.

Next Steps:

- Understand students' conceptions of data.
- Collaborate with CS educators (to the point of annoyance)
- Measure the interaction of computational and statistical thinking
- Bridge the gap.