Teaching & Learning with Geoscience Data

Kim Kastens

Work done at Education Development Center
& Lamont-Doherty Earth Observatory of Columbia University

Presented at:
Summit on Undergraduate Geoscience Education
8 January 2015, Austin, TX

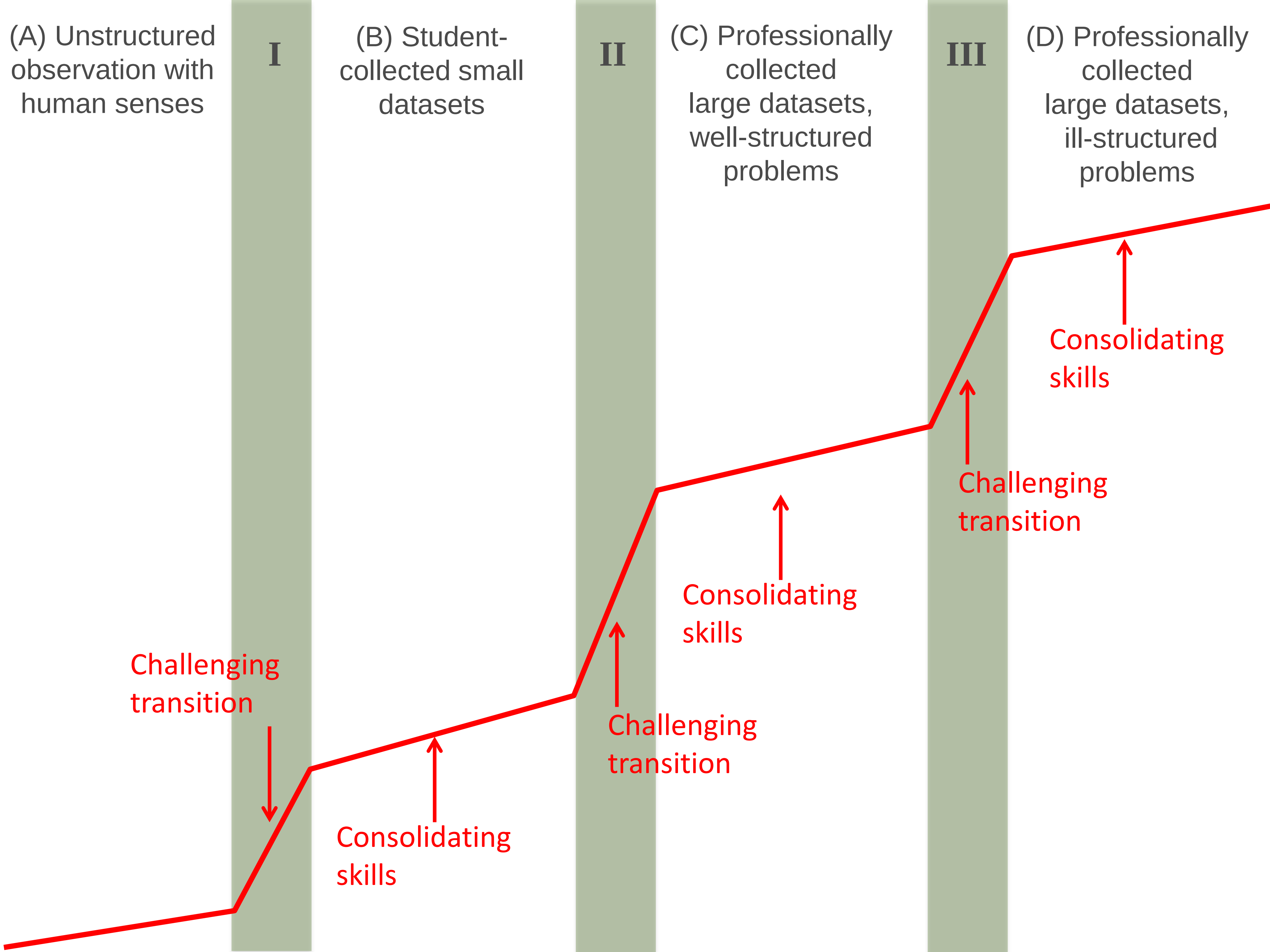
EDC OCEANS of DATA INSTITUTE

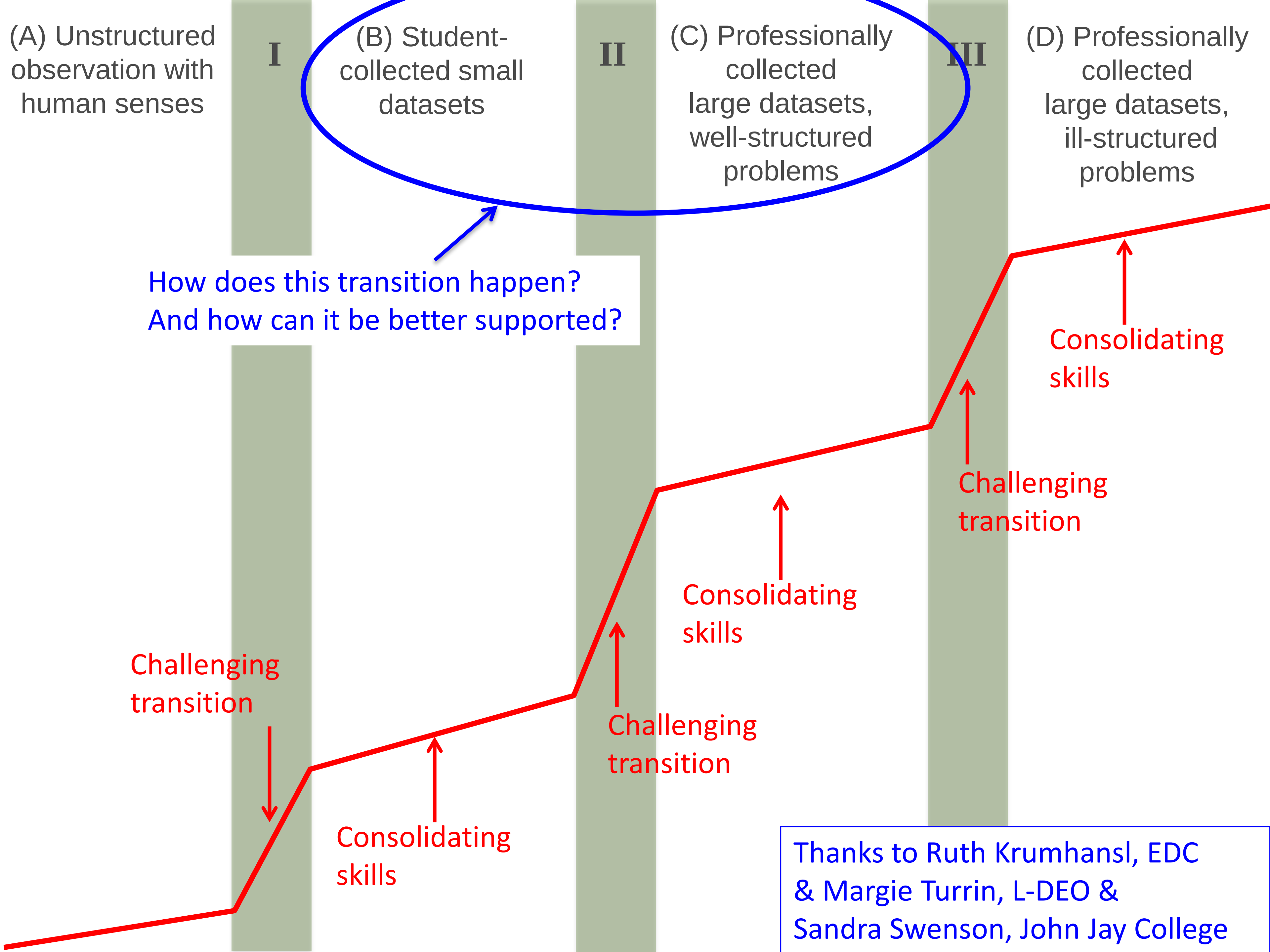
Learning Science from Scientific Data: Why bother?

Reason #1: Students can grasp the evidence base that underlies the big ideas of science, rather than having to take these ideas on authority.

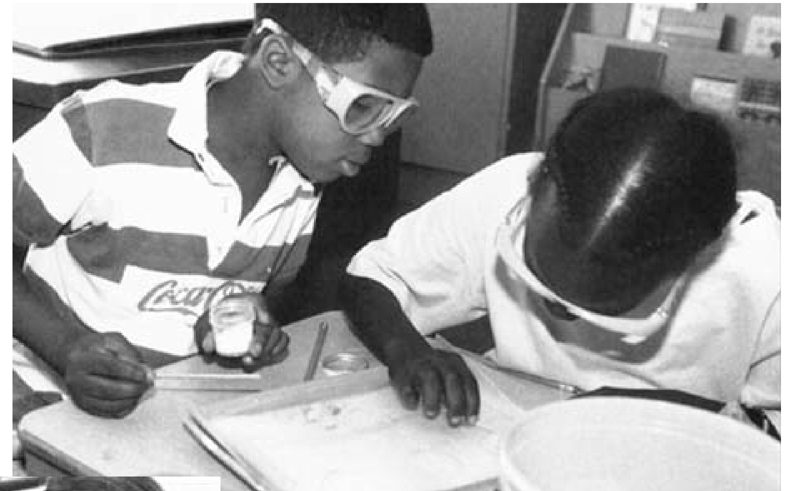
Reason #2: The world faces tough decisions and society is making some bad decisions. We want to raise up a generation who have the skills and disposition to make decisions based on evidence.

Reason #3: Those of our undergraduates who become scientists will need this skill set in their careers.


(A) Unstructured observation with human senses
I
(B) Student-collected small datasets
II
(C) Professionally collected large datasets, well-structured problems
III
(D) Professionally collected large datasets, ill-structured problems
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills


(A) Unstructured observation with human senses
I
(B) Student-collected small datasets
II
(C) Professionally collected large datasets, well-structured problems
III
(D) Professionally collected large datasets, ill-structured problems
How does this transition happen?
And how can it be better supported?
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills
Thanks to Ruth Krumhansl, EDC & Margie Turrin, L-DEO & Sandra Swenson, John Jay College

Most pre-college data experiences have been with small, student-collected data sets



Chapter 6
Science Content Standards

What is involved in this transition?

Student-collected data

Professionally-collected data

Day in the Life of the Hudson

Kim aboard *Joides Resolution*, Leg 107

Embodied, experiential grasp of the natural setting and data collection methods

Metadata

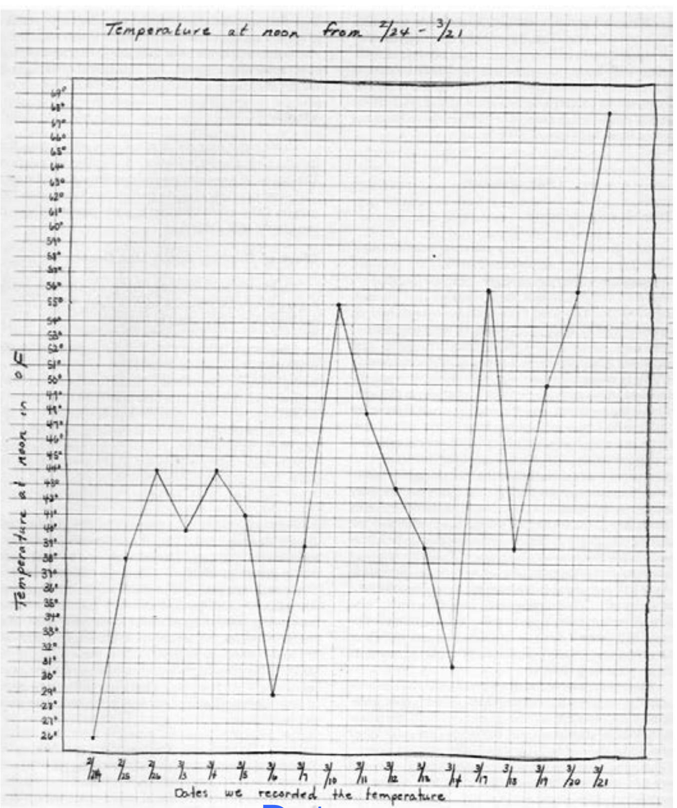
(from School in the Forest powerpoint, http://www.blackrockforest.org/docs/about-the-forest/schoolintheforest/)

(from Using a Digital Library to Enhance Earth Science Education, Rajul Pandya, Holly Devaul, and Mary Marlino)

Dozens of data points

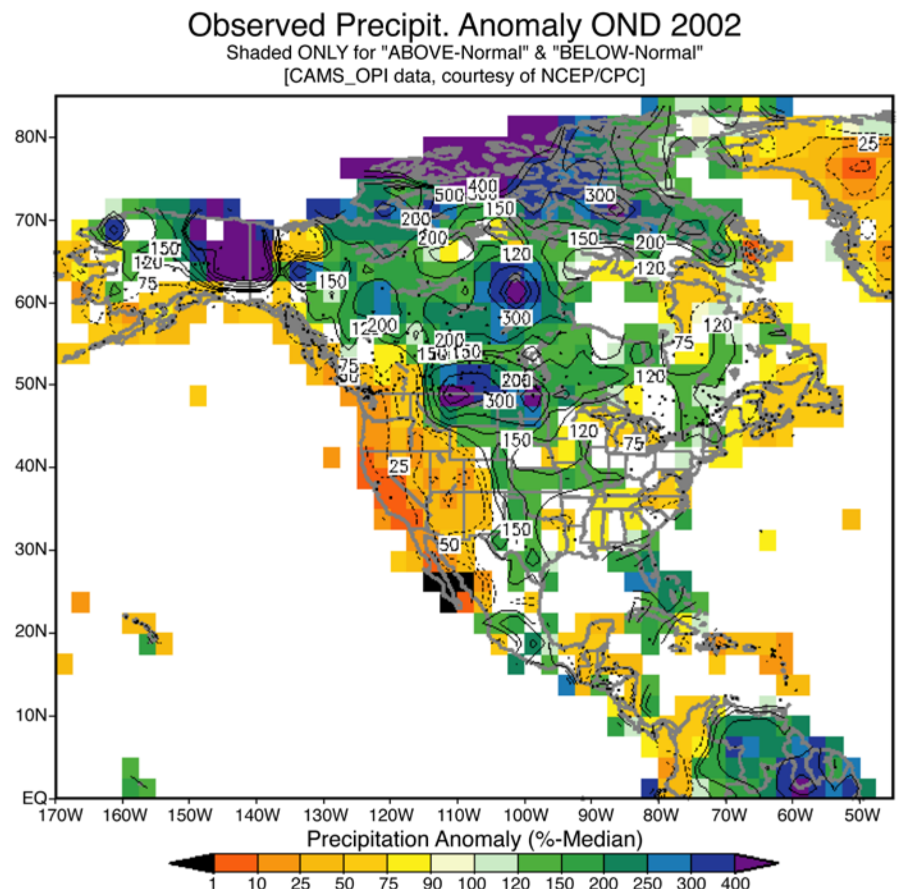
Megabytes

Air temperature at noon



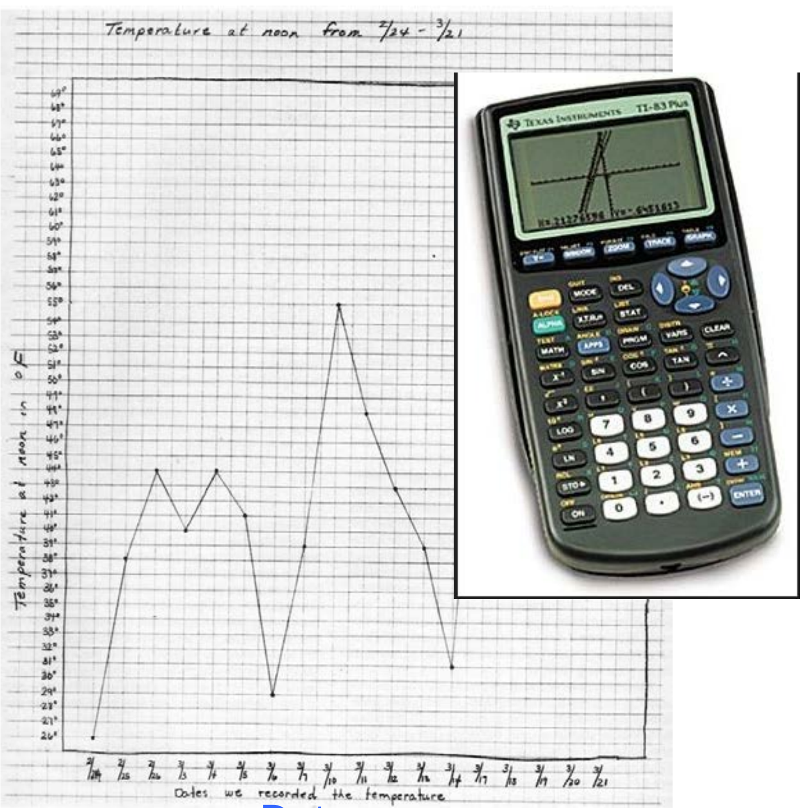
Date

(from Clement, 2002)



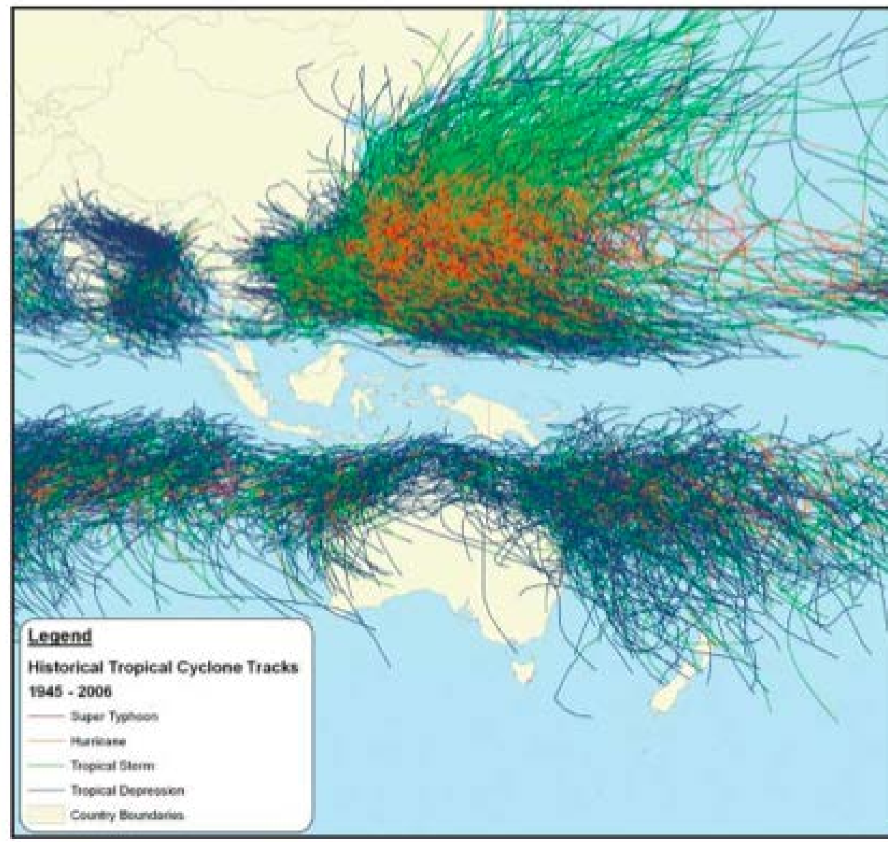
Simple, transparent tools and techniques

Sophisticated tools & techniques



Air temperature at noon

Date

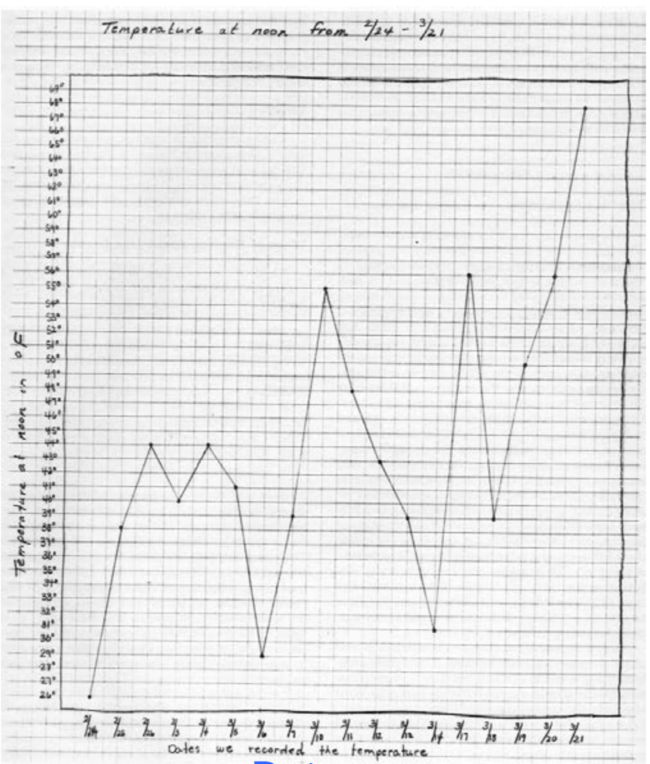


http://www.esri.com/library/ebooks/climate-change.pdf

Interpret one data set at a time

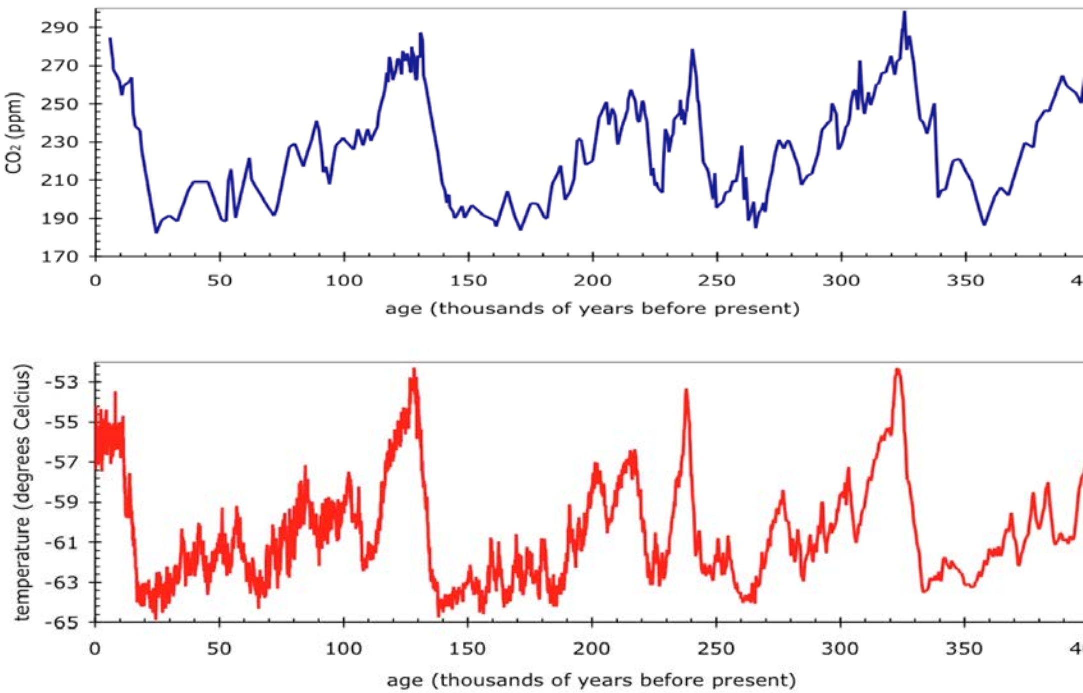
Multiple data sets with interactions; varying data types

Air temperature at noon



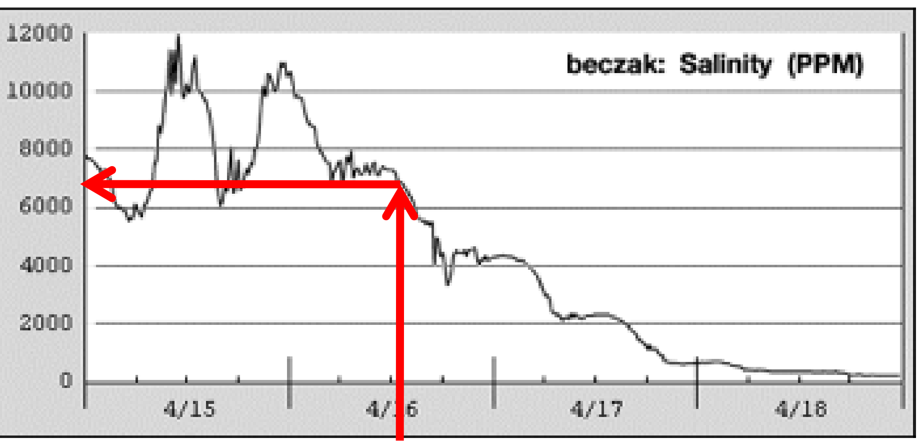
Date

(from Clement, 2002)

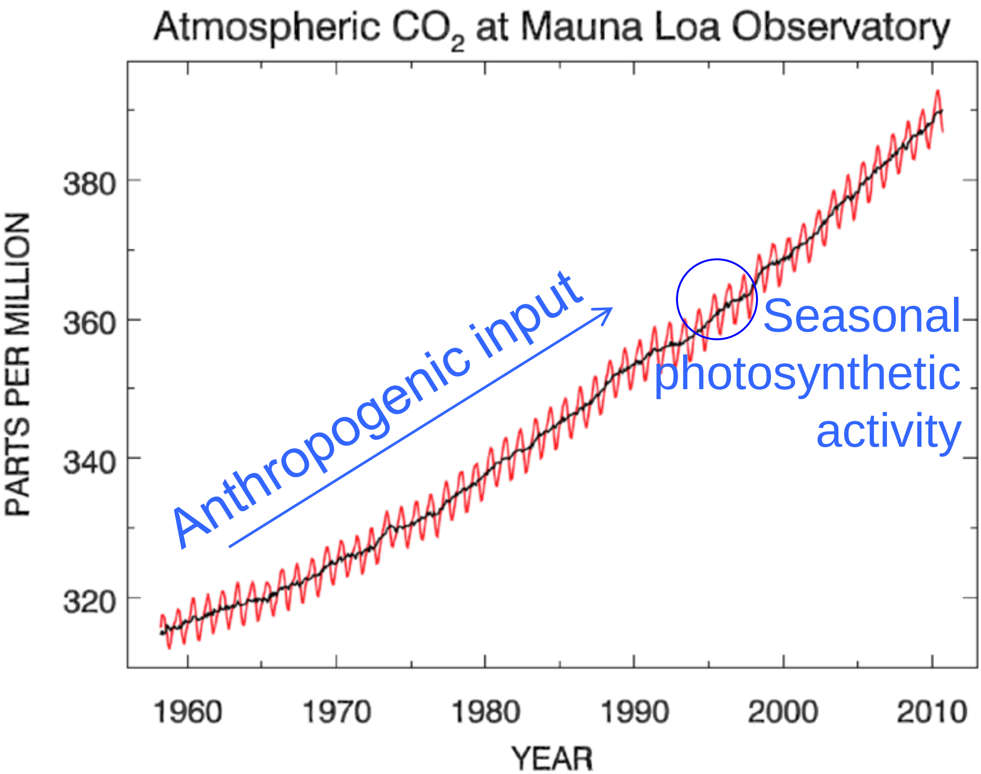


Looking up values

Seeing and interpreting patterns

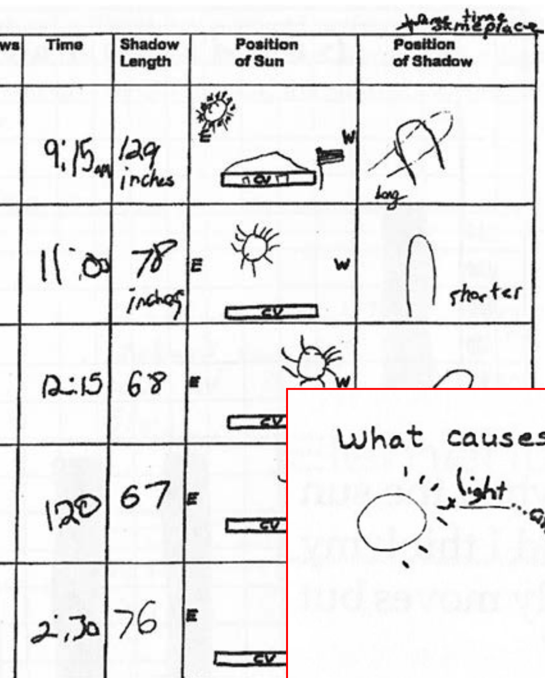


What was the salinity at noon on April 16?

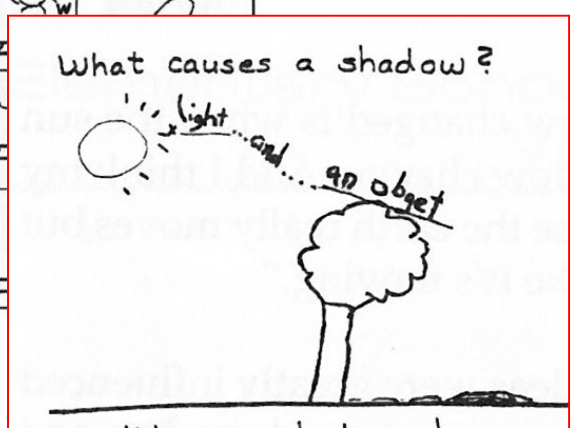


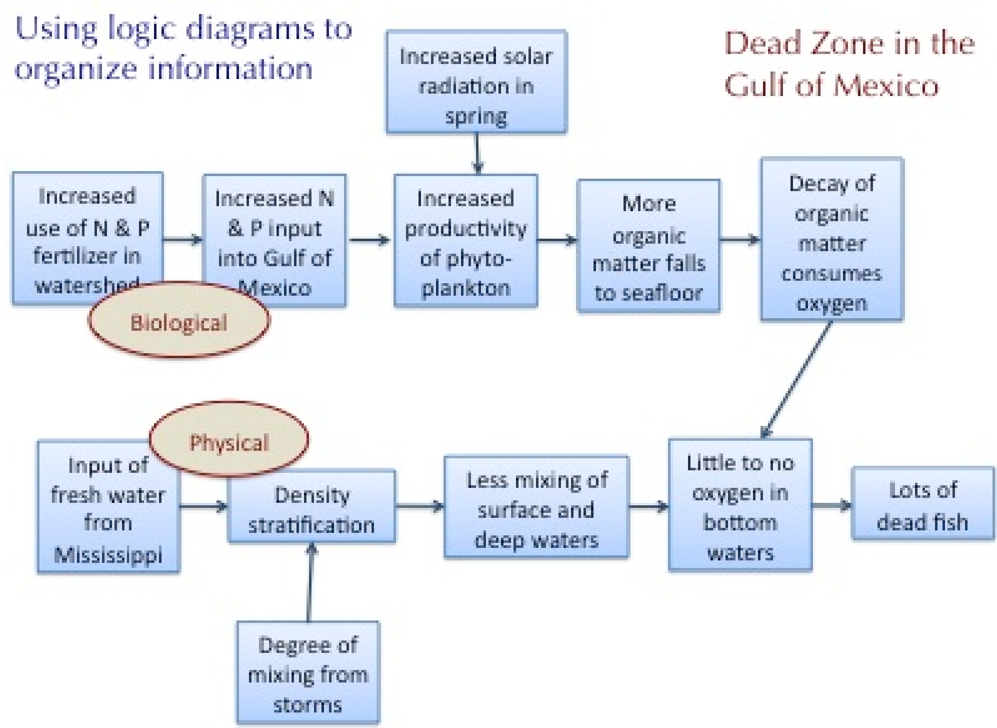
Common sense lines of reasoning

Spatial, temporal, statistical reasoning. Multi-step chains of reasoning

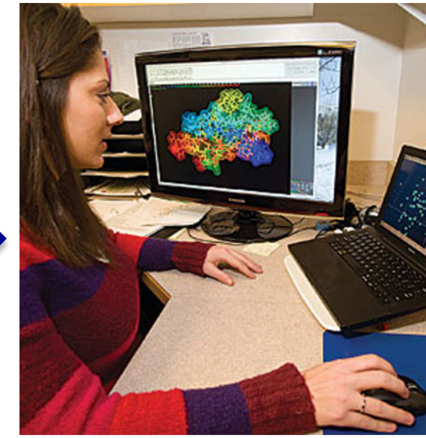


(Wainwright, 2002)





Day in the Life of the Hudson



http://www.umt.edu/urelations/rview/winter10/briknarova.html

- Student-collected data
- Embodied, experiential sense of circumstances
- Dozens to hundreds of data points
- Simple, transparent tools & techniques
- Interpret one data set at a time
- “Common sense” lines of reasoning
- Single step causal chains

- Professionally-collected data
- Sense of circumstances from metadata
- Megabytes
- Complex tools & techniques; black boxes
- Multiple data sets and their interactions
- Temporal, spatial, quantitative and other lines of reasoning
- Multi-step lines of reasoning

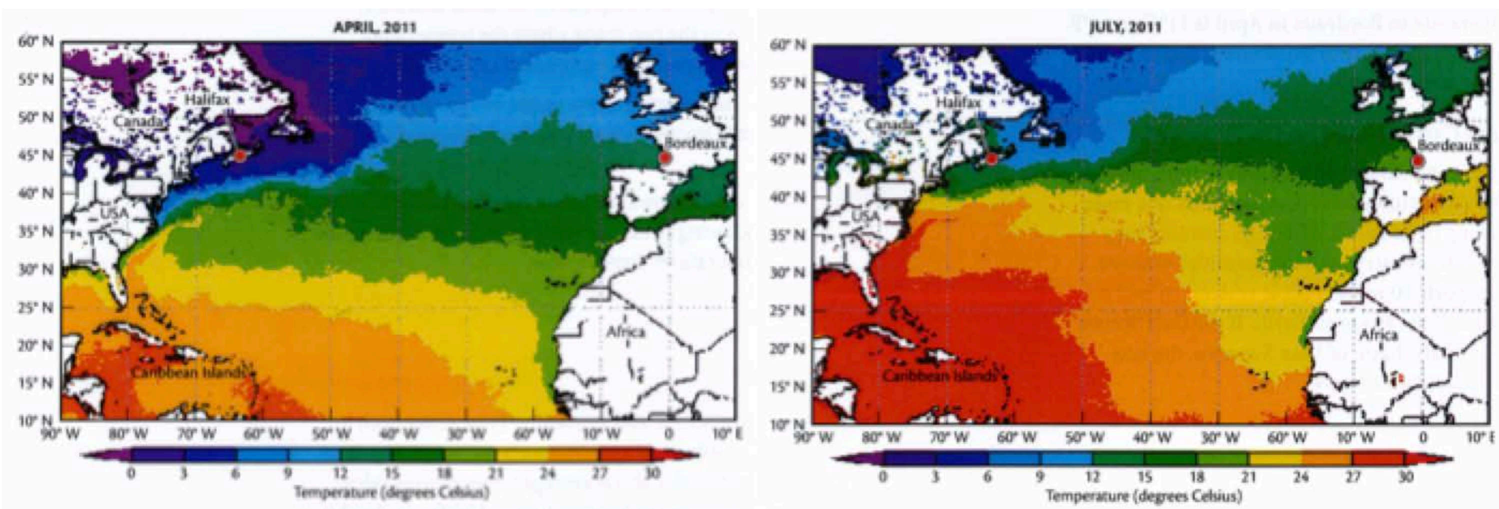
Four ways to scaffold students' transition from small, student-collected datasets to large, professionally-collected data bases

(1) Use pre-selected snippets of high insight:effort ratio data ("Data Puzzles").
(2) Nest a small student-collected data within a larger dataset.
(3) Ask students to commit to a prediction of what they will see before they start making data visualizations.
(4) Provide an array of candidate hypotheses.

High insight: effort ratio data snippet

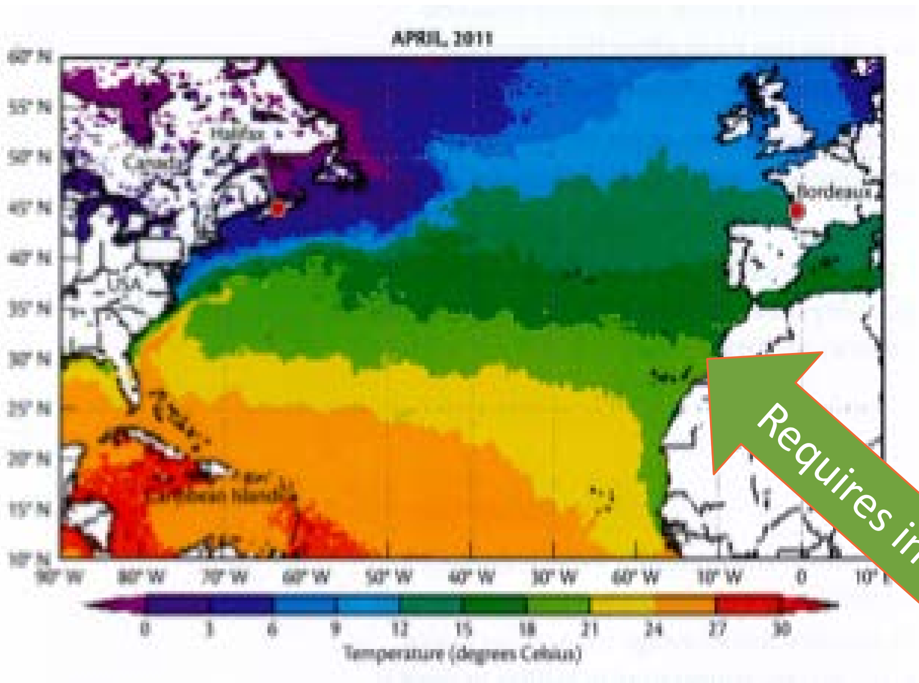
Procedure:

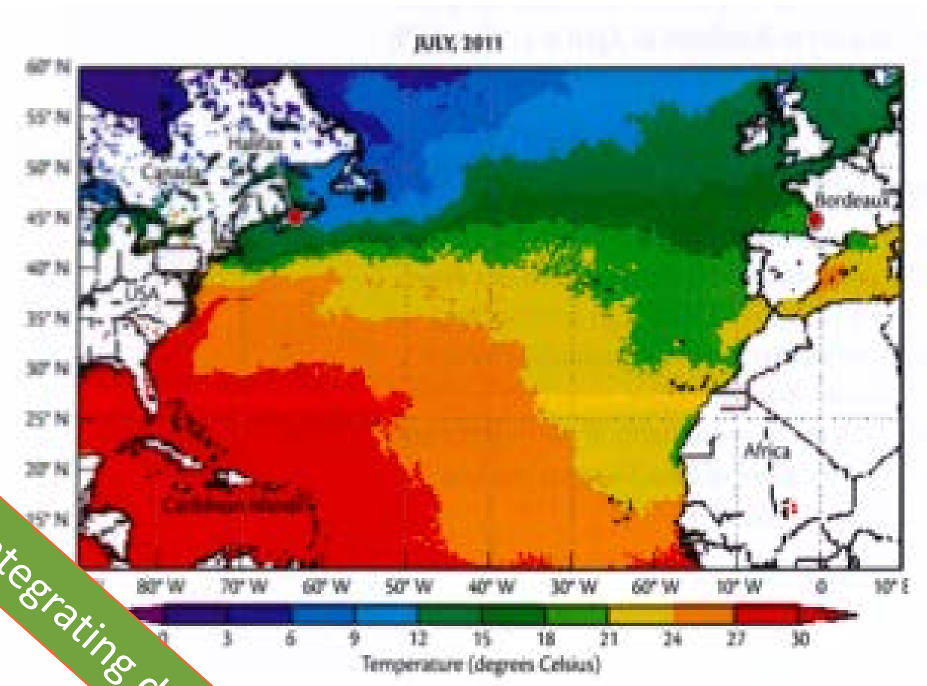
1) Curriculum developer identifies a small snippet of authentic data that embodies an important and widely-taught scientific concept, and develops data visualization(s) that foreground the patterns or relationships emerging from that concept.
2) Students view data visualizations on screen or paper, and answer guiding questions about the system represented by the data (not just about how to decode the data).
3) Students experience a rewarding "Aha! moment" of recognition when they see the process they have previously studied conceptually manifest in real world data.



From: Krumhansl, R., 2014, EDC Earth Science, Ronkonkoma, NY: Lab-Aids

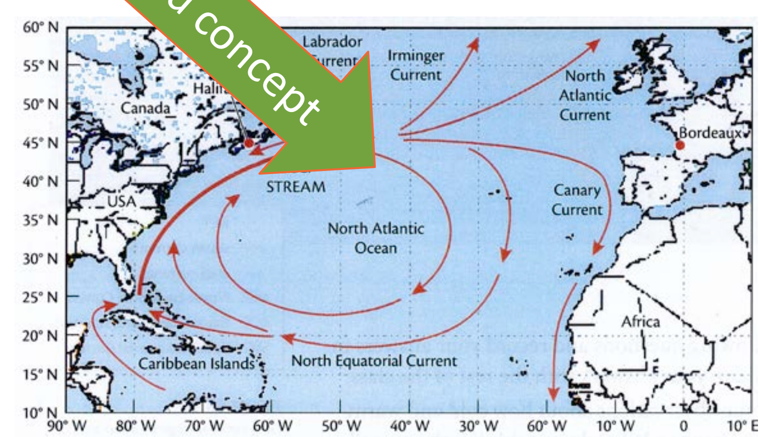
Decoding question: “What regions of the ocean had the warmest temperatures during April?”





Requires integrating data and concept

Aha! Question: Propose a hypothesis that might explain why the average temperature in Bordeaux is higher than in Halifax, even though they are at the same latitude.



(from Krumhansl, 2014, *EDC Earth Science*.)

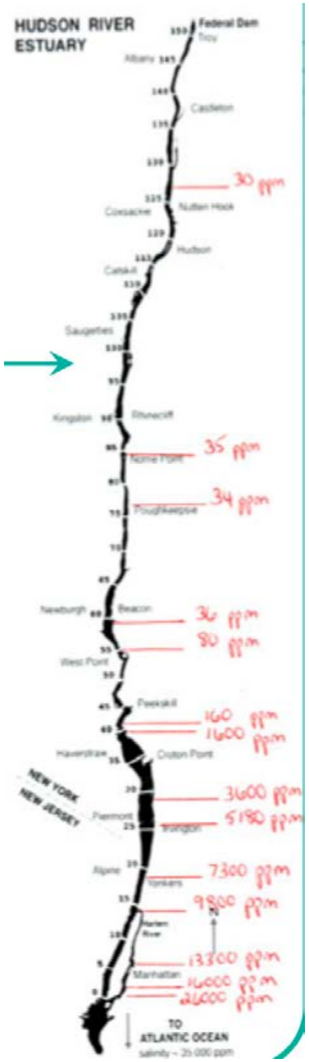
SSwithinLP

(Small, student-collected dataset within large, professionally collected dataset)

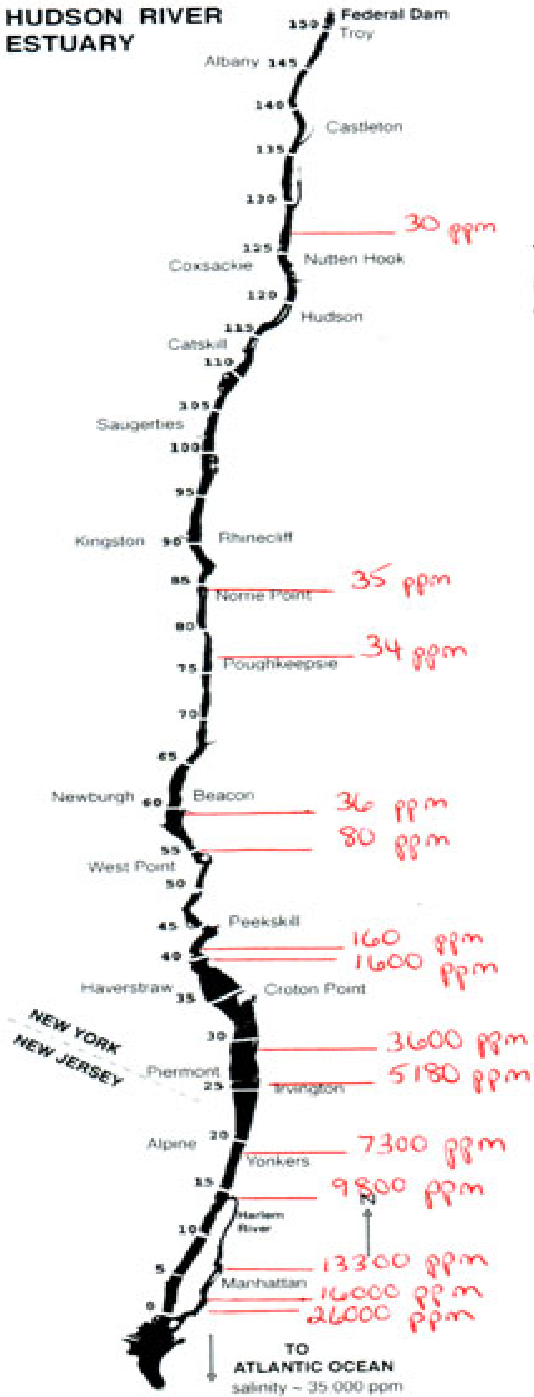
Procedure:

1) Students collect and interpret a local data set.
2) (optional) Students from multiple schools combine similar datasets to aggregate a larger sample or span a larger area.
3) Students interpret larger professionally collected dataset(s) which encompass and expand beyond the circumstances of their self-collected dataset.

Day in the Life of the Hudson

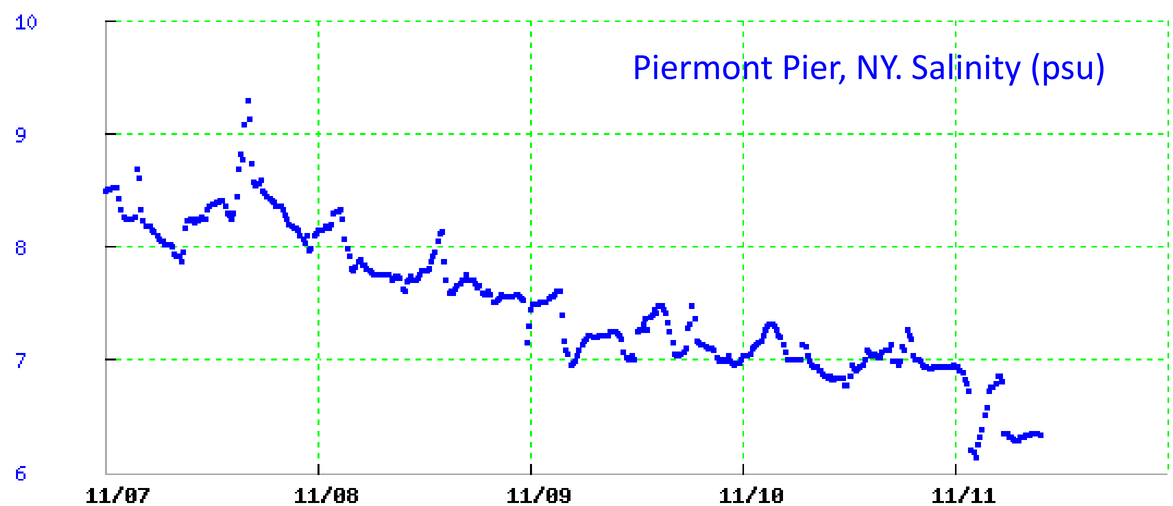


A Day in the Life of the Hudson: http://www.ldeo.columbia.edu/edu/k12/snapshotday/



Combine with other school groups' data to explore variation across space.

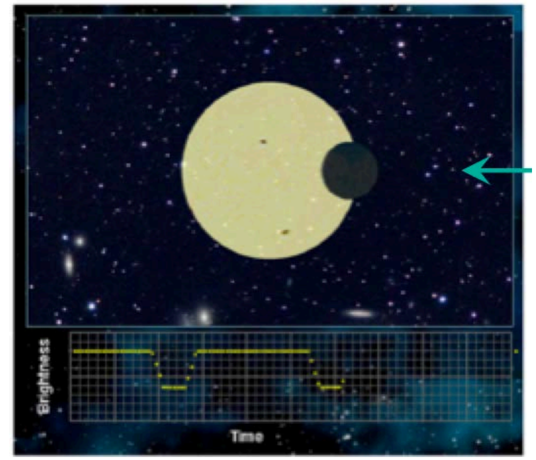
Combine with professionally collected data to explore changes through time.



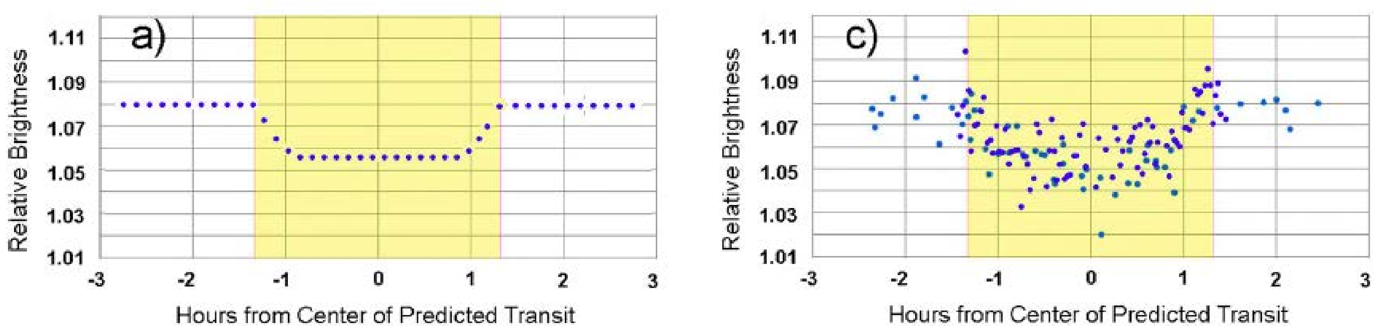
(from Turrin, M., & Kastens, K. A. (2010). In *Earth Science Puzzles: Making Meaning from Data* and http://www.hrecos.org/)

Prediction

Procedure:



1) Based on either a conceptual model, physical model or computational model, students predict what data from the system under consideration would look like under various conditions.
2) Students examine professionally collected data taken under a range of conditions, looking for the presence or absence of predicted patterns.

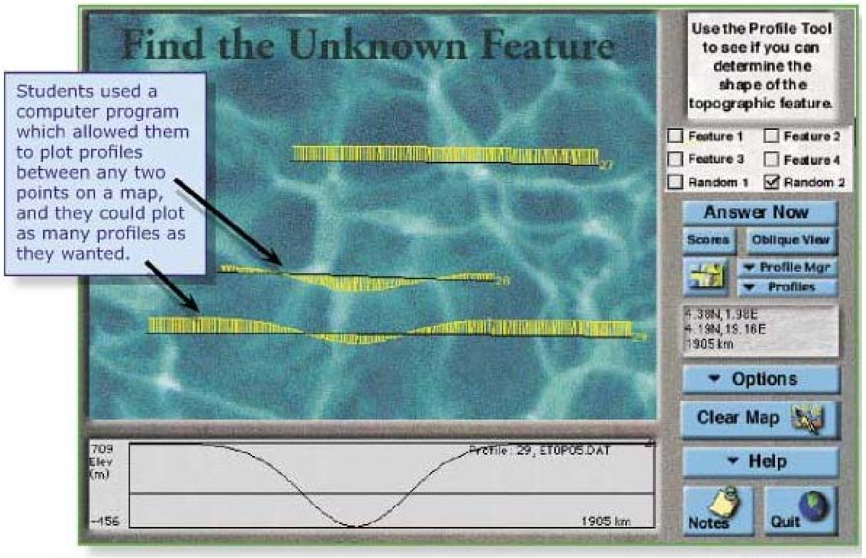


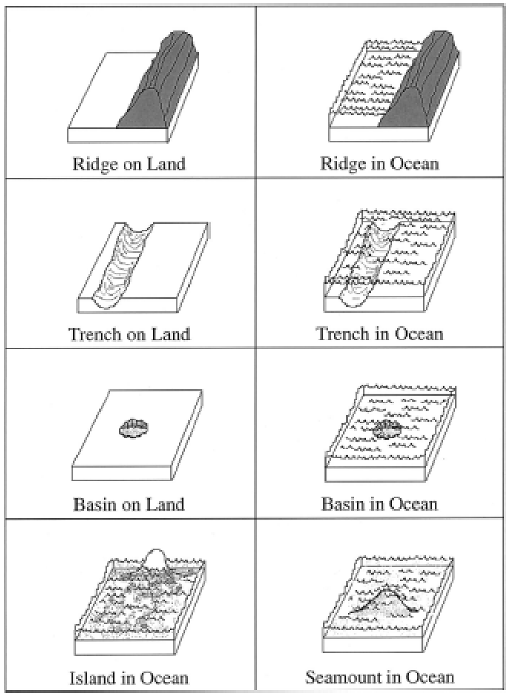
Gould, P., Sunbury, S., & Krumhansl, R. (2012). Using online telescopes to explore exoplanets from the physics classroom. *American Journal of Physics*, *80*(5), 445-451.

Hypothesis Array

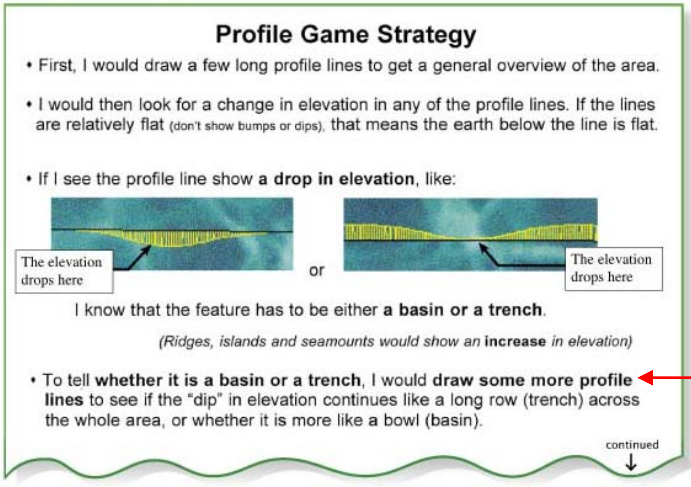
Procedure:

1) Students are provided with text descriptions or sketches of several alternative working hypotheses (the "choice array") that might depict a process or structure of the system under consideration.
2) Students explore a database of professionally collected data, seeking to assemble evidence in support of one of the hypotheses.

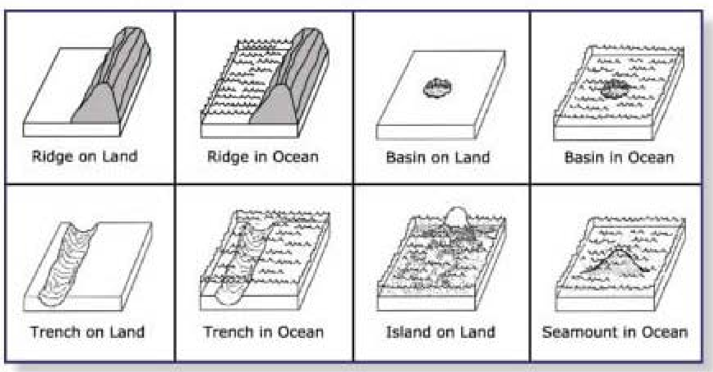




Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, *94*(1), 171-185.



Analyzing and clearly articulating the strategies used by experts…..



…. was not as valuable as providing a visual array of candidate answers.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, *94*(1), 171-185.

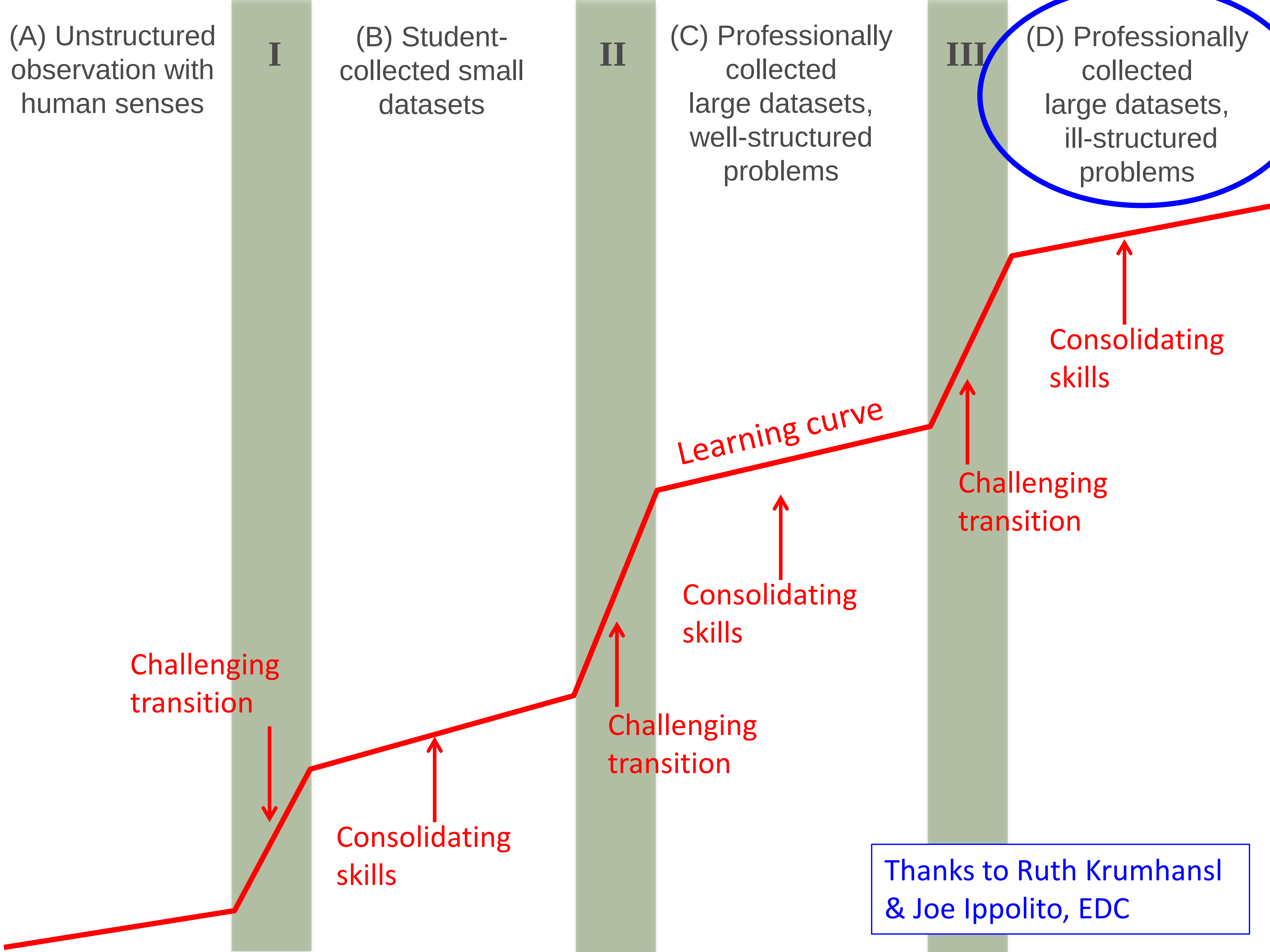
Four ways to scaffold students' transition from small, student-collected datasets to large, professionally-collected data bases

(1) Use pre-selected snippets of high insight:effort ratio data ("Data Puzzles").

(2) Nest a small student-collected data within a larger dataset.

(3) Ask students to commit to a prediction of what they will see before they start making data visualizations.

(4) Provide an array of candidate hypotheses.


(A) Unstructured observation with human senses
I
(B) Student-collected small datasets
II
(C) Professionally collected large datasets, well-structured problems
III
(D) Professionally collected large datasets, ill-structured problems
Learning curve
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills
Challenging transition
Consolidating skills
Thanks to Ruth Krumhansl & Joe Ippolito, EDC

DACUM: a process for Developing A CurriculUM

A methodology for occupational analysis

Premise: experienced and respected practitioners can best define and describe their job or profession

Product:

- Definition of the job/career/profession
- Duties & Tasks
- Knowledge, Skills, Tools & Behaviors

Expert Panel:
Aug 14-15, 2014

Developing an Occupational Profile

Medical Informatics

Astrophysics

Telecommunications

Utilities

Law Enforcement/ Forensics

Hydrology

What are the skills, knowledge and behaviors of a "big data-enabled specialist"?

Climate Modeling

Education

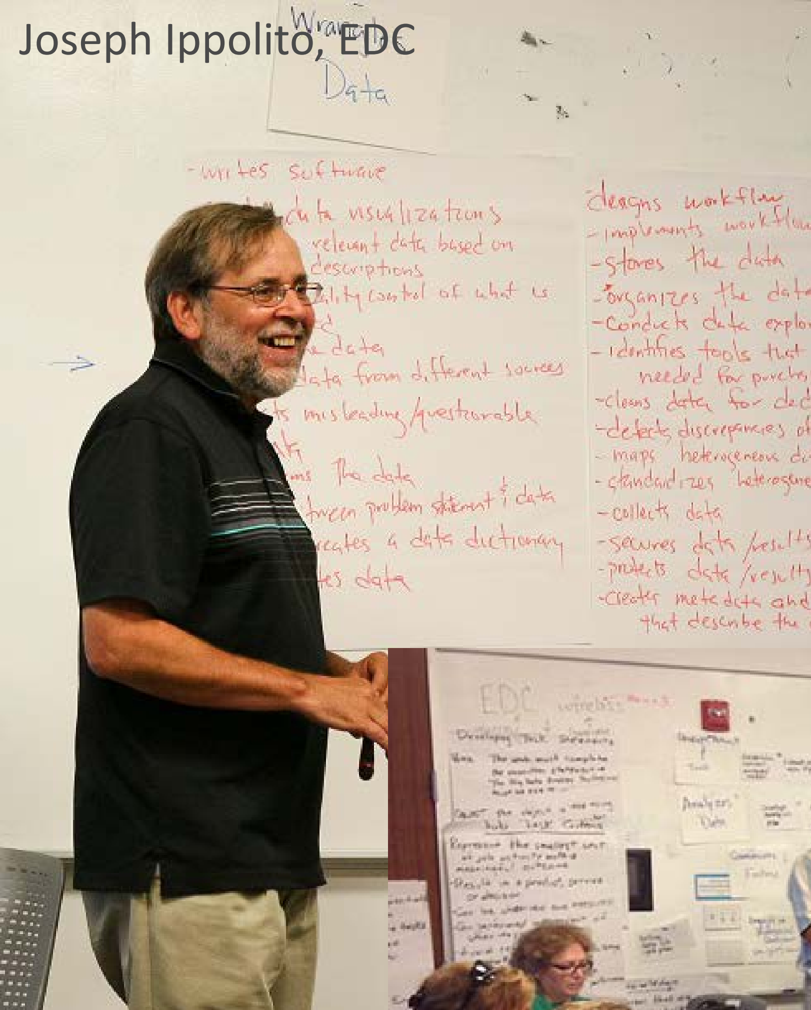
Hazard Analysis

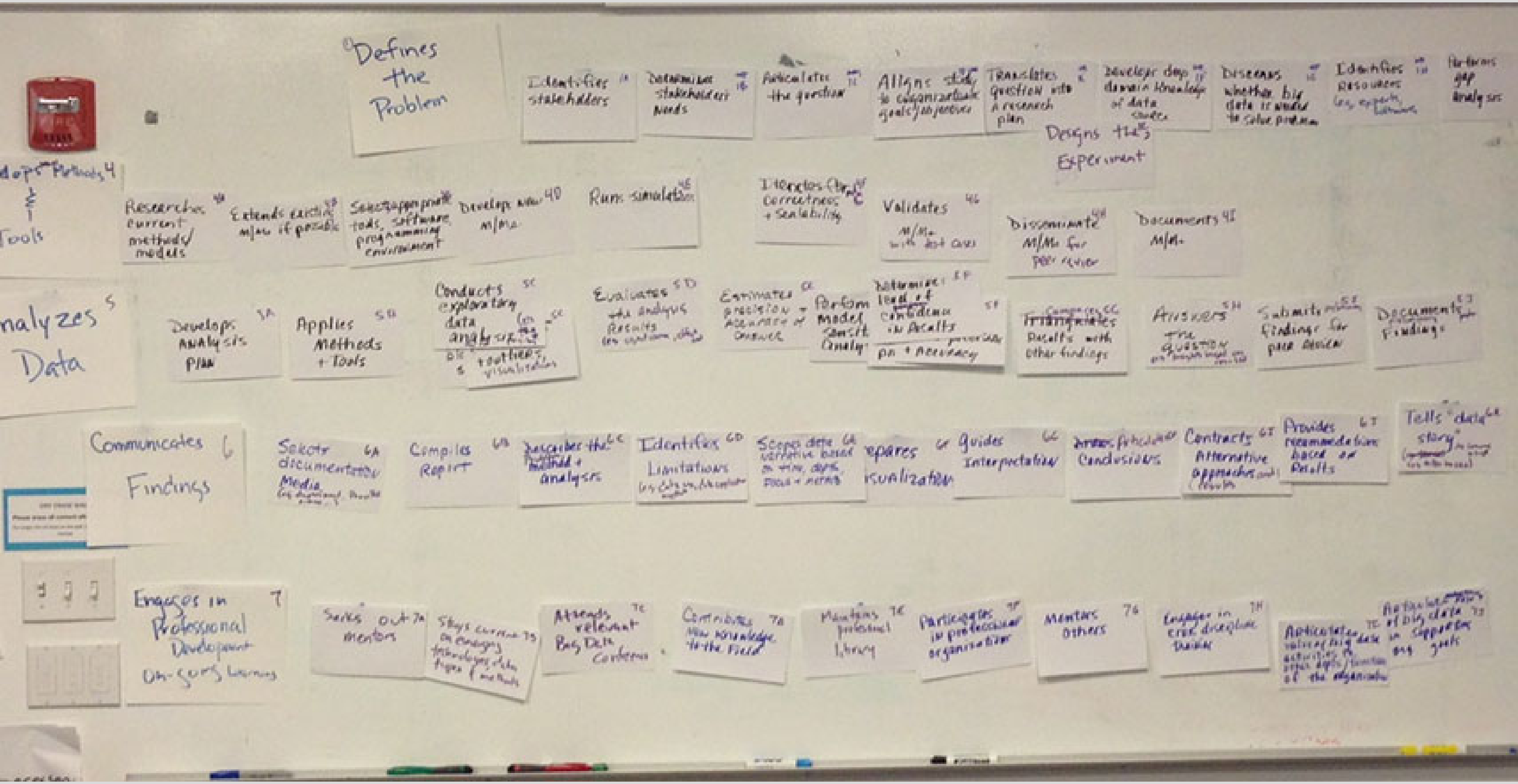
Marketing

Analytical Journalism

Bioacoustics

Joseph Ippolito, EDC




Defines the Problem
Identifies stakeholders
Determines stakeholders needs
Articulates the question
Aligns study to organizational goals/objectives
Translates question into a research plan
Develops deep domain knowledge of data source
Discerns whether big data is needed to solve problem
Identifies Resources
Performs gap analysis
Designs the Experiment
Tools
Researches current methods/models
Extends existing m/m if possible
Selects appropriate tools, software programming environment
Develops new m/m
Runs simulations
Iterates for correctness + scalability
Validates m/m with test cases
Disseminates m/m for peer review
Documents m/m
Analyzes Data
Develops analysis plan
Applies methods + tools
Conducts exploratory data analysis
Evaluates the analysis results
Estimates precision + accuracy of answer
Perform model sensitivity analysis
Determines level of confidence in results
Compares results with other findings
Answers the question
Submits findings for peer review
Documents findings
Communicates Findings
Selects documentation media
Compiles report
Describes the method + analysis
Identifies limitations
Scopes data version based on time, depth, focus + metrics
Prepares visualization
Guides interpretation
Draws particular conclusions
Contrasts alternative approaches and results
Provides recommendations based on results
Tells "data story"
Engages in Professional Development On-going learning
Seeks out mentors
Stays current on emerging technologies, data types & methods
Attends relevant Big Data conferences
Contributes new knowledge to the field
Maintains professional library
Participates in professional organizations
Mentors others
Engages in cross discipline dialog
Articulates value of big data activities to other depts/functions of the organization
Articulates the value of big data in supporting org goals

Occupational Profile

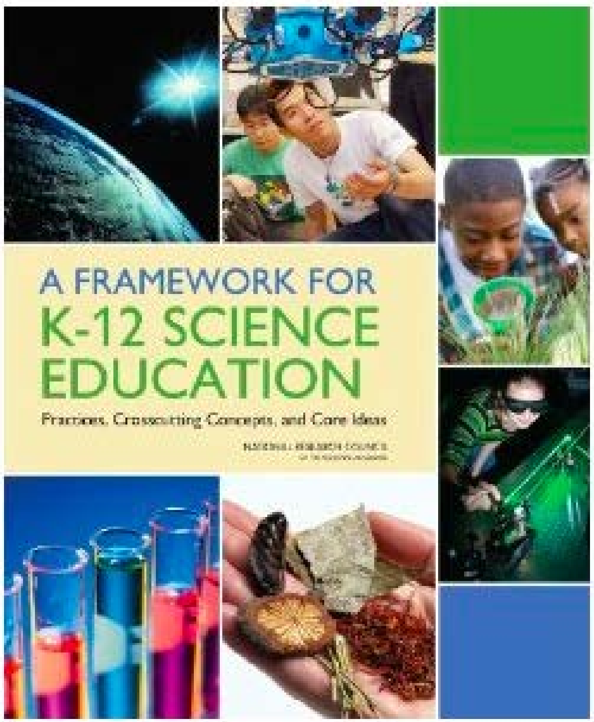
DUTIES	TASKS						
1. Defines the Problem	1A. Identifies stakeholders	1B. Determines stakeholders' needs	1C. Articulates the question	1D. Aligns study to organizational goals and objectives	1E. Translates question into a research plan	1F. Designs the experiment	1G. Develops deep domain knowledge of data source
	1M. Negotiates plan, including deadlines and budgets	1N. Creates requirement document (sign-off)					
2. Wrangles Data	2A. Performs data exploration	2B. Identifies data	2C. Creates the data dictionary	2D. Collects data	2E. Assesses the extent/methods to clean the data	2F. Maps data across heterogeneous sources	2G. Identifies outliers and anomalies
	2M. Writes software to automate tasks	2N. Documents the process					
3. Manages Data Resources	3A. Manages data life cycle	3B. Conducts capacity planning of resources	3C. Complies with legal obligations	3D. Applies ethical standards	3E. Identifies tools that may be needed for purchase or modification	3F. Protects data and results	3G. Determines access to the data
4. Develops Methods and Tools	4A. Researches current methods/models	4B. Extends existing methods/ models, if possible	4C. Selects tools/ software/ programming environment	4D. Develops new methods/models	4E. Runs simulations	4F. Iterates correctness and scalability of methods/models	4G. Validates methods/ models with test cases
5. Analyzes Data	5A. Develops analysis plan	5B. Applies methods and tools	5C. Conducts exploratory analysis (e.g., identifies anomalies, outliers, bias in sampling; visualizes)	5D. Evaluates results of the analysis (e.g., significance, effect, size)	5E. Estimates precision and accuracy of answer	5F. Determines level of confidence in results	5G. Compares results with other findings

Etc.

Etc.

Gap analysis

How well is the current education system doing at preparing students for the tasks and duties of the big-data-enabled specialist?



NEXT GENERATION
SCIENCE
STANDARDS
For States, By States

- Disciplinary Core Ideas
- Cross-cutting Themes
- Practices of Science & Engineering

  #4: Analyze & interpret data

Comparison of ODI occupational profile tasks with NGSS Performance Expectations

DUTIES	TASKS										
5. Analyzes Data	5A. Develops analysis plan	5B. Applies methods and tools	5C. Conducts exploratory analysis (e.g., identifies anomalies, outliers, bias in sampling; visualizes)	5D. Evaluates results of the analysis (e.g., significance, effect, size)	5E. Estimates precision and accuracy of answer	5F. Determines level of confidence in results	5G. Compares results with other findings	5H. Answers the question (e.g., insights drawn from results)	5I. Submits preliminary findings for peer review	5J. Documents preliminary findings	
6. Communicates Findings	6A. Selects documentation media (e.g., dashboard, PowerPoint, e-mail)	6B. Compiles report	6C. Describes problem, method, and analysis	6D. Identifies limitations (e.g., data use, data application methods)	6E. Scopes data narrative based on time, depth, and method	6F. Prepares visualizations	6G. Guides interpretation	6H. Articulates conclusions	6I. Contrasts alternative approaches and past results	6J. Provides recommendations based on results	6K. Tells "data story" to convey insight (e.g., talks to CEO)
7. Engages in Professional Development	7A. Seeks out mentors	7B. Stays current on emerging technologies, data types, and methods	7C. Attends relevant big data conferences	7D. Contributes new knowledge to the field	7E. Maintains professional library	7F. Participates in professional organizations	7G. Mentors others	7H. Engages in cross-discipline training	7I. Articulates value of big data activities to other departments/ functions of organization	7J. Articulates evolving role of big data in supporting organizational goals	

Abundant in NGSS

Potentially (implicitly) abundant in NGSS

Sparse in NGSS

Absent from NGSS

EDC OCEANS of DATA INSTITUTE

Occupational Profile tasks that are *well-represented* in NGSS

1. Defines the Problem
    - 1B. Determines stakeholders' needs
    - 1C. Articulates the question
    - 1E. Translates question into a research plan
    - 1F. Designs the experiment
    - 1G. Develops deep domain knowledge of data source

2. Wrangles Data
    - 2D. Collects data

5. Analyzes Data
    - 5A. Develops analysis plan
    - 5B. Applies methods and tools
    - 5D. Evaluates results of the analysis (e.g., significance, effect, size)
    - 5H. Answers the question (e.g., insights drawn from results)

Occupational Profile tasks that are *absent from* NGSS

2. Wrangles Data
    - 2A. Performs data exploration
    - **2G. Identifies outliers and anomalies**
    - 2N. Documents the process

3. Manages Data Resources
    - **3D. Applies ethical standards**
    - **3F. Protects data and results**

4. Develops Methods and Tools
    - 4F. Iterates correctness … of … models

5. Analyzes Data
    - **5F. Determines level of confidence in results**

6. Communicates Findings
    - **6D. Identifies limitations (e.g., data use, data application methods)**

Bottom line:

- It’s a long, complicated pathway to grow a populace that has the skills and disposition to use data as part of their tool-kit when confronted with a difficult question or problem.
- There are effective instructional templates to build on experience with small, student-collected datasets towards proficiency with large, complex, professionally-collected data.
- Big-data enabled professionals value a suite of skills around data quality, data safety, and data ethics that may be missing from today’s students.

kastens@ldeo.columbia.edu

For more information:

- Kastens, K. A., & Turrin, M. (2010). *Earth Science Puzzles: Making Meaning from Data*. Washington, D.C.: National Science Teachers Association.
- Kastens, K. A. (2011) *Learning to Learn from Data.* http://serc.carleton.edu/earthandmind/posts/datalearningpro.html
- Oceans of Data Institute (2014). Profile of a Big-Data-Enabled Specialist. http://www.oceansofdata.org/our-work/profile-big-data-enabled-specialist
- Kastens, K. (2015). Data Use in the Next Generation Science Standards. http://www.oceansofdata.org/our-work/data-next-generation-science-standards
- Kastens, K. A., Krumhansl, R., & Baker, I. (2015). Thinking Big: Transitioning your students from working with small student-collected data sets towards "big data". *The Science Teacher, 82*(5), 25-31.

EDC OCEANS of DATA INSTITUTE