Bringing Data Science to Middle and High Schools

Hollylynne S. Lee

Professor of Mathematics & Statistics Education, NC State University

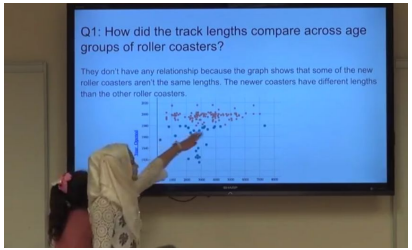
2018-19 University Scholar, RTI International

Peter Baumgartner

Research Data Scientist, RTI International

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION

Partially supported by grants from Hewlett Foundation, NSF DUE 1625713, and RTI International University Scholars program. Opinions and findings are those of the presenters.



Today's Goals

1. Generate interest and connections among folks who care about the pipeline towards data-intensive careers.
2. Share current efforts in school-level data science
3. Share current research about data science and AP Statistics curriculum and teaching practices
4. A glimpse into what data science skills could look like in schools

Go to Menti.com Use code 33 72 61

Let's start with crowdsourcing the idea of data science...

Enter words or brief phrases that describe the skills and ways of thinking used in data science.

How did webinar participants describe skills and ways of thinking in data science?

Creating Data Literate Citizenry

"Data are abundant, quantitative information about the state of society and the wider world is around us more than ever. Paradoxically, recent trends in the public discourse point towards a post-factual world that seems content to ignore or misrepresent empirical evidence. **As statistics educators we are challenged to promote understanding of statistics about society.** In order to re-root public debate to be based on facts instead of emotions and to promote evidence-based policy decisions, statistics education needs to embrace two areas widely neglected in secondary and tertiary education:

understanding of **multivariate phenomena** and

the **thinking with and learning from complex data**." (Engel, 2017, p. 1)

http://iase-web.org/documents/SERJ/SERJ16(1)_Engel.pdf

Where are opportunities for data science in middle and high school curricula?

Sample Common Core Mathematics Standards

CCSS.MATH.CONTENT.HSS.ID.A.2 Use statistics appropriate to the **shape of the data distribution to compare center** (median, mean) and spread (interquartile range, standard deviation) of **two or more different data sets.**

CCSS.MATH.CONTENT.HSS.ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, **accounting for possible effects of extreme data points (outliers)**.

CCSS.MATH.CONTENT.HSS.ID.B.6 **Represent data on two quantitative variables** on a scatter plot, and describe how the variables are related.

- CCSS.MATH.CONTENT.HSS.ID.B.6.A **Fit a function to the data;** use functions fitted to data to solve problems in the context of the data.
- CCSS.MATH.CONTENT.HSS.ID.B.6.B Informally **assess the fit** of a function by plotting and analyzing residuals.

CCSS.MATH.CONTENT.HSS.ID.C.7 **Interpret** the slope (rate of change) and the intercept (constant term) of a linear model in **the context of the data**.

Sample Next Generation Science Standards

MS-LS2-1 Ecosystems: Interactions, Energy, and Dynamics **Analyze and interpret data** to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem.

MS-ESS1-3 Earth's Place in the Universe **Analyze and interpret data** to determine scale properties of objects in solar system.

MS-ESS2-3 Earth's Systems **Analyze and interpret data** on the distribution of fossils and rocks, continental shapes, and seafloor structures to provide evidence of the past plate motions.

HS-PS3-1 Energy **Create a computational model** to calculate the change in the energy of one component in a system when the change in energy of the other component(s) and energy flows in and out of the system are known.

HS-PS3-4 Energy **Plan and conduct an investigation** to provide evidence that the transfer of thermal energy when two components of different temperature are combined within a closed system results in a more uniform energy distribution among the components in the system

HS-PS4-1 Waves and their Applications in Technologies for Information Transfer **Use mathematical representations to support a claim regarding relationships** among the frequency, wavelength, and speed of waves traveling in various media.

Sample ELA standards

CCSS.ELA-LITERACY.RH.6-8.7 Integrate **visual information** (e.g., in **charts, graphs,** photographs, videos, or maps) with print and digital texts.

CCSS.ELA-LITERACY.RST.11-12.3 **Follow precisely a complex multistep procedure when carrying out experiments, taking measurements, or performing technical tasks**; analyze the specific results based on explanations in the text.

CCSS.ELA-LITERACY.RST.11-12.5 **Analyze how the text structures information or ideas into categories or hierarchies,** demonstrating understanding of the information or ideas.

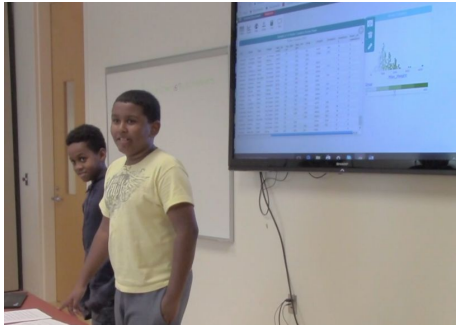
CCSS.ELA-LITERACY.RST.11-12.7 **Integrate and evaluate multiple sources** of information presented in diverse formats and media (e.g., **quantitative data, video, multimedia**) in order to address a question or solve a problem.

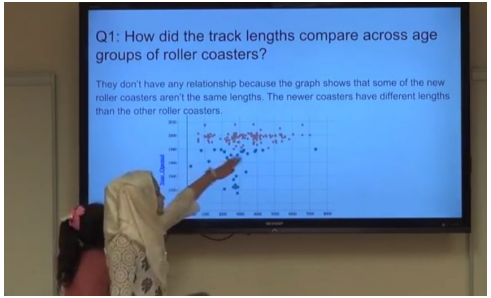
CCSS.ELA-LITERACY.RST.11-12.8 Evaluate the **hypotheses, data, analysis, and conclusions** in a science or technical text, **verifying the data** when possible and corroborating or challenging conclusions with other sources of information.

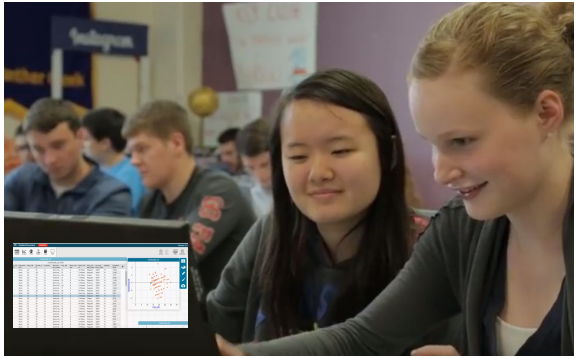
Sample Standards in Social Studies

1. **Define and frame questions about events and the world in which we live, form hypotheses** as potential answers to these questions, **use evidence** to answer these questions, and consider and analyze counter-hypotheses.
2. **Identify, describe, and evaluate evidence** about events from diverse sources (including written documents, works of art, photographs, **charts and graphs**, artifacts, oral traditions, and other **primary and secondary sources**).
3. **Identify causes and effects** using examples from different time periods and courses of study across several grade levels.
4. **Identify, analyze, and evaluate relationship** between multiple causes and effects
5. **Distinguish between long-term and immediate causes** and multiple effects (time, continuity, and change)

Big(ger) Data in Schools







Examples Efforts to bring Data Science to Schools

Census at Schools project from ASA--large messy data collected from kids!

Data Visualization poster competition ASA

Oceans of Data Institute at Educational Development Center

Zoom In! Learning Science with Data 3 Earth Science and 3 Biology modules for HS students.

Exploring Urban Mobility 2 modules on different urban mobility problems for HS students.

Data Modeling and Computational Thinking in middle schools (project Engage at Friday Institute NC State)

Introduction to Data Science course developed and taught in LA HSs. uses RStudio

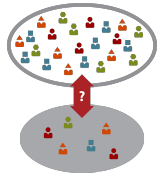
Bootstrap modules to integrate computational modeling into algebra, physics and data science

Concord Consortium Projects such as CodeR4Math, Data Science Games, InSPECT, and development of the platform CODAP for easy data moves and visualization.

Data Science Education webinars --many examples of K-12 efforts!

Teacher Preparation Efforts at FI and RTI

1. Two MOOCs for Educators and micro-credentials with a data focus offered since 2015 to over 4000 educators in middle, high, and collegiate settings!
   - *Teaching Statistics Through Data Investigations*
   - *Teaching Statistics Through Inferential Reasoning* **open now!**
   -



2. Other MOOCs for educators related to computational thinking
   - Teaching Beauty and Joy of Computing
   - Problem solving in the Digital Age
   - Computer Science Discoveries

More efforts in Teacher Preparation

1. ESTEEM-Enhancing Statistics Teacher Education through E-modules.
   - Developing modules to be integrated into teacher preparation courses.
   - Teaching teachers to use big(ger) data and investigate real world phenomena.
2. From RTI--Developing resources to introduce data science skills and habits to secondary teachers--to be completed in Fall 2019

Two Ongoing Efforts

1. Representing work of data scientists in ways that can help middle and high school students envision themselves in a data-intensive career
2. Use perspectives of data scientists to inform new definitions and frameworks
3. Examine opportunities for data science in curricula.
4. Outreach to teachers through RTI symposiums

Talking with Data Scientists

Interviews with data scientists (or folks who work with large data as part of their job!)

- Ongoing study, 5 interviews thus far
- Also using public interviews (blogs, articles, etc) with data scientists
- What are key ways of working with data
- How to explain data science to younger students
- How to get more data science in schools

In one sentence or phrase...

What would you tell a middle or high school teacher they should include in their instruction to give students experience with data to be prepared to pursue a data-intensive career?

Menti.com Code 33 72 61

What a (convenience) sample of data scientists said.....

“Case studies. Actual problems where you're measuring and observing data or sourcing it from somewhere where they have to get Hands-On and actually understand what the data means. This is such a crucial step.... Like I'm going with a group of students and we're standing on the corner where **we’re recording how many cars pass every day** for six weeks by hour or whatever. Now, since I've been doing this and **interacting with the data in this way, I've kind of got a couple hunches about what I'm going to see**. I know that it gets really really busy at 8 a.m. And then it gets really really busy at 4 pm. Now I’ve got some actual experience with what the data means. so now when I run my regression and I'm looking at my time series or whatever models I'm using I can look at the results and I can say that's what it looks like when I see a trend. **I've *felt* this trend, and now I know what it mathematically looks like when it manifests in this model.”**

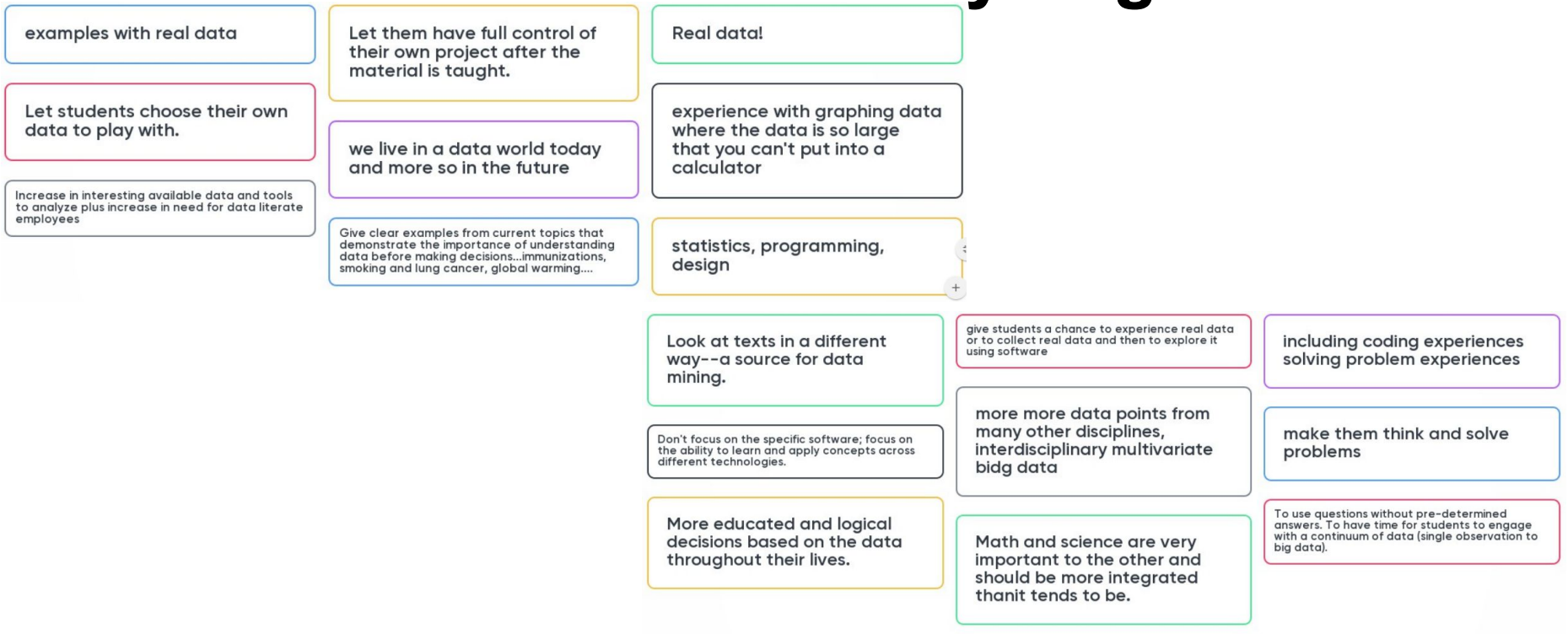
“I think, as a high schooler, they would enjoy **visualization**. So if you were doing some, **if you had a clean data set**, you’re not asking them doing boring, intensive cleaning-- you have a clean data set, maybe related to sports, or related to candy or **something they can be interested in** -- and you just have them kind of **do a few small coding task of visualization**.”

“I know, a lot of education right now is kind of catered towards standardized testing. Um, and **standardized testing is pretty much the exact opposite of doing what I described and what you're gonna have to do in the real world**....So do **more open ended, like project type approaches, or project type assignments that encourage students to kind of find their own way and, and come up with their own solutions and that sort of thing.”**

“So they could **gather their data and they need to understand at the very basic level what algorithm they use, what assumptions they make about the data**. And then kind of apply that to—and then they will be very excited to see that, you know how they’re **finding the most salient features of their data. “**

“I feel like **general computer programming** is an easy thing to be taught... The **concept of statistics**-- I think is important... getting some of those basic concepts of **how you might not know what one person is going to do, but if you've got 1000 people you can predict that a few of them are going to do this and a few of them are going to do that.** ... Getting comfortable with that kind of a thinking and understanding, **sometimes finding that trend is easy and sometimes it's really hard.** I guess, trying-- **developing resilience** is like trying a problem that, you know, you, you might even set them up to fail the first couple times, and then show them, “hey we learned this the first time, we learn that the second time and look at this third thing that we did, and it’s perfect.” And **getting them comfortable with that kind of scientific method of just keep on chipping away at a problem until you get to your goal**.”

What did advice did you give?

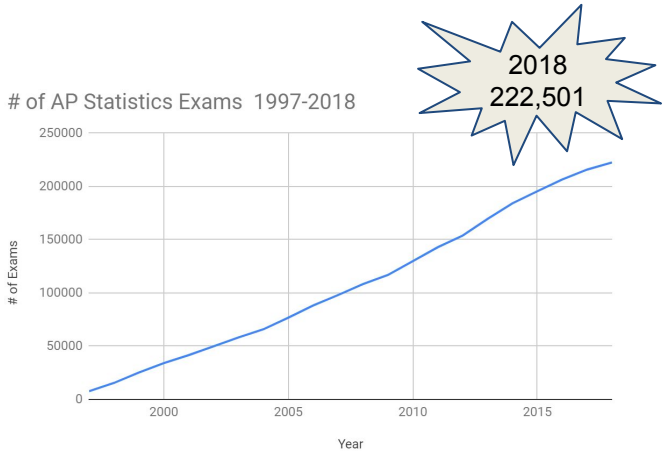


A closer look at AP Statistics

Last course guideline published 2010

Teacher Guide also provides info

Support networks and PD



Survey of current AP Statistics Teachers, Fall 2019

Of all the data sets students see in this course (AP Statistics), what portion of them meet the following criteria? n=428

#	AWA.FIELD	None		A few		About half		Most of them		All of them	
1	Real-world data	0.47%	2	7.71%	33	17.29%	74	61.21%	262	13.32%	57
2	Collected by students	3.04%	13	70.02%	299	20.84%	89	4.92%	21	1.17%	5
3	Contains between 100 and 1000 cases/observations/subjects	9.58%	41	51.17%	219	28.97%	124	9.58%	41	0.70%	3
4	Contains more than 1,000 cases/observations/subjects	35.36%	151	54.33%	232	9.13%	39	0.47%	2	0.70%	3
5	Contains between 2 and 10 variables/attributes	2.35%	10	44.13%	188	27.93%	119	22.77%	97	2.82%	12
6	Contains more than 10 variables/attributes	45.67%	195	42.62%	182	5.85%	25	5.15%	22	0.70%	3
7	Contains multiple types of attributes (quantitative, categorical, etc.)	3.98%	17	56.91%	243	24.36%	104	11.71%	50	3.04%	13

Content Analysis of AP Statistics

Examined Career Profiles (developed by ODI) of

- Big Data Enabled Specialists
- Data Practitioner

Data use in the Next generation Science Standards (report)

- Coded NGSS with tasks and skills from BDES profile

How well does the AP Statistics curriculum offer opportunities for students to experience aspects of tasks and skills needed by BDES and Data Practitioner careers?

Big Data Enabled Specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes

1. Defines the Problem (Skills 1A-1N)
2. Wrangles Data (Skills 2A-2N)
3. Manages Data Resources (Skills 3A-3K)
4. Develops Methods and Tools (Skills 4A-4I)
5. Analyzes Data (5A-5J)
6. Communicates Findings (Skills 6A-6K)
7. Engages in Professional Development (skills 7A-7J)

Data Practitioner in service of an organization and/or stakeholders, supports the data life cycle by collecting, transforming, and analyzing data, and communicating results in order to inform and guide decision-making.

1. Initiates the Project (Skills 1A-1I)
2. Sources the Data (Skills 2A-2I)
3. Transforms the Data (Skills 3A-3M)
4. Analyzes the Data (Skills 4A-4M)
5. Closes the Project (Skills 5A-5K)
6. Engages in Professional Development (Skills 6A-6I)

Tasks and skills in Career Profiles

Big Data Enabled Specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes

1. Defines the Problem (Skills 1A-1N)
2. Wrangles Data (Skills 2A-2N)
3. Manages Data Resources (Skills 3A-3K)
4. Develops Methods and Tools (Skills 4A-4I)
5. Analyzes Data (5A-5J)
6. Communicates Findings (Skills 6A-6K)
7. Engages in Professional Development (skills 7A-7J)

3 Skills well represented

Data Practitioner in service of an organization and/or stakeholders, supports the data life cycle by collecting, transforming, and analyzing data, and communicating results in order to inform and guide decision-making.

1. Initiates the Project (Skills 1A-1I)
2. Sources the Data (Skills 2A-2I)
3. Transforms the Data (Skills 3A-3M)
4. Analyzes the Data (Skills 4A-4M)
5. Closes the Project (Skills 5A-5K)
6. Engages in Professional Development (Skills 6A-6I)

3 skills well represented

Tasks & skills Well Represented in AP Stats

Big Data Enabled Specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes

1. Defines the Problem (Skills 1A-1N)
2. Wrangles Data (Skills 2A-2N)
3. Manages Data Resources (Skills 3A-3K)
4. Develops Methods and Tools (Skills 4A-4I)
5. Analyzes Data (5A-5J)
6. Communicates Findings (Skills 6A-6K)

17 Skills sparsely represented

Data Practitioner in service of an organization and/or stakeholders, supports the data life cycle by collecting, transforming, and analyzing data, and communicating results in order to inform and guide decision-making.

1. Initiates the Project (Skills 1A-1I)
2. Sources the Data (Skills 2A-2I)
3. Transforms the Data (Skills 3A-3M)
4. Analyzes the Data (Skills 4A-4M)
5. Closes the Project (Skills 5A-5K)

16 skills sparsely represented

Tasks & skills Sparsely in AP Stats

Big Data Enabled Specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes

1. Defines the Problem (Skills 1A-1N)
2. Wrangles Data (Skills 2A-2N)
3. **Manages Data Resources** (Skills 3A-3K)
4. Develops Methods and Tools (Skills 4A-4I)
5. Analyzes Data (5A-5J)
6. **Communicates Findings** (Skills 6A-6K)

49 Skills Absent

Data Practitioner in service of an organization and/or stakeholders, supports the data life cycle by collecting, transforming, and analyzing data, and communicating results in order to inform and guide decision-making.

1. Initiates the Project (Skills 1A-1I)
2. Sources the Data (Skills 2A-2I)
3. **Transforms the Data** (Skills 3A-3M)
4. Analyzes the Data (Skills 4A-4M)
5. **Closes the Project** (Skills 5A-5K)

35 skills absent

Tasks & skills Absent in AP Stats

Data Practitioner **in service of an organization and/or stakeholders, supports the data life cycle by collecting, transforming, and analyzing data, and communicating results in order to inform and guide decision-making.**

1. Initiates the Project (Skills 1A-1I)

   1D Identifies appropriate data

2. Sources the Data (Skills 2A-2I)

   2F. Cleans data (e.g. identifies outliers/errors)

3. Transforms the Data (Skills 3A-3M)

   3B. Splits data

   3C. Derives new variables

   3H. Changes data structure

4. Analyzes the Data (Skills 4A-4M)

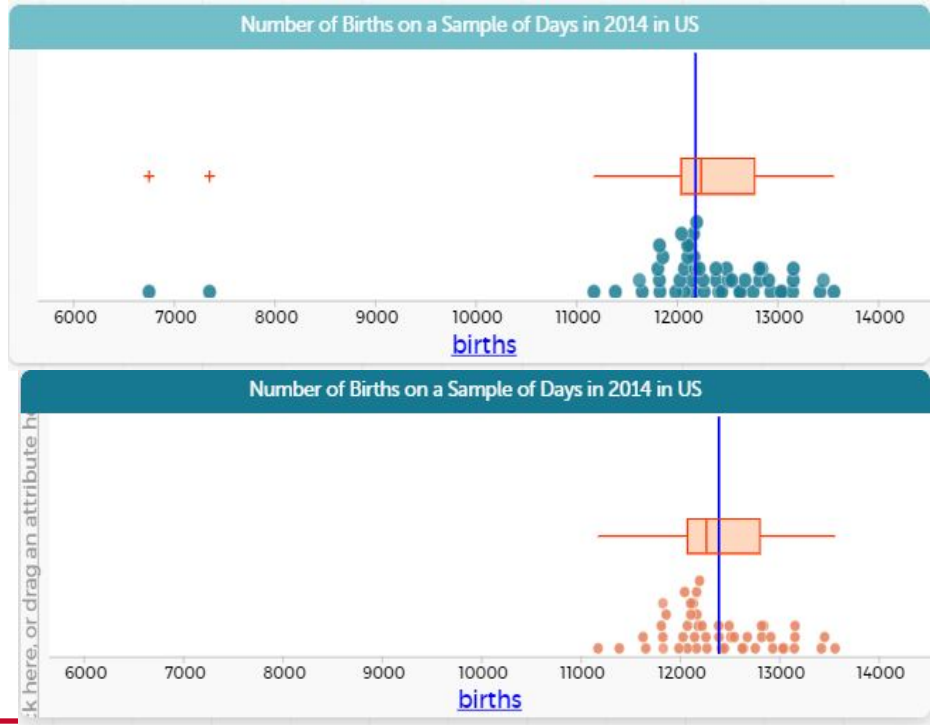
   4E. Performs data mining

   4F. Separates any anomalies

   4G. Interpret the results

Typical Data Experience in Schools

- Given small data set (or graph) of one (maybe 2) variables.
- Often graph by hand or on a graphing calculator
- Taught to think: "*there are outliers, they pull down my mean. Lets get rid of them and recompute mean.*"
- Report typical number of births per day as average of 12,387.



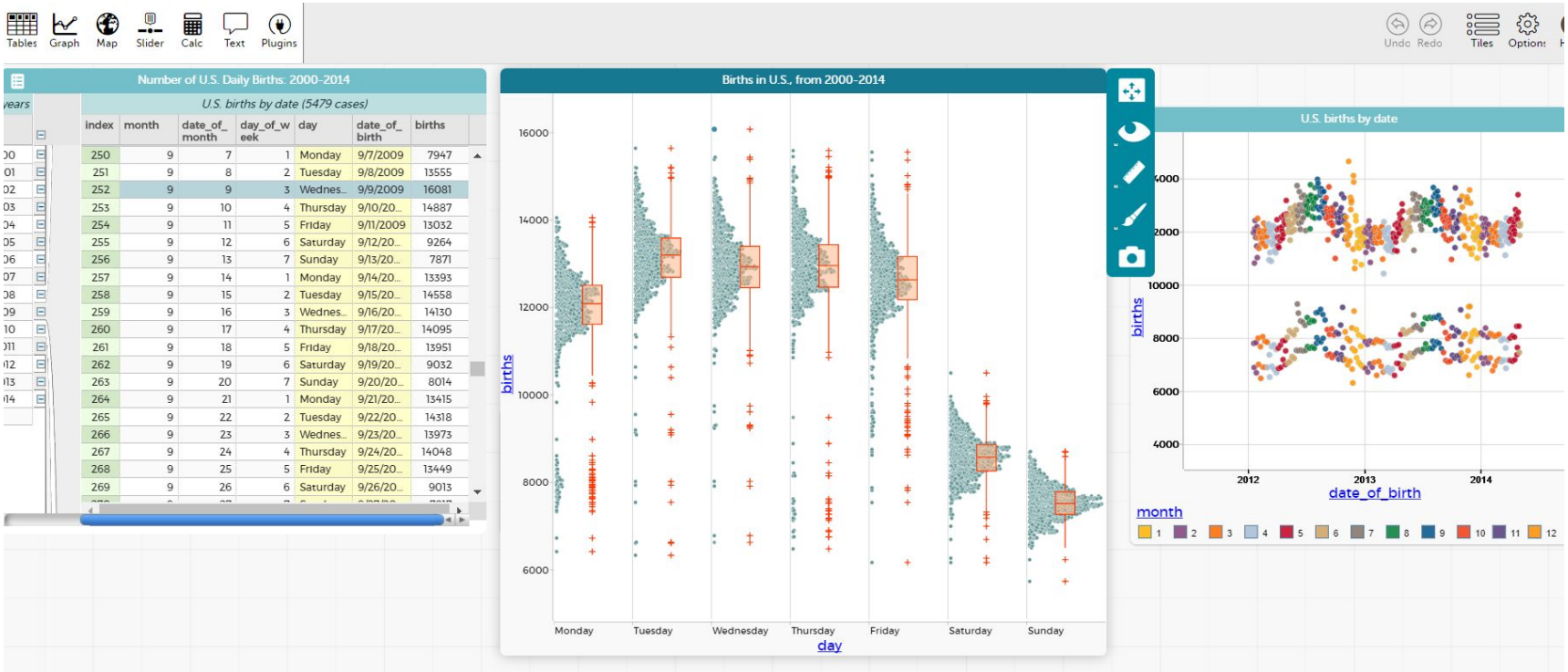
Questions about births per day

Are birthdays uniformly distributed throughout the year?

Is every birthday just as likely as any other birthday?

In the US, are there about the same number of births recorded on every day throughout the year?

bit.ly/2C3MJQU



Scheduling Births--A Convenience Event?

- Read articles related to births in the US (E.g., "time to have a baby"
- Have students write a report about their findings from the data investigation and compare with other sources.
- What do they wonder? Is the pattern from 2000-2014 the same if we look at births from 1950-1964? What if we examine births in other countries? How would source that data? What variables would you want in the data set?

Goal: Increase opportunities for learning data science skills and ways of working

within existing curricula with easy to use tools!

Design Principles for Bringing Data Science into Schools

- Data is real (collected by students or authenticated by teacher), multivariate (categorical & quantitative), “large”, and sometimes messy
- Data contexts are engaging to students

- Tools facilitate data moves, in tabular and graphical form
- Tools support visualizations and links among representations of data
- Tools to learn simple scripting/coding

- Tasks have multiple entry points for different levels of sophistication
- Tasks provoke curiosity and promote deep engagement with data and communicating results

Data Visualizations and Media

https://www.nytimes.com/column/whats-going-on-in-this-graph

Week of Feb 13: This graph shows the relative size of military budgets by country in 2015. It originally appeared elsewhere on NYTimes.com.

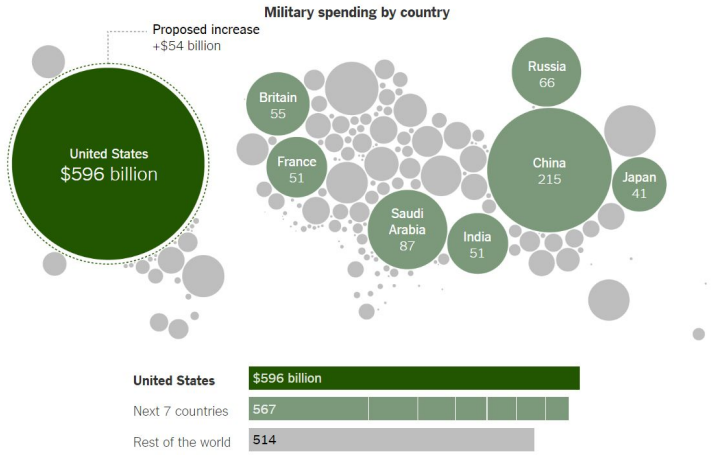
After looking closely at the graph above, think about these three questions:

- **What do you notice?**
- **What do you wonder?**

What are you curious about that comes from what you notice in the graph?

- **What might be going on in this graph?**

Write a catchy headline that captures the graph's main idea. If your headline makes a claim, tell us what you noticed that supports your claim.



NY Times Learning Network FREE Webinar March 20th 4pm
Explore how to teach and learn across subjects with "What is in this Graph"

A few videos to ponder:

Data Investigations with CODAP in the Classroom

Middle school Students' work with roller coasters in CODAP (first time using!)

AP Stats students examine linear models with vehicle data

Think about: What skills or ways of working with data did these students get to experience? What is still missing (from the lessons shown here) that we can do better with?

Contact Us!

hollylynne@ncsu.edu

HI-RiSE: A Hub for Innovation and Research in Statistics Education

http://hirise.fi.ncsu.edu

pbaumgartner@rti.org

https://www.rti.org/expert/peter-baumgartner