

Baseline Professional Learning Handout

M211: Data Management and Visualization

Baseline Professional Learning Handout	1
Selected Common Core High School Standards for Mathematics	2
Data Science Big Ideas	2
M211 Content and Practice Expectations	3
M211 Content and Practice Expectations and Indicators	4
Let's Investigate	6
The Animals Dataset	6
Data Visualization & Misleading Data	7
Let's do some math! Round 1	8
Task 1: Me & the Bees	8
Learning Principles	10
Let's do some math! Round 2	11
Task 2: Household Water	11
Criteria for Success	13

Selected Common Core High School Standards for Mathematics

Summarize, represent, and interpret data on a single count or measurement variable

1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
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.
3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
4. Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

Summarize, represent, and interpret data on two categorical and quantitative variables

1. Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.
2. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
 - a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.
 - b. Informally assess the fit of a function by plotting and analyzing residuals.
 - c. Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

1. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
2. Compute (using technology) and interpret the correlation coefficient of a linear fit.
3. Distinguish between correlation and causation.

Data Science Big Ideas

M211 Content and Practice Expectations

211.a	Identify features that make some graphs of quantitative and categorical data misleading.
211.b	Identify the subjects and variables in a data set.
211.c	Interpret and compare graphical representations of quantitative data and categorical data.
211.d	Describe key features of raw quantitative data and categorical data.
211.e	Create graphical representations of quantitative data and categorical data.
211.f	Prepare raw data for predictive modeling by organizing, cleaning, summarizing, aggregating, and feature engineering.
211.g	Understand where data comes from and how to collect primary and secondary sources of data.
211.h	Perform data practices (collecting, generating, analyzing, and disseminating data) ethically, and evaluate data practices on their use of ethics.

M211 Content and Practice Expectations and Indicators

Content and Practice Expectations	Indicators Choose an artifact where you...
211.a: Identify features that make some graphs of quantitative and categorical data misleading.	i. demonstrate understanding that axis scale, truncated axes, and pictographs can exaggerate differences between values.
211.b: Identify the subjects and variables in a data set.	i. given a data visualization, set of raw data, or study description, identify who or what the subjects are.
	ii. given a data visualization, set of raw data, or study description, identify what was collected or observed from the subjects.
	iii. demonstrate understanding in context of what the rows and columns represent in a set of raw data (i.e., rows represent subjects and columns represent variables measured or collected).
211.c: Interpret and compare graphical representations of quantitative data and categorical data.	i. summarize key features of graphs in context such as shape, center, outliers, variability, and direction of trends, if applicable.
	ii. compare key features of graphs in context.
211.d: Describe key features of raw quantitative data and categorical data.	i. calculate measures of center and measures of spread, and identify outliers and other summary statistics (such as maximum, minimum, and skew) of large data sets and explain their meanings.
211.e: Create graphical representations of quantitative data and categorical data.	i. select appropriate types of visual representations for the type of data.
	ii. use appropriate tools to create visual representations of large data sets.
211.f: Prepare raw data for predictive modeling by organizing, cleaning, summarizing, aggregating, and feature engineering.	i. merge multiple data sources and explain why this is necessary.
	ii. make decisions to remove data observations or fill in missing values of data and justify these decisions.
	iii. filter and/or sort data to identify key features of data.
	iv. aggregate data if necessary to create new variables for the modeling process and justify the need for these variables.

Content and Practice Expectations	Indicators Choose an artifact where you...
211.g: Understand where data comes from and how to collect primary and secondary sources of data.	i. distinguish between primary and secondary sources of data.
	ii. demonstrate understanding of types of sampling methods such as voluntary response, random sampling, stratified random sampling, systematic random sampling, and web scraping.
211.h: Perform data practices (collecting, generating, analyzing, and disseminating data) ethically, and evaluate data practices on their use of ethics.	i. define and apply terms surrounding data ethics, such as privacy, targeted advertising, location services, readability, and cookies.
	ii. demonstrate understanding of choosing sampling methods that responsibly prevent bias.

Let's Investigate

The Animals Dataset

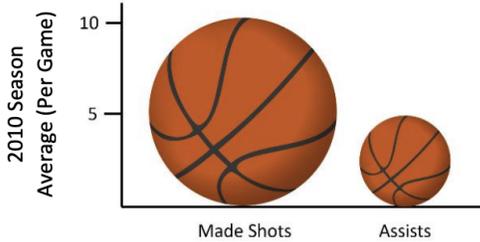
1. Open the [Animals Spreadsheet](#), specifically the pets tab. This table contains data from an animal shelter, listing animals that have been adopted.
 - a. What do you notice about this dataset?
 - b. Describe how the spreadsheet is organized. Consider the number of data rows, data columns, variables, and the type of data – categorical or quantitative.
2. Reviewing this data set may have sparked some questions. What do you wonder about this dataset?
3. There is an art to asking the right questions. Data Scientists take their time to think about the kind of questions that can and cannot be answered. Most questions can be broken down into one of four categories. Lookup questions can be answered by looking up a single value in a table. Compute questions can be answered by computing an answer across a single column. Relate questions answer whether or not there is a relationship between multiple columns. Can't answer are questions that cannot be answered based on the available data.
 - a. Create an example for each type of question.
 - b. Look back at what you wondered. Identify the type of question you may have posed.

Source: [Bootstrap](#)

Data Visualization & Misleading Data Task

- For each example, describe why the data visual might be misleading. Then, sketch a more transparent graph of the same data.

Example 1: "Kobe was a ball hog"

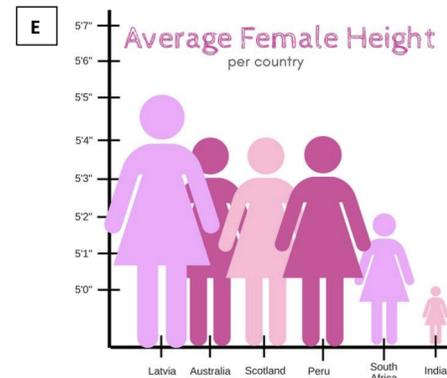
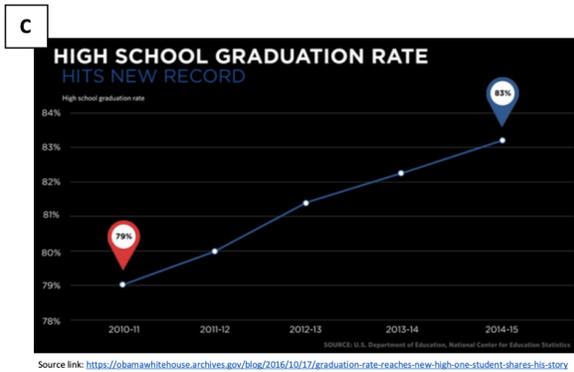


Example 2: "Chevy builds the most dependable trucks"



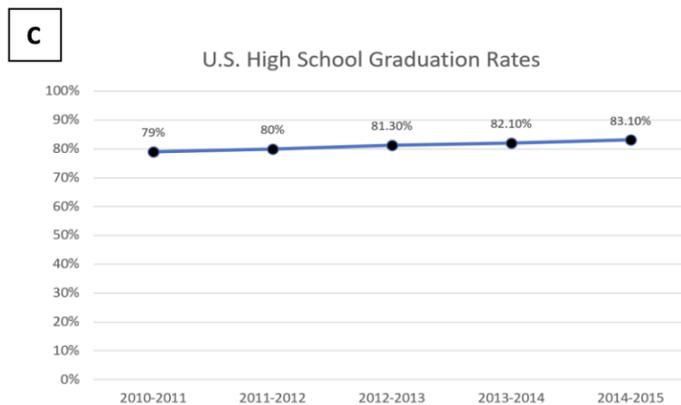
Example from TED-Ed: "How to spot a misleading graph"

- Graphic 'C' was published on the White House's official blog during the Obama Administration. It uses national school data prepared by the Department of Education. Graphic 'E' was shared on Twitter and has an unknown origin. For each graph, answer the following questions: Is the visual misleading? Why or why not? If it is misleading, how would you change it?

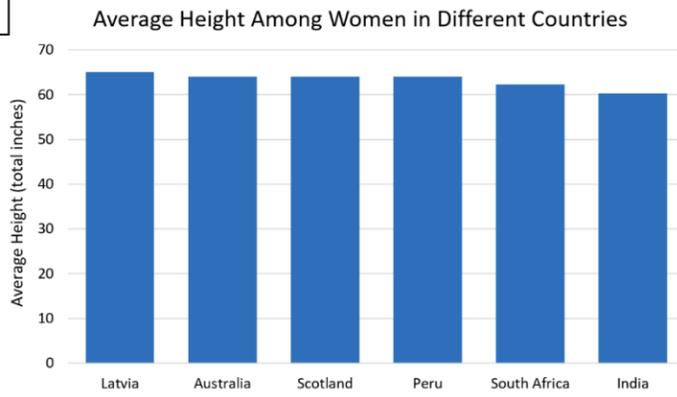


Graphic found by Sabah Ibrahim on [twitter](#)

- This page presents the same data, but with adjusted graphs. For each graph, answer the following question: What makes this version of the graph less misleading? Explain.



E



Source: [Skew the Script](#)

Let's do some math! Round 1

Task 1: Me & the Bees

With your partner, you will write a business report for Me & the Bees. During this time, you will analyze data, synthesize your discoveries, and communicate your findings. You will use the tools and strategies you learned for working in a spreadsheet to continue to look for patterns in the data. You will then write up your findings in a business report. In your report to Me & the Bees, you will respond to the question, "What are your recommendations for Me & the Bees? Make sure to include evidence."

Your business report should include:

- An introduction
- Discussion of findings with visual evidence as support
 - Visuals along with your analysis and synthesis of those visuals
- Recommendations and conclusions
- Business report writing features: concise, clear visuals, color, and graphs as supporting evidence

Prepare to share your findings with the class.

Adapted from YouCubed.org

Sample Learning Experience

To launch the lesson, ask the class if any of them are business owners or if they know any business owners. Share with them the idea that there is no age requirement to starting and running a business. Share with them the story of Me & the Bees through [this video](#), and explain that companies like Me & the Bees will have a team in their company or an outside consulting firm analyze their data to give recommendations on how to improve their business. Let students know that they will be creating a business report for Me & the Bees. They will be given [data to explore](#), model (find patterns within), and analyze. Finally, they will write a report to communicate their findings and describe recommendations for the company to better understand and improve their business.

In the week leading up to presentations, have class discussions using the following questions to guide their thinking:

- *What is the [spreadsheet](#) about?*
- *What does each row represent? What about each column?*
- *Brainstorm some things you could learn about the Me & The Bees business from the data in the spreadsheet.*
- *What are some questions you still have about it?*
- *What are additional things about the business you want to know?*
- *What steps would you need to take to merge the data into one spreadsheet?*

Give students some individual work time to explore the Me and the Bees Spreadsheet. Share this video to assist students in cleaning and preparing this data for their reports. Then have students pair with a partner to share their approach to this investigation. Provide partner pairs access to tools (such as chart paper, Google Slides, Sheets, Docs) to create a display of their findings. As students work, circulate and identify approaches to present to the whole class. Be mindful to emphasize different approaches to providing visual evidence for both univariate and bivariate data (bar charts, pie charts, boxplots, etc.). Encourage students to ask questions of each other as solutions are presented.

Learning Principles

Engage with cognitively demanding tasks in heterogeneous settings (LP 1). Students should be given opportunities to grapple with multistep, non-routine tasks that promote mathematical rigor. These experiences should be differentiated so that all students engage in appropriate challenges—for example, through tasks with multiple entry points and solution pathways. These experiences should continue to integrate knowledge and skills developed in grades 6-8 at the level of sophistication of high school mathematics.

Engage in social activities (LP 2). Students should have opportunities to work independently and to communicate with one another about mathematics by engaging in collective and collaborative learning activities. Explaining and having opportunities to revise one’s thinking has excellent value in solidifying one’s knowledge.

Build conceptual understanding through reasoning (LP 3). Students should be given the opportunity to reason, justify, and problem solve with critical thinking, reading, writing, speaking, and listening. Through reasoning and work with multiple representations, students learn why procedures work and build a conceptual understanding of key mathematical ideas.

Have agency in their learning (LP 4). Students should be able to choose tasks and learning experiences that align with their interests and aspirations. All students have rich and varied experiences and home lives. Learning mathematics should bring students’ identities and interests to the fore and build on the strengths that they bring to the learning space.

View mathematics as a human endeavor across centuries (LP 5). Students should understand that mathematical ideas emanated over time from civilizations around the world and have opportunities to explore these contributions to mathematics. Students should develop an appreciation of mathematics as a human endeavor: one in which they feel a sense of belonging, where they see themselves as mathematicians, and one that offers opportunities to broaden their ideas about what mathematics is, how it is used, and who it is for.

See mathematics as relevant (LP 6). Students should engage with mathematics in ways that authentically involve real-world situations. Problem-solving contexts should allow students to perceive mathematics as a tool for addressing the questions that arise in everyday life, as well as the ways it can model our world and address global economic, social, and environmental challenges. Students should also engage with mathematics in ways that connect to academic disciplines and future careers by doing the mathematics used by artists, designers, engineers, and other professionals.

Employ technology as a tool for problem-solving and understanding (LP 7). Research indicates that technology is a powerful tool for learning deeper mathematics by improving calculation efficiency and enabling more sophisticated analysis. Students should learn to use technology, with emphasis put on widely used tools and software, such as calculators and spreadsheets, to make sense of models. Technology use should not be limited to supporting “doing mathematics,” but should also be used as a tool for displaying and communicating results to appropriate audiences.

Task 1: Me & the Bees Student Work

Example 1:

Student work goes here

Example 2:

Student work goes here

Let's do some math! Round 2

Task 2: Household Water

Provide students with American Community Survey Data and a prompt as follows:

In this activity we will collect bivariate data to see if there is a relationship between two variables: number of gallons of water used and how many people are in a household.

After a few days or a weekend passes to allow students to collect water use data, resume the activity:

Your group will use the Colab program, [Box Plots – Explore Your Own Data](#), to tell the story of the variable(s) from the perspective of your local data. Then you'll explore the similarities and differences in your local data set from your survey and the state data set from ACS (American Community Survey Data). For each variable, your group will consider the reasons why the similarities and differences might exist and if they point to the same need.

Adapted from YouCubed.org

Sample Learning Experience

In pairs or groups, ask students to brainstorm a list of the different ways that water is used in their homes or living spaces. Share this as a class and generate a class list. Given this list, ask the class what they would estimate as their daily water usage in their household or living space in gallons of water per day. In their groups, students can make an estimate of household water usage based on one group member's data who is comfortable sharing with the group, as well as this group member's number of people within the household or living space. Have groups share their estimates with the class and record them.

Students are collecting this bivariate data to see if there is a relationship between the two variables: number of gallons used and how many people are in a household or living space. You can have a conversation with the class. Ask them:

- *Do you think there will be a relationship between these two variables?*
- *Why or why not? What relationship would you expect?*
- *What other factors might play into water usage?*

Let them know that this is an example of bivariate data. Give students a few days or a weekend to estimate their household water usage. Teachers may copy [this form](#) to collect data from students.

After some individual work time on the prompt where students develop an approach, encourage them to share and compare approaches. Look for student work that provides very different viewpoints about the data to bring up during the class discussion.

Bring the class together for a whole group discussion on what they noticed about the comparison between survey and state data. Invite students to share how the local data gave them new perspectives on their variable(s) and need or how there was no change at all. Ask them to think about why in some cases the data was very similar and in other cases there was a difference in the local and state level story.

Consider using this discussion to encourage wonderings about who is represented in each data set and how the difference we see in the local- and state-level data is a result of who was surveyed. Something to keep in mind is that not only does it matter who was in the survey but also the population the survey is meant to represent. This conversation will encourage students to be curious about the idea and ethics of sampling which they explore next, improving upon their answers from this activity to write a letter to their local state government.

Task 2: Household Water Student Work

Example 1:

Student work goes here

Example 2:

Student work goes here

Criteria for Success

Conference and Provide Revision Support	Accept with Revision	Accept
The student's artifact shows an emerging understanding of the expectations of the indicator(s). After conferencing and additional instruction/learning, the student may provide a revised or different artifact as evidence of the indicator(s).	The student's artifact is approaching a full understanding of the expectations of the indicator(s). The artifact may contain execution errors that should be corrected in revision. The student may revise the selected artifact or submit a different artifact.	The student's artifact demonstrates evidence that they have met the expectations of the indicator(s).