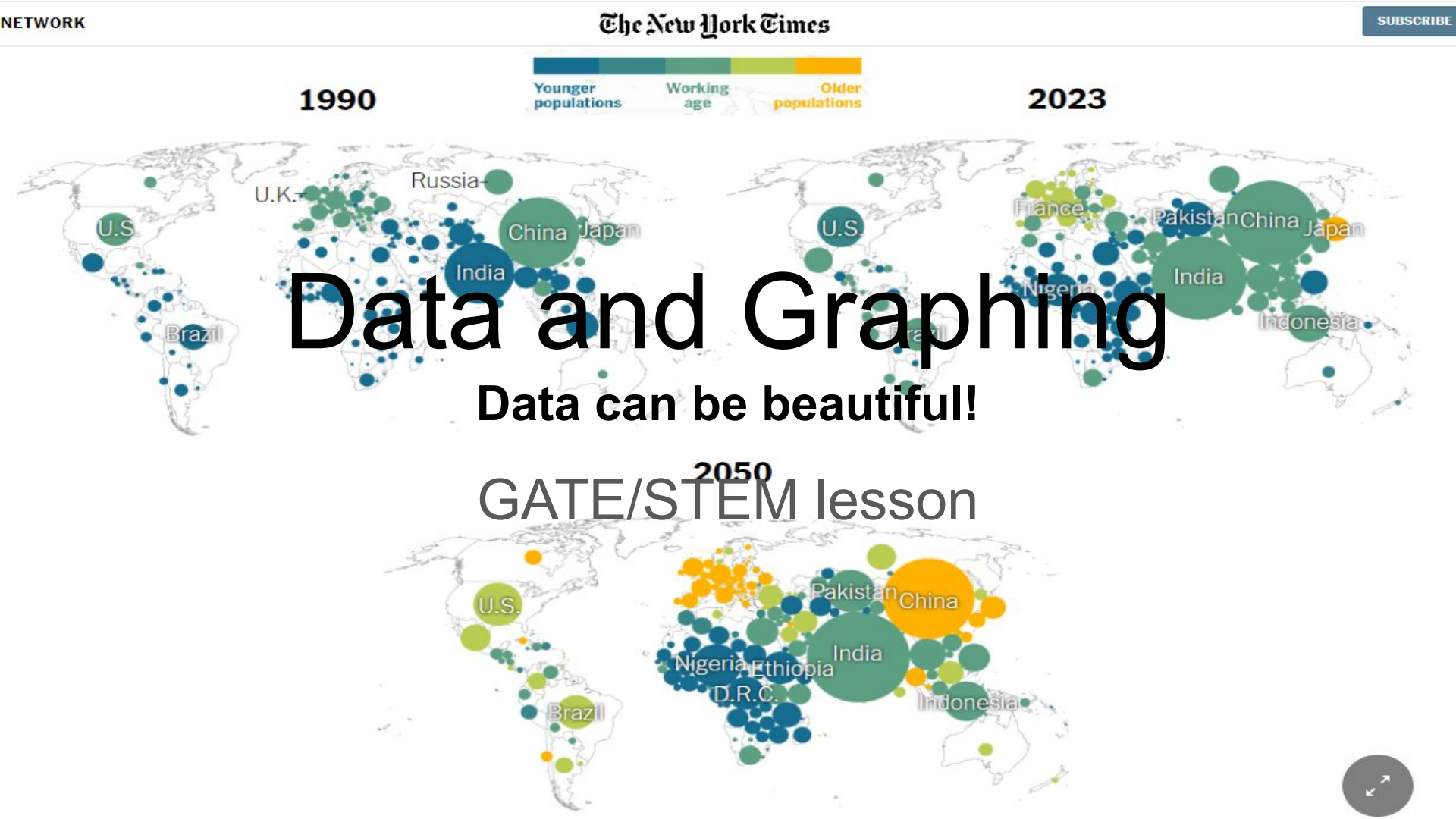


Data and Graphing

Data can be beautiful!

GATE/STEM lesson

Standards addressed:

7.SP.A.2 Use random sampling to draw inferences about a population.

Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.

7.SP.B Draw informal comparative inferences about two populations.

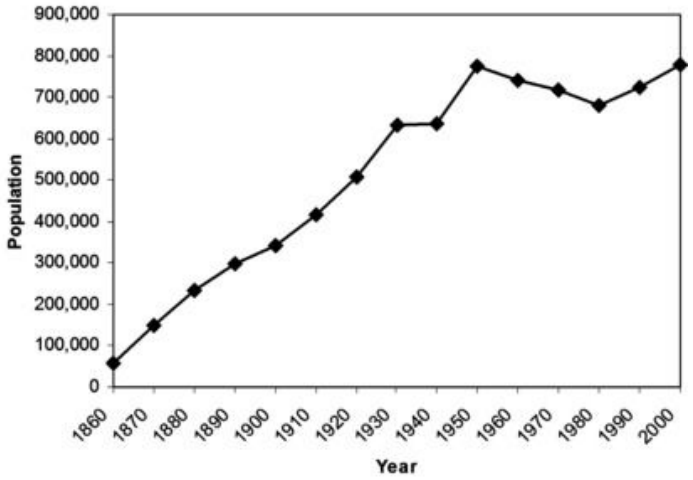
Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

Cross cutting concepts: patterns; SEPs: analyze and interpret data, engage in argument from evidence, obtain/evaluate/communicate information.

Why are we doing this?

Graphs present information visually, efficiently, powerfully. Trends/patterns readily seen. What did you look at first? What questions did you ask?

Year	Population
1860	56,802
1870	149,473
1880	233,959
1890	298,997
1900	342,782
1910	416,912
1920	506,676
1930	634,394
1940	634,536
1950	775,357
1960	740,316
1970	715,674
1980	678,974
1990	723,959
2000	776,733

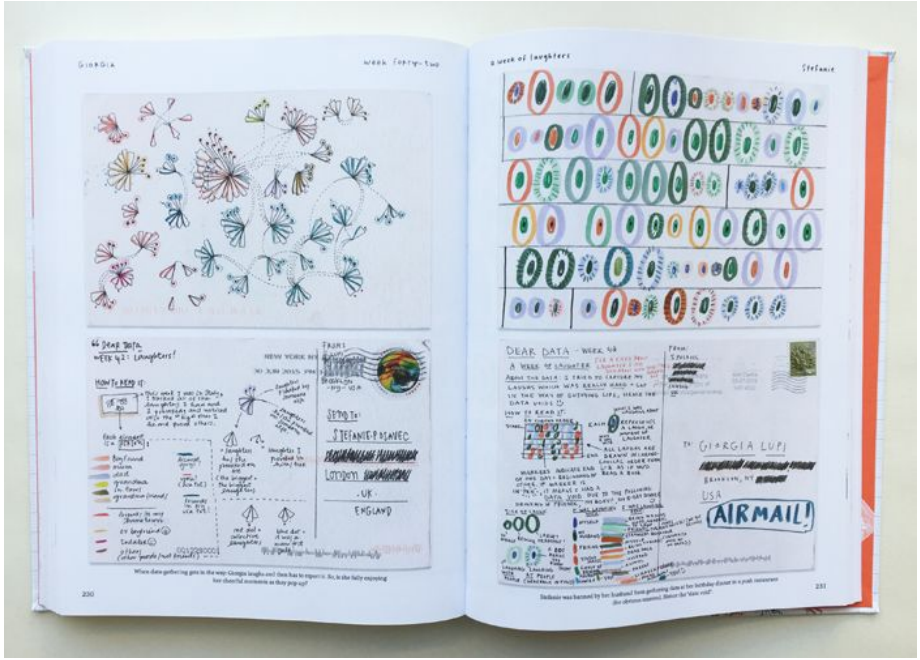


But that was still ‘boring…’ An interesting graph pauses the viewer, makes them think, ask and answer questions. Words as a graph (left), change over time graph in a photo (right).
https://www.nytimes.com/column/whats-going-on-in-this-graph?page=5

gas 904 milk 455 meat 402 beef 298
chicken 265 eggs 205 cars 150 hamburger 138 ground
beef 134 bread 131 haircut 127 groceries 123 food 116 vegetables 107
steak 94 produce 93 bacon 88 chicken wings 82 fruit 82 clothing 75 soda 73
cheese 66 beer 64 lumber 64 toilet paper 63 pork 62 restaurants 61 electricity 60
cereal 59 fast food 56 chicken breast 53 coffee 51 dairy 45 takeout 40 houses 39 used
cars 36 utilities 36 rent 35 water 35 fish 33 chips 29 ice cream 29 dog food 27 pizza 27 cigarettes 26
fresh vegetables 26 homes 24 alcohol 23 paper products 23 everything 22 frozen food 21 seafood 21 butter 20
electronics 20 new cars 20 shoes 20 snacks 20 wood 20 flights 19 pet food 18 apples 17 furniture 17 cat food 16 fresh
fruit 16 paper towels 16 wine 16 yogurt 16 insurance 15 mcdonald's 15 candy 14 canned goods 14 cleaning supplies 14
delivery 14 natural gas 14 apartment 13 drinks 13 fresh produce 13 juice 13 laundry detergent 13 medicine 13 paper goods 13
appliances 12 poultry 12 soft drinks 12 orange juice 11 propane 11 salmon 11 diapers 10 ground meat 10 heating 10 heating oil 10
phone 10 sausage 10 wings 10 avocado 9 computers 9 dining out 9 shrimp 9 travel 9 car repair 8 chicken thighs 8 cleaning
products 8 coke 8 hotels 8 household products 8 hygiene products 8 oil change 8 strawberries 8 trucks 8 tv 8 bananas 7 beef steak 7
cell phones 7 cookies 7 deli meat 7 dollar store 7 liquor 7 lunch meat 7 movies 7 oil 7 pasta 7 potatoes 7 red meat 7 shampoo 7

Inspired by:

https://www.dear-data.com/thebook



Objective(s):

Students will generate and record data.

Students will create 2 different graphs to show data: line, bar, or pie chart and one that shows data using central measures of tendency and in a visually effective and artful way.

Students will analyze the class data.

Students will interpret the data set.

Students will employ skills with a new data set.

Materials:

PBS Learning Media: How Long is a Minute?

https://az.pbslearningmedia.org/resource/mgbh.math.sp.minute/how-long-is-a-minute/

Use other data aggregation sources is you want! Suggestions:

https://www.abcwua.org/wp-content/uploads/Education/7_MightyMacroinvertebrates.pdf

https://datanuggets.org/

https://www.sciencejournalforkids.org/articles/lesson-ideas/graphing-scientific-data-for-students/

Notebook

Paper, art supplies (colored pencils, sharpies, etc.)

Dear Data by Giorgia Lupi and Stefanie Posavec

Hands-on/what to do

- Students go to PBS Learning media site and do “How Long is a Minute?” and record their guesses.
- Record what method they used to make their guess (counting, sing a song, just plain guess, etc.)
- Class data. If the class is large, you may want to consider breaking into two or three groups.
- Students use traditional method to graph data, first by hand, then using an online tool such as https://nces.ed.gov/nceskids/createagraph/ or Canva or other app
- Students create a visually engaging, informative graph using data for one of the central tendencies of data.
- Students will compare their ‘traditional’ graph with one they create inspired by Dear Data
- Mini lesson on central tendencies of data
- Write an analysis and interpretation based on their graphs.

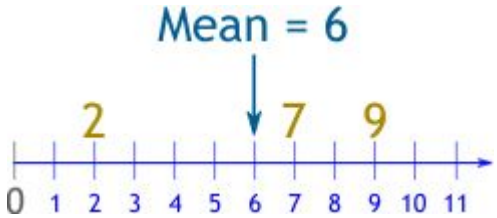
Measures of central tendency

Mean: also called average. What does it tell you?

Average value of a distribution of figures, such a the mean of test scores

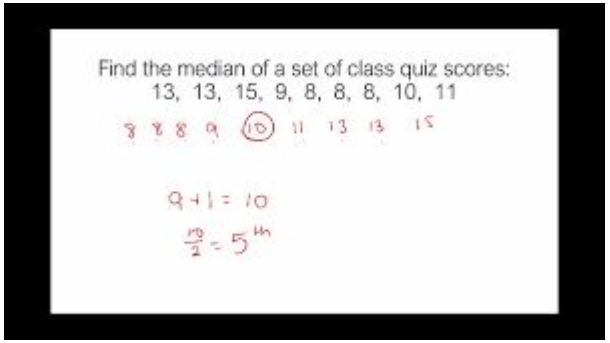
Advantages: commonly used, takes into account all values, useful for comparison, all data sets will have one mean, easy to calculate

Disadvantages: sensitive to outliers, can make data seem skewed



Median

How to calculate:



What does it tell you? Median separates the higher values from the lower values.

Advantages: easy to compute, not distorted by outliers, good measure of center

Disadvantages: does not use all data in calculation

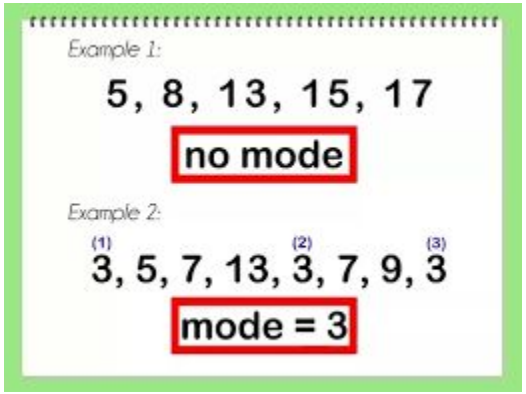
https://www.statology.org/advantages-disadvantages-of-median-in-statistics/

Mode

The most frequent value. Can have more than one mode or none at all. What can it tell you? Use when looking for ‘most popular’ category.

Advantages: can used for numerical and non-numerical data

Disadvantages: may not reflect central tendencies, absence of mode or more than one mode may not reflect center of distribution



Range

Difference between highest and lowest value. It tells you the distance between the values in a data set.

Advantages: identify outliers, shows variation in data set, allows statistician some guidelines in analysis (if value > range may ‘trigger’ warning)

Disadvantages: sensitive to extreme values even though rest of data is clustered

Difference between analysis and interpretation of data or graphs.

Analysis: looking for patterns or trends, what might the patterns/trends mean?

Interpretation: the story of the data, explain the patterns/trends that were uncovered.

Interpretation is like stitching a mystery together: using clues, evidence, background information, etc. to review data and make conclusions. This should lead to questions that prompt further investigation. What decisions will be based on interpretation? What are next steps? Can we make better decisions based on the data? Find trends→take action. How to better allocate resources, etc.

Dear Data Graphing

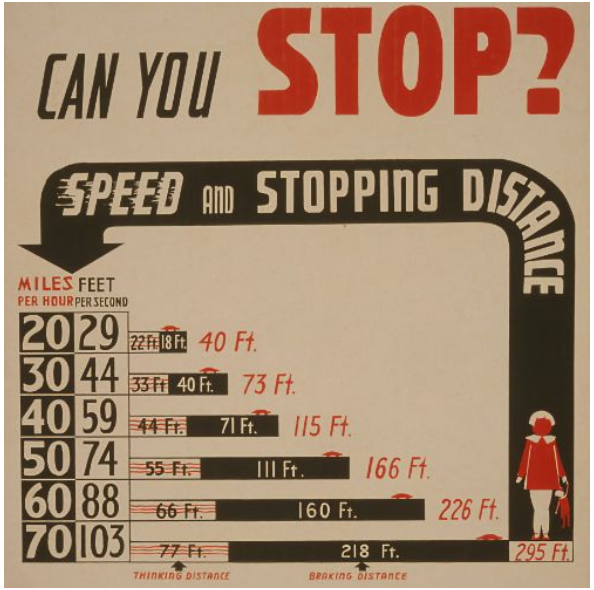
Students choose how they are going to portray data (mean, median, mode, range) and create a graph using all participant's data (not just their own).

Topic of data is time, but it's not necessary to incorporate that into the graph.

Students will also write their analyses of the data, using the graph to explain what information is being conveyed.

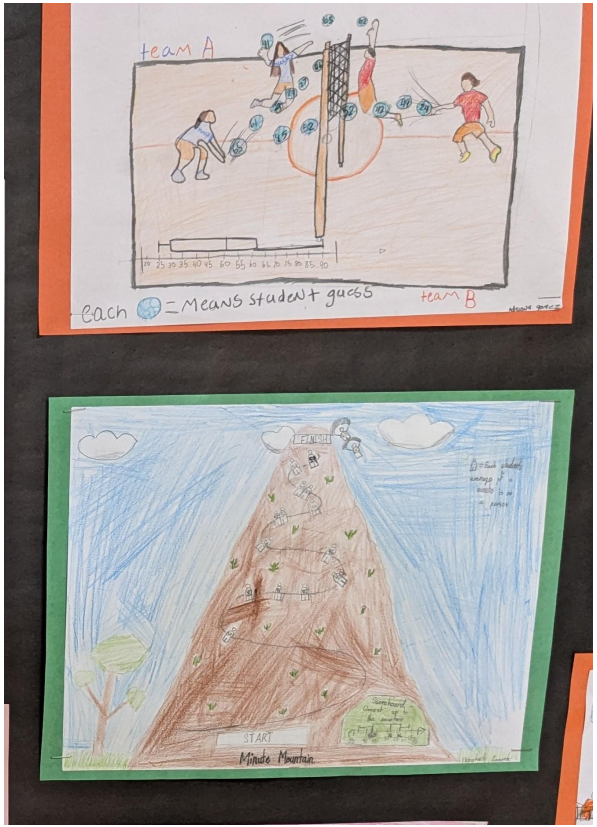
Students will then offer their interpretation of their graphs, data.

This is not a novel idea! Visual displays of data are powerful and effective ways to communicate information or a message.

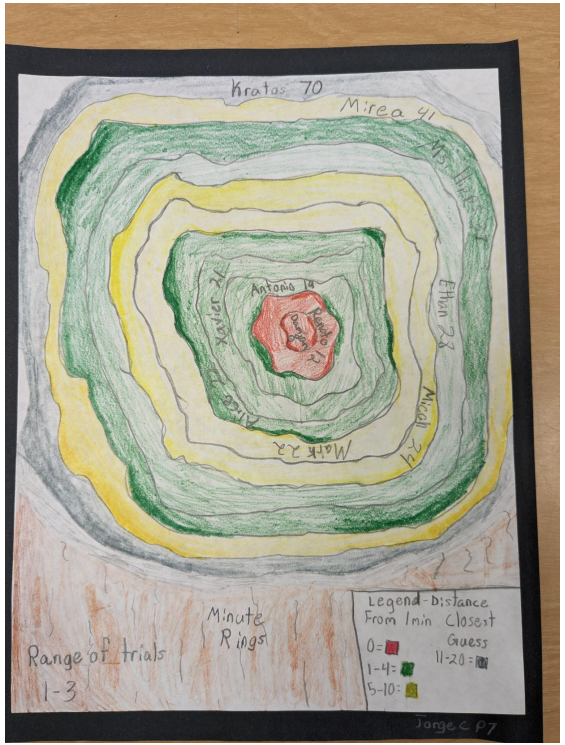


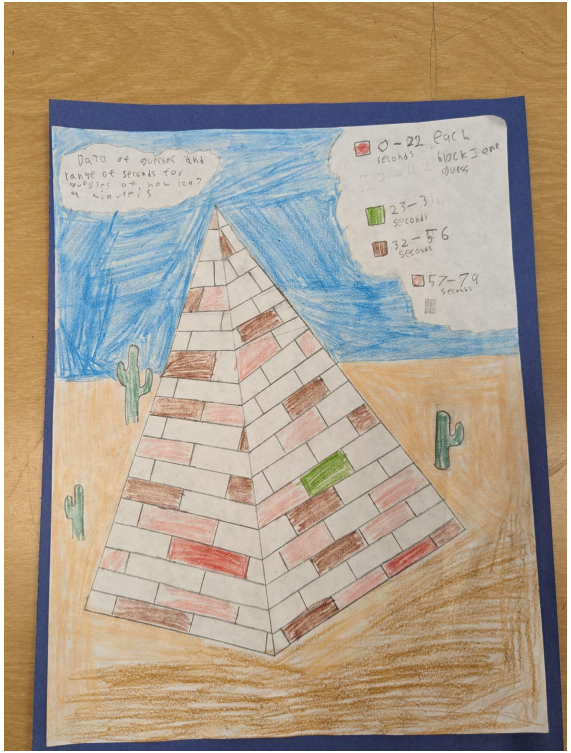
https://www.loc.gov/classroom-materials/charts-and-graphs/

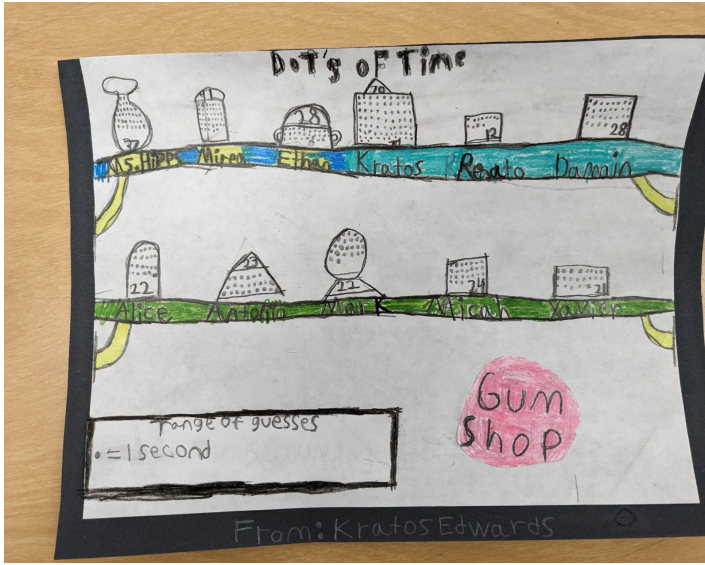
Example of student work. Graphed by method of ‘waiting’ for a minute, averages of those who used each method by trial.



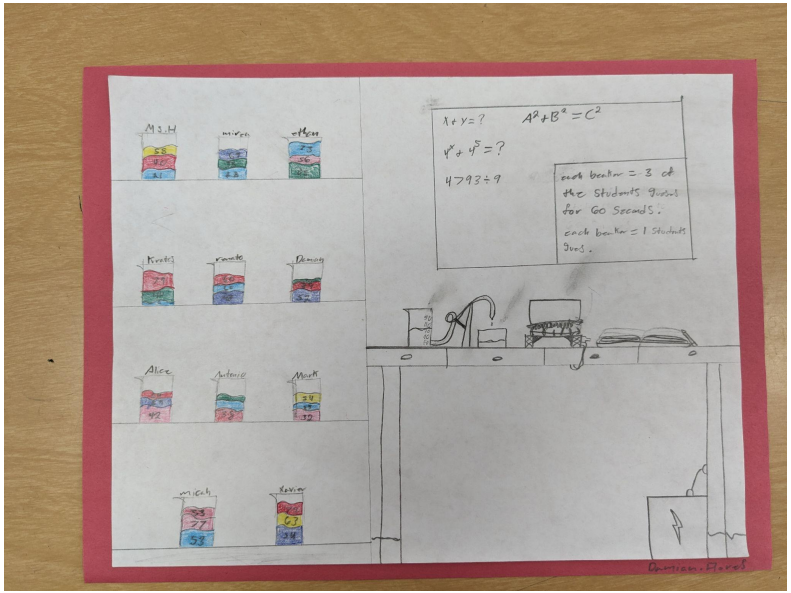
Student graphs showing range of data: better guesses closest to center of tree ring or each person's range of data.

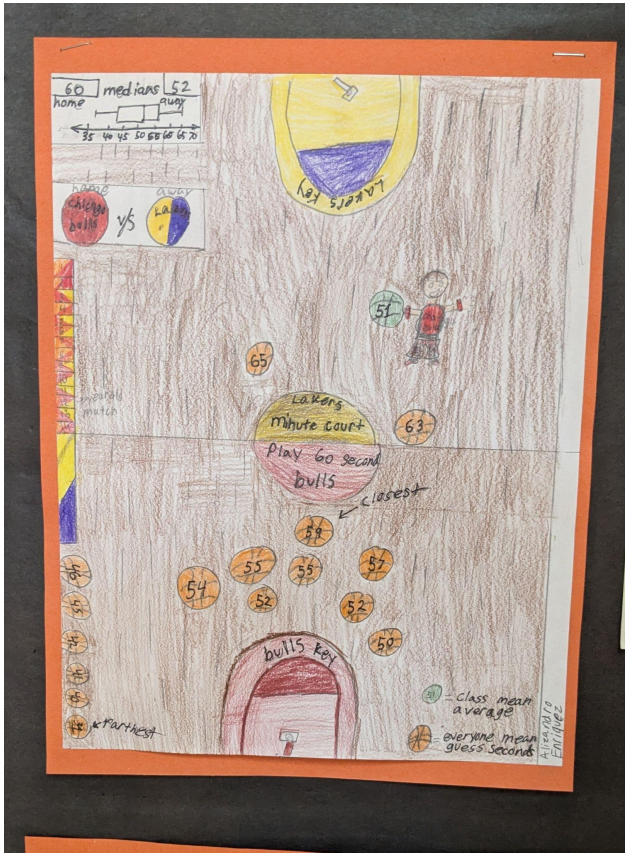




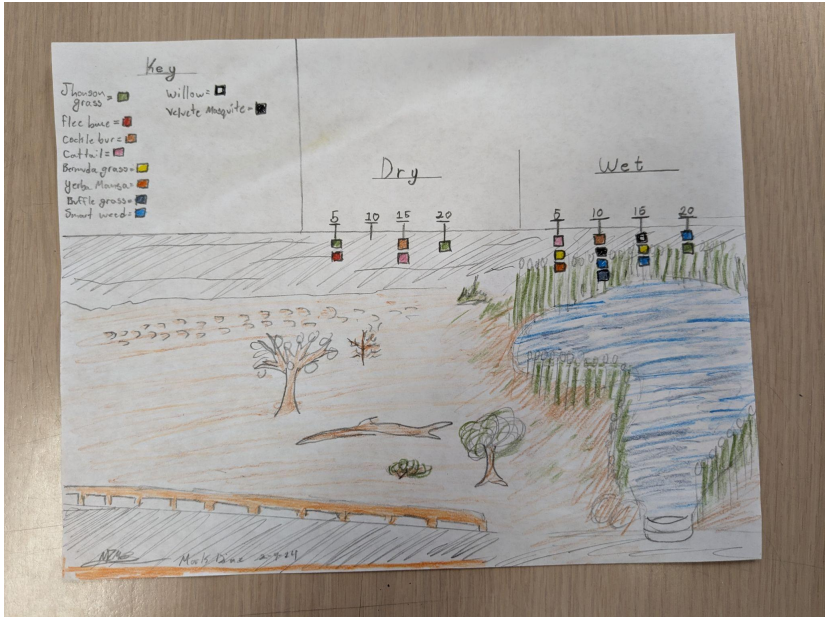


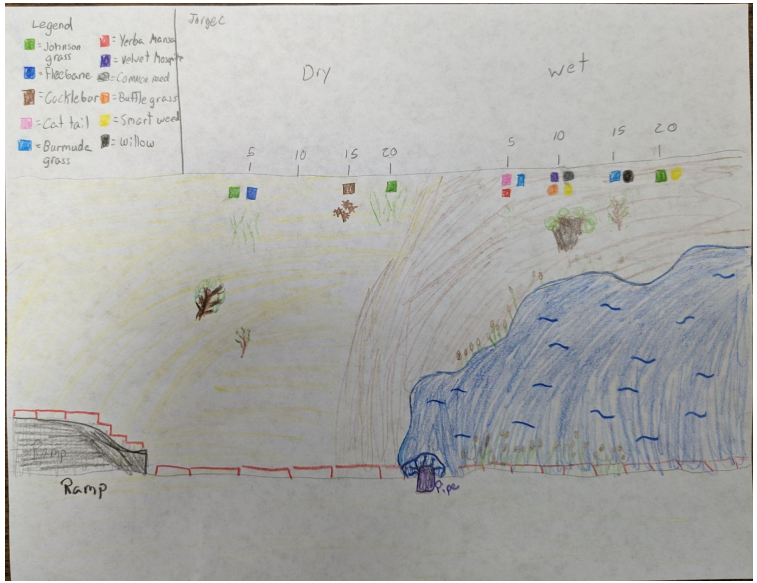
Student graph showing all data.





Graphed data from Santa Cruz River plant survey: dry vs. wet





Example student analysis and interpretation.

Analysis

Average:

During the first trial the average guess was about 41.45 sec, the second was 51.33 sec, & finally for trial 3 the average was 63.81 seconds.

Range:

The first attempts range was approximately 54 sec, which is not even close to the goal. The second trial was lower at a 45 sec range and the last trial was close to the goal at 26 seconds.

Mode:

The first attempts mode was 45 sec, the second 59 sec, and finally the third at 60 seconds.

Method:

For counting the average was about 48.125 sec, for singing it was 41.6 sec, & for guessing it is about 51.33 seconds.

Interpretation

Average: This means that as the trials went on the students arrived closer to the goal. This is because after every trial they got used to the challenge.

Range: The students performed more poorly as the methods most likely faltered. This is because they overestimated their gut feeling. This is also a bad measurement because even 1 small number can bring it down.

Mode: The mode was closer to the goal than the range and is because its more generalized. This is because mode is everyone's guesses and is considered like a popularity contest. It can be bad because even if it got guessed 2 times it overthrows the other numbers.

Method: The averages were scattered and varied between the people. Guessing gives you the best chance because it is more generalized. Guessing is more randomized and does not count on speed and or tempo like the other methods.

Optional assessment or concluding activity

Graph this data. You may not use a bar graph or pie chart. Provide an analysis and interpretation for the data as if you are the cafeteria manager at a school.

Favorite cafeteria food	Number of students	Percentage of students	Circle degrees (% x 360)
Pizza	8	40%	144
Hamburgers	5	25%	90
Chicken Nuggets	3	15%	54
Other	4	20%	72
Total	**20**	**100%**	**360**

Table 1: Table of favorite cafeteria foods

Data art rubric

Criteria	Exceeds	Meets	Approaching	Not met
Data	Data is accurate, clear, embedded	Data is accurate, clear, but inserted/after thought	Data is present but difficult to find. Unclear what data is.	Data is not present or unclear, inaccurate
Graphing	Relationship between variables is clear, show relationships or trends	Relationship between variables is present but lacks clarity	Relationship between variables is unclear or random	No relationship between variables is present or conspicuous
Art	Clear, neat, detailed, integrated with data. Theme is present. Composition, color, scale present.	Clear, neat, integrated with data. Composition and color present.	Rushed, somewhat careless. Lacking clear composition.	Rushed, careless, lack of color/tone and composition.