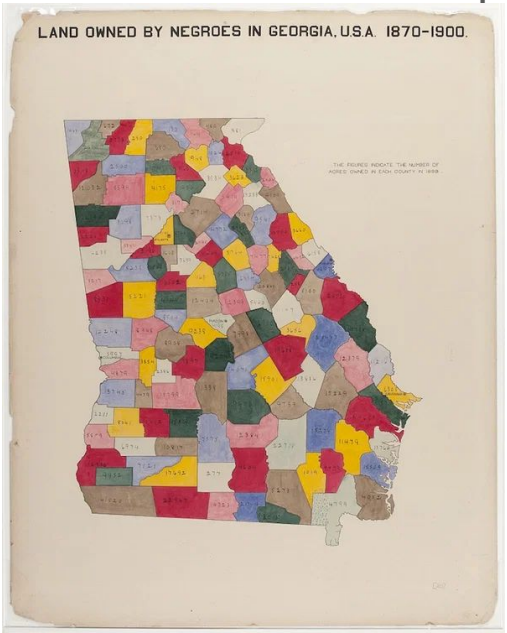
Data Can Be Beautiful

Presentation and Lesson for 6-8 Students

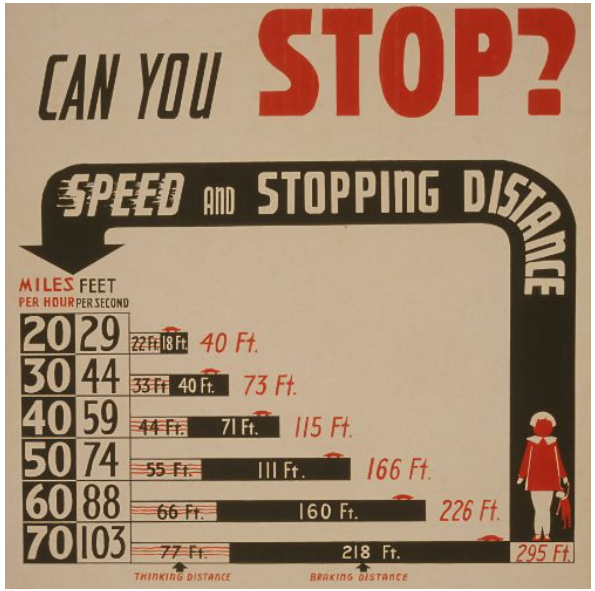
A long history of data as art

W.B. Du Bois' hand drawn graphs are powerful visualizations that informed the public about the African American experience. https://publicdomainreview.org/collection/w-e-b-du-bois-hand-drawn-infographics-of-african-american-life-1900/



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

What do you observe? What questions do you have? How does the imagery make the information more potent?



Standards

NGSS MS-LS4-3: Analyze displays of pictorial data to compare patterns of similarity…and identify relationships.

AZ State Standard 6.SP.B.4 Display and interpret numerical data

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.

What am I doing?

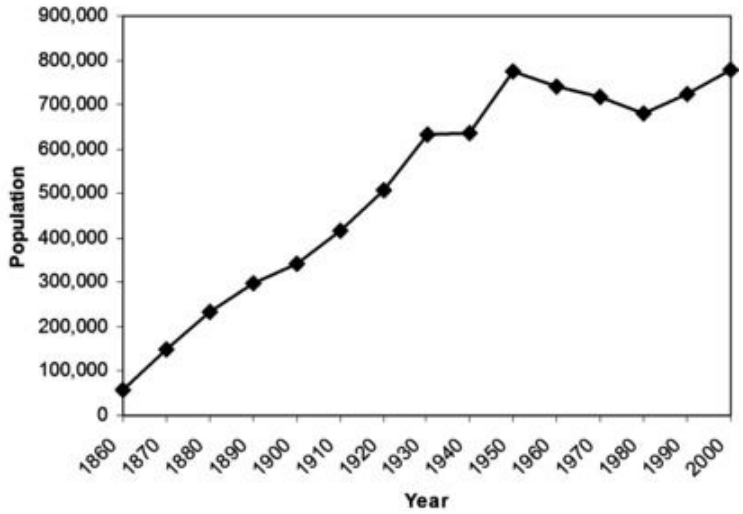
- Obtaining and tabulating data
- Creating a standard bar graph or pie chart
- Creating a more powerful, visually appealing graph/data art to convey the same information
- Analyze and interpret the information in the graph–what is the story of the data?

Why am I doing this?

- To organize information
- To make that information easy/easier to understand
- To make that information beautiful, interesting, powerful, add your personality and creativity

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

How will I know I've completed the assignment?

Rubric for data art

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.

Ready, set, go…

…to

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

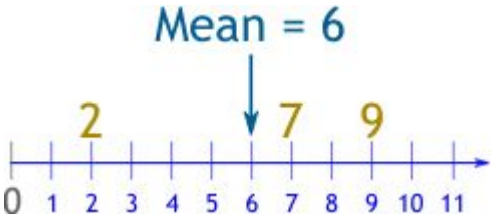
Get your data:

Trial 1, Trial 2, Trial 3

Aggregate class data.

Review of measures of central tendency

Mean: also called average.



How to calculate: add up all the values, divide by number of data points

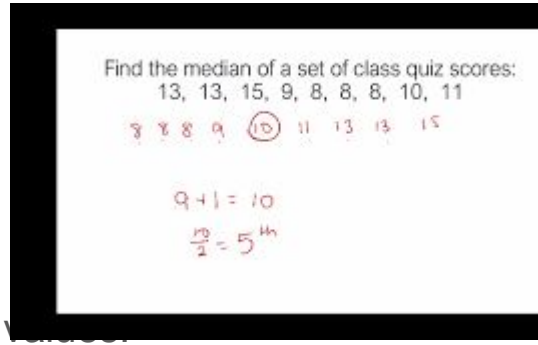
What does it tell you?

Average value of a distribution of figures, such as 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: put data in numeric order. It is the middle value, or average of 2 middle values

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/

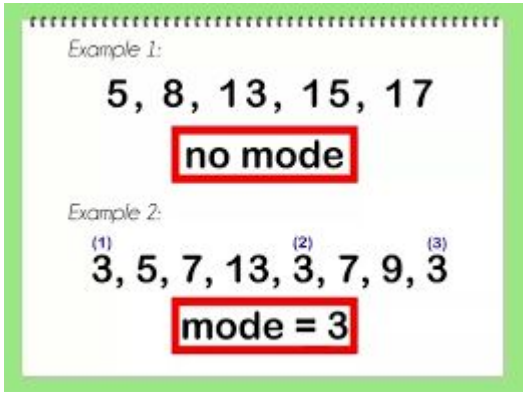
Mode

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.

How to calculate: set all values in numeric order, find most frequently occurring value(s)

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.

How to calculate: put data in numeric order, find the difference between greatest and least value

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

Which measure will you use?

Average of class’ guess by trial?

Each student’s average guess?

Median of class’ guess by trial?

Mode?

Range of student guesses?

It’s up to you, but you will have to tell the story of the data using a particular measure.

Create a bar graph or pie chart with the data.

You can try:

https://nces.ed.gov/nceskids/createagraph/

Plug in your data and create.

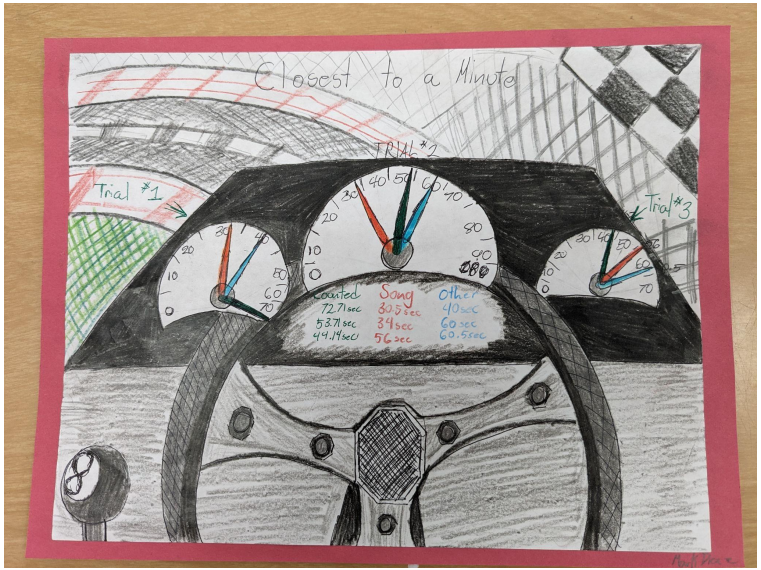
Now for the data as art graph!

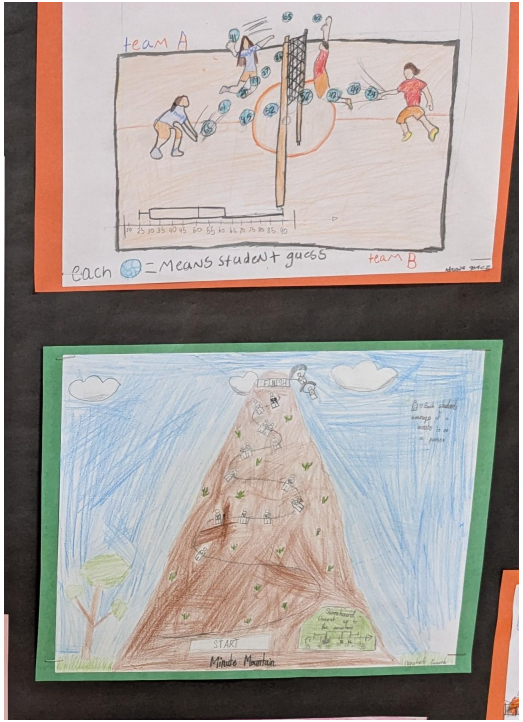
You will use the same data/measure and create a more artistic, visually appealing and powerful graph to convey the same information.

Your graph must be informative. The data must be visible and clear.

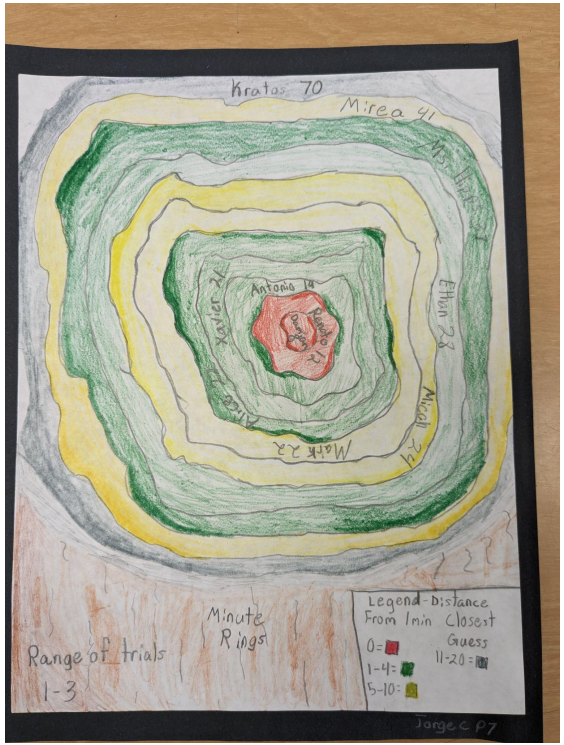
The graph should incorporate ‘time’ into the overall theme.

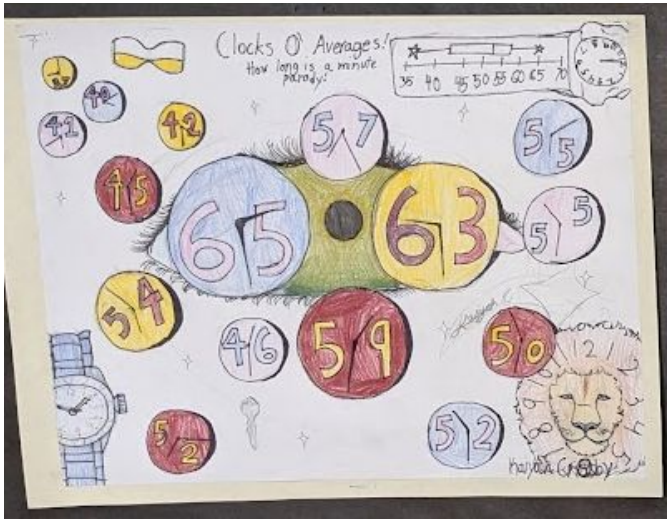
Some examples of student graphs

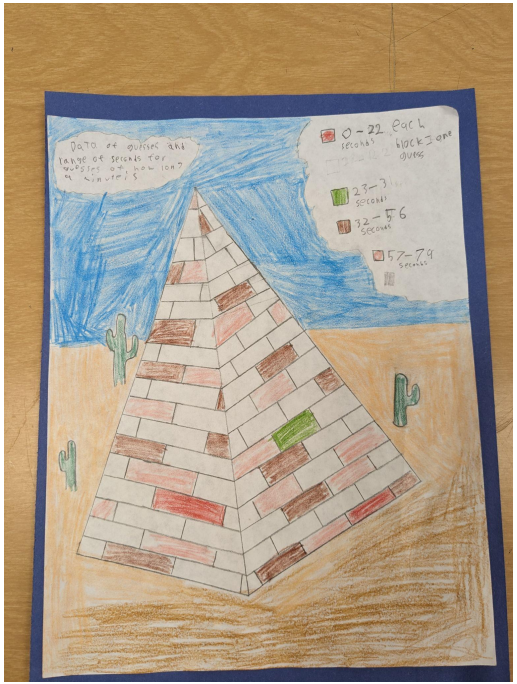




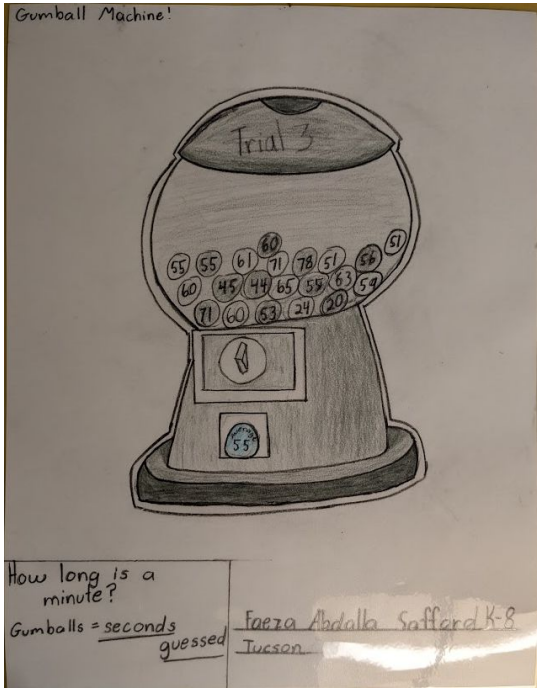
More examples







More samples



Difference between analysis and interpretation

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.

Analyze and interpret the data

- What patterns or trends do you see in your data? For example, is the class getting better at guessing how long is a minute with each trial? Why or why not?
- What would a most common guess (mode) tell you that is different from what the range would tell you?
- Was there a method to guessing? (singing a song, counting, etc.)
- Is there one particular data set/student trials that is interesting? What could that tell you?
- How could this data inform future decisions?

Write your summaries in your notebook.

Take a look at graphs via a gallery walk.

Write down your observations, questions about 3 different graphs. What works? What aspect of the graph leaves you with questions about the data? Does your graphed data tell the same or similar story as your classmates'?

For example, given the data to the right,

what kind of password would you create

for the highest security?