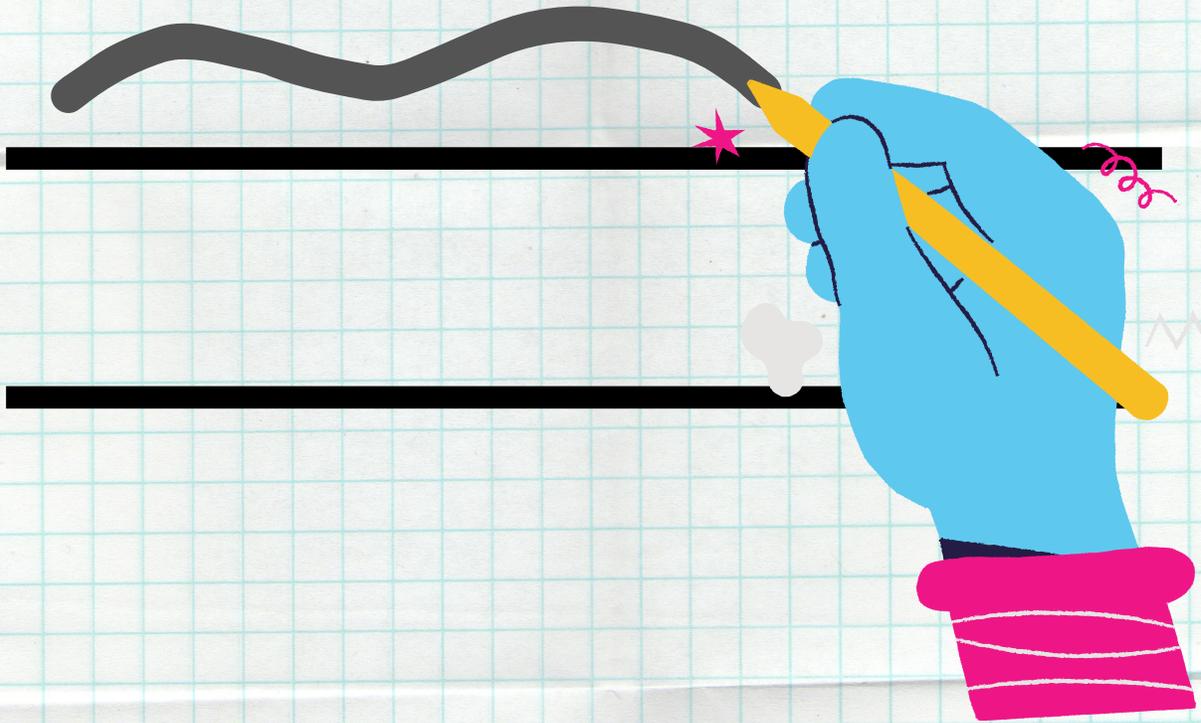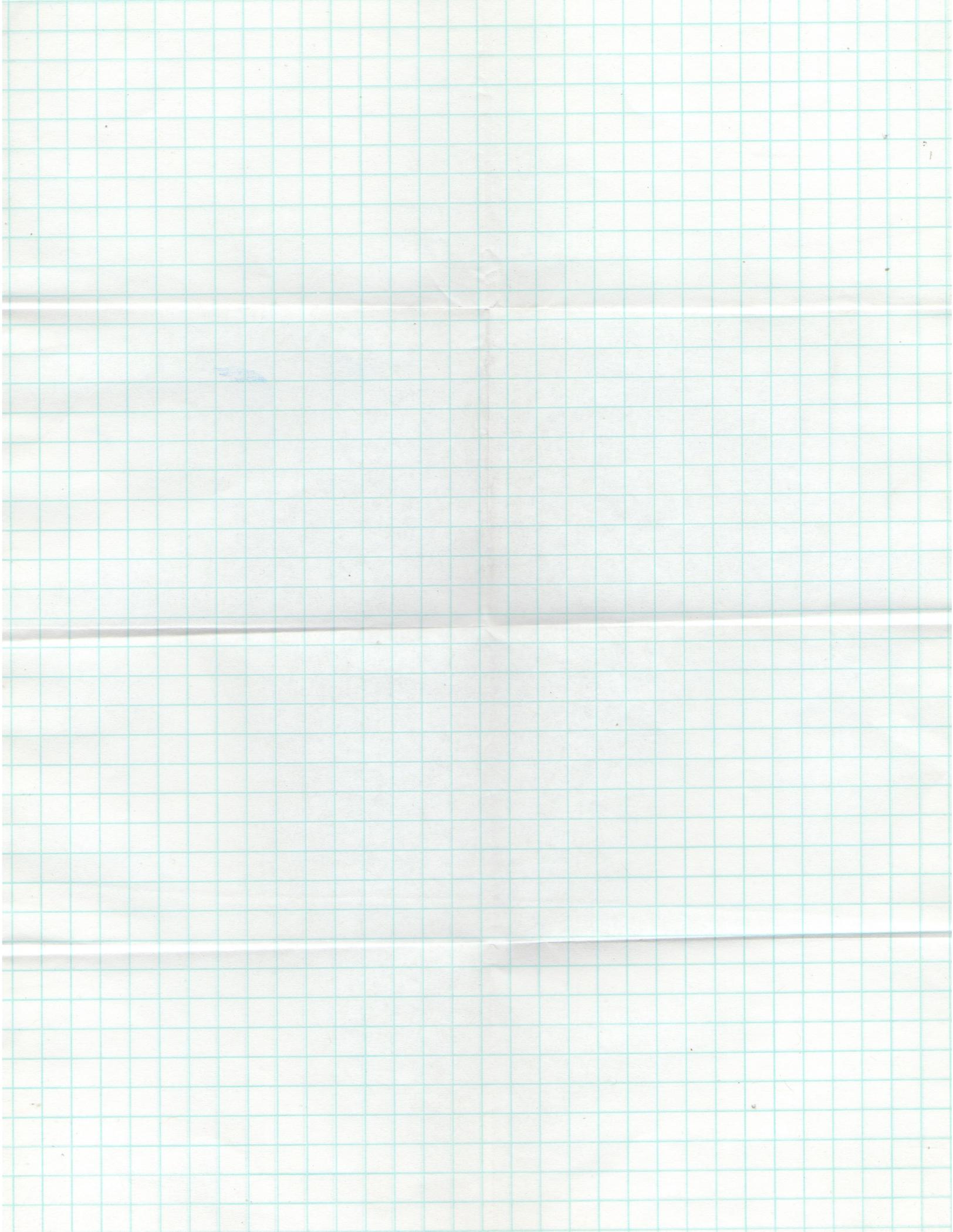# Data
Visualization

the great synthesizer

\#. _____

A student workbook to accompany data visualization activities

Foreword

Acknowledgements

We would like to thank the University of Rhode Island, the US Census Bureau, NERR (Natural Estuarine Research Reserve), Brown University, the Stroud Water Research Center for the use of their data, programs, visualizations, and other support they have provided in the making of this book.
A special thanks goes out to NSF (National Science Foundation) and RI C-AIM (Rhode Island Consortium for Coastal Ecology Assessment Innovation and Modelling) for brining all of us together.

Purpose

The purpose of this book, when used in conjunction with the teacher version, is to help students explore different aspects of data science in a beginner-friendly way. We hope you will learn about data visualization and that this knowledge can be useful throughout your professional and daily life.

Authors

This book was written by faculty and students at Rhode Island College. They are Dr. Sally Hamouda, Dr. Anabela Maia, Lauren Cenedella, Destiny Gonzalez, Meghan VanSchalkwyk, and Matthew Spaulding. You can contact us at: shamouda@ric.edu.

Table of Contents

Computer Science

Code.org's: Good vs. Bad Visualizations

Follow the teacher's directions on Code.org, then rate the visualizations and explain why you rated the visualizations the way you did.

The Ethics of Visualizations: The Challenger Space Shuttle Disaster

Link to source: https://www.adam-p-jones.com/why_is_visualization_important/ OR

link #2 in "Data Sets and Links"

Examine the image of O – Ring Damage in figure 1. What problems do you see with it?

_____

_____

_____

Why does this make the visualization confusing to try to understand?

_____

_____

_____

Why did it fail to delay the launch despite the engineers realizing there was a risk of O –

Ring failure?

_____

_____

_____

How would you organize a visualization to display the same information?
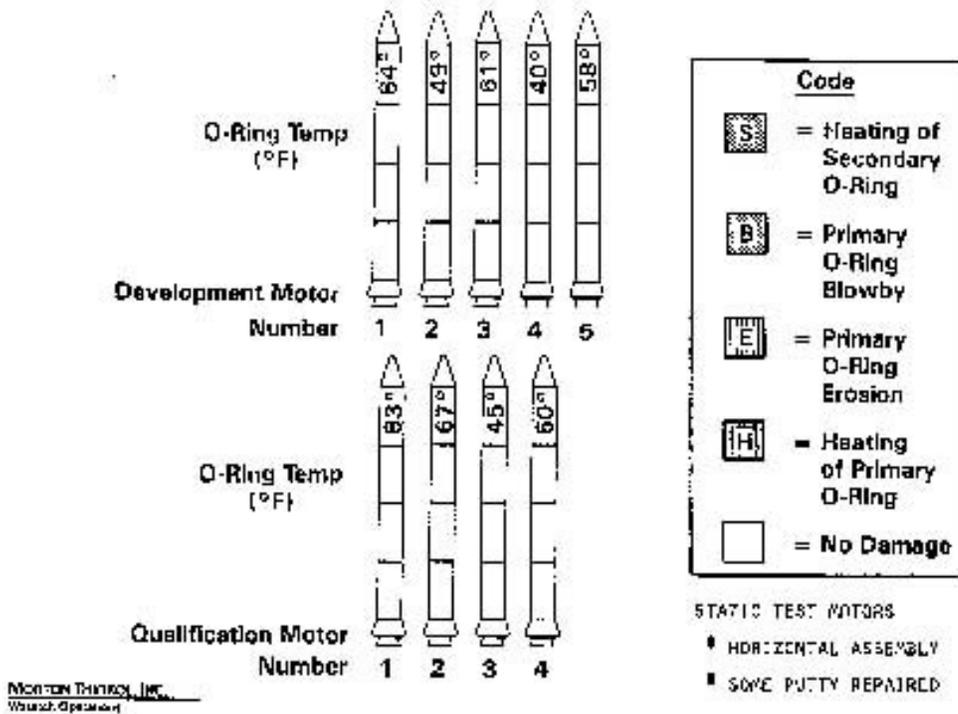
_____

_____

_____

How would you present it in a clear and compelling way? Remember public pressure

was on management to launch so a very clear argument must be made.

_____

_____

_____
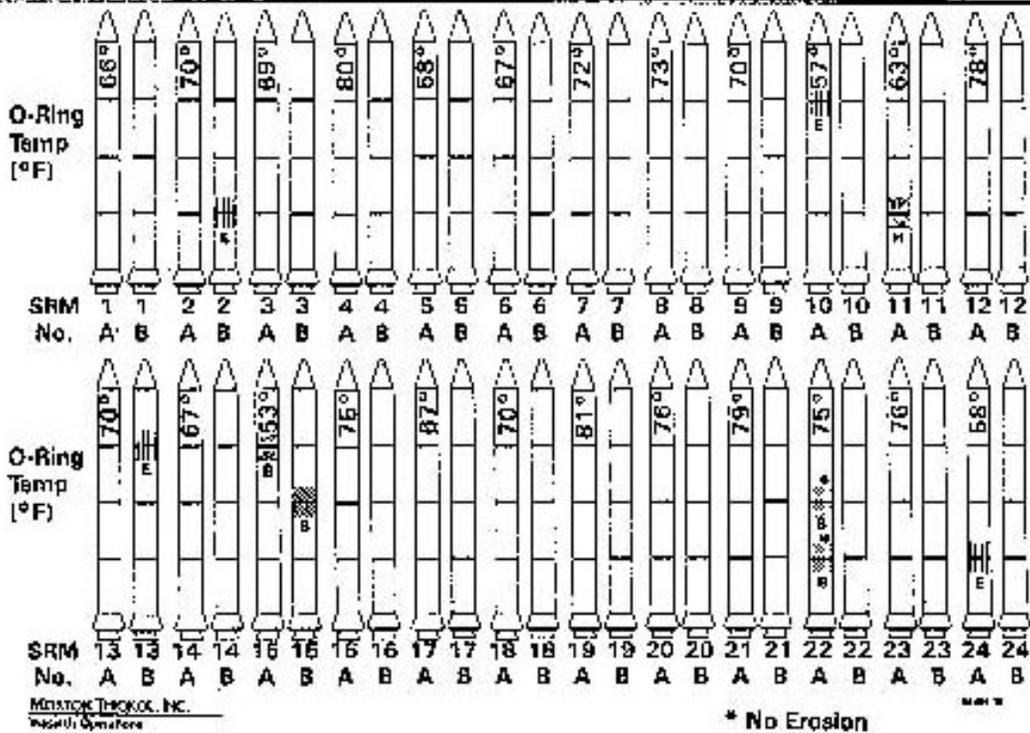
Figure 1. NASA's O - Ring damage visualization. Link to digital version in appendix.

Back to Basics With Bootstrap

Everyone needs to start somewhere, and this activity is meant to help students that need an introduction (or refresher) on creating visualizations from data. I used the Music, US Income, and RI Schools data sets from the provided link to create a bar chart, pie chart, line graph, and scatter plot.

Link to source: https://www.bootstrapworld.org/materials/spring2017/courses/data-science/resources/teachers/index.shtml OR link #3 in "Data Sets and Links"

1. Starting with the Music data set, I decided to clean the data by adding a column with data validation. Right-click on column J and insert a column to the right of it and call it "genre."

Figure 1. Excel's data validation tool. Used here to restrict a column's data.

2. Right-click on cell K2, scroll down and click on "Data validation."
3. In "Cell range," add ":K994" to the end (see *fig. 1*).
4. In "Criteria," click the dropdown menu and choose "List of items" and choose the genres you think are common (keep it simple and small!). I chose classical, country, cultural, dubstep, jazz, pop, r&b, rock, and other/unknown. Also be sure to check the "Show dropdown list in cell" option (see *fig. 1*).
5. Click save and we can begin to filter the data.
6. Select cell H, then click "Data" in the top toolbar, and then choose "Create a filter."
7. To make the list more manageable, click the dropdown arrow next to the column header called "familiarity." Choose "Filter by condition," "Greater than or equal to," and put 0.9. We will only have to categorize the songs that 90% of the public is familiar with.
8. After doing that, we can make a pivot table. Click "Data" again but choose "Pivot table" on a new sheet.
9. Add genre to "Rows" and uncheck "Show totals."
10. Add genre to "Values."
11. Add the same conditional filter from the sheet to "Filters."
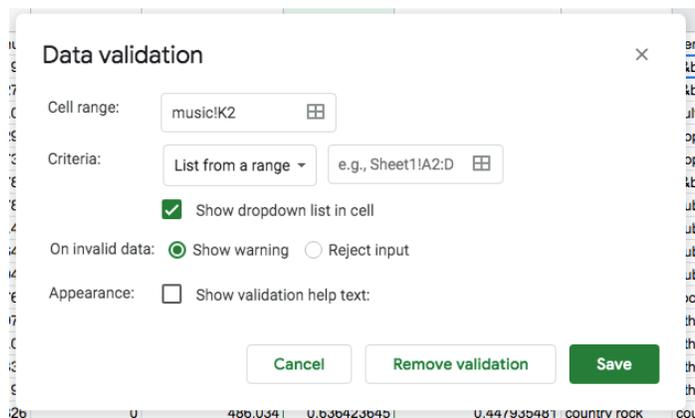
12. With cell A1 selected on the pivot table, click "Insert" in the top toolbar and choose "Chart."
13. Now that you have a chart, you can customize it as you please. I was able to make a bar and pie chart from this one pivot table (*fig. 2 and 3*).

Number of genres for songs that 90% of the public is familiar with
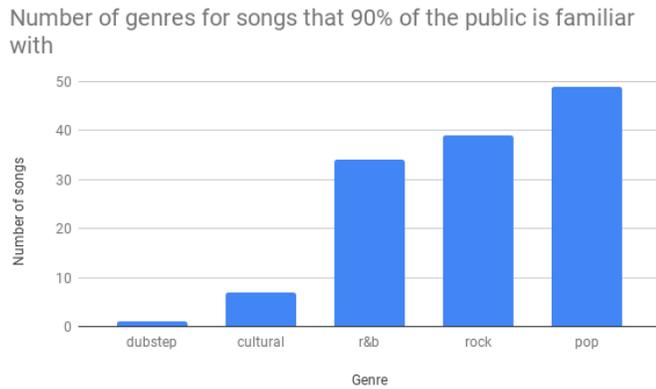
Figure 2. A bar chart of the number genres for the most popular songs.

Percent of genres for songs that 90% of the public is familiar with
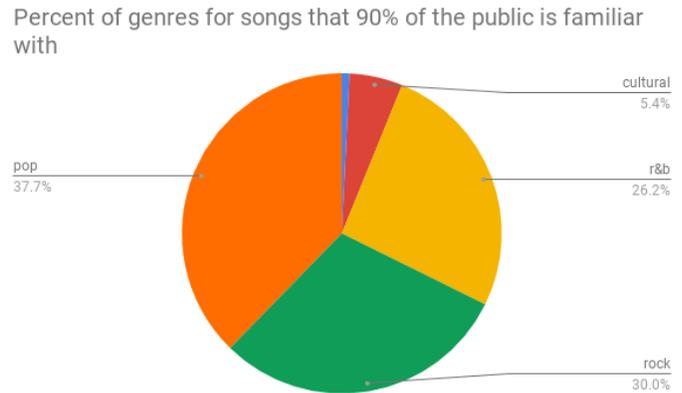
Figure 1. A pie chart of the most popular songs' genres.

1. Next, with the US Income data set, there is no cleaning needed.
2. Simply select columns A, L, and M, and then click "Insert," and "Chart."
3. It should come up with a line chart comparing the median and mean income of households over time. With this visualization, you can take the time to discuss why mean is so far above median (a few households with a ridiculous amount of money skews the data), and why you should always consider what is being graphed. My line graph, after customizing it, is seen below in figure 4.
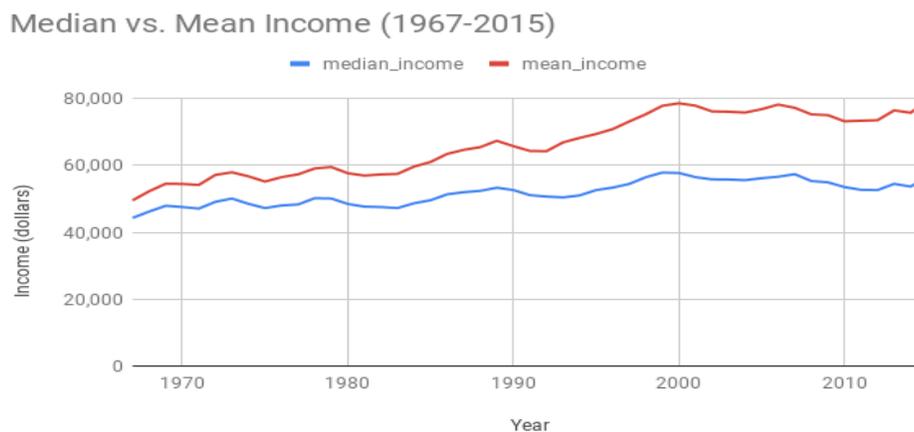
Figure 2. A line graph comparing the median and mean income of U.S. citizens over time.

Finally, with the RI Schools data set, you can play around with trends and correlations which are perfectly represented by scatter plots. For this example, I wanted to see if an increasing number of students correlated with the percentage passing English/language arts.

1. I added a filter to column P, and clicked "Sort A --> Z," which ordered the data from smallest student body to the largest.
2. I chose to make a chart of the 10 smallest schools, so I selected ranges P1:P11 (total students) and C1:C11 (percent passing) and made a chart.
3. In the chart editor, you want to change the X- and Y-axes, so click on each and choose the correct series (total students on X-axis and percent on Y). This is shown in figure 5.
4. On the series tab, click the 3 dots and choose "Add labels."
5. Click on the little box (see bottom right corner in *fig. 5*) to add the data range of A1:A11 (the school names).
6. Now in the "Customize" tab of the chart editor, scroll down to "Horizontal axis" and set the min and max bounds (0 and 200).
7. Go up to "Series" to show the trendline and R^2 value.

Figure 3. The chart editor showing the correct X- and Y-axis data. Also shows the "Add labels" box.

8. Customize the chart as you please. My visualization is shown below in figure 6. As you can see, Google Sheets has some trouble making the data labels visible.

Figure 4. A scatter plot of percent passing vs. number of students. Also shows the trendline and R-squared value.

What type of chart(s) did you choose for the music data and why?

_____

_____

_____

Do you think the type of graph you chose displays the data clearly? If it doesn't, why not? How could it be improved?

_____

_____

_____

For the US income data why is the mean so high?

_____

_____

What does that tell you about how outliers can affect data?

_____

_____

_____

Do you see any correlation between class size and passing English/Language arts? Why do you think that is?

_____

_____

_____

Do you think there would still be a correlation between class size and passing math?

_____

_____

_____

If you were to graph the data with math were you correct about there being a correlation or not?

_____

_____

_____

Fishy Data (Cleaning Data)

Download the URI water temperature data from Fox Island and view the data. (Note: the Excel sheet is 2,900 cells long, choosing a subset of the data that has some missing values that need to be cleaned is recommended for less experienced students)

Link to source: https://web.uri.edu/fishtrawl/data/ OR link #4 in "Data Sets and Links"

Clean the data and prepare it for a visualization. This will depend on the programming language or tool you are using. Figure 1 is an example of some of the data with missing values.

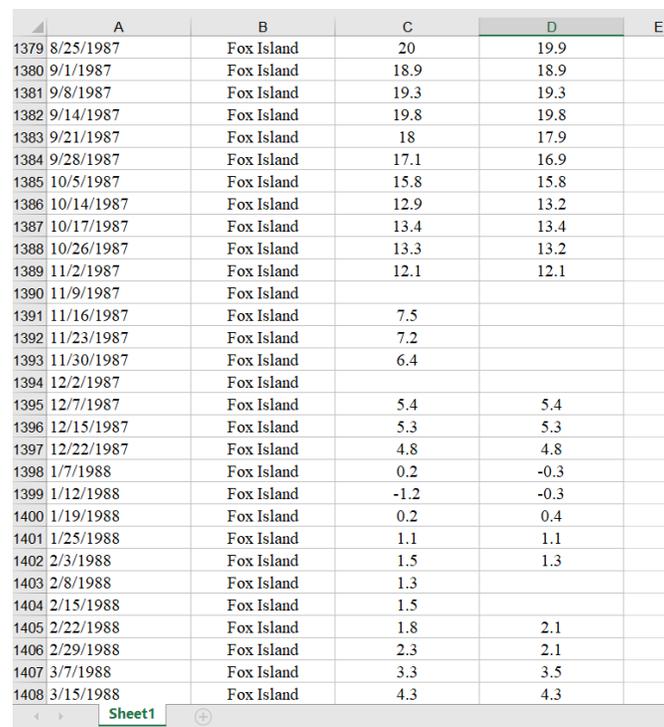|	A	B	C	D	E
1379	8/25/1987	Fox Island	20	19.9	
1380	9/1/1987	Fox Island	18.9	18.9	
1381	9/8/1987	Fox Island	19.3	19.3	
1382	9/14/1987	Fox Island	19.8	19.8	
1383	9/21/1987	Fox Island	18	17.9	
1384	9/28/1987	Fox Island	17.1	16.9	
1385	10/5/1987	Fox Island	15.8	15.8	
1386	10/14/1987	Fox Island	12.9	13.2	
1387	10/17/1987	Fox Island	13.4	13.4	
1388	10/26/1987	Fox Island	13.3	13.2	
1389	11/2/1987	Fox Island	12.1	12.1	
1390	11/9/1987	Fox Island			
1391	11/16/1987	Fox Island	7.5		
1392	11/23/1987	Fox Island	7.2		
1393	11/30/1987	Fox Island	6.4		
1394	12/2/1987	Fox Island			
1395	12/7/1987	Fox Island	5.4	5.4	
1396	12/15/1987	Fox Island	5.3	5.3	
1397	12/22/1987	Fox Island	4.8	4.8	
1398	1/7/1988	Fox Island	0.2	-0.3	
1399	1/12/1988	Fox Island	-1.2	-0.3	
1400	1/19/1988	Fox Island	0.2	0.4	
1401	1/25/1988	Fox Island	1.1	1.1	
1402	2/3/1988	Fox Island	1.5	1.3	
1403	2/8/1988	Fox Island	1.3		
1404	2/15/1988	Fox Island	1.5		
1405	2/22/1988	Fox Island	1.8	2.1	
1406	2/29/1988	Fox Island	2.3	2.1	
1407	3/7/1988	Fox Island	3.3	3.5	
1408	3/15/1988	Fox Island	4.3	4.3	

Sheet1

Figure 1. A preview of the raw data from the Fox Island water temperature data set.

How do you think missing values would affect a visualization?

_____

_____

_____

What can be done about the missing data points?

_____

_____

_____

What other issues do you think you might find in a data set? What effect do you think

they would have and how would you address the issues?

_____

_____

_____

How Are You Feeling?

This activity has you creating surveys, or using an existing survey to generate data, clean it, and visualize it.

Looking at the pre-made introduction survey data, you will have three challenges to clean the data. You will have to separate the timestamp column, make a data validation column for the types of entertainment, and use the "countif" function and wildcard characters to count how many instances of a particular de-stressor there are. The first challenge is explained in the science activity "Navigating NERR Data," and the second is explained in the computer science activity "Back to Basics With Bootstrap."

Link to resource: https://www.howtogeek.com/446398/how-to-count-data-matching-set-criteria-in-google-sheets/

Text to Columns: From "Navigating NERR Data"

1. The next problem you have to fix is changing the time stamp column into two separate columns for date and time. In Excel, click on column D, and choose "Insert."
2. Do this again so you have two blank columns next to the time stamp column (C).
3. Click on C so the whole column is selected, then click on "Data" in the top toolbar, and then click "Text to Columns."
4. Choose "Fixed Width" on the first page.
5. Make sure there is only one break (between the date and time) on the second page (*fig. 3*).
6. Change the column data format to date and make the destination the next column over (in this case D) on the last page (*fig. 4*).
7. Click "Finish" and you're back in the spreadsheet. Right-click on the column with the dates, choose "Format Cells...," and make sure they are set to "Date."
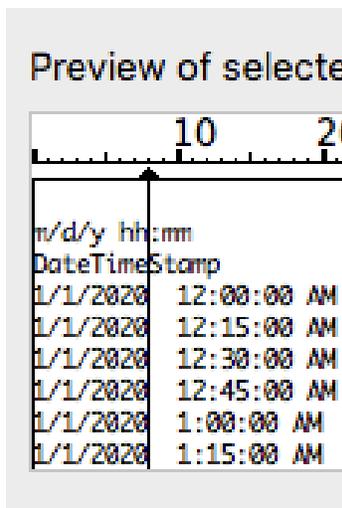
Figure 3. Page 2 of the "Text to Columns" tool. With the fixed width option, you choose where the text is broken up.
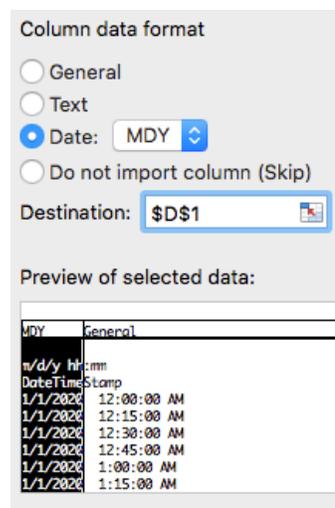
Figure 4. Page 3 of the "Text to Columns" tool. The highlighted column contains dates, so it is formatted as a date. The destination is the next column over.

Data Validation: From "Back to Basics With Bootstrap"

1. Starting with the Music data set, I decided to clean the data by adding a column with data validation. Right-click on column J and insert a column to the right of it and call it "genre."
2. Right-click on cell K2, scroll down and click on "Data validation."
3. In "Cell range," add ":K994" to the end (see *fig. 1*).

Figure 1. Excel's data validation tool. Used here to restrict a column's data.

4. In "Criteria," click the dropdown menu and choose "List of items" and choose the genres you think are common (keep it simple and small!). I chose classical, country, cultural, dubstep, jazz, pop, r&b, rock, and other/unknown. Also be sure to check the "Show dropdown list in cell" option (see *fig. 1*).
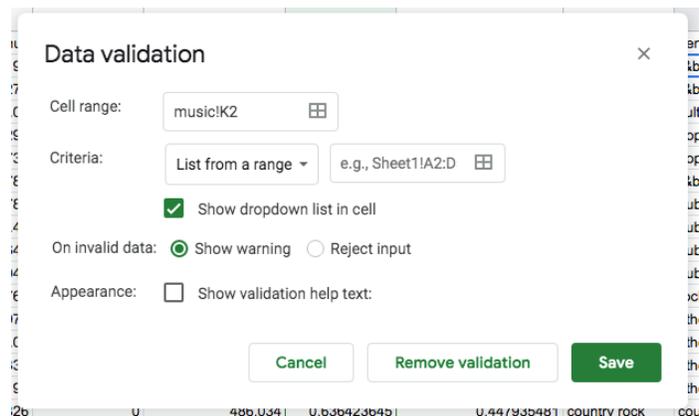
Countif Function

The last question allows users to pick more than one answer, meaning there will be several pieces of data in one cell instead of spread out individually.

1. We need to make a separate mini table of each option and its respective count so we'll need to use the "countif" function.

2. The first parameter of the function needs a range, so I selected the column with unsorted data.
3. The second parameter needs a logic statement in between quotes. I wanted to put in the option by itself, but that doesn't work. You have to add an asterisk (a wildcard character) so that the function looks for any instance with a space, no space, or comma before or after it.
4. You do not need to put the whole option between quotes, just something that is unique to it. For example, I have the options "Going outside" and "Going for a drive around town" as things that make you relaxed. I can keep "*Going outside*" in the function because it is short, but I can just put "*Going for*" as the parameter for the other option since it is a unique phrase found in no other option.

Figure 2, below, shows the data, the mini table I created, and the function that counts for instances of "Music."

```
=COUNTIF(I2:I101, "*Music*")
```

H	I	J		K	L	
Which pet do you prefer:	What are things that help			Thing	Count	
Dog	Music, Going outside, Jus			Music	2	
Cat	Going outside, Shopping,			Going outside	2	
birds	Music, Going for a drive a			Just chilling around the ho	1	
	other			Going for a drive around t	1	
				Shopping	1	
				Hanging out	0	
				Scented candles	2	
				Playing with your pets	1	
				Other	1	

Figure 2. A sample data set generated by using the introduction survey. With cell L2 highlighted, the "countif" function and its parameters are shown.

What did you learn about the process of collecting data through the surveys?

_____

_____

_____

What would you change or improve about how you collected data from surveys?

_____

_____

_____

13

What was the most common de-stressing tactic your classmates used?

_____

_____

_____

Are there any days of the week or times when your classmates are the most stressed?

The least stressed?

_____

_____

_____

What did you learn about data, in terms of how to collect it, clean it, and analyze it?

_____

_____

_____

Creating a Graph With Jupyter Notebook

Using Jupyter Notebook and any code your teacher provided, familiarize yourself with each variable and what each line of code does. Once you feel as though you understand this, try and customize your graph in some way. Then, find a dataset—a CSV file is best—and use that data instead of the hard-coded values.

Link to Jupyter Notebook: https://jupyter.org/try

Are there any problems you encountered?

_____

_____

_____

If you got your graph working with your data and everything works fine, can you foresee
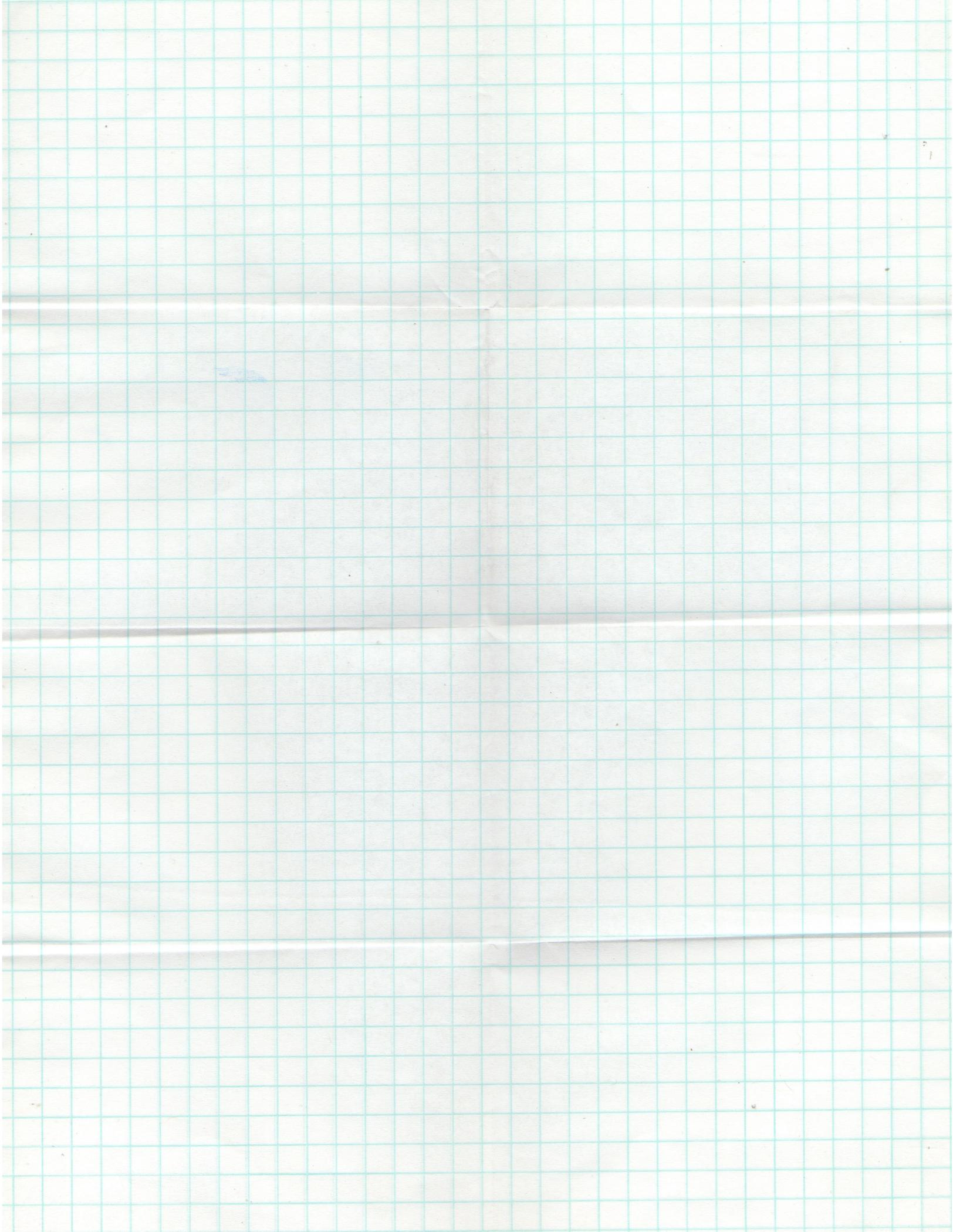
any problems with other datasets?

_____

_____

_____

Are there any benefits you see in representing data visually, compared to the raw CSV

file?

_____

_____

_____

English/Literature

What's Trending Now?

Go to trends.google.com and type in two different topics, cat and dog adoption in any US state that you choose and any time frame, just keep consistent. Now there are different ways that you can phrase this.

"cat adoption near me" and "dog adoption near me"; "buy dog near me" and "buy cat near me"; "animal shelter with cats" and "animal shelter with dog."

See how the search levels fluctuate for different phrasings.

See how the same topic is being looked up, but the phrasing makes a difference. This makes an important difference when you are doing research. Keep this in mind when you are doing research for certain topics.

What kind of map/graph is this most like?

_____

_____

_____

Do you think a different kind of graph would be easier to understand? Why and what kind of graph?

_____

_____

_____

What does this tell you about the potential to alter a visualization based on the search terms that are being compared?

_____

_____

_____

Graphing Grammar

1. Download the Grammarly Extension for Google Chrome
2. Type 2 paragraphs about any given topic. Use sentences of varying lengths and colorful language. Use a minimum of 100 words.
3. Allow Grammarly to make revisions on your verbiage, spelling etc.

How is this a data visualization tool?

_____

_____

_____

Do you see the potential of this being useful to you as a student in the future?

_____

_____

_____

What do you notice about your writing style that you did not notice before?

_____

_____

Do you notice anything teachers have commented on before?

_____

_____

What, if anything, do you think you could learn from a person just by looking at their grammar?

_____

_____

_____

Social Networking – Who Knows Who

This activity introduces the concept of a node diagram. You can find a well-written description here at datavizcatalogue.com/methods/network_diagram.html. By using flourish.studio/, you will be able to create node diagrams for the characters and their relationships of any book the class is currently reading.

For example, by assigning a link value (based on the strength of their relationship) figure 1 shows the characters in *Romeo and Juliet*.
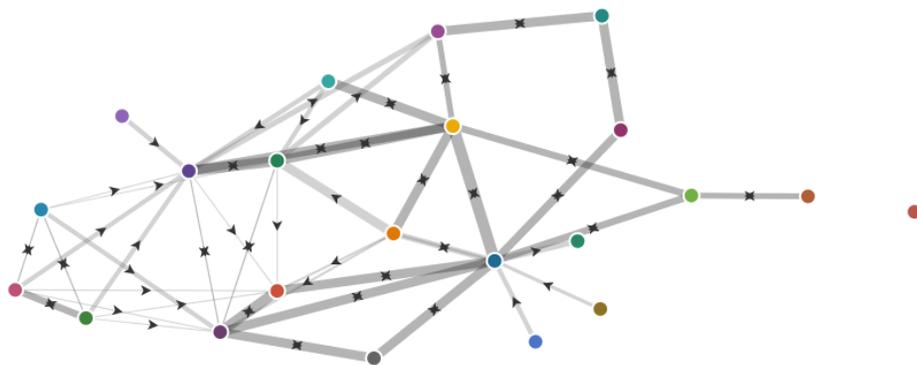
Figure 1. A node diagram of the relationships between characters in Romeo and Juliet.

1. To begin, make an account with the Flourish website. You can sign in with your Google account, or you make a new account.
2. On the My Projects page, click "New visualization," and then scroll all the way down until you see "Network graph" section. Choose the "Directional graph" option.
3. You will see a visualization that looks like the one in figure 2. Click the "Data" button and you will see 2 tables for the links and points.

Figure 2. The default node diagram when choosing the network graph template.

4. Delete all of the data by clicking the dropdown arrows next to the labels "Links" and "Points," and choose clear sheet.
5. You can then select the table headers (who, suing, who, etc.) and delete those, as well.
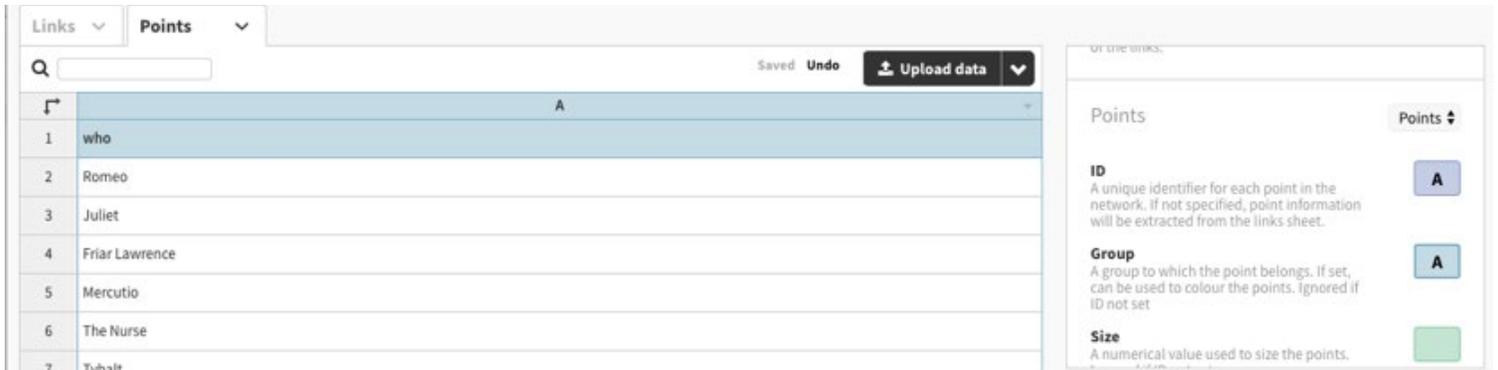
Figure 3. The correct "Points" table and side panel. Each row in the table is a character, and "ID" and "Group" in the panel are both set to "A."

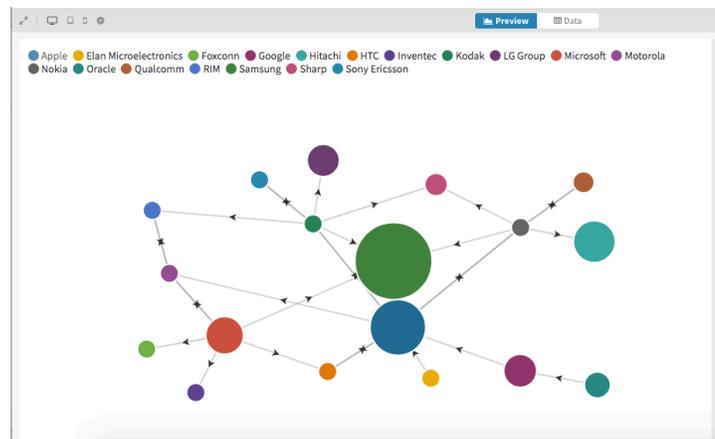6. Now we will set up our own tables. We want each point in the visualization to be 1 person, so we only need 1 column in the "Points" table.
7. To delete any extra columns, right click on their letter (A, B, C, etc.) and choose "Remove column."
8. We can label the A column "who" and start filling in the characters (*fig. 3*).
9. On the right side of the screen there will be a list of options. Scroll down to the "Points" section and make sure "ID" and "Group" are both A, so that each point will have the name you put in and its own color (*fig. 3*).
10. For the links, choose a scale to define the strength of the relationship. For my visualization, I had 1-10 representing true hate, hate, merchant, servant, acquainted, friend, good friends, related/familial love, married, and true love.
11. The "Links" table will require 3 columns. To add a column, right click on the B column and choose "Insert column right."
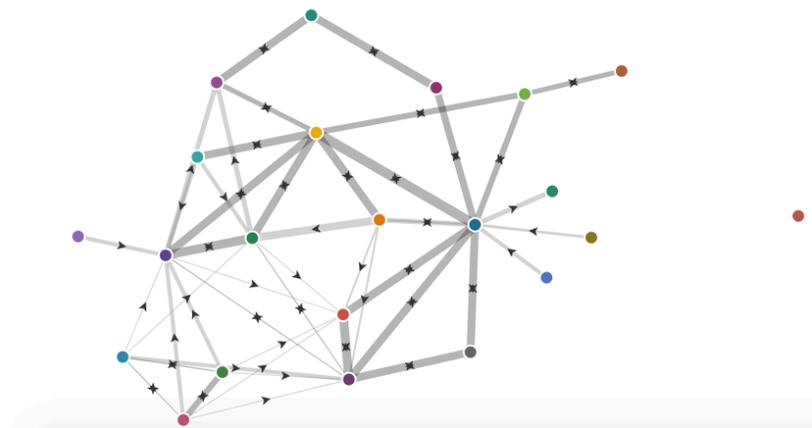
Figure 4. The same diagram as figure 1, but without a title.

12. Label A as "who," label B "to who," and C "relationship."
13. On the options panel on the right, source should be A, target should be B, and value should be C.
14. After filling in all the relationships (I found SparkNotes' list of characters to be very helpful in this part), your chart should look like the one in figure 4.
15. You can add a title by going to the "Preview" tab, scrolling down to the "Header" section, and adding one.

What is easier to understand in the graph compared to reading the book?

_____
_____
_____

What is immediately clear about the relationships the main character has with the other

characters?

_____
_____
_____

Did you notice anything unexpected or that you did not realize before seeing the graph?

What was it?

_____
_____
_____

Complete the Infographic

After you receive the incomplete infographic from your teacher, please answer these questions.

What can you already learn or guess from the incomplete visualization you have been provided?

_____

_____

_____

As you start to fill it in what do you notice about the book that you would not have seen without the visualization?

_____

_____

_____

Did you find the visualization more helpful than the same information would have been in written form?

_____

_____

_____

Visualize a Theme

As a class pick a theme or recurring event, such as every time a character is happy, or every time something is successful or fails. Track every instance in a spreadsheet, and at the end of the book create a visualization based on the data. This can be done in small groups or as a class. How you organize your spreadsheet to best track the data and what type of visualization you make is up to you. You can read more about this idea and how others have implemented it by using the provided link.

Link to resource: https://www.edutopia.org/blog/homer-high-tech-data-visualization-james-earle

As an example, I visualized themes and words from the first chapter in *The Great Gatsby*. I made a simple spreadsheet (*fig. 1*) since the data set was small, but complex tracking will result in a more complex sheet. You could include what chapter, page number, or quote the instance occurs in a collaborative spreadsheet.

Figure 1. A simple table for the number of occurrences of a word or theme in chapter 1 of The Great Gatsby.

Using Infogram, I chose a simple template and added a word chart. Depending on what you want to show, you will end up using different graphs and should be able to explain your choices. You should also add a caption to your project as to what the visualization is showing if your title doesn't fully explain it. My undecorated work ended up as shown in figure 2.
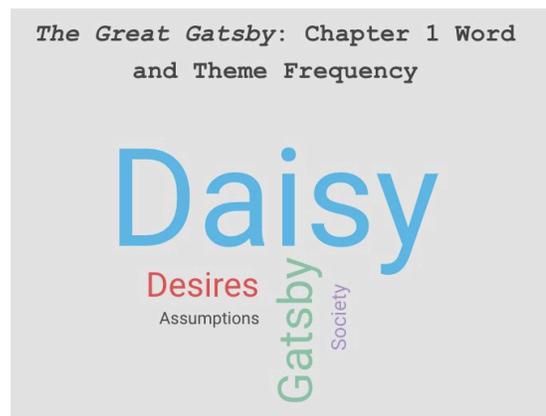
Figure 2. A word chart based off of the data in figure 1's table.

Why did your group or the class pick the theme that they did?

_____

_____

_____

How will you organize your spreadsheet to be clear and easy to read but still contain all the information necessary?

_____

_____

_____

How will you organize your visualization?

_____

_____

_____

Which is easier to understand, the spreadsheet or the visualization?

_____

_____

_____

 Is all of the data still displayed in the visualization?

_____

_____

_____

History of a Text

Providing context for a novel always enhances the experience of reading it. This activity provides an example infographic made with Infogram on what you could make when asked to research the context of a novel (see *fig. 1*). I researched *The Catcher in the Rye* and made a timeline of its relevance over time. You can view my infographic here: https://infogram.com/history-of-a-text-1h7z2l3mgejy2ow
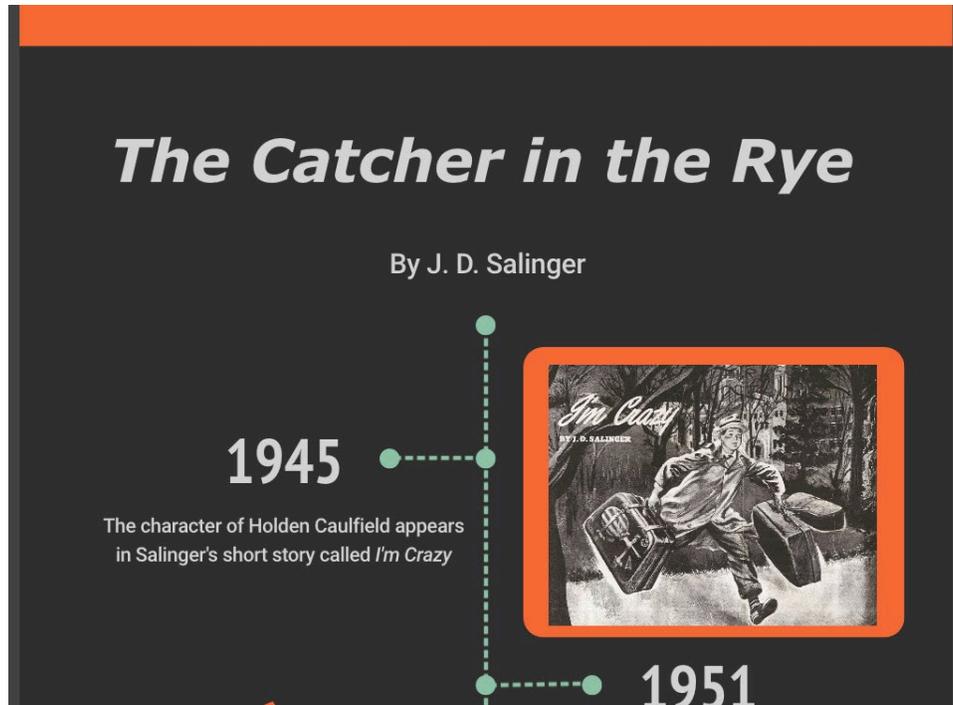
Figure 1. A snapshot of the infographic about The Catcher in the Rye.

You can include information about the author's inspiration or life at the time their book was written, what was happening when the book was released, the impact the novel had (e.g. Sinclair's *The Jungle* lead to food safety laws), or any other interesting facts. Instead of using Infogram, you could also make a PowerPoint/Google Slides presentation, or a physical poster/tri-fold board.

What do you think was the inspiration for the book?

_____

_____

_____

What were some relevant current events that were happening as the book was being written or around the same time it came out?

_____

_____

_____

Did the book have an impact when it was first published?

_____

_____

_____

Was the book controversial when it was published? If it was, explain why.

_____

_____

_____

What did you find interesting when you began research about your text?

_____

_____

_____
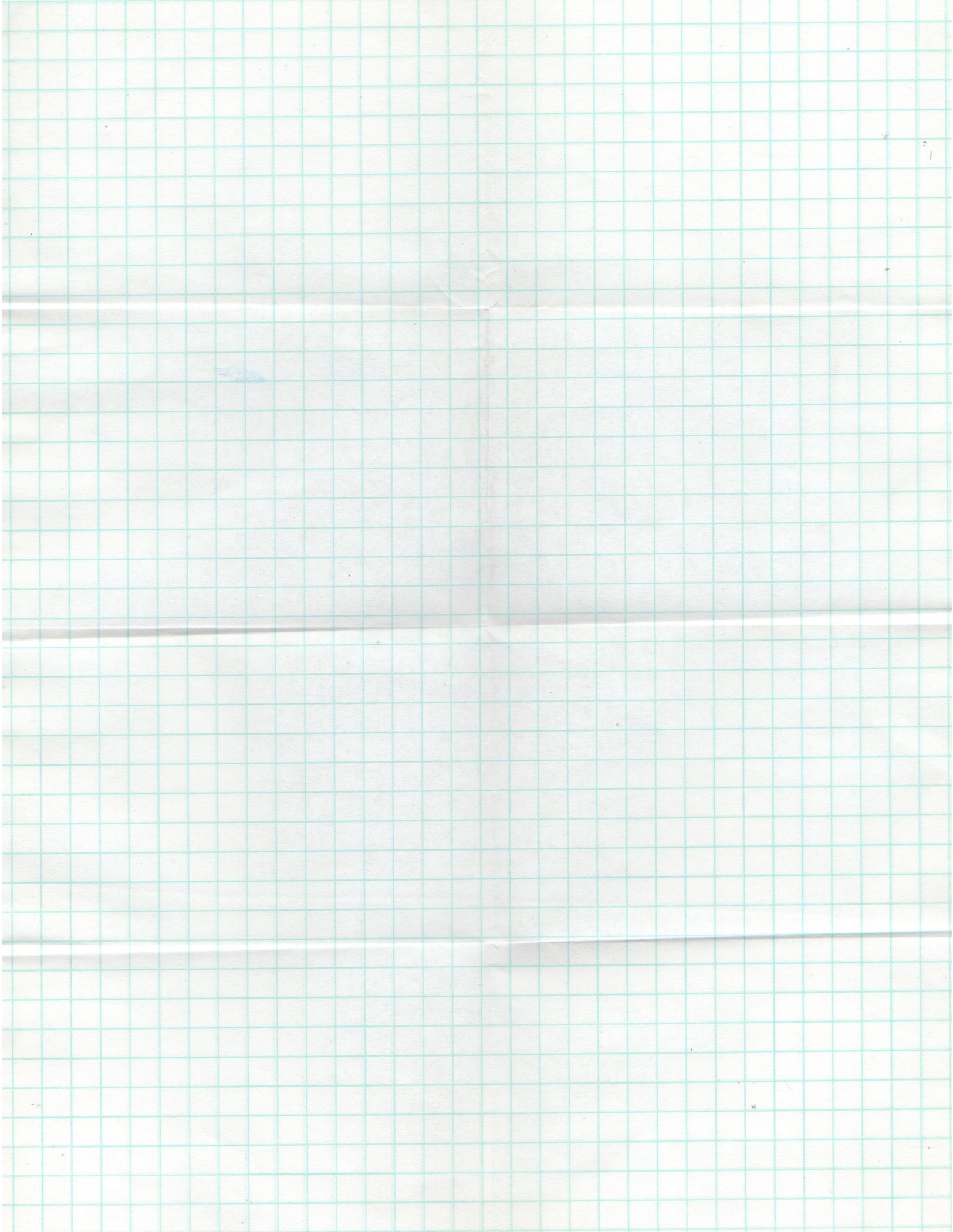
Mathematics

Analyze an Unfamiliar Visualization
Average Test Grade

Given the following graph can you determine the mean, median, mode and range of the data? Hint: you will need to find the average.
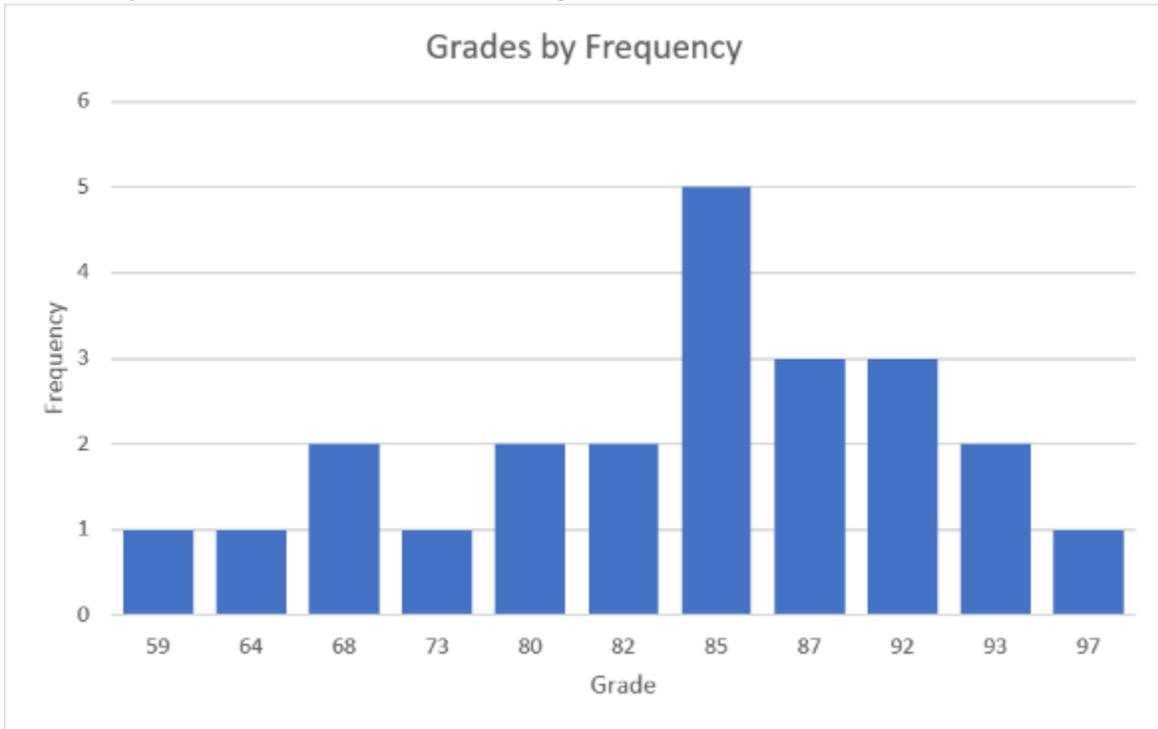
Figure 1. A bar graph of the frequency of grades.

Mean: _____   Mode: _____
Median: _____   Range: _____

What do you see that is familiar in this visualization?

_____
_____
_____

 What is unfamiliar?

_____
_____
_____

What, if anything, confuses you about it?

_____
_____
_____

Why do you think the creator chose to display the information in this manner?

_____
_____
_____

Do you think this is an effective way to display all of the information? Why or why not?

_____
_____
_____

Average Annual Snowfall 2019
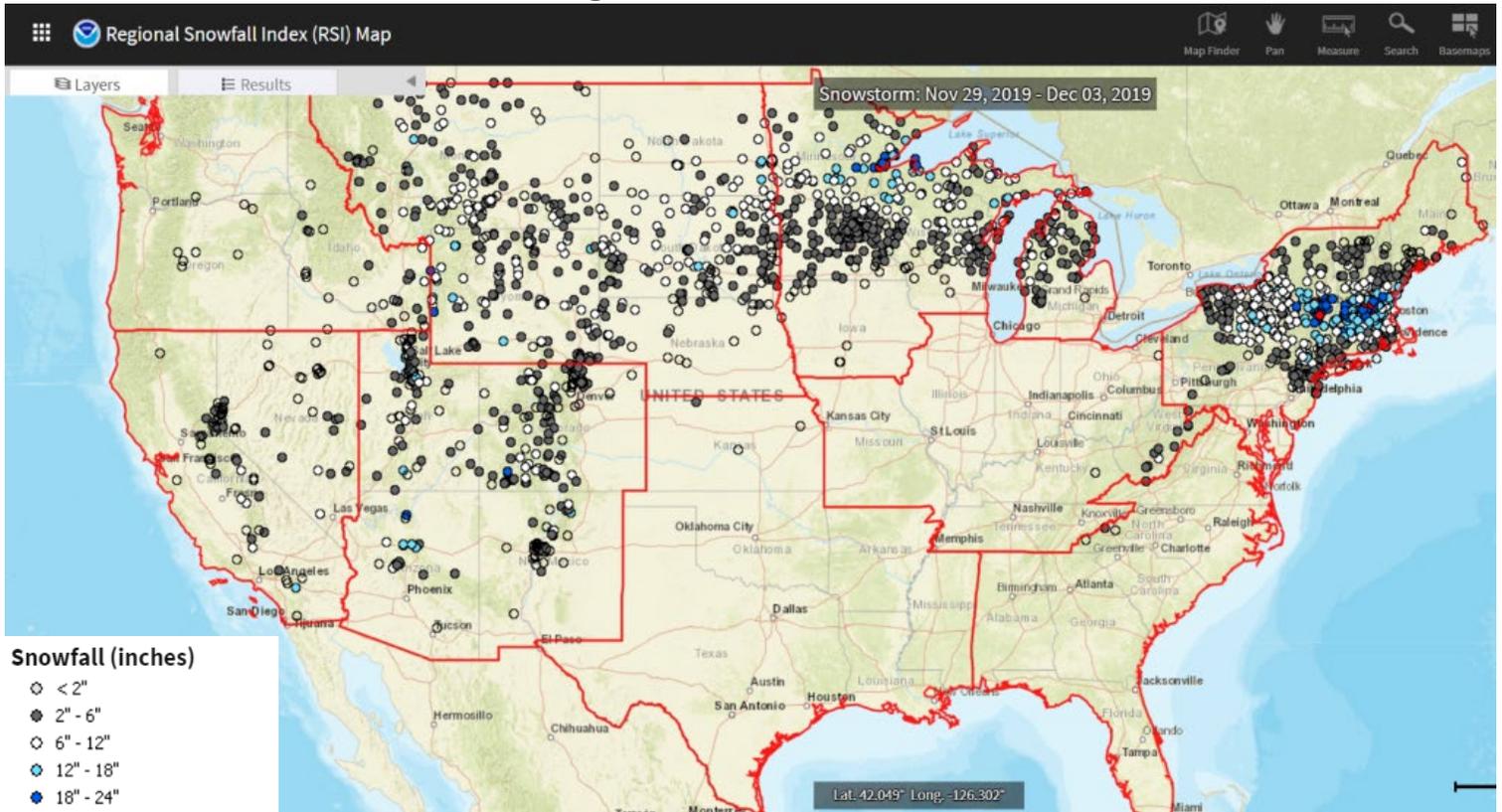
Figure 2. A map depicting snowfall in 2019 in different places with colored dots representing the amount of snow.

Link to digital map: https://gis.ncdc.noaa.gov/maps/ncei/rsi

How would you go about finding the mode for figure 2?

_____

_____

_____

Find the mode for the Regions with Phoenix and Oklahoma City in inches.

_____

_____

_____

Do any of the regions contain all the ranges of inches in the legend?

_____

_____

_____

Name a city in the region with the highest number of snowstorms.

_____

_____

_____

Let's Sample Some...

Pizza!

Below I have listed the prices of large cheese pizzas around Rhode Island as of 6/19/2020 (before tax and delivery) in table 1. Create a graph of these thirty+ places using Excel. Here's a refresher on how to graph distributions – https://www.extendoffice.com/documents/excel/2404-excel-template-bell-curve-chart.html - bell_curve_chart

On TI calculator.
1. STAT --> EDIT --> put numbers in a list
2. STAT --> CALC --> 1-Var Stats
3. Make sure "List" is the list where you entered the data, and choose "Calculate."

Xbar (x with a bar on top) = the avg of the numbers

$$\text{Standard Error of } \bar{X} = \frac{s}{\sqrt{n}}$$

We will be using S_x : because that is the sample standard deviation. $S_x = s$; where n = number of elements in list/set.

Table 1. A list of pizza places and the cost of pizza. (n=32)

Place	Cost
Big Tony's	11.35
Tommy's Pizza	15.10
Providence Coal Fire Pizza	21.00
Sicilia's	11.99
Caserta Pizza	17.50
Napolitano's	16.00
Pizza J	15.00
Twin's Pizza	16.85
Tom's Kitchen	12.45
Federal Hill Brick Oven	13.25
Antonio's	12.25
Warren House of Pizza	12.90
Piezonis	12.49
Papa Gino's	13.99
Balasco's Pizzeria	15.45
Pomodoro Pizzeria	13.00

Place	Cost
Cantazaro's Pizzeria	18.99
Nino's Pizza	14.50
Uncle Tony's	15.20
Sako's Pizza	11.99
Tony's Pizza Palace	12.50
Maria's Pizza and Wings	12.50
Boulevard Pizza	11.99
Domino's	10.99
Vivaldi's Pizza	10.95
Tomaselli's Pizza	13.49
Papa John's	11.99
DePetrillo's Pizza and Bakery	13.50
East Providence House of Pizza	10.59
Pizza Roma	10.99
Rosa Mia Pizzeria	16.00
Nice Slice	16.00

Does the price of pizzas from the listed local pizza spots fit a normal curve?

_____

_____

_____

What was the standard error of the mean?

_____

_____

_____

Do you think you would get the same result if you looked at pizza prices across the

United States?

_____

_____

_____

…Heights!

Say we have a group of women with the heights listed below, create a graph. Is it still normally distributed? Heights (in inches): 65, 48, 60, 52, 62, 60, 61, 63, 64, 60, 62, 63, 60, 62, 66, 67, 62, 63, 66, 63, 65, 63.

Optionally, measure the heights of everyone in the class.

Do the heights listed have a normal distribution?

_____

_____

What about the heights of the students?

_____

_____

Does gender make a difference?

_____

_____

_____

Does the number of data points?

_____

_____

_____

So Many People!

Using the given link choose 7 states and create a bar graph with the population in 2010 vs 2018 in whatever program that you feel comfortable.

Link to resource: https://www.census.gov/schools/facts OR link #6 in "Data Sets and Links"

Do you see the population increasing or decreasing in each state?

_____

_____

_____

If your seven states are close to each other geographically are they all decreasing or all increasing?

_____

_____

_____

If some are increasing and some are decreasing what factors do you think would cause states near each other to have different population change?

_____

_____

_____

If you also graph 2014 is the change consistent in each state?

_____

_____

_____

Geometry Jam – Understanding Scale

After you complete the activity about scale models your teacher presented, answer whichever questions are applicable.

Before this activity have you ever used a compass, geometry triangle or other tools before?

_____

_____

_____

How familiar are you with the metric system of measurements?

_____

_____

_____

What did you learn about the importance of accurate estimates?

_____

_____

_____

What did you learn about scale and proportions?

_____

_____

_____

Algebra Exposition

Whether you're preparing for a mid-term or final, having an interactive study guide is a great way to convey material in a different way to facilitate different memory formation. You can take a look at an example of a presentation you could make here: https://infogram.com/algrebra-exposition-1hdw2jogj9le6l0

Did you find a visual study guide to be more useful than a standard written one?

_____

_____

_____

Did you find the presentations easier to learn from than the book or just looking at text?

_____

_____

_____

Do you think it would be worth it to you to make your own visualization as a study guide?

_____

_____

_____

 Were the sources your classmates listed useful to you?

_____

_____

_____
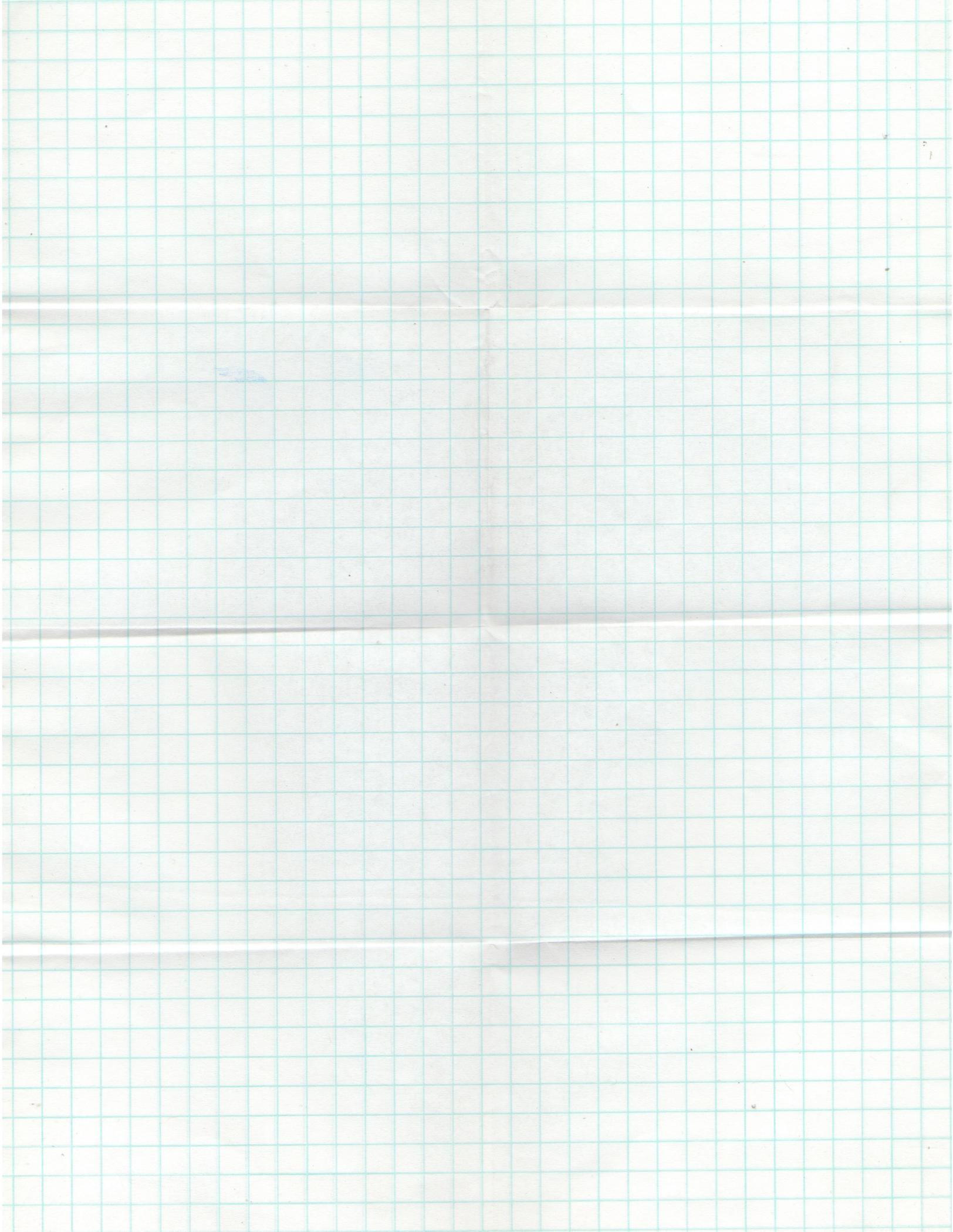
Science

Analyze a Visualization

Do you find the following visualization in figure 1 to be effective?

___Yes   ___No   ___Maybe

Figure 1. A visualization about the oxygen needs for different sea creatures.

Explain your reasoning. If you think it needs improving what would you improve or change and why?

_____

_____

_____

What problems do you see?

_____

_____

_____

Why do you think the air bubbles to the left were included?

_____

_____

_____

Was it a good decision to include the air bubbles?

_____

_____

_____

What other ways could the same data be represented?

_____

_____

_____

Graphing Over Time

If you haven't been provided with the data, then follow the steps below. Go to the URI data sets using the link below.
Link to data: https://web.uri.edu/fishtrawl/data/ OR link #4 in "Data Sets and Links"

1. Download the Fox Island and Whale Rock 2006 to 2017 data set.
2. Choose two or more categories to graph.
3. You will want to only select some of the data due to the large amount of data provided. An example would be only 2010 to 2017, a date every 6 months. Dates that have all data available are best to avoid the need to clean the data.
4. Create a visualization from your data.

What type of graph did you choose and why?

_____

_____

_____

Do you see any trends?

_____

If there are trends what are they and why do you think they are there? Remember that

correlation is not causation!

_____

_____

_____

Compare and contrast the visualizations. Is one type of graph clearer or easier to read?

_____

_____

_____

Model My Watershed

Going to the watershed app, you will see a page about Model My Watershed and the link to run it.

Link to app: https://wikiwatershed.org/model/ OR link #8 in "Data Sets and Links"

1. Run the browser application, click "Get Started," and select the area you want to look at.
2. In this example, I looked at different school districts in Rhode Island, so I zoomed the map in and chose school districts under the "Select boundary" option.
3. Choosing any school district will bring up two panels and a number of options within each (*fig. 1*).
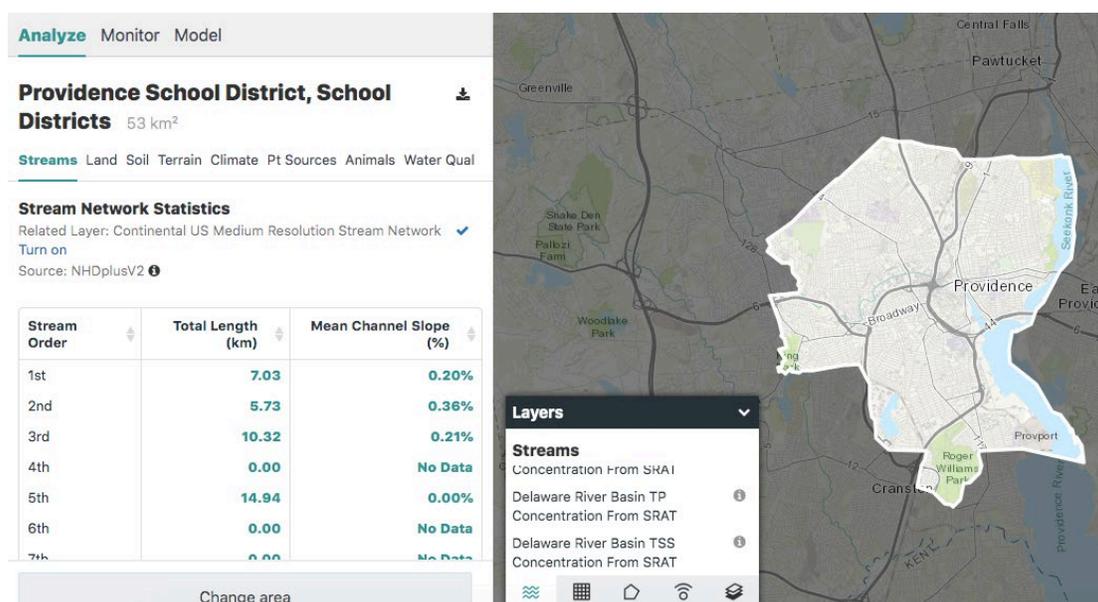
Figure 1. The Providence, RI school district and its data from the Model My Watershed app.

The left panel's "Analyze" tab lets you look at different types of data and download a CSV file for each one. The "Monitor" tab allows you to query for past or ongoing data collections, and "Model" lets you simulate natural and human-made impacts on the environment. The "Layers" panel on the right lets you apply different streams, coverage grids, boundaries, observation stations, and basemaps to the map.

I downloaded land data for the North Smithfield, Providence, and Narragansett and compared the percent of land coverage by type. Below (*fig. 2, 3, and 4*) are three visualizations I was able to make after downloading the three data sets and combining them into one workbook.
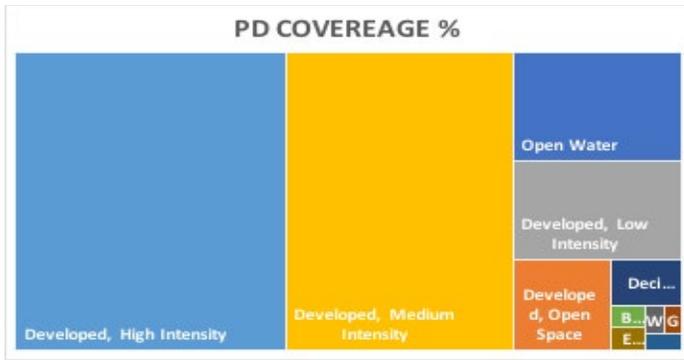
Figure 2. A treemap of Providence schools district's different land types.
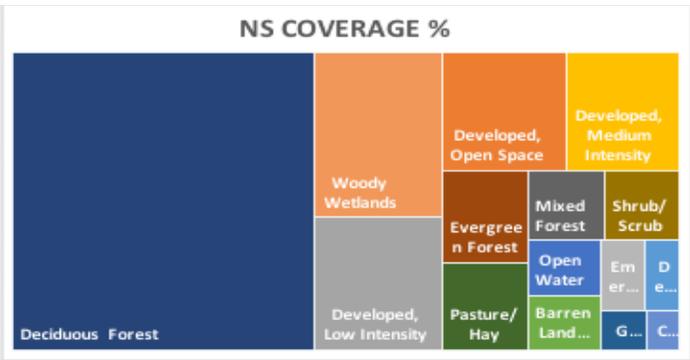
Figure 3. A treemap of North Smithfield school district's different land types.
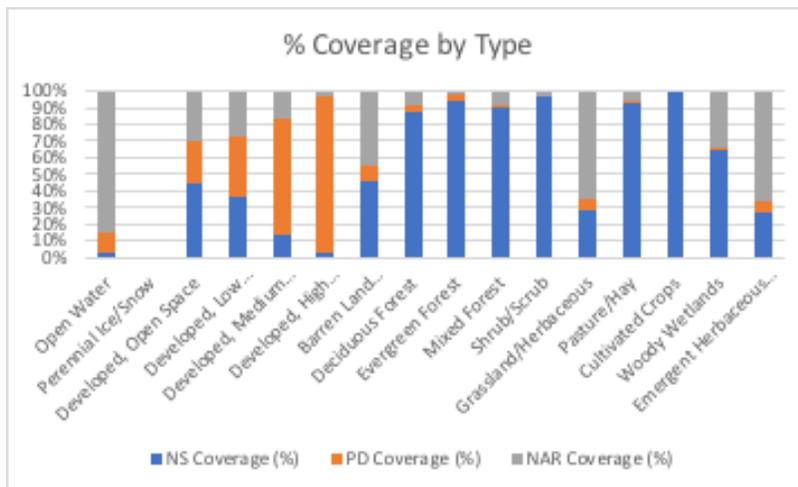
Figure 4. A 100% stacked bar chart comparing how much of each land type the North Smithfield, Providence, and Narragansett school districts have.

Do the three graphs tell you the same information?

_____

_____

_____

What type of graph do you find easier to read and understand?

_____

_____

_____

Are you familiar with this type of graph (a treemap)? What does it tell you about the makeup of North Smithfield?
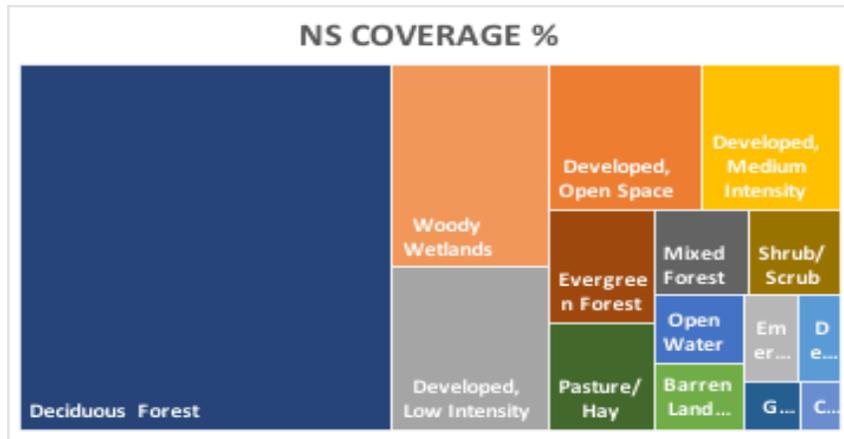
Figure 3. A treemap of North Smithfield school district's different land types.

_____

_____

_____

Are you familiar with this type of graph (100% stacked bar chart) What does it tell you about the makeup of North Smithfield, Providence, and Narragansett?
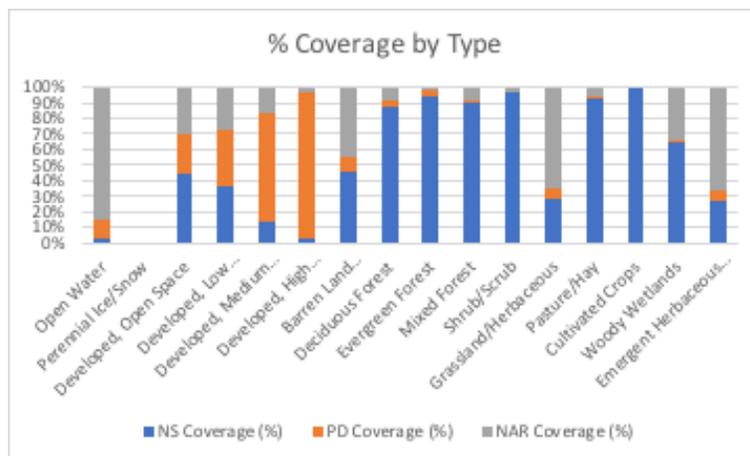
Figure 4. A 100% stacked bar chart comparing how much of each land type the North Smithfield, Providence, and Narragansett school districts have.

_____

_____

_____

After looking through the data yourself, what type of land coverage seems to have the most impact on the amount of runoff?

_____

_____

_____

What do you think is the best mix of land coverages that does not result in excess runoff?

_____

_____

_____

Which location has the least amount of runoff?

_____

_____

_____

Which location has the most amount of runoff?

_____

_____

_____

Navigating NERR Data

The National Estuarine Research Reserve has a wealth of data available for the taking, but it does take a little getting used to. Use the link below to get to a page to get data. The "Data Graphing and Export System" is intuitive to use, but the data and metadata gets emailed to you, so make sure you use an email that accepts correspondence from outside your network!

Link to data: https://cdmo.baruch.sc.edu/get/landing.cfm OR link #9 in "Data Sets and Links"

Opening up your CSV file, it will probably look something like figure 1. Reading the legend will explain the column names and what certain cells mean. In this example, the data revolved around air data, so I trimmed the file down to just the essentials: time stamp, air temperature and its data quality level, barometric pressure and its data quality level, and wind speed and its data quality level.

Tip: It's always best to copy the data to another sheet in the same workbook and edit it there.

Figure 1. The raw data from one of the National Estuarine Research Reserve's data sets.

1. Before we visualize anything, we must filter out data marked as questionable.
2. Click the box to the left of column A and above the first row to select everything, and then click "Data" in the top toolbar, and then "Filter."
3. Every column should now have a filter, so go to each "Flag Code" column and uncheck any code that is below 0 (see *fig. 2*).
4. The next problem you have to fix is changing the time stamp column into two separate columns for date and time. In Excel, click on column D, and choose "Insert."
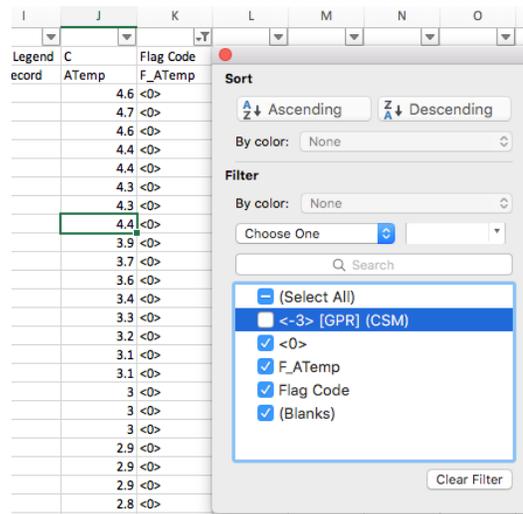
Figure 2. A filter to keep only data with acceptable flag codes.

5. Do this again so you have two blank columns next to the time stamp column (C).
6. Click on C so the whole column is selected, then click on "Data" in the top toolbar, and then click "Text to Columns."
7. Choose "Fixed Width" on the first page.
8. Make sure there is only one break (between the date and time) on the second page (*fig. 3*).
9. Change the column data format to date and make the destination the next column over (in this case D) on the last page (*fig. 4*).
10. Click "Finish" and you're back in the spreadsheet. Right-click on the column with the dates, choose "Format Cells…," and make sure they are set to "Date."
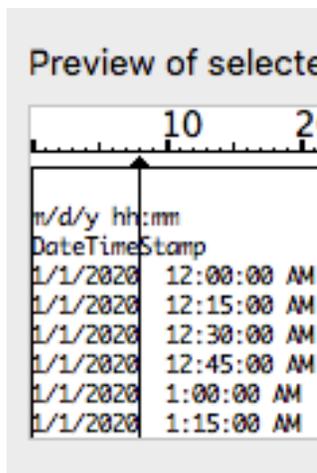
Figure 3. Page 2 of the "Text to Columns" tool. With the fixed width option, you choose where the text is broken up.
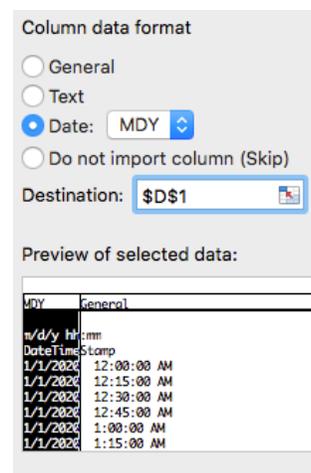
Figure 4. Page 3 of the "Text to Columns" tool. The highlighted column contains dates, so it is formatted as a date. The destination is the next column over.

You are now ready to make visualizations! My two examples (*fig. 5 and 6*) are air temperature over the course of 1 day and air temperature at 12 AM over the course of 7 months.
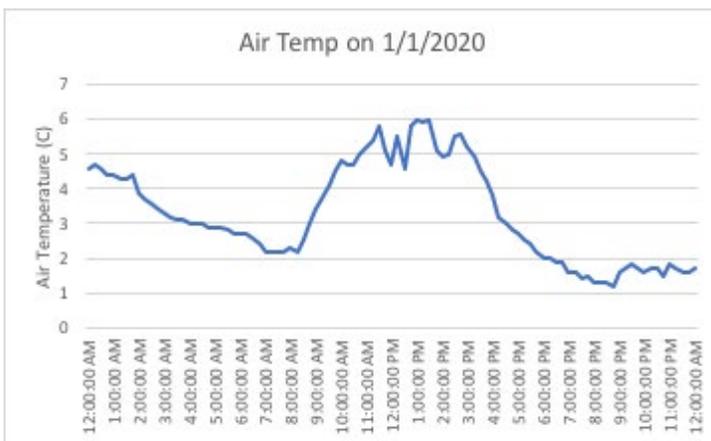
Figure 5. A line graph of air temperature over the course of a day.
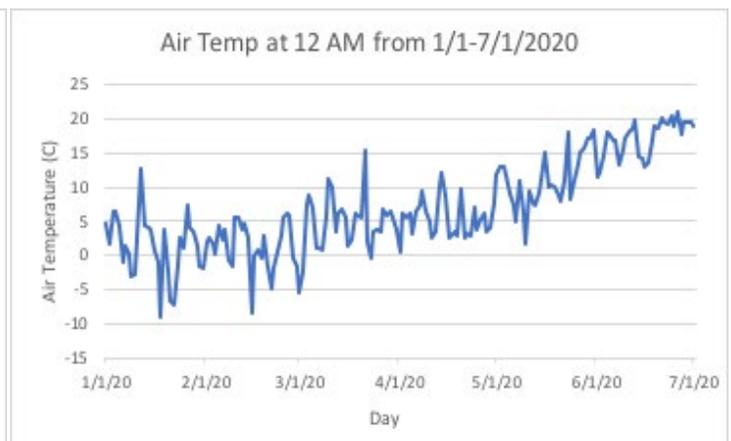
Figure 6. A line graph of air temperature over the course of 6 six months.

What do you think would happen if you did not clean the data and simply tried to make a visualization from it?

_____

_____

_____

What other issues do you think could be in a data set that would affect a visualization?

_____

_____

_____

What do the graphs show you that you did not see in the raw data?

_____

_____

_____

Combining Data Sets

Sometimes the answers you need can't be found in one neat data set, which means you have to comb through multiple sets, databases, or even make your own experiments to get all of the necessary data.

This activity looks at data from the Smithsonian Global Volcanism Program and the United States Geological Survey.

1. I downloaded data on eruptions in the United States from here: volcano.si.edu/search_eruption.cfm

2. I downloaded the earthquake data from here: https://earthquake.usgs.gov/earthquakes/search/

3. I set "Date & Time" to be this past decade.

4. I set "Geographic Region" to Conterminous U.S.

5. And I set "Output Options" to export the first 1000 data points as a CSV file.

ID	Year	Latitude	Longitude	Type
Cleveland	2020	52.8	-169.9	Eruption
Semisopochnoi	2019	51.9	179.6	Eruption
Shishaldin	2019	54.8	-164.0	Eruption
Semisopochnoi	2019	51.9	179.6	Eruption
Great Sitkin	2019	52.1	-176.1	Eruption
Semisopochnoi	2018	51.9	179.6	Eruption
Veniaminof	2018	56.2	-159.4	Eruption
Great Sitkin	2018	52.1	-176.1	Eruption
Bogoslof	2016	53.9	-168.0	Eruption
Cleveland	2016	52.8	-169.9	Eruption
Pavlof	2016	55.4	-161.9	Eruption
Pavlof	2014	55.4	-161.9	Eruption
Pavlof	2014	55.4	-161.9	Eruption
Ahyi	2014	20.4	145.0	Eruption
Shishaldin	2014	54.8	-164.0	Eruption
Veniaminof	2013	56.2	-159.4	Eruption
Pavlof	2013	55.4	-161.9	Eruption
Mariana Back-Arc Segment at 15.5°N	2013	15.4	144.5	Eruption
Pagan	2012	18.1	145.8	Eruption
Kanaga	2012	51.9	-177.2	Eruption
Cleveland	2011	52.8	-169.9	Eruption
Pagan	2011	18.1	145.8	Eruption
Cleveland	2010	52.8	-169.9	Eruption
South Sarigan Seamount	2010	16.6	145.8	Eruption
Pagan	2010	18.1	145.8	Eruption
uu60402867	7/21/20	37.5	-113.5	Earthquake
nc73428961	7/21/20	38.4	120.2	Earthquake

Figure 1. A combined spreadsheet of eruption and earthquake data.

6. After looking at both data sets, make a new workbook (*fig. 1*) to combine the two with the five columns you will need: ID, Date, Latitude, Longitude, and Type.

7. Copy and paste the necessary data, but to get the date from the earthquake data you will need to use the "Text to Column" option which is explained in "Navigating NERR Data." The necessary steps are provided at the end of this activity.

8. The only difference is that you will use the delimited option instead of fixed width, and with "T" as the marker (see *fig. 2*).

9. For Type, you just mark the data as being an eruption or earthquake.

Figure 2. Page 2 in the "Text to Columns" tool with the delimiter option. It will separate each cell at the letter "T."

10. Once you have the workbook ready, go to https://flourish.studio/ and sign in to make the necessary visualization.
11. Click new visualization, scroll down to the projection map section, and choose the world map option.
12. Go to the "Data" tab and delete all the values in the right panel under Regions (*fig. 3 and 4*).
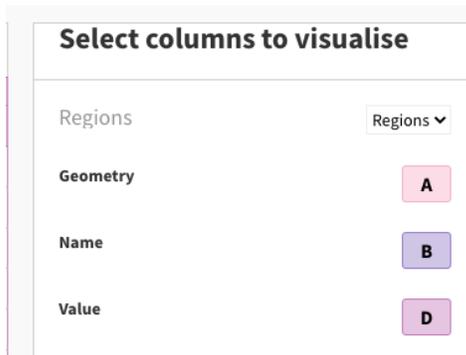
Figure 3. The side panel with data that needs to be changed.
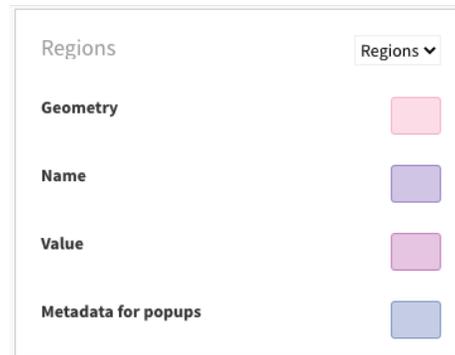
Figure 4. What the side panel should look like.

13. Go to the "Points" tab, and upload your combined workbook.
14. On the right panel, assign columns to the different options. It should look something like figure 5, which would result in the map (see *fig. 6*) also shown below.

Figure 5. What the "Points" data tab and side panel should look like. The columns are ordered as ID, Date, Latitude, Longitude, and Type. The side panel is ordered as A, D, C, E, and B (not shown).
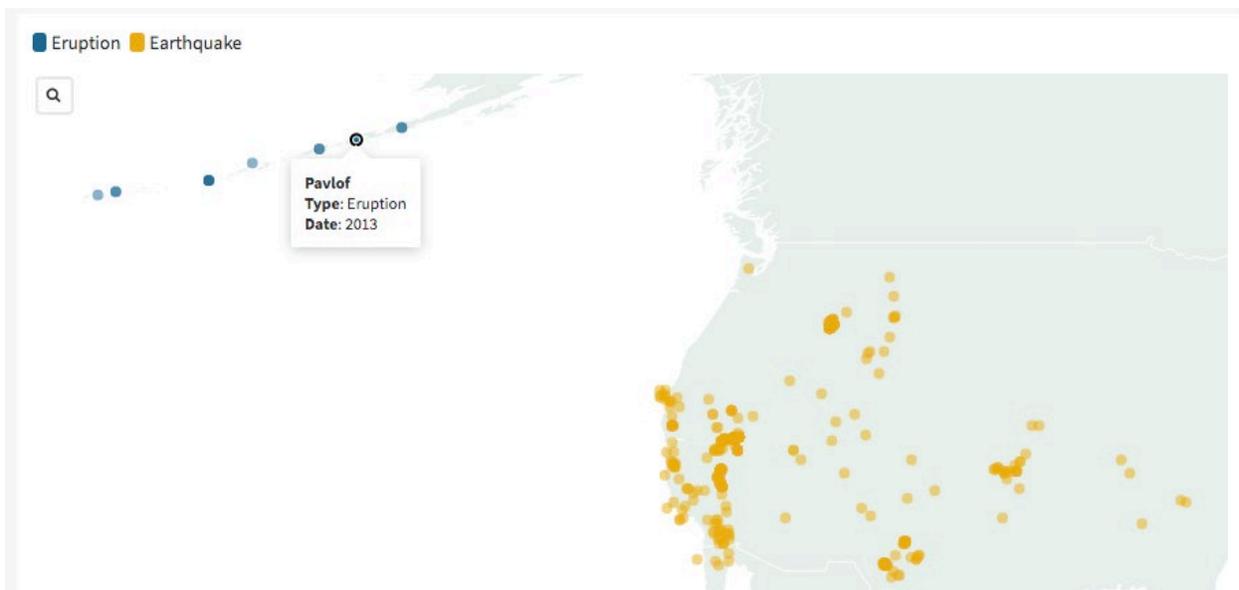
Figure 6. A snapshot of the map or eruptions in blue and earthquakes in yellow. The popup show the ID, Type, and Date.

Text to Columns: From "Navigating NERR Data"

8. The next problem you have to fix is changing the time stamp column into two separate columns for date and time. In Excel, click on column D, and choose "Insert."
9. Do this again so you have two blank columns next to the time stamp column (C).
10. Click on C so the whole column is selected, then click on "Data" in the top toolbar, and then click "Text to Columns."
11. Choose "Fixed Width" on the first page.
12. Make sure there is only one break (between the date and time) on the second page (*fig. 3*).
13. Change the column data format to date and make the destination the next column over (in this case D) on the last page (*fig. 4*).

14. Click "Finish" and you're back in the spreadsheet. Right-click on the column with the dates, choose "Format Cells...," and make sure they are set to "Date."
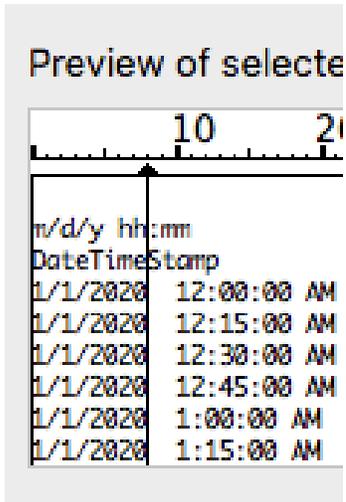
Figure 3. Page 2 of the "Text to Columns" tool. With the fixed width option, you choose where the text is broken up.
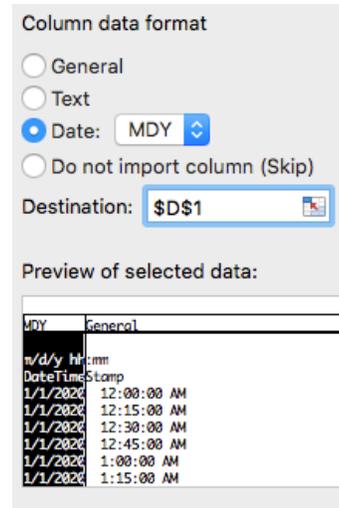
Figure 4. Page 3 of the "Text to Columns" tool. The highlighted column contains dates, so it is formatted as a date. The destination is the next column over.

How can you improve this visualization to display that there are more eruptions in the dataset, but they do not show up because they are from the same volcano?

_____

_____

_____

In terms of tectonic plates, the Alaska-Aleutian Megathrust and the Juan de Fuca are both subduction zones, but why is only the former resulting in volcanic activity?

_____

_____

_____

Are volcanoes in the Cascade Range dormant, or are they gearing up for another period of activity?

_____

_____

_____

Social Studies

National Geographic's Marshallese Stick Charts

This activity gives you a quick look at how some people in the South Pacific Ocean used to navigate the waters around their homes. National Geographic has a great explanation of these stick charts and a few questions to make sure you understand them.

National Geographic: www.nationalgeographic.org/media/micronesian-stick-chart/

If your teacher chooses, you can make a stick chart of your own. You can use whatever materials are usually around you, just like the Polynesians. The chart can be a map of whatever you want.

For example, if you make a map about "traffic" flow in a particular hallway at school, you could use pencils of different colors and lengths to show where people walk and where.

As another example, you could make a map about where you usually spend time in your house. You could use straws as the frame since those are in most people's home, and different objects could represent points of interest. If you spend a lot of time in your kitchen, you could attach little pictures of food to the chart. Anything works!

What location and activity did you choose to make a map of?

_____

_____

_____

Why did you choose that location?

_____

_____

_____

What material did you make your map out of and why did you choose that material?

_____

_____

_____

What did the other students make a map of and what materials did they use?

_____

_____

_____

Do you see common patterns in the what your classmates mapped and what materials

they used?

_____

_____

_____

Voting: Then and Now

Voting is an essential part of our country, and the idea of citizens having a voice extends all the way back to the 5th century B.C.E. The difference between Ancient Greece and modern America is the type of democracy – we have a representative democracy while they had a direct one. Read National Geographic's article about democracy, which can be found at www.nationalgeographic.org/encyclopedia/democracy-ancient-greece/.

Your teacher will then create a ballot box, and guide the class through a re-enactment of each democracy. You will vote on the same law and see if a different democracy results in a different outcome for the proposed law. You can then make a visualization based on the data, as seen in figure 1.

	Yes	No
Direct Democracy	17	8
Representative Democracy	2	4
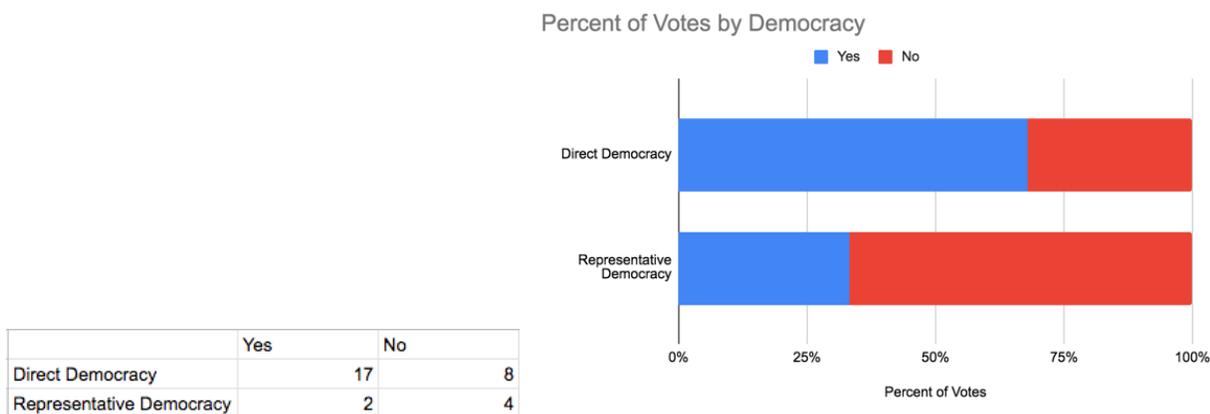
Figure 1. A table of votes and how it looks as a 100% stacked bar chart

After completing the exercise on the direct voting vs representative voting what do you think are the pros and cons of each system?

_____

_____

_____

_____

_____

_____

How do you think gerrymandering impacts the outcomes of elections? What should be done to combat it?

_____

_____

_____

How do you think people that don't vote impacts the outcomes of elections? What are the reasons you think that people don't vote, and what can be done to change their minds?

_____

_____

_____

_____

_____

_____

Also consider the Australian system for people who do not vote. Do you think that could work in America?

_____

_____

_____

Think about how it could be argued that your government and who is in power is a form of data visualization. Your vote is the raw data, and the resulting makeup of Legislative and Executive branches is the visualization.

Is this visualization always accurate to the raw data? For example, there have been 5 times where a candidate had the popular vote but lost the election because of the electoral college. How do the systems in place work to represent, or misrepresent, the data given to it?

_____

_____

_____

_____

_____

_____

Torres Strait Tallies – A Chain of Achievements

In the 19th century, Alfred Cort Haddon discovered that the Torres Strait Islanders used physical representations for social purposes. A "kupe" was a bundle of sticks tied to a string (see *fig. 1*) and could be used to tally a specific achievement, such as the number of turtles they caught.

Figure 1. A drawing of a kupe from the Torres Strait Islanders.

Come up with a list of things you have done more than once and are proud of. For example, it could be the number of good deeds you do in a week, the number of times you participated in/won a competition, how many languages you know. Your kupe will represent one of these list items.

Next, design a kupe of your own. It can have a clear representation or one only you can decipher, similar to the Marshallese stick charts. For example, a digital kupe of good deeds could be a PowerPoint slideshow with each slide representing one deed. The picture on each slide could represent what you did, such as a rake if you raked leaves. For a personal kupe, you could use color-coded popsicle sticks tied together with string.

What did you make your visualization out of?

_____

_____

_____

Why did you choose that material?

_____

_____

_____

What accomplishment did you visualize?

_____

_____

_____

Why is that accomplishment important to you?

_____

_____

_____

Mapping America

Geographic heat maps are a great way to visualize a range of data that can be categorized by state or country. The U.S Census Bureau has years' worth of data that you can download to look at in Excel or Google Sheets, look at in tables, and view as geographic heat maps they created. This activity will walk you through the process of making your own.

1. Let's make a map of population by state based off of data from 2019. To begin, go to the Census Bureau website, find the search bar and type "population by state".
2. On the results page, scroll down until you find the "State Population Totals: 2010-2019" link (*fig. 1*).

Figure 1. The results from searching "population by state" in the U.S. Census Bureau.

3. On the new page, scroll down until you find the "Datasets" section, and click on the Excel file link to download the data as a comma separated value (.csv) file (*fig. 2*).

Figure 2. The Excel file link for population data.

4. Open up your Google Drive, and upload the file by clicking on New --> File upload, or by dragging the file into your browser window.
5. After that, create a new Google Sheets file.
6. In the top left corner of your new sheet click File --> Import. It will bring up your Google Drive, and you can click on the data set you just uploaded.
7. Selecting the set will bring up a list of options, most of which can be left alone. The only one that matters is if you want the data used on the blank Sheet 1 you have, or to create a Sheet 2 and use it there.

8. There will be a lot of data! Don't worry, we will only be focusing on 2 columns: NAME and POPESTIMATE2019.
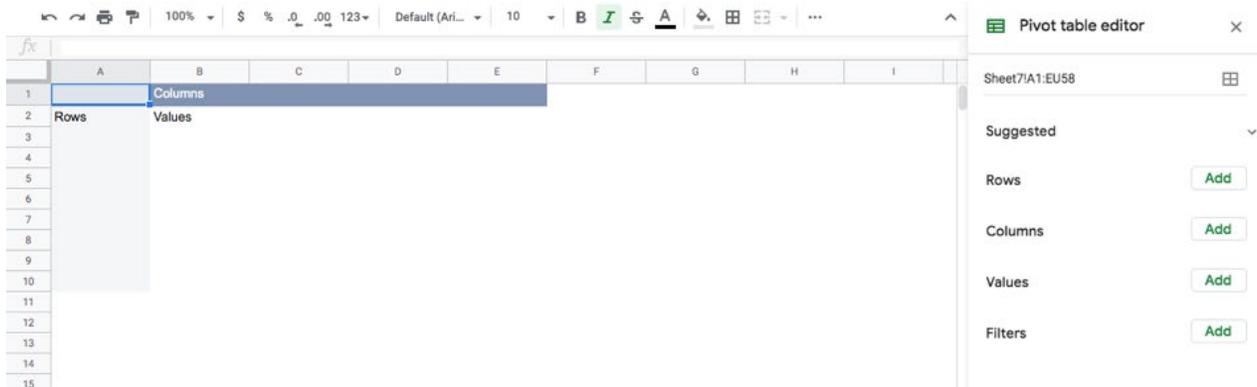
Figure 3. A generic pivot table with no customization.

9. Without anything selected, go to the top toolbar and click on Data --> Pivot Table --> Create (you don't have to worry about the options). Your screen should now look like figure 3.
10. Since we want to display 2019 populations, click on Values' Add --> POPESTIMATE2019.
11. Next, since we want those values organized by state, click on Rows' Add --> NAME.
12. Finally, since we don't want regions or national sums, click on Filters' Add --> NAME --> Showing all items.
13. Scroll through the list, or search, for Midwest Region, Northeast Region, Puerto Rico (the Google chart doesn't include it so we can't leave it to skew data), South Region, United States, and West Region. Uncheck them all, and you're now ready to make the map!

14. Highlight all the values from A1 to B52, go to the top toolbar and click Insert --> Chart.
15. In the Chart Editor, click "Chart Type" and scroll down until you see the "Map" section.
16. Click on the regular "Geo chart" option.
17. To change from a world view, click the Chart Editor's Customize --> Geo --> Region --> United States. Your chart should now look like figure 4.
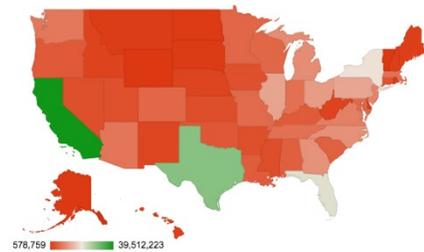
Figure 4. A heat map with the correct data, but no color correction.

18. You should also see color options for Min, Mid, Max, and No Value. By changing those, and generally adhering to the rule that colors should be a range of one shade, I made mine to look like figure 5. I used a Min of light pink, a Mid of light purple, and a Max of dark purple. I feel the pink helps differentiate the low and medium values, but feel free to experiment with your own charts!
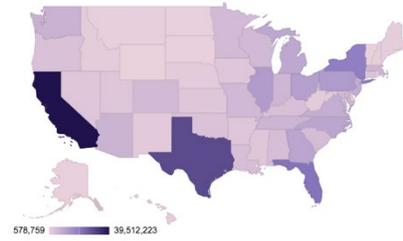
Figure 5. The same heat map as Figure 4 but with customized colors.

Looking at the maps everyone has made, do some colors seem more effective at displaying the same data than others?

_____

_____

_____

What other information would you want to include in a map of estimated population?

_____

_____

_____

How would you make a map showing population change from 2010 to 2019?

_____

_____

_____

Land of the Favored

Pew Research has many interactive data visualization tools that are available for free to play with. One of the interesting ones is a global map of questions asked to those across the world. Amongst those questions is "Please tell me if you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of the United States."

The results can be found here at

https://www.pewresearch.org/global/database/indicator/1

Pick a year, any year, and answer the following questions:

Are there any years that have the same percentage as an answer?

_____

_____

_____

What is the country that favors the US with the highest percentage?

_____

_____

_____

Why do you think that country feels that way?

_____

_____

_____

Make sure that you can also put the list in unfavorable, and answer the following:

Who favors the US the least (in percentage)?

_____

_____

Are the countries with the highest confidence in the president and who favor the US the most the same in your selected year? What year is that?

_____

_____

Why do you think that is?

_____

_____

_____
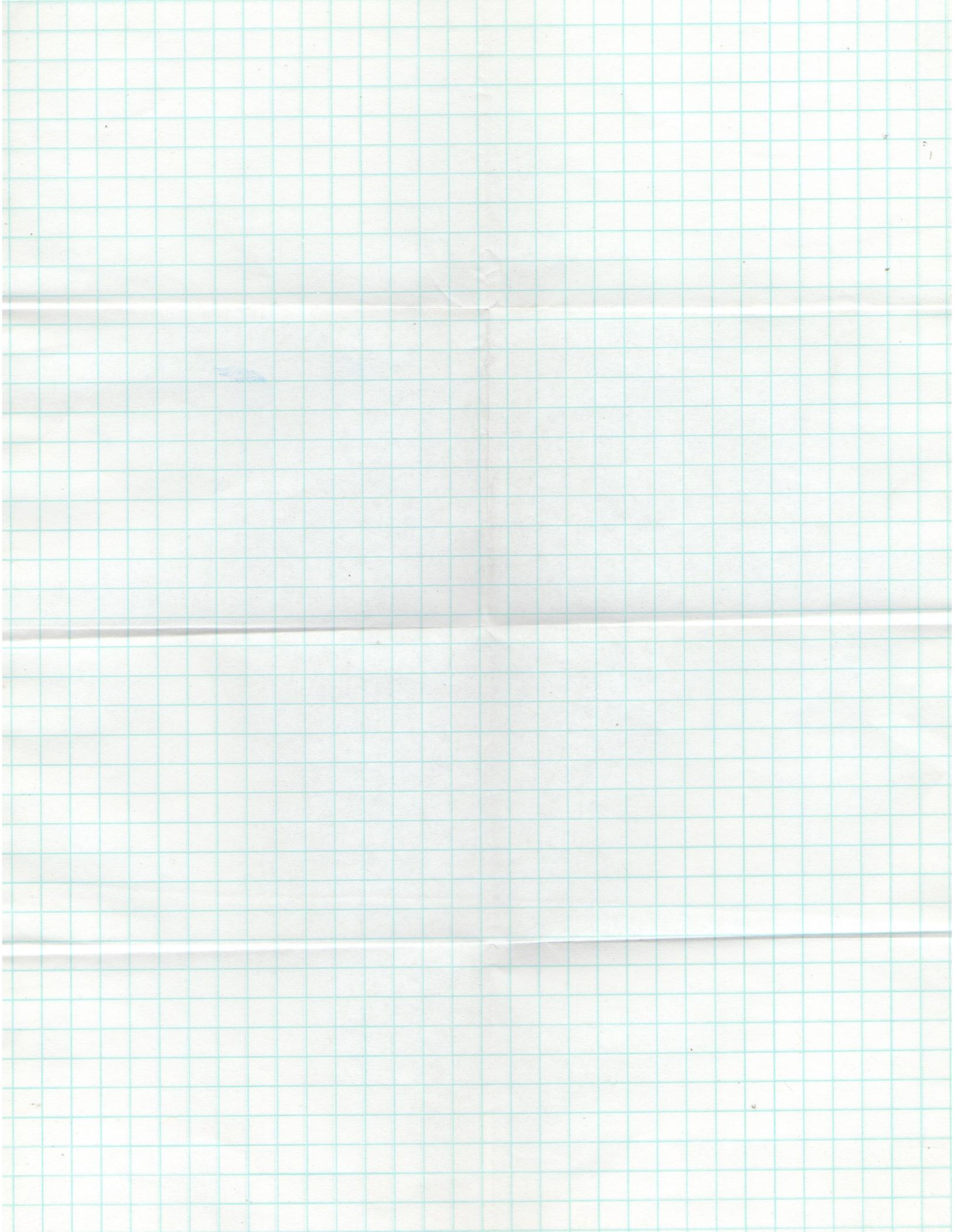
Miscellaneous

What Should I Be Eating?

Understanding what you should be eating at any given meal is always a little confusing. To make things easier, we have grown up looking at different visualizations of a healthy diet – the two most iconic being the pyramid and plate (see *fig. 1 and 2*). Each one has its pros and cons, and today we're going to discuss them.

Figure 1. The most recent diet visualization from the USDA. Source: "A Brief History of USDA Food Guides"
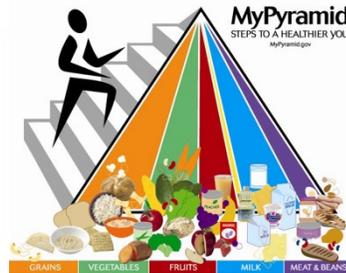
Figure 2. The second most recent diet visualization from the USDA. Source: "A Brief History of USDA Food Guides"

Think about the statistics of what your diet should actually look like. This website, www.center4research.org/myplate-new-alternative-food-pyramid/, details the facts the visualizations should be showing. Some points include getting exercise, varying protein sources, and making half of your grains whole grains. Discuss if either of the visualizations match the facts of a healthy diet.

Work in groups to come up with a diet visualization of your own. It can be an infographic with multiple data visualizations and facts, or one visualization focusing on a single message.

To help you get inspiration for your design, look at www.choosemyplate.gov/eathealthy/brief-history-usda-food-guides. It shows the USDA's old food guides (see *fig. 3*) and explains the drawbacks that led to new and improved ones. Some of them are more of an infographic, some try to use the common chart styles we know about today, and some create new visualization styles (the modern pyramid and plate).
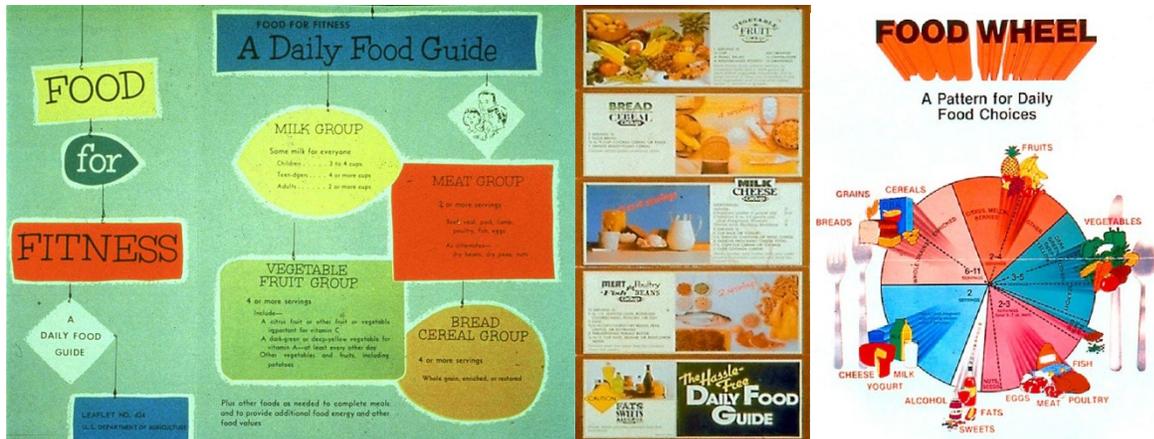
Figure 3. Three of the older food guides that are no longer in use.

How effective do you find the visualizations in figures 1 and 2?

_____

_____

_____

Would you change anything about either of them?

_____

_____

_____

What would you change and why?

_____

_____

_____

What does each visualization tell you about your diet?

_____

_____

_____

Which one is easier to understand in terms of real-life application?

_____

_____

_____

Do either of them look similar to common forms of data representation (bar charts, line graphs, pie charts, etc.)?

_____

_____

_____

Do either of the visualizations match the facts of a healthy diet?

_____

_____

_____

What elements do you like and not like about the older food guides?

_____

_____

_____

Can you see from the older food guides why the guides were changed?

_____

_____

_____

What's So Good About It?

Packaging is one of the most important aspects of a product. It conveys information to consumers about the product's abilities, the brand's stance on certain issues, or how great the product will make your life. All of this information, or data, has to be represented well in order to make a sale; that is exactly Michael Sandberg talked about in his "Datavis as Art: Are Toys Data Visualizations?" article. Taking a look through his blog post with the link below, you'll see he talks about the DIKW Pyramid (*fig. 1*) and how it relates to processing the information given on a toy's packaging.
Link to blog: https://datavizblog.com/2017/06/25/dataviz-as-art-are-toys-data-visualizations/

Figure 1. The Data Information Knowledge Wisdom (DIKW) pyramid. It shows the hierarchy of knowledge processing in people.

After this, work independently to design packaging for a new toy you create, or one that already exists. When everyone is done, you can work in small groups to discuss each other's creations. Did your peers convey the same information in their packaging as in the professionally made packaging? What visual strategies and tactics did they use to convince people to buy their toy?

If your teacher chooses, you can also look at how different packaging correlates to sales over time. Be sure to remember that correlation is not causation! As an example, let's look at Barbie.

From the bar chart found at www.statista.com/statistics/370361/gross-sales-of-mattel-s-barbie-brand/, it shows that prior to 2015 Barbie's sales were declining. After 2015, though, the company started diversifying the Fashionistas line to appeal to more people according to the article found at www.cnn.com/style/article/barbie-vitiligo-hairless-dolls-wellness-scli-intl/index.html. Not only is the product now appealing to people, but

the packaging must convey the new outlook on giving kids a toy that looks like them. A quick Google search of "barbie wheelchair doll" yields results showing the packaging for different women in wheelchairs being prominently displayed, the Barbie logo, and a small infographic about the chair itself (*fig. 2 and 3*).

Figure 2. One of the new products from the Barbie Fashionistas line.

Figure 3. Another product from the Barbie Fashionistas line.

Other students packaging review – what information does it display?

_____

_____

_____

Does it make the toy look fun?

_____

_____

_____

Would you consider purchasing the toy if you saw it advertised?

_____

_____

_____

How much of an effect do you think packaging has on sales and the success of a product?

_____
_____
_____

What do toys tell you about the product?

_____
_____
_____

Do different types of toys emphasize different data?

_____
_____
_____

Without looking at any of the words could you still tell what a toy is and what age group it is aimed at?

_____
_____

Does the packaging you and your classmates created show the same information as the professionally created packaging?

_____
_____

What visual strategies and tactics did you use to convince people to buy your toy?

_____
_____
_____

Self-Portrait in Data

Take a look at the slideshow presentation "Art as Information, Information as Art" by Loren Madsen, a data artist. It can be found here at www.slideshare.net/lorenmadsen/information-as-art. The self-portrait part starts as slide 55.

After looking at how some people define themselves (as a years' worth of receipts, the TV guide on the day they were born, different patterns of movement, etc.), how does it relate to conventional portraits you've seen before? Do feel these ones convey more, or less, information about a person than regular self-portraits?

If you want to visualize yourself as data, but aren't sure what to do, here are some examples. It could be the number of hours you played for different video games, the amount of different foods you consume or products you buy in a week, or what your daily schedule looks like. A simple chart could work, or you could make an infographic with multiple points of data. Infogram is one free source you could use to create a visualization with multiple data sets and some text.

What data did you chose to make your self-portrait?

_____

_____

_____

Do you think it shows more or less than a standard self-portrait?

_____

_____

_____

What information did you choose to make your self-portrait?

_____

_____

_____

Taking a Look at Treemaps

Being able to properly infer information about spending habits and then make data-driven decisions is important for anyone trying to balance their budget. It is also important in a work environment if you're tasked with figuring out their expenditures, sales, or other hierarchical information. Treemaps are perfect for displaying this type of data because categories can be nested inside of their parents. For example, say you have a category for food spending. You can have leaf nodes to show your two sub-categories for eating out and groceries. You can choose to show the nested categories, or just their sum/parent node (food spending). Figure 1 is an example of a visualization you might have to make in the future. We will be finding data and making something similar with Flourish and Infogram.
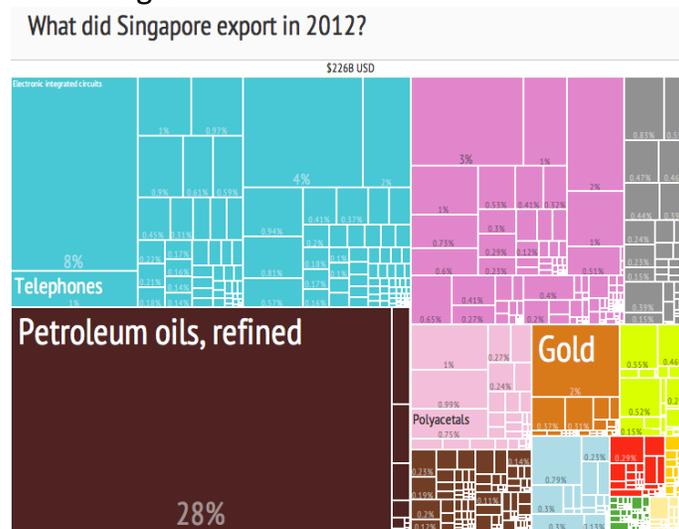
Figure 1. A treemap of Singapore's exports in 2012.

1. Go to wits.worldbank.org, hover over "Trade Stats" in the top left corner, and click "By Country" (*fig. 2*). I chose to look at North America, but you can look at whichever country they want and compare with peers afterwards.
2. In the new screen, choose the second link (Trade Summary) in the "Quick Links" panel on the left. You will see a basic treemap that you should aim to replicate but style in your own way.
3. On the top right, click "Download" to generate a .csv file.

Figure 2. The World Integrated Trade Solution (WITS) website. The "Trade Stats" navigation menu is shown.

4. In the file, there will be data for several years but since we are just focusing on the 2018 data, you can hide the columns or delete the data.
5. Space out the columns so you can read what is in each cell, and then you can start filtering.
6. I wanted export data, so I added a filter to column D (Indicator Type) (*fig. 3*).
7. After this, the data used in the website's treemap emerges (see *fig. 4*).

Partner	Product categories	Indicator Typ	Indicator		2018
…	…	Export		Indicator Type	31
…	…	Export			14
Canada	All Products	Export	**Sort**		58
China	All Products	Export	Ascending	Descending	86
Germany	All Products	Export			
Japan	All Products	Export	By color: None		.8
Korea, Rep.	All Products	Export			
Mexico	All Products	Export	**Filter**		75
United Kingdom	All Products	Export	By color: None		
United States	All Products	Export			54
World	Animal	Export	Equals	Export	09
World	Capital goods	Export	And Or		59
World	Chemicals	Export			53
World	Consumer goods	Export	Choose One		31
World	Food Products	Export			59
World	Footwear	Export	Q Search		58
World	Fuels	Export	(Select All)		58
World	Hides and Skins	Export	Development		32
World	Intermediate goods	Export			48
World	Mach and Elec	Export	✓ Export		31
World	Metals	Export	Import		39
World	Minerals	Export	Tariff		41
World	Miscellaneous	Export	Trade Indicator		28
World	Plastic or Rubber	Export			04
World	Raw materials	Export			97

Figure 3. The filter applied to the "Indicator Type" column. You can either filter by deselecting everything except "Export," or you can filter by only allowing data that equals "Export."

22	United Kingdom	All Products	Export	Trade (US$ Mil)-Top 5 Export Partner	
24	United States	All Products	Export	Trade (US$ Mil)-Top 5 Export Partner	337828.64
27	World	Animal	Export	Export(US$ Mil)	41077.09
28	World	Capital goods	Export	Export(US$ Mil)	620831.69
29	World	Chemicals	Export	Export(US$ Mil)	199201.63
30	World	Consumer goods	Export	Export(US$ Mil)	551352.81
31	World	Food Products	Export	Export(US$ Mil)	61564.69
32	World	Footwear	Export	Export(US$ Mil)	2459.58
33	World	Fuels	Export	Export(US$ Mil)	291995.68
34	World	Hides and Skins	Export	Export(US$ Mil)	4893.82
35	World	Intermediate goods	Export	Export(US$ Mil)	434106.48
36	World	Mach and Elec	Export	Export(US$ Mil)	424267.31
37	World	Metals	Export	Export(US$ Mil)	108613.89
38	World	Minerals	Export	Export(US$ Mil)	19985.41
39	World	Miscellaneous	Export	Export(US$ Mil)	224973.28
40	World	Plastic or Rubber	Export	Export(US$ Mil)	97935.04
41	World	Raw materials	Export	Export(US$ Mil)	309453.97
42	World	Stone and Glass	Export	Export(US$ Mil)	93884.81
43	World	Textiles and Clothing	Export	Export(US$ Mil)	30480.39
44	World	Transportation	Export	Export(US$ Mil)	347745.47
45	World	Vegetable	Export	Export(US$ Mil)	94364.52
46	World	Wood	Export	Export(US$ Mil)	72160.94
47	World	All Products	Export	Exports (in US$ Mil)	2115603.55
50	Canada	All Products	Export	Partner share(%) Top 5 Export Partner	14.17

Figure 4. The export data that will be used in the treemap. It has the product type and amount of money it made.

8. It will be easiest to save this highlighted data into a new workbook along with the column names. If you also want to include the exports' product share percentage, scroll down and copy that data into the new workbook.

9. Going to flourish.studio/, sign in with a Google account and choose the treemap option from the template list.

10. You will see an already made treemap, and going into the "Data" tab, you will see how they did it.

11. The panel on the right describes the different nesting, sizing, filtering, and popup options you can utilize.

12. Upload your new workbook, and then we can condense everything down into 3 columns (Product categories, 2018 Export (US$ Mil), and Export Product share (%)).

13. In the panel on the right, nest with column A, size by column B, and add C to the popup info (*fig. 5*). Your visualization should look like figure 6.

Figure 5. The correct "Data" tab. The table has an extra column for each product's share percentage. The right panel shows how "Size by" and "Info for popups" should be assigned.
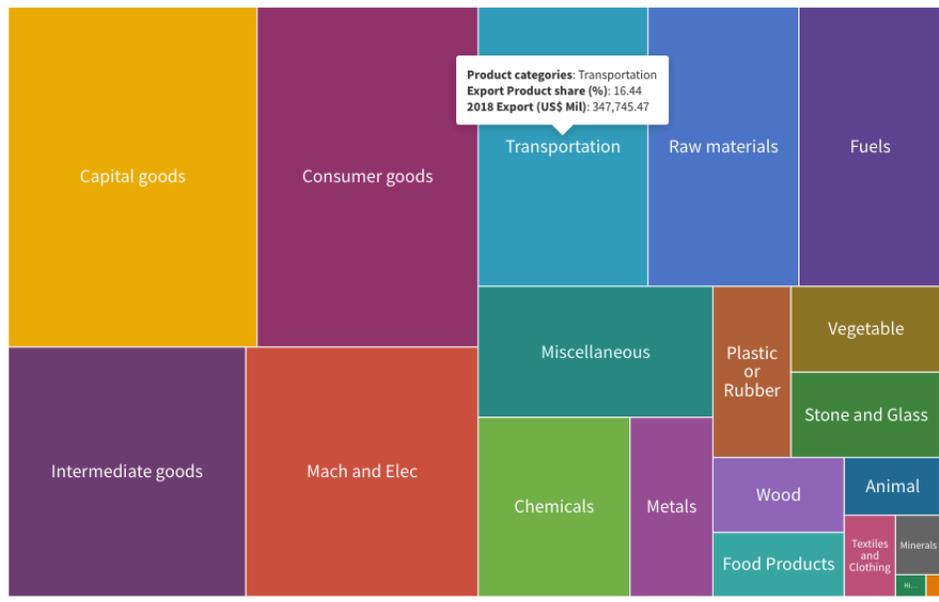
Figure 6. The resulting treemap and popup information based on the data and settings of Figure 5.

Putting the same data from Flourish into Infogram yields a slightly different visualization. Instead of a popup with the share percentage, it just makes the same treemap in a different tab (see *fig. 8*). While these visualizations are good, they still don't show the nesting capabilities that treemaps specialize in, so I will add some more generalized categories in Flourish to sort this data.

14. Taking a quick look at the product categories, I added a column to the left called "Category" and made "product categories" into "Product."

15. My categories were Fuel, Goods, Organic, and Other (not the best but this is just for demonstrating purposes). I organized the products and the treemap came out to look like figure 7.

16. The popups change to only show the lowest category information. Also, the nesting feature also does not work on Infogram.

17. Remember to add appropriate titles and metadata (where you got the data or how it was collected) (*fig. 9*)!
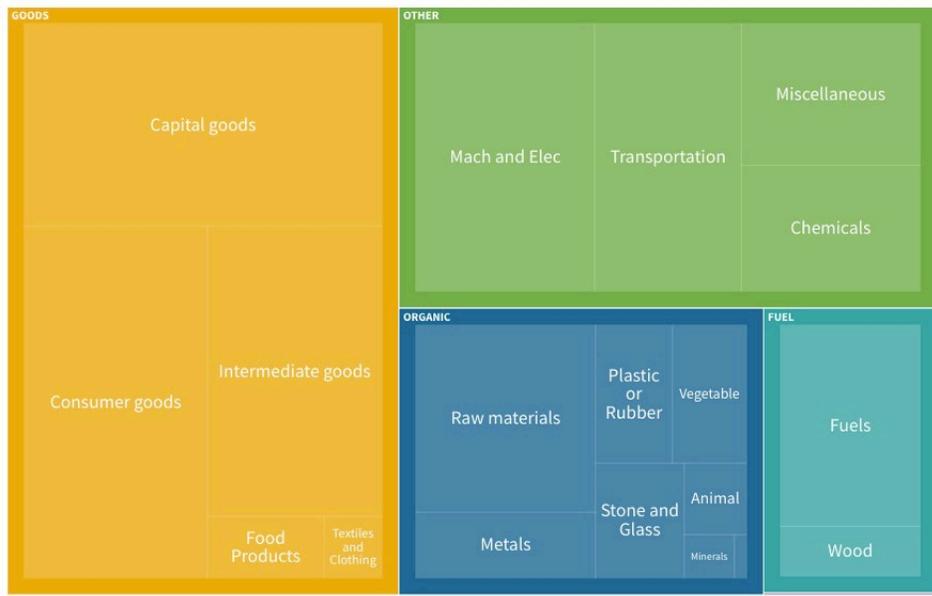
Figure 7. A treemap where the data has been categorized into broader categories. The visualization shows the hierarchical nature of treemaps.

Figure 6. A treemap with the same data as Figure 5, but made in Infogram instead of Flourish. The share percentage column becomes a separate treemap instead of being used for popup information.

Figure 5. An example of a title and explanation of where the data came from for this treemap.

What data did you choose for your treemap?

_____

_____

_____

Does it tell you anything about the data that you did not see by looking at the numbers?

If so, what does it tell you?

_____

_____

_____

How useful do you find treemaps?

_____

_____

_____

What other information do you think you could use treemaps for?

_____

_____

_____

Have you seen treemaps before?

_____

_____

_____

Wearable Data

The goal of this activity is not to make clothes like the ones in figure 1, but instead something much simpler. You can convey data meaningfully with just a scarf, for example. So, we will be making a temperature scarf here to introduce the idea of data that you wear.

Figure 1. Laura Perovich's collection of clothes visualizing the chemicals found in different homes.

1. Temperature data can be found here at www.timeanddate.com/weather/usa/providence/historic?month=1&year=2019. I set the location to Providence, but you can change that. You can also change what date you want to look at.
2. To start the project, choose a range of time you want the scarf to cover. It could be 1 day, 1 month, 3 months, or anything you think would work.
3. Next, choose a ratio of scarf rows to temperature point. For example, if doing 1 day and there are 26 recorded data points, you could have a ratio of 3 rows to 1 point. This would give you 78 rows of yarn for your scarf.
4. Look at the range of temperatures you have and create color codes for specific subsets of that range.
5. For example, looking at January 1st, 2019, the range of temperatures is 34-58. Since the difference is 24, you could have 6 colors to represent subsets of 4 (34-38, 39-42, etc.) or 4 colors to represent subsets of 6. To show change over time, using 6 colors might be better with such a small range. Keep in mind that adding more colors to your scarf does mean buying more yarn, though.
6. Some color codes would be what you see on a regular weather map. Dark blue or purple for extremely cold or freezing, lighter blues for cooler temperatures, and yellows, oranges, and reds for warmer temperatures.

If you would like to see more of the designs shown at the beginning of this activity, you can check them out here at www.lauraperovich.com/projects/dressedindata.html.

If you're more interested in the Home Economics side, you could make pieces for a child based on their data like Seung Lee did. He made a blanket out of his son's sleeping pattern over his first year of life (*fig. 2*). His Twitter thread describing the process can be found here at https://twitter.com/Lagomorpho/status/1149754592579600384. If that link doesn't work, you can also look up "Seung Lee child blanket."

Figure 2. Seung Lee's sleep blanket. He tracked his son's sleeping pattern and made a blanket out of the data.

What data would you chose to make a wearable visualization? (personal data, public data)

_____

_____

Why did you choose the data set?

_____

_____

_____

Is there a reason behind the colors you chose? If yes what is the reason?

_____

_____

_____

Appendix

Data Sets and Links

1. The 2020 Data Visualization Best Practices Workshop and resources.
    a. https://drive.google.com/drive/folders/1uzdPAscZDi77AazK0Lgwkl-NxZO_o9pm?usp=sharing
2. Article about visualizations and ethics, had a good image of the Challenger Space shuttle O – ring image.
    a. https://www.adam-p-jones.com/why_is_visualization_important/
3. Bootstrap, from Brown University. Math, science, and computer science related.
    a. https://www.bootstrapworld.org/materials/spring2017/courses/data-science/resources/teachers/index.shtml
4. Fish trawl, water temperature and related data from the University of Rhode Island.
    a. https://web.uri.edu/fishtrawl/data/
5. Direct link to the snowfall totals map
    a. https://gis.ncdc.noaa.gov/maps/ncei/rsi
6. The math activity "So Many People!" uses this link to Census Bureau data
    a. https://www.census.gov/schools/facts/
7. Narragansett Bay water quality, links to other related sites as well as visualizations and research done on the water.
    a. http://www.watershedcounts.org/marine_water_quality.html
8. A watershed modeling app from the Stroud Water Research Center.
    a. https://wikiwatershed.org/model/
9. NERR – National Estuarine Research Reserve data, multiple data sets included.
    a. https://coast.noaa.gov/digitalcoast/data/nerr.html
    b. Or also here is a link to NERR data
    c. https://cdmo.baruch.sc.edu/get/landing.cfm
10. Links provided by Dr. Peter Meyer from RIC's Physical Science Department
    a. U.S. Geological Survey: https://www.usgs.gov/
    b. National Oceanic and Atmospheric Administration: https://www.noaa.gov/education/resource-collections/data
    c. Smithsonian Institution Global Volcanism Program: https://volcano.si.edu/
11. Census bureau data has links to specific math related problems with visualizations.
    a. https://www.census.gov/programs-surveys/sis/activities/math.Grades_9-12.html