ESR 173 Lab 4: Data Visualization
Storytelling with Science

Purpose:

The purpose of this exercise is to give you the opportunity to explore data visualization tools and their applications. You'll be able to practice deciding what type of tool to use for data you find, and data you have collected. You'll also learn more about strategies to interpret visual data such as charts and graphs.

Background:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data (Tableau, 2019). Data visualization is a tool for visual storytelling which can be especially important in the realm of science, to ensure that the message or narrative being told is clear to not only the scientific community but the public. Not all graphs or charts help in telling a story - too simple or too elaborate and the message can be lost.

The creation and use of data visualization are skills to be practiced, and worth learning.

What is data?

There are two main types of data that scientists collect when exploring different scientific concepts - qualitative and quantitative. Data is commonly the product of measurements or observations.

Qualitative - quality - information not numerical in nature; descriptive, and regards phenomenon which can be observed but not measured.
Ex. Public opinions about climate action

Quantitative - quantity - information that is numerical in nature; measurable
Ex. Temperature, the amount of CO_2 in the atmosphere, ocean pH

NOTE: In some cases and particularly in Chemistry, quantitative data can be collected through modeling or calculations, rather than direct measurements, depending on the **scale** (i.e. length of a bond, quantity of a byproduct in a chemical reaction)

ESR 173 Lab 4: Data Visualization
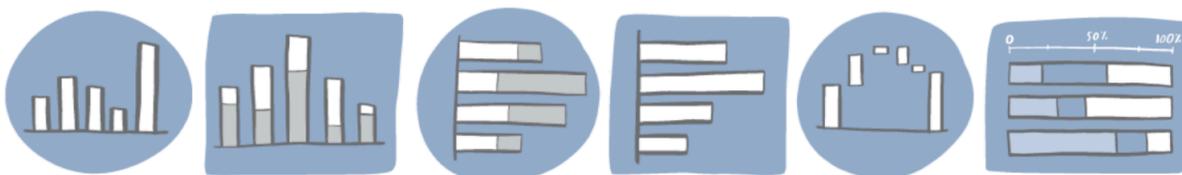
Storytelling with Science

What are the major types of data visualization tools, and how are they used?

Review some great overviews by Storytellingwithdata.com to explore the following:

Line graphs

Bar c0harts

Pie charts

MYTH:
PIE CHARTS are EVIL

When USING a PIE, ASK yourself WHY?

Studies prove that people read pies & donuts by comparing AREA, not ANGLE

If you think pies make sense for your data & audience, test them out to see!

Instructions:

☐ Follow the instructions in this packet to explore different visual tools for representing data.

☐ Reflect on applications of data visualization tools as a means of storytelling in Chemistry.

☐ Reflect on applications of data visualization tools for your Perspectives Project topic.

1. Explore one or more the following websites for examples of data visualization:

ESR 173 Lab 4: Data Visualization
Storytelling with Science

- http://www.storytellingwithdata.com/
- https://informationisbeautiful.net/
- https://pudding.cool/
- https://www.census.gov/dataviz/

    a. What types of elements (images, formats, features, animations, etc.) are used to visually represent **quantitative data**? Include a description of at least one (1) example, including the site from which you observed the visualization (1 point)

A horizontal bar chart, as a chart type for graphing categorical data or data, is sorted into groups. Used to visually represent quantitative data. It consists of multiple rectangles aligned to a common baseline that visually represent the percentage of top countries with coronavirus doses given to the people. The bars are separated into categories with different colors and themes. The length of each is proportional to the value it represents. When objects are aligned, making this graph easy to interpret. It's from COVID-19 #CoronaVirus Infographic Datapack — Information is Beautiful .

    b. What types of elements (images, formats, features, animations, etc.) are used to visually represent **qualitative data**? Include a description of at least one (1) example, including the site from which you observed the visualization. (1 point)

Table, organizes data into a tabular structure, consisting of rows and columns. Used to visually represent qualitative data. It consists of multiple categories made into low, medium and high-risk activities with colors to differentiate the three, aligned along the line that visually represent order of increasing risk. The size of each is proportional to the value it represents, making this graphic easy to interpret. It's from COVID-19 #CoronaVirus Infographic Datapack — Information is Beautiful

    c. What about these examples surprised you if anything? OR what questions were raised when viewing them? (1 point)

ESR 173 Lab 4: Data Visualization
Storytelling with Science

> Tables can be great when you have a diverse audience. But I am imagining If there is a pattern or something significant about the shape of my data, a graph will be a better choice over a table. Or is there a way to limit unnecessary rows and columns or direct attention to a subset of the data, making the table faster to digest? Sometimes we need to communicate related metrics with multiple units of measure or varying ranges.

2. As mentioned previously, not all visual tools tell a story that is accessible for everyone. Explore Kaiser Fung's website JunkCharts.typepad.com and describe, briefly, an example of a poor data representation based on your own interpretations.

- a. Include the following information (where applicable) (1 point):
- Topic of the data visualization
- Form (graph, bar chart, infographic, line graph, or other imagery)
- Representing quantitative data, qualitative data or both
  - o What data in particular did you observe

> The topic of data visualization is visualizing fertility rates around the globe. The chart form is a column chart that is curled up into a ball that tries to do a lot. It represents Both qualitative and qualitative data. The data visualization doesn't tell you anything important. Some information is injected in those arrows that connect the columns to the data labels. Beyond this basic setup, the designer embellishes the chart with a trove of information observed only to close up view. The column chart can only take a single data series. The data set contains fertility rates from over 200 countries over 20 years.

- b. What about the example selected made sense or clearly connected with the message it was trying to portray? (1 point)

> In the example, a threshold is about the births necessary to avoid a population decline. If that's the case, then comparing each country's fertility rate to a global fertility rate threshold is too simplistic because fertility is only one of several key factors driving a country's population growth. A more sophisticated model should generate country-level thresholds.

    c. Why was the example unsuccessful in your opinion? (1 point)

> Because the main insight I obtained from it is a rough ranking of continents, with African countries predominantly having higher fertility rates, followed by Asia and Oceania, then Americas, and finally, Europe which has the lowest fertility rates. If this is the key message, a standard choropleth map brings it out more directly.

3. Explore three (3) of the following websites that host web-based access to data visualization and infographics:

    A. https://www.compoundchem.com/category/environmental-chemistry/
    B. https://fivethirtyeight.com/features/the-40-weirdest-and-best-charts-we-made-in-2020/
    C. https://infogram.com/examples/maps
    D. https://www.acs.org/content/acs/en/pressroom/reactions/infographics.html
    E. https://public.tableau.com/en-us/s/blog/2020/12/2020-vizinreview-year-viz-days

    a. **How is this data relevant to our study of environmental science**? (1 point each)

Include the following information (where applicable):
- Which site you explored
- Topic of the data visualization related to environmental science or chemistry
- Form (graph, bar chart, infographic, line graph, or other imagery)
- Representing quantitative data, qualitative data or both
  - What data in particular did you observe

Site 1: https://www.compoundchem.com/category/environmental-chemistry/

The basic form is an infographic that is developed into illustrations of the AC-working mechanism and chemical compounds. The illustrations are given colors and arrows to show directions of air movement. Some information is written and labeled on the sides. The illustration can take more than one data series. The illustration represents qualitative data. I have seen the information on the model and important components (chemical compounds) that are released from air condition. Including CFCs, HCFCs and HFCs. The use of air conditioning is increasing pollution in the environment by releasing poisonous gases into the environment. They harm the environment as they are part of the greenhouse gases that trap heat and lead to the depletion of the ozone layer.

Site 2: [_Water Scarcity - Infogram](Water Scarcity - Infogram)

The topic of data visualization is Global water stress rankings. The basic form is map imagery that developed into a drawing of the world map. The illustrations have given colors and percentages to show stress levels. Some information is embedded into the map and countries are colored depending on stress level. The drawing can take more than one data series. Drawings represent both quantitative and qualitative data. I have seen information on the level of stress from extremely high, high, medium-high, low–medium, and low. Prolonged water stress can have devastating effects on public health and economic development. More communities worldwide lack access to safe drinking water.

Site 3: https://www.acs.org/content/acs/en/pressroom/reactions/infographics.html

Storytelling with Science

The topic of data visualization is Cast iron. The basic form is an infographic that developed into paragraphs and illustrations. The illustrations have given colors to show emphasis. Some information is embedded into real pictures and images. The paragraphs contain more than one data series. The infographic represents qualitative data. I have seen information on how cast iron is prone to rust and on how to prevent it. Improper handling of materials can have devastating effects on materials made by cast iron. Earth crust composed of iron contents, at right conditions chemical reactions occurs to form rust.

    a. Of the sites you observed, which one resonates with you most personally, and why? (1 point)

This one https://www.compoundchem.com/category/environmental-chemistry/ struck me with most personally because we probably love the feel of freshly chilled air on a hot day, but don't forget that law called the conservation of energy! There's always a price to pay for getting something good in our universe. In this case, the price is the energy you have to use to run the air conditioning unit; using energy means there's an impact on your pocket and the planet too, in the shape of environmental problems like global warming.

    b. Reflect on your engagement with these sites and in particular those that most resonated with your learning style. What parallels, if any, do you see between this type of data presentation or "storytelling" and what you have encountered in terms of Chemistry content? (1 point)

Some of these insights are fairly straightforward while others are less obvious. But overall, these data and coverage allow you to better organize, visualize, and verbalize the messages and contents with success and ease, no matter the background you're coming from.

4. In upcoming labs, you'll be developing your own data visualizations using selected datasets related to a variety of environmental science topics. Additional guidance for the use of Google Sheets will be provided.

ESR 173 Lab 4: Data Visualization
Storytelling with Science

Before we apply this learning directly in Google Sheets, consider what data visualization tools would be best to communicate the following stories, and why (datasets from DataNuggets) (3 points)

HINT: Revisit Background section to recall different uses for various types of data visualizations including line graphs, bar charts, and pie charts.

    a. Scenario/dataset 1:

corals from	tank type	number of bleached corals
inshore reef	normal	0/15
offshore reef	normal	0/15
inshore reef	heated	5/15
offshore reef	heated	10/15

Type of data visualization you could make and why

An area graph would be my ideal selection because it emphasizes a part-to-whole relationship over time when one part is very large or changes from being very large to very small. Our eyes are very good at comparing and contrasting when objects have aligned in an opposite form such as heated and normal or inshore and offshore, making it easy to interpret.

    b. Scenario/dataset 2:

Treatment	# of plants surviving to flower	calendar date when first flower appeared	average # of days until first flower (since start of experiment)	standard error (SE)
Normal	28	May 20, 2013	16.11	0.66
Heated	25	May 10, 2013	6.08	1.65

* Standard error (SE) tells us how confident we are in our estimate of the mean, and depends on the number of replicates in an experiment and how much variation is in the data. A large SE means we are not very confident, while a small SE means we are more confident.

Type of data visualization you could make and why

Table I will choose this type because we have varied data and need to communicate related metrics with multiple units of measure. In this case, the end-user will likely reference these items to look up specific values, so a table is an effective choice. Also, tables aren't subject to the same two-dimensional space constraints as graphs, so it can be easier to display different units

c. Scenario/dataset 3:

Location	Ecosystem Type	Organic Matter (%)	Total Phosphorus (ug P/g d.w.)
Wintergreen	Lake	24%	528
Douglas Lake	Lake	55%	523
Jackson Hole	Lake	5%	105
Whitford Lake	Lake	1%	28
Pond 9	Pond	21%	556
Pond 18	Pond	24%	512
Pond 10	Pond	17%	537
Pond 23	Pond	19%	366
Pond 6	Pond	6%	177
Loosestrife Fen	Wetland	40%	773
FCTC	Wetland	80%	1441
Osprey Bay	Wetland	8%	167
Turkey Marsh	Wetland	15%	459
Sheriffs Marsh	Wetland	30%	1909
Brook Lodge	Wetland	18%	443
Eagle Marsh	Wetland	4%	130

Type of data visualization you could make and why

Bubble chart/scatter plot would be the best bet since, data has at least three dimensions, making it easier to display different units and ranges within a single view. From this scenario/datasheet 3, it can be easy to display multiple units with two sets at the X and the Y-axis compared to graphs. It's possible to incorporate visual elements to help draw attention to the most important data or to communicate a pattern compared to other types of forms.

5. Reflect on data you may include in your Perspectives Project. In what ways might you use data visualization to strengthen your message? (1 point)

To strengthen my message, I will;
1. Establish the goal of my visualization.
Decide where your visualization falls on the spectrum between exploration and communication. Then determine what message you want the audience to take away from the visualization.

2. Clean up and understand my dataset.
Make sure you thoroughly understand the data you're trying to explain and you have a clean set of data to work with.

3. Know my audience.
Your audience will help determine what elements to use within the visualization, such as text, colors, and other effects.

4. Choose a type of chart.
Experimentation is part of the process of selecting a chart type, so try visualizing your data in different ways.

5. Don't try to pack too much into one chart.

Some Additional tools to consider when planning for your project:

https://www.chartjs.org/
https://public.tableau.com/s/
https://www.springboard.com/blog/free-public-data-sets-data-science-project/