

A (quick) Introduction to Using Python Notebooks for Ocean Science Research Day 2

Sage Lichtenwalner

2020 Data Labs Virtual REU

June 2020

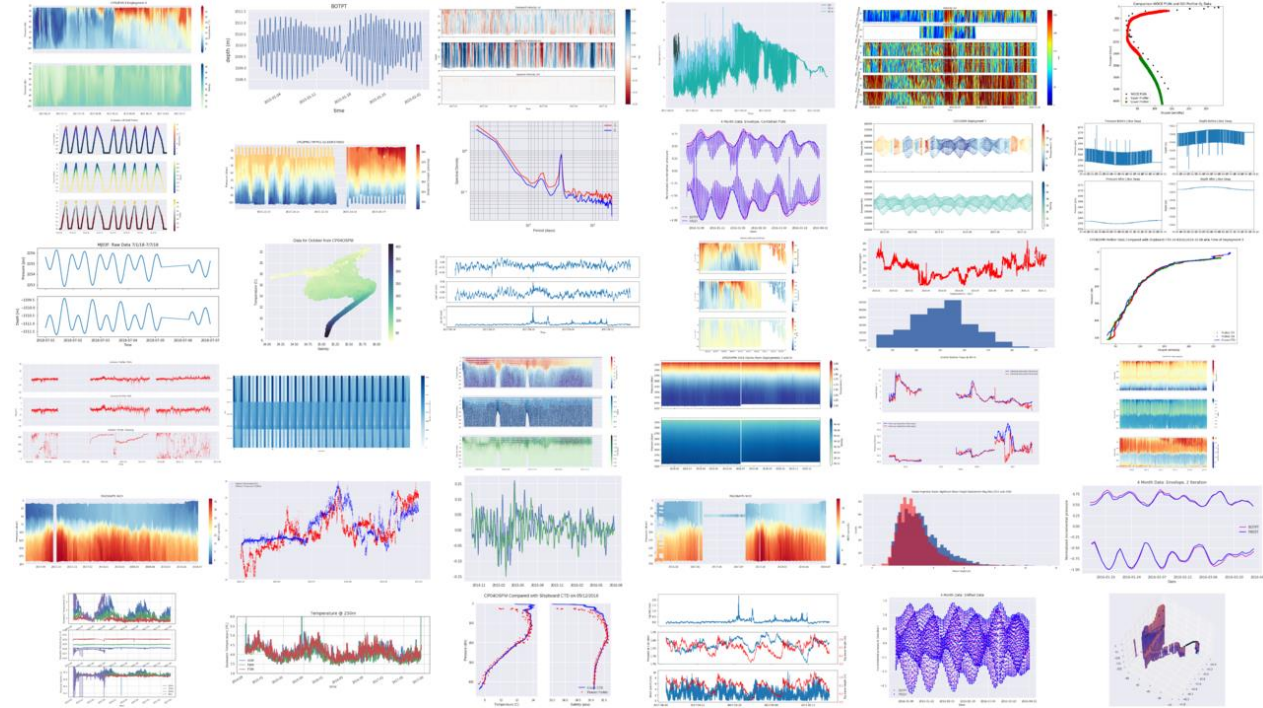


datalab.marine.rutgers.edu

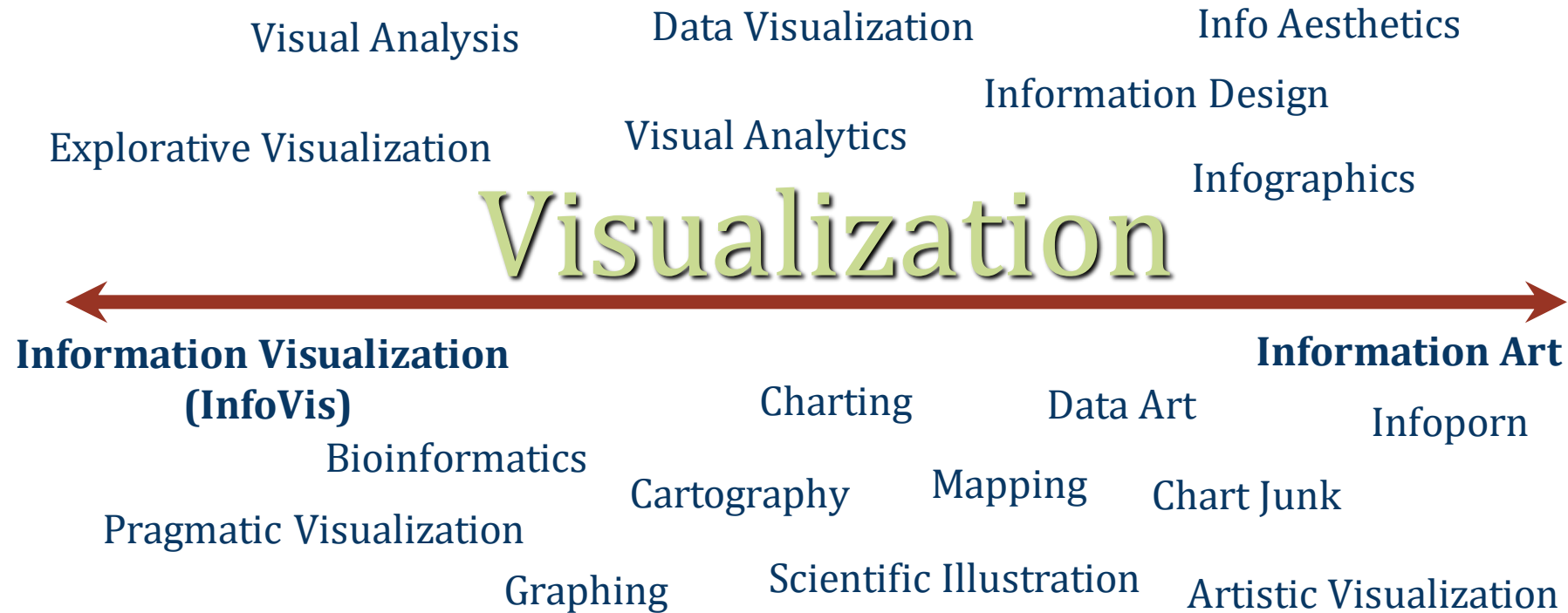


Today's Agenda

- Quick Intro to Data Visualization
- Data Visualization in Python
 - Customizing Plots
 - More visualization types
- Bonus Activity 4 – More plots
 - A new dataset – Argo Floats
 - Profile plots
 - T-S Diagrams
 - Maps



What is “Visualization” ?

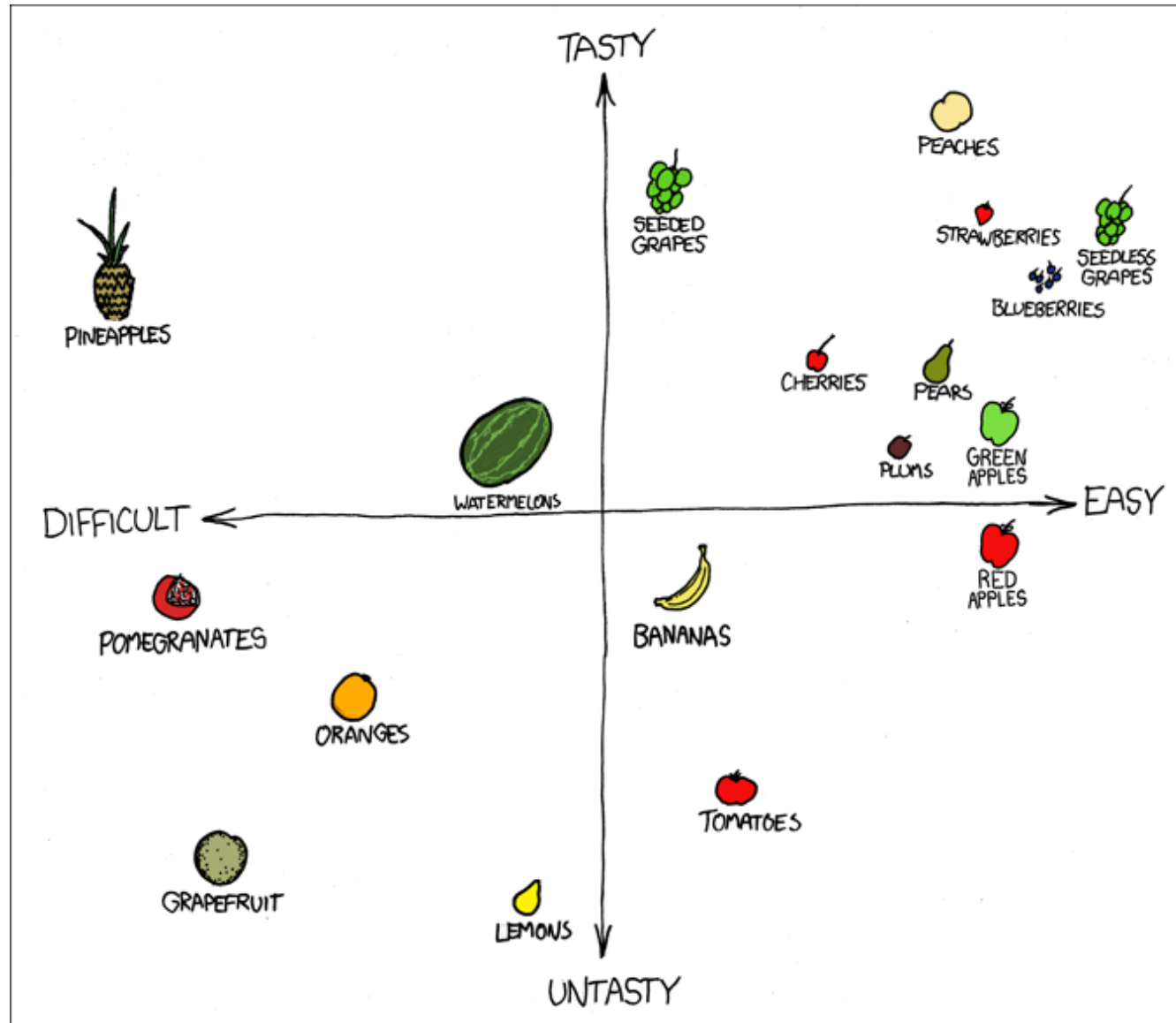


“Visual analysis is not primarily about the pictures, but about finding ways to use our powerful visual systems to analyze data. It's analysis done in a visual way. It's visual exploration, visual data analysis, and visual presentation of results.”

Robert Kosara, eagereyes.org

The World's Greatest Chart?

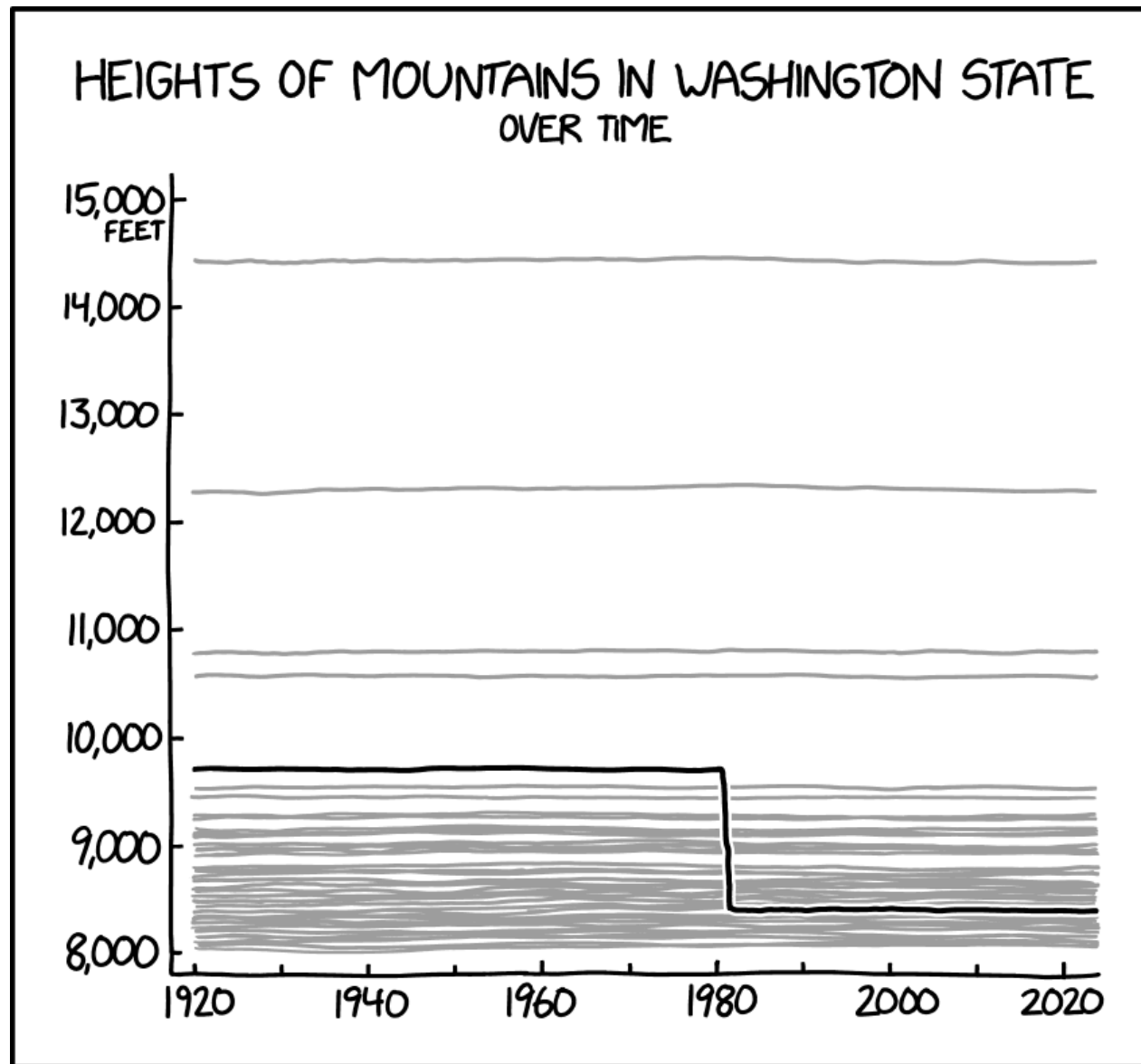




Reference: <https://xkcd.com/388/>

Lots more xkcd charts:

<https://www.explainxkcd.com/wiki/index.php/Category:Charts>

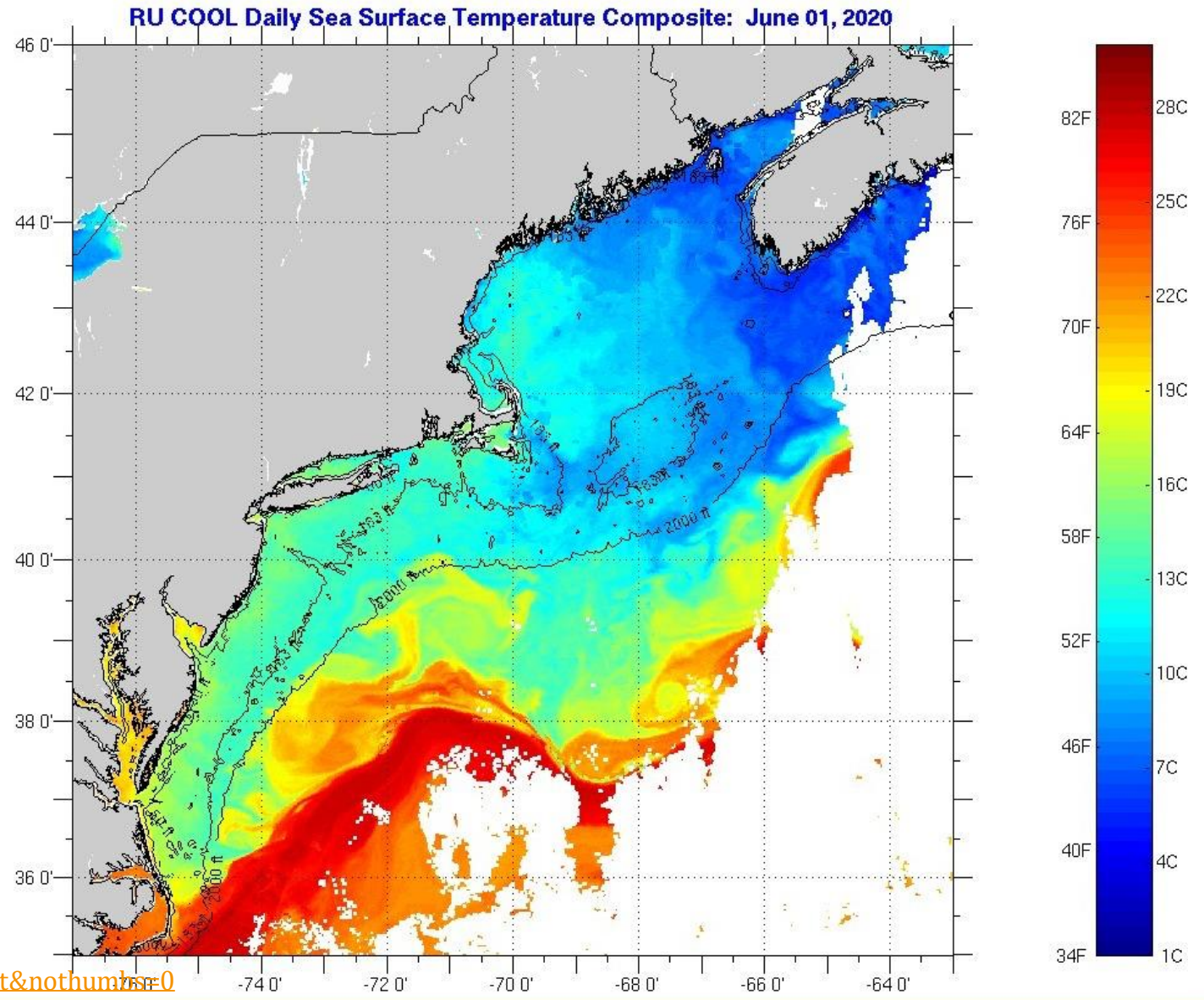


Mount St. Helens

"It's a good mountain, but it really peaked in the 80s."

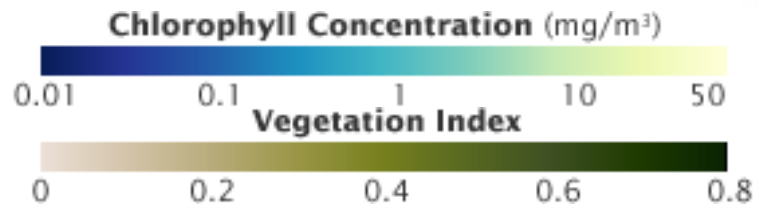
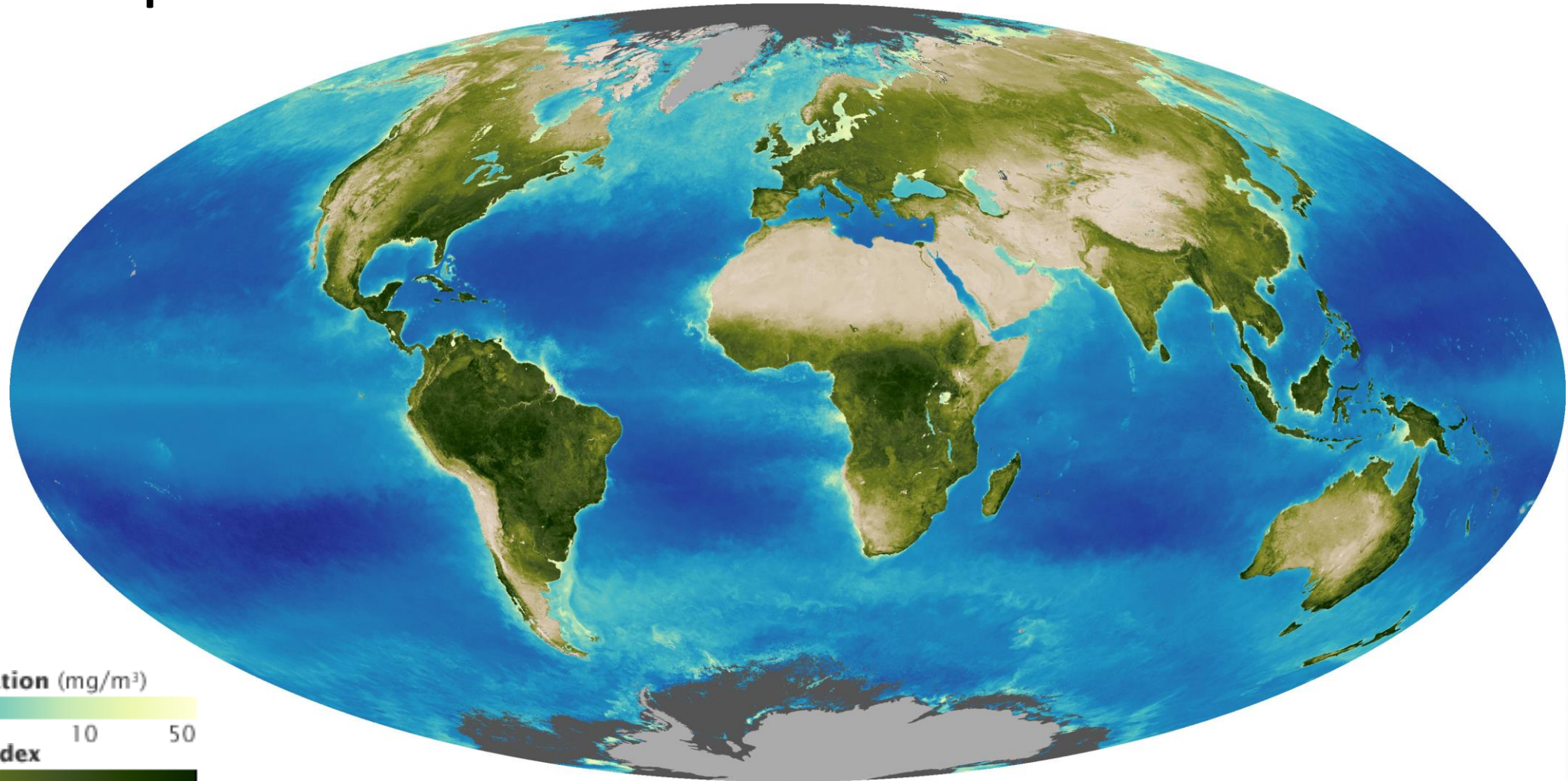
<https://xkcd.com/2308/>
https://www.explainxkcd.com/wiki/index.php/2308:_Mount_St_Helens

Sea Surface Temp



https://marine.rutgers.edu/cool/sat_data/?product=sst_comp®ion=bigbight¬humb=0

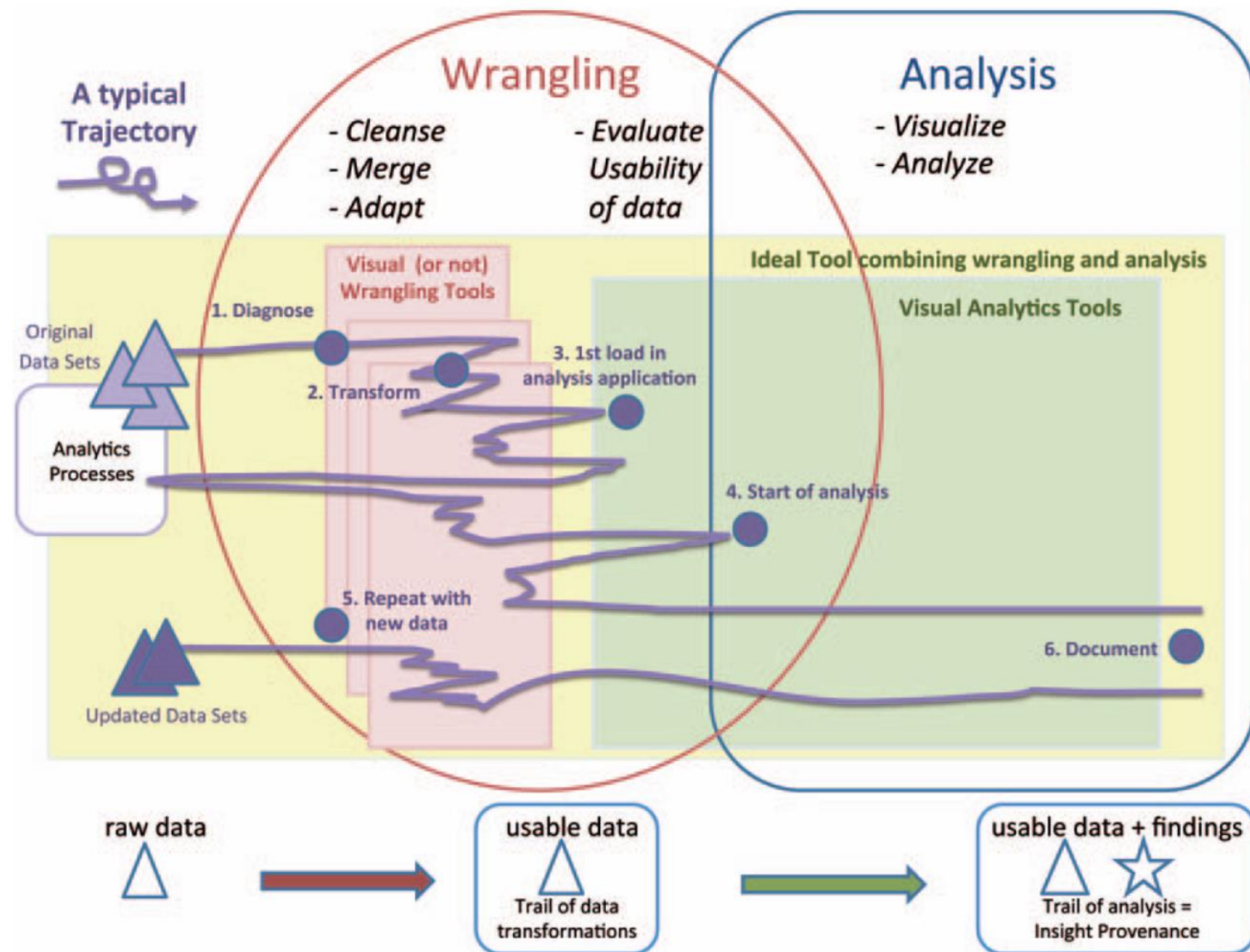
Global Biosphere 2008



<https://earthobservatory.nasa.gov/world-of-change/Biosphere>

Data Wrangling

A “linear” process?

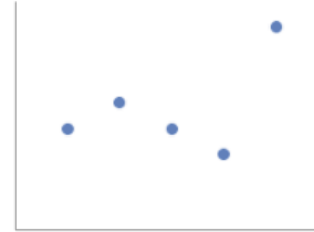


Kandel, S., et al (2011), Research directions in data wrangling: Visualizations and transformations for usable and credible data, *Inf. Vis.*, 10(4), 271–288.

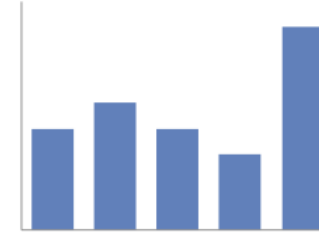
Common Data Visualizations

91%

Simple text



Scatterplot



Vertical bar



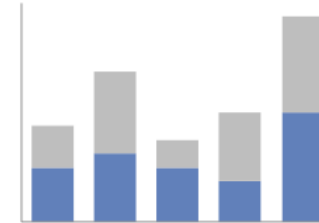
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

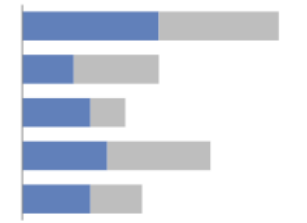
Table



Line



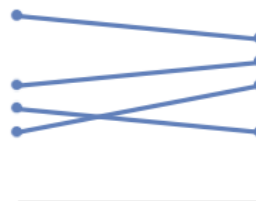
Stacked vertical bar



Stacked horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

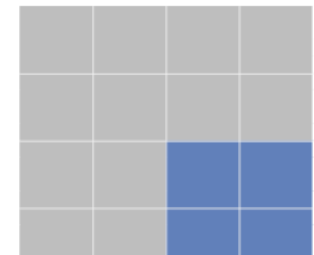
Heatmap



Slopegraph



Waterfall



Square area

<http://www.storytellingwithdata.com/books>

Data Visualization Taxonomy

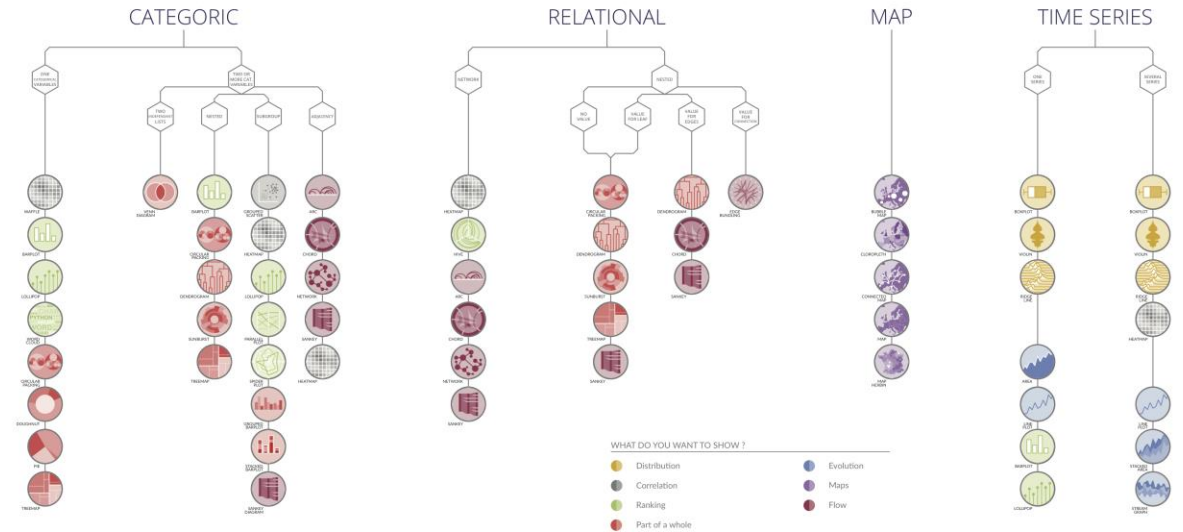


"From Data to Viz" is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps:

- 1 Identify what type of data you have.
- 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
- 3 Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

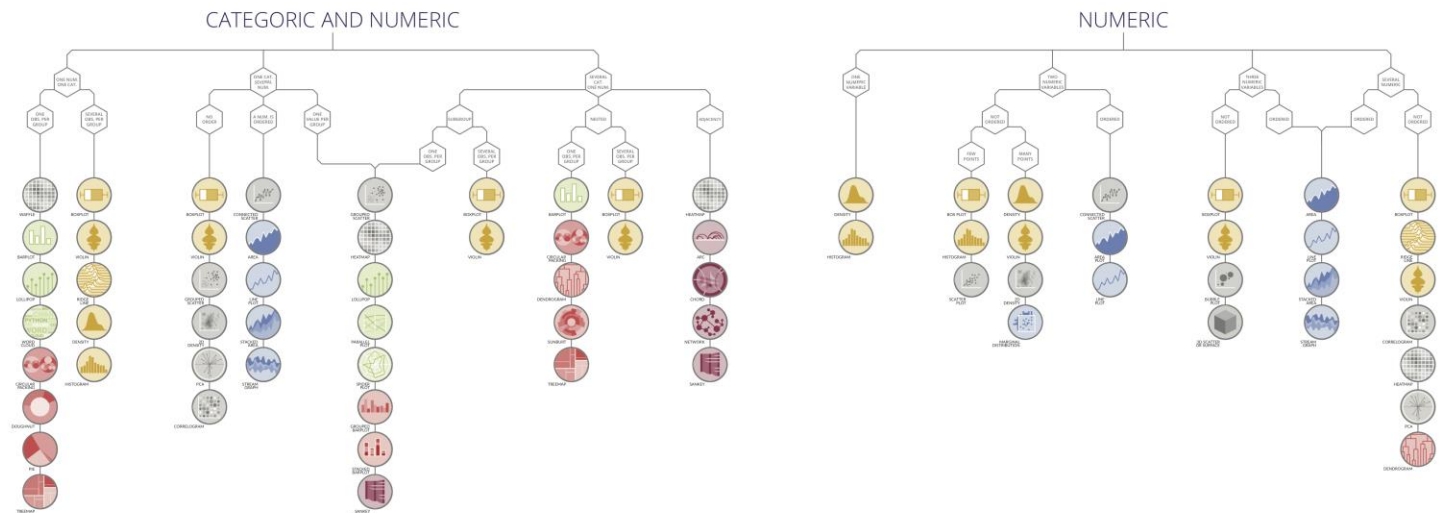


Data to Viz Poster

<https://www.data-to-viz.com>

Python Gallery

<https://python-graph-gallery.com>



2018 © Sam Hickey & Connor Hickey for www.data-to-viz.com

Visual Encodings



Position



Length



Angle/Slope



Area



Volume



Difference



Color hue



Color Saturation



Contrast



Texture

<https://blog.glik.com/visual-encoding>

Visual Encoding & Cognitive Research

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

<https://www.oreilly.com/library/view/designing-data-visualizations/9781449314774/ch04.html>

Pre-attentive Attributes



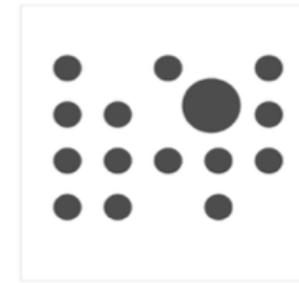
Length



Width



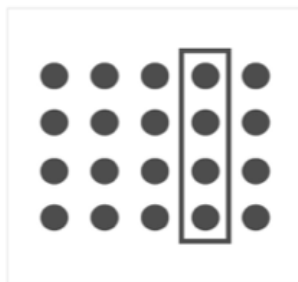
Orientation



Size



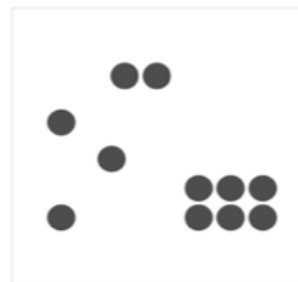
Shape



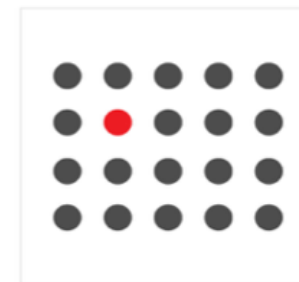
Enclosure



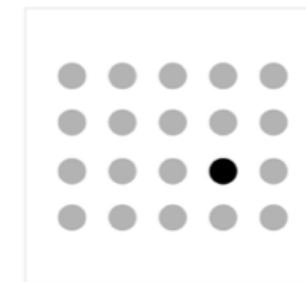
Position



Grouping



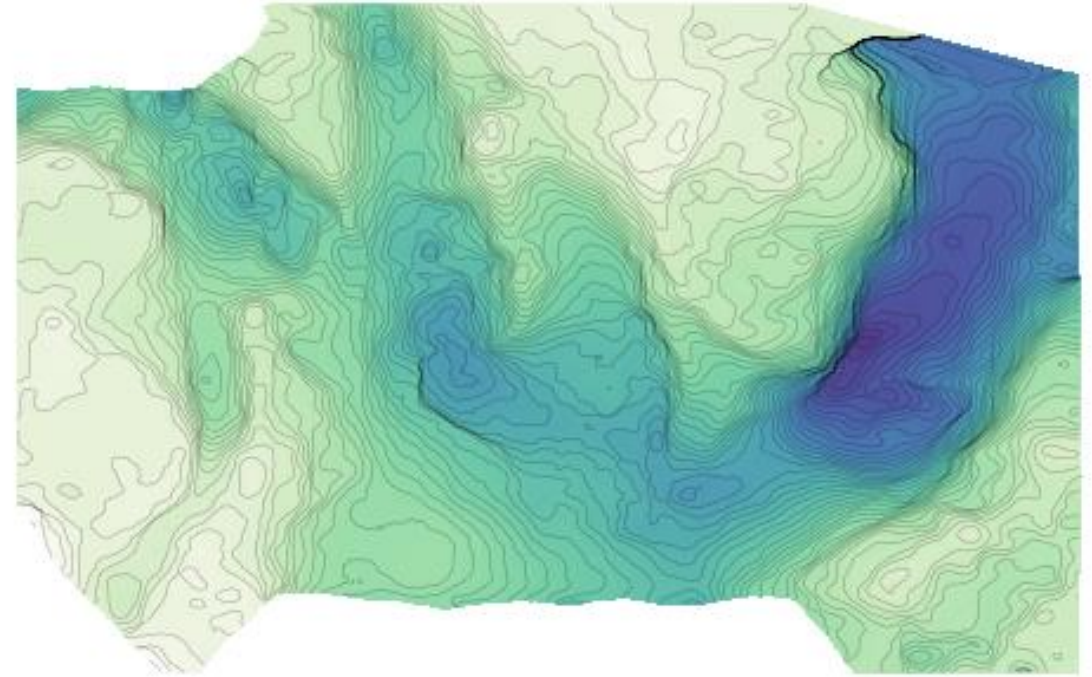
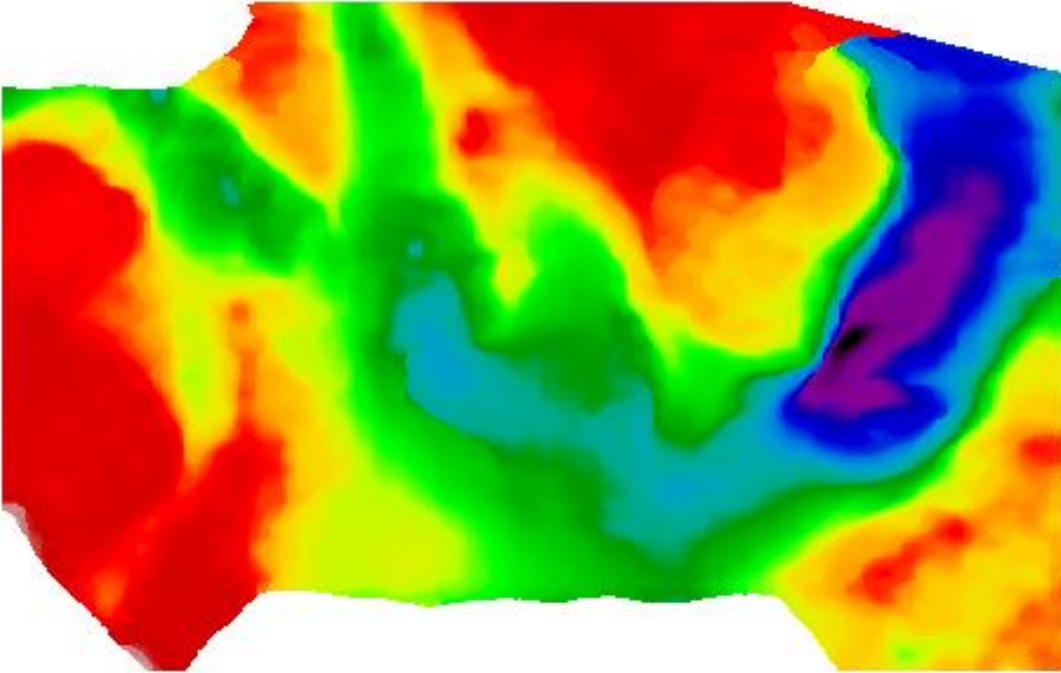
Color Hue



Color Intensity

https://help.tableau.com/current/blueprint/en-us/bp_why_visual_analytics.htm

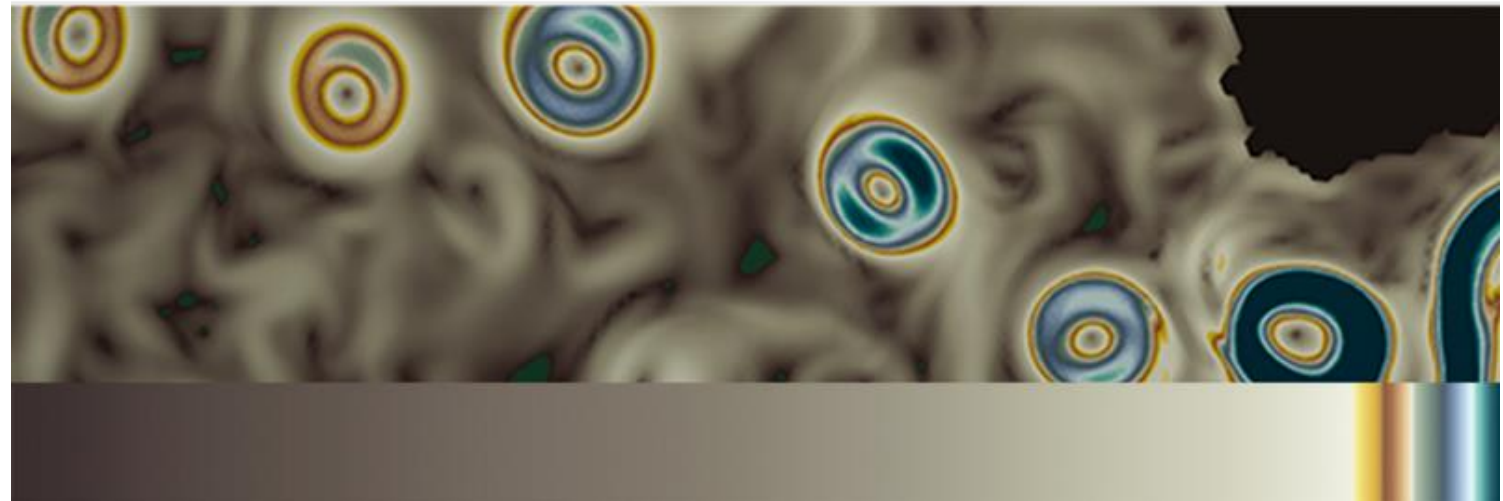
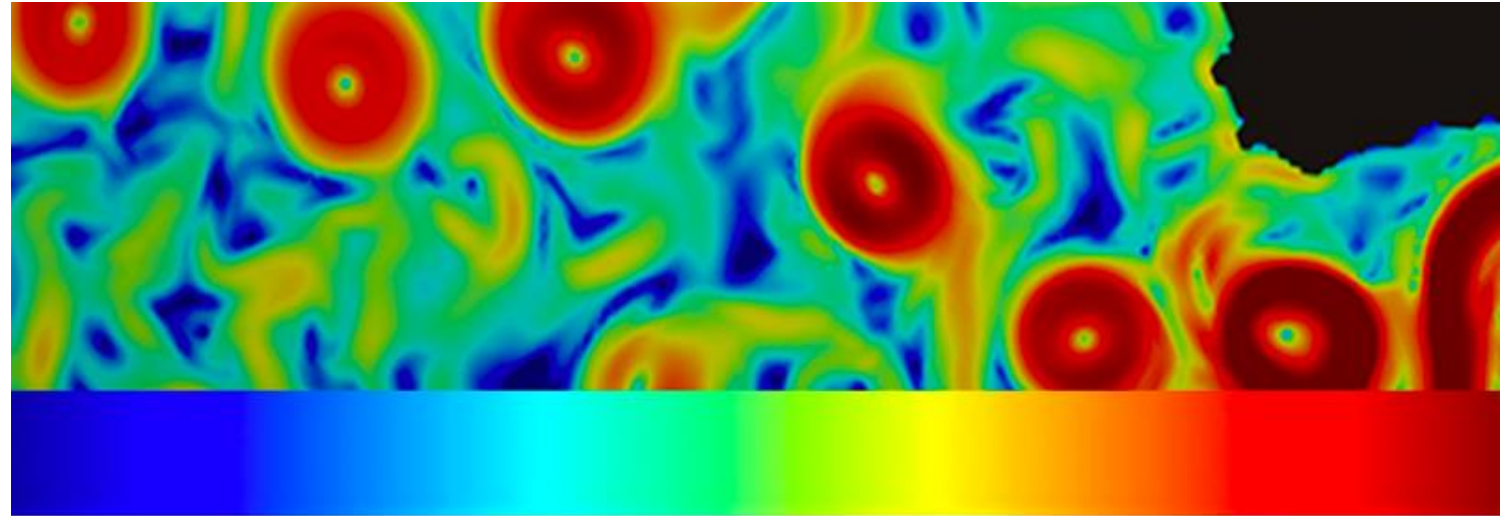
Choose Appropriate Colors & Colormaps



<https://agilescientific.com/blog/2017/12/14/no-more-rainbows>

New Research: Color Scaling

- Feature identification
 - Pinpointing outliers
 - Determining relationships
- Exploration
- Communication

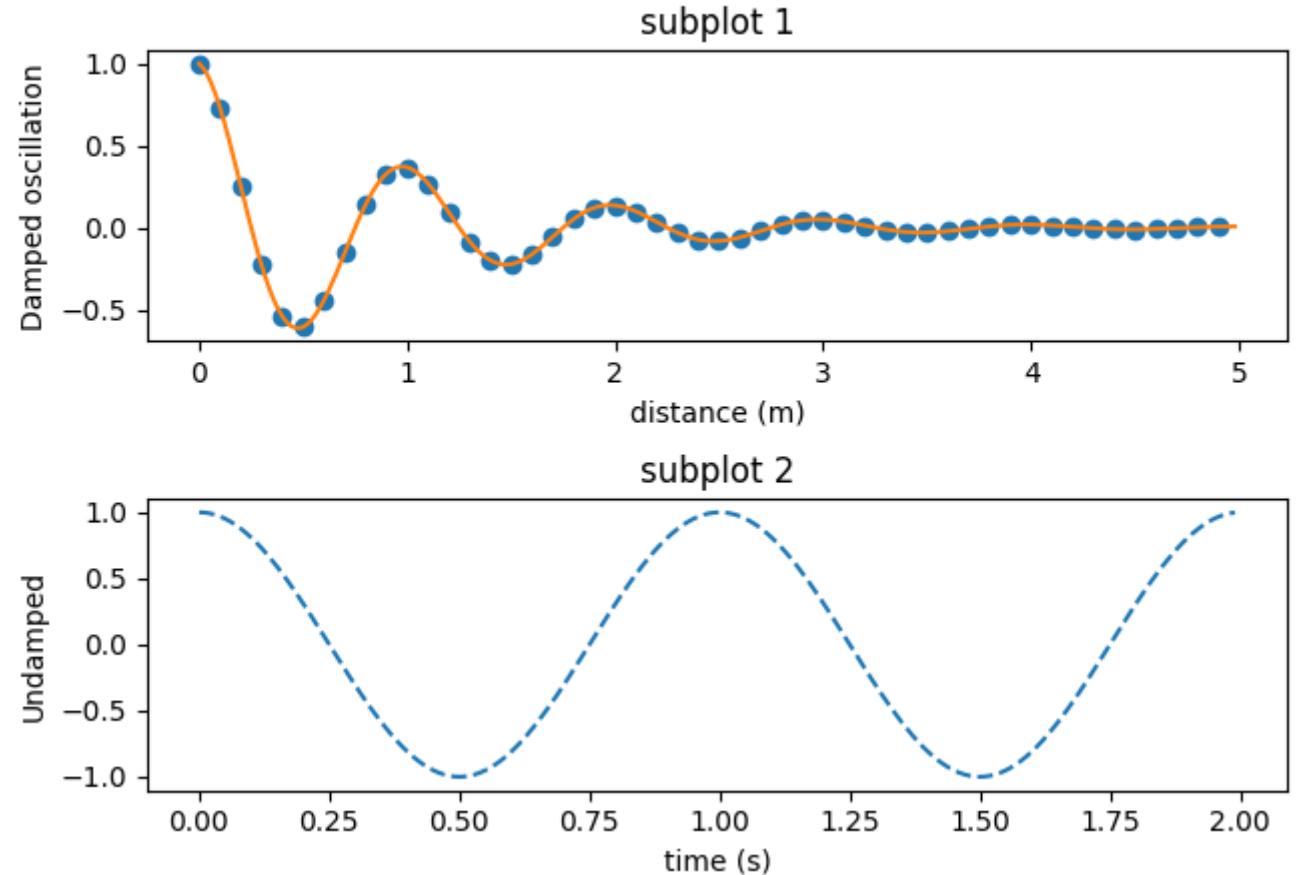


<https://eos.org/features/visualizing-science-how-color-determines-what-we-see>

Key Chart Elements

- Chart Title(s)
- X & Y Labels
- X & Y Tickmarks
 - Often automatic
- Grid lines
- Line/Marker color
- Line/Marker style, width
- Legend
- Subplots

This is a somewhat long figure title



https://matplotlib.org/gallery/subplots_axes_and_figures/figure_title.html#sphx-glr-gallery-subplots-axes-and-figures-figure-title-py

Mini-Project Expectations

- Work in your group of 3
- Choose one or more NDBC stations
 - You can try either “stmet” or “ocean” [station types](#)
- Find some events (e.g. storm), processes (e.g. daily, seasonal, long-term) or other comparisons (e.g. geographic) that interest you
- Create figures to tell a story of what you found
 - If helpful, search the web or literature for more background
- Explain your analysis and conclusions
- Create a ~10 min presentation where each of you can contribute

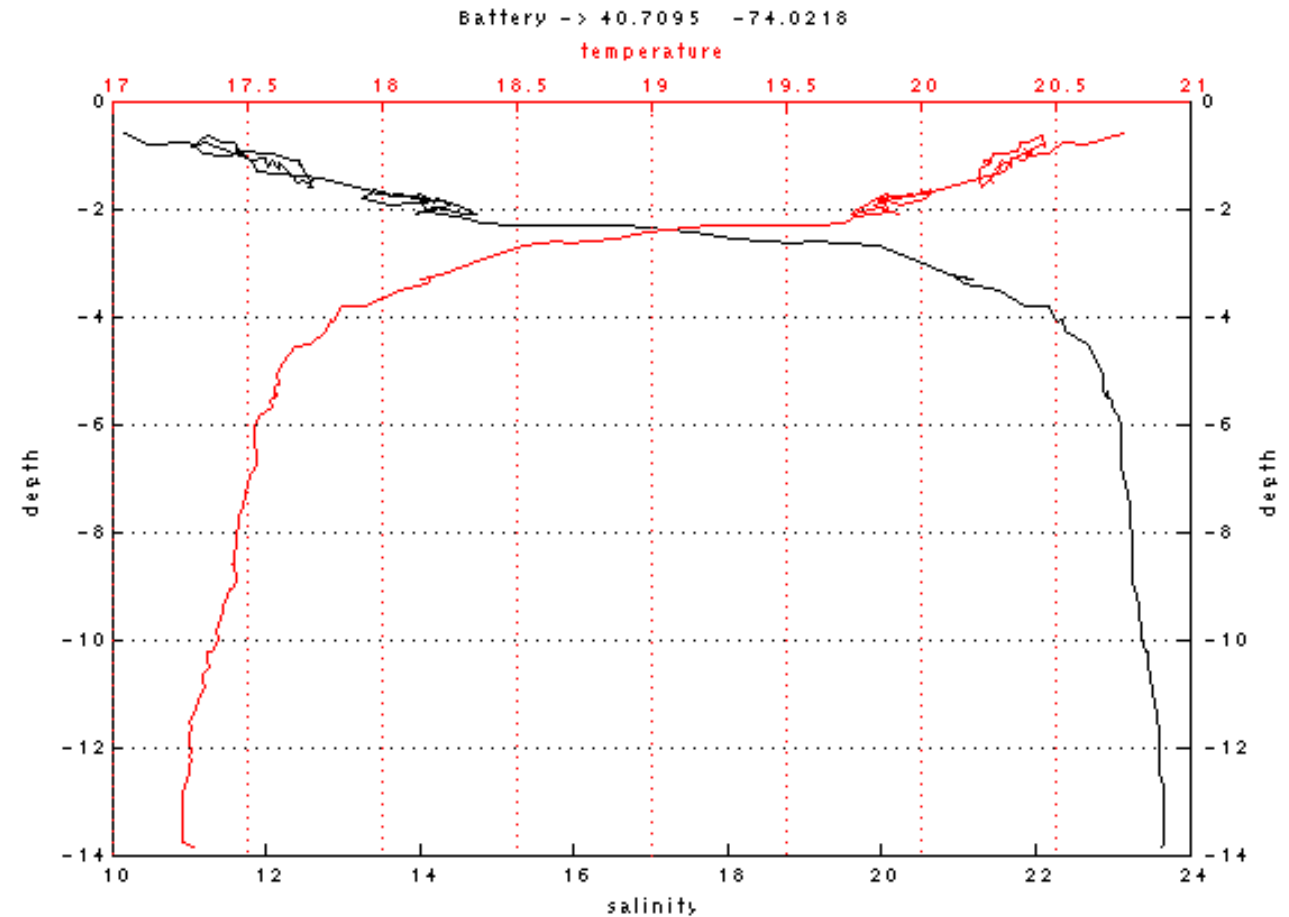
Mid-workshop Group Report

Here are a few topics your group should think about for our report out on Monday to the full team. Feel free to use this slide as a template.

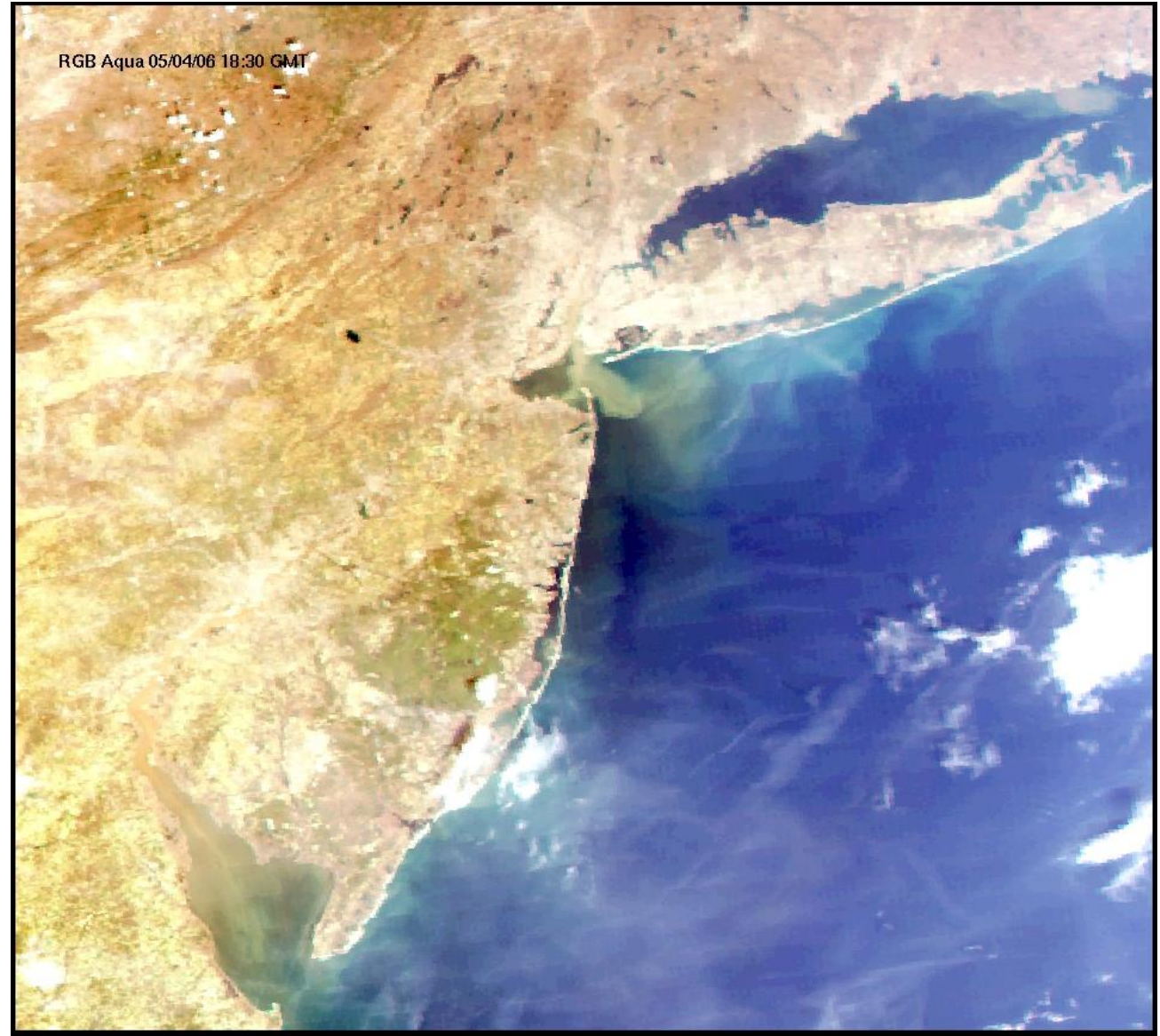
- Group members (name?)
- What is your question? Is it testable or more descriptive?
- What dataset(s) will you use?
- What analysis will you do?
- What is your expectation of what you hope to find after analyzing the data?
- Other things to discuss, but not needed for the report:
 - What are the tasks for each group member?
 - What challenges do you have you may need help with?



Conductivity Temperature Depth (CTD) Profiles



NY Harbor Estuary



CDOM is conservative which allows for water mass discrimination

