

What's Going On in this Graph?

Environmental Data Visualization Literacy Workshop
With R and RStudio

<https://tinyurl.com/EnvDataViz>

Presented by Jajwalya Karajgikar

Workshop Slides, Illustrations, & Materials from [Dr. Allison Horst](#)

Land Acknowledgement

"We recognize and acknowledge that the University of Pennsylvania stands on the Indigenous territory known as "Lenapehoking," the traditional homelands of the Lenape, also called Lenni-Lenape or Delaware Indians. These are the people who, during the 1680s, negotiated with William Penn to facilitate the founding of the colony of Pennsylvania. Their descendants today include the Delaware Tribe and Delaware Nation of Oklahoma; the Nanticoke Lenni-Lenape, Ramapough Lenape, and Powhatan Renape of New Jersey; and the Munsee Delaware of Ontario." [UPenn Association of Native Alumni](#)

Visit the [Native American and Indigenous Studies Program](#) at Penn (Coordinator: Associate Professor of Anthropology, Margaret M. Bruchac)

Follow [Natives at Penn](#) (NAP) and learn about how NAP "creates awareness of Native history and contemporary issues."

Jajwalya Karajgikar

Applied Data Science Librarian

R | Python | Text Data Mining | Natural Language Processing | Machine Learning

@JajRK
jajk@upenn.edu





Q

Research Data and Digital Scholarship

<https://www.library.upenn.edu/about/exhibits-events/earth-week-data-jam>

Facilitating data-driven and data-literate research and scholarship across the disciplines.

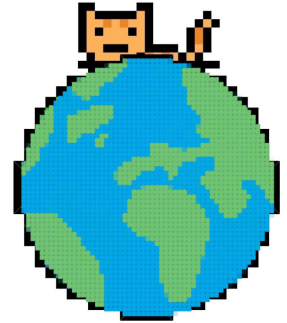
Fostering informed and ethical data communities at Penn.

Interested in data, computational research, digital humanities, or open and public scholarship?

Join the Research Data & Digital Scholarship mailing list to be first in the know about upcoming workshops, talks, and other events!



EARTH WEEK DATA JAM 2022



April 18

Introduction to Data Ethics

Using NVivo to Analyze Perceptions of Environmental Change

Visualizing and Analyzing Renewable Energy Data

April 19

Embodied Information: Local Trees, Datasets, and You

Audubon in Action: Creative Approaches to Data

April 20

What's Going On in This Graph?

Visualizing Botanical Variety in The Woodlands with Mapbox

April 21

Visual Design Tactics for Data and Information Visualization

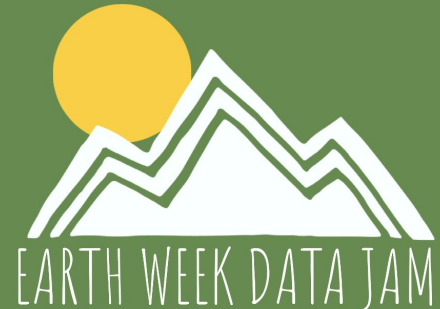
Freeplay: Eco Gaming Showcase

April 22

Mapathon

<https://www.library.upenn.edu/about/exhibits-events/earth-week-data-jam>

RDDS AT PENN LIBRARIES



Learning Outcome

- Rethink the process of Visualizing Data.
- Analyze graphs with the “noticing and wondering” skills.
- Build confidence and acquire new conceptual understanding.
- Critique how different chart design decisions present or obscure uncertainty in research findings.
- Consider diverse viewpoints that may be ignored.

Code of Conduct

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Be graceful with constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Remember!

There are no wrong answers. This is a space to learn.

Let's Play Connect the Dots

Who?

BuzzFeed News' science news desk reporters Peter Aldhous and Zahra Hirji

How?

R scripts to handle the data. Most charts were built with **Datawrapper** using BuzzFeed News' customised design.

What?

"To shift public perception on this issue, Aldhous explained it is vital to ensure such stories focus on the **human cost of the crisis** and **put pressure on those in power**."

Where?

You can use this map to view NASA's historical temperature analysis for any location shown: Click or tap on the map or enter a location into the search box to see the annual chart redraw for that location.

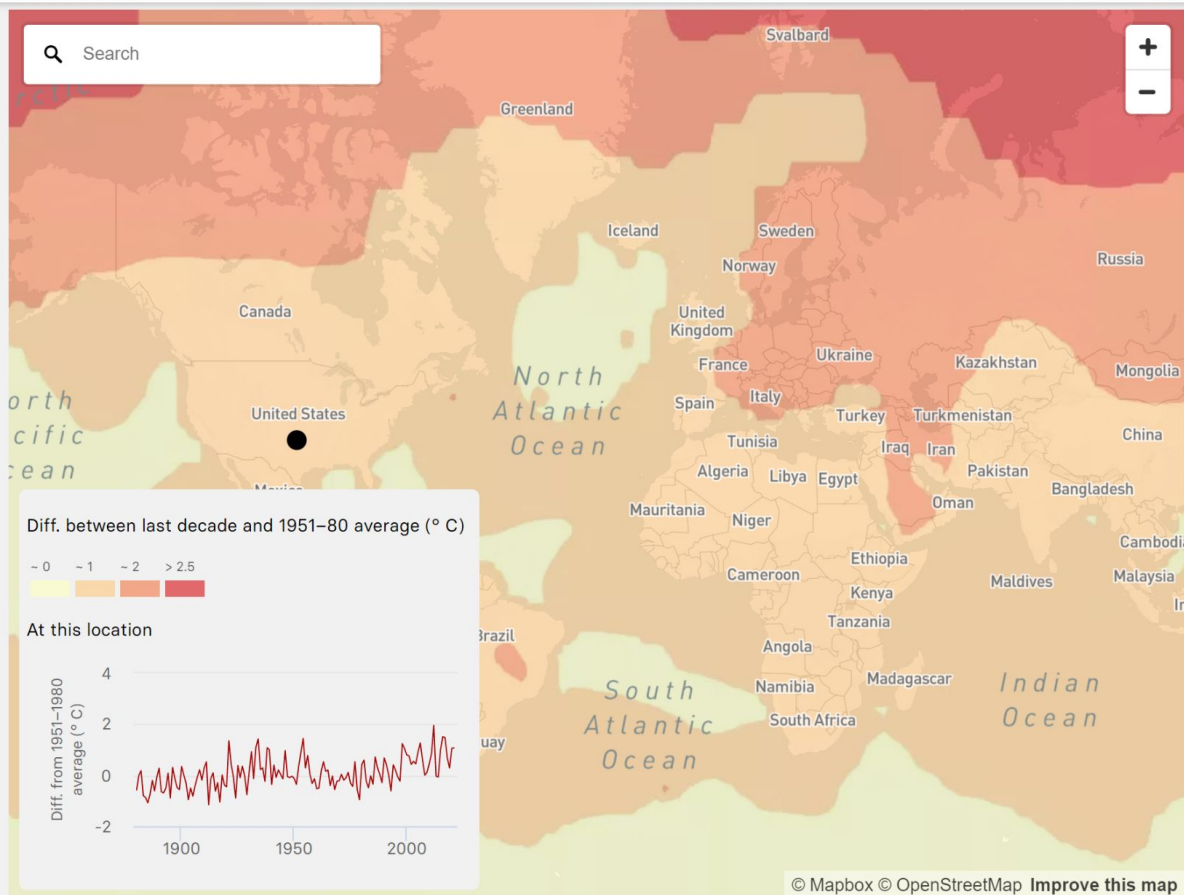
How does the visualization perform?

What do you notice?

What do you wonder?

What do you think is going on in this graph?

What could be improved upon? If anything?



Let's Play Connect the Dots

Who?

BuzzFeed News' science news desk reporters Peter Aldhous and Zahra Hirji

How?

R scripts to handle the data. Most charts were built with **Datawrapper** using BuzzFeed News' customised design.

What?

"To shift public perception on this issue, Aldhous explained it is vital to ensure such stories focus on the **human cost of the crisis** and **put pressure on those in power**."

Where?

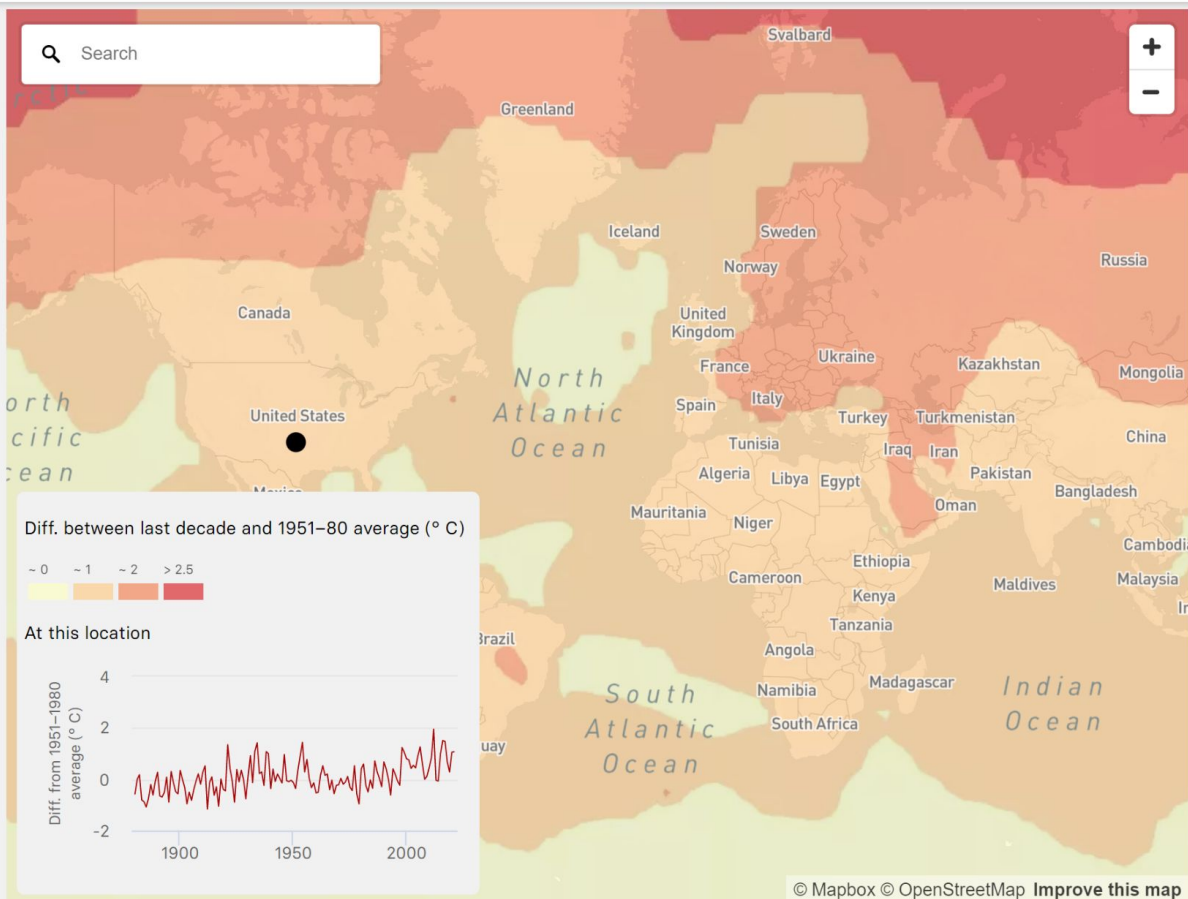
You can use this map to view NASA's historical temperature analysis for any location shown: Click or tap on the map or enter a location into the search box to see the annual chart redraw for that location.

How does the visualization perform?

- Specific details
- "Improve this map"
- Interactivity to engage the reader
- *Read* the visualization
- Clear continuous data labels
- Search for locations

What's going on in this graph? What story can it tell?

What could be improved upon? If anything?

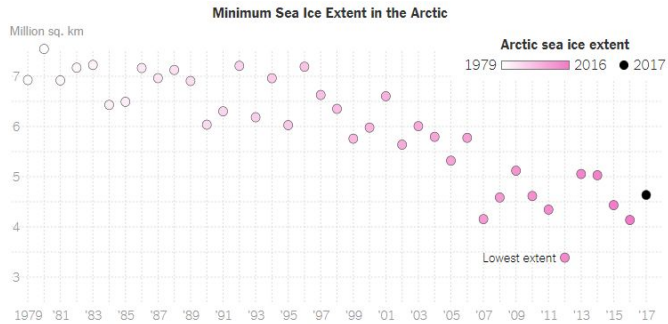


Let's Play Spot the Differences

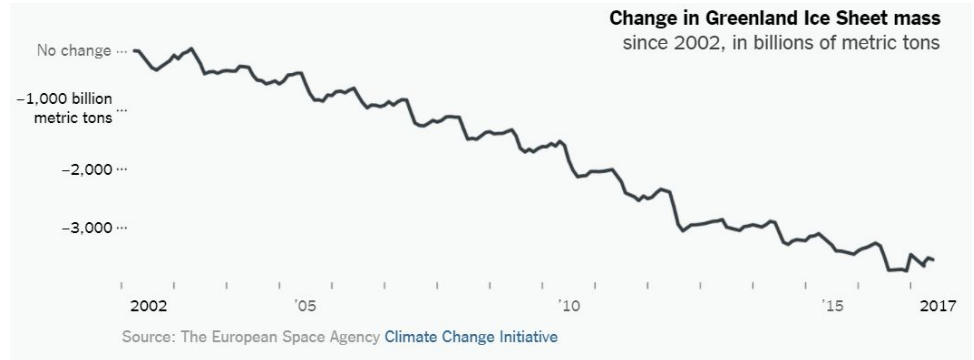
Select the one graph you think is most valuable for teaching the general public about Earth's changing climate. Why?

What are the graphs trying to convey?

Is one graph more effective for the message than the other?



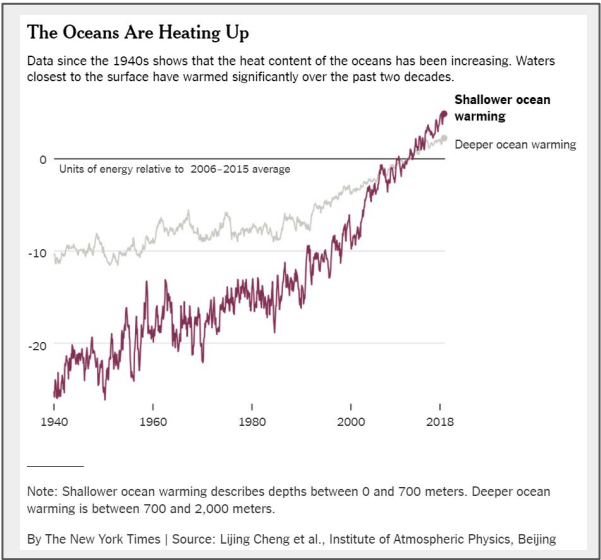
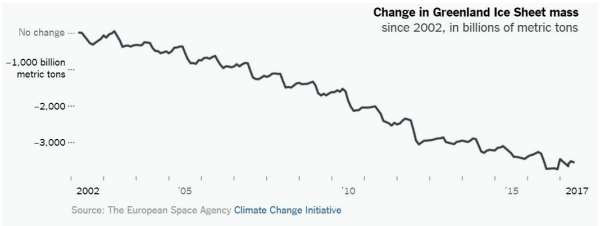
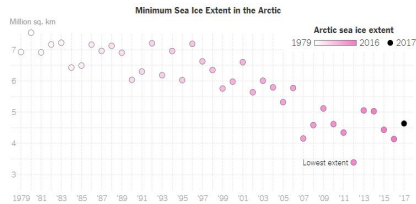
<https://www.nytimes.com/interactive/2017/09/22/climate/arctic-sea-ice-shrinking-trend-watch.html>



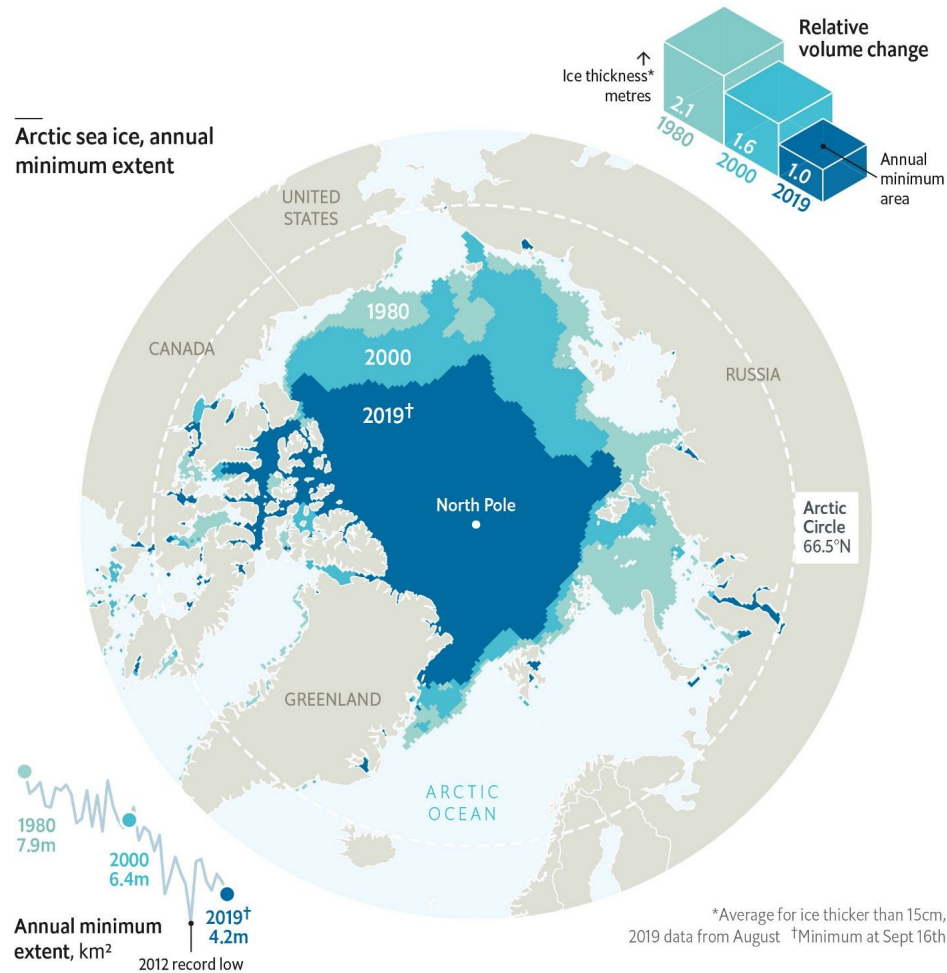
<https://www.nytimes.com/interactive/2017/12/05/climate/greenland-ice-melting.html>

Let's Play Spot the Differences

Select the one graph you think is most valuable for teaching the general public about Earth's changing climate. Why?



Arctic sea ice, annual minimum extent

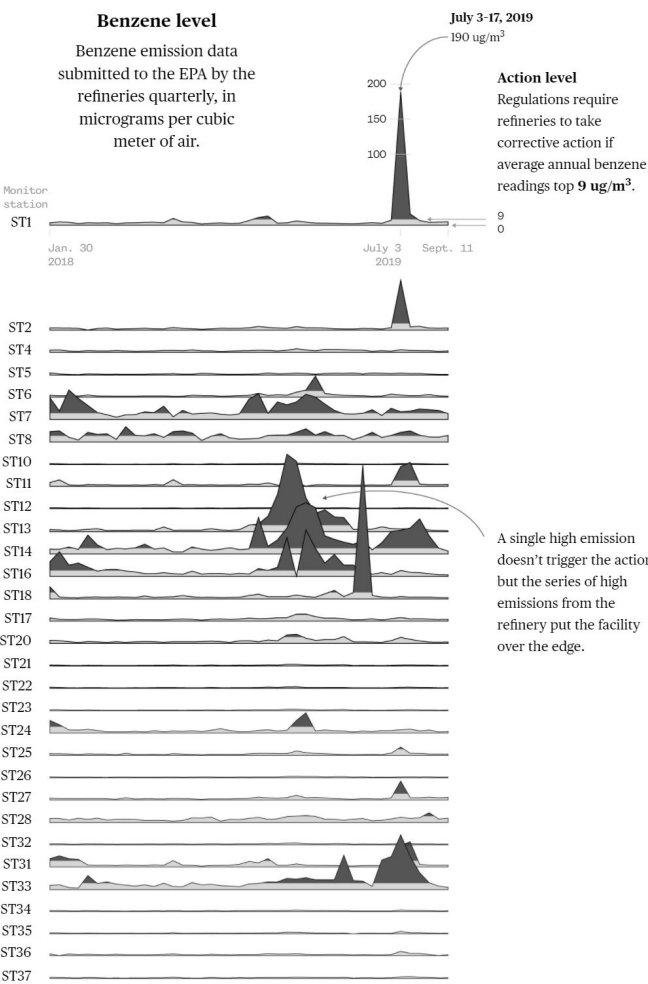
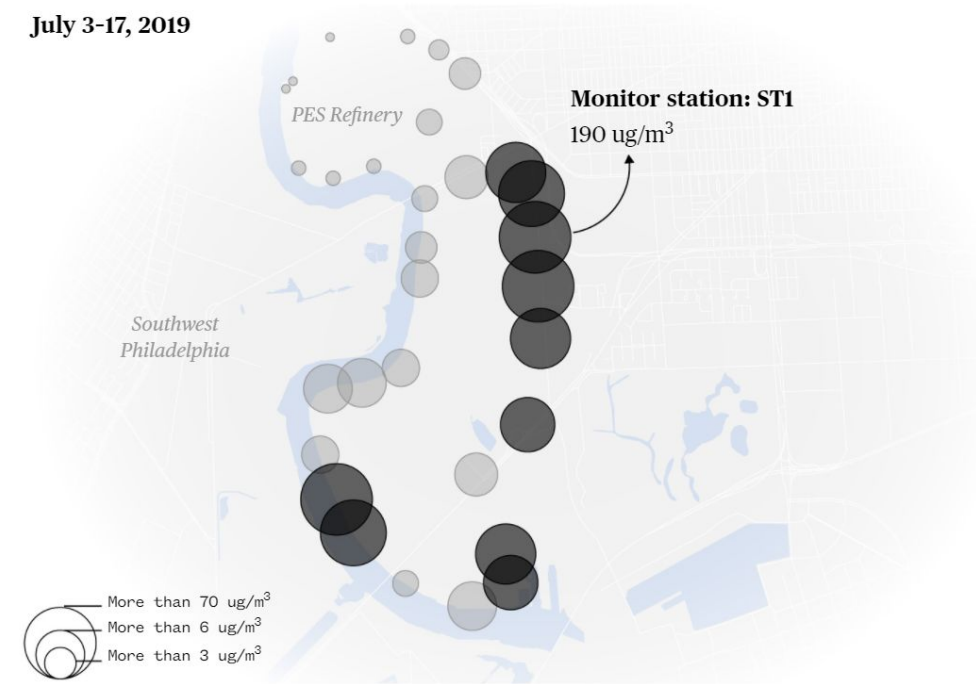


Massive oil refinery leaks toxic chemical in the middle of Philadelphia

Benzene levels at the PES refinery

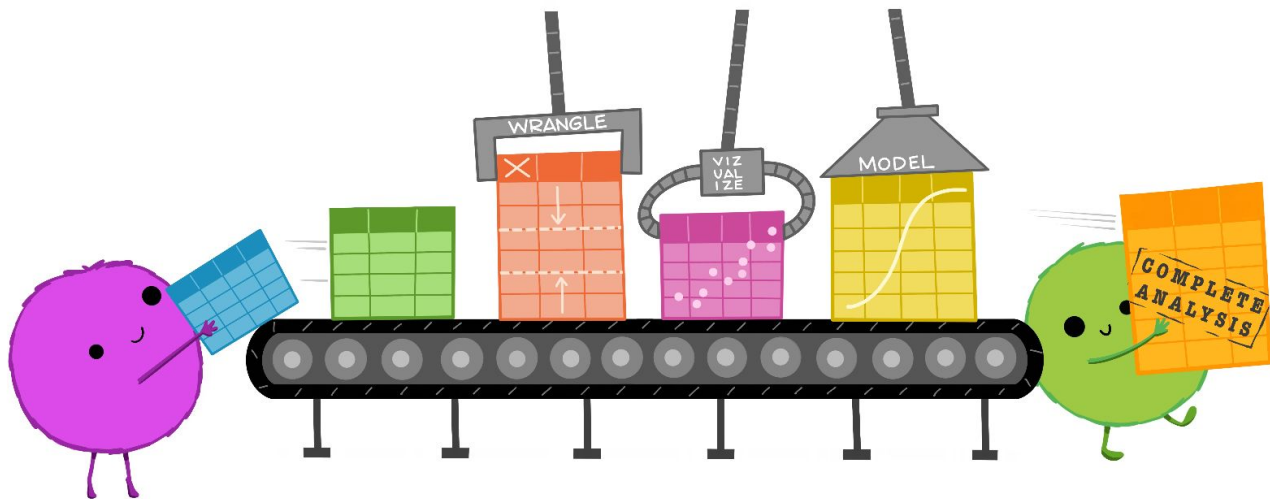
● Over EPA action level ● Under EPA action level

July 3-17, 2019



“Data science is the process of turning data into understanding.”

RStudio Chief Scientist and [R for Data Science](#) co-author [Hadley Wickham](#)



Developing Data Visualization Literacy

Critiquing an infographic or data visual by evaluating its effectiveness is an exercise in developing data literacy

■ **Consider the graphic**
What is the visualization trying to say?
Is the messaging effective?

■ **Understand the source**
Is the data source traceable?
Is the context and background of the subject matter clear?

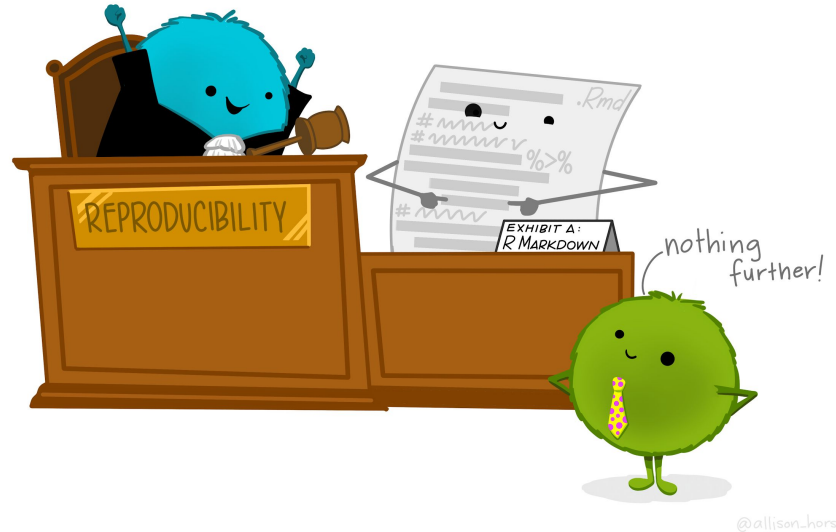
■ **Recognize the functionality**
Are the axes starting from the minimum?
Are they at equal intervals and ranges?
Is the design indicative of the subject matter?

REPRODUCIBILITY

■ **Convert to other forms to verify & validate**
Is the chart type supporting the message in the graph?

“U.S. National Science Foundation (NSF) subcommittee on replicability in science: *reproducibility* refers to **the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator.**”

[Goodman et al., Science Translational Medicine 01 Jun 2016: Vol. 8, Issue 341, pp. 341ps12](#)
[DOI: 10.1126/scitranslmed.aaf5027](#)

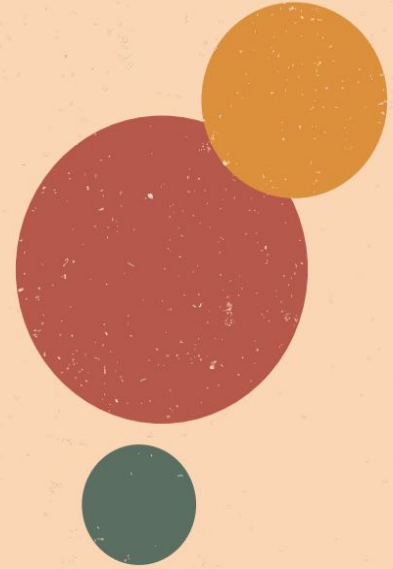


Why (and how) you should include public lands, tourism and Indigenous issues in your climate change coverage

CLIMATE CHANGE FOR EVERYONE

- Climate change is one of the most important topics we cover -- it permeates everything
- Human-land relationship mindset
- How are you covering climate change already?
 - Focusing on land, add humans, vice versa
- For BOTH urban and rural reporters
- Ask where! Place is so important
- Water is life.
- People don't "believe" in climate change, they just don't *understand* it.

K. Sophie Will
Senior National Parks Reporter, Editor
The Spectrum/USA Today/Report for America



Considering and Including Indigenous perspectives

From <https://www.ksophiewill.com/>

Indigenous Data Sovereignty: Implications for Data Journalism

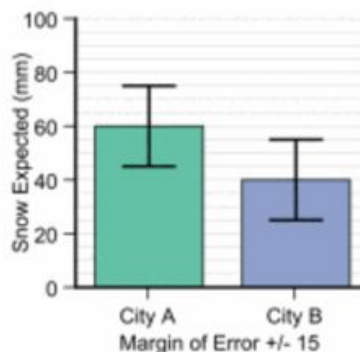


For each element and design decision, ask:

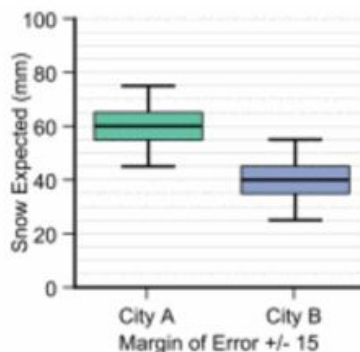
- Is this element necessary/helpful for **audience understanding**?
- Is it presented in the **best way** to encourage audience understanding & retention of the things that I have decided are the most critical?

Presenting uncertainty

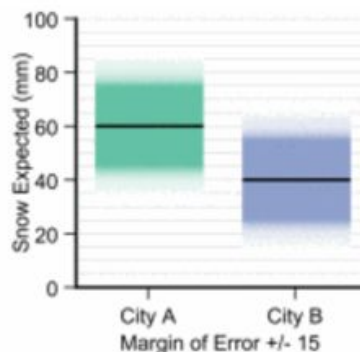
Common techniques include error bars, gradients and ranges, as well as showing individual data points (dot plots) (Correll and Gleicher 2014).



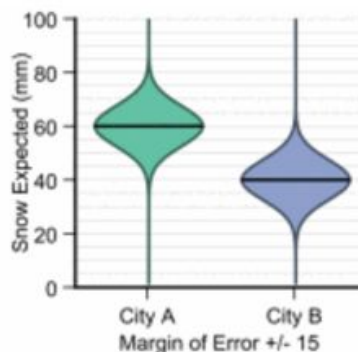
(a) **Bar chart** with error bars: the height of the bars encodes the sample mean, and the whiskers encode a 95% t-confidence interval.



(b) **Modified box plot**: The whiskers are the 95% t-confidence interval, the box is a 50% t-confidence interval.

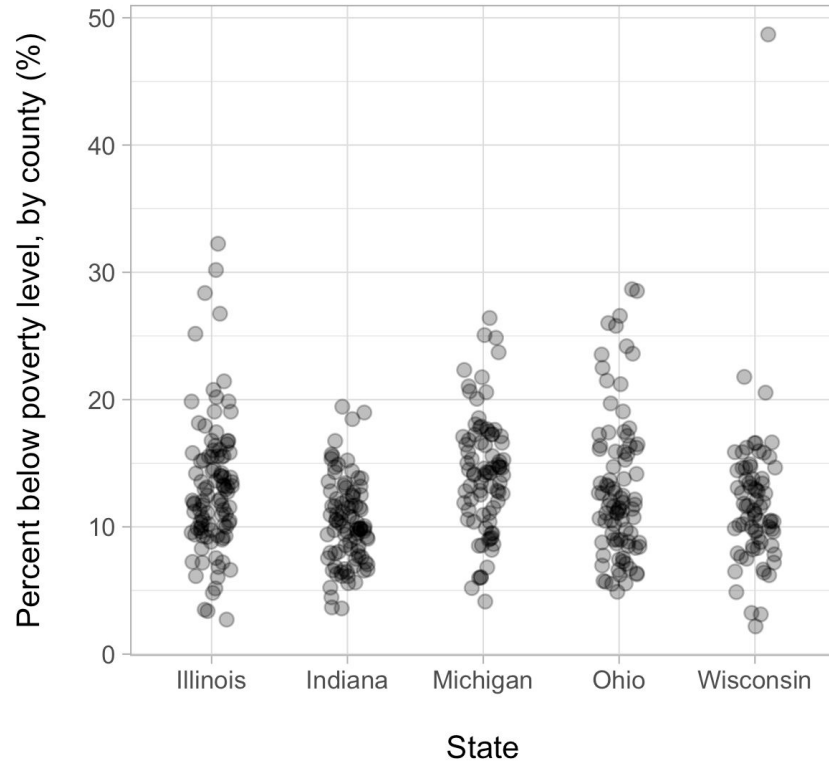
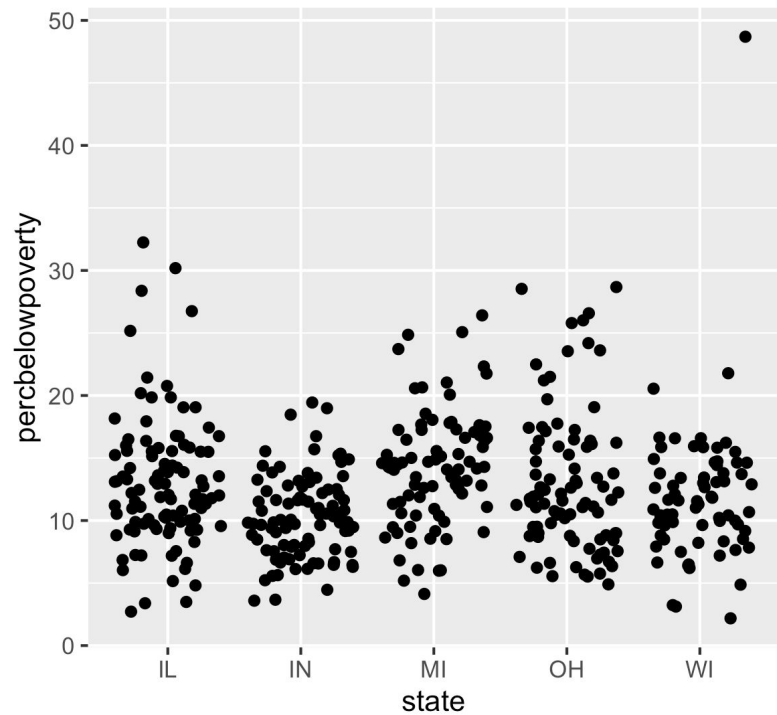


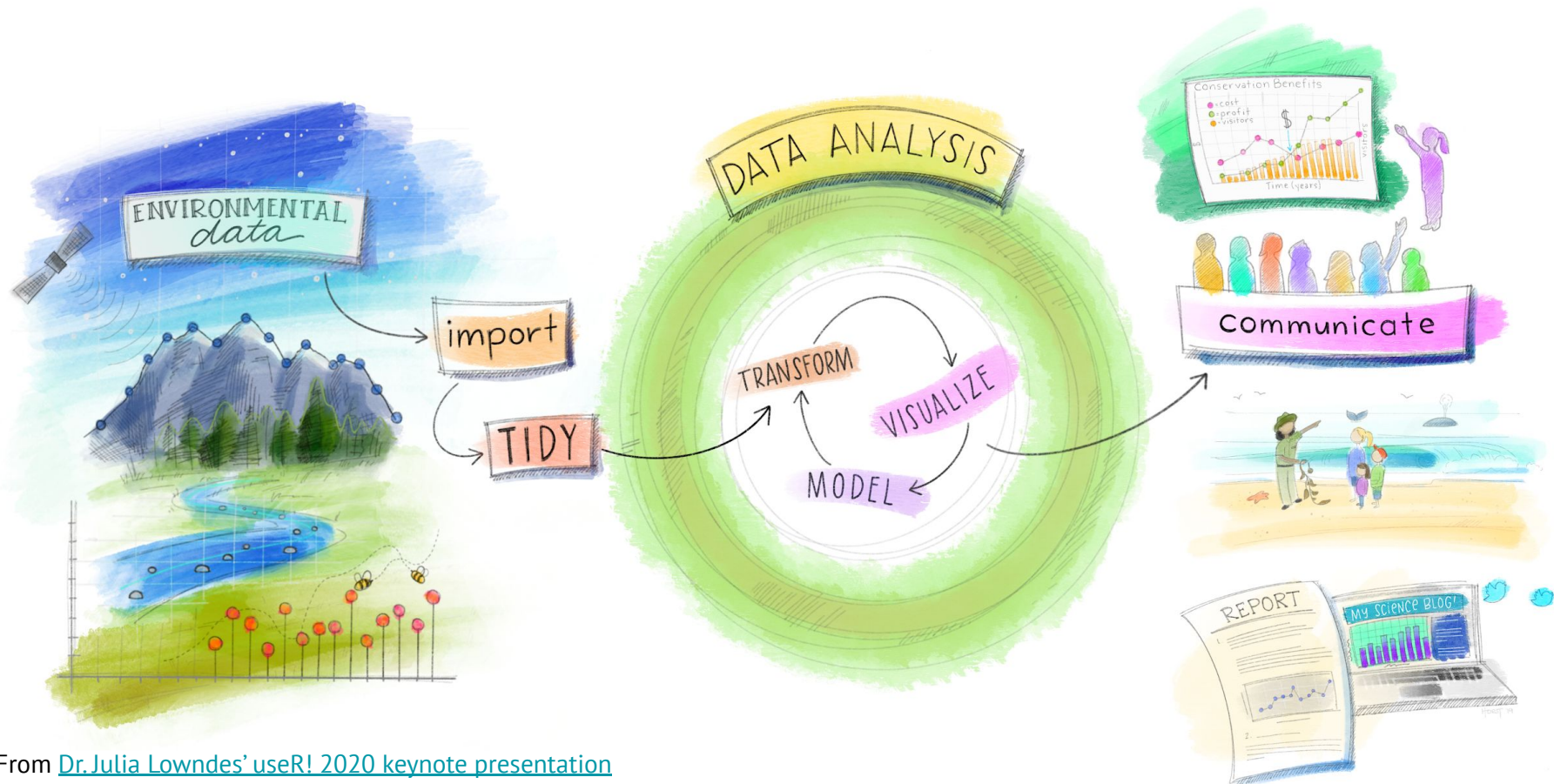
(c) **Gradient plot**: the transparency of the colored region corresponds to the cumulative density function of a t-distribution.



(d) **Violin plot**: the width of the colored region corresponds to the probability density function of a t-distribution.

OR





From [Dr. Julia Lowndes' useR! 2020 keynote presentation](#)

Artwork: [@allison_horst](#)



“The tidyverse is a coherent system of packages for data manipulation, exploration and visualization that share a common design philosophy.”

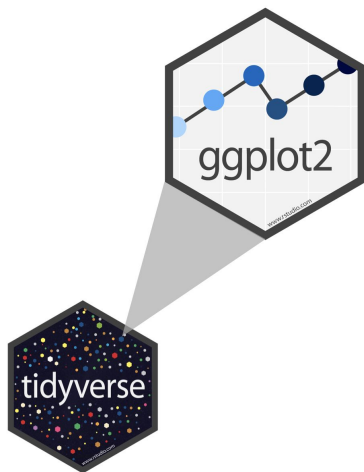
- Joseph Rickert
[What is the tidyverse?](#)
RStudio Community Blog



ggplot2:

Build a data MASTERPIECE

The Grammar of Graphics is a formal guideline which describes the dependencies between all elements of a statistical graphic.



Environmental Data Literacy Tools:

- [New Data Tools and Tips for Investigating Climate Change](#) by Global Investigative Journalism Network
- [Environmental Resources by the Society of Professional Journalists](#)
- [Toolbox for Teaching Climate & Energy](#)
- [Data Literacy](#) CEEDAR
- [The Data Literacy Project](#)
- Culturally Relevant Education in Environmental Data Science

Readings:

- Bahlai, C et al. (2019). Open science isn't always open to all scientists. American Scientist 107 (2): 786
- Ch 14 in Indigenous Data Sovereignty, Building a data revolution in Indian Country by Dr. Desi Rodriguez-Lonebear
- Cheruvellil, KS and PA Soranno (2018). Data-intensive ecological research is catalyzed by open science and team science. BioScience 68 (10): 813 - 822
- Hampton et al. (2015). The Tao of open science for ecology. Ecosphere 6 (7): 1 - 13C
- Lowndes et al. (2017): Our path to better science in less time using open data science tools
- Mah, Alice. (2016) Environmental justice in the age of big data : challenging toxic blind spots of voice, speed, and expertise. Environmental Sociology.doi: 10.1080/23251042.2016.1220849
- Martha C. Monroe, Richard R. Plate, Annie Oxarart, Alison Bowers & Willandia A. Chaves (2019) Identifying effective climate change education strategies: a systematic review of the research, Environmental Education Research, 25:6, 791-812, DOI: 10.1080/13504622.2017.1360842
- Nyman, M., Ellwein, A. L., Daniel, M., and Connealy, S., Using Data-Rich Instruction for Climate Change Education: Road Blocks and Pathways, vol. 2011, 2011.
- [The Next Generation of Environmental Scientists are Data Scientists by Jenny Seifert and Kathryn Meyer](#)
- Wilke, C. (2019). Fundamentals of data visualization: A primer on making informative and compelling figures.
- Wilson et al. (2017): Good enough practices in scientific computing

Examples of Data Visualization Literacy:

- Gonchar, M. (2019, February 28). *Teach about climate change with these 24 New York Times graphs*. The New York Times. Retrieved April 8, 2022, from <https://www.nytimes.com/2019/02/28/learning/teach-about-climate-change-with-these-24-new-york-times-graphs.html>
- [Climate Analytics](#)

People:

- [Environmental data science study group at UCSB](#)
- #rstats,
- #tidytuesday
- @allison_horst
- @dataandme
- @drob
- @hadleywickham
- @WeAreRLadies
- @jennybryan
- @ROpenSci
- @juliasilge
- @apreshill
- R-Ladies
- R-Users groups
- R-Philly
- Mine Cetinkaya-Rundel
- Tiffany Timbers
- Julia Lowndes
- Jenny Bryan
- Alison Hill
- Greg Wilson
- Kelly Bodwin

Helen Fillmore
Charitie Ropati
Sumak Helena
Dallas Goldtooth
Winona Laduke
Lydia Jennings
Nina Rose Berglund
Tara Houska
Xiuhtezcatl Martinez
Quannah Chasinghorse

