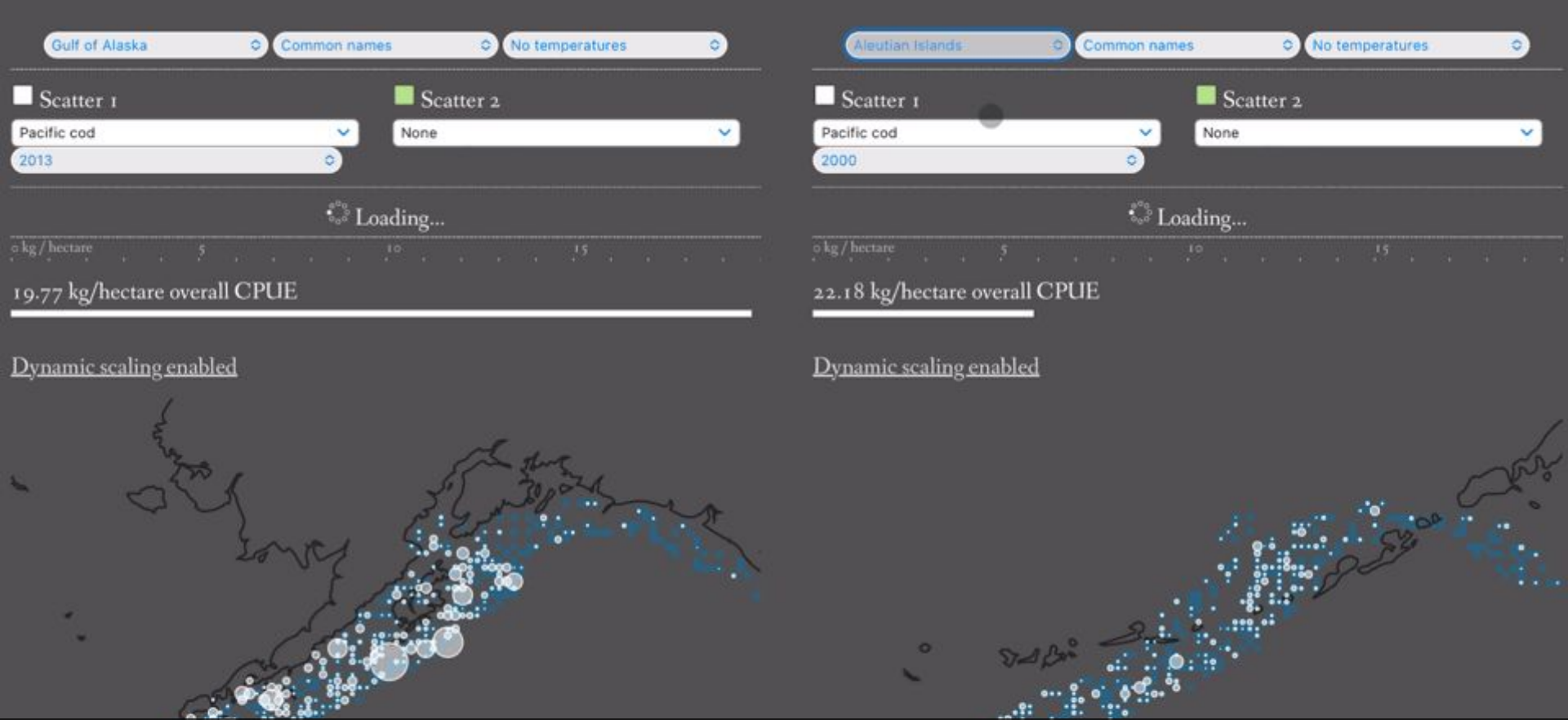
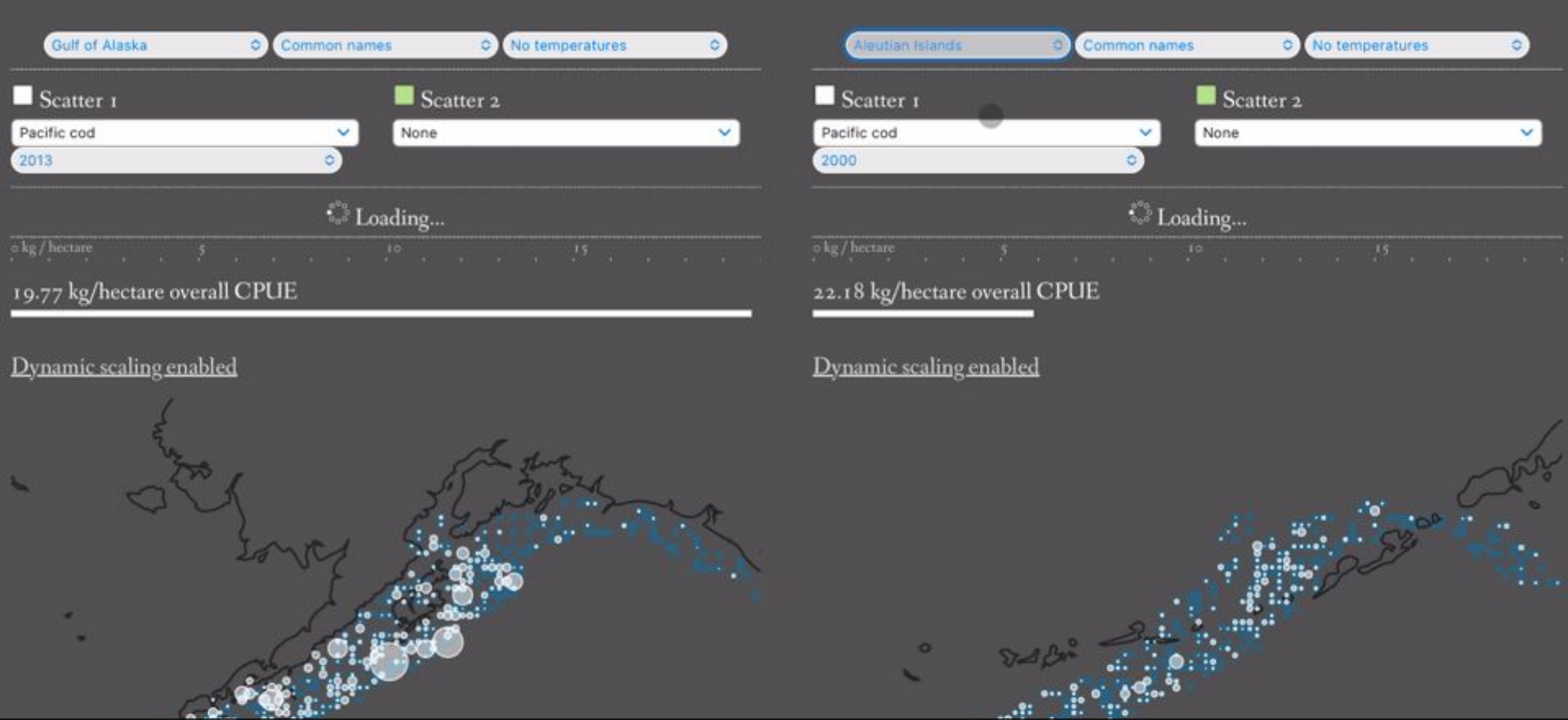


Hello!

Please find a seat. We will start at about 2:05 - 2:10



Data Visualization

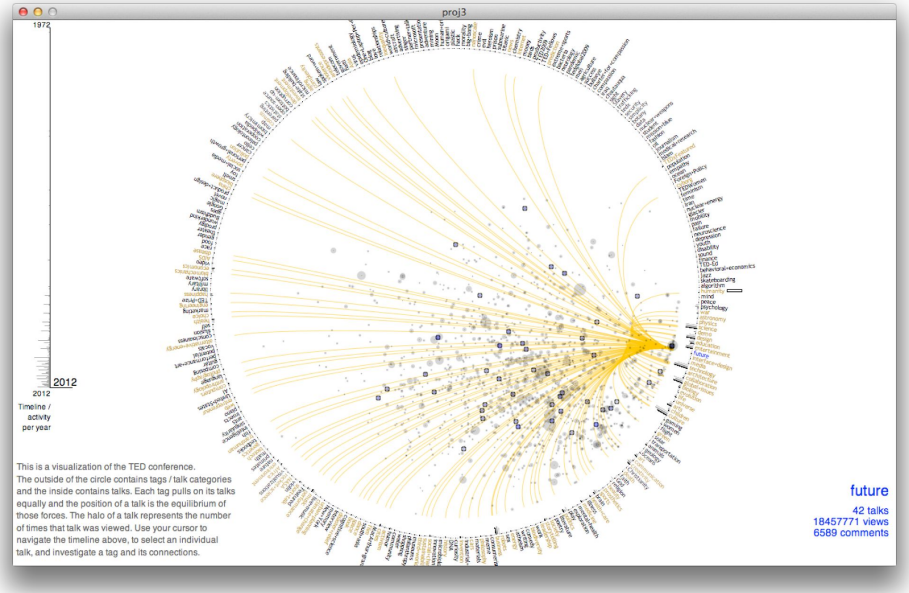
A Samuel Pottinger
Sept 15, 2025
Stat 159
Collaborative and Reproducible Data Science

Hello! I'm a data scientist, software engineer, and information designer.

Sam Pottinger
A more human-centered AI/ML
https://gleap.org

UC Berkeley | Data + Environment
EVERY | Data + Synthetic Biology
IDEO | Data + Design
Plenty | Data + Indoor Agriculture
Apple | Data + Engineering
Google | Data + Visualization
LabJack | Data + Hardware

Processing | Data + Love in Java
Sketchingpy | Data + Love in Python



Today

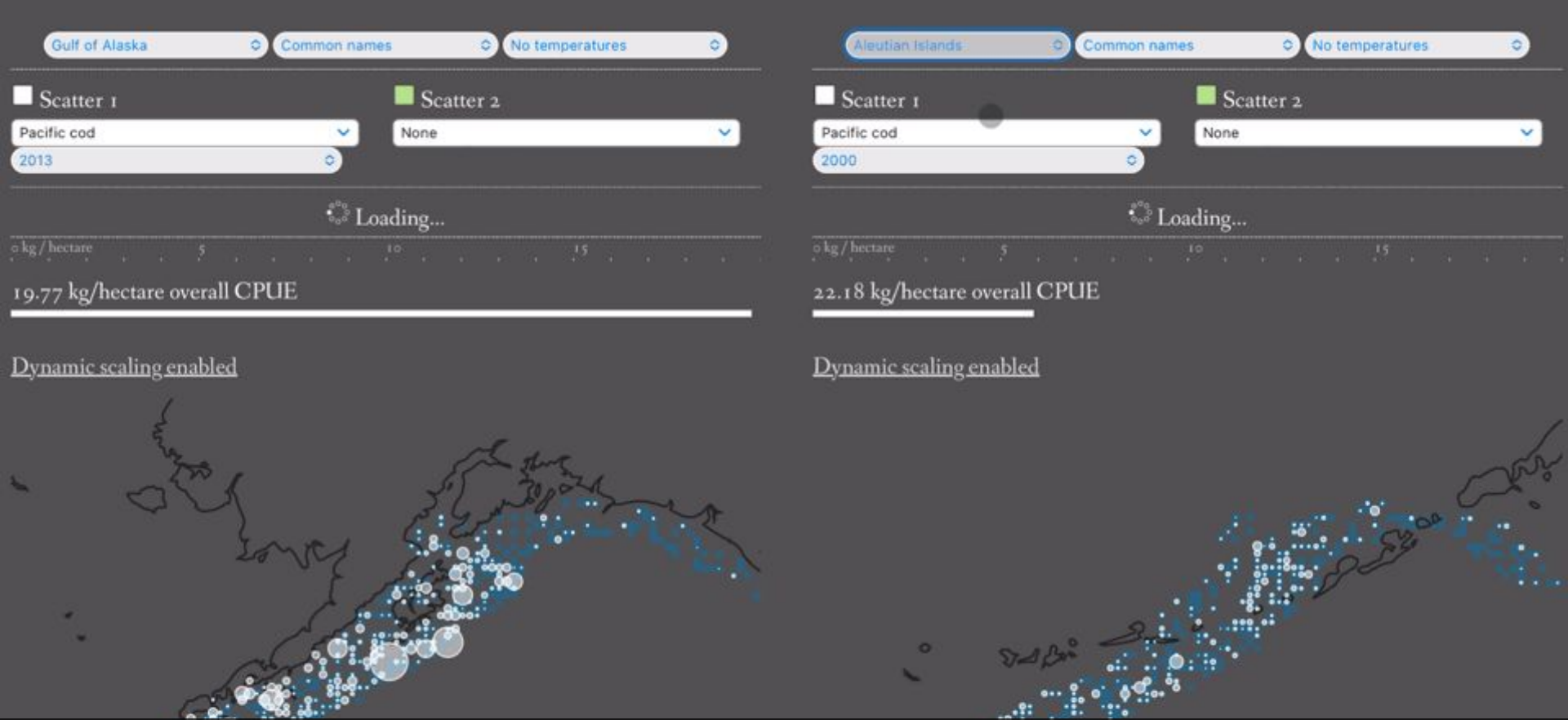
Why care about visualization

How to design data visualizations

How to think about data visualization

How to implement visualizations

How to continue your journey



Why care about visualization?

Year	Number of Wolves	Number of Moose
1980	50	664
1982	14	700
1984	24	811
1986	20	1025
1988	12	1653
1990	15	1216
1992	12	1600
1994	15	1800
1996	22	1200
1998	14	700
2000	29	850
2002	17	1000
2004	29	750
2006	30	385
2008	23	650
2010	19	510
2012	9	750
2014	9	1050
2016	2	1300
2018	2	1500

Premise: The human visual system is good at spotting patterns.

What year saw the most moose?

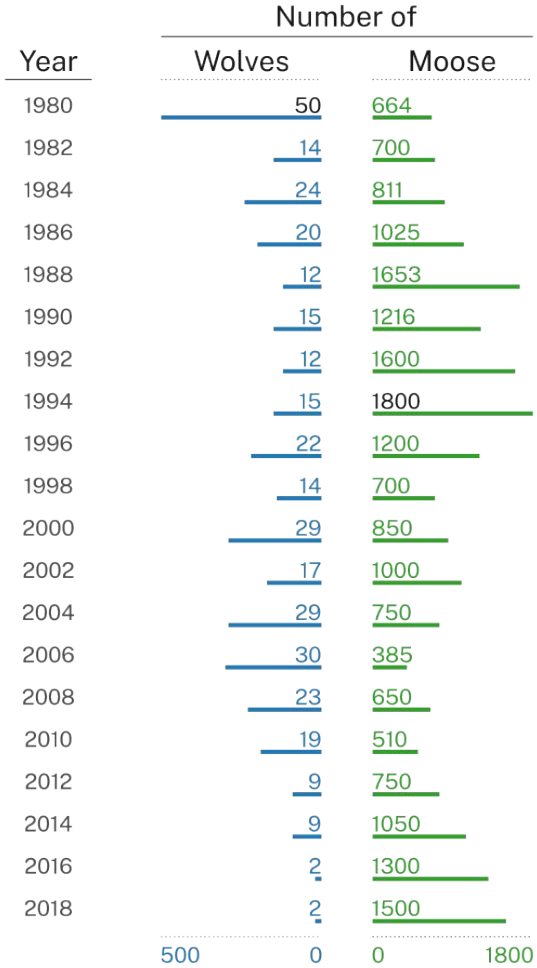
🎮 Raise your hand when you have it.

Year	Number of Wolves	Number of Moose
1980	50	664
1982	14	700
1984	24	811
1986	20	1025
1988	12	1653
1990	15	1216
1992	12	1600
1994	15	1800
1996	22	1200
1998	14	700
2000	29	850
2002	17	1000
2004	29	750
2006	30	385
2008	23	650
2010	19	510
2012	9	750
2014	9	1050
2016	2	1300
2018	2	1500

Premise: The human visual system is good at spotting patterns.

What year saw the most wolves?

🎮 Raise your hand when you have it.

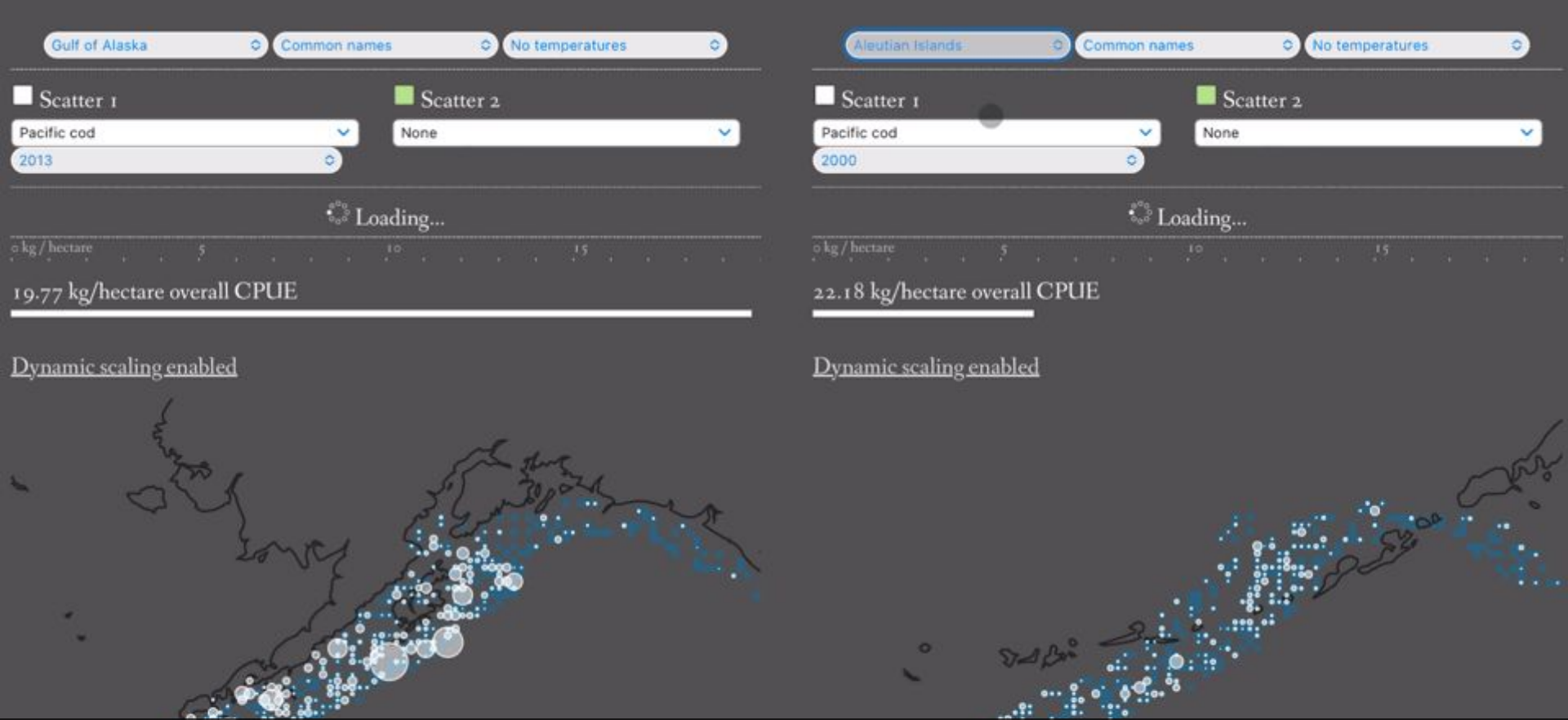


Year	Number of Wolves	Number of Moose
1980	50	664
1982	14	700
1984	24	811
1986	20	1025
1988	12	1653
1990	15	1216
1992	12	1600
1994	15	1800
1996	22	1200
1998	14	700
2000	29	850
2002	17	1000
2004	29	750
2006	30	385
2008	23	650
2010	19	510
2012	9	750
2014	9	1050
2016	2	1300
2018	2	1500
	500 0	0 1800

Premise: The human visual system is good at spotting patterns.

What is the relationship between wolves and moose in Isle Royale?

🎮 Raise your hand when you have it.



How to design visualizations

Cleveland and McGill

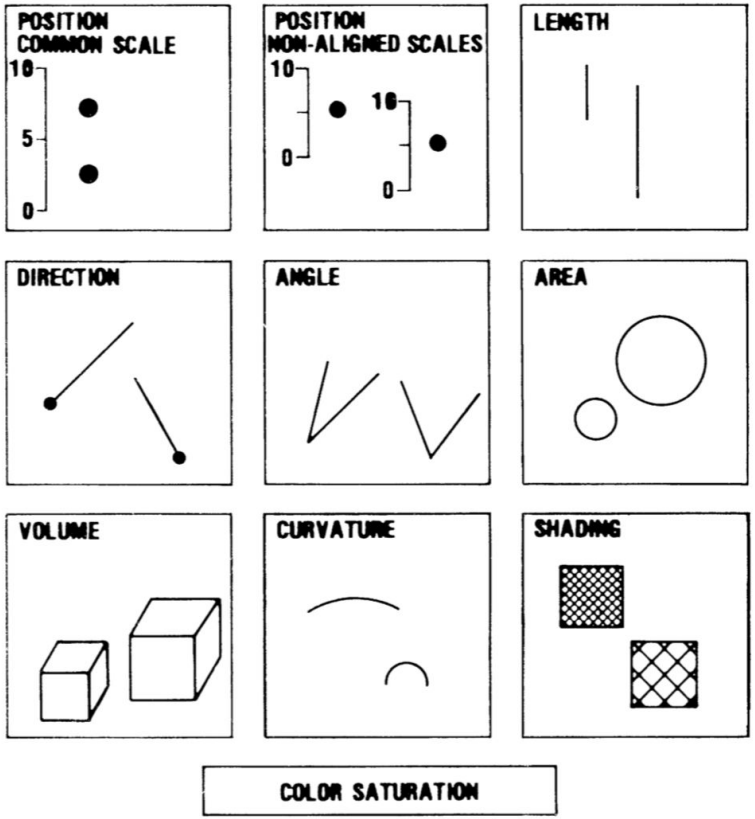


Figure 1. Elementary perceptual tasks.

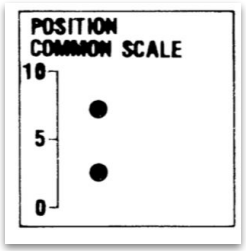
Evolution gives us preattentive features.

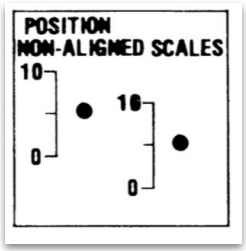
However, what is the right design of using those features?

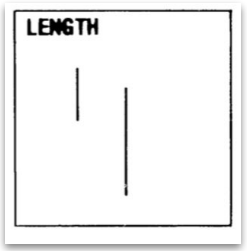
Presenting on Cleveland and McGill in addition to some work that came after as cited.

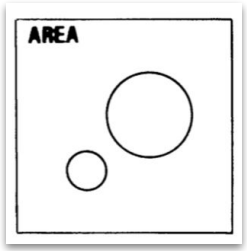
Fairly robust hierarchy

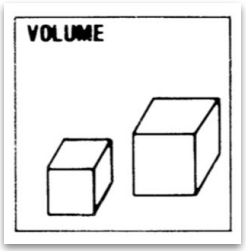
Higher Accuracy — Lower or Inconsistent

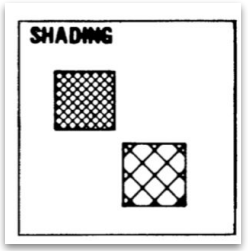


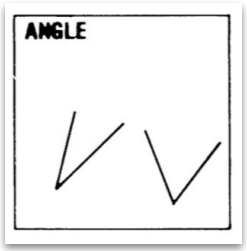




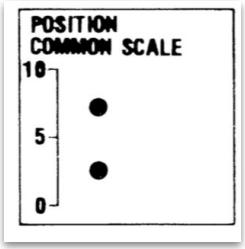




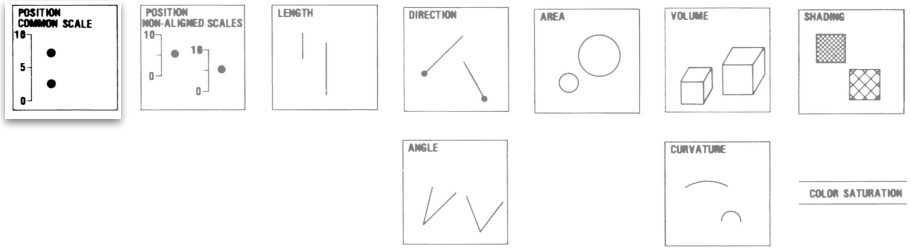




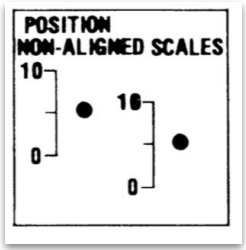
Fairly robust hierarchy

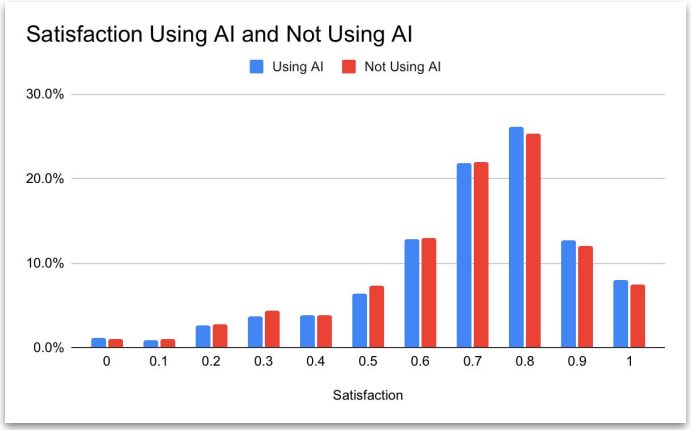


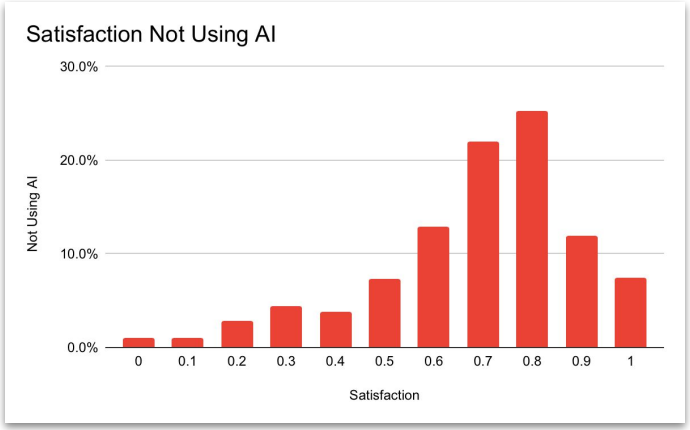
The highest accuracy encoding device is potentially not surprising as it underpins common patterns like scatter plots.

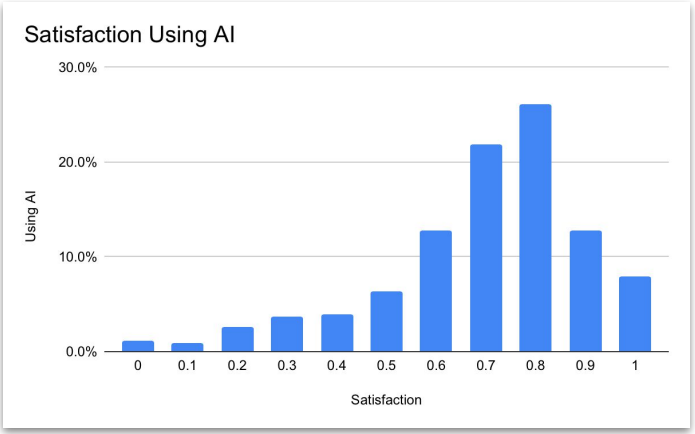


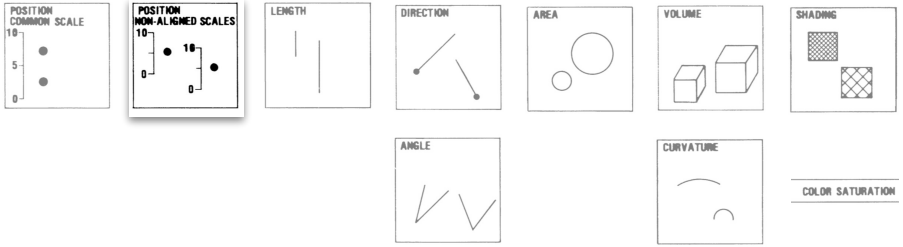
Fairly robust hierarchy



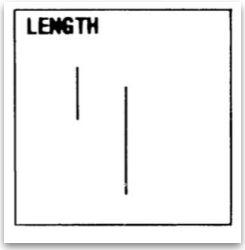


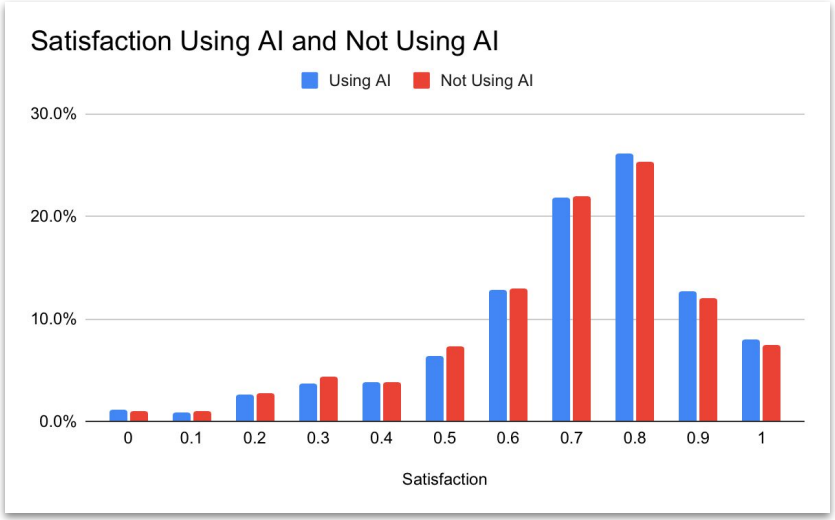


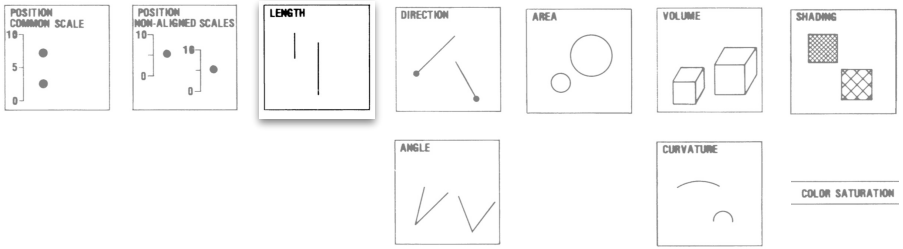




Fairly robust hierarchy



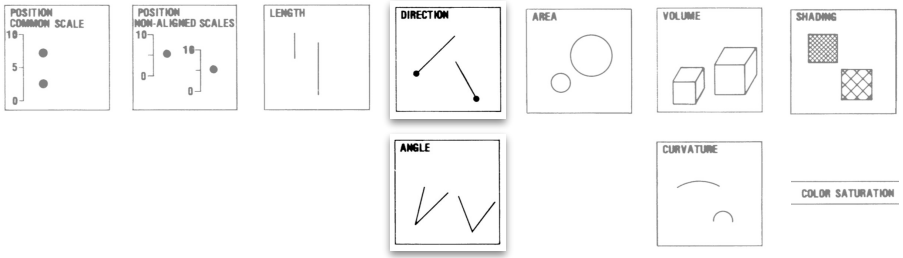




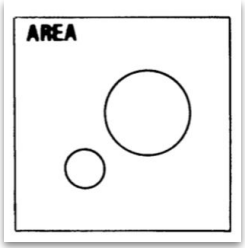
Fairly robust hierarchy

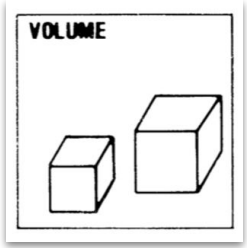
I am placing length higher than direction (slope) and angle because of consistency. There's evidence that we do better with angle closer to cardinal directions.

This is why pie charts may perform relatively poorly. Length typically has an easy fix: align against a common axis.



Fairly robust hierarchy

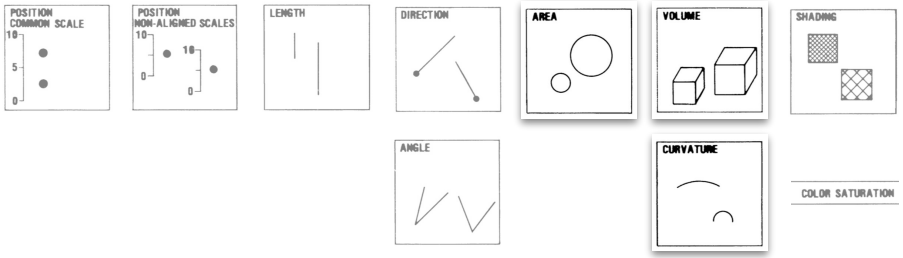




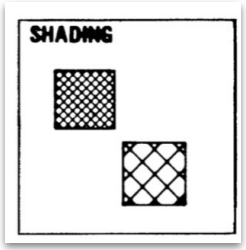
Volume generally goes poorly. This may be partially due to 3D representation within 2D media.

In general, area is good for less important “contextualizing” metrics.

Area has an issue: area vs radius.



Fairly robust hierarchy

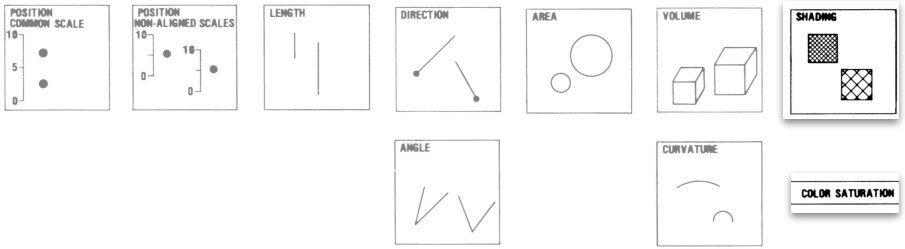


Color is fairly unreliable. It is often better for branding or complementing a message through emotion and aesthetic than it is for conveying quantitative information.

It may still serve a purpose for a limited number of qualitative groups.

Lightness generally better than hue.

Accessibility issues: readability and color-blindness.



Group Activity

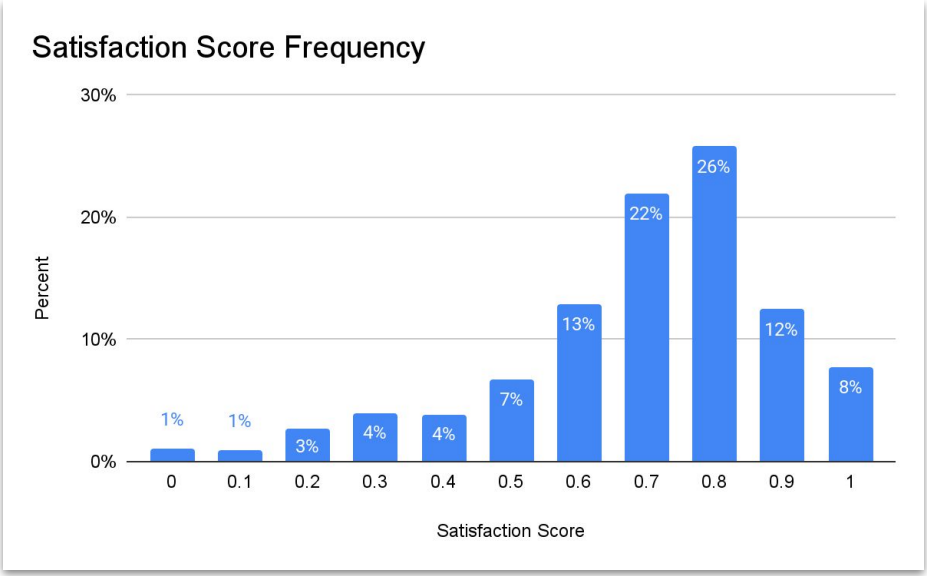
🎮

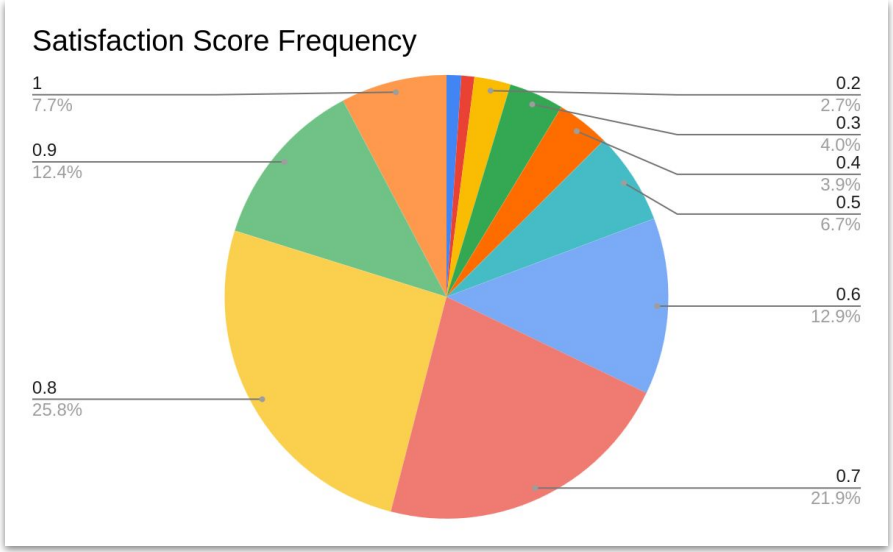
Which graphic is more likely to be read accurately?

---

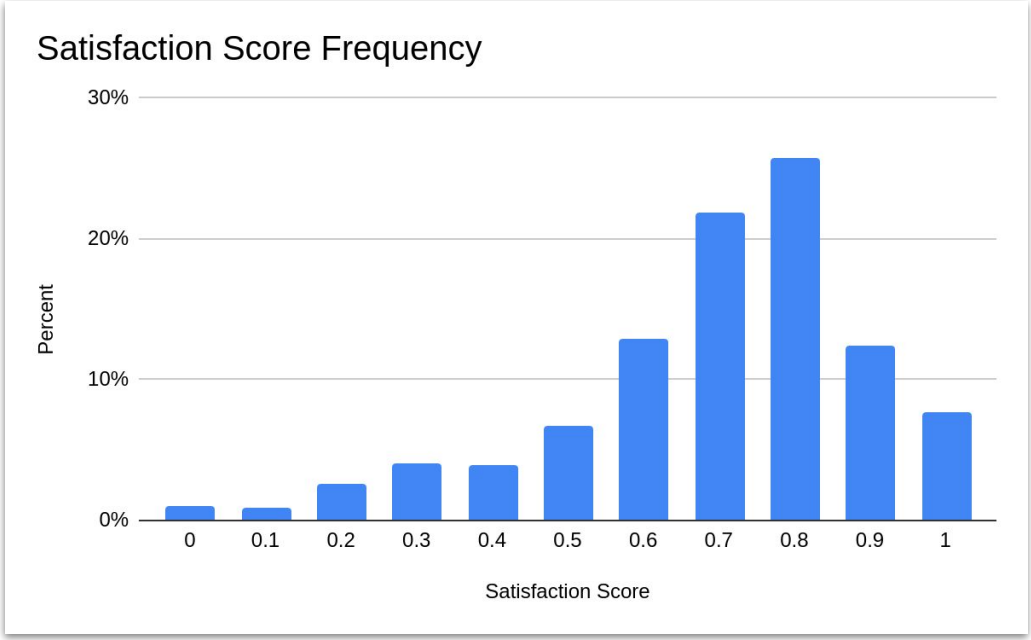
https://interactivedatascience.courses/stat159.pdf

Which one is more likely to be successful?



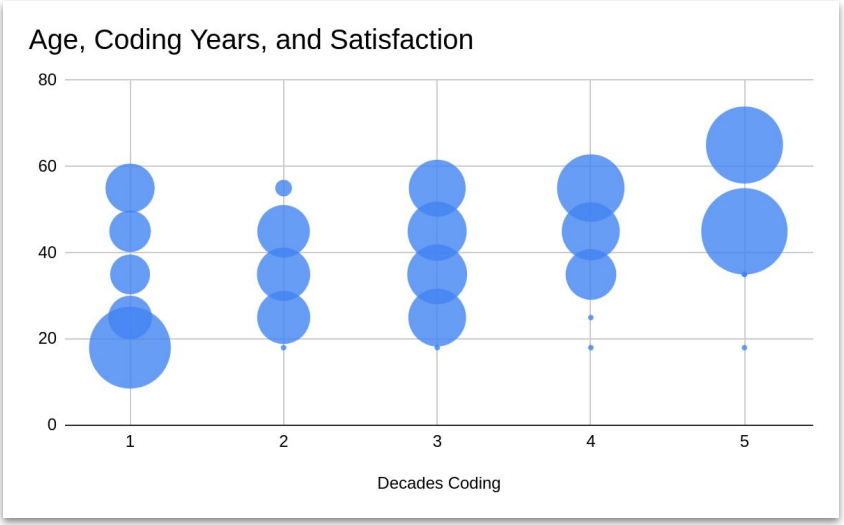


Which one is more likely to be successful?



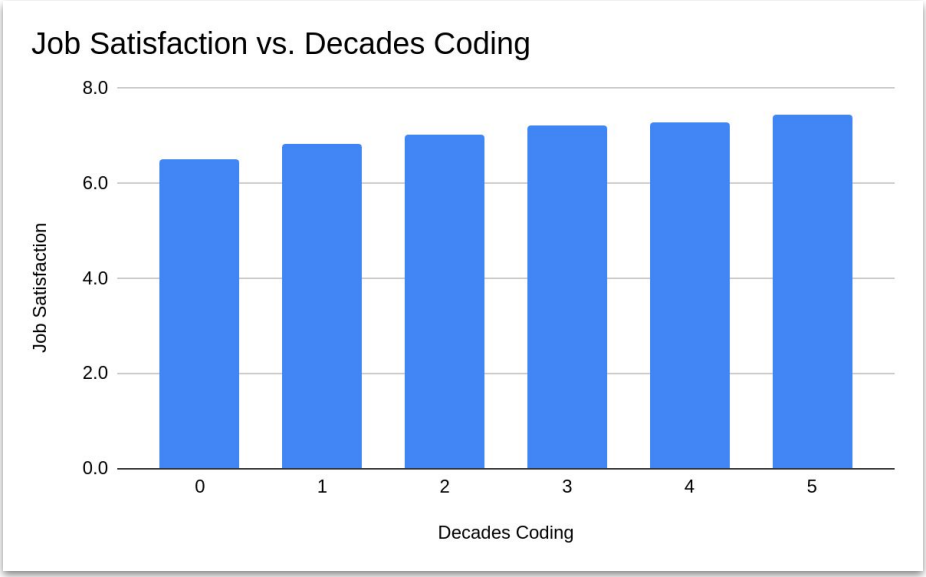
Satisfaction Score	Percent
0	
0.1	
0.2	
0.3	
0.4	
0.5	
0.6	
0.7	
0.8	
0.9	
1	

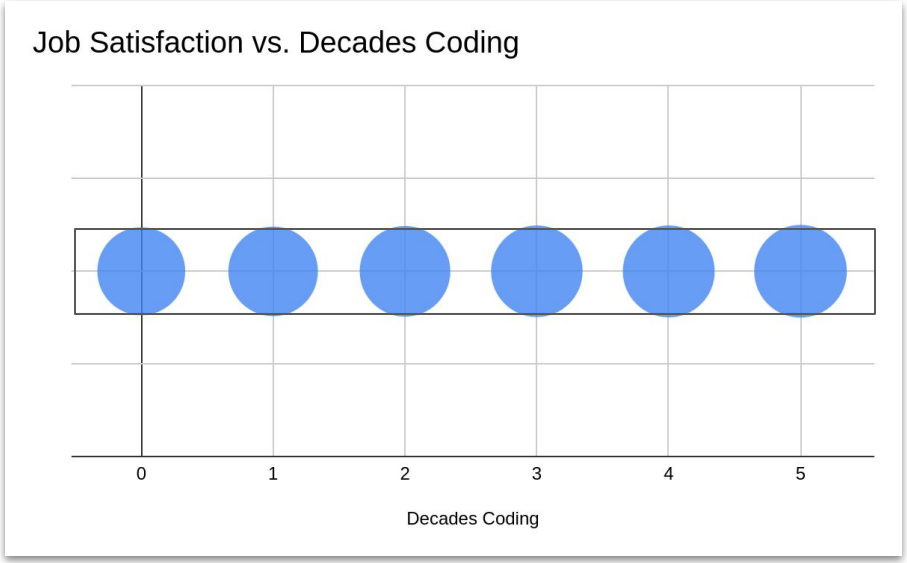
Which one is more likely to be successful?



Age Group	Decades Coding				
	1	2	3	4	5
65					
55					
45					
35					
25					
18					

Which one is more likely to be successful?

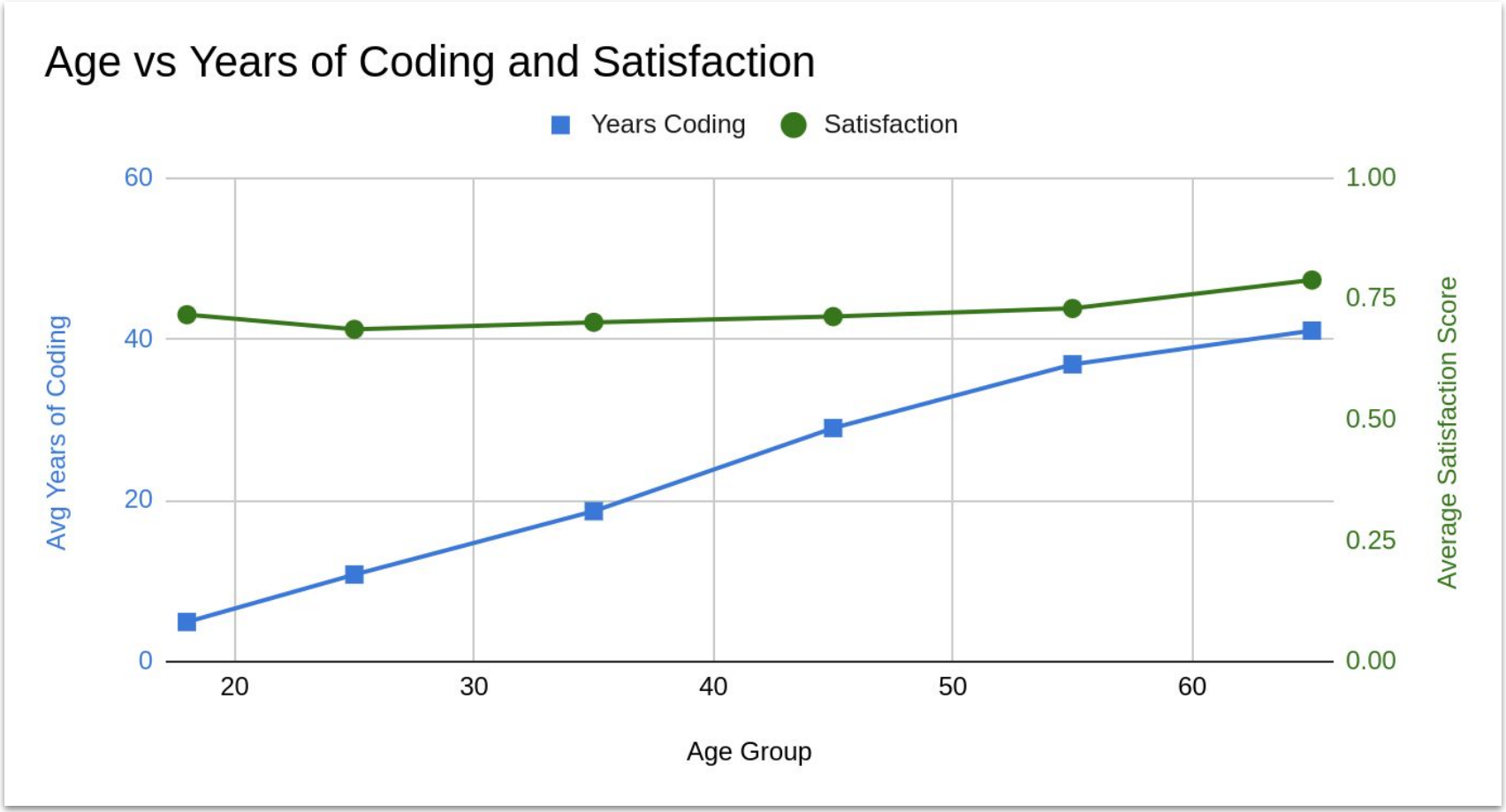




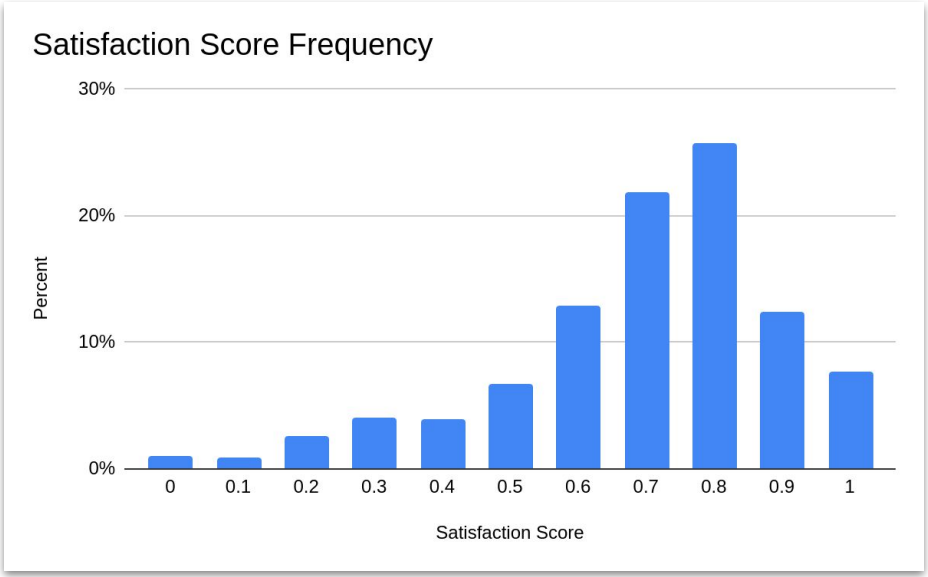
Working in limitations

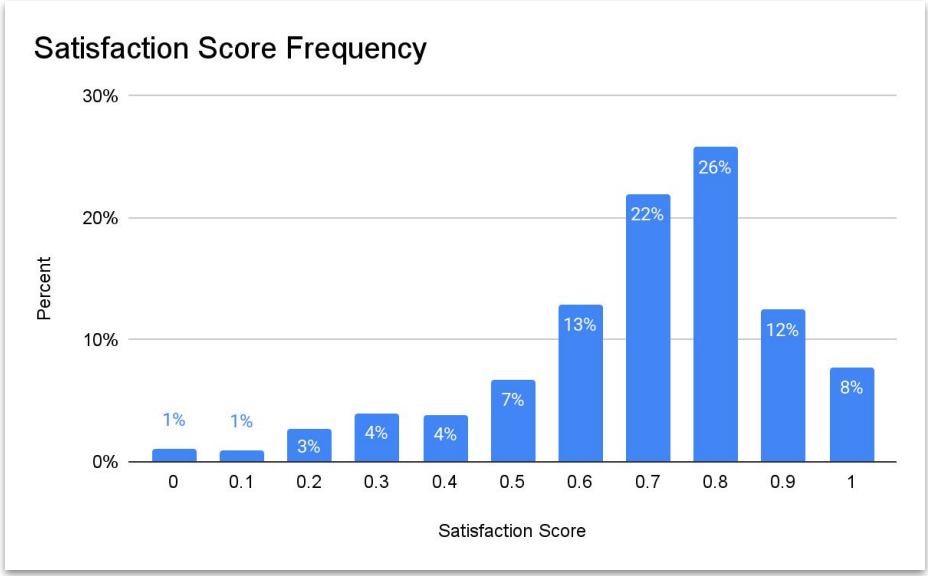
Shared axes, dual axes, and direct labeling.

Dual Axes

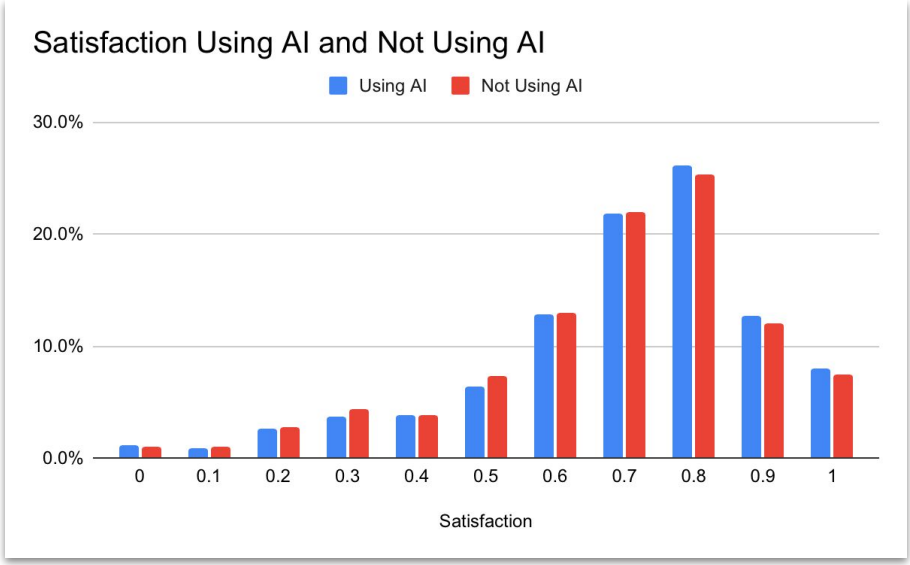

Age vs Years of Coding and Satisfaction
Years Coding
Satisfaction
60
40
20
0
1.00
0.75
0.50
0.25
0.00
Avg Years of Coding
Average Satisfaction Score
20
30
40
50
60
Age Group

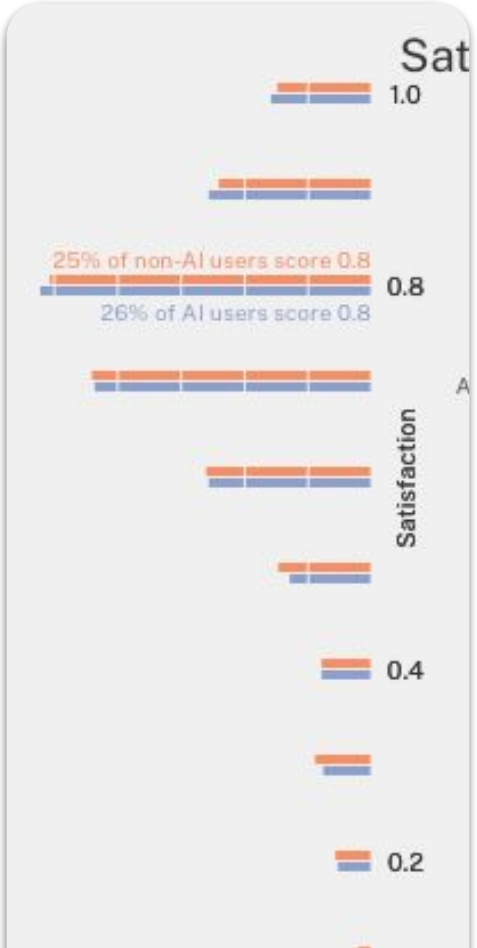
Direct Labeling



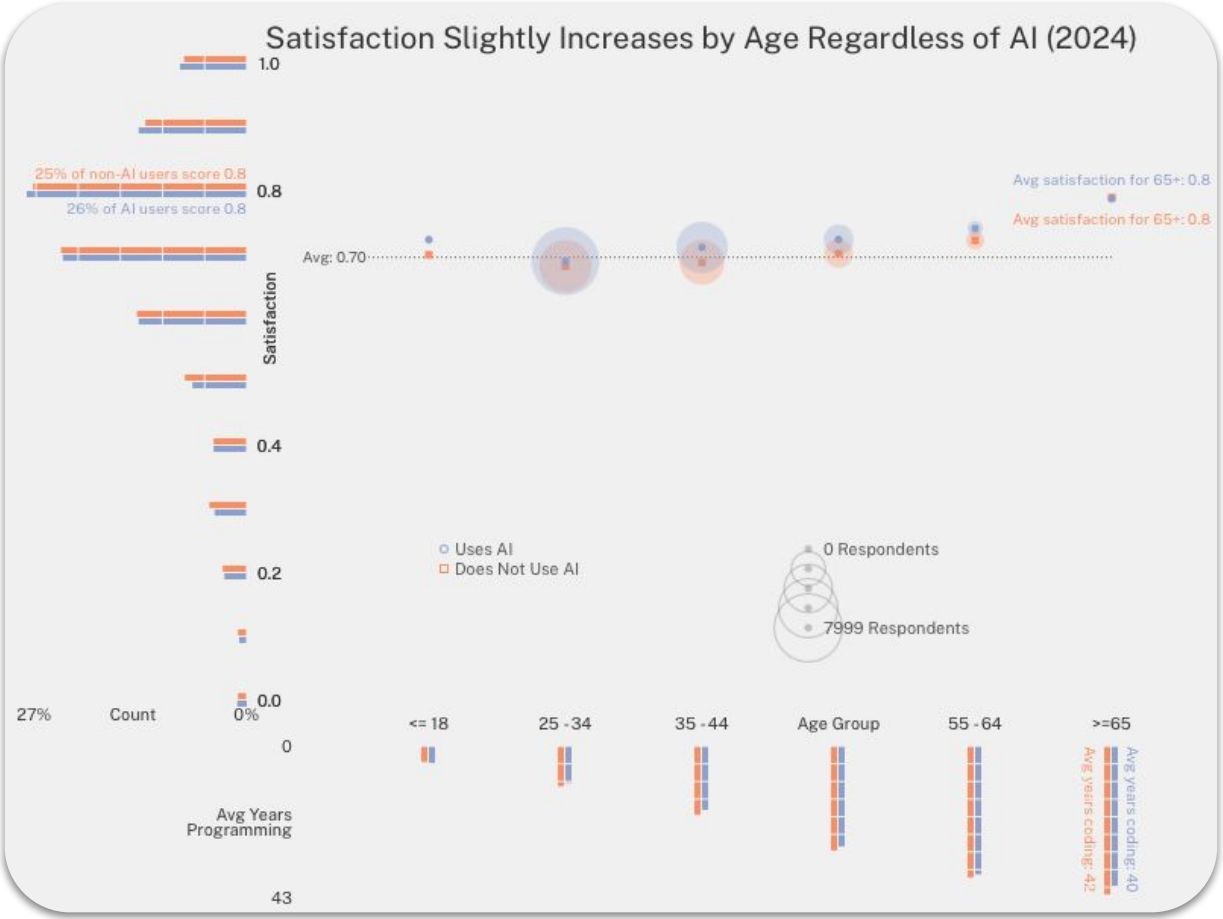


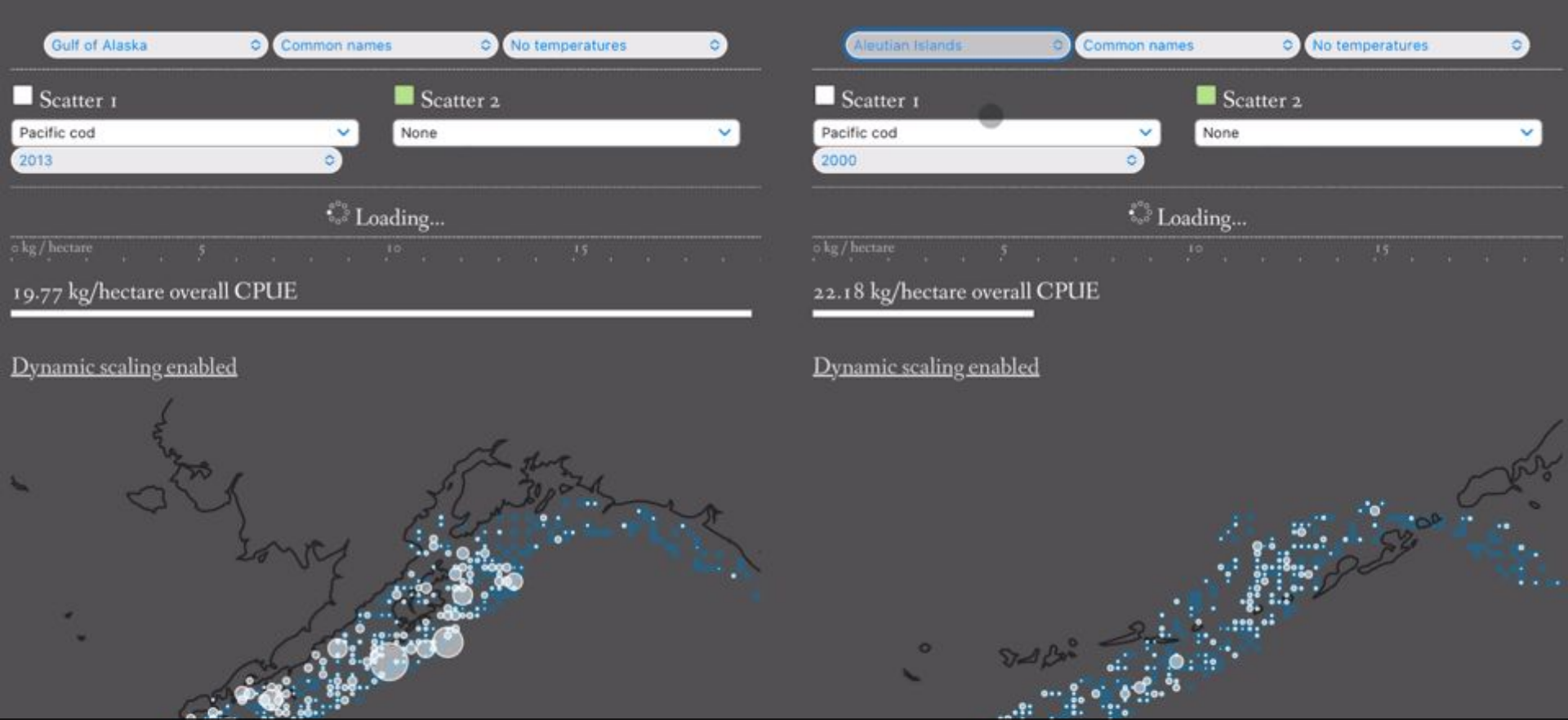
Keeping channels clear



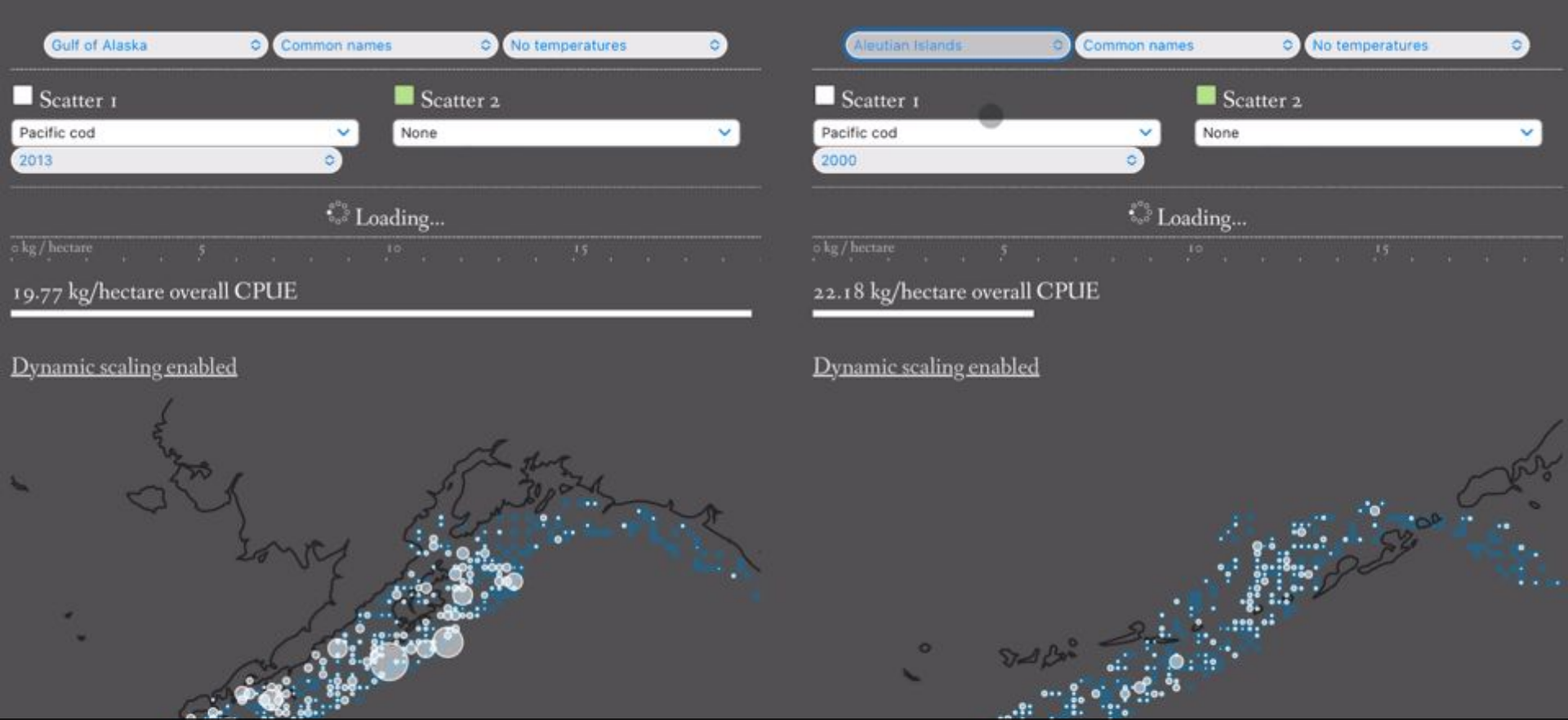


Tie it together




Gulf of Alaska
Common names
No temperatures
Scatter 1
Pacific cod
2013
Scatter 2
None
Loading...
0 kg / hectare
5
10
15
19.77 kg/hectare overall CPUE
Dynamic scaling enabled
Aleutian Islands
Common names
No temperatures
Scatter 1
Pacific cod
2000
Scatter 2
None
Loading...
0 kg / hectare
5
10
15
22.18 kg/hectare overall CPUE
Dynamic scaling enabled

<Break>



How to think about visualization

How can data become visible?

Data -> Graphic

4 Perspectives

| As representation

As task

As message

As dialogue

Why and how different groups do data visualization. How you can think about it in your work.

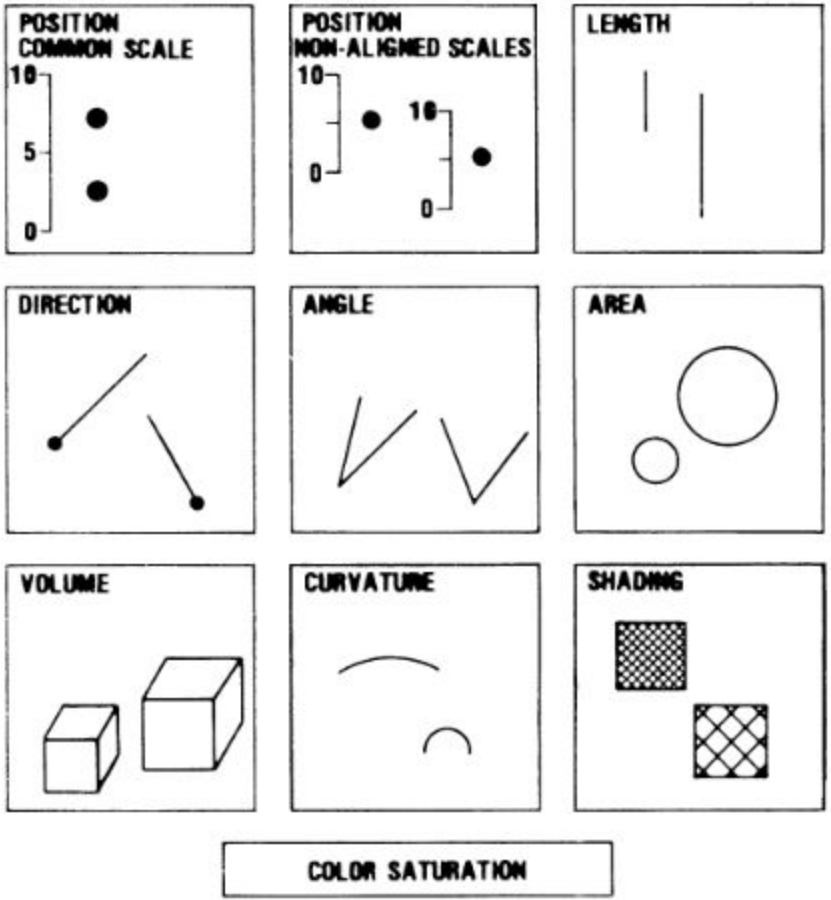


Figure 1. Elementary perceptual tasks.

Example: This first way of thinking about data visualization focuses on encoding.

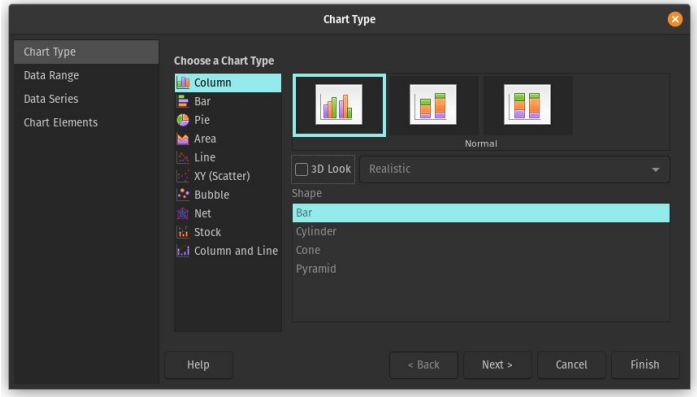
How do we “map” attributes of data to visual attributes?

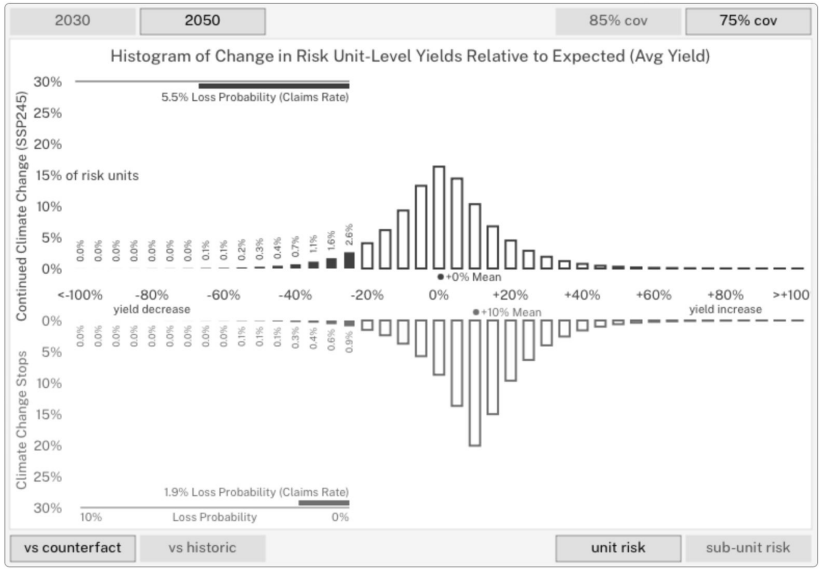
What visual encodings are better than others?

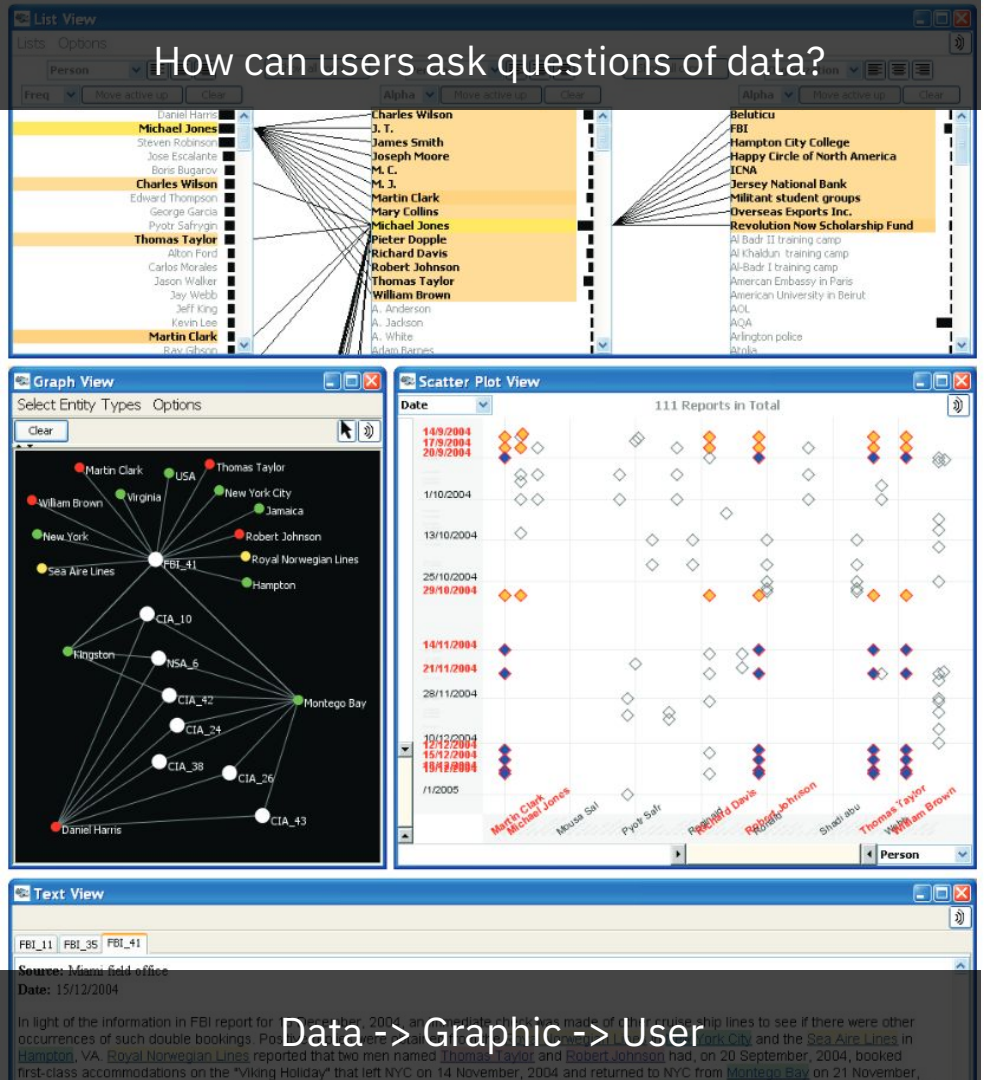
How do we make visualizations accessible?

Offers: Flexibility beyond the chart wizard but principles to guide us.

Gives us the basic building blocks for how humans process visual information but lets us use that understanding in many different ways.







4 Perspectives

As representation

| As task

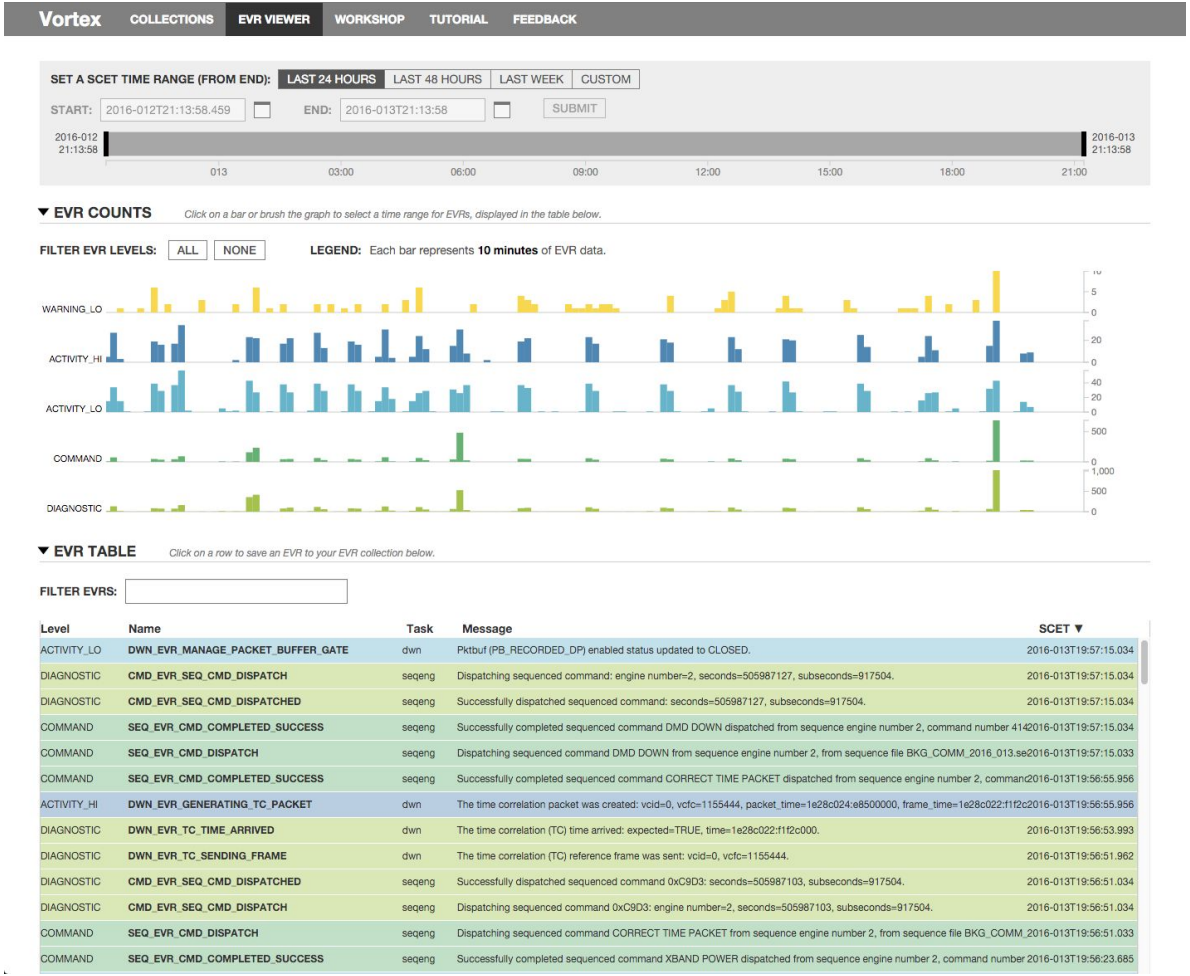
As message

As dialogue

Why and how different groups do data visualization. How you can think about it in your work.

Premise: Visualizations are part of a broader user journey.

A structured way to think about the user in the context of data visualization.



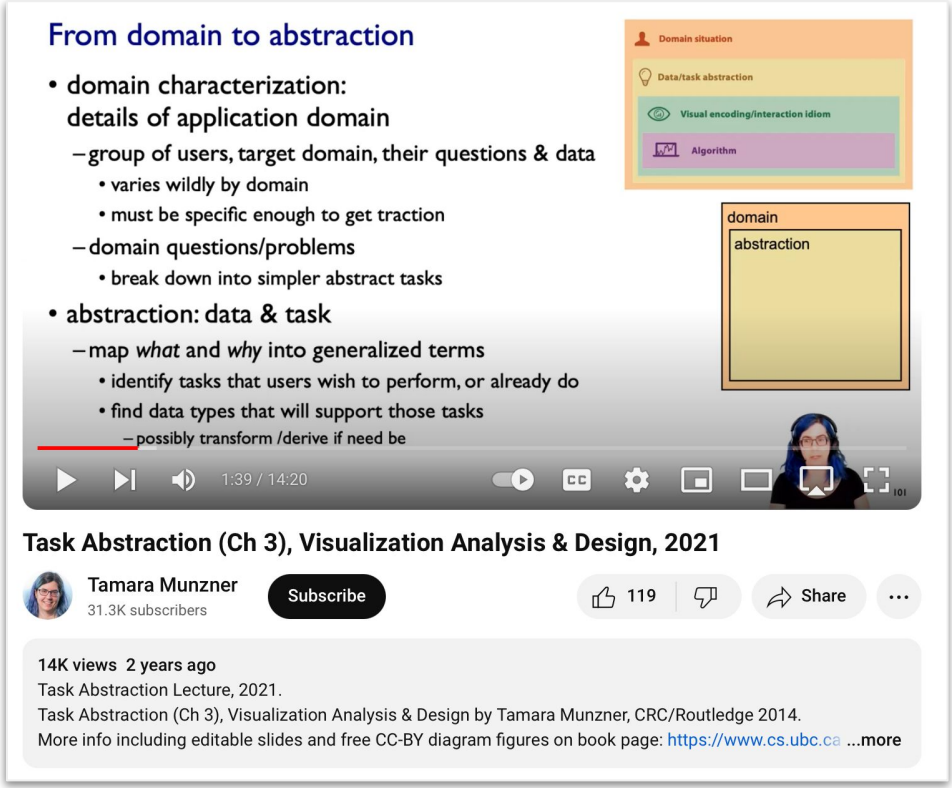
Example: Rachel Binx at NASA.

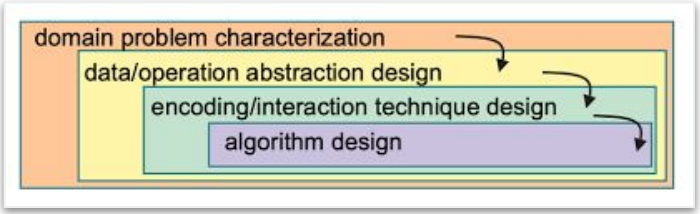
Looking at “event records” sent from spacecraft to NASA.

Interviewed a bunch of users to figure out how they worked with these data previously (log files).

Binx talks about how people had never seen their data before visually and the periodicity of events was revelatory for example.

Boils down into “tasks” the user executes and build user experiences to support those tasks.





Offers: Structured evidence-based understanding of the user to support them in their tasks.

Oritents around domains, tasks, questions, and data.

Fits within a broader modern user experience design dialogue.

How can data tell stories?

Data -> Graphic -> Audience

4 Perspectives

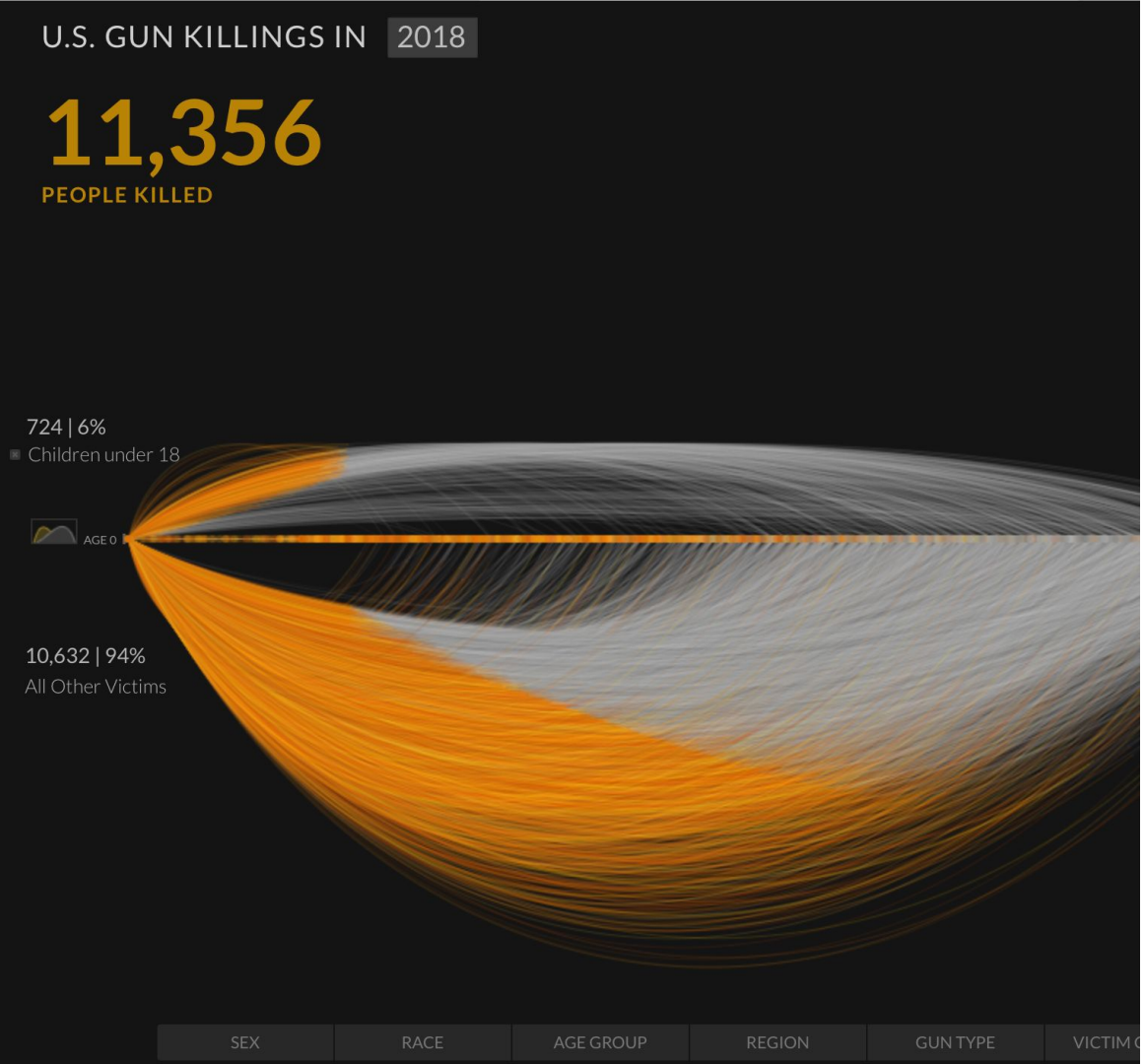
As representation

As task

| As message

As dialogue

Why and how different groups do data visualization. How you can think about it in your work.



Premise: Forms given to data enable authors to convey a message to a reader.

How does the reader feel when going through a visualization?

Where is efficiency helpful but where does it conflict with the message of the piece?

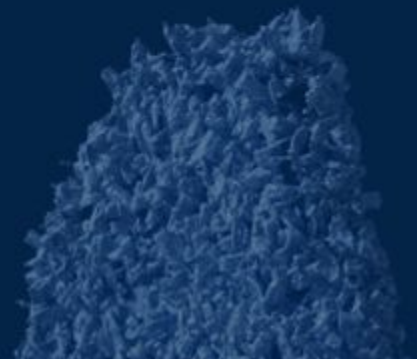
How might we defy reader expectations or have them confront prior held beliefs?

guns.periscopic.com

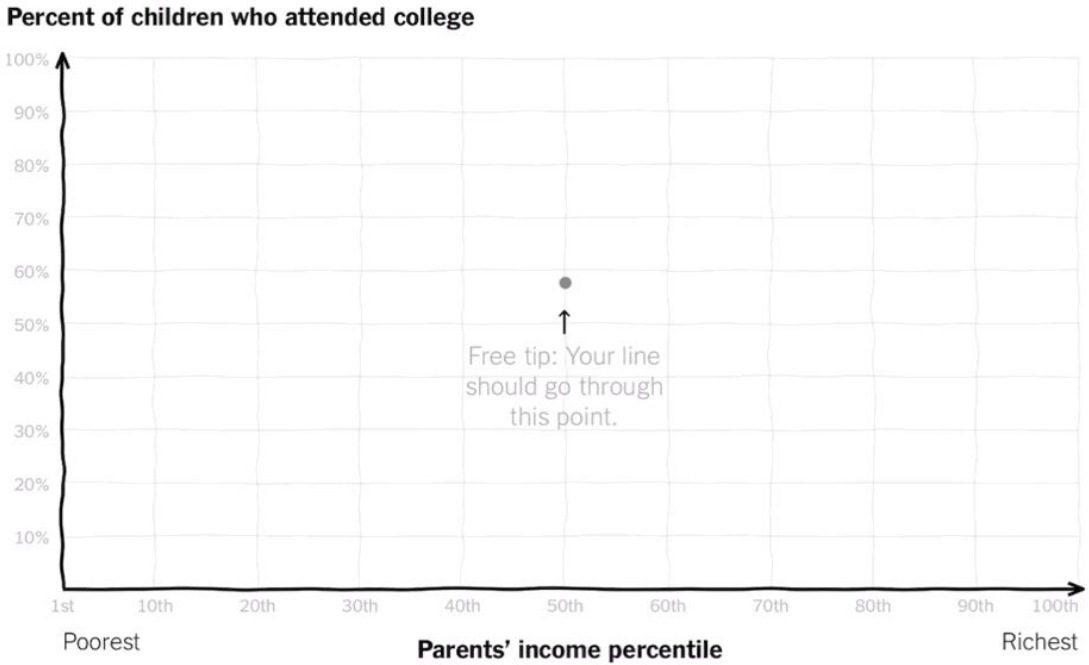
A Treaty To End Plastic Pollution. Forever.

World leaders must take steps to drastically limit the impact of plastics on the environment and human health

CHOOSE OUR FUTURE



Draw your line on the chart below



I'm done

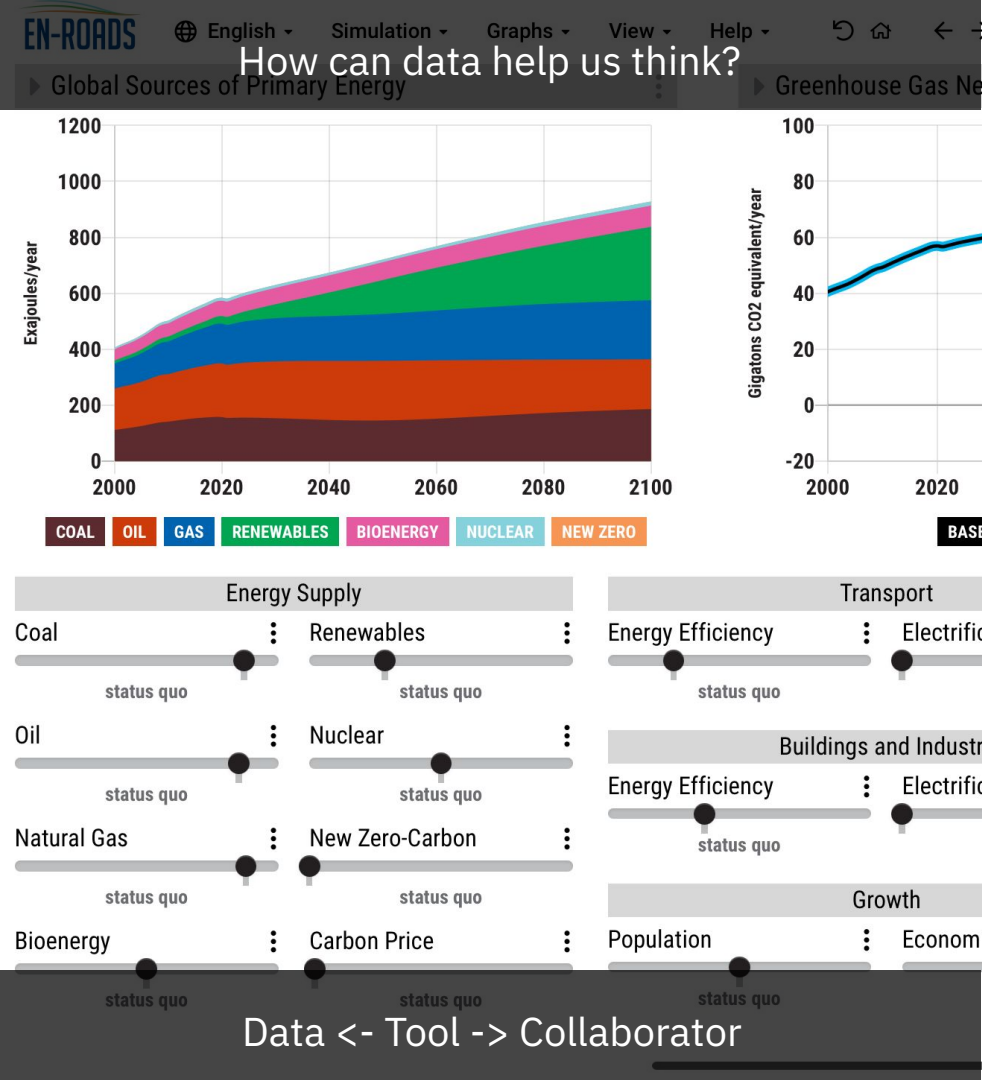
Start over

Offers: A way to convey messages with logos and pathos.

How to invoke emotional response.

How to challenge reader assumptions.

How to understand the process by which messages and meaning are interpreted.



4 Perspectives

As representation

As task

As message

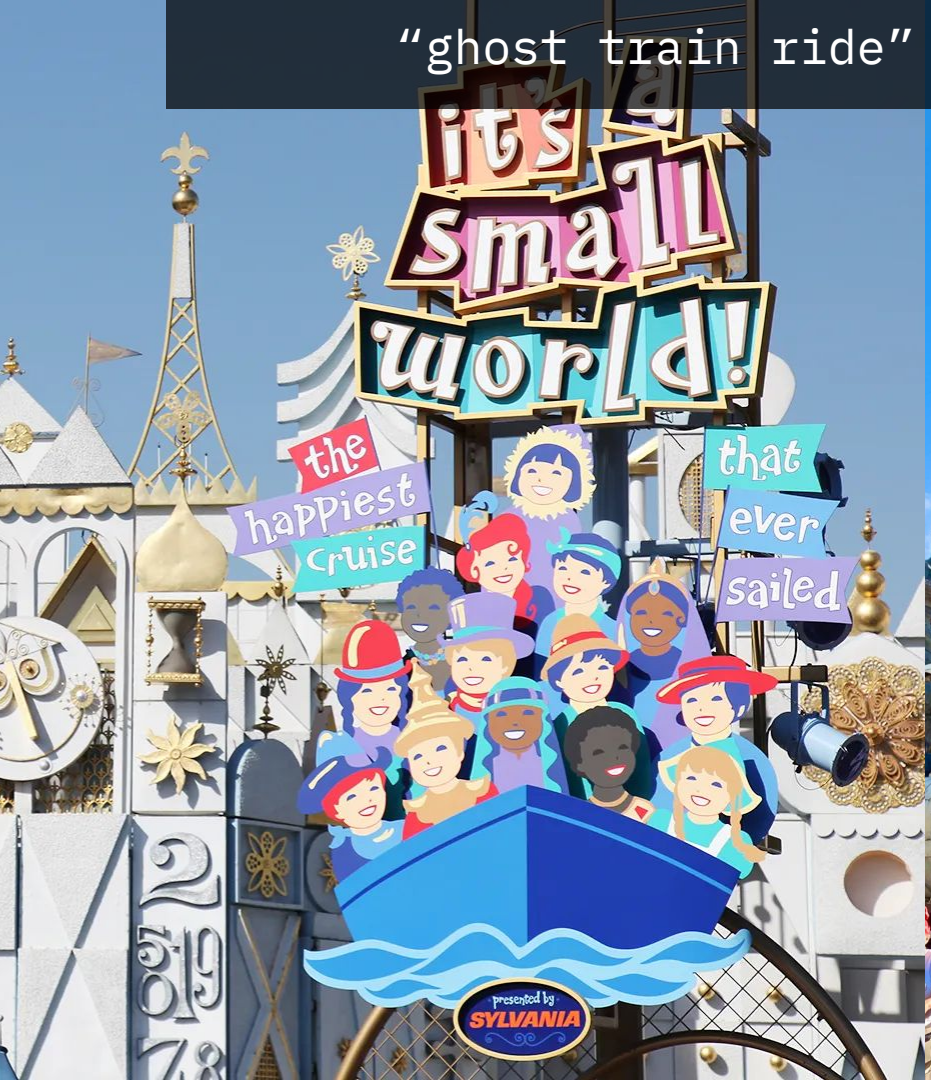
| As dialogue

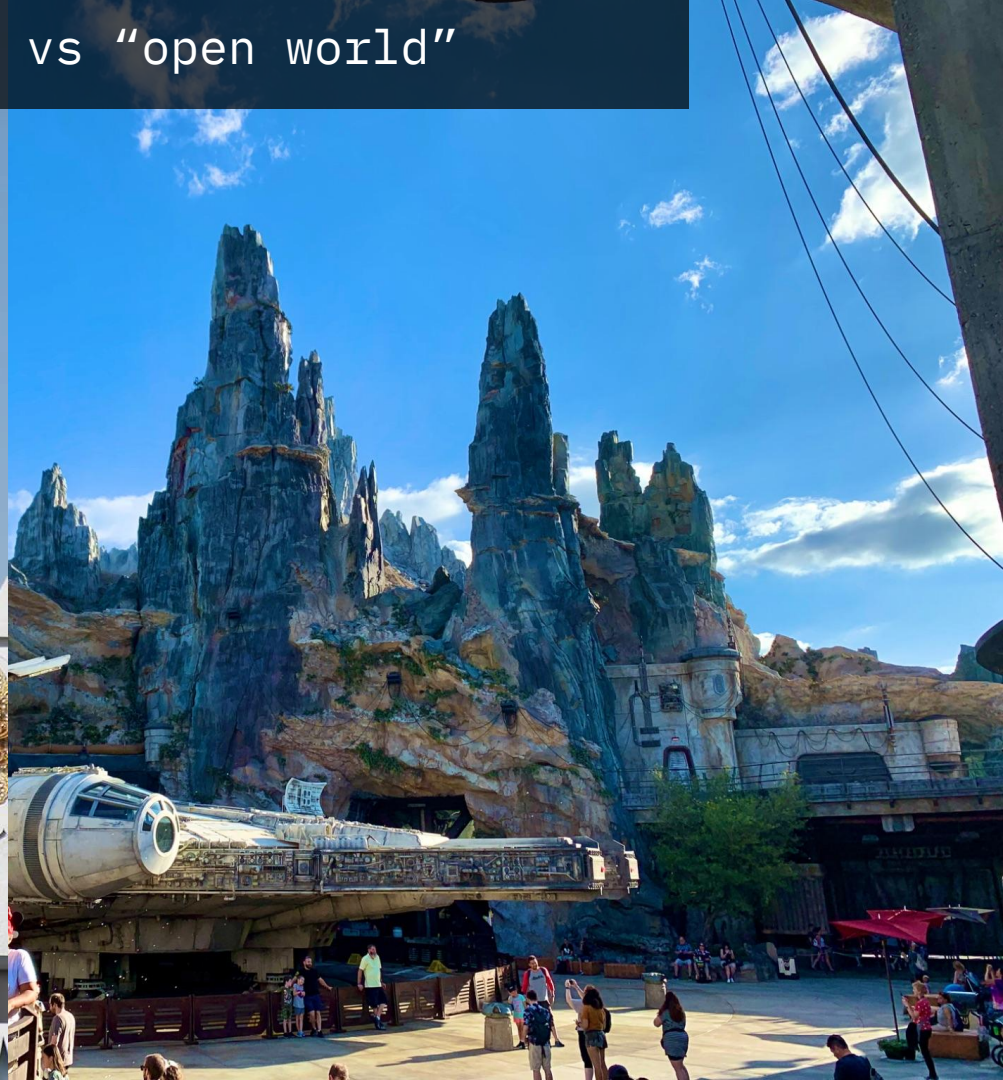
Why and how different groups do data visualization. How you can think about it in your work.

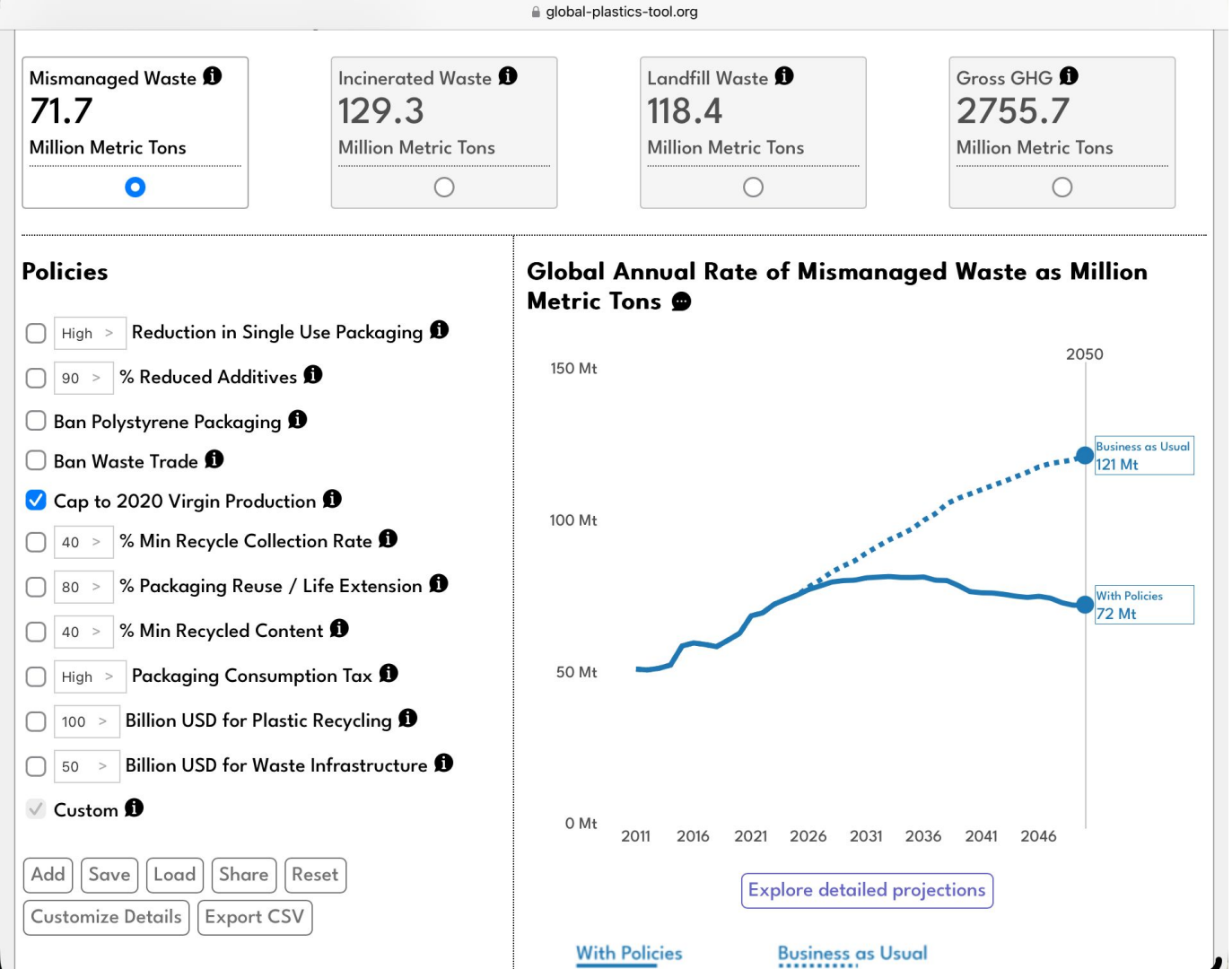
Premise: Data as humane dynamic media.

The designer creates media for thought, elevating the reader to an author of tools and co-creator of meaning.

“ghost train ride” vs “open world”







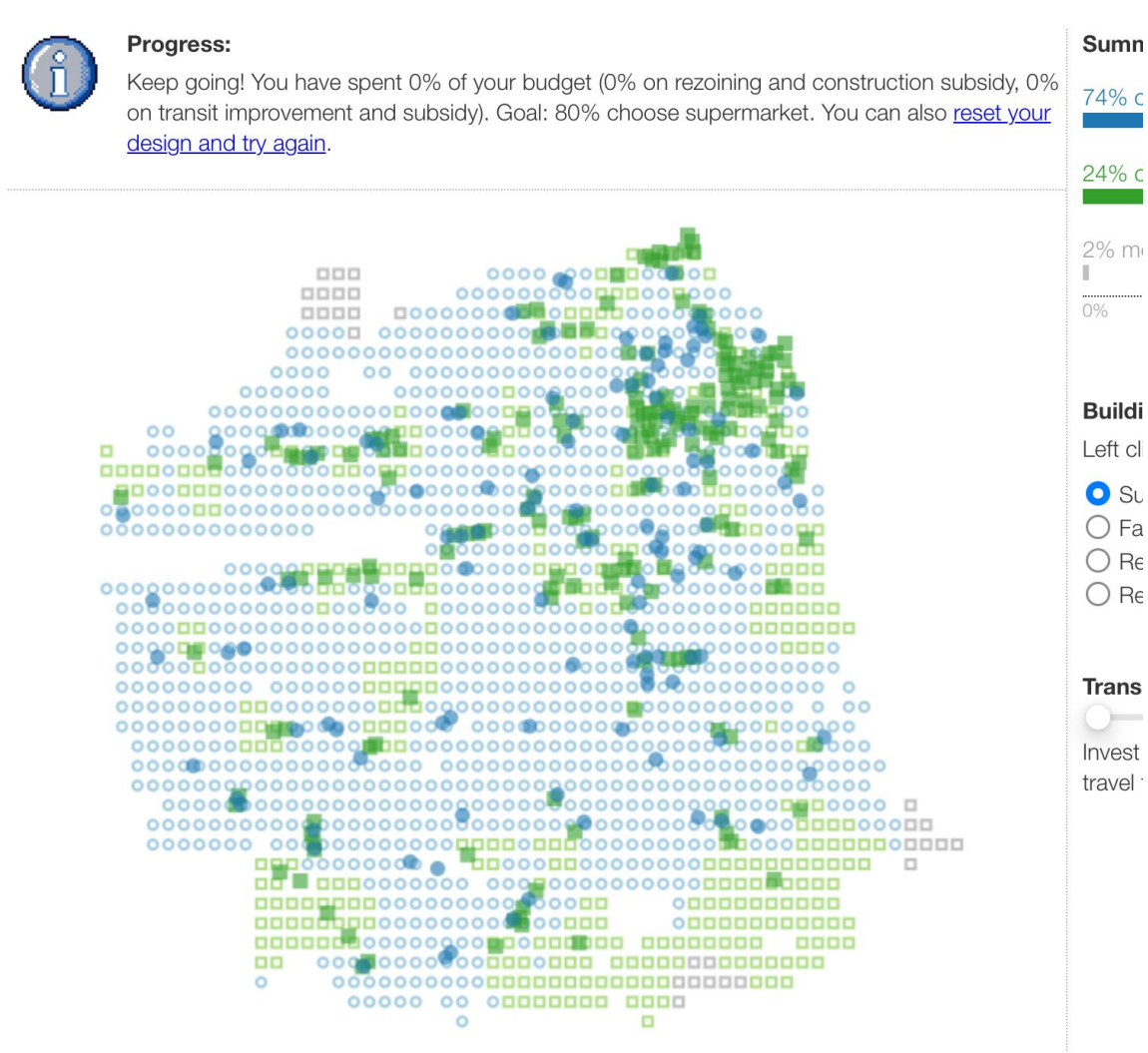
Example:
Finding a solution to the plastics crisis.

A layered experience in which the user can simulate different policies.

An invitation to build outside the original designer's intention.

🎮

https://global-plastics-tool.org



Offers: Co-creation and user agency.

Often leaning on game design concepts.

How to teach with/without tutorializing.

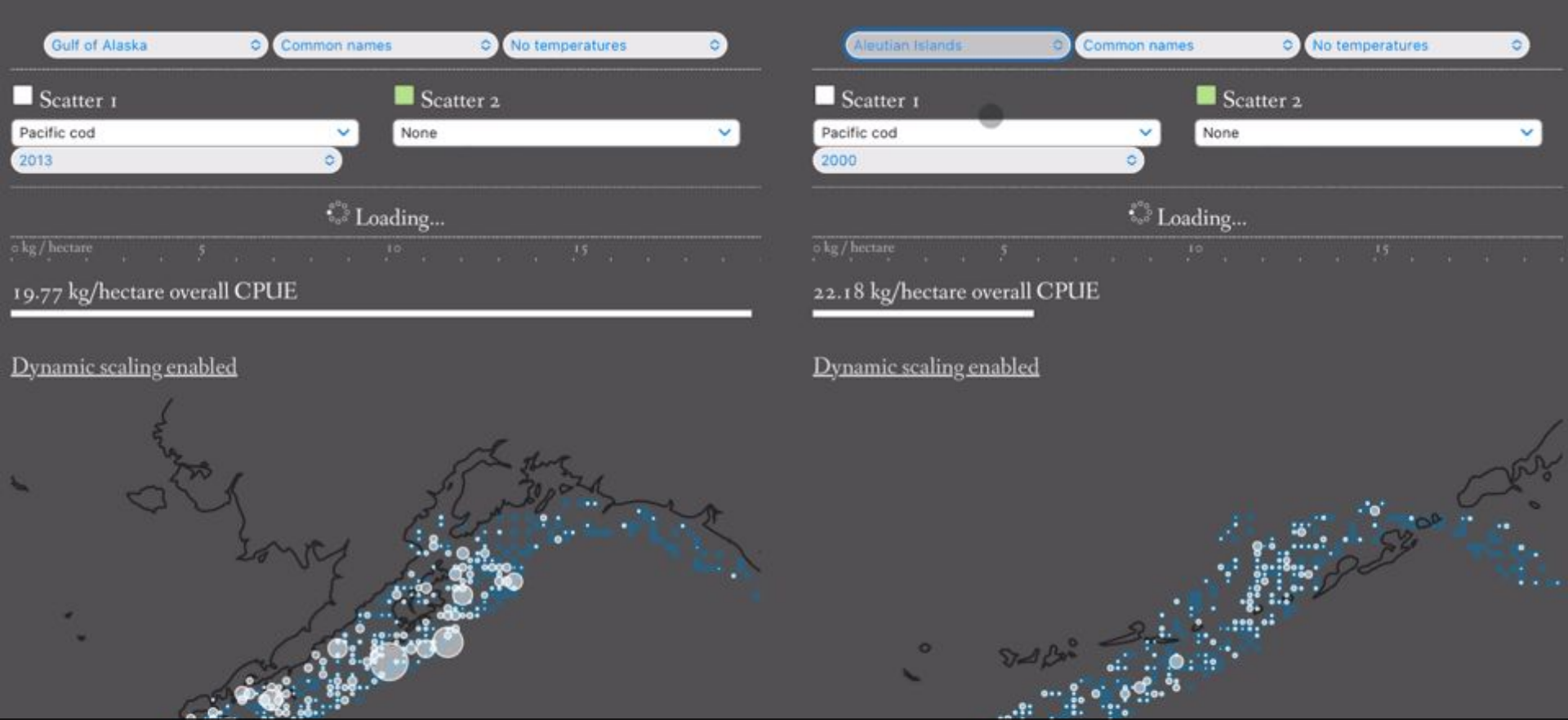
How to create spaces to interrogate assumptions.

How to build media to be repurposed.

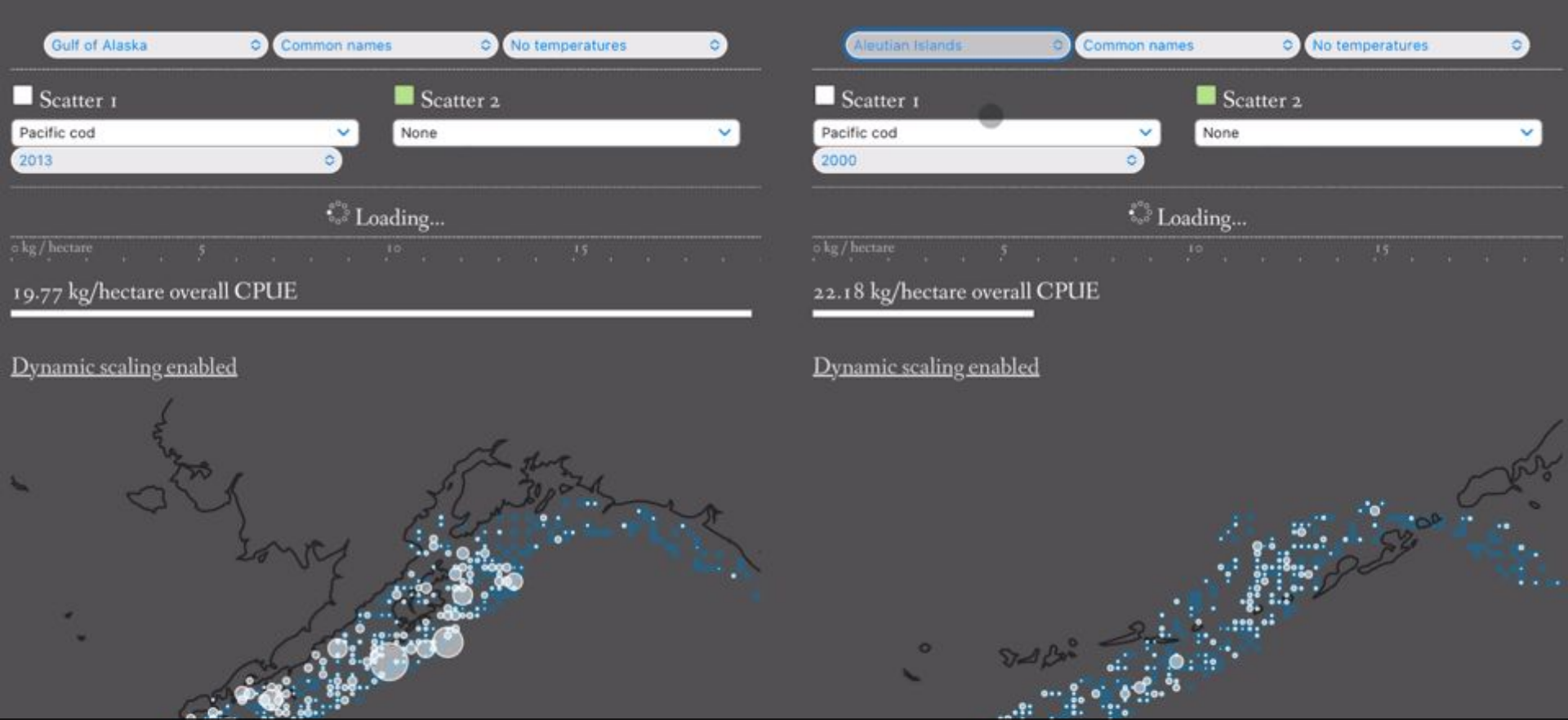
How to design experiences where the user is co-author.

Keep it going: Check-in before we go to code.

What’s something you learned today?


Gulf of Alaska
Common names
No temperatures
Scatter 1
Pacific cod
2013
Scatter 2
None
Loading...
0 kg / hectare
5
10
15
19.77 kg/hectare overall CPUE
Dynamic scaling enabled
Aleutian Islands
Common names
No temperatures
Scatter 1
Pacific cod
2000
Scatter 2
None
Loading...
0 kg / hectare
5
10
15
22.18 kg/hectare overall CPUE
Dynamic scaling enabled

<Break>



How to implement visualizations

Let’s start with a pre-built chart

Data:
https://
interactivedatascience.courses
/labs
/wolvesMoose.csv

Let’s start with a pre-built chart

```
import matplotlib.pyplot
import pandas

# Load data into pandas DataFrame
df = pandas.read_csv('wolvesMoose.csv')
```

Let’s start with a pre-built chart

```
# Create figure with subplots
fig, (ax1, ax2) = matplotlib.pyplot.subplots(2, 1, figsize=(15, 6))
```

Let’s start with a pre-built chart

```
# Plot wolves data
ax1.bar(df['Year'], df['Wolves'], color='#d95f02', alpha=0.8)
ax1.set_title('Wolf Population Over Time', fontsize=14, fontweight='bold')
ax1.set_xlabel('Year', fontsize=12)
ax1.set_ylabel('Number of Wolves', fontsize=12)
ax1.grid(True, alpha=0.3)
```

Let’s start with a pre-built chart

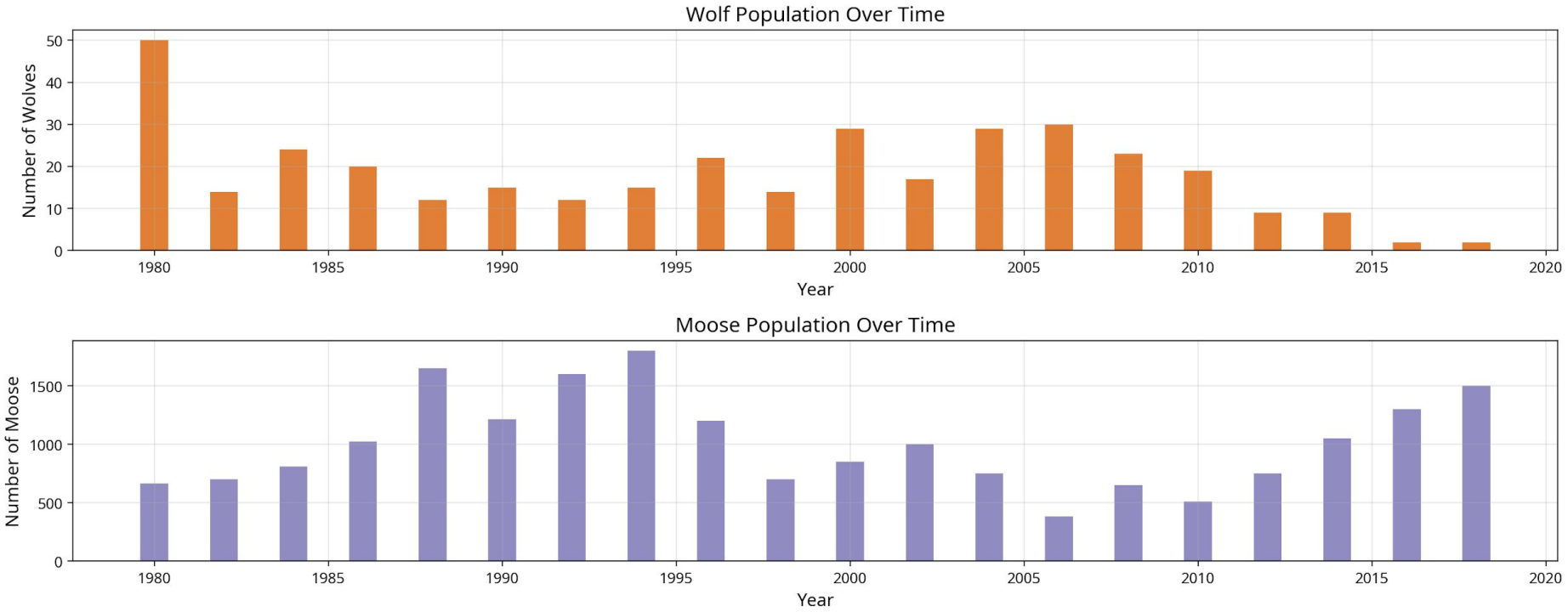
```
# Plot moose data
ax2.bar(df['Year'], df['Moose'], color='#7570b3', alpha=0.8)
ax2.set_title('Moose Population Over Time', fontsize=14, fontweight='bold')
ax2.set_xlabel('Year', fontsize=12)
ax2.set_ylabel('Number of Moose', fontsize=12)
ax2.grid(True, alpha=0.3)
```

Let’s start with a pre-built chart

```
# Adjust layout to prevent overlap
matplotlib.pyplot.tight_layout()

# Show the plot
matplotlib.pyplot.show()
```

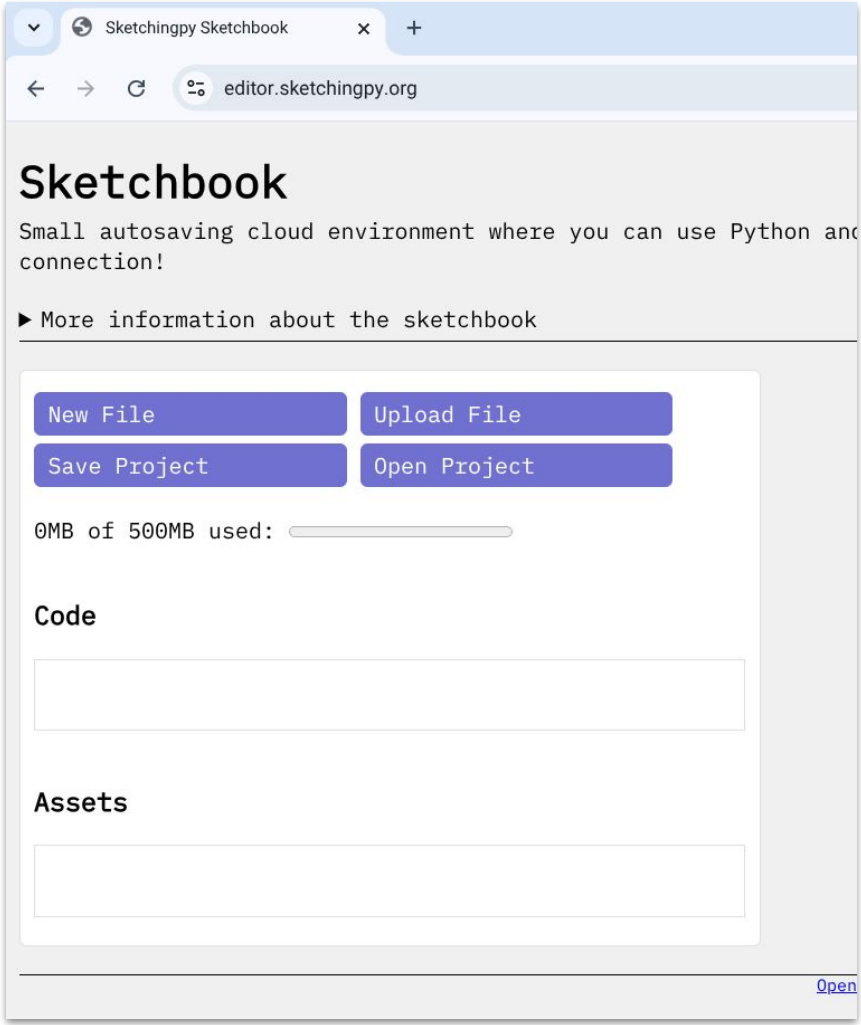
Let’s start with a pre-built chart



Where did the colors come from?

https://colorbrewer2.org/

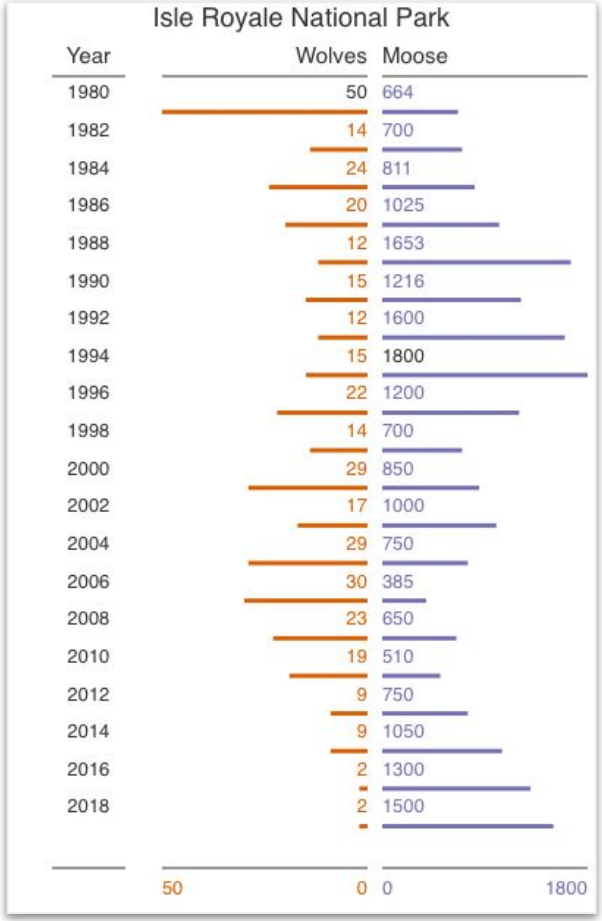
https://webaim.org/resources/contrastchecker/



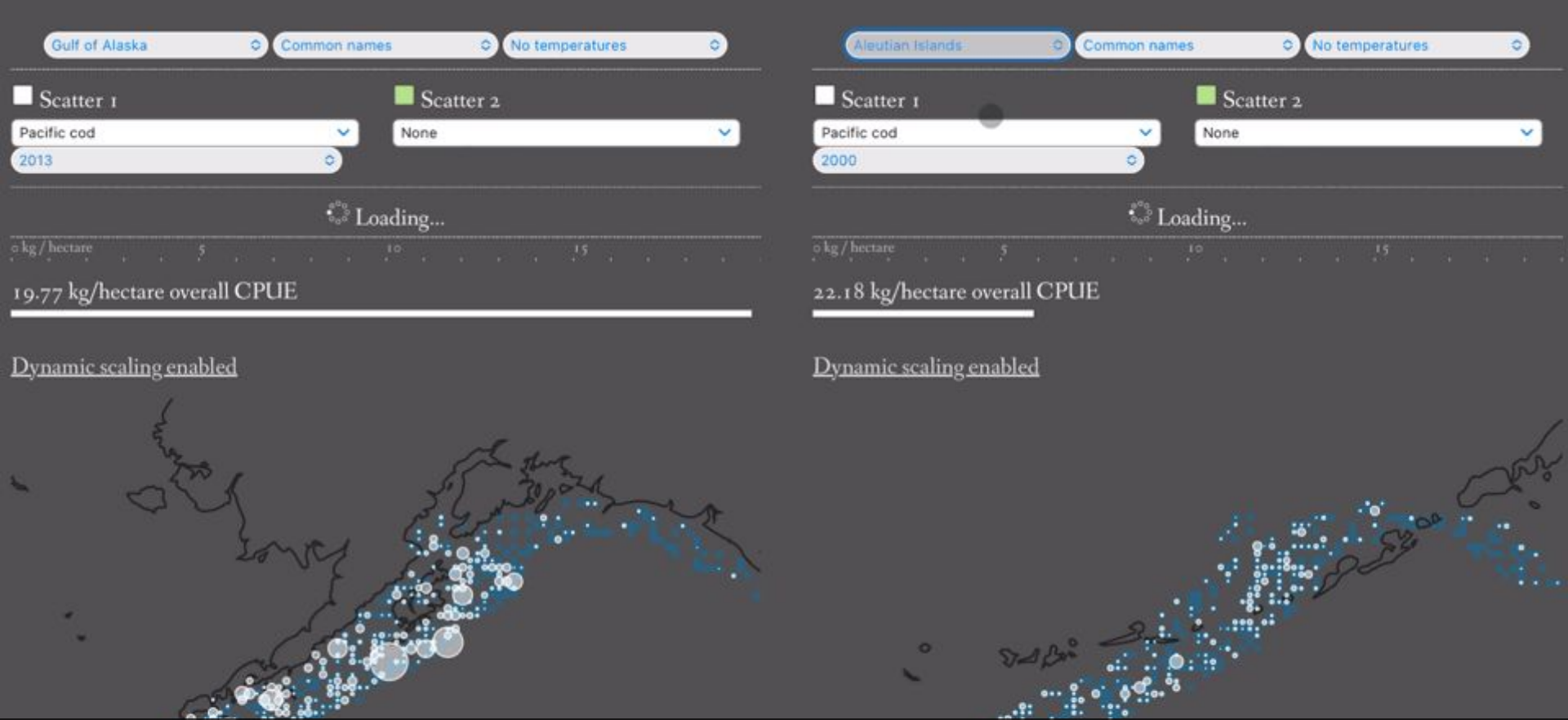
Tool:
https://
editor
.sketchingpy
.org

Data:
https://
interactivedatascience.courses
/labs
/wolvesMoose.csv

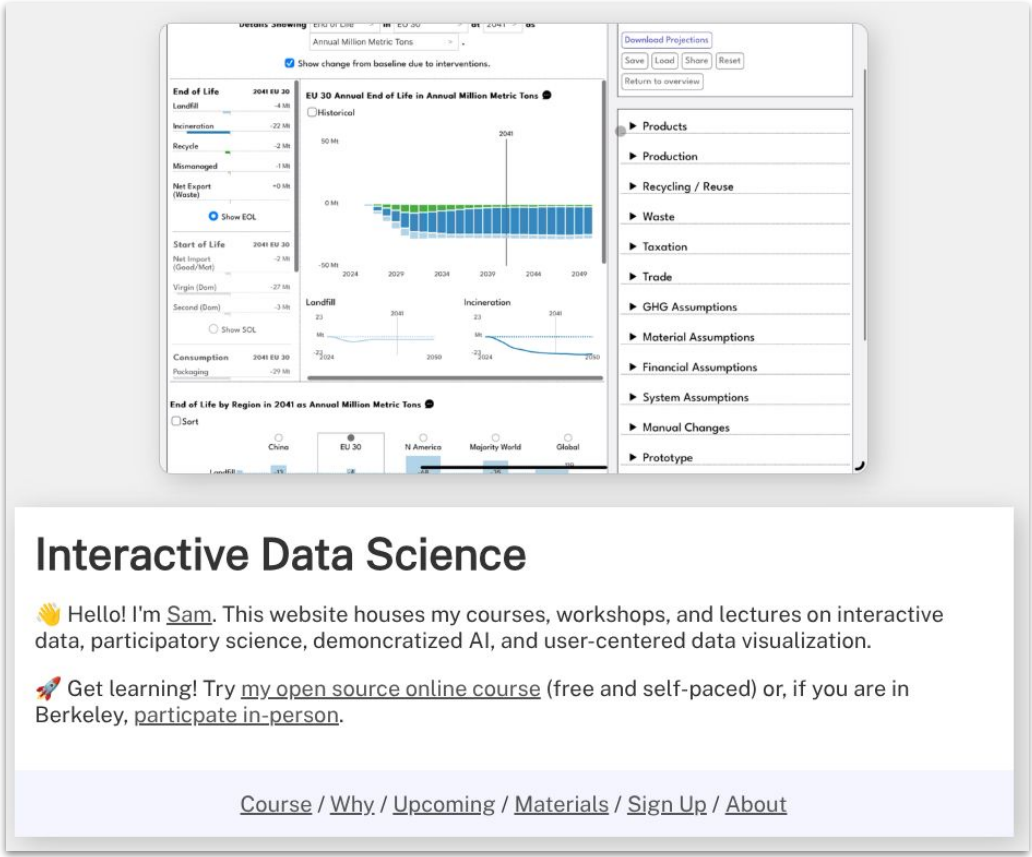
An alternative



https://gist.github.com/sampottinger/8721ba6d20f9ff5387900e393f5a9b3d

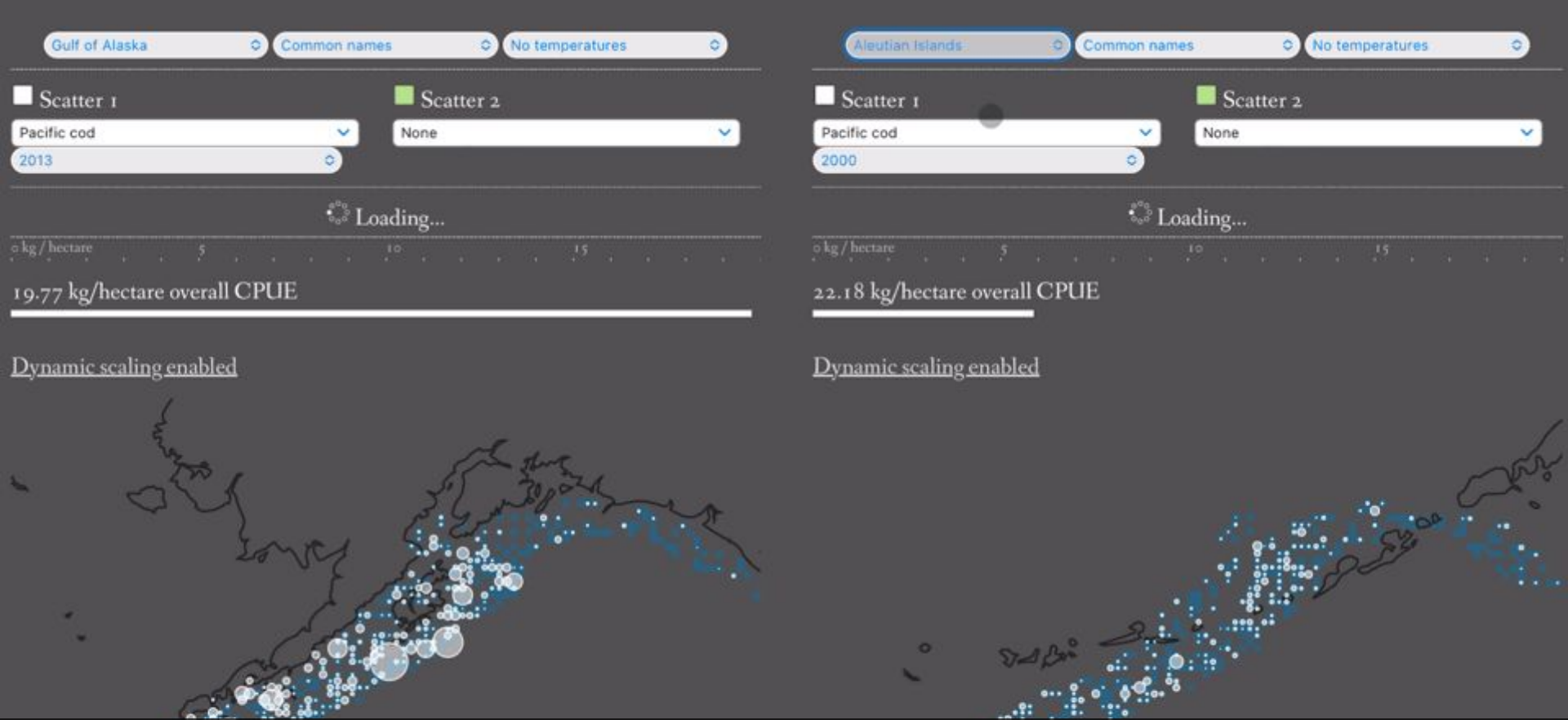


How to continue your journey



Keep going:
Resources online.

https://
interactive
data
science
.courses



Works Cited / Further Reading

[1] G. Aisch, A. Cox, and K. Quealy, “You Draw It: How Family Income Predicts Children’s College Chances.” The New York Times Company, May 28, 2015. [Online]. Available: https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html

[2] R. Binx, “Vortex.” 2015. [Online]. Available: https://rachelbinx.com/data-visualization/vortex

[3] R. Binx, “Designing for Realtime Spacecraft Operations.” BocoupLLC, Apr. 2016. [Online]. Available: https://www.youtube.com/watch?v=HuYKhSHcRSQ

[4] Book Books Book, “We Feel Fine.” YouTube, Dec. 2009. [Online]. Available: https://www.youtube.com/watch?v=vi8WrnWNSzU

[5] W. S. Cleveland and R. McGill, “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, Sep. 1984, doi: 10.1080/01621459.1984.10478080.

[6] Flipflops, “I feel... everything.” Flipflops.org, Oct. 2007. [Online]. Available: https://www.flipflops.org/category/thoughtful/page/2/

[7] V. Hart and N. Case, “Parable of the Polygons.” Nicky Case, 2022. [Online]. Available: https://ncase.me/polygons/

[8] Isle Royale National Park MIchigan, “Wolf & Moose Populations.” National Parks Service, Mar. 29, 2024. [Online]. Available: https://www.nps.gov/isro/learn/nature/wolf-moose-populations.htm

[9] S. D. Kamvar and J. Harris, “We feel fine and searching the emotional web,” in *Proceedings of the fourth ACM international conference on Web search and data mining*, Hong Kong China: ACM, Feb. 2011, pp. 117–126. doi: 10.1145/1935826.1935854.

[10] A. KIRK, *DATA VISUALISATION: a handbook for data driven design*. S.l.: SAGE PUBLICATIONS, 2024.

[11] A. Kirk, “Visualizing Data.” Visualising Data Ltd, 2024. [Online]. Available: https://visualisingdata.com/

[12] T. Munzner, “A Nested Model for Visualization Design and Validation,” *IEEE Trans. Visual. Comput. Graphics*, vol. 15, no. 6, pp. 921–928, Nov. 2009, doi: 10.1109/TVCG.2009.111.

[13] T. Munzner, “Visualization Analysis and Design.” AK Peters Visualization Series, 2014. [Online]. Available: https://books.apple.com/us/book/visualization-analysis-and-design/id1567434451

[14] T. Munzner, *Visualization analysis and design*. in A.K. Peters visualization series. Boca Raton: CRC Press, Taylor & Francis Group, CRC Press is an imprint of the Taylor & Francis Group, an informa business, 2015.

[15] T. Munzner, “Task Abstraction (Ch 3), Visualization Analysis & Design, 2021.” YouTube, 2021. [Online]. Available: https://www.youtube.com/watch?v=pHljd-cgICY

[16] M. Nix, *Visual simplexity: die Darstellung großer Datenmengen*. Frankfurt am Main: entwickler.press, 2013.

[17] Periscopic, “U.S. Gun Deaths.” Periscopic, 2018. [Online]. Available: https://guns.periscopic.com/

[18] A. Pottinger, “FoodSim: San Francisco.” 2023. [Online]. Available: https://foodsimsf.com/

[19]

[19]
A. Pottinger, “Income Gaps.” 2023. [Online]. Available: https://incomegaps.com/

[20]
A. Pottinger, “Interactive Data Science.” 2024. [Online]. Available: https://interactivedatascience.courses/

[21]
A. S. Pottinger *et al.*, “Combining Game Design and Data Visualization to Inform Plastics Policy: Fostering Collaboration between Science, Decision-Makers, and Artificial Intelligence,” 2023, *arXiv*. doi: 10.48550/ARXIV.2312.11359.

[22]
A. S. Pottinger, L. Connor, B. Guzder-Williams, M. Weltman-Fahs, and T. Bowles, “Climate-Driven Doubling of Maize Loss Probability in U.S. Crop Insurance: Spatiotemporal Prediction and Possible Policy Responses,” 2024, *arXiv*. doi: 10.48550/ARXIV.2408.02217.

[23]
A. S. Pottinger and G. Zarpellon, “Pyafscgap.org: Open source multi-modal Python-basedtools for NOAA AFSC RACE GAP,” *JOSS*, vol. 8, no. 86, p. 5593, Jun. 2023, doi: 10.21105/joss.05593.

[24]
J. N. Rooney-Varga, F. Kapmeier, J. D. Sterman, A. P. Jones, M. Putko, and K. Rath, “The Climate Action Simulation,” *Simulation & Gaming*, vol. 51, no. 2, pp. 114–140, Apr. 2020, doi: 10.1177/1046878119890643.

[25]
J. Schell, *The art of game design: a book of lenses*, Third edition. Boca Raton: CRC Press/Taylor & Francis Group, 2020.

[26]
J. Snow, *On the mode of communication of cholera*. London: John Churchill, 1855. [Online]. Available: https://archive.org/details/b28985266/page/n57/mode/2up

[27]
J. Stasko, C. Gorg, Z. Liu, and K. Singhal, “Jigsaw: Supporting Investigative Analysis through Interactive Visualization,” in *2007 IEEE Symposium on Visual Analytics Science and Technology*, Sacramento, CA, USA: IEEE, Oct. 2007, pp. 131–138. doi: 10.1109/VAST.2007.4389006.

[28]
The Document Foundation, *LibreOffice*. (2024). The Document Foundation.

[29]
ThoughtLab, The Wendy and Eric Schmidt Center for Data Science and Environment, and Benioff Ocean Science Laboratory, “A Treaty to End Plastic Pollution. Forever.” University of California, 2023. [Online]. Available: https://plasticstreaty.berkeley.edu/

[30]
B. Victor, “Inventing on Principle.” CUSEC, 2012. [Online]. Available: https://www.youtube.com/watch?v=PUv66718DII

[31]
B. Victor, “Media for Thinking the Unthinkable.” MIT Media Lab, Apr. 04, 2013. [Online]. Available: https://vimeo.com/67076984

[32]
Visual Computing BLOG, “Tamara Munzner discussed quantification in terms of a nested model of visualization design and evaluation.” Transregional Collaborative Research Center. [Online]. Available: https://www.visual-computing.org/2018/10/17/computerscienceconferenceweek/201810_conferenceweek_munzner-2/

[33]
C. Ware, “Colin Ware | The Data Visualzation Research Lab.” University of New Hampshire. [Online]. Available: https://ccom.unh.edu/vislab/people/colin_ware/

[34]
C. Ware, *Information visualization: perception for design*, Fourth edition. Cambridge, MA: Morgan Kaufmann, 2021.

[35]
Wikipedia Contributors, “Snow-cholera-map-1.jpg.” Wikimedia Foundation, Inc., 2020. [Online]. Available: https://en.wikipedia.org/wiki/File:Snow-cholera-map-1.jpg

[36]

[36] Wikipedia Contributors, “Bret Victor.” Wikimedia Foundation, Inc., Jun. 22, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Bret_Victor

[37] Wikipedia Contributors, “Star Wars: Galaxy’s Edge.” Wikimedia Foundation, Inc., Sep. 21, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Star_Wars:_Galaxy%27s_Edge

[38] Wikipedia Contributors, “It’s a Small World.” Wikimedia Foundation, Inc., Sep. 24, 2024. [Online]. Available: https://en.wikipedia.org/wiki/It%27s_a_Small_World

[39] N. Yee, “Motivations for Play in Online Games,” *CyberPsychology & Behavior*, vol. 9, no. 6, pp. 772–775, Dec. 2006, doi: 10.1089/cpb.2006.9.772.

Thanks to https://unsplash.com/photos/DHl49oyrn7Y

A. Pottinger, "TED Visualization," Gleap.org. Available: https://gleap.org/content/ted_visualization

W. Cleveland and R. McGill, “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” Journal of the American Statistical Association, 1984. Available: https://www.jstor.org/stable/2288400

“Stack Overflow Annual Developer Survey 2024,” Stack Exchange Inc, 2024. Available: https://survey.stackoverflow.co/

C. Ware, “Information Visualization: Perception for Design (Interactive Technologies),” Morgan Kaufmann, 2012.

A. Cairo, “The Truthful Art,” New Riders, 2016.

NYT Opinion, “10 Columnists and Writers Rate What Mattered in Trump’s First Full Month,” New York Times Company, 2025. Available: https://www.nytimes.com/interactive/2025/03/01/opinion/trump-administration-first-month.html.


CC BY-NC-SA 4.0