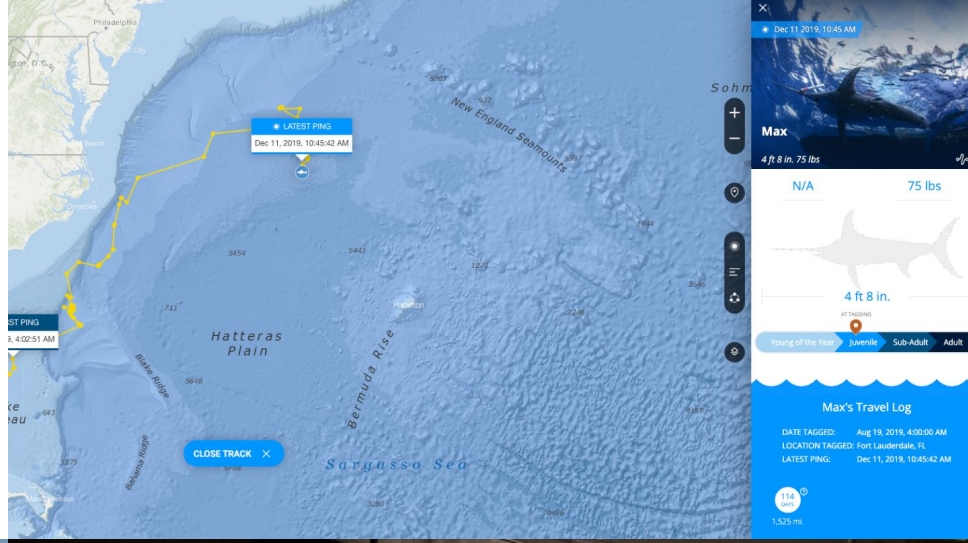
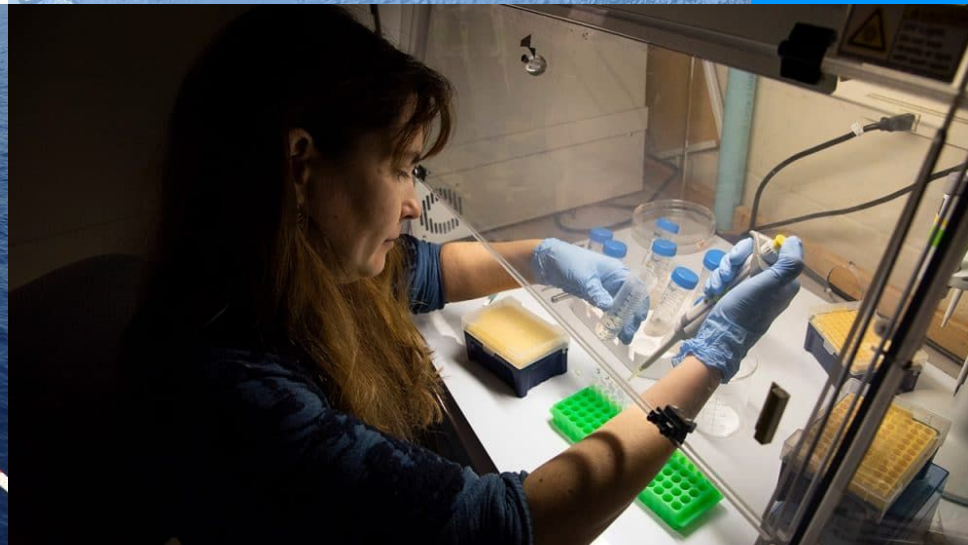
Enhancing Scientific Work

Caroline Berger
PhD Candidate
Aarhus University
23 October 2025

caroline.berger@cs.au.dk
carolineberger.github.io





Interview, Observation, Code Review with Scientists

Scientists and Code: Programming as a Tool, Berger et al. PLATEAU 2024

Workshop and Prototype Review with Marine Biologists

Under submission

Who programs?

Software Engineers

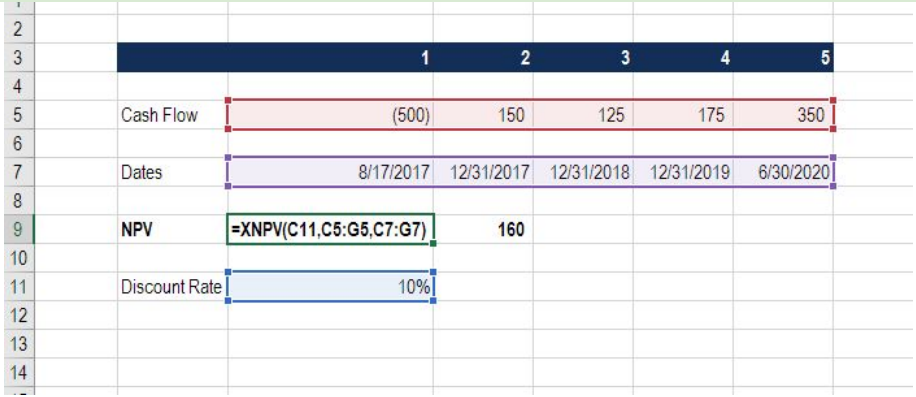
System Administrators

Cybersecurity Specialists

But Also

Finance Professionals

End-User Programmers



	1	2	3	4	5
Cash Flow	(500)	150	125	175	350
Dates	8/17/2017	12/31/2017	12/31/2018	12/31/2019	6/30/2020
NPV	=XNPV(C11,C5:G5,C7:G7)	160			
Discount Rate	10%				

Bloggers

What is End-User Programming?

The key to understanding non-programmers' interaction with computers is to recognize that non-programmers are not simply under-skilled programmers who need assistance learning the complexities of programming. Rather, *they are not programmers at all.* They are business professionals or scientists or other kinds of domain specialists whose jobs involve computational tasks.
— *Nardi et al.*

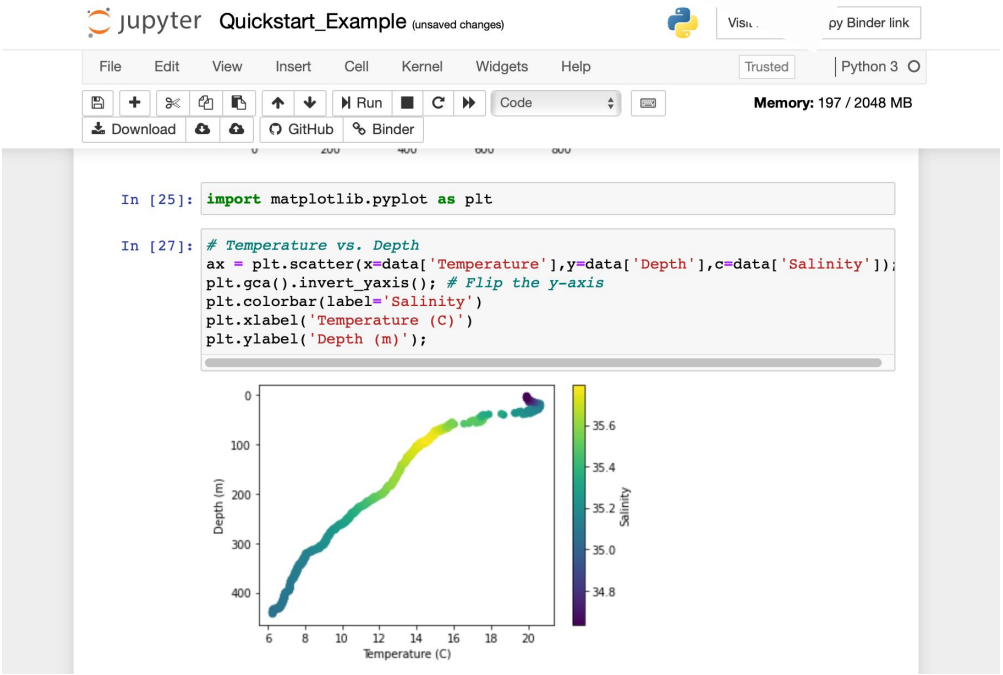
End-user programming is done by a motivated power user who may or may not be a professional programmer, modifying or creating small ad-hoc tools for their own use or perhaps to share with a handful of colleagues.
— *Ink & Switch*

We then define *end-user programming* as "programming to achieve the result of a program primarily for personal, rather public use."
— *Ko et al.*

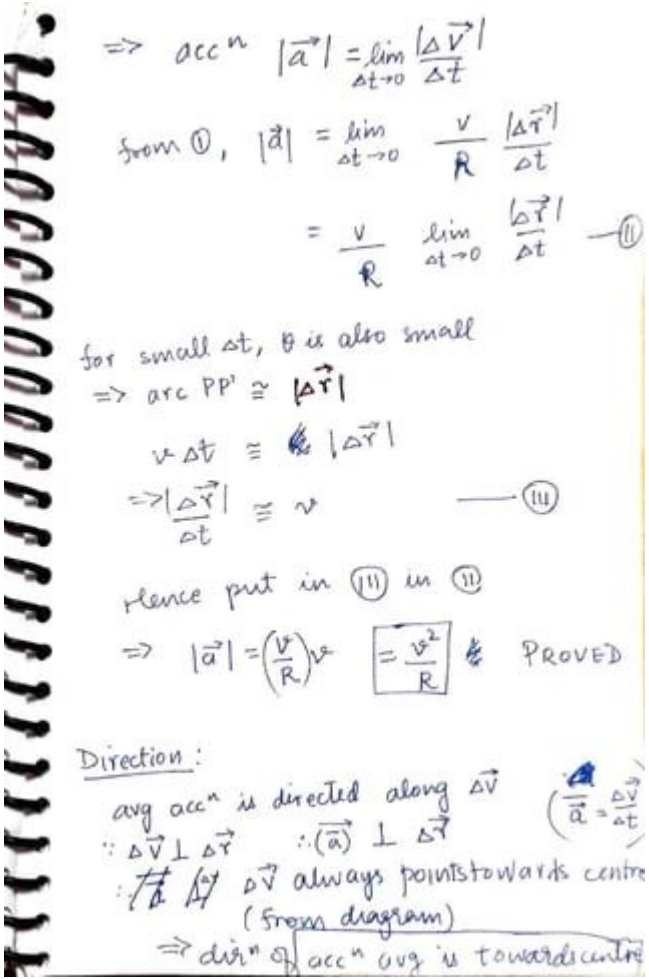
Our Definition

End-user programmers are domain specialists that modify or create programs to achieve the result of a program for their own use or to share with a handful of colleagues.

Scientific Work



Also Scientific Work



Real Scientific Work

jupyter BaSnF4_ELFCAR_slice Last Checkpoint: 04/09/2023 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

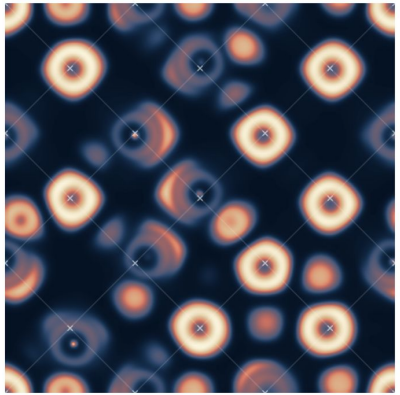
In [1]:
```
from pymatgen.io.vasp import Chgcar
from cmcrameri import cm
import numpy as np

import matplotlib.pyplot as plt
%matplotlib inline
%config InlineBackend.figure_format='retina'
```

In [8]:
```
chgcar = Chgcar.from_file('BaSnF4_ELFCAR.gz')
chgcar.data['diff'].shape
```

Out[8]: (140, 140, 140)

In [ ]:
```
ax = plt.subplot()
plane = chgcar.data['diff'][47]+chgcar.data['total'][47]
shifted_plane = np.roll(plane, 47, axis=0)
ax.imshow(shifted_plane, interpolation="spline16", cmap=cm.lipari)
for shift in range(1,13,2):
    ax.plot([0, (shift+1)/6], [(6-(shift+1))/6, 1],
            transform=ax.transAxes, color="white", alpha=0.15, linestyle='-', linewidth=1)
    ax.plot([0, (shift+1)/6], [(shift+1)/6, 0],
            transform=ax.transAxes, color="white", alpha=0.15, linestyle='-', linewidth=1)
for i in range(7):
    for j in range(7):
        if not (i+j)%2:
            ax.plot(1/6*i, 1/6*j, 'x', color='white',
                    markeredgewidth=None, markersize=4, alpha=0.5, transform=ax.transAxes)
plt.axis('off')
plt.tight_layout()
plt.savefig('BaSnF4_elfcar.pdf', dpi=300)
plt.show()
```



What do you think?

Which programming languages/frameworks did scientists report using for their work? (select all that apply)

- ❑ Vanilla Javascript
- ❑ Python
- ❑ R
- ❑ Vega-Lite
- ❑ D3

Join at menti.com | use code 8626 0629

What do you think?

Which programming languages/frameworks did scientists report using for their work? (select all that apply)

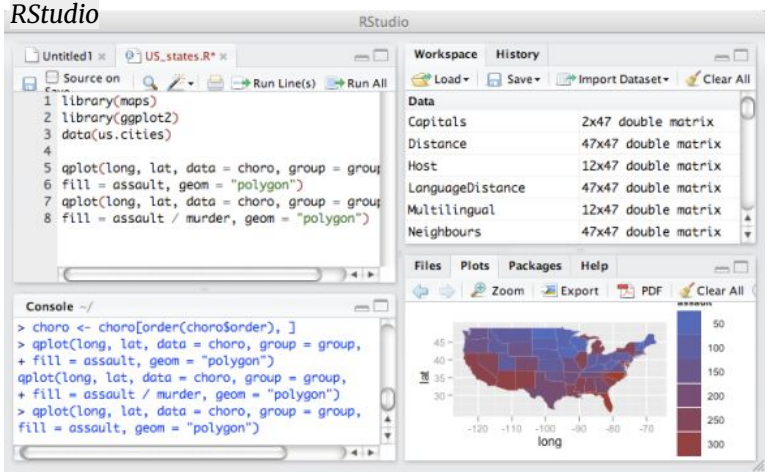
- ❑ Vanilla Javascript
- ❑ Python
- ❑ R
- ❑ Vega-Lite
- ❑ D3

Join at menti.com | use code 8626 0629

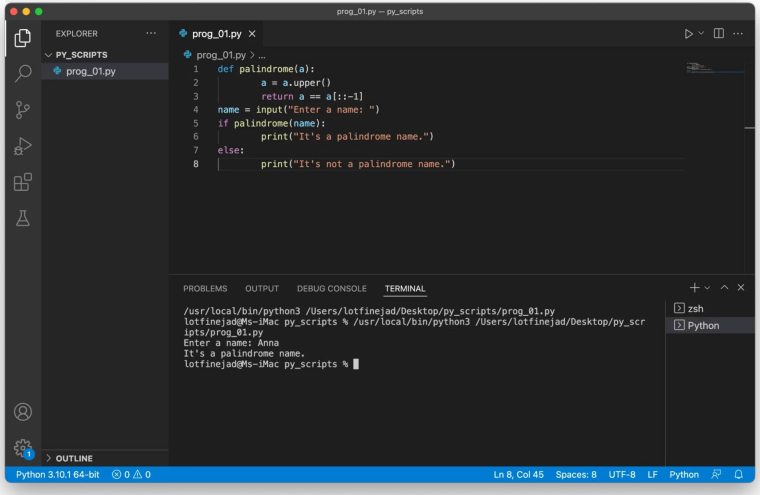
R & Python

- ❌ *Vanilla JS*
- ❌ *D3*
- ❌ *Vega-Lite*

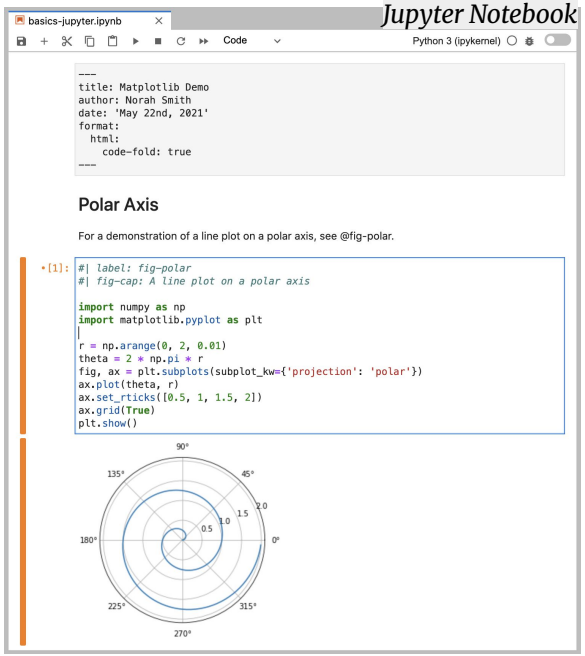
RStudio



VS Code

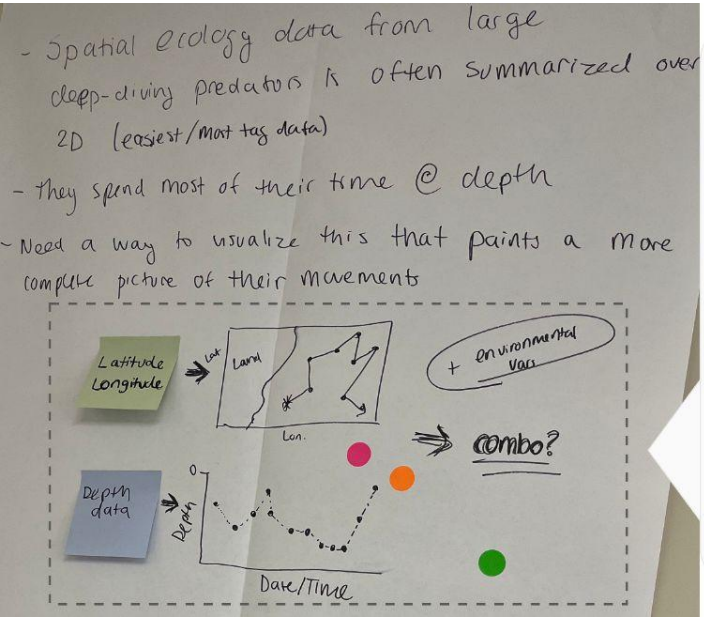


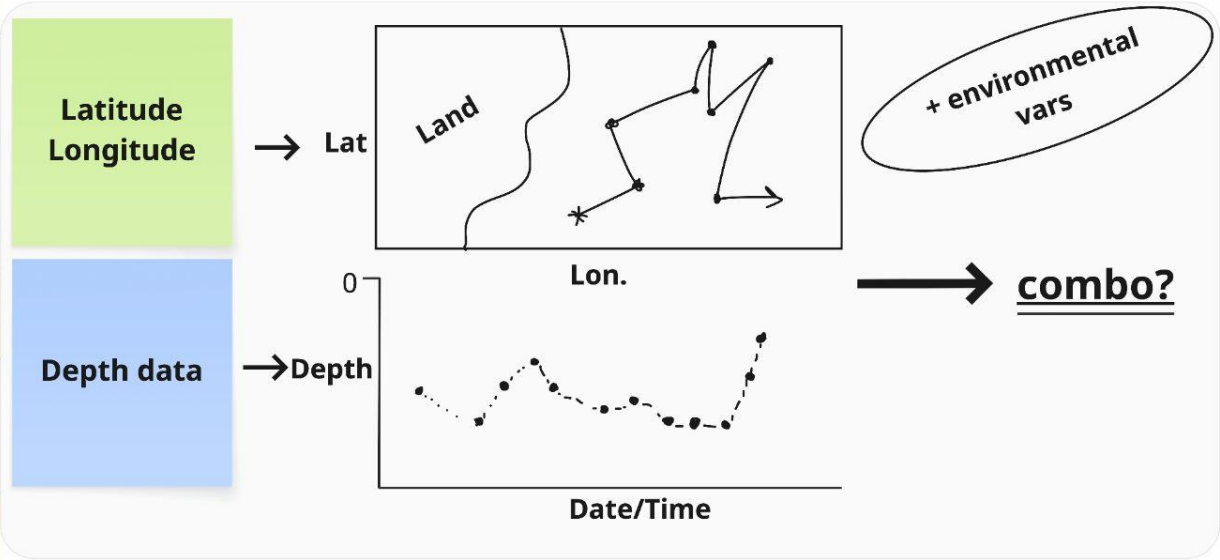
Jupyter Notebook



Google Colab

Computational tools & practices limit what scientists can express



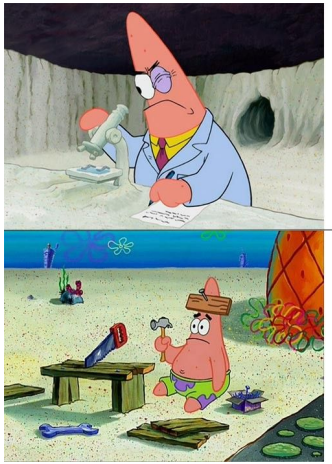


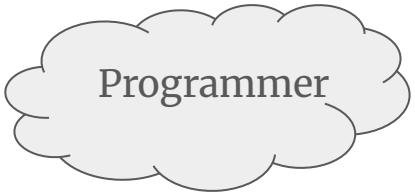
"I think one of our big issues is how [large predatory fish like tuna and sharks] spend so much time at depth that ideally we want a cool way to show their tracks at depth and how they're interacting with oceanographic variables throughout the actual tracks that they're taking"

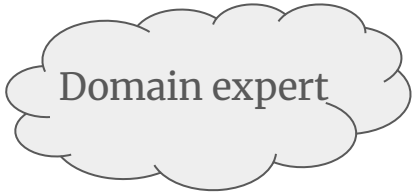
Workshop Excerpt, February 2025

Scientific Programming

- **Many reasons to code**: Design RNA structures (Nouwens et al.), model climate data (Ziegler et al.), analyze ecology data (Brousil et al.)
- **But, aren't formally taught**: Little instruction, and instead learn from peers or through self-study (Hannay et al.)
- → **Disempowerment**: Variety of languages, debugging difficult (Nouwens et al.)
- **+ Steep learning curve**: Tools and languages are technically complex (Berger et al.)

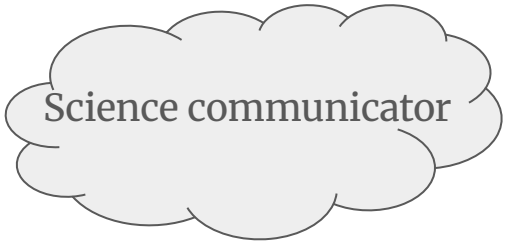


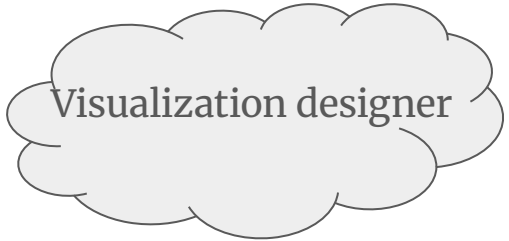




“I don’t have much of a computer science background...” [and]

“I’m supposed to be an expert in too many things.”

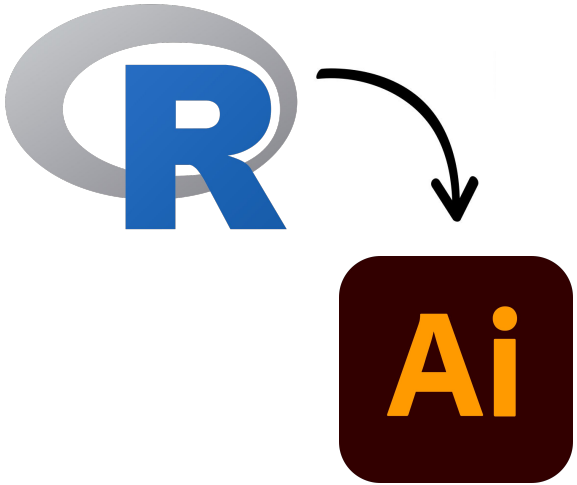


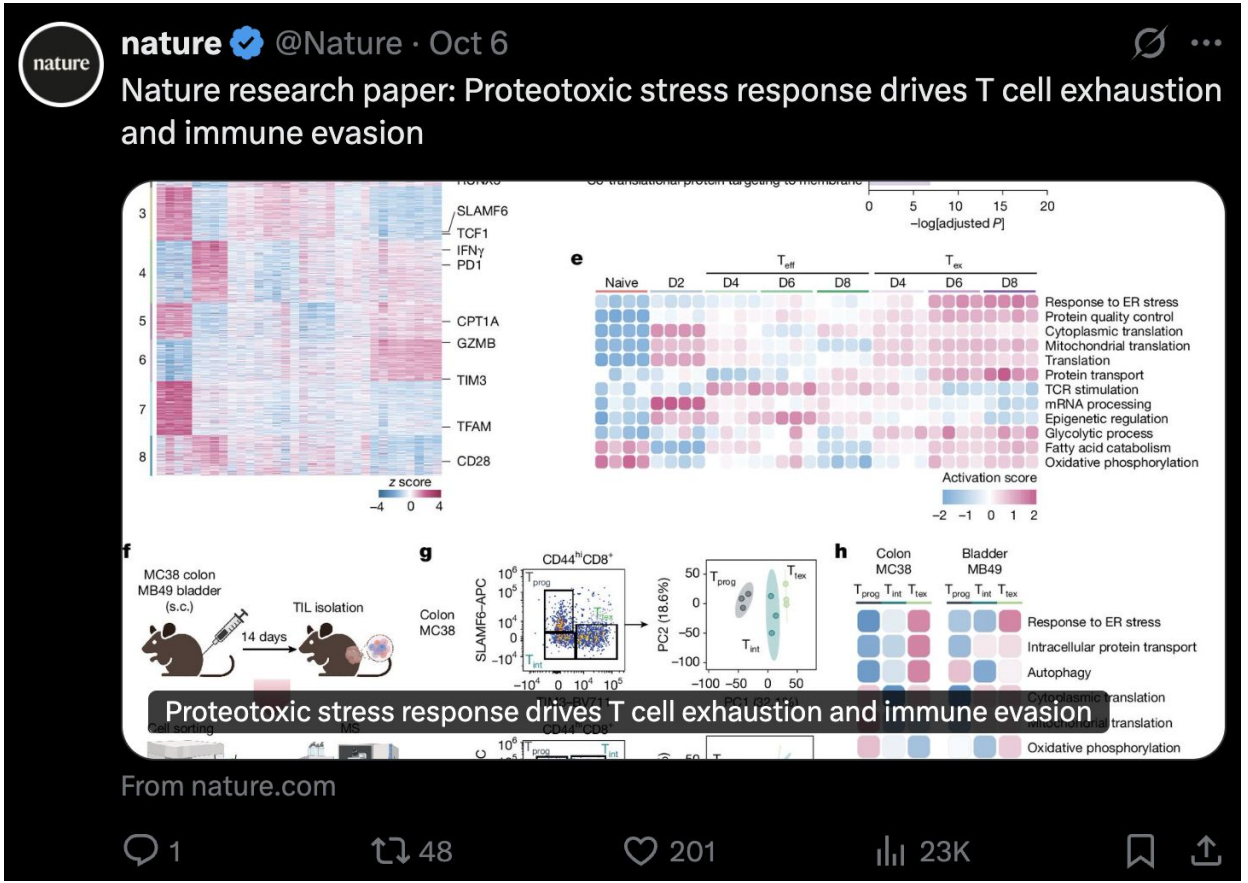


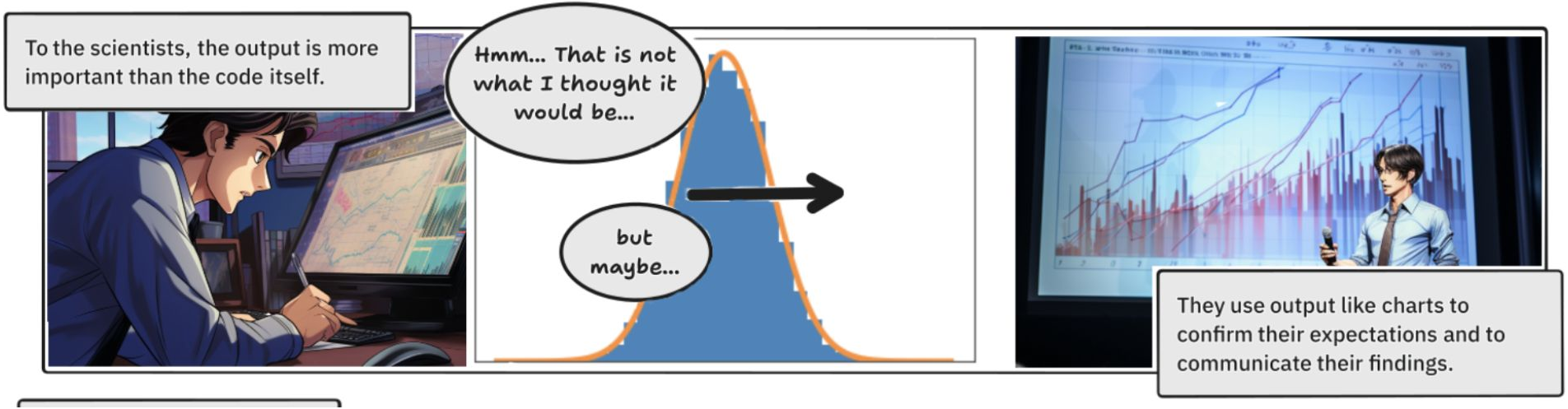
Prototype Review Excerpt, July 2025

“[figures] generate buzz”

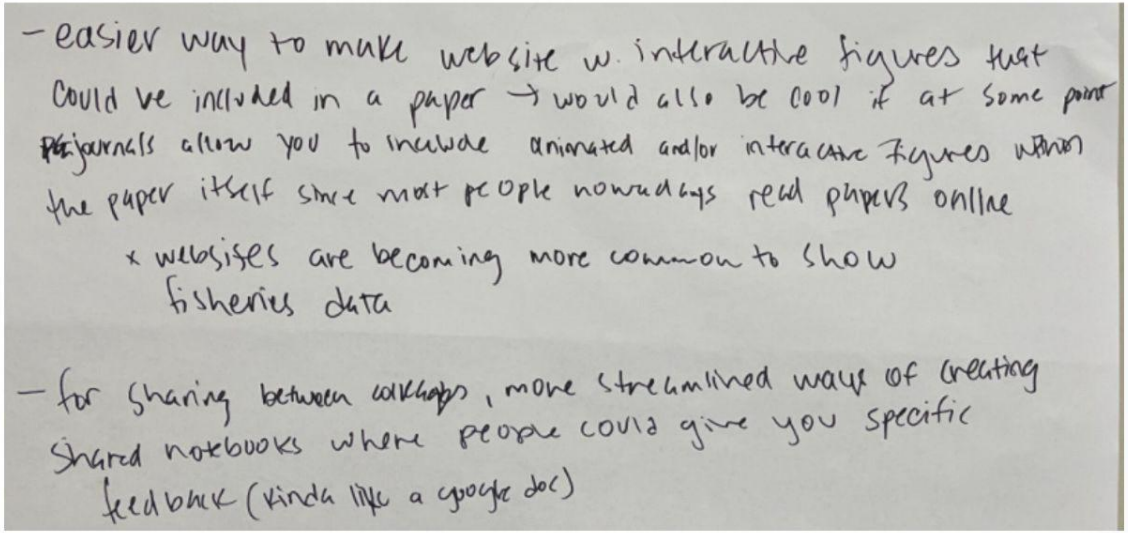
Prototype Review Excerpt, July 2025







Interview/Observation/Code Review Excerpt, October 2023

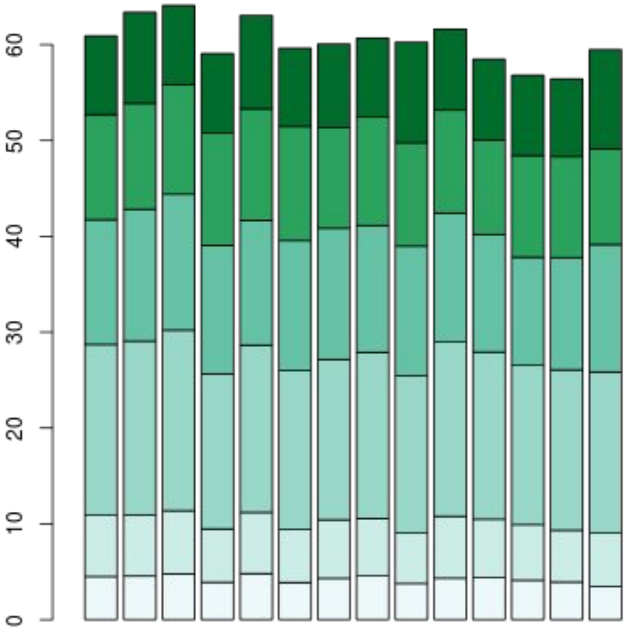


- easier way to make website with interactive figures that could be included in a paper -> would also be cool if at some point journals allow you to include animated and/or interactive figures within the paper itself since most people nowadays read papers online
    - websites are becoming more common to show fisheries data
- for sharing between colleagues, more streamlined ways of creating shared notebooks where people could give you specific feedback (kind of like a google doc)

"It would be really cool to have an interactive component where people could, when you're reading a paper, zoom in on a graph, click the points and stuff. I think that would solve a lot of the issues, either with the volume of data that we have, or the nuance of each data point that you lose in the 2D printed version that they're still publishing"

Workshop Excerpt, February 2025

“[Plotting] at a higher taxonomic rate like phylum or class... would put me and my dog in the same color and we know that me and my dog have different niches... I don’t want to lose my fine grained data at the taxonomy level”



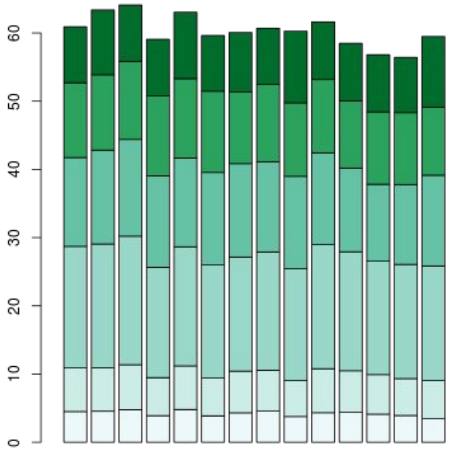
← human & dog

human →

← dog

Workshop Excerpt, February 2025

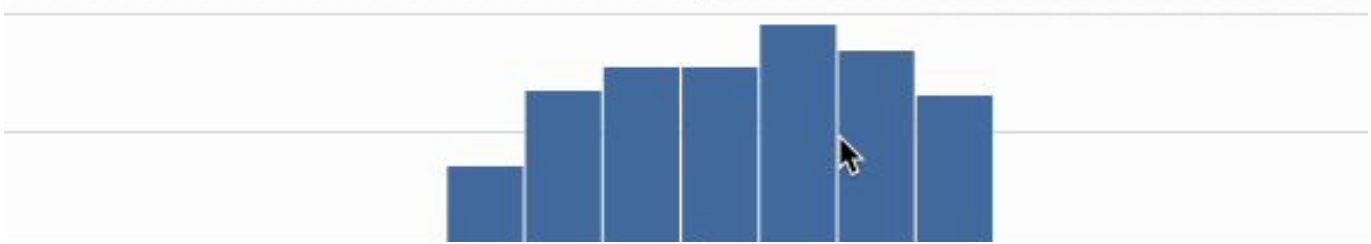
“[Plotting] at a higher taxonomic rate like phylum or class... would put me and my dog in the same color and we know that me and my dog have different niches... I don’t want to lose my fine grained data at the taxonomy level”



← human & dog

human →

← dog



Workshop Excerpt, February 2025

What is interactivity?

Potluck

Plant Watering Tracker

Plants

🌱 Fiddle leaf: every 6 days, last watered on 8/19/2022

🌱 Montsera: every 4 days, last watered on 8/21/2022

🌱 Yuzu tree: every 5 days, last watered on 8/17/2022

- The Yuzu tree needs some holes in the pot so that water can drain.

🌱 Pine Bonsai: every 4 days, last watered on 8/19/2022

Ten Brighter Ideas

Premise

Suppose 85% of US households always turned off lights in unoccupied rooms.

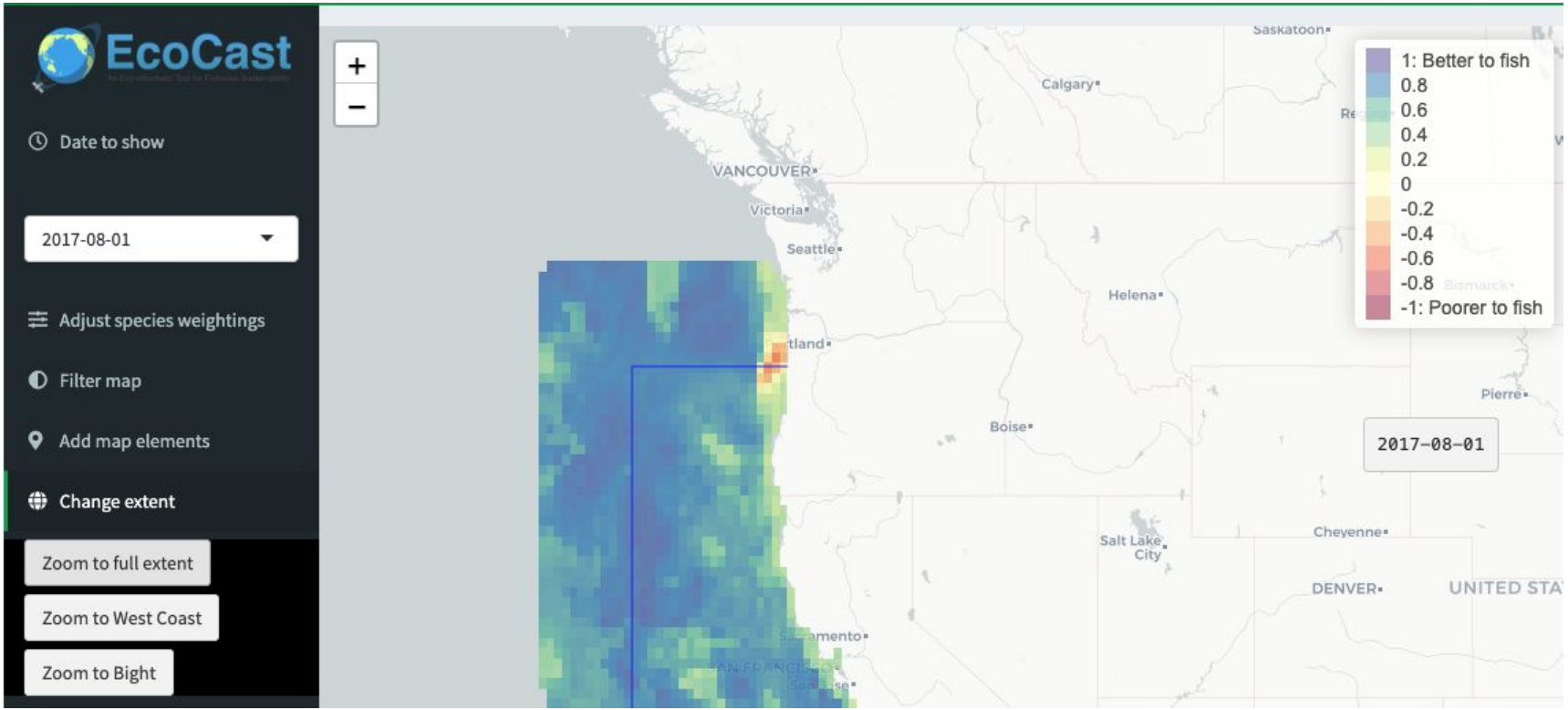
Result

This would save **41.8 TWh** per year.

Interactivity in science

"...more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble."
– *Engelbart*

“The spectacle metaphor of instruments is replaced by a metaphor where the instruments become a material playground that provides us with a way to learn a lot about the world and about phenomena produced by these instruments. Observation as a source of empirical knowledge is extended to include doing, by interacting and intervening with the world through our instruments.” - *Boon*



What about Zooming in on the San Francisco Bay?

So what can we do?

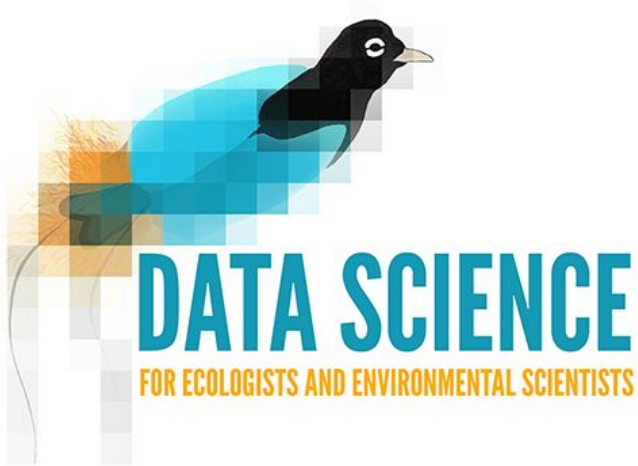
What do you think?

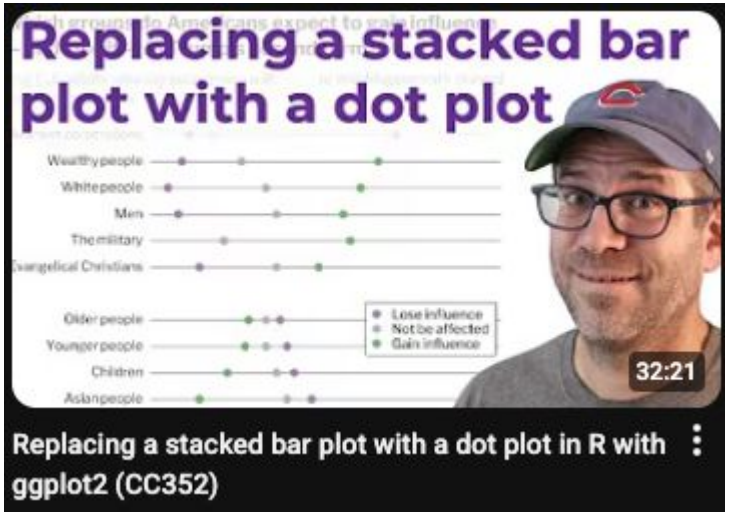
How can we give scientists the ability to create interactive elements that they could potentially then use to enhance their analysis, communication or other work activities?

Join at menti.com | use code 8626 0629

So what can we do?

We can provide scientists training...





So what can we do?

We can provide scientists training...

Or dedicated staff...

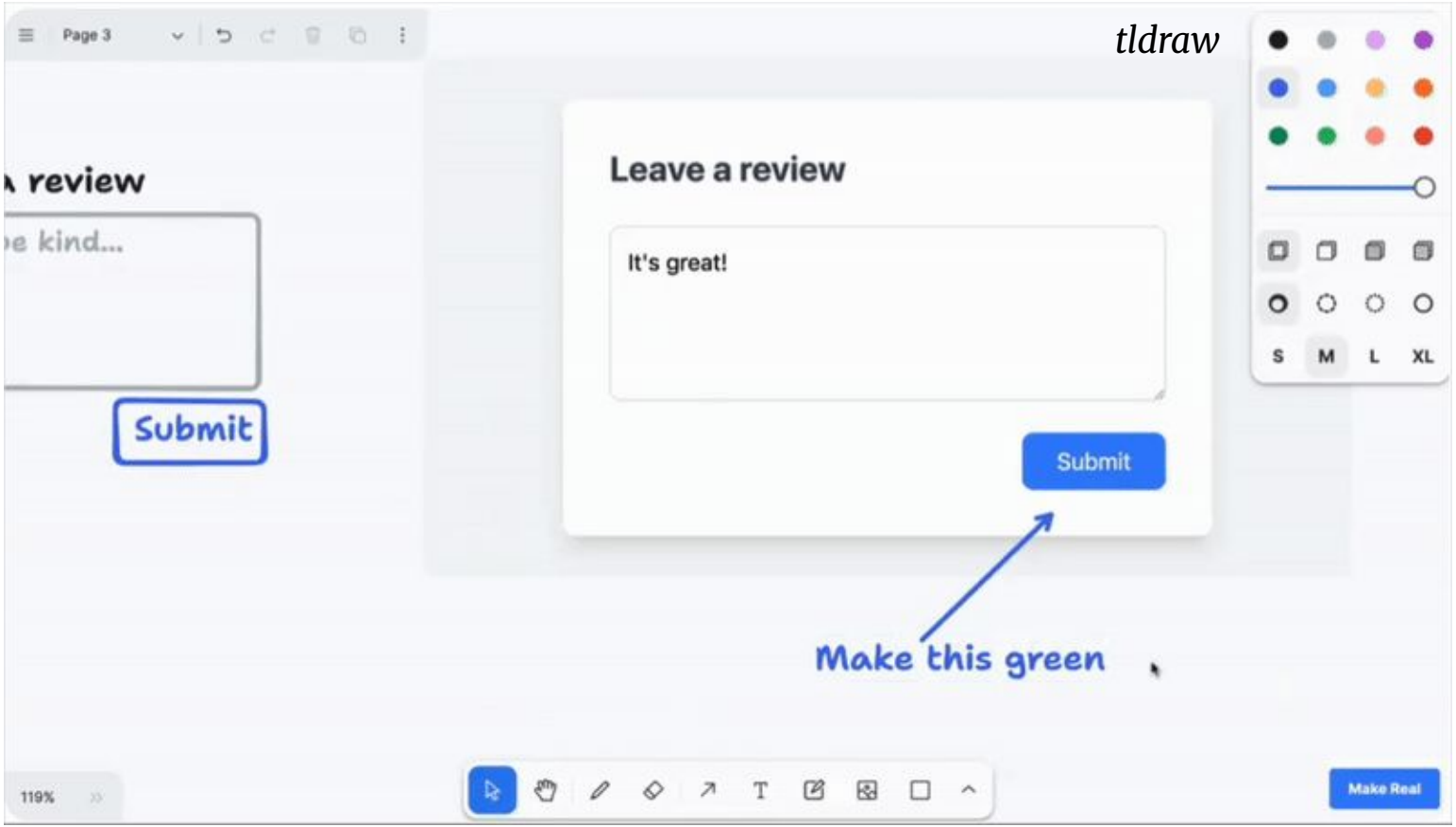
> “There's a person in our lab, she set it up so that we can run R through the terminal.” *Prototype Review, July 2025*

So what can we do?

We can provide scientists training...

Or dedicated staff...

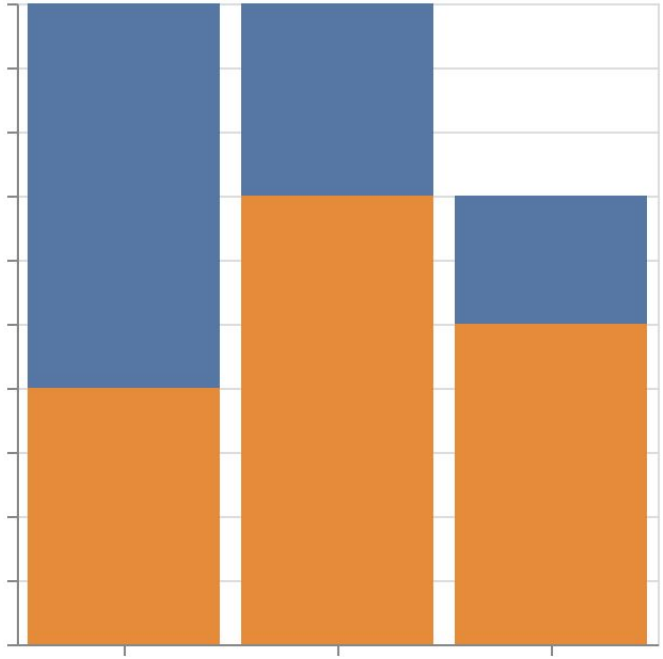
Or improve tooling

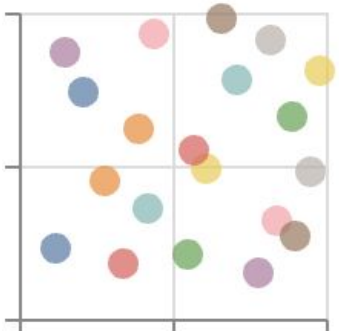

tldraw
Page 3
review
e kind...
Submit
Leave a review
It's great!
Submit
S
M
L
XL
Make this green
119%
Make Real

“What is the best way to display data? Should I separate it... How do I know which [visualization] to pick?”

“Sometimes [visualizing data] can be very overwhelming. What do we include? How do I group things?”

Workshop Excerpt, February 2025





Here is the R code that performs these steps:

R Copy code

```r
# Load required libraries
library(osmdata)
library(mapview)

# Get the bounding box for Pittsburgh
pittsburgh_bbox <- getbb("Pittsburgh, Pennsylvania, USA")

# Create an Overpass query for pharmacies in Pittsburgh
query <- opq(bbox = pittsburgh_bbox) %>%
  add_osm_feature(key = "amenity", value = "pharmacy")

# Extract the data as an sf object
pharmacies_sf <- osmdata_sf(query)

# Visualize the pharmacies on a map
mapview(pharmacies_sf$osm_points, zcol = "name")
```

Vibe coding with AI agents is not for scientists

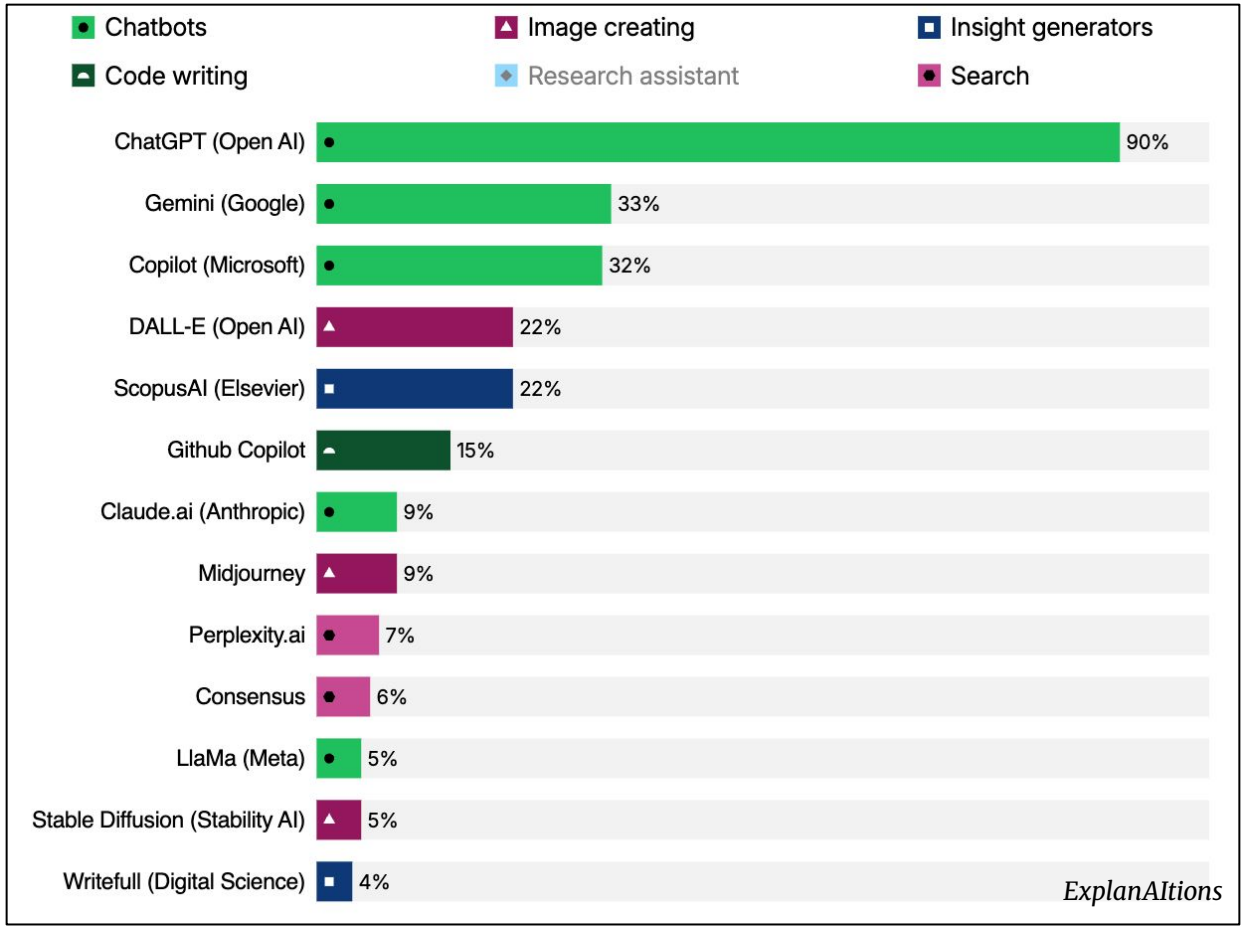
RSTATS GENAI

What do you think?

Based on a 2024 survey of researchers, which AI tool is most commonly used?

- ❏ Github Copilot (code)
- ❏ Midjourney (images)
- ❏ ChatGPT

Join at menti.com | use code 8626 0629


Chatbots
Image creating
Insight generators
Code writing
Research assistant
Search
ChatGPT (Open AI) 90%
Gemini (Google) 33%
Copilot (Microsoft) 32%
DALL-E (Open AI) 22%
ScopusAI (Elsevier) 22%
Github Copilot 15%
Claude.ai (Anthropic) 9%
Midjourney 9%
Perplexity.ai 7%
Consensus 6%
LlaMa (Meta) 5%
Stable Diffusion (Stability AI) 5%
Writefull (Digital Science) 4%
ExplanAItions

How does scientific work change when scientists author interactive elements?

Defining Interactivity

Jupyter Notebook

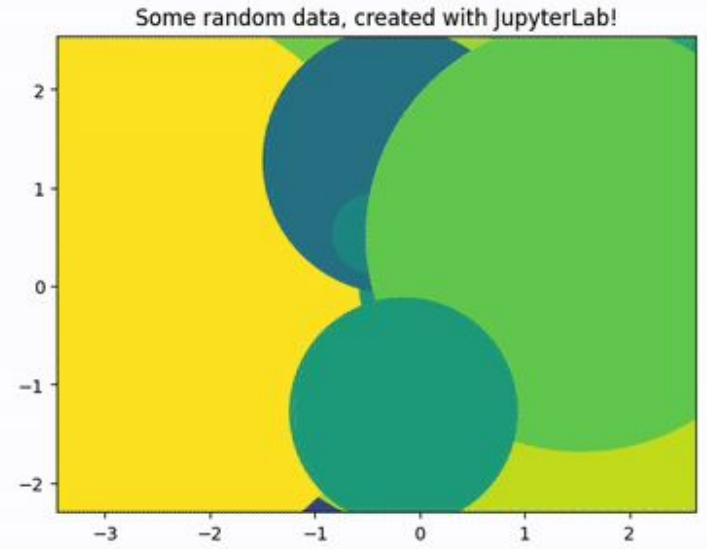
```
[10]: from matplotlib import pyplot as plt
      import numpy as np

      x, y, scale = np.random.randn(3, 100)


      fig, ax = plt.subplots()

      #ADJUST THIS VALUE
      size=50000

      ax.scatter(x=x, y=y, c=scale, s=np.abs(scale)*size)
      ax.set(title="Some random data, created with JupyterLab!")
      plt.show()
```



Ten Brighter Ideas

Premise

Suppose 85% of US households always turned off lights in unoccupied rooms.

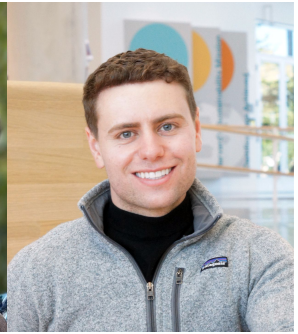
Result

This would save **41.8 TWh** per year.

- Ease of creation
- Ease of using
- Fossilized
- Malleable

Clemens N Klokmose | Magnus Madsen | Matthew Lutze

Arvind Satyanarayan | Josh Pollock | Dylan Wooton



Enhancing Scientific Work

Caroline Berger

PhD Candidate

caroline.berger@cs.au.dk
carolineberger.github.io

Questions/Comments:

Join at menti.com | use code 8626 0629

References

Nardi et al. 1990. The spreadsheet interface: A basis for end user programming. In IFIP TC13.
Ink & Switch. 2019. End-user Programming. Blog Post.
Ko et al. 2011. The state of the art in end-user software engineering. ACM CSUR.
Hannay et al., "How do scientists develop and use scientific software?," 2009 ICSE Workshop
Engelbart. 1962. Augmenting human intellect: A conceptual framework.
Boon. 2015. The scientific use of technological instruments. Philosophical perspectives.
https://www.seascapemodels.org/posts/2025-06-08-vibe-coding-not-for-scientists/
https://www.wiley.com/en-us/about-us/ai-resources/ai-study/

educational resources
https://ourcodingclub.github.io/
https://riffomonas.org/
https://software-carpentry.org/

tools
https://www.inkandswitch.com/potluck/demo/?openDocument=plant-watering
https://worrydream.com/TenBrighterIdeas/
https://myoceanlearn.marine.copernicus.eu/
https://tldraw.substack.com/p/make-real-the-story-so-far
https://coastwatch.pfeg.noaa.gov/ecocast/explorer.html

images
https://www.whoi.edu/ocean-learning-hub/multimedia/dna-detective/
https://www.ocearch.org/tracker/detail/max
https://share.google/images/ozbWTT8aOEdvgUyln
https://www.sun-sentinel.com/2024/03/22/teens-get-up-close-and-personal-with-wild-sharks-to-encourage-careers-in-science/
https://sharkresearch.earth.miami.edu/
https://www.reddit.com/r/physicsmemes/comments/y3p8c3/somebody call my momma/
https://x.com/deniseearle/status/1465971899415023617
https://www.r-bloggers.com/2023/04/how-to-r-code-faster-with-chatgpt/
https://pythoninchemistry.org/ch40208/notebooks introduction/what is a jupyter notebook.html
https://www.slideshare.net/slideshow/physics-derivations-class-xi-ncert-all-chapters-and-important-questions/266808400
https://quarto.org/docs/tools/jupyter-lab.html
https://www.dataquest.io/blog/how-to-set-up-visual-studio-code/
https://amitness.com/posts/google-colab-tips
https://statisfaction.wordpress.com/2011/04/29/rstudio-is-good-for-you/