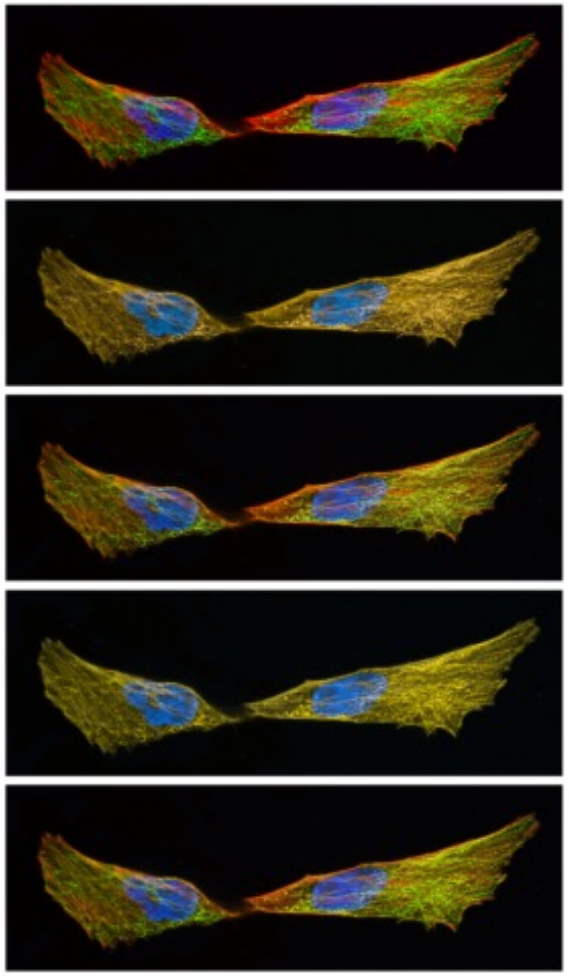
Access & Accessibility

Communicating with data: visualizations and more

Andy Boughton
Center for Statistical Genetics

Scott-Boehnke Group Meeting
June 17, 2021

Are your visualizations being seen?

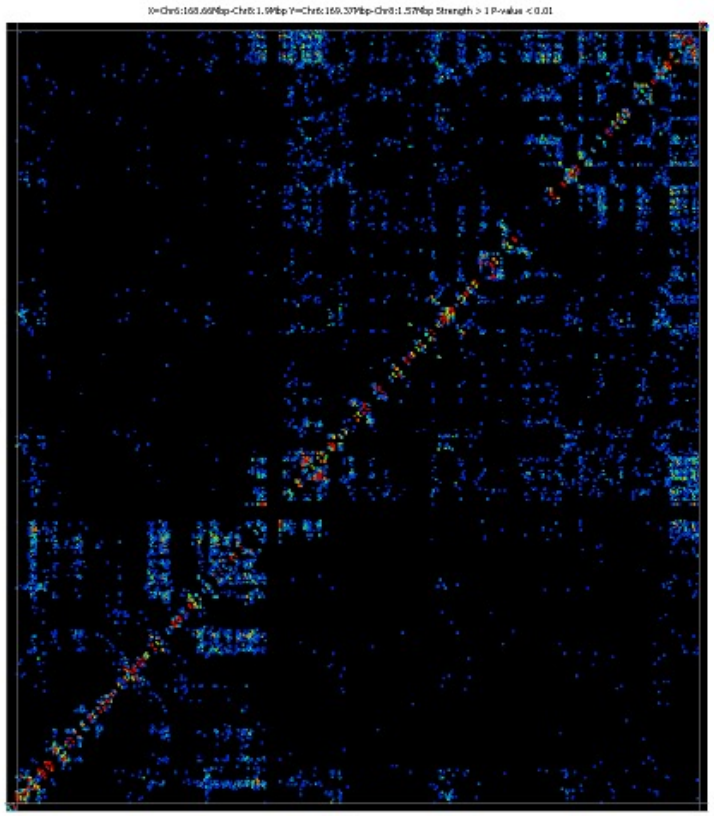


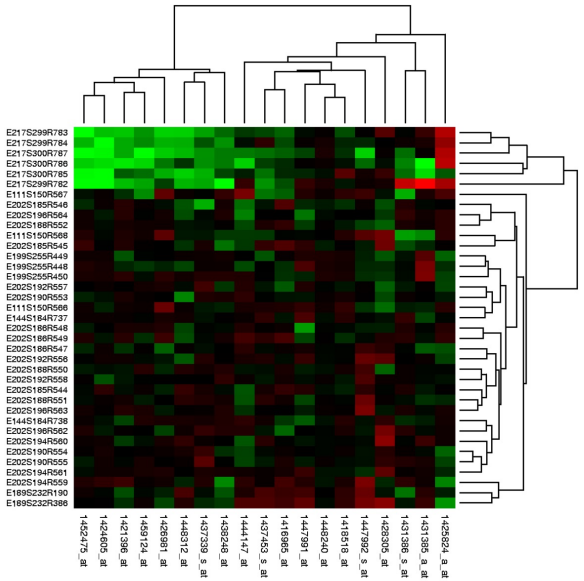
"The purpose of scientific visualization is to graphically illustrate scientific data to enable scientists to understand, illustrate, and glean insight from their data."

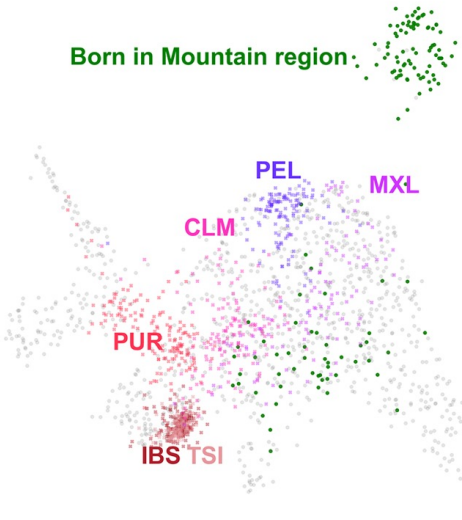
Ask: how does your audience see the world?

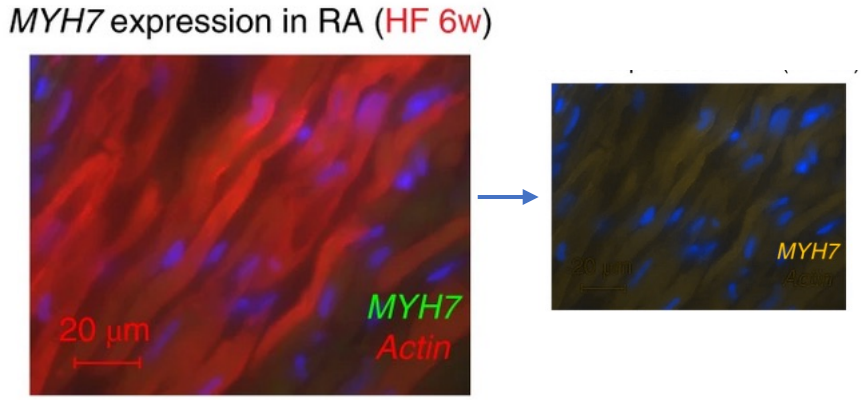
https://www.ascb.org/science-news/how-to-make-scientific-figures-accessible-to-readers-with-color-blindness/
https://en.wikipedia.org/wiki/Scientific_visualization

This happens a lot









[1] “HiC Heatmap”: https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/Help/3%20Visualisation/3.2%20Figures%20and%20Graphs/3.2.12%20The%20HiC%20Heatmap%20Plot.html
[2] “Gene expression heatmap”: https://en.wikipedia.org/wiki/Heat_map#/media/File:Heatmap.png
[3] UMAP plots of populations structure (2020): https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008432
[4] It’s annual review season! https://www.nature.com/articles/s41588-018-0171-3/figures/4

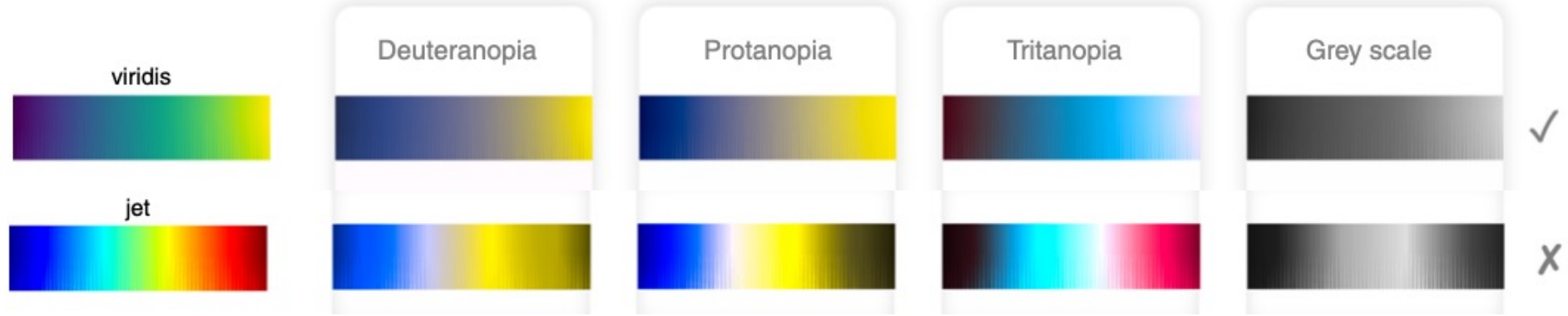
Forms of colorblindness

- Three basic categories, each with several forms [1]:
  - Red-green: ~8% of males and ~0.5% of females of N. Eur descent
  - Blue-yellow
  - Complete
- Heritable and X-linked (more common in males)
  - *"If a submitted manuscript happens to go to three male reviewers of Northern European descent, the chance that at least one will be color blind is 22 percent." [2]*
- There is a continuum: some people have *partial* loss

[1] https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness/types-color-blindness
[2] Wong, B. Points of view: Color blindness. *Nat Methods* **8**, 441–441 (2011). (link)

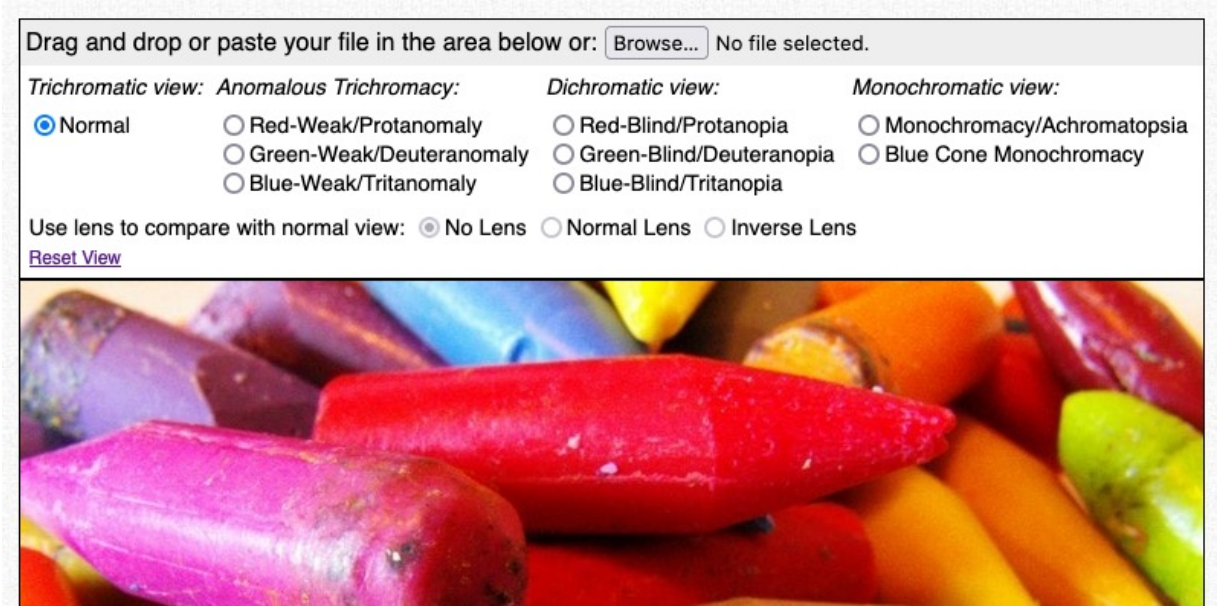
Accessibility Helps Everyone

- Often, accessibility has benefits beyond the intended audience
  - If it works without color, then it works in black and white (or on old projectors, bad technology setups, etc)
  - Good color choices can aid interpretation and better conclusions



Experiencing colorblindness

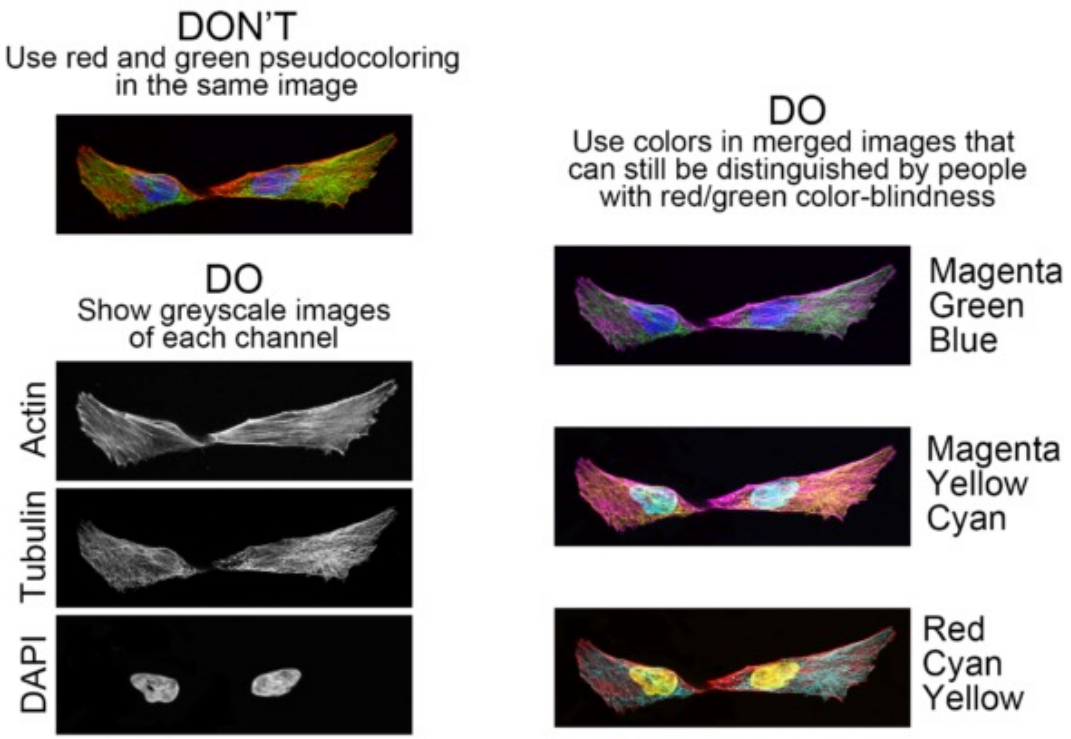
- Online “color-blind simulator” tools let you try any image
- Firefox web browser has a built-in tool for web pages



Tool: https://www.color-blindness.com/coblis-color-blindness-simulator/

Improving colors: content matters

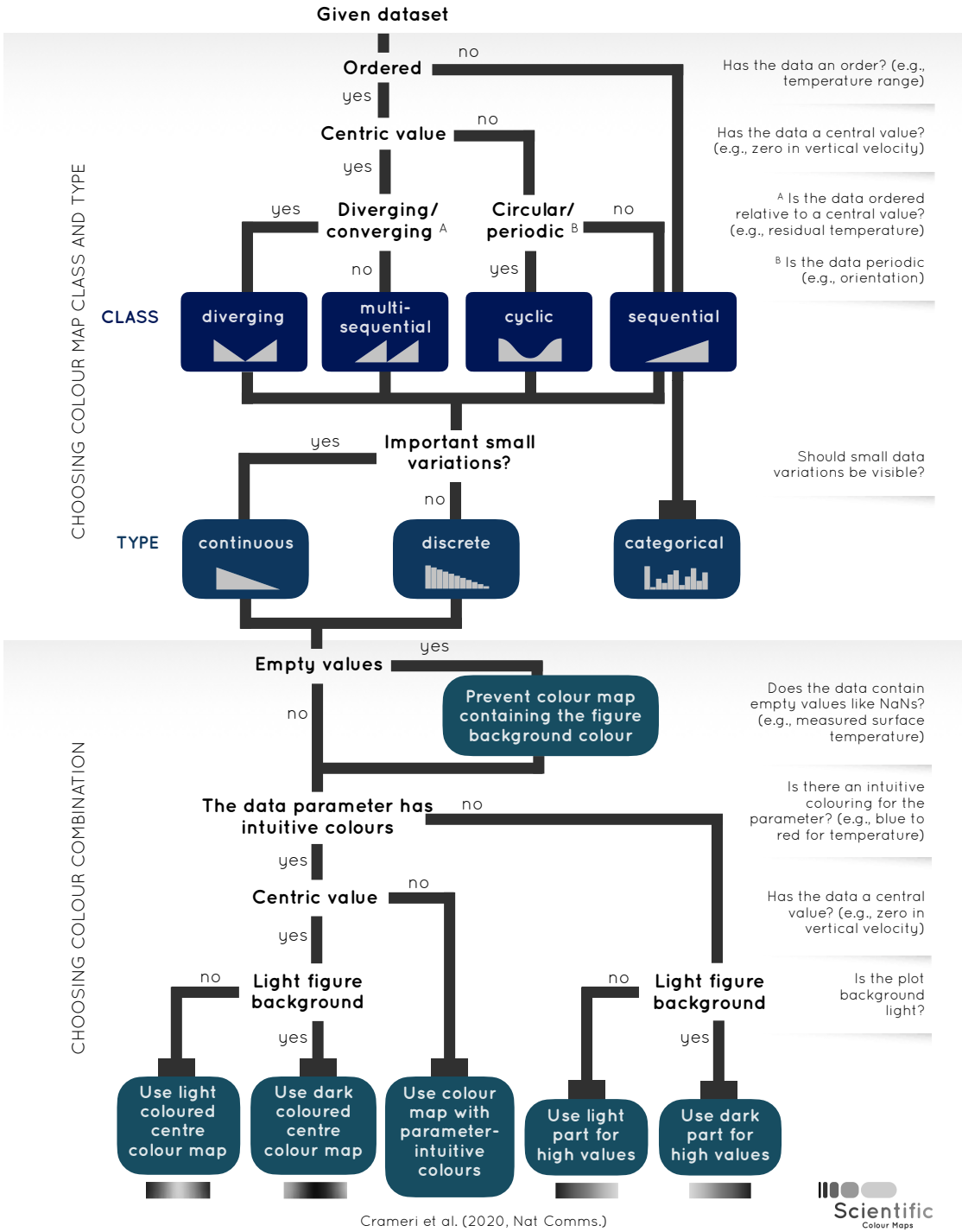
- Consider what you are using color to convey
  - Combined image or separate pieces of information?
- Choose a color scheme that reflects the type of data
  - Continuous
  - Sequential
  - Categorical



[1] https://www.ascb.org/science-news/how-to-make-scientific-figures-accessible-to-readers-with-color-blindness/
[2] "Semiology of graphics": https://karlsluis.medium.com/before-tufte-there-was-bertin-63af71ceaa62
[3] "Grammar of Graphics": https://www.amazon.com/Grammar-Graphics-Statistics-Computing/dp/0387245448

Choosing a color scheme

- It should reinforce intuitions about data
  - Items adjacent in a sequence should be more similar in color
  - Unrelated categories should be as distinct as possible
  - Trends in brightness should follow trends in values



[1] Crameri, F., Shephard, G. E. & Heron, P. J. The misuse of colour in science communication. *Nat Commun* **11**, 5444 (2020). (link)

A  E

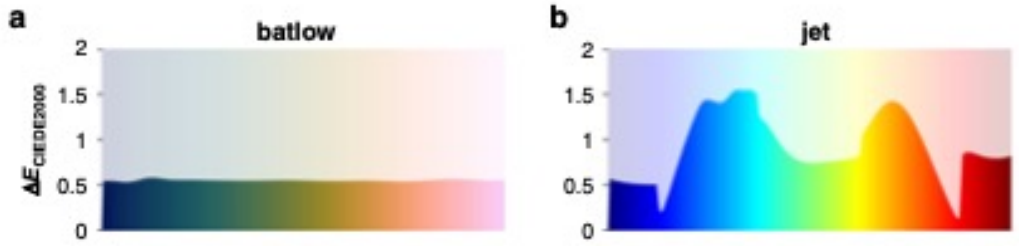
Categorical

0  1

Sequential

-1  +1

Diverging



Does brightness increase consistently as values change?

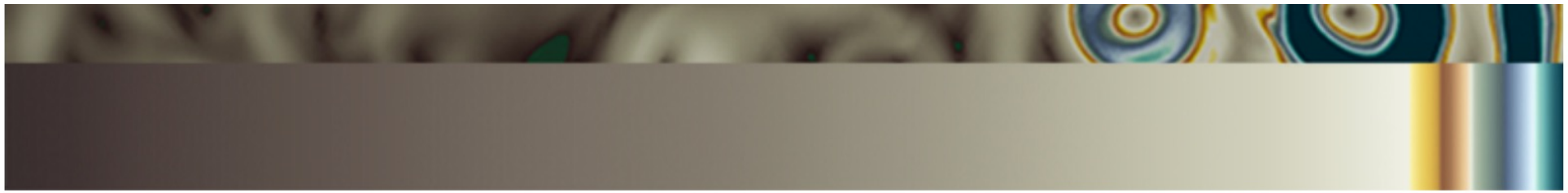


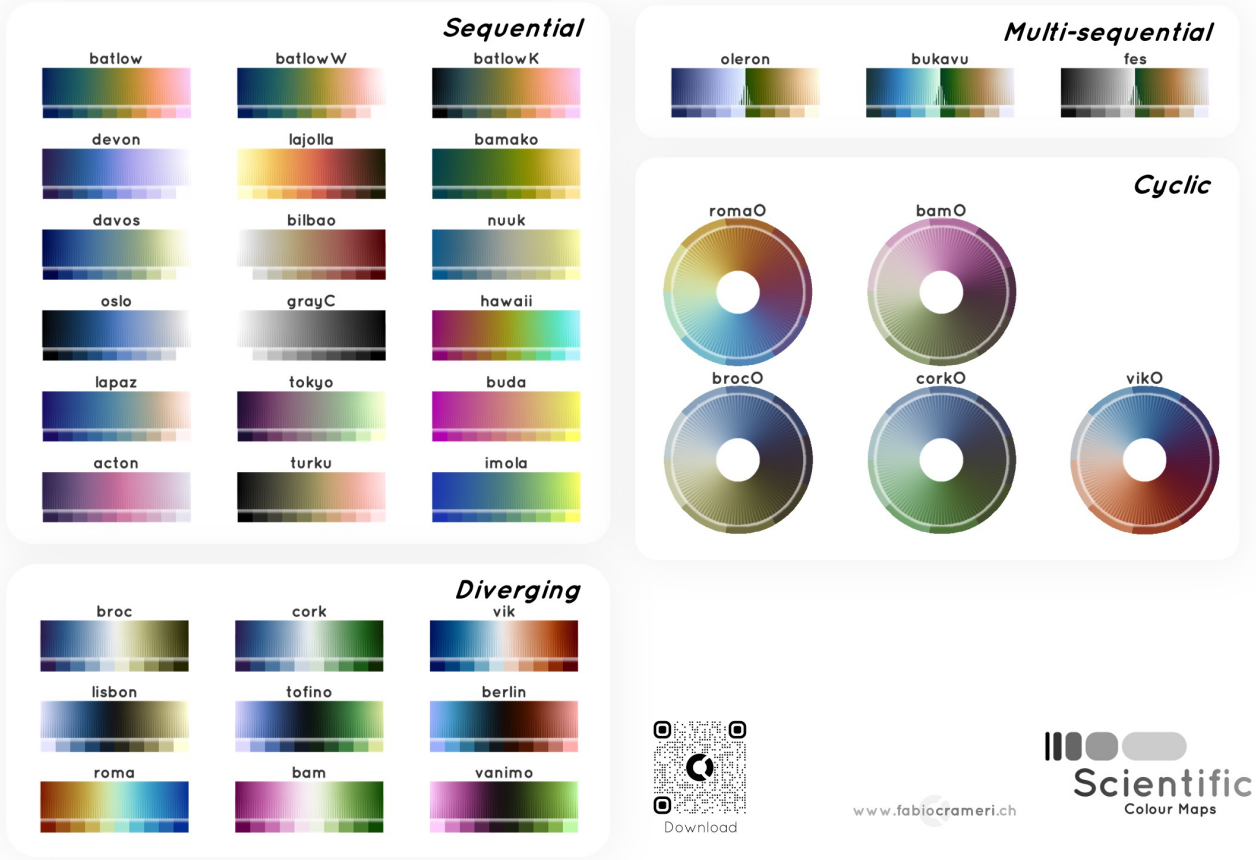
Fig. 3. Comparison of the same data portrayed with a desaturated rainbow colormap (top), which is often used to increase visual detail, and with a colormap designed to provide detail in outlying data ranges (bottom). The specialized colormap provides detailed information about the kinetic energy within eddies in the Agulhas Current, a major current in the Indian Ocean, whereas the rainbow colormap simply identifies the eddies. Credit: Graphic created by Francesca Samsel with data processed and provided by M. Petersen, LANL, using MPAS-Ocean

https://colorbrewer2.org/

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

Color schemes: many options

- Premade color schemes
  - Popular (maps): https://colorbrewer2.org/
  - Scientific (various): https://www.fabiocrameri.ch/colourmaps/
  - Accessible-focused: https://davidmathlogic.com/colorblind/
  - Turbo (improved rainbow): https://ai.googleblog.com/2019/08/turbo-improved-rainbow-colormap-for.html
  - "Other resources": https://www.kennethmoreland.com/color-advice/
  - http://soliton.vm.bytemark.co.uk/pub/cpt-city/
- Tools to generate
  - Free-form selection: https://color.adobe.com/create/color-accessibility
  - Generate n random colors based on set priorities: http://vrl.cs.brown.edu/color
- Thematic: https://github.com/karthik/wesanderson

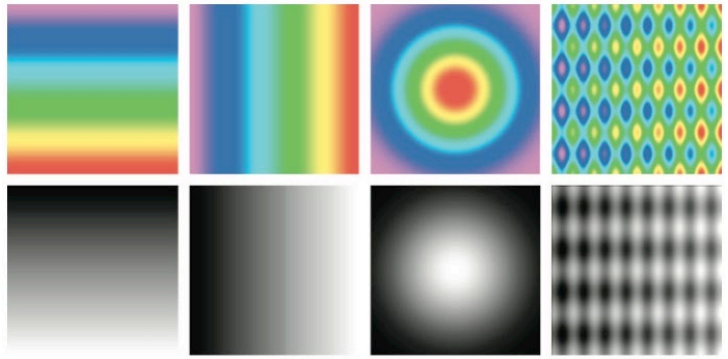


[1] Crameri, F., Shephard, G. E. & Heron, P. J. The misuse of colour in science communication. *Nat Commun* **11**, 5444 (2020). (link)

Seriously, don't trust the defaults

MS Excel pre-2007 (kids, ask your parents)

People *published* these graphs



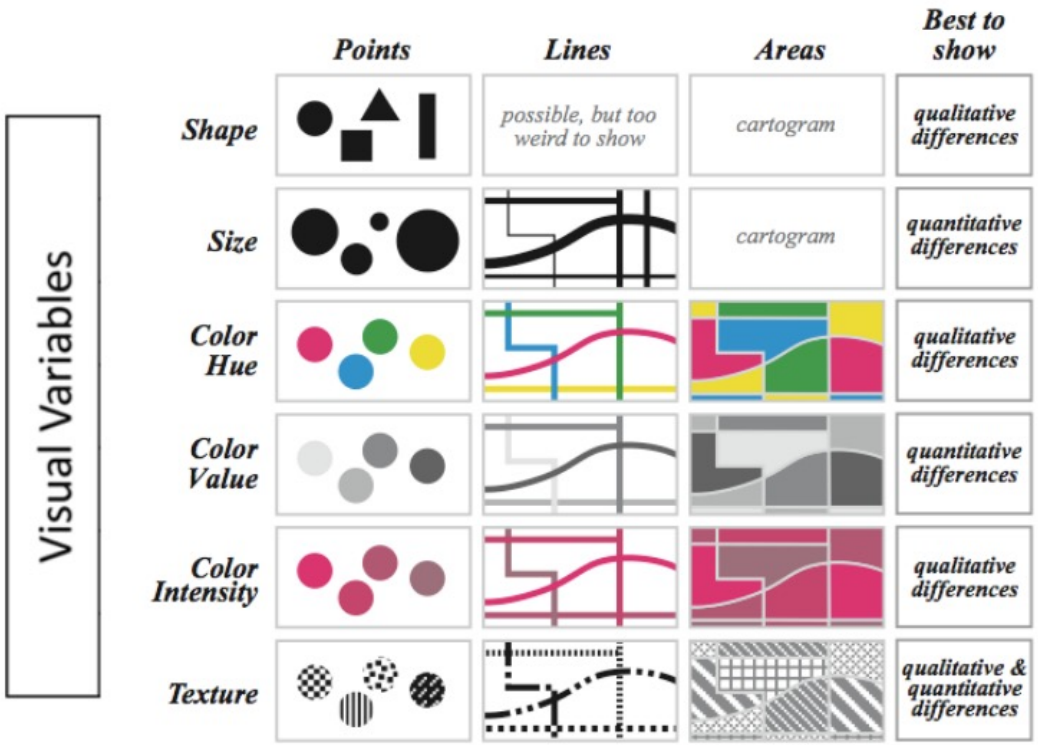
"Jet": MATLAB pre-2014 and Matplotlib pre-2016

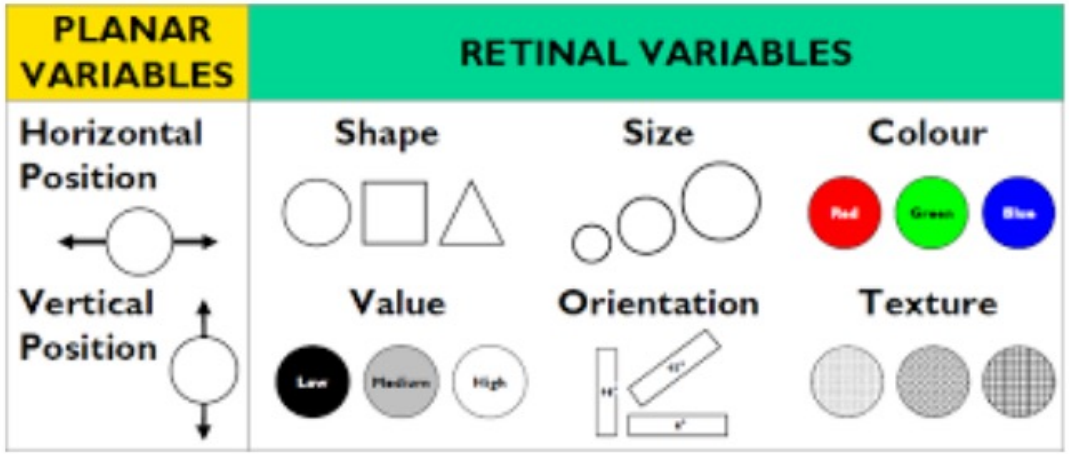
^ "Matlab's default 'Jet' colormap will obscure real patterns while showing you imaginary ones" (commentary)

Borkin et al: "The rainbow scheme (A) was preferred by most since it is what they are accustomed to viewing… a perceptually appropriate color map leads to fewer [clinical] diagnostic mistakes than a rainbow color map… task completion times were nearly twice as slow with the rainbow color map " [1]

[1] Borkin, M. A. *et al.* Evaluation of artery visualizations for heart disease diagnosis. *IEEE Trans Vis Comput Graph* **17**, 2479–2488 (2011). (link)

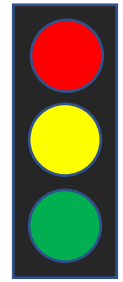
Go Beyond Color



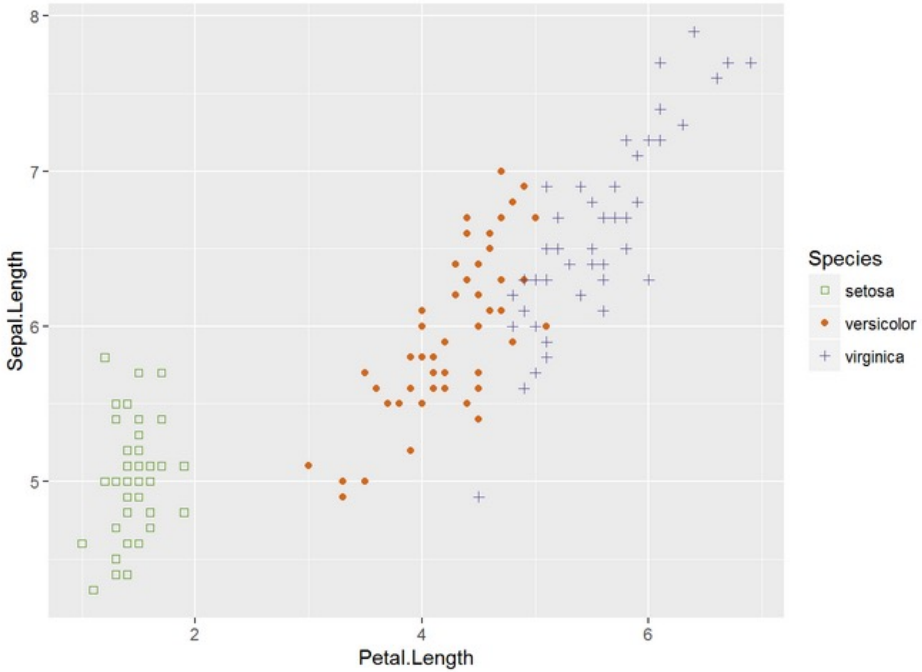


http://euclid.psych.yorku.ca/www/psy6135/lectures/Language.pdf

Do Repeat Yourself



- Convey the same information in multiple ways when possible:
  - In a simple plot, can differentiate points by multiple encodings: color, shape, size, position…

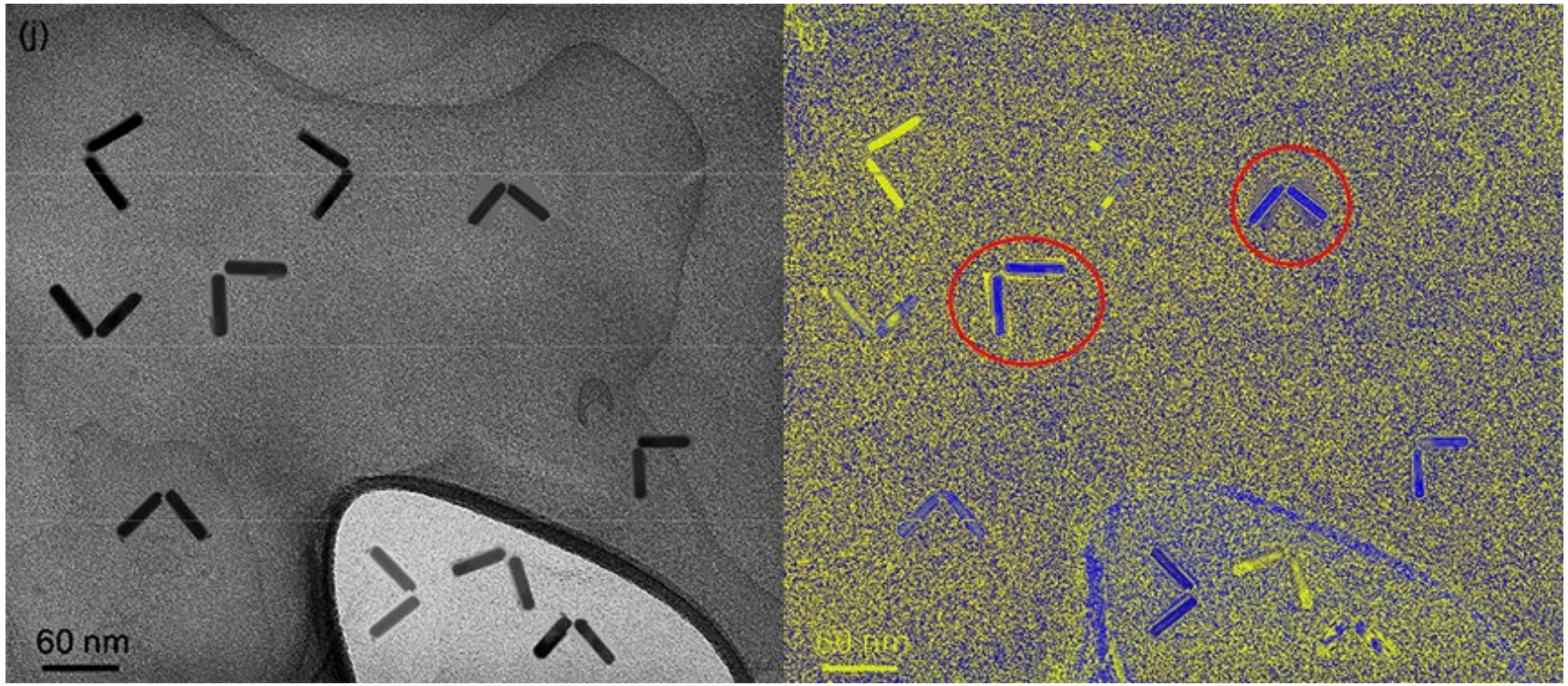


"Even though these are all nominally "cat Pokémon," it's easy to recognize them as different kinds of characters."

← Or even from just the silhouette

https://betterposters.blogspot.com/2021/05/solving-pokemon-problem.html

Know when to break the rules



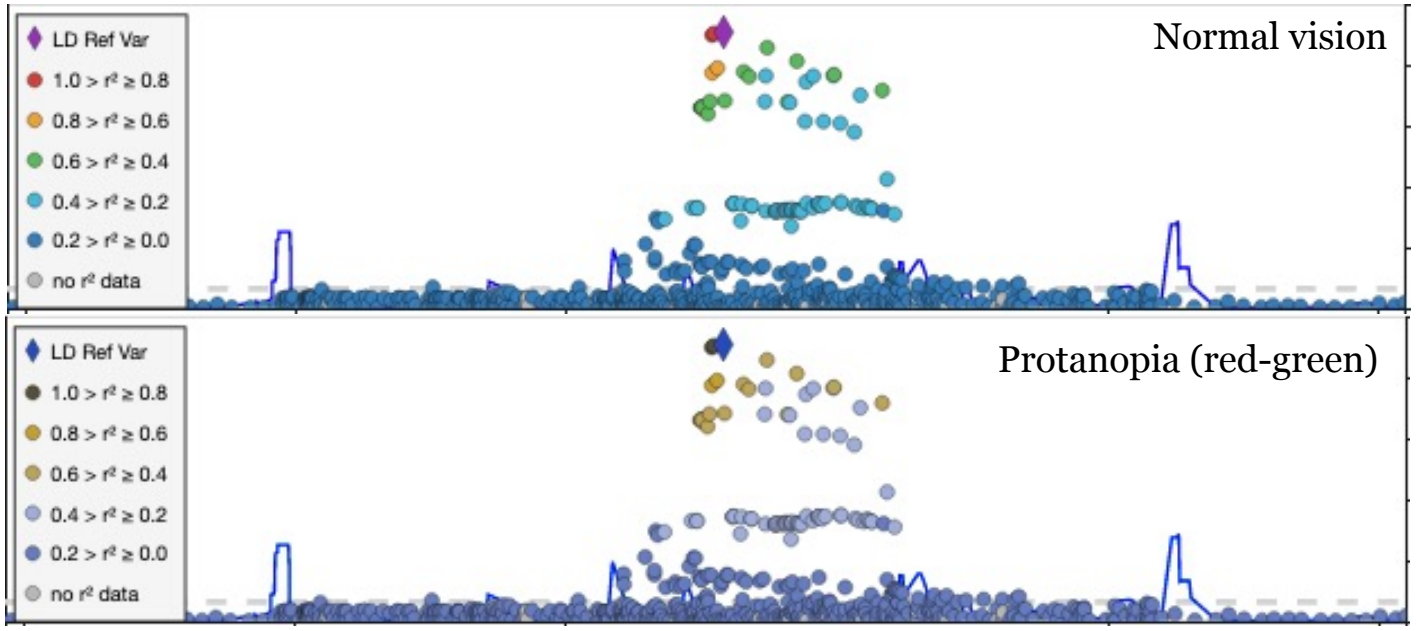
Small variations can be useful!

Example [1]: image fraud analysis; slight background differences are magnified

Example [2, 3]: Patterns in specialized data- color scheme that emphasizes extreme values

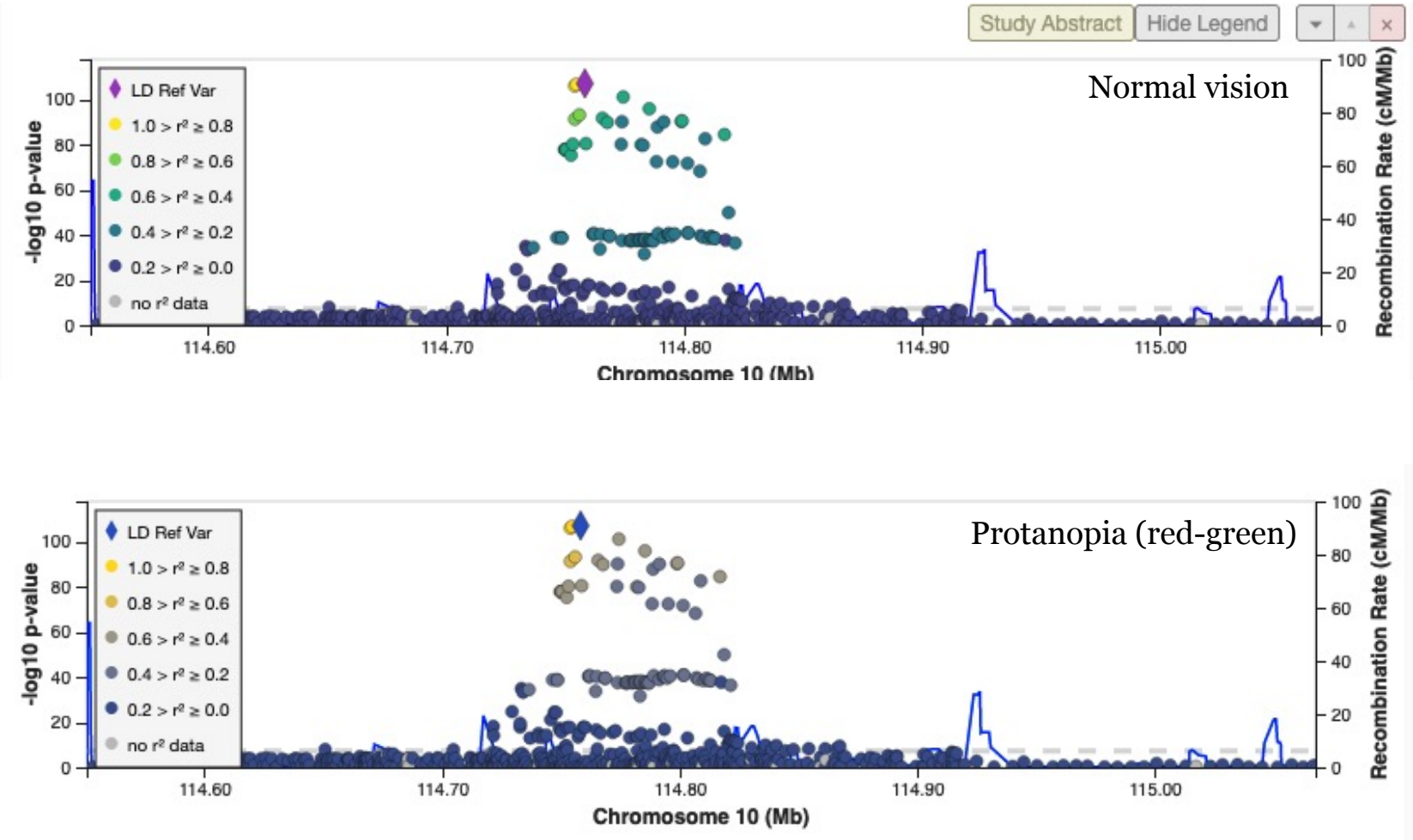
[1] https://github.com/abought/ori_openforensicactions/
[2] https://eos.org/features/visualizing-science-how-color-determines-what-we-see
[3] "Ramps": http://soliton.vm.bytemark.co.uk/pub/cpt-city/imagej/index.html

Application to our own tools: LocusZoom.js



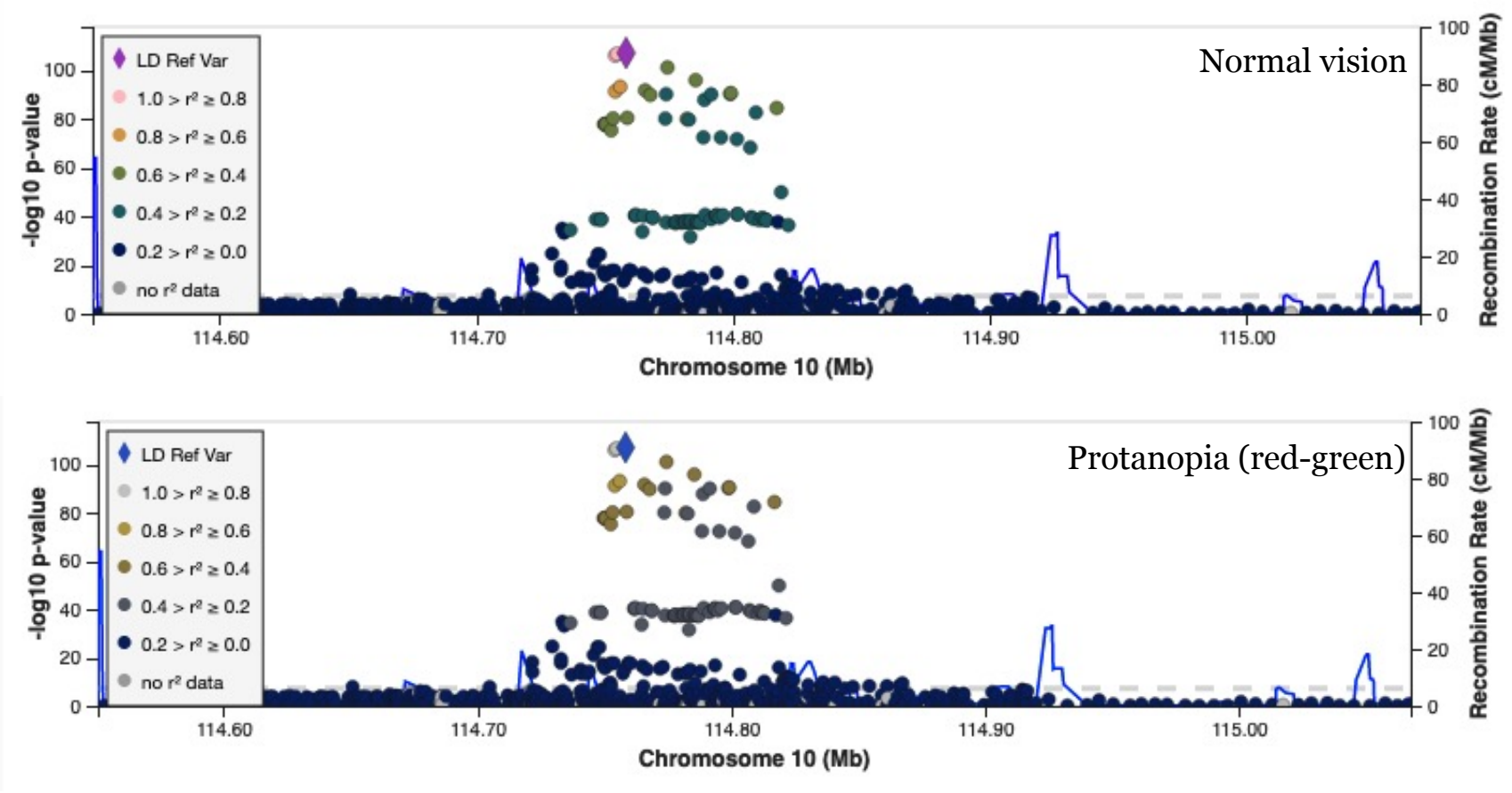
- Default LocusZoom color palette is ok, but not great with red-green colorblindness
- Can we find a better color palette?

Candidate colormap 1: Viridis



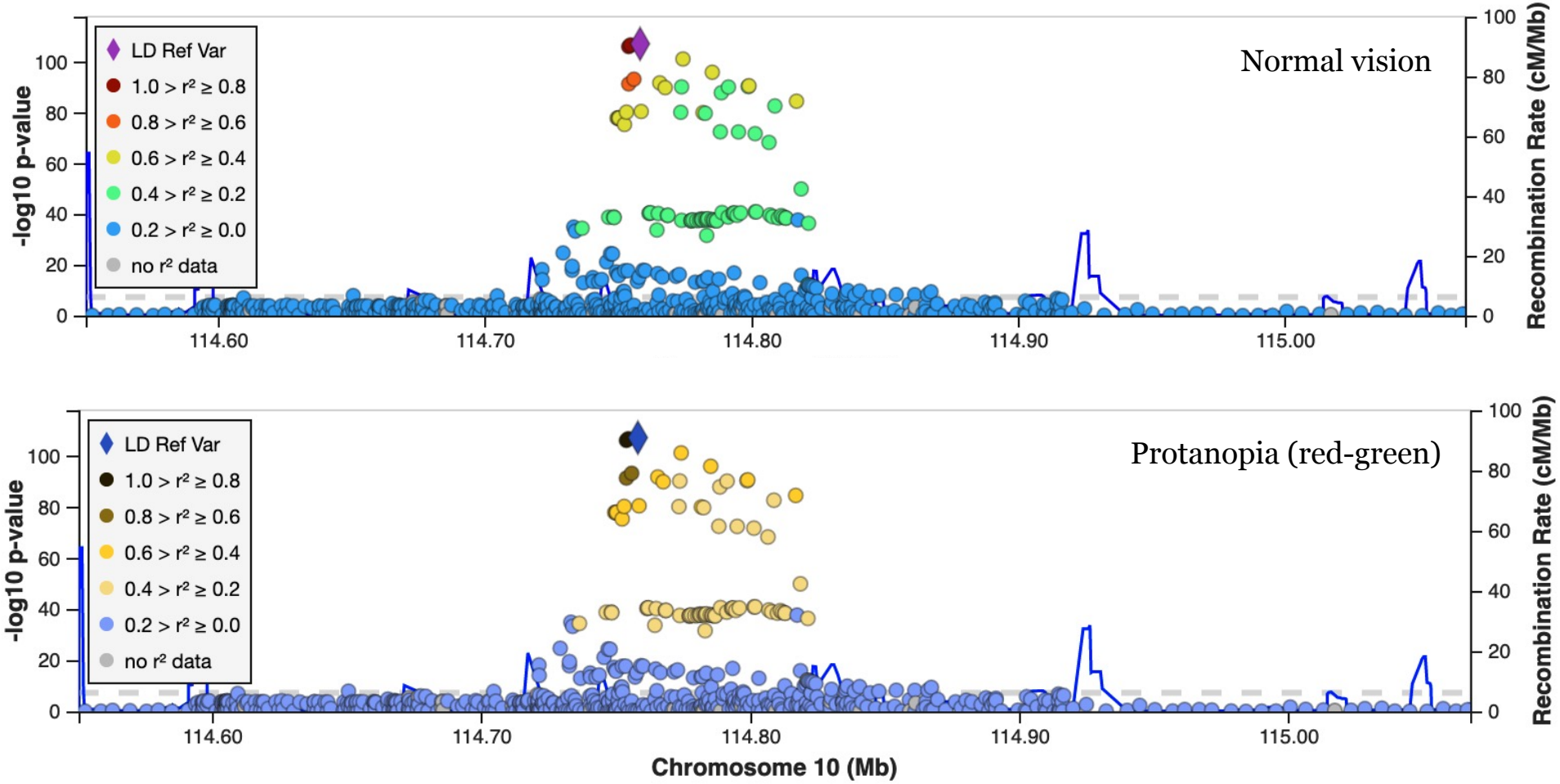
- Popular colormap (eg matplotlib default)
- Decent uniformity and colorblind-support
- Less contrast between adjacent values
- Changes meaning for existing users: Yellow means “high”; there is no red
  - This could be confusing

Candidate color map 2: Batlow10



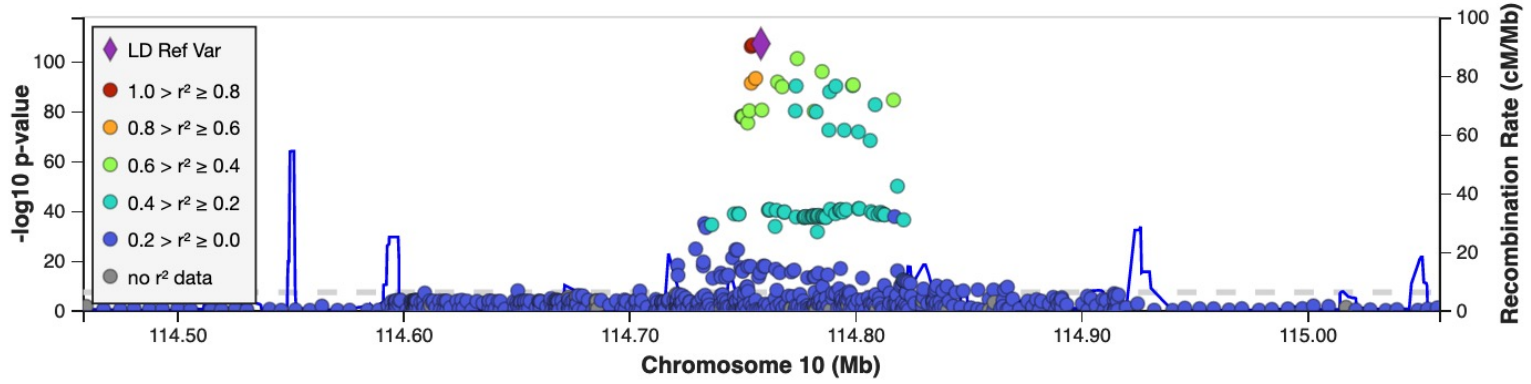
- Relatively new
- A perceptually uniform rainbow scheme
- Good black and white and colorblind support
- Kind of depressing: A bit more muted than Jet/ Rainbow

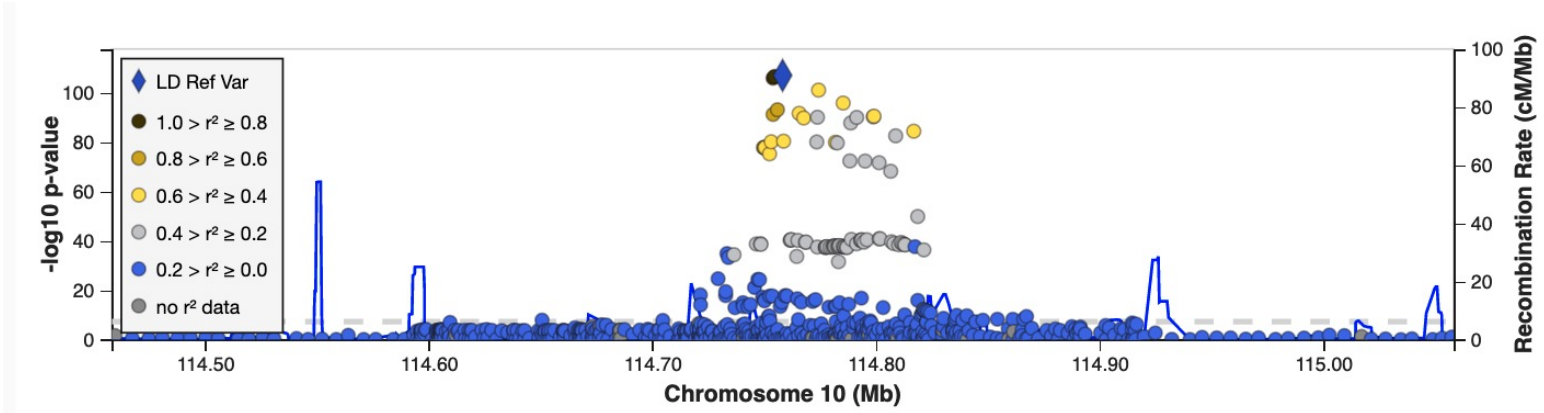
Candidate color map 3a: Turbo



- A re-balanced version of Jet
  - Peppy with more consistent brightness
- Ok for some colorblindness
- Not great in black and white
- 0.2, 0.4…

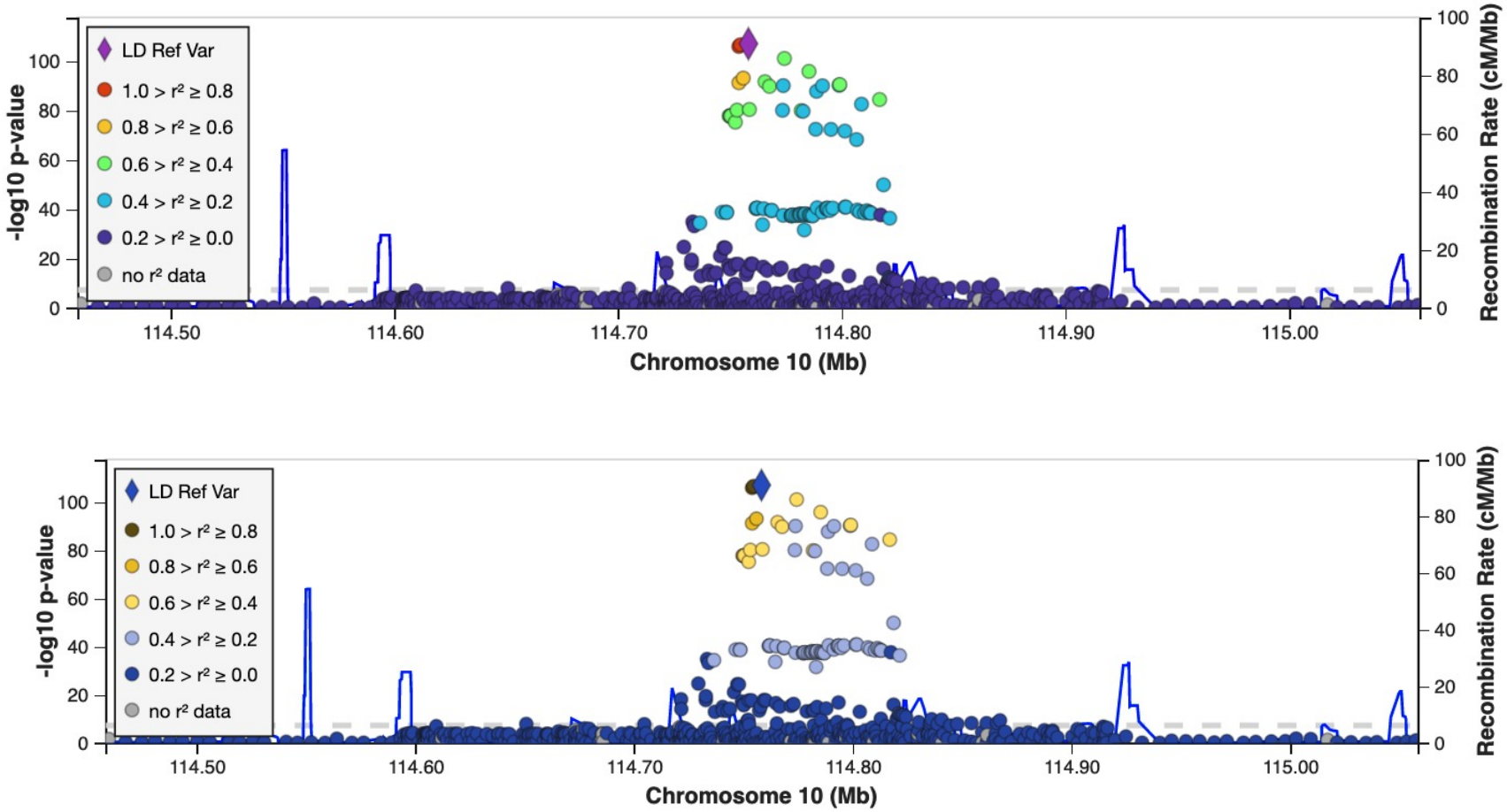
Candidate color map 3b: Turbo(b)





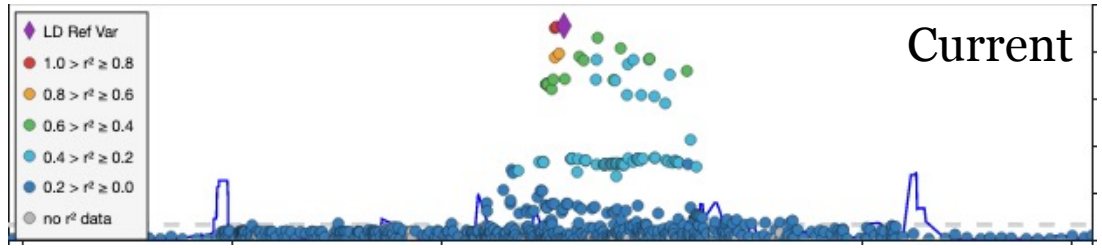
- Same colormap as prior slide
- Different interpolation points (0.1, 0.3…)
- Slightly better colorblindness support (0.4->0.6)

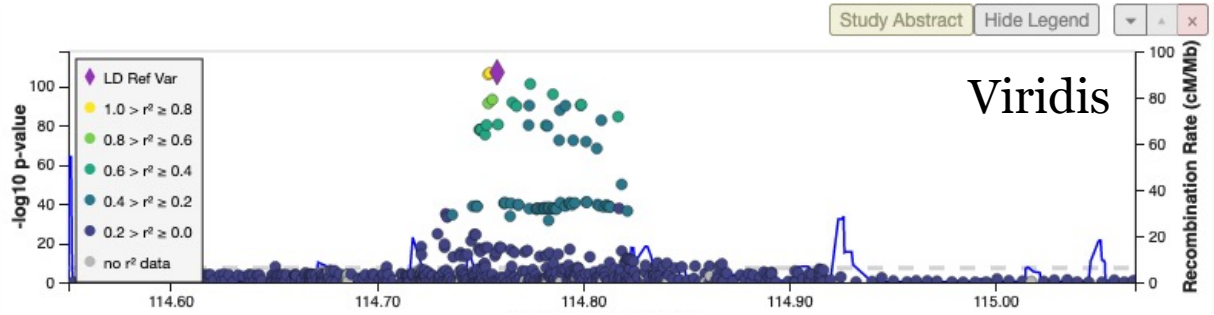
Candidate color map 3c: Turbo(c)

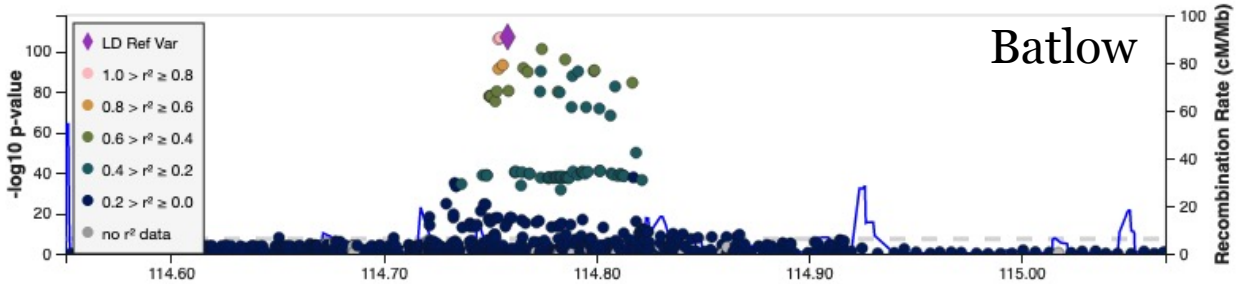


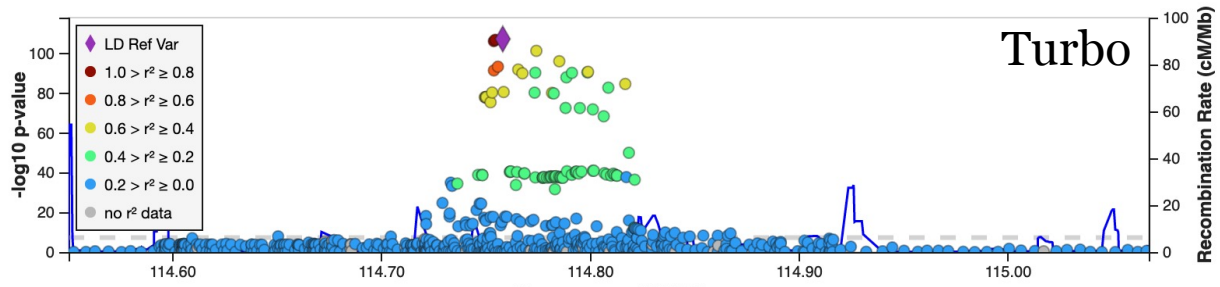
- Same colormap as prior slide
- Different interpolation points (0.05, 0.25…)

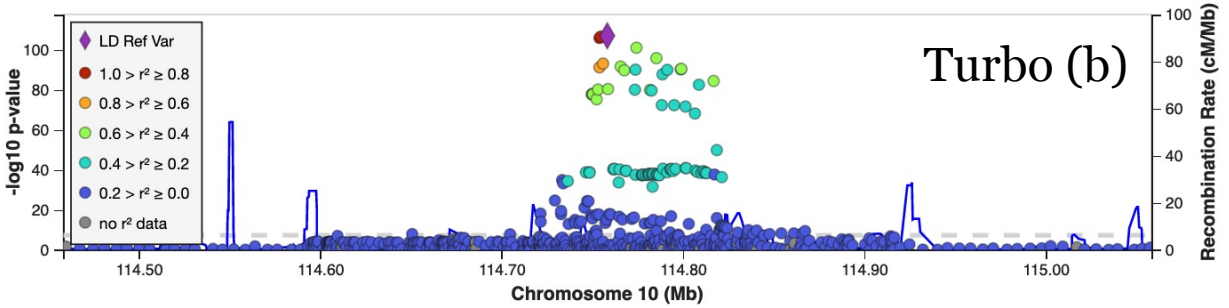
Summary: New LZ.js color options

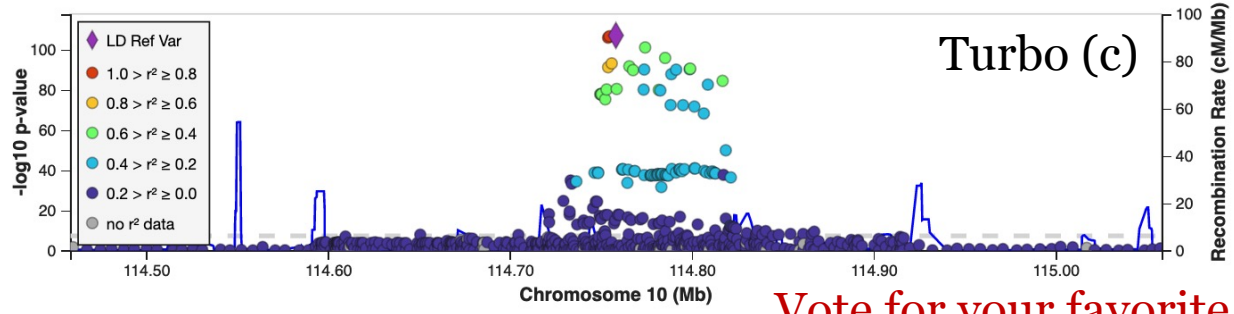












Vote for your favorite!

Further reading

- ggplot2: Intro to data visualization in R (“grammar of graphics”)
  - https://learning.oreilly.com/videos/data-visualization-in/9781491963661/9781491963661-video256239
- Posters: figures in context
  - Better posters (Blog): https://betterposters.blogspot.com/2011/05/epic-logo-post.html
  - Ten Simple Rules for Better Figures (2014) https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833
  - Why We Use Bad Color Maps and What You Can Do About It https://www.osti.gov/servlets/purl/1338147
  - NYU Poster guide: https://guides.nyu.edu/poster

Other ways to see: Screen Readers

- Demonstration videos
  - https://twitter.com/Kristy_Viers/status/1287189581926981634
- How to use macOS “VoiceOver” (built in):
  - https://dequeuniversity.com/screenreaders/voiceover-keyboard-shortcuts
- Windows 10 “Narrator”:
  - https://support.microsoft.com/en-us/windows/complete-guide-to-narrator-e4397a0d-ef4f-b386-d8ae-c172f109bdb1

Beyond visualization

- Sonification (turning patterns into music)
  - https://www.youtube.com/watch?v=3EXvR1shVFQ
- WebPlotDigitizer: Extract datapoints by tracing a figure (*)
  - https://apps.automeris.io/wpd/
- Show the data!
  - Tables
  - Supplemental files / data repositories

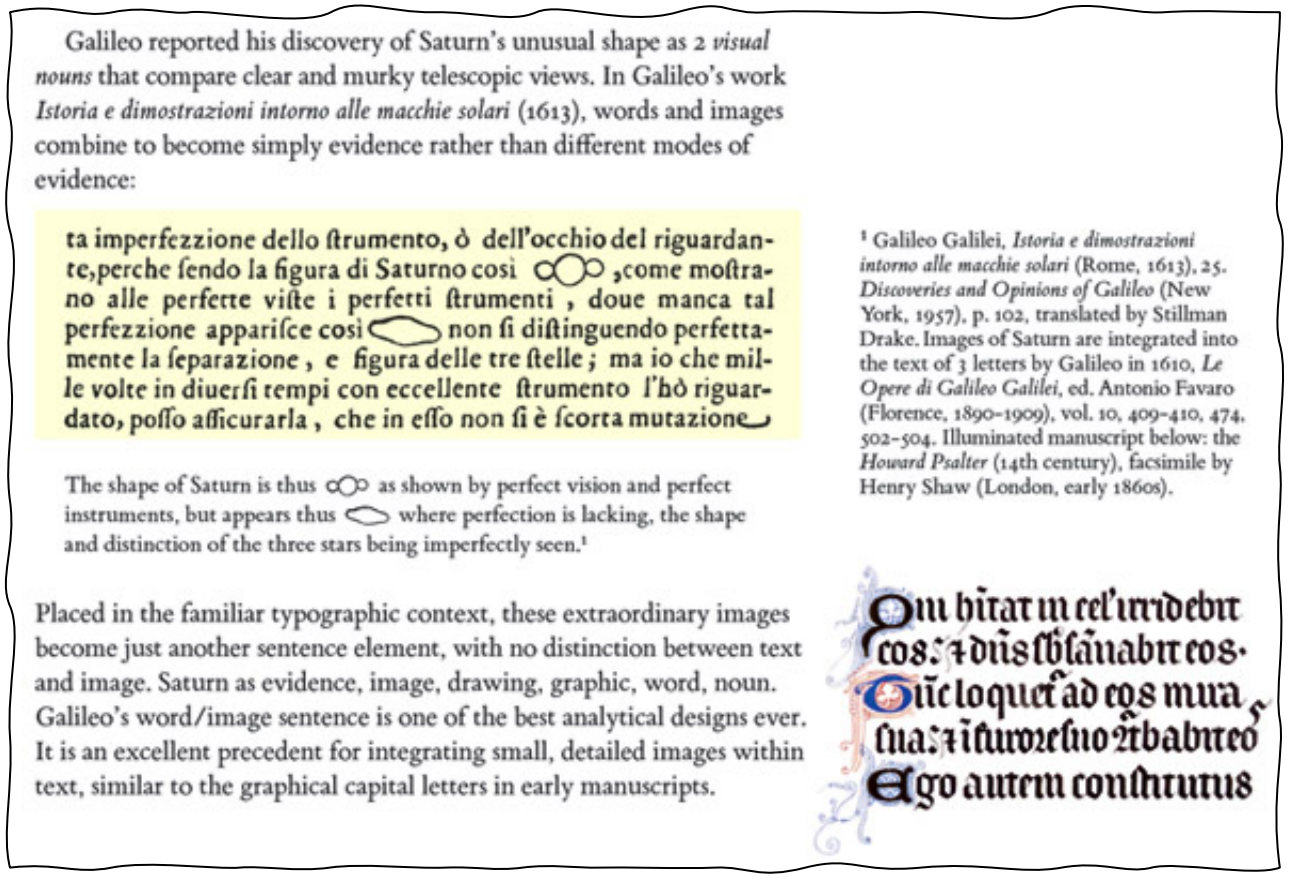
(*) Some authors get *really* mad when you access raw data this way, even just to check the results in that paper.

Sometimes, you just want data

Resources for sharing and reusing data

Claims and evidence

- Papers have long incorporated claims and evidence together
- As methods become more complex, it is hard to follow up just from figures
  - Access to the raw data is needed



Galileo reported his discovery of Saturn's unusual shape as 2 *visual nouns* that compare clear and murky telescopic views. In Galileo's work *Istoria e dimostrazioni intorno alle macchie solari* (1613), words and images combine to become simply evidence rather than different modes of evidence:

ta imperfezzione dello ſtrumento, ò dell'occhio del riguardante, perche ſendo la figura di Saturno così ○○○ ,come moſtrano alle perfette viſte i perfetti ſtrumenti , doue manca tal perfezzione appariſce così ⬭ non ſi diſtinguendo perfettamente la ſeparazione , e figura delle tre ſtelle ; ma io che mille volte in diuerſi tempi con eccellente ſtrumento l'hò riguardato, poſſo aſſicurarla , che in eſſo non ſi è ſcorta mutazione

The shape of Saturn is thus ○○○ as shown by perfect vision and perfect instruments, but appears thus ⬭ where perfection is lacking, the shape and distinction of the three stars being imperfectly seen.[1]

[1] Galileo Galilei, *Istoria e dimostrazioni intorno alle macchie solari* (Rome, 1613), 25. *Discoveries and Opinions of Galileo* (New York, 1957), p. 102, translated by Stillman Drake. Images of Saturn are integrated into the text of 3 letters by Galileo in 1610, *Le Opere di Galileo Galilei*, ed. Antonio Favaro (Florence, 1890–1909), vol. 10, 409–410, 474, 502–504. Illuminated manuscript below: the *Howard Psalter* (14th century), facsimile by Henry Shaw (London, early 1860s).

Placed in the familiar typographic context, these extraordinary images become just another sentence element, with no distinction between text and image. Saturn as evidence, image, drawing, graphic, word, noun. Galileo's word/image sentence is one of the best analytical designs ever. It is an excellent precedent for integrating small, detailed images within text, similar to the graphical capital letters in early manuscripts.

Edward Tufte, Beautiful Evidence (2006)

"We selected 41 open access papers… and executed the R code….

46 out of 97 reproduced figures deviated from the original figures on a deeper level, e.g. graphs had different curves, and key numbers were missing or different…

The code of two papers ran without any issues, 33 had resolvable issues, and two were partially executable, i.e. the code produced output but also had issues that we could not resolve… The code of 15 papers contained issues that required contacting the corresponding author."

Konkol, M., Kray, C. & Pfeiffer, M. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science* **33**, 408–429 (2019). (link)

Reproducibility Challenge

Pick an interesting paper in your field. Can you find the code and data to reproduce it after 1 year? After 5?

If the links in the paper don't work, try doing this for a paper from your own group. Can the person who sits next to you reproduce your paper? What if you give them the code, data, and instructions?

Considerations for sharing

- FAIR Principles
  - **F**indable: Indexed by search, and persistent links
  - **A**ccessible: Respect access rules, but not just “available upon request”
  - **I**nteroperable: Integrate with tools and workflows
  - **R**eusable: Maintain provenance and vocabulary
- Resources and best practices vary by field; ask around!

Data Sharing and Management Snafu in 3 Short Acts: https://www.youtube.com/watch?v=66oNv_DJuPc

Where to share: Methods and code

- Genetic data (ask Heather how to use these!)
  - dbSNP
  - dbGaP
  - …others?
- Summary results
  - EBI eQTL Catalogue
  - EBI GWAS Catalogue
- Source code/ methods
  - GitHub (“living” version of source code + provenance)
  - Zenodo (auto-generate author list and citable DOI from a GitHub repo; archives maintained by CERN)

Example from collaborators: https://pubmed.ncbi.nlm.nih.gov/32213349/ - https://github.com/ParkerLab/ataqv/

Preparing to share: Source code

- Good Enough Practices in Scientific Computing: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510
- JOSS Review checklist: https://joss.readthedocs.io/en/latest/review_checklist.html
- Researcher guides: https://www.software.ac.uk/resources/guides/guides-researchers
- US Research Software Engineer Association: https://us-rse.org/

Summary

- Data visualization is a mix of art and science; the best images take the domain and type of data into account
  - A good visualization reveals patterns that are hidden; a bad visualization implies things that are not there
- Multivariate data can be shown in many overlapping ways
- When in doubt, make the raw data accessible – don't restrict people to only your preset view
- Be inspired by examples!

Possible discussion topics

- What are some unique challenges for visualizing the data you work with?
- How do we share data in our group?
- What are some visualization tools you like working with? What do they do well? What doesn't work so well?