SC6.9
DataViz: Visualise your data effectively and avoid common pitfalls

Edoardo Martini (University of Leipzig, Germany)
Paola Mazzoglio (Politecnico di Torino, Italy)
Epari Ritesh Patro (University of Oulu, Finland)
Roshanak Tootoonchi (University of Trento, Italy)
Debasish Mishra (Texas A&M University, USA)

Young Hydrologic Society

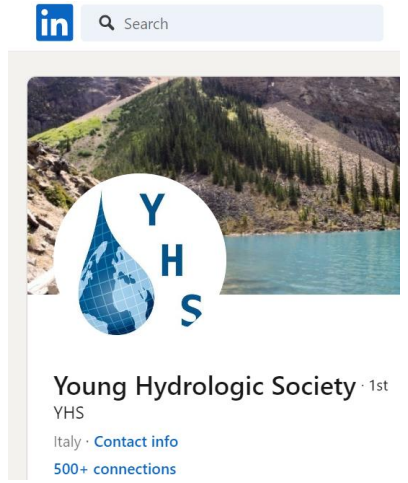
younghs.com
@YoungHydrology

EGU General Assembly
#EGU24
14 -19 April 2024

Young Hydrologic Society

News | About | National initiatives | Blog | Resources | Past events | Contact | Get involved

YHS



The Young Hydrologic Society (YHS) is a bottom-up initiative to stimulate the interaction and participation of young hydrologists within the hydrological community.

Founded in October 2012, the YHS is currently run by a team of PhD's and post-doc's from several universities across the world.

For more info, please visit our website! https://younghs.com

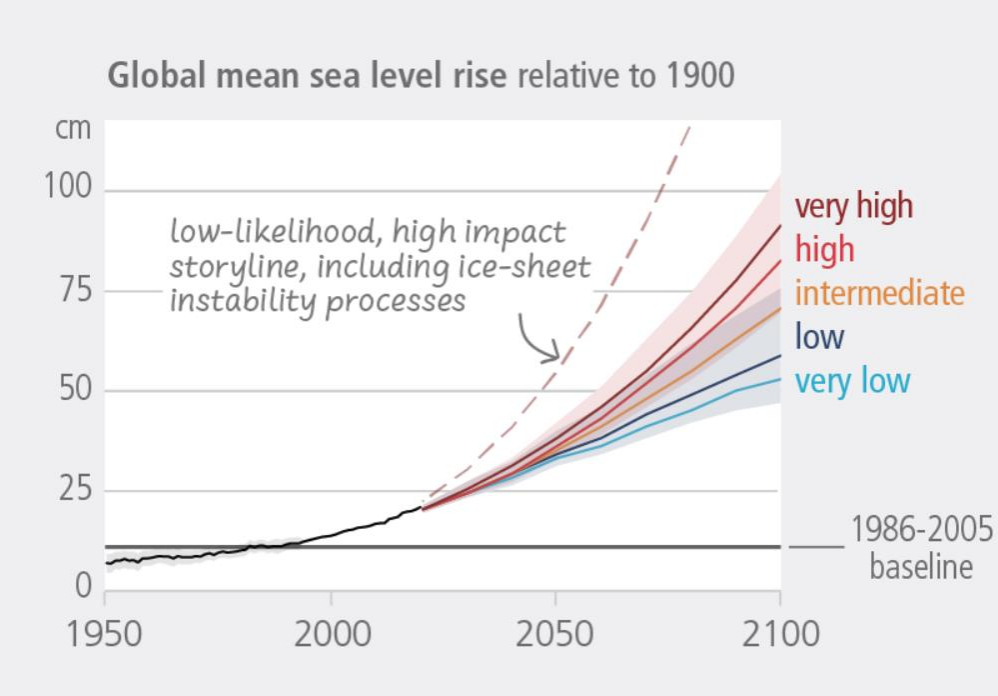
Today's menu

1. DataViz: fundamentals
2. Colour schemes
3. Publication compliance
4. Tutorial: DataViz with R
5. Tutorial: DataViz with NCL

What is DataViz?

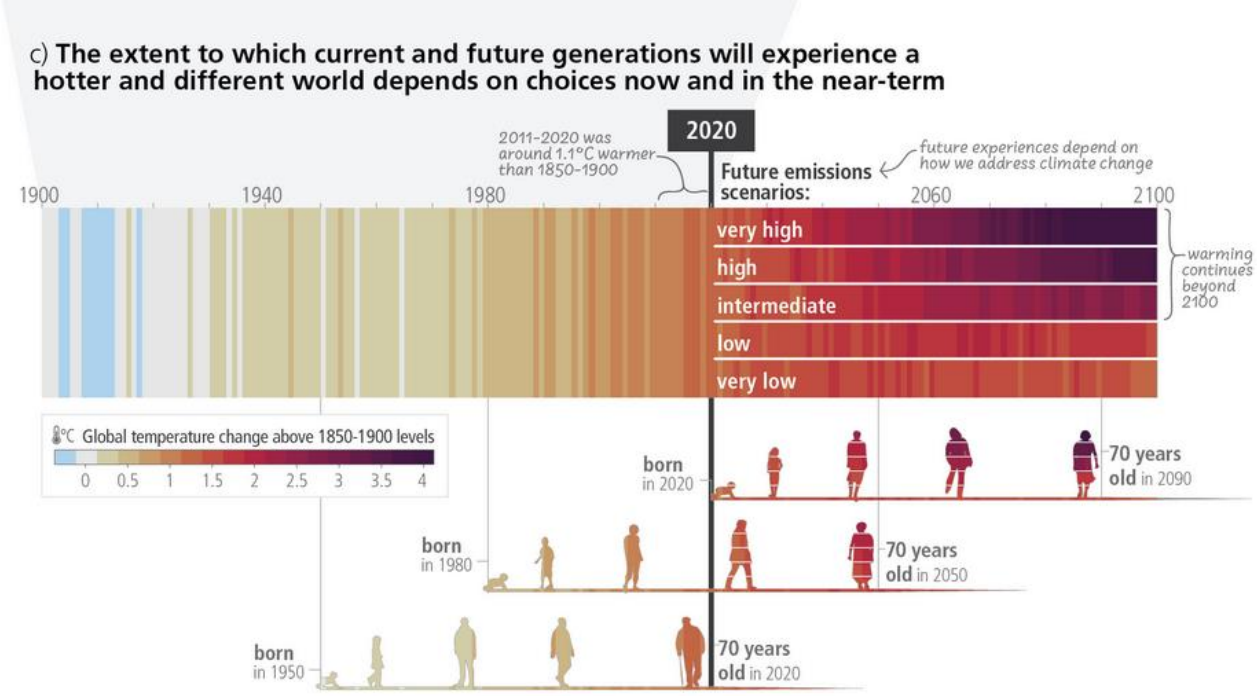
“DataViz” = Data Visualization

graphical representation of scientific data



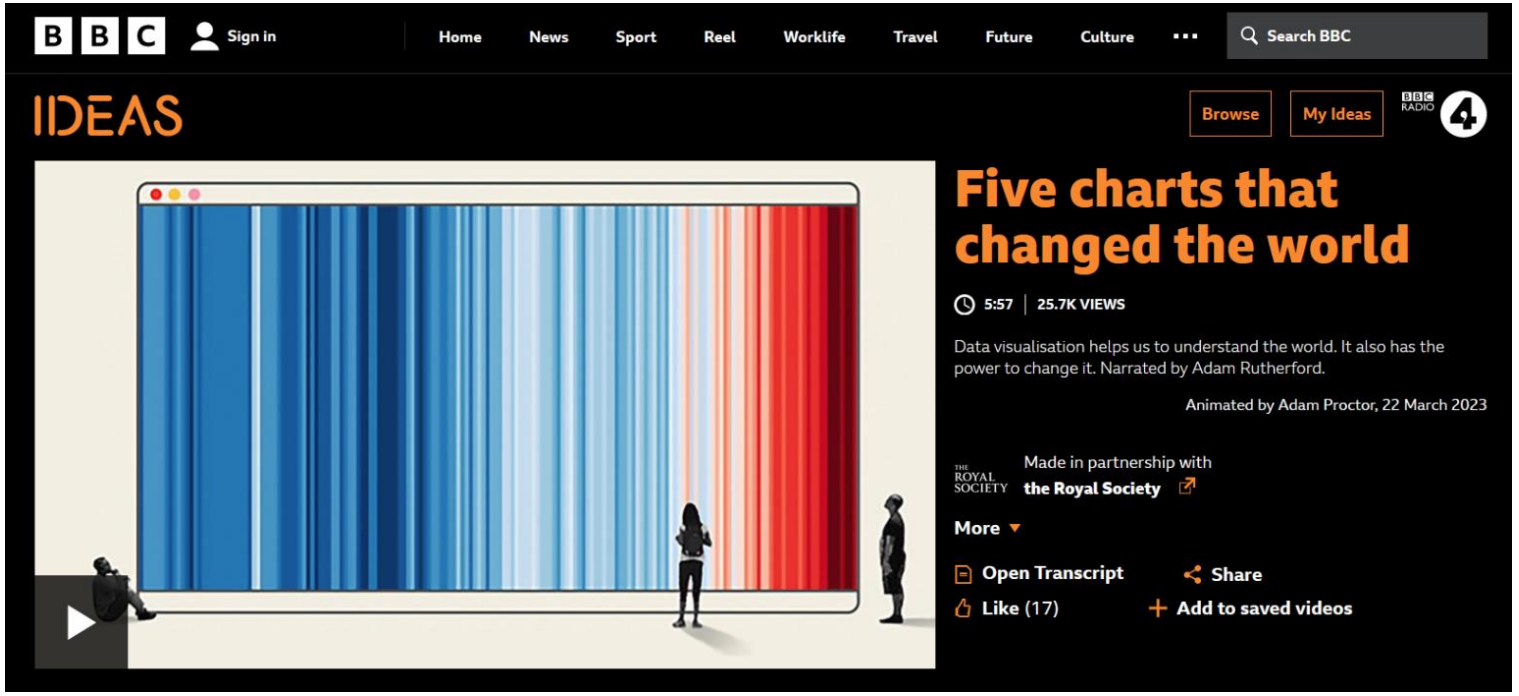
“InfViz” = Information Visualization

graphical representation of scientific concepts, incl. abstract concepts



Source: IPCC AR6 Synthesis Report, Climate Change 2023

Brief history of DataViz



bbc.co.uk/ideas/videos/five-charts-that-changed-the-world/p0fb69c1

1. DataViz: how to

Edoardo

Why DataViz?

Visualizing data helps us to comprehend huge amounts of information by compressing them into a simple, easy to understand visualization.

It helps us to find hidden patterns or see underlying problems in the data itself which might not have been obvious without a good chart.

Why DataViz?

We use DataViz to understand our data and communicate them to the audience.

The goal of a DataViz is to convey information in a clear and concise format.

The human brain processes information better and quickly when it is presented visually.

How to DataViz

Good DataViz:

- Correct
- Effective
- Accessible

Bad DataViz:

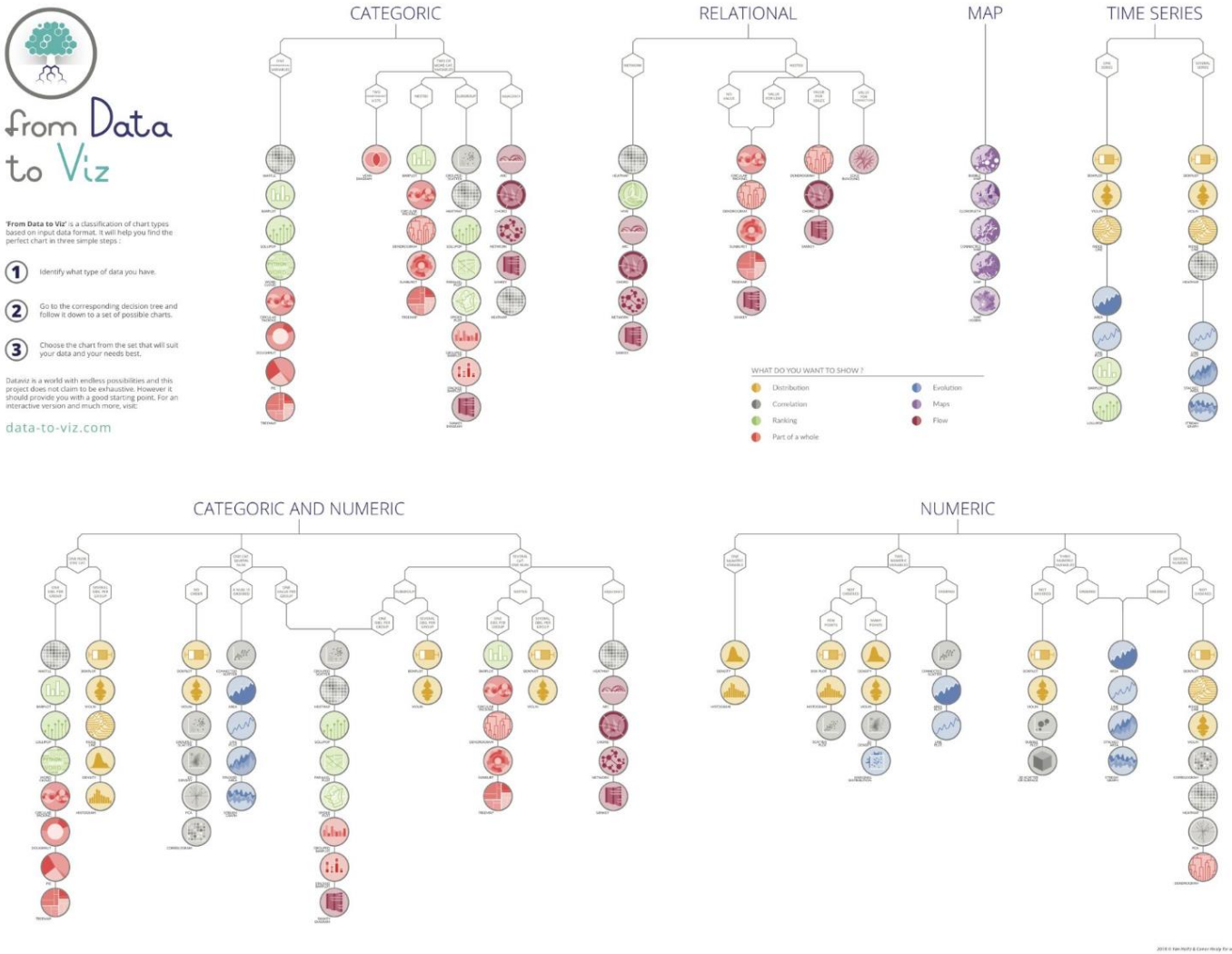
- Misrepresent the data
- Use inappropriate data
- Too much or too less information
- Inconsistent
- Ignore limits of human perception

How to DataViz

Before you DataViz, think:

- Purpose — *Why am I making this visualization?*
- Audience — *Who am I making it for?*
- Medium — *How will I use and share it?*
- Tools — *What can I use to make it?*
- Message — *What story does it tell?*
- Critical approach — *Who does it affect? Who is left out?*

Choosing the right plot type



Useful resources:

data-to-viz.com

datavizcatalogue.com

datavizproject.com

Source: datavizuniverse.substack.com

Graphical elements of a plot

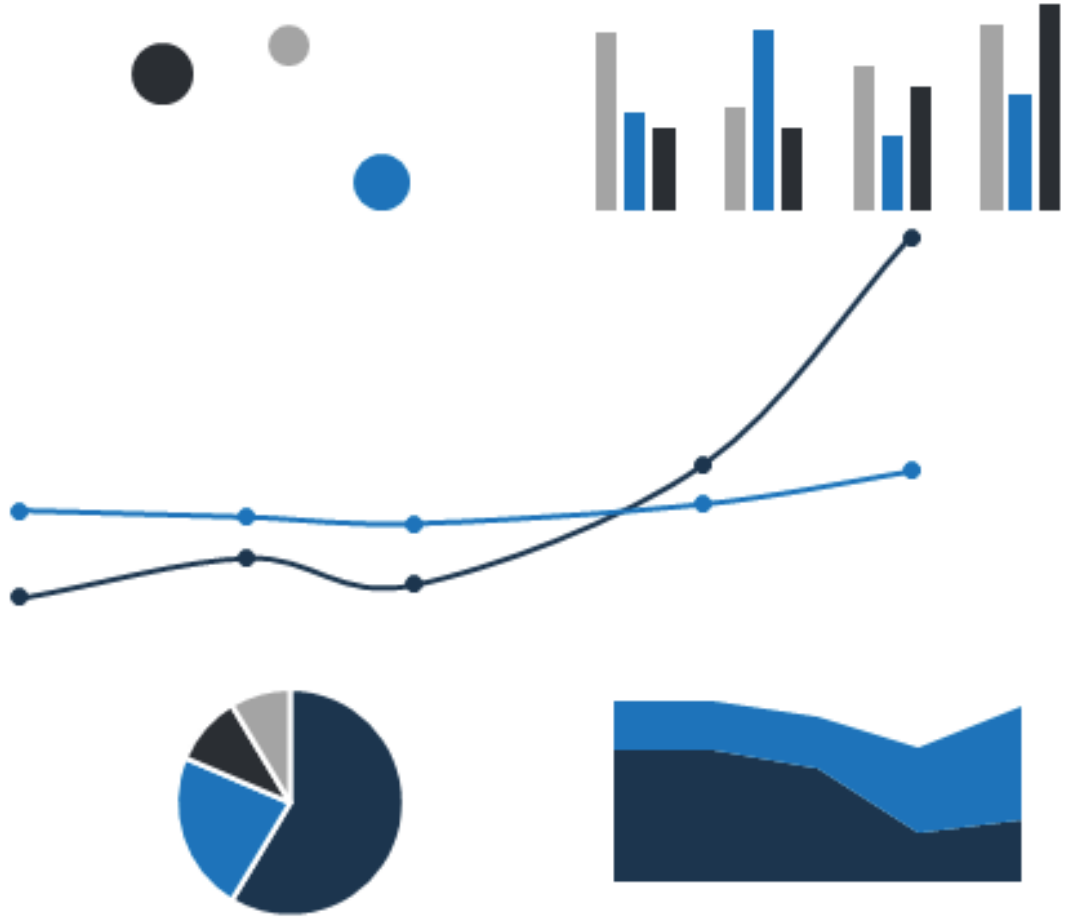
Marks and their attributes are the building blocks of all data visualizations.

Marks:

- Points
- Lines
- Areas
- Volumes

Attributes:

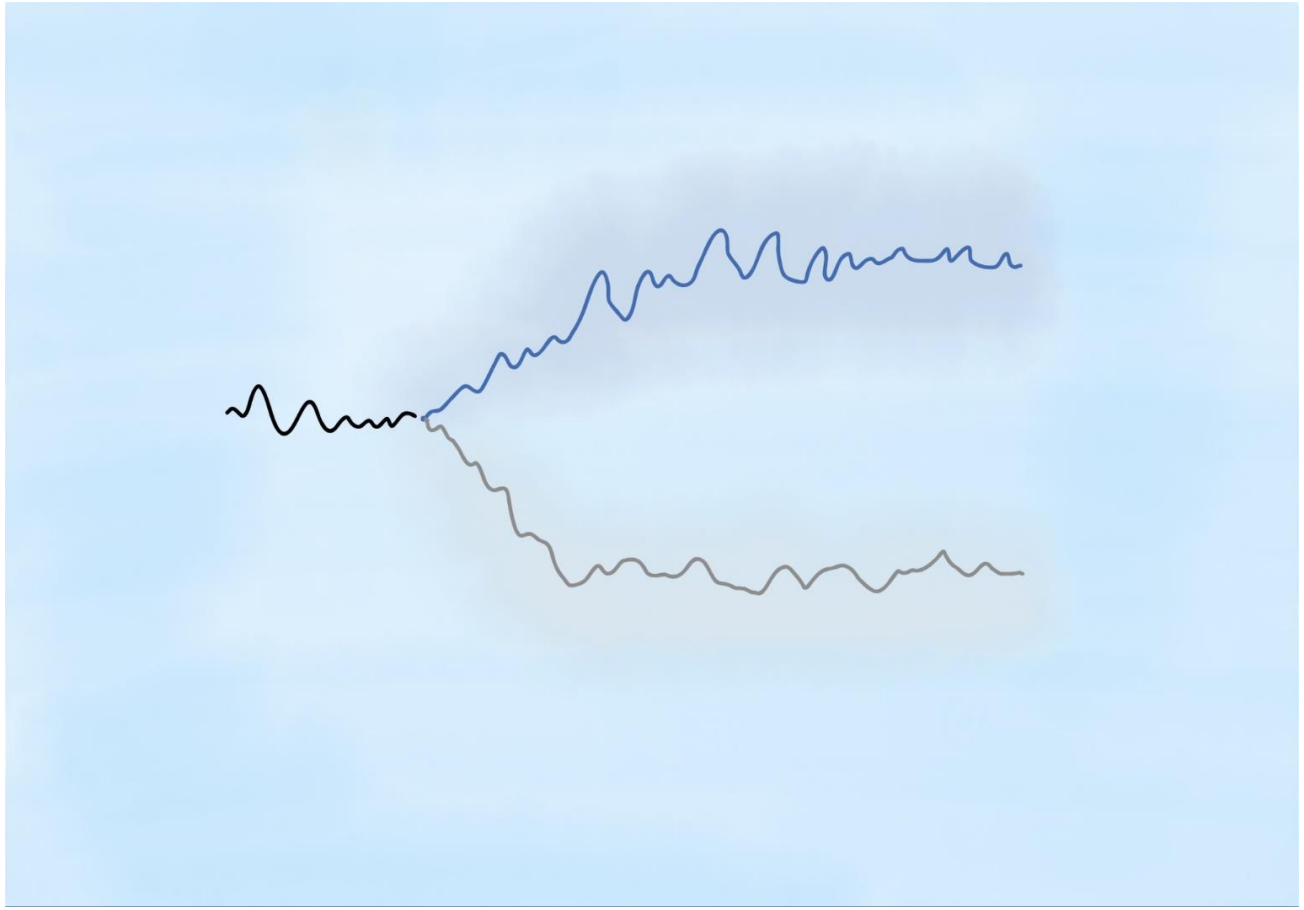
- Size
- Shape
- Orientation
- Colour



Graphical elements of a plot

Marks and their attributes are the building blocks of all data visualizations.

Alone, they are not sufficient to convey the message



Graphical elements of a plot

- Axis
- Axis label(s)
- Axis title(s)
- Grid lines & ticks
- Title
- Caption
- Data labels
- Layout (blank spaces)

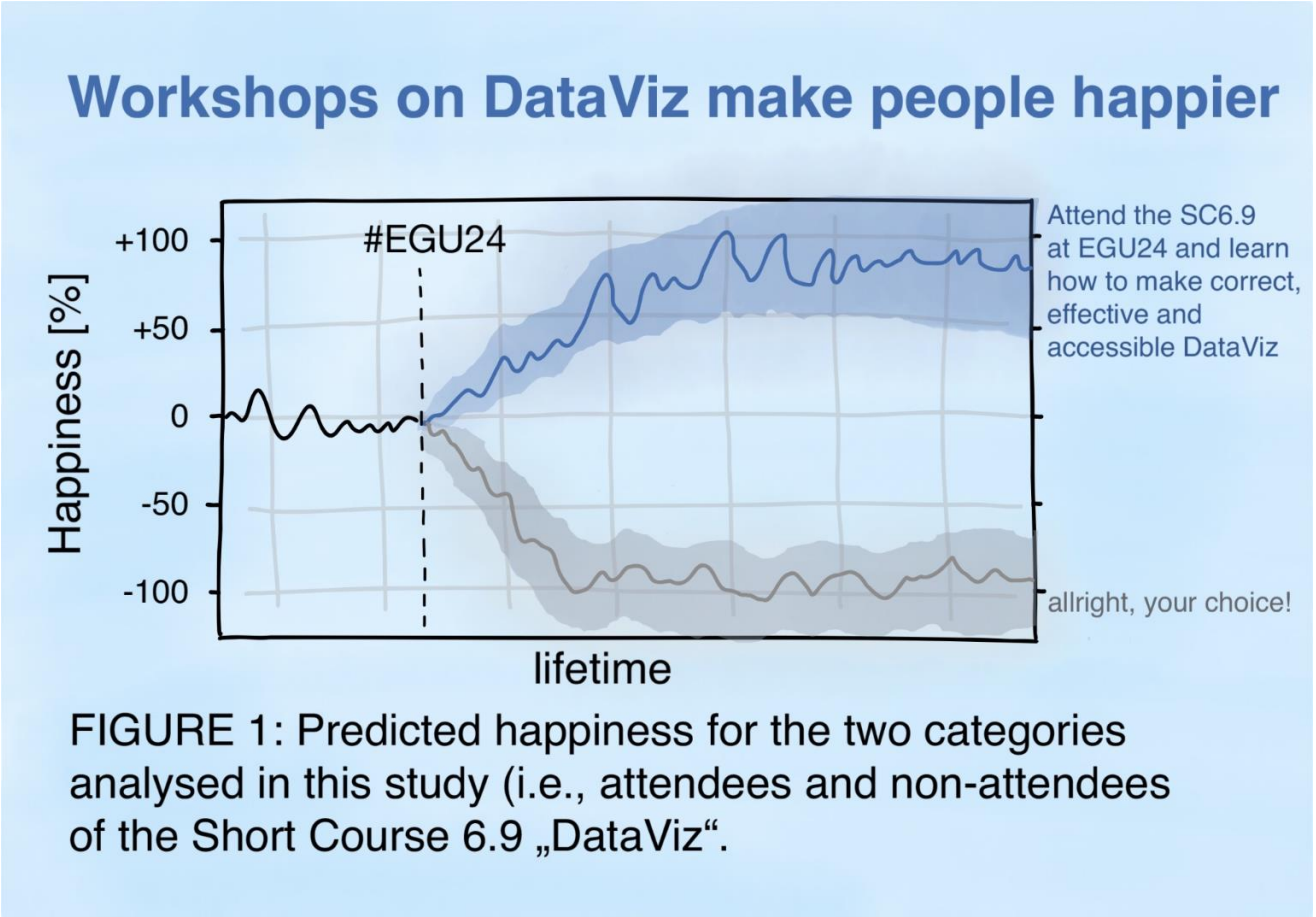


FIGURE 1: Predicted happiness for the two categories analysed in this study (i.e., attendees and non-attendees of the Short Course 6.9 „DataViz“.

Choose the right font type

this is the “serif”

Sans-Serif font types are:

- more clean
- less formal
- less affected by poor resolution
- better readable?

Font size: at least 9p on paper, 18p on screen

Choose the right font type

Choose the right font type — Arial

Choose the right font type — Calibri

Choose the right font type — Verdana

Choose the right font type — Times New Roman

Choose the right font type — Georgia

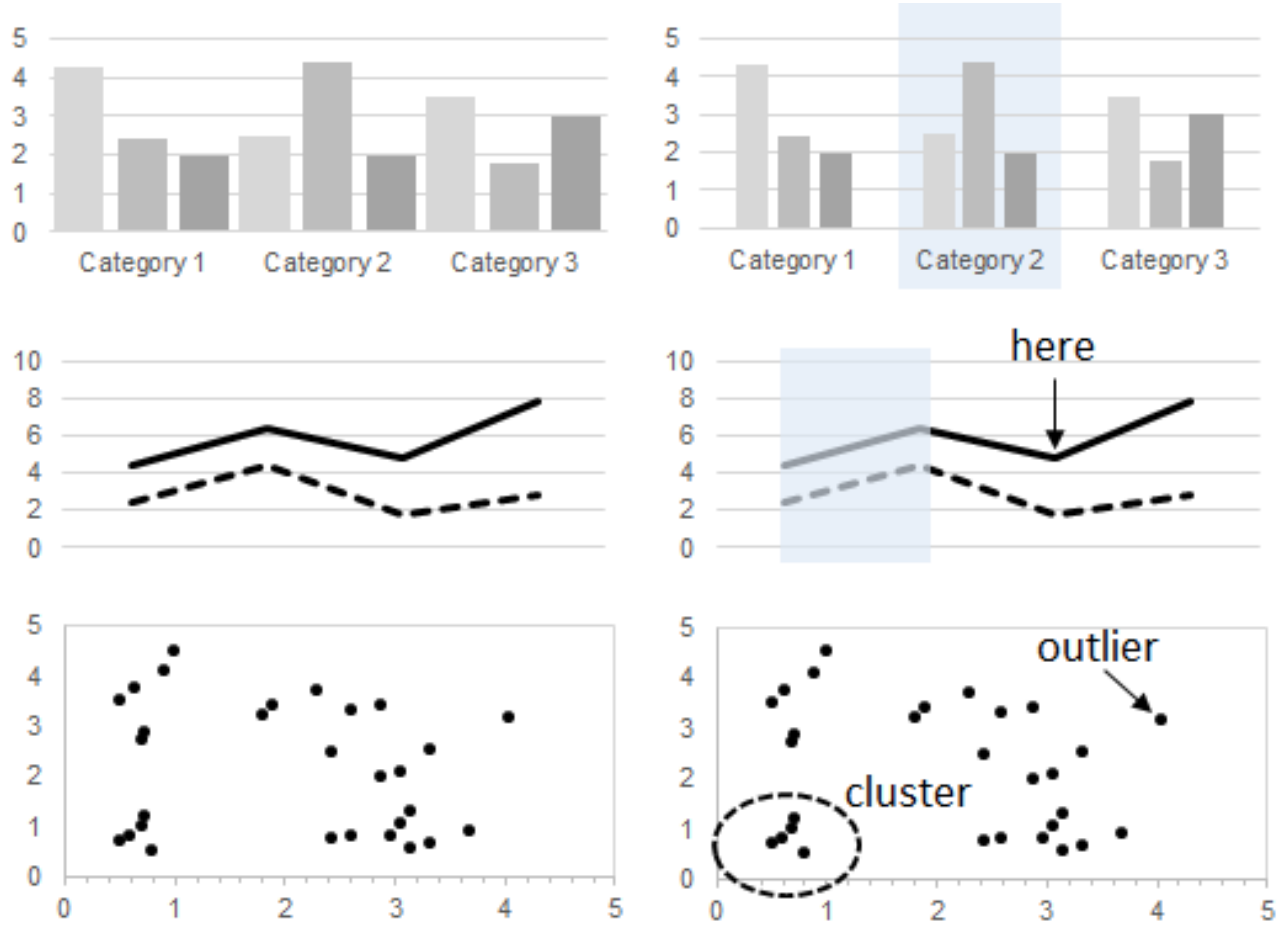
Choose the right font type — Courier New

Choose the right font type — Comic Sans

Layout (single graph)

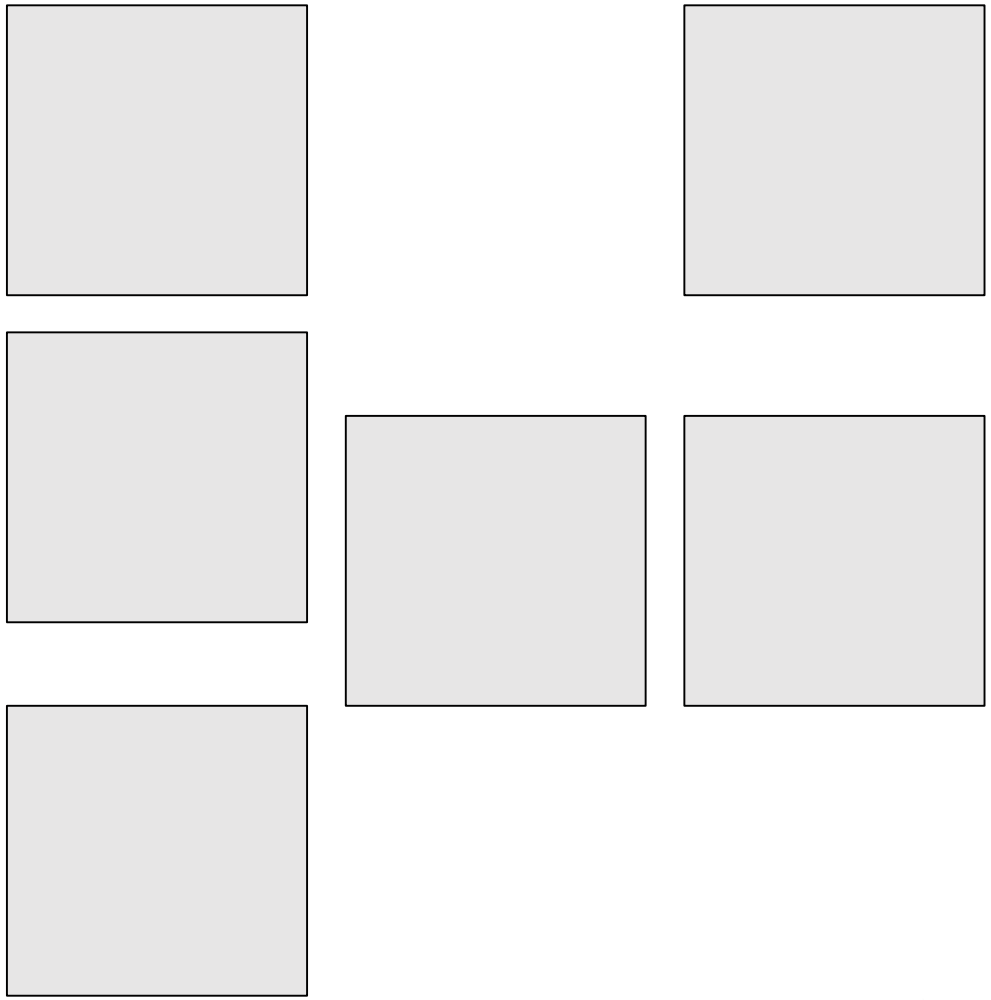
Individual chart elements (including the spatial arrangement) work together to reinforce a unified takeaway message.

Use layout and annotation to highlight and guide the reader.



Layout (multiple graphs)

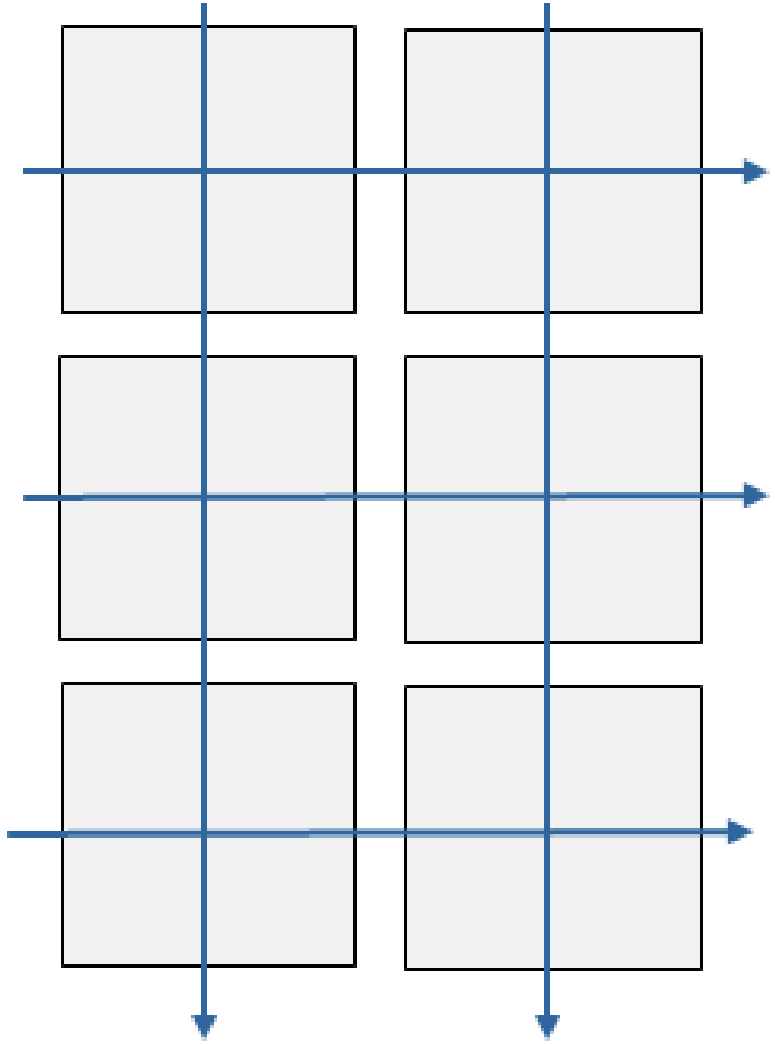
Improper arrangement of graph elements can confuse and/or mislead the readers.



Layout (multiple graphs)

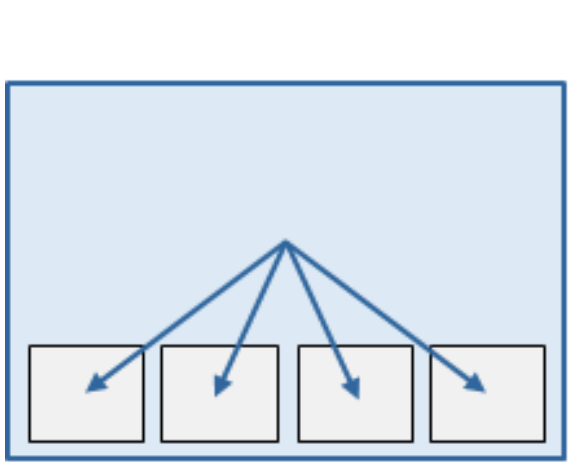
Improper arrangement of graph elements can confuse and/or mislead the readers.

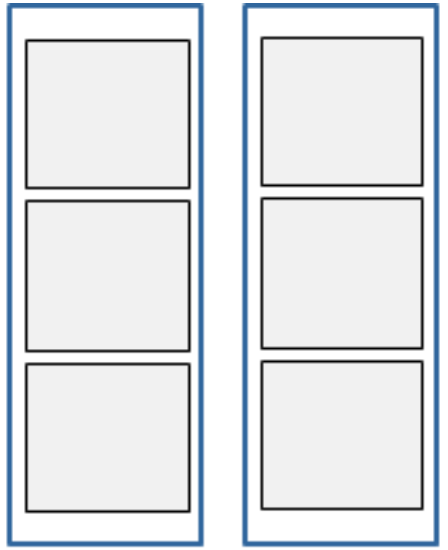
- Make it intuitive to the reader



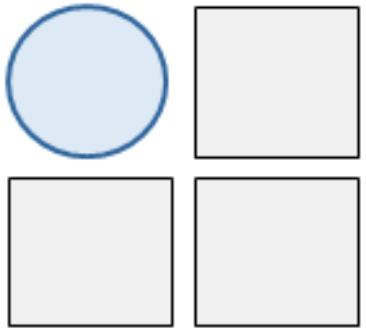
Layout (multiple graphs)

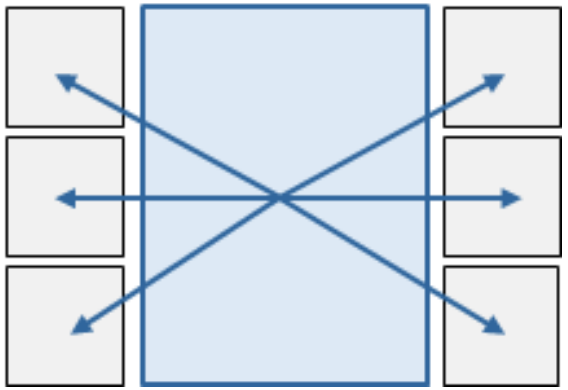
Improper arrangement of graph elements can confuse and/or mislead the readers.





- Make it intuitive to the reader
- Use layout to set priorities

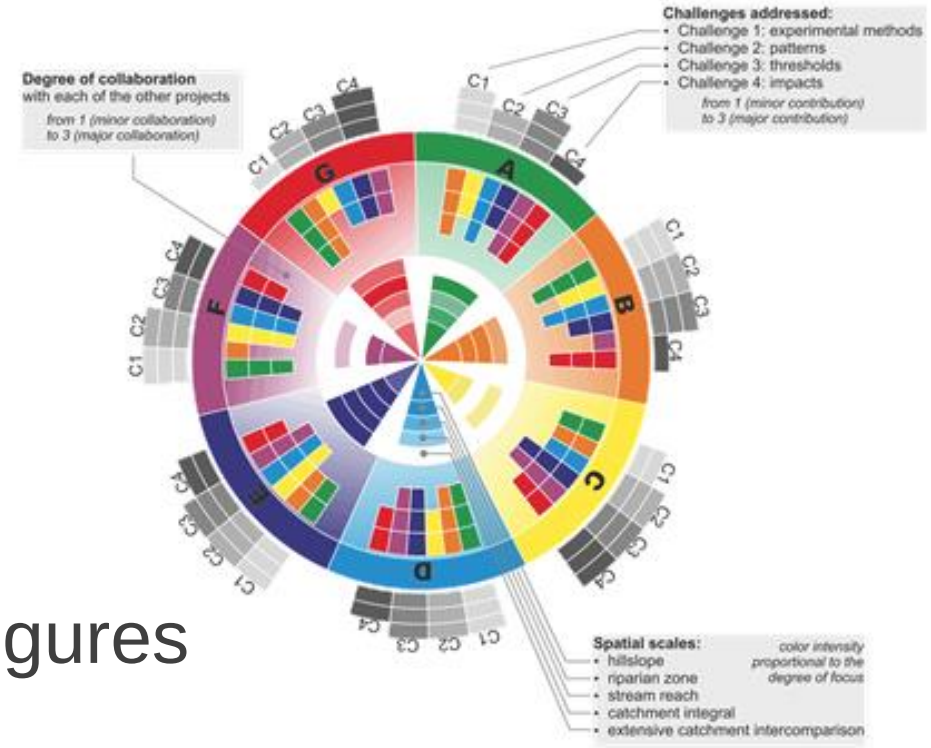




Simplicity vs creativity

Keep it simple!

- Do not overload
- Aim to one clear message vs many
- “Must have” vs “nice to have”



Or help the reader navigate the graph

- Split the content into n figures or sub-figures
- Create a hierarchy/sequence (group, highlight, annotate, ...)

Simplicity vs creativity

Beyond data plots, there are many more types of scientific visualization (posters, presentations, brochures, graphical abstracts, videos, web platforms, ...)

When we visualize scientific concepts rather than data:

- More creative approaches are accepted/recommended
- Different rules apply there

Golden rules for good DataViz

Choose the right font type:
Sans-Serif
The text should be readable (use different font sizes to convey a hierarchy)

Choose colours wisely!

Keep it simple, do not overload (must-have vs nice-to-have)

Choose the most appropriate plot type

Think before DataViz

Labels and legend concise and informative, avoid redundancy, when possible position the data labels near the data rather than in a separate legend

The axes should be clear and self-explaining

Use layout and annotation to highlight/guide

Blank spaces are your friends

Useful resources

Choose the right plot type, get inspired etc…

- data-to-viz.com
- dataviz-inspiration.com
- datavizuniverse.substack.com
- datavizcatalogue.com
- datavizproject.com
- s-ink.org
- datawrapper.de

Follow the experts:

- yan-holtz.com
- albertocairo.com
- lisacharlottemuth.com
- fabiocrameri.ch

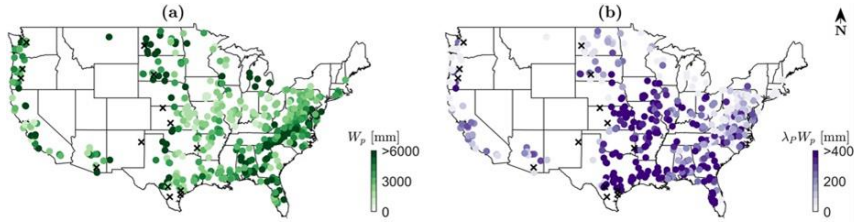
2. Colour schemes

Paola

Colour in scientific visualization

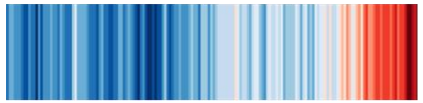
Choose the right color scale for the right reason!

Sequential (0, 1, 2, 3...)



Gnann et al, 2019, WRR

Diverging (-2, -1, 0, 1, 2)

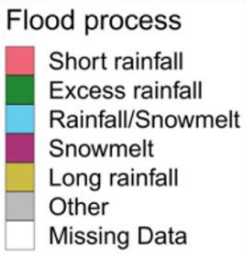


Ed Hawkins, Climate Stripes

Cyclic

Blöschl et al, 2017, Science

Categorical



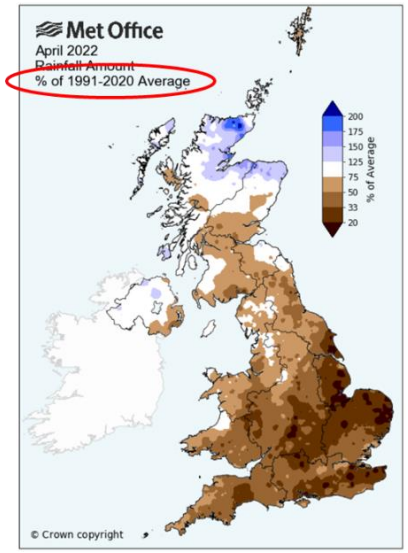
Stein et al, 2020, HP

Diverging and sequential color scale

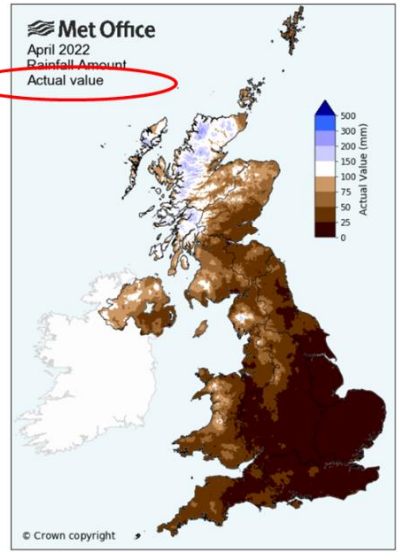
A **diverging color scale** indicates a relevant change-point.

A **sequential color scale** is needed when there is no change-point but you would like to highlight a mid-point.

Diverging values with relevant mid-point

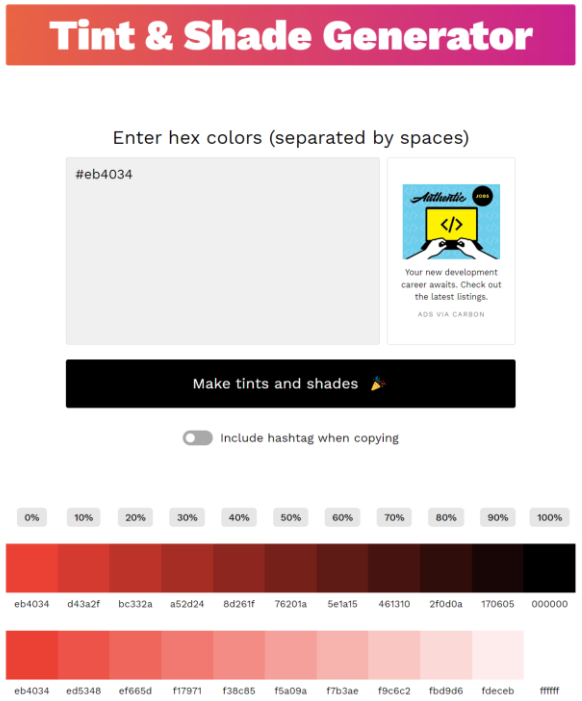


Non-diverging values indicate a mid-point that is not there



Diverging and sequential color scale

Selecting a **sequential color scale**: Tint & Shade Generator (https://maketintsandshades.com)



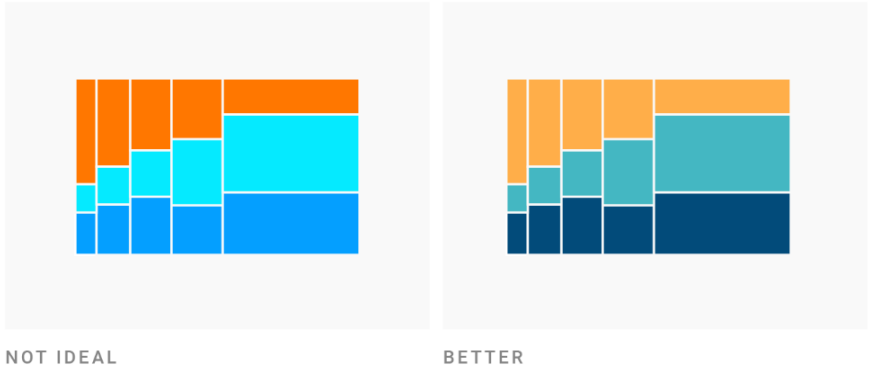
Categorical color scale

EF LEAF	EF AQUA	EF ROSE
RGB: 125.196.98 **HEX:** #7DC462 **CMYK:** 55.0.82.0	**RGB:** 13.149.208 **HEX:** #0D95D0 **CMYK:** 78.27.2.0	**RGB:** 231.47.82 **HEX:** #E72F52 **CMYK:** 3.95.62.0
EF IRIS	**EF GOLD**	**EF RUST**
RGB: 119.79.160 **HEX:** #774FA0 **CMYK:** 63.81.0.0	**RGB:** 239.183.67 **HEX:** #EFB743 **CMYK:** 6.29.85.0	**RGB:** 212.70.39 **HEX:** #D44627 **CMYK:** 11.87.100.2

Categorical colors for data visualizations defined by **Eurofound**.

Your colors should be distinguishable at every size you'll use them in and when you print them in black and white.

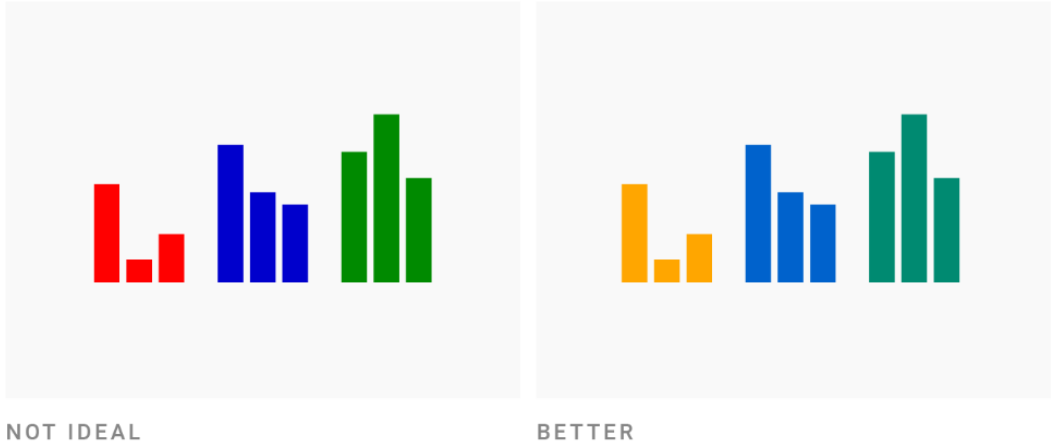
Use saturation and lightness to create new and more interesting colors for your plots.



NOT IDEAL BETTER

Source: https://blog.datawrapper.de/colors-for-data-vis-style-guides/ and https://blog.datawrapper.de/beautifulcolors/

Categorical color scale



NOT IDEAL BETTER

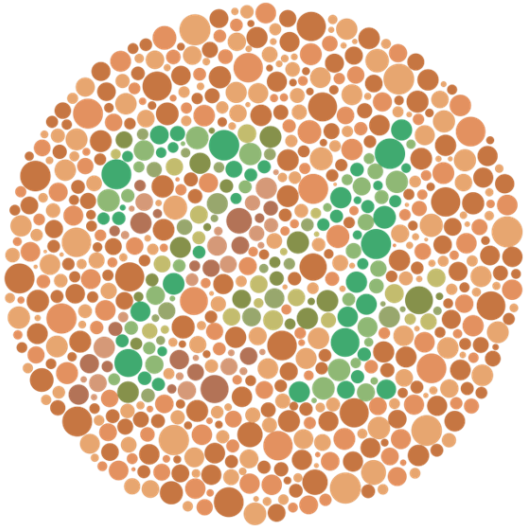
If possible, avoid pure colors: to make your colors look more natural and pleasing to your readers' eyes, you can either tone down the saturation of pure colors or make them darker.

More info on how to pick more beautiful colors for your data visualizations: https://blog.datawrapper.de/beautifulcolors/

Source: https://blog.datawrapper.de/colors-for-data-vis-style-guides/ and https://blog.datawrapper.de/beautifulcolors/

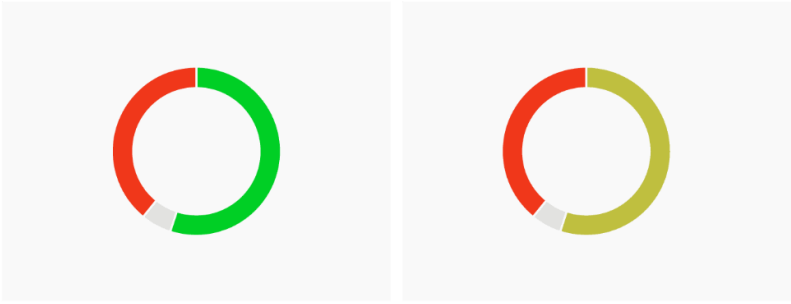
Colour in scientific visualization

Please, check your colors!



Color combinations to avoid:

- Red - orange - brown - green (red - blue could be an option);
- Pink, turquoise and grey
- Purple and blue



Source: https://blog.datawrapper.de/beautifulcolors/

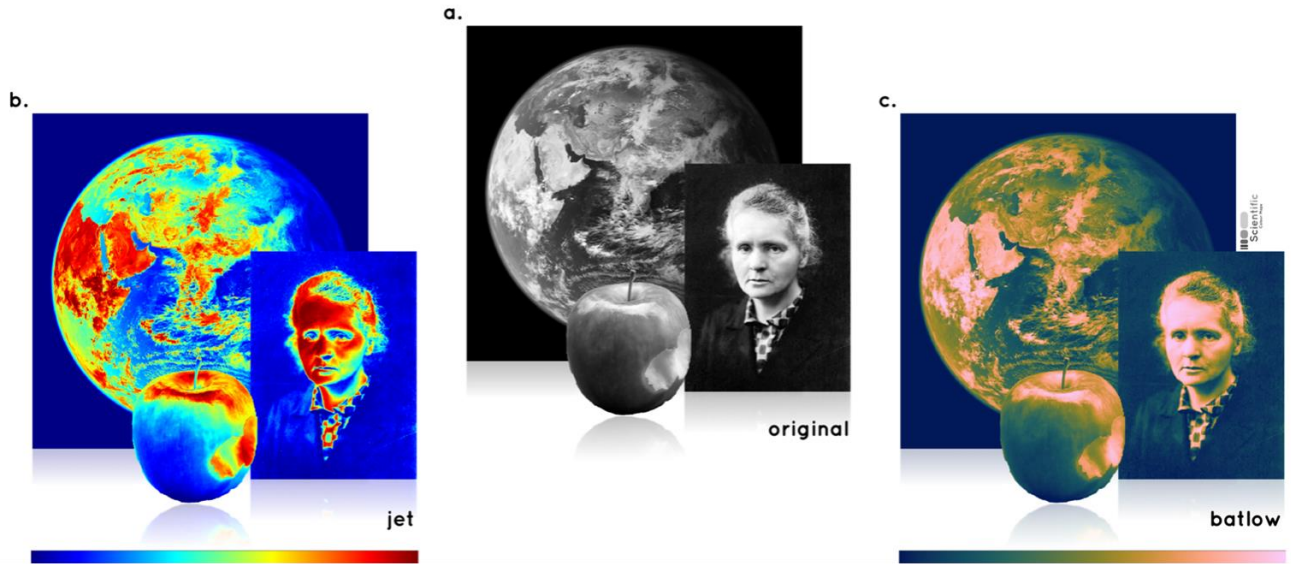
Check your figures:

- https://www.color-blindness.com/
- R package colorblind
- https://contrastchecker.com/

More info: https://blog.datawrapper.de/colorblindness-part1/

Colour in scientific visualization

Avoid rainbow and jet color scales, they lead to uneven color perception!



Crameri et al, 2020, Nature communications

Hydrol. Earth Syst. Sci., 25, 4549–4565, 2021
https://doi.org/10.5194/hess-25-4549-2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Open Access
EGU

Rainbow color map distorts and misleads research in hydrology – guidance for better visualizations and science communication

Michael Stoelzle[1] **and Lina Stein**[2]

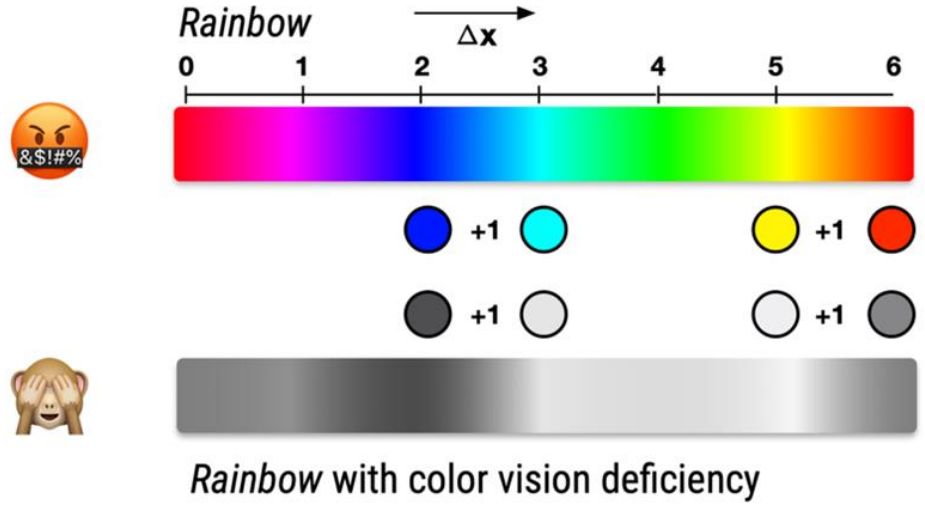
[1]Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany
[2]Department of Civil Engineering, University of Bristol, Bristol, UK

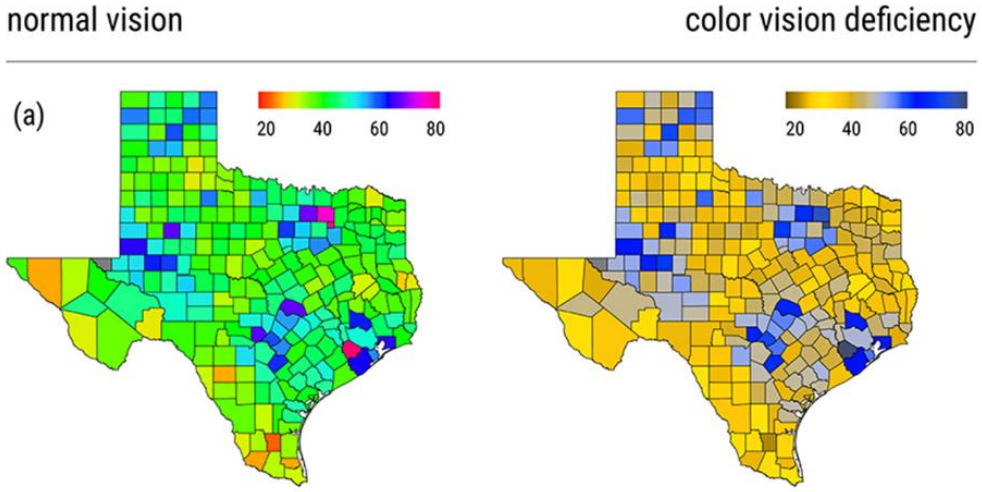
Correspondence: Michael Stoelzle (michael.stoelzle@hydro.uni-freiburg.de)

Received: 27 February 2021 – Discussion started: 9 March 2021
Revised: 21 June 2021 – Accepted: 30 June 2021 – Published: 24 August 2021

Colour in scientific visualization

Avoid rainbow and jet color scales, they do not guarantee accessibility to Color Vision Deficiency (CVD)!



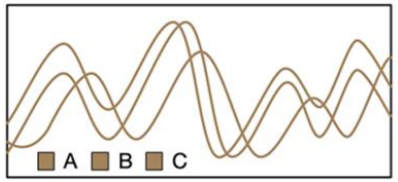


Source: Stoelzle and Stein, 2021, HESS

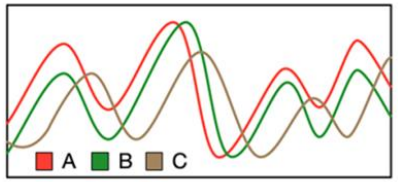
Colour in scientific visualization

What are the alternatives?

With color vision deficiency



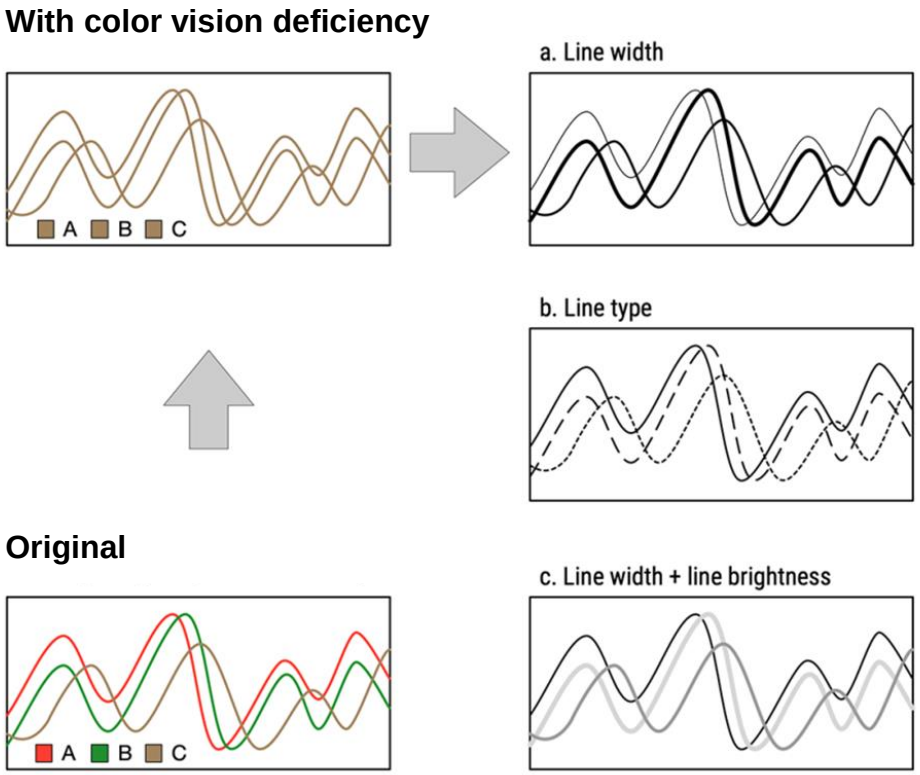
Original



Source: Stoelzle and Stein, 2021, HESS

Colour in scientific visualization

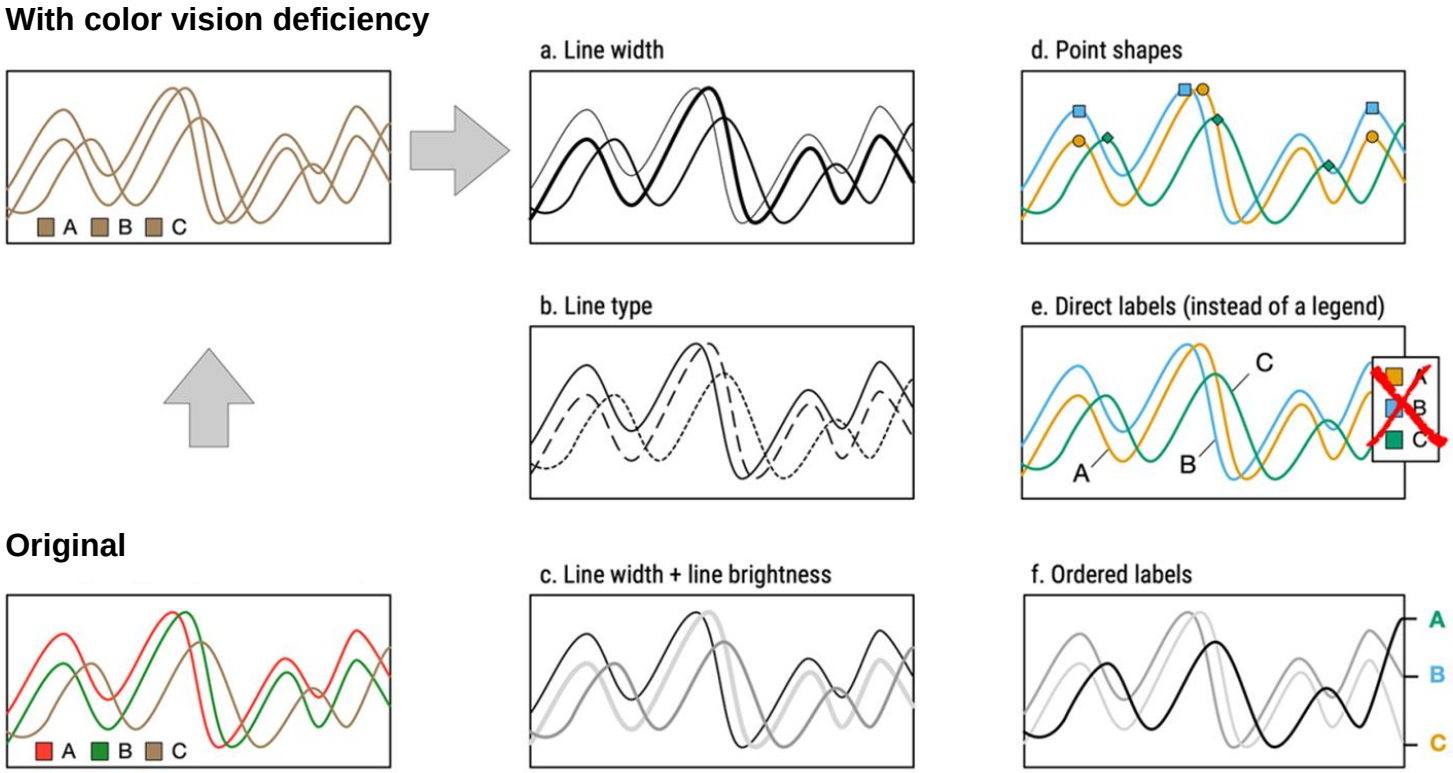
What are the alternatives? **Avoid color use, if not necessary**



Source: Stoelzle and Stein, 2021, HESS

Colour in scientific visualization

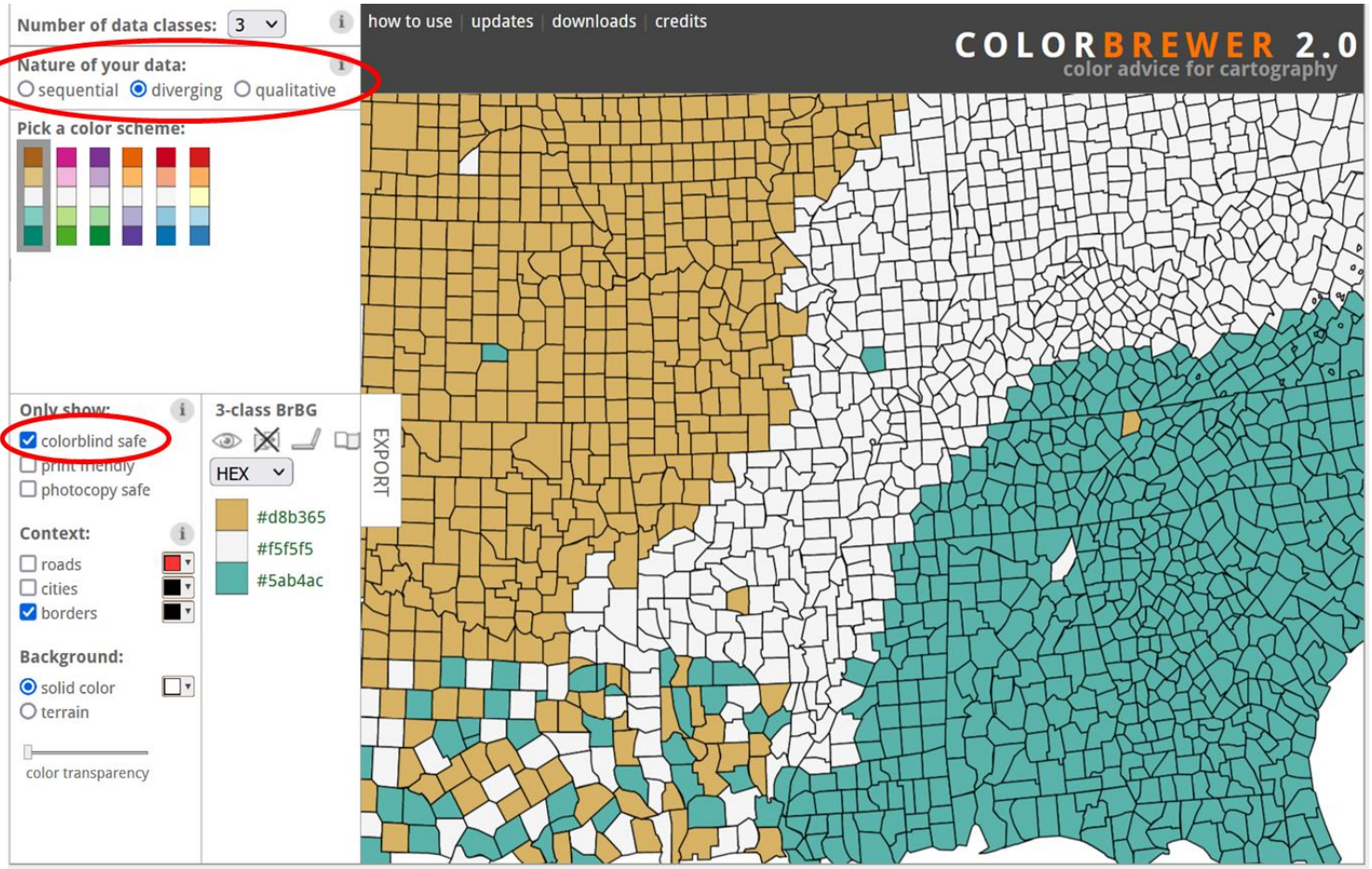
What are the alternatives? **Avoid colors and improve labels**



Source: Stoelzle and Stein, 2021, HESS

Colour in scientific visualization

Online resources: Colorbrewer (colorbrewer2.org)



RColorBrewer

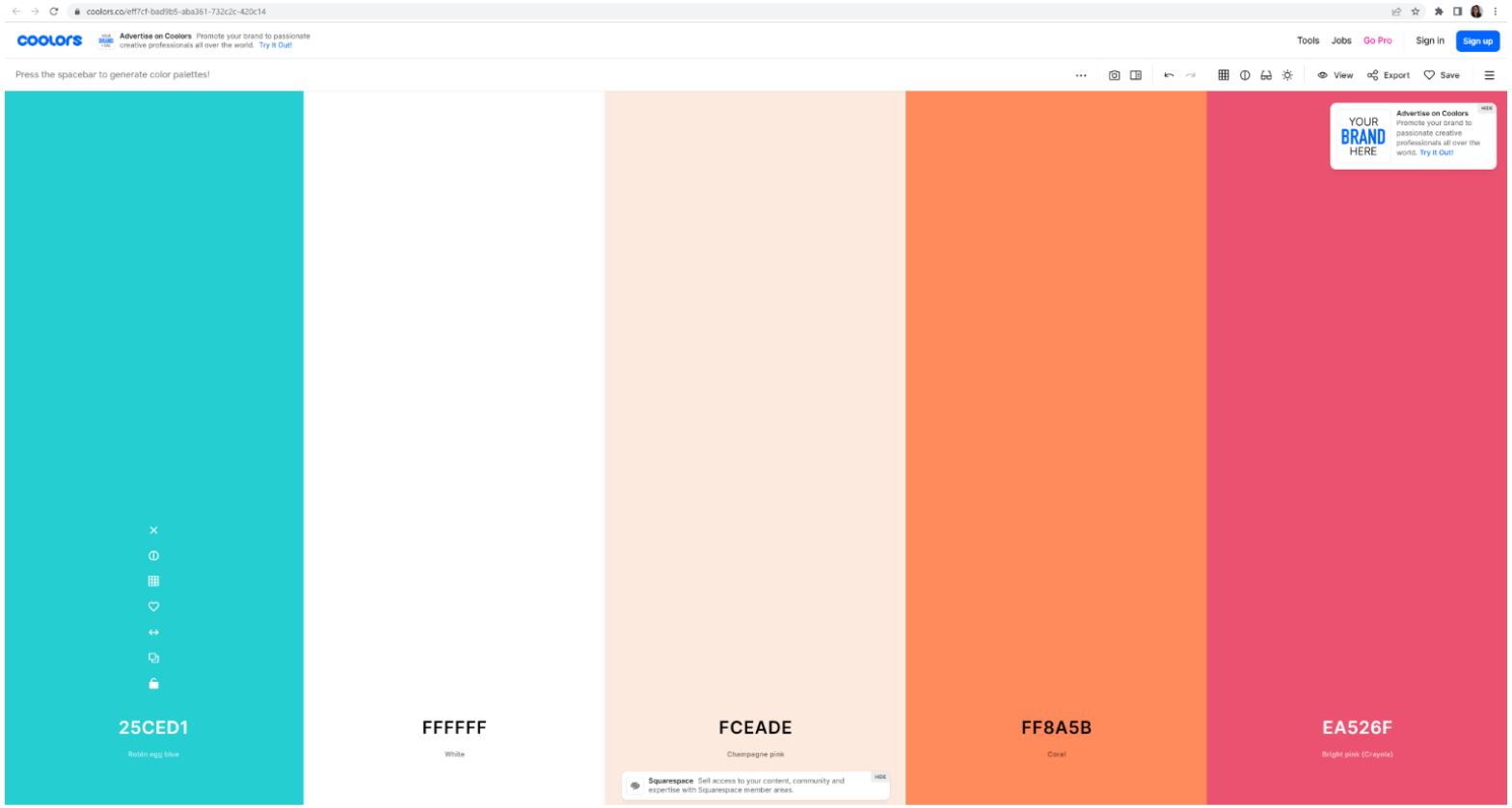
MATLAB®

BREWERMAP Function

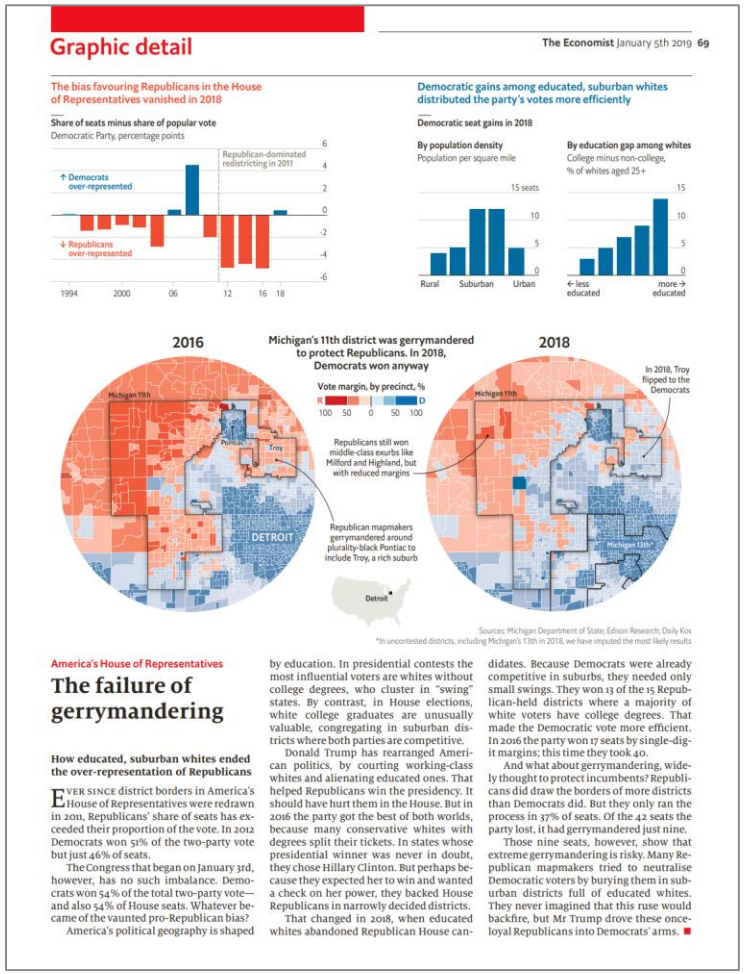
palettable

Colour in scientific visualization

Online resources: Coolors (https://coolors.co) - a color palette generator.



Interesting examples



Graphic detail

The Economist January 5th 2019 69

The bias favouring Republicans in the House of Representatives vanished in 2018

Share of seats minus share of popular vote

Democratic Party, percentage points

↑ Democrats over-represented

↓ Republicans over-represented

Republican-dominated redistricting in 2011

1994 2000 06 12 16 18

Democratic gains among educated, suburban whites distributed the party's votes more efficiently

Democratic seat gains in 2018

By population density

Population per square mile

Rural Suburban Urban

By education gap among whites

College minus non-college, % of whites aged 25+

← less educated more → educated

2016

2018

Michigan's 11th district was gerrymandered to protect Republicans. In 2018, Democrats won anyway

Vote margin, by precinct, %

R D

100 50 0 50 100

Michigan 11th

Pontiac

Troy

DETROIT

Republicans still won middle-class exurbs like Milford and Highland, but with reduced margins

Republican mapmakers gerrymandered around plurality-black Pontiac to include Troy, a rich suburb

In 2018, Troy flipped to the Democrats

Detroit

Sources: Michigan Department of State; Edison Research; Daily Kos

*In uncontested districts, including Michigan's 13th in 2018, we have imputed the most likely results

America's House of Representatives

The failure of gerrymandering

How educated, suburban whites ended the over-representation of Republicans

EVER SINCE district borders in America's House of Representatives were redrawn in 2011, Republicans' share of seats has exceeded their proportion of the vote. In 2012 Democrats won 51% of the two-party vote but just 46% of seats.

The Congress that began on January 3rd, however, has no such imbalance. Democrats won 54% of the total two-party vote—and also 54% of House seats. Whatever became of the vaunted pro-Republican bias?

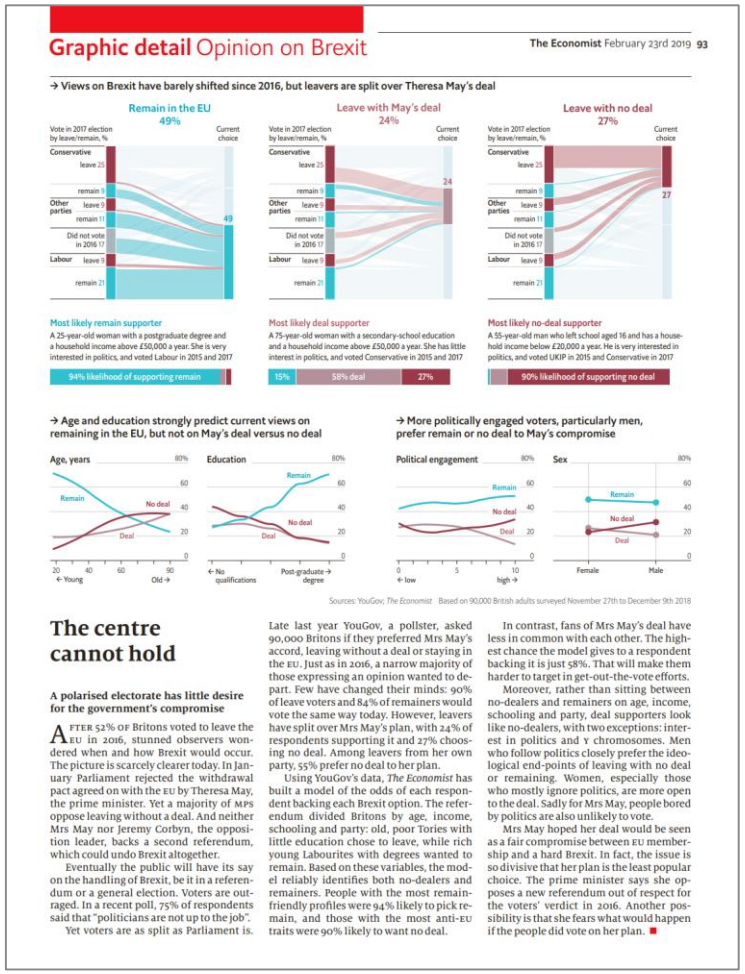
America's political geography is shaped by education. In presidential contests the most influential voters are whites without college degrees, who cluster in "swing" states. By contrast, in House elections, white college graduates are unusually valuable, congregating in suburban districts where both parties are competitive.

Donald Trump has rearranged American politics, by courting working-class whites and alienating educated ones. That helped Republicans win the presidency. It should have hurt them in the House. But in 2016 the party got the best of both worlds, because many conservative whites with degrees split their tickets. In states whose presidential winner was never in doubt, they chose Hillary Clinton. But perhaps because they expected her to win and wanted a check on her power, they backed House Republicans in narrowly decided districts.

That changed in 2018, when educated whites abandoned Republican House candidates. Because Democrats were already competitive in suburbs, they needed only small swings. They won 13 of the 15 Republican-held districts where a majority of white voters have college degrees. That made the Democratic vote more efficient. In 2016 the party won 17 seats by single-digit margins; this time they took 40.

And what about gerrymandering, widely thought to protect incumbents? Republicans did draw the borders of more districts than Democrats did. But they only ran the process in 37% of seats. Of the 42 seats the party lost, it had gerrymandered just nine.

Those nine seats, however, show that extreme gerrymandering is risky. Many Republican mapmakers tried to neutralise Democratic voters by burying them in suburban districts full of educated whites. They never imagined that this ruse would backfire, but Mr Trump drove these once-loyal Republicans into Democrats' arms. ■



Graphic detail Opinion on Brexit

The Economist February 23rd 2019 93

→ Views on Brexit have barely shifted since 2016, but leavers are split over Theresa May's deal

Remain in the EU 49%

Leave with May's deal 24%

Leave with no deal 27%

Vote in 2017 election by leave/remain, %

Current choice

Conservative leave 25 remain 9

Other parties leave 9 remain 11

Did not vote in 2016 17

Labour leave 9 remain 21

49 24 27

Most likely remain supporter

A 25-year-old woman with a postgraduate degree and a household income above £50,000 a year. She is very interested in politics, and voted Labour in 2015 and 2017

94% likelihood of supporting remain

Most likely deal supporter

A 75-year-old woman with a secondary-school education and a household income above £50,000 a year. She has little interest in politics, and voted Conservative in 2015 and 2017

15% 58% deal 27%

Most likely no-deal supporter

A 55-year-old man who left school aged 16 and has a household income below £20,000 a year. He is very interested in politics, and voted UKIP in 2015 and Conservative in 2017

90% likelihood of supporting no deal

→ Age and education strongly predict current views on remaining in the EU, but not on May's deal versus no deal

Age, years

Remain

No deal

Deal

20 40 60 80

← Young Old →

Education

No qualifications Post-graduate degree →

→ More politically engaged voters, particularly men, prefer remain or no deal to May's compromise

Political engagement

← low high →

Sex

Female Male

Sources: YouGov; *The Economist* Based on 90,000 British adults surveyed November 27th to December 9th 2018

The centre cannot hold

A polarised electorate has little desire for the government's compromise

AFTER 52% OF Britons voted to leave the EU in 2016, stunned observers wondered when and how Brexit would occur. The picture is scarcely clearer today. In January Parliament rejected the withdrawal pact agreed on with the EU by Theresa May, the prime minister. Yet a majority of MPs oppose leaving without a deal. And neither Mrs May nor Jeremy Corbyn, the opposition leader, backs a second referendum, which could undo Brexit altogether.

Eventually the public will have its say on the handling of Brexit, be it in a referendum or a general election. Voters are outraged. In a recent poll, 75% of respondents said that "politicians are not up to the job".

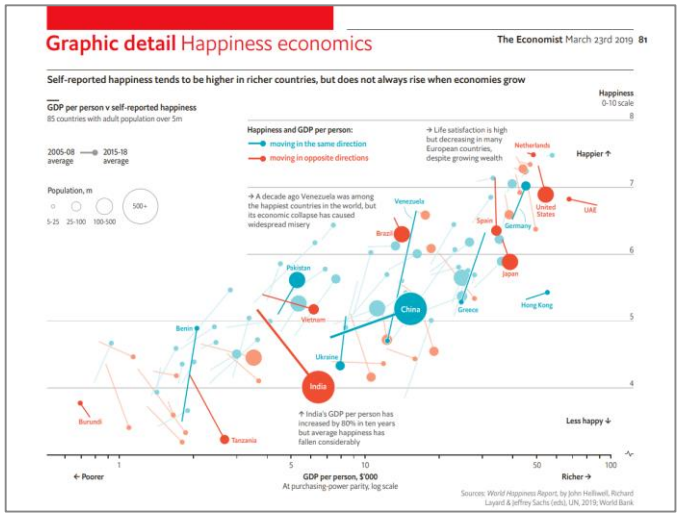
Yet voters are as split as Parliament is. Late last year YouGov, a pollster, asked 90,000 Britons if they preferred Mrs May's accord, leaving without a deal or staying in the EU. Just as in 2016, a narrow majority of those expressing an opinion wanted to depart. Few have changed their minds: 90% of leave voters and 84% of remainers would vote the same way today. However, leavers have split over Mrs May's plan, with 24% of respondents supporting it and 27% choosing no deal. Among leavers from her own party, 55% prefer no deal to her plan.

Using YouGov's data, *The Economist* has built a model of the odds of each respondent backing each Brexit option. The referendum divided Britons by age, income, schooling and party: old, poor Tories with little education chose to leave, while rich young Labourites with degrees wanted to remain. Based on these variables, the model reliably identifies both no-dealers and remainers. People with the most remain-friendly profiles were 94% likely to pick remain, and those with the most anti-EU traits were 90% likely to want no deal.

In contrast, fans of Mrs May's deal have less in common with each other. The highest chance the model gives to a respondent backing it is just 58%. That will make them harder to target in get-out-the-vote efforts.

Moreover, rather than sitting between no-dealers and remainers on age, income, schooling and party, deal supporters look like no-dealers, with two exceptions: interest in politics and Y chromosomes. Men who follow politics closely prefer the ideological end-points of leaving with no deal or remaining. Women, especially those who mostly ignore politics, are more open to the deal. Sadly for Mrs May, people bored by politics are also unlikely to vote.

Mrs May hoped her deal would be seen as a fair compromise between EU membership and a hard Brexit. In fact, the issue is so divisive that her plan is the least popular choice. The prime minister says she opposes a new referendum out of respect for the voters' verdict in 2016. Another possibility is that she fears what would happen if the people did vote on her plan. ■



Graphic detail Happiness economics

The Economist March 23rd 2019 81

Self-reported happiness tends to be higher in richer countries, but does not always rise when economies grow

GDP per person v self-reported happiness

85 countries with adult population over 5m

Happiness 0-10 scale

Happiness and GDP per person:

moving in the same direction

moving in opposite directions

2005-08 average → 2015-18 average

Population, m

5-25 25-100 100-500 500+

→ A decade ago Venezuela was among the happiest countries in the world, but its economic collapse has caused widespread misery

→ Life satisfaction is high but decreasing in many European countries, despite growing wealth

Happier ↑

Less happy ↓

Netherlands United States UAE Venezuela Brazil Spain Germany Japan Pakistan Vietnam China Greece Hong Kong Benin Ukraine India Burundi Tanzania

→ India's GDP per person has increased by 80% in ten years but average happiness has fallen considerably

← Poorer

GDP per person, $'000

At purchasing-power parity, log scale

Richer →

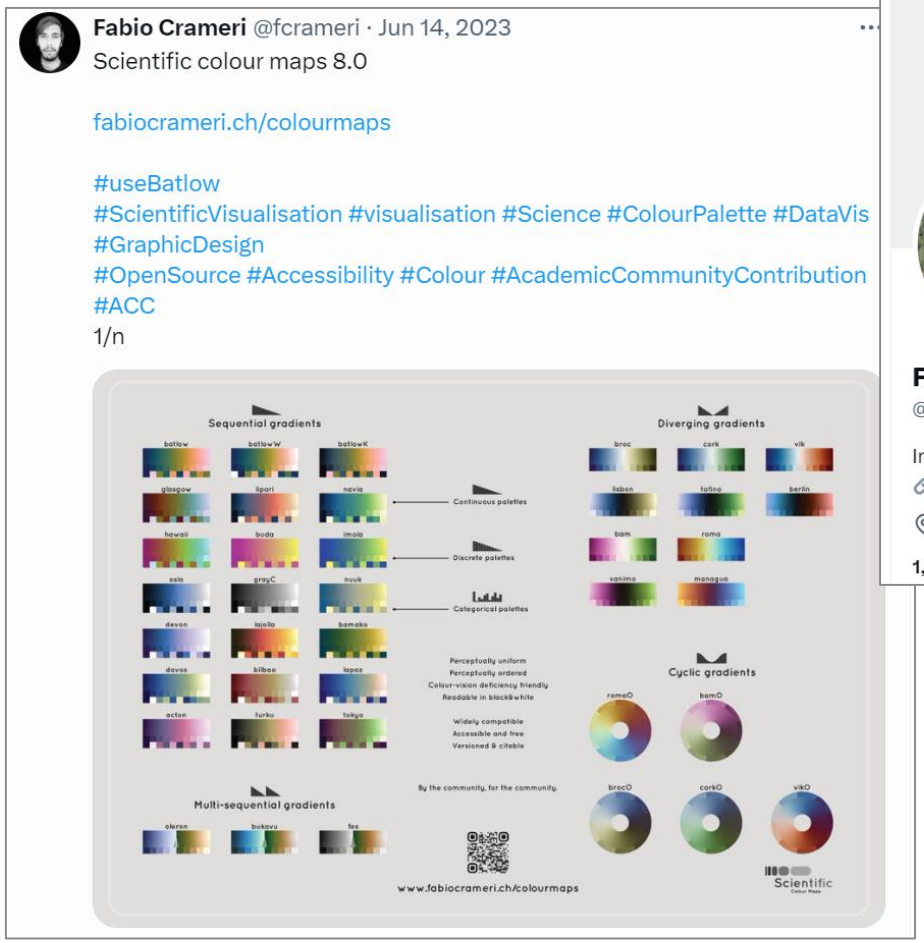
Sources: *World Happiness Report*, by John Helliwell, Richard Layard & Jeffrey Sachs (eds); UN, 2019; World Bank

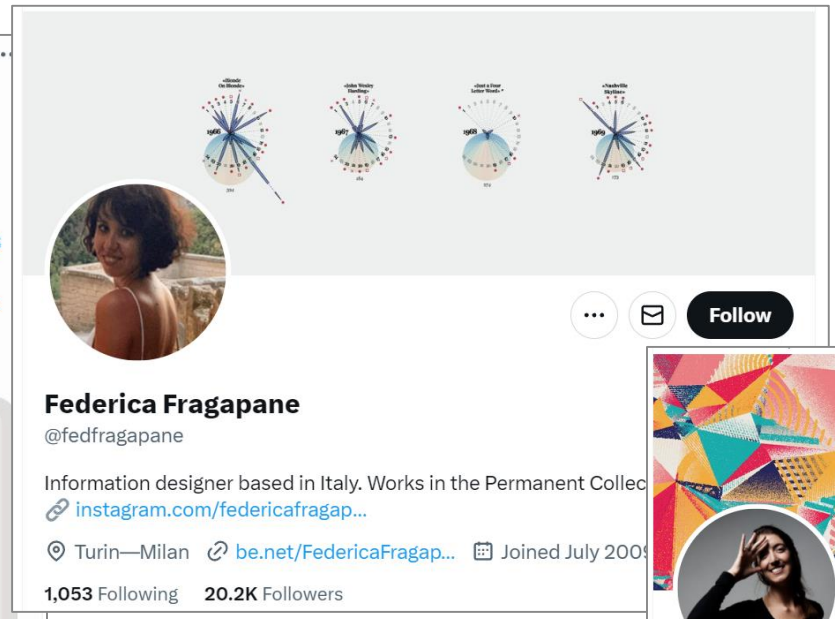
SOURCE: https://infographics.economist.com/2019/AChristmasGiftForYou/AYearInGraphicDetail.pdf

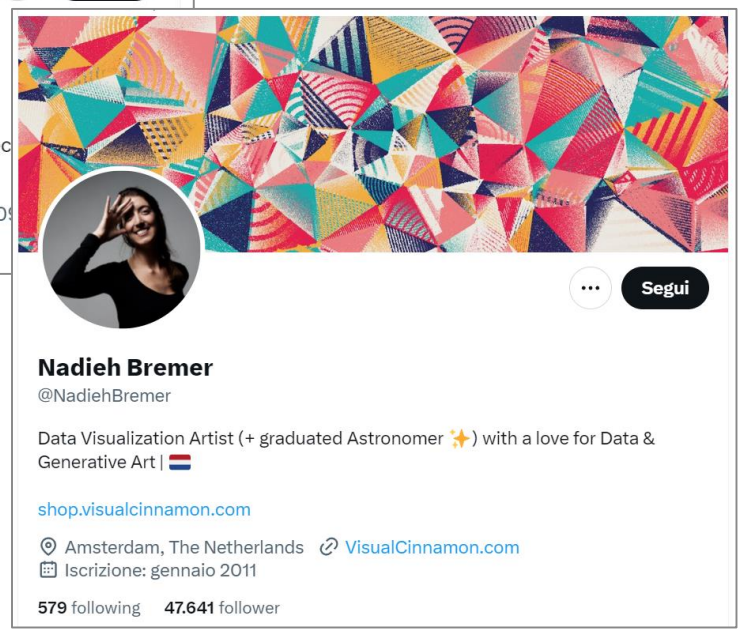
Interesting examples

SOURCE: https://multimedia.scmp.com/culture/article/SCMP-printed-graphics-memory/

Interesting examples







Colour in scientific visualization

Do we even need a fixed color palette for our data visualizations?

This choice is up to you!

There are lots of good reasons to set colors in your data vis style guide. Once a good color palette is decided:

- visualizations will look **more consistent** no matter which tool you're using to create them;
- you and your team members will be able to **think less about colors while creating charts** for your works;
- consistent colors can look better than what you would choose in the three minutes before a deadline.

Source: https://blog.datawrapper.de/colors-for-data-vis-style-guides/

Colour in scientific visualization

Not having a strict color palette is an option, too.

There are good reasons for it, e.g. when it’s not important that your visualizations have a strong visual identity, for whatever reason.

In this case, choose the colors following the basic principles of data viz!

Source: https://blog.datawrapper.de/colors-for-data-vis-style-guides/

3. Publication compliance

Edoardo

Rights, Permissions & Licenses

Open Access vs Subscription-Based

Creative Commons Attribution 4.0 International (CC BY 4.0)

You are free to:

- Share (copy and redistribute the material in any medium or format)
- Adapt (remix, transform, and build upon the material for any purpose, even commercially)

Under the following terms:

- Attribution (you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use)
- No additional restrictions (you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits)

Rights, Permissions & Licenses

Open Access vs Subscription-Based

Who owns the copyrights? Most likely not you (the author)…

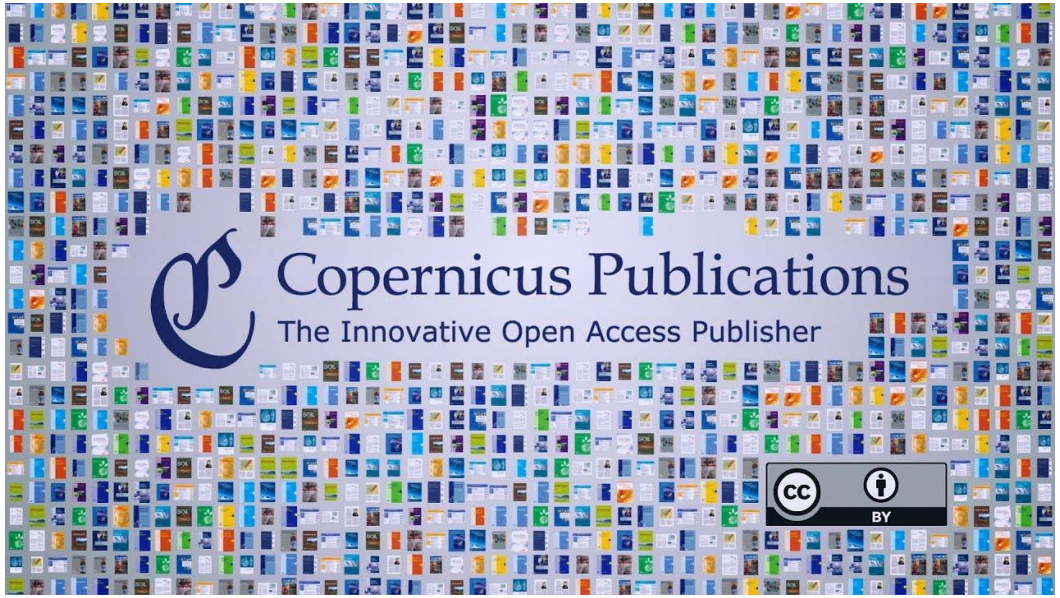
You must get permission from the publisher!

(*Copyright Clearance Center's RightsLink® service*)

- During the publication process, authors may be asked to sign the publisher's "Copyright Transfer Agreement."
- Read the terms of the Copyright Transfer Agreement or Exclusive License Form
- When in doubt, ask for support!

DataViz from the publisher’s perspective

Sarah Schneemann & Natascha Töpfer



Copernicus, Editorial Support (typesetting, image processing)

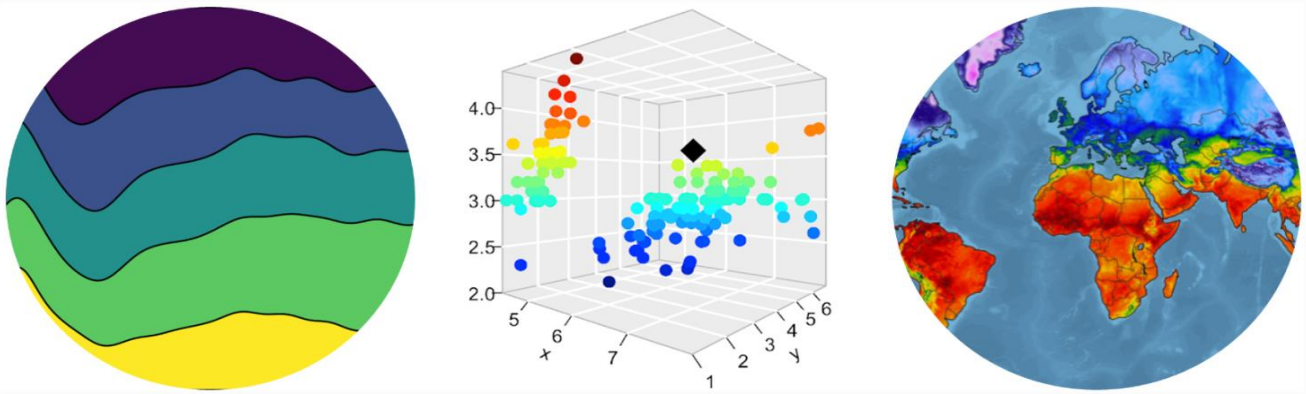
4. DataViz with R

Debasish

Details of Large-Scale Geospatial Analysis

Contact Details:
debmishra@tamu.edu

Data Visualization and Geospatial Analysis With R



https://rpubs.com/Vinit_Sehgal/lgar23

Thinking Parallel: High Performance Computing (HPC) for Researchers

Debasish Mishra
Doctoral Research Student @ Vadose Zone Research Group, Texas A&M University | ESIP Community Fellow | Silver Medalist @IIT Kgp

3 articles

5. DataViz with NCL

Roshanak

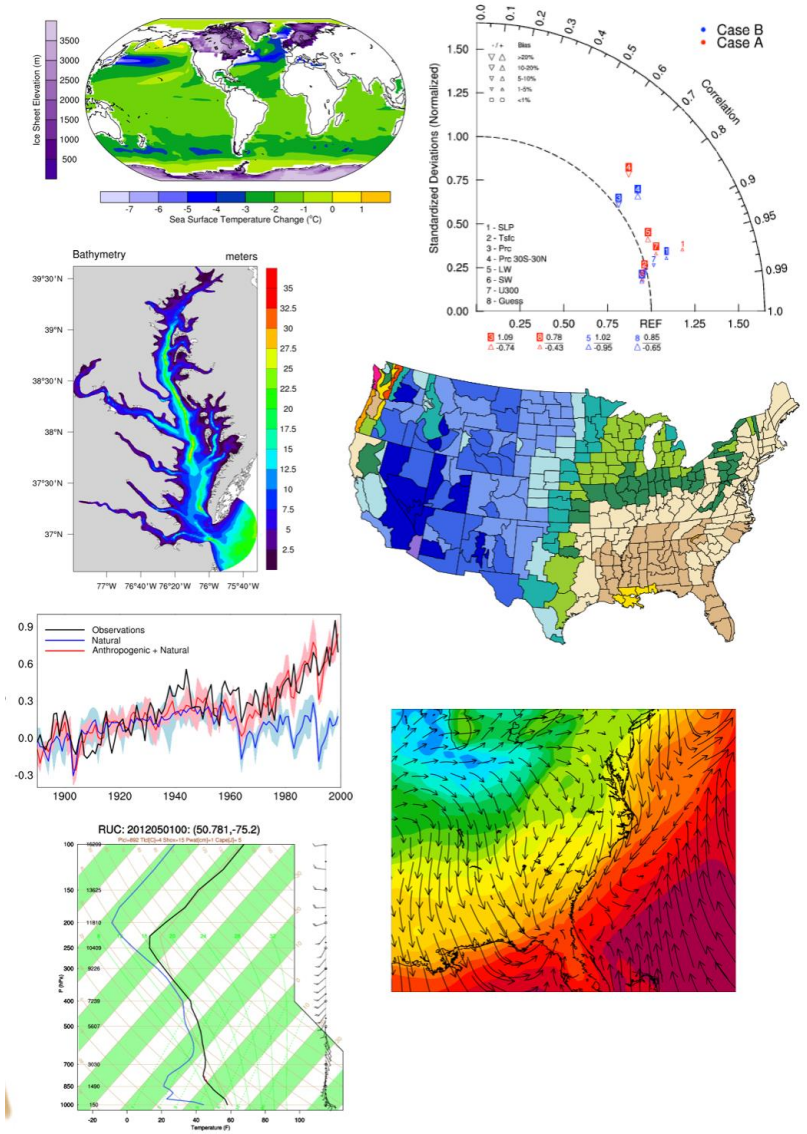
NCAR Command Language (NCL)

NCL is an interpreted language designed for data analysis and visualization.

NCL is open source; It is available for MacOS, Linux, and Windows systems running the Windows 10 Linux subsystem.

It supports NetCDF, GRIB, Shapefile, etc.

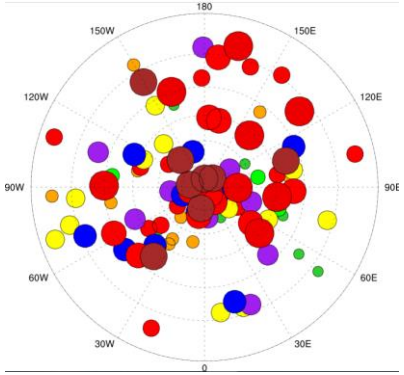
There are a lot of useful built-in functions. It's got many graphic resources, and high quality graphics can be created.



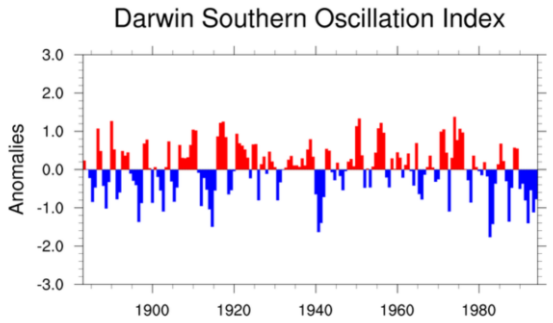
NCAR Command Language (NCL)

Examples and Application

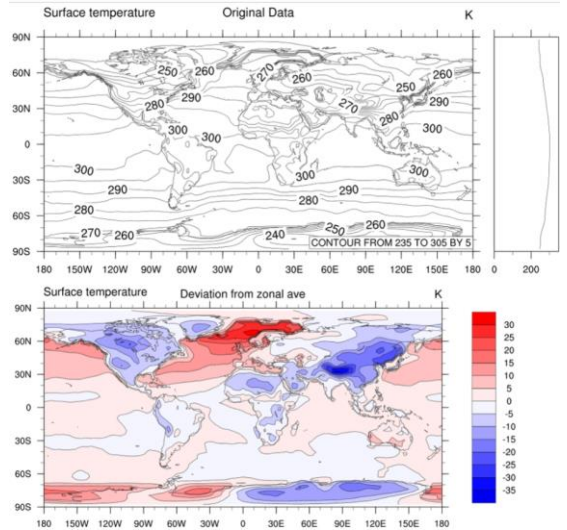
Scatter plot



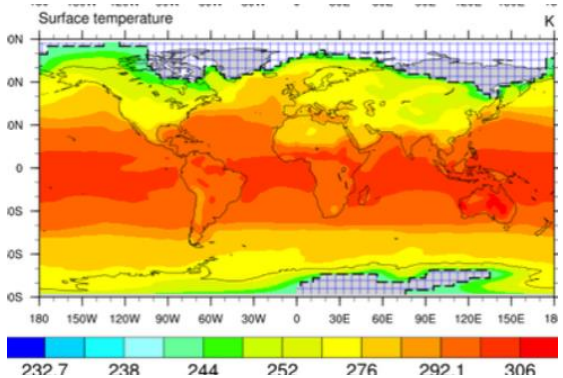
Bar charts



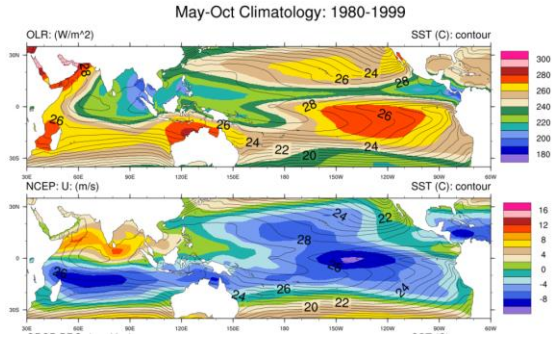
Anomalies



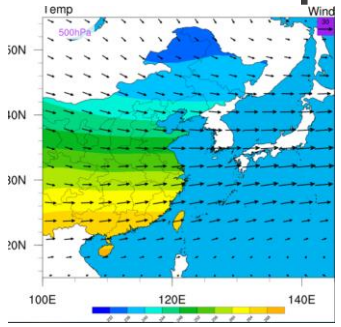
Contour plot



Climatology



Vector map



NCAR Command Language (NCL)

Where and how?

SCAN ME

or

Go to this address: https://www.ncl.ucar.edu/

Applications and examples: https://www.ncl.ucar.edu/Applications/

Thank you!

Edoardo Martini (University of Leipzig, Germany)
Paola Mazzoglio (Politecnico di Torino, Italy)
Epari Ritesh Patro (University of Oulu, Finland)
Roshanak Tootoonchi (University of Trento, Italy)
Debasish Mishra (Texas A&M University, USA)

Young Hydrologic Society

younghydrologicsociety@gmail.com
www.younghs.com
@younghydrology