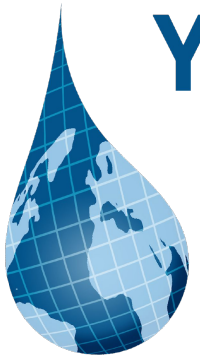
SC3.9
DataViz: Visualise your data effectively and avoid common pitfalls

Swamini Khurana (Stockholm University, Sweden)
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)



Young Hydrologic Society

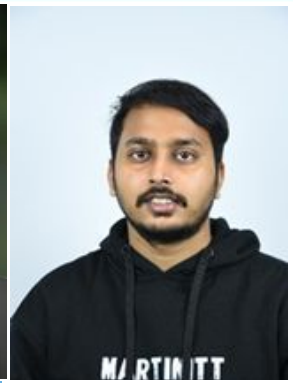
EGU General Assembly 2023

younghs.com

@YoungHydrology

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. Fundaments of DataViz

Interactive block #1

1. Different approaches for visualization - publication figures vs posters vs graphical abstracts
2. Publication compliance

Interactive block #2

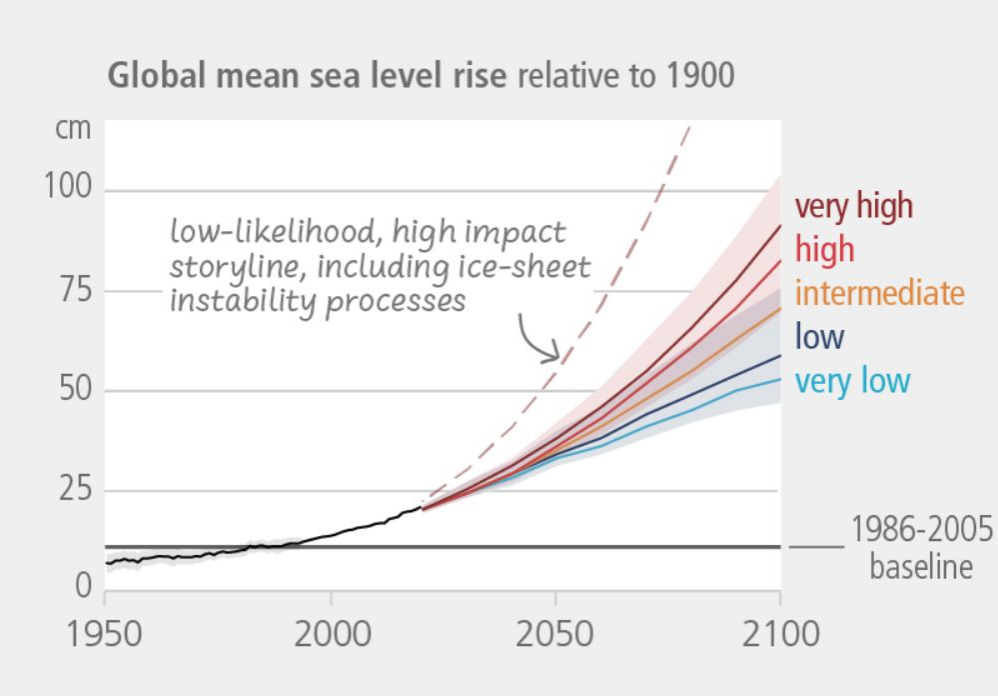
1. Colour schemes
2. Panoply

Interactive block #3

What is DataViz?

“DataViz” = Data Visualization

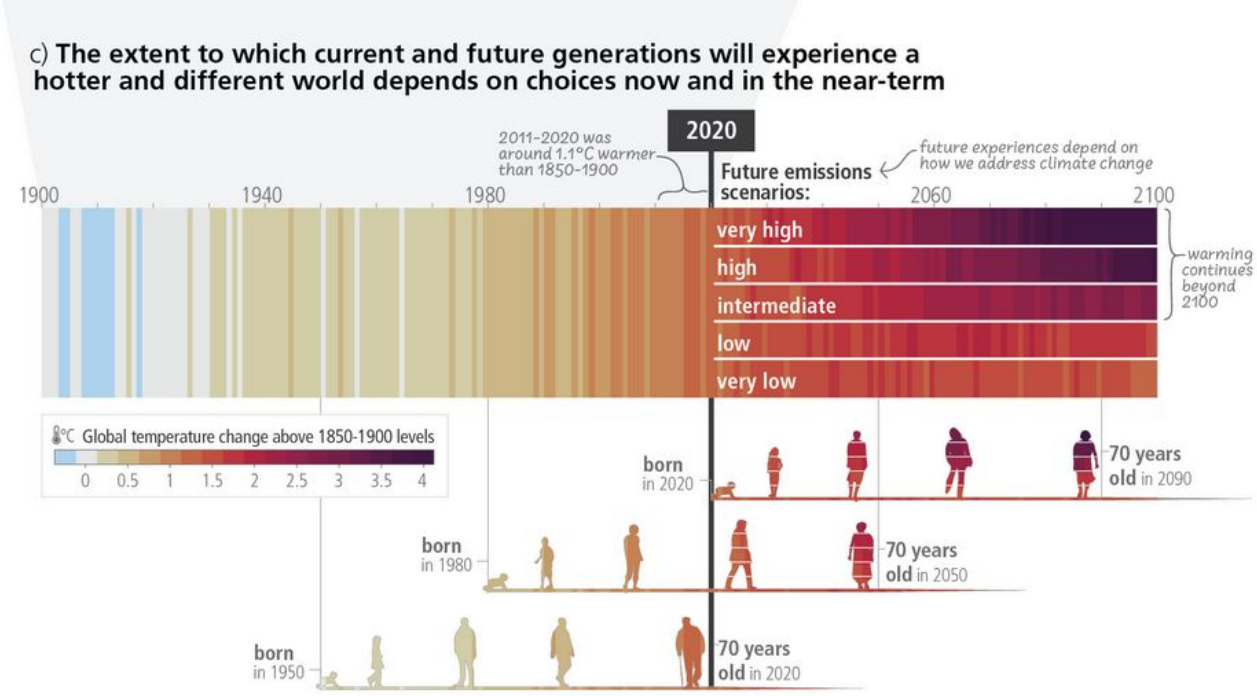
graphical representation of scientific data



Source: IPCC AR6 Synthesis Report, Climate Change 2023

“InfViz” = Information Visualization

graphical representation of scientific concepts, incl. abstract concepts



Brief history of DataViz

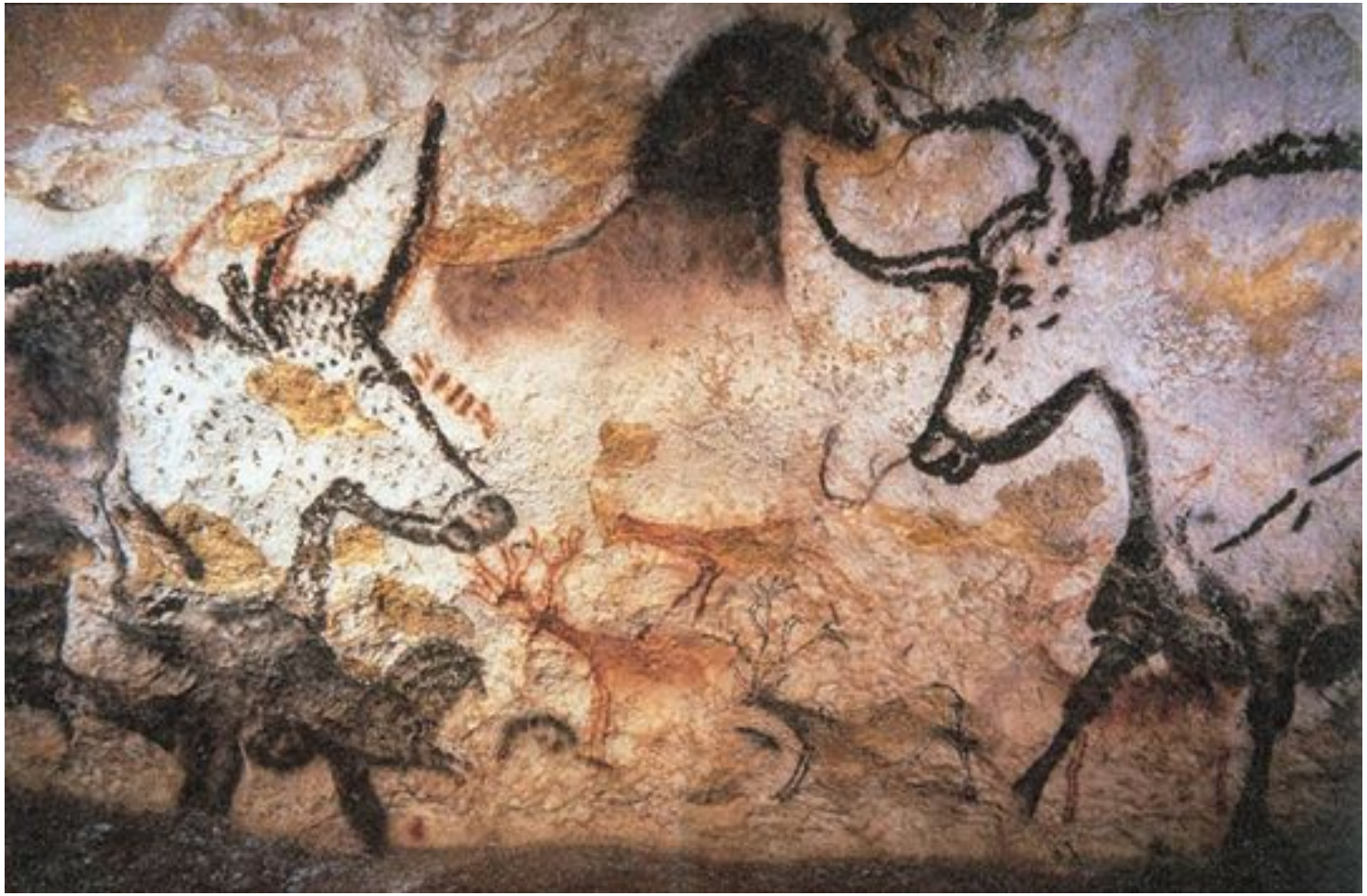
Edoardo

Brief history of DataViz

Upper Paleolithic
(50-12k years ago)

Lascaux Cave, France

Parietal wall painting



Source: https://en.wikipedia.org/wiki/Lascaux#/media/File:Lascaux_painting.jpg

Brief history of DataViz

700-500 BC, Mesopotamia

Imago Mundi

Babylonian map of the „world“

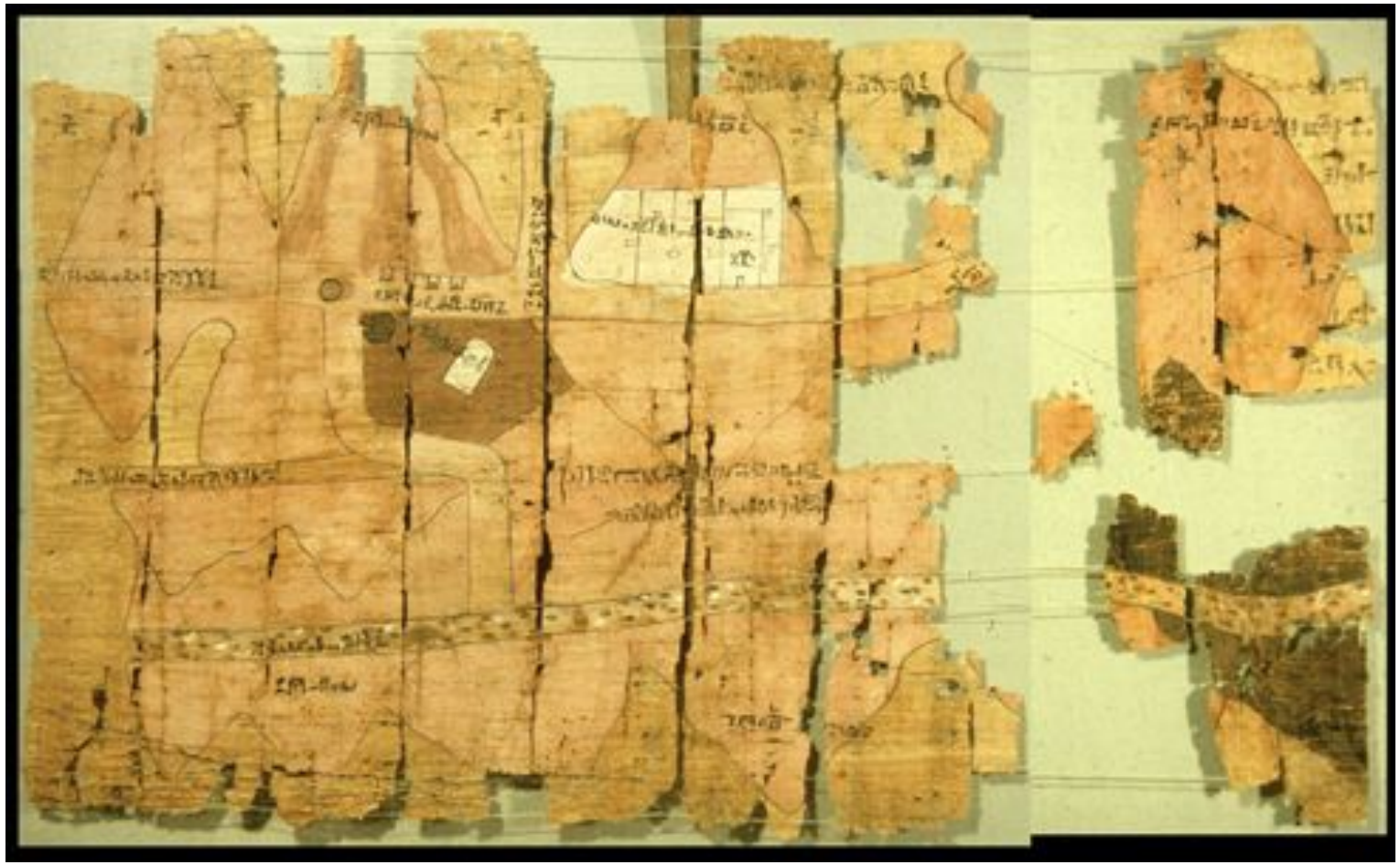
Source: https://en.wikipedia.org/wiki/Babylonian_Map_of_the_World#/media/File:The_Babylonian_map_of_the_world,_from_Sippar,_Mesopotamia..JPG

Brief history of DataViz

Ca. 1150 BC, Amennakhte, Egypt

Turin Papyrus Map

The oldest surviving map of geological interest from the ancient world



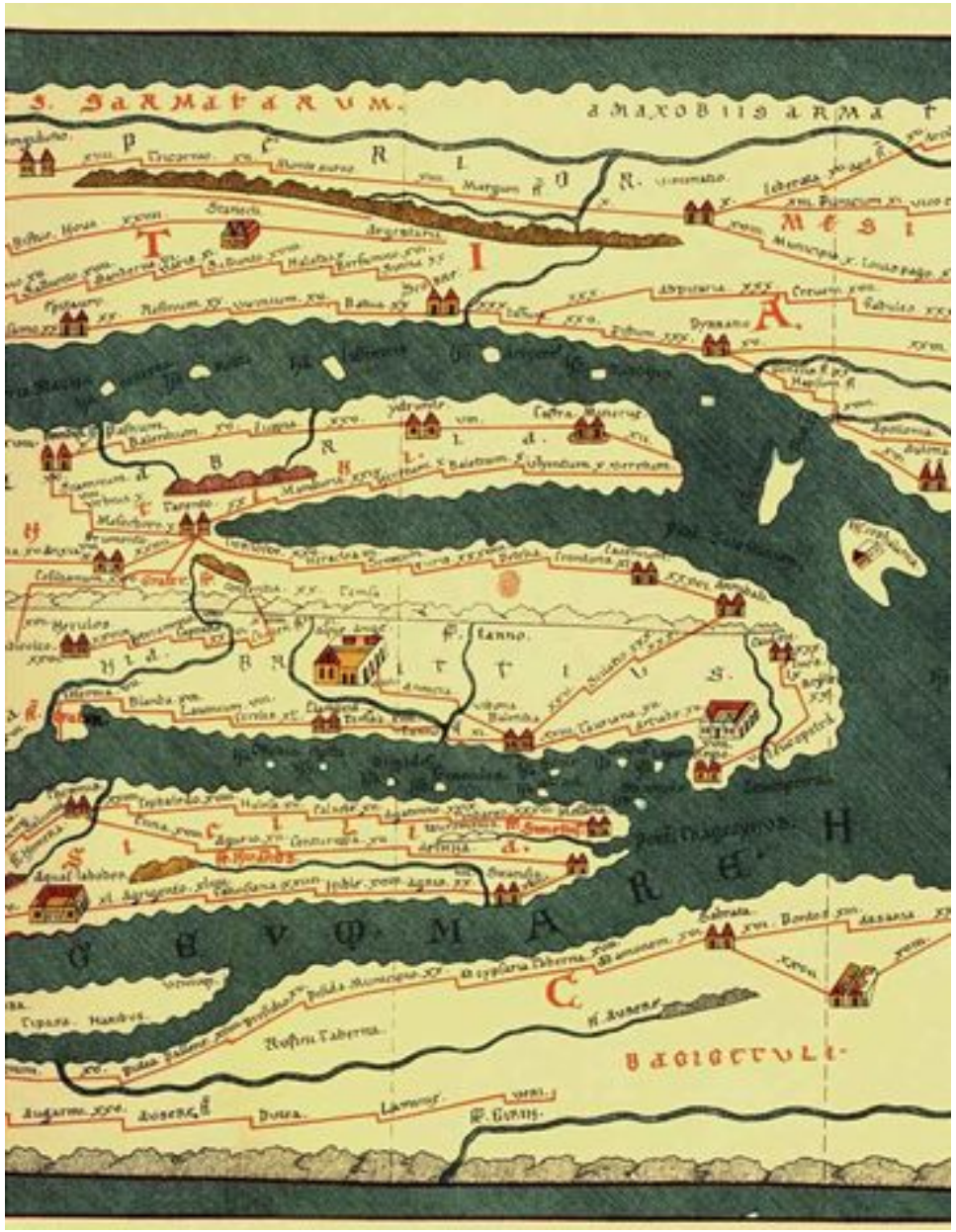
Source: https://en.wikipedia.org/wiki/Turin_Papyrus_Map#/media/File:TurinPapyrus1.jpg

Brief history of DataViz

27 BC – AD 14, Agrippa, Rome

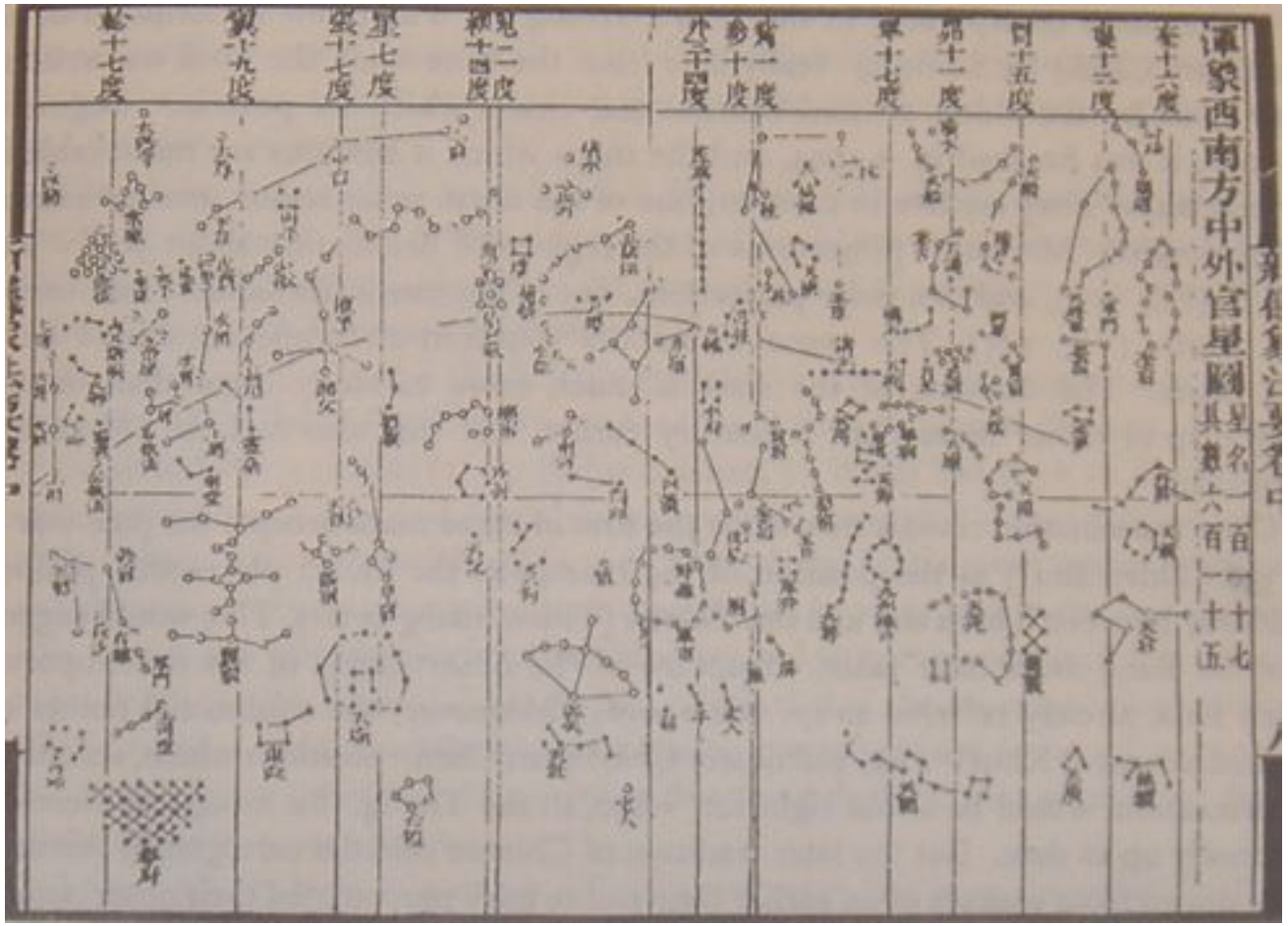
Tabula Peutingeriana

The road network of the Roman Empire



Source: https://en.wikipedia.org/wiki/Tabula_Peutingeriana#/media/File:Part_of_Tabula_Peutingeriana.jpg

Brief history of DataViz



Source: https://en.wikipedia.org/wiki/Chinese_astronomy#/media/File:Su_Song_Star_Map_1.JPG

1092, Su Song, China

Celestial atlas

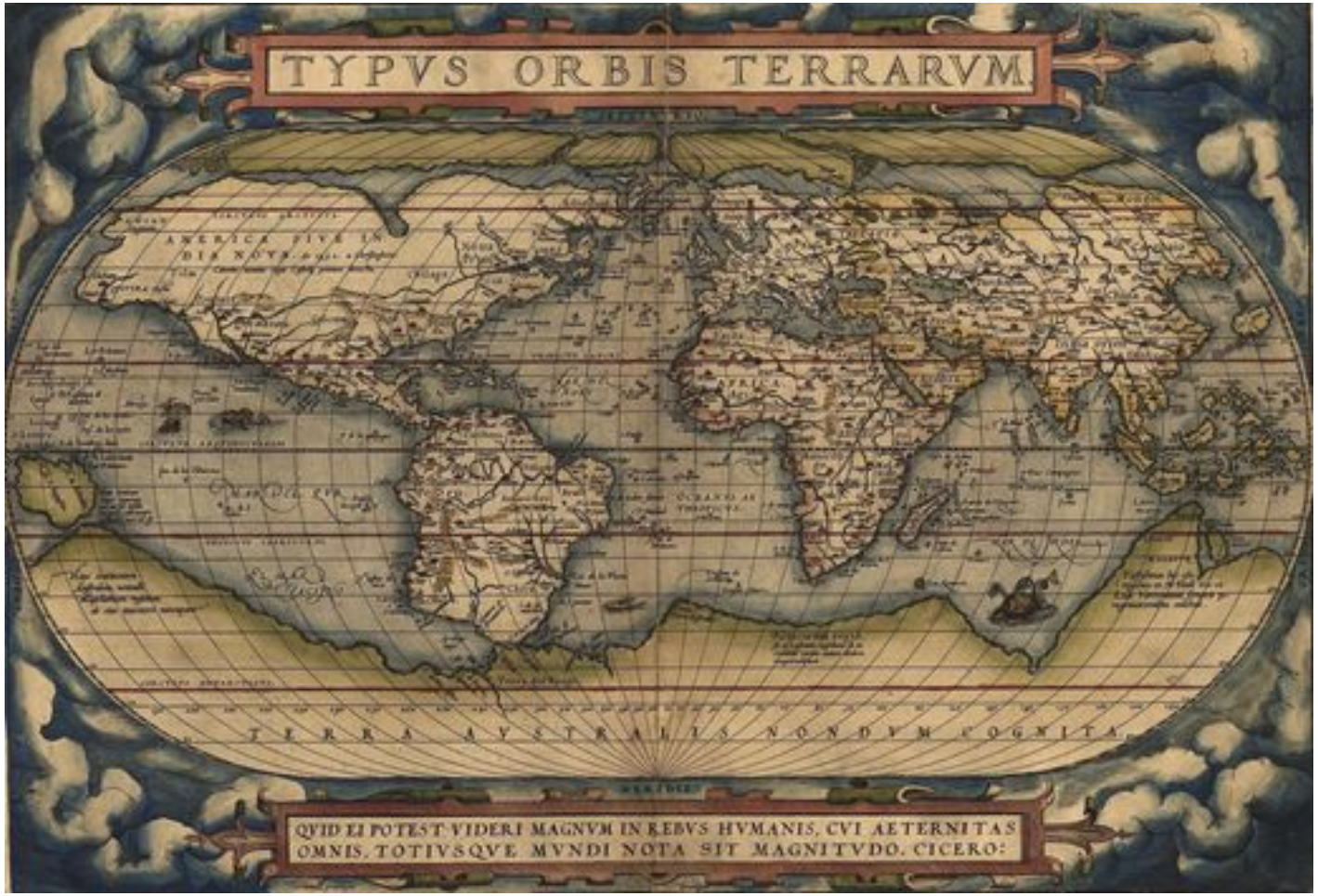
One of the oldest existent star maps in printed form

Brief history of DataViz

1570, Abraham Ortelius, Belgium

Typus Orbis Terrarum

The first world map

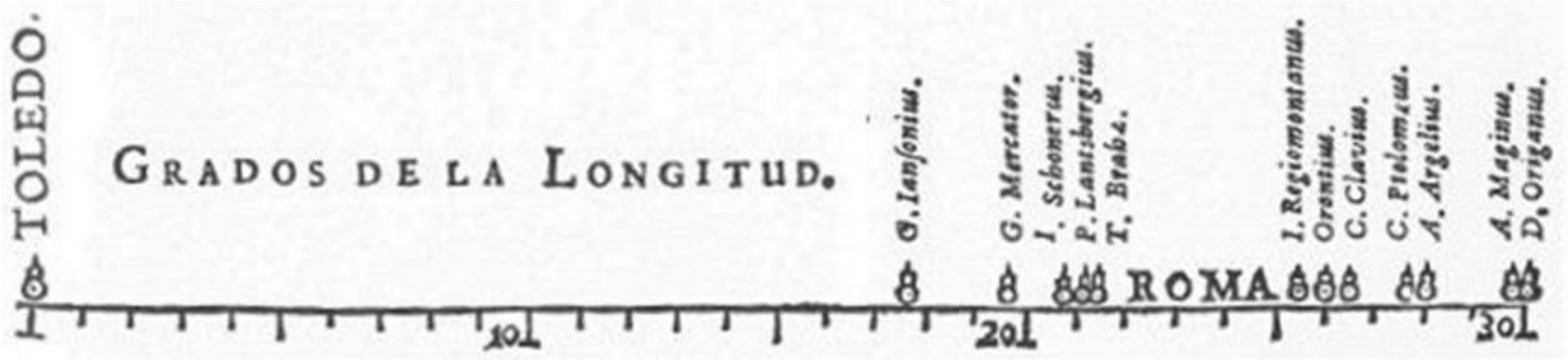


Source: https://en.wikipedia.org/wiki/Abraham_Ortelius#/media/File:OrteliusWorldMap1570.jpg

Brief history of DataViz

1644, Michael Florent van Langren, Belgium

The first known graph of statistical data: difference in longitude between Toledo and Rome, many estimates

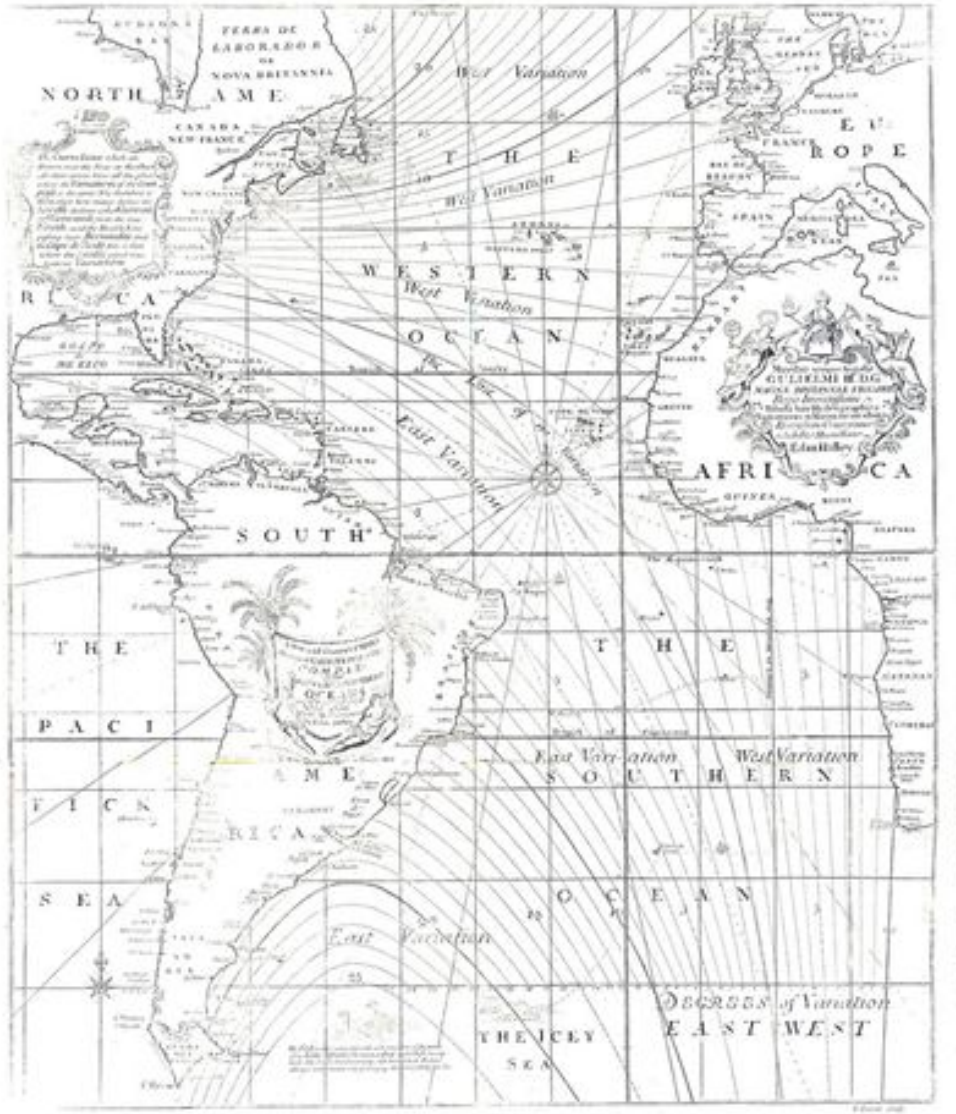


Source: https://en.wikipedia.org/wiki/Michael_van_Langren#/media/File:Grados_de_la_Longitud.jpg

Brief history of DataViz

1701, Edmond Halley, England

The first known isolines (contour maps)

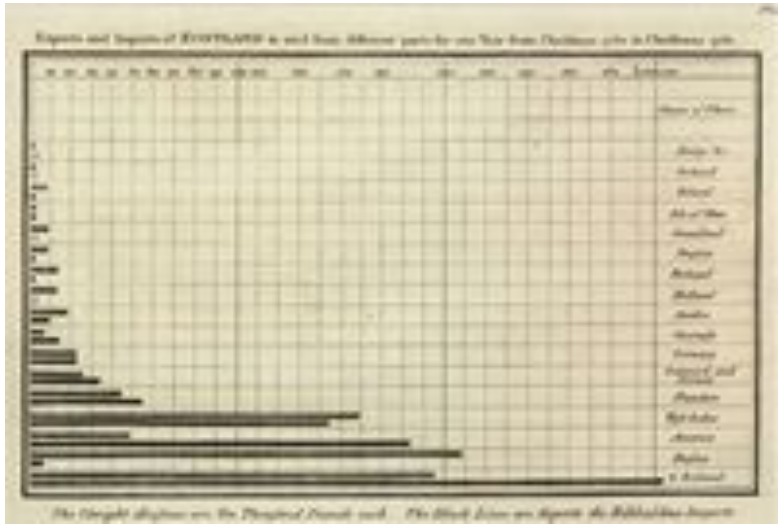


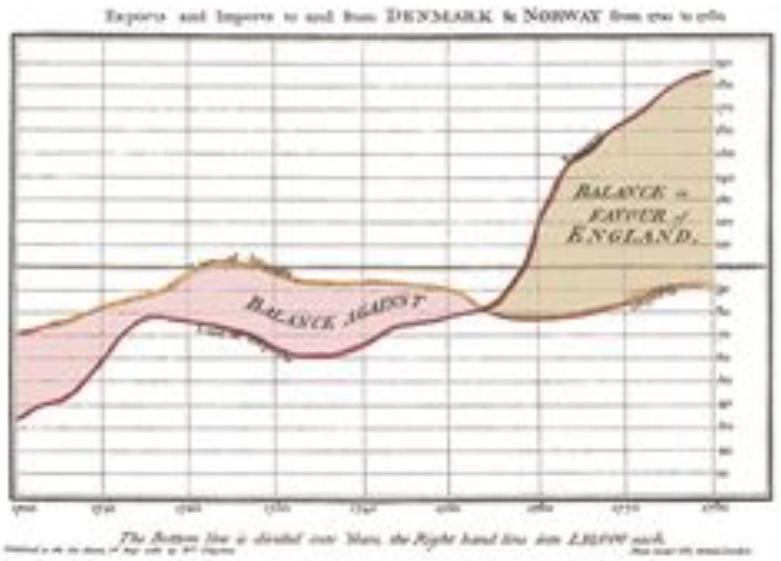
Source: https://en.wikipedia.org/wiki/Edmond_Halley#/media/File:Halley_isogonic_1701.jpg

Brief history of DataViz

1759-1823, William Playfair, Scotland/England

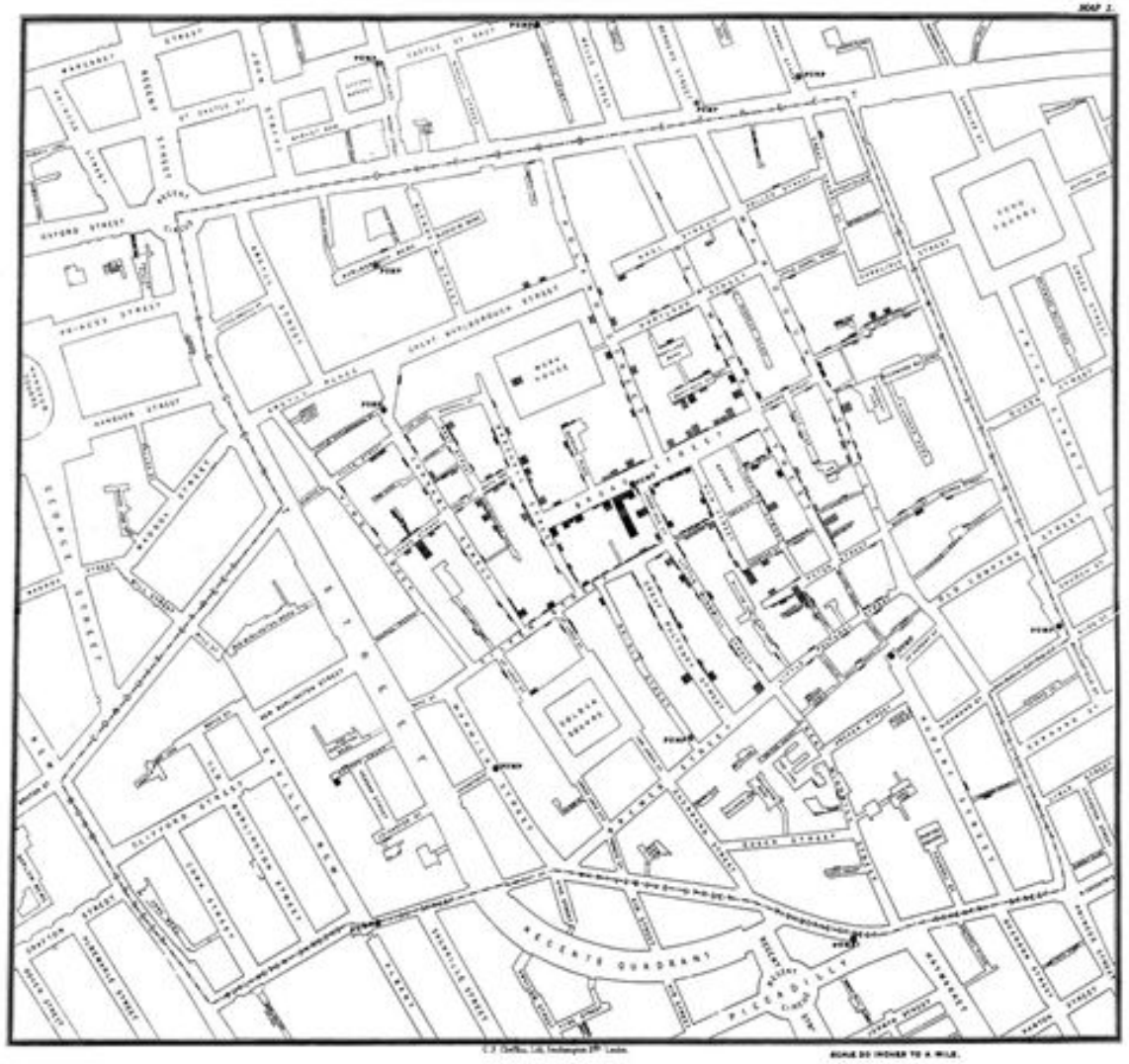
The father of graphical visualization of statistical data





Source: https://en.wikipedia.org/wiki/William_Playfair

Brief history of DataViz



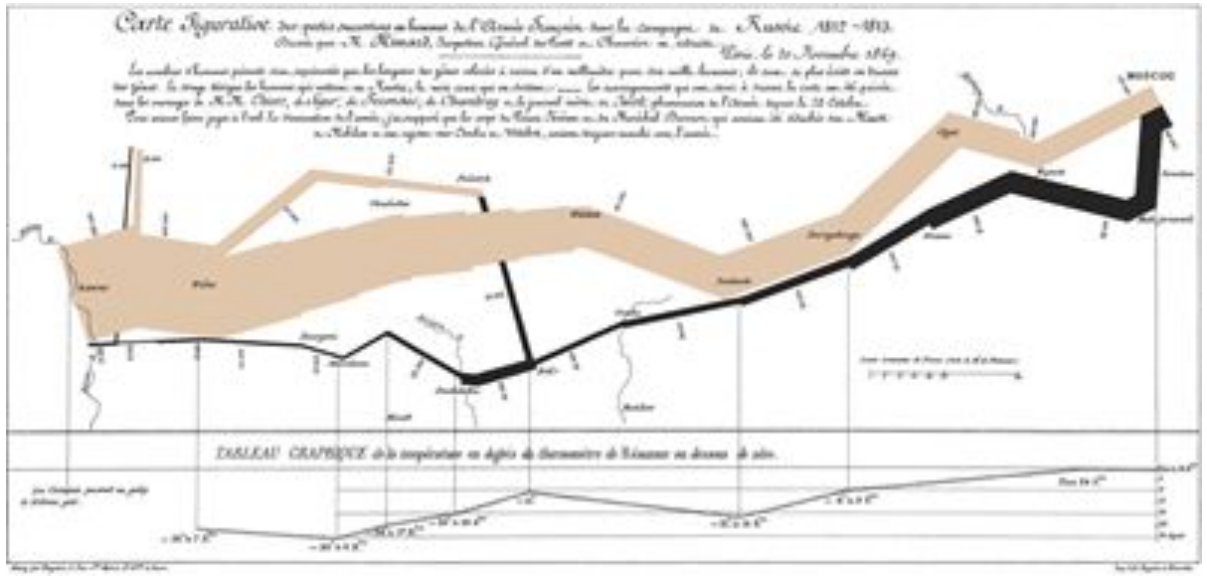
1855, John Snow, England

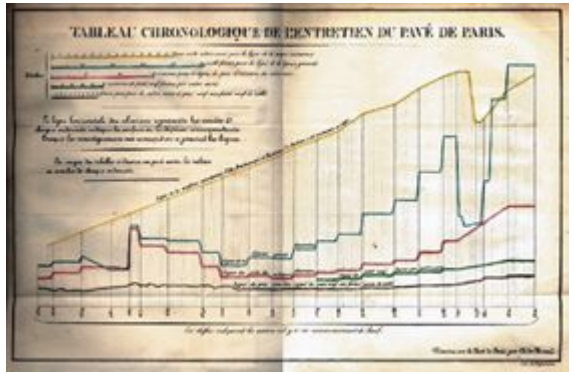
Broad Street cholera outbreak (1854)

Applied DataViz

Source: https://en.wikipedia.org/wiki/John_Snow#/media/File:Snow-cholera-map-1.jpg

Brief history of DataViz





1781-1870, Charles Minard, France

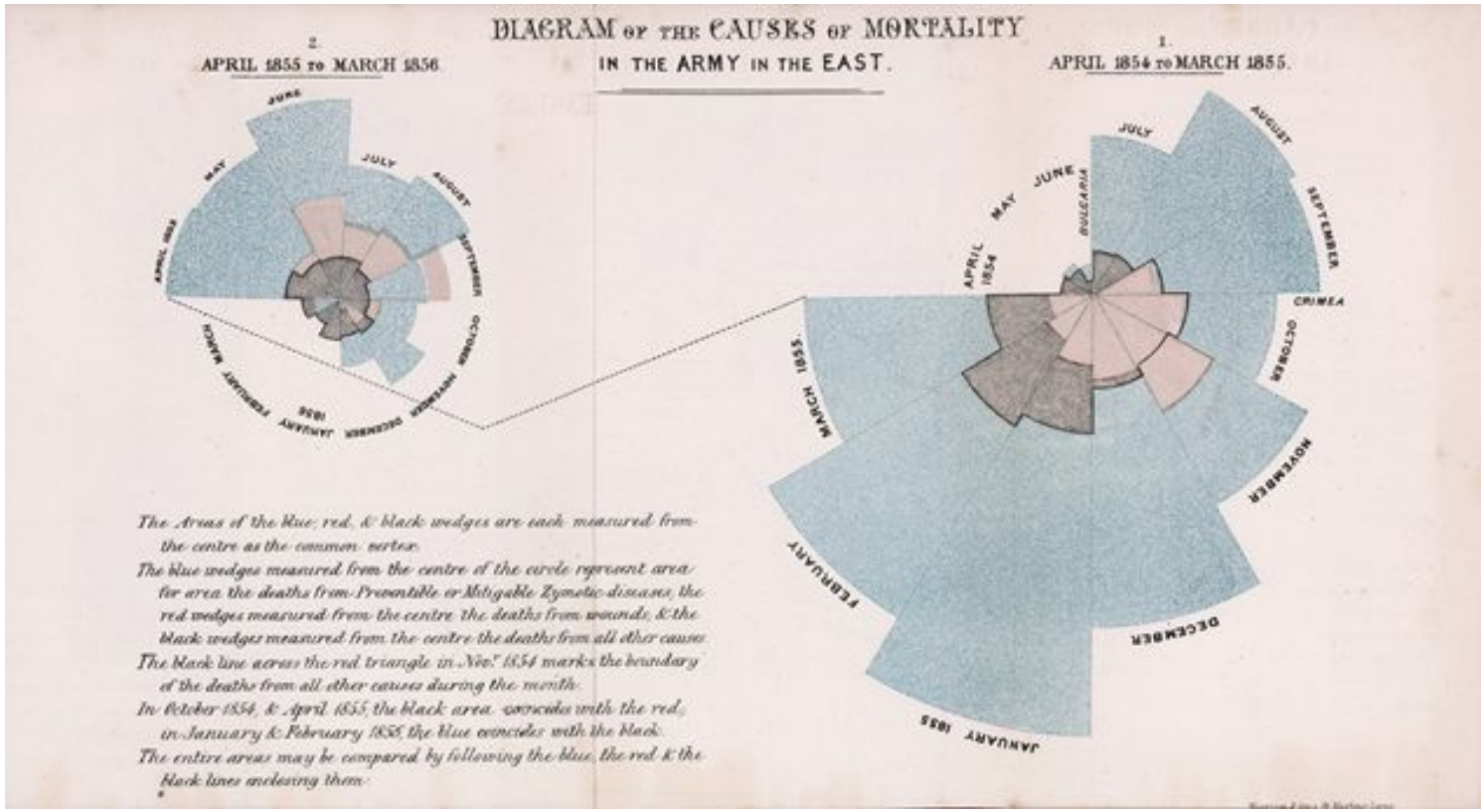
The origin of modern infographics

Source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

Brief history of DataViz

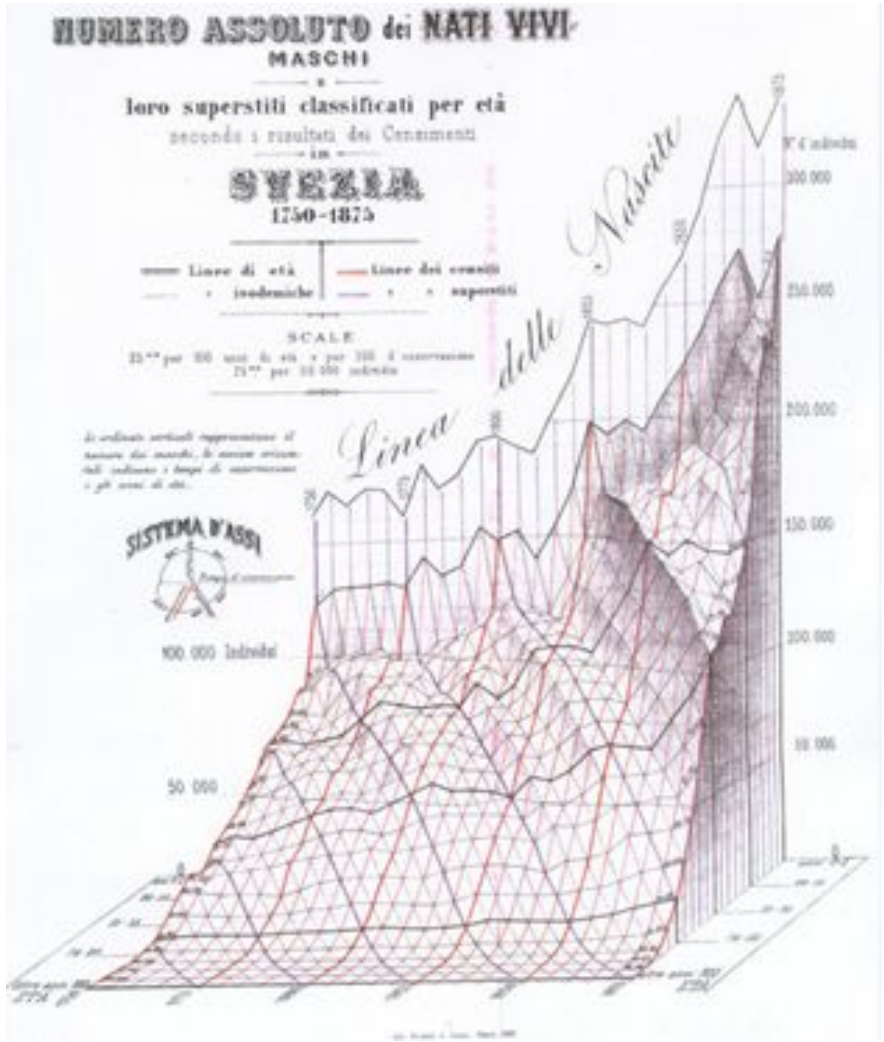
1820-1910, Florence Nightingale, England

From DataViz to science communication



Source: https://en.wikipedia.org/wiki/Florence_Nightingale#/media/File:Nightingale-mortality.jpg

Brief history of DataViz



1856-1916, Luigi Perozzo, Italy

Population pyramid

The first known 3D visualization

Source: https://it.wikipedia.org/wiki/Luigi_Perozzo#/media/File:Stereogram_(three-dimensional_population_pyramid)_modeled_on_actual_data_(Swedish_census,_1750-1875).jpg

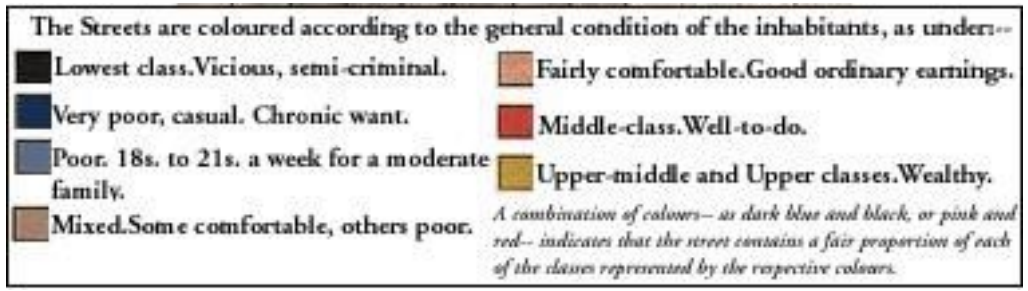
Brief history of DataViz

1840-1916, Charles Booth, England

London poverty maps

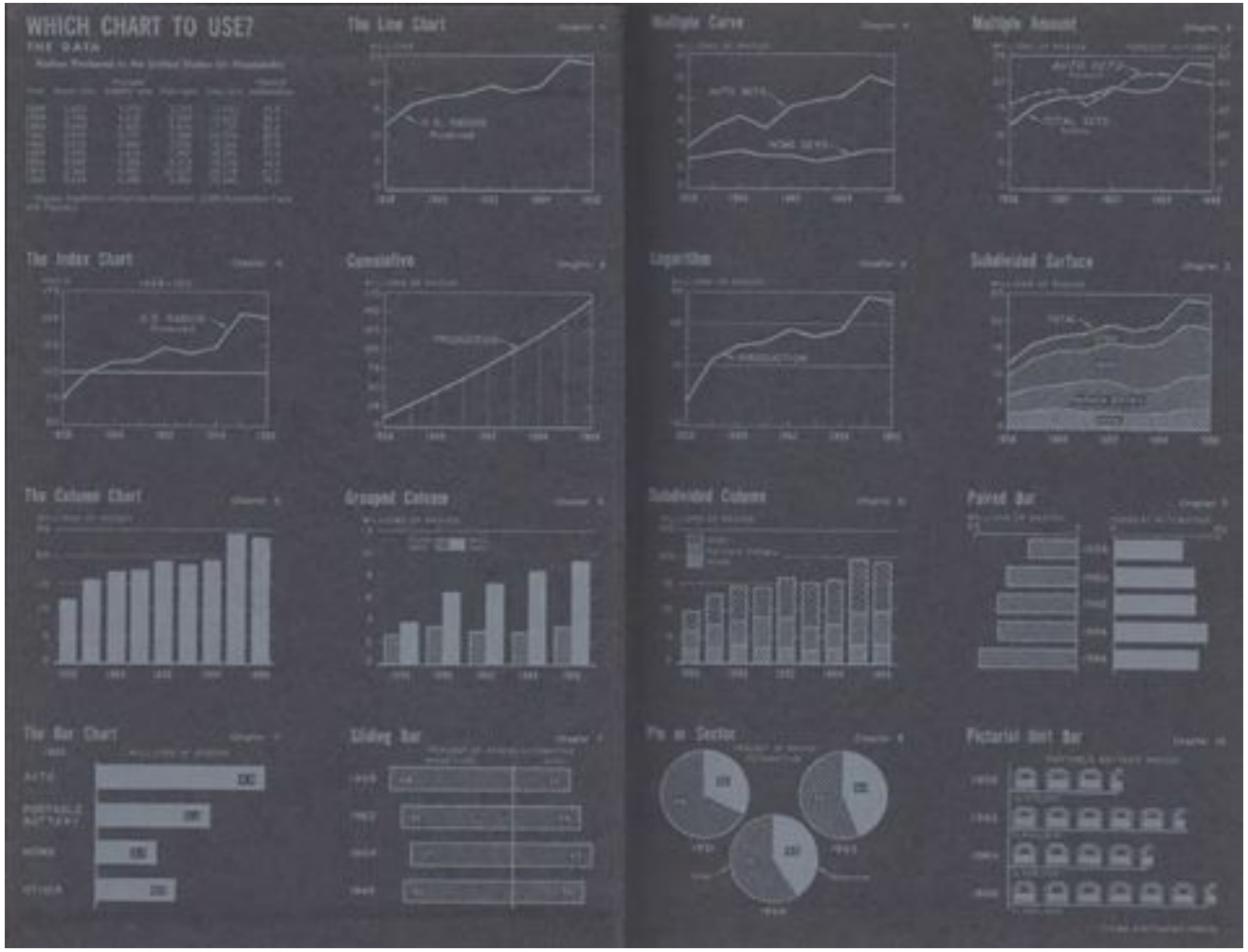
DataViz meets colours (incl. palette)

Source: https://en.wikipedia.org/wiki/Whitechapel#/media/File:Commercial_Road,_Whitechapel,_Charles_Booth,_Map_Life_London_nek56.jpg



Source: https://en.wikipedia.org/wiki/Charles_Booth_(social_reformer)#/media/File:Booth_poverty_map_colour_key.jpg

Brief history of DataViz



1897-1986, Mary Eleanor Spear, USA

1952 "*Charting Statistics*"

1969 "*Practical Charting Techniques*"

Practical guidelines, box-plot explained, cheating with DataViz, data presentation, …

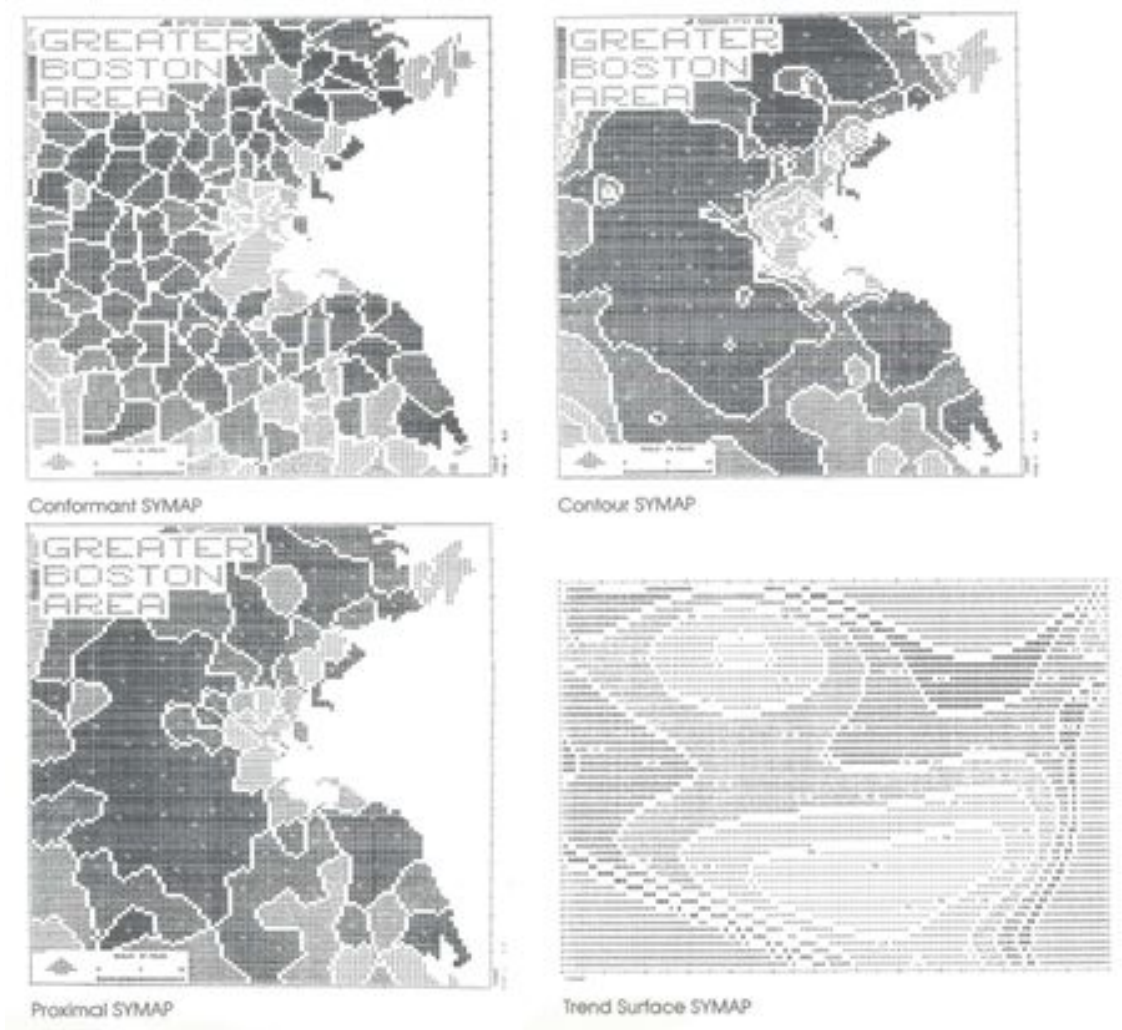
Source: https://medium.com/nightingale/credit-where-credit-is-due-mary-eleanor-spear-6a7a1951b8e6

Brief history of DataViz

Late 1960's, Harvard Laboratory for Computer Graphics and Spatial Analysis, USA

SYMAP (Synergistic Mapping)

Computer software for the analysis and graphic display of spatial data: first (low cost) printed output



Source: https://en.wikipedia.org/wiki/Harvard_Laboratory_for_Computer_Graphics_and_Spatial_Analysis#/media/File:SYMAP_-_LAB-LOG_1980.png

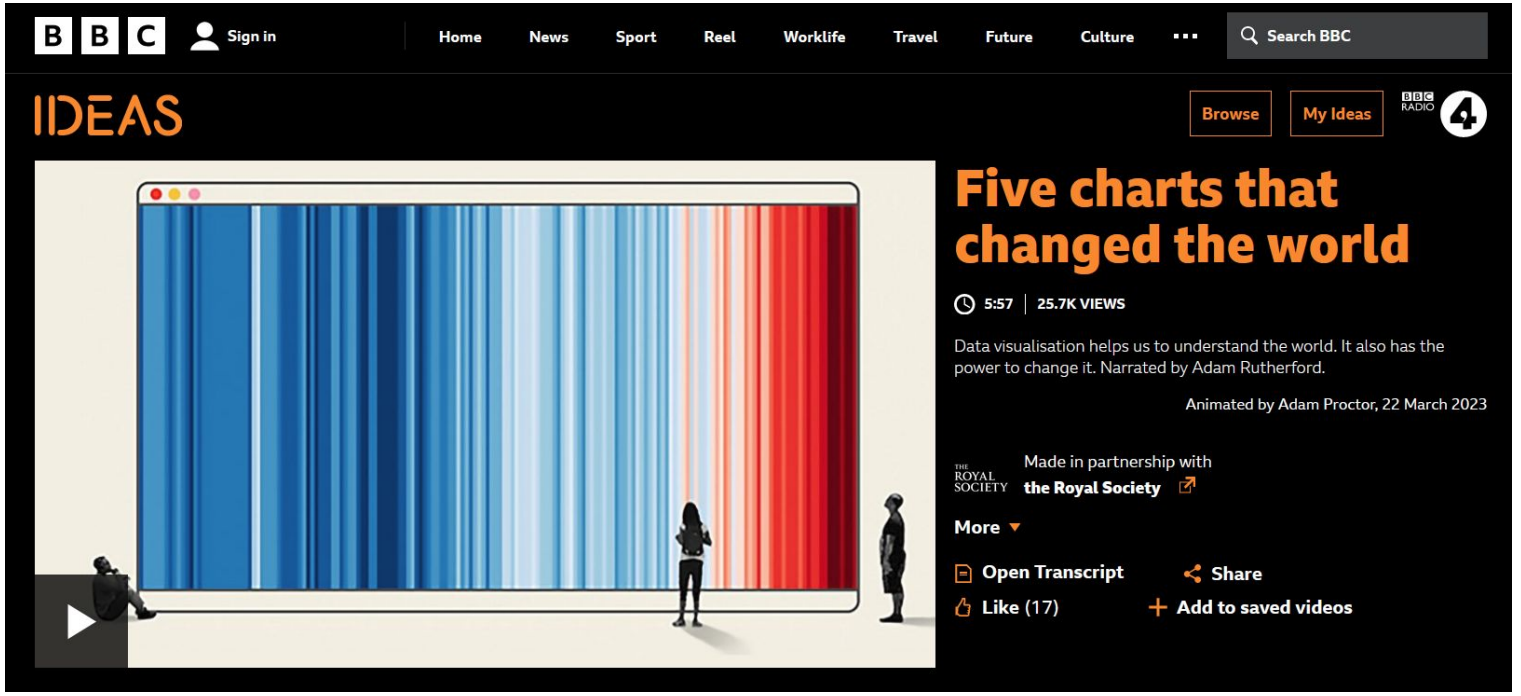
Brief history of DataViz

1985, Microsoft Excel 1.0

The game changer!

Source:
https://www.flickr.com/photos/microsoftsweden/5395285812/in/photostream/lightbox/

Brief history of DataViz



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

1. Fundaments of DataViz

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

A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE

A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE
A good DataViz is CORRECT, EFFECTIVE and ACCESSIBLE

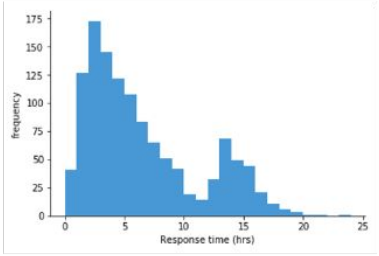
if possible, also visually appealing

How to DataViz

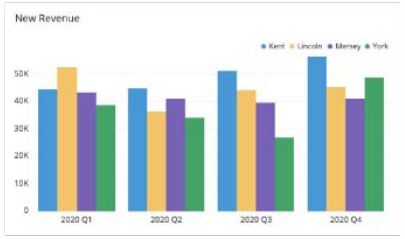
Before you DataViz, think:

- 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?*

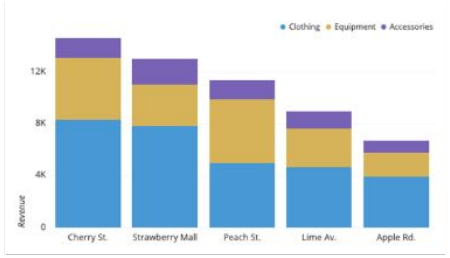
Choosing the right plot type



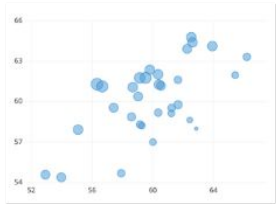
histogram



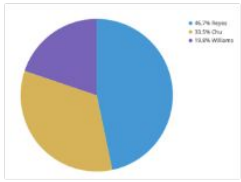
grouped bar chart



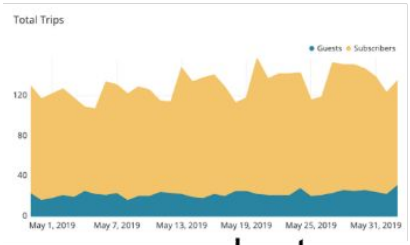
stacked bar chart



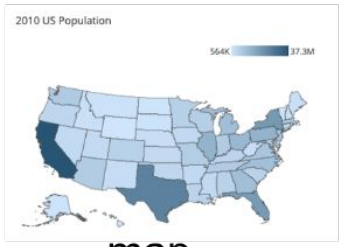
bubbles chart



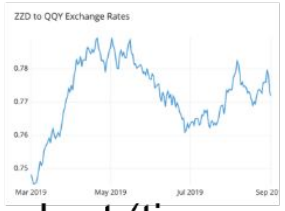
pie chart



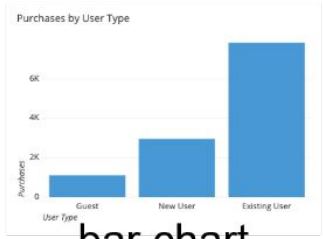
area chart



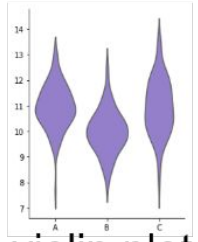
map



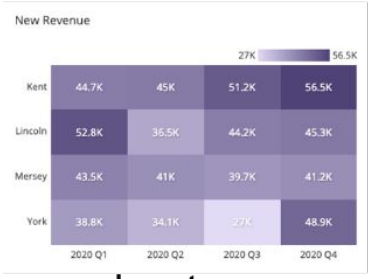
line chart (time series)



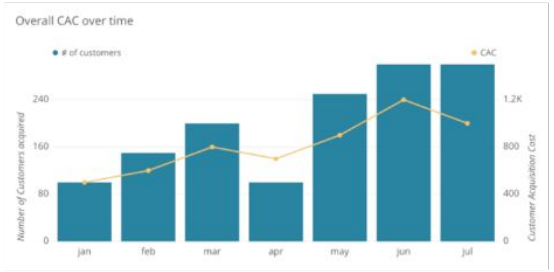
bar chart



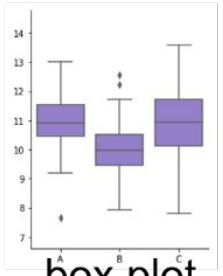
violin plot



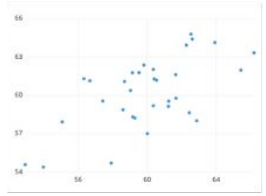
heatmap



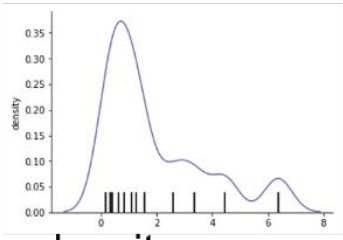
combined (dual axes)



box plot



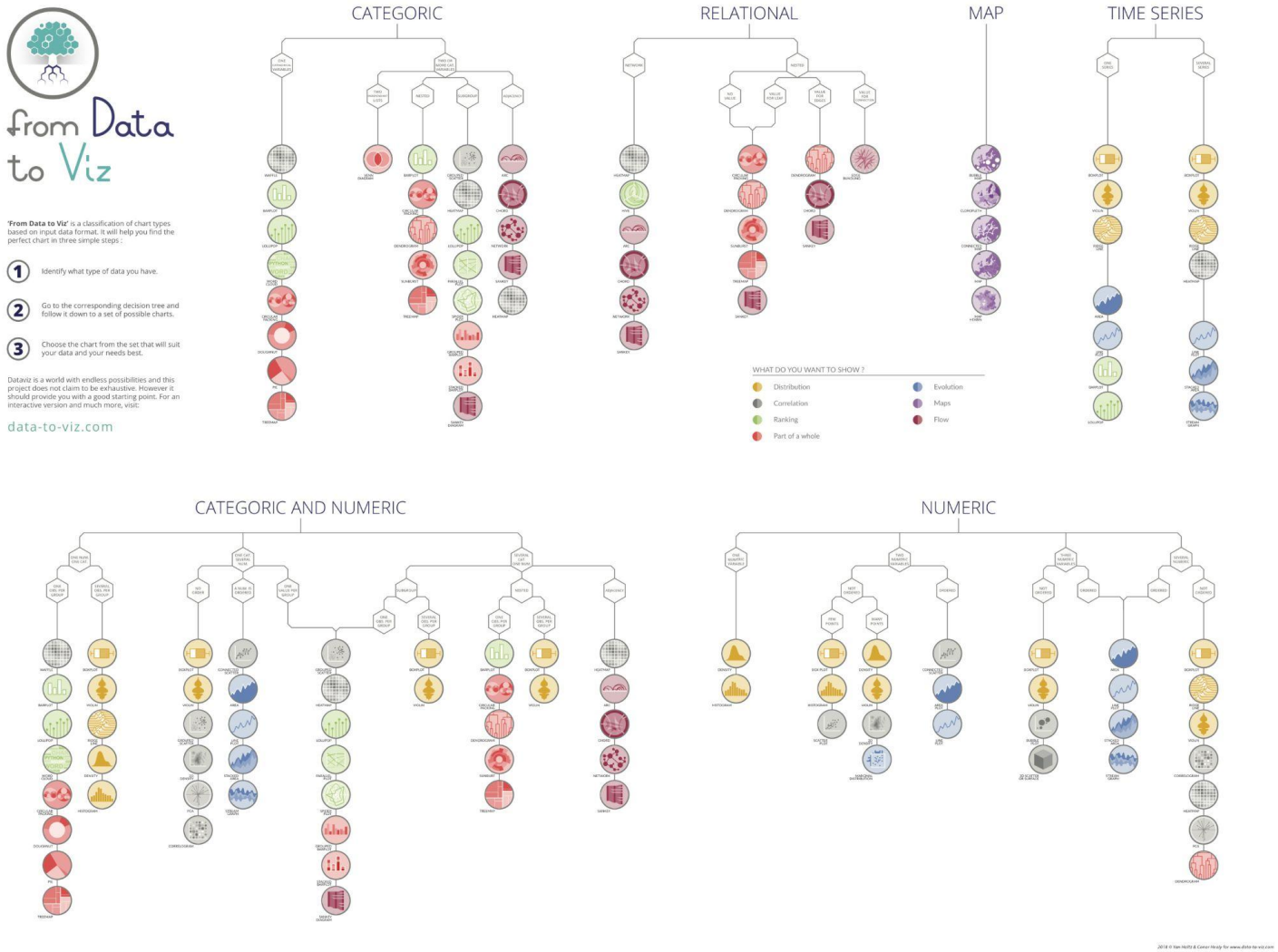
scatter plot



density curve

Source: chartio.com

Choosing the right plot type



Useful resources:

data-to-viz.com

datavizcatalogue.com

datavizproject.com

Source: datavizuniverse.substack.com

Graphical elements of a plot

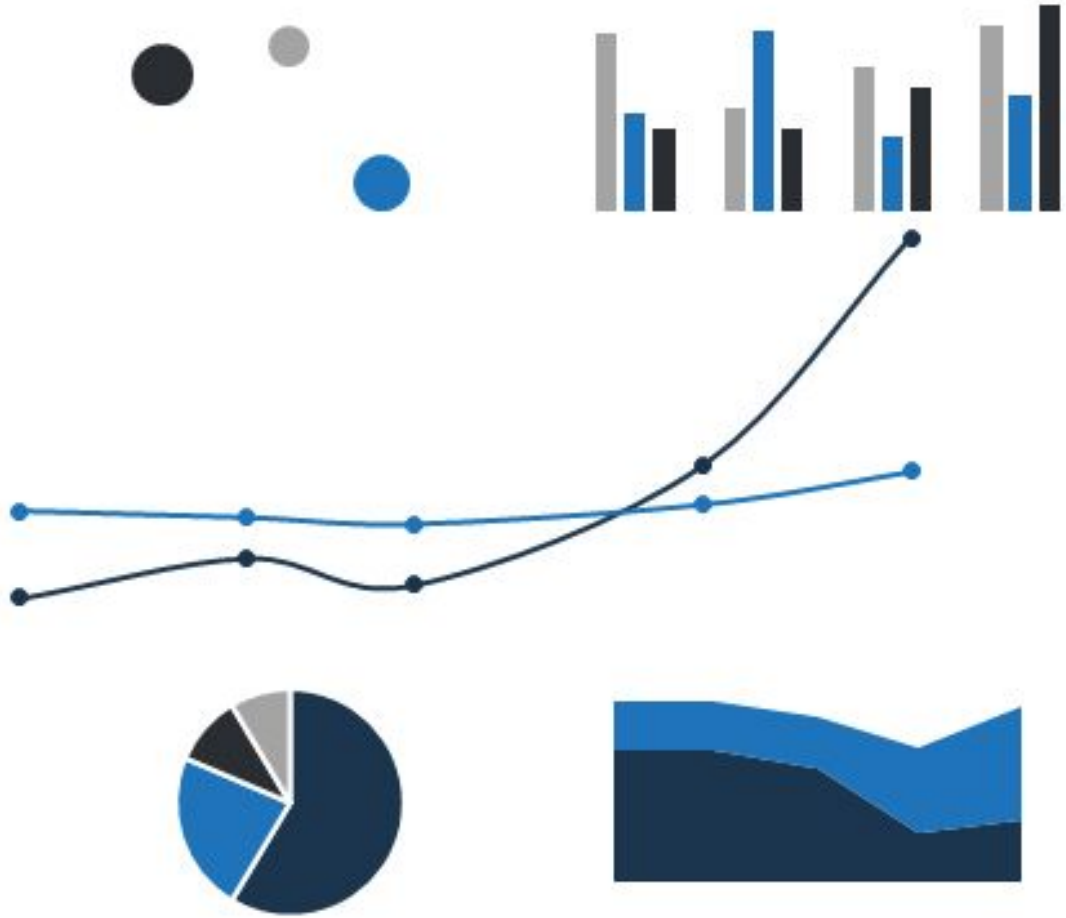
Marks and their attributes are the building blocks of all data visualizations.

Marks:

- Points
- Lines
- Areas
- Volumes

Attributes:

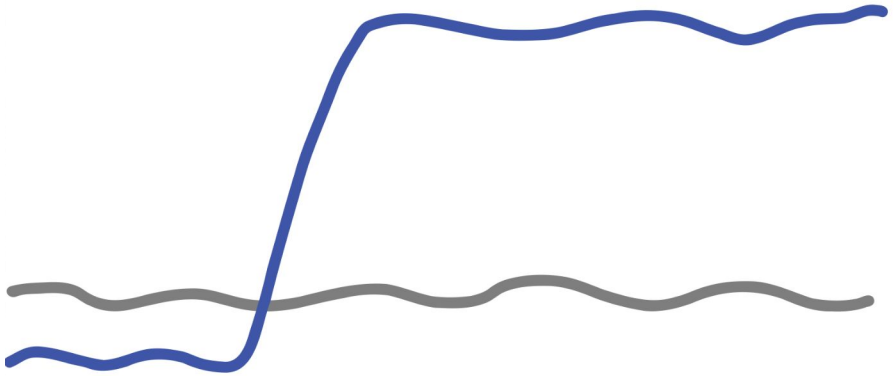
- Size
- Shape
- Orientation
- Colour



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
- (Legend)
- (Data labels)
- (Title)
- Caption
- Blank spaces
- Layout

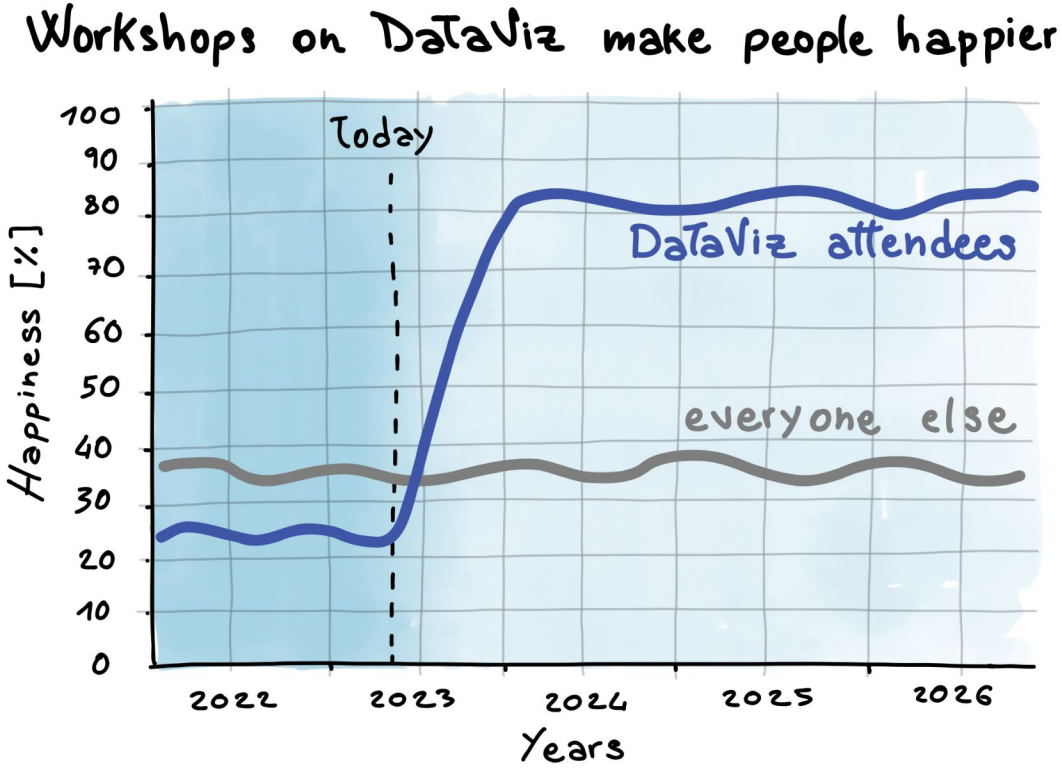


Fig. 1: Time series of the happiness factor for the two samples analysed. The data will be made available upon request.

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

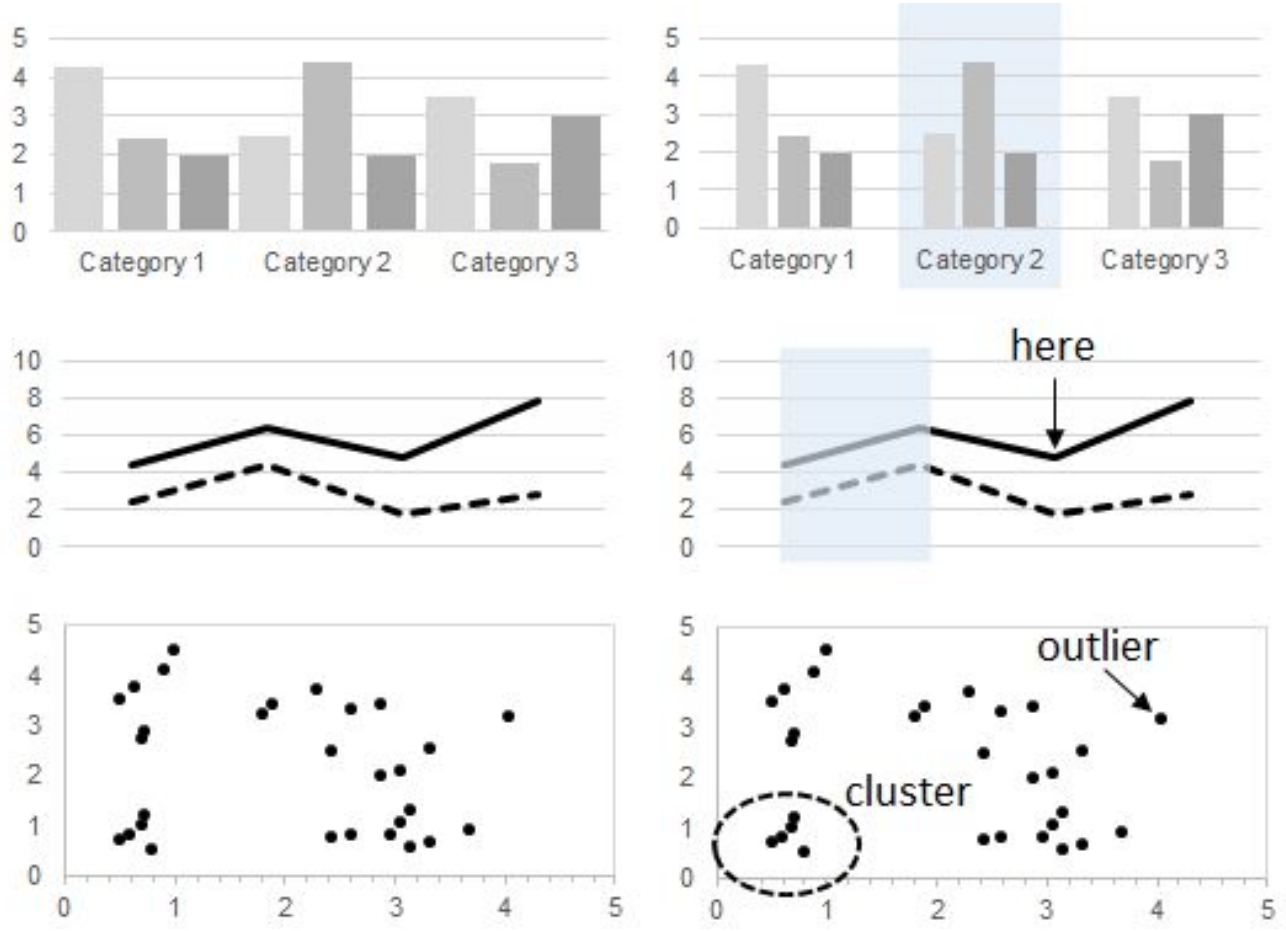
Choose the right font type

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.

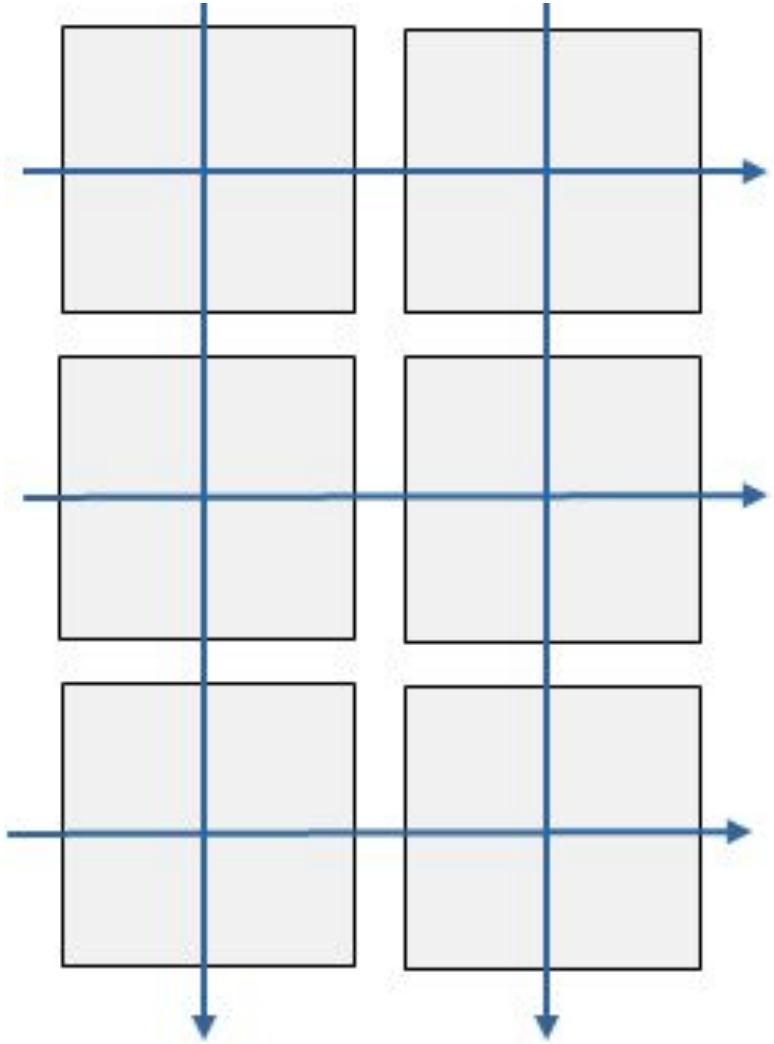
Use layout and annotation to highlight and guide the reader.



Layout (multiple graphs)

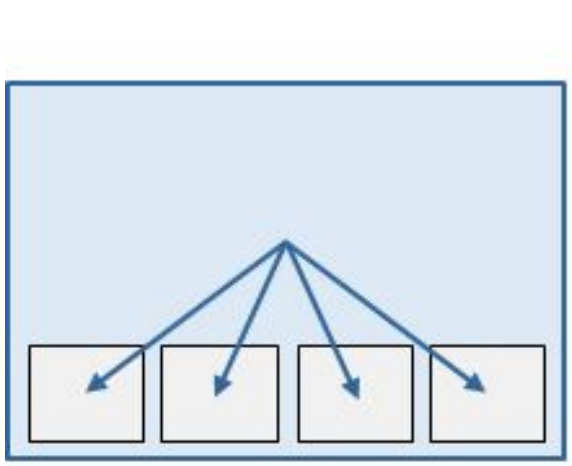
Improper arrangement of graph elements can confuse and/or mislead the readers.

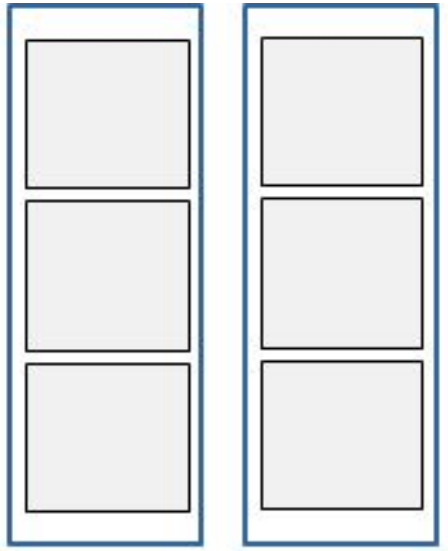
- Make it intuitive to the reader



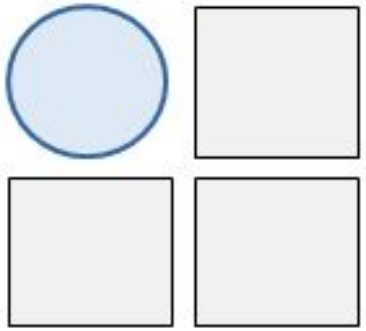
Layout (multiple graphs)

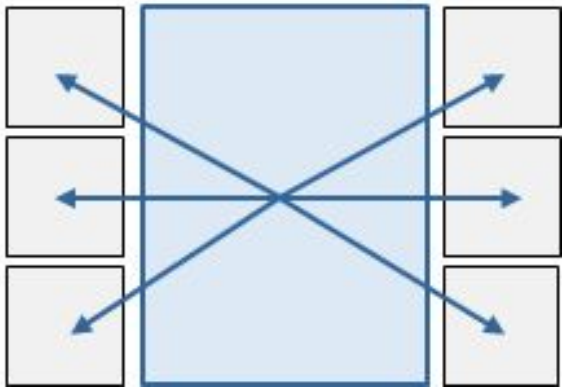
Improper arrangement of graph elements can confuse and/or mislead the readers.





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

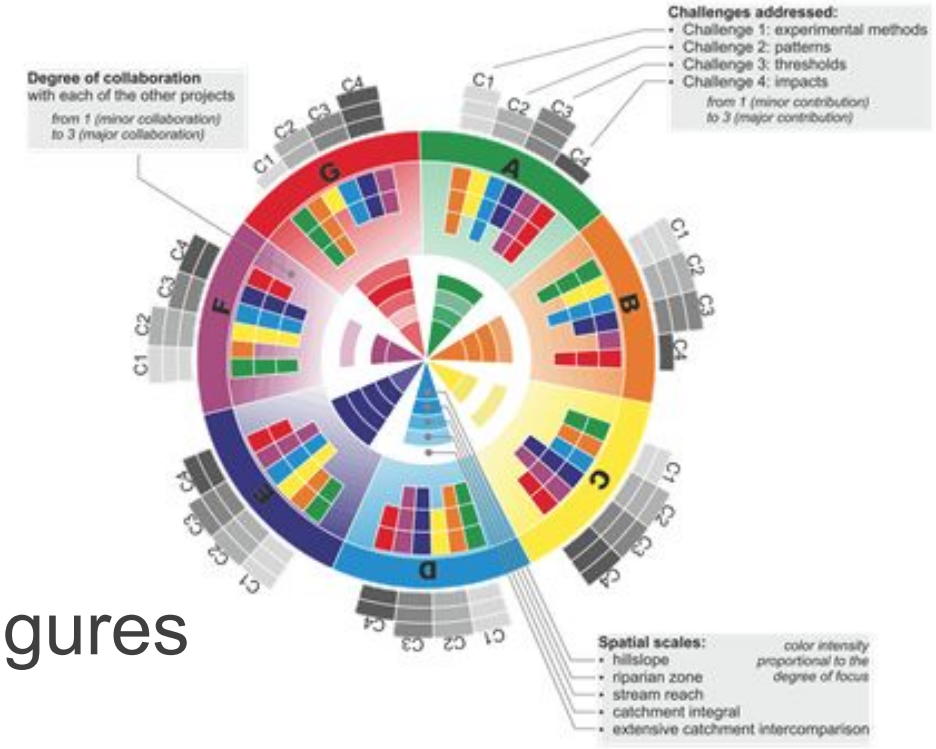




Simplicity vs creativity

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 colours wisely!

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

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
- dorotheepost.de
- fabiocrameri.ch

2. Different approaches for visualization - publication figures vs posters vs graphical abstracts

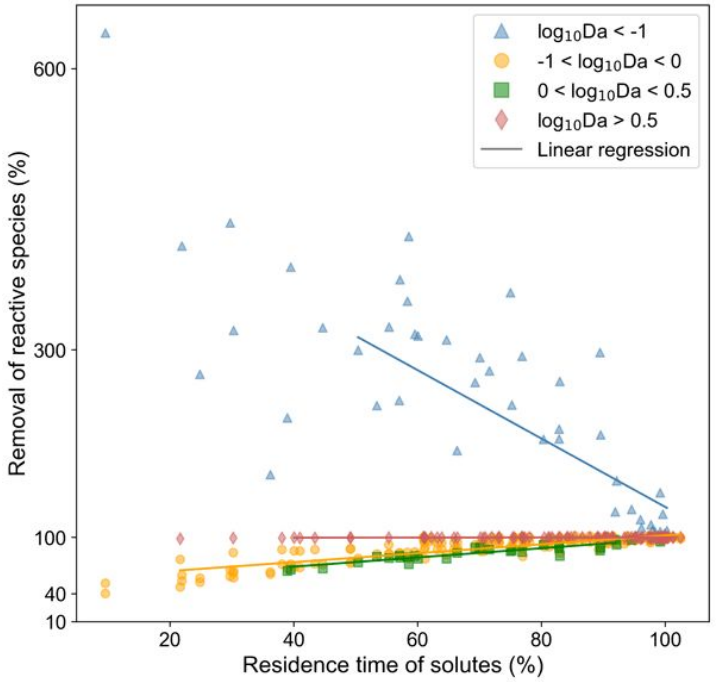
(Swamini/?)

Publications, posters, presentations

- Publications and posters are static and should be standalone - enough information in caption and text
- Posters:
  - 10 guidelines for an awesome poster
  - Presents a snapshot for casual observers to encourage engagement
  - Enough supporting material for an involved conversation with an expert

- Presentations: Dynamic
  - Modify according to audience
  - Leverage *simple* animations to prevent overwhelming the audience

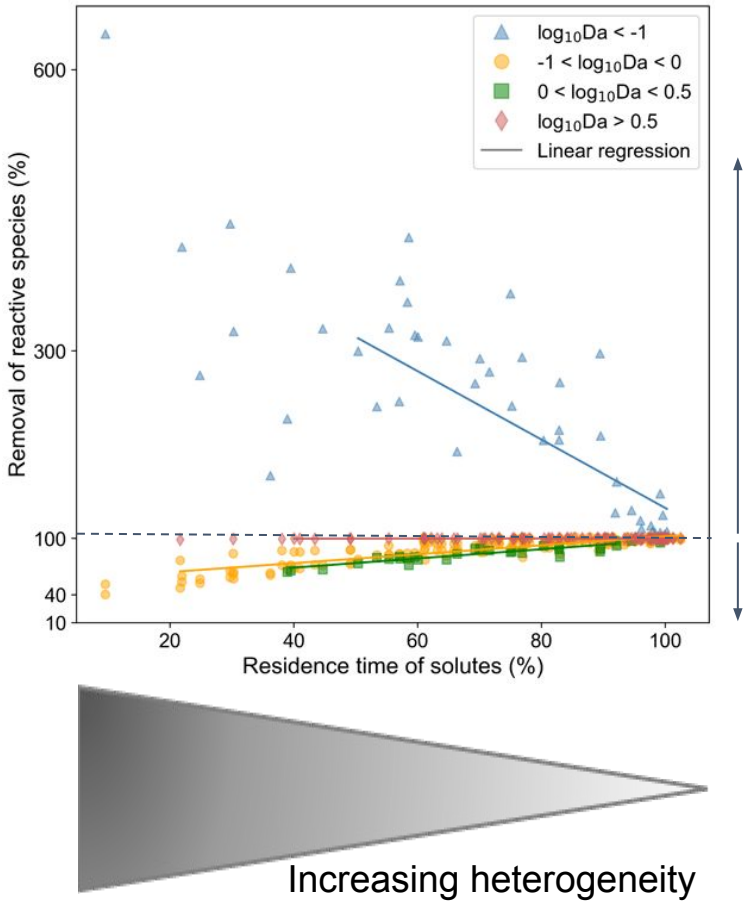
Example of a publication figure



Khurana et al. (2022)

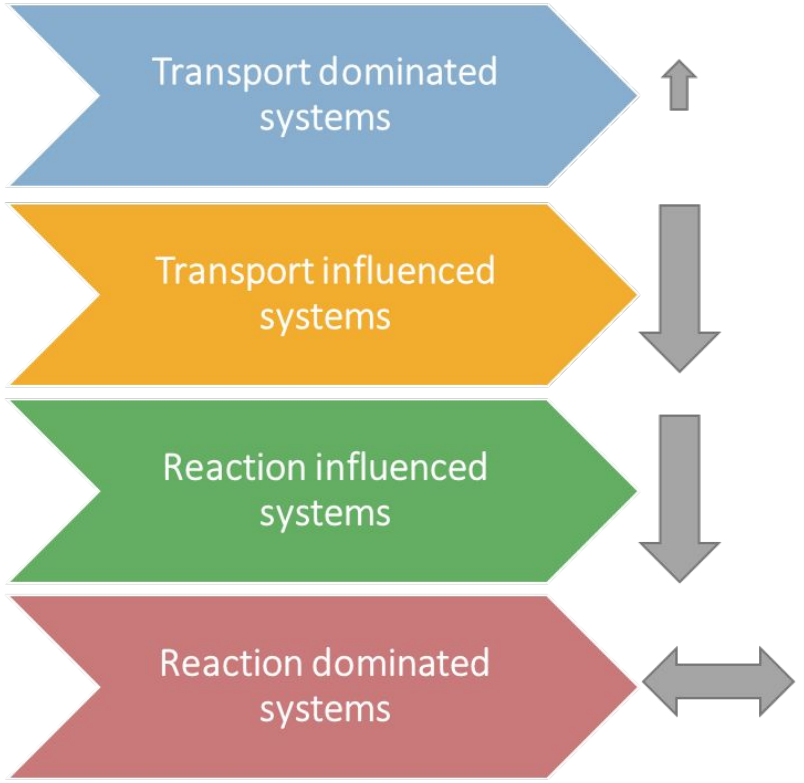
Figure 6. Regression analysis: predicting impact of spatial heterogeneity on chemical species removal in the different reaction regimes indicated by log10Da. Value on the y axis indicate the removal of chemical species in heterogeneous domains normalized by that in the corresponding base case. Spatial heterogeneity is plotted on the x axis, indicated by the breakthrough time in the heterogeneous domain normalized by that in the base case (homogeneous domain). A value of 100 % on the y axis indicates that the removal of the chemical species is the same as that in the corresponding base case (homogeneous domain). A value of 50 % indicates that the removal of the chemical species was reduced by half in the corresponding heterogeneous domain. A value of 600 indicates that the removal of the chemical species in the heterogeneous domain was 6 times that in the homogeneous domain.

Example of a modified publication figure for presentation



Removal of chemicals increases

Removal of chemicals decreases



Example of a publication figure

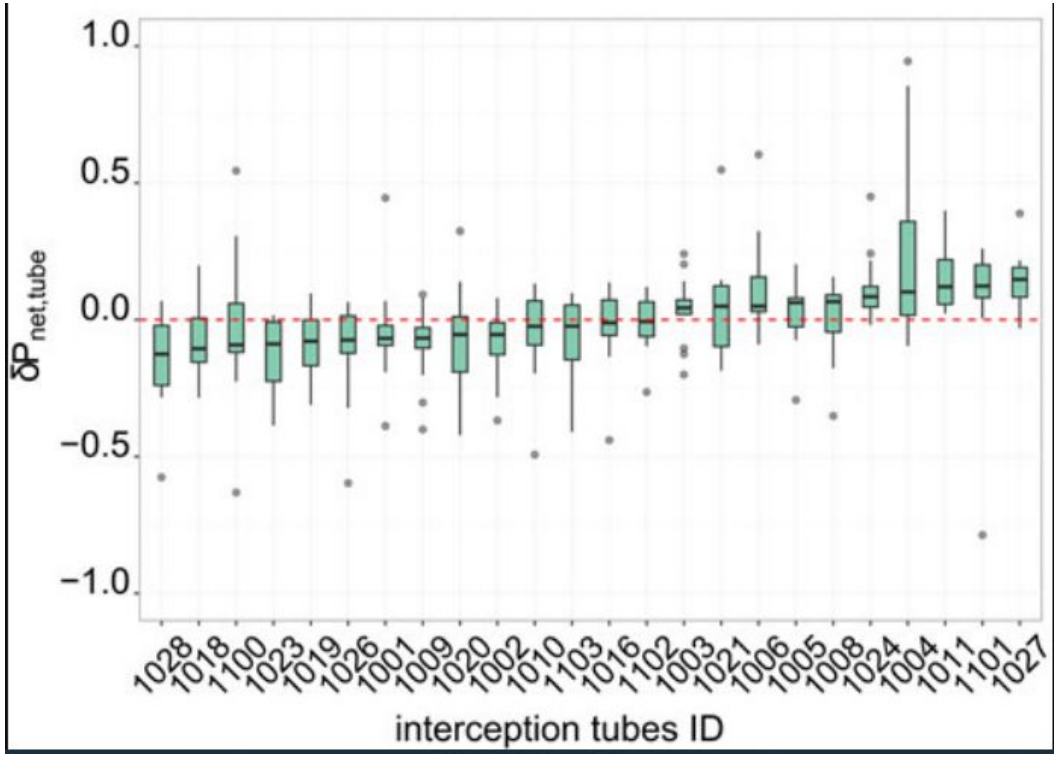
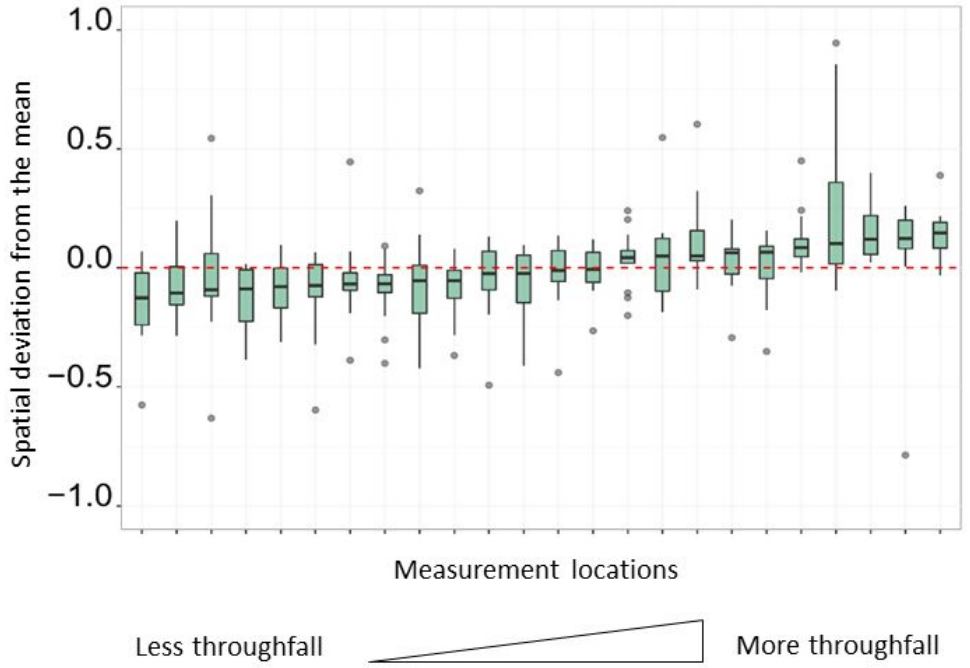


FIGURE 8. Temporal stability of precipitation patterns based on the spatial deviation of the interception tube measurements in 2019 (April–August) for (top) uncovered period and (bottom) covered period. *X* axis indicates location identification, and *Y* axis gives the spatial deviation from the mean of the tubes' measurements for the particular location for all sampling weeks.

Demir et al. (2022)

Example of a modified publication figure for presentation



Grassland canopy introduces spatial variability in throughfall patterns which are temporally stable

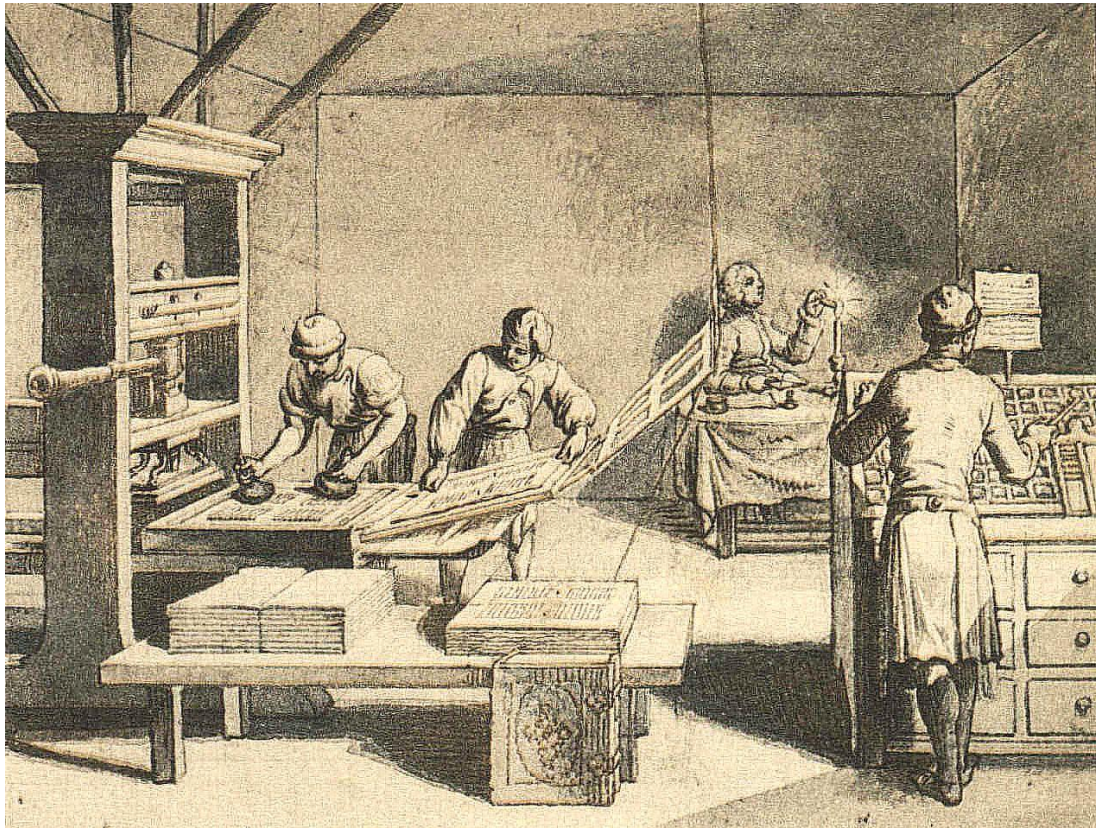
Demir et al., 2022a

3. Publication compliance

Swamini

Always check the publisher guidelines

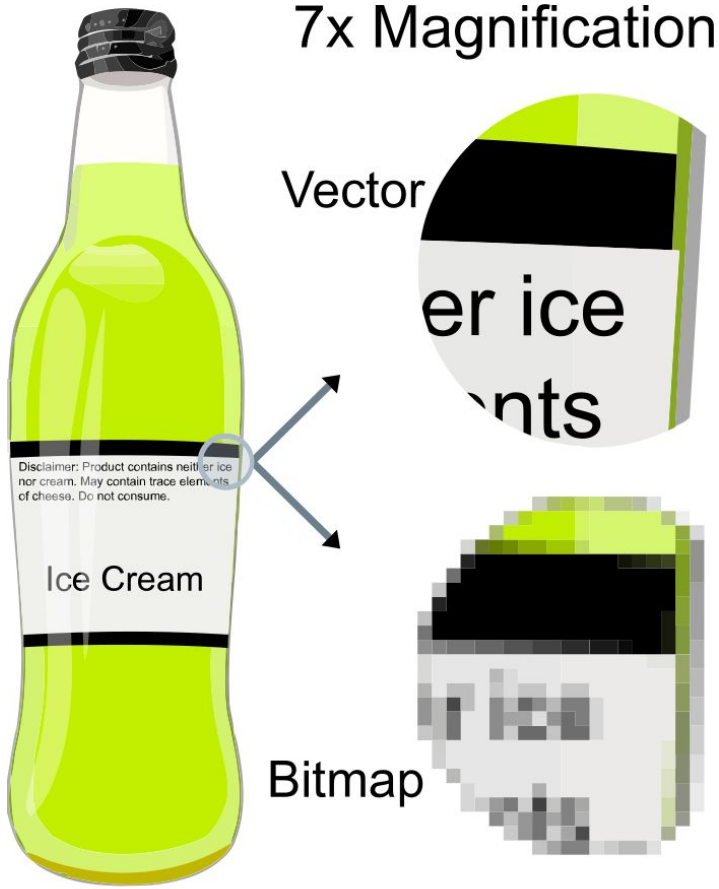
- Formats
- Image sizes
- Resolution



Formats: Raster vs vector graphics

- **Raster:**
  ***.jpg, *.png, or *.tif**
  - Better for photographs and subtle color variation
  - *.png recommended over *.jpg except for photos

- **Vector**
  ***.pdf, *.ps, *.eps**
  - Better for everything else



Resolution: 300 dpi

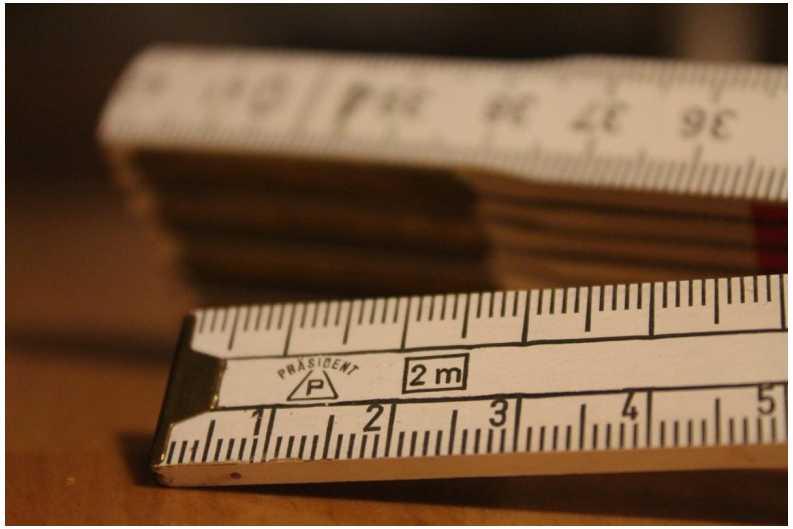
- dpi = dots-per-inch
- Resolution depends on size of final figure, but only for raster images.

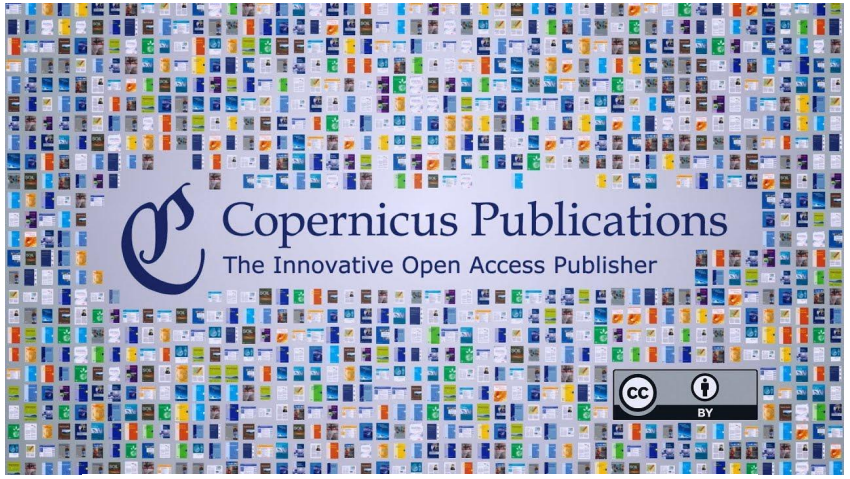
563 pixel / 4.5 inch = 125 dpi

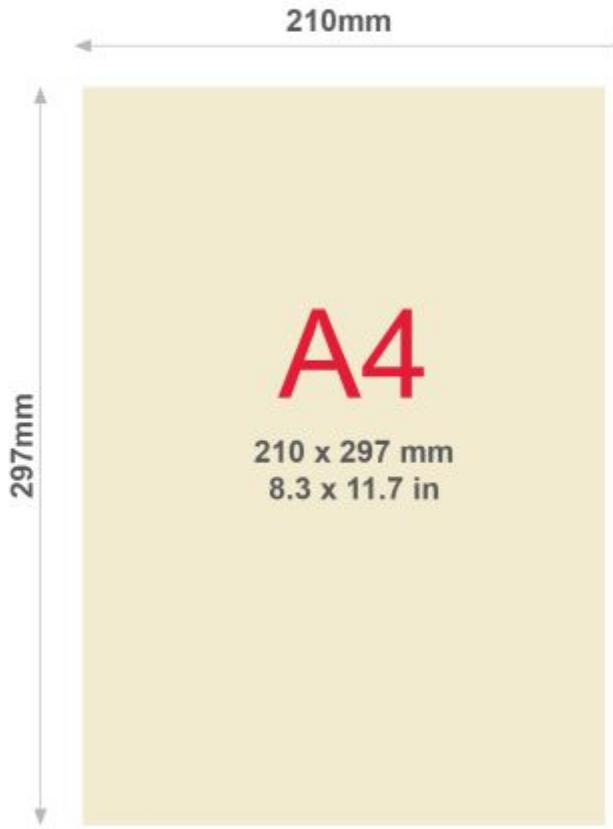
96 pixel / 4.5 inch = 21 dpi

Size and ratio:

- Copernicus recommends to assume a width of at least 8 cm
- Typesetter decides figure placement and size, but can take author requests into account







Source: Paper Sizes UK Guide | A2, A3, A4, A5 Paper Size | Viking UK (viking-direct.co.uk)

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)

The licensor cannot revoke these freedoms as long as you follow the license terms.

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!

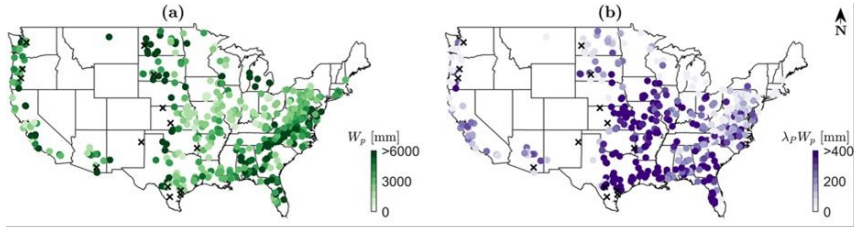
4. Colour schemes

Paola

Colour in scientific visualization

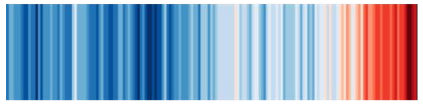
Choose the right color scale for the right reason!

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



Gnann et al, 2019, WRR

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

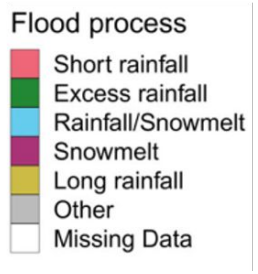


Ed Hawkins, Climate Stripes

Cyclic

Blöschl et al, 2017, Science

Categorical



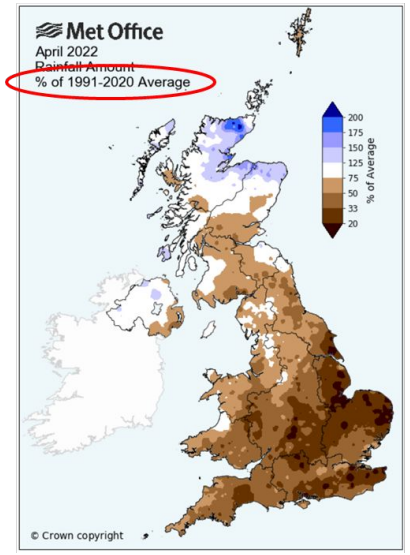
Stein et al, 2020, HP

Colour in scientific visualization

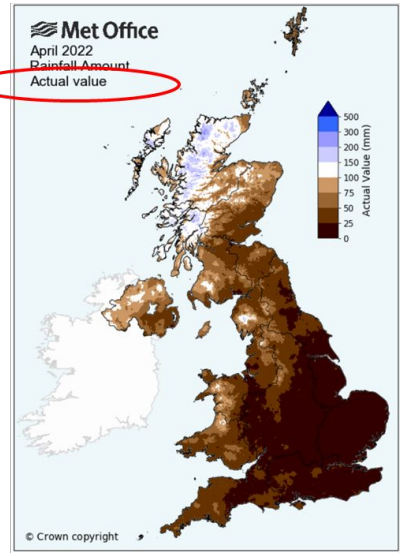
A **diverging color scale** indicates a relevant change-point.

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

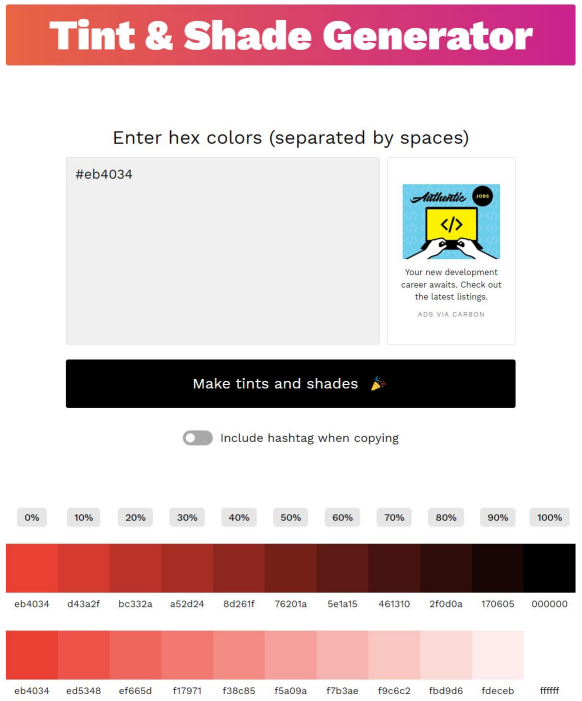


Non-diverging values indicate a mid-point that is not there



Colour in scientific visualization

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



Colour in scientific visualization

Visualizing **categorical data**.

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.

Your colors should have different brightnesses.

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

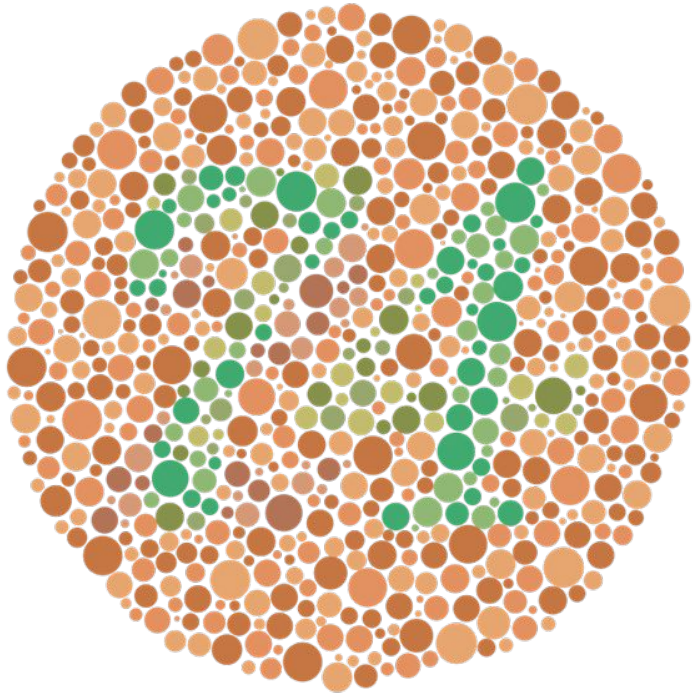
Source: https://blog.datawrapper.de/colors-for-data-vis-style-guides/

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



Check your figures:

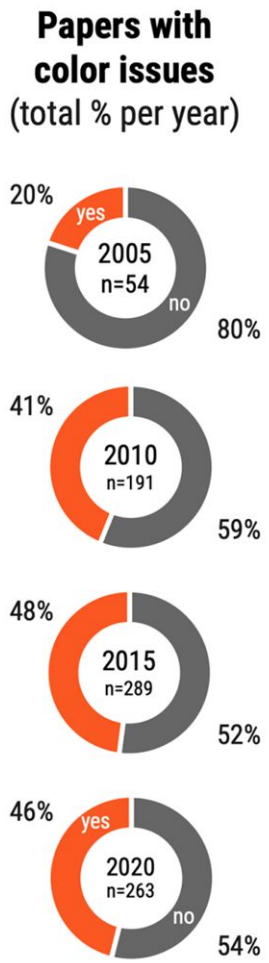
- https://www.color-blindness.com/
- R package colorblind
- https://contrastchecker.com/

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

Colour in scientific visualization

Survey conducted on papers published in HESS in 2015, 2010 and 2005.

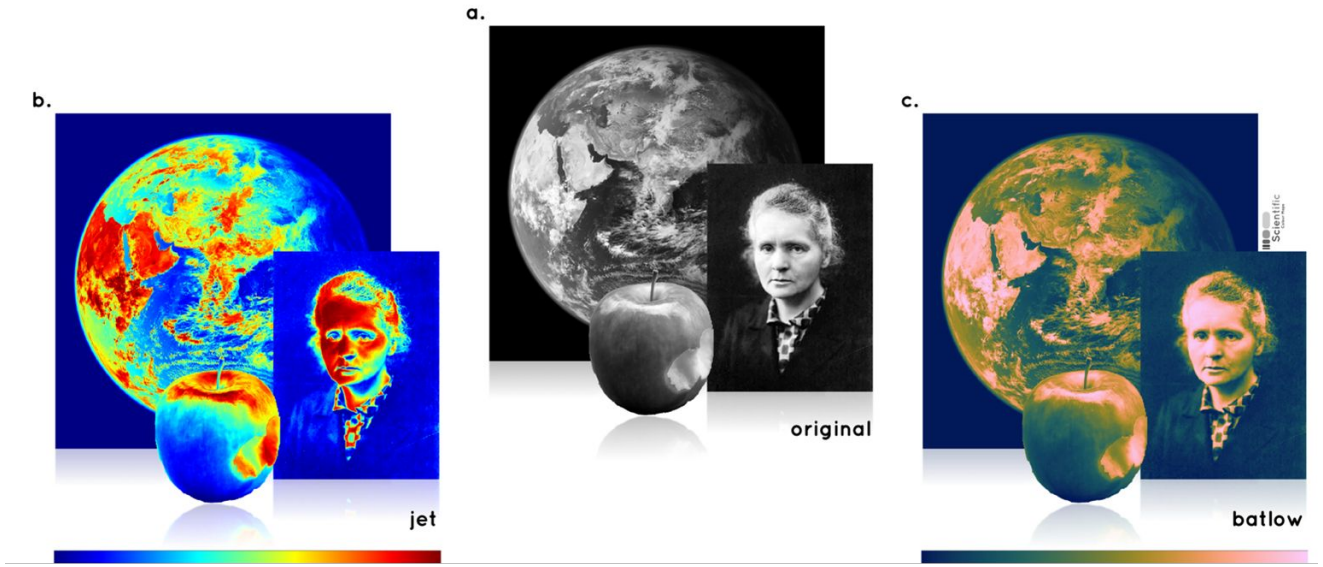
Scope: better understand if there is or was a tendency towards more or fewer rainbow-colored visualizations in scientific publications.



Source: Stoelzle and Stein, 2021, HESS

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

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]

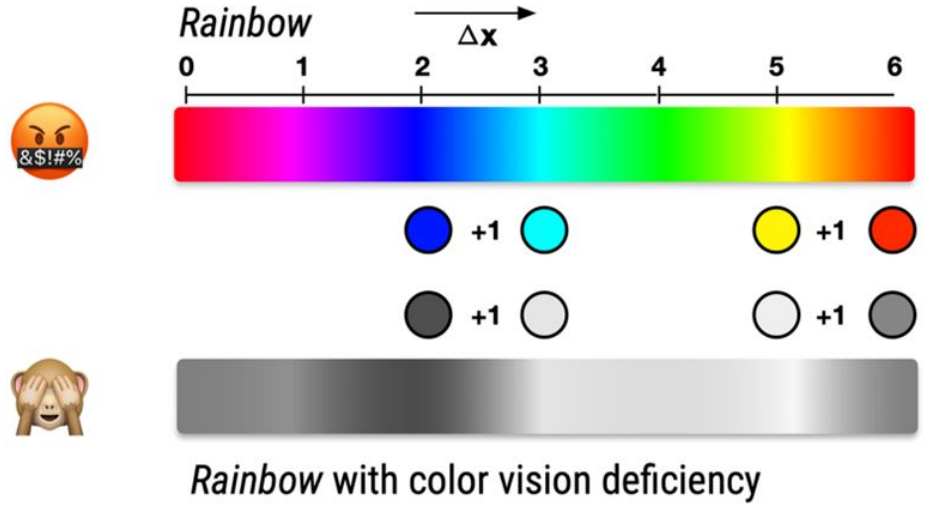
[1]Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany
[2]Department of Civil Engineering, University of Bristol, Bristol, UK

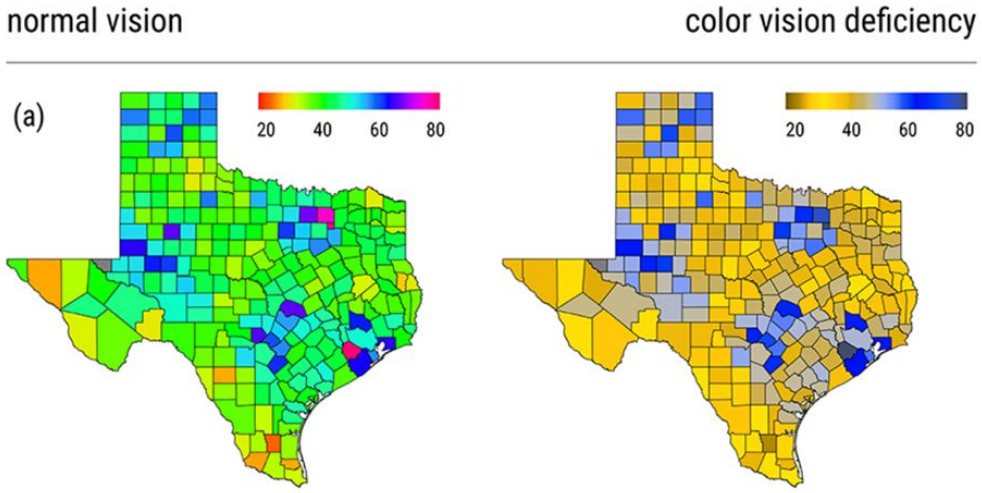
Correspondence: Michael Stoelzle (michael.stoelzle@hydro.uni-freiburg.de)

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)!

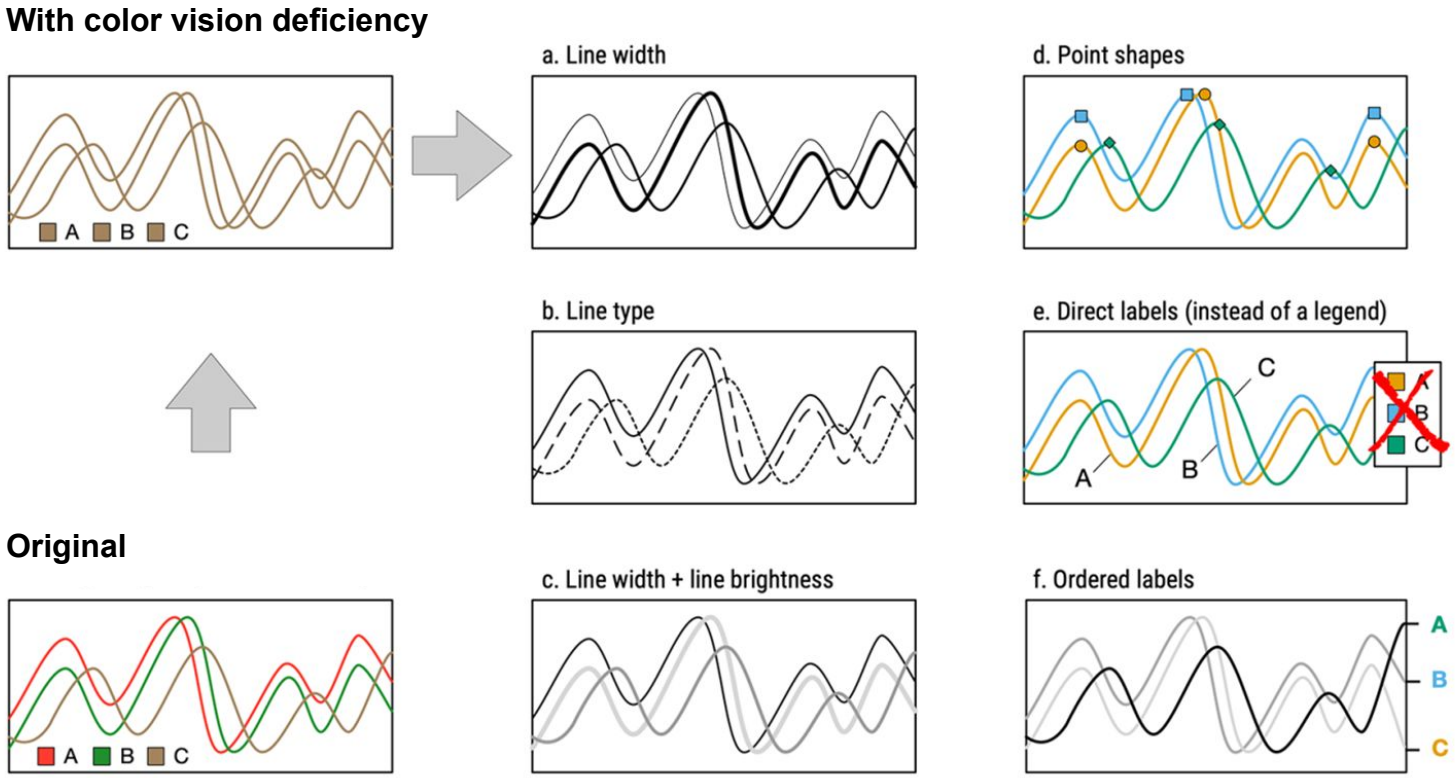




Source: Stoelzle and Stein, 2021, HESS

Colour in scientific visualization

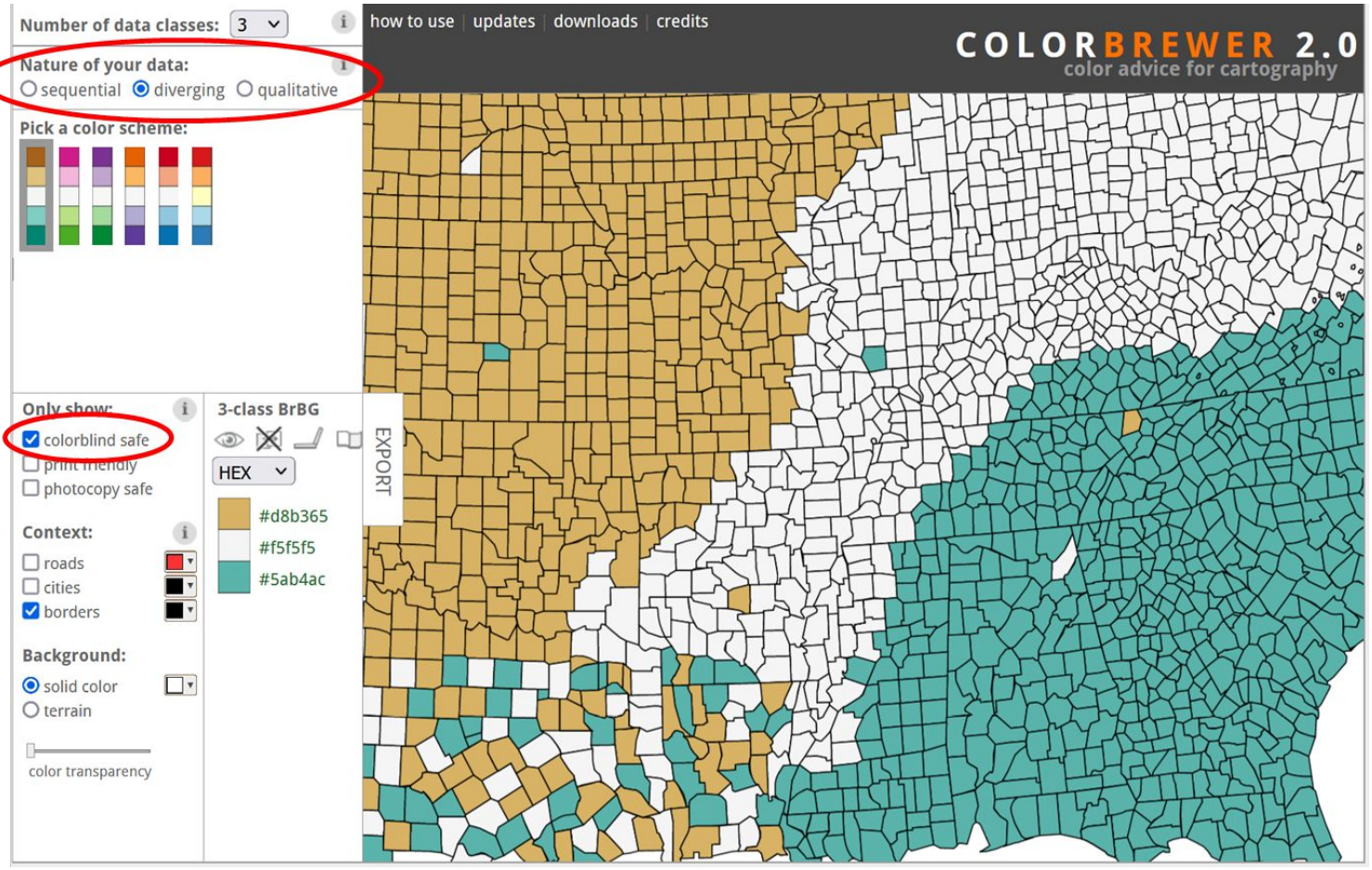
What are the alternatives? **Avoid colors and improve labels**



Source: Stoelzle and Stein, 2021, HESS

Colour in scientific visualization

Online resources: Colorbrewer (colorbrewer2.org)



RColorBrewer

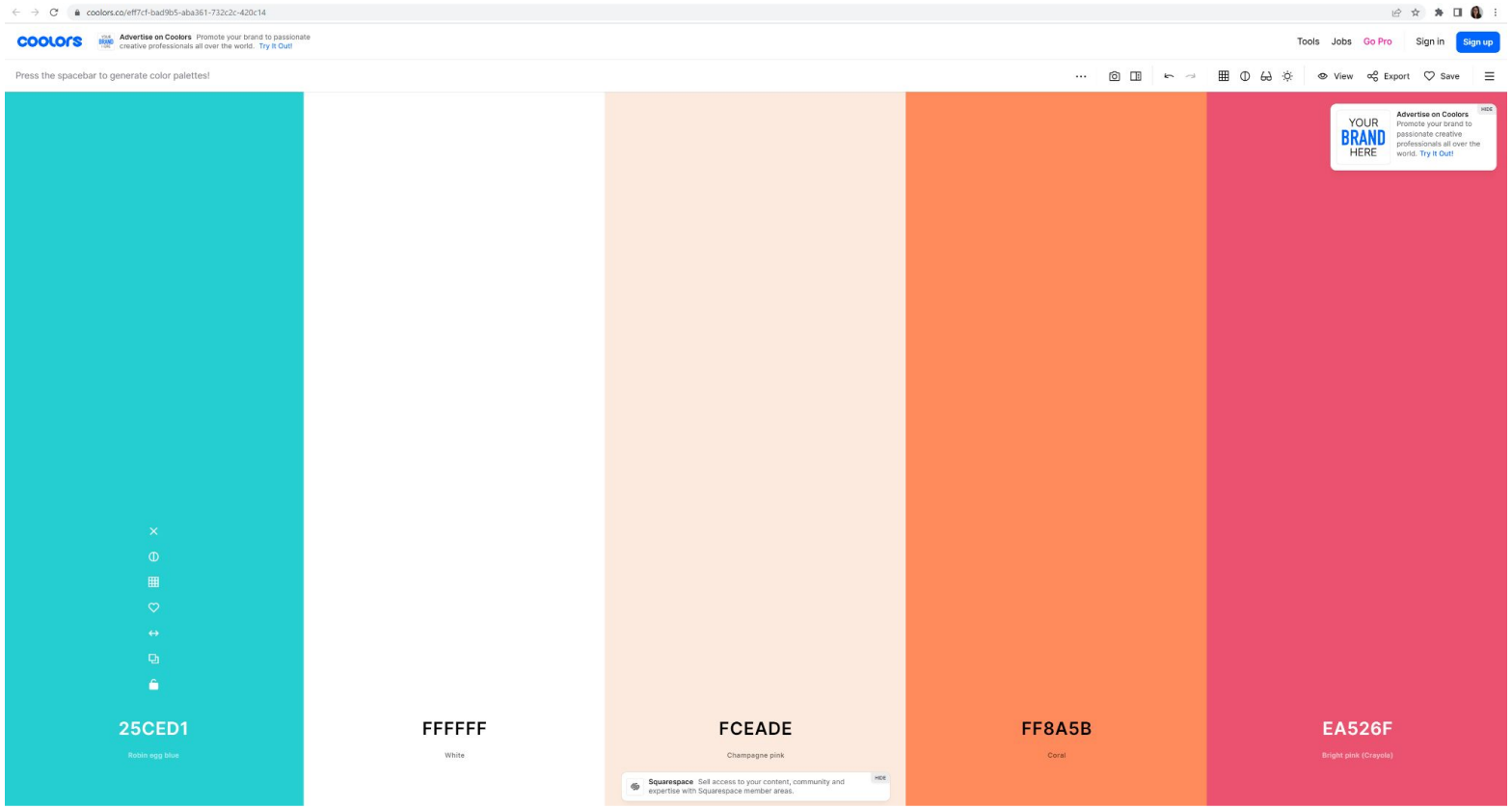
MATLAB®

BREWERMAP Function

palettable

Colour in scientific visualization

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



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/

Colour in scientific visualization

To summarize:

- take some time to read this very interesting guidelines https://blog.datawrapper.de/colors-for-data-vis-style-guides/ (it is **full of info** and **additional resources**);
- choose the **right** kind of **color scale**;
- do **not** use rainbow colors or other non-accessible color combinations;
- improve on your labels, avoid color completely or pick the correct color scales;
- point out color issues as a reviewer;
- give friendly suggestions to colleagues.

5. Panoply

Roshanak

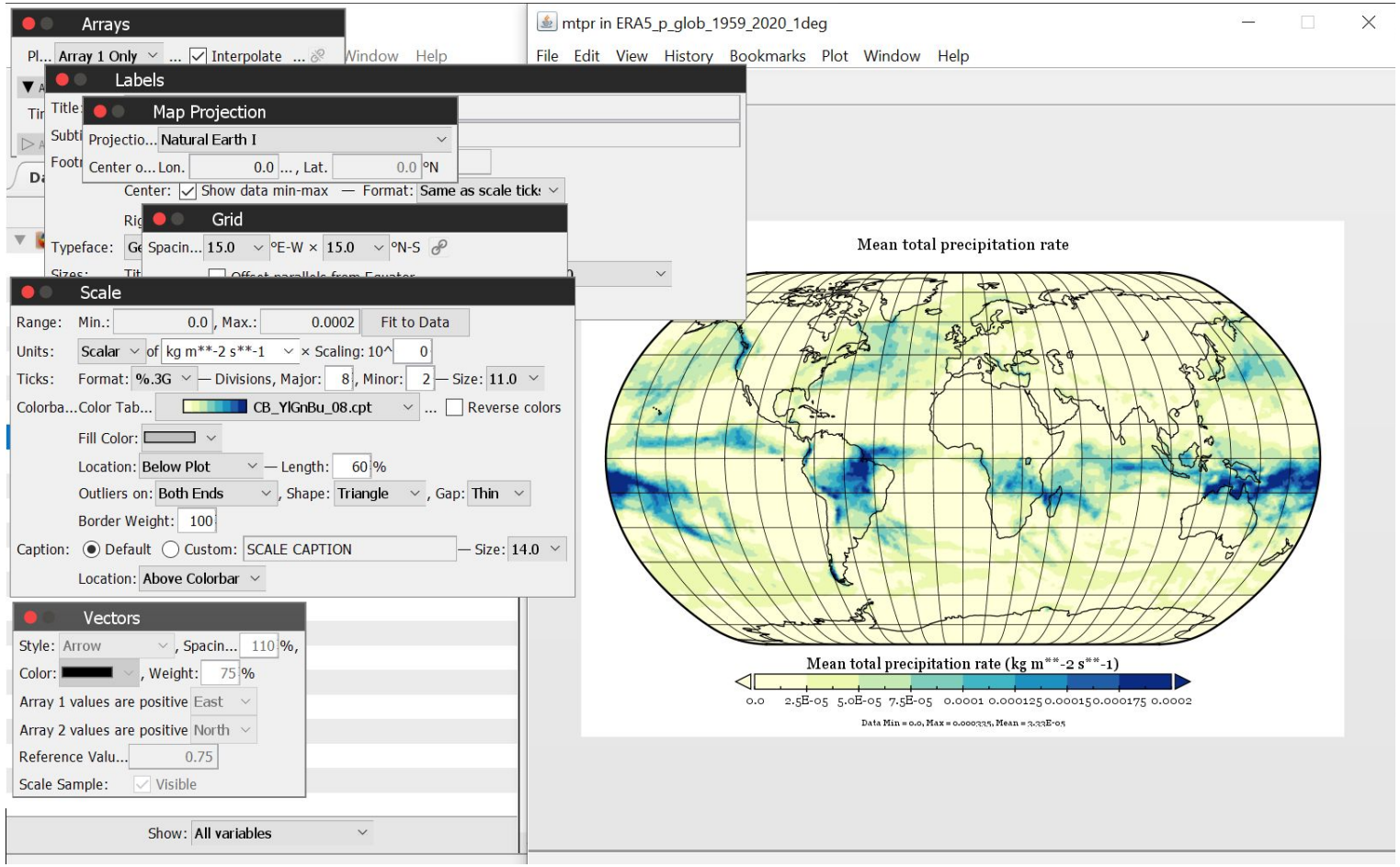
Panoply netCDF, HDF and GRIB Data Viewer

panoply \PAN-uh-plee\, noun: 1. A splendid or impressive array. ...

Panoply plots georeferenced and other arrays from netCDF, HDF, GRIB, etc.

Panoply is a cross-platform application; runs on Macintosh, Windows, Linux and other desktop computers.

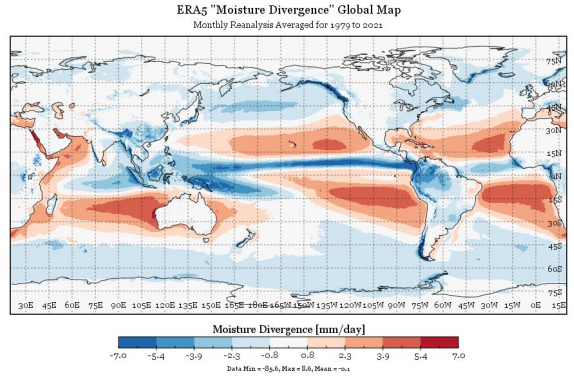
Must have: a compatible **Java 11** (or later version) JRE or JDK installed.

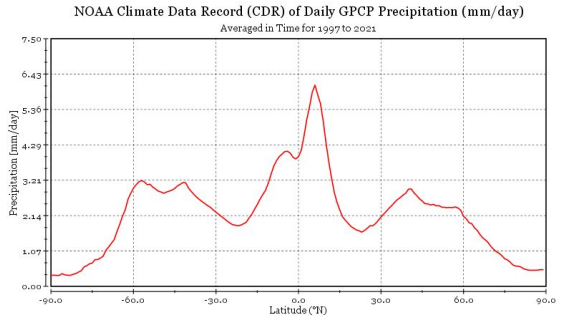


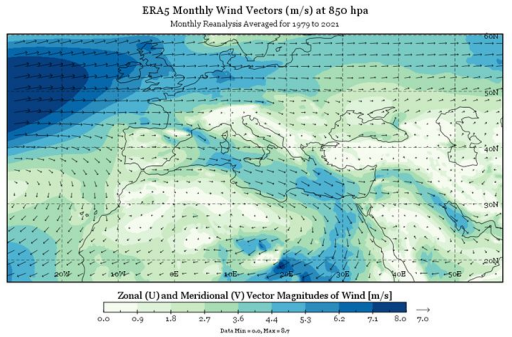
Panoply netCDF, HDF and GRIB Data Viewer

With ***Panoply 5*** you can:

- Create color contour plots of geo-referenced lat-lon, lat-vertical, lon-vertical, time-lat or time-vertical arrays from 2D or larger multi-D variables.
- Create line plots of data from 1D or larger multi-D variables.
- Combine two geo-referenced arrays in one plot by subtracting, summing, averaging, merging, etc.
- Plot lon-lat data on a global or regional map using over 200 map projections or make a zonal average line plot.
- Use numerous color tables for the scale colorbar, or apply your own custom RGB color table.
- Export animations as MP4 video, or collection of individual frames.





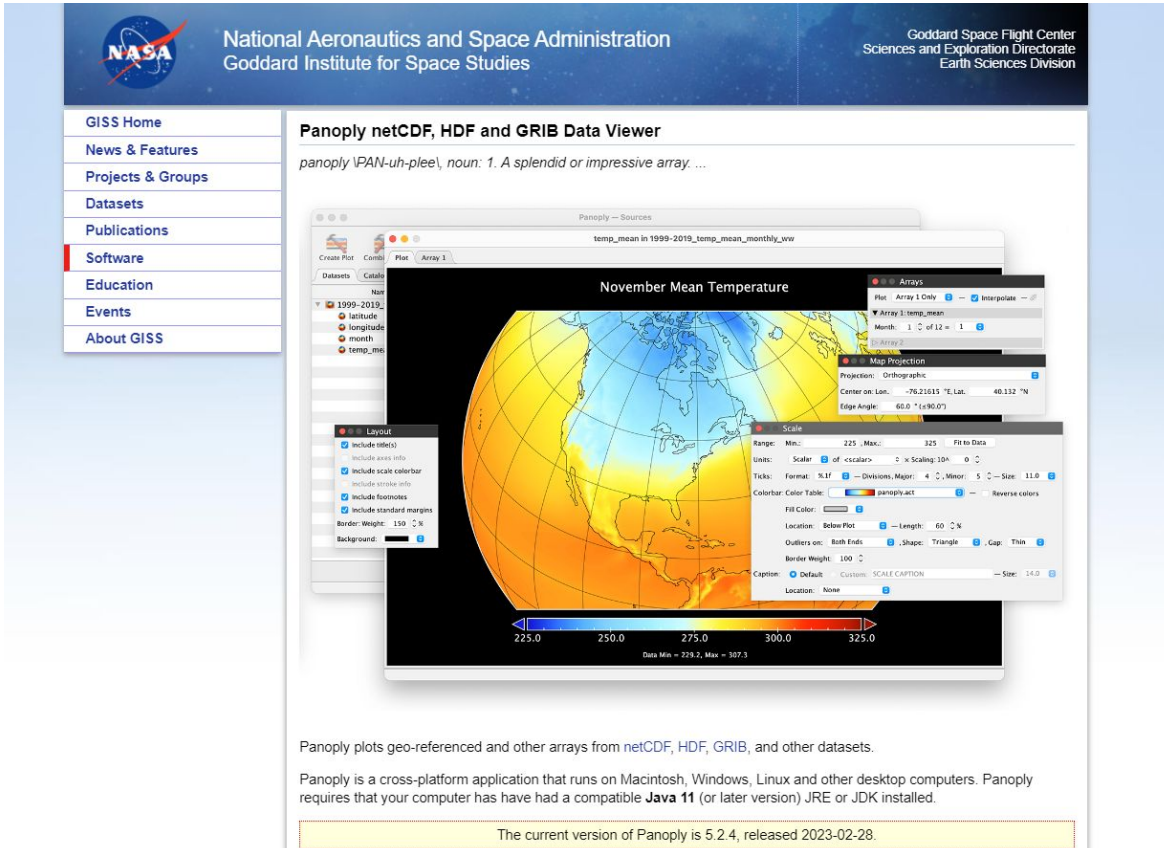


Panoply netCDF, HDF and GRIB Data Viewer

Where and how??

Google: **Panoply data viewer**

https://www.giss.nasa.gov/tools/panoply/



Scroll down and find the **Download Panoply** option.

Thank you!

Swamini Khurana (swamini.khurana@gmail.com)
Edoardo Martini (edoardo.martini@uni-leipzig.de)
Paola Mazzoglio (paola.mazzoglio@polito.it)
Epari Ritesh Patro (ritesh.patro@oulu.fi)
Roshanak Tootoonchi (roshanaktootoonchi94@gmail.com)

Young Hydrologic Society

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