HAS Tools:
Dos and don’ts of data vis.

October 23, 2024

Why is it important to think about data visualization?

When reading a new paper, it is one of the first, and most important things I will look at:

1. Paper title

Journal of Hydrology 645 (2024) 132080

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

ELSEVIER

Research papers

A blueprint for coupling a hydrological model with fine- and coarse-scale atmospheric regional climate change models for probabilistic streamflow projections

Chandra Rupa Rajulapati[a,b,*], Zelalem Tesemma[b], Kevin Shook[b], Simon Michael Papalexiou[b,c], John W Pomeroy[b]

[a] Department of Civil Engineering, University of Manitoba, MB, Canada
[b] Centre for Hydrology, University of Saskatchewan, SK, Canada
[c] Department of Civil Engineering, University of Calgary, AB, Canada

ARTICLE INFO

This manuscript was handled by Nandita Basu, Editor-in-Chief, with the assistance of Juliane Mai, Associate Editor

Keywords:
Streamflow projections
Hydrological modelling
Uncertainty
Floodplain mapping
Flow duration curves

ABSTRACT

In cold regions, climate change, including warming and changes in snowmelt dynamics, have profound impacts on streamflow patterns, often leading to flooding events. Understanding the projected changes in hydrometeorological factors contributing to streamflow, such as snowmelt runoff and rain-on-snowmelt, is crucial for effective adaptation and mitigation strategies. It is also essential to consider uncertainty in streamflow projections and incorporate this uncertainty into flood predictions. This study presents a blueprint for calculating probabilistic future streamflow and flow duration curves in a mountainous cold region. The methodology integrated forcings from an atmospheric model (WRF) with a hydrological model (MESH) at fine spatial (4 km) and temporal (3-hourly) resolutions. To account for uncertainty, an ensemble of 15 CanRCM4 regional climate simulations with varying boundary conditions for the RCP 8.5 scenario was utilized. A novel method was developed to perturb the CanRCM4 simulations' precipitation using a WRF Pseudo Global Warming run to incorporate uncertainty. This comprehensive approach revealed significant uncertainty in streamflow driven by internal variability. WRF-driven simulations without uncertainty showed a decrease in future streamflow extremes, while the perturbed simulations demonstrated the potential for substantially higher values under climate change. Moreover, the study derived probabilistic flow duration curves from the streamflow projections, aiding in estimating flood frequency for floodplain mapping when driving local hydraulic models. The methodology developed in this work can be extended to other river basins in Canada and elsewhere where gridded downscaled climate model outputs are available.

1. Introduction

Understanding historical changes in streamflow and accurately predicting future responses are vital for effective adaptation and mitigation strategies. Reliable streamflow projections are essential for developing appropriate measures to address these challenges. In cold regions, climate change impacts, including warming and alterations in snowmelt dynamics, have significant implications for streamflow patterns, often leading to flooding events (Fang & Pomeroy, 2007; Rasouli et al., 2022; Trenberth, 2011). Flooding in these regions can arise from

Accurate floodplain mapping and delineation are critical for urban planning and development in proximity to river floodplains. Additionally, it is important to account for uncertainty in streamflow projections and propagate this uncertainty in flood predictions (Table 1).

Hydrological modelling in cold regions presents significant challenges due to the intricate nature of systems influenced by heat transfer processes within snowpacks, glaciers, and soils, as well as the interplay between snow cover and freeze-thaw cycles (Wheater et al., 2022; Yang et al., 2021). Moreover, the availability of data for model evaluation and calibration is limited, primarily due to sparse ground-based observa-

Why is it important to think about data visualization?

When reading a new paper, it is one of the first, and most important things I will look at:

1. Paper title
2. Author list

Journal of Hydrology 645 (2024) 132080

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

ELSEVIER

Research papers

A blueprint for coupling a hydrological model with fine- and coarse-scale atmospheric regional climate change models for probabilistic streamflow projections

Chandra Rupa Rajulapati [a,b,*], Zelalem Tesemma [b], Kevin Shook [b], Simon Michael Papalexiou [b,c], John W Pomeroy [b]

[a] Department of Civil Engineering, University of Manitoba, MB, Canada
[b] Centre for Hydrology, University of Saskatchewan, SK, Canada
[c] Department of Civil Engineering, University of Calgary, AB, Canada

ARTICLE INFO

This manuscript was handled by Nandita Basu, Editor-in-Chief, with the assistance of Juliane Mai, Associate Editor

Keywords:
Streamflow projections
Hydrological modelling
Uncertainty
Floodplain mapping
Flow duration curves

ABSTRACT

In cold regions, climate change, including warming and changes in snowmelt dynamics, have profound impacts on streamflow patterns, often leading to flooding events. Understanding the projected changes in hydrometeorological factors contributing to streamflow, such as snowmelt runoff and rain-on-snowmelt, is crucial for effective adaptation and mitigation strategies. It is also essential to consider uncertainty in streamflow projections and incorporate this uncertainty into flood predictions. This study presents a blueprint for calculating probabilistic future streamflow and flow duration curves in a mountainous cold region. The methodology integrated forcings from an atmospheric model (WRF) with a hydrological model (MESH) at fine spatial (4 km) and temporal (3-hourly) resolutions. To account for uncertainty, an ensemble of 15 CanRCM4 regional climate simulations with varying boundary conditions for the RCP 8.5 scenario was utilized. A novel method was developed to perturb the CanRCM4 simulations' precipitation using a WRF Pseudo Global Warming run to incorporate uncertainty. This comprehensive approach revealed significant uncertainty in streamflow driven by internal variability. WRF-driven simulations without uncertainty showed a decrease in future streamflow extremes, while the perturbed simulations demonstrated the potential for substantially higher values under climate change. Moreover, the study derived probabilistic flow duration curves from the streamflow projections, aiding in estimating flood frequency for floodplain mapping when driving local hydraulic models. The methodology developed in this work can be extended to other river basins in Canada and elsewhere where gridded downscaled climate model outputs are available.

1. Introduction

Understanding historical changes in streamflow and accurately predicting future responses are vital for effective adaptation and mitigation strategies. Reliable streamflow projections are essential for developing appropriate measures to address these challenges. In cold regions, climate change impacts, including warming and alterations in snowmelt dynamics, have significant implications for streamflow patterns, often leading to flooding events (Fang & Pomeroy, 2007; Rasouli et al., 2022; Trenberth, 2011). Flooding in these regions can arise from

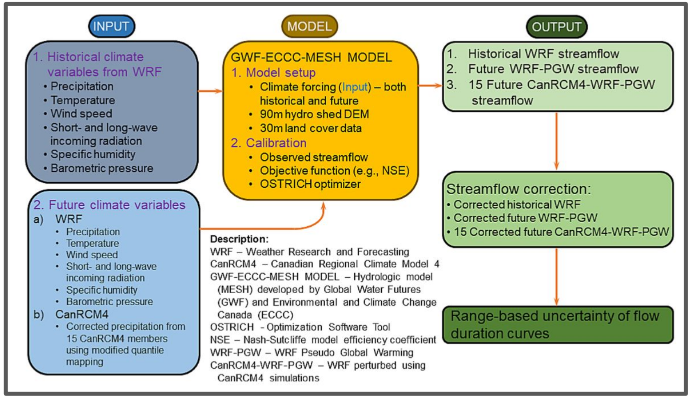
Accurate floodplain mapping and delineation are critical for urban planning and development in proximity to river floodplains. Additionally, it is important to account for uncertainty in streamflow projections and propagate this uncertainty in flood predictions (Table 1).

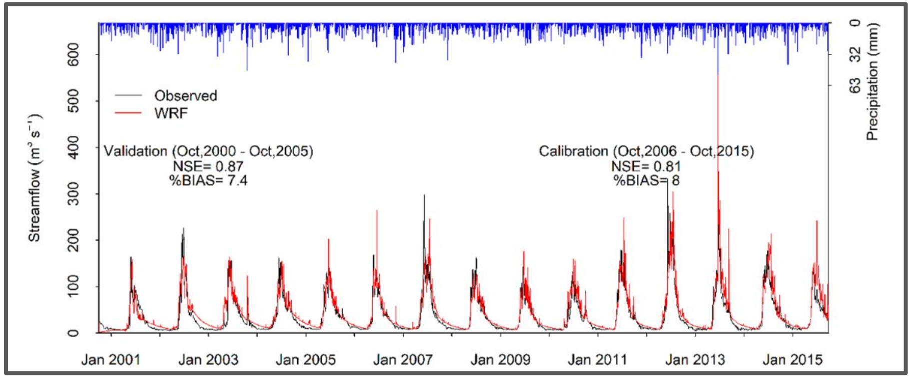
Hydrological modelling in cold regions presents significant challenges due to the intricate nature of systems influenced by heat transfer processes within snowpacks, glaciers, and soils, as well as the interplay between snow cover and freeze-thaw cycles (Wheater et al., 2022; Yang et al., 2021). Moreover, the availability of data for model evaluation and calibration is limited, primarily due to sparse ground-based observa-

Why is it important to think about data visualization?

When reading a new paper, it is one of the first, and most important things I will look at:

1. Paper title
2. Author list
3. If I know the authors' work:
    a. Straight to the figures
    b. Otherwise, read the abstract





How do we judge data visualizations?

Let’s start with a bad example.

How do we judge data visualizations?

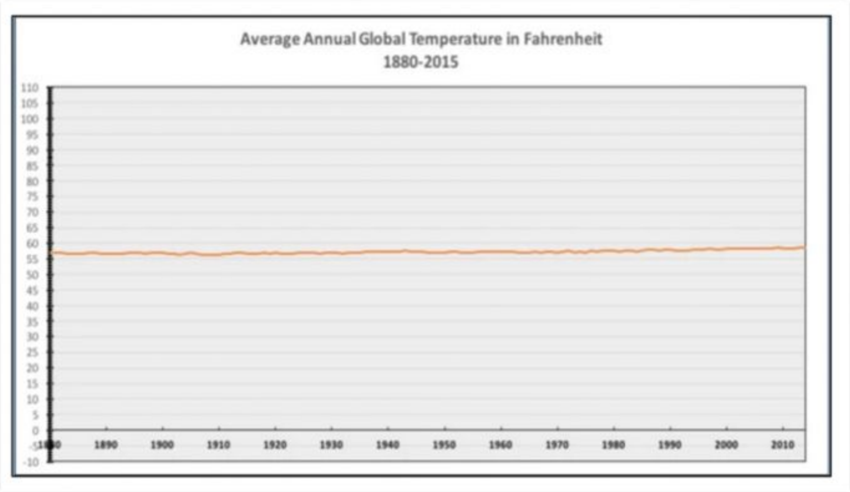
Let's start with a bad example.

National Review @NRO

Follow

The only #climatechange chart you need to see. natl.re/wPKpro

(h/t @powerlineUS)



1:36 PM - 14 Dec 2015

How do we judge data visualizations?

Let’s start with a bad example.

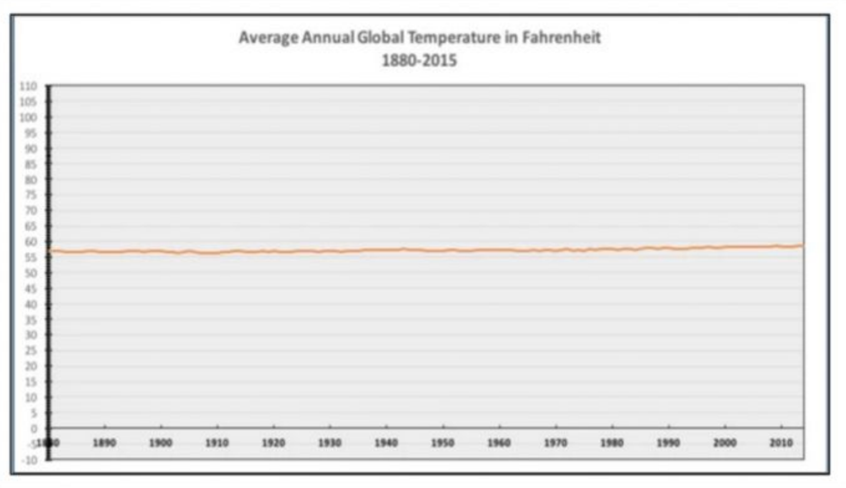
So, *why* is this bad?

National Review
@NRO

Follow

The only #climatechange chart you need to see. natl.re/wPKpro

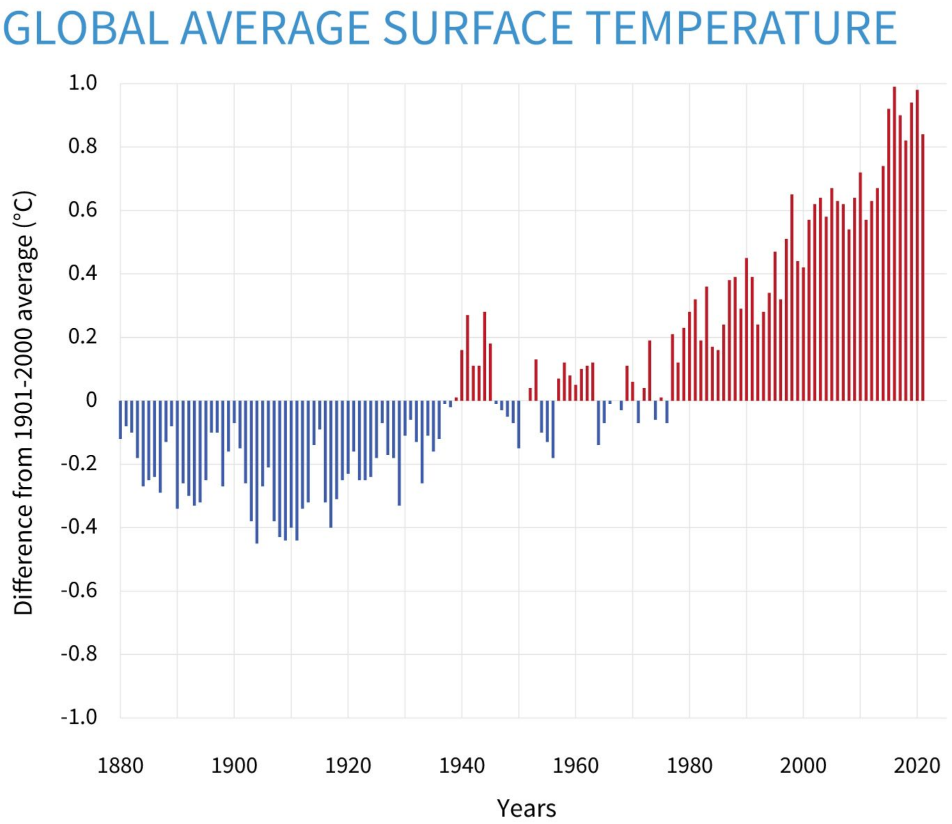
(h/t @powerlineUS)



1:36 PM - 14 Dec 2015

How do we judge data visualizations?

How about an example of a better visualization?

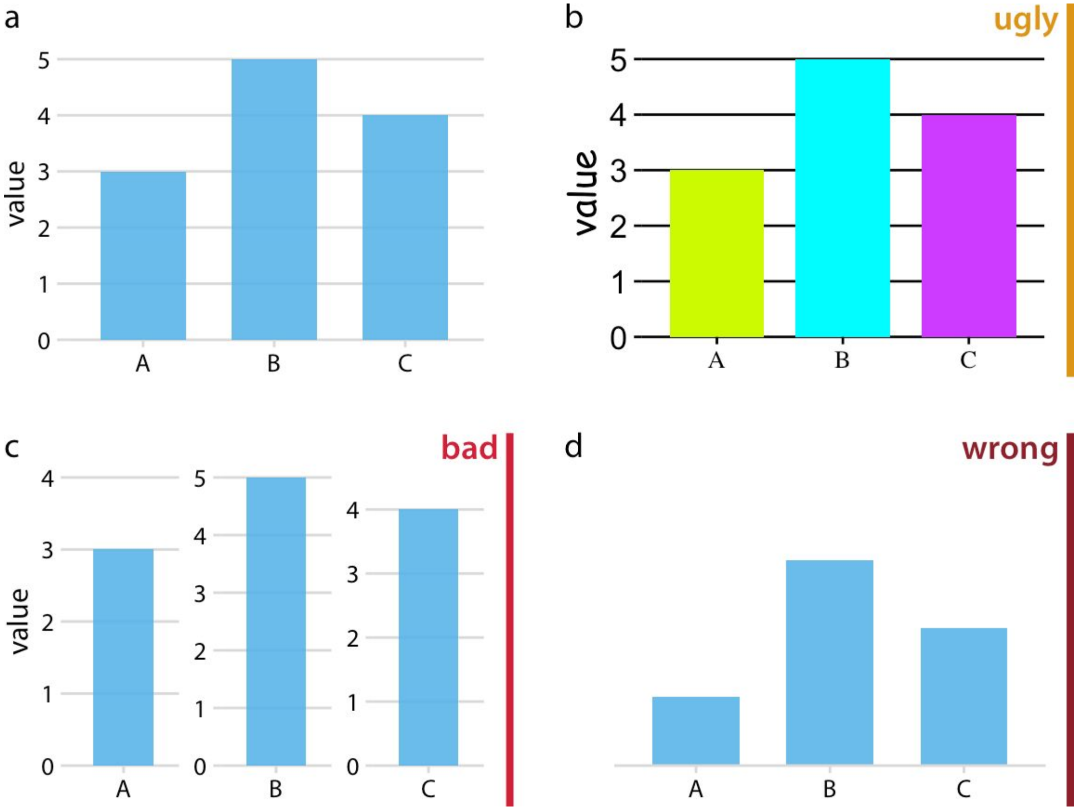


When creating data visualizations you have many choices

Oftentimes, we hope that the defaults of the software will do us good

Inevitably, you will need to go beyond the defaults though

The purpose of this lecture is to provide some tools/concepts to make good decisions when presenting your data



Visualization is a form of communicating complex information

Let’s start with an example

Visualization is a form of communicating complex information

Let’s start with an example

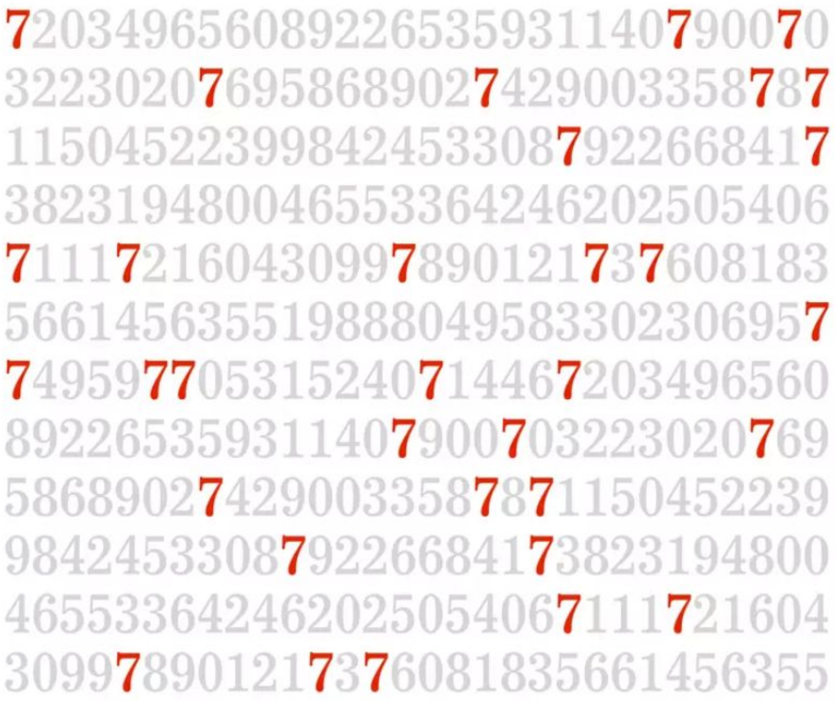
How many ‘7’s are there?

720349656089226535931140790070
322302076958689027429003358787
115045223998424533087922668417
382319480046553364246202505406
711172160430997890121737608183
566145635519888049583302306957
749597705315240714467203496560
892265359311407900703223020769
586890274290033587871150452239
984245330879226684173823194800
465533642462025054067111721604
309978901217376081835661456355

Visualization is a form of communicating complex information

Let's start with an example

How many '7's are there?

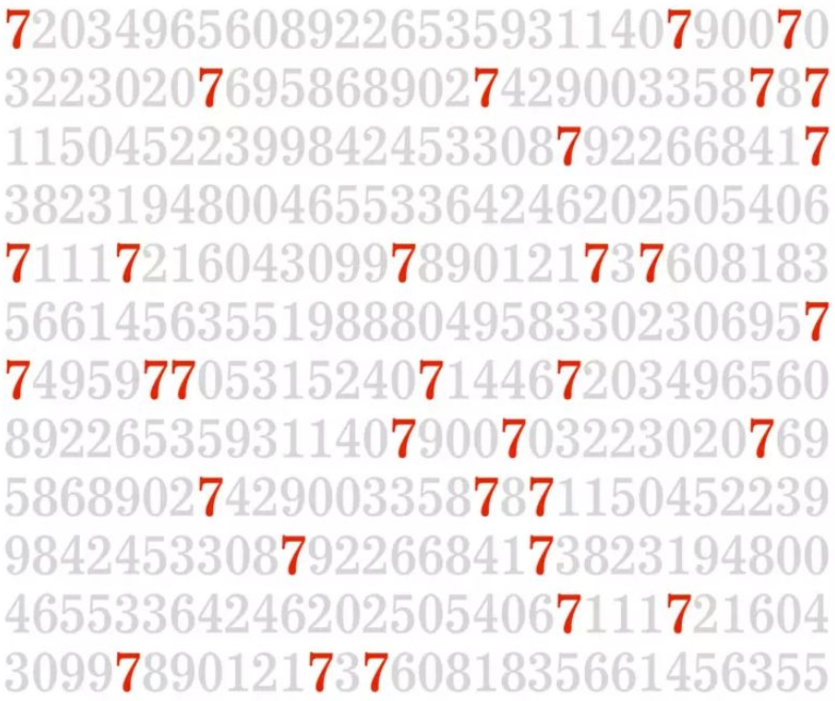


Visualization is a form of communicating complex information

Let’s start with an example

How many ‘7’s are there?

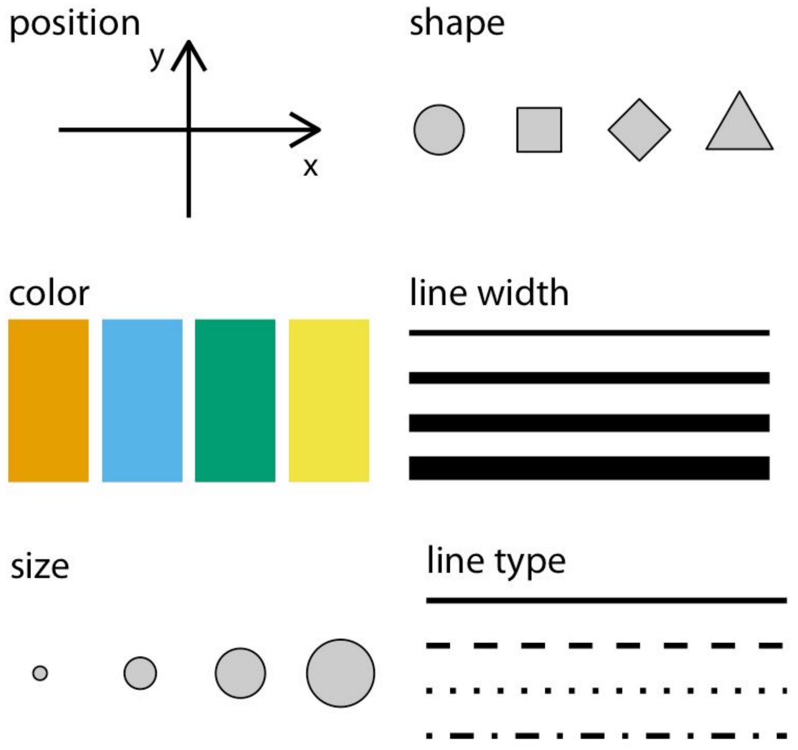
Much easier, but how did that work?



On aesthetics: tools you have to create good visualizations

Visualizations of data are made up of come common components, or use various shared concepts

Choosing the right combination of these components is the challenge at hand for making good plots

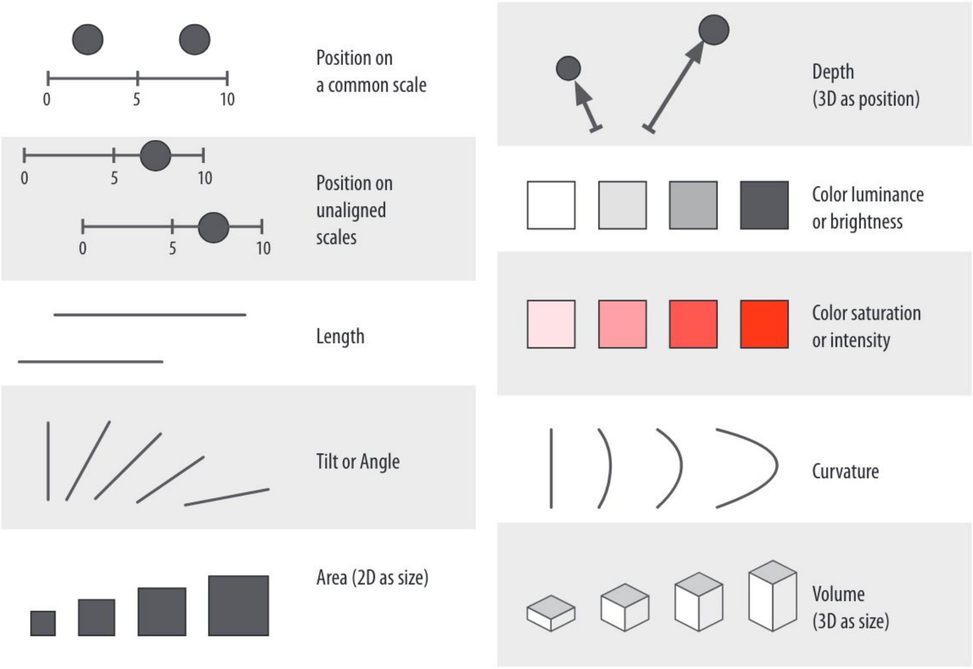


On aesthetics: tools you have to create good visualizations

Visualizations of data are made up of come common components, or use various shared concepts

Choosing the right combination of these components is the challenge at hand for making good plots

Of course, this is not a strict science, so there are multiple other interpretations of these components



https://assets.press.princeton.edu/chapters/s13826.pdf

From these tools you have almost infinite possibilities

A corollary of this is that there are many types of data visualizations - nearly endless!

There are a few basic types that I think are most commonly used in scientific data, and geoscience more specifically

While other chart types may look cool, I advise sticking to standards whenever possible

https://datavizcatalogue.com/home_list.html

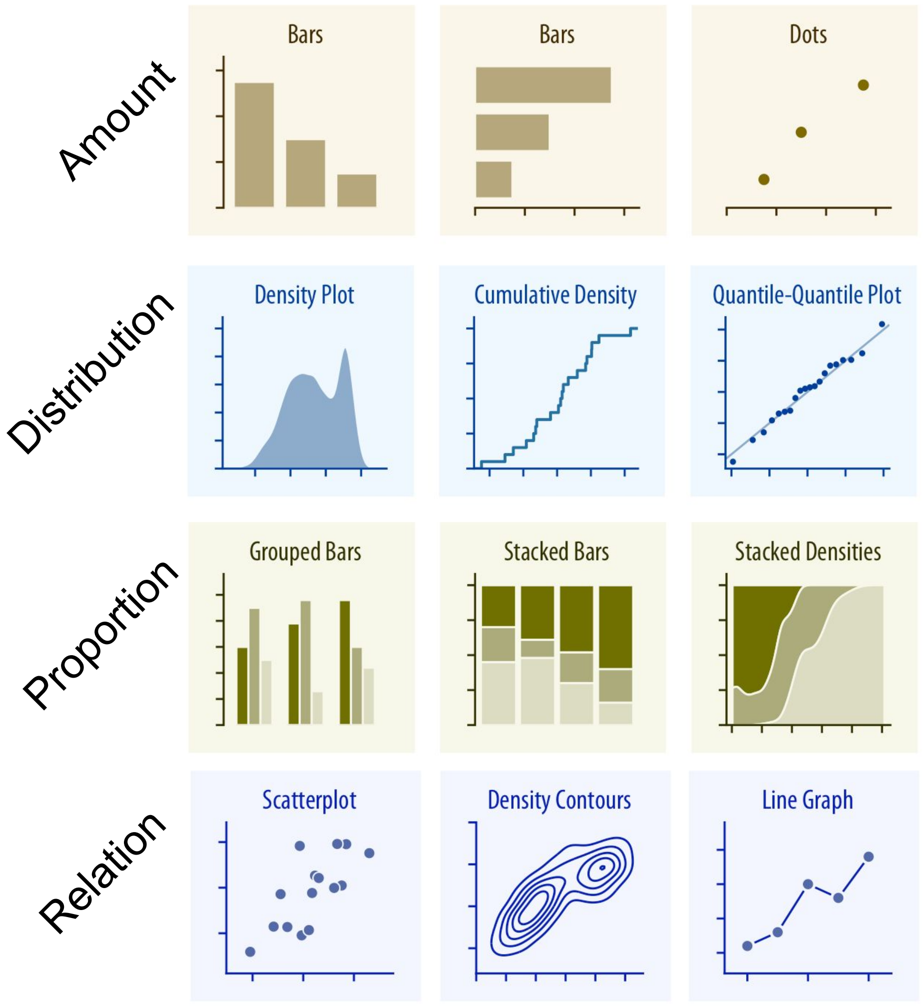
In scientific data vis, there tend to be 4 styles of plots

Amount: Pretty self explanatory - how much of a thing is there

Distribution: What is the spread, or how much stuff is in a given region/time/place

Proportion: How does the amount of one thing compare to another

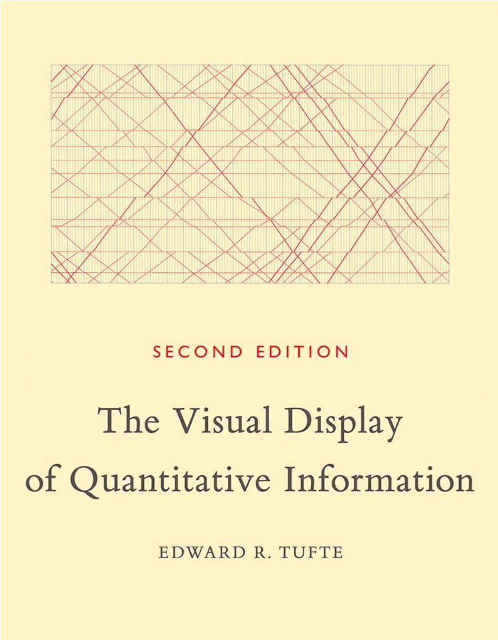
Relation: Do changes in one thing seem to be associated with a change in another thing



Even once you have chosen a chart type, there is much work to be done.

Some guiding principles for making effective scientific visualizations:

1. "Create the simplest graph that conveys the information you want to convey" - Edward Tufte
2. Select an appropriate color scheme based on the type of data



nature COMMUNICATIONS

PERSPECTIVE

Check for updates

https://doi.org/10.1038/s41467-020-19160-7 OPEN

The misuse of colour in science communication

Fabio Crameri[1✉], Grace E. Shephard[1] & Philip J. Heron[2]

The accurate representation of data is essential in science communication. However, colour maps that visually distort data through uneven colour gradients or are unreadable to those with colour-vision deficiency remain prevalent in science. These include, but are not limited to, rainbow-like and red-green colour maps. Here, we present a simple guide for the scientific use of colour. We show how scientifically derived colour maps report true data variations, reduce complexity, and are accessible for people with colour-vision deficiencies. We highlight ways for the scientific community to identify and prevent the misuse of colour in science, and call for a proactive step away from colour misuse among the community, publishers, and the press.

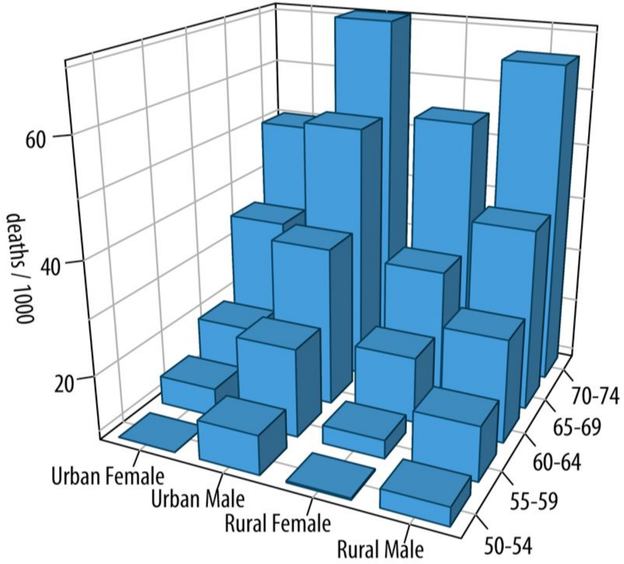
Vision is one of the most fundamental means of communication. It is (or should be) in every scientist's best intention to make figures and their content as accurate and easily understandable as possible. One of the most powerful aspects of images is colour, which in turn transforms information into meaning. The visual evaluation of a colour gradient is important to a variety of different fields such as the first direct impression of a black hole[1], the mapping of votes cast in political elections[2,3], the planning of an expensive rover route on Martian topography[4], the essential communication of climate change[5,6], or the critical diagnosis of heart disease[7]. However, when colours are used incorrectly, this can lead to the effective manipulation of data (e.g., by highlighting some data over others), the oversight of the needs of those with colour vision deficiencies, and the removal of meaning when printed in black and white (Supplementary Note 1).

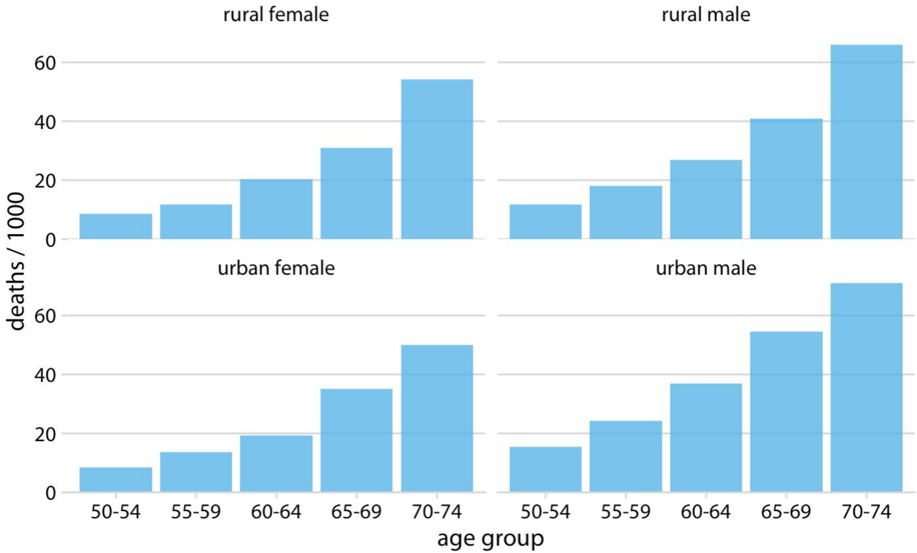
As science has become more prevalent in mainstream culture, it is not only the scientific community that suffers due to the use of poor colour choices, but also the wider public. Colour maps, therefore, are a crucial intersection between science and society. For instance, weather forecasts and hazard maps are two examples of immediately societal-relevant data sets that are also repeat offenders for use of the rainbow-like colour maps. Given the (daily) importance of these scientific topics, the underlying data should be conveyed in a universal manner. However, the colour-vision deficient fraction of the population is excluded and therefore unable to process this critical information. Furthermore, zones of danger, such as the boundaries of a hurricane track or current virus spread, are often based on uneven colour gradients to accentuate their importance. Using an uneven colour gradient is not an action without consequences, including those with significant financial or life-threatening consequences. Decisions based on data being 'unfairly' represented could produce, for instance, a Martian rover being sent over terrain that is too steep as the topography was inaccurately visualised, or a medical worker making an incomplete or inaccurate diagnosis based on uneven colour gradients.

Although some scientific communities have largely moved away from using distorting colour maps, such as rainbow, there are numerous signs of bad habits returning en masse[8]. Unfortunately, the previous efforts within specific disciplines to discredit rainbow-like maps appear to

[1] Centre for Earth Evolution and Dynamics (CEED), University of Oslo, Postbox 1028, Blindern, 0315 Oslo, Norway. [2] Department of Earth Sciences, Durham University, Durham, UK. ✉email: fabio.crameri@geo.uio.no

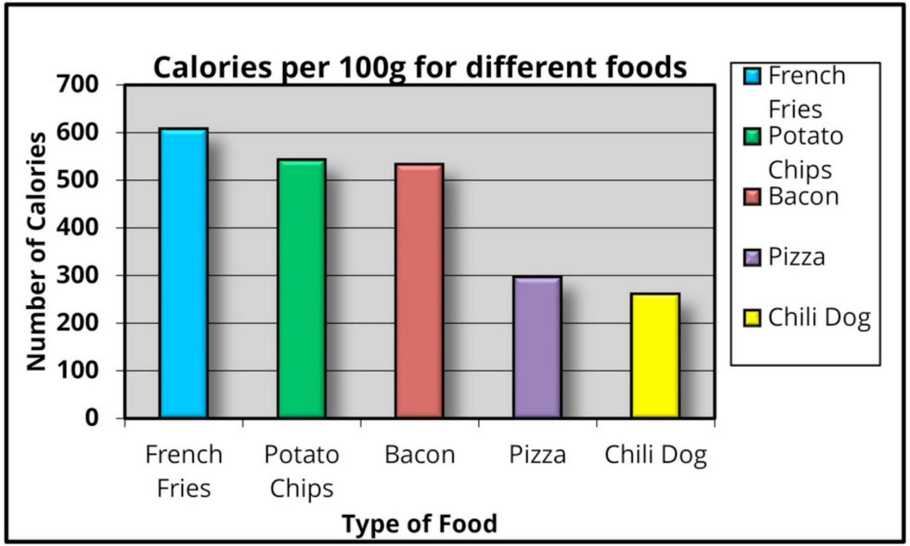
Create the simplest graph that conveys the information you want to convey

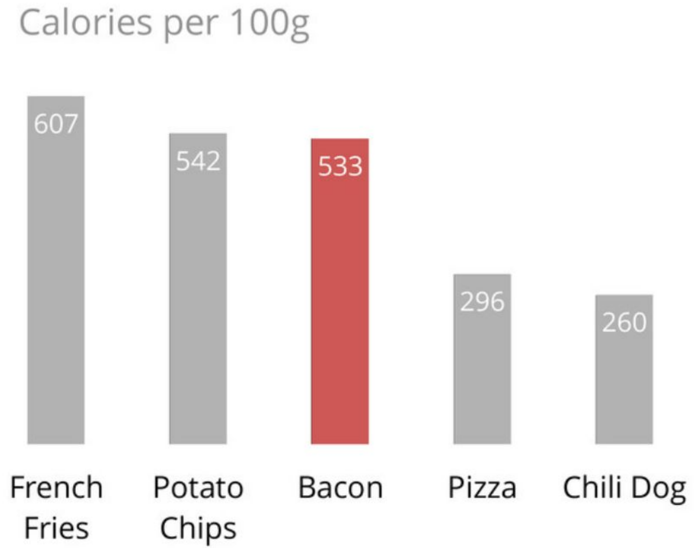




Create the simplest graph that conveys the information you want to convey

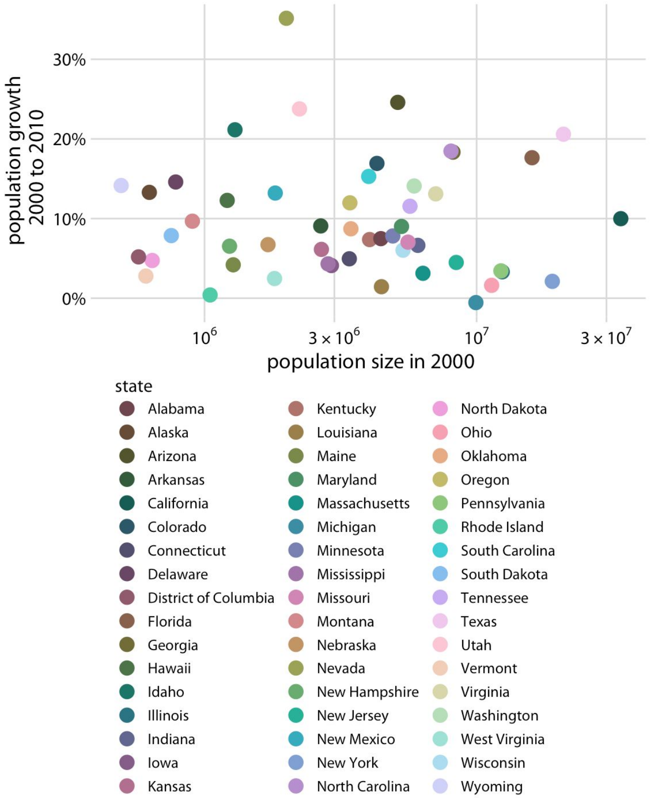
https://www.darkhorseanalytics.com/blog/data-looks-better-naked

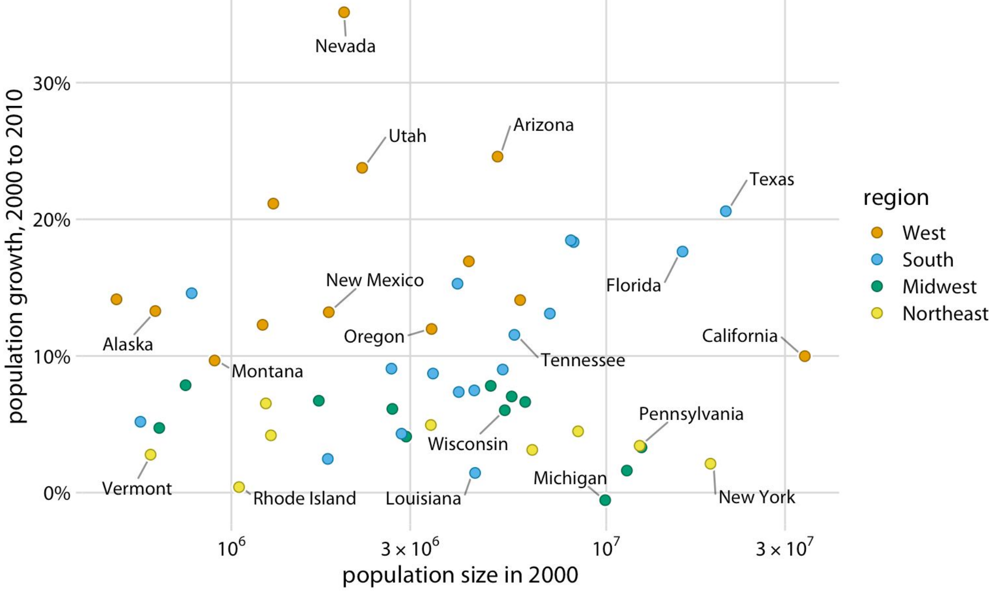




Select an appropriate color scheme based on the type of data

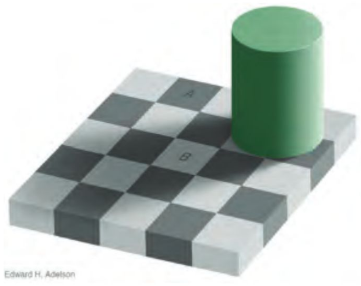
Don’t inundate your audience with too much information

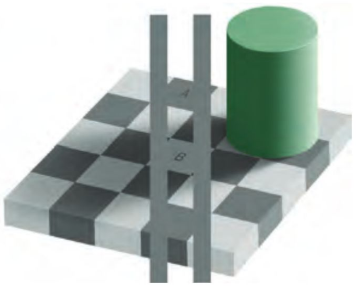


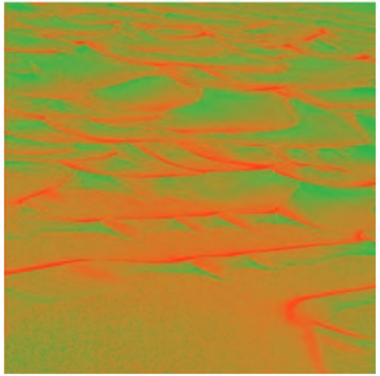


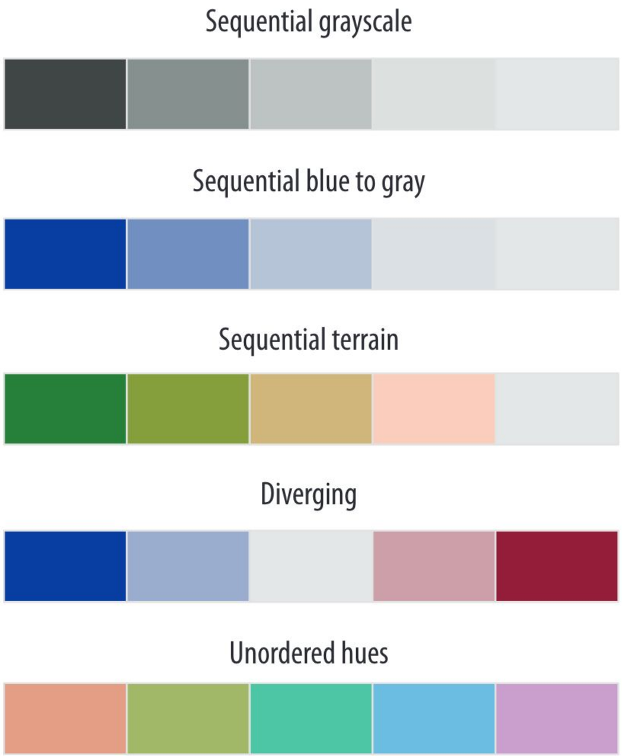
Select an appropriate color scheme based on the type of data

Our perceptual systems have biases, choose colormaps that avoid pitfalls



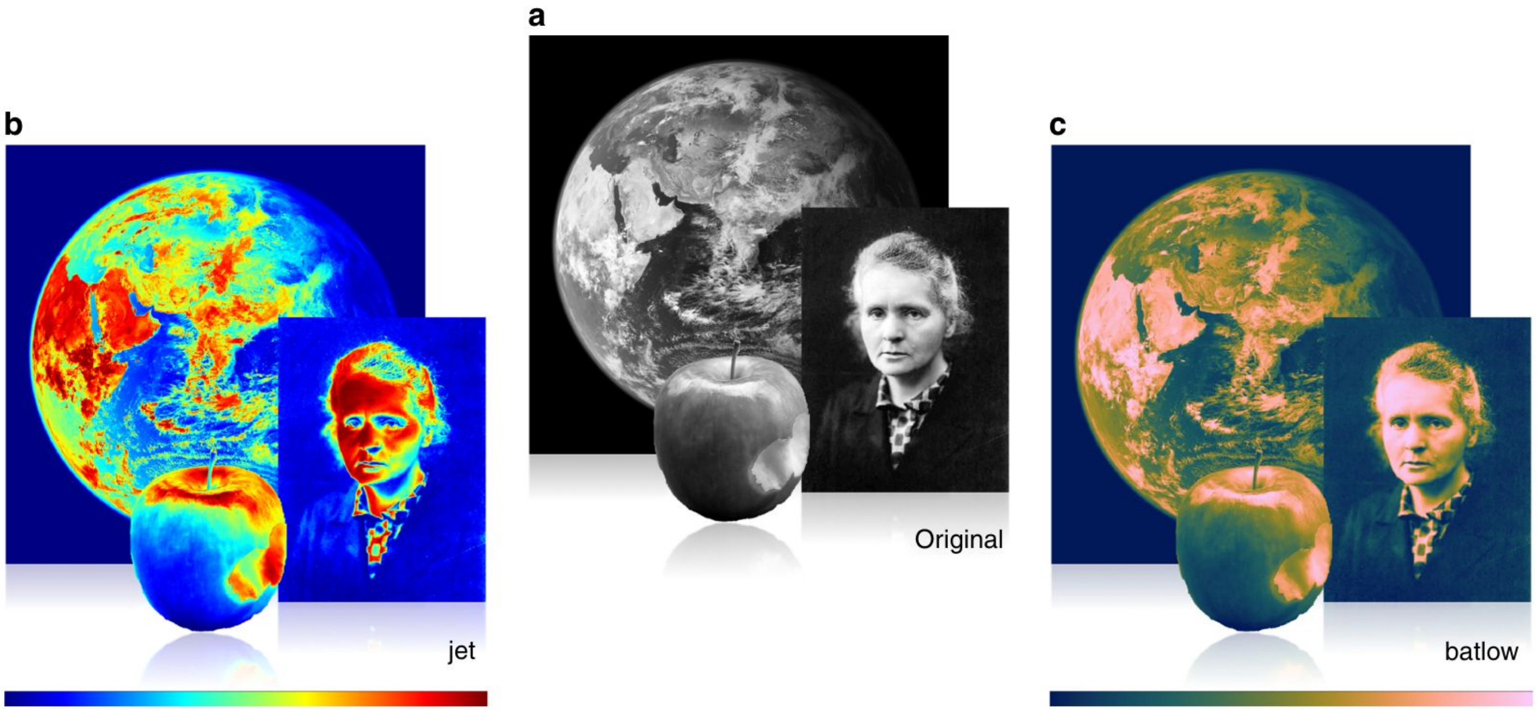






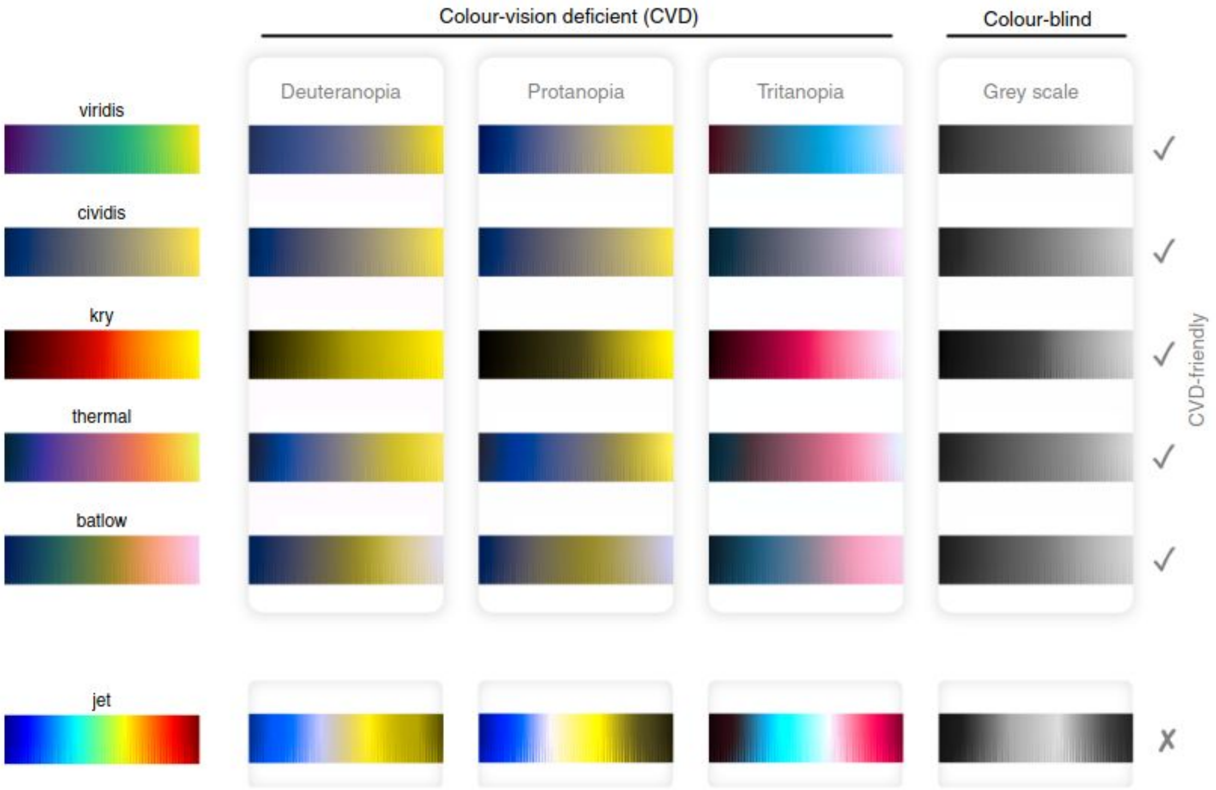
Select an appropriate color scheme based on the type of data

Our perceptual systems have biases, choose colormaps that avoid pitfalls



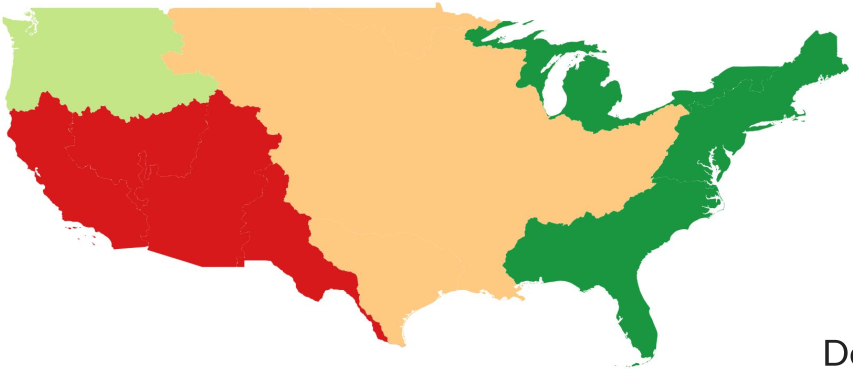
https://doi.org/10.1038/s41467-020-19160-7

Color choice is not just a matter of taste, but accessibility too!



Why are Colorblind-friendly Palettes Important?

Original

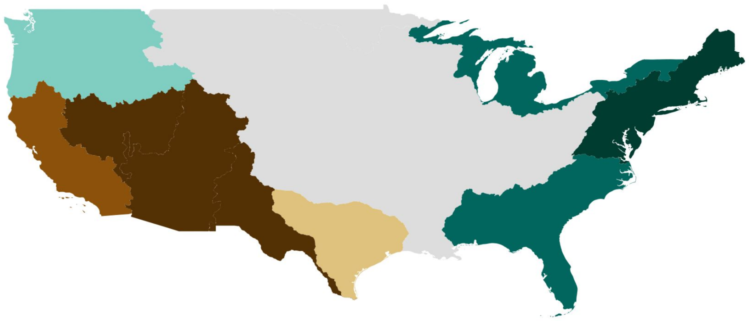


Protanomaly

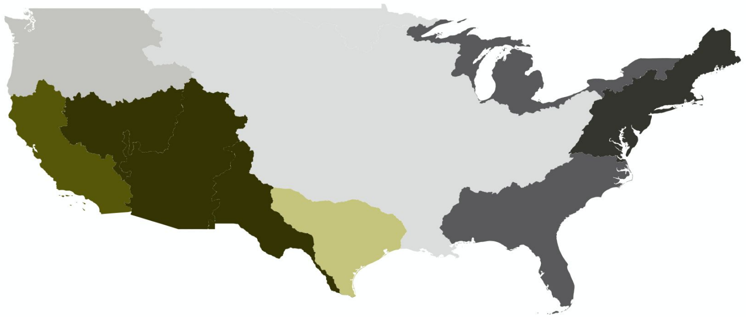
Deuteranomaly

Why are Colorblind-friendly Palettes Important?

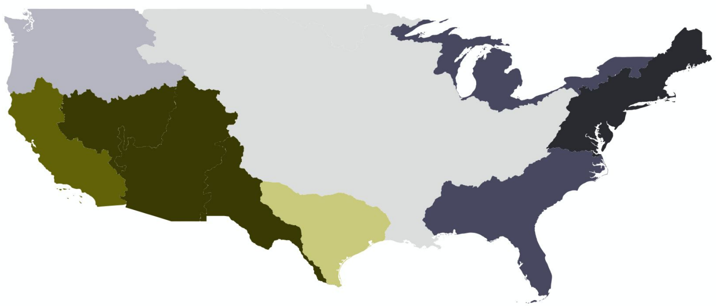
Original



Protanomaly



Deuteranomaly



Some resources for guidance

https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3

https://sashamaps.net/docs/resources/20-colors/

https://davidmathlogic.com/colorblind/

https://www.tpgi.com/color-contrast-checker/

https://www.perceptualedge.com/articles/ie/the_right_graph.pdf

https://clauswilke.com/dataviz/