NREL
Transforming ENERGY

A Picture is Worth a Thousand Data Points: Introduction to Visualization

Andy Berres (they/them)
National Renewable Energy Laboratory

ACM Tapia Conference 2023
September 14, 2023

Photo by Dennis Schroeder, NREL 55200

Quick Bio

- BS, MS, PhD Computer Science
  - Minors: Mathematics (BS)/Biology (MS)
  - Research: Visualization, Geometry, Topology
- These days: Applied Data Science
  - Climate
  - Traffic
  - Buildings
  - Power Grid

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Los Alamos
NATIONAL LABORATORY
EST.1943

OAK
RIDGE
National Laboratory

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Contents

1 Why should you care

2 What not to do

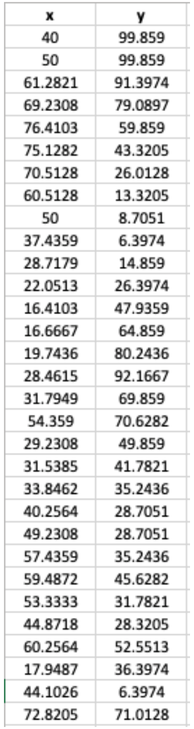
3 What to do instead

4 Hands-on practice

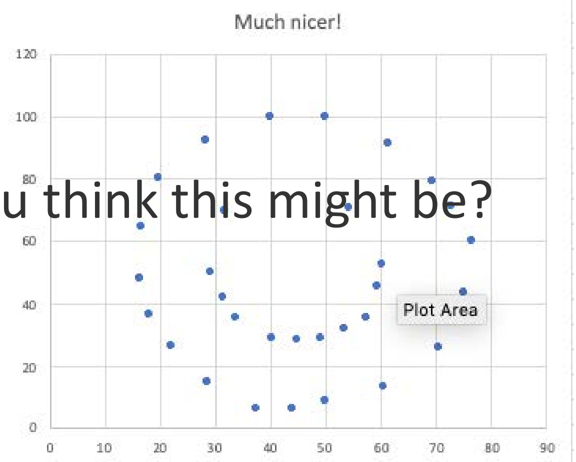
5 Resources

Why should you care?

Showcase 1: Turn that frown upside down!



x	y
40	99.859
50	99.859
61.2821	91.3974
69.2308	79.0897
76.4103	59.859
75.1282	43.3205
70.5128	26.0128
60.5128	13.3205
50	8.7051
37.4359	6.3974
28.7179	14.859
22.0513	26.3974
16.4103	47.9359
16.6667	64.859
19.7436	80.2436
28.4615	92.1667
31.7949	69.859
54.359	70.6282
29.2308	49.859
31.5385	41.7821
33.8462	35.2436
40.2564	28.7051
49.2308	28.7051
57.4359	35.2436
59.4872	45.6282
53.3333	31.7821
44.8718	28.3205
60.2564	52.5513
17.9487	36.3974
44.1026	6.3974
72.8205	71.0128

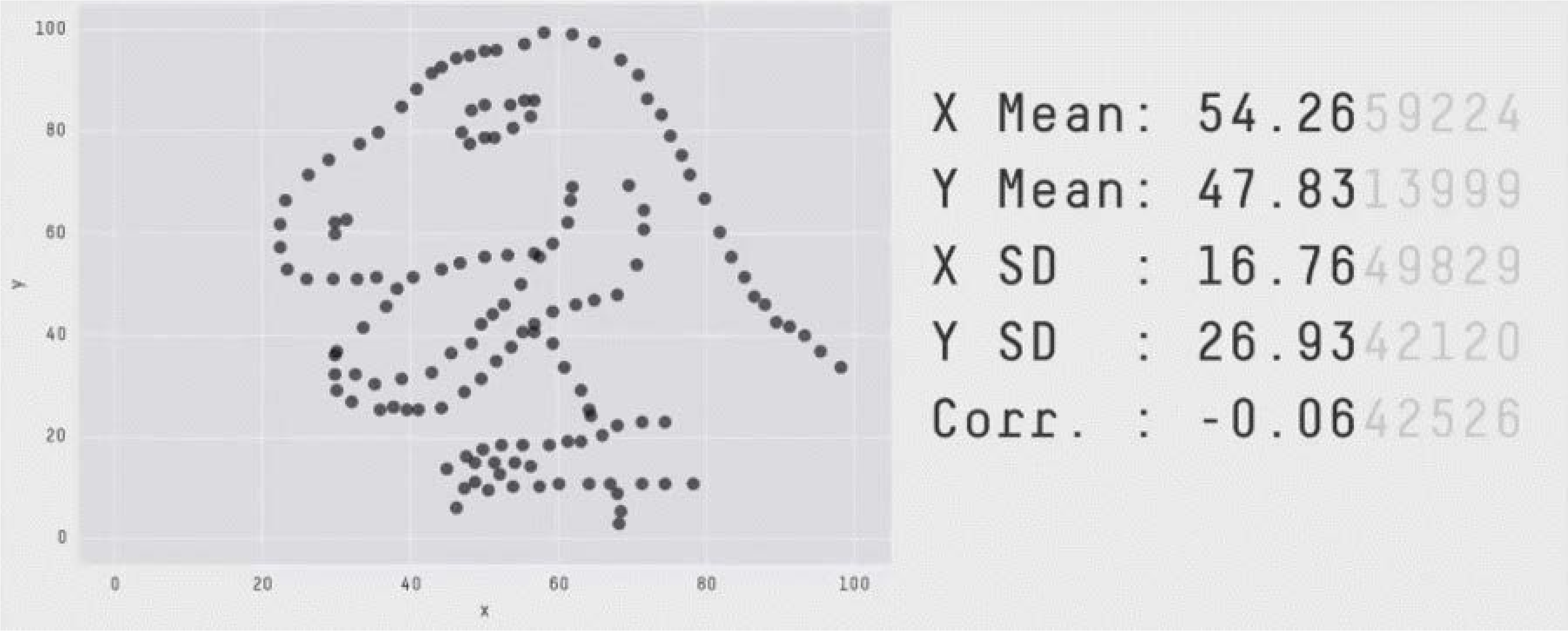


What do you think this might be?

☹

☺

Showcase 2: Datasaurus Dozen



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Matejka, Justin, and George Fitzmaurice. "Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1290-1294. ACM, 2017.

Why Should You Care? – Recap

Visualization will help you understand your data better.

Visualization gives you information at a glance.

Visualization is a form of communication.

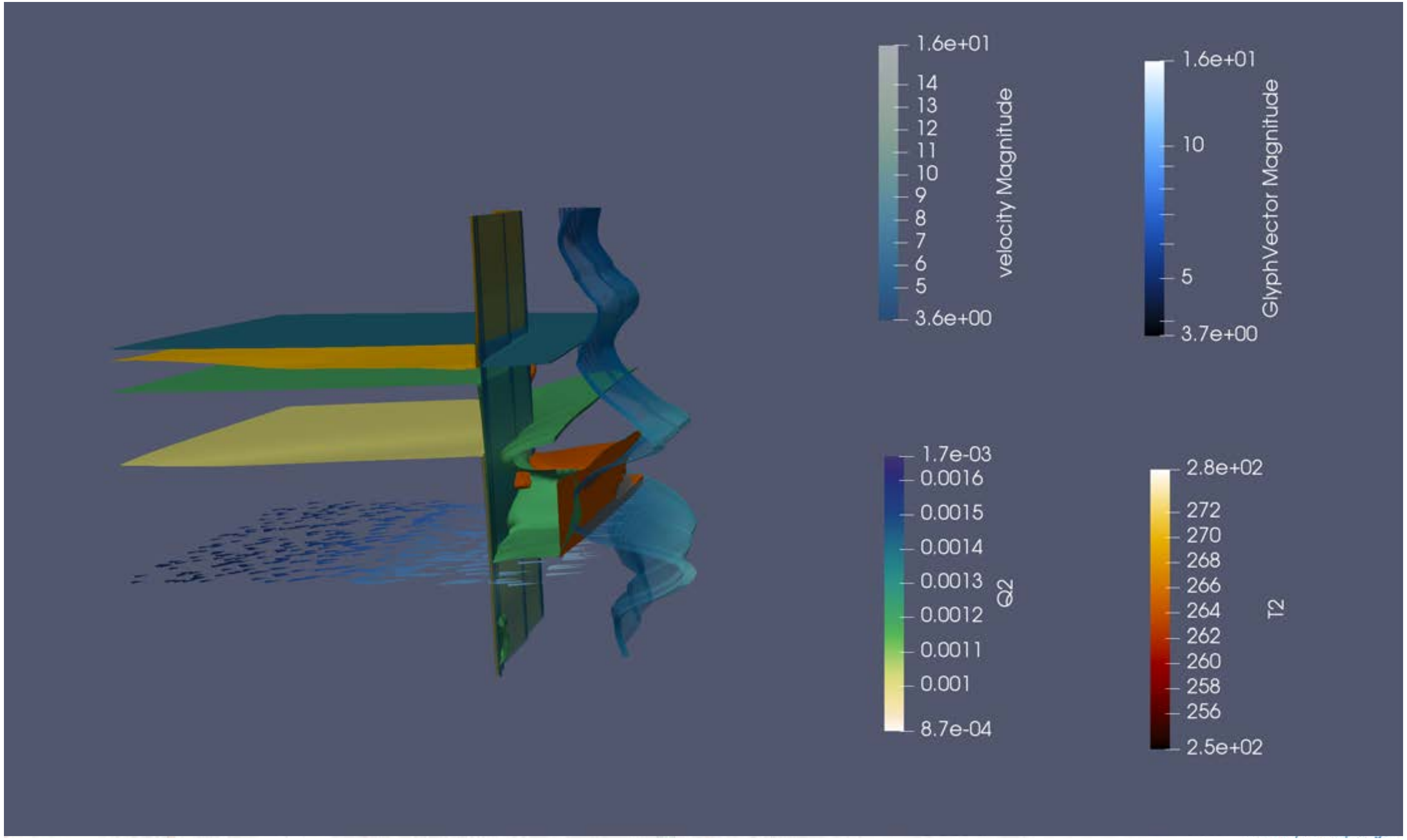
What not to do

- The many ways you can mess things up
- Don’t try this at home (except do)
- A game of “What’s wrong with this?”

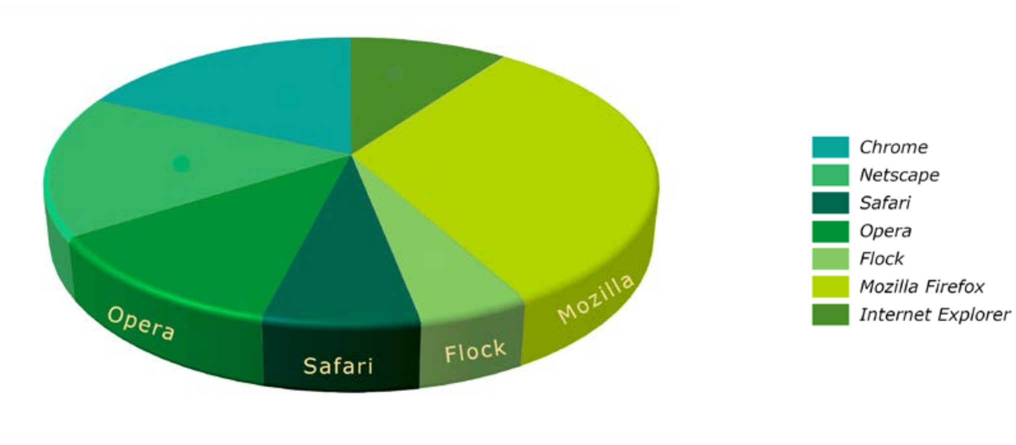
Disclaimer

- Examples shown serve illustrative purposes.
  - Some of them are actually pretty good
  - Some commit a whole list of sins
  - Some of them are used to advertise tool features

Too much information

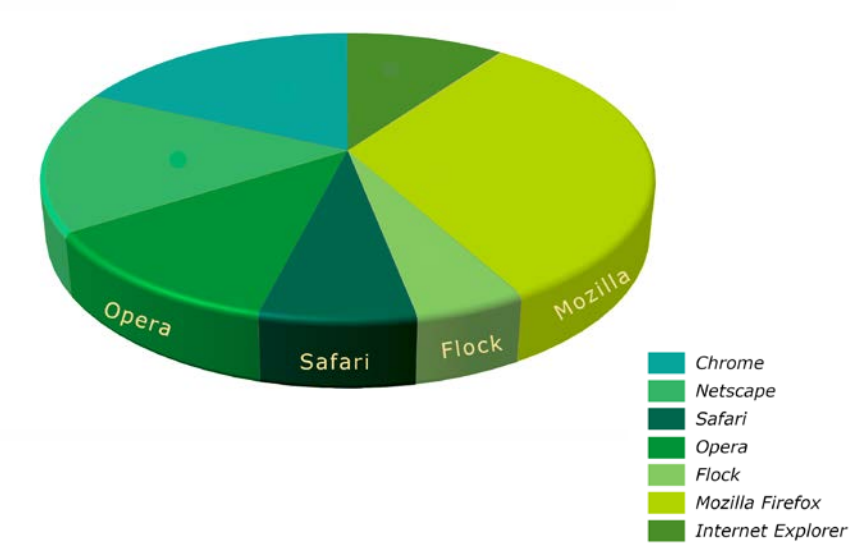


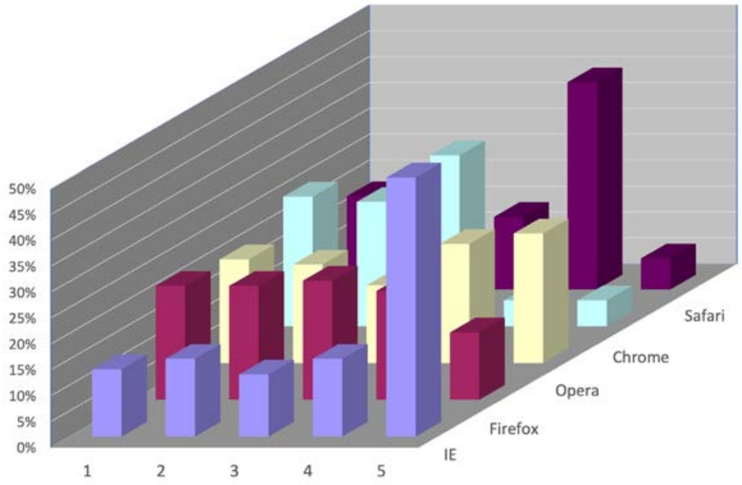
Inappropriate Color Choice



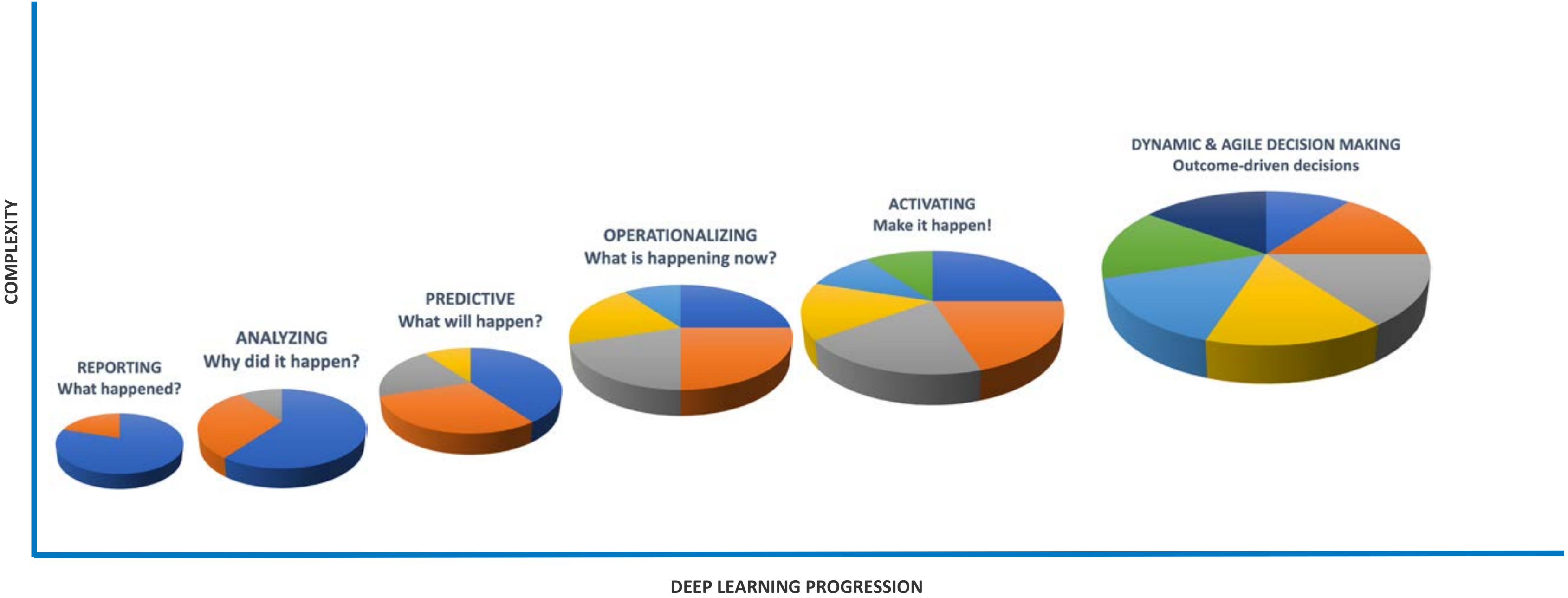
Useless 3D

- Humans are phenomenally bad at comparing volumes
- 3D often leads to occlusion



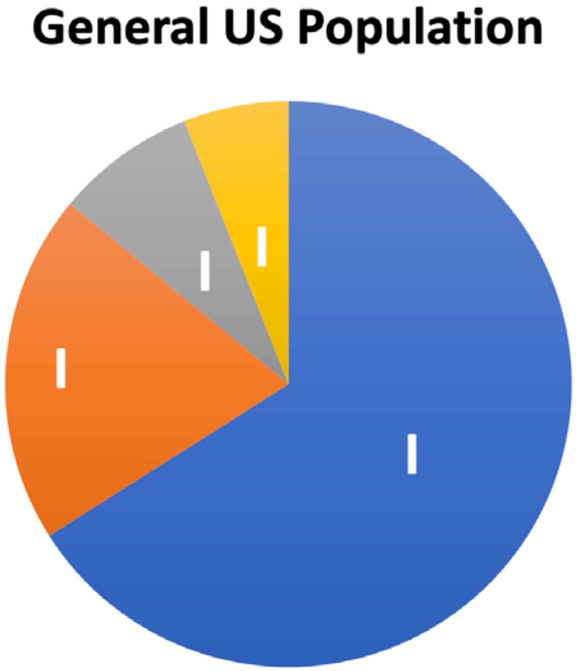


Useless 3D – yes, there's worse

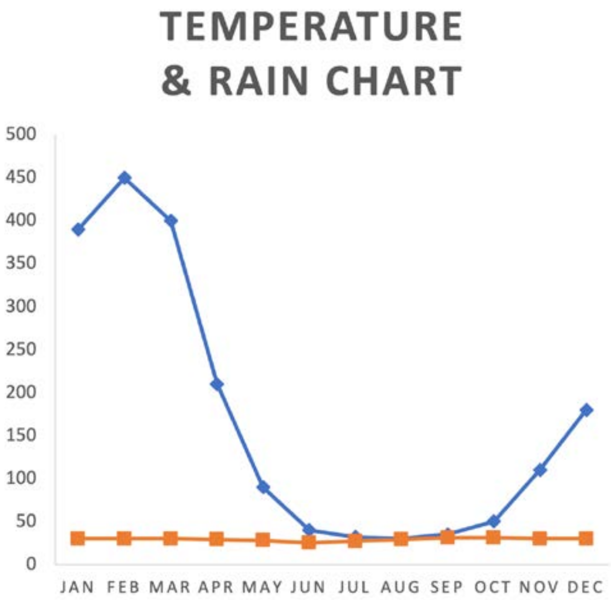


Lack of Context

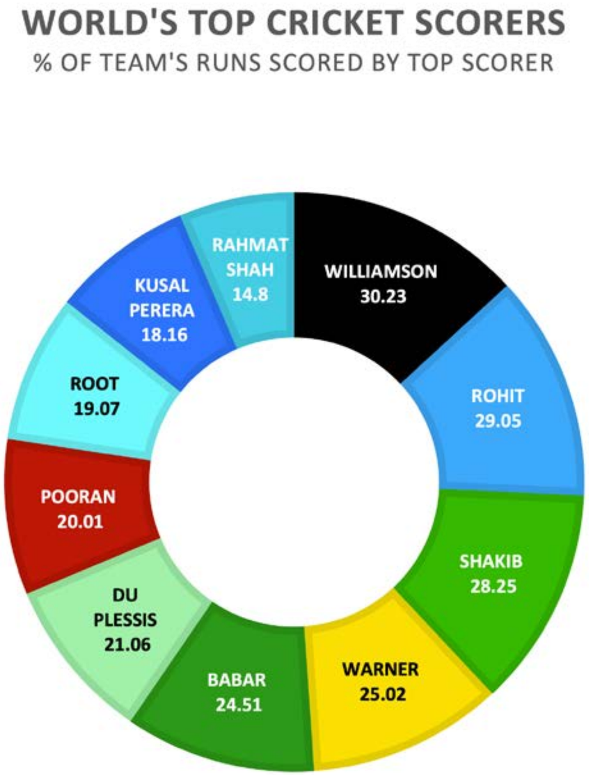
- Missing/useless labels
- Missing axis ticks
- Missing legend
- Missing caption



Inappropriate Representation

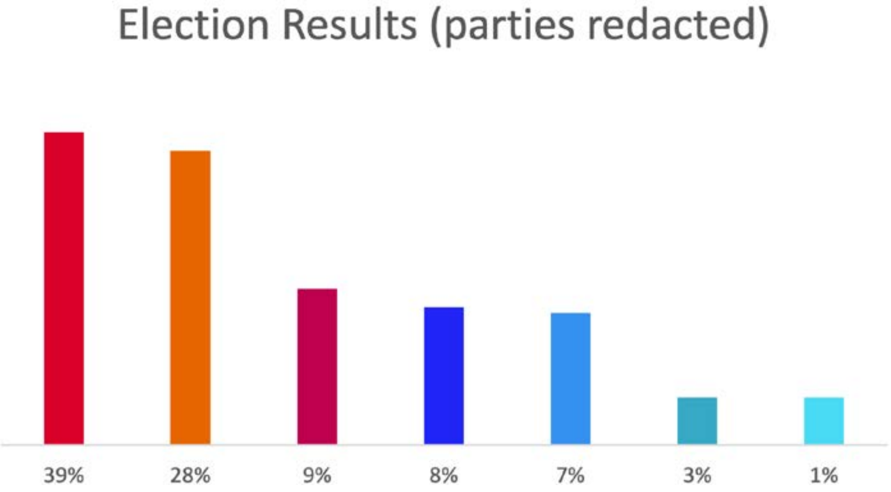


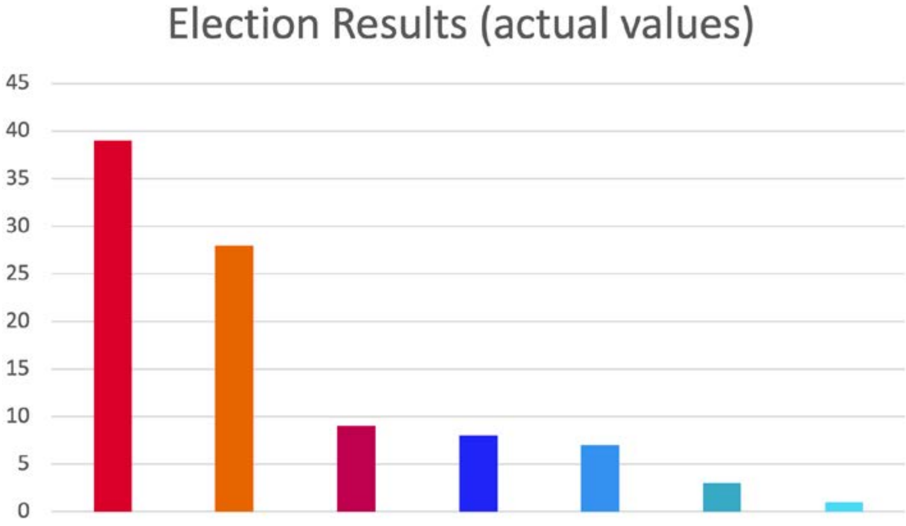
Can you tell the temperature?



Numbers add up to more than 100%!

Misleading Information





- Axis offset
- Wrong scale

- Choice of projection
- Choice of colormap

Manual Labor

- No reproducibility
- Error-prone
  - What settings did you use?
  - Where did you store the data?
  - Which of the 5 versions is it, really?
- What if your data changed? Or you have new data?
  - “Oops, there was an error in my spreadsheet!”
  - “This chart is great! Can you make one for each of the 300 intersections?”
  - “What exact zoom level did I use for my screenshot?”

What not to do-- Recap

- Information Overload
- Inappropriate Color Choice
- Useless 3D
- Lack of Context
- Inappropriate Data Representation
- Misleading Information
- Manual Labor

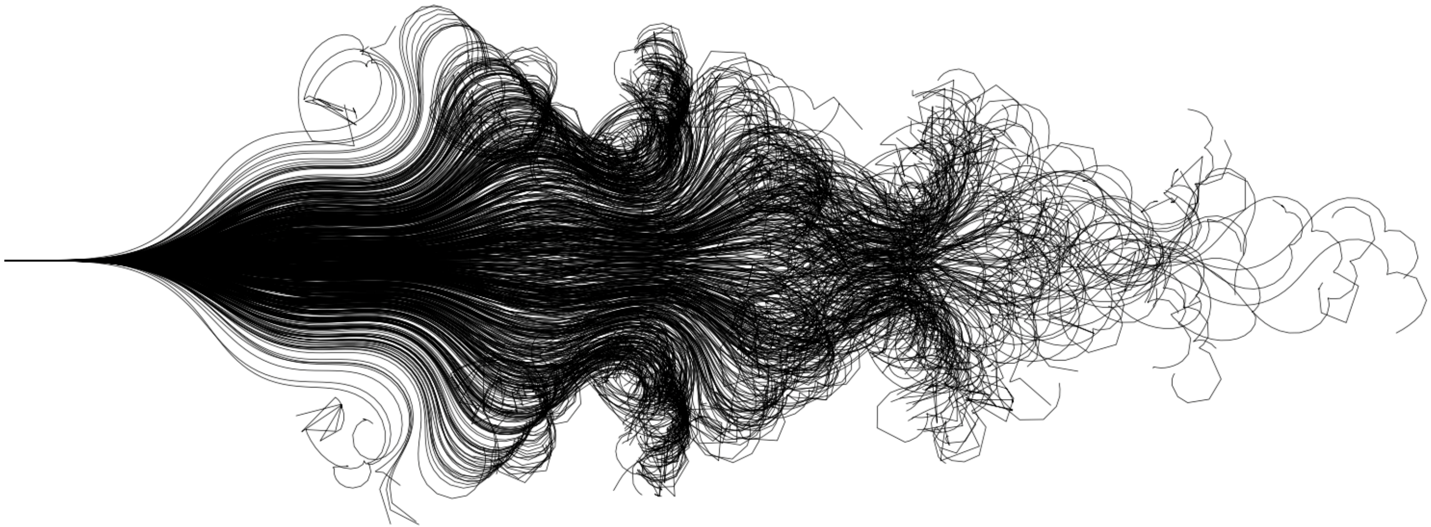
What to do instead

- Know Your Audience!
- Tell a Story
- Human Factor
- Follow the Rules
- Break the Rules

Know Your Audience

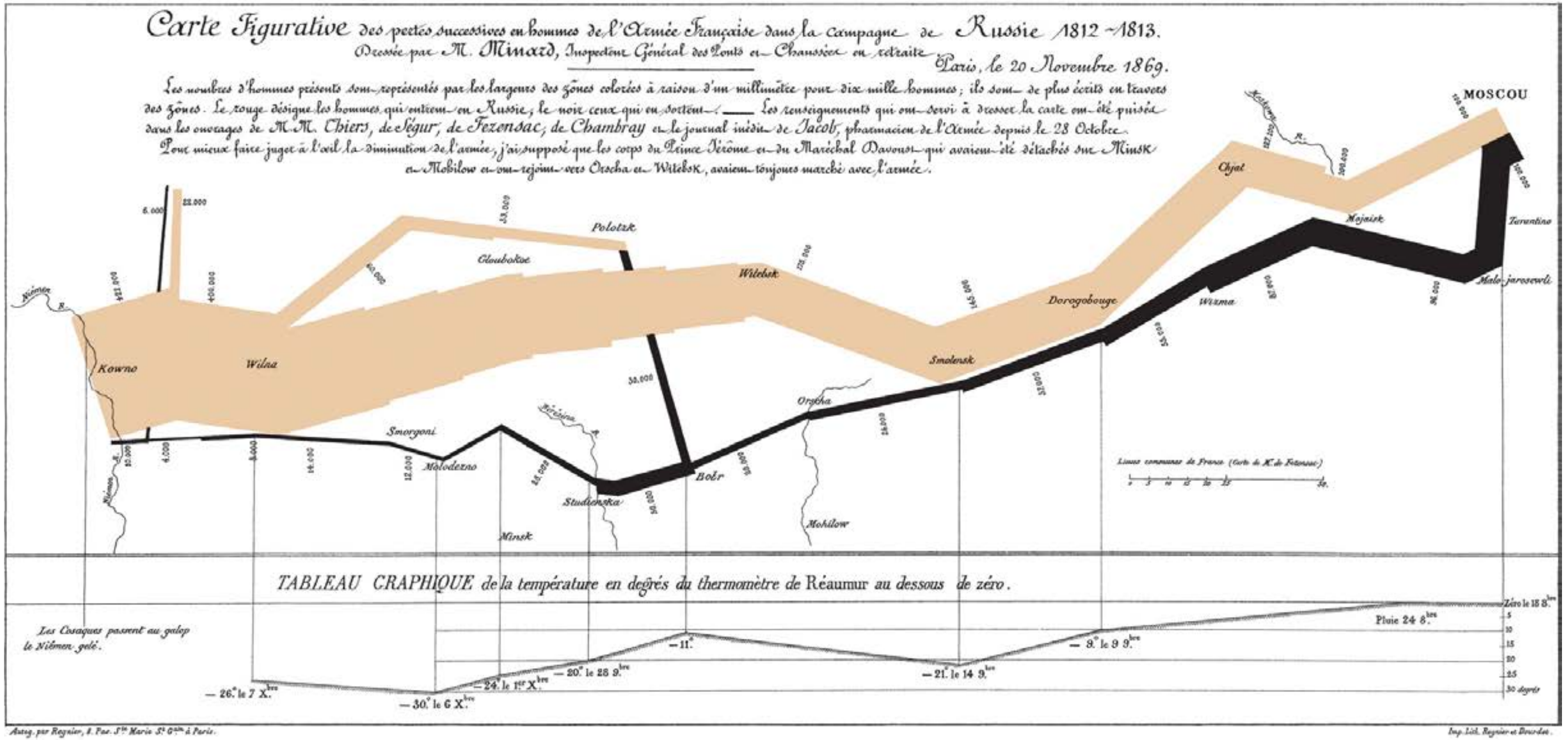
- Objective of the visualization
  - High-level overview (infographic)
  - Expert user (more options, more complexity)
- Visualization literacy
  - Understand which types of visualization will work for your audience – and which types won’t
  - For a general audience, aim for simplicity and known concepts
  - For an expert audience (daily interaction), your visualization can be more complex
- Presentation format
  - Paper vs slides vs interactive

Tell a Story



Cook, Matthew. "It takes two neurons to ride a bicycle." *Demonstration at NIPS* 4 (2004).

Tell a Story



Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. —— Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

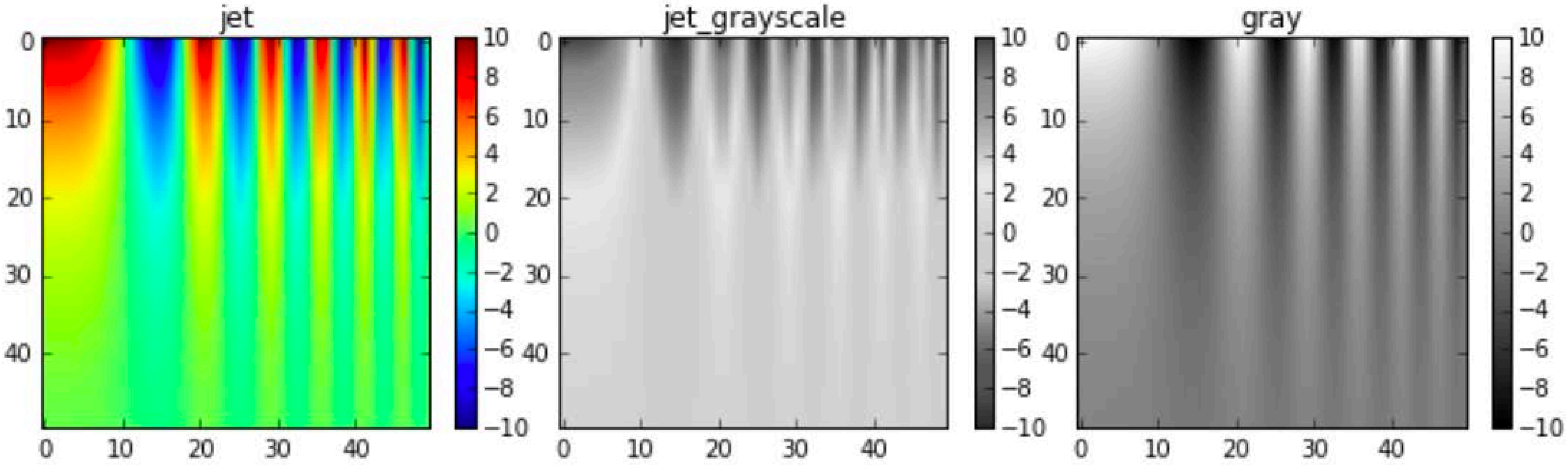
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.

Charles Joseph Minard (1869): Napoleon’s failed attempt at conquering Russia

Human Factor: Perception

- Not all colormaps are perceptually smooth
- To check how your colormap is doing, convert to grayscale and compare with a grayscale map

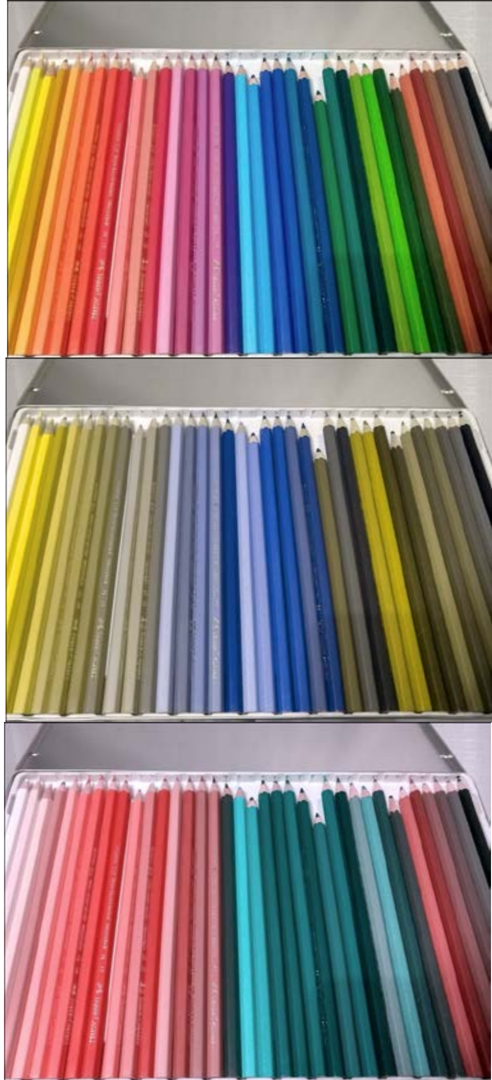


https://jakevdp.github.io/blog/2014/10/16/how-bad-is-your-colormap/

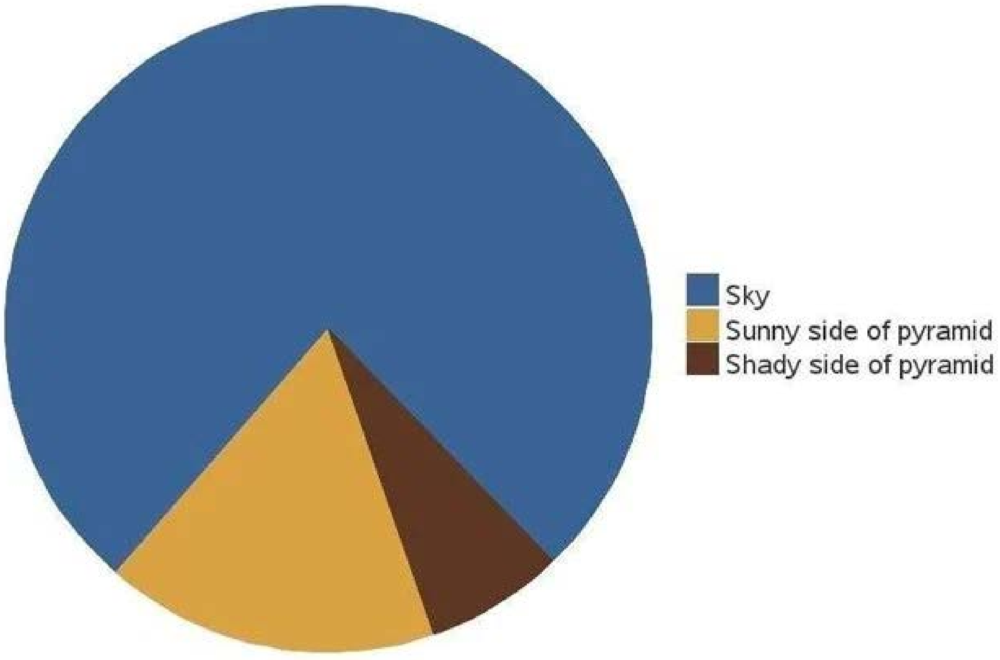
Human Factor: Colorblindness

- Colorblindness is fairly common
  - 8% in people with one X chromosome
  - 0.5% in people with two X chromosomes
- Many colormaps don’t consider this!
  - Choose one that was designed with colorblind people in mind
  - Test your visualization:

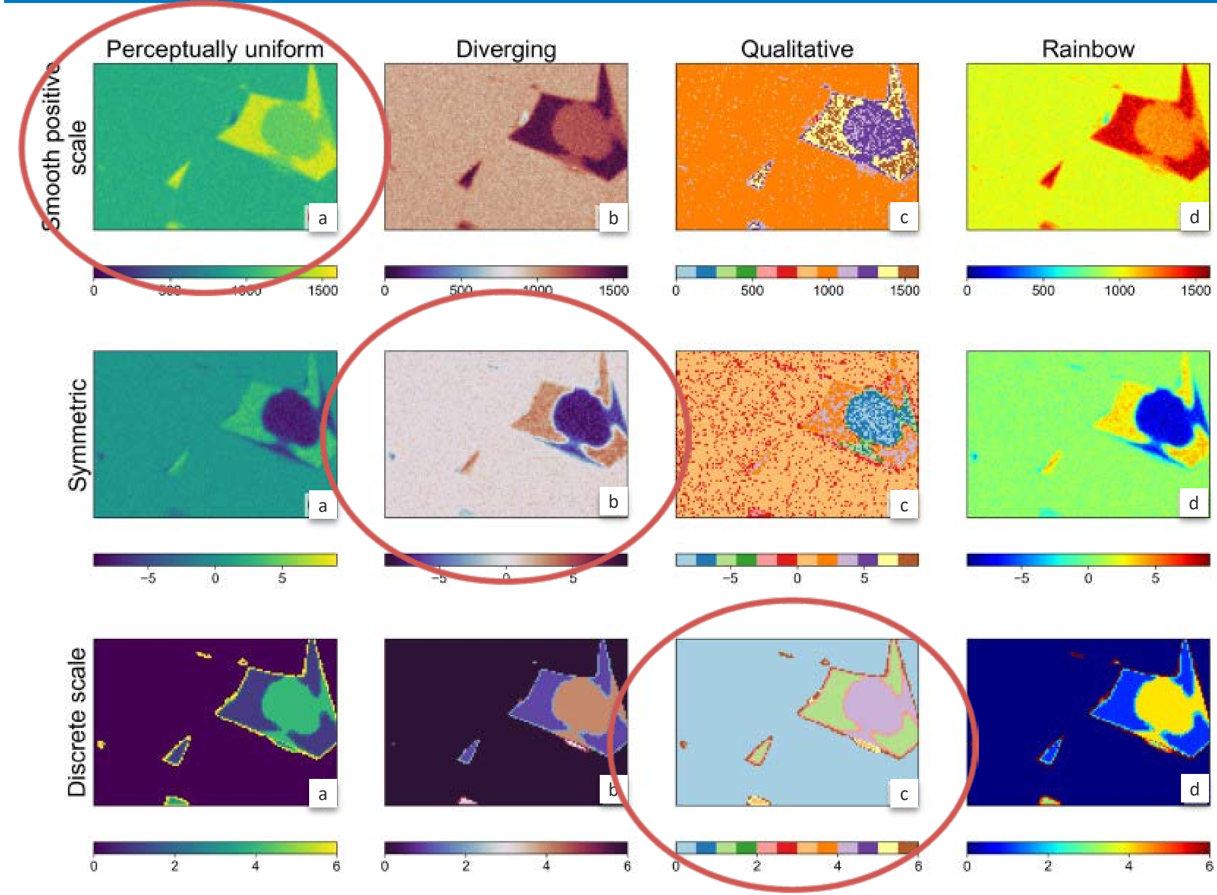
    https://www.color-blindness.com/coblis-color-blindness-simulator/



Meaningful Colors



Meaningful Colors: Colormaps That Make Sense



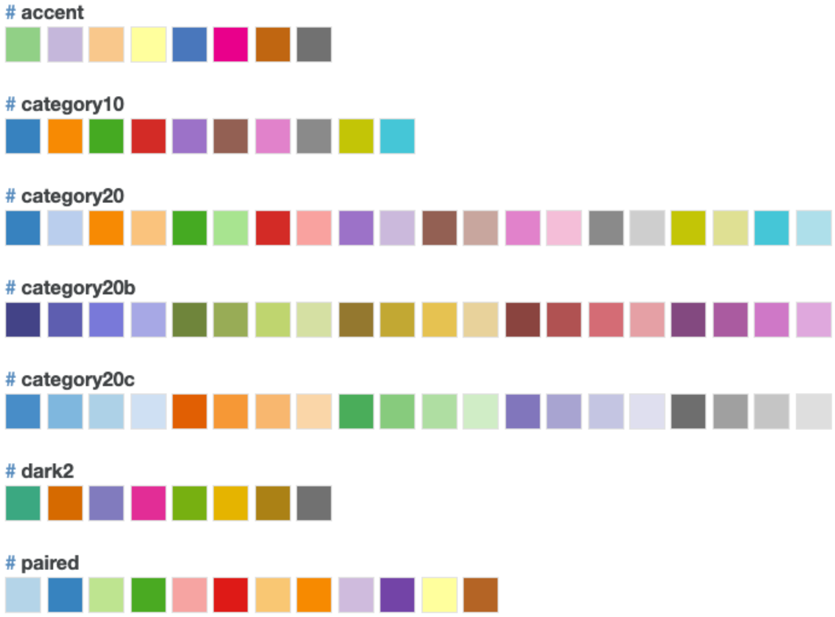
What's the best colormap?

Parish, Chad M., and Philip D. Edmondson. "Data visualization heuristics for the physical sciences." *Materials & Design* 179 (2019): 107868.

Meaningful Colors: Colormap Choice

***Categorical* or *discrete* colormap**

- Your data consists of distinct categories
- If you have groups of similar meaning, there are colormaps for that, too (category20b, category20c)

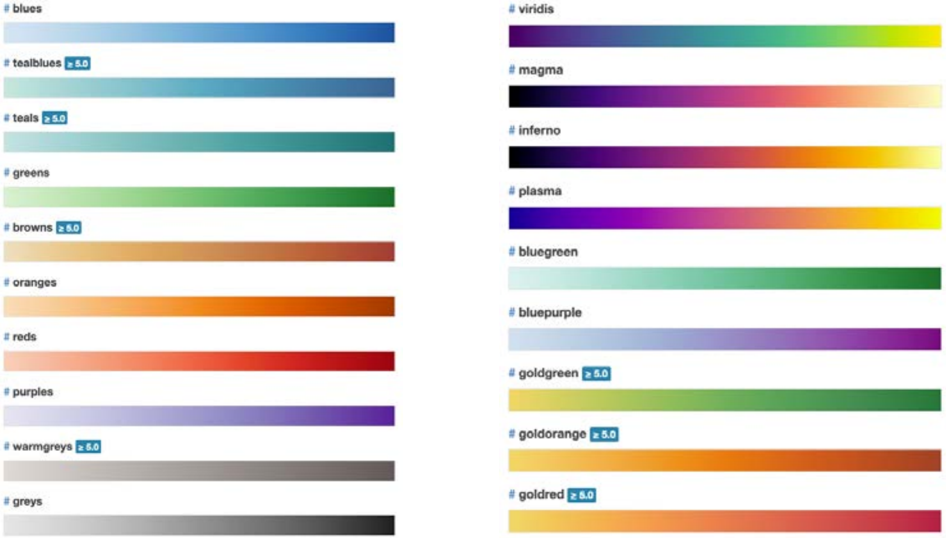


https://vega.github.io/vega/docs/schemes/

Meaningful Colors: Colormap Choice

Sequential **or** ***linear*** **colormap**

- Your data has numerical values along one scale
- Your data doesn't have a meaningful midpoint
- Single-hue/multi-hue

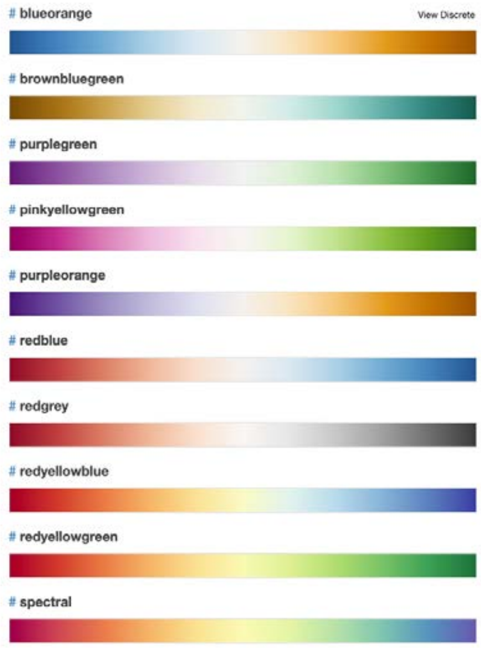


https://vega.github.io/vega/docs/schemes/

Meaningful Colors: Colormap Choice

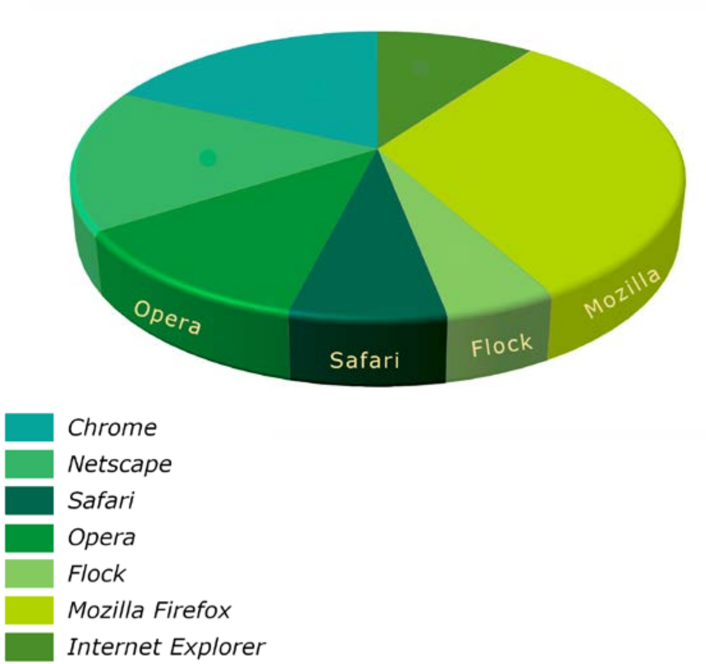
***Divergent* colormap**

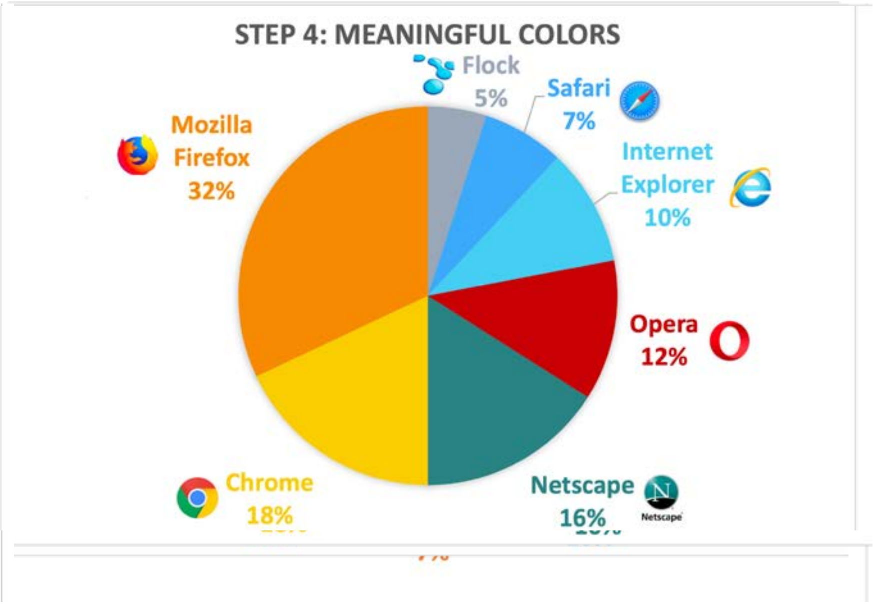
- Quantitative data
- Meaningful mid-point
  - Zero
  - Average value
- *Usually* refers to white in the middle
- *Convergent* colormap *usually* refers to black in the middle
- Sometimes you'll see other colors in the middle



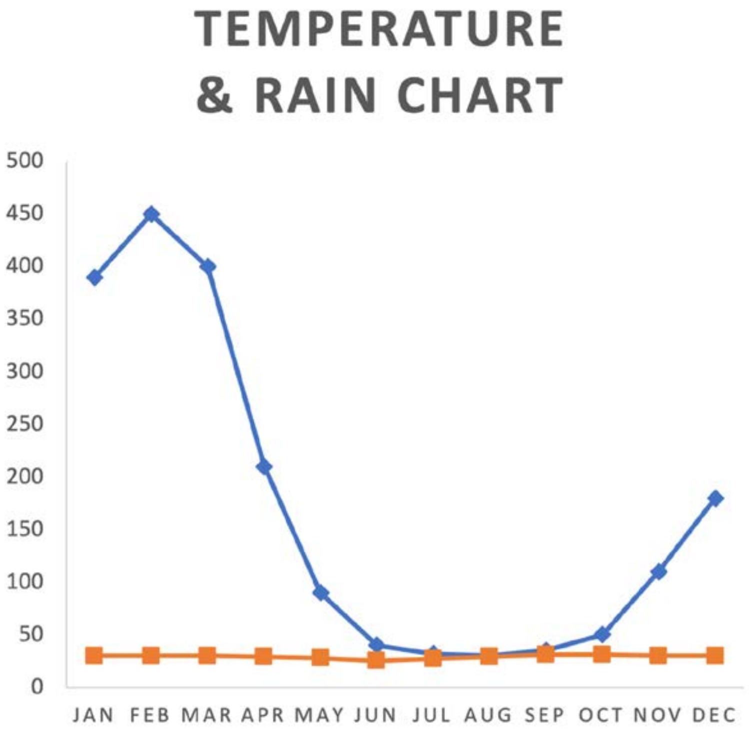
https://vega.github.io/vega/docs/schemes/

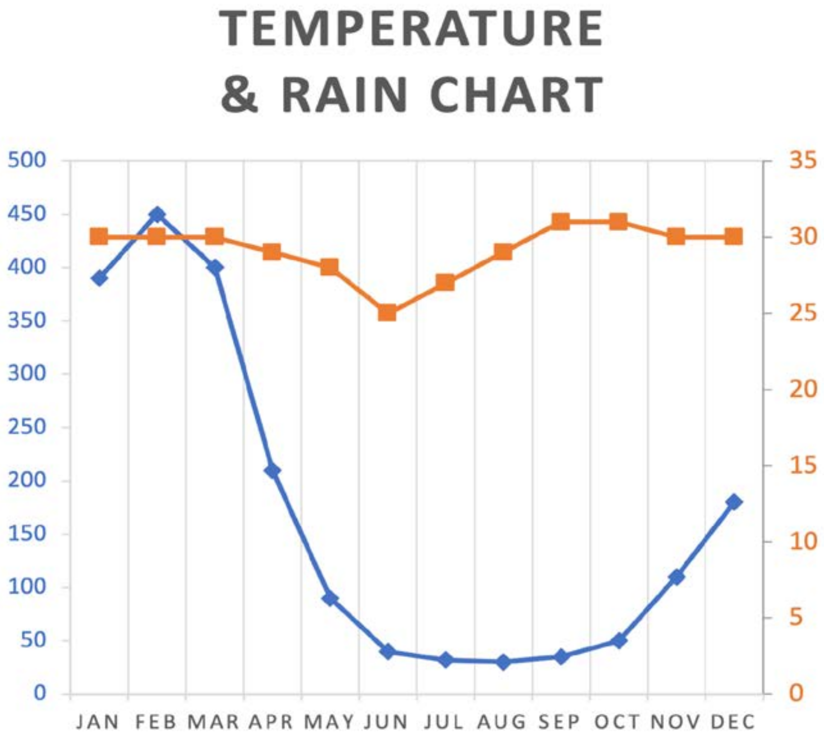
Demo: How to Fix a Bad Chart





Demo: How to Fix a Bad Chart

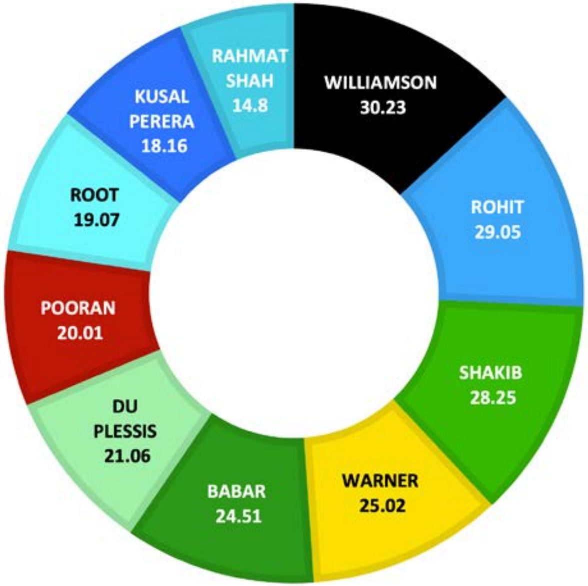


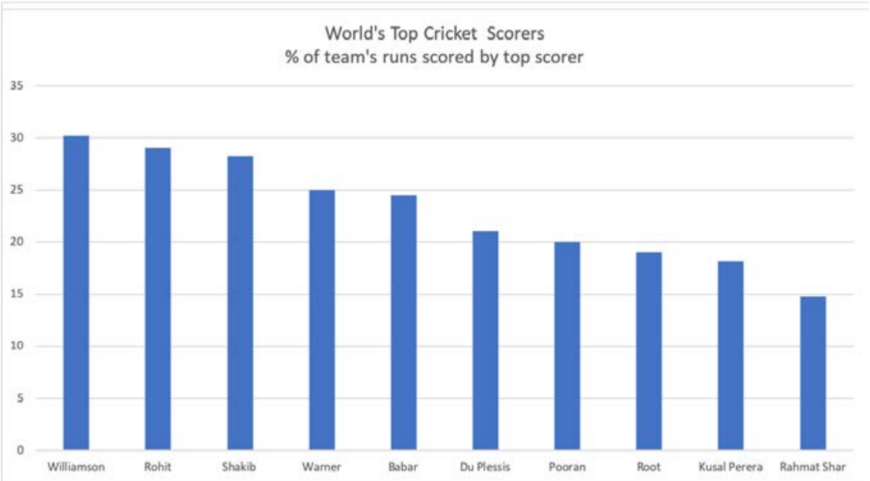


Demo: How to Fix a Bad Chart

WORLD'S TOP CRICKET SCORERS

% OF TEAM'S RUNS SCORED BY TOP SCORER





World's Top Cricket Scorers

% of team's runs scored by top scorer

Key to Better Visualizations

- Faithful representation of numbers
- Appropriate colormap
  - Correct type for task
  - Colorblind-friendly
- Be consistent
- Focus on data
- Provide context
- Be consistent
- Be minimal

Sometimes It’s Okay to Break the Rules

- Axis scaling – make it abundantly clear and be consistent
- Domain-specific conventions
    - Habits are hard to break (people love their colorful visualizations)
    - Different colors/representations can be confusing, e.g., traffic lights
    - Incremental improvement is key!
- If your data is 3D, it’s ok to make a 3D visualization!
- Manual labor – if it’s a one-off or illustrative, doing it manually is fine

Hand-on practice

https://tinyurl.com/TapiaVis2023

Resources

Reading

- Parish, Chad M., and Philip D. Edmondson. "Data visualization heuristics for the physical sciences." *Materials & Design* 179 (2019): 107868.
- Shneiderman, Ben, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.
  - Search for “Ben Shneiderman 8 golden principles”
- Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. Cheshire, CT: Graphics press, 2001.
  - Search for “Edward Tufte visualization principles”
- Moreland, Kenneth. "Why we use bad color maps and what you can do about it." *Electronic Imaging* 2016, no. 16 (2016): 1-6.

Colormaps

- Free E-Book on Color Blindness Essentials: http://www.color-blindness.com/2010/02/23/color-blind-essentials/
- Finding appropriate colormaps: https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3
- Many other color and visualization tools: https://sciviscolor.org/tools/

Data Visualization Tools: Python

- Matplotlib
  - Essentially trying to recreate Matlab's plotting
  - Lots of options but steep learning curve
- ggplot
  - Essentially trying to recreate R's plotting
- Bokeh
  - Interactive plots but steep learning curve
- Seaborn
  - Built on top of Matplotlib, uses pandas input
  - Simple to use
  - Caution: changed dramatically between versions
- PySAL
  - Geospatial data, great for maps
- PyLeaflet
  - Python library to include LeafletJS
  - Geospatial data

…and many more!

Data Visualization Tools: JavaScript

- LeafletJS
  - Maps
  - Easy to use
- OpenLayers
  - Maps
  - Medium difficulty
- WebWorldWind
  - 3D globe
  - Medium difficulty
  - Comes with layer switching
- Cesium
  - Fancy-looking 3D visualizations
  - Advanced
- WebGL
  - Fast compute in browser
- D3.js
  - Charts
- Highcharts
  - Charts

…and many more!

Coding Optional

Thank you

www.nrel.gov

Andy.Berres@nrel.gov

NREL/PR-5700-87388

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Photo from iStock-627281636