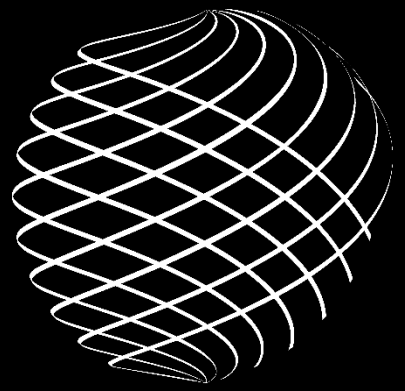


DATA SCIENCE
CENTER

University of Bremen

DataNord

Data Visualization and Communication Using Python and Jupyter

NEWSLETTER

Sign Up!

CONTACT
DSC Data Scientists
anolte@uni-bremen.de

DataNord Data Scientists for Training & Consultation

DataNord

DATA SCIENCE CENTER

RESEARCH

Hub for data-intensive research at U Bremen.

QUALIFI-CATION

Provides targeted training to improve data literacy.

SERVICE

Offers comprehensive “data services” and consultation.

Interdisciplinary collaboration

M.Sc. SARAH BÜKER (Marine Environmental Sciences)

M.Sc. ANNIKA NOLTE (Environmental Sciences)

Dr. MARYAM MOVAHEDIFAR (Statistics)

Dr. SUSANNE DE VOGEL (Sociology)

M.A. NELE FUCHS (Transcultural Studies, Philosophy)

Dipl.-Soz. HEIKE THÖRICHT (Sociology)

- Marine & Environmental Sciences
- Health Sciences
- Humanities
- Material Sciences & Engineering
- Social Sciences
- All disciplines

Outline

1. Motivation and elements behind effective data visualizations
2. “How to Python” in this workshop
3. Quickstart: Python programming (~30 min)
4. Basic plotting workflows and customization options

← Coffee Break (~15 min)

5. Plot types & plotting techniques

← Lunch Break (~1h)

6. Mapping geospatial data
7. Animations

← Coffee Break (~15 min)

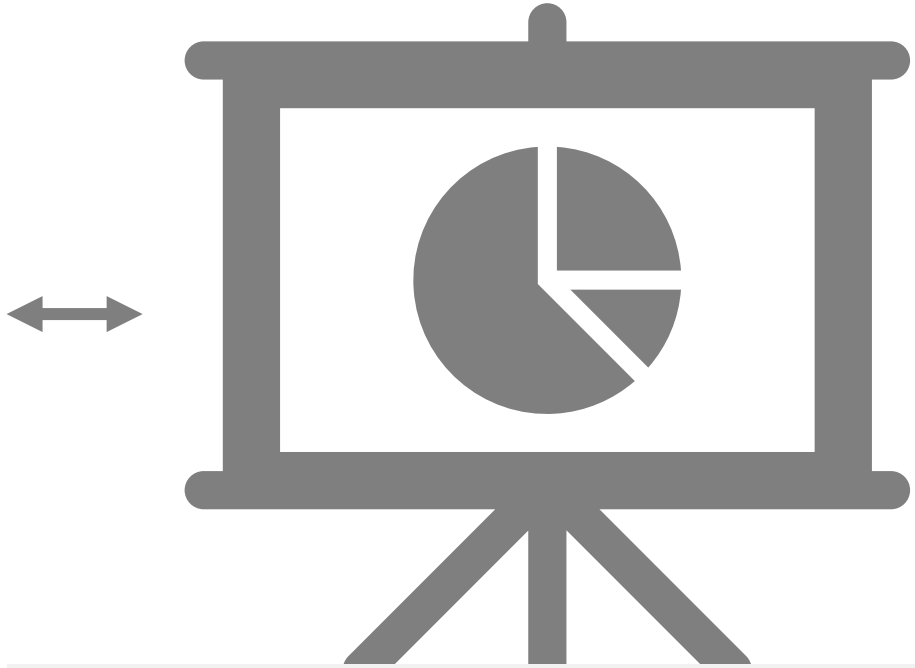
8. Interactive visualizations
9. Wrap-up

Goals

- ✓ Understand the role of data visualization in **storytelling**
- ✓ Recall the **key principles and best practices** that contribute to impactful data visualizations
- ✓ Acquire **skills for advanced data visualization** tailored to various data types and analytical objectives in the Earth Sciences using essential Python libraries
- ✓ Learn how to bring your data visualization **ideas to life**
- ✓ Understand the **interplay between data, plotting methods, and the messages** conveyed by visuals
- ✓ Promote **reproducibility and shareability** in data analysis
- ✓ (Getting started with Python via data visualization tasks – personal experience: Learning Python with visuals is fun!)
- × In-depth theoretical analysis of specific data types for visualization, such as time series, geospatial, or image data.

Data Literacy & Visualizations

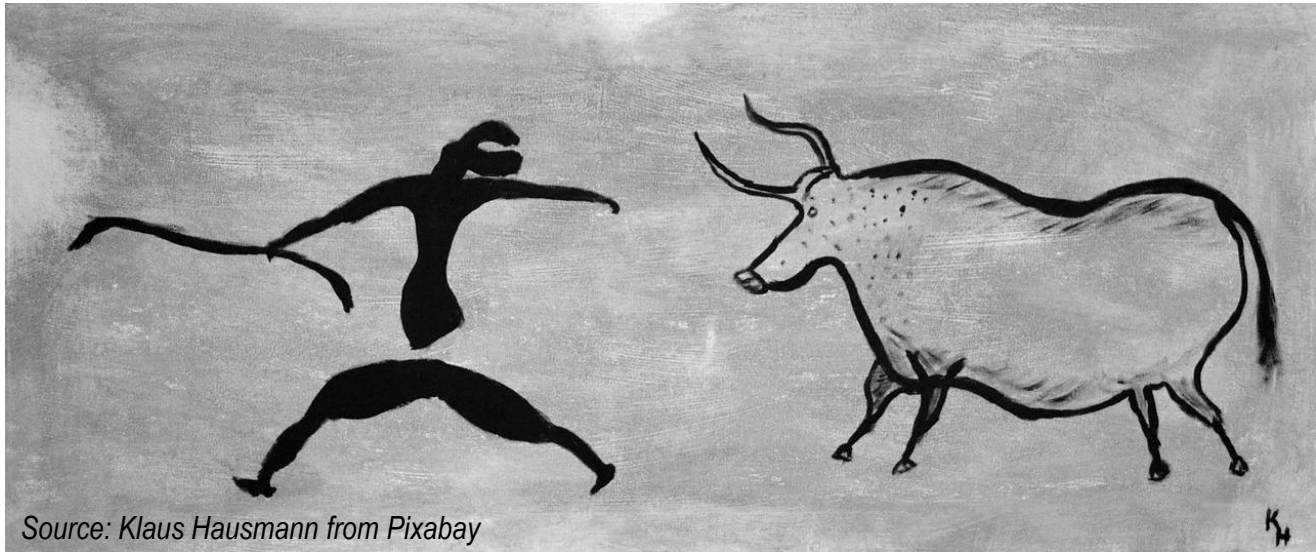
Overview of the Data Lifecycle. Adapted from RDMkit



Data visualization is the representation and presentation of information and data to make it more accessible, understandable, and actionable.

Storytelling

- … is the most natural form of **education**
- … help us **to make sense of things**, e.g. developments in time; human connections
- … has changed with the **inventions** over the course of history:
  - printed stories in newspapers, magazines, and books
  - motion picture camera made movies possible
  - TV and streaming
  - data visualization tools

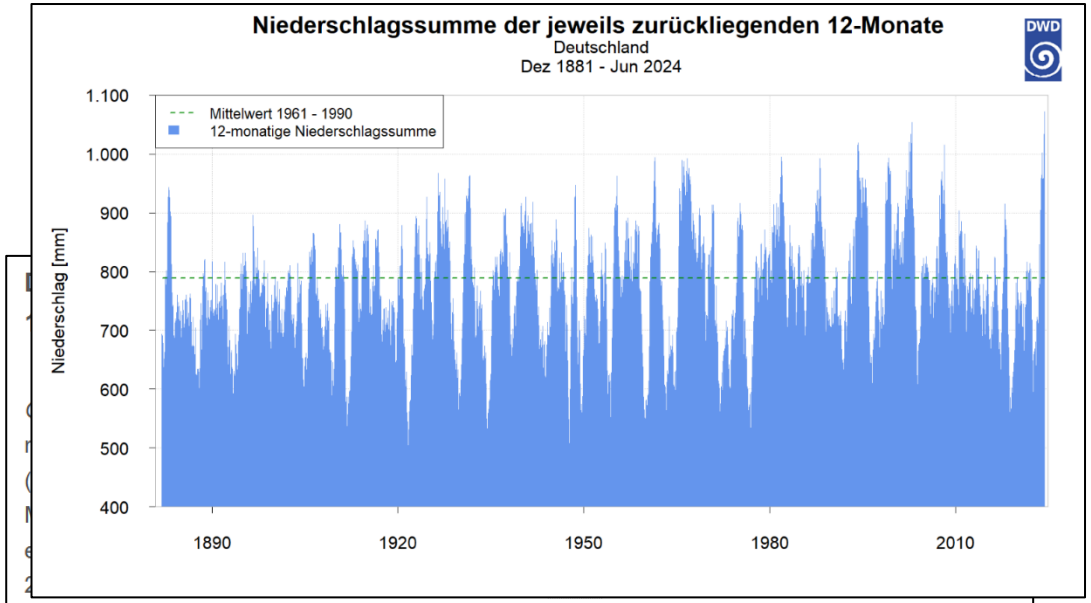


Source: Klaus Hausmann from Pixabay

Data storytelling is communicating the meaning of a dataset with visuals and a story that are customized for each particular audience.

Motivation for Data Visualization

- Quickly and clearly **understand** scientific data:
  - reducing confusion, mistakes and time-effort
  - uncover trends, patterns, relationships, and outliers.



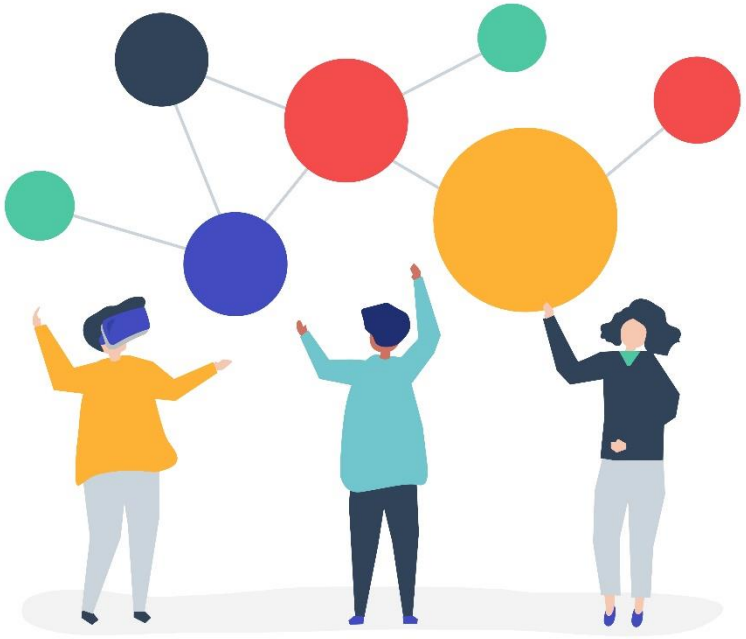
Litern pro Quadratmeter. Im Vergleich dazu beträgt der vieljährige Mittelwert der Referenzperiode 1961-1990 rund 789 Liter pro Quadratmeter im Jahr.

Ein Blick ins nationale Klimaarchiv des DWD zeigt: Während der vergangenen zehn Jahre wurden überwiegend eher unterdurchschnittliche jährliche Niederschlagssummen beobachtet. (Abb. 2) Betrachtet man die Zeitreihe seit Messbeginn 1881 zeigen die Daten aber eine leichte Zunahme der jährlichen Niederschlagsmengen in Deutschland. Allerdings kommt es dabei laut DWD immer wieder zu einem Wechsel zwischen trockenen und feuchten Perioden. Kaspar: "Der Niederschlag zeichnet sich durch eine hohe Variabilität sowohl von Jahr zu Jahr als auch über längere Zeiträume hinweg aus." (Abb. 3) Die Trockenheit der vergangenen Jahre wurde nun durch eine sehr feuchte 12-monatige Phase abgelöst, die das Niederschlagsdefizit mit jedem Monat weiter reduziert hat.

Source: https://www.dwd.de/DE/presse/pressemitteilungen/DE/2024/20240703_die-zwoelf-nassesten-monaten-seit-messbegin_news.html?nn=16210

Motivation for Data Visualization

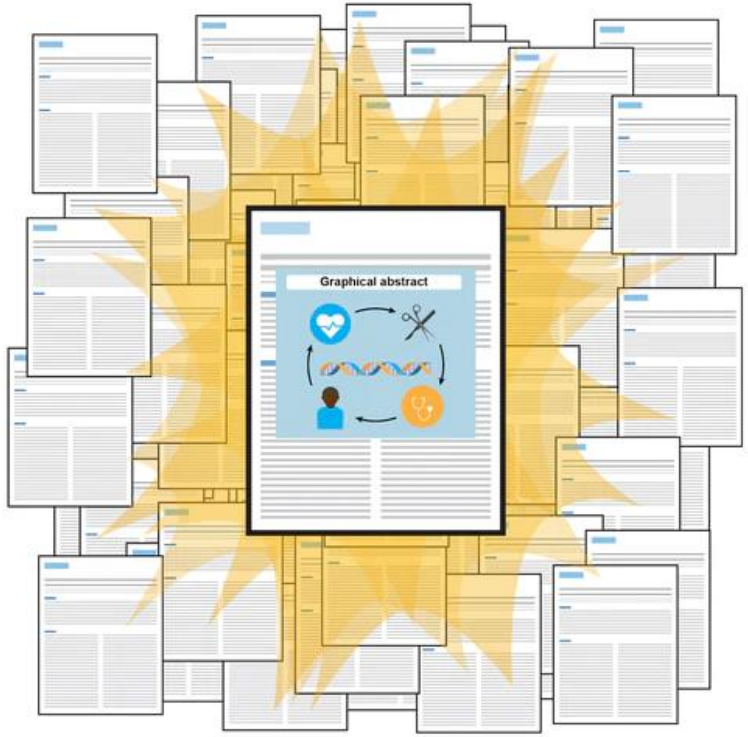
- Quickly and clearly **understand** scientific data:
  - reducing confusion, mistakes and time-effort
  - uncover trends, patterns, relationships, and outliers.
- Boosted **collaboration** by creating a shared narrative that fosters understanding and effective teamwork, especially in interdisciplinary teams – simplification of complex data makes it easier to work together.



Source: rawpixel.com from Freepik

Motivation for Data Visualization

- Quickly and clearly **understand** scientific data:
  - reducing confusion, mistakes and time-effort
  - uncover trends, patterns, relationships, and outliers.
- Boosted **collaboration** by creating a shared narrative that fosters understanding and effective teamwork, especially in interdisciplinary teams.
- Improve grant proposals, streamline the peer review process (e.g., through visual abstracts), elevate presentations, and improve social media **outreach**.



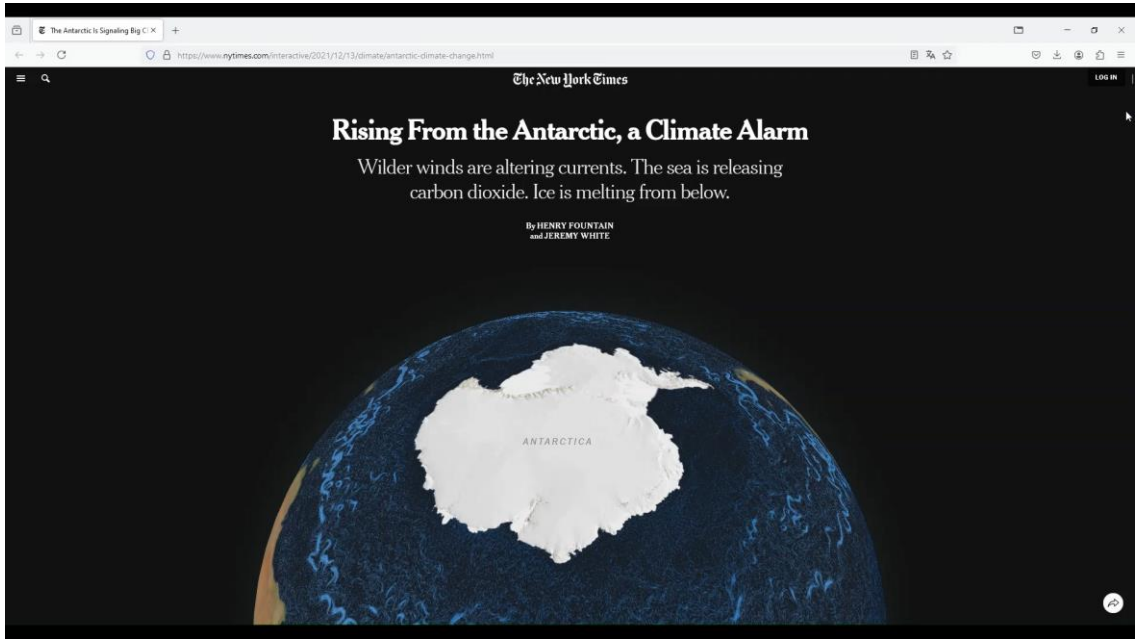
Source: https://solutions.springernature.com/blogs/visibility/what-is-a-graphical-abstract-and-why-do-i-need-one-for-my-paper

A visual abstract is shared 8 times more often on social media

People remember visuals 6 times better than text

Motivation for Data Visualization

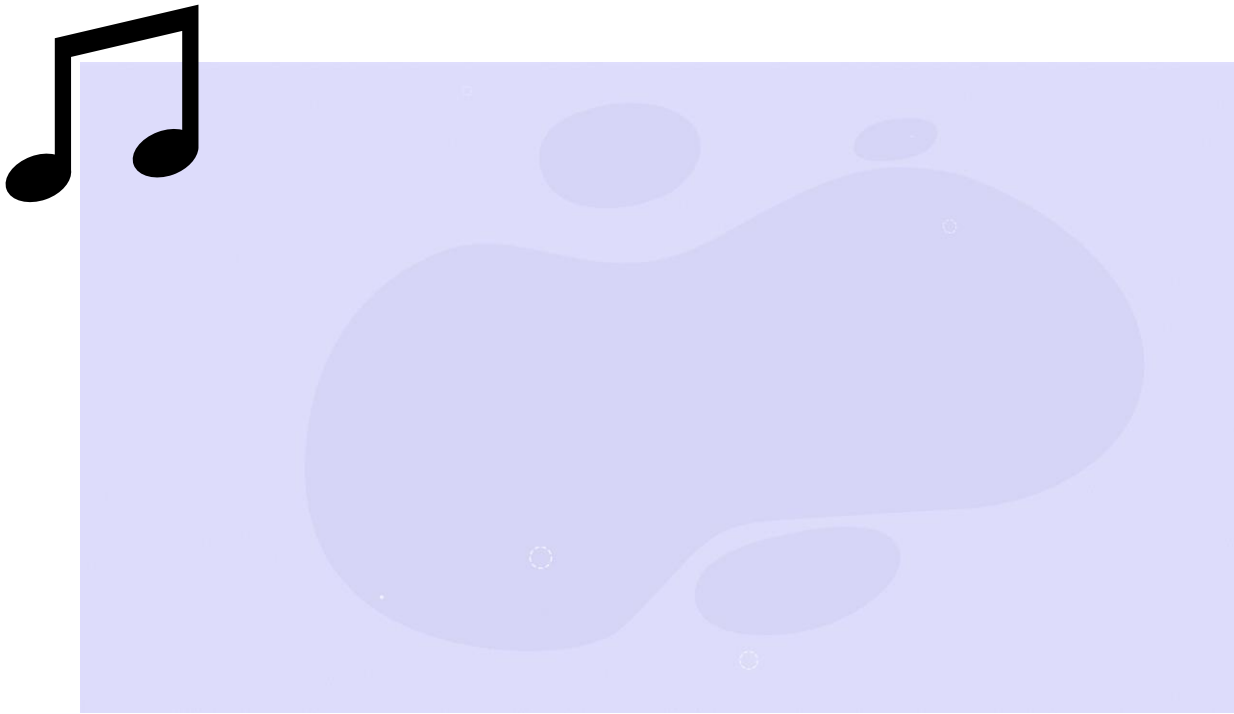
- Quickly and clearly **understand** scientific data:
  - reducing confusion, mistakes and time-effort
  - uncover trends, patterns, relationships, and outliers.
- Boosted **collaboration** by creating a shared narrative that fosters understanding and effective teamwork, especially in interdisciplinary teams.
- Improve grant proposals, streamline the peer review process (e.g., through visual abstracts), elevate presentations, and improve social media **outreach**.
- Making scientific information **more engaging and accessible** for everyone, including policymakers, students, and the general public.



Source: https://www.nytimes.com/interactive/2021/12/13/climate/antarctic-climate-change.html

Data Storytelling in the Music Streaming Industry

- Because data storytelling is so efficient, industry makes use of it a lot
- Example from the music streaming industry: Annual reviews of the customers listening behavior with animated visuals
  - Identifies and highlights: Songs, artists, genres, etc.
  - Concratulations for being a top listener
  - How much time spend enjoying the service

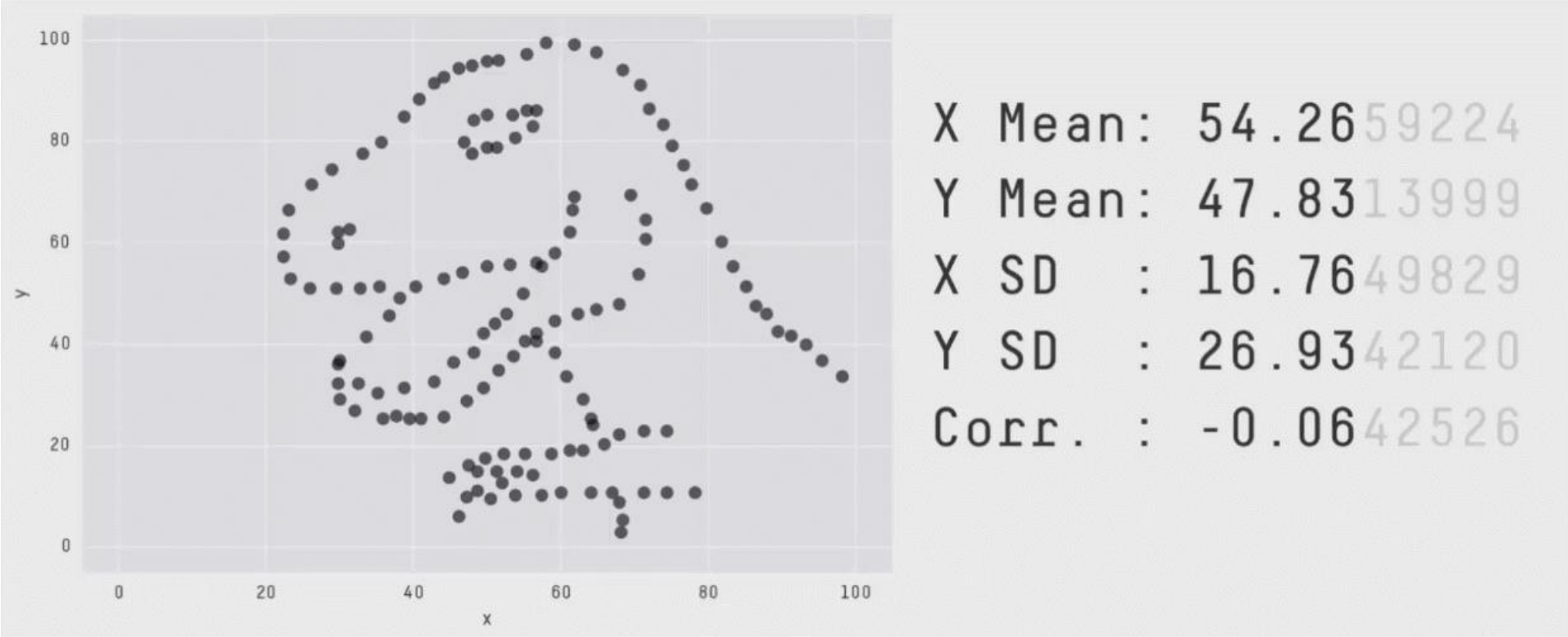


Source: Freepik (www.freepik.com)

Anscombe’s Quartet: Same statistics – different visual patterns

University of Bremen

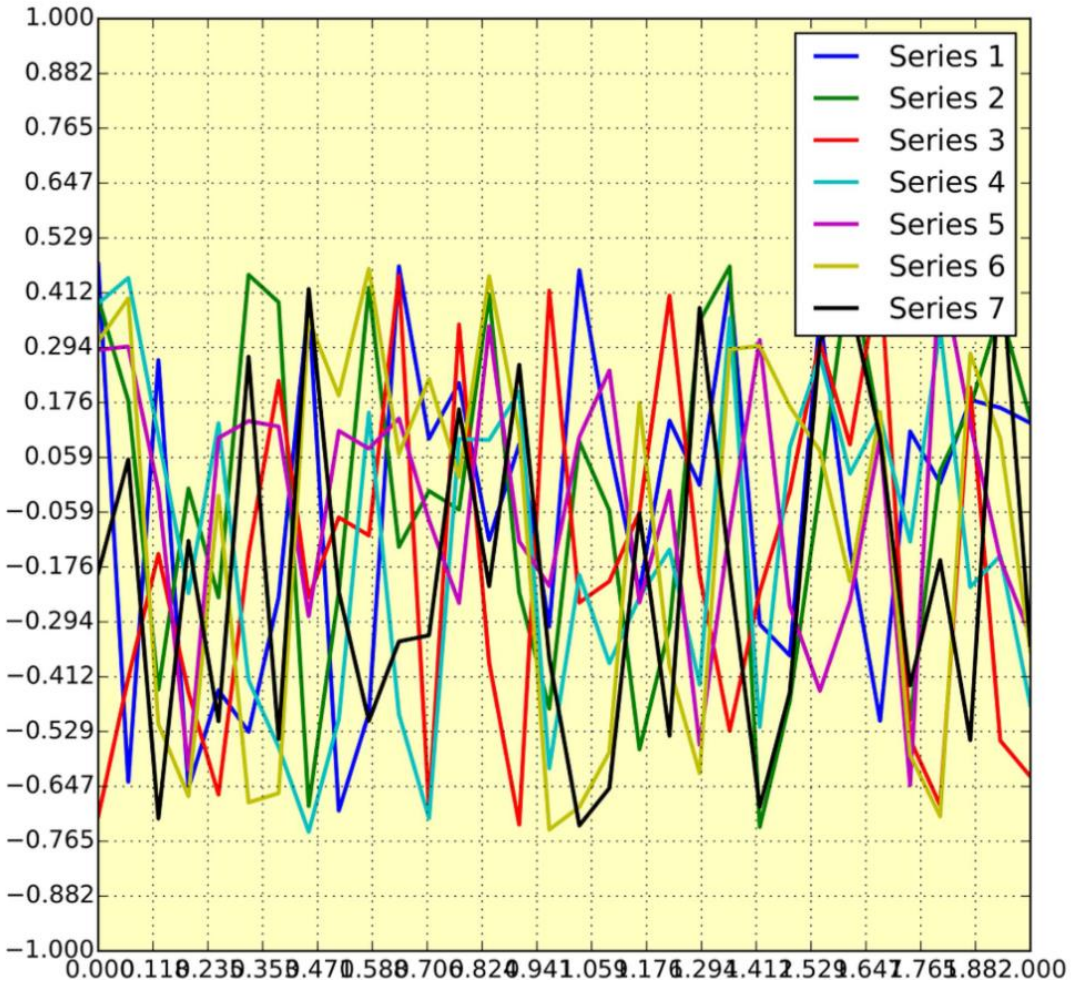
Never trust summary statistics alone. Always visualize your data to tell the whole story.

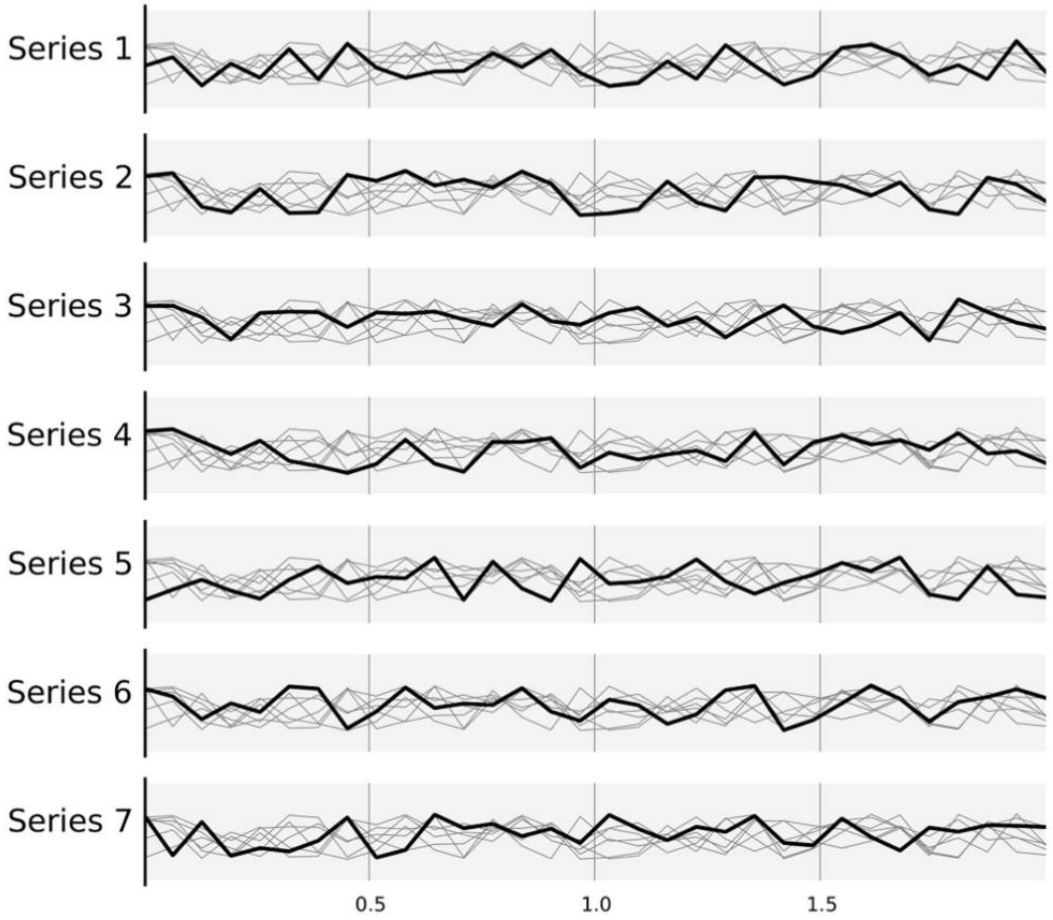


Sources: https://www.research.autodesk.com/publications/same-stats-different-graphs/; Albert Cairo (http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html)

Some Worst and Best Practices

“Avoid chartjunk”

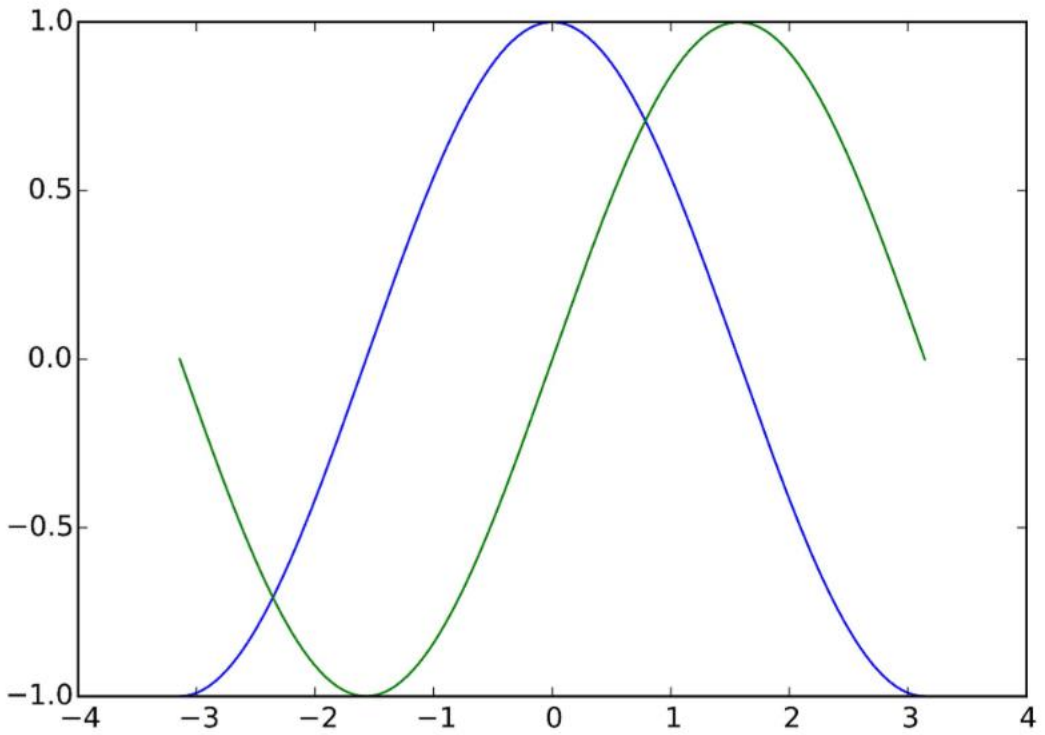


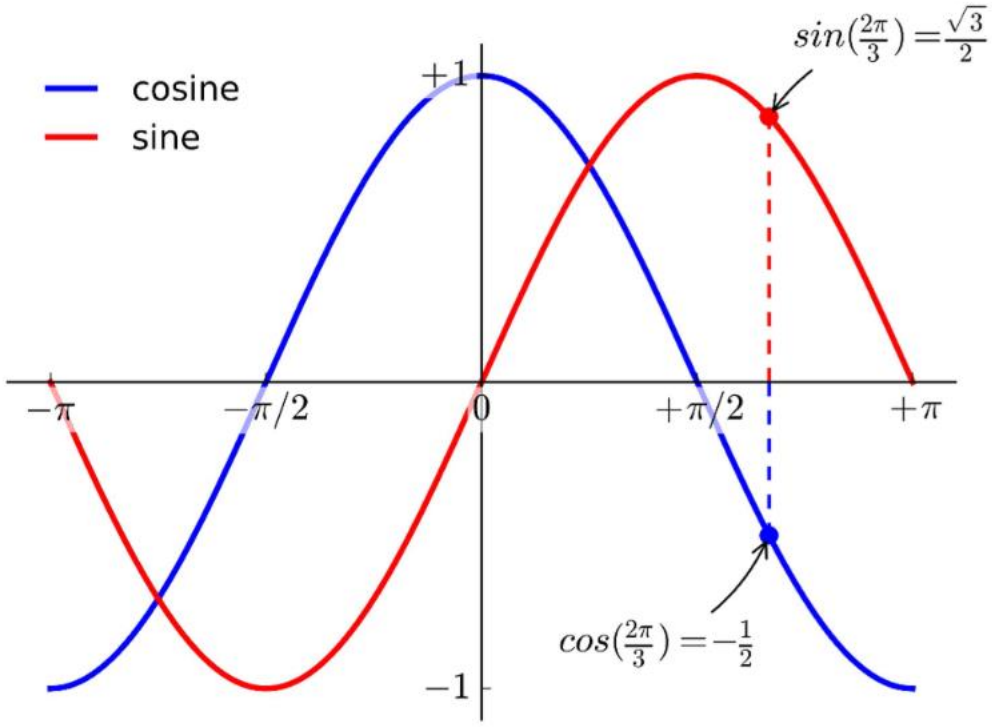


Source: Rougier, Nicolas P.; Droettboom, Michael; Bourne, Philip E. (2014): Ten simple rules for better figures: Public Library of Science (10) (9).

Some Worst and Best Practices

“Do go beyond the defaults”

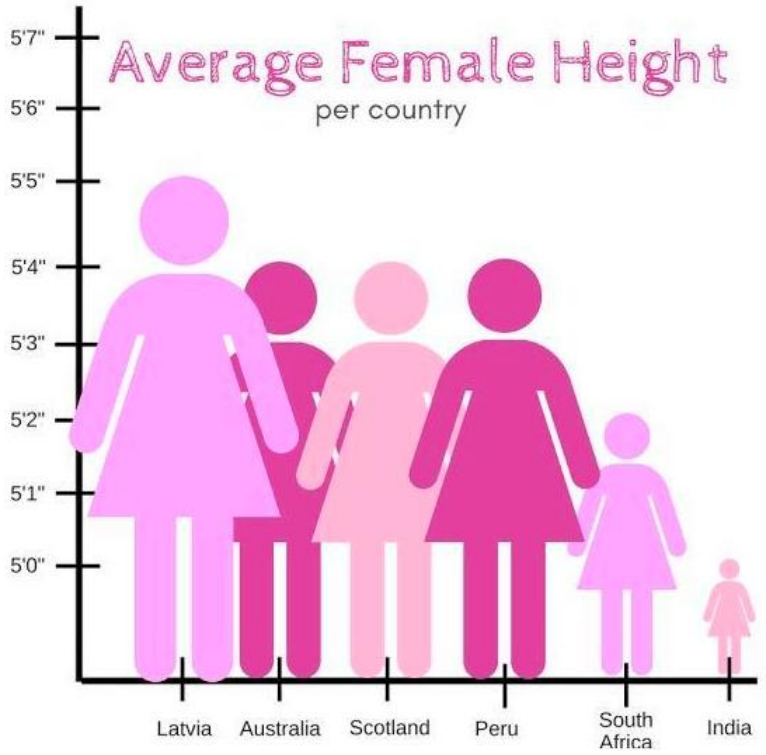




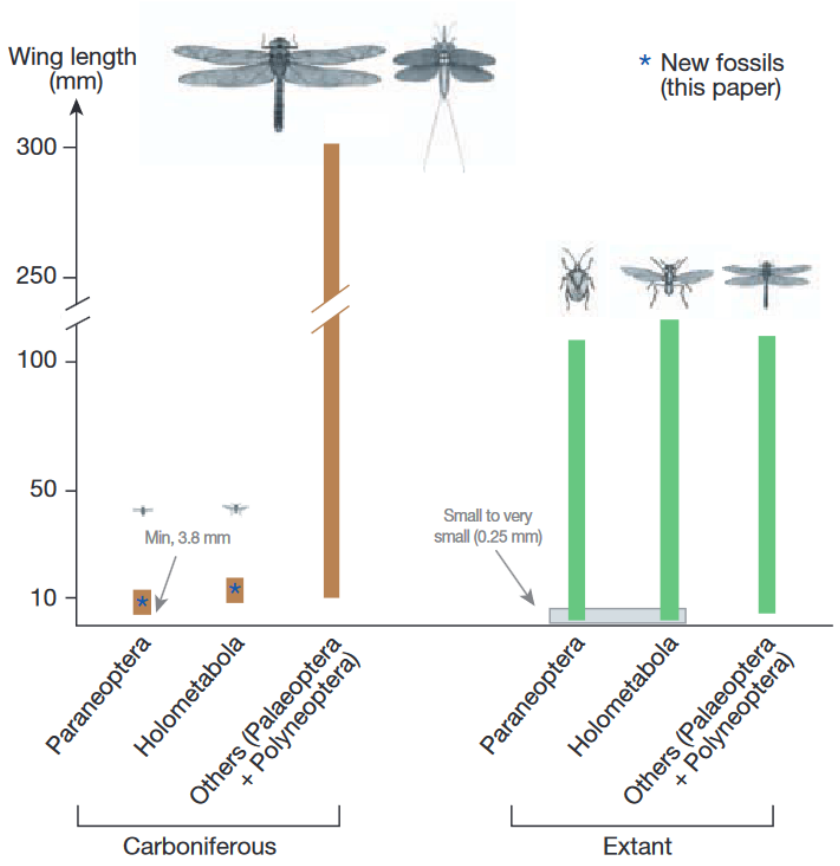
Source: Rougier, Nicolas P.; Droettboom, Michael; Bourne, Philip E. (2014): Ten simple rules for better figures: Public Library of Science (10) (9).

Some Worst and Best Practices

“Stick to universal rules”



Source: https://ninjatables.com/bad-data-visualization-examples/

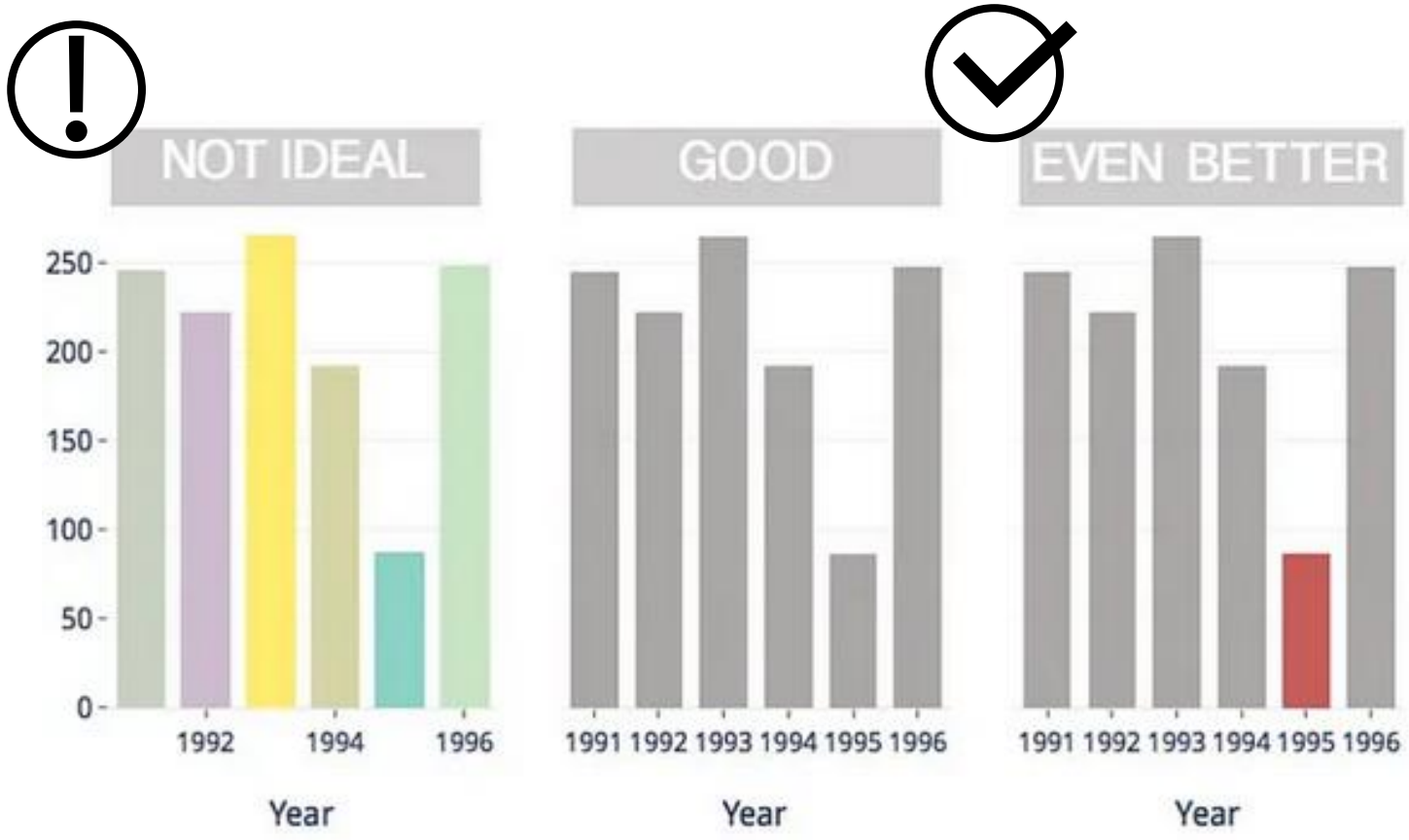


Source: Nel, André; Roques, Patrick; Nel, Patricia; Prokin, Alexander A.; Bourgoin, Thierry; Prokop, Jakub et al. (2013): The earliest known holometabolous insects. In Nature 503 (7475), pp. 257–261.

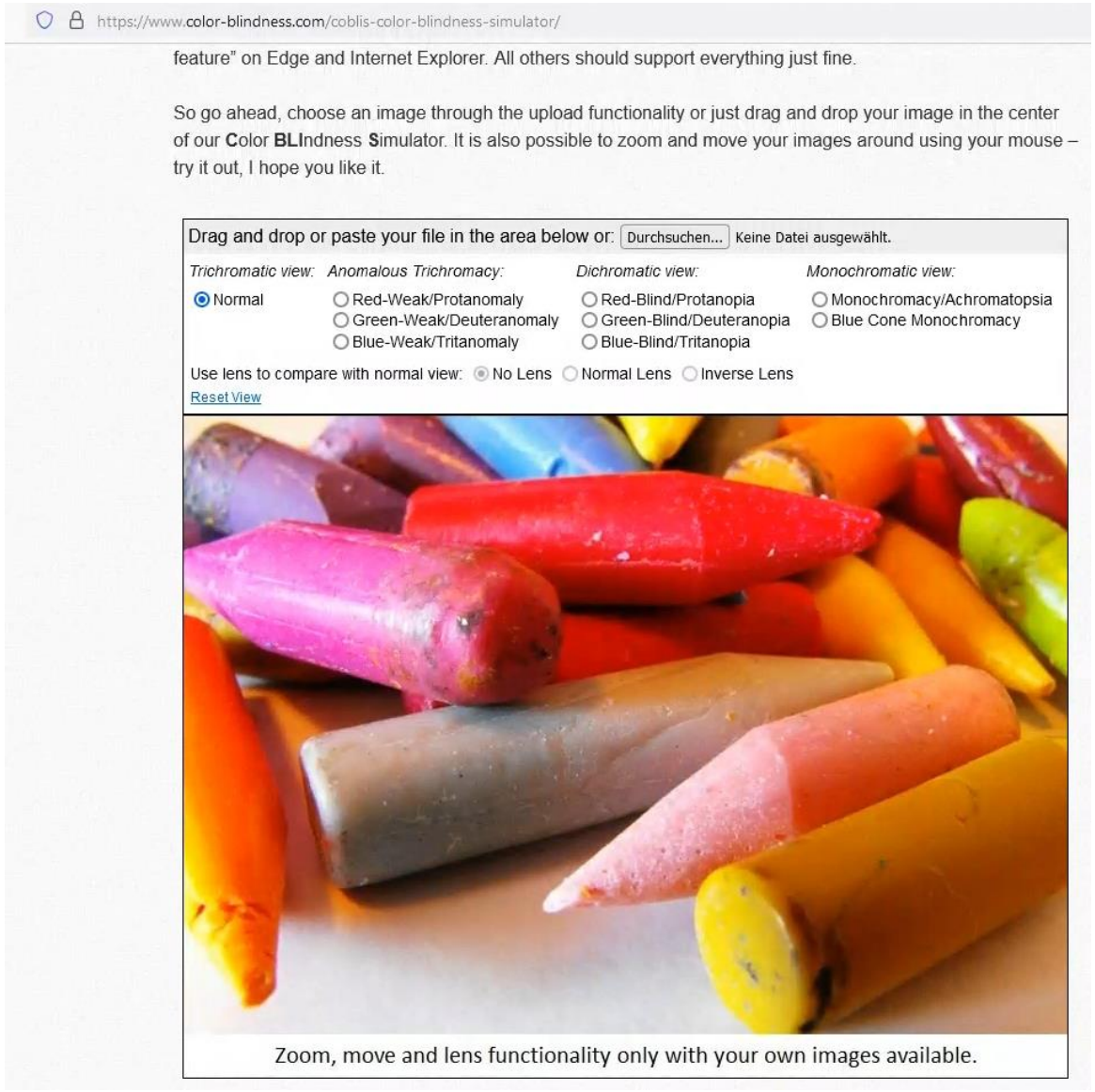
Some Worst and Best Practices

- Color were it makes sense
- Sensitive to cultural contexts
- Consistency across multiple visualizations
- Color as part of the storytelling
- Differentiability and accessibility (e.g. to people with color vision deficiencies)

“Use colors wisely”



Color Blindness Simulator

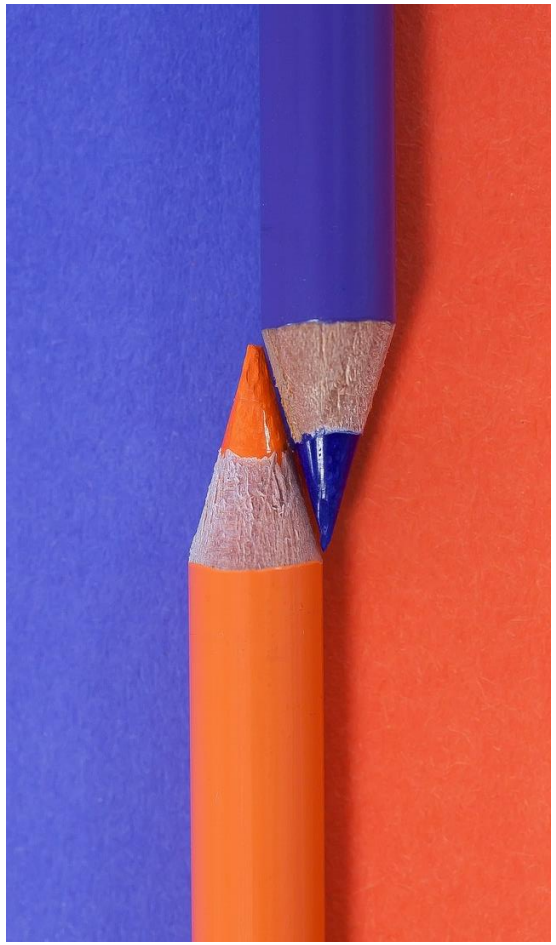


https://www.color-blindness.com/coblis-color-blindness-simulator/

The role of color in data visualizations

University of Bremen

Differentiation



Source: stux from Pixabay

Emphasis

Source: schuetz-mediendesign from Pixabay

Aesthetics

Source: Alexas_Fotos from Pixabay

Emotions

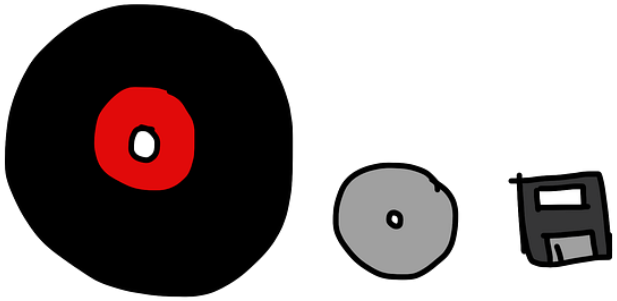
Source: Adam Filipowicz from Shutterstock

Summary: What Makes a Good Data Visualization?

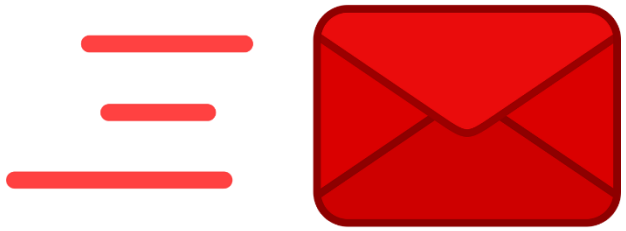
University of Bremen

Best Practice Layout

Medium Adapted



Message Focused



Audience-Centric

The Python Universe for Data Visualization

plotly

Fundamental Plotting Library

matplotlib

seaborn

Fundamental Libraries

pandas

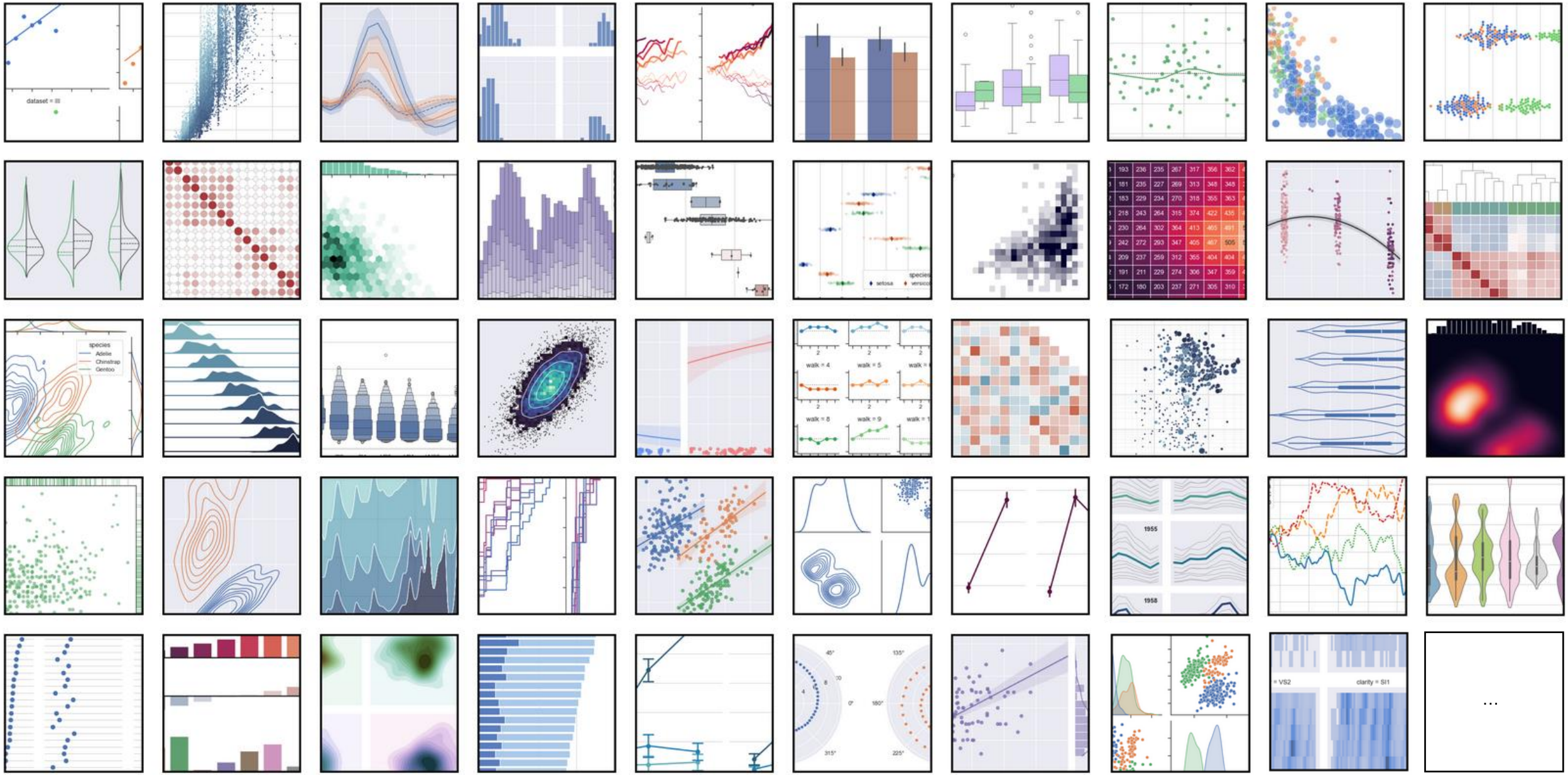
NumPy

...

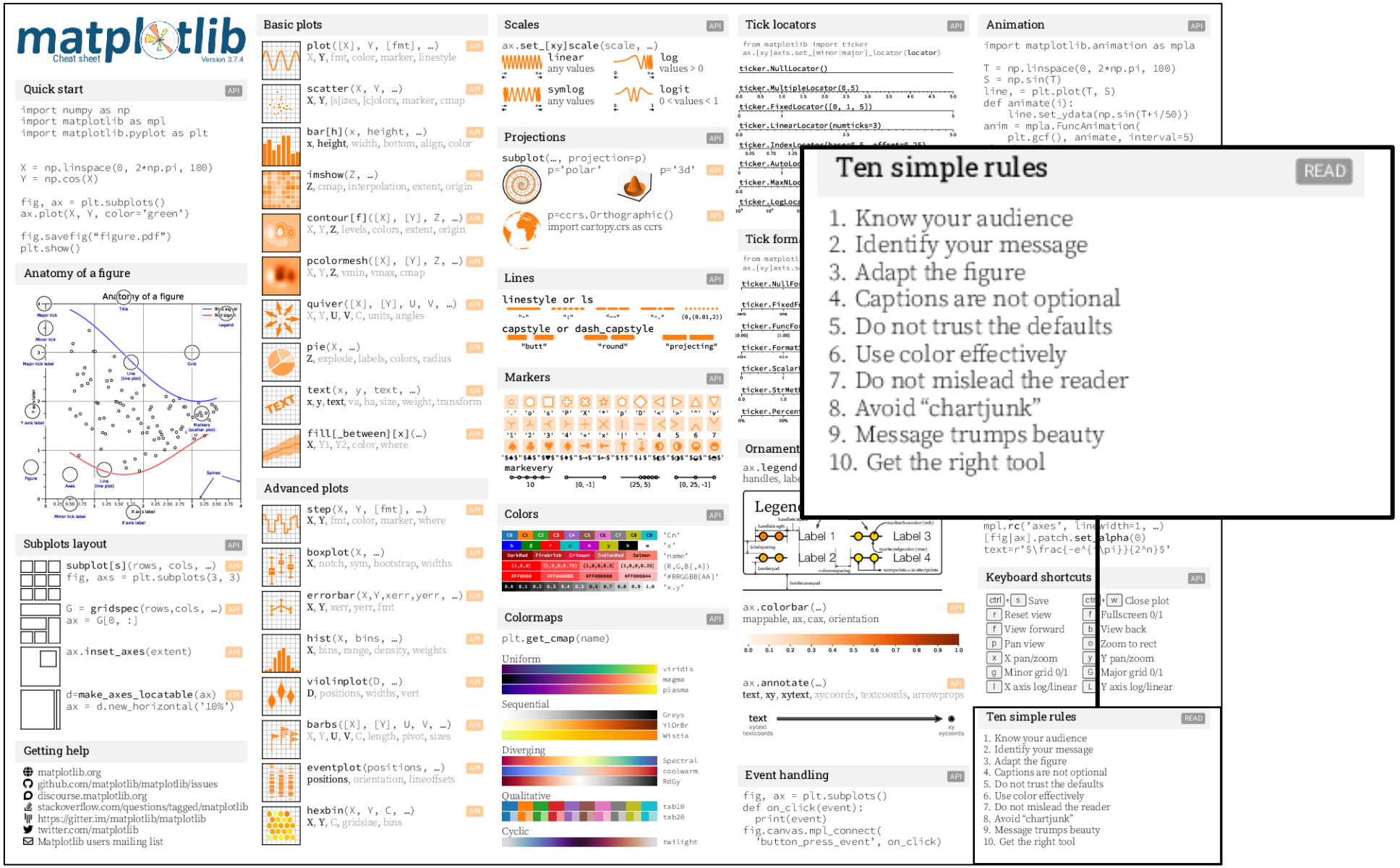
Choosing Types of Visualizations

University of Bremen

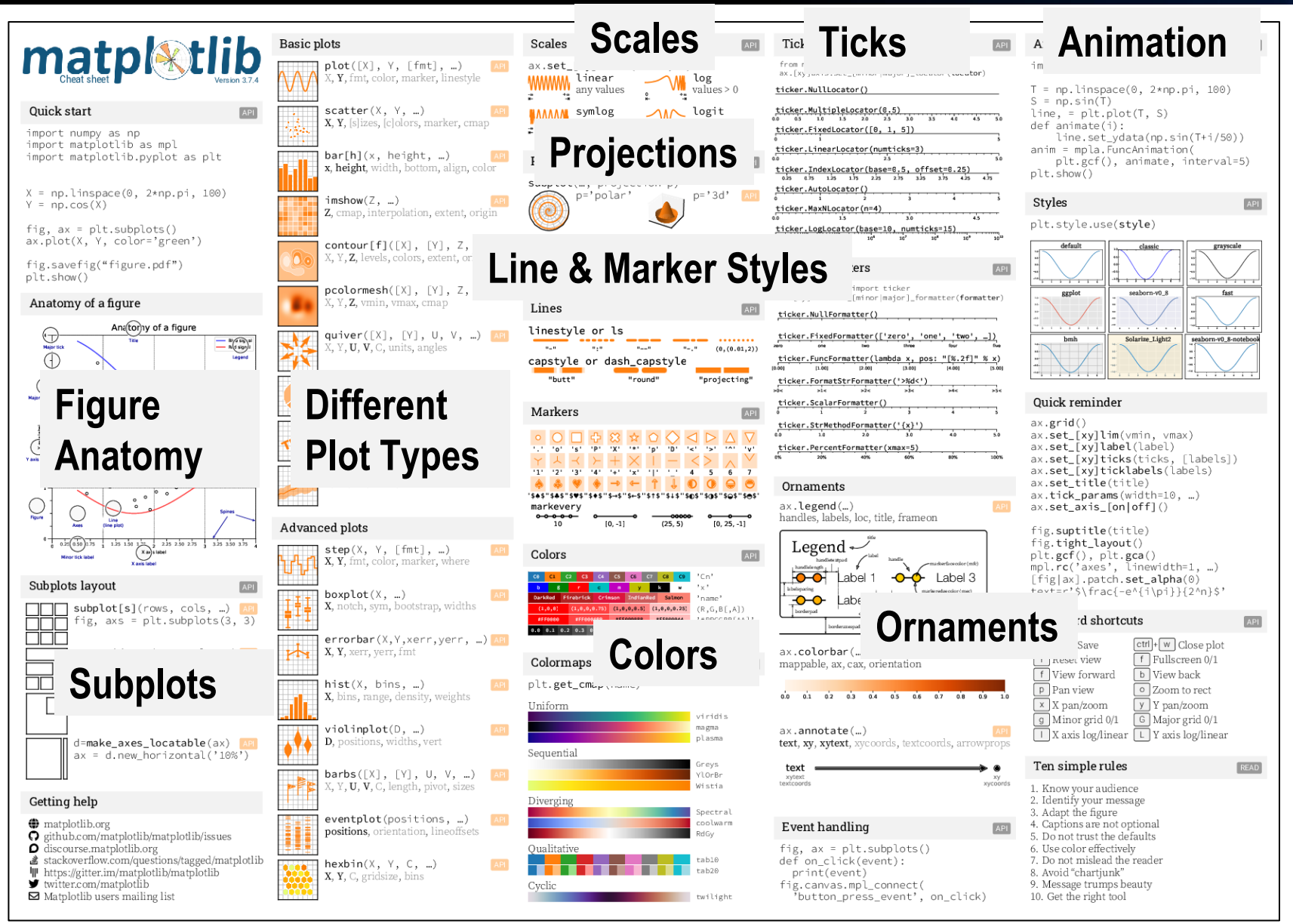
Source: https://seaborn.pydata.org/examples/index.html



Python Packages Cheat Sheets



Python Topics for Data Visualization



Scales

Ticks

Animation

Projections

Line & Marker Styles

Figure Anatomy

Different Plot Types

Subplots

Colors

Ornaments

Other packages:

Spatial Data Visualizations Using Maps

Interactive Visualizations