

# Better Data visualization for biology

Helena Jambor  
Dresden/DE & Chur/CH  
@helenajambor

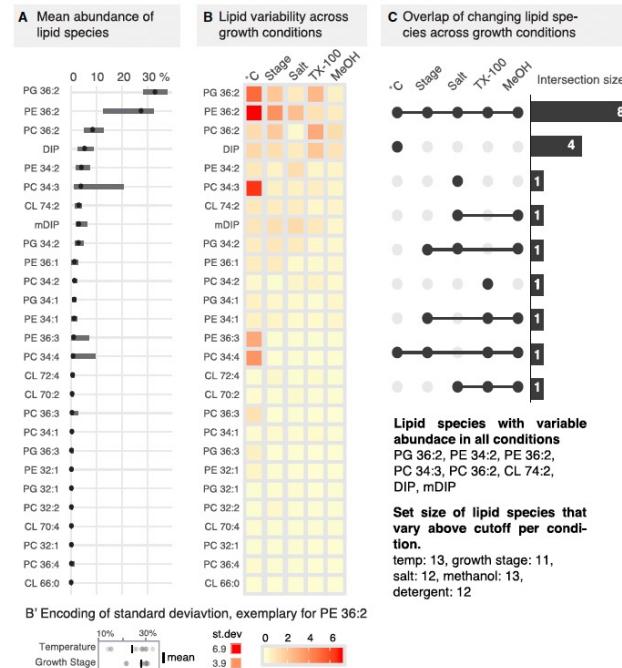
# Visualization for...

## Data exploration



Cell Reports  
Resource

CellPress  
OPEN ACCESS



Chwastek et al 2020

# Visualization for...

Data exploration



Data presentation



# Visualization for...

## Data exploration

Schmied&Jambor:  
From microscope to  
publication image,  
2020

### Publishing images for papers & posters: CHEAT SHEET

**Image check list**  
Before publishing, ensure images are informative, truthful, and legible.

- Color visibility: Grayscale and dark colors on light background have highest contrast. Test visibility in color blind mode (left) and grayscale (right).
- Annotations complete? Explain annotations and abbreviations in figure legend. Hint: best avoid abbreviations!
- Image and text size on page: Test image size and resolution. Text is test legible?
- Image resolution: Be careful with image compression when saving, avoid pixelated images.
- No manipulation: No individual image feature may be enhanced, obscured, removed, or introduced.

**Magnification**  
Evidence in images must be visible to be convincing. Scientists should decide on the necessary detail and fill the frame to show the information clearly.

**Zoom Insets**  
Different magnified regions-of-interest are used when two magnifications are needed. Indicate inset position in original image, do not obstruct key image features.

**Color**  
Color or not? Visibility depends on color lightheft and background color.

**Annotation**  
Scale information Every image needs a scale bar to relate it to reality. Options:

- Micrograph of a colored stain (e.g. histology)
- Micrographs of fluorescent stain, imaged in grayscale; consider grayscale.
- Micrographs of fluorescent stain color, use color-blind safe combination.
- Electron micrographs, grayscale by default, use grayscale.

**Layout**  
Reading direction For multi-panel figures, choose either column or row layout for best readability. Mixed layouts are hard for audience and appear cluttered.

**Annotation on a busy background**  
X Poor visibility: scale bars, overlapping scale bars, image help visibility.

**Point out details**  
Arrows: Lines: Regions of interest: Labels: Regions of interest: Enter code: Arrows: Tips: Lines: Regions of interest: Labels: Regions of interest: Enter code: Arrows: Points to structures, direct point to labels. Align, avoid arrow crossings. Lines: Label at line end: Choose suitable point size for labels. Align, avoid line crossings. Labels: Label many objects: Choose suitable font size for labels.

**Resources for color**  
Channel colors are adjustable in ImageJ/Fiji, referred to as "LUTs". These LUTs can also be inverted and new LUTs defined.

**TIPP**  
Highest contrast: dark objects on light background, and in grayscale.

**Figure legend**  
Figure legends inform briefly about result, experiments, and figure elements.

- Include here: short journal guidelines
- Annotate images directly if not possible indicate scale and colors in legend.
- State species (e.g. with **BBID**, tissue/cell type, and growth conditions).
- Avoid method details and discussion.

**Cropping**  
Draw selection in toolbar.

- Image > Crop... - Ctrl + Shift + X
- Image > Duplicate... - Ctrl + Shift + D

**Attribution**  
CC BY SA Helena Jambor, Christopher Schmied, and others. All rights reserved. Images © Helena Jambor, Andreas Müller (EM), Tom F. Kuehne, Template Results.

## Data presentation

### Processing images for papers & posters: CHEAT SHEET

**Open & Save**  
Download: [www.fiji.sc](http://www.fiji.sc)  
Update and install:  
Help > Update...  
Image>Color>Manage update sites then add specific site

**Typical problems in images:**

- Offset Gap from Peak at 255
- Background not cropped
- High values cropped (clipping)
- Scaling artefact

**Color**  
TIPP In composite images, switch between channels for viewing details:  
Image > Color > Channels Tool  
To adjust color, split composite image to separate channels:  
Image > Color > Split

**Annotate**  
Set scale:  
Analyze > Tools > Scale Bar...

- Width: 1/5/10 steps
- Color: highest contrast add background
- Overlay: separate layer from image (lost in png)
- Hide text and add later if resolution of image too low

**Image Processing**  
Gaussian blur:  
Process > Filters > Gaussian Blur...  
Projection:  
Image > Stacks > Z Project...

**Rotation & Resizing**  
Image > Transform > Rotate 90 Degrees...  
Image > Transform > Flip...  
Ideally measure before rotation

**Original** 90 Degrees 45 Degree

**TIPP** sketch, draw, use post-it & pen until it works. Only then start an electronic version.

**Figure legend**  
Figure legends inform briefly about result, experiments, and figure elements.

- Include here: short journal guidelines
- Annotate images directly if not possible indicate scale and colors in legend.
- State species (e.g. with **BBID**, tissue/cell type, and growth conditions).
- Avoid method details and discussion.

**Cropping**  
Draw selection in toolbar.

- Image > Crop... - Ctrl + Shift + X
- Image > Duplicate... - Ctrl + Shift + D

**TIPP** Get feedback! Test your figure by asking a colleague to explain it back to you: this quickly uncovers missing text, insufficient annotations etc.

**Merge channels with predefined LUT:**  
Image > Color > Merge Channels...

- tick "ignore source LUT's to merge with custom LUT's"
- no merge
- Create composite
- Save source image
- Save merged image

**Cropping**  
Create your own LUT:  
Image > Color > Edit LUT...  
**TIPP** Get feedback! Test your figure by asking a colleague to explain it back to you: this quickly uncovers missing text, insufficient annotations etc.

**Color blindness**  
Test color blind safety: Most common form color blindness: deutanopia. Test general visibility in grayscale or "Monochromacy". Required:  
Image > Color > Stack to RGB...  
Image > Color > Simulate Color Blindness...

@helenajambor



# Visualization for...

## Data exploration

Quarep initiative;

Schmied et al: Community checklists for i publication, 2024

**nature methods**

Explore content ▾ About the journal ▾ Publish with us ▾

nature > nature methods > perspectives > article

Perspective | Published: 14 September 2023

## Community-developed checklists for publishing images and image analyses

Christopher Schmied  , Michael S. Nelson, Sergiy Avilov, Gert-Jan Bakker, Cristina Bertocchi, Johanna Bischof, Ulrike Boehm, Jan Brocher, Mariana T. Carvalho, Catalin Chiriteanu, Jana Christopher, Beth A. Cimini, Eduardo Conde-Sousa, Michael Ebner, Rupert Ecker, Kevin Eliceiri, Julia Fernandez-Rodriguez, Nathalie Gaudreault, Laurent Gelman, David Grunwald, Tingting Gu, Nadia Halidi, Matthias Hammer, Matthew Hartley, ... Helena Klara Jambor  + Show authors

[Nature Methods](#) 21, 170–181 (2024) | [Cite this article](#)

32k Accesses | 19 Citations | 451 Altmetric | [Metrics](#)

## Data presentation

### Checklist for image publishing

#### Image format

- Focus on relevant image content (e.g. crop, rotate, resize)
- Separate individual images
- Show example image used for quantifications
- Indicate position of zoom-view/inset in full-view/ original image
- Show images of the range of described phenotype

Minimal

#### Image colors and channels

- Annotation of channels (staining, marker etc.) visible
- Min Max Adjust brightness/contrast, report adjustments, use uniform color-scales
- Image comparison: use same adjustments
- Channel colors high visibility on background Best visibility: grayscale
- Multi-colors: provide grayscale for each color channel
- Multi-color: if channels are merged, make accessible to color
- Provide intensity scales (calibration bar) for grayscale, color, color...)
- Pseudo-colored images: additionally provide grayscale version comparison
- Gamma adjustments: additionally provide linear-adjusted image comparison

Minimal

#### Image annotation

- Add scale information (scale bar, image length; in figure/figure)
- Explain all annotations (in figure/figure legend)
- Annotations should be legible (line width, size/point size, color)
- Annotations should not obscure key data
- Annotate imaging details important for interpreting the figure (Depending on the main message and imaging technique this e.g., image pixel size, imaging intervals (time-lapse in movies time, or anatomical section...))
- C



Ideal

#### Image availability

- Images are shared (lossless compression/microscope images)
- Image files are freely downloadable (public database)
- Image files are in dedicated image database (added value database or image archive)

@helenajambor

# Visualization for...

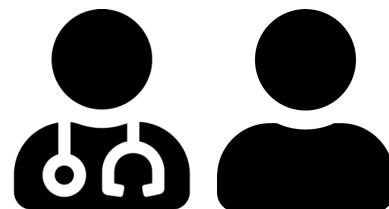
Data exploration



Data presentation



**Data communication**

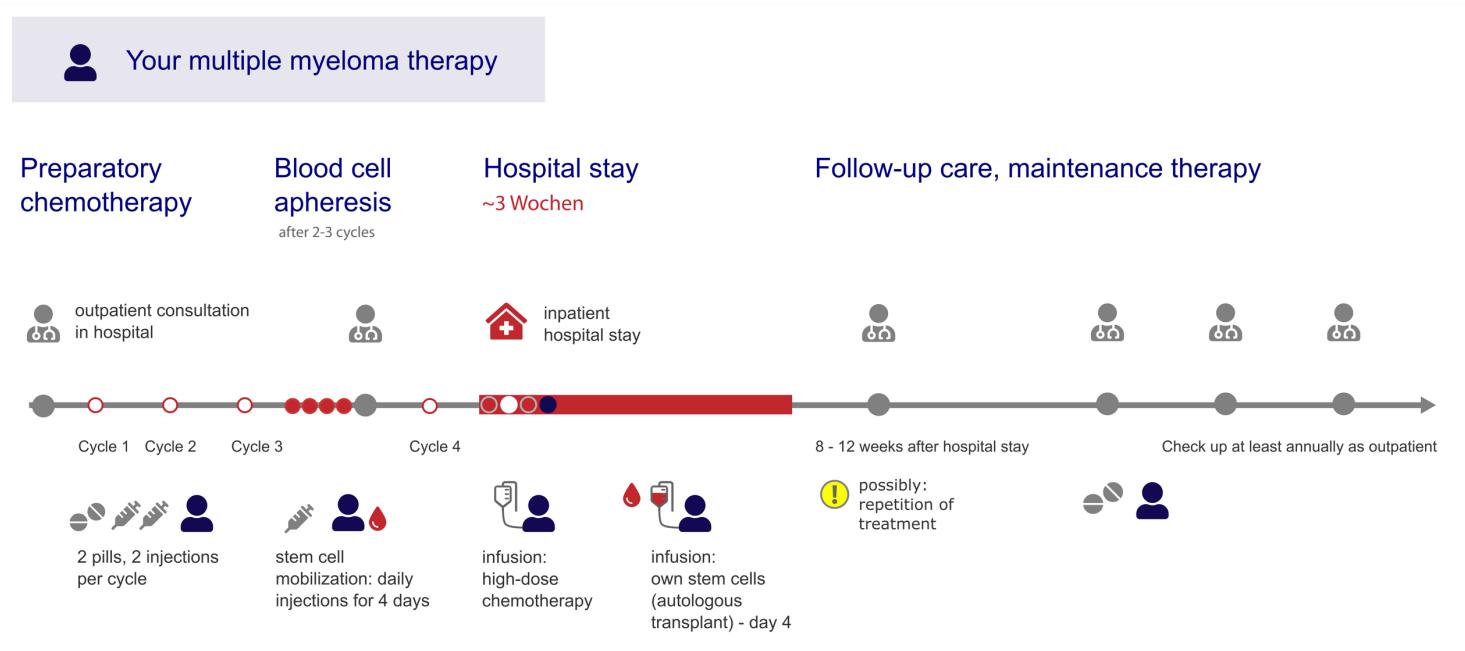


# Patient communication

Patients & MD: mismatch in health literacy.

Result: patients have low recall, which leads to non-adherence, stress, reduced quality of care.

Solution: include visual information to support verbal communication.



Why we visualize

K Z R

Q B T

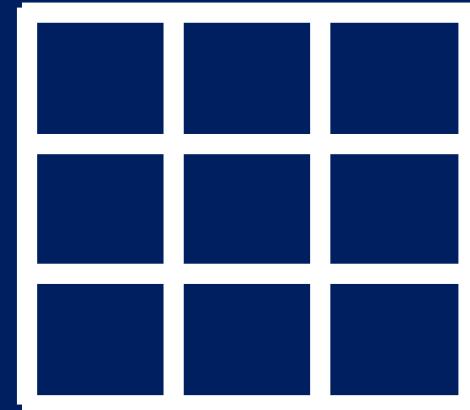
S G N

# 1. Visual perception is fast

In 0.3 seconds....

Letters of the  
alphabet

K Z R  
Q B T  
S G N



## 2. Visual perception is innate

We use eye to compare...



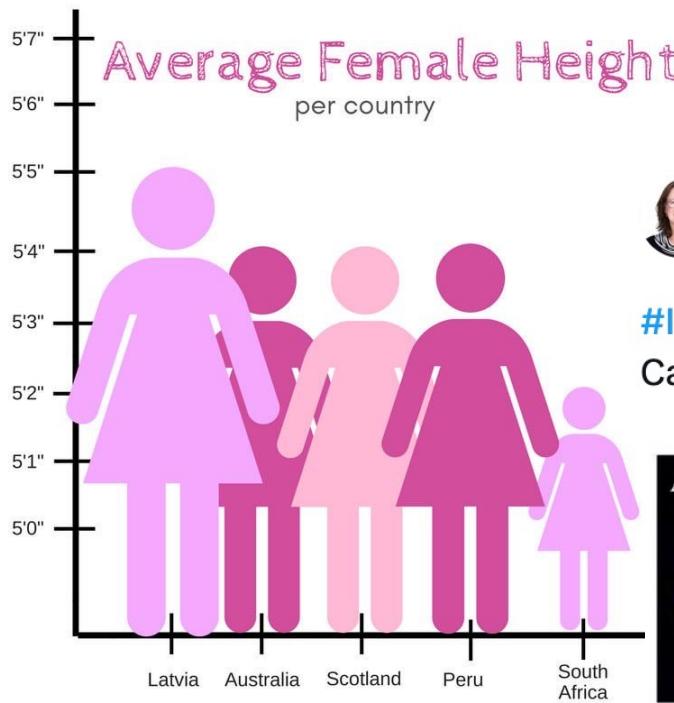
### 3. Visualizations are learned

We use eye to compare...



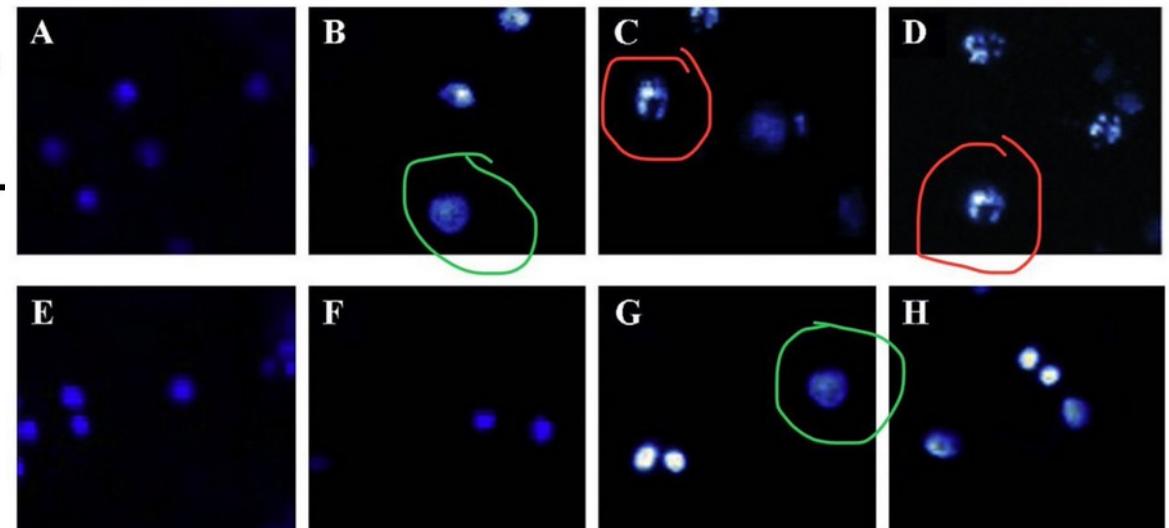
# Visualizations can fail

## Misleading



**Elisabeth Bik**  
@MicrobiomDigest

#ImageForensics - WifiOnATrain edition.  
Can you spot the overlap(s)?



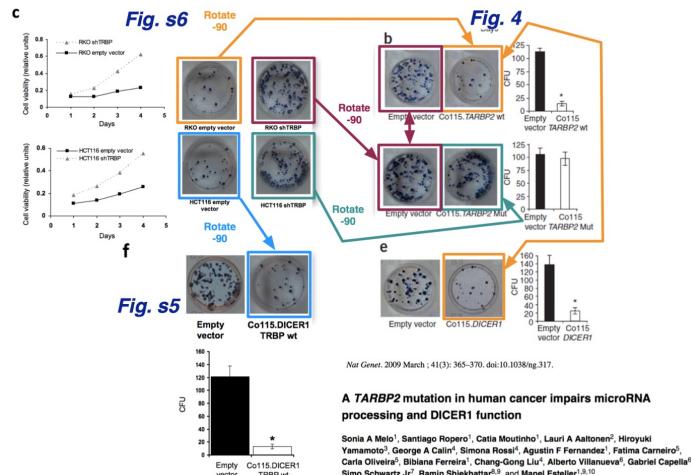
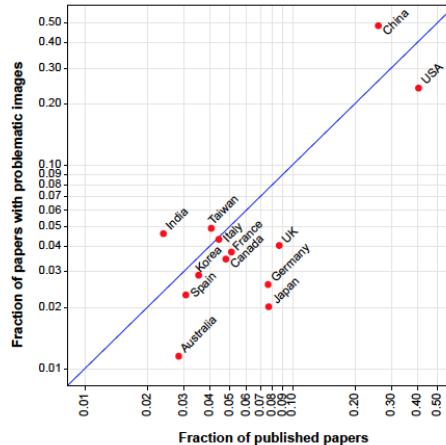
# Visualizations can fail

## Misleading QUANTIFICATION

Rossner, 2004; Elisabeth Bik 2016 & blog;

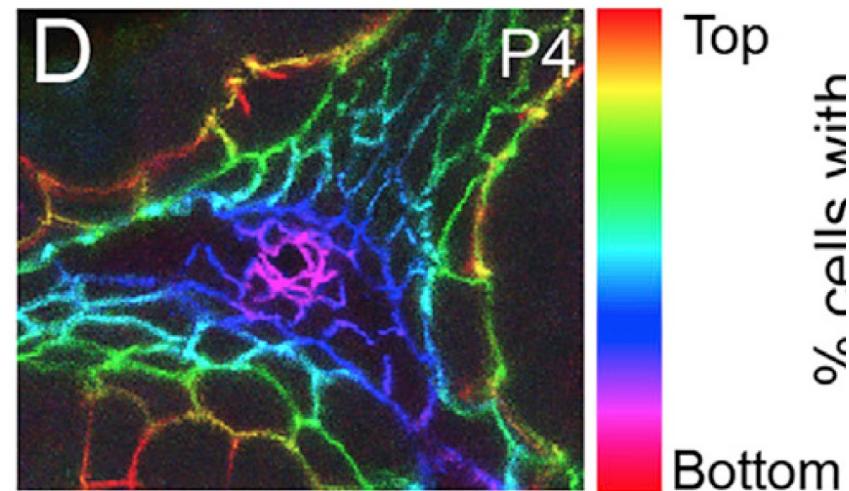
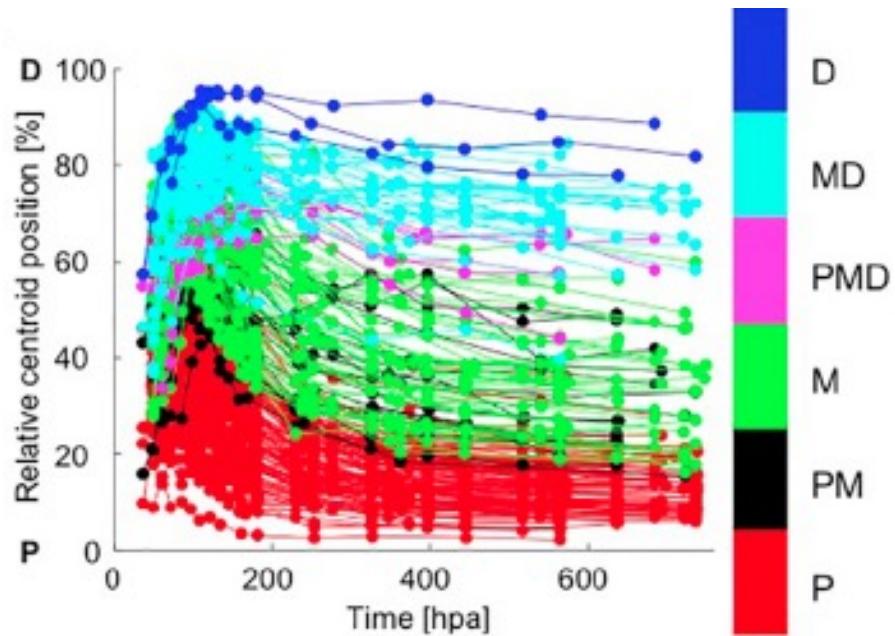
Bik EM... *The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications.* *mBio.* 2016;7(3).

Rossner M, Yamada KM. What's in a picture? The temptation of image manipulation. *J Cell Biol.* 2004



# Visualizations can fail

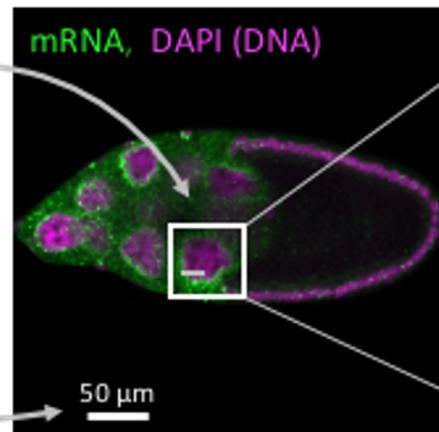
Confusing



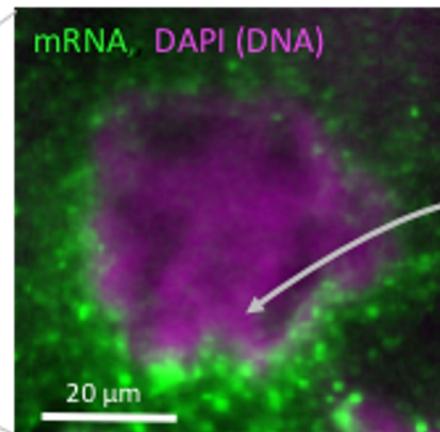
# Visualizations can fail

## Confusing QUANTIFICATION

**Inset** incorrectly or not at all annotated: 35-45% of published images.



**Scale bar** missing / incomplete: 45-70% of published images.



**Annotation** not color blind save: 15-30% of published images.

**Color** not color blind safe: 30-50% of published images.

**Figure legend** incomplete: 15-40% of published images.

PLOS BIOLOGY

BROWSE PUBLISH ABOUT

OPEN ACCESS PEER-REVIEWED

META-RESEARCH ARTICLE

Creating clear and informative image-based figures for scientific publications

Helena Jambor, Alberto Antonietti, Bradley Alicea, Tracy L. Audisio, Susann Auer, Vivek Bhardwaj, Steven J. Burgess, Iuliia Ferling, Małgorzata Anna Gazda, Luke H. Hoeppner, Vinodh Ilangoian, Hung Lo, Mischa Olson, [ ... ], Tracey L. Weissgerber [ view all ]

Version 2 Published: March 31, 2021

<https://doi.org/10.1371/journal.pbio.3001161> See the preprint

Article	Authors	Metrics	Comments	Media Coverage	Peer Review
---------	---------	---------	----------	----------------	-------------

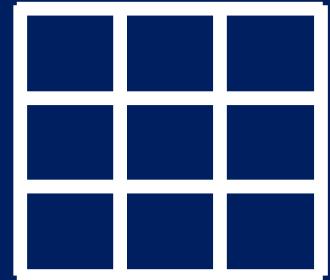


Take a sad plot,  
and make it better



**What?**  
(Display  
Type)

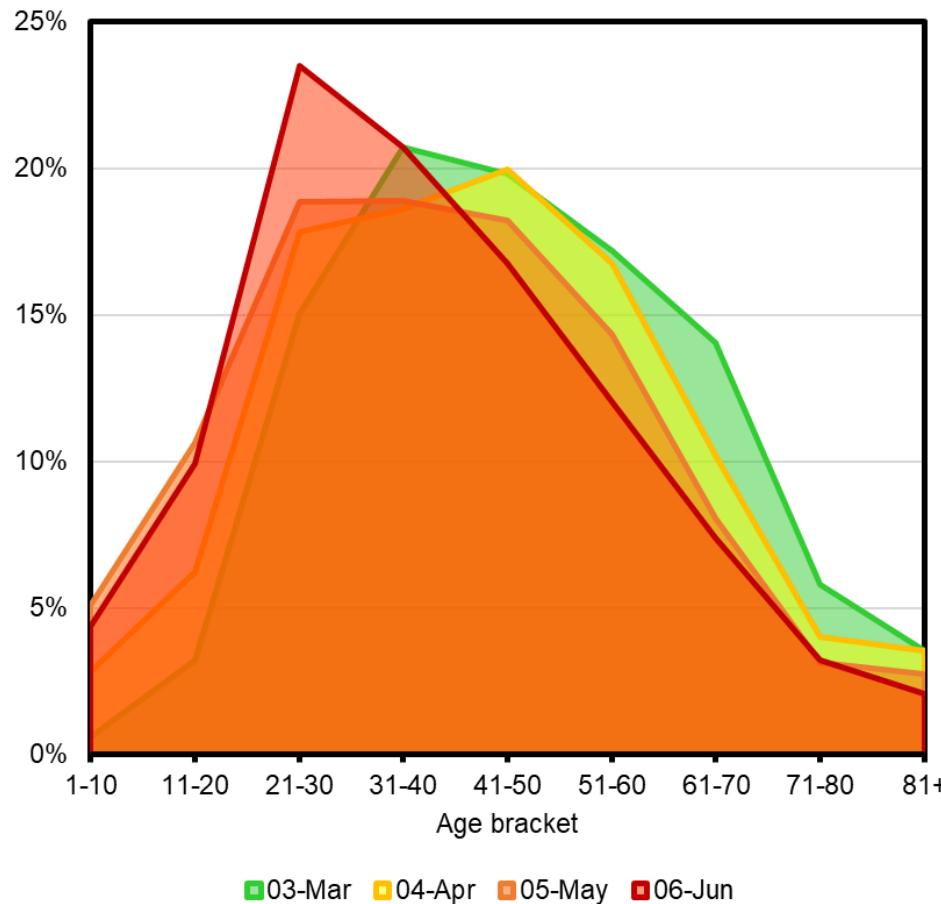
K Z R  
Q B T  
S G N



K Z R  
Q B T  
S G N

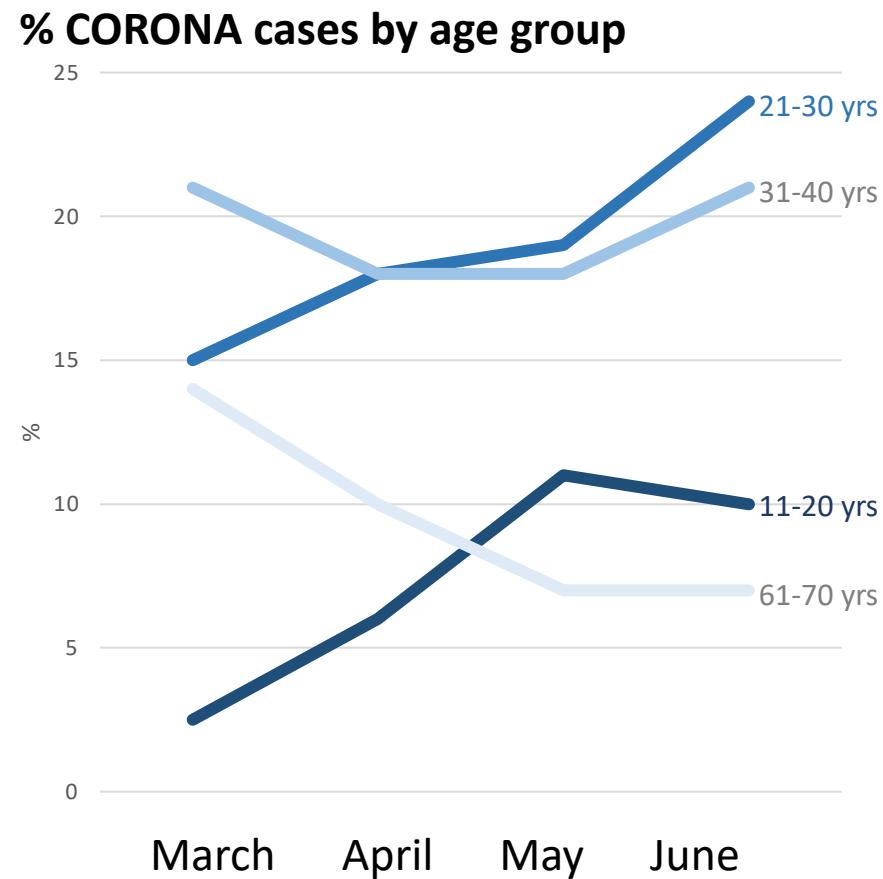
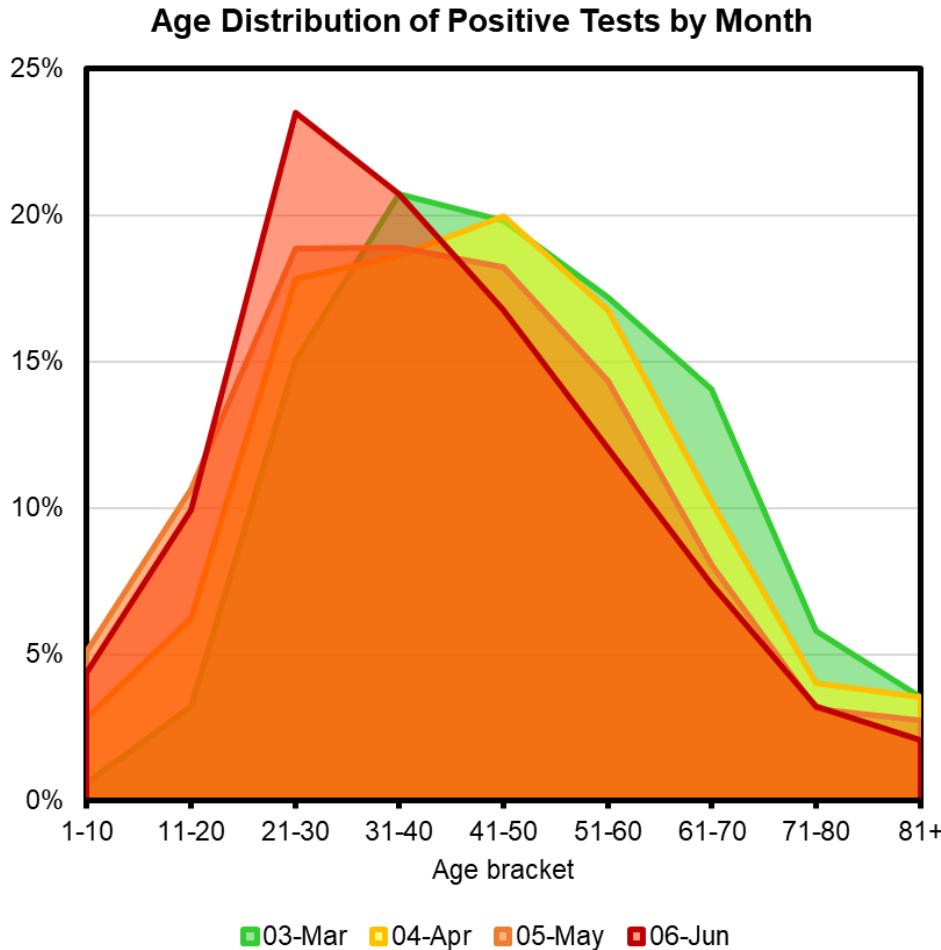
# 1. Pick a suitable chart type

**Age Distribution of Positive Tests by Month**



*personal communication*

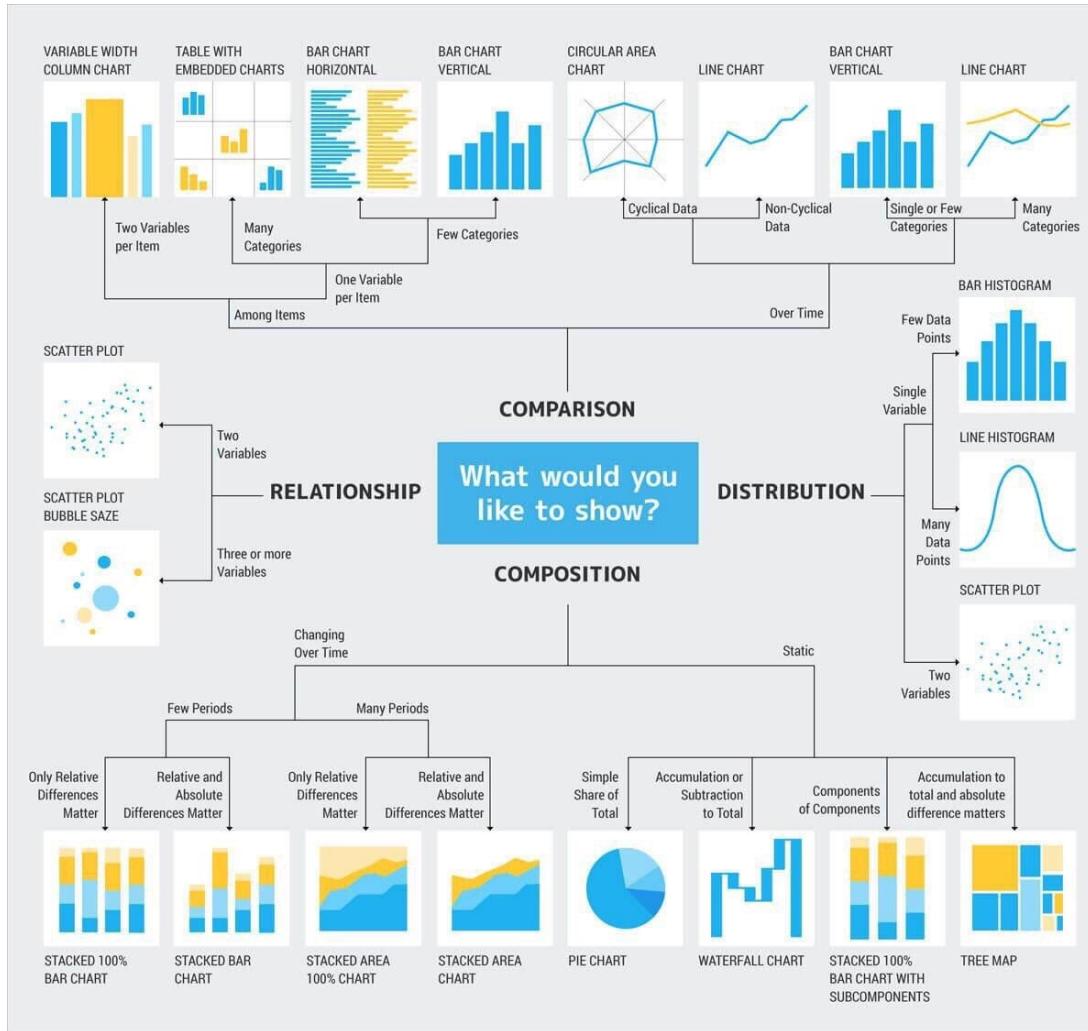
# What is the problem?



*personal communication*

# 1. Pick a suitable chart type

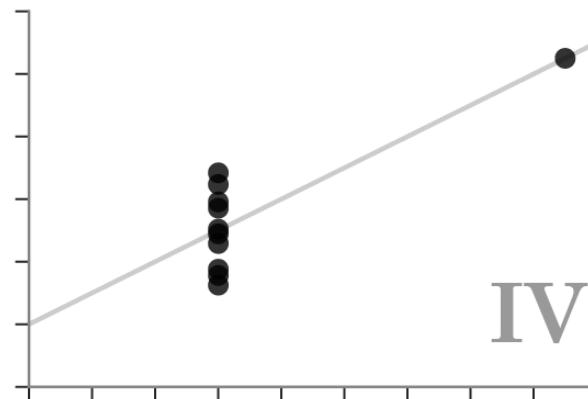
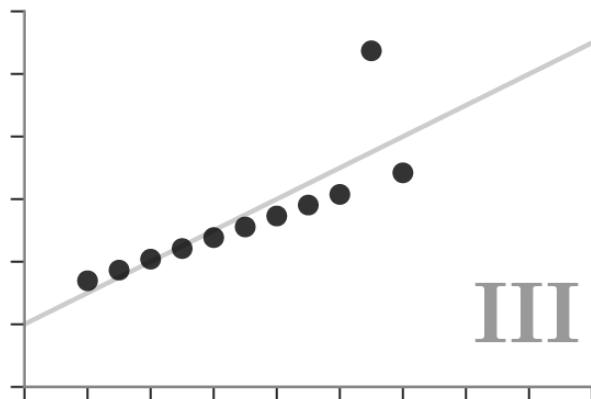
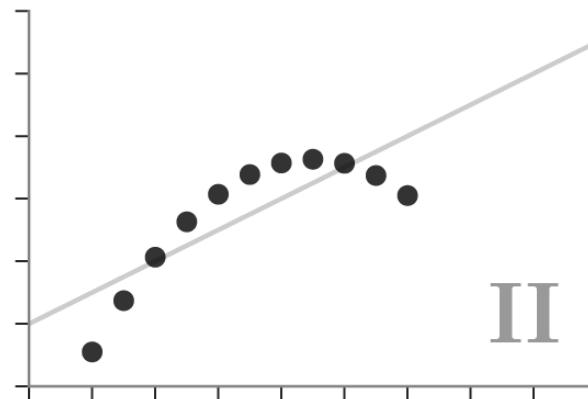
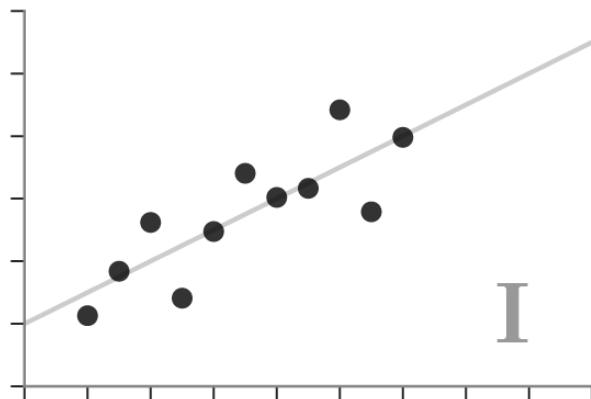
Resources for choosing charts:



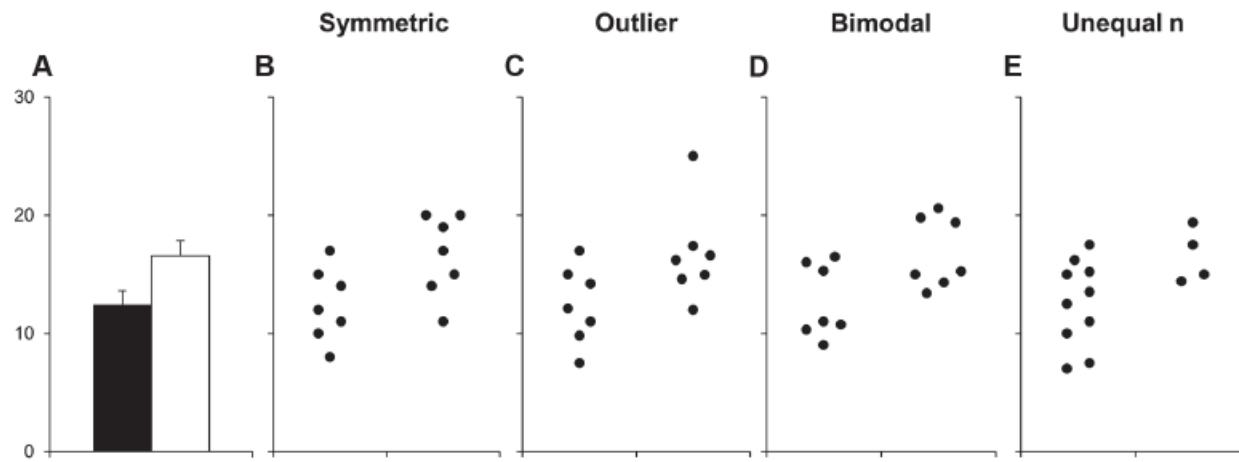
## 2. Show your data

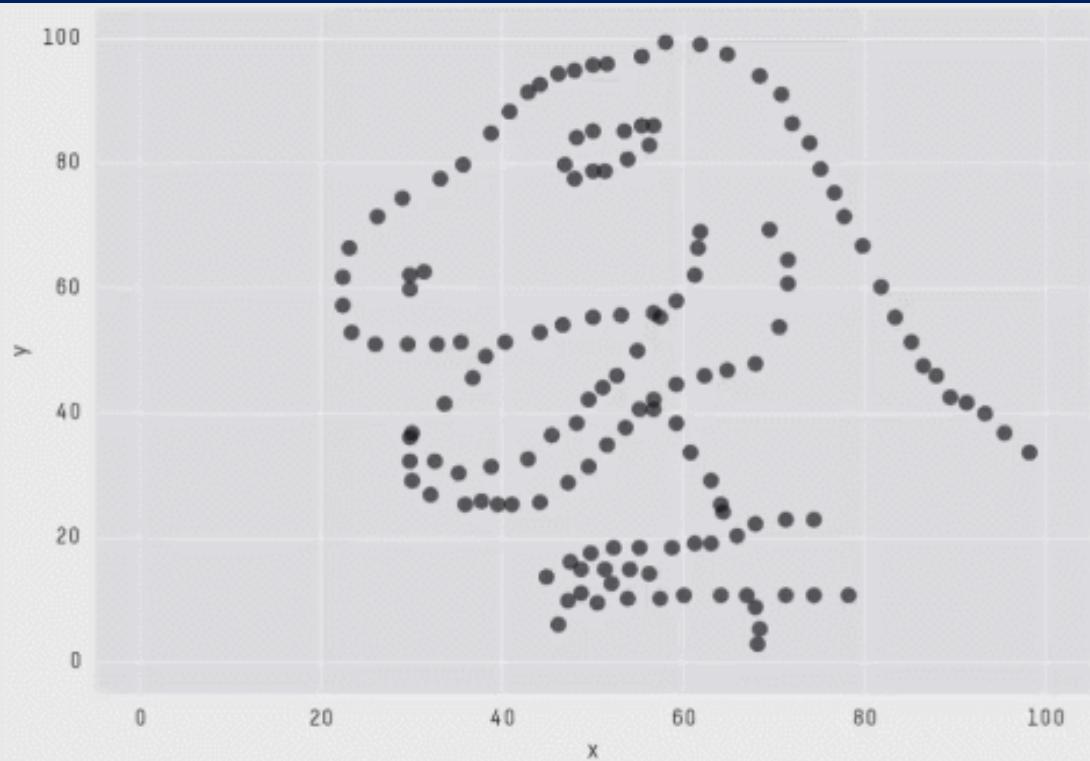
### Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



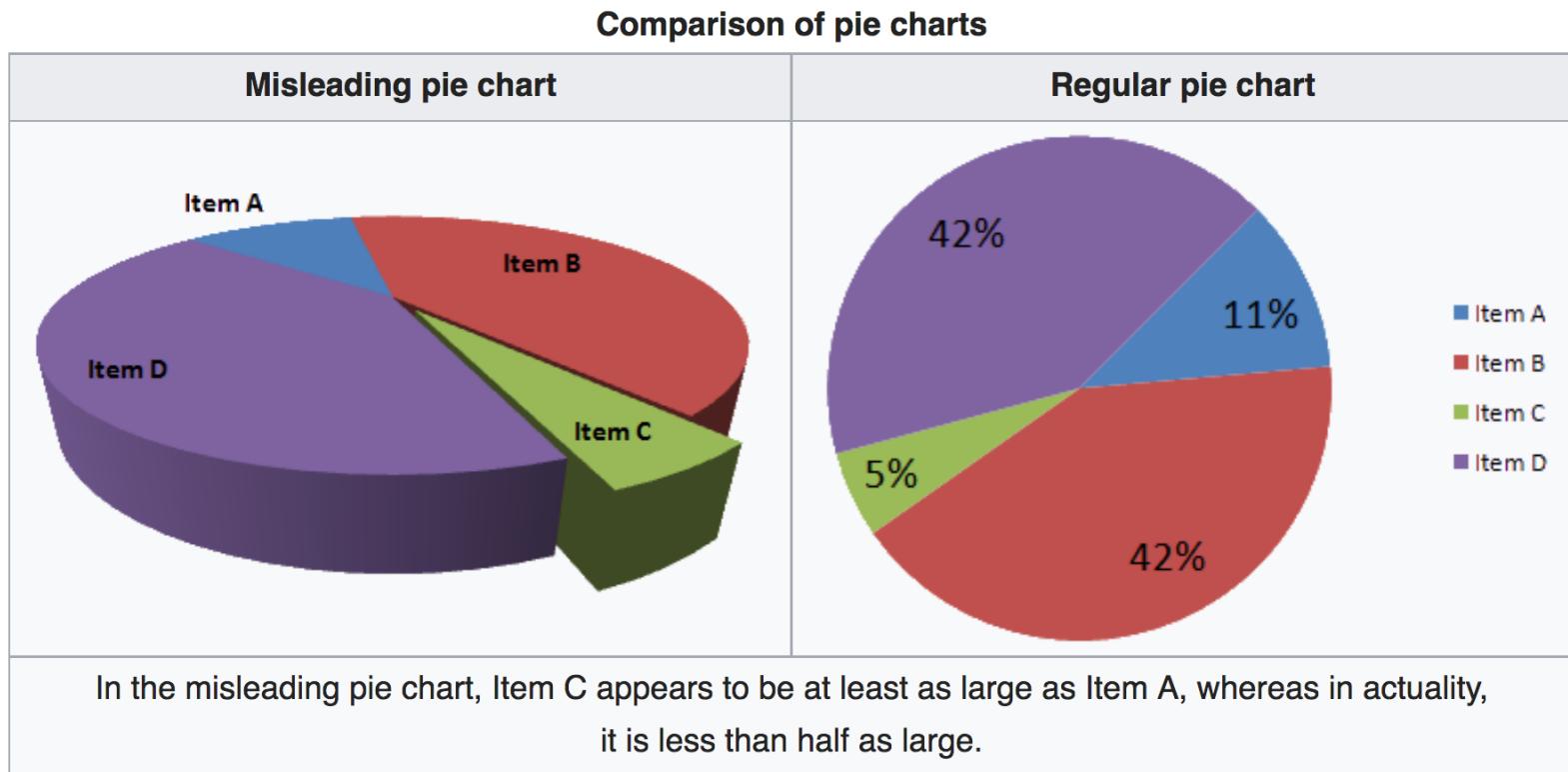
## 2. Show your data





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

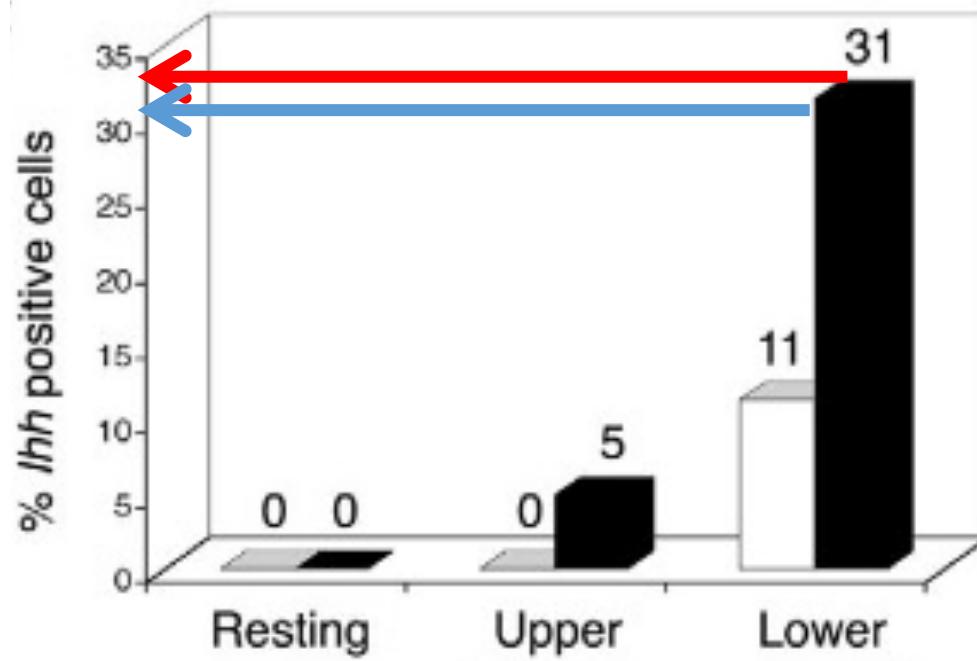
### 3. Avoid 3D charts



### 3. Avoid 3D charts

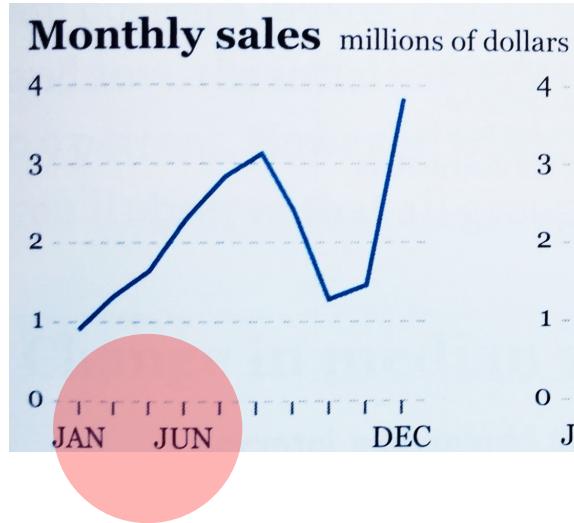
Eye sees this...

...but should read here

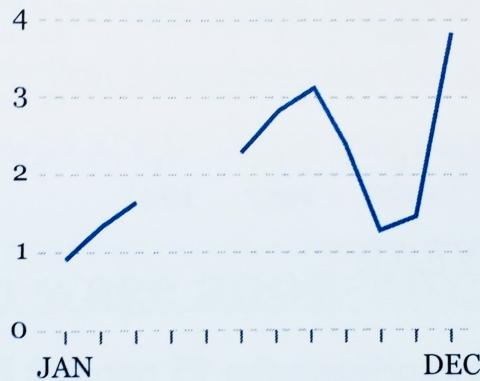


# 4 . Indicate missing data

Misleading choice



Better choices



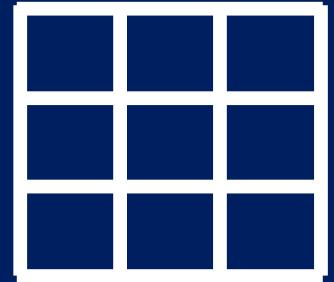
No line for highly variable data



Dashed line for continuous data, e.g. growth of plants.

**What?**  
(Display  
Type)

**Text**



# 5. Label

What are the axes?

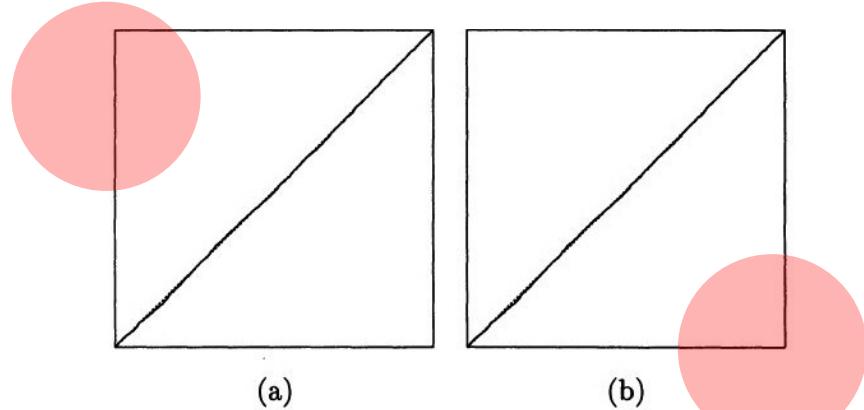
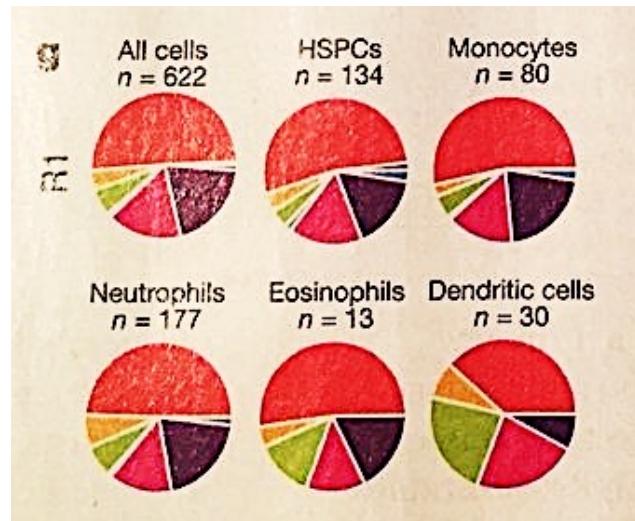


Figure 1. SRQ Plots of  $T_i/T_n$  (Vertical Axes) Against  $i/n$  (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

Color code?



**What?**  
(Display  
Type)

**Text**

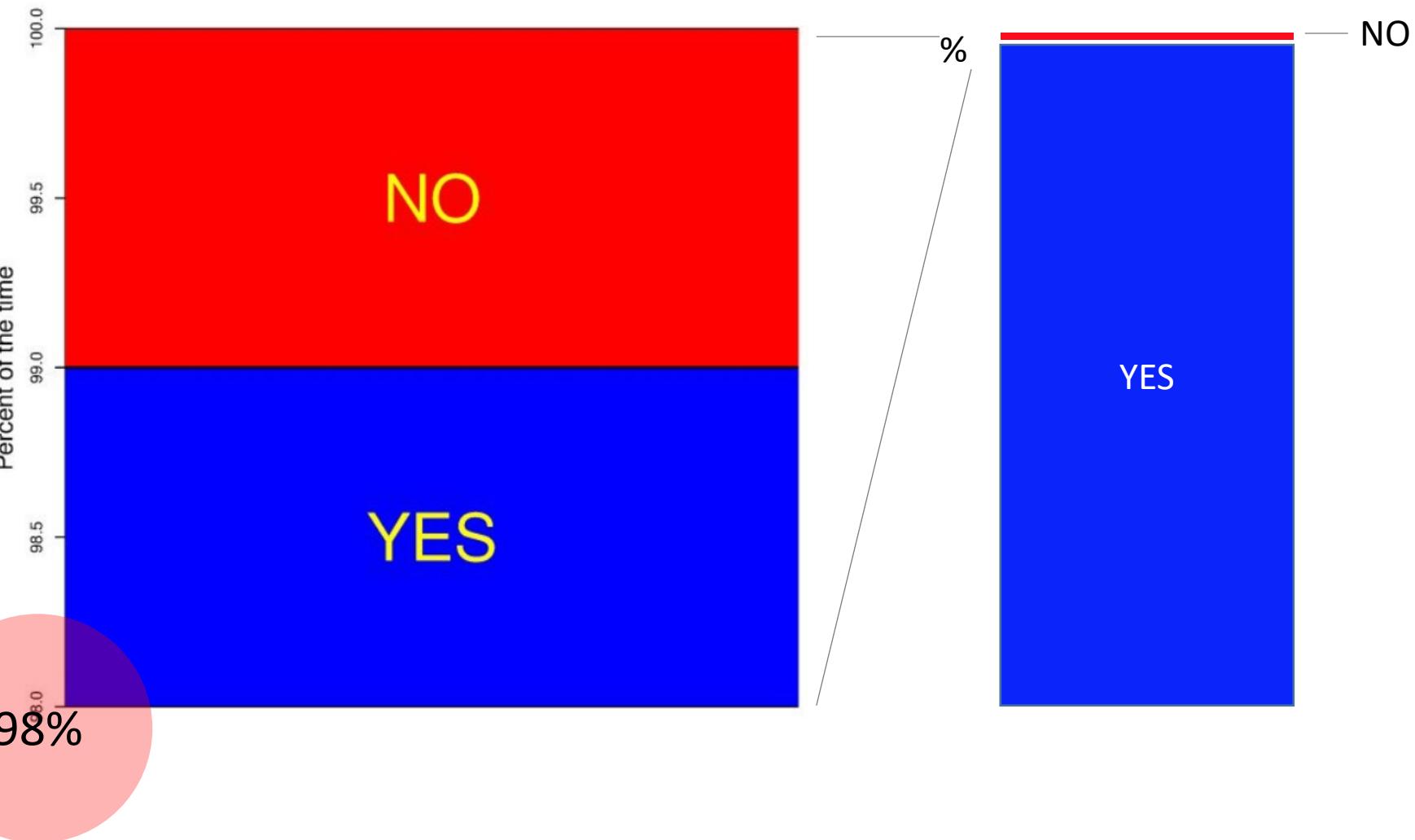
**Layout**



## 6. Bar charts need zero-baseline

Is Finland the prettiest country in the world?

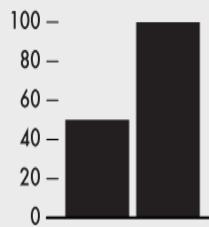
Half say 'yes', half say 'no'



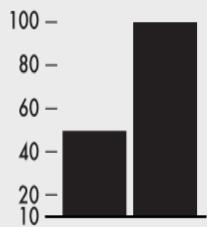
# 6. Bar charts need zero-baseline

Summary of the effect:

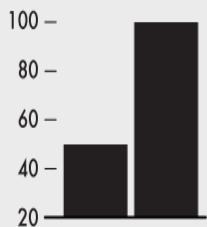
Baseline at 0



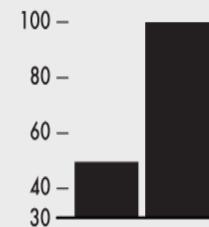
Baseline at 10



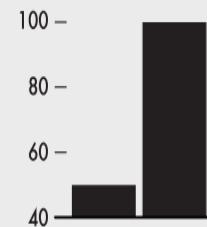
Baseline at 20



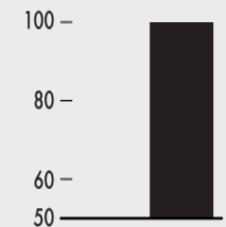
Baseline at 30



Baseline at 40



Baseline at 50



*This is correct.*

*Hm, first bar got shorter.*

*It's going...*

*...going...*

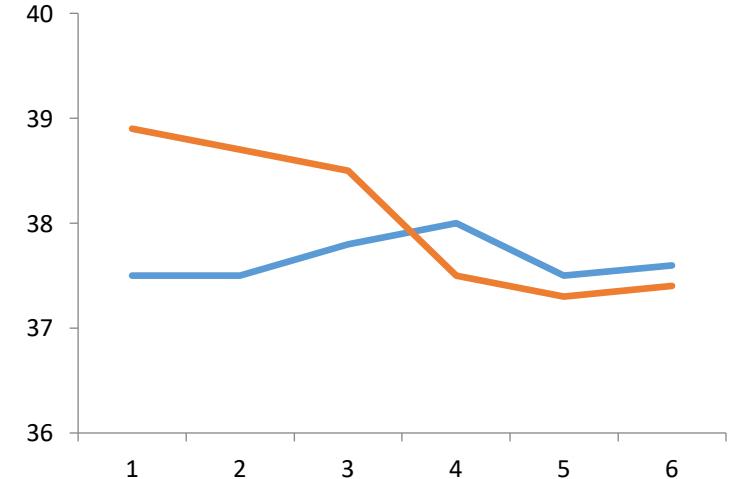
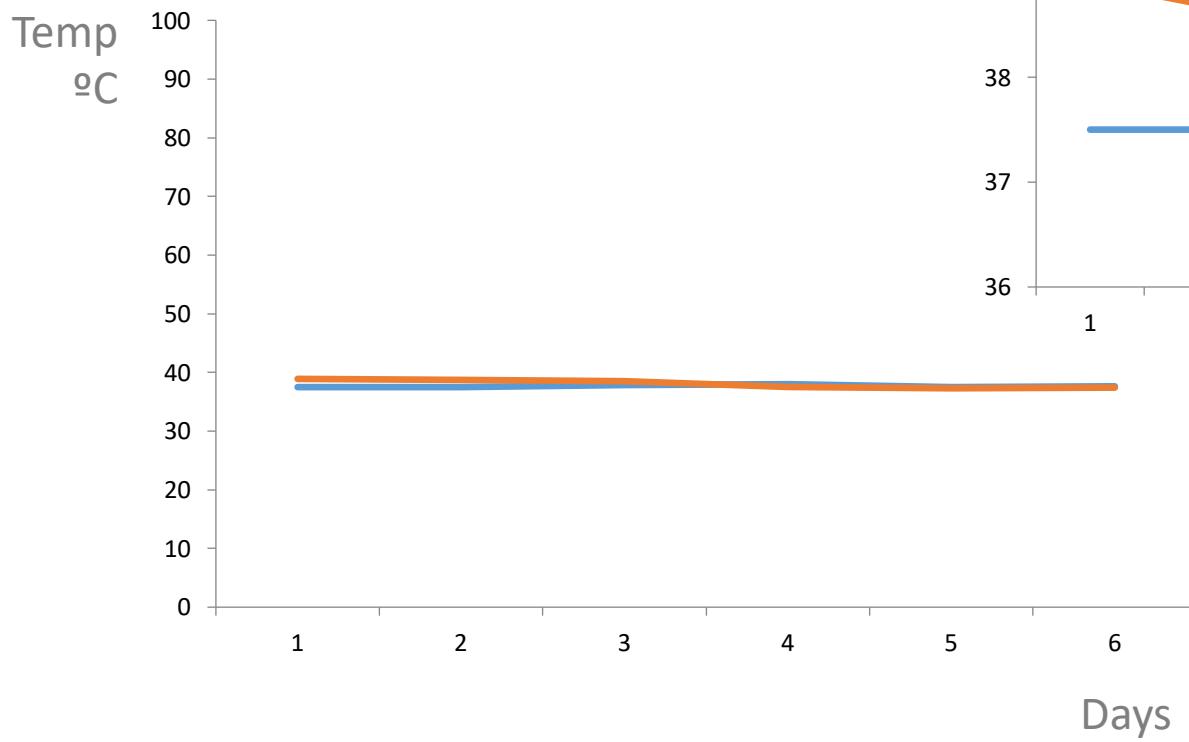
*...going.*

*First bar is gone.*

## 7. Line charts **may mislead** with zero-baselines

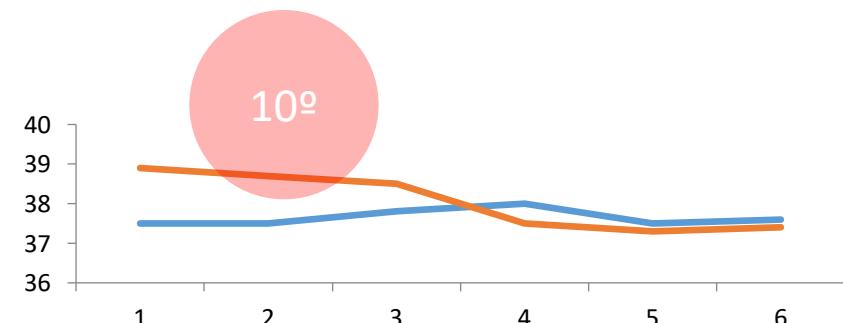
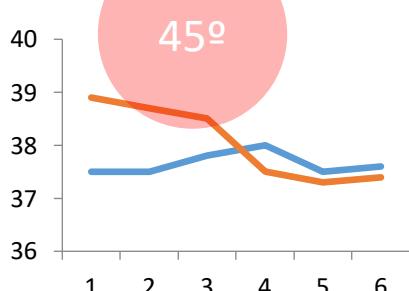
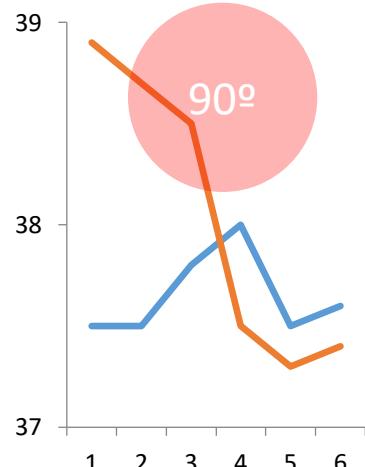
Patient fever chart

*...nobody has fever here!*



## 8 . Think about axis layout

Patient fever chart, correctly focused on physiological range

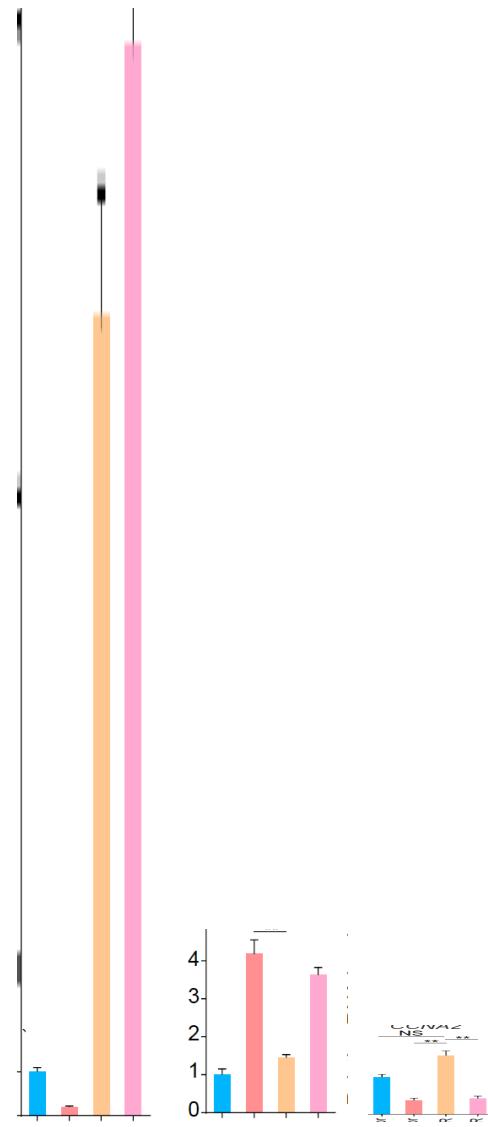
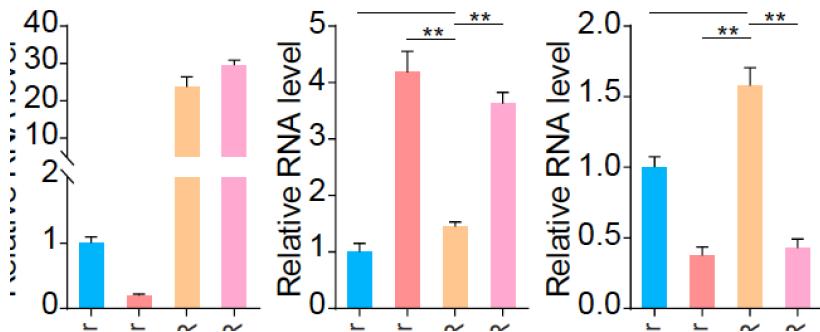


Trend visible best at  $\sim 45^\circ$

## 9. Avoid axis breaks

- not necessary
- never intuitive

Solutions: log-scale, normalization, 2 charts



**What?**  
(Display  
Type)

**Text**

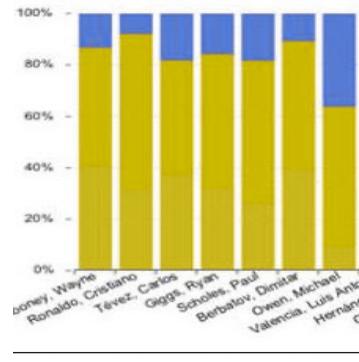
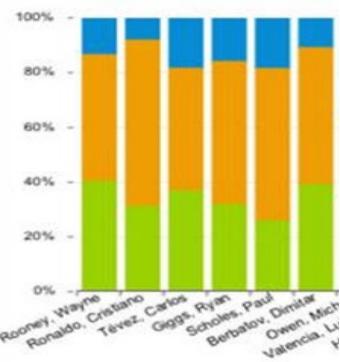
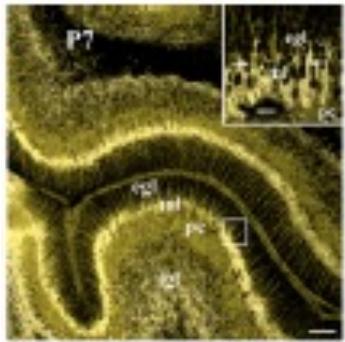
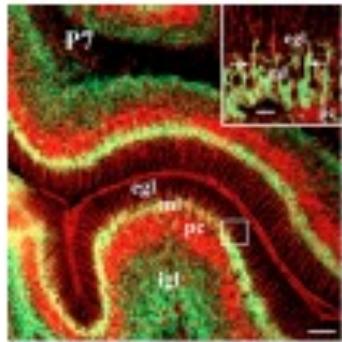
**Layout**

**Color**

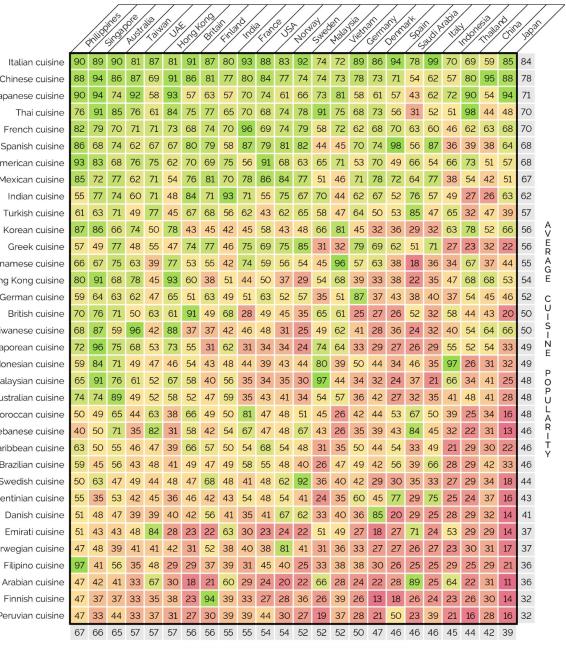
# 10. Pick suitable color

Sometimes data is encoded with **green** and **red** colors.

10% of male audience sees it like this:



Italian, Chinese and Japanese cuisines are the world's most popular  
% of people who have tried that cuisine in each country that say they like it



YouGov | yougov.com

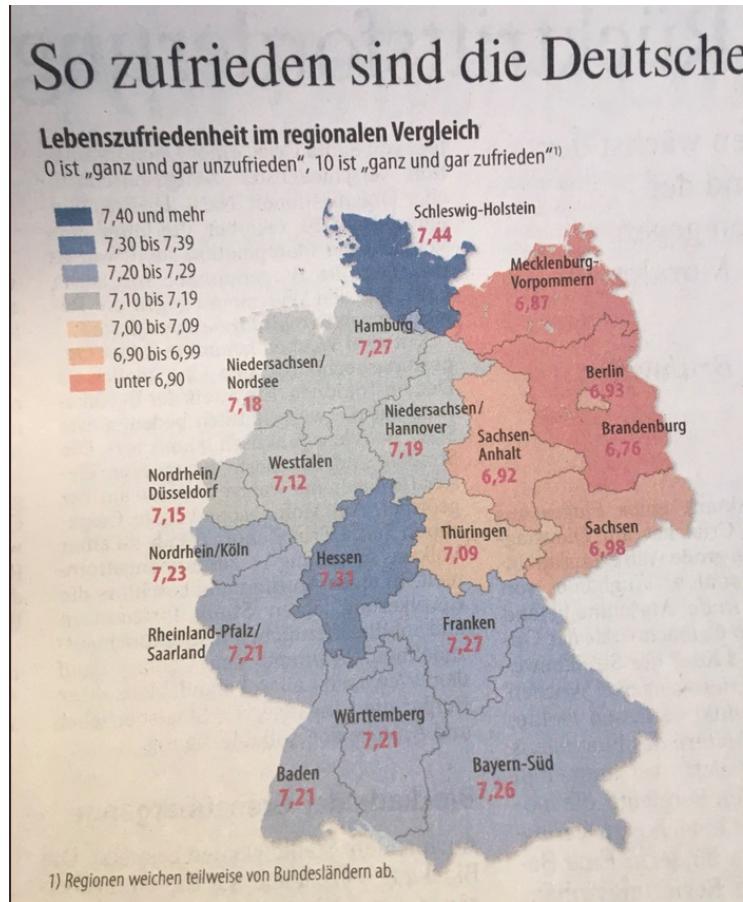
Nations most likely to enjoy foreign cuisine Nations least likely to enjoy foreign cuisine

May-December 2018

Be colorblind friendly & do not combine red & green

# Color

Happiness divide between East and West Germany?



Sequential values:  
1 Color **vary saturation**



2 Colors for **diverging** values



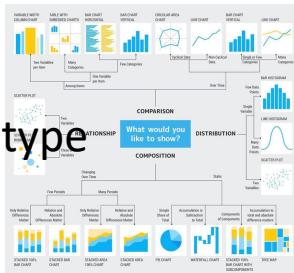
Categorical data  
vary the hue



# 10 tips to not lie with charts

1

Pick suitable chart type



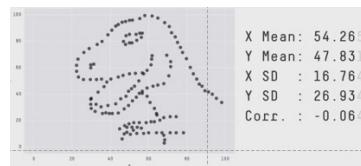
2

No 3D charts



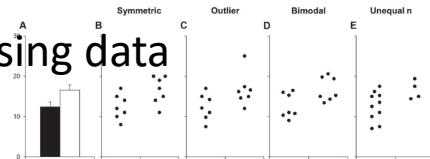
3

Show the data



4

Indicate missing data

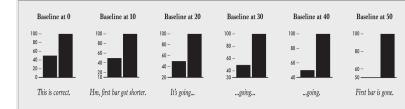


5

Label with text, avoid abbreviations

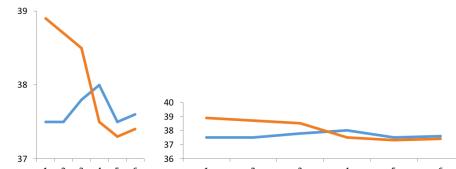
6

Bar charts need zero-baseline



7

Line charts: focus on relevant range



8

Choose suitable axis layout

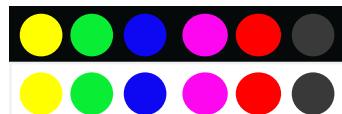


9

Avoid axis breaks

10

Pick accessible colors that are suitable to data type



Improve  
charts

Clean up (declutter)

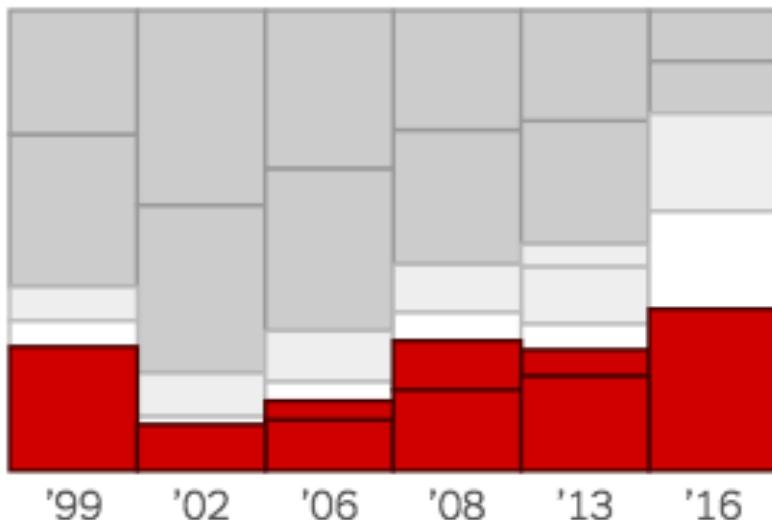
**Remove**  
to improve  
(the **data-ink** ratio)

# Focus attention

Too many messages? Too much data? **Highlight one aspect**

Party ideology in parliamentary elections\*     Center-left, center-right     Other parties     Right-wing

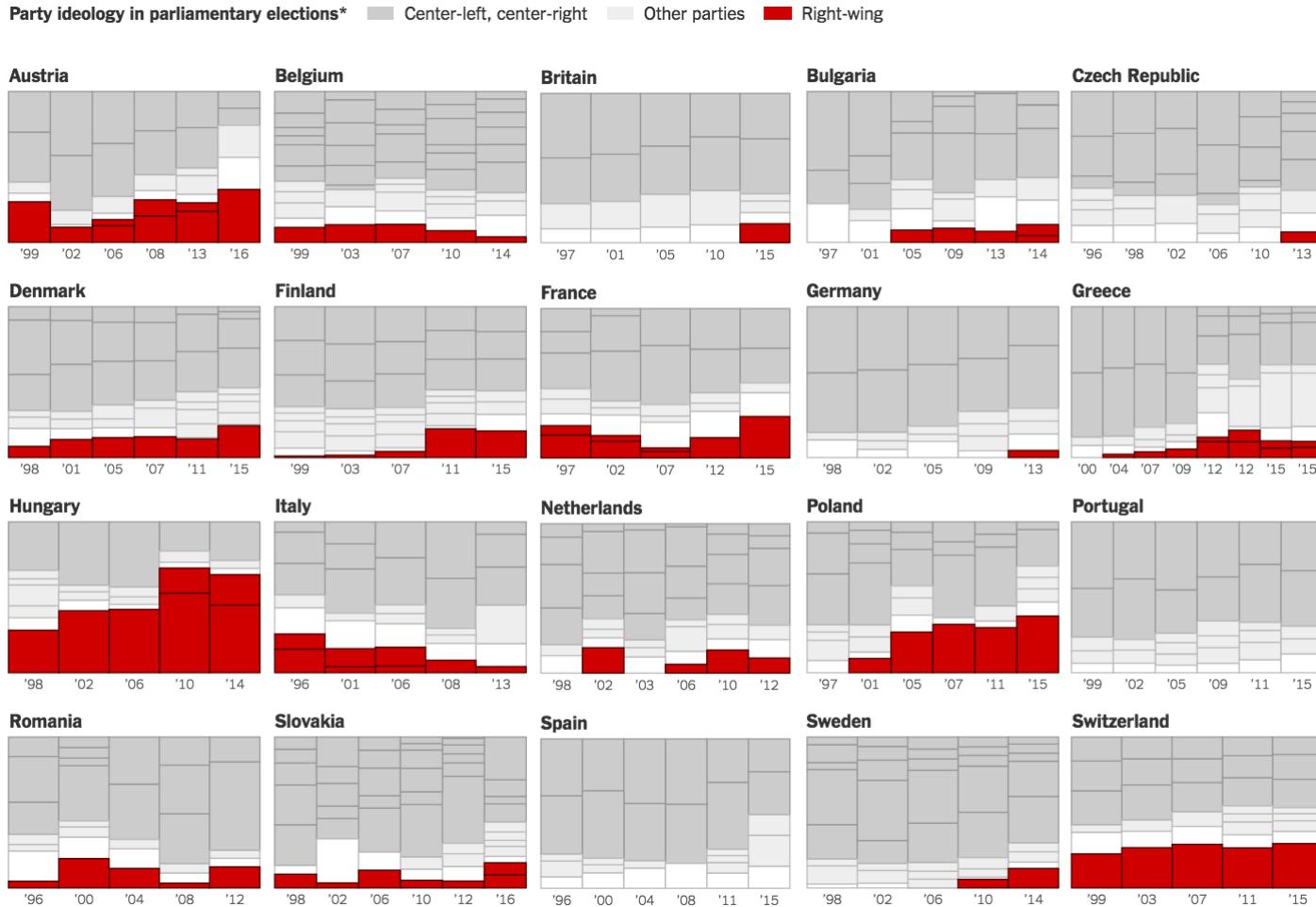
## Austria



>30 data  
points!

# Focus attention

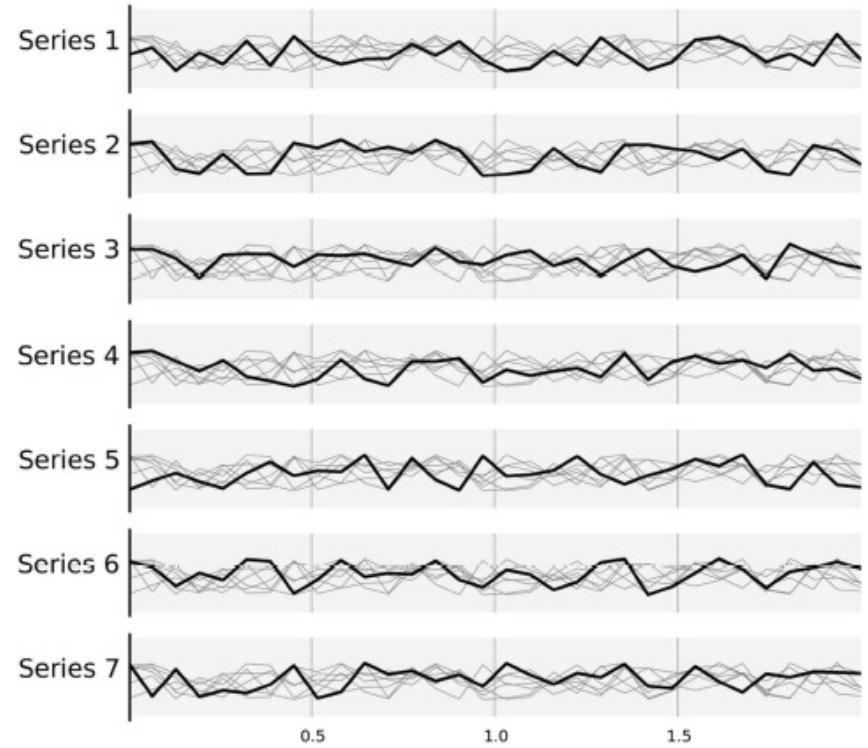
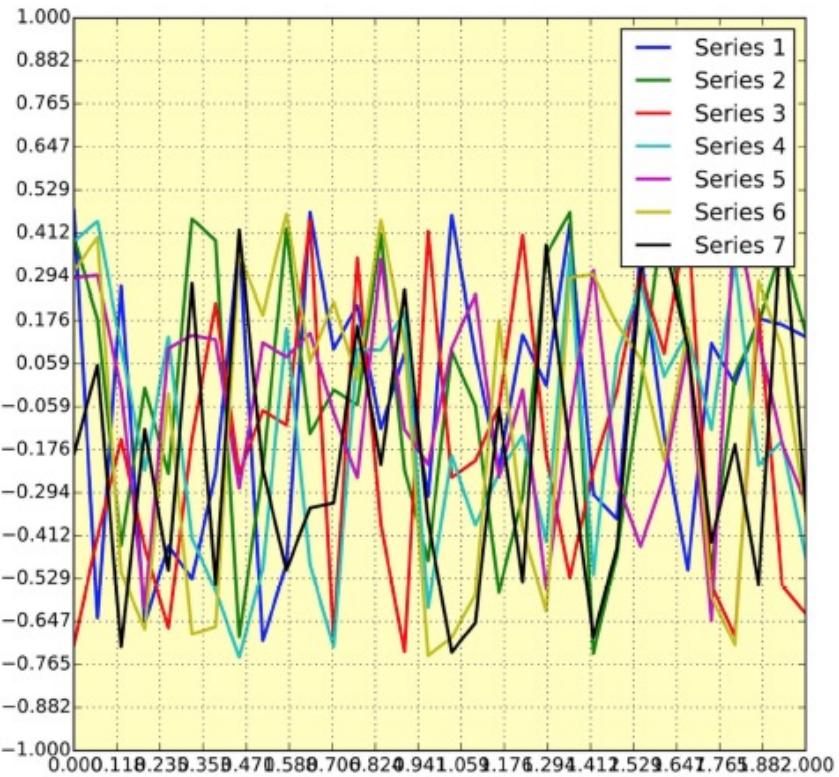
Too many messages? Too much data? **Highlight one aspect**



\*The 2016 Austrian presidential election and the 2015 French regional elections are included to add a more recent result for those countries.

# Small multiples

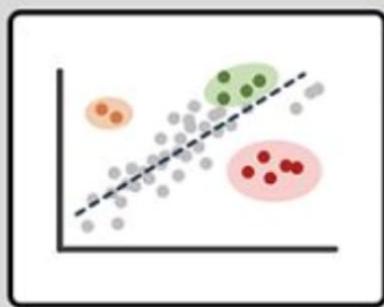
Too many messages? Too much data? **Many charts of same kind**



# Animation

Too many messages? Too much data? **Show step by step**

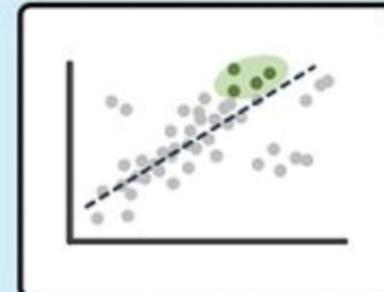
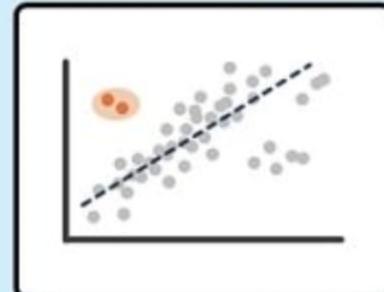
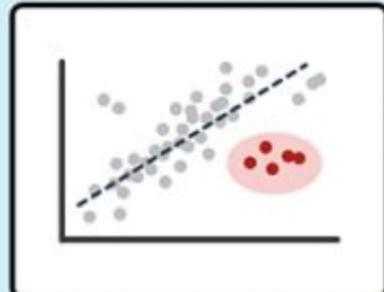
**Everything,  
everywhere,  
all-at-once  
approach**



Complexity can  
overwhelm your  
audience

Break it up  
into separate  
micro scenes

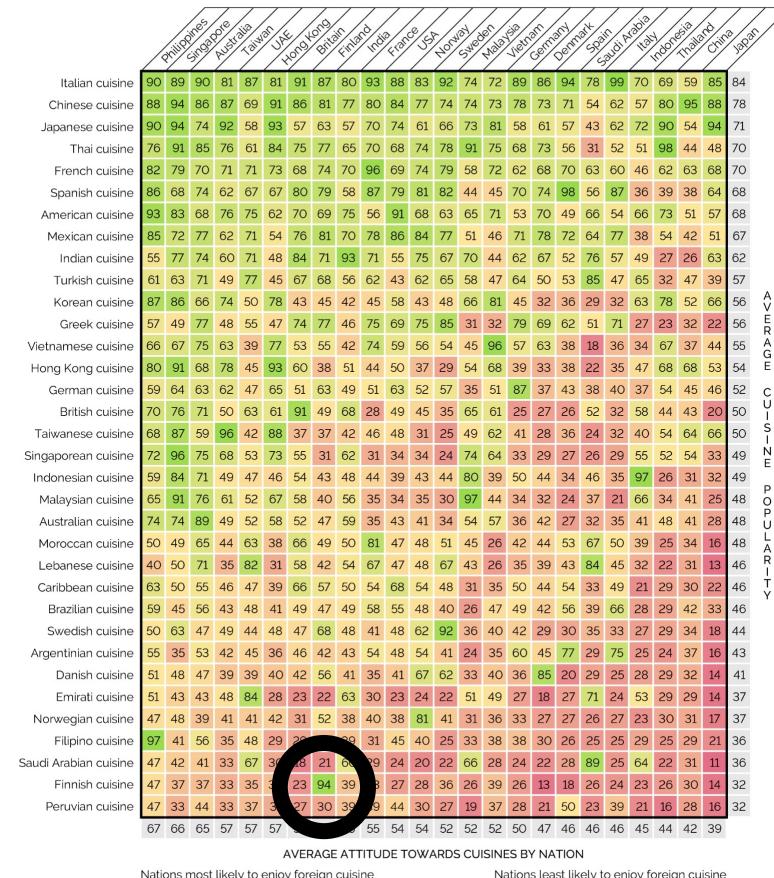
**Content  
staging  
approach**



# Make it a table!

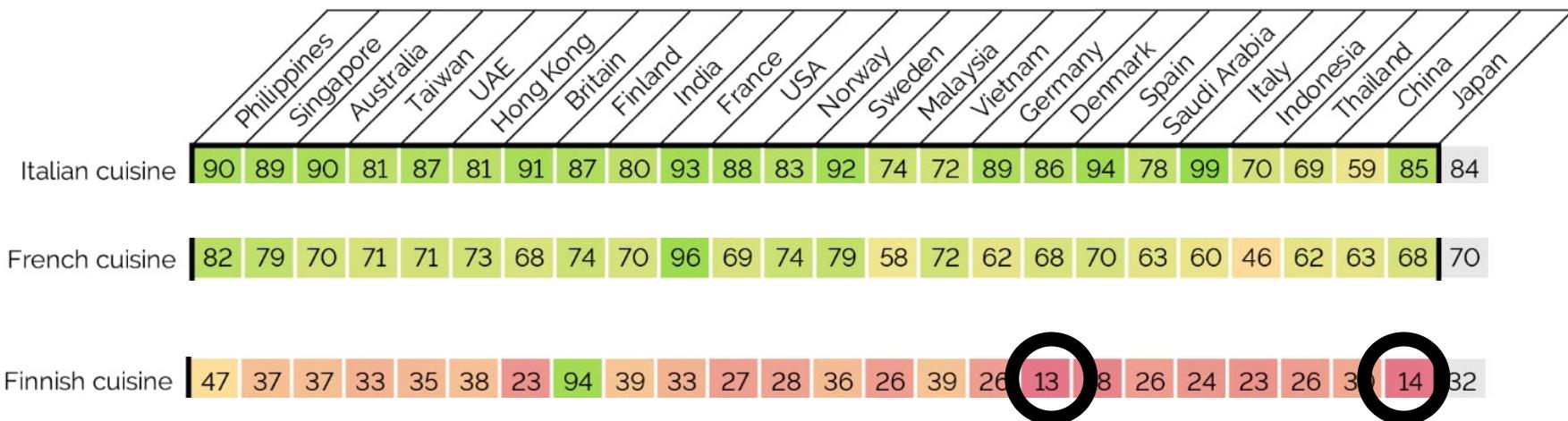
Too many messages? Too much data? **Tables better than messy chart**

**Italian, Chinese and Japanese cuisines are the world's most popular**  
 % of people who have tried that cuisine in each country that say they like it



# Make it a table!

Too many messages? Too much data? **TABLES** 😊



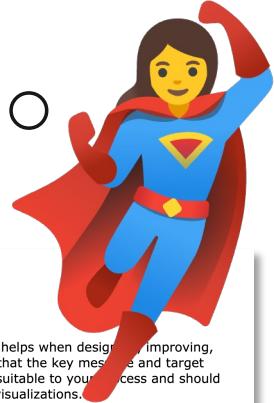
# Thank you & be a figure hero

[osf.io/preprints/osf/xgbyr](https://osf.io/preprints/osf/xgbyr)

Blog (resources etc) :  
[helenajamborwrites.netlify.app/](http://helenajamborwrites.netlify.app/)

Social: LinkedIn | Mastodon  
[helena.jambor@fhgr.ch](mailto:helena.jambor@fhgr.ch)

Helena Jambor, helena.jambor@fhgr.ch



## From zero to figure hero

Effective figures communicate scientific insights. This checklist helps when designing, improving, and reviewing figures for publications. Before starting, ensure that the key message and target audiences are defined. The checklist can be used in any order suitable to your process and should be consulted iteratively as necessary to create impactful data visualizations.

### Feedback



#### The 1-second test

- Ask 2-5 people: What do you see first?  
Evaluates if data is visible at first glance

#### Reverse feedback

- Ask: Explain to me what you see. Get feedback on chart type, text, layout, colors

#### Focus the attention

- Axes, boxes, tick marks. Remove or mute?
- Legend necessary? Direct data label or title possible?
- Gridlines. Remove or mute? Necessary for log-scales/ precise values
- Use color sparingly. Remove unnecessary colors, use grey instead of black
- Align chart elements and multi-panel figures ticks, text, titles, legends, axes, labels
- Aim for symmetry, evenly filled space, separate elements in multi-panel figures gaps/white space

### Basics



#### Choose chart

- Suitable for data and message?
- Suitable for audiences?

Consult resources for chart types:

Datavizcatalogue.com  
Python-graph-gallery.com  
Datavizproject.com

#### Simplify charts

Too much data:

- Split data across 2 charts
- Consider small multiples for many categories/ observations
- Animate information stepwise
- Uncommon chart type:
- Include help on how-to-read e.g., in title, subtitle, legend....
- Use intuitive guides e.g., direct data labels, color similarity, regions-of-interest

#### Text in charts

- Label axes
- Label tick marks, choose easy intervals  
good: 0, 5, 10; poor: 0, 7, 14
- Explain colors, marks, shades e.g. legends
- Use title, subtitle to orient readers
- Avoid abbreviations if necessary: test
- Typography: choose legible font and style regular > bold, italic

### Design



#### Layout

- Aim for horizontal text, avoid text rotation
- Label data directly. Otherwise place legend close to data
- Align text elements title best: top left/center
- Use white space to visually separate panel elements

#### Encode data with color, color schemes

- Are all colors needed?
- Color palettes for data. Sequential data: 1 color, vary saturation; multi-colors necessary: palette with homogeneous lightness, e.g., viridis (not jet/rainbow). Diverging data: -2 colors diverging from central/neutral point in white/grey; Categorical data: several colors possible.
- Use consistent color schemes
- Ensure colors are accessible. Color-blind safe, high contrast foreground/background. Double encode color information. Best, information is also visible in greyscale.
- Explain all colors e.g., legend

#### On beauty

- Align and organize with a grid
- Use white space to separate elements, but do not leave gaps in multi-panel figures
- Strive for symmetry
- Consider using icons. Sources:  
General: Nounproject, SVGrepo,  
Bio: Bioicons, Phylotic, SciDraw  
Medicine: Healthicons, SmartServier

### Specials



#### Tables

- Left align text, right align numbers, align each column header with its content.
- Use font with same width for all numbers
- Eliminate unnecessary cell borders to gain space in cells
- For overview, consider color-coding numbers heatmap

#### Image data

- Select a suitable image frame
- Add scale bar with dimensions if possible
- Explain colors scale for quantitative data
- Ensure colors are color blind safe
- Explain annotations symbols, arrows, regions