

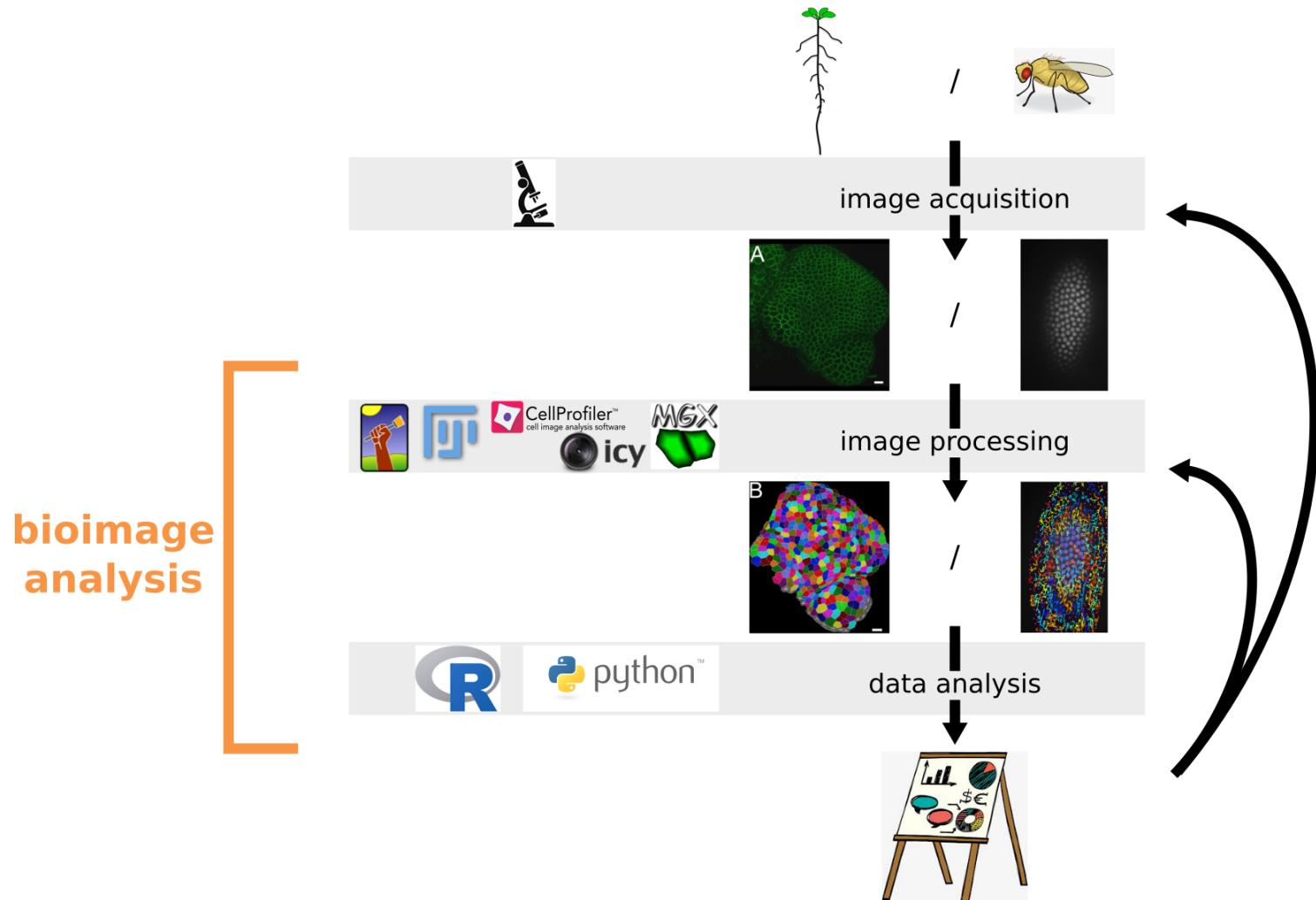
NEUBIAS - TS15

A brief introduction to data visualization

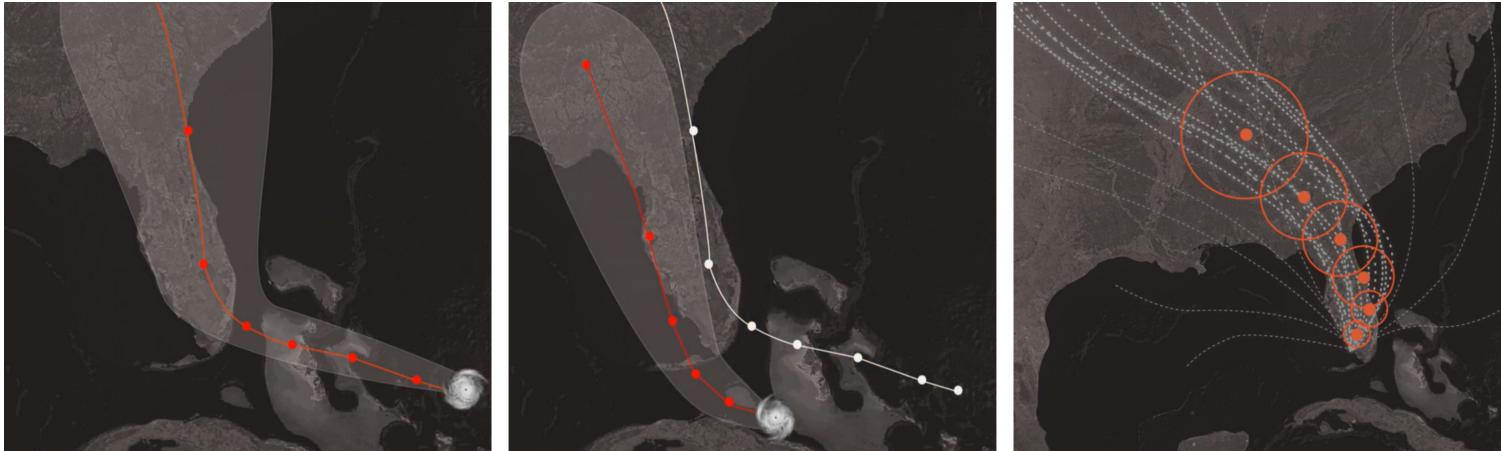
Marion Louveaux

2020-03-01

From images... to images



Graphicacy



"Scientists [...] ought to create visualizations that are as transparent as possible [...] but readers often have the unrealistic expectation that any graphic should be understandable without effort. [...] Audiences often lack knowledge of graphic symbols, grammar, and conventions" => "help increase readers's graphicacy." *Albert Cairo*

Graphicacy: the ability to understand and present information in the form of sketches, photographs, diagrams, maps, plans, charts, graphs and other non-textual formats.

<http://www.thefunctionalart.com/2019/09/explaining-visualizations-in-new-york.html>

<https://www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html>

<https://en.wikipedia.org/wiki/Graphicacy>

Graphs

- Choice of graphical representation tightly related to your question
- Presentation of information should be **clear, accessible, reliable and transparent**

Distribution

Is there a difference between these two groups (treatments, mutants...)?



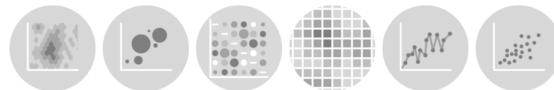
Ranking

Is there more cells in condition B than in condition A?



Correlation

Are these two variables linked?



Part of a whole

How does A behaves compared to B, C, D... ?



Flow

How are different individuals connected over space and/or time?
Lineage, cell fate changes...



Map

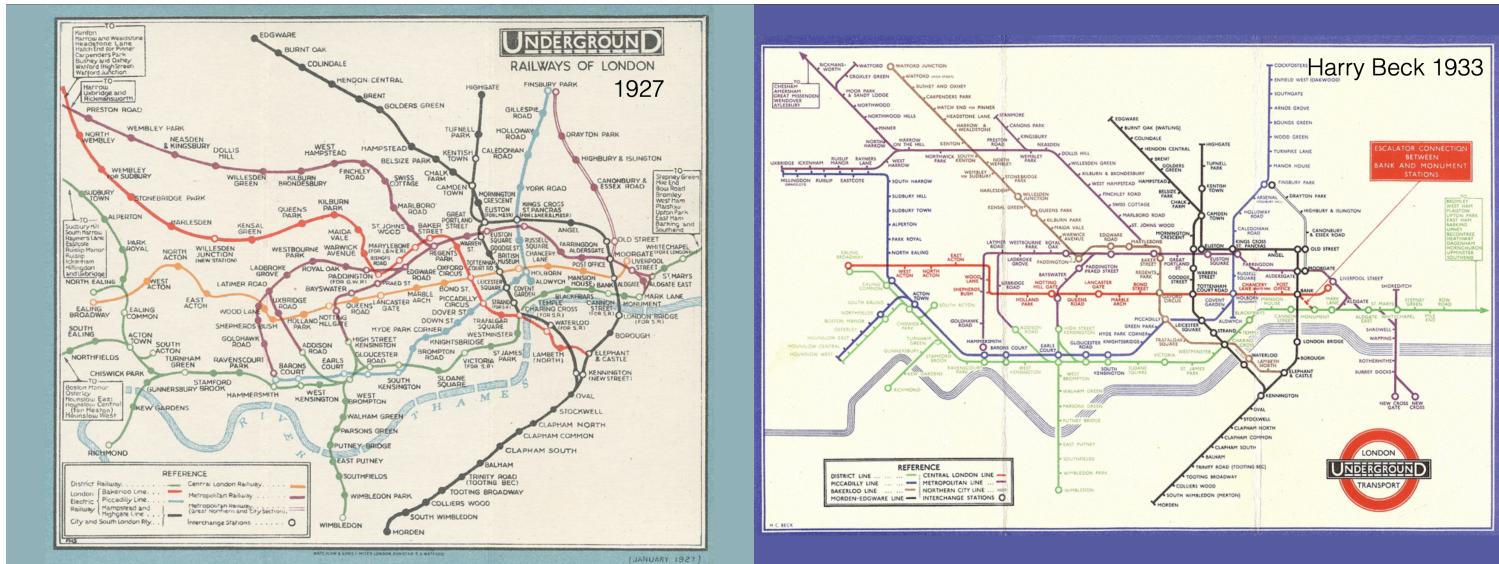
What is the spatial distribution of two fluorescent markers in a tissue?



Adapted from <https://www.r-graph-gallery.com/>
See also <https://www.data-to-viz.com/>

Clarity

Less is more



Original London subway map from 1927 takes geography into account.

<https://towardsdatascience.com/the-power-of-visualization-in-data-science-1995d56e4208>

Clarity

Less is more

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

Created by Darkhorse Analytics

www.darkhorseanalytics.com

- **Color:** Reduce colors, remove shadows and glossy effects for decoration
- **Accessory elements:** Remove or lighten background, borders, grid lines
- **Information:** Remove redundant information & prioritize info

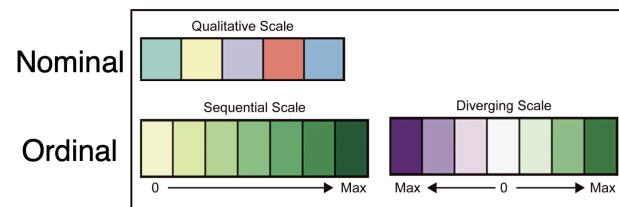
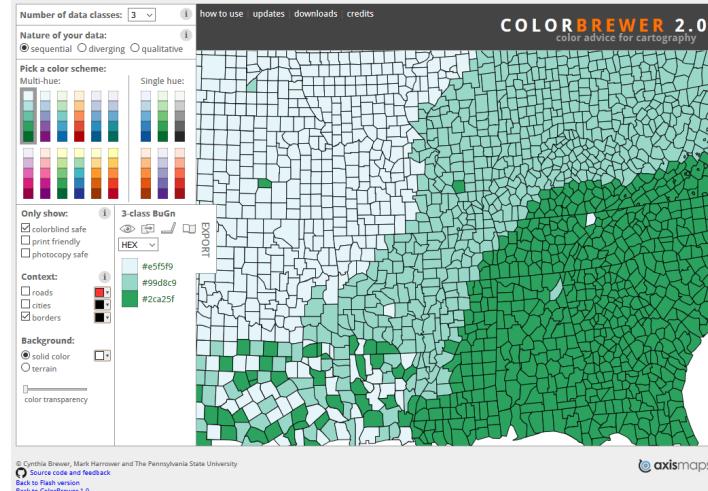
The data-ink ratio is a concept introduced by Edward Tufte in *The Visual Display of Quantitative Data* (1983).

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

Clarity

Colors matter

- No more than 6 to 7 colors
- Easily **separable**
 - Color blind compatible
 - Printer-friendly
 - Contrasted
 - Grey color is favoured
- **Meaningful**
 - Add an extra information / highlight specific elements (not redundant or random)
 - Light colors -> low values, dark colors -> high values
- **Consistent**
 - Same color for the same variable between all graphs

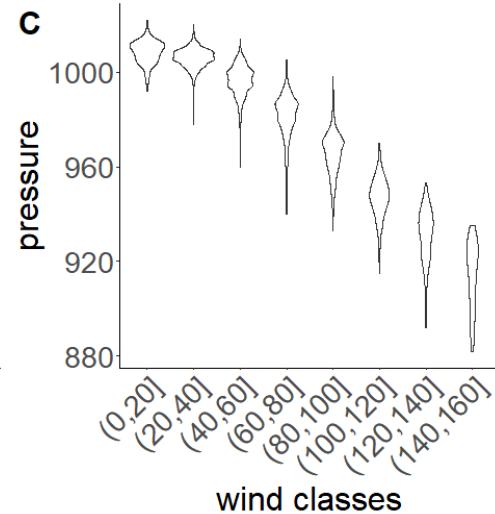
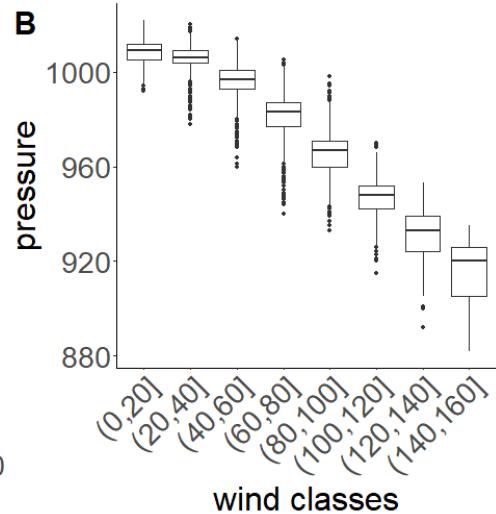
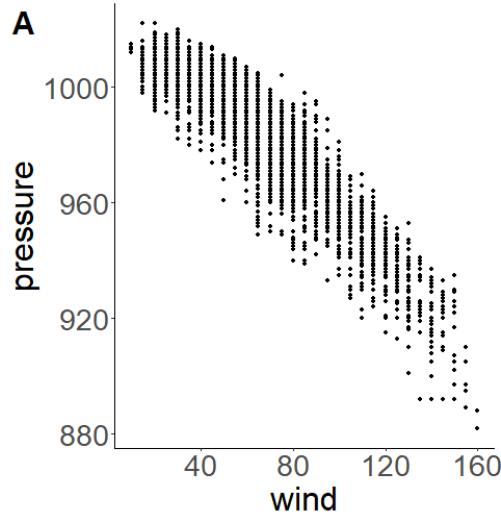


<https://blog.datawrapper.de/colors/>

ColorBrewer: <http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>

Accessibility

The right graph

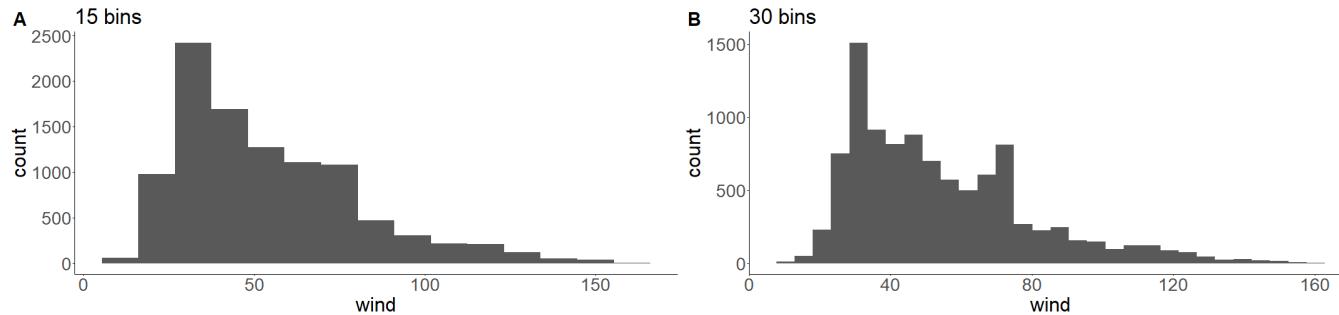


<https://www.data-to-viz.com/caveat/boxplot.html>
<http://www.storytellingwithdata.com/blog?offset=1569328920602>

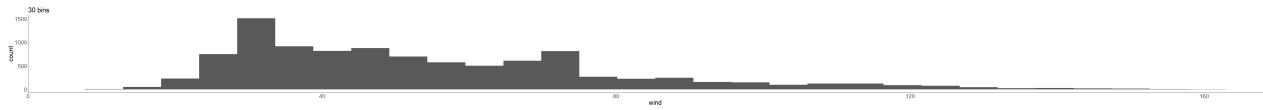
Accessibility

The right scale

- Histogram bin size



- Aspect ratio



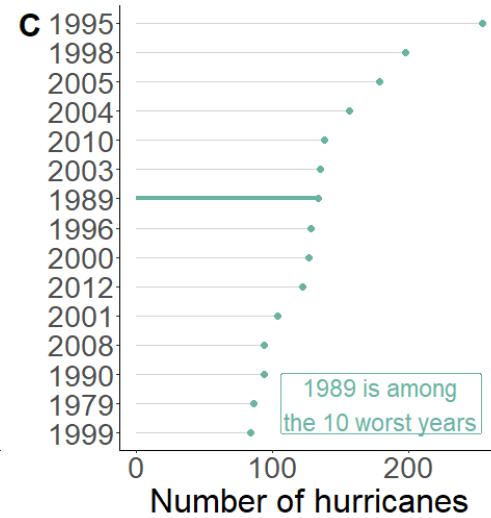
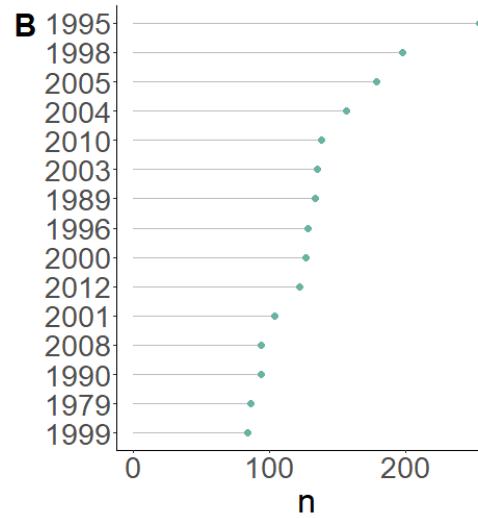
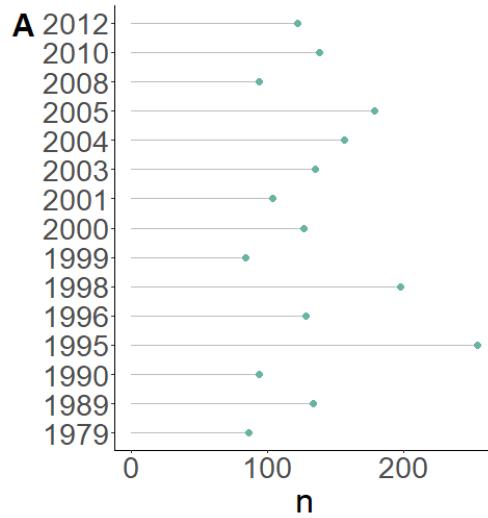
https://www.data-to-viz.com/caveat/bin_size.html

https://www.data-to-viz.com/caveat/aspect_ratio.html

Accessibility

A tidy graph

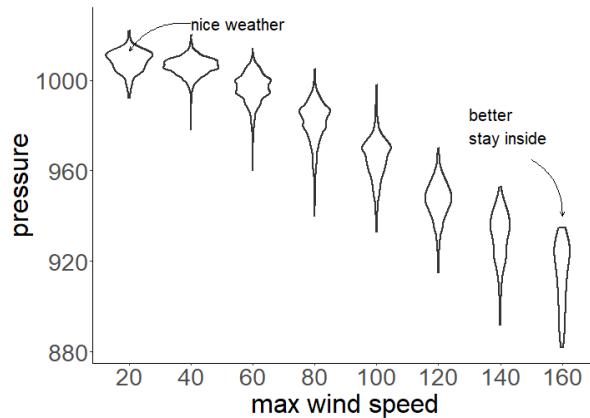
- Order data
- Highlight some elements of the graph



<https://www.data-to-viz.com/caveat/spaghetti.html>
<https://www.data-to-viz.com/caveat/overplotting.html>
https://www.data-to-viz.com/caveat/order_data.html
https://www.data-to-viz.com/caveat/connect_your_dot.html
https://www.data-to-viz.com/caveat/grouped_bar.html

Accessibility Annotations

- Annotate the graph
- Shorten labels
- If necessary, switch axes for categorical variables

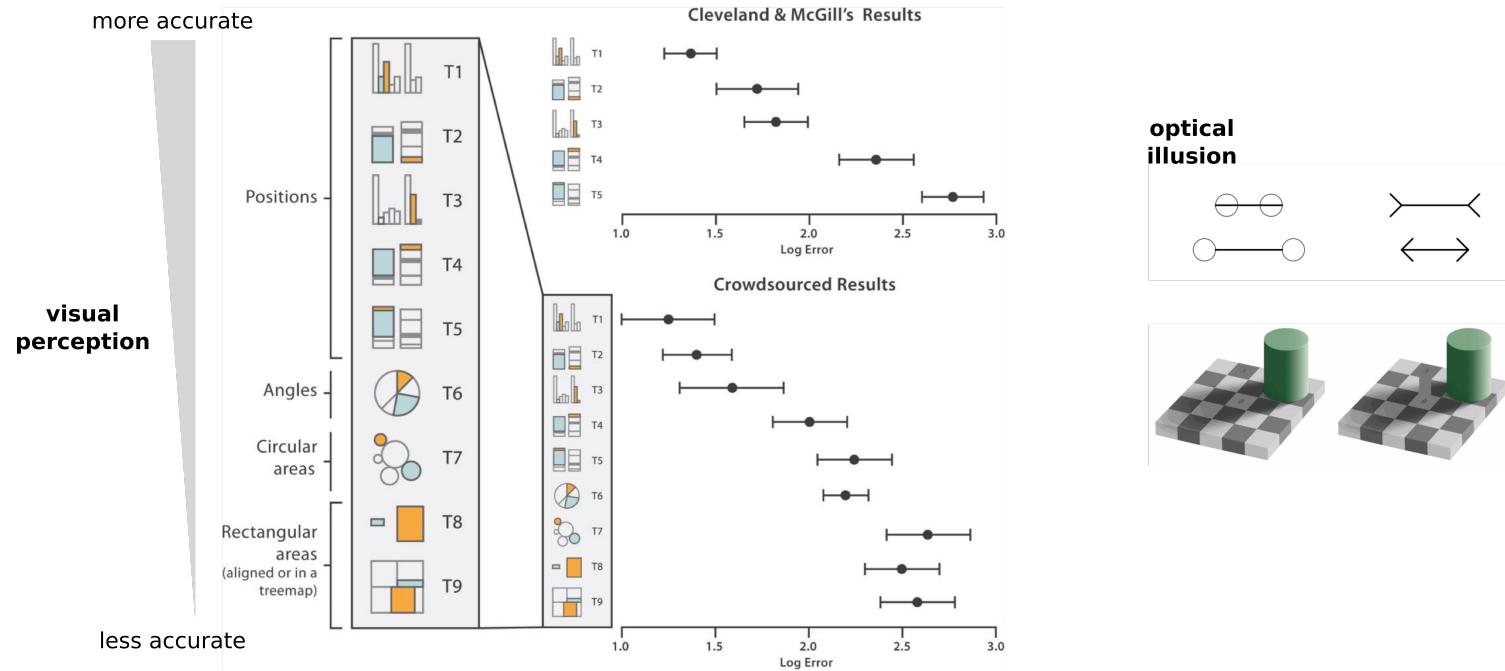


<https://www.data-to-viz.com/caveat/annotation.html>
https://www.data-to-viz.com/caveat/hard_label.html

Reliability

No mental arithmetic

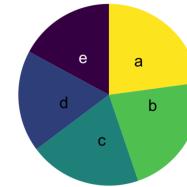
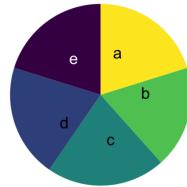
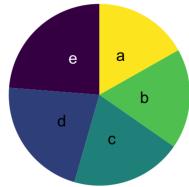
- Visual perception: better for length than areas and volumes
- Optical illusions: context matters



https://hackmd.io/zS8OW22LR3mGArfX_9gi2A
<https://homepage.divms.uiowa.edu/~luke/classes/STAT4580/percep.html>

Reliability

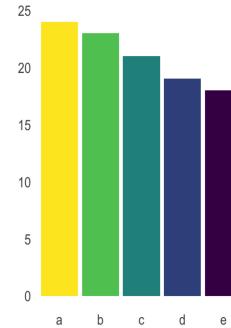
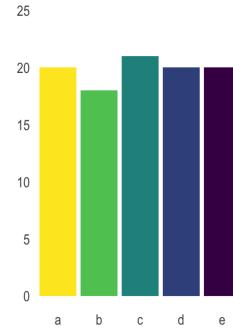
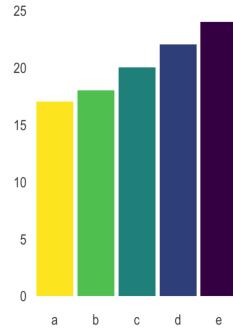
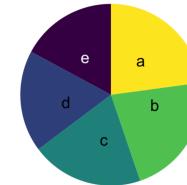
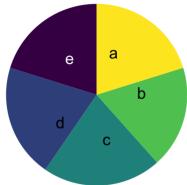
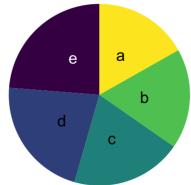
No mental arithmetic



<https://www.data-to-viz.com/caveat/pie.html>
https://www.data-to-viz.com/caveat/radius_or_area.html
https://www.data-to-viz.com/caveat/area_hard.html

Reliability

No mental arithmetic



<https://www.data-to-viz.com/caveat/pie.html>

https://www.data-to-viz.com/caveat/radius_or_area.html

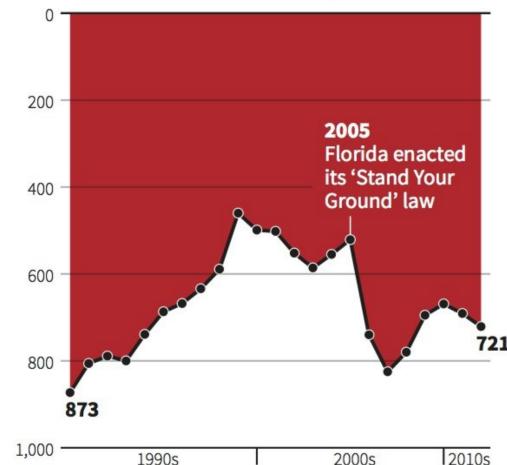
https://www.data-to-viz.com/caveat/area_hard.html

Reliability

Don't be counter-intuitive

Gun deaths in Florida

Number of murders committed using firearms



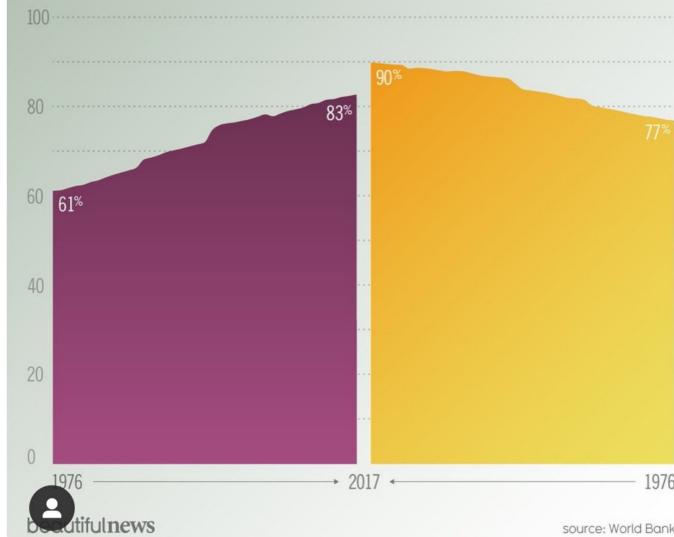
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Female and Male Literacy Rates Have Almost Equalised

% who can read & write globally



source: World Bank



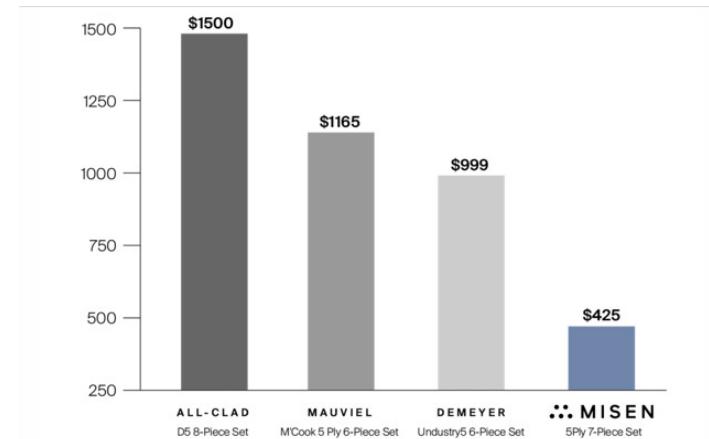
beautifulnews

https://www.data-to-viz.com/caveat/counter_intuitive.html
https://junkcharts.typepad.com/junk_charts/simplicity/

Reliability

Don't use misleading representations

- Avoid distorted graphs and 3D
- Don't cut y axis on barplots

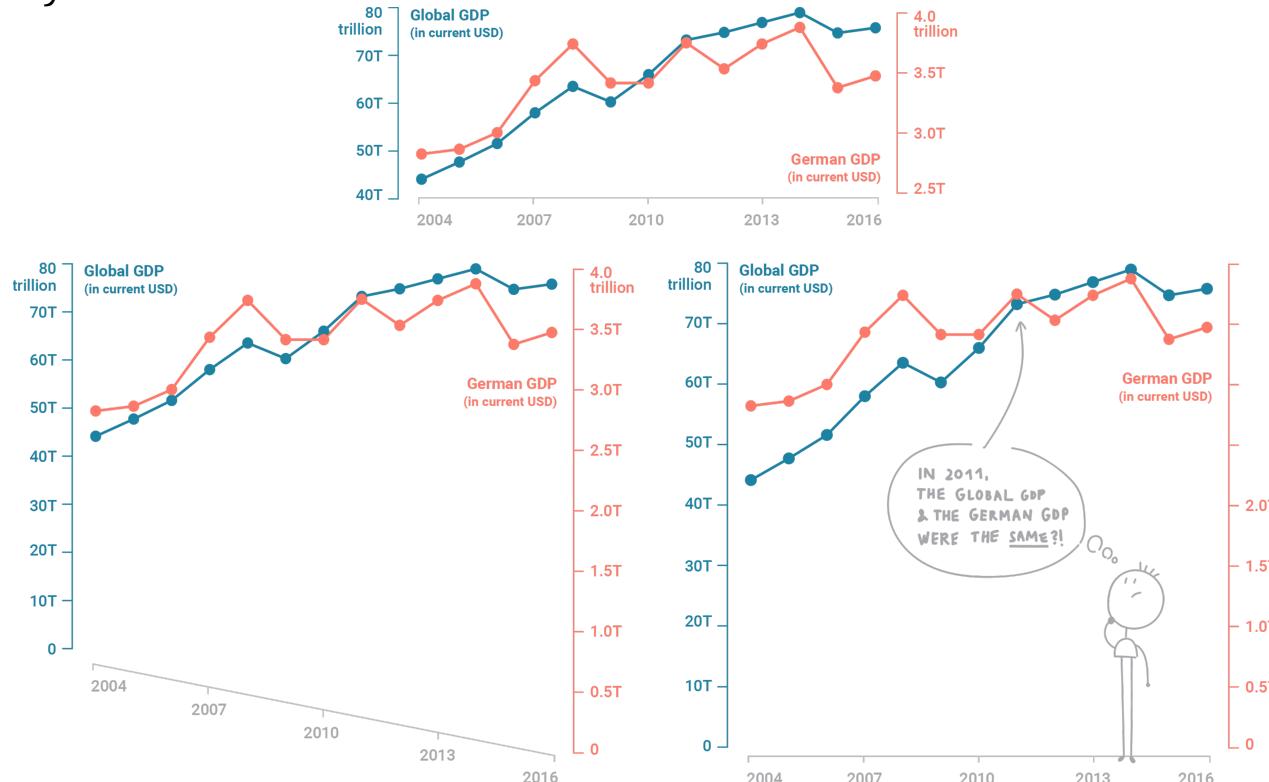


https://www.data-to-viz.com/caveat/circular_barplot_accordeon.html
<https://viz.wtf/post/185041906238/not-as-cheap-as-youd-think>
https://www.data-to-viz.com/caveat/cut_y_axis.html
<https://www.data-to-viz.com/caveat/3d.html>

Reliability

Don't use misleading representations

- Dual y axis

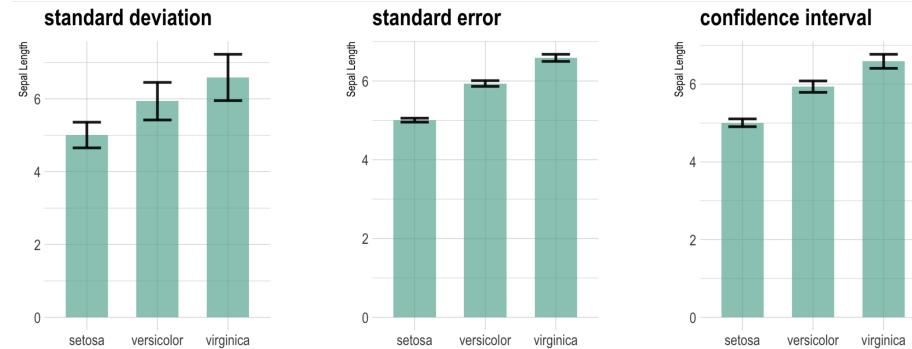
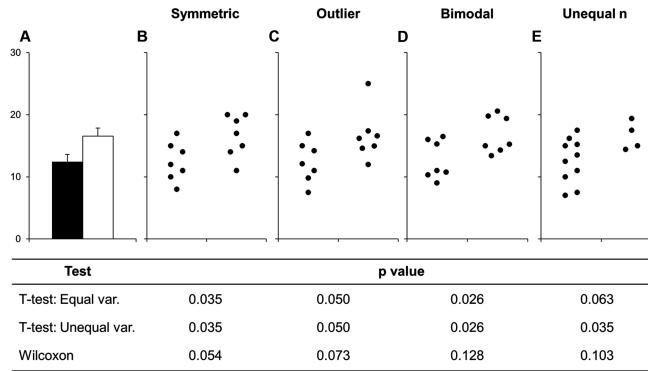


https://junkcharts.typepad.com/junk_charts/2019/11/graph-literacy-in-a-sense.html
<https://blog.datawrapper.de/dualaxis/>
<https://edaemeklioglu.wordpress.com/2015/02/15/wtf-visualizations/>

Reliability

Don't use misleading representations

- Barbarplot



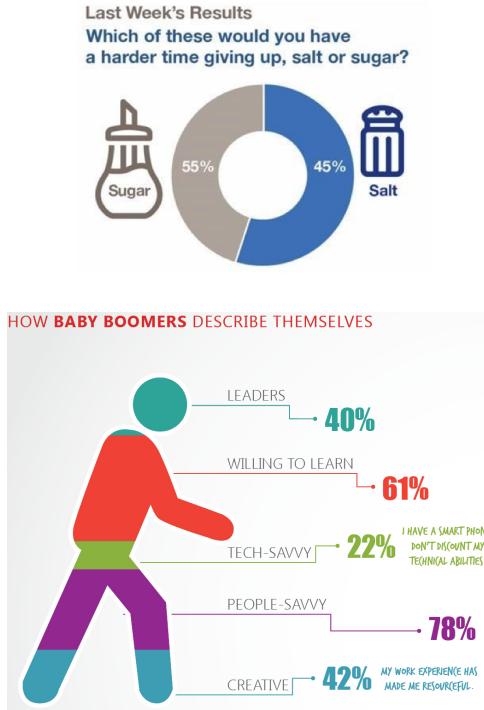
https://www.data-to-viz.com/caveat/error_bar.html

<https://rtask.thinkr.fr/data-visualisation-pitfalls-how-to-avoid-barbarplots/>

Weissgerber et al. 2015

Transparency

Automate graph production, avoid introducing mistakes



No manual manipulation of raw data or graphical representations!!!

https://www.data-to-viz.com/caveat/calculation_error.html
<https://viz.wtf/post/169392205697/fewer-people-would-have-a-hard-time-giving-up>
<https://viz.wtf/post/59697293967/hes-243-baby-boomer>

Transparency

Share your code

Graphs should be reproducible from raw data + your code

- Upload your code on GitHub or GitLab
- Share the raw data
- Reference both with Zenodo (DOI)



<https://github.com/>
<https://about.gitlab.com/>
<https://zenodo.org/>

Acknowledgements

- **Local organisers:** Fabrice Cordelières and Florian Levet.
- **NEUBIAS Vice-Chair and TS15 co-organiser:** Kota Miura.
- **NEUBIAS Chair:** Julien Colombelli.