A brief introduction to data visualisation

Marion Louveaux

2020-03-02

Data Viz

Why it matters

https://junkcharts.typepad.com/junk_charts/2019/09/as-dorian-confounds-meteorologists-we-keep-our-minds-clear-on-hurricane-graphics-and-discover-correl.html

Graph families

Choice of graphical representation matters
It is tightly related to your question. What is your message? To which question do you answer?
It also affects your data exploration (you might miss some important info in your data with the wrong data representation).
https://www.r-graph-gallery.com/

https://www.data-to-viz.com/#explore

Clarity

=> simplify the graph as much as possible

Clarity

Less is more

- Reduce colors, remove shadows and glossy effects for decoration
- Remove background, borders
- Lighten or remove grid lines
- Remove redundant information
- Prioritize info : font and text size, grey VS black text

Clarity

Colors matter

https://blog.datawrapper.de/colors/ ColorBrewer: http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3

- No more than 6 to 7 colors
- Color palette matter Color blind compatible and printer-friendly colors

Pay attention to contrast and distance between colors in the graph

- Add color to add an extra information or highlight specific elements of your graph

Color is distracting when redundant with another graph element or when random. Grey color is the most important color in DataViz.

- Choose consistent colors between graphs in a report Use the same color for the same variable all along your report
- Intuitivity Light colors -> low values, dark colors -> high values Are your data sequential = ordered from low to high diverging = ordered from low to high, mid-range value matters (for instance, mid-range = average value) qualitative = non ordered data (categories, groups...)

Accessibility

=> no visualisation effort needed to read the graph

Accessibility

The right graph & the right scale

- The right graph https://www.data-to-viz.com/caveat/boxplot.html boxplots summarize a lot the info, they are good when distribution is known and easy to understand when distrib is gaussian, otherwise, they can hide the distribution of data When looking at distributions, violin plots or ridges when there is many distributions give more info xs

http://www.storytellingwithdata.com/blog?offset=1569328920602 barplots for numerical ~ categorical variables line plots for numerical ~ numerical

- The right scale

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

Play with the aspect ratio of the graph to highlight the best the trend in your data. In histograms, bin size also matters.

Accessibility

A tidy graph

- Order data https://www.data-to-viz.com/caveat/order_data.html
- Highlight some elements of the graph https://www.data-to-viz.com/caveat/spaghetti.html https://www.data-to-viz.com/caveat/overplotting.html

If one group matters more than the others, or has a different behavior than the others, depends on your message/question

- Connect dots and group bars https://www.data-to-viz.com/caveat/connect_your_dot.html For variable in x that are continuous

https://www.data-to-viz.com/caveat/grouped_bar.html for groups and subgroups

Accessibility

Annotations

- Annotate the graph https://www.data-to-viz.com/caveat/annotation.html and legend Add arrows, pay attention to the legend, depends on your message

https://www.data-to-viz.com/caveat/hard_label.html shorten labels as much as possible, switch axes if necessary for categorical variables

Reliability

=> No misleading representation

Reliability

No mental arithmetic

- stacked representations require to recalculate raw data https://www.data-to-viz.com/caveat/multi_distribution.html https://www.data-to-viz.com/caveat/stacking.html
- graphs using area or volumes are hard to understand https://www.data-to-viz.com/caveat/radius_or_area.html

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

about length VS area https://hackmd.io/zS8OW22LR3mGArfX_9gi2A https://hackmd.io/zS8OW22LR3mGArfX_9gi2A

Reliability

No counter intuitive or confusing representations

- Counter intuitive visualization https://www.data-to-viz.com/caveat/counter_intuitive.html

https://junkcharts.typepad.com/junk_charts/2019/11/graph-literacy-in-a-sense.html

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

https://blog.datawrapper.de/dualaxis/

Reliability

Avoid 3D and distorted graphs

- Distorted graphs https://www.data-to-viz.com/caveat/circular_barplot_accordeon.html

https://www.data-to-viz.com/caveat/cut_y_axis.html Distort the increase over time

- 3D
  https://www.data-to-viz.com/caveat/3d.html

Transparency

Automate graph production

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

No manipulation later on Illustrator!!!

Transparency

Share your code

Provide the code you used to create your graphs.
Graphs should be reproducible from raw data.
No manual manipulation of raw data.

Acknowledgements