

# SCIENTIFIC VISUALIZATION

---

Nicolas P. Rougier – Nicolas.Rougier@inria.fr

Inria - University of Bordeaux

ASPP - August 2021

# ON THE IMPORTANCE OF VISION

*... about 50 percent of the cerebral cortex of primates is devoted exclusively to visual processing, and the estimated territory for humans is nearly comparable.*

The MIT Encyclopedia of the Cognitive Sciences

# ANSCOMBE'S QUARTET

What is common to these data sets?

Mean of x	9
Sample variance of x	11
Mean of y	7.5
Sample variance of y	4.12
Linear regression	$y=3.00+0.500*x$
R squared	0.666
p value	0.0021

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

# ANScombe'S QUARTET

What is common to these data sets?

Mean of x 9

Sample variance of x 11

Mean of y 7.5

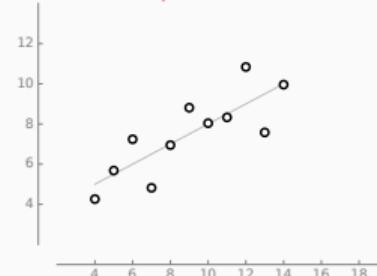
Sample variance of y 4.12

Linear regression  $y=3.00+0.500*x$

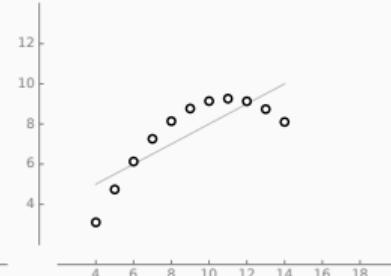
R squared 0.666

p value 0.0021

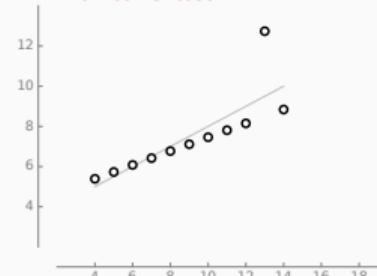
What we expect...



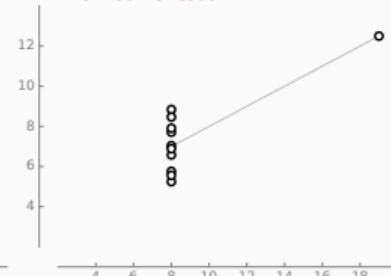
The non-linear case



The Y outlier case



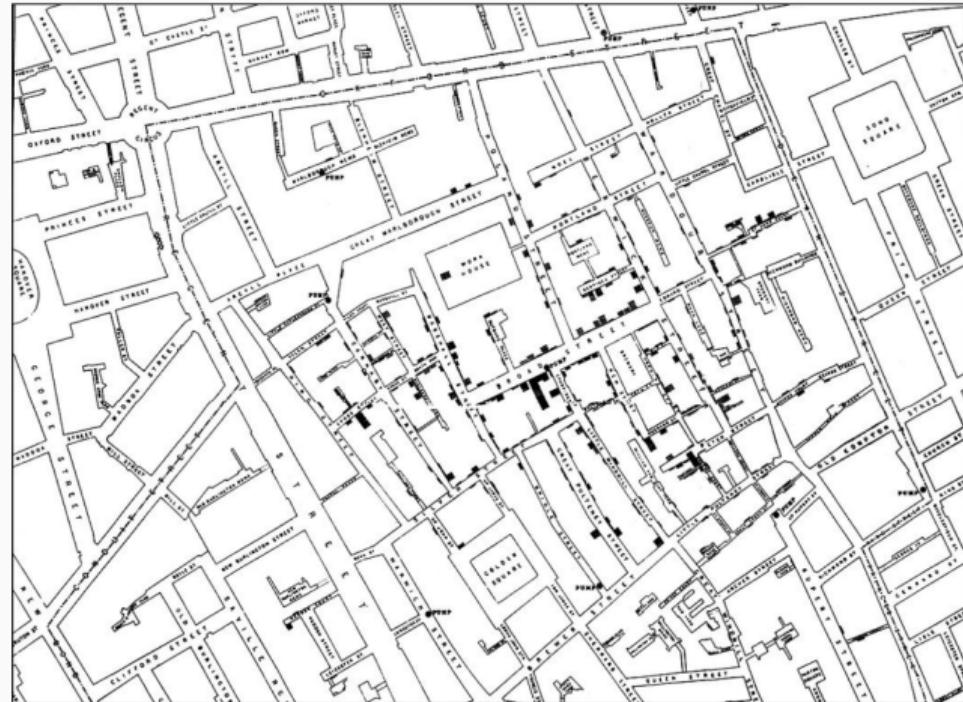
The X outlier case



A computer should make both calculations and graphs – *Francis Anscombe (1918-2001)*

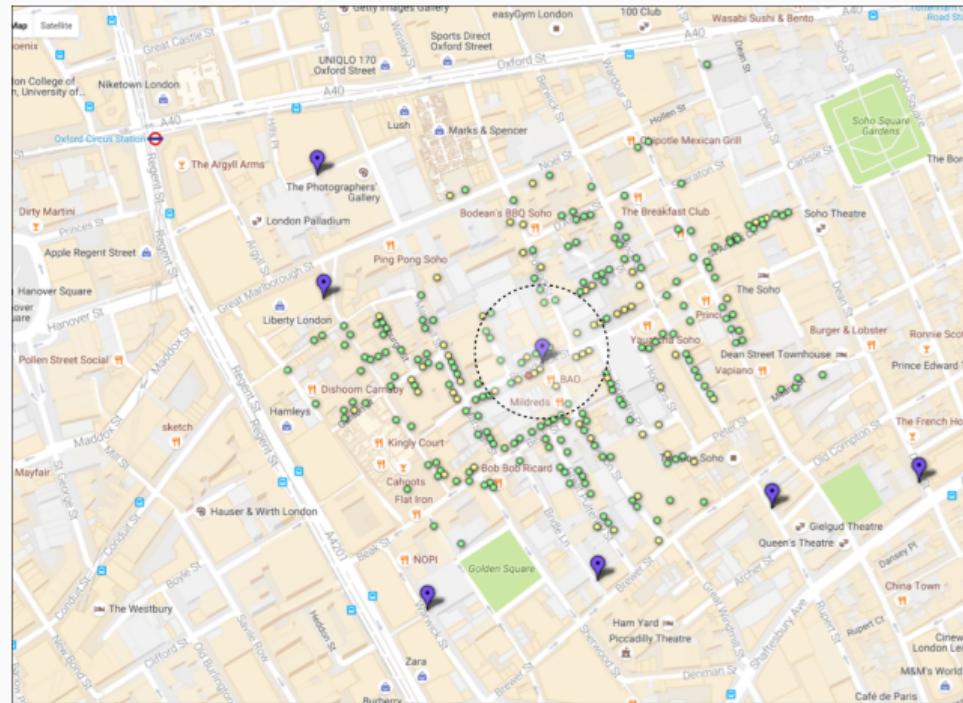
# CHOLERA EPIDEMIC, LONDON, 1854

John Snow (1813-1858) is considered one of the fathers of modern epidemiology, in part because of his work in tracing the source of a cholera outbreak in Soho, London, in 1854.



# CHOLERA EPIDEMIC, LONDON, 1854

John Snow (1813-1858) is considered one of the fathers of modern epidemiology, in part because of his work in tracing the source of a cholera outbreak in Soho, London, in 1854.

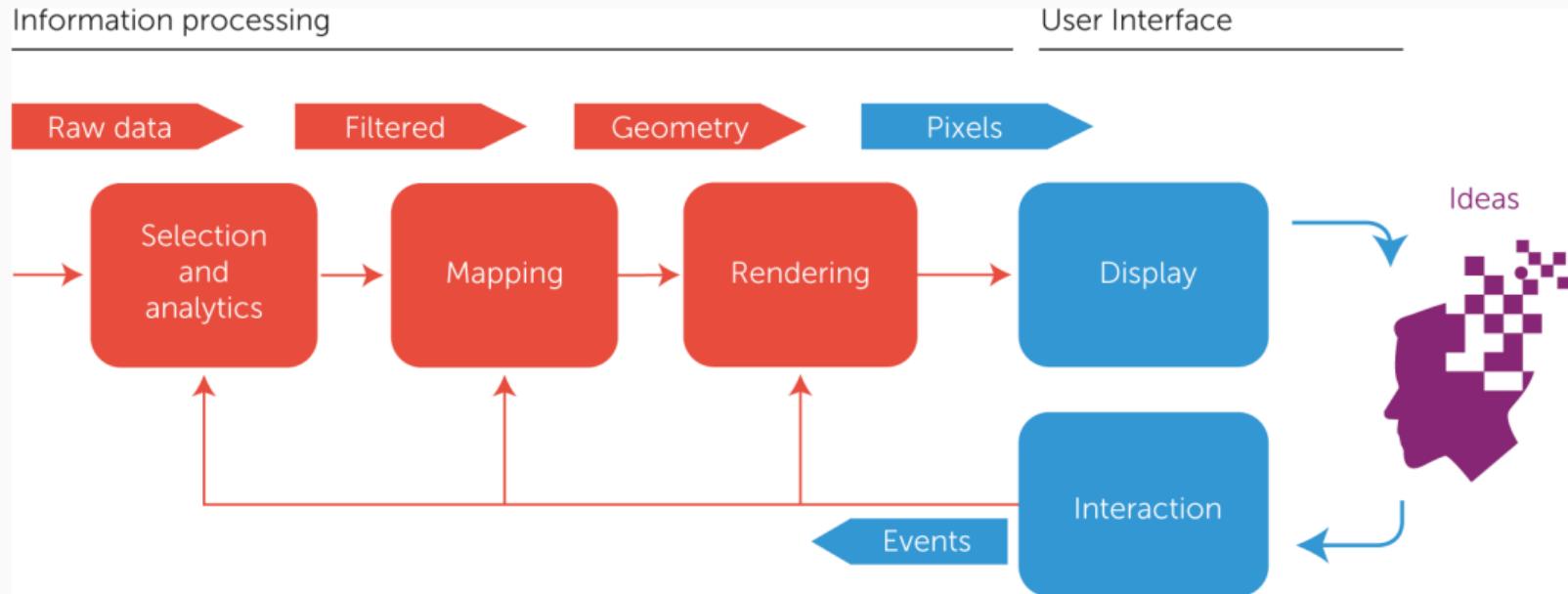


# WHAT IS DATA VISUALISATION?

*Visualisation is a method of computing. It transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualisation offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights.*

Visualisation in Scientific Computing, NSF report, 1987

# THE VISUALIZATION PIPELINE



From Scalable Real-Time Visualization Using the Cloud, Holliman & Watson, 2015.

# QUANTITATIVE VS QUALITATIVE DATA

**Quantitative** (values or observations that can be measured)

- Continuous (e.g. temperature)
- Discrete (e.g. number of inhabitants)

**Qualitative** (values or observations that can be sorted into groups or categories)

- Nominal (e.g. nationality)
- Ordinal (e.g. months)
- Interval (e.g. age groups)

# GRAPHICAL ELEMENTS

A scientific figure can be fully described by a set of graphic primitives with different attributes:

- Points, markers, lines, areas, ...
- Position, color, shape, size, orientation, curvature, ...
- Helpers, text, axis, ticks, ...
- Interaction, animation, ...

Questions is thus how to organize and link them to the underlying data.

# PRINCIPLES OF VISUAL PERCEPTION

Objects that are close together will be grouped together visually.

## PROXIMITY



## CLOSURE

The Brain is good at filling in gaps to create a whole.

## CONTINUATION

A line will always appear to continue travelling in the same way.

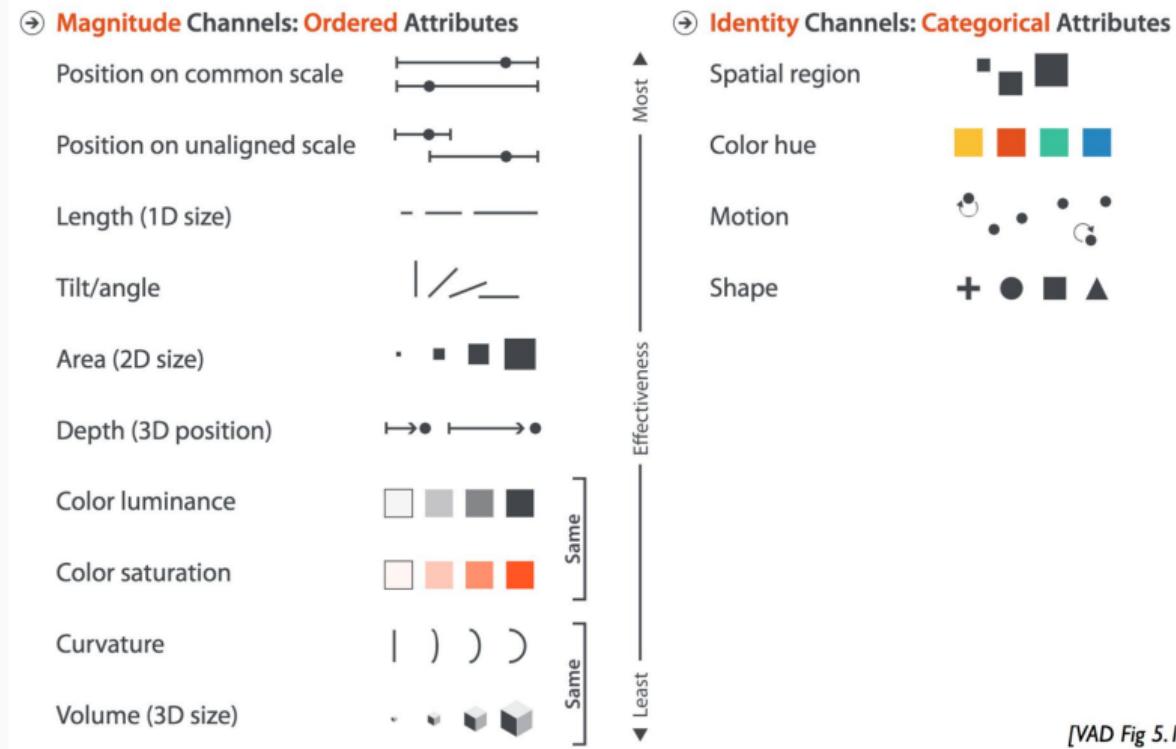
Two items that share attributes will be visually grouped together.

## SIMILARITY

## FIGURE & GROUND

Sometimes, the blank space is just as important as the filled space.

# VISUALIZATION ANALYSIS AND DESIGN (T. MUNZNER)



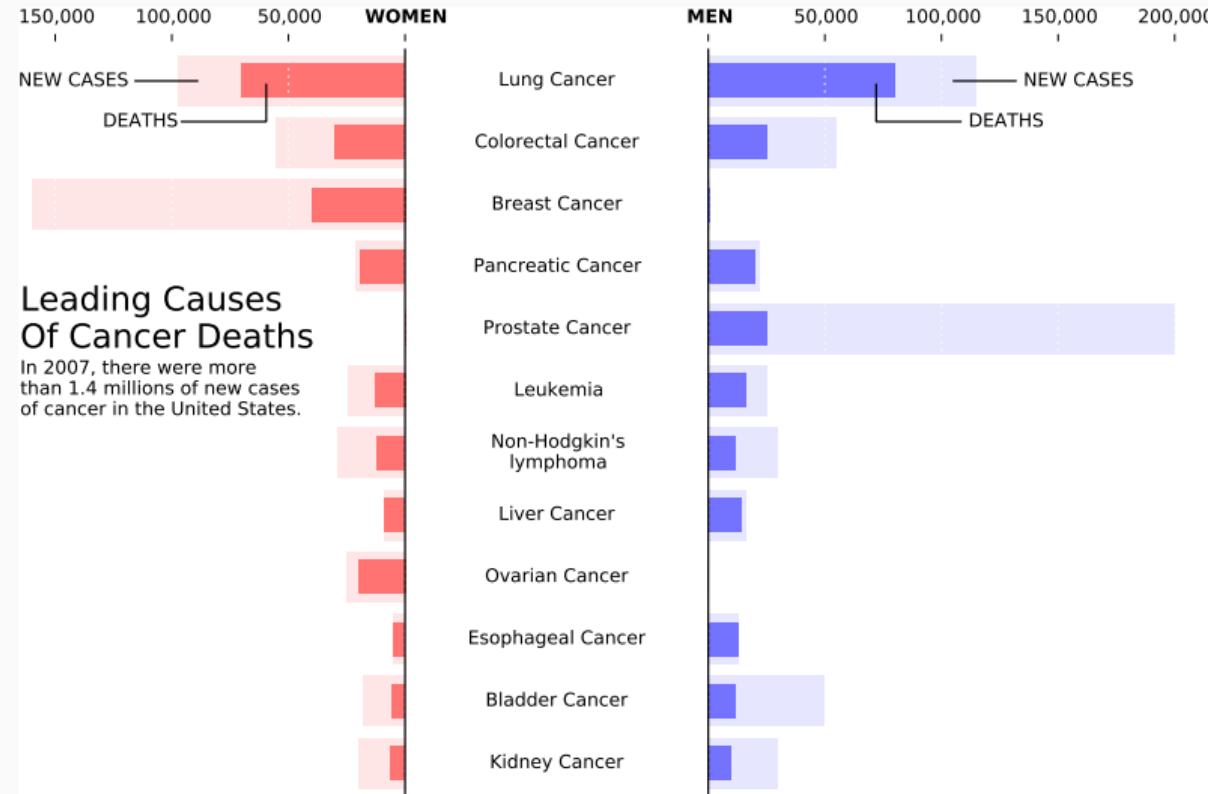
# DATA VISUALIZATION CATALOGUE (S. RECEBBA)



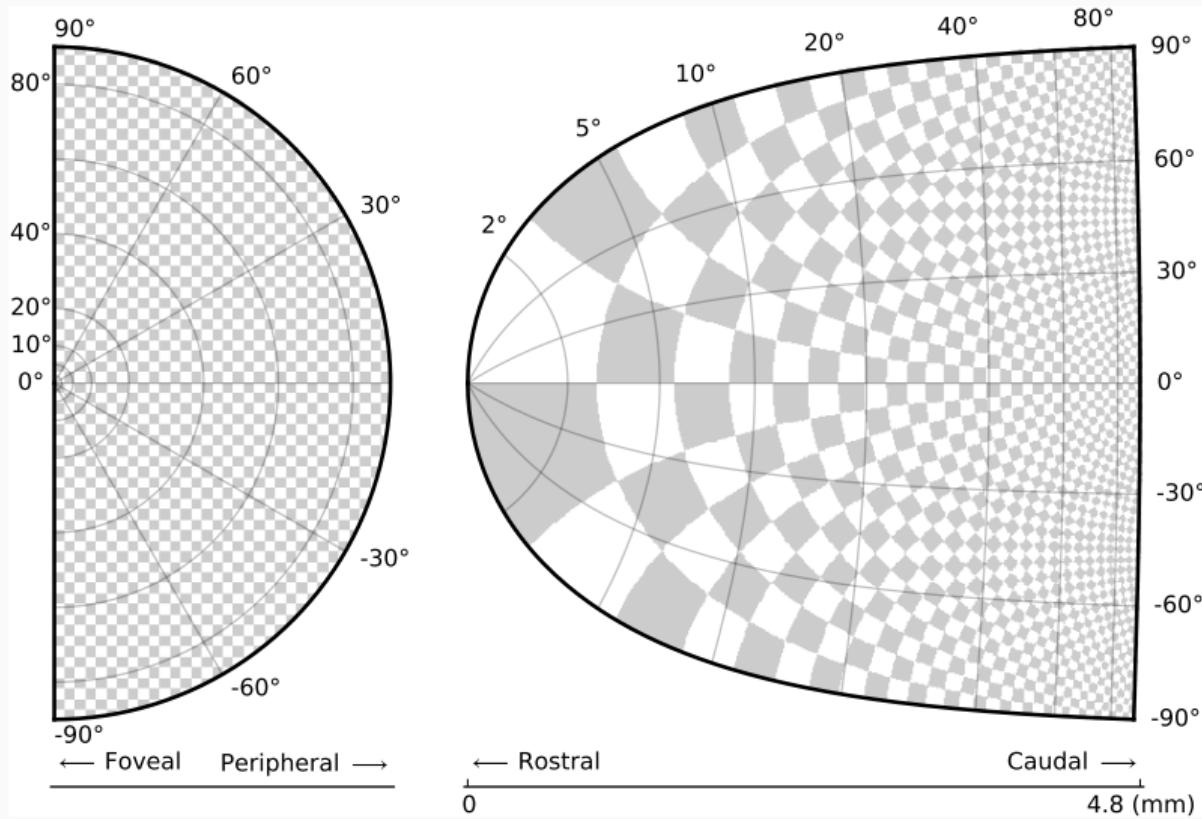
## Ten simple rules for better figures

---

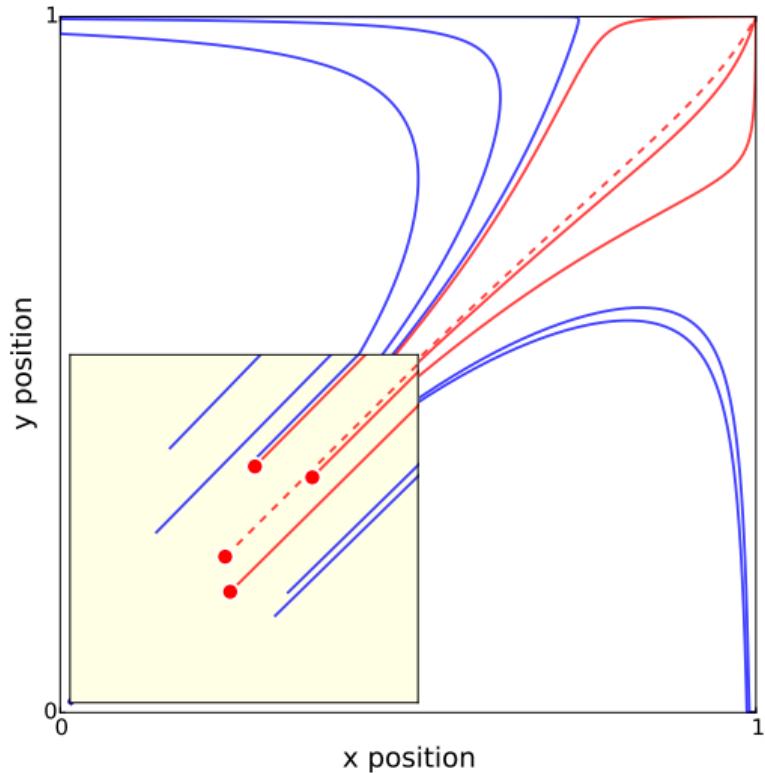
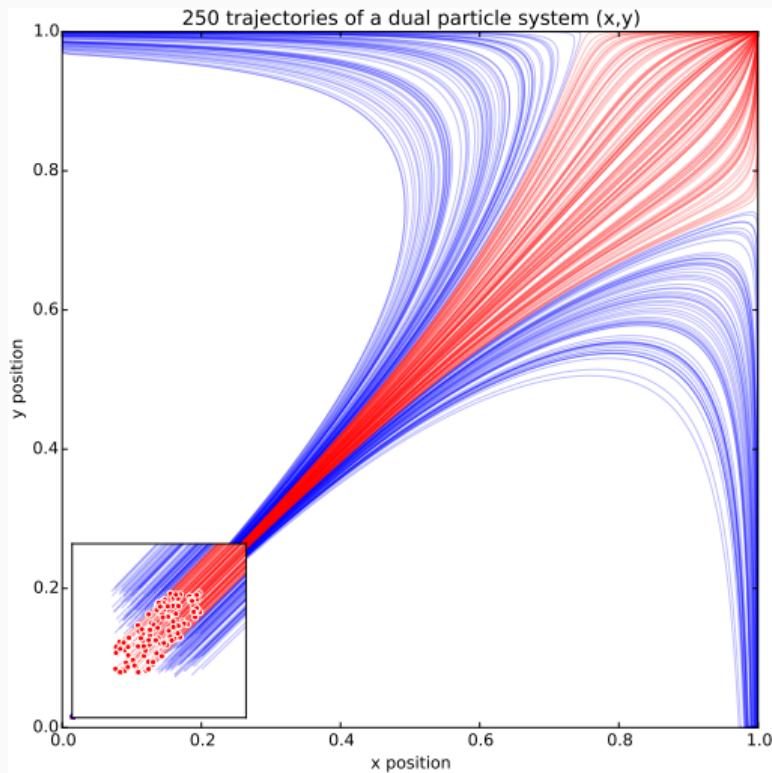
# RULE 1: KNOW YOUR AUDIENCE



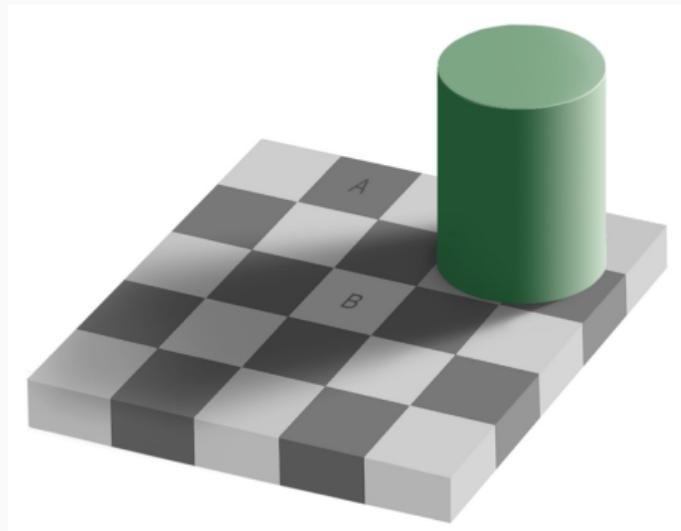
## RULE 2: IDENTIFY YOUR MESSAGE



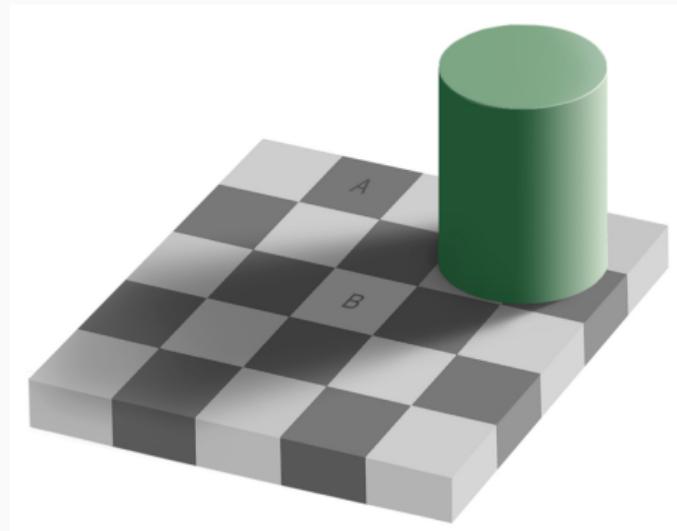
## RULE 3: ADAPT THE FIGURE



## RULE 4: CAPTIONS ARE NOT OPTIONAL

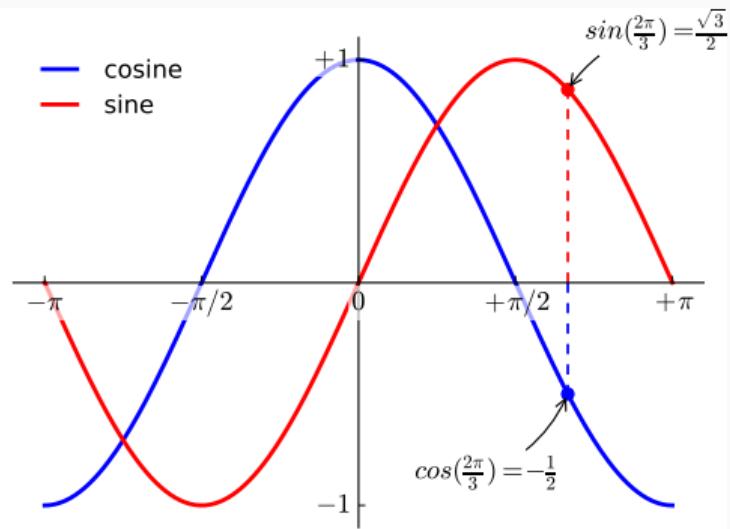
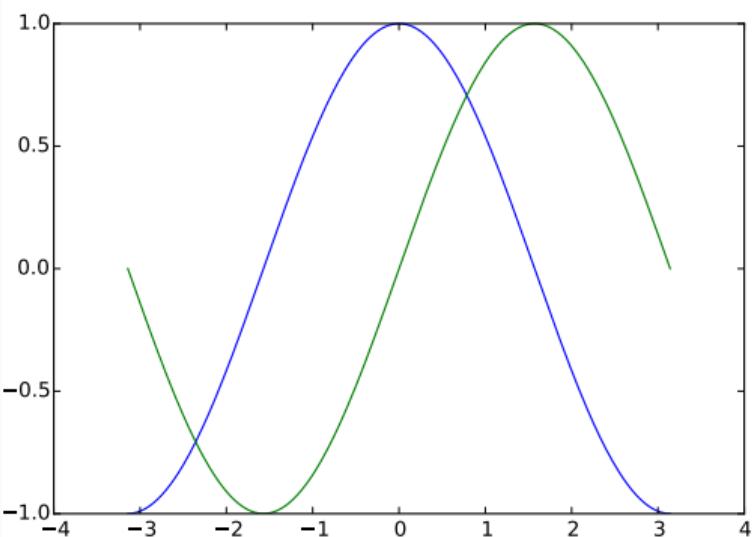


Optical illusion

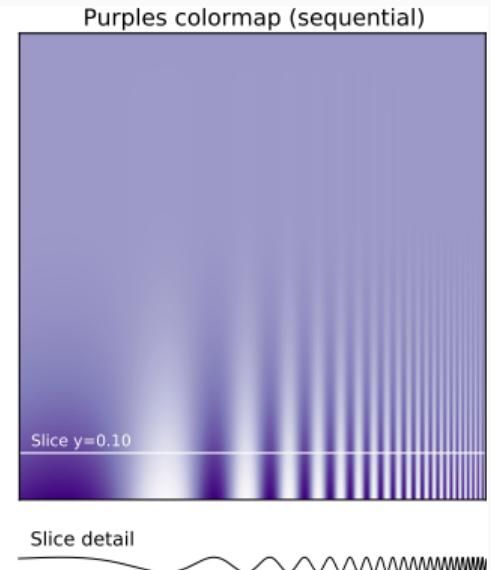
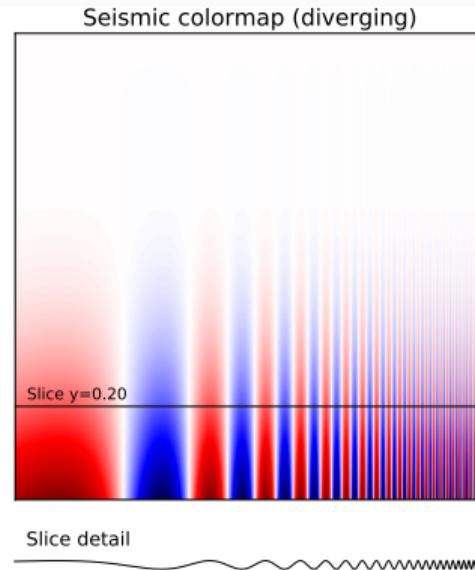
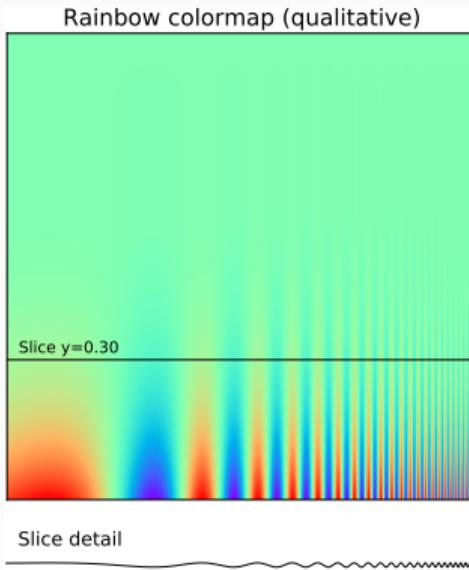


The A and B patches are actually the same color even though we perceived them at being different color.

## RULE 5: DO NOT TRUST THE DEFAULTS



# RULE 6: USE COLOR EFFECTIVELY

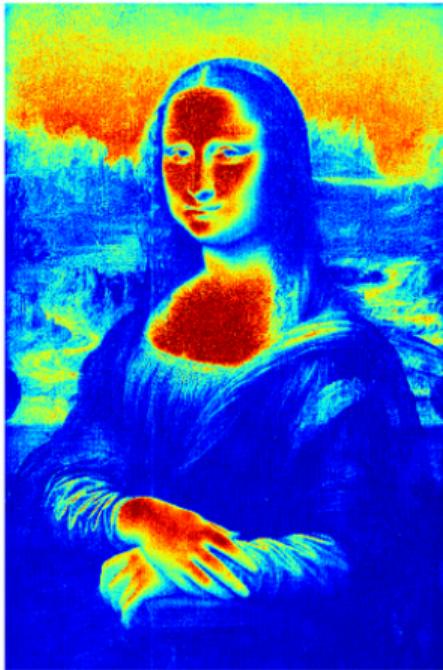


## RULE 6 BIS: ABOVE ALL, NO JET. EVER.

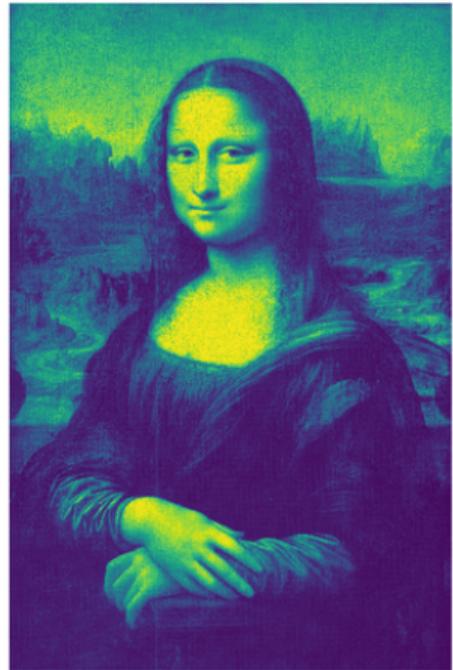
Colour



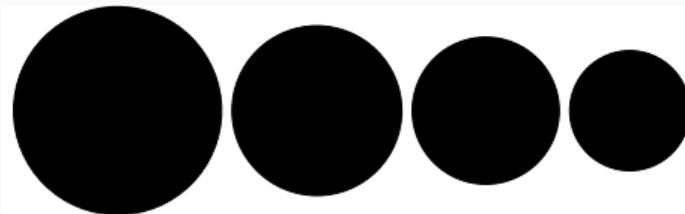
Jet



Viridis

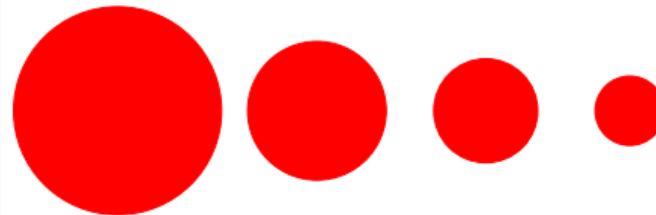


## RULE 7: DO NOT MISLEAD THE READER



Relative size using disc area

Relative size using disc radius

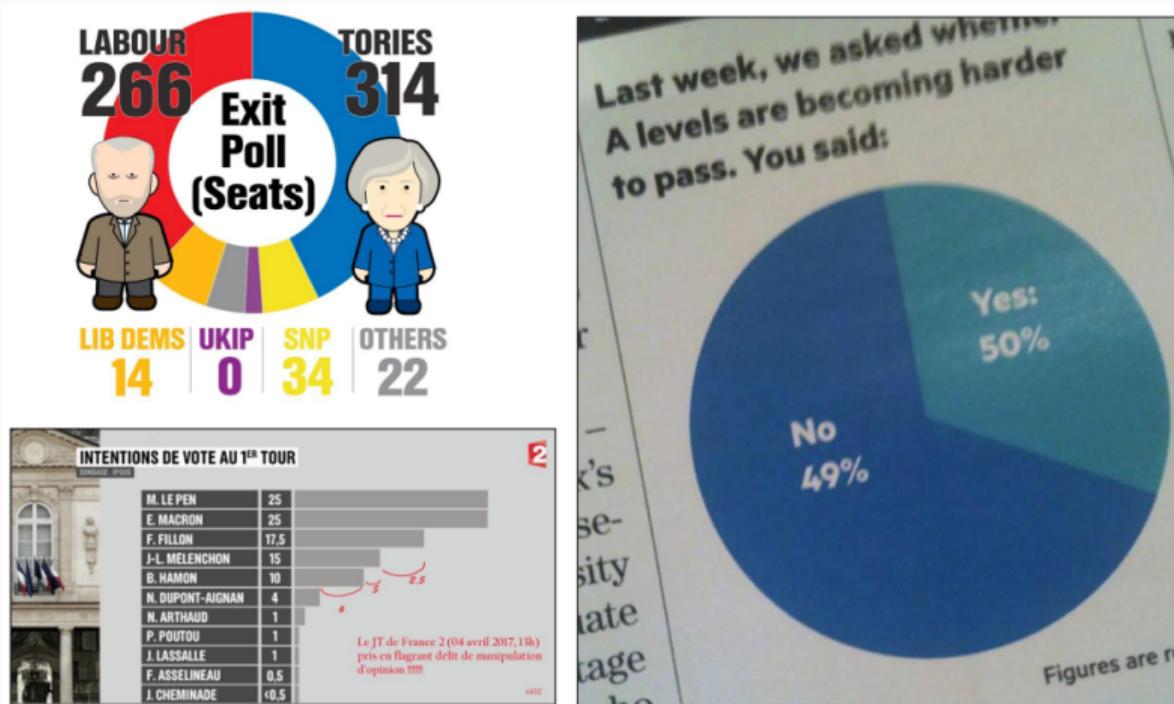


Relative size using full range

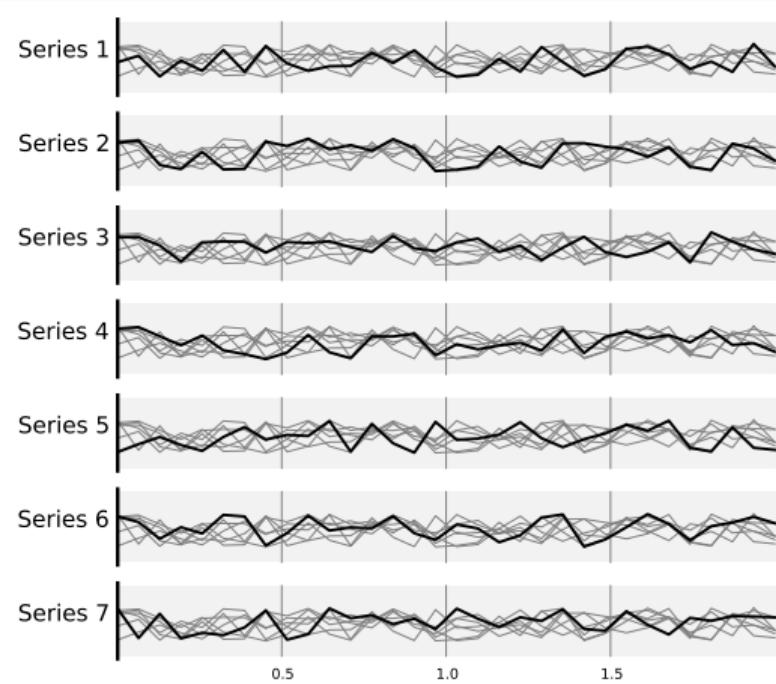
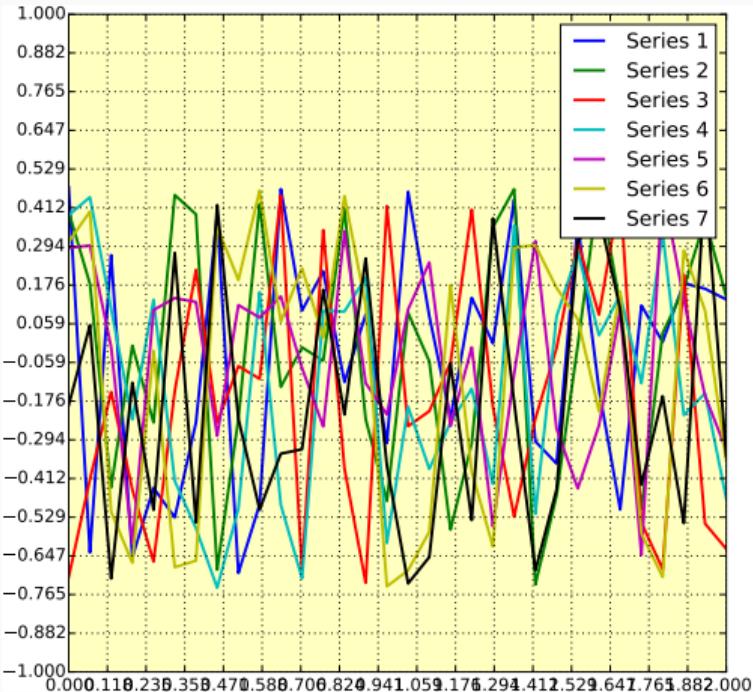
Relative size using partial range



# RULE 7: DO NOT MISLEAD THE READER. REALLY.

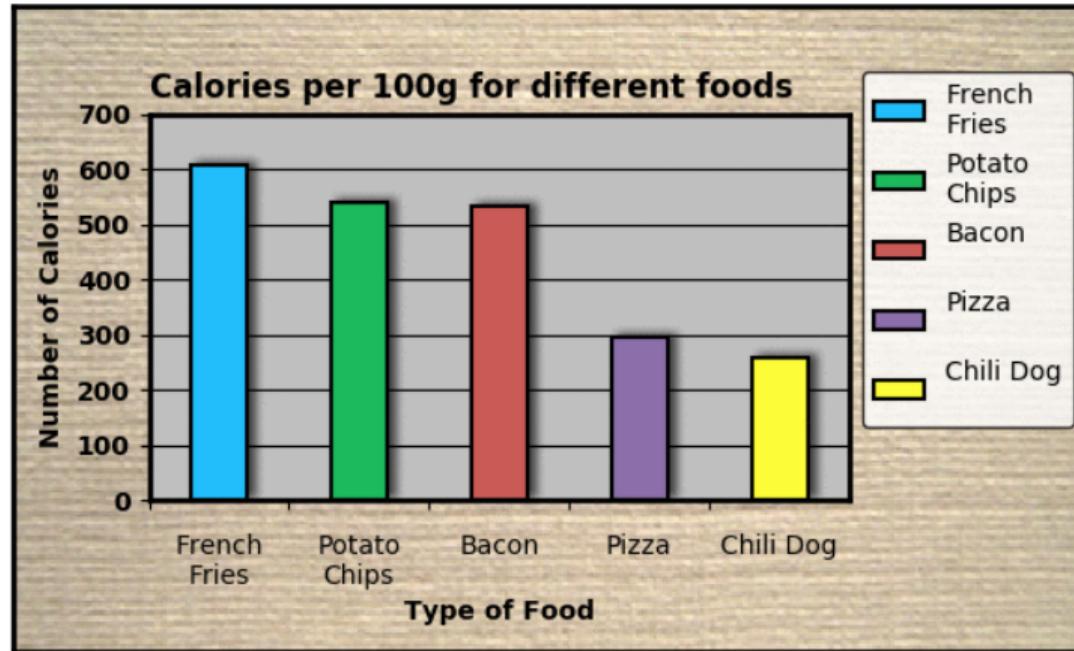


## RULE 8: AVOID CHARTJUNK

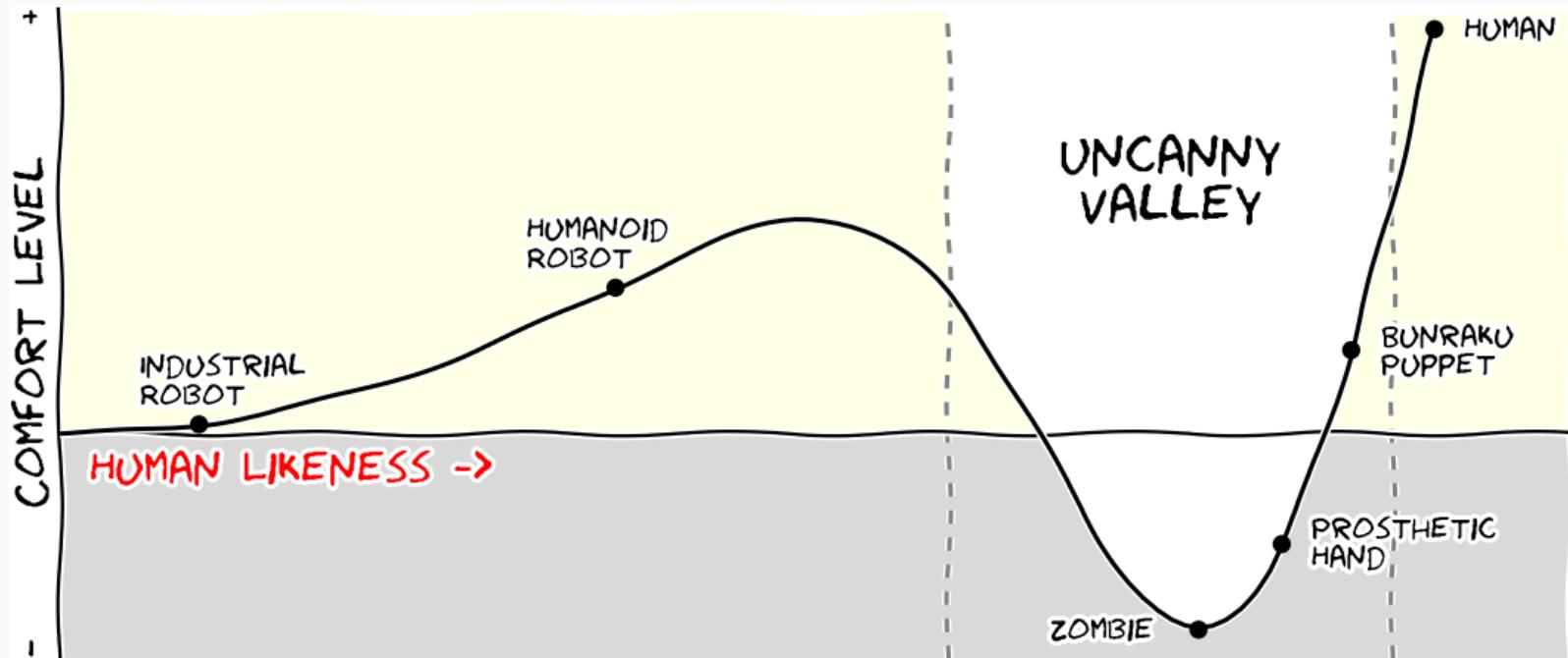


## RULE 8 BIS: LESS IS MORE

Less is More



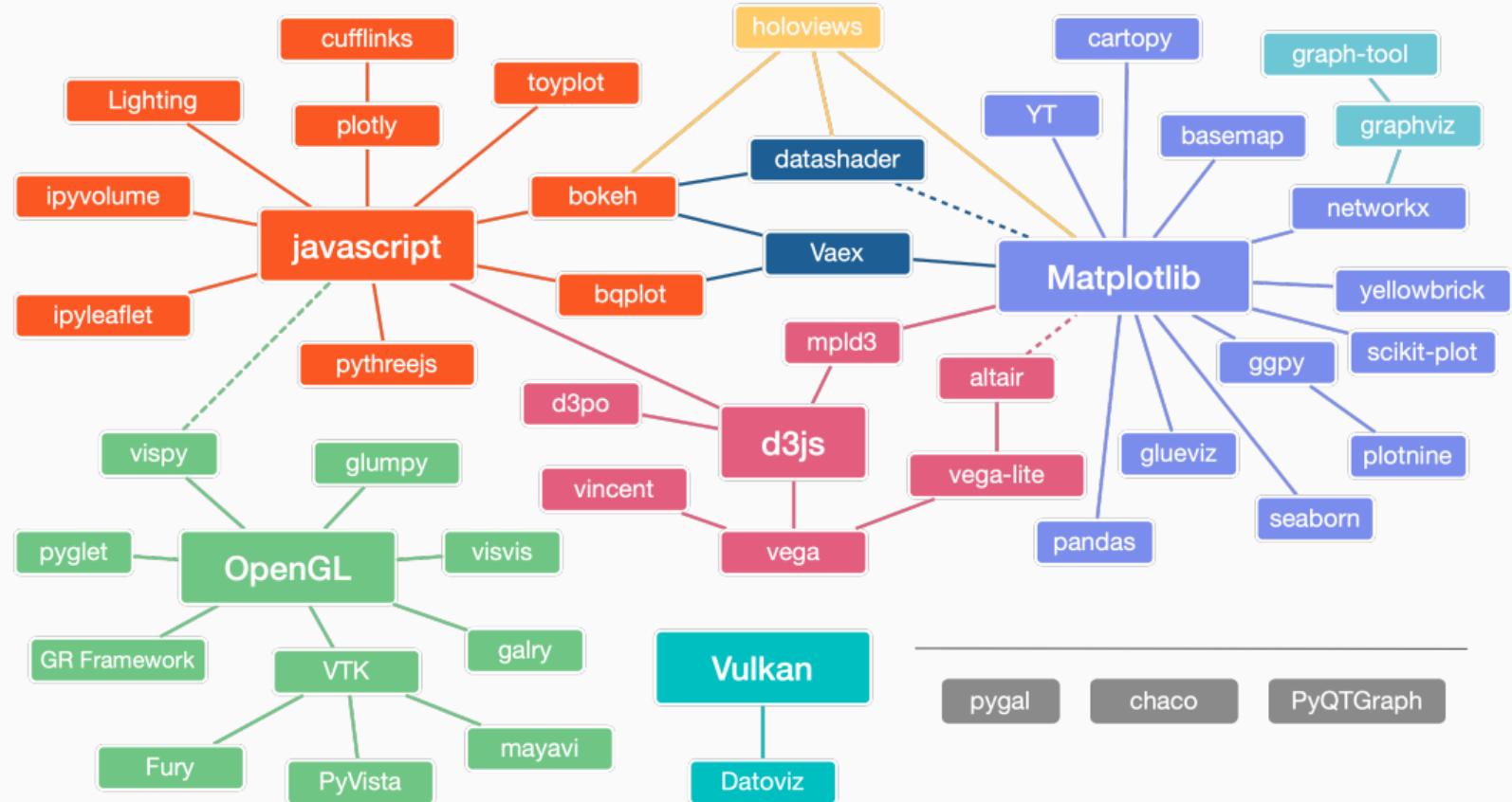
## RULE 9: MESSAGE TRUMPS BEAUTY



## RULE 10: GET THE RIGHT TOOL

- PDFCrop (remove white borders)  
<http://pdfcrop.sourceforge.net>
- GraphViz (easy graph)  
<http://www.graphviz.org>
- ImageMagick (scripted image processing)  
<http://www.imagemagick.org/script/index.php>
- Gimp (bitmap image manipulation)  
<https://www.gimp.org>
- Inkscape (vector image manipulation)  
<https://www.inkscape.org>
- Tikz (scripted vector art)  
<http://www.texample.net/tikz/examples/all/>
- And many, many, many others ...

# RULE 10: GET THE RIGHT TOOL



# CONCLUSION

“Above all, do no harm.” *Edward Tufte*.

# REFERENCES

## Books

- Scientific Visualization – Python & Matplotlib, Nicolas P. Rougier (2021)
- Fundamentals of Data Visualization, Claus O. Wilke (2018)
- Visualization Analysis and Design, T. Munzner (2014)
- Trees, maps, and theorems, J.-L. Doumont (2009)

## Other resources

- Matplotlib Cheatsheets, Nicolas P. Rougier (2020)
- Python Graph Gallery, Yan Holtz (2017)
- Matplotlib Gallery, Matplotlib team
- Ten simple rules for better figures,  
Nicolas P. Rougier, Michael Droettboom, Philip E. Bourne (2014)