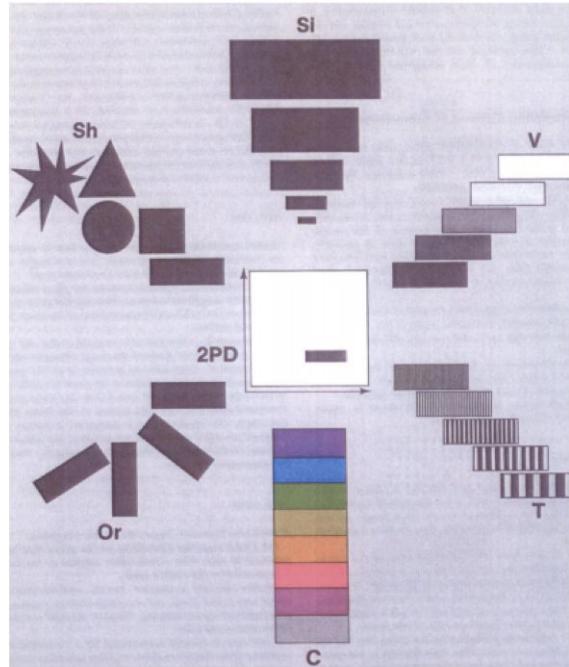


Graphic vocabulary

LES VARIABLES DE L'IMAGE

	POINTS	LIGNES	ZONES
XY 2 DIMENSIONS DU PLAN	x x x	/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
Z TAILLE		/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
VALEUR		/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
	LES VARIABLES DE SÉPARATION DES IMAGES		
GRAIN	■ ■ ■	/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
COULEUR		/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
ORIENTATION		/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9
FORME	▲ ●	/ \ S /	15 9 14 1 2 18 21 2 2 14 15 1 1 1 2 9

Graphical features



- Size
- Value (Density)
- Texture
- Color
- Orientation
- Shape

- 3D
- Animation/Time

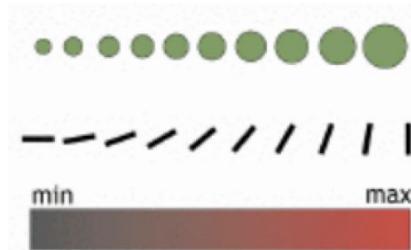
Bertin, *Semiology of Graphics*, 1967 Gauthier-Villars, 1983 UW Press

Data types

- **Continuous:** distance to the closest star (can take any value)
 - **Continuous Ordinal:** Earthquakes (notlinear scale)
 - **Interval:** F temperature - interval size preserved
 - **Ratio:** Car speed - 0 is naturally defined
- **Discrete:** any countable, e.g. number of brain synapses
 - **Counts:** number of bacteria at time t in section A
 - **Ordinal:** survey response Good/Fair/Poor
- **Categorical:** type of medication taken
- **Censored:** individual dead/alive
- **Missing:** “Prefer not to answer” (NA / NaN)

Data types

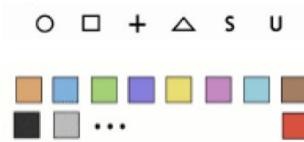
Continuous



Ordered

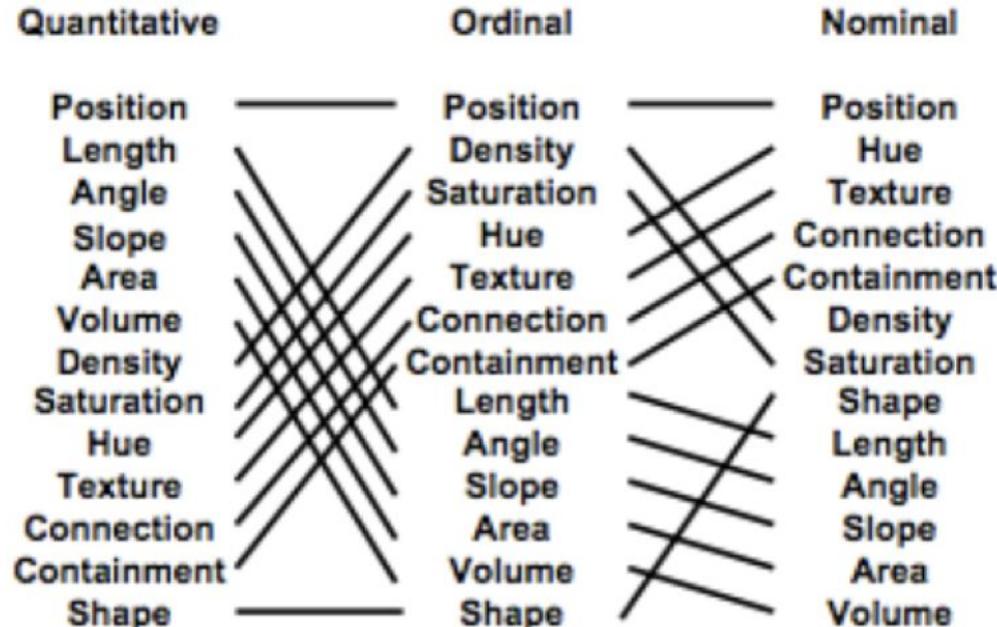


Categorical



graphical elements
work differently on
different data types

Data types



[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]

Colors

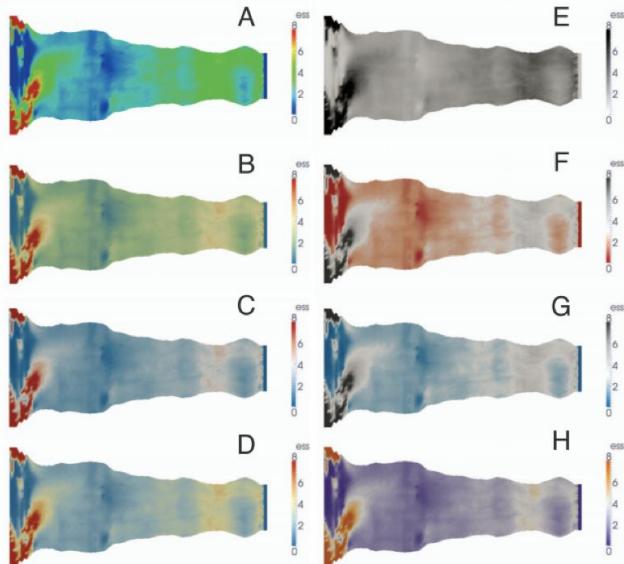


Fig. 4. Color schemes presented during the qualitative user study. The rainbow scheme (A) was preferred by most since it is what they are accustomed to viewing. The next most popular scheme was the red-black diverging scale (F). The grayscale image (E) was unanimously disliked since participants assume black-and-white images to be raw radiological data, while color indicates that the data has been processed or simulated.

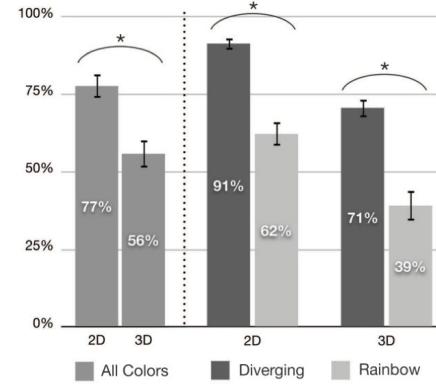


Fig. 7. Average percent of low ESS regions identified broken down by 2D and 3D representation, and color. Error bars correspond to the standard error and the asterisks indicate results of statistical significance. Participants were more accurate in 2D and when using the diverging color map.

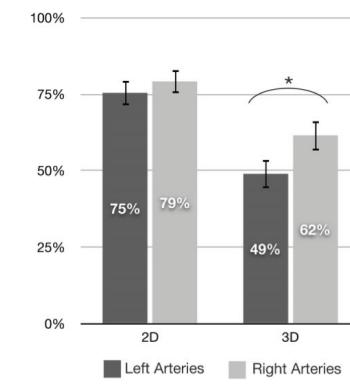


Fig. 8. Average percent of low ESS regions identified broken down by 2D and 3D representation, and left and right artery systems. Error bars correspond to the standard error and the asterisks indicate results of statistical significance. In 3D, users were less accurate identifying regions in the most complex data sets (i.e., left artery systems). Whereas in 2D, performance was the same regardless of task complexity.

Evaluation of artery visualization for heart disease diagnostic - identification of endothelial shear stress (ESS)

<http://www.eecs.harvard.edu/~kgajos/papers/2011/borkin11-infoviz.pdf>

Colors

- 1) Never use rainbow



- 2) Use *diverging* color maps for data for which the center value is “special” (e.g., 0 with data ranging from positive to negative. In a diverging colormap the center of the range is white or black range is white or black



- 3) Choose a *perceptually uniform* color map for continuous data that does not have a focal point (a special point inside the range)



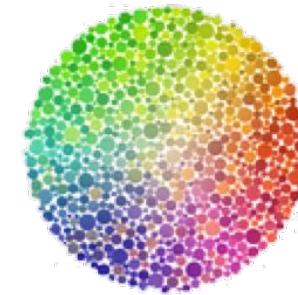
- 4) Choose a sequential cm if your data range represents a progression (reflects some property of the data)



Color blindness



use the <http://colororacle.org/>
app to test your plots for
color-blindness



<http://www.colourblindawareness.org/colour-blindness/>

HS2022 Big Data Analysis in Biomedical Research (376-1723-00L)

Several tools to pick discrete colors/color palettes

<https://colorbrewer2.org/>

<https://www.color-hex.com/color-palettes/>

<https://htmlcolorcodes.com/>

Finding colors: named colors, RGB and HEX

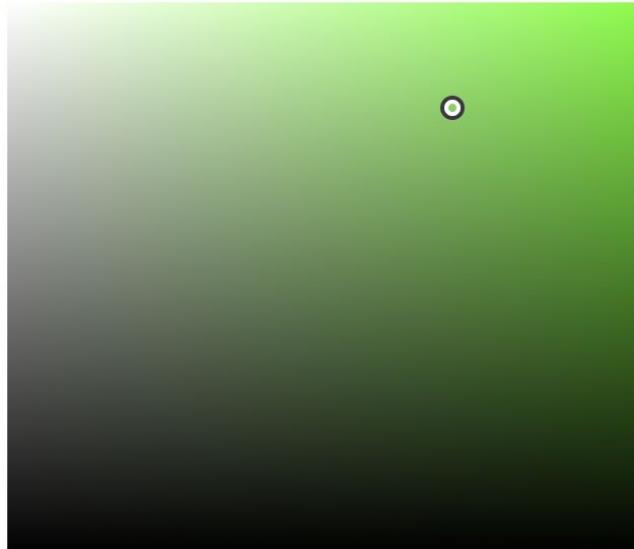


Web Safe Color Chart

#CCFF00 rgb(204, 255, 51)	#CCCC00 rgb(204, 204, 0)	#CC9900 rgb(204, 165, 0)	#CC6600 rgb(204, 132, 0)	#CC3300 rgb(204, 93, 0)	#CC0000 rgb(204, 0, 0)	#660000 rgb(165, 0, 0)	#663300 rgb(165, 51, 0)	#666600 rgb(165, 102, 0)	#669900 rgb(165, 153, 0)	#66CC00 rgb(165, 204, 0)	#66FF00 rgb(165, 255, 0)	#00FF00 rgb(0, 255, 0)	#00CC00 rgb(0, 165, 0)	#009900 rgb(0, 93, 0)	#006600 rgb(0, 61, 0)	#003300 rgb(0, 31, 0)	#000000 rgb(0, 0, 0)
#CCFF33 rgb(204, 255, 93)	#CCCC33 rgb(204, 204, 93)	#CC9933 rgb(204, 165, 93)	#CC6633 rgb(204, 132, 93)	#CC3333 rgb(204, 93, 93)	#CC0033 rgb(204, 0, 93)	#660033 rgb(165, 0, 93)	#663333 rgb(165, 51, 93)	#666633 rgb(165, 102, 93)	#669933 rgb(165, 153, 93)	#66CC33 rgb(165, 204, 93)	#66FF33 rgb(165, 255, 93)	#00FF33 rgb(0, 255, 93)	#00CC33 rgb(0, 165, 93)	#009933 rgb(0, 93, 93)	#006633 rgb(0, 61, 93)	#003333 rgb(0, 31, 93)	#000033 rgb(0, 0, 93)
#CCFF66 rgb(204, 255, 162)	#CCCC66 rgb(204, 204, 162)	#CC9966 rgb(204, 165, 162)	#CC6666 rgb(204, 132, 162)	#CC3366 rgb(204, 93, 162)	#CC0066 rgb(204, 0, 162)	#660066 rgb(165, 0, 162)	#663366 rgb(165, 51, 162)	#666666 rgb(165, 102, 162)	#669966 rgb(165, 153, 162)	#66CC66 rgb(165, 204, 162)	#66FF66 rgb(165, 255, 162)	#00FF66 rgb(0, 255, 162)	#00CC66 rgb(0, 165, 162)	#009966 rgb(0, 93, 162)	#006666 rgb(0, 61, 162)	#003366 rgb(0, 31, 162)	#000066 rgb(0, 0, 162)
#CCFF99 rgb(204, 255, 193)	#CCCC99 rgb(204, 204, 193)	#CC9999 rgb(204, 165, 193)	#CC6699 rgb(204, 132, 193)	#CC3399 rgb(204, 93, 193)	#CC0099 rgb(204, 0, 193)	#660099 rgb(165, 0, 193)	#663399 rgb(165, 51, 193)	#666699 rgb(165, 102, 193)	#669999 rgb(165, 153, 193)	#66CC99 rgb(165, 204, 193)	#66FF99 rgb(165, 255, 193)	#00FF99 rgb(0, 255, 193)	#00CC99 rgb(0, 165, 193)	#009999 rgb(0, 93, 193)	#006699 rgb(0, 61, 193)	#003399 rgb(0, 31, 193)	#000099 rgb(0, 0, 193)
#CCFFCC rgb(204, 255, 234)	#CCCCCC rgb(204, 204, 234)	#CC99CC rgb(204, 165, 234)	#CC66CC rgb(204, 132, 234)	#CC33CC rgb(204, 93, 234)	#CC00CC rgb(204, 0, 234)	#6600CC rgb(165, 0, 234)	#6633CC rgb(165, 51, 234)	#6666CC rgb(165, 102, 234)	#6699CC rgb(165, 153, 234)	#66CCCC rgb(165, 204, 234)	#66FFCC rgb(165, 255, 234)	#00FFCC rgb(0, 255, 234)	#00CCCC rgb(0, 165, 234)	#0099CC rgb(0, 93, 234)	#0066CC rgb(0, 61, 234)	#0033CC rgb(0, 31, 234)	#0000CC rgb(0, 0, 234)
#CCFFFF rgb(204, 255, 255)	#CCCCFF rgb(204, 204, 255)	#CC99FF rgb(204, 165, 255)	#CC66FF rgb(204, 132, 255)	#CC33FF rgb(204, 93, 255)	#CC00FF rgb(204, 0, 255)	#6600FF rgb(165, 0, 255)	#6633FF rgb(165, 51, 255)	#6666FF rgb(165, 102, 255)	#6699FF rgb(165, 153, 255)	#66CCFF rgb(165, 204, 255)	#66FFFF rgb(165, 255, 255)	#00FFFF rgb(0, 255, 255)	#00CCFF rgb(0, 165, 255)	#0099FF rgb(0, 93, 255)	#0066FF rgb(0, 61, 255)	#0033FF rgb(0, 31, 255)	#0000FF rgb(0, 0, 255)
#FFFFFF rgb(255, 255, 255)	#FFCCFF rgb(255, 204, 255)	#FF99FF rgb(255, 165, 255)	#FF66FF rgb(255, 132, 255)	#FF33FF rgb(255, 93, 255)	#FF00FF rgb(255, 0, 255)	#9900FF rgb(165, 0, 255)	#9933FF rgb(165, 51, 255)	#9966FF rgb(165, 102, 255)	#9999FF rgb(165, 153, 255)	#99CCCC rgb(165, 204, 255)	#99FFFF rgb(165, 255, 255)	#00FFFFFF rgb(0, 255, 255)	#00CCFF rgb(0, 165, 255)	#0099FF rgb(0, 93, 255)	#0066FF rgb(0, 61, 255)	#0033FF rgb(0, 31, 255)	#0000FF rgb(0, 0, 255)
#FFFFCC rgb(255, 255, 255)	#FFCCCC rgb(255, 204, 255)	#FF99CC rgb(255, 165, 255)	#FF66CC rgb(255, 132, 255)	#FF33CC rgb(255, 93, 255)	#FF00CC rgb(255, 0, 255)	#9900CC rgb(165, 0, 255)	#9933CC rgb(165, 51, 255)	#9966CC rgb(165, 102, 255)	#9999CC rgb(165, 153, 255)	#99CCCC rgb(165, 204, 255)	#99FFCC rgb(165, 255, 255)	#00FFCC rgb(0, 255, 255)	#00CCCC rgb(0, 165, 255)	#0099CC rgb(0, 93, 255)	#0066CC rgb(0, 61, 255)	#0033CC rgb(0, 31, 255)	#0000CC rgb(0, 0, 255)
#FFFF99 rgb(255, 255, 193)	#FFCC99 rgb(255, 204, 193)	#FF9999 rgb(255, 165, 193)	#FF6699 rgb(255, 132, 193)	#FF3399 rgb(255, 93, 193)	#FF0099 rgb(255, 0, 193)	#990099 rgb(165, 0, 193)	#993399 rgb(165, 51, 193)	#996699 rgb(165, 102, 193)	#999999 rgb(165, 153, 193)	#99CCCC rgb(165, 204, 193)	#99FF99 rgb(165, 255, 193)	#00FF99 rgb(0, 255, 193)	#00CC99 rgb(0, 165, 193)	#009999 rgb(0, 93, 193)	#006699 rgb(0, 61, 193)	#003399 rgb(0, 31, 193)	#000099 rgb(0, 0, 193)
#FFFF66 rgb(255, 255, 162)	#FFCC66 rgb(255, 204, 162)	#FF9966 rgb(255, 165, 162)	#FF6666 rgb(255, 132, 162)	#FF3366 rgb(255, 93, 162)	#FF0066 rgb(255, 0, 162)	#990066 rgb(165, 0, 162)	#993366 rgb(165, 51, 162)	#996666 rgb(165, 102, 162)	#999966 rgb(165, 153, 162)	#99CCCC rgb(165, 204, 162)	#99FF66 rgb(165, 255, 162)	#00FF66 rgb(0, 255, 162)	#00CC66 rgb(0, 165, 162)	#009966 rgb(0, 93, 162)	#006666 rgb(0, 61, 162)	#003366 rgb(0, 31, 162)	#000066 rgb(0, 0, 162)
#FFFF33 rgb(255, 255, 93)	#FFCC33 rgb(255, 204, 93)	#FF9933 rgb(255, 165, 93)	#FF6633 rgb(255, 132, 93)	#FF3333 rgb(255, 93, 93)	#FF0033 rgb(255, 0, 93)	#990033 rgb(165, 0, 93)	#993333 rgb(165, 51, 93)	#996633 rgb(165, 102, 93)	#999933 rgb(165, 153, 93)	#99CCCC rgb(165, 204, 93)	#99FF33 rgb(165, 255, 93)	#00FF33 rgb(0, 255, 93)	#00CC33 rgb(0, 165, 93)	#009933 rgb(0, 93, 93)	#006633 rgb(0, 61, 93)	#003333 rgb(0, 31, 93)	#000033 rgb(0, 0, 93)
#FFFF00 rgb(255, 255, 0)	#FFCC00 rgb(255, 204, 0)	#FF9900 rgb(255, 165, 0)	#FF6600 rgb(255, 132, 0)	#FF3300 rgb(255, 93, 0)	#FF0000 rgb(255, 0, 0)	#990000 rgb(165, 0, 0)	#993300 rgb(165, 51, 0)	#996600 rgb(165, 102, 0)	#999900 rgb(165, 153, 0)	#99CCCC rgb(165, 204, 0)	#99FF00 rgb(165, 255, 0)	#00FF00 rgb(0, 255, 0)	#00CC00 rgb(0, 165, 0)	#009900 rgb(0, 93, 0)	#006600 rgb(0, 61, 0)	#003300 rgb(0, 31, 0)	#000000 rgb(0, 0, 0)

Finding colors: named colors, RGB and HEX

black	linen	forestgreen	slategray
k	bisque	limegreen	lightsteelblue
dimgrey	darkorange	darkgreen	cornflowerblue
dimgray	burlwood	green	royalblue
grey	antiquewhite	g	ghostwhite
gray	tan	lime	lavender
darkgray	navajowhite	seagreen	midnightblue
darkgrey	blanchedalmond	mediumseagreen	navy
silver	papayawhip	springgreen	darkblue
lightgrey	moccasin	mediumspringgreen	mediumblue
lightgray	orange	mediumaquamarine	blue
gainsboro	wheat	aquamarine	slateblue
whitesmoke	oldlace	turquoise	darkslateblue
white	florawhite	lightseagreen	mediumslateblue
w	darkgoldenrod	mediumturquoise	mediumpurple
snow	goldenrod	azure	rebeccapurple
rosybrown	cornsilk	lightcyan	blueviolet
lightcoral	gold	paleturquoise	indigo
indianred	lemonchiffon	darkslategray	darkorchid
brown	khaki	darkslategrey	darkviolet
firebrick	palegoldenrod	teal	mediumorchid
maroon	darkkhaki	darkcyan	thistle
darkred	ivory	c	plum
red	beige	cyan	violet
r	lightyellow	aqua	purple
mistyrose	lightgoldenrodyellow	darkturquoise	darkmagenta
salmon	olive	cadetblue	m
tomato	yellow	powderblue	magenta
darksalmon	olivedrab	lightblue	fuchsia
coral	yellowgreen	deepskyblue	orchid
orangered	darkolivegreen	skyblue	mediumvioletred
lightsalmon	greenyellow	lightskyblue	deeppink
sienna	chartreuse	steelblue	hotpink
seashell	lawngreen	aliceblue	lavenderblush
chocolate	honeydew	dodgerblue	palevioletred
saddlebrown	darkseagreen	lightslategray	crimson
sandybrown	palegreen	lightslateblue	pink
peachpuff	lightgreen	slategray	lightpink
peru			



More examples of coded colors available

X11/CSS4	xkcd	
#00FFFF	#13EAC9	aqua
#7FFFAD	#04D8B2	aquamarine
#FFFFFF	#069AF3	azure
#F5F5DC	#E6DAA6	beige
#000000	#000000	black
#0000FF	#0343DF	blue
#A52A2A	#653700	brown
#7FFF00	#C1F80A	chartreuse
#D2691E	#3D1C02	chocolate
#FF7F50	#FC5A50	coral
#DC143C	#8C000F	crimson
#00FFFF	#00FFFF	cyan
#00008B	#030764	darkblue
#006400	#054907	darkgreen
#FF00FF	#ED0DD9	fuchsia
#FFD700	#DBB40C	gold
#DAA520	#FAC205	goldenrod

X11/CSS4	xkcd	
#008000	#15B01A	green
#808080	#929591	grey
#4B0082	#380282	indigo
#FFFFFF	#FFFFCB	ivory
#F0E68C	#AAA662	khaki
#E6E6FA	#C79FEF	lavender
#ADD8E6	#7BC8F6	lightblue
#90EE90	#76FF7B	lightgreen
#00FF00	#AAFF32	lime
#FF00FF	#C20078	magenta
#800000	#650021	maroon
#000080	#01153E	navy
#808000	#6E750E	olive
#FFA500	#F97306	orange
#FF4500	#FE420F	orangered
#DA70D6	#C875C4	orchid
#FFC0CB	#FF81C0	pink

X11/CSS4	xkcd	
#DDA0DD	#580F41	plum
#800080	#7E1E9C	purple
#FF0000	#E50000	red
#FA8072	#FF796C	salmon
#A0522D	#A9561E	sienna
#C0C0C0	#C5C9C7	silver
#D2B48C	#D1B26F	tan
#008080	#029386	teal
#FF6347	#EF4026	tomato
#40E0D0	#06C2AC	turquoise
#EE82EE	#9A0EEA	violet
#F5DEB3	#FBDD7E	wheat
#FFFFFF	#FFFFFF	white
#FFFF00	#FFFF14	yellow
#9ACD32	#BBF90F	yellowgreen

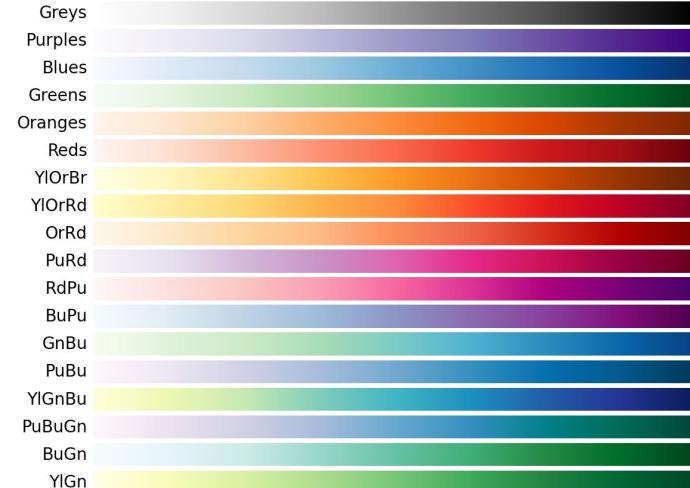
Color palettes

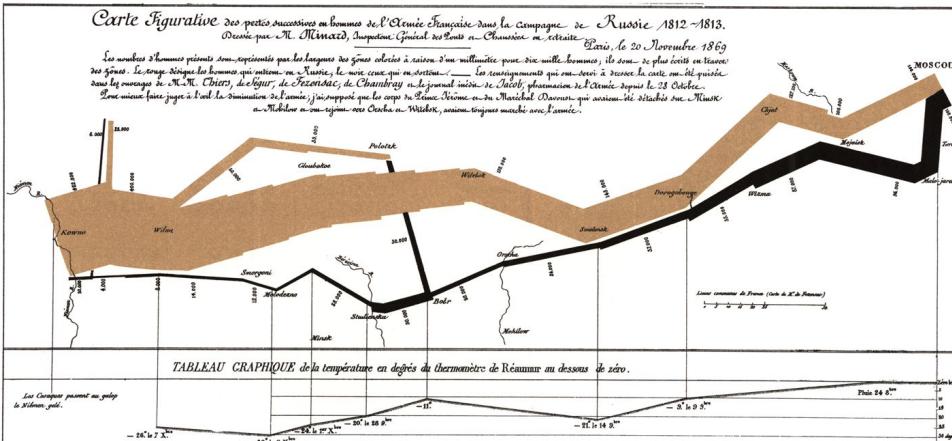


Perceptually Uniform Sequential colormaps



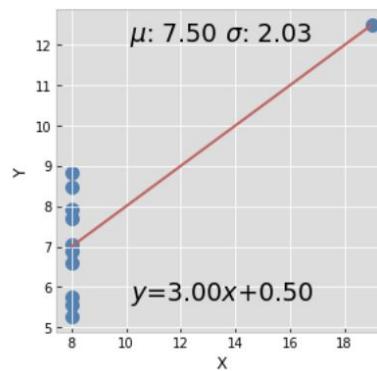
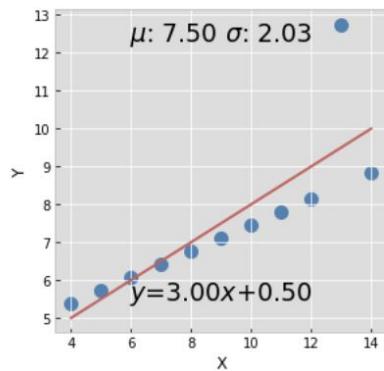
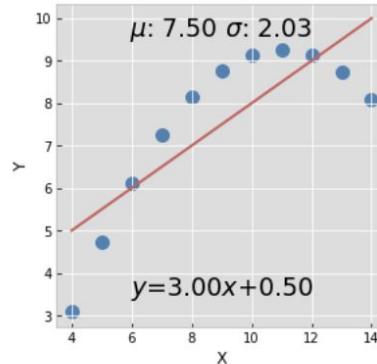
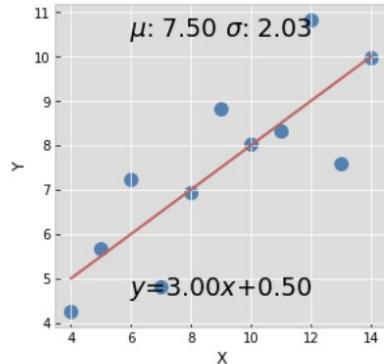
Sequential colormaps





Data Visualization

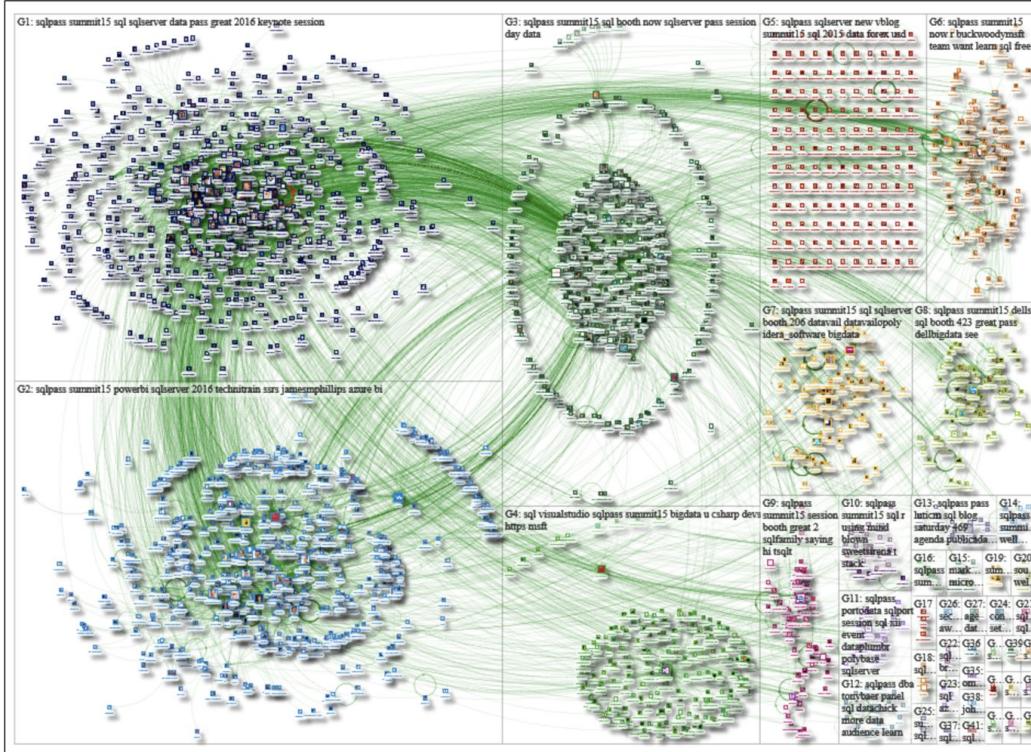
The importance of visualization



Anscombe's quartet

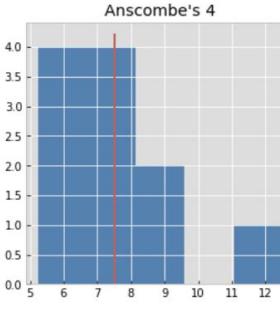
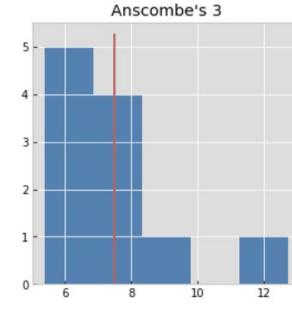
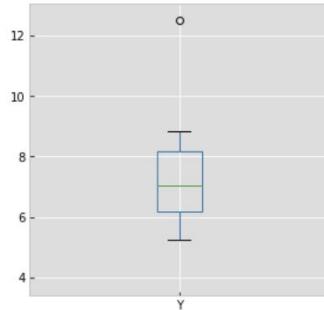
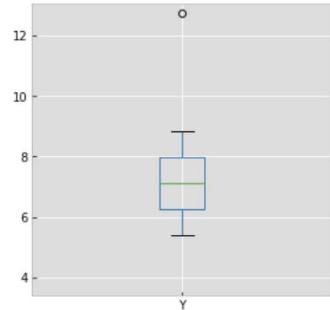
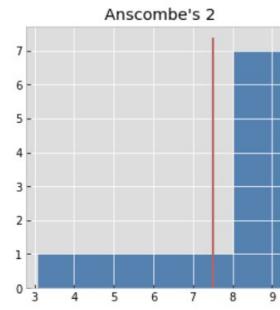
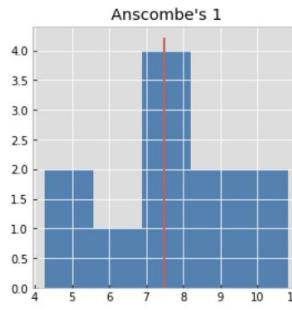
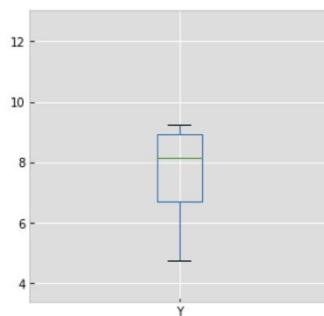
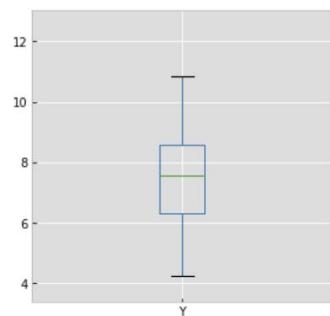
(Francis Anscombe, 1973) comprises four datasets that have **nearly identical simple descriptive statistics**, yet appear very different when graphed. Each dataset consists of eleven (x,y) points.

The problem with big data



the larger the data, and especially
the higher the number of
dimensions, the harder to design a
visualization that is effective

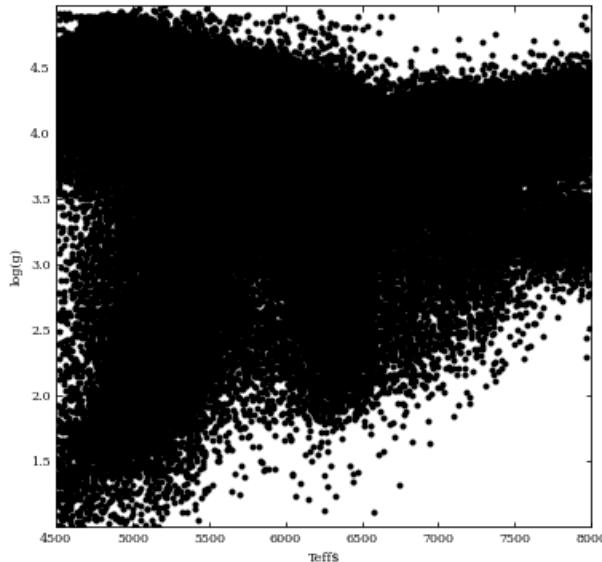
Visualization of derived data products may be effective



... the larger the data the harder it is to collectively understand the descriptive statistics

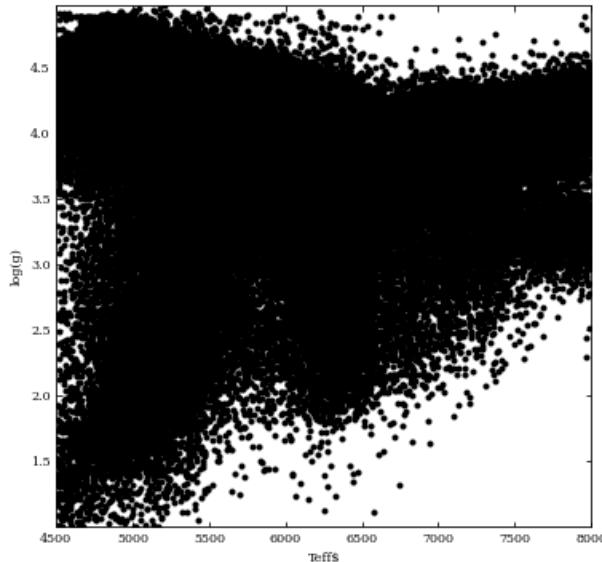
A common problem: too many points

```
plt.plot(Teff, logg, 'k.')
```

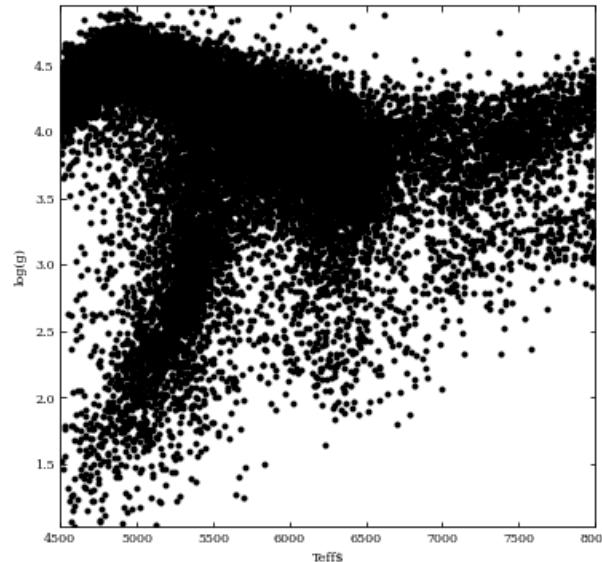


A common problem: too many points

```
plt.plot(Teff, logg, 'k.')
```



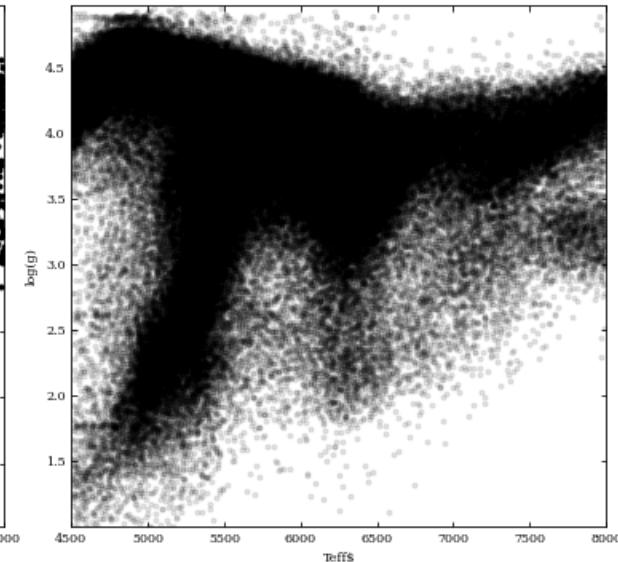
```
plt.plot(Teff[::10], logg[::10], 'k.')
```



solution: subsample

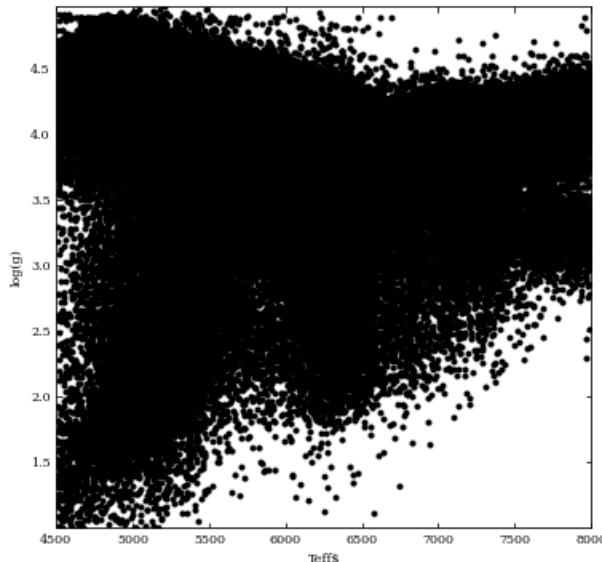
solution: alpha

```
plt.plot(Teff, logg, 'k.', alpha=0.1)
```



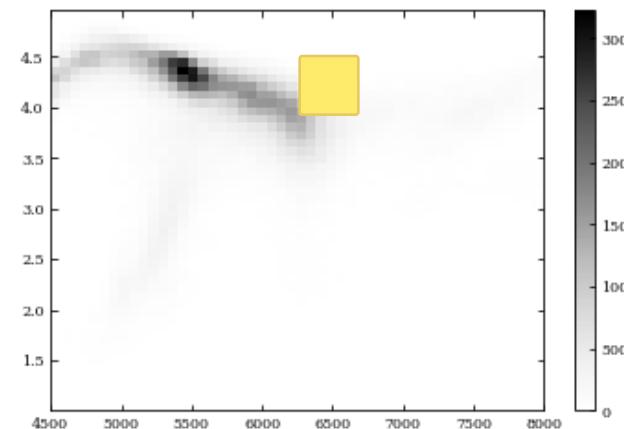
A common problem: too many points

```
plt.plot(Teff, logg, 'k.')
```



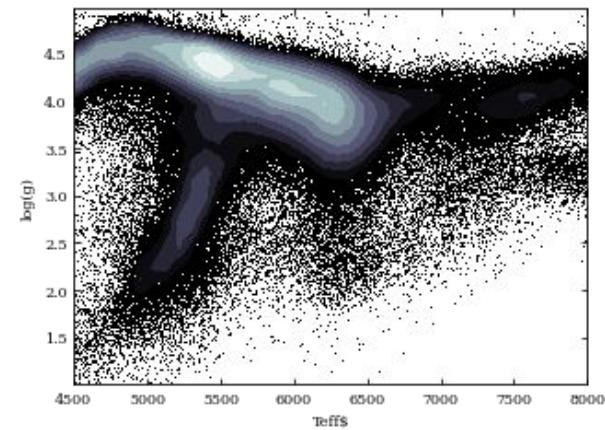
solution: density histograms

```
plt.hist2d(Teff, logg, bins=(50, 50),  
cmap=plt.cm.Greys)
```

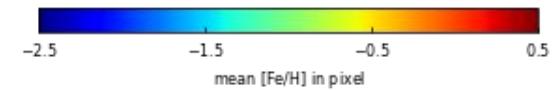
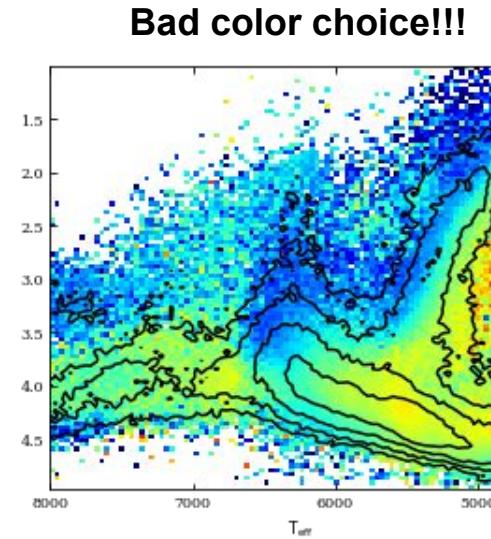
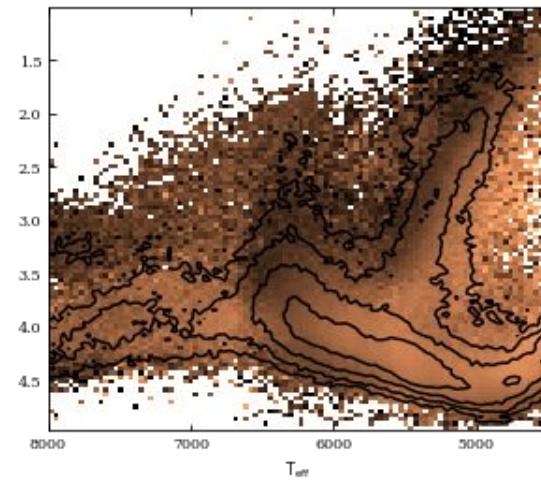
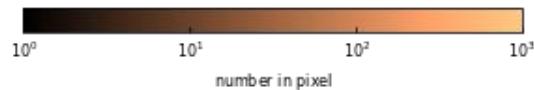
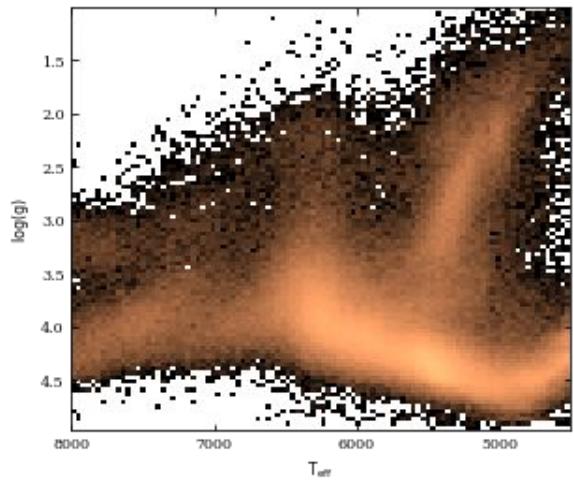


solution: scatter contours

```
plt.plot(Teff, logg, 'k.', alpha=0.1)
```



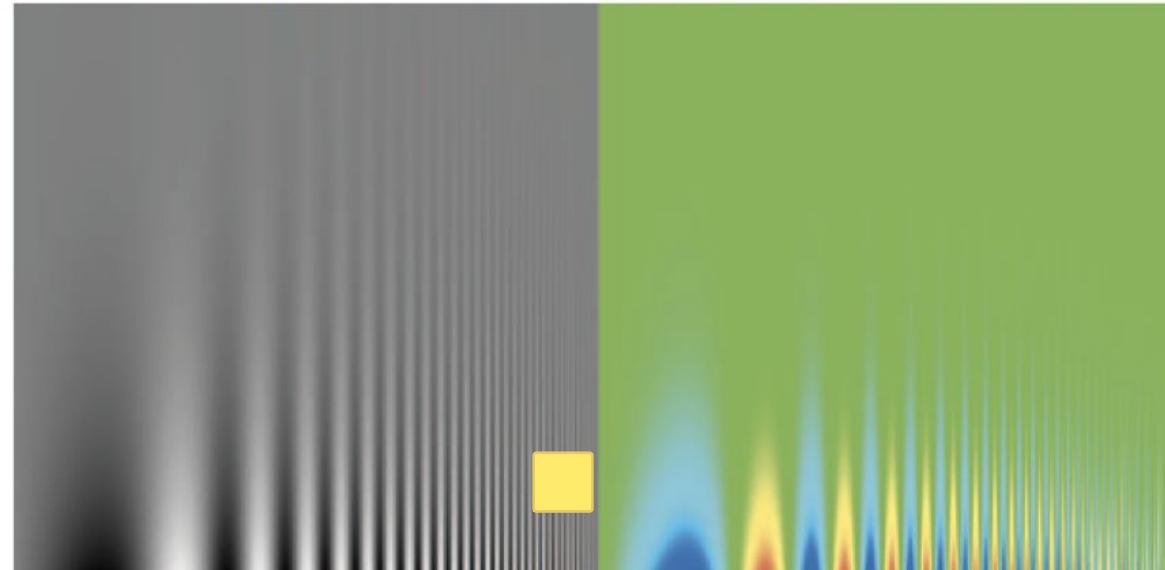
A common problem: too many points



A common problem: too many points

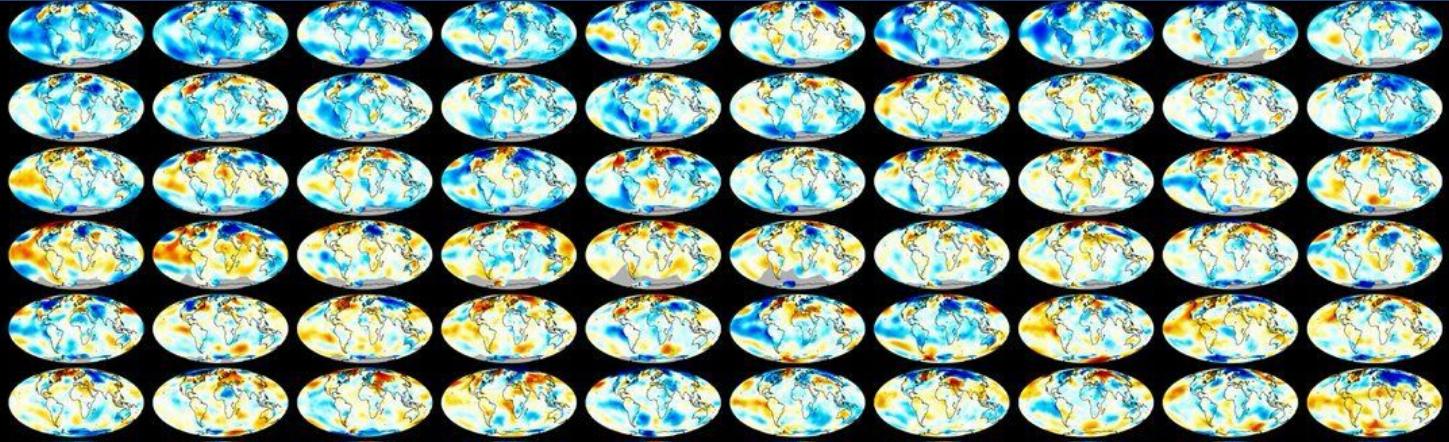
5. Detail is actually harder to see in a rainbow.

The logic that it is easier to see detail in a range when you add colors seems to make sense, but in reality, more detail can be seen in a single hue image with a high brightness range.



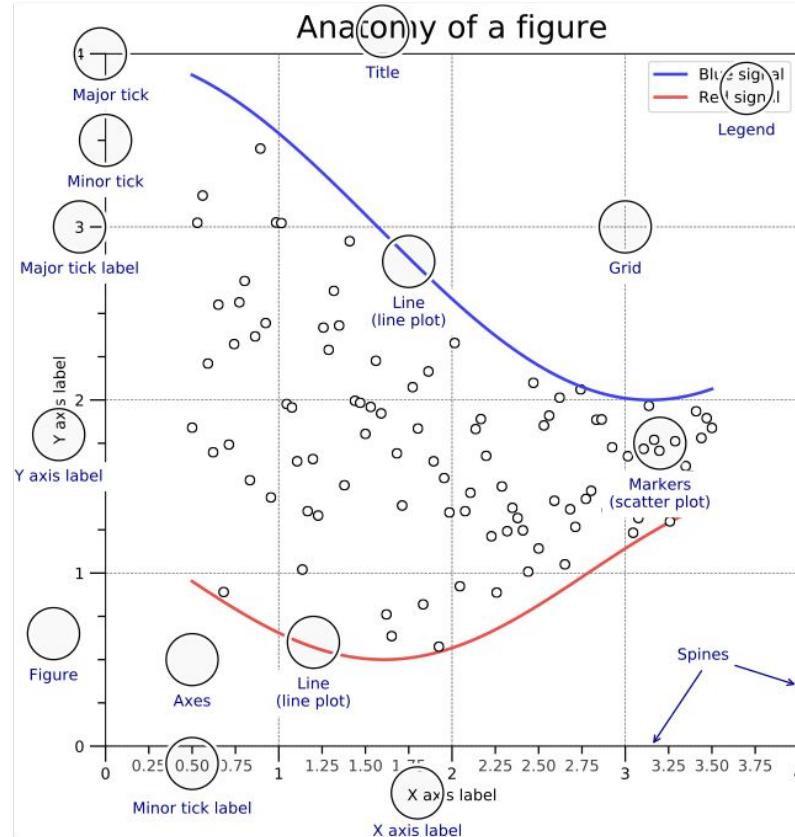
(source)

<https://scienzenode.org/visualization/dear-nasa-no-more-rainbows-please.php>



What makes a good plot

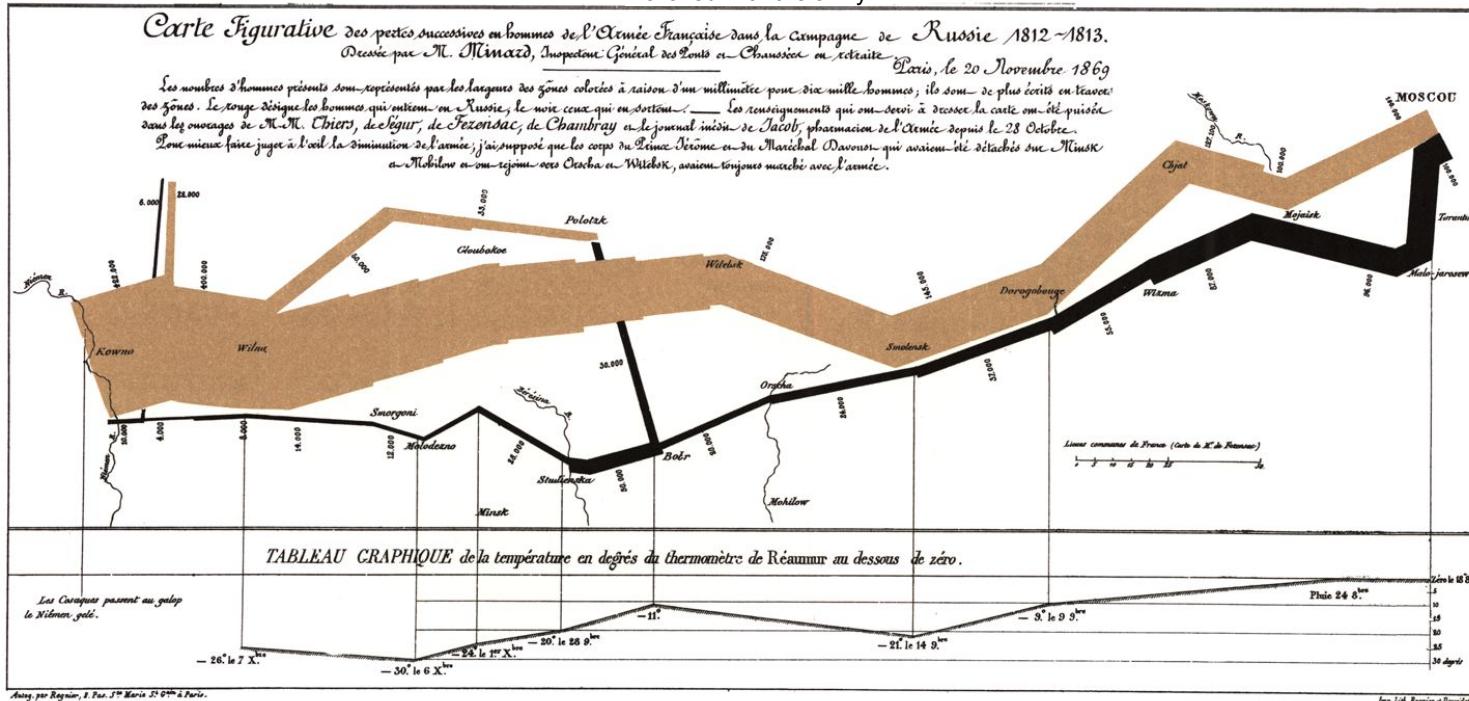
Anatomy of a figure



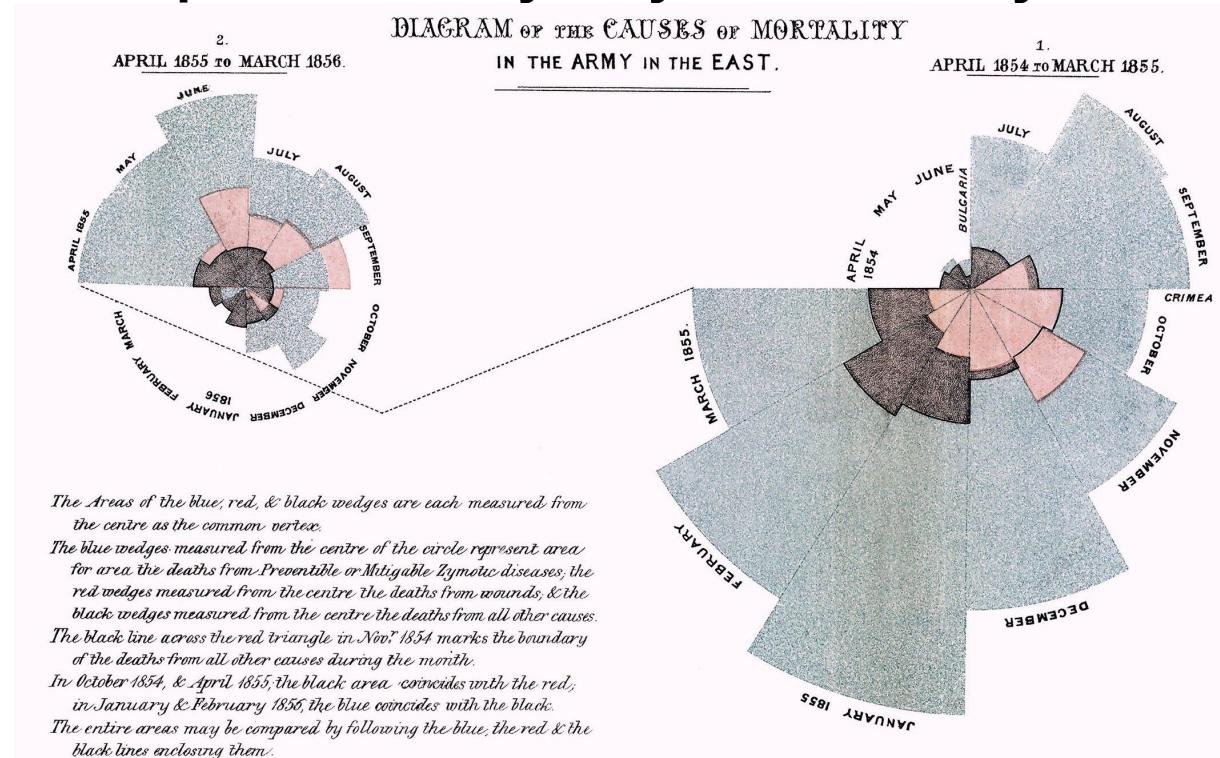
A few historical plots and why they made history

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.

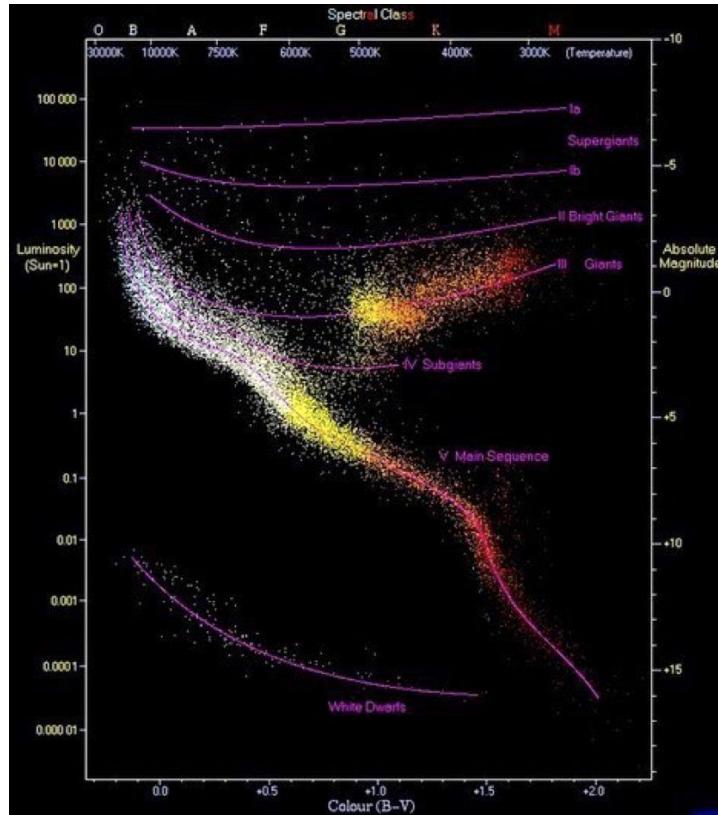


A few historical plots and why they made history



Florence Nightingale Coxcombs : http://timelyportfolio.github.io/rCharts_micropolar/nightingale/index.html

A few historical plots and why they made history



H-R diagram: the life of a star (~1910)

https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram



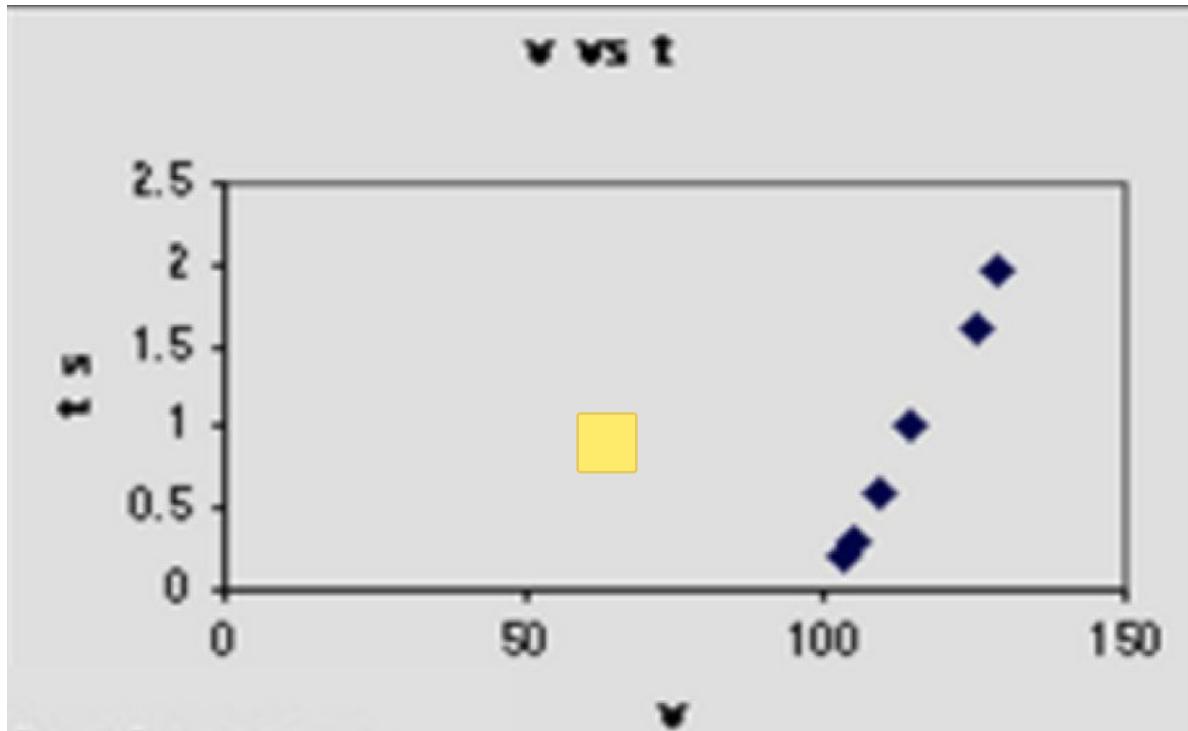
What makes a bad plot?

Ambiguity | distortion | distraction.



Ambiguity | distortion | distraction.

What is wrong with this plot?



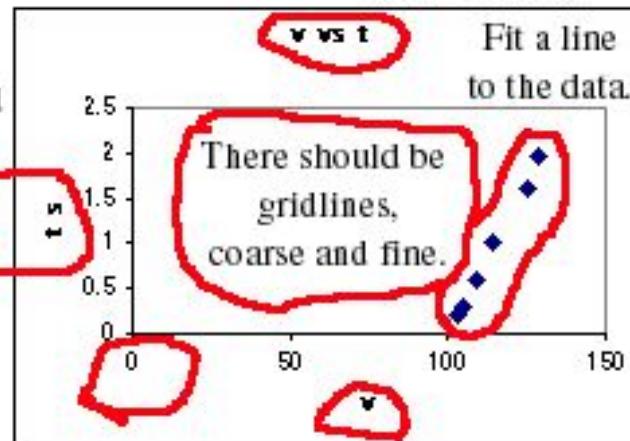
<http://www.geol.lsu.edu/jlorenzo/geophysics/graphing/badlinear1.html>

What is wrong with this plot?

The entire graph is too small.

The title should be better.
This graph is t versus v ,
not v versus t .

The axis label should have words, and the units should be in parentheses.

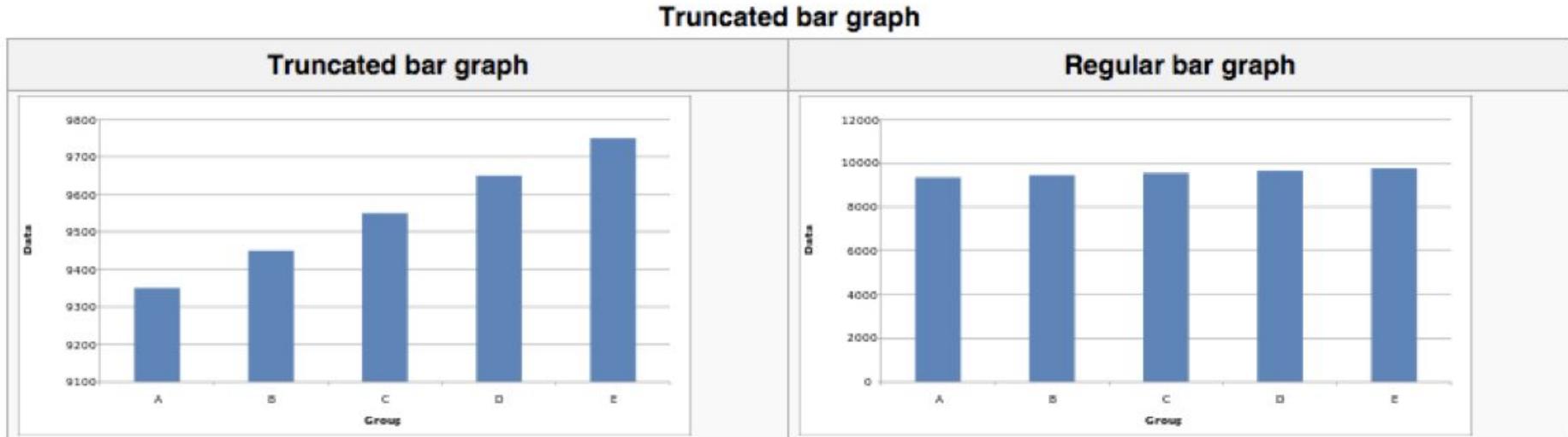


The first data occurs at $v = 100$, so the scale can begin at 100.

The axis label should have words and units in parentheses.

Ambiguity | distortion | distraction.

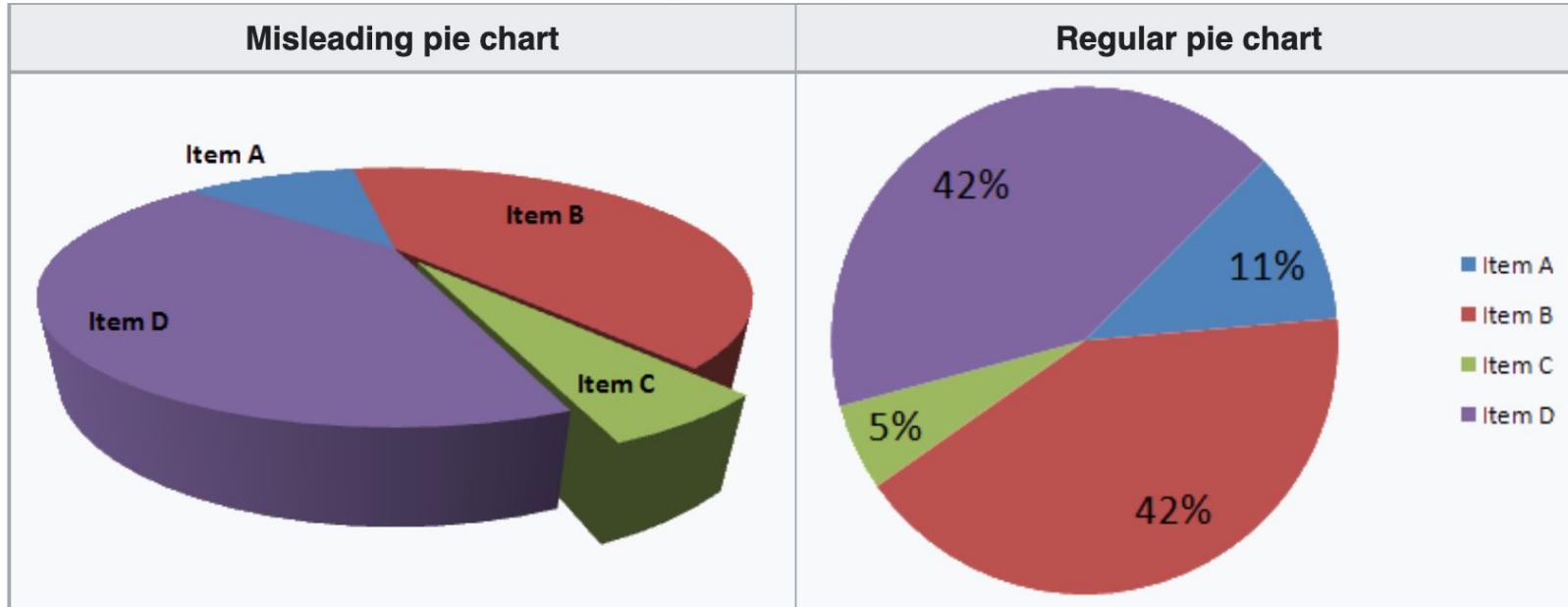
Distortion = misleading representation



Note that both of these graphs display *identical data*; however, in the truncated bar graph on the left, the data *appear* to show significant differences, whereas in the regular bar graph on the right, these differences are hardly visible.

Distortion = misleading representation

Comparison of pie charts

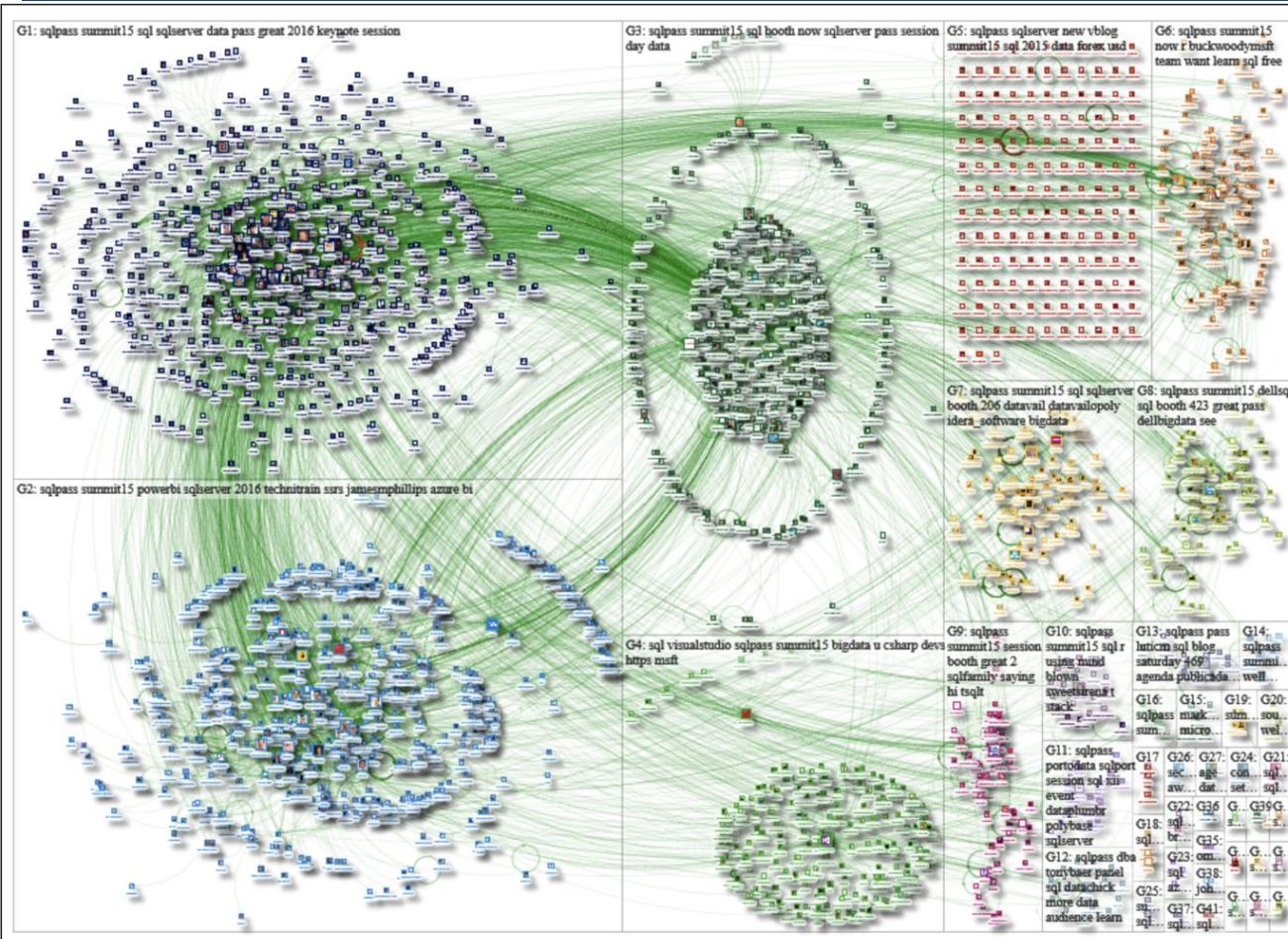


Item C appears to be at least as large as Item A in the misleading pie chart, whereas in actuality, it is less than half as large.

https://en.wikipedia.org/wiki/Misleading_graph

Ambiguity | distortion | **distraction.**

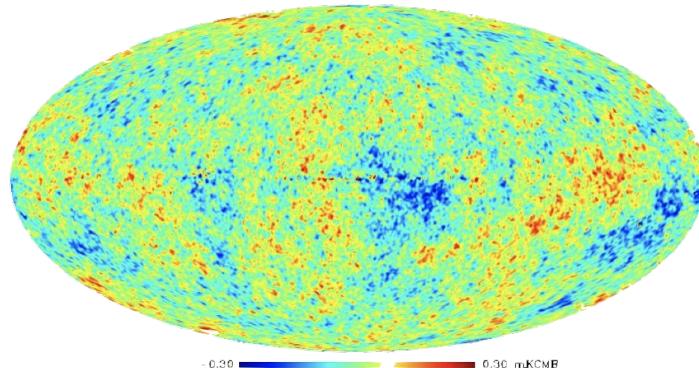




Sometime the distraction is a consequence of the complexity of the data.

<http://www.nodexlgraphgallery.org/Pages/InteractiveGraph.aspx?graphID=56967>

WMAP ILC 5 years



What makes a good visualization?

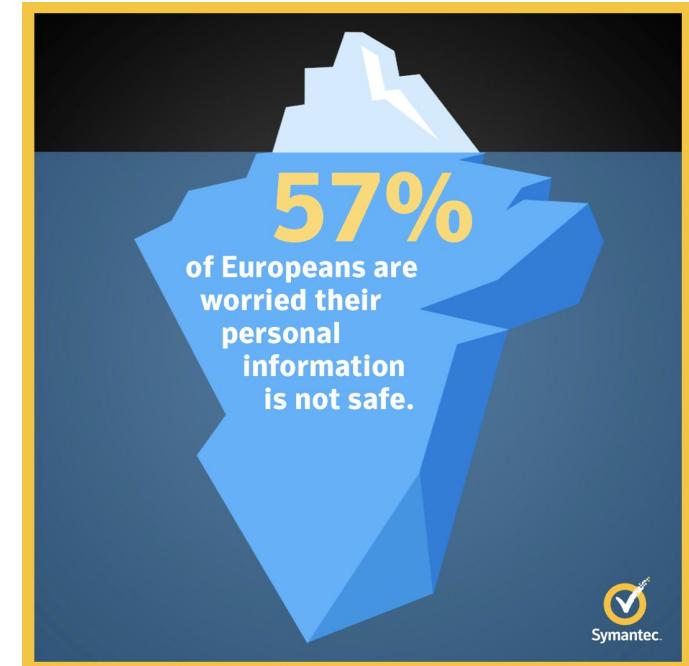
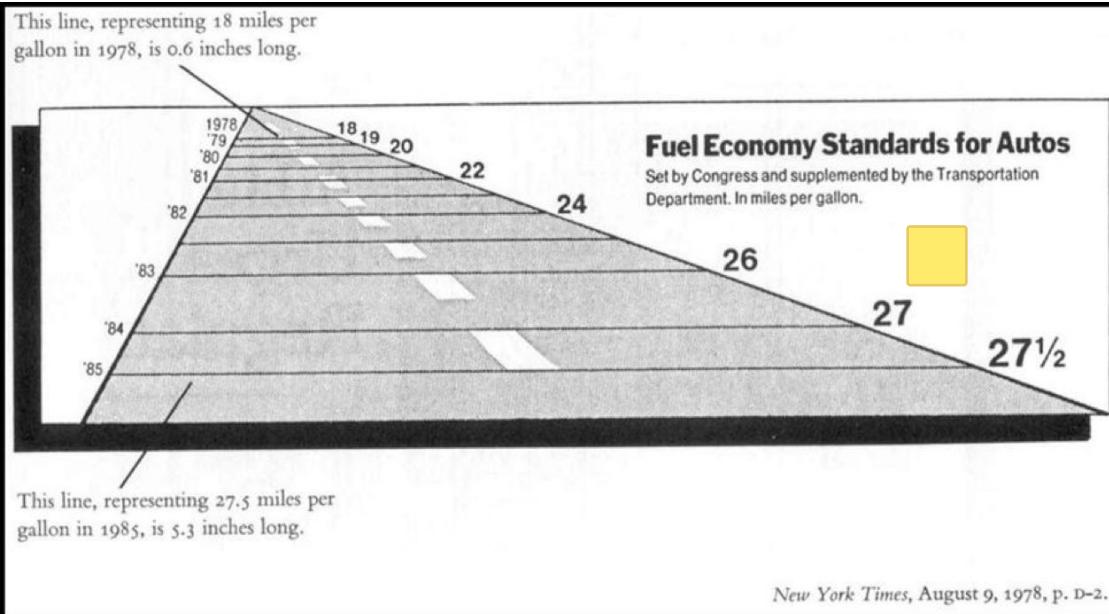
Tufte's rules

Tufte's rules

Lie factor =

size of the effect in the graphic

size of the effect in the data



Tufte's rules

- 1) The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented ("lie factor")

Effect size ~ 1

Data-ink ratio

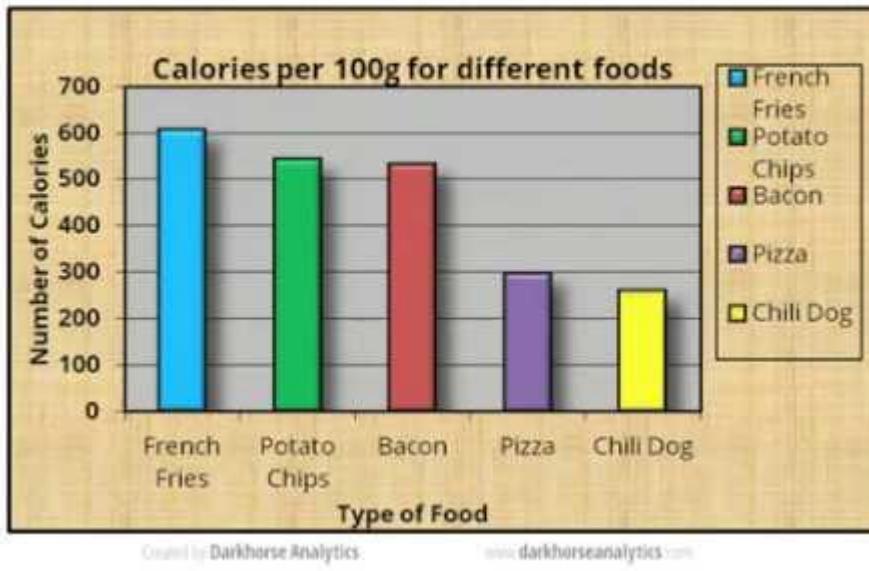
The data-ink ratio is the proportion of Ink that is used to present actual data compared to the total amount of ink (or pixels) used in the entire display. (Ratio of Data-Ink to non-Data-Ink).

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

- = proportion of a graphic's ink devoted to the non-redundant display of data-information
- = 1.0 – proportion of a graphic that can be erased

Data-ink ratio

Remove backgrounds



Tufte's rules

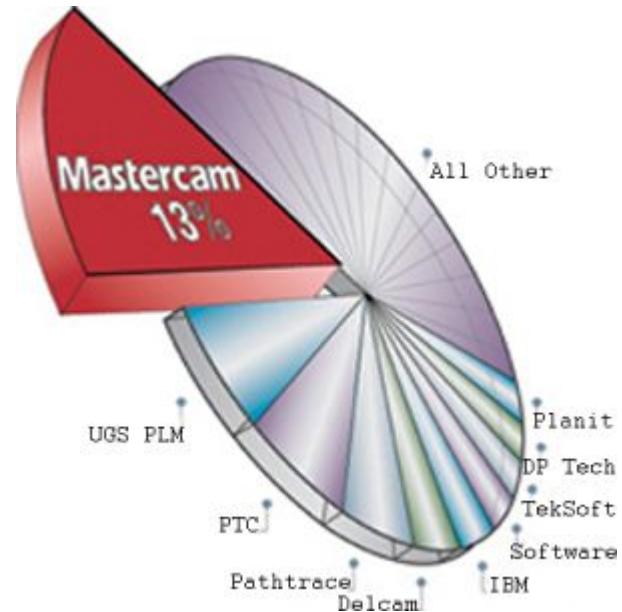
- 1) The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented ("lie factor")
- 2) Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.

Effect size ~ 1

data/ink -> large

Chart junk

The excessive and unnecessary use of graphical effects



Tufte's rules

- 1) The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented ("lie factor")
- 2) Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.
- 3) Show data variation, not design variation

Effect size ~ 1

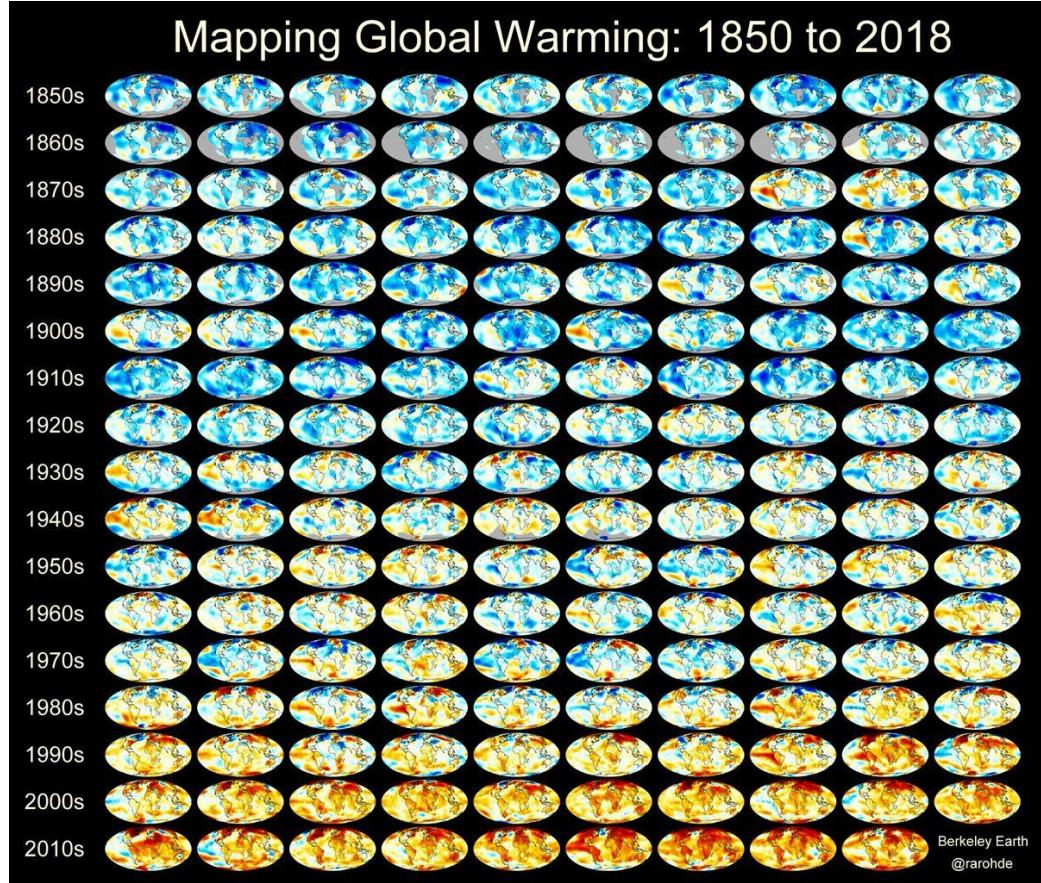
data/ink -> large

no chart junk

Small multiples

encourage comparison

<https://mahb.stanford.edu/whats-happening/167-tiny-maps-tell-major-story-climate-change/>



Tufte's rules

- 1) The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented ("lie factor")
- 2) Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.
- 3) Show data variation, not design variation
- 4) In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

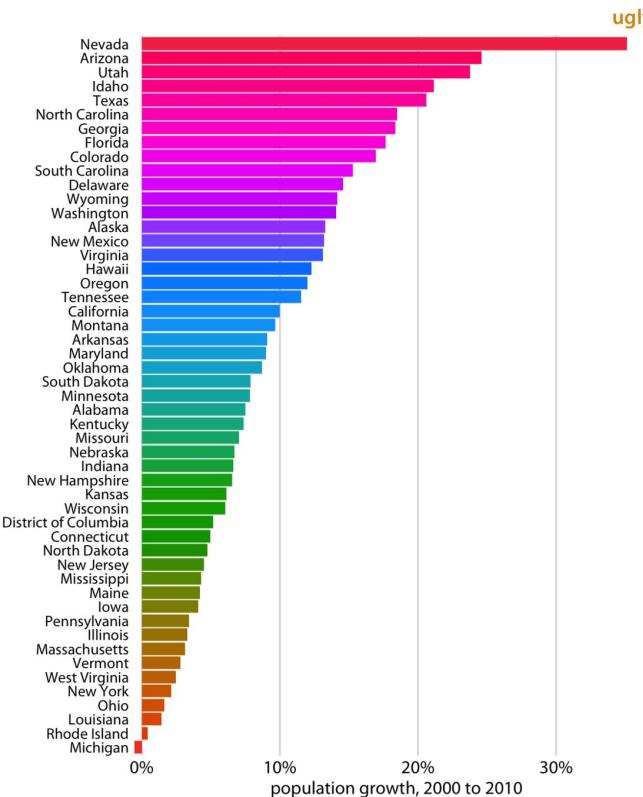
Effect size ~ 1

data/ink -> large

no chart junk

use small multiples

Redundancy of information

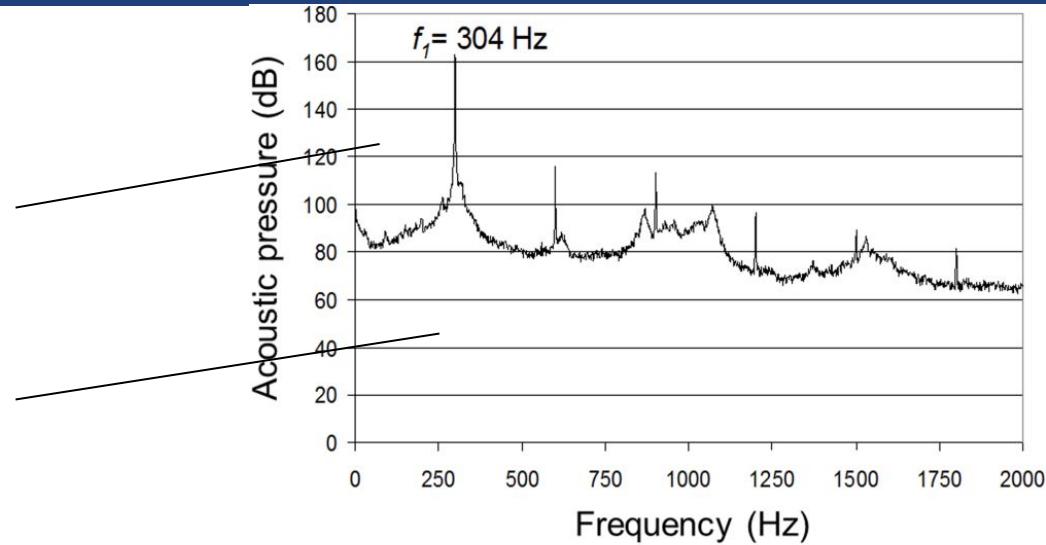


Tufte's rules

- 1) The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented ("lie factor")
Effect size ~ 1
- 2) Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.
data/ink -> large
- 3) Show data variation, not design variation
no chart junk
- 4) In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
use small multiples
- 5) The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data. Graphics must not quote data out of context.
avoid redundant communication

Tufte's rules

low data/ink ratio

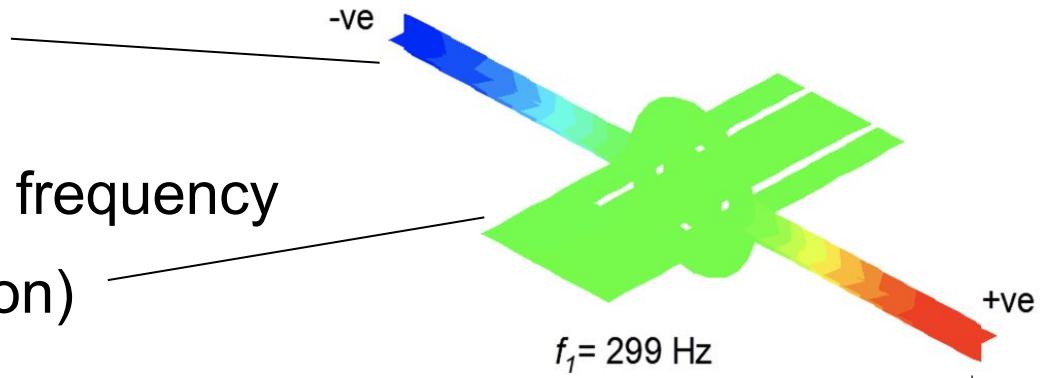


no comparison

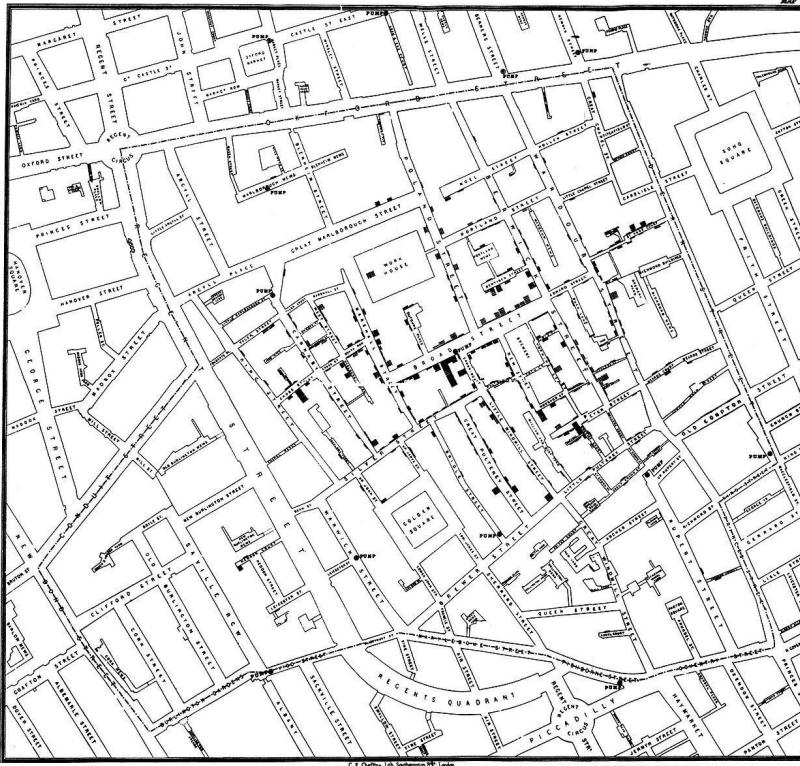
chart junk

2 graphical elements for frequency

(color and position)



Use visualization to understand data, not to communicate a result



John Snow's map of cholera,
considered the first
"data science project"
uses "clustering" to drive
causal inference

[https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak#cite_ref-FOOTNOTES now1855\[httpsarchiveorgstreamb28985266page38mode1up_38\]_19-0](https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak#cite_ref-FOOTNOTES now1855[httpsarchiveorgstreamb28985266page38mode1up_38]_19-0)

More rules

Function first, Form next
no unjustified beauty

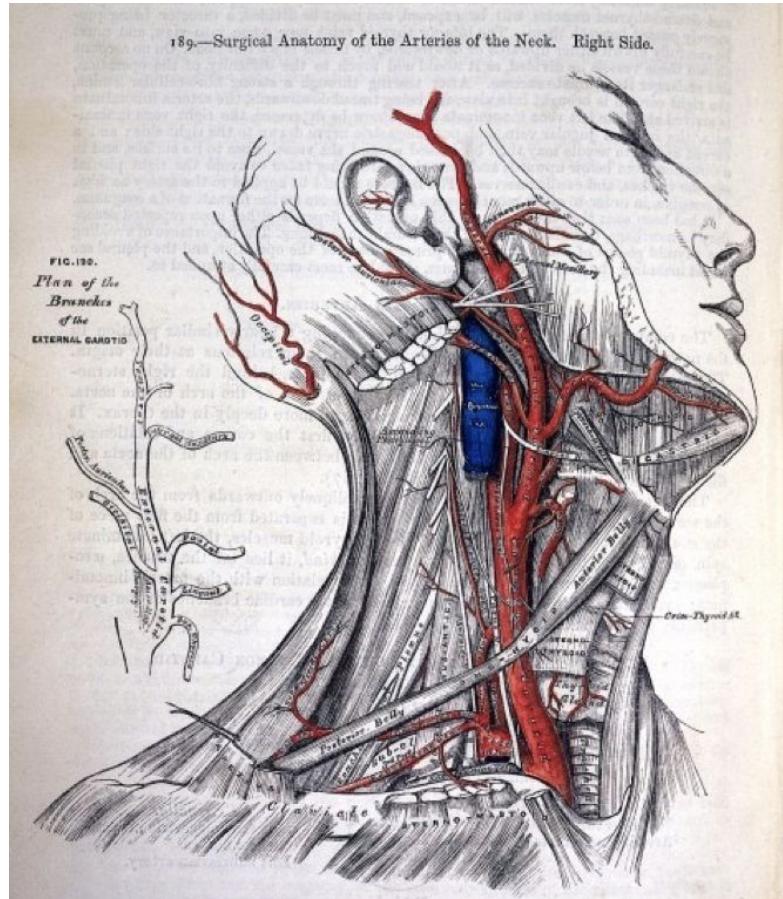
Get it right in Black & White
no unjustified color
consider designing your plot in BW first

No Unjustified 3D

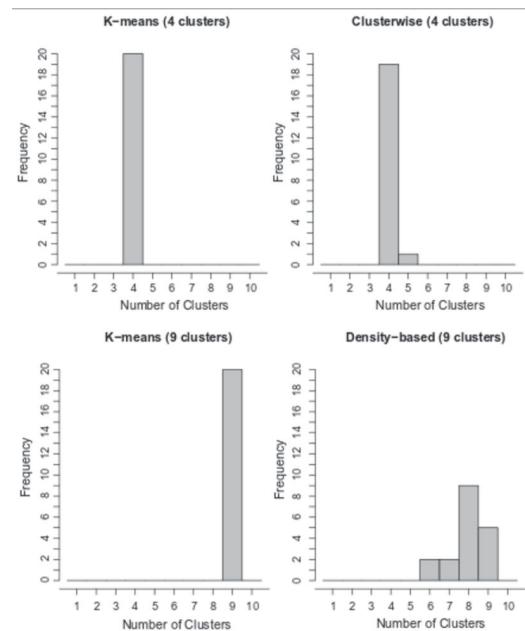
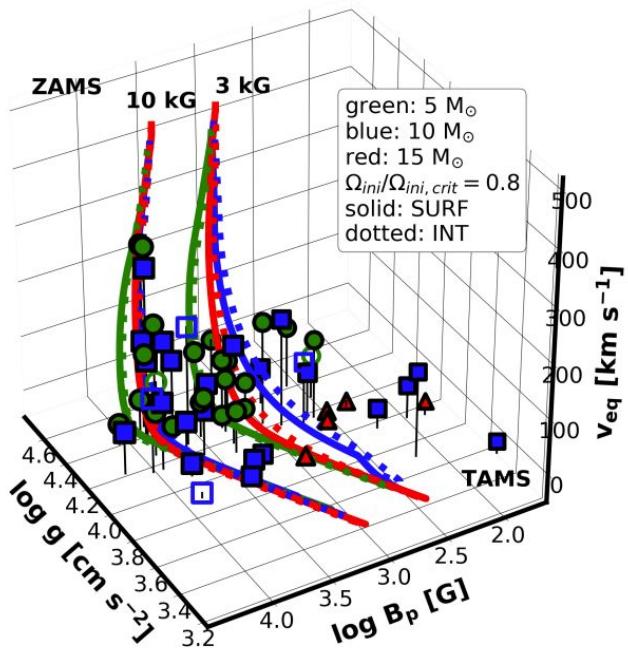
Eyes over Memory
no unjustified animation

More examples

Functional use of colors (and distortion)

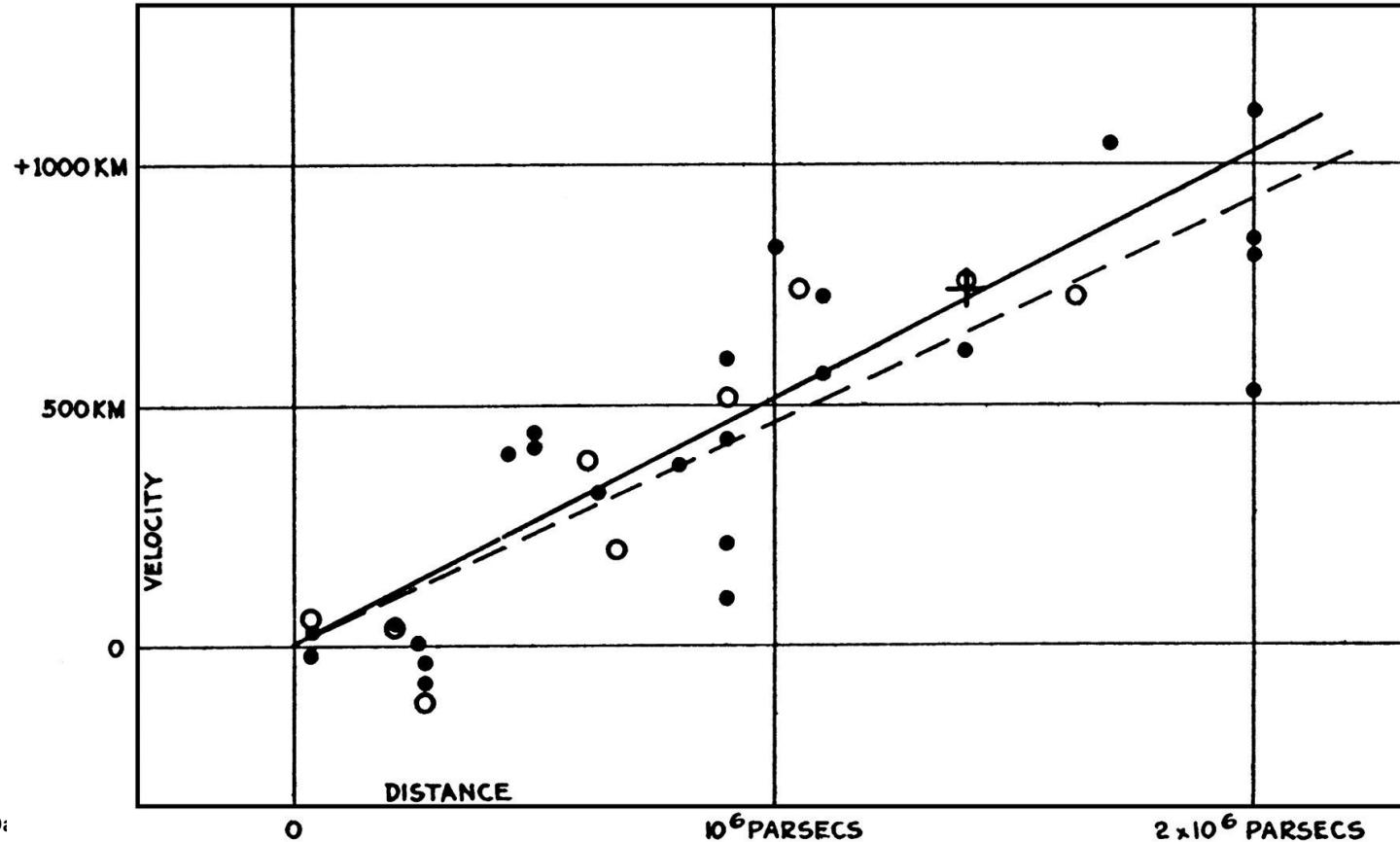


More examples

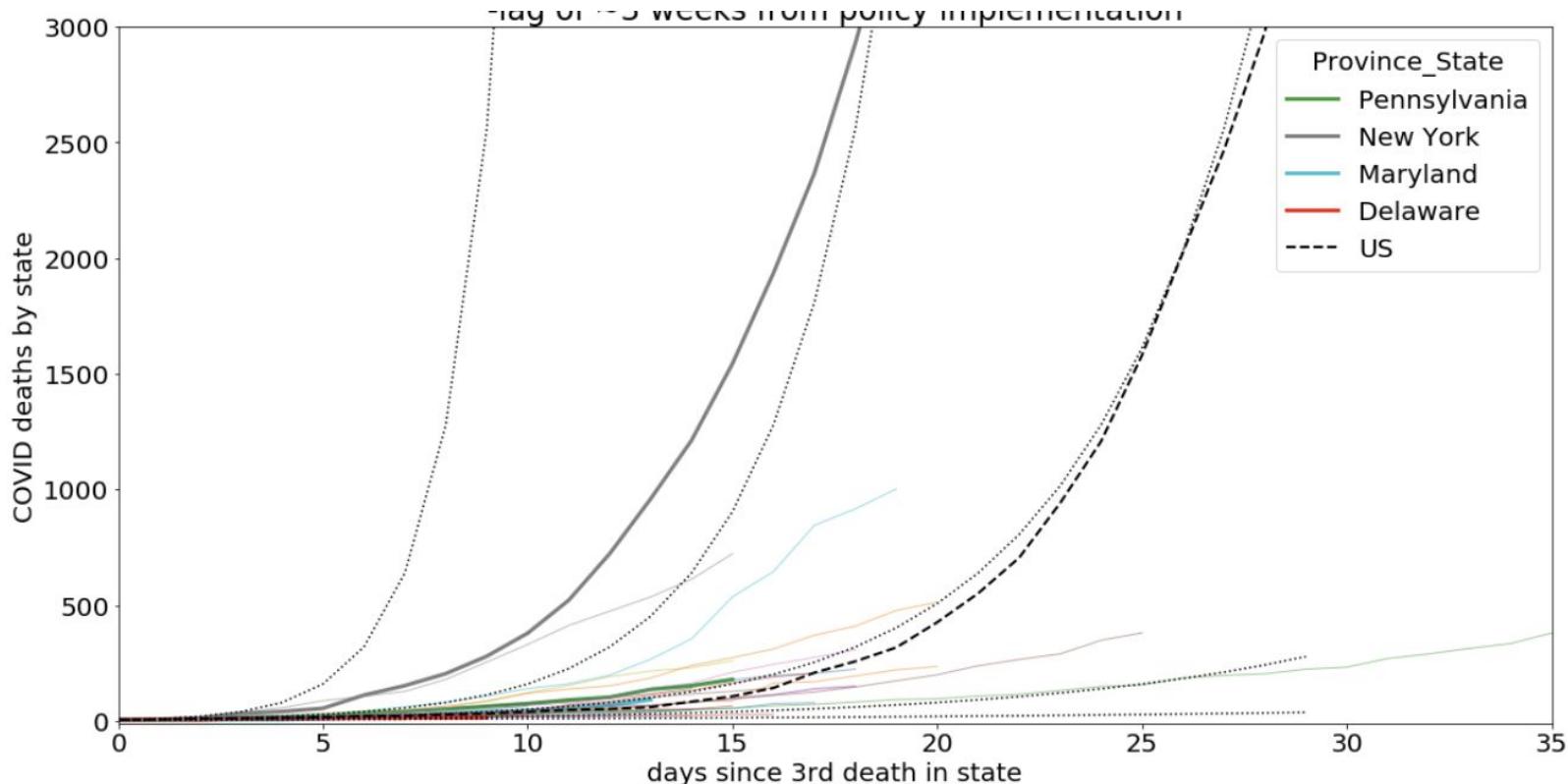


No unjustified 3D
(or 2D either!)
obstruction
clutter
deformation

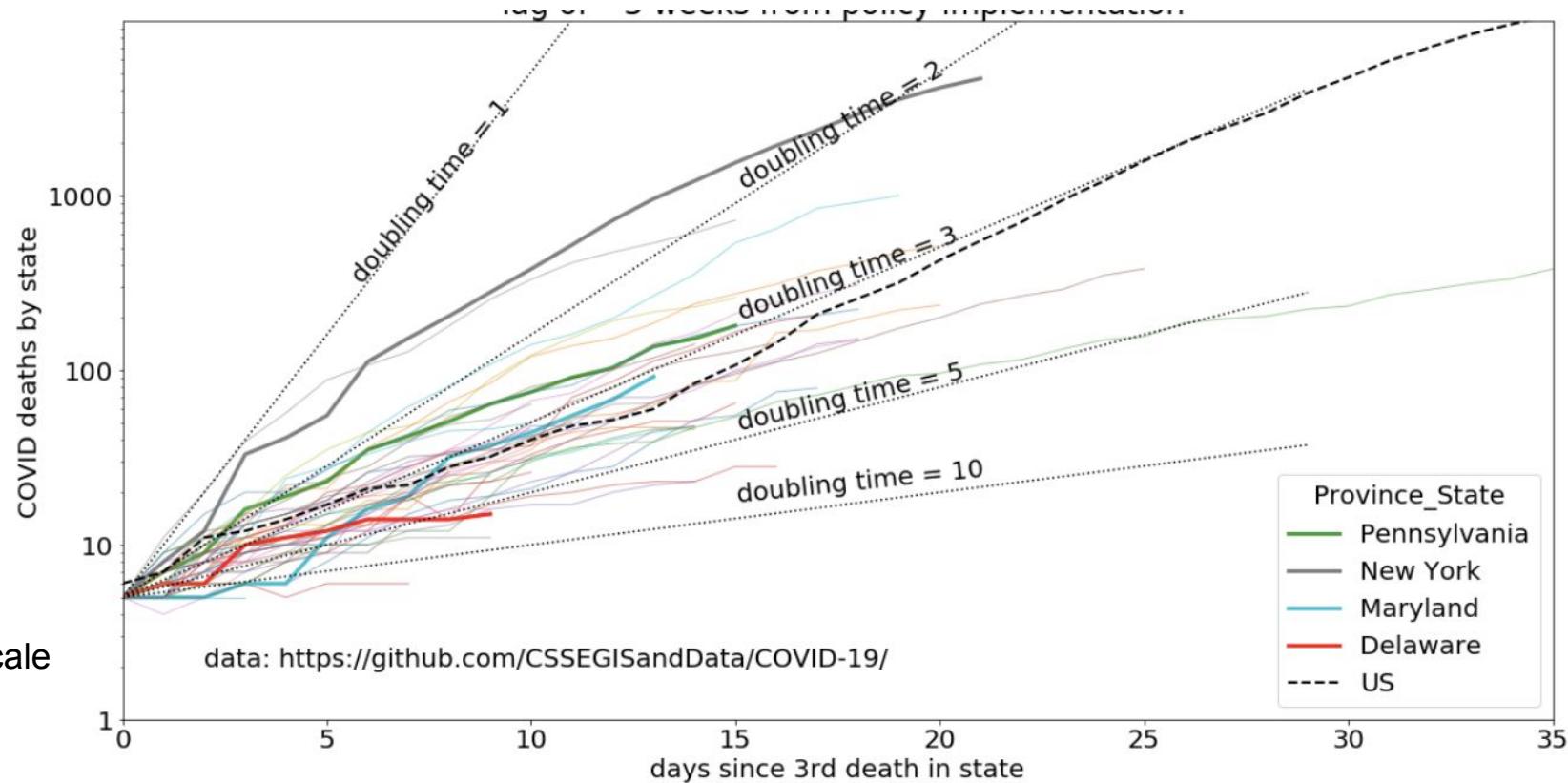
How would you improve these plots?



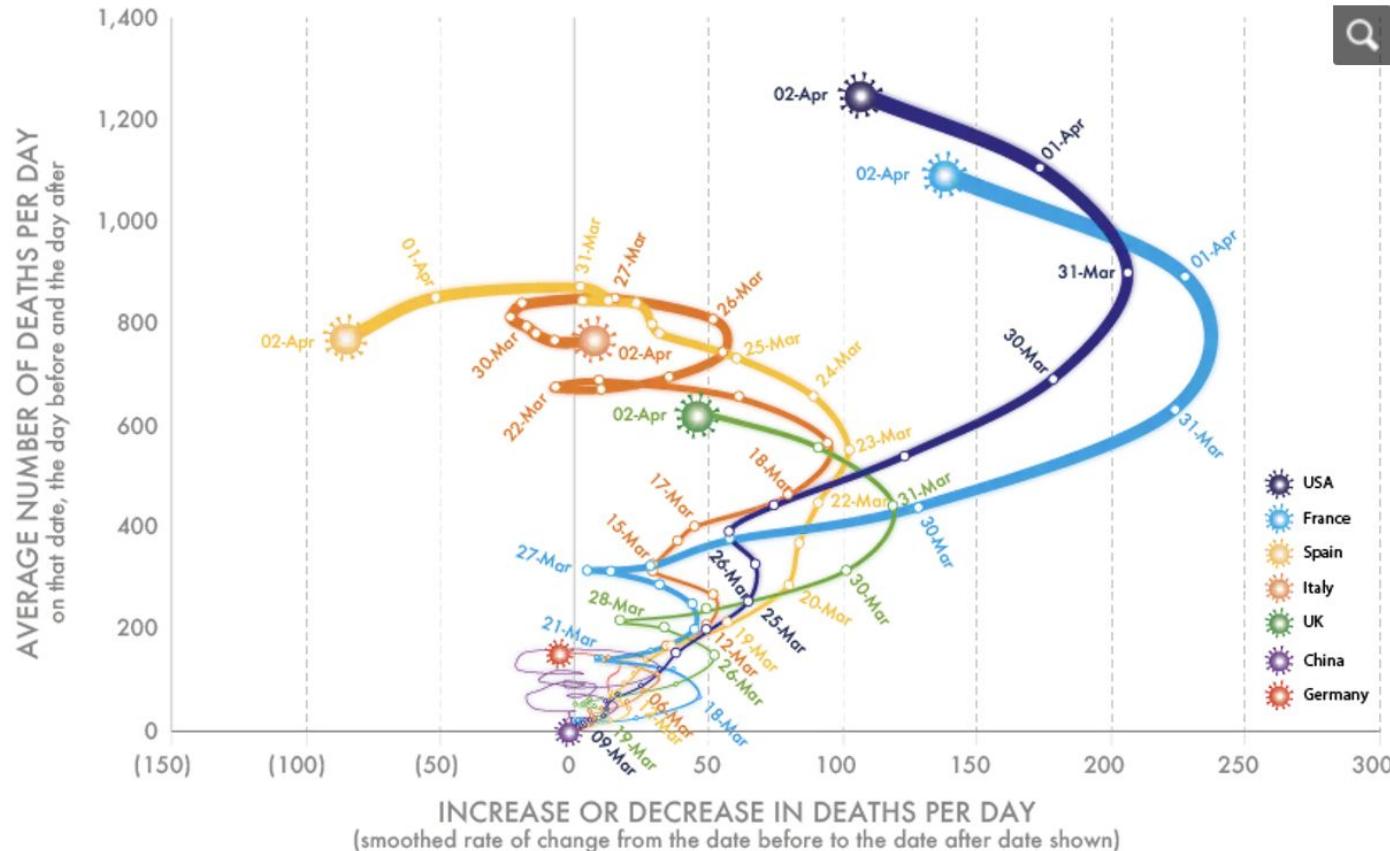
How would you improve these plots?



How would you improve these plots?

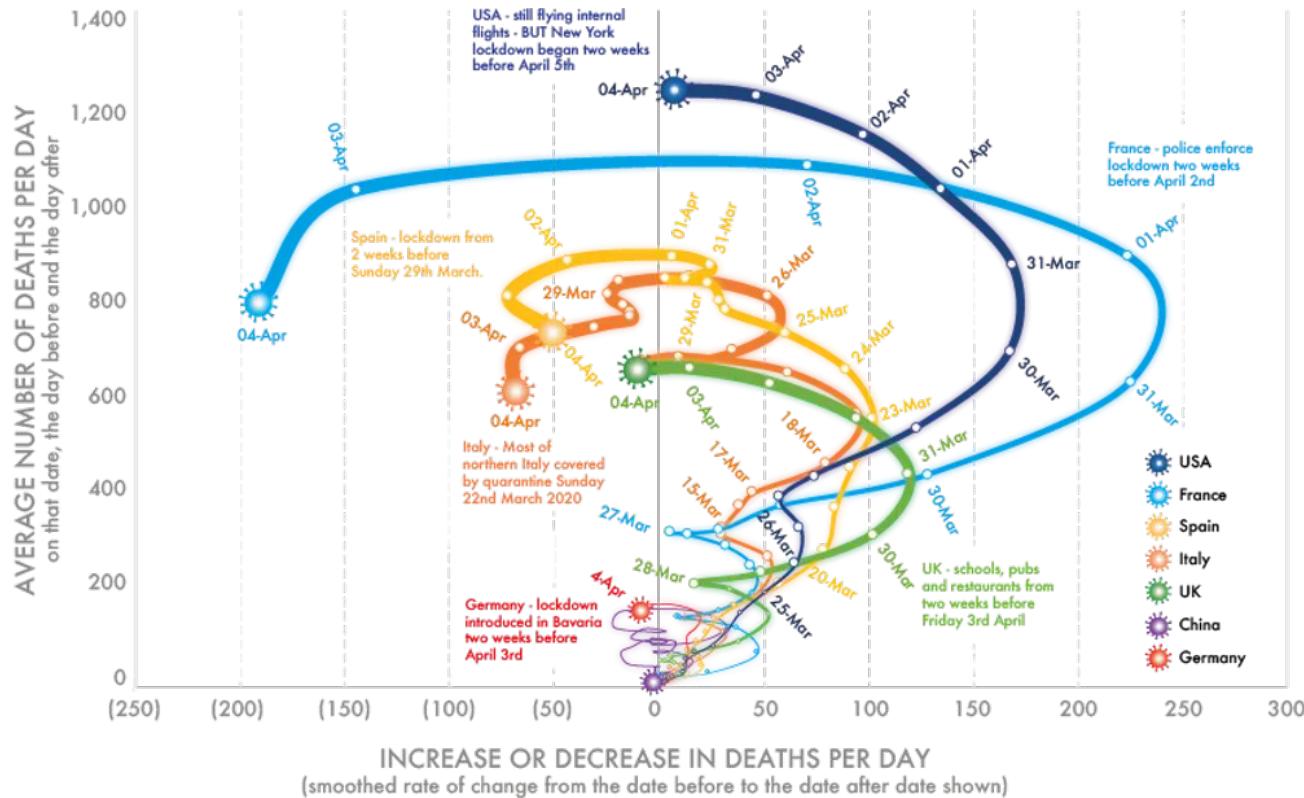


How would you improve these plots?

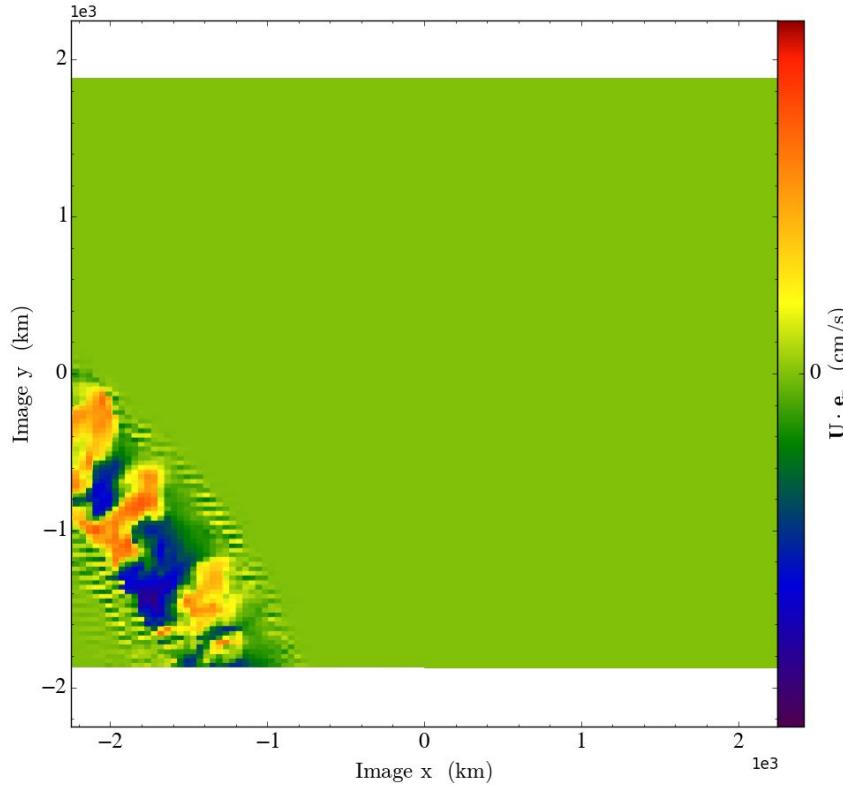


How would you improve these plots?

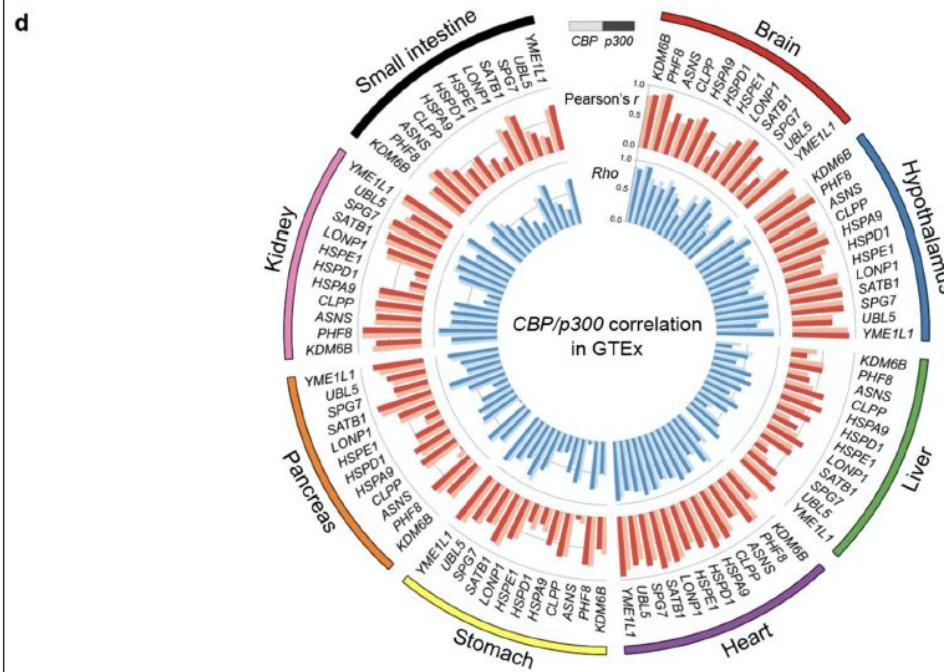
I would say this plot is at the limit of confusion (information saturation)



How would you improve these plots?



How would you improve these plots?



To conclude:

Beautiful is better than ugly.

Explicit is better than implicit.

Simple is better than complex.

Complex is better than complicated.

Flat is better than nested.

Sparse is better than dense.

Readability counts.