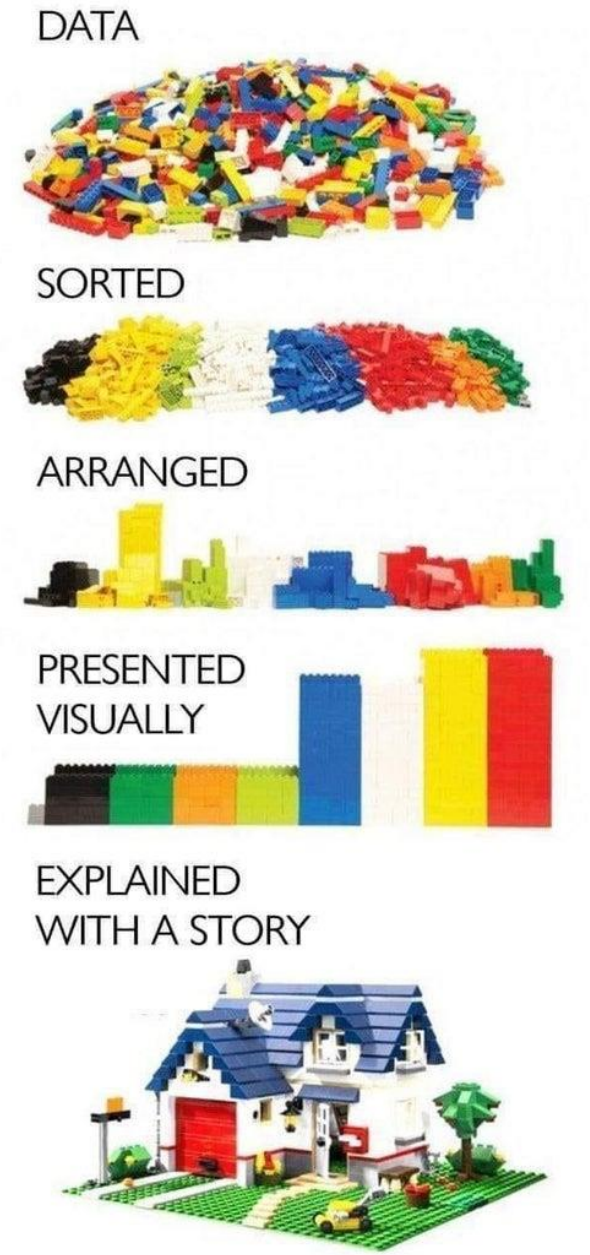
Library Training Series
SESSION 10:

Better Data Visualizations

Jess Newman, MSIS



"By visualizing information, we turn it into a landscape that you can explore with your eyes. A sort of information map. And when you're lost in information, an information map is kind of useful."

– David McCandless

Learning Objectives

- Recognize common data visualizations, their best uses, and their limitations
- Understand the importance of effective visualizations for communicating research results
- Become familiar with software and tools for data visualization available to researchers

What is data visualization?

Anything that converts data into a visual representation

```
>K00136:63:H3LCNBBXX:5:1101:1304:1439#0/1 t0000001
GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCT
>K00136:63:H3LCNBBXX:5:1101:2603:1439#0/1 t0000003
GCATTGGTGGTTCAGTGGTAGAATTCTCGCCT
>K00136:63:H3LCNBBXX:5:1101:5709:1439#0/1 t0000009
TACCCTGTAGAACCGAATTTGTGA
>K00136:63:H3LCNBBXX:5:1101:5831:1439#0/1 t0000002
TACCCTGTAGAACCGAATTTGT
>K00136:63:H3LCNBBXX:5:1101:7496:1439#0/1 t0000001
GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCT
>K00136:63:H3LCNBBXX:5:1101:9221:1439#0/1 t0000010
GCTTCTGTAGTGTAGTGGTTATCACGTTCGCCT
>K00136:63:H3LCNBBXX:5:1101:9830:1439#0/1 t0000002
TACCCTGTAGAACCGAATTTGT
>K00136:63:H3LCNBBXX:5:1101:9850:1439#0/1 t0000001
GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCT
>K00136:63:H3LCNBBXX:5:1101:10155:1439#0/1 t0000011
TCGGGCCTGGTTAGTACTTGGATGGGAGACCGCC
>K00136:63:H3LCNBBXX:5:1101:10317:1439#0/1 t0000004
GCATTGGTGGTTCAGTGGTAGAATTCTCGCC
```

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe UUC UUA Leu UUG	UCU UCC UCA Ser UCG	UAU Tyr UAC UAA STOP UAG STOP	UGU Cys UGC UGA STOP UGG Trp	U C A G
C	CUU CUC CUA Leu CUG	CCU CCC CCA Pro CCG	CAU His CAC CAA Gln CAG	CGU CGC CGA Arg CGG	U C A G
A	AUU AUC AUA Ile AUG Met	ACU ACC ACA Thr ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU GUC GUA Val GUG	GCU GCC GCA Ala GCG	GAU Asp GAC GAA Glu GAG	GGU GGC GGA Gly GGG	U C A G

Including tables!

Why visualize?
(Anscombe's Quartet)

These groups have almost identical summary statistics at first glance.

- X & Y mean
- X & Y variance
- X-Y correlation
- X-Y linear regression

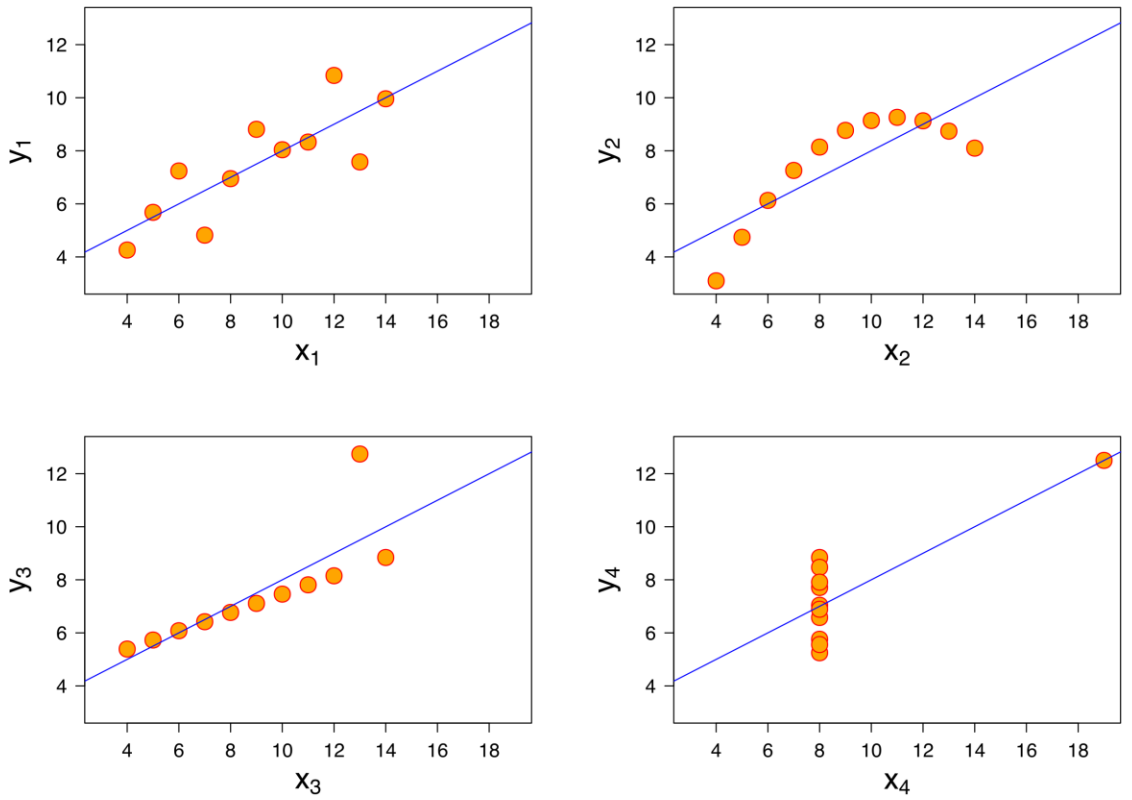
1		2		3		4	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	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

https://en.wikipedia.org/wiki/Anscombe's_quartet

But when plotted, we see patterns

1		2		3		4	
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

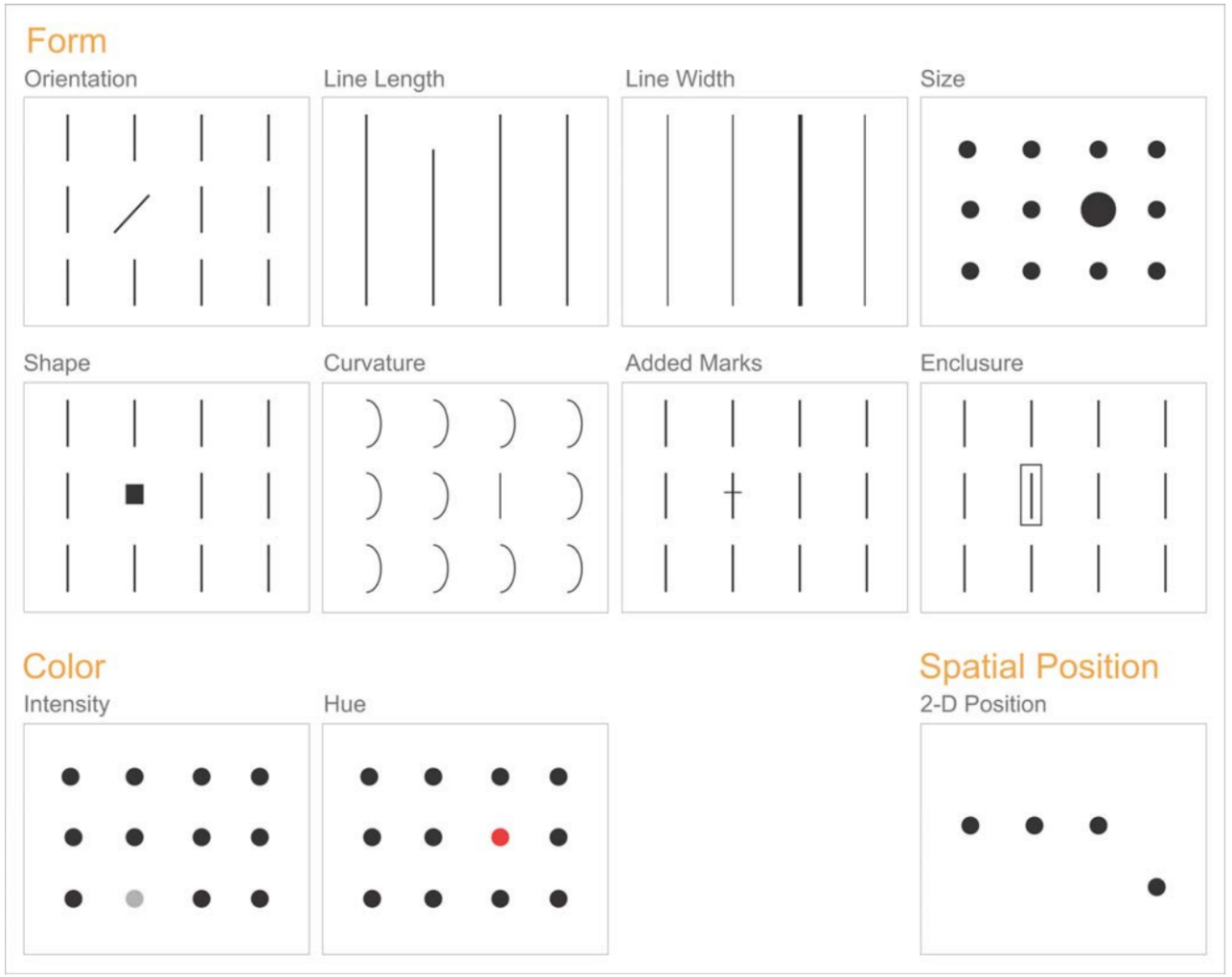
VS



https://en.wikipedia.org/wiki/Anscombe's_quartet

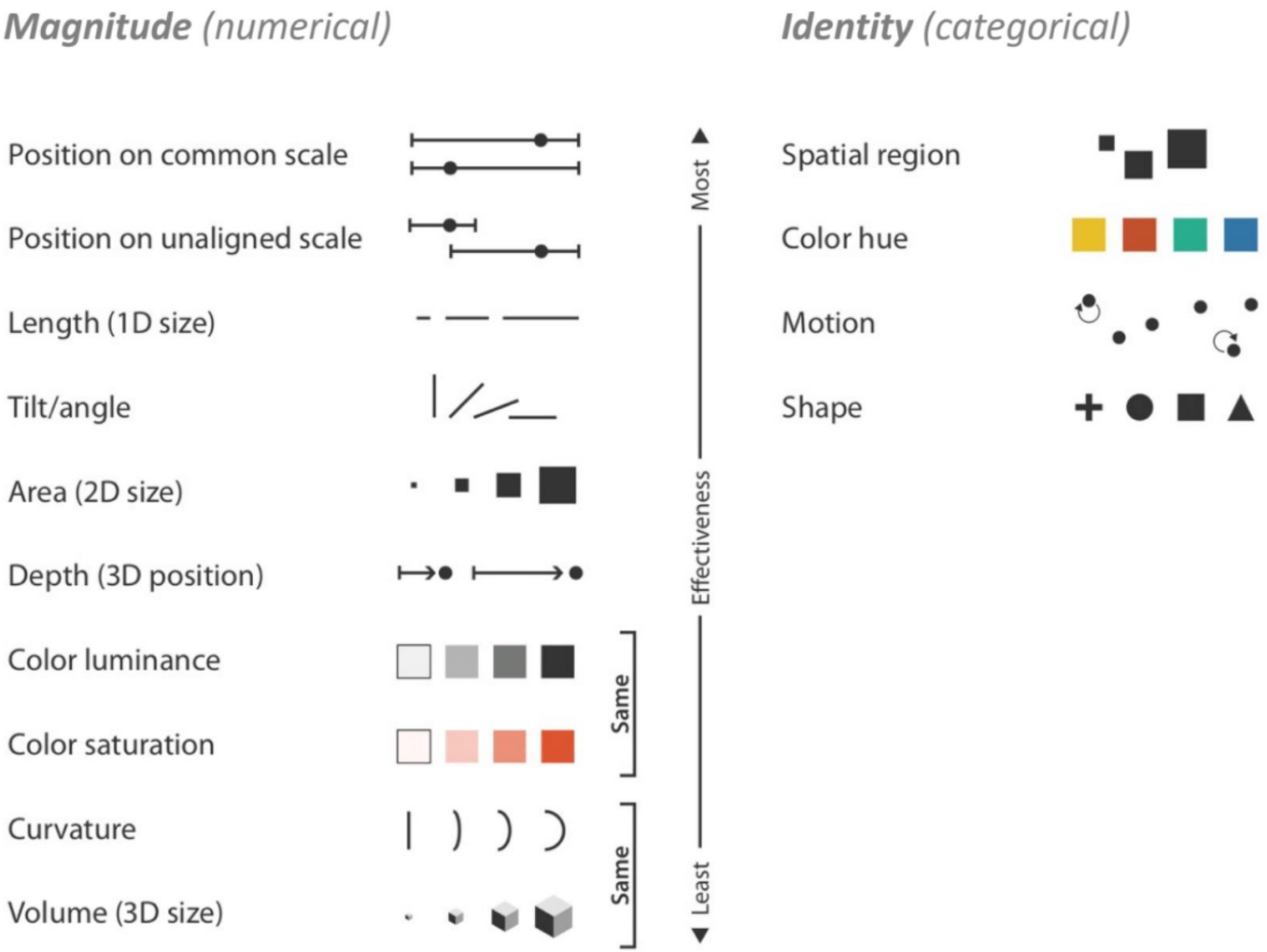
https://flat.io/score/60a8d8653374193bc2aa3633-anscombe-s-quartet

Visual cues used in data visualization make interpretation even quicker



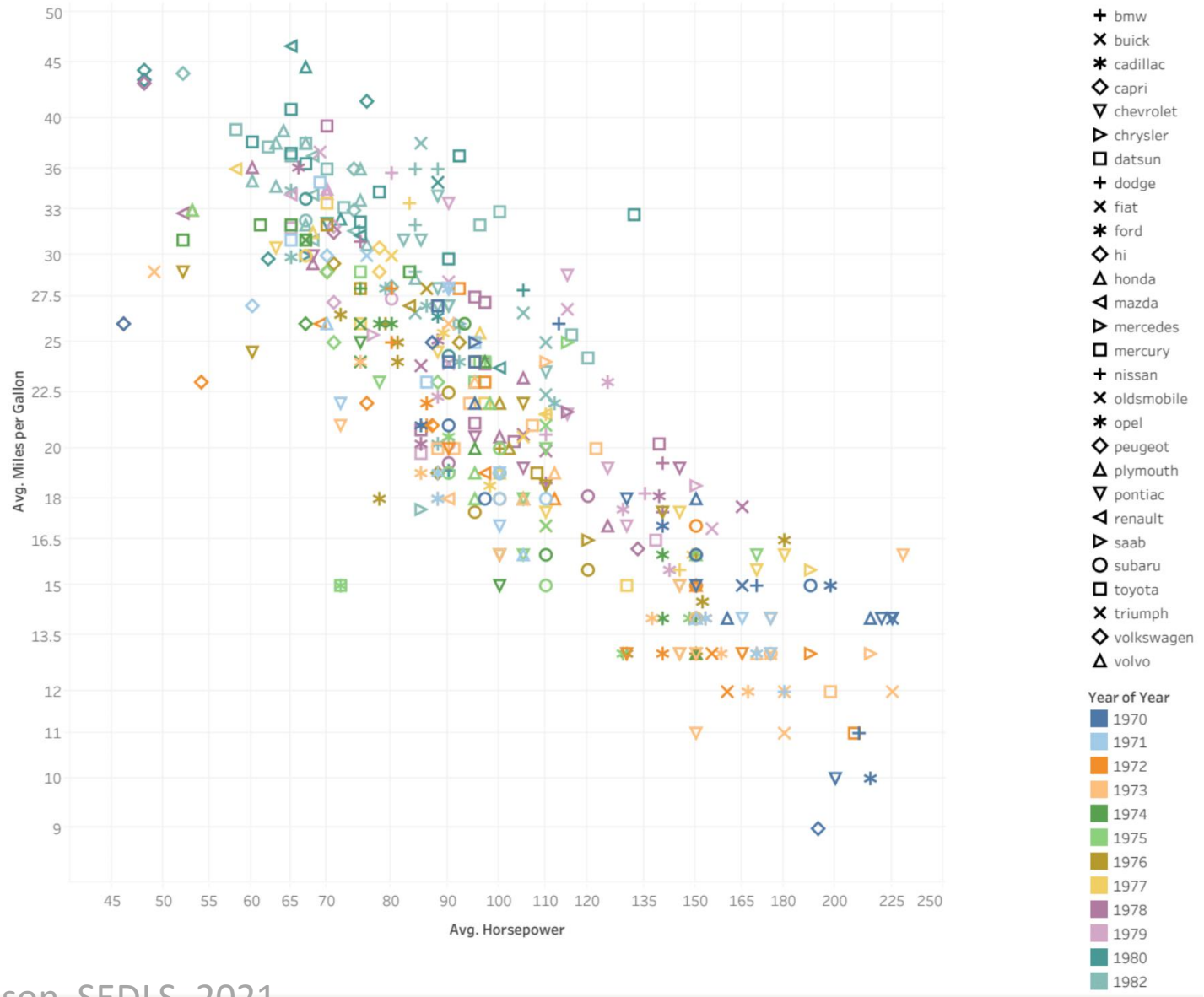
https://www.perceptualedge.com/articles/ie/visual_perception.pdf

Some cues are more effective than others



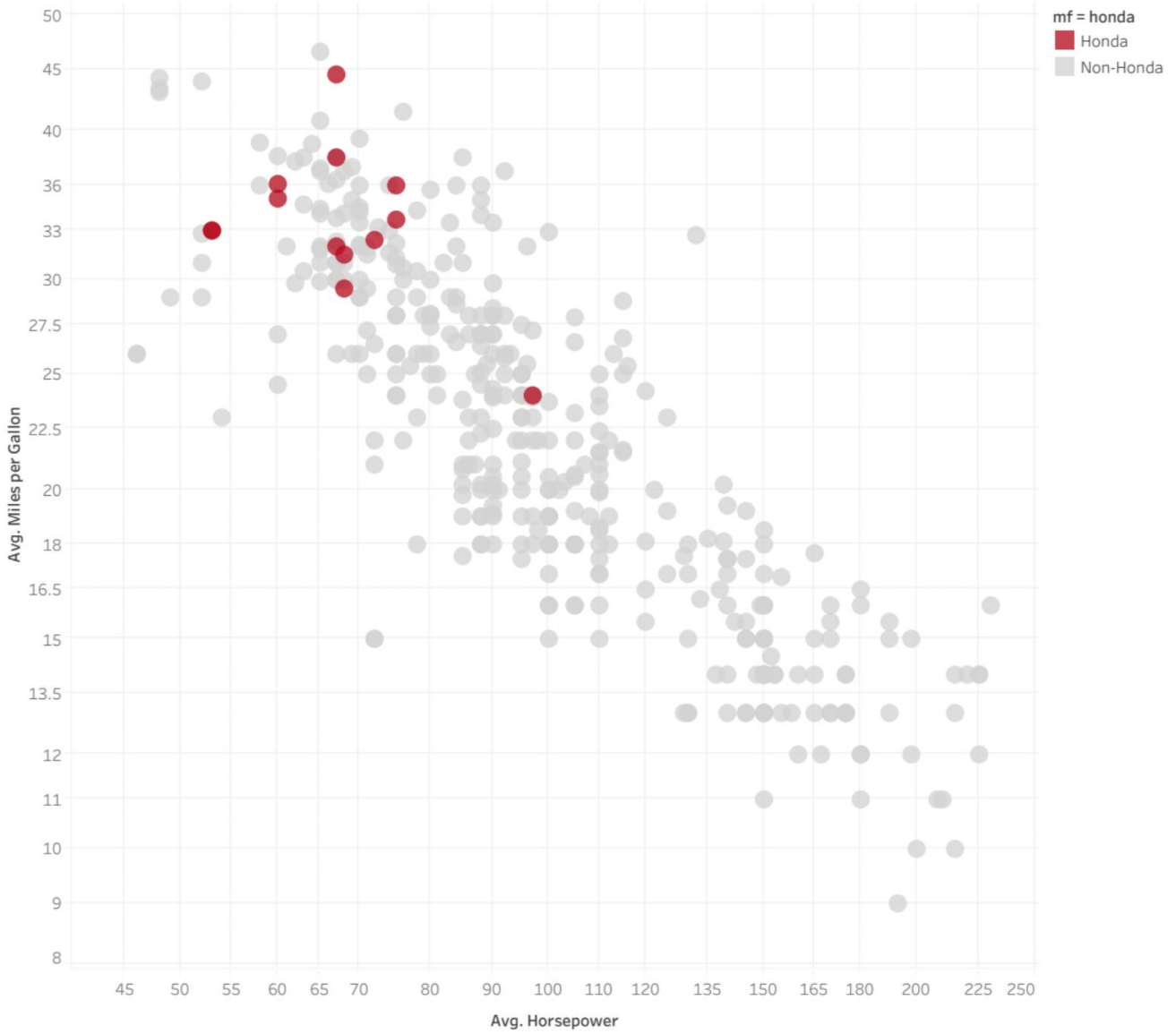
https://www.cs.ubc.ca/~tmm/vadbook/eamonn-figs/fig5.1.pdf

Too many cues
=
Visual overload



"Tips for Effective Data Visualization." Eric E Monson. SEDLS, 2021.

Picking a story or focus is much more effective

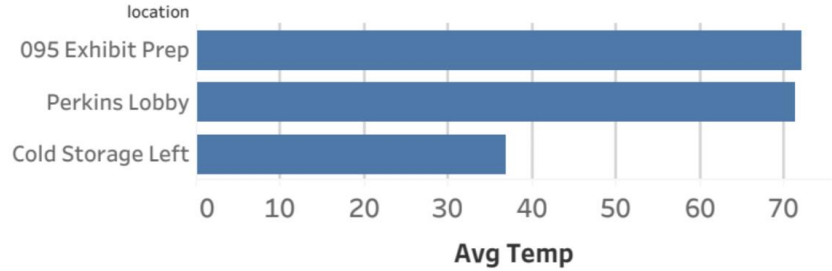


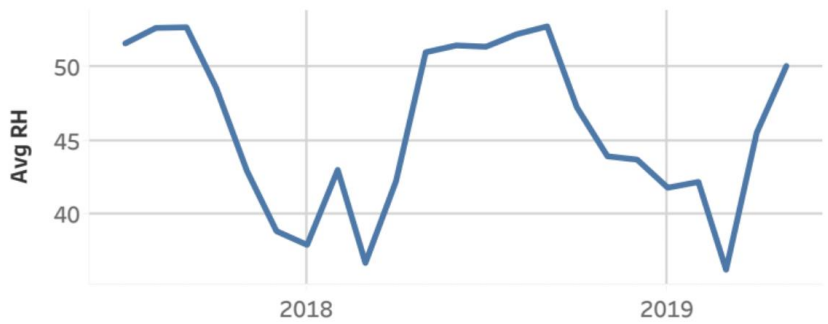
"Tips for Effective Data Visualization." Eric E Monson. SEDLS, 2021.

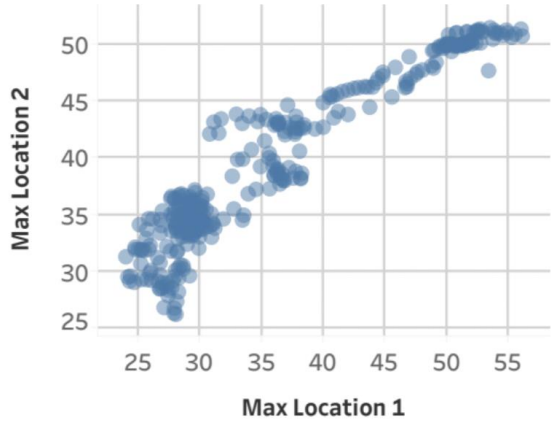
You have your data, now what?

Start here with your visualization:

- Category + Number = Bar
- Date/Time + Number = Line
- Number + Number (correlation) = Scatter





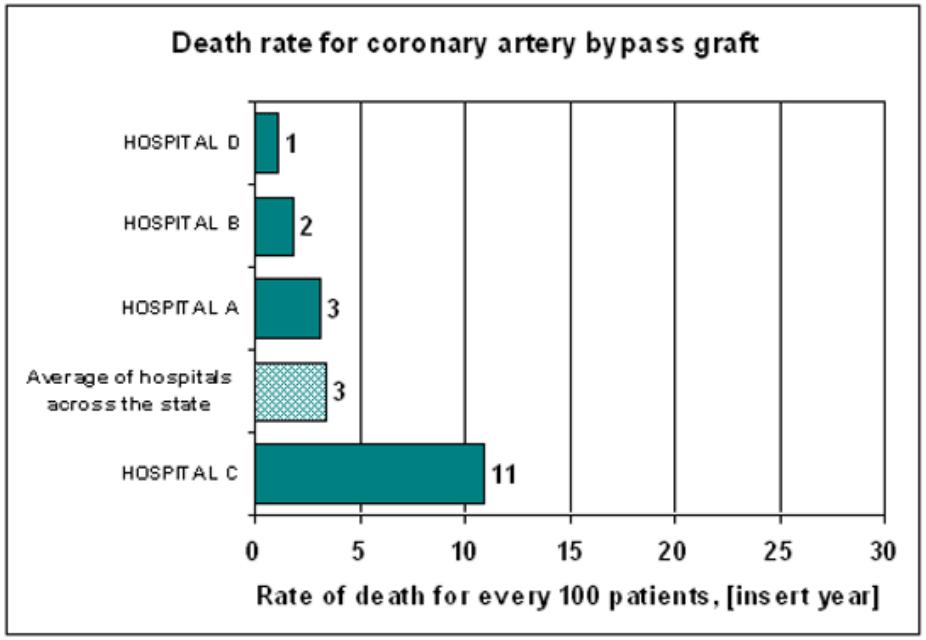


"Tips for Effective Data Visualization." Eric E Monson. SEDLS, 2021.

Let's learn about bar graphs

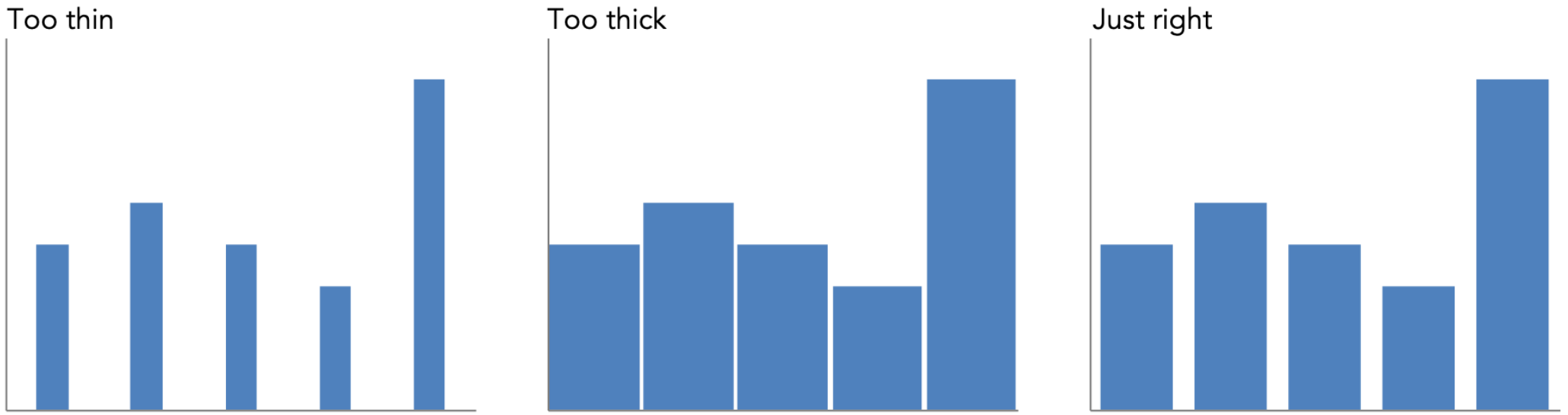
Bar graphs are great for categorical data

- Useful for comparing values across categories (categorical data)
- Easily understood
- Don’t use with paired or non-independent data



https://pubmed.ncbi.nlm.nih.gov/28974579/

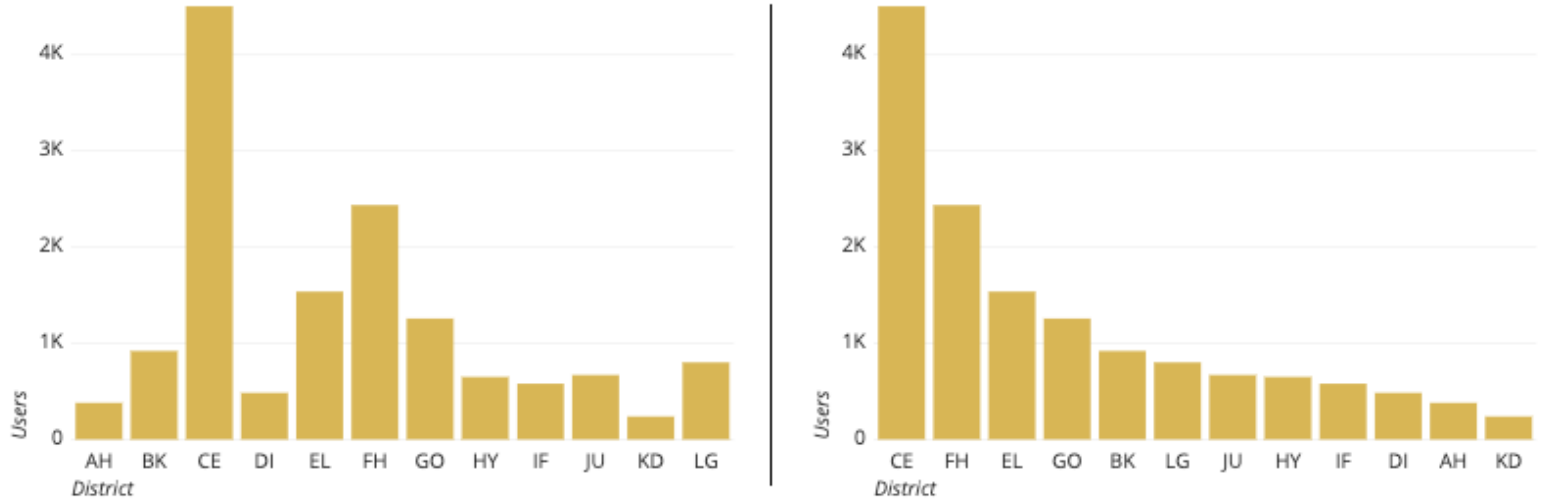
Tip: Mind the spacing between bars



https://www.storytellingwithdata.com/blog/2020/2/19/what-is-a-bar-chart

Tip: Add order if possible

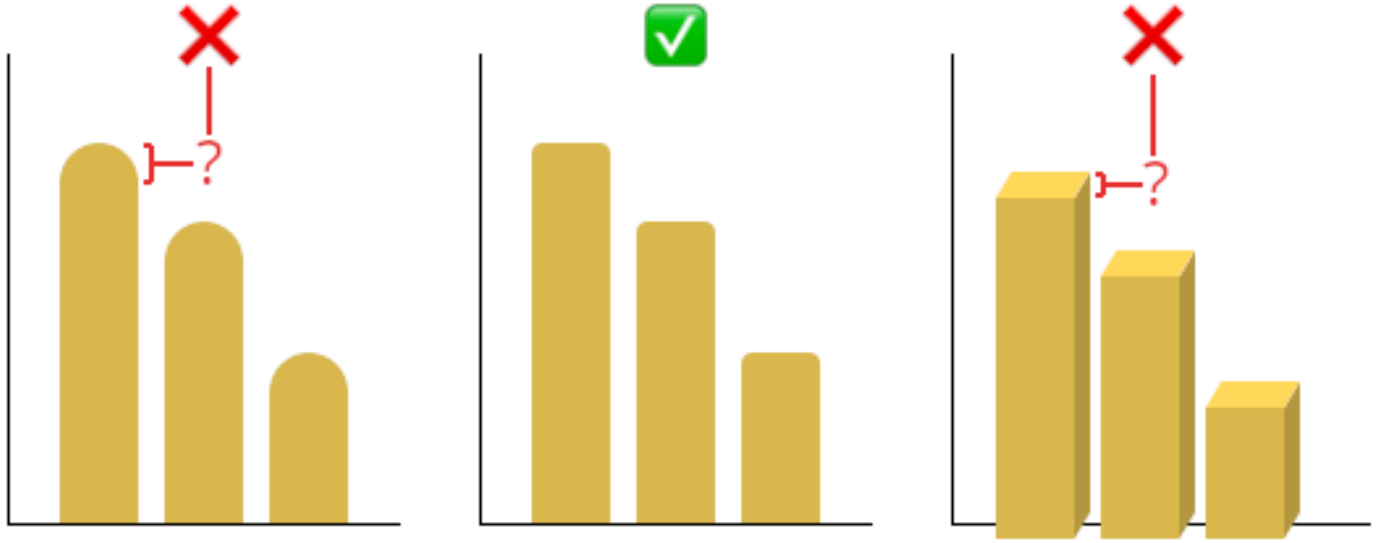
- If variables are not inherently ordered (e.g. date), consider ordering by value



https://chartio.com/learn/charts/bar-chart-complete-guide/

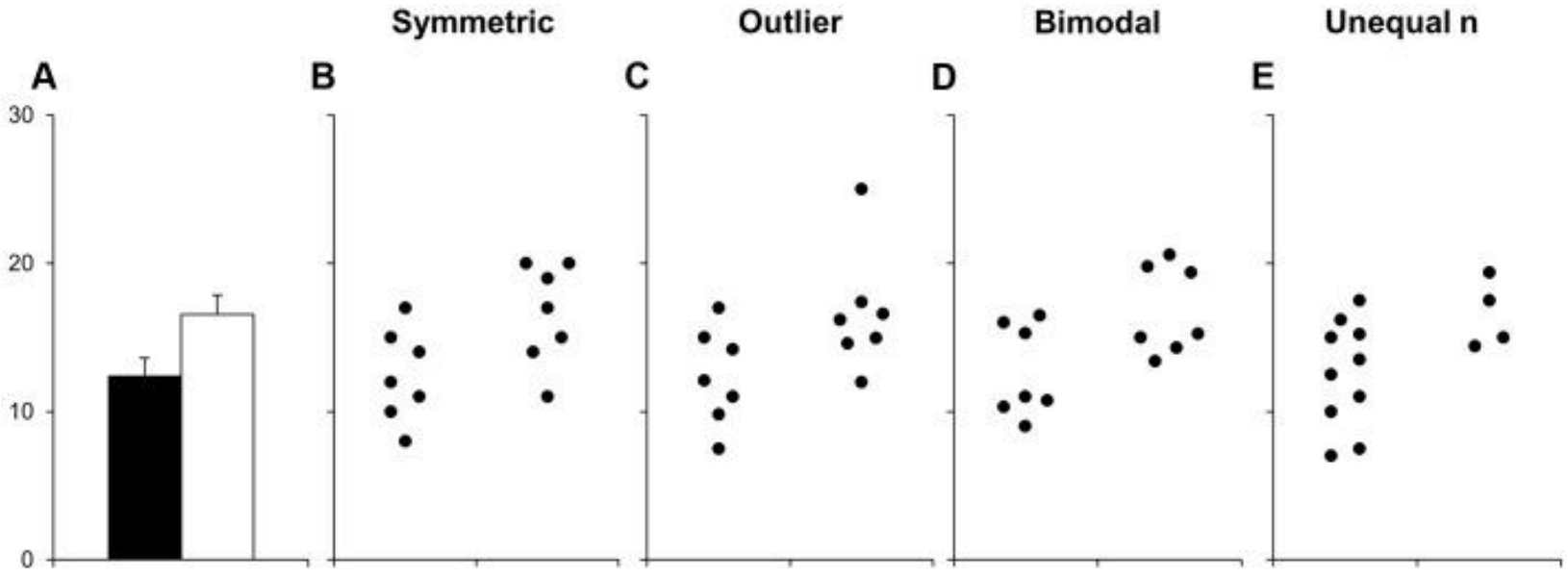
Tip: Use rectangular bars

- 3D bars, rounded bars, and other shapes can make it difficult to accurate assess height and may imply volume



https://chartio.com/learn/charts/bar-chart-complete-guide/

Caution: Many datasets can lead to the same bar graph

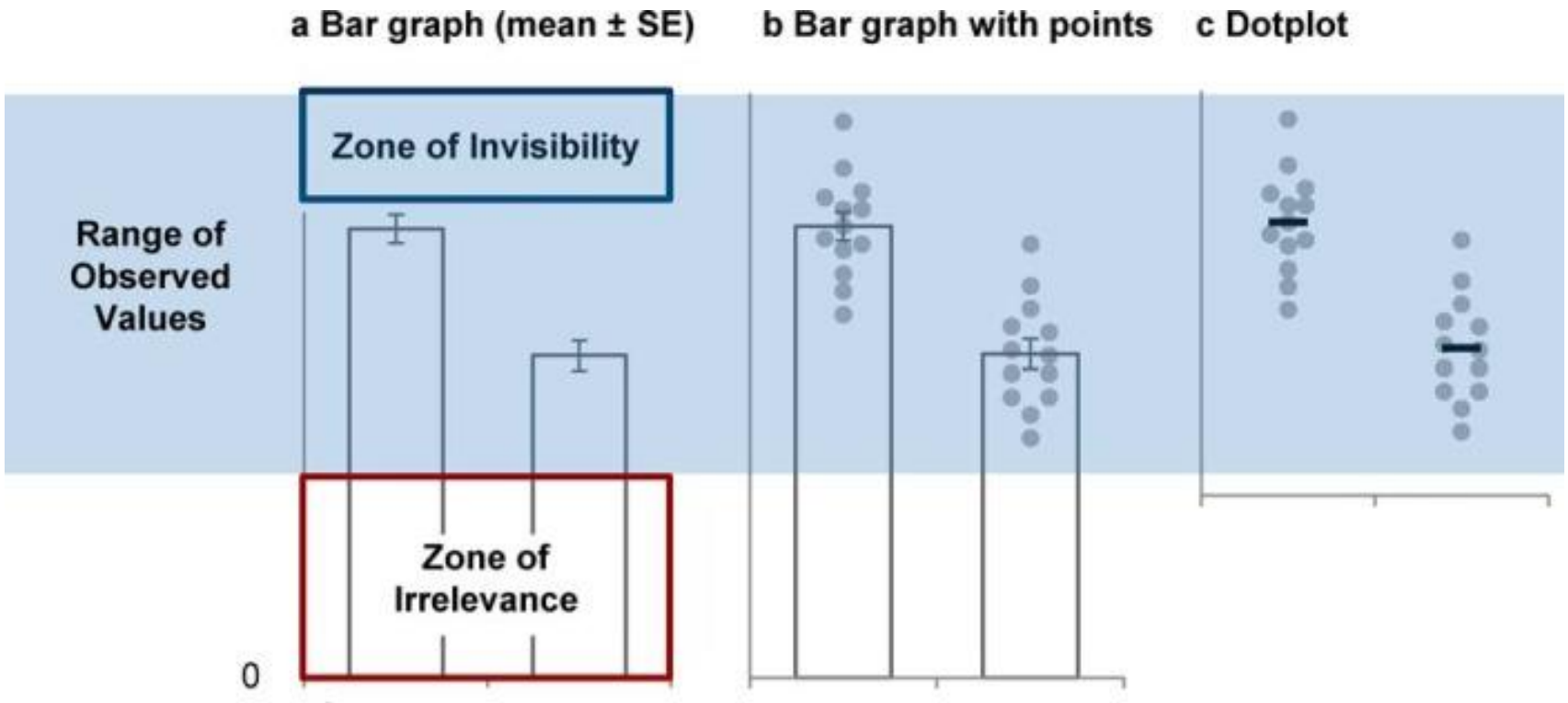


Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

https://doi.org/10.1371/journal.pbio.1002128

Scatter plots for numerical variables

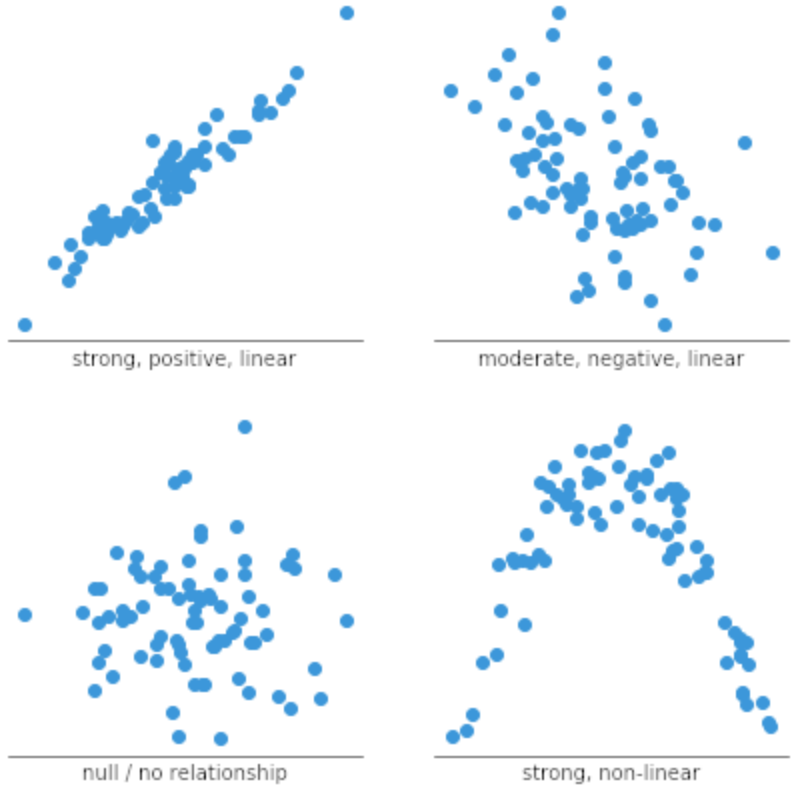
Scatter plots may reveal data that is hidden by bar charts



https://pubmed.ncbi.nlm.nih.gov/28974579/

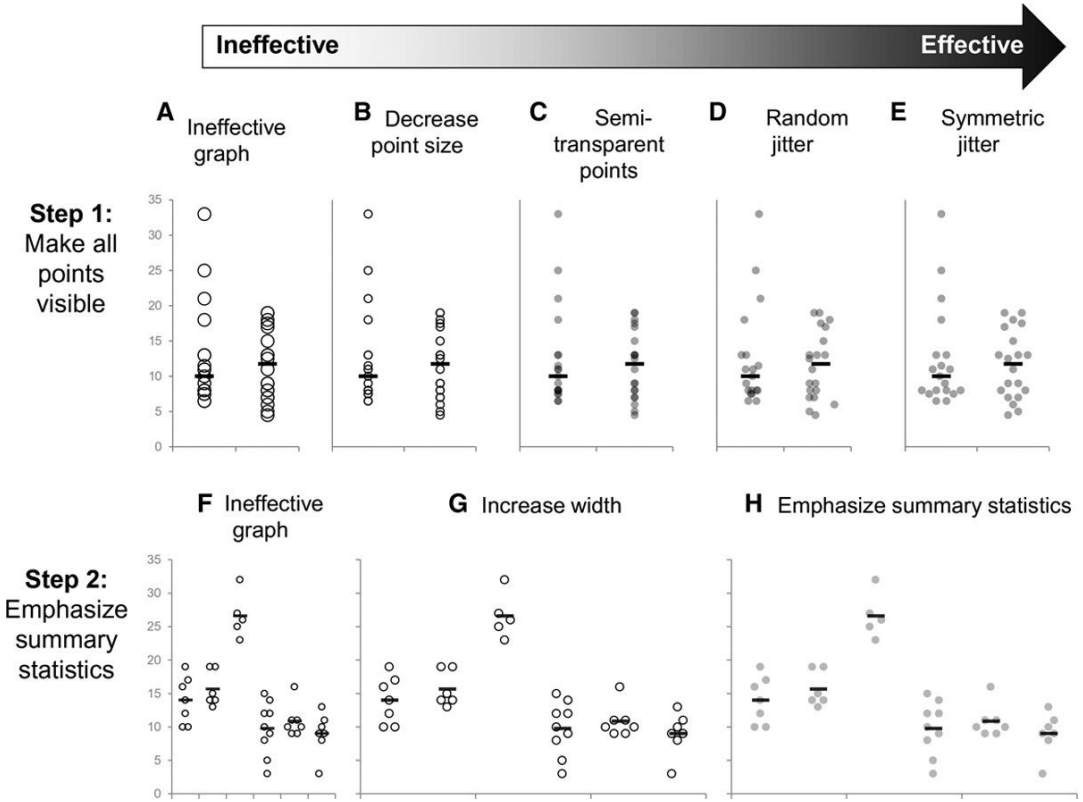
Scatter Plots

- Best Use
  - Identifying patterns in data (groups, gaps, outliers?)
  - Illustrate relationship(s) between two *numeric* variables
  - Continuous scale data
- Be Wary
  - With categorical data
  - Too few values in one variable



https://chartio.com/learn/charts/what-is-a-scatter-plot/

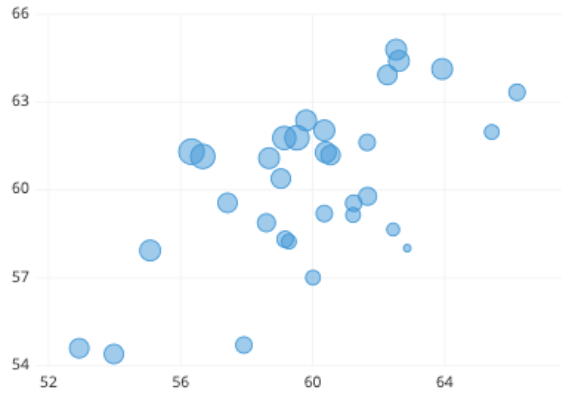
Caution: Too many values in one variable may lead to overplotting

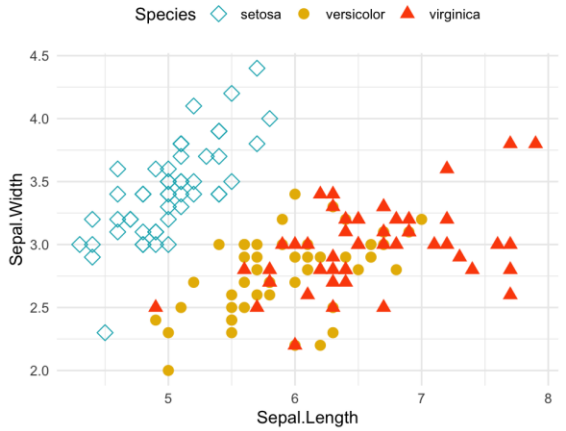


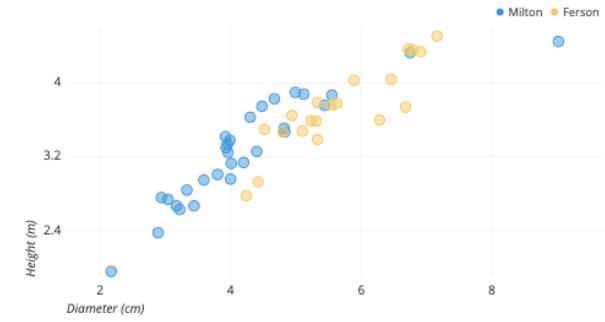
https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.037777

Scatter plots may have a categorical third variable

- You may use visual cues (color, shape, size) to encode a third variable, but be careful of becoming too busy







https://chartio.com/learn/charts/what-is-a-scatter-plot/

Quick note on pie charts

… Avoid them

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

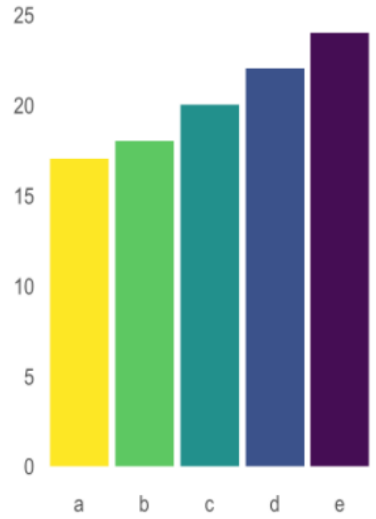
They might also be hiding patterns

A

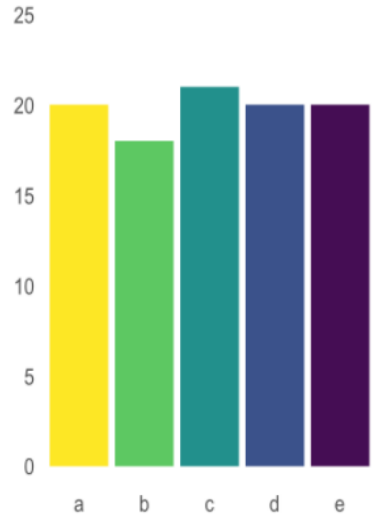
B

C

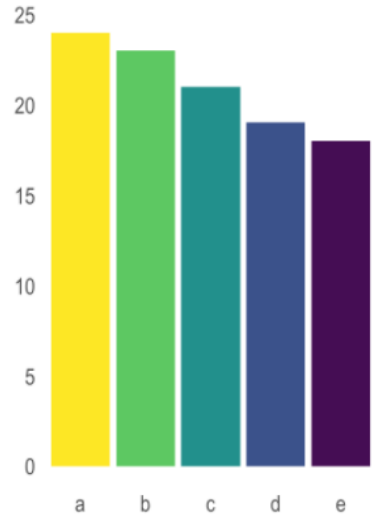
A



B



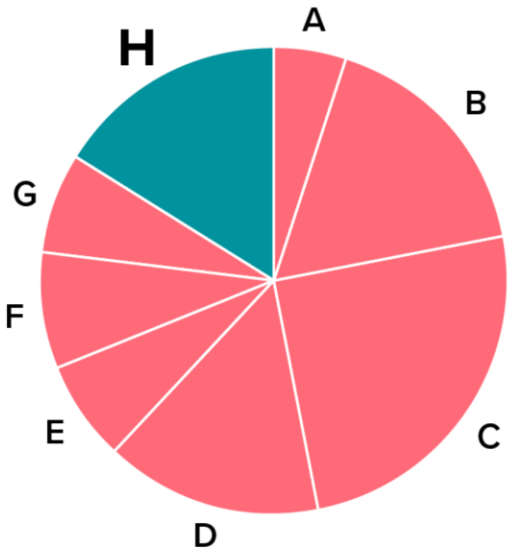
C



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

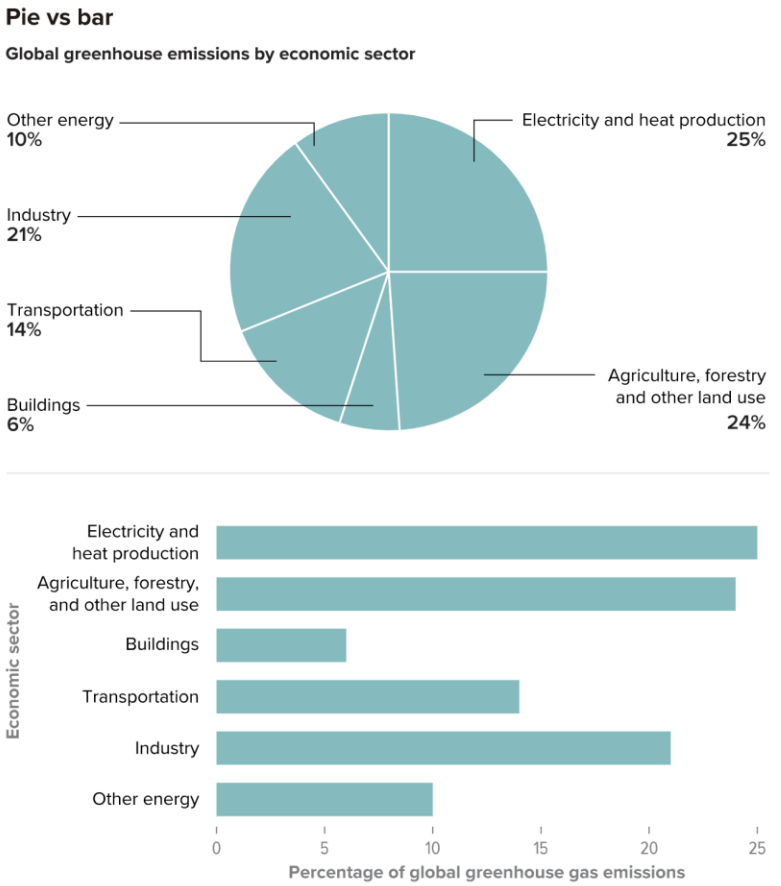
We aren't great at estimating volume of a circle

(And they often aren't used for their intended purposes - to represent parts of a whole)



Which is the third largest segment in the pie chart?

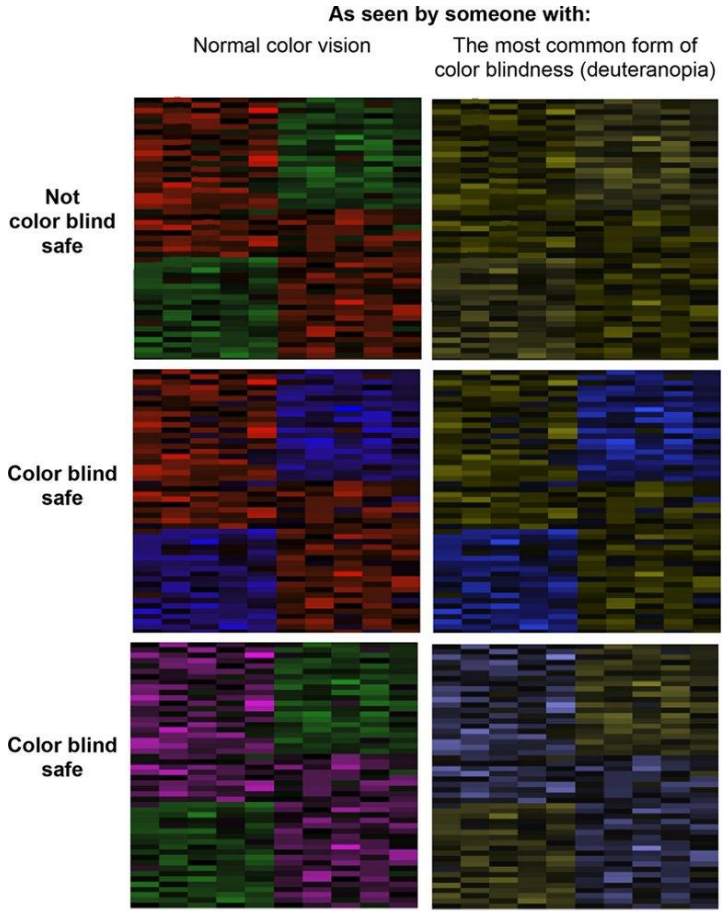
- A
- H
- B
- D

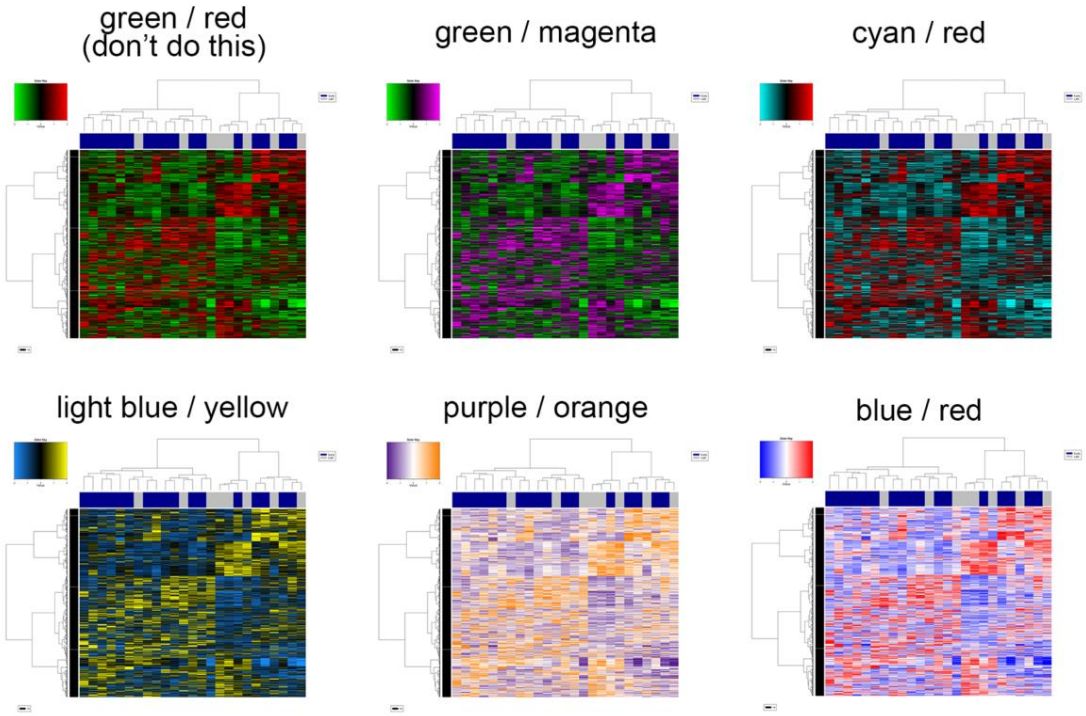


Try a bar chart instead!

Color and accessibility considerations

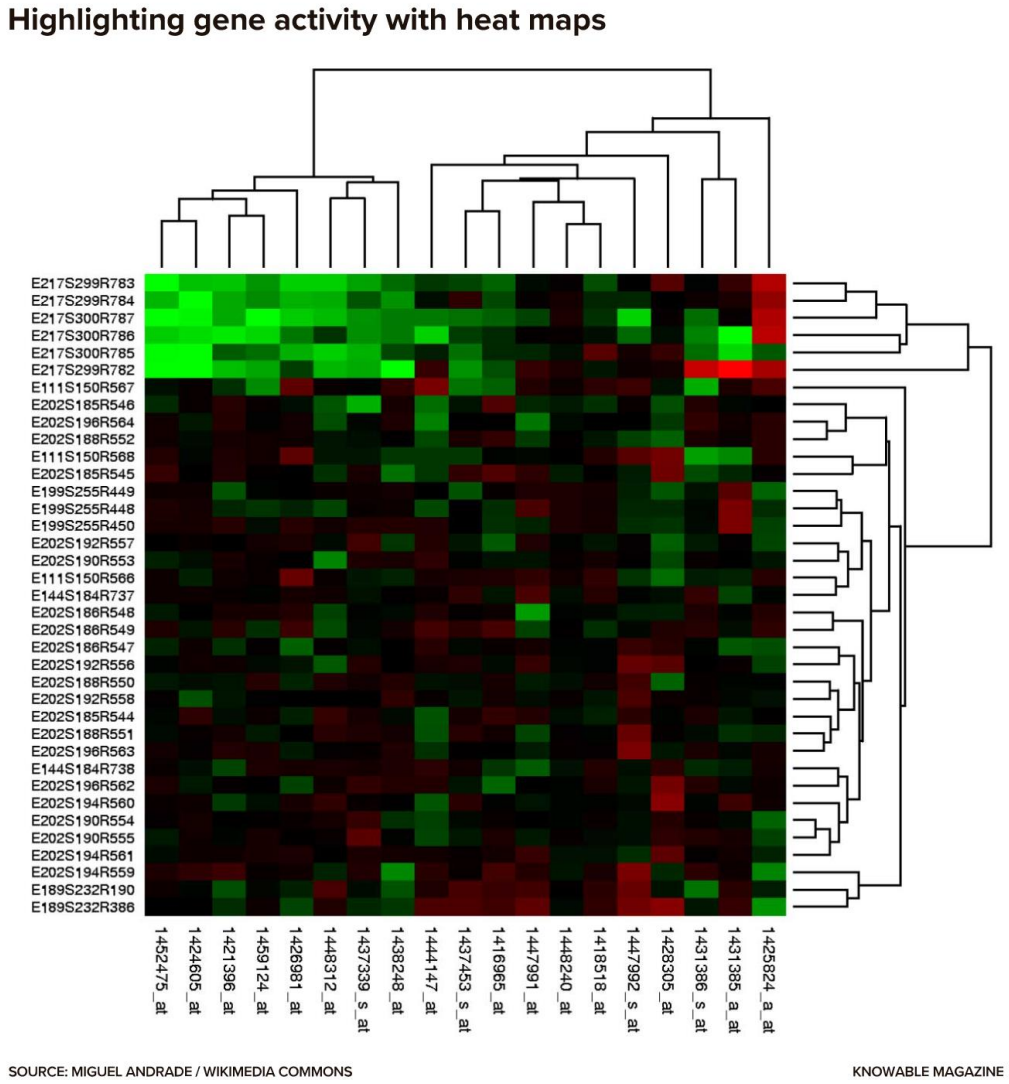
Colorblind safe: Avoid red and green





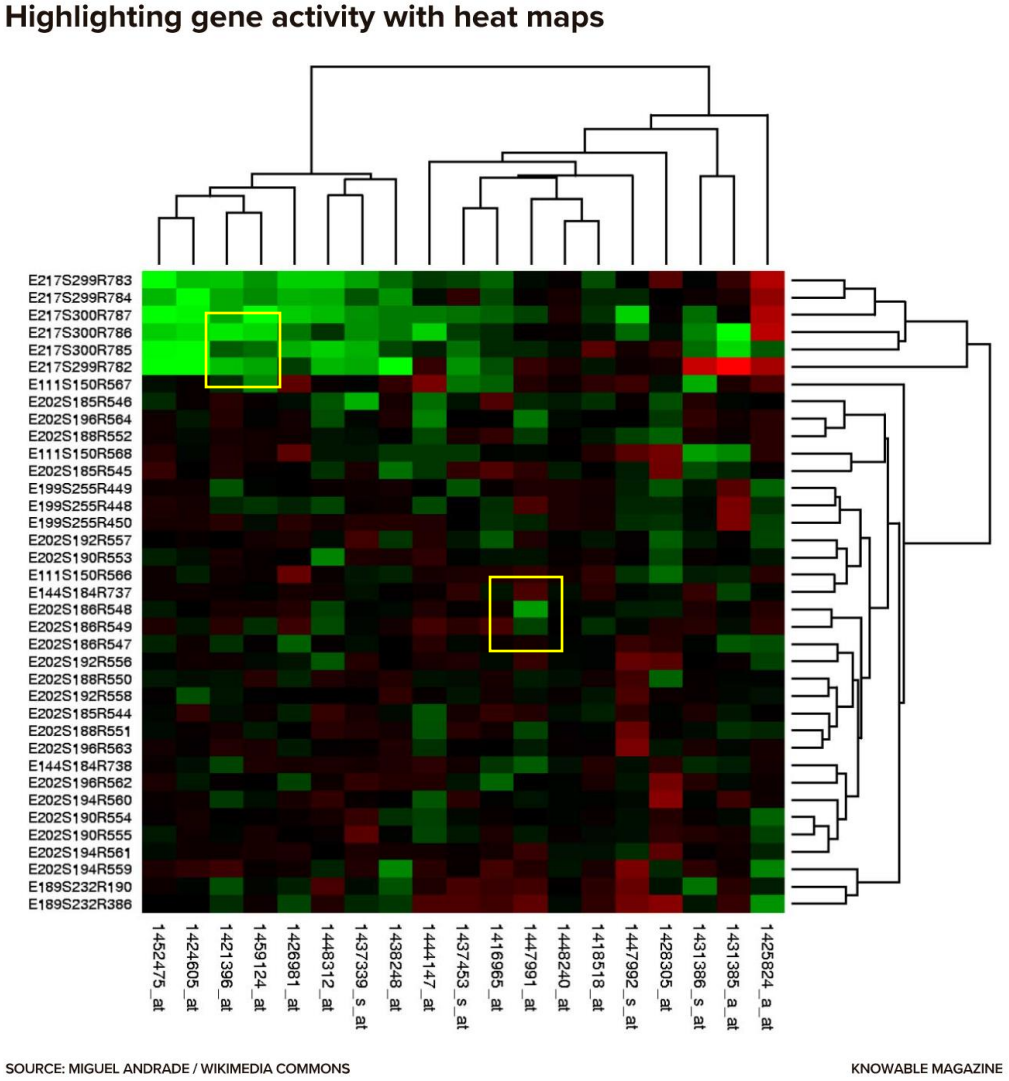
https://www.ascb.org/science-news/how-to-make-scientific-figures-accessible-to-readers-with-color-blindness/

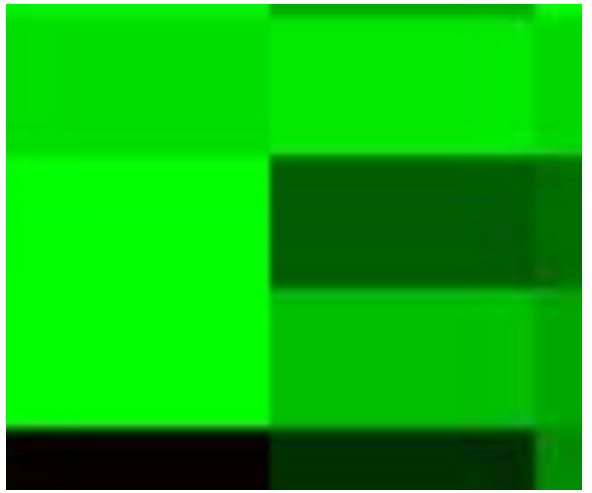
Relative contrast can create distortion

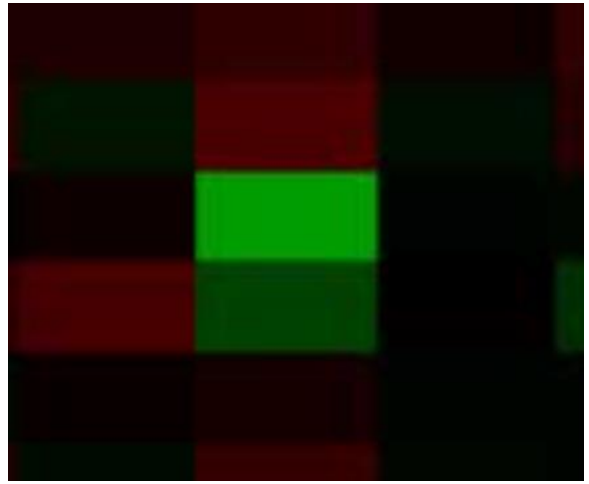


https://knowablemagazine.org/article/mind/2019/science-data-visualization

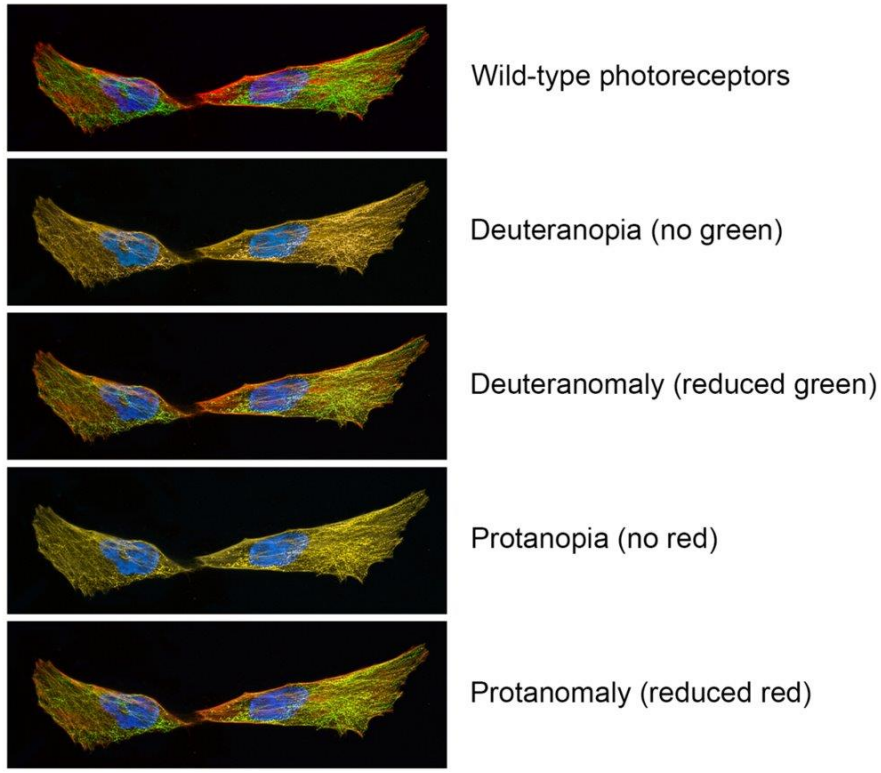
Relative contrast can also create distortion



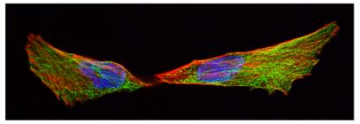




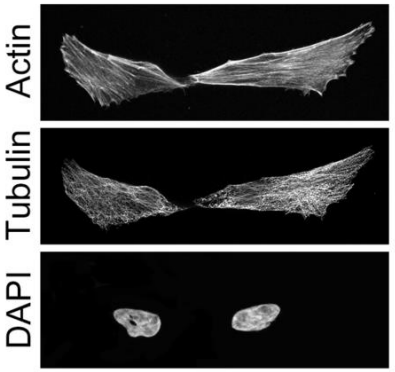
Alternatives to red and green in imaging



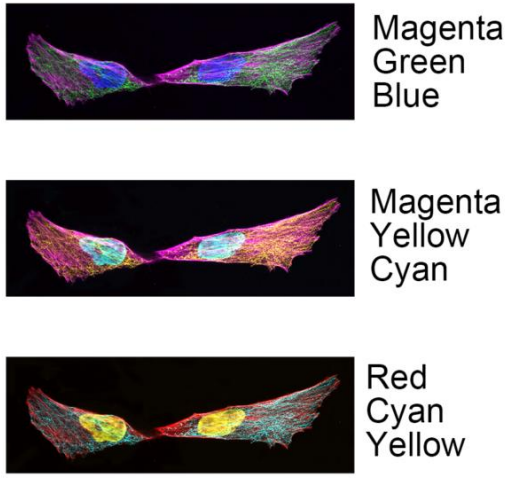
DON'T
Use red and green pseudocoloring in the same image



DO
Show greyscale images of each channel



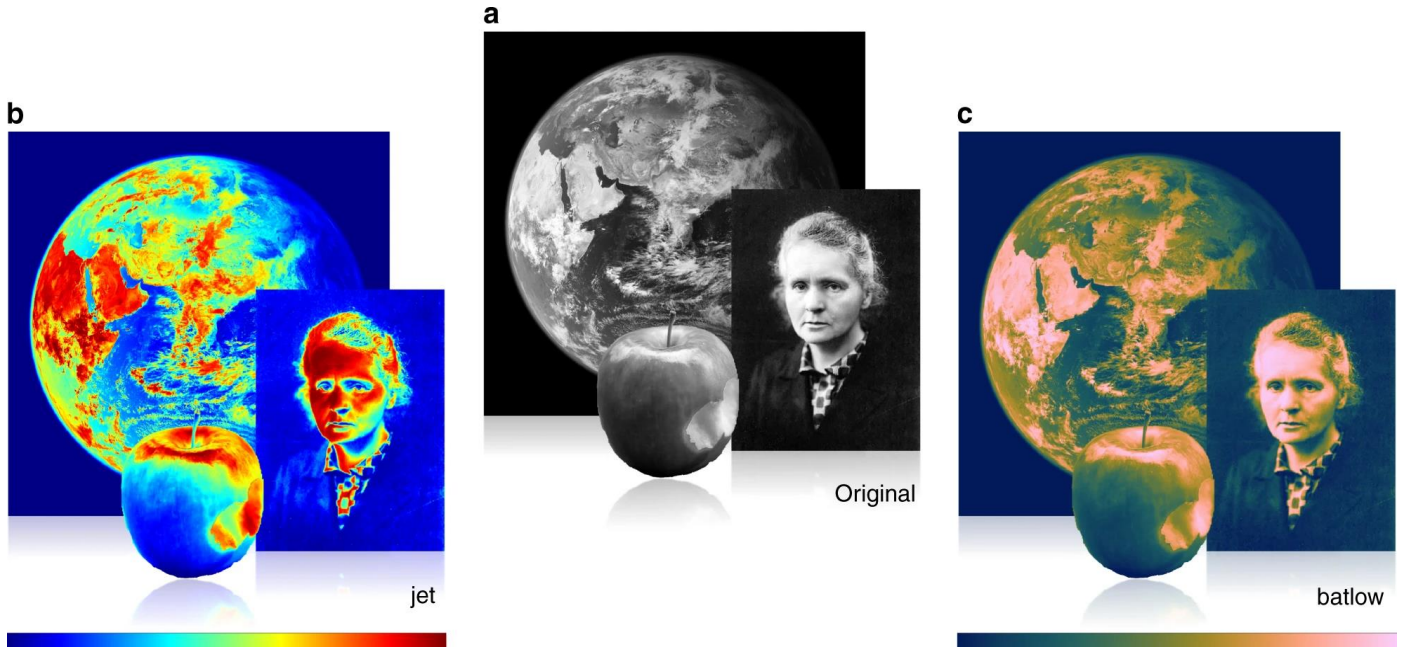
DO
Use colors in merged images that can still be distinguished by people with red/green color-blindness

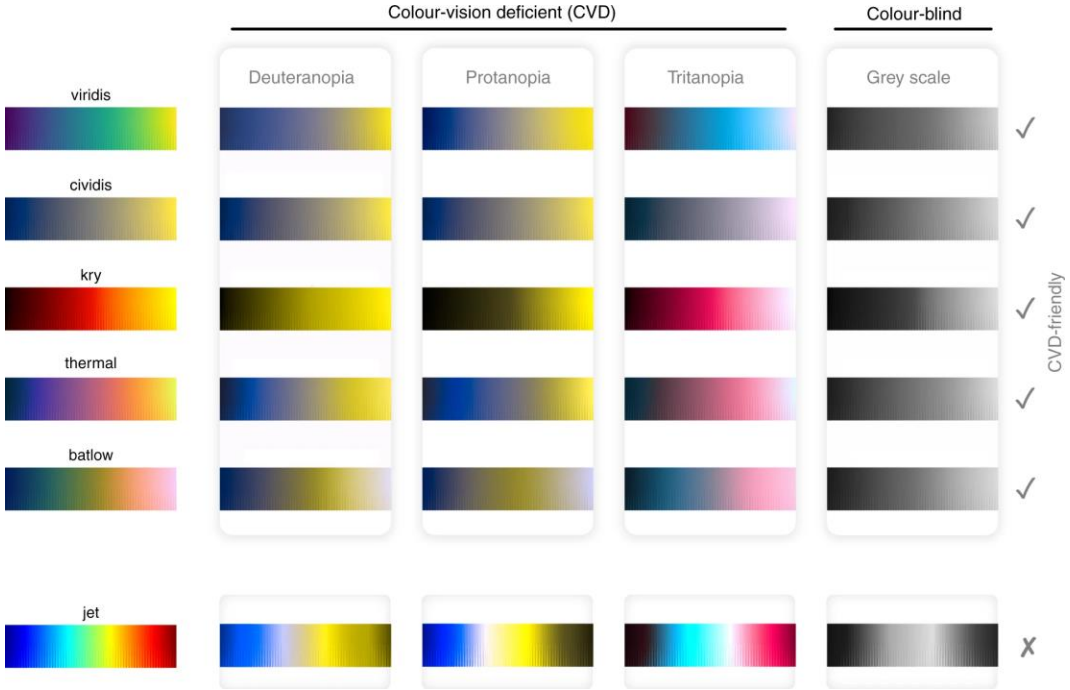


ColorBrewer:
tool for picking color palettes

https://www.ascb.org/science-news/how-to-make-scientific-figures-accessible-to-readers-with-color-blindness/

Jet ('Rainbow') colormaps also create distortion

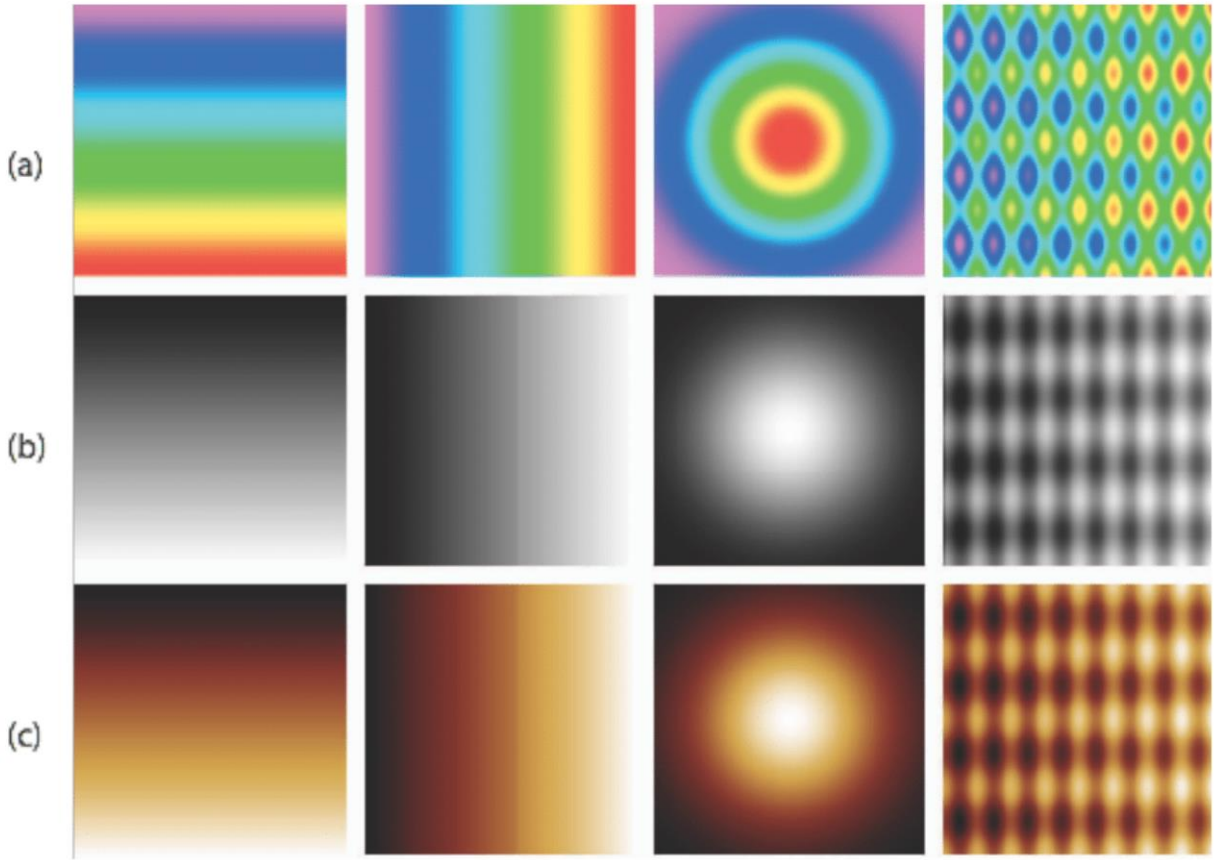




https://doi.org/10.1038/s41467-020-19160-7

A bit more about rainbow maps…

- Have no intuitive color ordering
- Make data look striped / banded



Borland, David, and Russell M. Taylor Ii. "Rainbow color map (still) considered harmful." *IEEE computer graphics and applications* 27.2 (2007). https://ieeexplore.ieee.org/document/4118486

Choose categorical colors carefully

- Only use gradient color for ordered categories (implies relation)
- Use the same color for the same variables if you have multiple visuals

Not ideal

Better

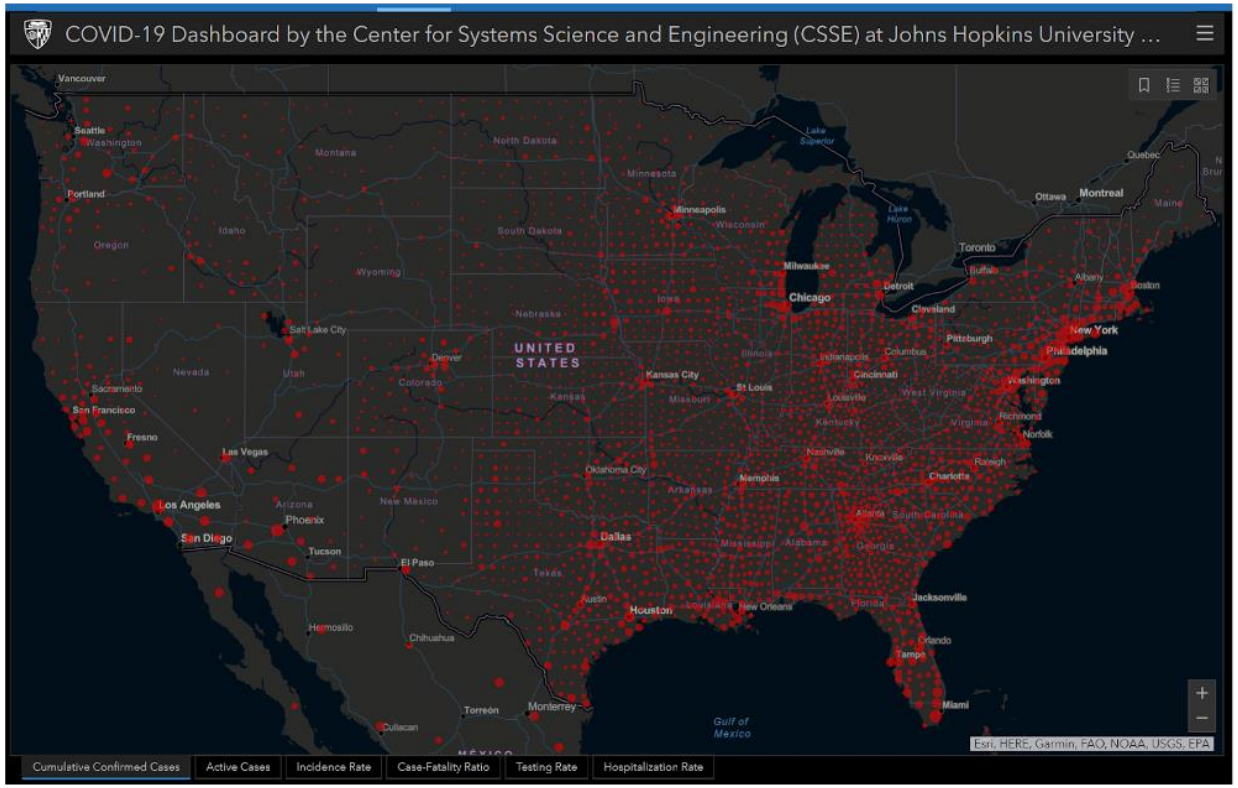
Not ideal

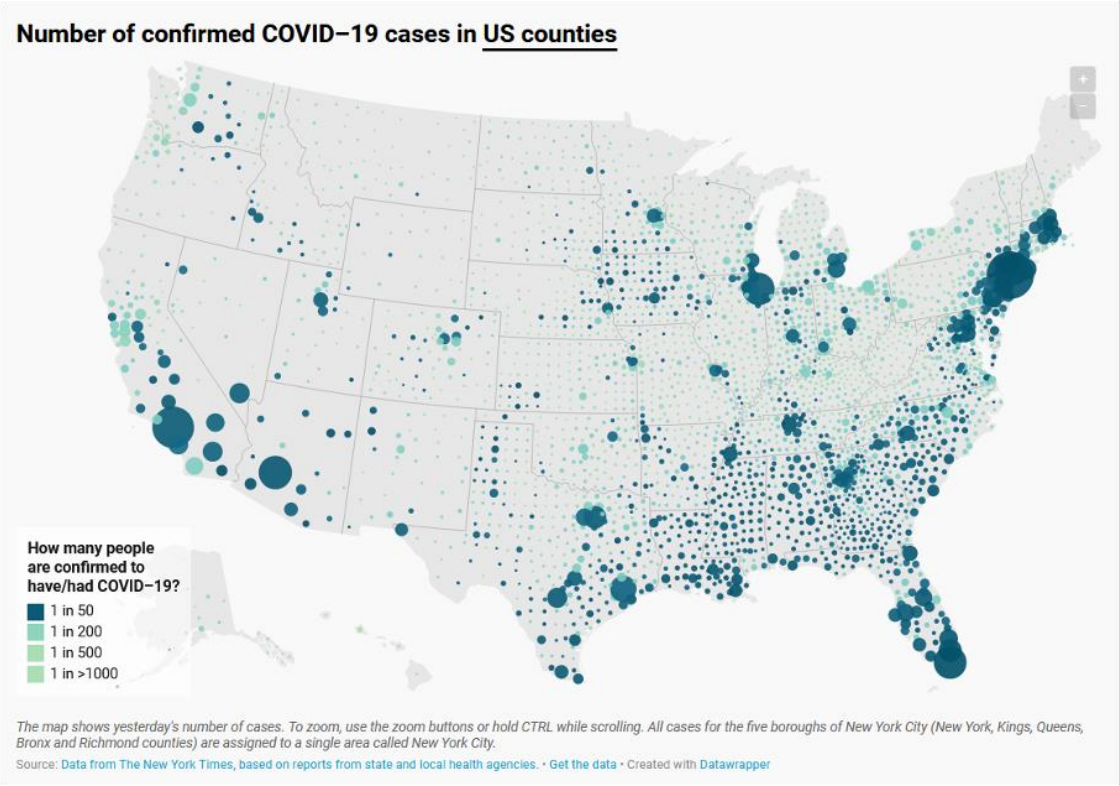
Better

"Tips for Effective Data Visualization." Eric Monson, SEDLS 2021.

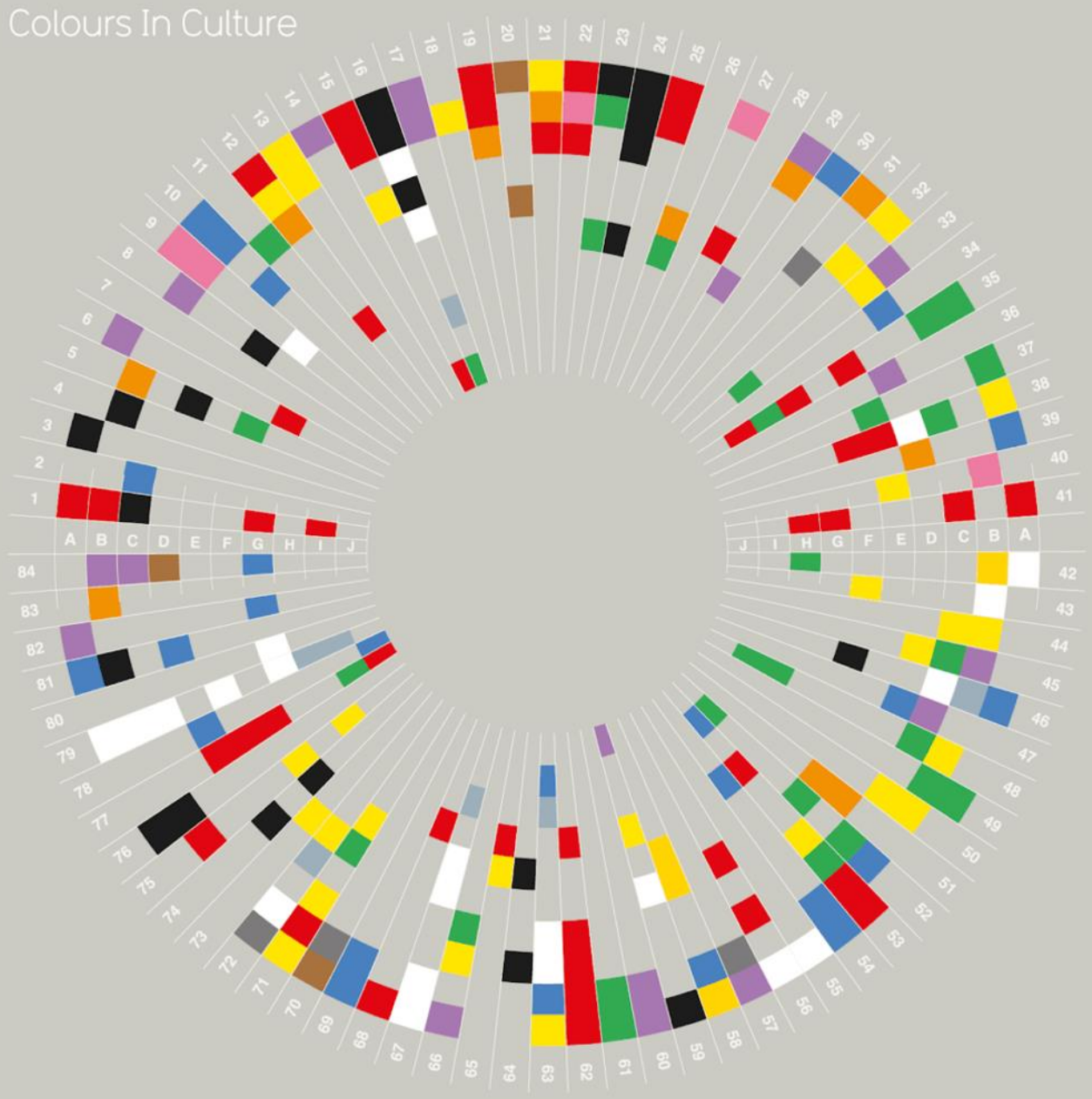
Be aware of emotional affect of color

- Colors can have meaning, and our emotional response is culturally variable





Colours In Culture



A Western / American	1 Anger	19 Desire
B Japanese	2 Art / Creativity	20 Earthy
C Hindu	3 Authority	21 Energy
D Native American	4 Bad Luck	22 Erotic
E Chinese	5 Balance	23 Eternity
F Asian	6 Beauty	24 Evil
G Eastern European	7 Calm	25 Excitement
H Arab	8 Celebration	26 Family
I African	9 Children	27 Femininity
J South American	10 Cold	28 Fertility
	11 Compassion	29 Flamboyance
	12 Courage	30 Freedom
	13 Cowardice	31 Friendly
	14 Cruelty	32 Fun
	15 Danger	33 God
	16 Death	34 Gods
	17 Decadence	35 Good Luck
	18 Deceit	36 Gratitude

37 Growth	55 Luxury	73 Royalty
38 Happiness	56 Marriage	74 Self-cultivation
39 Healing	57 Modesty	75 Strength
40 Healthy	58 Money	76 Style
41 Heat	59 Mourning	77 Success
42 Heaven	60 Mystery	78 Trouble
43 Holiness	61 Nature	79 Truce
44 Illness	62 Passion	80 Trust
45 Insight	63 Peace	81 Unhappiness
46 Intelligence	64 Penance	82 Virtue
47 Intuition	65 Power	83 Warmth
48 Religion	66 Personal power	84 Wisdom
49 Jealousy	67 Purity	
50 Joy	68 Radicalism	
51 Learning	69 Rational	
52 Life	70 Reliable	
53 Love	71 Repels Evil	
54 Loyalty	72 Respect	

Yellow Grey
Gold Silver

David McCandless & AlwaysWithHonor.com // v1.0 // Apr 09 // InformationIsBeautiful.net

source: Pantone, ColorMatters & web sources

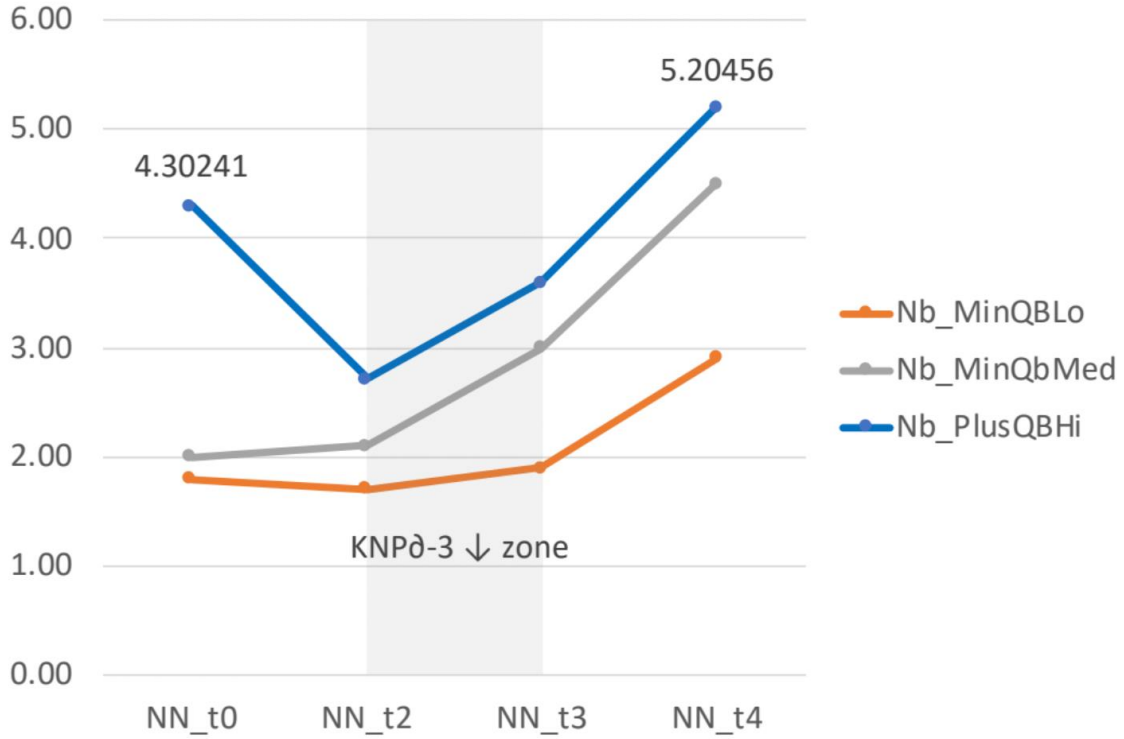
General tips for better data visualization

Remove unnecessary detail

Use human-readable labels

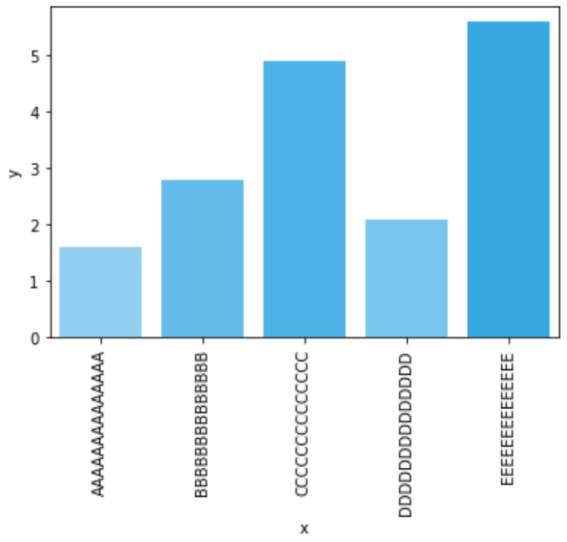
Avoid:

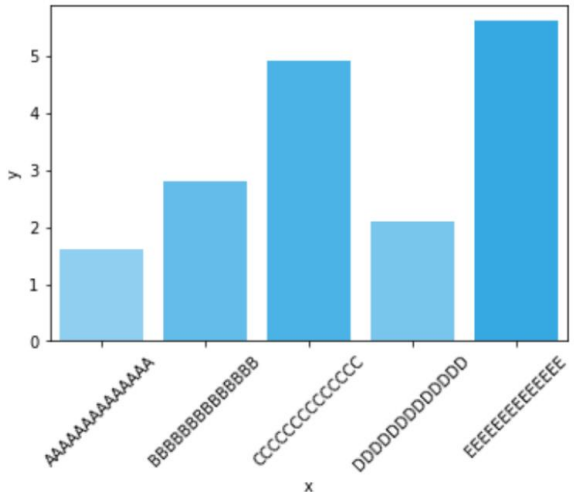
- Jargon
- Abbreviations
- Variable names
- Useless decimals
- Too many labels



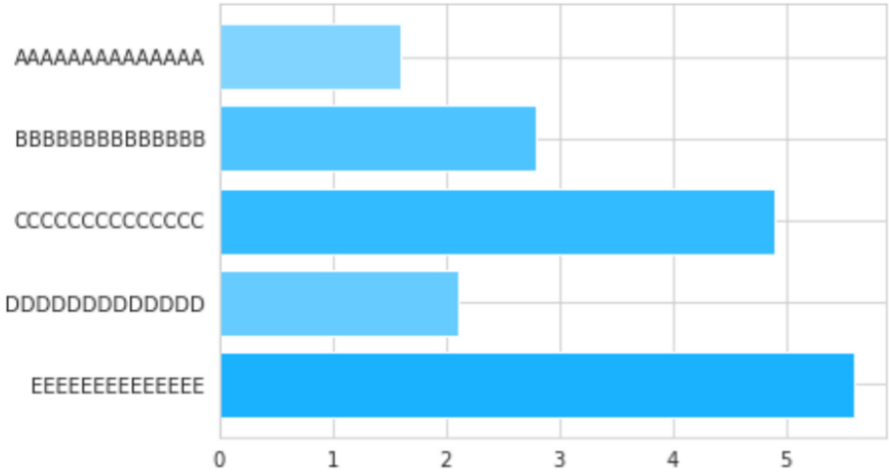
"Tips for Effective Data Visualization." Eric Monson, SEDLS 2021.

Don't use vertical labels



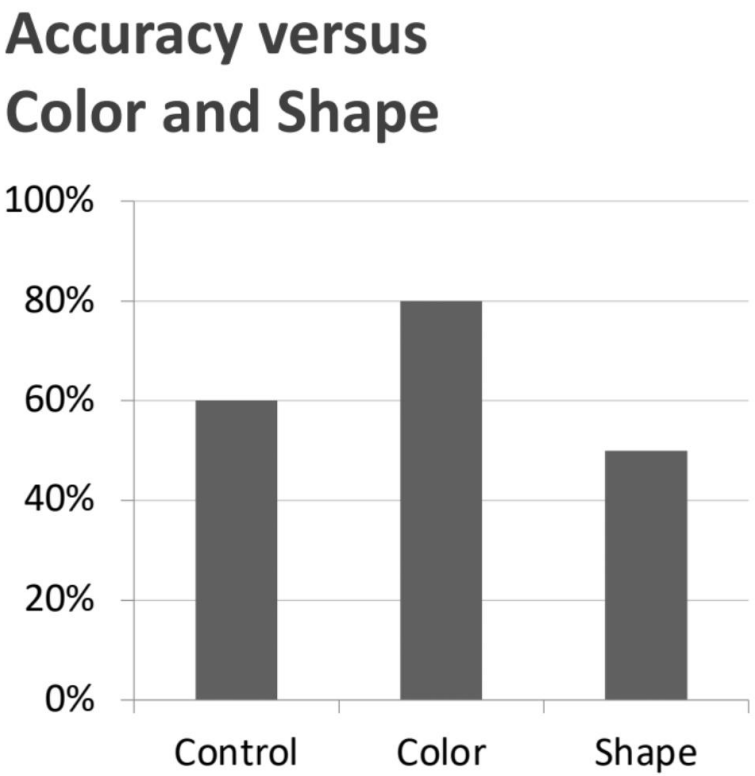


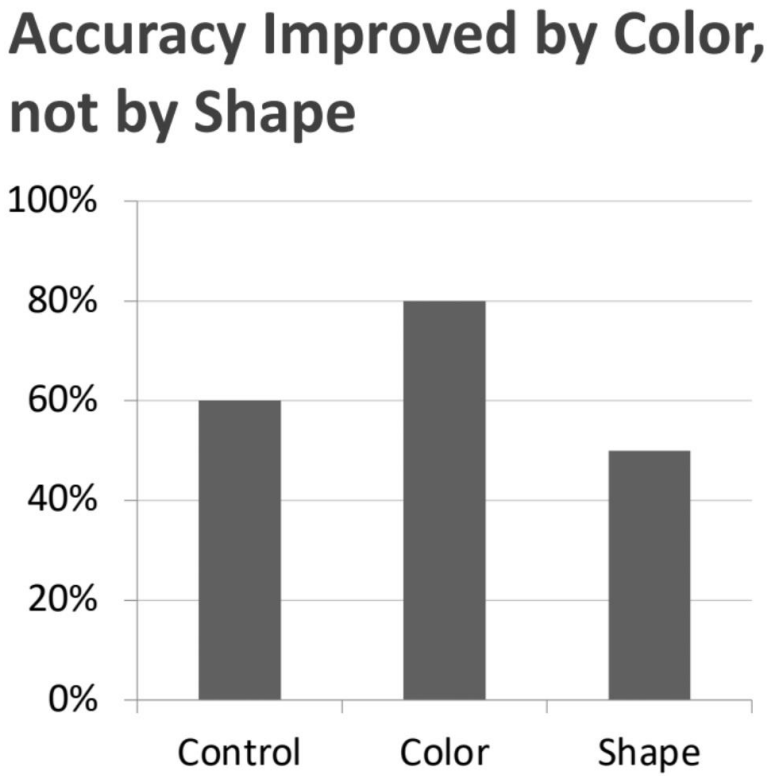
Better



Best!

Use active titles

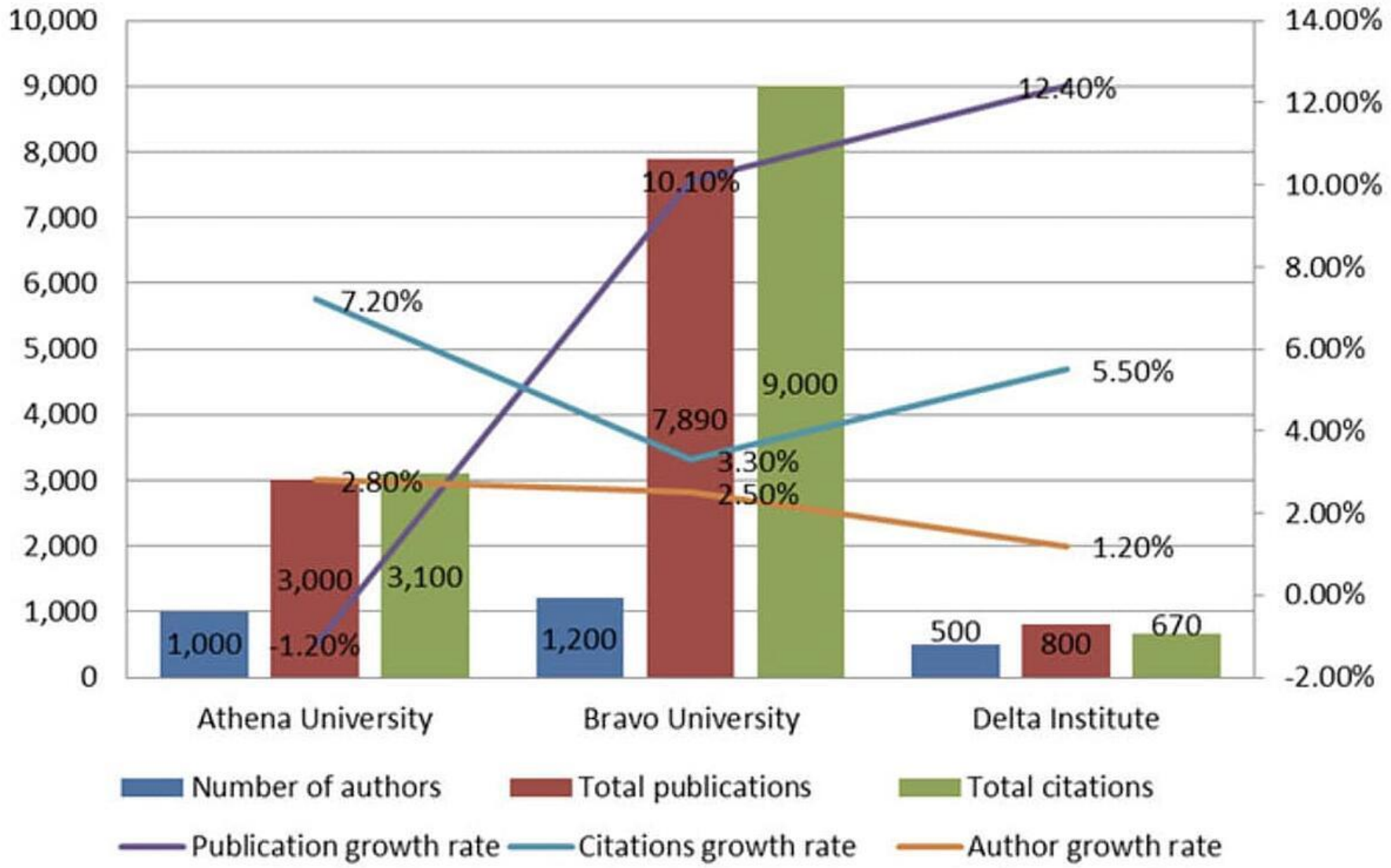


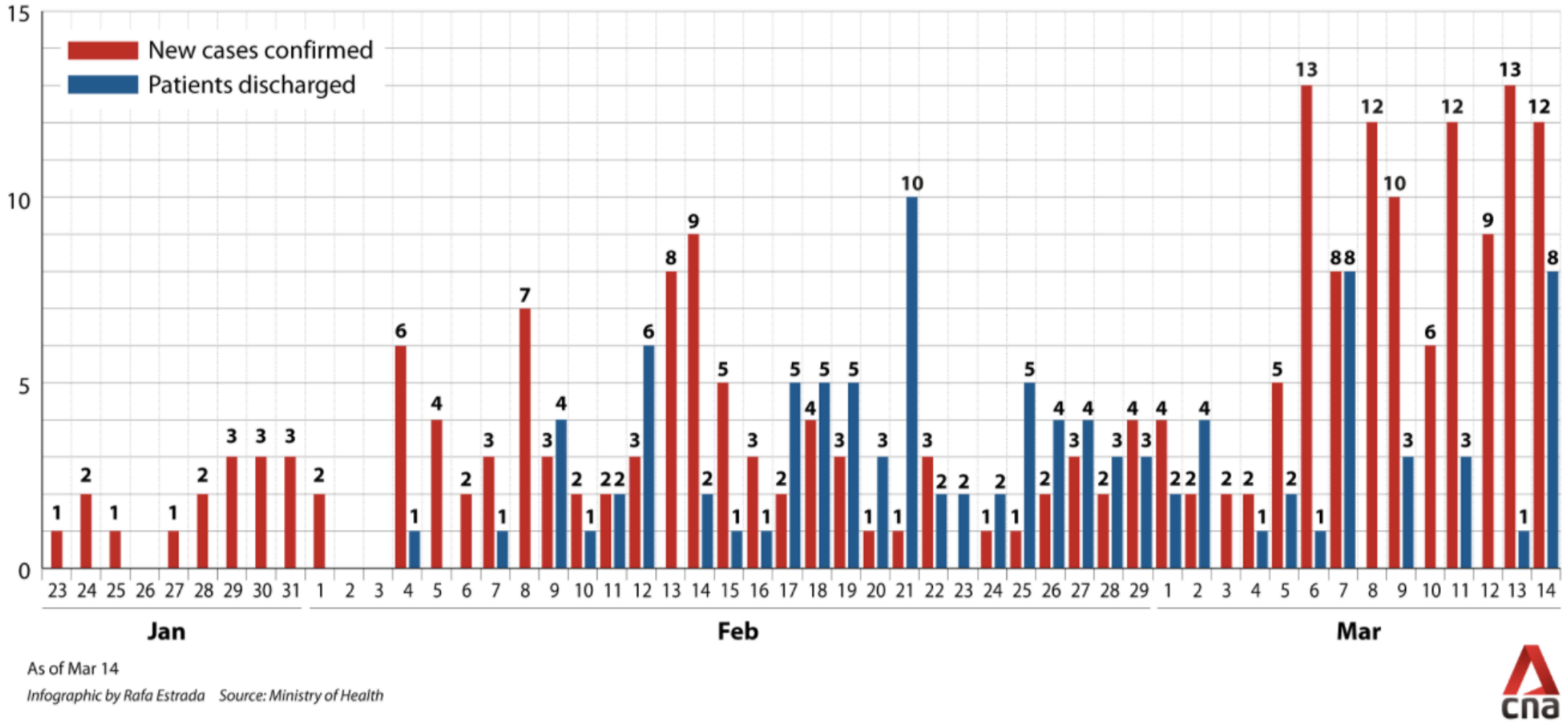


"Tips for Effective Data Visualization." Eric Monson, SEDLS 2021.

Let’s give it a try!

What suggested edits do you have for these visualizations?


10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
-2.00%
1,000
3,000
3,100
1,200
7,890
9,000
500
800
670
-1.20%
7.20%
2.80%
10.10%
3.30%
2.50%
12.40%
5.50%
1.20%
Athena University
Bravo University
Delta Institute
Number of authors
Total publications
Total citations
Publication growth rate
Citations growth rate
Author growth rate



https://analythical.com/blog/covid19-in-charts

Data Visualization Tools

Tools: Access Through UTHSC

- Excel
- Mathematica
- ArcGIS
- NVivo

https://libguides.uthsc.edu/data/tools

Tools: Additional (Not UTHSC)

- Tableau (free 1 year educator and student licenses, renewable annually)
- D3.js
- ggplot2
- NodeBox
- RAWGraphs
- Spotfire
- Plotly

https://libguides.uthsc.edu/data/tools

Library Research Guide for Clinical Researchers

https://libguides.uthsc.edu/clinicalresearchers

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER