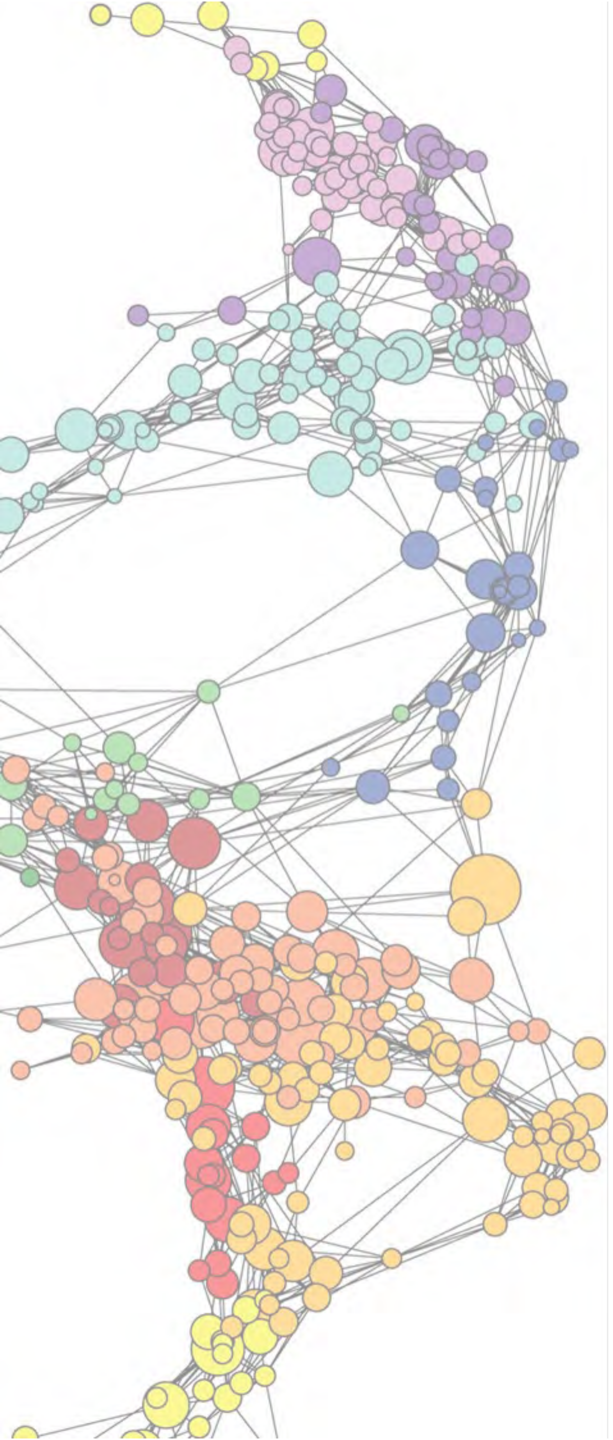
Tips for Effective Data Visualization

Eric E Monson, PhD
Data Visualization Specialist
Duke University Libraries
Center for Data and Visualization Sciences

Quantitative Summer Internship in HIV/AIDS · May 2023

Slides: https://bit.ly/HIVQuantEffVisMay2023



What is data visualization?

Anything that converts data sources into a visual representation

charts, graphs, maps – even just tables!

Why do we visualize?

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

Almost identical summary statistics:

x & y mean

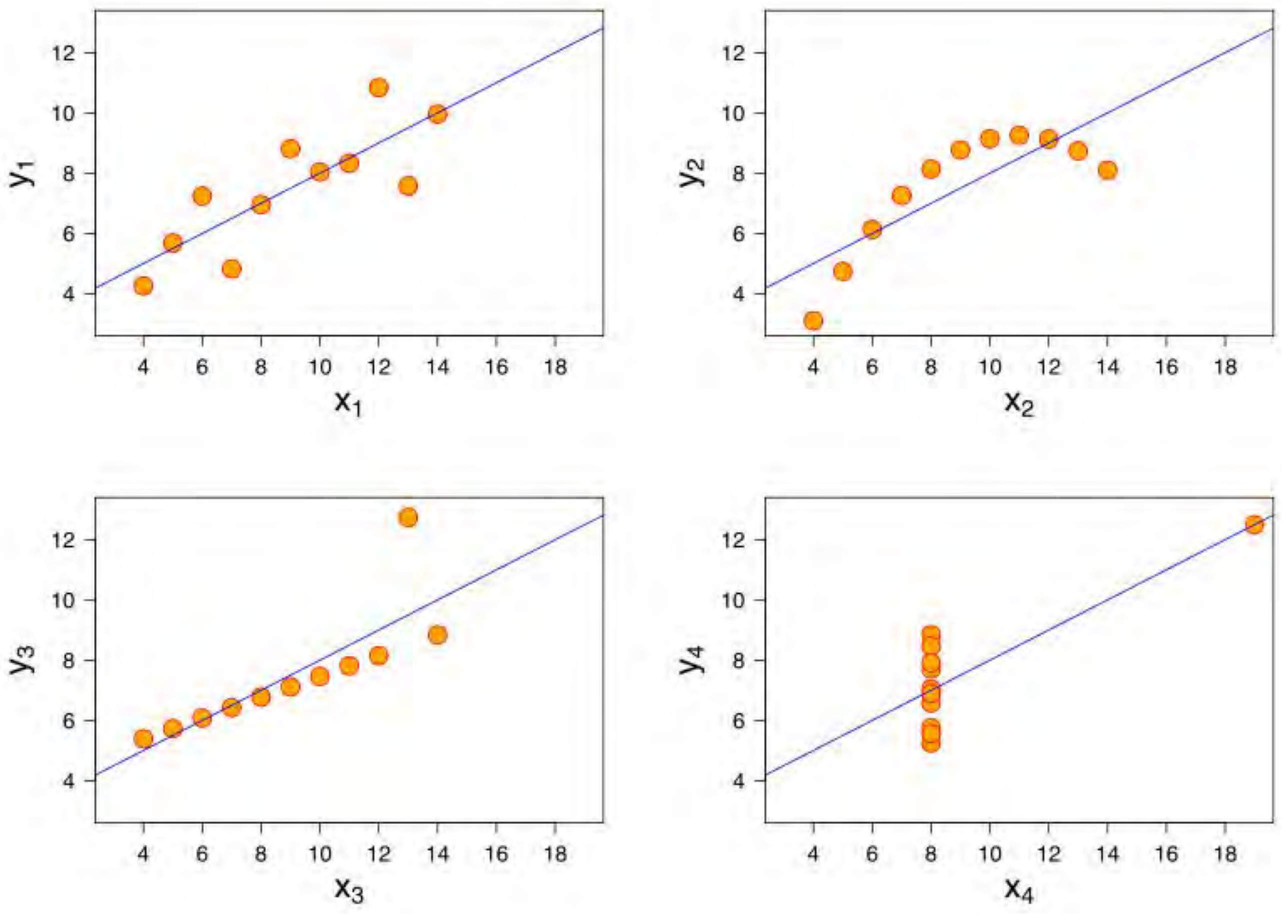
x & y variance

x-y correlation

x-y linear regression

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

We visualize to see patterns



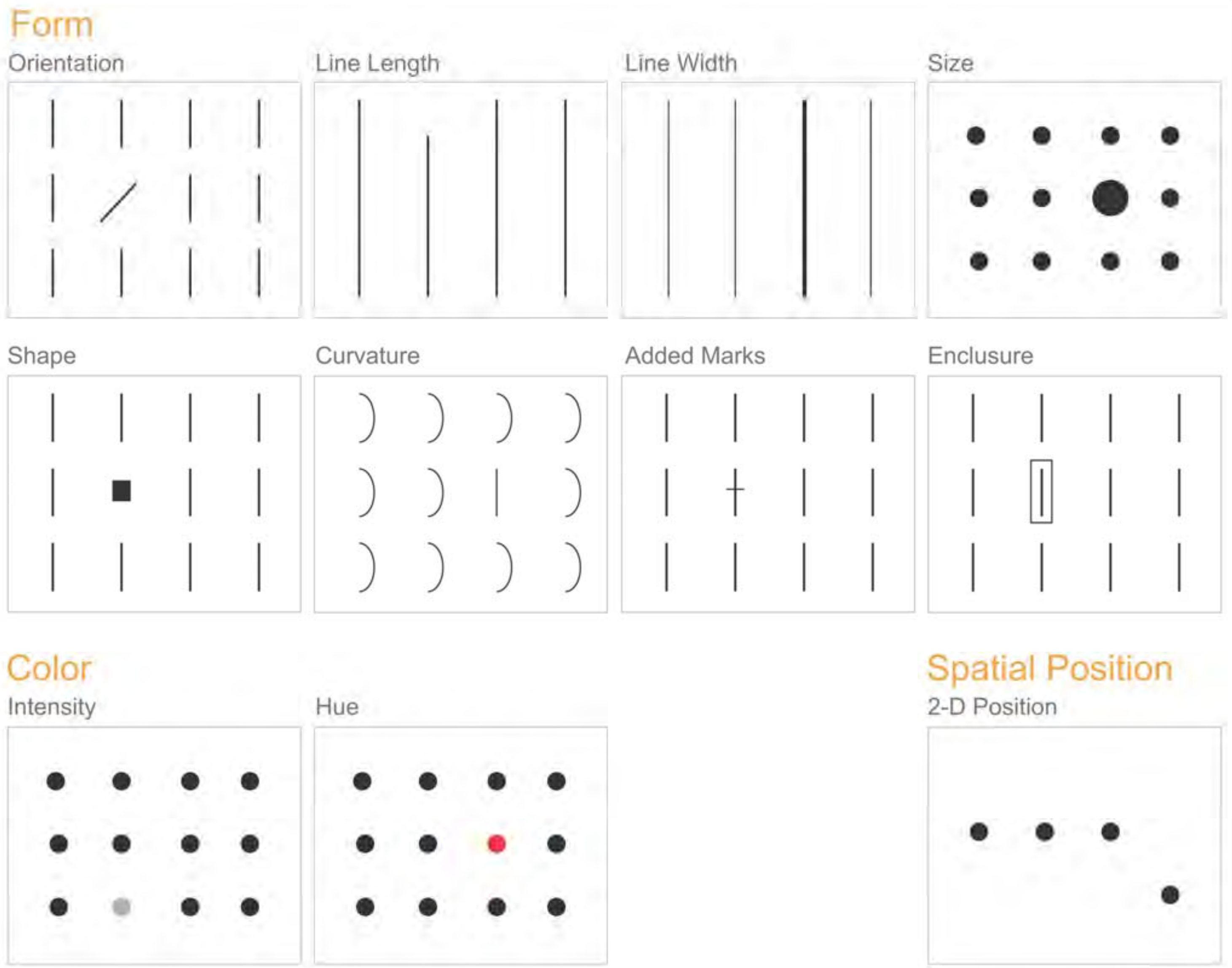
Anscombe's Quartet

http://en.wikipedia.org/wiki/Anscombe%27s_quartet

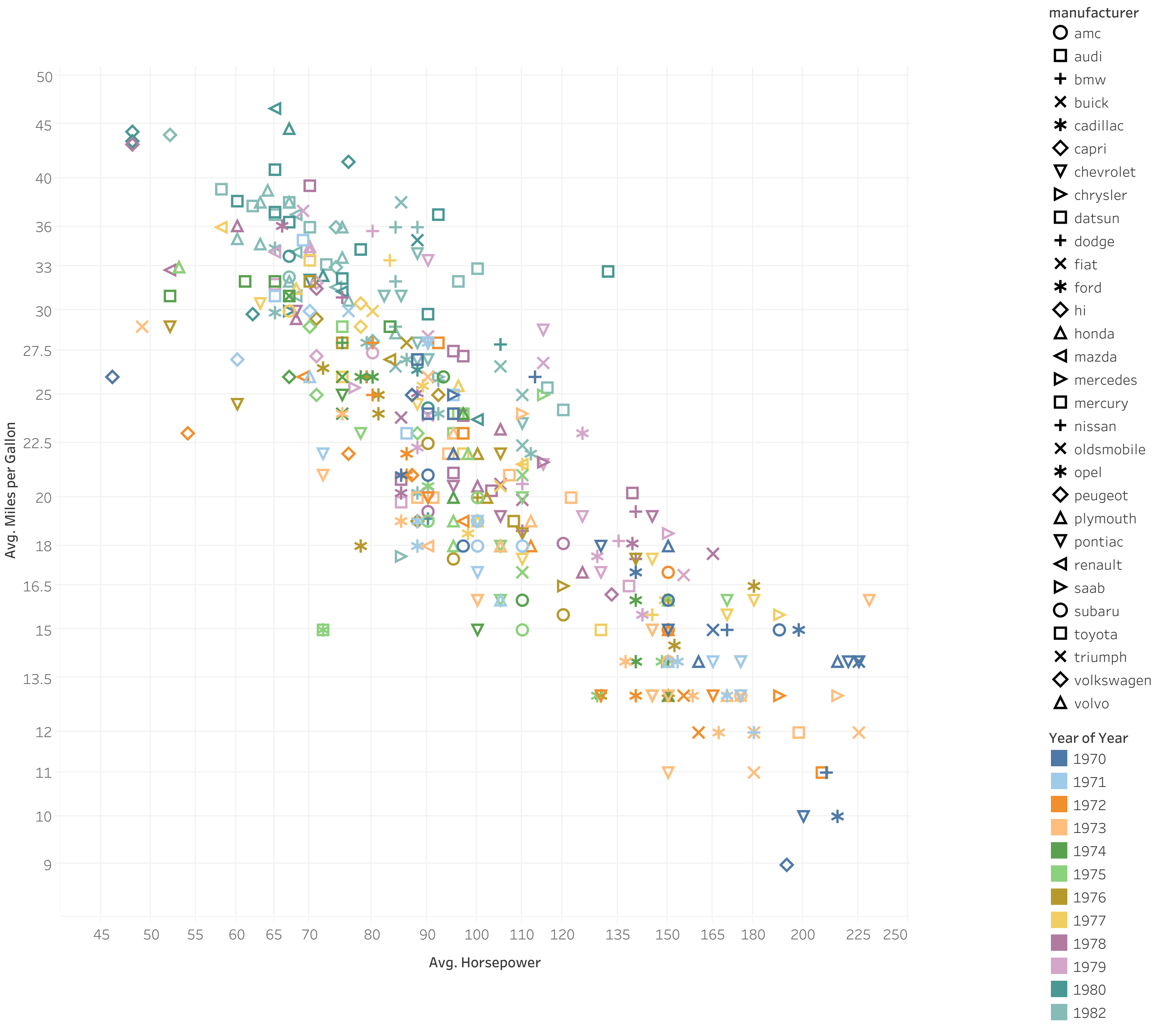
Visualization:

Starting points

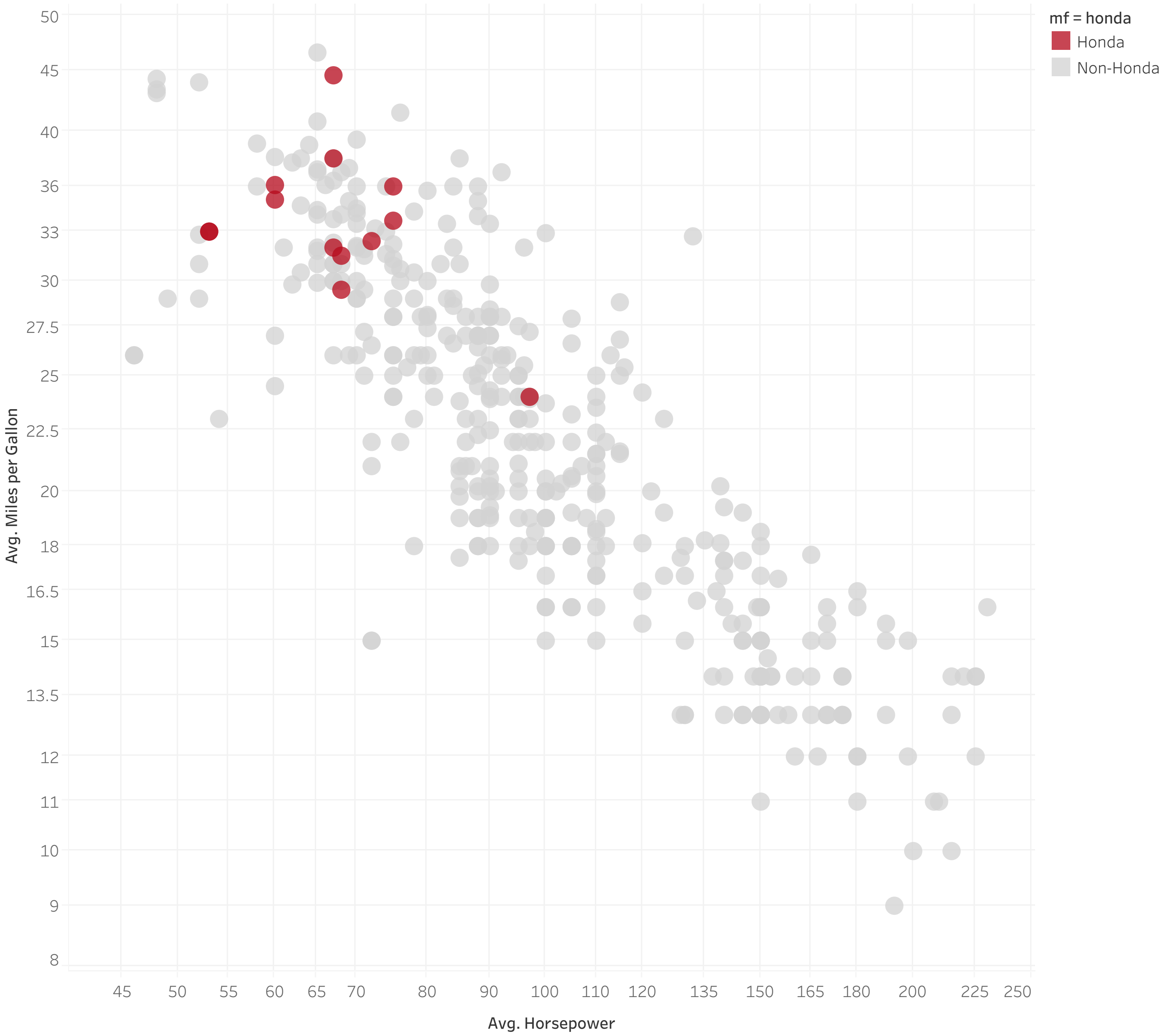
Pre-attentive visual attributes will encode our data



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



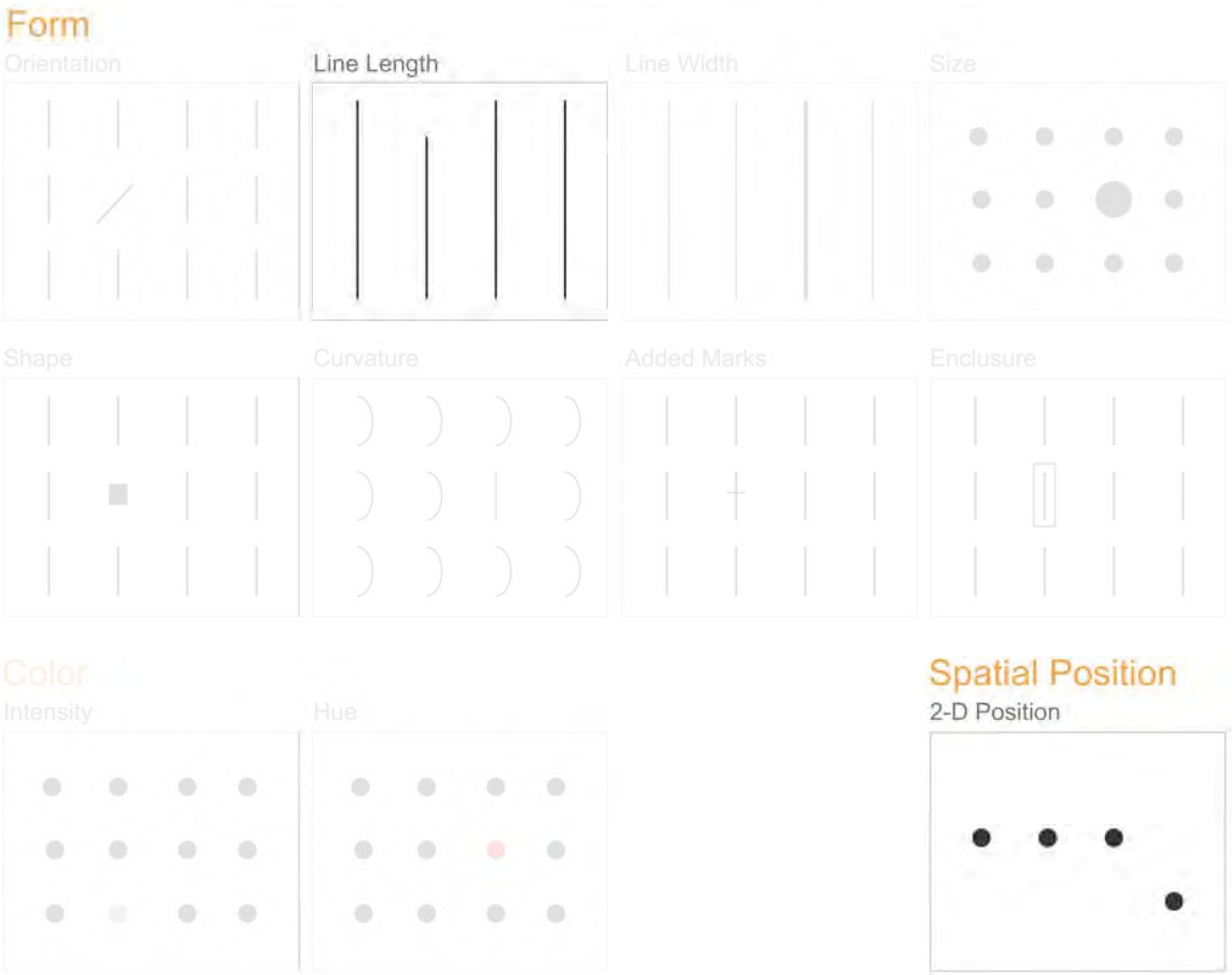
Avg. Miles per Gallon
50
45
40
36
33
30
27.5
25
22.5
20
18
16.5
15
13.5
12
11
10
9
45
50
55
60
65
70
80
90
100
110
120
135
150
165
180
200
225
250
Avg. Horsepower
manufacturer
amc
audi
bmw
buick
cadillac
capri
chevrolet
chrysler
datsun
dodge
fiat
ford
hi
honda
mazda
mercedes
mercury
nissan
oldsmobile
opel
peugeot
plymouth
pontiac
renault
saab
subaru
toyota
triumph
volkswagen
volvo
Year of Year
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1982



mf = honda
Honda
Non-Honda
Avg. Miles per Gallon
50
45
40
36
33
30
27.5
25
22.5
20
18
16.5
15
13.5
12
11
10
9
8
45
50
55
60
65
70
80
90
100
110
120
135
150
165
180
200
225
250
Avg. Horsepower

Pre-attentive visual **attributes** will encode our data

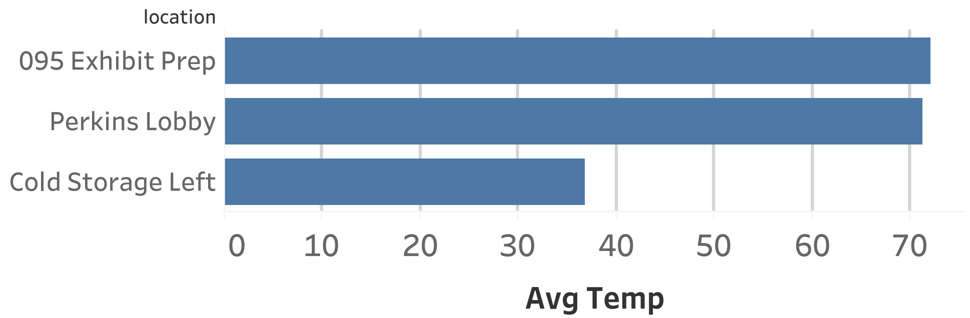
Quantitative comparisons easiest for these attributes



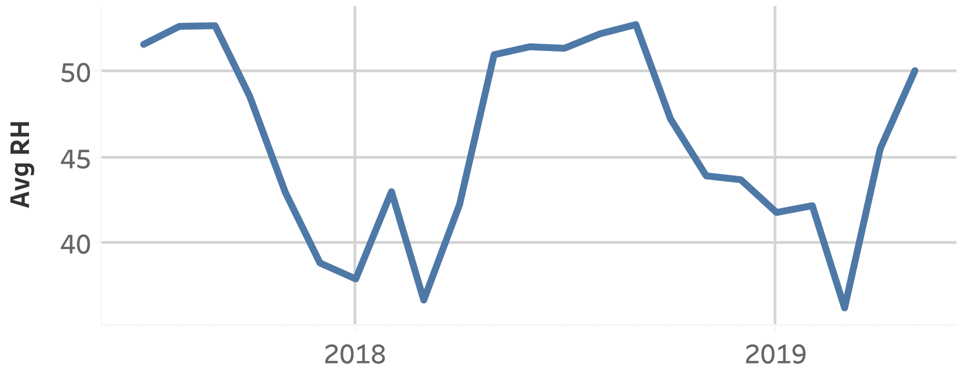
https://www.perceptualedge.com/articles/ie/visual_perception.pdf

Classic charts because they work well
– good starting point!

Category + Numbers
Bar



Date/time + Numbers
Line



Two Numerical (correlation)
Scatter

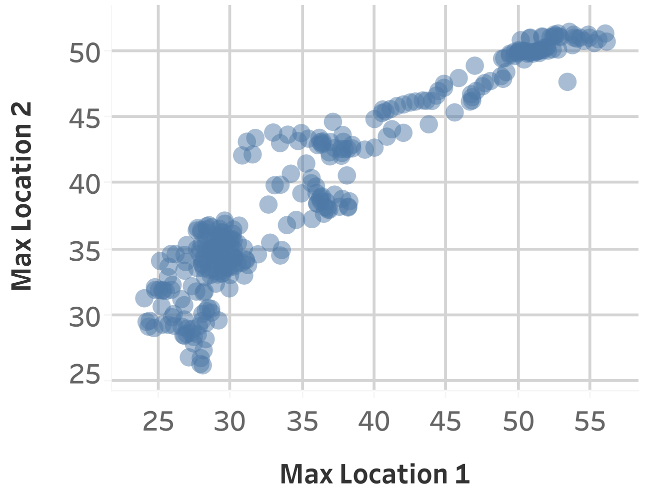


Chart choosing:

Make the most important comparisons easy

*There are a huge variety of potential plots, even with a simple data set, and **many possible stories to notice**.*

You must decide what's important and design to reveal that!

Steven Franconeri (Northwestern) – reading word vs paragraph

	Green	Yellow	Cheap	Tasty	Gross
Corn	6	29	18	30	7
Squash	8	27	17	13	11
Brussel sprouts	10	21	16	4	19
Green beans	20	17	16	9	7
Peas	23	5	15	19	2

Inspired by: https://www.nature.com/articles/nmeth.2807

Story: Not clear…

	Green	Yellow	Cheap	Tasty	Gross
Corn	6	29	18	30	7
Squash	8	27	17	13	11
Brussel sprouts	10	21	16	4	19
Green beans	20	17	16	9	7
Peas	23	5	15	19	2

Table

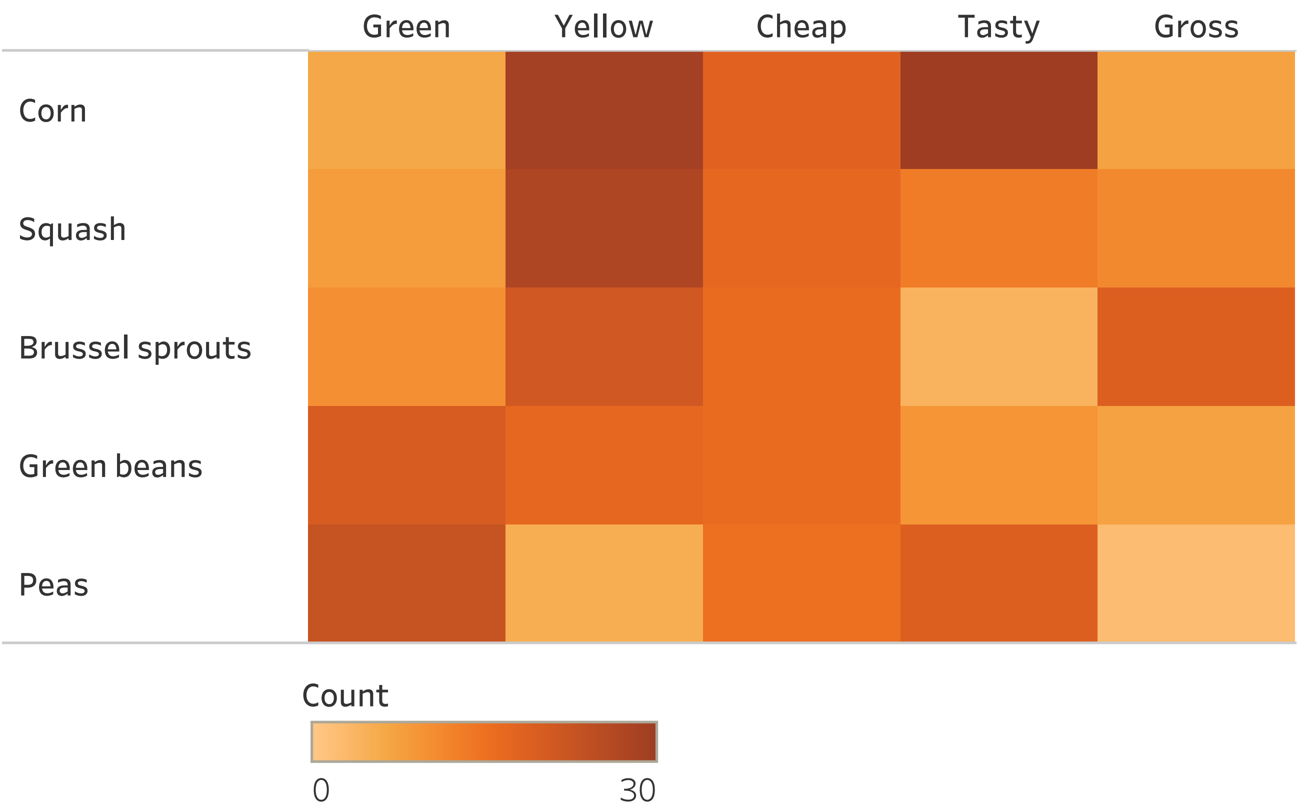
Pro:

- Compact
- **Precise value lookup**

Con:

- **Hard to see patterns**
- Not favoring any specific comparison

Story: Not clear…



Heatmap

Pro:

- Compact
- Eye-catching
- **See blocks of light and dark**

Con:

- Can’t see small differences
- Eyes fooled by nearby colors
- Not great quantitatively
- Not favoring any specific comparison/story

Story: How characteristics vary across the vegetables

Proportional size symbols

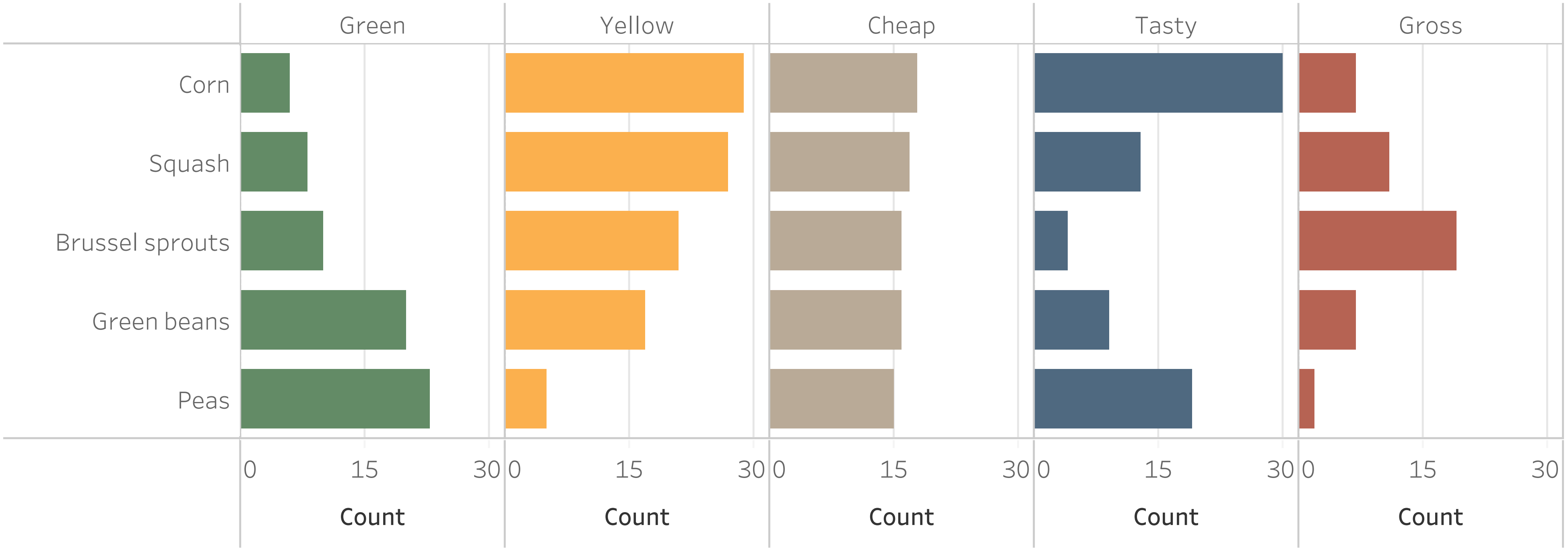
Pro:

- Compact
- Eye-catching
- **Color biases to seeing columns**
- **See ramps in size**

Con:

- Can’t see small differences
- Not great quantitatively

Story: How characteristics vary across the vegetables



Small multiples

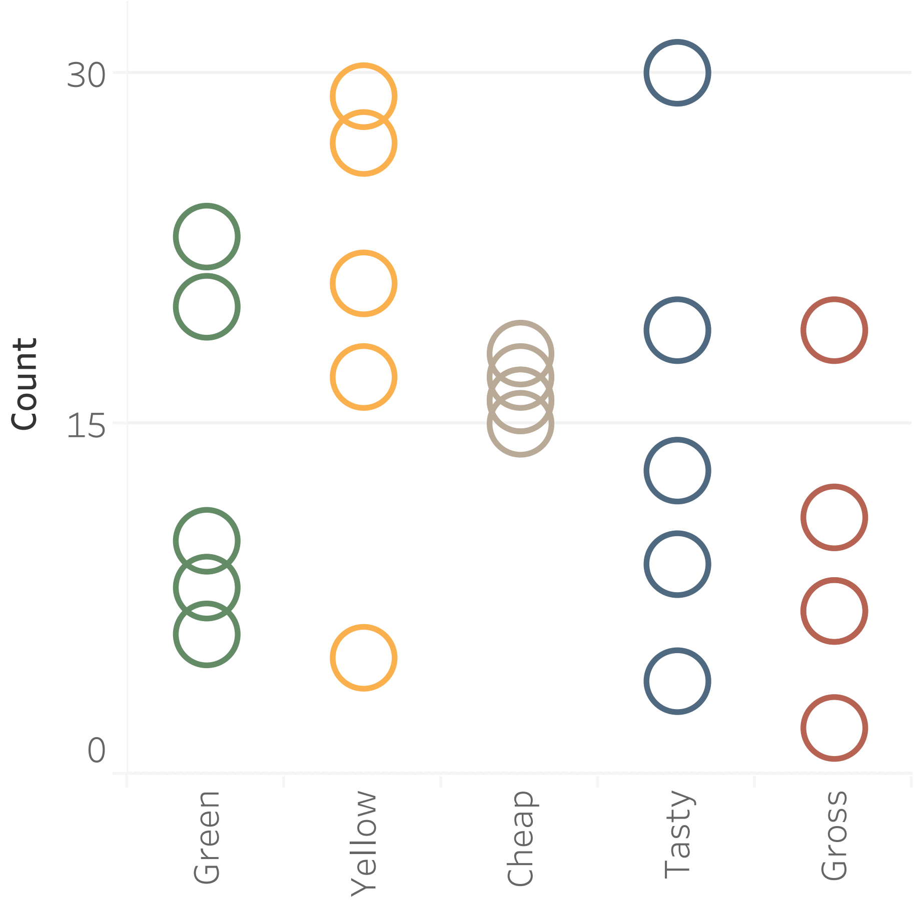
Pro:

- **Easy to compare within categories with common baselines**
- Can see small differences
- Everything directly labeled

Con:

- Comparisons across harder
- **Some software can't do faceting**

Story: How characteristics vary across the vegetables



Dot distribution plot

Pro:

- Directly see numbers and distribution of **individual values, not just summary**

Con:

- Hard to judge density if overlap
- **Not all software can jitter or pack points to reveal density**

Story: Characteristics of each vegetable

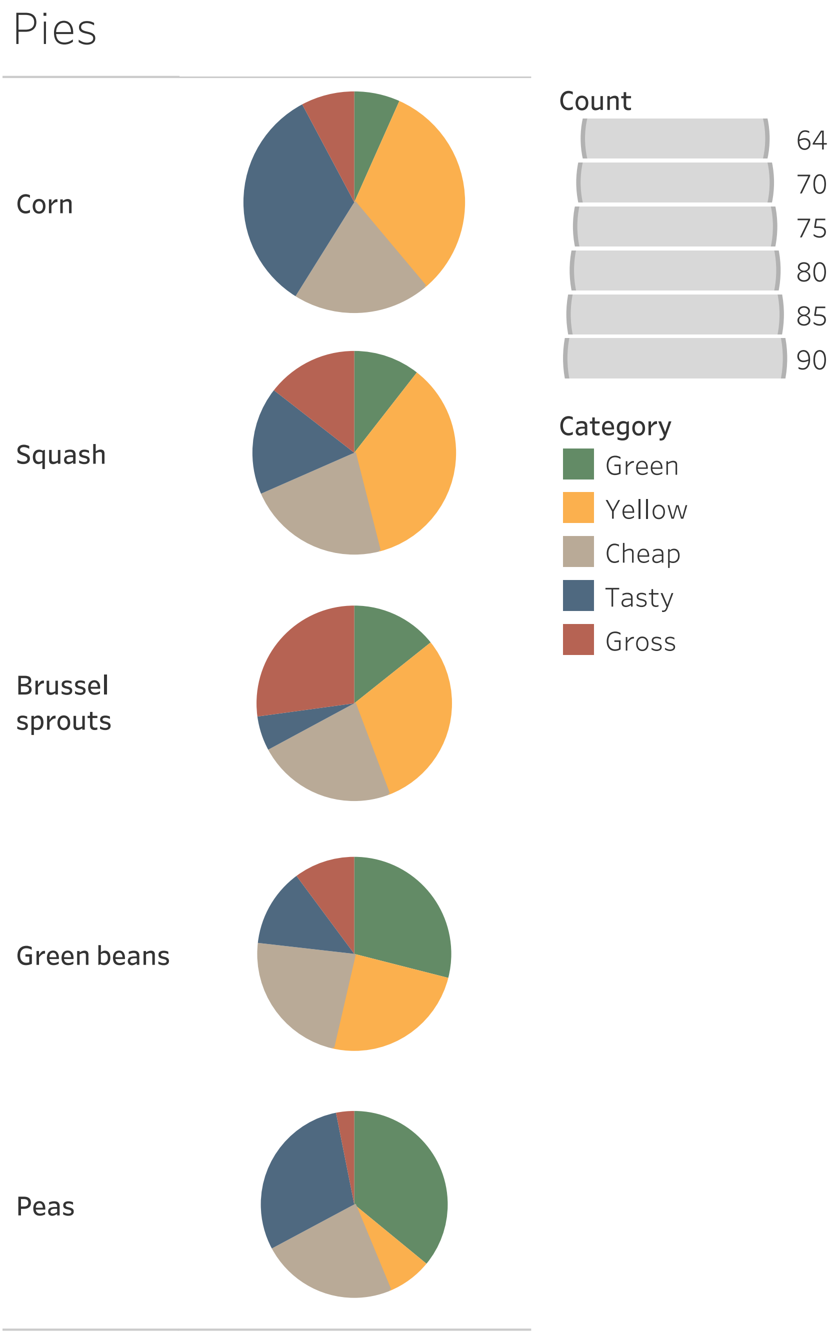
Pie chart grid

Pro:

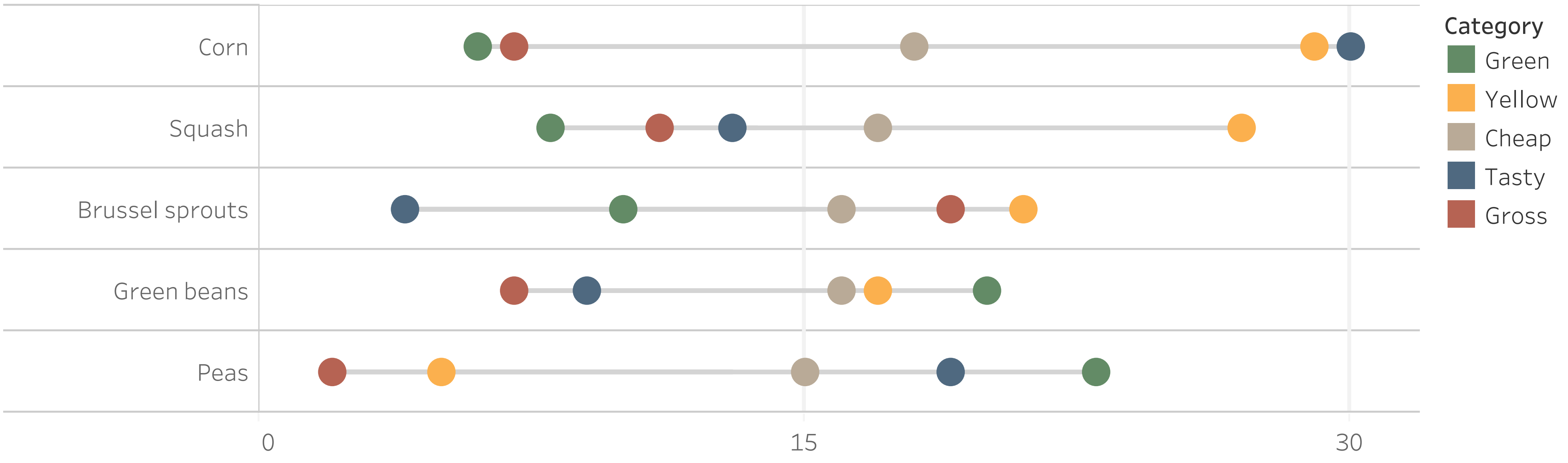
- Familiar
- **Fine if not too many slices**
- Fine if “parts of a whole” metaphor holds
- Best if sort slices large to small

Con:

- **Slices starting at 12:00 easiest**
- **Other floating slices hard to compare**
- **Hard to compare across pies**
- Can’t see small differences
- Not great quantitatively



Story: Characteristics of each vegetable



Dot plot

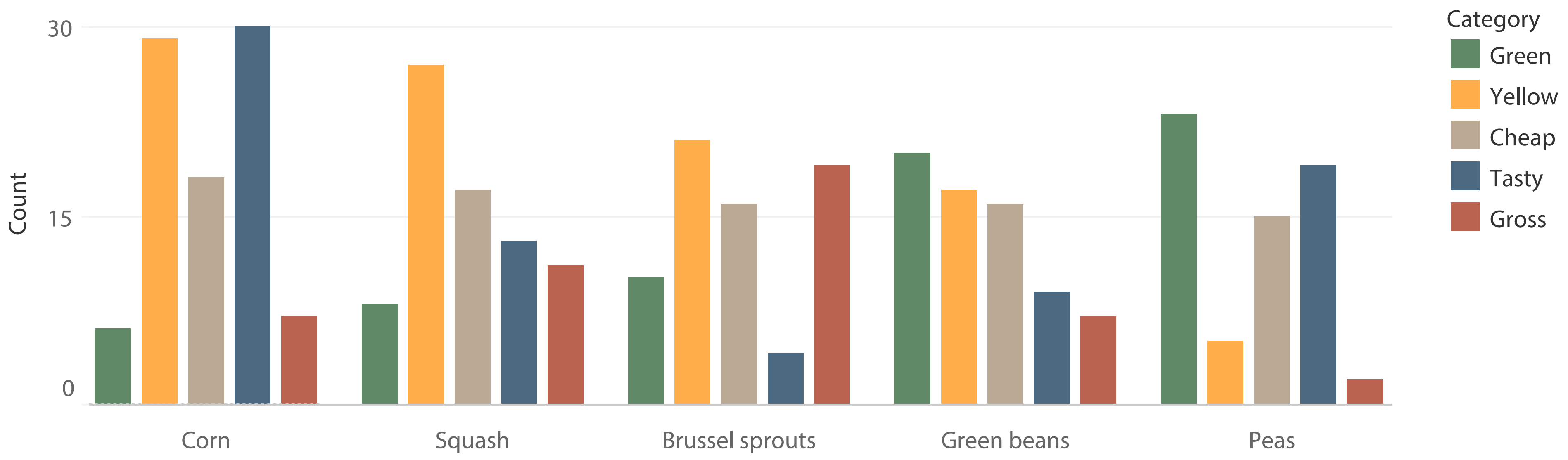
Pro:

- Easy to see small differences
- Works on a log scale
- **Great for two categories (dumbbell plot)**

Con:

- Five categories too many with large value variations

Story: Characteristics of each vegetable



Grouped bars

Pro:

- Common-baseline bars
- Easy within groups

Con:

- **Hard to visually filter and compare across groups**
- “Color strobing” hard to look at
- Still need legend

Story: Characteristics of each vegetable

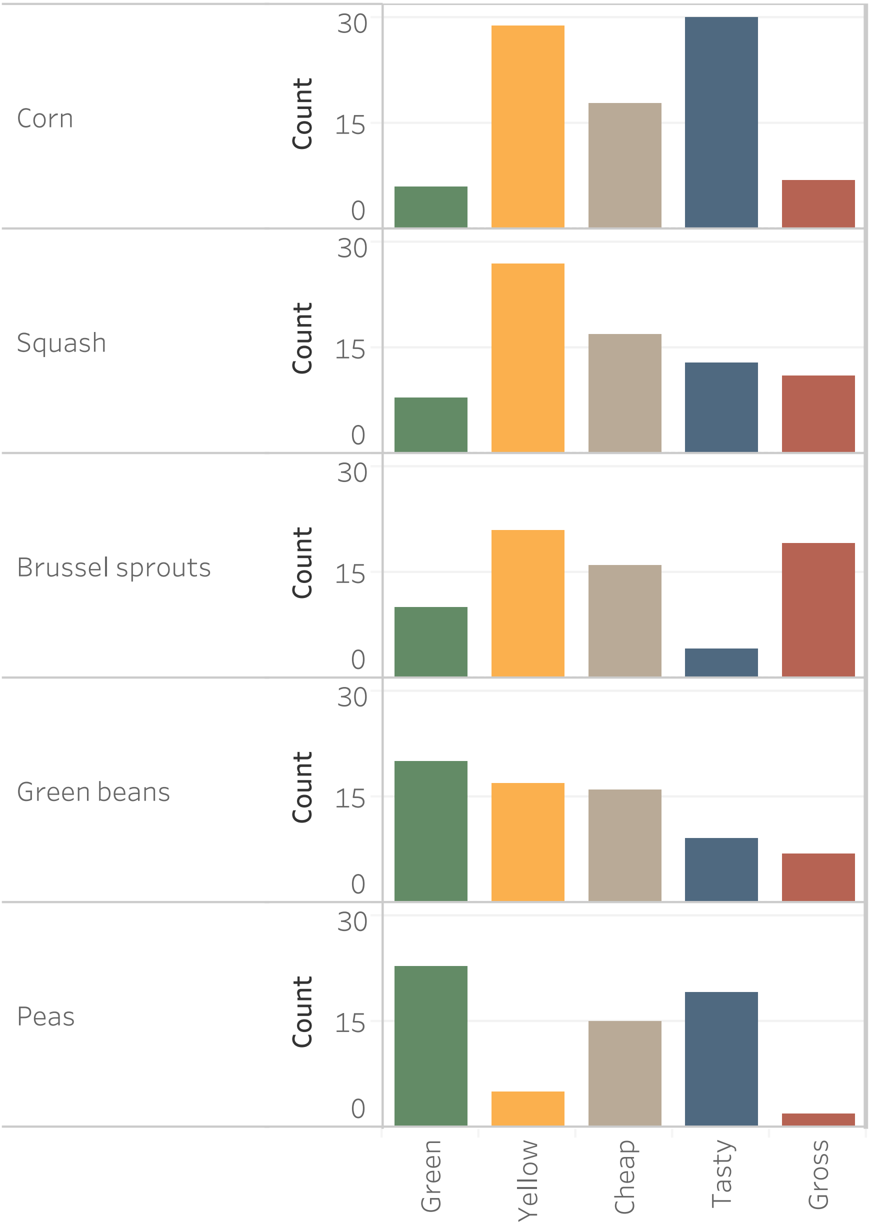
Small multiple bars

Pro:

- Facets or “small multiples” – nice approach
- **Common baseline easy to compare across**
- **Everything directly labeled (no legend)**

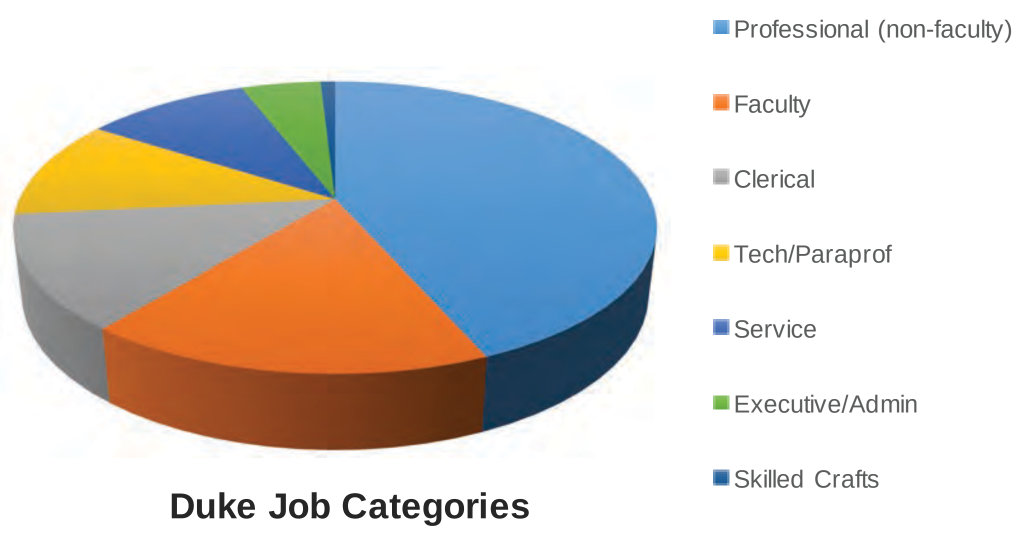
Con:

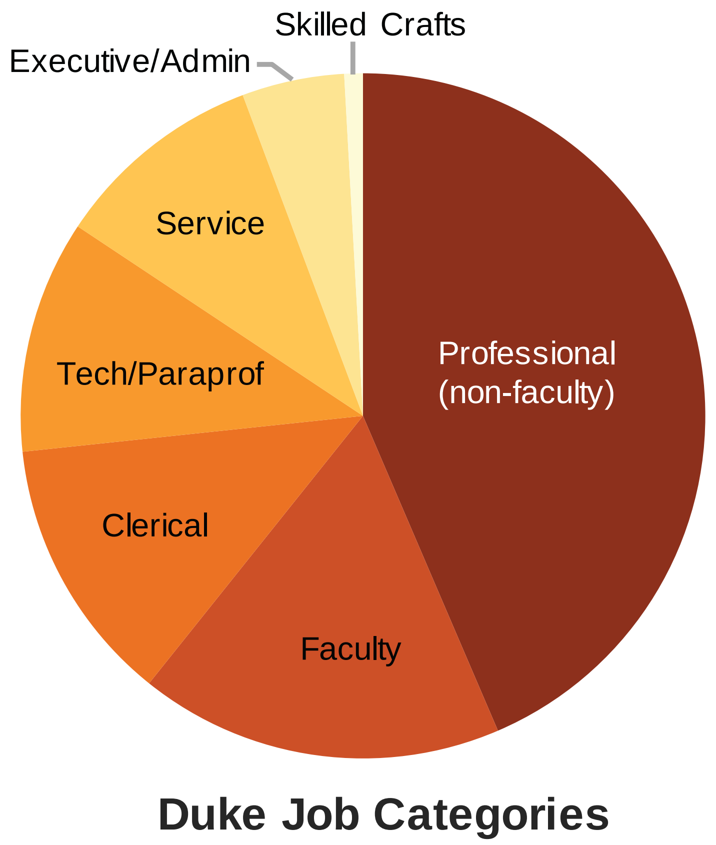
- Comparison up and down possible, but harder
- **Some software can’t do faceting**



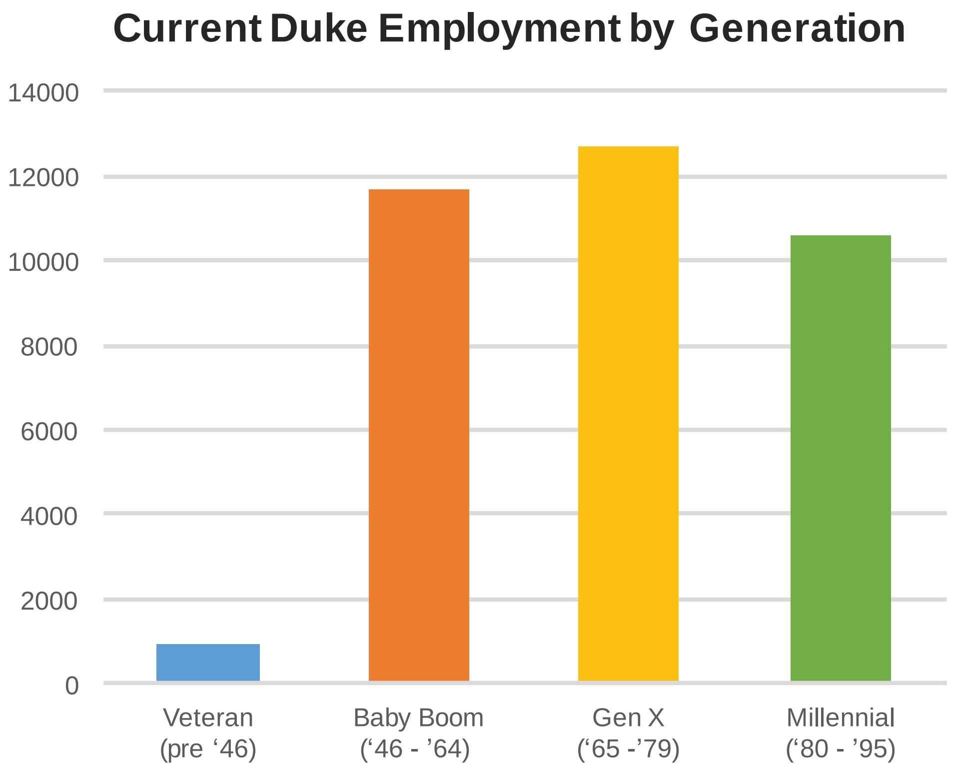
Three tips for designing effective visualizations

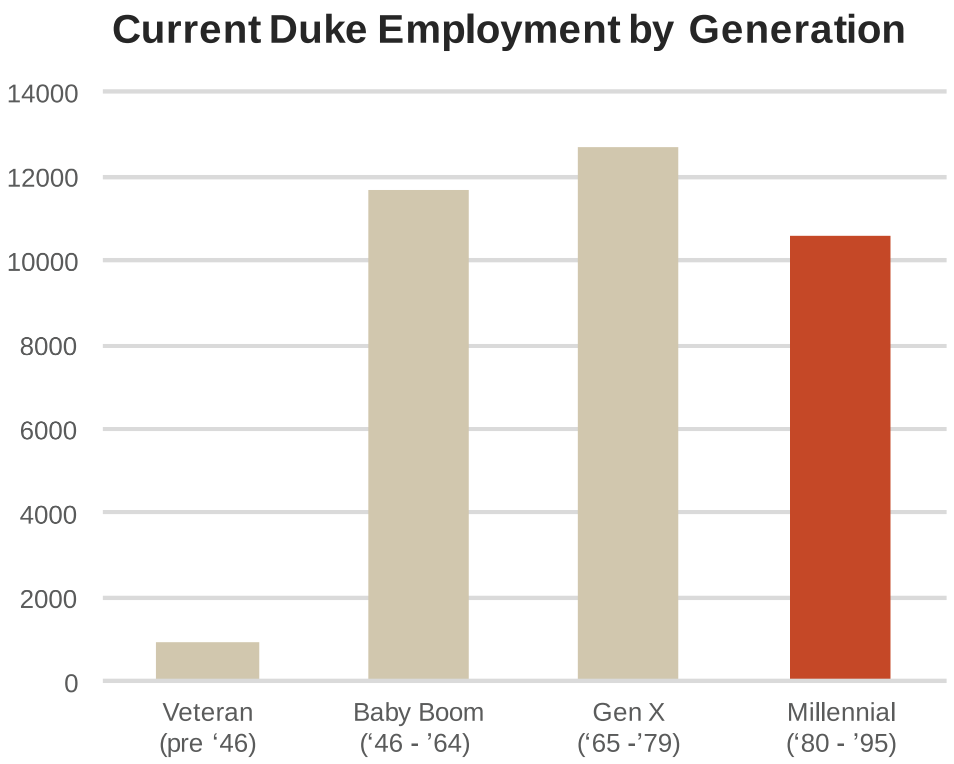
Avoid distortion & legends



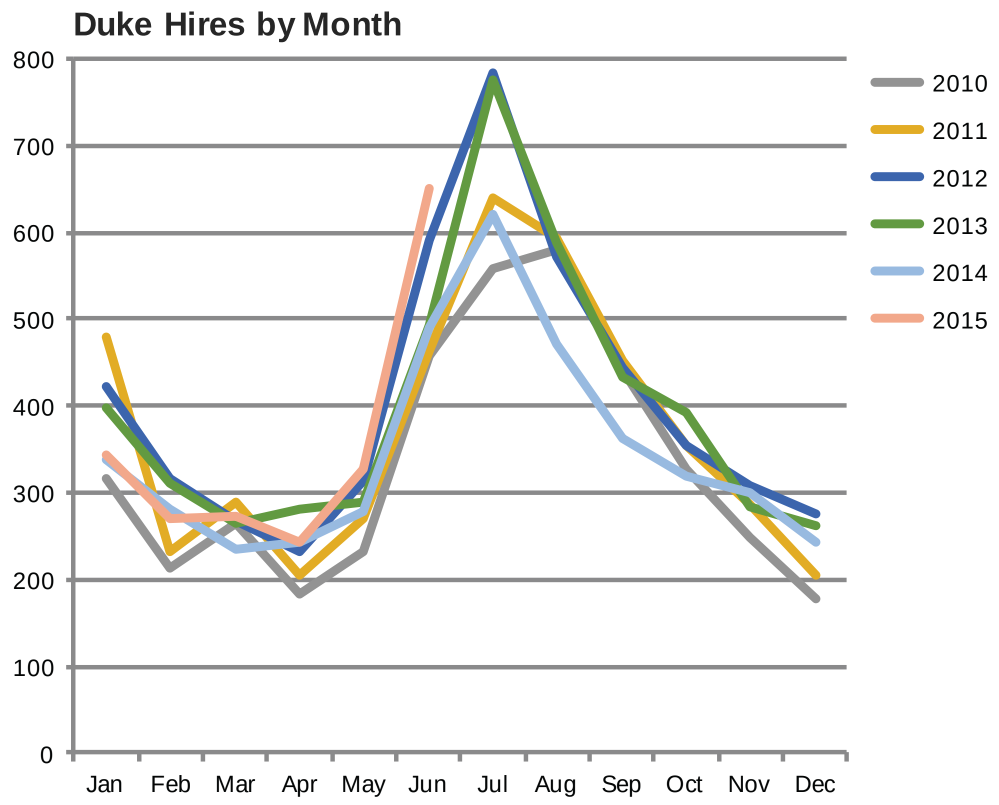


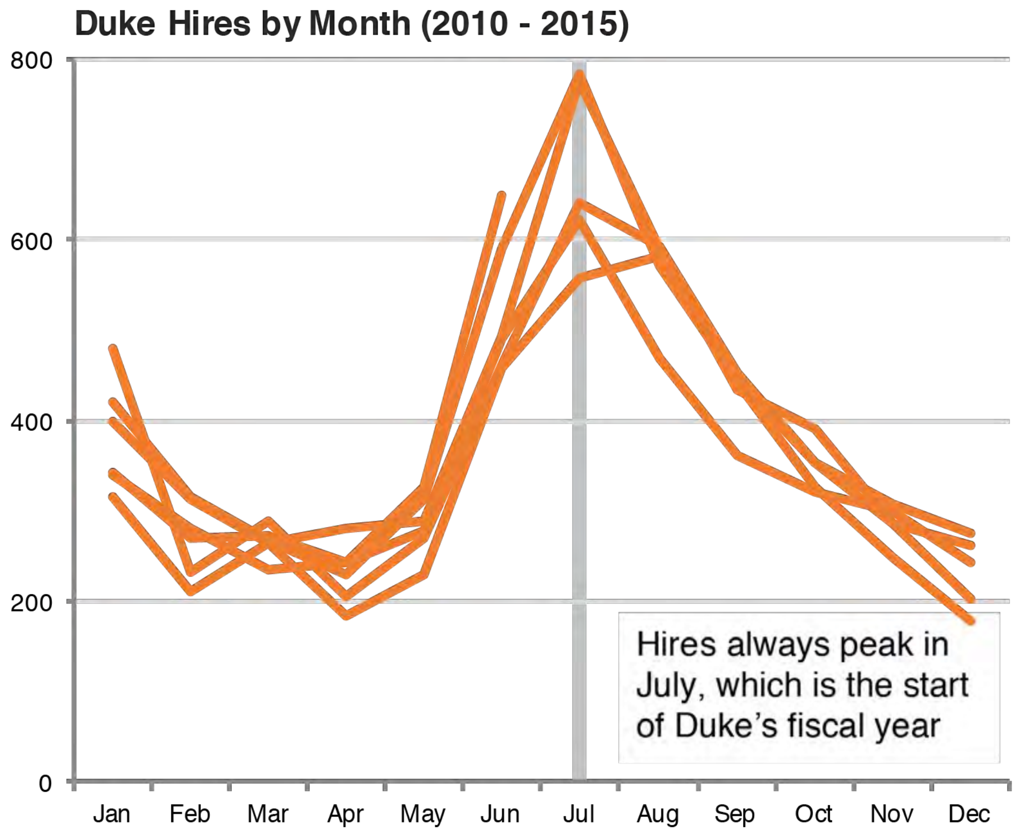
Don’t waste color – use it to draw attention!



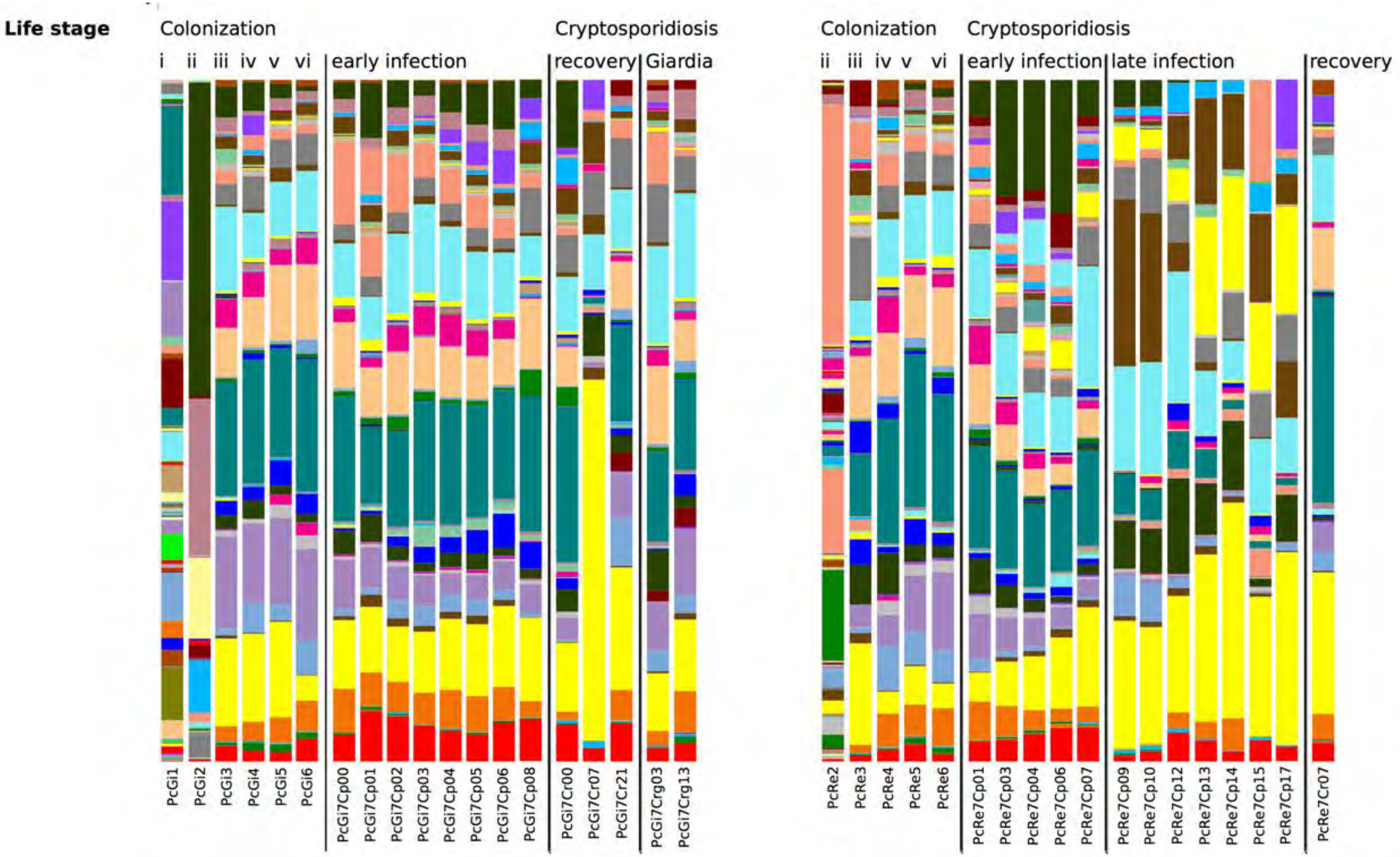


Don't just show the data – tell a story!

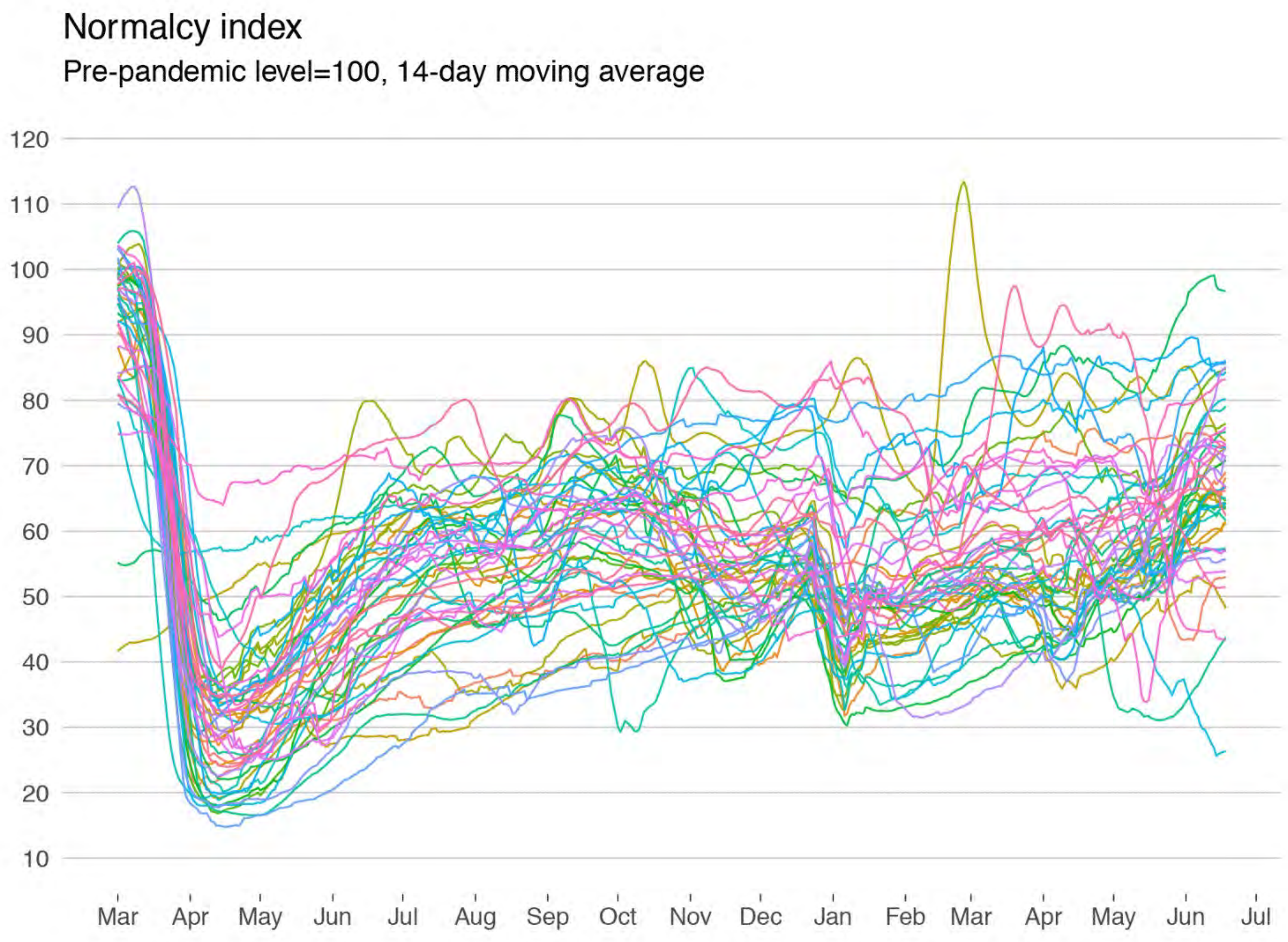




All the data doesn't tell a story

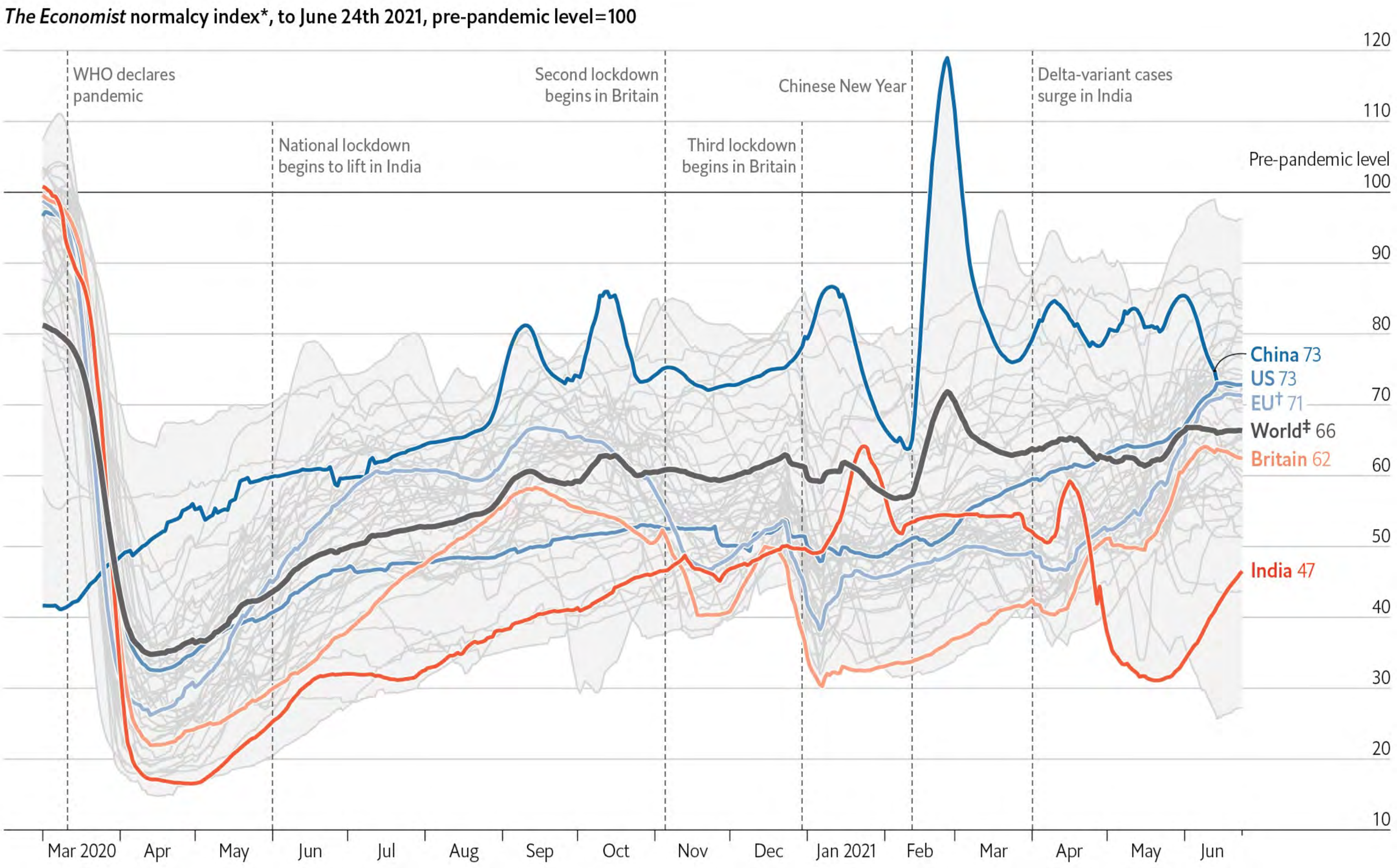


All the data doesn't tell a story



The Economist: Off the Charts newsletter – Aug 10, 2021
Between the lines: How to declutter a chart
Marie Segger, Data Journalist

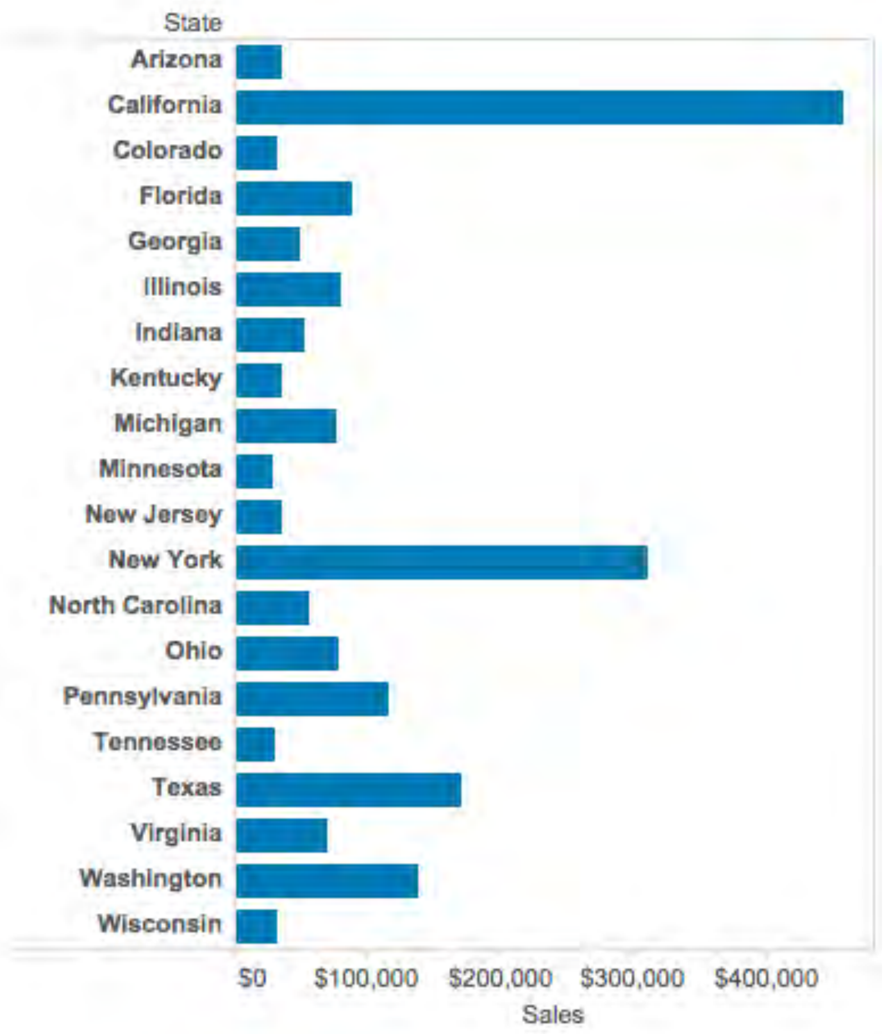
https://view.e.economist.com/?qs=2a8a99a7c5829c773a15e1b8a20305bee3f0832c13cba5acd5029208d271be68b4f6c48a2a5026368b033da213ae2b0665fabba975d24e568b9612d1d35885839287043cbbc8ca91e89742d62bad0554



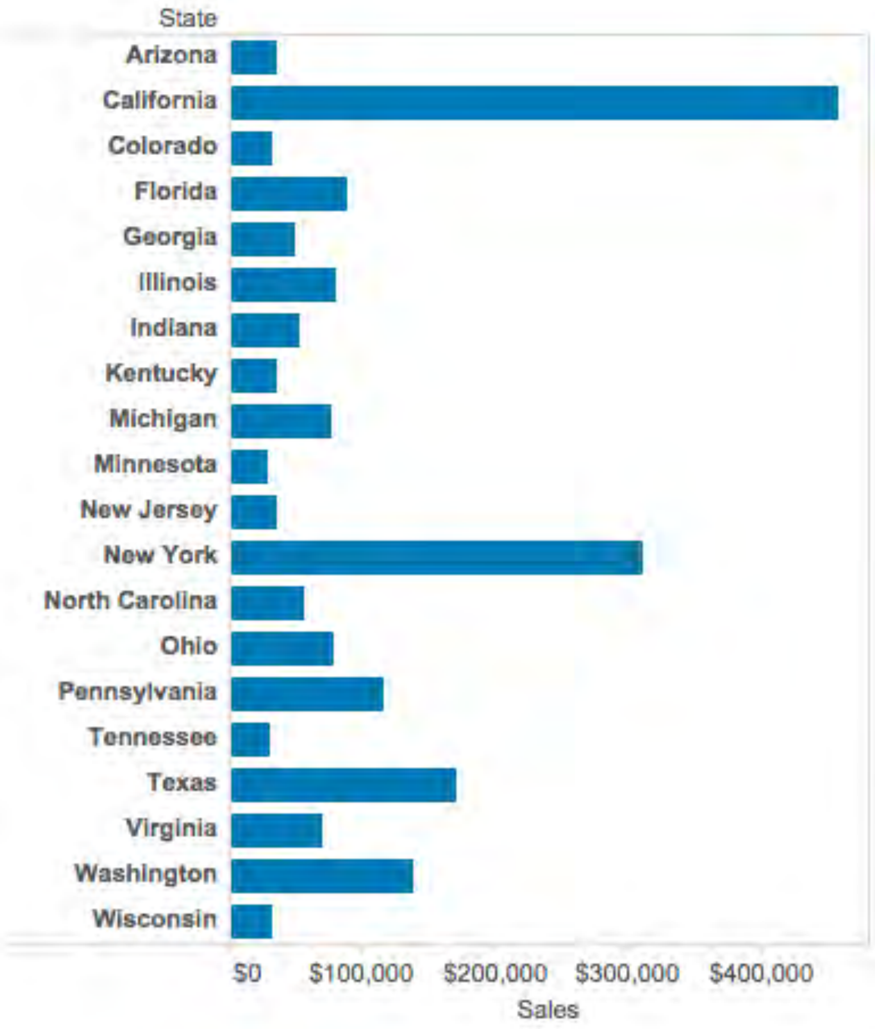
https://www.economist.com/graphic-detail/tracking-the-return-to-normalcy-after-covid-19

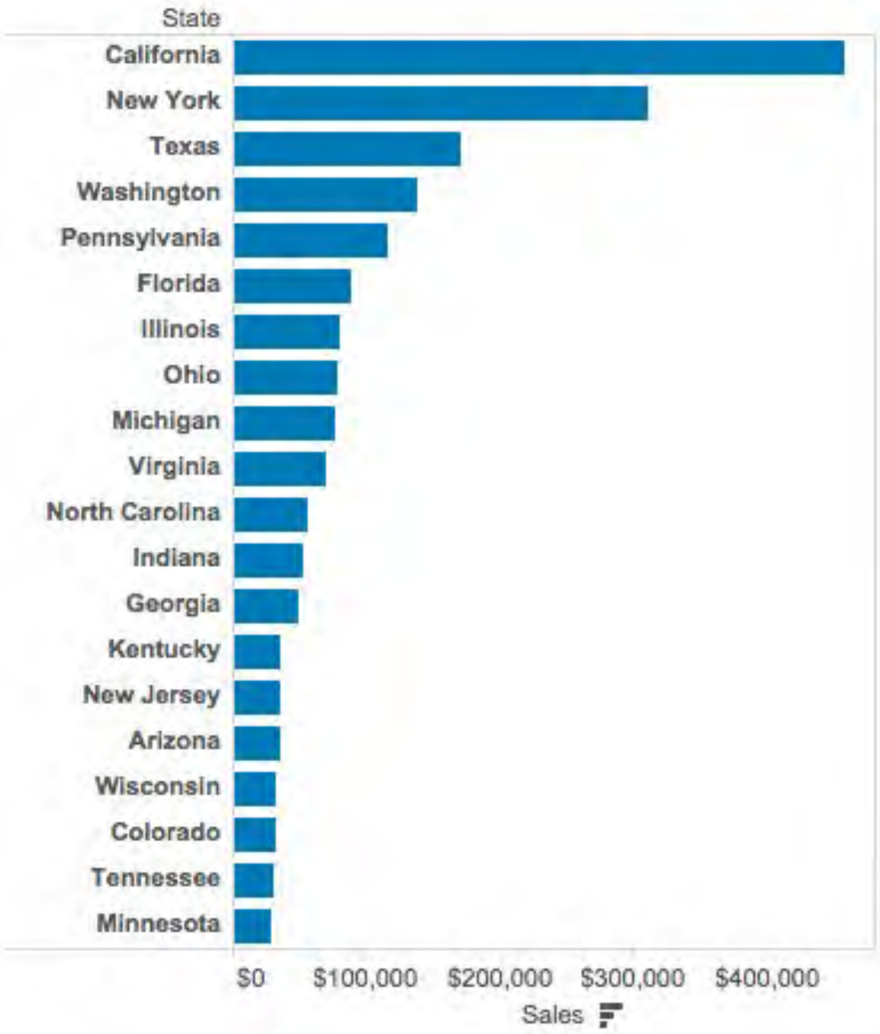
Common missteps

Default ordering hides patterns



Sorting reveals patterns



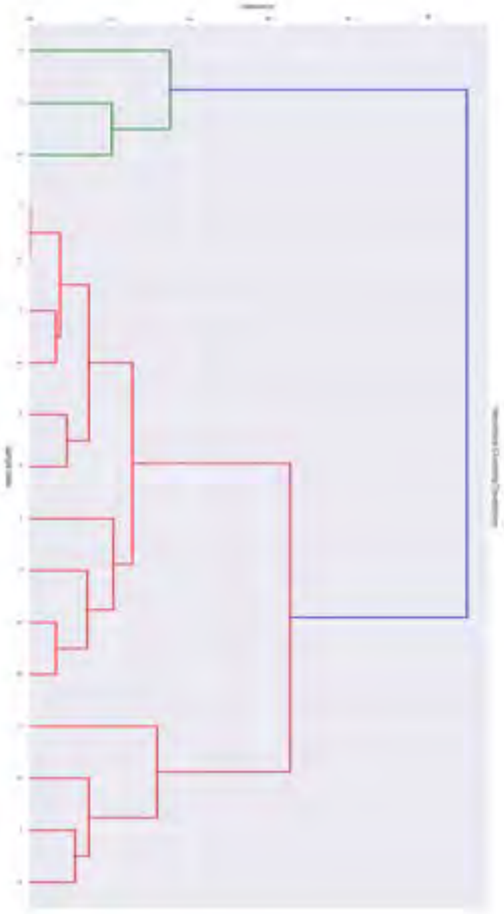


Alphabetical again hides patterns

Phylum	HgBp	HgBs	HgBv	HgBw	AfHe	AfNu	AmTt	AmMx	AmBb
Acidobacteria			0.42						
Actinobacteria			0.55	1.01	0.42	1.21			0.80
Bacteroidetes	2.36	2.58	2.51	2.52	0.87	1.07	0.72	0.86	0.97
Chlorobi	0.45								
Cyanobacteria	1.11	1.11	1.10	1.20	2.23	2.93	1.51	1.77	1.11
Fibrobacteres	0.58	1.11	0.61	1.03					
Firmicutes	2.37	2.46	2.47	2.31	2.88	1.34	2.42	2.40	2.49
Lentisphaerae	1.17	1.38	1.30	1.00	0.49	0.47	0.47	0.45	0.47
Nitrospirae				0.79					
Planctomycetes	0.94			0.81					
Proteobacteria	1.79	1.68	1.90	1.79	1.52	2.09	2.79	2.74	2.77
Spirochaetes	1.52	1.40	1.40	1.55	0.45				
Synergistetes			0.42						
TM7	1.97	0.85	0.95	1.61	0.58	0.70	0.56	0.64	0.69
Tenericutes	1.17		0.74	0.99	0.38				
Verrucomicrobia	0.89		0.89	1.08					0.42
unclassified_Haloplasm							0.49		

Clustering to see response groups

Phylum	HgBp	HgBs	HgBv	HgBw	AfHe	AfNu	AmTt	AmMx	AmBb
Cyanobacteria	1.11	1.11	1.10	1.20	2.23	2.93	1.51	1.77	1.11
Firmicutes	2.37	2.46	2.47	2.31	2.88	1.34	2.42	2.40	2.49
Proteobacteria	1.79	1.68	1.90	1.79	1.52	2.09	2.79	2.74	2.77
Acidobacteria			0.42						
Synergistetes			0.42						
Chlorobi	0.45								
unclassified_Haloplasm							0.49		
Nitrospirae				0.79					
Planctomycetes	0.94			0.81					
Actinobacteria			0.55	1.01	0.42	1.21			0.80
Fibrobacteres	0.58	1.11	0.61	1.03					
Tenericutes	1.17		0.74	0.99	0.38				
Verrucomicrobia	0.89		0.89	1.08					0.42
Bacteroidetes	2.36	2.58	2.51	2.52	0.87	1.07	0.72	0.86	0.97
TM7	1.97	0.85	0.95	1.61	0.58	0.70	0.56	0.64	0.69
Lentisphaerae	1.17	1.38	1.30	1.00	0.49	0.47	0.47	0.45	0.47
Spirochaetes	1.52	1.40	1.40	1.55	0.45				



Tables are notorious for hiding data patterns!

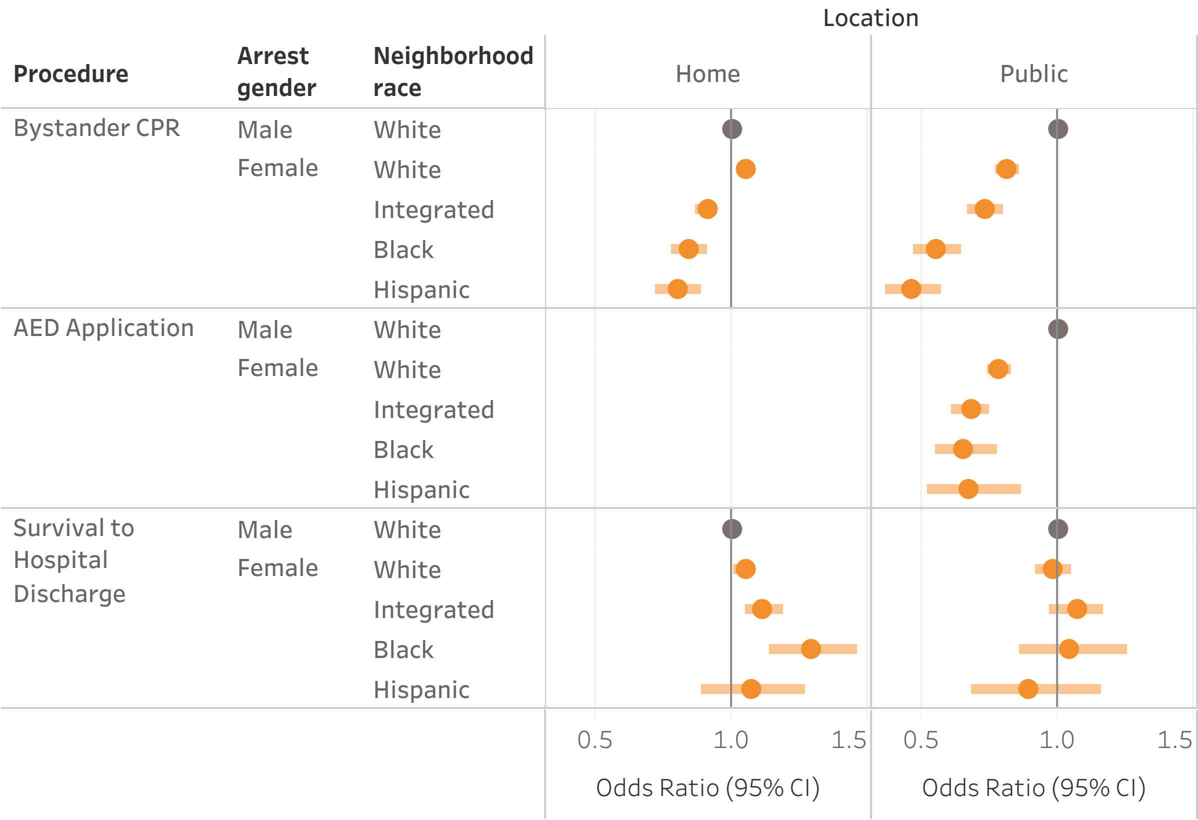
	Home	*Public*
Bystander CPR	**OR (95% CI)**	**OR (95% CI)**
Female arrest in a White neighborhood	1.05 (1.02,1.07)	0.81 (0.77, 0.86)
Female arrest in a Black neighborhood	0.84 (0.78,0.91)	0.55 (0.47, 0.65)
Female arrest in a Hispanic neighborhood	0.80 (0.72,0.89)	0.46 (0.37, 0.57)
Female arrest in an Integrated neighborhood	0.91 (0.87,0.95)	0.73 (0.67, 0.80)
Male arrest in a White neighborhood	reference	reference
AED Application		
Female arrest in a White neighborhood	-	0.78 (0.74, 0.83)
Female arrest in a Black neighborhood	-	0.65 (0.55, 0.78)
Female arrest in a Hispanic neighborhood	-	0.67 (0.52, 0.87)
Female arrest in an Integrated neighborhood	-	0.68 (0.61, 0.75)
Male arrest in a White neighborhood	reference	reference
Survival to Hospital Discharge		
Female arrest in a White neighborhood	1.05 (1.01, 1.09)	0.98 (0.92, 1.05)
Female arrest in a Black neighborhood	1.29 (1.14, 1.46)	1.04 (0.86, 1.26)
Female arrest in a Hispanic neighborhood	1.07 (0.89, 1.27)	0.89 (0.68, 1.16)
Female arrest in an Integrated neighborhood	1.11 (1.05, 1.19)	1.07 (0.97, 1.17)
Male arrest in a White neighborhood	reference	reference

Blewer et al (manuscript in preparation for submission)

	Home	*Public*
Bystander CPR	**OR (95% CI)**	**OR (95% CI)**
Female arrest in a White neighborhood	1.05 (1.02,1.07)	0.81 (0.77, 0.86)
Female arrest in a Black neighborhood	0.84 (0.78,0.91)	0.55 (0.47, 0.65)
Female arrest in a Hispanic neighborhood	0.80 (0.72,0.89)	0.46 (0.37, 0.57)
Female arrest in an Integrated neighborhood	0.91 (0.87,0.95)	0.73 (0.67, 0.80)
Male arrest in a White neighborhood	reference	reference
AED Application		
Female arrest in a White neighborhood	-	0.78 (0.74, 0.83)
Female arrest in a Black neighborhood	-	0.65 (0.55, 0.78)
Female arrest in a Hispanic neighborhood	-	0.67 (0.52, 0.87)
Female arrest in an Integrated neighborhood	-	0.68 (0.61, 0.75)
Male arrest in a White neighborhood	reference	reference
Survival to Hospital Discharge		
Female arrest in a White neighborhood	1.05 (1.01, 1.09)	0.98 (0.92, 1.05)
Female arrest in a Black neighborhood	1.29 (1.14, 1.46)	1.04 (0.86, 1.26)
Female arrest in a Hispanic neighborhood	1.07 (0.89, 1.27)	0.89 (0.68, 1.16)
Female arrest in an Integrated neighborhood	1.11 (1.05, 1.19)	1.07 (0.97, 1.17)
Male arrest in a White neighborhood	reference	reference

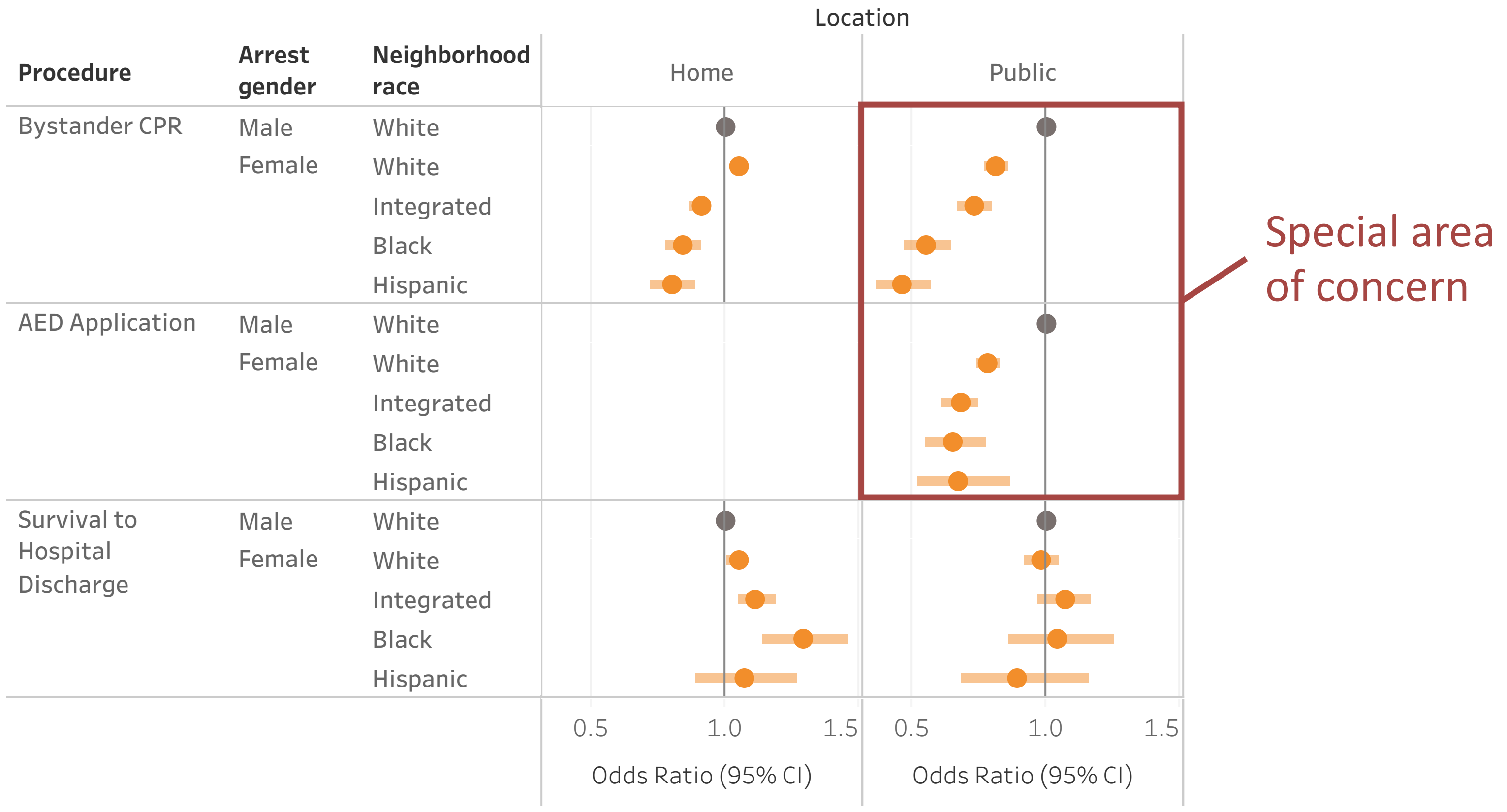
	Home	***Public***
Bystander CPR	**Odds Ratio** (95% CI)	**Odds Ratio** (95% CI)
Female arrest in a **White** neighborhood	1.05 (1.02,1.07)	0.81 (0.77, 0.86)
Female arrest in a **Black** neighborhood	0.84 (0.78,0.91)	0.55 (0.47, 0.65)
Female arrest in a **Hispanic** neighborhood	0.80 (0.72,0.89)	0.46 (0.37, 0.57)
Female arrest in an **Integrated** neighborhood	0.91 (0.87,0.95)	0.73 (0.67, 0.80)
Male arrest in a **White** neighborhood	reference	reference
AED Application		
Female arrest in a **White** neighborhood	-	0.78 (0.74, 0.83)
Female arrest in a **Black** neighborhood	-	0.65 (0.55, 0.78)
Female arrest in a **Hispanic** neighborhood	-	0.67 (0.52, 0.87)
Female arrest in an **Integrated** neighborhood	-	0.68 (0.61, 0.75)
Male arrest in a **White** neighborhood	reference	reference
Survival to Hospital Discharge		
Female arrest in a **White** neighborhood	1.05 (1.01, 1.09)	0.98 (0.92, 1.05)
Female arrest in a **Black** neighborhood	1.29 (1.14, 1.46)	1.04 (0.86, 1.26)
Female arrest in a **Hispanic** neighborhood	1.07 (0.89, 1.27)	0.89 (0.68, 1.16)
Female arrest in an **Integrated** neighborhood	1.11 (1.05, 1.19)	1.07 (0.97, 1.17)
Male arrest in a **White** neighborhood	reference	reference

Blewer et al (manuscript in preparation for submission)



	Home	***Public***
Bystander CPR	**Odds Ratio** (95% CI)	**Odds Ratio** (95% CI)
Female arrest in a **White** neighborhood	1.05 (1.02,1.07)	0.81 (0.77, 0.86)
Female arrest in a **Black** neighborhood	0.84 (0.78,0.91)	0.55 (0.47, 0.65)
Female arrest in a **Hispanic** neighborhood	0.80 (0.72,0.89)	0.46 (0.37, 0.57)
Female arrest in an **Integrated** neighborhood	0.91 (0.87,0.95)	0.73 (0.67, 0.80)
Male arrest in a **White** neighborhood	reference	reference
AED Application		
Female arrest in a **White** neighborhood	-	0.78 (0.74, 0.83)
Female arrest in a **Black** neighborhood	-	0.65 (0.55, 0.78)
Female arrest in a **Hispanic** neighborhood	-	0.67 (0.52, 0.87)
Female arrest in an **Integrated** neighborhood	-	0.68 (0.61, 0.75)
Male arrest in a **White** neighborhood	reference	reference
Survival to Hospital Discharge		
Female arrest in a **White** neighborhood	1.05 (1.01, 1.09)	0.98 (0.92, 1.05)
Female arrest in a **Black** neighborhood	1.29 (1.14, 1.46)	1.04 (0.86, 1.26)
Female arrest in a **Hispanic** neighborhood	1.07 (0.89, 1.27)	0.89 (0.68, 1.16)
Female arrest in an **Integrated** neighborhood	1.11 (1.05, 1.19)	1.07 (0.97, 1.17)
Male arrest in a **White** neighborhood	reference	reference

Blewer et al (manuscript in preparation for submission)



Blewer et al (manuscript in preparation for submission)

Color can be tricky

Rainbow colormaps distort

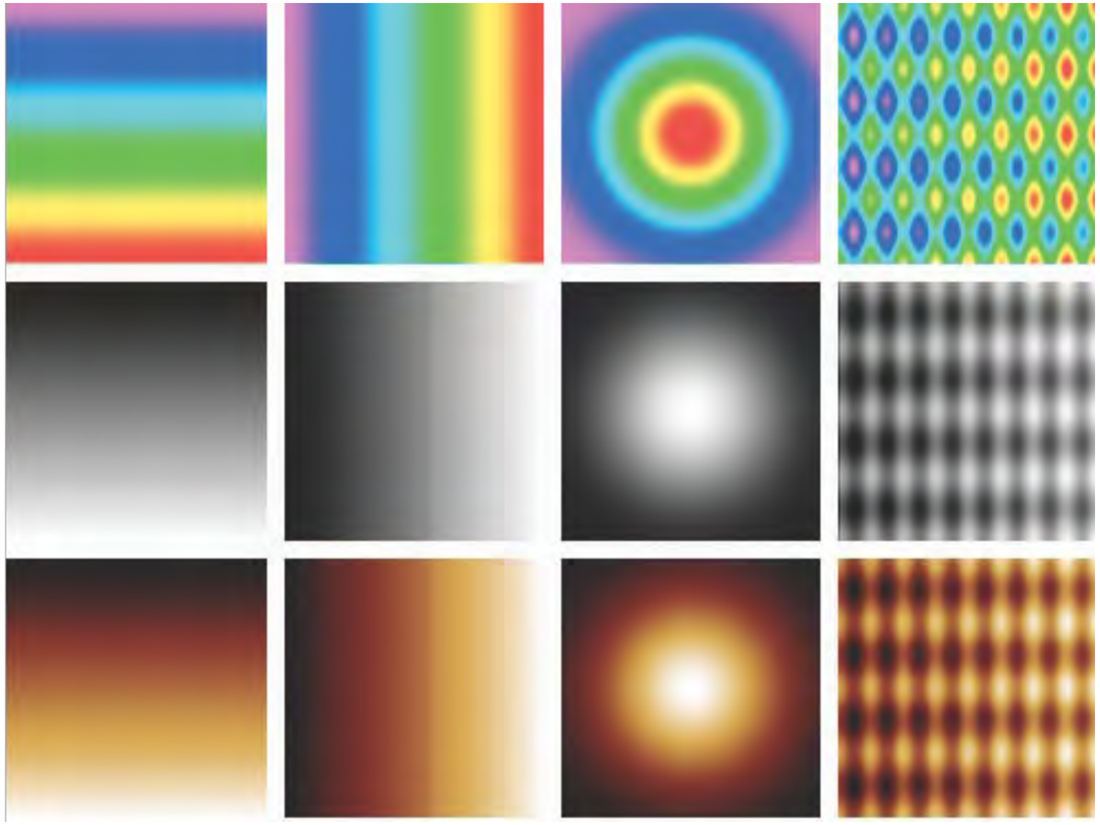
Bad because:

- No intuitive color ordering
- Makes the data look striped / banded

(a)

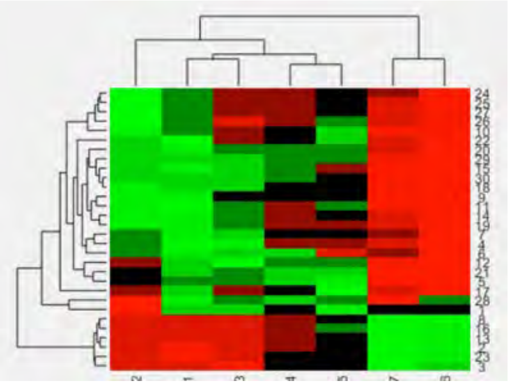
(b)

(c)

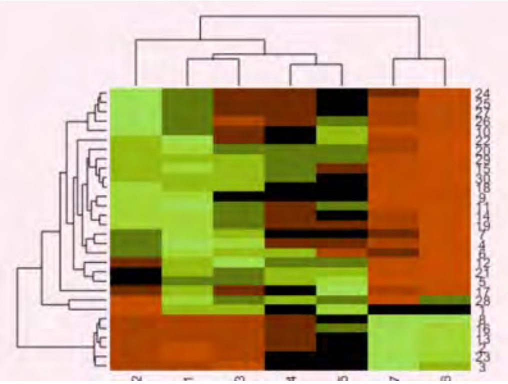


Borland, David, and Russell M. Taylor Ii. "Rainbow color map (still) considered harmful." *IEEE computer graphics and applications* 27.2 (2007). https://ieeexplore.ieee.org/document/4118486

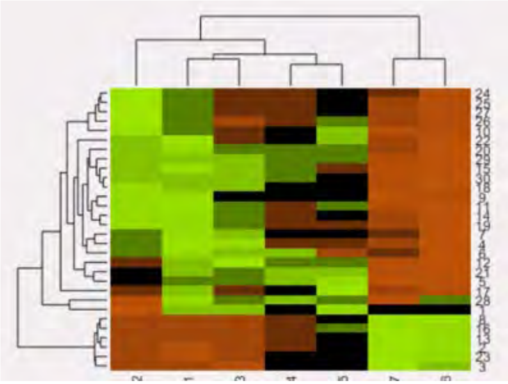
Red-green bad for common color deficiencies



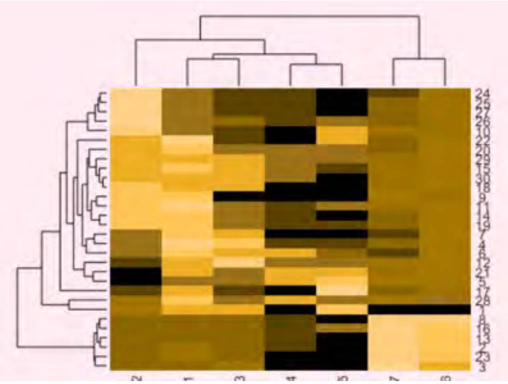
Normal



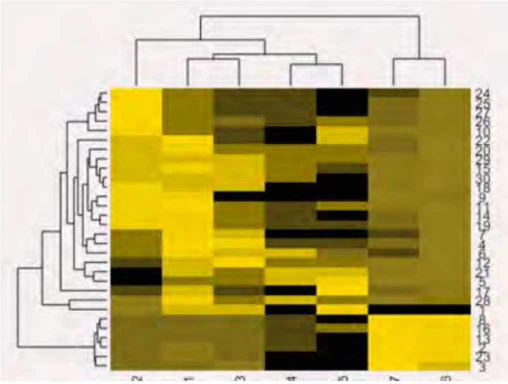
Green-weak/Deuteranomaly



Green-weak/Deuteranopia

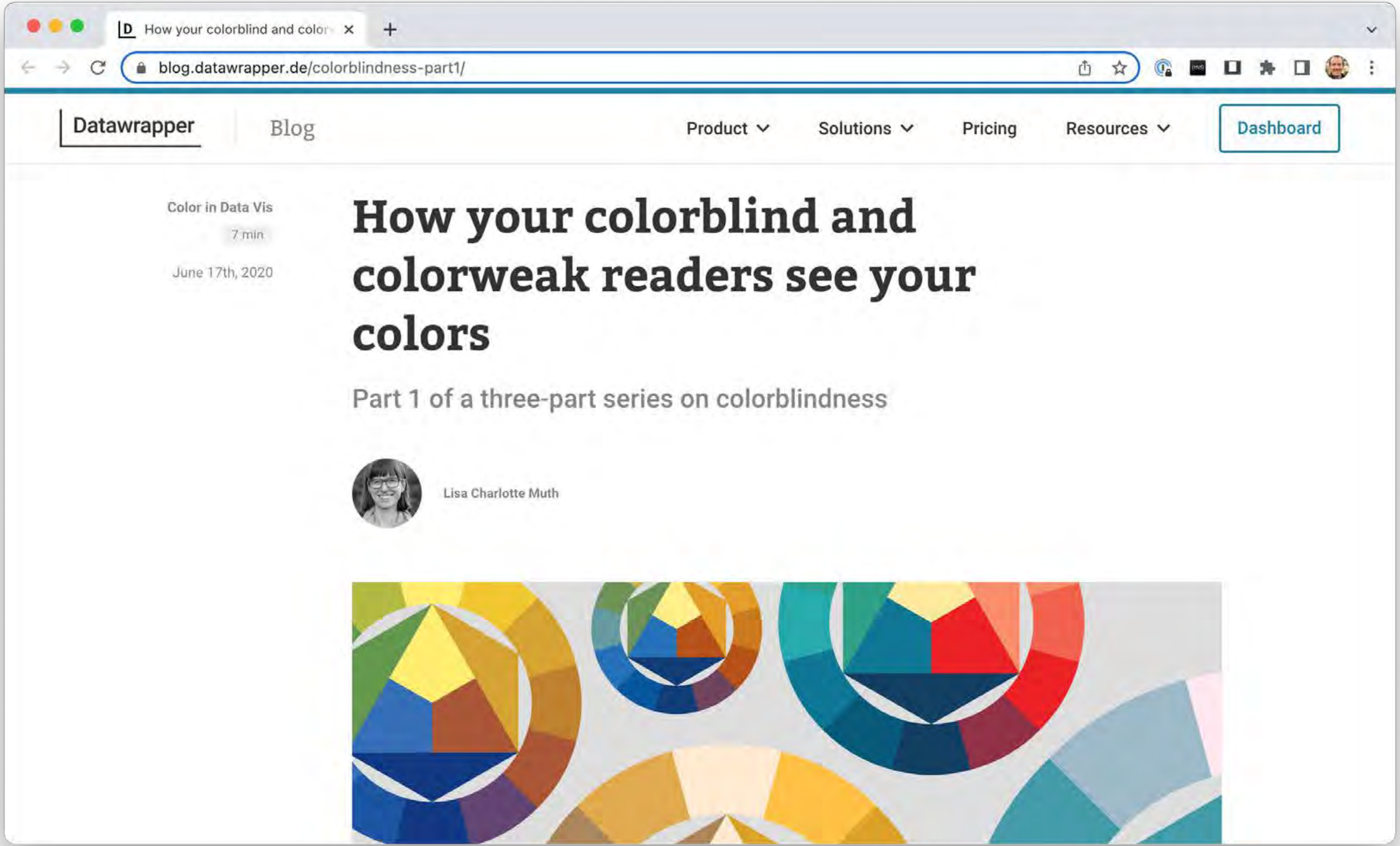


Red-weak/Protanomaly



Red-weak/Protanopia

https://www.color-blindness.com/coblis-color-blindness-simulator/



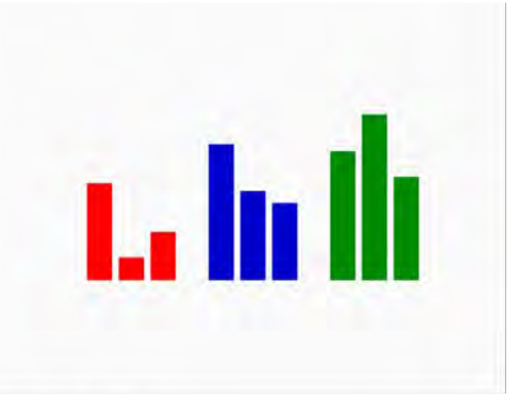
https://blog.datawrapper.de/colorblindness-part1/

Avoid pure saturated colors

How to pick more beautiful colors for your data visualization

https://blog.datawrapper.de/beautifulcolors/

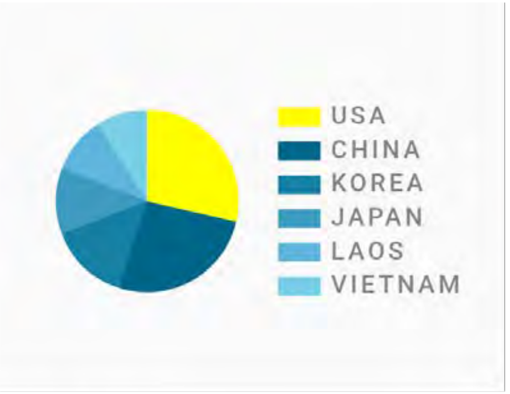
Avoid pure colors



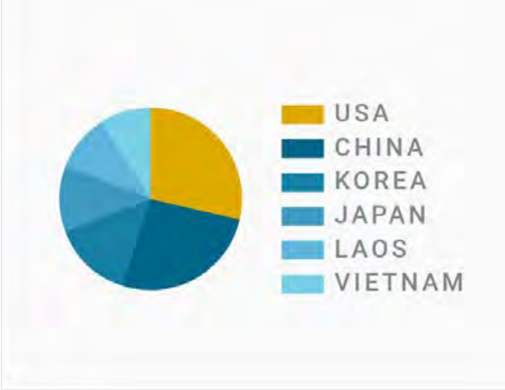
Not ideal

Better

Avoid bright, saturated colors



Not ideal



Better

Choose different colors for unordered sets

What to consider when choosing colors for data visualization
https://blog.datawrapper.de/colors/

Only use a gradient color palette for ordered categories

Not ideal *Better*

Keep your colors consistent across figures

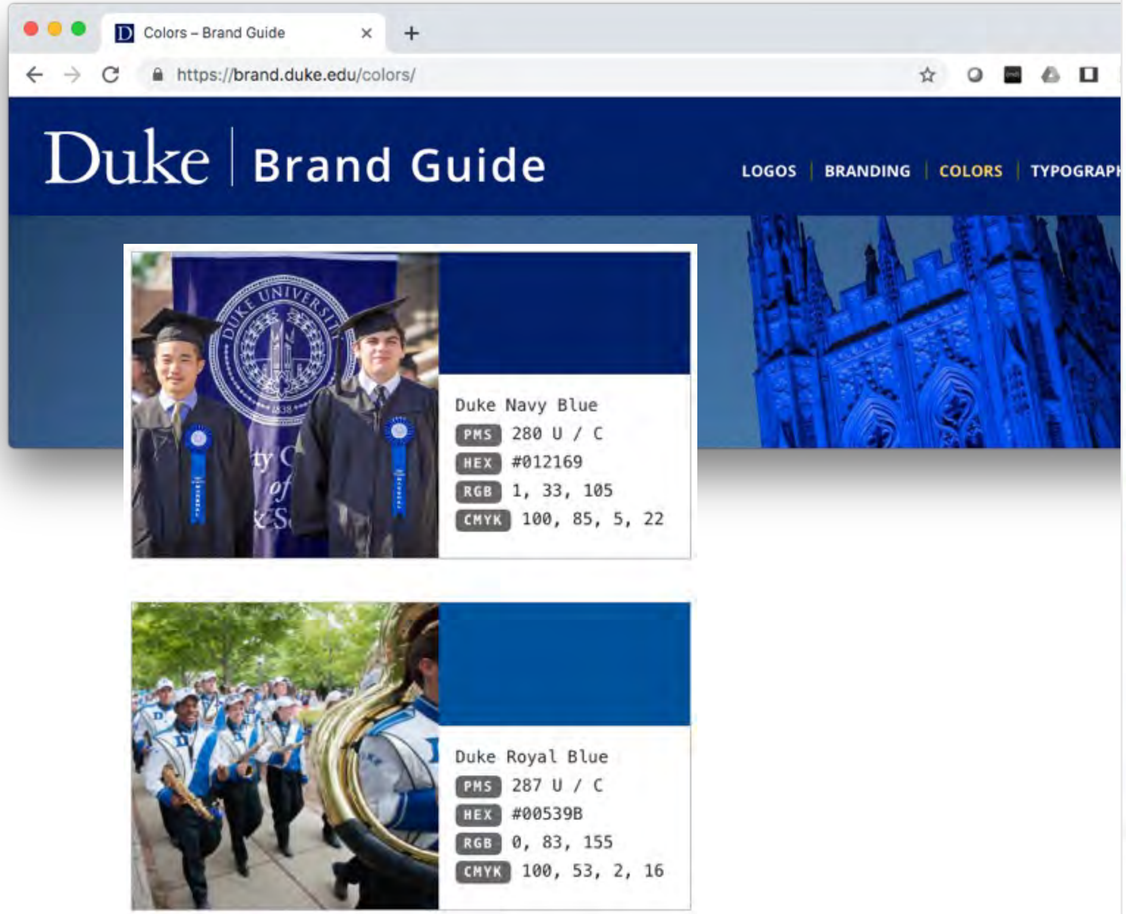
What to consider when choosing colors for data visualization
https://blog.datawrapper.de/colors/

Consider using the same color for the same variables

Not ideal

Better

Color schemes · design style/brand guides



Extended Palette

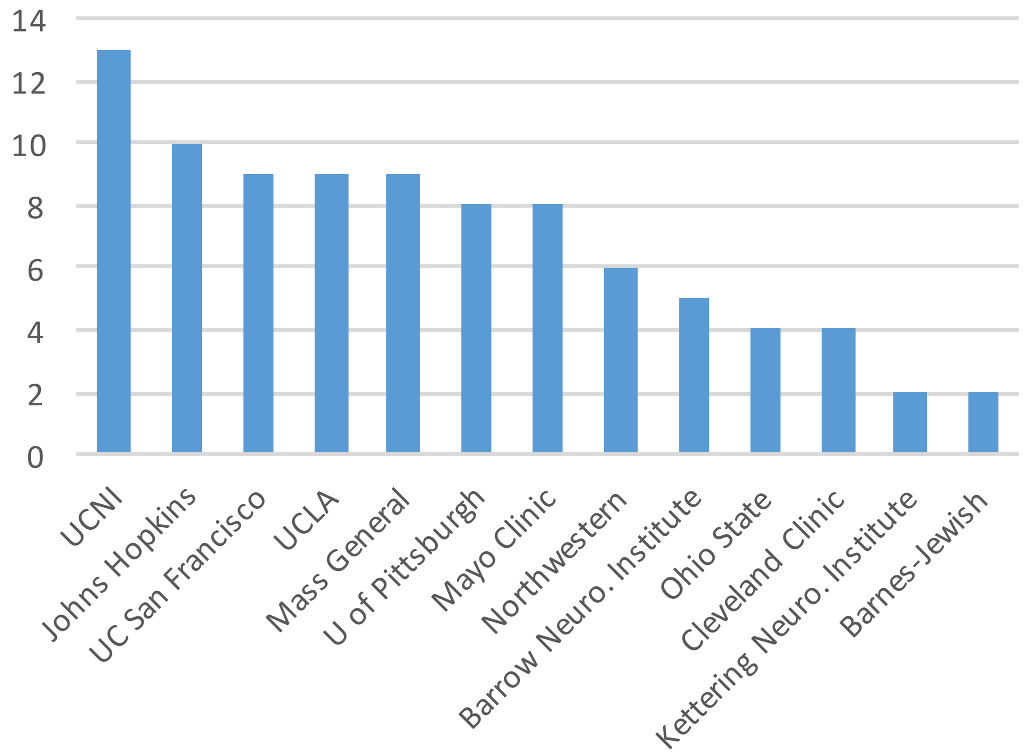
The colors in Duke's extended palette are intended for use as secondary and tertiary colors in design projects. They were selected to complement Duke Navy Blue and should be used for a range of elements including graphic accents, typography, backgrounds, call-to-action buttons and more.

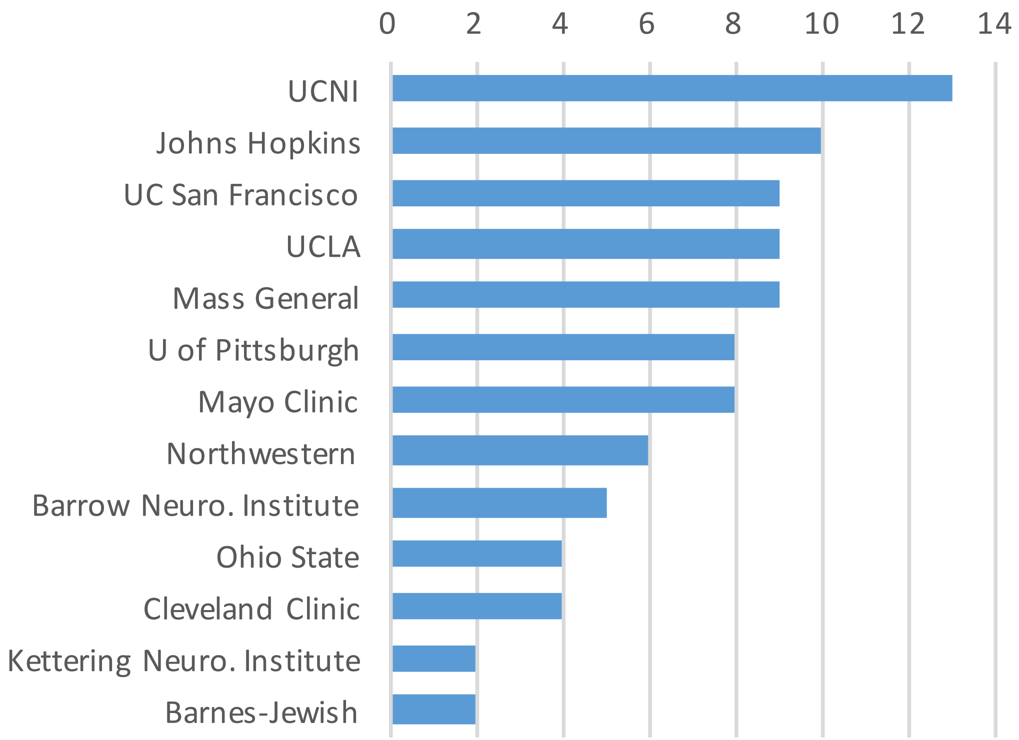
Copper	PMS 166 U / C	HEX	#C84E00	CMYK	0, 76, 100, 0	RGB	200, 78, 0
Persimmon	PMS 1375 U / C	HEX	#E89923	CMYK	0, 45, 95, 0	RGB	232, 153, 35
Dandelion	PMS 114 U / 121 C	HEX	#FFD960	CMYK	0, 8, 70, 0	RGB	255, 217, 96
Piedmont	PMS 382 U / 376 C	HEX	#A1B70D	CMYK	54, 0, 100, 0	RGB	161, 183, 13
Eno	PMS 3262 U / 326 C	HEX	#339898	CMYK	81, 0, 39, 0	RGB	51, 152, 152
Magnolia	PMS 328 U / 323 C	HEX	#1D6363	CMYK	96, 16, 42, 57	RGB	29, 99, 99
Prussian Blue	PMS 301 U / 7692 C	HEX	#005587	CMYK	100, 45, 0, 45	RGB	0, 85, 135
Shale Blue	PMS Pantone Process Blue U / 7461 C	HEX	#0577B1	CMYK	100, 0, 1, 3	RGB	5, 119, 177
Ironweed	PMS Pantone Purple U / 248 C	HEX	#993399	CMYK	35, 95, 0, 0	RGB	153, 51, 153
Hatteras	PMS 649 U / 656 C	HEX	#E2E6ED	CMYK	10, 2, 0, 0	RGB	226, 230, 237
Whisper Gray	PMS Cool Gray 1 U / C	HEX	#F3F2F1	CMYK	4, 2, 4, 8	RGB	243, 242, 241
Ginger Beer	PMS 9060 U / C	HEX	#FCF7E5	CMYK	0, 2, 15, 0	RGB	252, 247, 229
Dogwood	PMS 7530 U / C	HEX	#988675	CMYK	10, 18, 25, 32	RGB	152, 134, 117
Shackleford	PMS 7527 U / 2527 C	HEX	#DAD0C6	CMYK	3, 4, 14, 8	RGB	218, 208, 198
Cast Iron	PMS Black 3 U / C	HEX	#262626	CMYK	67, 44, 67, 95	RGB	38, 38, 38
Graphite	PMS Cool Gray 10 U / C	HEX	#666666	CMYK	40, 30, 20, 66	RGB	102, 102, 102
Granite	PMS 421 U / C	HEX	#B5B5B5	CMYK	13, 8, 11, 26	RGB	181, 181, 181
Limestone	PMS Cool Gray 2 U / C	HEX	#E5E5E5	CMYK	5, 3, 5, 11	RGB	229, 229, 229

https://brand.duke.edu/colors/

Minimal, readable text
to tell your story

Horizontal text is more readable





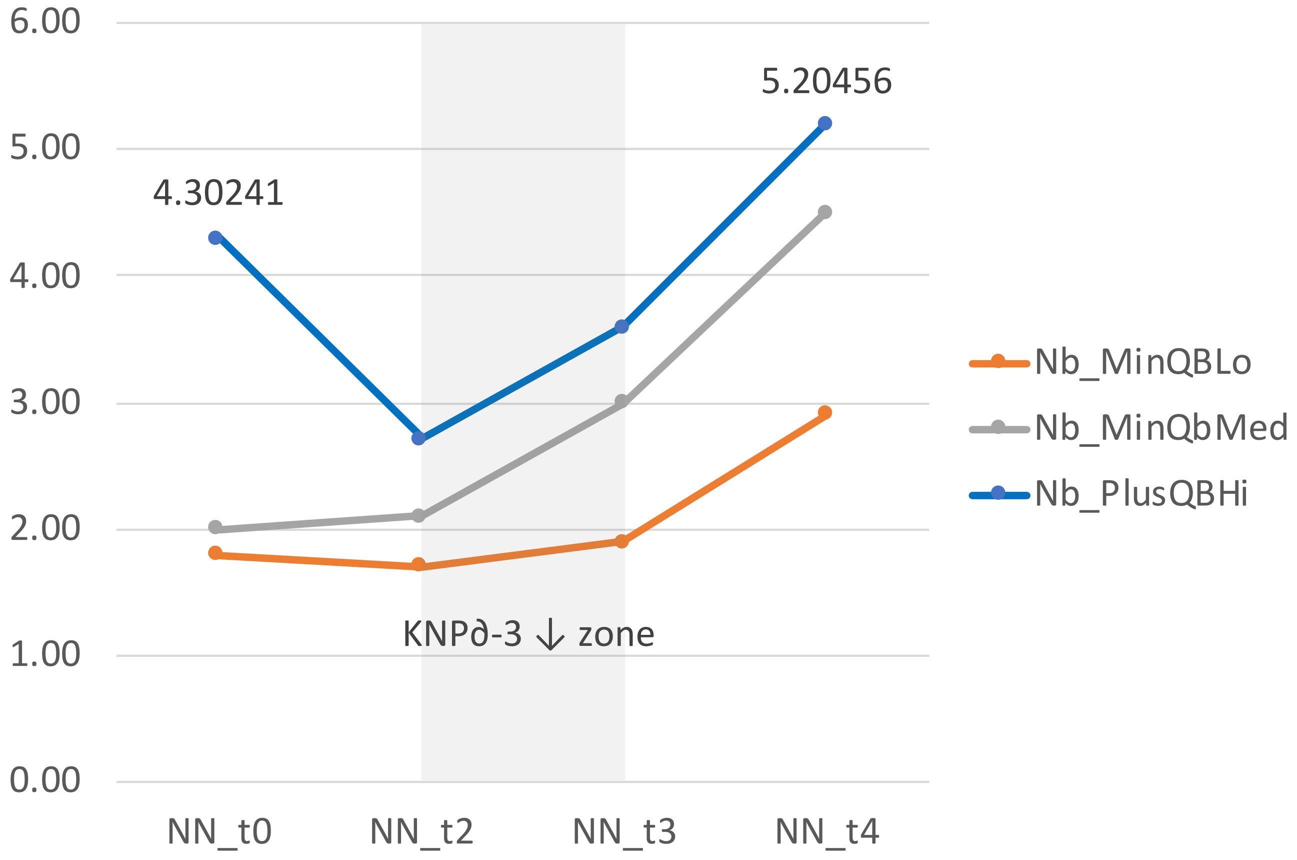
http://www.storytellingwithdata.com/2012/09/some-finer-points-of-data-visualization.html

Use human-readable labels

& Order legend same as visual when possible

Avoid:

- Abbreviations
- Jargon
- Variable names
- Useless decimal places



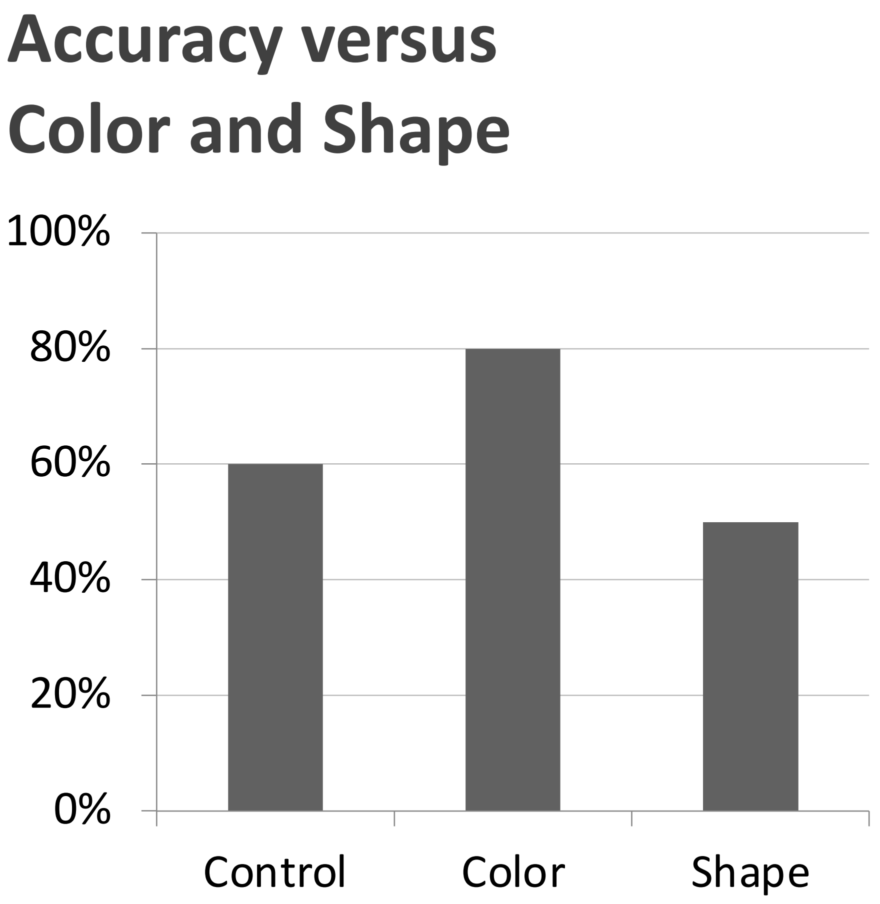
Direct stats output doesn't tell a story

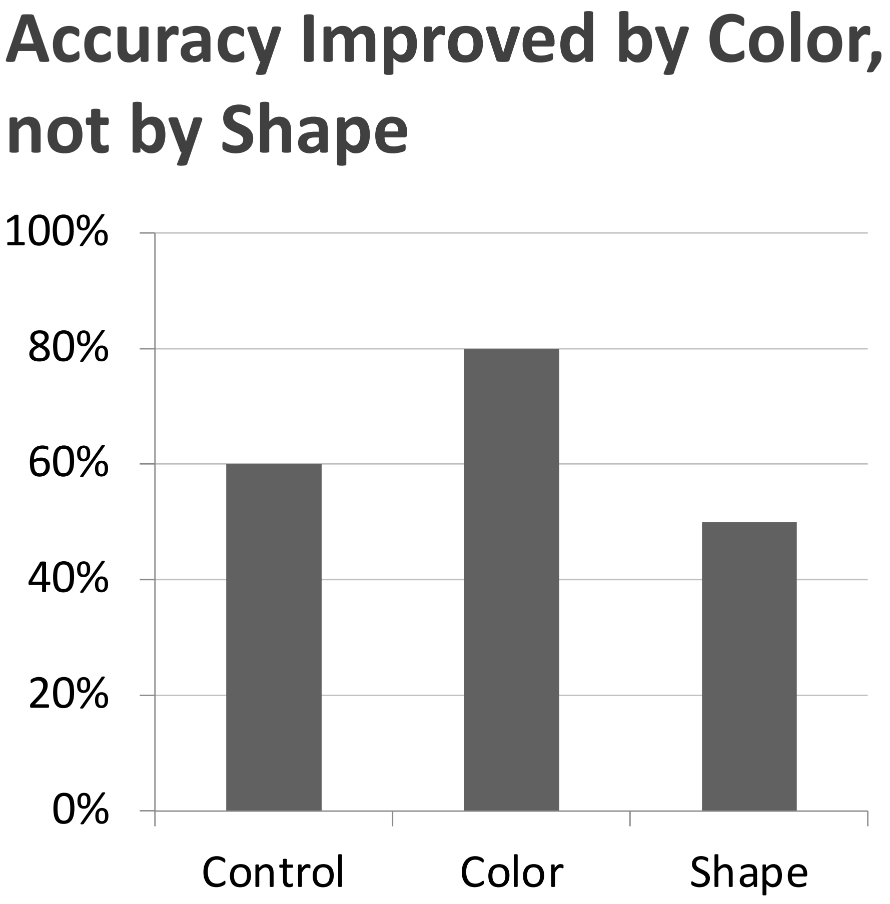
```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.28391    0.87438   9.474 1.44e-12 ***
cars$dist    0.16557    0.01749   9.464 1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1.1197	0.1446	-7.7454	0.0000
ageCent	0.1220	0.0376	3.2467	0.0017
gpCent	-0.0289	0.0103	-2.8166	0.0061
w_pctCent	3.6909	1.0096	3.6556	0.0005
def_ratingCent	0.0726	0.0359	2.0222	0.0464
ast_toCent	-0.4962	0.2592	-1.9145	0.0590
ast_ratioCent	0.0617	0.0203	3.0303	0.0033
dreb_pctCent	3.9947	2.1491	1.8588	0.0666
logsalaryCent	0.7050	0.1611	4.3753	0.0000
ptsCent	0.1180	0.0271	4.3611	0.0000
gpCent:logsalaryCent	-0.0137	0.0063	-2.1900	0.0314

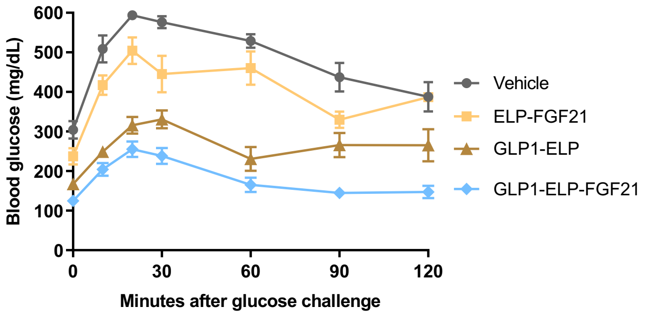
```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -3.307e+03  6.643e+02  -4.978 8.46e-07
rel.compact                  3.147e+03  4.466e+02   7.046 5.20e-12
surface.area                 1.793e+01  1.635e+00  10.964  < 2e-16
wall.area                   -1.021e+01  5.177e-01 -19.718  < 2e-16
height                      -6.623e+02  3.566e+01 -18.572  < 2e-16
glazing.area                 3.708e+01  2.714e+00  13.660  < 2e-16
glazing.dist0               -9.623e+00  1.661e+00  -5.793 1.13e-08
glazing.dist1               -5.659e-01  1.084e+00  -0.522 0.601978
glazing.dist2               -1.611e+00  1.077e+00  -1.496 0.135242
glazing.dist3               -6.769e-01  1.058e+00  -0.640 0.522463
glazing.dist4               -1.021e+00  1.077e+00  -0.948 0.343498
wall.area:roof.area          4.328e-02  1.812e-03  23.883  < 2e-16
wall.area:glazing.area       5.907e-02  6.603e-03   8.946  < 2e-16
wall.area:glazing.dist0     -1.387e-02  3.809e-03  -3.642 0.000294
wall.area:glazing.dist1      1.982e-04  2.489e-03   0.080 0.936555
wall.area:glazing.dist2      1.133e-03  2.650e-03   0.428 0.669065
wall.area:glazing.dist3     -5.624e-04  2.555e-03  -0.220 0.825873
wall.area:glazing.dist4      4.101e-04  2.576e-03   0.159 0.873600
rel.compact:surface.area    -5.160e+00  4.796e-01 -10.758  < 2e-16
surface.area:height          5.532e-01  3.135e-02  17.648  < 2e-16
surface.area:roof.area      -4.763e-02  2.784e-03 -17.110  < 2e-16
surface.area:wall.area      -4.940e-03  4.643e-04 -10.640  < 2e-16
surface.area:glazing.area   -5.800e-02  3.271e-03 -17.734  < 2e-16
surface.area:glazing.dist0   1.487e-02  2.001e-03   7.429 3.89e-13
surface.area:glazing.dist1   1.266e-03  1.303e-03   0.972 0.331390
surface.area:glazing.dist2   2.269e-03  1.281e-03   1.771 0.077056
surface.area:glazing.dist3   1.304e-03  1.265e-03   1.031 0.302897
surface.area:glazing.dist4   1.646e-03  1.287e-03   1.279 0.201406
rel.compact:height           1.916e+02  2.715e+01   7.055 4.88e-12
```

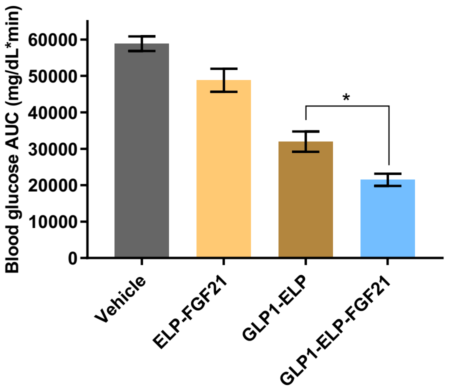
Active titles tell your story



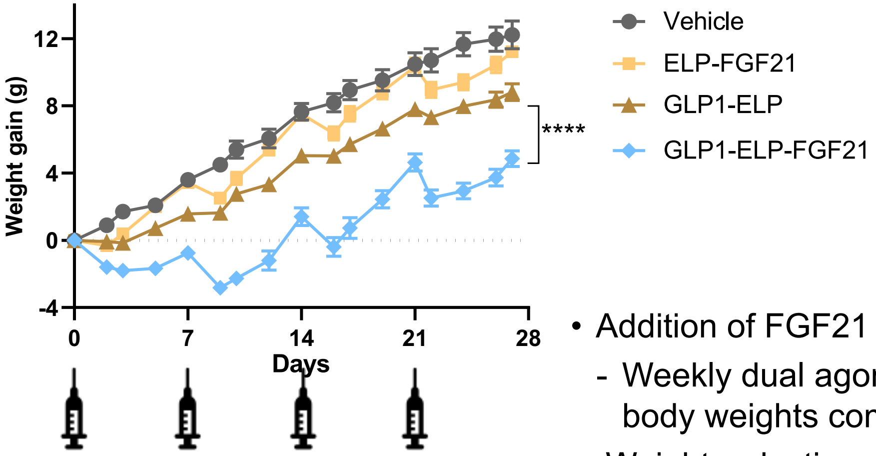


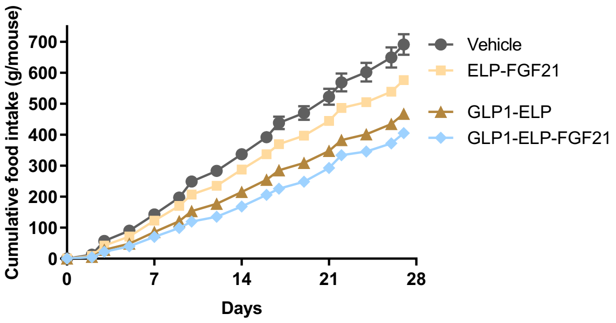
Dual agonist outperforms GLP1 receptor agonist





- GLP1-ELP-FGF21 treated mice display superior response to glucose challenge
  - Single treatment to *db/db* mice followed by fasted glucose bolus
  - Dual agonist group returns to baseline more quickly than equimolar dose of GLP1-ELP



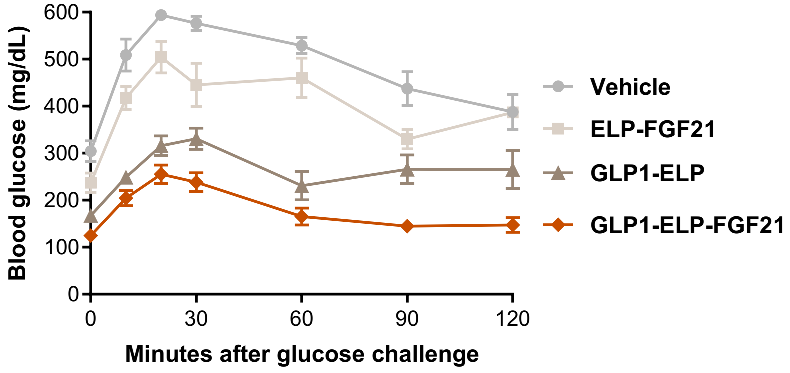


- Addition of FGF21 to GLP1-ELP results in substantial weight effect
  - Weekly dual agonist treatments to *db/db* mice significantly reduces body weights compared to equimolar GLP1-ELP treatments
- Weight reduction attributed to factor distinct from feeding
  - GLP1-ELP-FGF21 treated mice eat at same rate as GLP1-ELP

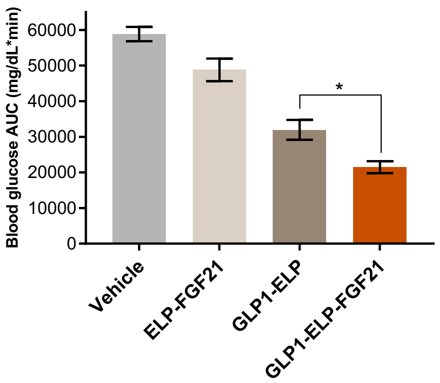
Caslin Gilroy

Dual agonist outperforms long-acting GLP-1 receptor agonist

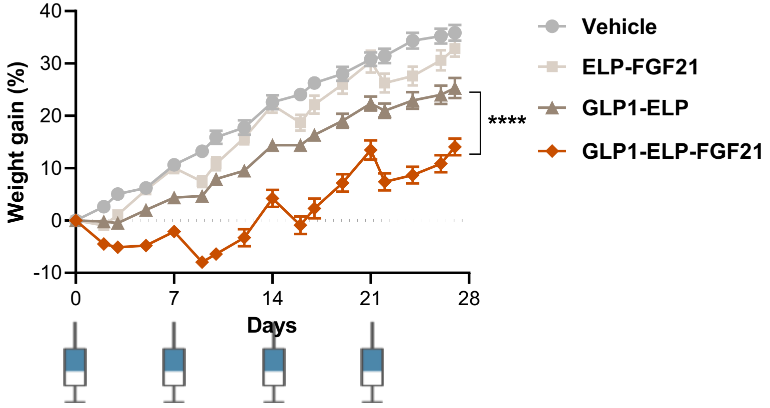
Hyperglycemic *db/db* mice challenged with a fasted glucose bolus



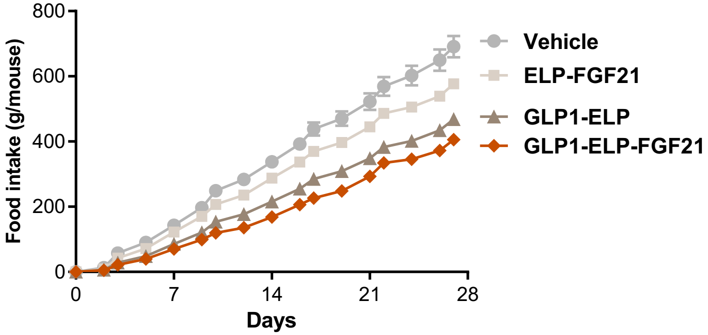
Dual agonist-treated group **responds to glucose spike more efficiently** than an equimolar dose of GLP1-ELP



Weekly dual agonist treatments to obese *db/db* mice results in **significantly lower body weights** compared to equimolar GLP1-ELP treatments

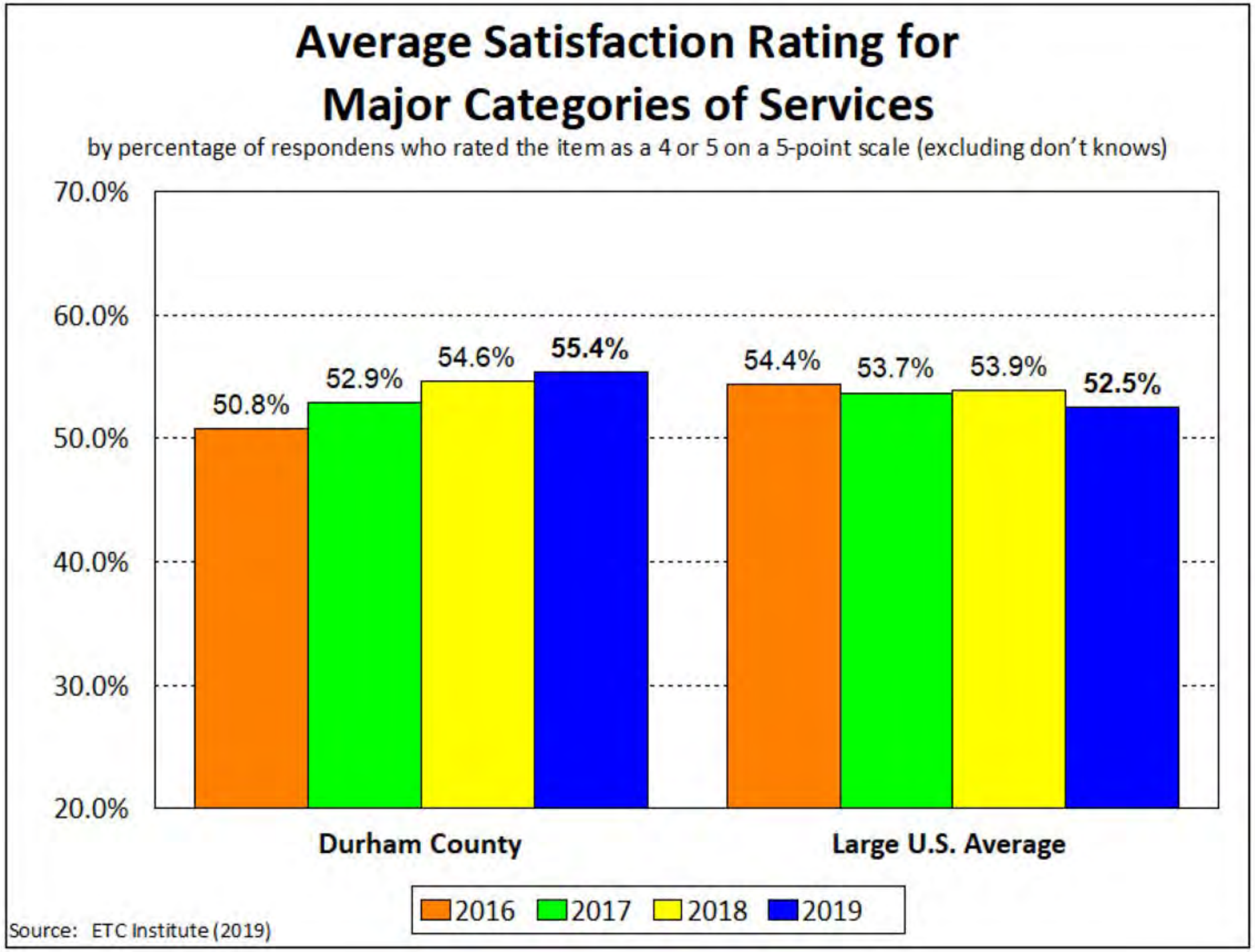


Dual agonist **inhibits weight gain without decreasing feed rate** compared to GLP1-ELP → altered energy balance



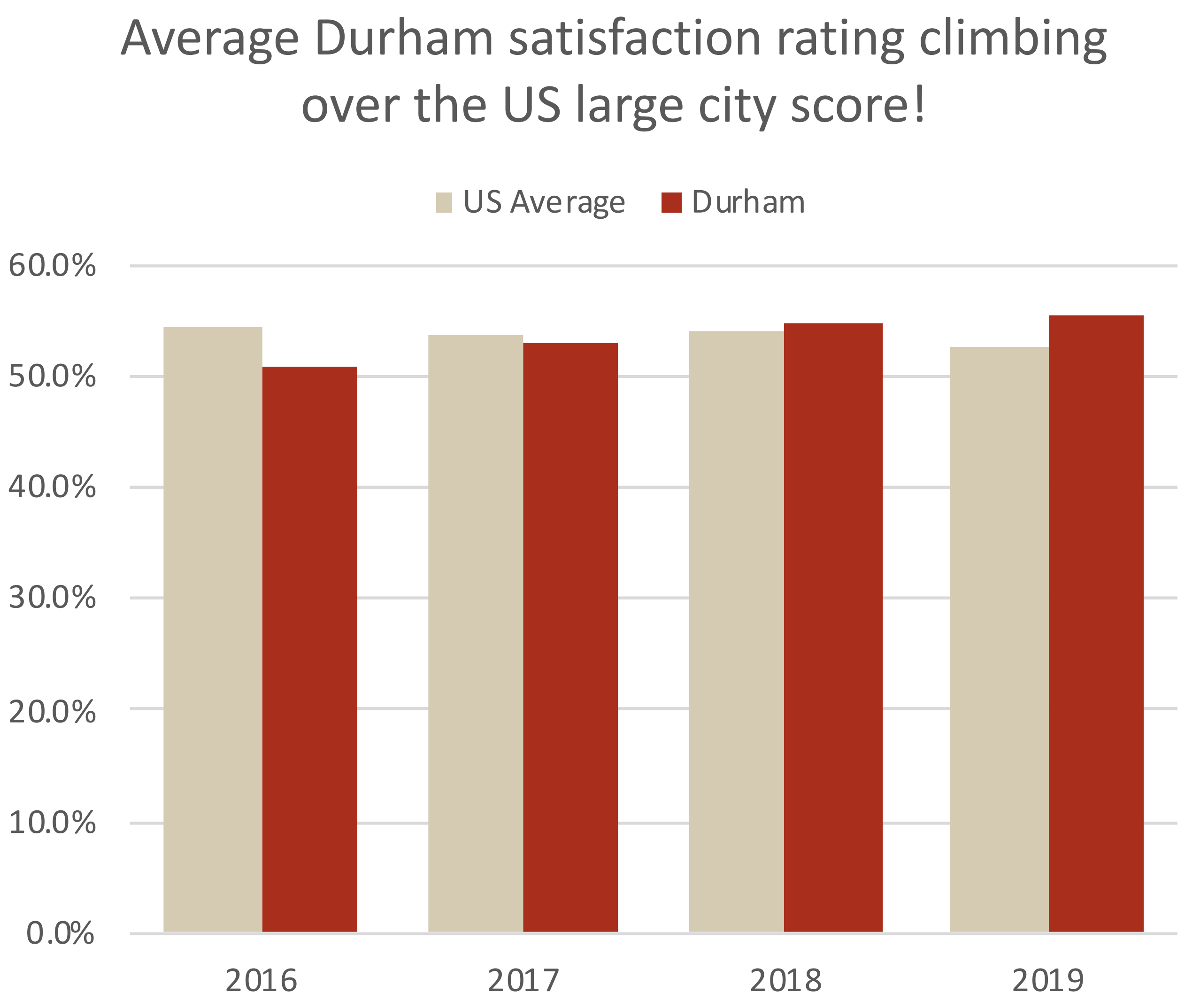
Caslin Gilroy

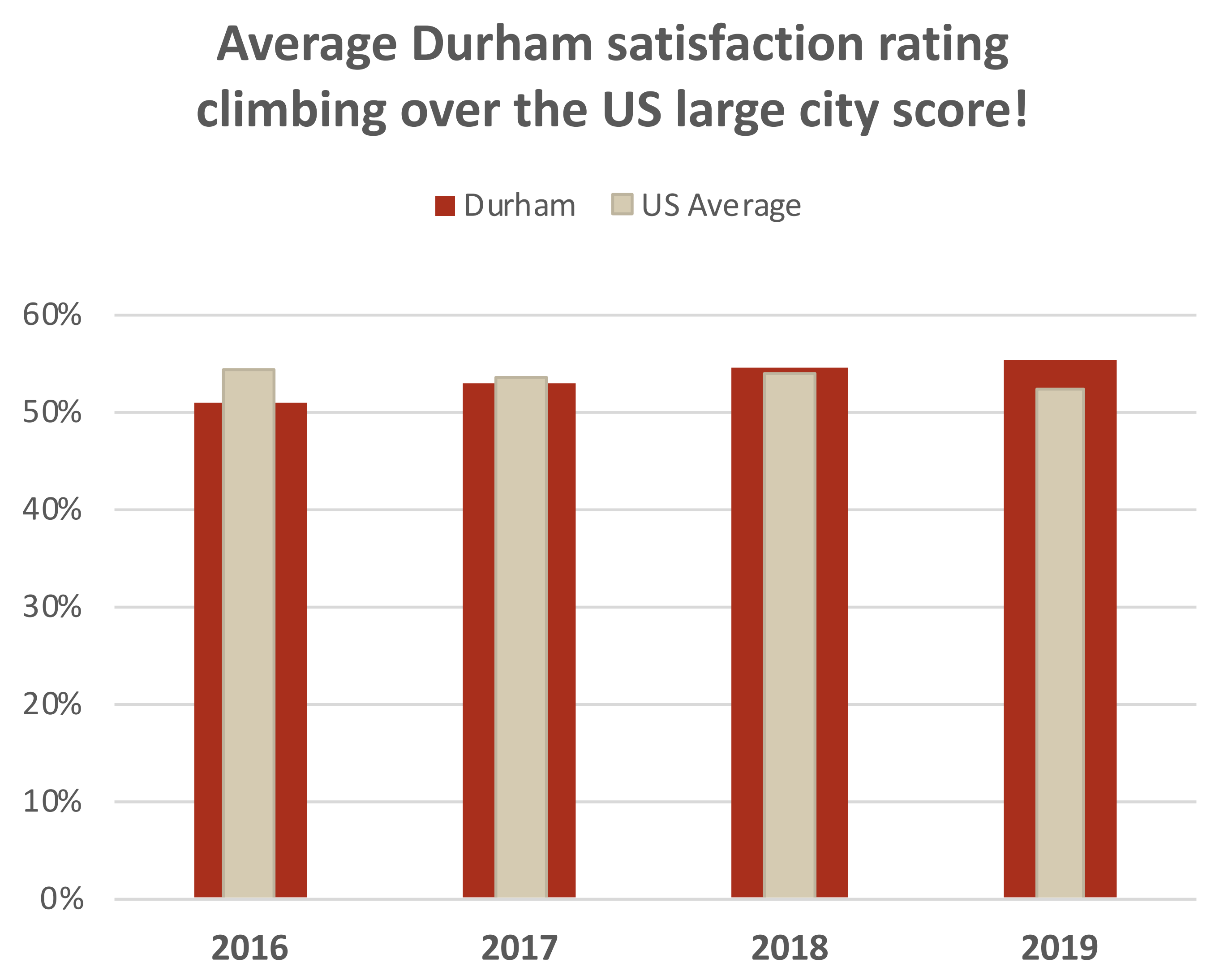
Figure critique & reworks

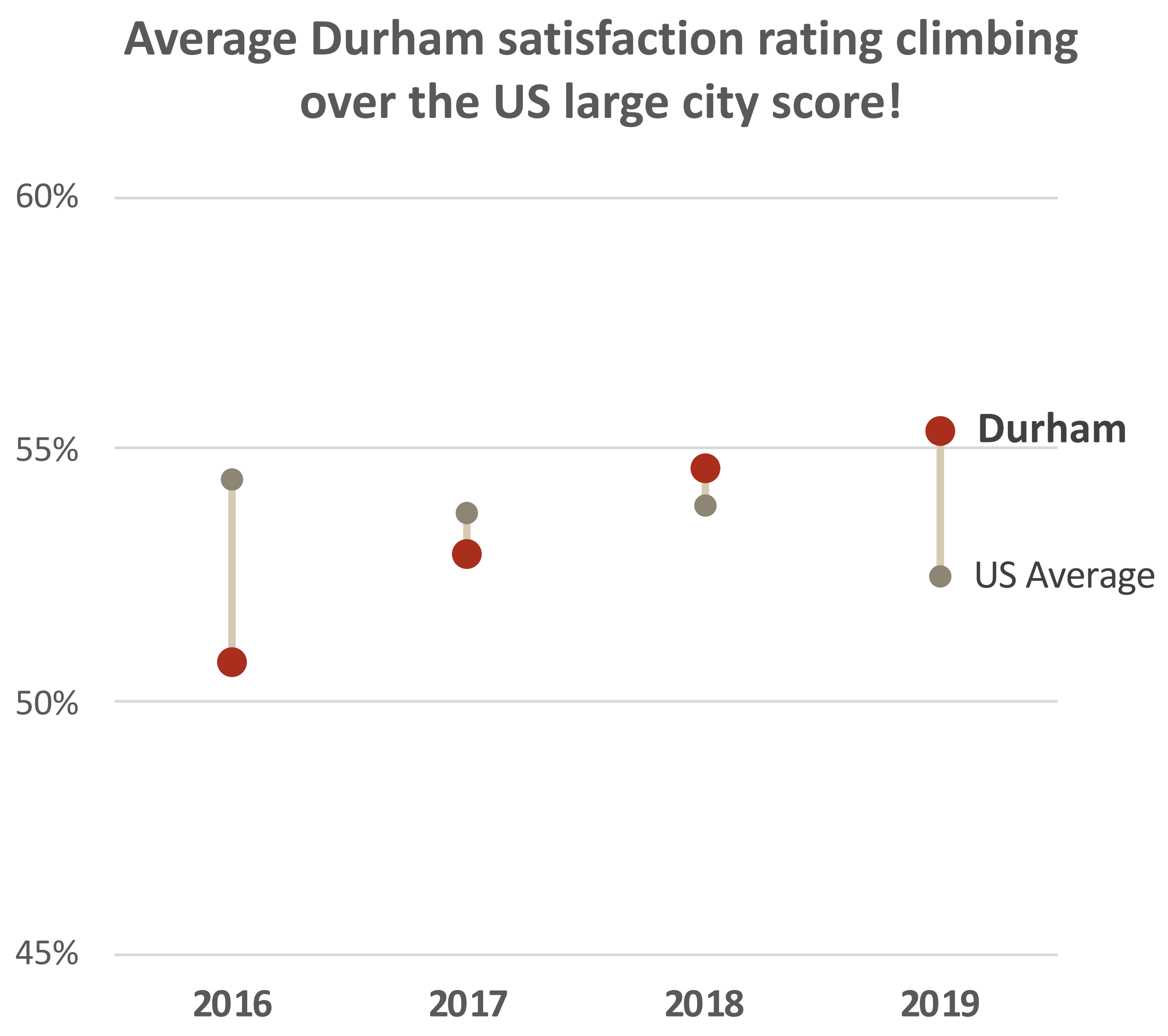


from Durham County 2019 Resident Survey Findings Report

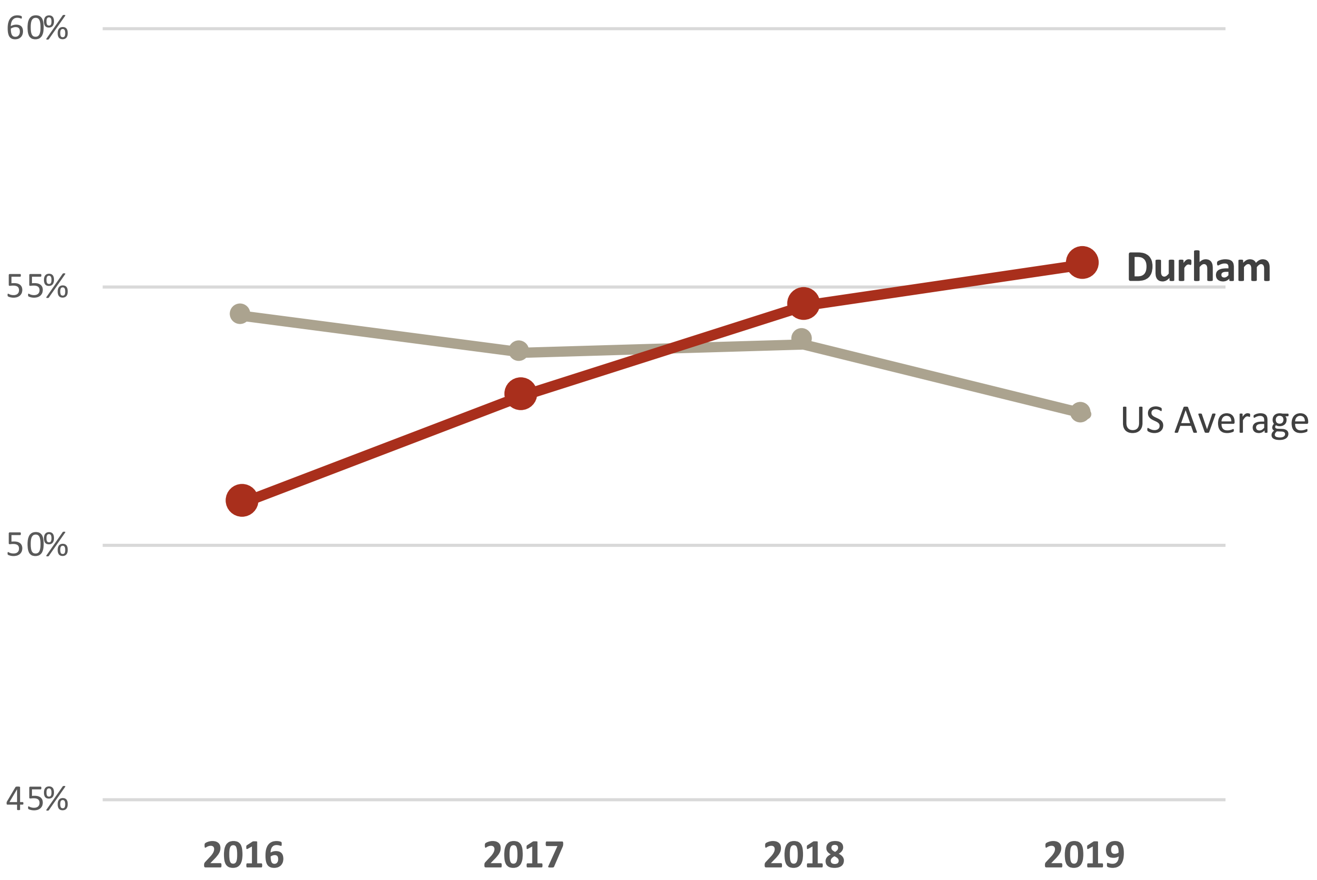
https://www.dconc.gov/Home/ShowDocument?id=30130

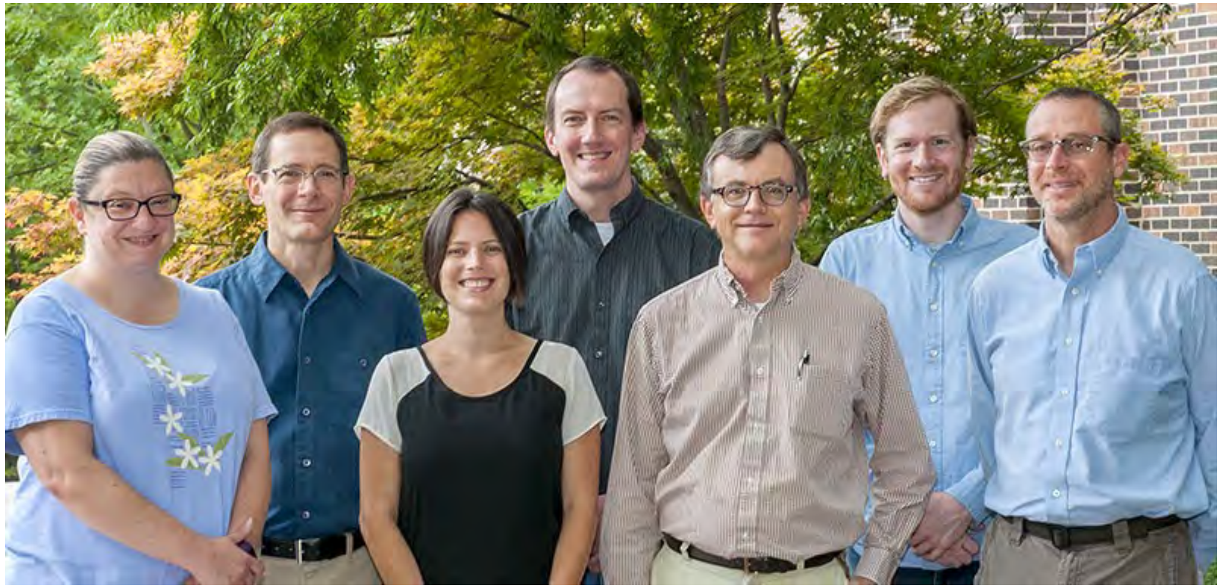

Average Durham satisfaction rating climbing over the US large city score!
US Average
Durham
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
2016
2017
2018
2019


Average Durham satisfaction rating
climbing over the US large city score!
Durham
US Average
60%
50%
40%
30%
20%
10%
0%
2016
2017
2018
2019


Average Durham satisfaction rating climbing
over the US large city score!
60%
55%
50%
45%
Durham
US Average
2016
2017
2018
2019

Average Durham satisfaction rating climbing over the US large city score!





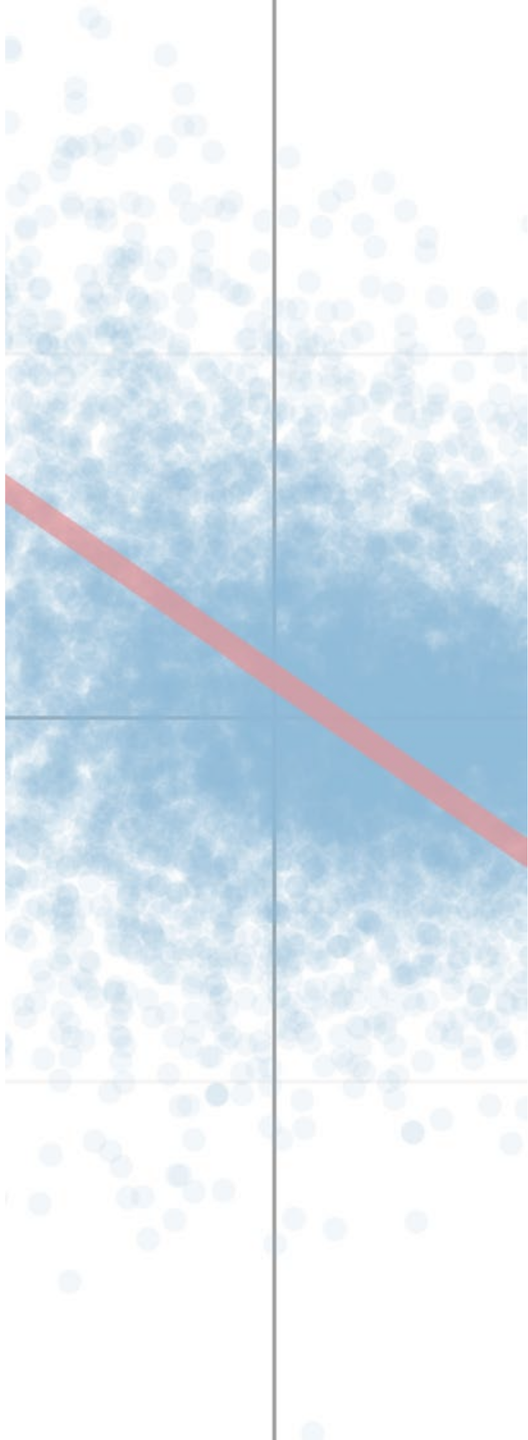
Center for Data and Visualization Sciences

- Data Sources
- Data Science
- Mapping and GIS
- Data Management
- Data Visualization

http://library.duke.edu/data
askdata@duke.edu

Types of visualization consulting

- Look at data and brainstorm about the best visualization
- Recommend appropriate tools
- Troubleshoot software problems
- Help with cleaning and structuring data
- Offer graphic design advice for figures, diagrams, slides and posters



Many free workshops every semester!

Duke University Libraries

Center for Data and Visualization Sciences

Spring 2023 Workshop Series

Registration
https://library.duke.edu/data/workshops

For online workshops, a Zoom link will be sent via email to registered participants to join the workshop.

Workshop	*Date*	*Time*	*Mode*
Tools for Data Management	Tue, Jan 17	1:00pm – 3:00pm	Online
Intro to ArcGIS Pro	Wed, Jan 18	10:00am – 12:00pm	Online
R for data science: getting started, EDA, data wrangling	Tue, Jan 24	10:00am – 12:00pm	Online
R for data science: visualization, pivot, join, regression	Thu, Jan 26	10:00am – 12:00pm	Online
R for data science: custom functions and iteration	Tue, Jan 31	10:00am – 11:30am	Online
Effective Data Visualization	Tue, Jan 31	5:00pm – 6:30pm	Online
Creating dashboards with R: flexdashboards and Shiny	Thu, Feb 02	10:00am – 12:00pm	In-Person
Designing Thematic Maps	Tue, Feb 07	10:30am – 12:00pm	Online
Prep for Data Publishing: Standards & Disciplinary Repositories	Tue, Feb 14	10:00am – 12:00pm	Online
Intro to QGIS	Wed, Feb 15	10:00am – 12:00pm	Online
Meeting Data Management Plan Requirements	Mon, Feb 20	1:00pm – 3:00pm	Online
Quarto: a first look	Thu, Feb 23	10:00am – 11:00am	Online
Geospatial Data in R: Mapping	Thu, Feb 23	1:00pm – 3:00pm	Online
Ethics of Data Management and Sharing	Thu, Mar 02	10:00am – 12:00pm	Online
Make a horizontal dot (forest) plot in Excel	Fri, Mar 03	10:00am – 11:00am	Online
Open Scholarship: Practices and Principles	Wed, Mar 22	1:00pm – 3:00pm	Online
Effective Academic Posters	Tue, Mar 28	5:00pm – 6:30pm	Online
Python for Data Science: Pandas 103 – groupby & aggregation	Thu, Apr 06	10:00am – 12:00pm	Online
An Introduction to Reproducible Research Practices	Wed, Apr 19	10:00am – 12:00pm	Online

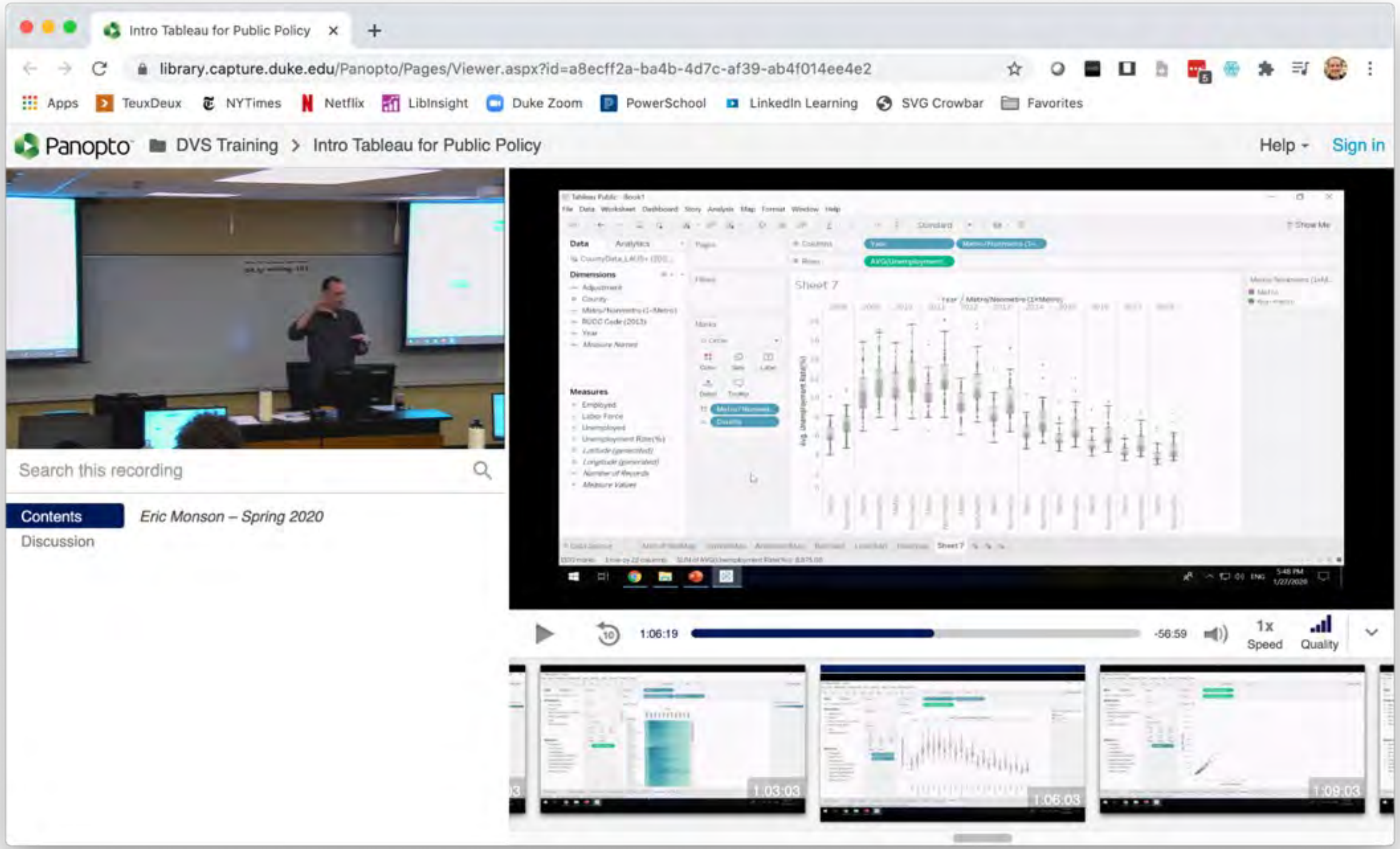
Asynchronous online learning
https://library.duke.edu/data/tutorials

Questions
askdata@duke.edu

https://library.duke.edu/data/workshops

Videos of past CDVS workshops

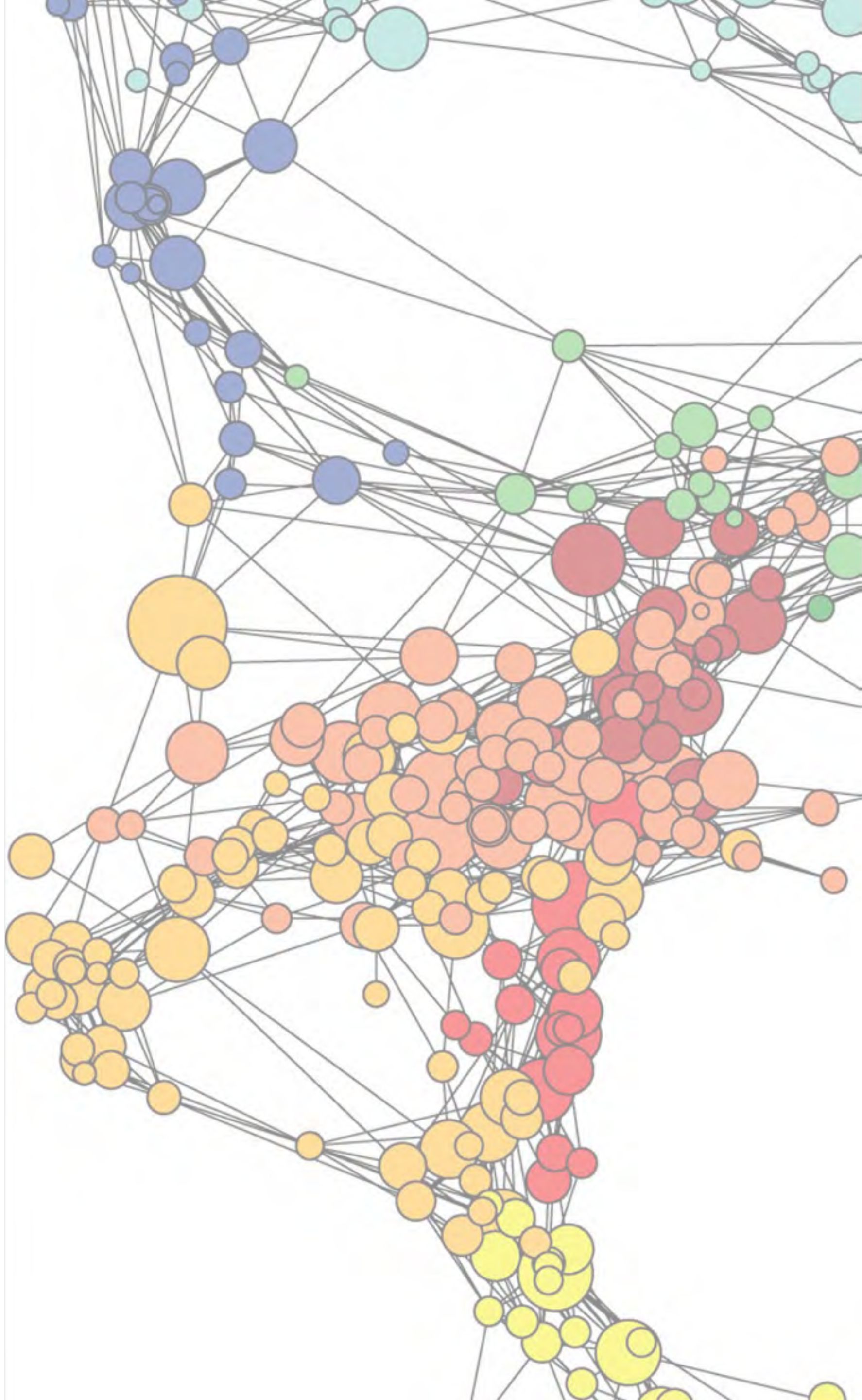
Online Learning: https://library.duke.edu/data/tutorials



Questions

askdata@duke.edu

Slides: https://bit.ly/HIVQuantEffVisMay2023



Extra slides

Encoding Choices

Magnitude *(numerical)*

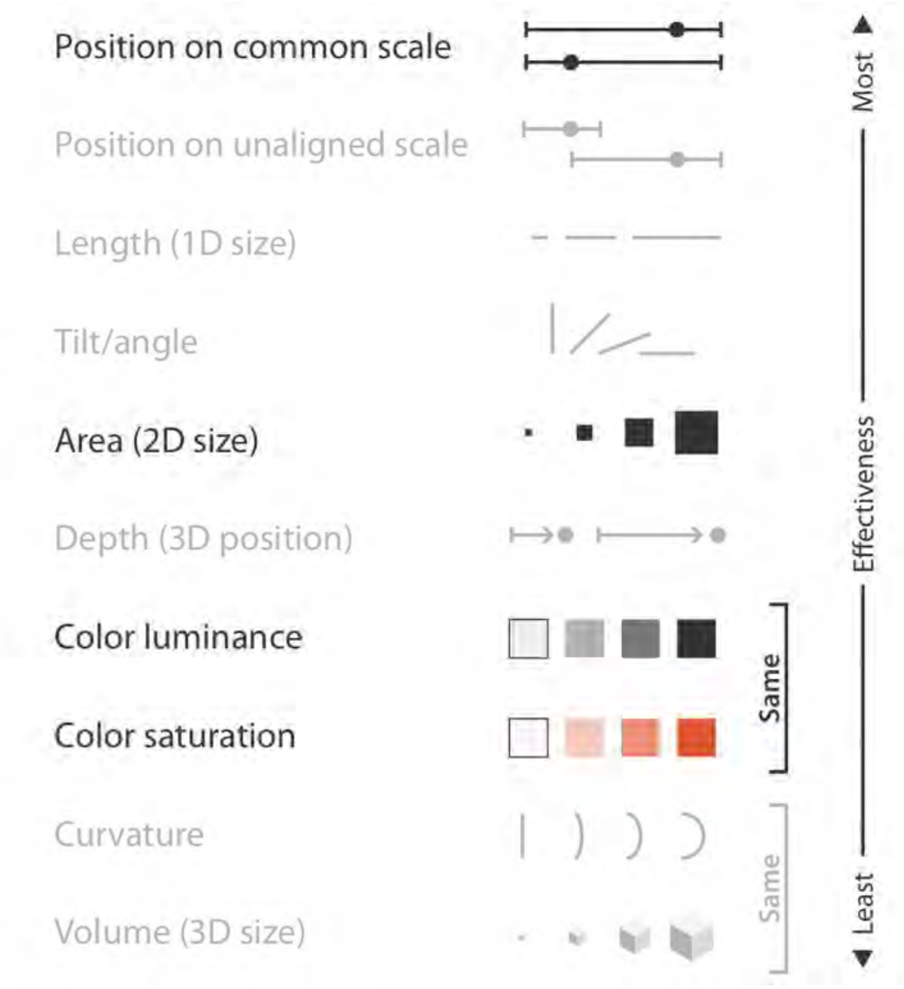
Identity *(categorical)*

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

Encoding Choices

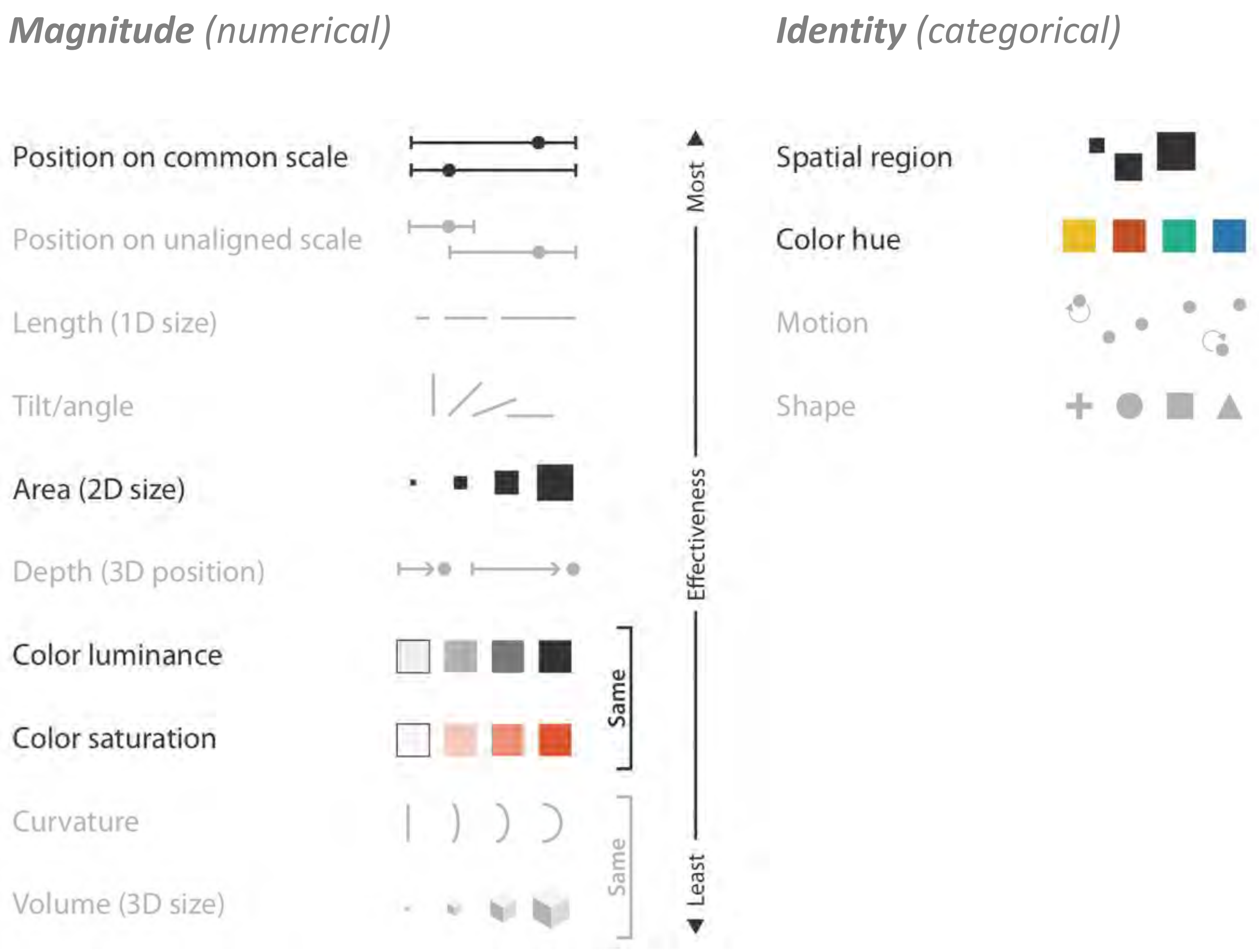
***Magnitude** (numerical)*

***Identity** (categorical)*



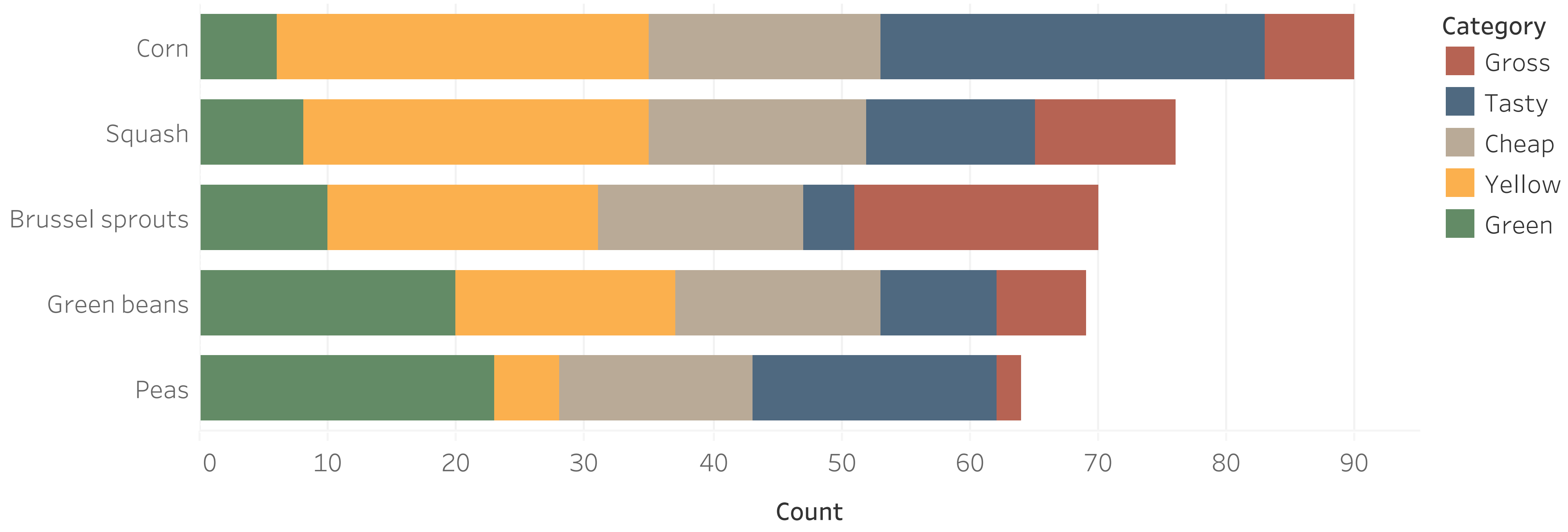
Tamara Munzner: https://www.cs.ubc.ca/~tmm/vadbook/eamonn-figs/fig5.1.pdf

Encoding Choices



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

Stacked bars



Pro:

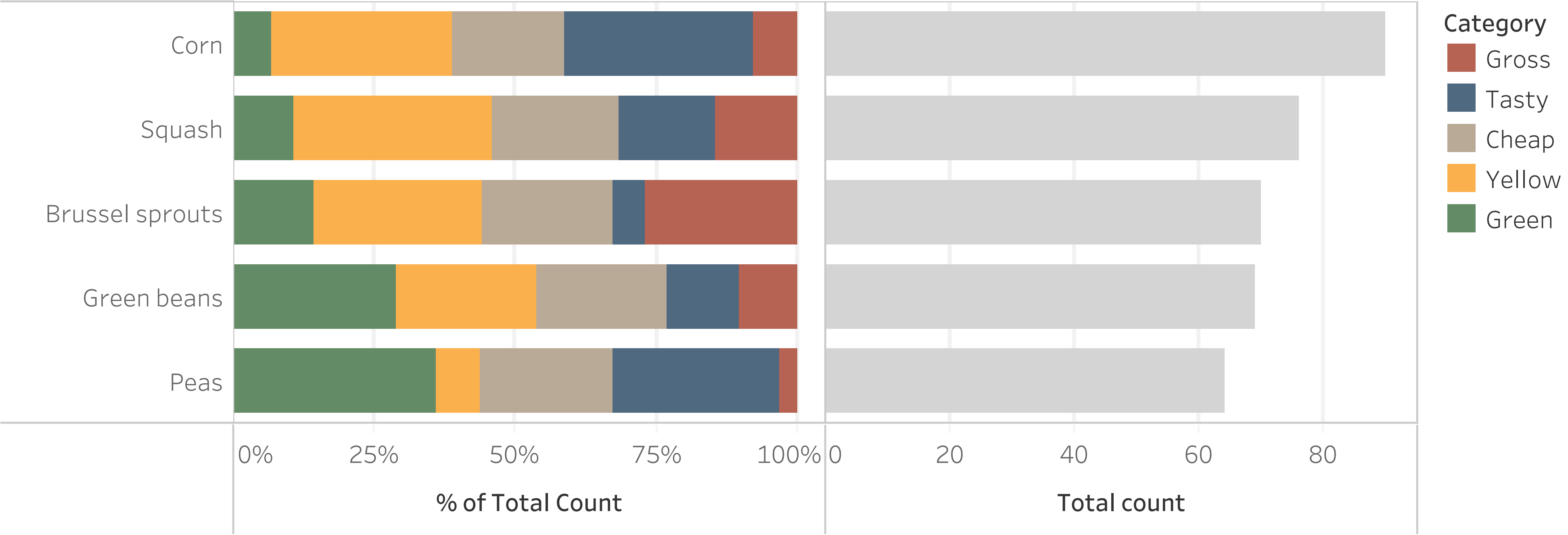
- **Great if totals are most important**

Con:

- Floating bars (no common baseline) are hard to compare

Inspired by: https://www.nature.com/articles/nmeth.2807

100% stacked bars with totals

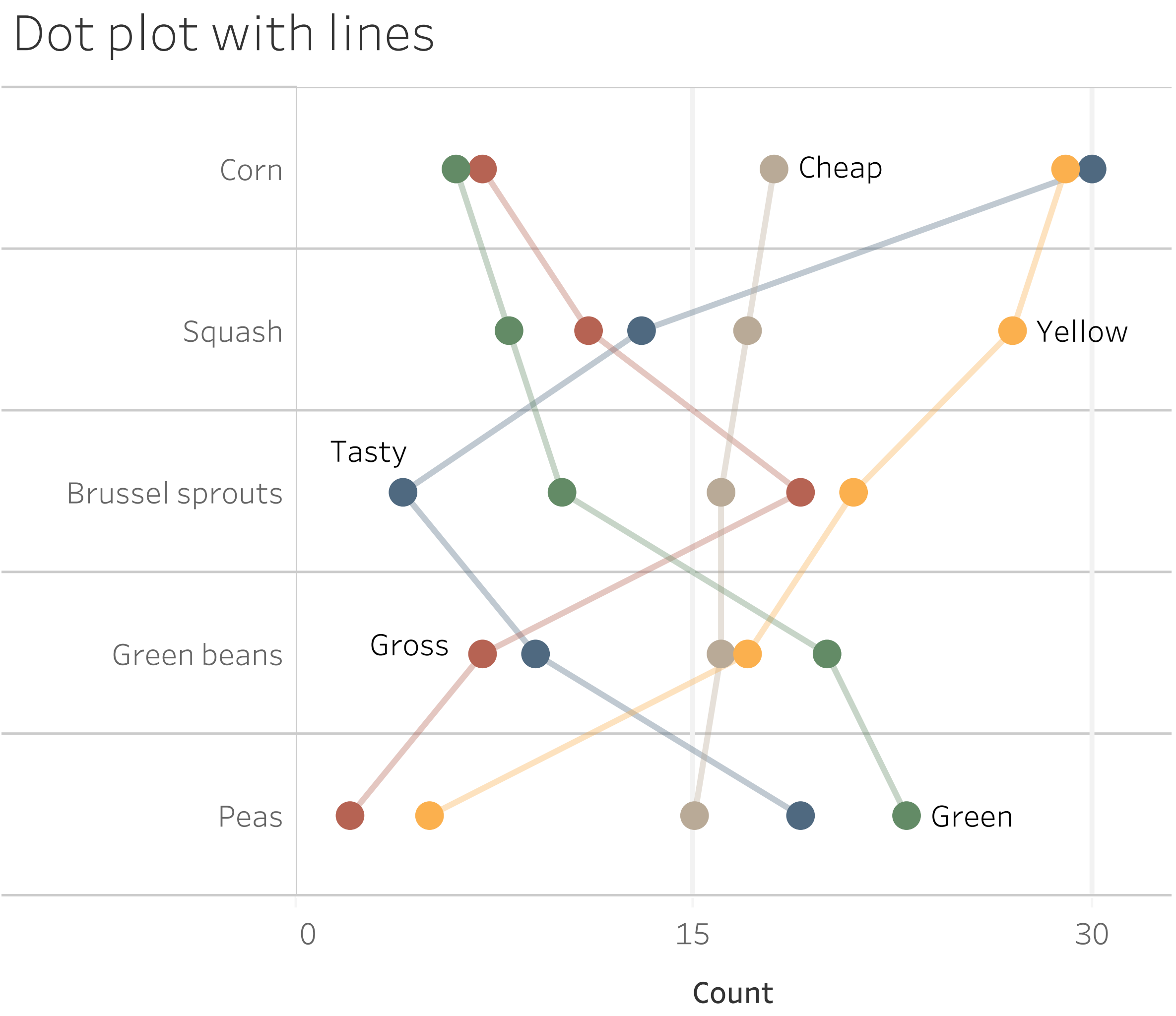


Pro:

- **Compact alternative to pies**
- Works well for survey data

Con:

- Floating bars (no common baseline) are hard to compare
- Often need separate totals bars

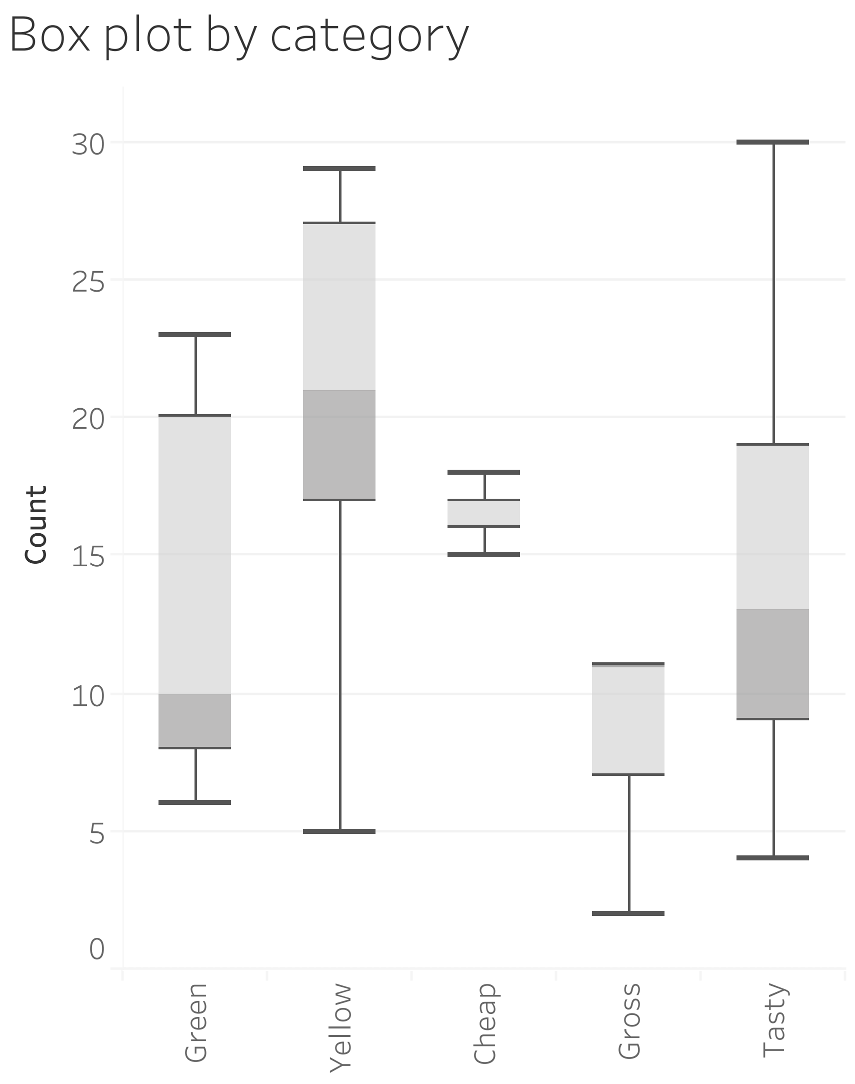


Pro:

- Easier to follow with eyes
- Can directly label lines

Con:

- Problematic to connect categories with lines (people sometimes make strange interpretations)



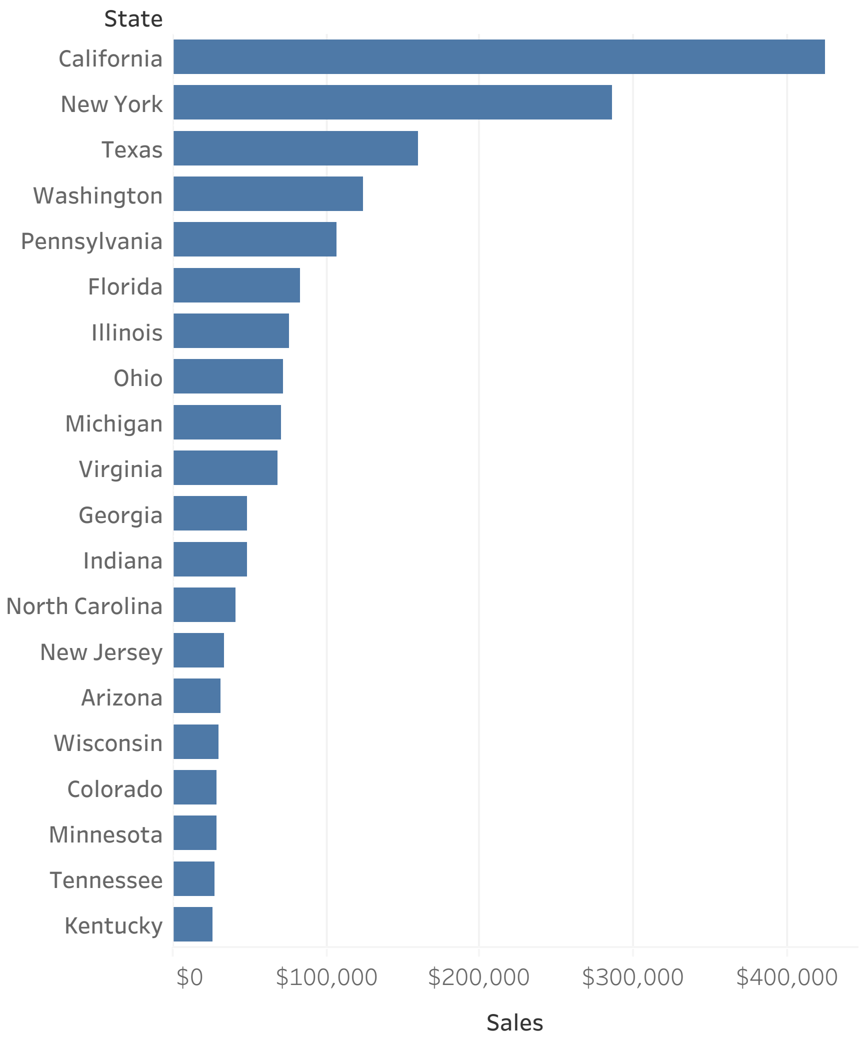
Pro:

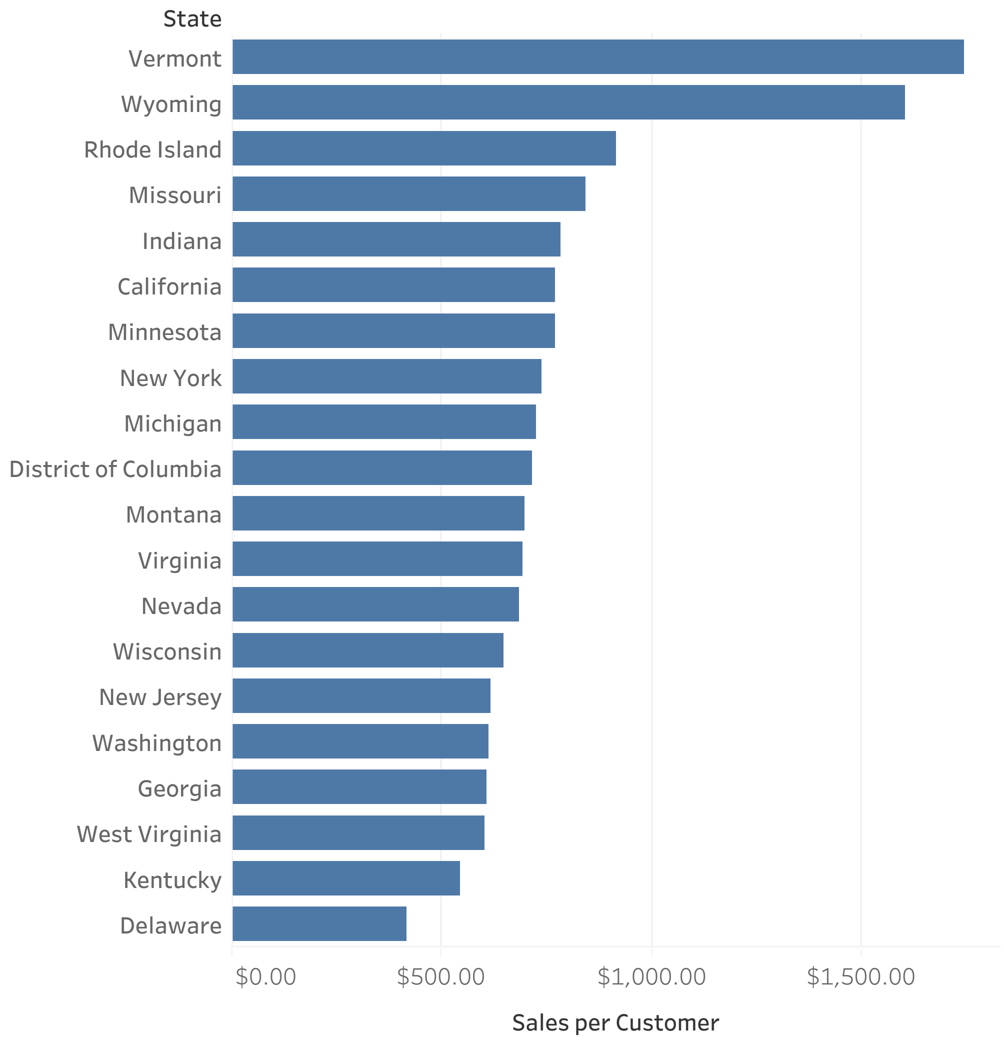
- **Simpler summaries of distributions can make comparisons easier**
- Great for large number of points

Con:

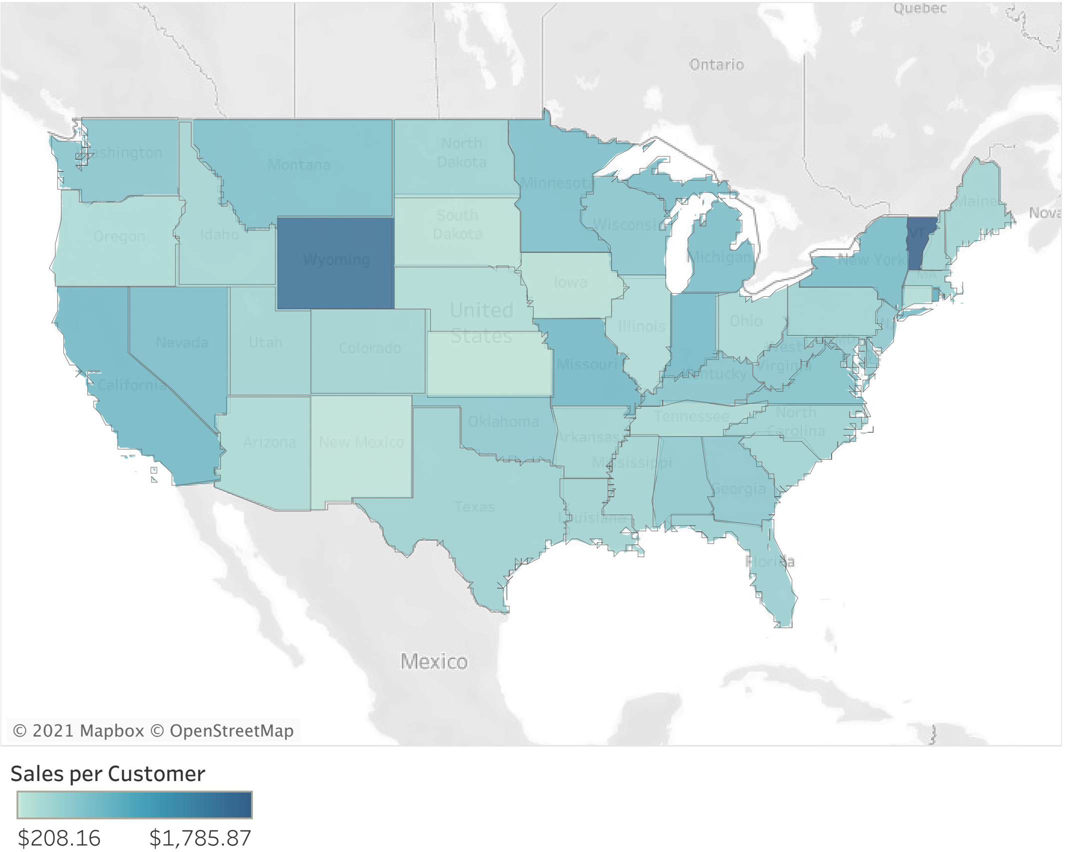
- **Summaries hide number of points and subtleties of distribution**
- Bad for small number of points

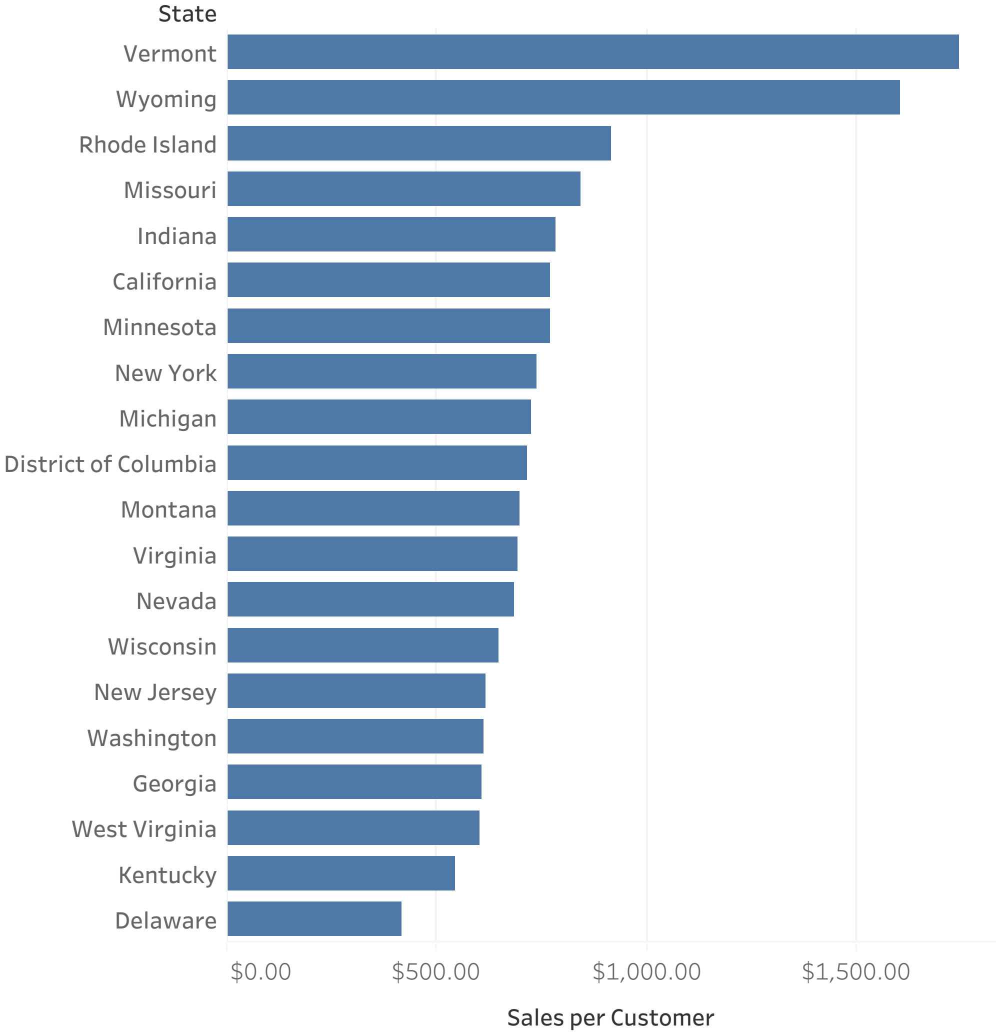
Some patterns are just population!



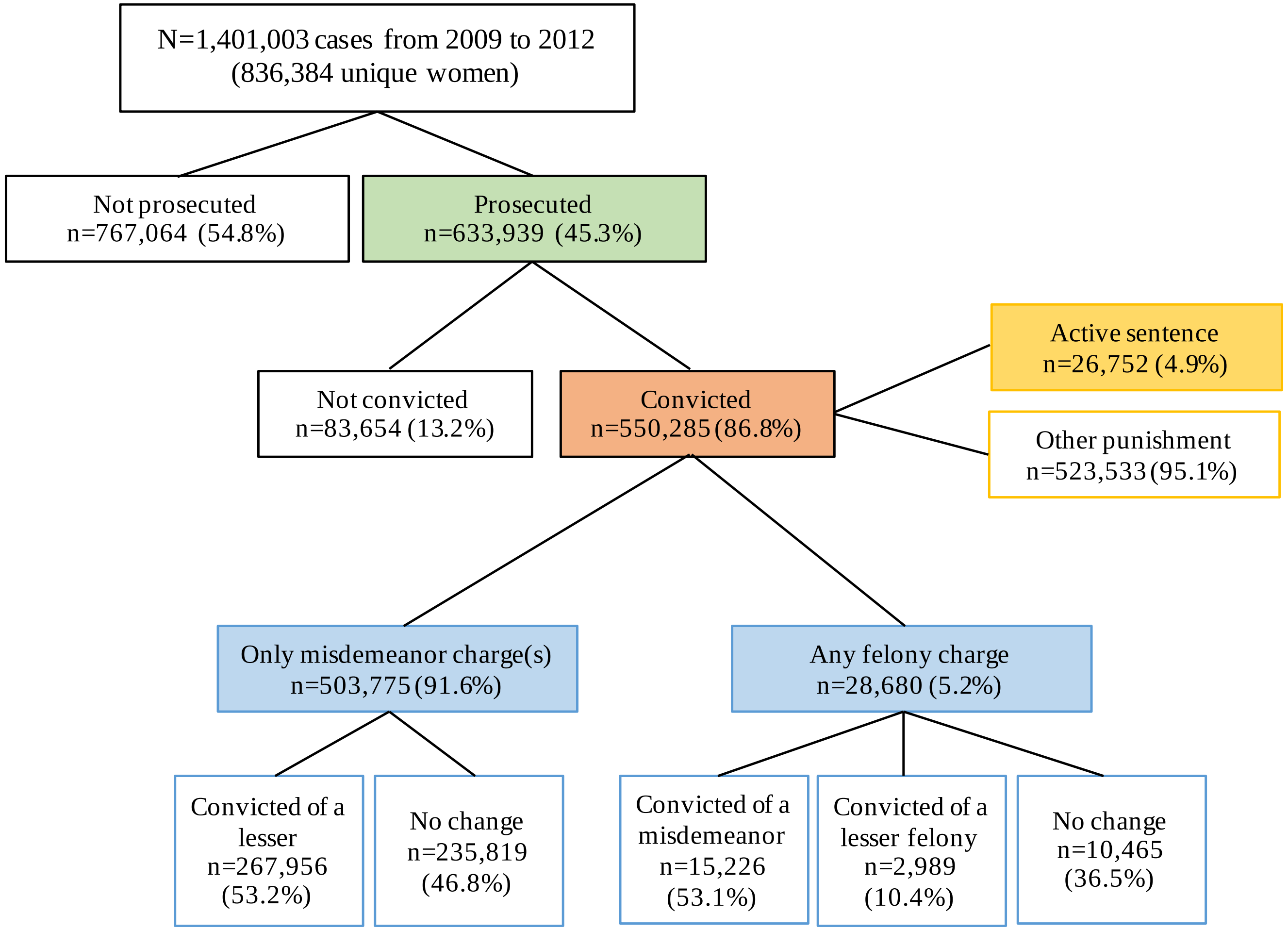


Maps are not always best for geo data



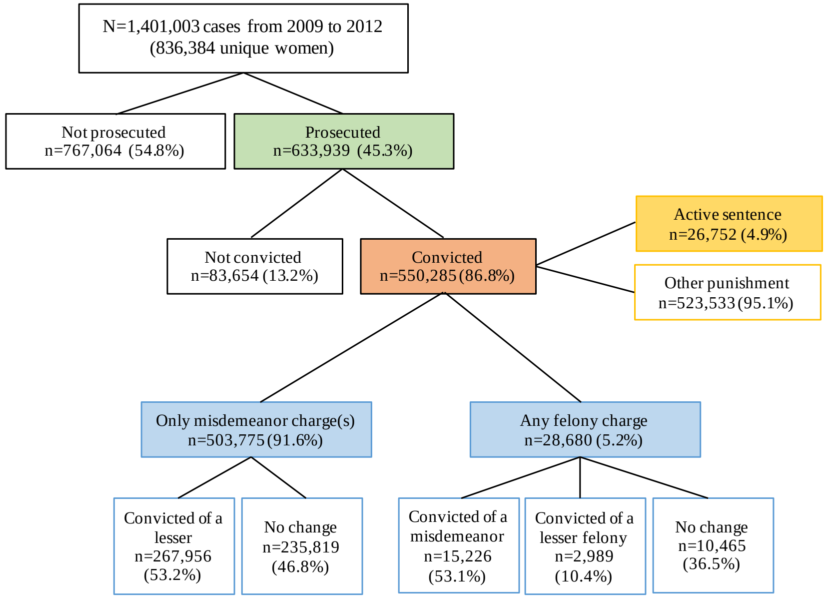


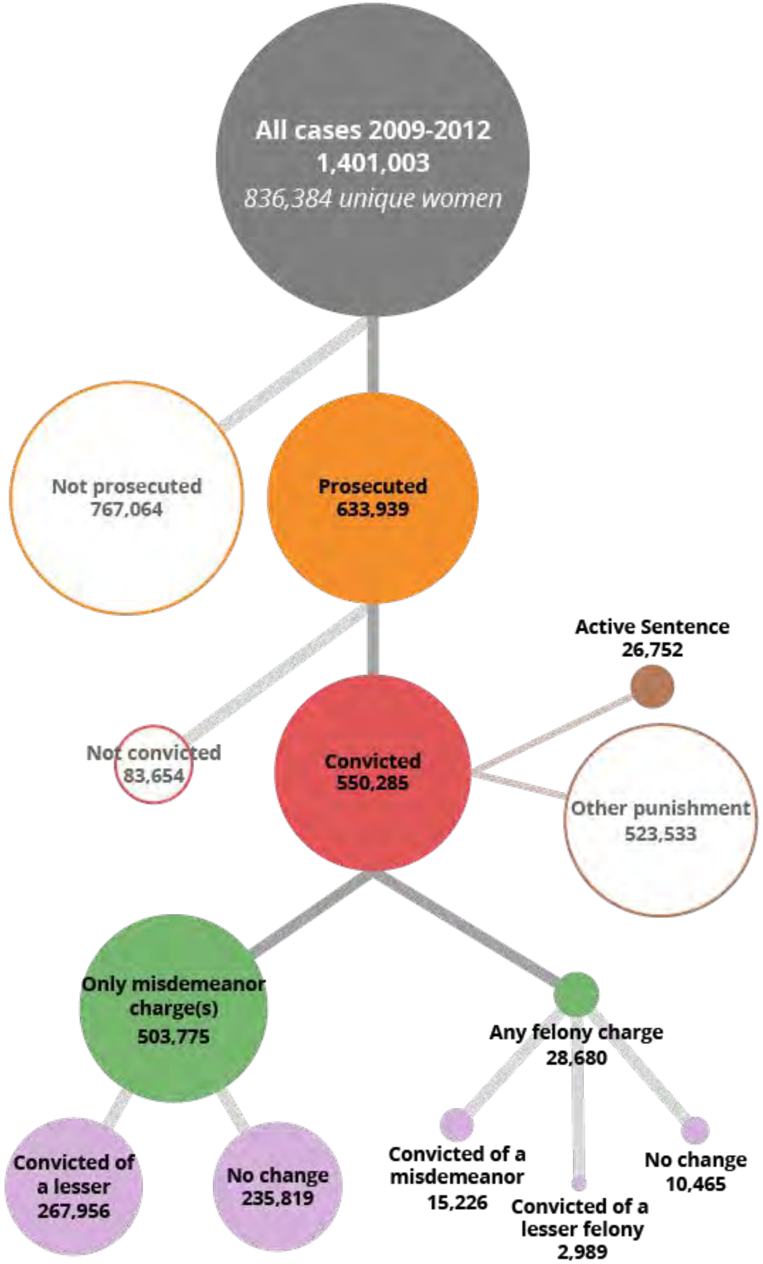
Numbers just written out hides patterns



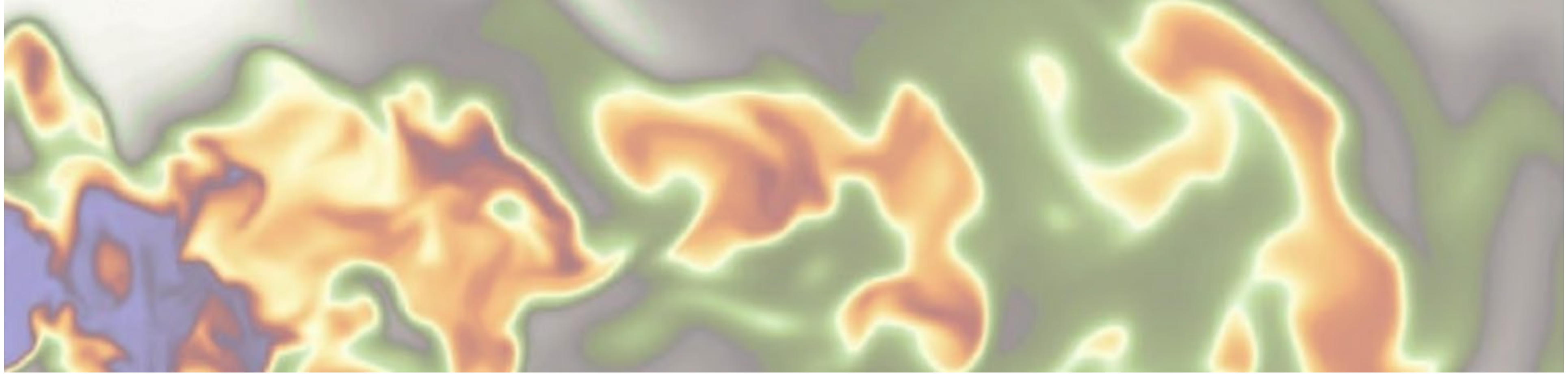
Beth Gifford

Try to make numbers visual





Beth Gifford



Another, more recent article on the problems with a rainbow colormap:

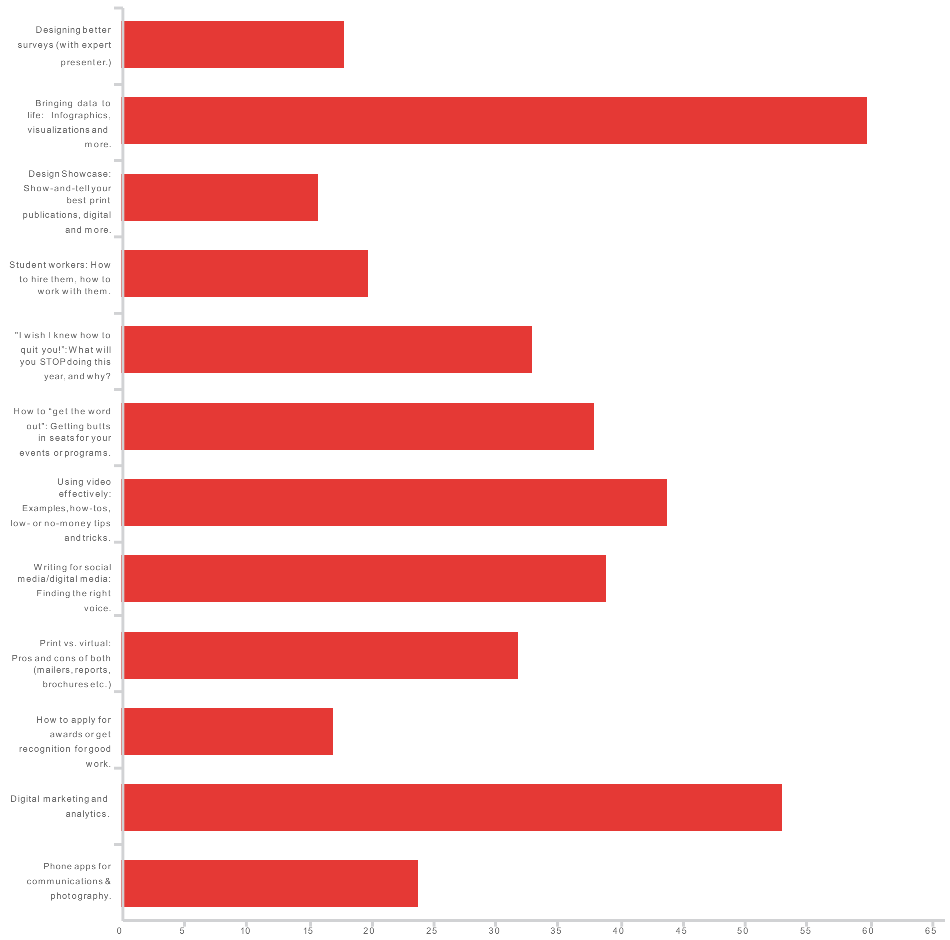
- The misuse of colour in science communication – 2020, Fabio Crameri, Grace Shephard & Philip Heron

And I love Francesca Samsel's work on better colormaps:

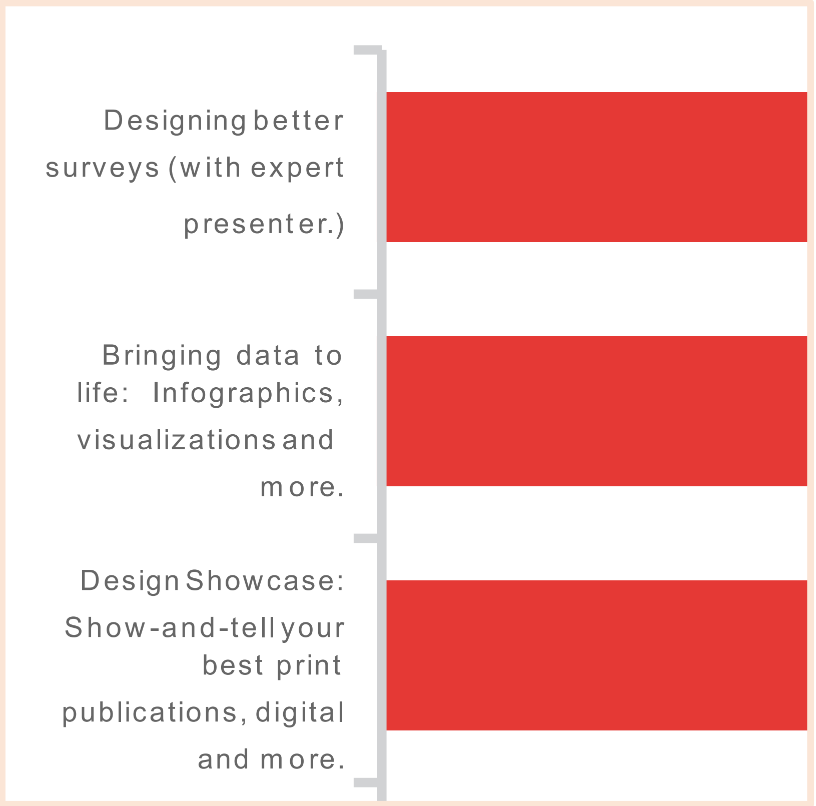
- Visualizing Science: How Color Determines What We See 2 – 2020, Stephanie Zeller & David Rogers
- ColorMoves: Real-time Interactive Colormap Construction for Scientific Visualization – 2018, with Sebastion Klaassen & David Rogers
- Colormaps Constructed with an Artist in the Loop 1 – 2015, with Utkarsh Ayachit

Default sizes may not be legible

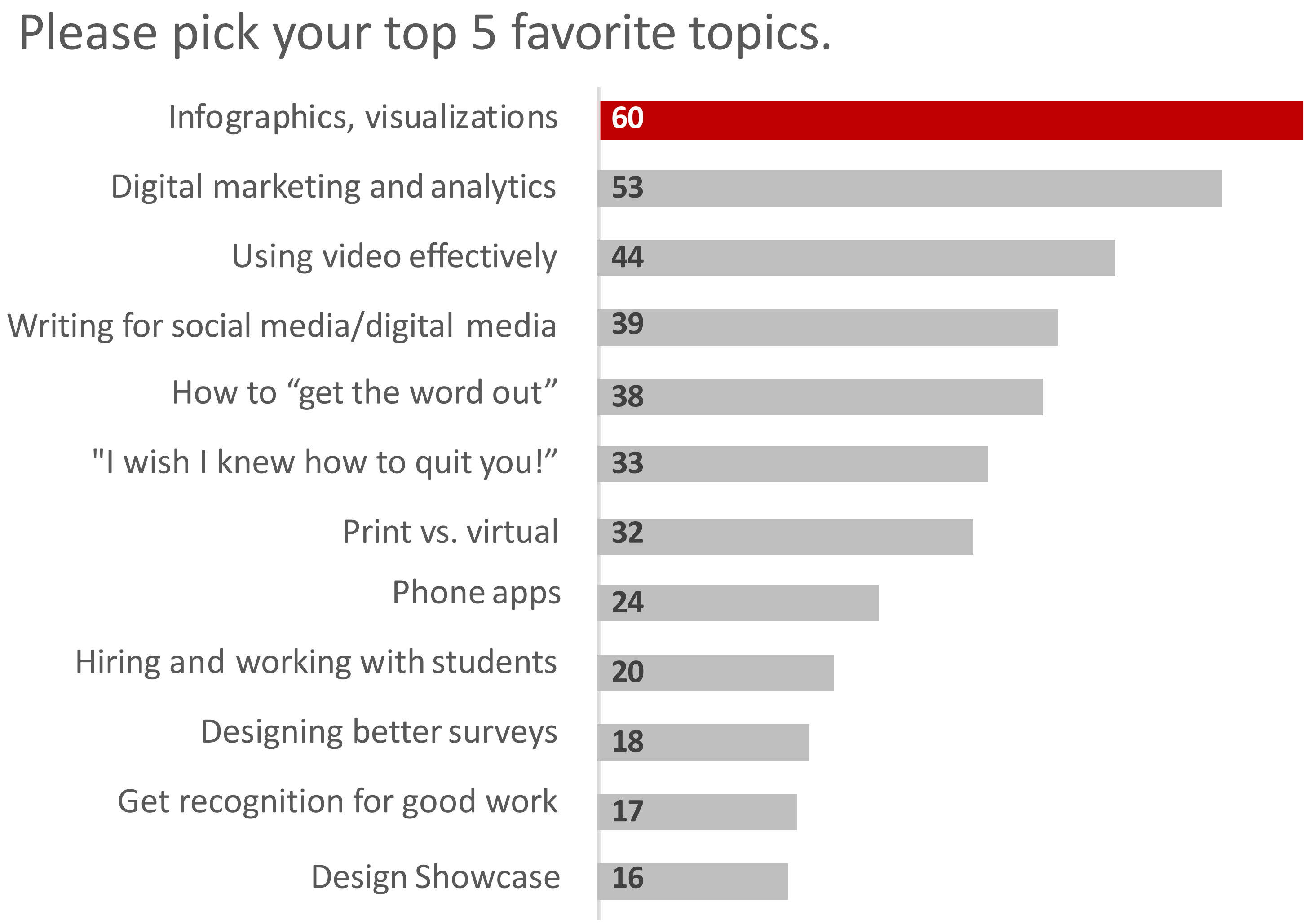
Q2 - Please pick your top 5 favorite topics.



Default Qualtrics output



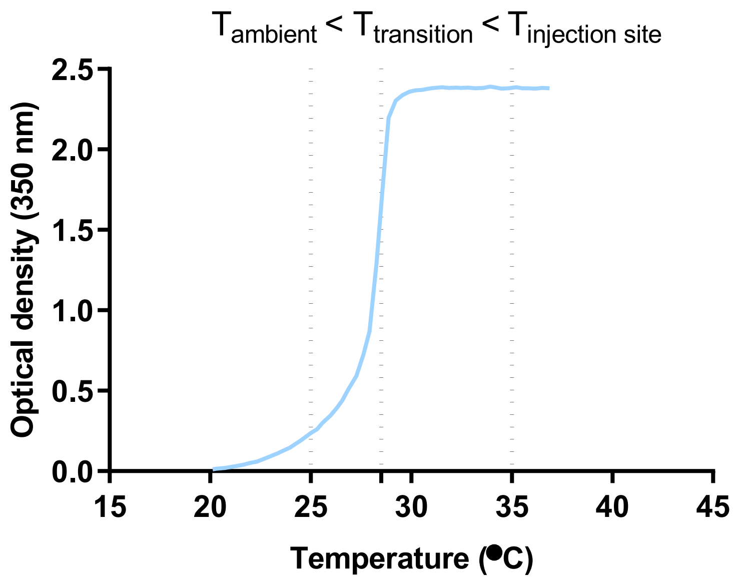
Summarize, sort & highlight

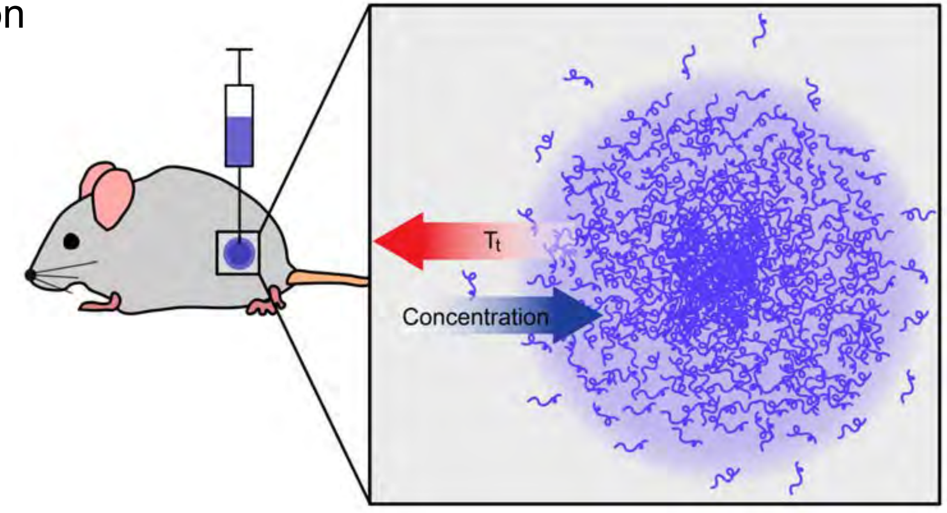


Replace text with visuals

Depot formation

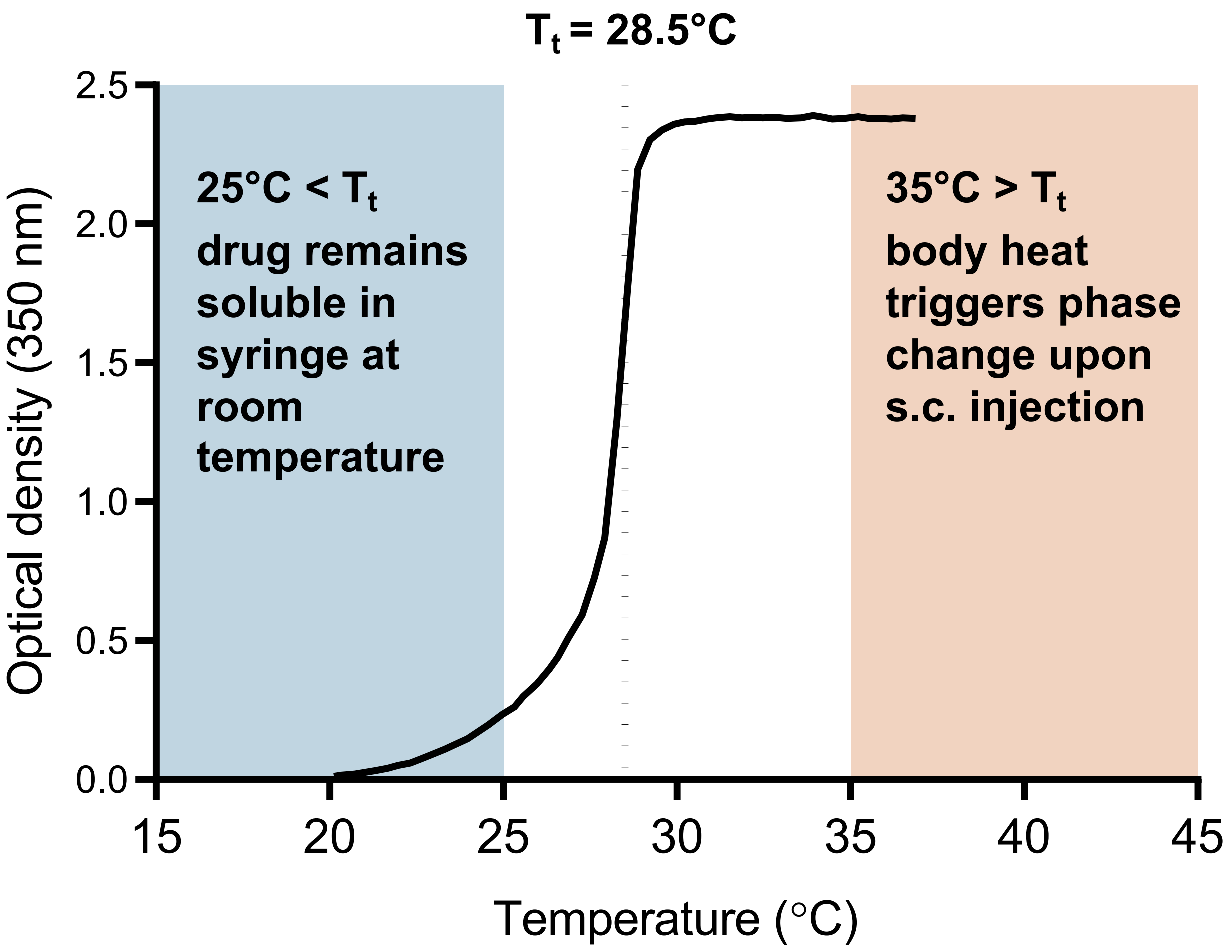
- GLP1-ELP-FGF21 designed to form an *in vivo* drug depot
  - GLP1-ELP and ELP-FGF21 previously optimized as depot-forming single agonist treatments [2,3]
  - 25°C < T_t → drug remains soluble in syringe at room temperature
  - 35°C > T_t → body heat triggers phase change upon s.c. injection
  - T_t identified by monitoring ELP solution turbidity during temperature ramping
- ELP T_t inversely dependent on concentration
  - Core of depot represents injection concentration
  - Depot boundary slowly hydrated
  - Concentration decreases → T_t increases
  - When T_t increases above 35°C, fusion unimers resolubilize and leave depot

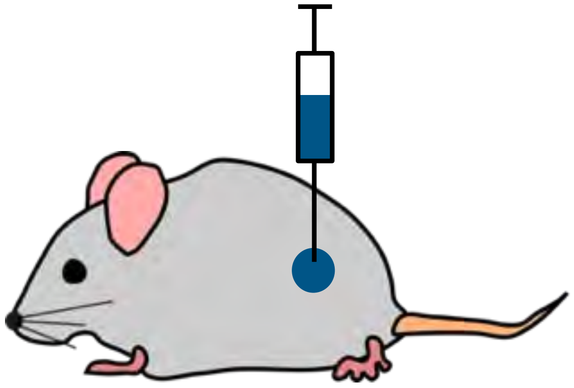




Caslin Gilroy

Dual agonist designed to form an *in vivo* drug depot

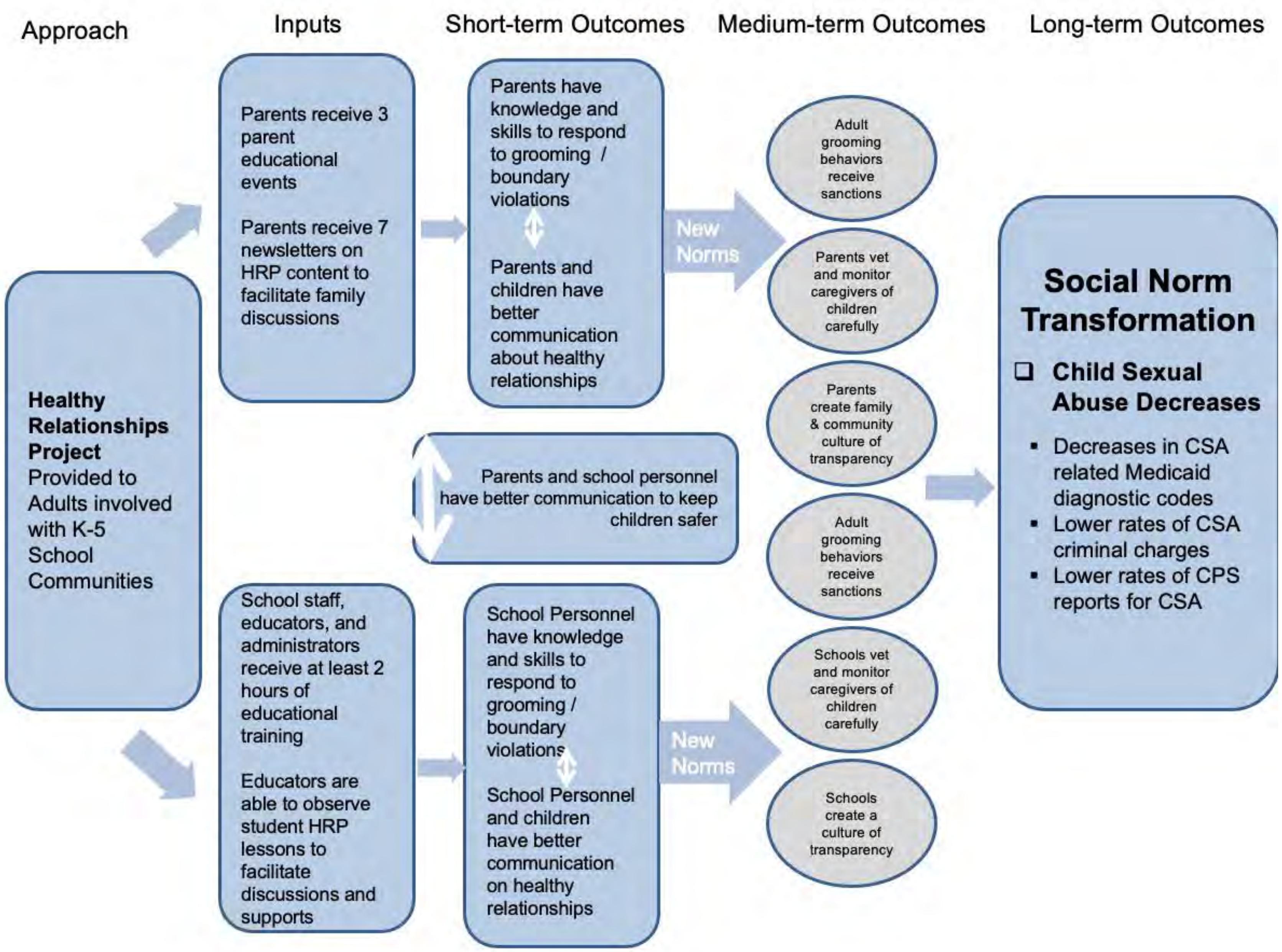




Depot slowly dissolves as it gets hydrated

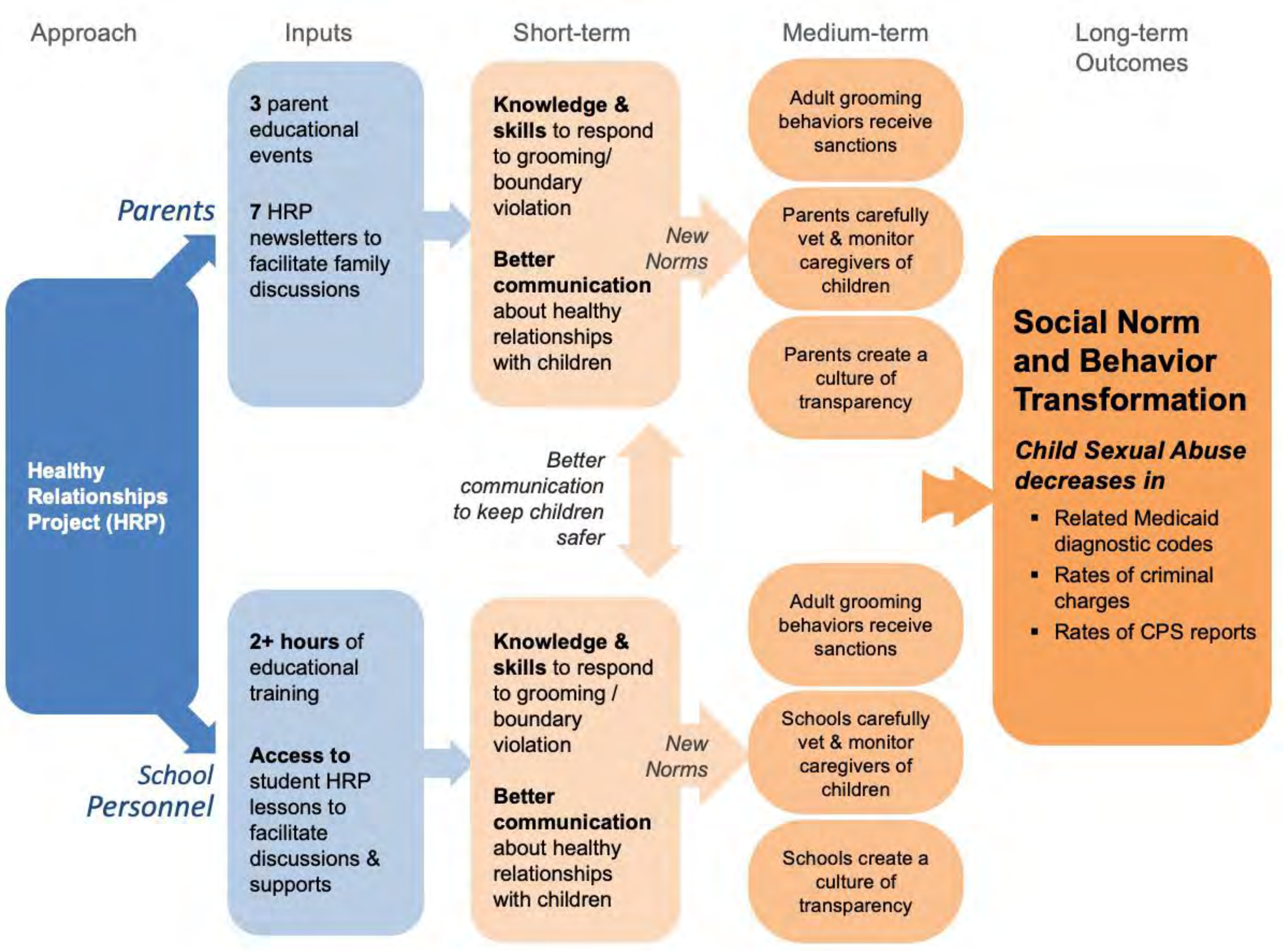
Caslin Gilroy

Remove distractors & add hierarchy



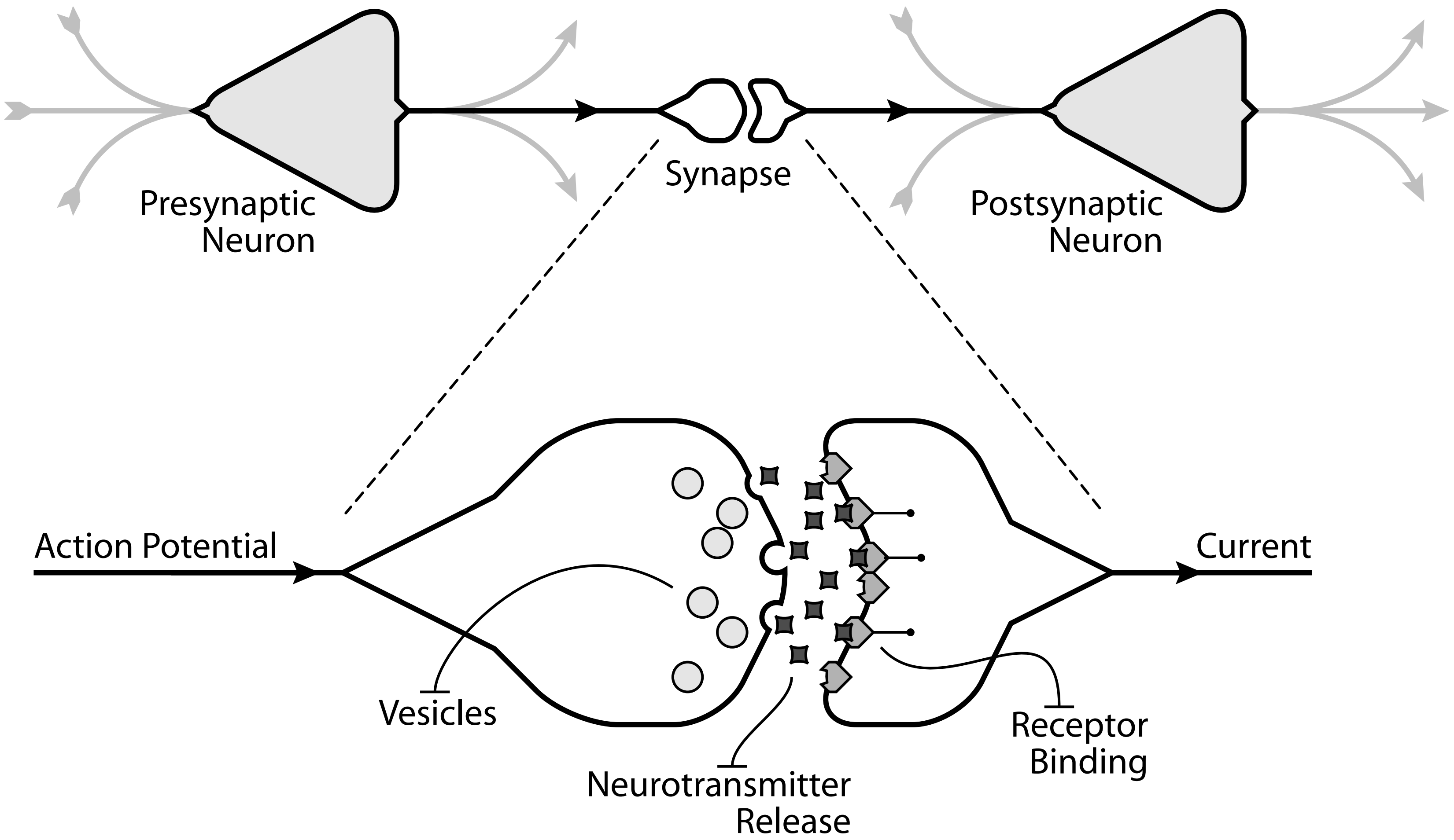
YuerongLiu

Remove distractors & add hierarchy

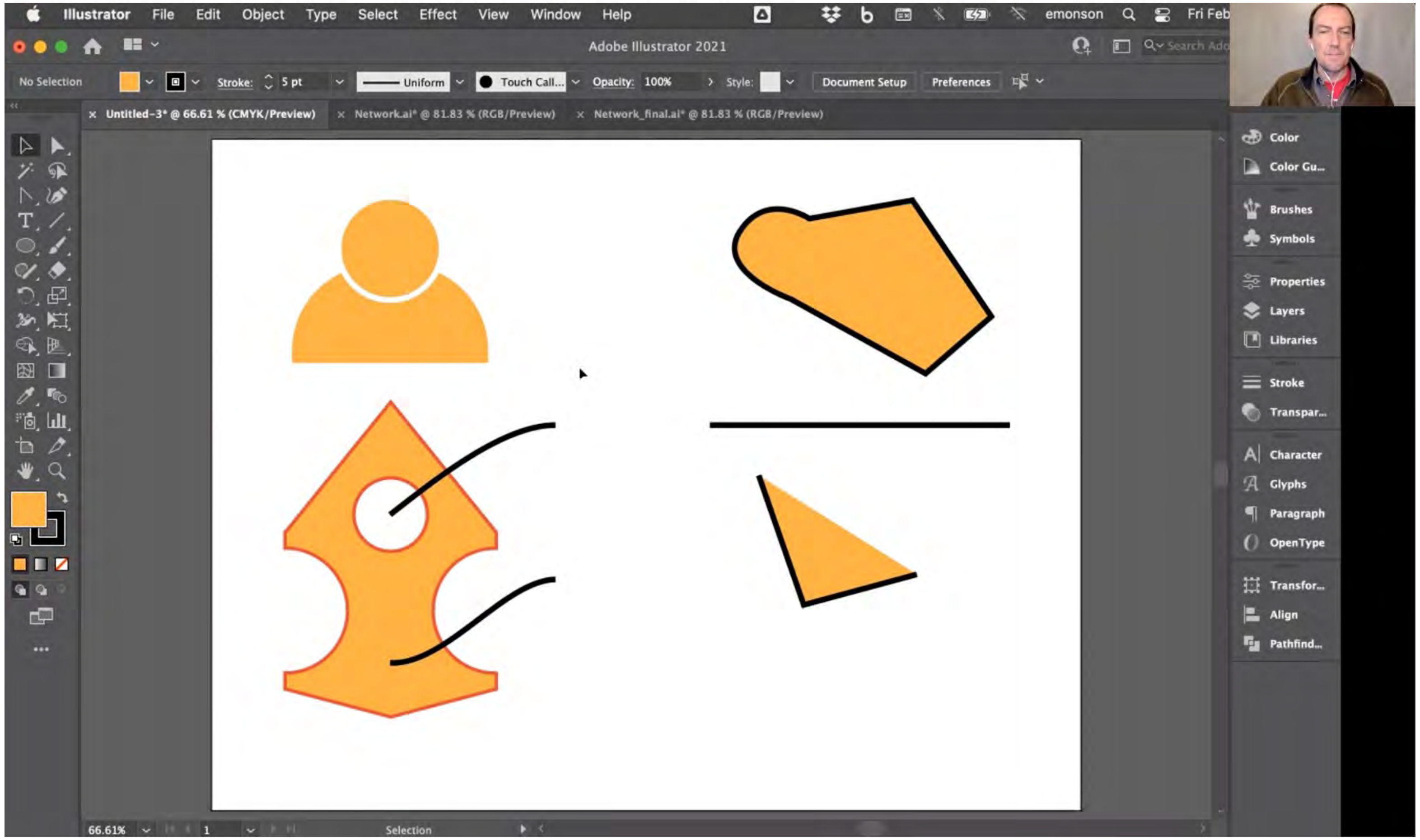


YuerongLiu

Adobe Illustrator for figures

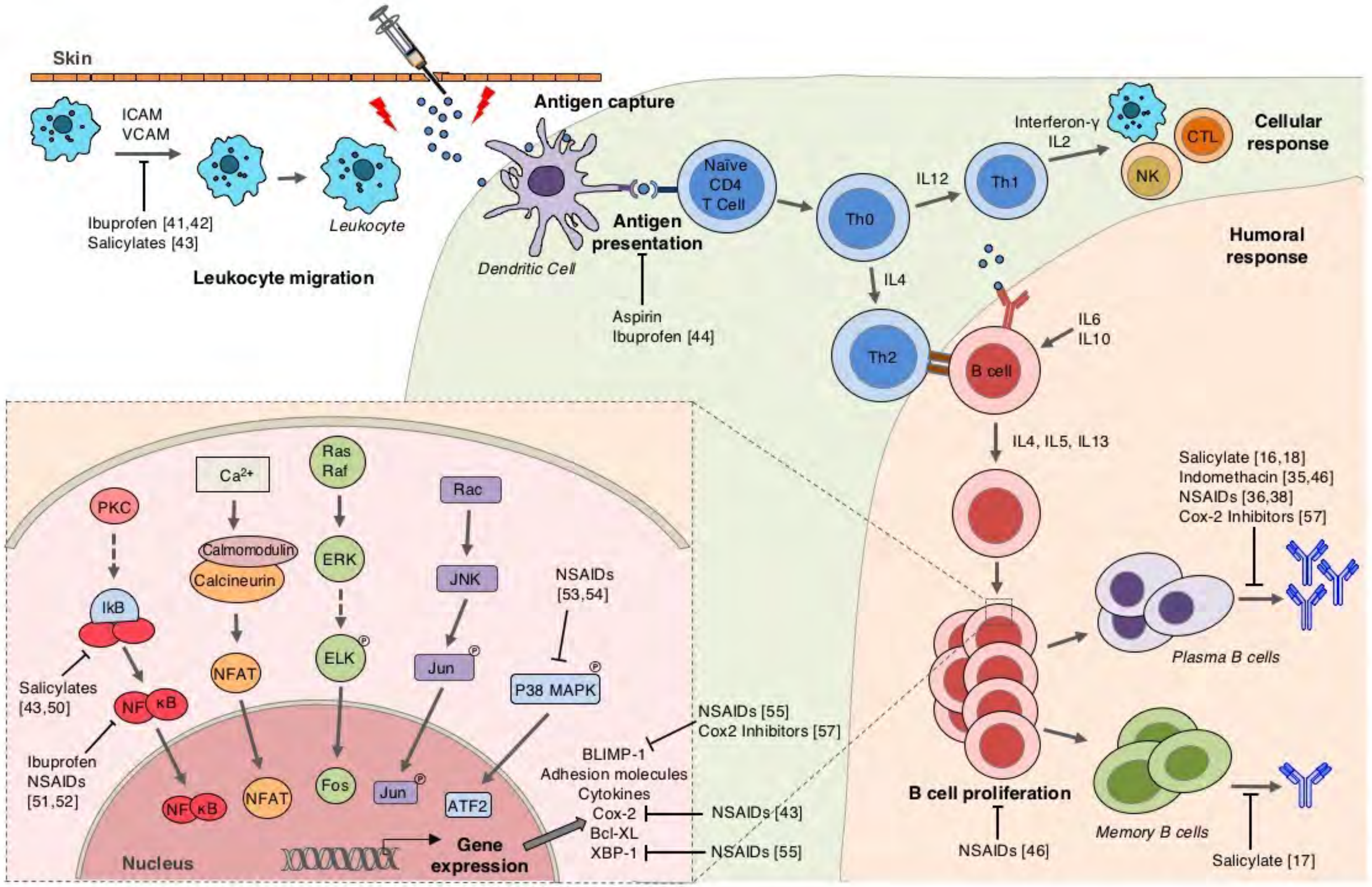


Adobe Illustrator for Diagrams



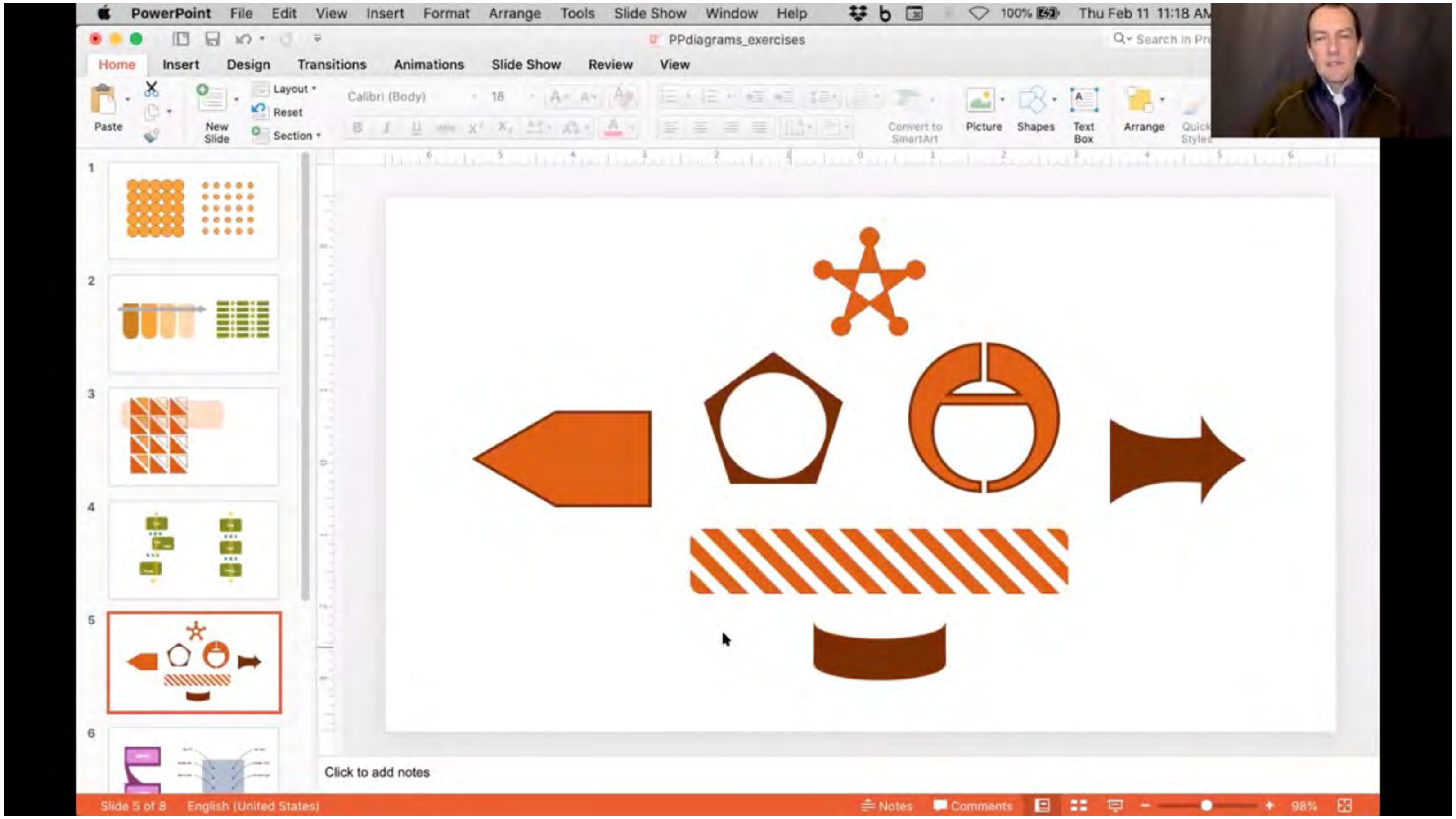
https://warpwire.duke.edu/w/_bIGAA/

PowerPoint for figures



Ezzeldin Saleh

PowerPoint Skills for Diagrams



https://warpwire.duke.edu/w/s0sFAA/

Brandaleone Family Lab for Data and Visualization Services

http://library.duke.edu/data/about/lab

See our website for remote access options.

- **The Edge** *(1st floor of Bostock Library, West Campus)*
- Open whenever the library is open
- 12 high-powered Dell workstations
- 3 Bloomberg financial workstations
- Software for data analysis, GIS, and visualization

Data and Visualization Lab

Entrance to Bostock