Information Visualization I

UM Big Data Summer Institute 2019

Matthew Kay
Assistant Professor
School of Information & Computer Science and Engineering
University of Michigan

A little bit about me

Master's and Bachelor's in CS (Fine Art minor) from the University of Waterloo

PhD in CSE from the University of Washington

My work draws upon human–computer interaction, visualization, design, and statistics

What I would like to do today

Motivate why visualization is important.

Give you grounding to help you systematically create effective visualizations:

Perceptual principles

Design principles

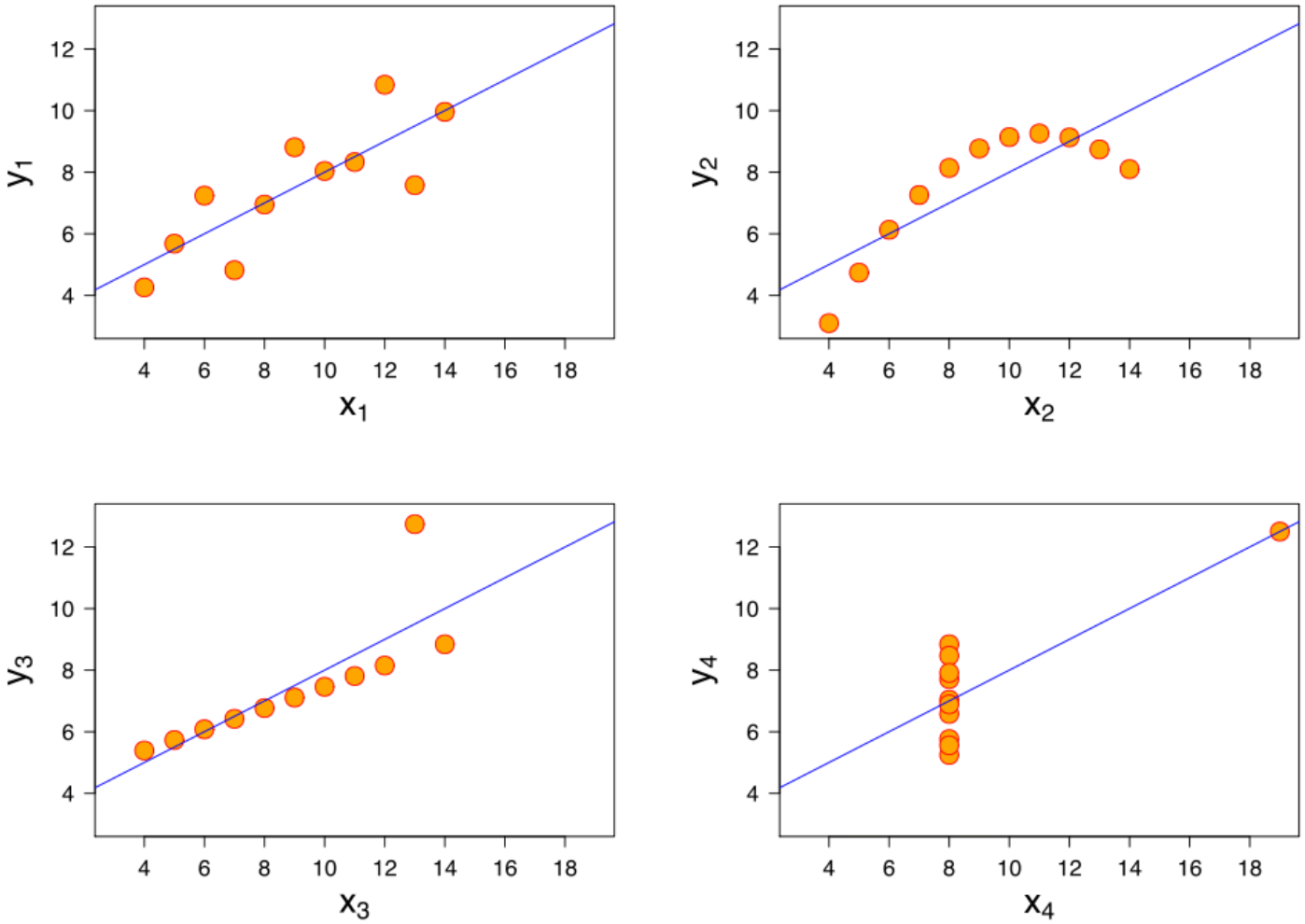
Why visualize?

Anscombe's quartet

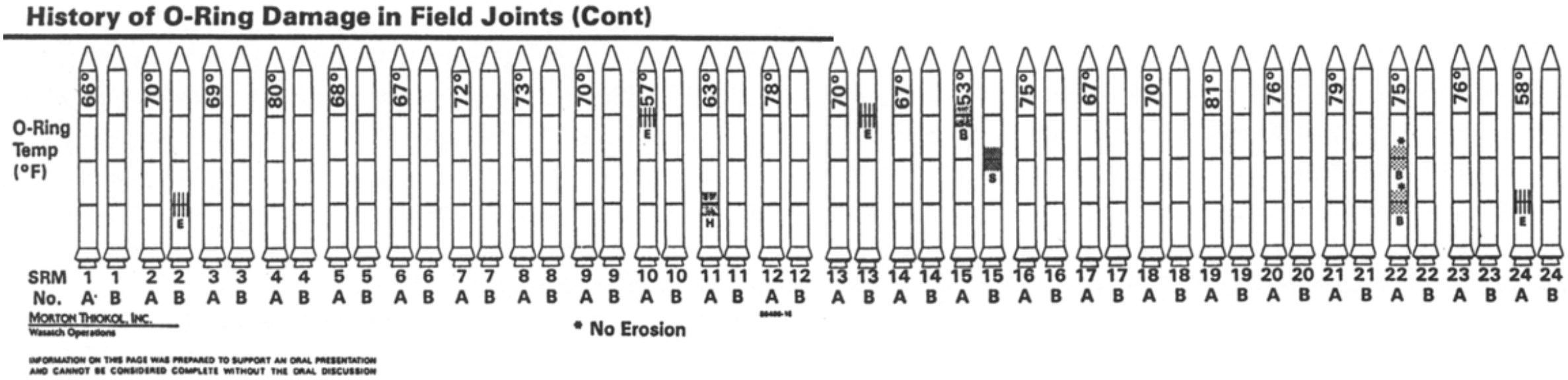
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0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

4 datasets, same means, variances, correlation

Anscombe's quartet

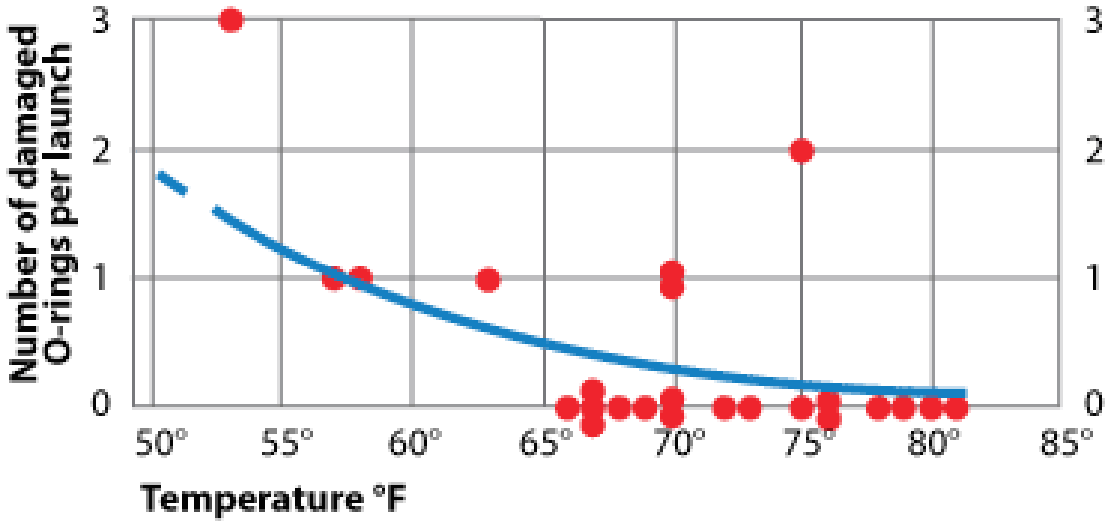


O-ring failure in the Challenger



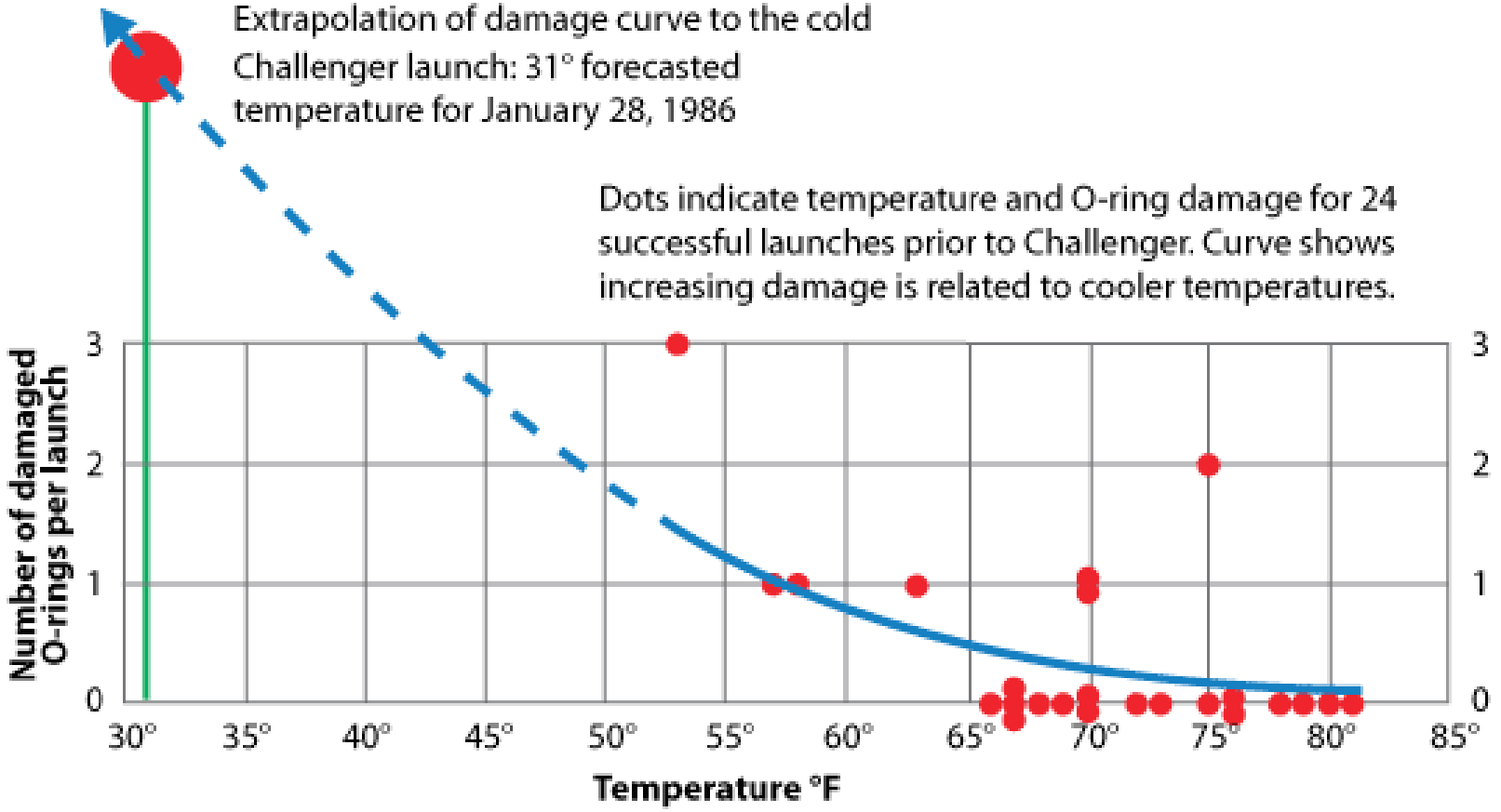
O-ring failure, Morton-Thiokol

O-ring failure in the Challenger



O-ring failure, Tufte

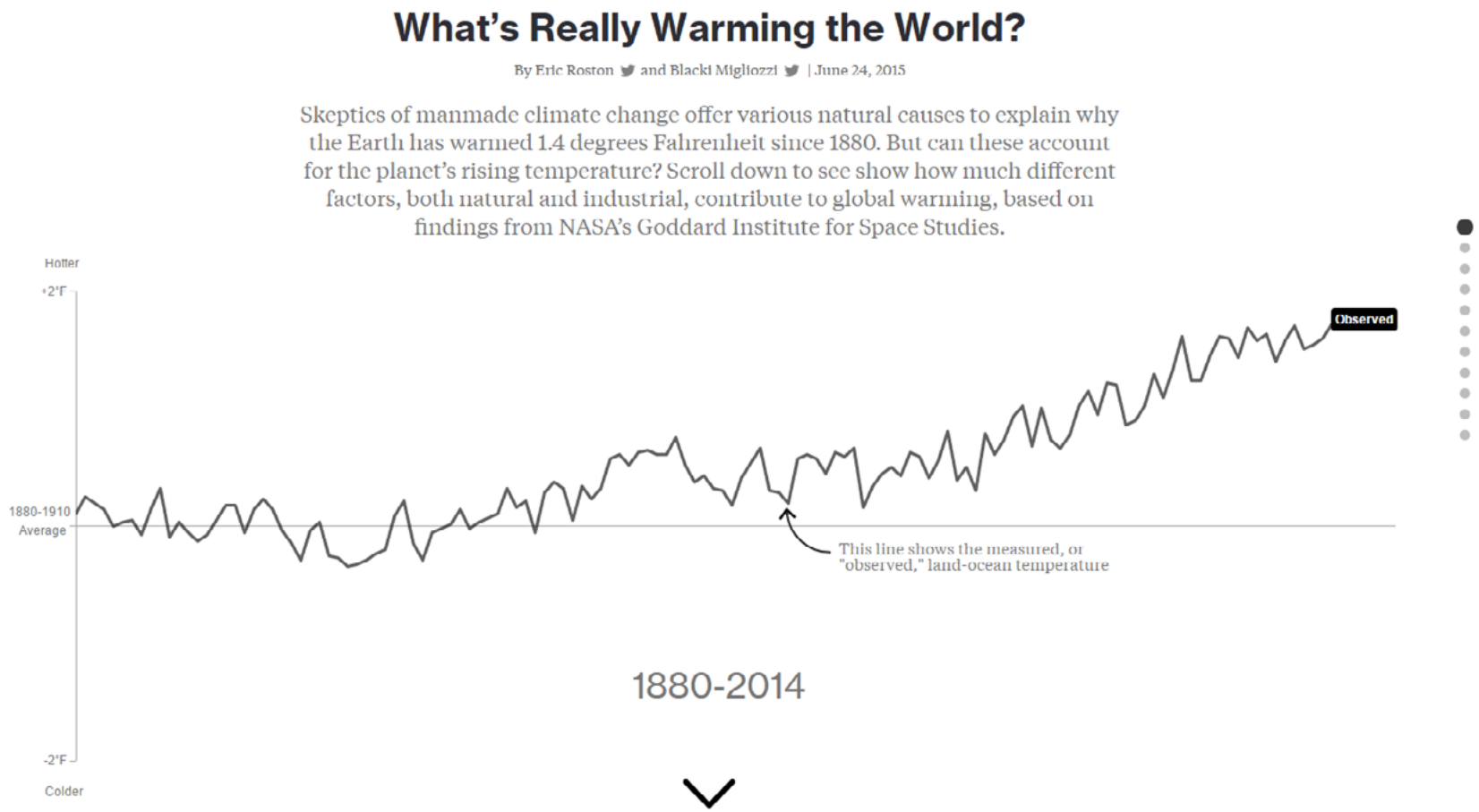
O-ring failure in the Challenger



O-ring failure, Tufte

Visualize to see patterns you wouldn't otherwise

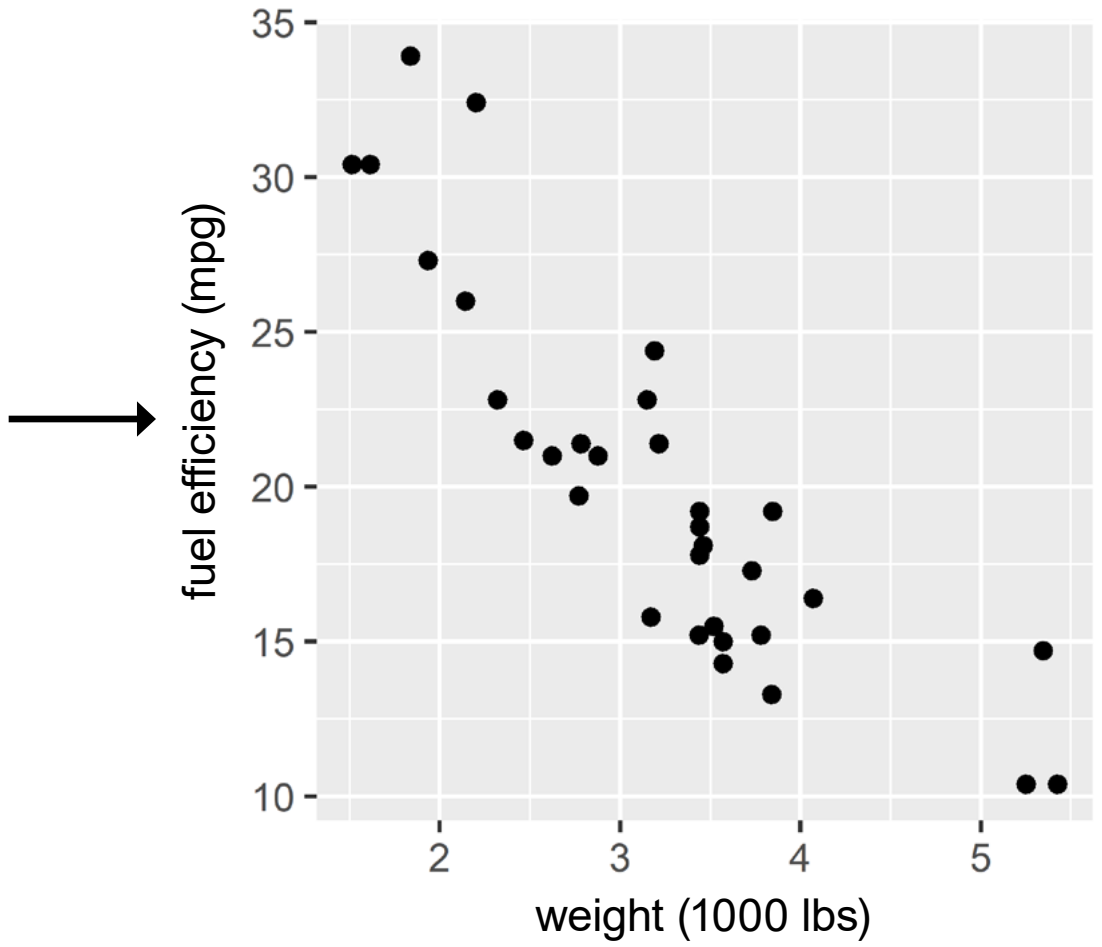
Visualize for communication



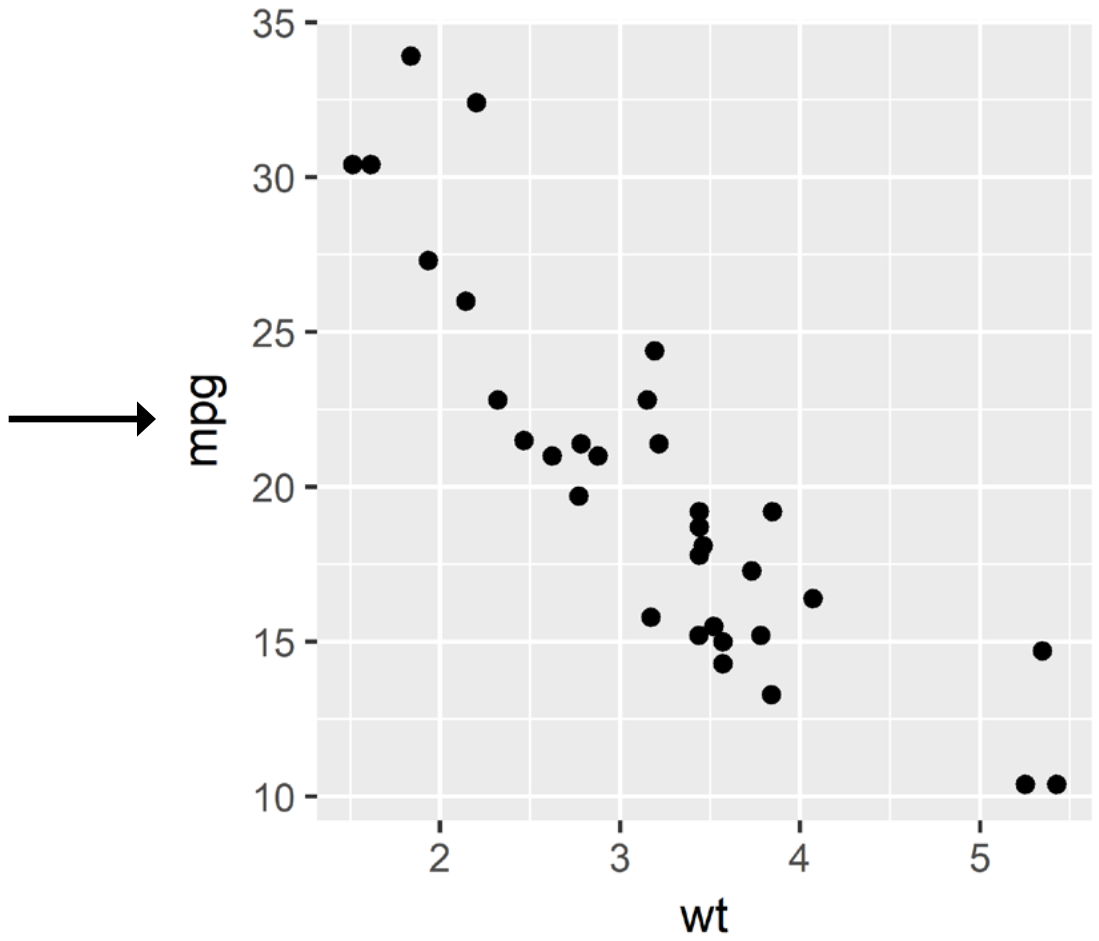
[https://www.bloomberg.com/graphics/2015-whats-warming-the-world/]

How do we turn data into visualizations?

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1



	mpg	cyl	disp	hp	drat	wt	qsec	vs	am
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1



Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

??? = some vis API
= some way of thinking about vis systematically

Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

??? = New function for every chart type:

scatter_plot(data, ...)

bar_chart(data, ...)

...

Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

??? = New function for every chart type:

scatter_plot(data, ...)

bar_chart(data, ...)

...

Every new chart is a new adventure!

Too many specs! — Too high level!

Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

??? = ~~New function for every chart type~~
= Low-level drawing functions
draw_point(...)
draw_rectangle(...)

Let's systematize "turning data into a vis"

data -> ??? -> marks on the screen (or paper)

??? = ~~New function for every chart type~~
= Low-level drawing functions
draw_point(...)
draw_rectangle(...)

Too low level!

Let's systematize "turning data into a vis"

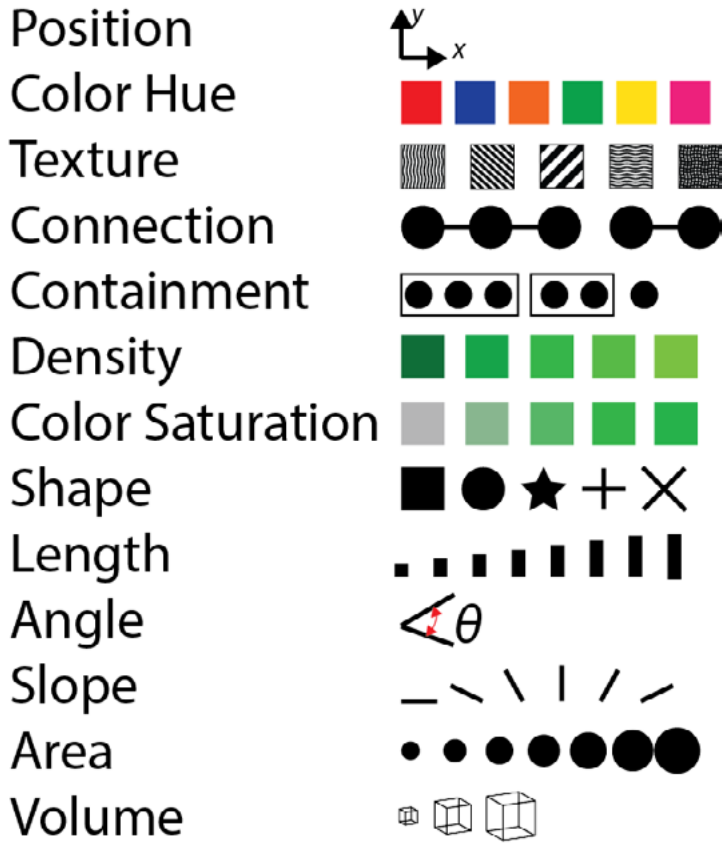
data -> ??? -> marks on the screen (or paper)

??? = ~~New function for every chart type~~
= ~~Low-level drawing functions~~
= Grammar of graphics

Encode data with visual channels
Display encodings with marks

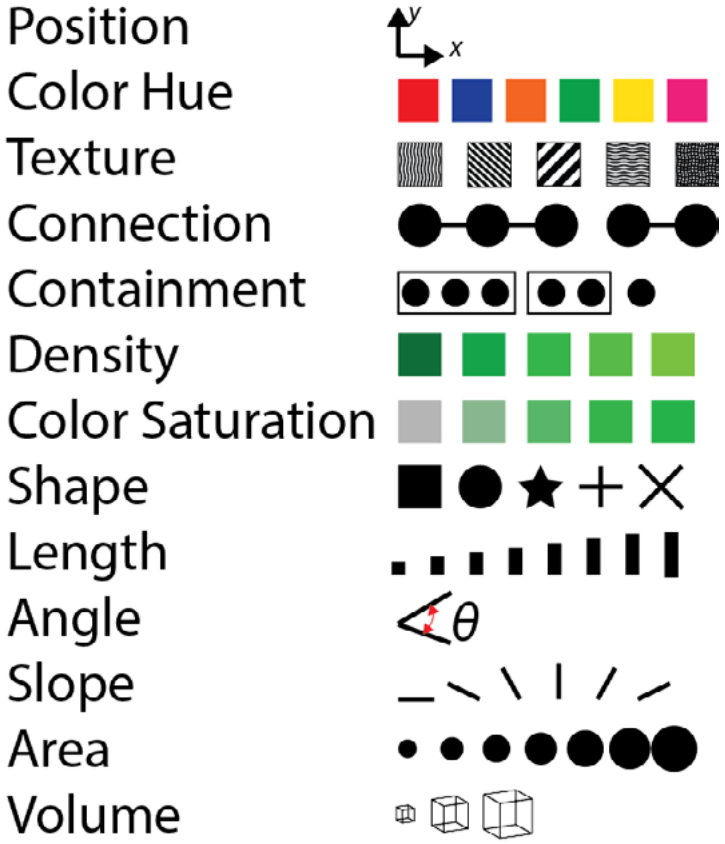
Visual channels

(ggplot “aesthetics”)

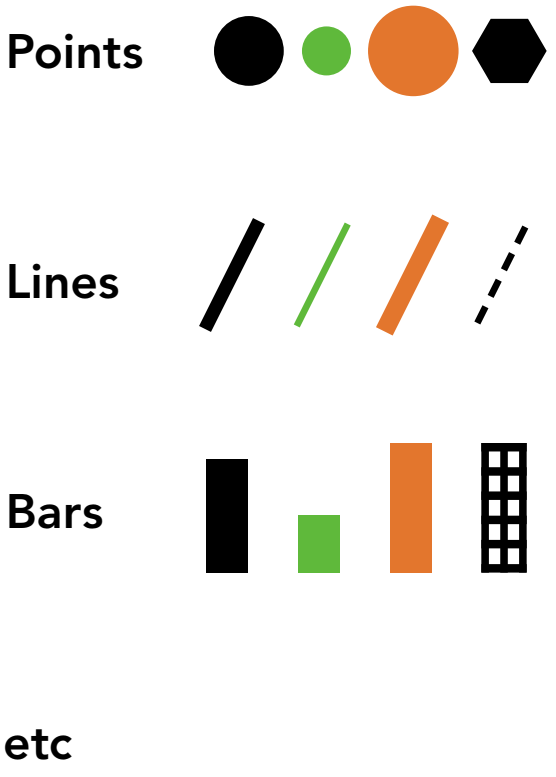


Visual channels ------> Marks

(ggplot "aesthetics")



(ggplot "geometries")



Grammar of graphics

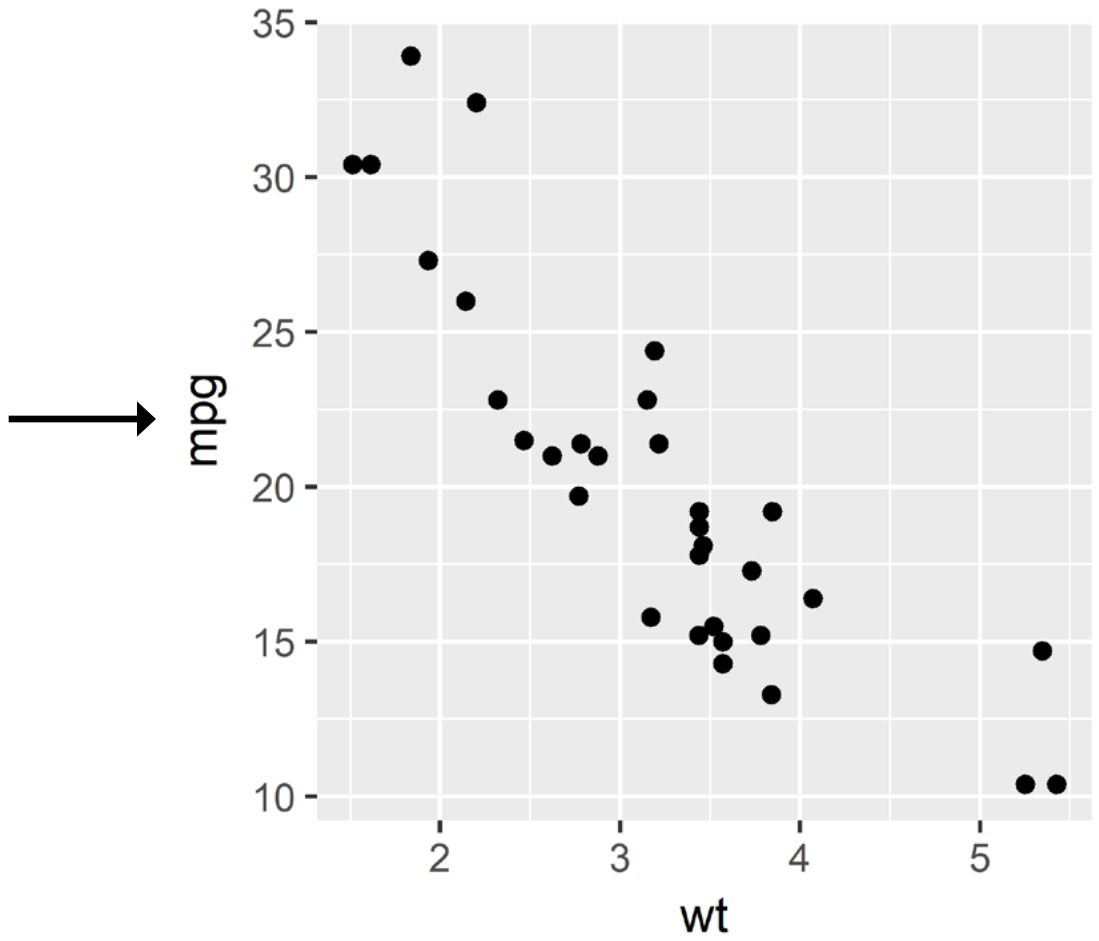
Codifies data types, encodings/channels, marks

Maps data -> channels -> marks

Makes visualization specification straightforward

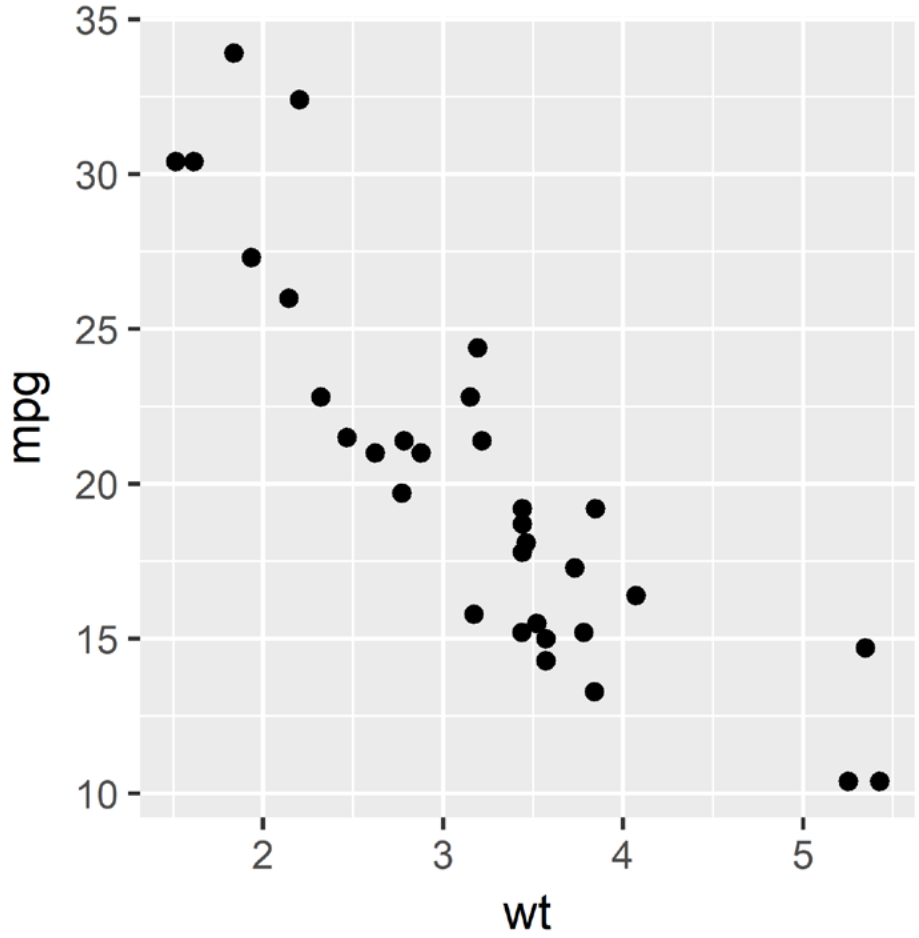
Undergirds ggplot, Tableau, Vega-Lite, Altair,...
(terms may vary)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1



Grammar of graphics

(data types, channels, marks)

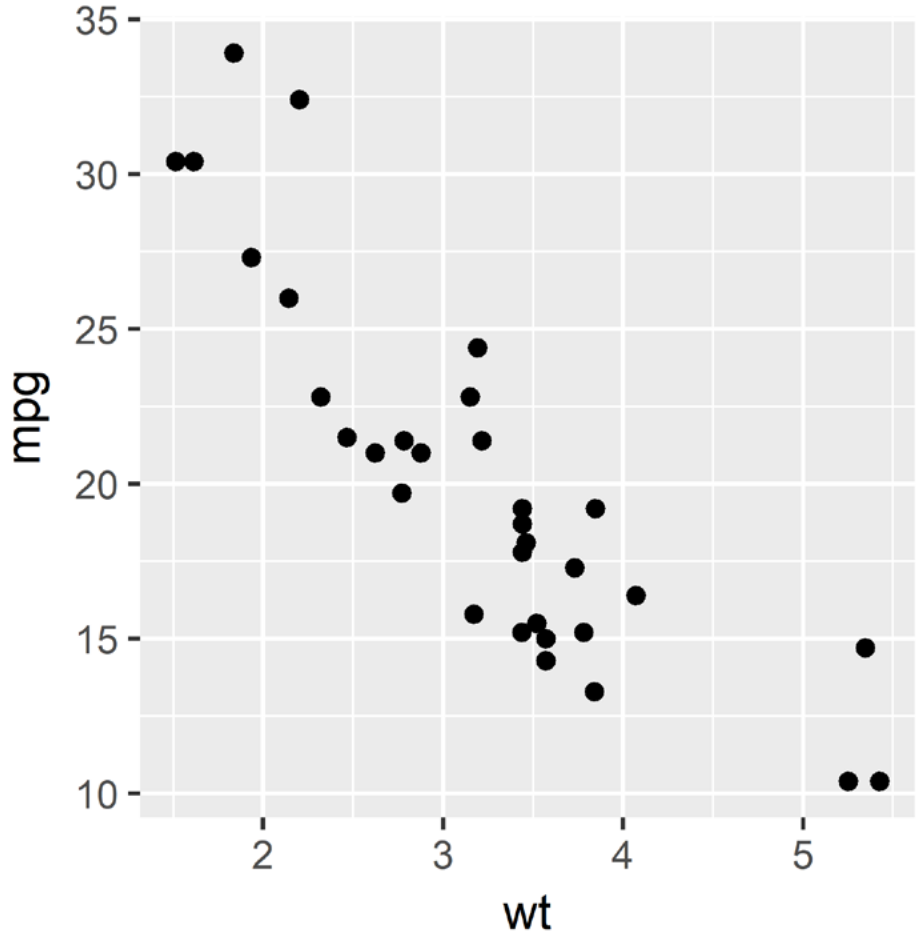


Grammar of graphics

(data types, channels, marks)

mpg: numeric

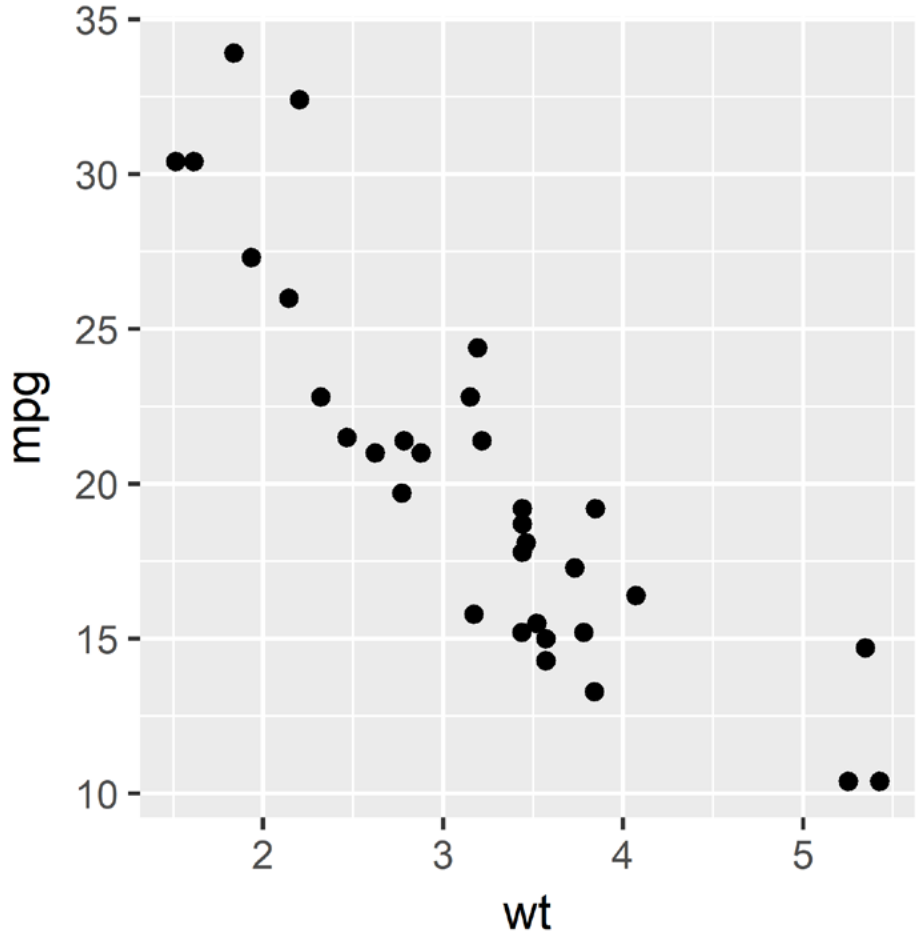
wt: numeric



Grammar of graphics

(data types, channels, marks)

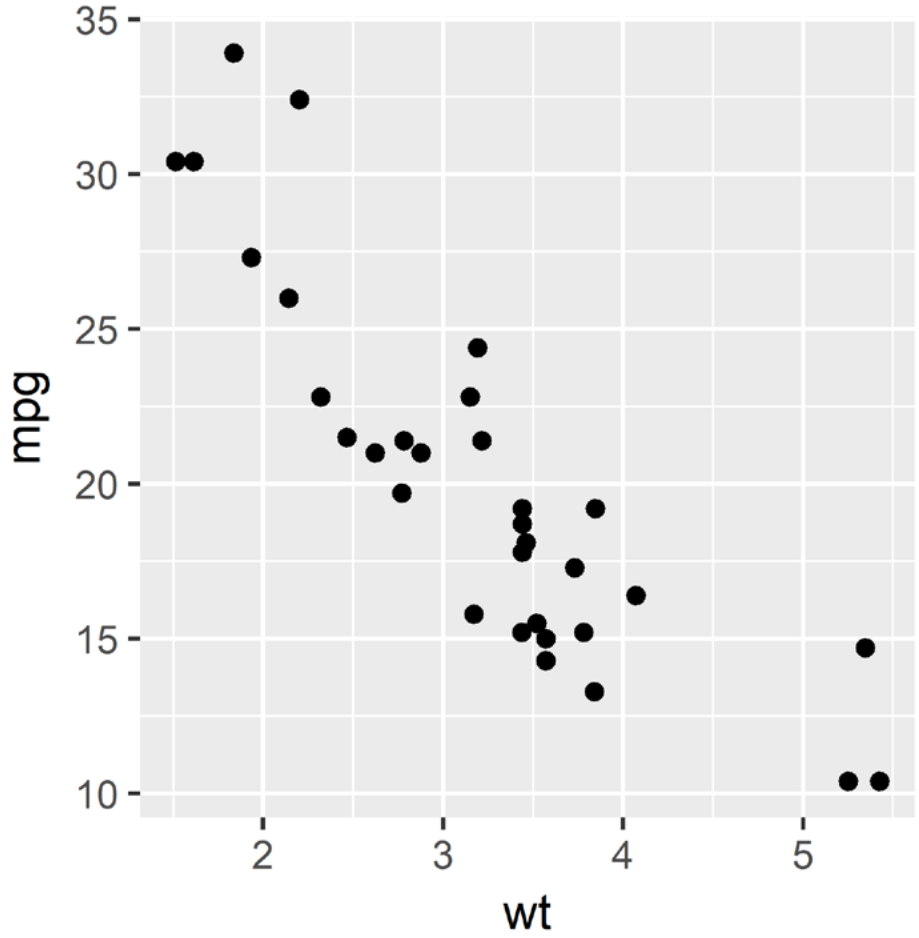
mpg: numeric
wt: numeric
wt -> x position
mpg -> y position



Grammar of graphics

(data types, channels, marks)

mpg: numeric
wt: numeric
wt -> x position
mpg -> y position
mark: point



Grammar of graphics

(data types, channels, marks)

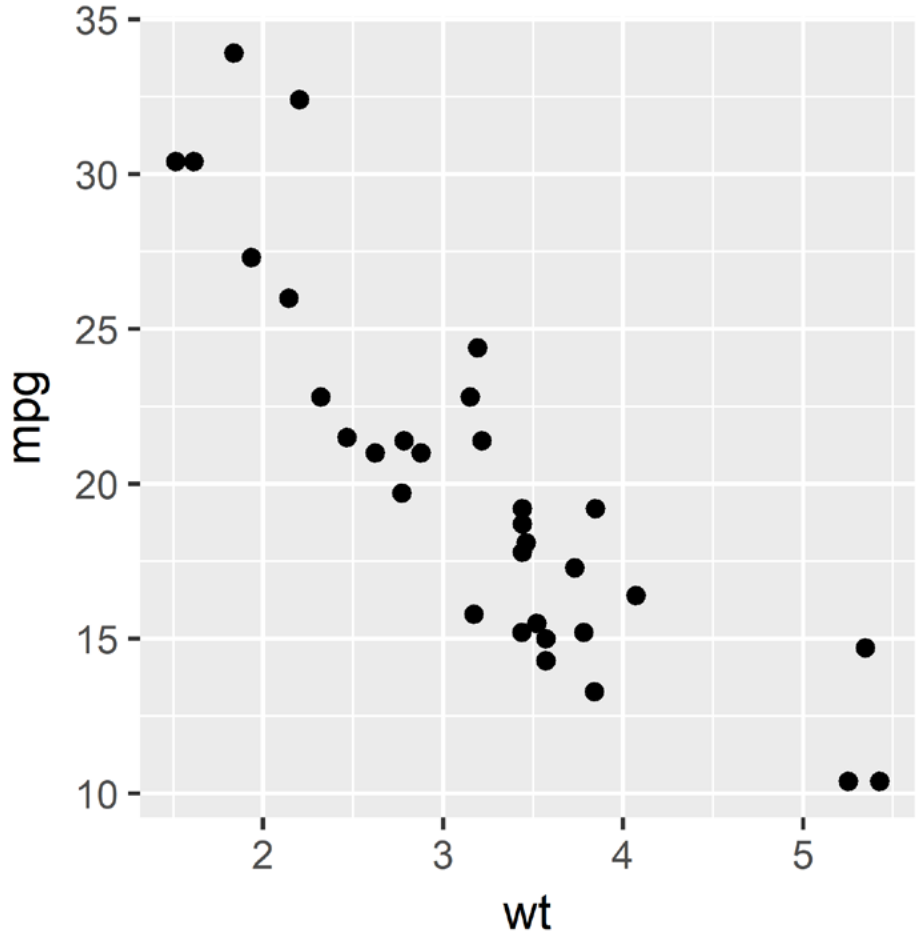
mpg: numeric

wt: numeric

wt -> x position

mpg -> y position

mark: point



Grammar of graphics

(data types, channels, marks)

mpg: numeric

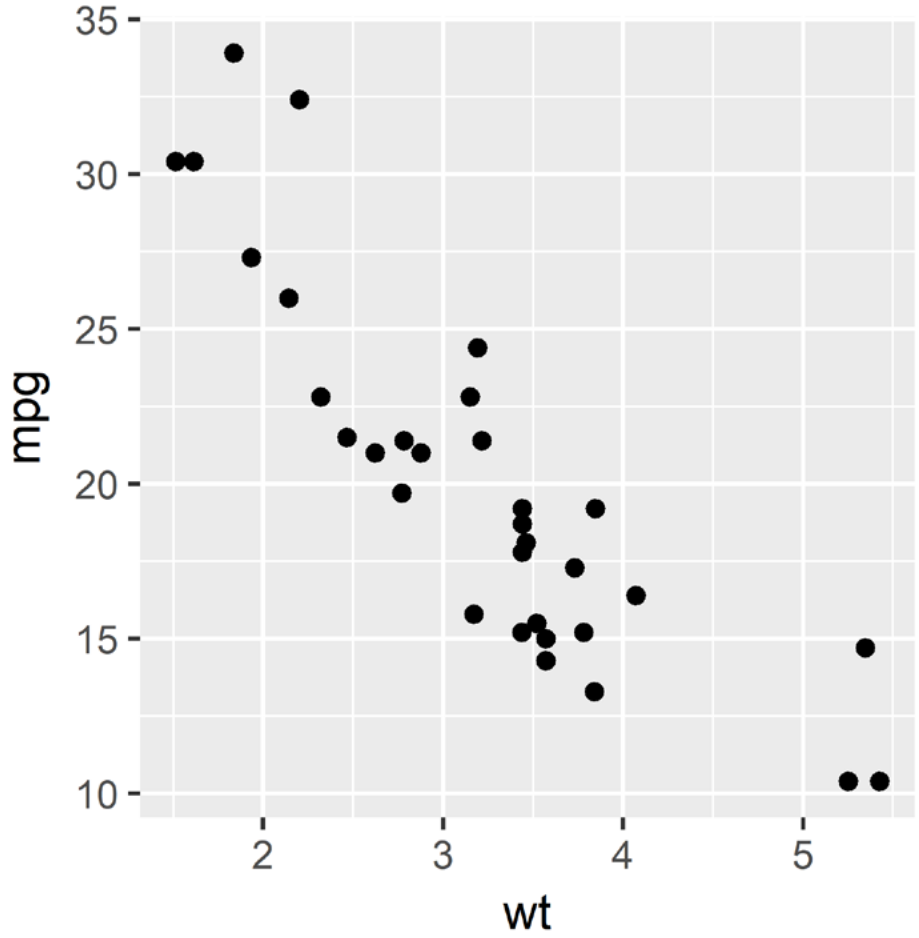
wt: numeric

manual: nominal

wt -> x position

mpg -> y position

mark: point



Grammar of graphics

(data types, channels, marks)

mpg: numeric

wt: numeric

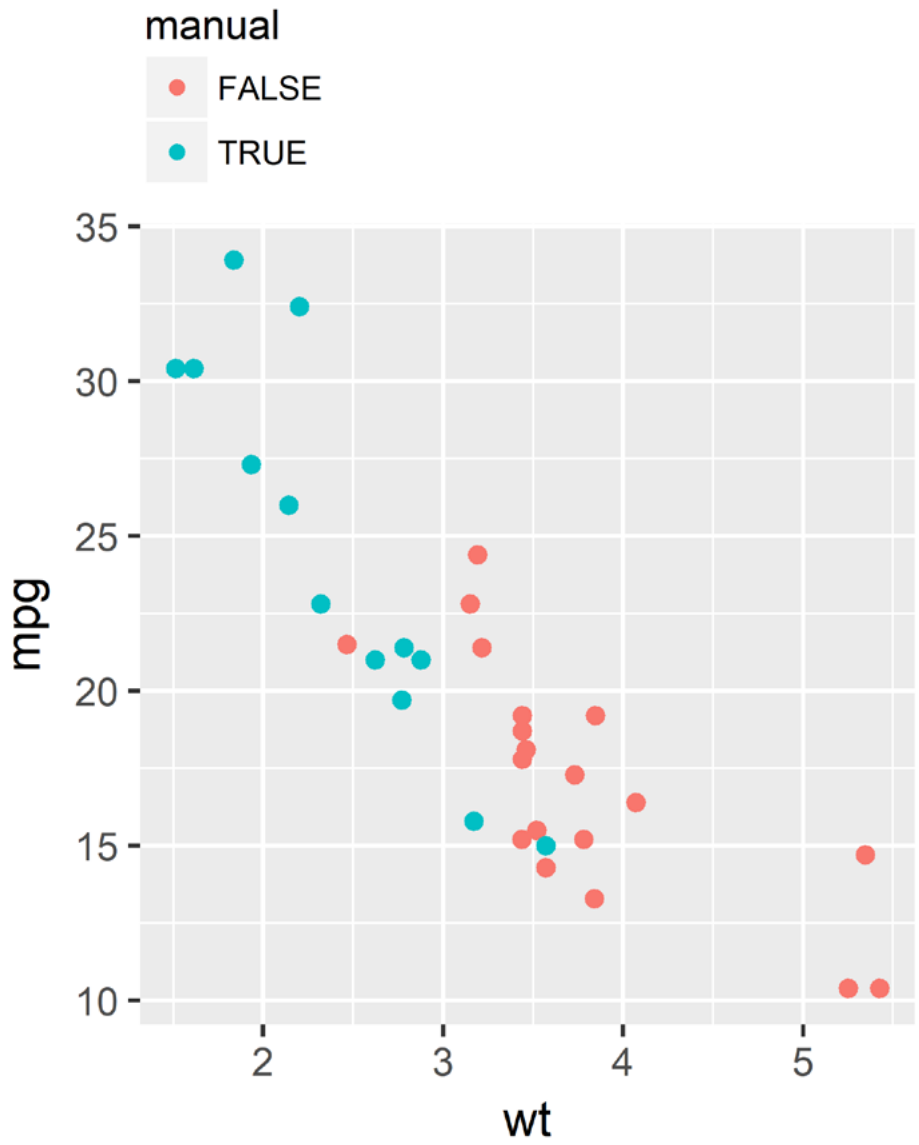
manual: nominal

wt -> x position

mpg -> y position

manual -> color

mark: point



Grammar of graphics

(data types, channels, marks)

mpg: numeric

wt: numeric

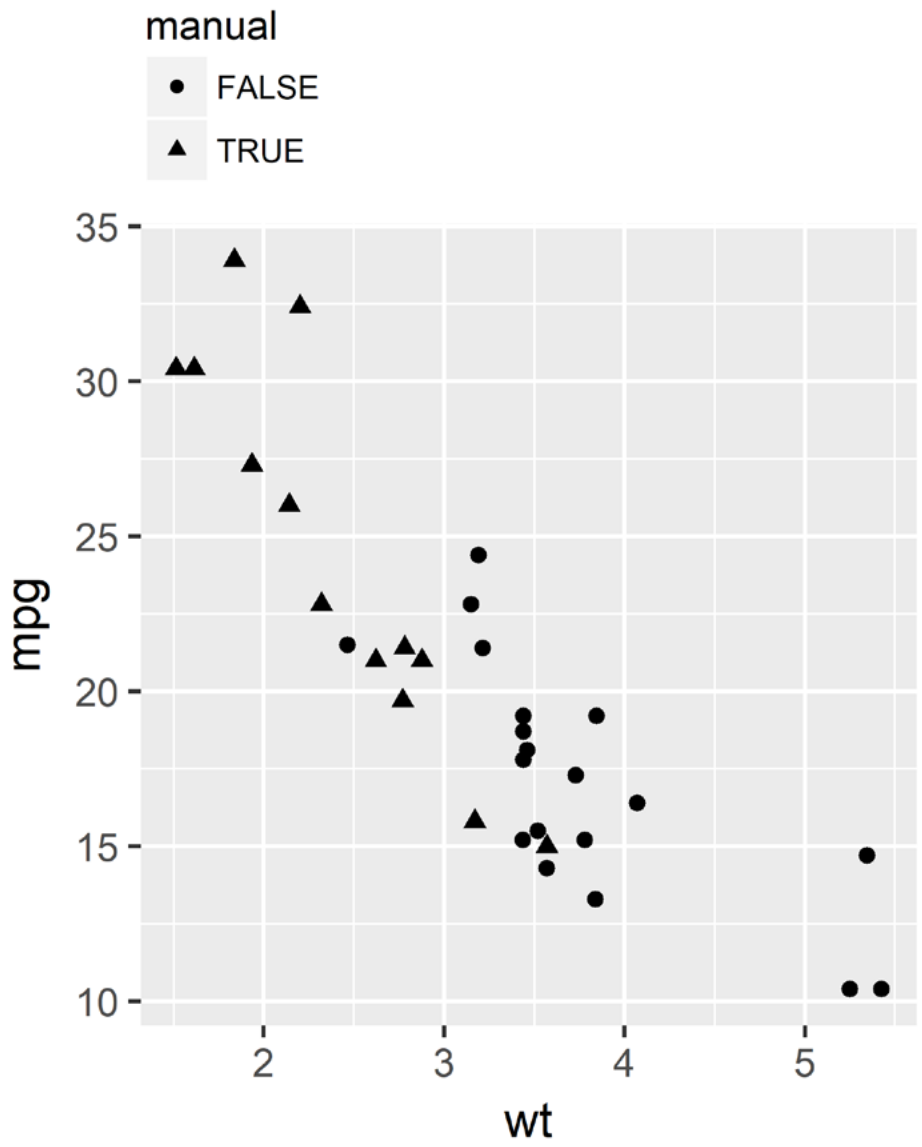
manual: nominal

wt -> x position

mpg -> y position

manual -> shape

mark: point



Grammar of graphics

(data types, channels, marks)

mpg: numeric

wt: numeric

manual: nominal

wt -> x position

mpg -> y position

manual -> color

manual -> shape

mark: point

Grammar of graphics

(data types, channels, marks)

mpg: numeric

wt: numeric

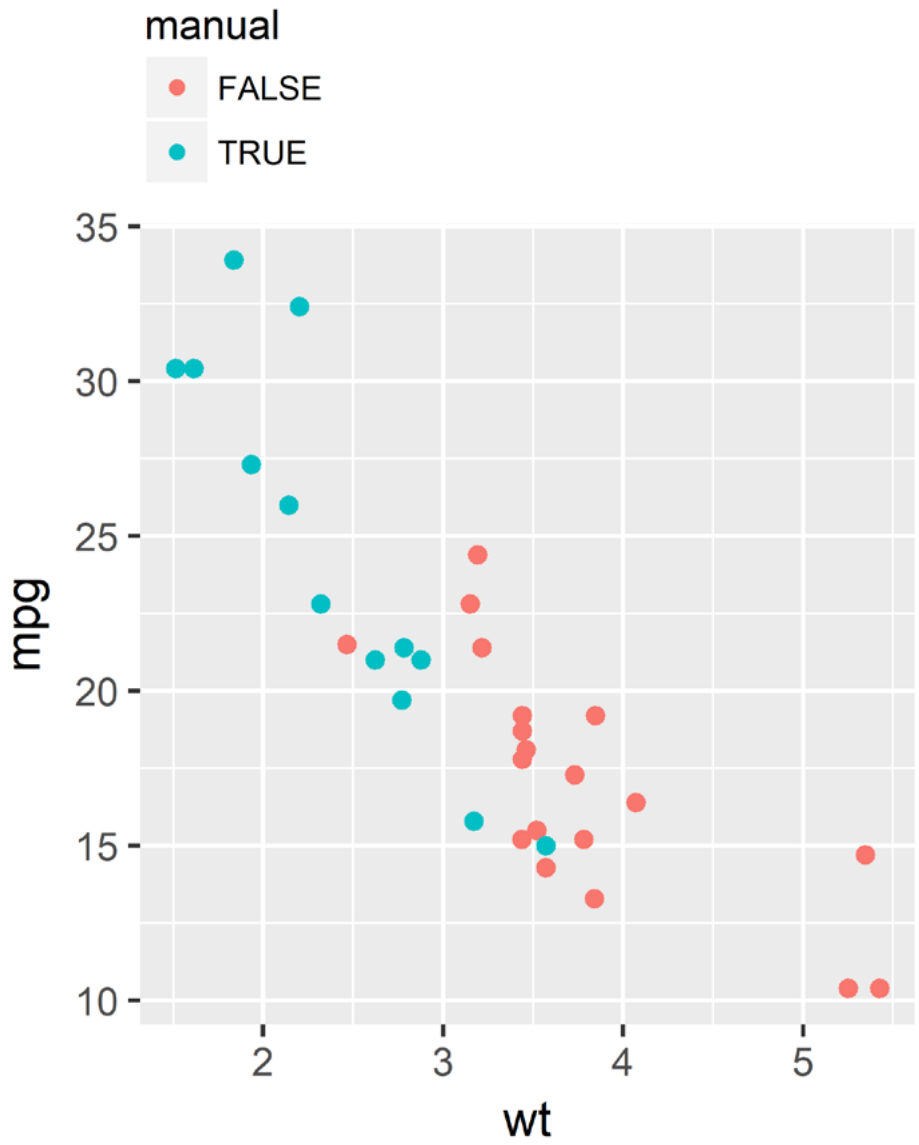
manual: nominal

wt -> x position

mpg -> y position

manual -> color

mark: point

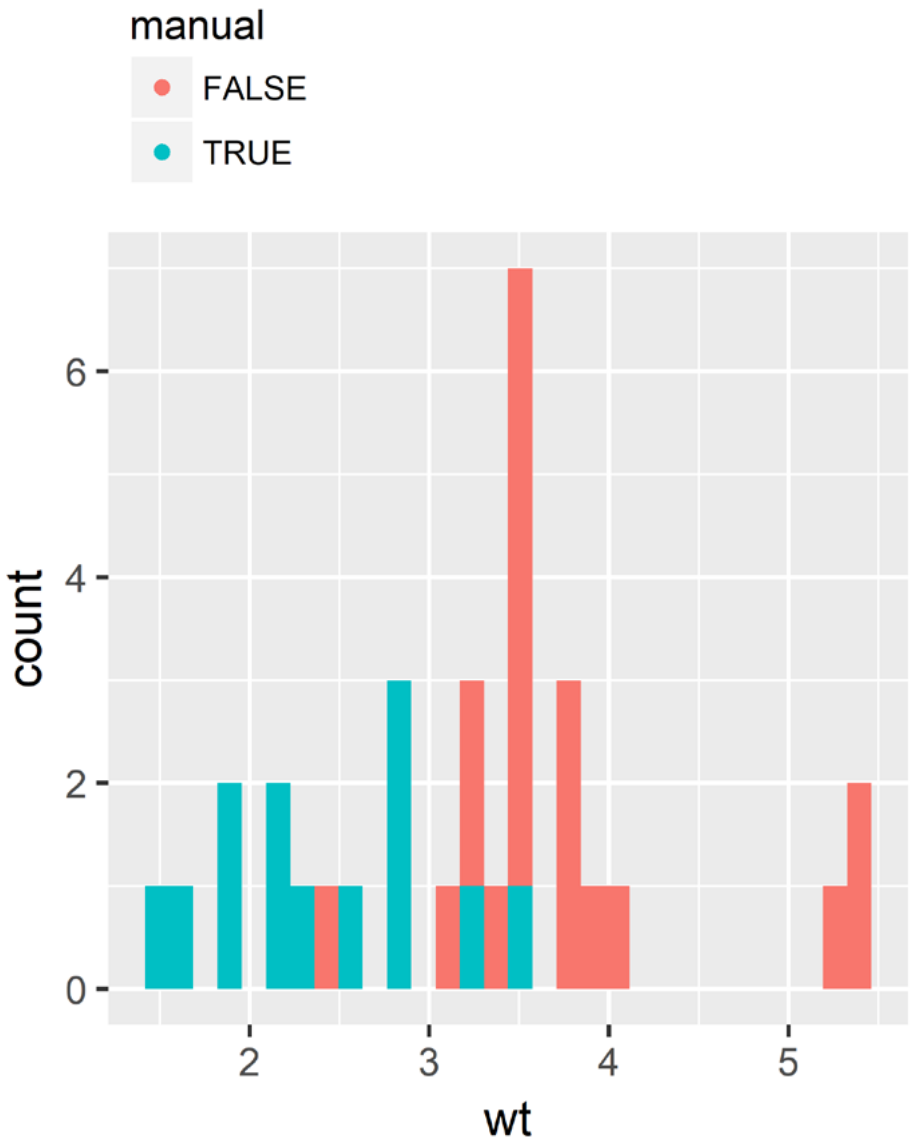


Grammar of graphics

(data types, channels, marks)

wt: numeric
manual: nominal
bin(wt) -> x position
count(wt)-> height
manual -> color

mark: bar



Why is the grammar of graphics useful?

1. Easier to specify many charts, combinations

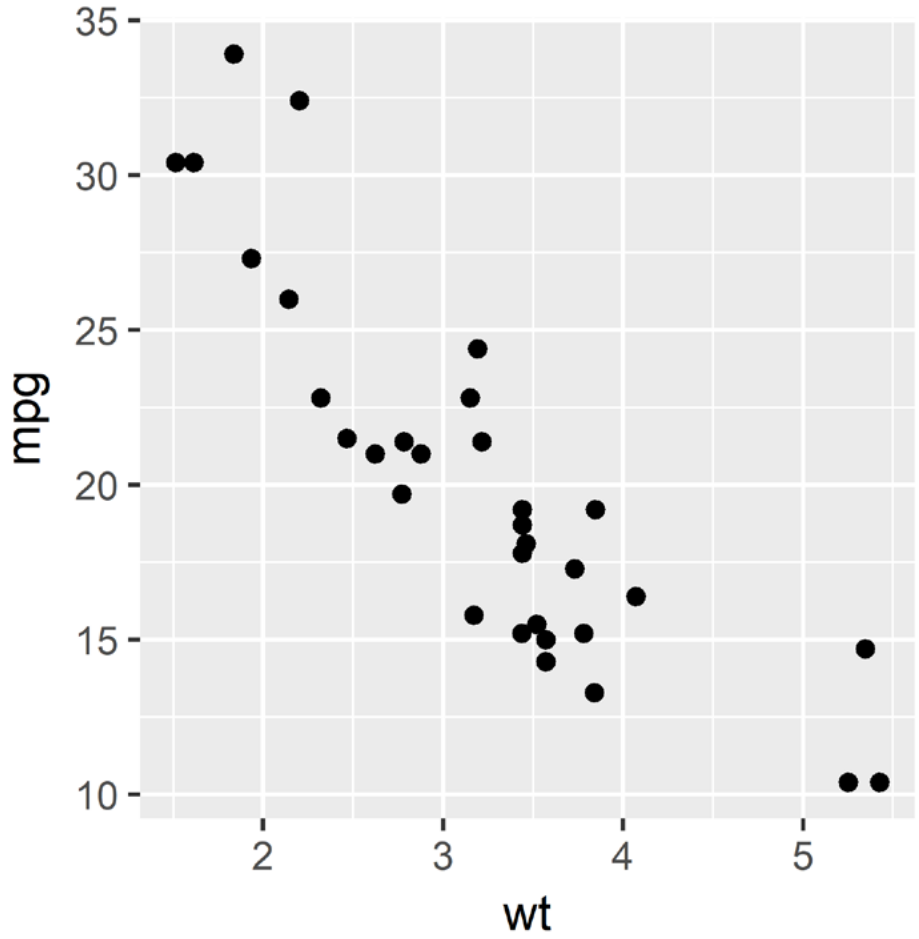
2. Helps you design/evaluate charts systematically

1. Easier to specify many charts, combinations

mpg: numeric
wt: numeric

wt -> x position
mpg -> y position

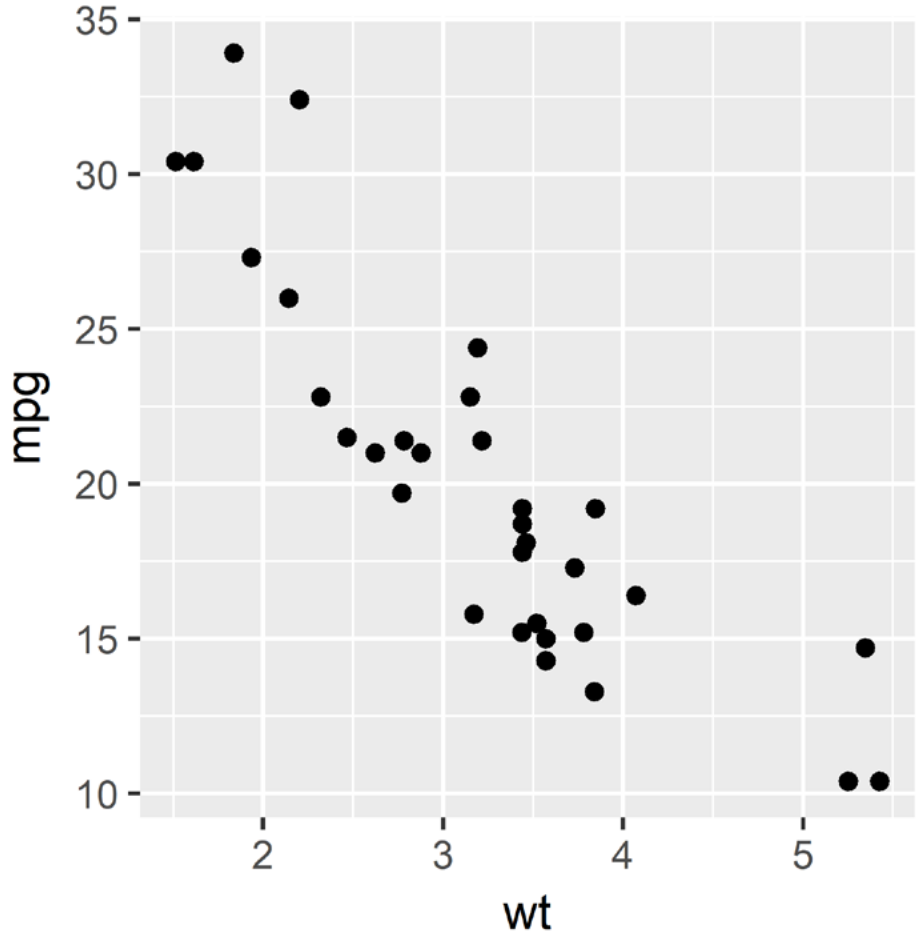
mark: point



1. Easier to specify many charts, combinations

Not:

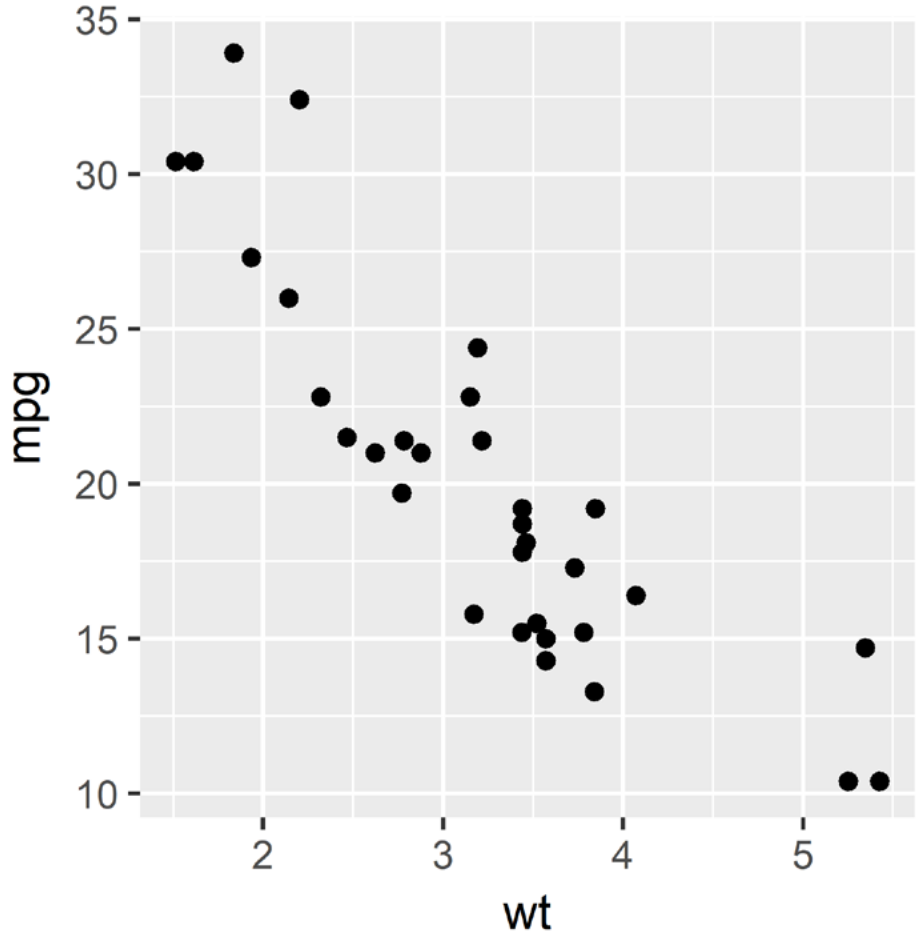
```
some_big_function_to_make_scatterplots(
    my_data,
    a_bunch_of_options
)
```



1. Easier to specify many charts, combinations

Not:

```
some_function_to_draw_grid()
some_function_to_draw_axes()
for (row in data) {
   draw_point(data[i]["x"], ...)
}
 ...
```

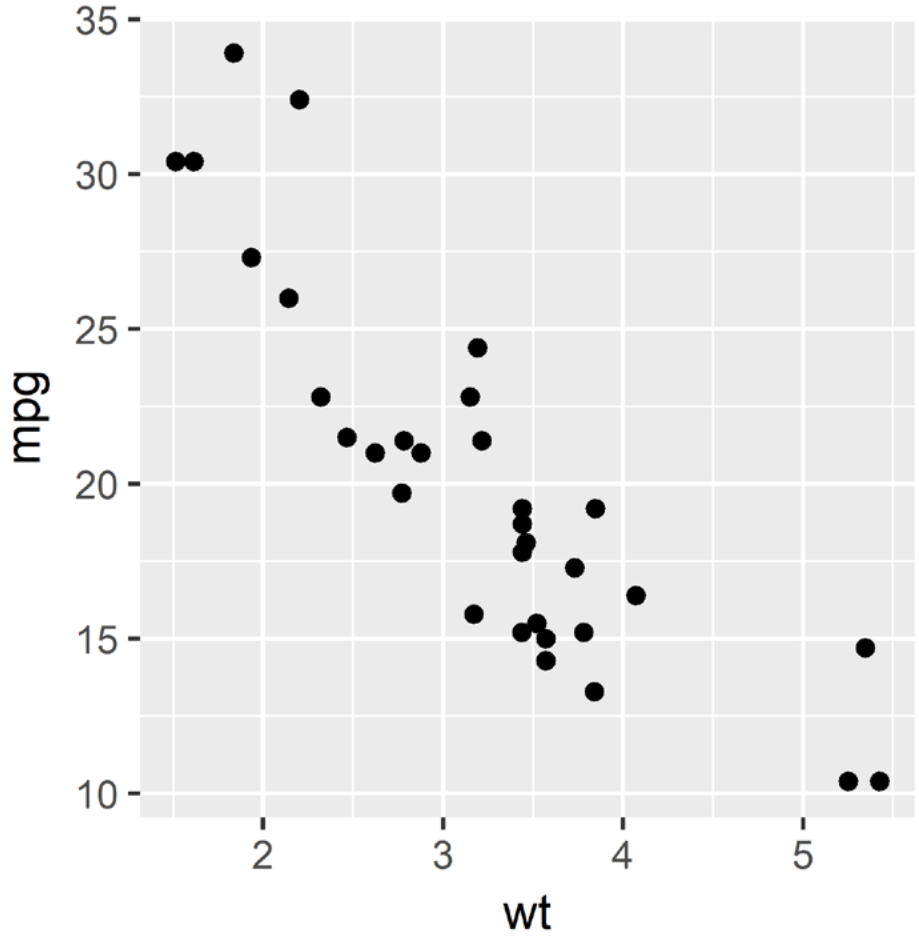


1. Easier to specify many charts, combinations

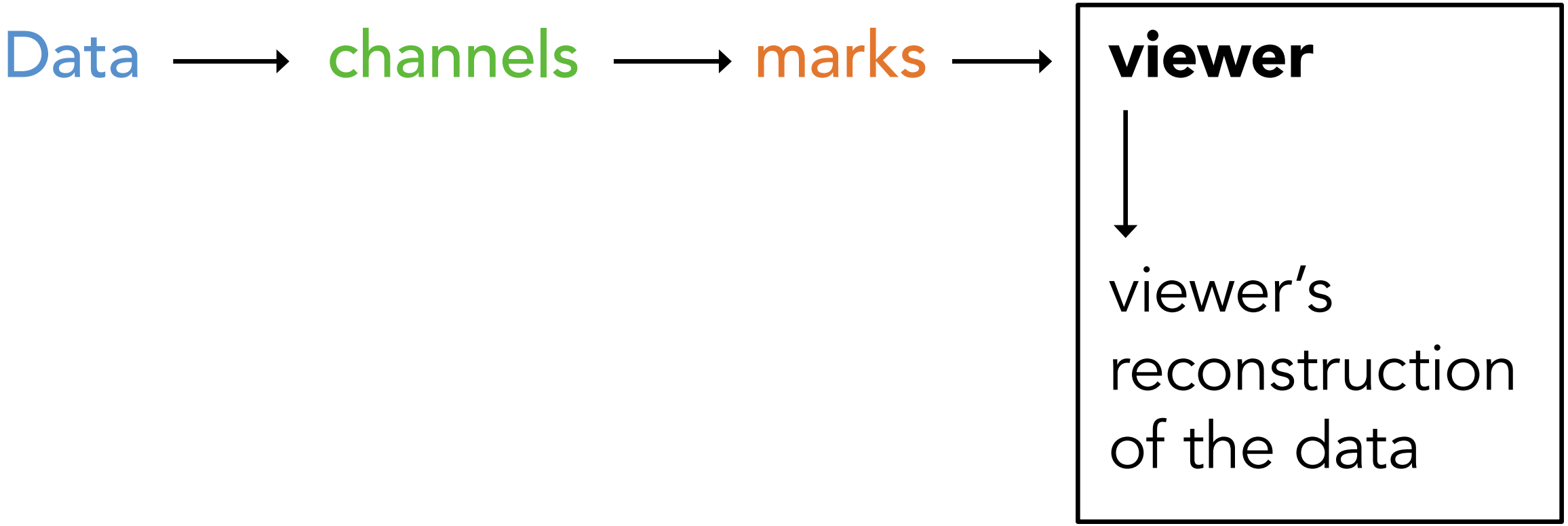
e.g., in ggplot

(data, channels, marks):

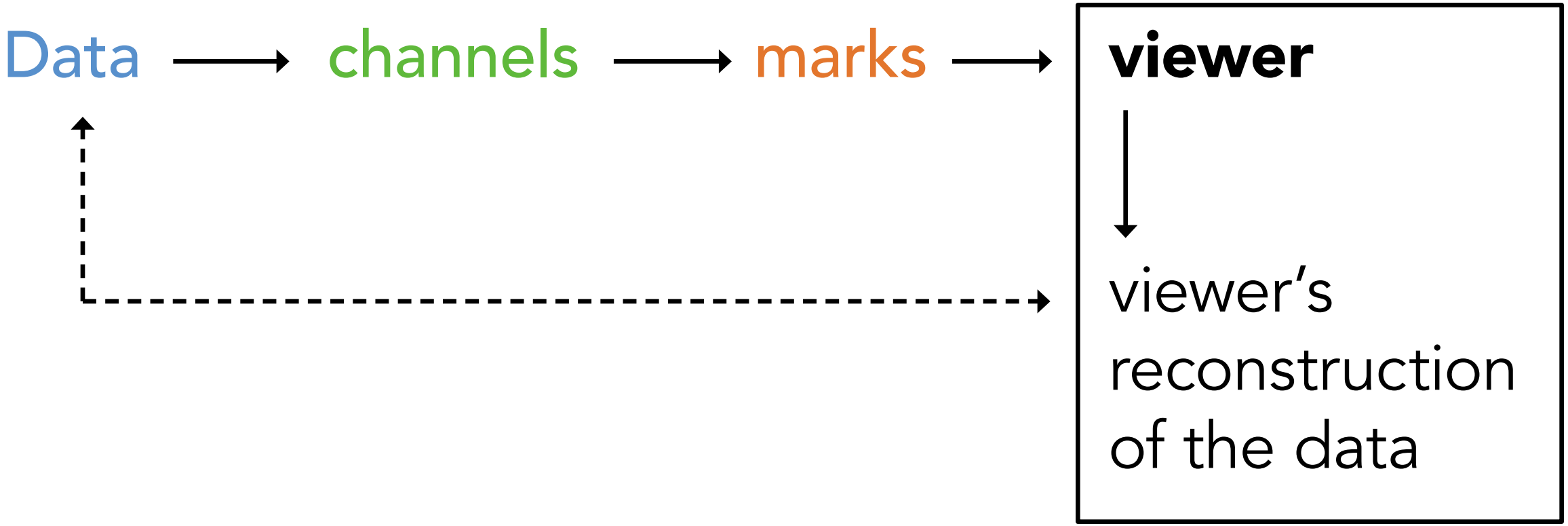
```
ggplot(mtcars, aes(
      x = wt,
      y = mpg
   )) +
   geom_point()
```



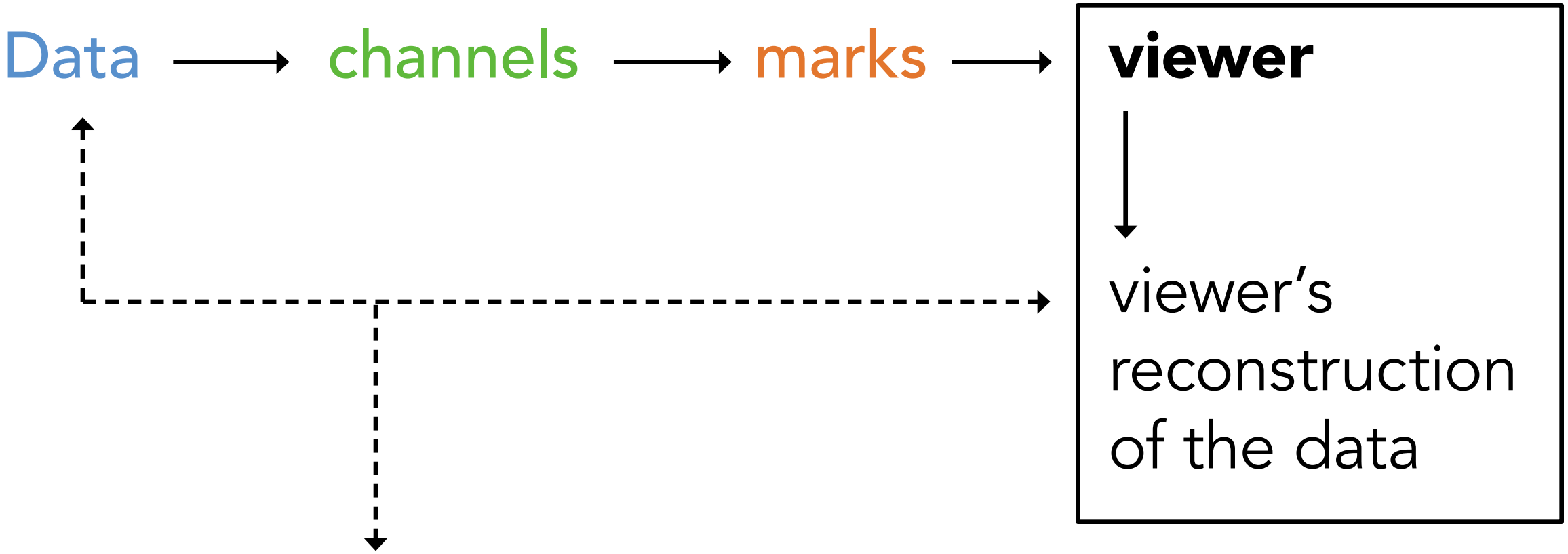
2. Helps you design charts systematically



2. Helps you design charts systematically



2. Helps you design charts systematically



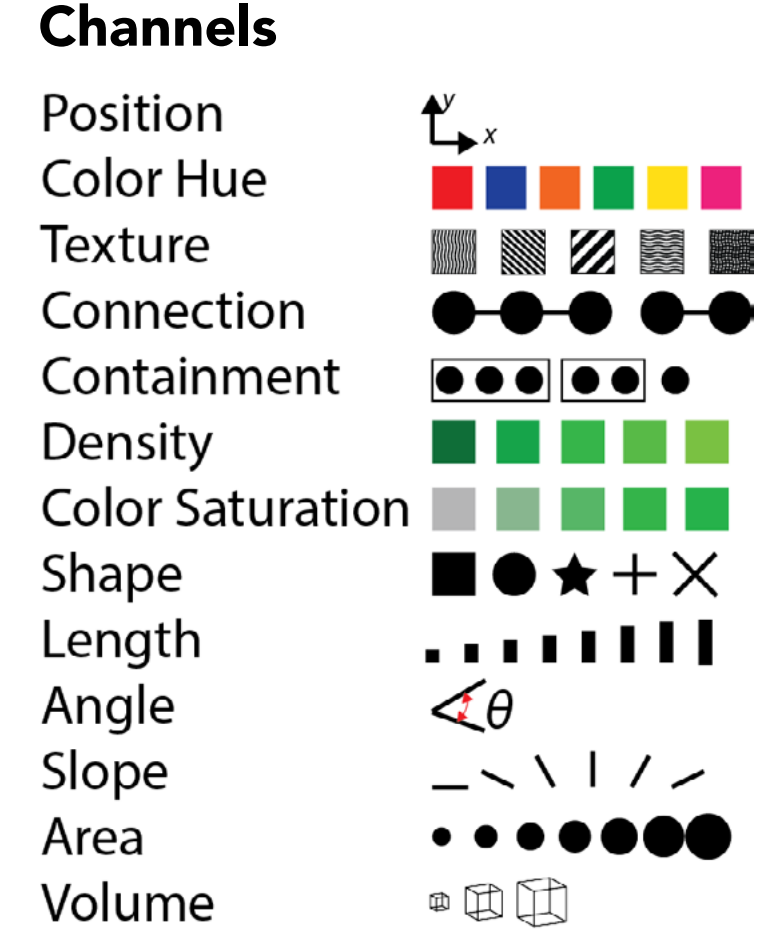
How well do these match, given the channel used?

2. Helps you design charts systematically

E.g.,

How accurately do people perceive position?

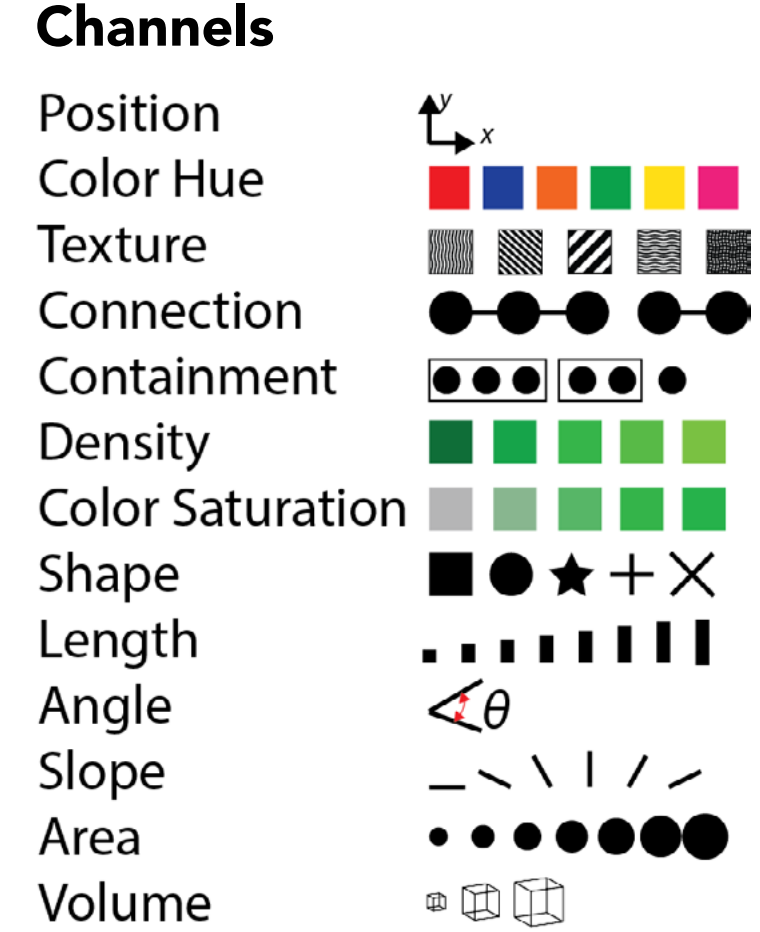
How accurately do people perceive area?



2. Helps you design charts systematically

E.g.,

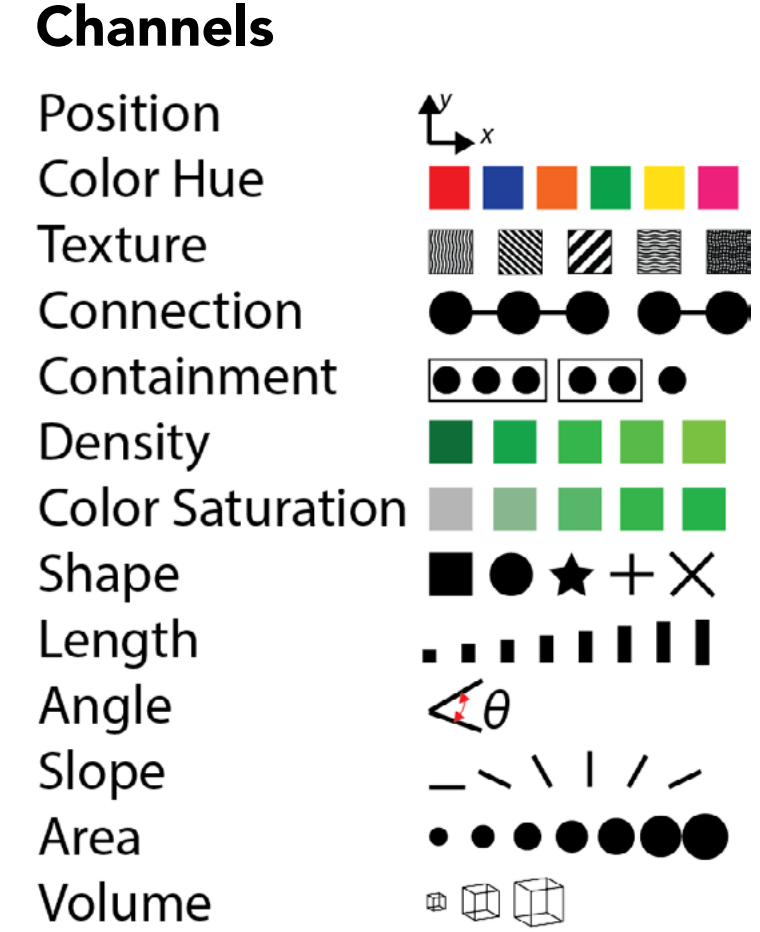
How accurately do people perceive position for quantitative data?
...for ordered data?
...for nominal data?
etc.



2. Helps you design charts systematically

E.g.,

What channel is best for quantitative data?
...for ordered data?
...for nominal data?
etc.



How good is a visual channel / encoding?

Length encoding:

- 
- 

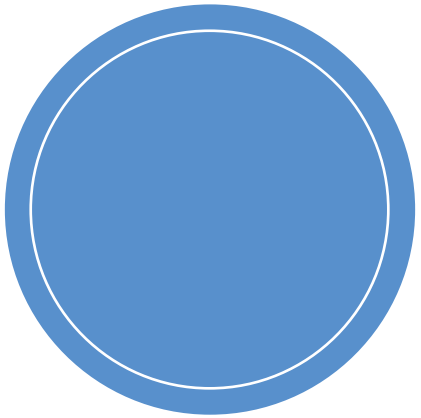
How good is a visual channel / encoding?

Length encoding:

How good is a visual channel / encoding?

Length encoding:

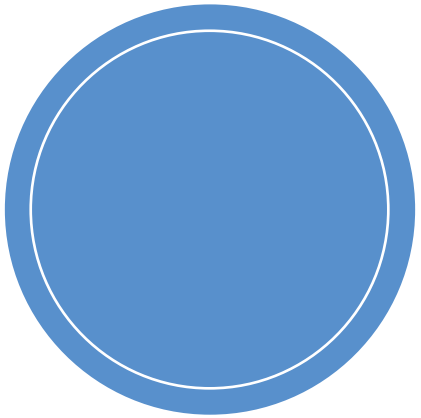
Area encoding:

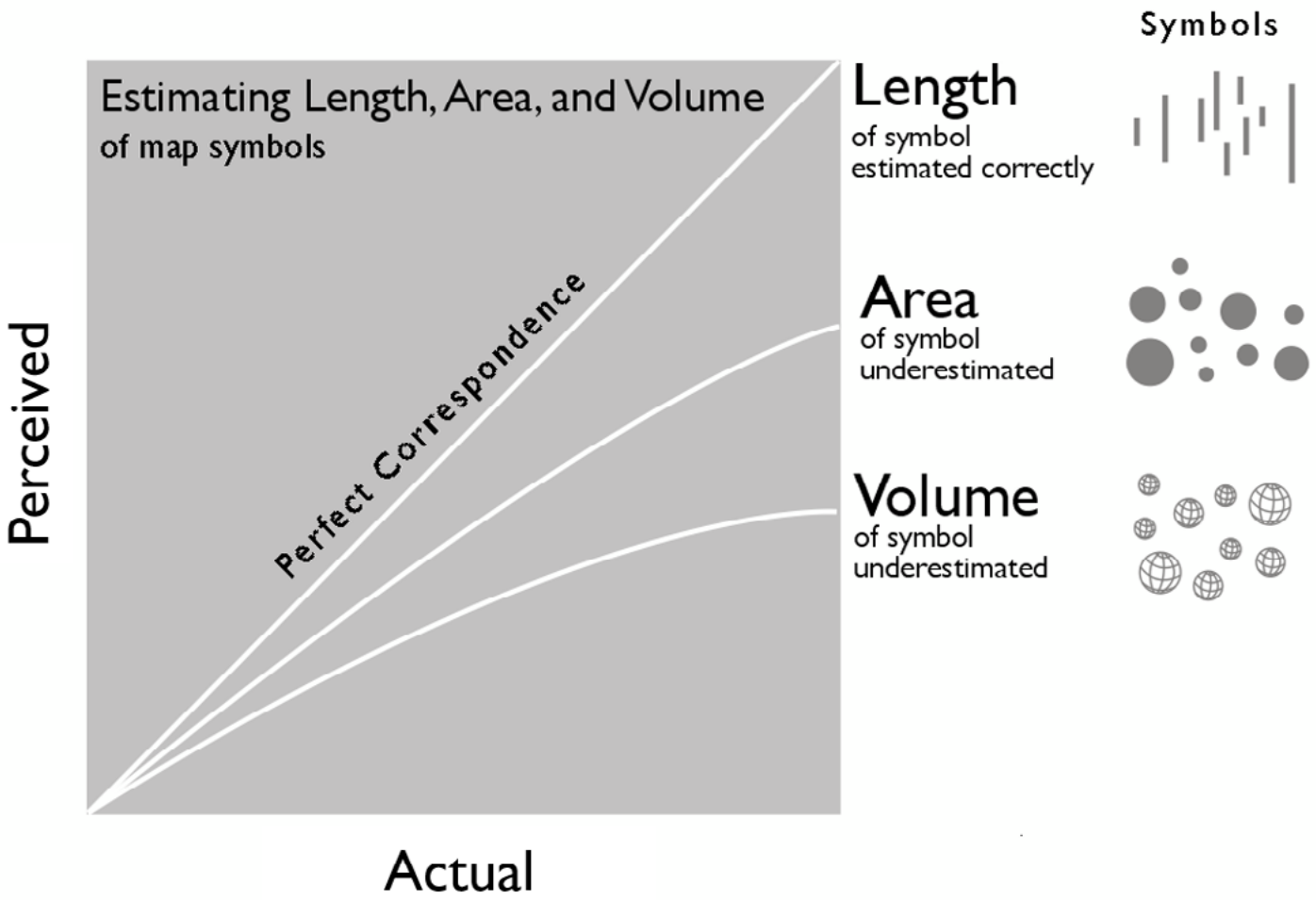


How good is a visual channel / encoding?

Length encoding:

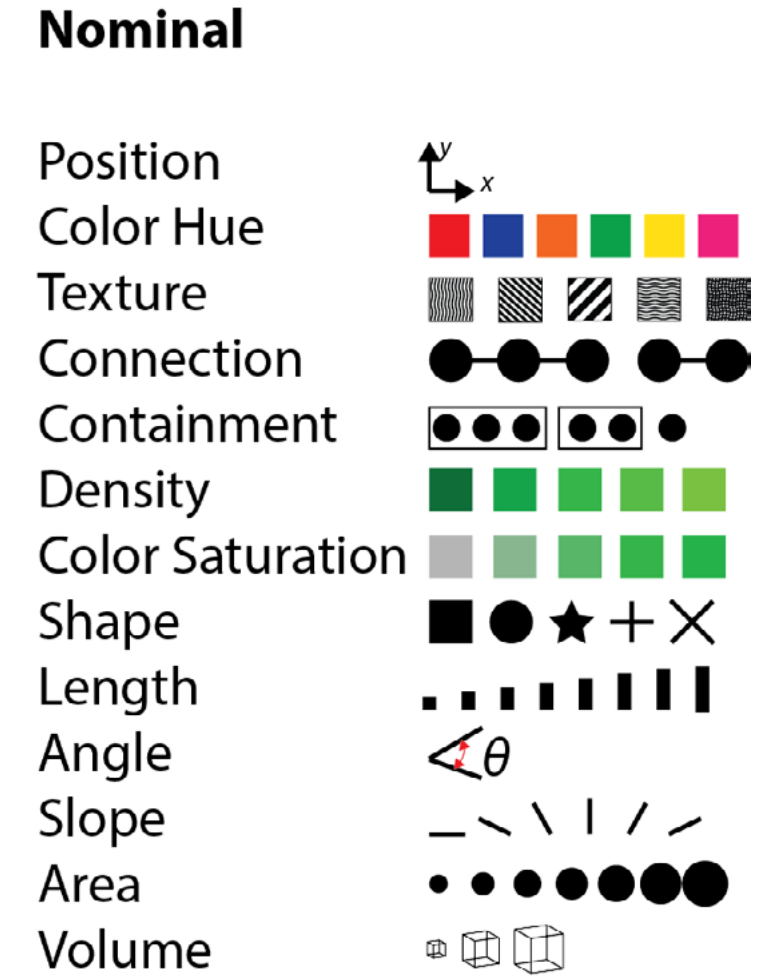
Area encoding:



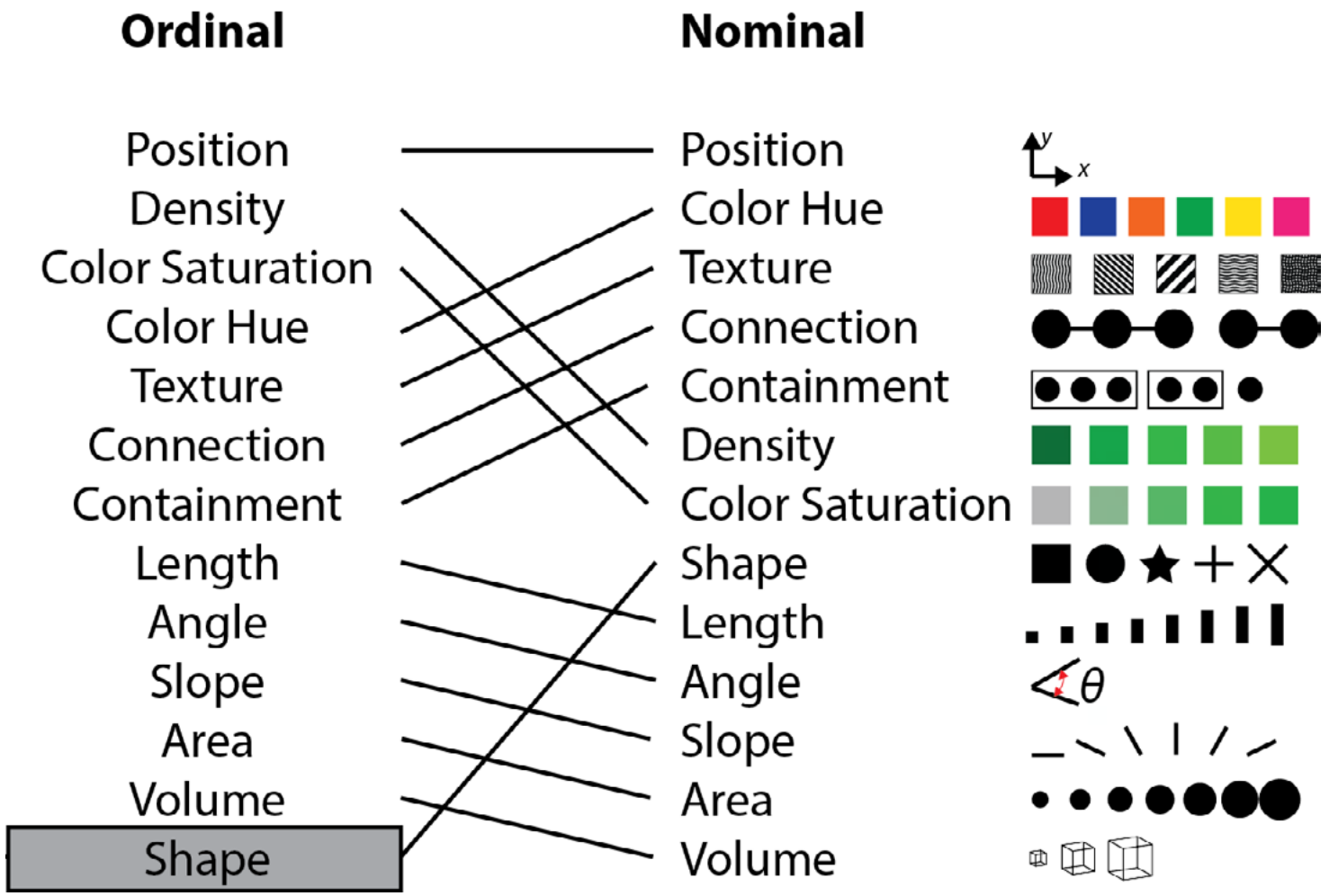


Encodings help us judge chart effectiveness

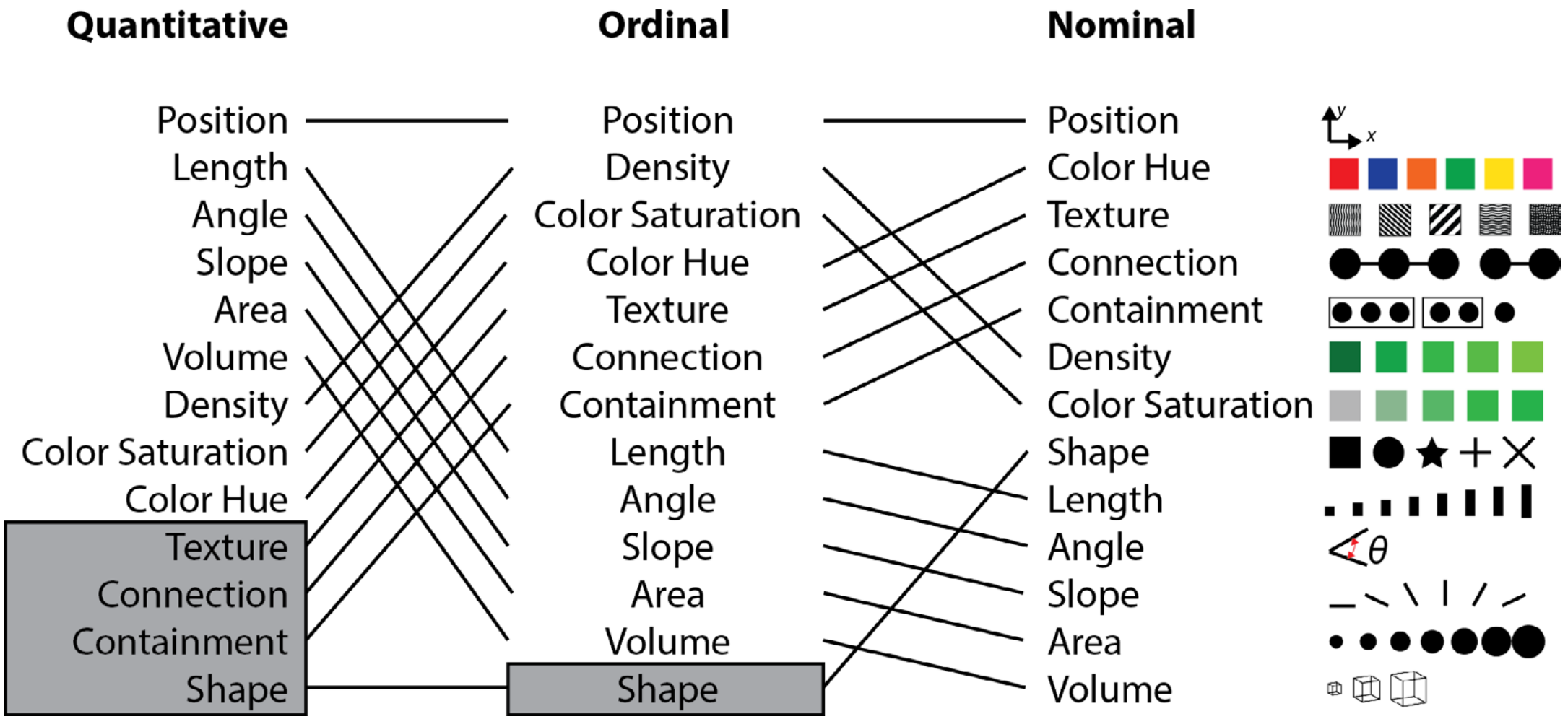
What encoding is best for quantitative data?
...for ordered data?
...for nominal data?
etc.



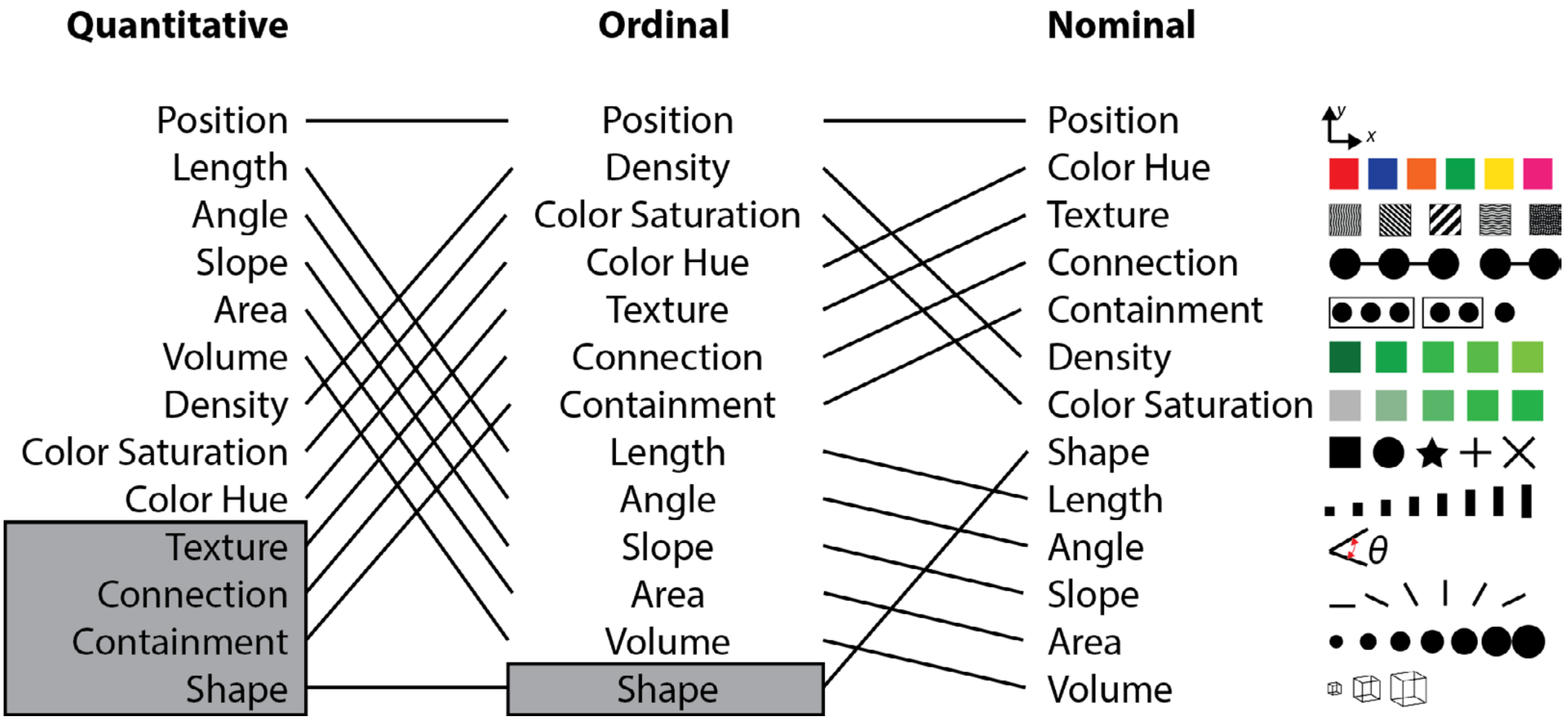
Encodings help us judge chart effectiveness



Encodings help us judge chart effectiveness

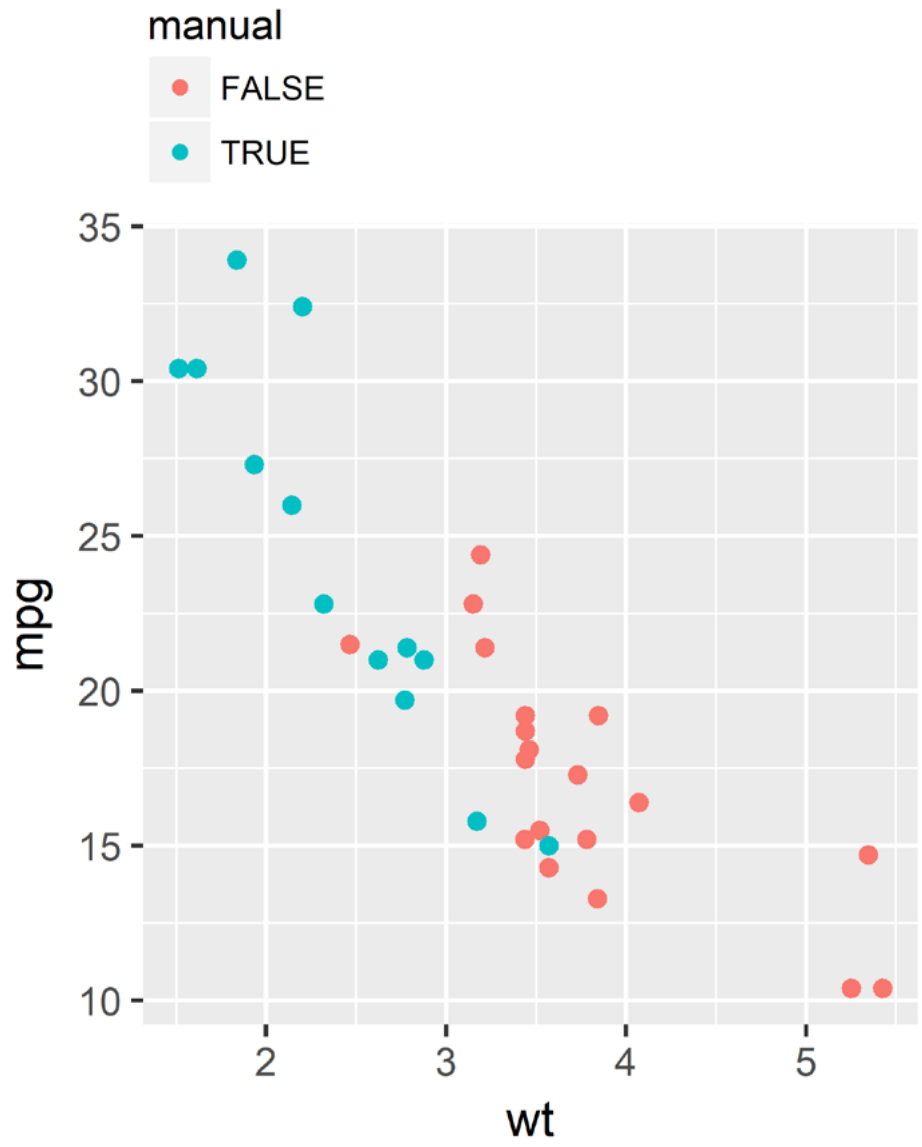


Pick one, cross it off...



Pick one, cross it off...

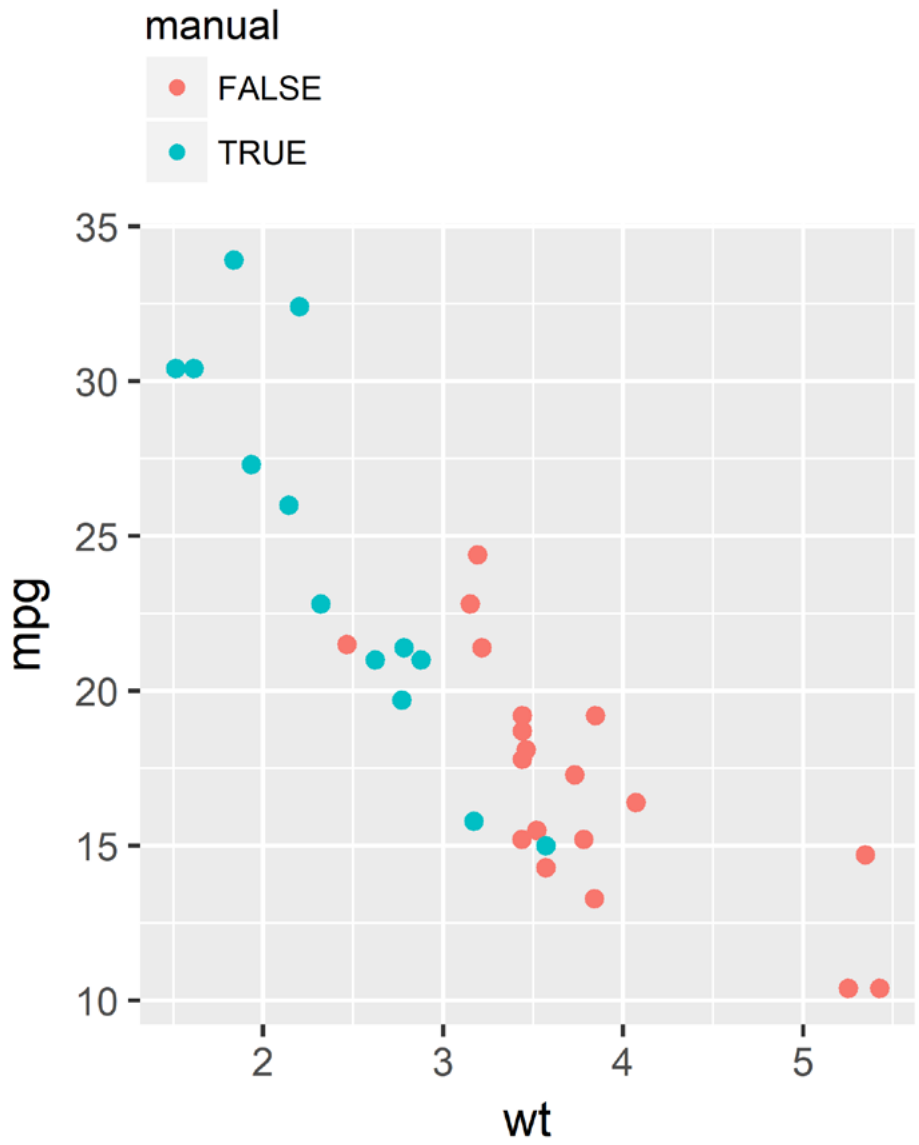
Quantitative	Nominal
Position	Position
Length	Color Hue
Angle	Texture
Slope	Connection
Area	Containment
Volume	Density
Density	Color Saturation
Color Saturation	Shape
Color Hue	Length
Texture	Angle
Connection	Slope
Containment	Area
Shape	Volume



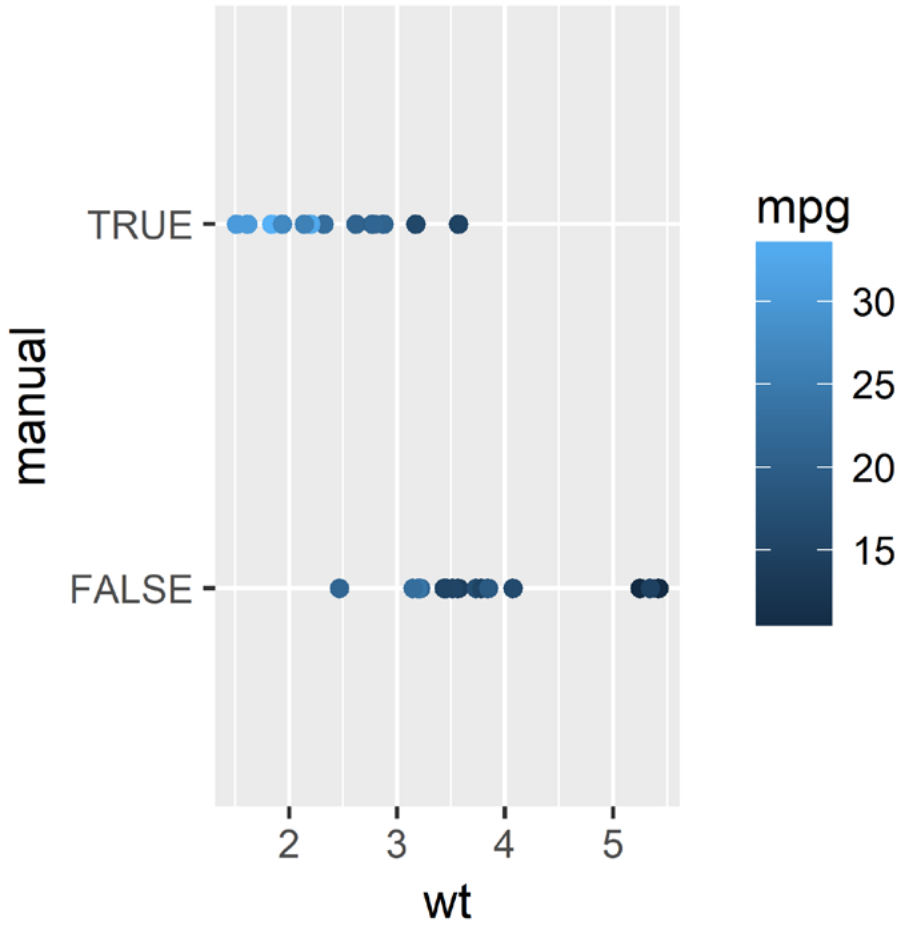
Effectiveness

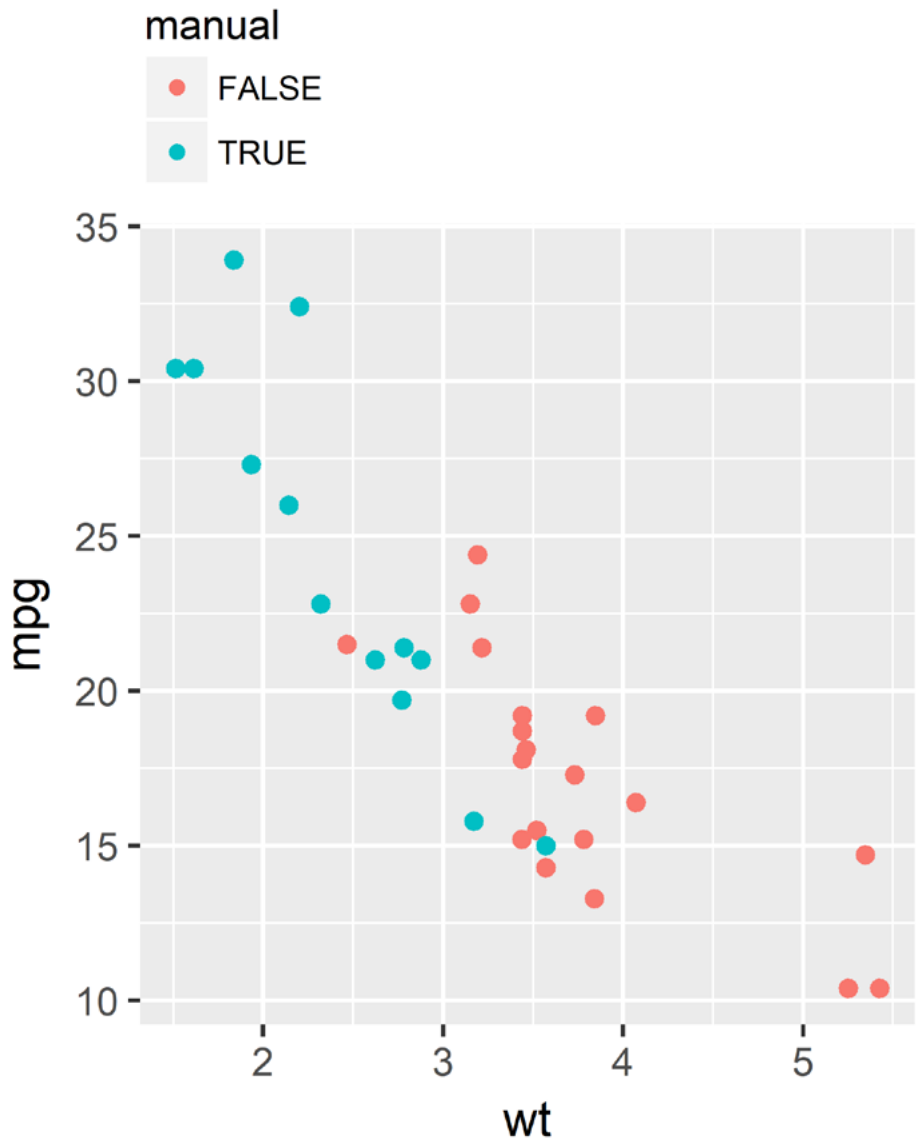
This chart works because it uses accurate channels (ones with low estimation error).

This is the essence of effectiveness.

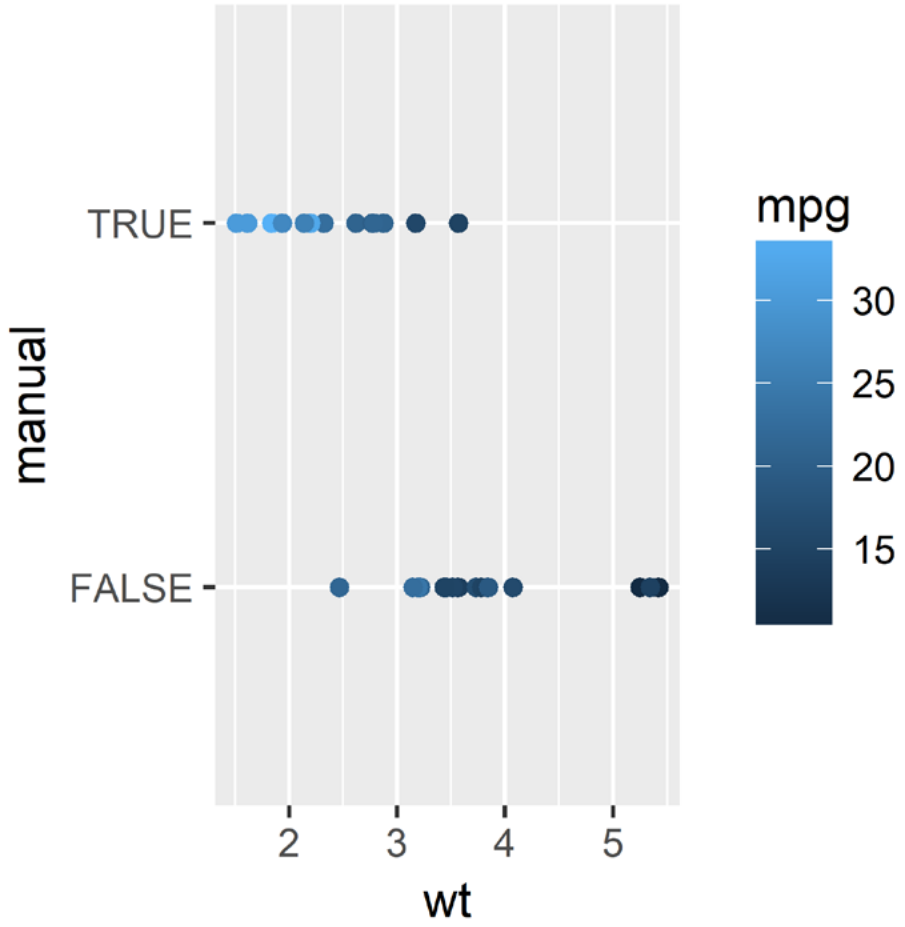


What about this?





What about this?



Quantitative

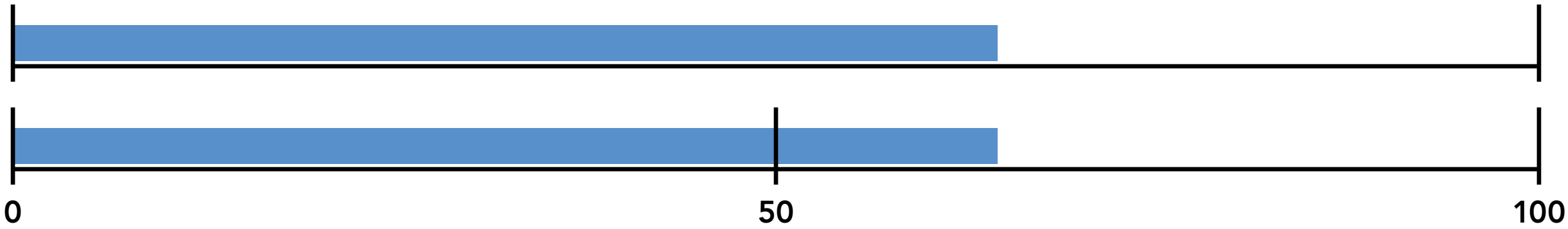
Position
Length
Angle
Slope
Area
Volume
Density
Color Saturation
Color Hue

Texture
Connection
Containment
Shape

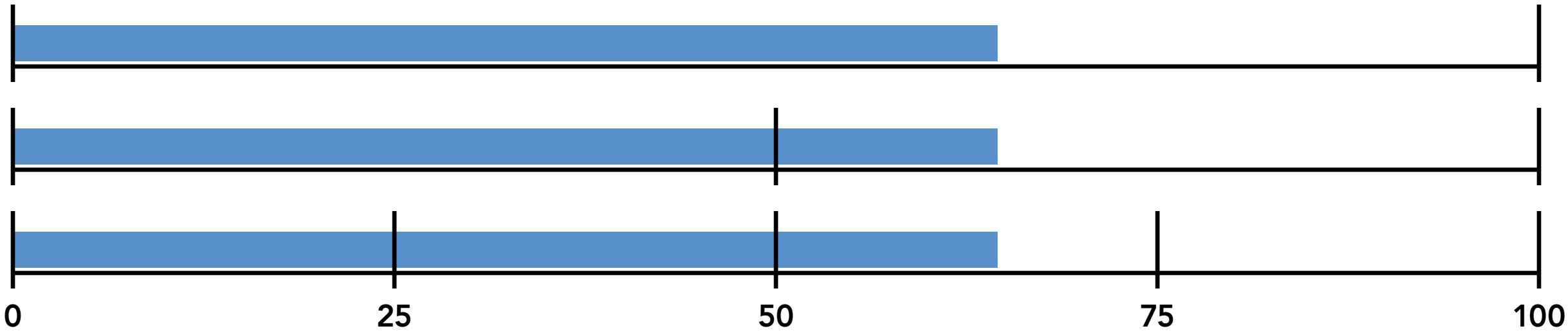
Other insights from perception

Reference lines can help…

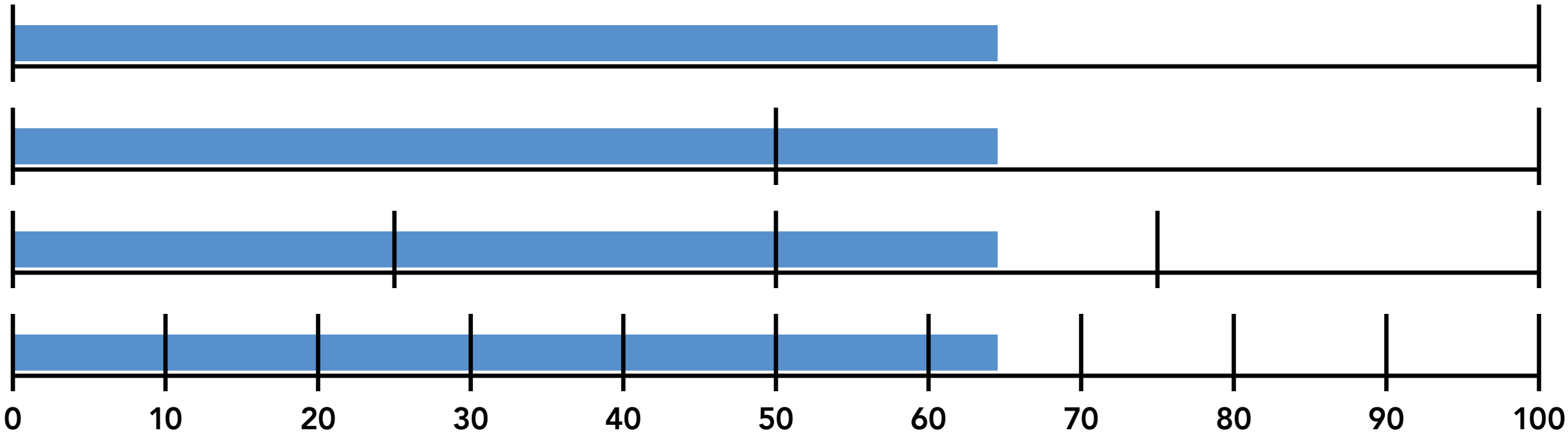
Reference lines can help…



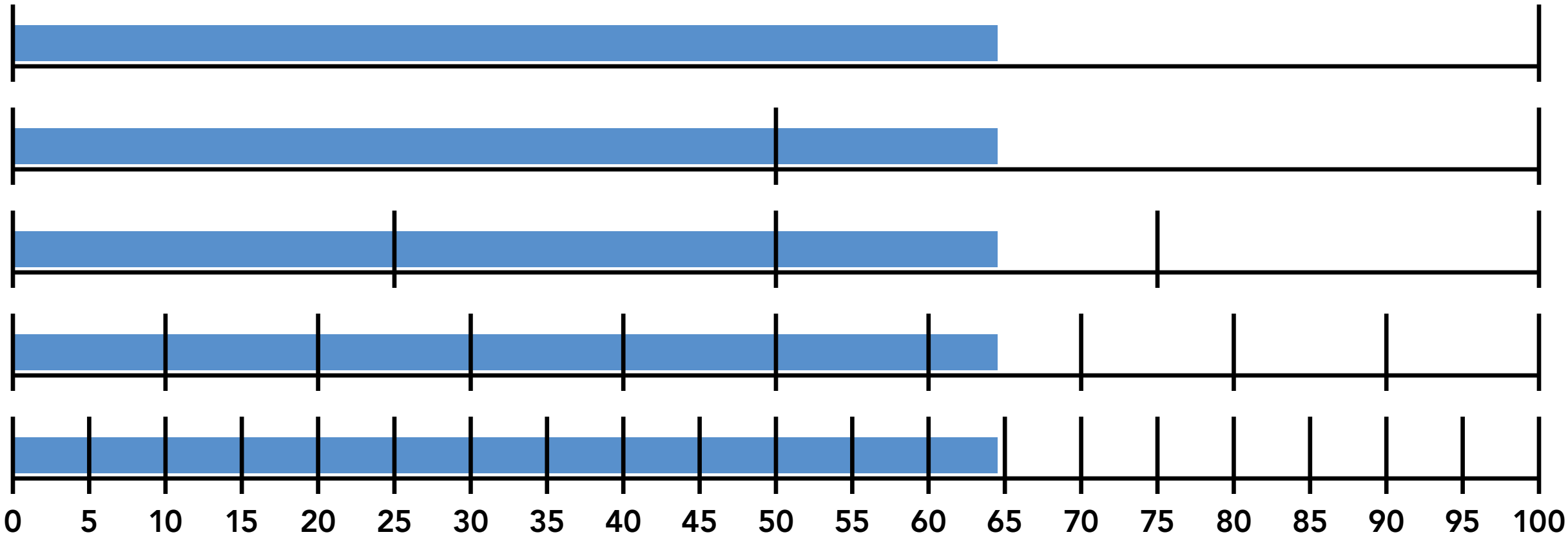
Reference lines can help...



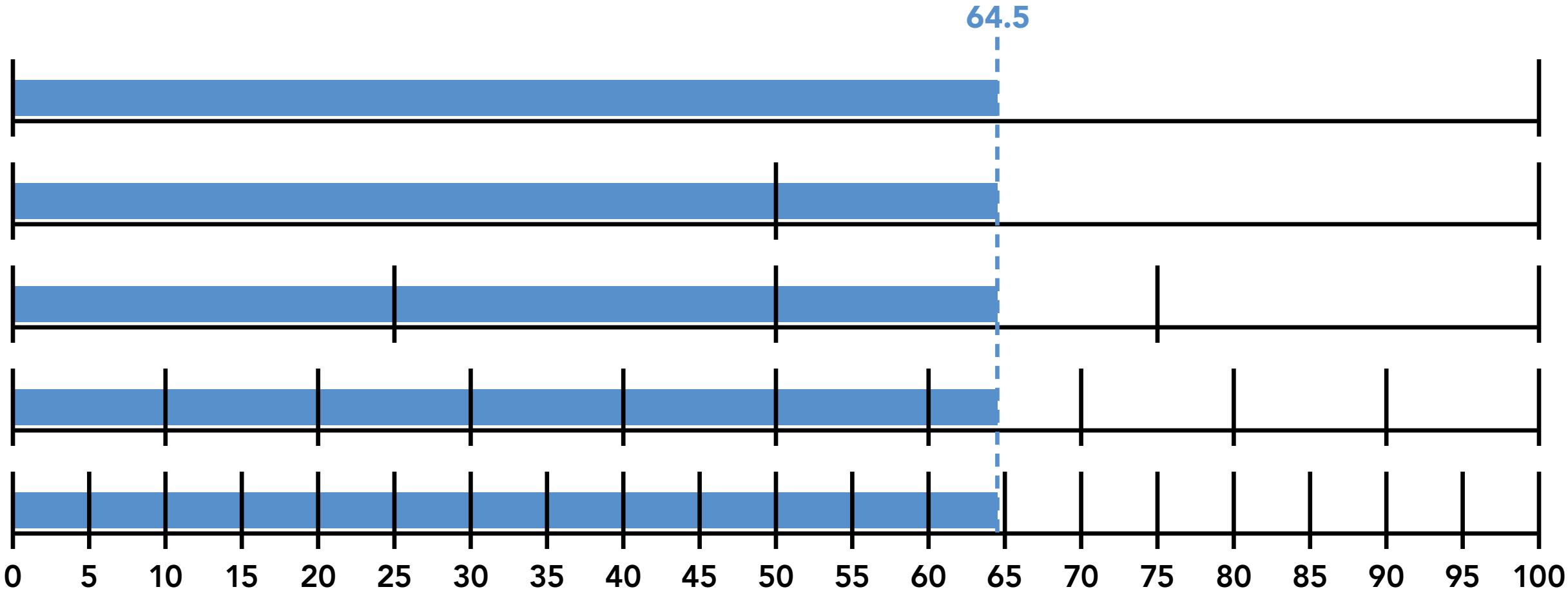
Reference lines can help…



Reference lines can help...



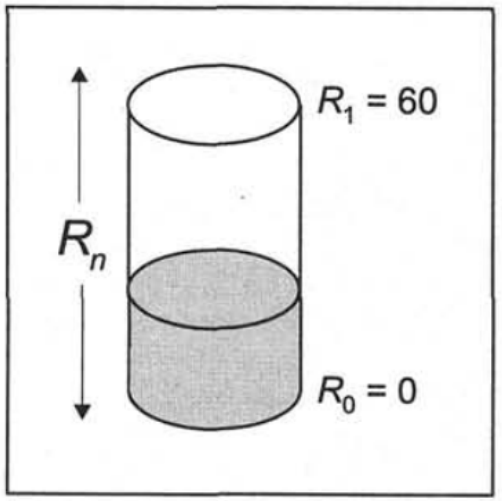
Reference lines can help...

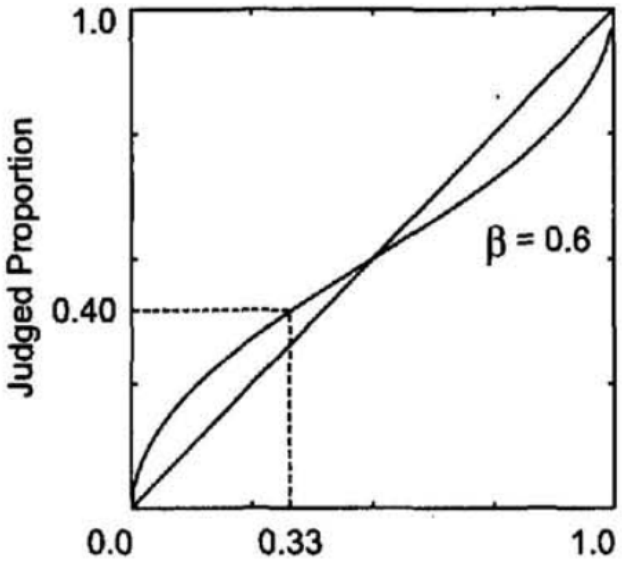


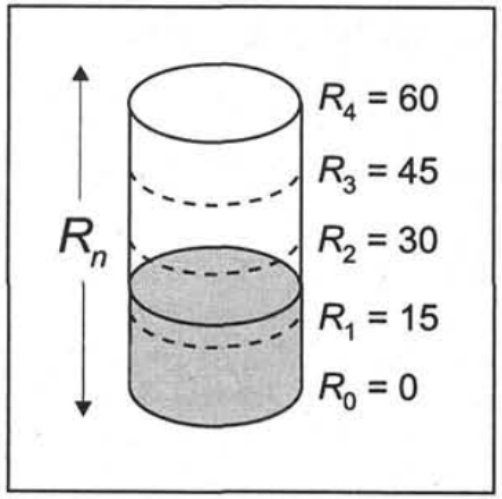
Reference lines

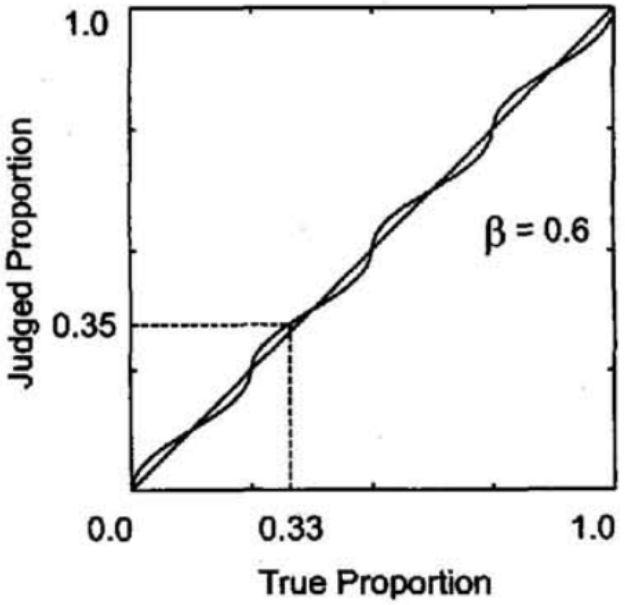
Induce bias...

...but can be used to decrease error

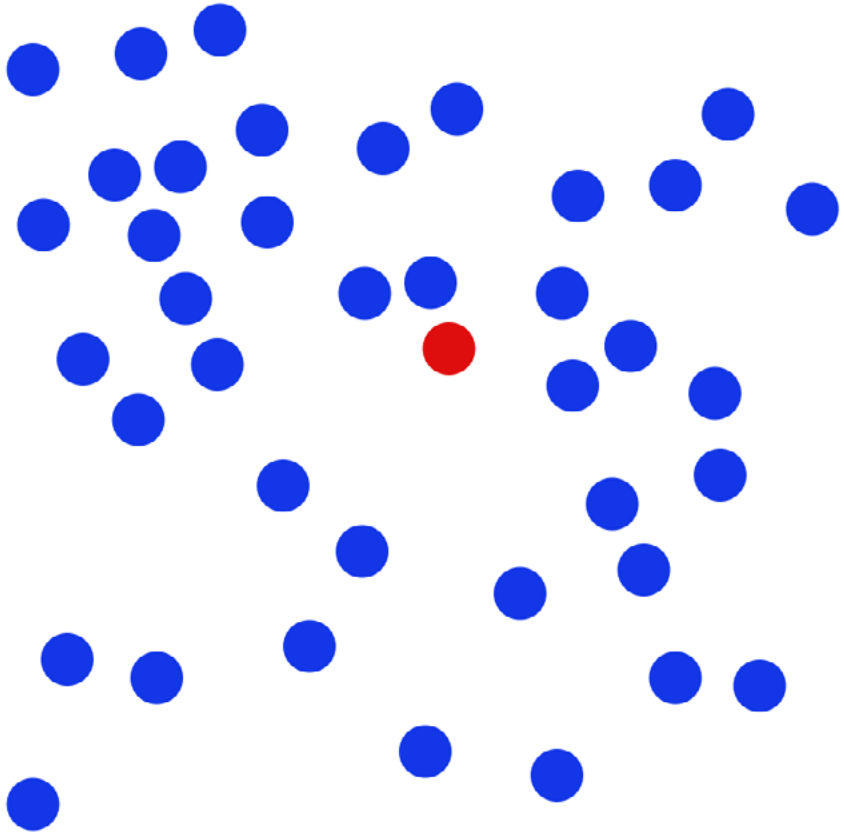






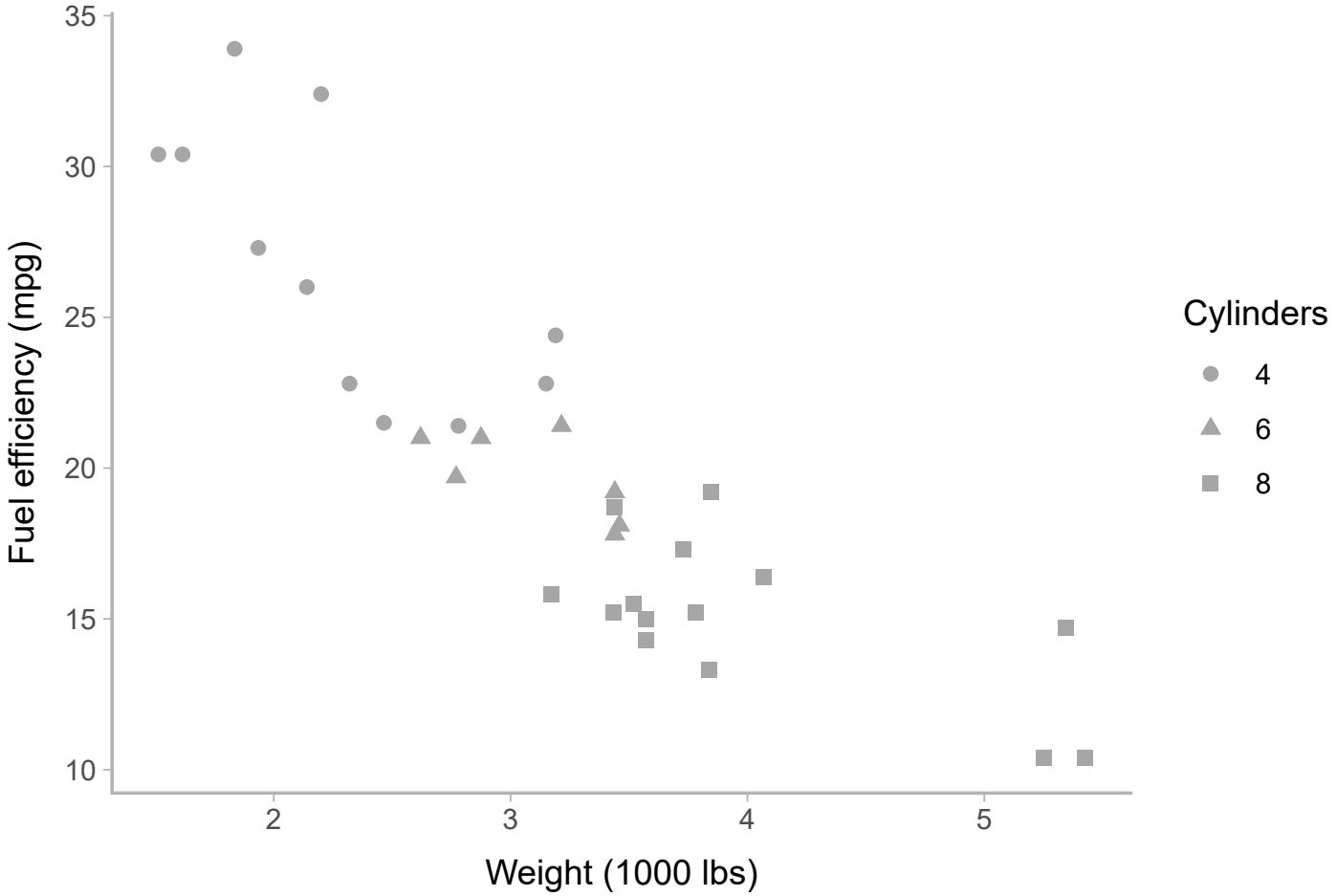


Popout and preattentiveness



https://www.csc2.ncsu.edu/faculty/healey/PP/

Popout and preattentiveness



Preattentiveness

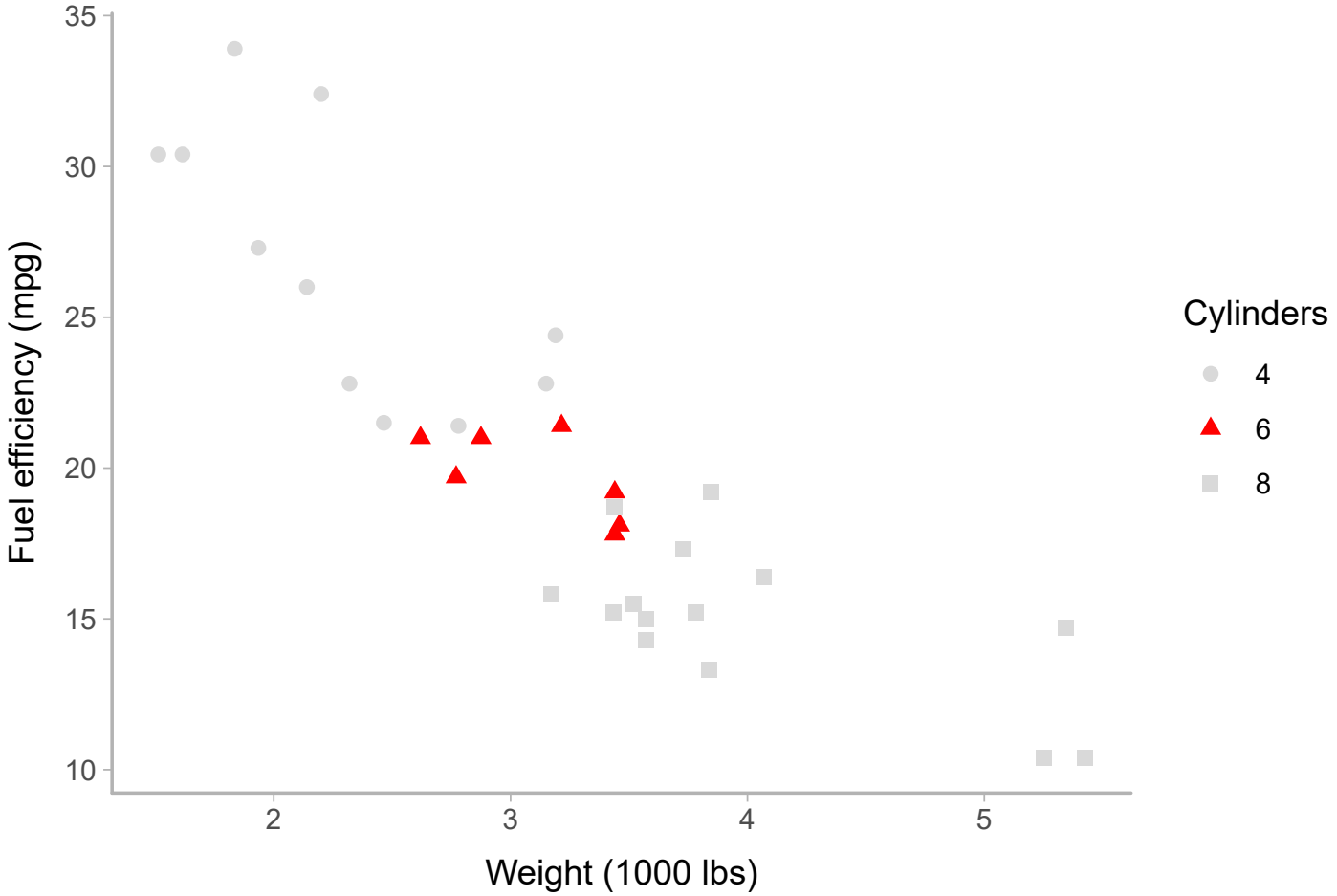
-> popout

-> layering

What do people see first?

What can people see separately?

Popout and preattentiveness



Preattentiveness

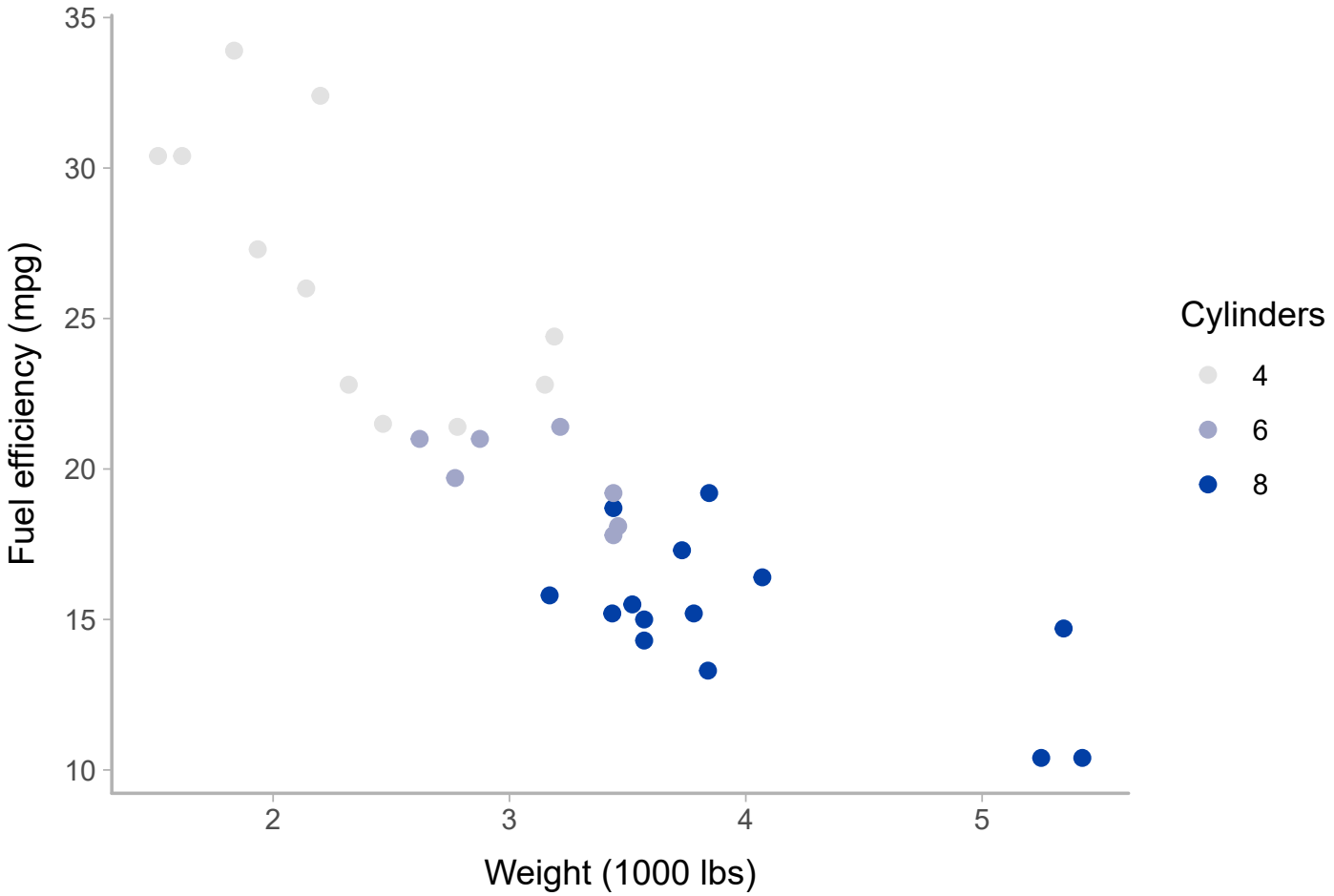
-> popout

-> layering

What do people see first?

What can people see separately?

Popout and preattentiveness



Preattentiveness

-> popout

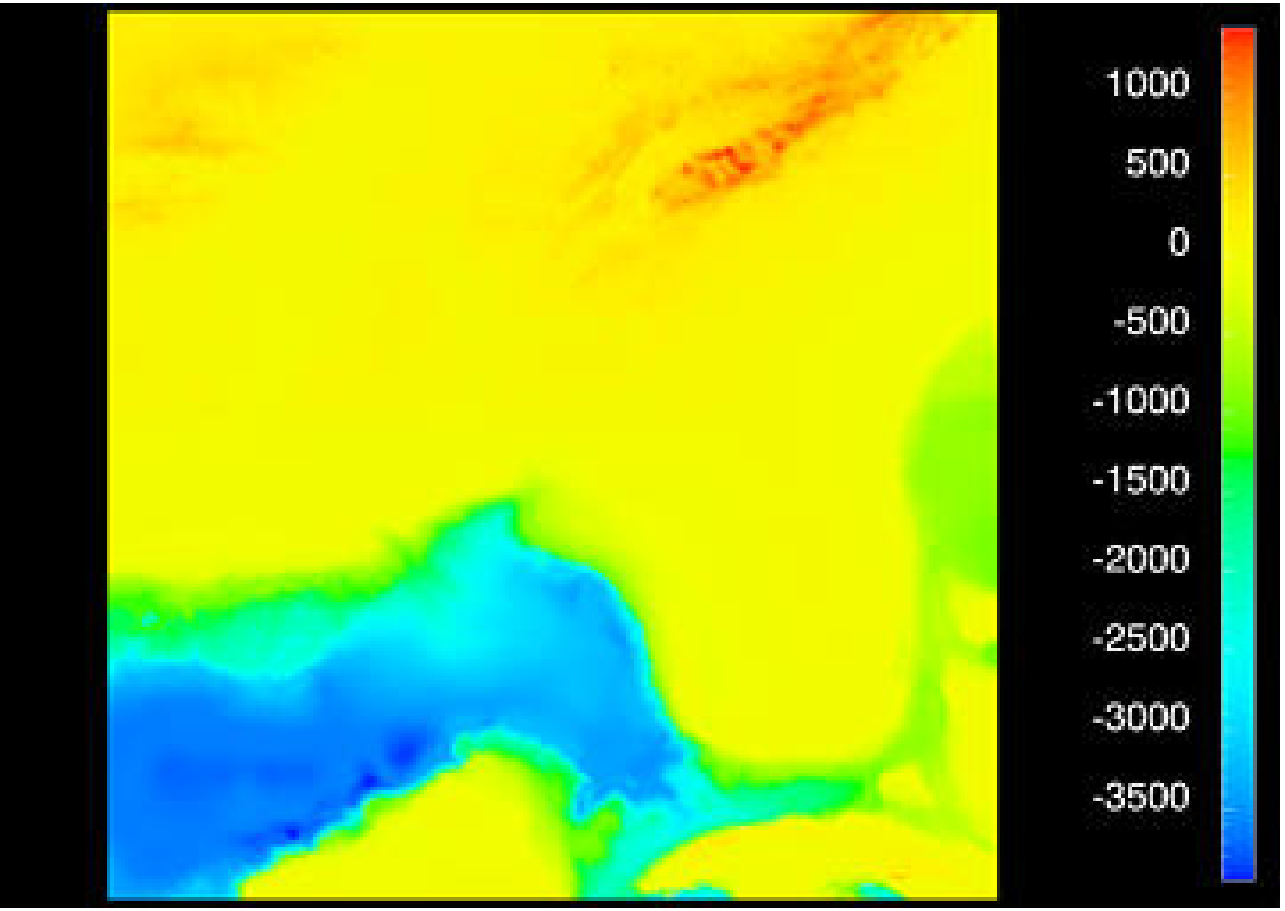
-> layering

What do people see first?

What can people see separately?

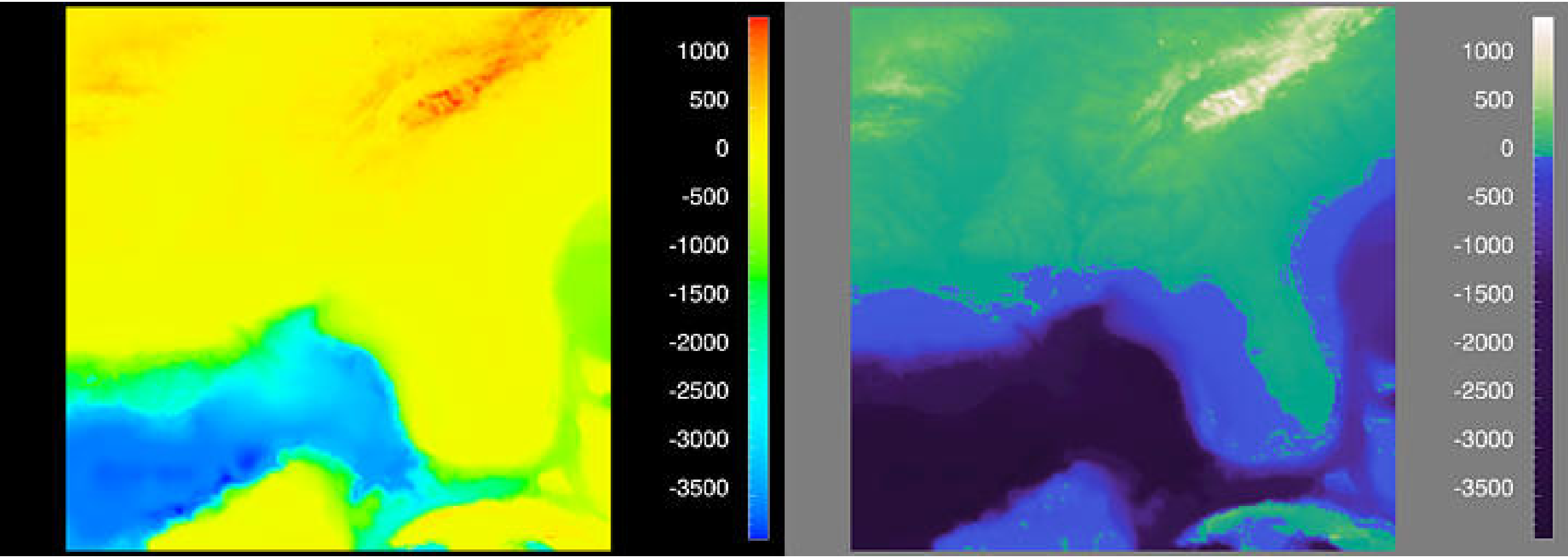
Color

Sequential / diverging data



[http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Sequential / diverging data



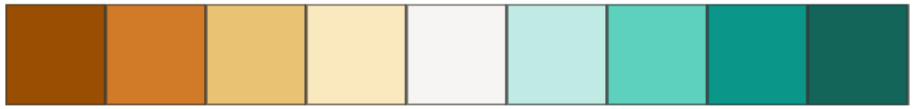
[http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Sequential / diverging scales

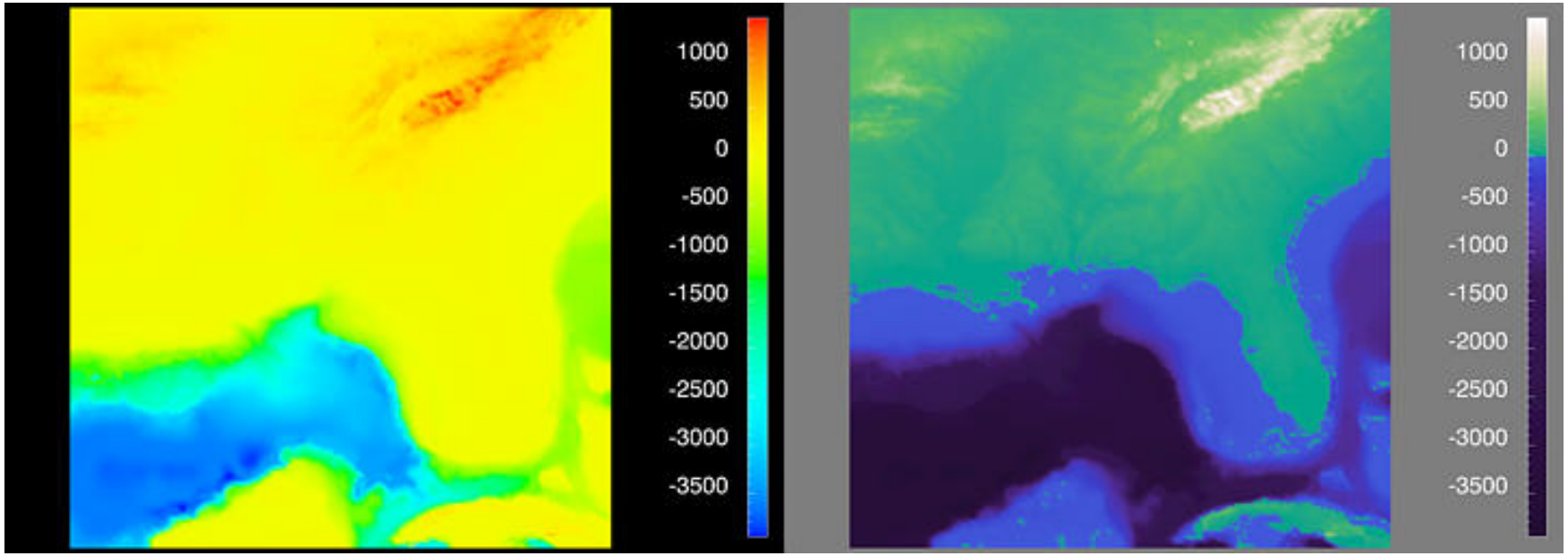
Ordered / quantitative data may
be sequential or diverging

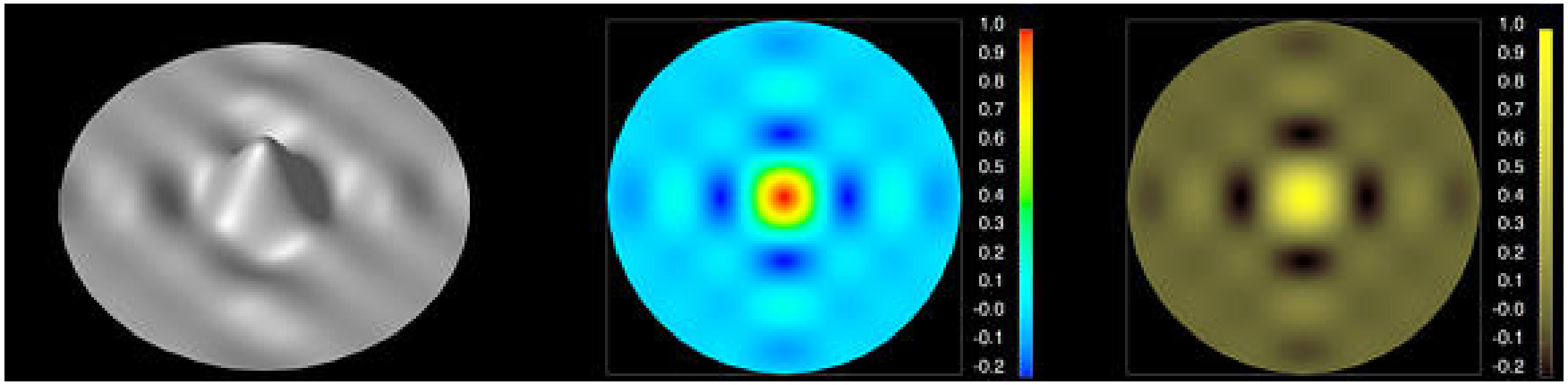
This impacts encoding choice, for example:

Sequential color scale: 

Diverging color scale: 

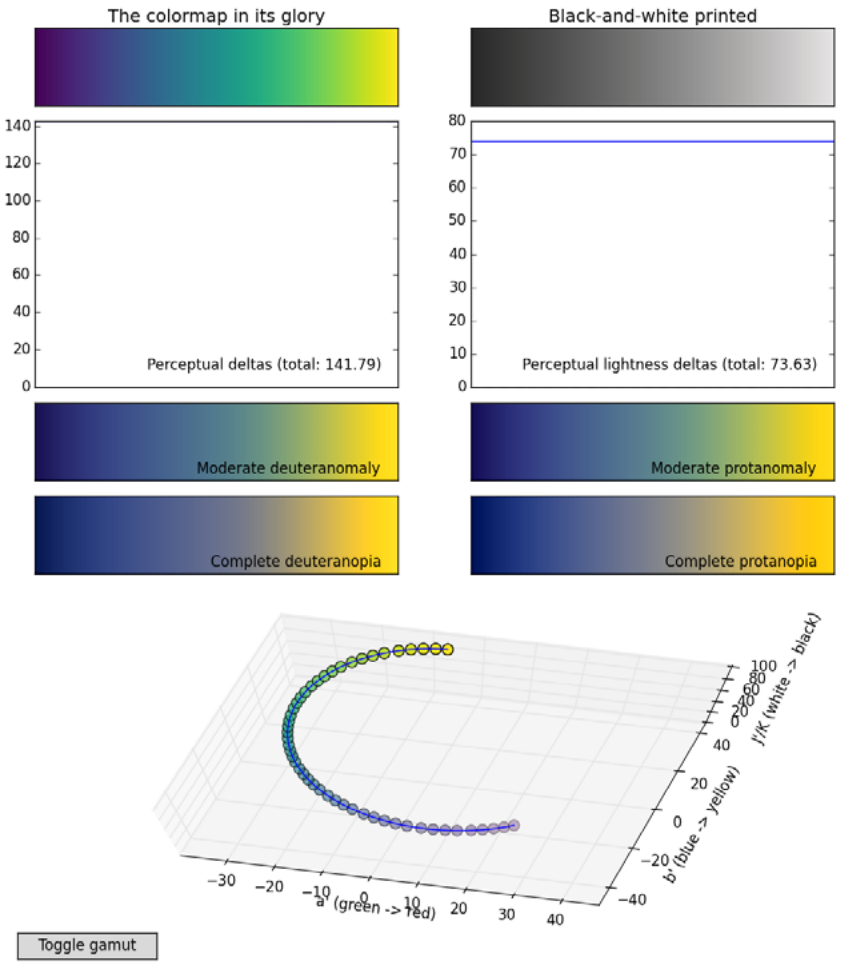
Use perceptually uniform colormaps

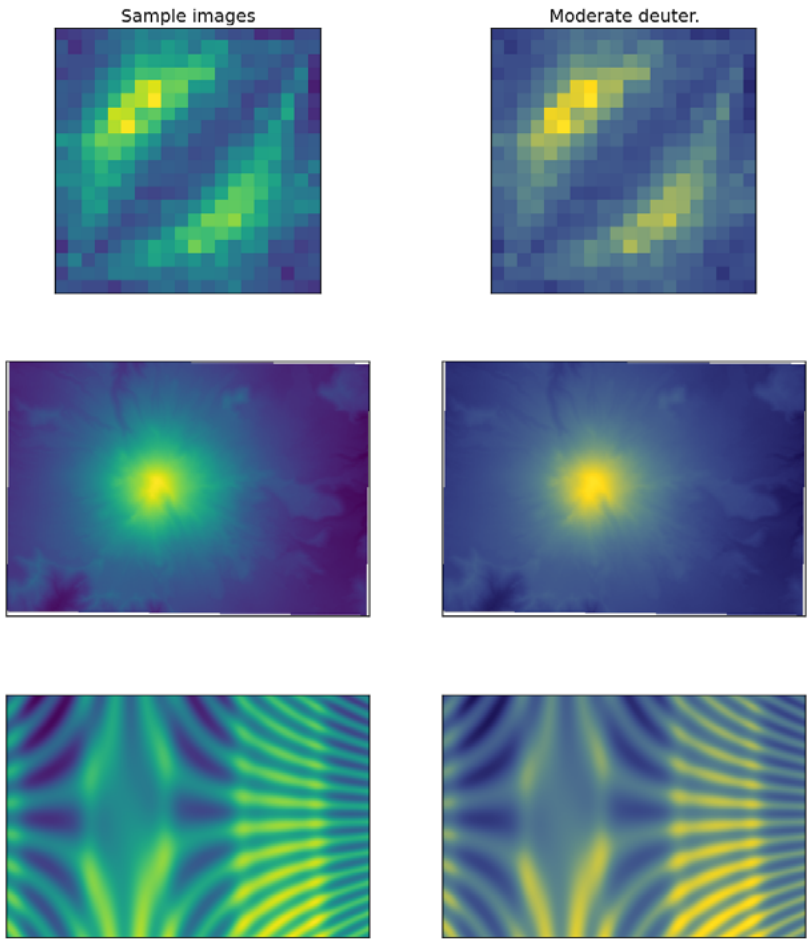




[Bernice E Rogowitz and Lloyd A Treinish. 1993. Why Should Engineers and Scientists Be Worried About Color? IBM Thomas J. Watson Research Center. Retrieved May 11, 2013 from http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

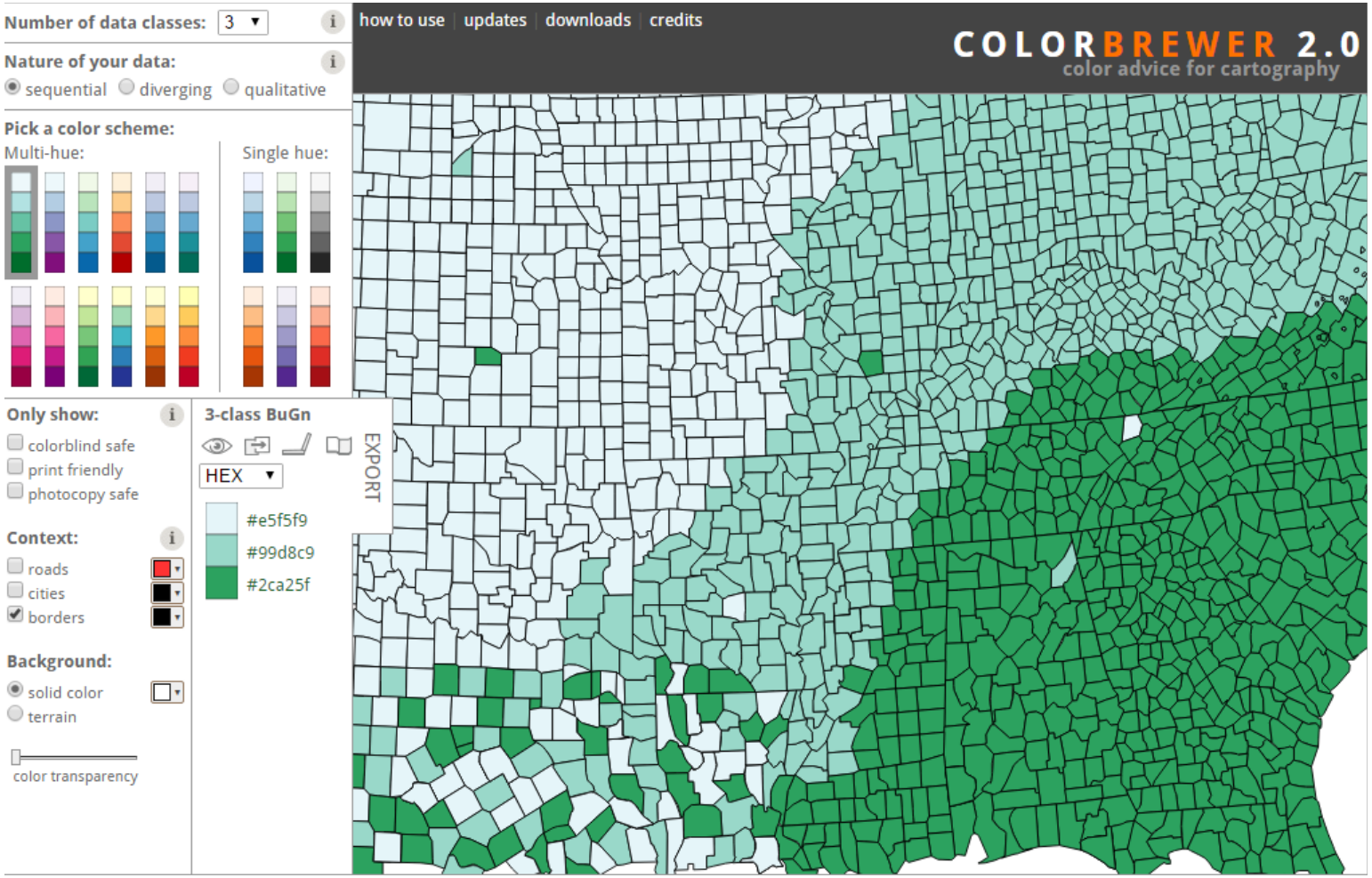
For continuous color maps, Viridis (and co)...





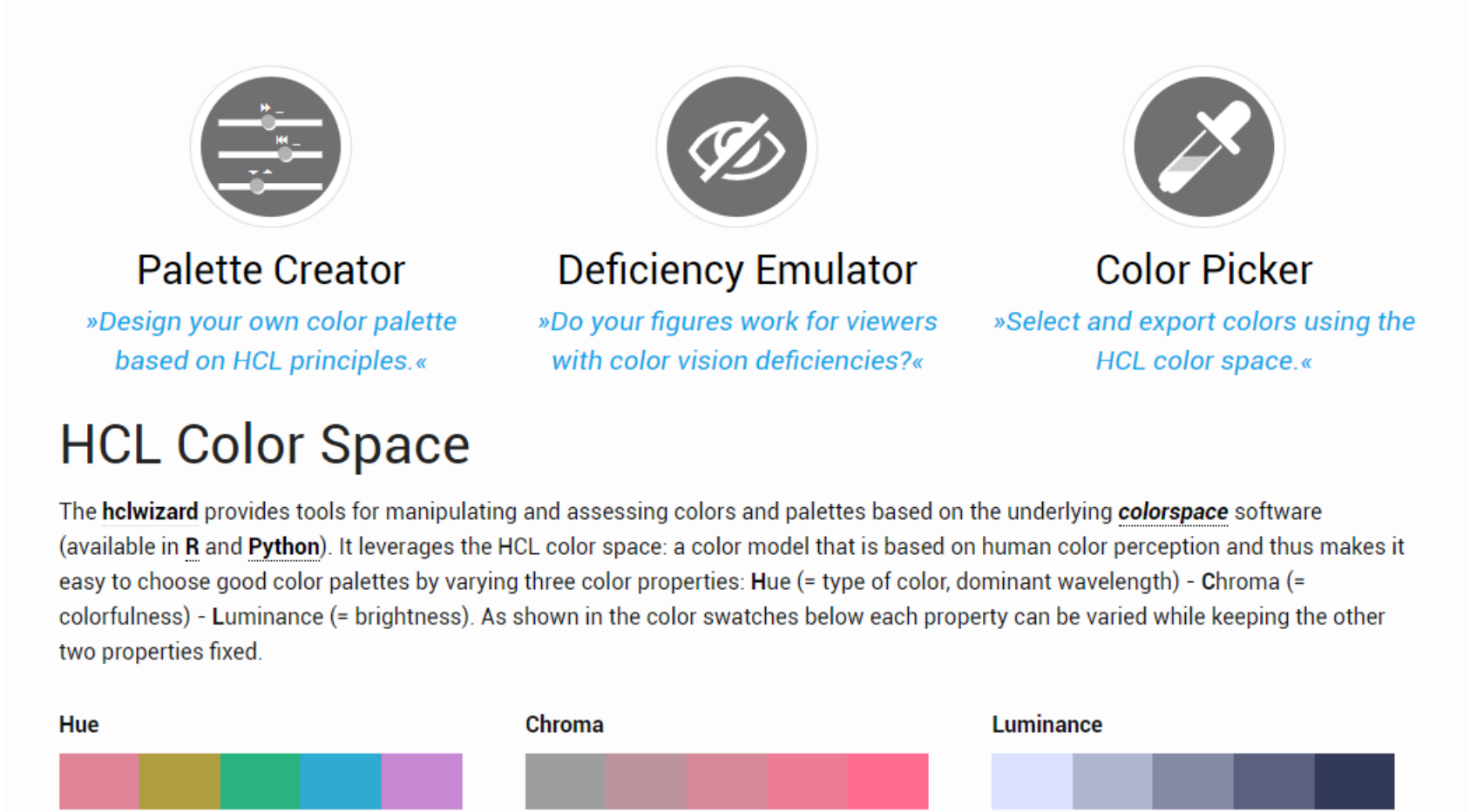
[http://bids.github.io/colormap]

For discrete colormaps, Color Brewer…



[http://colorbrewer2.org]

For more, hclwizard / colorspace R package



[http://hclwizard.org/]

Grammar of graphics + Perception
helps us design more effective charts

Grammar of Graphics + Perception

Think in data types, channels, and marks.

Helps you specify and design charts using perceptually effective channels.

Consider sequential / diverging nature of data.

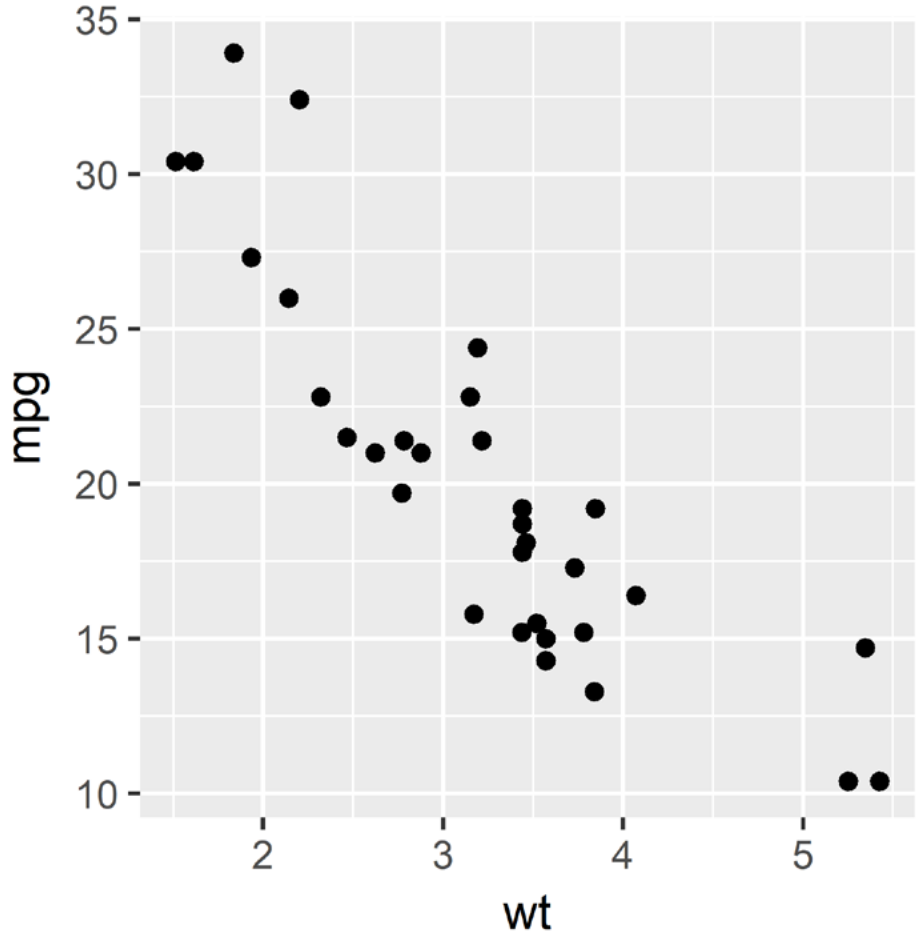
Use perceptually uniform colormaps.

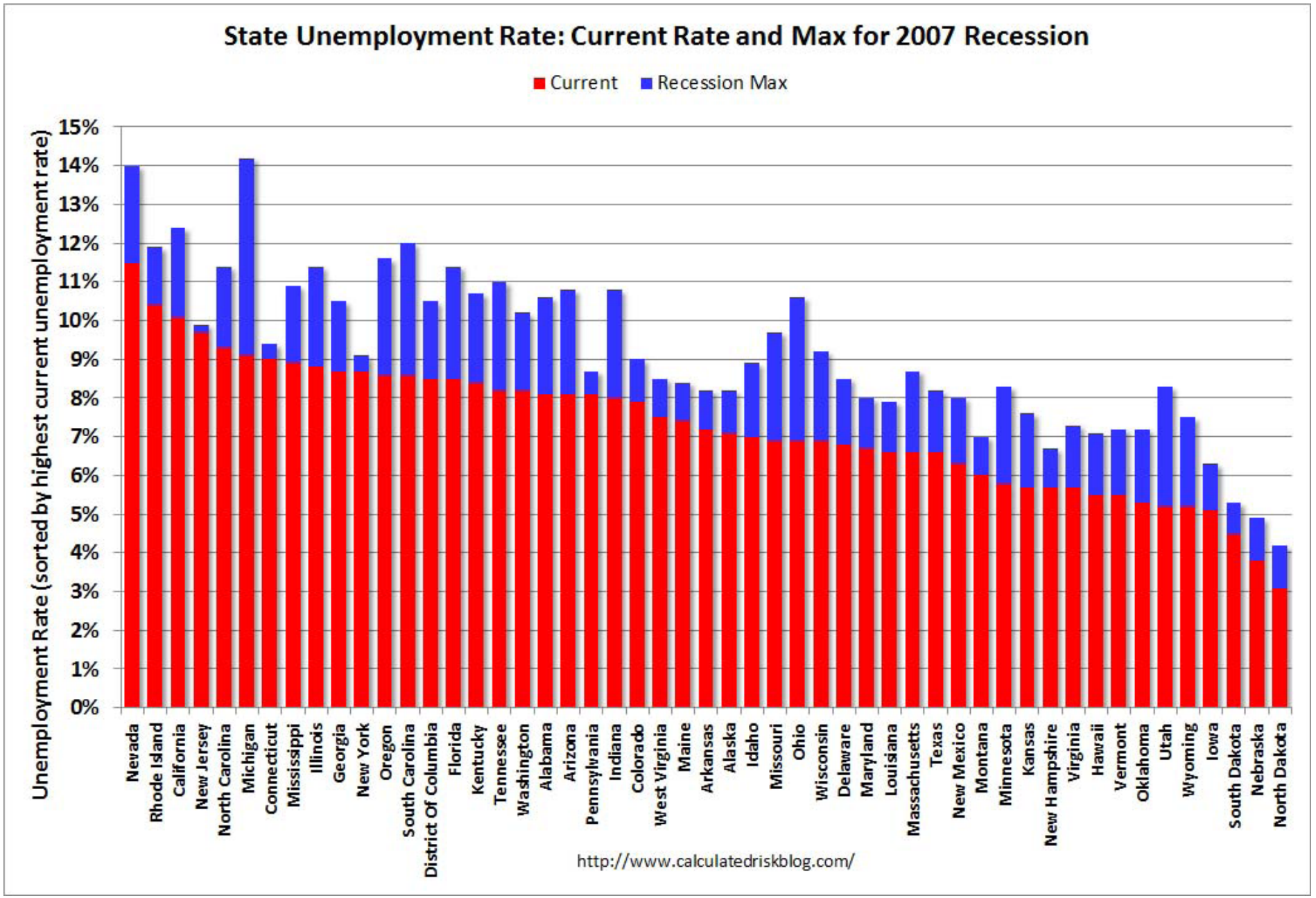
Questions about the grammar of graphics?

Grammar of graphics

(data types, channels, marks)

mpg: numeric
wt: numeric
wt -> x position
mpg -> y position
mark: point





Group activity

What are the variables / types?

Channels / encodings?

Marks?

Is this effective?

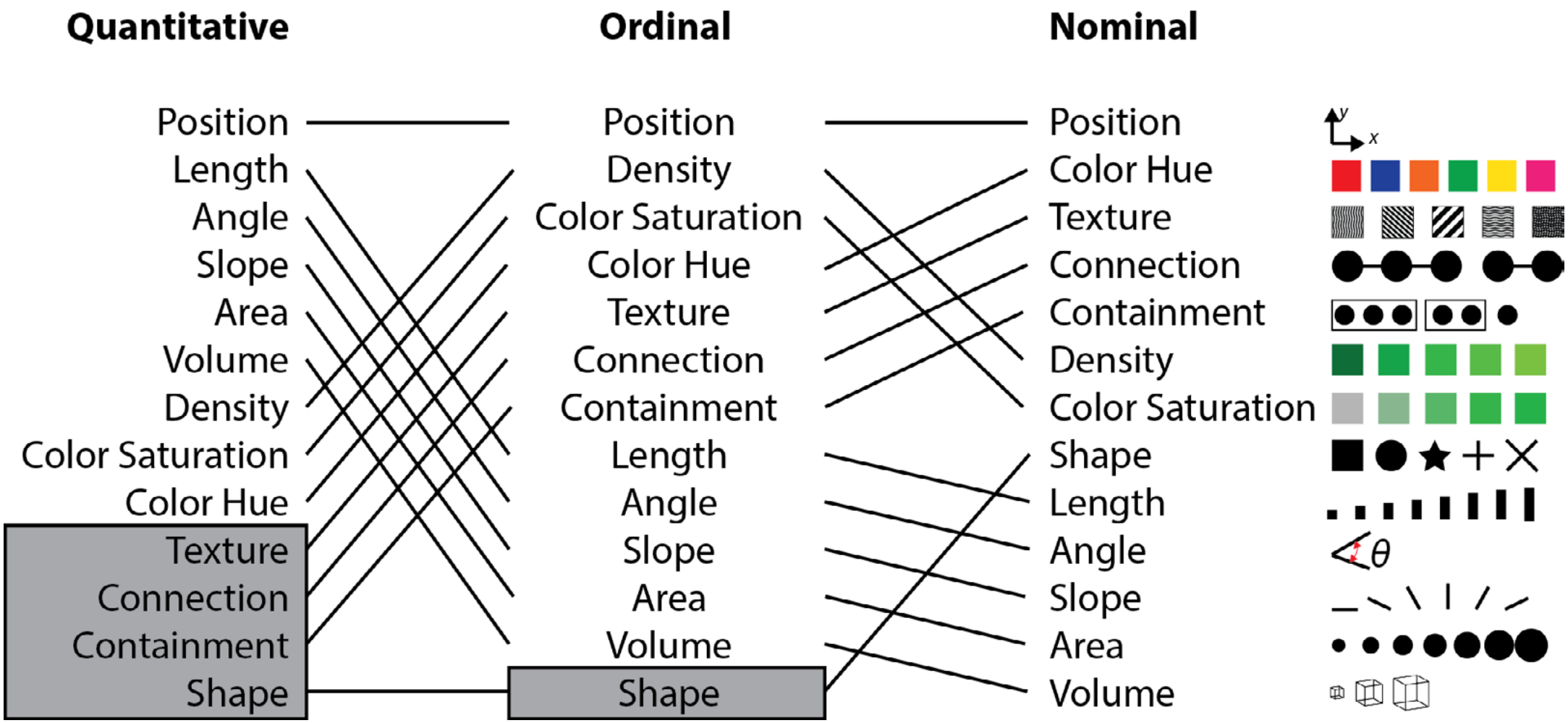
Design guidelines

Some rough design guidelines*

1. Match effectiveness with importance
2. Avoid ambiguity
3. Locality is king / eyes beat memory
4. Establish viewing order
5. Layer, layer, layer
6. When in doubt, grid
7. Treat visual attributes like adjectives

* These guidelines are drawn largely from my experience + personal preferences + the literature. Design is messy, these are not perfect, others will disagree with me, etc. *Caveat emptor.*

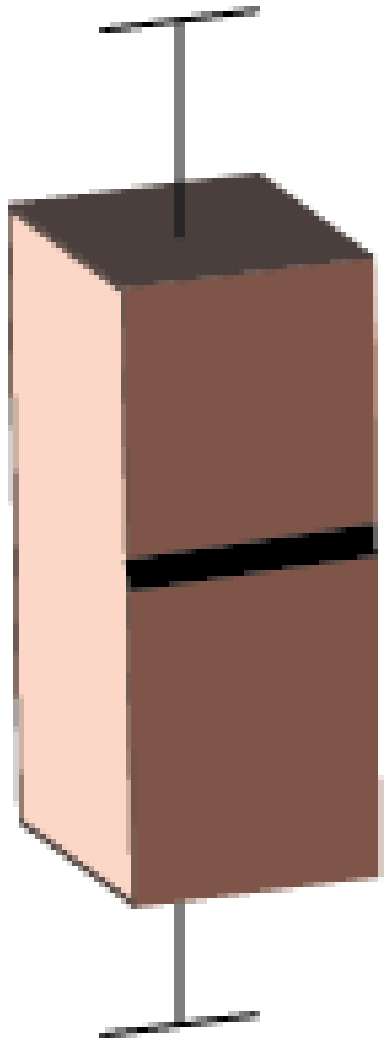
1. Match effectiveness with importance



2. Avoid ambiguity

Does the 3D mean anything here?

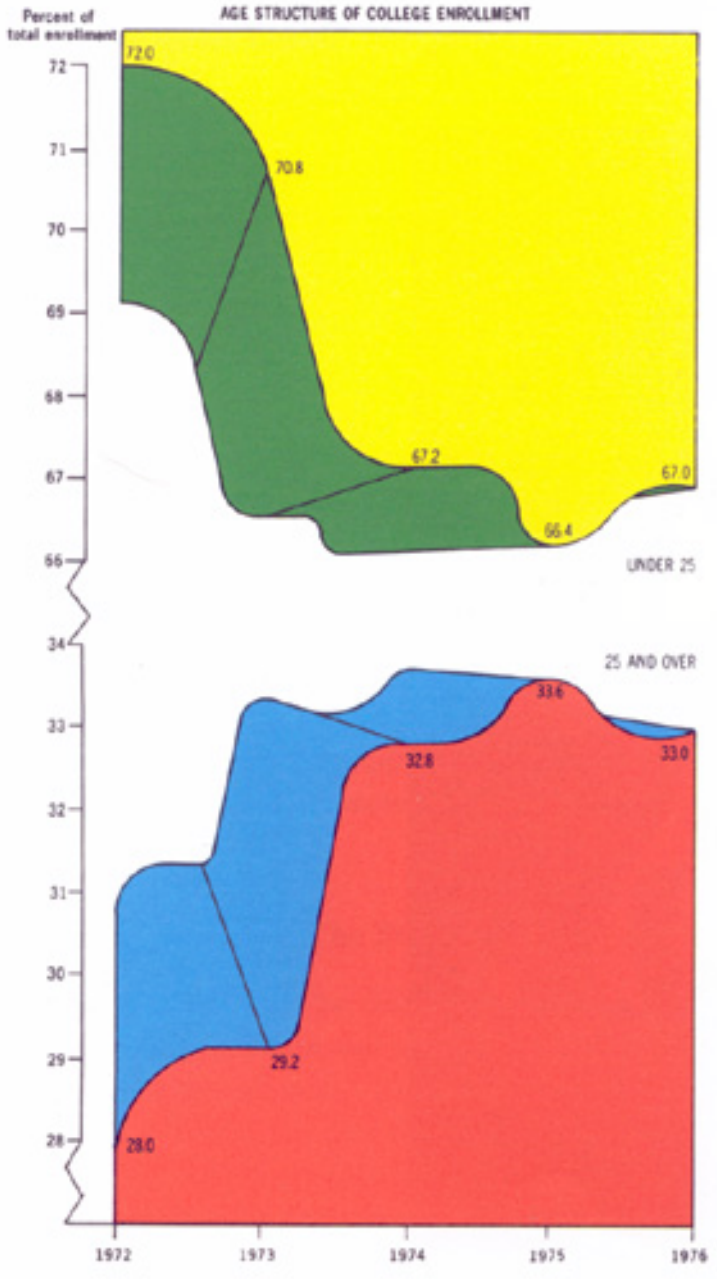
(Hint: No)



2. Avoid ambiguity

Marks should not have multiple reasonable interpretations

If it looks like it could come from data, it should come from data



3. Locality is king / eyes beat memory

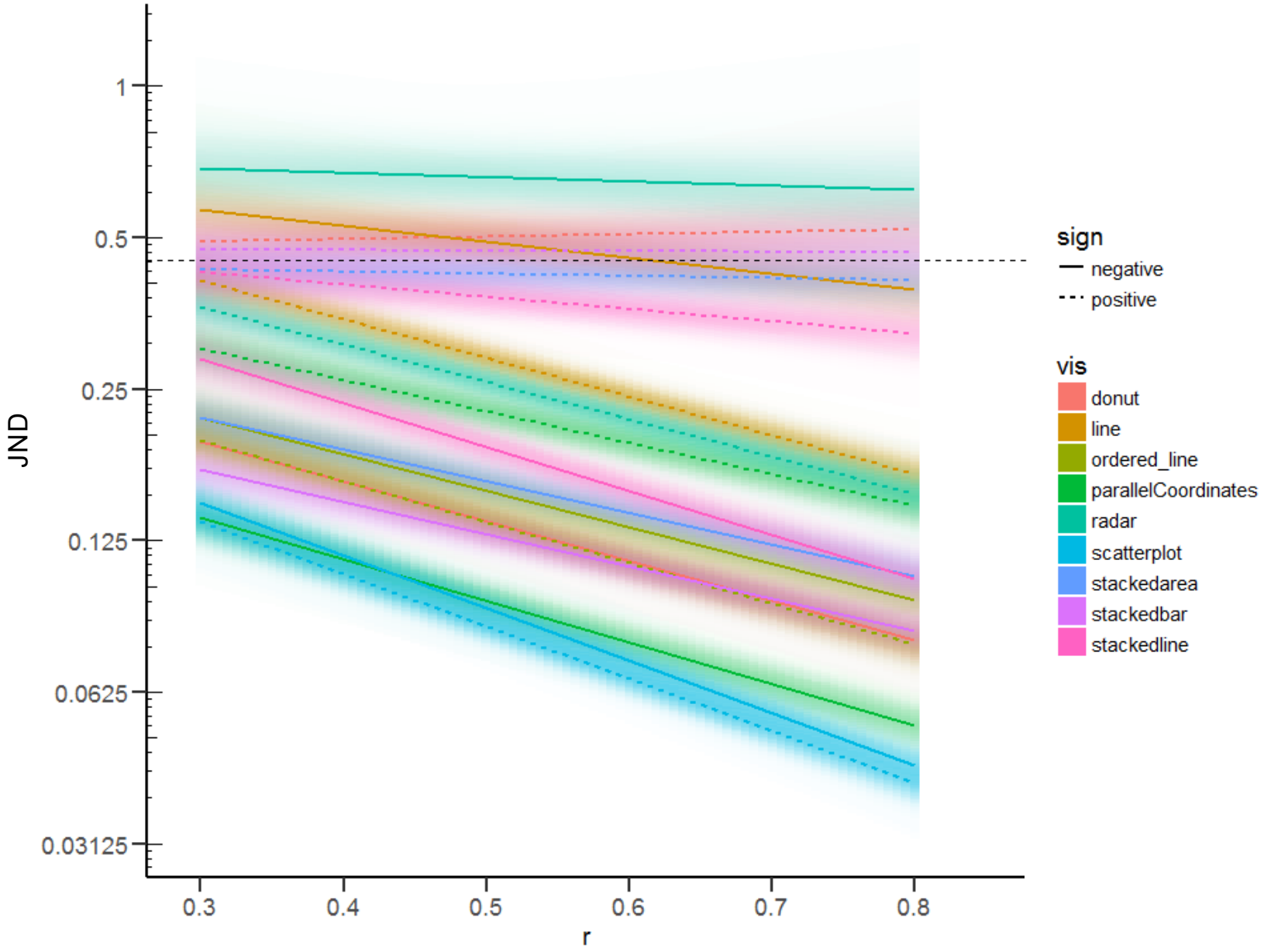
No:

Thing ⟷ Information I need to understand thing

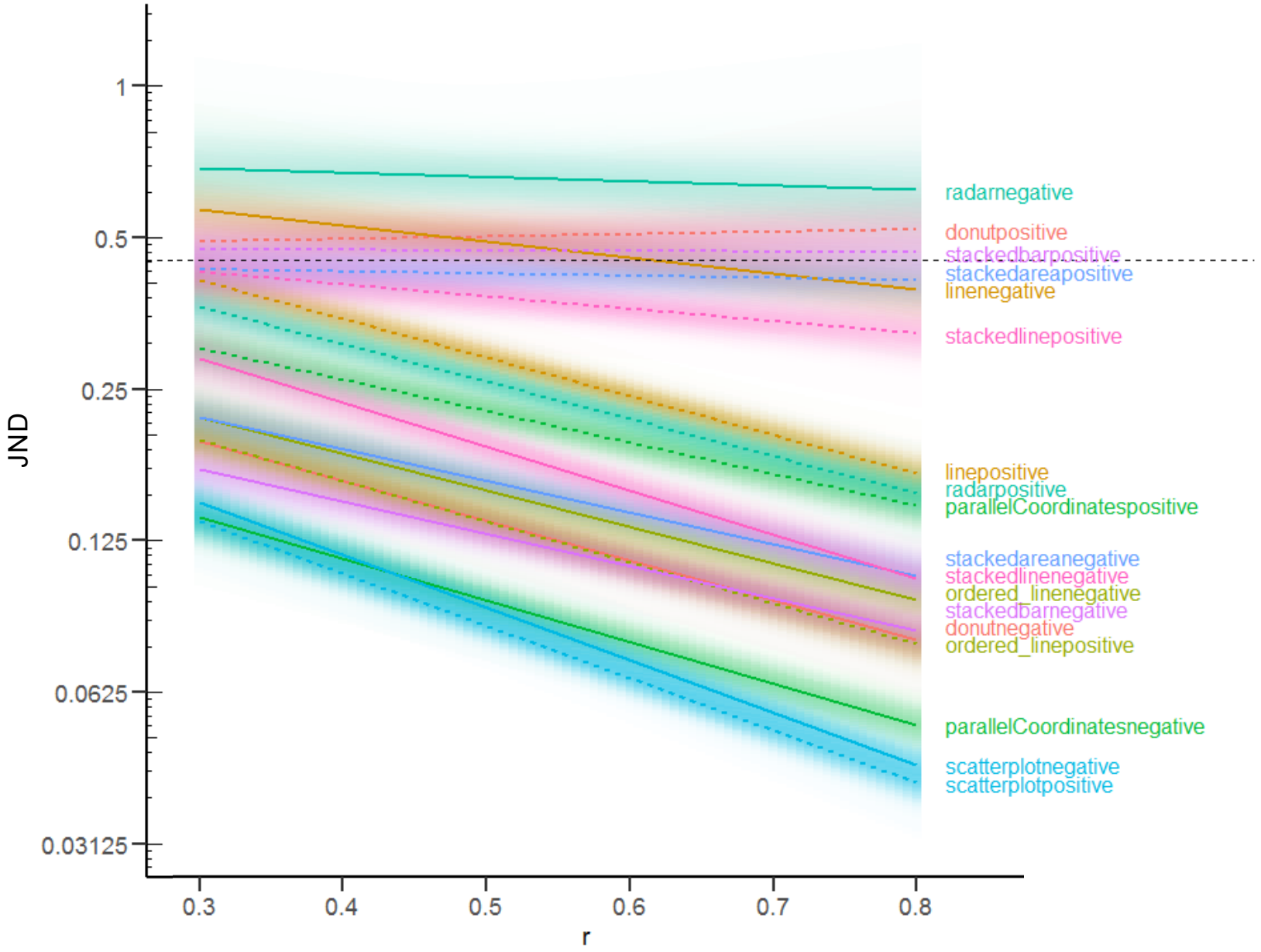
Yes:

Thing ↔ Information I need to understand thing

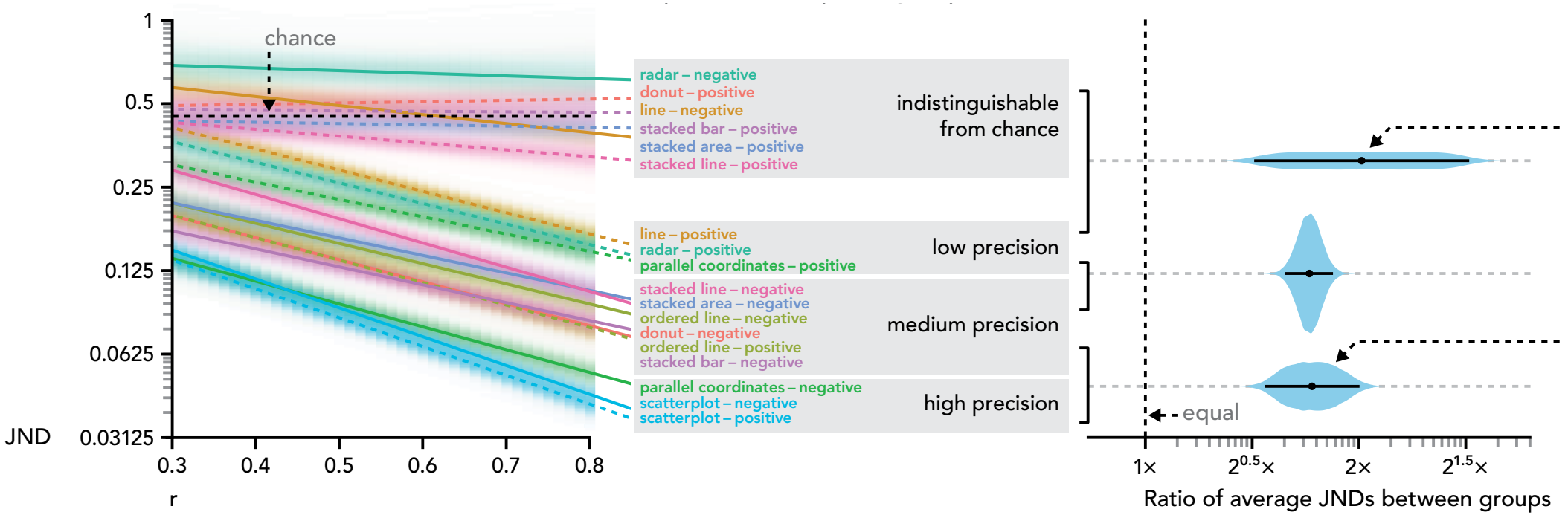
No


JND
1
0.5
0.25
0.125
0.0625
0.03125
0.3
0.4
0.5
0.6
0.7
0.8
r
sign
negative
positive
vis
donut
line
ordered_line
parallelCoordinates
radar
scatterplot
stackedarea
stackedbar
stackedline

Yes

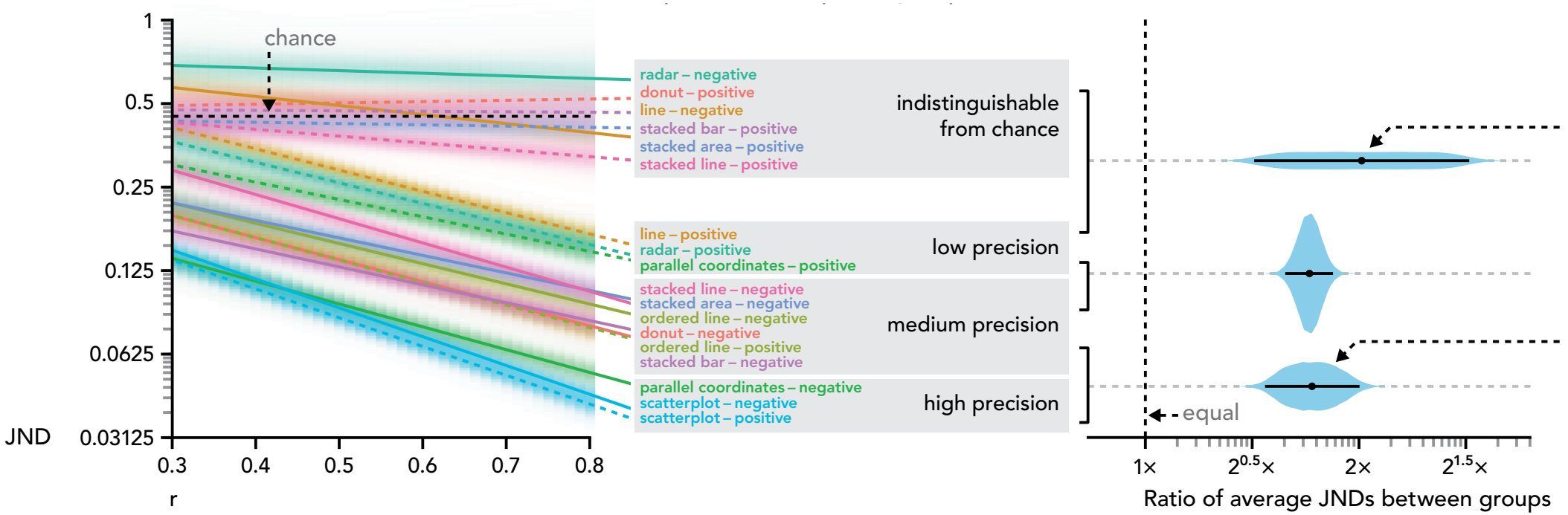

JND
1
0.5
0.25
0.125
0.0625
0.03125
0.3
0.4
0.5
0.6
0.7
0.8
r
radarnegative
donutpositive
stackedbarpositive
stackedareapositive
linenegative
stackedlinepositive
linepositive
radarpositive
parallelCoordinatespositive
stackedareanegative
stackedlinenegative
ordered_linenegative
stackedbarnegative
donutnegative
ordered_linepositive
parallelCoordinatesnegative
scatterplotnegative
scatterplotpositive

No



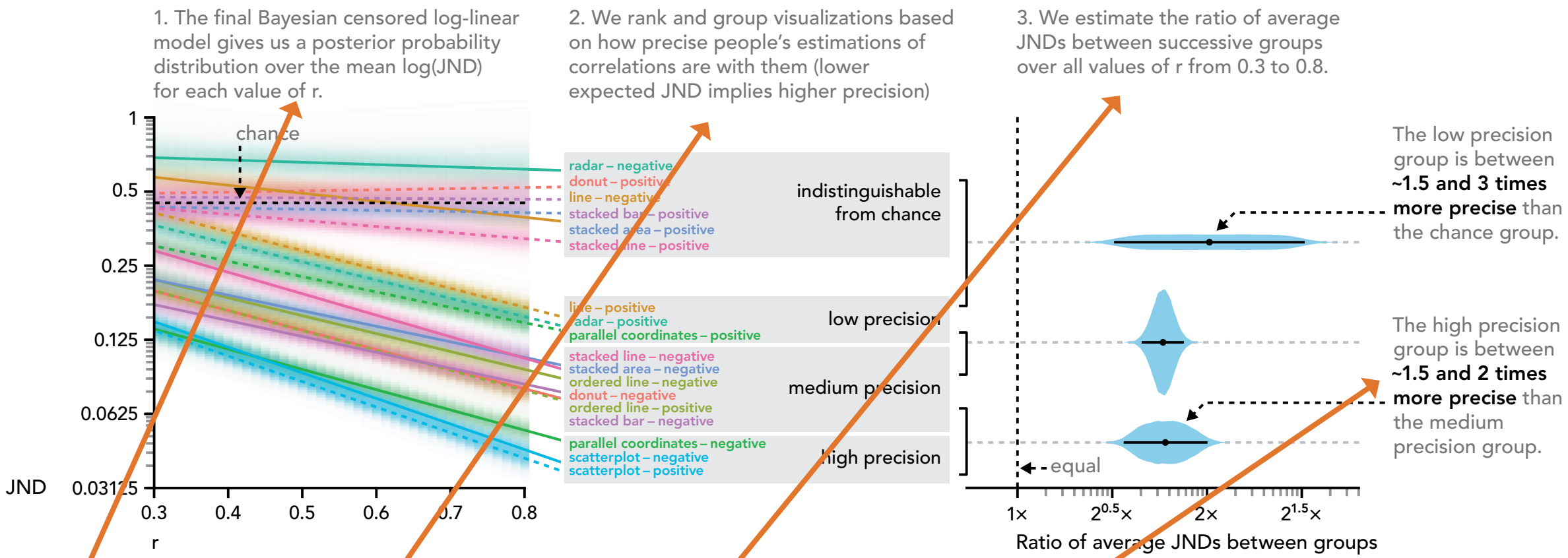
The left panel shows the Bayesian censored log-linear model, which gives us a posterior probability distribution over the mean log(JND) for each value of r. In the center panel we rank and group visualizations based on how precise estimations of correlations are with them (lower expected JND implies higher precision). In the right panel we estimate the ratio of average JNDs between succesive groups over all values of r from 0.3 to 0.8. The low precision group is between ~1.5 and 3 times more precise than the chance group. The high precision group is between ~1.5 and 2 times more precise than the medium precision group.

No



The left panel shows the Bayesian censored log linear model, which gives us a posterior probability distribution over the mean log(JND) for each value of r. In the center panel we rank and group visualizations based on how precise estimations of correlations are with them (lower expected JND implies higher precision). In the right panel we estimate the ratio of average JNDs between succesive groups over all values of r from 0.3 to 0.8. The low precision group is between ~1.5 and 3 times more precise than the chance group. The high precision group is between ~1.5 and 2 times more precise than the medium precision group.

Yes



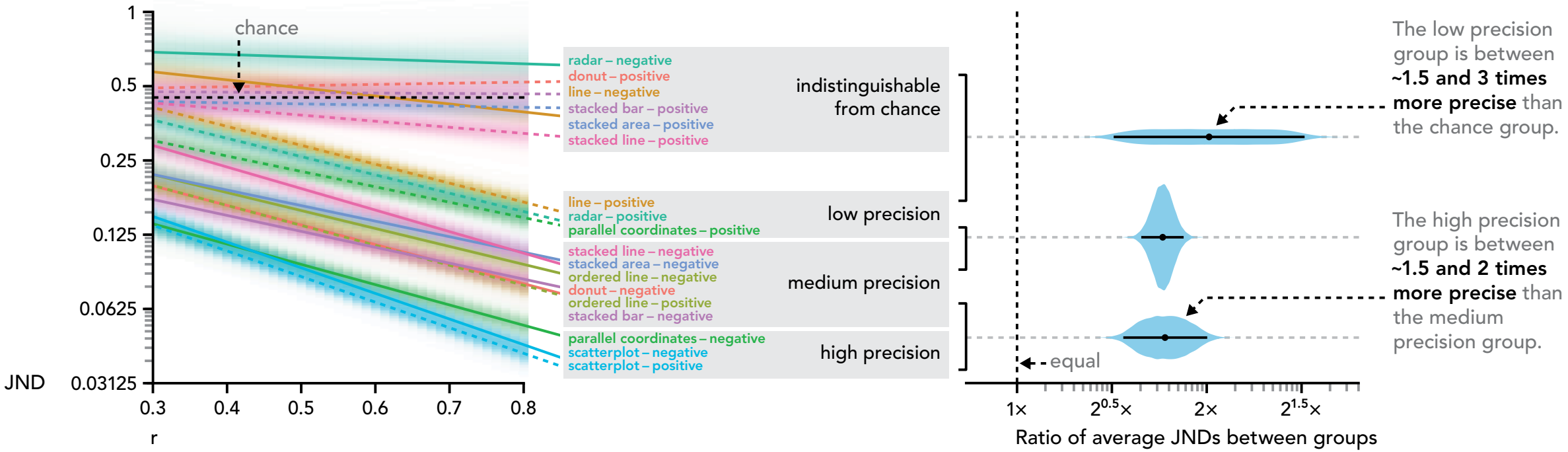
The left panel shows the Bayesian censored log-linear model, which gives us a posterior probability distribution over the mean log(JND) for each value of r. In the center panel we rank and group visualizations based on how precise estimations of correlations are with them (lower expected JND implies higher precision). In the right panel we estimate the ratio of average JNDs between succesive groups over all values of r from 0.3 to 0.8. The low precision group is between ~1.5 and 3 times more precise than the chance group. The high precision group is between ~1.5 and 2 times more precise than the medium precision group.

Yes

1. The final Bayesian censored log-linear model gives us a posterior probability distribution over the mean log(JND) for each value of r.

2. We rank and group visualizations based on how precise people's estimations of correlations are with them (lower expected JND implies higher precision)

3. We estimate the ratio of average JNDs between successive groups over all values of r from 0.3 to 0.8.

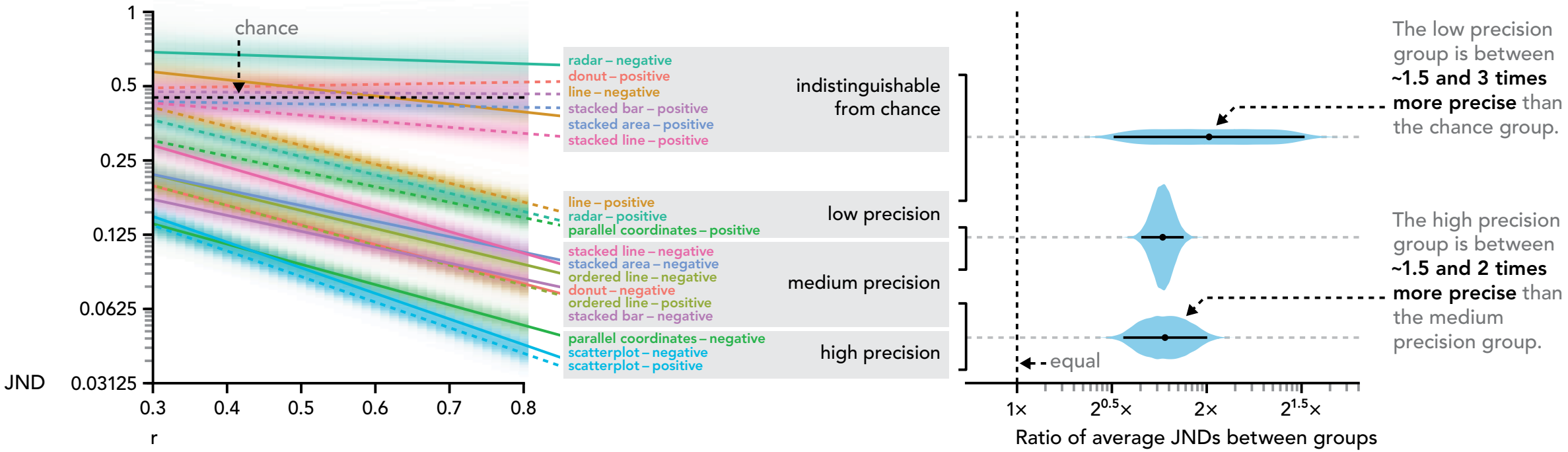


Yes

1. The final Bayesian censored log-linear model gives us a posterior probability distribution over the mean log(JND) for each value of r.

2. We rank and group visualizations based on how precise people's estimations of correlations are with them (lower expected JND implies higher precision)

3. We estimate the ratio of average JNDs between successive groups over all values of r from 0.3 to 0.8.



Count lookups!

4. Establish viewing order

Know where your audience will look first, second.

Think like a movie director. Are you telling a story?

https://www.youtube.com/watch?v=v4seDVfgwOg

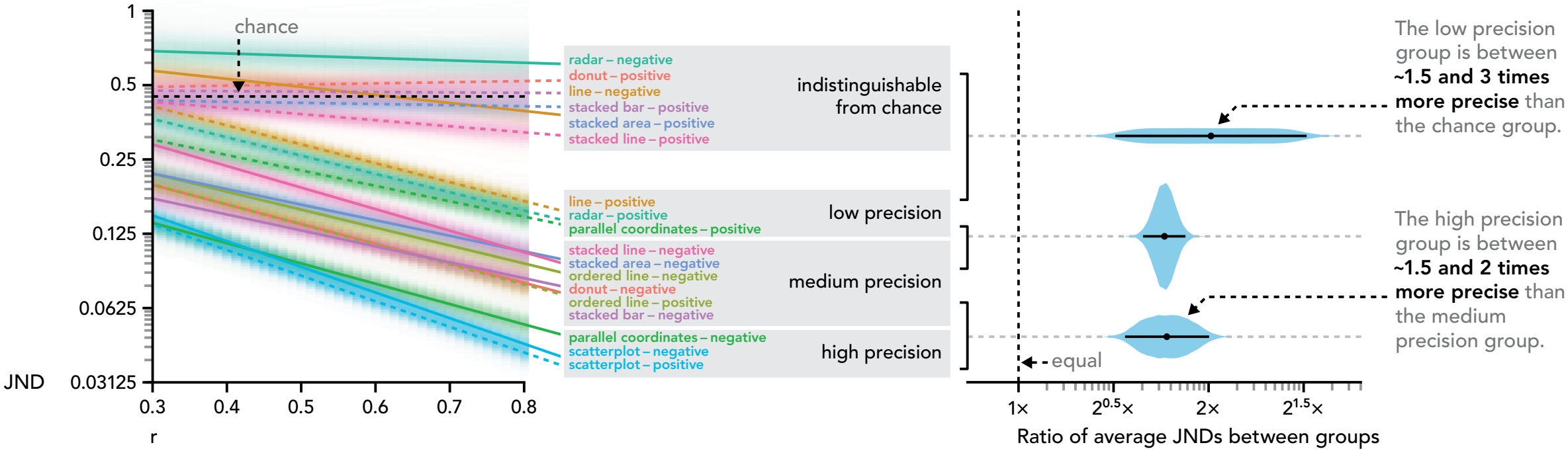
4. Establish viewing order

Can be as simple as some numbers...

1. The final Bayesian censored log-linear model gives us a posterior probability distribution over the mean log(JND) for each value of r.

2. We rank and group visualizations based on how precise people's estimations of correlations are with them (lower expected JND implies higher precision)

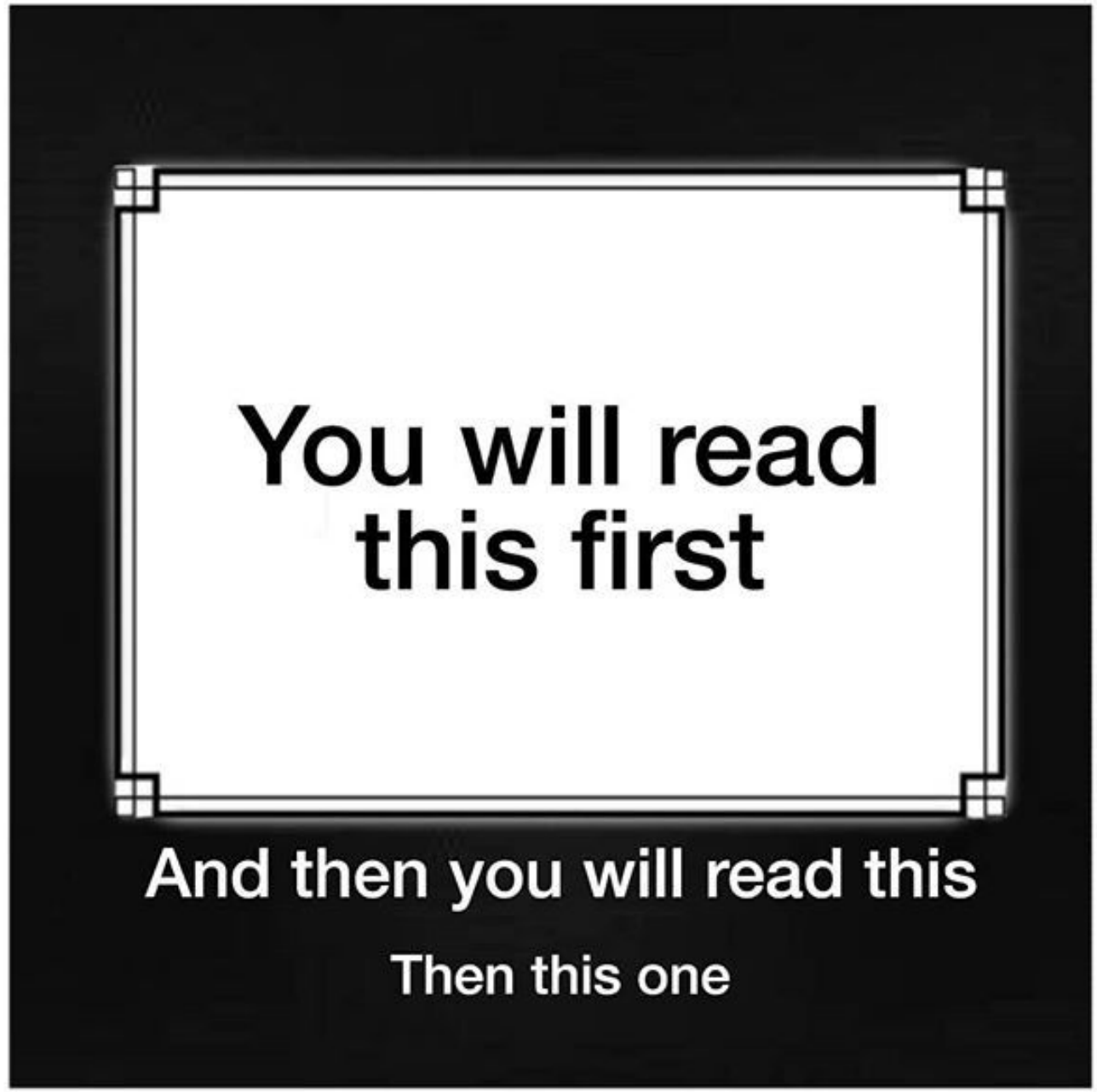
3. We estimate the ratio of average JNDs between successive groups over all values of r from 0.3 to 0.8.



4. Establish viewing order

Or more complex,
relying on salience,
other visual cues,
viewer expectations
(maybe) ...

And you will read this at the end

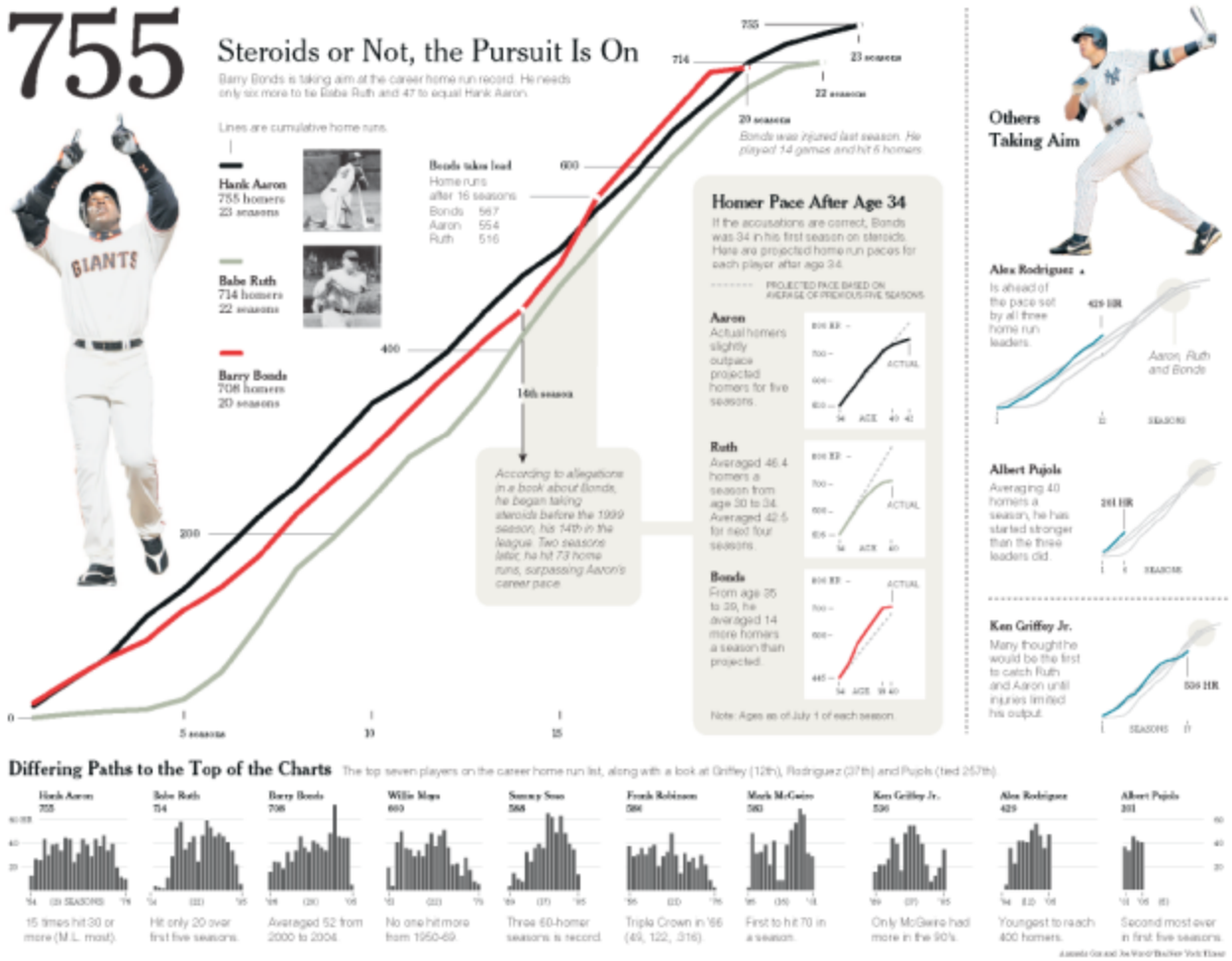


4. Establish viewing order

Or more complex, relying on salience, other visual cues, viewer expectations (maybe) ...

4. Establish viewing order

Or more complex, relying on salience, other visual cues, viewer expectations (maybe) …

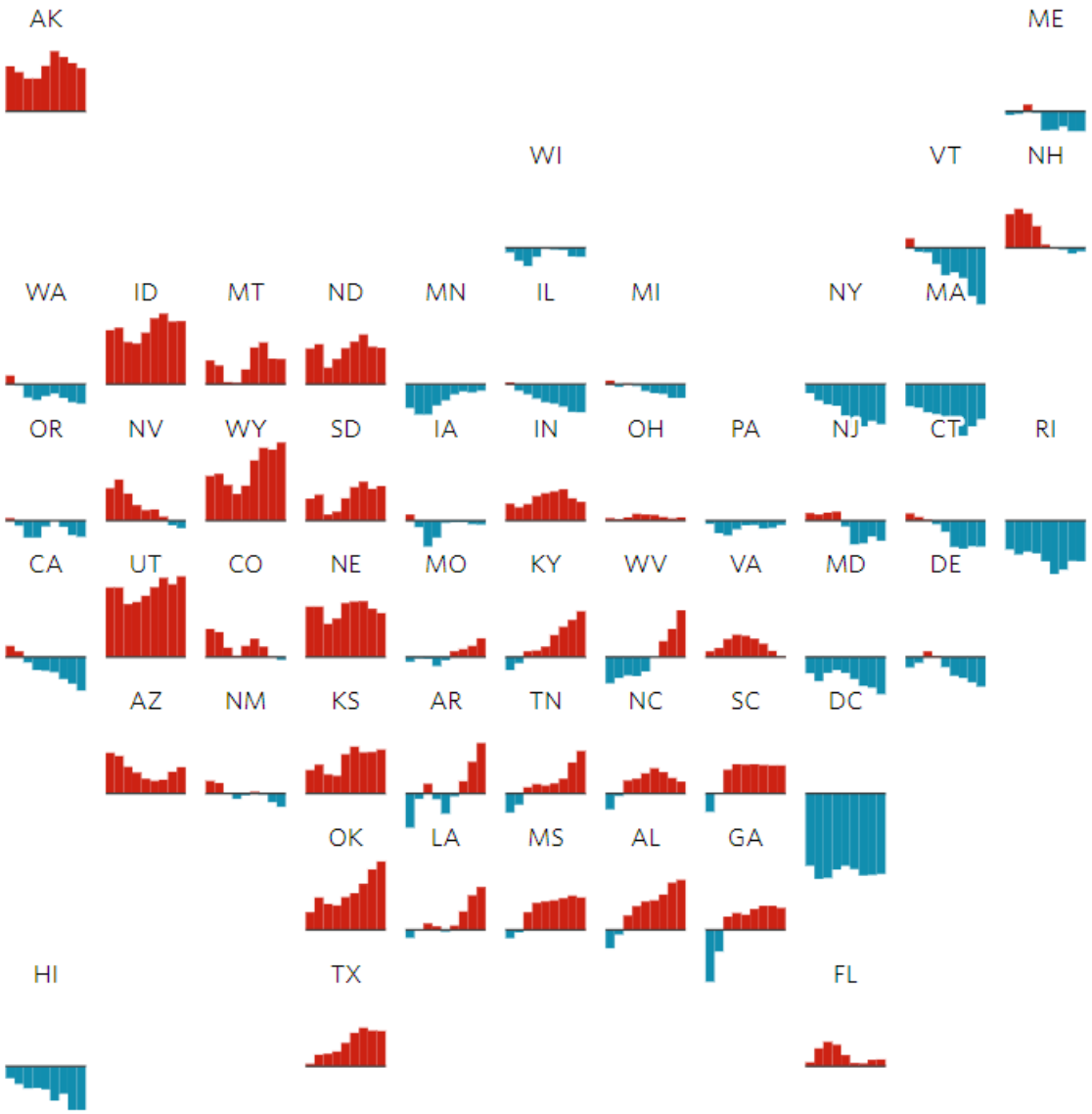


5. Layer, layer, layer

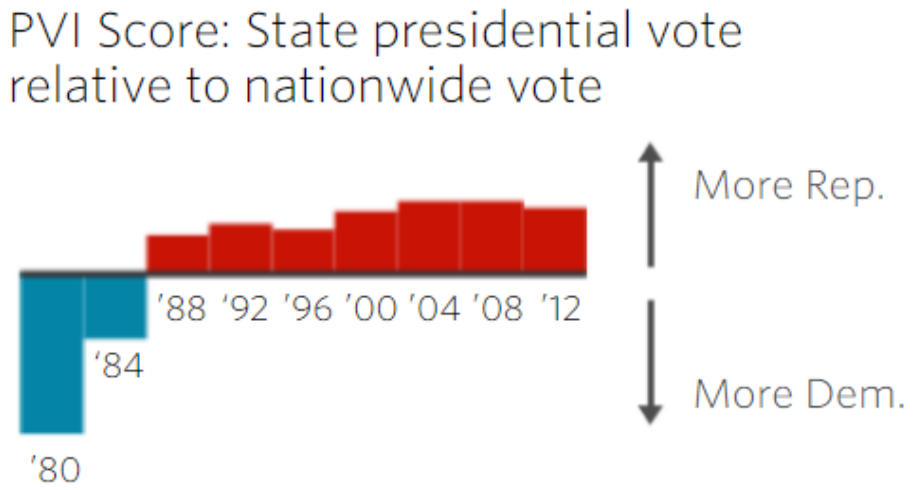
Design for micro-macro reading

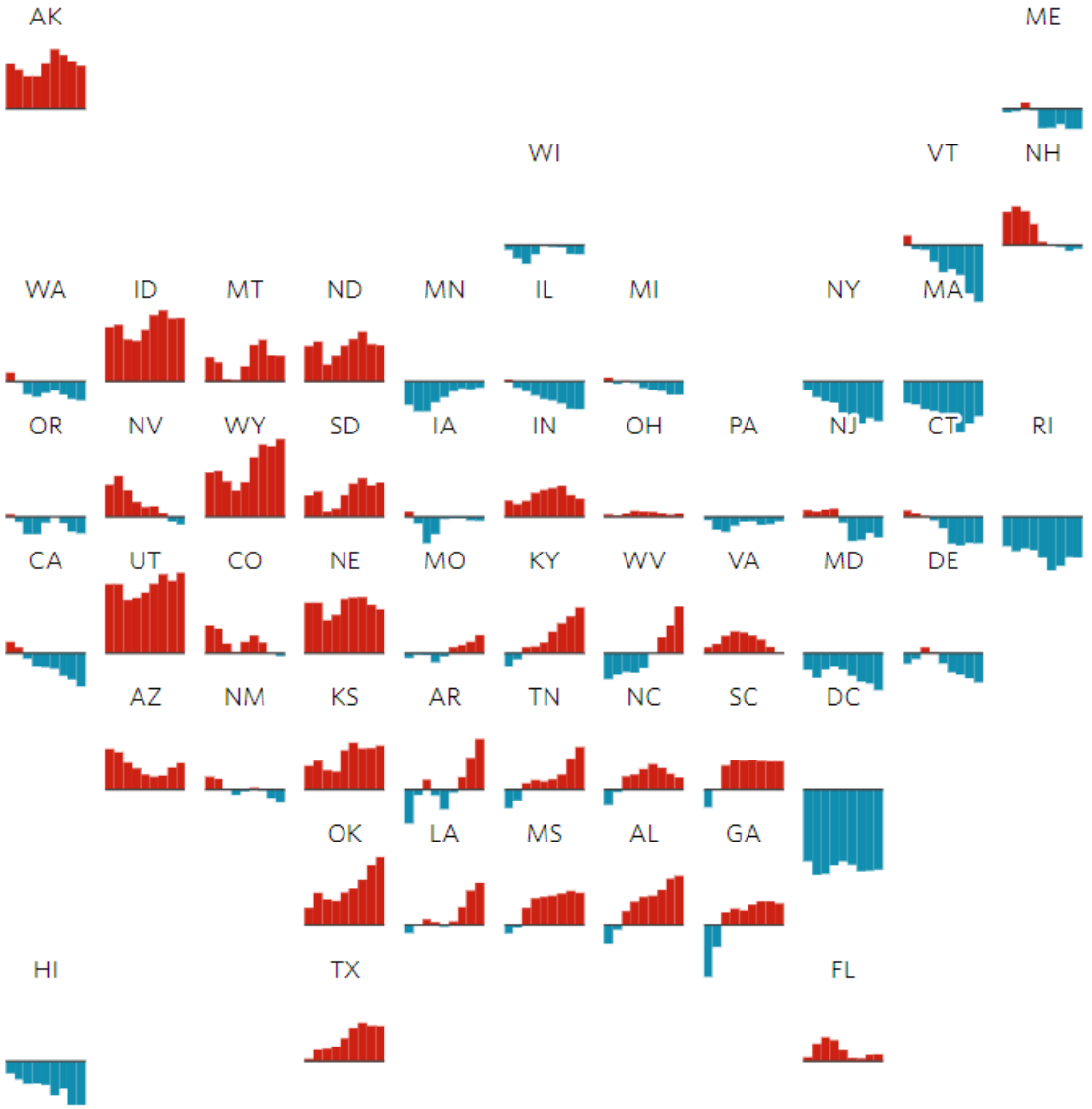
Pre-attentive attributes help

[http://graphics.wsj.com/elections/2016/field-guide-red-blue-america/]



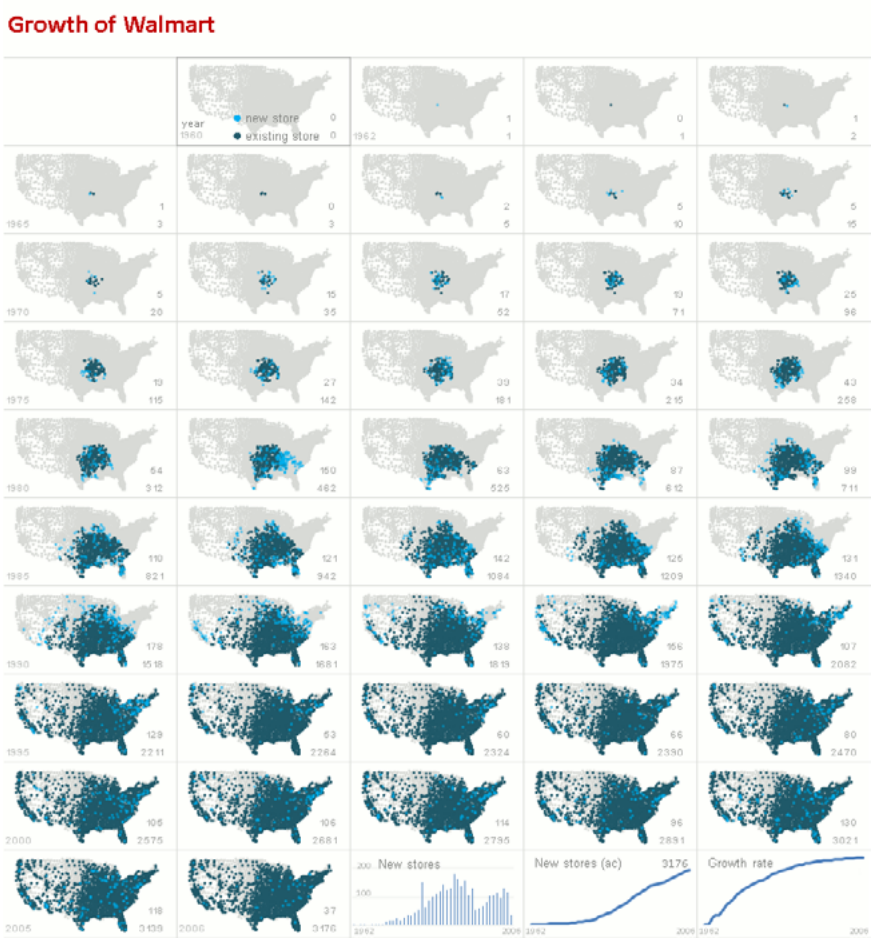
5. Layer, layer, layer



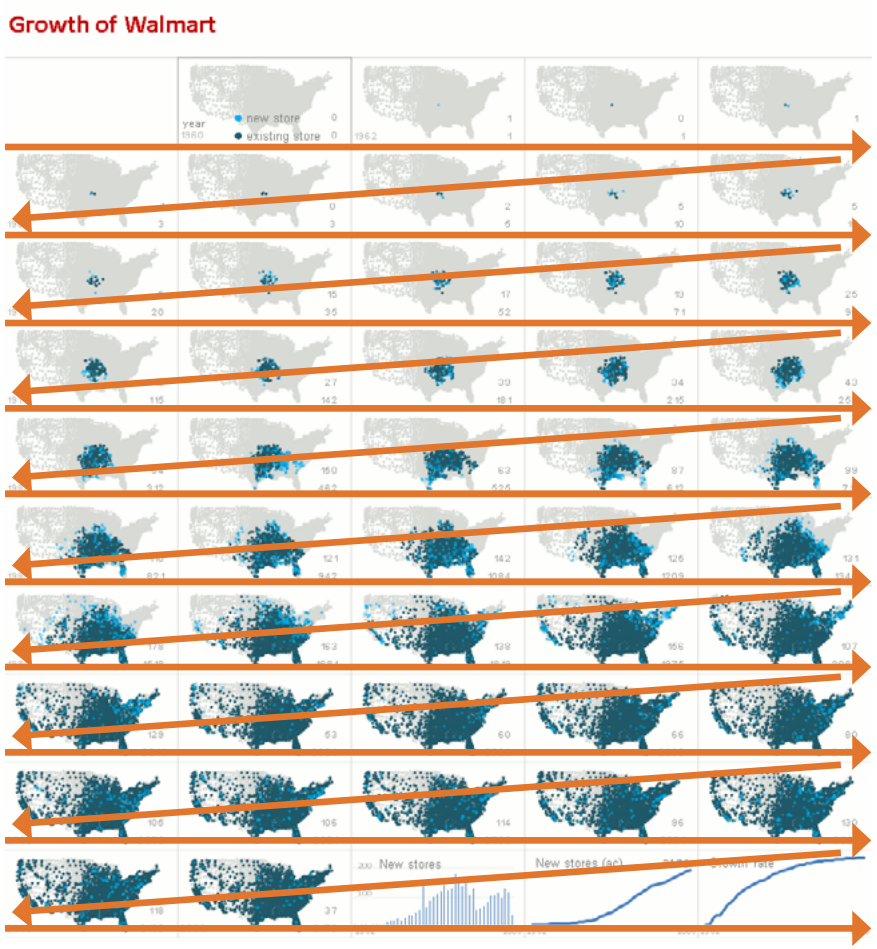


[http://graphics.wsj.com/elections/2016/field-guide-red-blue-america/]

(small multiples)

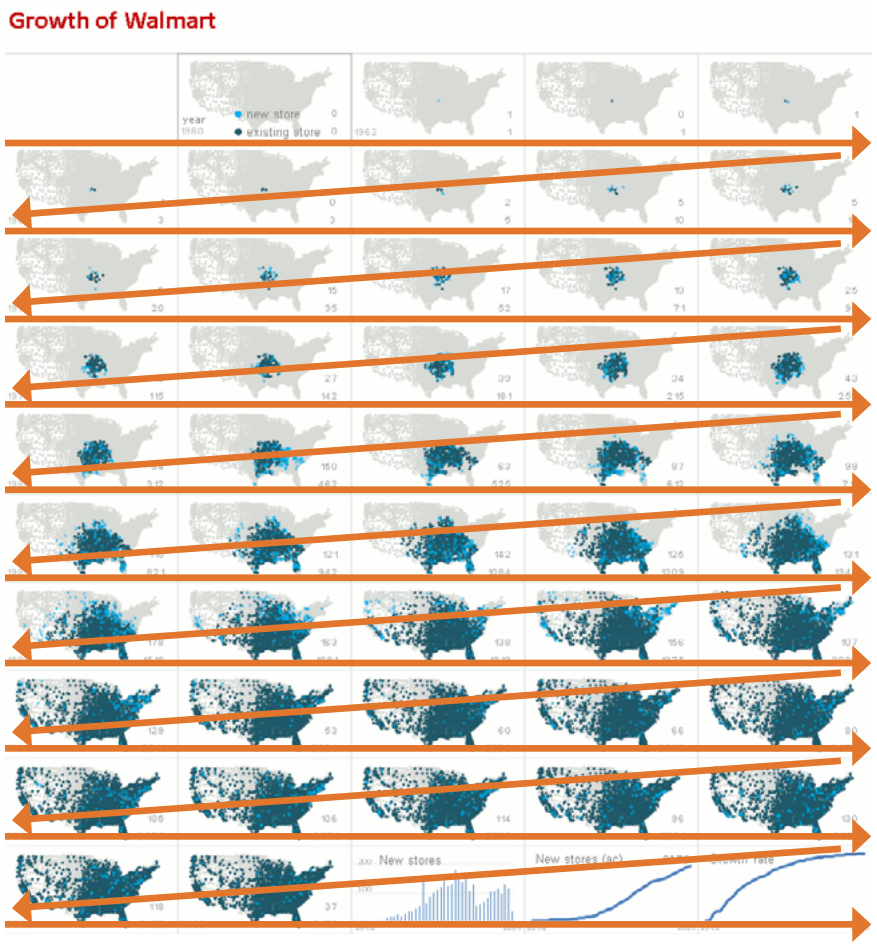


(small multiples = double use of position)

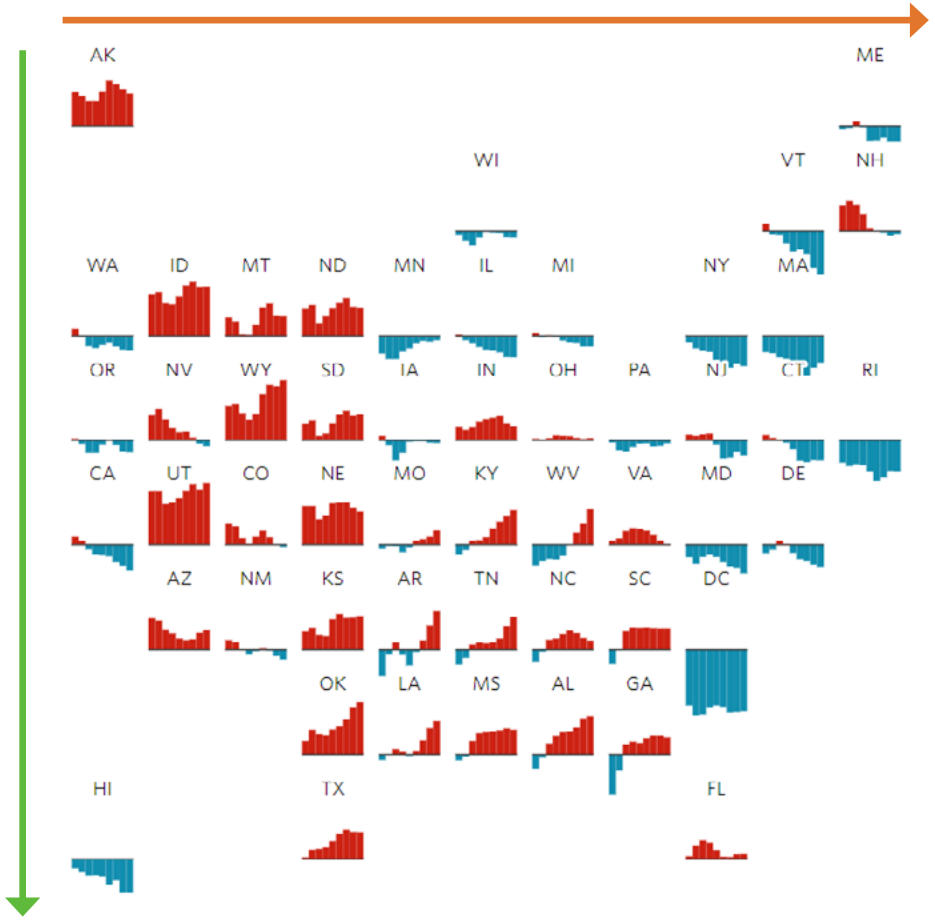


year -> wrapped column (x position)

(small multiples = double use of position)



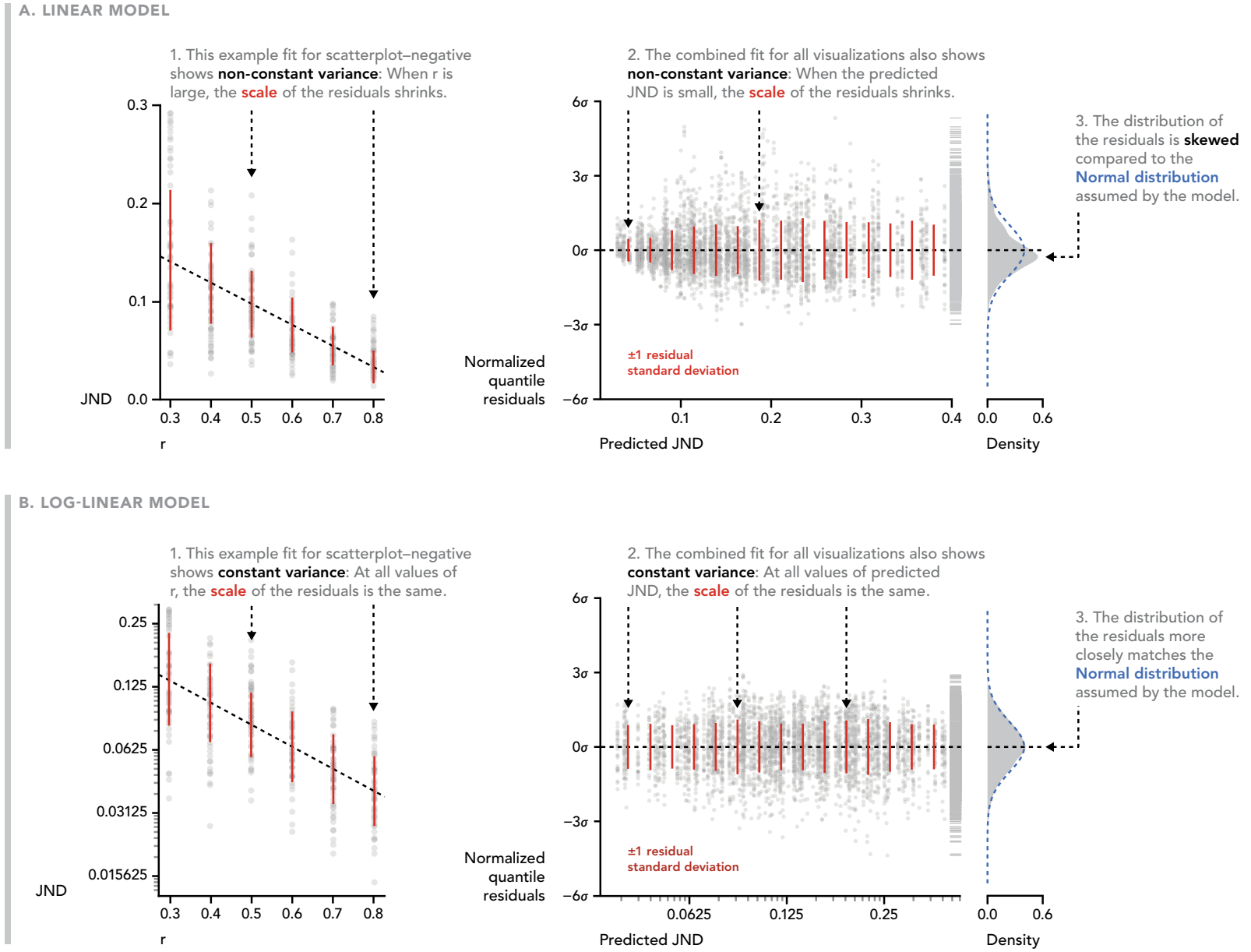
year -> wrapped column (x position)



~longitude -> column (x position)
~latitude -> row (y position)

6. When in doubt, grid

And get synchronized axes as a bonus

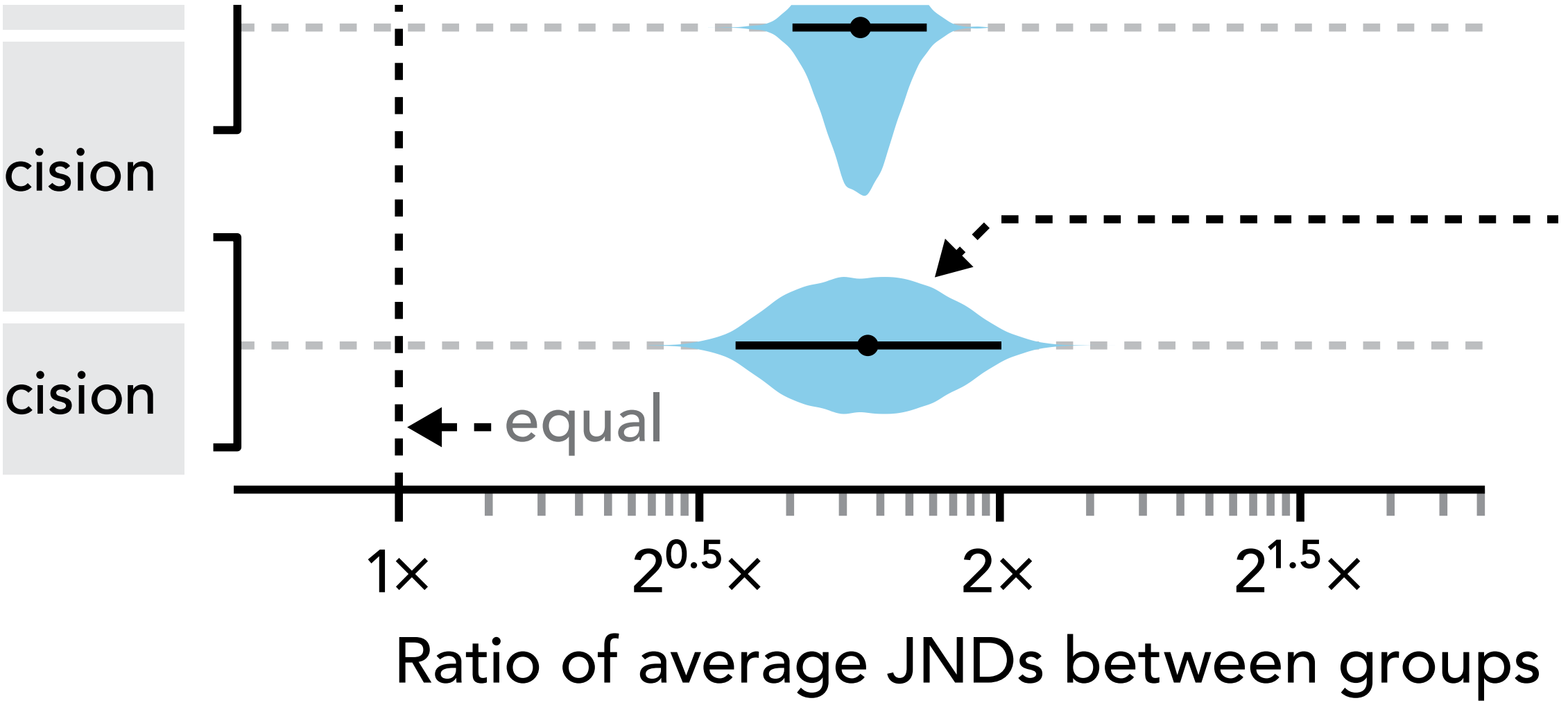


7. Treat visual attributes like adjectives

Don’t use three attributes (size, color, shape, ...) to create emphasis where one or two will do.

The very tall building is very extremely tall.

(7b. Obey the pen)

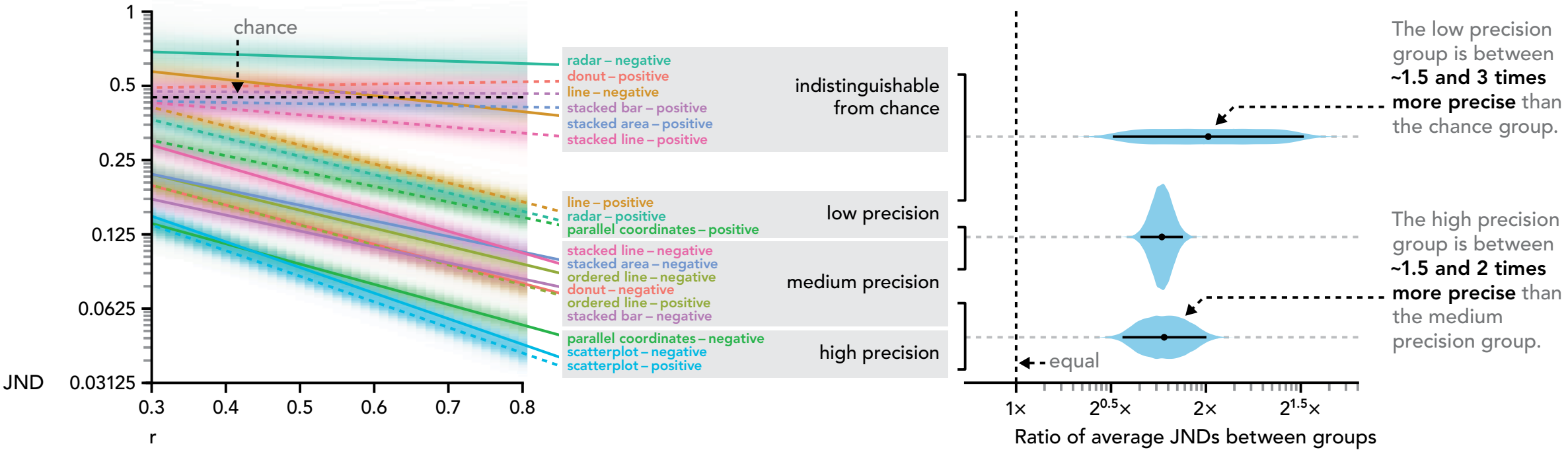


(7b. Obey the pen)

1. The final Bayesian censored log-linear model gives us a posterior probability distribution over the mean log(JND) for each value of r.

2. We rank and group visualizations based on how precise people's estimations of correlations are with them (lower expected JND implies higher precision)

3. We estimate the ratio of average JNDs between successive groups over all values of r from 0.3 to 0.8.



(7b. Obey the pen)

Even visual texture is pleasing

Also makes it easier to create visual hierarchy and call out something important when you need to

Some rough design guidelines*

1. Match effectiveness with importance
2. Avoid ambiguity
3. Locality is king / eyes beat memory
4. Establish viewing order
5. Layer, layer, layer
6. When in doubt, grid
7. Treat visual attributes like adjectives

* These guidelines are drawn largely from my experience + personal preferences + the literature. Design is messy, these are not perfect, others will disagree with me, etc. *Caveat emptor.*

In sum

Understand the effectiveness of different aesthetics (encodings).

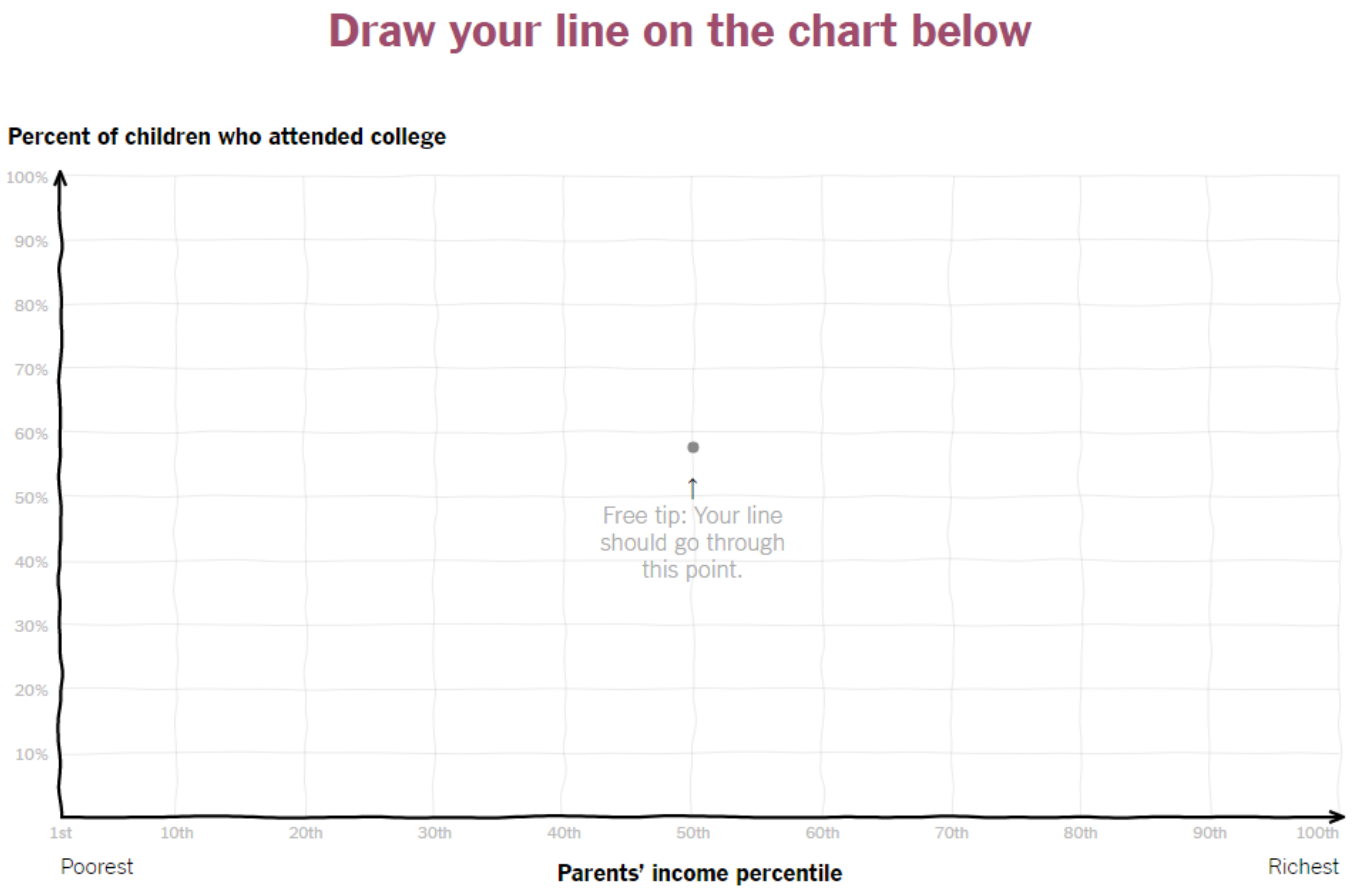
Understand where your viewer will look and what they want to do (tasks).

Visualize as a reflex during analysis.

Questions?

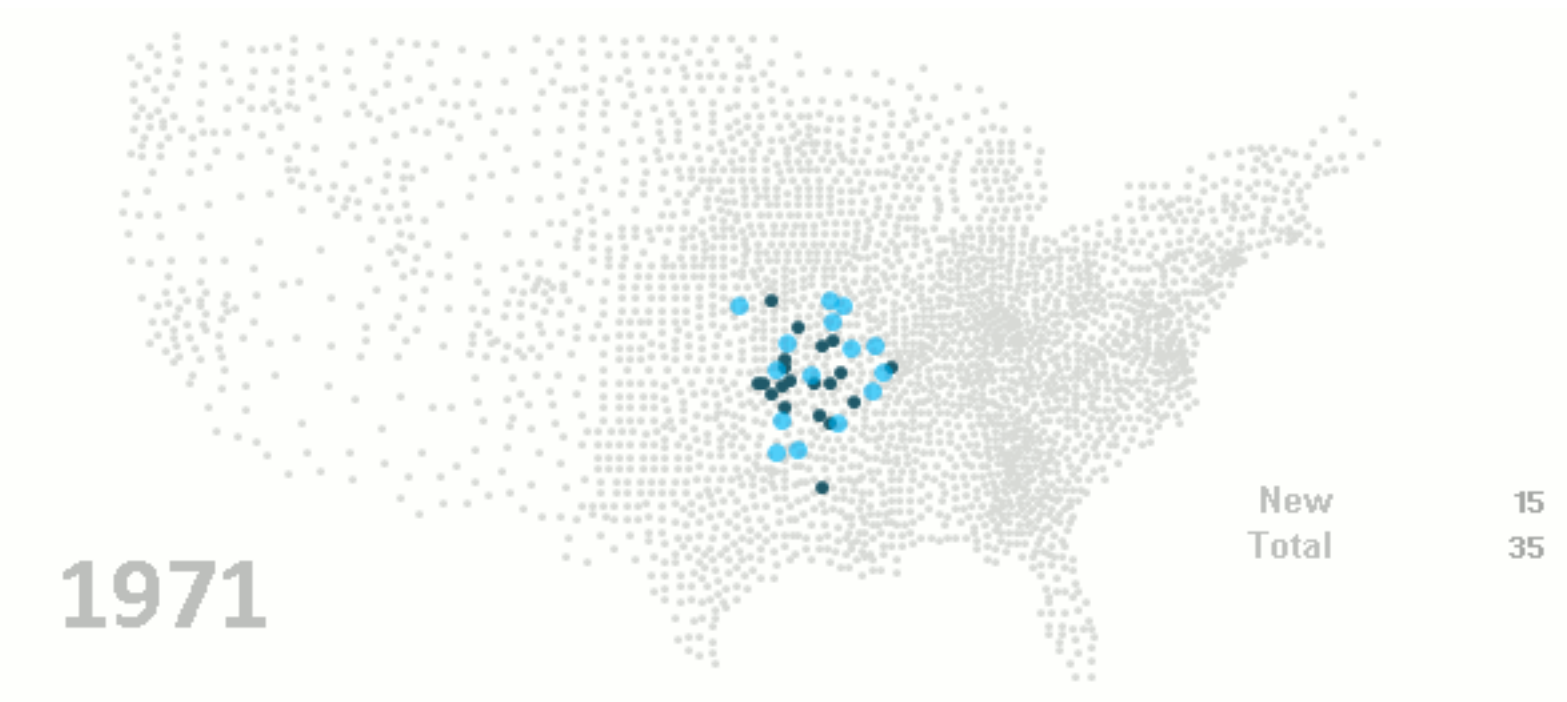
Examples / exercises

Prediction and memory

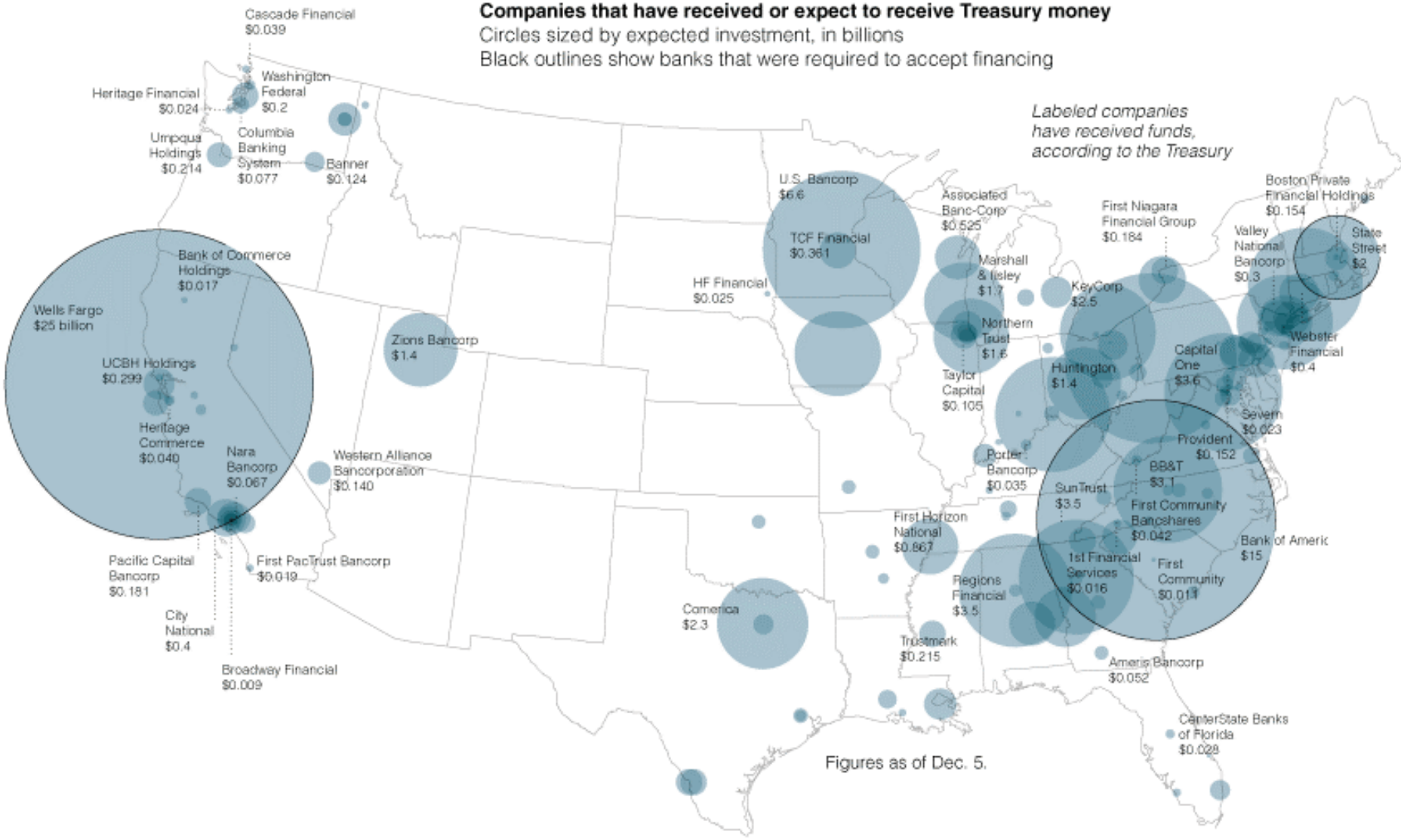


[https://nyti.ms/2jX8zue]

Small multiples



[https://excelcharts.com/animation-small-multiples-growth-walmart-excel-edition/]



Group activity

What are the variables / types?

Channels / encodings?

Marks?

Is this effective?

Hack Your Way To Scientific Glory

You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

1 CHOOSE A POLITICAL PARTY

Republicans | Democrats

2 DEFINE TERMS

Which politicians do you want to include?

- [ ] Presidents
- [x] Governors
- [x] Senators
- [x] Representatives

How do you want to measure economic performance?

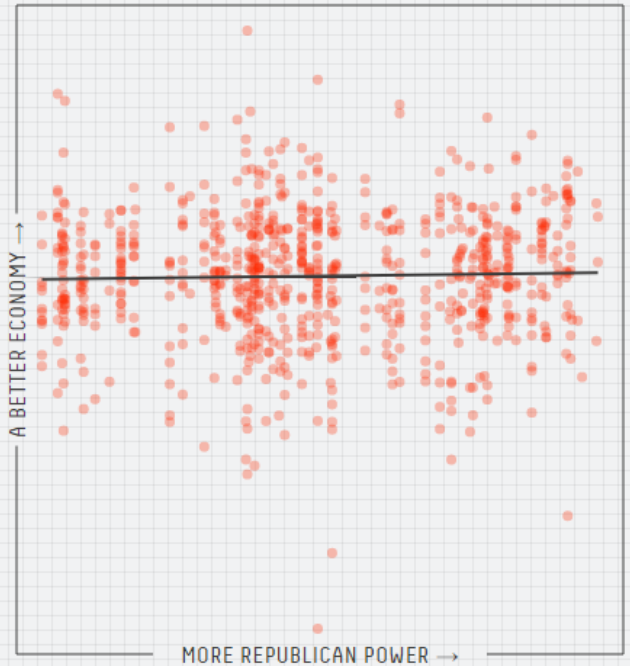
- [ ] Employment
- [ ] Inflation
- [ ] GDP
- [x] Stock prices

Other options

- [x] **Factor in power**
  Weight more powerful positions more heavily
- [ ] **Exclude recessions**
  Don't include economic recessions

3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Republicans are in power? Each dot below represents one month of data.



4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a **p-value of 0.05 or less** to get published.

Result: Unpublishable

With a p-value of **0.43**, your findings are not statistically significant. Try defining your terms differently.

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.

[https://fivethirtyeight.com/features/science-isnt-broken/]

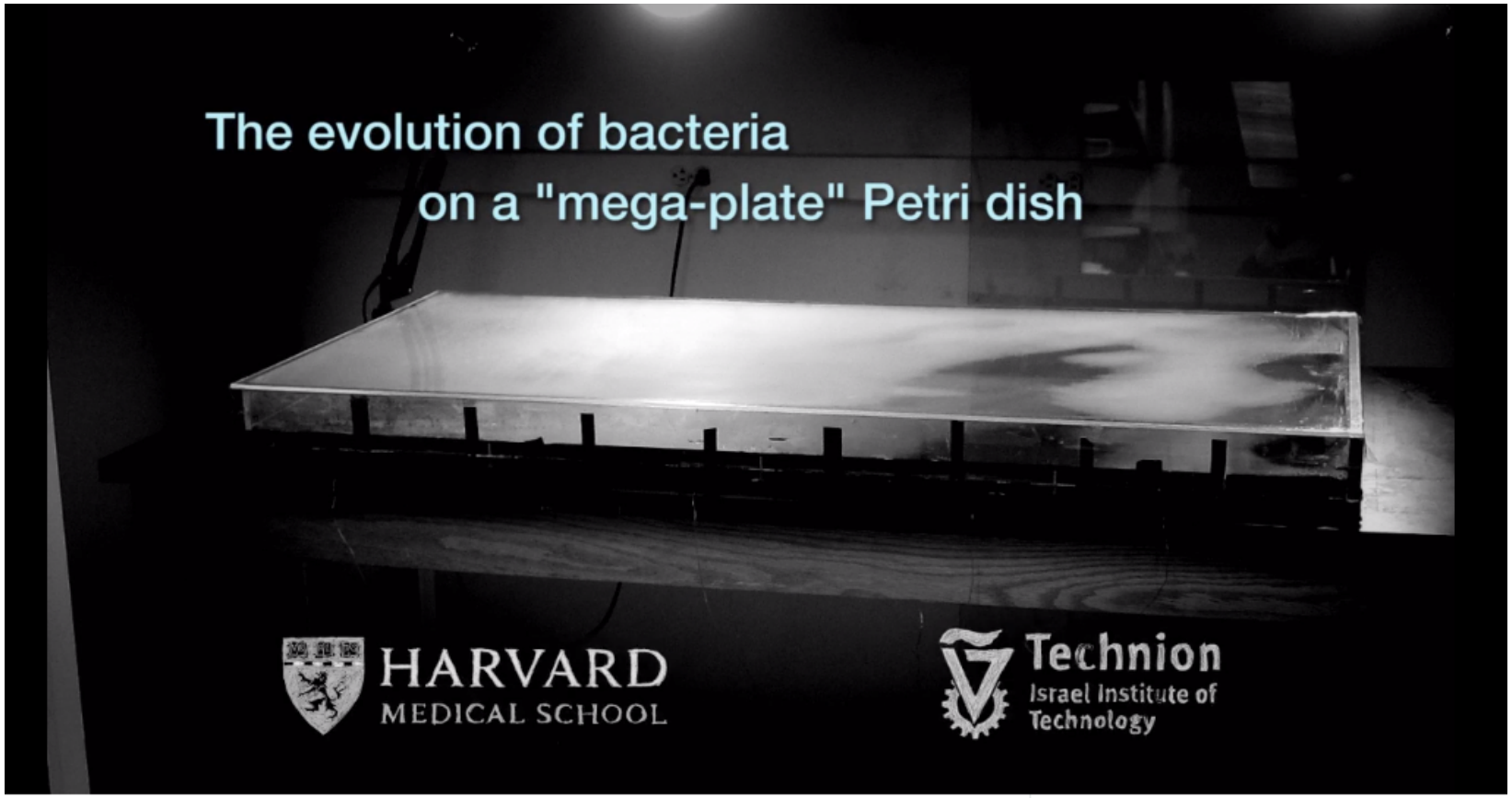
SPLOM: Scatter plot matrix

[https://bl.ocks.org/mbostock/4063663]

Hyberbolic trees

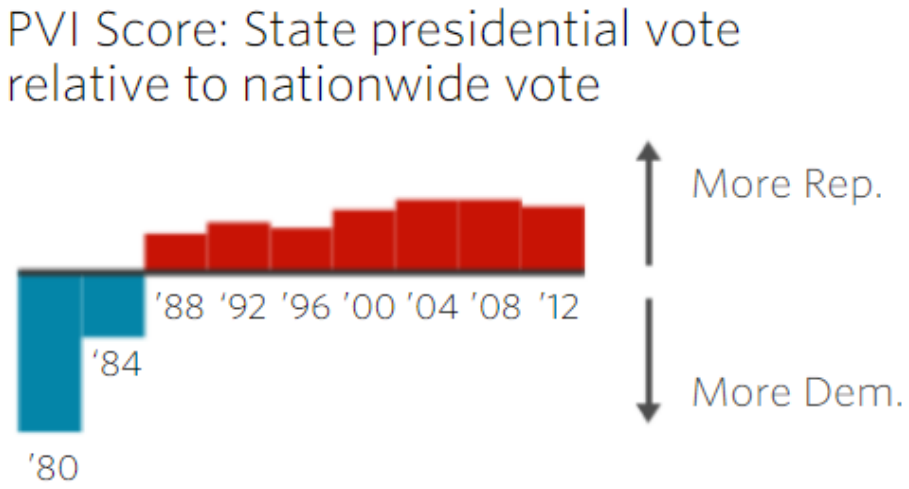
[https://youtu.be/fhbQy_NCwWI]

Evolution of bacteria



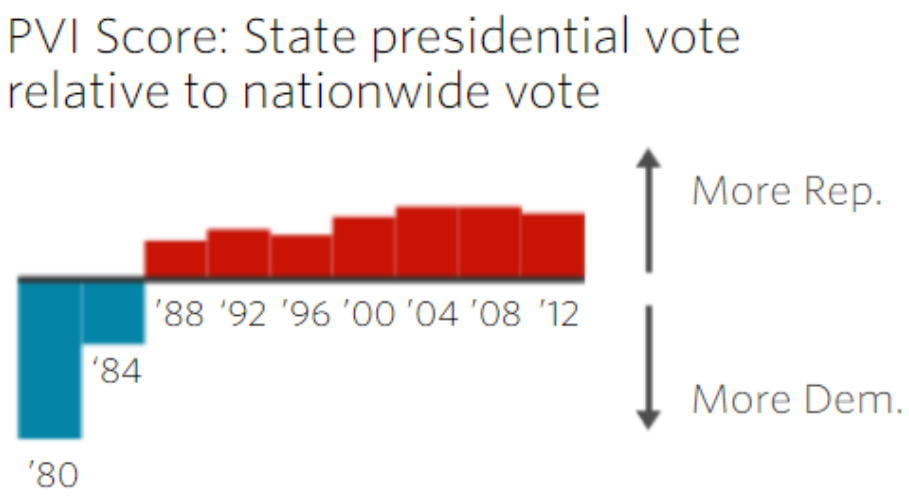
https://vimeo.com/180908160

Small multiples



[http://graphics.wsj.com/elections/2016/field-guide-red-blue-america/]

Small multiples



[http://graphics.wsj.com/elections/2016/field-guide-red-blue-america/]