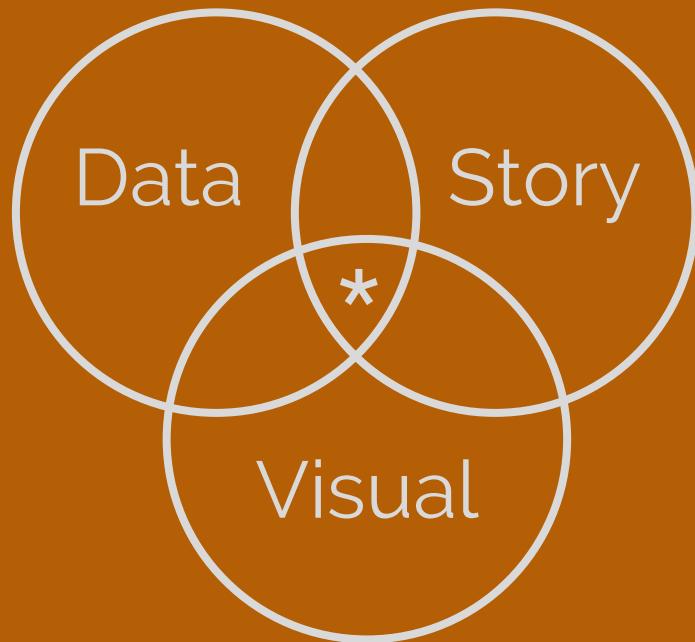


Data Visualisation

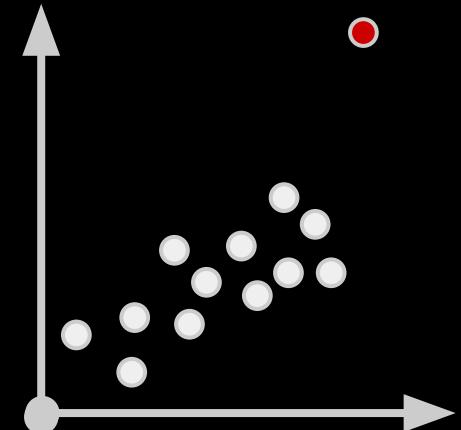
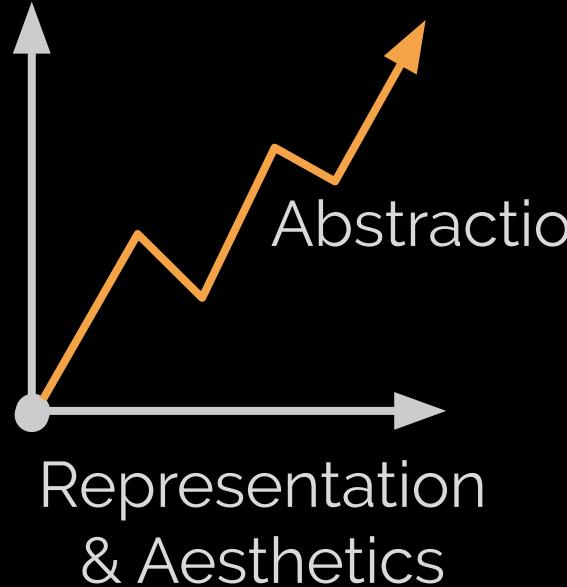
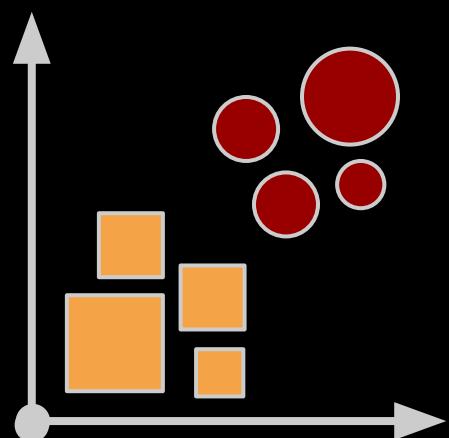
Amit Kapoor
@amitkaps

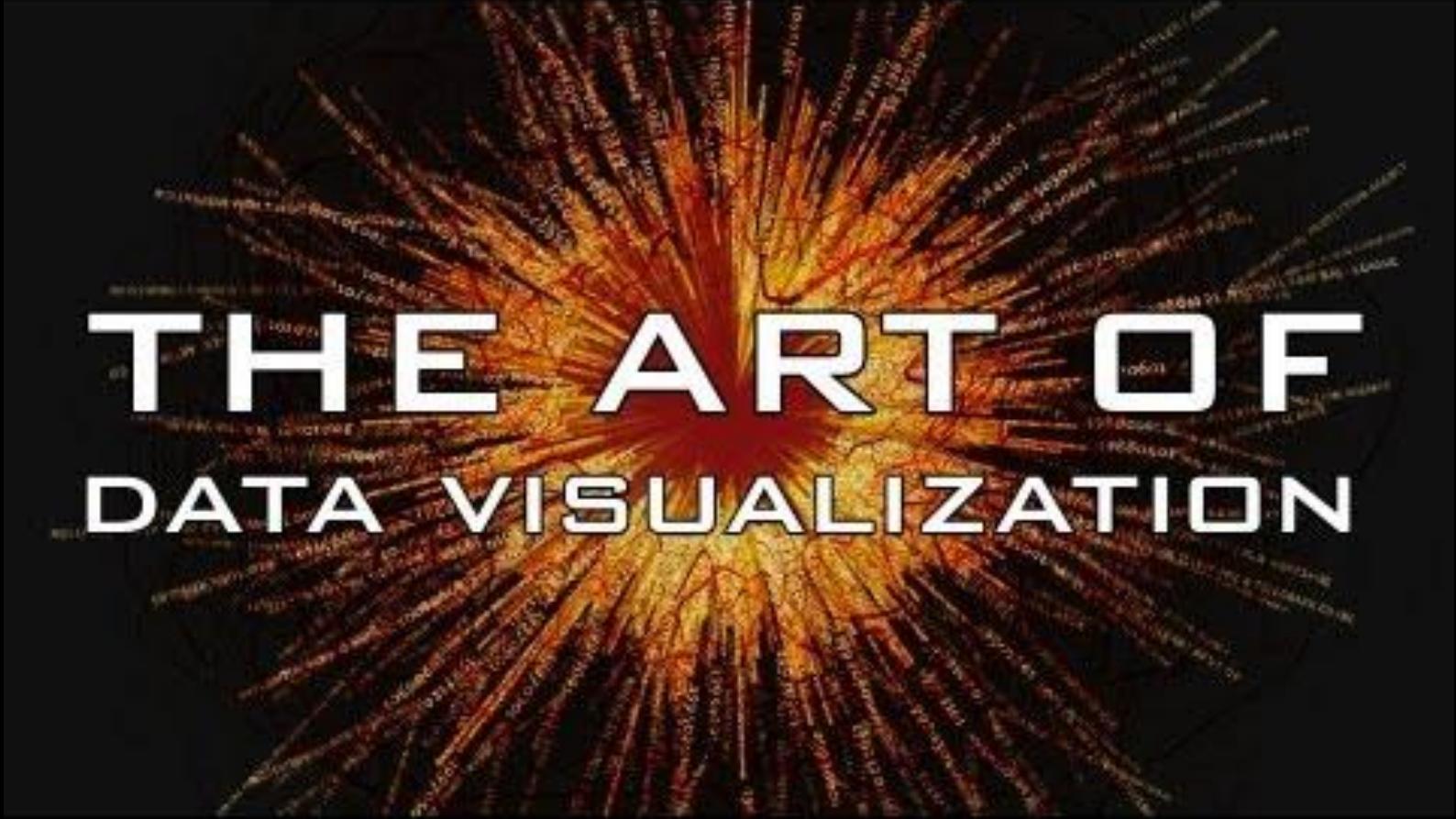
Aesthetics



3

Aesthetics





THE ART OF DATA VISUALIZATION

Source: [The Art of Data Visualization](#)

How many 5's can you find?

142536789251364789245369178

419356728495126783149356728

245369178145672893145672938

495126783149356728423698517

359164782145672938451672938

465132978423698517459163782

145762938451672938359164782

431567298459163782431567298

Proximity

142 5 367892 5 136478924 5 369178
4193 5 672849 5 1267831493 5 6728
24 5 36917814 5 67289314 5 672938
49 5 1267831493 5 6728423698 5 17
3 5 916478214 5 6729384 5 1672938
46 5 132978423698 5 174 5 9163782
14 5 7629384 5 16729383 5 9164782
431 5 672984 5 9163782431 5 67298

Alignment

555 142367892136478924369178
555 419367284912678314936728
555 243691781467289314672938
555 491267831493672842369817
555 391647821467293841672938
555 461329784236981749163782
555 147629384167293839164782
555 431672984916378243167298

Repetition

Enclosure

142 [5] 367892 [5] 136478924 [5] 369178
4193 [5] 672849 [5] 1267831493 [5] 6728
24 [5] 36917814 [5] 67289314 [5] 672938
49 [5] 1267831493 [5] 6728423698 [5] 17
3 [5] 916478214 [5] 6729384 [5] 1672938
46 [5] 132978423698 [5] 174 [5] 9163782
14 [5] 7629384 [5] 16729383 [5] 9164782
431 [5] 672984 [5] 9163782431 [5] 67298

Contrast

142536789251364789245369178

419356728495126783149356728

245369178145672893145672938

495126783149356728423698517

359164782145672938451672938

465132978423698517459163782

145762938451672938359164782

431567298459163782431567298

Subtraction

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

5

Preattentive Process

Subtraction

Contrast

Repetition

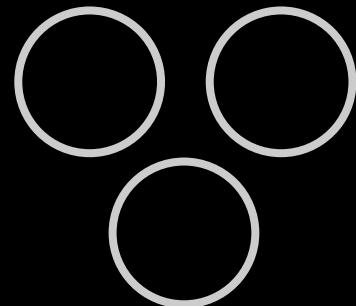
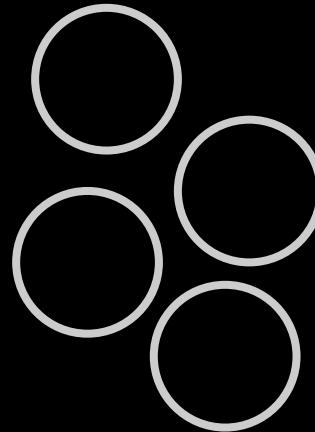
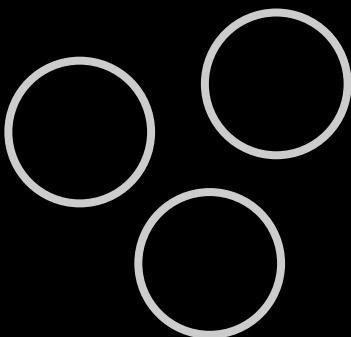
Alignment

Proximity

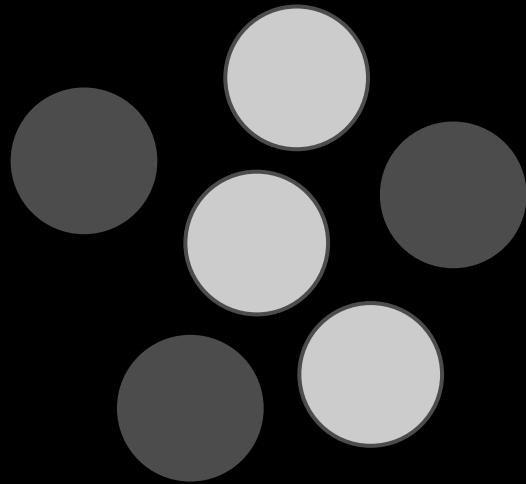
Enclosure

Gestalt Principles of Visual Perception

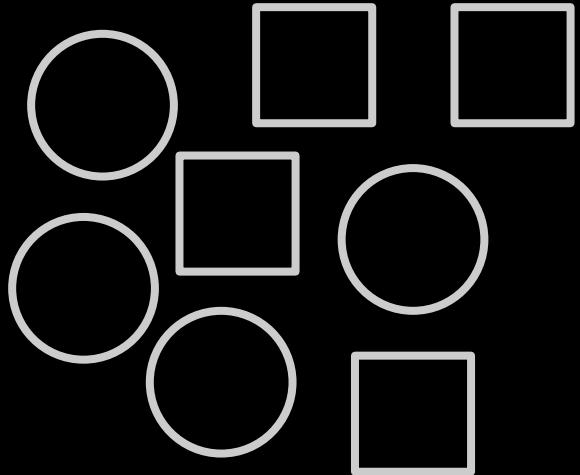
Proximity



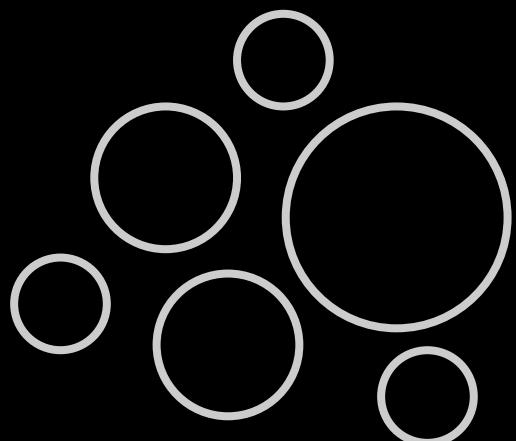
Similarity



Colour

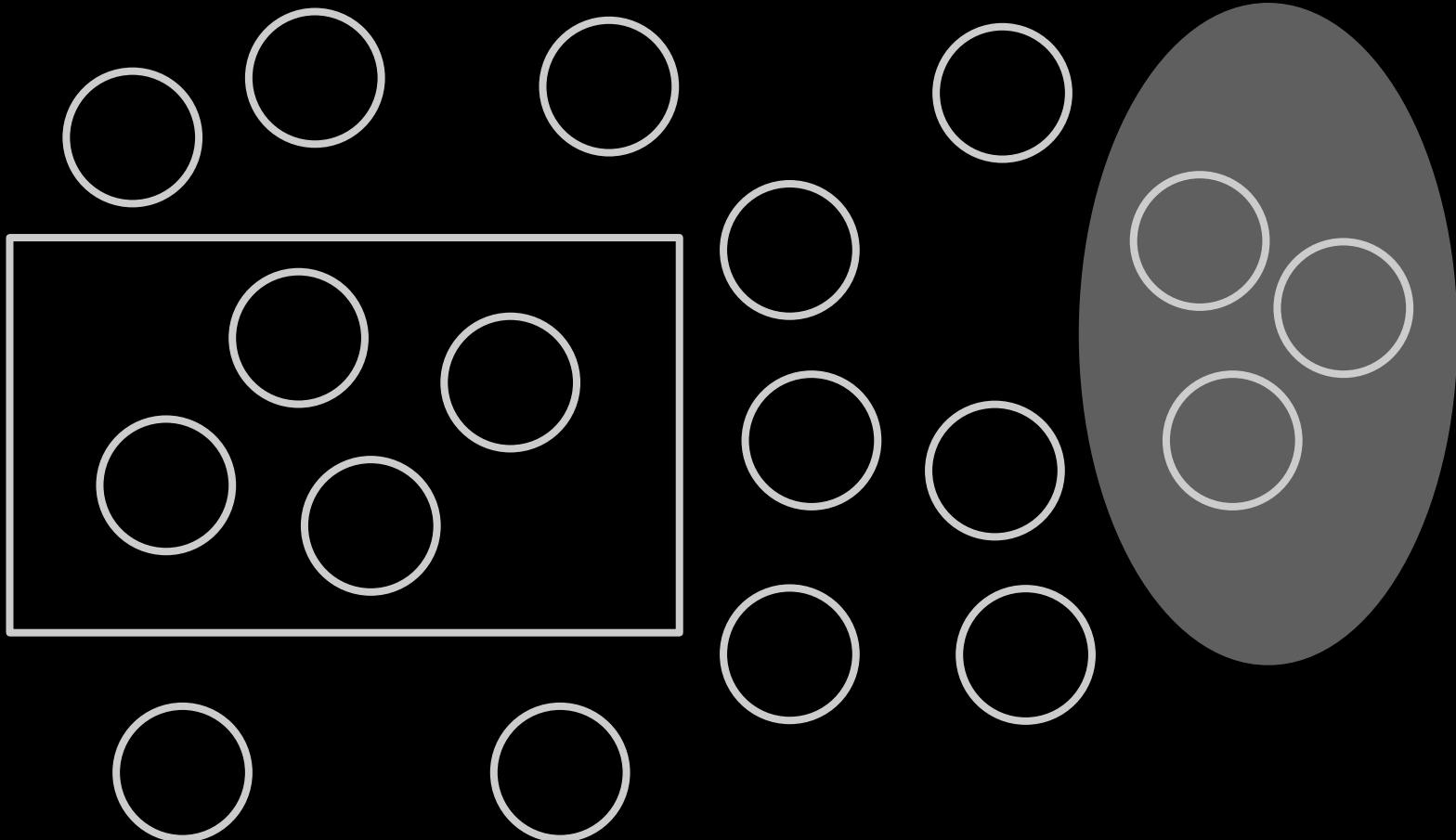


Shape

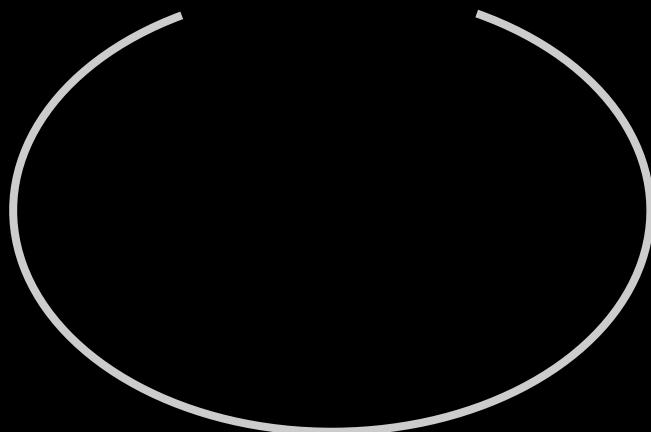


Size

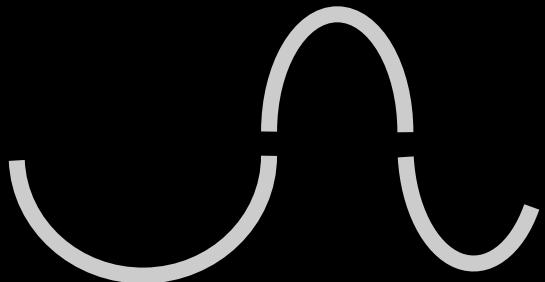
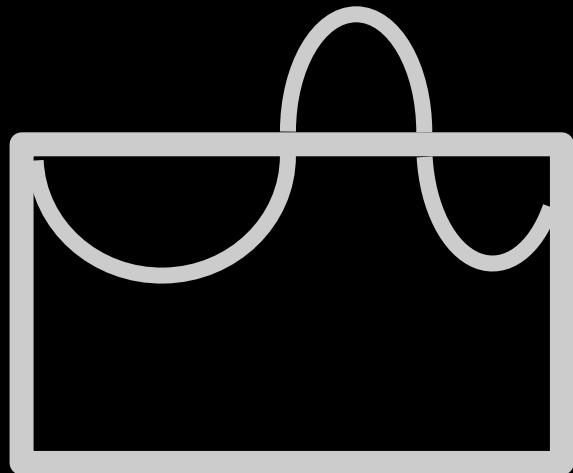
Enclosure



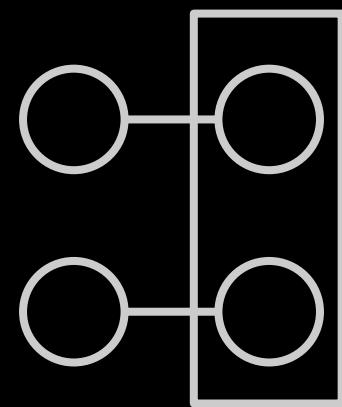
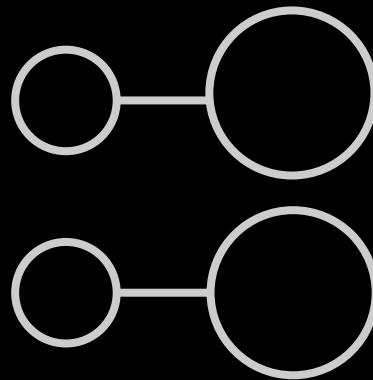
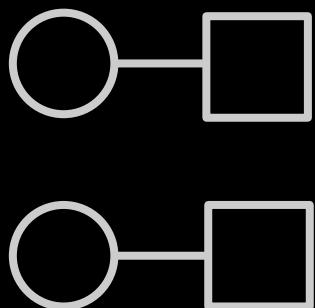
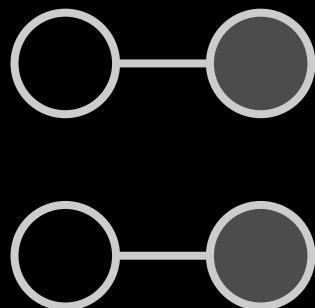
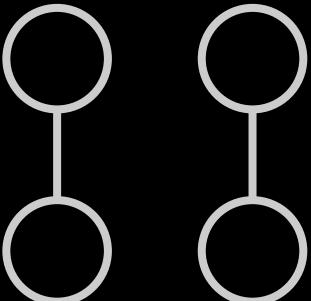
Closure



Continuity



Connection



Visual Language



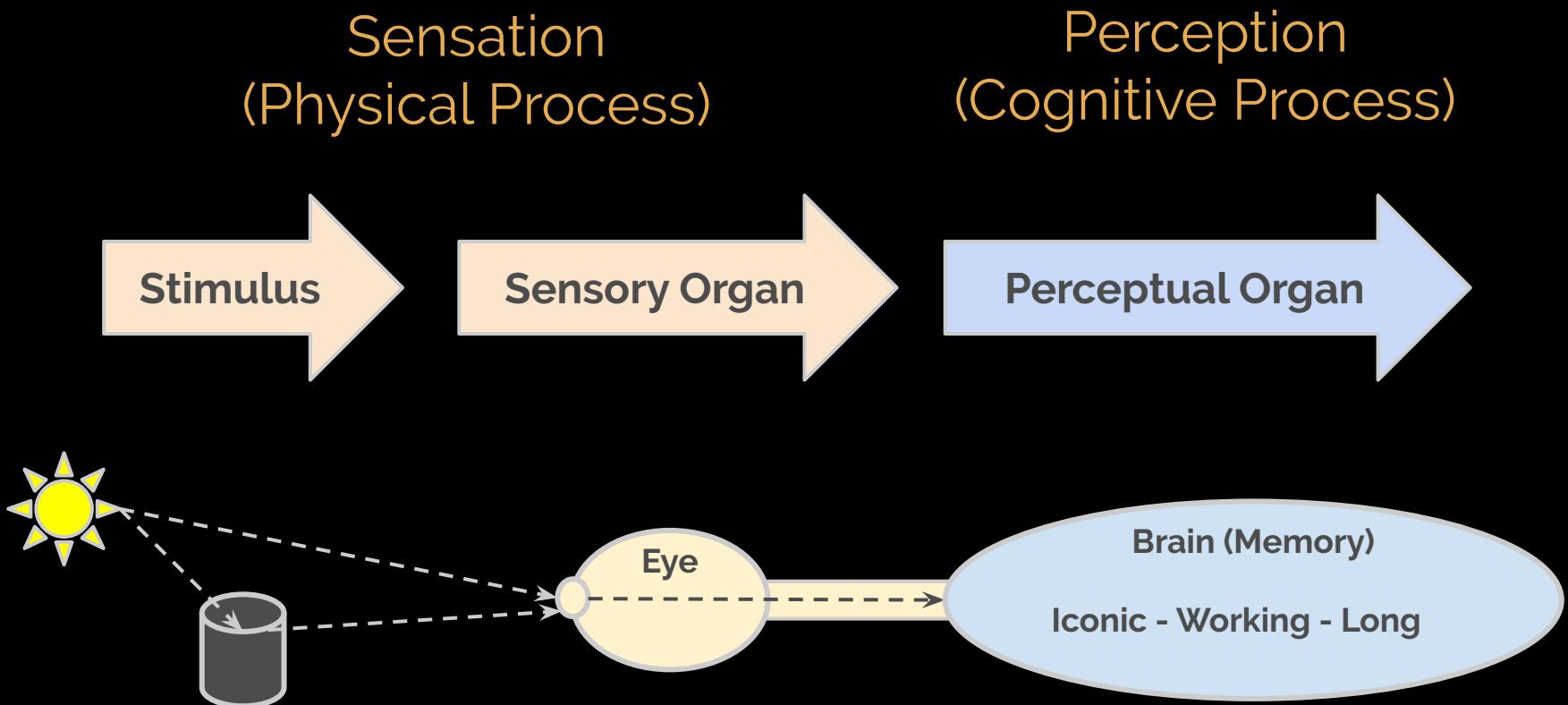
Visual Language is a Sign Language

“... finding the artificial memory that best supports our natural means of perception.” - Bertin

Graphical Perception

The ability of viewers to interpret visual (graphical) encodings of information and therefore decode information in graphs

Perception



Sensation - Eye

Fovea

Small focus, blurry outside

Saccadic Eye Movement

Eyes jump from one place to another (Saccades)

Perception - Memory

Iconic

Working

Long Term

Pre-attentive

Attentive

Attentive

< 1s - Fast

1-3s - Medium

Long

Temporary

Temporary

Permanent

Unconscious

Conscious

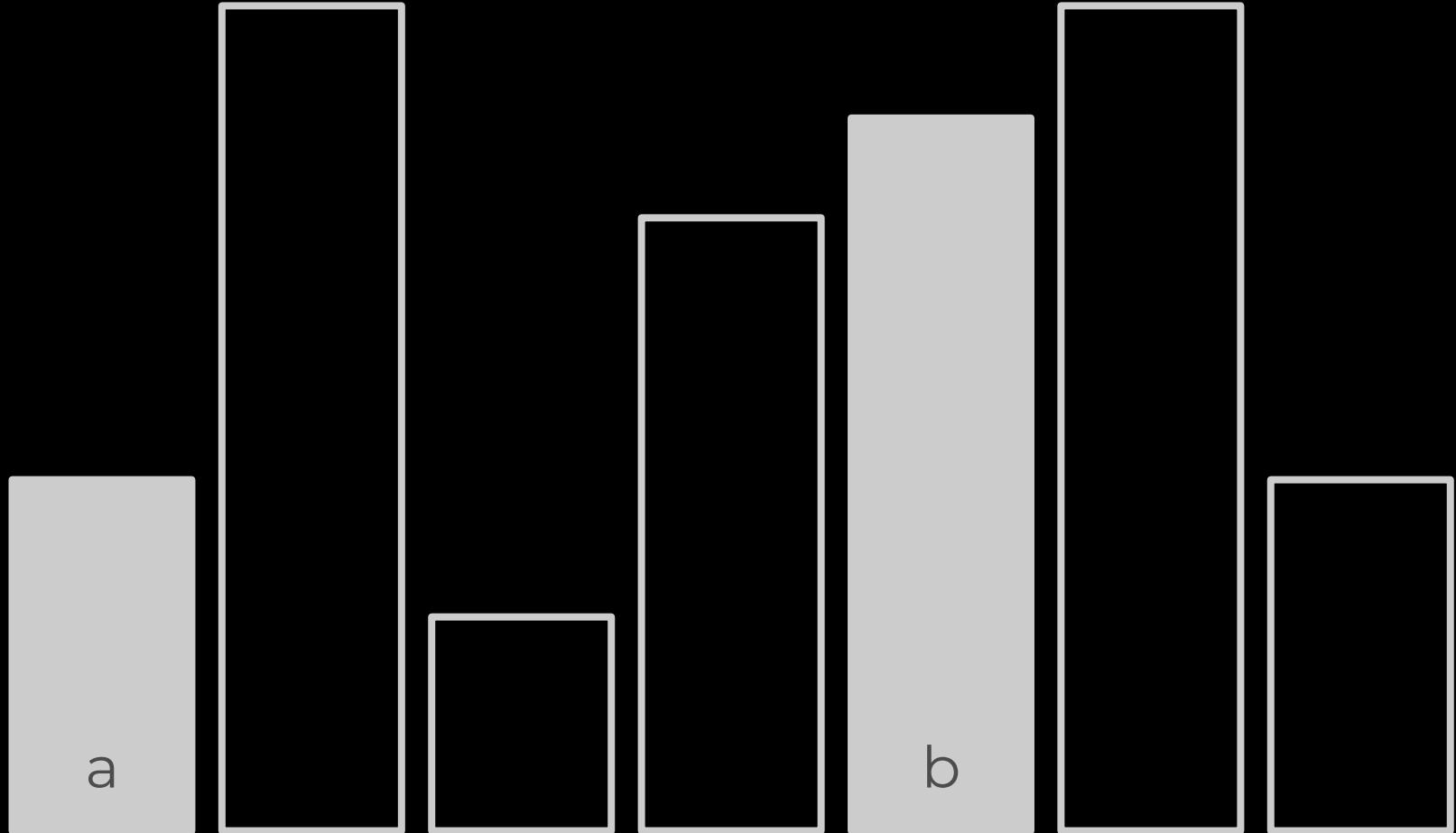
Deliberate

In-device

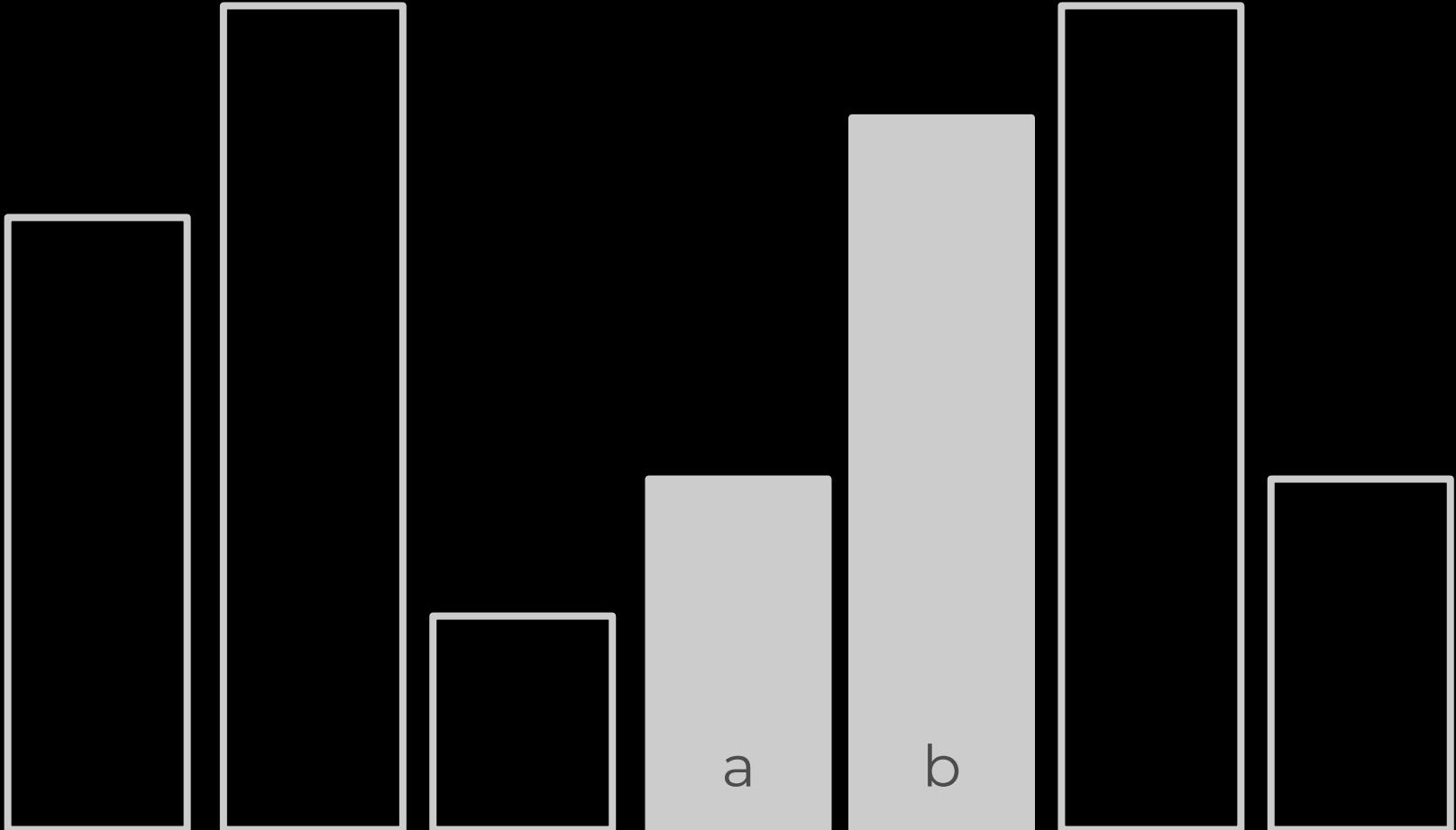
RAM

SSD

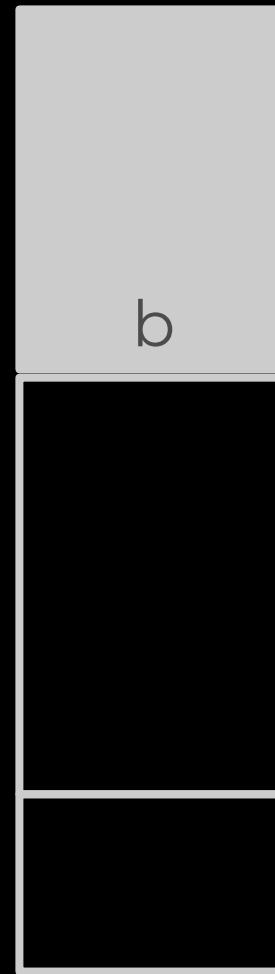
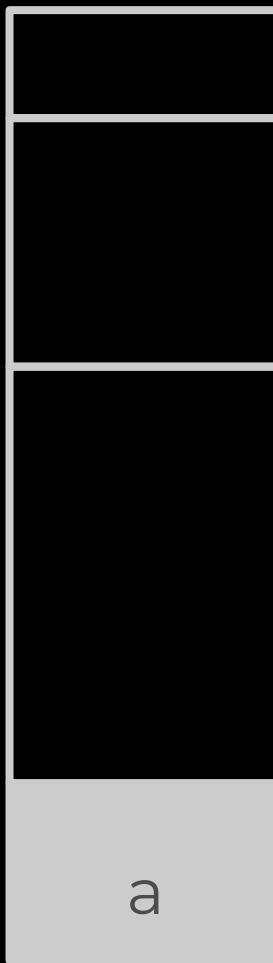
We compare relatively



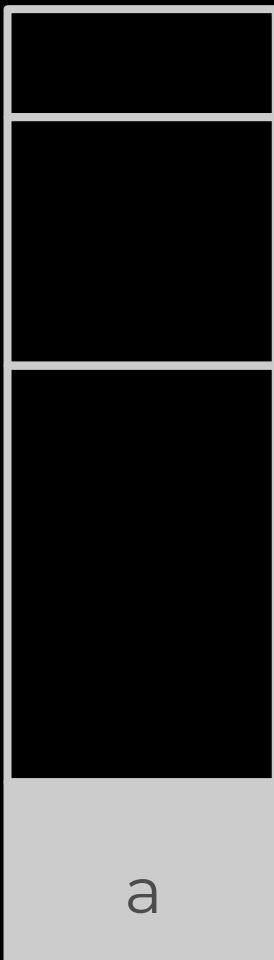
Proximity helps



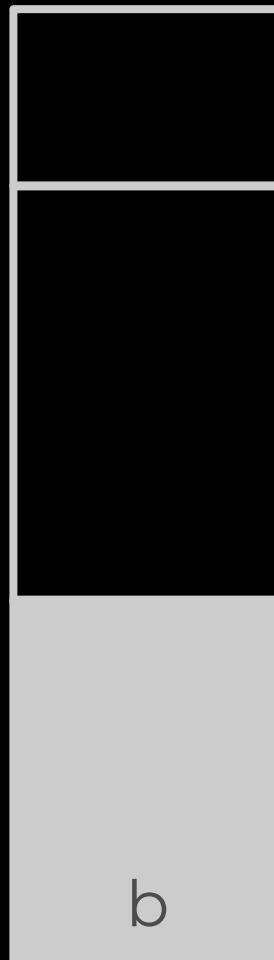
This is hard



Make comparisons easy

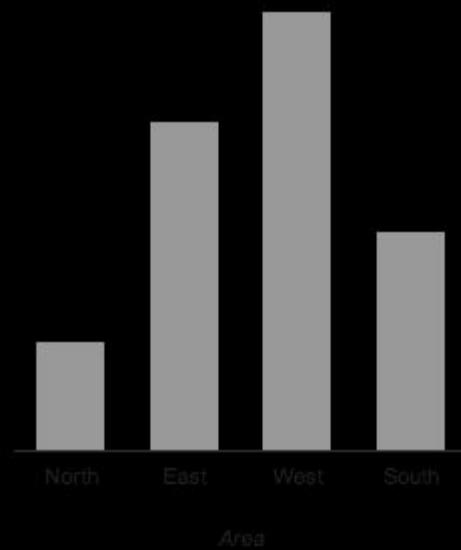


a

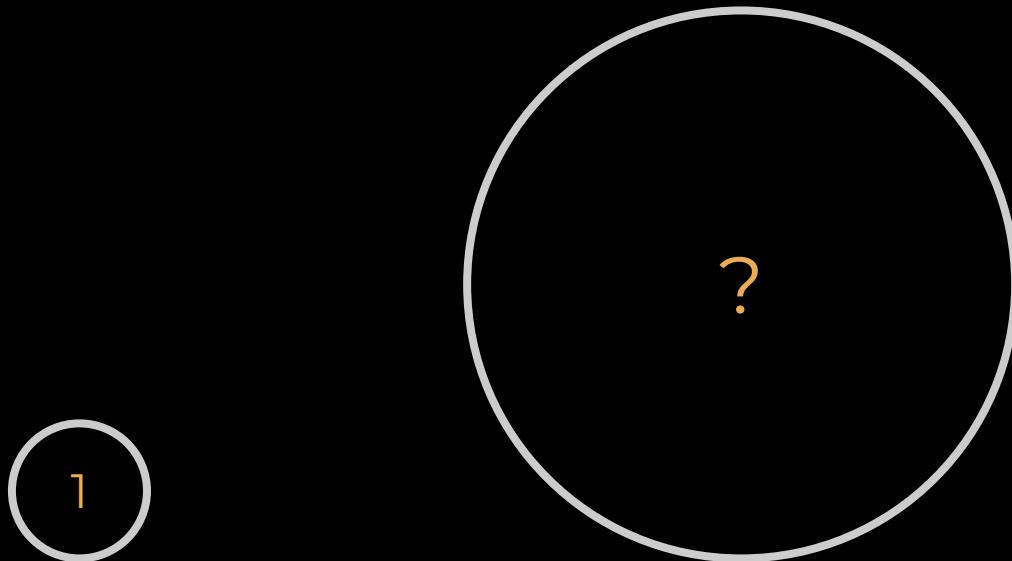


b

Angles vs. Length



Area Comparison



Preattentive Attributes

Form

Length, Width, Orientation,
Shape, Size, Enclosure

Color

Hue, Intensity

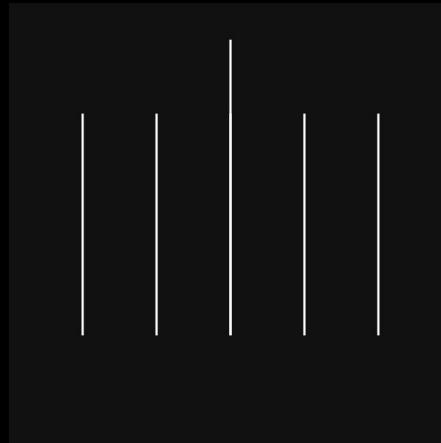
Spatial

2d position

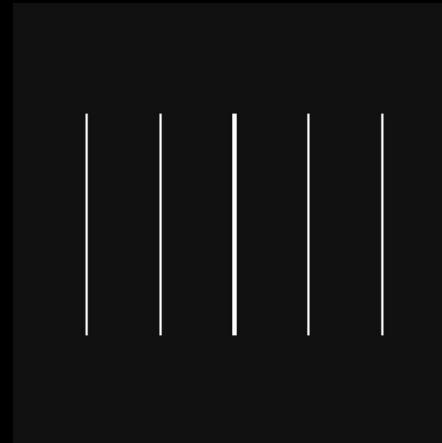
Motion

Preattentive - Form

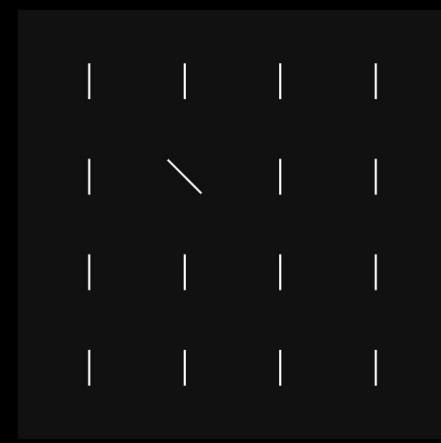
Length



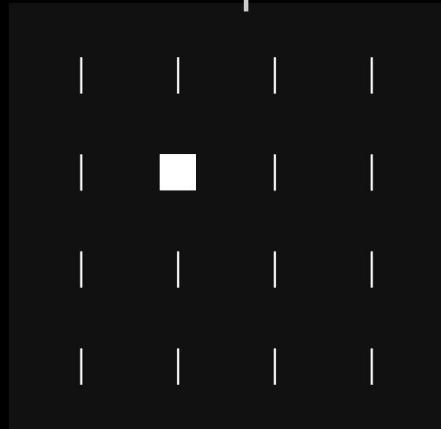
Width



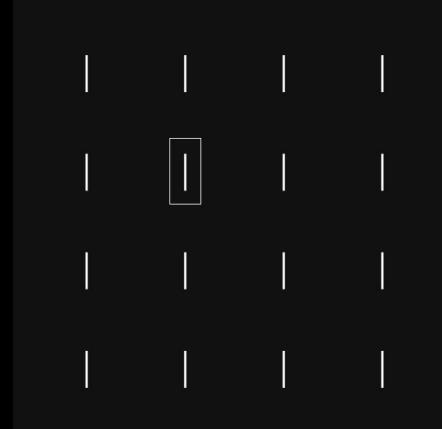
Orientation



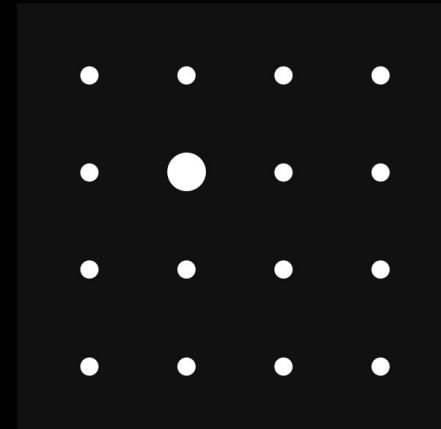
Shape



Enclosure

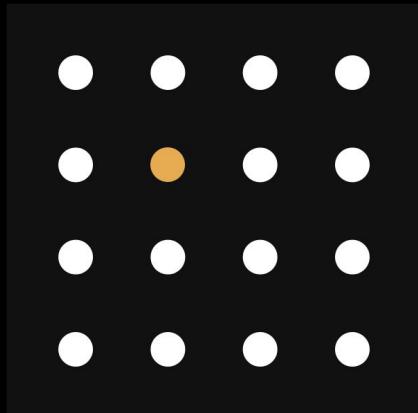


Size

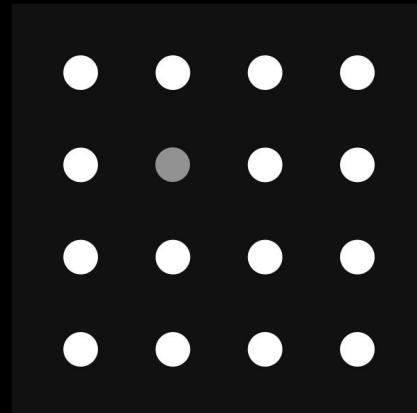


Preattentive - Color

Hue

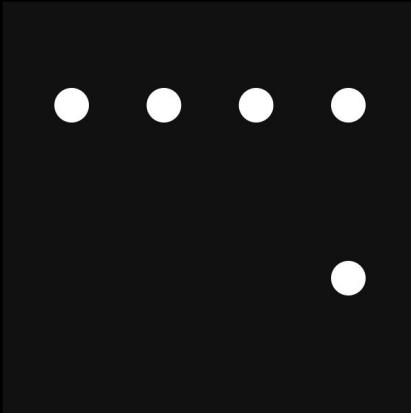


Saturation



Preattentive - Spatial

2D Position



3D Position

vs.

2.05D Vision

Preattentive - Motion

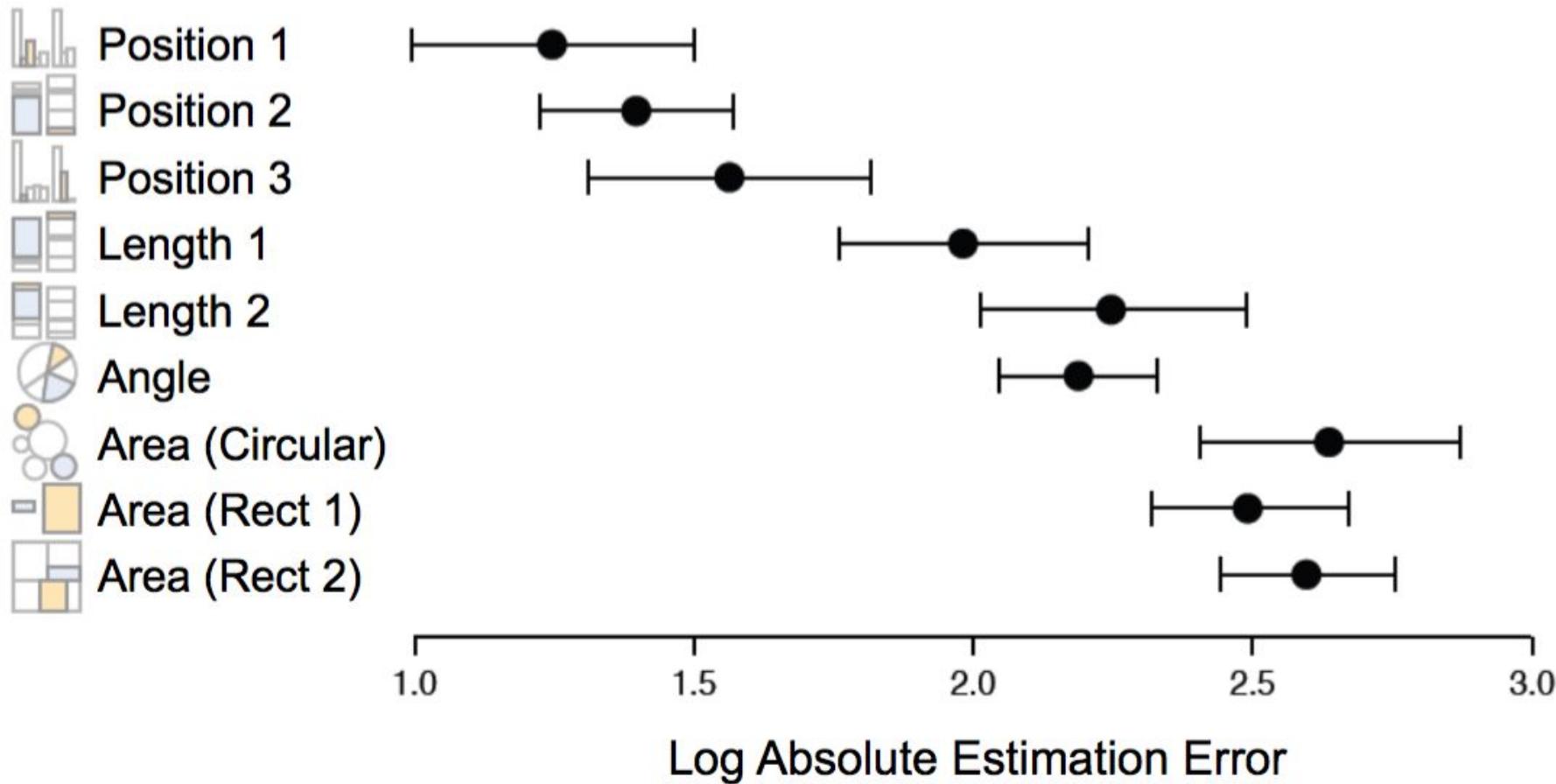


Ranking of Encoding

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Color Saturation	Color Hue
Angle	Color Hue	Connection
Slope	Connection	Enclosure
Size	Enclosure	Color Saturation
Volume	Length	Shape
Color Saturation	Angle	Length
Color Hue	Slope	Angle
Connection	Size	Slope
Enclosure	Volume	Area
Shape	Shape	Volume

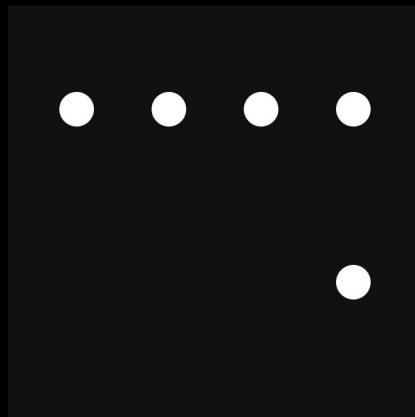
Ranking of Encoding

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Color Saturation	Color Hue
Angle	Color Hue	Connection
Slope	Connection	Enclosure
Size	Enclosure	Color Saturation
Volume	Length	Shape
Color Saturation	Angle	Length
Color Hue	Slope	Angle
Connection	Size	Slope
Enclosure	Volume	Area
Shape	Shape	Volume

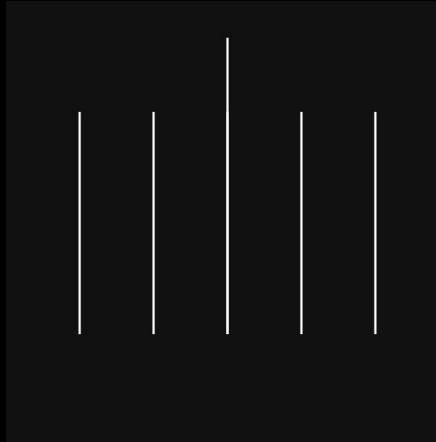


Quantitative Encoding

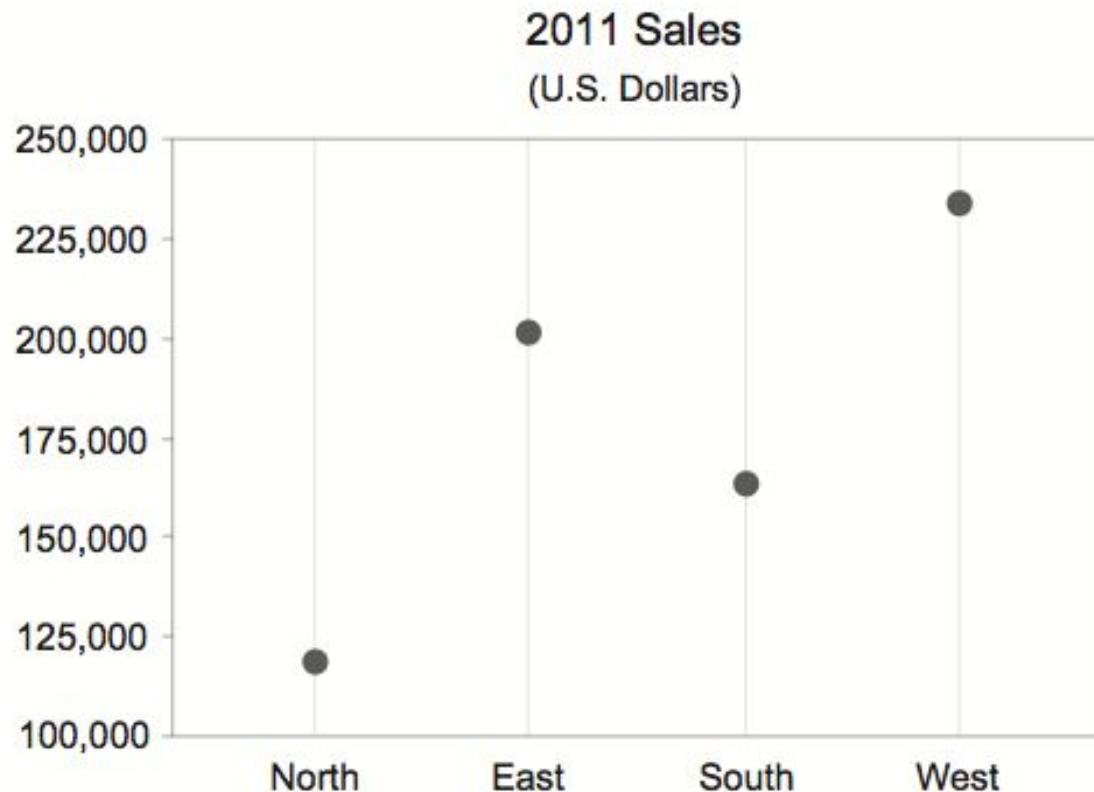
2D Position



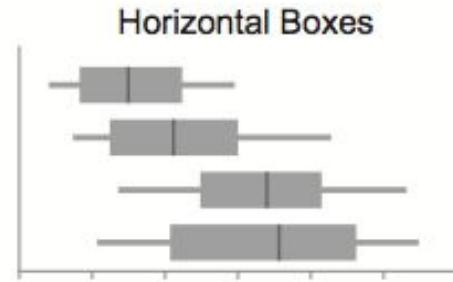
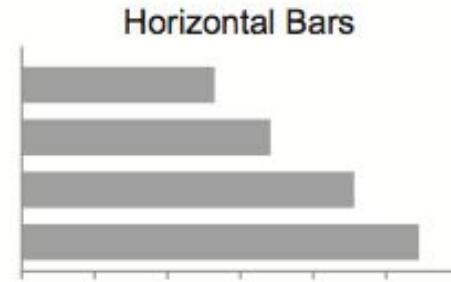
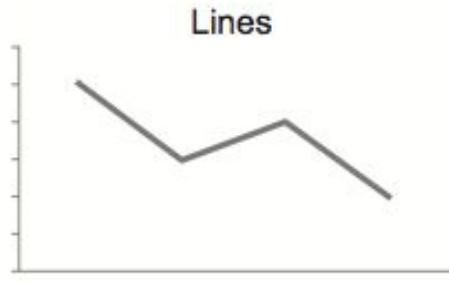
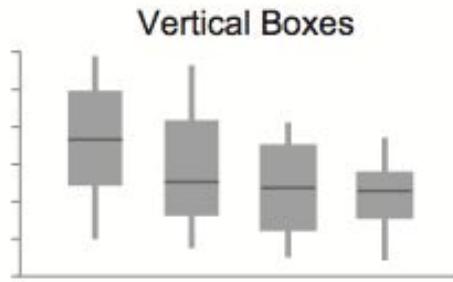
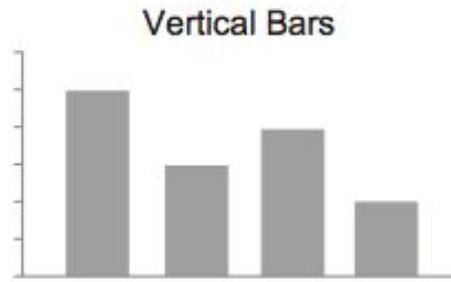
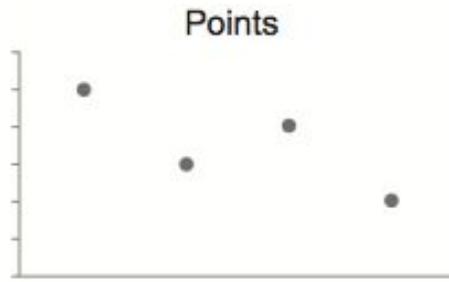
Length



Position can do it all



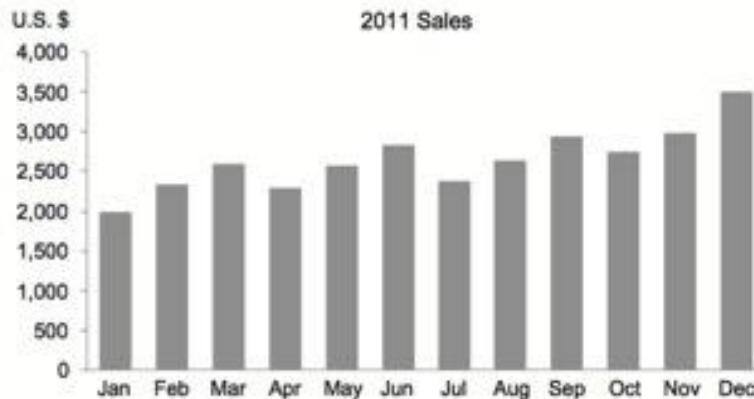
Quantitative Encoding



Use Position and Length mainly

Categorical Encoding

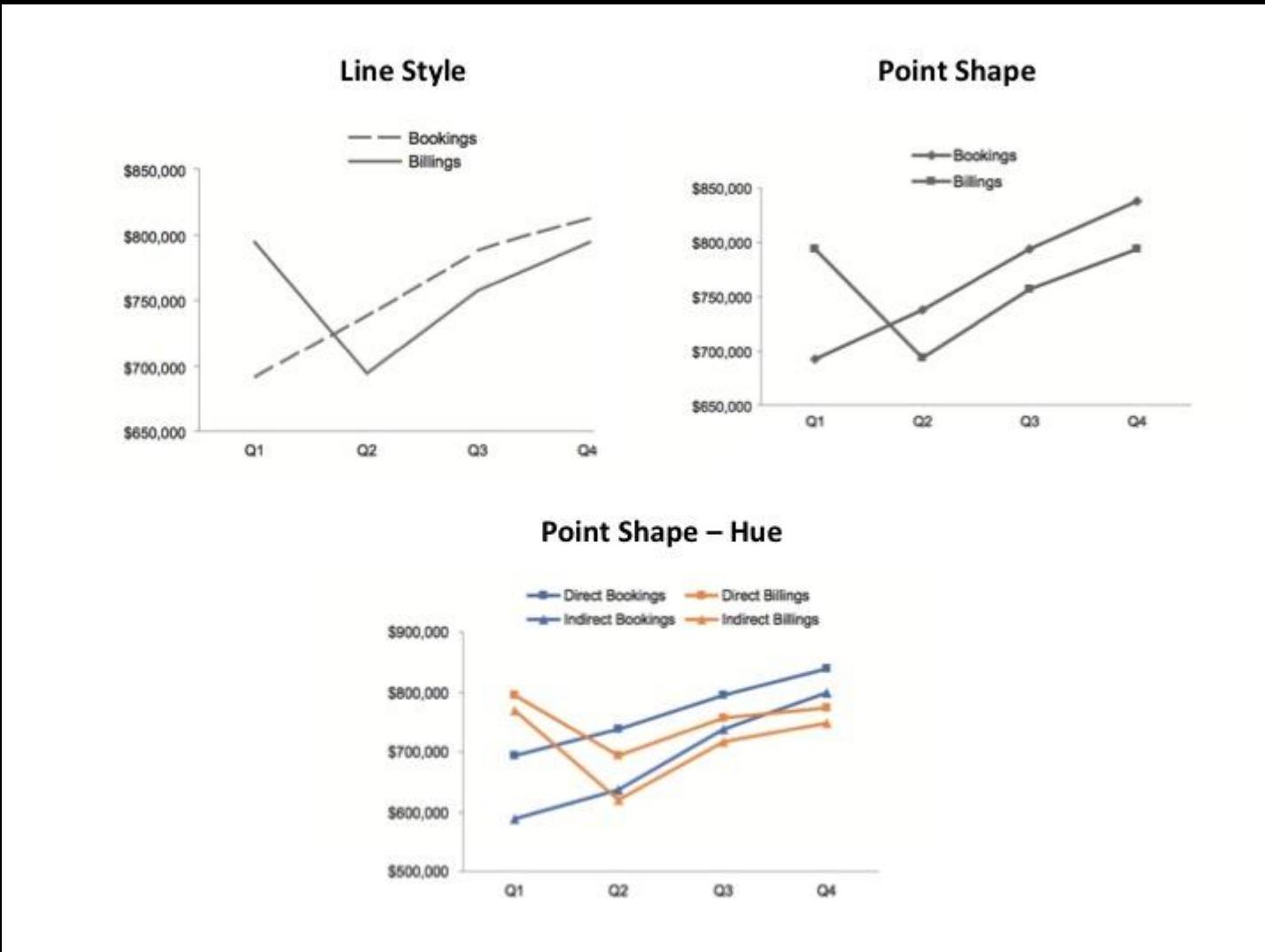
2D – Space



Hue

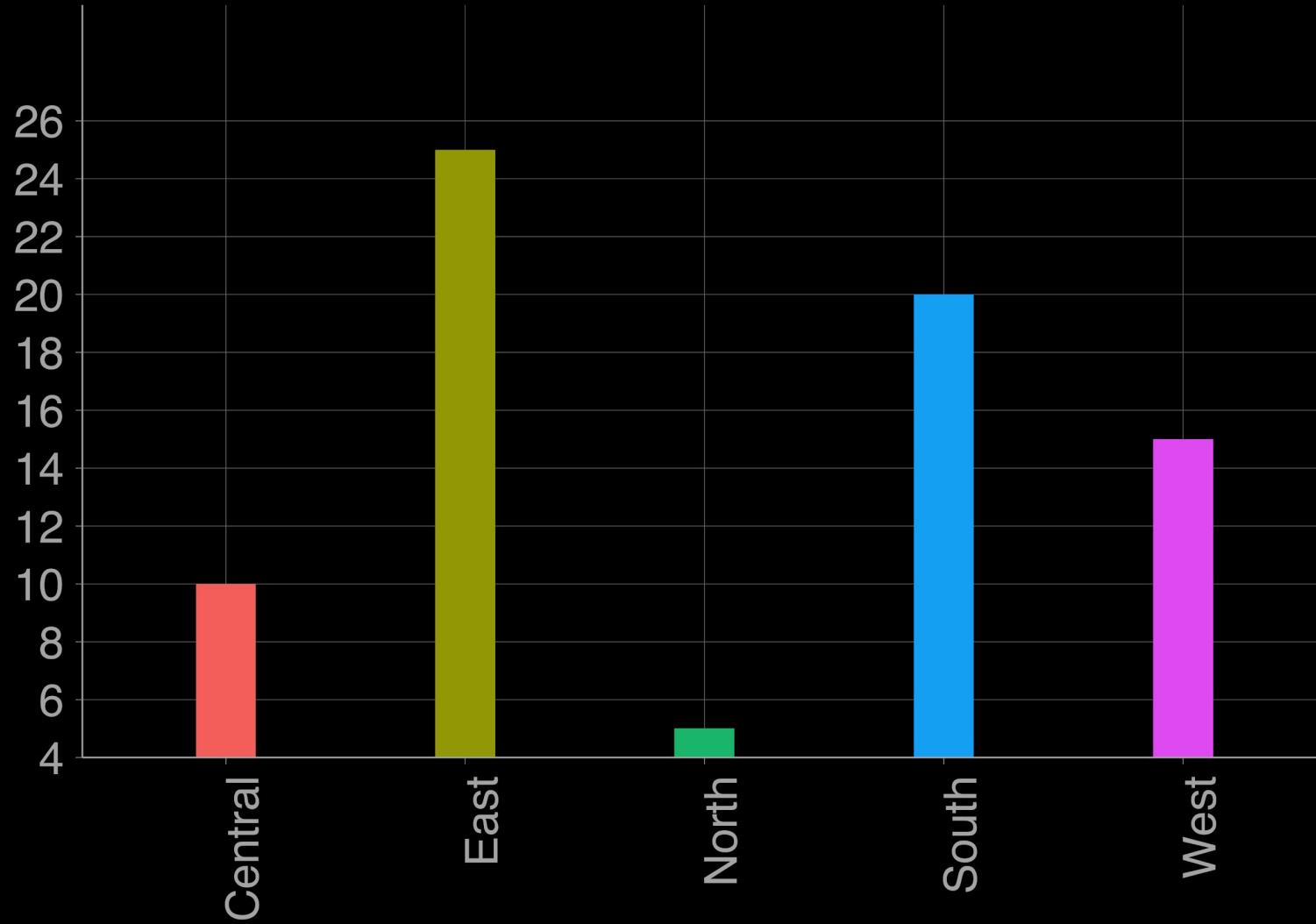


Categorical Encoding



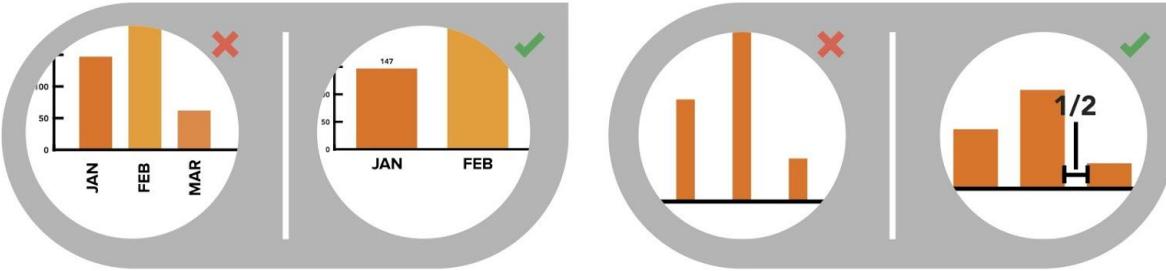
Aesthetic Best Practices

Review - Bar Chart



BAR CHART

DESIGN BEST PRACTICES

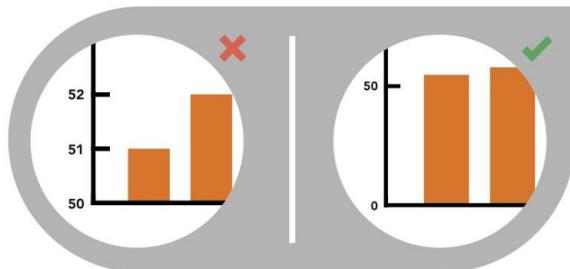


USE HORIZONTAL LABELS

Avoid steep diagonal or vertical type, as it can be difficult to read.

SPACE BARS APPROPRIATELY

Space between bars should be $\frac{1}{2}$ bar width.



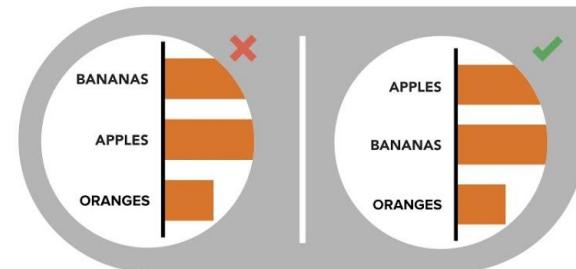
START THE Y-AXIS VALUE AT 0

Starting at a value above zero truncates the bars and doesn't accurately reflect the full value.



USE CONSISTENT COLORS

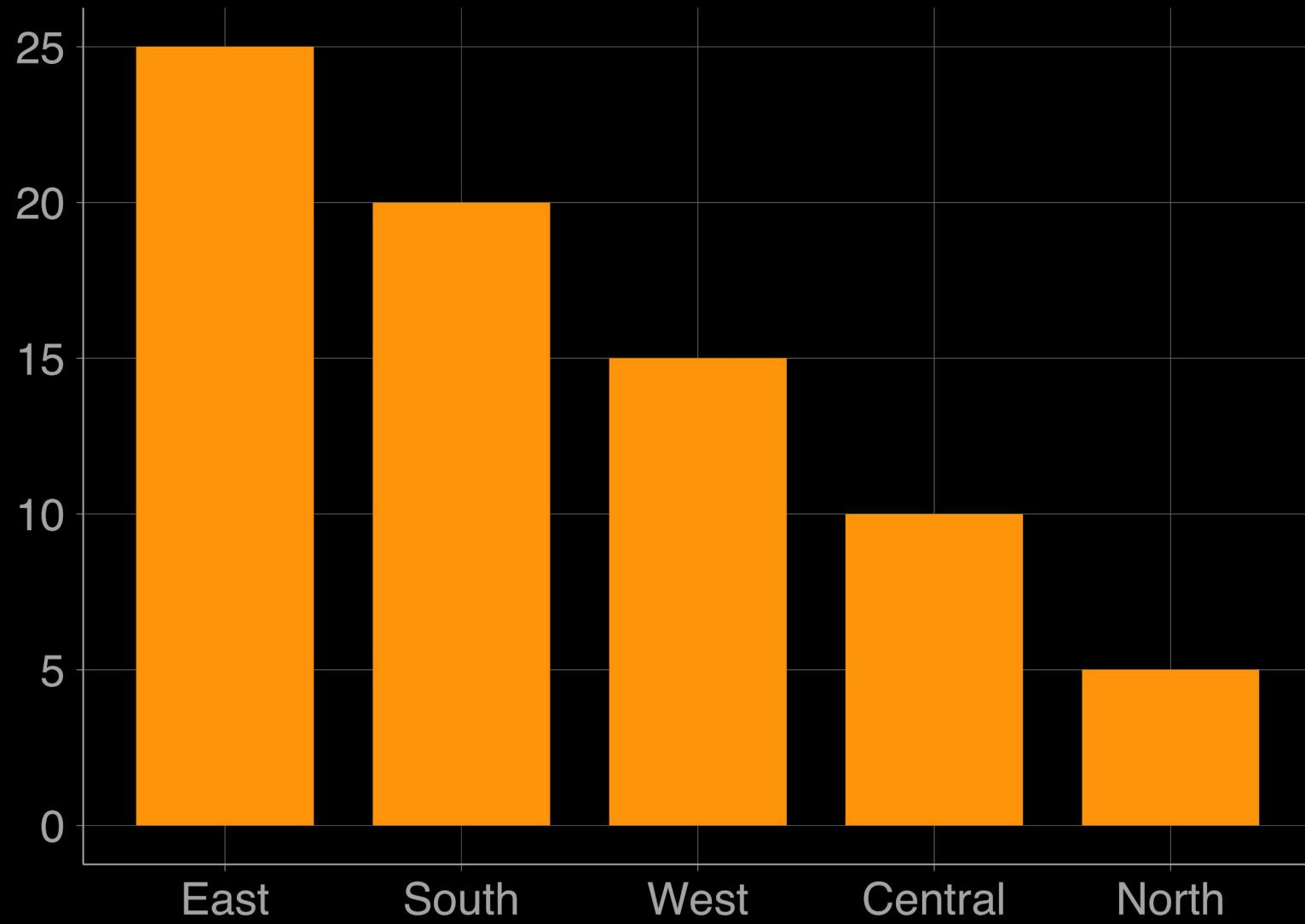
Use one color for bar charts. You may use an accent color to highlight a significant data point.



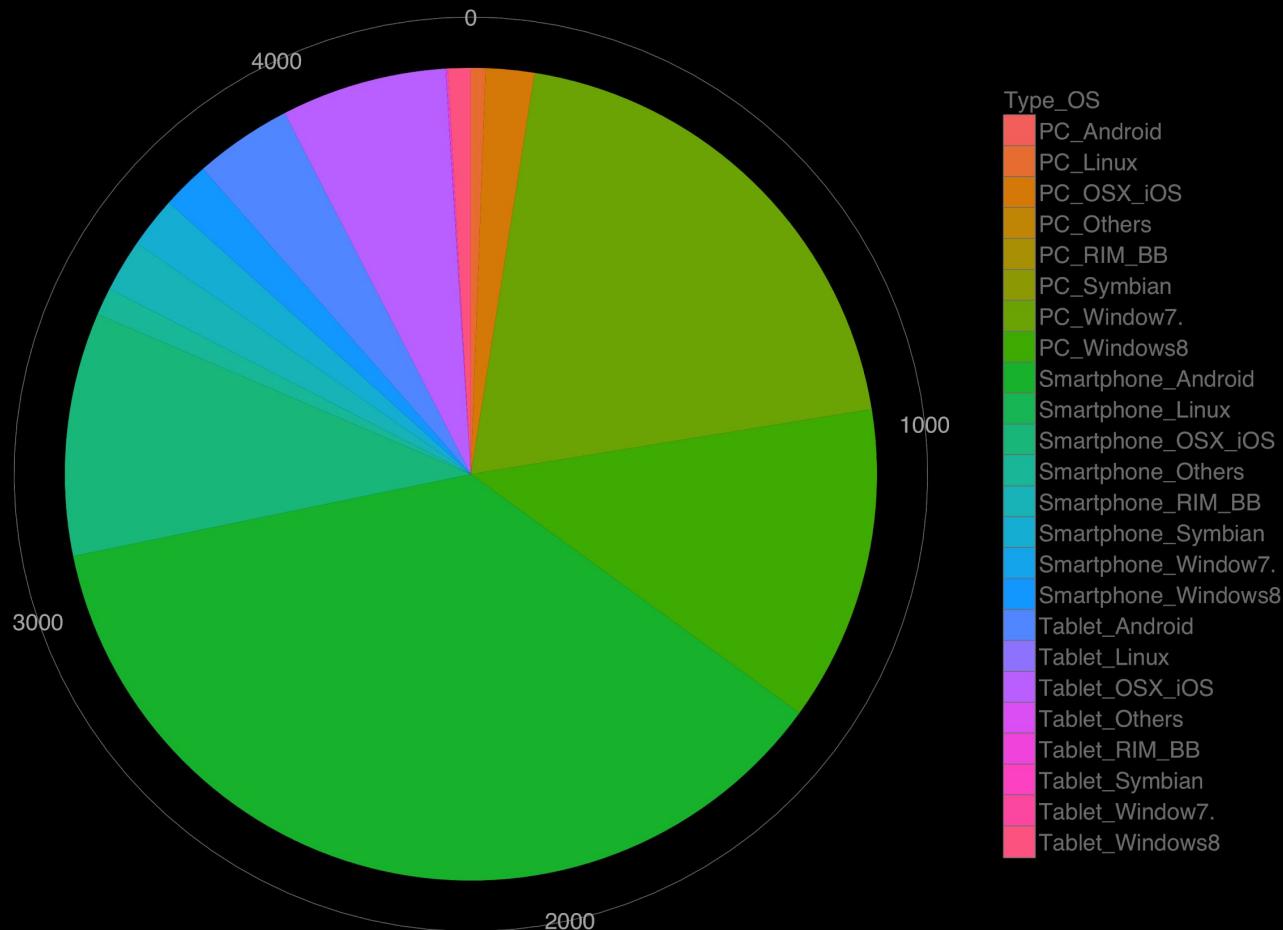
ORDER DATA APPROPRIATELY

Order categories alphabetically, sequentially, or by value.

Bar Chart - Aesthetics



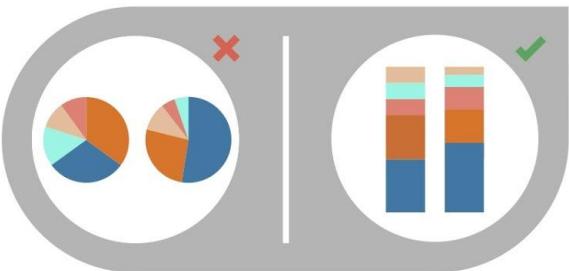
Review - Pie Chart





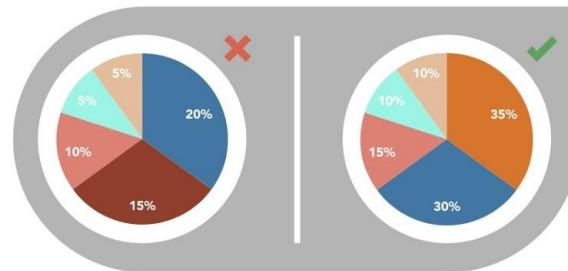
VISUALIZE NO MORE THAN 5 CATEGORIES PER CHART

It is difficult to differentiate between small values; depicting too many slices decreases the impact of the visualization. If needed, you can group smaller values into an “other” or “miscellaneous” category, but make sure it does not hide interesting or significant information.



DON'T USE MULTIPLE PIE CHARTS FOR COMPARISON

Slice sizes are very difficult to compare side-by-side. Use a stacked bar chart instead.



MAKE SURE ALL DATA ADDS UP TO 100%

Verify that values total 100% and that pie slices are sized proportionate to their corresponding value.



PIE CHART

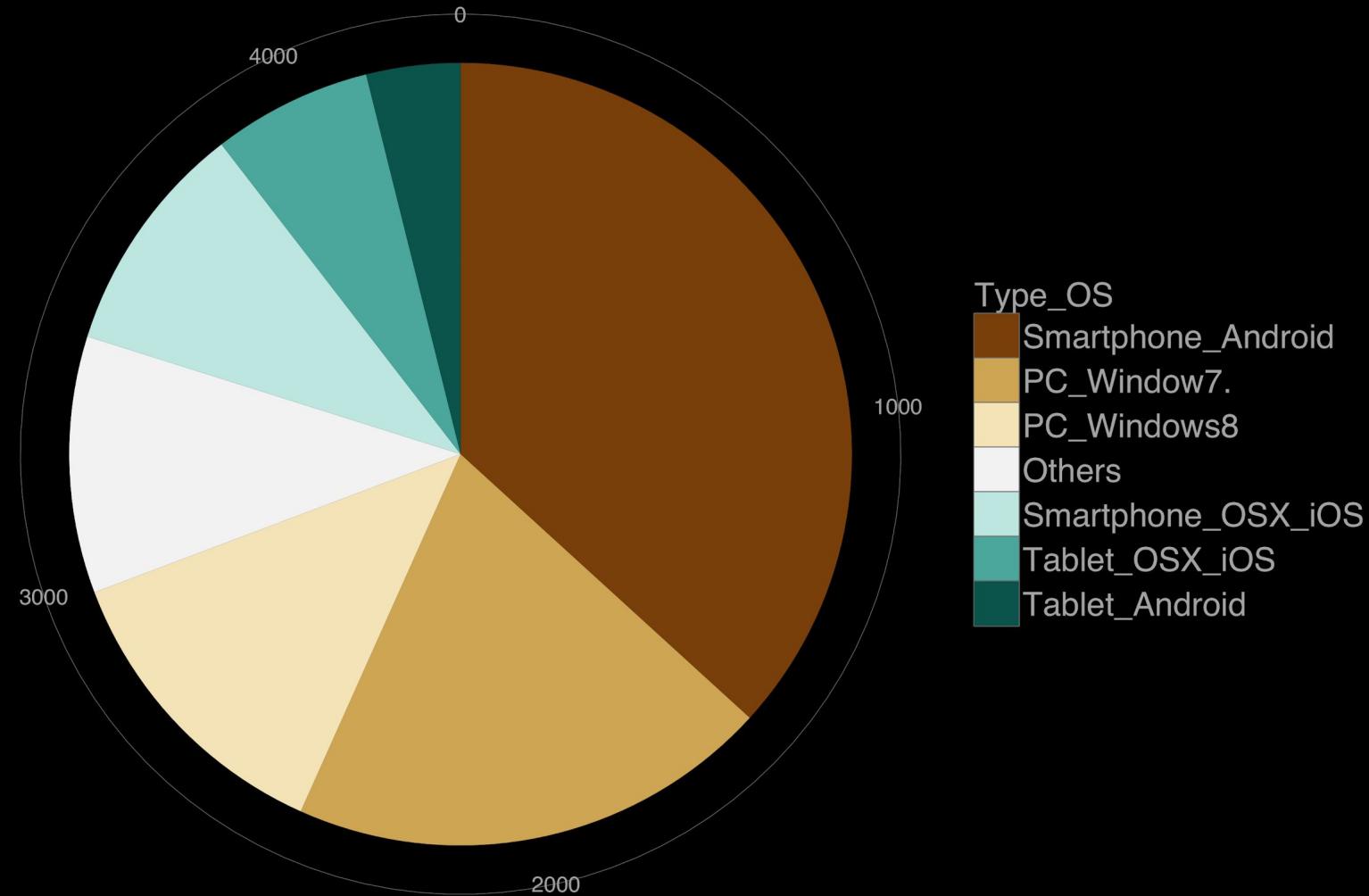
DESIGN BEST PRACTICES

ORDER SLICES CORRECTLY
There are two ways to order sections, both of which are meant to aid comprehension:

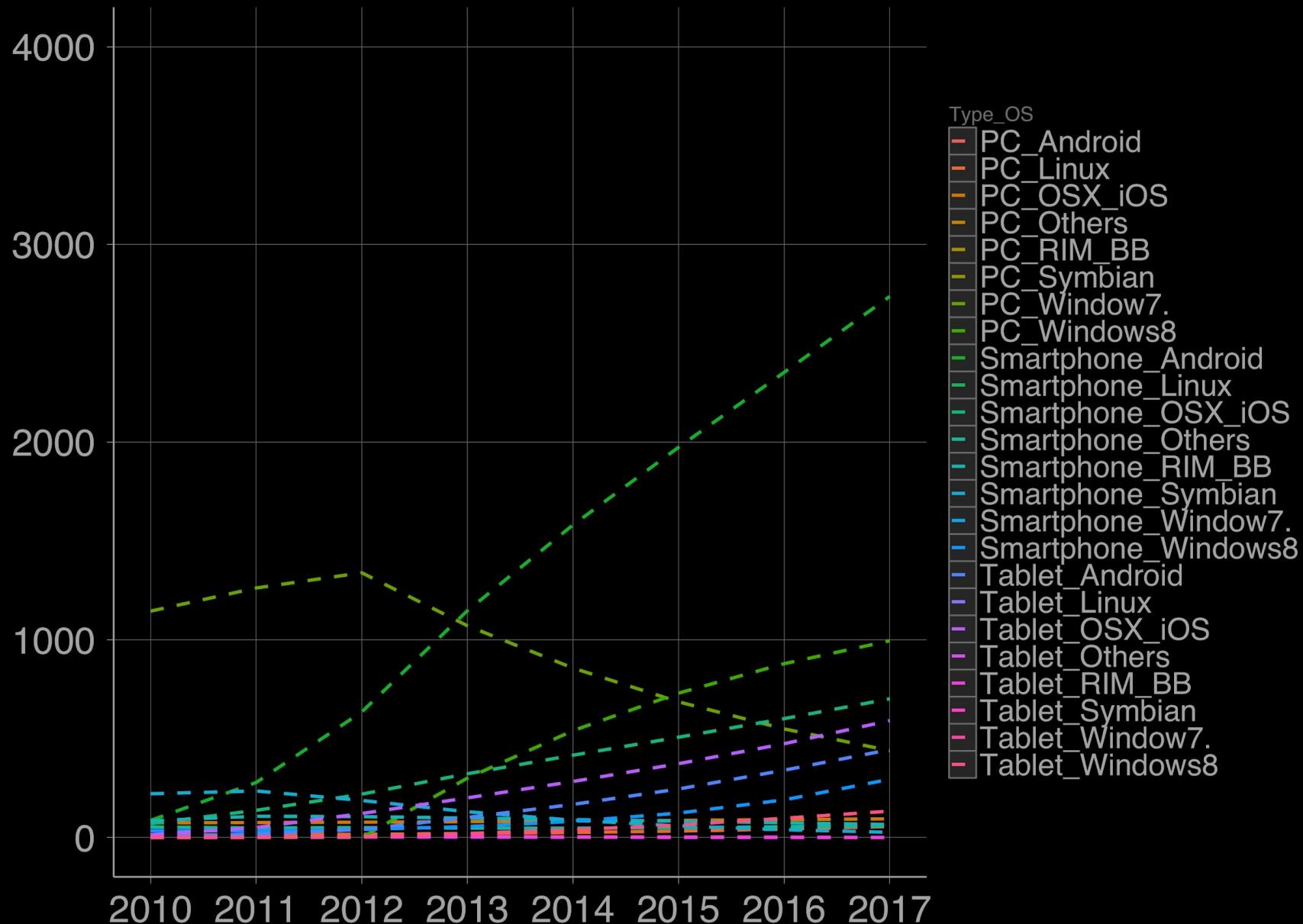
OPTION 1
Place the largest section at 12 o'clock, going clockwise. Place the second largest section at 12 o'clock, going counterclockwise. The remaining sections can be placed below, continuing counterclockwise.

OPTION 2
Start the largest section at 12 o'clock, going clockwise. Place remaining sections in descending order, going clockwise.

Pie Chart - Aesthetics

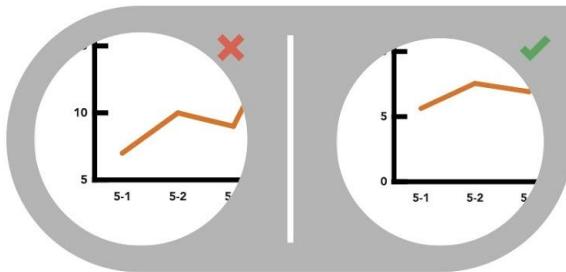


Review - Line Chart



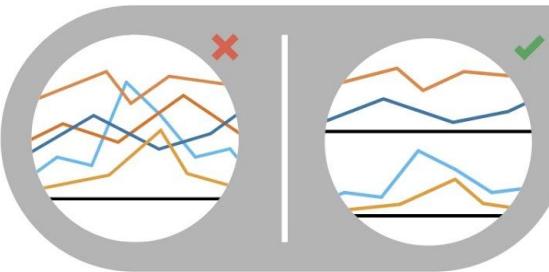
LINE CHART

DESIGN BEST PRACTICES



INCLUDE A ZERO BASELINE IF POSSIBLE

Although a line chart does not have to start at a zero baseline, it should be included if possible. If relatively small fluctuations in data are meaningful (e.g., in stock market data), you may truncate the scale to showcase these variances.



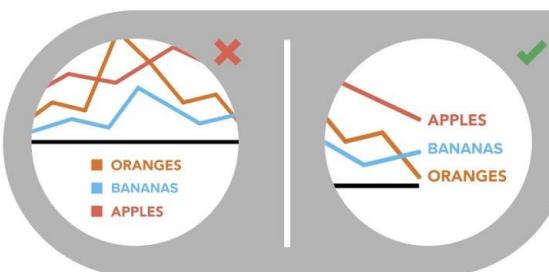
DON'T PLOT MORE THAN 4 LINES

If you need to display more, break them out into separate charts for better comparison.



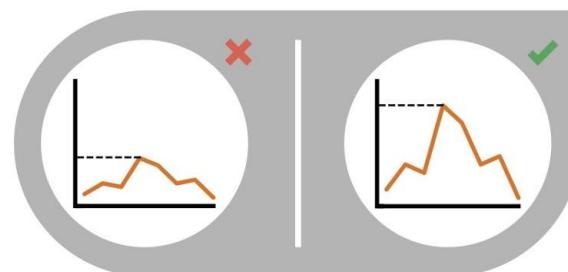
USE SOLID LINES ONLY

Dashed and dotted lines can be distracting.



LABEL THE LINES DIRECTLY

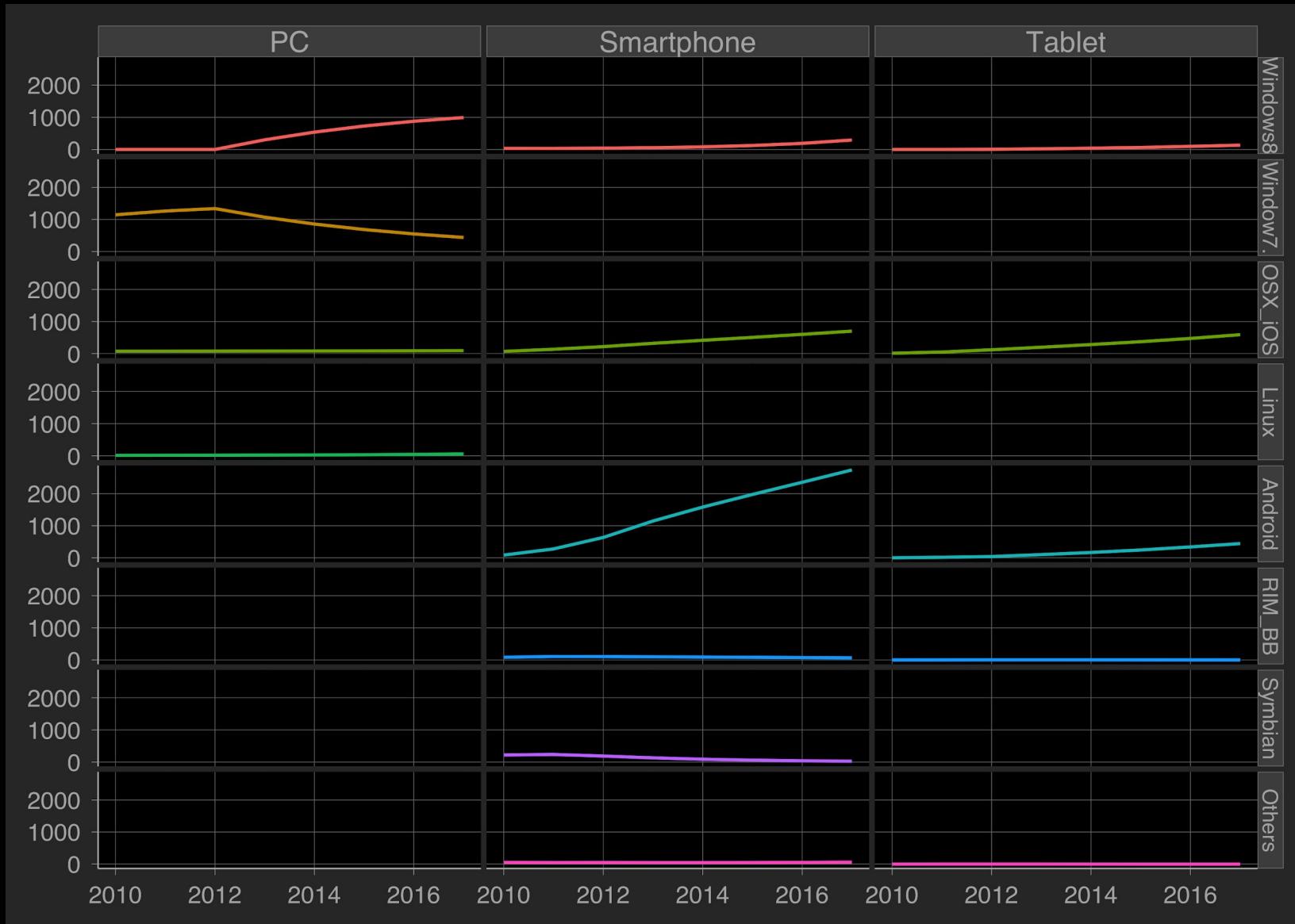
This lets readers quickly identify lines and corresponding labels instead of referencing a legend.



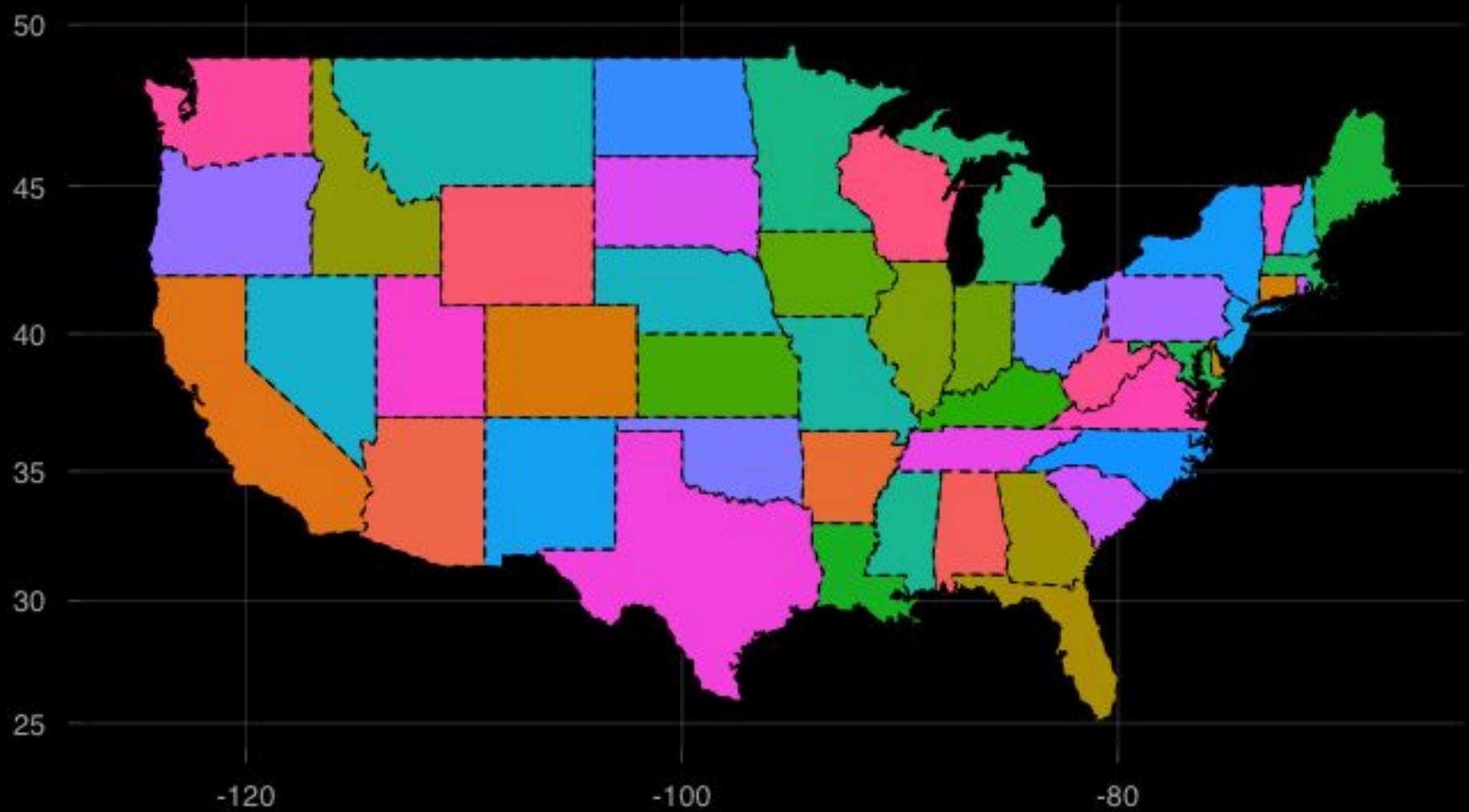
USE THE RIGHT HEIGHT

Plot all data points so that the line chart takes up approximately two-thirds of the y-axis' total scale.

Line Chart - Aesthetics

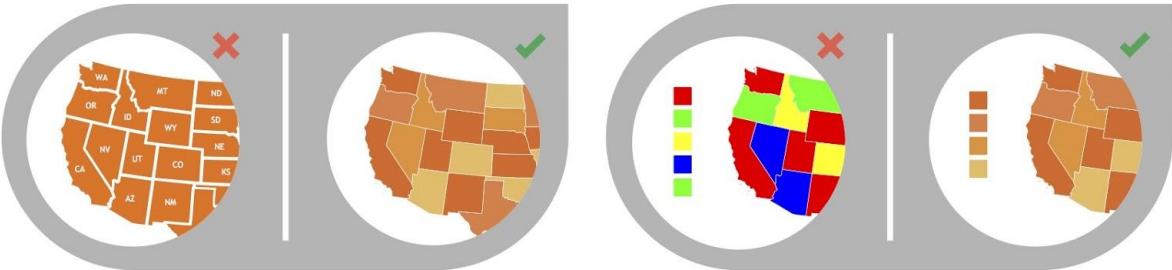


Review - Choropleth



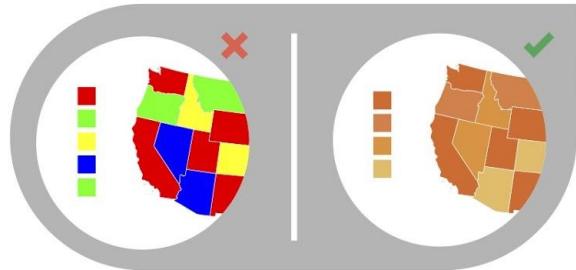
HEAT MAP

DESIGN BEST PRACTICES



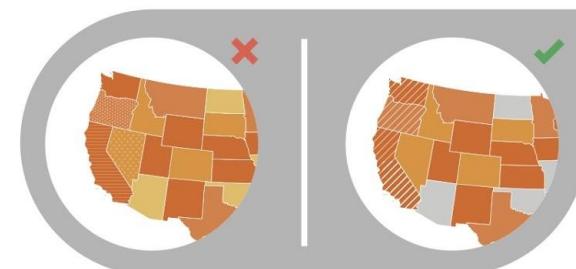
USE A SIMPLE MAP OUTLINE

These lines are meant to frame the data, not distract.



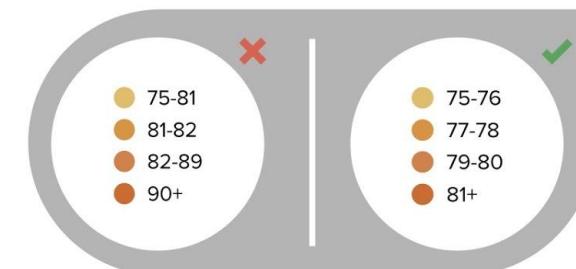
SELECT COLORS APPROPRIATELY

Some colors stand out more than others, giving unnecessary weight to that data. Instead, use a single color with varying shade or a spectrum between two analogous colors to show intensity. Also remember to intuitively code color intensity according to values.



USE PATTERNS SPARINGLY

A pattern overlay that indicates a second variable is acceptable, but using multiple is overwhelming and distracting.

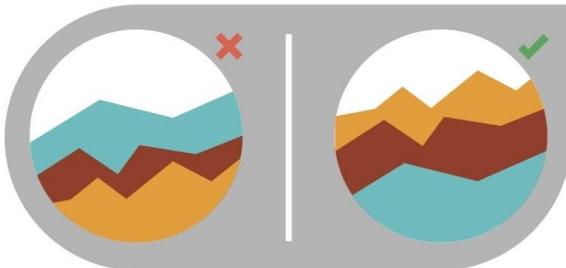


CHOOSE APPROPRIATE DATA RANGES

Select 3-5 numerical ranges that enable fairly even distribution of data between them. Use +/- signs to extend high and low ranges.

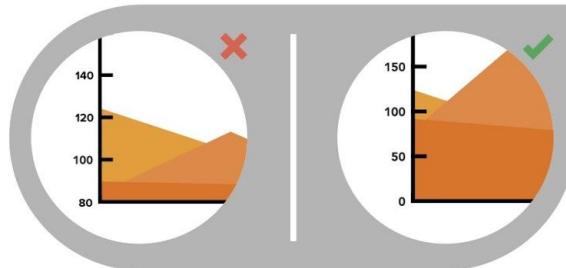
AREA CHART

DESIGN BEST PRACTICES



MAKE IT EASY TO READ

In stacked area charts, arrange data to position categories with highly variable data on the top and low variability on the bottom.



START Y-AXIS VALUE AT 0

Starting the axis above zero truncates the visualization of values.



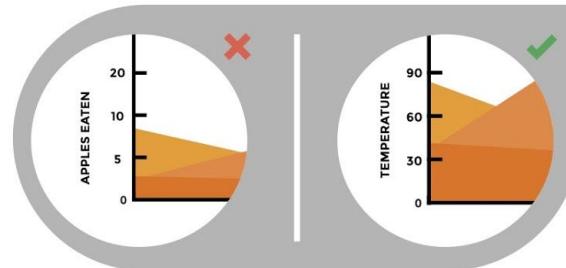
DON'T DISPLAY MORE THAN 4 DATA CATEGORIES

Too many will result in a cluttered visual that is difficult to decipher.



USE TRANSPARENT COLORS

In standard area charts, ensure data isn't obscured in the background by ordering thoughtfully and using transparency.

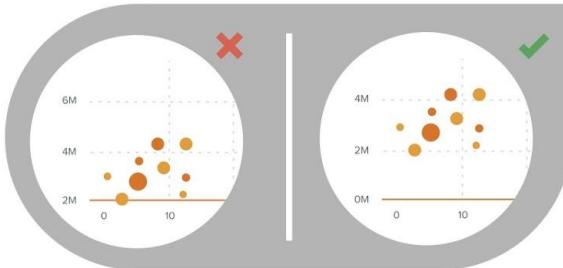


DON'T USE AREA CHARTS TO DISPLAY DISCRETE DATA

The connected lines imply intermediate values, which only exist with continuous data.

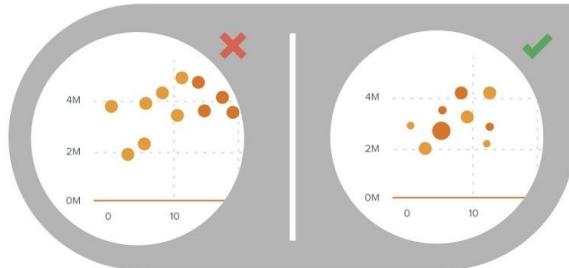
SCATTER PLOT

DESIGN BEST PRACTICES



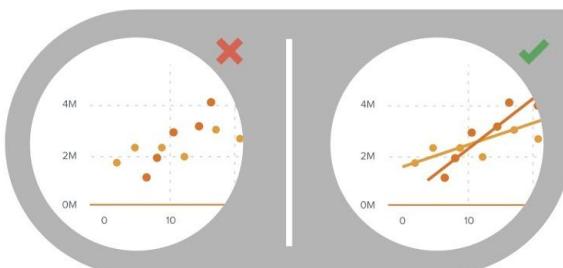
START Y-AXIS VALUE AT 0

Starting the axis above zero truncates the visualization of values.



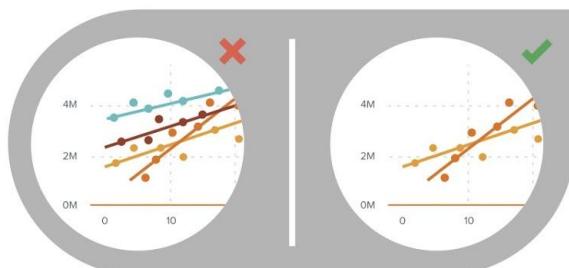
INCLUDE MORE VARIABLES

Use size and dot color to encode additional data variables.



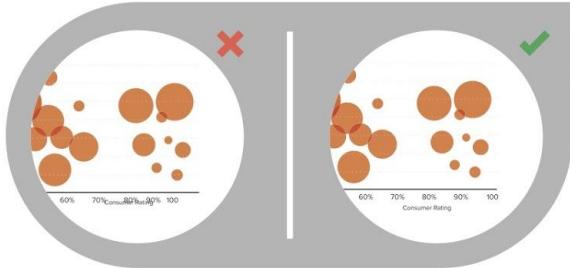
USE TREND LINES

These help draw correlation between the variables to show trends.



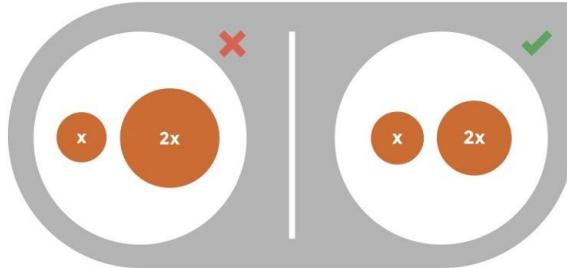
DON'T COMPARE MORE THAN 2 TREND LINES

Too many lines make data difficult to interpret.



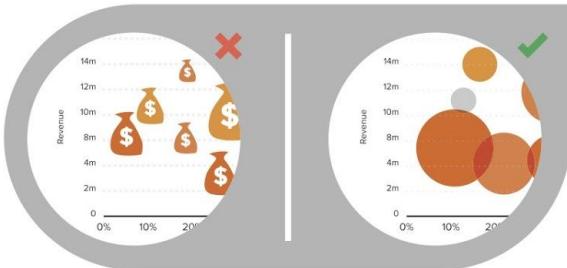
MAKE SURE LABELS ARE VISIBLE

All labels should be unobstructed and easily identified with the corresponding bubble.



SIZE BUBBLES APPROPRIATELY

Bubbles should be scaled according to area, not diameter.



DON'T USE ODD SHAPES

Avoid adding too much detail or using shapes that are not entirely circular; this can lead to inaccuracies.

BUBBLE CHART

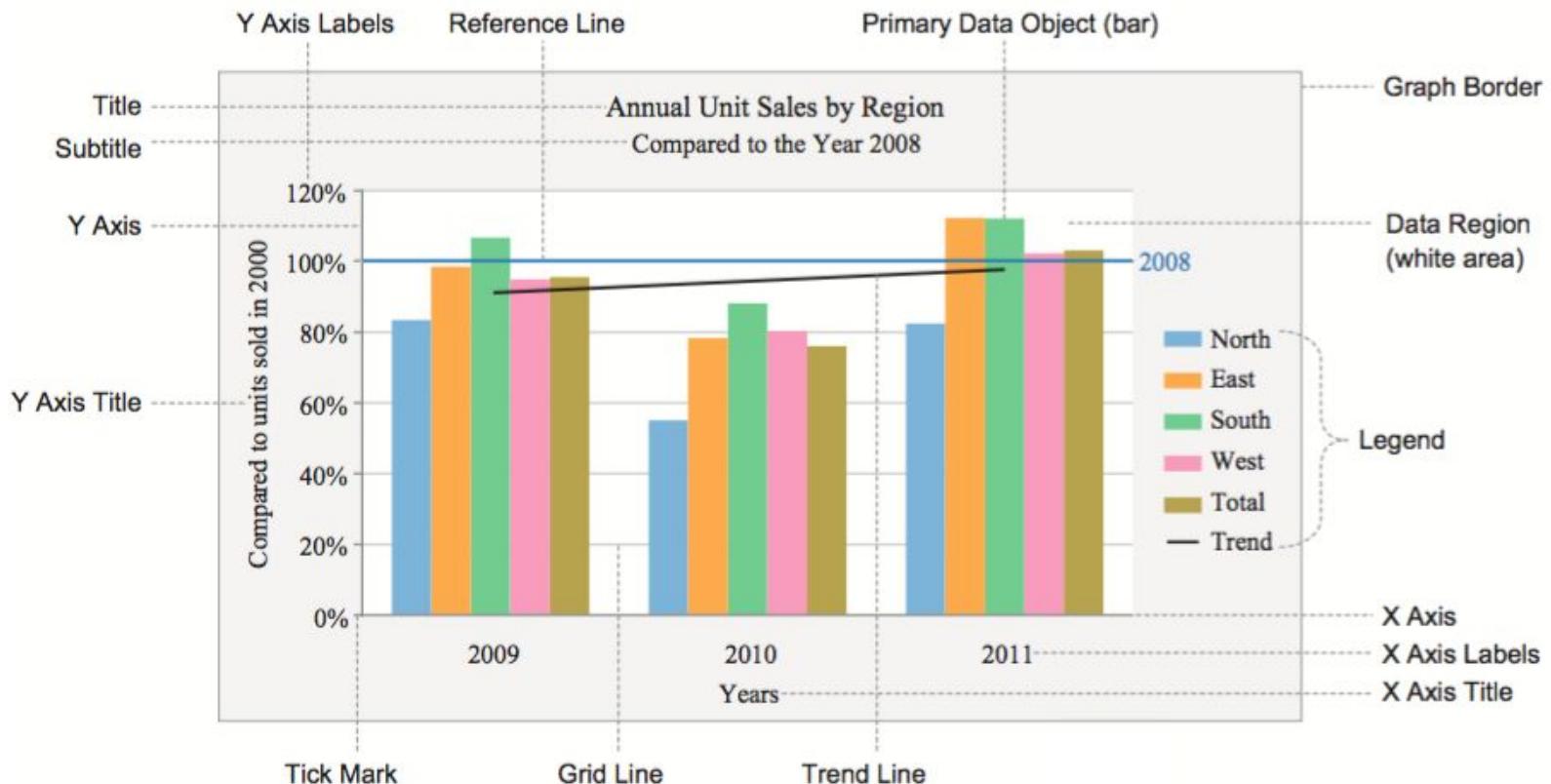
DESIGN BEST PRACTICES

Data-Ink & Chart Junk Principles

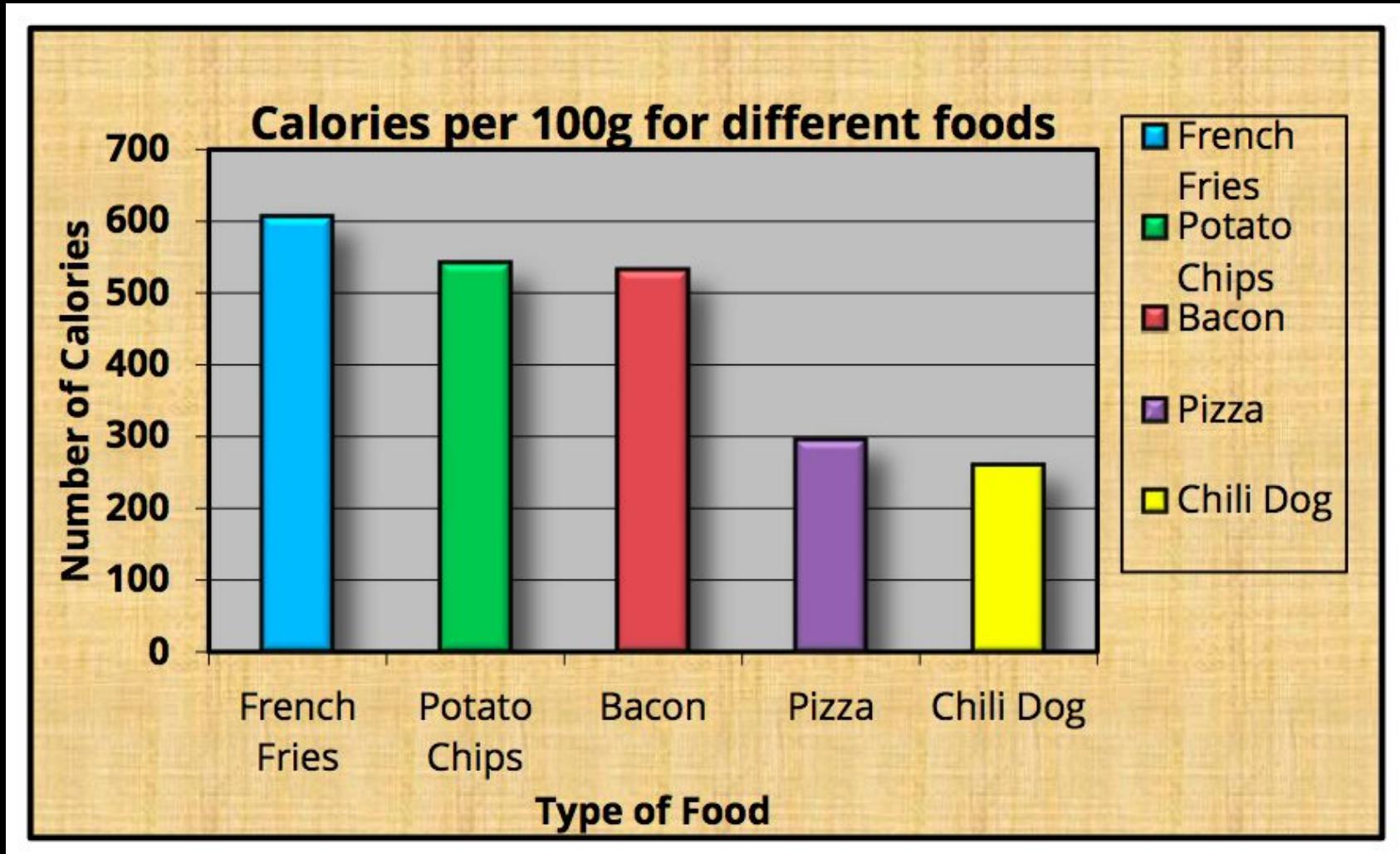
“Above all else show
the data”

- Edward Tufte

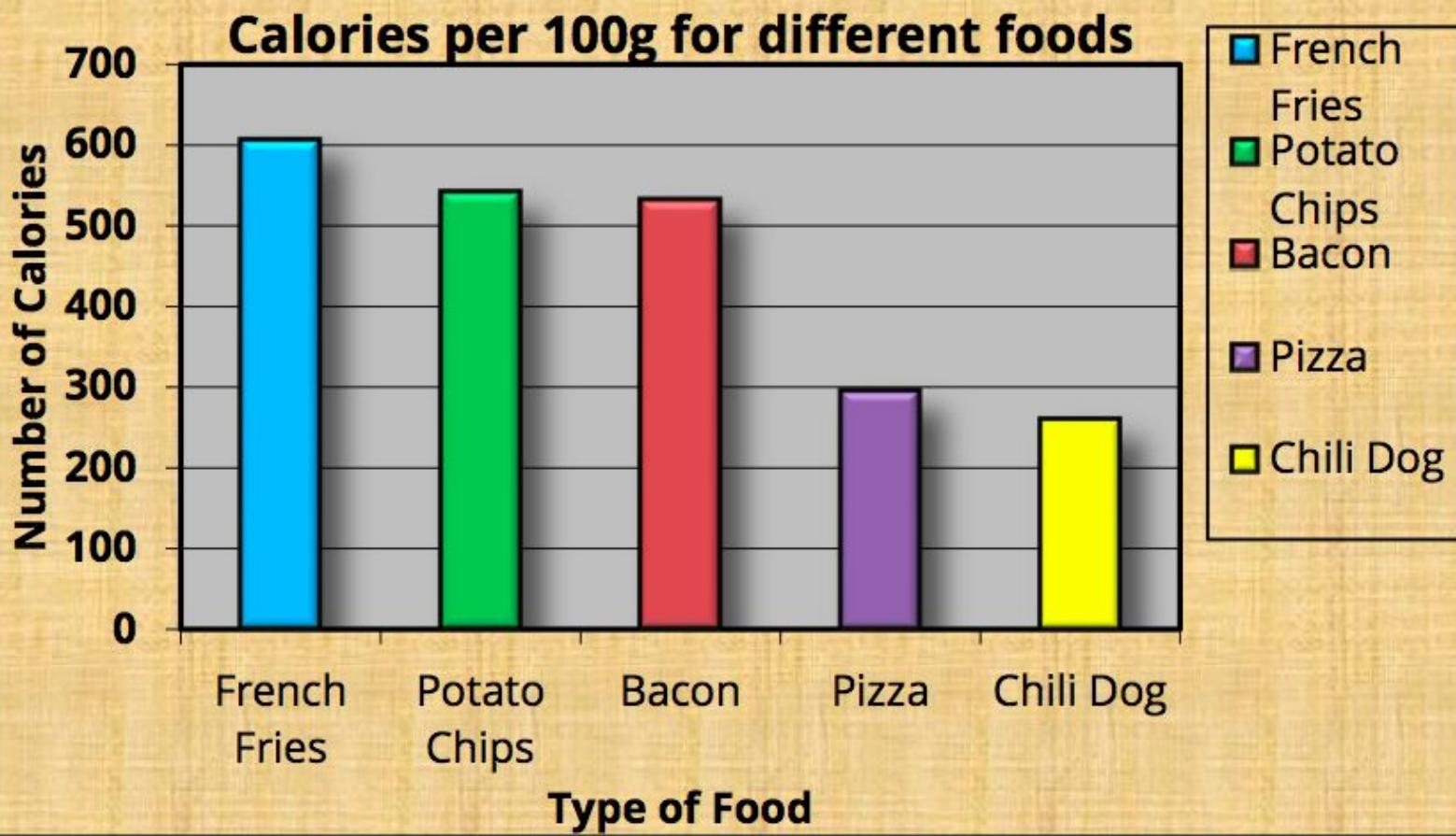
Anatomy of a Chart



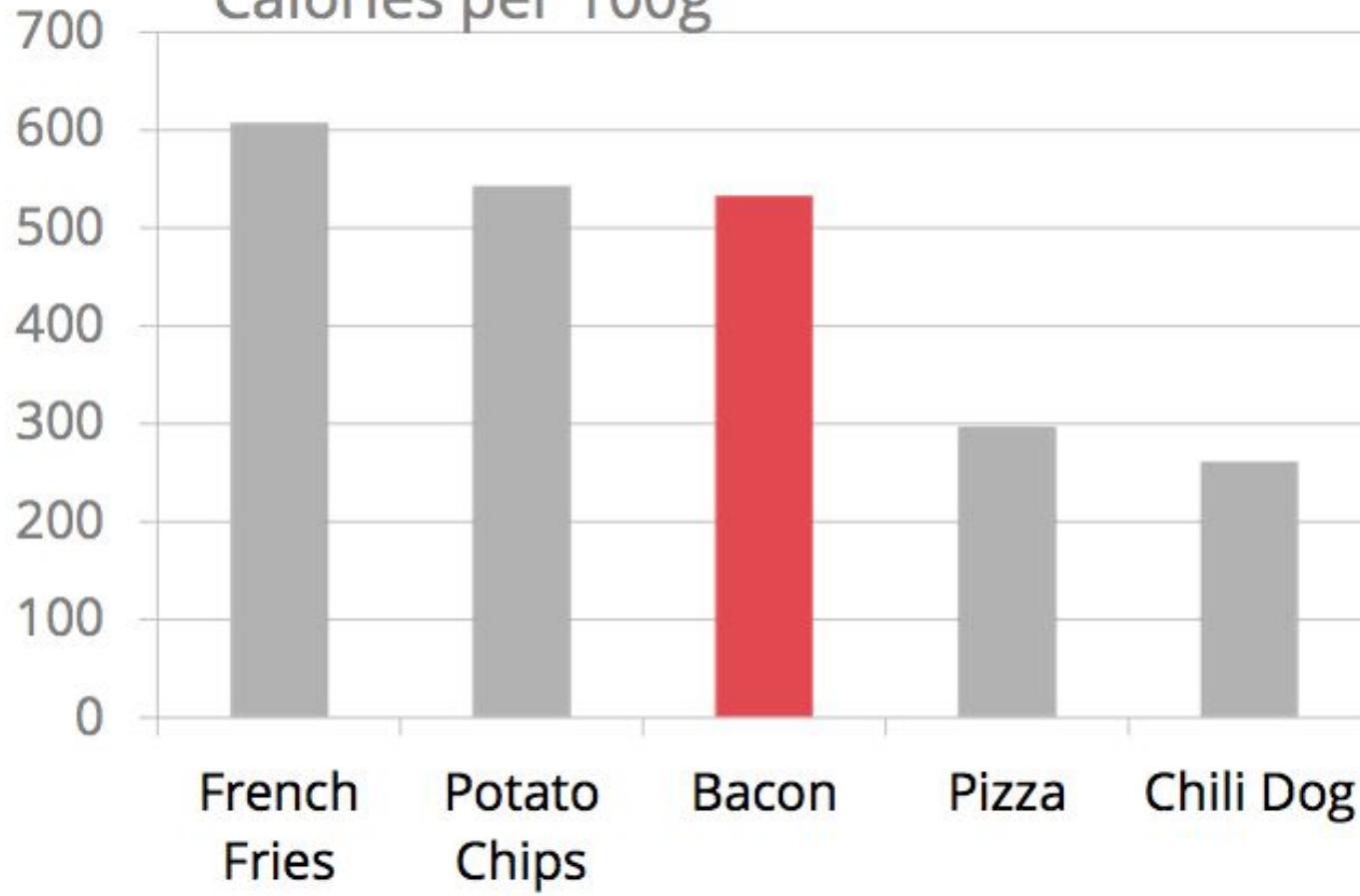
How can we improve this?



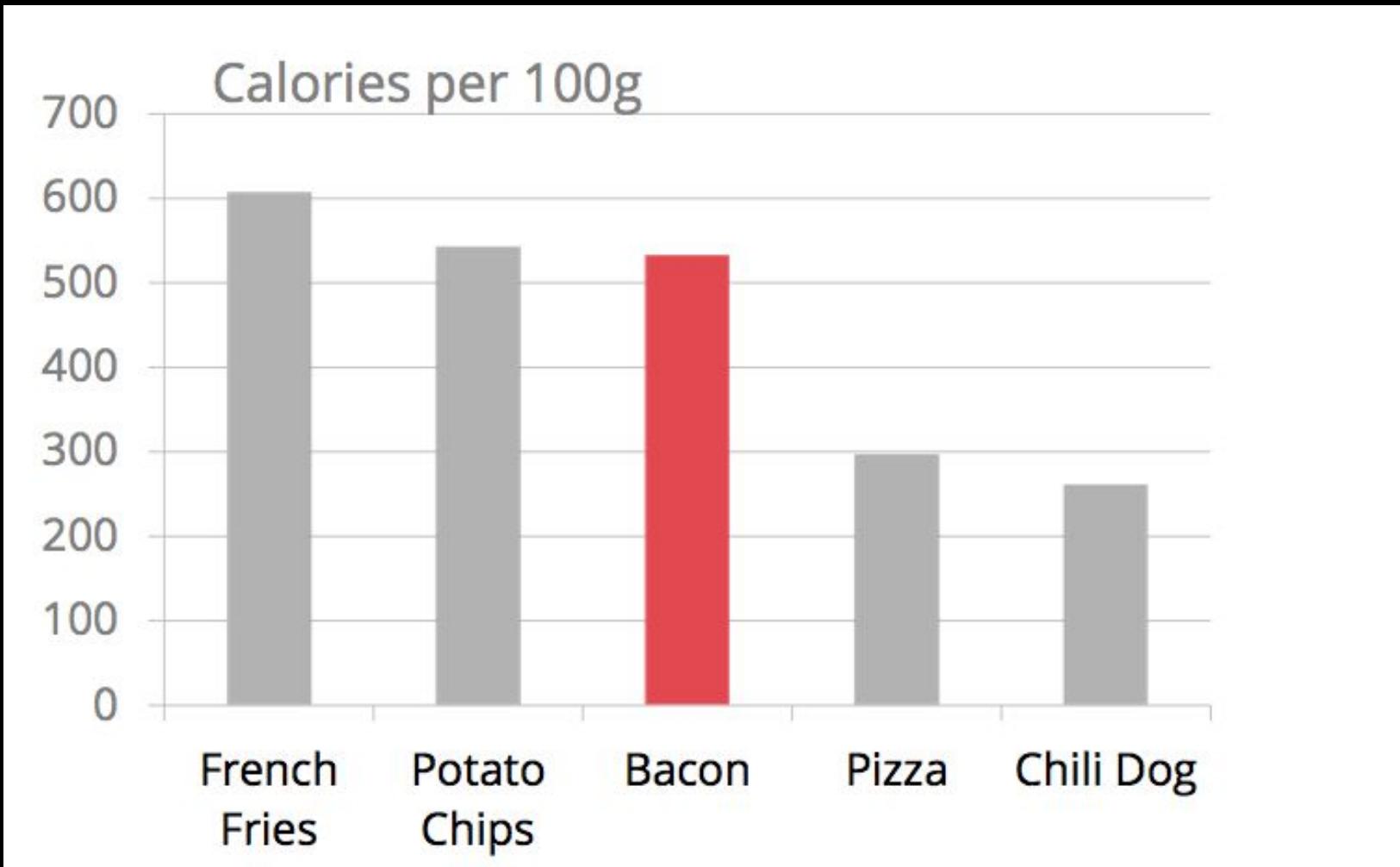
Source: [Darkhorse Analytics](#)



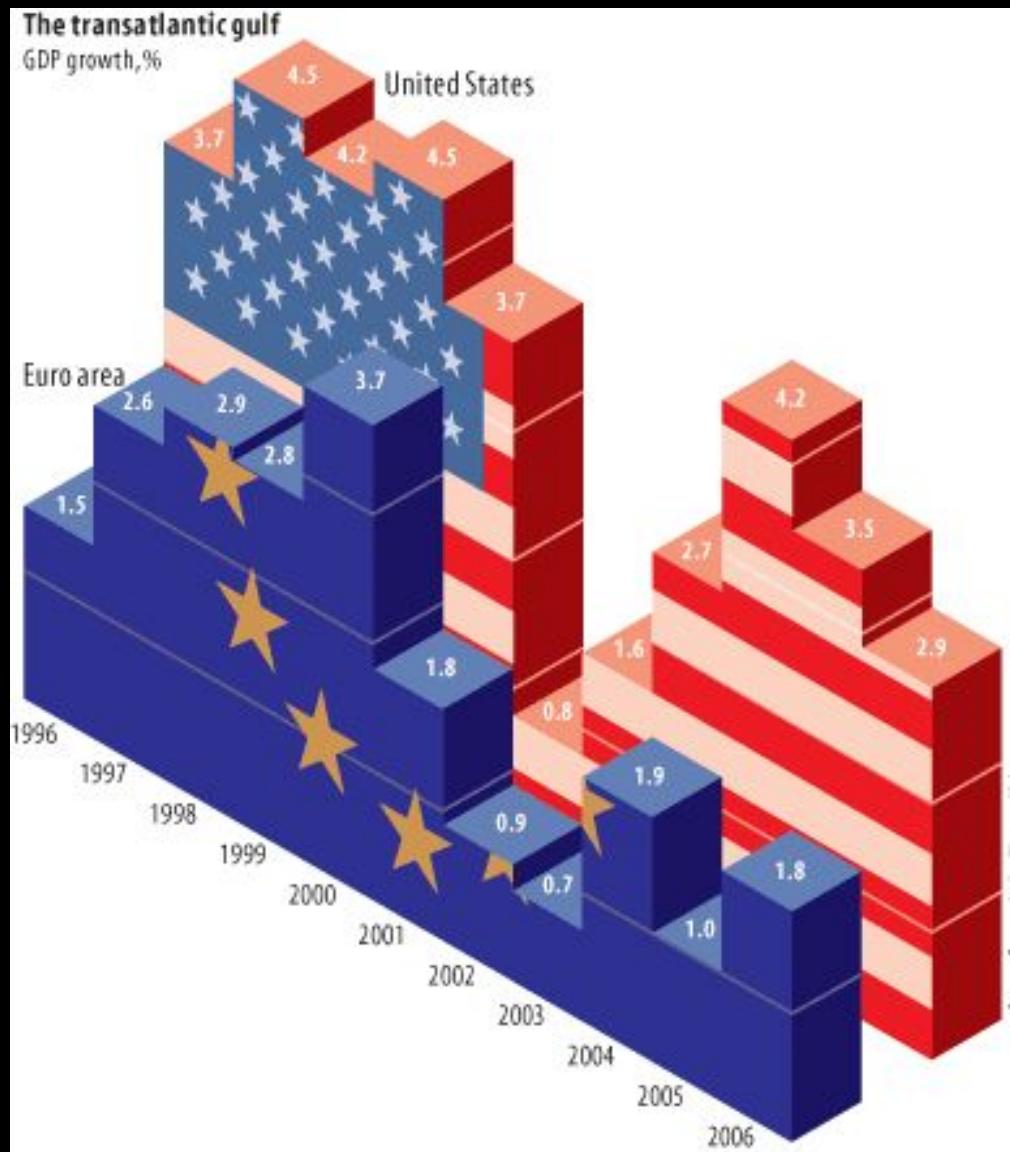
Calories per 100g



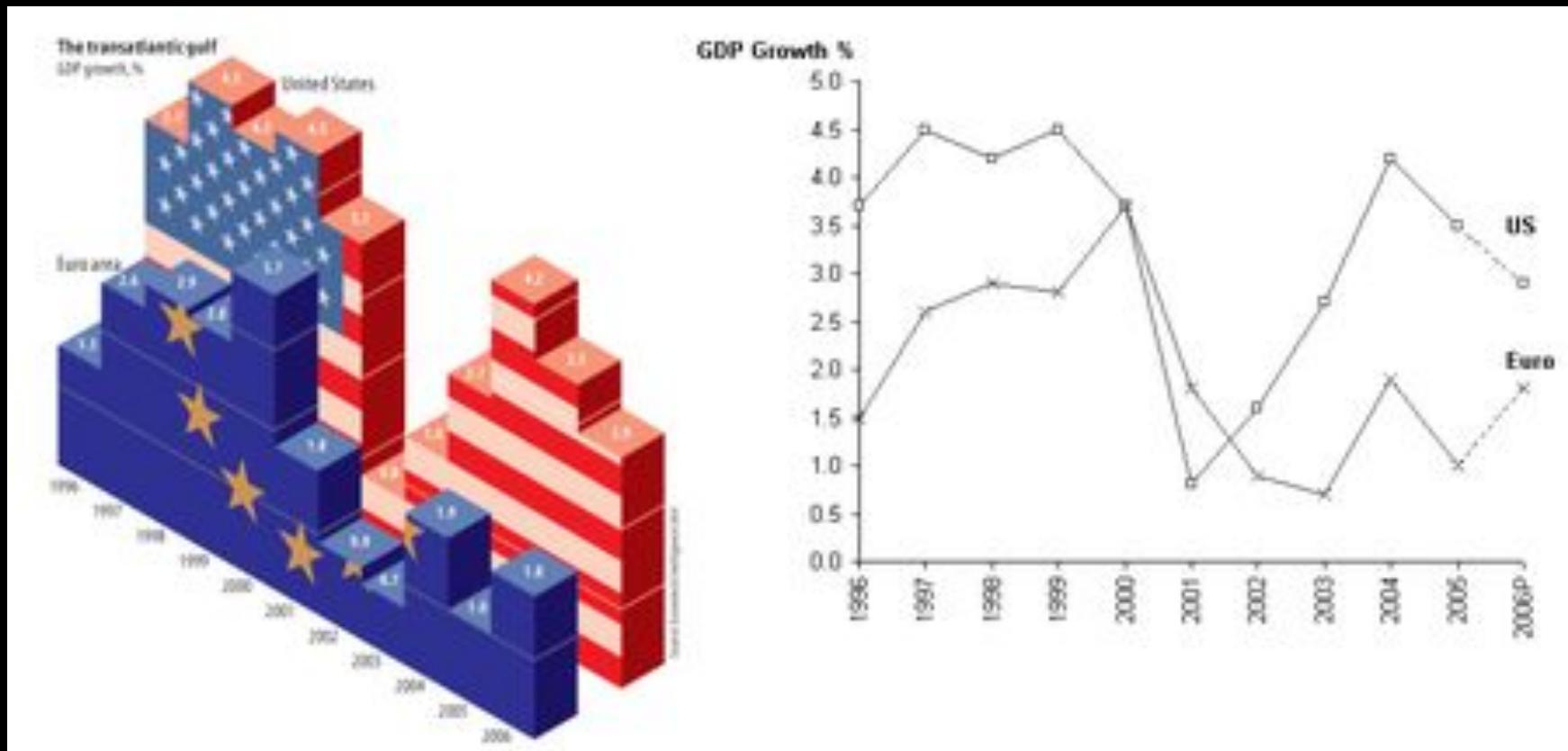
Data-Ink Improved



Avoid Chart Junk



Avoid Chart Junk



Avoid Chart Junk

Big Mac index

Local currency under (-)/ over (+) valuation against the dollar, %

Big Mac price*, \$

7.02	Norway
6.30	Switzerland
4.84†	Euro area
3.98	Australia
3.97	Canada
3.86	Hungary
3.83	Turkey
3.67	Britain
3.58‡	United States
3.50	Japan
3.19	Singapore
2.99	United Arab Emirates
2.98	South Korea
2.86	Poland
2.50	Mexico
2.46	South Africa
2.38	Egypt
2.36	Taiwan
2.34	Russia
2.24	Indonesia
2.11	Thailand
2.08	Malaysia
1.83	China

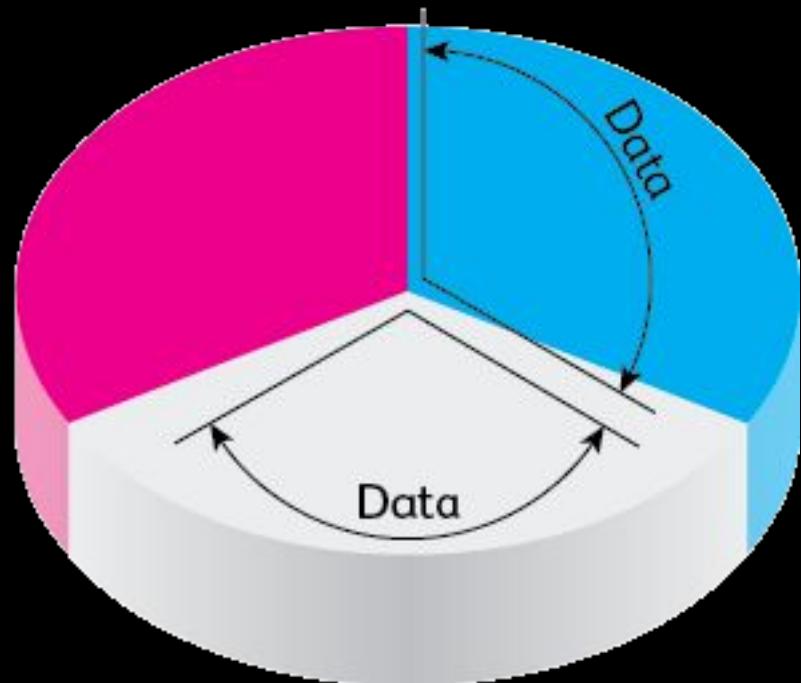
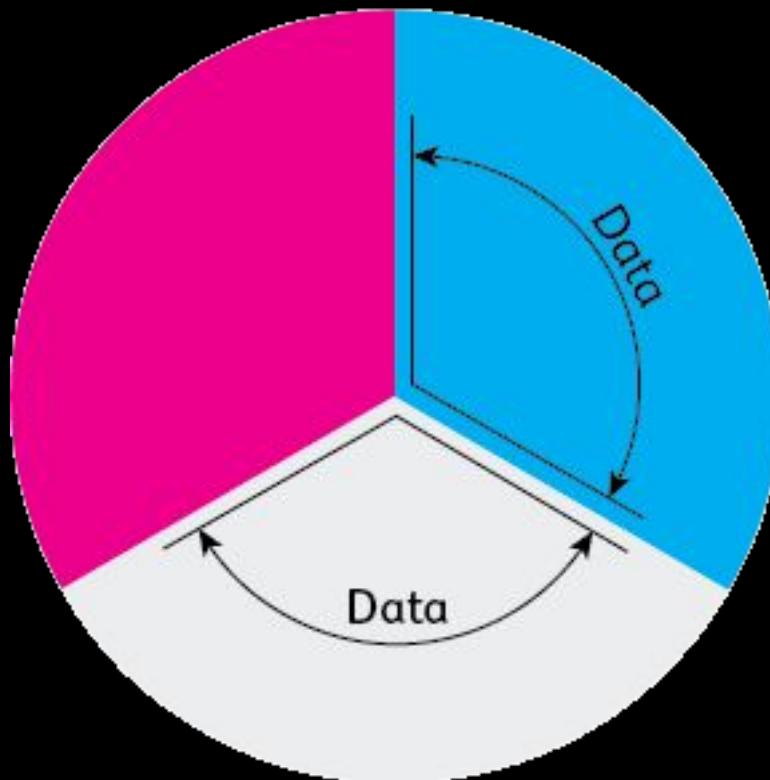


*At market exchange rate (January 5th)

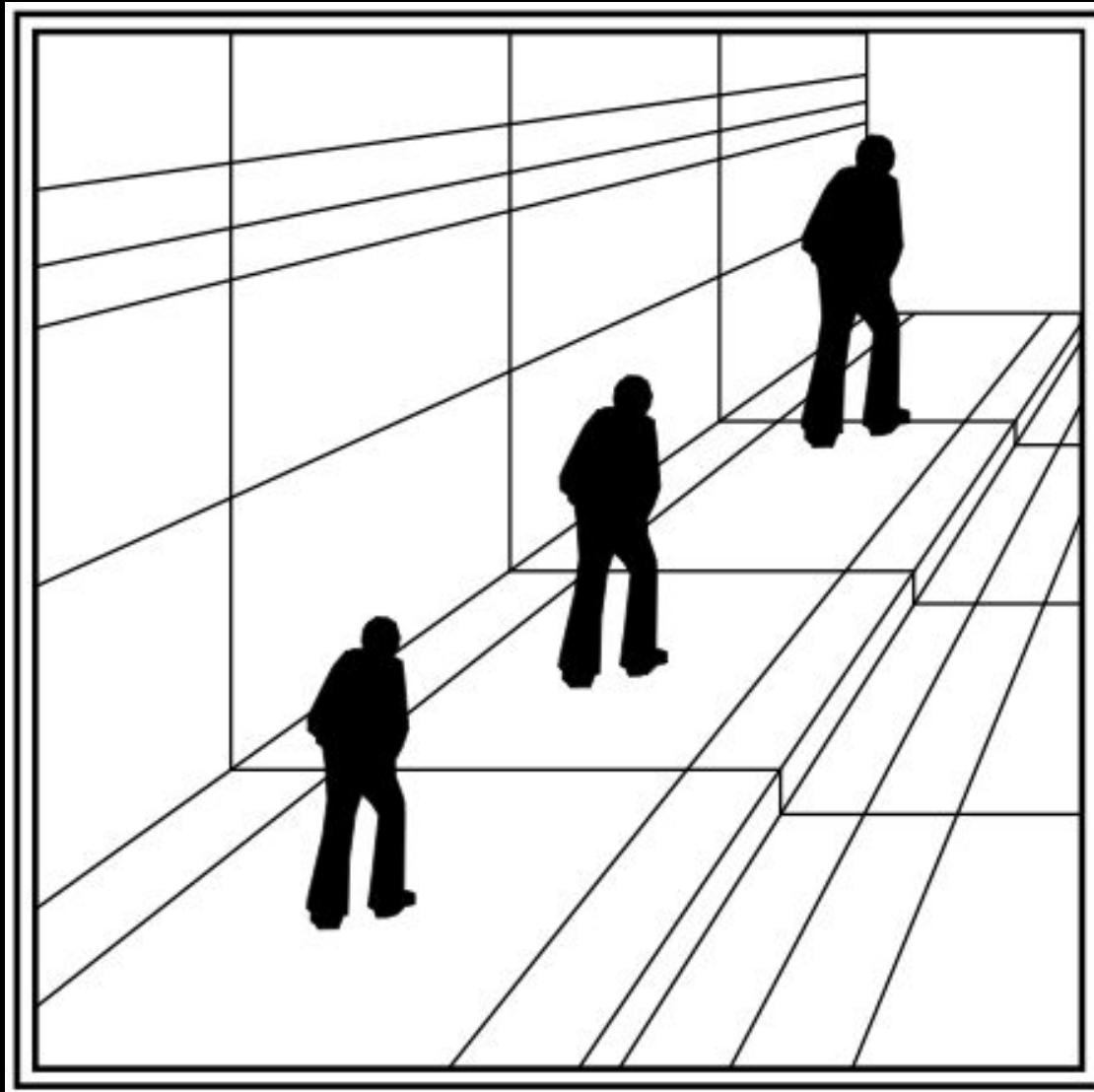
†Weighted average of member countries ‡Average of four cities

Sources: McDonald's; *The Economist*

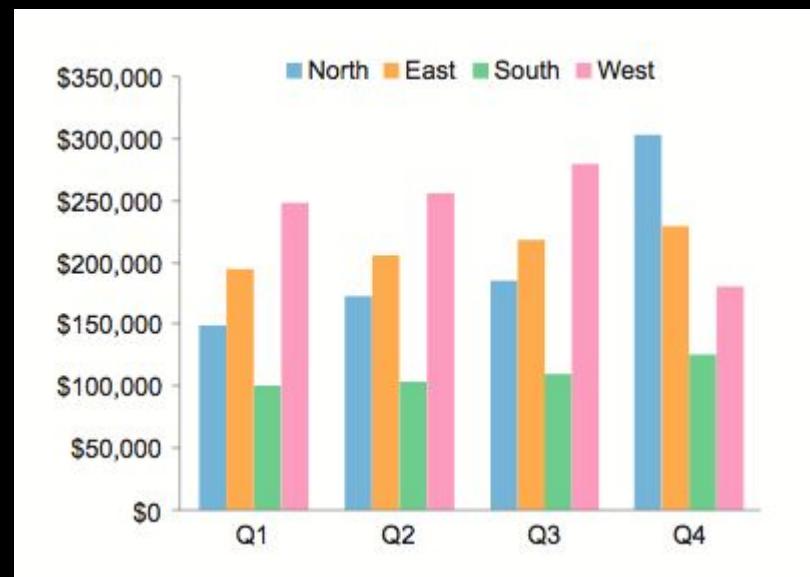
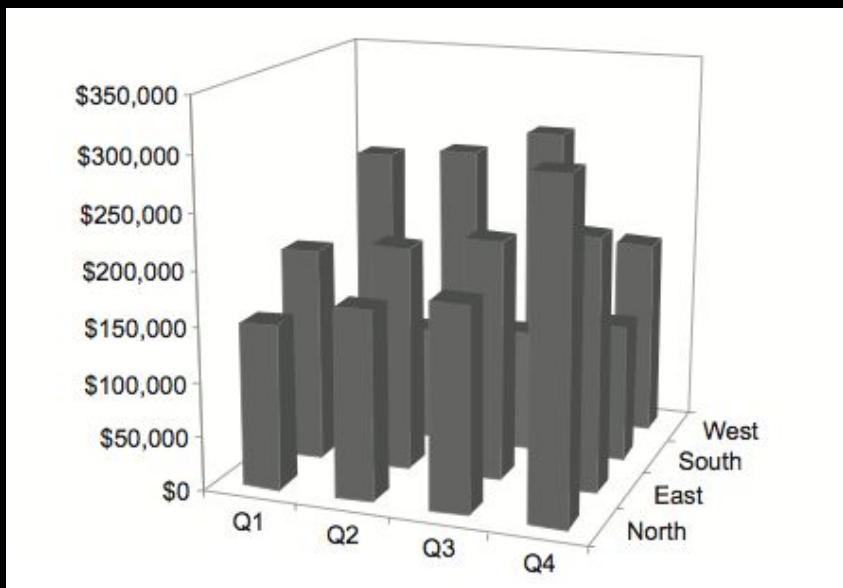
Avoid 3D Distortion



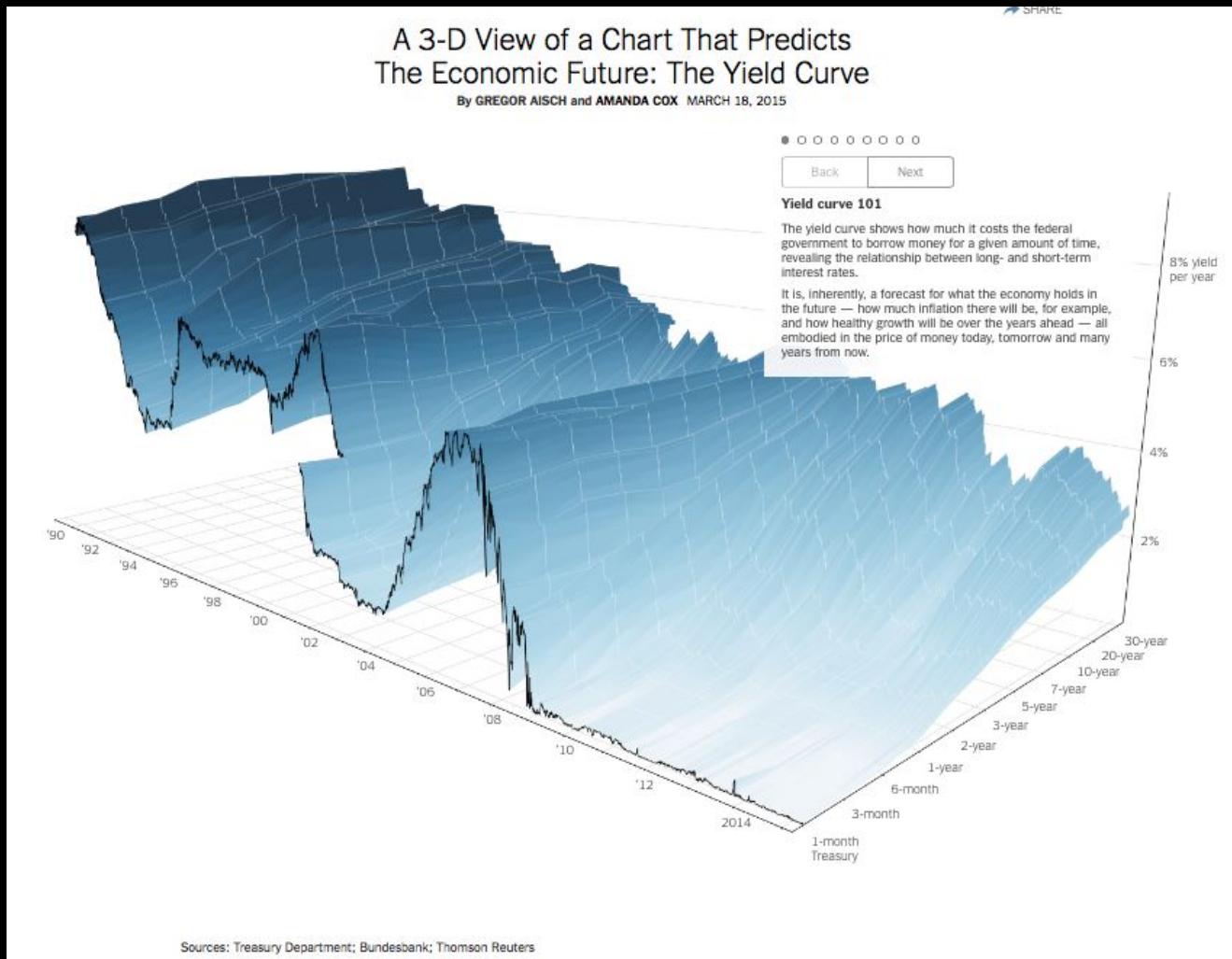
3D vs 2.05D



You do not need 3D



Interactive 3D can work



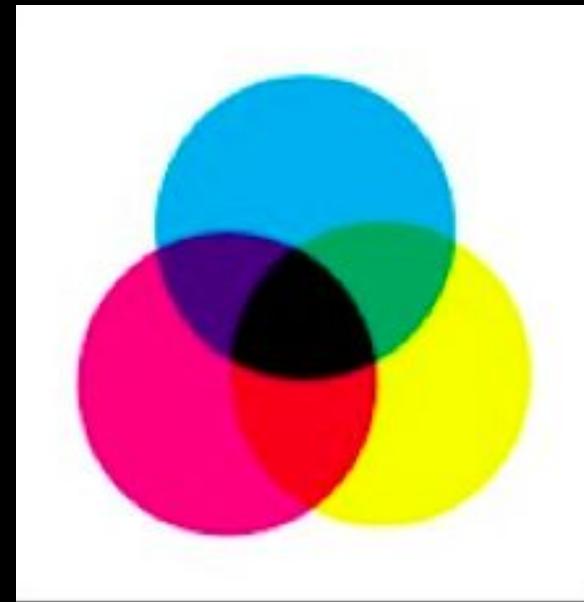
Source: [NYTimes Upshot](#)

Color

Color - RGB / CMYK



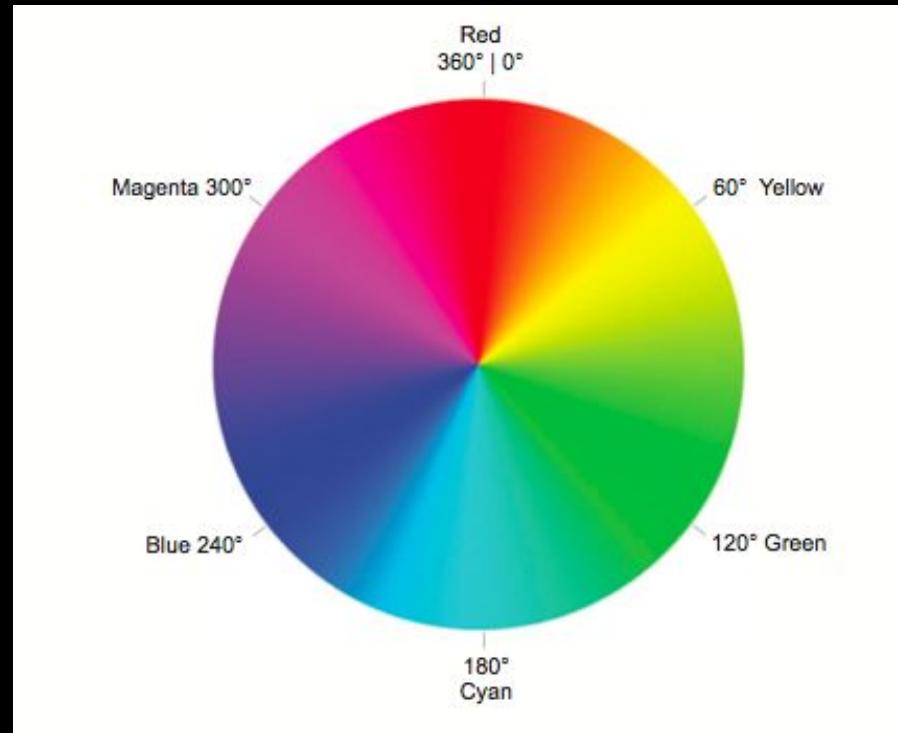
Additive
(Digital Display)



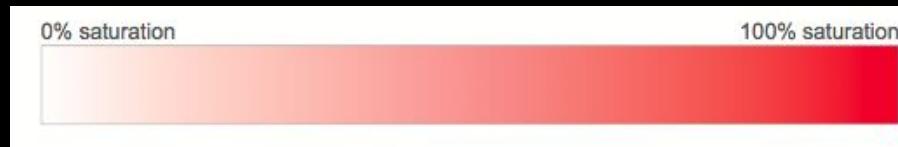
Subtractive
(Print)

Color Perceptual

Hue



Saturation



Lightness



Use of Color

To label

To measure

To represent and imitate

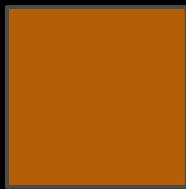
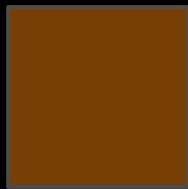
To enliven and decorate

“Above all, do no harm” - Tufte

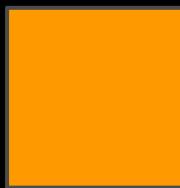
“Position is everything, Color is difficult” - Stefaner

Color

Saturation for Ordinal / Nominal



Hues for Nominal only



Grey

Blue

Orange

Green

Pink

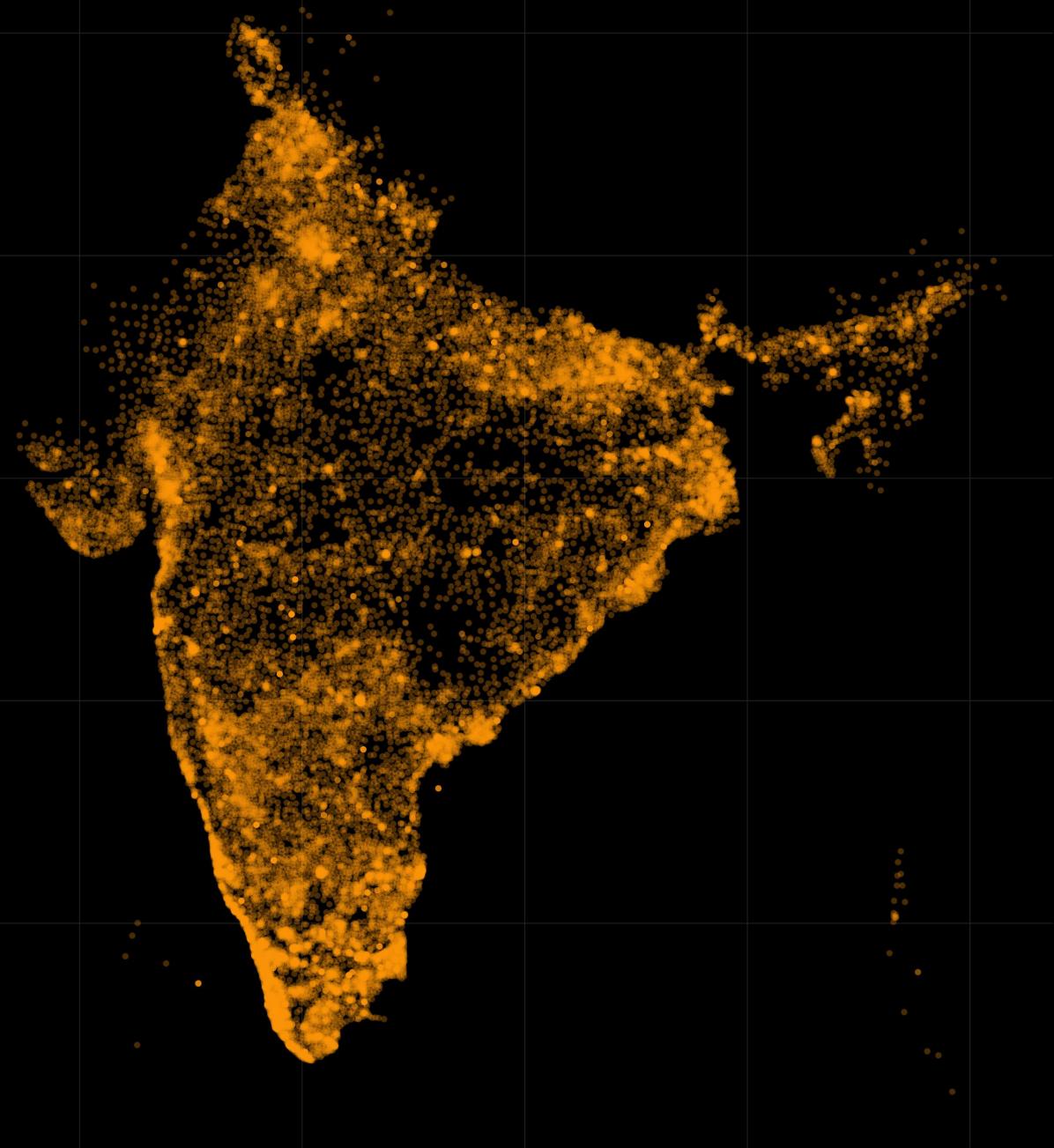
Brown

Purple

Yellow

Red

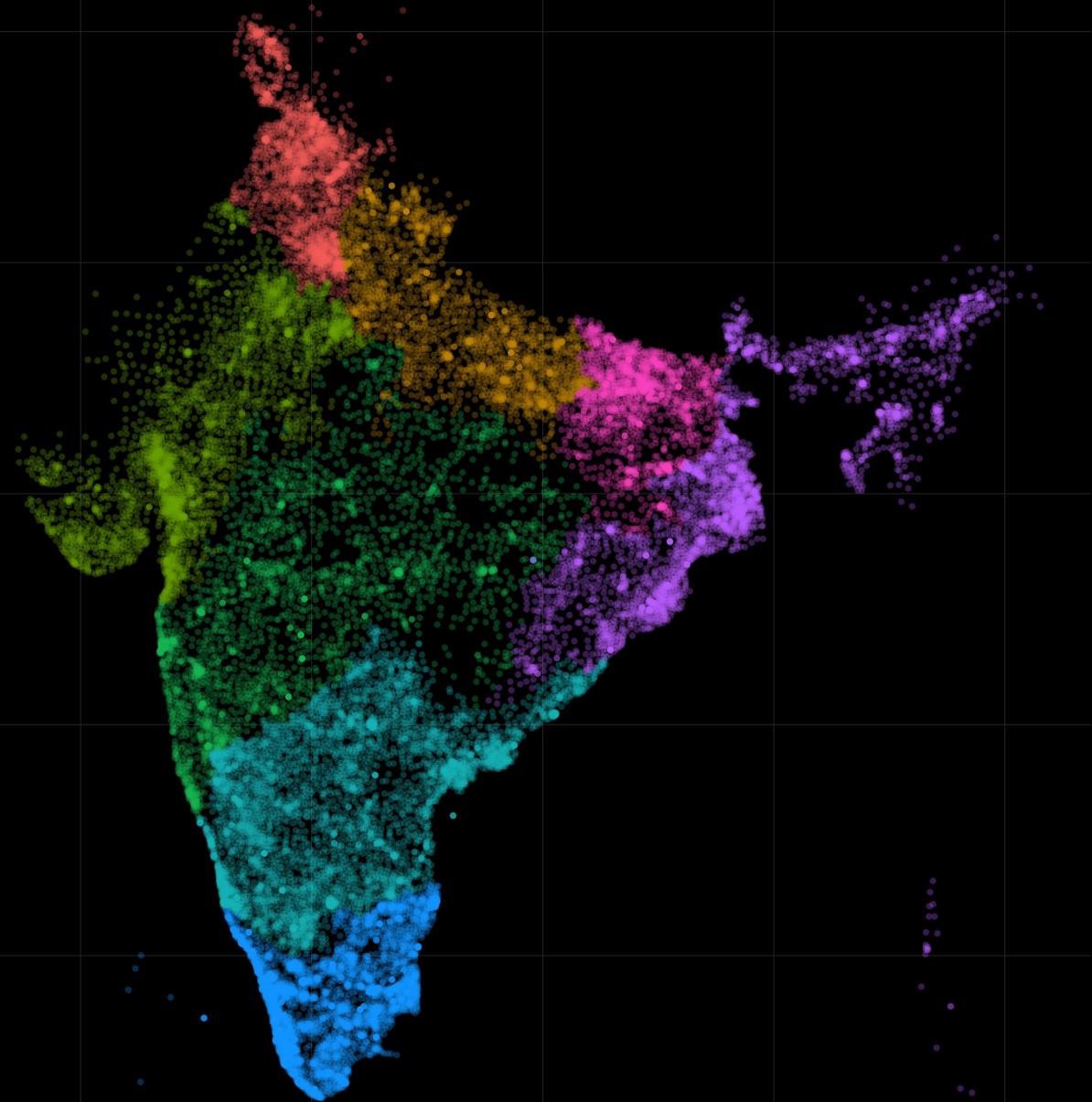
Pincode Map



Scatter plot,
play with alpha
to show density

Single Color

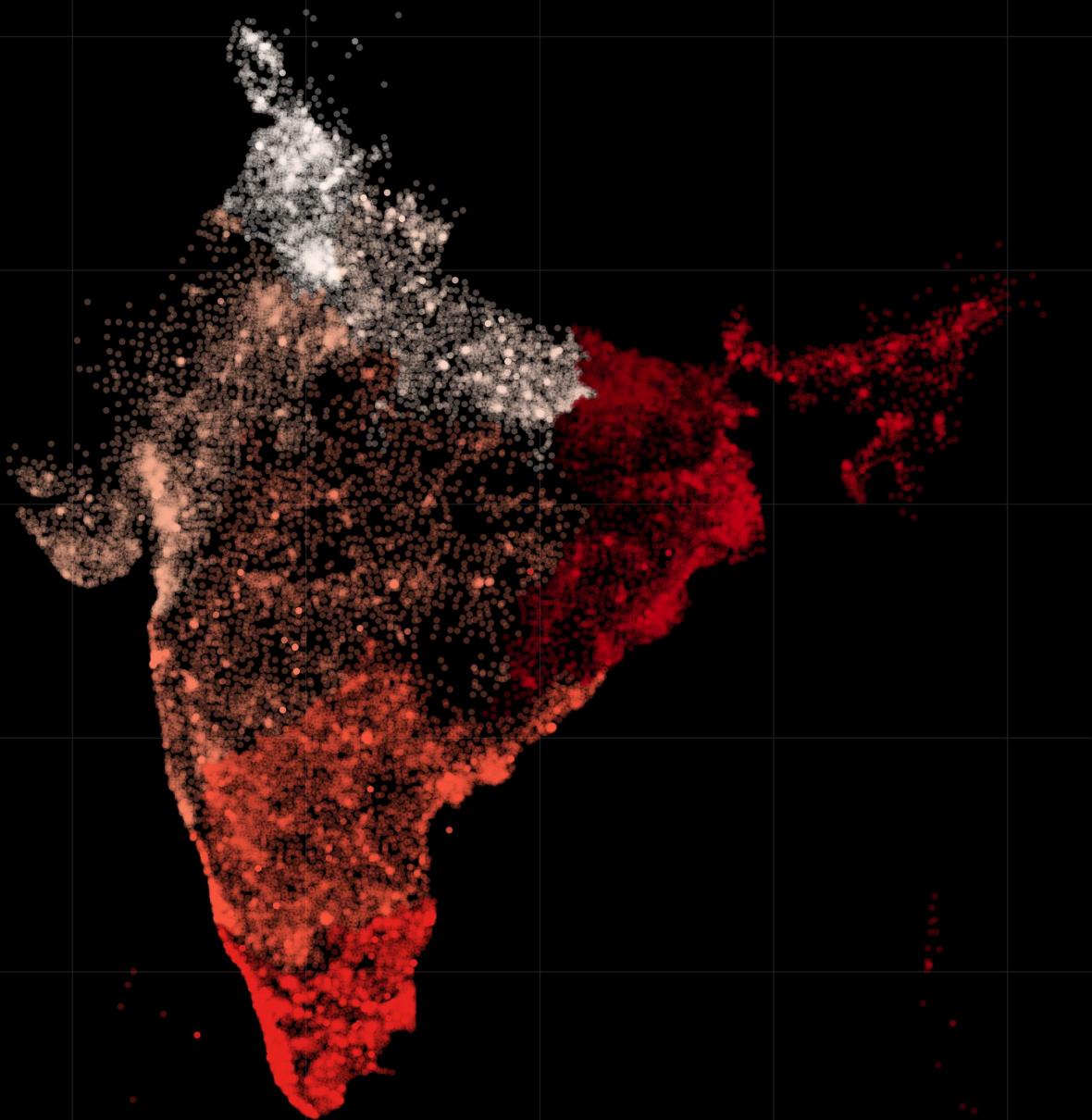
Pincode Map ++



Scatter plot,
play with alpha
to show density

**Hue for each
starting
pincode digit**

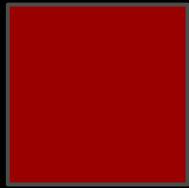
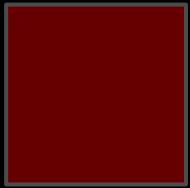
Pincode Map ++



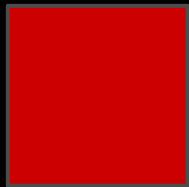
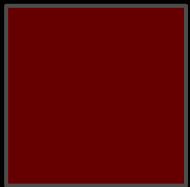
Scatter plot,
play with alpha
to show density

Saturation for
each starting
pincode digit

Perception Limit



Too Close



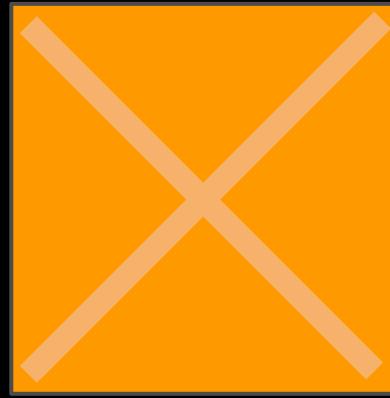
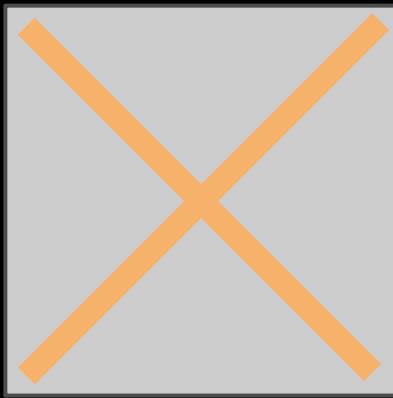
Distinct

Limit the number of colors

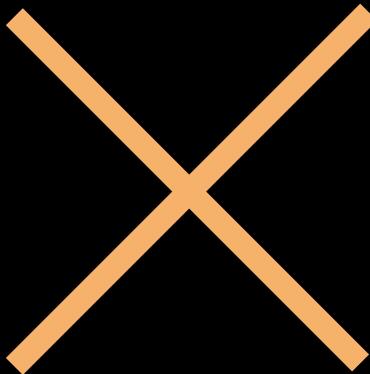
**UNITED
STATES
FREQUENCY
ALLOCATIONS
THE RADIO SPECTRUM**



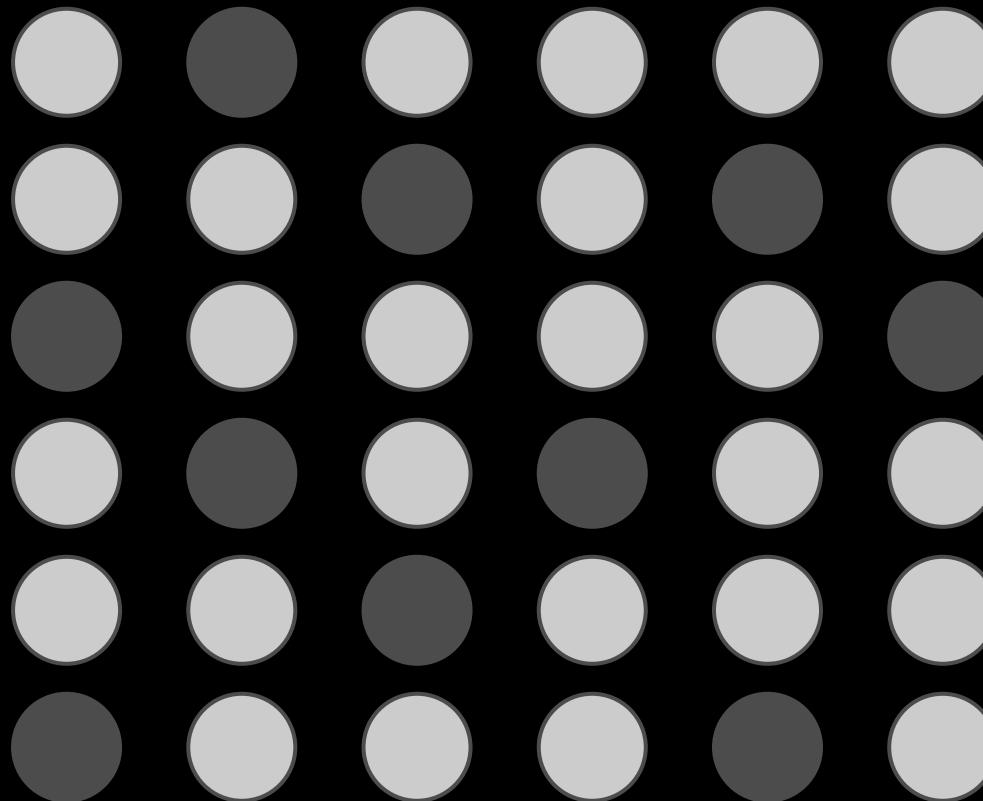
Context Matters



Context Matters

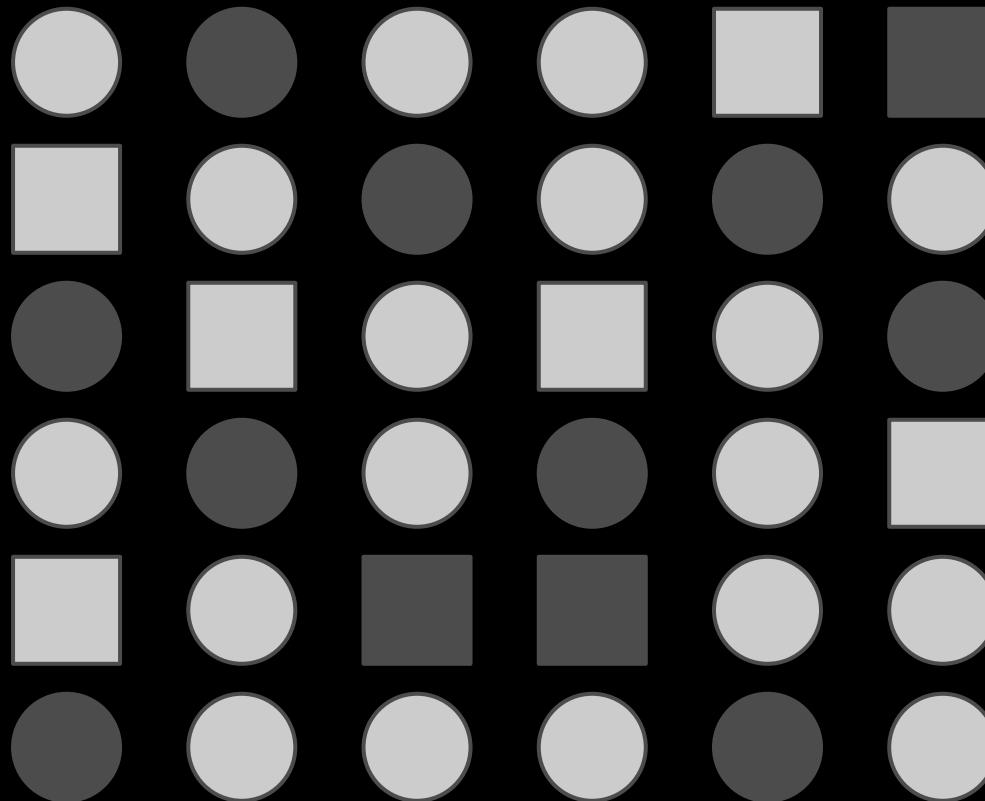


Perception Limit



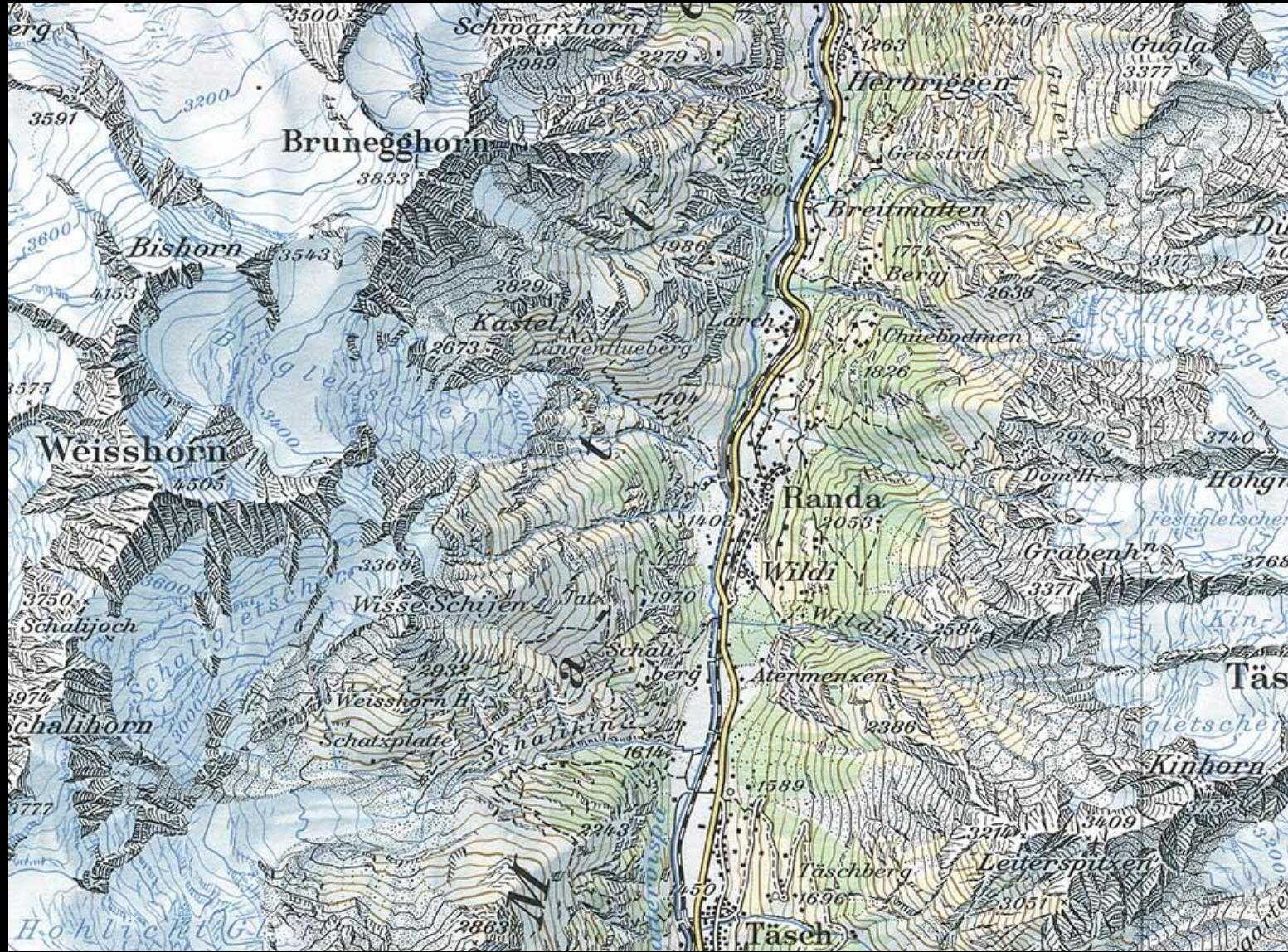
Intensity: One is distinct

Perception Limits

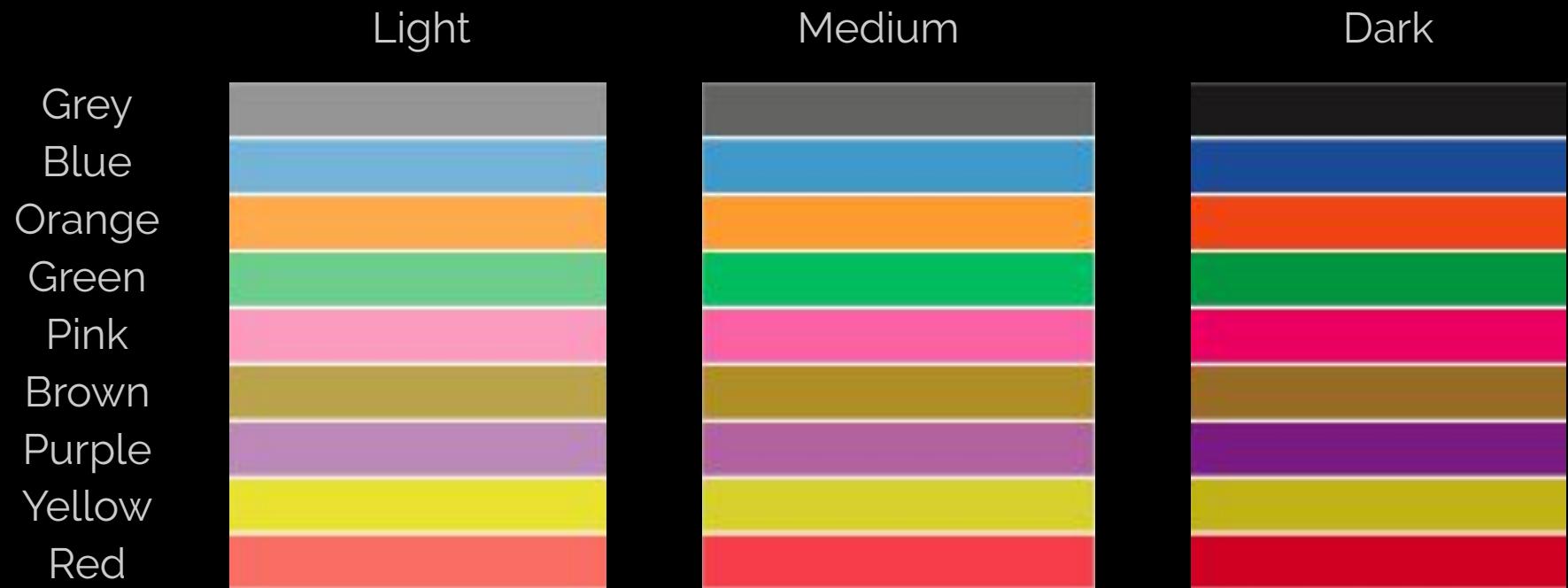


Intensity + Shape: Two is too many

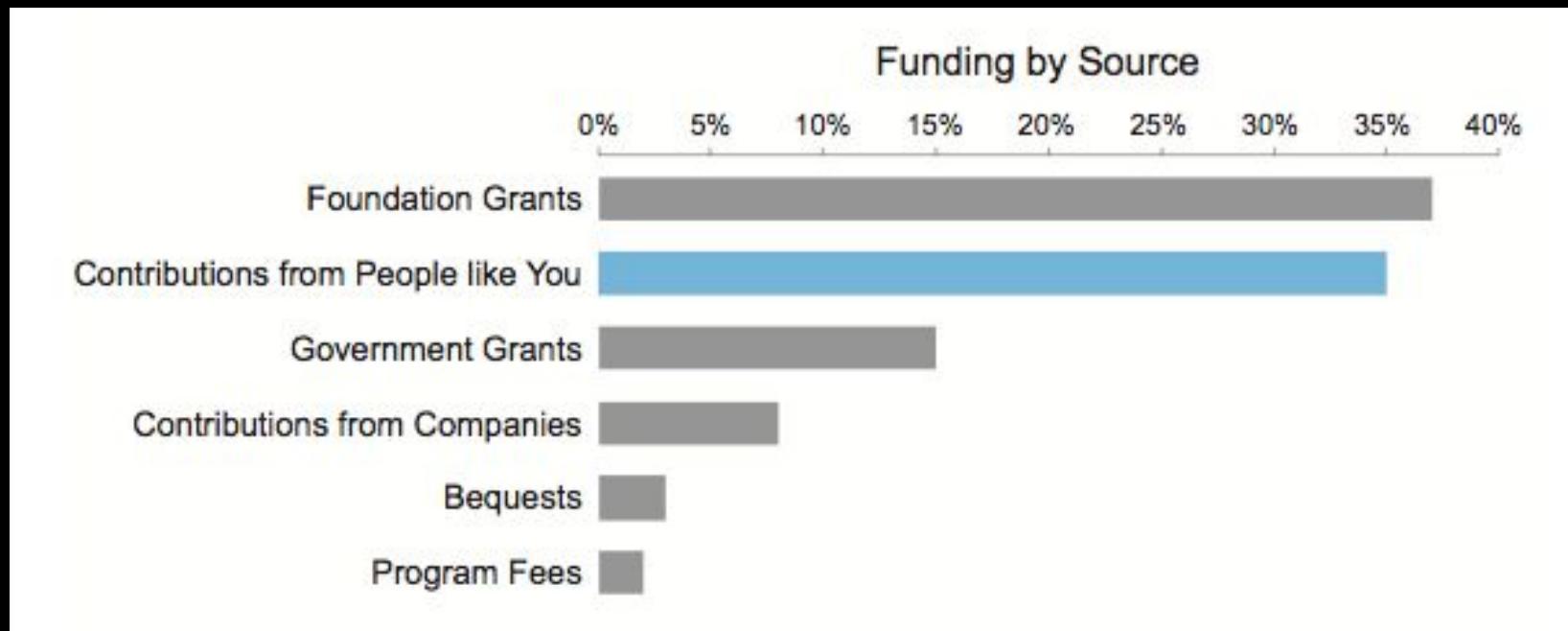
Harmony - Natural Colors



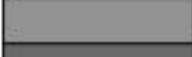
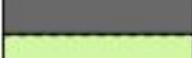
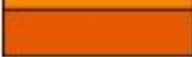
Use Distinct & Named



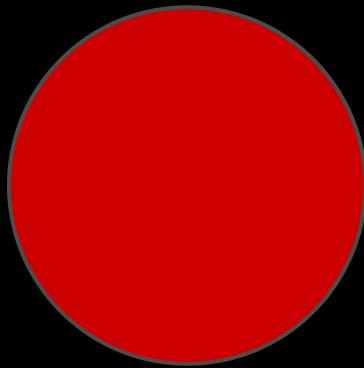
One Contrast Color



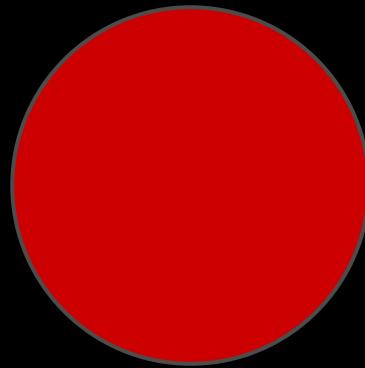
Example Palette

217 / 13 / 57		Signal Color
6 / 107 / 176		Blue 1
76 / 155 / 220		Blue 2
118 / 178 / 228		Blue 3
165 / 204 / 237		Blue 4
220 / 220 / 220		Grey 1
191 / 191 / 191		Grey 2
147 / 147 / 147		Grey 3
105 / 105 / 105		Grey 4
218 / 240 / 168		Green 1
175 / 224 / 110		Green 2
125 / 185 / 53		Green 3
96 / 139 / 45		Green 4
243 / 207 / 116		Orange 1
239 / 182 / 67		Orange 2
241 / 137 / 23		Orange 3
210 / 99 / 8		Orange 4
255 / 255 / 255		White
240 / 240 / 240		Grey 0
60 / 60 / 60		Grey 5
0 / 0 / 0		Black

Cultural Convention

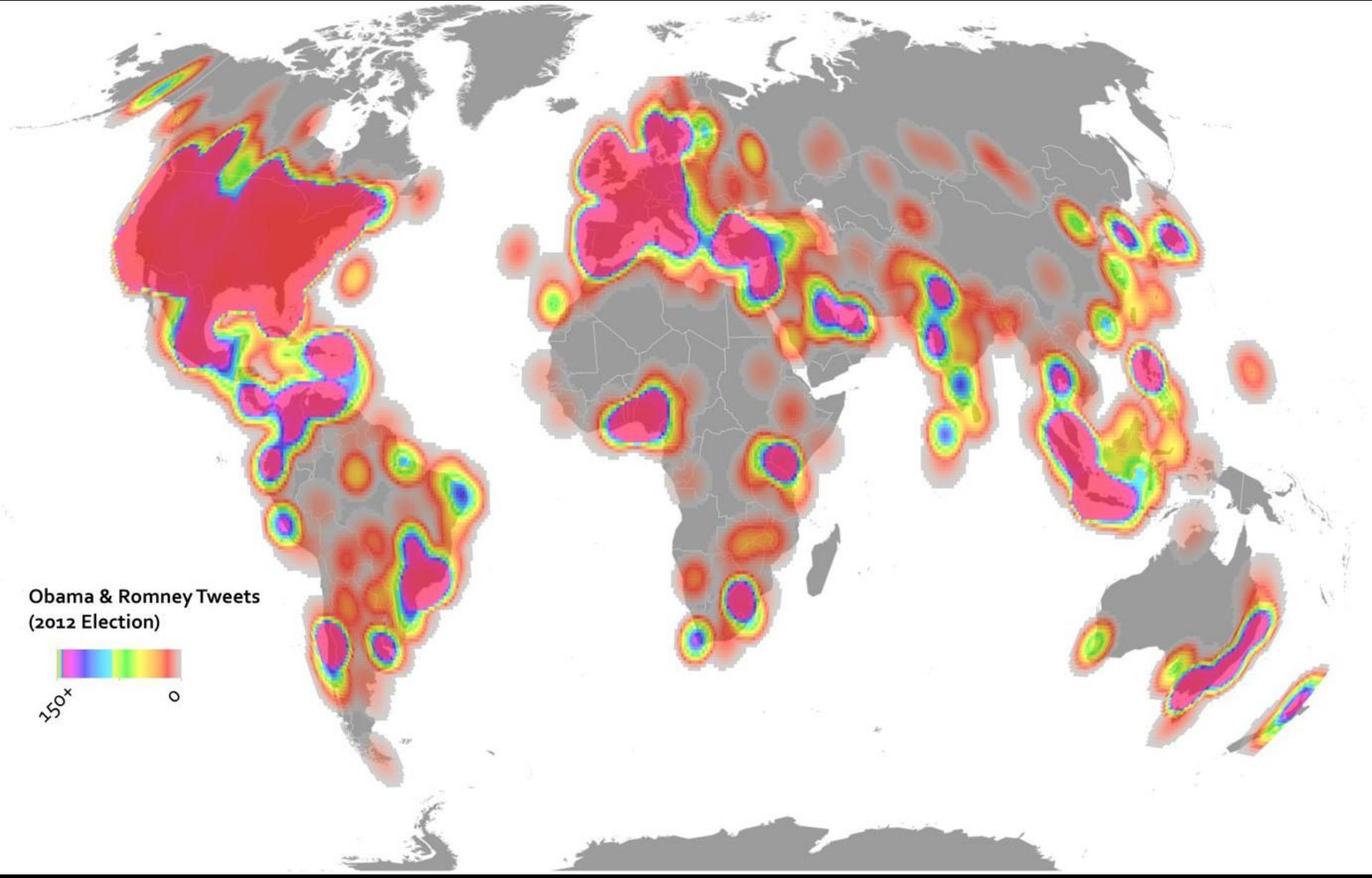


Danger
(West)

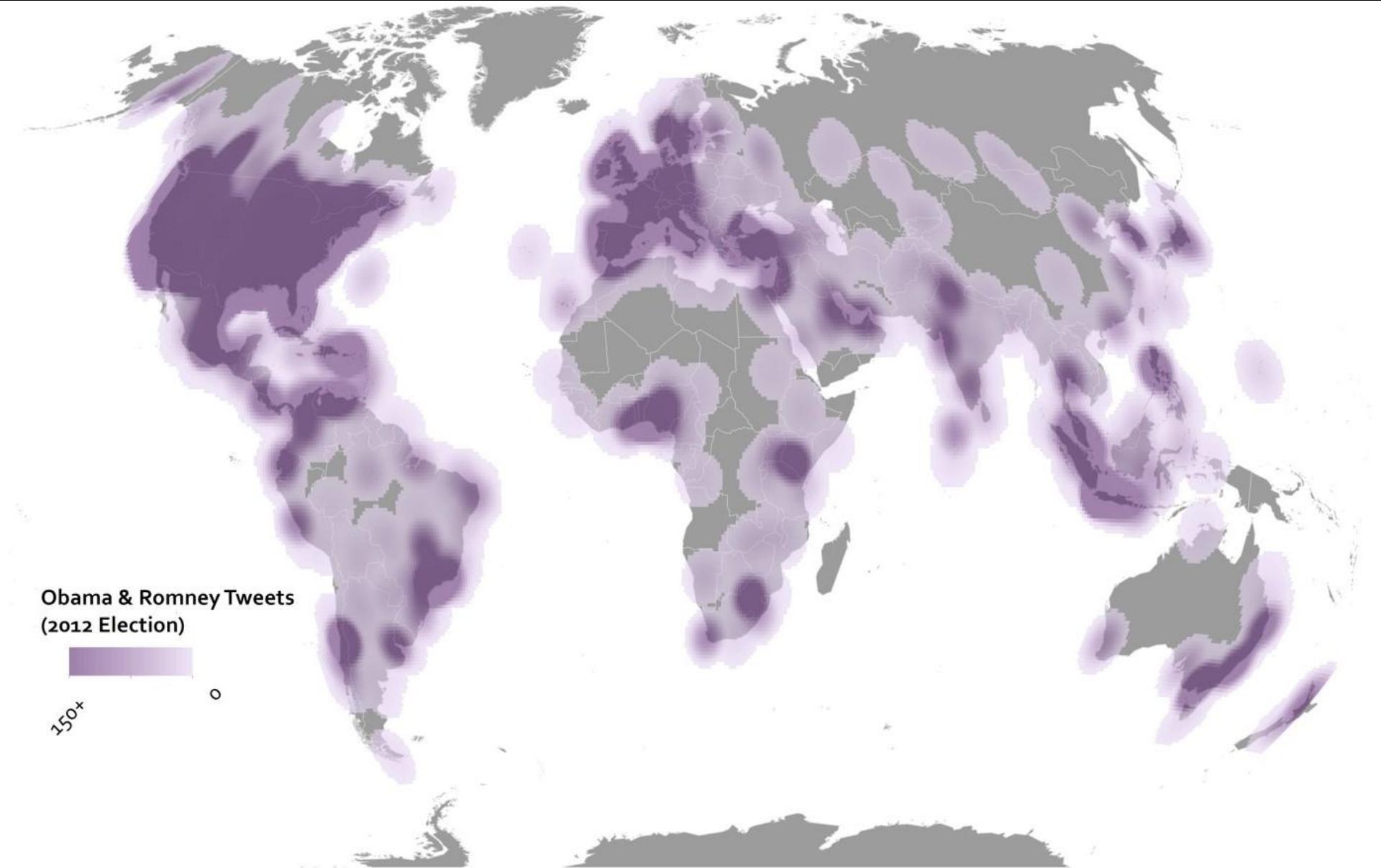


Fortune
(China)

Avoid Rainbow Colors



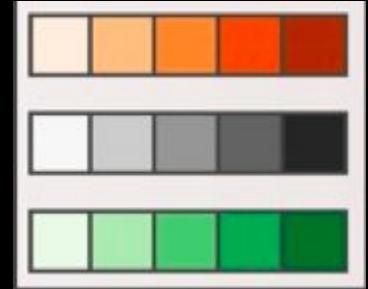
Avoid Rainbow Colors



Quantitative Encoding

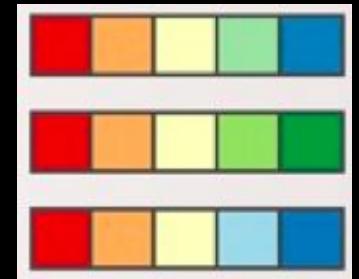
Sequential

- Constrain hue, change sat/lightness
- Map higher value to darker shade



Divergent

- Used when data is a meaningful midpoint
- Use neutral color (grey) at midpoint
- Use saturated color at endpoint

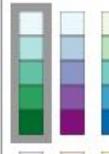
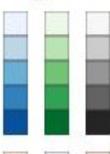


Limit the number of steps from 3 to 9

Tool - Color Brewer

Number of data classes: 3 [i](#) how to use | updates | downloads | credits

Nature of your data: [i](#)
 sequential diverging qualitative

Pick a color scheme:
Multi-hue: 
Single hue: 

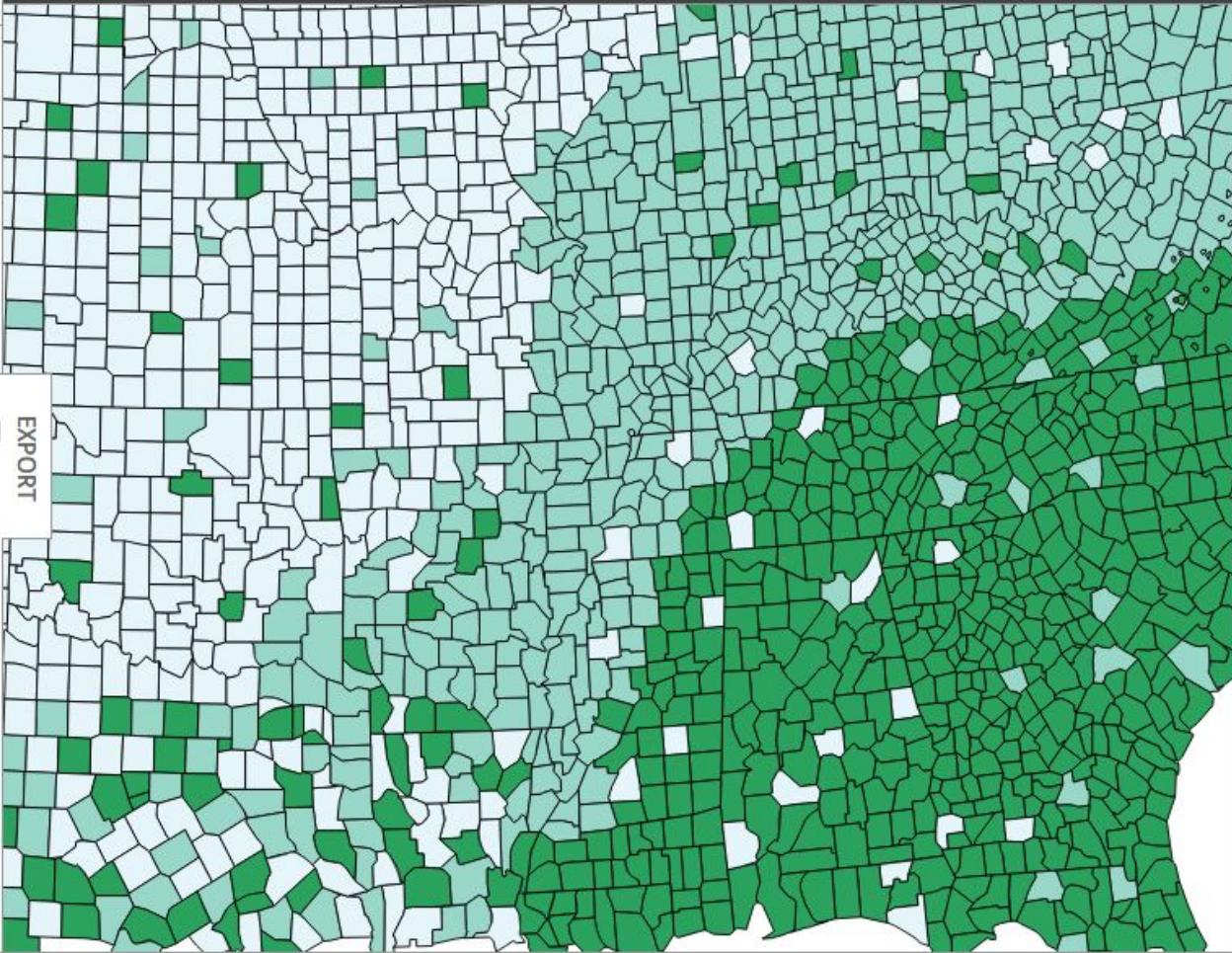
Only show: [i](#)
 colorblind safe
 print friendly
 photocopy safe

Context: [i](#)
 roads
 cities
 borders

Background: [i](#)
 solid color 
 terrain 
color transparency 

COLORBREWER 2.0
color advice for cartography

EXPORT [i](#)
3-class BuGn
  
HEX [i](#)
 #e5f5f9
 #99d8c9
 #2ca25f



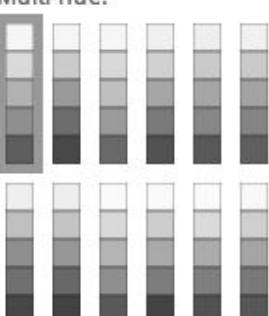
Source: [Color Brewer 2](#)

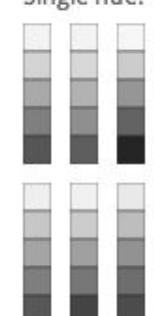
Test in Black & White

Number of data classes: 3

Nature of your data: sequential diverging qualitative

Pick a color scheme:

Multi-hue: 

Single hue: 

Only show:

colorblind safe
 print friendly
 photocopy safe

Context:

roads
 cities
 borders 

Background:

solid color terrain 

color transparency 

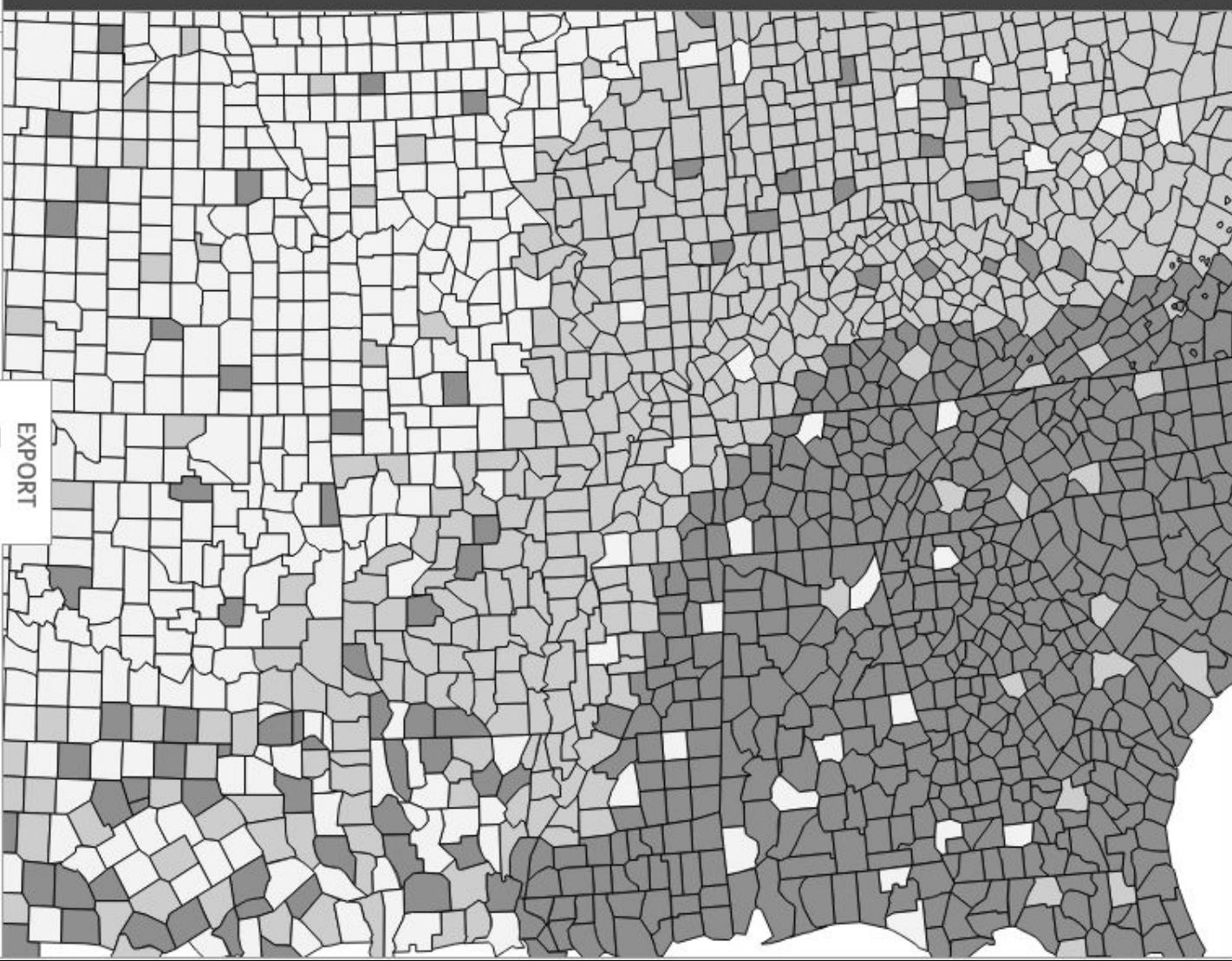
how to use | updates | downloads | credits

COLORBREWER 2.0 color advice for cartography

3-class BuGn     

HEX

#e5f5f9
#99d8c9
#2ca25f



Tool - I want hue

I want hue Tutorials Examples Theory Experiment Old version ▾ GitHub Issues + Médialab Tools

 i want hue

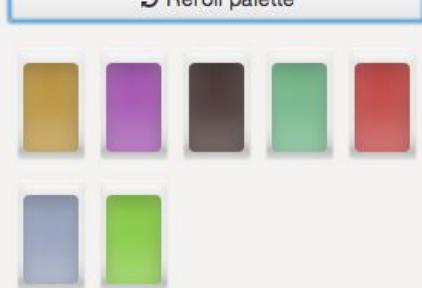
Colors for data scientists. Generate and refine palettes of optimally distinct colors.

Color space

Presets... ▾

H 0 ➔ 360
C 0 ➔ 3
L 0 ➔ 1.5

Dark background



Palette

7 colors soft (k-Means) ▾

Reroll palette

Colors



JSON

HEX json

```
["#RD9849"]
```

CSS

HEX list for CSS

```
#RD9849
```

Source: [I want hue](#)

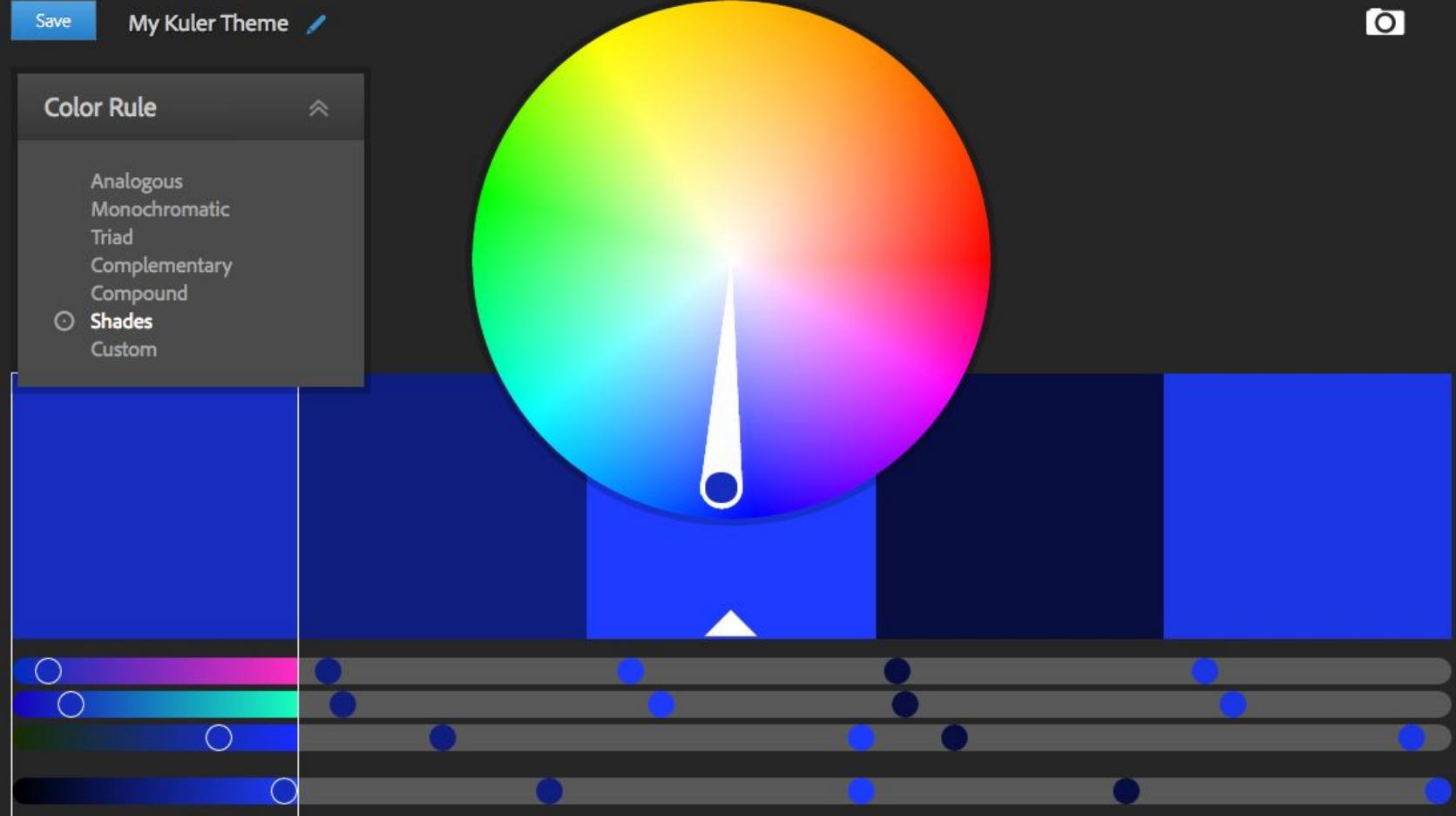
Tool - Adobe Kuler

Adobe Kuler Create Explore My Themes Sign up | Sign in

Save My Kuler Theme 

Color Rule 

- Analogous
- Monochromatic
- Triad
- Complementary
- Compound
- Shades**
- Custom



The interface features a large color wheel at the top center, divided into a gradient from yellow to red, green, blue, and purple. Below the wheel is a horizontal color palette consisting of four vertical bars of varying shades of blue. At the bottom, there are two rows of horizontal sliders, each with a series of colored dots (ranging from light to dark) and small circles for adjustment.

Source: [Kuler Adobe](#)

“In anything at all, perfection is finally attained not when there is no longer anything to add, but when there is no longer anything to take away.”

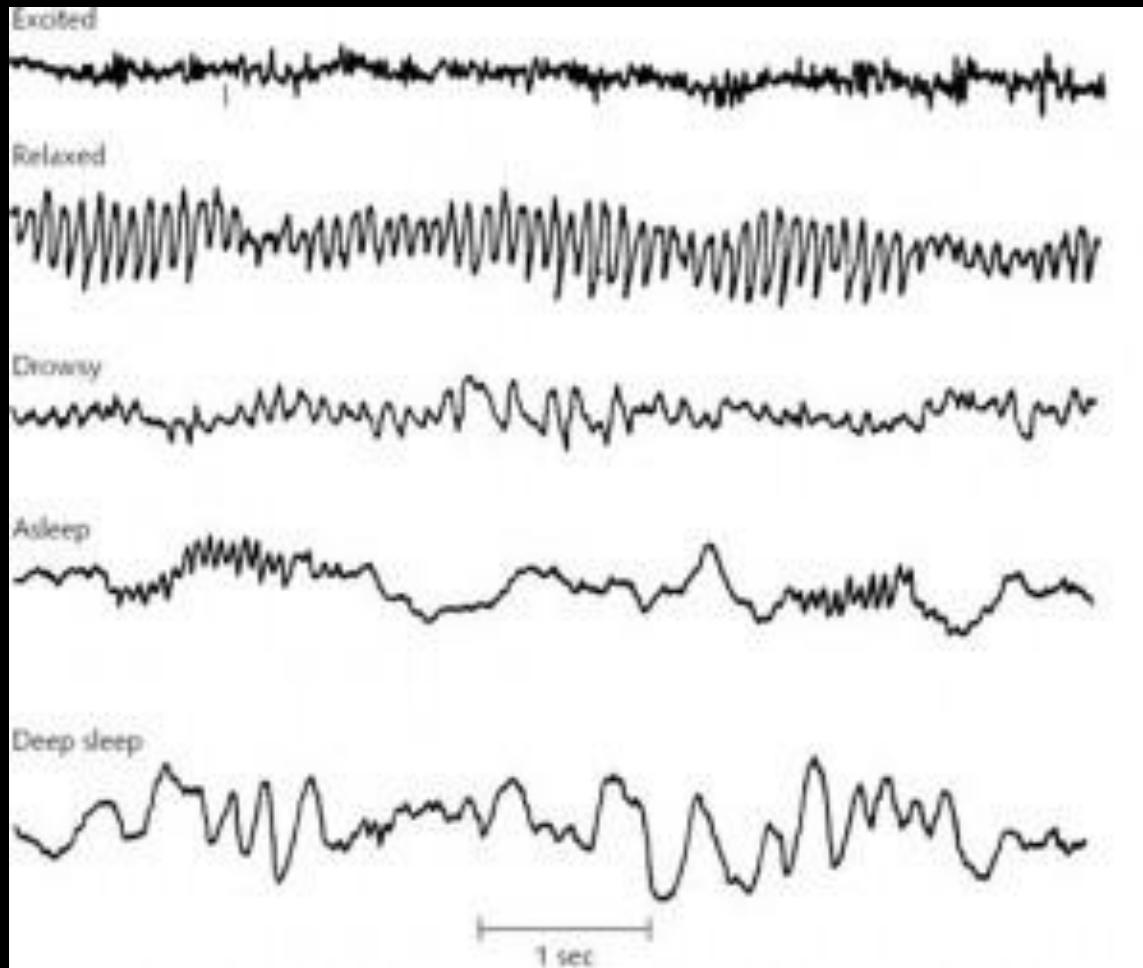
- Antoine de Saint Exupery

Data-Ink Ratio

Data-Ink

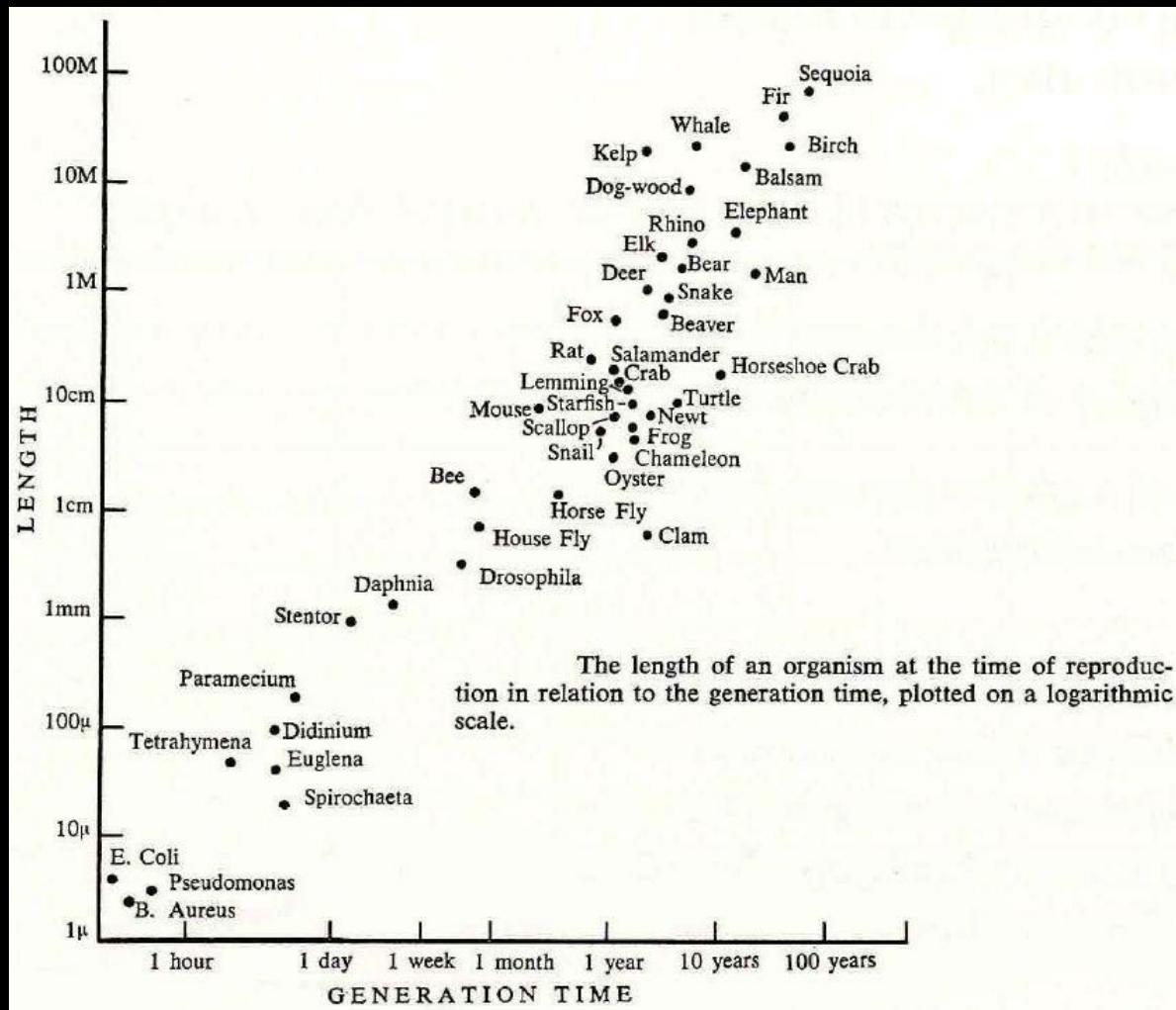
Total Ink used to Print

Data-Ink ~ 1.0



Electroencephalogram

High Data-Ink ~ 0.8



Structure of Biology

Enhance Data-Ink

1. Above all else show the data
2. Maximize the data-ink ratio
3. Erase non-data ink
4. Erase redundant data-ink
5. Revise and Edit

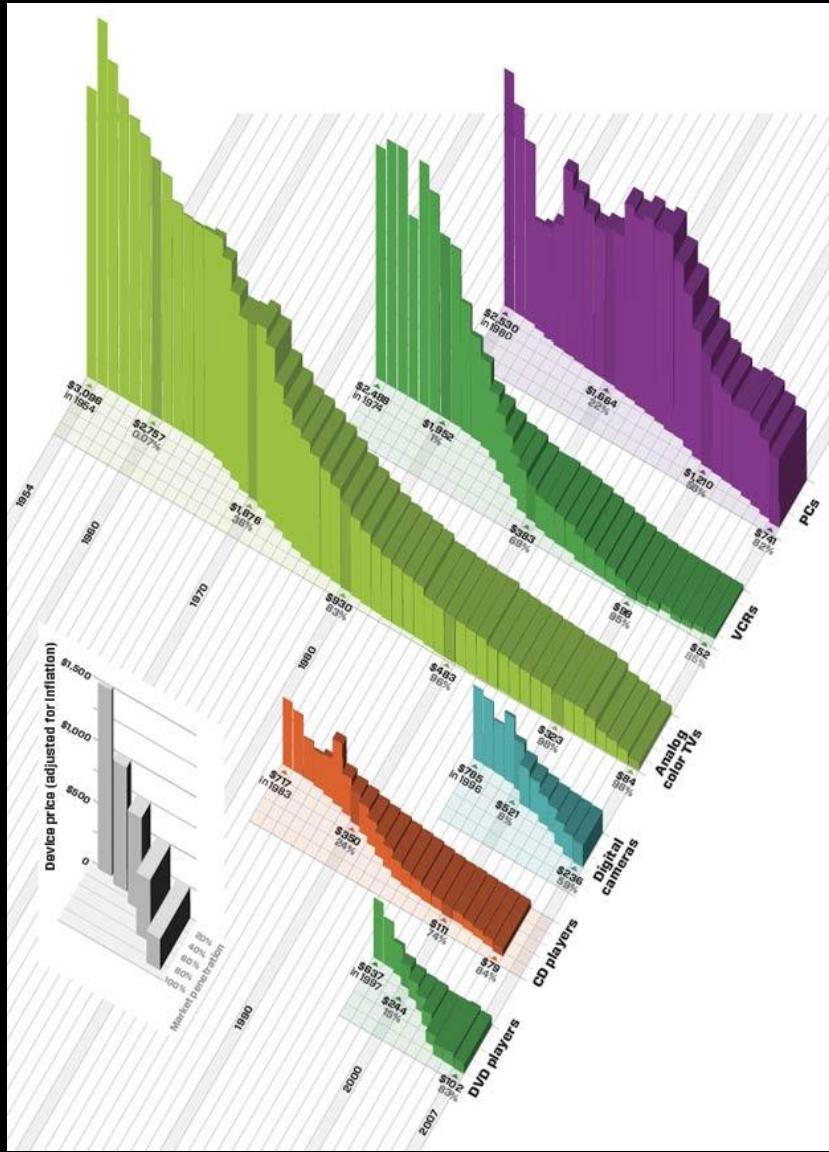
Examples



BBC FOUR



3D Charts - Avoid!



Amit Kapoor
@amitkaps

amitkaps.com

