

Data Analytics and Decision Sciences

Part 2

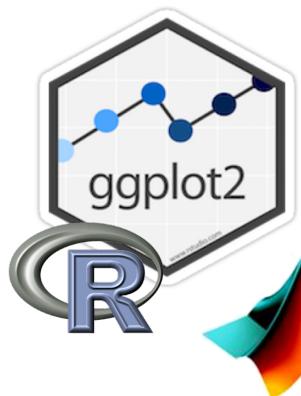
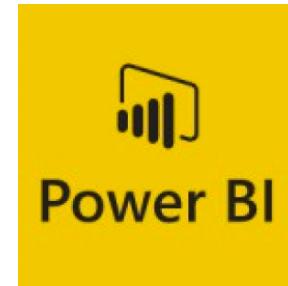
Data and Results Visualization

Daniele Loiacono

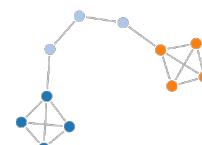


VISUALIZATION TOOLS

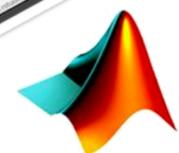
Overview



Seaborn

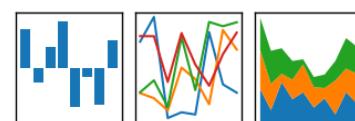


MATLAB®



pandas

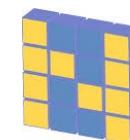
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib



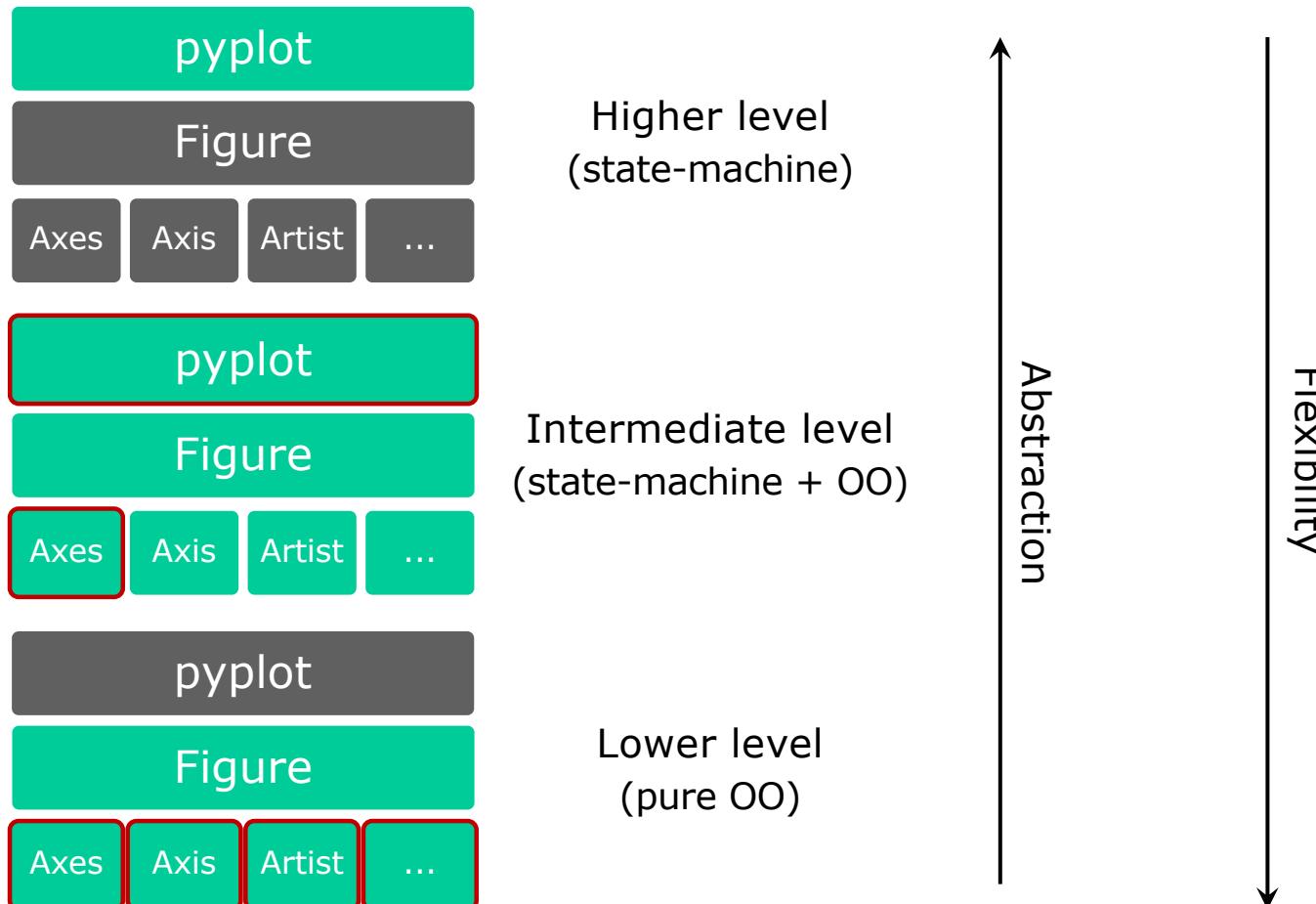
plotly



NumPy

Matplotlib and Seaborn

Using Matplotlib



Seaborn

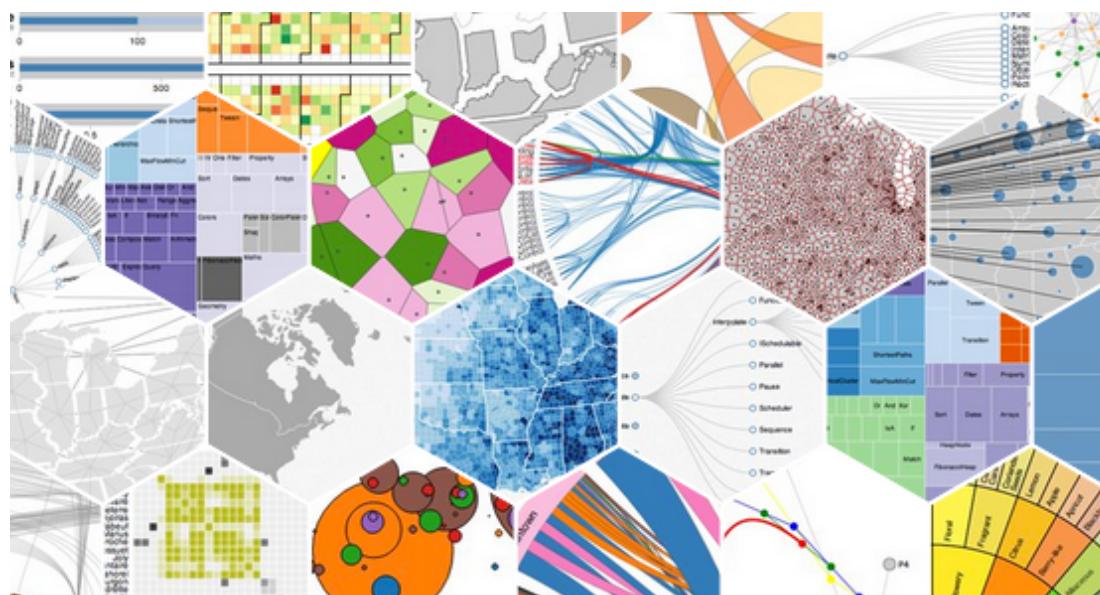
- ❑ Built on top of Matplotlib
- ❑ Integrate with Pandas
- ❑ Allows to use Matplotlib functions to customize the outcome



Alternatives

D3.js

- JavaScript library for manipulating documents based on data
- Uses Web standards HTML, SVG, CSS
- Has powerful visualization components
- Suitable for interactive design



ggplot2

- It is a package for R, a scripting language widely used for data and statistical analysis.
- It produces high-quality (statistical) graphics, using a sort of “grammar” to design graphs component-by-component rather than providing standard graphs.
- It allows to build a graph from concepts rather than recall of commands and options.

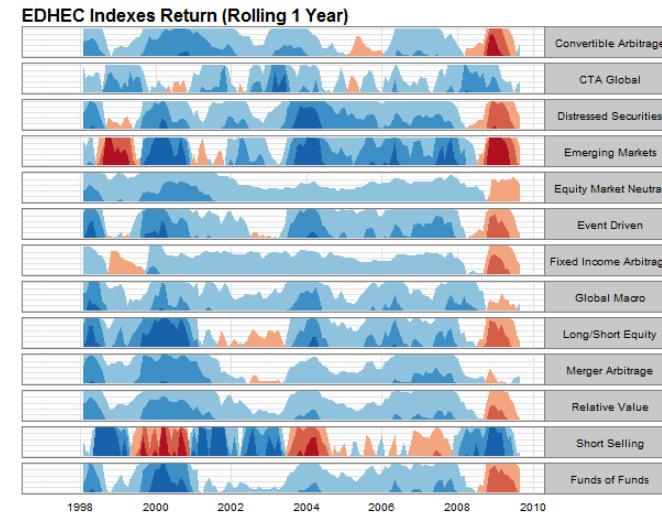
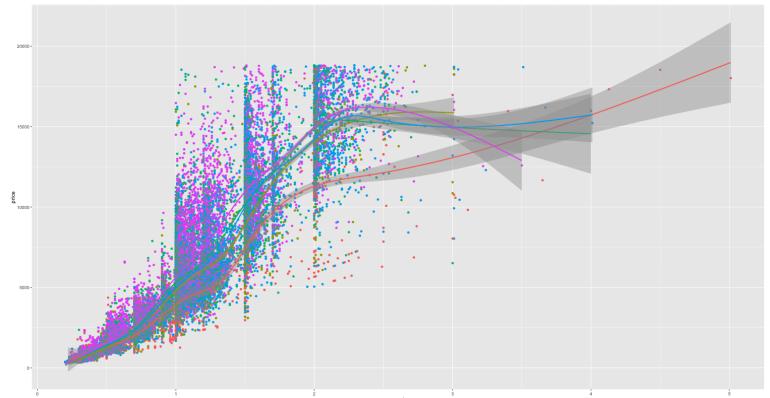
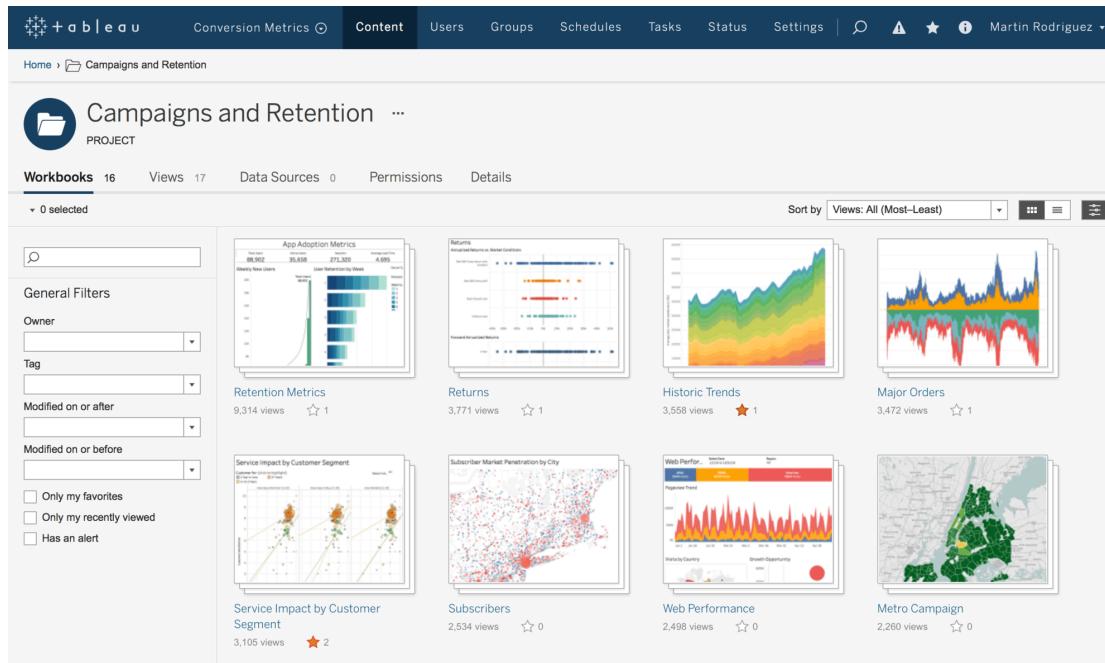


Tableau and Powerbi

- ❑ Commercial integrated solutions.
- ❑ Does not require programming.
- ❑ Allow to create good visualization (also interactive) but not fully customizable.



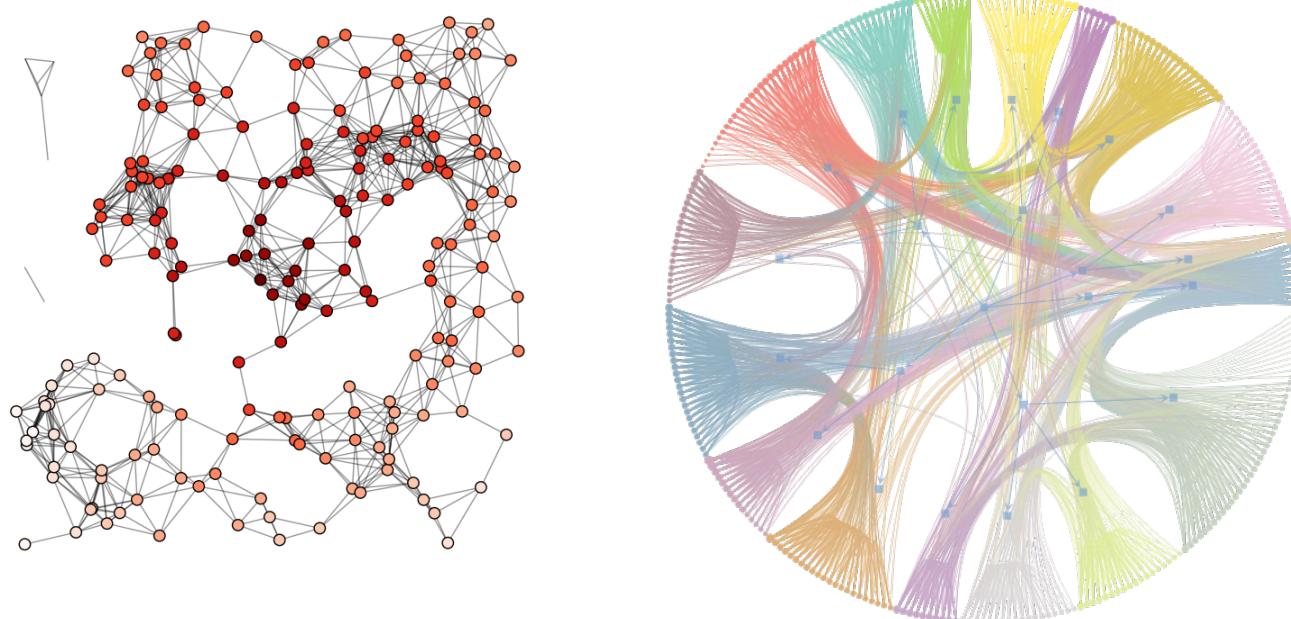
Plotly

- ❑ Premium cloud-based platform for commercial and integrated solutions.
- ❑ Open source libraries in different languages (js, python, R).
- ❑ Solutions for interactive visualization and dashboards.



Networkx and graph-tool

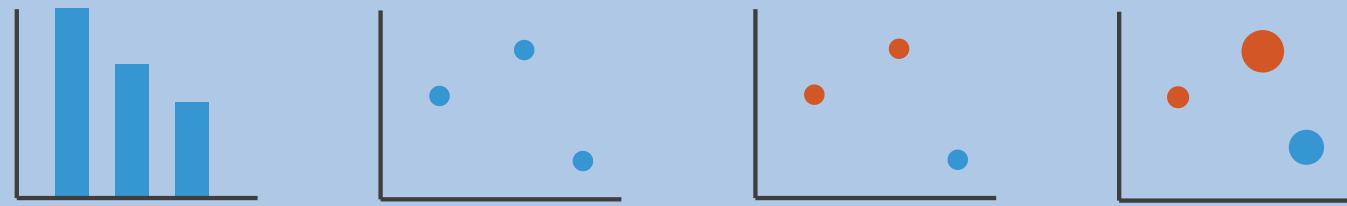
- ❑ Popular graph libraries in Python.
- ❑ Networkx is based on Matplotlib.



Hands-On

Introduction to Pandas
Introduction to Matplotlib
Introduction to Seaborn

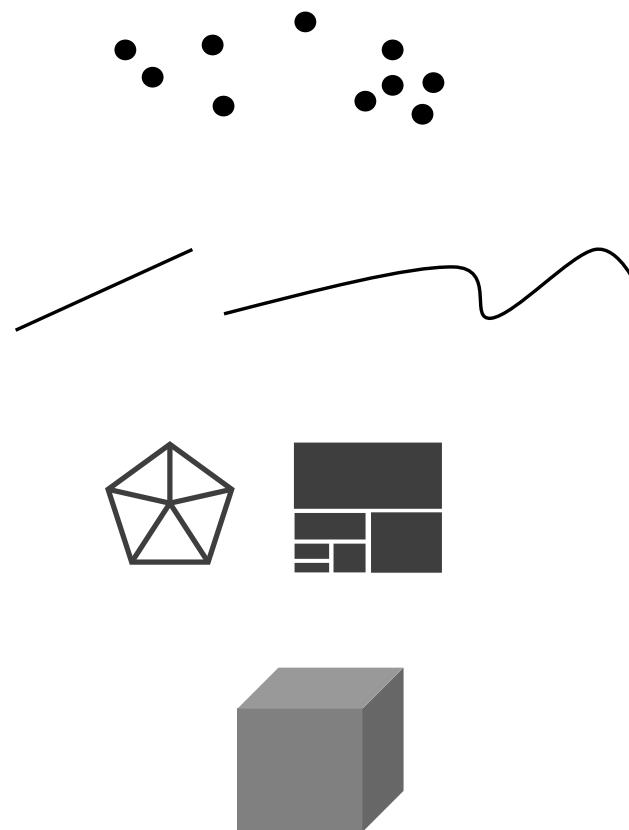
MARKS AND CHANNELS



Elements of visual encodings

Definition: marks

- Marks are the geometric primitives used to visualize data:
 - ▶ Points (0-D)
 - ▶ Lines (1-D)
 - ▶ Areas (2-D)
 - ▶ Volumes (3-D)
- Marks are used to represent items and links



Definition: channels

- Visual channels are visual variables that allow to control the appearance of the marks:

④ Position

→ Horizontal → Vertical



→ Both



④ Color



④ Shape



④ Tilt



④ Size

→ Length



→ Area



→ Volume



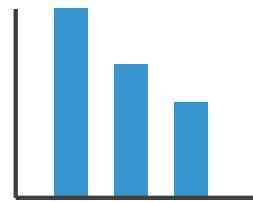
Visual encoding with marks and channels

- ❑ Marks are used to encode **items, nodes, grids, and links**.
- ❑ Channels are used to encode **attributes and positions**.
- ❑ We distinguish two types of visual channels:
 - ▶ **identity channels** encode categorical attributes and positions
 - ▶ **magnitude channels** encode ordered attributes

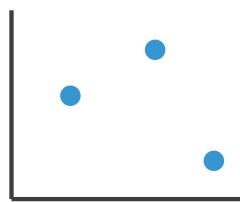
A simple example of analysis

- Identify

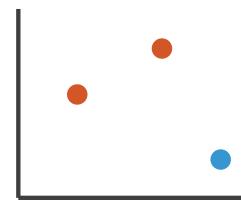
- ▶ Marks used for items
- ▶ Channels of **categorical** and **ordered** attributes



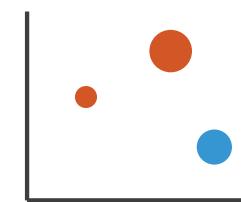
lines
horiz. position
vert. position



points
horiz. position
vert. position



points
horiz. position
vert. position
color

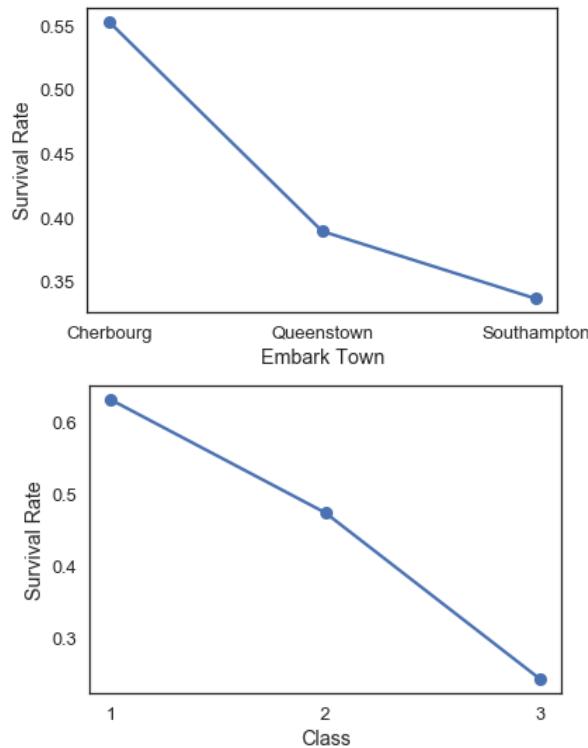


points
horiz. position
vert. position
color
size (area)

Visual encoding principles

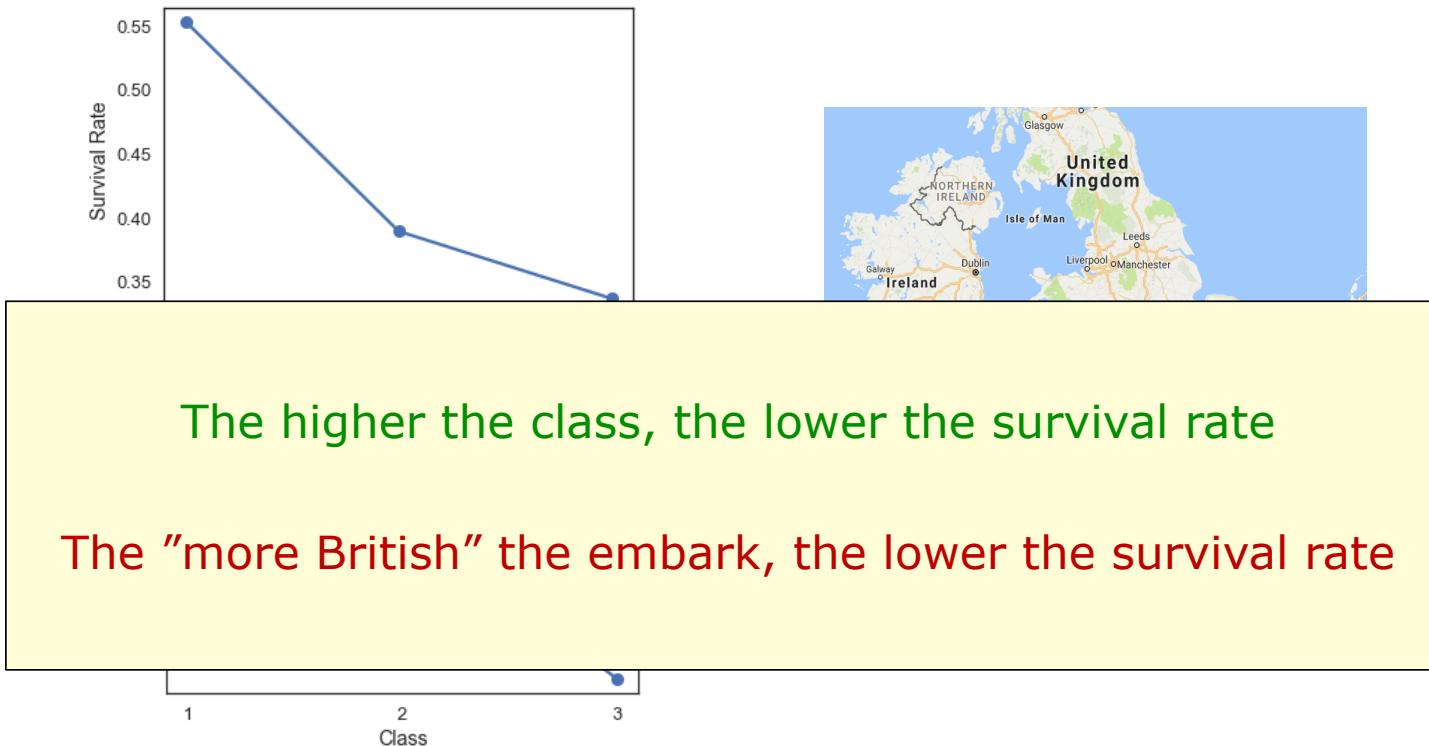
Expressiveness

- Visual encoding should express all and only the information in the data.



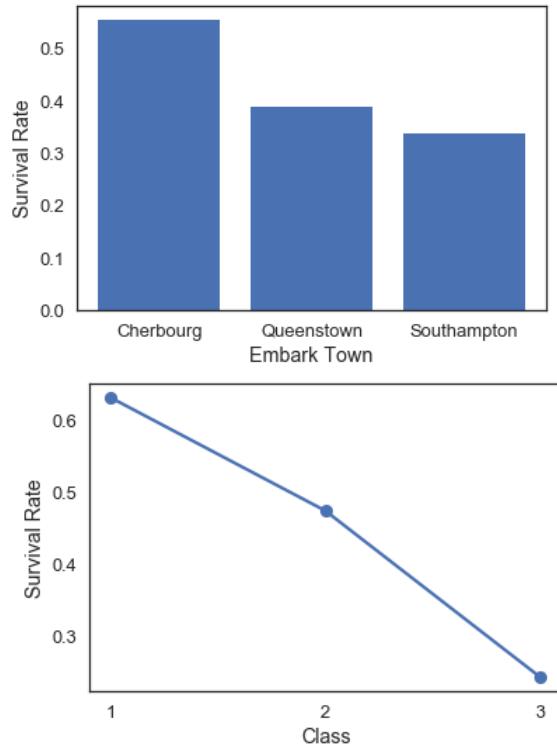
Expressiveness

- Visual encoding should express all and only the information in the data.



Expressiveness

- Visual encoding should express all and only the information in the data.



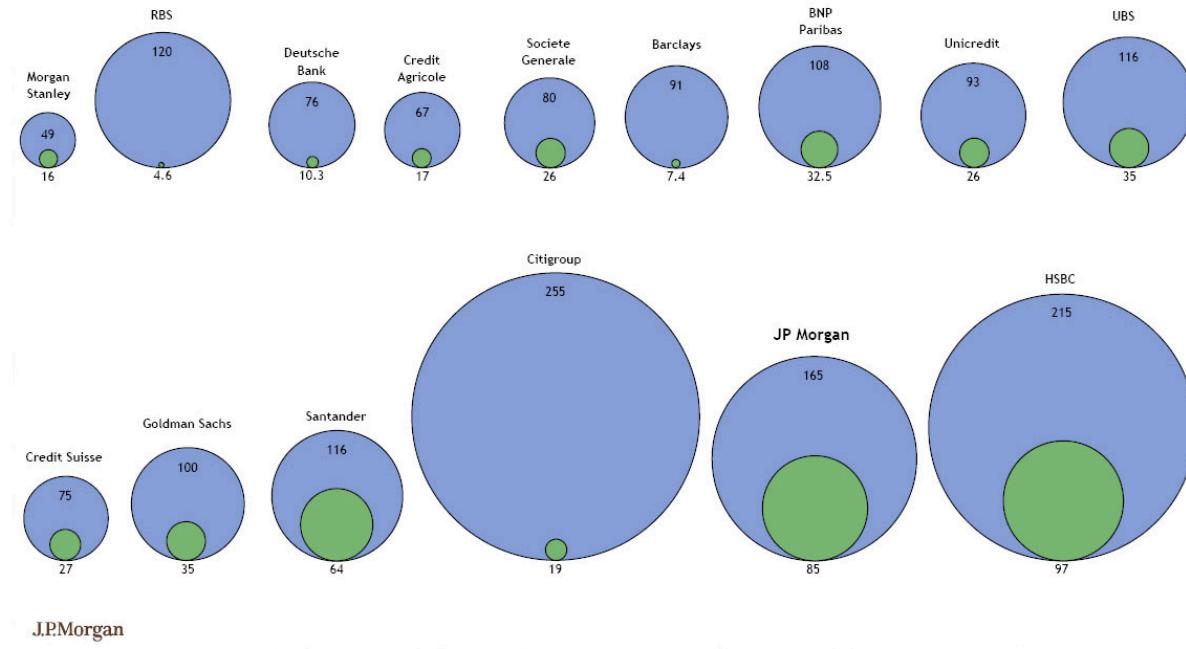
Effectiveness

- The importance of attributes should match the **salience** of the channels used to encode them!

Banks: Market Cap

● Market Value as of January 20th 2009, \$Bn

● Market Value as of Q2 2007, \$Bn



J.P.Morgan

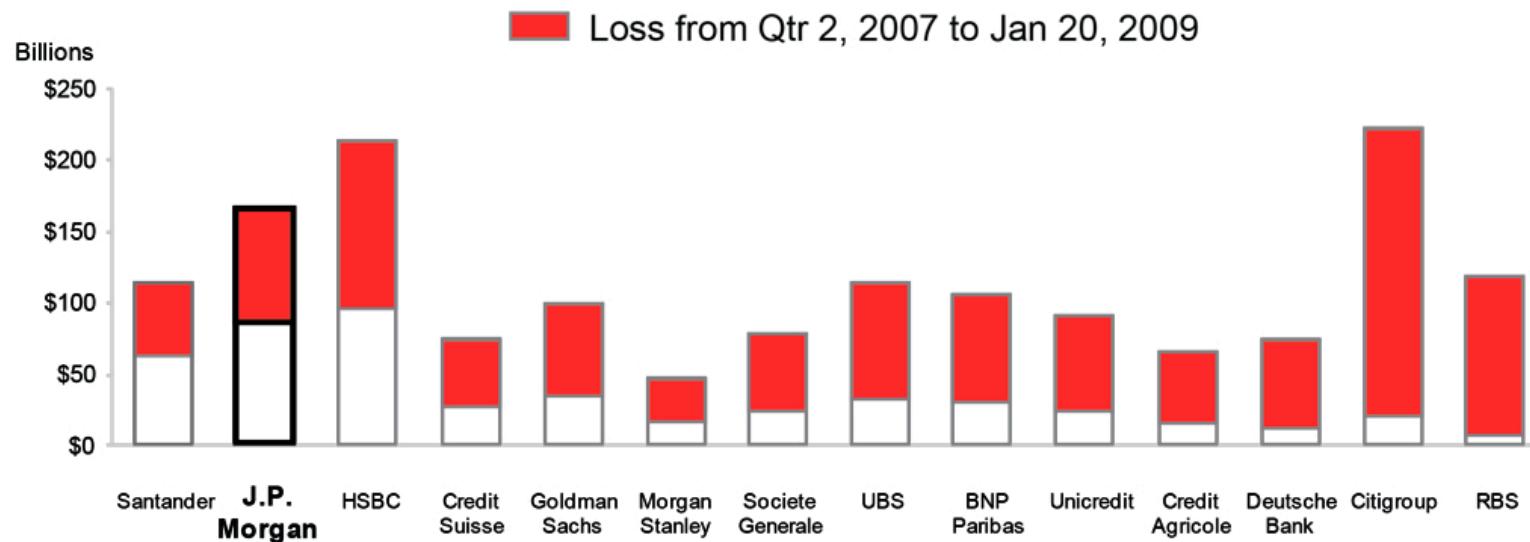
While JPMorgan considers this information to be reliable, we cannot guarantee its accuracy or completeness

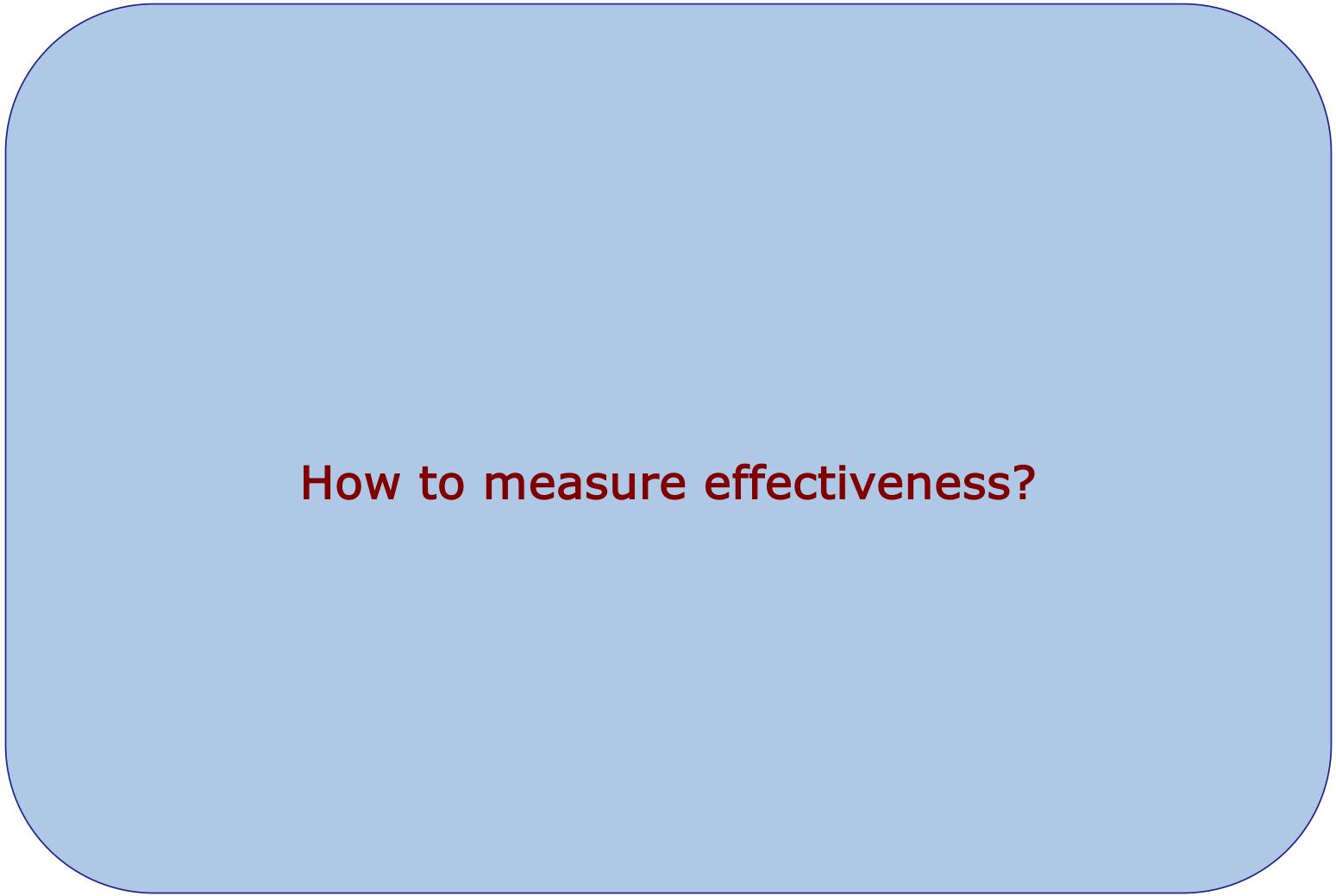
Source: Bloomberg, Jan 20th 2009

Effectiveness

- The importance of attributes should match the **salience** of the channels used to encode them!

Declines in Bank Market Values Since the Financial Crisis Began





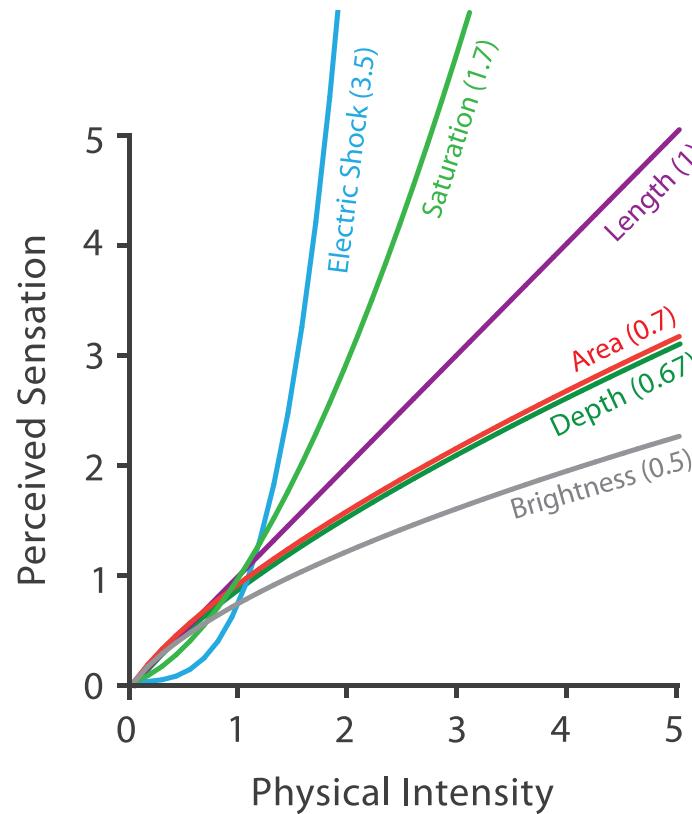
How to measure effectiveness?

Channel Effectiveness

- ❑ **Accuracy:** How accurately values can be estimated.
- ❑ **Discriminability:** How many different values can be perceived.
- ❑ **Separability:** How much interaction there is with multiple encodings.
- ❑ **Popout:** How easy it is to spot some values from the rest.
- ❑ **Grouping:** How good a channel is in conveying groups.

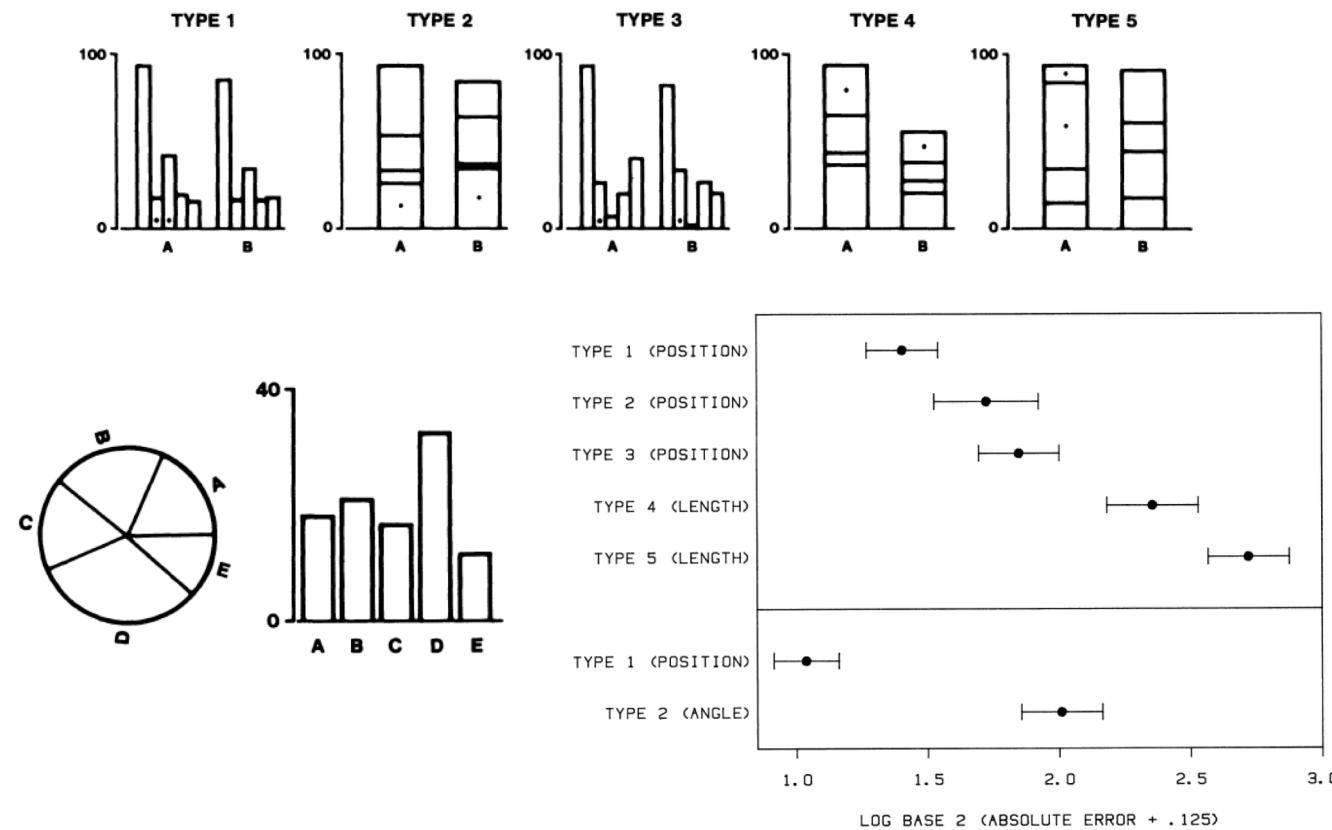
Accuracy

- ❑ How do we perceive magnitude through different channels?



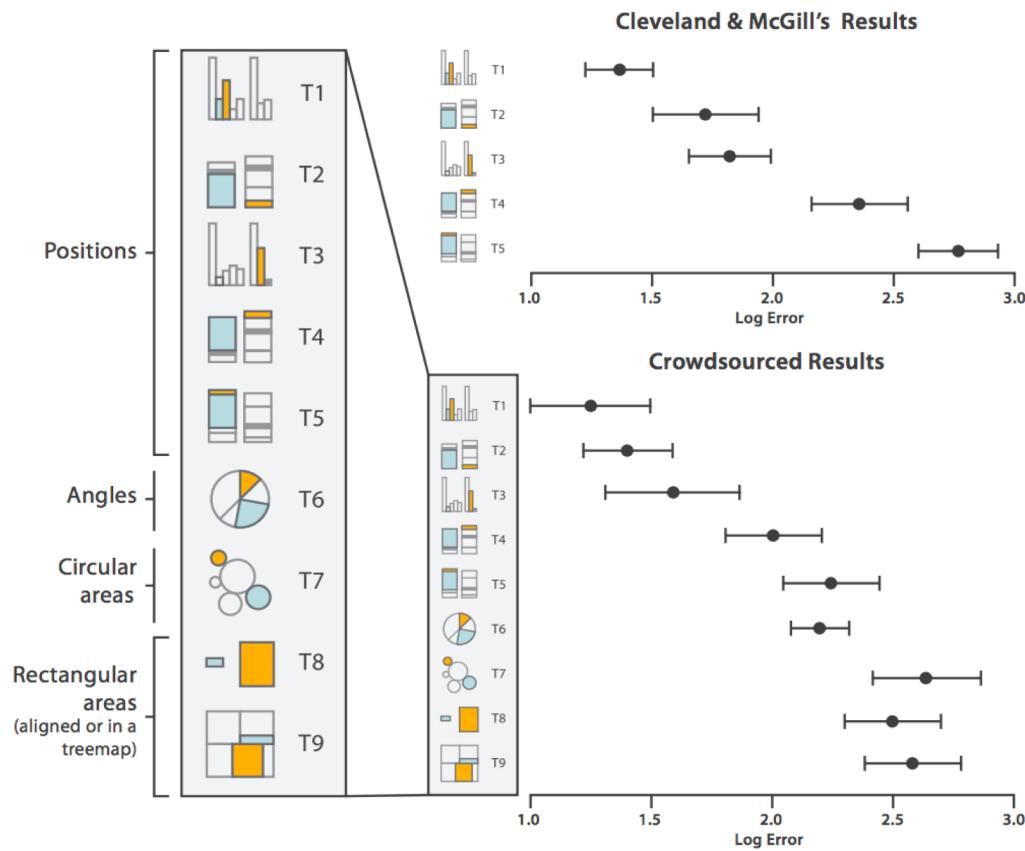
Accuracy: experimental results

Cleveland, William S., and Robert McGill. "[Graphical perception: Theory, experimentation, and application to the development of graphical methods.](#)" (1984)

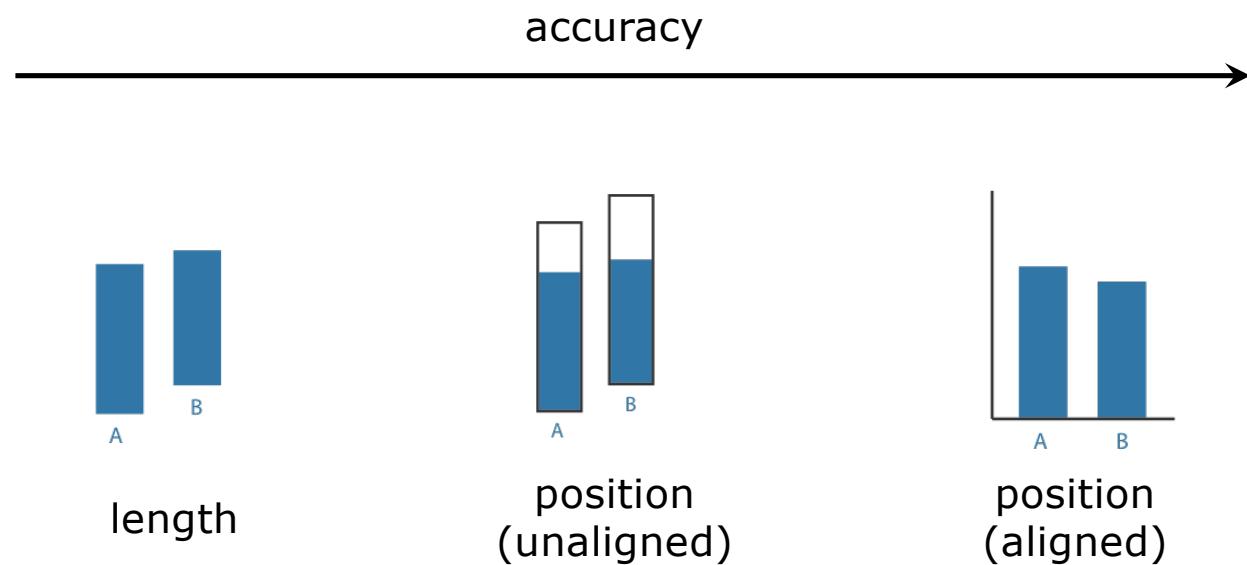


Accuracy: experimental results

Jeffrey Heer and Michael Bostock. "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design." (2010)

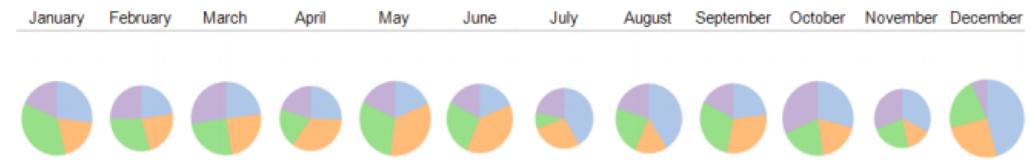
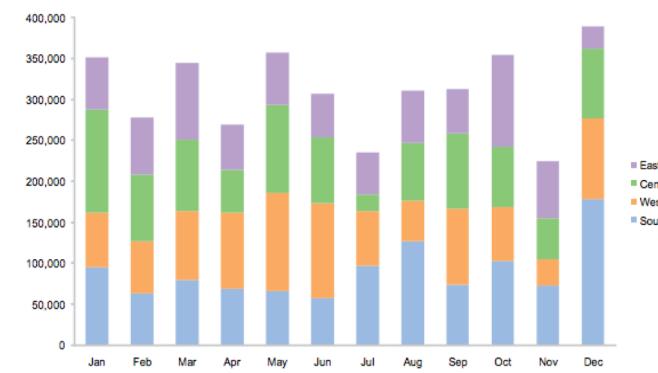
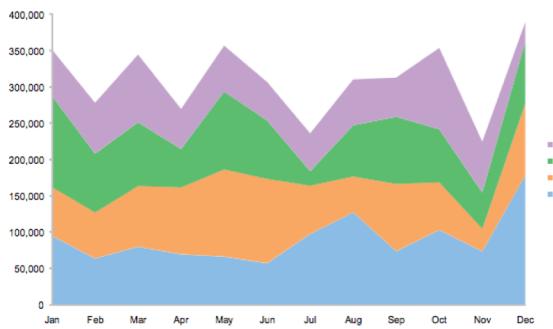
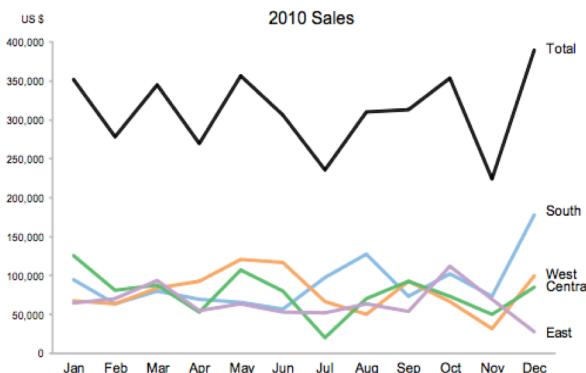


Accuracy: an example



Accuracy: an example

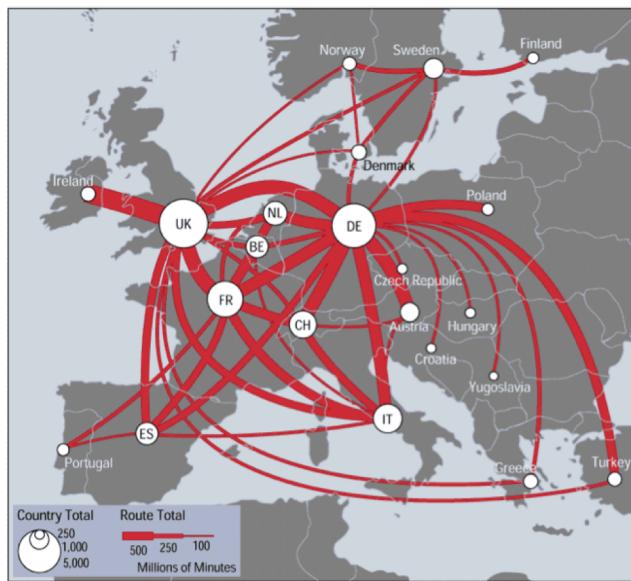
- Which one is more effective to compare sale trends across regions?



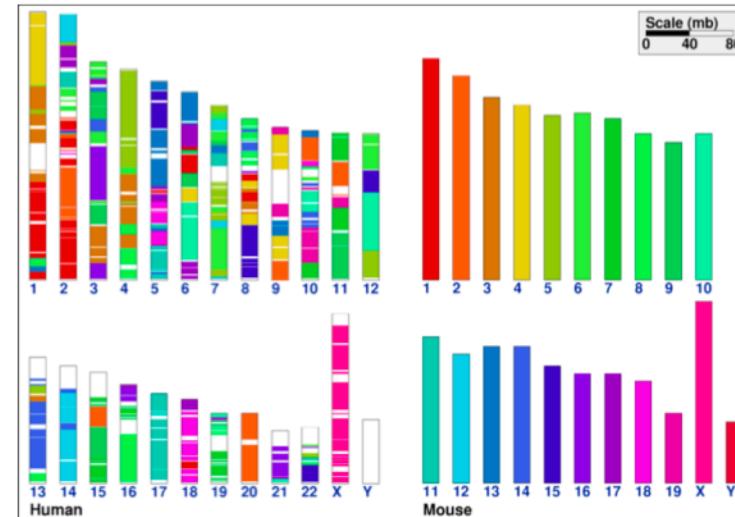
Example taken from Stephen Few: [link](#)

Discriminability

- ❑ How many values (or bins) can we distinguish for any given channel?



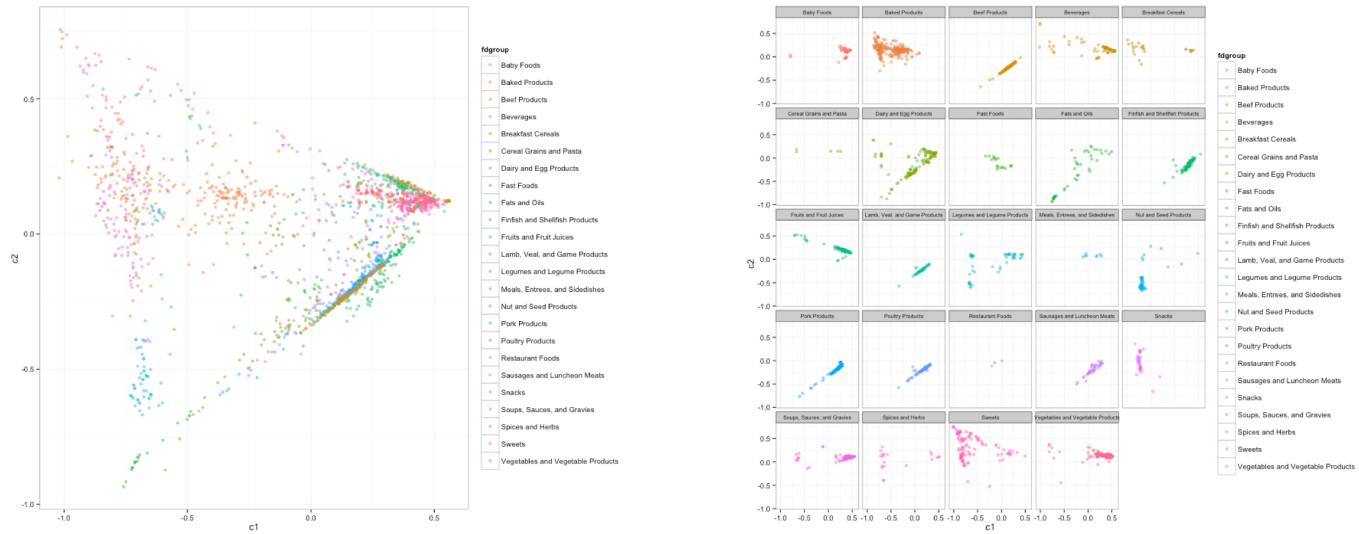
Line width



Color (hue)

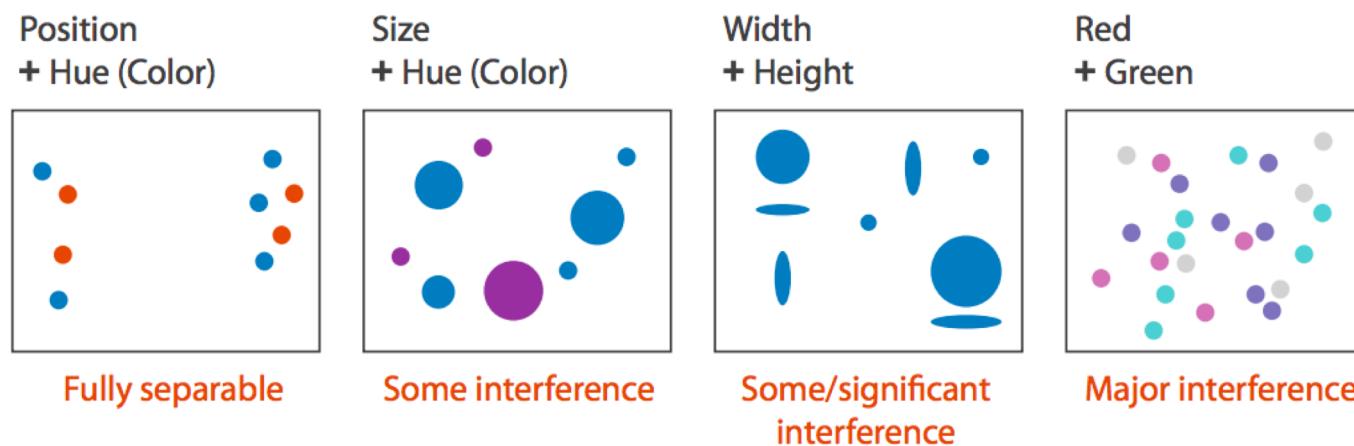
Discriminability: guidelines

- Most channels can deal with a **very low number of bins**.
- If you have too many bins
 - ▶ group data further
 - ▶ change channel
 - ▶ trade-off space with other channels



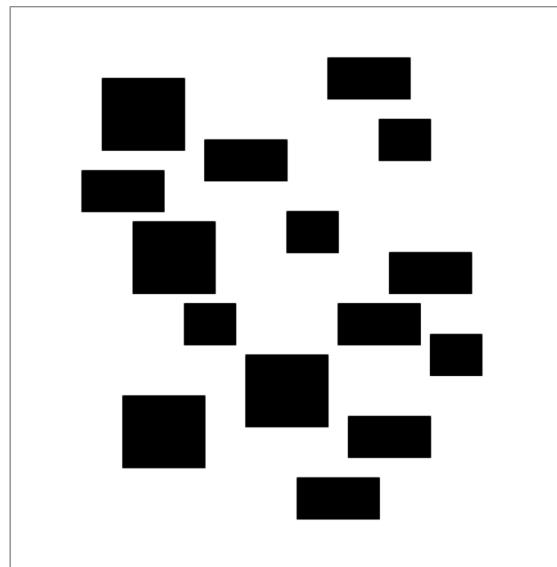
Separability

- Visual channels are not independent: they can be fully **separable** or almost **integral**.
- Designer must take into account dependencies between channels.
- Examples:



Separability: example

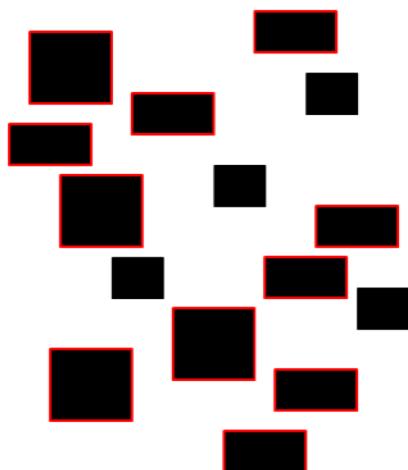
Try group these by same width or height ...



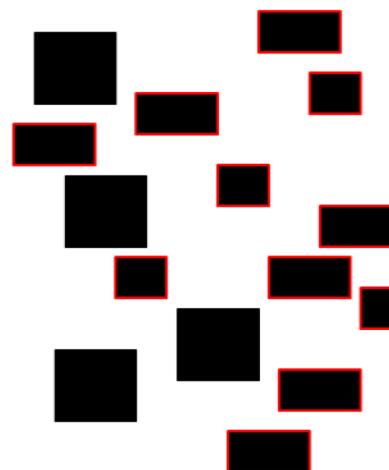
It's hard! Because width and height interfere one to another

Separability: example

Same Width

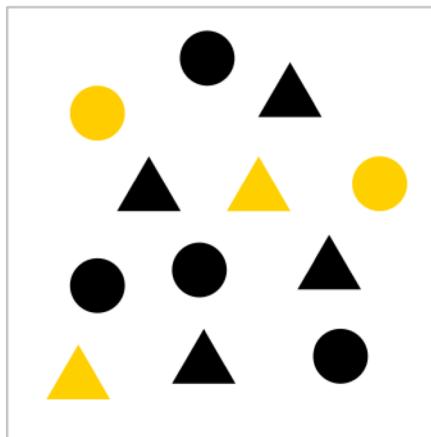


Same Height

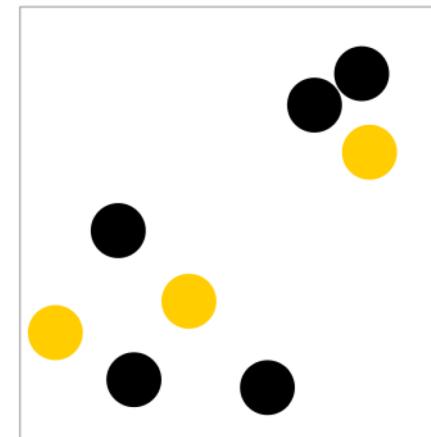


Separability: example

Color and Shape

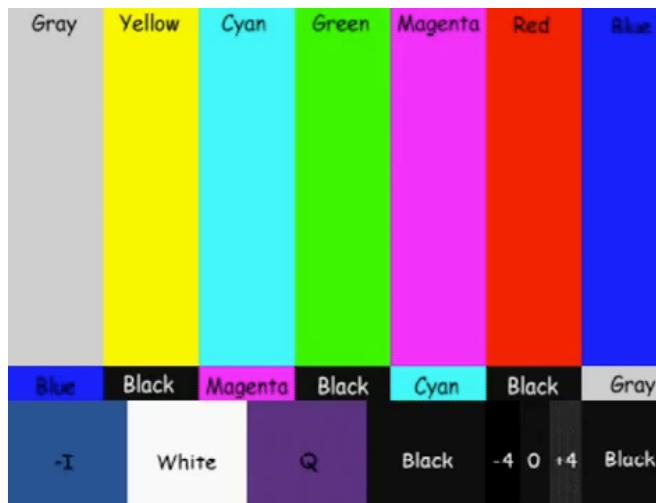


Position and Color



Popout

- Some channels provide a visual popout: makes a distinct item standing out from others.



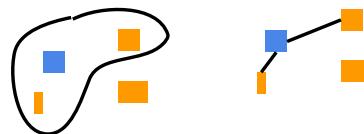
<http://www.csc.ncsu.edu/faculty/healey/PP/>

Separability and popout: guidelines

- Do not encode data with too many non-spatial visual channels (3/4 at max).
- Integral vs separable channels:
 - ▶ Use separable channels to encode different attributes.
 - ▶ Use integral channels to encode values of a single attribute (in a more effective way).
- To direct attention (e.g., selection) use channels that provide a good visual popout.

Grouping

- ❑ The perception of **grouping** is visually represented with three major approaches:
 - ▶ Containment and Connection (**marks**)

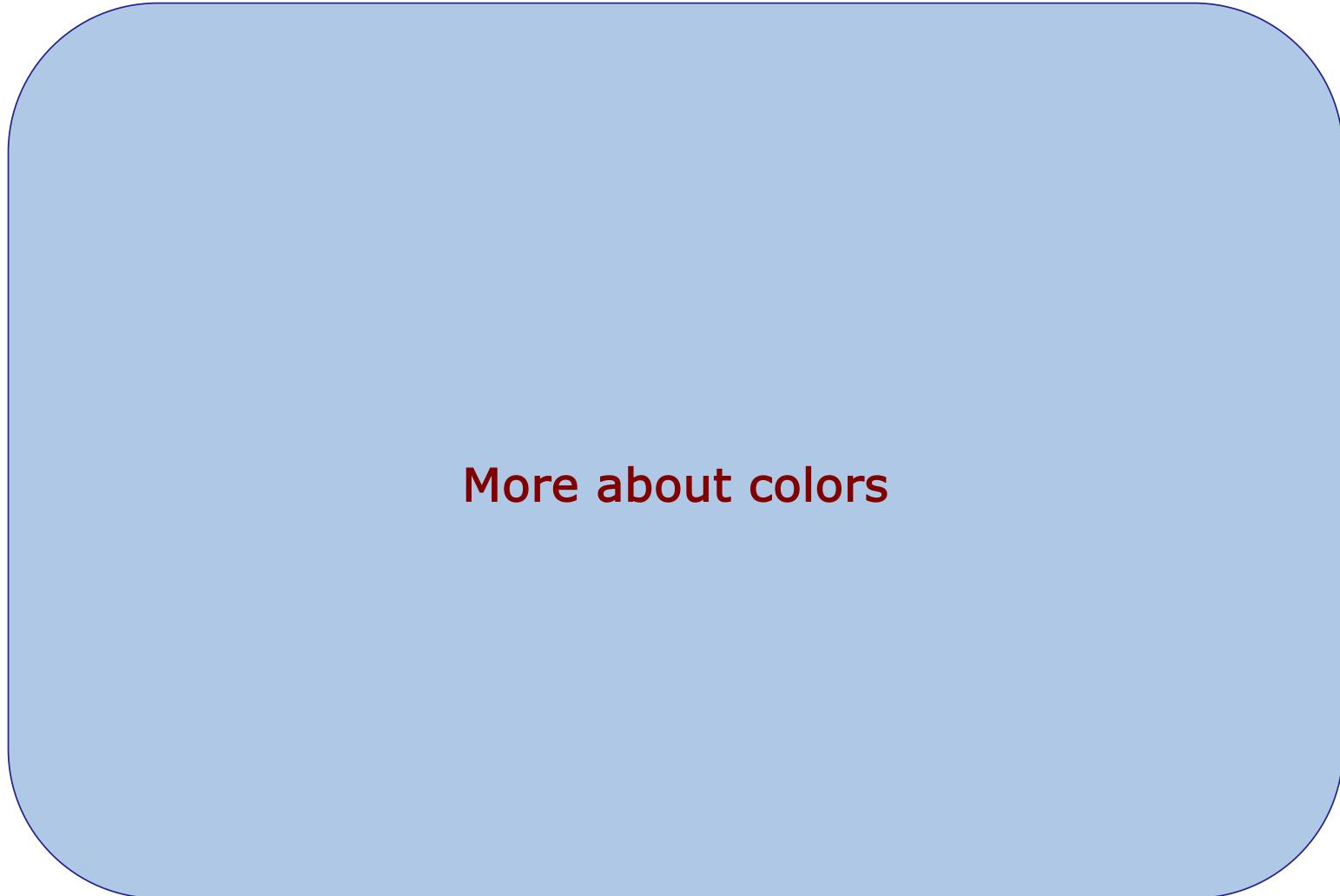


- ▶ Proximity (**spatial channels**)



- ▶ Similarity (**identity channels**)





More about colors

How do we perceive color?

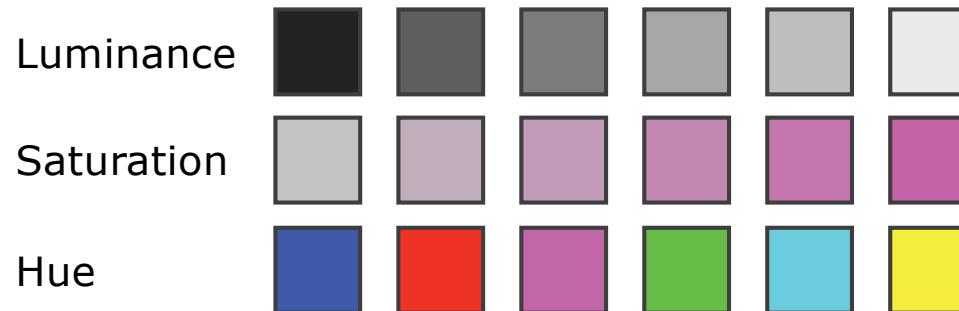
- Our visual system process signals into 3 opponent color channels:
 - ▶ Red-Green
 - ▶ Blue-Yellow
 - ▶ Black-White

} chromaticity

luminance
- Luminance channel conveys more information than chromaticity channels.
 - ▶ Chromaticity is what we informally call **color**.
 - ▶ Luminance allows **edge** detection.
- People with color deficiency (**color blindness**) has reduced perception of Red-Green or (more rarely) Blue-Yellow channels.

Encoding data with colors

- We decompose colors in three channels:
 - ▶ Luminance: how bright is?
 - ▶ Saturation: how *colorful* is?
 - ▶ Hue: which *color* is?



Encoding with luminance

- ❑ It is a **magnitude** channel suitable to encode **ordered** data.
- ❑ It allows to **discriminate** about 5 different (non-contiguous) values.
- ❑ When used to encode data, it might not be used to provide enough contrast to detect edges.
 - ▶ Luminance is the most important channel to provide contrast.
 - ▶ E.g., for making text readable standard guideline is a 10:1 ratio (where 3:1 is the minimum).

2:1 ratio

3:1 ratio

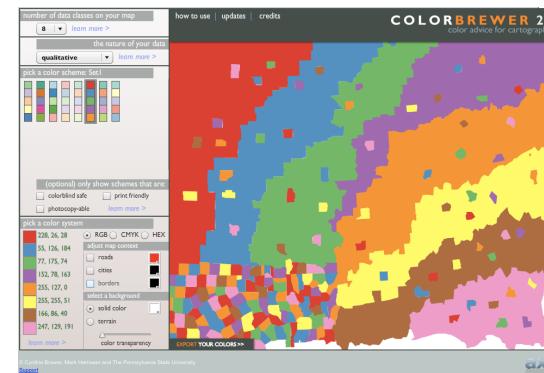
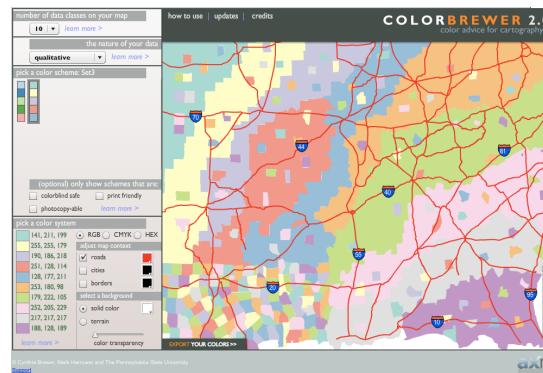
5:1 ratio

8:1 ratio

10:1 ratio

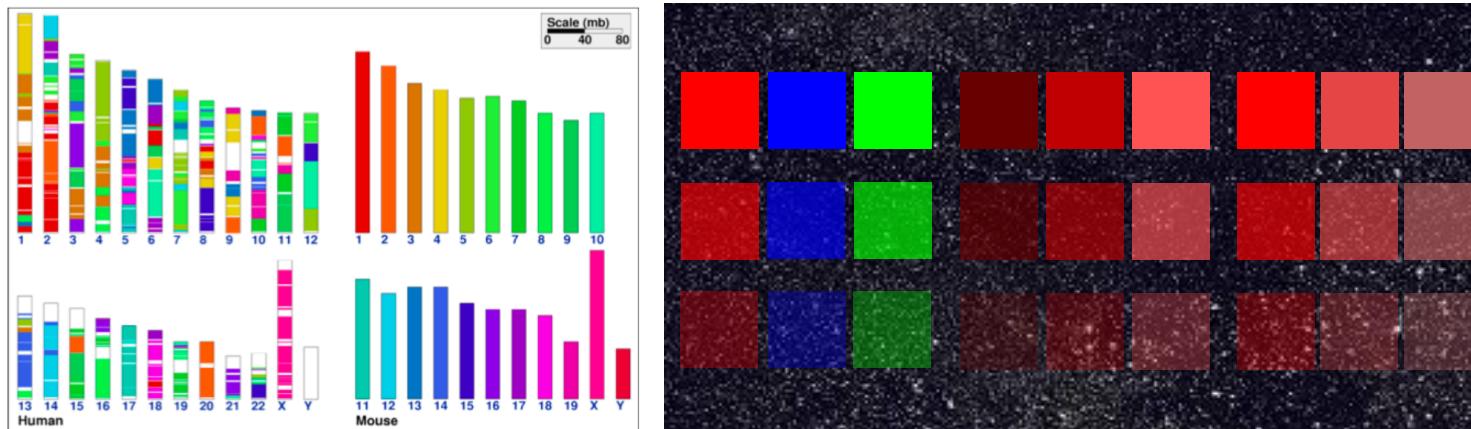
Encoding with saturation

- It is a **magnitude** channel suitable to encode **ordered** data.
- It is less accurate than luminance and allows to discriminate about 3 different (non-contiguous) values.
- Separability issues
 - ▶ Difficult to perceive in small areas.
 - ▶ Hue and saturation are not well separable.
- Guidelines
 - ▶ Use high-saturated colors for small areas.
 - ▶ Use low-saturated colors suits better large areas.



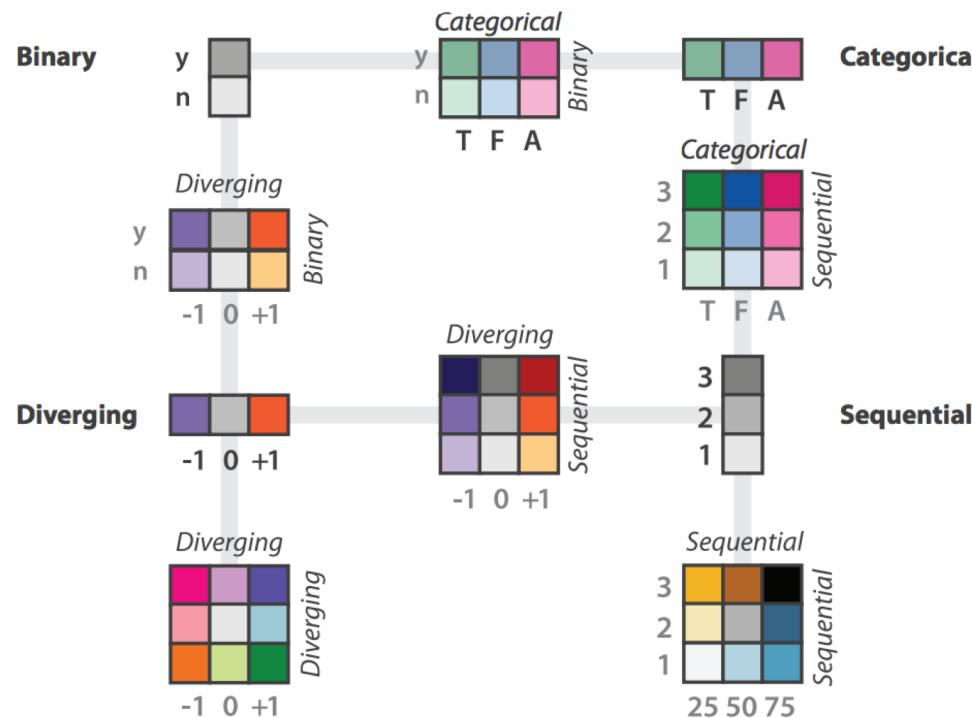
Encoding with hue

- ❑ Hue is an effective **identity** channels to encode **categorical** values but it does provide any implicit perceptual ordering.
- ❑ It allows to **discriminate** about 7 different (non-contiguous) values.
- ❑ It can be used with **transparency**, which is instead has strong interaction with luminance and saturation.



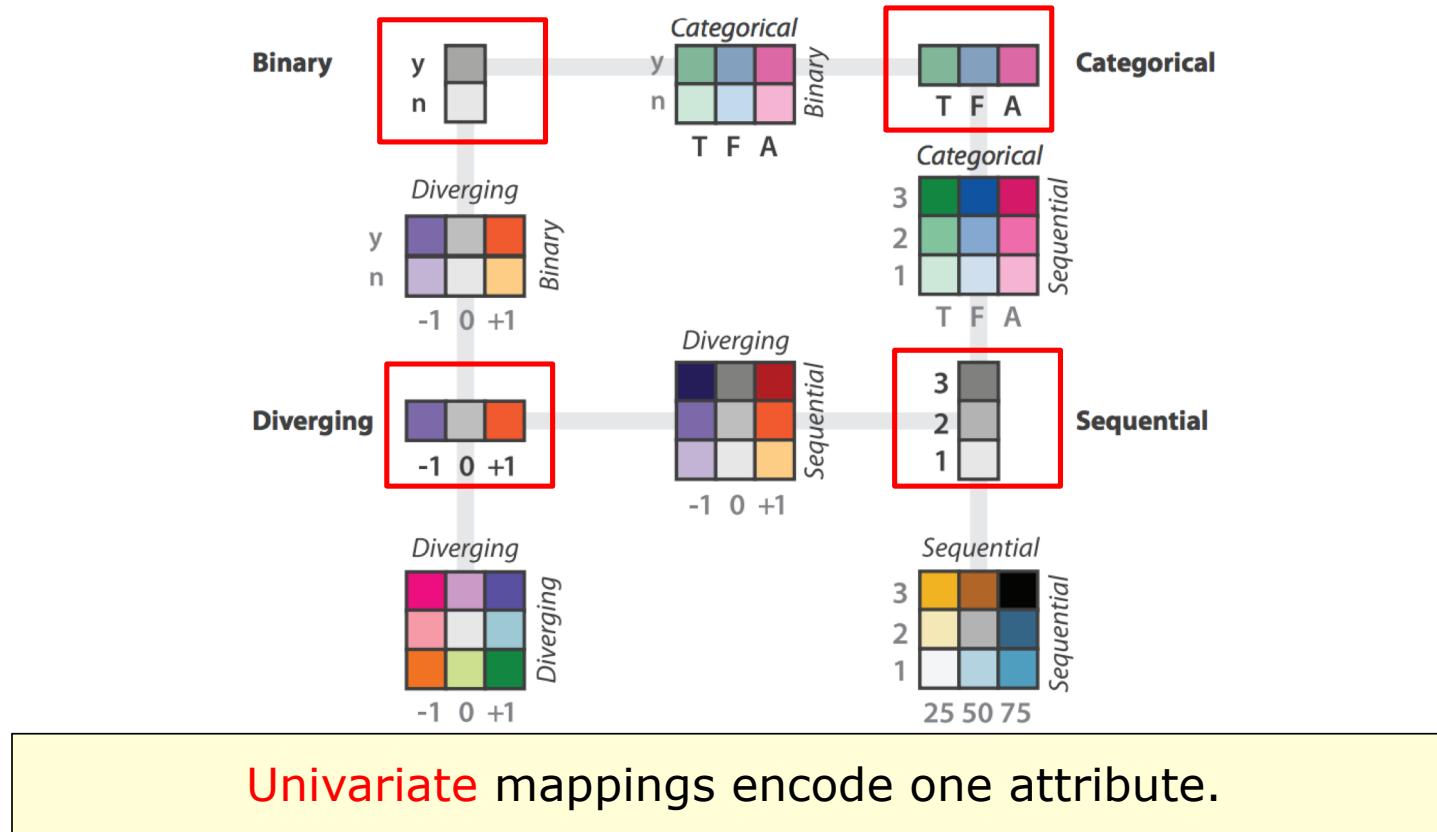
Colormaps

- ❑ A colormap is the mapping chosen to encode one (or two) attributes with color.



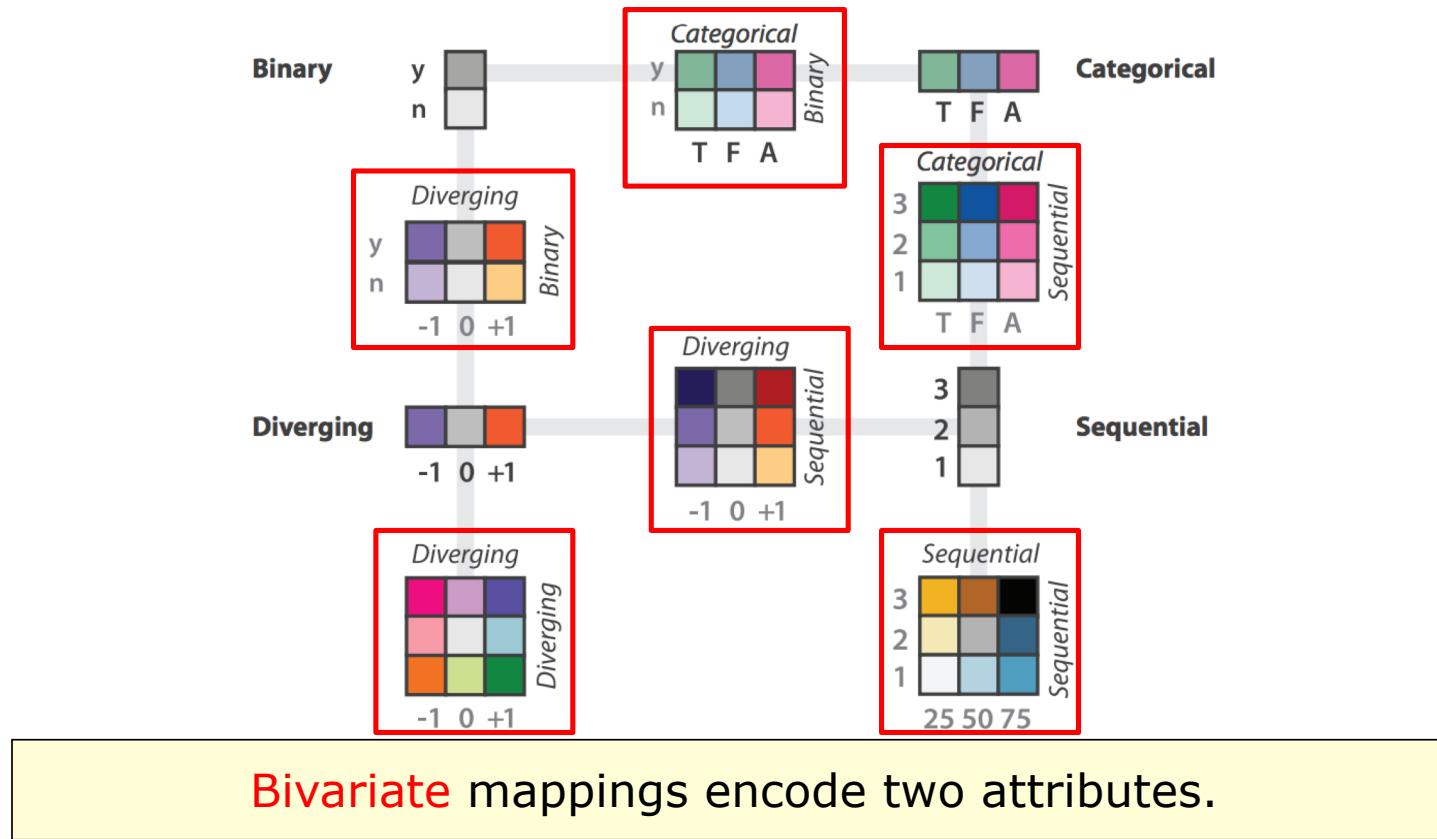
Colormaps

- ❑ A colormap is the mapping chosen to encode one (or two) attributes with color.



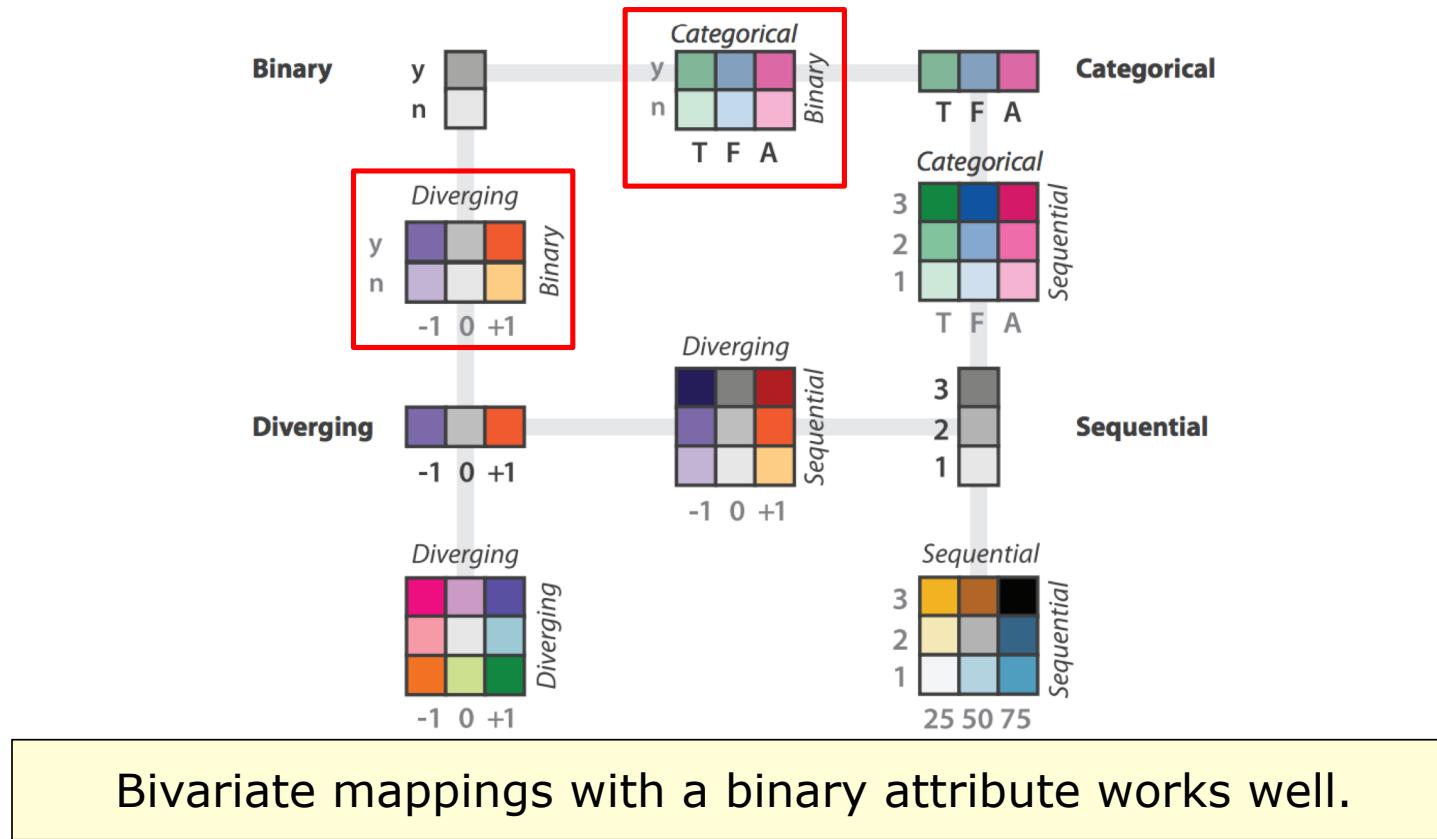
Colormaps

- ❑ A colormap is the mapping chosen to encode one (or two) attributes with color.



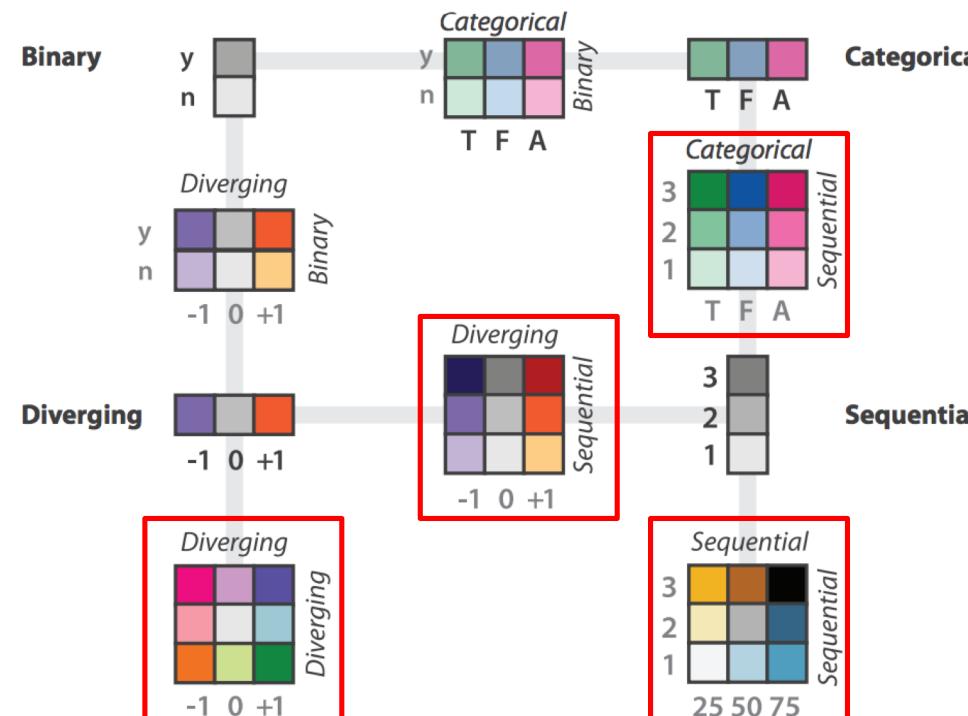
Colormaps

- ❑ A colormap is the mapping chosen to encode one (or two) attributes with color.



Colormaps

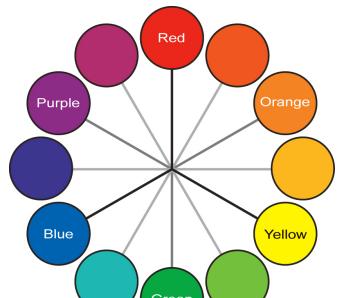
- ❑ A colormap is the mapping chosen to encode one (or two) attributes with color.



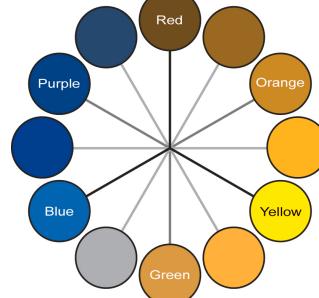
Other bivariate mappings are possible but less effective

Dealing with color deficiency

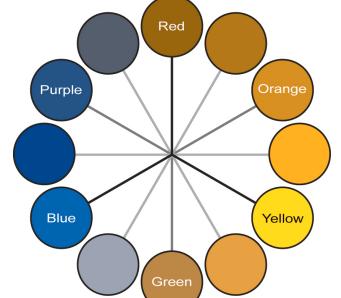
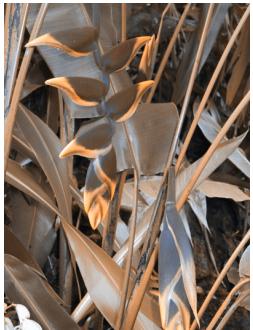
- ❑ Use only two channels



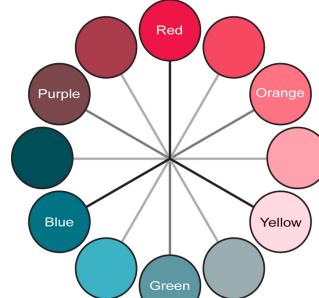
Normal



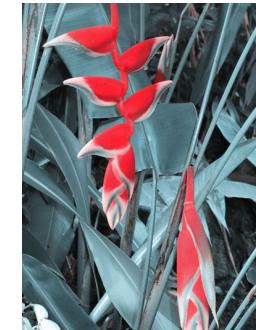
Protanope



Deutanope

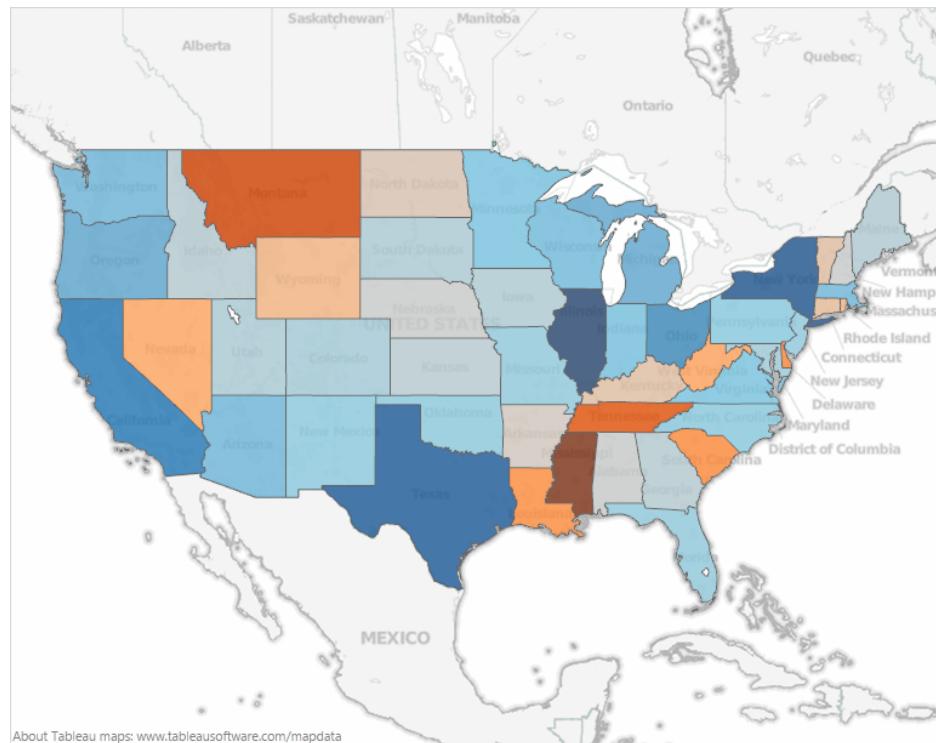


Tritanope



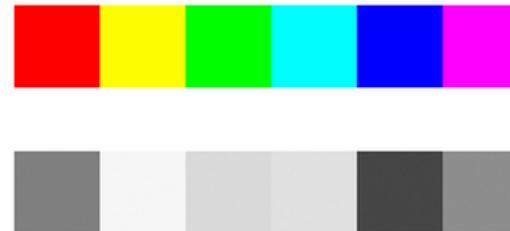
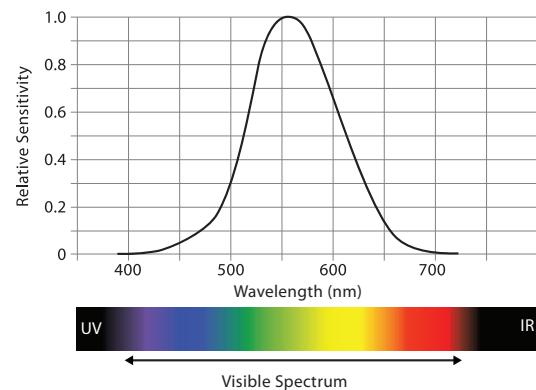
Dealing with color deficiency (2)

- ❑ Use colorblind-safe mappings



Design colormaps: the color space

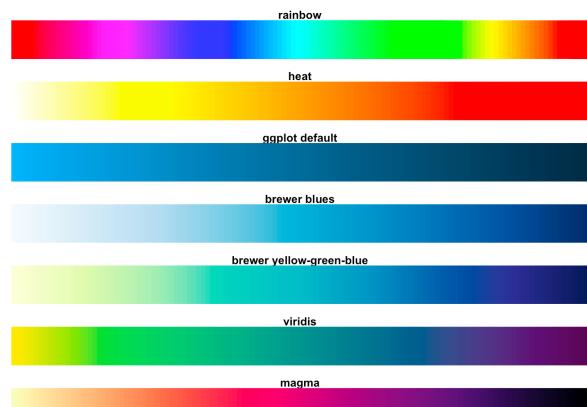
- ❑ RGB
 - ▶ red, green, blue channels
 - ▶ good for display hardware but not useful for encoding
- ❑ HSL/HSV
 - ▶ hue, saturation and lightness/value channel
 - ▶ much better for encoding...
 - ▶ ... but lightness/value \neq luminance



Colors with
same L/V
converted
to grayscale

Infact...

- ❑ Some standard mappings (e.g., rainbow) are not perceptually uniform and might not be effective to encode ordered attributes.

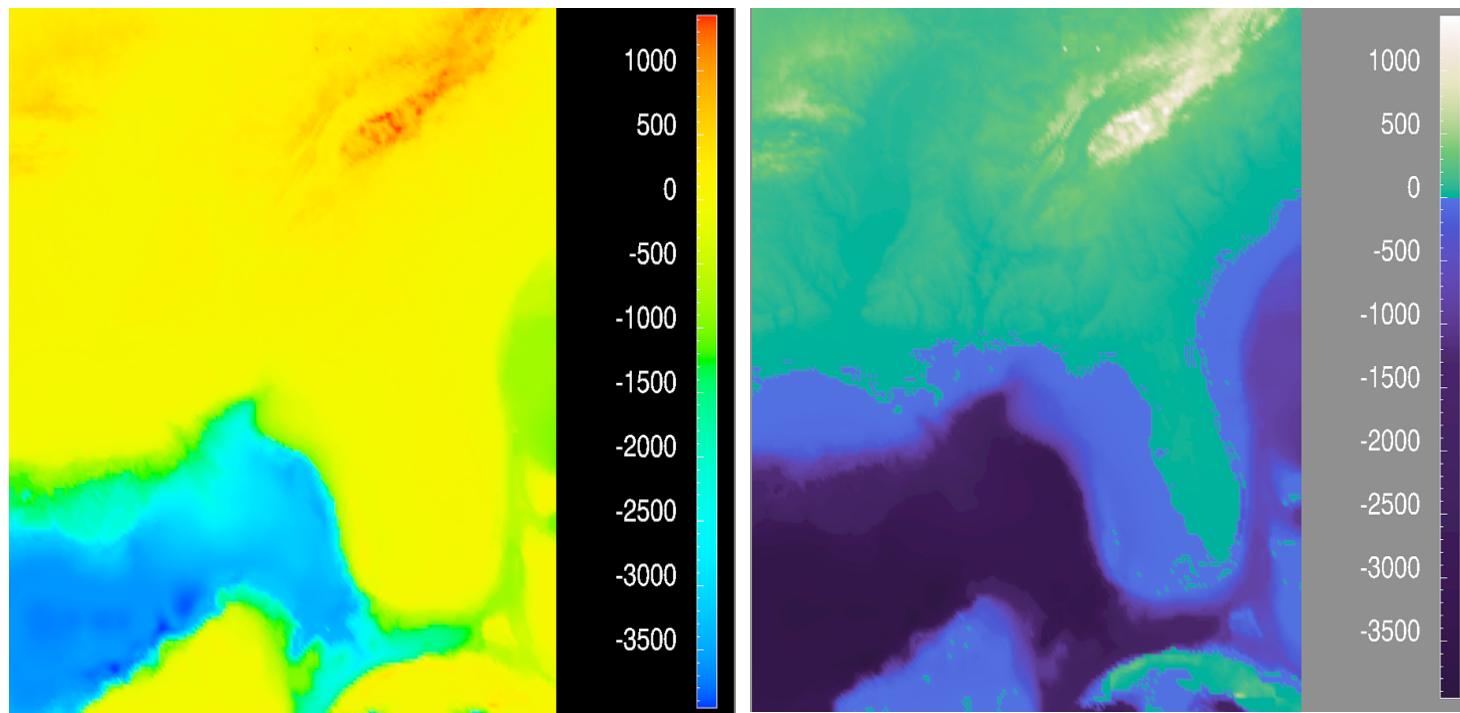


Colormaps



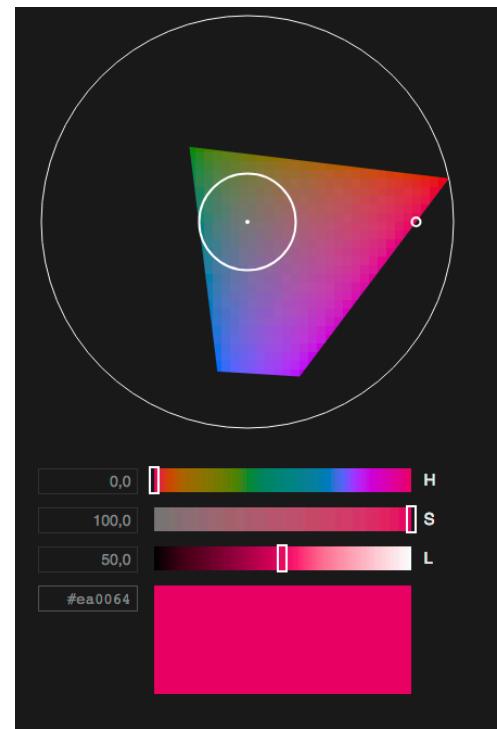
Converted to
grayscale

An example: design of diverging colormap



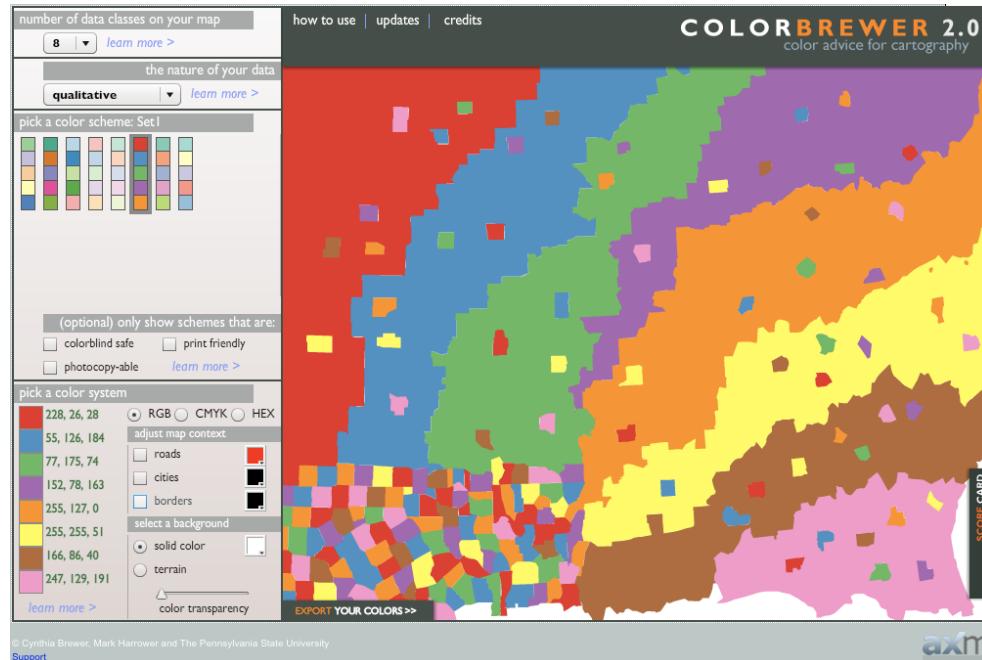
Design colormaps: get it right

- ❑ Use CIELab and CIELUV spaces that are perceptually uniform
 - ▶ Human friendly HSL (HSL_{uv}): <http://www.hsluv.org/>



Design colormaps: tools

- Use CIELab and CIELUV spaces that are perceptually uniform
 - ▶ Human friendly HSL (HSL_{uv}): <http://www.hsluv.org/>
 - ▶ Color Brewer: <http://colorbrewer2.org>



Design colormaps: tools

- Use CIELab and CIELUV spaces that are perceptually uniform
 - ▶ Human friendly HSL (HSL_{uv}): <http://www.hsluv.org/>
 - ▶ Color Brewer: <http://colorbrewer2.org>
 - ▶ Several libraries provide some helps (e.g., see https://seaborn.pydata.org/tutorial/color_palettes.html)

Additional visual channels

Size channel

- ❑ It is a magnitude channel
- ❑ Interacts with shape and color channels
- ❑ Accuracy depends on dimensionality:
 - ▶ length (1D size) is very accurate
 - ▶ area (2D size) is moderately accurate
 - ▶ volume (3D size) is not accurate

Angle/Tilt channel

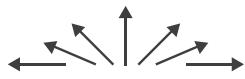
- Tilt vs angle



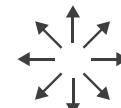
- It is a magnitude channel



Sequential ordered
line mark or arrow glyph



Diverging ordered
arrow glyph



Cyclic ordered
arrow glyph

- Perceptual accuracy is not uniform

- ▶ More accurate around 0° , 90° , and 45°

Shape channel

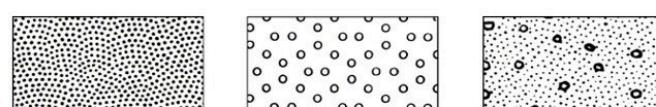
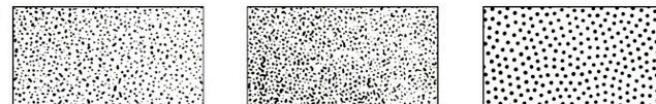
- ❑ It is an identity channel
- ❑ Generally used with point marks
- ❑ Major interaction with size and color channels

Curvature channel

- ❑ It is a magnitude channel.
- ❑ It can be used only with line marks.
- ❑ It is very inaccurate channel.

Texture/Stippling channels

- ❑ Encoding data with patterns based on three dimensions:
 - ▶ orientation
 - ▶ size
 - ▶ density
- ❑ Combined together, they can encode a categorical attribute.
- ❑ Otherwise, a dimension can be also used to encode an ordered attribute.
- ❑ Heavily used in the past (with older printing technologies) are now mainly used with **line marks**.



Motion channels

- They are identity channels that can be used for animated/interactive visualization:
 - ▶ Direction
 - ▶ Speed
 - ▶ Flickering
- Extremely salient and well **separable from static channels**.
- Separability of different motion channels as well the possibility to encode different values is not well studied.
- Often used to encode binary attribute or to highlight items in interactive visualization.

Summary on visual channels

A visual channels summary

Categorical
what/where

- planar position 
- color hue 
- shape 
- stipple pattern 

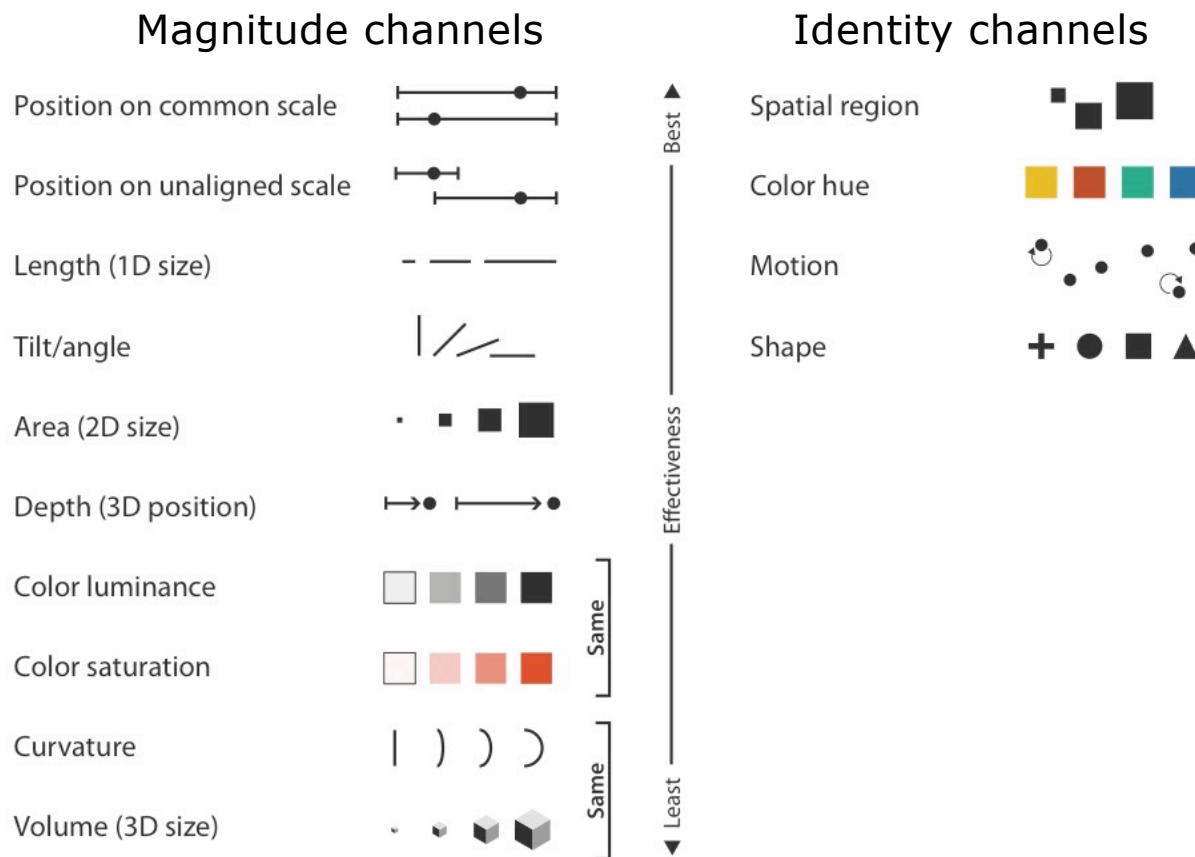
Relationship, Same Category
Grouping

- Containment (2D) 
- Connection (1D) 
- Similarity (other channels) 
- Proximity (position) 

Ordered: Ordinal/Quantitative
How much

- position on common scale 
- position on unaligned scale 
- length (1D size) 
- tilt, angle 
- area (2D size) 
- curvature 
- volume (3D size) 
- lightness black/white 
- color saturation 
- stipple density 

Ranking

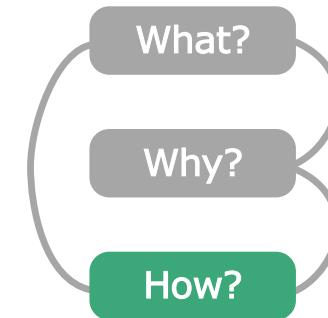
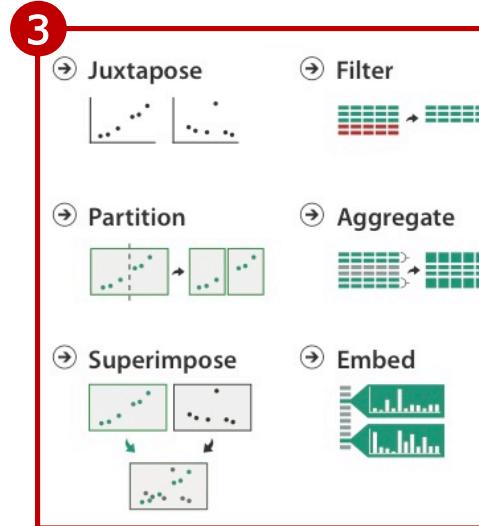
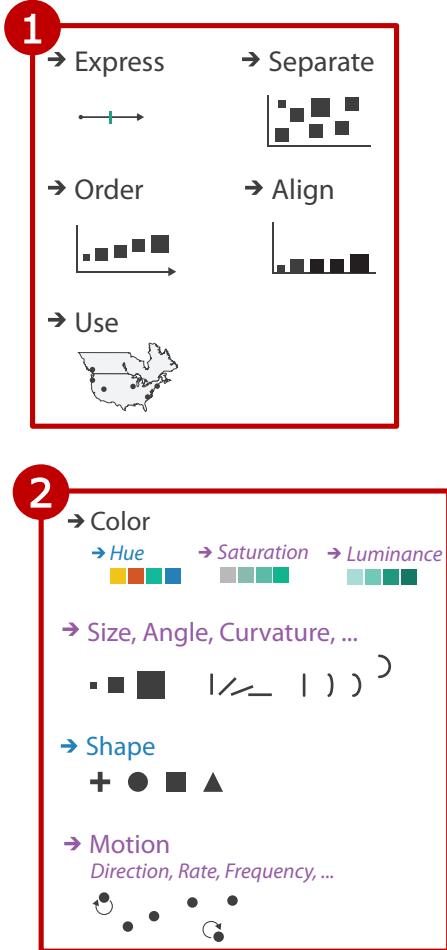


Hands-On

Working with Colors in Seaborn

GENERAL GUIDELINES

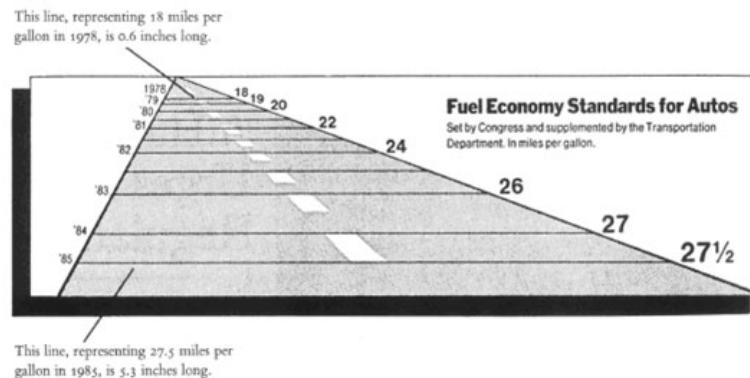
Idiom design: an overview



- Encode data
 - ▶ Arrange data in space **1**
 - ▶ Map data to other channels **2**
- Deal with complexity **3**

Tufte's Principles

- Graphical integrity (e.g., no data distortion)



Tufte's Principles

- Graphical integrity (e.g., no data distortion)
- Maximize data-ink

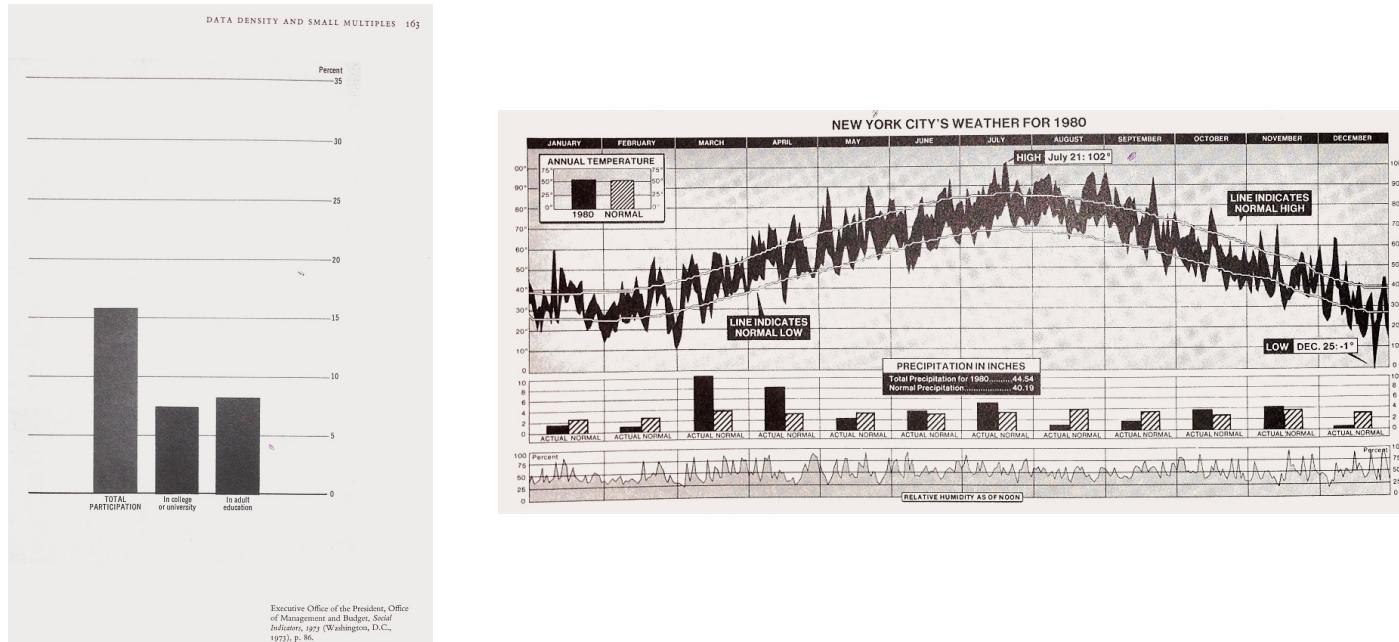
Remove
to improve
(the **data-ink** ratio)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

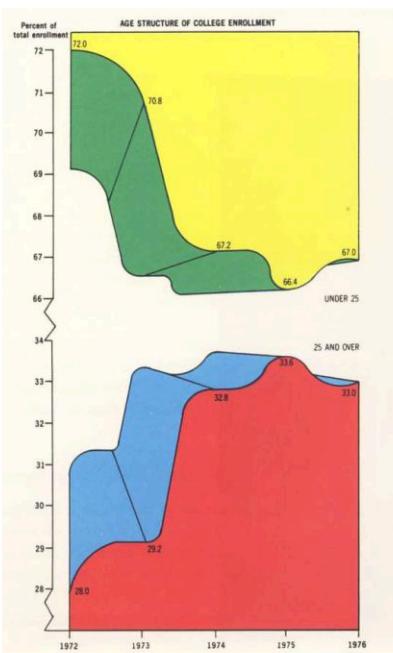
Tufte's Principles

- Graphical integrity (e.g., no data distortion)
- Maximize data-ink
- Create high-density graphs



Tufte's Principles

- Graphical integrity (e.g., no data distortion)
- Maximize data-ink
- Create high-density graphs
- Avoid unnecessary graphical effects



No unjustified 3D

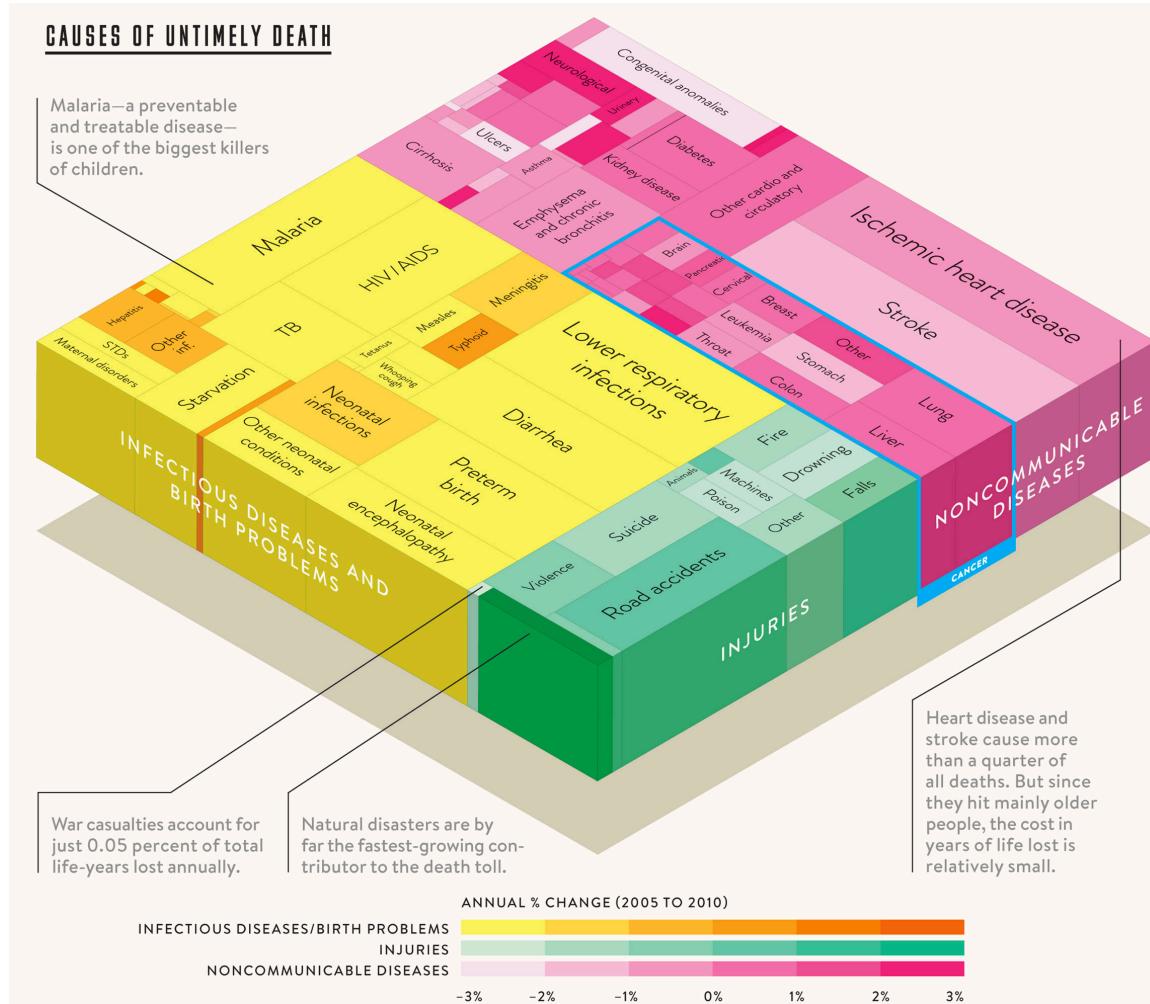
❑ Cons

- ▶ Low accuracy on depth channel
- ▶ Dangers of perspective distortion
- ▶ Occlusion of information
- ▶ Text legibility

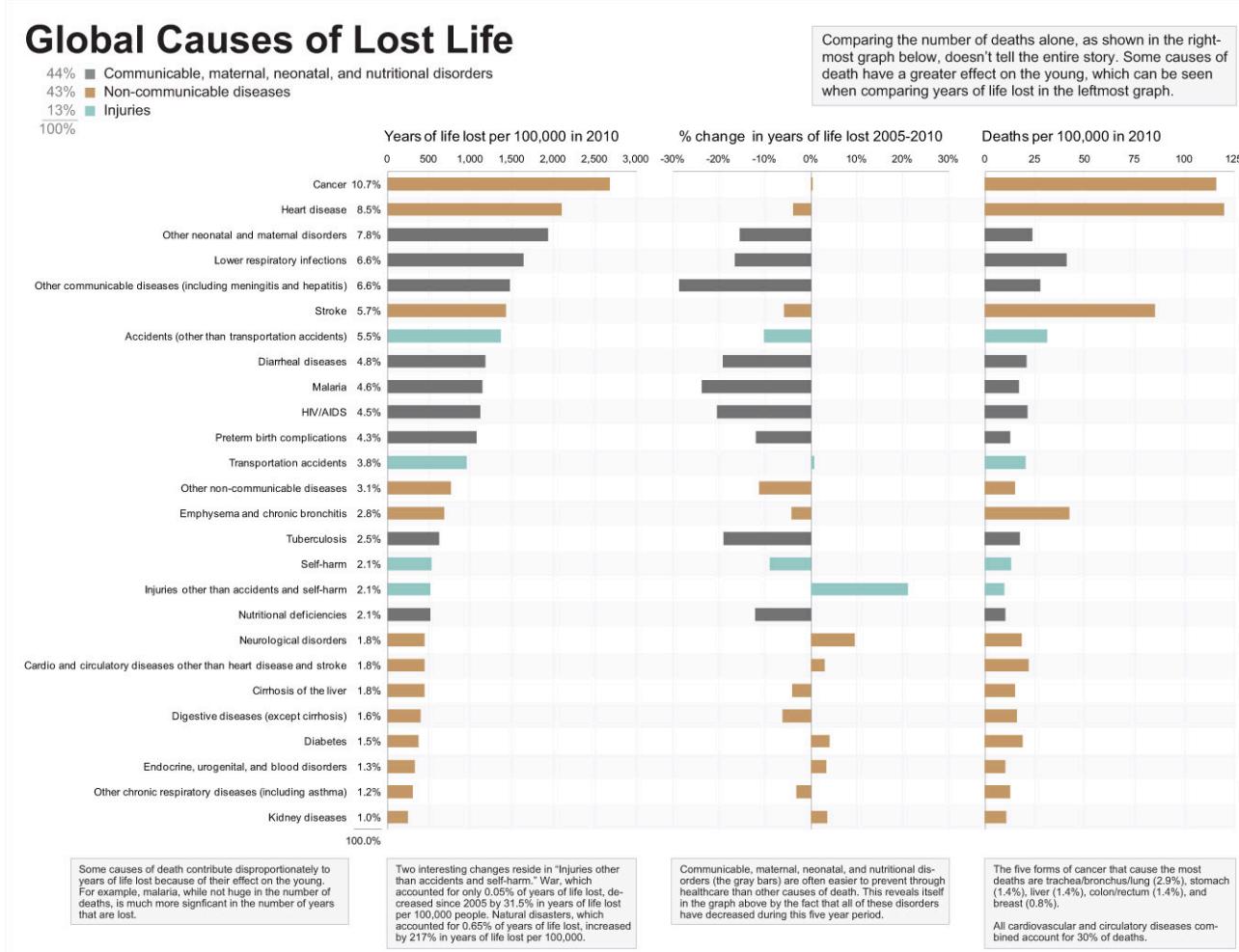
❑ Pro

- ▶ Better understanding of shapes

No unjustified 3D: an example



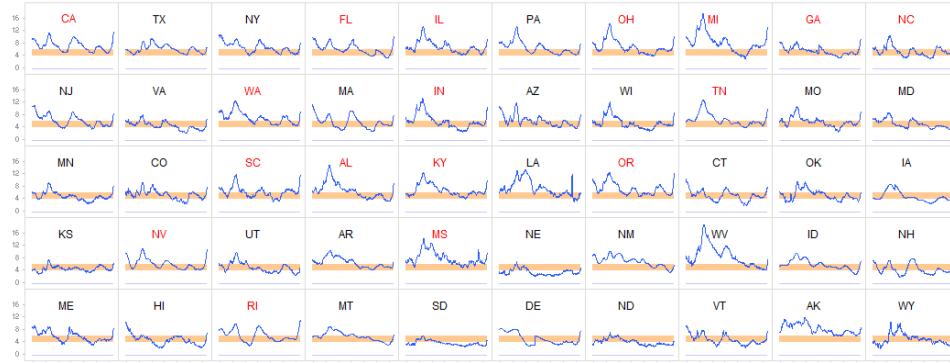
No unjustified 3D: an example



Eyes beat memory

- When dealing with complex data do not **underestimate** visual perception and/or **overestimate** working memory.
- Transitions or not connected views lead to a bigger cognitive load
- Trade space for memory! (high-density graph)
 - ▶ Multiple connected views
 - ▶ Superimpose
 - ▶ Aggregation

Monthly Unemployment Rates by State, Jan 1976 - Apr 2009

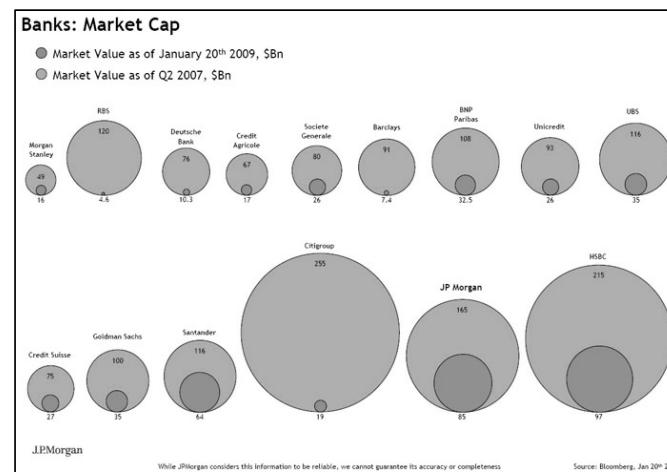
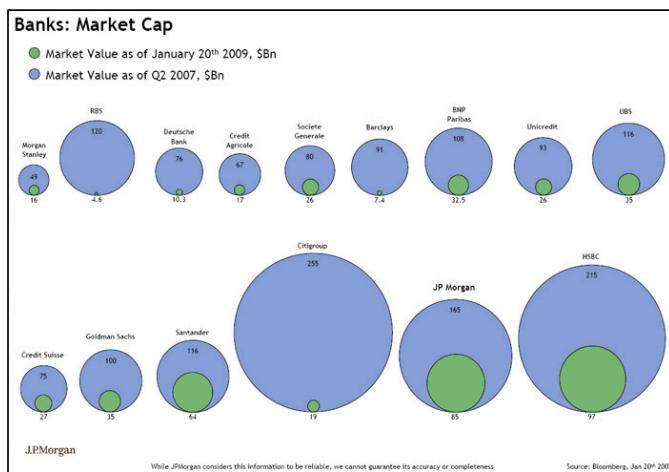


Source: Bureau of Labor Statistics

Notes: The orange band denotes a "normal" unemployment rate (4%-6%);
State code in red: unemployment rate in April 2009 is higher than the US average

Get it right in black and white

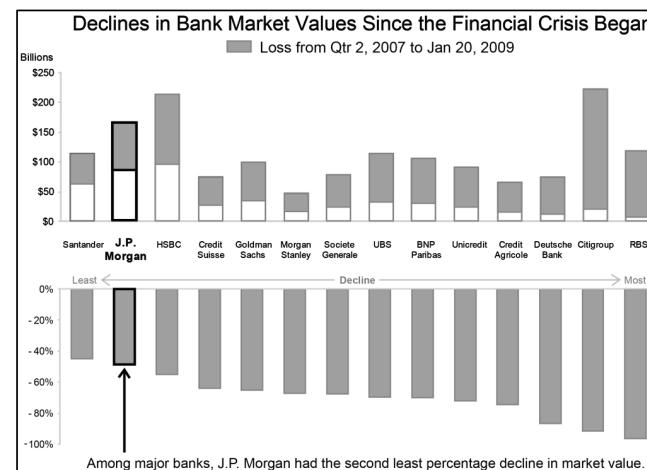
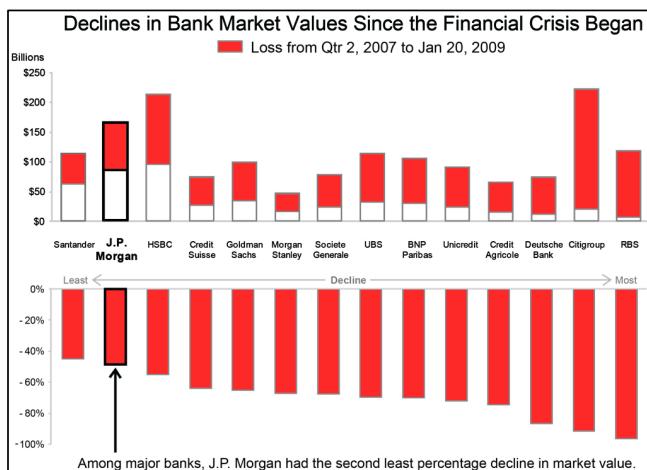
- ❑ Make sure that most important data is visible also when image is transformed to black and white (**check it!**).
 - ▶ use space and luminance/saturation as major source of information
 - ▶ others channels are to be used for less important data



Stephen Few: <http://www.perceptualedge.com/example18.php>

Get it right in black and white

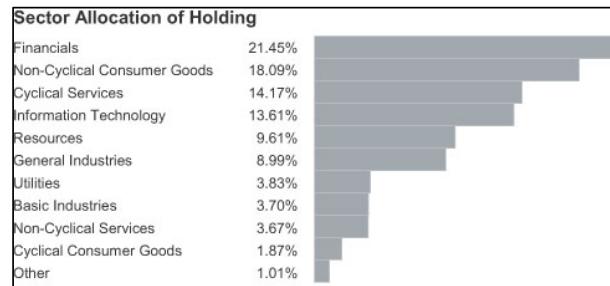
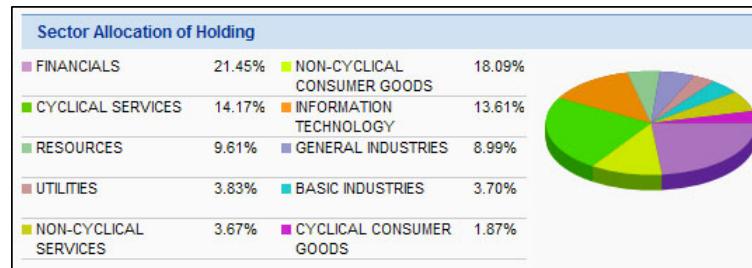
- ❑ Make sure that most important data is visible also when image is transformed to black and white (**check it!**).
 - ▶ use space and luminance/saturation as major source of information
 - ▶ others channels are to be used for less important data



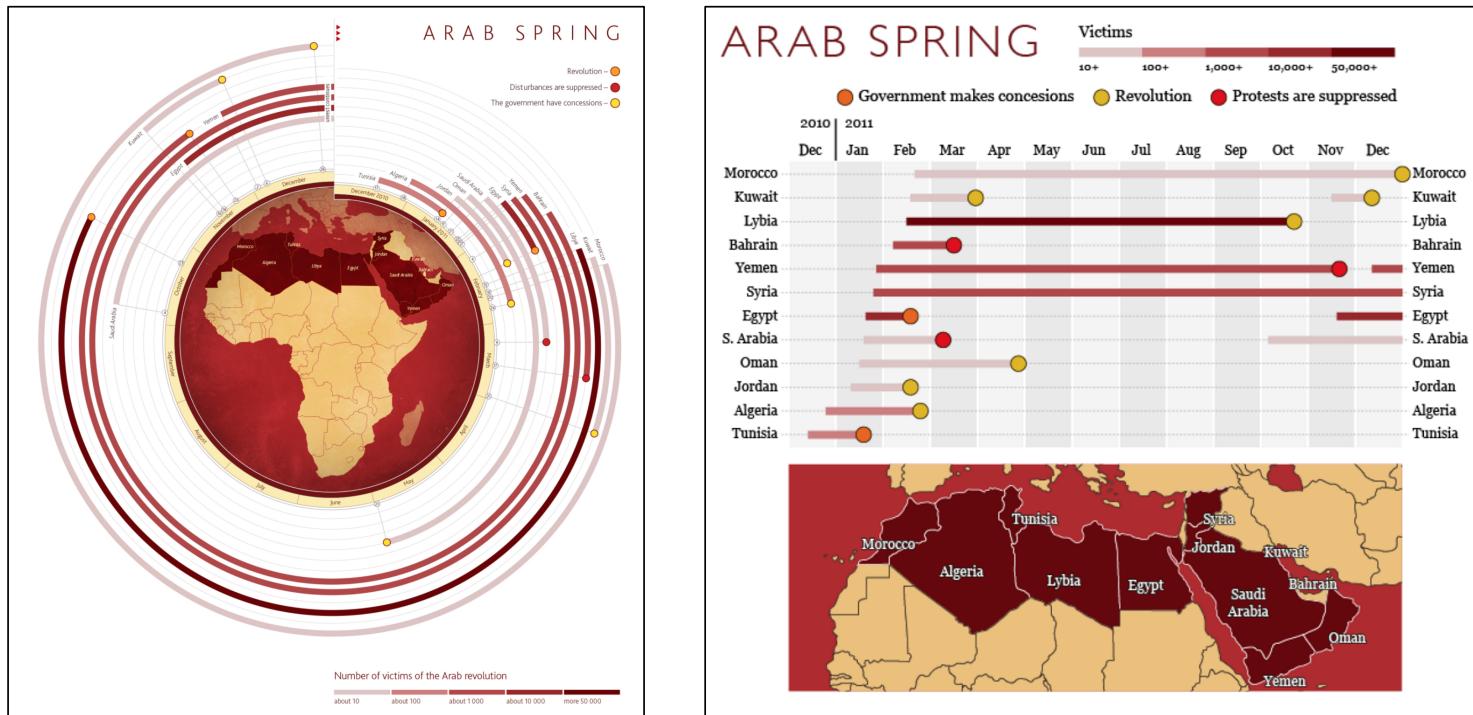
Stephen Few: <http://www.perceptualedge.com/example18.php>

Function first!

- ❑ A beautiful visualization design leads to more engagement.
- ❑ However, function is **essential** and form can be more easily improved **iteratively**.



Function vs form: an example



<http://www.thefunctionalart.com/2015/02/redesigning-circular-timeline.html>