



CS-GY 6313 B: Information Visualization

9/26/2024

Logistics

- Assignment 1 due next week
- Google Colab okay
 - Export final notebook to a Jupyter file or python script
- Dataset sources coming soon
 - Will be an announcement on Brightspace, after class today.



Encoding Data Visually

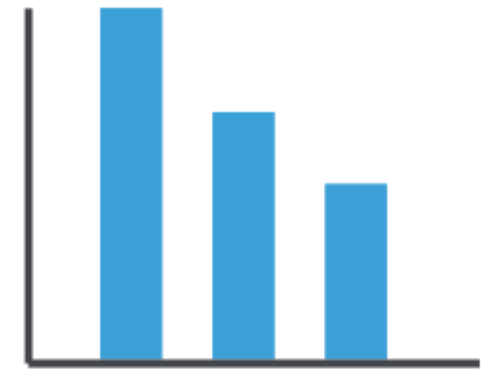
Visual Encoding

- We know how vision works (mostly)
- What is the best way to present information visually?
 - Image-based representations, symbols
- Goals:
 - Efficient extraction of information
 - Minimal ambiguity



Visual Encoding

- How do we systematically analyze these charts?
- Marks and channels
 - Marks: represent entities or links between entities
 - Channels: change the appears of marks based on attribute values



Marks for items

- Marks are basic geometric elements

➔ Points



0D

➔ Lines



1D

➔ Interlocking Areas

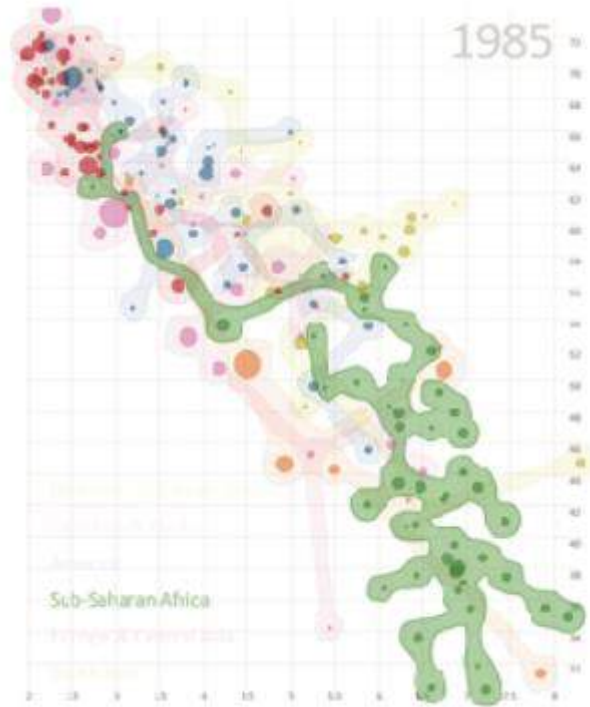


2D

- 3D marks are rare

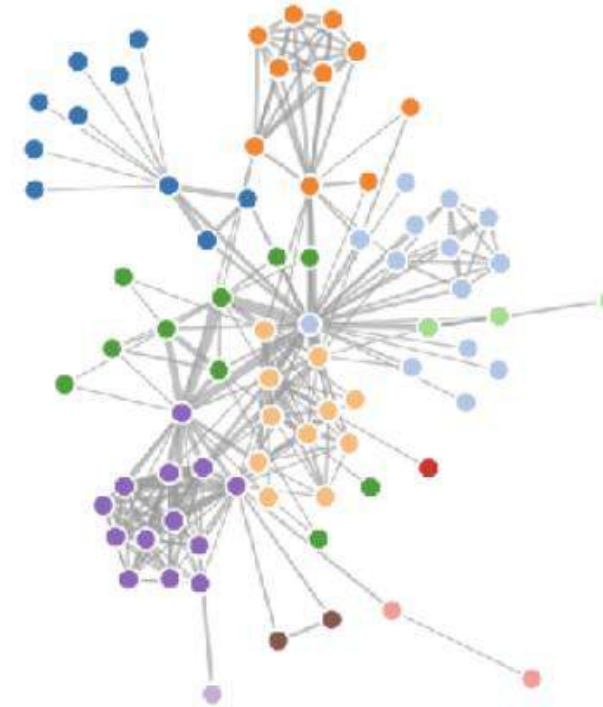
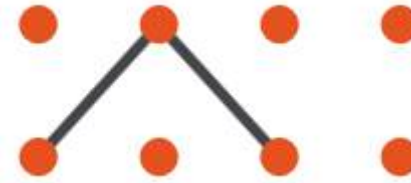
Marks for links

- Containment



<https://vialab.ca/research/bubble-sets>

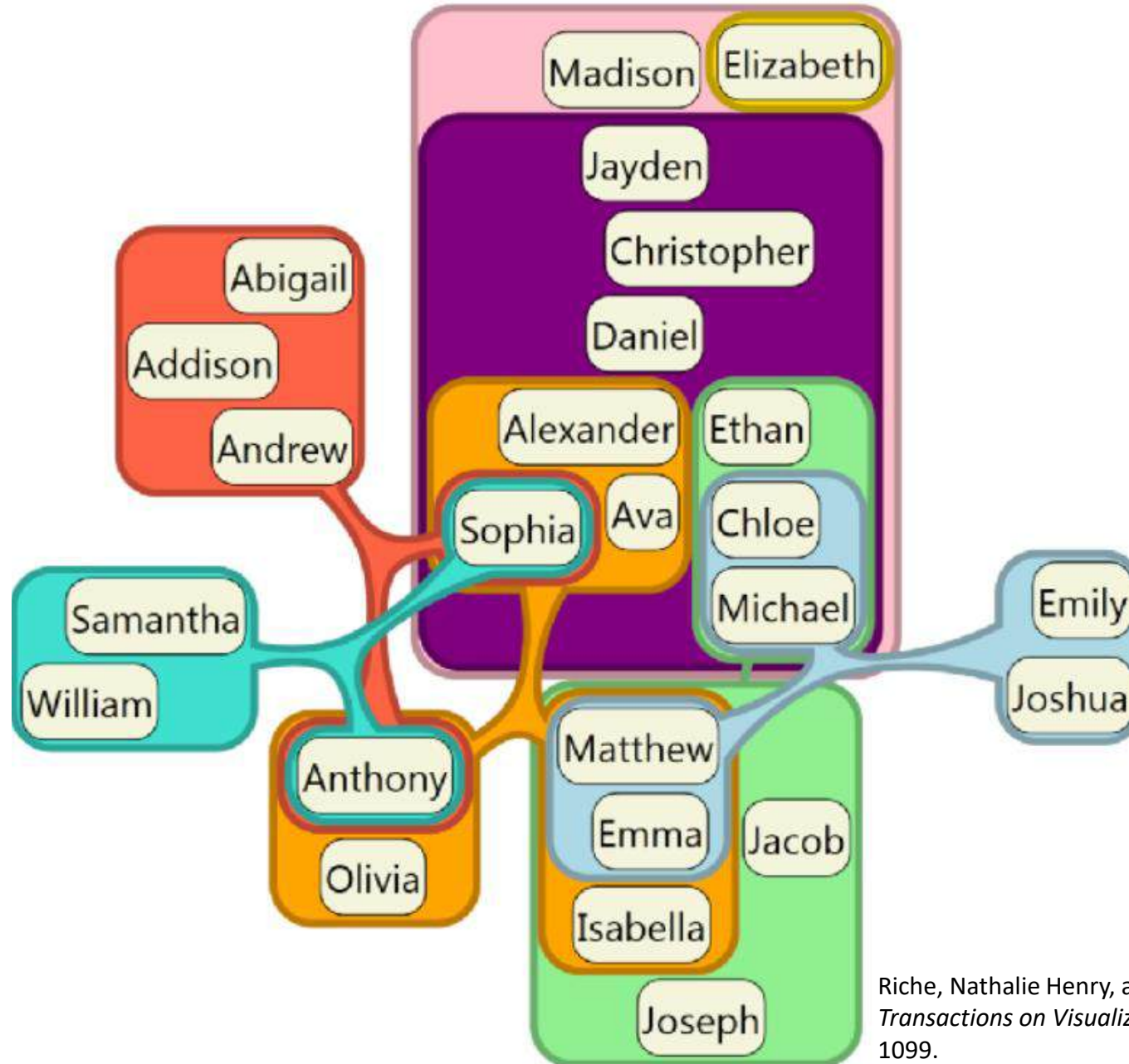
- Connection



<https://observablehq.com/@d3/force-directed-graph-component>

Containments can be nested

- Hierarchical!



Channels

- Control the appearance of marks

- Proportional to or based on attributes

→ Position

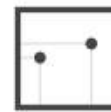
→ Horizontal



→ Vertical



→ Both



→ Color



- Goes by many different names

- Visual channels
- Visual variable
- Retinal channel
- Visual dimension
- ...

→ Shape



→ Tilt



→ Size

→ Length



→ Area



→ Volume



Channels

- Channel properties differ

- Type and amount of info that is conveyed to the human's visual system
- Not all channels are equal

→ Position

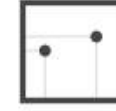
→ Horizontal



→ Vertical



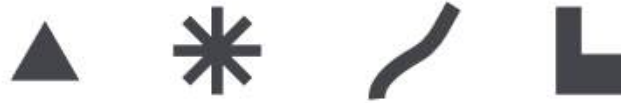
→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area

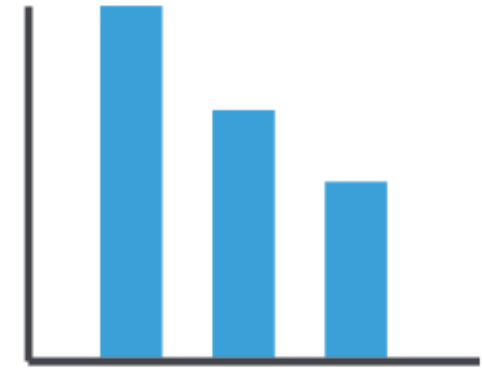


→ Volume



Visual Encoding

- Analyze charts as a **combination of marks and channels**



1 channel:
Vertical position

Mark: line



2 channels:
Vertical position
Horizontal position

Mark: point



3 channels:
Vertical position
Horizontal position
Color hue

Mark: point

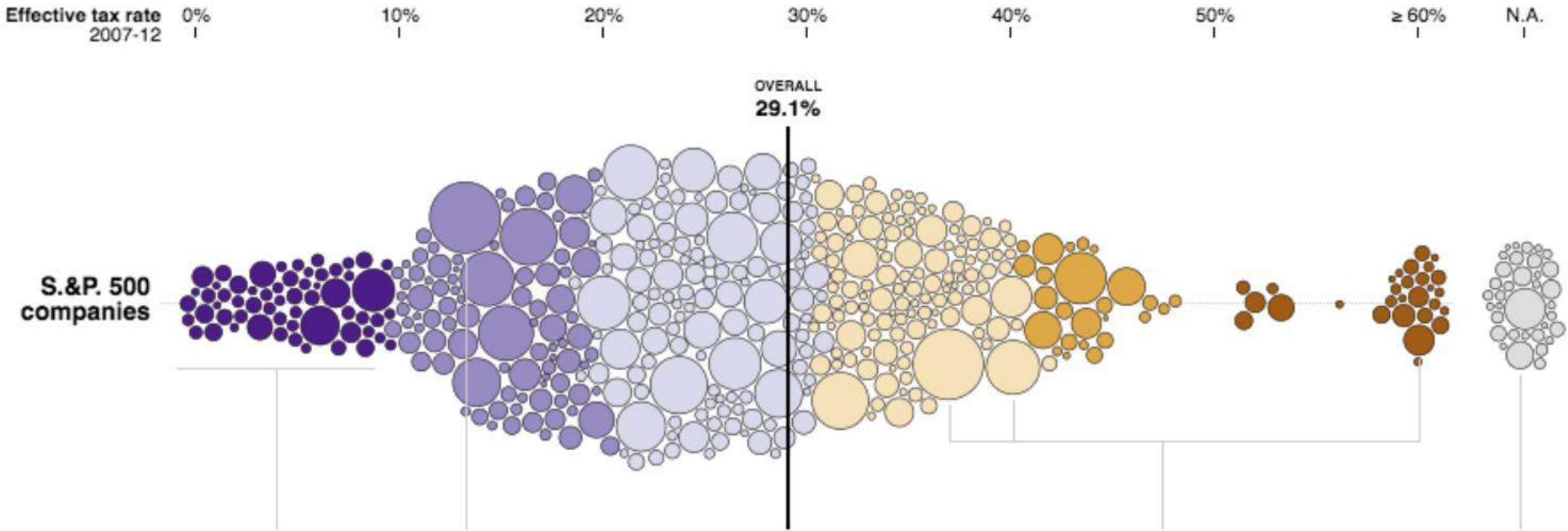


4 channels:
Vertical position
Horizontal position
Color hue
Size (area)

Mark: point

Exercise: Name the marks and channels

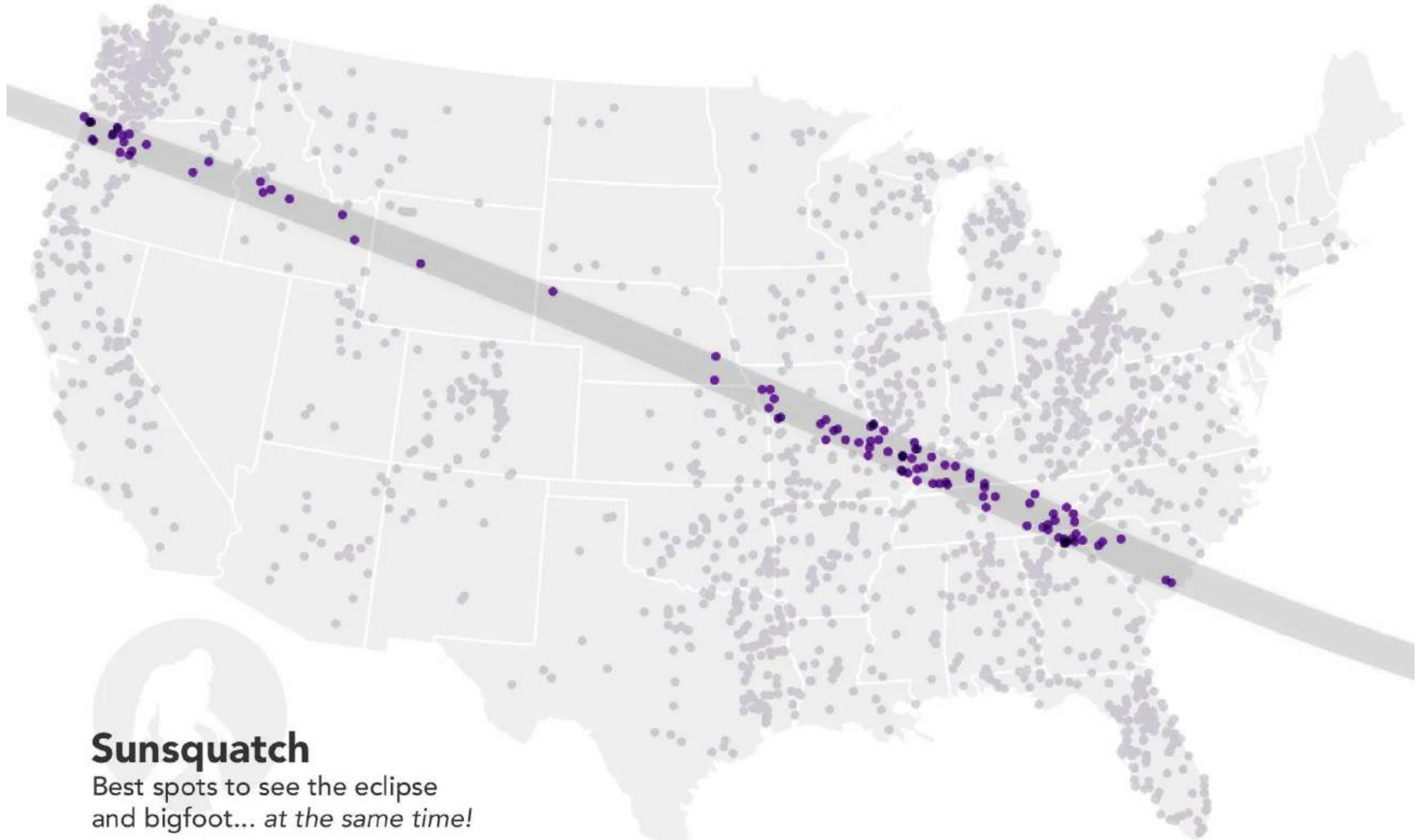
- Tax rates



<https://archive.nytimes.com/www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html>

Exercise: Name the marks and channels

- Sunsquatch



Sunsquatch

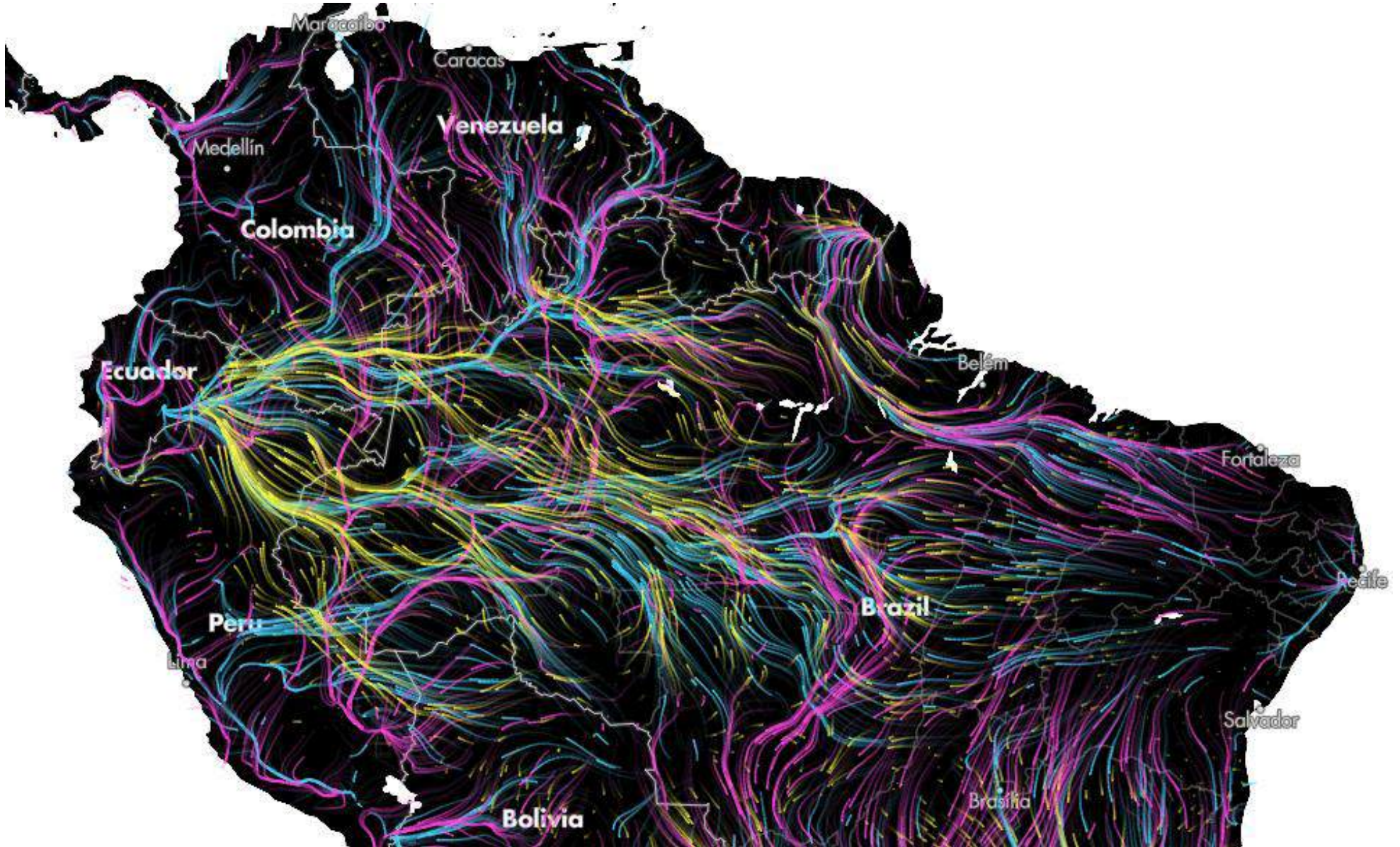
Best spots to see the eclipse
and bigfoot... *at the same time!*

Data: Bigfoot Field Researchers Organization | NASA Scientific Visualization Studio

<https://flowingdata.com/2017/08/20/sunsquatch-the-only-eclipse-map-you-need/>

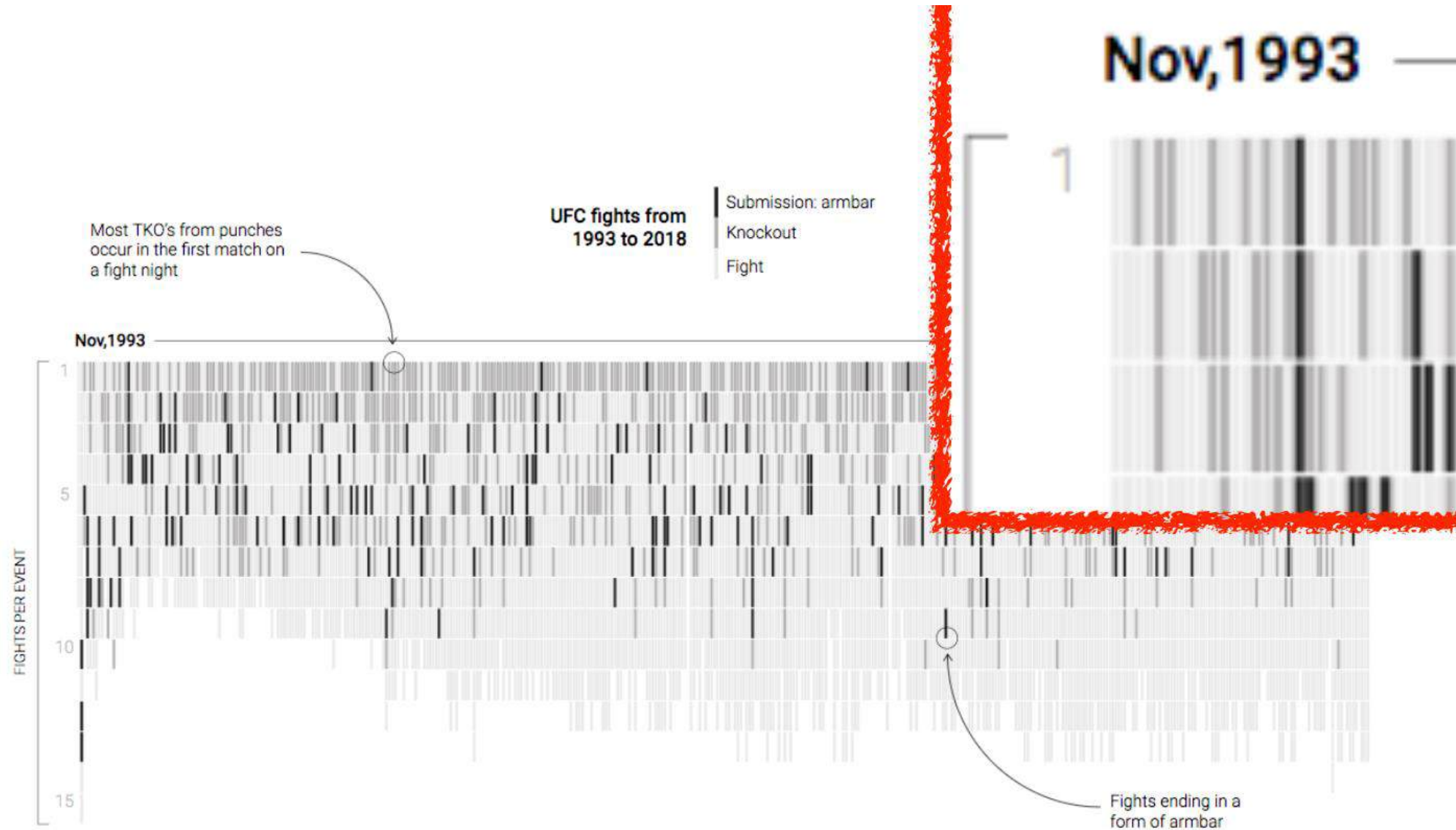
Exercise: Name the marks and channels

- Migration



Exercise: Name the marks and channels

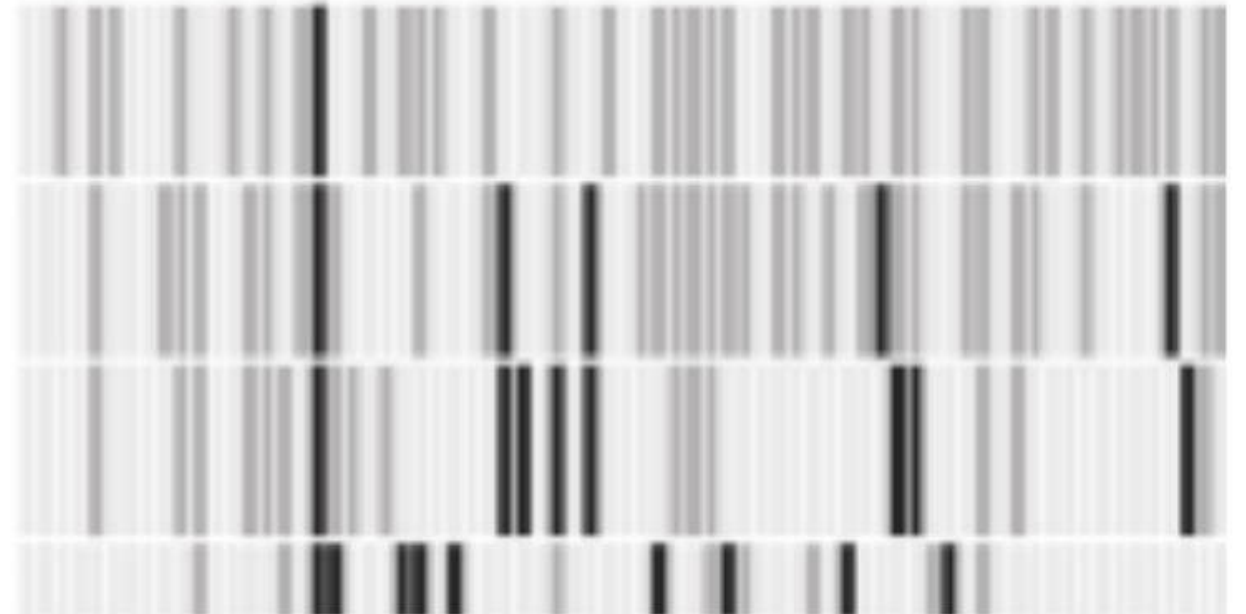
- UFC fights



Exercise: Name the marks and channels

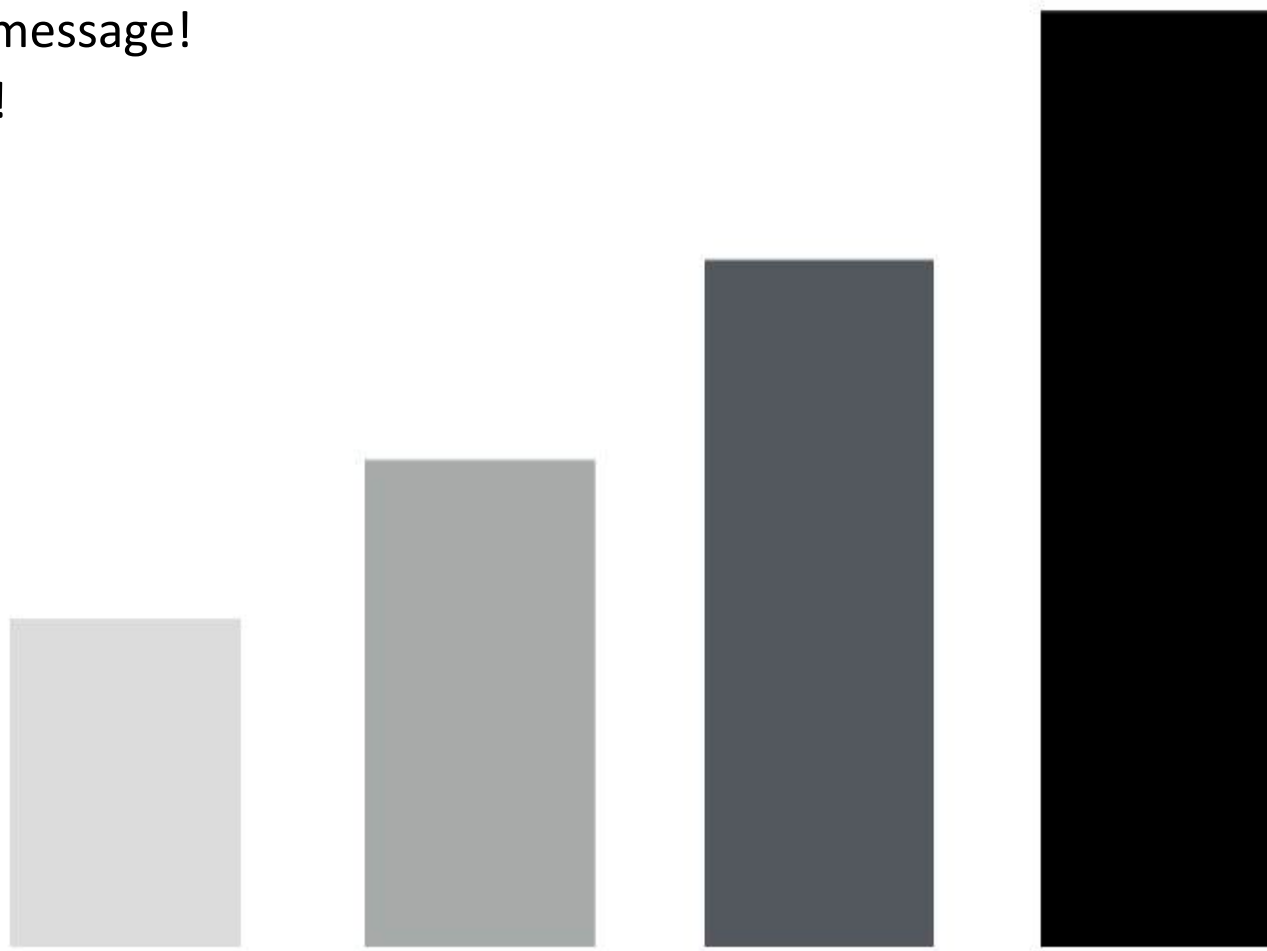
- UFC fights
 - Line?
 - No, not length encoded
 - Area?
 - No, area/shape does not convey any meaning
 - Point?
 - Yes! Just has a rectangular shape

Nov, 1993



Redundant Encoding

- Multiple channels can encode the same information
 - Sends a stronger message!
 - Uses up channels!



Length and Luminance

Marks as constraints

- Math view: marks are geometric primitives



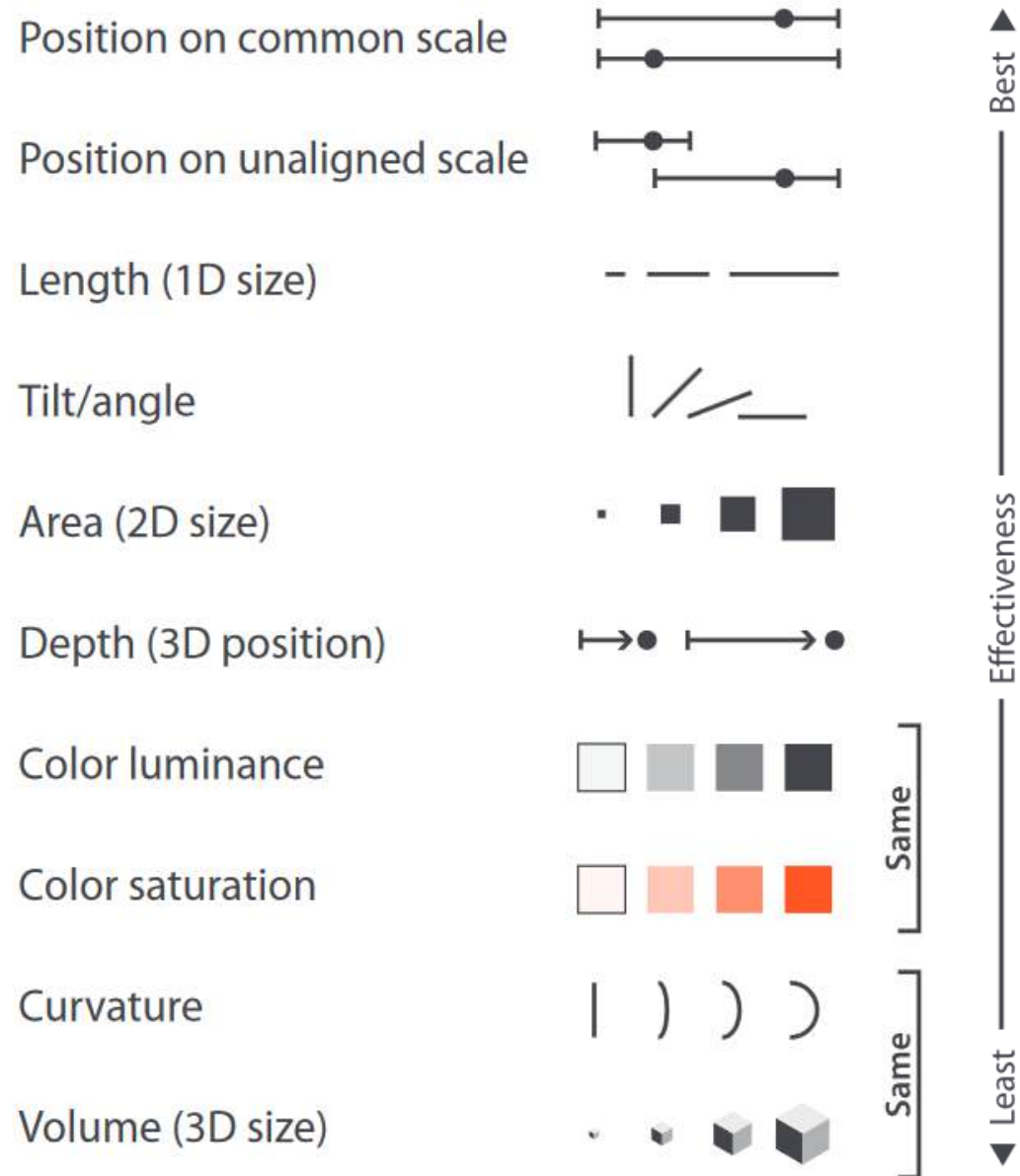
- Constraints view: Mark type constrains what else can be encoded
 - Points: 0 constraints on size, can encode more attributes with size & shape
 - Lines: 1 constraint on size (length), can encode something else another way (width)
 - Areas: 2 constraints on size (length & width), can't encode using size or shape
- Quick check: ask “Can I size-encode another attribute?”
 - If shape/size is already in use, the answer is “no”

When to use which channel?

- **Expressiveness:** Match channel type to data type
- **Effectiveness:** Some channels are better than others

Channel ranking

- **Magnitude channels:** have an ordering

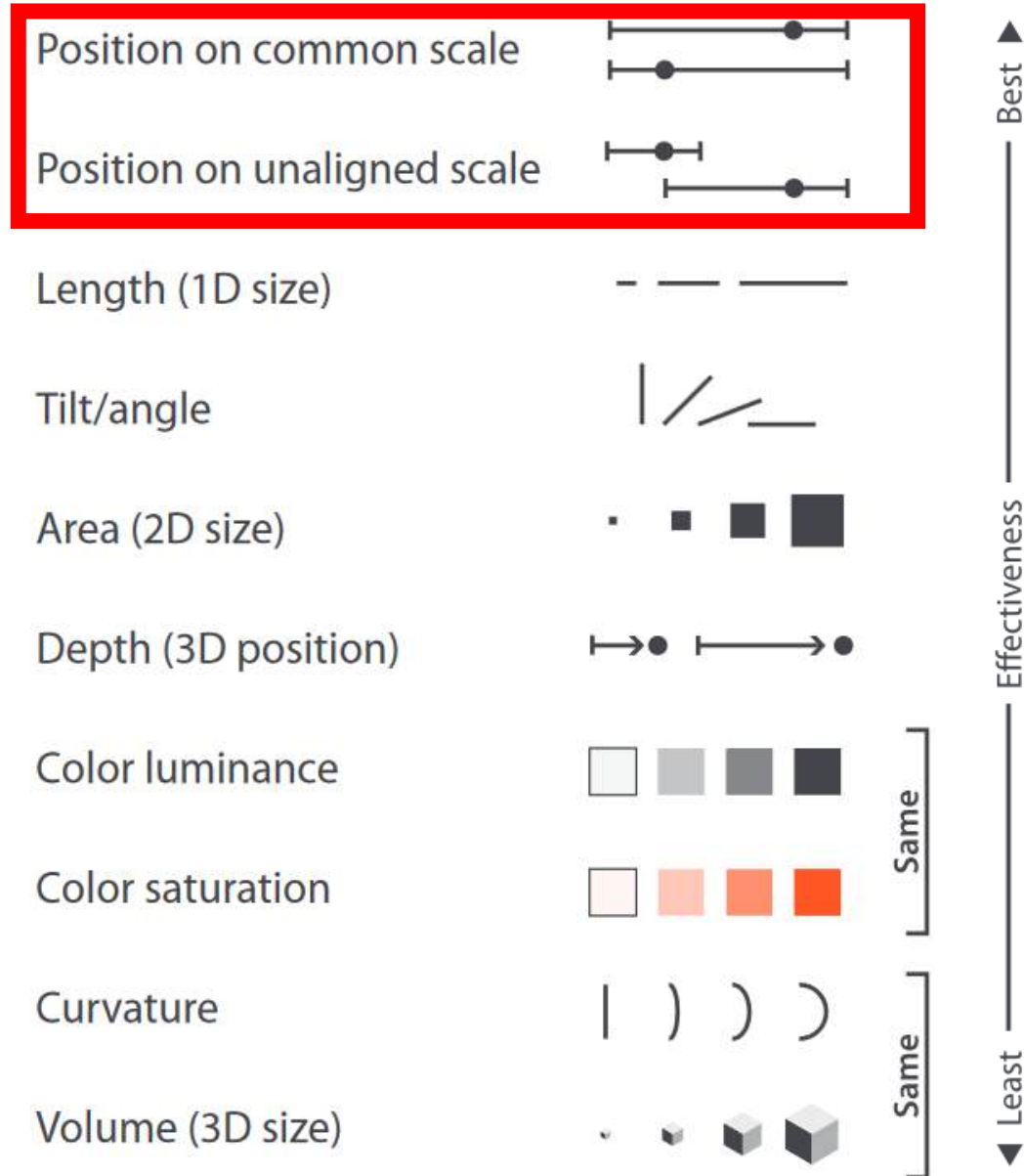


- **Identity channels:** no ordering

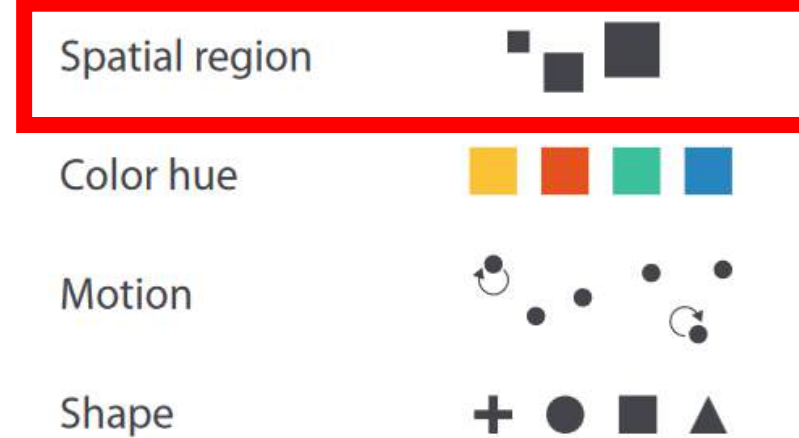


Channel ranking

- **Magnitude channels:** have an ordering



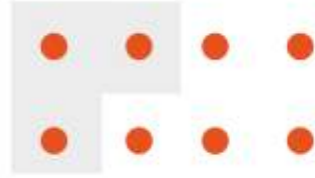
- **Identity channels:** no ordering



Grouping

- Marks as links:

- Containment
- Connection



- Identity channels:

- Proximity
- Similarity

Spatial region



Color hue



Motion



Shape



Channel Effectiveness

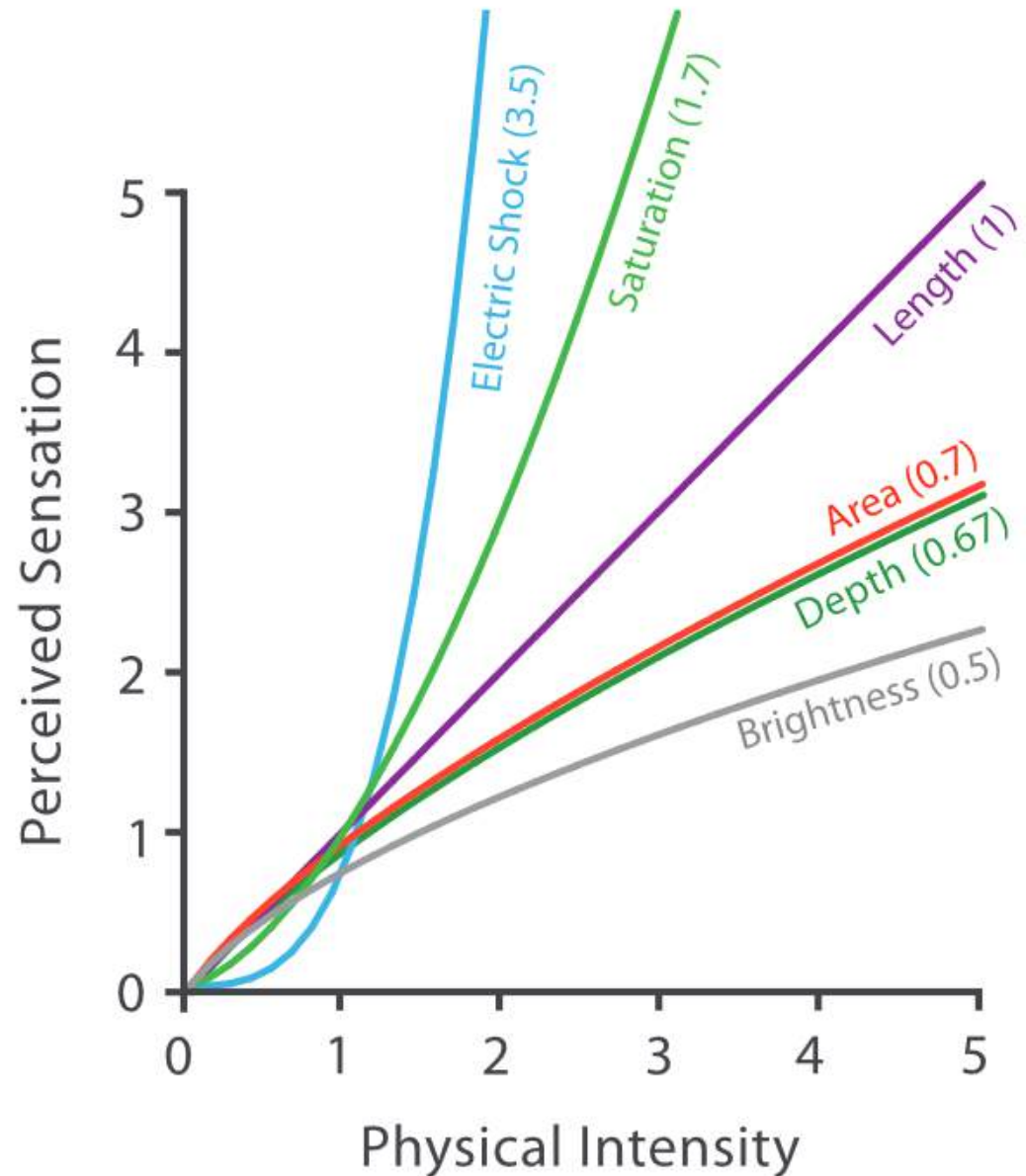
- Accuracy: how precisely can we tell the difference between encoded items?
- Discriminability: how many unique steps can we perceive?
- Separability: is our ability to use this channel affected by another one?
- Popout: can things jump out using this channel?

Accuracy: Fundamental Theory

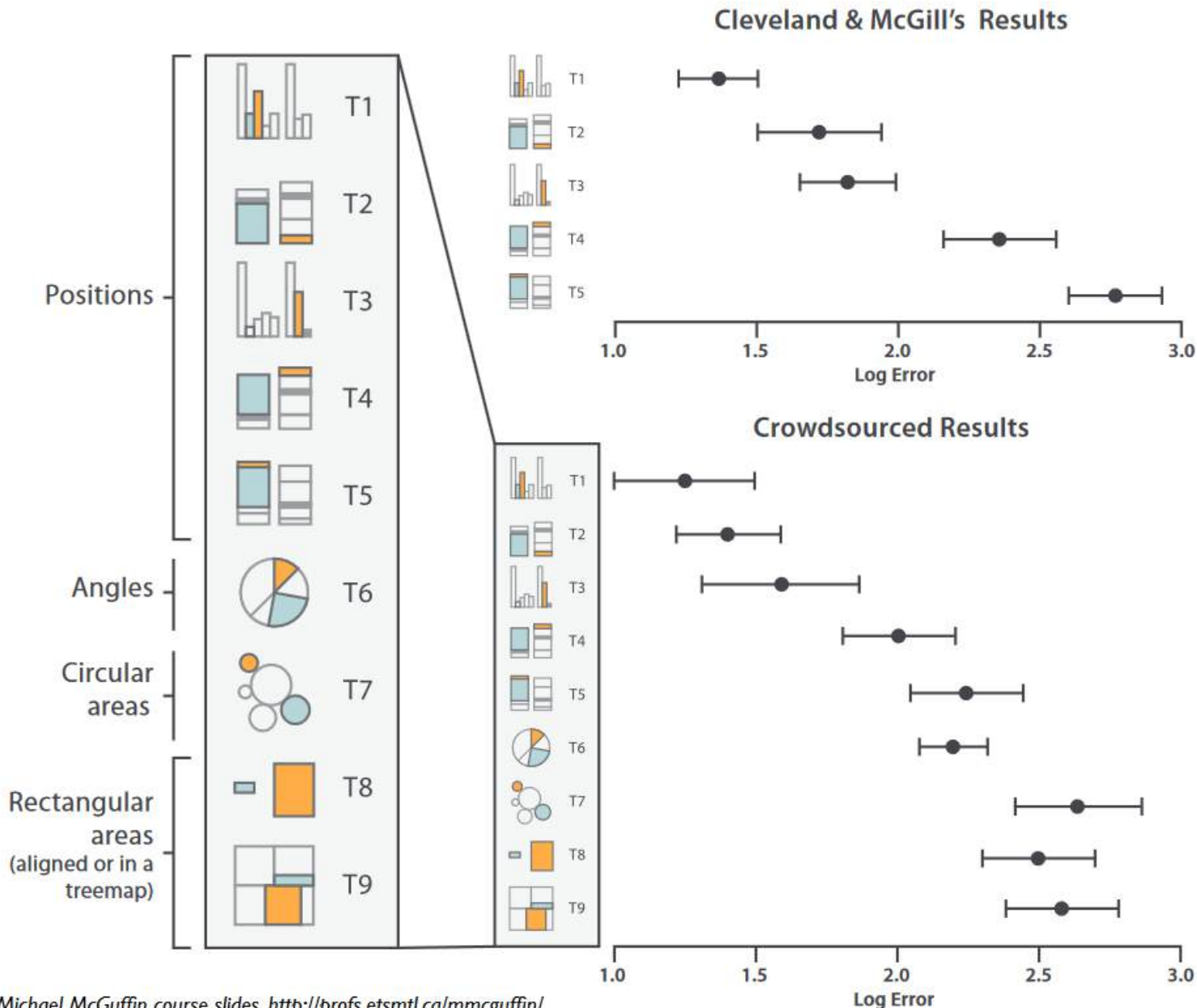
Steven's Psychophysical Power Law: $S = I^N$

S = sensation

I = intensity



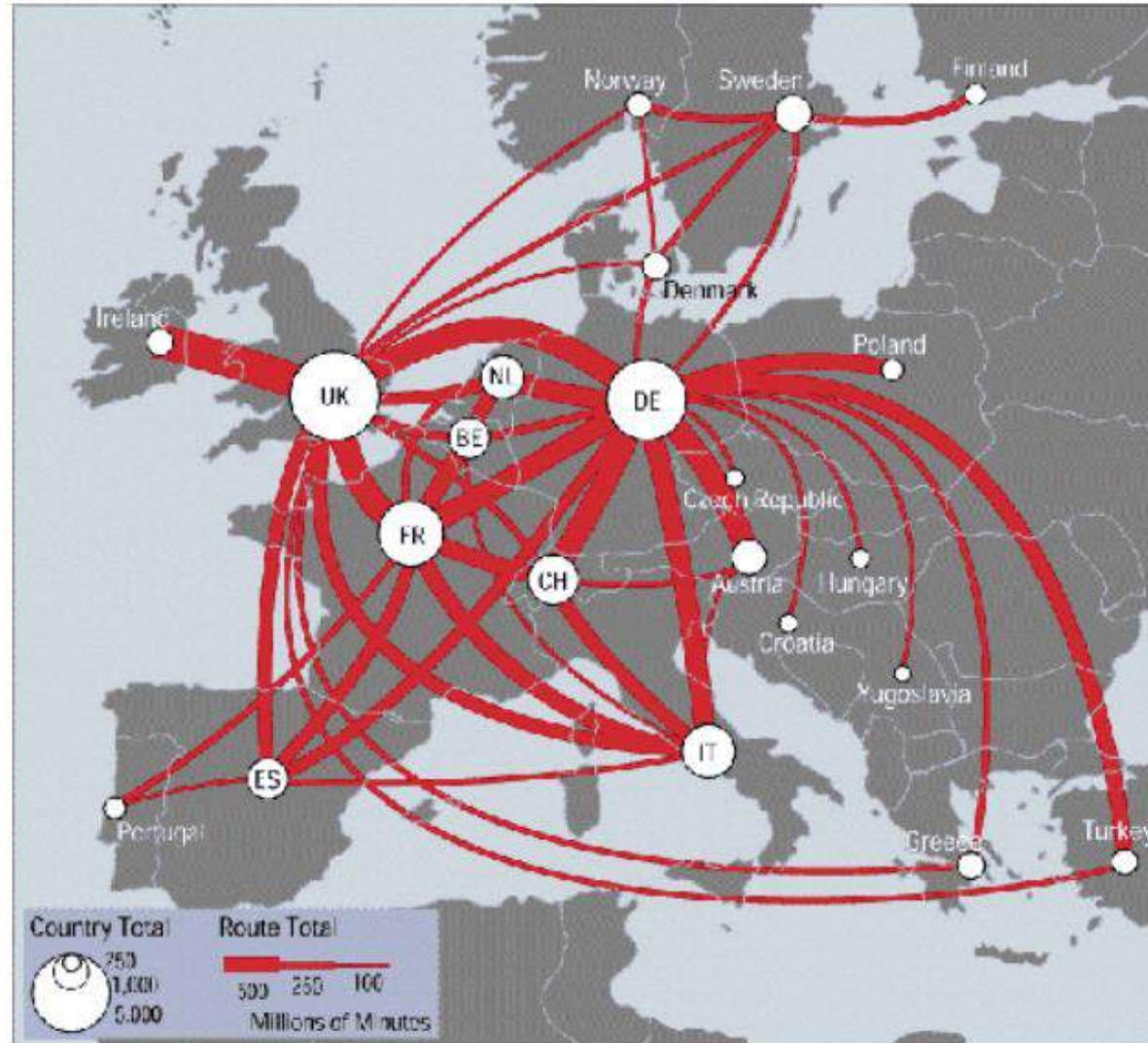
Accuracy: Viz experiments



Heer, Jeffrey, and Michael Bostock. "Crowdsourcing graphical perception: using mechanical turk to assess visualization design." *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010.

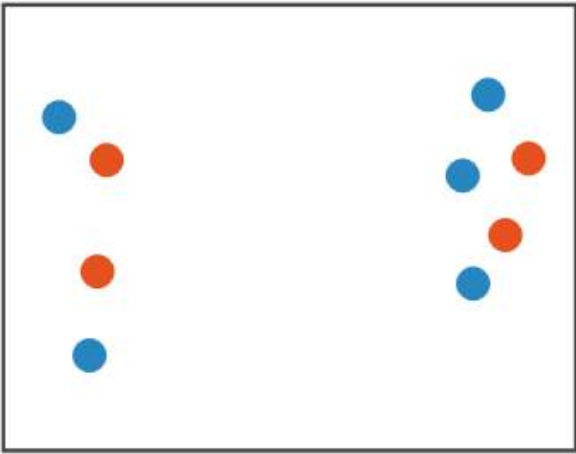
Discriminability: How many usable dimensions?

- Need a sufficient number of attribute levels to show details of the data
 - E.g.: Line width has only a few bins → can't make lines arbitrarily wide



Separability vs Integrality

Position
+ Hue (Color)



Fully separable

2 groups each

2 groups each

3 groups total:
Integral area

4 groups total:
Integral hue

Popout

45929078059772098775972655665110049836645
27107462144654207079014738109743897010971
43907097349266847858715819048630901889074
25747072354745666142018774072849875310665

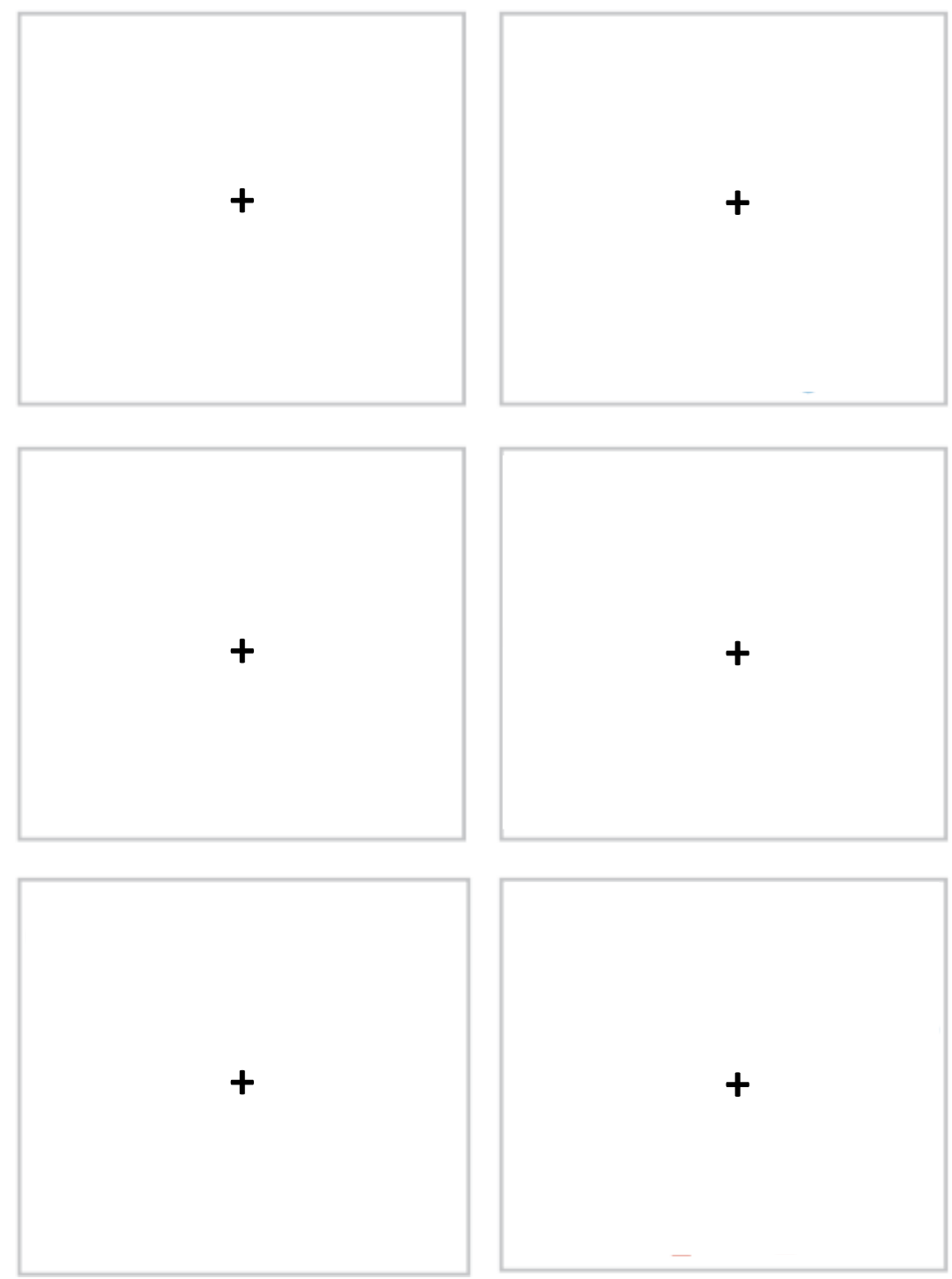
(a)

459290780597720987759726556651100498**3**6645
271074621446542070790147**3**8109743897010971
4**3**907097**3**49266847858715819048630901889074
25747072**3**54745666142018774072849875**3**10665

(b)

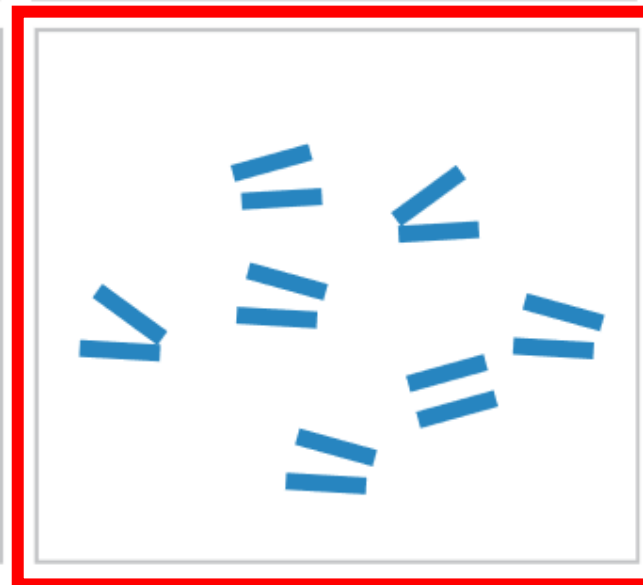
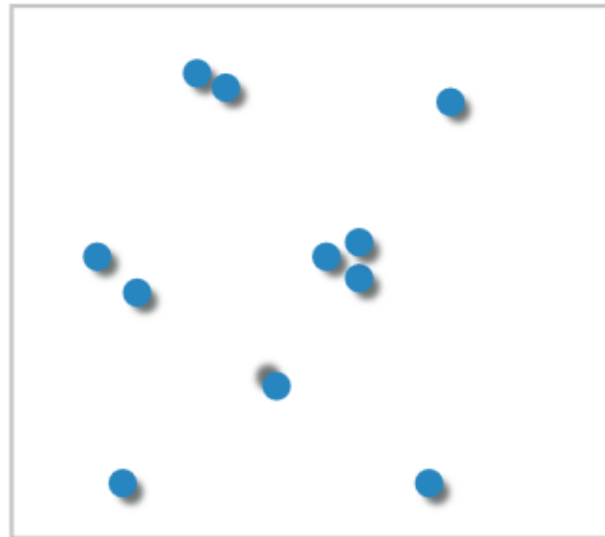
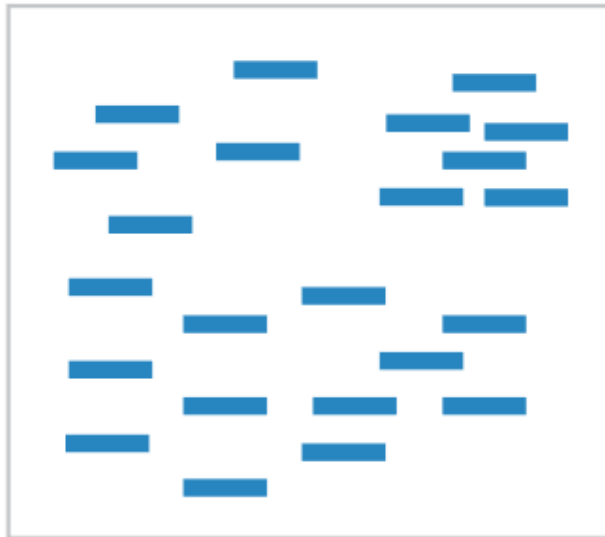
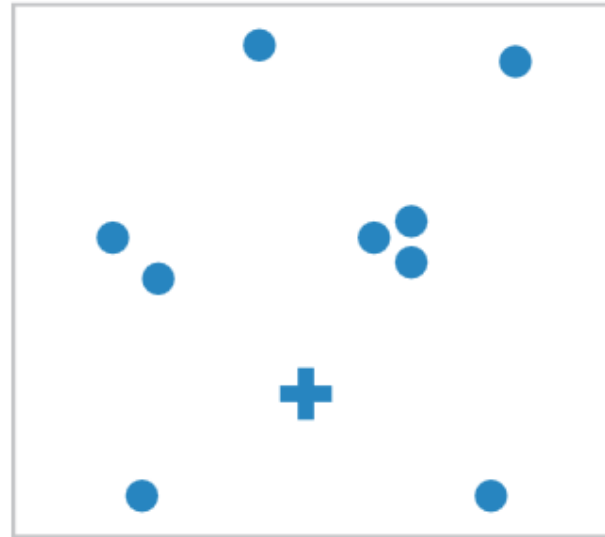
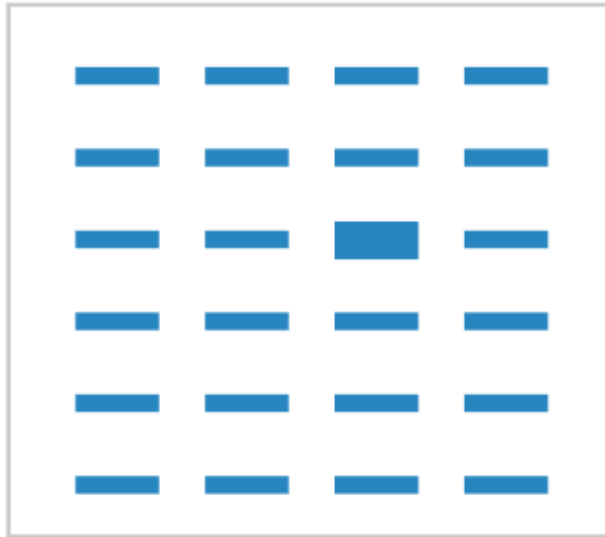
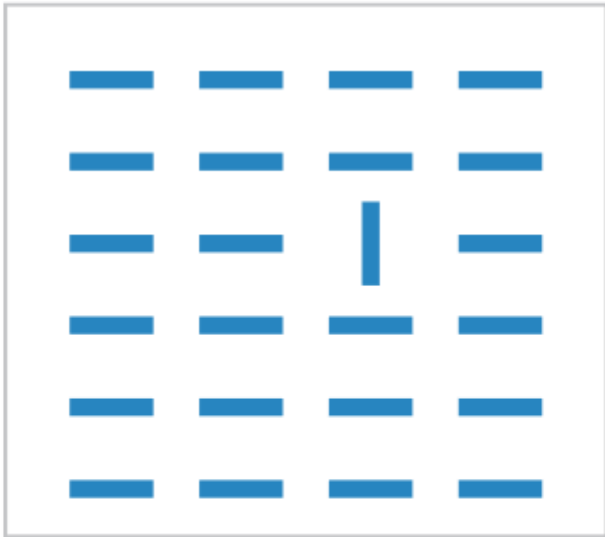
Popout

- Find the red dot
 - How long does it take?
- Parallel processing on individual channels
 - Popout → speed independent of distractor count
 - Speed depends on channel and similarity to distractors
- Serial search for combined channels
 - Speed depends on the number of distractors



Popout

- Works for many different channels



- Parallel lines don't pop out from tilted lines

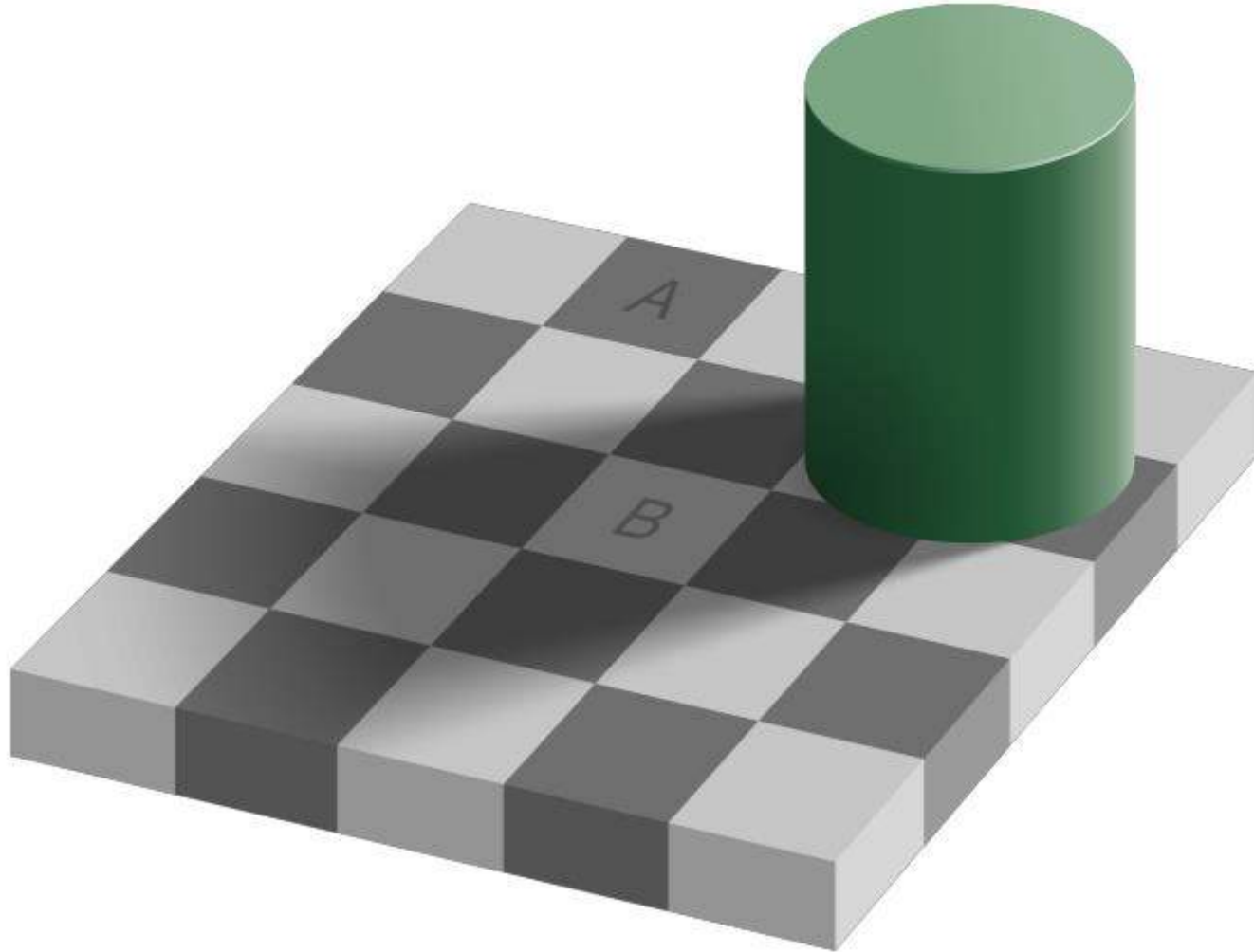
Factors affecting accuracy

- Alignment
- Distractors
- Distance



Relative vs absolute judgements

- Visual system mostly operates using relative judgements



Relative vs absolute judgements

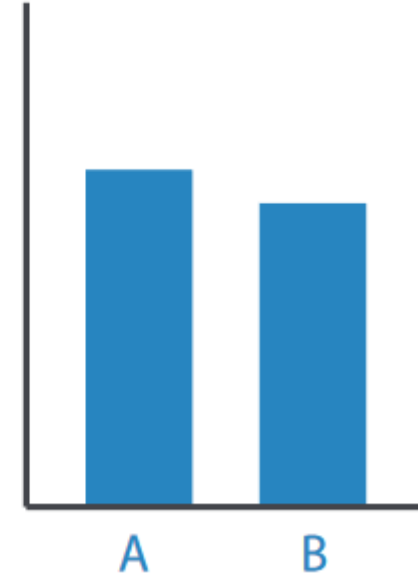
- Visual system mostly operates using relative judgements



Length



Position along
unaligned
common scale



Position along
aligned
common scale

Color

- Focus so far has been on spatial arrangement

Encode

➔ Arrange

➔ Express



➔ Separate



➔ Order



➔ Align



▸ Use



➔ Map

from **categorical** and **ordered** attributes

➔ Color

➔ Hue



➔ Saturation



➔ Luminance



➔ Size, Angle, Curvature, ...



➔ Shape



➔ Motion

Direction, Rate, Frequency, ...



Decomposing color

- Do not just talk about color!
 - It's confusing if treated as monolithic
- Recall the three channels of color:
 - Ordered (can show magnitude)
 - **Luminance**: how bright (black/white)
 - **Saturation**: how colorful
 - Categorical (can show identity)
 - **Hue**: what color
- Color channels have different properties
 - What they convey directly to the visual system
 - How much they can convey
 - How many discriminable bins can we use

Luminance



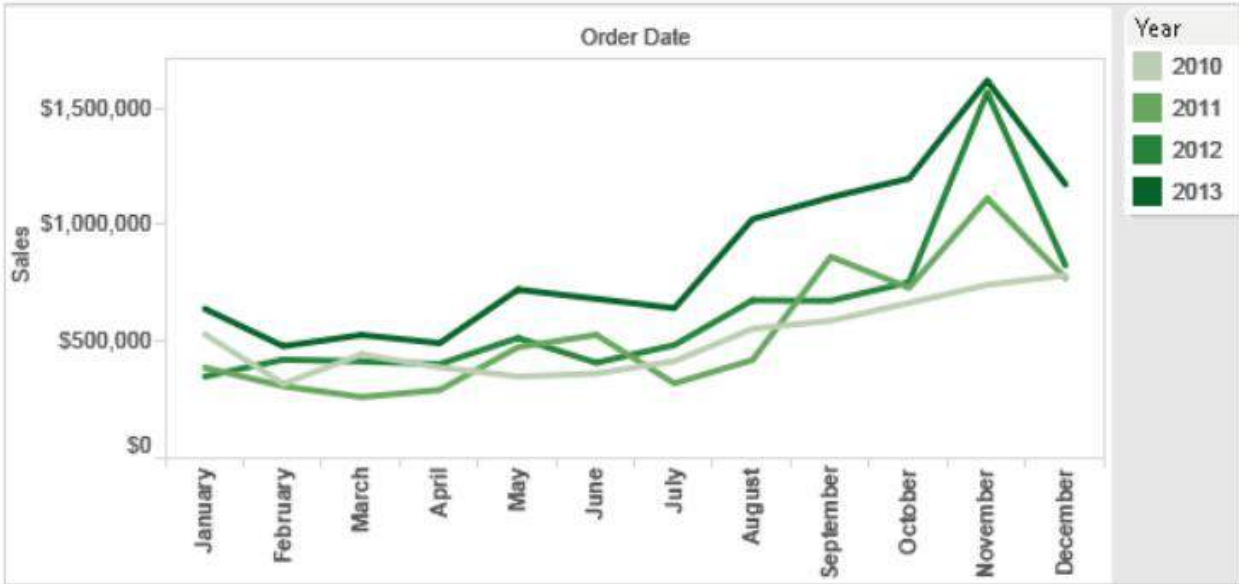
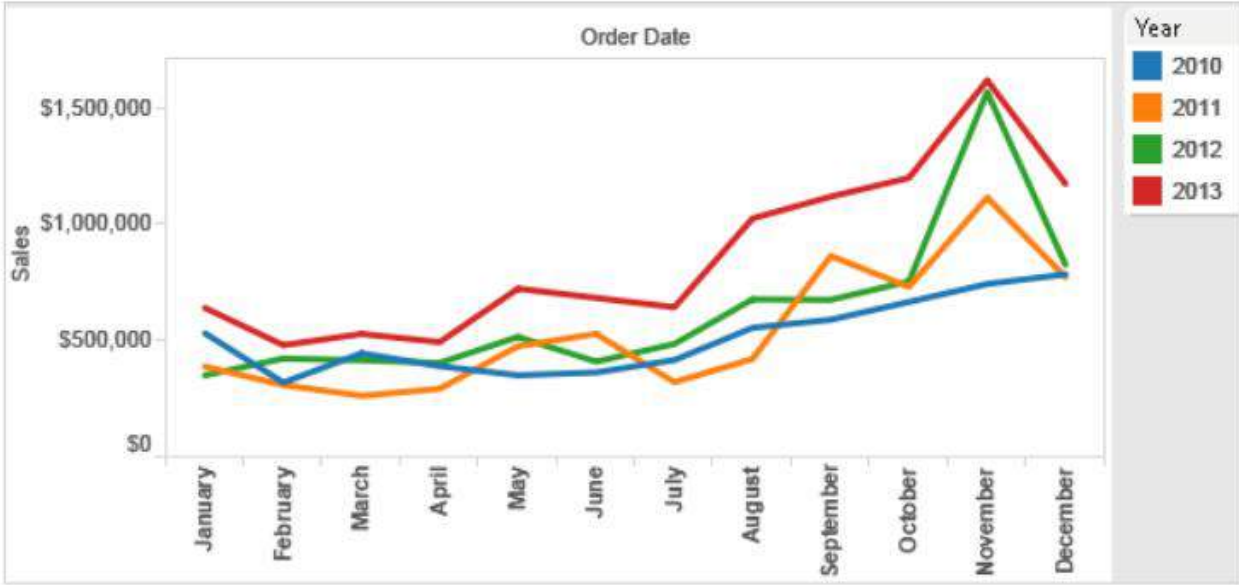
Saturation



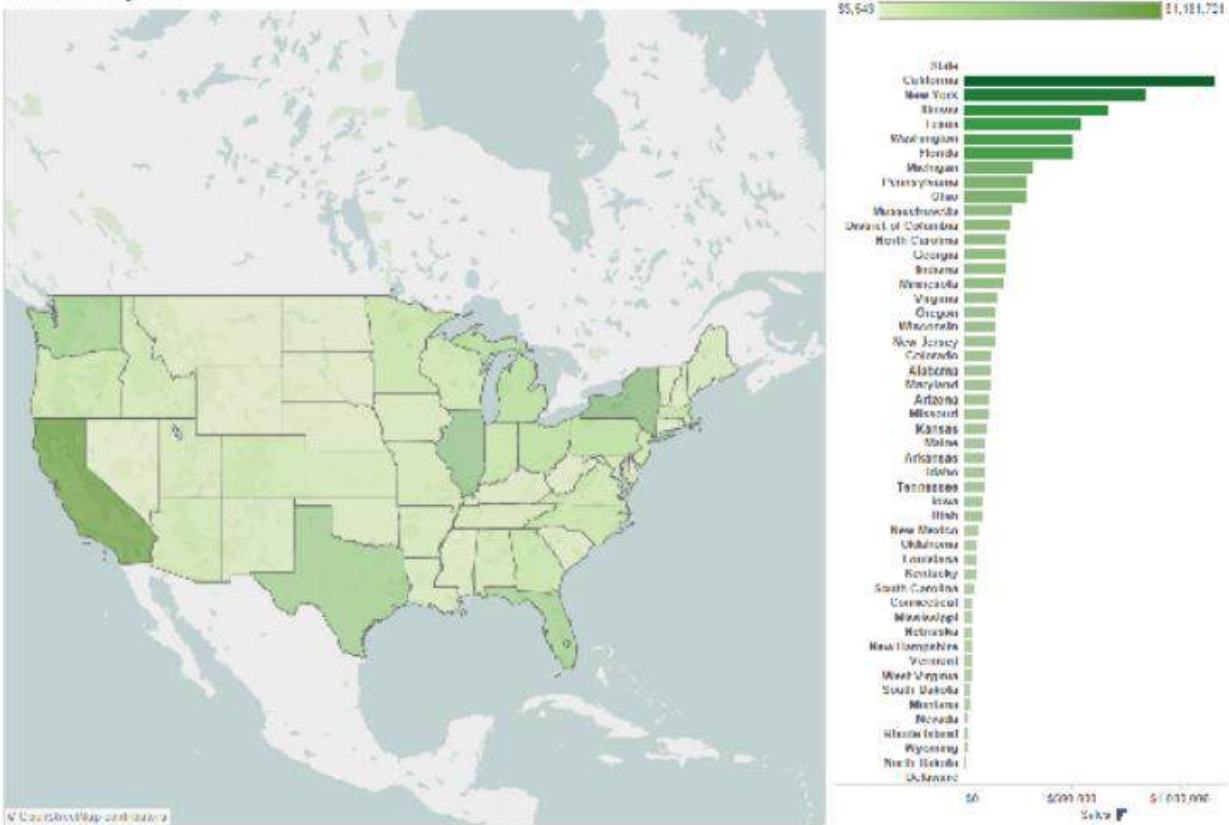
Hue



Categorical vs ordered color

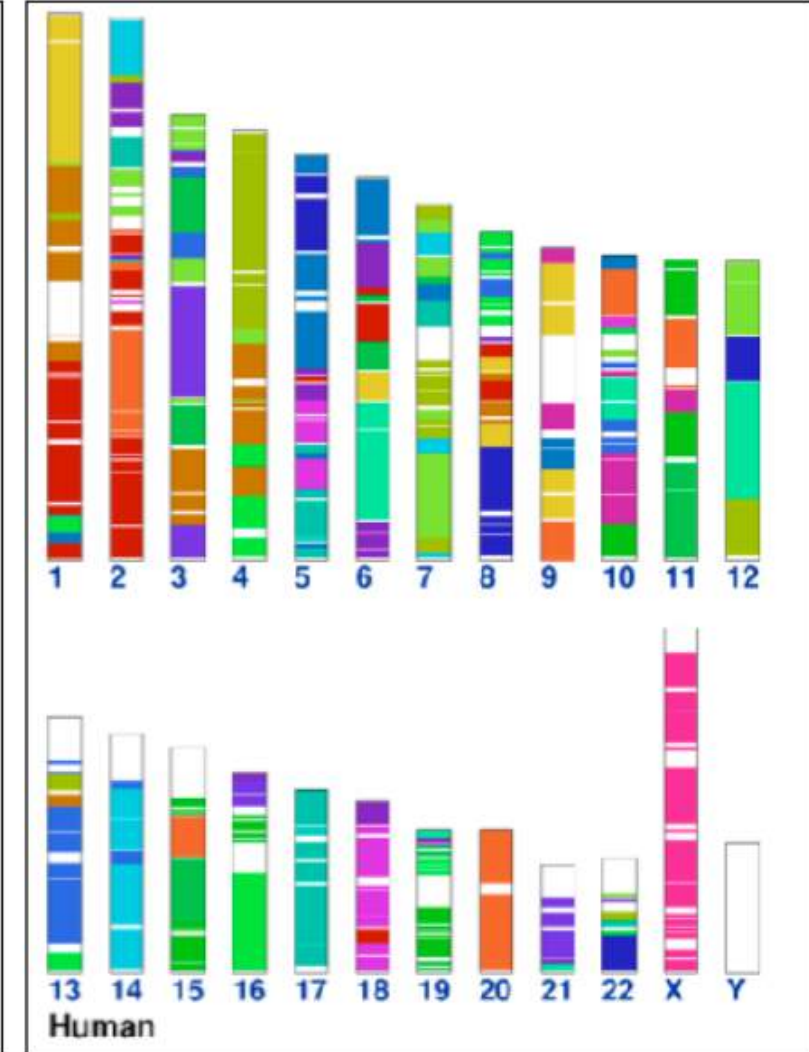
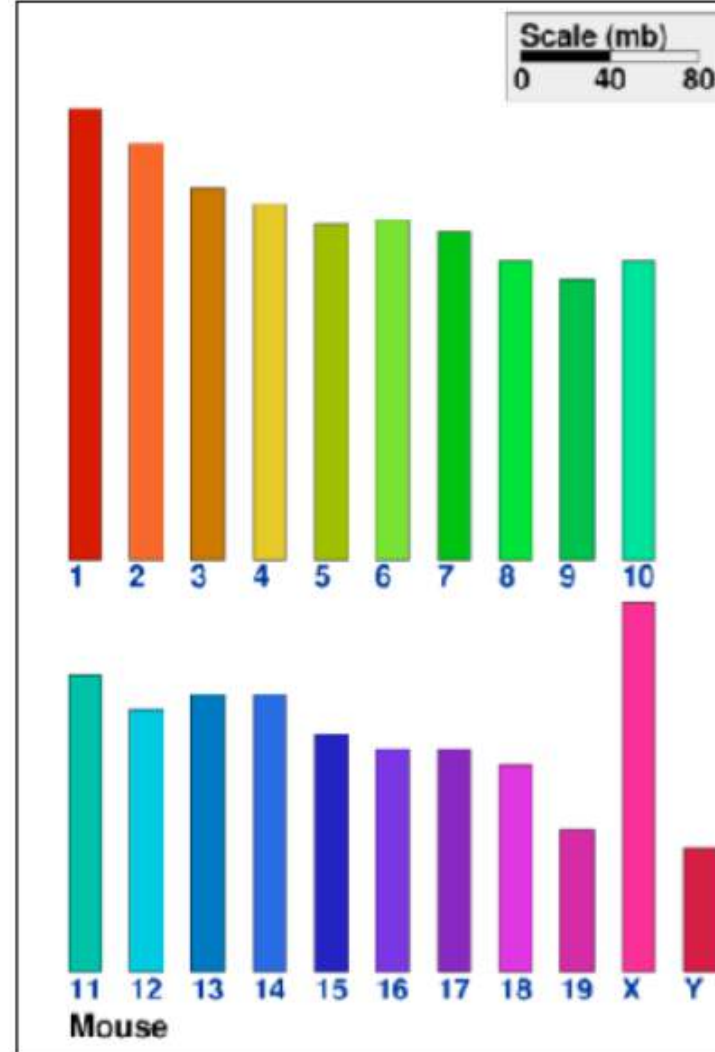


Annual sales by state



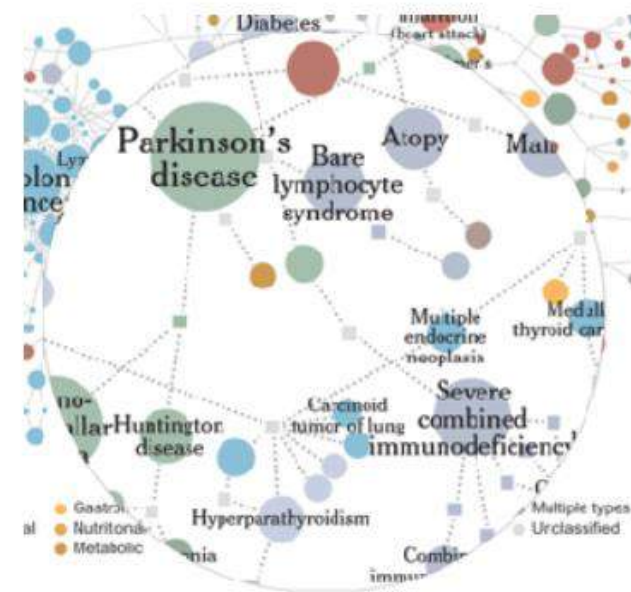
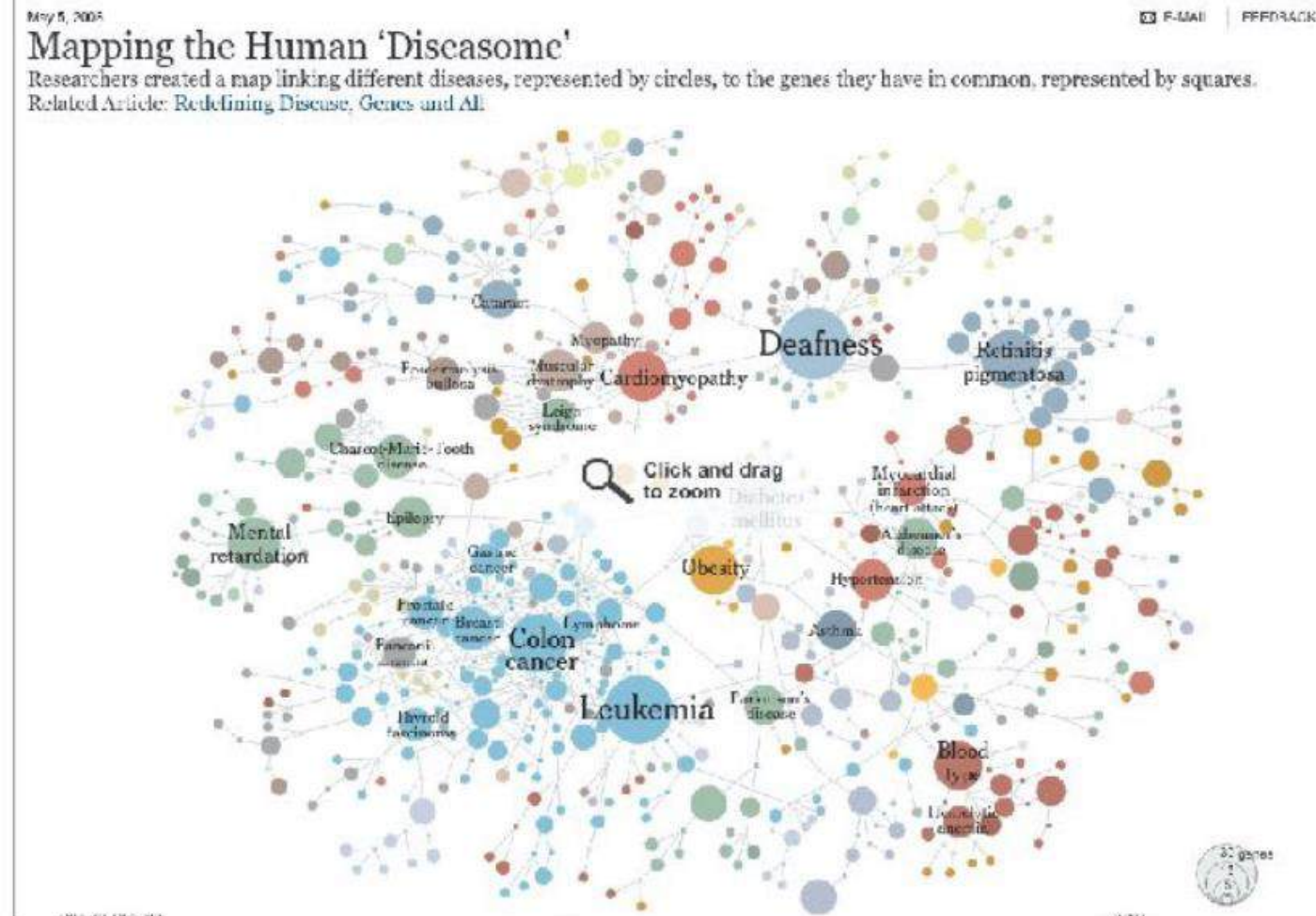
Categorical color: limited # of discriminable bins

- Human perception uses *relative* comparisons
 - Good if colors appear contiguously
 - Bad for absolute comparisons
- Noncontiguous small regions of color:
 - **Rule of thumb:** Can only use 6 – 12 different colors (bins), **including** background and highlights
 - Advice: deliberately, manually bin categories first, then map them to colors

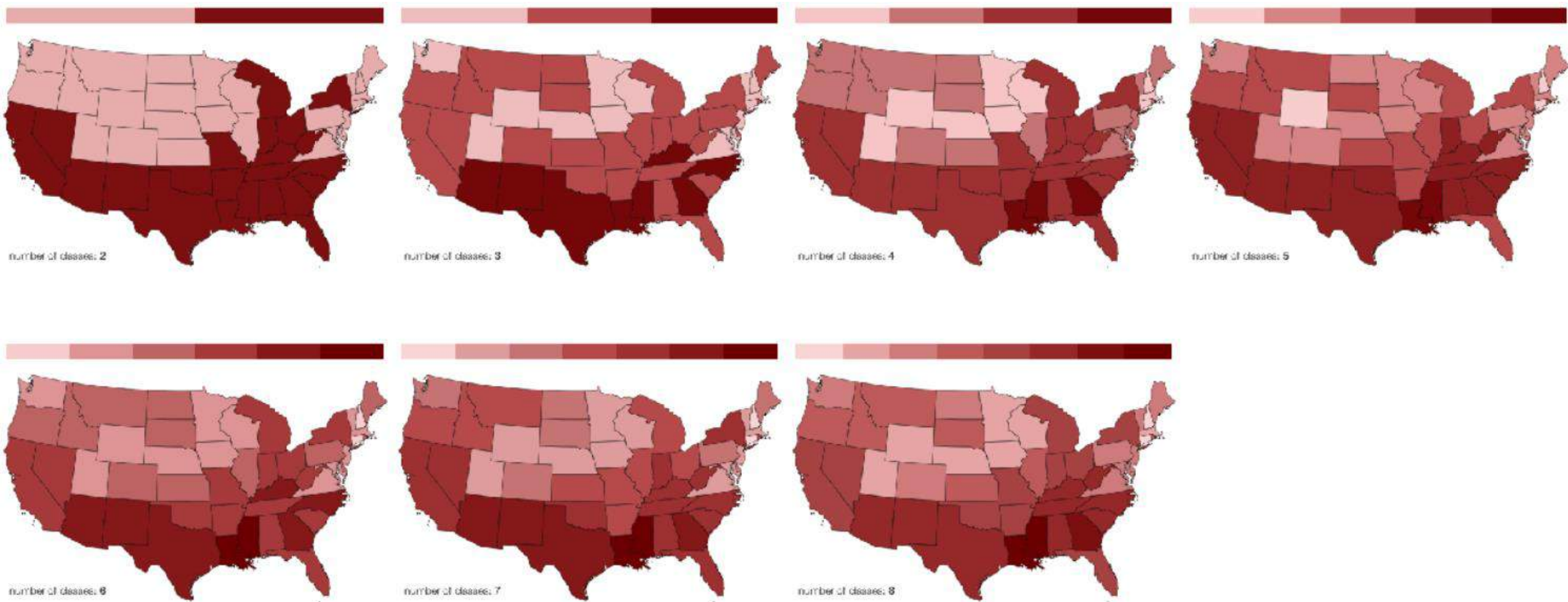


Sinha, Amit U., and Jaroslaw Meller. "Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms." *BMC bioinformatics* 8 (2007): 1-9.

Categorical color: limited # of discriminable bins

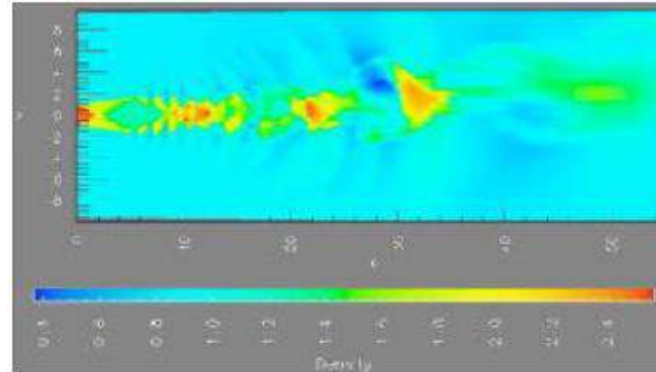
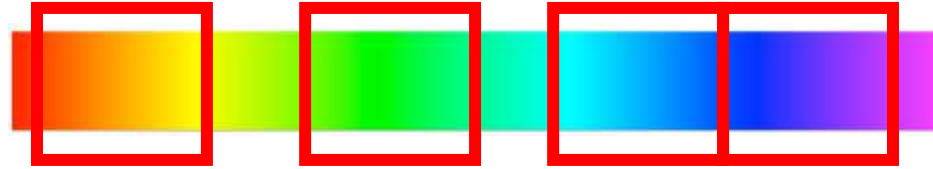


Ordered color: limited # of discriminable bins

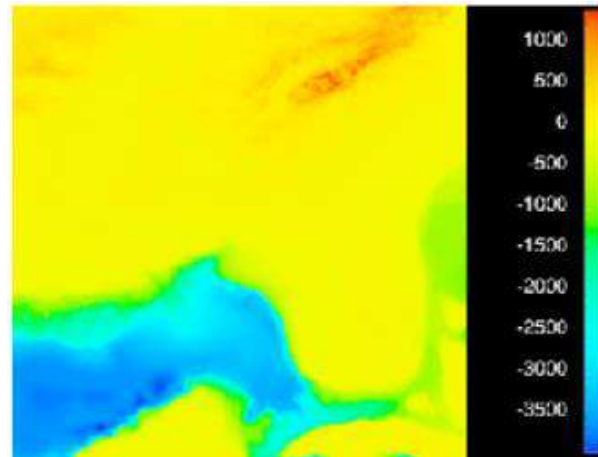


Ordered color: rainbow color map is often bad

- Problems:
 - Perceptually unordered
 - Perceptually nonlinear
- Not all bad:
 - Fine-grained structure is easily visible
- Alternatives:
 - Large-scale structure: use fewer hues
 - Fine structure: multiple hues with monotonically increasing luminance (e.g. viridis color map)



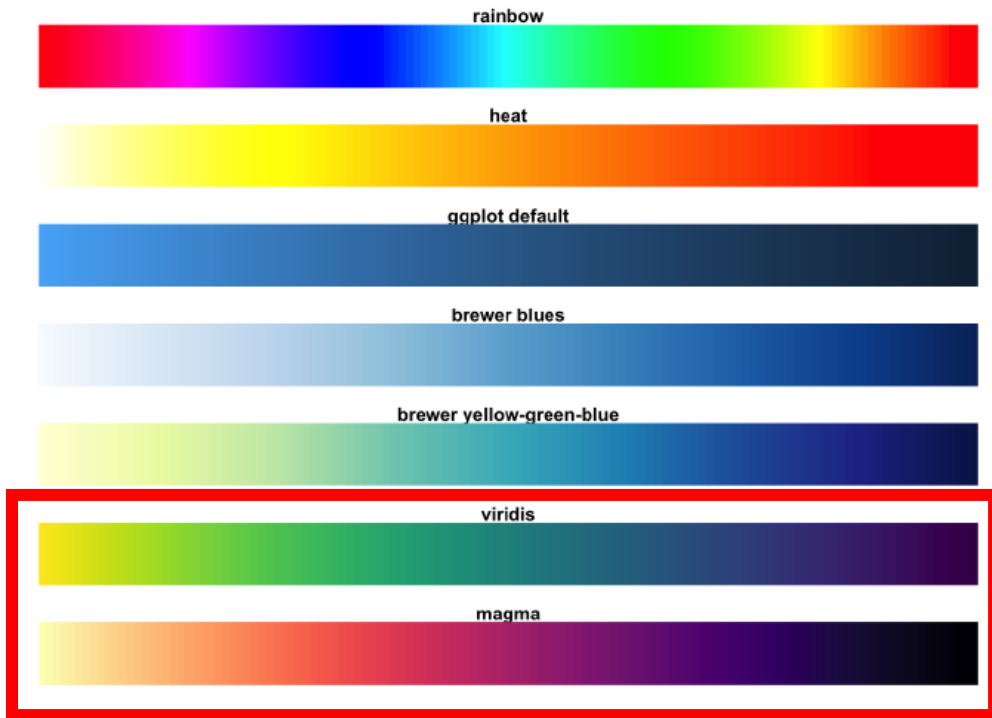
Bergman, Lawrence D., Bernice E. Rogowitz, and Lloyd A. Treinish. "A rule-based tool for assisting colormap selection." *Proceedings Visualization'95*. IEEE, 1995.



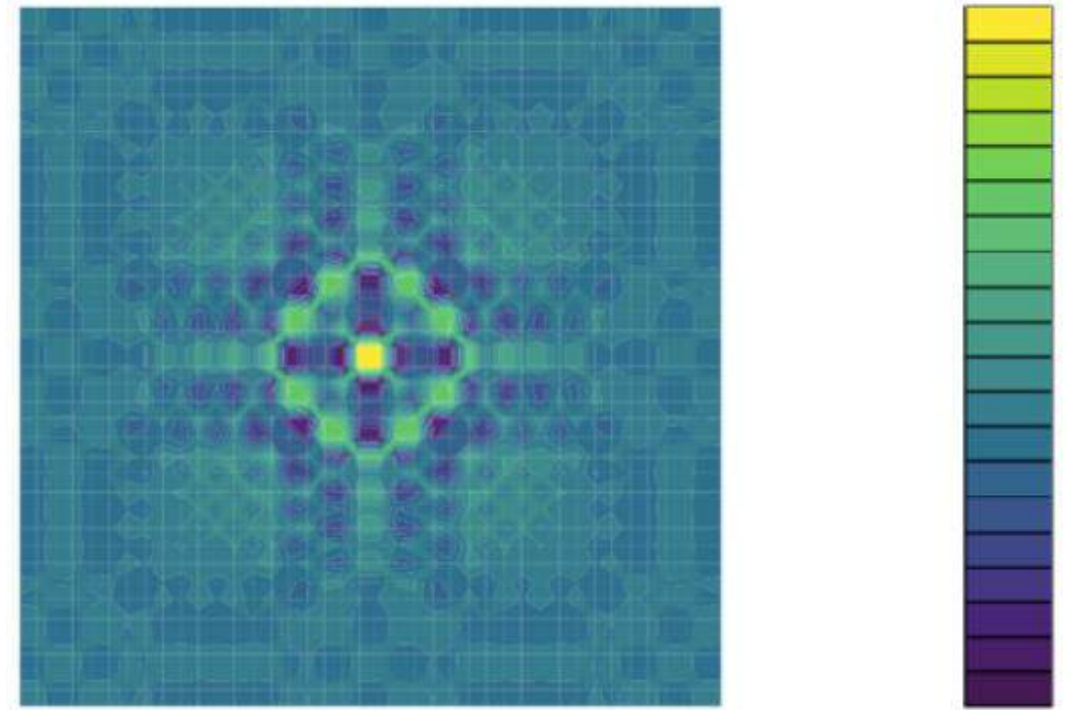
Rogowitz, Bernice E., and Lloyd A. Treinish. "Why should engineers and scientists be worried about color." See URL <http://www.research.ibm.com/people/l/lloyd/color/color.HTM> (1996).

Viridis / Magma: sequential colormaps

- Monotonically increasing luminance, perceptually uniform
- Colorblind-safe, too

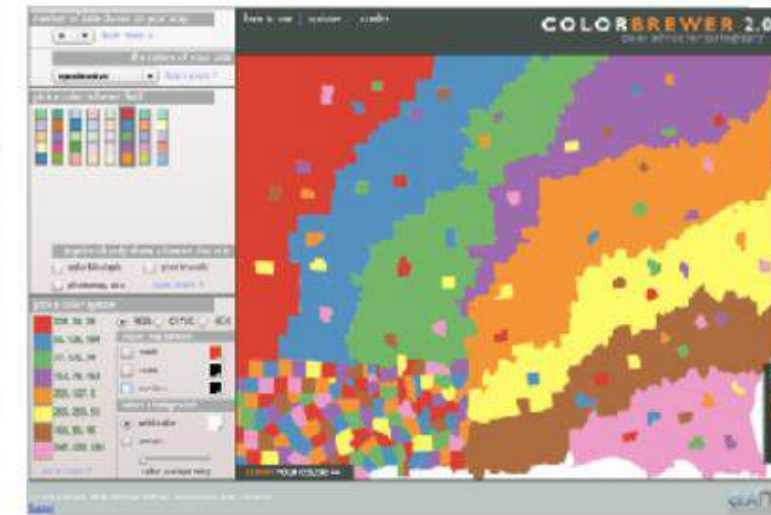


<https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>

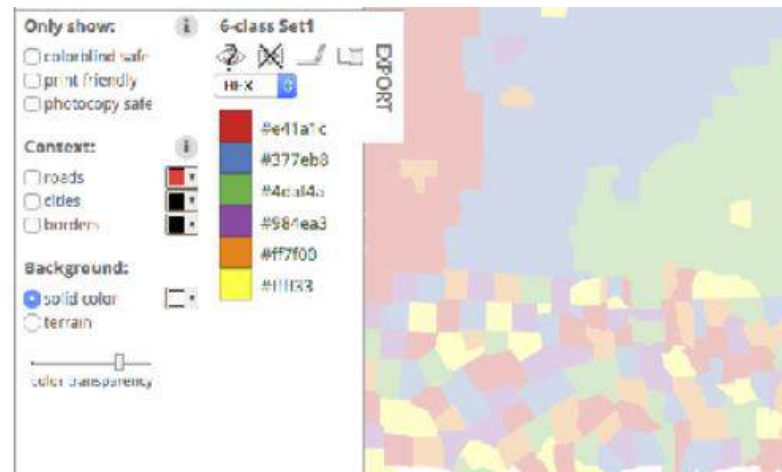


Interaction between channels: Not fully separable

- Color channel interactions
 - Size heavily influences salience
 - **Rule of thumb:** small regions need high saturation, large regions need low saturation
- Saturation & luminance
 - Not separable from each other!
 - Also not separable from transparency
 - **Rule of thumb:** Small separated regions → use 2 bins (3-4 max)
 - **Rule of thumb:** Only use one of saturation, luminance, or transparency

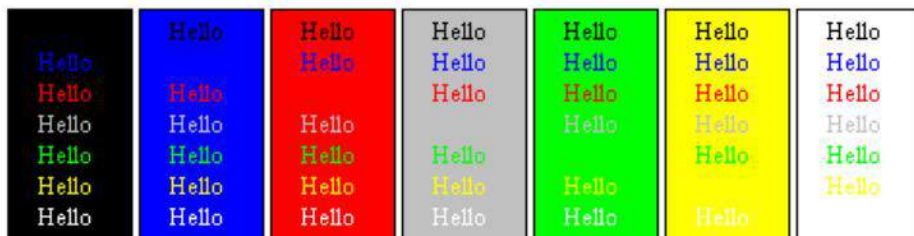


<https://colorbrewer2.org>



Color deficiency: Luminance

- We need luminance for edge detection
 - Fine-grained details only visible via luminance contrast
 - Legible text requires luminance contrast!



Luminance information

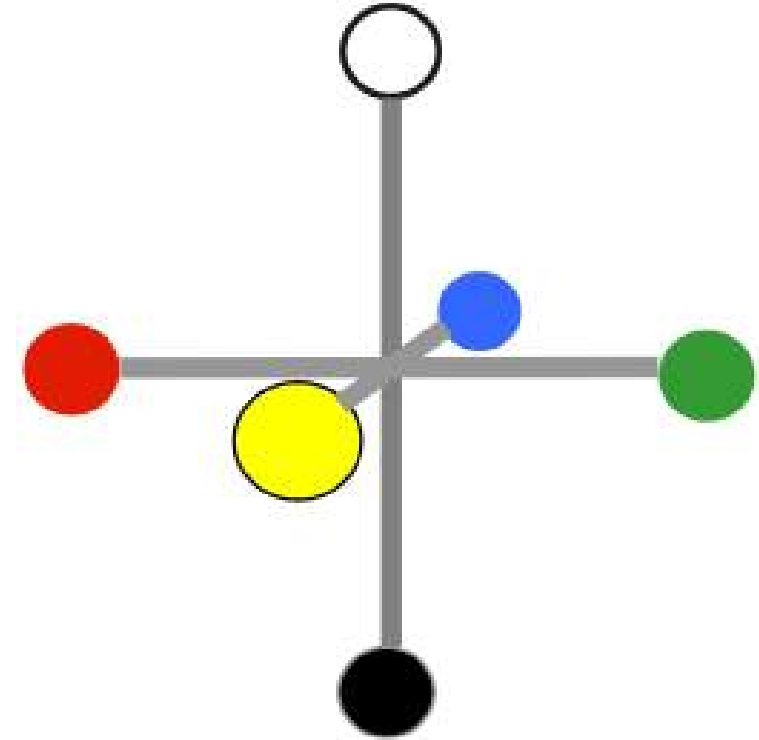


Saturation/hue information



Color deficiency

- “Colorblind” → degraded acuity along one axis
 - 8% of men are red/green color deficient
 - Blue/yellow deficiency is rare



Color deficiency: Check with simulator



**Normal
vision**



**Deuteranope
*green-weak***



**Protanope
*red-weak***



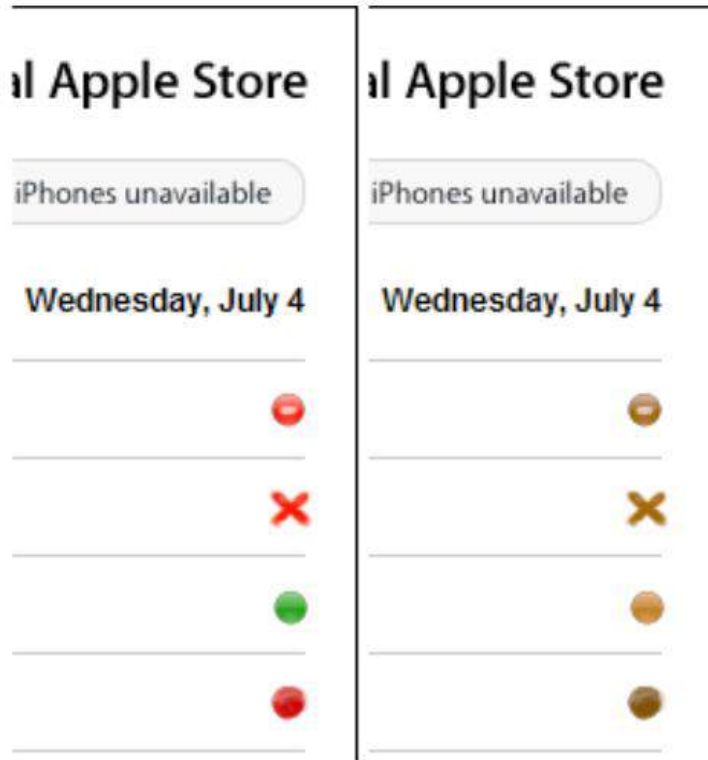
**Tritanope
*blue-weak***



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

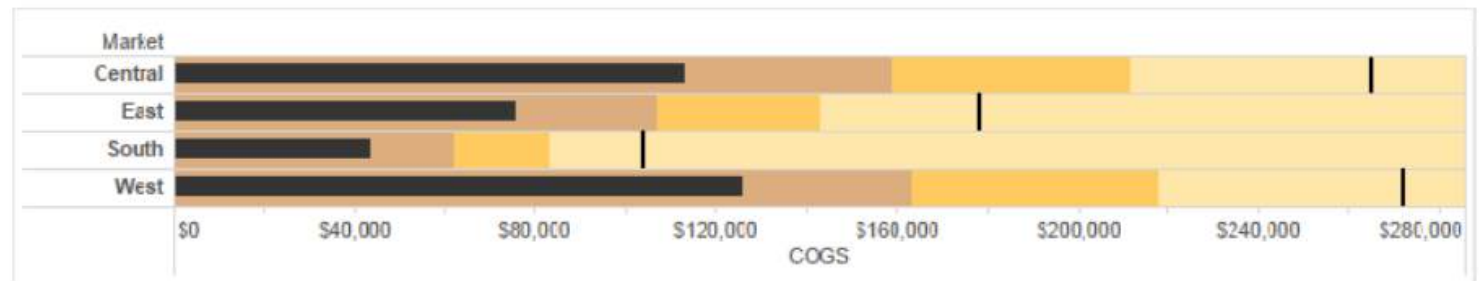
Color deficiency: avoid encoding by hue alone

- Redundantly encode information!
 - Vary luminance
 - Change shape



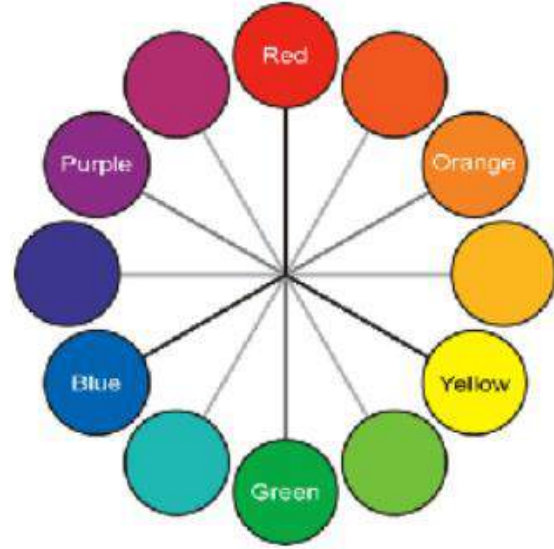
} Change the shape

} Vary the luminance

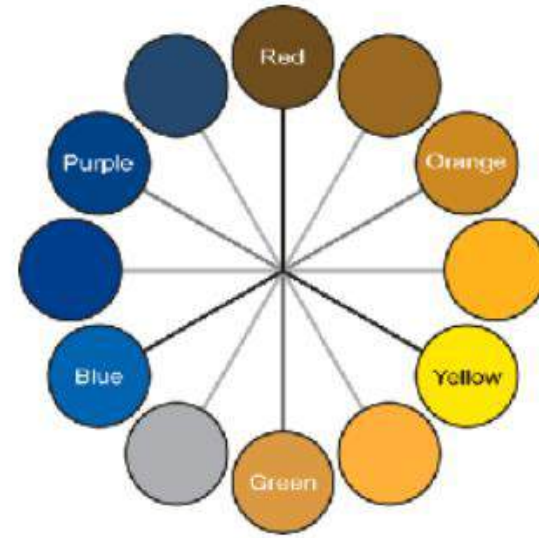


Deuteranope simulation

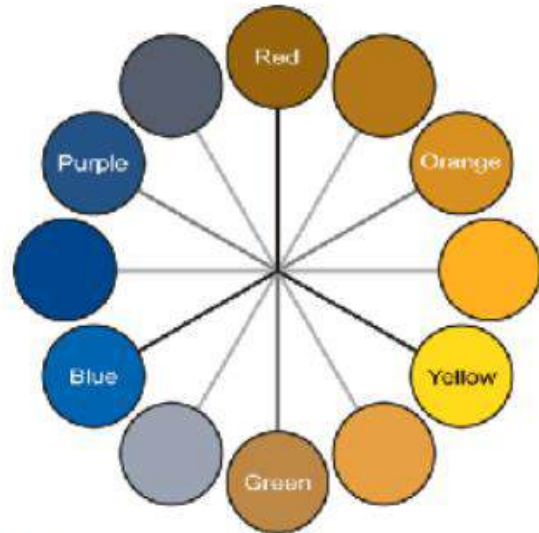
Color deficiency: Reduces color to 2 dimensions



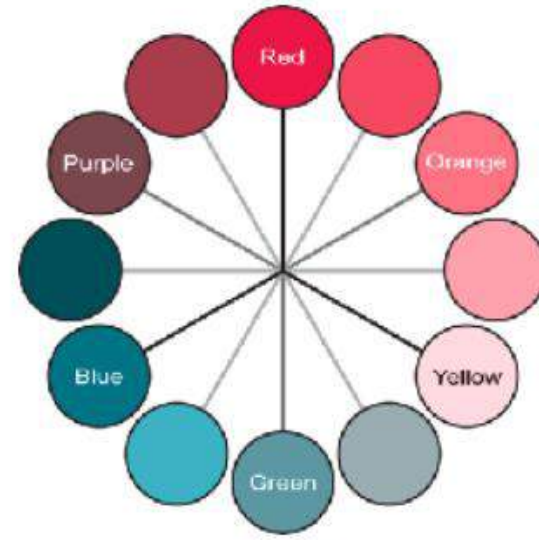
Normal



Protanope

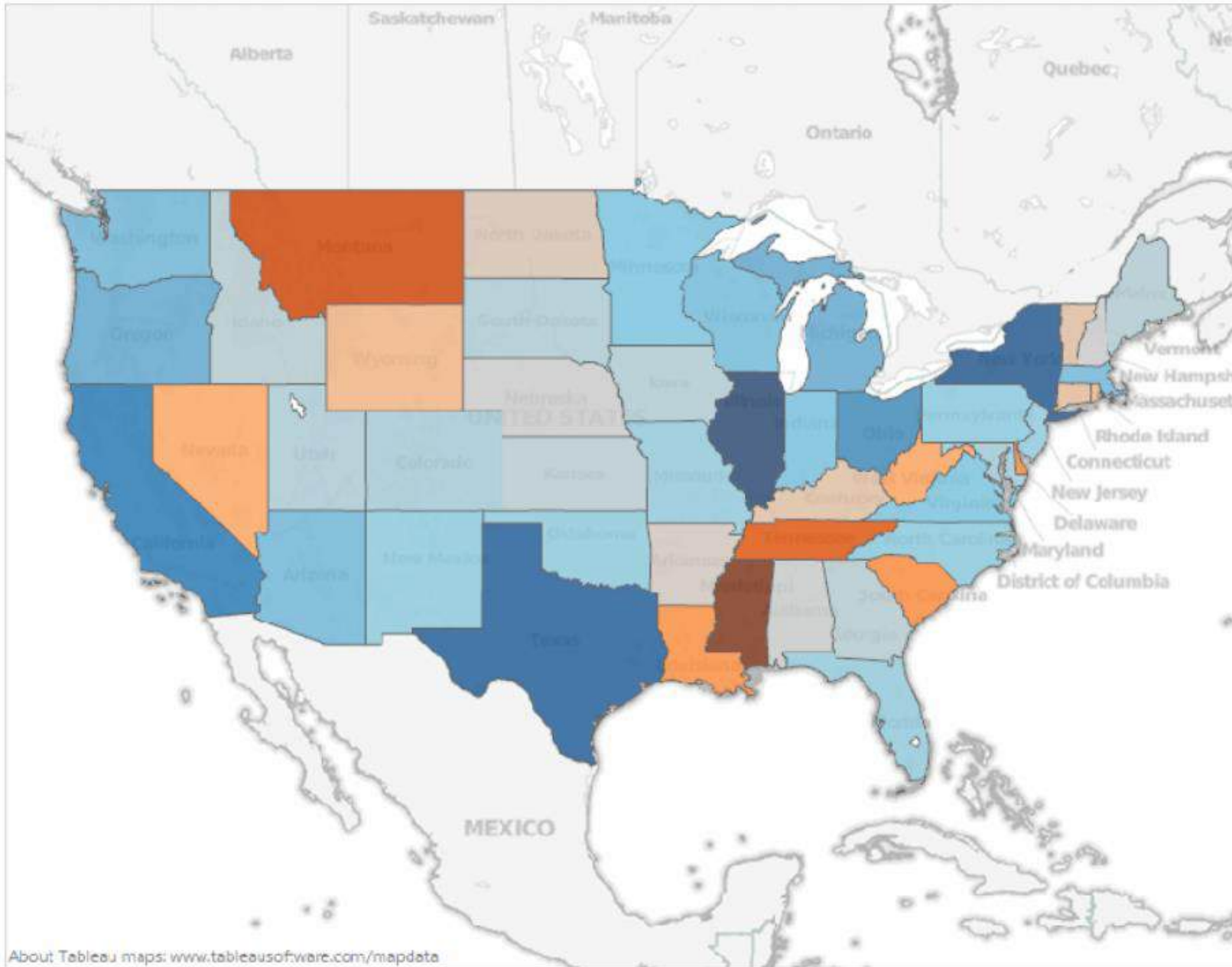


Deuteranope

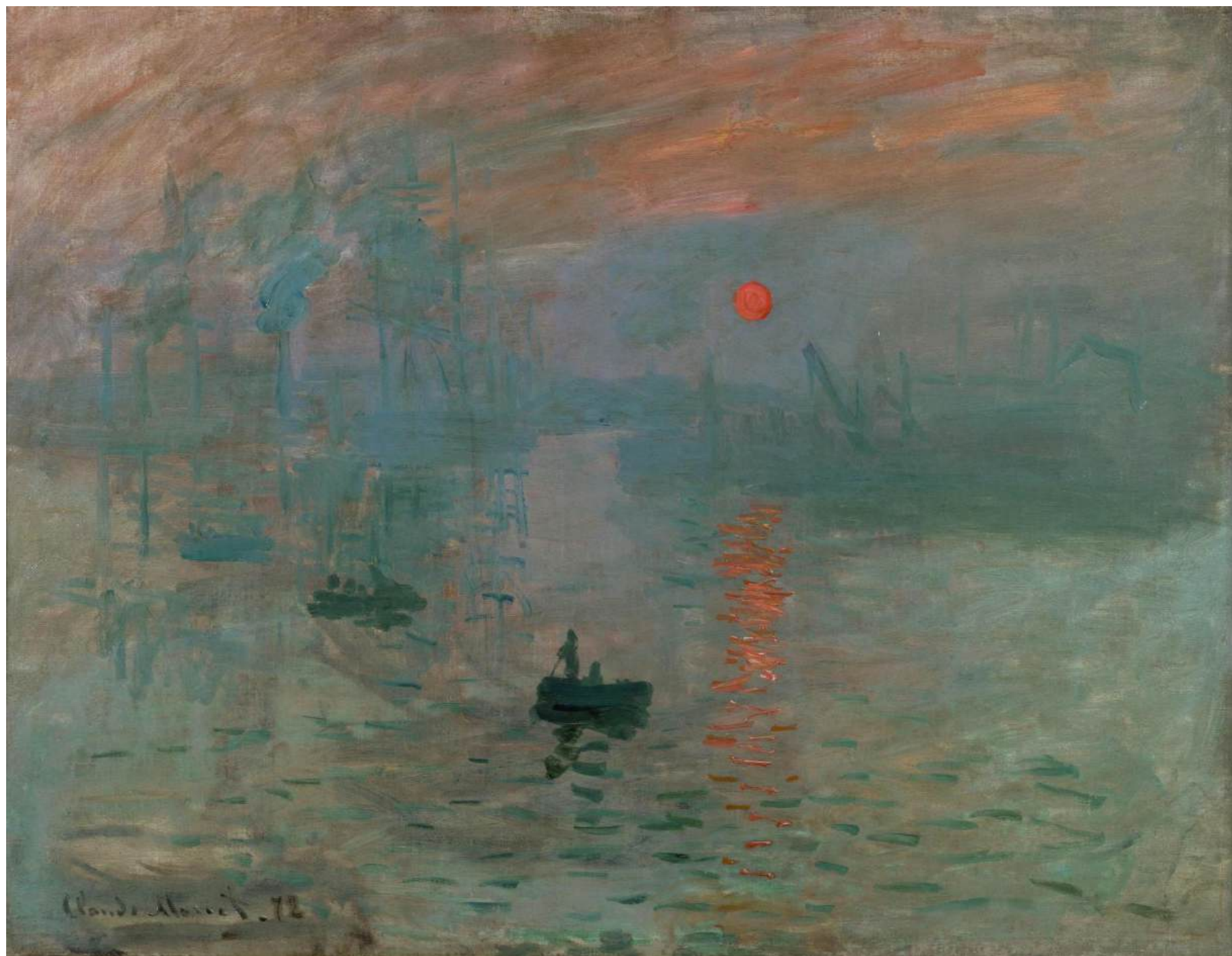


Tritanope

Color deficiency: Blue/orange is safe



Impression, Sunrise,
Claude Monet, 1872



Impression, Sunrise,
Claude Monet, 1872

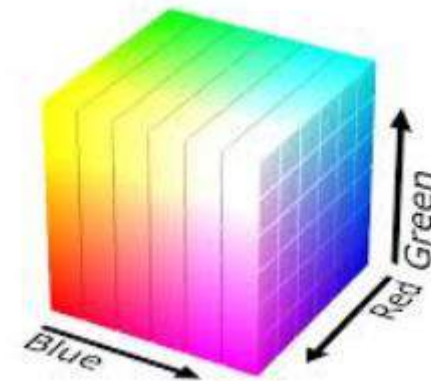
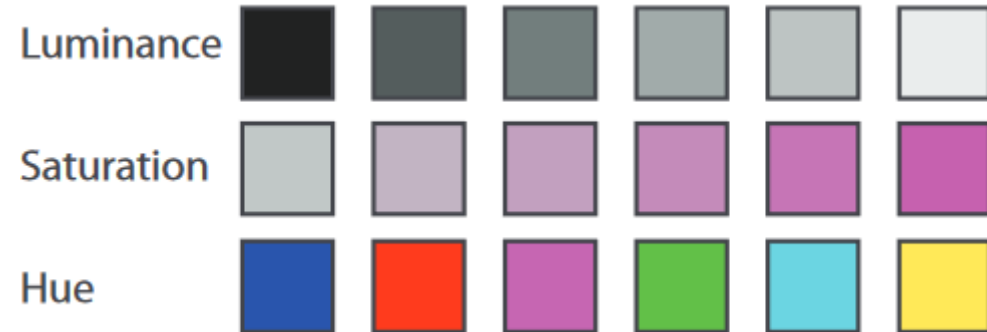




Break
back at 12:31pm

Color spaces

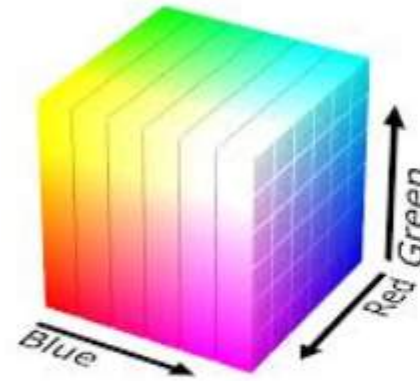
- There are many color spaces
- Luminance (L^*), hue (H), saturation (S)
 - Good for encoding
 - Not standard in graphics and tooling
- RGB
 - Good for display hardware



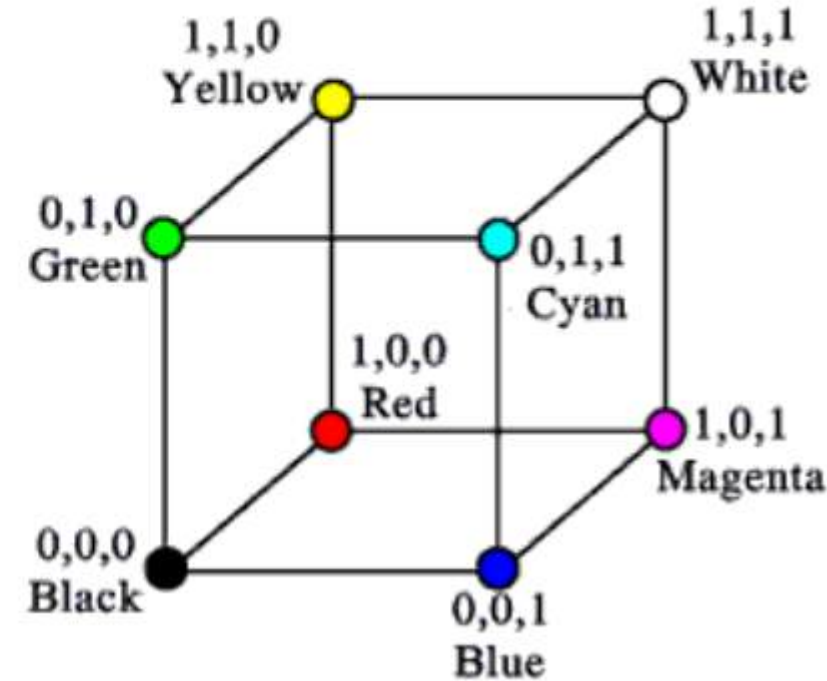
Color spaces: RGB

- RGB
 - Good for display hardware

Corners of the RGB color cube



https://commons.wikimedia.org/wiki/File:RGB_color_solid_cube.png



- Bad for encoding & interpolating

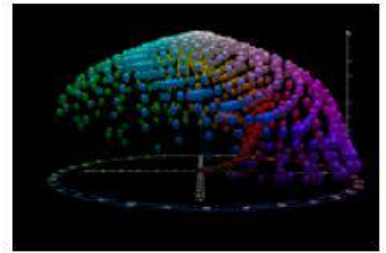
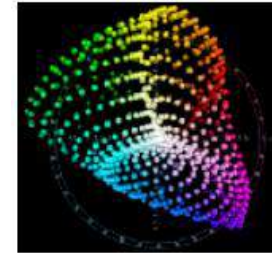
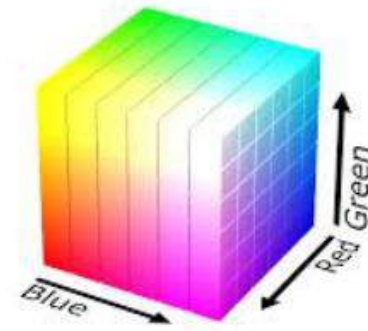
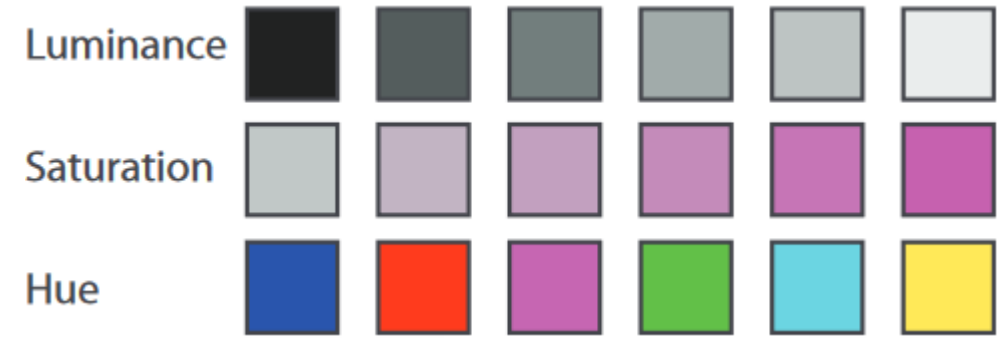
Red
+ Green



Major interference

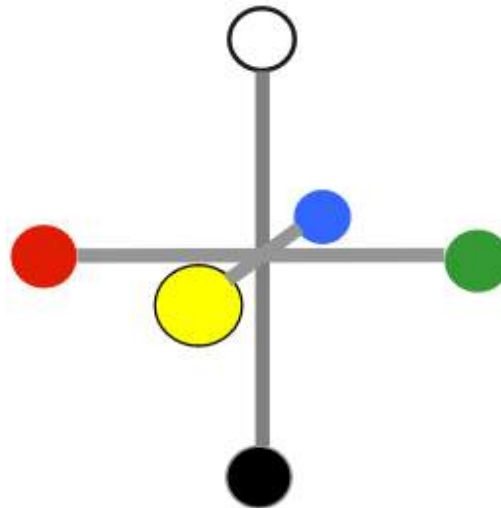
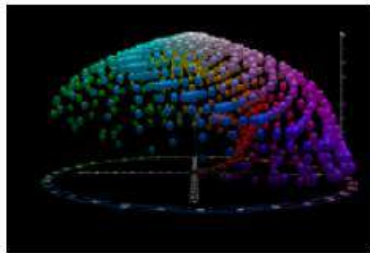
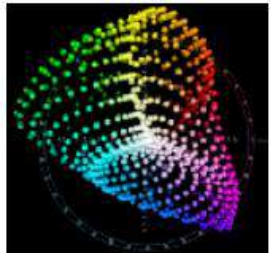
Color spaces

- There are many color spaces
- Luminance (L^*), hue (H), saturation (S)
 - Good for encoding
 - Not standard in graphics and tooling
- RGB
 - Good for display hardware
- CIE LAB ($L^*a^*b^*$)
 - Good for interpolation
 - Hard to interpret, poor for encoding



Perceptual color space: $L^*a^*b^*$

- Visual processing:
 - one achromatic luminance channel (L^*)
- 2 chroma channels
 - Red-green (a^*) & yellow-blue axis (b^*)
- CIE LAB
 - Perceptually uniform
 - Great for interpolating
 - Complex shape
 - Poor for encoding (hard to understand)



Luminance information

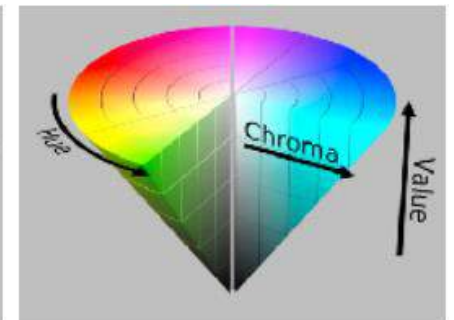
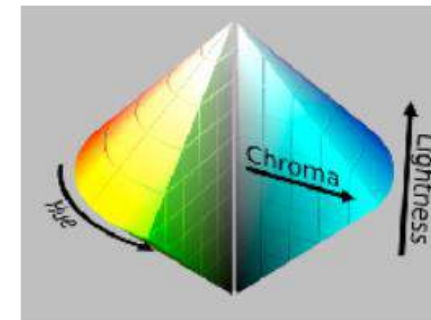
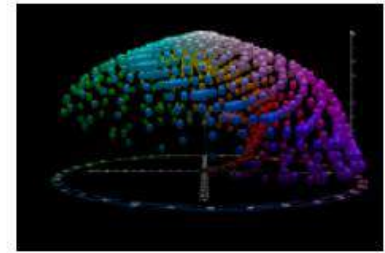
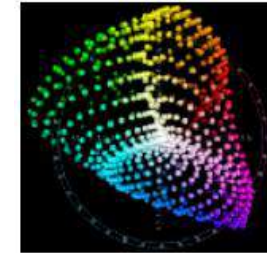
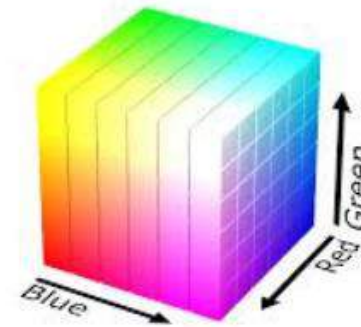
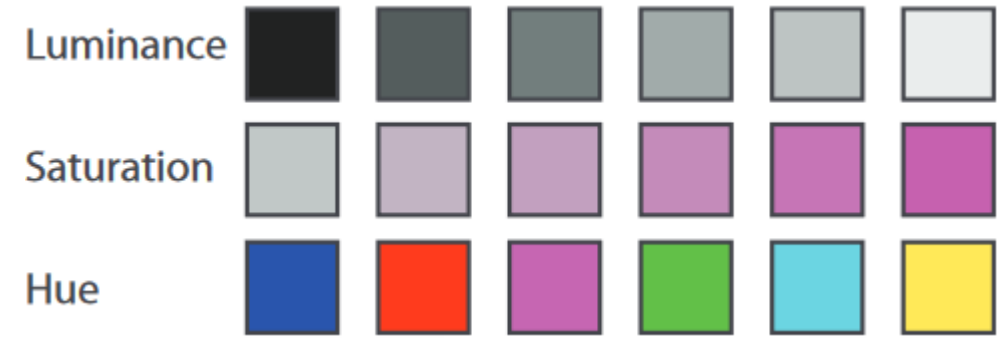


Saturation/hue information



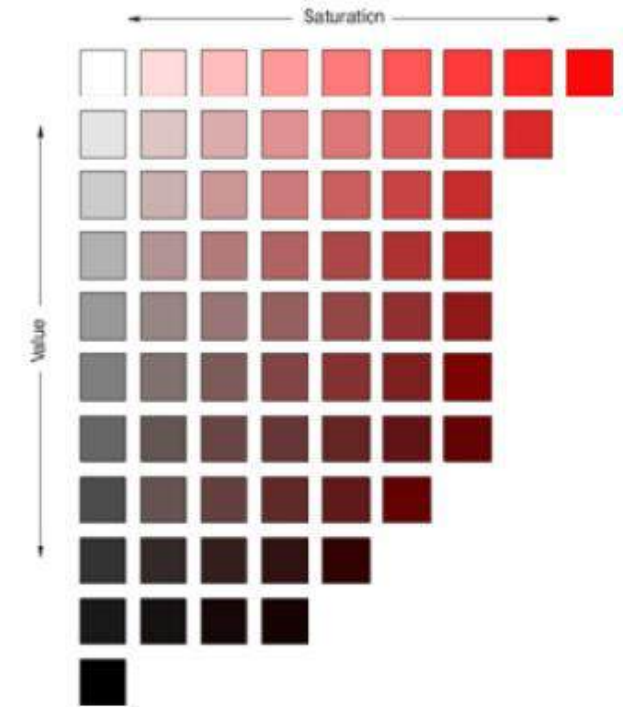
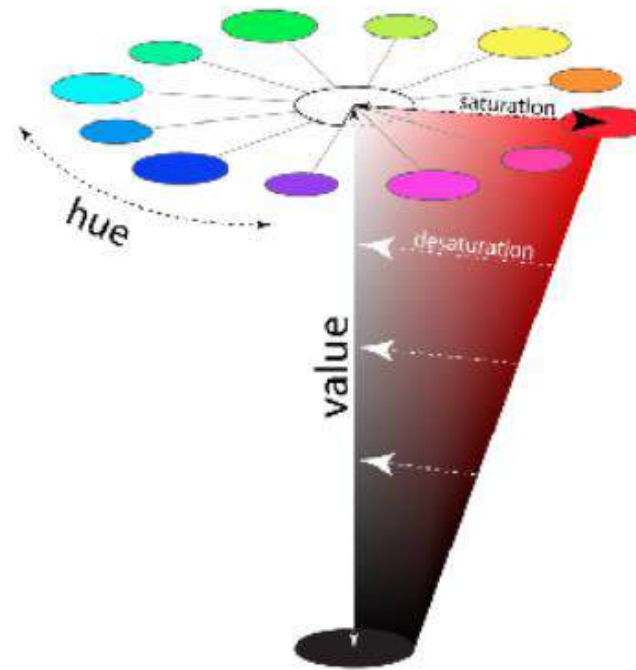
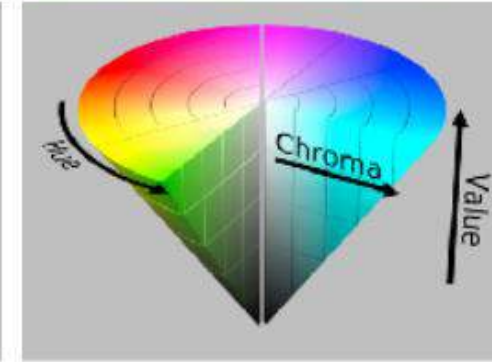
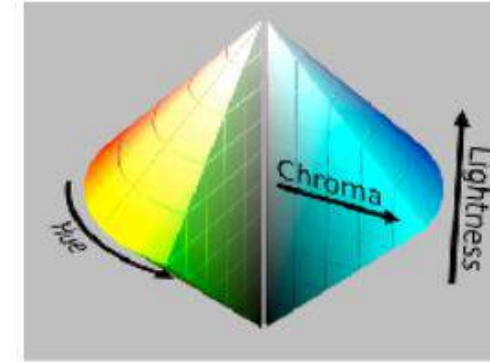
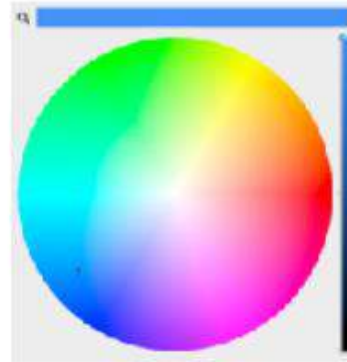
Color spaces

- There are many color spaces
- Luminance (L^*), hue (H), saturation (S)
 - Good for encoding
 - Not standard in graphics and tooling
- RGB
 - Good for display hardware
- CIE LAB ($L^*a^*b^*$): good for interpolation
 - Hard to interpret, poor for encoding
- HSL/HSV: somewhat better for encoding
 - Hue/saturation wheel is intuitive
 - Beware: only pseudo-perceptual
 - Lightness (L) or Value (V) \neq Luminance (L^*)



HSL/HSV

- Somewhat better for encoding
 - Hue/saturation wheel intuitive
- Saturation
 - In HSV (single cone): desaturated = white
 - In HSL (double cone): desaturated = grey
- Luminance vs saturation
 - These channels are not separable
 - We care about hue vs luminance/saturation for viz



HSL/HSV: Pseudo-perceptual

- HSL better than RGB for encoding but beware
 - L lightness \neq L* luminance

Corners of the RGB
color cube



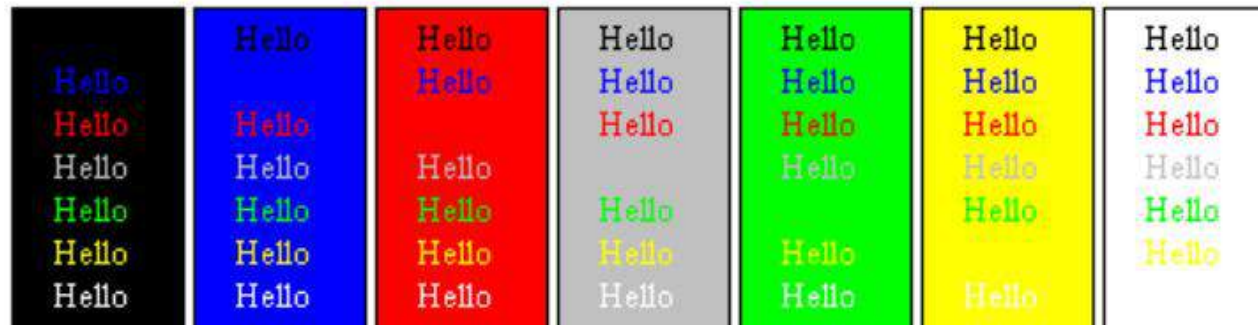
L from HLS
All the same



Luminance values

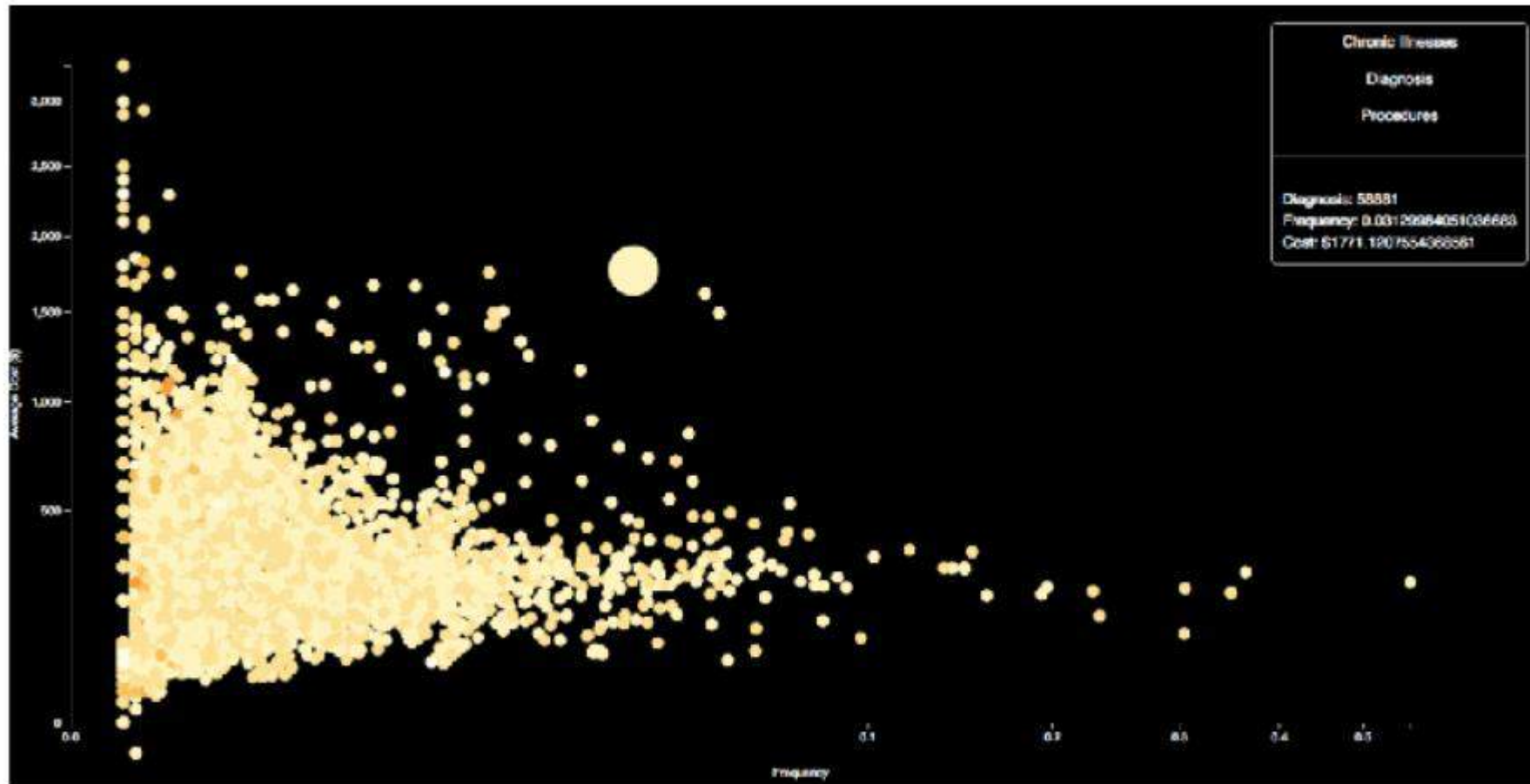


Color Contrast



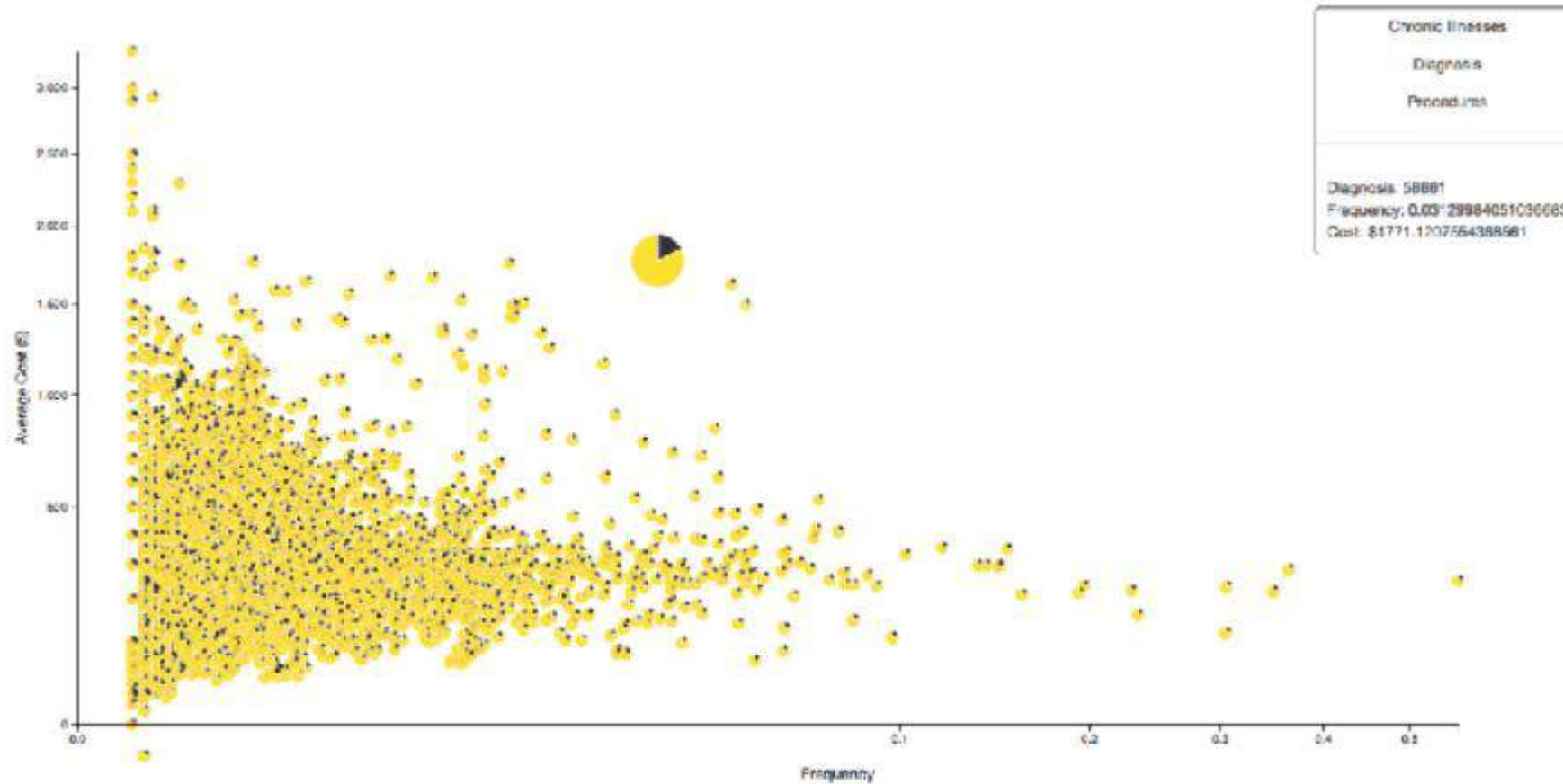
Color contrast: interaction with background

- Marks with high luminance on a background with low luminance



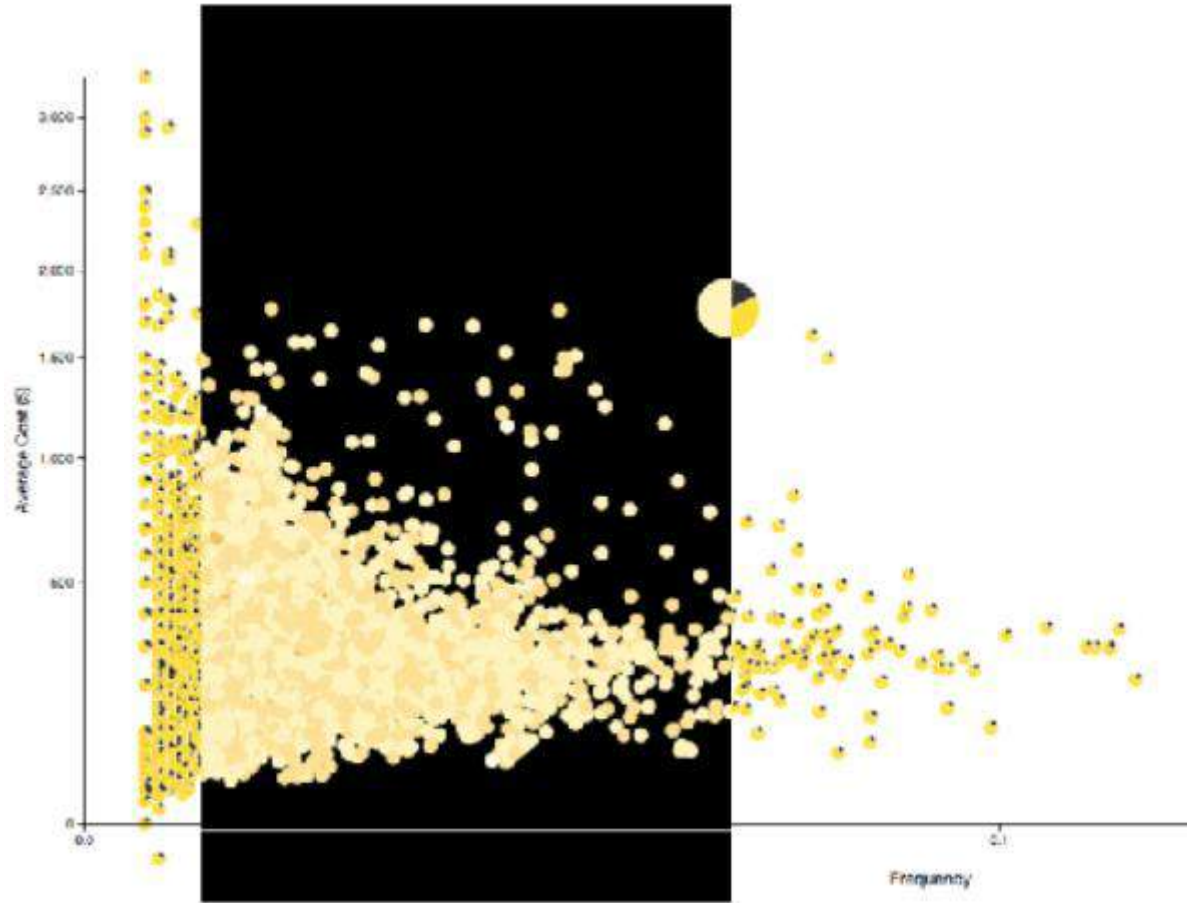
Color contrast: interaction with background

- Marks with medium luminance on a background with high luminance

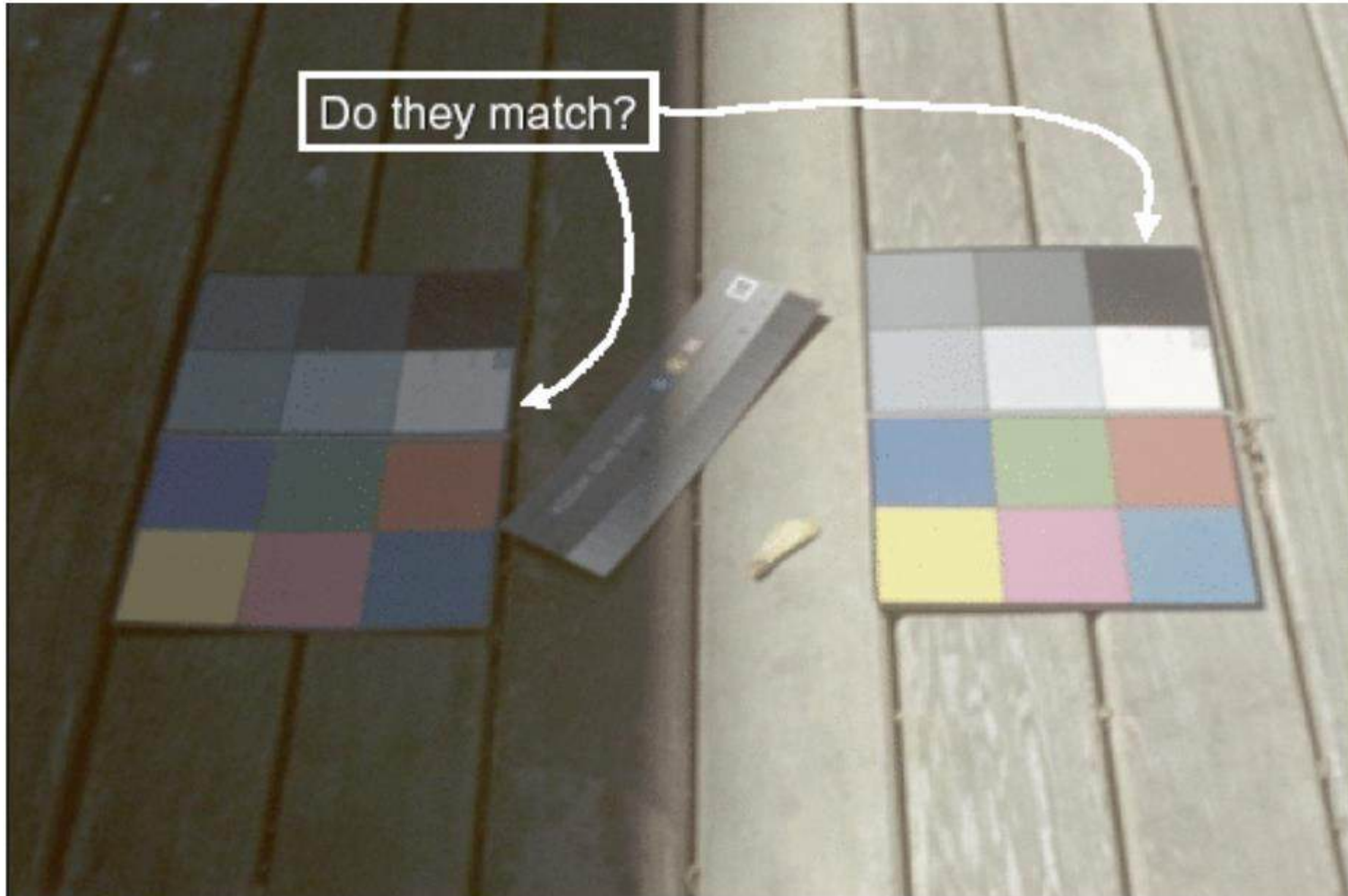


Color contrast: interaction with background

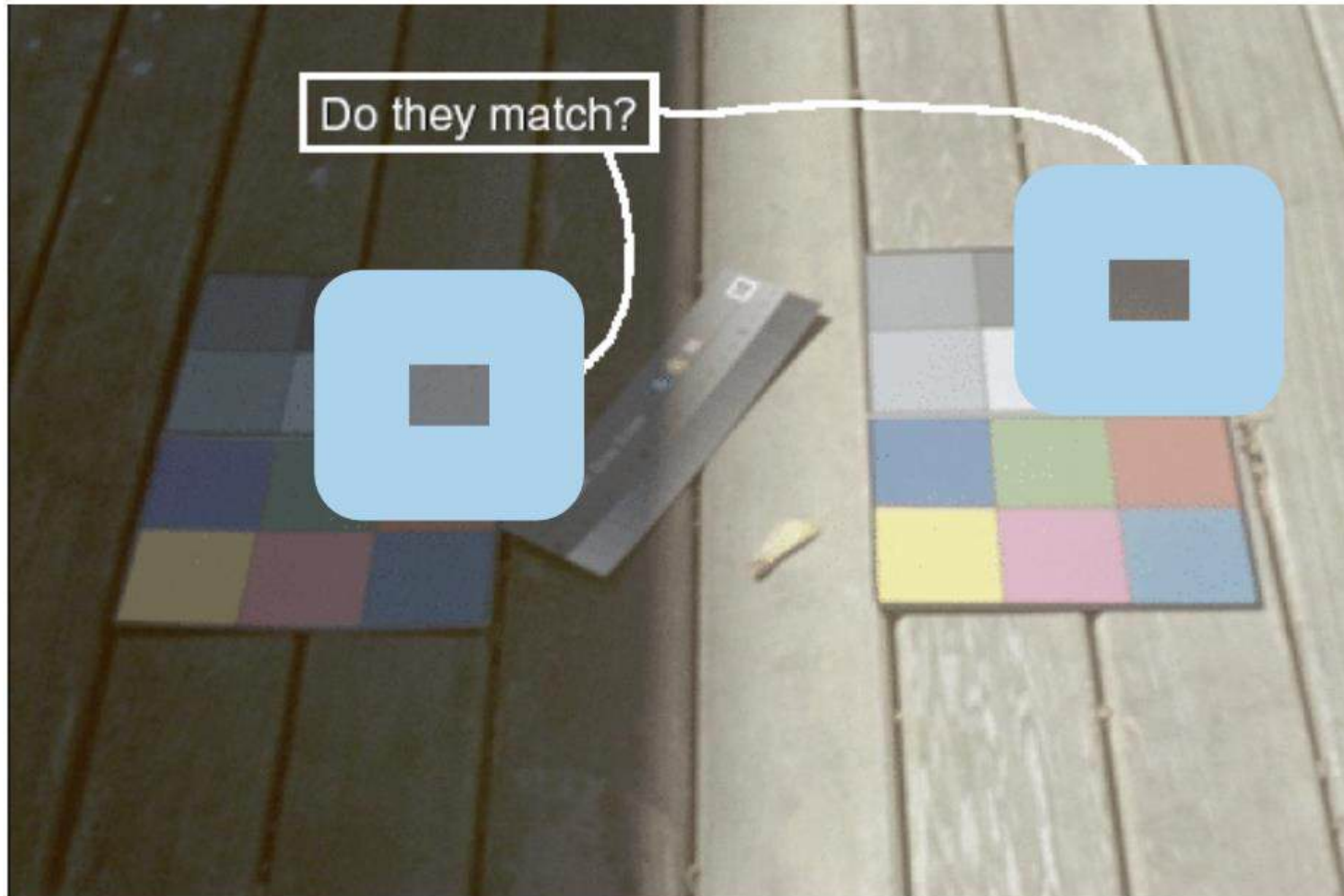
- Change the mark luminance based on the background!



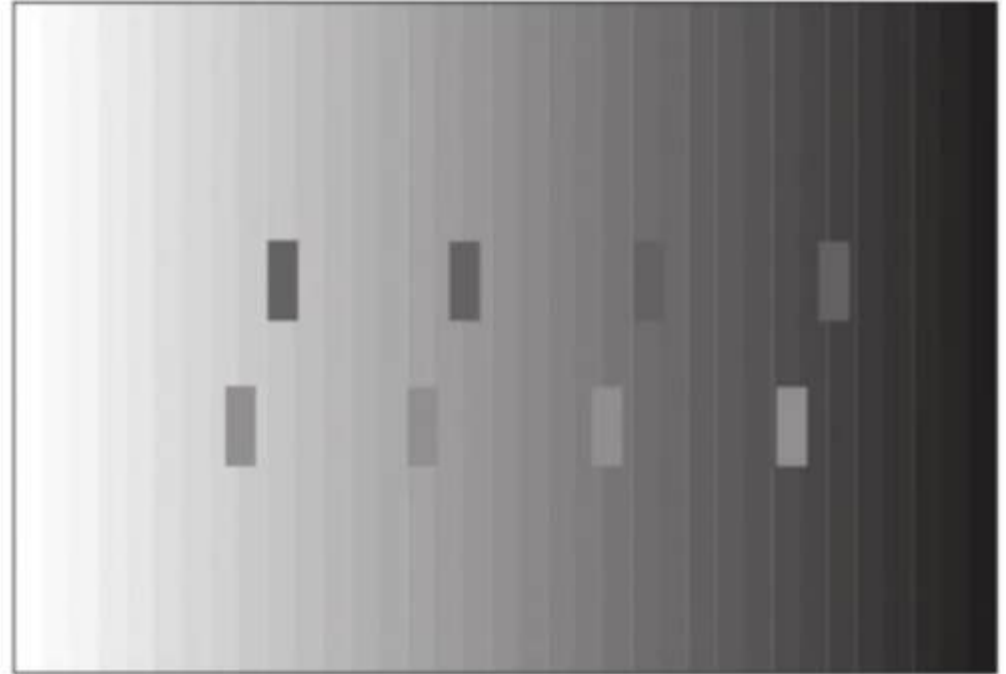
Color/lightness constancy



Color/lightness constancy



Color/lightness constancy: contrast w/background



Color contrast: outlines matter (Bezold Effect)

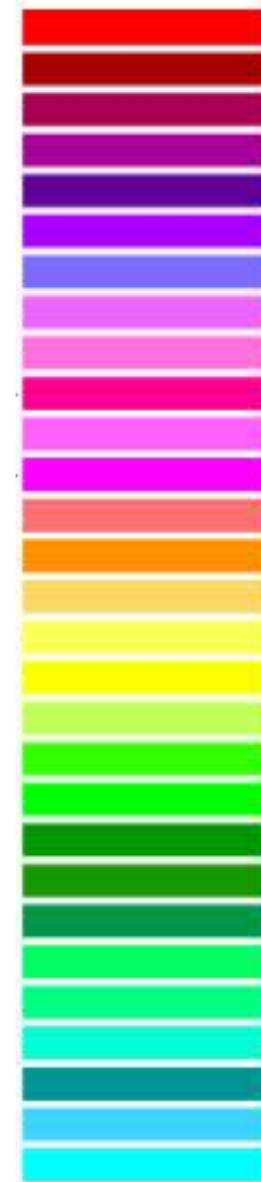


Color appearance

- Not fully specified by any color space
- Depends on the surrounding context!
 - Chromatic adaptation
 - Luminance adaptation
 - Simultaneous contrast
 - Spatial effects
 - Viewing angle
 - ...

Color names

- Has implications on communication and memorability

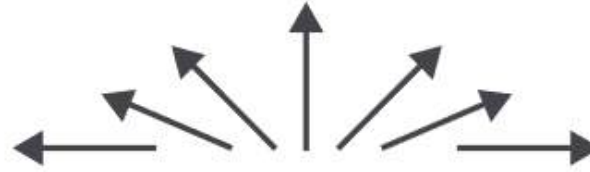


Angle/tilt/orientation channel

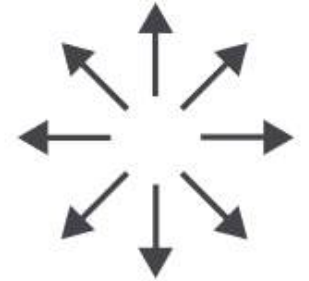
- Angle can also be used for encoding information
- Choose a different mapping based on the data range



Sequential ordered
line mark or arrow glyph



Diverging ordered
arrow glyph



Cyclic ordered
arrow glyph

- Nonlinear accuracy:
 - High: exact horizontal, vertical, diagonal
 - Lower: other orientations (eg 23 vs 24 degrees)

Mapping other channels

- Size

- Aligned lengths is best
- Length is good in general
- 2D area ok
- 3D volume is bad

➞ Size

➞ Length



➞ Area



➞ Volume



- Shape

- Complex combination of lower-level primitives
- Many bins

➞ Shape



- Motion

- Highly separable against static elements
 - Great for highlighting (binary)
- Use carefully to avoid annoying the user

➞ Motion

➞ Motion
*Direction, Rate,
Frequency, ...*



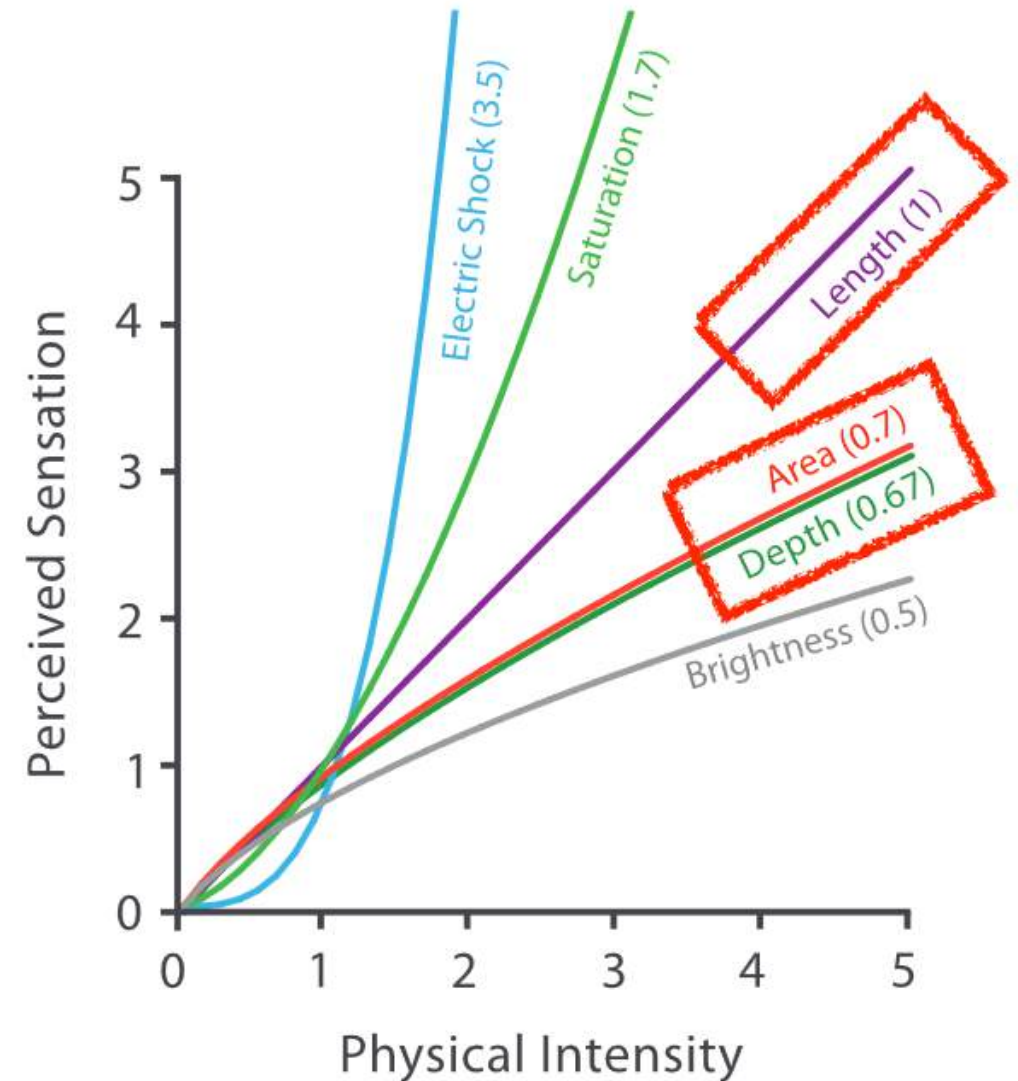
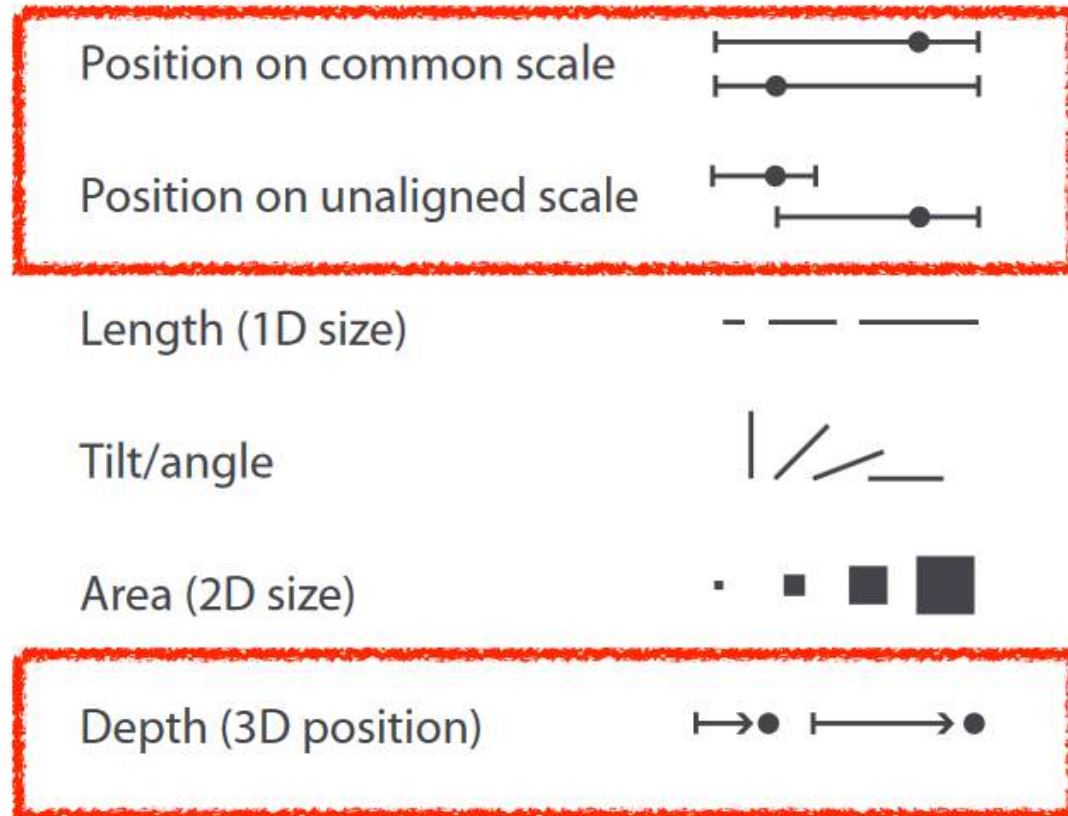
Rules of thumb

- No unjustified 3D
- Eyes beat memory
- Resolution over immersion
- Overview first, zoom and filter, details on demand
- Responsiveness is required
- Function first, form next

No unjustified 3D

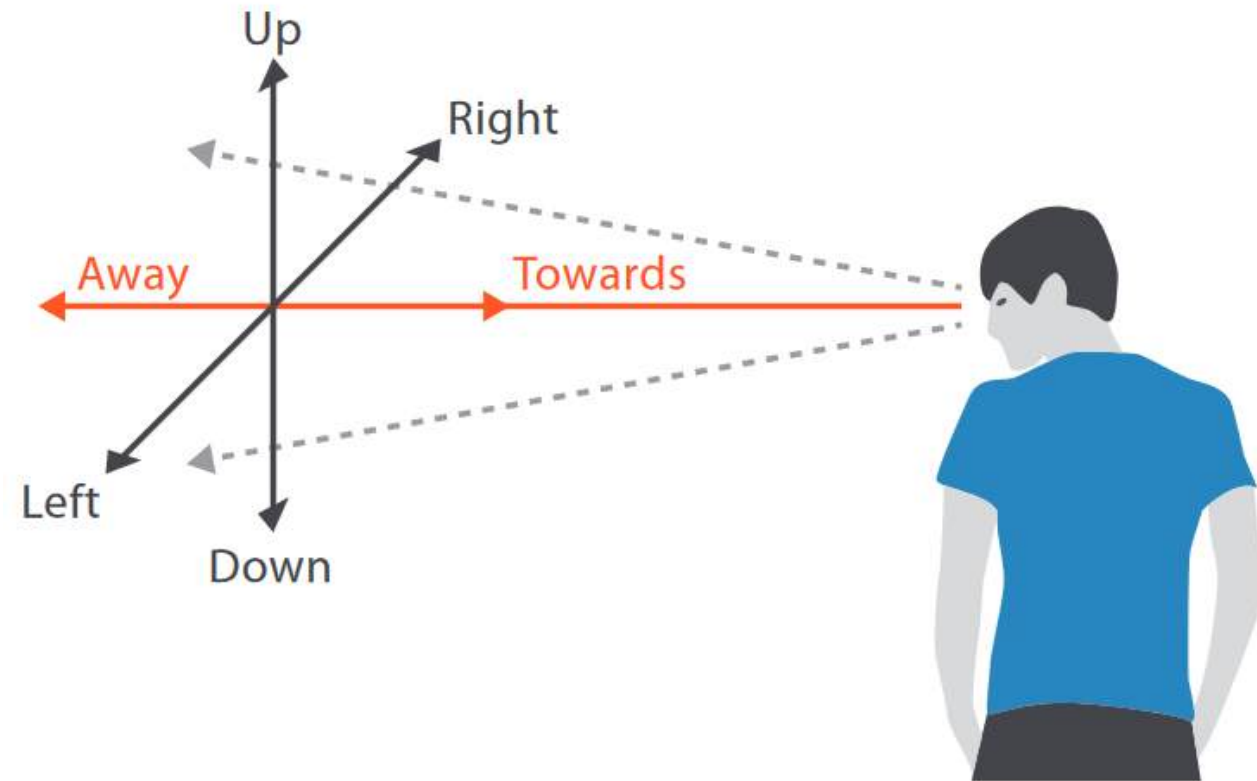
- **Planar** spatial position is good
 - Depth (3D position) is bad!

➔ Magnitude Channels: Ordered Attributes

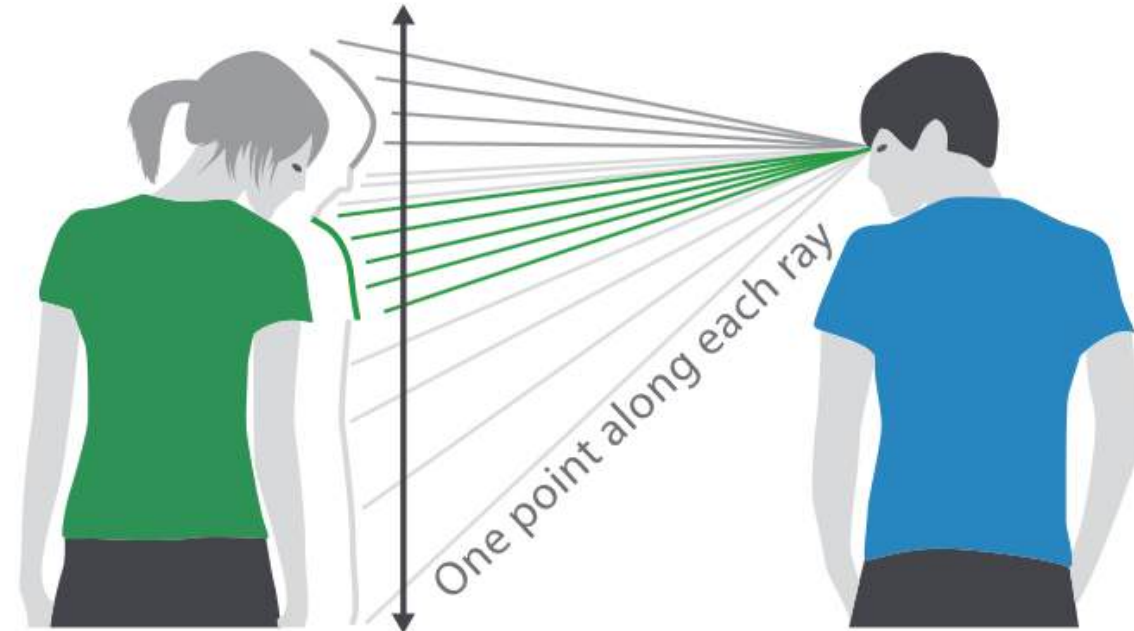


No unjustified 3D

- 3D visual processing is slower
 - A simple fact of the visual system



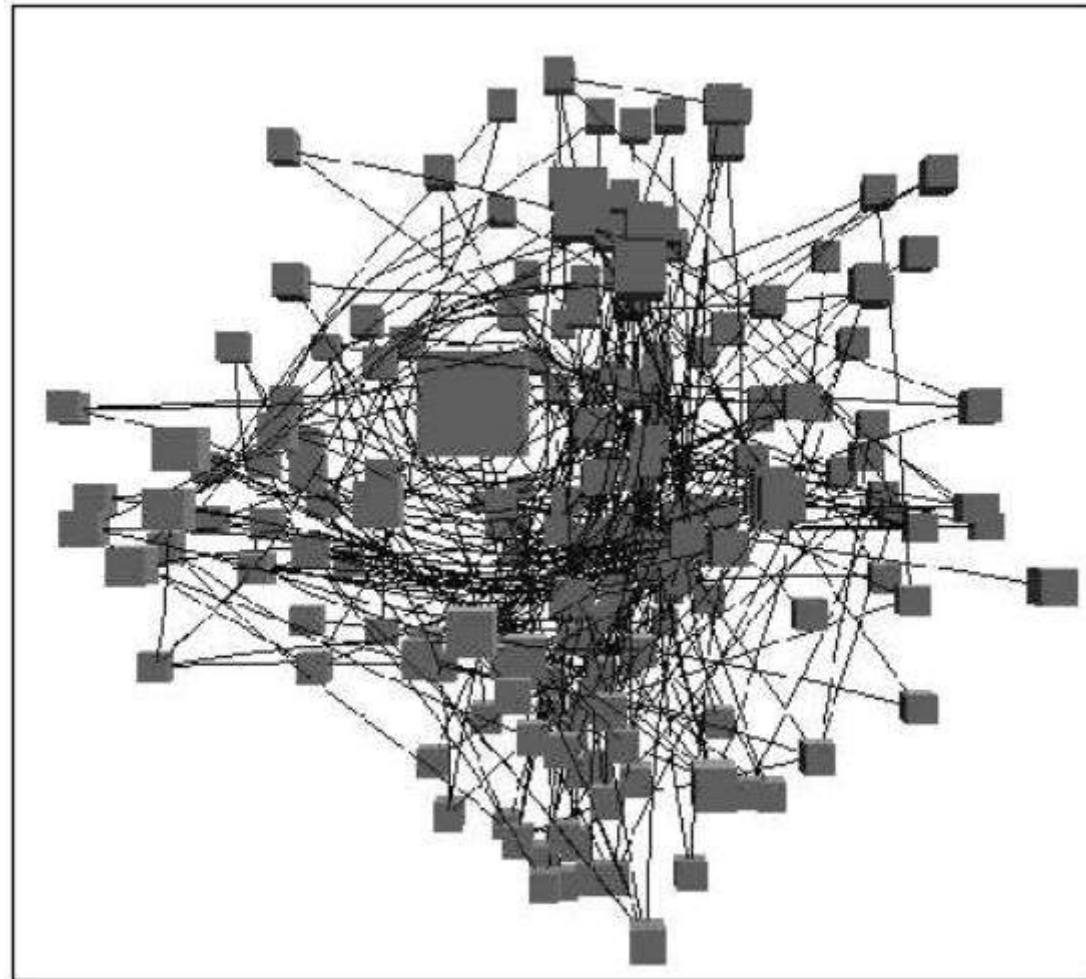
Thousands of points up/down and left/right



We can only see the outside shell of the world

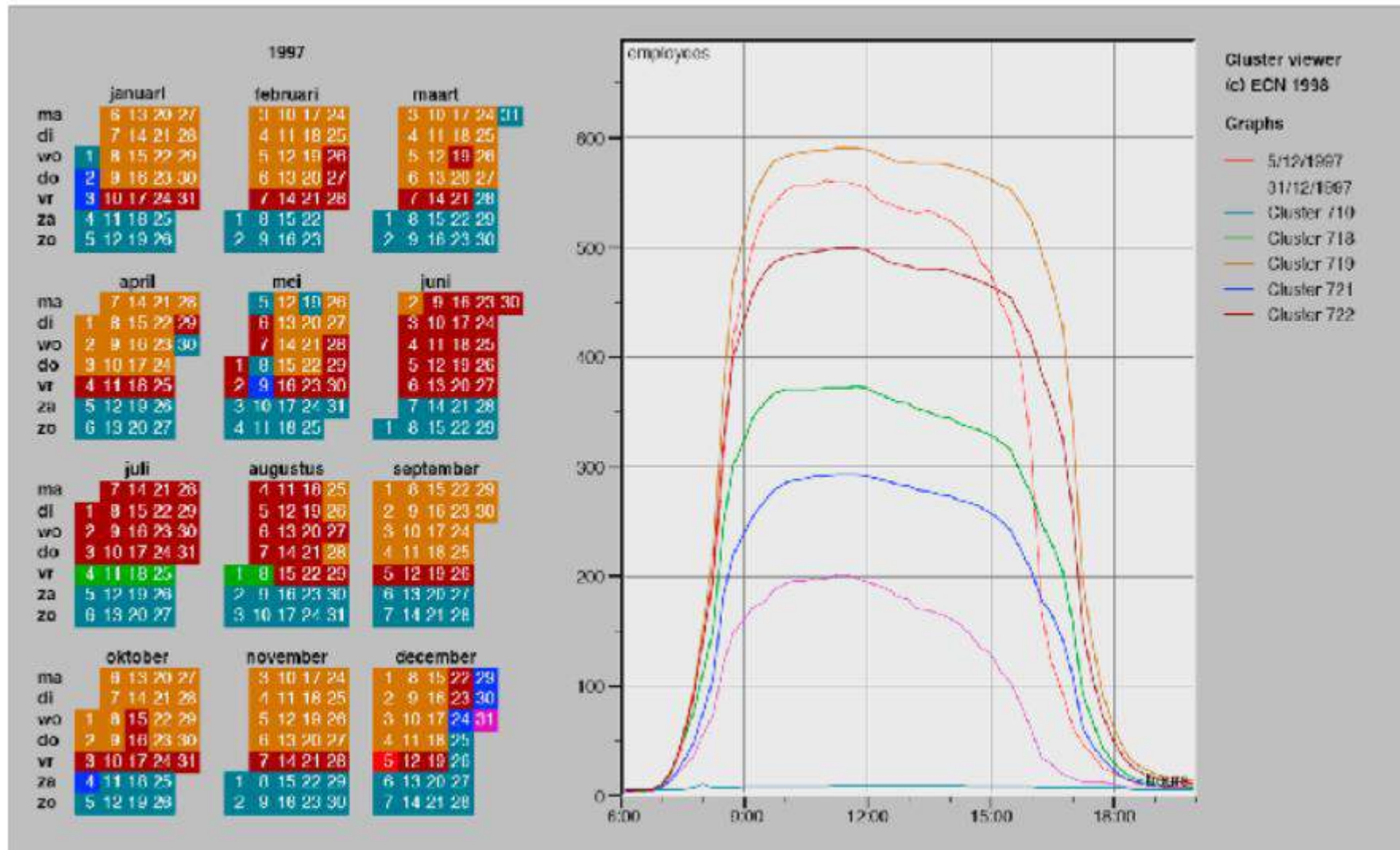
No unjustified 3D

- Occlusions hide information
- Interaction can solve this, but it costs time and cognitive load



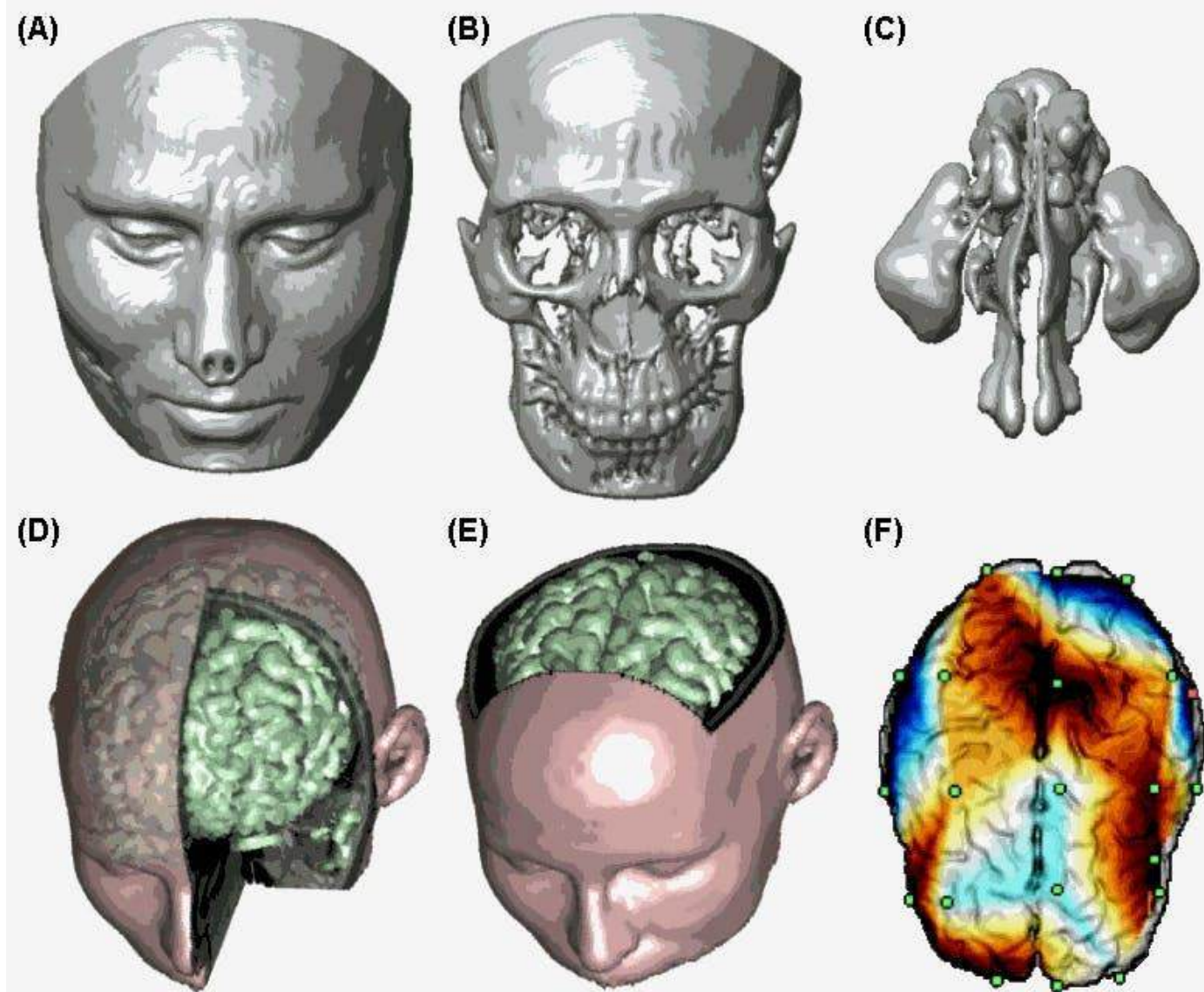
No unjustified 3D: solutions

- Juxtapose multiple views
 - E.g. calendar, superimposed 2D curves



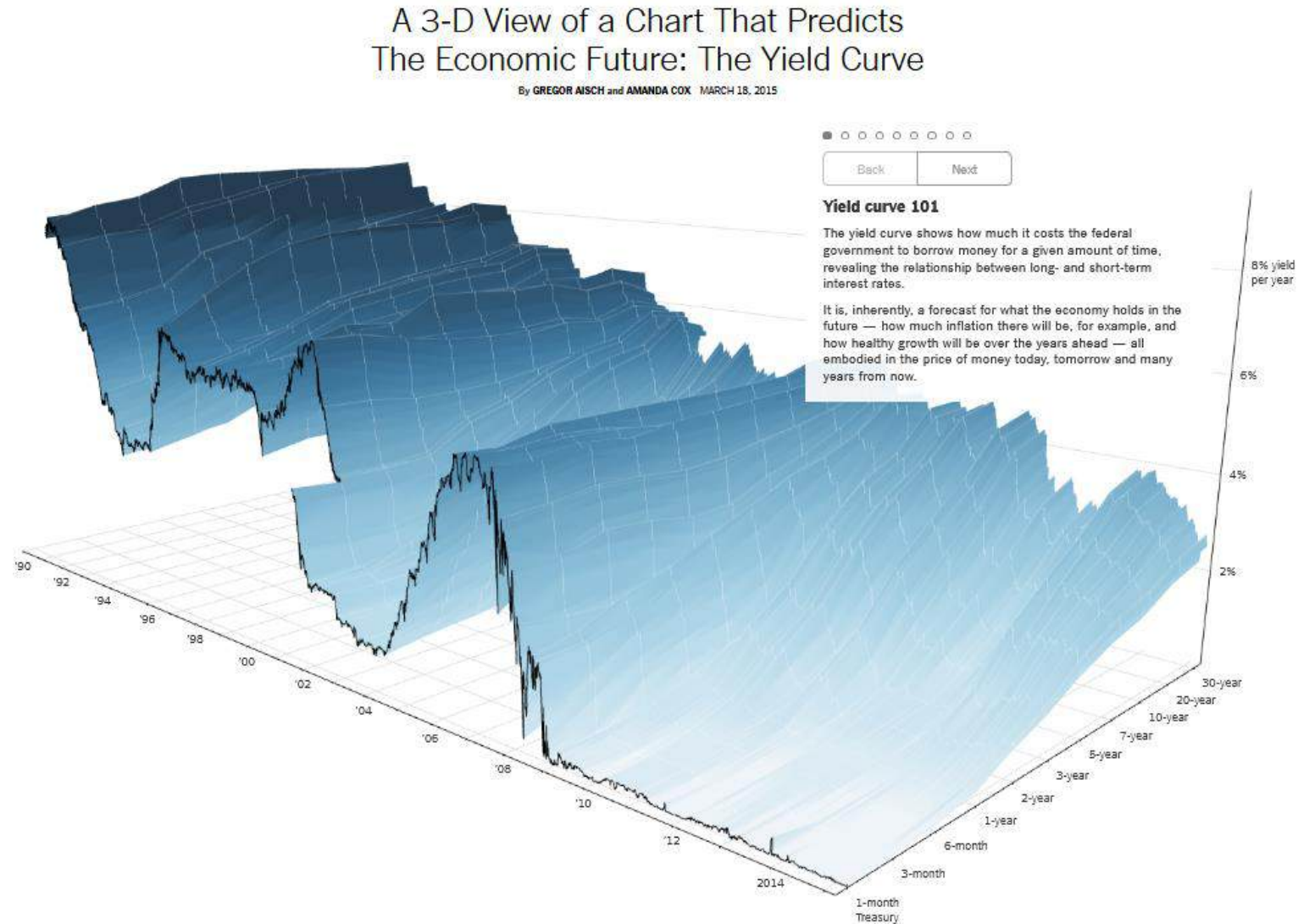
Justified 3D: Shape Perception

- If shape perception is the main task, 3D is often justified



Justified 3D: controlled views

- Carefully controlled viewpoints
- Optimized for the best views of the 3D data



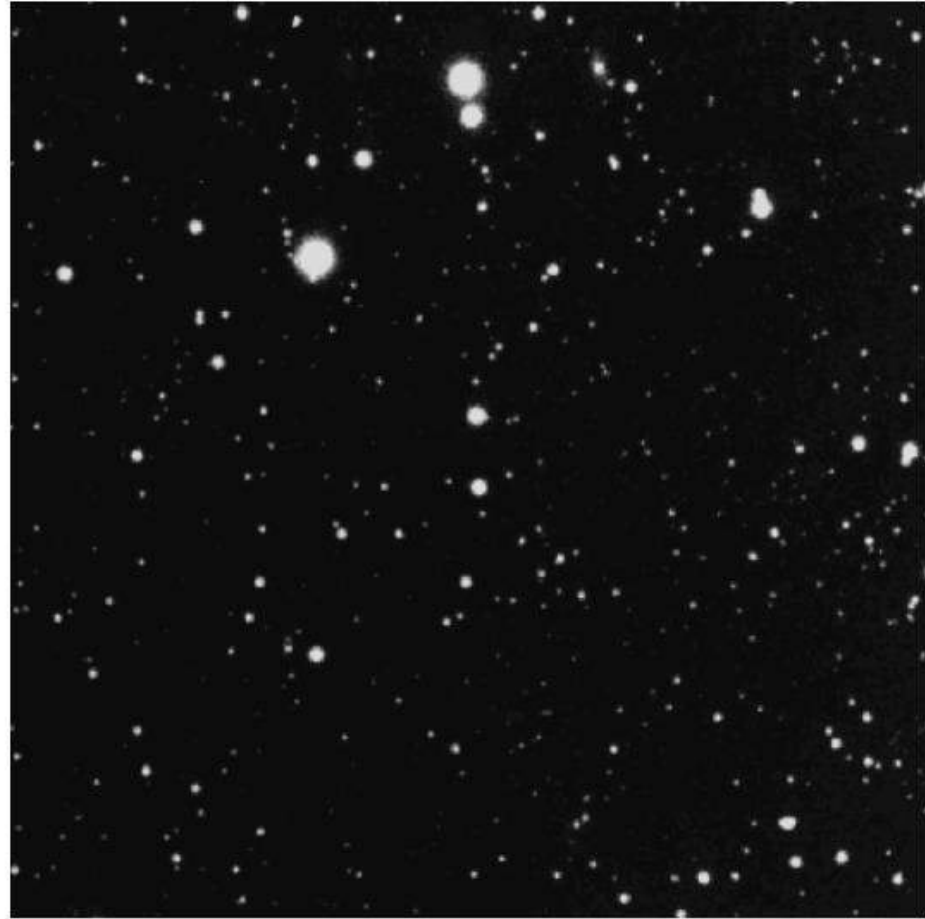
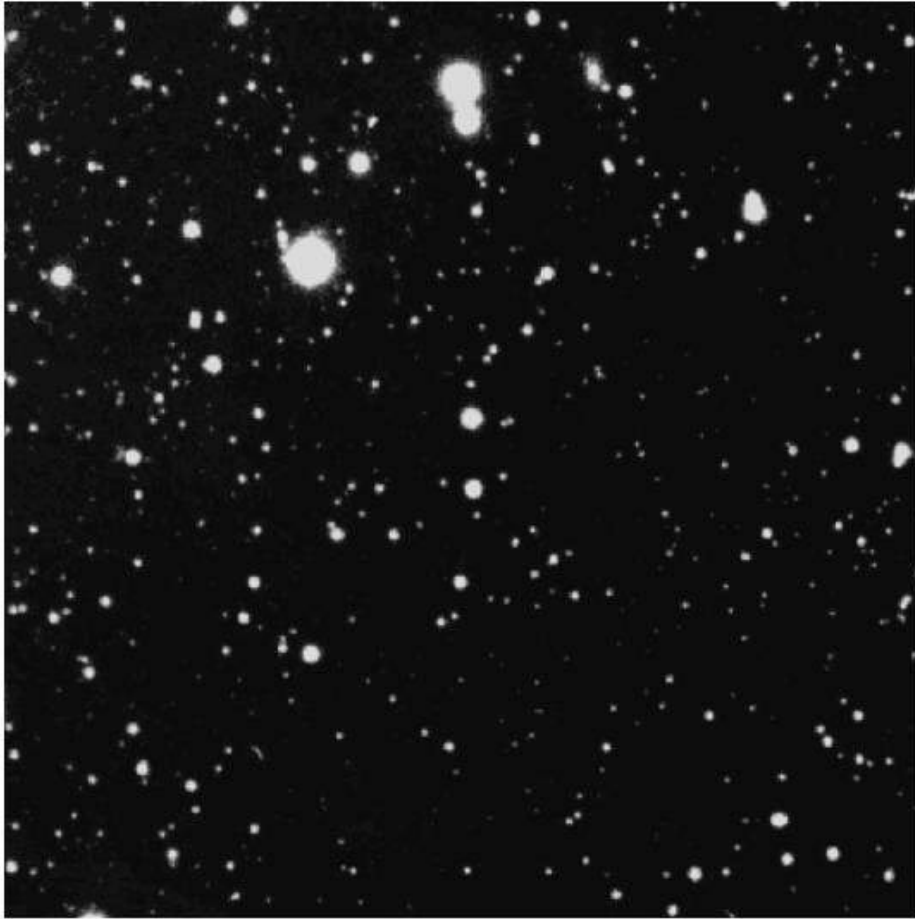
<https://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html>

Eyes beat memory

- Principle: external cognition vs internal memory
 - Easy to compare side-by-side items
 - Hard to compare a visible item to your memory of another item
- Implications for animation:
 - Good for transitions and choreographed storytelling
 - Bad for data with many states/views

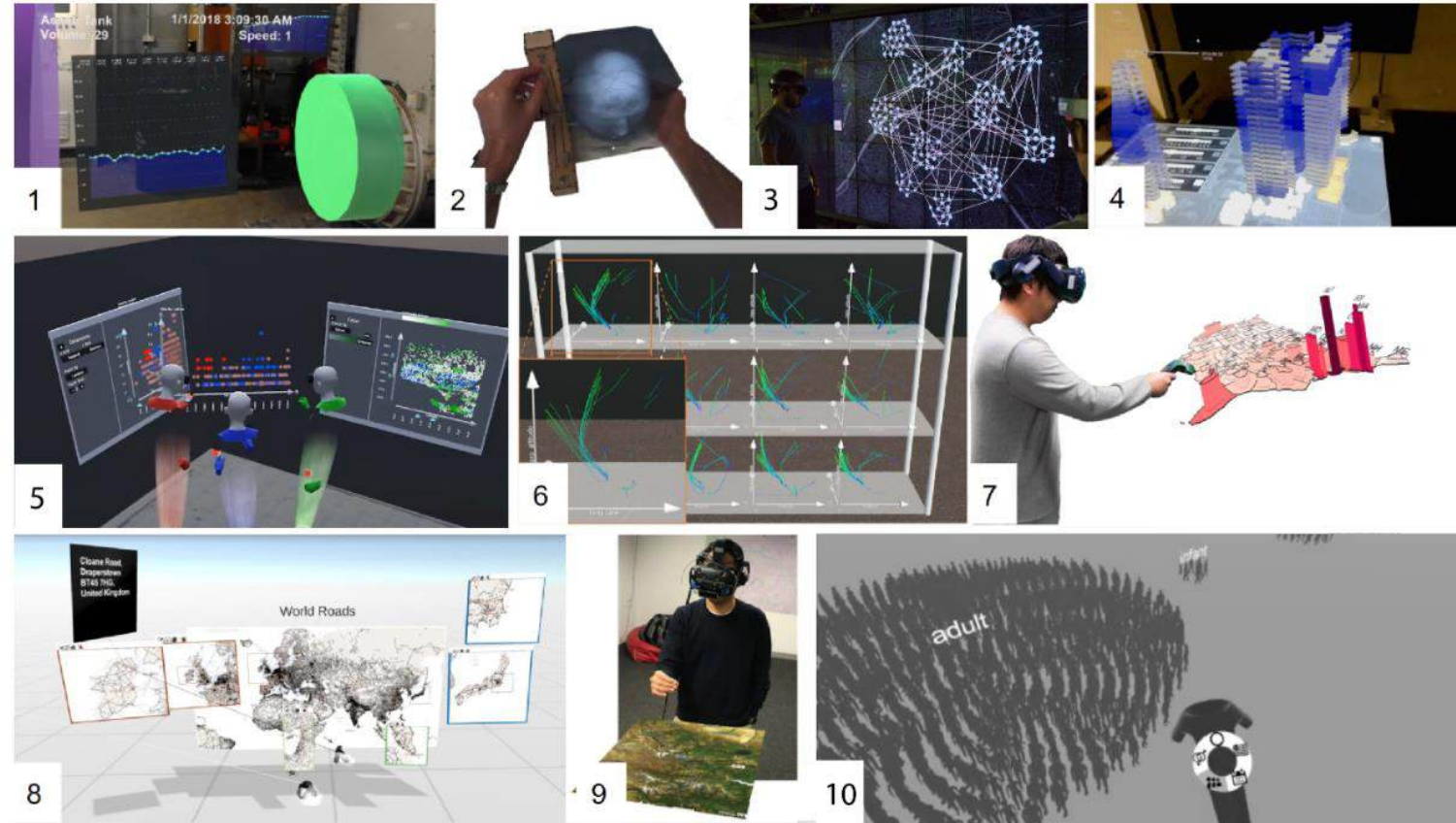
Eyes beat memory? Blink comparison

- Special case where animation (memory) beats side-by-side comparison (eyes)
 - Blink comparator used to discover Pluto
 - Change blindness!



Resolution beats immersion

- Immersion is usually not needed **for abstract data**
 - No need for things like virtual reality
- Resolution is more important
 - Pixels are a scarce resource
- Not completely useless, though...
 - Spatial data (augmented reality)
 - Collaborative analytics
 - Complex interactions with data



Ens, Barrett, et al. "Grand challenges in immersive analytics." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021.

Overview first, zoom and filter, details on demand

- Provide a summary/high-level view of the data
- Allow users to zoom into specific views of the data
 - Requires filtering the dataset to only render specified data
- Show more details when the users asks for it
 - Tooltips and other interactions
- Influential paper: Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." *The craft of information visualization*. Morgan Kaufmann, 2003. 364-371.

Responsiveness is required

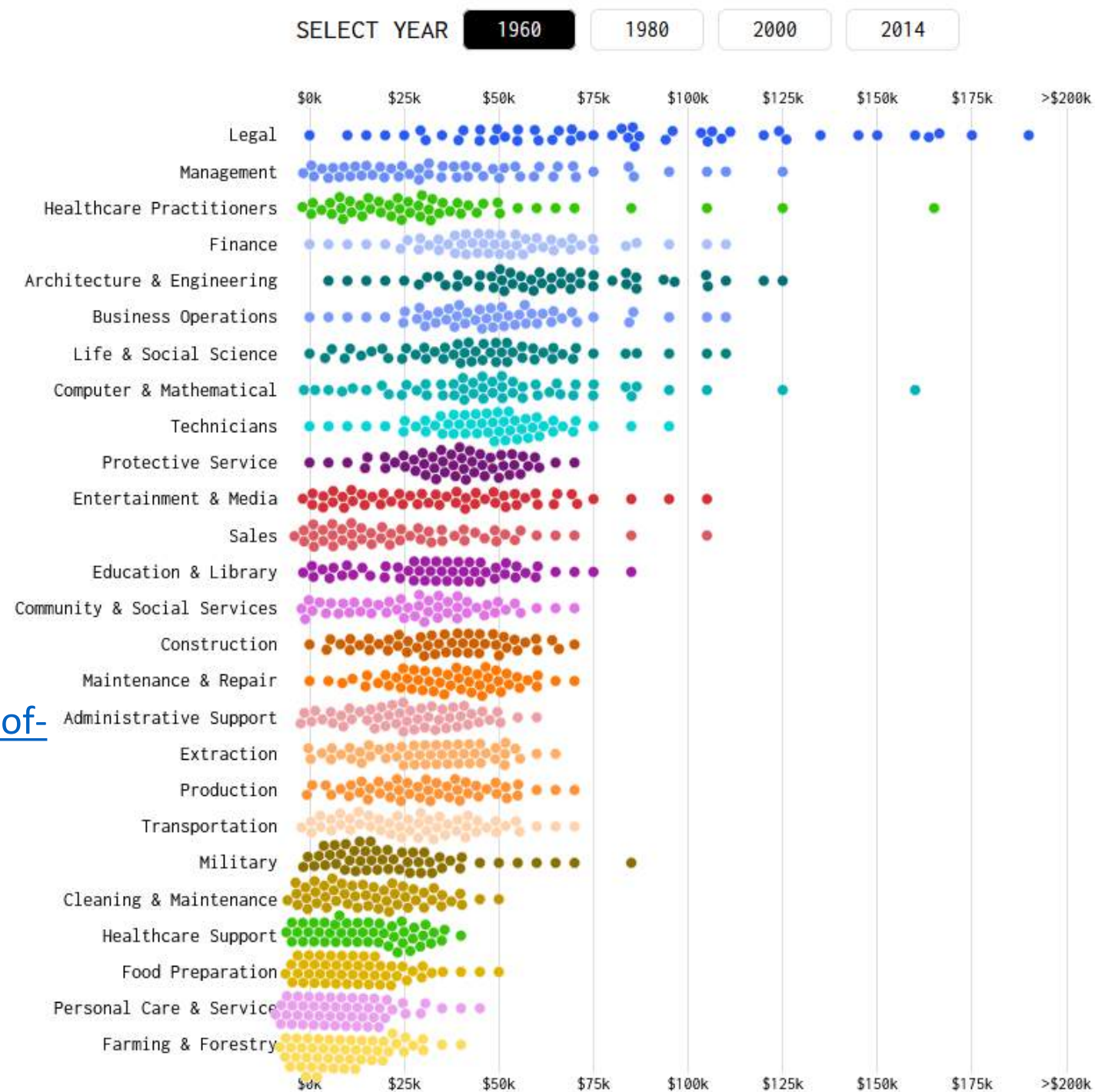
- Three categories of visual feedback:
 - 0.1 seconds: perceptual response
 - subsecond response for mouseover highlighting - ballistic motion
 - 1 second: immediate response
 - fast response after mouse click, button press - Fitts' Law limits on motor control
 - 10 seconds: brief tasks
 - bounded response after dialog box - mental model of heavyweight operation (file load)
- Considerations:
 - Need 60 FPS rendering! → computationally expensive
 - Visual indicators for interactions: hourglass, progress bar, etc.

Function first, form next

- Start with functionality
 - Aesthetics can be changed afterwards, as refinement
 - Aesthetics are still important!
 - Secondary level of function
- Don't start with aesthetics
 - Can't retrofit functionality

Exercise

- Visual channels used?
 - E.g. channel X encodes attribute Y
- Marks used?
 - E.g. mark of type X encodes item Y
- Any suggested improvements?

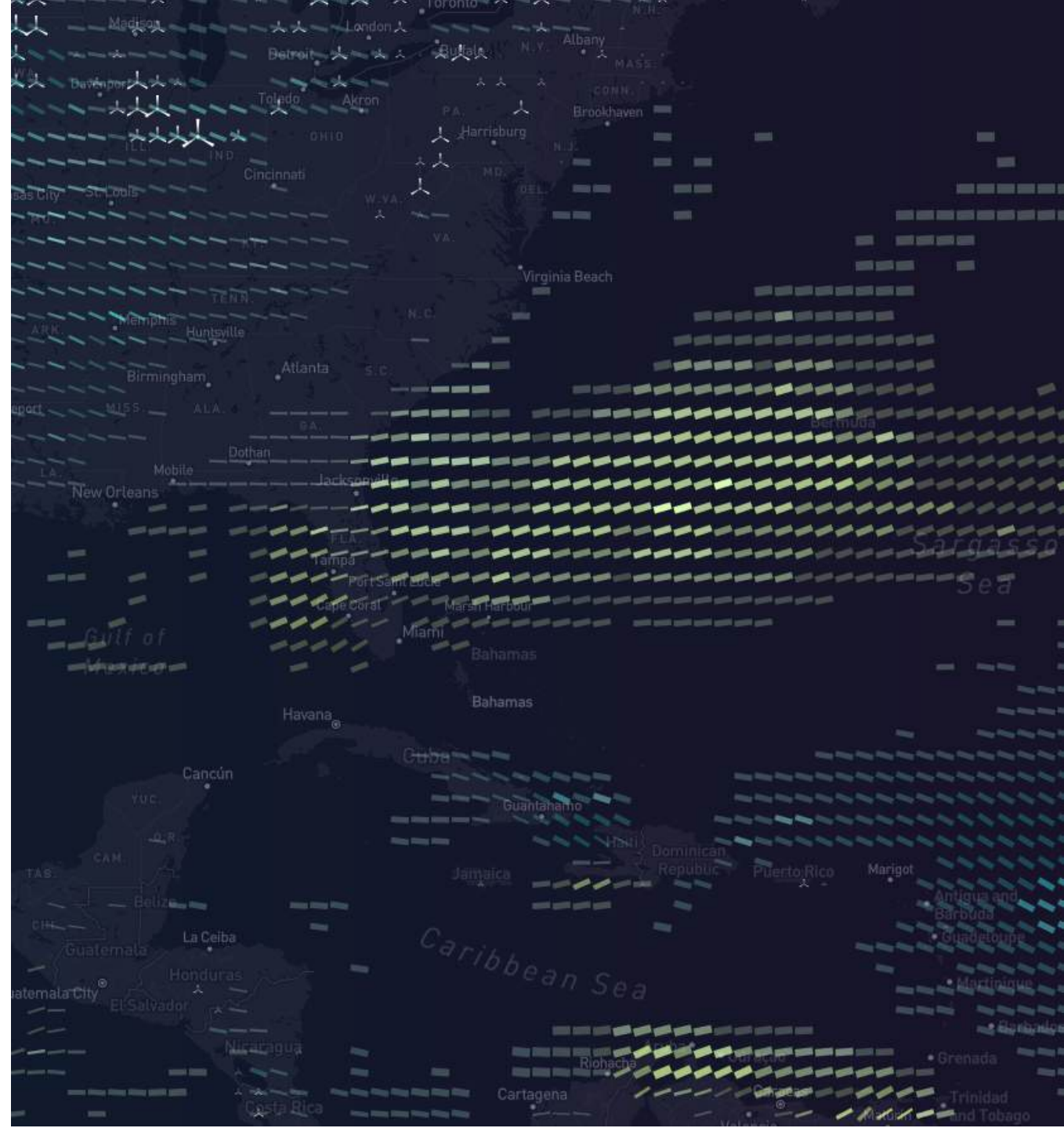


<https://flowingdata.com/2016/06/28/distributions-of-annual-income/>

Exercise

- Visual channels used?
 - E.g. channel X encodes attribute Y
- Marks used?
 - E.g. mark of type X encodes item Y
- Any suggested improvements?

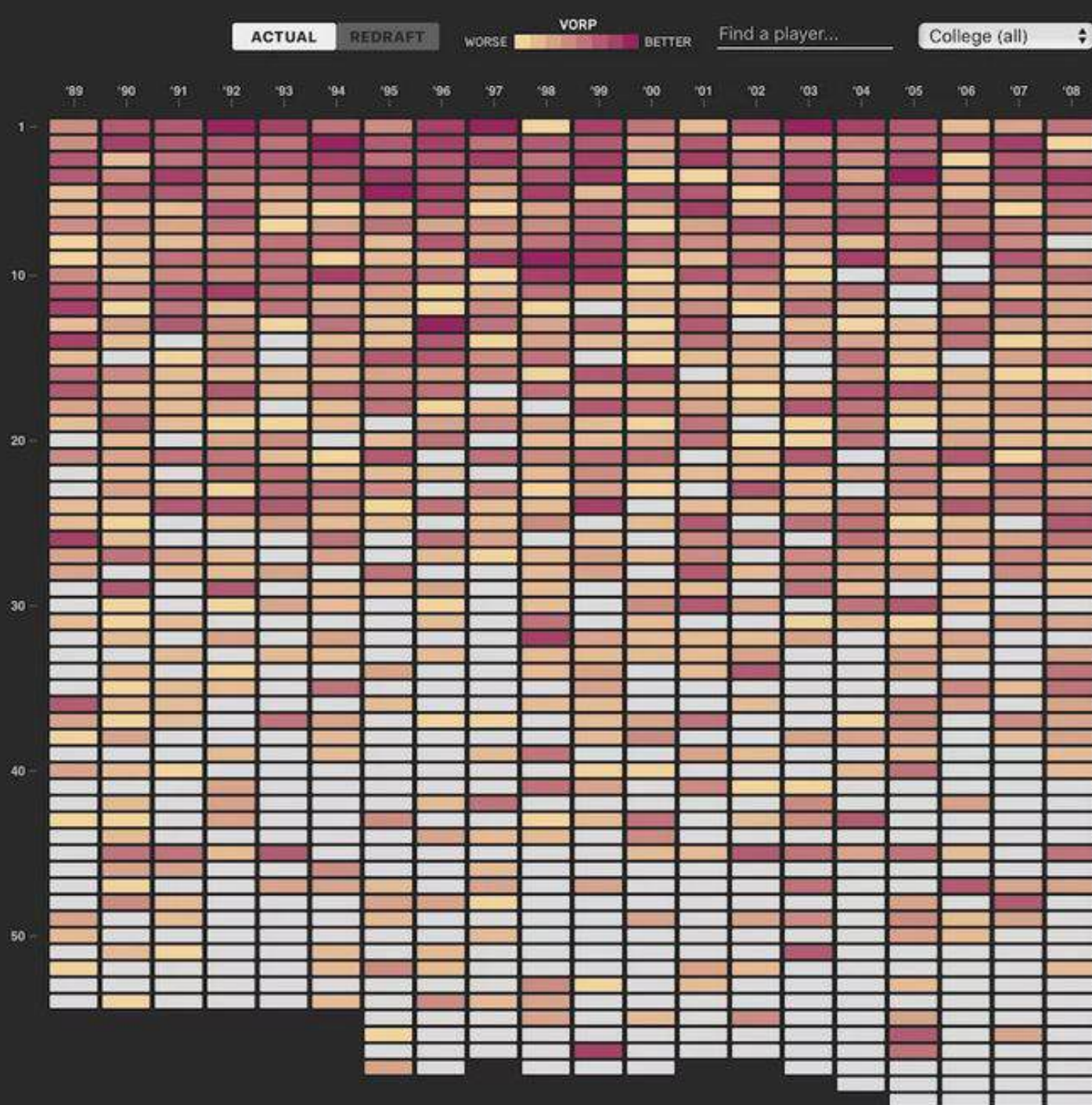
<https://project-ukko.net/map.html>



Exercise

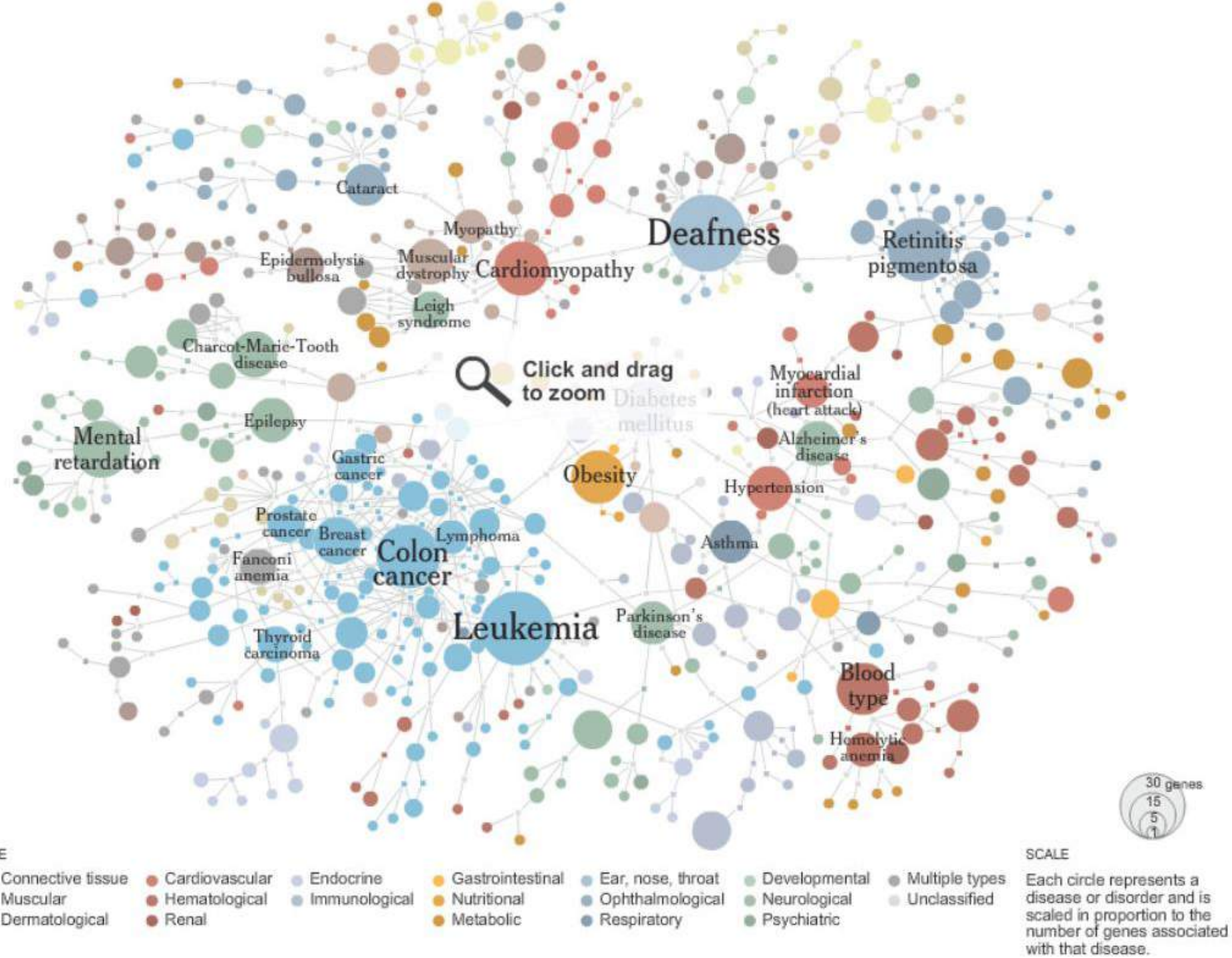
- Visual channels used?
 - E.g. channel X encodes attribute Y
- Marks used?
 - E.g. mark of type X encodes item Y
- Any suggested improvements?

<https://pudding.cool/2017/03/redraft/>



Exercise

- Visual channel
 - E.g. channel
- Marks used?
 - E.g. mark of
- Any suggestions?



https://archive.nytimes.com/screenshots/www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.jpg

Exercise

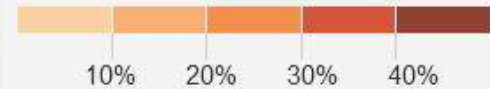
All Government Benefits >

ACCOUNTS FOR

17.6%

OF PERSONAL INCOME IN 2009

Government payments to individuals in more than 50 benefit programs, from food stamps to Medicare.



Social Security >

Medicare >

Medicaid >

Income Support >

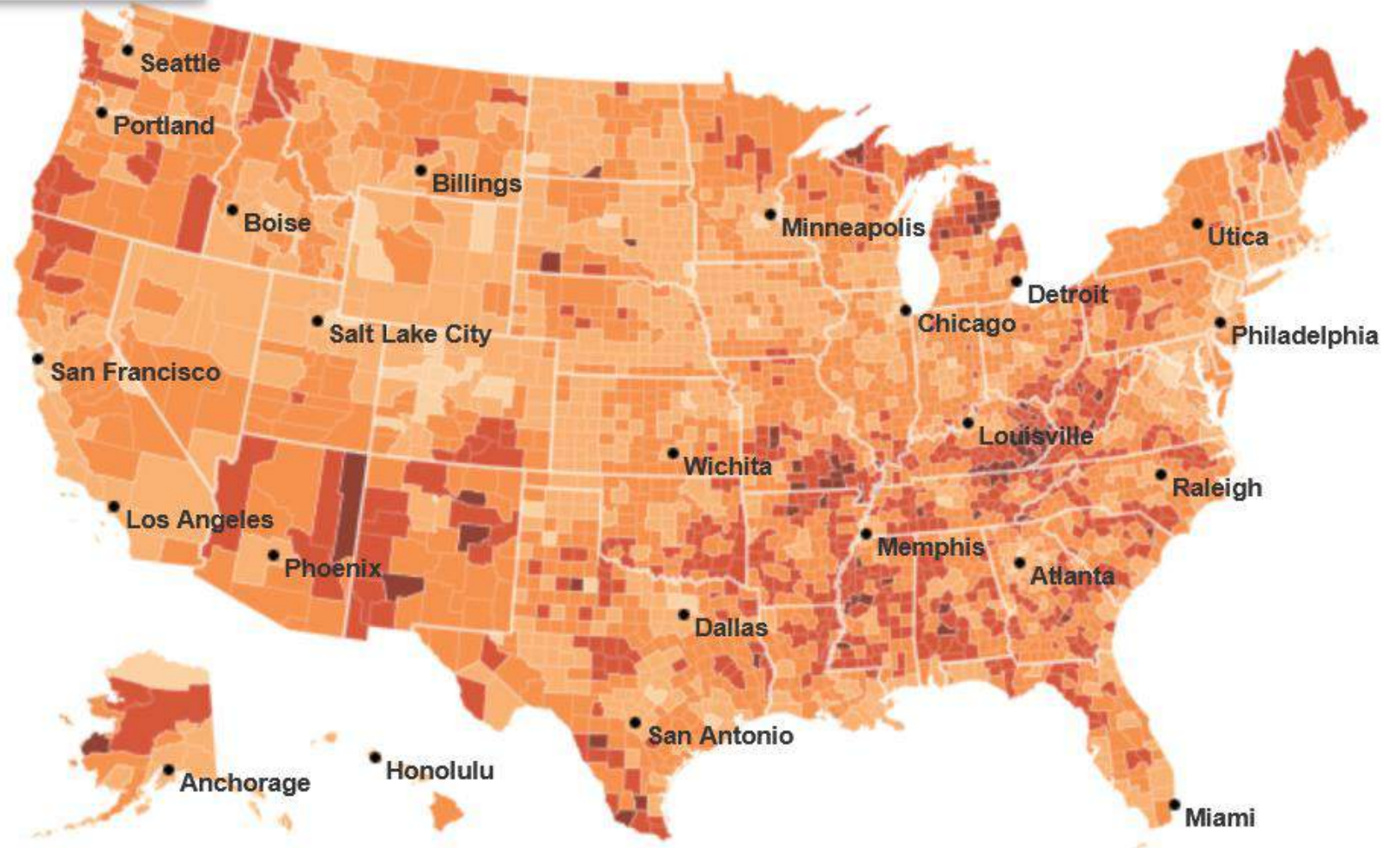
Veterans Benefits >

Unemployment Insurance >

GUIDE TO KEY TRENDS ▾

◀ 2009 ▶

Zoom to U.S.



<https://archive.nytimes.com/www.nytimes.com/interactive/2012/02/12/us/entitlement-map.html>