Visualization for Data Science

Color

Adminstrivia – Office Hours This week

Monday

- 5 - 7pm in Lecture Room Dr. K

Tuesday

- 2 - 4pm online and in ICCS 227 Dr. K
- 4 - 6pm in Lab Room Matt, Chi and William

Wednesday 2 - 3pm

- 1 - 3pm online Dr. K
- 5 - 7pm in Lecture Room Anna

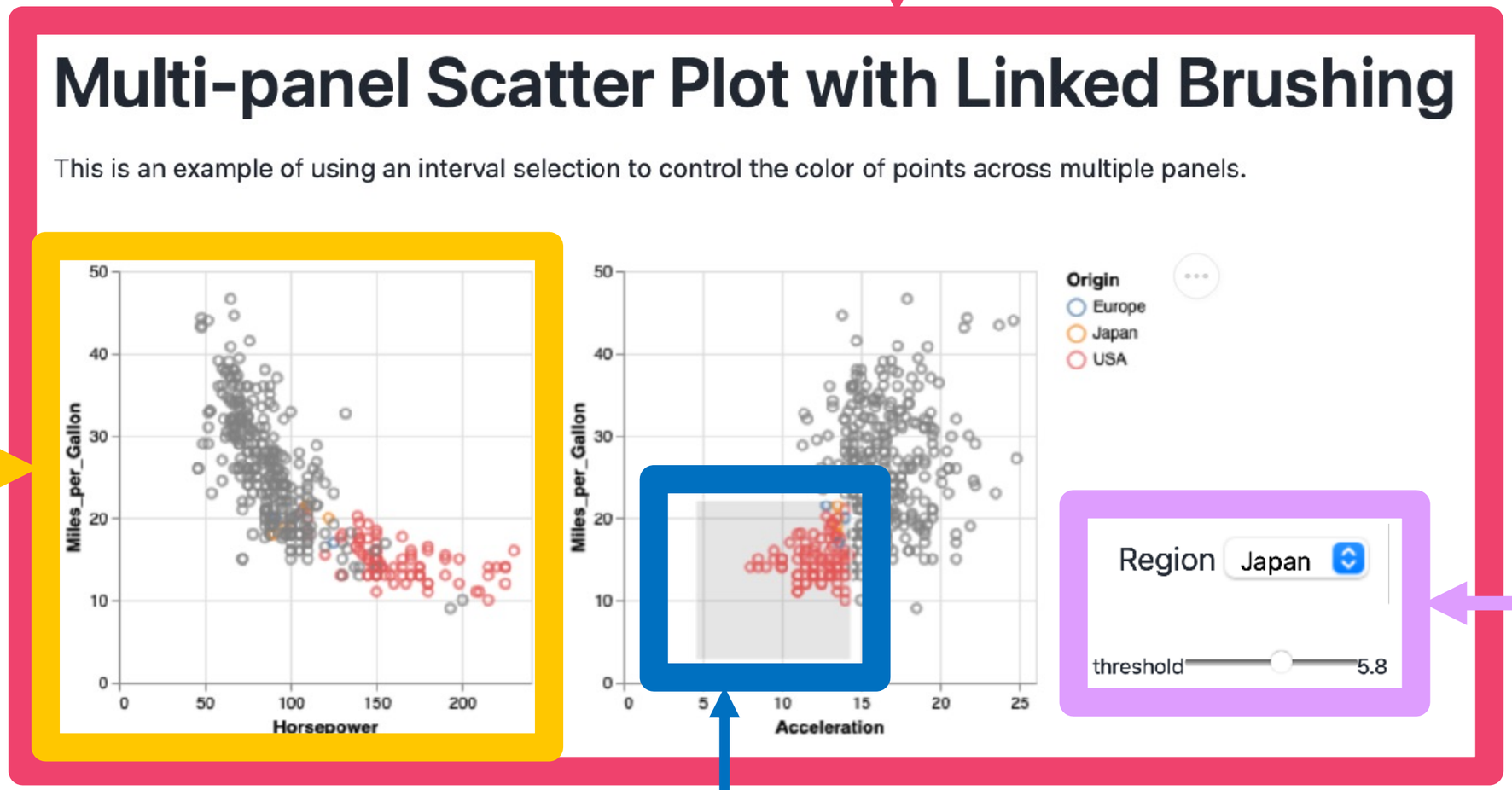
Thursday

- 3 - 6pm Lab Room Matt, Chi and William, Anna

Short Feedback Survey.
See Canvas (yup Canvas)
Lecture 13A Feedback

"View" or "Dashboard"

"Viz" or "Plot" or "Chart"



Widgets
Indirect
Manipulation
interaction

Advanced Direct Manipulation Interaction

All requirements listed below are per group member !
Each analytic question should be answered by one "view" !

	C	B	A	A+
Report	• Markdown file • Contribution summary	• Markdown file • Contribution summary • Significant insights & reflections • Reasonable workload distribution	• Markdown file • Contribution summary • Deep & thoughtful insights & reflections • Equitable workload distribution	• Publish viz on a website (dashboard) • Contribution summary • Deep & thoughtful insights & reflections • Equitable workload distribution
Visualizations	• 1 view • View has at least 2 altair visualizations (vizzes)	• 1 view • The view must have 3 vizzes	• 1 view (with 3 vizzes) • An additional view (with at least 2 vizzes) • Thoughful layout in views • At least 1 advanced, high-quality, high-effort visualizations. • >=5 vizzes in total	• 2 views (3 viz on 1, 2 viz on other) • An additional view (with at least 2 vizzes) • An additional advanced viz • >=7 vizzes in total
Interactivity	• 1 meaningful interaction (that is more than zoom, pan, tooltip, hover with no propagation)	• 1 indirect manipulation interaction (IMI) that is a UI widget (e.g., drop down, check box, radio button, slider, search box) • 1 advanced direct manipulation interaction (DMI) (e.g., brushing, details on demand)	• 2 IMI/UI widgets (unique) • 2 advanced direct manipulation interactions • 1 bi-directional interaction	• 3 IMI/UI widgets (unique) • 3 advanced DMI • 2 bi-directional interaction • Beyond bean counting - extra flair and effort that supports sensemaking.
Complexity	• Simple analytic tasks • Data wrangling	• Sophisticated analytic tasks • Additional wrangling and processing	• Deep and complex analytic tasks • Deep analysis and insightful commentary	• Deep and complex analytic tasks • Deep analysis and insightful commentary • Extra flair and effort
Cohesiveness	• Separate reports • Mini projects	• Single report • Coordinated styling	• Single report • Coordinated styling	• Single published website • Coordinated styling

Interaction

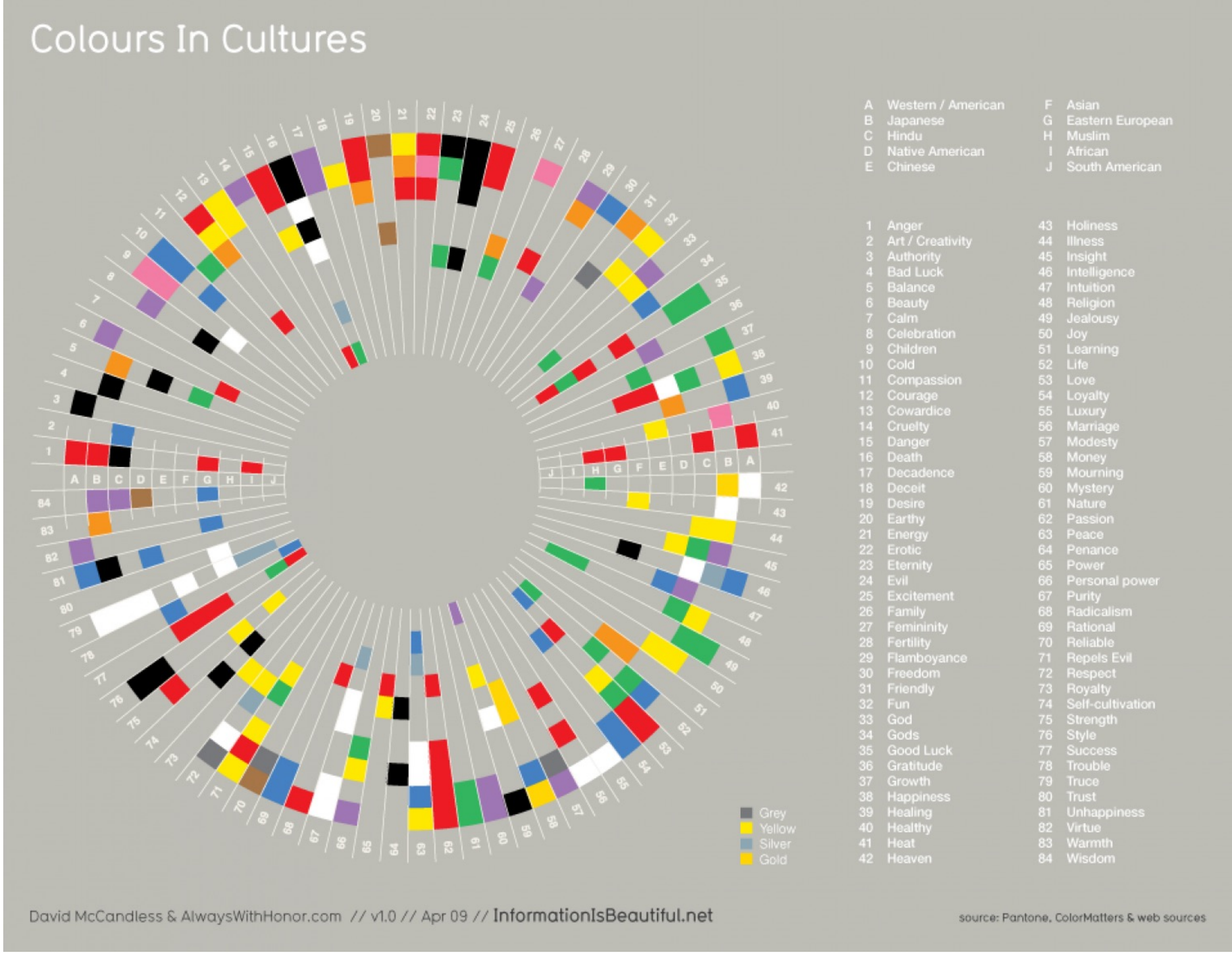
- Indirect Manipulation Interaction (Widgets)
  - Slider, Checkbox, Radio button, Dropdown menu, Search https://altair-viz.github.io/gallery/multiple_interactions.html
- Direct Manipulation Interaction
  - Trivial (non-meaningful): Panning, Standard Zooming, Tooltips of text. Meaningful Interaction: Hovering/highlight
  - Advanced:
    - Propagated Multi—Select that filters AND highlights
    - Details on Demand- for example. propagated Hovering: going beyond: https://altair-viz.github.io/gallery/scatter_point_paths_hover.html
    - Brushing: https://altair-viz.github.io/gallery/interval_selection_map_quakes.html
    - Interactive Calculation: https://altair-viz.github.io/gallery/selection_layer_bar_month.html OR https://altair-viz.github.io/gallery/interactive_column_selection.html (this second one is both interactive calculation and propagated multi-select
  - Bidirectional: https://altair-viz.github.io/gallery/scatter_with_layered_histogram.html

Learning Outcomes

- ~~List the main three color channels used in visualization~~
- Explain how color differences affect the user's perception of the visualization
- Detail strategies for making visualizations color accessible
- Describe the various types of color palettes
- Use examples to explain how light and background influence our perceptions
- Discuss how we can use color to effectively represent data attribute values

What does the color mean to you or to them?

https://blogs.sas.com/content/sastraining/2017/06/29/colors-in-different-cultures/



P: Ordering Colors?

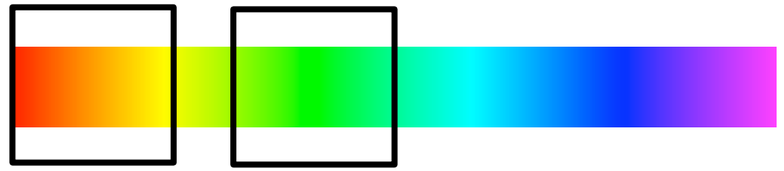
Which ordering of colors is best?

A. Red Blue Green Yellow Purple
B. Red Yellow Green Purple Blue
C. Purple Blue Green Yellow Red
D. Green Red Blue Purple Yellow
E. Yellow Red Green Blue Purple

Ordered color: Rainbow is poor default

Perceptually unordered

- It's unclear how hues relate: e.g. is red **greater** than purple, is green **greater** than blue?.
- There is no common agreed upon (intrinsic conceptualization) of each hue's value

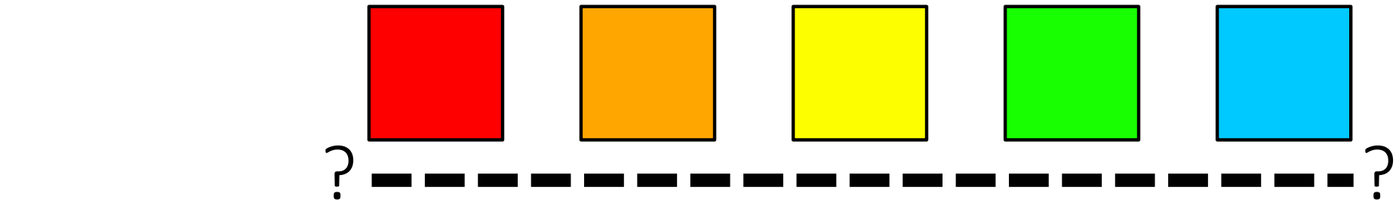


Perceptually nonlinear

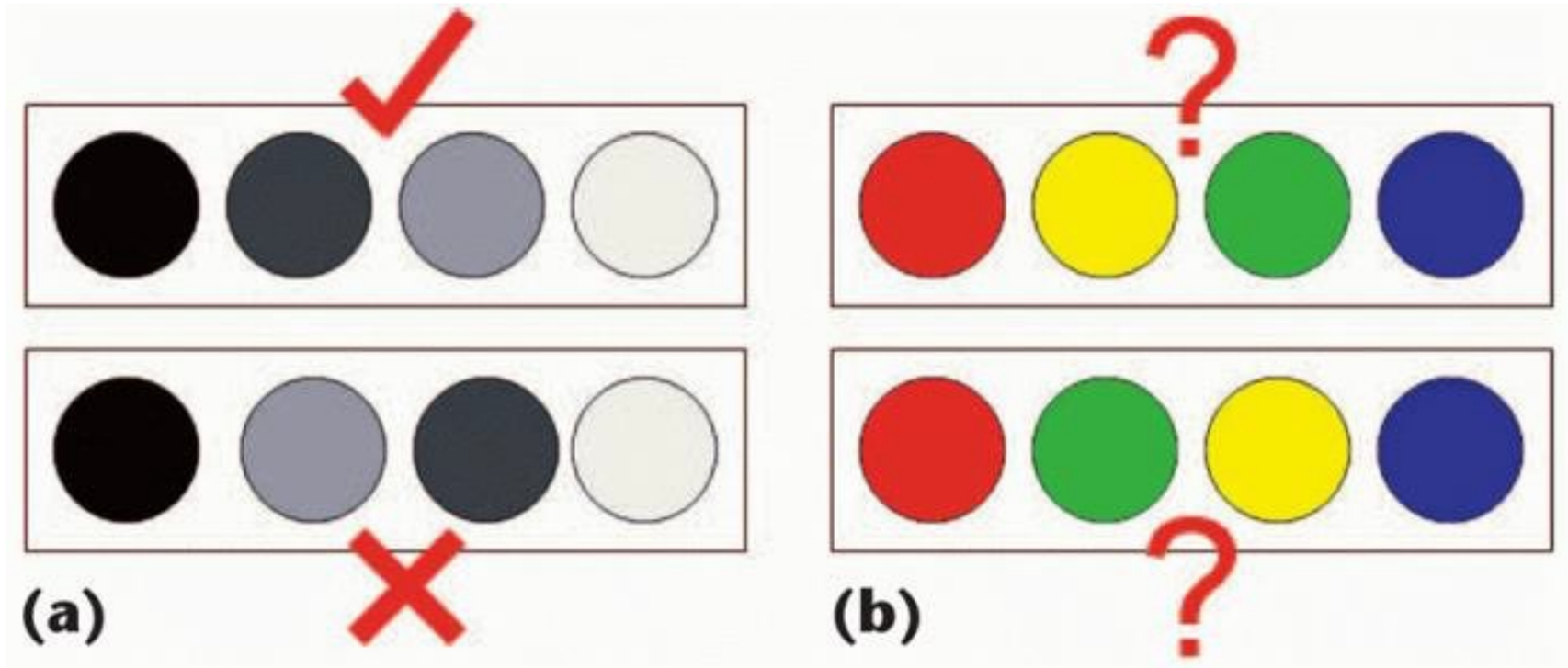
- Colors do not occupy equal perceptual space: the span of "red, orange and yellow" covers more perceptual variation than the large region occupied by "green".

Why this color map is a poor choice for quantitative data…

- No consistent perceptual ordering (confusing)
- Minimal variation in lightness (obscures details)
- Sharp color transitions imply sharp changes in data even where none exists (misleading)



No perceptual ordering (confusing)

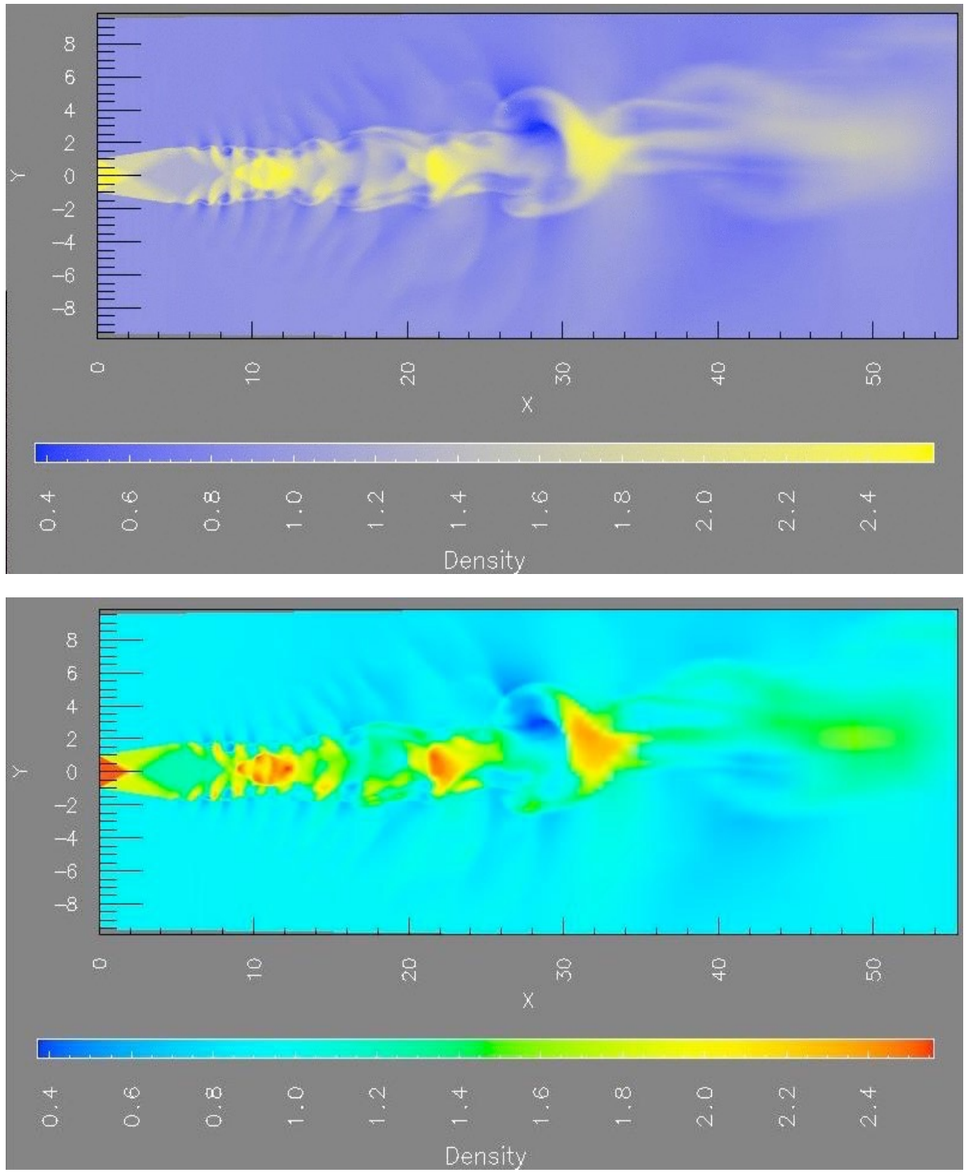


Borland & Russell, 2007

Ordered color: Rainbow is poor default

Benefits

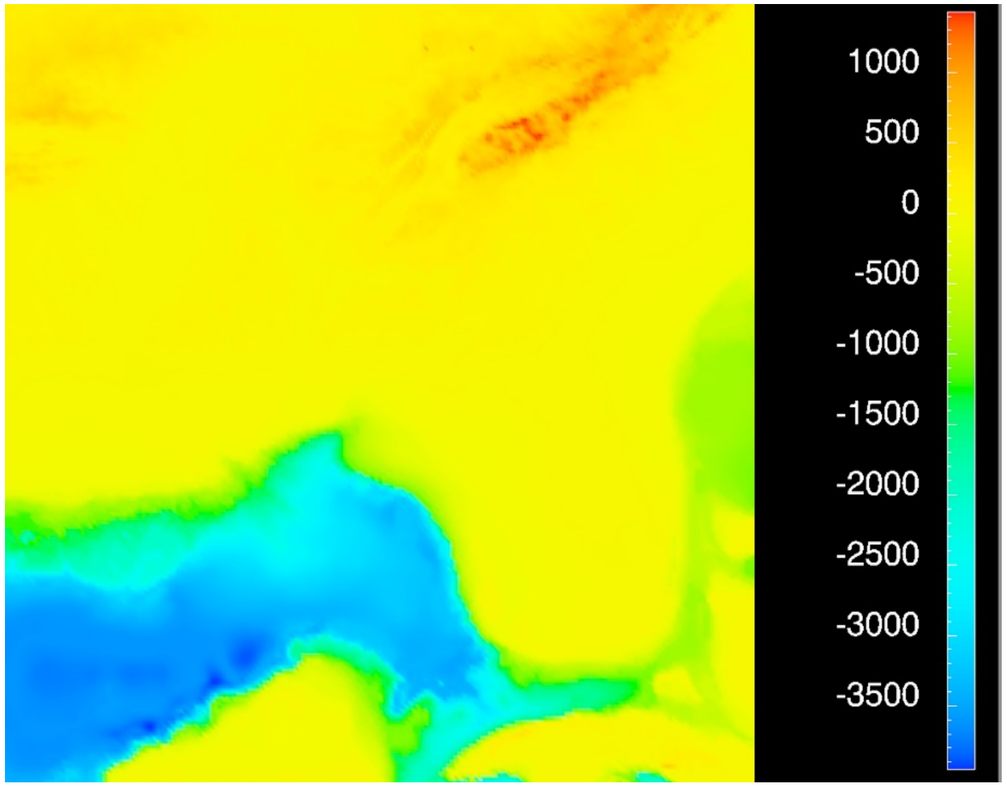
- Having a wide range of colors allows us to emphasize different aspects of a structure
- For instance, in the top image, we use 3 hues and can get a sense of the large-scale structure
- In the lower image, we use more hues and fine-grained structure are visible and nameable

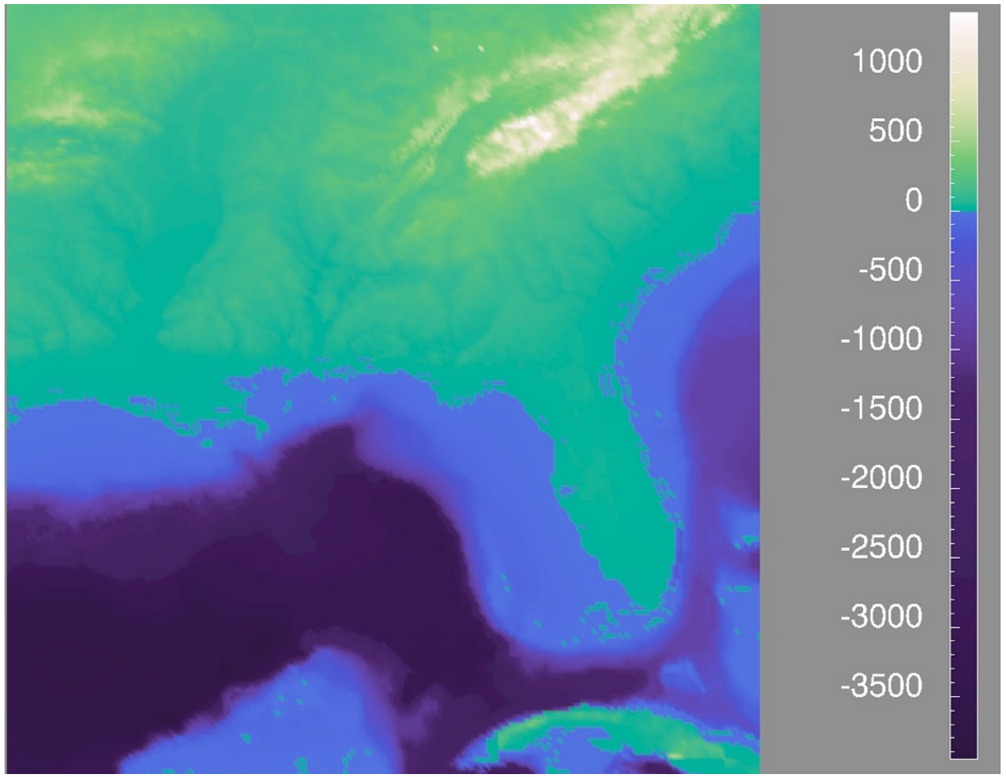


[A Rule-based Tool for Assisting Colormap Selection. Bergman,. Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

Ordered color: Rainbow is poor default

Effectiveness is at the discretion of the designer



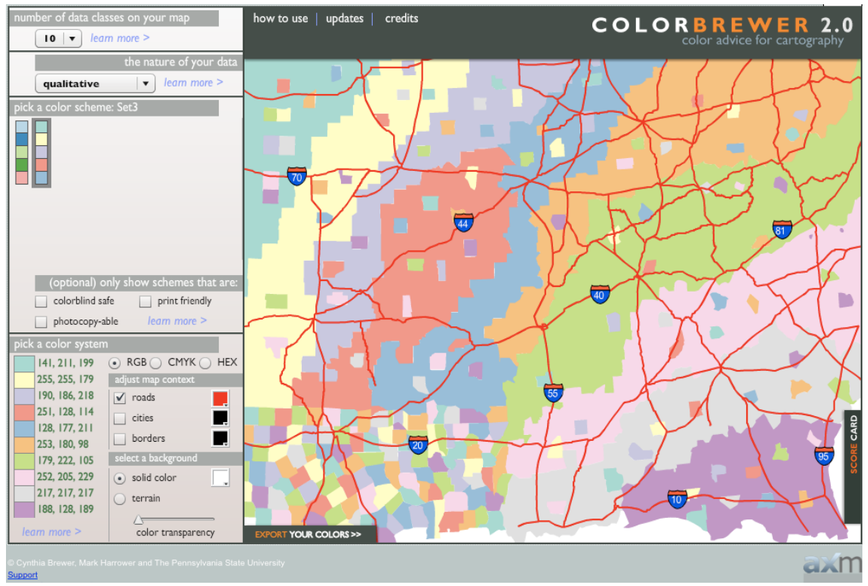


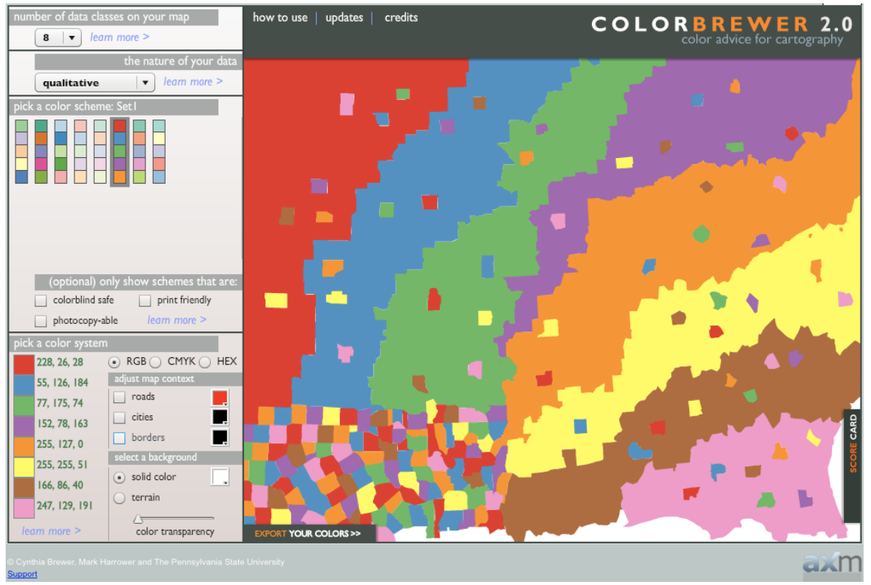
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Interaction between channels: Not fully separable

color channel interactions

- size heavily affects salience
- small regions need high saturation
- large regions need low saturation



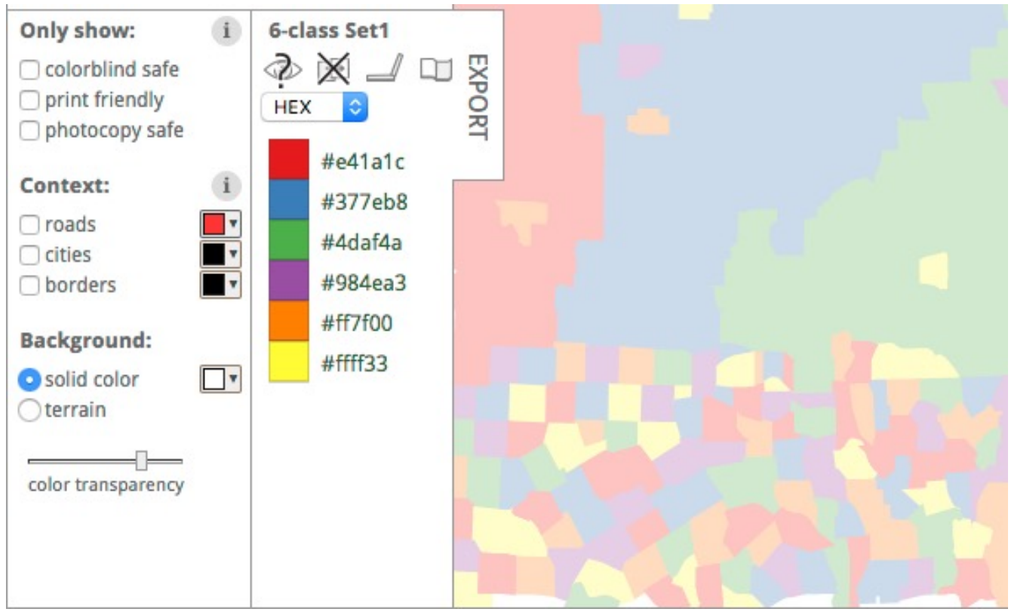


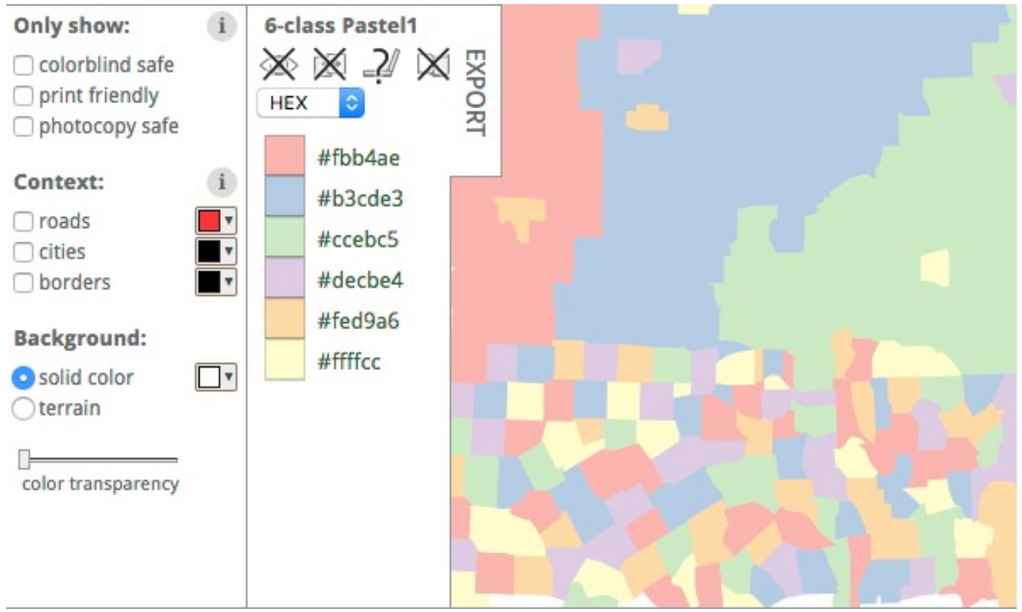
http://colorbrewer2.org/

Interaction between channels: Not fully separable

saturation & luminance:

- not separable from each other!
- also not separable from transparency
- small separated regions: 2 bins safest (use only one of these channels), 3-4 bins max
- contiguous regions: many bins (use only one of these channels)



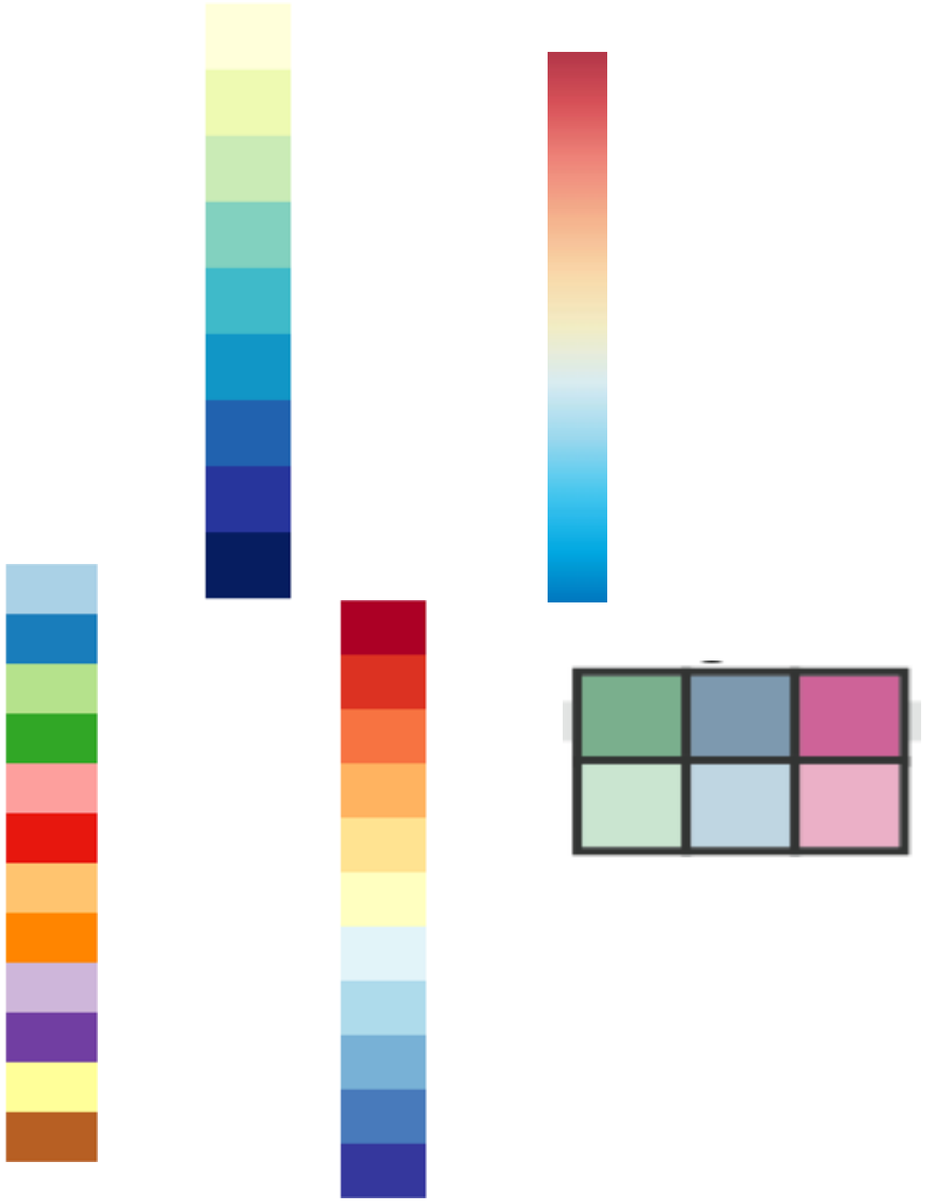


http://colorbrewer2.org/

What is a colormap/ color palette?

specifies a mapping between color and values

- categorical vs ordered
- sequential vs diverging
- segmented vs continuous
- univariate vs bivariate

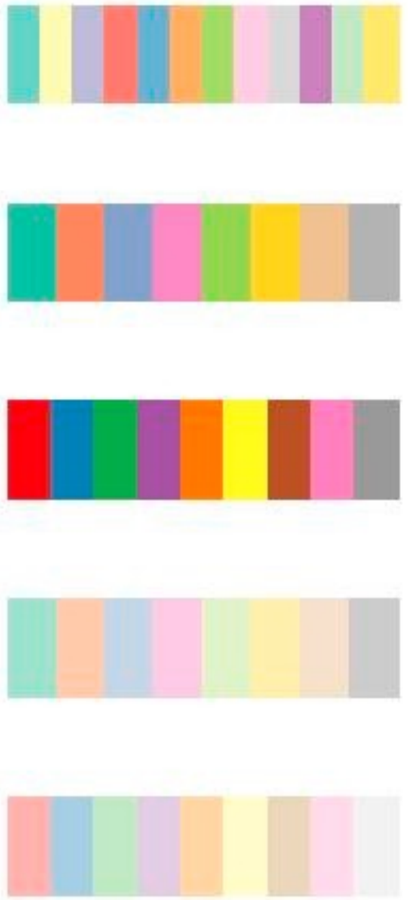


Design Tip: Match colormap to attribute characteristics!

Color palettes: univariate

➔ Categorical

- aim for maximum distinguishability
- aka *qualitative*, *nominal*

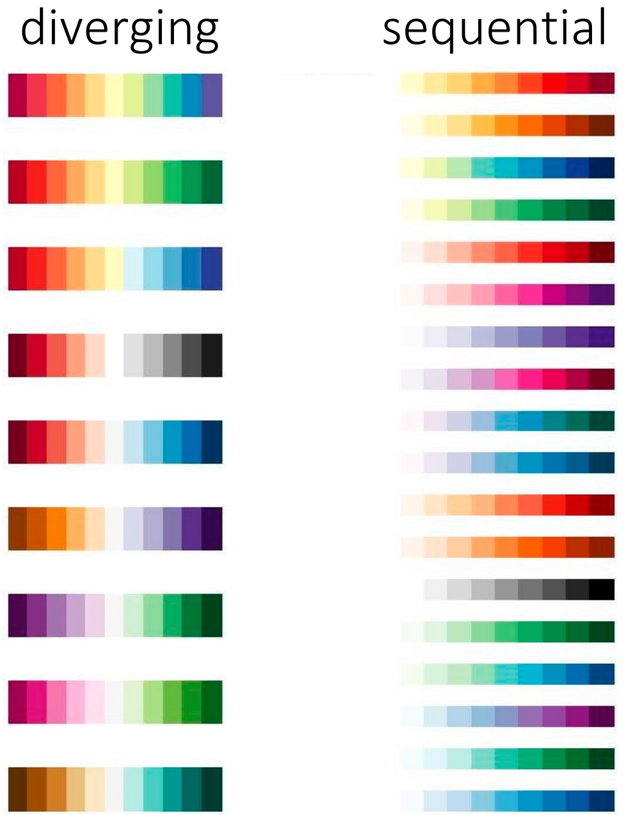


categorical

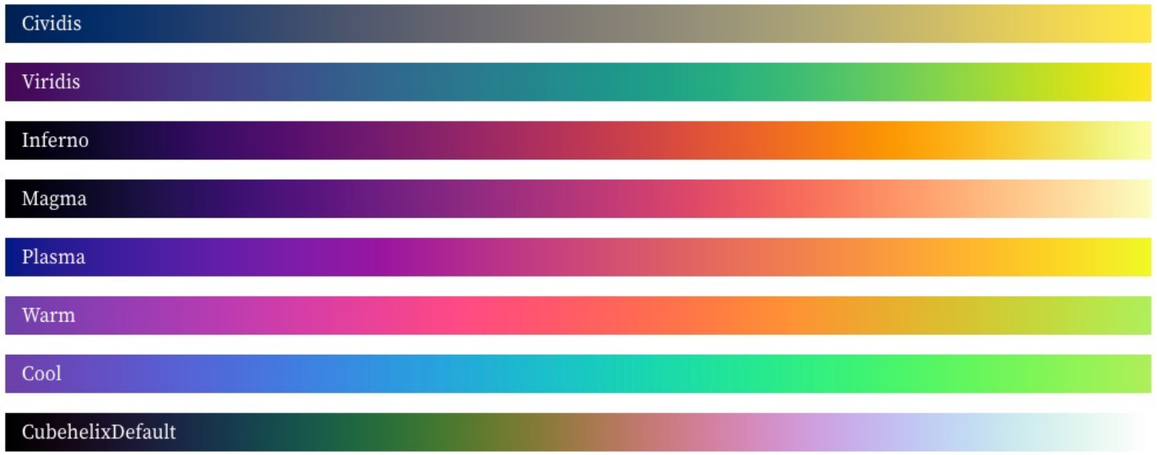
after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Color palettes: univariate

- ➔ Ordered
  - ➔ *Sequential*
  - ➔ *Diverging*



- diverging
  - useful when data has meaningful "midpoint
  - use **neutral color for midpoint**
    - white, yellow, grey
  - use saturated colors for endpoints
- sequential
  - ramp luminance or saturation
  - if multi-hue, good to order by luminance

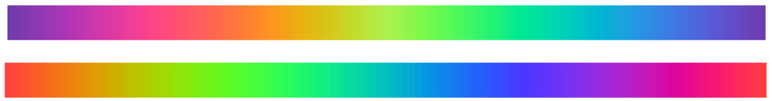


after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Color palettes: univariate

➔ Cyclic

cyclic multihue



after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Color map/palette design considerations

- segmented or continuous?
- diverging or sequential or cyclic?
- single-hue or two-hue or multi-hue?
- perceptually linear?
- ordered by luminance?
- colorblind safe?

Application: Example

Categorical

Does not imply magnitude differences (categorical/nominal data)

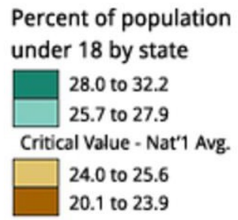
Distinct hues with similar emphasis

Sequential

Best for ordered data that progresses from low to high (ordinal, quantitative data)

Darkness (lightness) channel effectively employed

Diverging

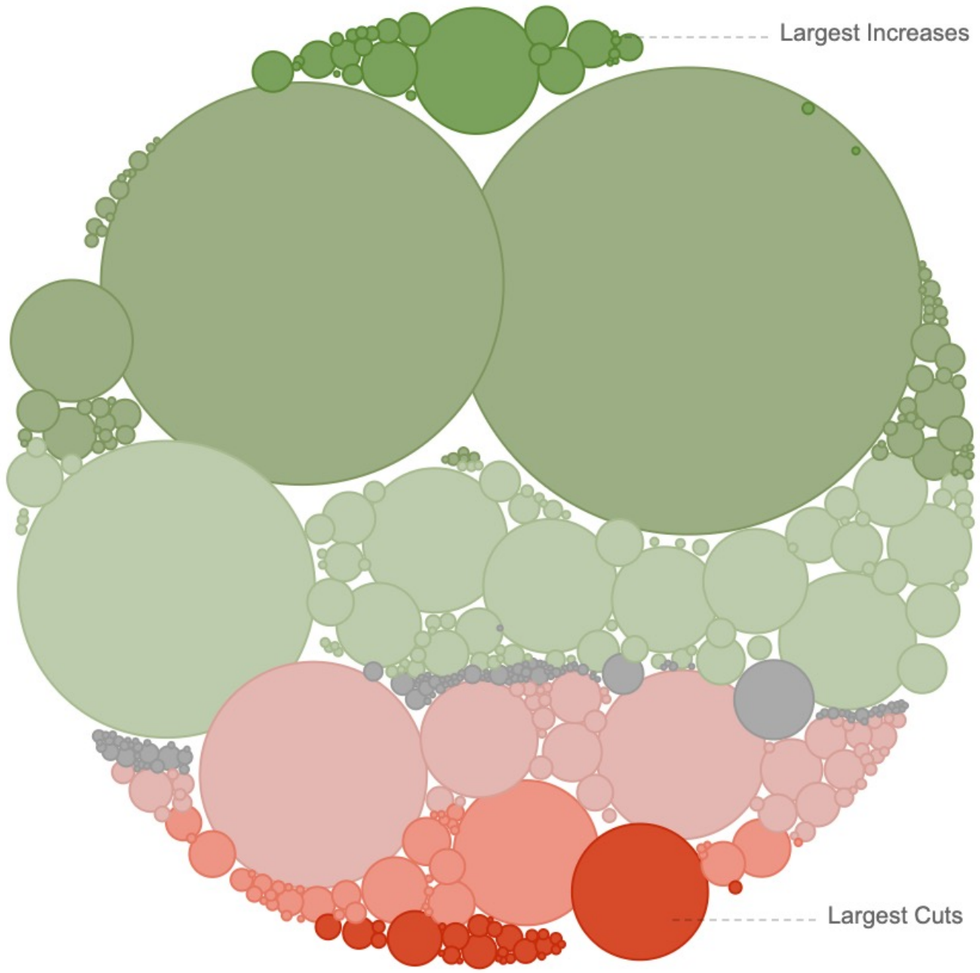


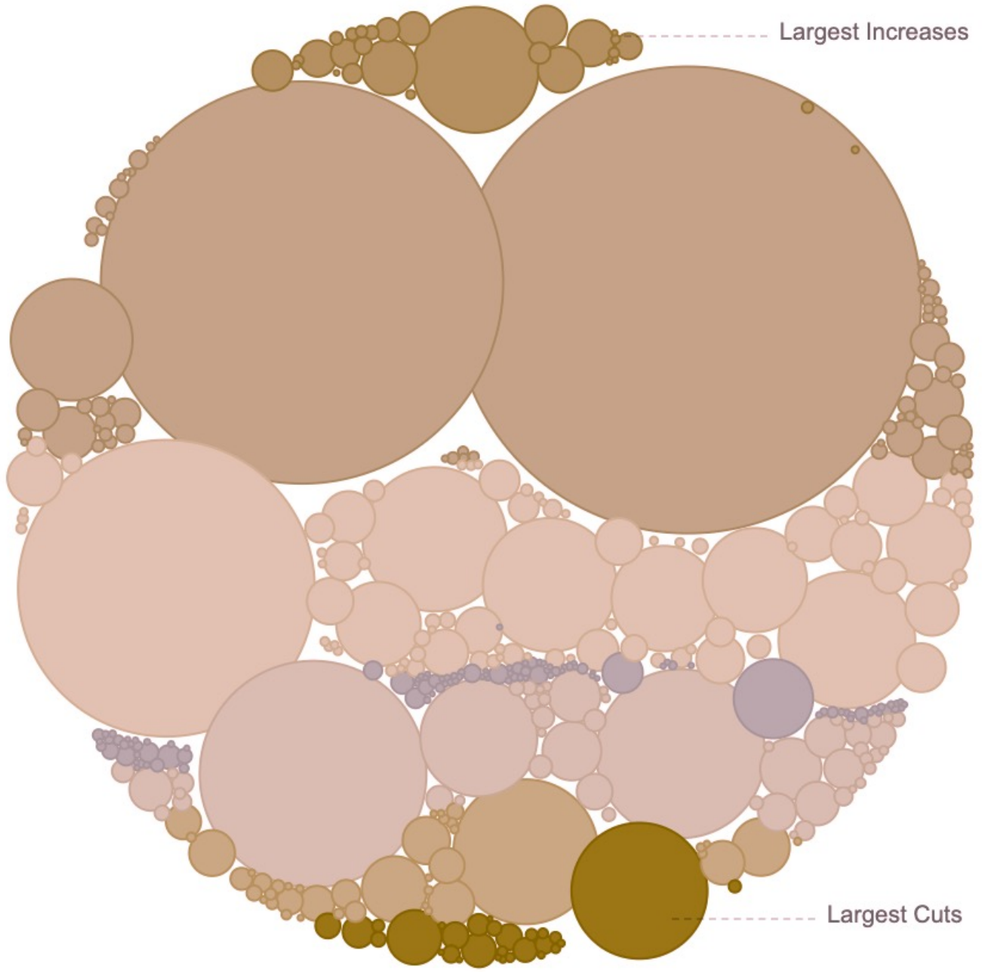
For data with a “diverging” (mid) point (quantitative data)

Equal emphasis on mid-range critical values and extremes at both ends of the data range

Brewer, 1994

What we see, what others see





https://pilestone.com/pages/color-blindness-simulator?

Luminance

need luminance for edge detection

- fine-grained detail only visible through luminance cont
- legible text requires luminance contrast!

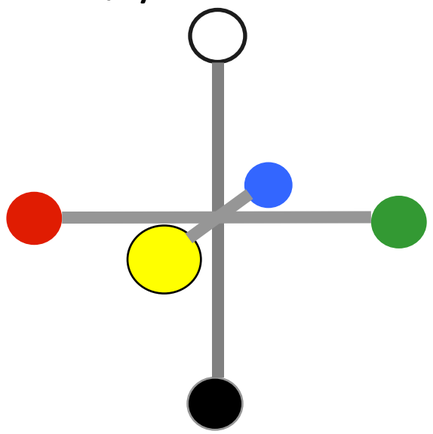
Luminance information

Saturation/hue information

[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

Opponent color and color deficiency

- perceptual processing before optic nerve
  - one achromatic luminance channel (L*)
    - edge detection through luminance contrast
  - 2 chroma channels
    - red-green (a*) & yellow-blue axis (b*)
- "colorblind": degraded acuity, one axis
  - ~10% of men are red/green color deficient
  - blue/yellow is rare

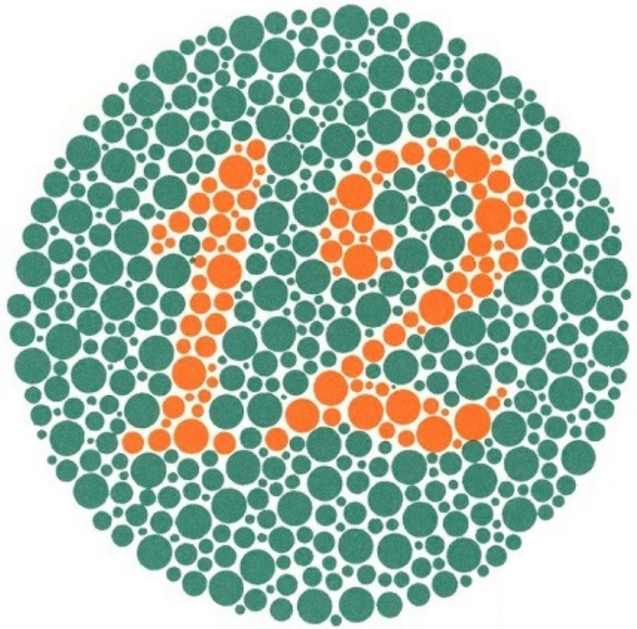


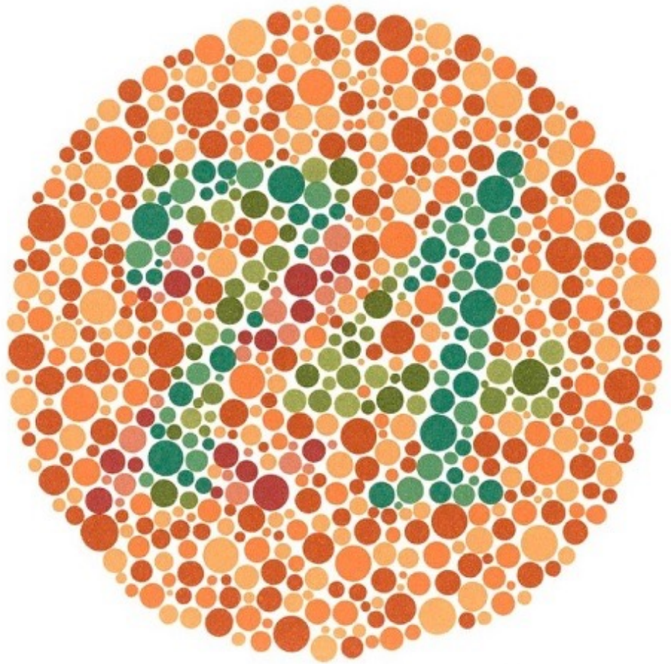
Luminance information

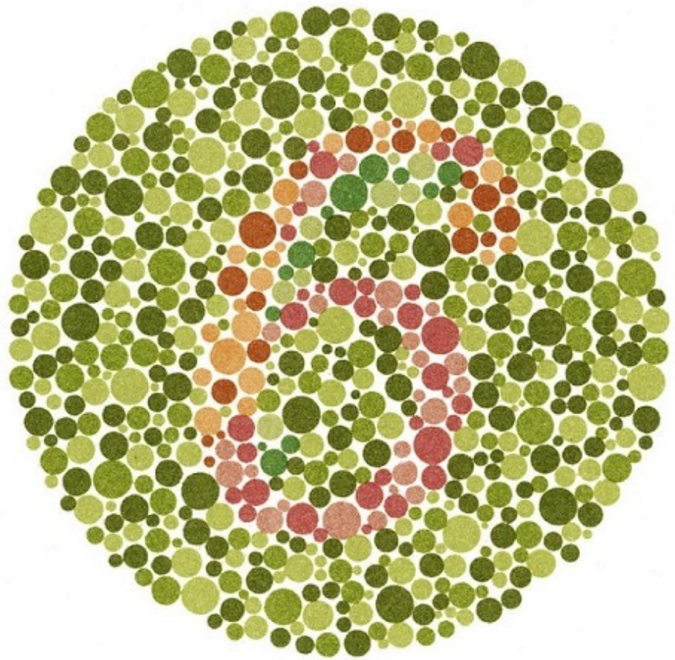
Chroma information

[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

Color Deficiency Tests: Ishihara Test







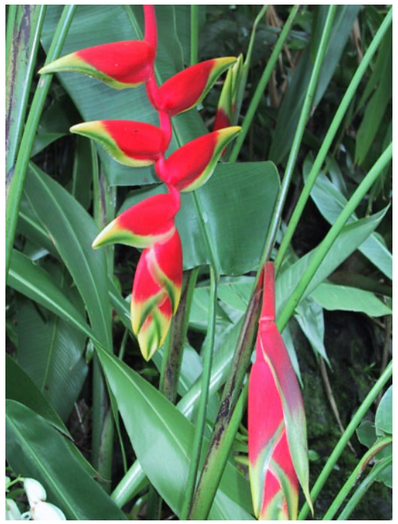
https://www.colorlitelens.com/ishihara-test.html

Color Deficiency

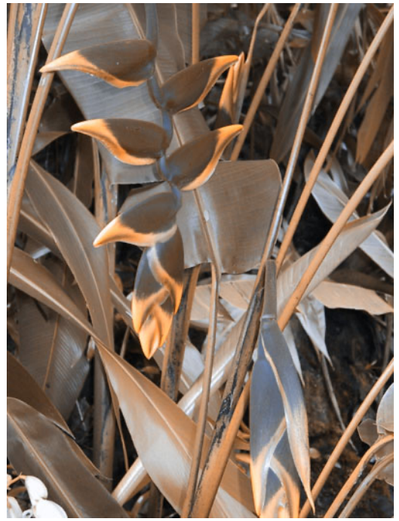
- ~10% of males, 1% of females (probably due to X- chromosomal recessive inheritance)
- Most common: red-green weakness / blindness
- Reason: lack of medium or long wavelength receptors, or altered spectral sensitivity (most common: green shift)

Normal Color Perception | Deuteranopia (no green receptors) | Protanopia (no red receptors)

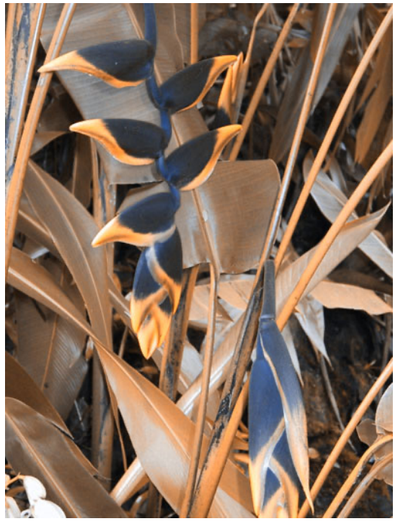
Designing for color deficiency: Check with simulator



Normal vision

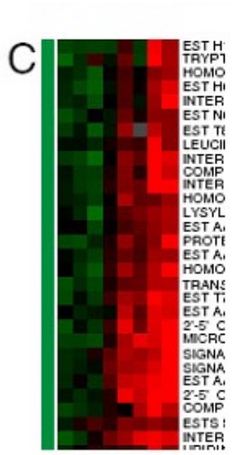


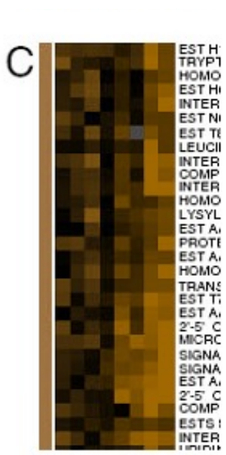
Deuteranope
green-weak



Protanope
red-weak

Tritanope
blue-weak



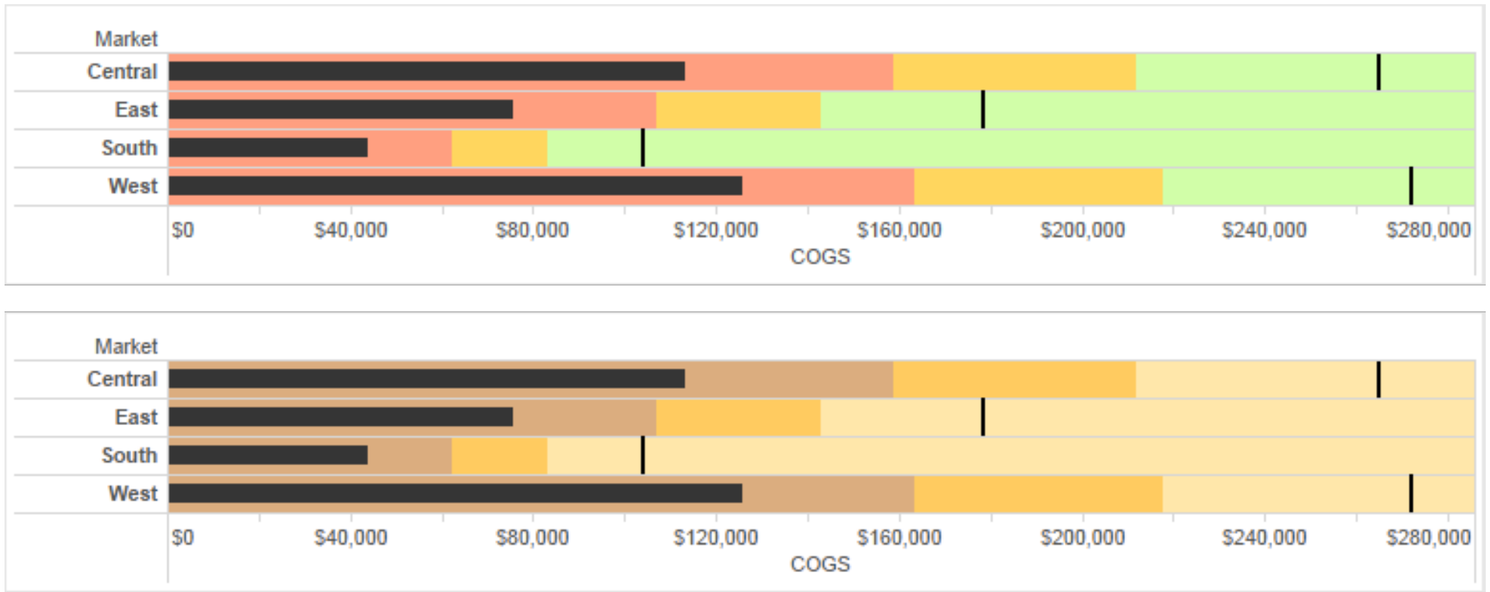


[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

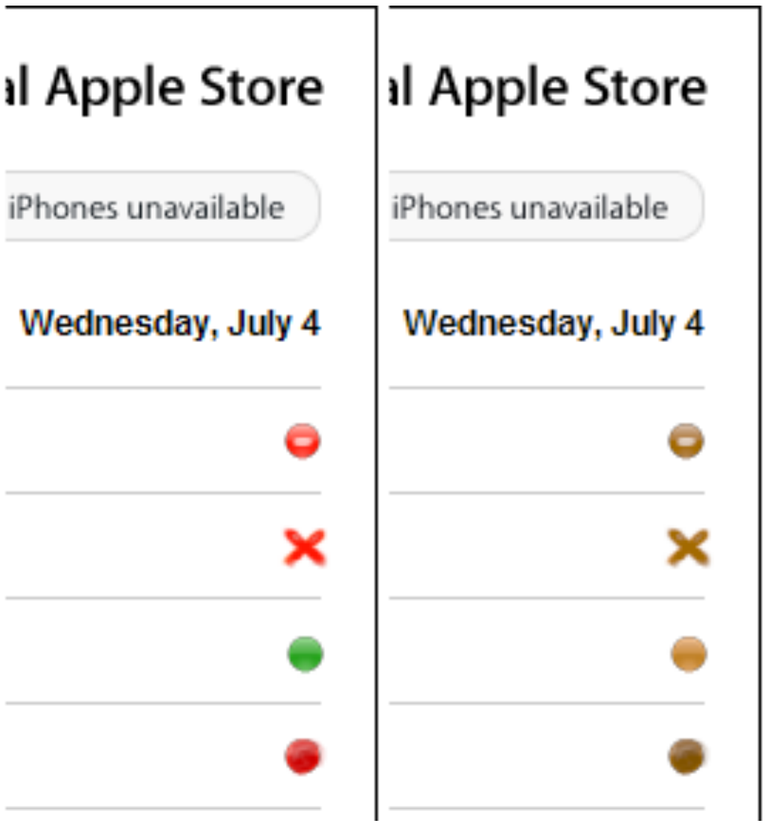
https://www.color-blindness.com/coblis-color-blindness-simulator/

Designing for color deficiency: Avoid encoding by hue alone

- redundantly encode
  - vary luminance
  - change shape



Deuteranope simulation

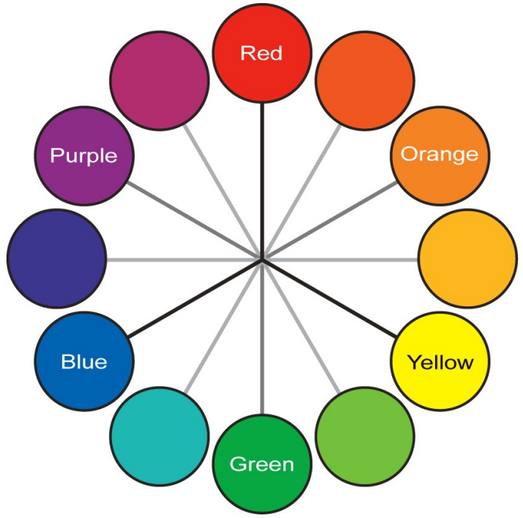


Change the shape

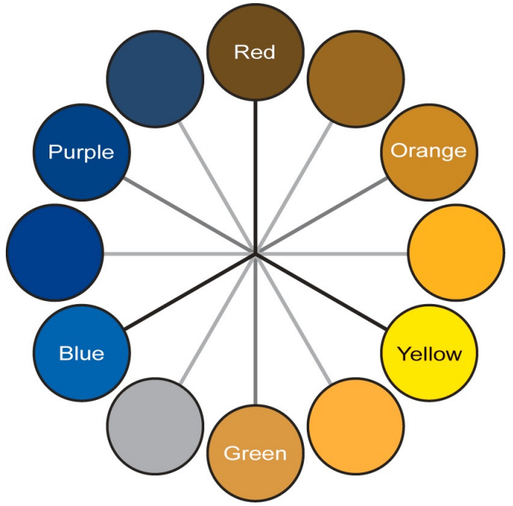
Vary luminance

[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

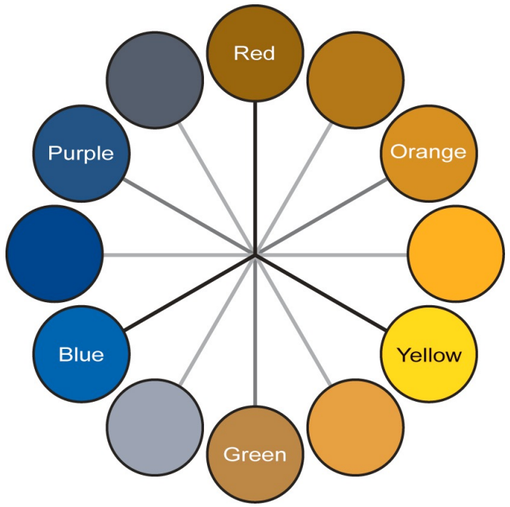
Color deficiency: Reduces color to 2 dimensions



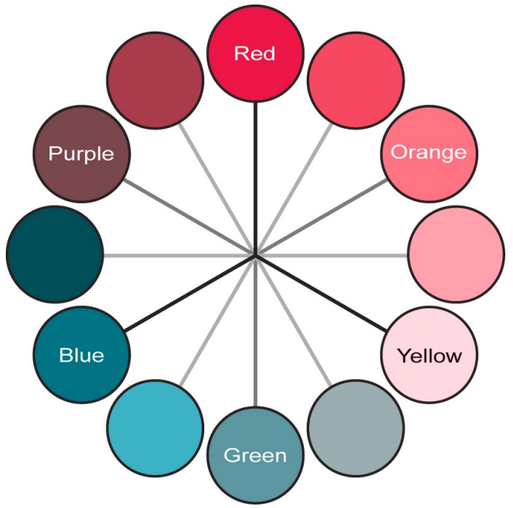
Normal



Protanope



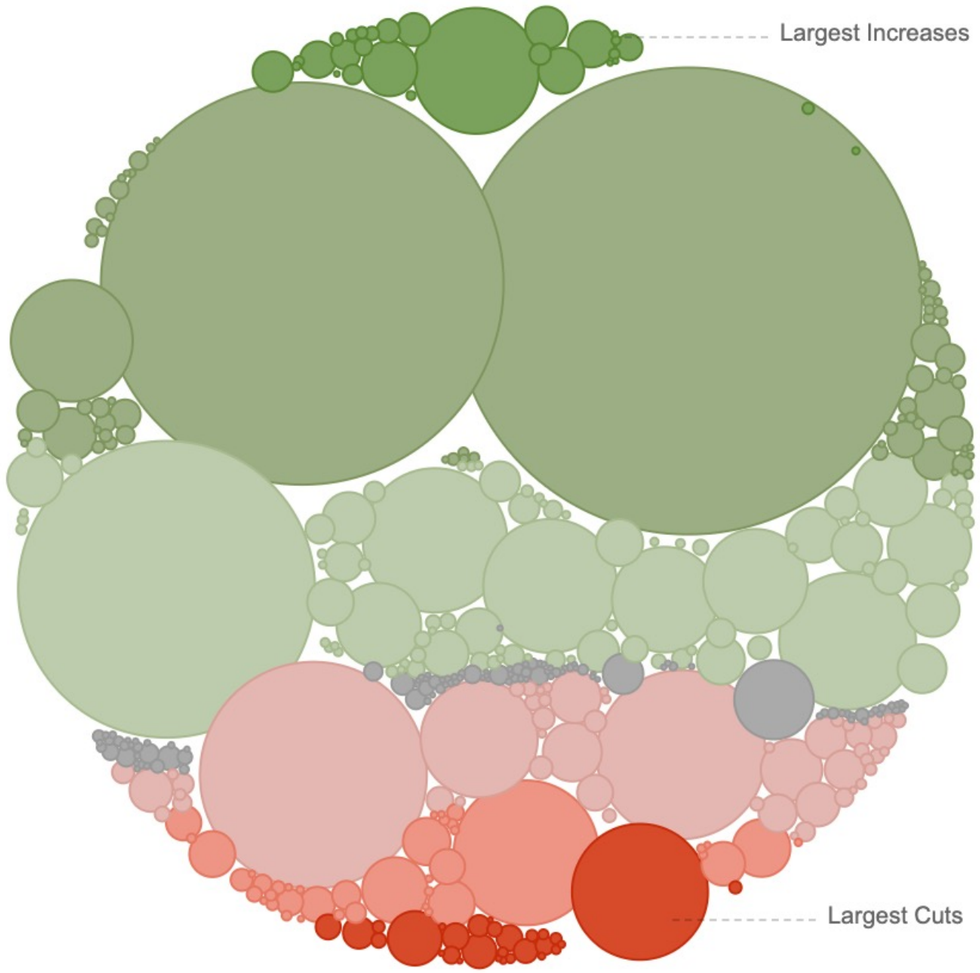
Deuteranope

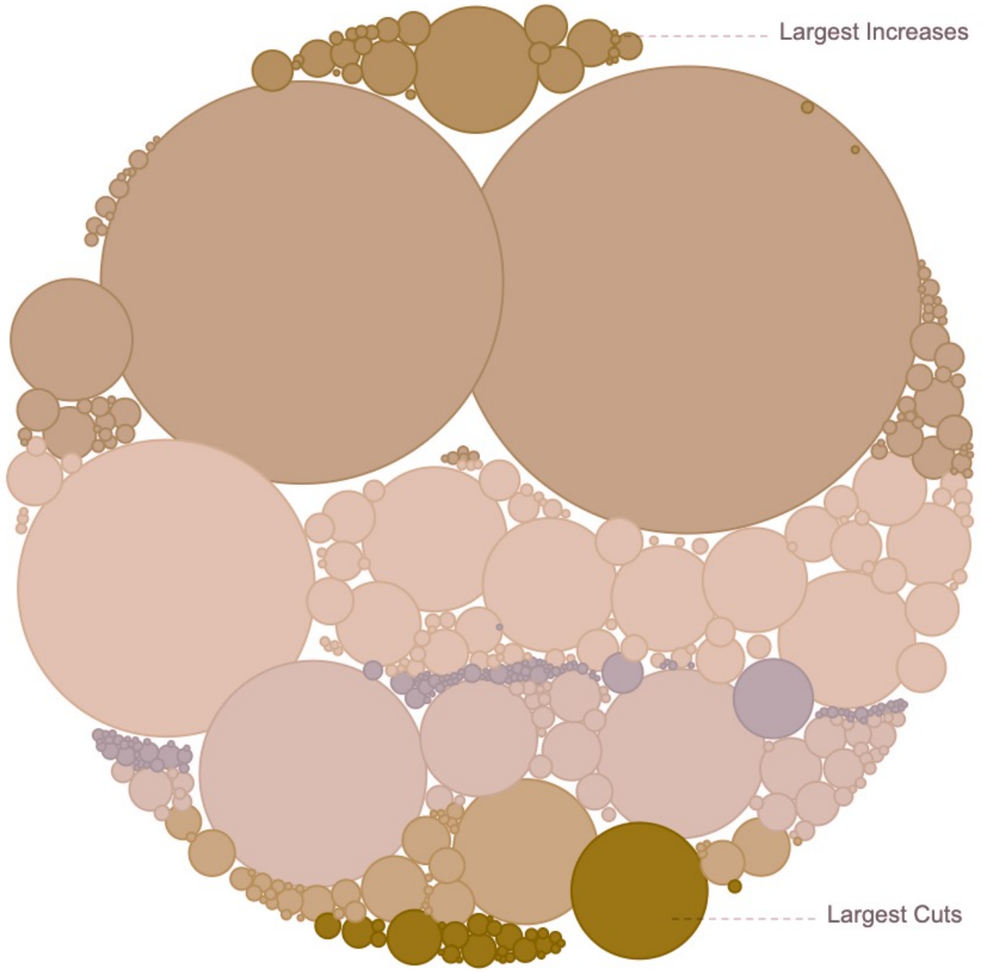


Tritanope

[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

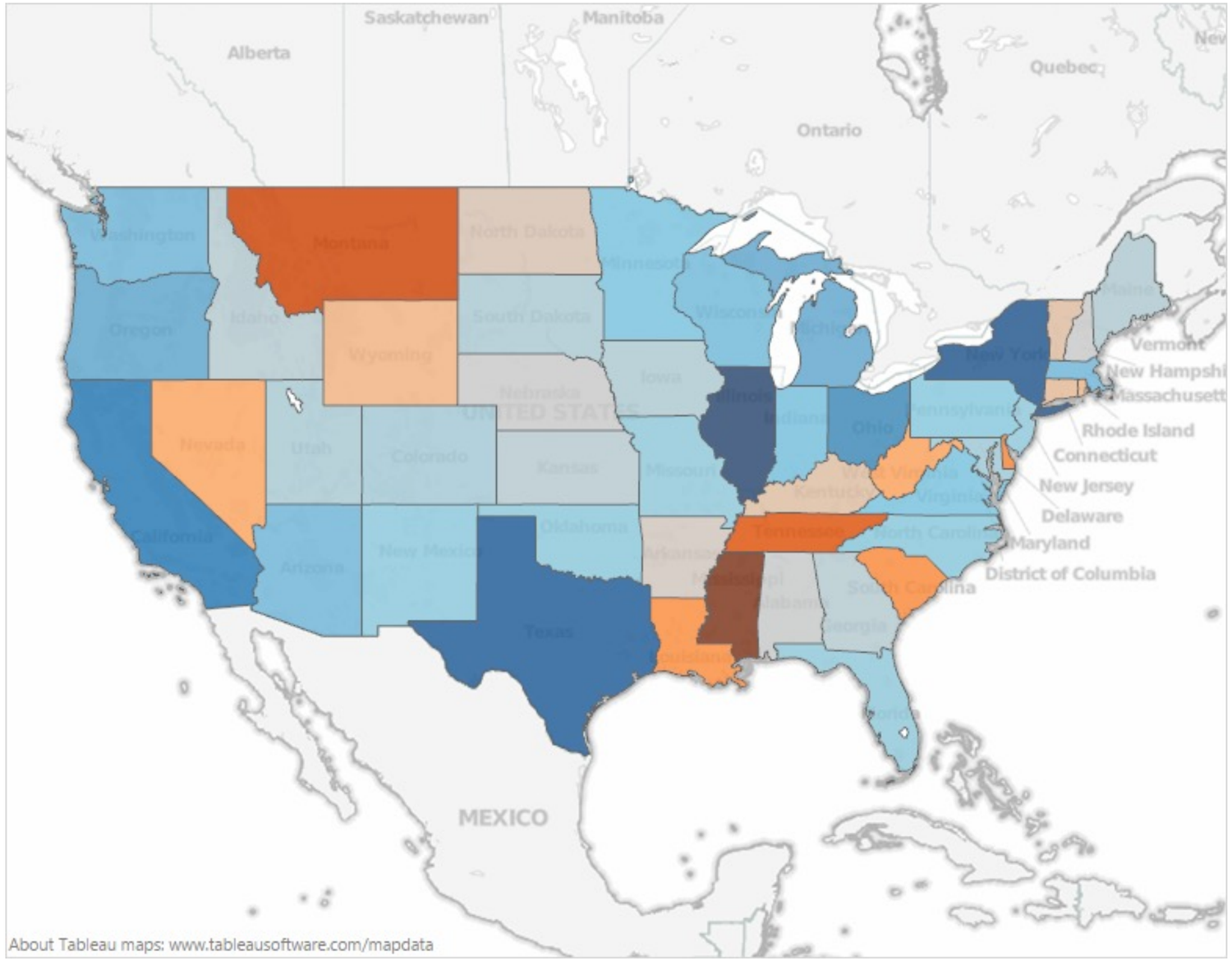
What we see, what others see





https://pilestone.com/pages/color-blindness-simulator?

Designing for color deficiency: Blue-Orange is safe



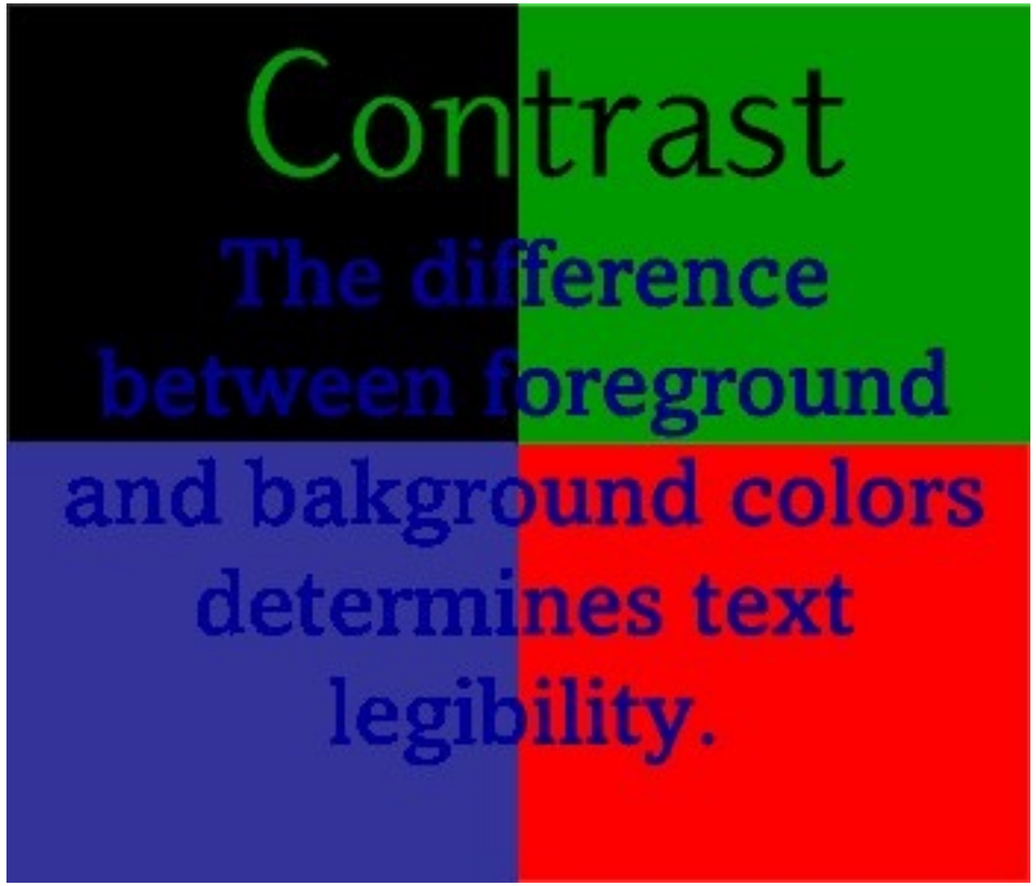
[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

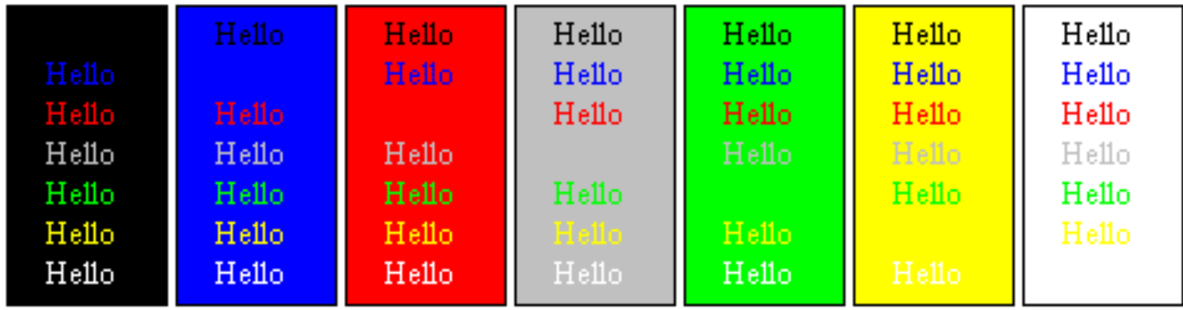
Luminance Contrast

When adding text to a visualization make sure that there is a distinct difference between the foreground and the background.

Luminance Contrast is Critical

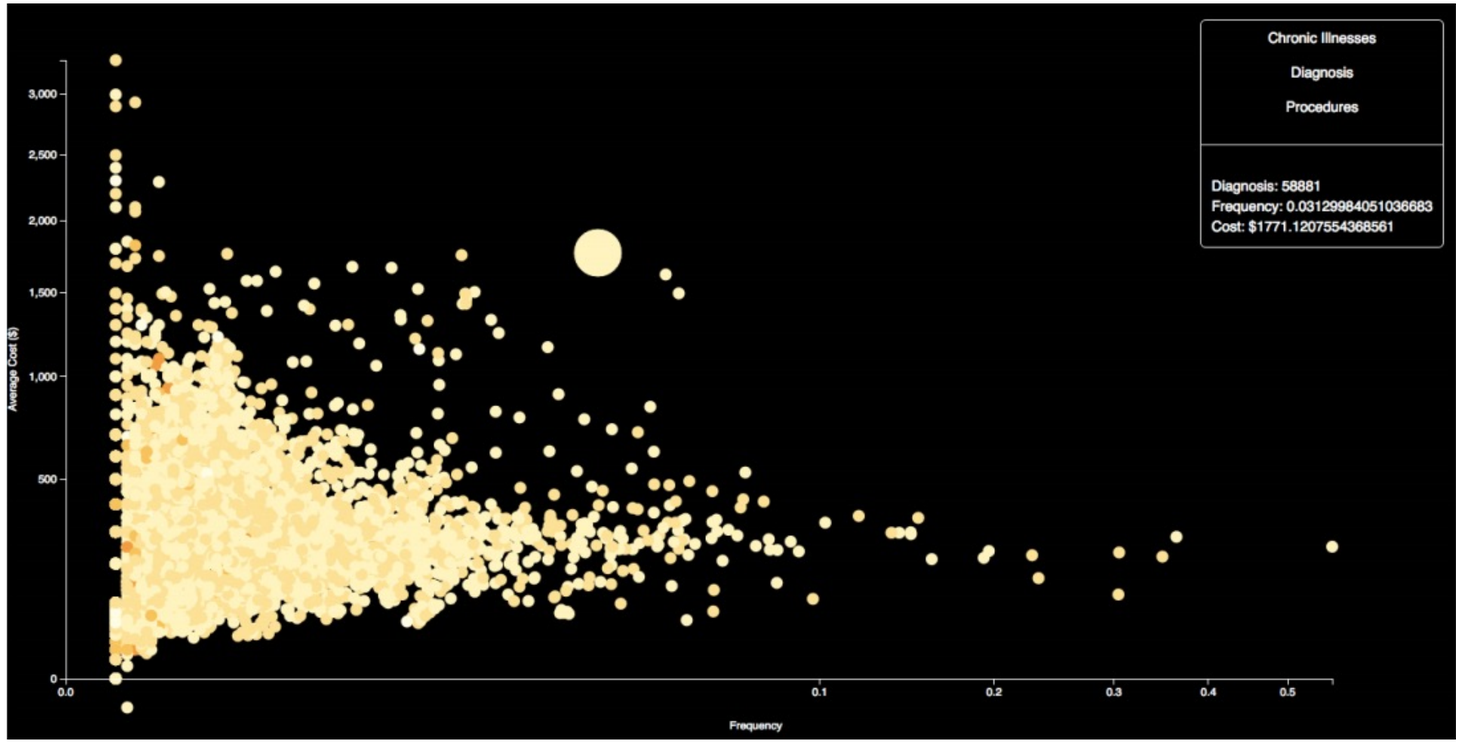
When you try to use Hue difference for edge detection it is hard to perceive





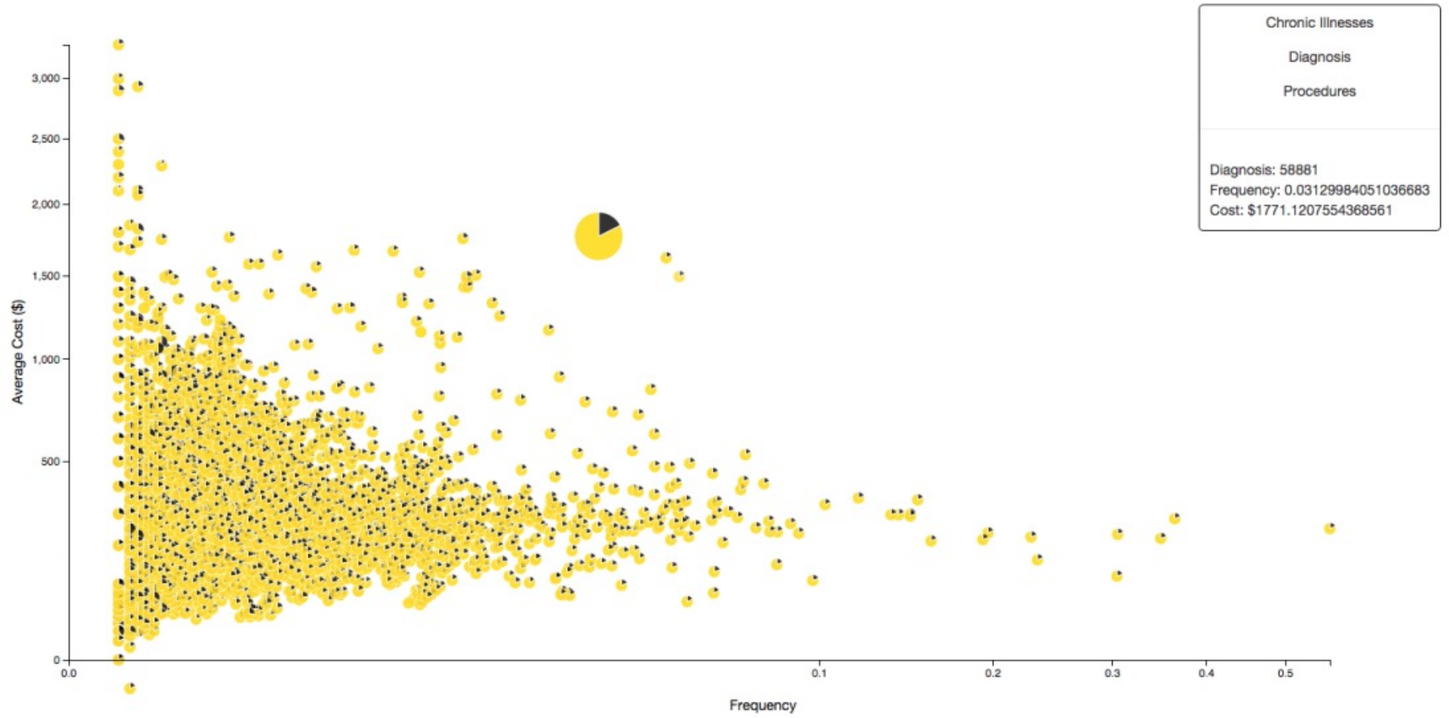
Interaction with the background: tweaking yellow for visibility

marks with high luminance on a background with low luminance



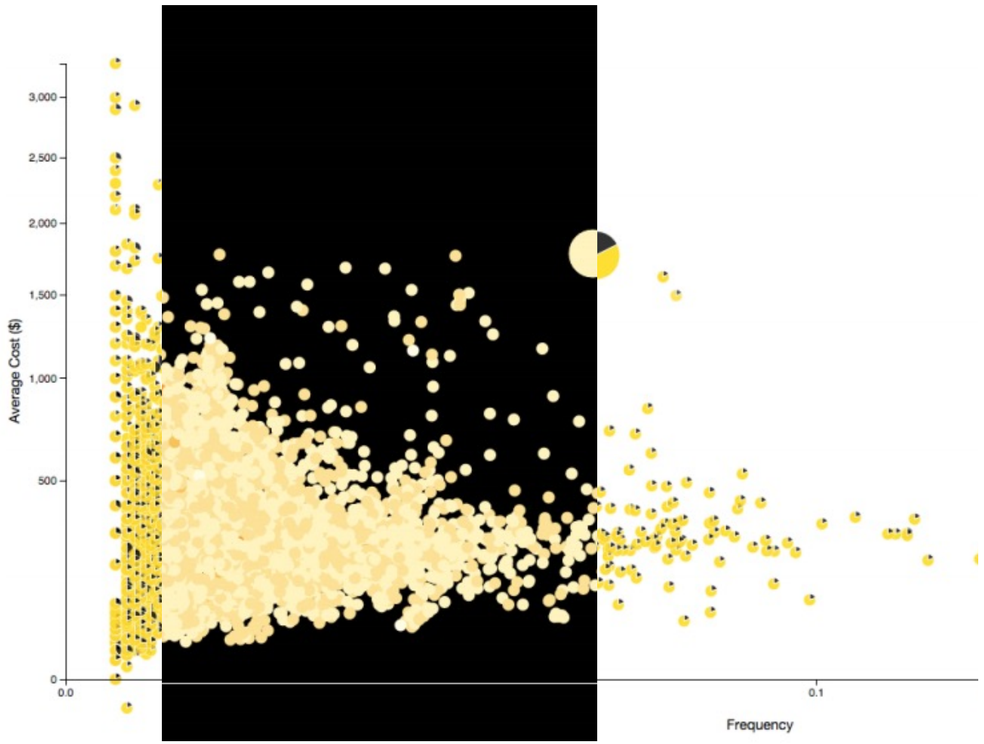
Interaction with the background: tweaking yellow for visibility

marks with medium luminance on a background with high luminance



Interaction with the background: tweaking yellow for visibility

change luminance of marks depending on background



Color/Lightness constancy: Illumination conditions

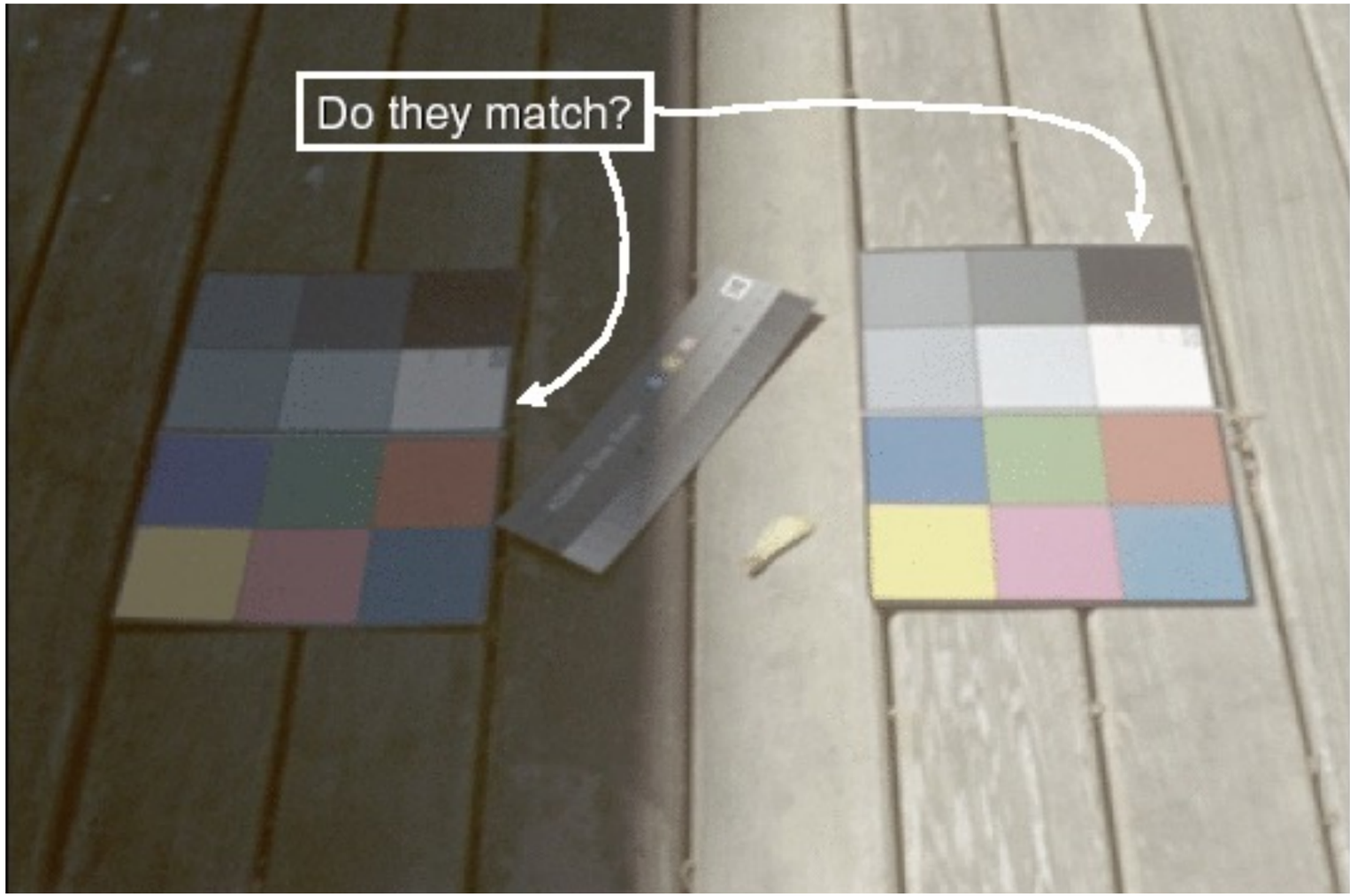


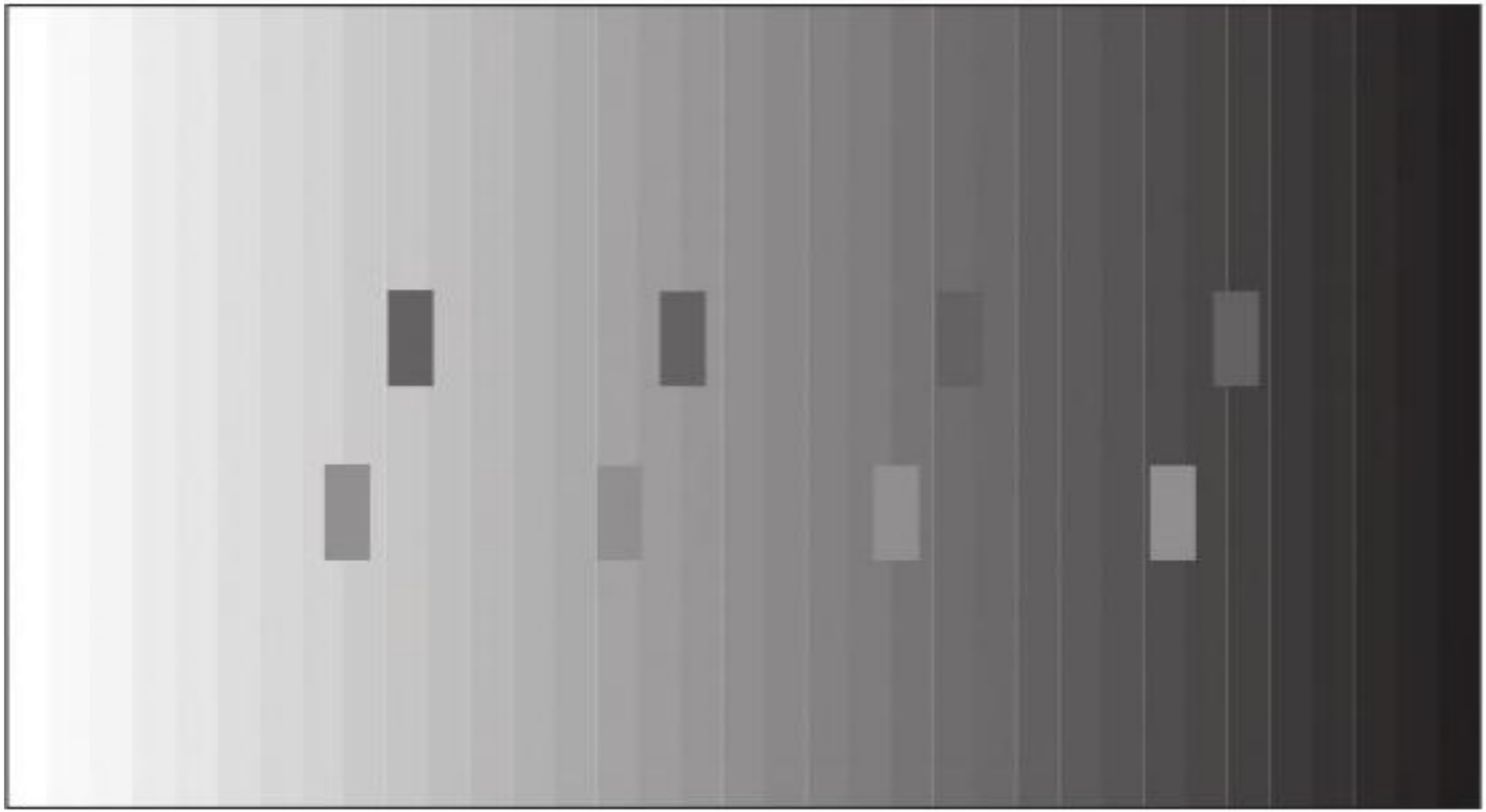
Image courtesy of John McCann via Maureen Stone

Which bar is the darkest

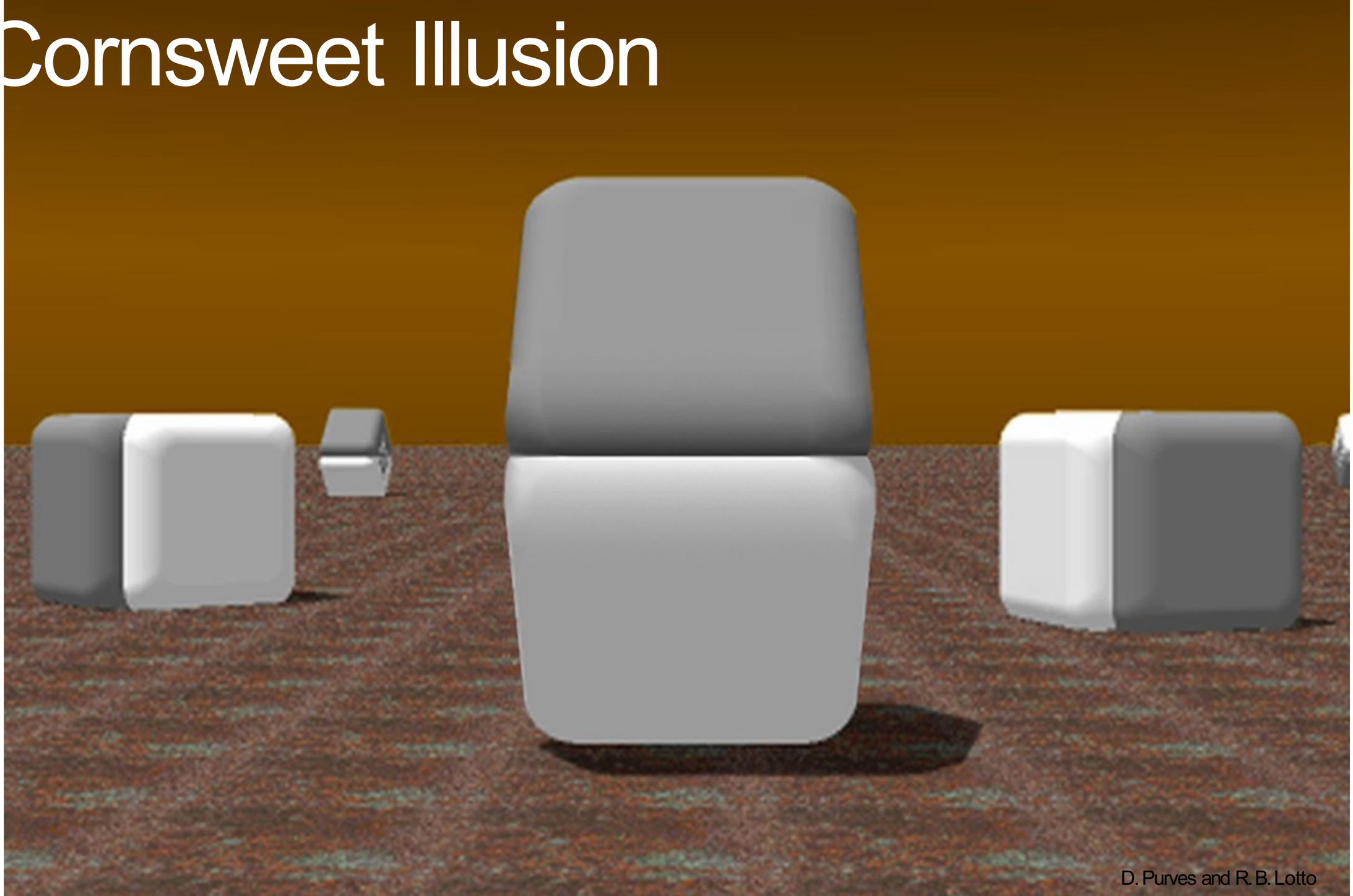
Which bar is the lightest

Top Row A B C D

Bottom Row E F G H

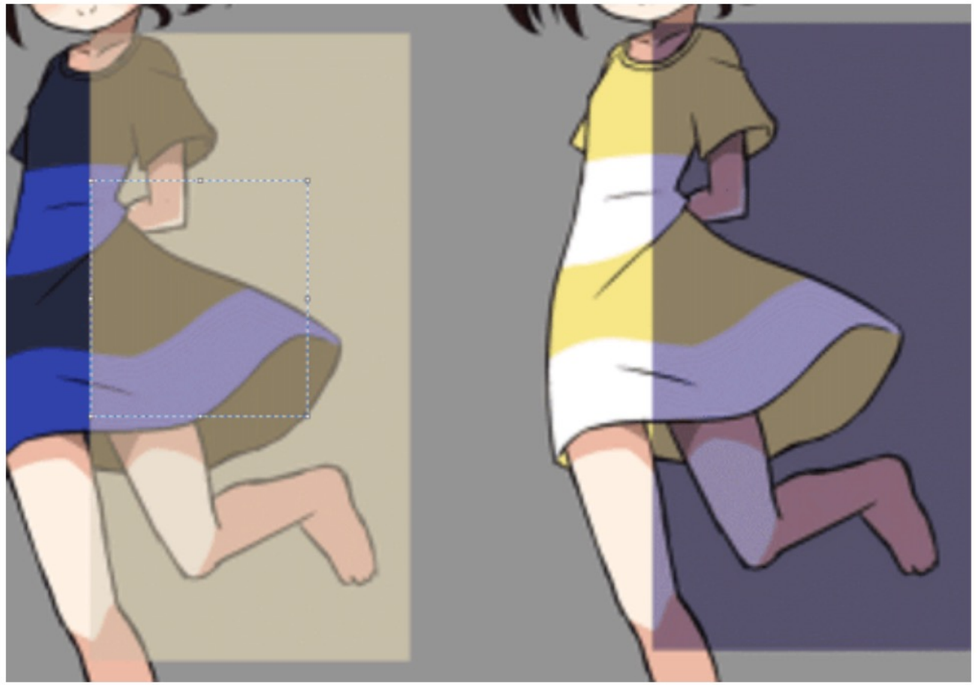


Cornsweet Illusion



D. Purves and R. B. Lotto

Contrast with background



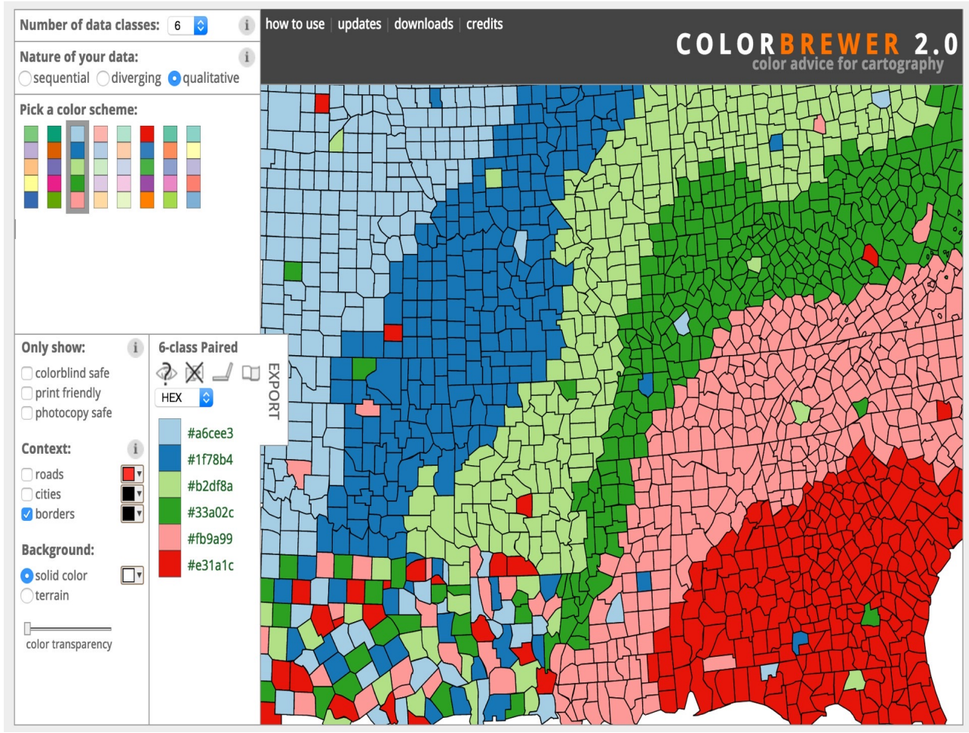
Black and blue? White and gold?

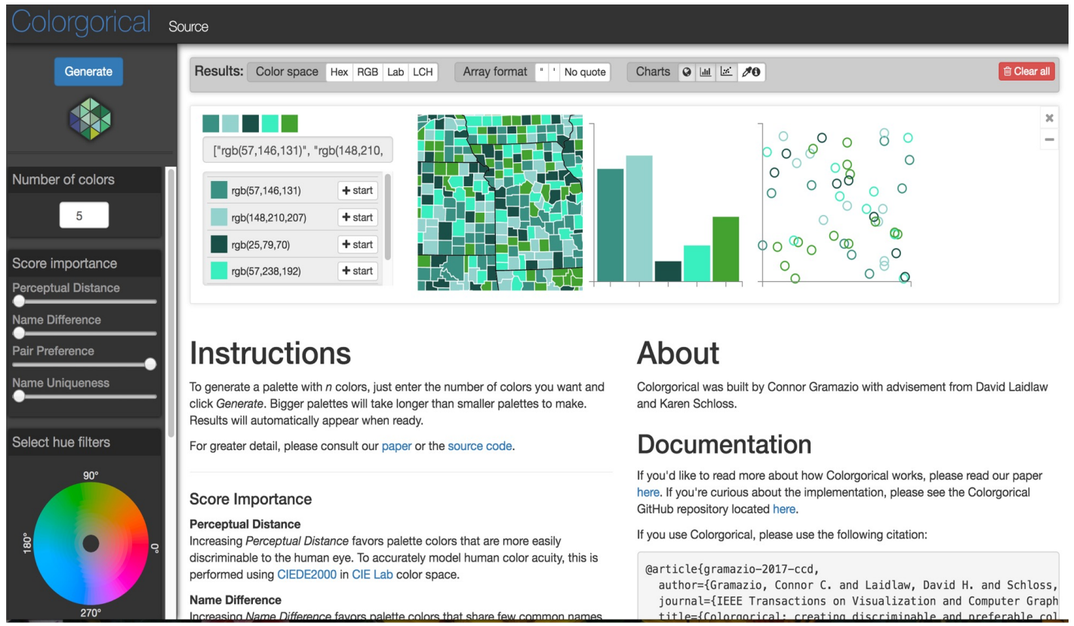
https://imgur.com/hxJjUQB

https://en.wikipedia.org/wiki/The_dress

Tools for picking Color

- Color Brewer
- Colorgorical
- Programming Libraries



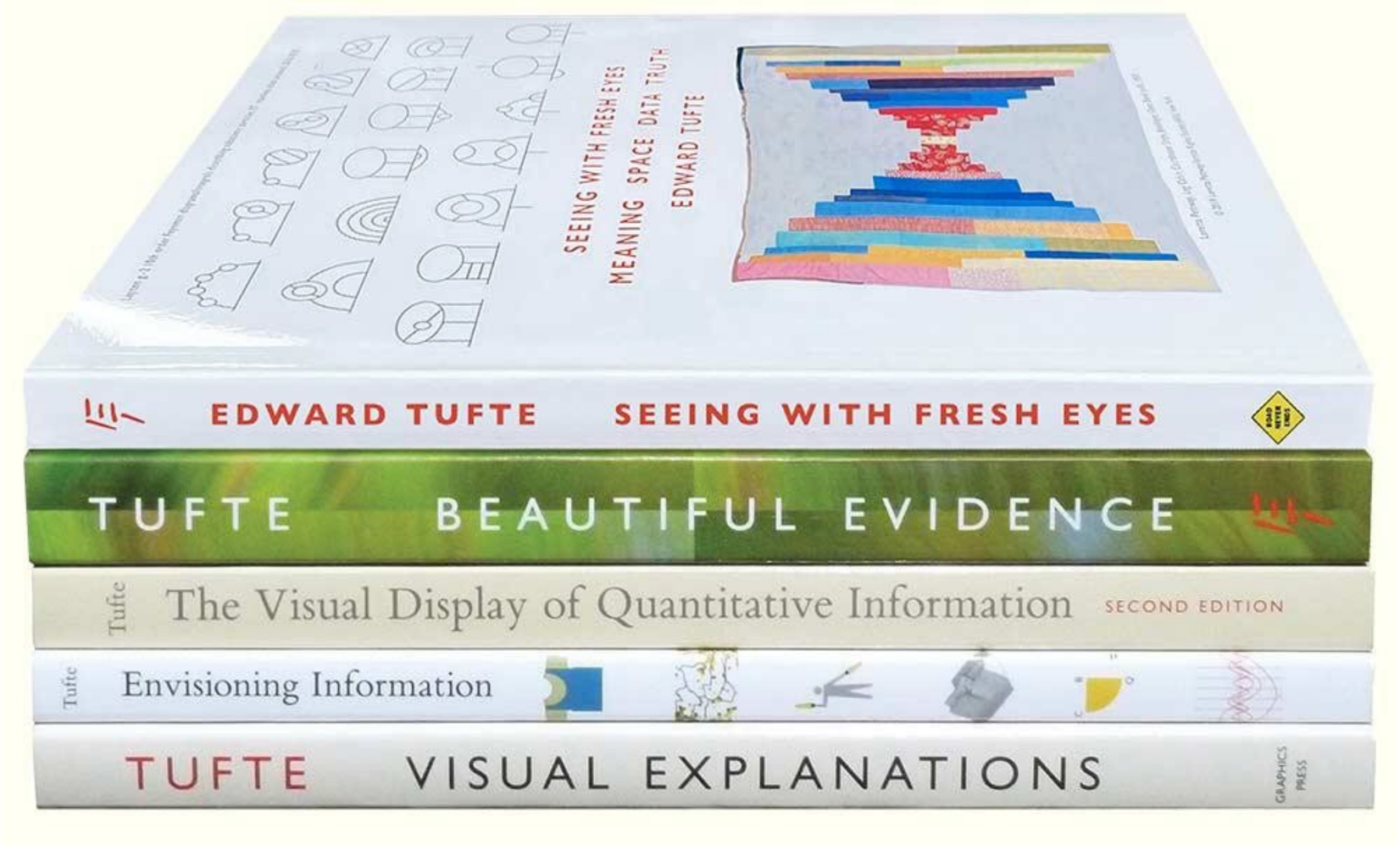


Visualization for Data Science
Design Rules of Toe

Principles of Effective Visualizations

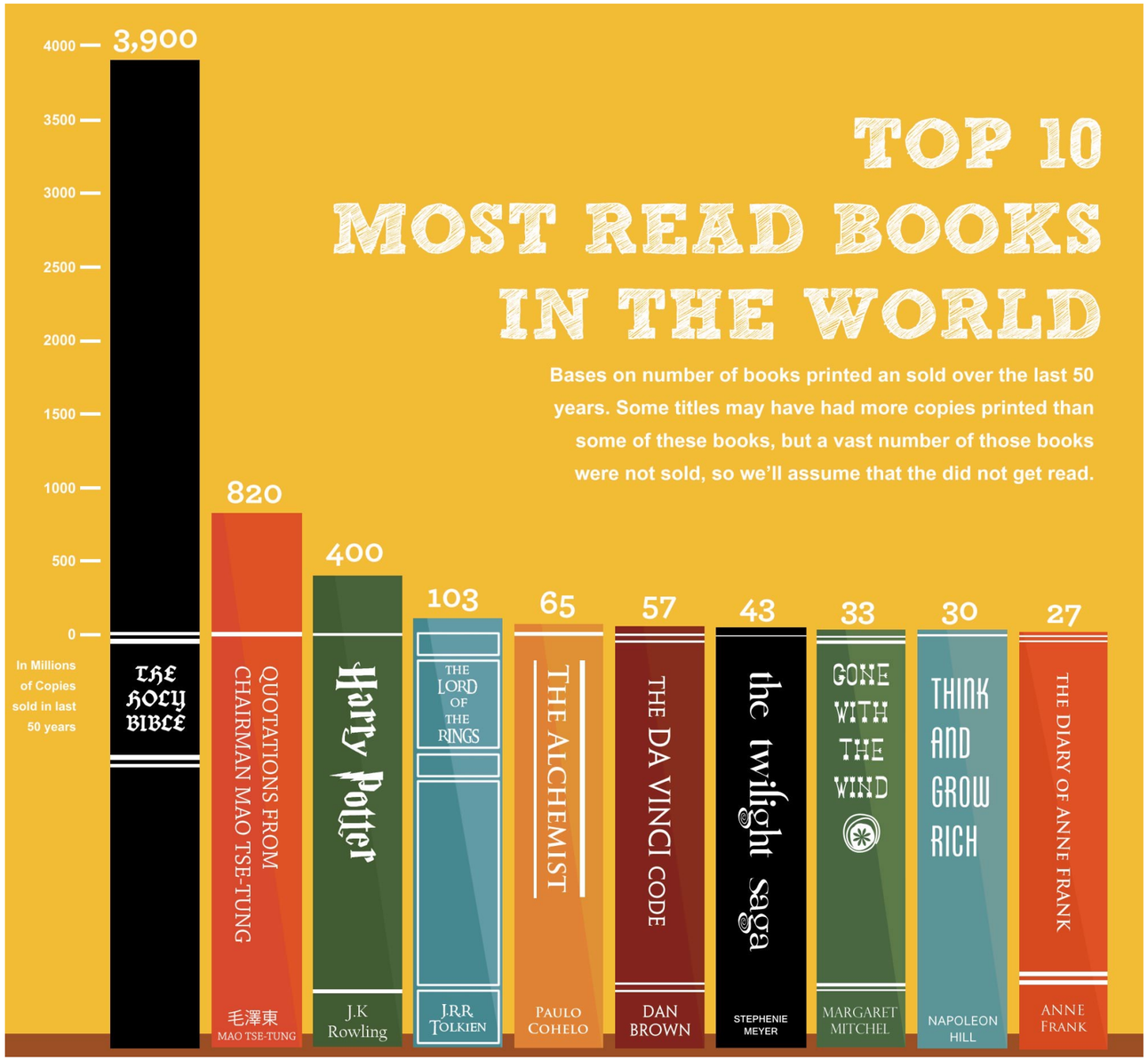
Principle	Definition	Examples
Proportional Ink	The amount of ink used to indicate a value should be proportional to the value itself	Truncating a position channel on a bar chart to exaggerate the difference between bars violates the principle of proportional ink.
Data:ink ratio	Remove distracting visual elements to focus attention on the data	Lighten line weights, remove backgrounds, never use 3D or special effects, avoid unnecessary/redundant labels.
Labels & legends	Use axes labels and titles to highlight/ communicate data	Never leave your data column names as axes labels! Generally good to add a title.
Colour & Accessibility	Colour can be used to encode information or for aesthetics/style/ design. However, colour can also be distracting if used inappropriately or poorly.	Choose a perceptually uniform colour palette; can be sequential or diverging for quantitative data. Opt for colour-blind friendly palettes. Categorical data can use qualitative colour schemes.

Edward Tufte



Principle of Proportional Ink

When a shaded region is used to represent a numerical value, the area of that shaded region should be ***directly proportional*** *to the corresponding value.* In other words, the amount of ink used to indicate a value should be proportional to the value itself.



https://www.businessinsider.com/the-top-10-most-read-books-in-the-world-infographic-2012-12

Data-Ink Ratio Principle

- Data-ink: Non-erasable ink that presents data. Removing it eliminates content.
- Non-data-ink: ink that does not transport the information but it is used for scales, labels and edges.
- Data-ink ratio: Proportion of ink presenting actual data vs. total ink used.

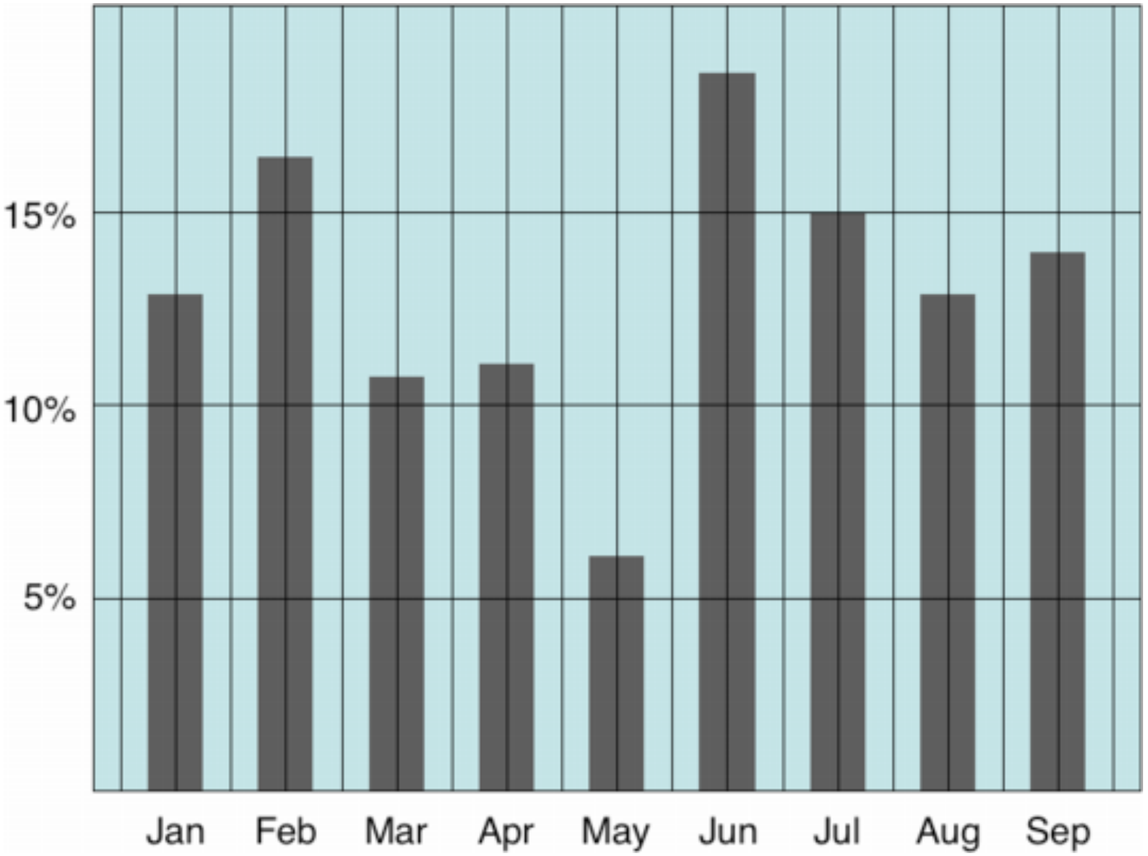
$$Data\_ink\ ratio = \frac{Data\_ink}{Total\ ink\ used\ to\ print\ the\ graphic}$$

- Principle: Maximize data ink, minimize non-data-ink

Maximize data-ink ratio (Edward Tufte)

- **Efficiency of Communication**: Design data visualizations efficiently, ensuring each mark on a graph conveys useful information.
- **Minimizing Non-Essential Elements**: Eliminate "chartjunk" like excessive grid lines, unneeded labels, and decorative features that don't enhance understanding.
- **Emphasis on Data**: Prioritize the data itself over artistic design elements, making it the central focus.
- **Clarity and Precision**: Reduce non-essential elements to enhance graph clarity and precision, aiding quick and accurate interpretation.
- **Balance**: While advocating minimalism, maintain essential elements like axis labels and keys for comprehension.

Data Ink Maximization



Activity – (5 minutes)

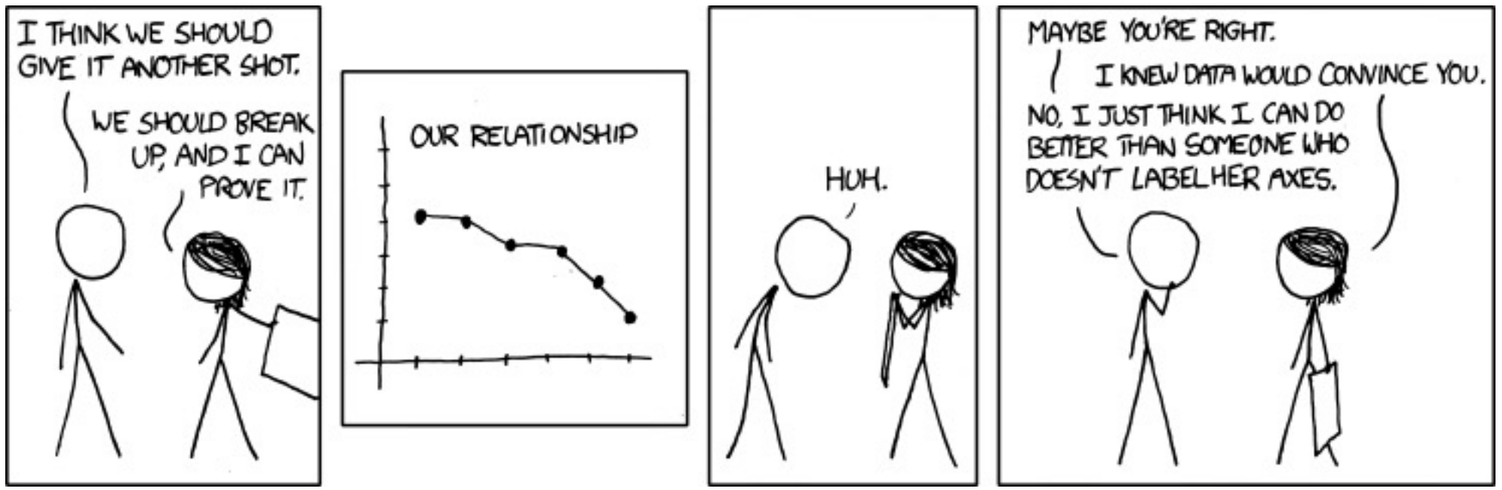
Redraw the chart removing the necessary ink that is present.

ongoing,Tim Brey: https://www.tbray.org/ongoing/data-ink/di1

Labelling Principles

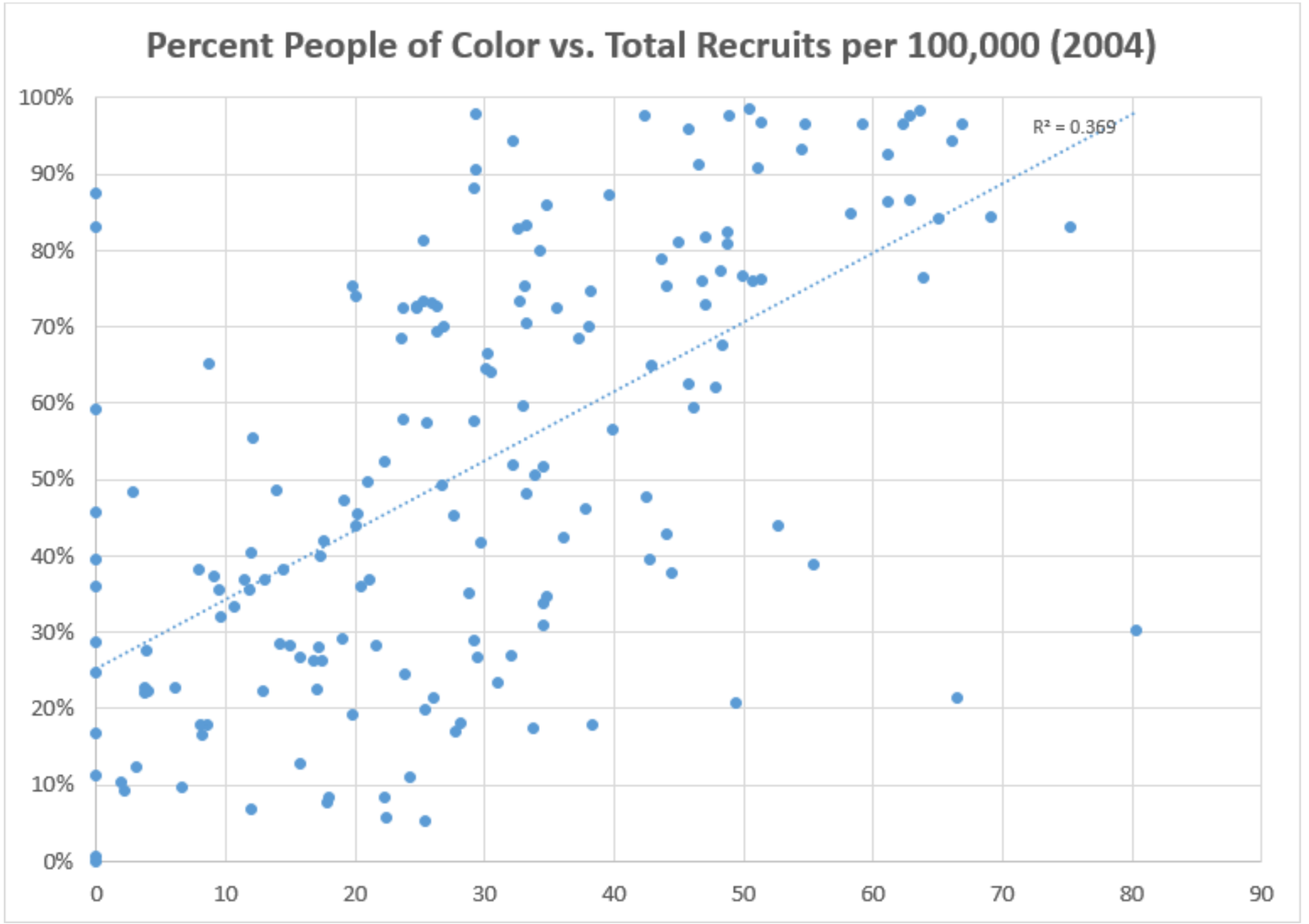
make visualizations as **self-documenting** as possible

- meaningful & useful title, labels, legends
  - axes and panes/subwindows should have labels
    - and axes should have good mix/max boundary tick marks
  - everything that's plotted should have a legend
    - and own header/labels if not redundant with main title
  - use reasonable numerical format
    - avoid scientific notation in most cases

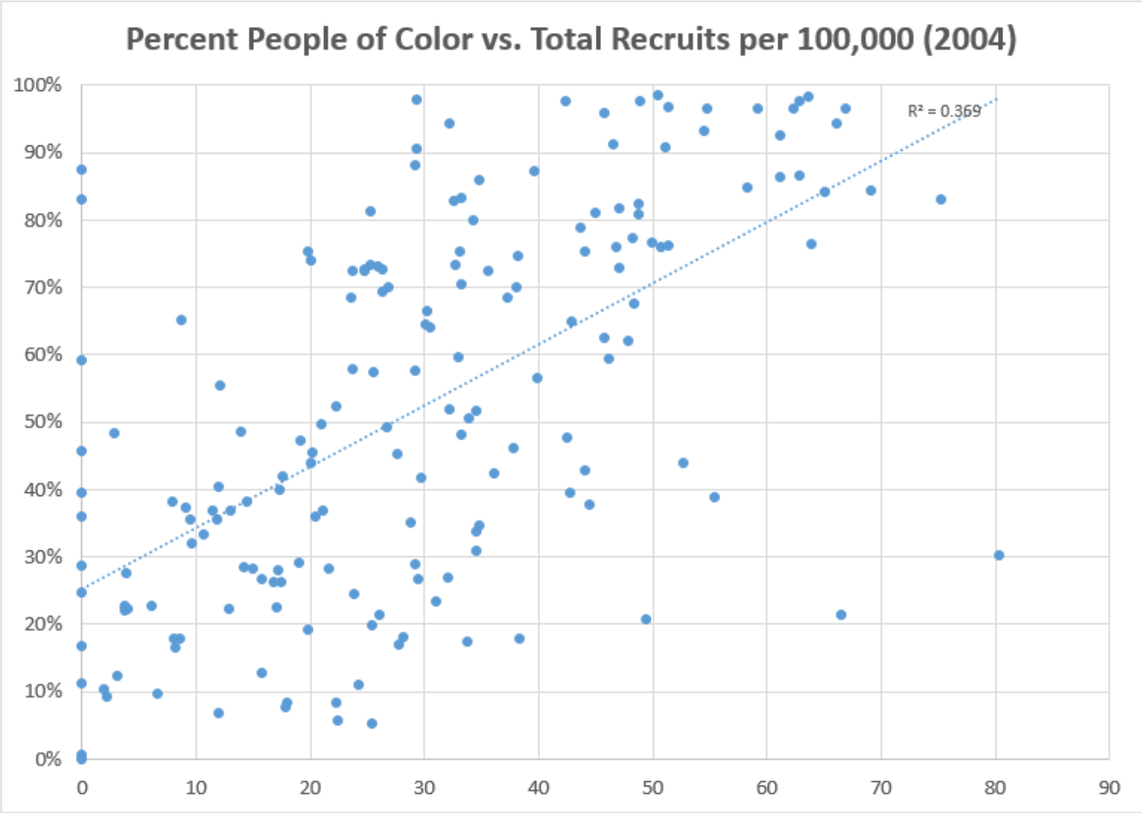


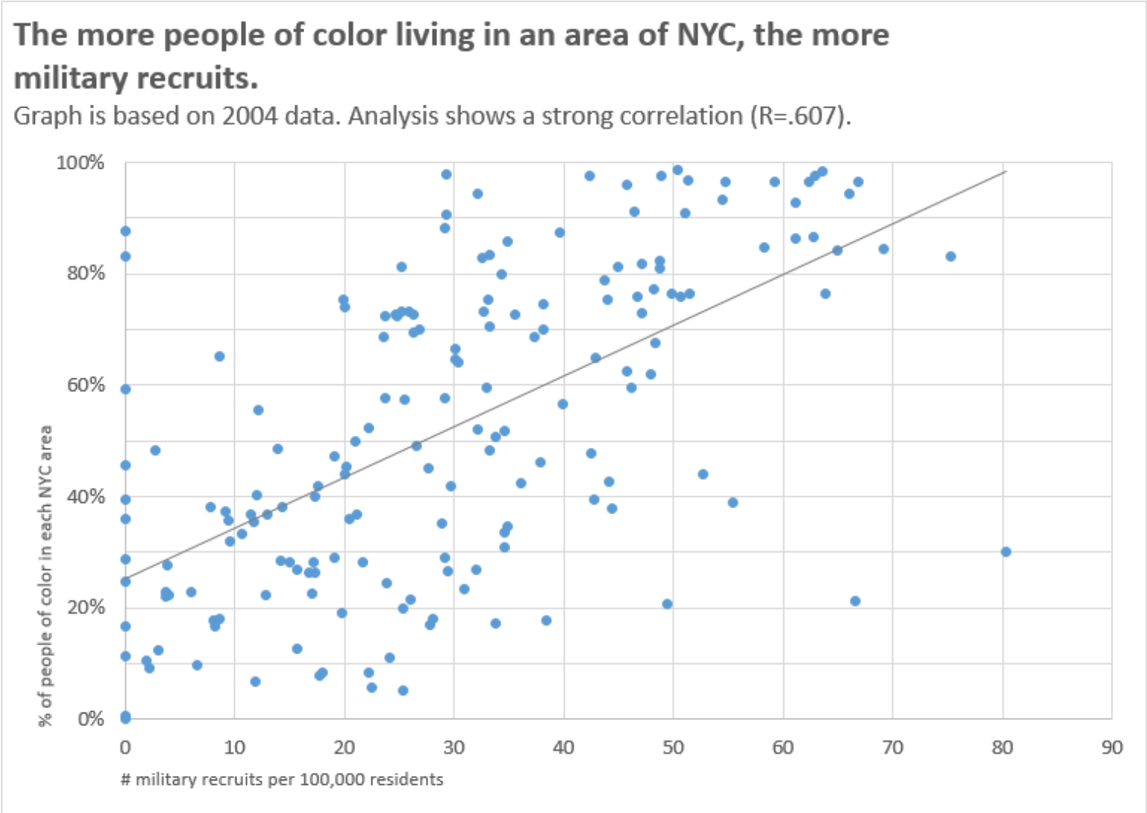
[https://xkcd.com/833/]

Labelling


Percent People of Color vs. Total Recruits per 100,000 (2004)
R² = 0.369
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
10
20
30
40
50
60
70
80
90

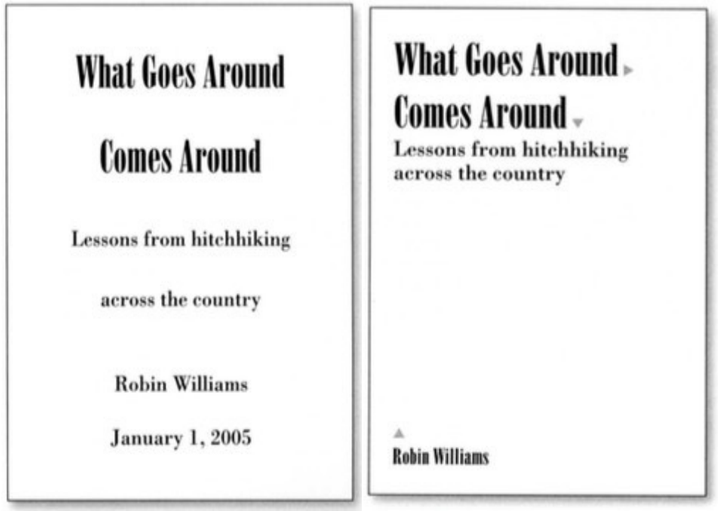
Labelling


Percent People of Color vs. Total Recruits per 100,000 (2004)
R² = 0.369
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
10
20
30
40
50
60
70
80
90


The more people of color living in an area of NYC, the more military recruits.
Graph is based on 2004 data. Analysis shows a strong correlation (R=.607).
% of people of color in each NYC area
100%
80%
60%
40%
20%
0%
0
10
20
30
40
50
60
70
80
90
military recruits per 100,000 residents

Labelling Principles

- **Label axes with units**: Essential for interpretability and measurement clarity
- **Titles convey insight**: Go beyond axis labels to communicate findings
- **Add subtitles for context**: Clarify methodology, timeframe, or conditions
- **Optimize data-ink ratio**: Balance information density with visual clarity (Tufte)
- **Align purposefully**: Avoid default centering; use intentional positioning

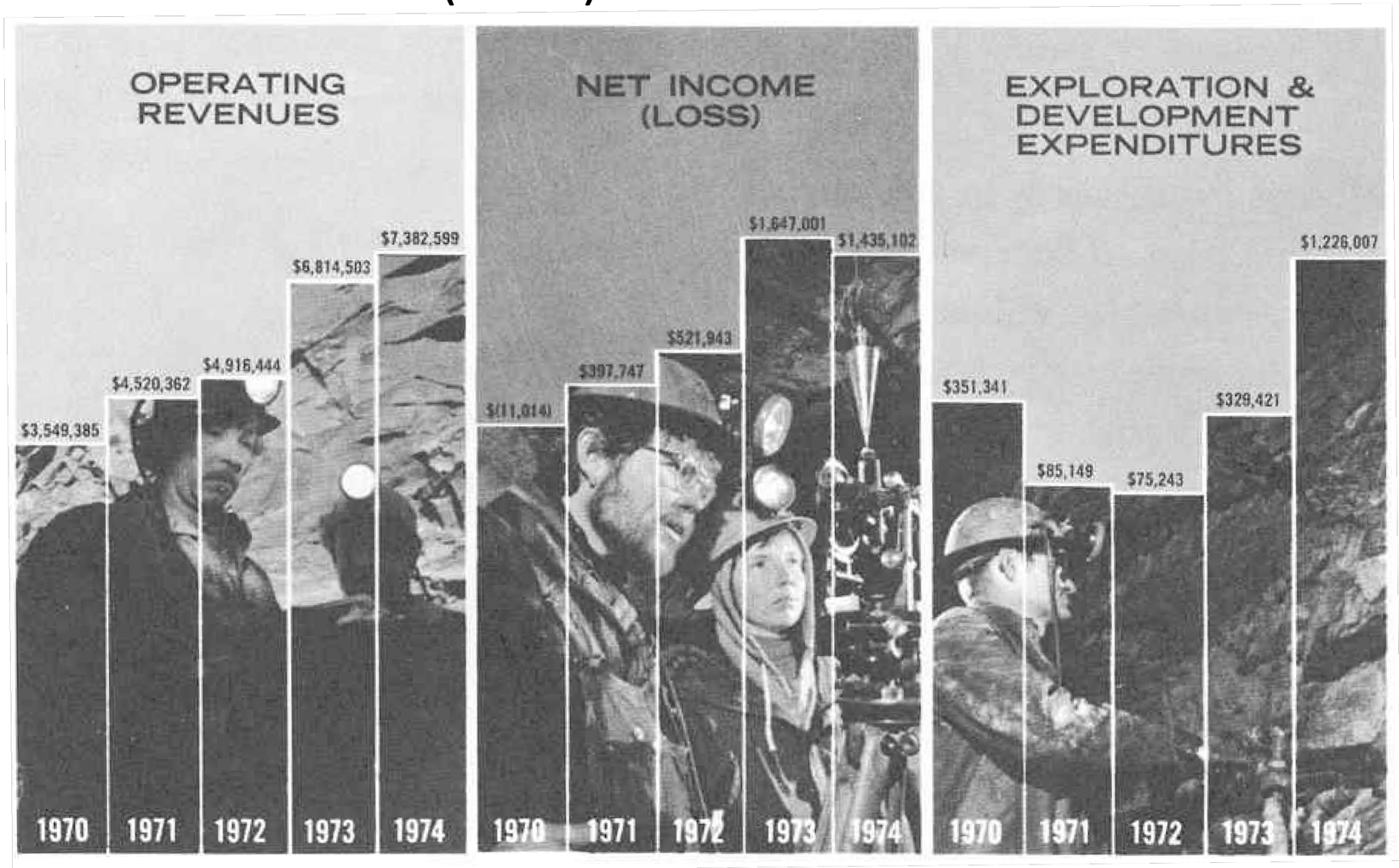


Principles of Effective Visualizations

Principle	Definition	Examples
Proportional Ink	The amount of ink used to indicate a value should be proportional to the value itself	Truncating a position channel on a bar chart to exaggerate the difference between bars violates the principle of proportional ink.
Data:ink ratio	Remove distracting visual elements to focus attention on the data	Lighten line weights, remove backgrounds, never use 3D or special effects, avoid unnecessary/redundant labels.
Labels & legends	Use axes labels and titles to highlight/ communicate data	Never leave your data column names as axes labels! Generally good to add a title.
Colour & Accessibility	Colour can be used to encode information or for aesthetics/style/ design. However, colour can also be distracting if used inappropriately or poorly.	Choose a perceptually uniform colour palette; can be sequential or diverging for quantitative data. Opt for colour-blind friendly palettes. Categorical data can use qualitative colour schemes.

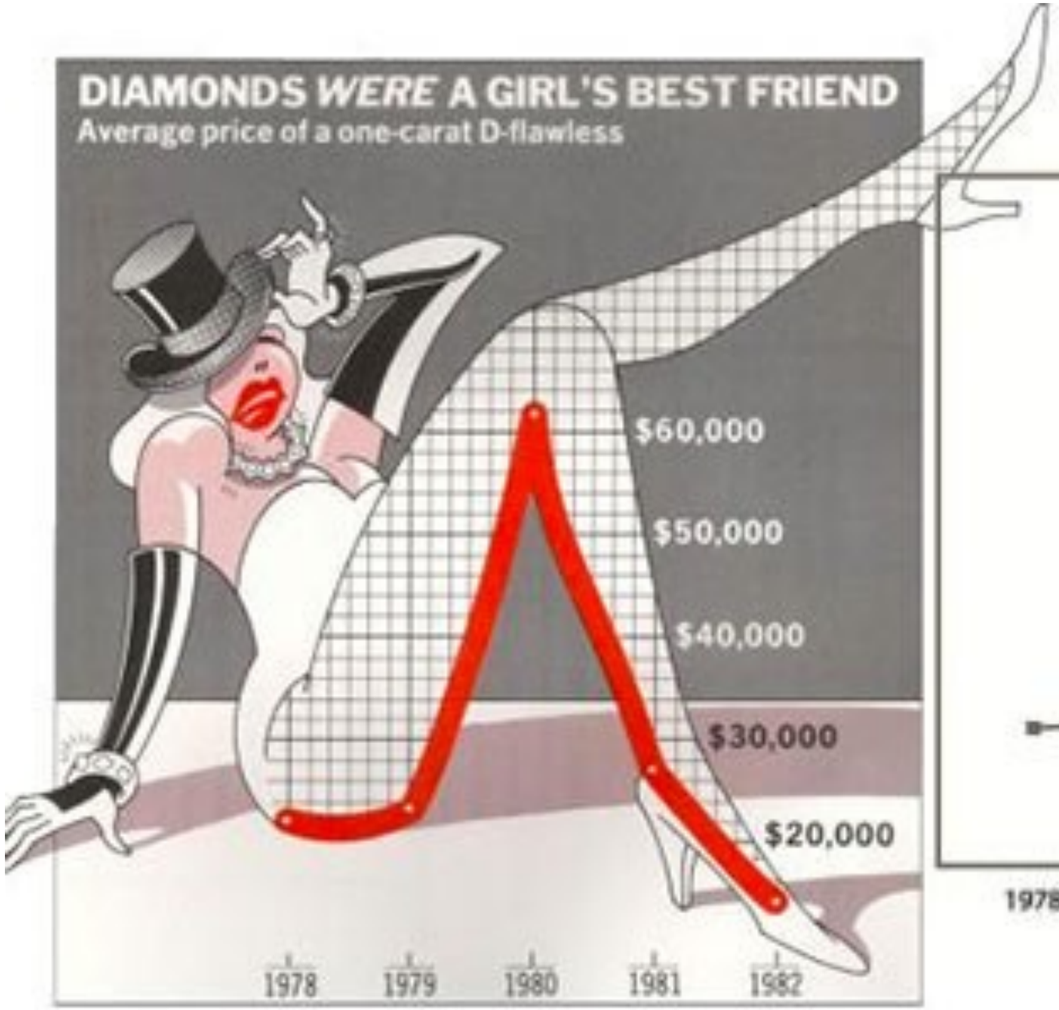
Function first, form next

“Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.” Tufte (1983)

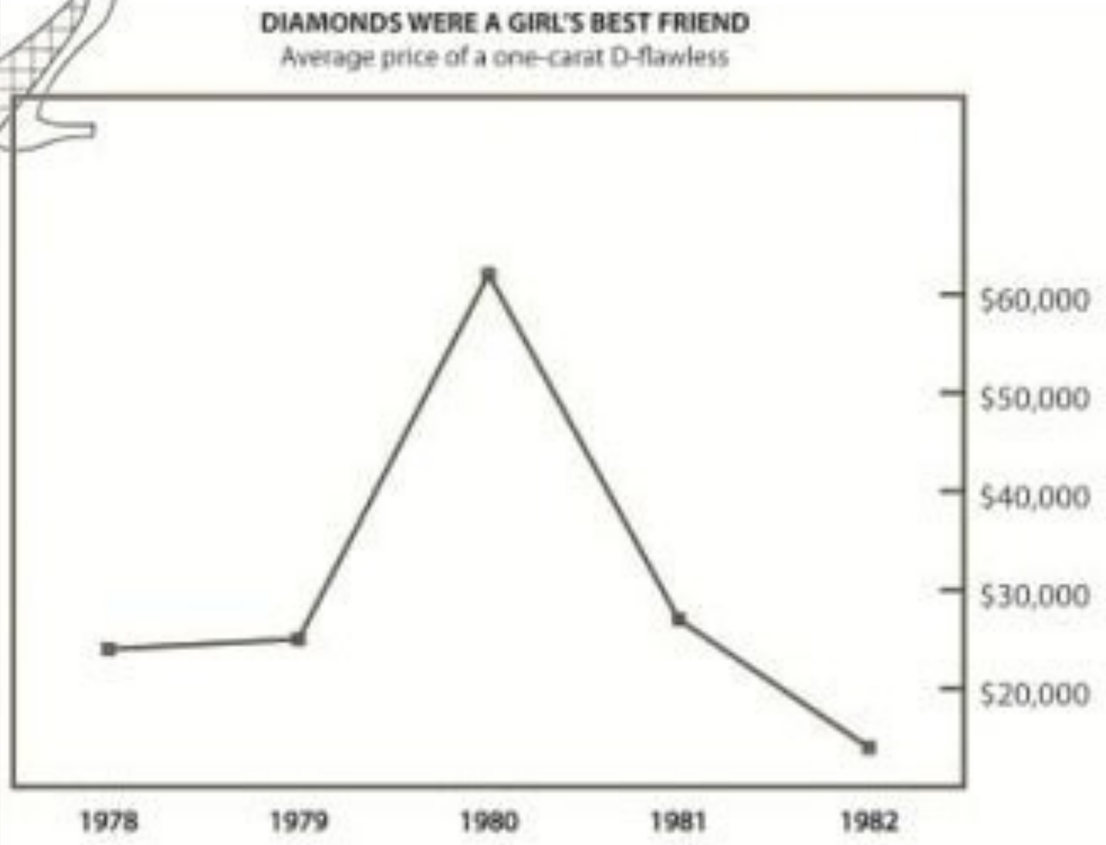


Which is better? Why?

A



B

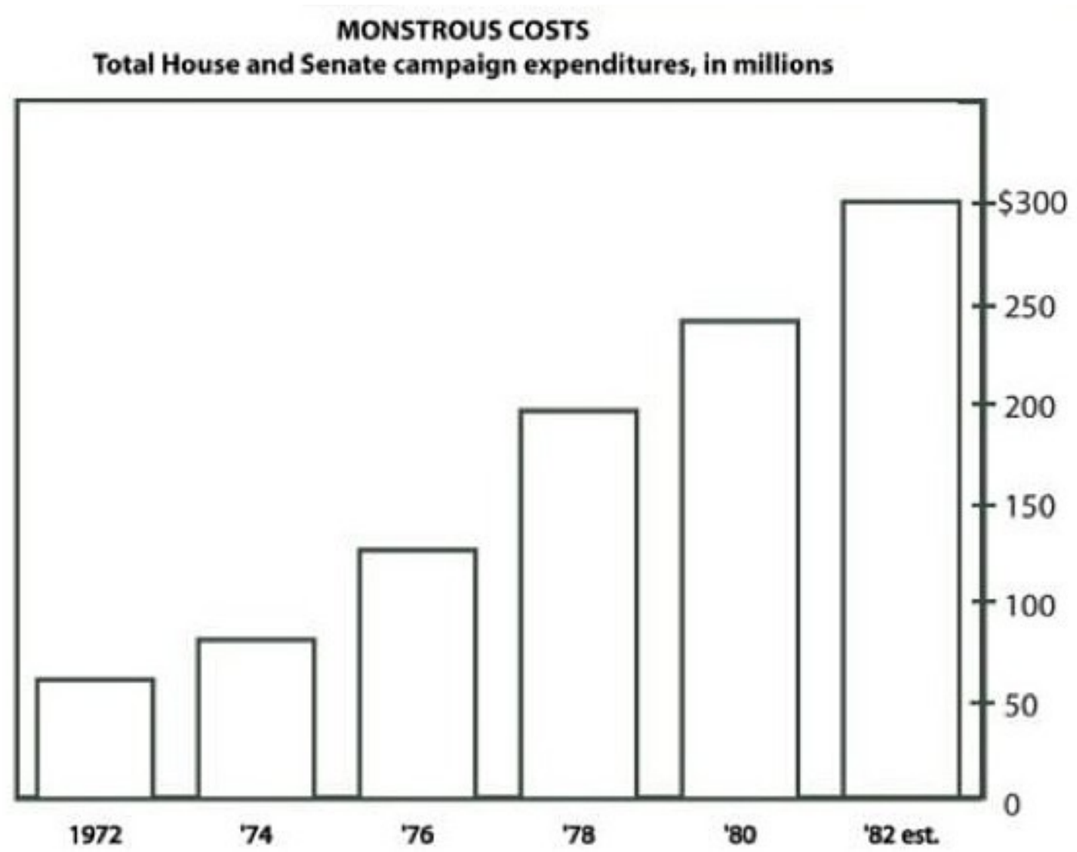


[Bateman et al. 2010]

Which is better? Why?

A

B



[Bateman et al. 2010]

https://eagereyes.org/criticism/chart-junk-considered-useful-after-all

Use Embellished Charts? It Depends!

PROS

CONS

Use Embellished Charts? It Depends!

PROS	CONS
persuasion	biased analysis
memorability	trustworthiness
engagement	interpretability
	space efficiency
	effort

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce 'chart junk' – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

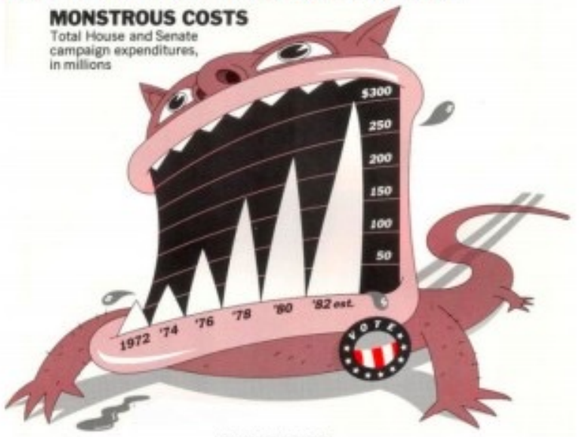
Author Keywords
Charts, information visualization, imagery, memorability.

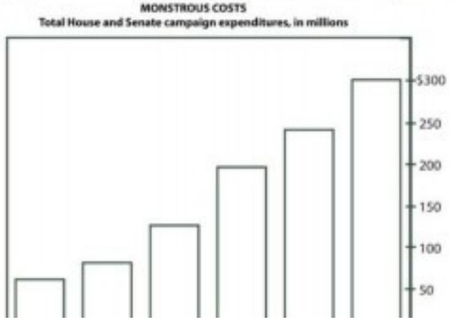
ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors

INTRODUCTION
Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk,* decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates *data-ink* – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).



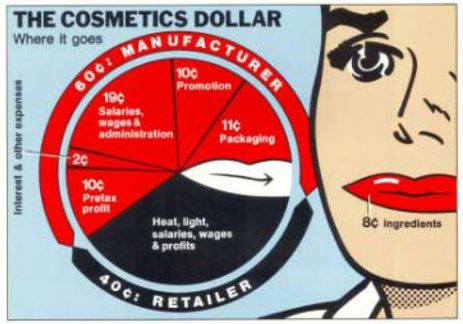


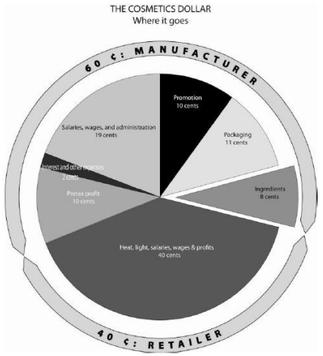
Experimental results

1. No difference for **interpretation accuracy**
2. No difference in **recall accuracy after a five-minute gap**
3. Significantly **better recall for Holmes charts** of both the chart topic and the details (categories and trend) **after long-term gap** (2-3 weeks).
4. Participants **saw value messages** in the Holmes charts significantly more often than in the plain charts.
5. Participants found the Holmes charts **more attractive, most enjoyed** them, and found that they were **easiest and fastest to remember**.

“Chart Embellishments Debate”

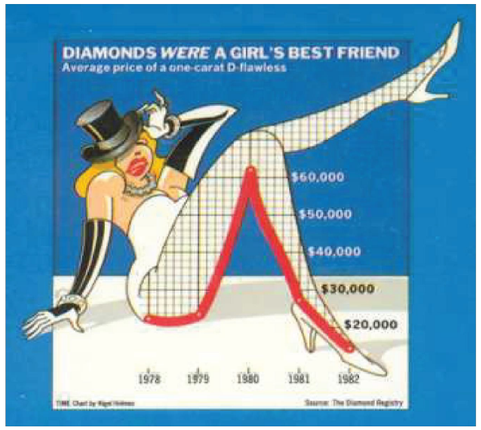
Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts





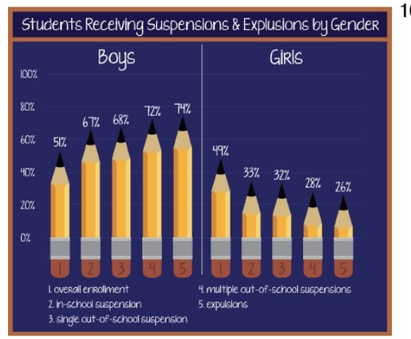
Bateman, et al. (2010)

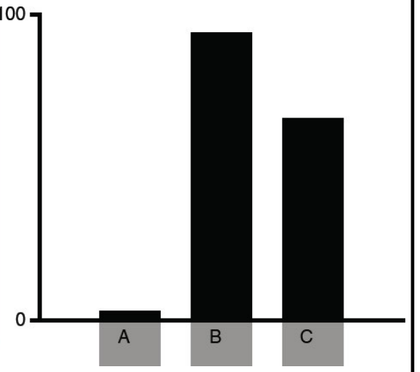
Benefitting InfoVis with Visual Difficulties



Hullman, et al. (2011)

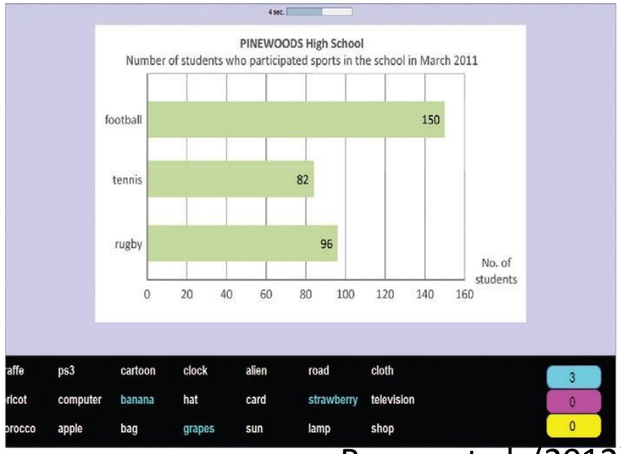
An Evaluation of the Impact of Visual Embellishments in Bar Charts





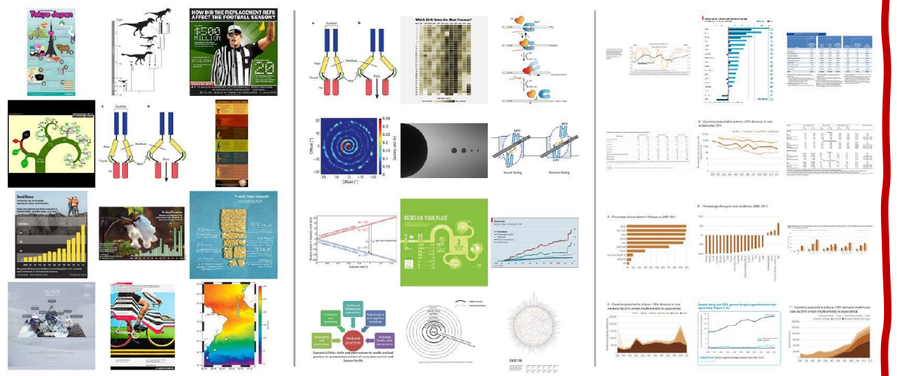
Skau, et al. (2015)

An Empirical Study on Using Visual Embellishments in Visualization



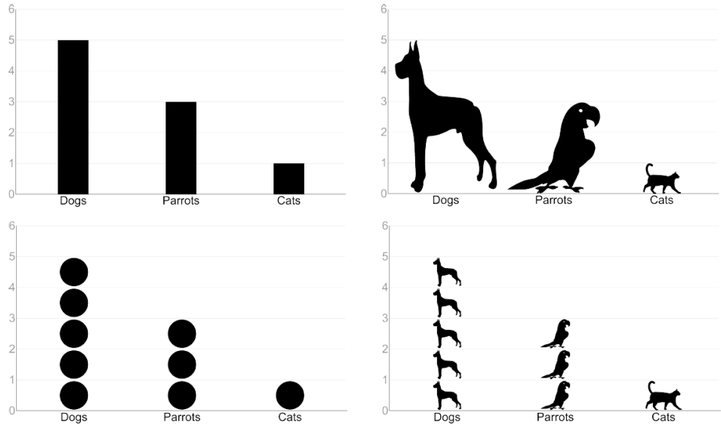
Borgo, et al. (2012)

What makes a visualization memorable?



Borkin, et al. (2013)
Borkin, et al. (2015)

ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs



Haroz, et al. (2015)

Function first, form next

- dangerous to start with aesthetics
  - usually impossible to add function retroactively
- start with focus on functionality
  - possible to improve aesthetics later on, as refinement
  - if no expertise in-house, find good graphic designer to work with
  - aesthetics do matter! another level of function
    - visual hierarchy, alignment, flow
    - Gestalt principles in action