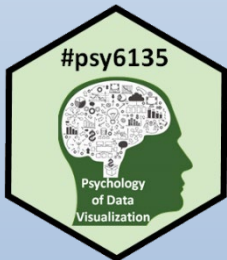
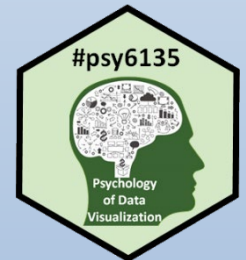


# Graphical Perception



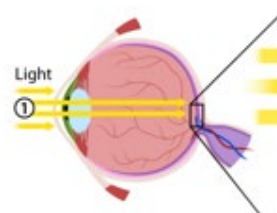
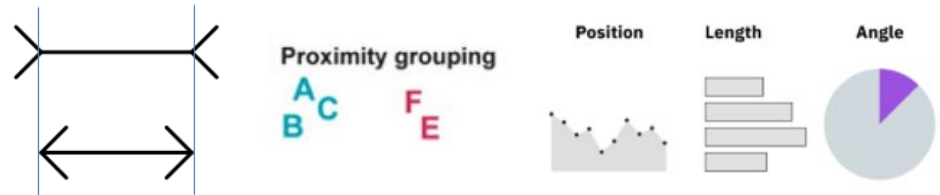
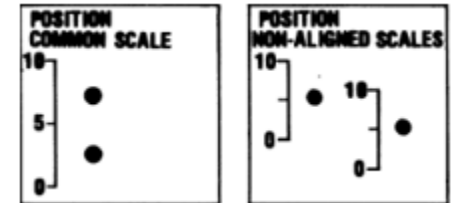
Michael Friendly  
Psych 6135

<https://friendly.github.io/6135>



# Topics

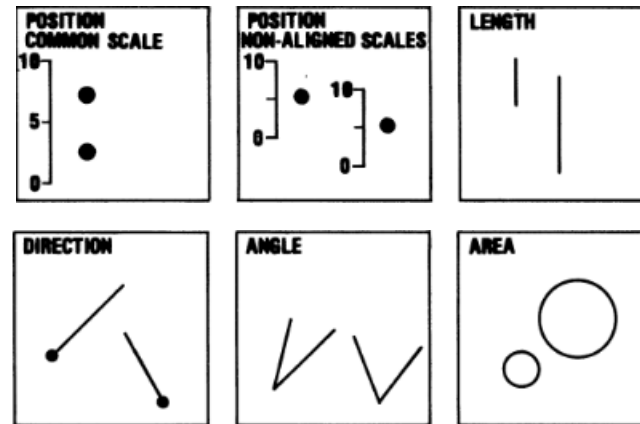
- Perception & Cognition
  - Encoding, decoding
  - Top-down vs. bottom-up processing
- Perceptual aspects
  - Illusions
  - Gestalt factors
  - Accuracy of decoding
- Cognitive aspects
  - Memory
  - Color



# Graphical Perception

- In constructing a graph, **quantitative** and **categorical** information is **encoded** by visual attributes:

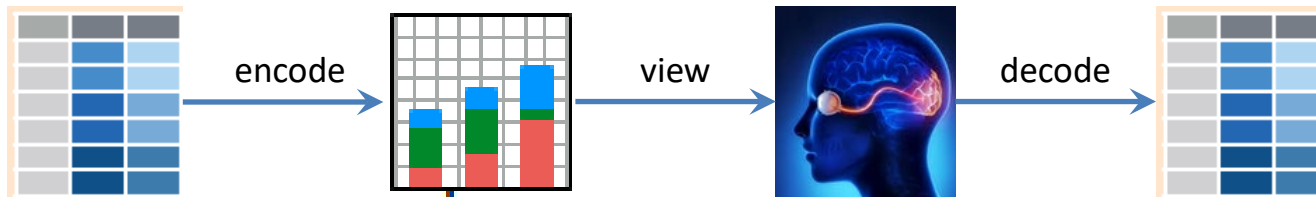
- Length
- Position along axis
- Angle
- Area
- Color, shape, line style



- What determines the ability of graph viewers to:
  - Make **comparisons** (which is larger?)
  - **Estimate** a magnitude?
  - See **patterns**, trends, unusual features?

# Encoding & decoding

- When we construct a graph, we **encode** a numerical or categorical variable as a graphical attribute
- When we view a graph, the goal is to **decode** the graphical attributes and extract information about the data that was encoded

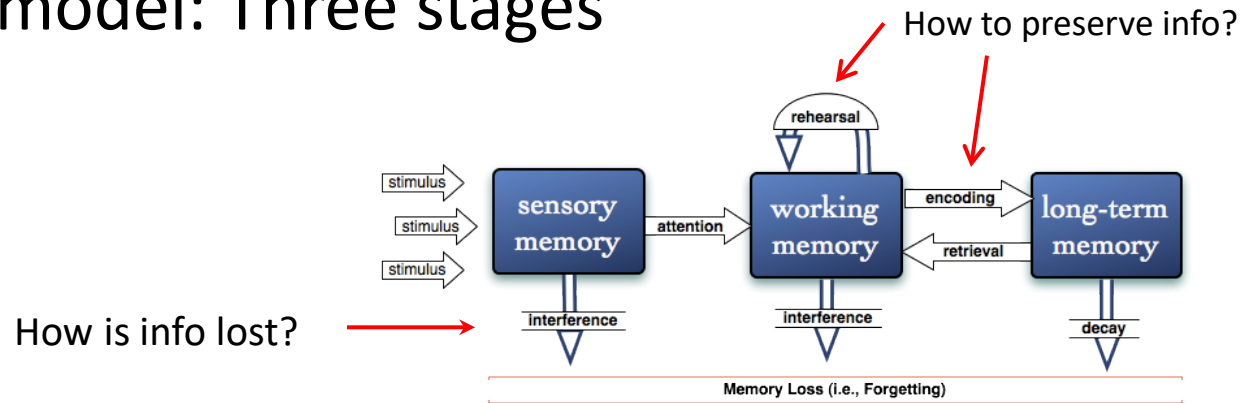


- → Encoding should rely on features that can easily be decoded
- → Designing effective graphs depends on understanding these ideas!



# Visual & cognitive systems

- A simplified model: Three stages

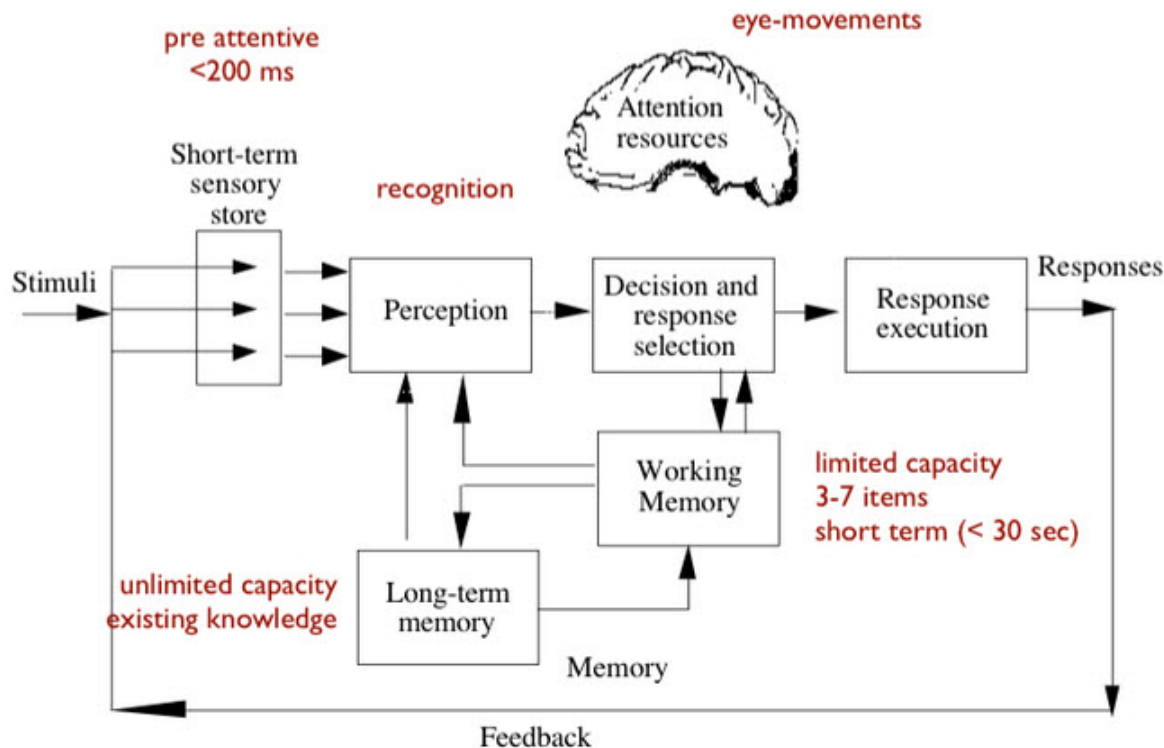


- Sensory (iconic) memory
  - pre-attentive, automatic, feature detection
  - massively parallel, short duration, easily fooled (“thinking fast”)
- Working memory
  - requires attention, limited capacity (~ 4-6 “chunks”)
  - memory aids: rehearsal; imagery
- Long-term memory
  - real-world knowledge, ~ unlimited capacity, inference (“thinking slow”)

# Perception vs. cognition

Another coarse distinction:

- **Perception:** Processing of the signals coming in: what you “see”
- **Cognition:** How you **understand** and **interpret** what you see



A nice scientific or textbook diagram

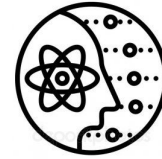
But where is cognition?



# Perception: Bottom-up & Top-down

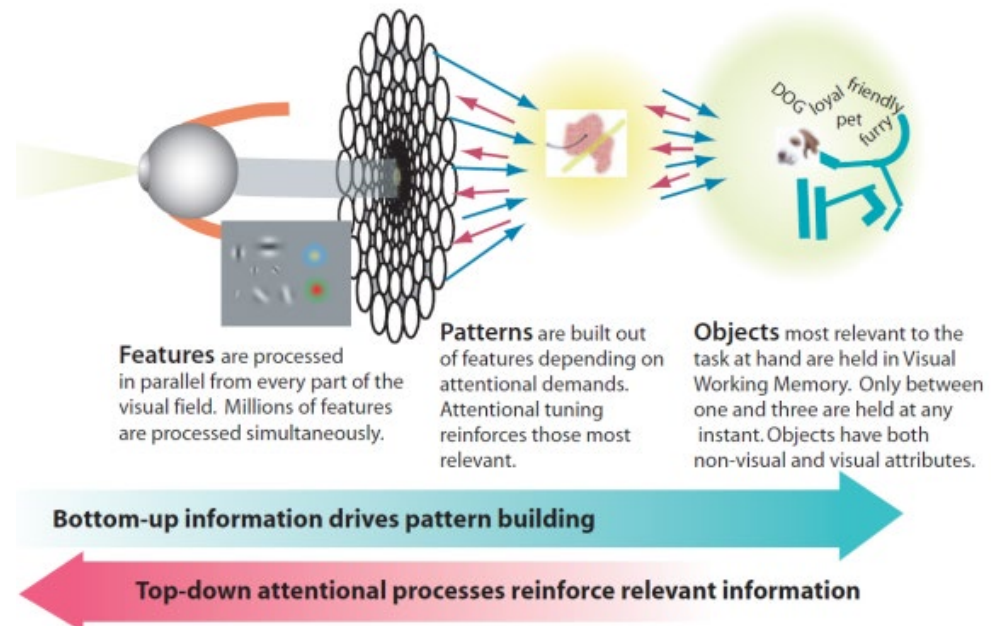
- Bottom-up processing
  - Low level: features → pattern → object
  - Detect edges, contours, color, motion
- Top-down processing
  - Driven by goals, expectations
  - Uses prior knowledge, experience, filters what we “see”

## Epistemology



Empiricism  
Aristotle

Rationalism  
Plato



# Perception: Bottom-up

How many 5s in this display?

1561321203658413076510374627  
4173127527327592732990709742  
1703707774179527931749270973  
4019743217909370945179279417

How many 5s in this display?

1561321203658413076510374627  
4173127527|327592732990709742  
1703707774179527931749270973  
4019743217909370945179279417

Numerals differ only in **shape**, and are high-level symbols

You have to literally scan them **all** & count the 5s.

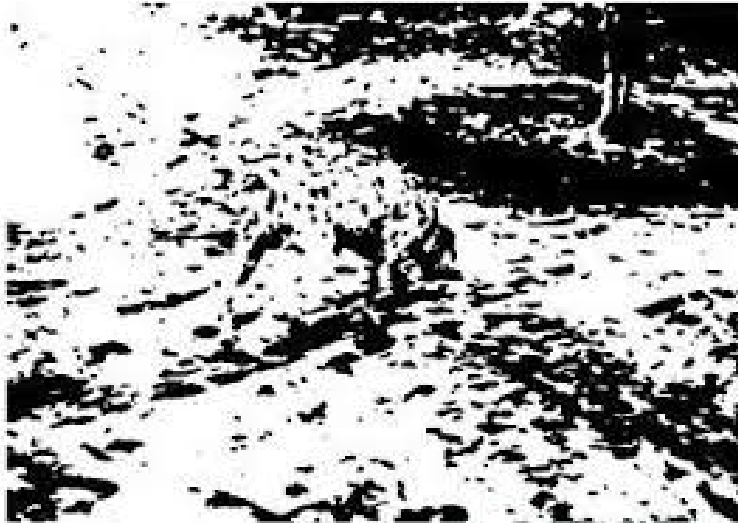
The distinction of **color** is immediate & **pre-attentive**

You only have to scan for 5s & count them.

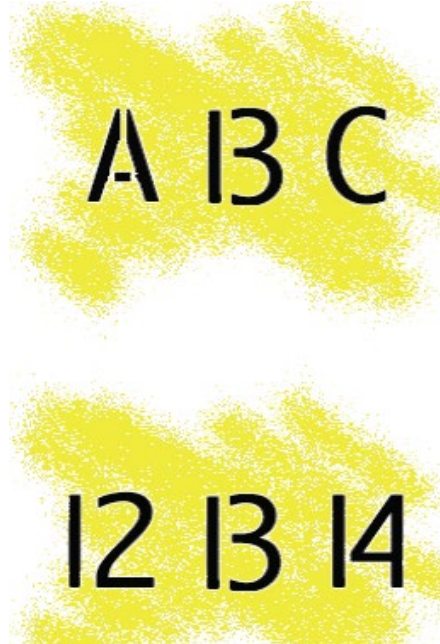
This is why **color** is an important visual attribute for a **categorical** variable in graphs

# Perception: Top-down

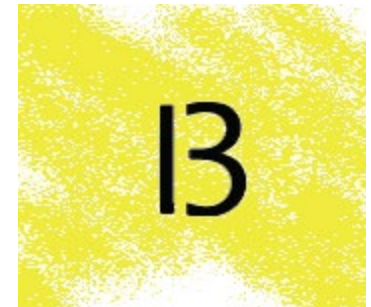
What is in this scene?



What is the middle character?



What here?



An ambiguous figure!

What is the middle letter in each word?

THE CAT

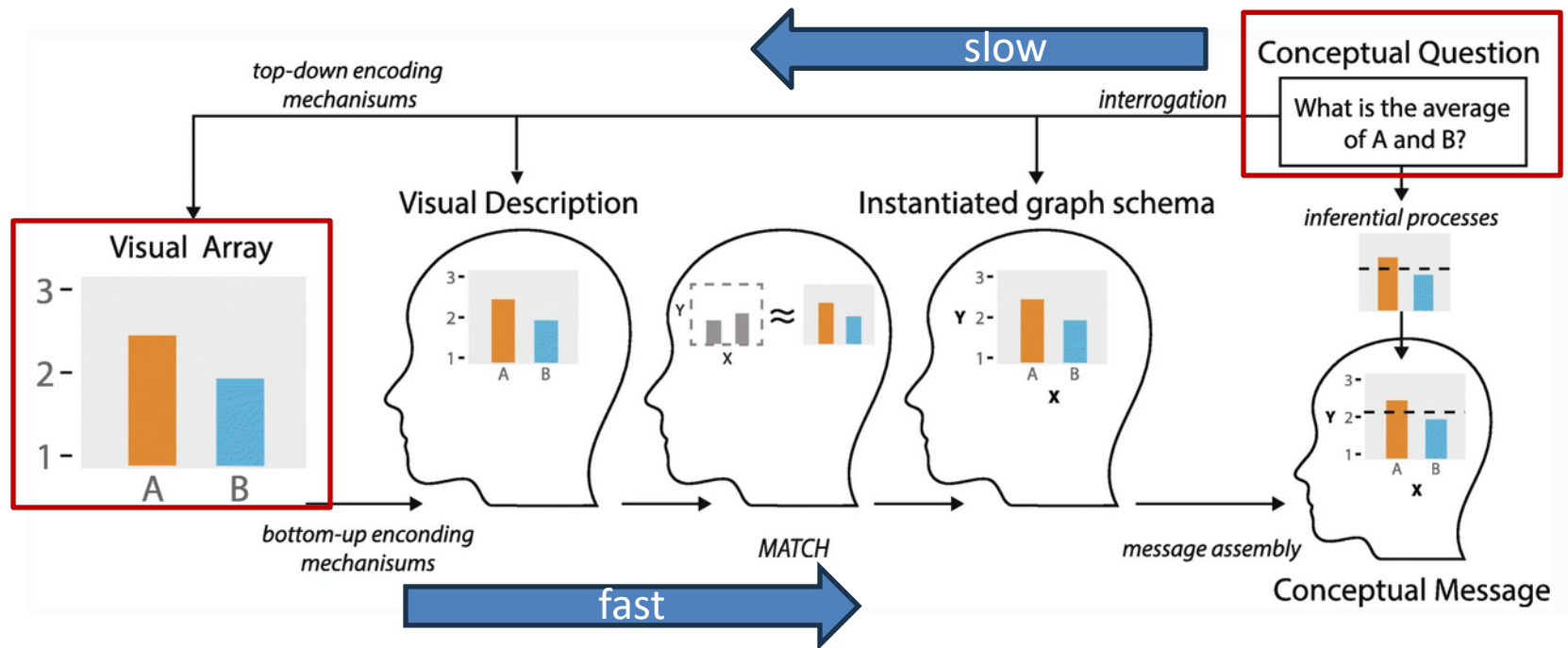
All of these are demonstrations of the role of **expectations** (top-down) in determining what we “see”

Gregory ('70): perception as **constructive**, depends on prior knowledge

# Cognition: Graph reading

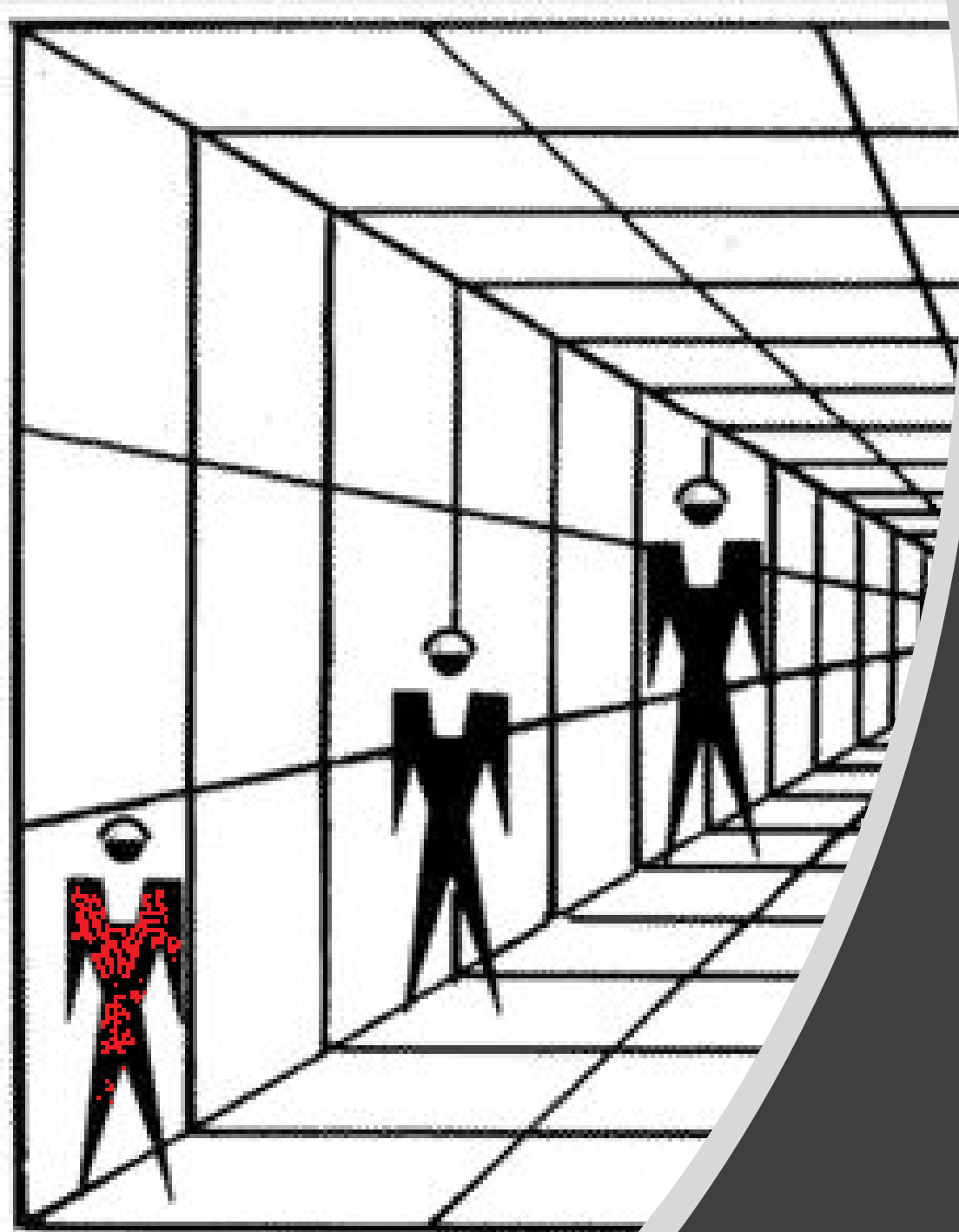
Pinker (1990): cognitive model that integrates top-down and bottom-up encoding mechanisms in understanding of data graphs

- What do we do, given a graph and a question?



Pinker, S. (1990). A theory of graph comprehension. In *Artificial intelligence and the future of testing*.

Image from: Padilla et al. (2018) <https://bit.ly/4atEwEo>



# Illusions:

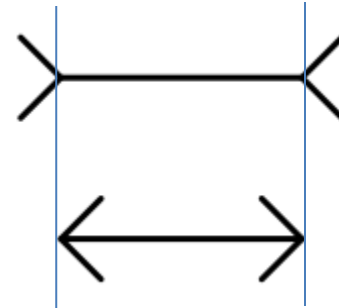
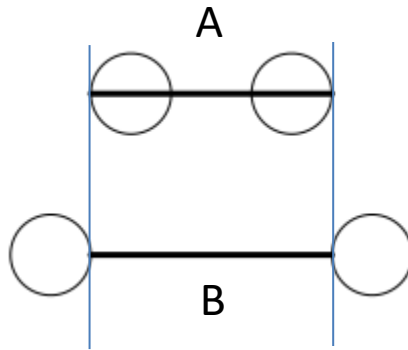
## The Eye-Brain Barrier

Perceptual illusions give some guidance on what **not to do** in data graphics

# Illusions: Length

Surrounding **context** matters in judging the **length** of objects.

Which **line** is longer? Or are they the same?

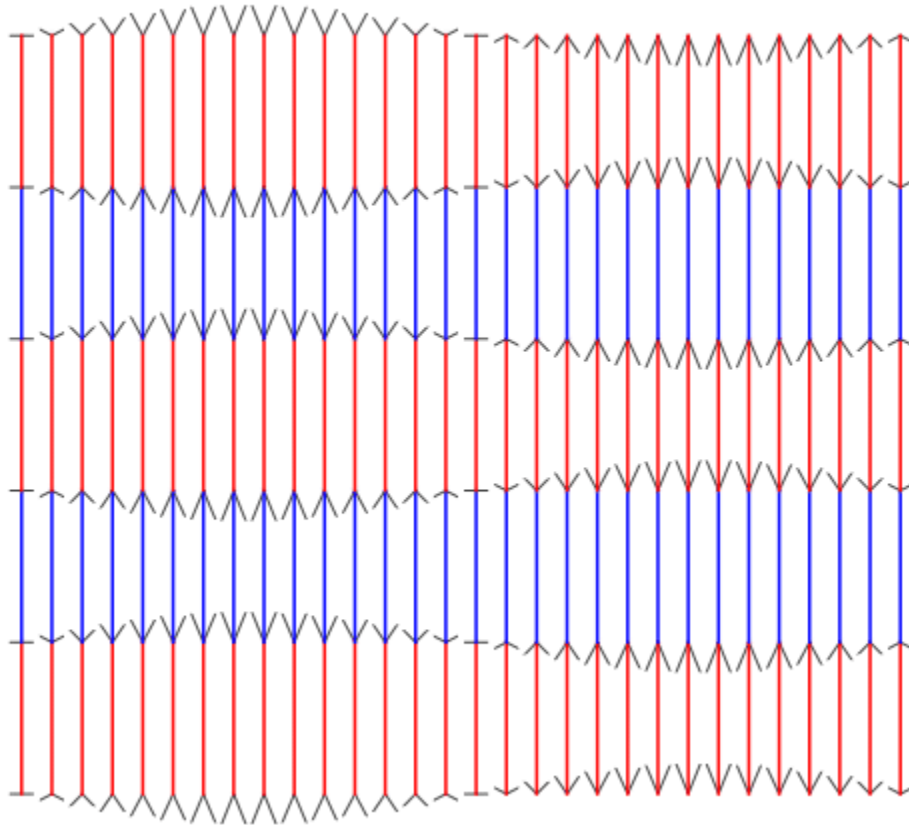


Surrounding context pulls perception of length in its direction  
This is the famous **Müller-Lyre** illusion



# Are these lines changing length?

Sarcone's Dynamic Müller-Lyer Illusion  
involving sinusoidal wave motion with dynamic color spreading effects



1. Focus your eyes on the center, and describe the pattern
2. Now, focus on **red bar** at top, right. Is it changing?
3. Now, focus on the **gray arrows** surrounding these bars. What do you see?

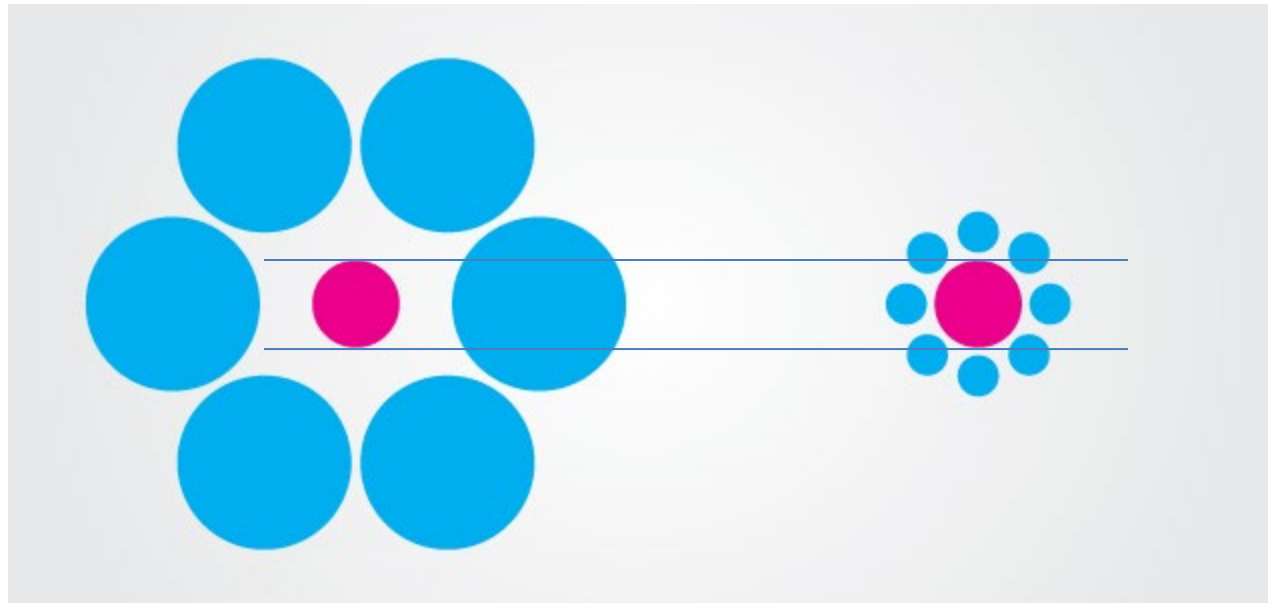
Arrows are often useful in data graphs.

Always use **same directions & angles** for arrow heads  $\longleftrightarrow$

# Illusions: Area

Surrounding **context** matters in judging the **area** of objects.

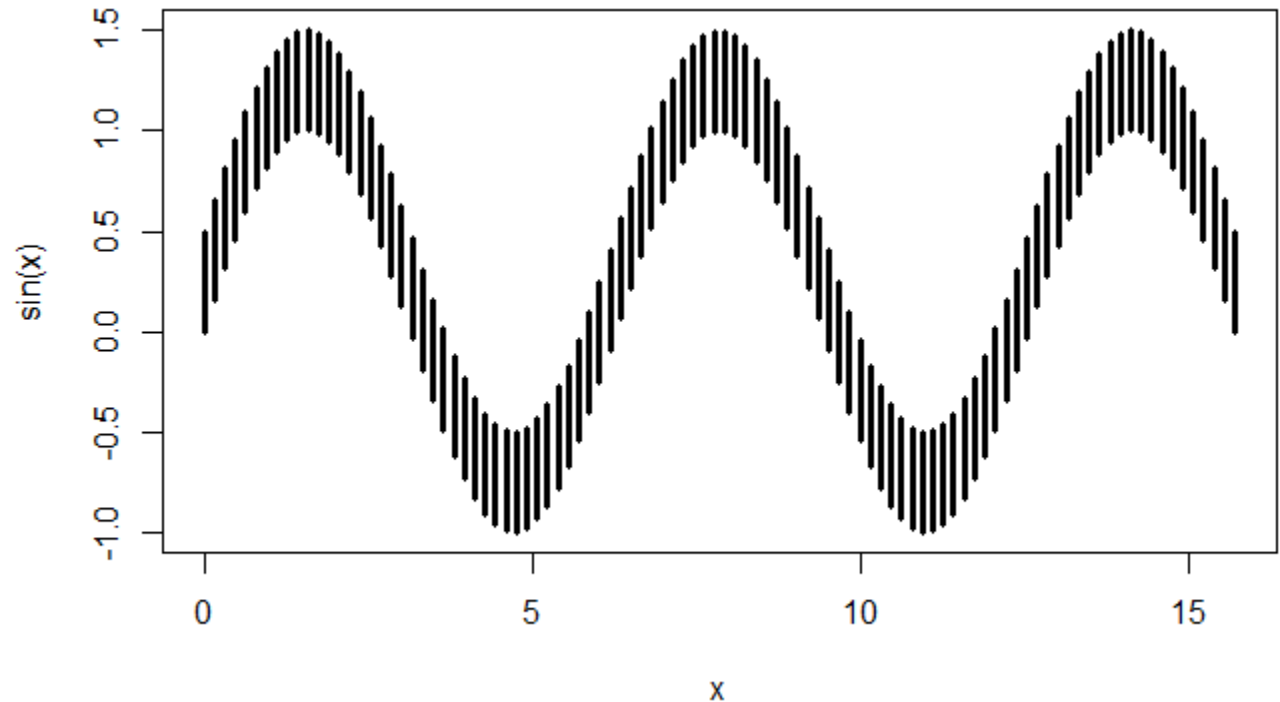
Which **red** circle is larger? Or are they the same?



Surrounding **context** pulls perception of area against the background  
This is often called the **Ebbinghaus** illusion or the **Tichener** illusion

# Illusions: Length

Which of the bars are longer? Or, are they all the same length?

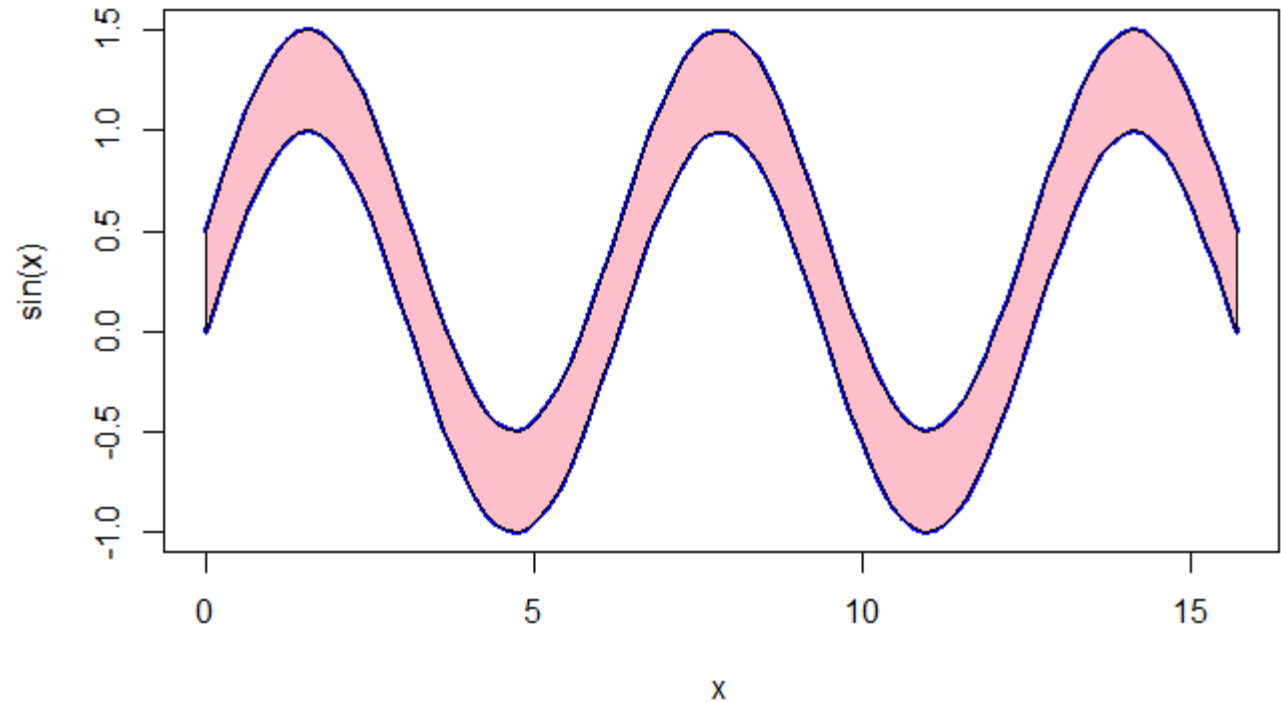


R code:

```
x <- seq(0, 5 * pi, length.out = 100)
w <- 0.5
plot(x, sin(x), ylim = c(-1, 1 + w), type = "n")
segments(x0 = x, y0 = sin(x), y1 = sin(x) + w, lwd = 3)
```

# Illusions: Difference

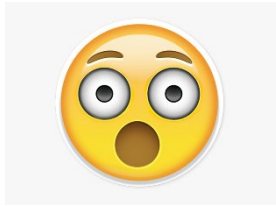
Where are **differences** between curves are larger? Or, are they all the same?



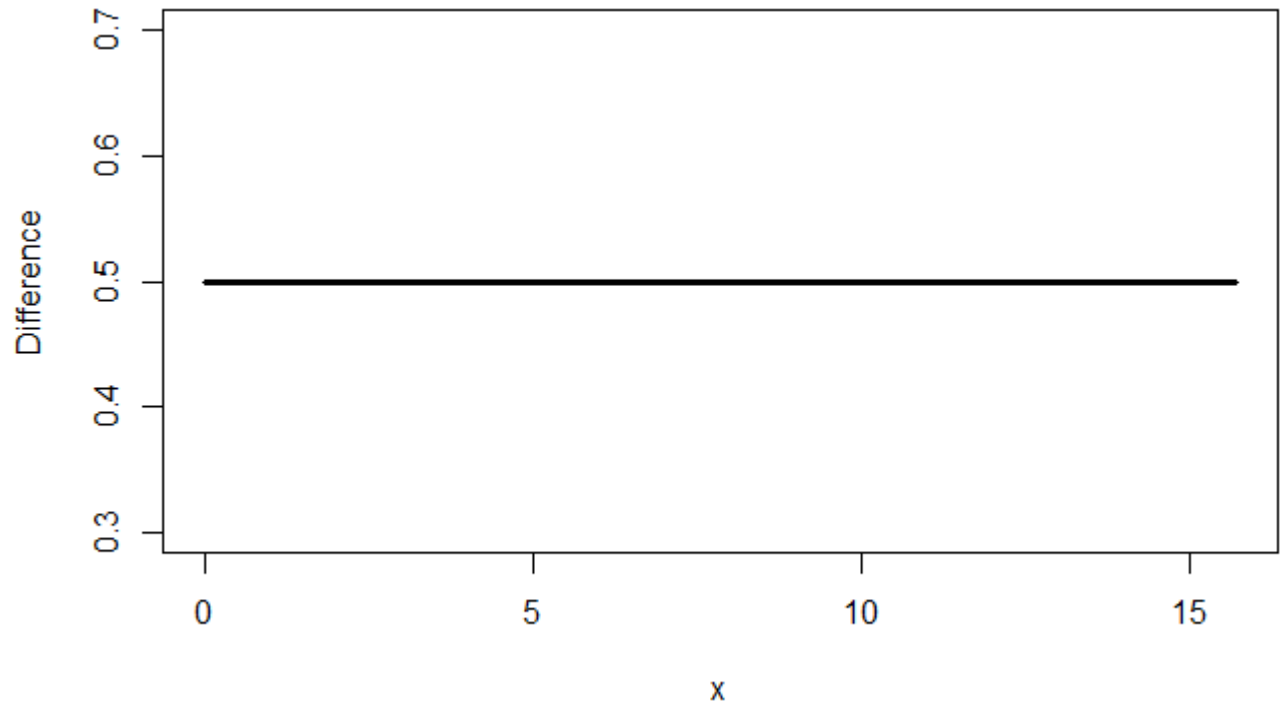
This is sometimes called the “sine illusion”

# Illusions: Difference

Plotting the **difference** directly gives the answer.



OMG!

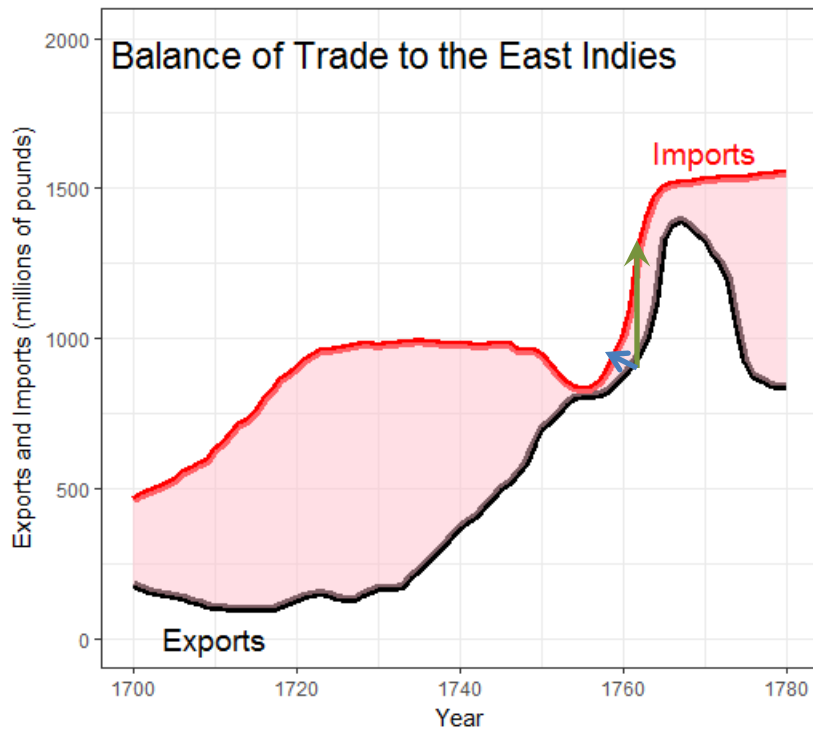


Why does this matter?

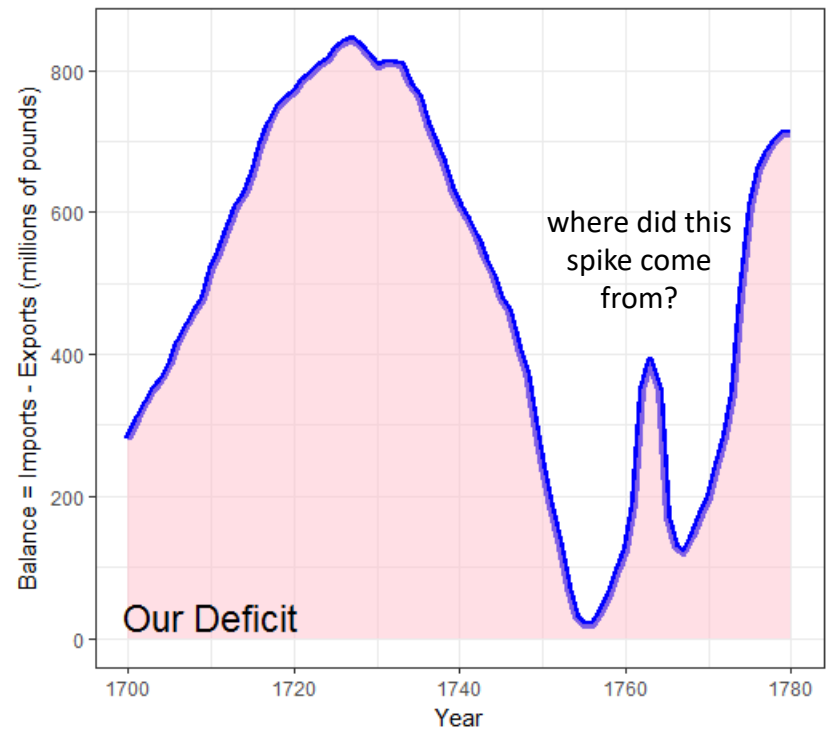
# Distances between curves

Playfair didn't know that judgments of distance between curves are **biased**  
We tend to see the **perpendicular** distance rather than the **vertical** distance

Original graph

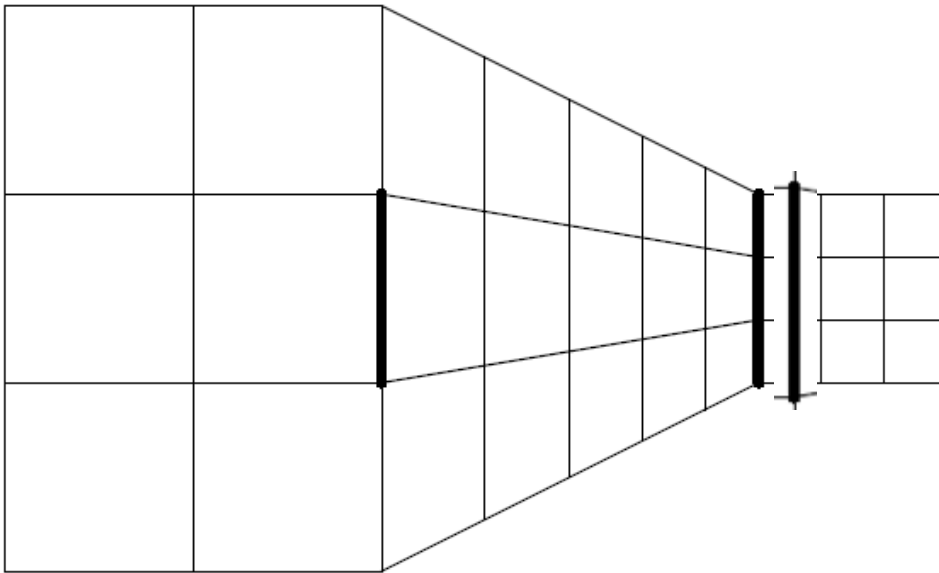


Plot of difference

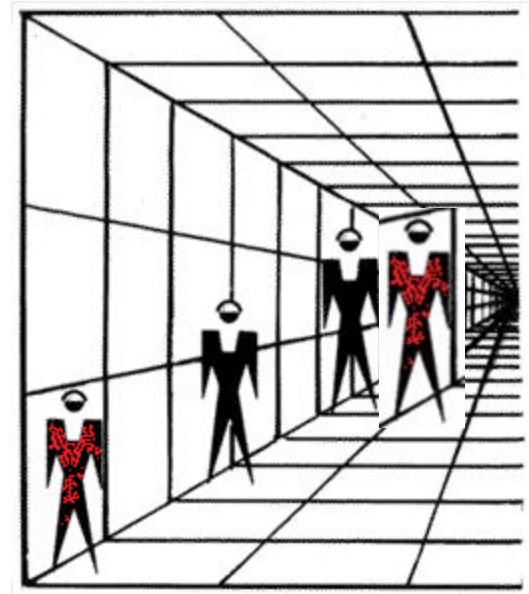


# Illusions: Perspective

Which **thick** line is longer? Or, both the same?



Which figure is tallest?  
Or, all the same?

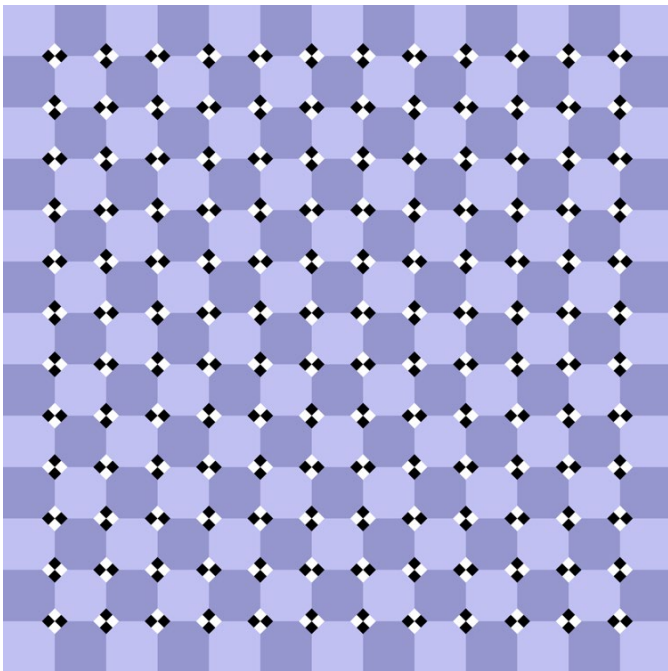


This is often called the **Ponzo** illusion: We judge the **size** of real-world objects relative to their **background** and **perspective**.

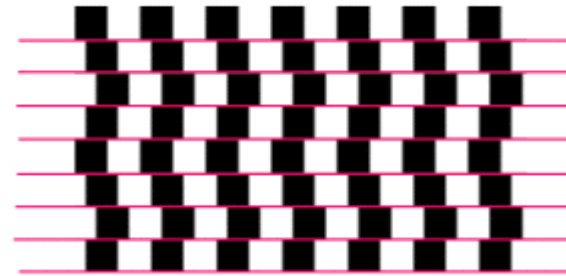
# Context illusions: Lines, shapes

Perception of elements of a scene is affected by context, background, etc.

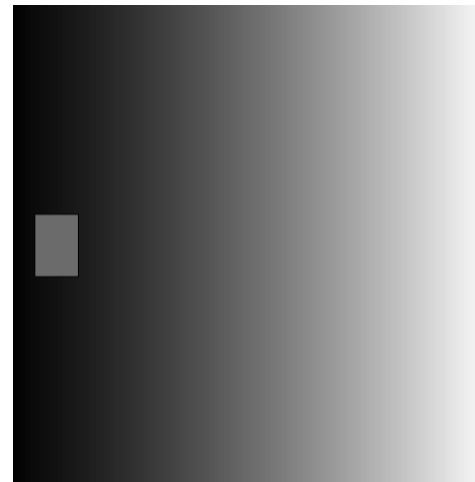
Are the squares straight or tilted?



Are the pink lines straight or curved?



Does the rectangle change in darkness?

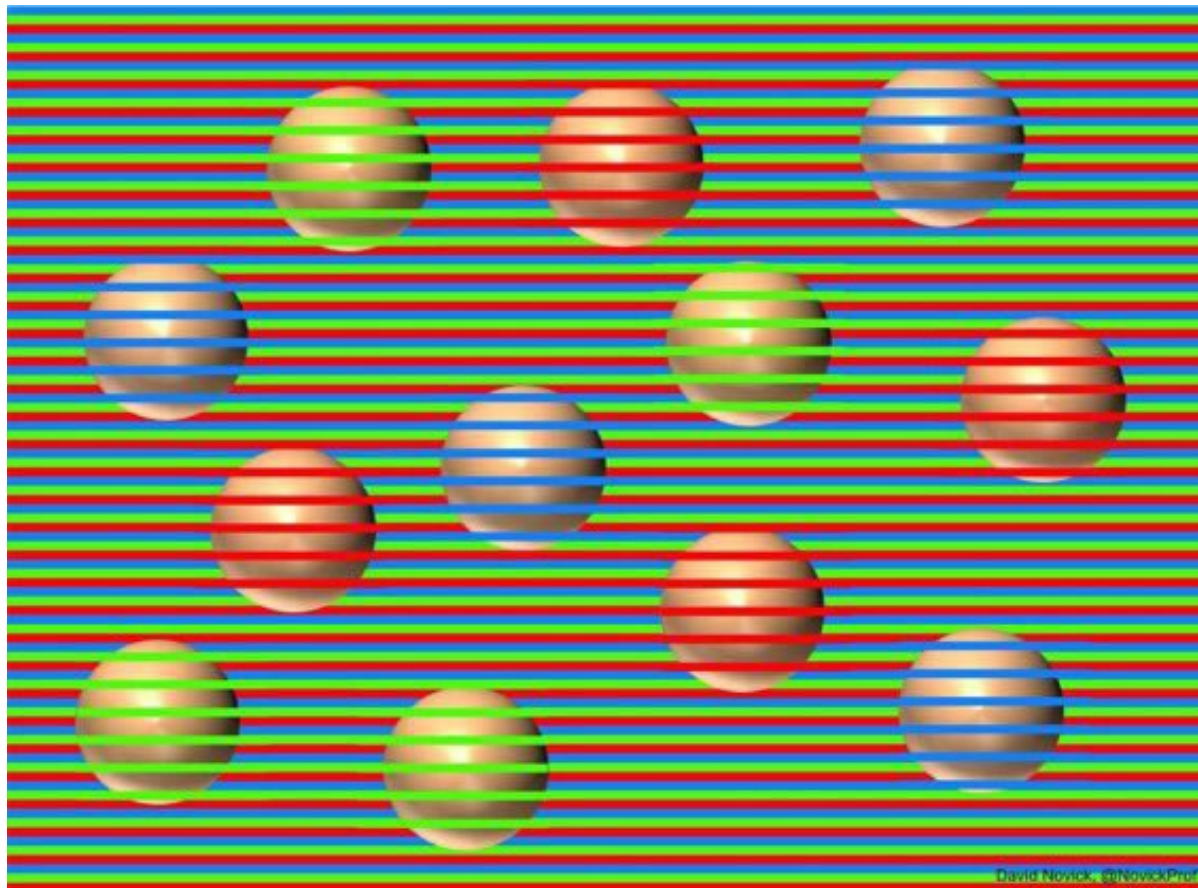


It is hard not to be fooled by these!



# Context illusions: Color

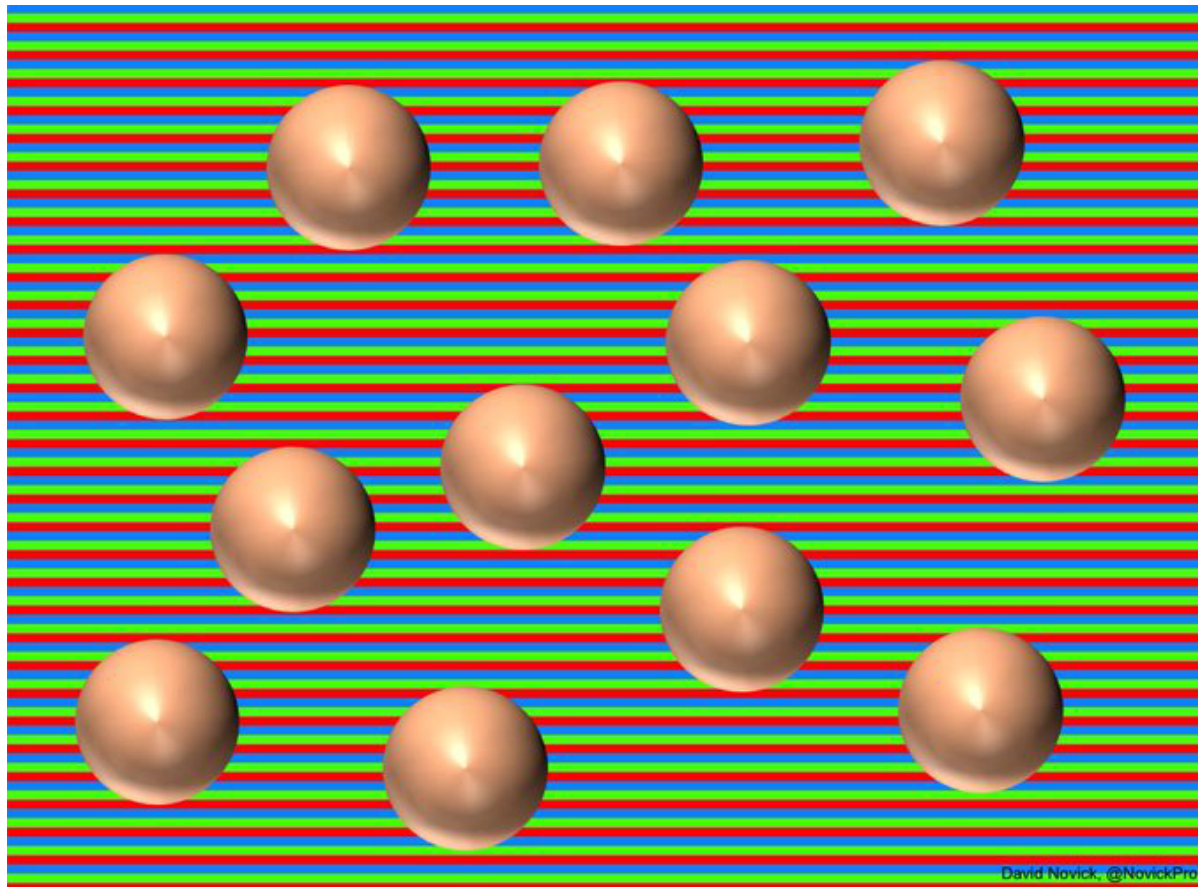
Are the balls **different colors** or are they all the **same color**?



David Novick, @NovickProf

# Context illusions: Color

Removing the foreground stripes shows them all the **same**  
(Munker illusion: perception of color is influenced by neighborhood )



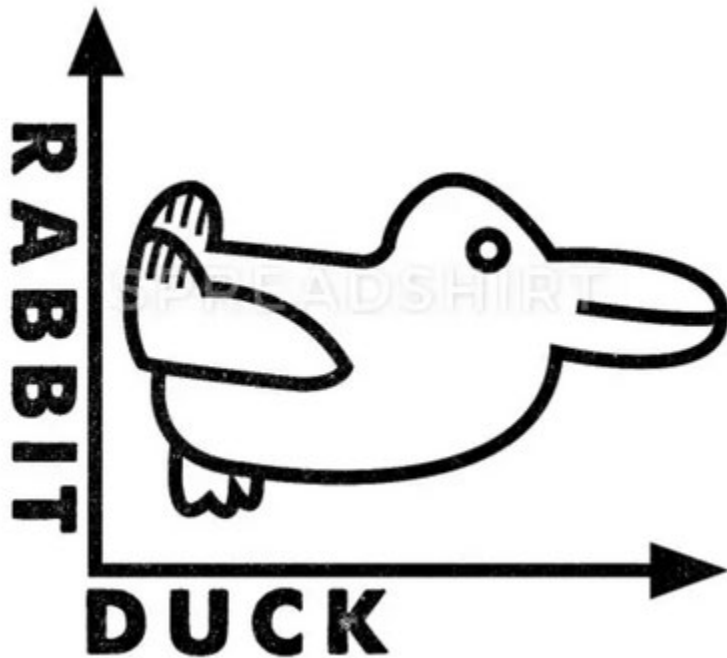
David Novick, @NovickProf

More examples: <http://engineering.utep.edu/novick/colors/>

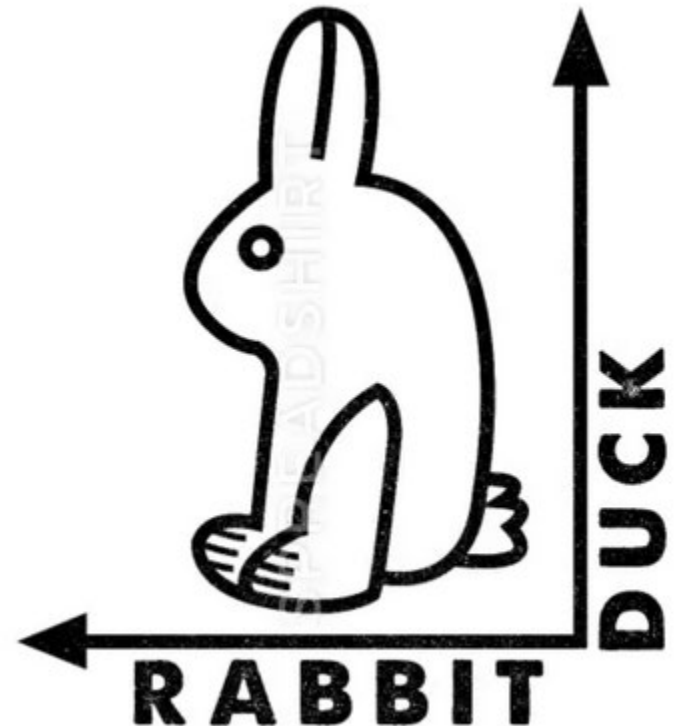
# Illusions: Semantic/cognitive

Perception of object figures often shows a preference for **orientation** in nature

Is this image a duck, or a rabbit?



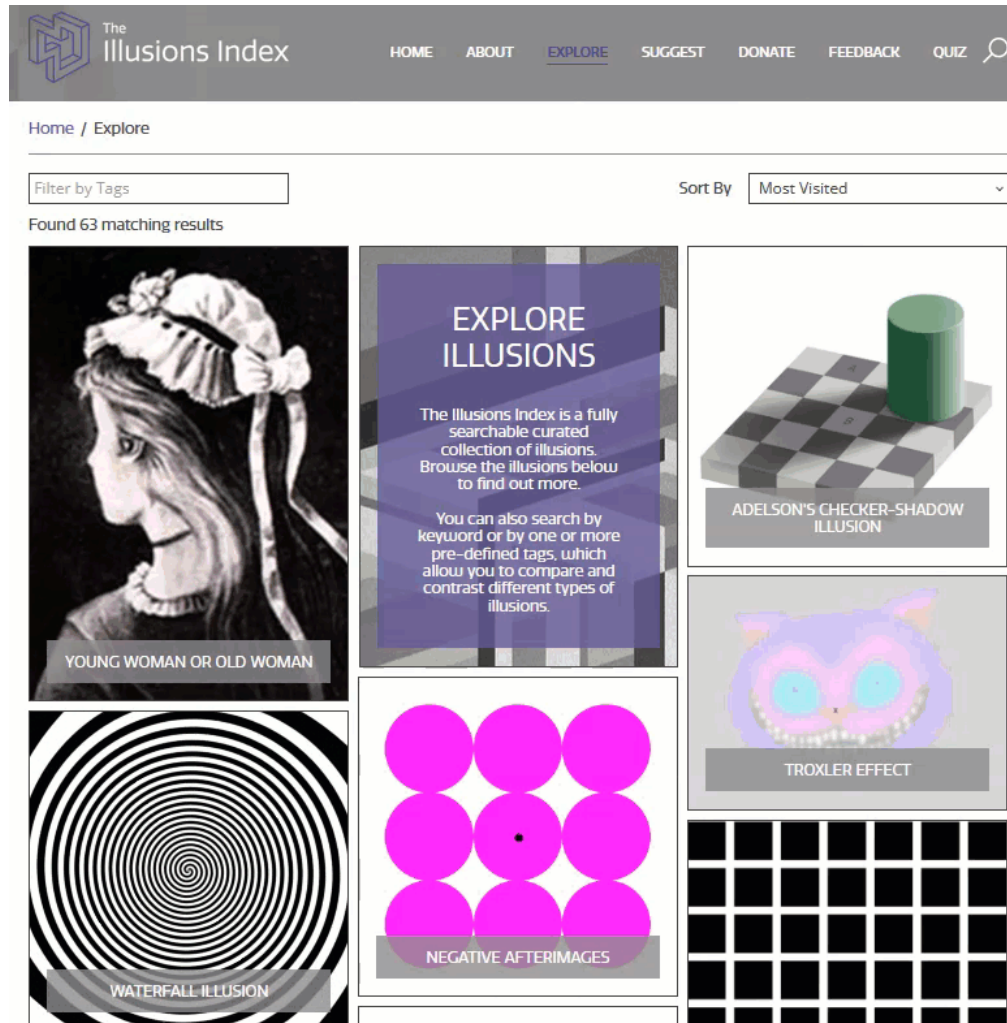
Duck or rabbit?



# Catalog of visual/auditory illusions

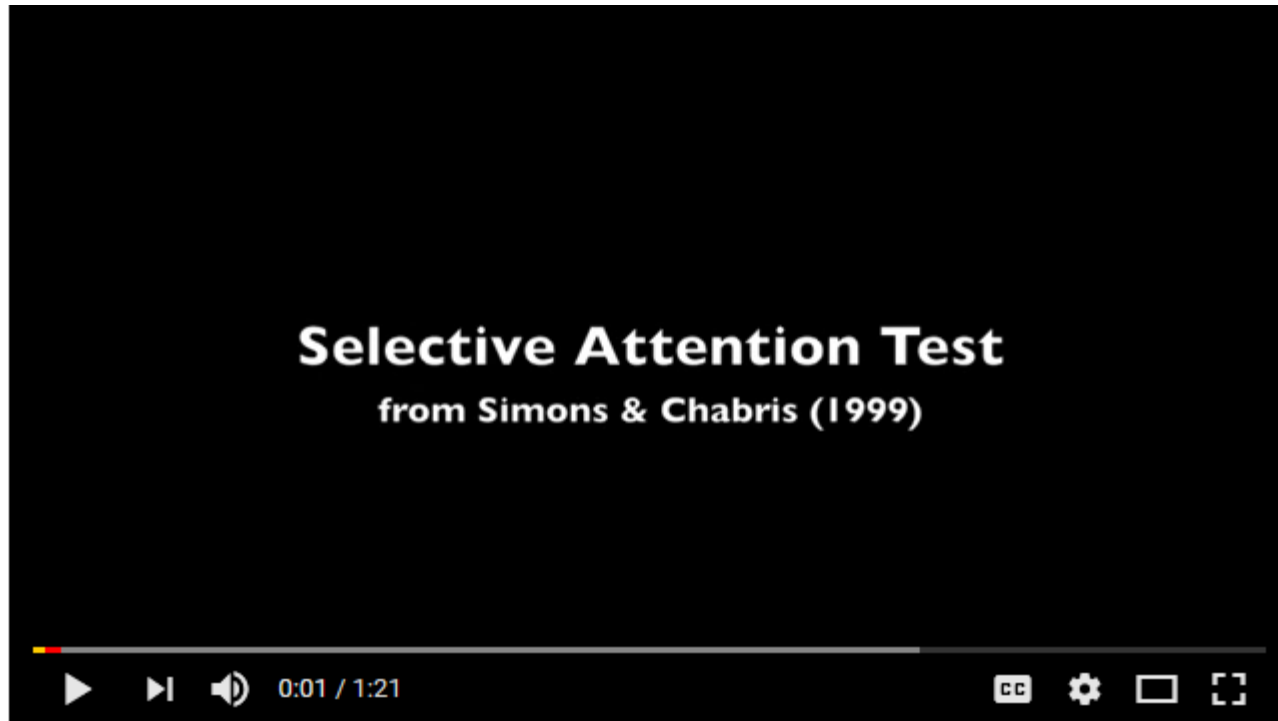
<https://www.illusionsindex.org/illusions>

A large collection of illusions with references to research studies.



# Selective attention

Attention strongly **focused** on some feature(s) **steals** attention from others



<https://www.youtube.com/watch?v=vJG698U2Mvo>

# Magnitude estimation

How large are transport accidents?

How much bigger than non-transport accidents?

transport

non-transport



Estimation of **length** or ratios of length are more accurate than the same judgments of **area**.

# Area vs. length judgments

Easy: Which is **larger**— South Africa or Egypt?

Harder: **How much larger** is South Africa than Egypt? (% or ratio)

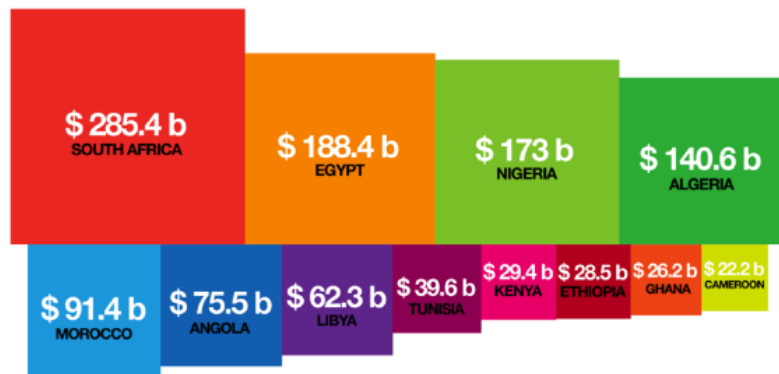
## African Countries by GDP

### TOP COUNTRIES BY GDP IN U.S. \$ BILLIONS

Gross domestic product (GDP) refers to the market value of all final goods and services produced within a country in a given period (2005 - 2009).

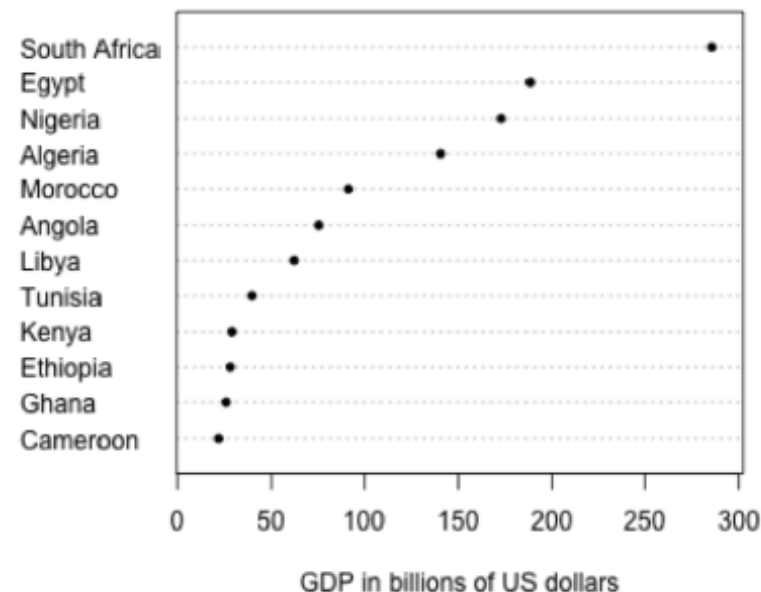
### GDP CALCULATION

private consumption + gross investment + government spending + (exports - imports)



Judgments here based on area

## African Countries by GDP



Judgment here based on position along a scale



# Theory: Stevens' Power Law

- How does perceived magnitude of a **sensation** relate to stimulus **intensity**?
- S. S. Stevens (1957) showed that, for many domains

$$\text{Sensation} \propto \text{Intensity}^p$$

- These provide ways to assess the **accuracy** of **magnitude estimation** for visual encodings
  - length: most accurate ( $p \approx 1$ )
  - area, depth: sub-sensitive ( $p < 1$ )
  - electric shock: hyper-sensitive ( $p > 1$ )
- But: graph perception is not always a matter of estimating **magnitudes**.

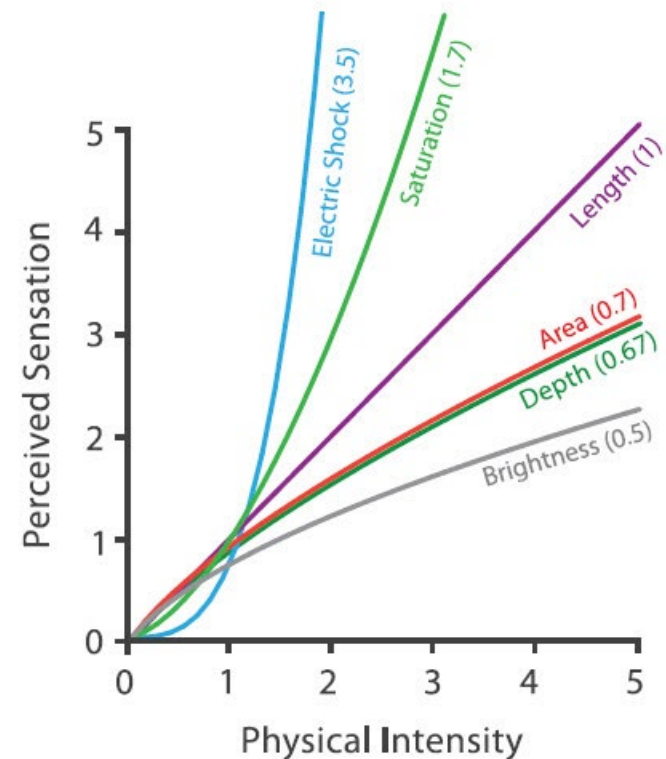


Fig. 5.7 from: Munzner, *Visualization Analysis & Design*

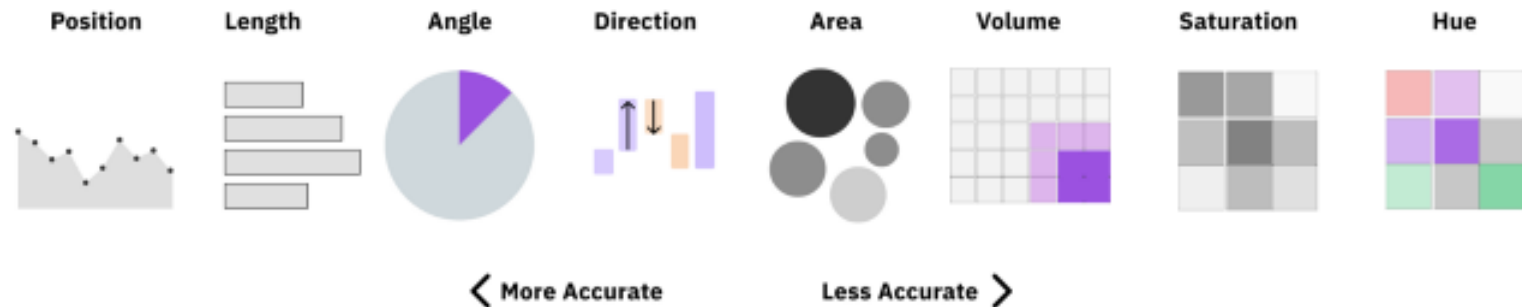


# Theory→Practice: Scale of accuracy

A commonly used “scale” of accuracy of magnitude judgments of relative size

- How much smaller/larger is A compared to B?

## Accuracy Of Visual Cues



Not necessarily the same for other tasks (Part-whole: What % is A of total?)

# Accuracy: Experimental evidence

Cleveland & McGill (1984) and later Heer & Bostock (2010) carried out experiments to assess the relative accuracy of magnitude judgments for different visual encodings

The task here is to estimate the %age of the smaller highlighted portion.

The details of these studies are interesting & important – more next week

The graph of these results is a great model for data display



Fig. 5.8 from: Munzner, *Visualization Analysis & Design*

# Encodings: Types & ranks

Based on this, Munzner (2015) proposes a ranking of visual attributes for **ordered** & **categorical** variables in data displays, with different **channels**

These hold when the task is to estimate a **magnitude**.

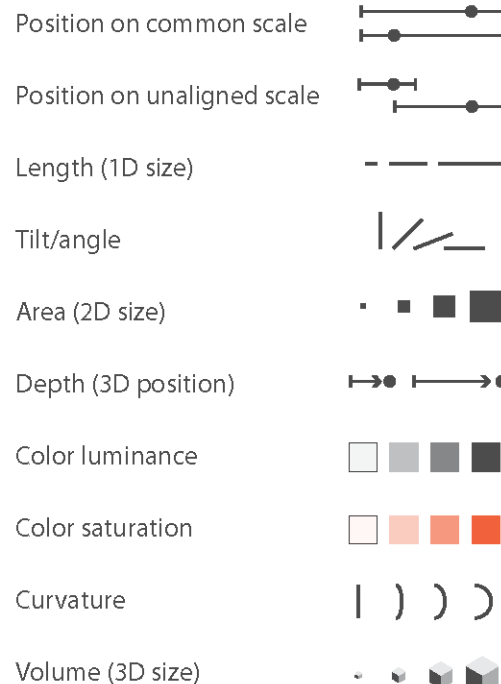
A different ranking may occur for other graph-based tasks.

**angle** (pie charts) – good for % of total judgments

**color** (mosaic plots) – good for pattern perception

Channels: Expressiveness Types and Effectiveness Ranks

## ➡ **Magnitude** Channels: **Ordered** Attributes



## ➡ **Identity** Channels: **Categorical** Attributes



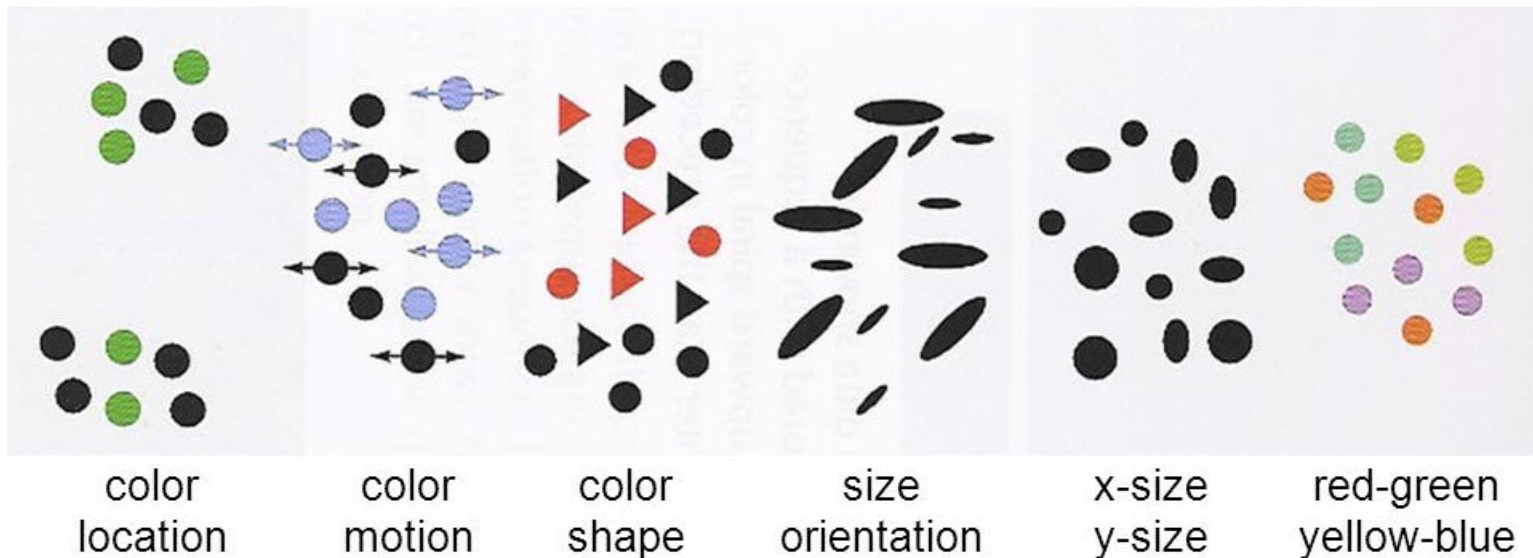
Fig. 5.6 from: Munzner, *Visualization Analysis & Design*

# Integral & separable encodings

- Some encodings can be viewed **independently**
  - two different variables **can** be decoded separately
- Some **combine** with each other to some degree.
  - different variables **cannot** be easily decoded separately

← Separable

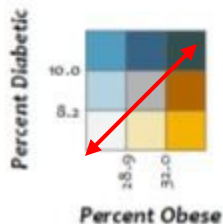
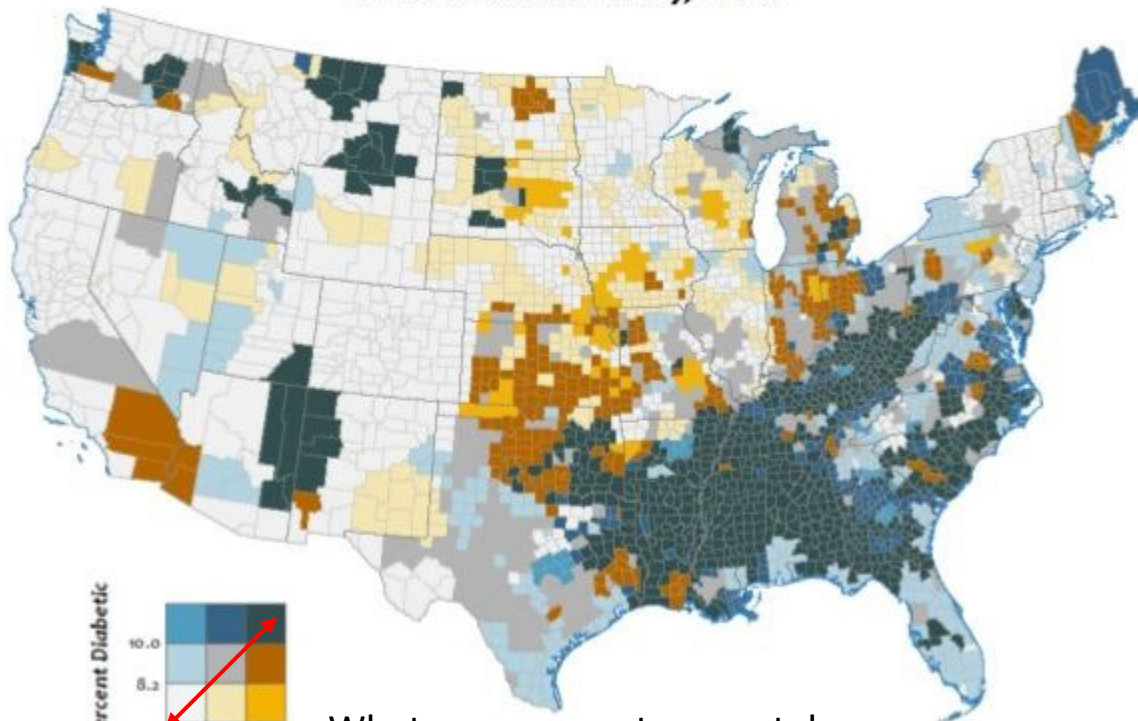
Integral →



# Integral dimensions

A bivariate U.S. county-level map showing:  
% diabetic (**saturation**) and % obese (**hue**)

*Diabetes and Obesity, 2009*



What we see most accurately

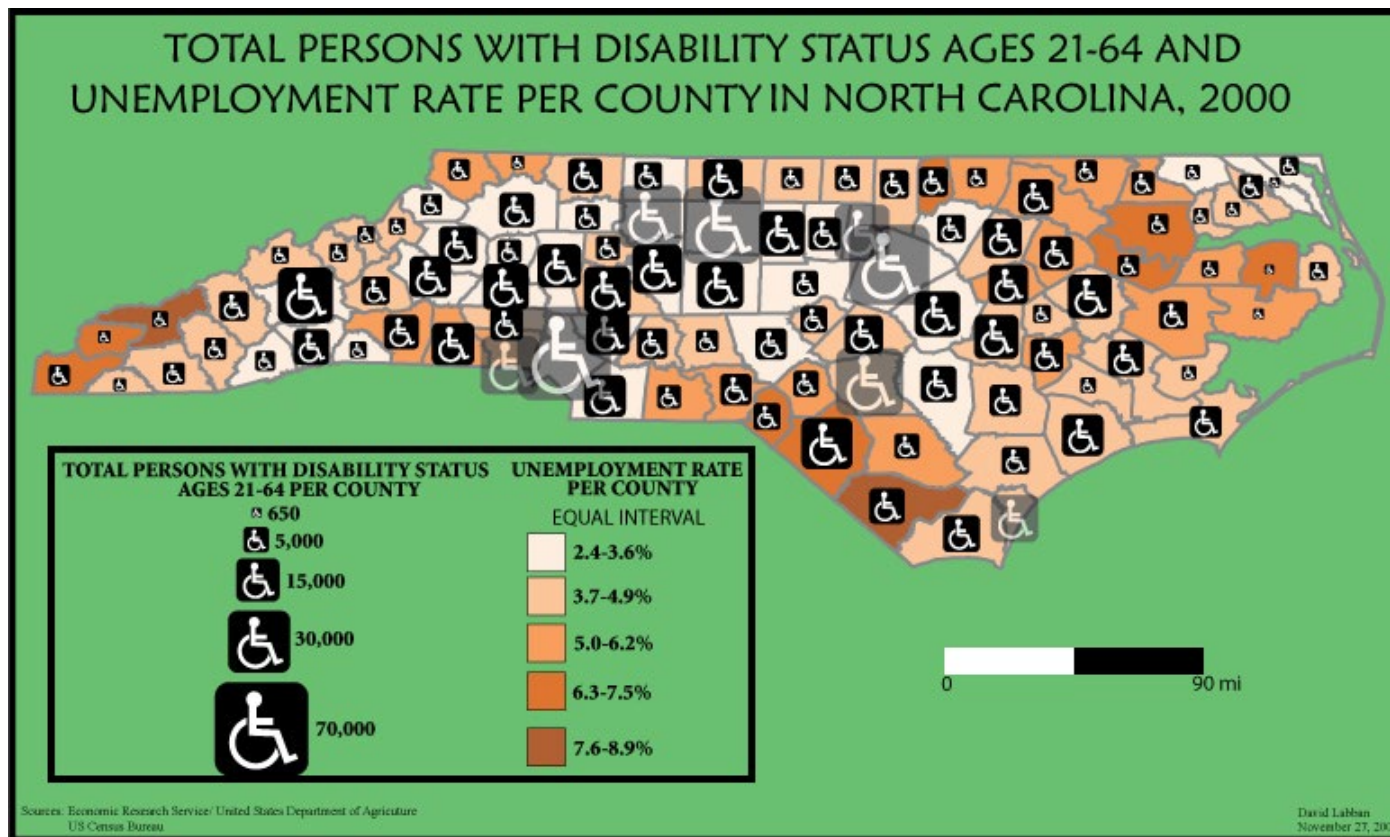
It is difficult to see variations in **diabetes** separately from **obesity**

The eye is tuned to the positive correlation between these dark (**blue**, **tan**) vs. light color

# Separable dimensions

Bivariate map of N. C.: disability (icon **size**) and unemployment rate (background **saturation**)

- These can be seen separately
- (However, total disability is confounded with population density)



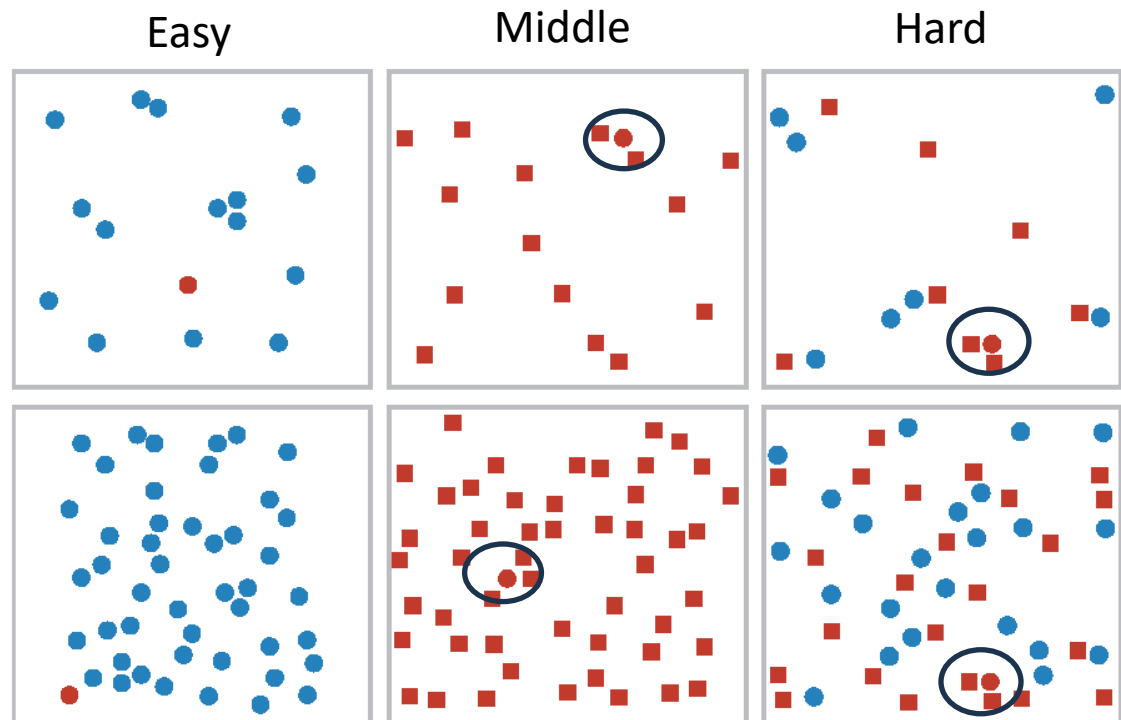
# Anomaly detection (outliers)

Find the red dot ● in each of the following displays

- This task is easiest when all the rest are blue dots ●
- Next easiest when **only shape** distinguishes the red dot ■
- Hardest when both **color and shape vary** ● ■

Sometimes called  
“popout” effect.  
Not a good term.

This is important in  
designing graphs to  
**highlight** some points.





# Anomaly detection

For each display, find the anomaly shown at the left

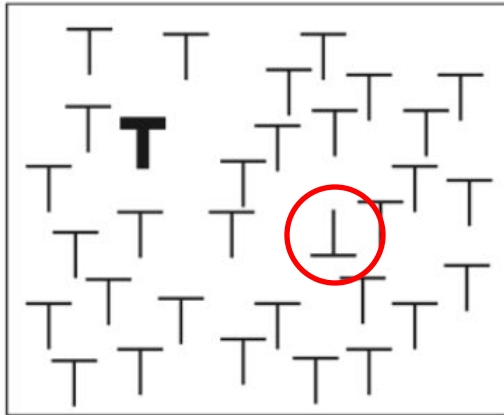
Color and shape: What is easy or hard depends on the background



difficult



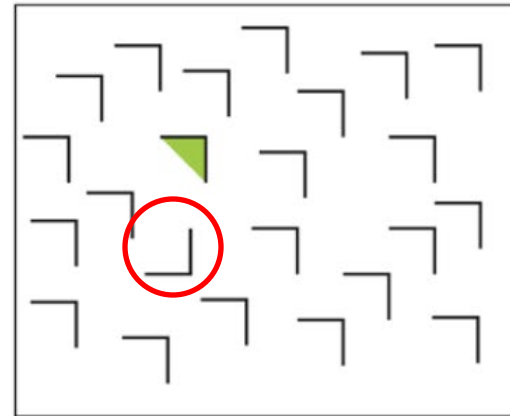
easy



difficult



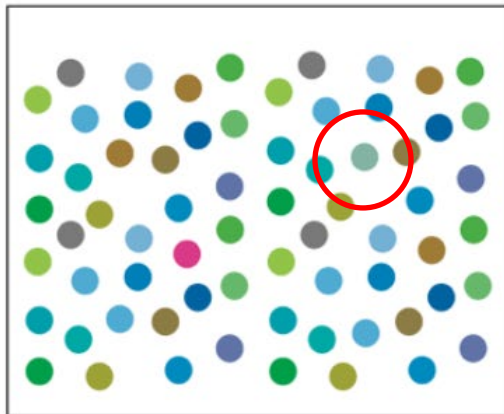
easy



difficult



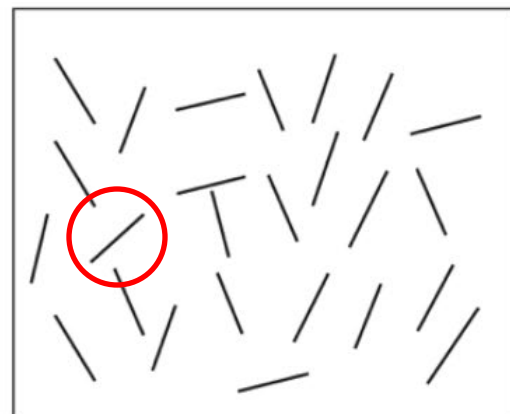
easy



difficult

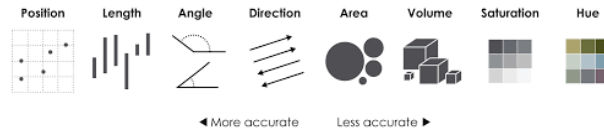


easy





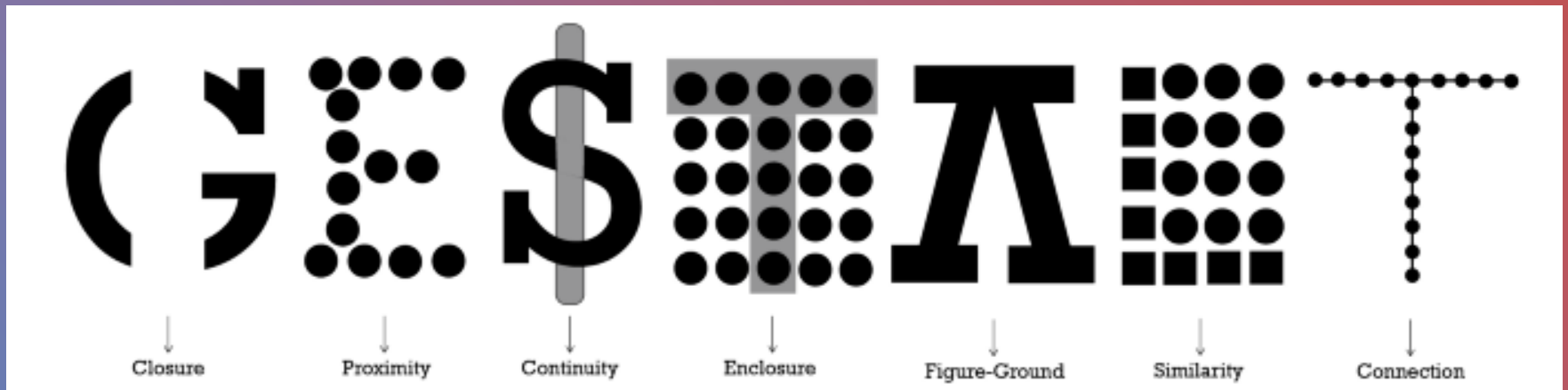
# Encodings: Lessons



- Ordered variables
  - Prefer encodings at the top of the hierarchy (position along a scale) to those at the bottom (color saturation, curvature)
- Favor **separable** encodings
  - Use color and another attribute--- shape, size, orientation ● ■
  - Don't overload symbols--- probably two at most
  - Avoid mixing two aspects of color or two aspects of shape
- Small multiples
  - Reduces the need for multiple encodings within a panel
  - But, makes direct comparison more difficult
- **Highlighting**: to draw attention to one group, use a pre-attentive attribute

# Encodings: Lessons

- Best to show quantitative variables with **position** or **length**
- Bar charts:
  - Best encoding via length → start at 0
  - Avoid stacked bars (not aligned), where possible
- Dot charts:
  - Best encoding **via** position along a scale → start at 0
- Frequency data:
  - area/color encoding to show patterns
  - sqrt or log scale often useful to show magnitude
- Color: choose **sensibly ordered hues** or saturation
- Arrangement
  - make comparisons easier by placing things to be compared nearby



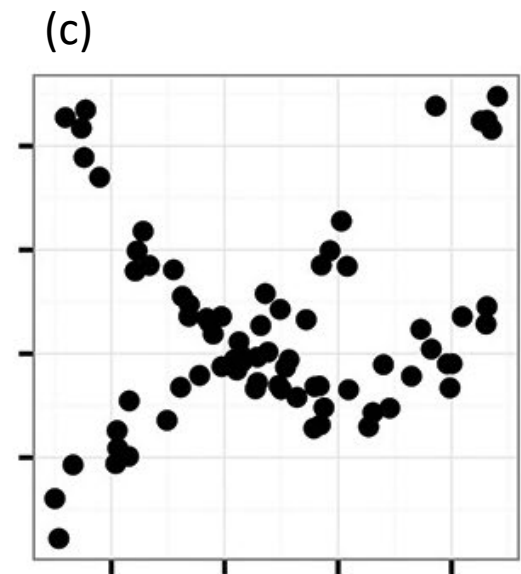
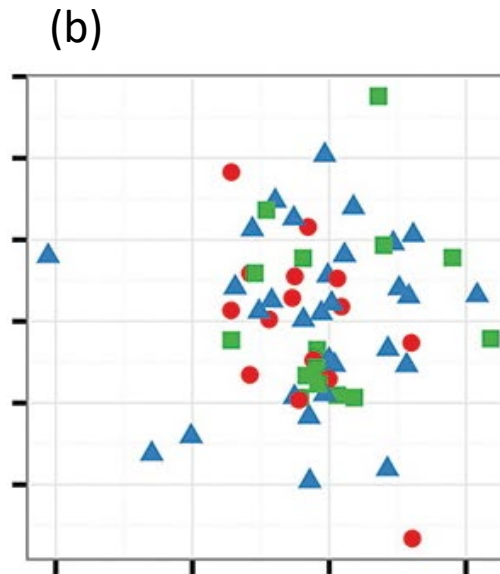
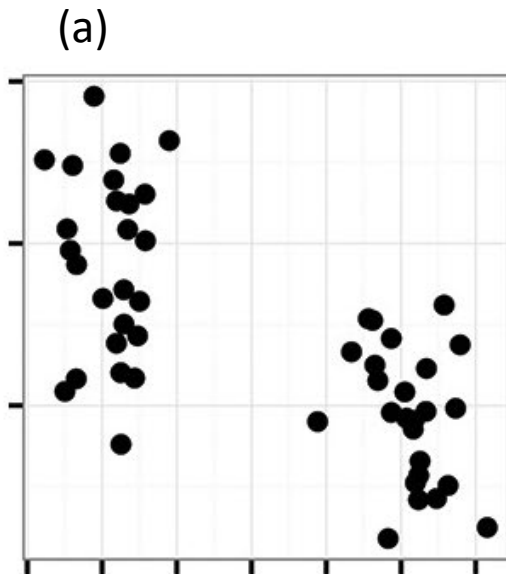
# Gestalt principles of graph perception

# Gestalt principles

- Perception as **top-down** process governed by holistic principles. “Gestalt” = “form”
  - **proximity**: elements close together likely to belong to the same unit
  - **similarity**: more common visual elements increases belonging together
  - **good continuation**: elements that blend together are likely in the same unit
  - **common region**: elements in the same region likely belong together
  - **closure**: elements that make a meaningful whole belong together

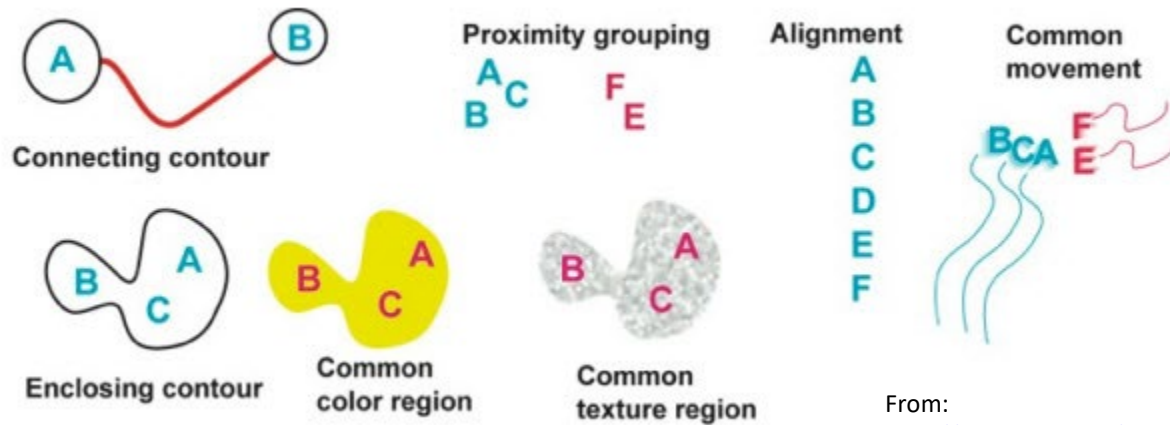
# Gestalt principles

- (a) **proximity** creates impression of 2 groups
- (b) **similarity**: 3 groups via color & shape
- (c) **good continuation** gives impression of 2 groups



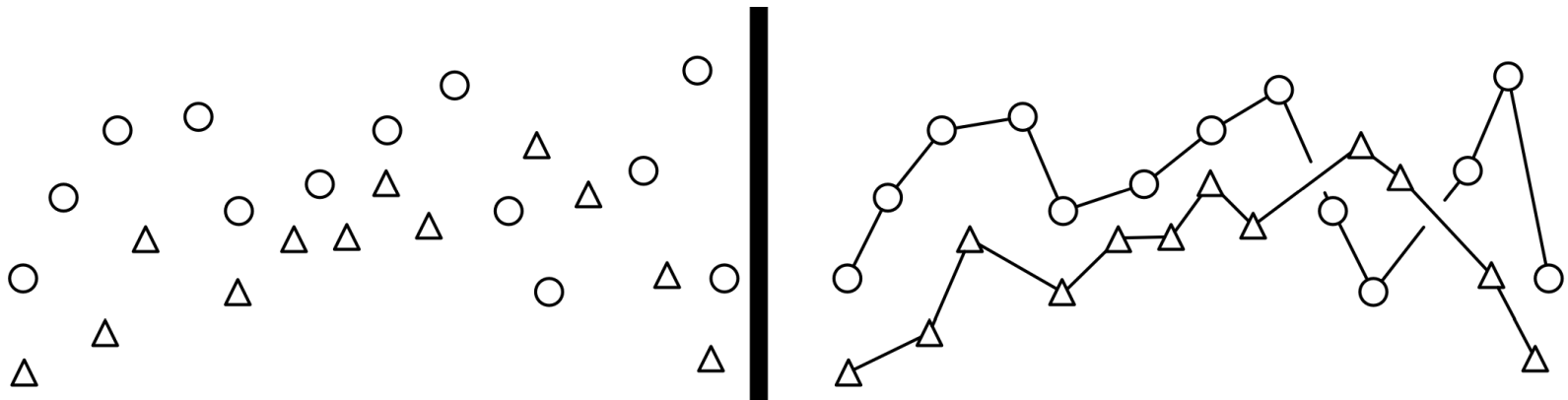
# Using Gestalt ideas in data graphs

Make grouping apparent



From:  
<http://blog.yhwu.me/notes/visualizations/cs171.html>

Why lines are good in time series graphs



# Closure

The Gestalt brain “wants” to make perception simpler by joining up disparate elements into coherent, meaningful wholes

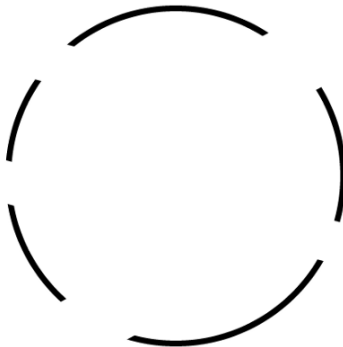
Logos: Empty space or a symbol?



Closure



5 arcs or a circle?



blobs or an animal?

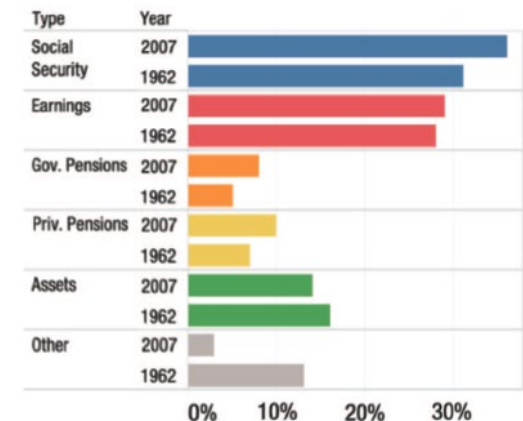
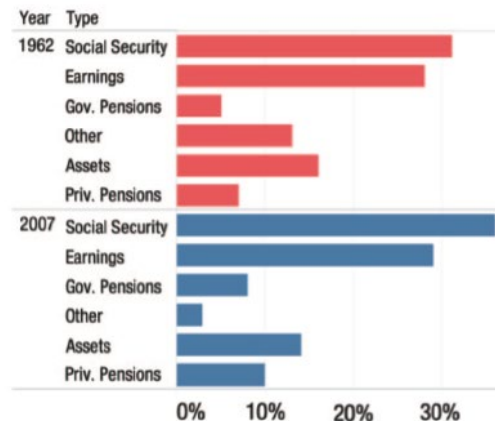


Q: Are these top-down or bottom-up?

# Visual grouping & comparisons

Combination of color & proximity grouping → different visual comparisons are easier or harder

Did earnings increase or decrease from 1962-2007?



What was the most frequent baseball term?  
Which terms go together?

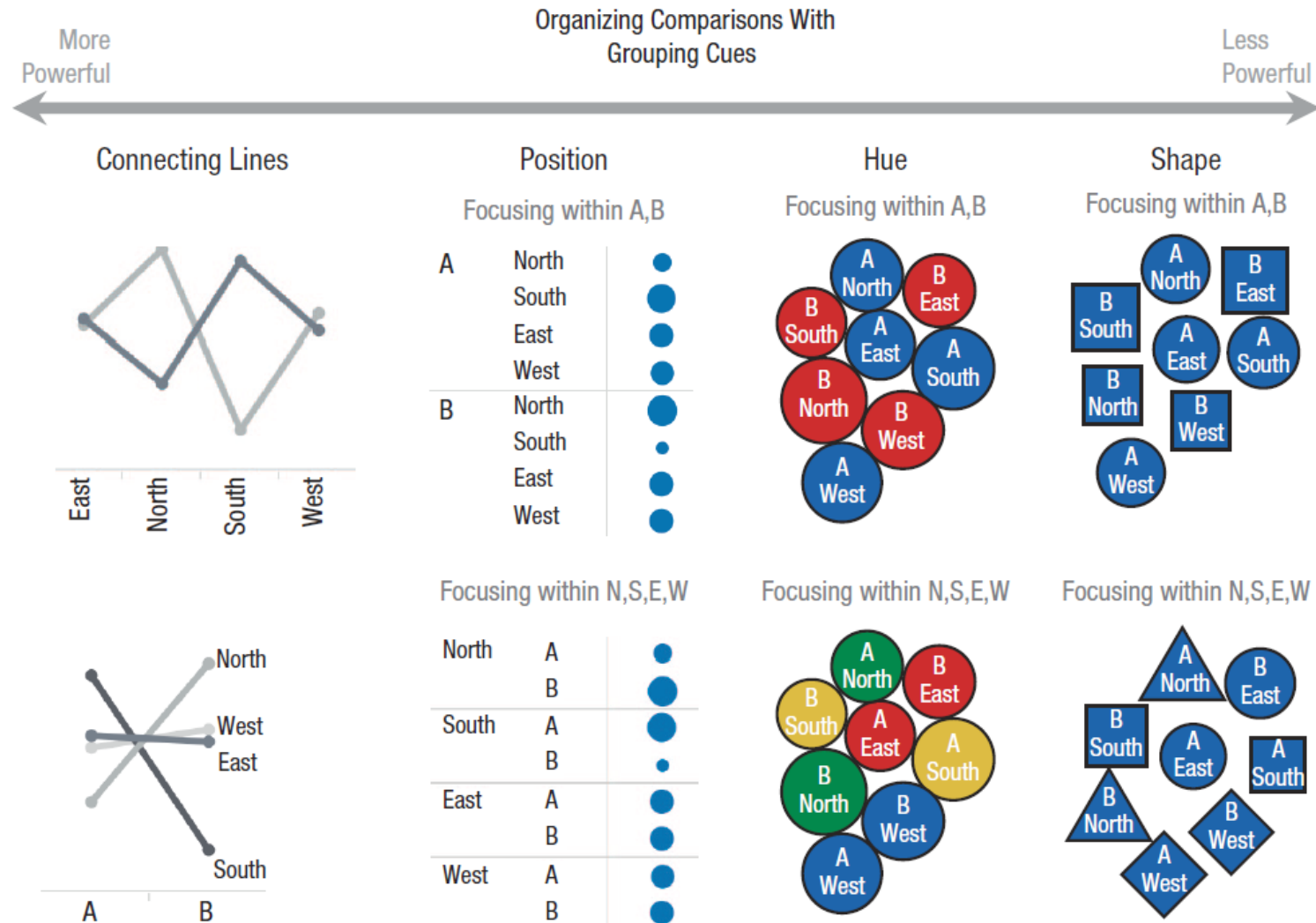


Comparisons in a word cloud:

- weakly controlled by **color**;
- better controlled by **proximity**



# Visual grouping & comparisons

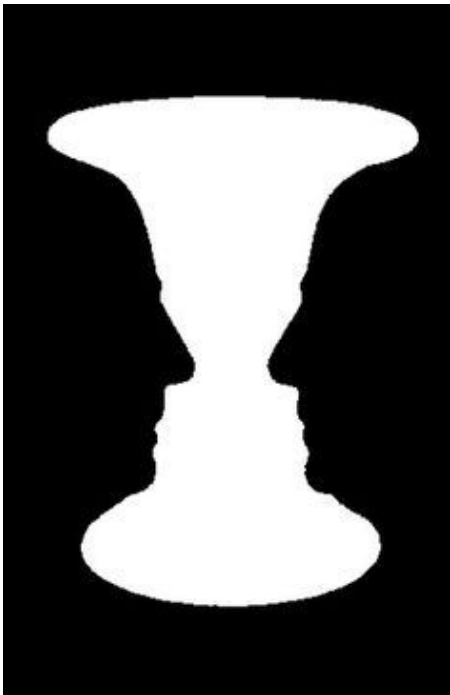


**Fig. 8.** Several grouping cues that can control how data values are compared. Connecting lines are particularly powerful cues, followed by proximity, color, and shape (Brooks, 2015).

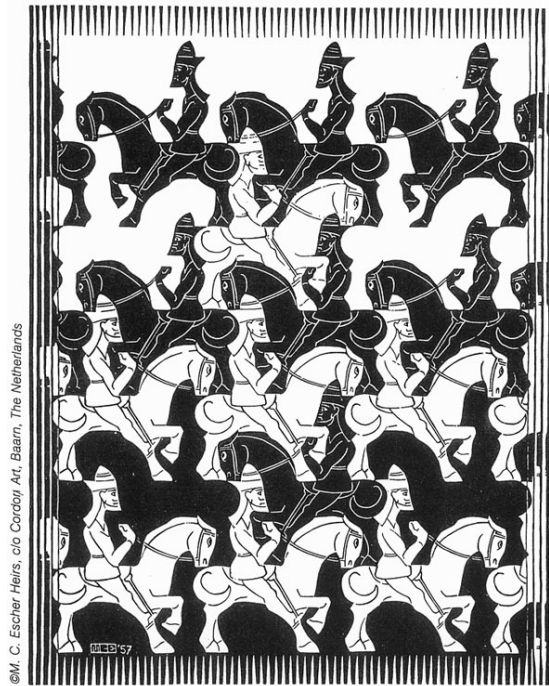
# Figure - Ground

What is the figure? What is the background?

Face or vase?



Black or white soldiers?



Face or park?



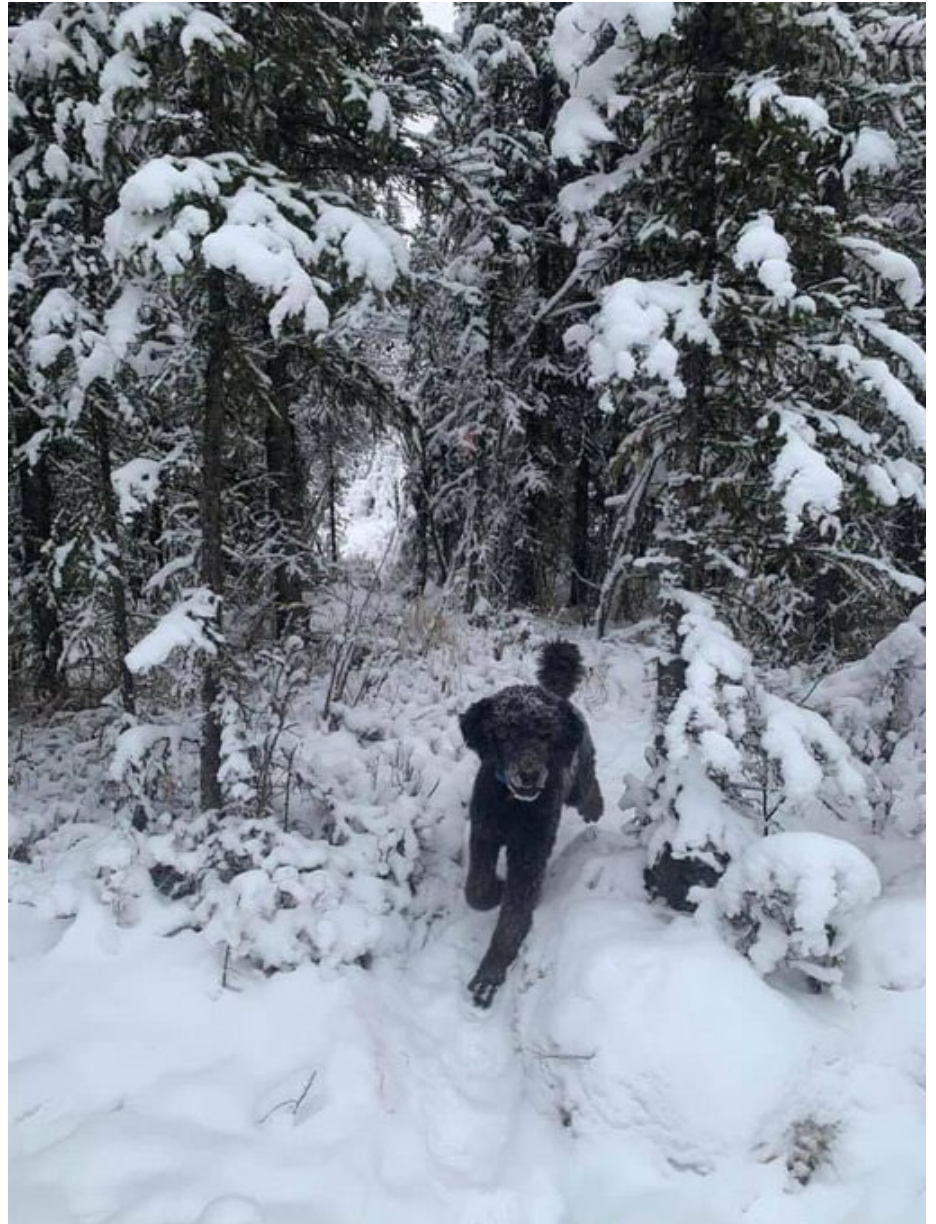
These examples all use different techniques to create ambiguous figures

# Ambiguous figures: Priming

Can you see the **poodle** in this scene?

What about the **man**?

**Semantic priming:** Suggestion increases likelihood of perception



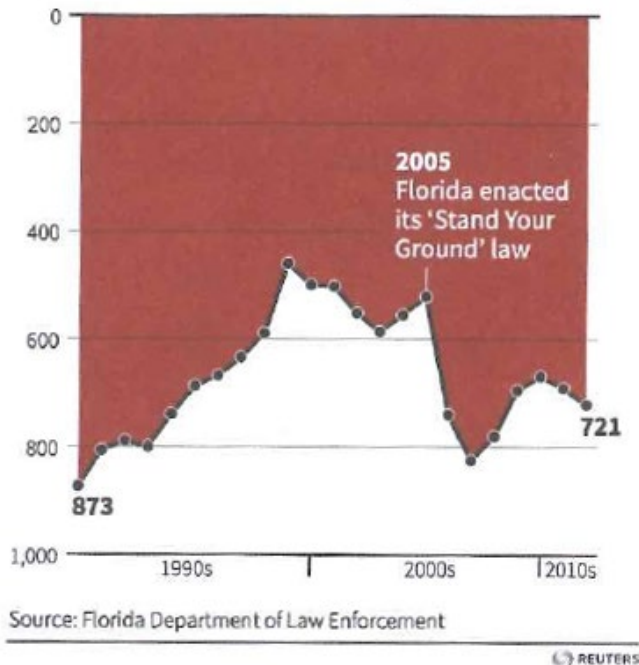


# Figure - Ground

This graph inverts the y-axis, and shades the area above the curve

## Gun deaths in Florida

Number of murders committed using firearms

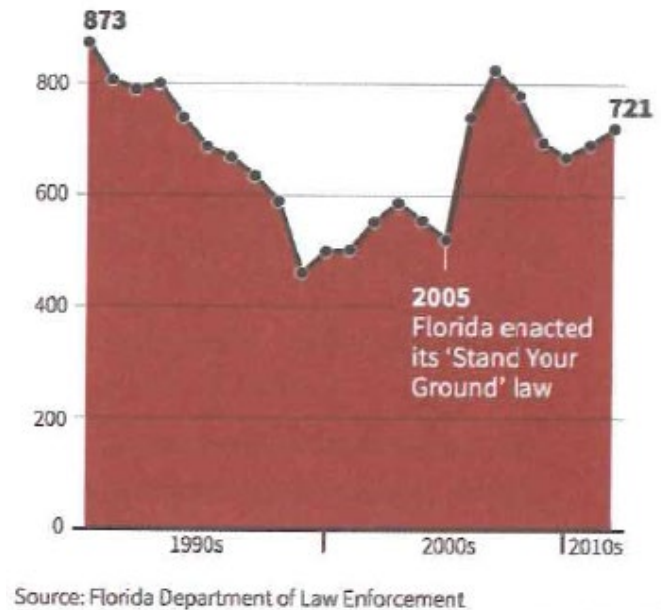


We tend to see 1999 & 2005 as high points

A more conventional version of the same graph

## Gun deaths in Florida

Number of murders committed using firearms



Gun deaths increased after the  
'Stand your ground' law

# Discussion

What perceptual features or principles are involved in your reading or understanding of these figures?



What about this?

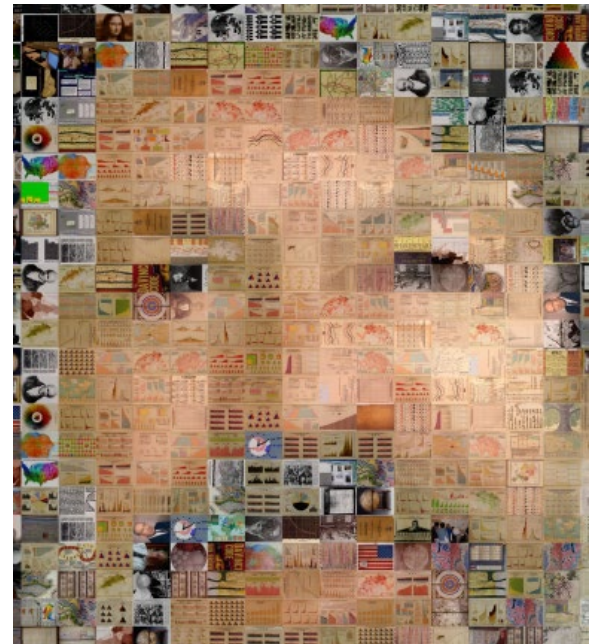


Top-down ?  
Bottom-up?  
Gestalt?

or this?



or this?





# Color



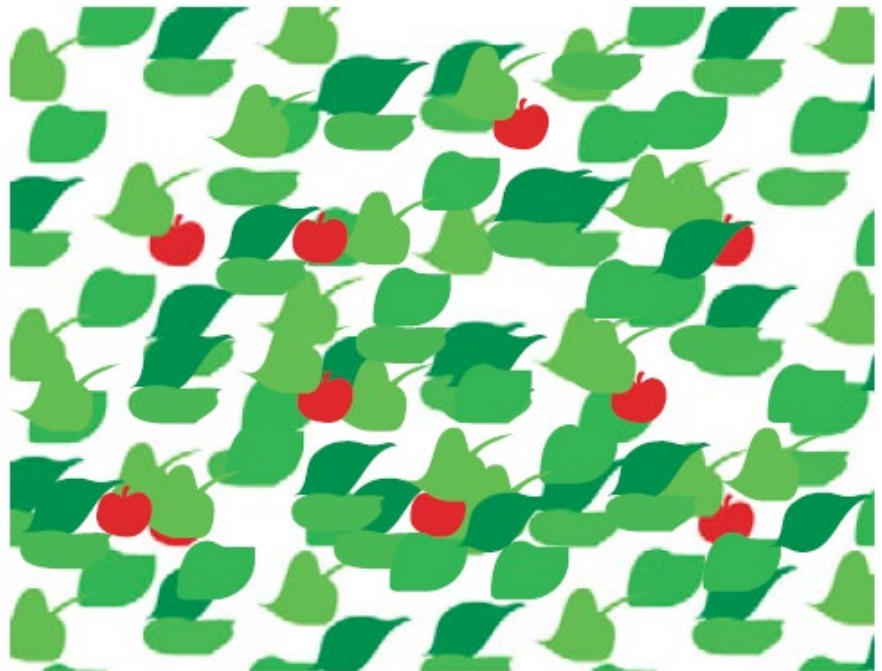
# Color: Functions in data graphics

Color serves to: **highlight**, **identify**, and **group elements** in a visual display

Find the cherries in this display:



Color acts as a **preattentive** attribute here



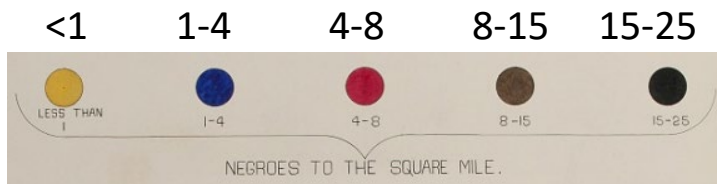
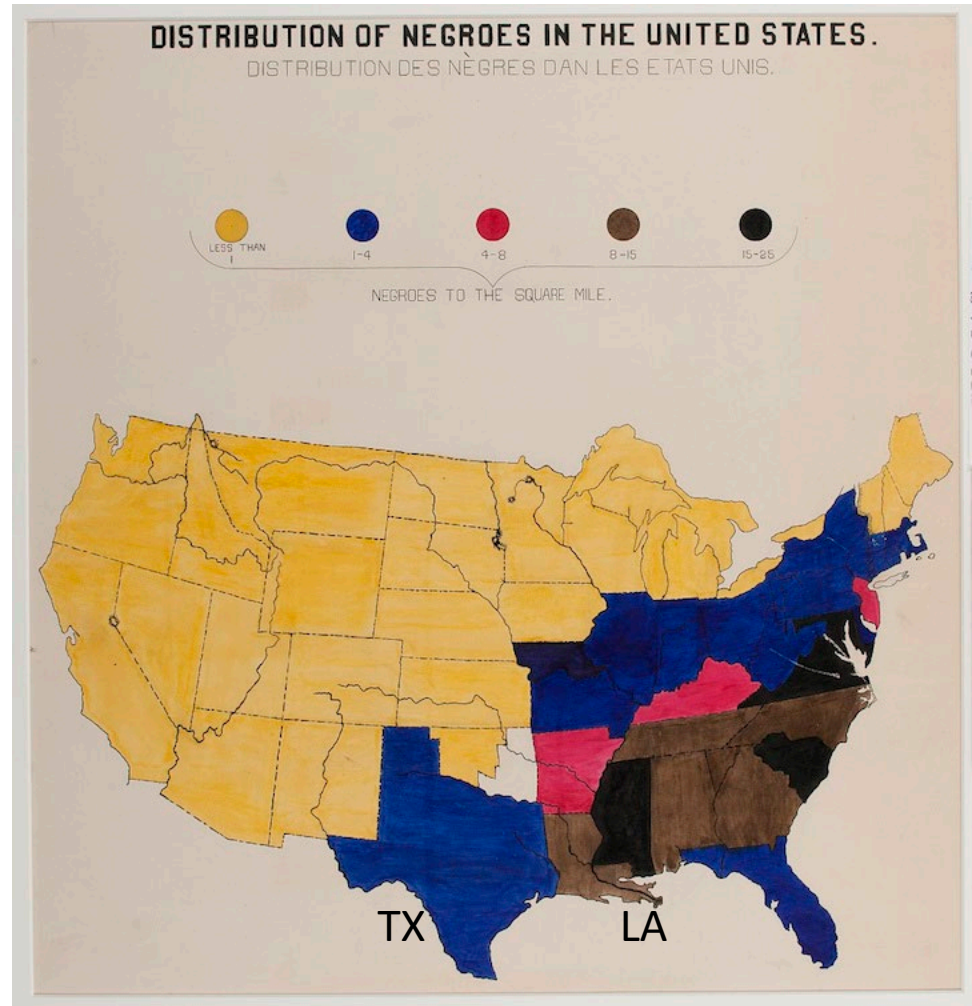


# Nice graphic, naïve about color

W.E.B. DuBois presented this as part of an exhibition on The American Negro at the 1900 Paris Exposition.

It is a landmark graphic, but shows no understanding of the use of color for a **quantitative** variable.

Q: Are there more Negroes per sq. mile in Texas (TX) or Louisiana (LA)?

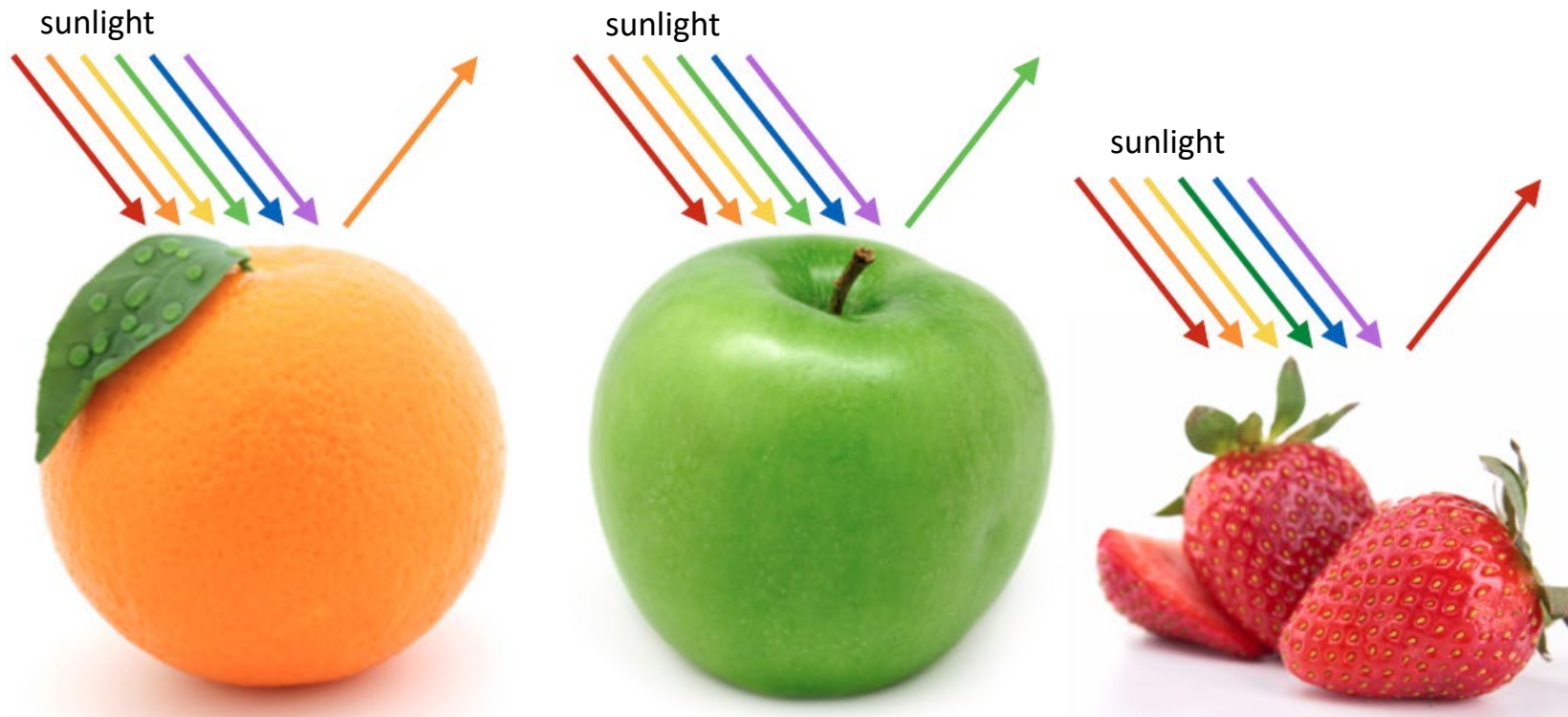




# Object color

What makes an orange look **orange**, a green apple look **green**, or a strawberry **red**?

Objects absorb colors from the rainbow, but **reflect** their own



# Is this a color photo?



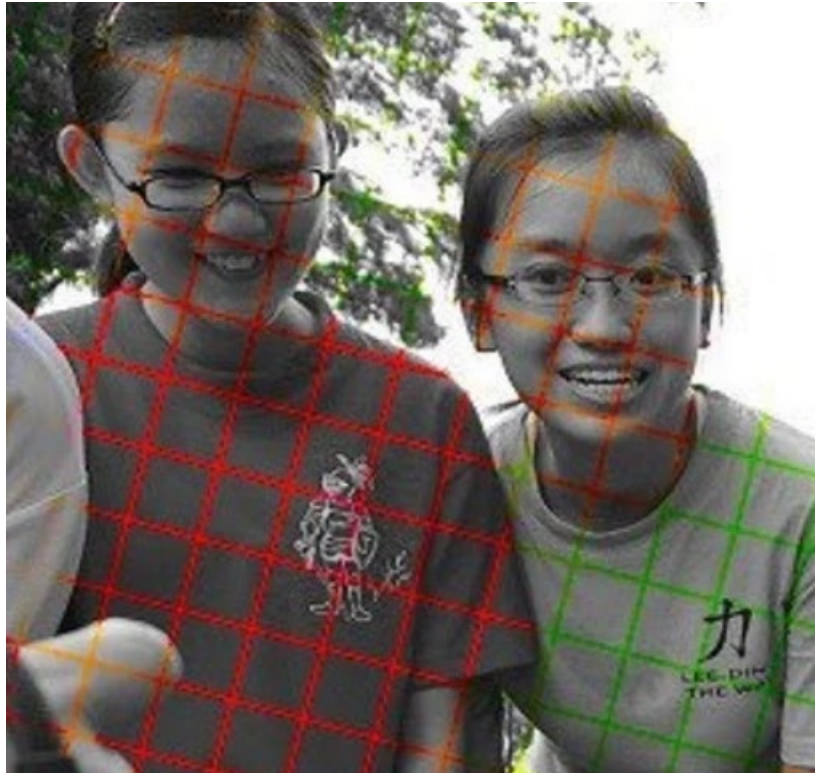
"2011-07-12-Railway People" by Chuwa (Francis) is licensed under CC BY-SA 2.0, color illusion remix by <http://pippin.gimp.org/>



# Is this a color photo?

Graphic artist Øyvind Kolås overlaid red, orange, yellow, blue, and green grid lines over a B/W photo.

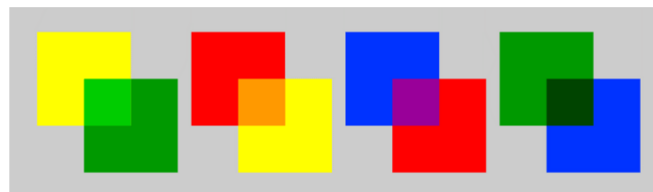
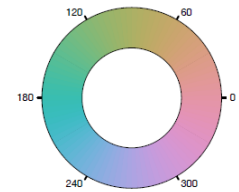
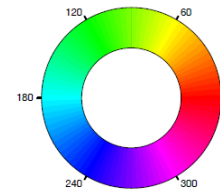
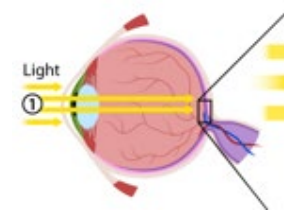
The colored grid causes us to perceive it as having smooth color



Color perception is influenced by top-down processes of **scene recognition**

# Color: Aspects in data graphics

- Perception: trichromatic theory
  - How the eye sees color
- Color spaces:
  - RGB (additive), CMYK (subtractive)
  - HSV, HCL: perceptually based
- Color palettes for computer graphics
  - ColorBrewer: sequential, diverging, qualitative
  - Color-blind safe ?
  - Photocopy safe ?
- Transparency

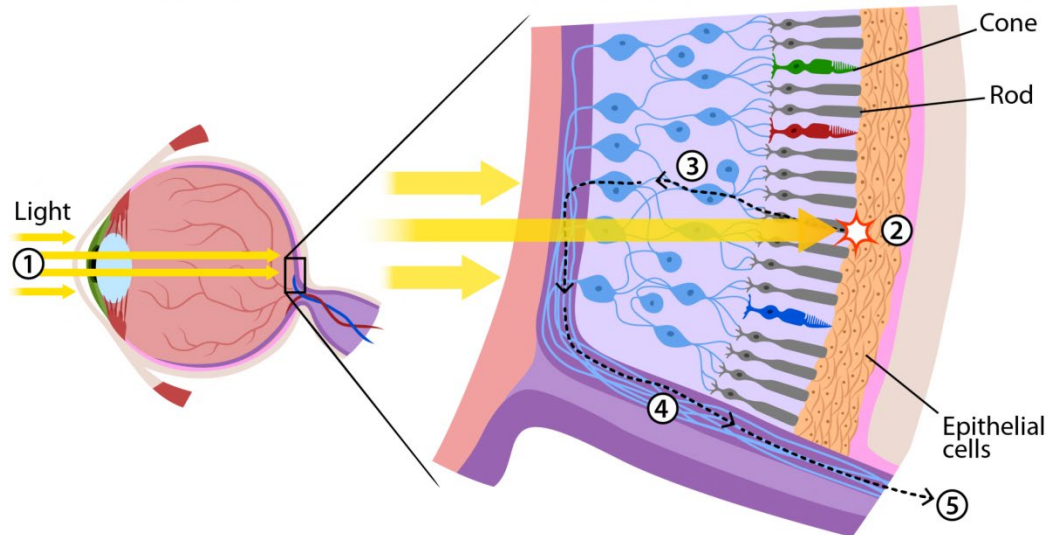
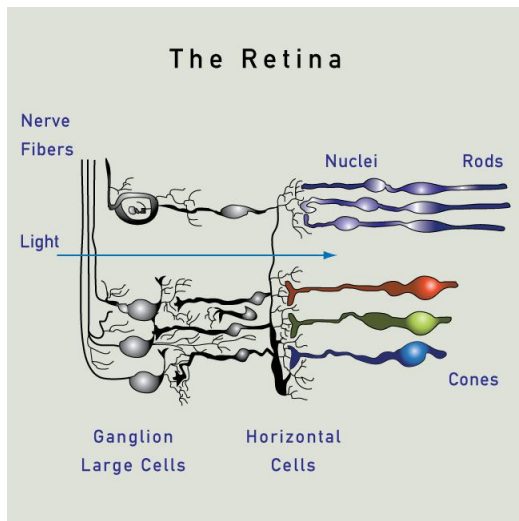


# Perception: The human eye

- Retina:
  - rods (monochromatic),
  - cones (R, G, B)

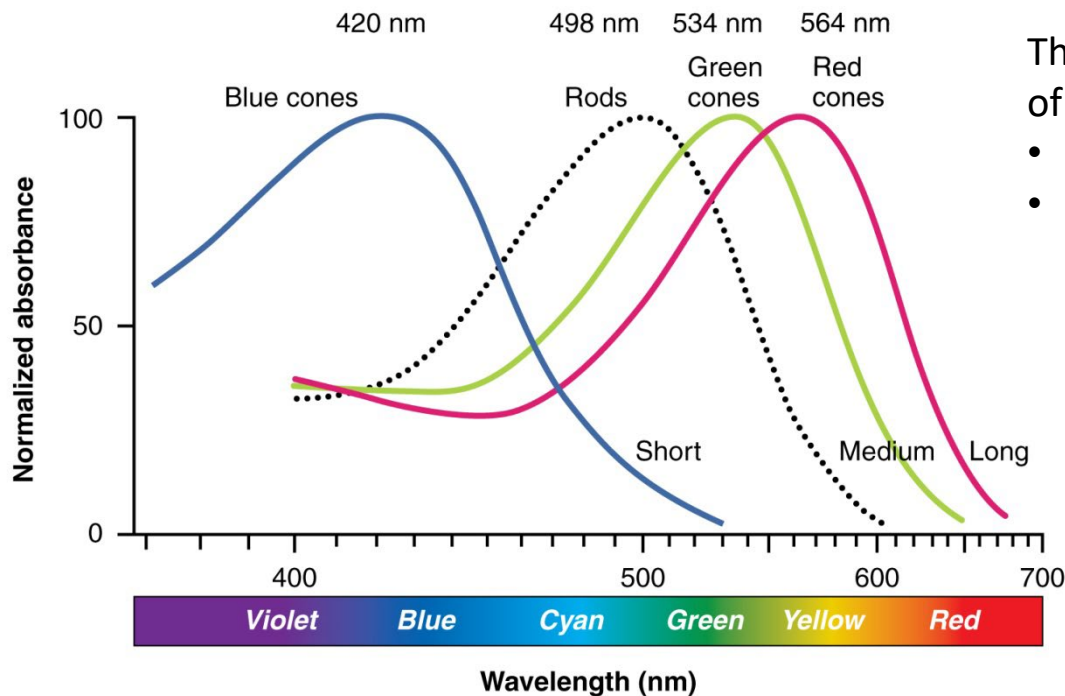


It is of interest to see the wide variety of ways this is conveyed in scientific diagrams:



# Perception: color sensitivity

- Cells in the retina are differentially sensitive to colors of different wavelength
  - Each have a **distribution** of sensitivity for **short**, **medium** & **long**
  - Their **peaks** are used to name them as **Blue**, **Green**, **Red** or Rods



This figure also stimulates questions of scientific visualization

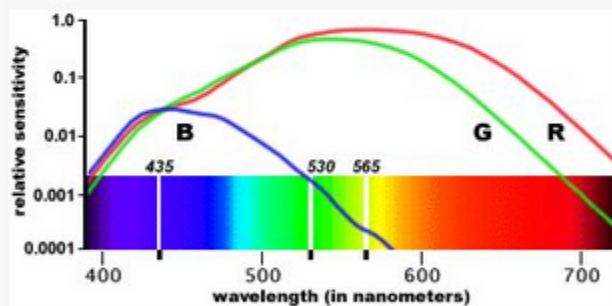
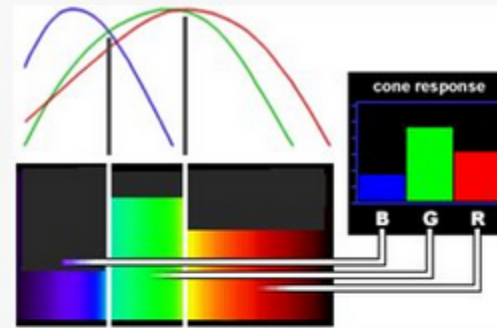
- Rods & cones are “normalized”
- Are they all equal in what we see?



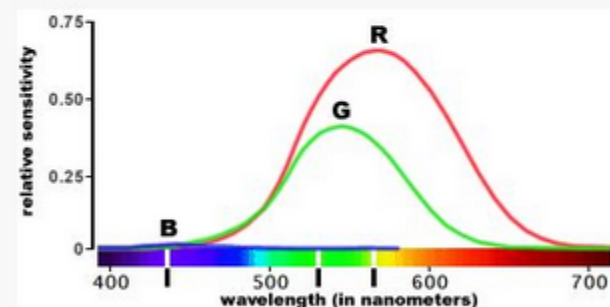
# Perception: color sensitivity

Color sensitivity, shown on three different scales

- Cone receptors least sensitive to (least output for) to blue
- most sensitive to red



Relative sensitivity curves for the three types of cones, log vertical scale, cone spectral curves from Vos & Walraven, 1974



Relative sensitivity curves for the three types of cones, the Vos & Walraven curves on a normal vertical scale

from: <http://slideplayer.com/slide/6329532/>

# Perception: Contrast

Color perception, even of gray, is influenced by **contrast** against a **background**

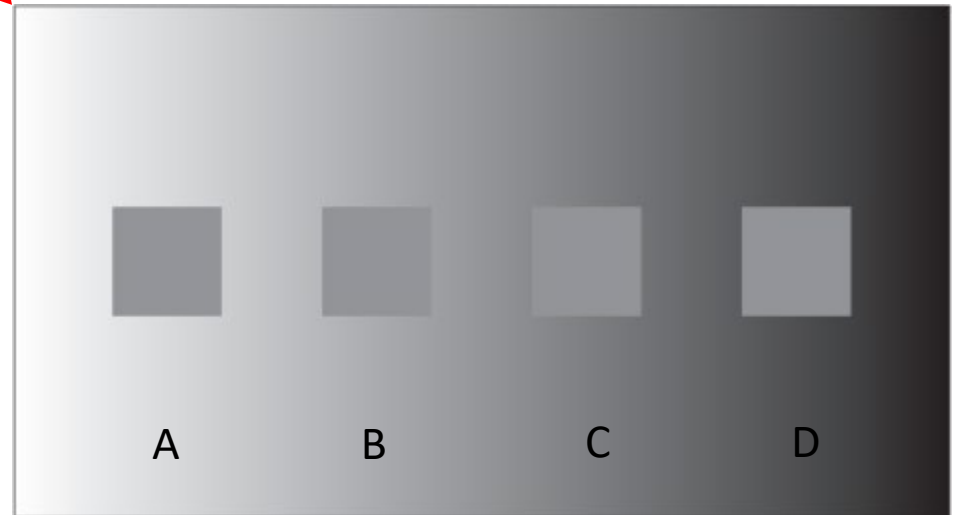
Q: Which gray square at right is most similar to that at the left?

gray  
square



Most people say **A**, because it is shown on a light background

A: it is the **same** gray square against a changing background





# Luminance contrast

Showing blue text on a black background doesn't work very well. There is insufficient luminance contrast.

Showing blue text on a white background works better. There is sufficient luminance contrast.

Showing yellow text on a white background doesn't work very well. There is insufficient luminance contrast.

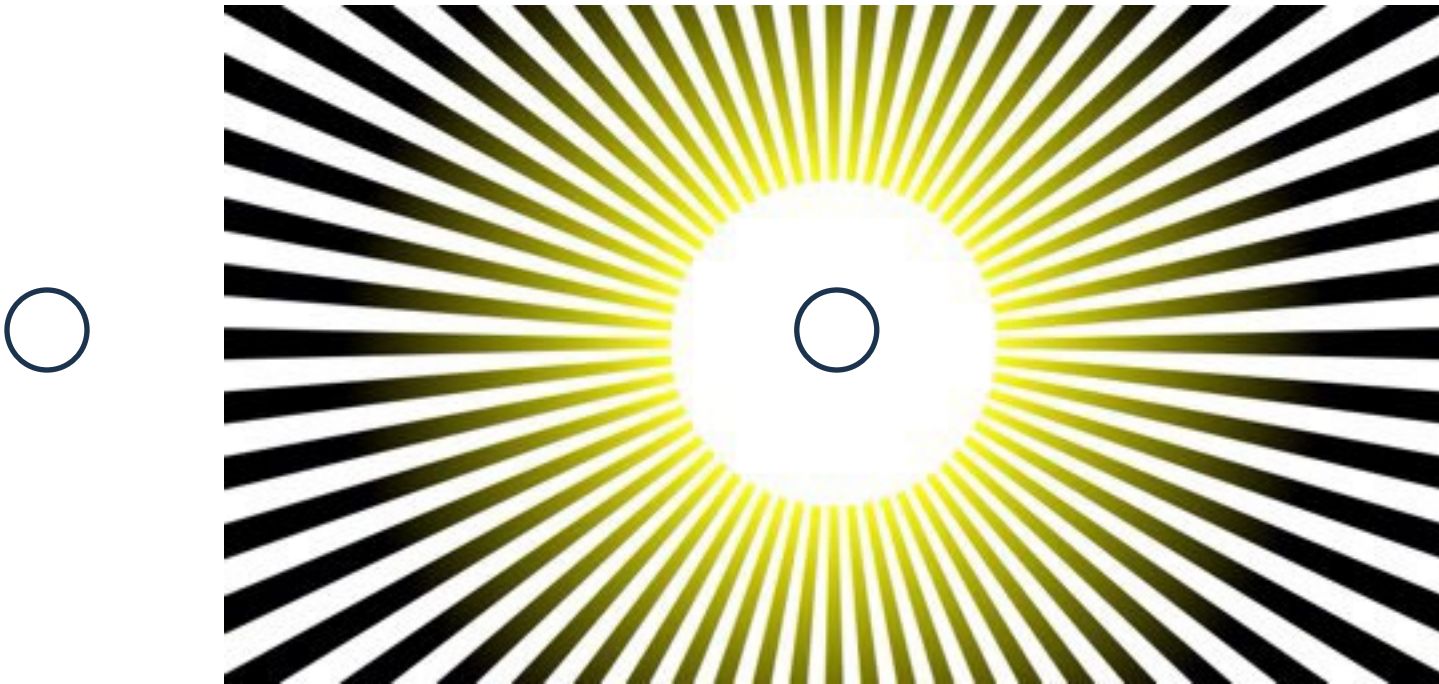
Showing yellow text on a black background works better. There is sufficient luminance contrast.

TIP: For presentations, light text on a **dark background** is often preferred. I don't do this, because I'm also concerned with printing slides. (With LaTeX Beamer, it is easy to have separate setups for presentation & print)

# Brightness illusion

Is the white at the center the **same** brightness as the white at the edges?  
Or, is it **brighter**?

They are the **same**. We interpret the center as a source of light relative to the surrounding bright yellow.



The illusion also affects pupil size! <https://www.pnas.org/content/109/6/2162>

# Discovery of color

Feb 8, 1672: Isaac Newton reads his Optics paper to the Royal Society of London  
“light is not homogeneous,... but rather consists of rays of different forms”





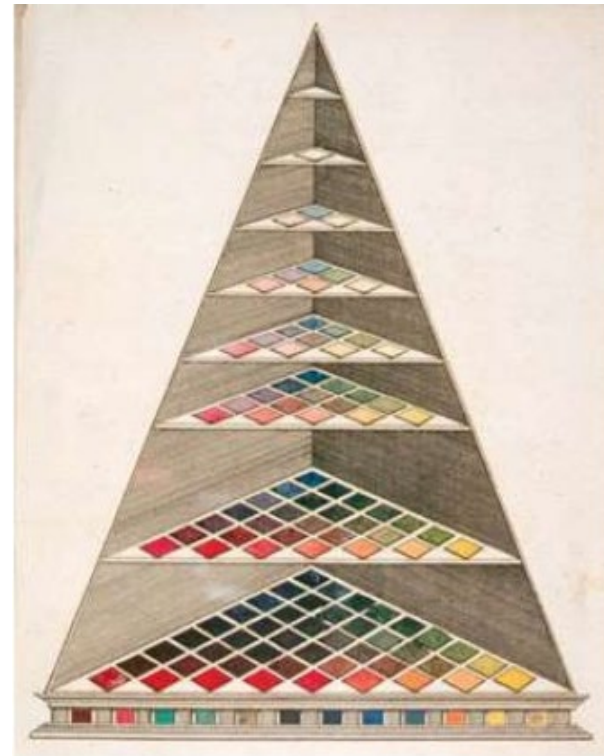
# Early color theory

Tobias Mayer (1755) – color theory composed of (blue, red, yellow) as basic colors



Introduces the idea of color “primaries”

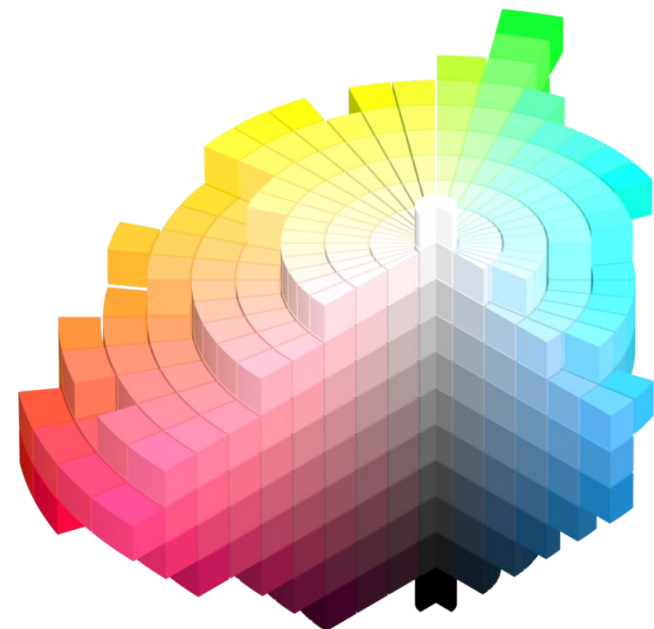
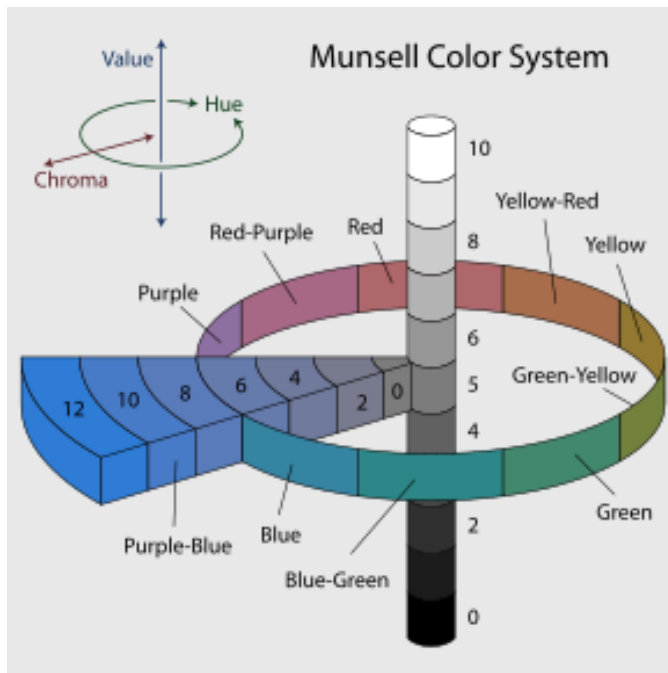
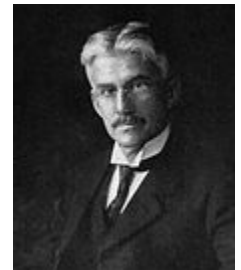
Johannes Lambert (1772) – A color pyramid, composed of 7 layers



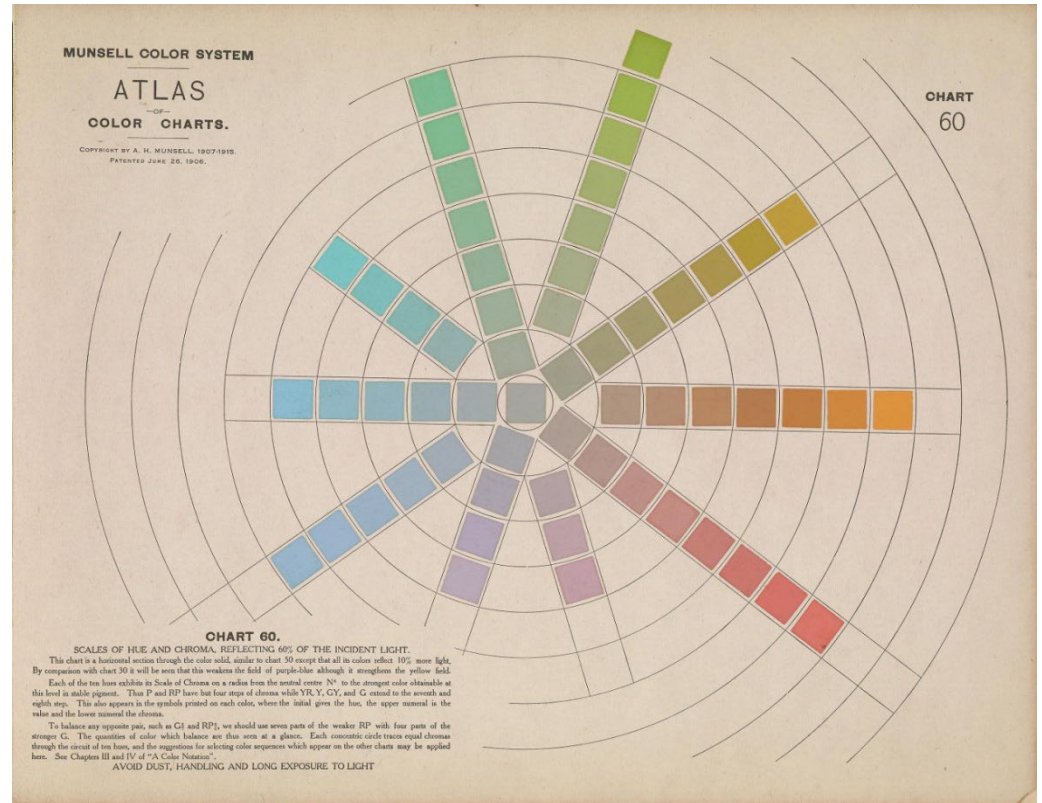
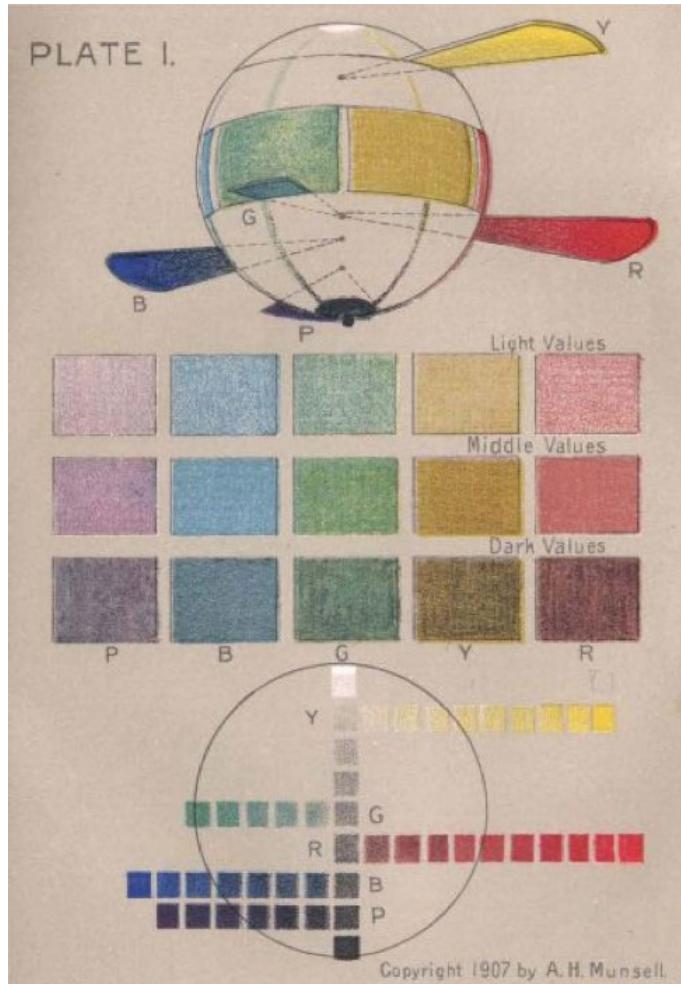
Introduces the idea of color saturation

# Color space: Munsell colors

- Color space is 3D
  - How to specify a given color in **perceptual** terms?
  - Albert Munsell (~1930): **hue, chroma, lightness** (HCL)
  - These form **perceptually uniform** & **independent** dimensions



Munsell's color scheme was highly influential in Psychology research  
Nearly every lab investigating color used standard sets of Munsell color chips

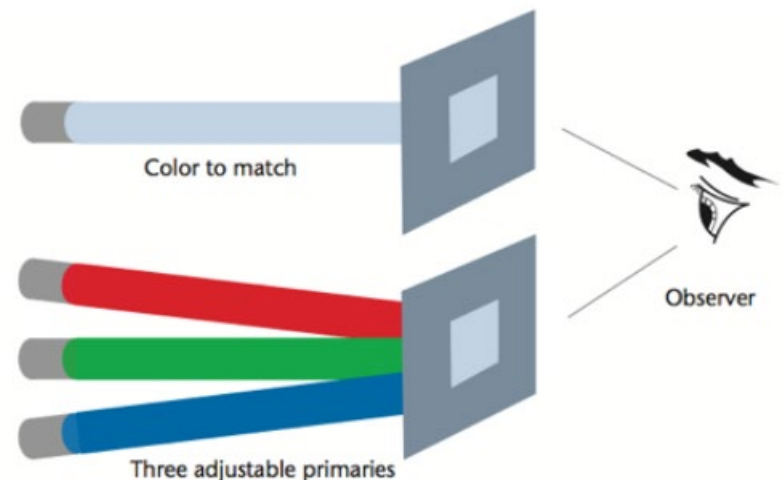


# CIE color space

- How do we know about the **perceptual** properties of colors, taking spectral sensitivity into account?

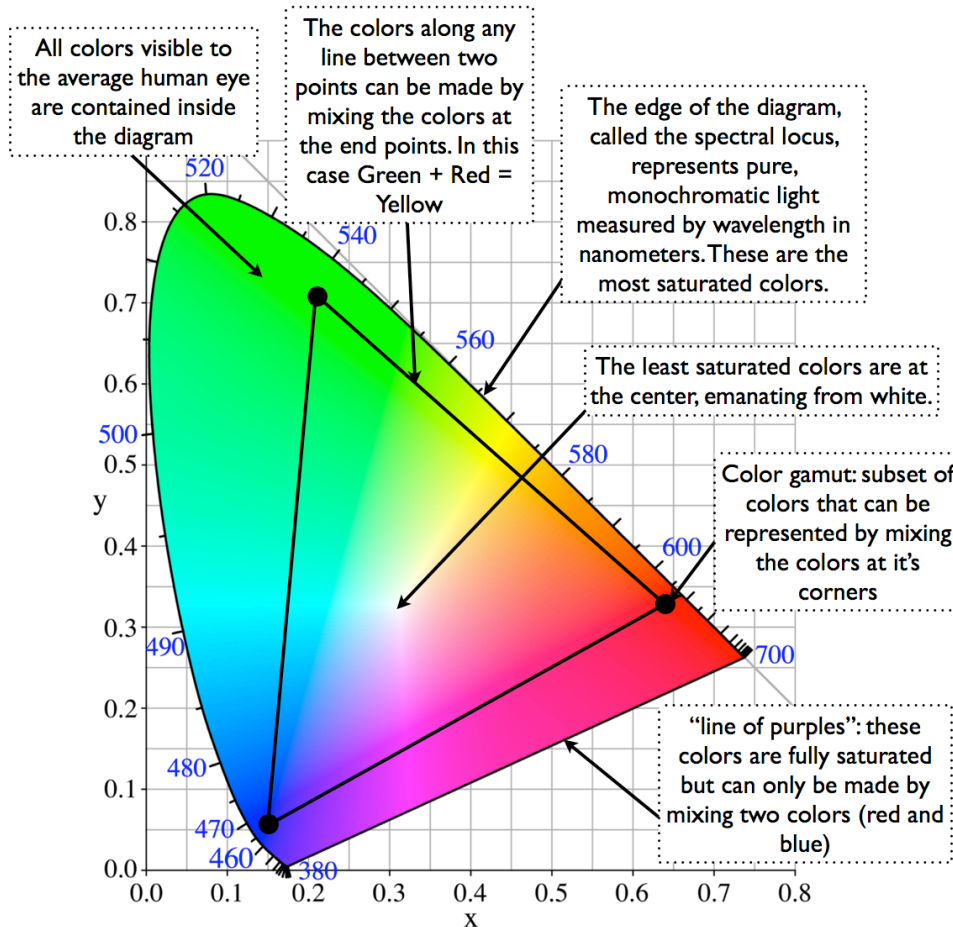
Experiments used a **color-matching** task:

- Adjust the intensity of pure R, G, B lights to match a given color
- This defines a new color theory connecting **physical** properties and human **perception** (spectral sensitivity).
- The CIE (International Commission on Illumination) becomes the standard to calibrate color in scientific instruments and human experiments





# CIE color space



Anatomy of a CIE Chromaticity Diagram

The International Commission on Illumination (CIE) in 1931 defines a color space of  $(x,y,z)$  coordinates based on color-matching experiments combining **R**, **G**, **B** light sources in additive mixtures, and a “standard colorimetric observer”

This defines a new color theory connecting **physical** properties and human **perception** (spectral sensitivity).

There are eventually a variety of CIE color spaces (CIElab, CIEluv, ...) and lots of formulas for converting among them.



# Color space: RGB & CMYK

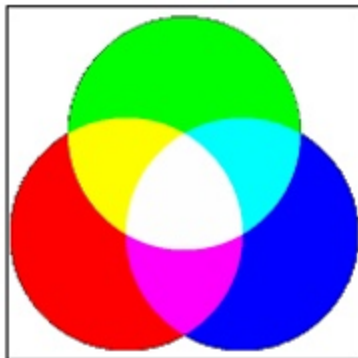
## Enter technology: how to produce color?

- RGB:
  - Combine **light**: R + G + B = white
  - Used in computer monitors, TV, film
- CMYK:
  - Combine **ink**: Cyan + Magenta + Yellow = Black
  - Used in color laser printers, the print industry



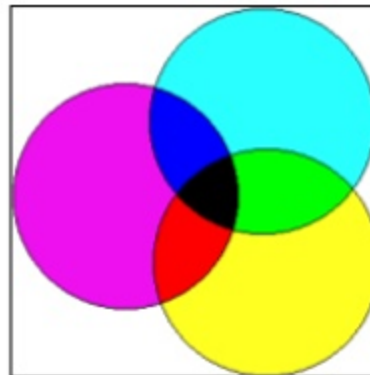
**Caution:** R, SAS, SPSS use RGB by default

### Additive colour system



Mixture of primary light colours-  
**White**

### Subtractive Colour System



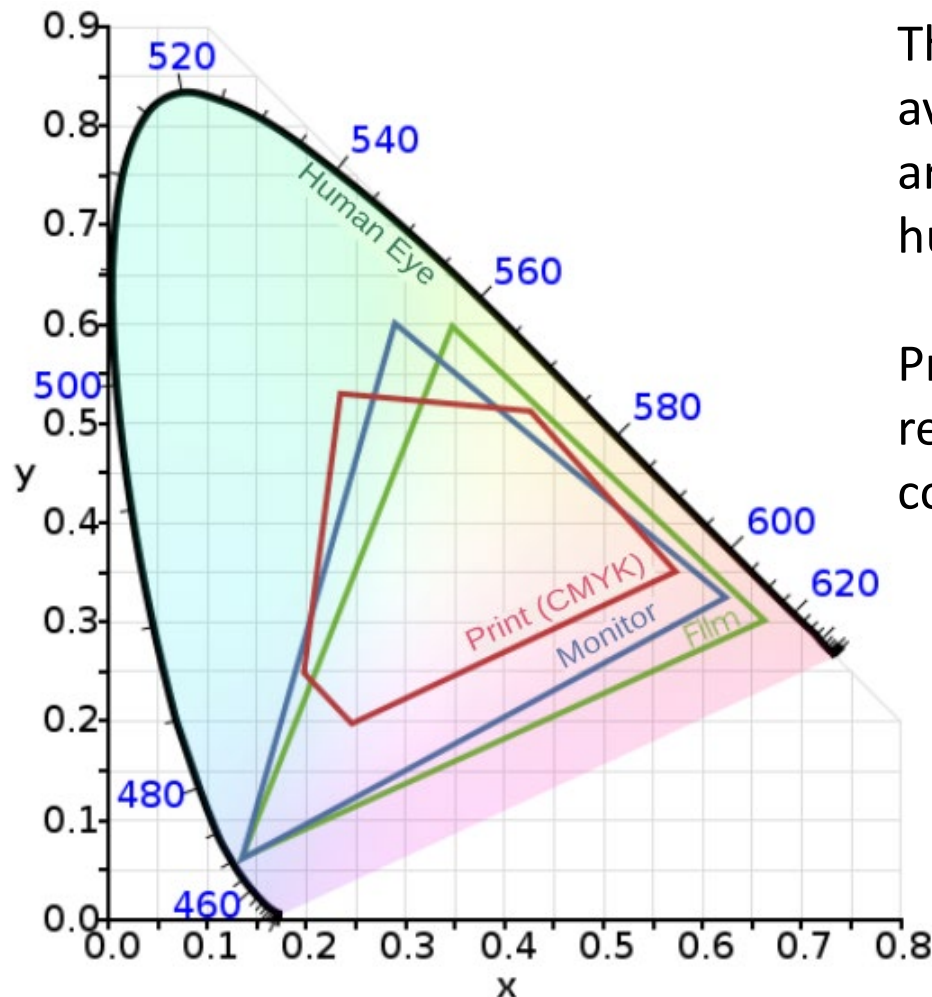
Mixture of primary pigment colours  
**Black**

**TIP:** for publishing, you may need to convert graphics from RGB to CMYK.

Some software offer useful tools for this:

- Adobe Acrobat Pro
- ImageMagik

# Color space: RGB & CMYK



The standard gamut of colors available for different **display media** are a restricted subset of what the human eye can see.

Print (CMYK) is most restricted, and requires a more careful choice of color in graphics



Always check the display on different media.

# Color in DataViz Design

There are several commonsense rules & guidelines for use of color in statistical graphics, maps & info vis

## RULES

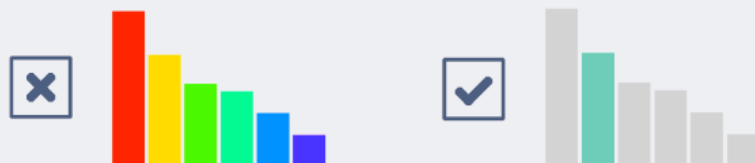
### INTUITIVENESS

Use intuitive colors. When choosing them, consider what associations do they evoke. If possible, use colors that audience will associate with your data anyway.



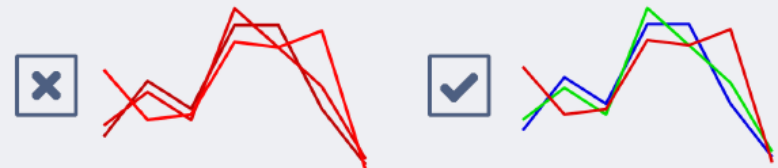
### MODERATION

Use colors in moderation. For a simple dataset, a single color is preferable. Use color as a strategic tool to highlight the important parts of your visual.



### CLARITY

Use colors to make the data easier to read. Make sure your audience will be able to distinguish between the items shown in the visualization.



### CLASSIFICATION

Don't use a gradient color palette for categories. And the other way round - different colors for same measurement.



# Software: Color specification

- Color is often hard to use effectively in software, because the ways to specify it are so varied:
  - **Color names:** “black”, “red”, “green3”, “skyblue“, “cyan”
  - **RGB:** black=(0,0,0); green3=(0, 205, 0), cyan=(0, 255, 255)
  - **Hex:** black=“#000000”; cyan=“#00FFFF”

18	beige	#F5F5DC	245	245	220
19	bisque	#FFE4C4	255	228	196
20	bisque1	#FFE4C4	255	228	196
21	bisque2	#EED5B7	238	213	183
22	bisque3	#CDB79E	205	183	159
23	bisque4	#8B7D6B	139	125	107
24	black	#000000	0	0	0
25	blanchedalmond	#FFEBCD	255	235	205

68	cyan	#00FFFF	0	255	255
69	cyan1	#00FFFF	0	255	255
70	cyan2	#00EEEE	0	238	238
71	cyan3	#00CDCD	0	205	205
72	cyan4	#008B8B	0	139	139
73	darkblue	#00008B	0	0	139
74	darkcyan	#008B8B	0	139	139
75	darkgoldenrod	#8B860B	184	134	11

26	blue	#0000FF	0	0	255
27	blue1	#0000FF	0	0	255
28	blue2	#0000EE	0	0	238
29	blue3	#0000CD	0	0	205
30	blue4	#00008B	0	0	139
31	blueviolet	#8A2BE2	138	43	226
32	brown	#A52A2A	165	42	42

76	darkgoldenrod1	#FFB90F	255	185	15
77	darkgoldenrod2	#EAD0E	238	173	14
78	darkgoldenrod3	#CD950C	205	149	12
79	darkgoldenrod4	#8B650B	139	101	8
80	darkgray	#A9A9A9	169	169	169
81	darkgreen	#006400	0	100	0
82	darkgrey	#A9A9A9	169	169	169

See: <http://research.stowers.org/mcm/efg/R/Color/Chart/> for R color charts

# Software: Color specification

WTF! Give me a break, please:

- Make it easier to **compute** with colors: define blends of colors or a color ramp
- Make it easier to specify color **schemes** with decent **perceptual** properties
- Make it easier to map colors to **data features** I want to show



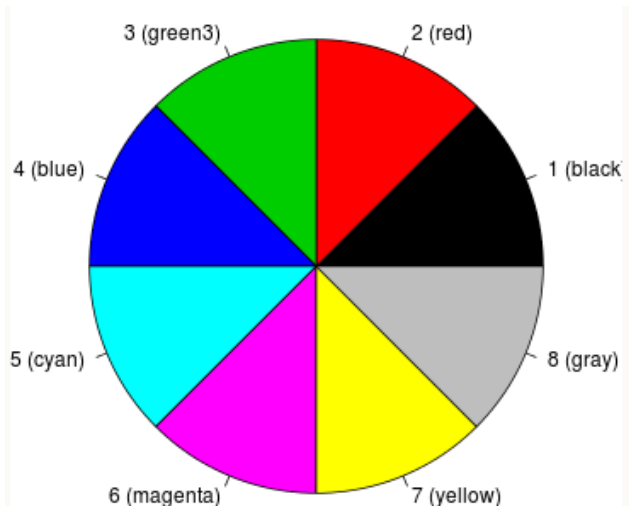
*Every time you are  
forced to say  
“#008B8B” or “cyan4”  
a puppy dies  
somewhere*

*-- MF, 2018*

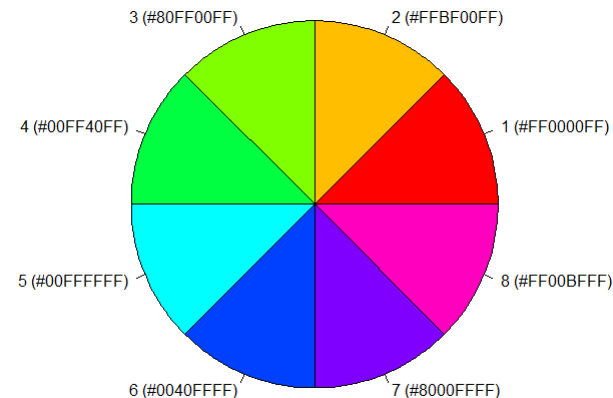
# Software: palettes

- R (and other software) provide palettes of colors used for defaults in graphs
  - Not all are nice— depends on your purpose
  - There are lot of choices – maybe too many!
  - You can change them **once** for all graphs in a session or paper

```
> (pal <-palette())  
[1] "black" "red" "green3" "blue" "cyan" "magenta" "yellow" "gray"  
> pie(rep(1, length(pal)), labels = sprintf("%d (%s)", seq_along(pal), pal), col = pal)
```

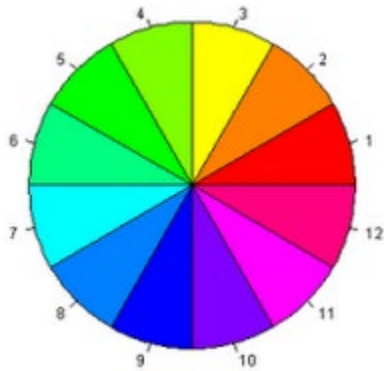


```
palette(rainbow(8)); pie(...)
```

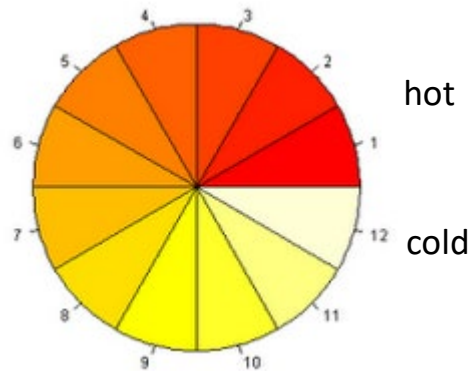


# R: basic palettes

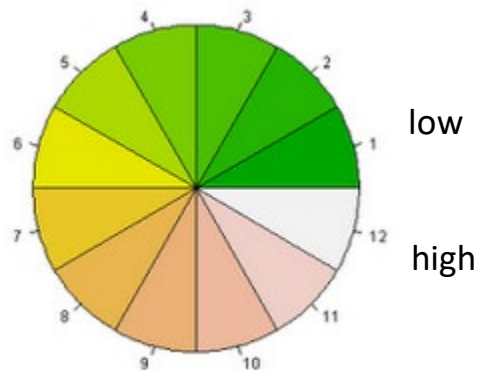
```
n <- 12  
pie(rep(1, n), col=rainbow(n))
```



```
pie(rep(1, n), col=heat.colors(n))
```



```
pie(rep(1, n), col=terrain.colors(n))
```



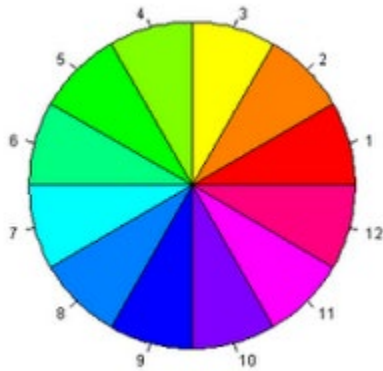
```
pie(rep(1, n), col=topo.colors(n))
```





# R: basic palettes

rainbow



heat



terrain



topo



## Discussion Q:

- Which of these seem better for **quantitative** variables?
- Which for **categorical**?

These are shown for **area** fill. How effective would they be for:

- **point** colors
- **line** colors

E.g., **yellow** is bright as an area, but nearly invisible as points (●) or lines (→) or **text on a white background**



# palettes: ColorBrewer

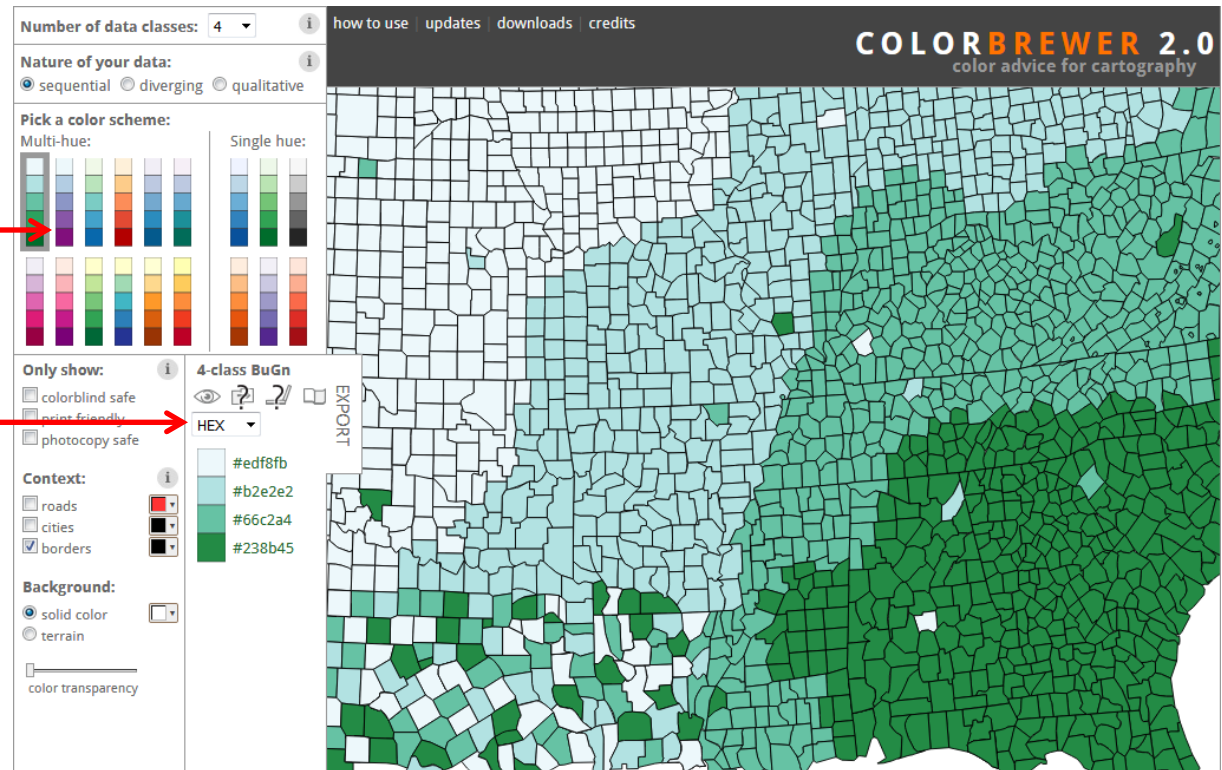
ColorBrewer, by Cynthia Brewer provides an interactive application for choosing color palettes, <http://colorbrewer2.org>

This is one example of a **multi-hue** scheme for a **quantitative, sequential** variable, shown from low to high with 4 color classes

variable type

choose different versions of the scheme

export color specs to HEX, RGB, CMYK



This example: <http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=4>

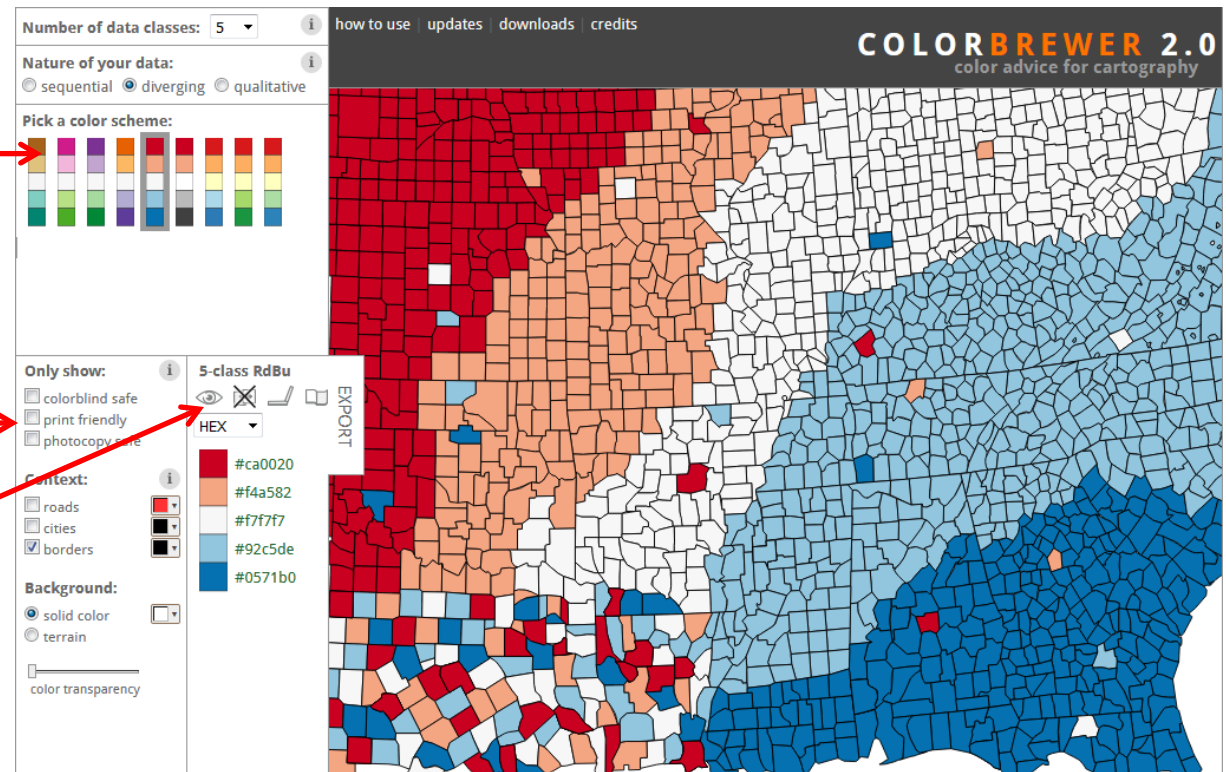
# palettes: ColorBrewer

**Diverging schemes** are designed to show a **quantitative** variable, where we want to see what is low vs. what is high, leaving the middle of less visual impact – difference from average, residuals, ...

there are different schemes within this rubric

there are tools to filter for colorblind, print & B/W

Warnings when not friendly



This example: <http://colorbrewer2.org/#type=diverging&scheme=RdBu&n=5>

# palettes: ColorBrewer

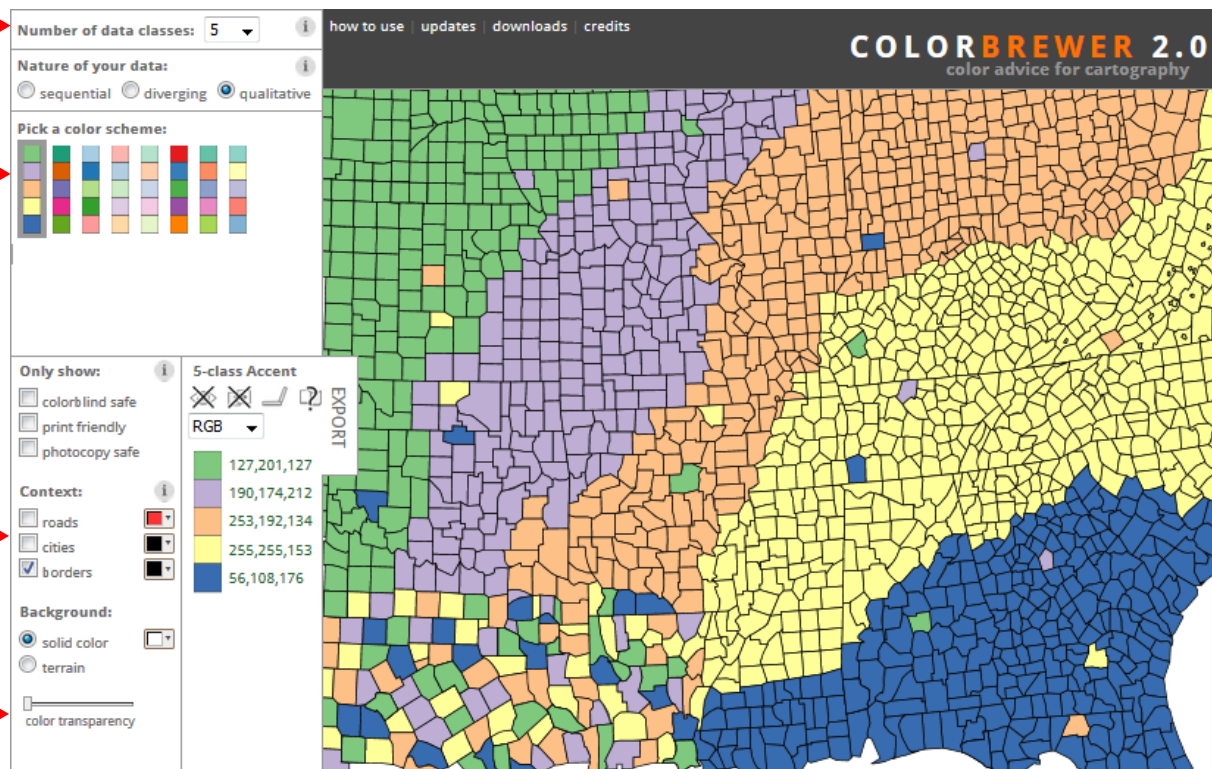
Qualitative schemes are designed to show a **categorical** variable, where we want to see differences among **unordered** categories

choose # classes

various schemes

see other context

add transparency



These are all available in the RColorBrewer package

This example: <http://colorbrewer2.org/#type=qualitative&scheme=Accent&n=5>

# palettes: RColorBrewer

```
RColorBrewer::display.brewer.all()
```

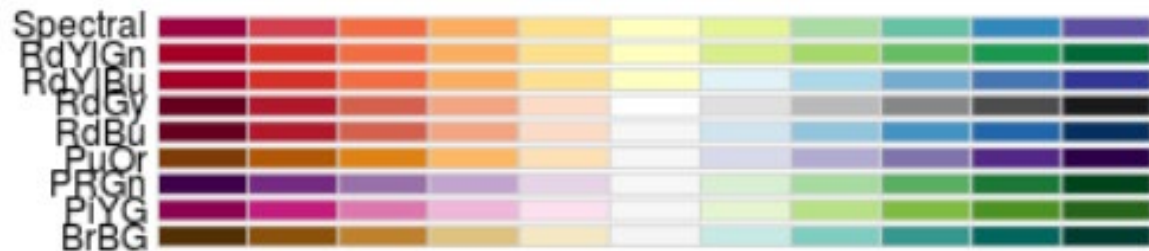
sequential



qualitative



diverging



# R: choose\_palette()

The colorspace package in R has an interactive palette widget.

It also provides functions for many kinds of color manipulations.

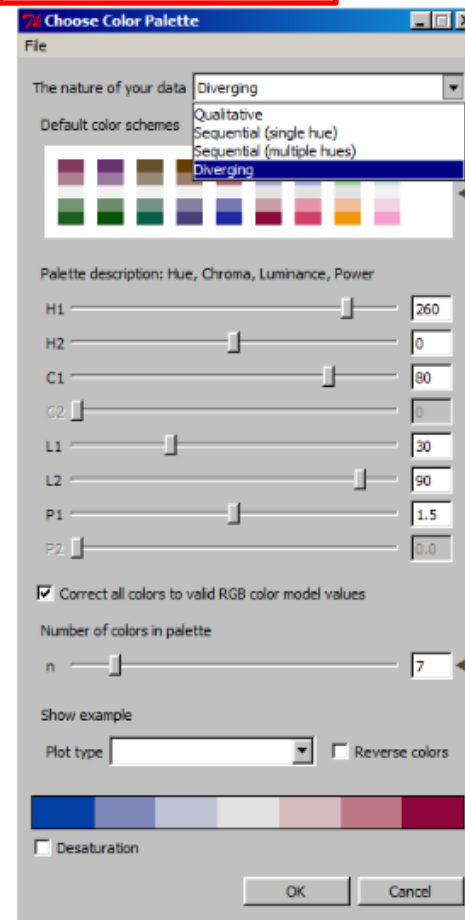
The R Color cheatsheet, by Malcolm Fraser is a goto source for all aspects of color in R:

<https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf>

## R color cheatsheet

Overview of colorspace palette selector

```
library("colorspace")  
pal <- choose_palette()
```

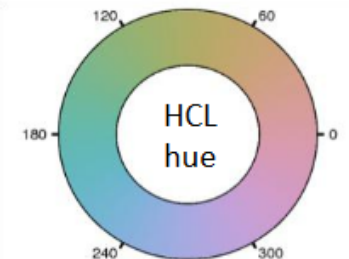


Select the type of color scheme based on the type of data

Default color schemes – can be used “as is” or as a starting point for modification

Interactively select:

- hue: color
- chroma: low chroma = gray
- luminance: high luminance = pastel
- power: how the color changes along a gradient



Select # of colors in palette

Save palette for future R sessions:

- txt file with hex codes
- .R file with a function describing how to generate the palette.

[source](#) can be used to import the

# Viridis palettes

Designed by Stéfan van der Walt and Nathaniel Smith for Python; ported to R in the [viridis](#) package.

Goals:

- **Colorful**, spanning as wide a palette as possible so as to make differences easy to see
- **Perceptually uniform**: values close to each other have similar-appearing colors and values far away from each other have more different-appearing colors
- **Robust to colorblindness**: these properties hold true for people with common forms of colorblindness, as well as in grey scale printing
- **Pretty**: much nicer as a defaults in software

These assertions are largely **untested**. Perhaps a good research topic!



# Comparing quantitative palettes

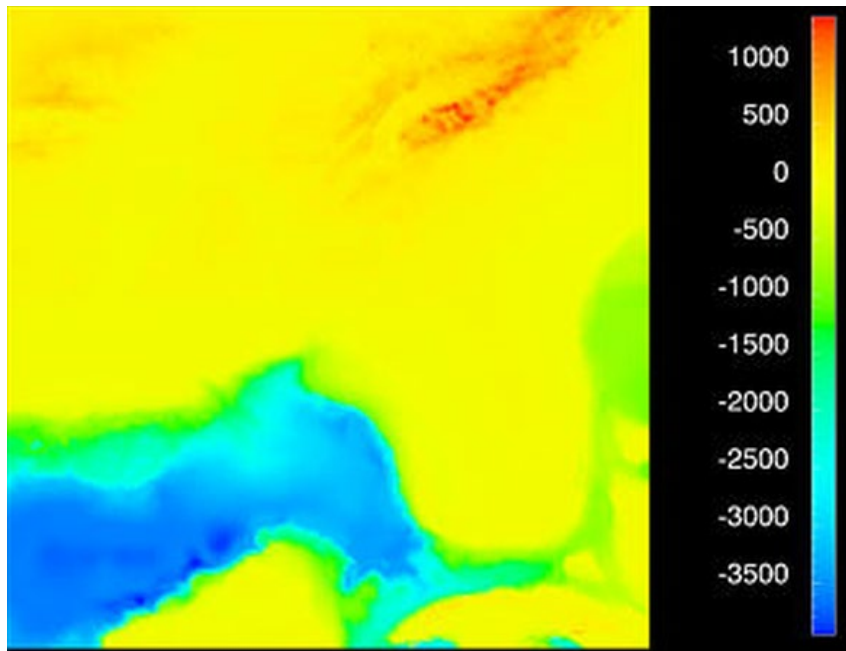
For a **quantitative** variable and a **continuous** color scale, there are many choices. How well do they work?



This is a bit tricky: ideally, we want a **wide range** of color

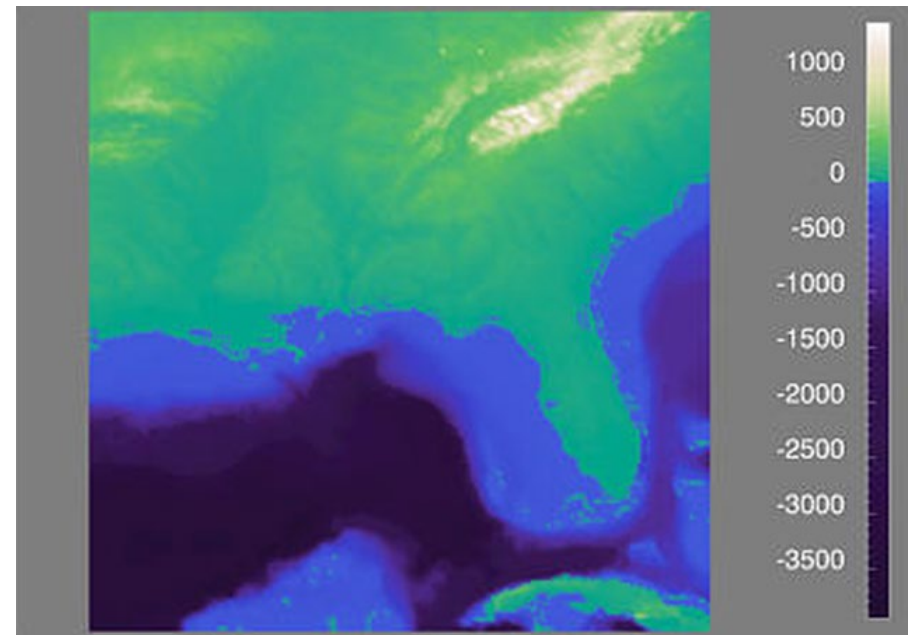
# Comparing palettes

What is shown in this map?



The rainbow color scale obscures the main features

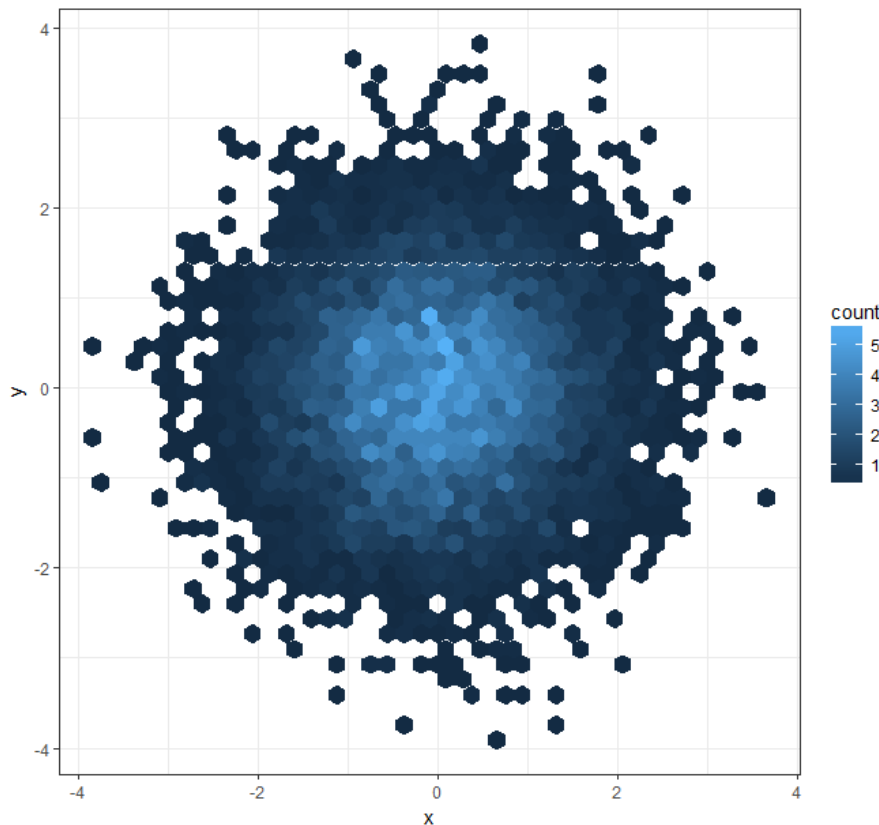
Now we can see it— elevation in the Florida coast: above or below 0



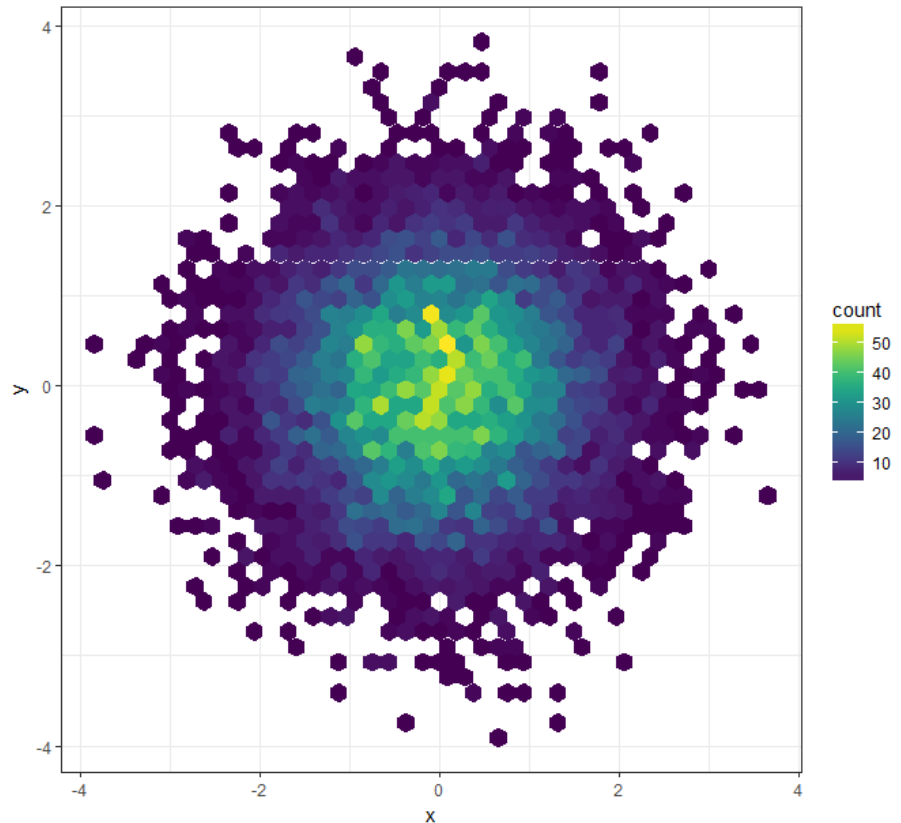
This color scheme was designed to reveal the essential topography of the map & to have perceptually equal elevation steps

# Comparing palettes

ggplot default palette



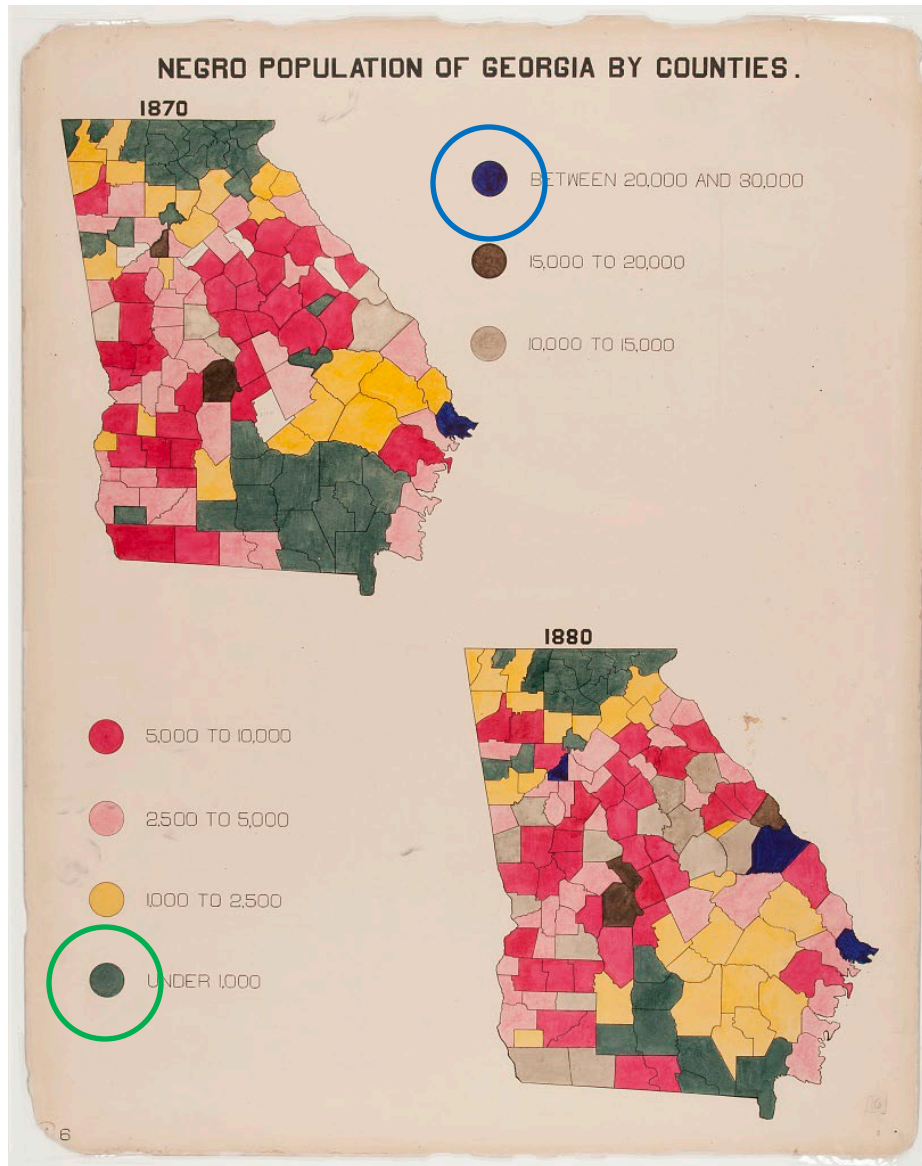
viridis default palette



```
df <- data.frame(x = rnorm(10000), y = rnorm(10000))  
g <- ggplot(df, aes(x = x, y = y)) +  
  geom_hex(bins=40) + coord_fixed() + theme_bw()  
g
```

```
library(viridis)  
g + scale_fill_viridis()
```

# WEB Du Bois' sense of color



Goal: Show Negro population of counties in Georgia from 1870 to 1880 & change

Which counties had largest Negro pop?

Which had lowest?

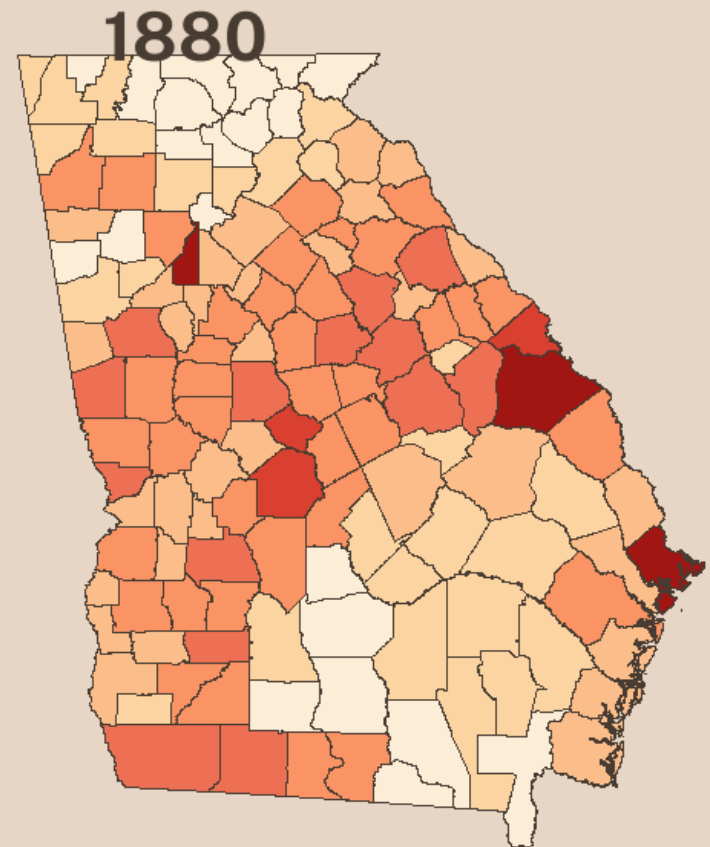
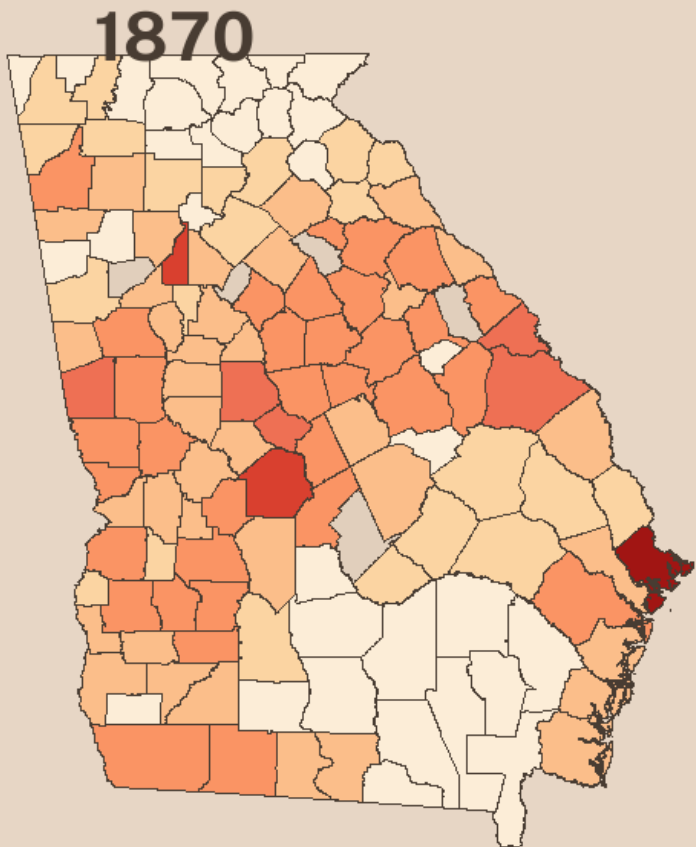
Where was the change greatest?  
(HARD: need to plot the diff<sup>ce</sup>)

# RColorBrewer to the rescue

```
library(RColorBrewer)
colors <- brewer.pal(n = 7, name = "OrRd")
colors <- c('#fef0d9', '#fdd49e', '#fdbb84', '#fc8d59',
            '#ef6548', '#d7301f', '#990000')
```

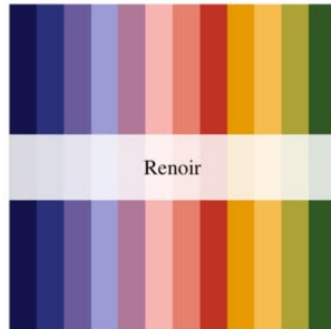
OrRd (sequential)

<1k  
1-2.5k  
2.5-5k  
5-10k  
10-15k  
15-20k  
20-30k



# Metbrewer palettes

Artistic palettes inspired by works at the Metropolitan Museum of Art in New York

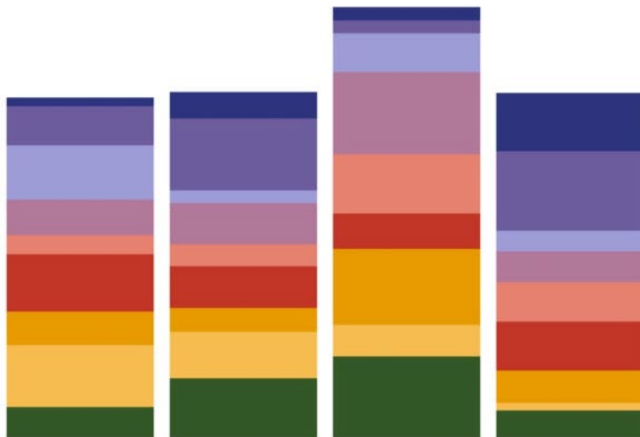


**Reinoir**

Nini in the Garden



Cypresses, 1889, Vincent van Gogh



Cap Nègre, 1909, Henri-Edmond Cross



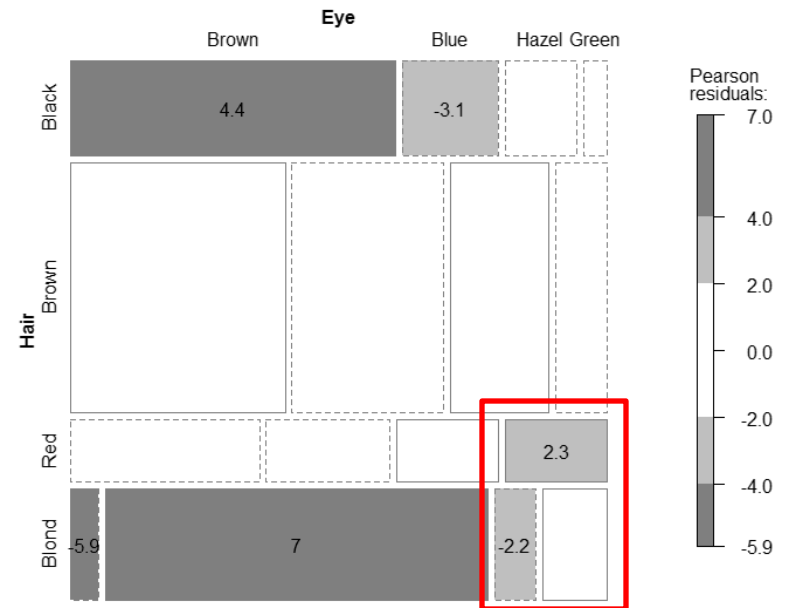
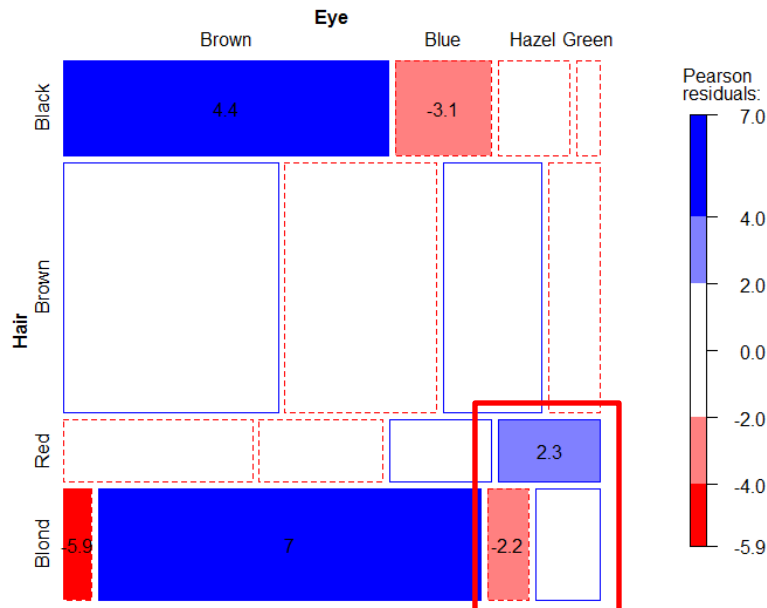


# Color → B/W ?

Graphics designed in color often must consider what happens when graphs are reproduced in B/W: **grayscale**

- This is particularly hard for a **diverging** color scale
- My original design for mosaic plots used solid vs. dashed lines to distinguish + vs. -

```
mosaic(haireye, labeling=labeling_residuals, gp=shading_Friendly)
```



# Color → B/W ?

The design of this graphic table was crafted to preserve readability if printed in B/W.  
NB: text for numbers changes from black to white depending on background color.

Figure 9: Section 37 benefits by type (1998–2015)

	1998– 2002	2003– 2005	2006– 2009	2010– 2013	2014– 2016	Scale	
Roads, streetscapes	30	35	54	83	15	0 - 10	
Culture, community, recreation	26	50	59	47	16	11 - 20	
Parks	27	41	41	52	20	21 - 30	
Affordable housing	17	26	38	56	11	31 - 40	
Public art	26	25	41	32	4	41 - 50	
Heritage	16	13	26	18	3	51 - 60	
Transit	11	7	10	20	3	61 - 70	
Libraries	6	2	5	11	1	71 - 80	
Other	3	6	7	8	3	81 - 90	

Figure 9: Section 37 benefits by type (1998–2015)

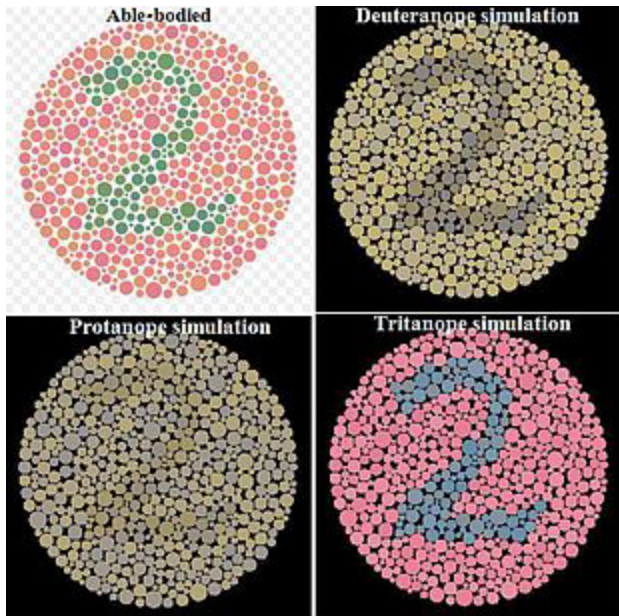
	1998– 2002	2003– 2005	2006– 2009	2010– 2013	2014– 2016	Scale	
Roads, streetscapes	30	35	54	83	15	0 - 10	
Culture, community, recreation	26	50	59	47	16	11 - 20	
Parks	27	41	41	52	20	21 - 30	
Affordable housing	17	26	38	56	11	31 - 40	
Public art	26	25	41	32	4	41 - 50	
Heritage	16	13	26	18	3	51 - 60	
Transit	11	7	10	20	3	61 - 70	
Libraries	6	2	5	11	1	71 - 80	
Other	3	6	7	8	3	81 - 90	

Background shading works equally well in color or B/W  
A+ for visual design!

Source: Friendly, A. R. (2017). *Land Value Capture and Social Benefits: Toronto and São Paulo Compared*. IMFG Papers on Municipal Finance and Governance, No 33, University of Toronto, <https://munkschool.utoronto.ca/imfg/>

# Colorblindness

Most common forms are genetic, and involve a deficiency in one of the cone type sensitivities



- Protanopia (**red** deficient)
- Deuteranopia (**green** deficient)
- Tritanopia (**blue** deficient)

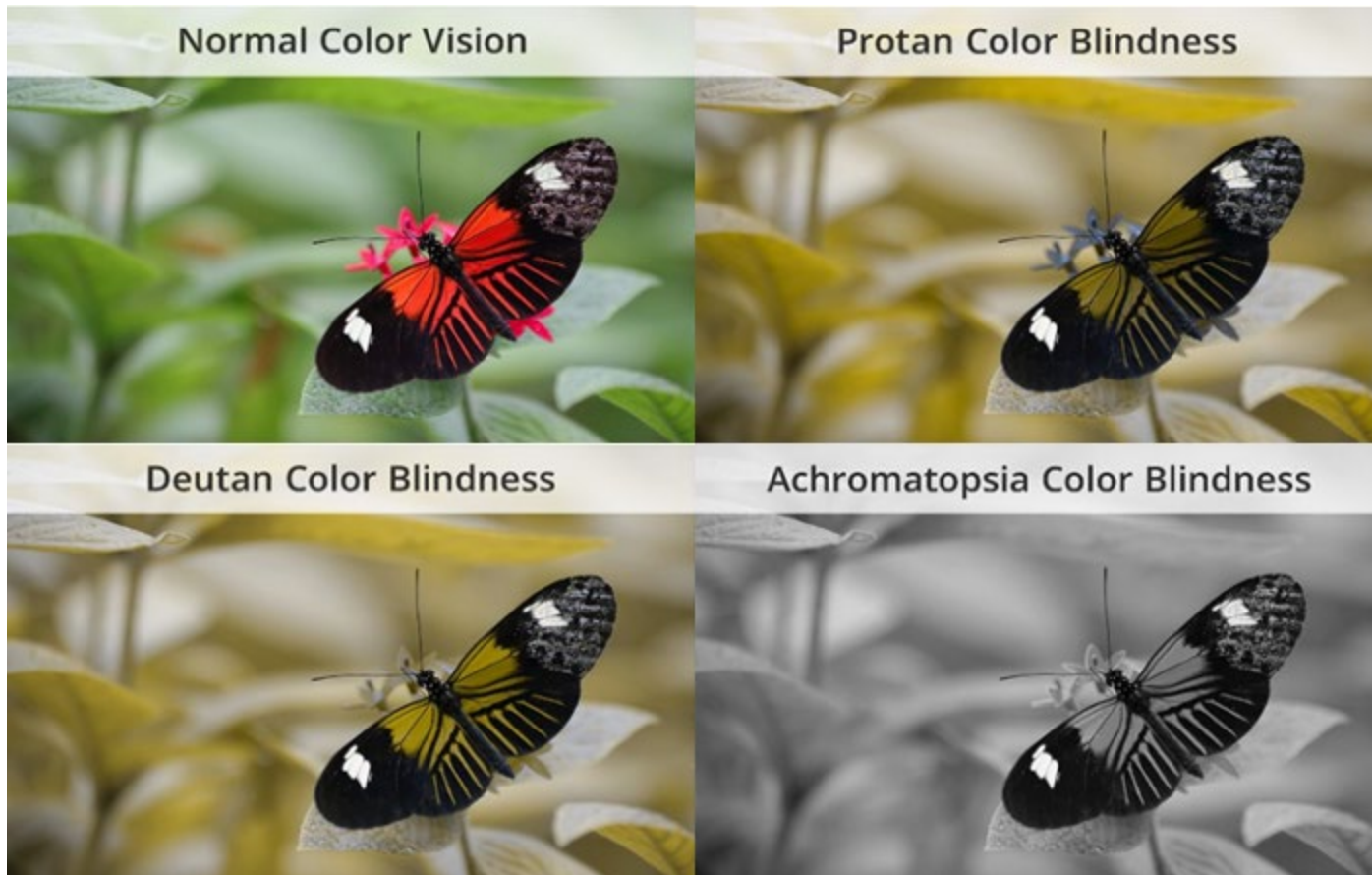
Some form of **red-green** insensitivity is most common

- about 6-8% of population
- more common in males

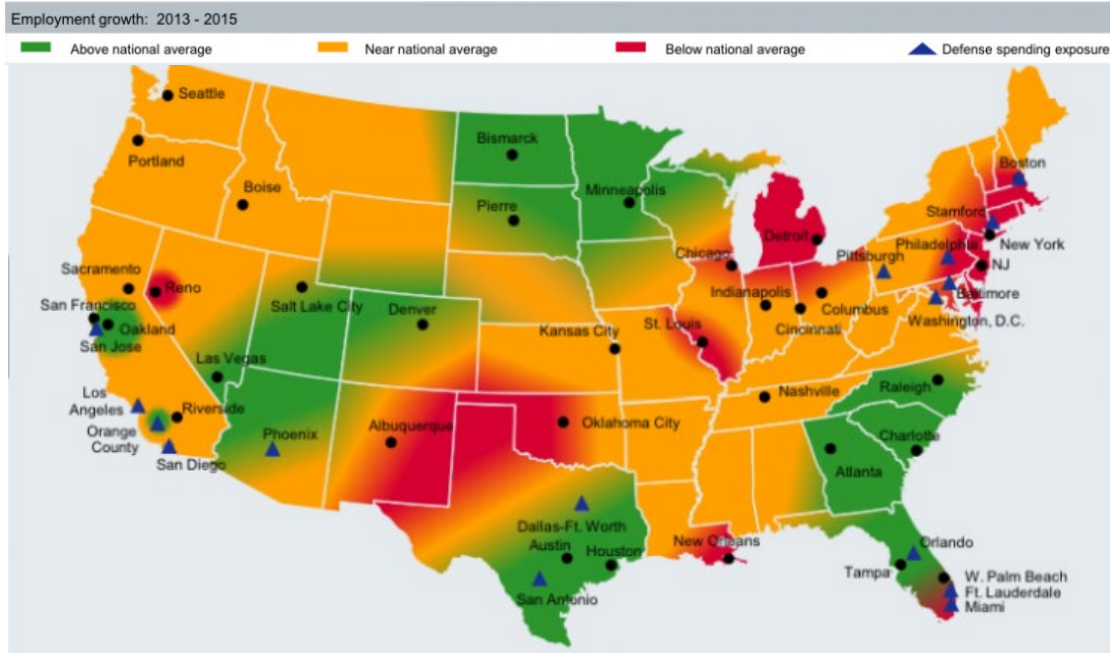
TIP: Avoid color scales with main variation between **red** & **green**

# Colorblindness

What an image looks like with various forms of color deficiency



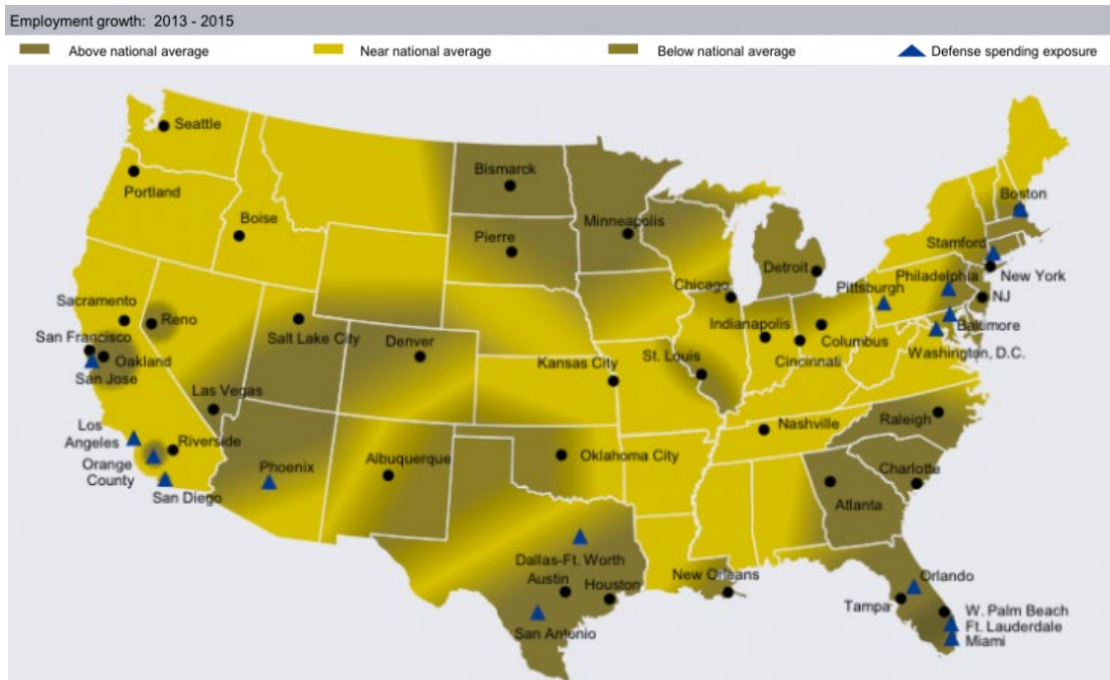
Red-green colorblindness:  $-R$  (protan)  $\approx$   $-G$  (deutan)



Goal: Show employment growth, 2013—2015

Original design, using

- green: above average
- red: below average



How this looks to someone with red-green colorblindness

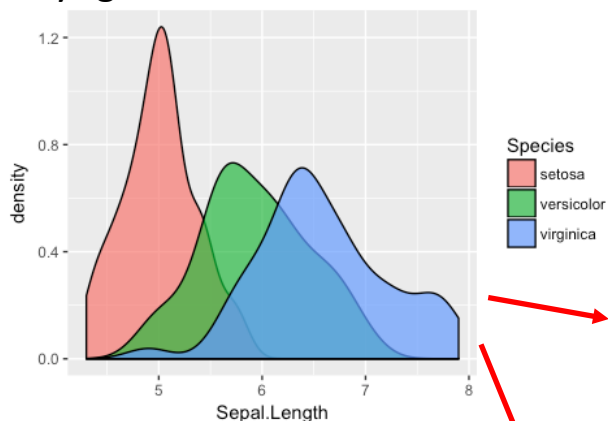
red & green become indistinguishable

From: <http://www.mena-forum.com/category/u-s-a/>



# colorblindr package

myfig



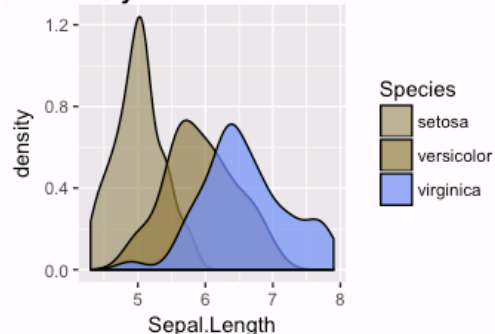
What would my graph look like to someone with color deficiency?

colorblindr simulates a graph under various conditions

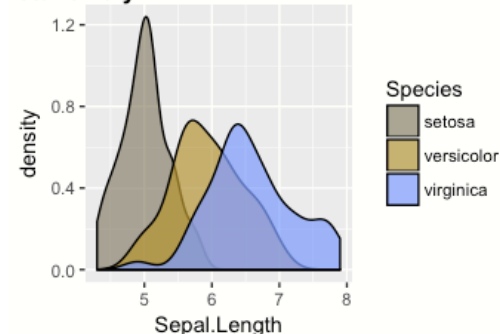
```
library(colorblindr)
cvd_grid(myfig)
```

NB: The default ggplot colors aren't colorblind or B/W friendly

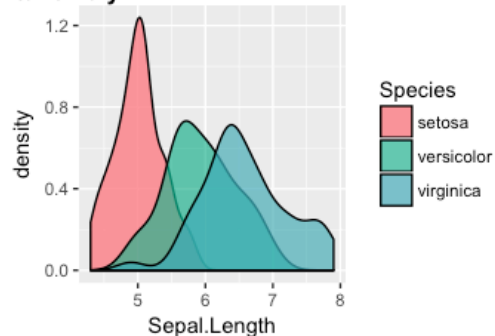
Deutanomaly



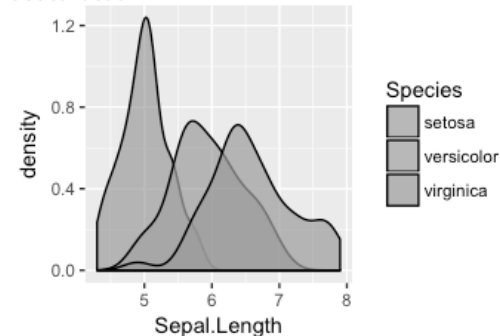
Protanomaly



Tritanomaly



Desaturated





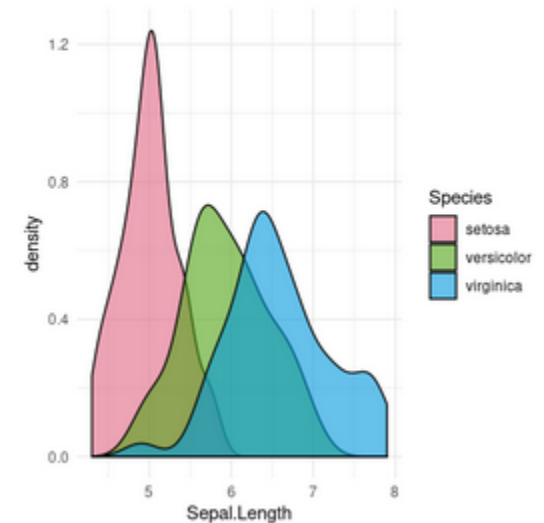
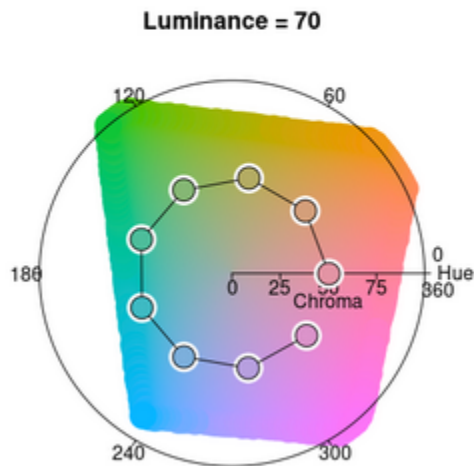


## A Toolbox for Manipulating and Assessing Colors and Palettes

Color spaces

HCL-based palettes

ggplot2 scales




See: <http://colorspace.r-forge.r-project.org/>

# colorspace: palette visualization

swatchplot(): display collections of palettes in flexible ways

Hue



Effect of varying hue, chroma and luminance individually

Chroma



Luminance



```
swatchplot(  
  "Hue" = sequential_hcl(5, h = c(0, 300), c = c(60, 60), l = 65),  
  "Chroma" = sequential_hcl(5, h = 0, c = c(100, 0), l = 65, rev = TRUE, power = 1),  
  "Luminance" = sequential_hcl(5, h = 260, c = c(25, 25), l = c(25, 90), rev = TRUE, power = 1),  
  off = 0  
)
```

Original



Deuteranope



Protanope



Tritanope



Desaturated



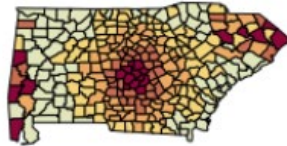
Emulate different types of color vision deficiency for one or more palettes

```
swatchplot(palette.colors(), cvd = TRUE)
```

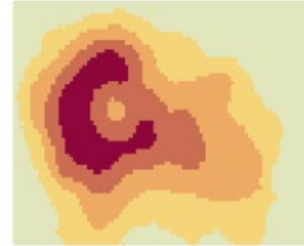
# colorspace: demoplot()

```
demoplot(sequential_hcl(5, "Heat"))
```

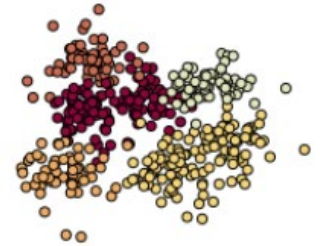
map



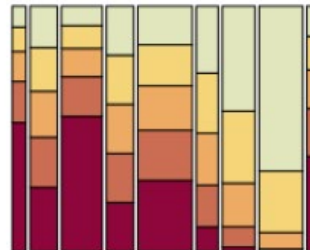
heatmap



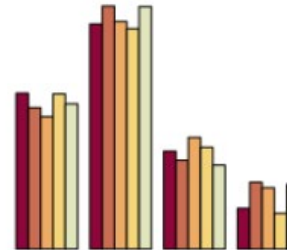
scatter



spine



bar

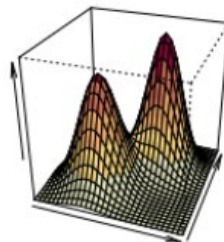


pie

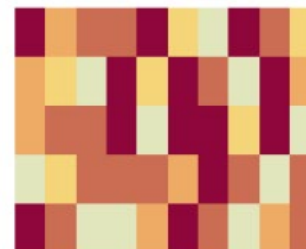


See how color palettes  
work in different kinds  
of statistical displays

perspective



mosaic

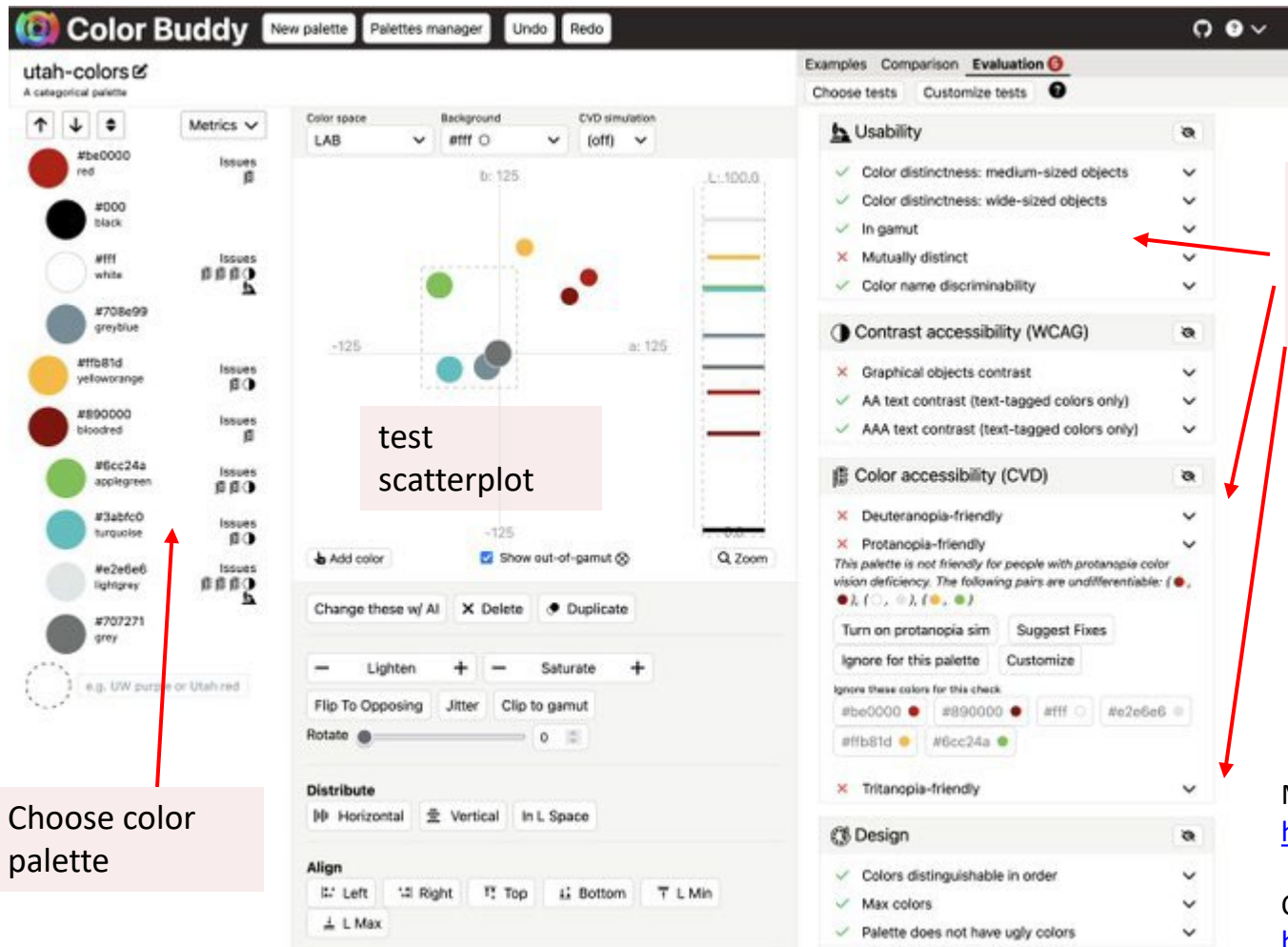


lines



# Color Buddy

McNutt et al (2024) Mixing Linters with GUIs: A Color Palette Design Probe  
→ Color Buddy Application for color palette editing, with usability testing



Palette is evaluated along a set of criteria

McNutt et al.  
<https://arxiv.org/abs/2407.21285>

Color Buddy App: <https://color-buddy.netlify.app/>

# Color: Lessons

- Use colors to represent differences in meaning
  - Avoid gratuitous use of multiple colors
  - Use consistent color scheme across multiple graphs of the same data (set your palette())
- Consider presentation goal:
  - Highlight one subset against the rest?
  - Group a categorical variable
  - Encode a quantitative variable
- Consider differences in color perception, B/W printing

# Color: Lessons

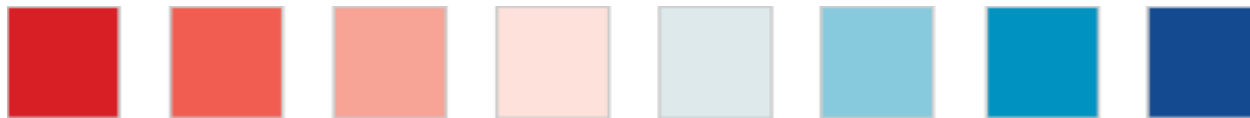
- Consider encoding scheme:
  - Categorical: Use a wide range of hues, of ~ same saturation



- Sequential: use a small range of hues of varying lightness



- Diverging: Use two sequential schemes, decreasing toward the middle





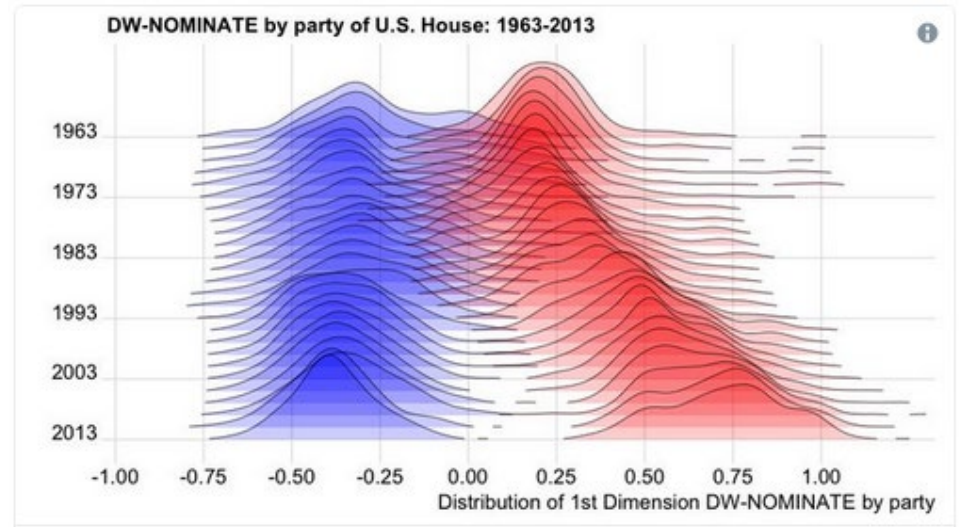
# Transparency

Colors can be made partially **transparent**, by adding an “alpha” channel,

$$0 \leq \alpha \leq 1 \text{ (opaque)}$$

Filled areas combine to look more saturated

What do you see here?

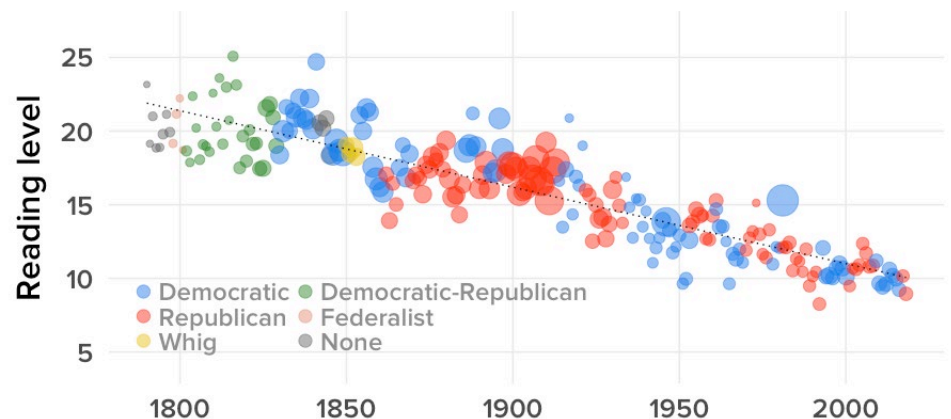


Increasing polarization of votes in the US House

This also works well with filled point symbols, which would otherwise be obscured when they overlap

Different colors “blend”

What do you see here?

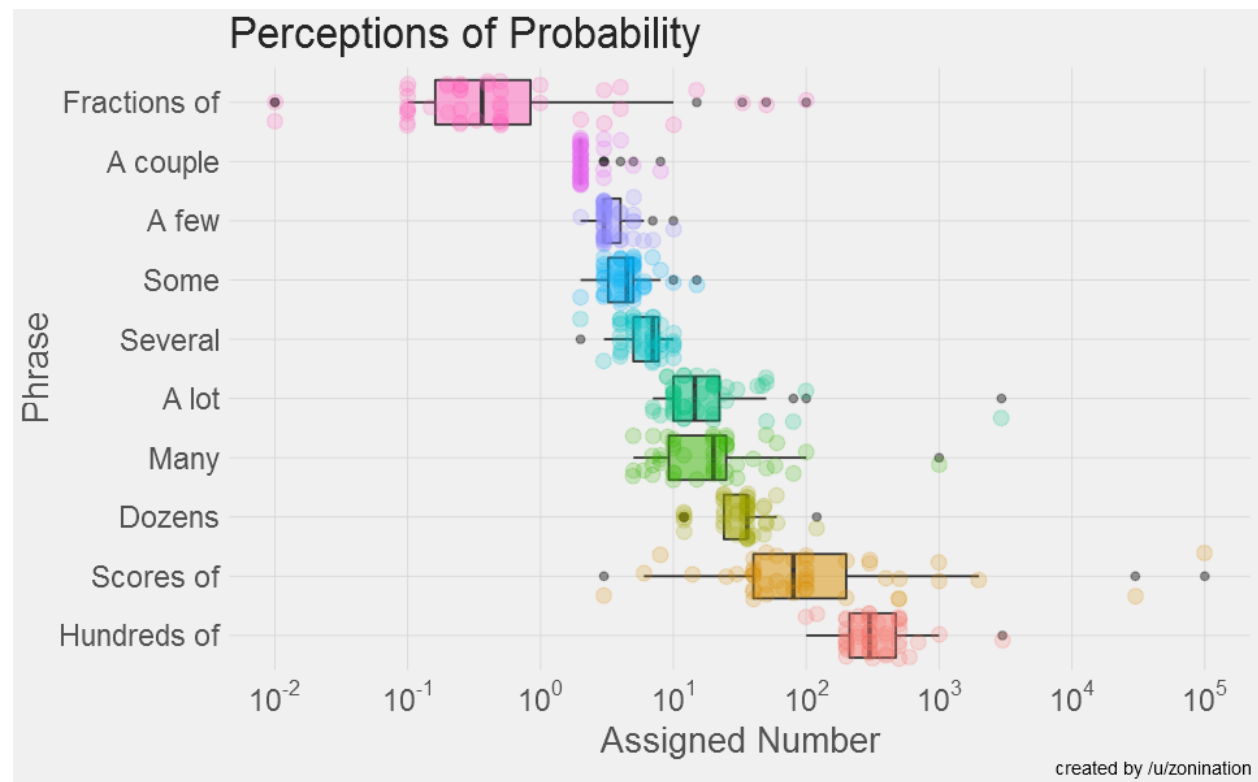


Reading level of US State of the Union Addresses 102

# Transparency: Adding another layer

Transparency also works well to combine different graphical features in a plot  
Here, a filled boxplot and dots representing individual observations

What number would you assign to the following phrases?



From: <https://github.com/zonination/perceptions>

# Summary

- In designing data graphics, consider the viewer
  - Info → encoding → image → decoding → understanding
- Perception: much is known, with ~ links to graphics
  - ↑ Bottom up: perceptual features, what grabs attention
  - ↓ Top down: expectations provide a context
  - 👁 Encoding attributes must consider what is to be seen
- Color: What is the presentation goal?
  - Color palettes for different purposes
  - Transparency increases the effective use of color
- Cognition: What is the viewer's TASK?