

# Data Visualization

```
graph TD; DV[Data Visualization] --> DA([Data Analysis]); DV --> D([Design]); DV --> B([Build]); DA --> SK[Statistics Knowledge]; D --> CG[Computer Graphics]; B --> P[Programming]; style D stroke:#f00,stroke-width:4px; style CG stroke:#f00,stroke-width:4px;
```

Data  
Analysis

Design

Build

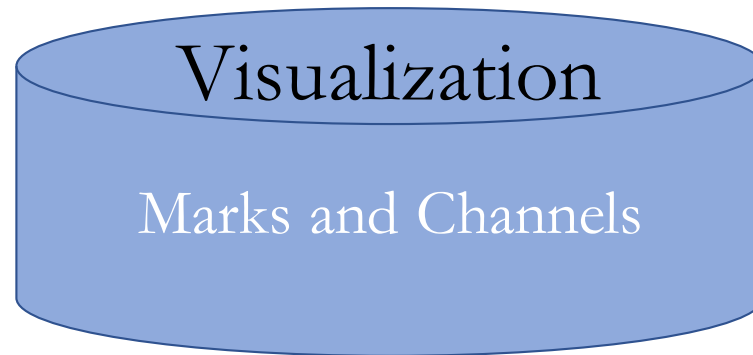
Statistics  
Knowledge

Computer  
Graphics

Programming

# Design data visualization

Graphics components for data visualization



- “Design is a plan for **arranging elements** in such a way as best to accomplish a **particular purpose**.”- Charles Eames, designer

# Marks & Channels

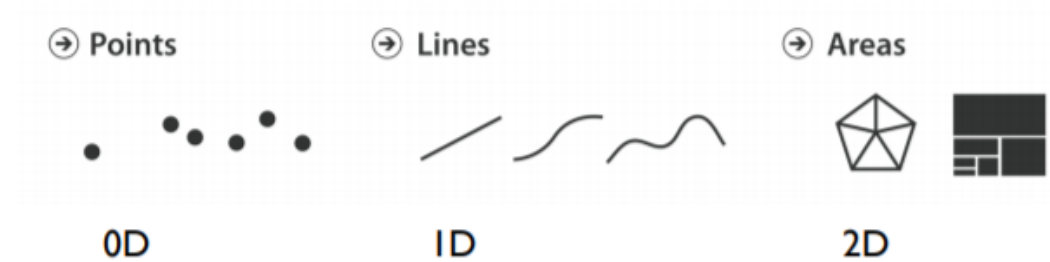
**Marks:** represent **items** or **links**

**Channels:** change **appearance** based on **attribute**

**Channel = Visual Variable**

# Marks for Items

Basic geometric elements



3D mark: Volume, but rarely used

# Marks for Links

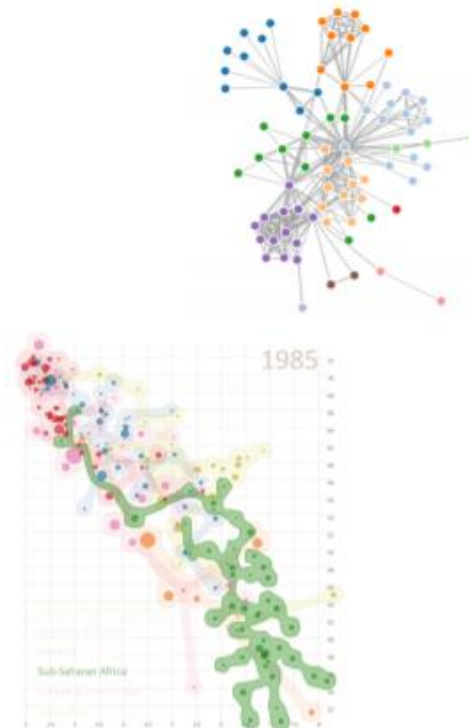
→ Containment



→ Connection



*Marks can be nested.*



# Channels (aka Visual Variables)

Control appearance  
proportional to or  
based on attributes

→ Position

→ Horizontal → Vertical → Both

→ Color

→ Shape

→ Tilt

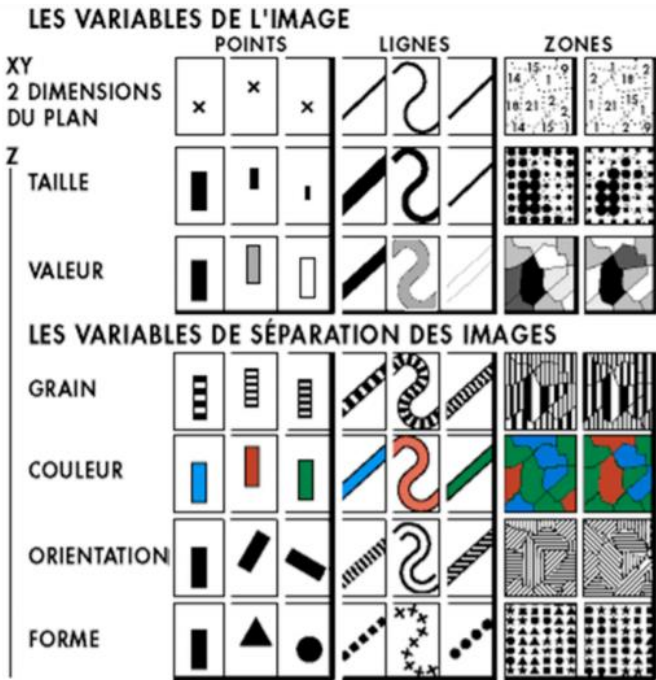
→ Size

→ Length → Area → Volume

Text: Font

Channels

- Position
- Size
- (Grey)Value
- Texture
- Color
- Orientation
- Shape



# Channels: Rankings

## ➔ **Magnitude** Channels: **Ordered** Attributes

Position on common scale 

Position on unaligned scale 


Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

Best


Effectiveness

Least

## ➔ **Identity** Channels: **Categorical** Attributes

Spatial region 

Color hue 

Motion 


Shape 

- **expressiveness principle**
  - match channel and data characteristics
- **effectiveness principle**
  - encode most important attributes with highest ranked channels

# Channels: Rankings

## ➔ Magnitude Channels: Ordered Attributes

Position on common scale 

Position on unaligned scale 

Length (1D size) 

Tilt/angle 


Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

Best

Effectiveness

Least

## ➔ Identity Channels: Categorical Attributes

Spatial region 

Color hue 

Motion 

Shape 

- expressiveness principle
  - match channel and data characteristics
- effectiveness principle
  - encode most important attributes with highest ranked channels



# Color

The psychology and theory of color is extremely relevant to marketing. After all, people decide how they feel about a product within 90 seconds—and researchers have found up to 90% of that judgement is solely based on color. Plus, further studies have found consumers are more likely to buy when a color feels “right” for the brand.

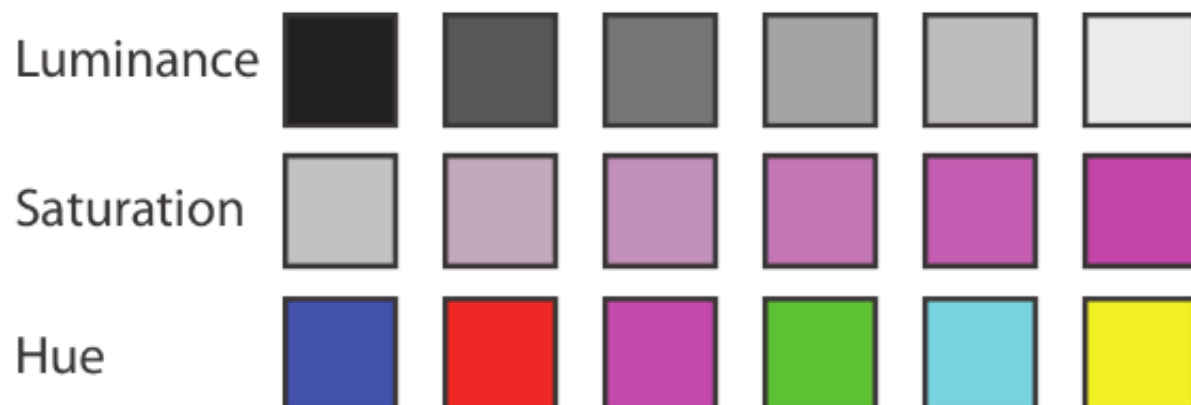
**Warm colors**, like red, orange, yellow, and their variations, project passion, warmth, happiness, power, and energy.

**Cool colors**, including blue, green, and purple, are calmer and more soothing.

**Neutral colors** include white, black, and gray. Technically, they don't have an emotional effect (but I'd argue “no effect” is an effect in and of itself!)

# Decomposing color

- first rule of color: do not talk about color!
  - color is confusing if treated as monolithic
- decompose into three channels
  - ordered can show magnitude
    - luminance: how bright
    - saturation: how colorful
  - categorical can show identity
    - hue: what color
- channels have different properties
  - what they convey directly to perceptual system
  - how much they can convey: how many discriminable bins can we use?



# Color: Luminance, Saturation, Hue

- 3 channels:

Identity for categorical

Hue

Magnitude for ordered

Luminance (Lightness), Saturation

- Hue

Hue is a degree on the color wheel from 0 to 360. 0 is red, 120 is green, 240 is blue.

- Saturation

Saturation is a percentage value; 0% means a shade of gray and 100% is the full color.

- Lightness

Lightness is also a percentage; 0% is black, 100% is white.

**HSL (hue, saturation, lightness)** and **HSV (hue, saturation, value, also known as HSB or hue, saturation, brightness)**

# Color models

- CMYK Colors

CMYK colors is a combination of CYAN, MAGENTA, YELLOW, and BLACK. CMYK is not supported in HTML, but it is suggested as a new standard in CSS4.

- RGB Colors

An RGB color value is specified with: `rgb(red, green, blue)`. RGB color values are supported in all browsers. e.g. `rgb(0, 0, 255)`

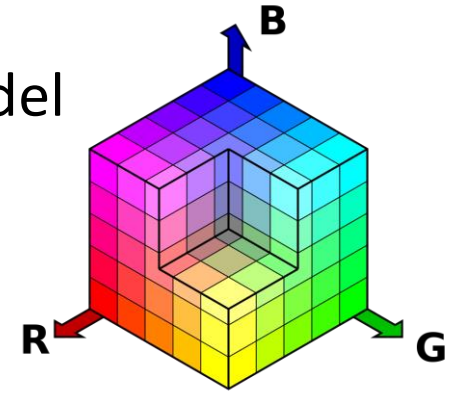
- Hexadecimal Colors

Hexadecimal color values are also supported in all browsers. A hexadecimal color is specified with: `#RRGGBB`. e.g. `#0000FF`

- HSL Colors

HSL color values are supported in IE9+, Firefox, Chrome, Safari, and in Opera 10+. HSL stands for hue, saturation, and lightness. HSL color values are specified with: `hsl(hue, saturation, lightness)`.

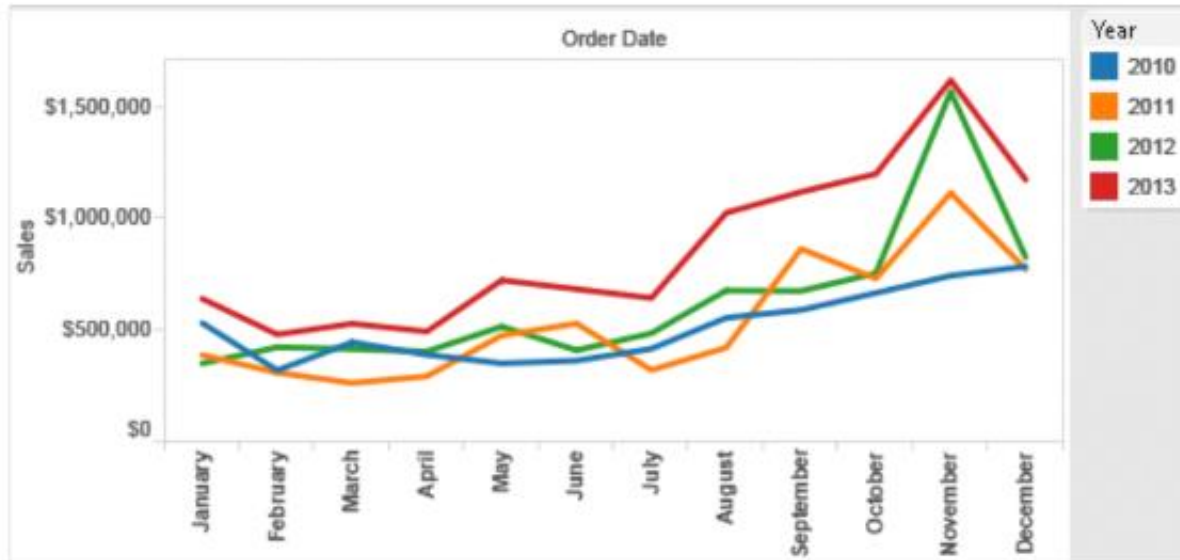
RGB model



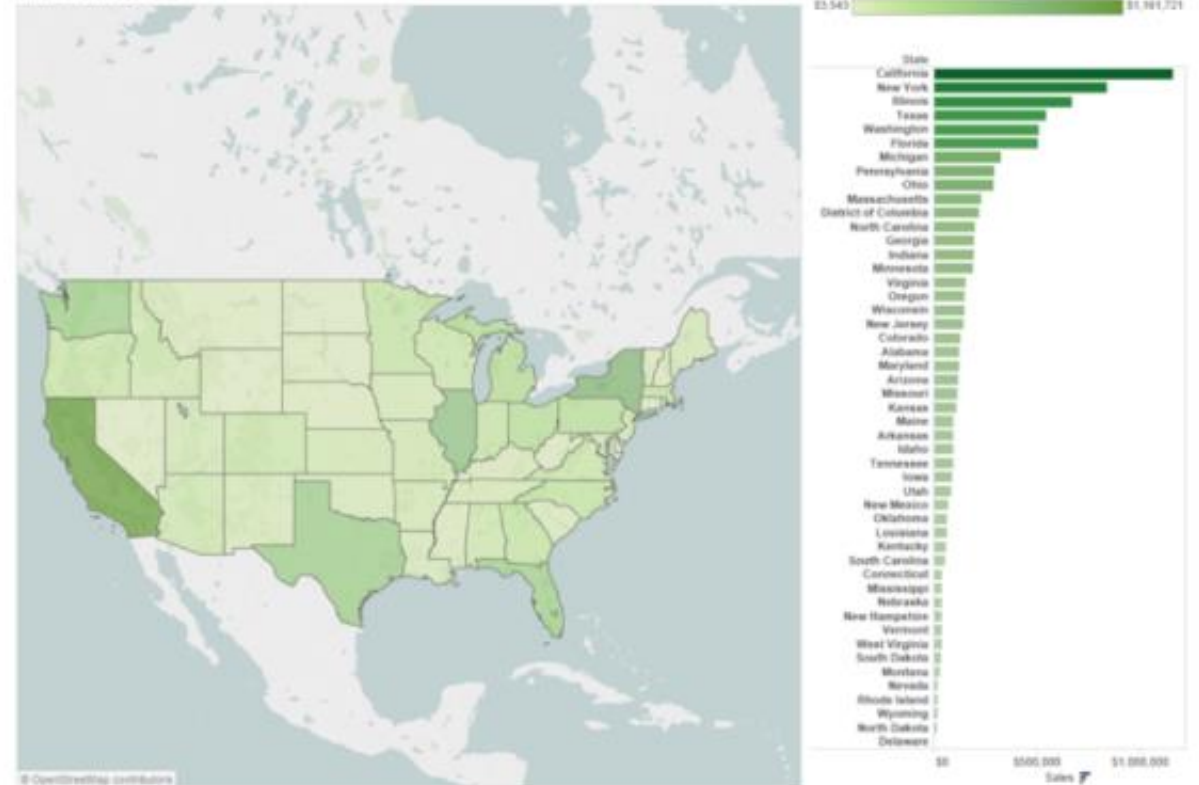
*Color picker:*

[https://www.google.com/search?safe=active&rlz=1C1GCEA\\_enUS921US922&sxsrf=ALeKk02ixjELih12vN1VnSMHbQLSTJUzig:1603637772090&q=HEX+color&spell=1&sa=X&ved=2ahUKEwiPt4OegNDsAhWEZ80KHXQNBoUQirwEKAB6BAgMECo&biw=1536&bih=698](https://www.google.com/search?safe=active&rlz=1C1GCEA_enUS921US922&sxsrf=ALeKk02ixjELih12vN1VnSMHbQLSTJUzig:1603637772090&q=HEX+color&spell=1&sa=X&ved=2ahUKEwiPt4OegNDsAhWEZ80KHXQNBoUQirwEKAB6BAgMECo&biw=1536&bih=698)

# Categorical vs ordered color



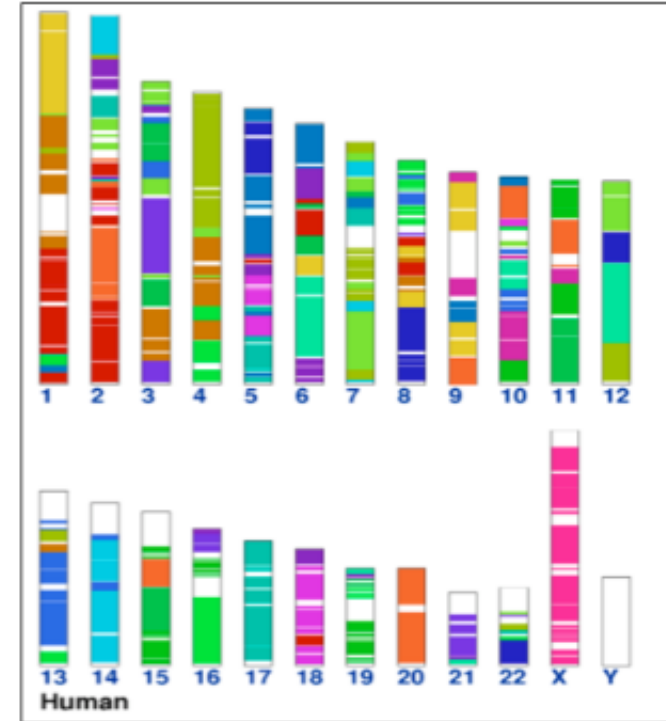
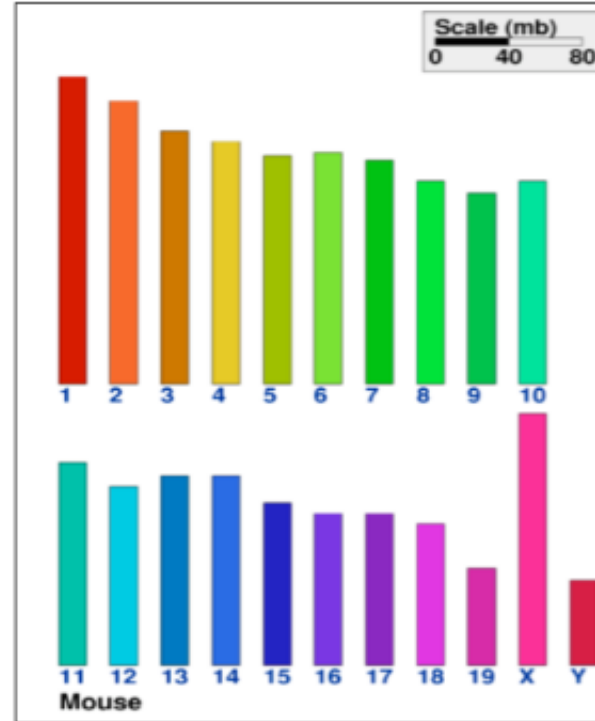
Annual sales by state



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

# Categorical color: limited number of discriminable bins

- Human perception built on relative comparisons
  - great if color contiguous
  - Bad for absolute comparison
- Noncontiguous small regions of color
  - fewer bins than you want
  - Rule of thumb: 6-12 bins, including background and highlights

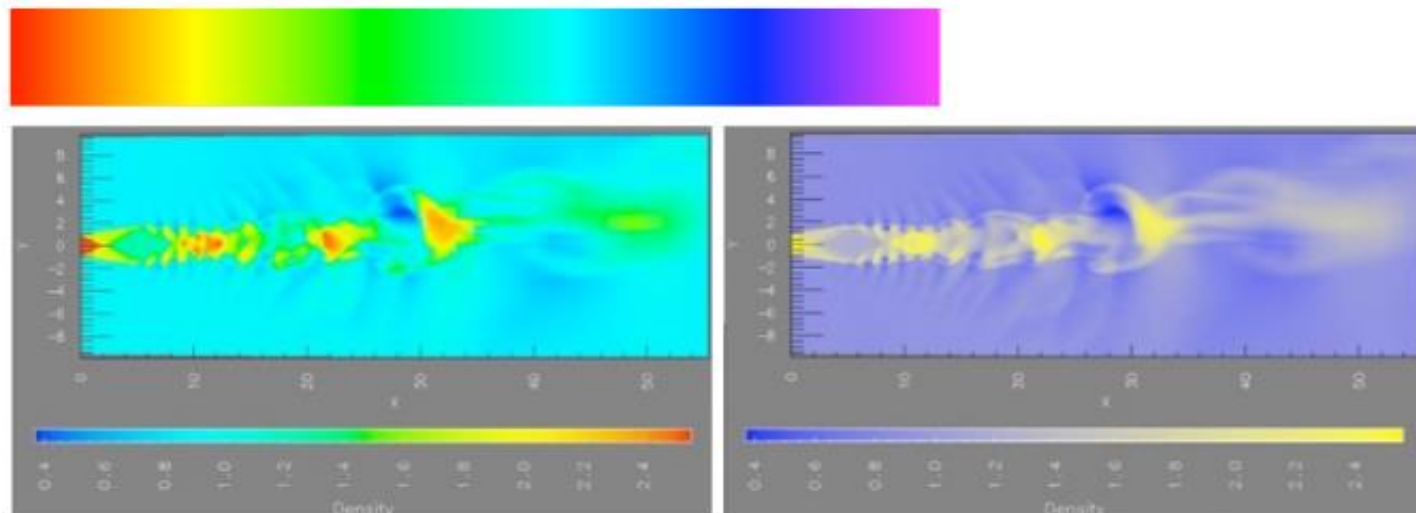


[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

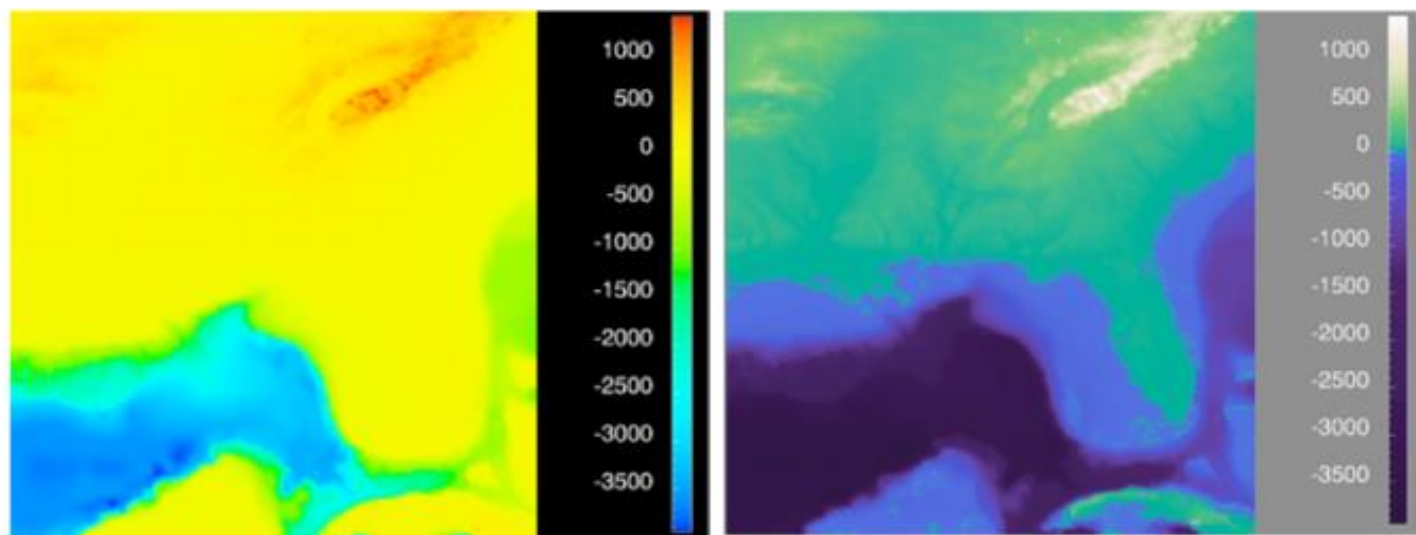


# Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - large-scale structure: fewer hues
  - fine structure: multiple hues with monotonically increasing luminance [eg viridis R/python]
  - segmented rainbows for binned or categorical



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

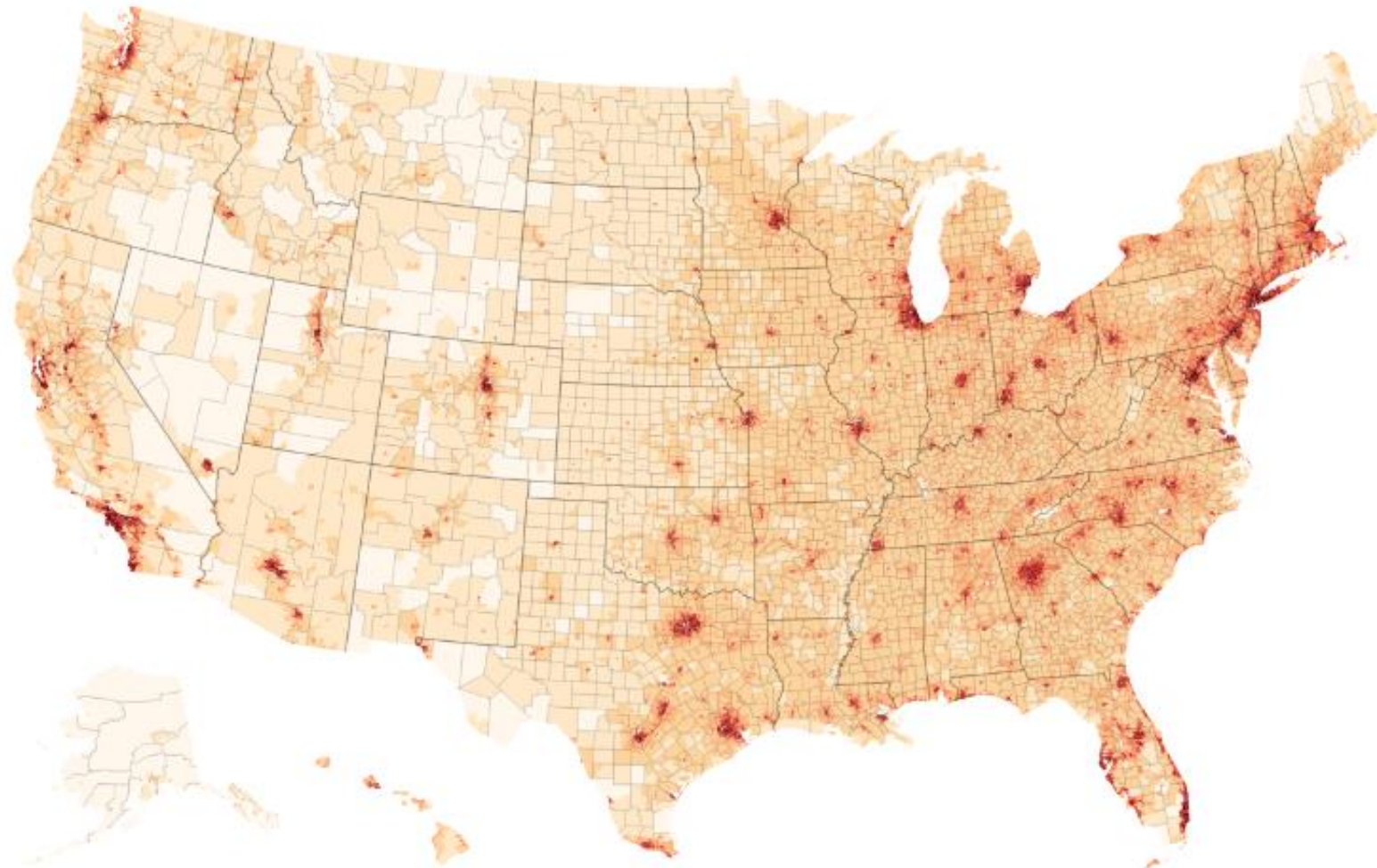


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/lloyd/color/color.HTM>]

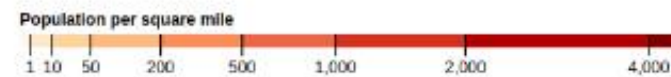
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. *SIGGRAPH 2002 Course Notes*]



# U.S. Population Density



By Mike Bostock

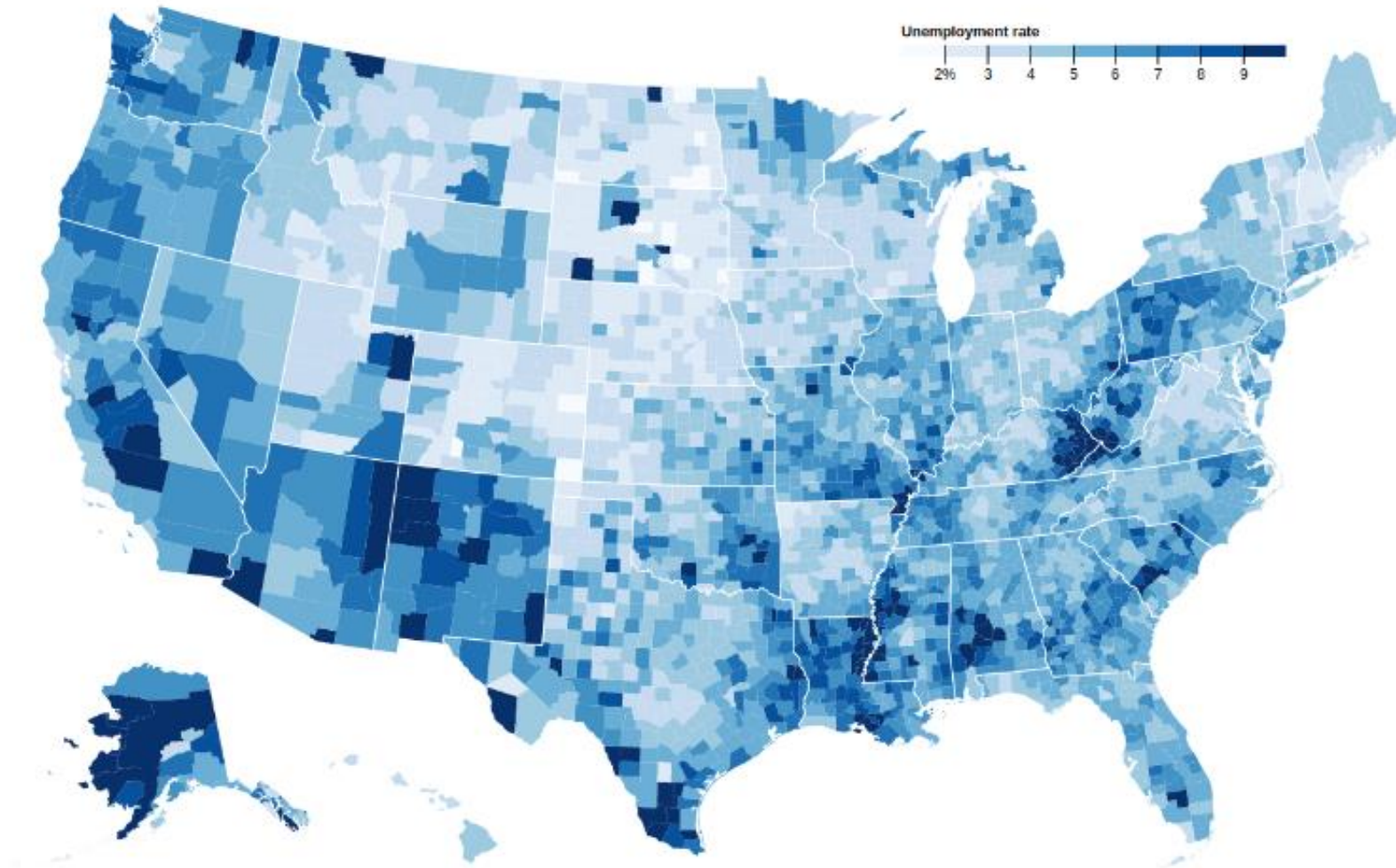


Source: [American Community Survey](#), 2014 5-Year Estimate

[Open](#)



# Choropleth



By Mike Bostock

Source: [Bureau of Labor Statistics](#), [Census Bureau](#)

[Open](#)

This choropleth shows unemployment rates as of August, 2016 with a [threshold scale](#). I employed a mix of command-line tools to transform the fixed-width text file into a CSV, including [dsv2dsv](#):



# Color in JavaScript

Everywhere in the library, wherever it is required to set color, it can be mentioned in “hex Code” , “rgb values” or by writing “Color-Names”.

## Color Values

HTML Colors are defined using a hexadecimal notation (HEX) for the combination of Red, Green, and Blue color values (RGB).

HEX values are specified as 3 pairs of two-digit numbers, starting with a # sign. Some of examples are.

Color Names	Hex Values	Color RGB
Black	#000000	rgb(0,0,0)
Red	#FF0000	rgb(255,0,0)
Green	#00FF00	rgb(0,255,0)
Blue	#0000FF	rgb(0,0,255)
White	#FFFFFF	rgb(255,255,255)

# Choosing Text Fonts for Data Visualization

- A minimum **font size** of 9pt for screens and 6pt for print. Though there are no rigid requirements for compliance regarding **font sizes**, striving for **font sizes** as close to 16pt as possible will ensure your **data visualizations** are as readable and user-friendly as possible.
- Here are some ideas to consider when you choose font sizes and formats:
  - Make important text a little bigger.
  - Consistency is key.
  - Avoid using all caps.
  - Each user's web browser may be different.
  - Avoid fancy or custom fonts.

# Expressiveness + Effectiveness

- expressiveness principle:
  - visual encoding should express all of, and only, the information in the dataset attributes
  - lie factor
- effectiveness principle:
  - importance of the attribute should match the salience of the channel
  - data-ink ratio

# Expressiveness + Effectiveness

- expressiveness principle:
  - visual encoding should express all of, and only, the information in the dataset attributes
  - lie factor
- effectiveness principle:
  - importance of the attribute should match the salience of the channel
  - data-ink ratio