Visualization for Data Science
Encoding Channels II

Administrivia – Office Hours

TA Office Hours (online)

Monday 11:00am – 12pm

Wednesday 10:00 – 11am

Instructor Office Hours

Monday and Wednesday 5 – 6pm (currently in this room)

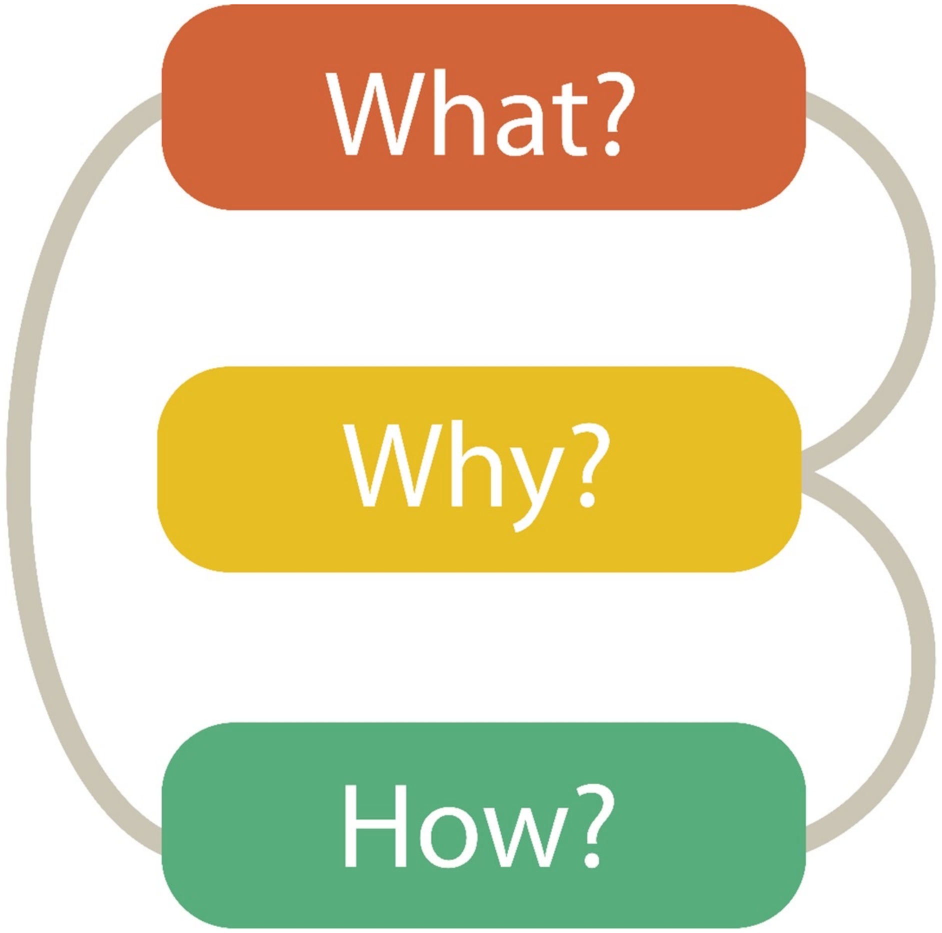
Tuesdays 2 – 3pm ICCS 227

Participation is Clickers, not PL (unless otherwise stated)

Learning Outcomes

- **Analyze and critique** visualizations by evaluating their encoding channels using the five core properties (accuracy, discriminability, pop-out, separability, grouping)
- **Identify and explain** how each encoding channel is implemented across different visualization types and contexts
- **Select appropriate encoding channels** for specific data attributes by applying theoretical principles to match channel characteristics with data types

Data Visualization Ecosystem



What data is shown?

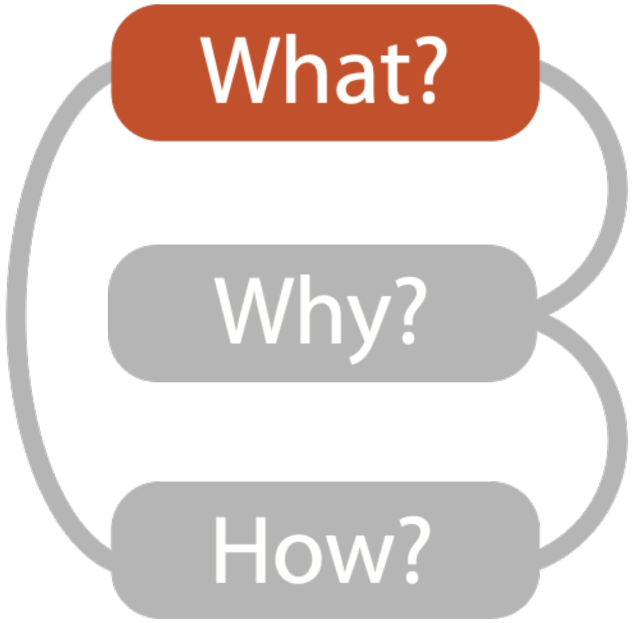
DATA ABSTRACTION

Why is the user analyzing / viewing it?

TASK ABSTRACTION

How is the data presented?

VISUAL AND INTERACTION ENCODING



What?

Datasets

Data Types

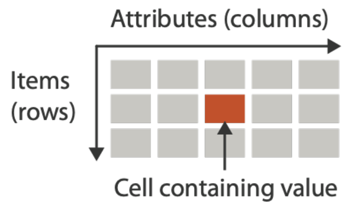
→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

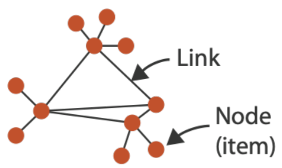
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

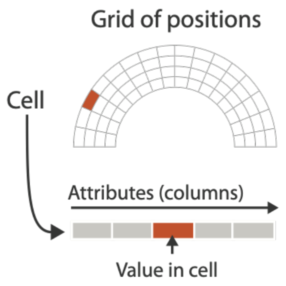
→ Tables



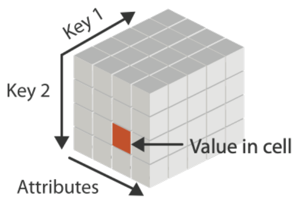
→ Networks



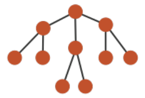
→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*



→ Geometry (Spatial)

Dataset Availability

→ Static

→ Dynamic

Attributes

Attribute Types

→ Categorical

→ Ordered

→ *Ordinal*

→ *Quantitative*

Ordering Direction

→ Sequential

→ Diverging

→ Cyclic

How?

Encode

Arrange
- Express
- Separate
- Order
- Align
- Use

Map
from categorical and ordered attributes
- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion
  Direction, Rate, Frequency, ...

Manipulate

Change

Select

Navigate

Facet

Juxtapose

Partition

Superimpose

Reduce

Filter

Aggregate

Embed

What?
Why?
How?

RECAP Clicker

A data visualization represents years of experience (an ordered attribute) using different shapes (an identity channel). Why might this design choice be criticized?

A. Shapes are always more engaging than other channels and adhere to the expressiveness principle

B. Shapes, as identity channels, do not inherently convey an order, violating the expressiveness principle when used for ordered data.

C. Using shapes ensures that the visualization is universally understood and keeps with the expressiveness principle

D. Years of experience should always be represented by color hue to capitalize on the expressiveness principle

RECAP Clicker

In a chart displaying different brands (a categorical attribute), the designer decided to use length (a magnitude channel) to differentiate them. What potential issue might arise from this design choice?

A. Length is too effective and will overshadow other elements in the visualization.
B. Brands, being categorical, do not have an intrinsic order, and using length might mislead by implying an unwarranted hierarchy or order.
C. Length is universally recognized, and therefore the best choice for any data type.
D. Brands are best represented in 3D visualizations only.

Why These Characteristics Matter: The Science Behind Good Design

Accuracy - *How precisely can we see quantitative differences?*

Foundation: Stevens' Power Law & Cleveland/McGill experiments
Principle: Position is most accurate, area systematically fools us
Design Guideline: Use position/length for precise comparisons, avoid area when accuracy matters

Discriminability - *How many levels can we actually distinguish?*

Foundation: Weber's Law (just noticeable differences)
Principle: We have built-in perceptual limits - only ~6 color lightness levels, ~10 area sizes
Design Guideline: Don't create more categories than people can actually see

Pop-out - *What grabs attention immediately?*

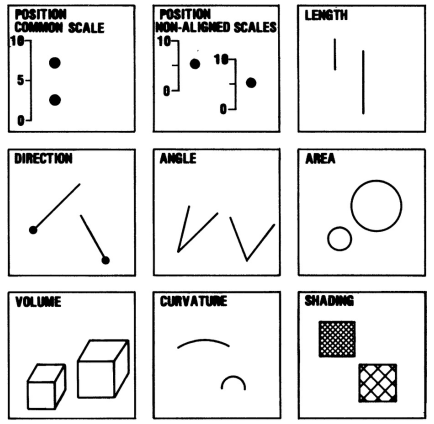
Foundation: Treisman's Feature Integration Theory
Principle: Color and motion jump out instantly, position requires focused attention
Design Guideline: Use color hue to highlight important categories, not for precise values

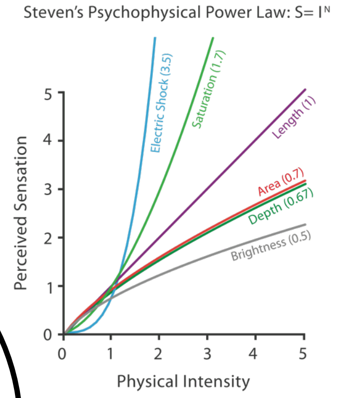
Separability - *Do channels interfere with each other?*

Foundation: Garner's Integral vs Separable Dimensions
Principle: Color hue + lightness interfere; position + shape work independently
Design Guideline: Be careful combining integral channels, use separable ones for multiple variables

Grouping - *How do we naturally organize/group visual information?*

Foundation: Gestalt Principles + Bertin's Visual Variables
Principle: Similar colors/shapes automatically create groups in our minds
Design Guideline: Use consistent encoding to show categories, vary it to show differences





Accuracy

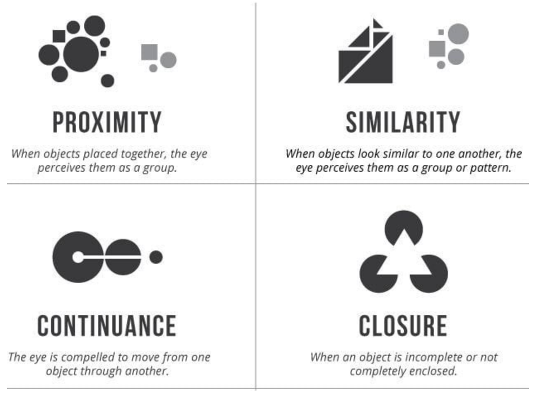
Grouping

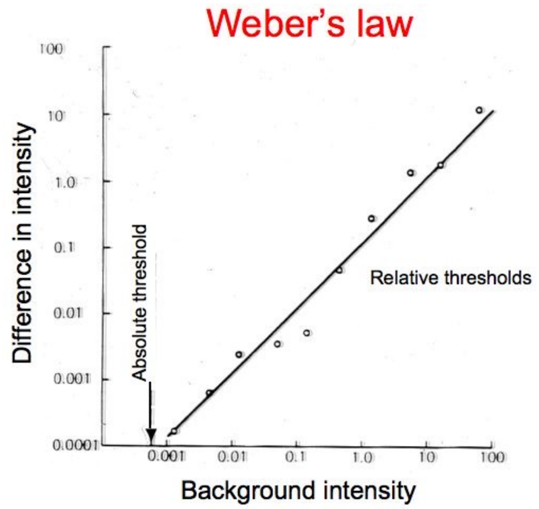
Discriminability

Channel

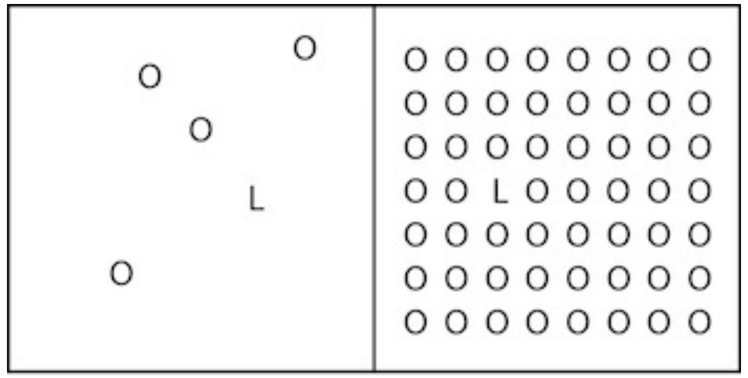
Separability

Pop-out






(a)



Channel Characteristic Analysis

Position on common scale

Position on unaligned scale

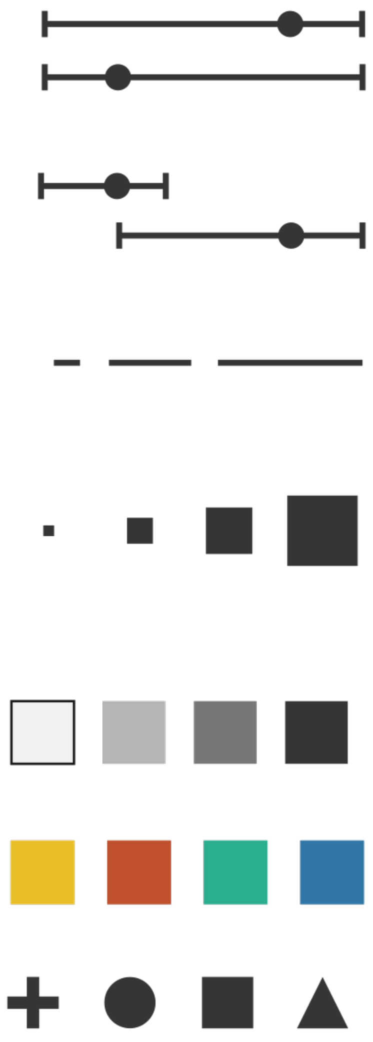
Length (1D size)

Area (2D size)

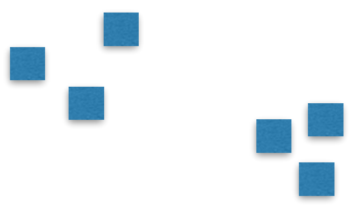
Color luminance

Color hue

Shape



Position



Mapping data values to spatial location along horizontal or vertical axes

Theory

- Cleveland & McGill: Most accurate channel for quantitative comparison
- Stevens' Power Law: Near-perfect linear relationship (exponent ≈ 1.0)
- Weber's Law: Can distinguish ~100+ position levels

Design Impact

- Reserve for most important variables - position is too valuable to **waste**
- Always your first choice for quantitative comparisons
- Combine with other channels for additional dimensions

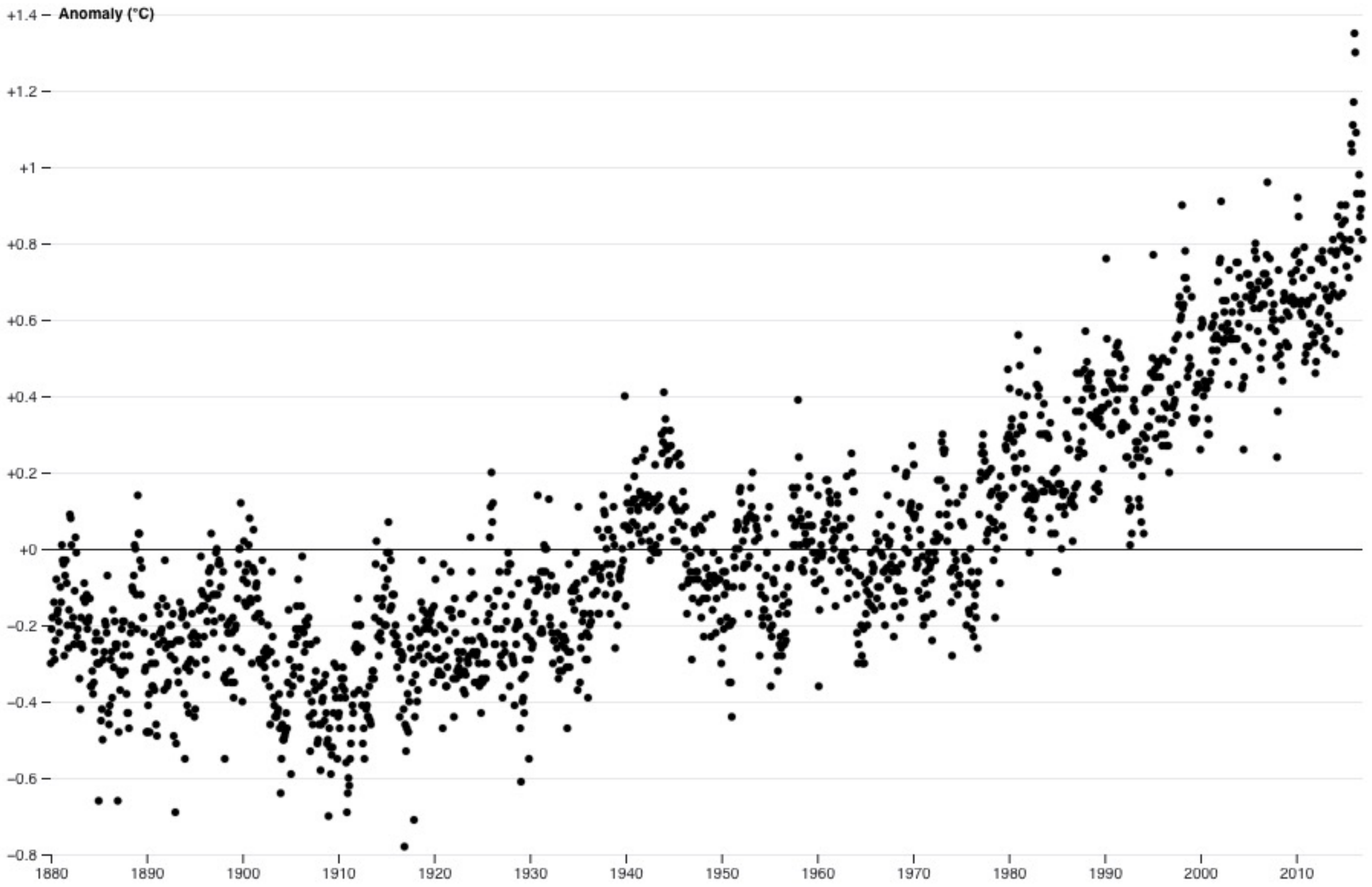
Problems with using it

- Limited real estate: Only 2 primary position channels available
- Scale sensitivity: Requires careful attention to axis ranges & zero baselines
- Overplotting: Points can overlap and obscure patterns in dense datasets

Example: Scatterplot

Global temperature trends

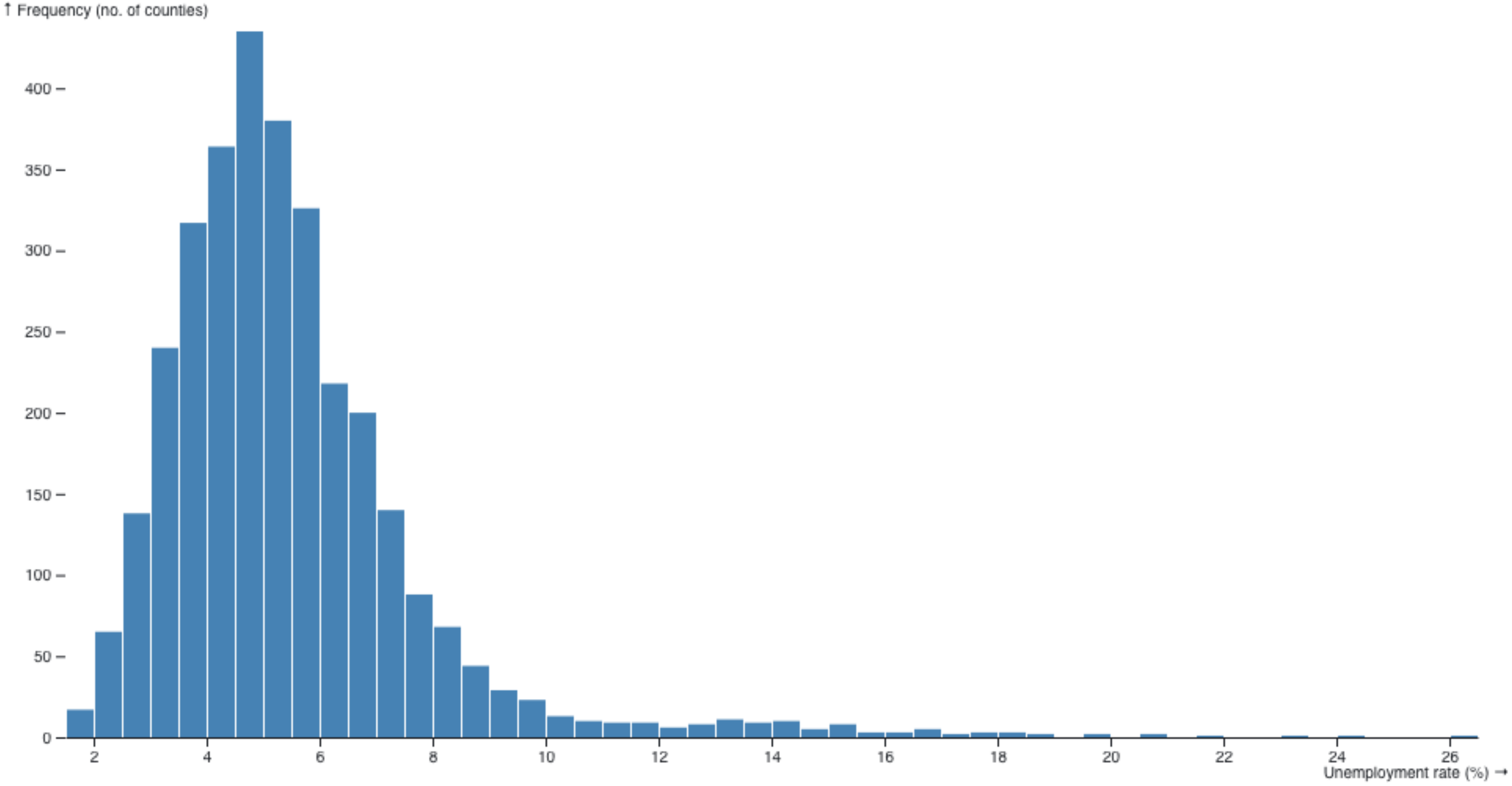
Based on "How 2016 Became Earth's Hottest Year on Record" by Jugal K. Patel. Data: NASA Goddard Institute for Space Studies



Example: Bar Chart (WAIT WHAT?)

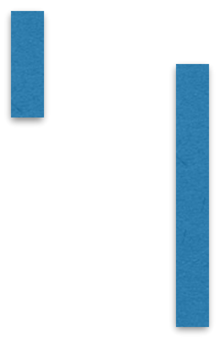
Histogram

A histogram visualizes a one-dimensional distribution by grouping continuous values into discrete bins. This example shows the unemployment rate of U.S. counties as of August 2016. Data: Bureau of Labor Statistics



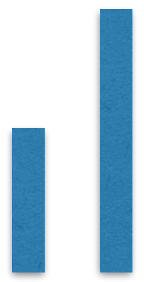
Length

Mapping data values to the linear extent of visual marks



Theory

- Cleveland & McGill: Second most accurate channel after position
- Stevens' Power Law: Nearly linear response (exponent ≈ 1.0)
- Gestalt: Strong comparison through parallel alignment
- Weber’s Law: ~20+ levels (Weber constant ≈ 2-3%)



Design Impact

- Excellent for bar charts and quantitative comparisons
- Use when position channels are occupied by other variables

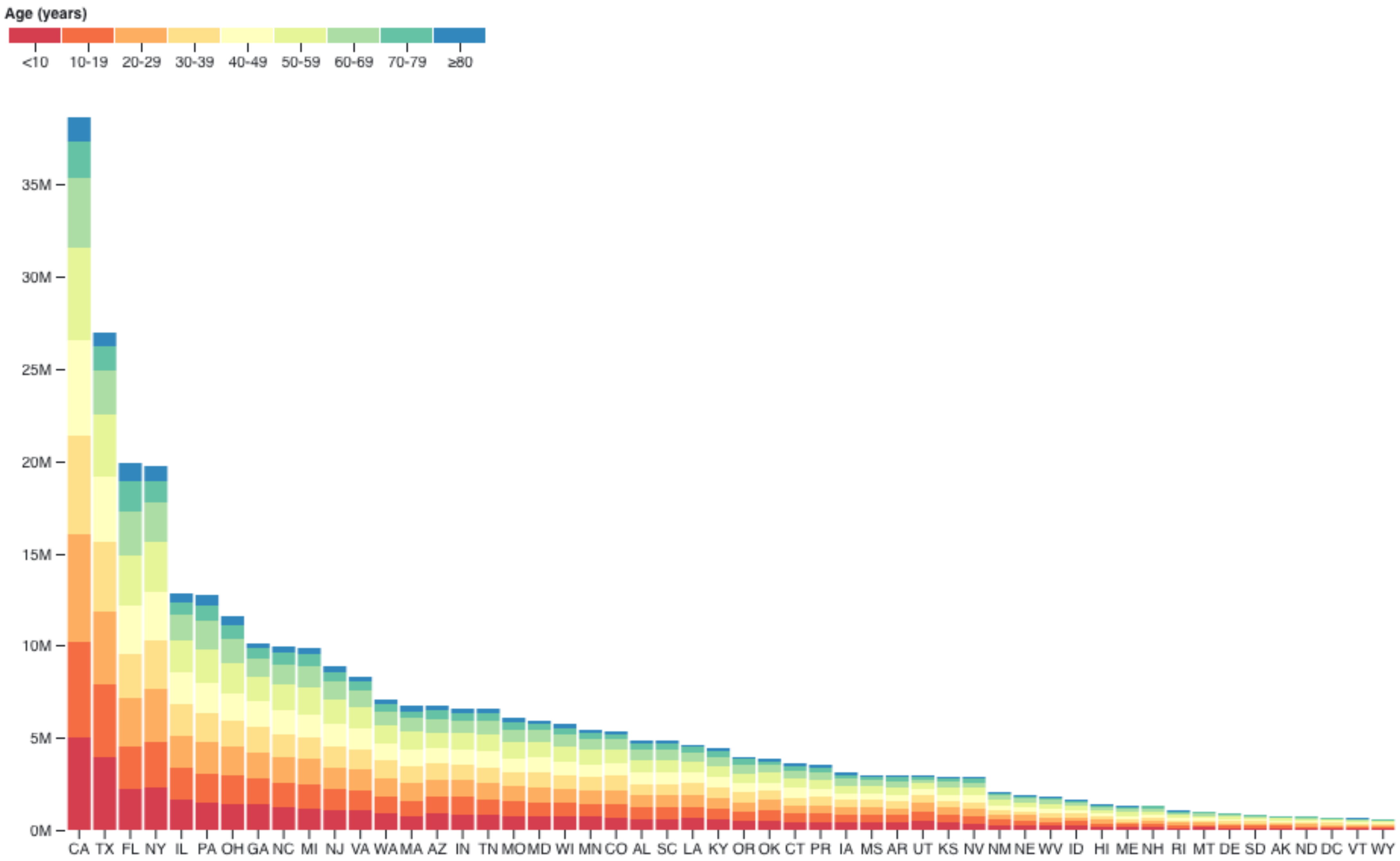
Problems with using it

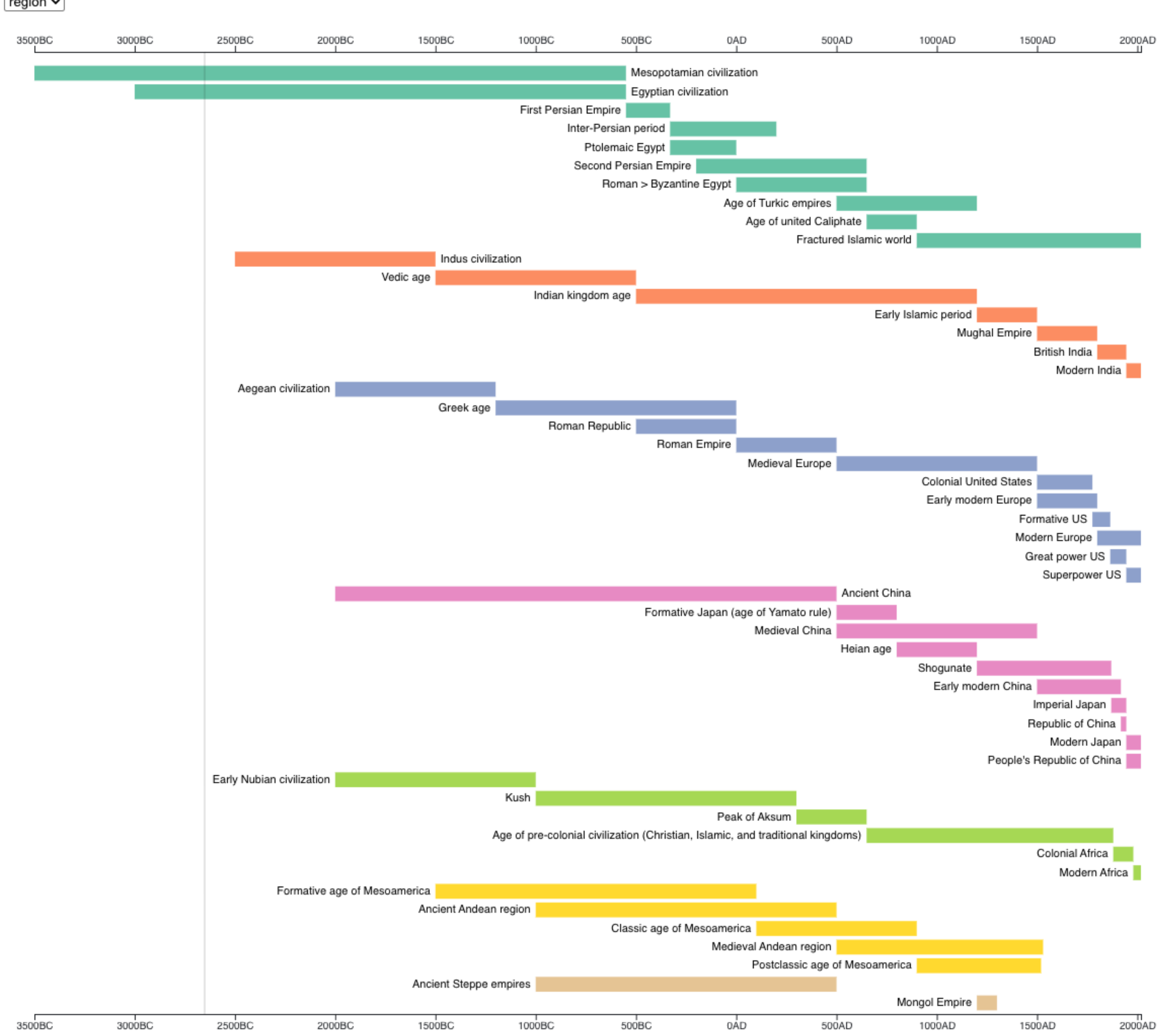
- Baseline dependency: Bars must start from zero or comparisons difficult to percieve
- Orientation effects: Horizontal vs vertical bars can be perceived differently
- Limited space: Can't show too many bars without overcrowding
- Accuracy is influenced by distance, alignment, distractors etc.

Example: Stacked Bar Chart

Stacked bar chart

Population by age and state. Compare to horizontal stacked bars, normalized stacked bars, grouped bars and a dot plot. Data: American Community Survey




region
3500BC
3000BC
2500BC
2000BC
1500BC
1000BC
500BC
0AD
500AD
1000AD
1500AD
2000AD
Mesopotamian civilization
Egyptian civilization
First Persian Empire
Inter-Persian period
Ptolemaic Egypt
Second Persian Empire
Roman > Byzantine Egypt
Age of Turkic empires
Age of united Caliphate
Fractured Islamic world
Indus civilization
Vedic age
Indian kingdom age
Early Islamic period
Mughal Empire
British India
Modern India
Aegean civilization
Greek age
Roman Republic
Roman Empire
Medieval Europe
Colonial United States
Early modern Europe
Formative US
Modern Europe
Great power US
Superpower US
Ancient China
Formative Japan (age of Yamato rule)
Medieval China
Heian age
Shogunate
Early modern China
Imperial Japan
Republic of China
Modern Japan
People's Republic of China
Early Nubian civilization
Kush
Peak of Aksum
Age of pre-colonial civilization (Christian, Islamic, and traditional kingdoms)
Colonial Africa
Modern Africa
Formative age of Mesoamerica
Ancient Andean region
Classic age of Mesoamerica
Medieval Andean region
Postclassic age of Mesoamerica
Ancient Steppe empires
Mongol Empire
3500BC
3000BC
2500BC
2000BC
1500BC
1000BC
500BC
0AD
500AD
1000AD
1500AD
2000AD

Difference between the channels

Position Encoding: Where is this mark along the scale?

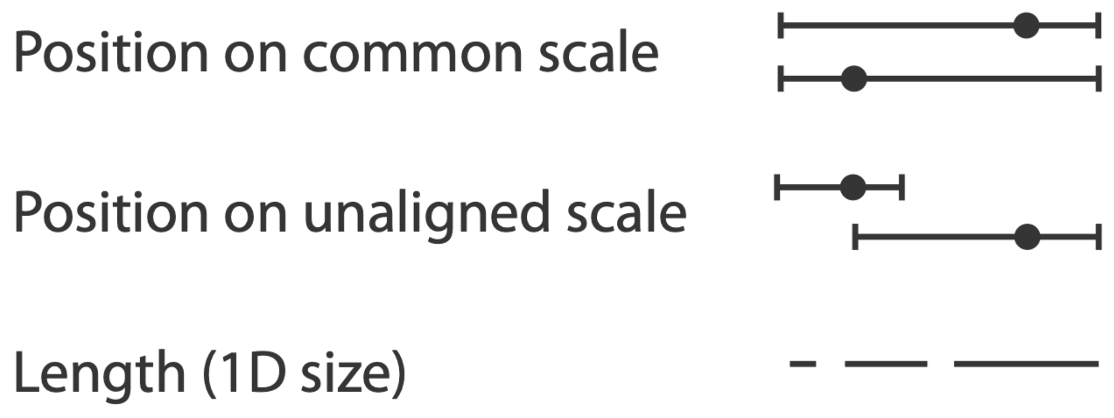
Length Encoding: How long is this bar?

Common scale – must have same baseline and then same scale

Unaligned Scale – you either have a different scale or a different baseline

"Aligned/Common Scale" = Same value is at same position across charts

"Unaligned Scale" = Same position represents different values



Area

Mapping data values to the 2D extent of visual marks (circle sizes, bubble areas)

Theory

- Stevens' Power Law: Area has exponent ≈ 0.7 (underestimation bias)
- Cleveland & McGill: Ranked lower than position/length for accuracy
- Weber's Law: Only ~10-15 distinguishable area levels (Weber constant ≈ 15%)

Design Impact

- Use sparingly and only when position/length unavailable
- Avoid for precise comparisons - reserve for general magnitude sense
- Consider square root scaling to compensate for perceptual bias

Problems with using it

- Systematic underestimation: Humans consistently underperceive area ratios
- Size limitations: Very large differences create unusable visualizations
- Overlapping issues: Large areas obscure smaller ones

Example : Bubbles

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.

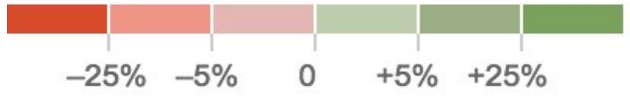
All Spending | Types of Spending | Changes | Department Totals

How $3.7 Trillion Is Spent

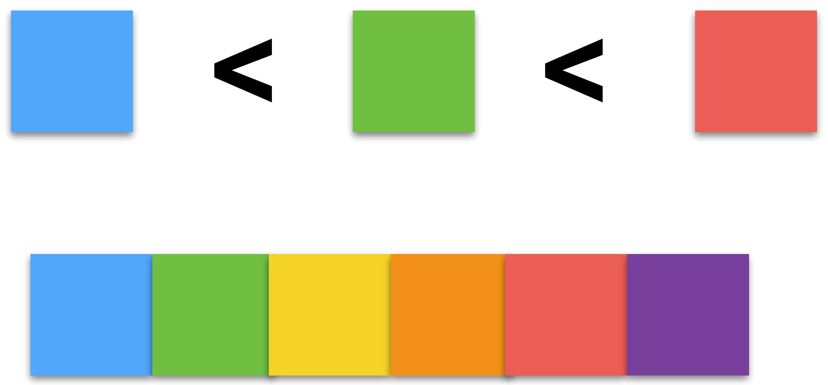
Mr. Obama's budget proposal includes $3.7 trillion in spending in 2013, and forecasts a $901 billion deficit.

Circles are sized according to the proposed spending.

Color shows amount of cut or increase from 2012.



Color Hue



Mapping data values to different colors (red, blue, green, etc.)

Theory

- **Treisman's Feature Integration**: Strong preattentive pop-out effect
- **Bertin's Visual Variables**: Highly selective and associative
- **Garner**: Integral with lightness - hard to perceive independently

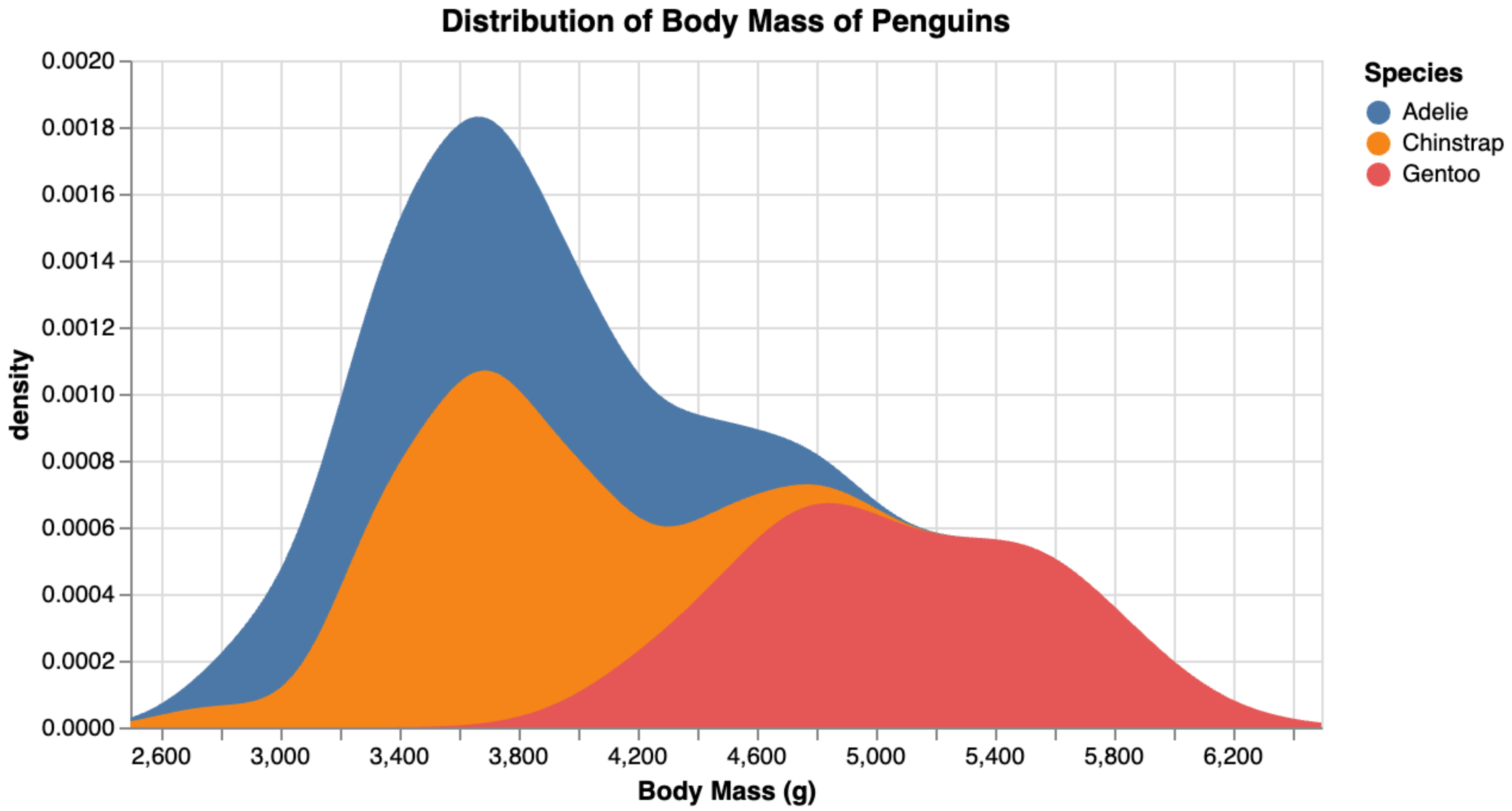
Design Impact

- **Perfect for nominal/categorical data** - creates instant groups
- **Never use for quantitative data** - no natural ordering
- **Always provide alternative encodings** for accessibility

Problems with using it

- **No natural ordering**: Red isn't inherently "more" than blue
- **Accessibility issues**: 8% of men have red-green color blindness
- **Cultural meanings**: Colors carry unintended associations

Good Example: Layered Density Chart



Color Lightness

Mapping data values to how light or dark a color appears (dark blue → light blue)

Theory

- **Bertin**: Naturally ordered visual variable
- **Weber's Law**: Limited just-noticeable differences in lightness, ~4-6 levels (Weber constant ≈ 2%)
- **Garner**: Integral with hue - difficult to separate perceptually

Design Impact

- **Good for ordinal data** with few categories
- **Use high contrast** for better discrimination
- **Test across devices** and printing conditions

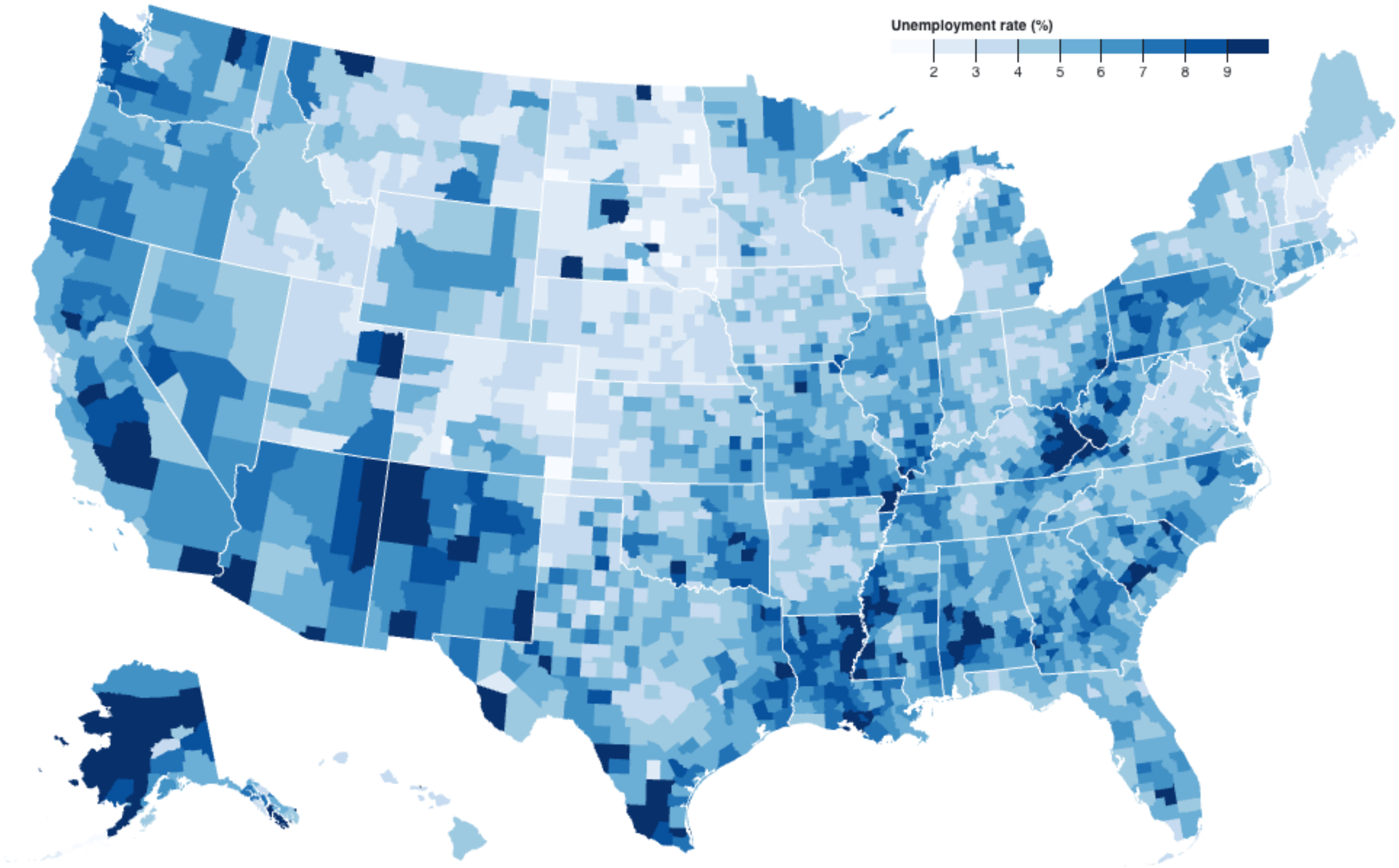
Problems with using it

- **Limited discriminability**: Only ~4-6 distinguishable levels
- **Context dependency**: Same color looks different against different backgrounds
- **Printer/screen variations**: Colors display differently across media

Choropleth Map

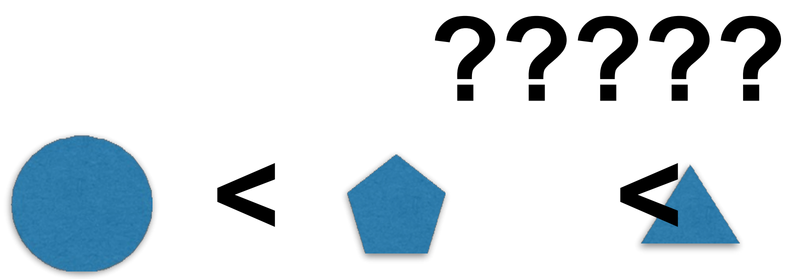
Choropleth

Unemployment rate by U.S. county, August 2016. Data: Bureau of Labor Statistics.



Shape

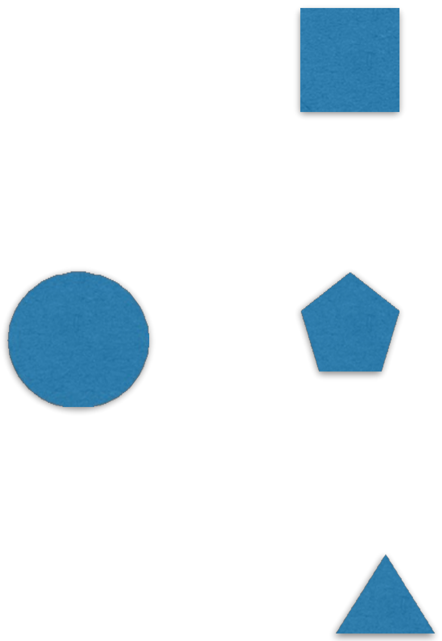
Mapping data values to different geometric forms



Theory

- Treisman: Some shapes create moderate pop-out effects
- Gestalt: Shapes naturally group similar categories
- Cognitive load: Shape recognition requires processing time

Design Impact

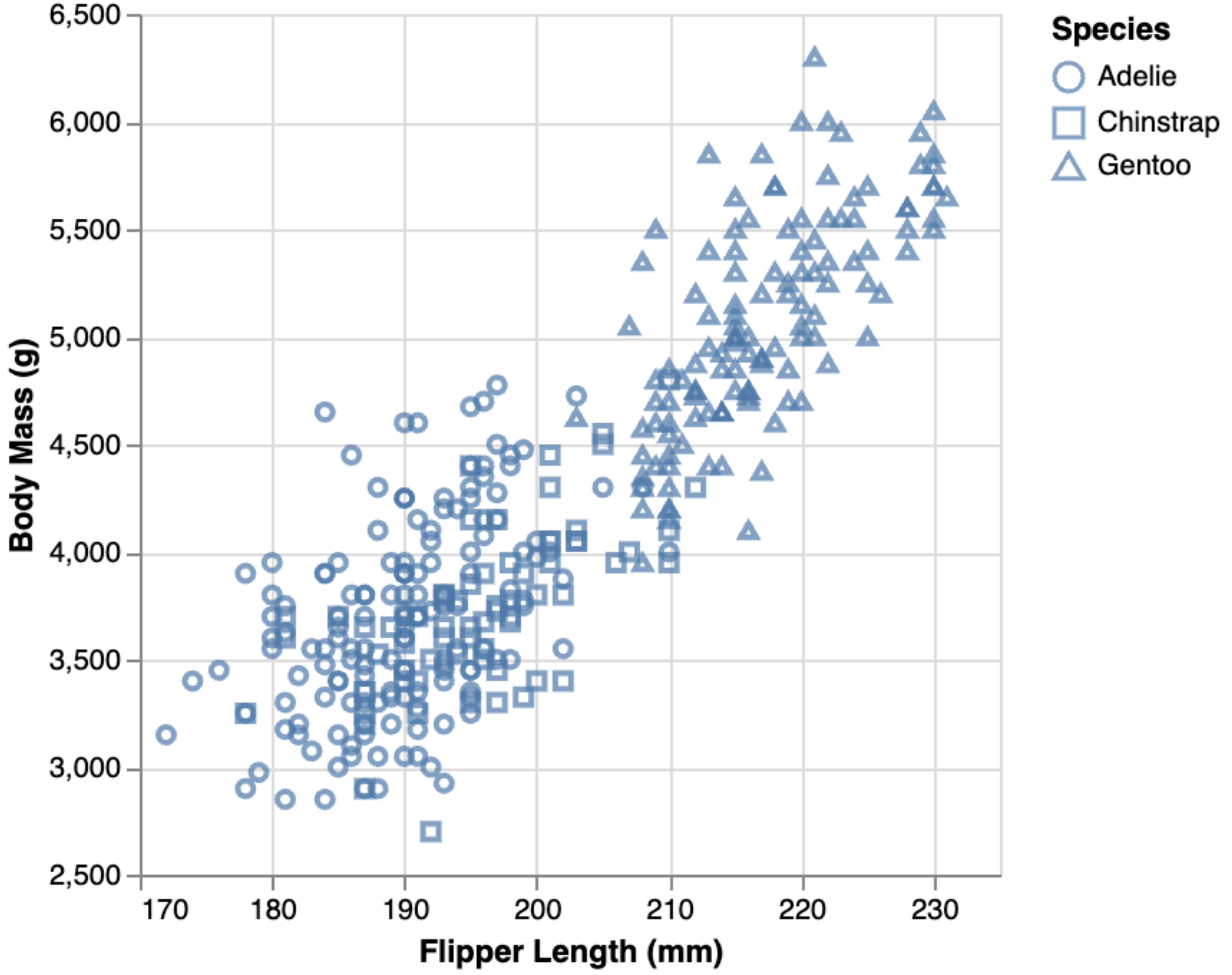


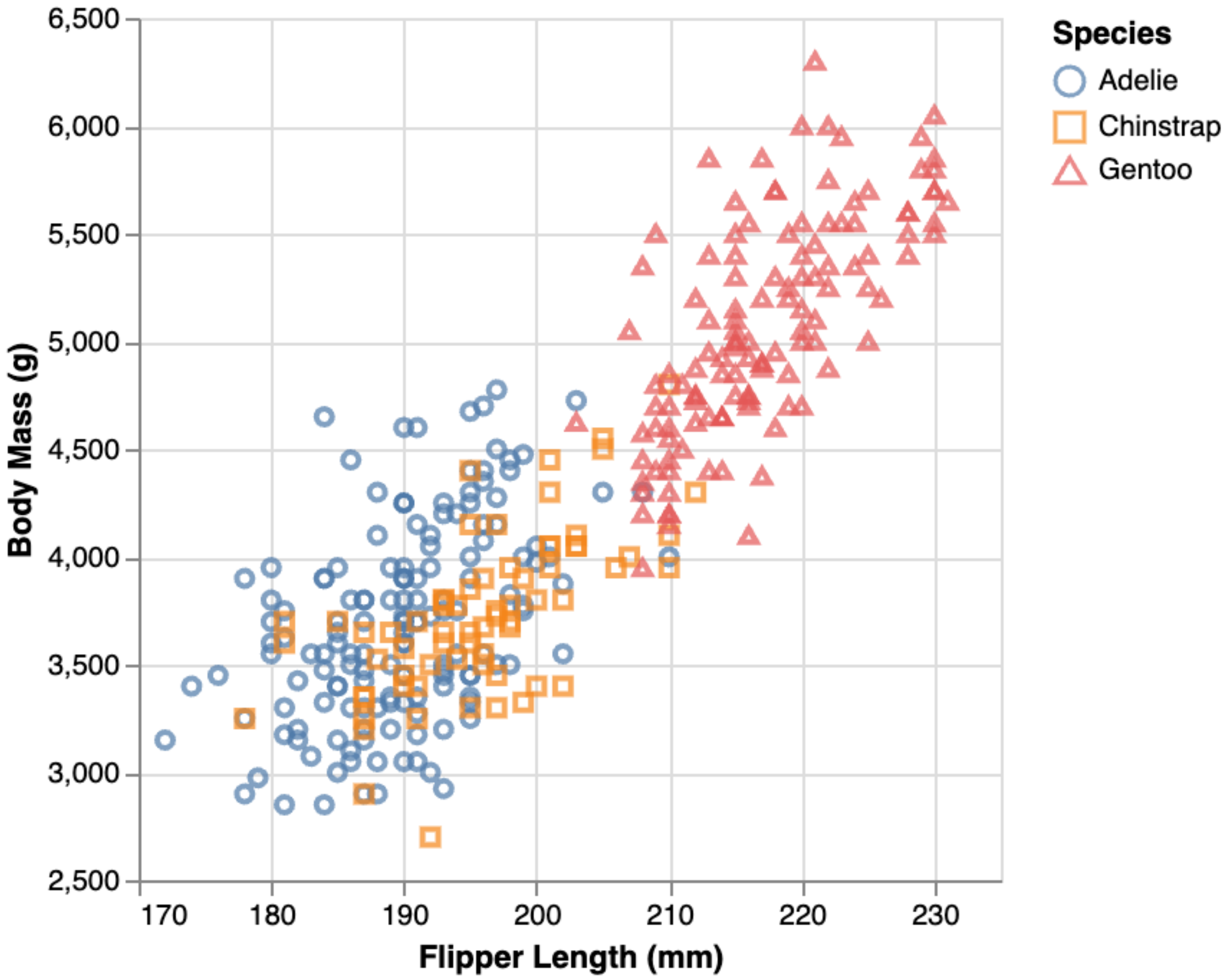
- Excellent for small numbers of categories (2-6 groups)
- Combine with color for redundant encoding
- Use familiar, simple shapes for quick recognition

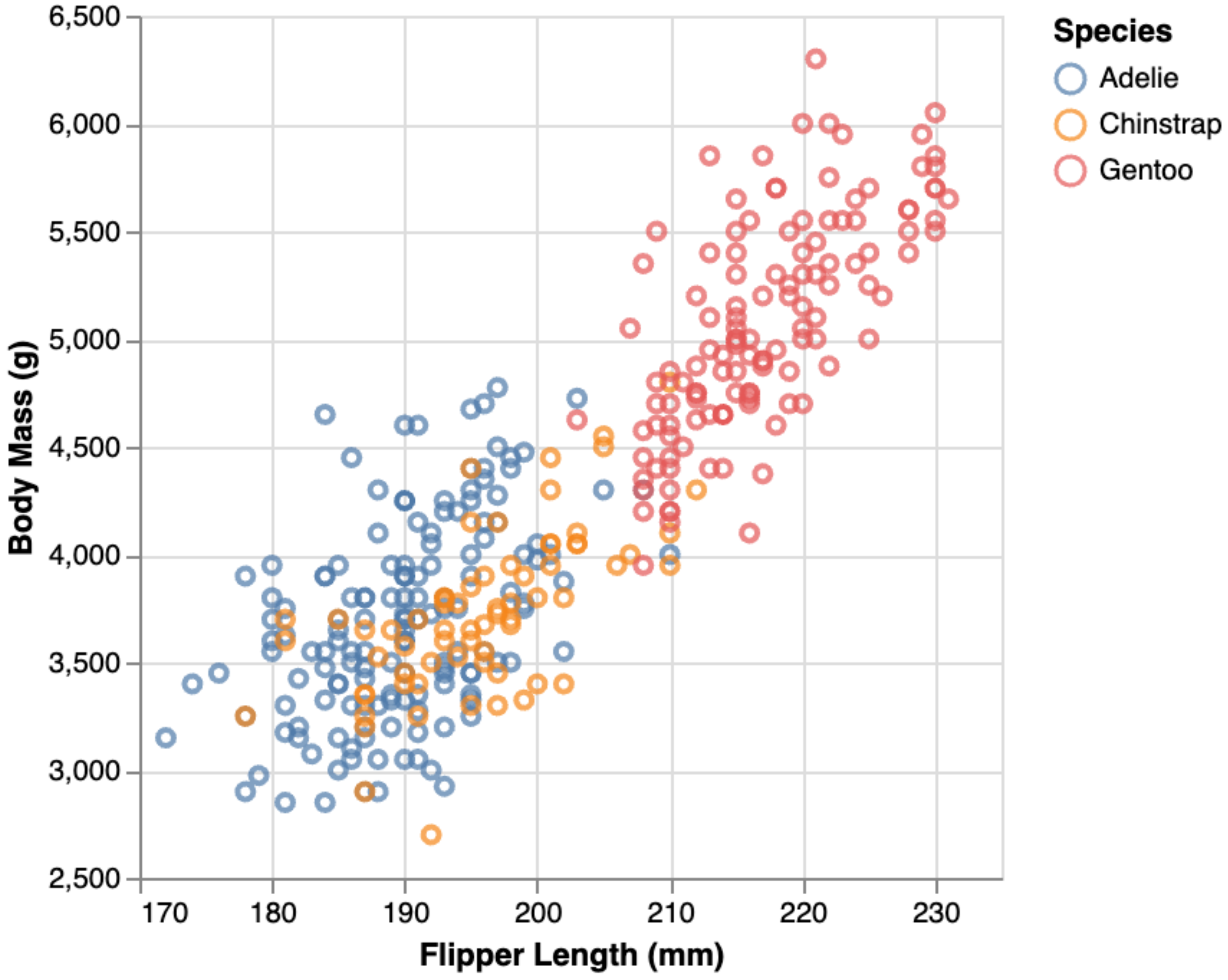
Problems with using it

- Limited discriminability: Only ~6-10 easily distinguishable shapes
- Size interactions: Shape perception changes with size
- Cultural interpretation: Shapes may have unintended meanings

Scatter plot




Species
Adelie
Chinstrap
Gentoo
Body Mass (g)
Flipper Length (mm)
6,500
6,000
5,500
5,000
4,500
4,000
3,500
3,000
2,500
170
180
190
200
210
220
230


Species
Adelie
Chinstrap
Gentoo
Body Mass (g)
Flipper Length (mm)
6,500
6,000
5,500
5,000
4,500
4,000
3,500
3,000
2,500
170
180
190
200
210
220
230

Channel	Accuracy	Discriminability	Pop-out	Separability	Grouping	Best Data Types	Notes
Position	Highest	Very high	No	Separable	Strong spatial	Quantitative, Ordinal	Most effective channel for precise comparison
Length	High	~20	No	Separable	Moderate	Quantitative, Ordinal	Second most accurate for quantitative data
Area	Medium	~10	Weak	Separable	Moderate	Quantitative	Prone to systematic perceptual bias
Color Hue	Low (quantity) / High (identity)	~10	Strong	Integral with lightness	Excellent categorical	Nominal	Perfect for categories, poor for quantities
Color Lightness	Medium-Low	~4-6	No	Integral with hue	Strong ordinal	Ordinal, Quantitative	Natural ordering, limited precision
Shape	N/A (categorical)	~<10	Moderate	Separable	Excellent categorical	Nominal	Limited by shape recognition

Table Talk

Deconstruct the viz and have a conversation about the following

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

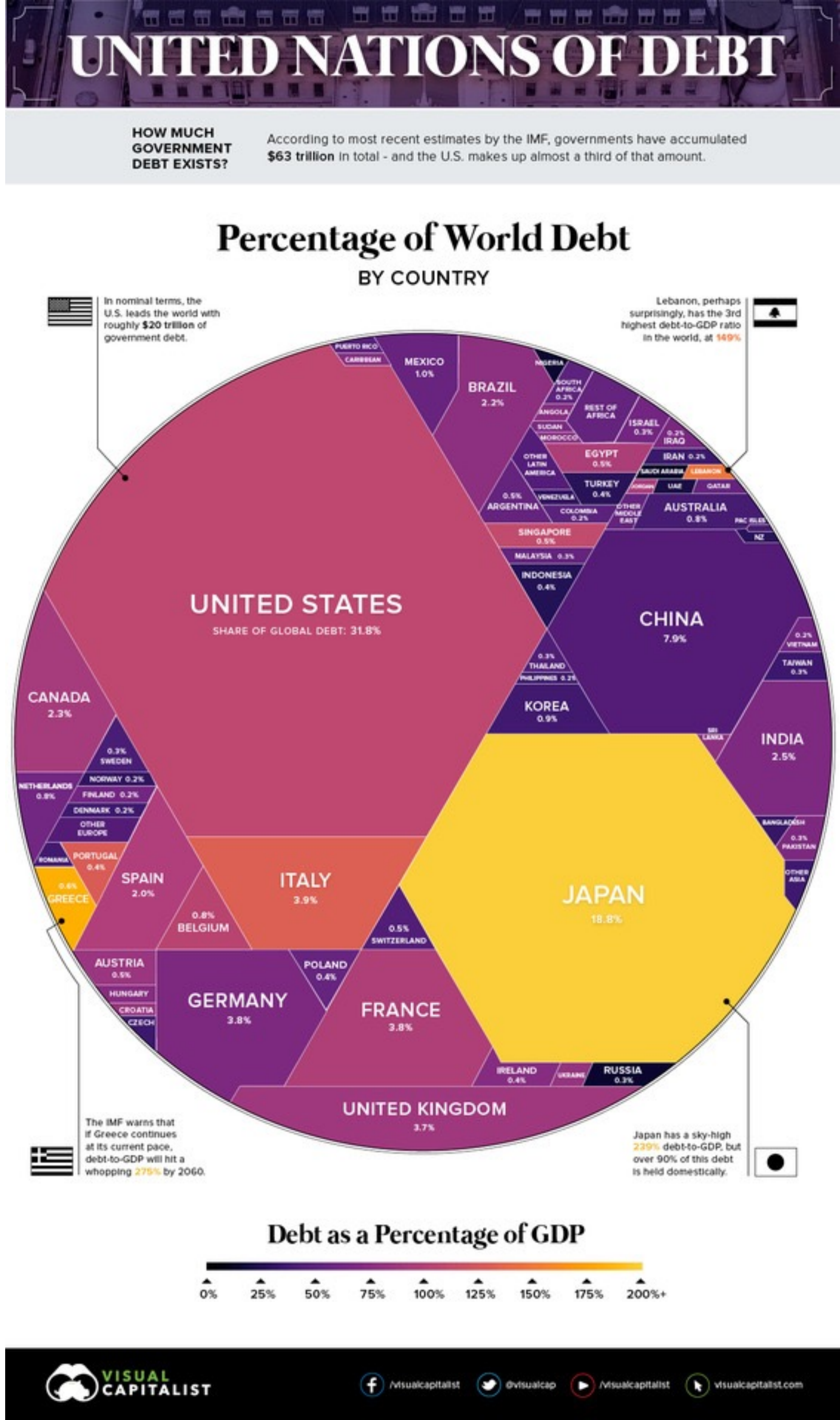
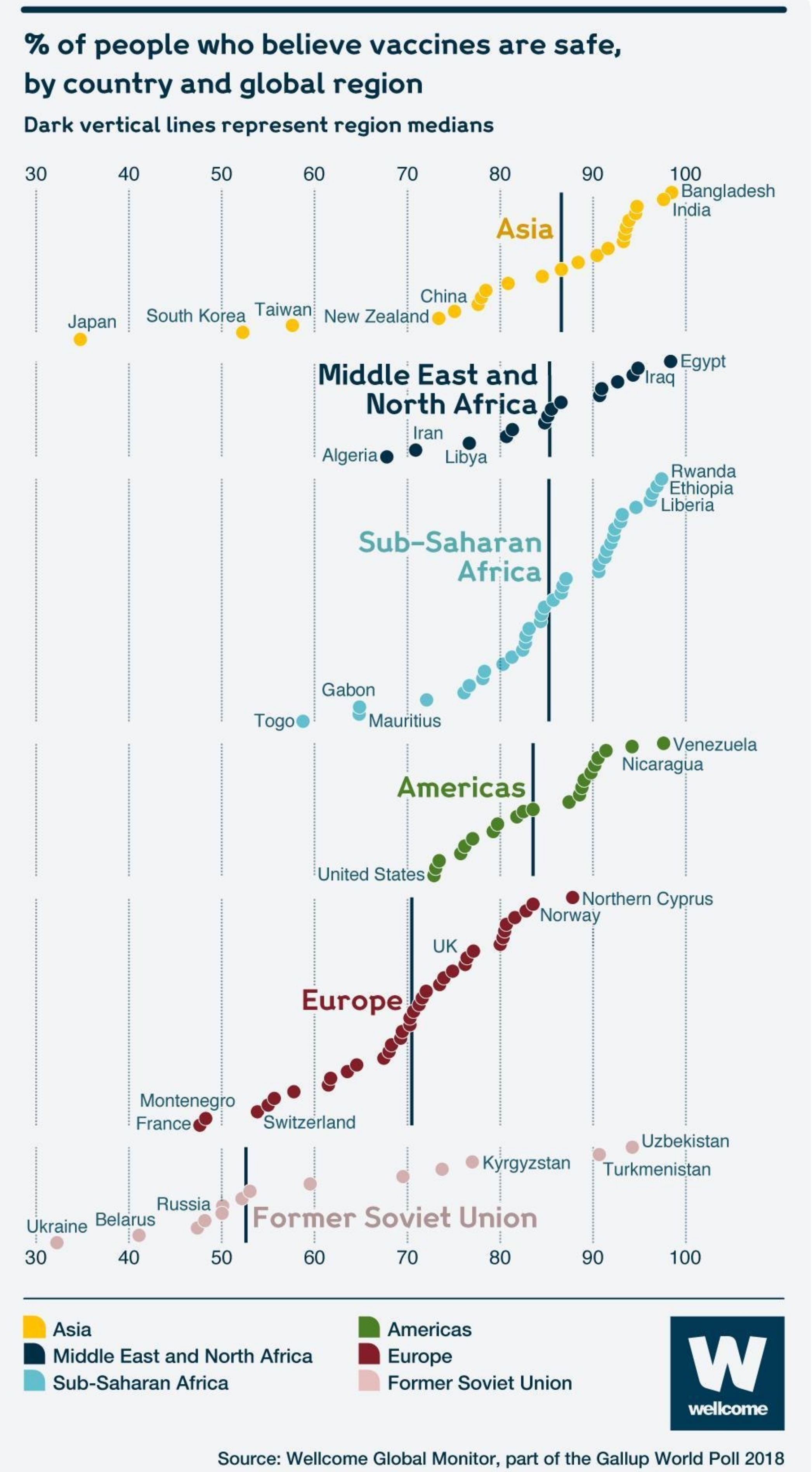


Table Talk

Deconstruct the viz and have a conversation about the following

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



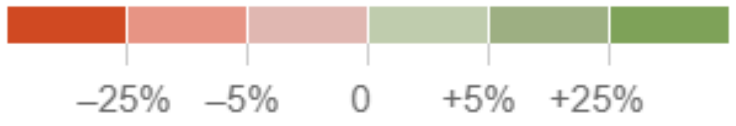
All Spending | Types of Spending | Changes | Department Totals

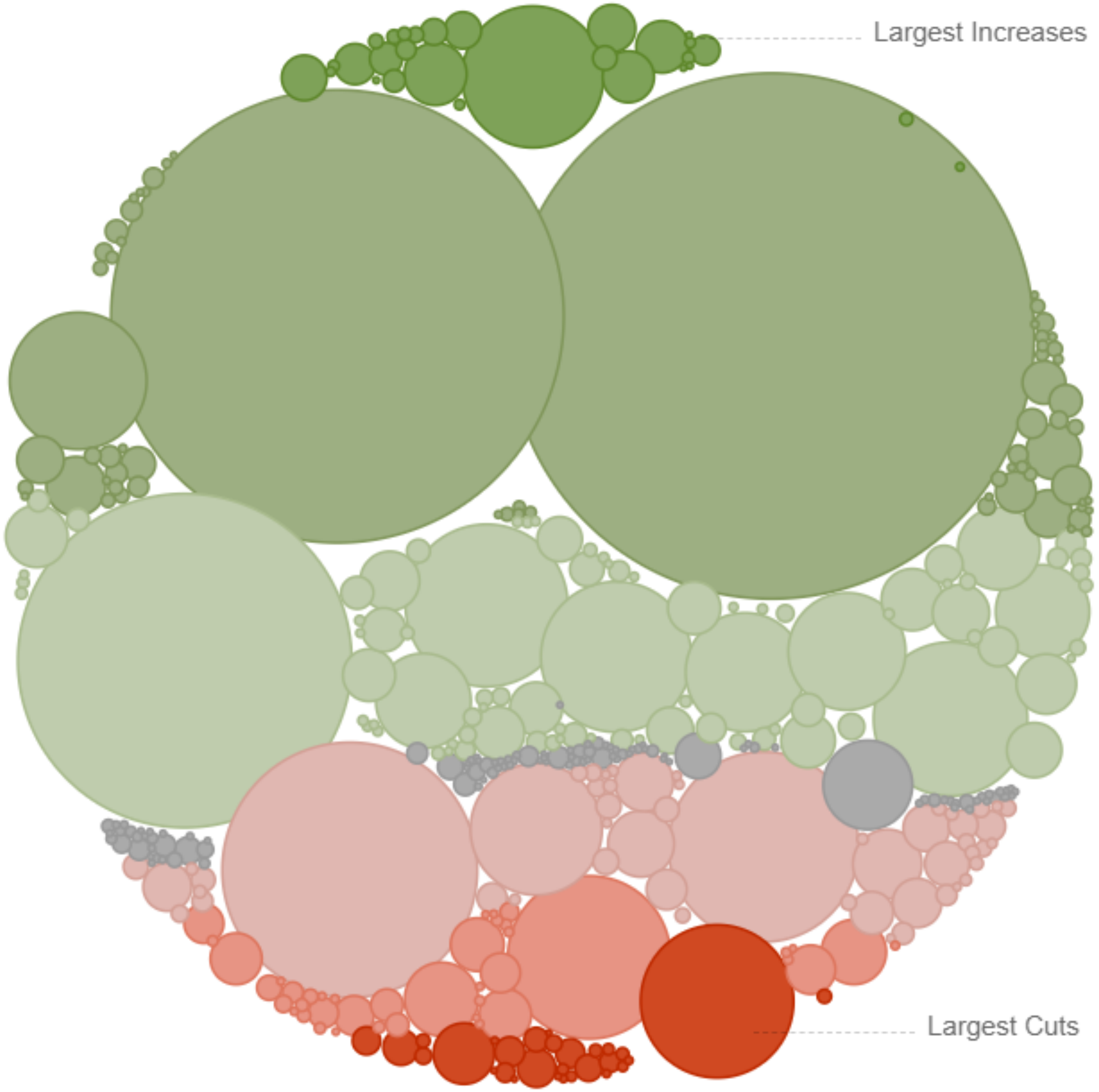
How $3.7 Trillion Is Spent

Mr. Obama's budget proposal includes $3.7 trillion in spending in 2013, and forecasts a $901 billion deficit.

Circles are sized according to the proposed spending.

Color shows amount of cut or increase from 2012.



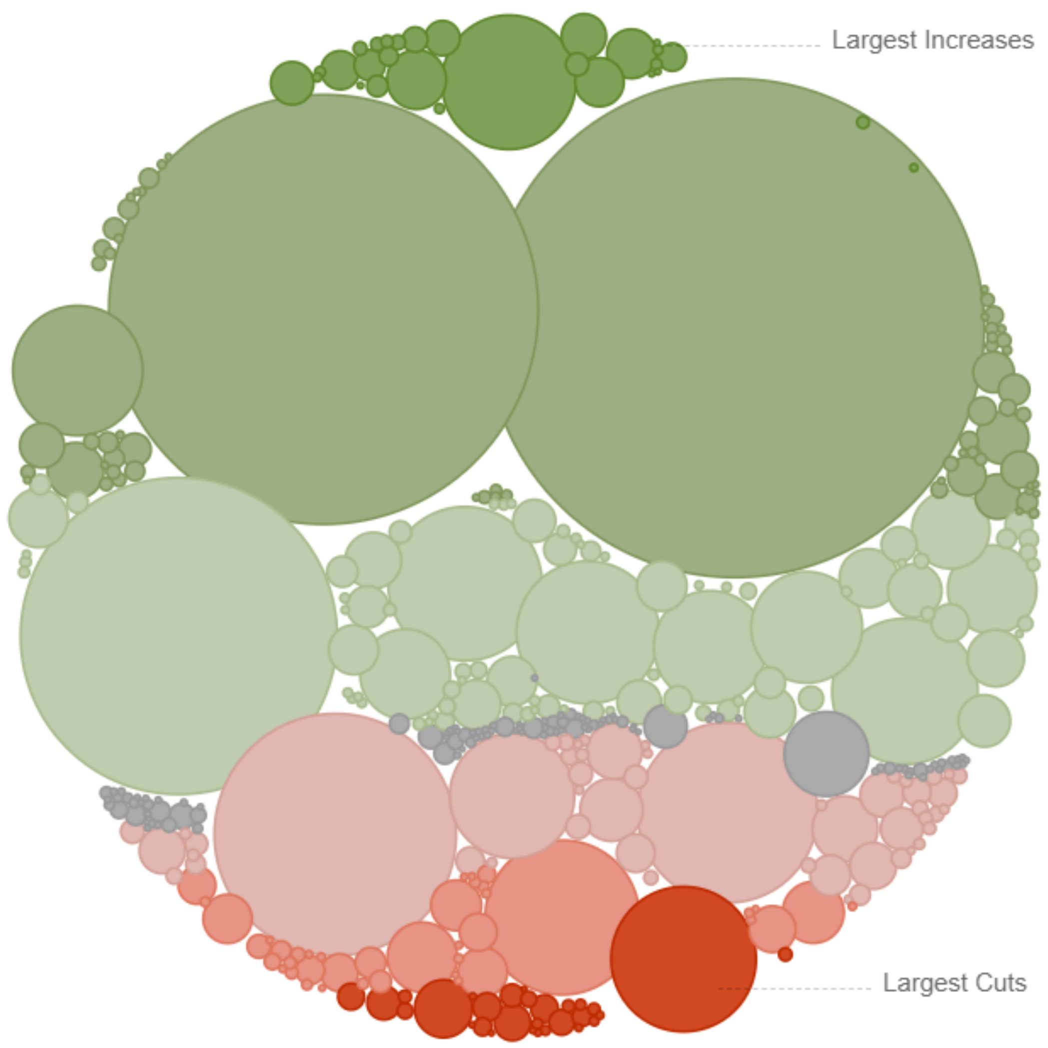


https://archive.nytimes.com/www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html

Table Talk

Deconstruct the viz and have a conversation about the following

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



Learning Outcomes

- **Analyze and critique** visualizations by evaluating their encoding channels using the five core properties (accuracy, discriminability, pop-out, separability, grouping)
- **Identify and explain** how each encoding channel is implemented across different visualization types and contexts
- **Select appropriate encoding channels** for specific data attributes by applying theoretical principles to match channel characteristics with data types

Questions We Still Need to Answer

We’ve only scratched the surface — there’s more to explore in the coming weeks:

- ~~Which channels are best suited for which types of data attributes?~~
- How can we combine multiple channels effectively without causing conflicts?
- Which channels are most effective for specific analytical tasks?

Get Stepping

- Work through Tutorial 4 this week, it builds on last week and you will be tested on it. There is no in-class session for it. Attend office hours if stuck.
- Work through Tutorial 5 before next week Wednesday’s class, we start our dive in EDA vizzes.
- Take a walk for it is starting to smell rain.