Data Visualization & Design

Week 12

1. **Review:** Key Concepts
2. **Studio:** Hierarchical Circle Packing in D3.js

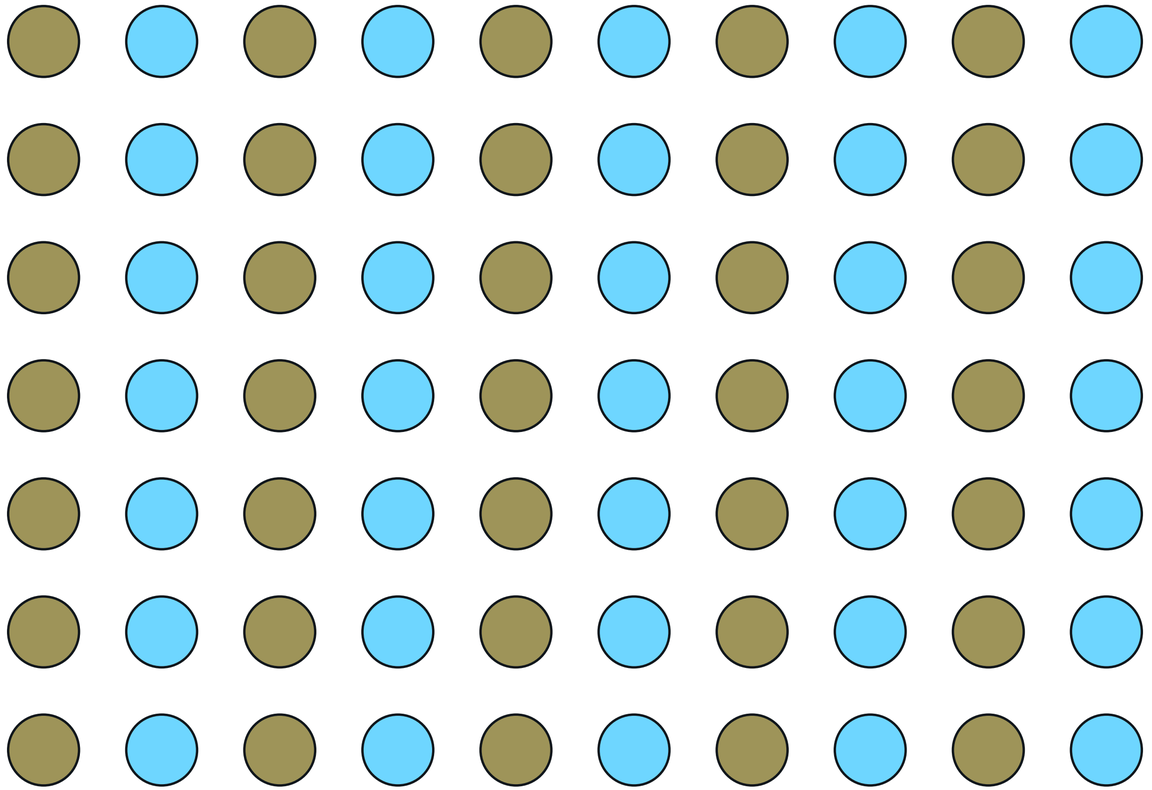
1. **Review:** Key Concepts
2. **Studio:** Hierarchical Circle Packing in D3.js

- **Gestalt Principles**
- **Marks and Channels**
- **Color**
  - Hue, Saturation, and Lightness
  - Practical Tips
- **Charts and Data Types**

3 Gestalt Principles

Visualization designers can **leverage perceptual tendencies** to better express meaning.

Gestalt Principle 1: **Similarity**

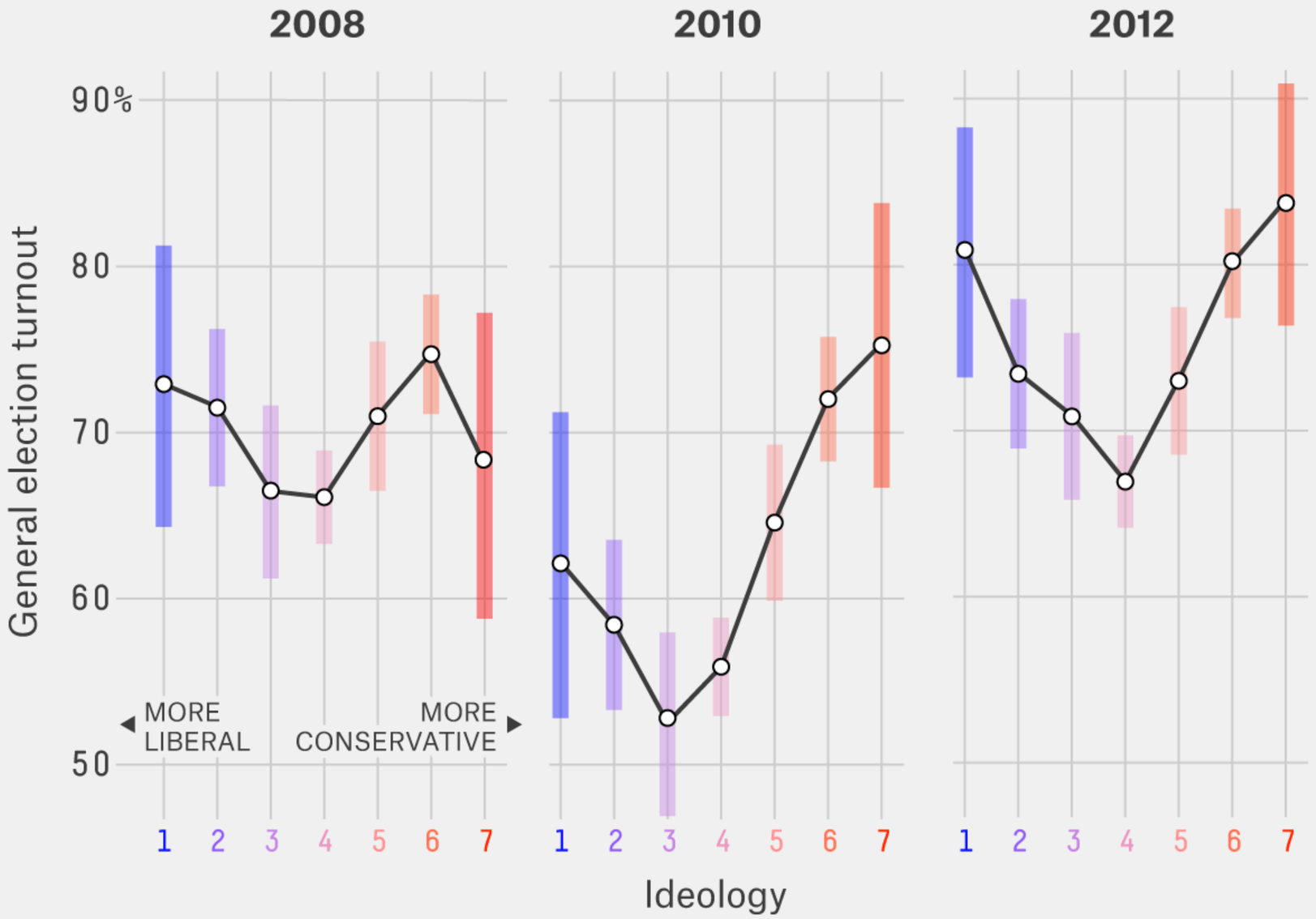


Gestalt Principle 1: **Similarity**

- The most intuitive (+ principle by which **color-coding** works)
- Graphical elements with shared visual properties are perceived as belonging to the same group
- In the previous image, we detect two classes of objects, denoted by **gold** and **blue**

Conservatives (and liberals) vote more

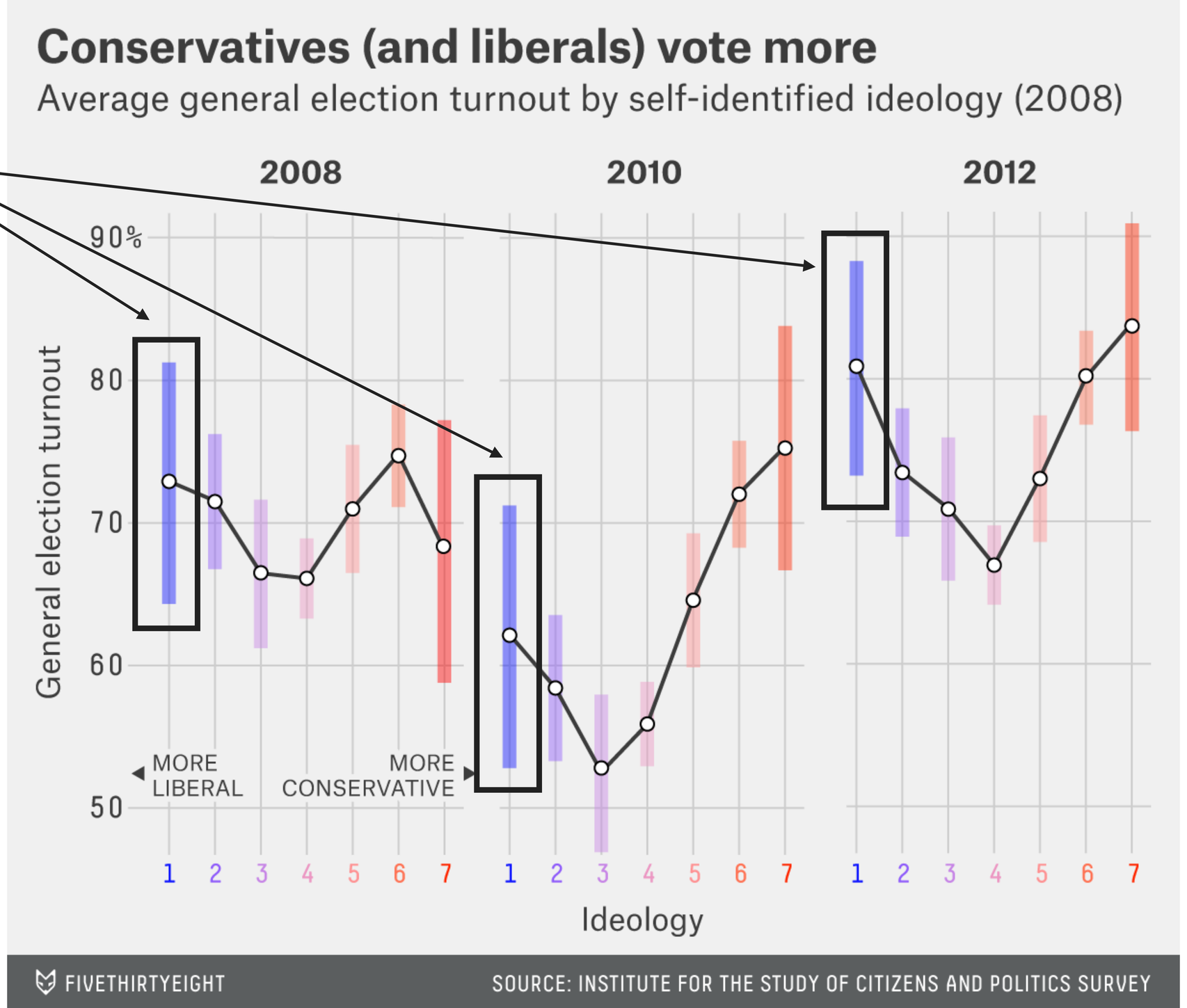
Average general election turnout by self-identified ideology (2008)



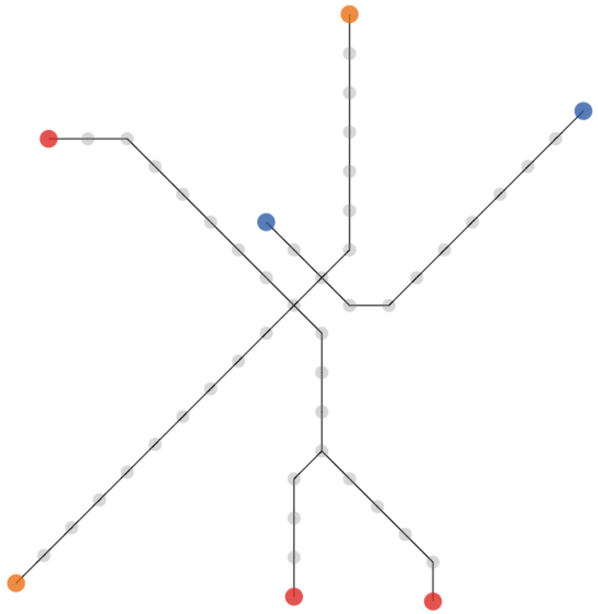
FIVETHIRTYEIGHT

SOURCE: INSTITUTE FOR THE STUDY OF CITIZENS AND POLITICS SURVEY

Related!


Conservatives (and liberals) vote more
Average general election turnout by self-identified ideology (2008)
2008
2010
2012
90%
80
70
60
50
General election turnout
MORE LIBERAL
MORE CONSERVATIVE
1 2 3 4 5 6 7
1 2 3 4 5 6 7
1 2 3 4 5 6 7
Ideology
FIVETHIRTYEIGHT
SOURCE: INSTITUTE FOR THE STUDY OF CITIZENS AND POLITICS SURVEY

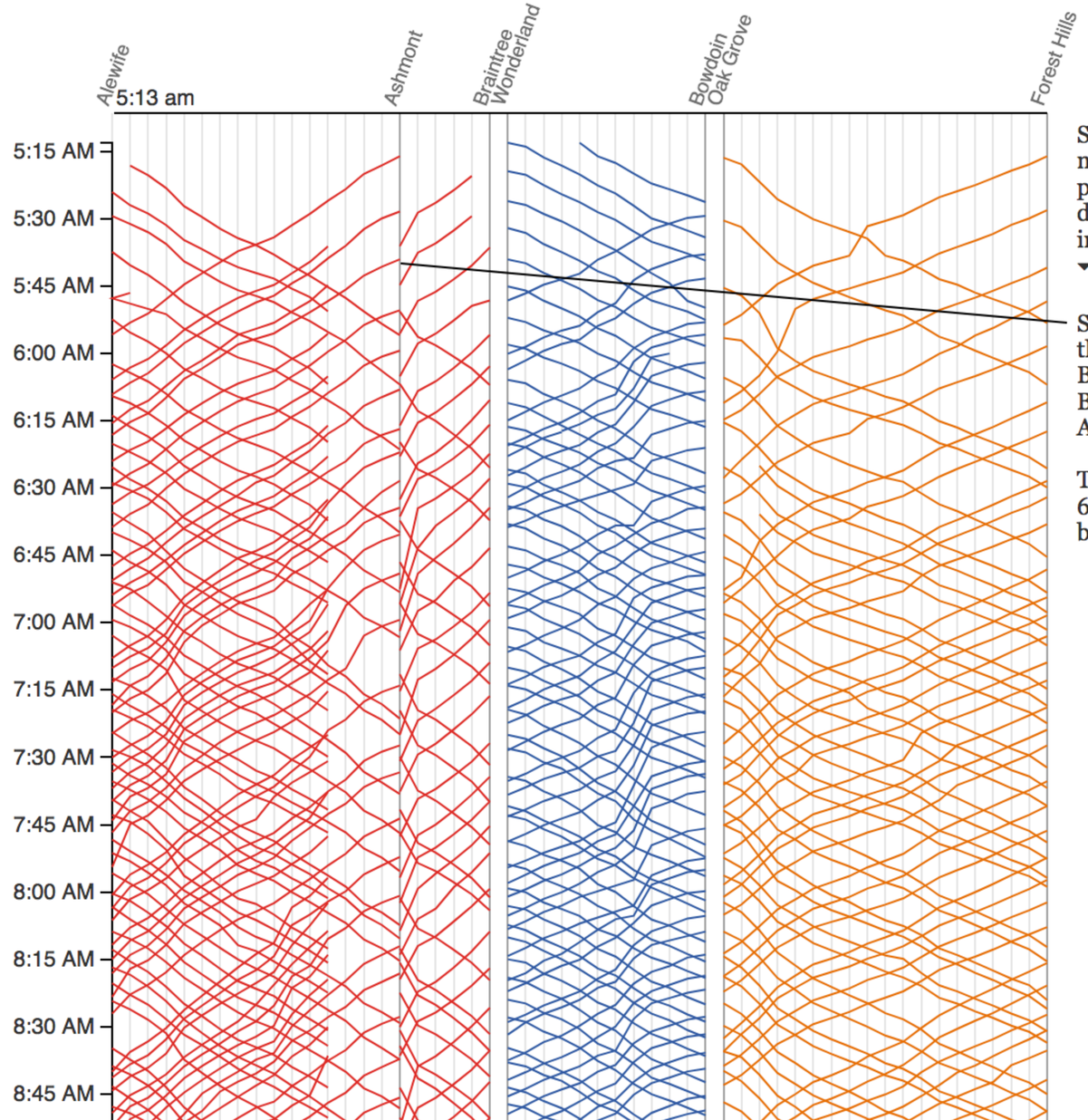
Subway Trips on Monday February 3, 2014



Locations of each train on the red, blue, and orange lines at 5:13 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

See the morning rush-hour, midday lull, afternoon rush-hour, and the evening lull.



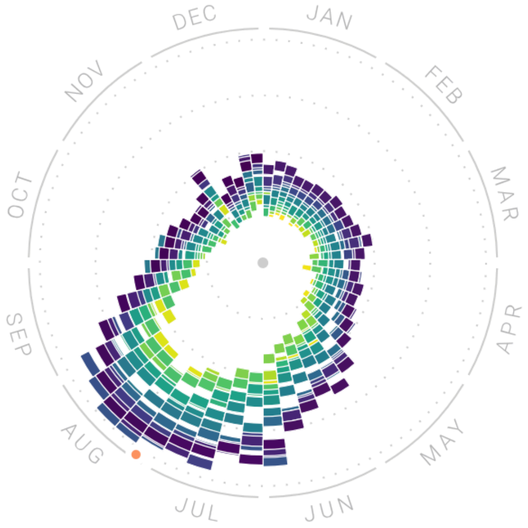
Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

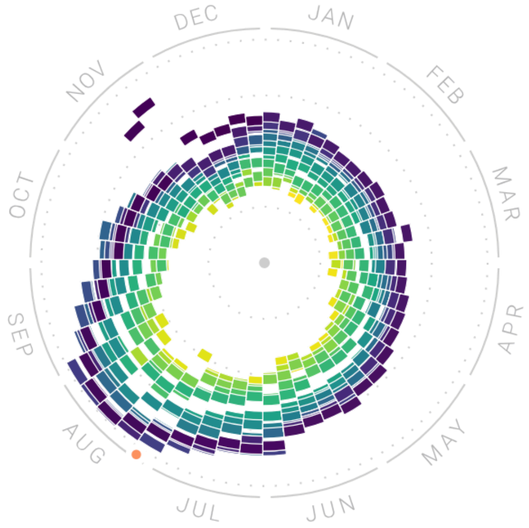
Train frequency increases around 6:30AM as morning rush hour begins.

It's August! What's asked for right now?

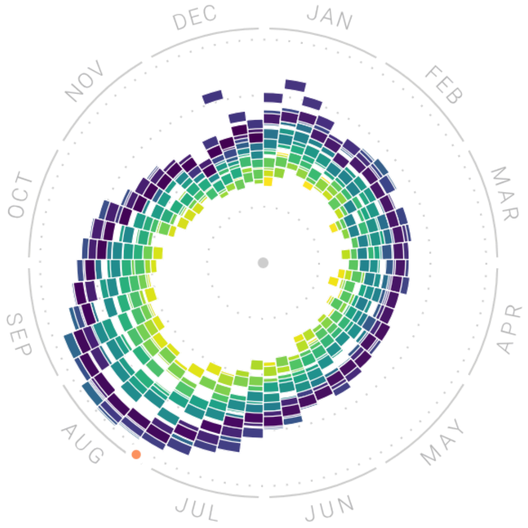
January | February | March | April | May | June | July | **August** | September | October | November | December



PEACH

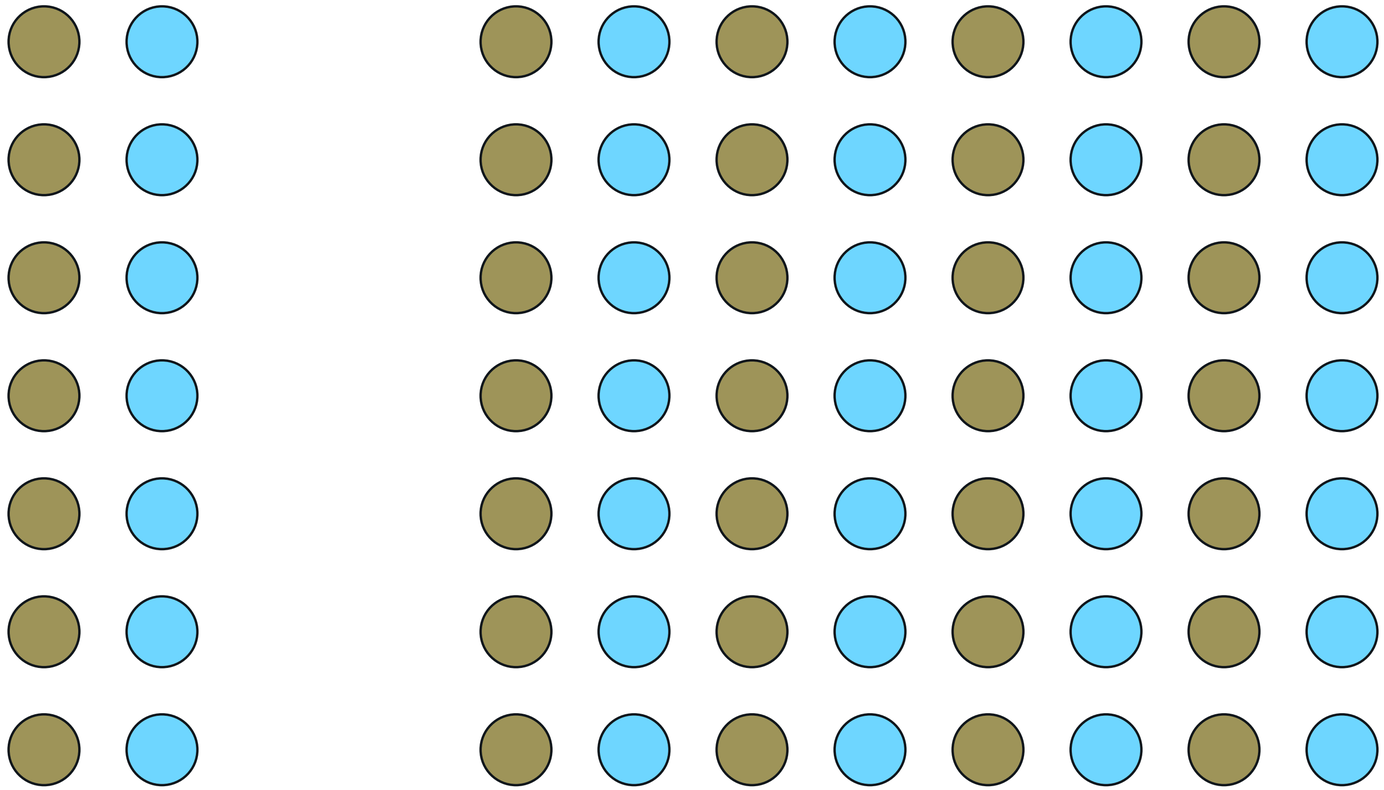


GARDEN TOMATO



EGGPLANT

Gestalt Principle 2: **Proximity**

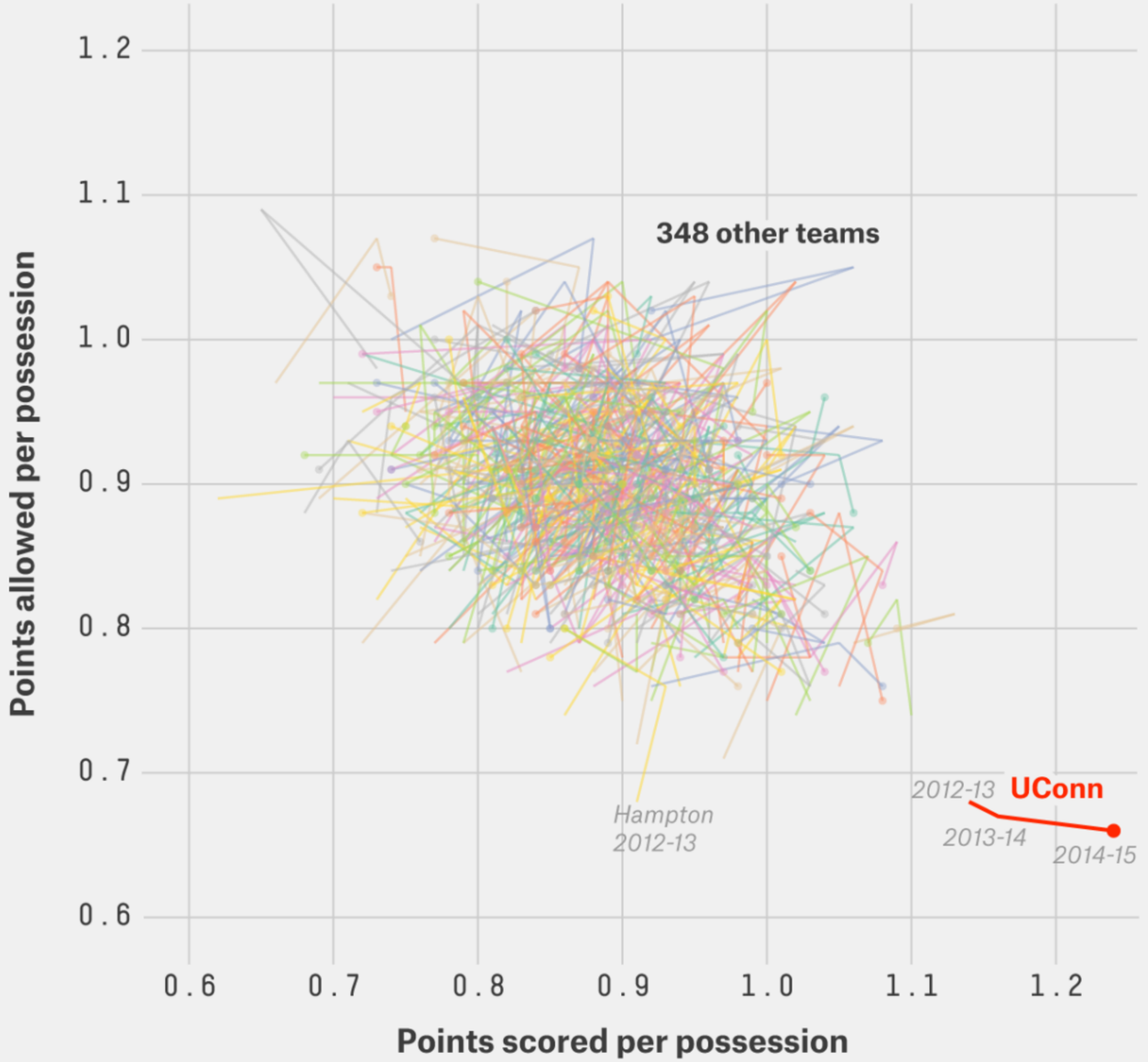


Gestalt Principle 2: **Proximity**

- ***Always* present** in data visualization charts
- A graphical element being close to another graphical element is a strong indicator of similarity (ex. pie charts, bar charts)
- In the previous image, we detect **two groups of objects**, because the two columns of circles on the left are closer to each other than to the eight columns of circles on the right

Huskies Have Separated From The Pack

Change in points scored and allowed per possession for 349 Division I women's teams for the past three seasons

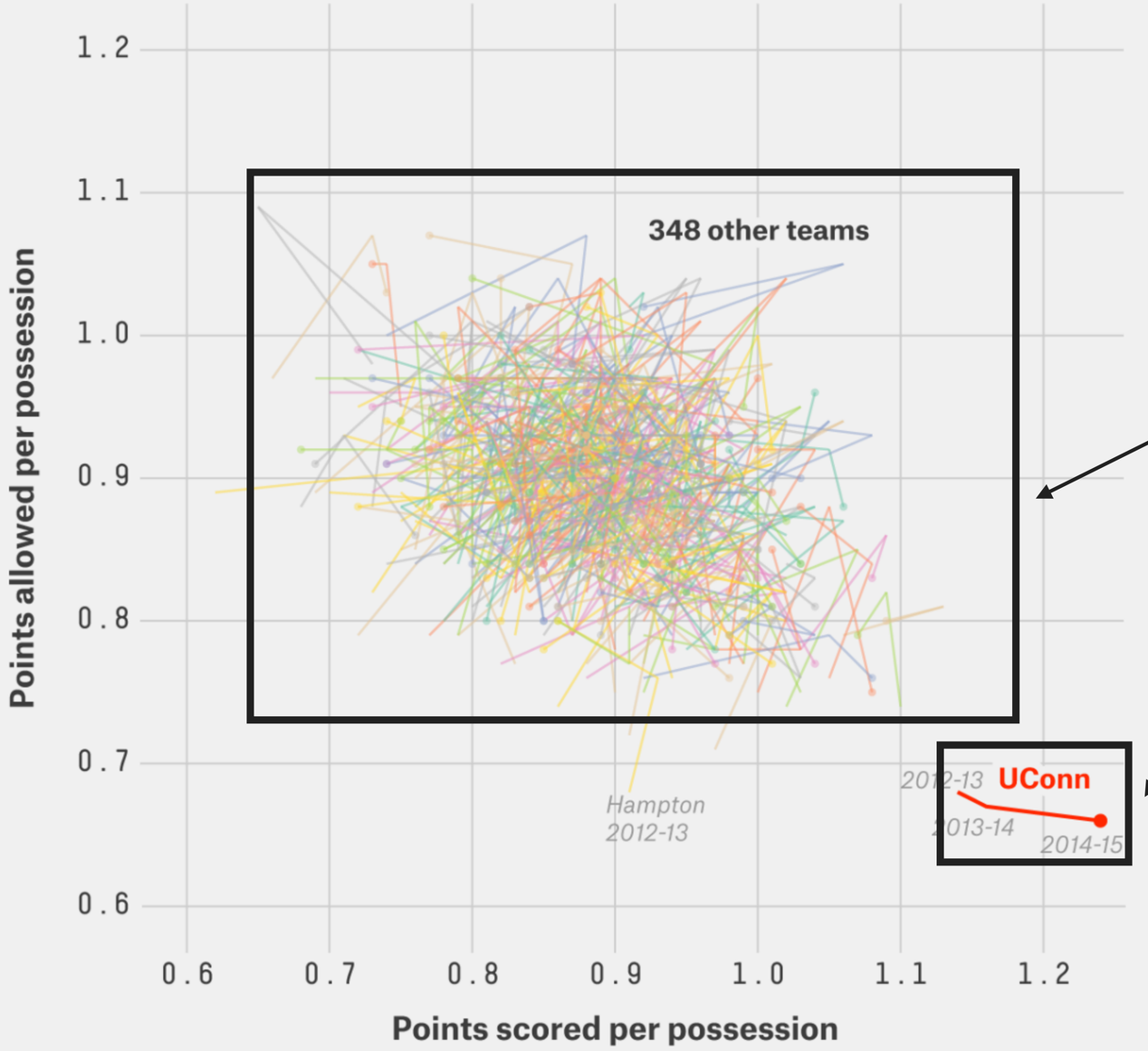


FIVETHIRTYEIGHT

SOURCE: WBBSTATE.COM

Huskies Have Separated From The Pack

Change in points scored and allowed per possession for 349 Division I women's teams for the past three seasons



FIVETHIRTYEIGHT SOURCE: WBBSTATE.COM

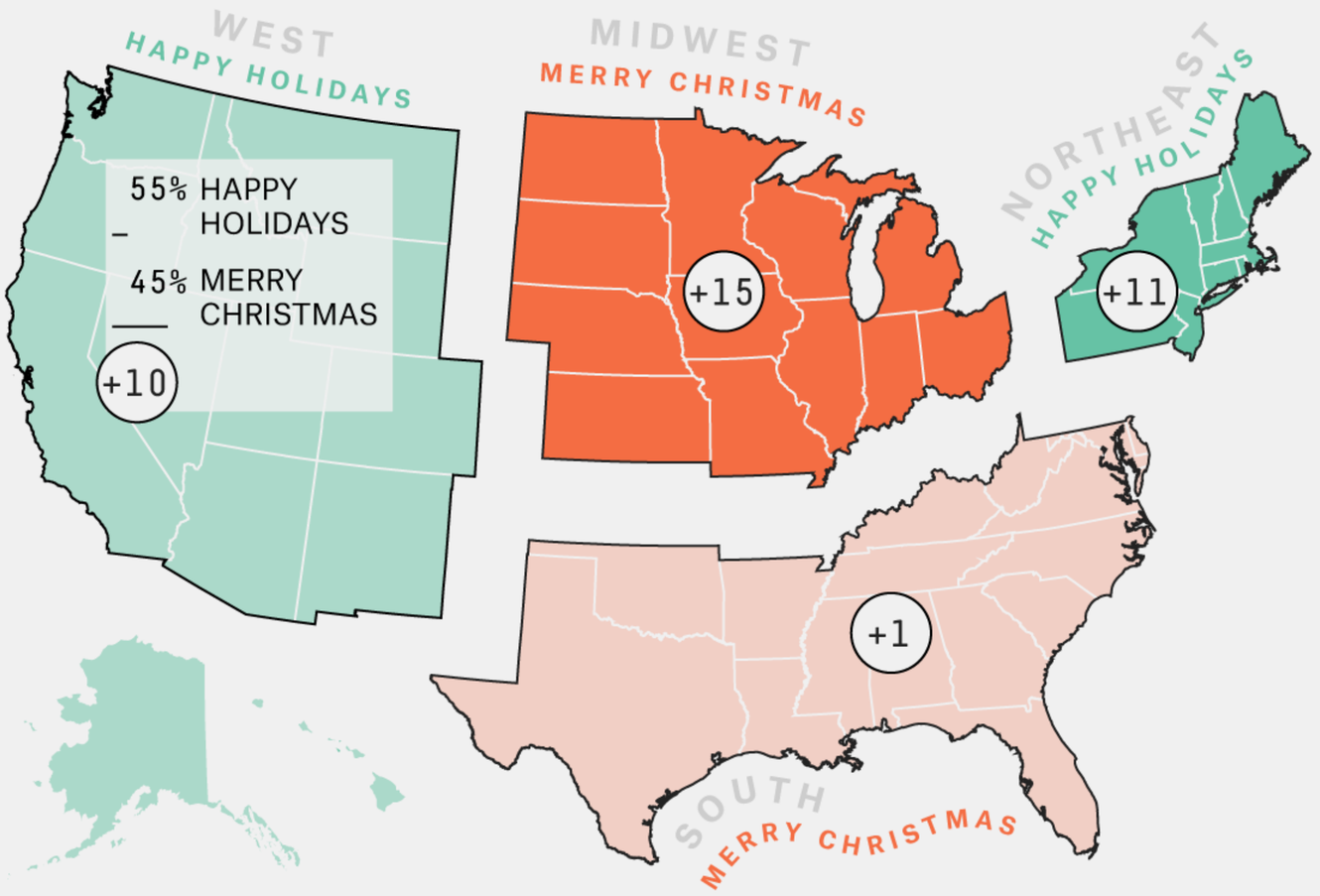
Distinct

Mapping the 'War on Christmas'

Difference between percentage of people who favor "merry Christmas" vs. "happy holidays" by Census region

SURVEY DATES
12/7 - 12/11/2016

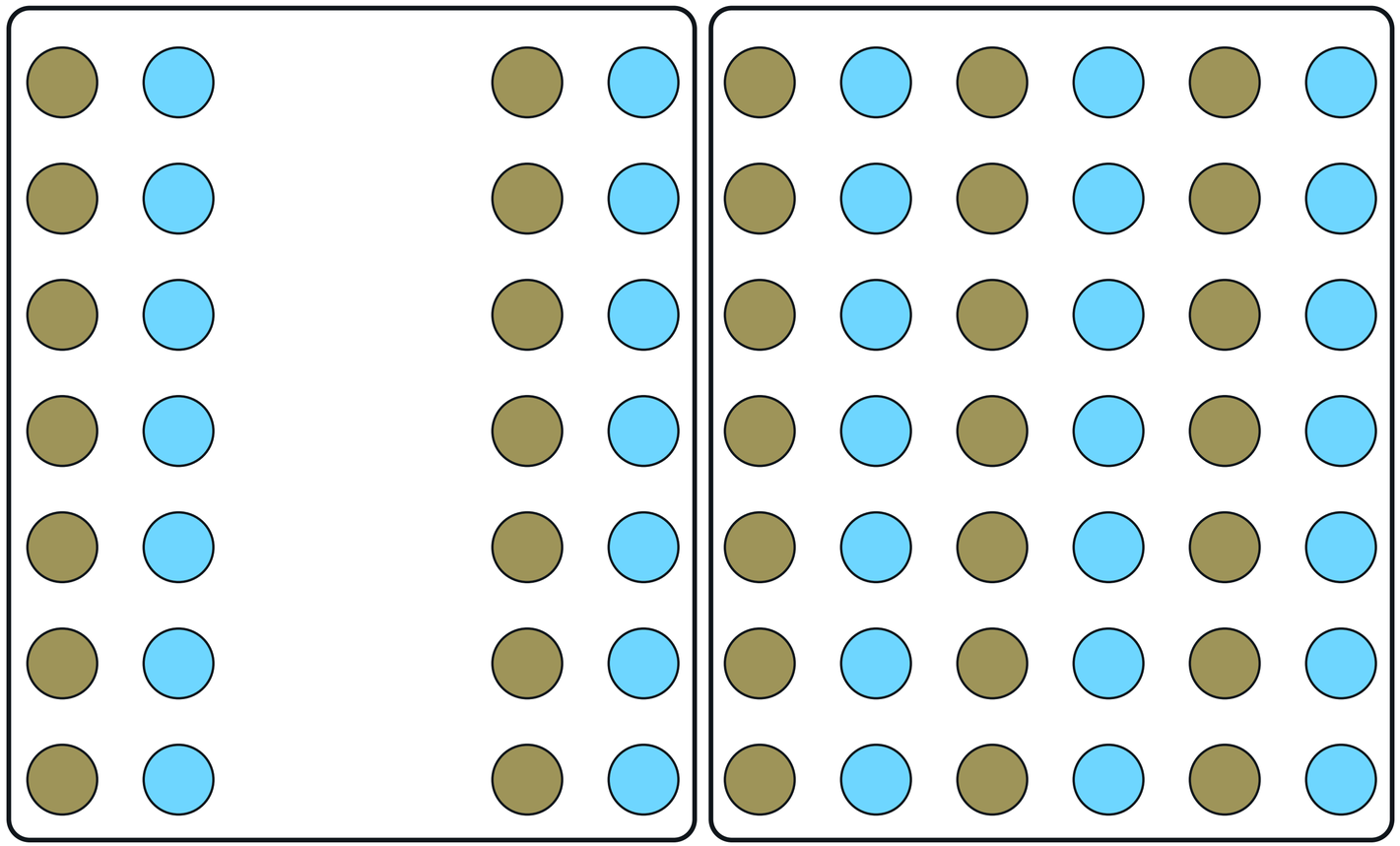
NO. OF RESPONDENTS
1,004



FiveThirtyEight

SOURCE: PUBLIC RELIGION RESEARCH INSTITUTE

Gestalt Principle 3: **Enclosure**

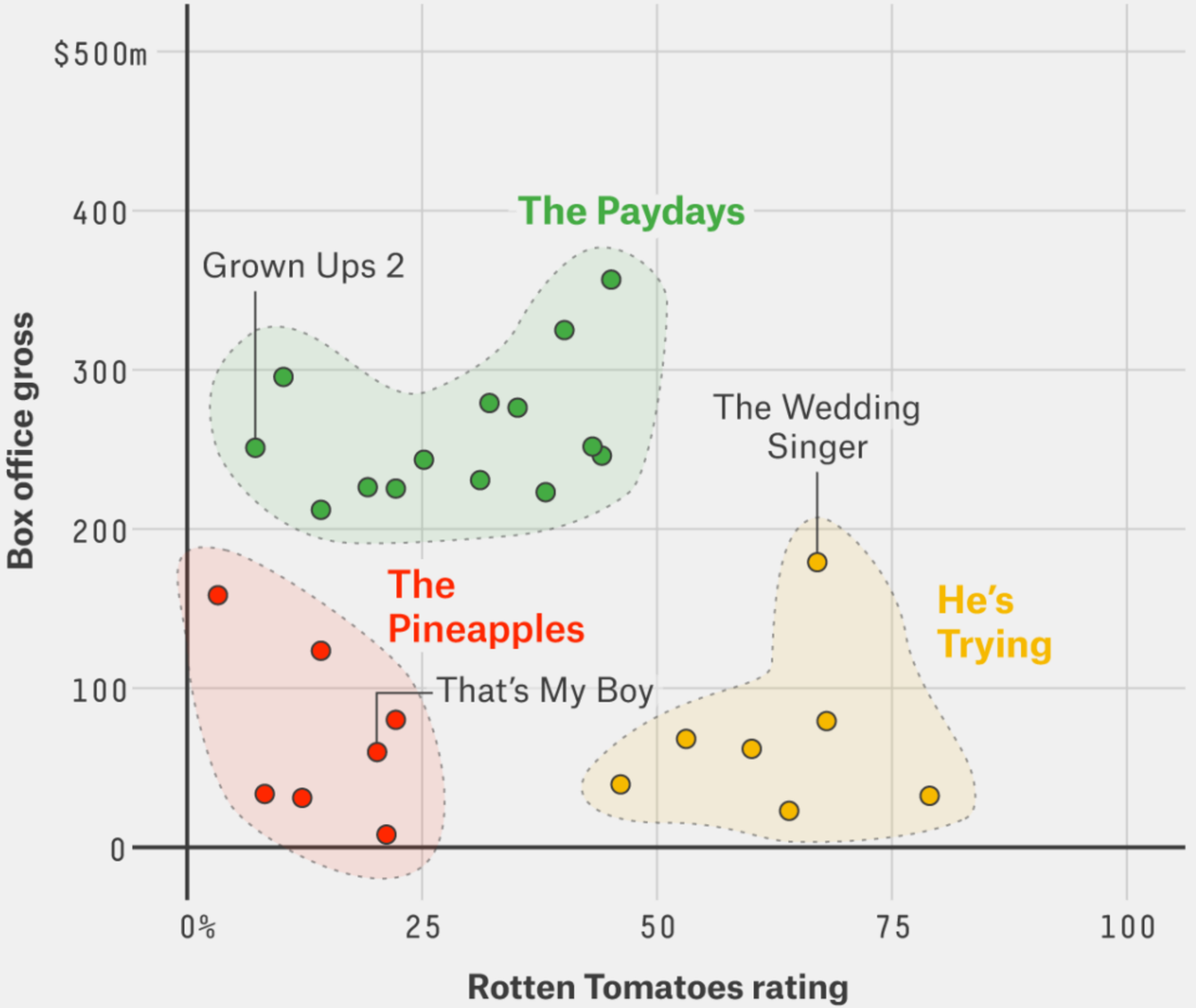


Gestalt Principle 3: **Enclosure**

- Surrounding a group of elements with a visual element
- Uncommon, but very powerful (correlated with annotations)
- In the previous image, we detect **two groups of objects** that override the two groups of objects we discerned through the principle of proximity

The Three Types Of Adam Sandler Movies

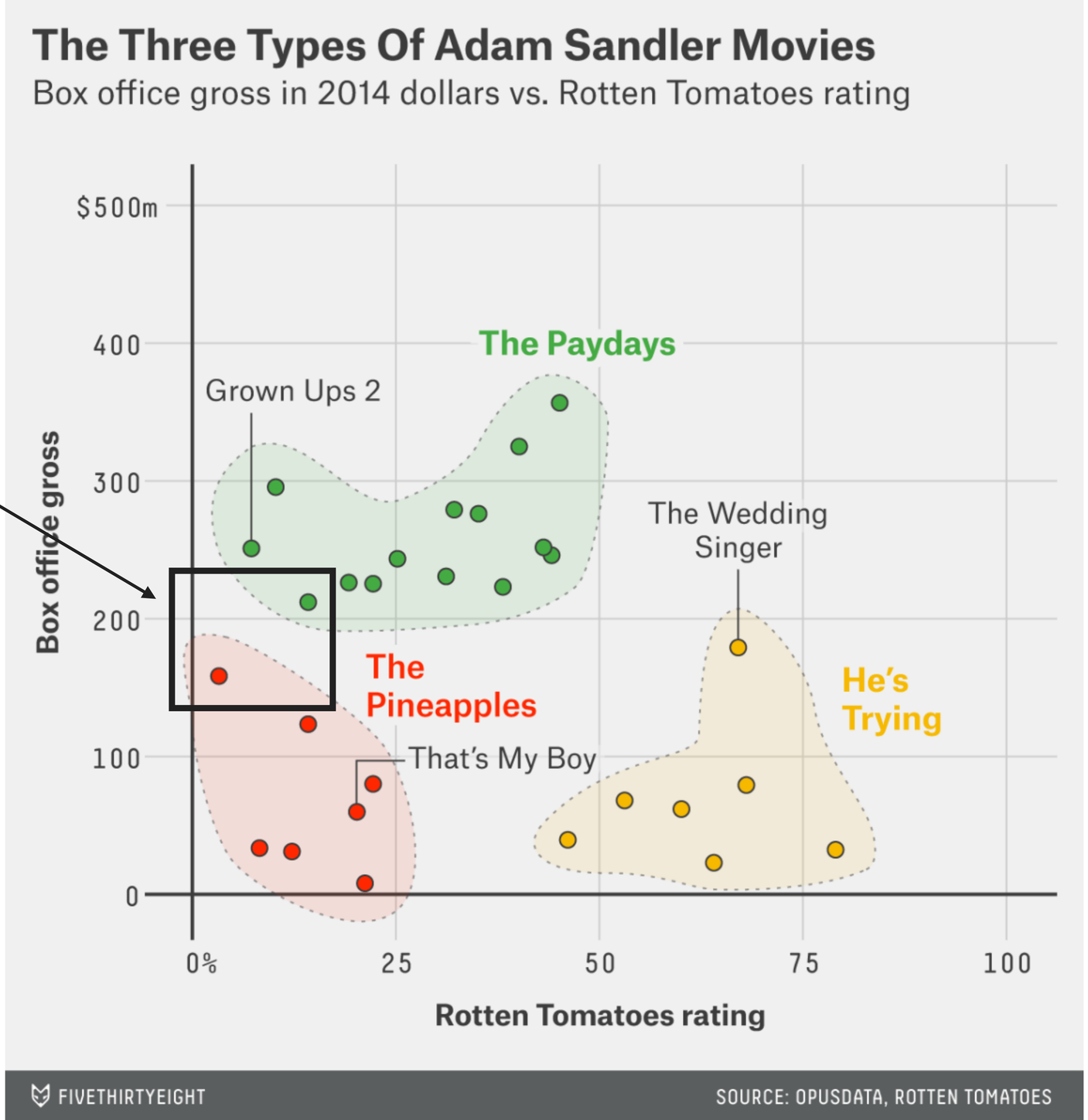
Box office gross in 2014 dollars vs. Rotten Tomatoes rating



FIVETHIRTYEIGHT

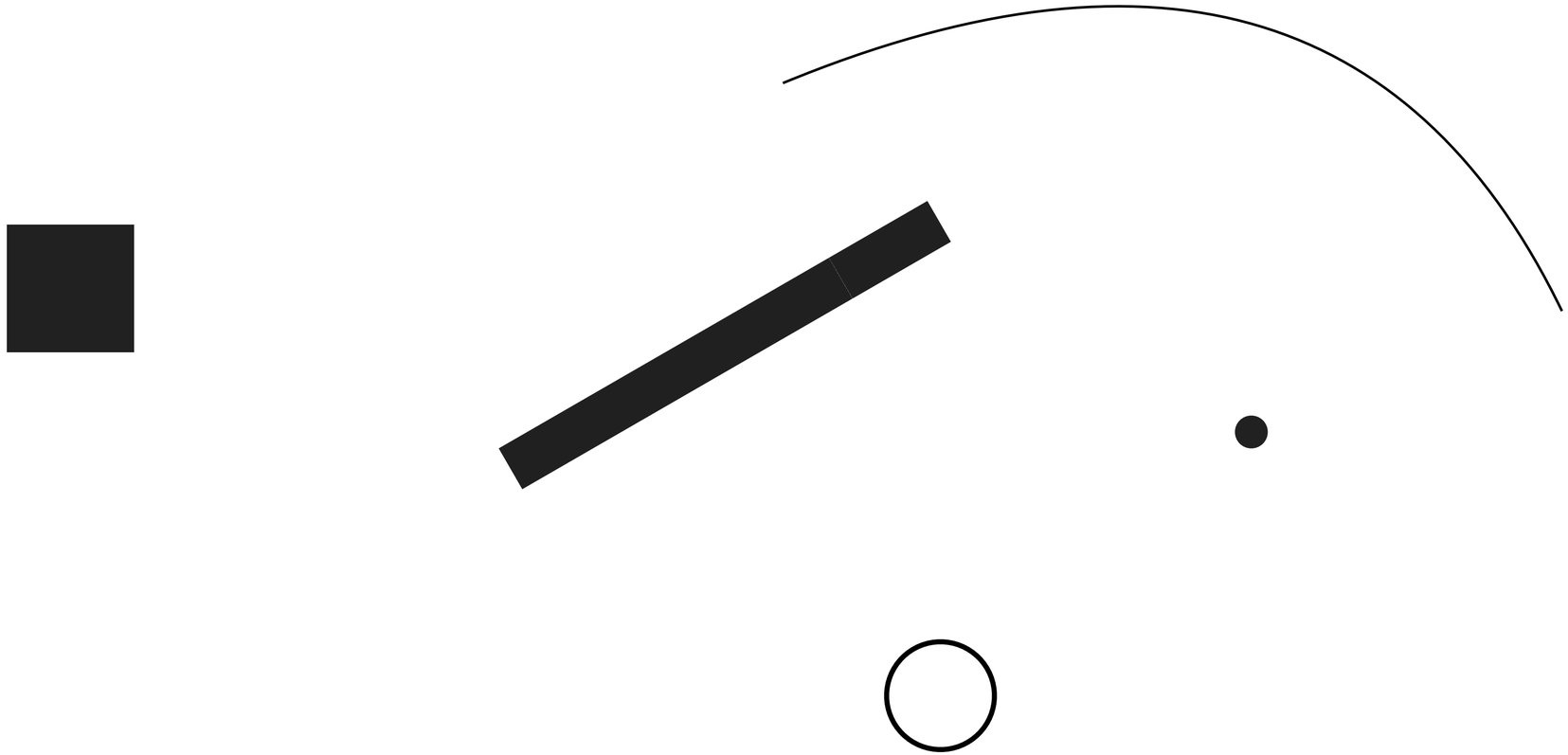
SOURCE: OPUSDATA, ROTTEN TOMATOES

Close, but not grouped

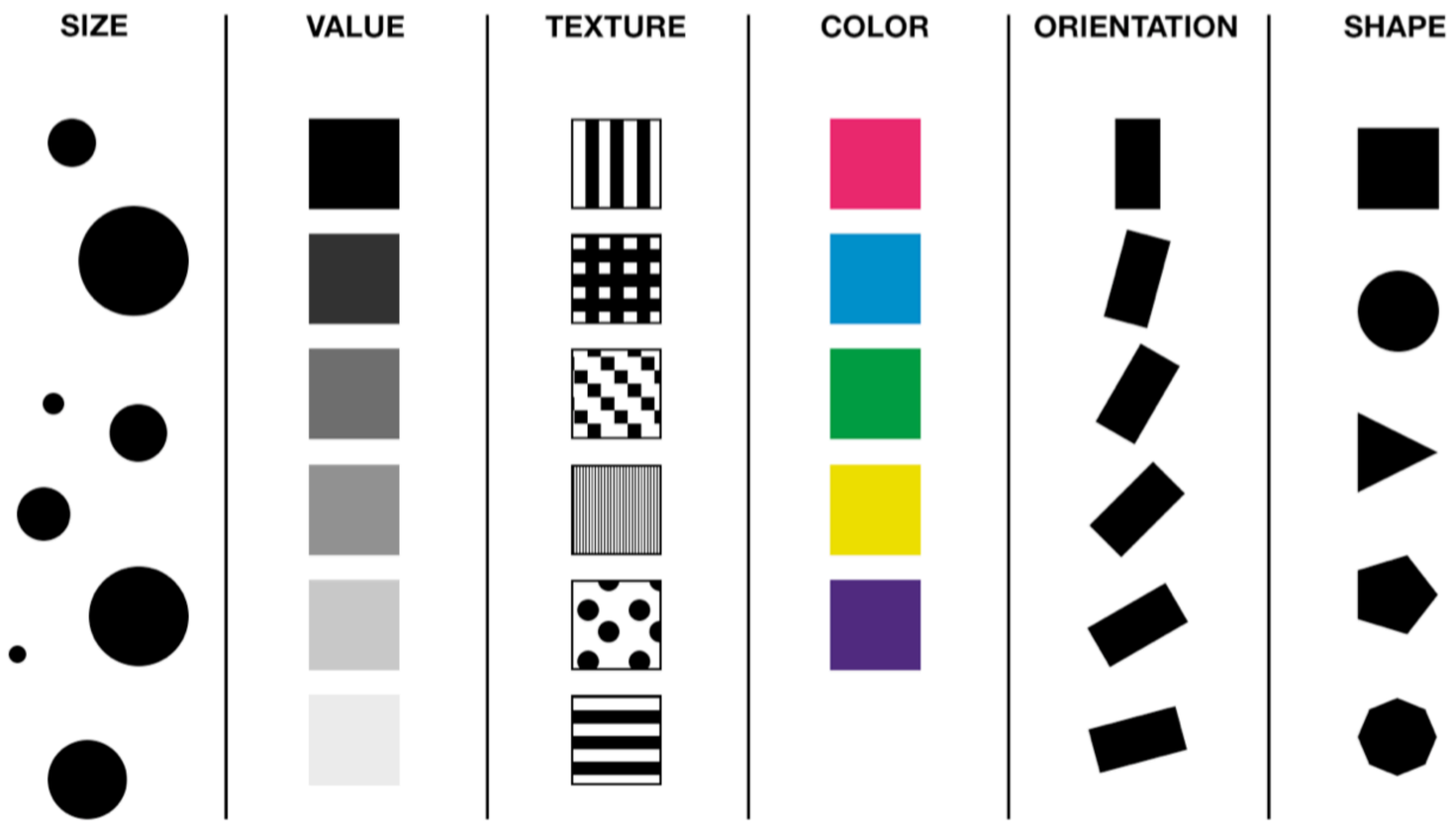


Marks & Channels

Marks



Channels

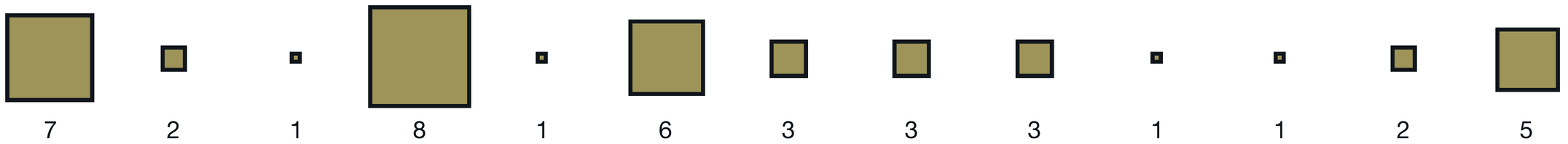


Jacques Bertin defines ***marks*** and ***channels*** as the primary components of a visualization.

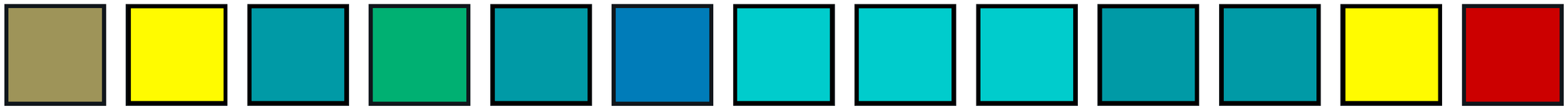
In his terms, each **channel *(= visual variable)*** modifies a **mark**, based on a given attribute.

(Undifferentiated marks)

Channel: **Size**



Channel: **Color**



Channel: **Saturation**

Jock Mackinlay **ranks channels** in terms of effectiveness (i.e. which can be read most quickly).

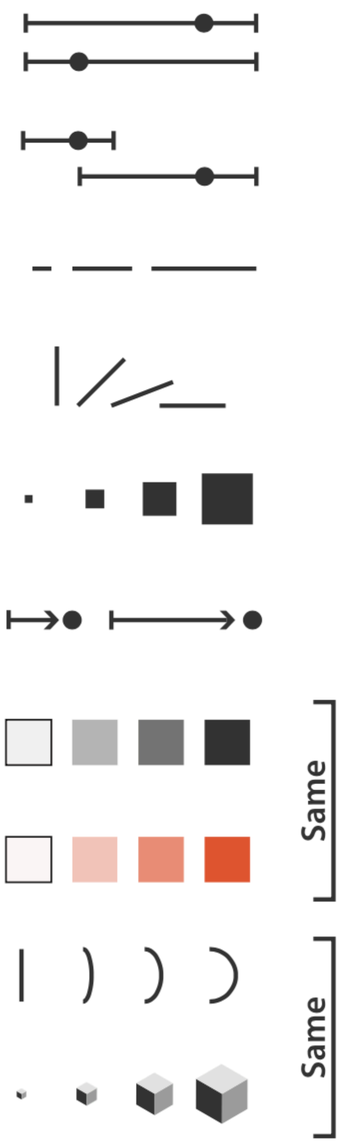
He also delineates the **different data types** that correspond to different channels.

Mackinlay's ***effectiveness principle—***

The most important attributes in a visualization should be encoded with the highest-ranked channels.

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)



Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

Color—

Hue, **Saturation**, & **Lightness**

- **Hue —** The actual color
- Saturation
- Luminance

- Hue
- **Saturation —** The amount of grey in a color
- Luminance

- Hue
- Saturation
- **Luminance —** The amount of white or black in a color

Color—

Practical Tips

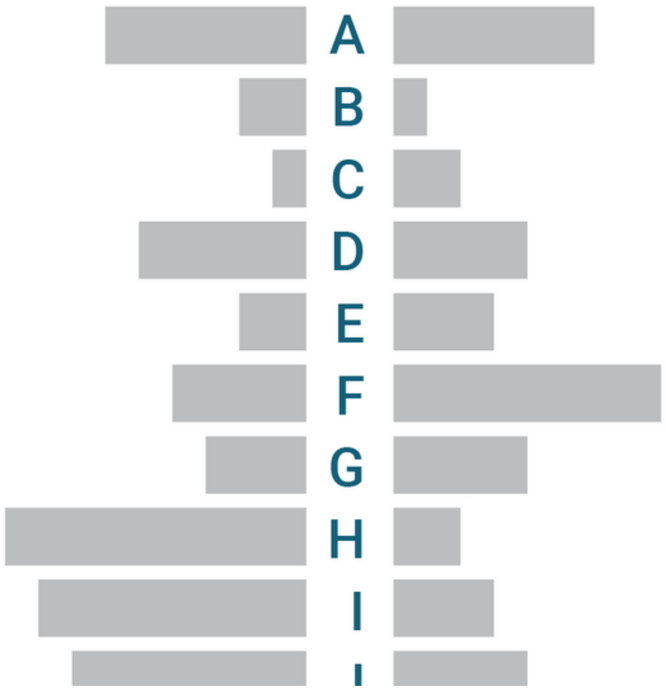
Lisa Charlotte Rost for ***Datawrapper*** put together a handy guide for using color in visualization.

→ *LINK:* **https://blog.datawrapper.de/colors/**

If you need more than **seven colors** in a chart, consider using another chart type or to group categories together.

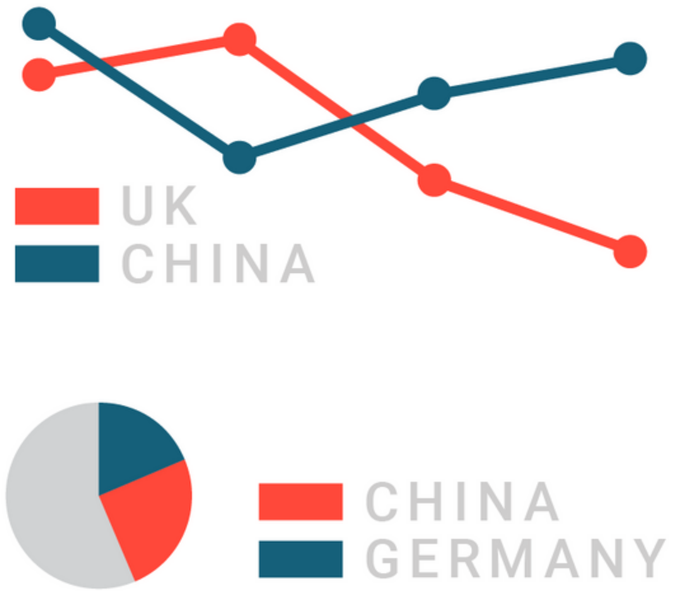
NOT IDEAL

BETTER

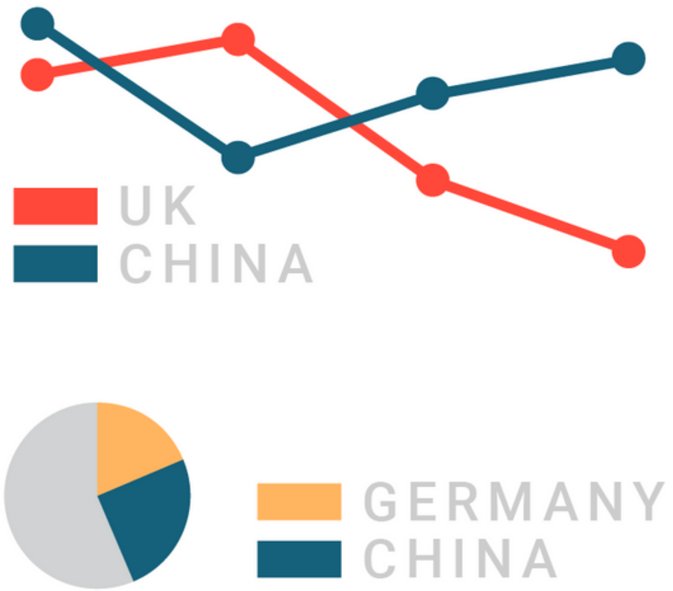


Consider using the **same color** for the **same variables**.

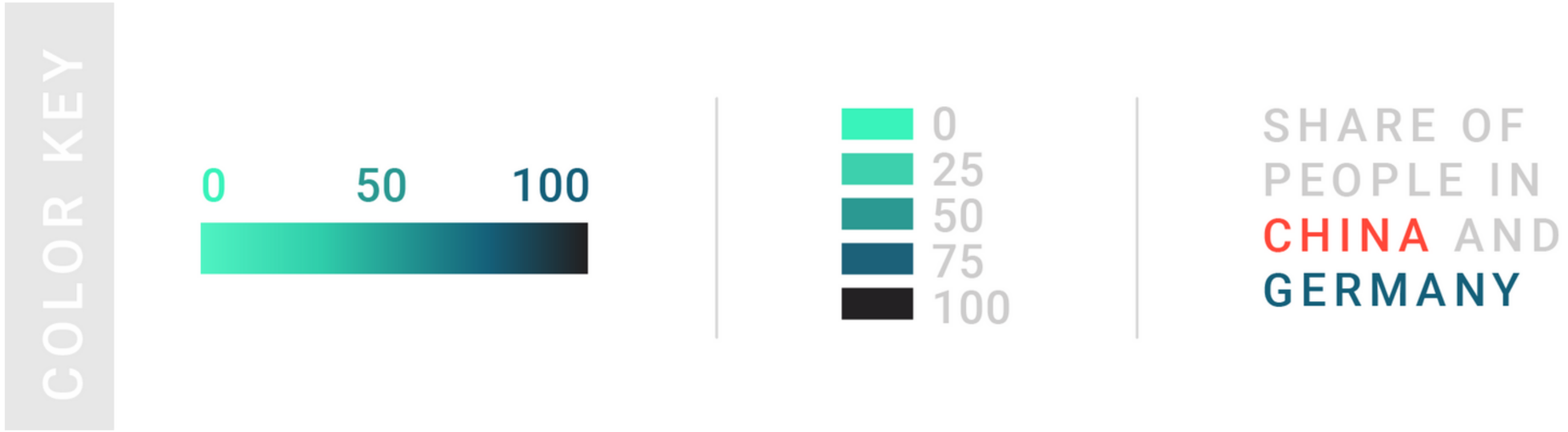
NOT IDEAL



BETTER

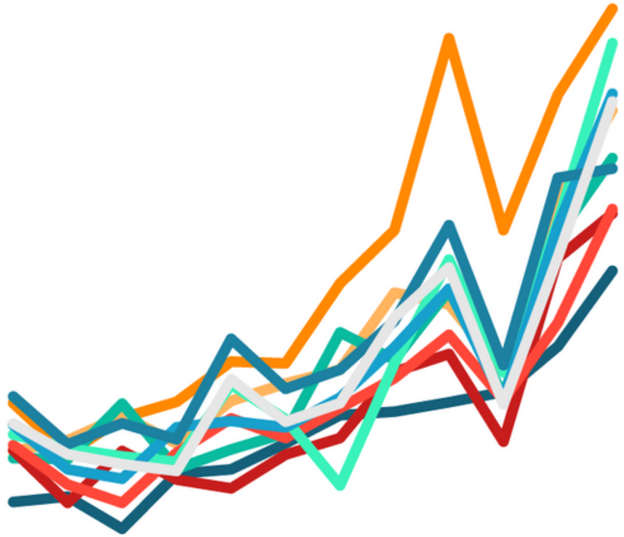


Make sure to **explain to readers** what your colors encode.

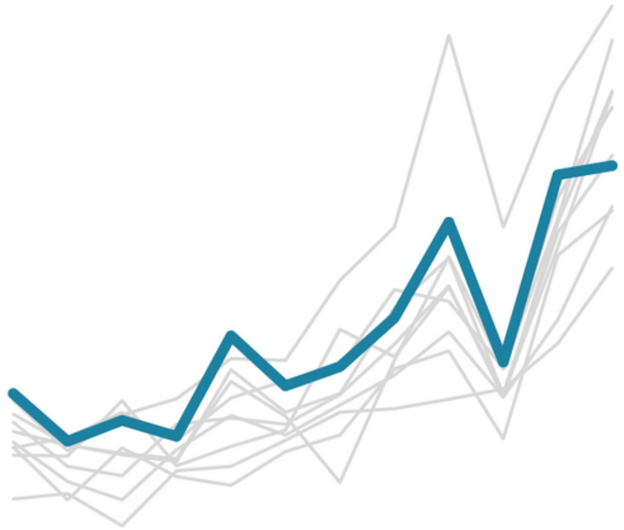


Consider the color grey as the **most important color** in visualization.

NOT IDEAL



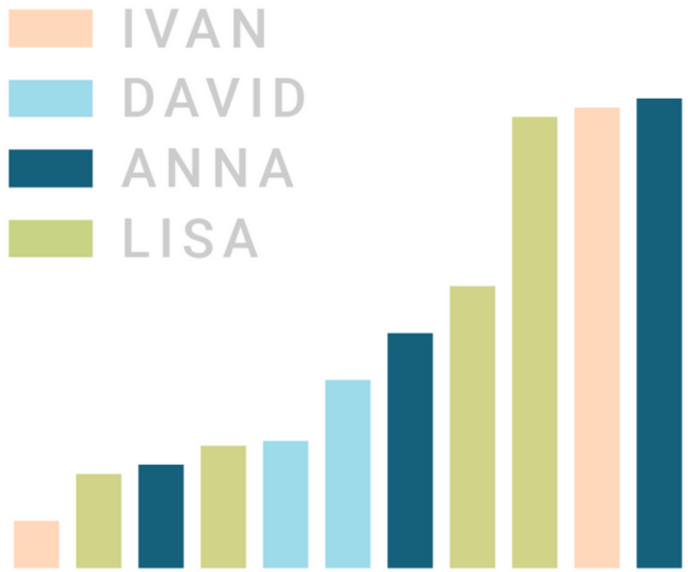
BETTER



Don’t use a **gradient color palette** for categories and the other way around.

NOT IDEAL

BETTER

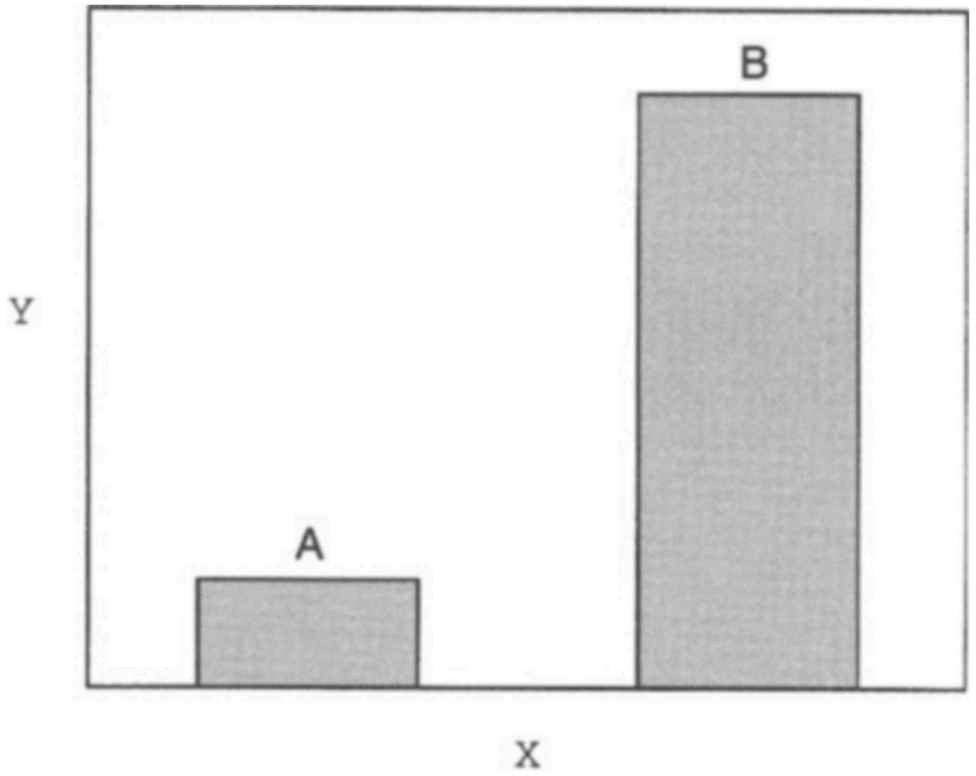


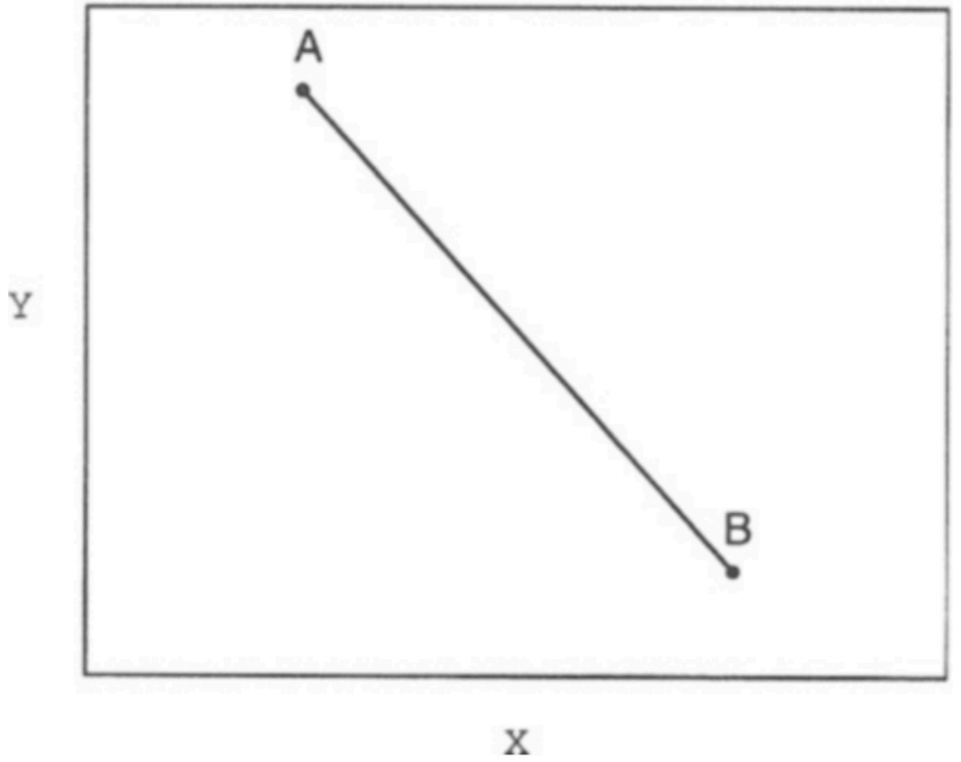
Charts & Data Types

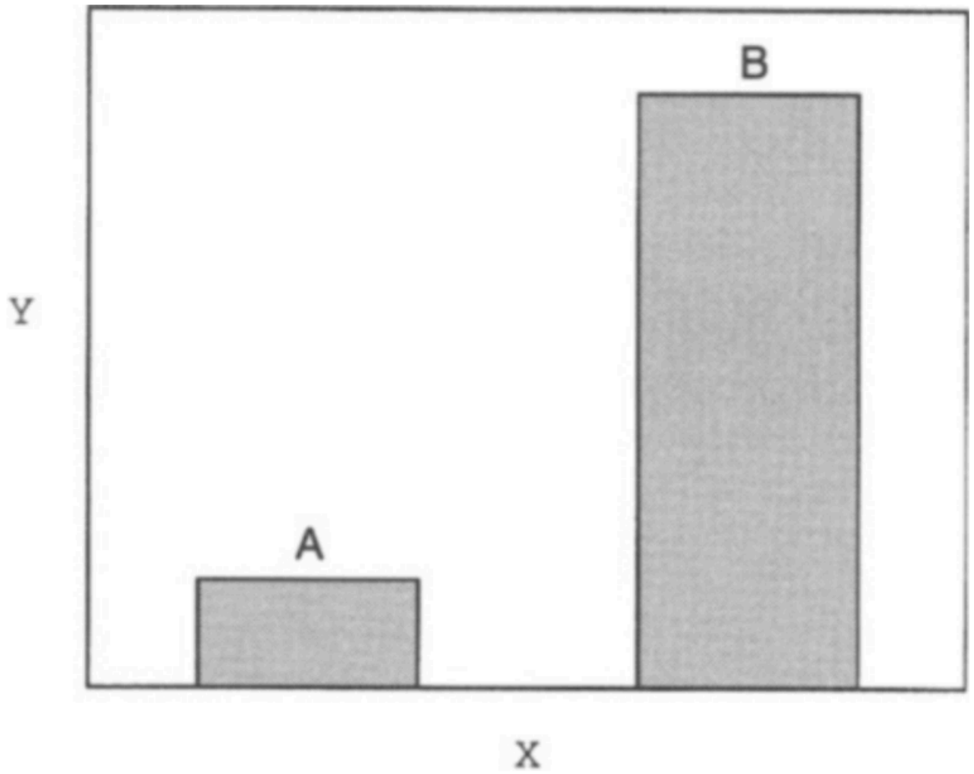
When to use a **bar chart**, and when to use a **line chart**?

Bar-line message correspondence—

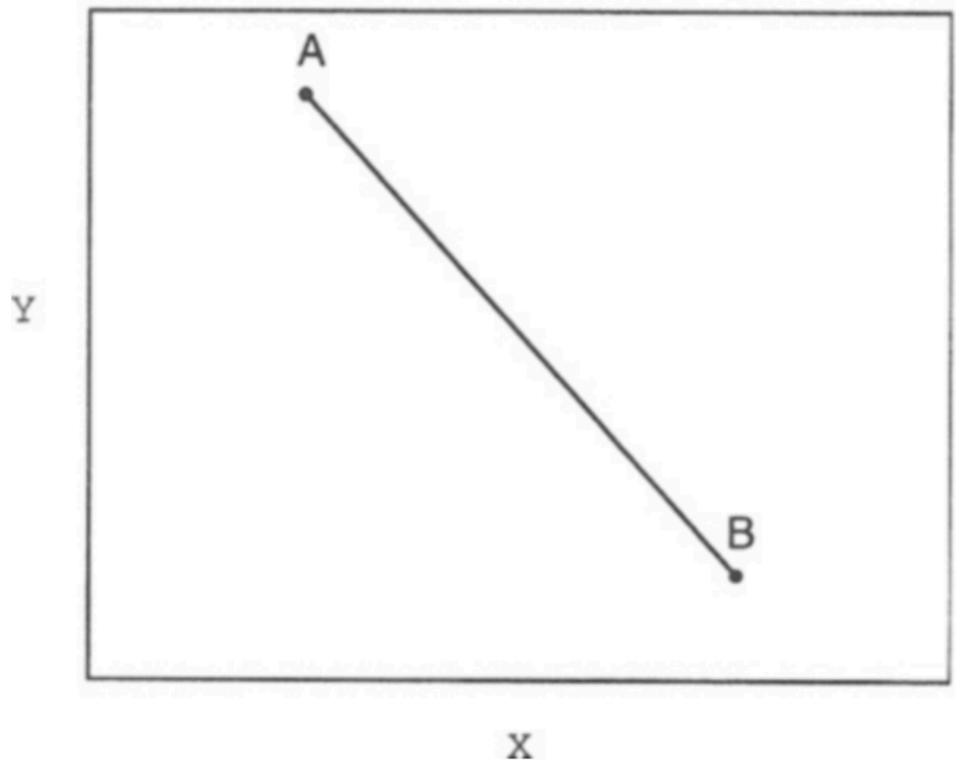
- People more readily associate **bars with discrete comparisons** between data points because bars are discrete entities and facilitate point estimates
- They more readily associate **lines with trends** because lines connect discrete entities and directly represent slope
- This correspondence does *not* seem to depend on knowledge of ‘rules’


B
A
Y
X


A
B
Y
X



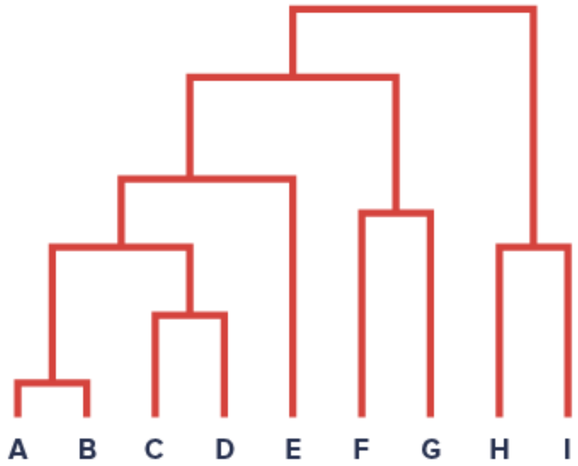
“B is **higher than** A.”



“A is **decreasing**.”

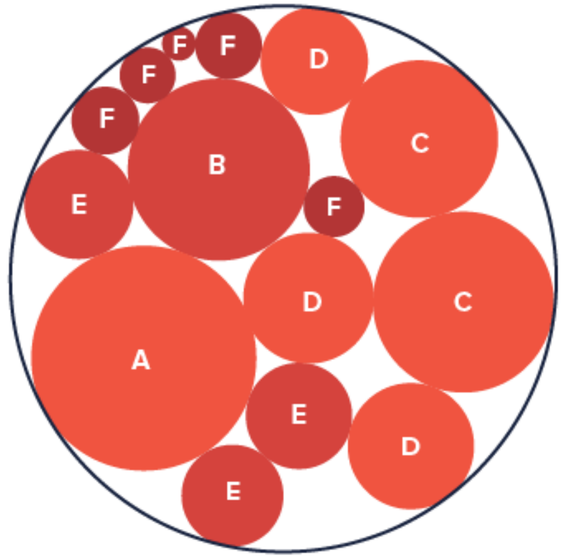
What visualization methods express **hierarchy**?

Dendrogram



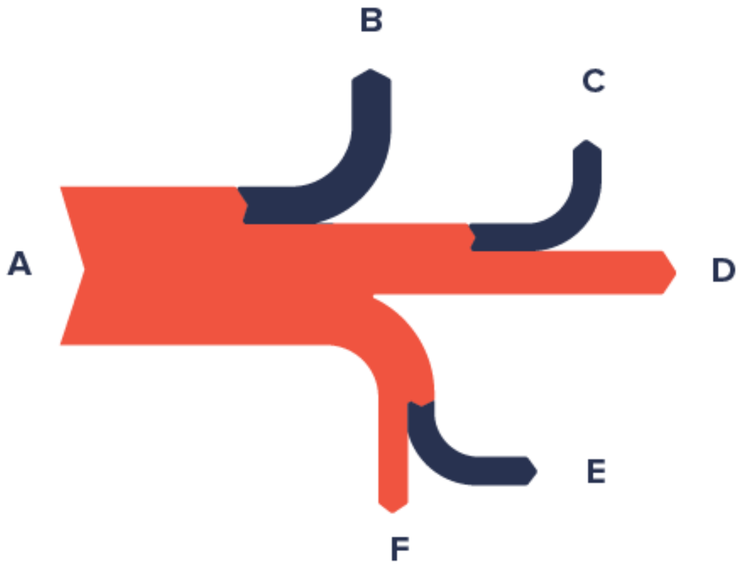
Sunburst

Packed Circle Chart

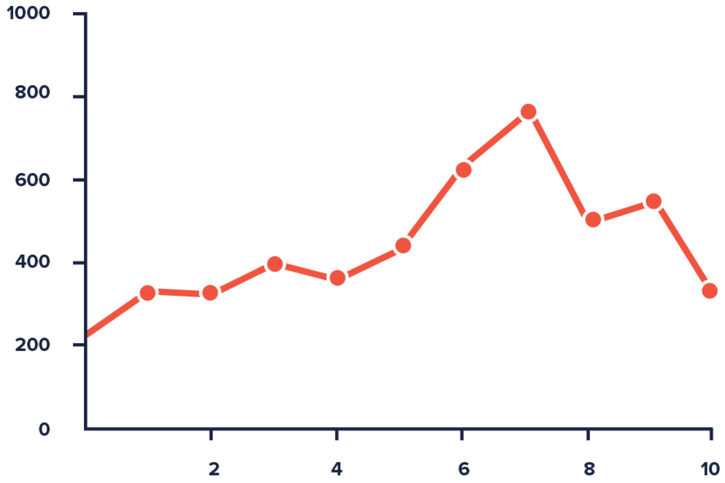


What visualization methods express **change over time**?

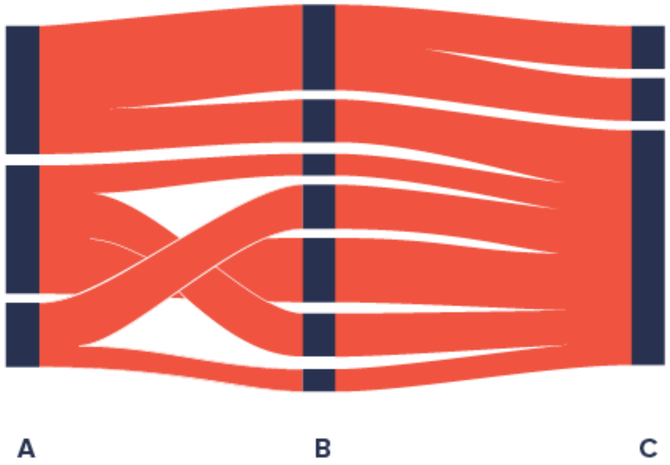
Sankey Diagram



Line Chart

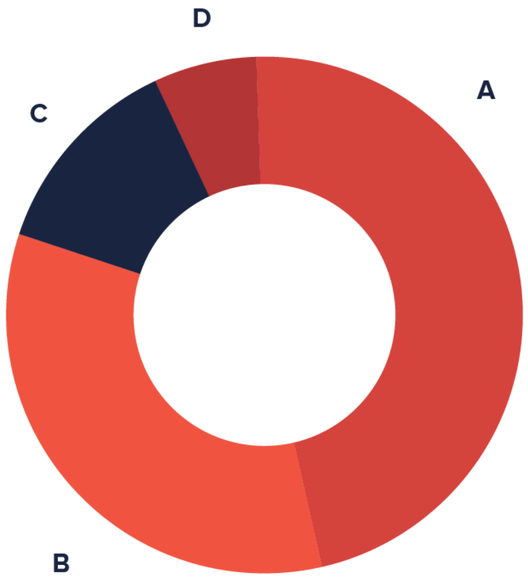


Alluvial Diagram

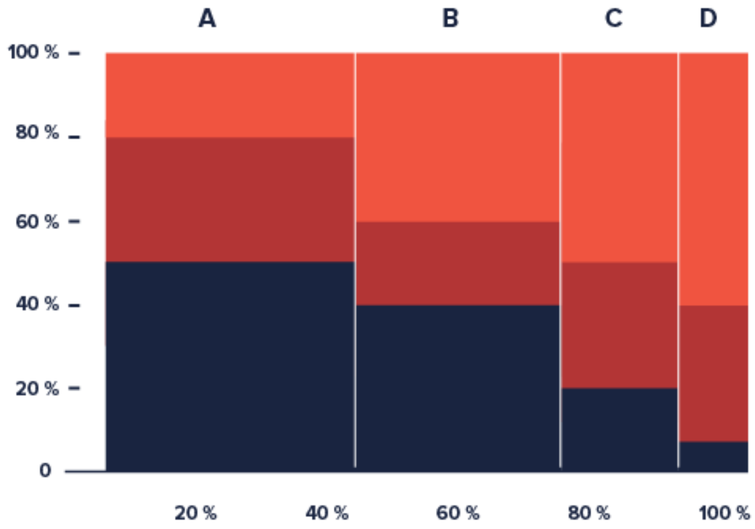


What visualization methods express **part-to-whole** relationships?

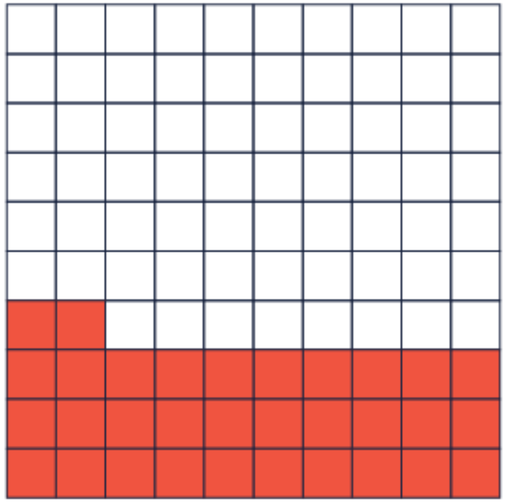
Donut Chart



Marimekko Chart



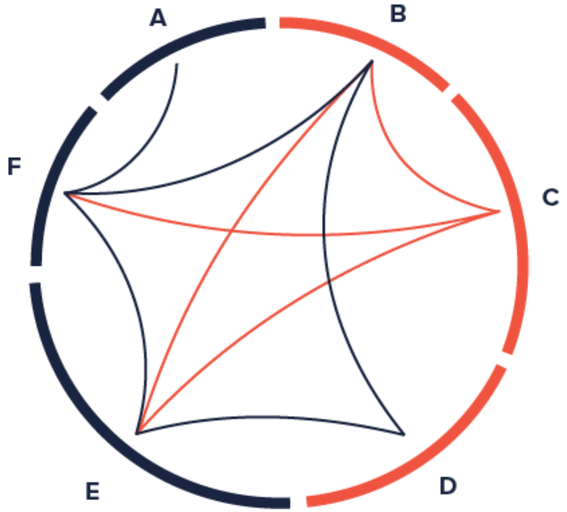
Waffle Chart



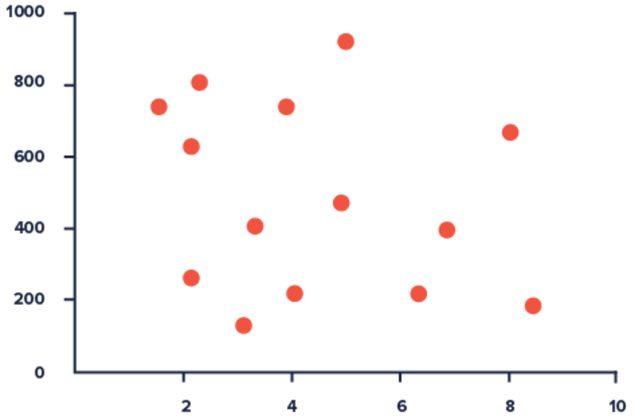
32%

What visualization methods express **interconnectivity**?

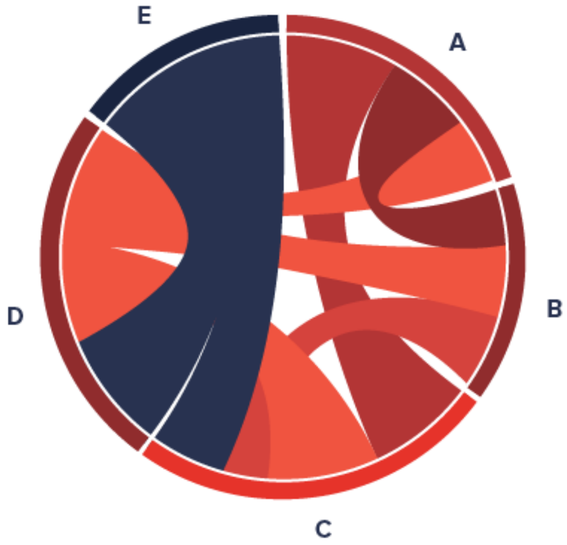
Non-Ribbon Chord



Scatterplot

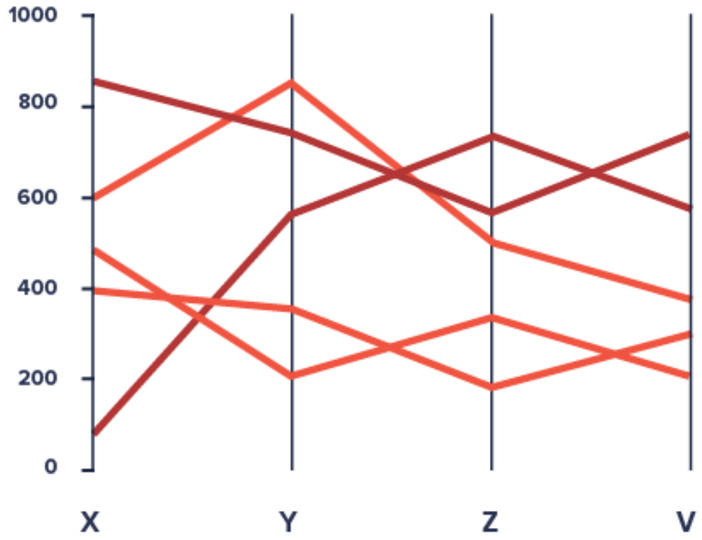


Chord Diagram

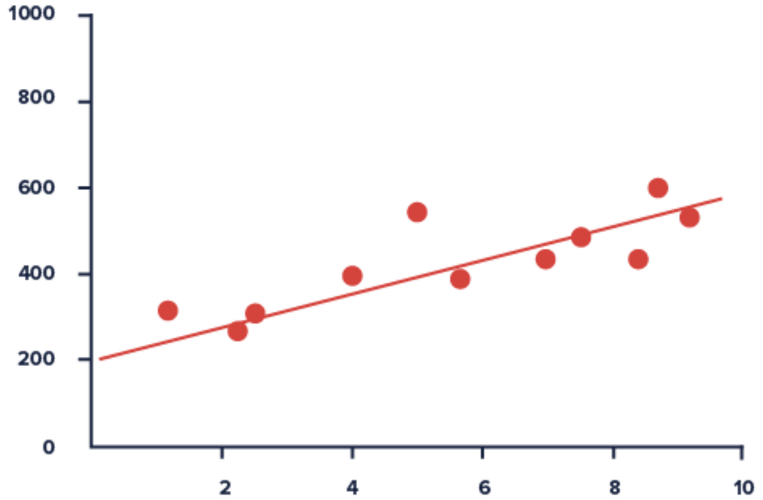


What visualization methods express **correlation**?

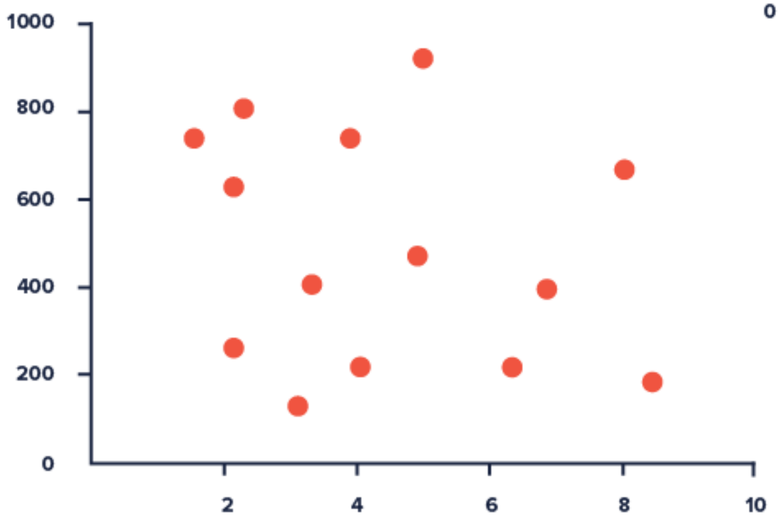
Parallel Coordinates



Trendline



Scatterplot



1. **Review:** Key Concepts
2. **Studio:** Hierarchical Circle Packing in D3.js