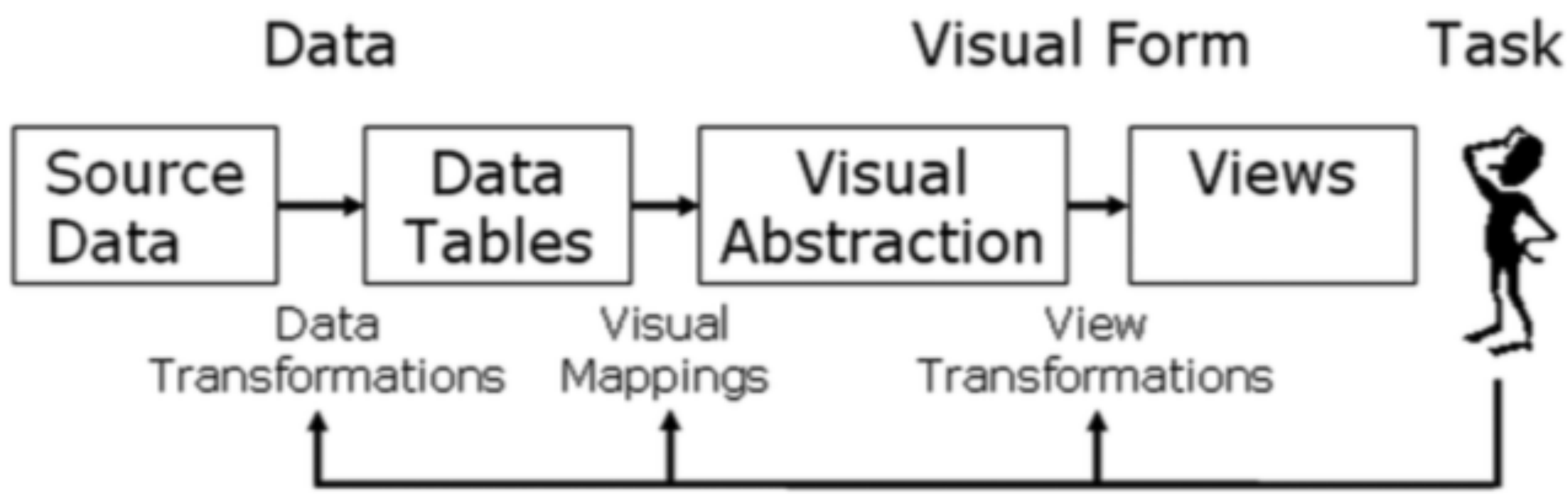
CS-GY-9223
Data Visualization Fundamentals

Course materials based on CS 8395-03 Visual Analytics & Machine Learning
By Matt Berger, Vanderbilt University
https://matthewberger.github.io/teaching/vaml/spring2019/

Agenda: Data Vis 101

- Visual encodings
- Interactions
- Sensemaking

Visualization Design



[Heer 2006]

- Visualization design is all about *choices*.
- Creating a data visualization is easy! (once you are adept with a visualization library 😅)
- Creating good visualizations is *hard*! The visualization design space is **huge**! It all comes down to making good choices for the problem at hand.

Principles of Visualization

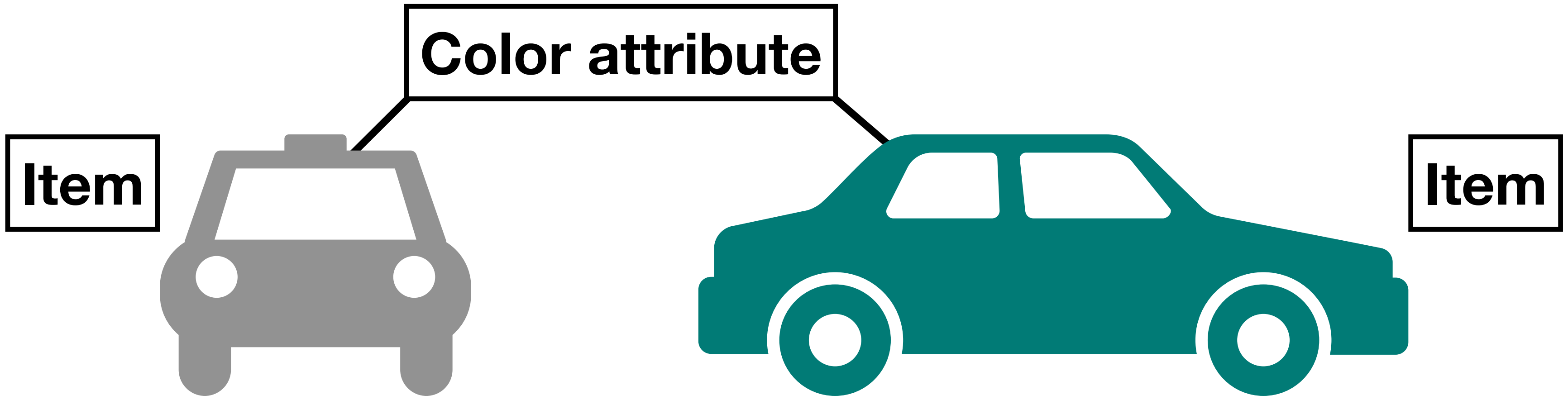
- Data
- Visual Marks
- Visual Channels
- Interaction
- Perceptual Principles and Guidelines

Data

- **Item:** an individual, discrete entity - record, data point, etc..

- **Attribute:** property of an item that can be measured, observed, or logged

- Example: a <u>car</u> is an item, a car's horsepower, year, color are its <u>attributes</u>.



Datasets: Tabular Data

Attributes

Car	Horsepower	Year	Color
Car 1	60	2013	silver
Car 2	86	2015	green
Car 3	55	1999	red
Car 4	50	1990	blue

Items

Cell

Attribute Types

- How do we work with attributes?
- Useful to group attributes into different *types*.

[Munzner 2014]

Categorical: Nominal

- Attribute draws from a discrete set

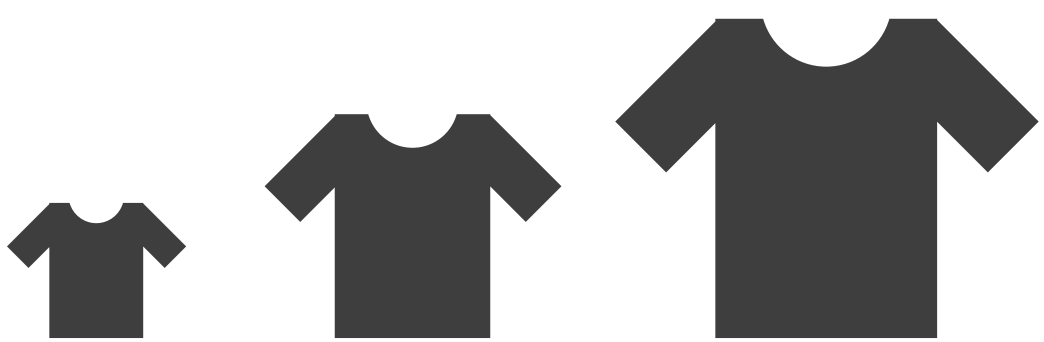
- But there may exist hierarchy in categories that is useful to preserve

Fruits

Vegetables

Ordered: Ordinal

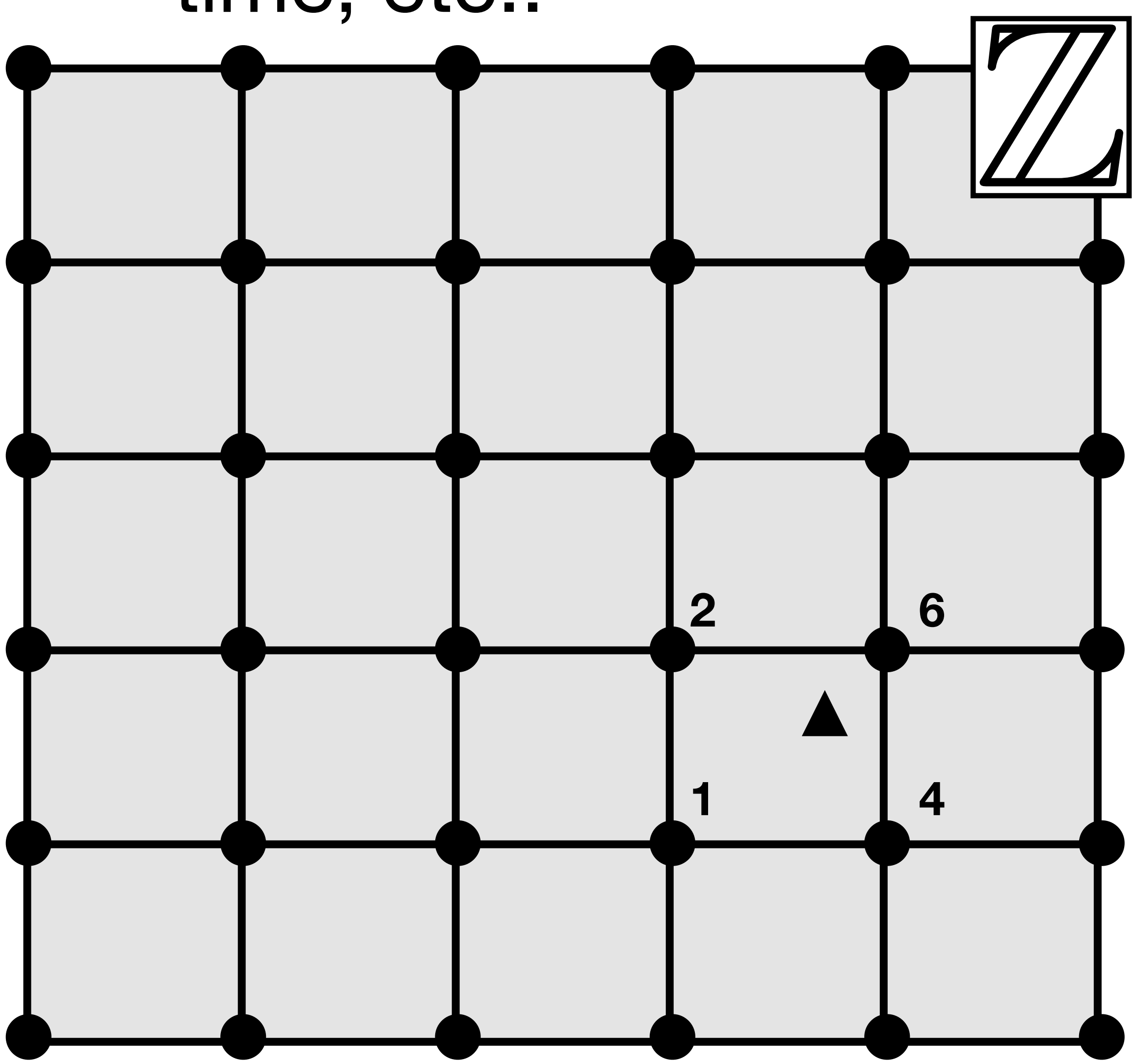
- There exists a natural *ordering* in the attribute
- Can’t necessarily perform mathematical operations though!

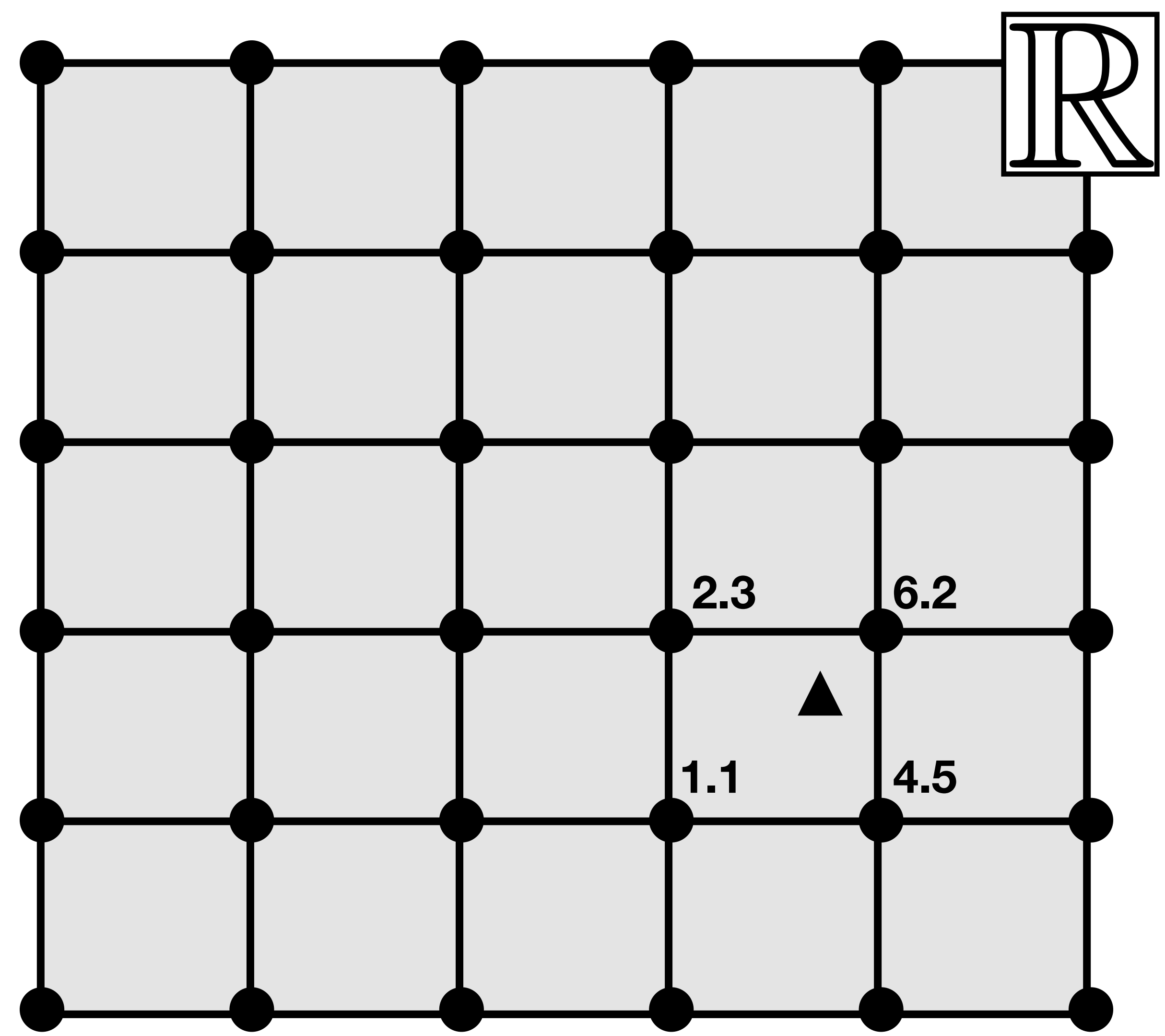


1	Virginia
2	Villanova
3	Xavier
4	Kansas
5	MI State
6	Cincinnati
7	Michigan
8	Gonzaga
9	Duke
10	UNC
11	Purdue
12	Arizona
13	Tennessee

Ordered: Quantitative

- Mathematical operations make sense*
- Height, temperature, pressure, density, population size, time, etc..





Visual Marks

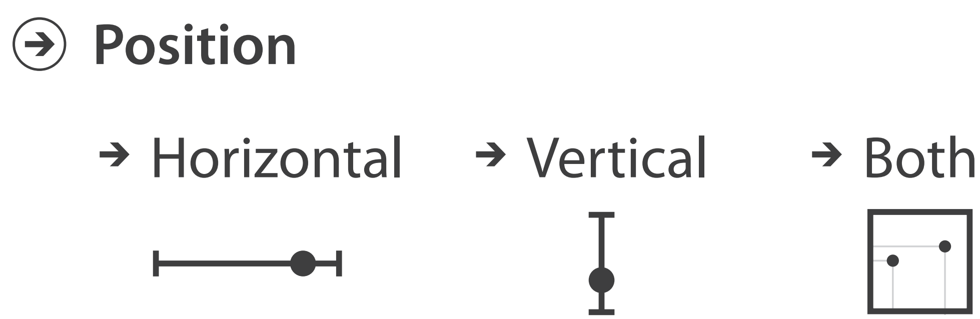
- So how do we visually encode data? Let us start with…
- **Visual marks:** geometric primitives

[Munzner 2014]

Note: marks do not have a visual instantiation!
Purely geometric types.

Visual Channels

- Controls the appearance of visual marks



Marks and Channels

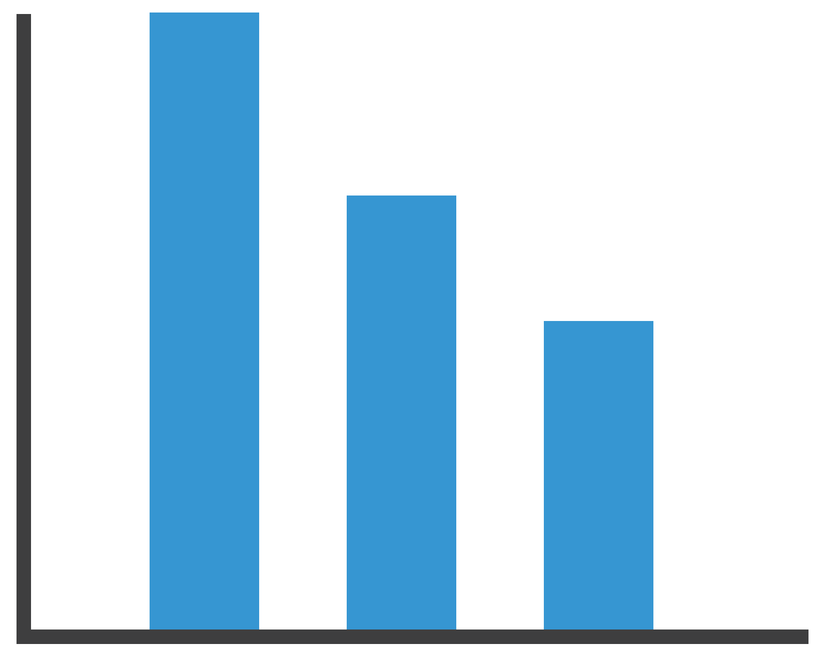
- Frequently (but not always), we associate tabular data with visual marks and channels as follows:

Attributes mapped to channels

Items mapped to marks

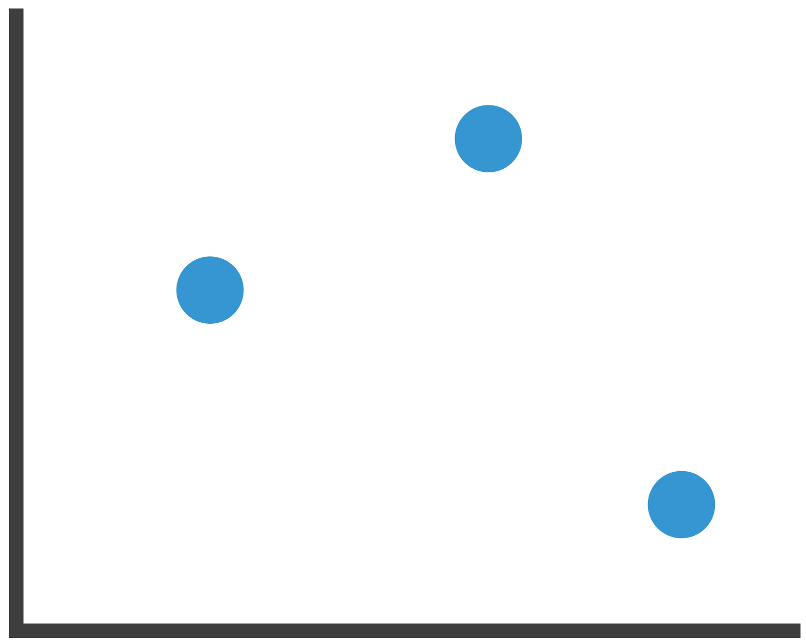
Car	Horsepower	Year	Color
Car 1	60	2013	silver
Car 2	86	2015	green
Car 3	55	1999	red
Car 4	50	1990	blue

Data, Marks, Channels, Pt. 1



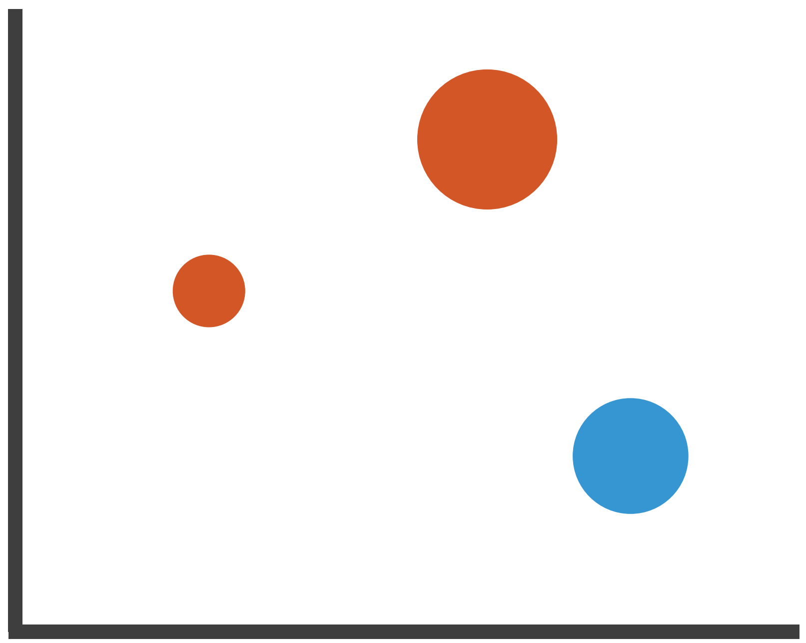
- Marks: 1D
- Channels: spatial position of bars, line height, … anything else?
- Data: each bar is an **item**, **quantitative attribute** mapped to y spatial channel, **categorical attribute** mapped to x spatial channel.

Data, Marks, Channels, Pt. 2



- Marks: ?
- Channels: ?
- Data: ?

Data, Marks, Channels, Pt. 3



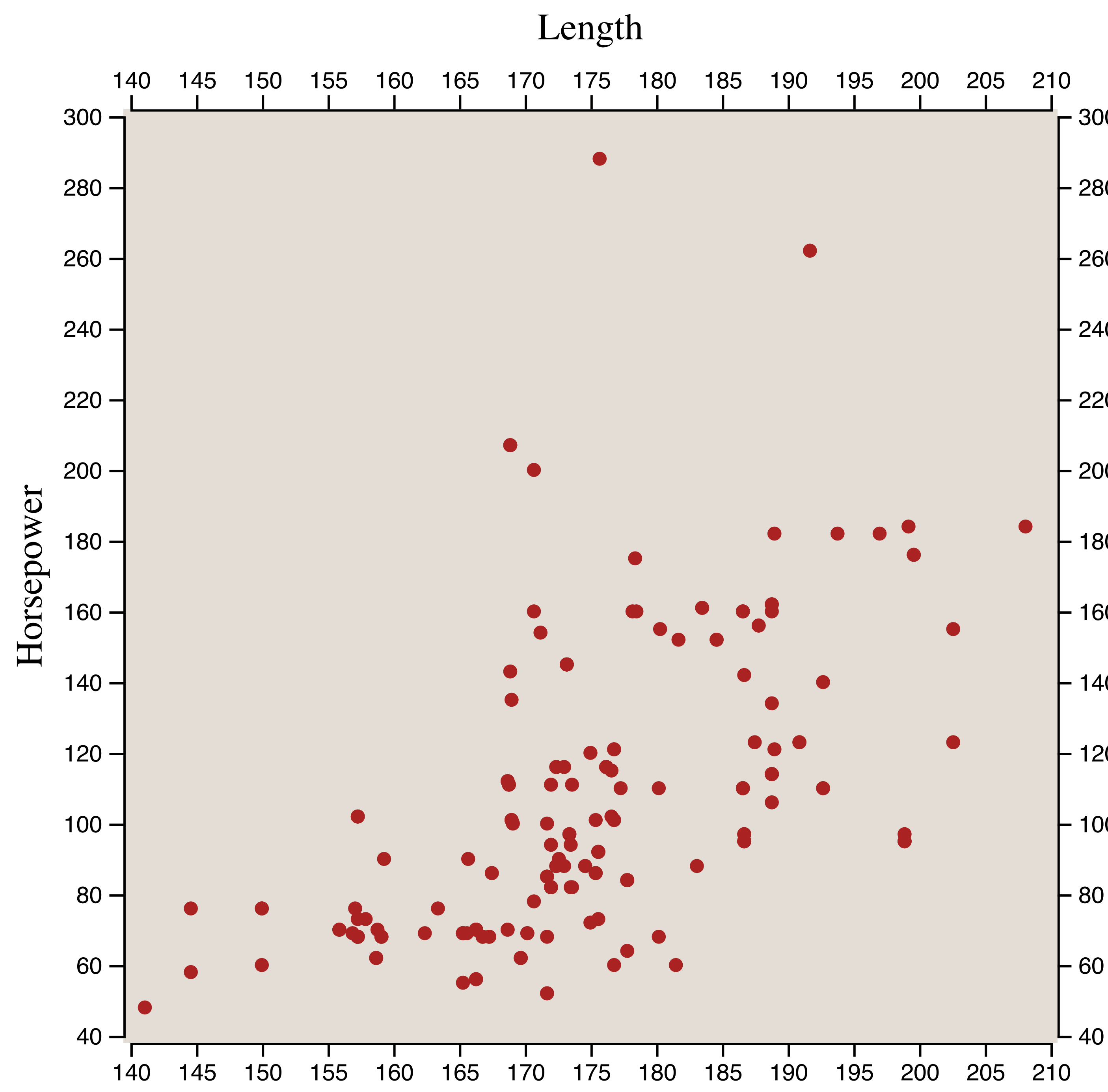
- Marks: ?
- Channels: ?
- Data: ?

Mapping Data to Visual Marks and Channels

Car	Horsepower	Price	Length	Body Style	Make
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge
...	...	...	...	...	...

Mapping Data to Visual Marks and Channels, Pt. 1

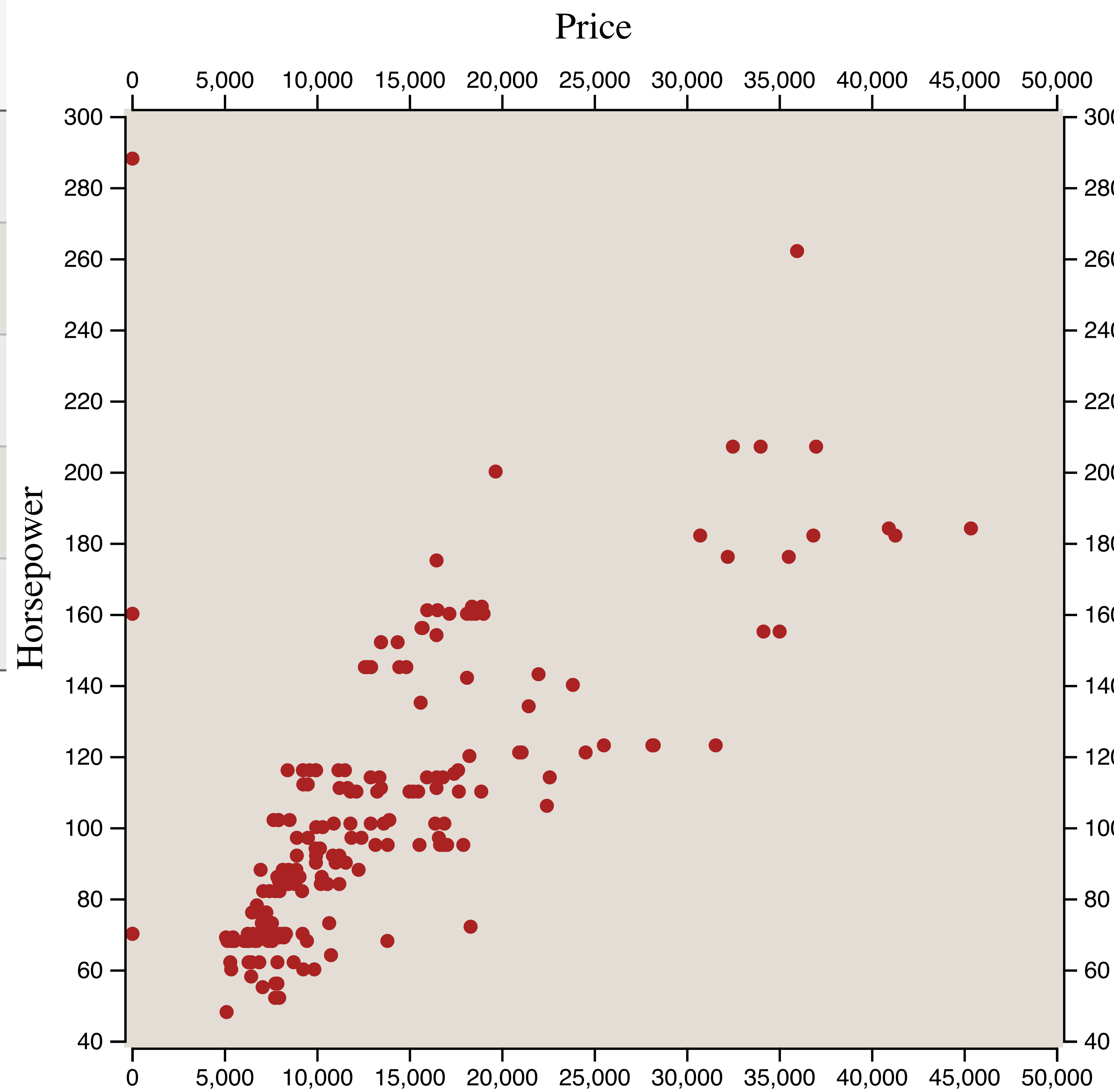
Car	Horsepower	P
Car 1	60	10
Car 2	86	12
Car 3	55	11
Car 4	50	20
...	...	



- Marks: ?
- Channels: ?

Mapping Data to Visual Marks and Channels, Pt. 2

Car	Horsepower
Car 1	60
Car 2	86
Car 3	55
Car 4	50
...	...

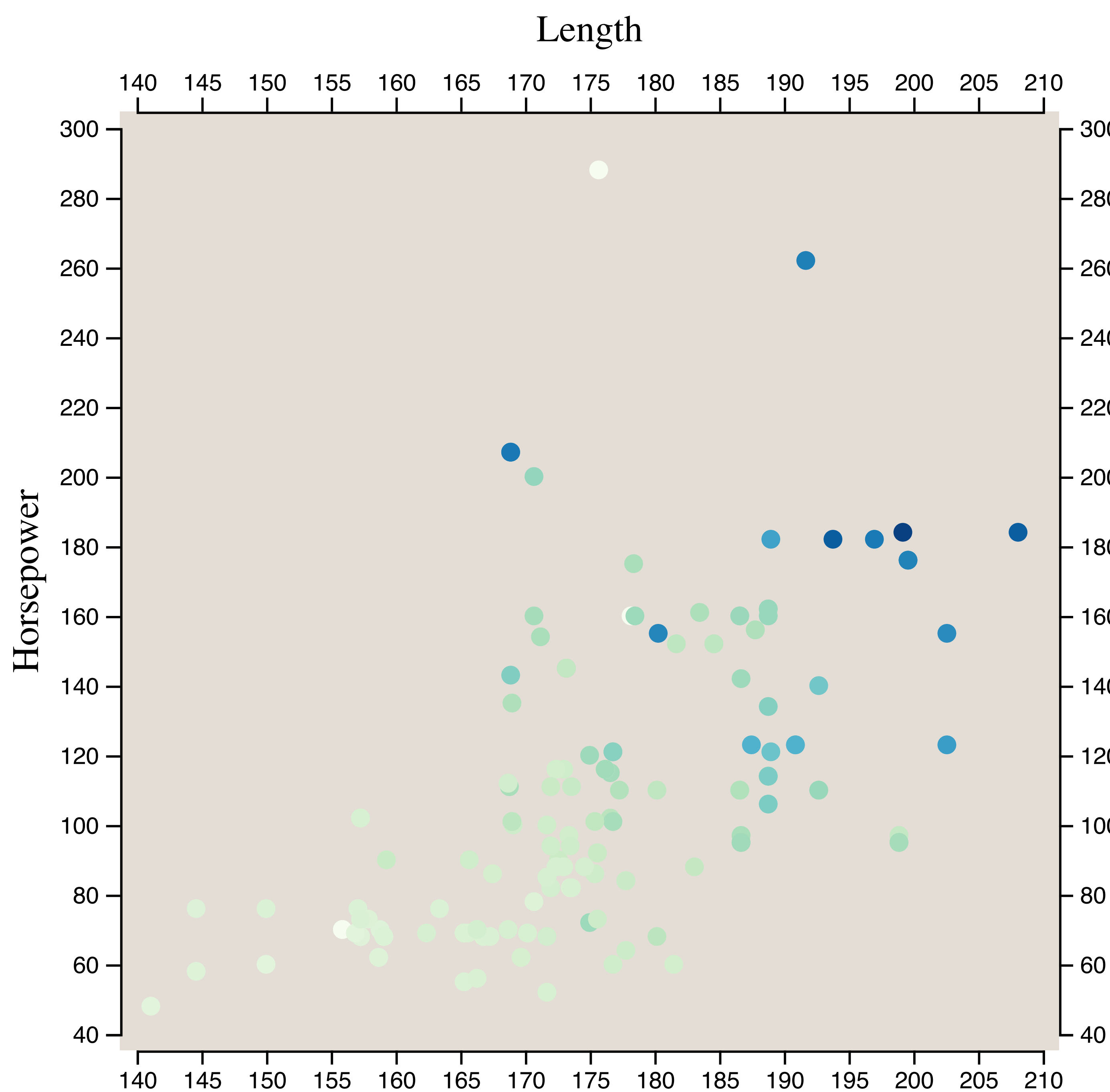


- Marks: ?
- Channels: ?

Mapping Data to Visual Marks and Channels, Pt. 3

Car	Horsepower
Car 1	60
Car 2	86
Car 3	55
Car 4	50
...	...

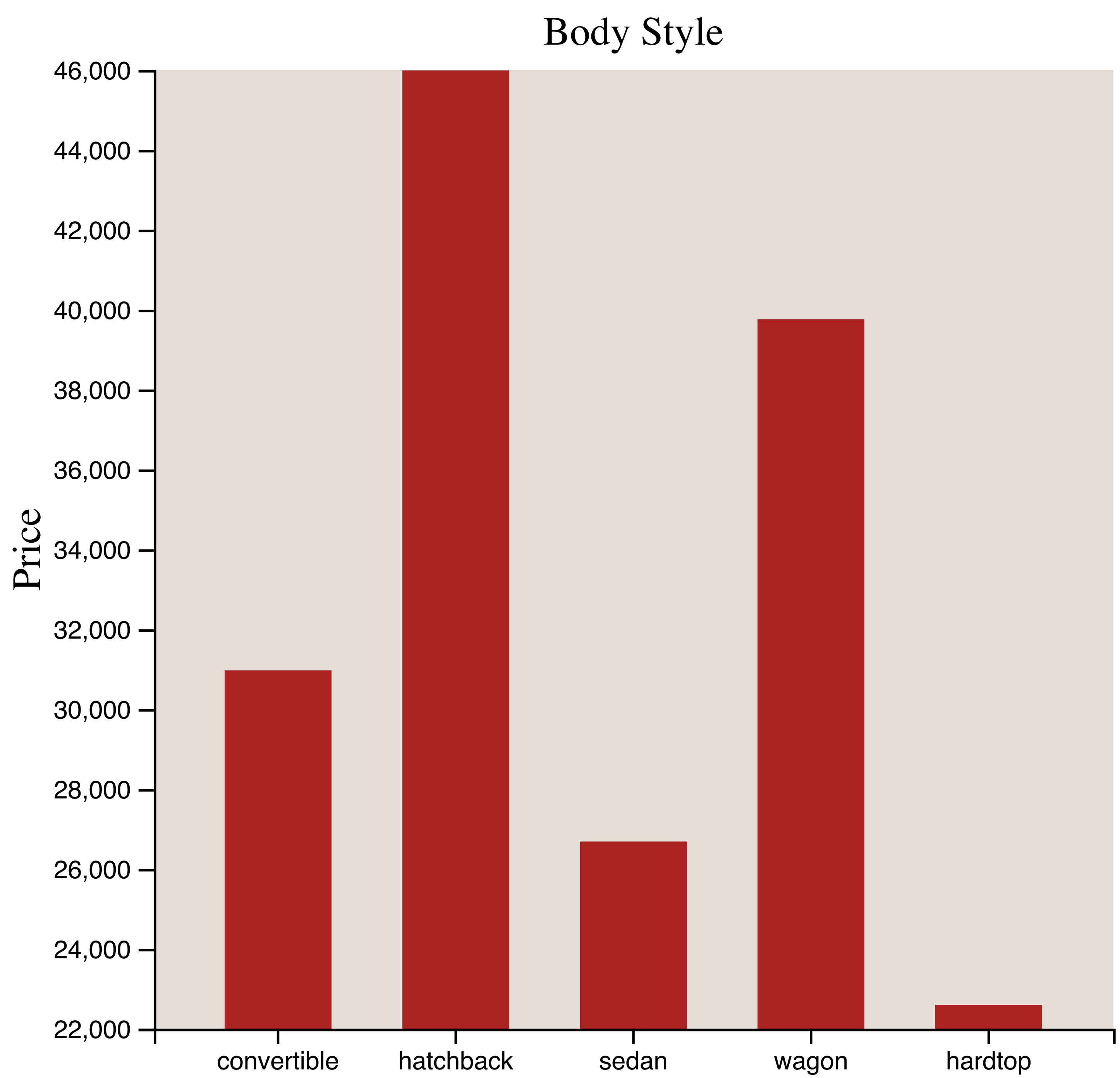
- Marks: ?
- Channels: ?



Mapping Data to Visual Marks and Channels, Pt. 4

Car	Horsepower
Car 1	60
Car 2	86
Car 3	55
Car 4	50
...	...

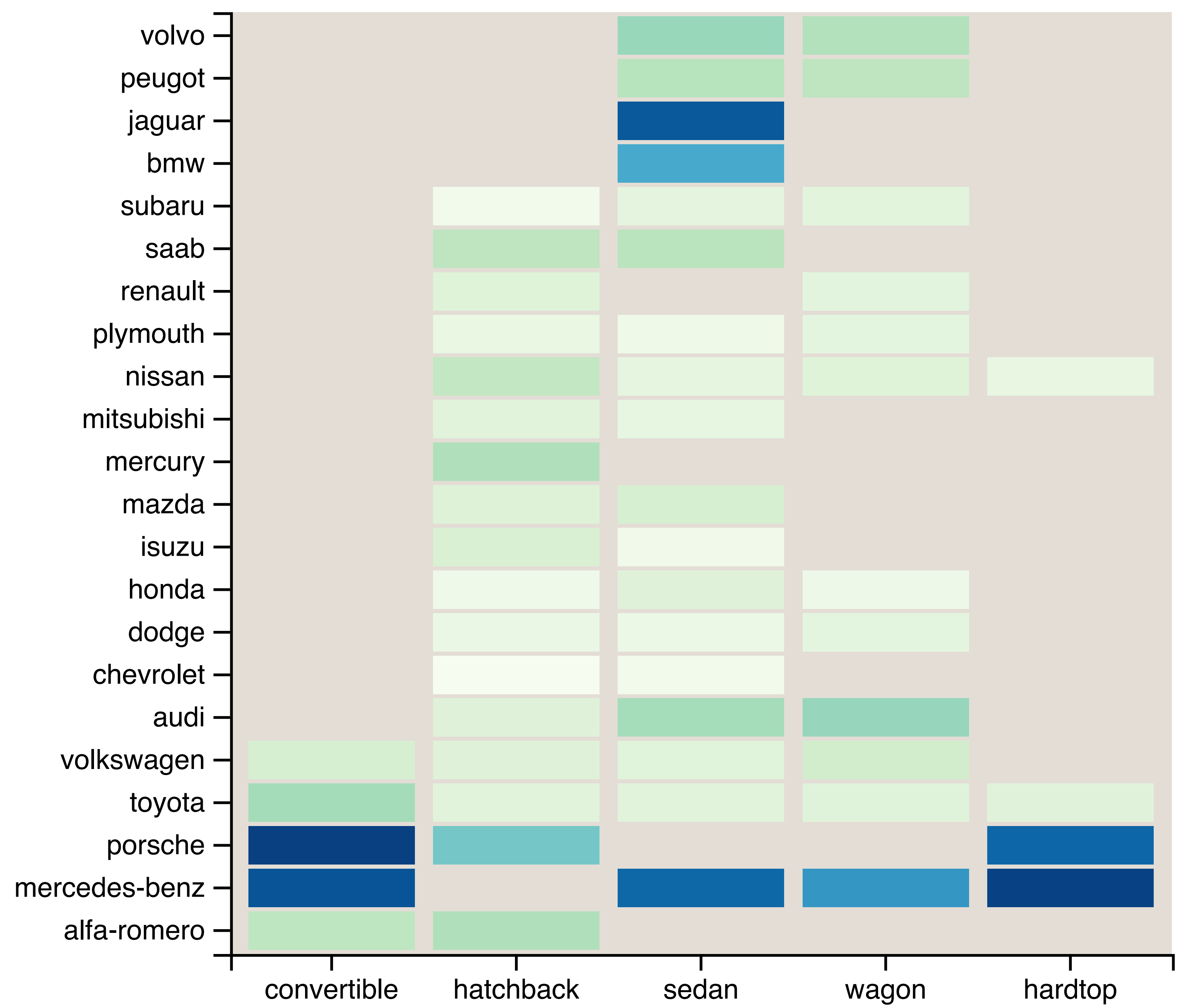
- Marks: ?
- Channels: ?



Mapping Data to Visual Marks and Channels, Pt. 5

Car	Horsepower
Car 1	60
Car 2	86
Car 3	55
Car 4	50
...	...

- Marks: ?
- Channels: ?



Project Interlude

- Designing good visualizations requires practice.
- The baseline portion of the project will give you the opportunity to experiment with high-level authoring tools.
- I encourage a good deal of experimentation as part of the baseline, so that you become more comfortable with visualization design.

Interaction

- Showing static views will only get us so far when working with complex data
    - Large number of data items (thousands or higher)
    - Large number of data attributes (tens or higher)
- **Interaction**: essential to visual exploration - *we want to enable the user to drive visualization*
- As the visualization designer, the challenge is in determining what interactions we should provide to the user.

Different types of Interaction

- Manipulating View
  - Changing our viewpoint, animated transitions, adjusting arrangements
- Brushing
  - Selecting items, highlighting, filtering attributes/data
- Coordinated Views
  - Linked selection/manipulation between many views

Information Seeking Mantra

“Overview first, zoom and filter, then details-on-demand”

- Ben Shneiderman

Interacting with Data

- **Overview first**
  - Obtain an overview of the entire dataset: general trends, patterns, outliers
- **Zoom**
  - Zoom in on items of interest
- **Filter**
  - Filter out uninteresting items
- **Details-on-demand**
  - Selection of item/group, reveal their details when necessary

Google Maps

- Exemplary of these principles

https://maps.google.com

Overview & Zoom?

- Manipulating our view of the data!
- How do we manipulate our view?

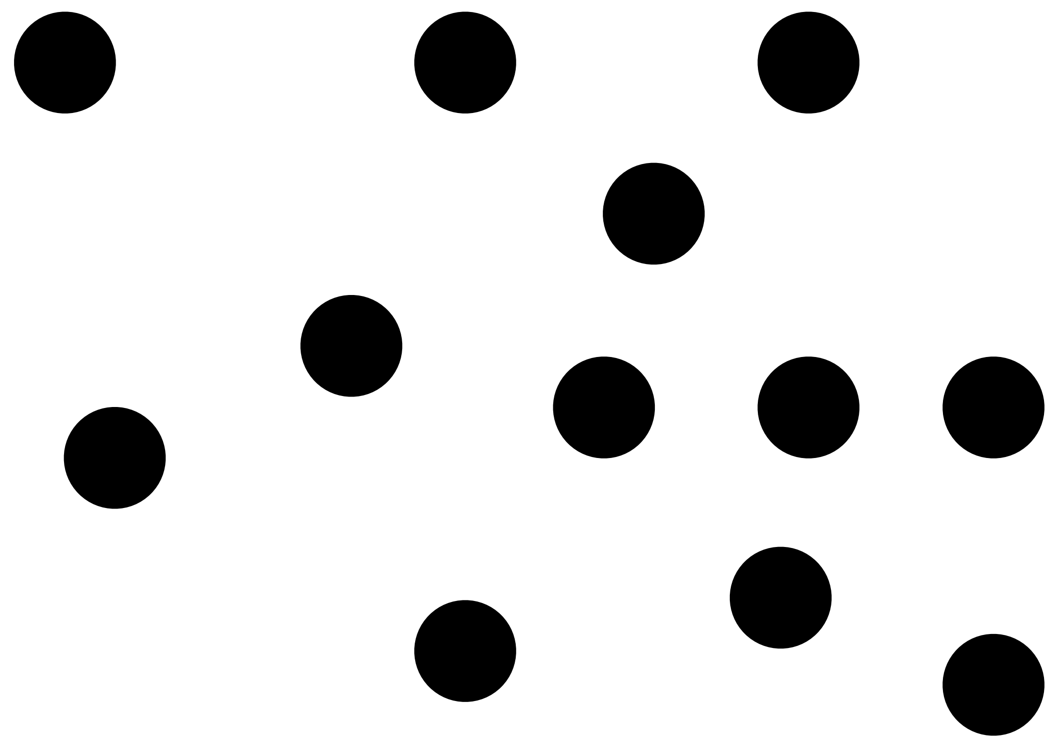
d3-zoom

https://bl.ocks.org/mbostock/3680999

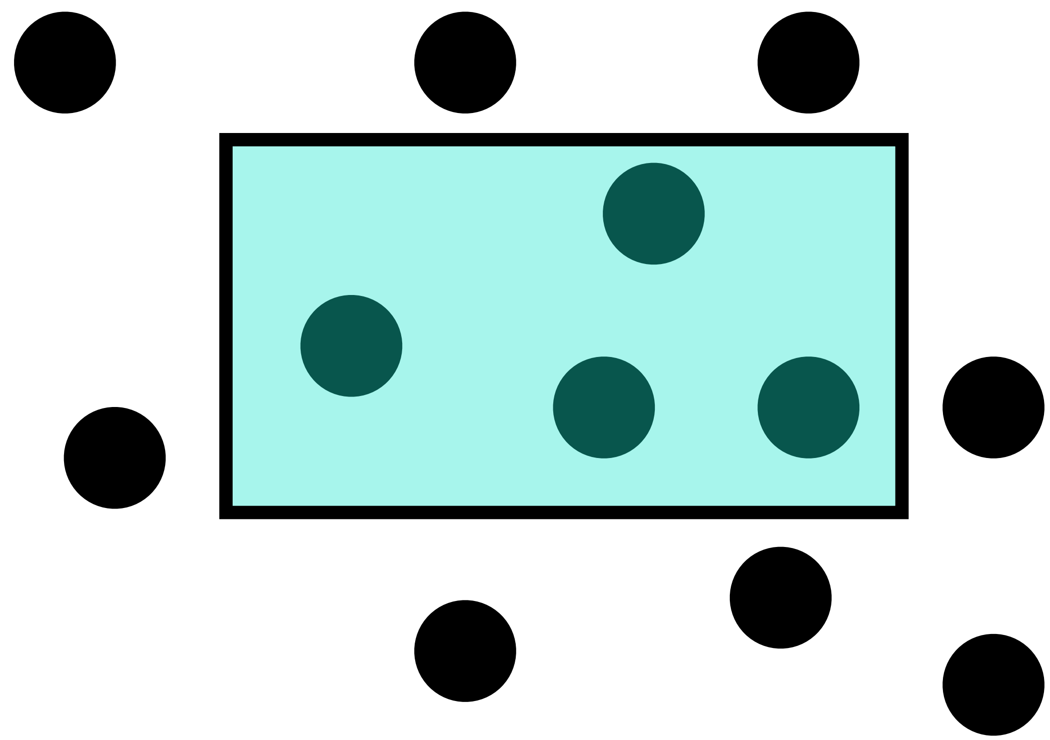
Selection

- **Selection**: a fundamental action for interactive vis
- Selection is realized by *brushing*: a user performs some action on marks
  - Mouse click
  - Draw a shape with mouse: simple shape like rectangle/circle, more complex shape like lasso
  - Simply hover mouse cursor over marks
- **Output**: selection, that often serves as **input** for subsequent operation

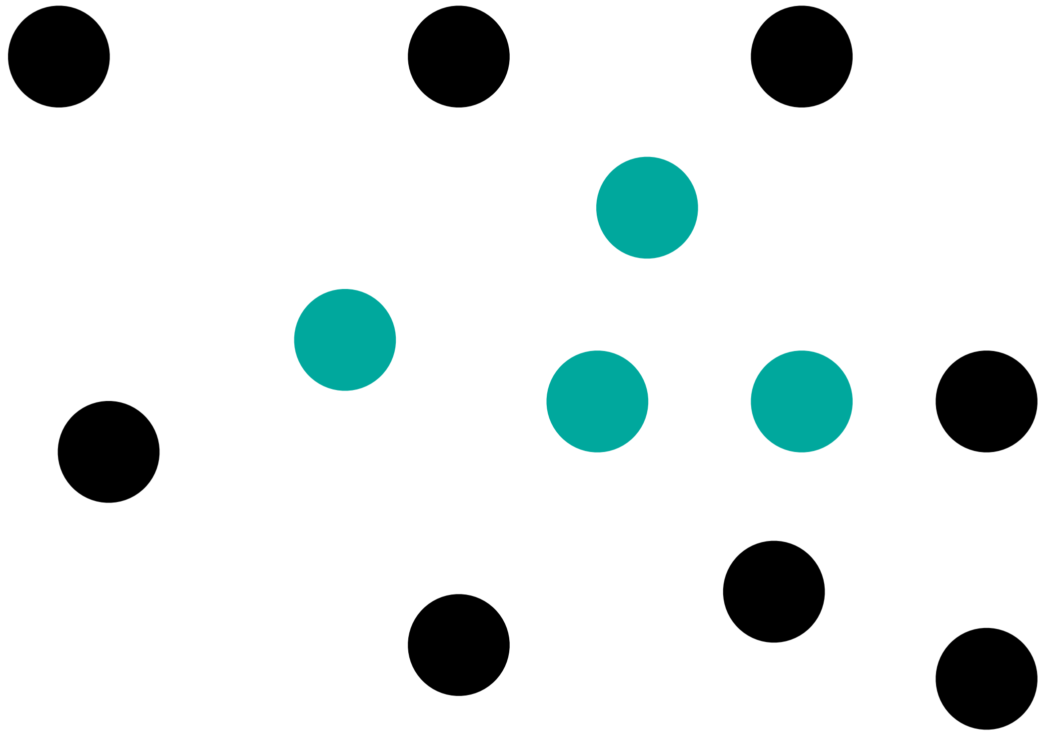
Brushing Data Items



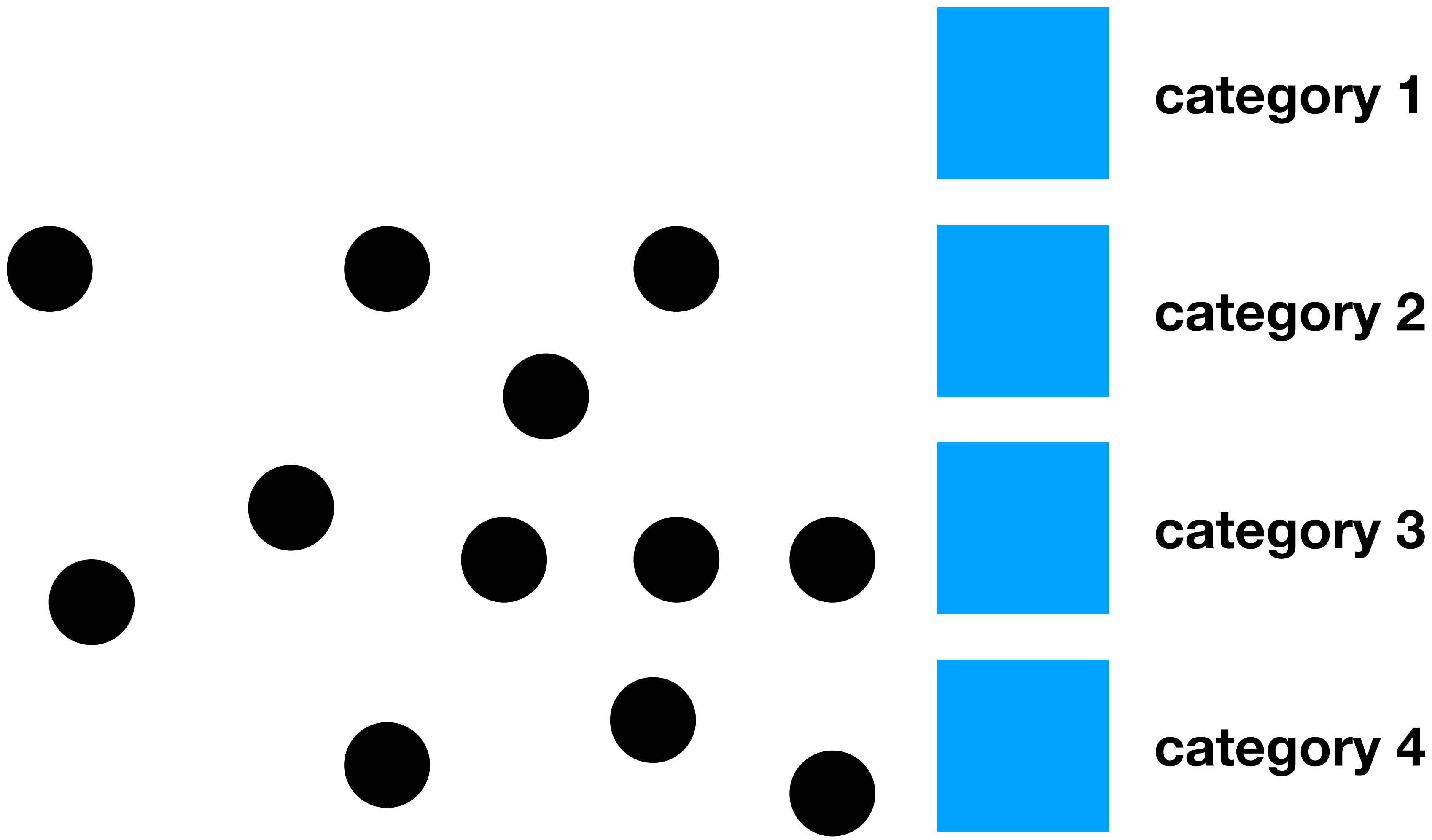
Brushing Data Items



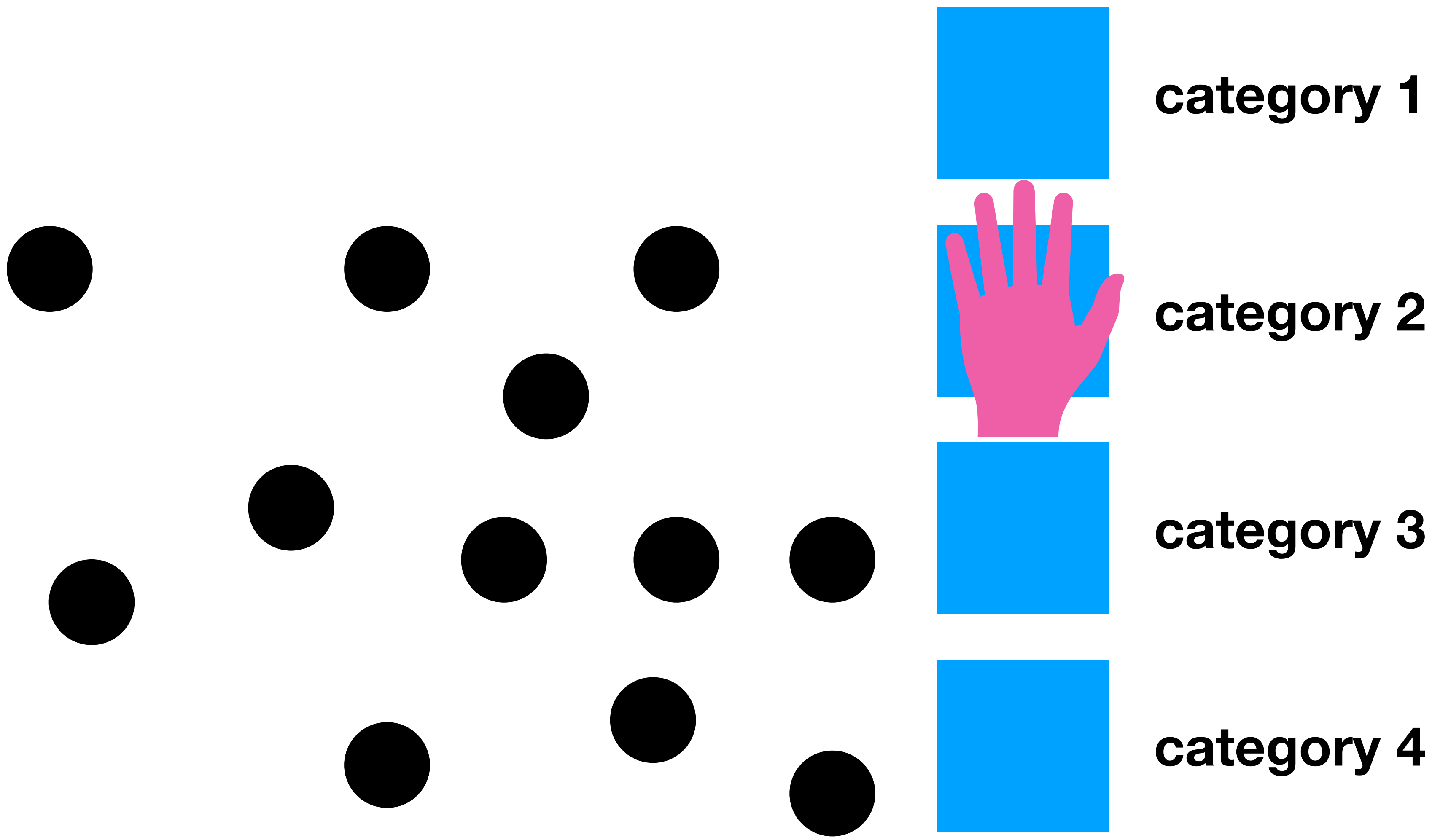
Brushing Data Items



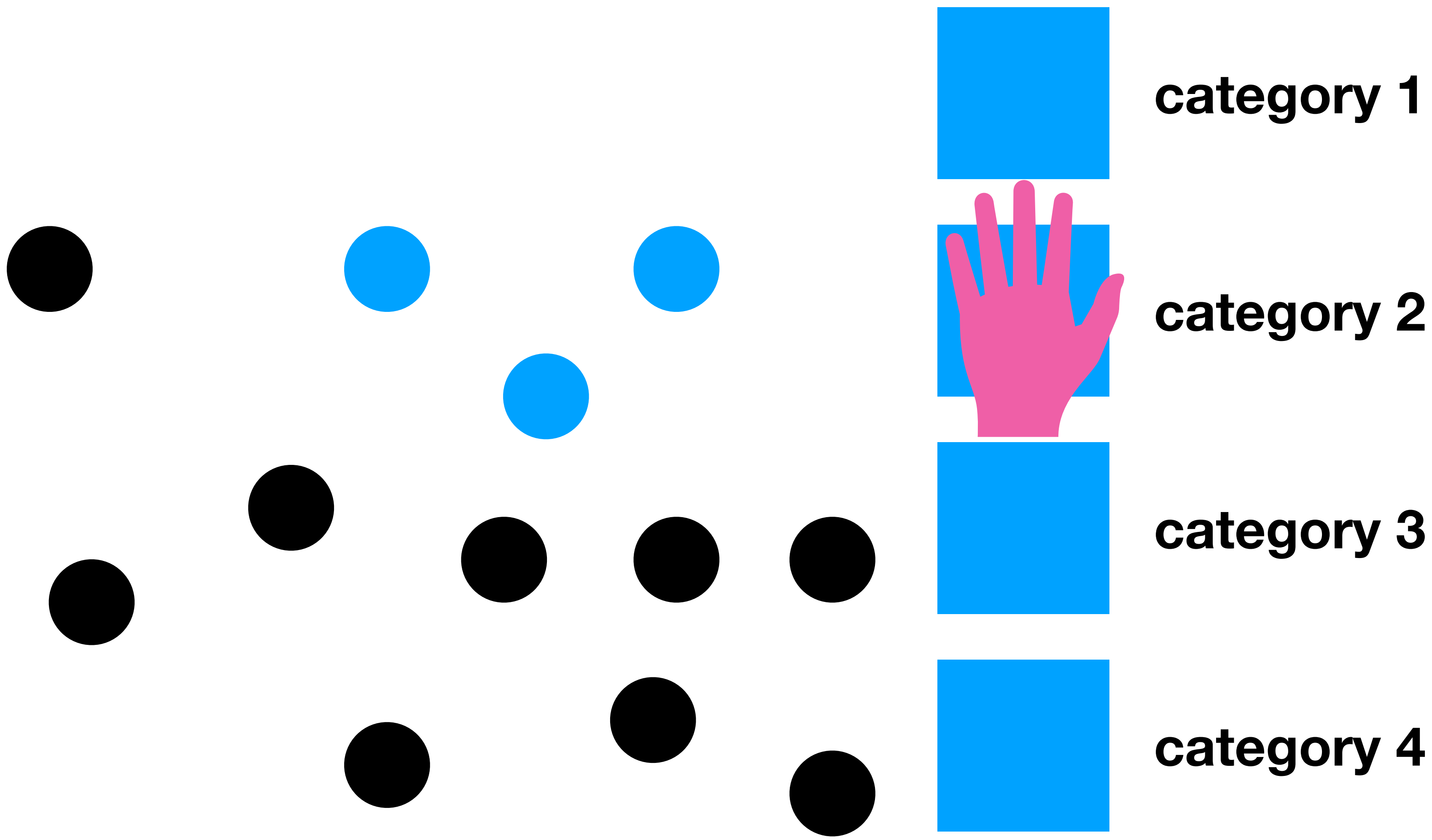
Brushing Attributes



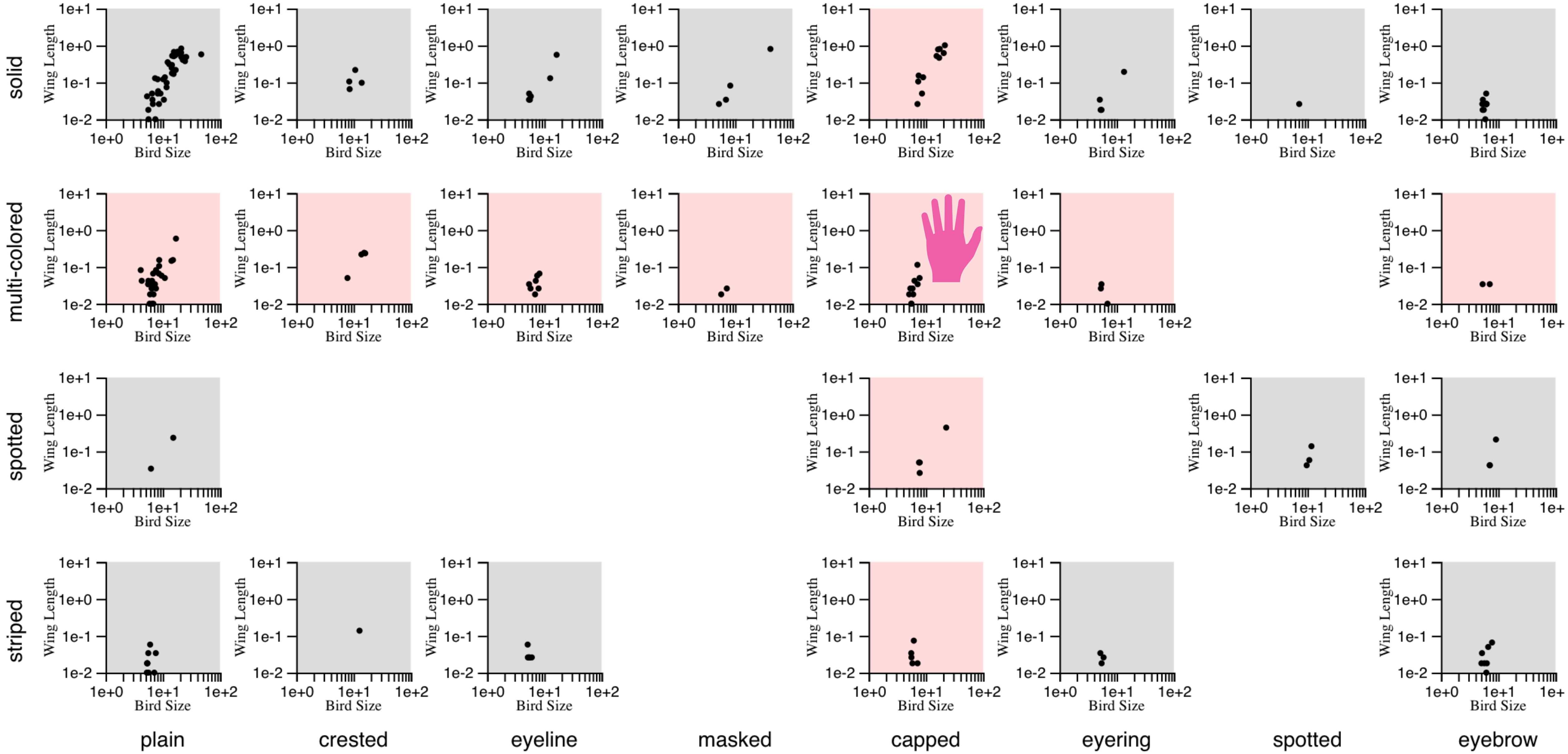
Brushing Attributes



Brushing Attributes



Multi-Attribute Brushing



Brush Lifecycle

- Initiate a brush
- Begin brushing (drag a shape)
  - Update view of selected data (**highlight**)
  - Trigger other selections (e.g. multiple views)
- Stop brushing
  - Change viewport, filter, details-on-demand
- Clearing a brush (and clearing everything else?)

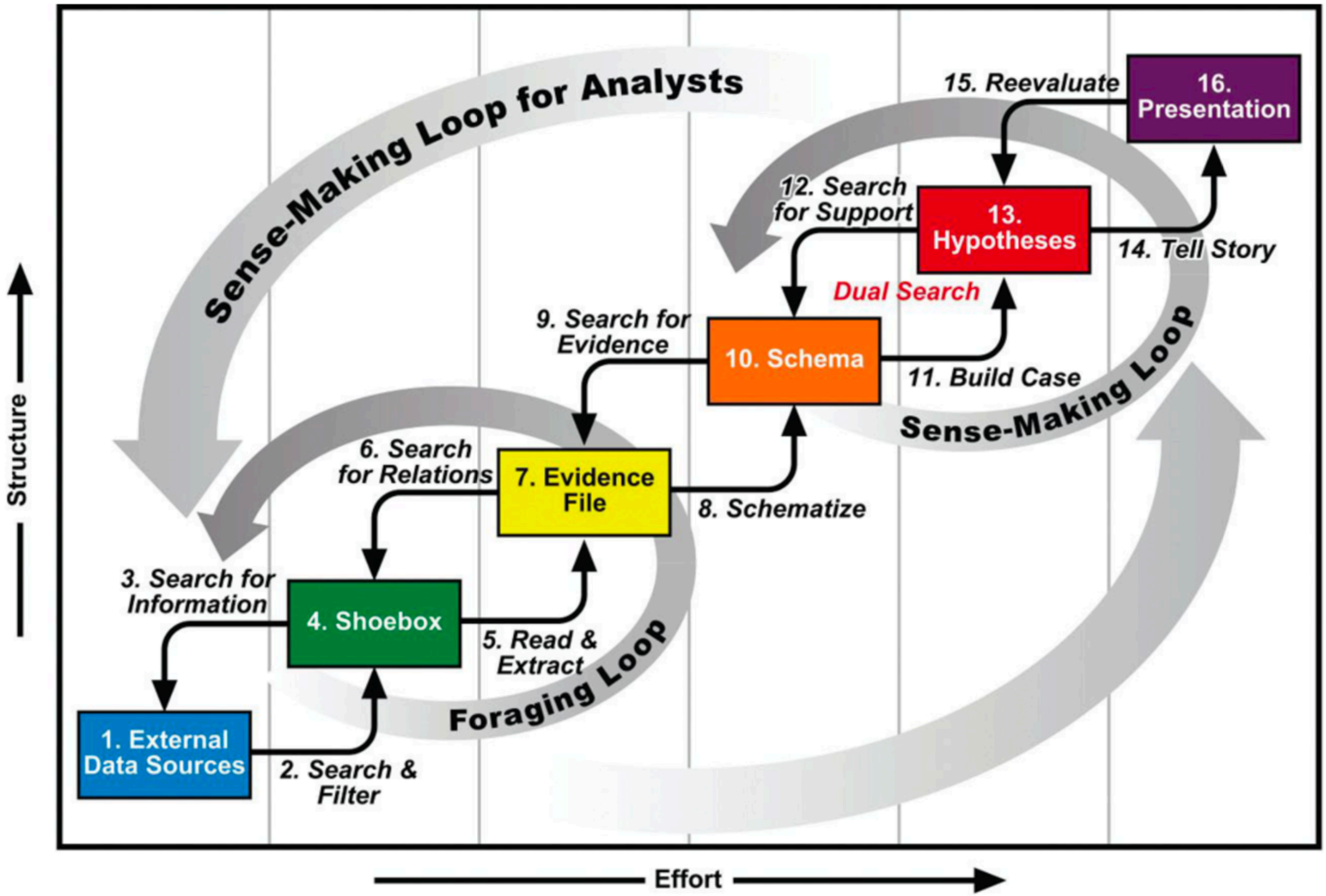
d3: Brushing Parallel Coordinates

https://bl.ocks.org/jasondavies/raw/1341281/

From Visual Encodings to Visual Analytics

- So, we now know *how* to create visualizations, but we are still left with a **ton of decisions**!
- We want to use visualization to help humans make sense of a problem.
    - Finding and assembling data into meaningful explanations
    - Searching, retrieving, organizing, indexing, storing information.
    - Formulating & testing hypotheses, iterating

Sensemaking Model

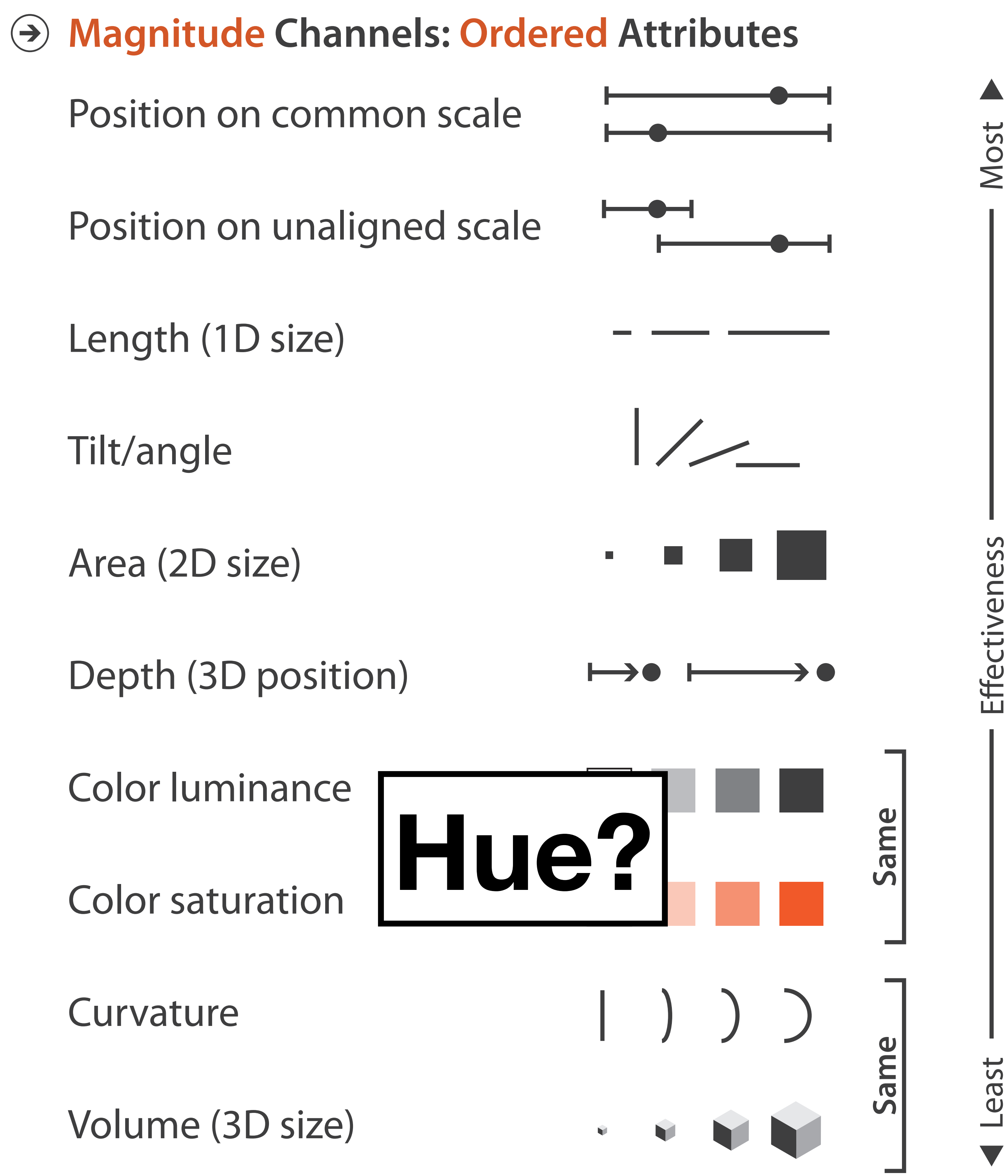


[Pirolli & Card 2005]

Visualization for Sensemaking

- Our design decisions should support how a human goes about making sense of data.
    - Visualization enables us to: <u>reduce search for necessary info</u>, <u>substitute perceptual inferences for more demanding logical inferences</u>
- We have a limited work memory. Visualization helps us to externalize our thoughts.
- We should make decisions that adhere to human perception.
- Let us start with visual channels. Are certain ones better than others?

Channel Rankings

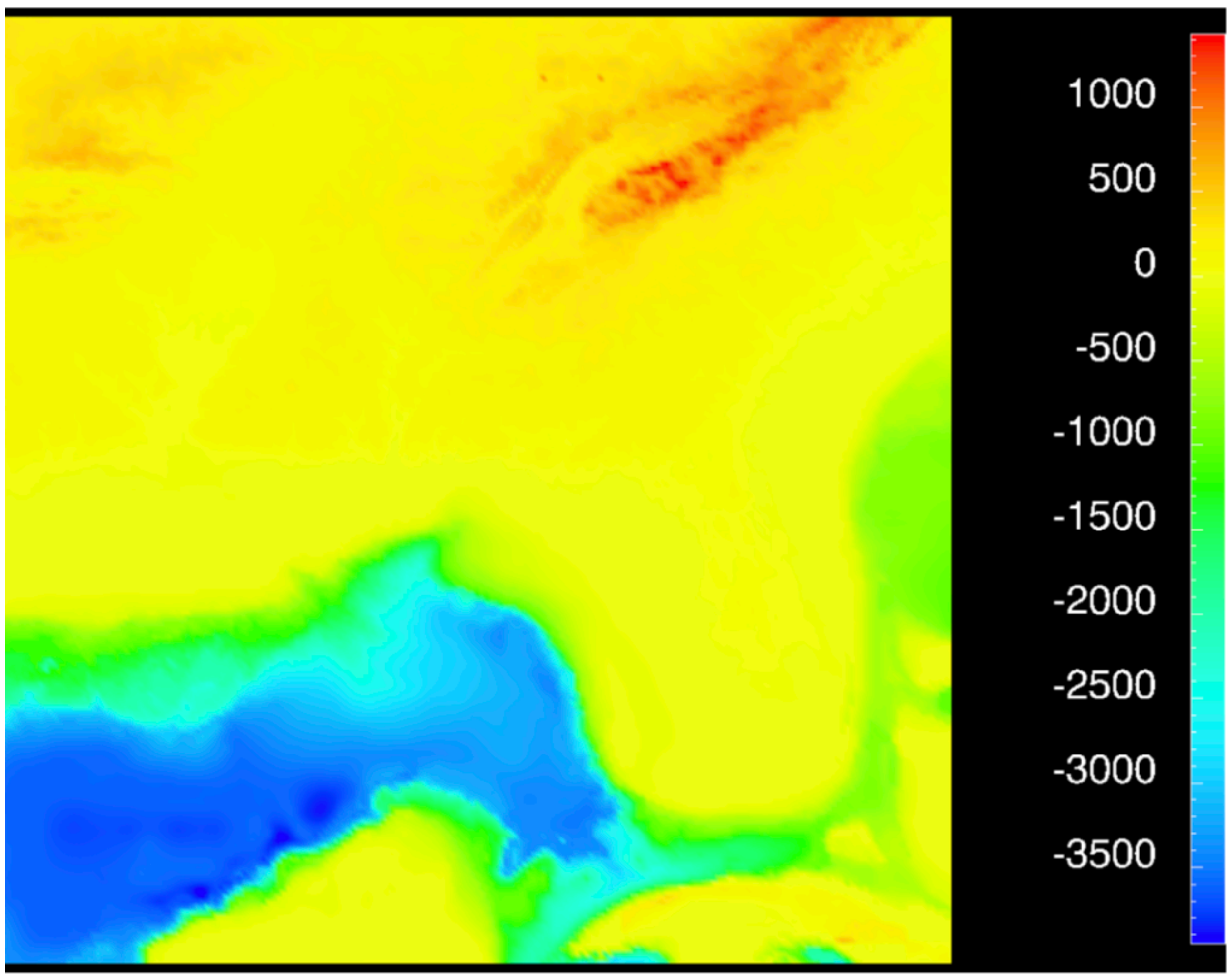


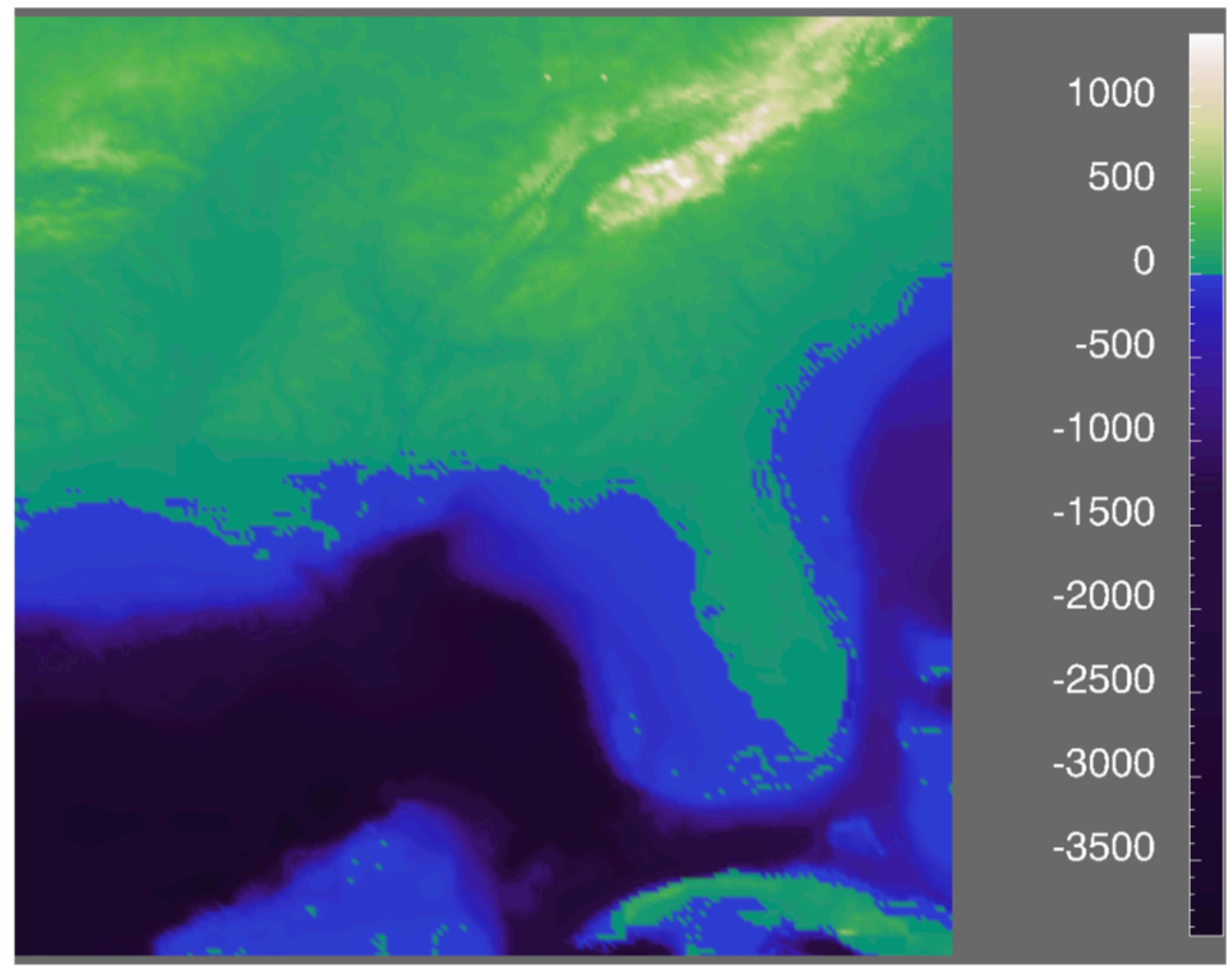
[Munzner 2014]

On Color

- “Get it right in black and white” - Maureen Stone
- Lots of different color spaces: RGB, HSL, etc..
- For visualization design: use a *perceptually uniform color space*.
- LAB: lightness (L), pure color / saturation (AB)
- HCL: hue (H), chroma (C), lightness (L)

Color for Quantitative Data

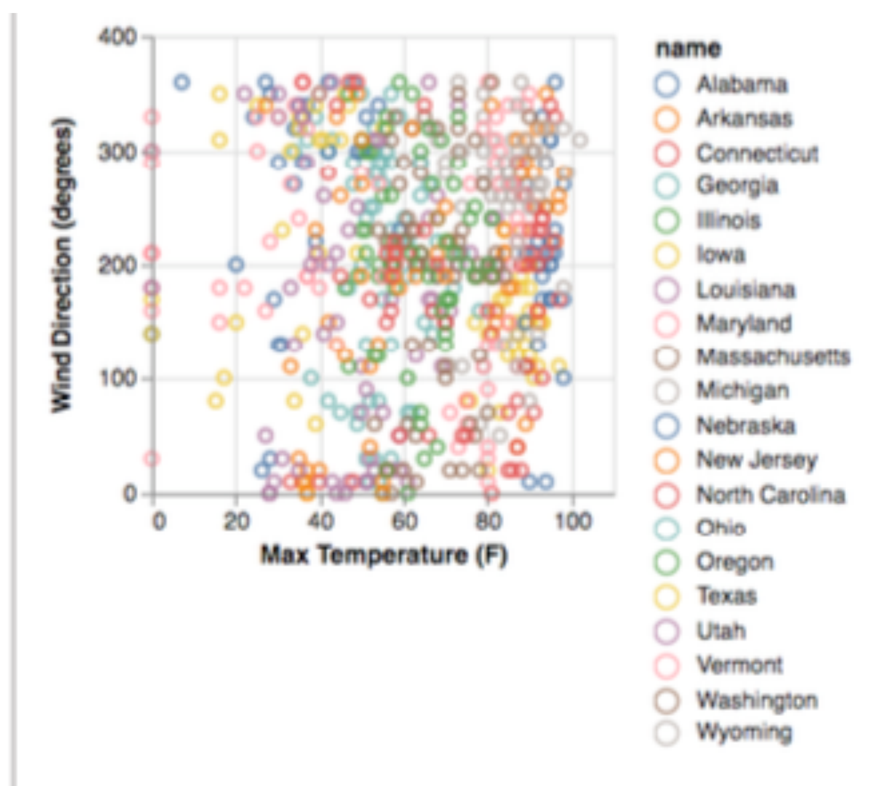




[Rogowitz and Treinish 1998]

Color for Categories

- Can only use around 5-6 colors, without interference [Healey 1996].
- Consider the following scatterplot

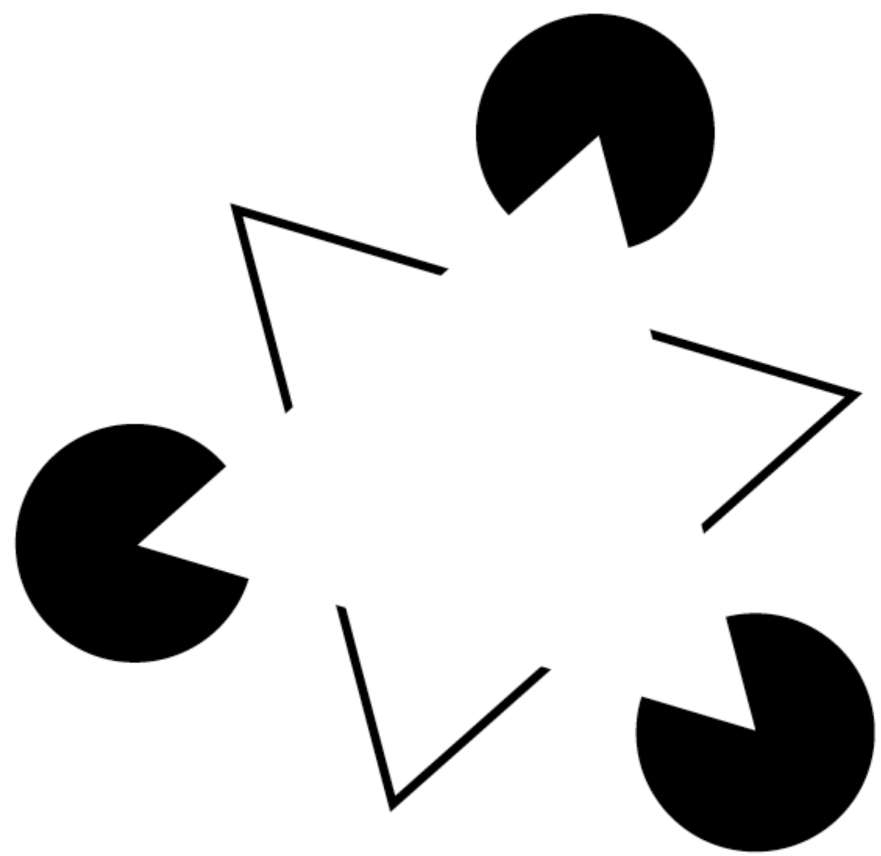


Gestalt Principles

- Visualization is a two-way street:
  - We (the vis designer) bring something to the table.
  - The human (end user) brings their *prior experience*.
- Design should take such prior experience into account!
- What is prior experience? **Gestalt laws**.

Closure

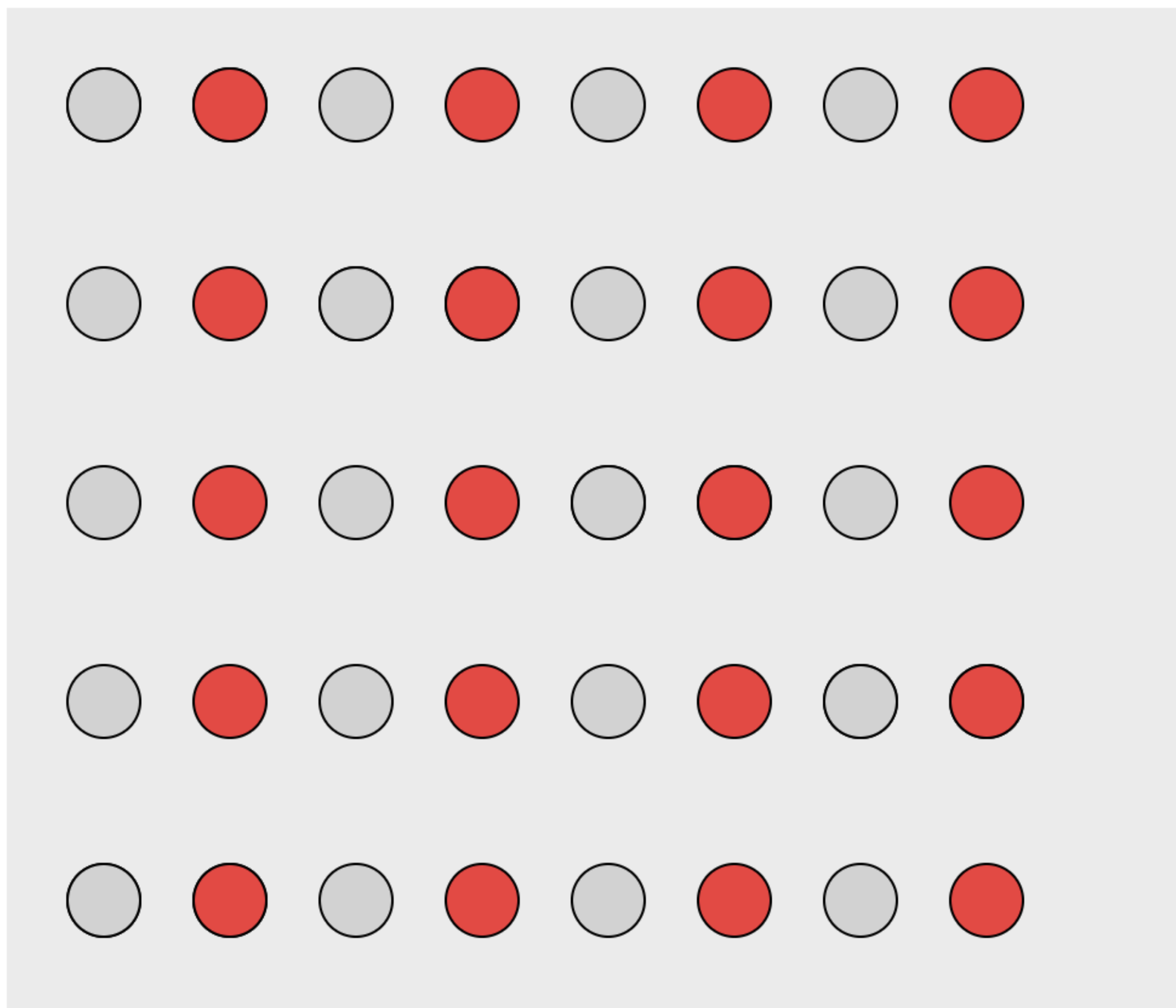
- We can complete incomplete shapes



- Implications: visualizations can be unintentionally misleading! Conversely: sometimes only necessary to show sparse set of marks to convey trend (dot plot)

Similarity

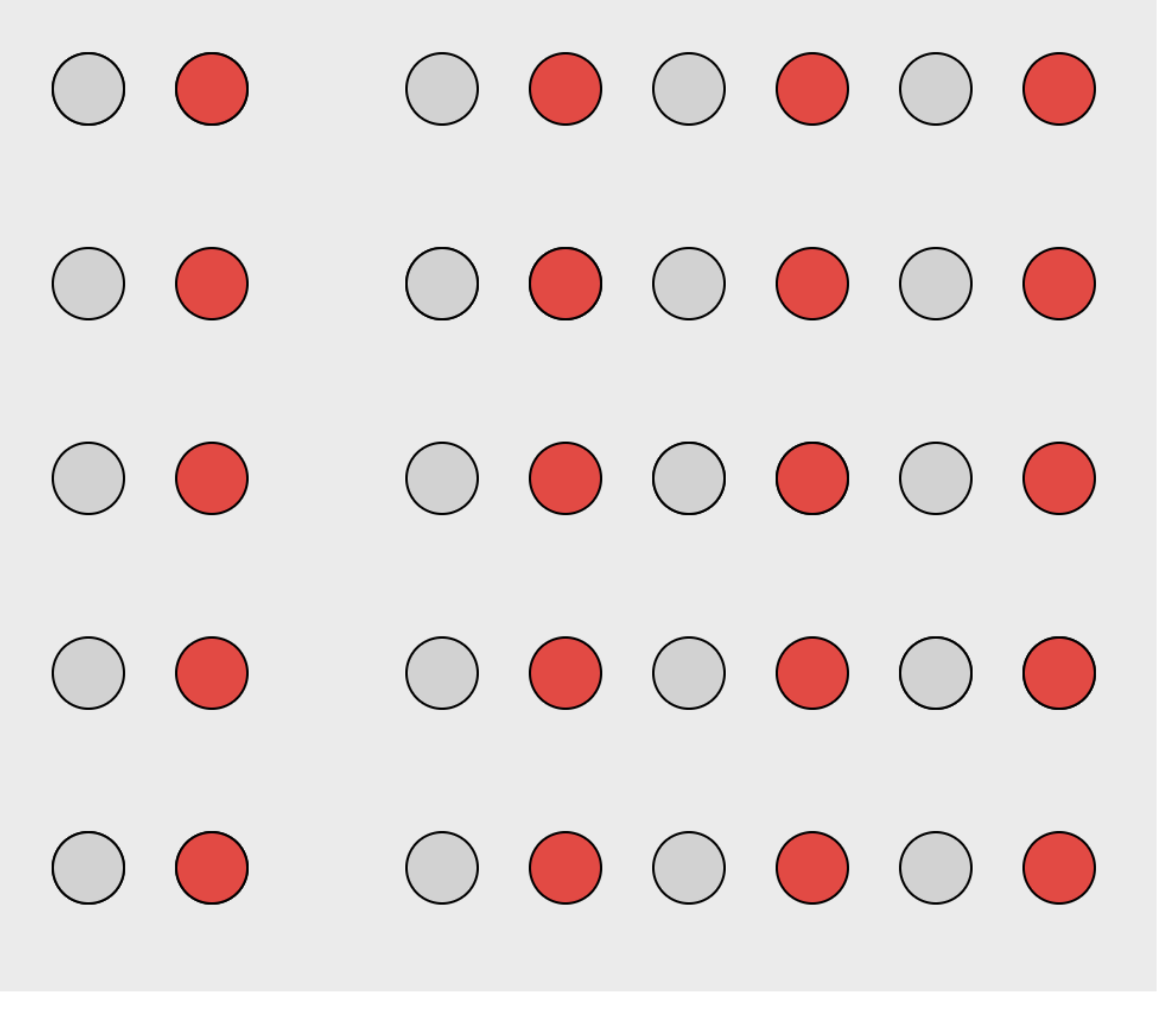
- Elements with the same visual properties considered to be grouped.



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Proximity

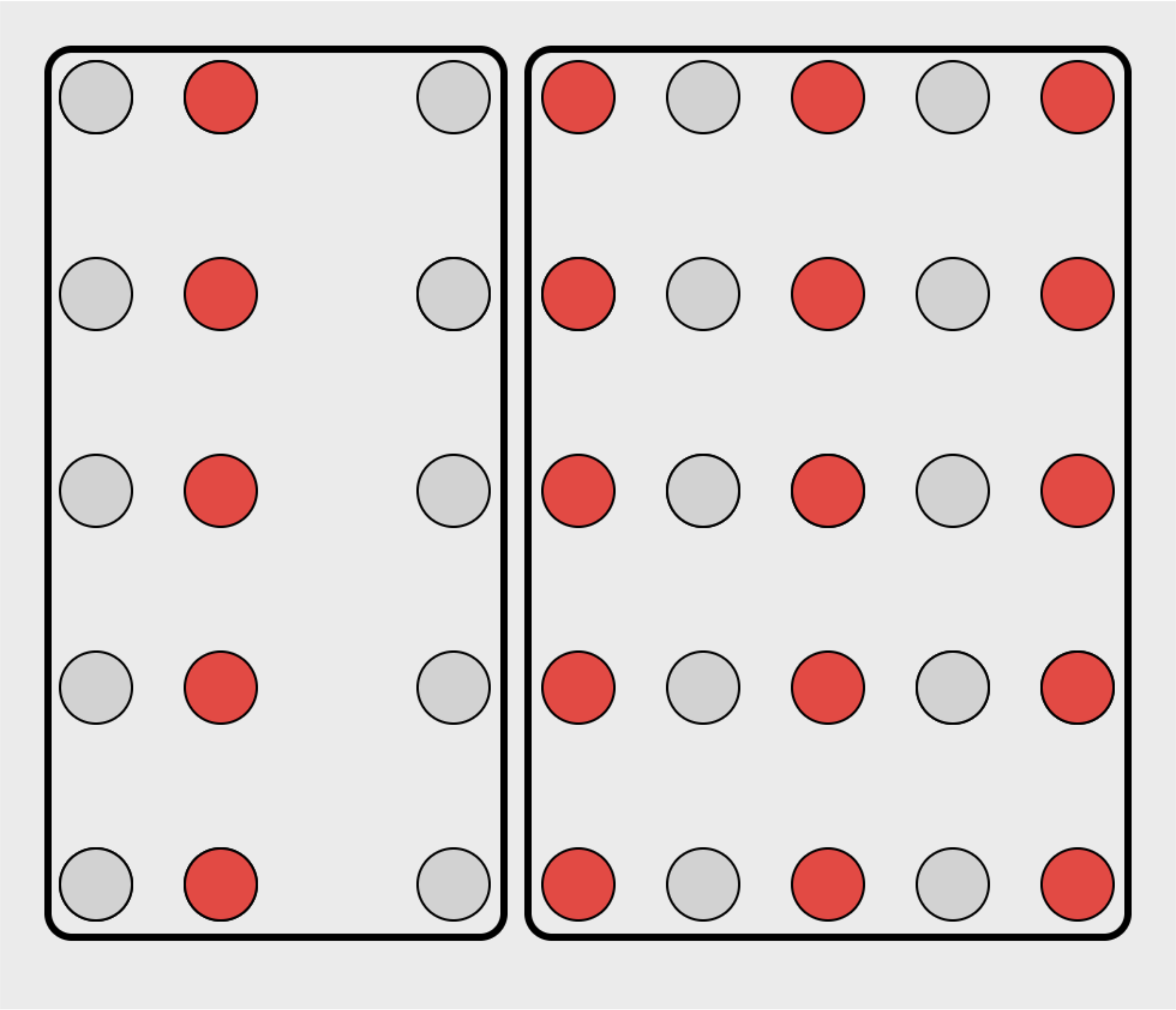
- Elements that are of close spatial proximity are somehow grouped. (we will repeatedly see this in dimensionality reduction, for better or worse)



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Enclosure

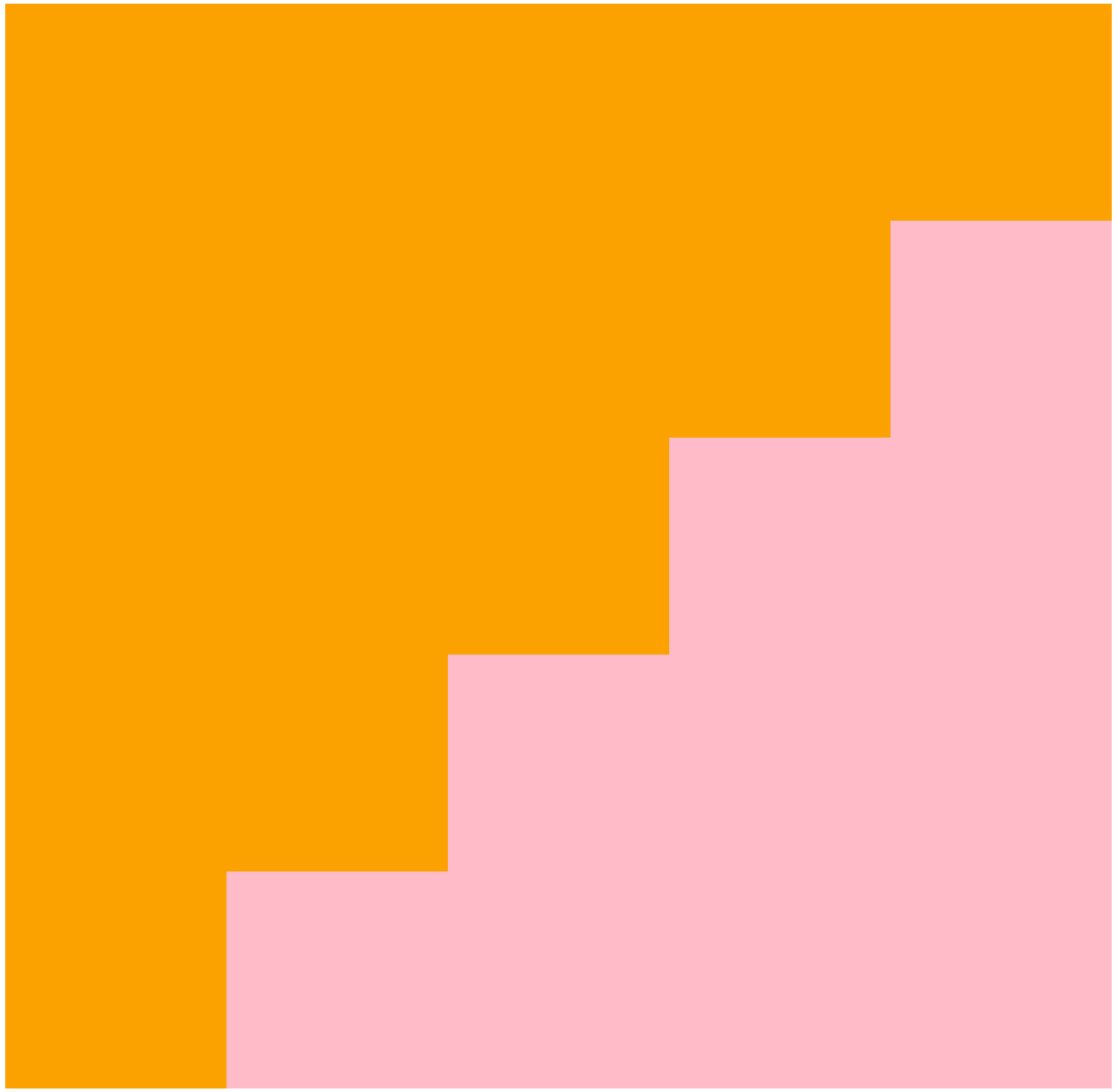
- Explicit visual encoding of enclosure also depicts grouping.



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Figure/Ground

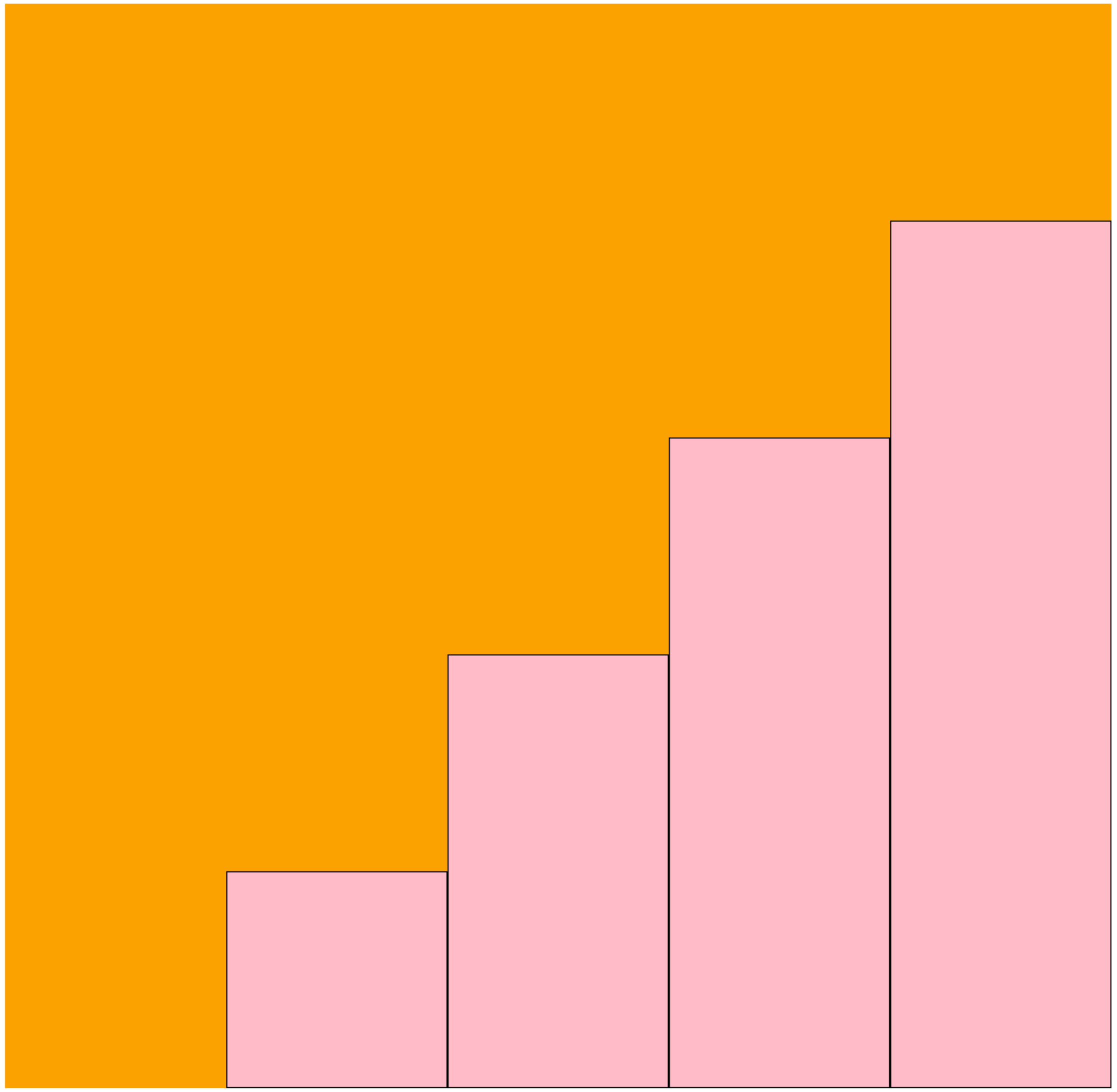
- Foreground from background



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Figure/Ground

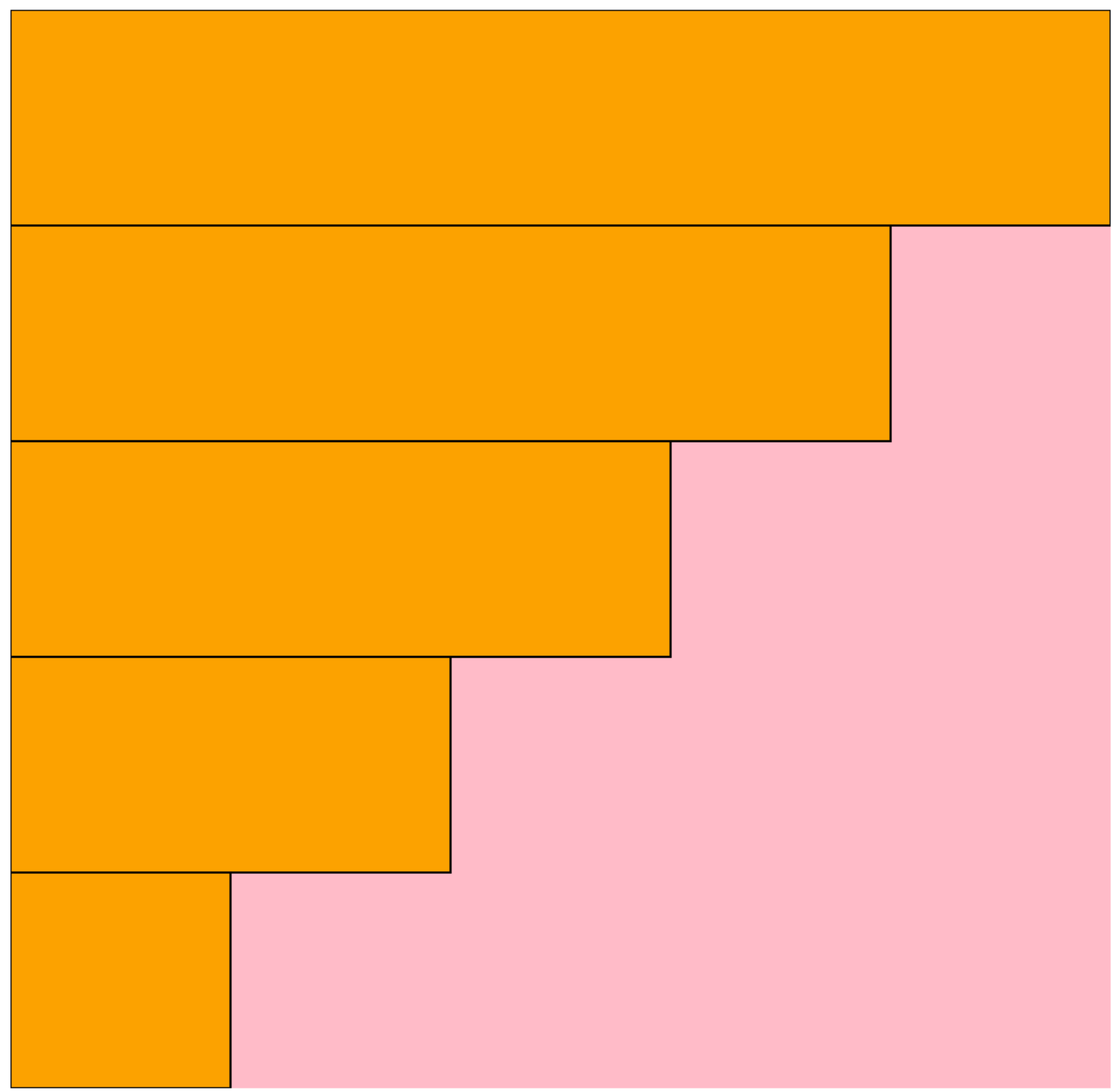
- Foreground from background



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Figure/Ground

- Foreground from background



[Meeks, https://emeeks.github.io/gestaltdataviz/]

Graphical Integrity

- From Edward Tufte:
  - Representation of numbers should match the true proportions.
  - Labeling should be clear and detailed.
  - Design should not vary for some ulterior motive, only show data variation.
  - The number of dimensions represented should be the same as the number of dimensions in the data.
  - Representations should not imply unintended context.

Data-Ink Ratio

- (more Tufte)

$$\frac{\textbf{Data-ink}}{\textbf{Total ink used to print the graphic}}$$

- Let’s illustrate this via examples…