Lecture 2

Creating Visualizations: What & How

CS5044 – Information Visualization

University of
St Andrews

where do I find lecture/exercise materials?

- https://studres.cs.st-andrews.ac.uk/CS5044/

outline for today

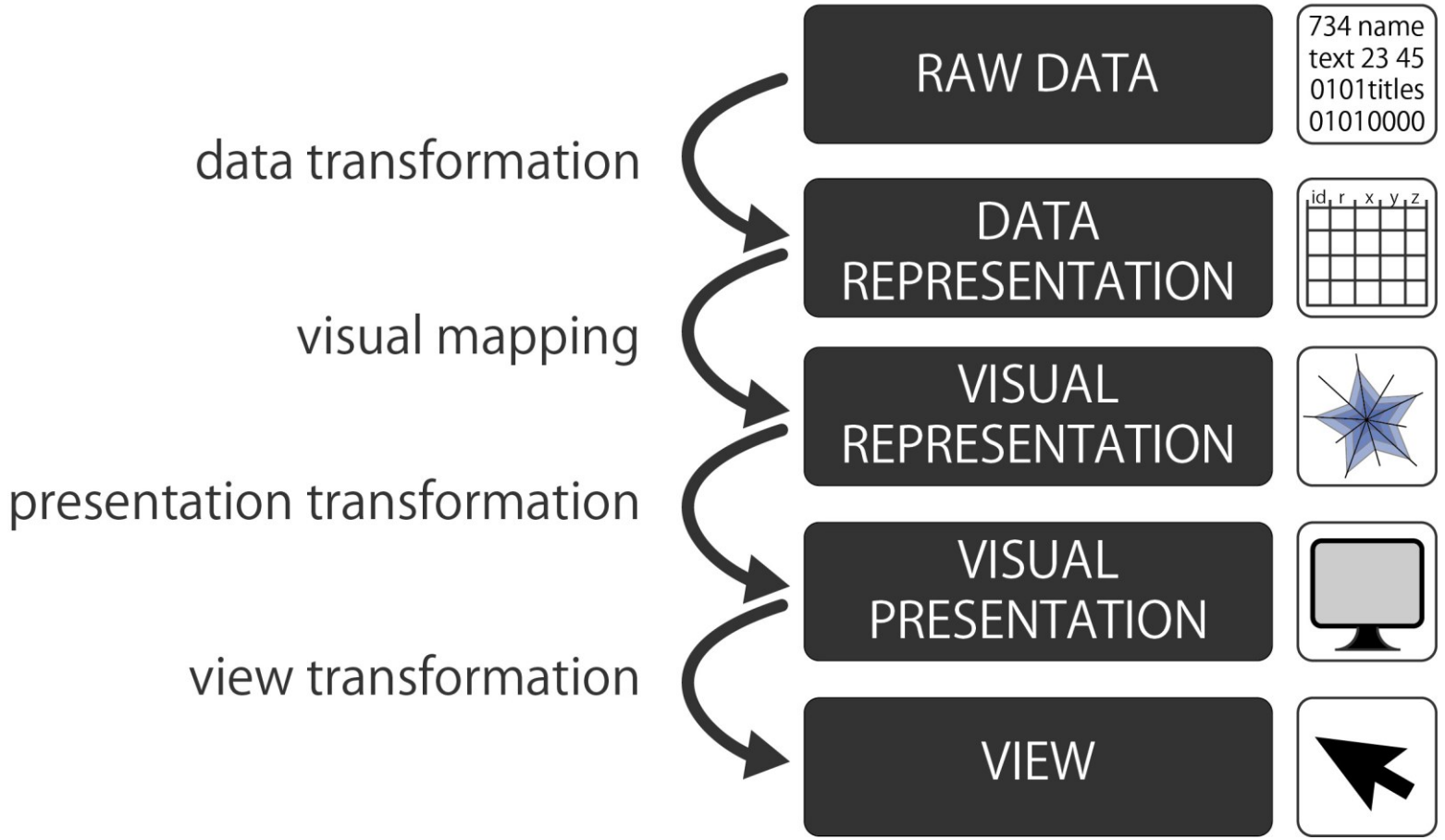
- Process of creating visualizations
- What? – data types
- How? – visual marks and variables

by the end of this lecture you should know

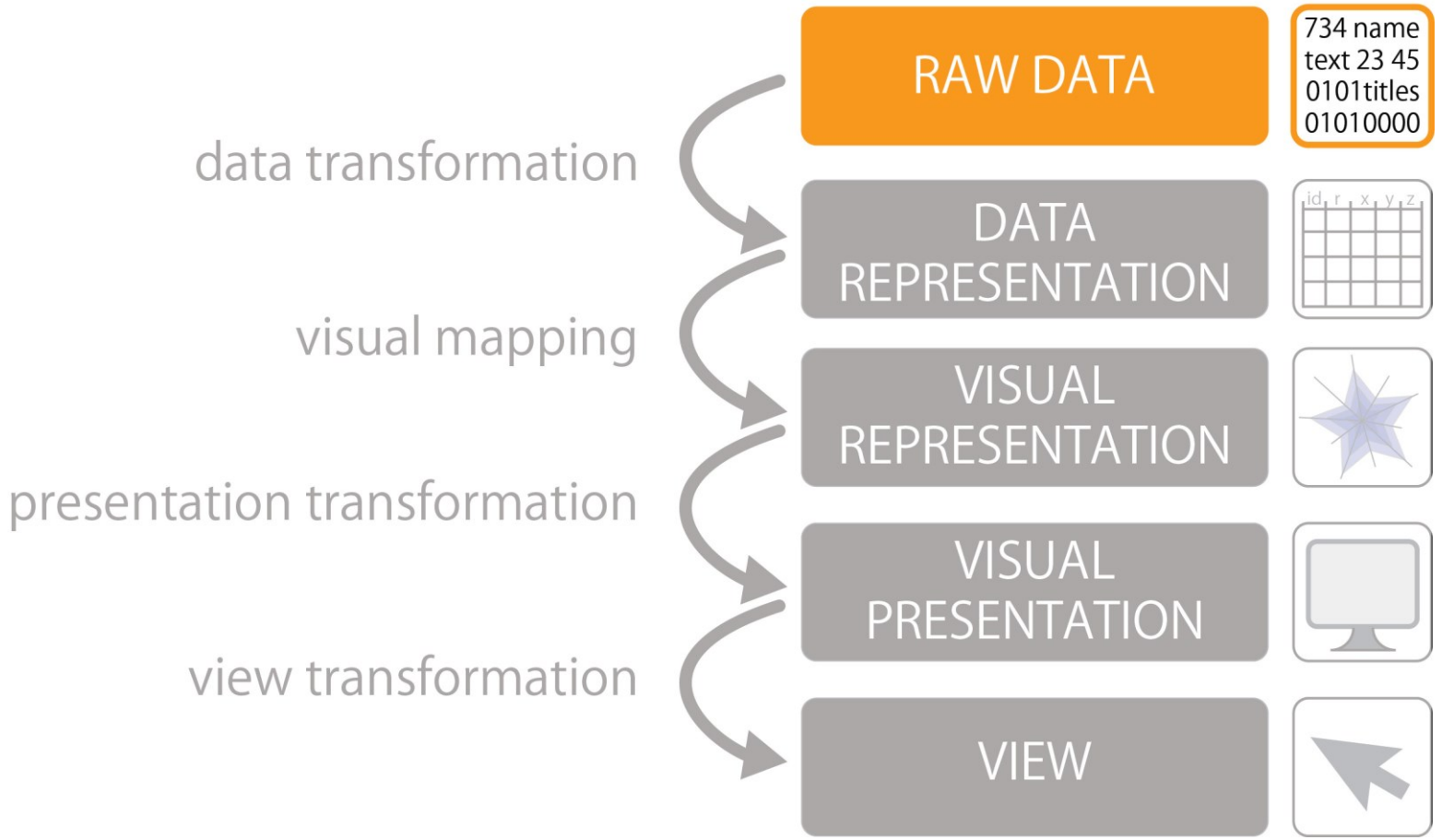
- The step-by-step process of creating an information visualization (theory)
- How to classify data types in the context of visualization
- Why the consideration of data types is useful when creating visualizations
- How to apply visual variables to create a visualization

the visualization process

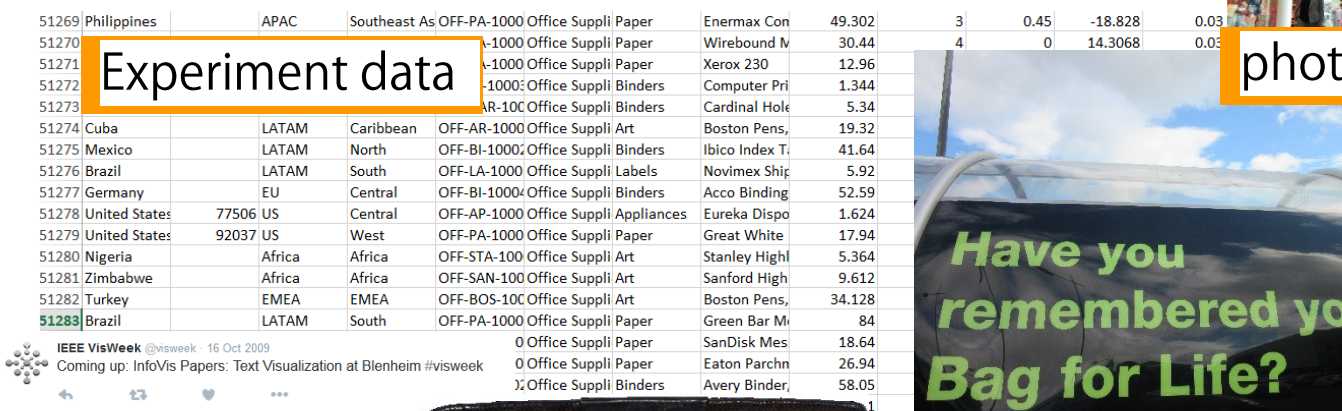
visualization pipeline



visualization pipeline

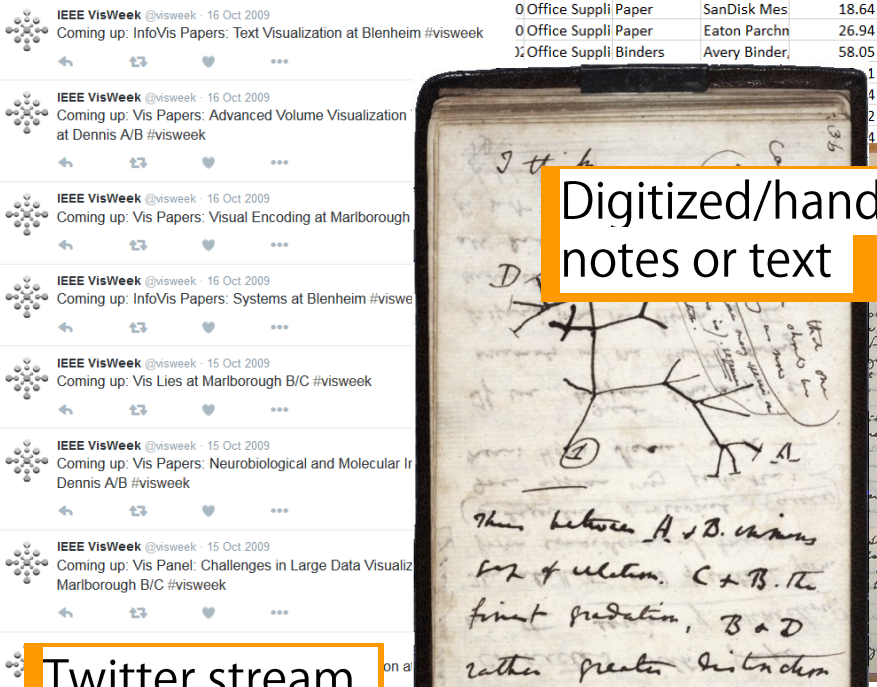


example: raw data



Experiment data

photographs

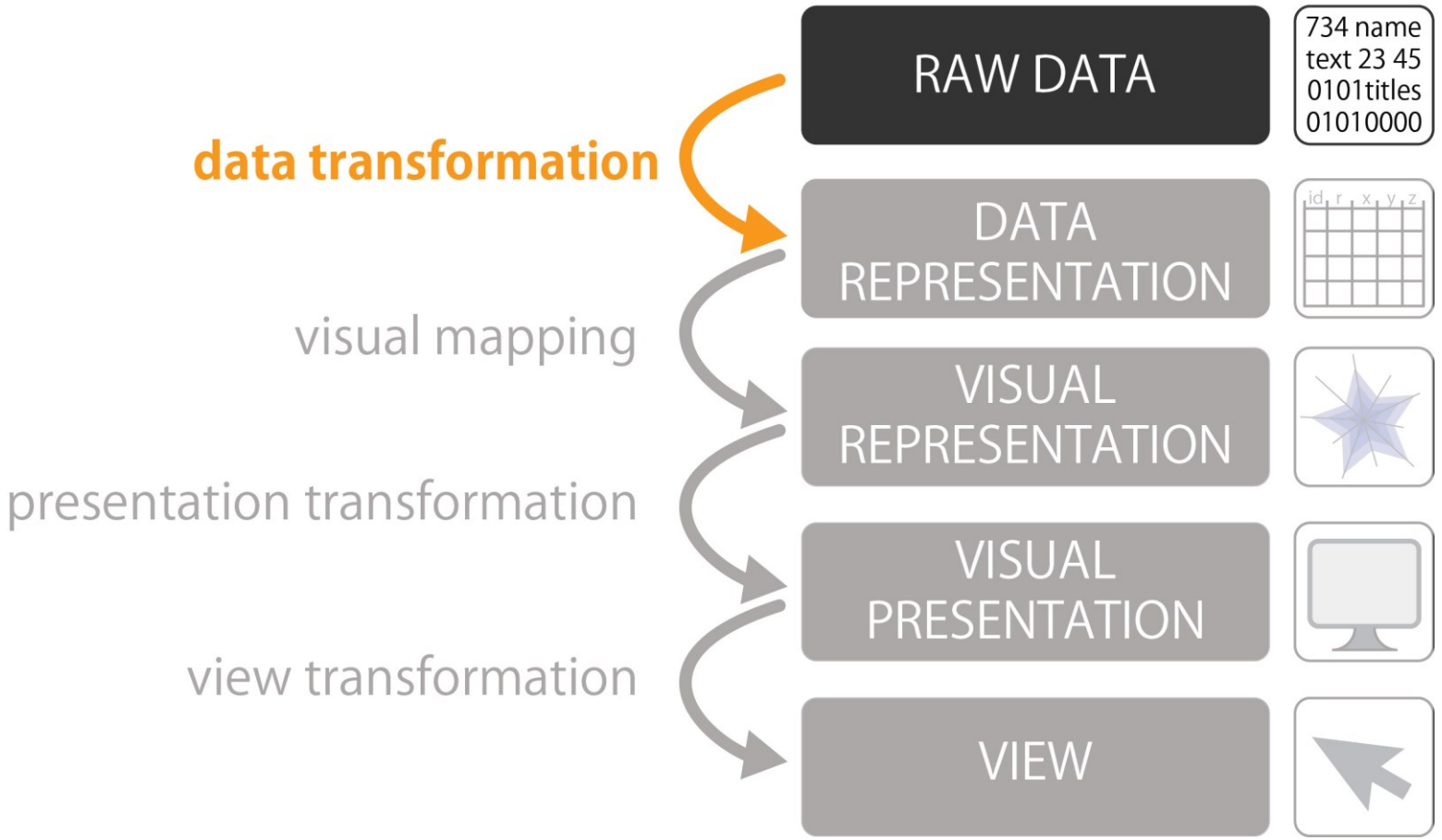


Twitter stream

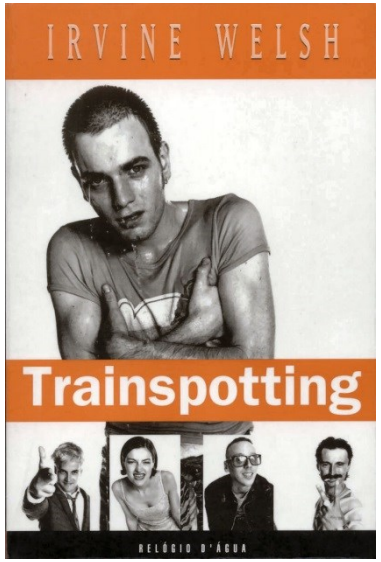
Digitized/handwritten notes or text

Personal log data

visualization pipeline



example: data transformation



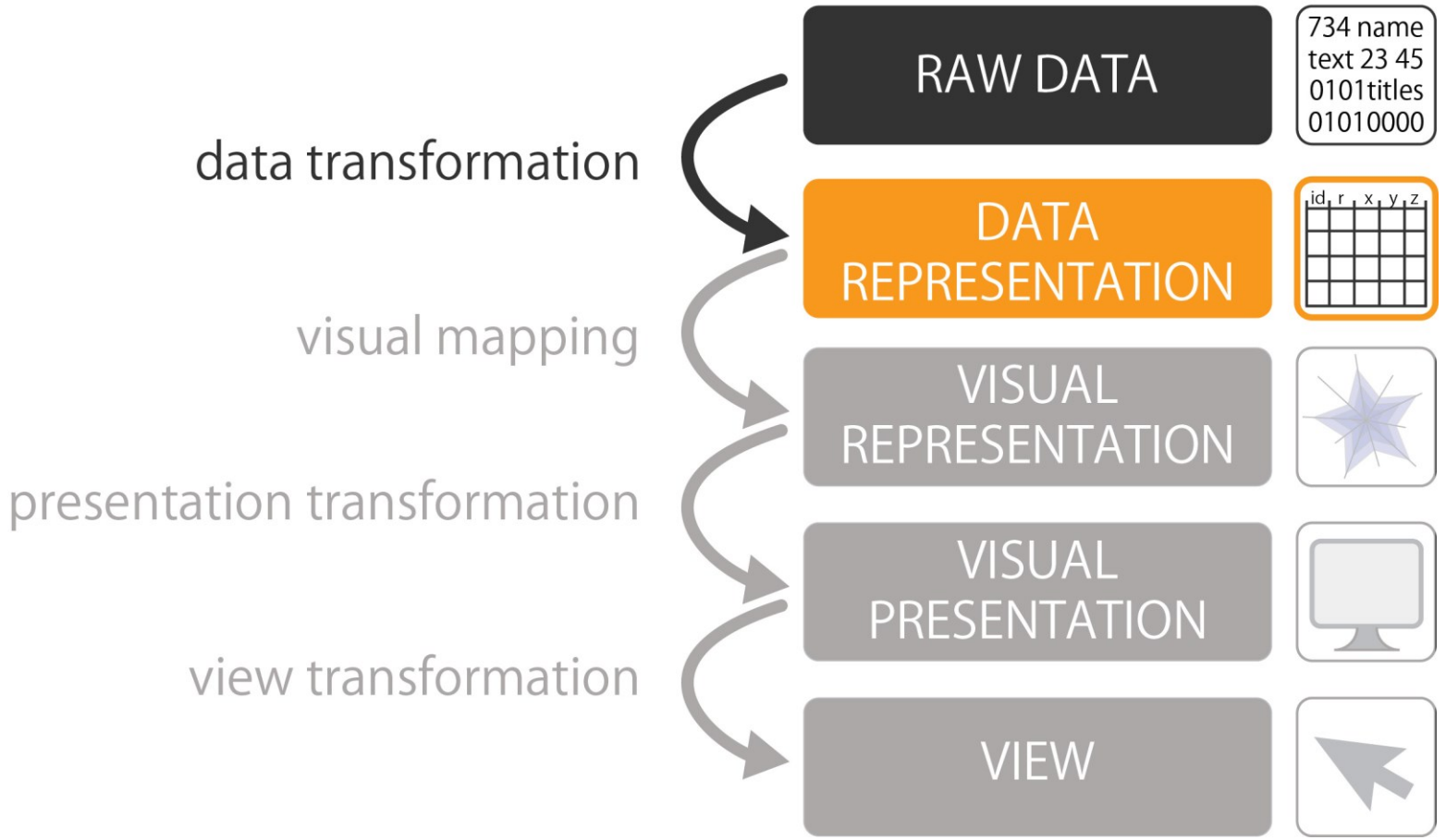
‘Mother superior' wis Johnny swan; also kent as the white swan, a dealer whae wis based in Tollcross and covered the Sighthill and Wester Hailes schemes. ah preferred tae score fi swanney, or his sidekick raymie, rather than seeker n the muirhoose-leith mob, if ah could.

Irvine Welsh 1993

Trainspotting

Author	Title	Publication year	Locations	Text
I. Welsh	Trainspotting	1993	Tollcross, Sighthill, Wester Hailes	Mother superior' wis Johnny swan; also kent ...
…	…	…	…	…

visualization pipeline



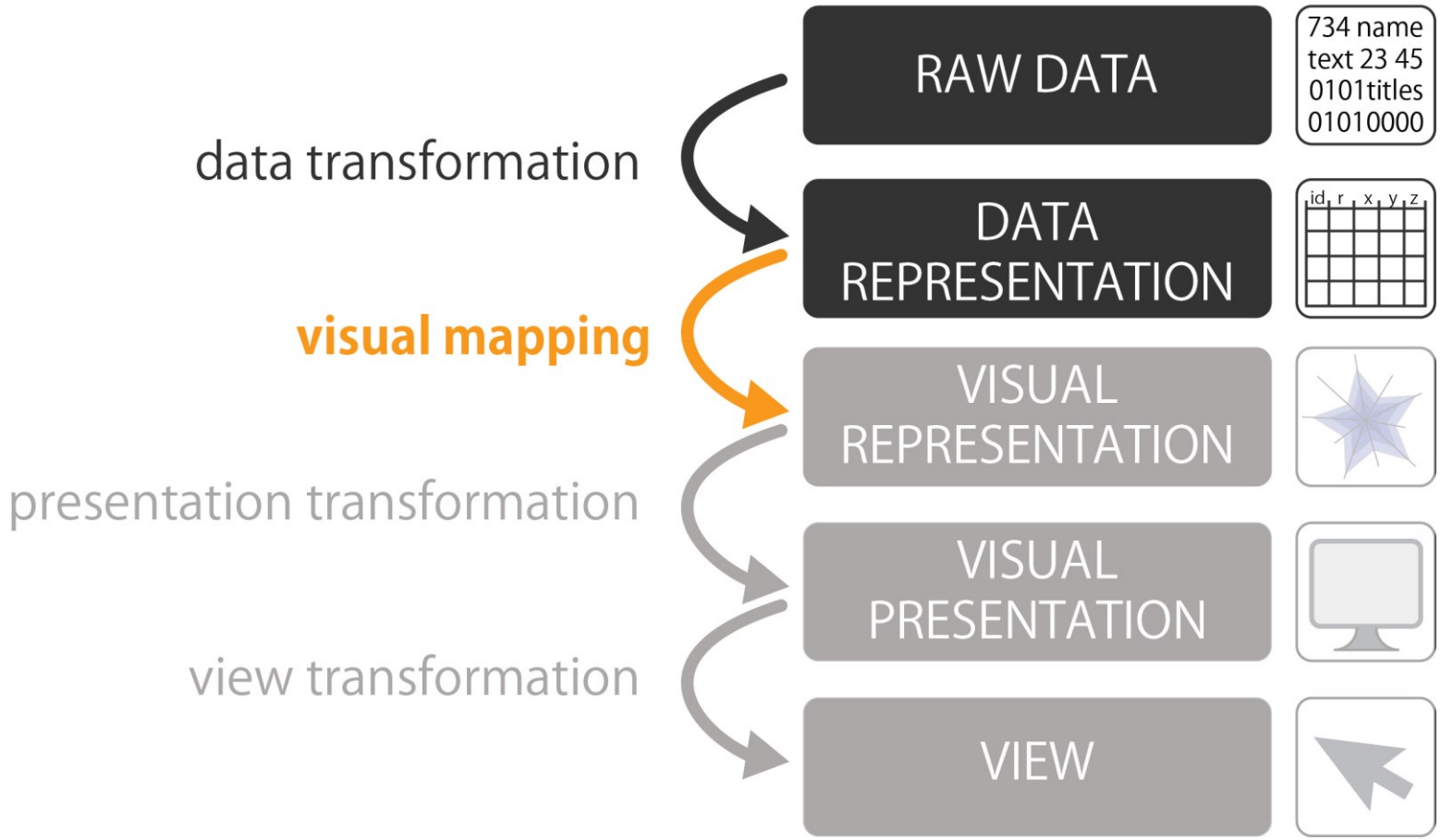
data representation

analytical abstraction

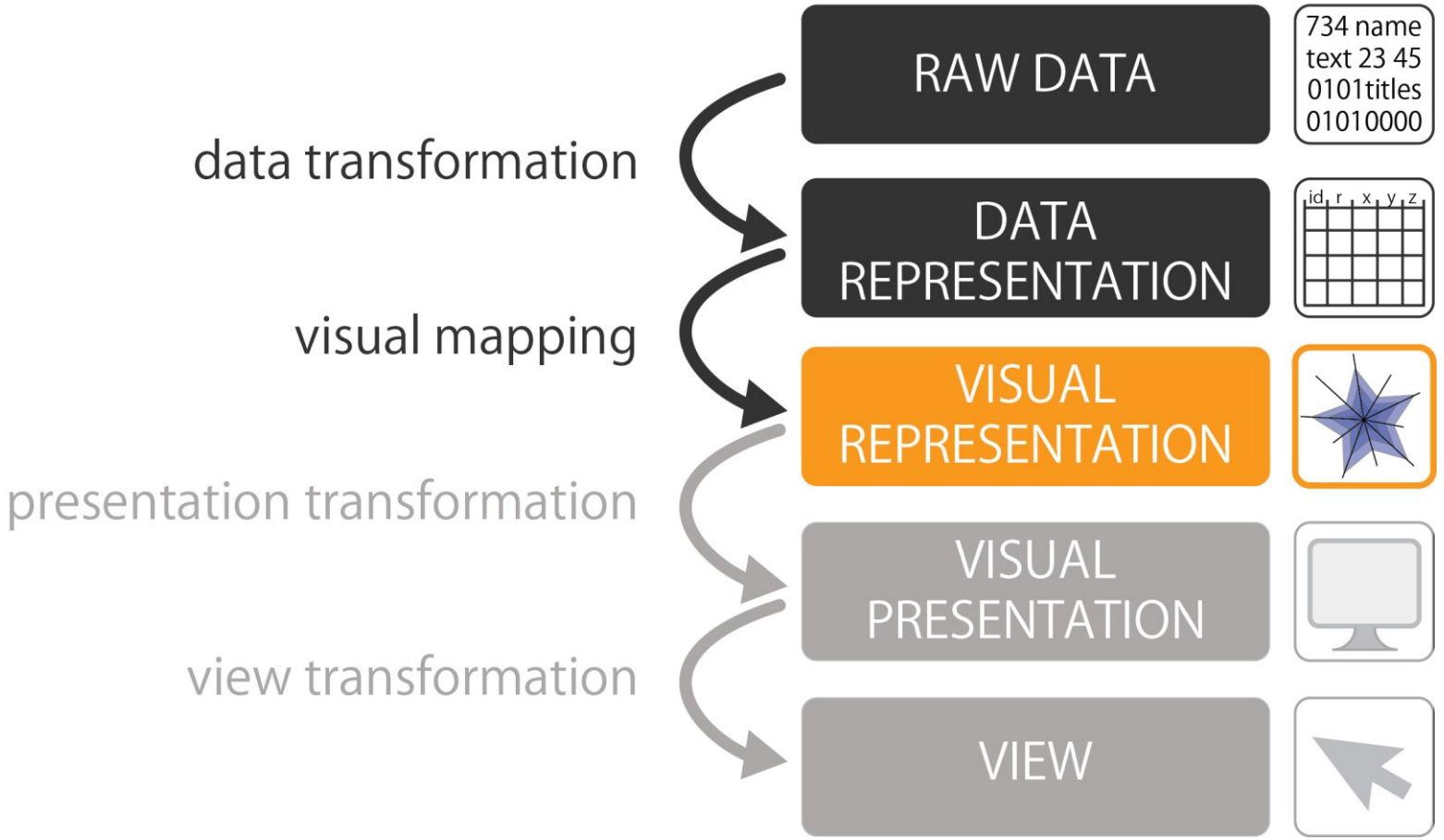
- Data in a structured, tidy, machine readable form, ready to be visualized
- Examples of data representations
  - Relational tables with metadata
  - Adjacency matrix (e.g., for graph structures)
  - Vector space model
  - ...

Author	Title	Publication year	Locations	Text
I. Welsh	Trainspotting	1993	Tollcross, Sighthill, Wester Hailes	Mother superior' wis Johnny swan; also kent ...
…	…	…	…	…

visualization pipeline



visualization pipeline

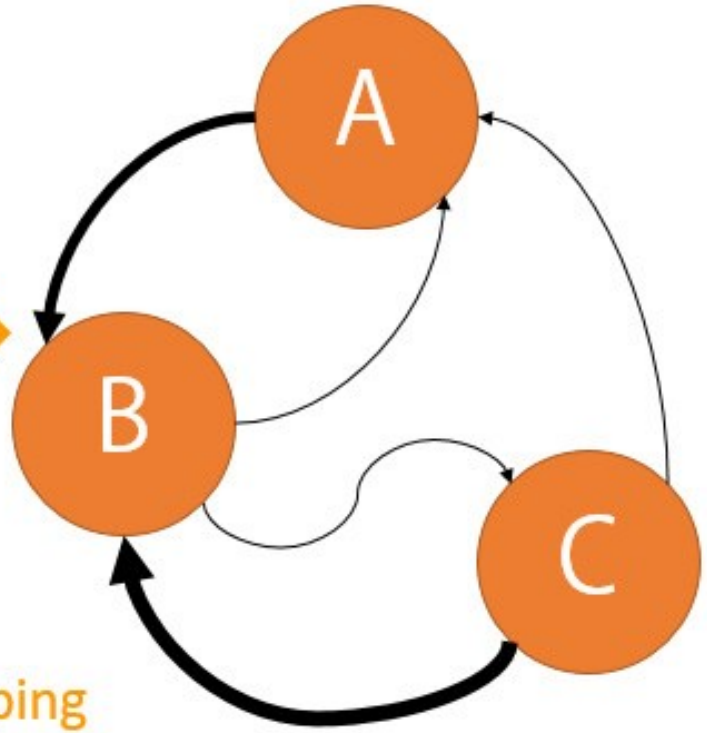


example: data representation

Data Represenatation

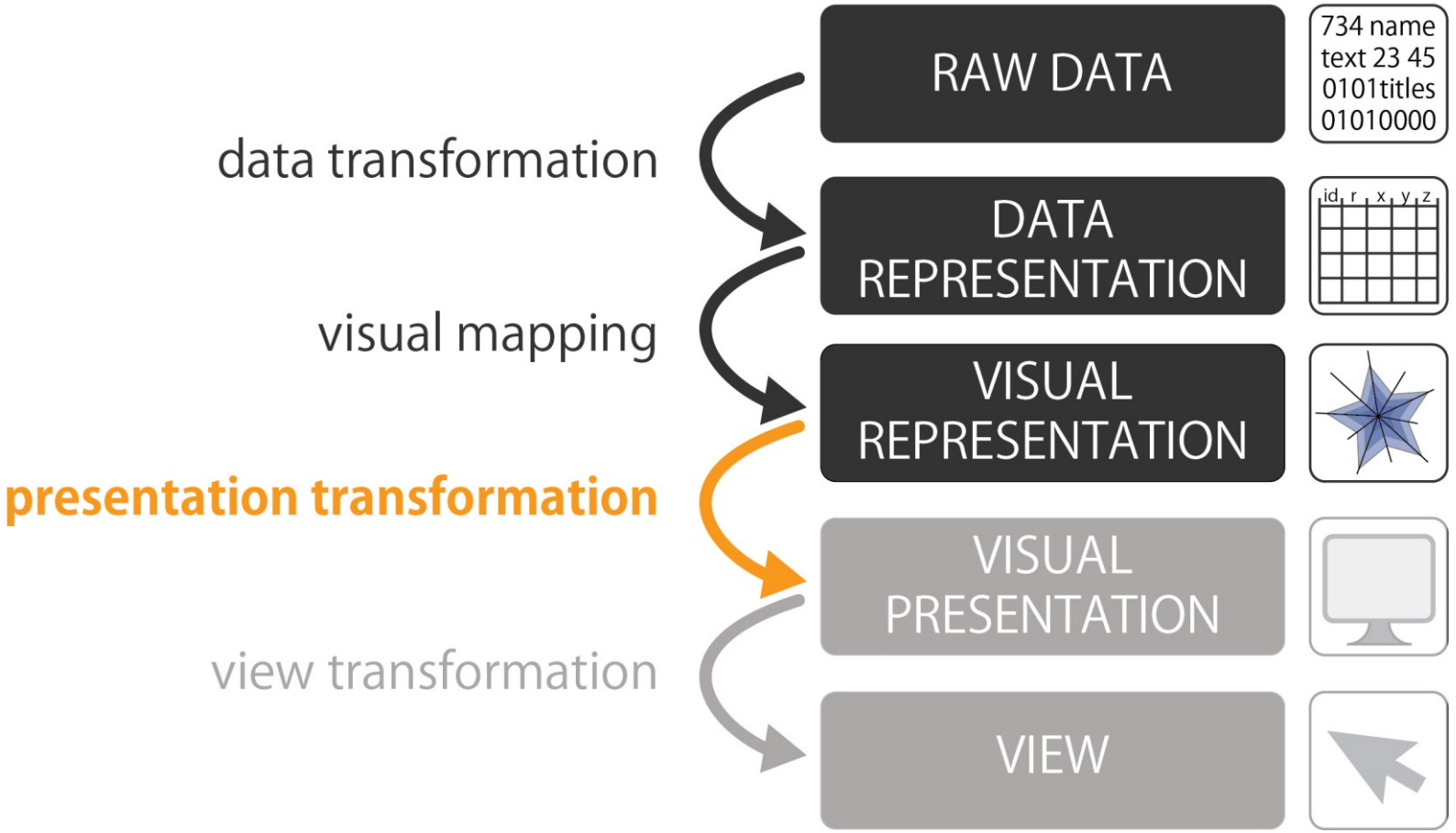
Visual Represenatation

	A	B	C
A	0	2	0
B	1	0	1
C	1	3	0



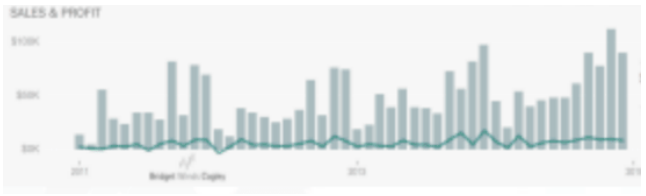
Visual Mapping

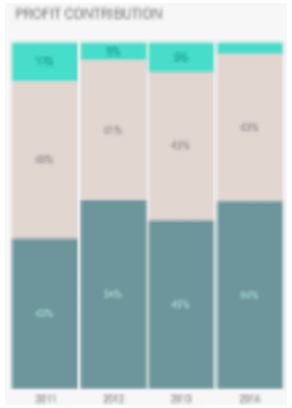
visualization pipeline

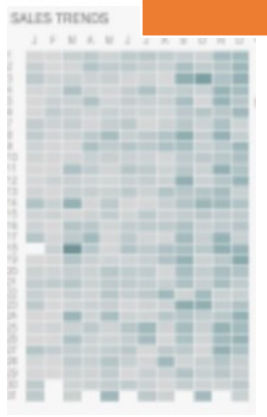


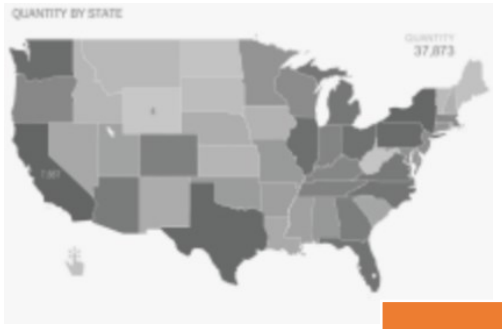
presentation transformation

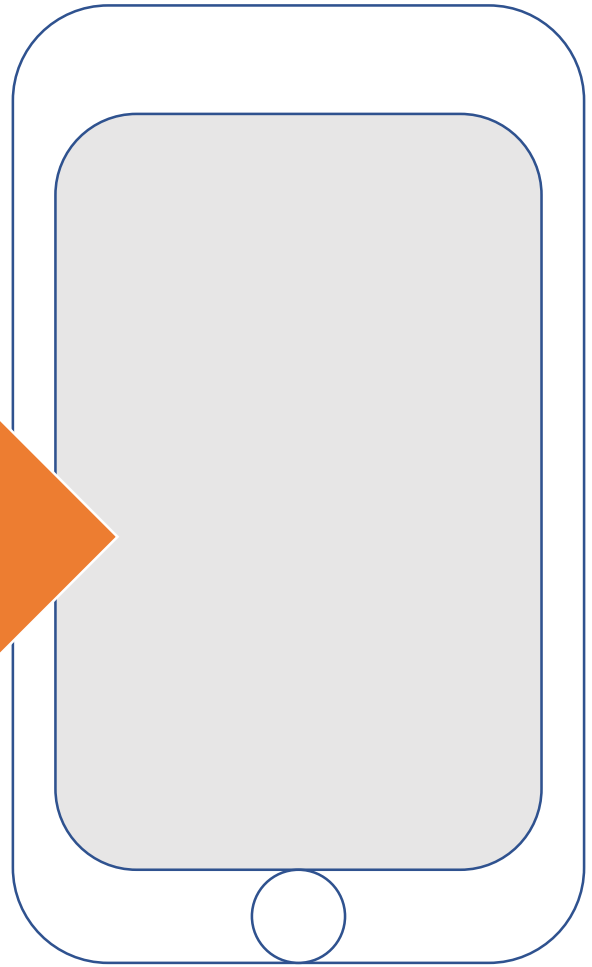
- Considering display resolution
- Layout: arrangement + size
- Considering possible interactions
  - Filtering
  - Searching
  - Linking of different visualizations



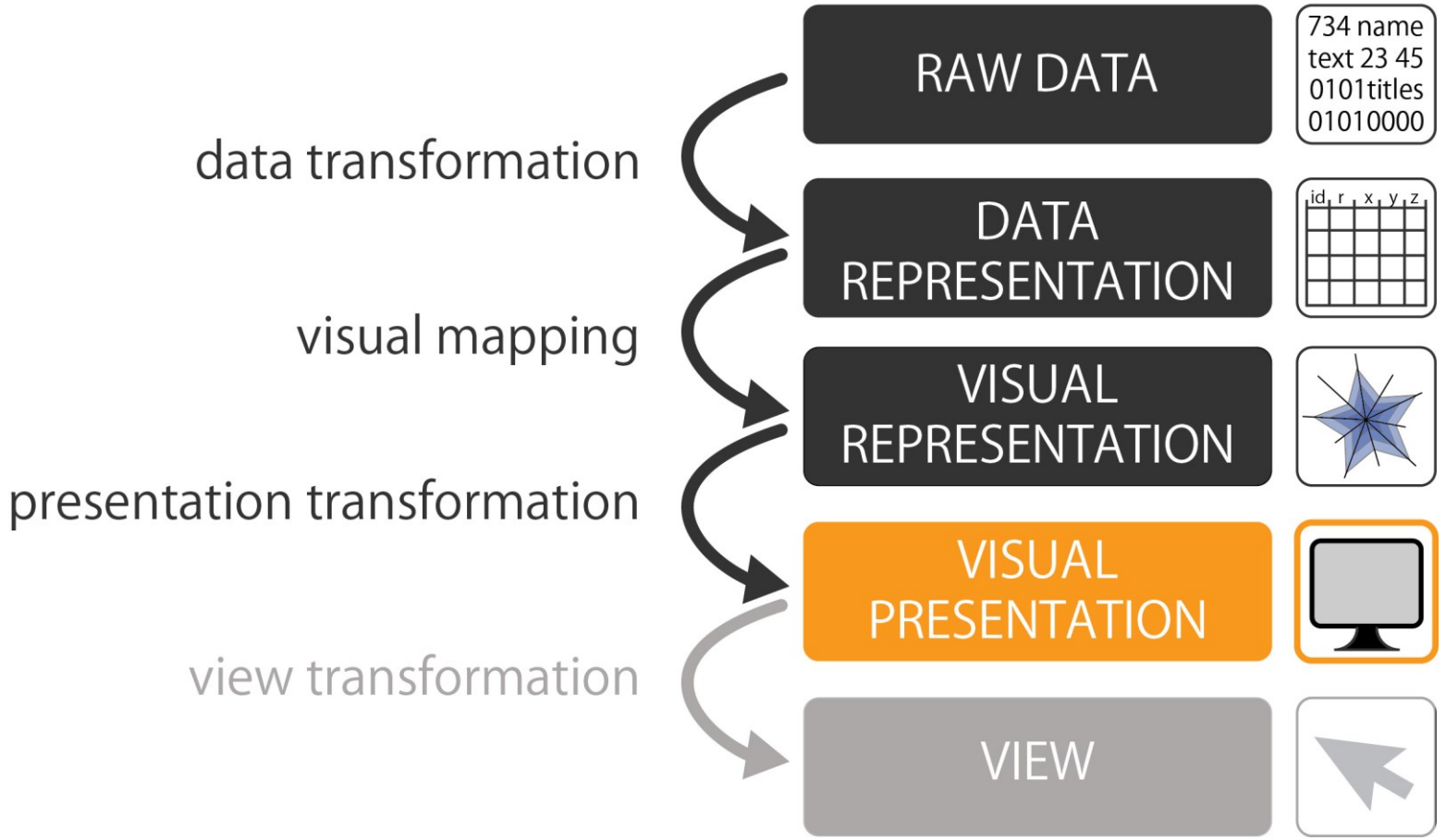




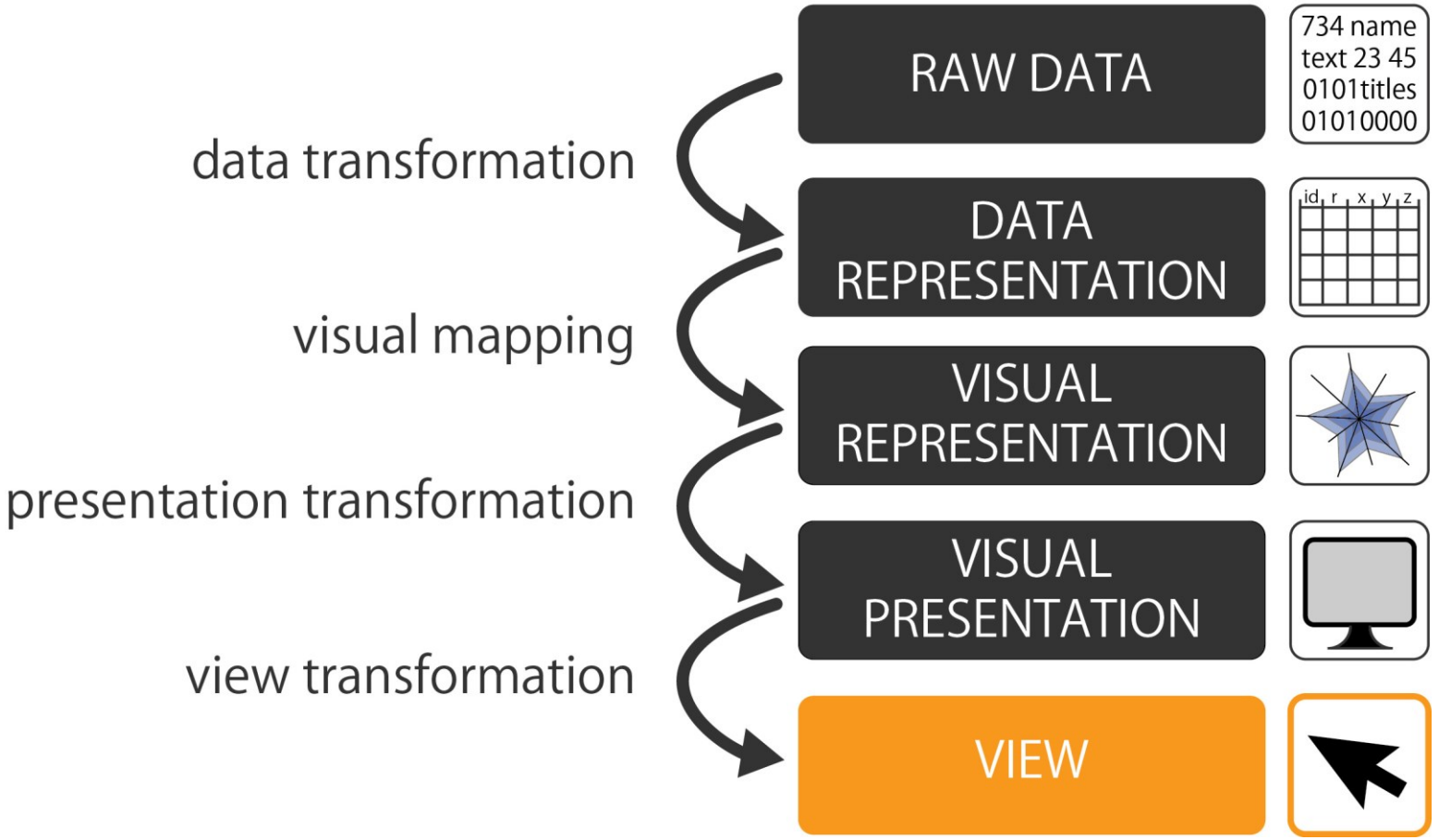




visualization pipeline

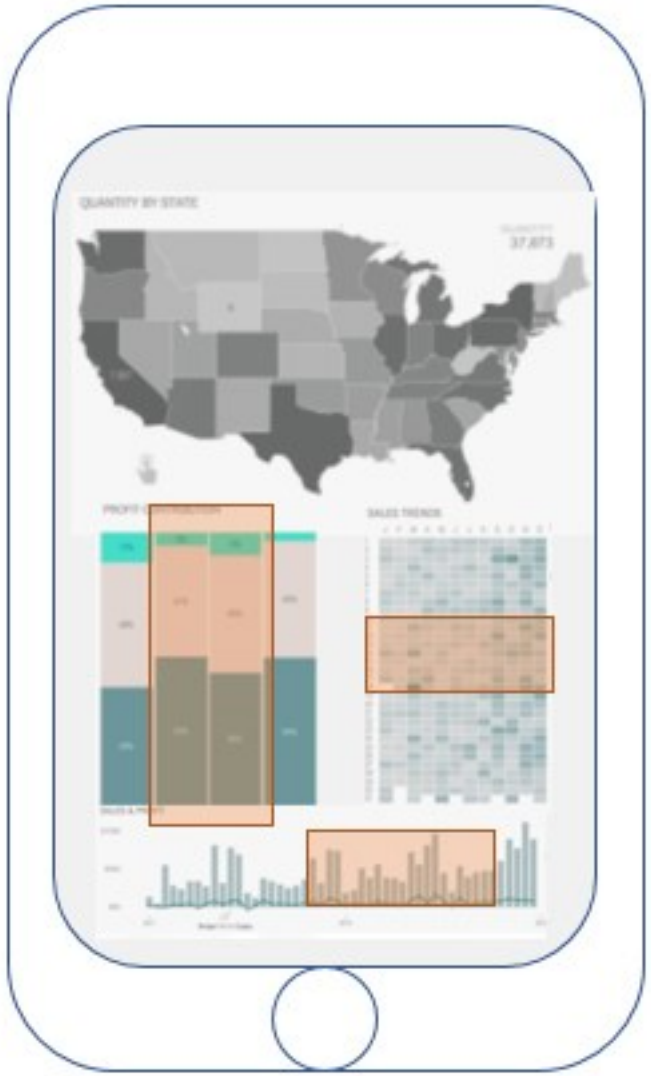


visualization pipeline



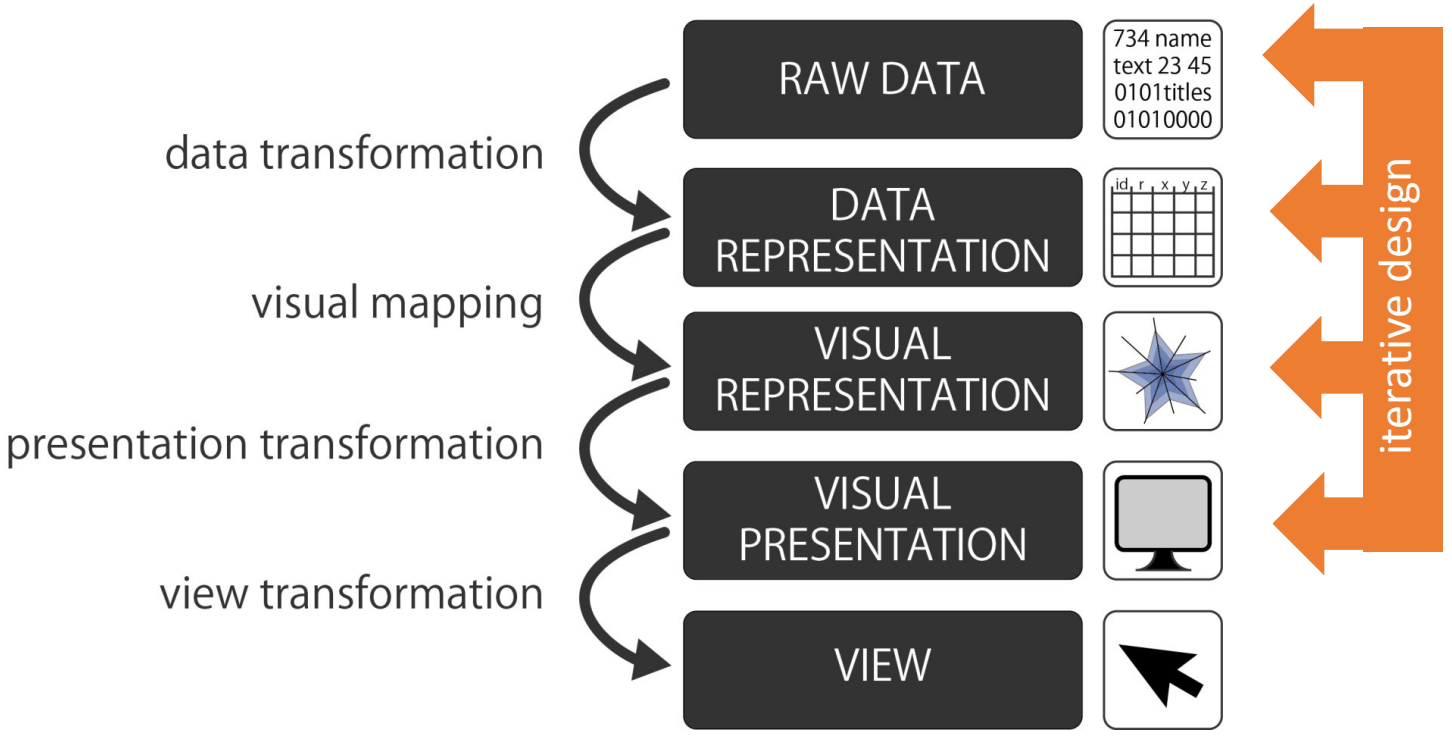
view

- The rendered visualization visible to the user
- Can change depending on interactions
  - Selecting
  - Filtering
  - Searching
  - Zooming
  - ...



visualization pipeline – design perspective

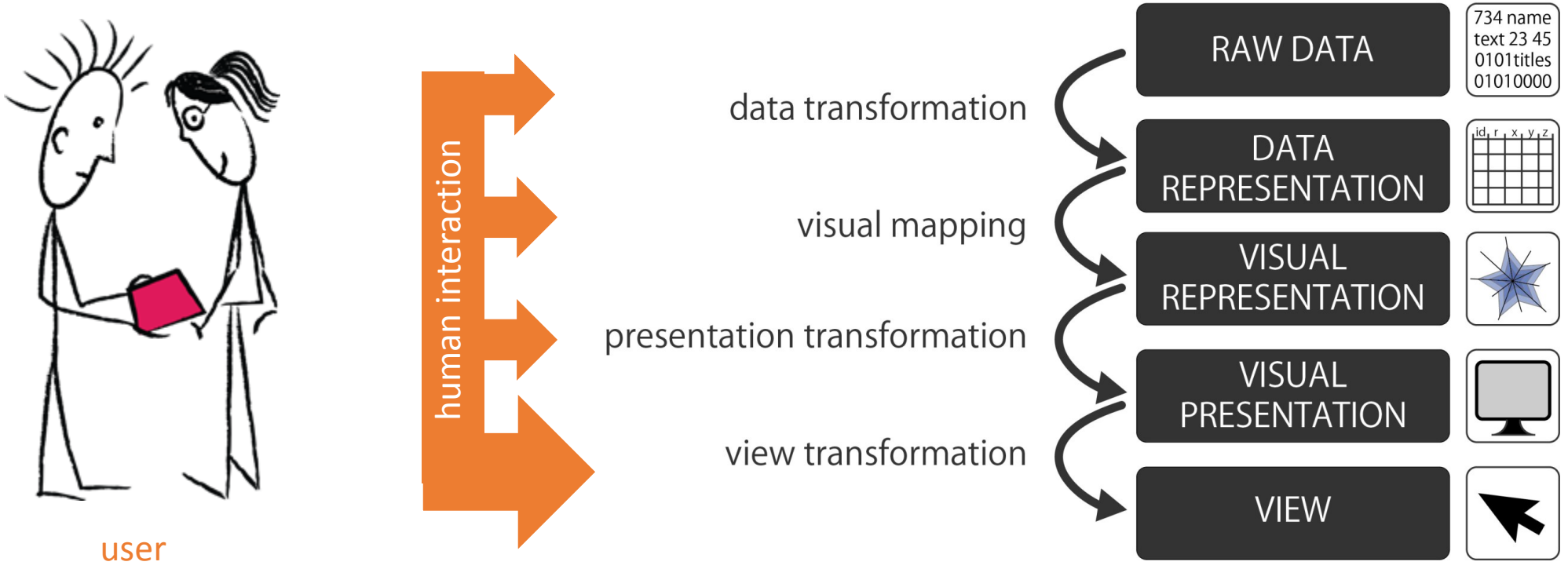
- Guides the visualization design process



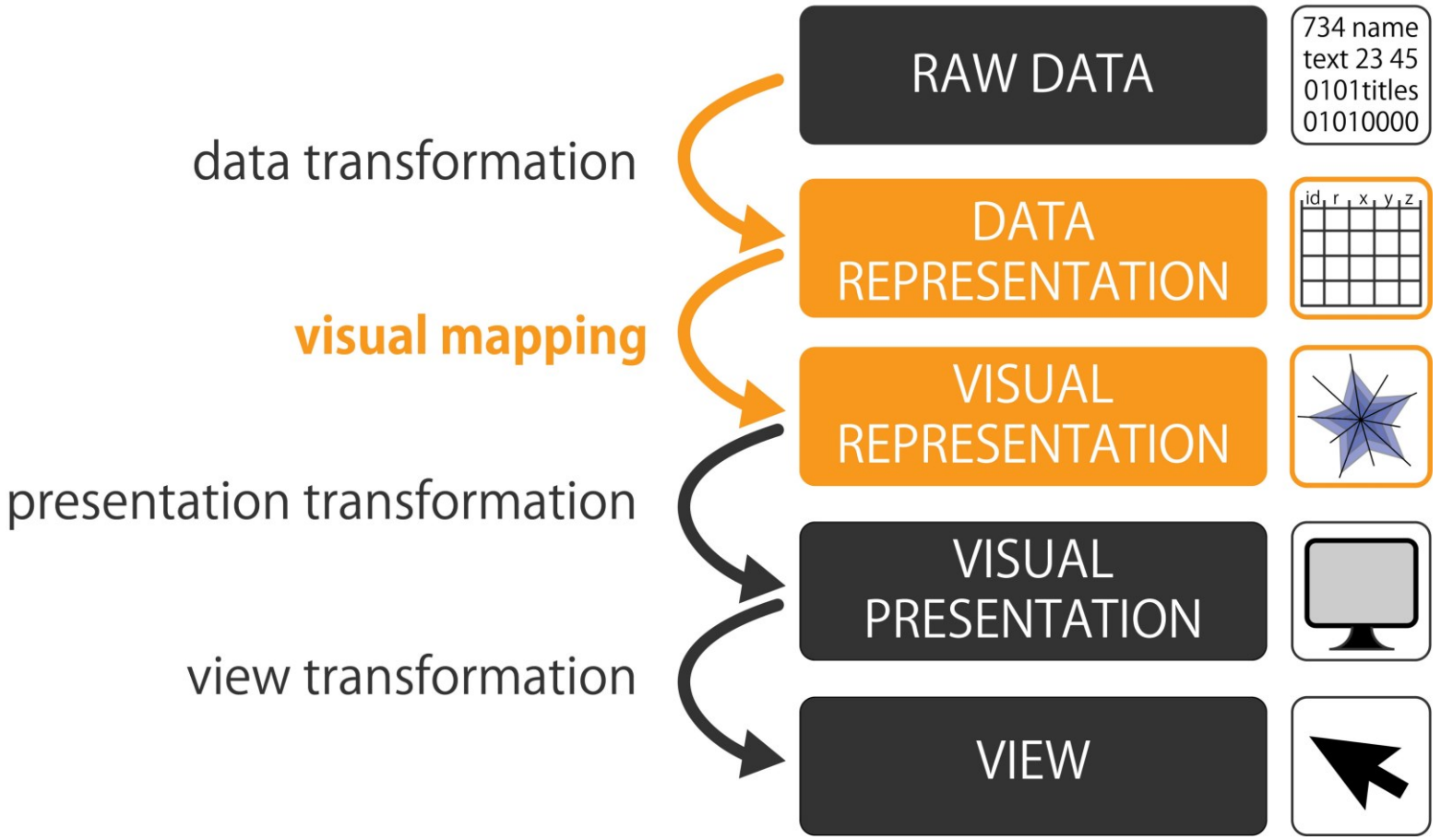
designer

visualization pipeline – user perspective

- Guides the exploration process

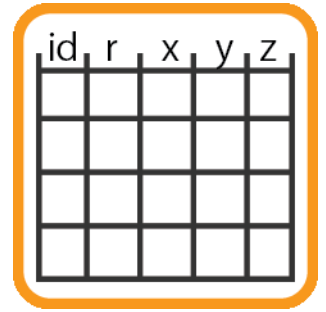


visualization pipeline



what can be visualized?

Data Abstraction



what to consider about the data?

- Data semantics
  - Real world meaning
  - What does a data point actually represent?
- Data types
  - Structural or mathematical interpretation of the data
- Why important?
  - Data semantics and type will influence our choice of visual representation

1	Apple	8	S	Banana
2	Basil	7	S	Pear

some terminology

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

some terminology

- Data attributes

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

some terminology

- Items or *data points*

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

some terminology

- Cells or *values* of an individual attribute/data point pair

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

some terminology

- Relations or links between items (or data points)

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

the dataset

- Collection of information/datapoints/items

- Table (tabular data)
- Networks and trees (hierarchical data)
- Clusters, sets, lists (rankings, timelines...)
- ...

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

	Frank	Ruth	Clara
Frank	0	1	0
Ruth	1	0	1
Clara	1	1	0

- Complex combination of data types are possible
- Can be dynamic or static

→ Thinking about the characteristics of the dataset is important for coming up with a good visualization design

attribute types

→ Important for coming up with a good mapping between data and visuals

- Categorical
  - also called: nominal [in name only]
- Ordered
  - Ordinal
  - Quantitative

ID	Name	Age	Shirt Size	Favourite Fruit
1	Apple	8	S	Banana
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Fanny	10	S	Lychee

categorical attributes

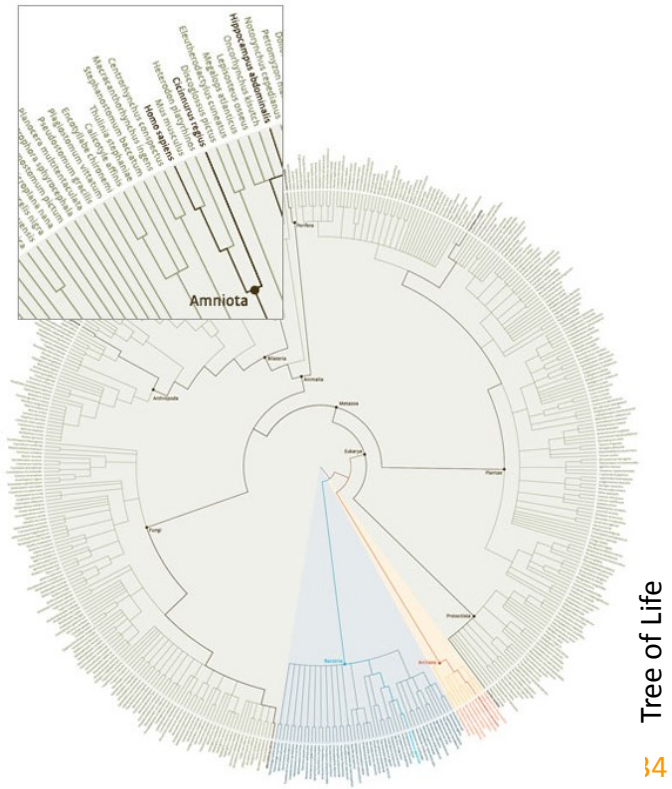
- Discrete values in a single category
- No intrinsic ordering
  - Although an external ordering can be imposed (e.g., by alphabet)
- Operations
  - Is the same as (=)
  - Is not the same as (≠)
- Examples
  - Movie genre
  - City names
  - Student names
  - Types of pets
  - ...

https://dribbble.com/shots/1056368-Movie-Genre-Icons

hierarchies: relationships between categorical attributes

categorical data

- Multiple categories that are associated in a “parent-to-child” relationship
- Each item is associated with exactly one higher-level category, but can be associated with several lower-level categories
- Examples
  - Family relationships: father, mother, daughter, aunt, grandma...
  - Company structures
  - Keyword hierarchies
    - Technology
    - Transportation & Travel
    - Sea Transport & Travel
    - Submarines
  - Time: hour, day, week, month, year...



Tree of Life
https://www.khanacademy.org/partner-content/amnh/human-evolutio/human-evolution-the-evidence/a/tree-of-life

ordinal attributes

ordered attributes

- Discrete values in a single category
- Have an implicit, well-defined ordering
- Ranking is possible BUT mathematical operations do not make sense

- Examples
  - Movie rankings
  - Customer satisfaction rankings

quantitative data

ordered attributes

- Measurements of magnitude
- Math operations make sense
- Typical relations: ranking, ratio, correlation

- Examples
  - Height
  - Temperature
  - Stock prices
  - ...

interval: common groupings of ordered attributes

ordered attributes

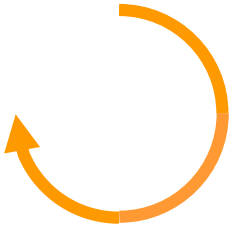
- Series of numerical ranges
- Subdivide a range of quantitative values into smaller ranges
- Sequential with implicit ordering

- Examples
  - Age categories [18-25], [26-35], [36-50], [>50]
  - Order sizes [>=0 and < 1000], [>=1000 and < 2000], [>=2000 and < 3000]
  - Units of time: Months, years, decades...

Different types of ordering

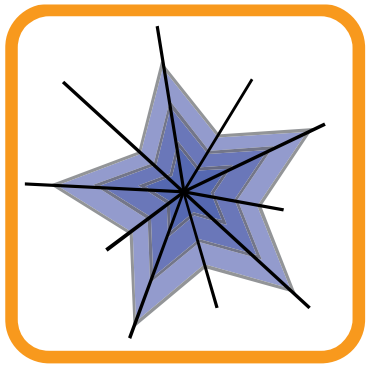
ordered attributes

- Sequential
  - range from minimum to maximum
  - Example: age, mountain height from sea level to peak
- Diverging
  - Two sequences pointing in opposite directions and meet at one point
  - Example: temperature
- Cyclic
  - Values wrap back to starting point
  - Examples: time



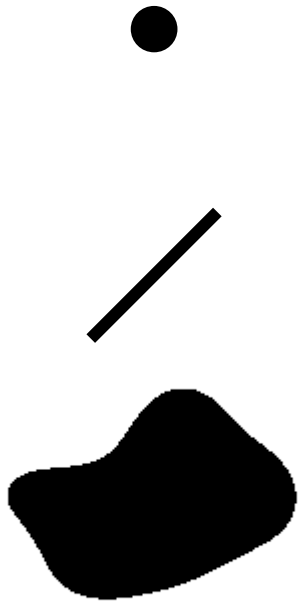
from data to visuals

Marks & Visual Variables



marks

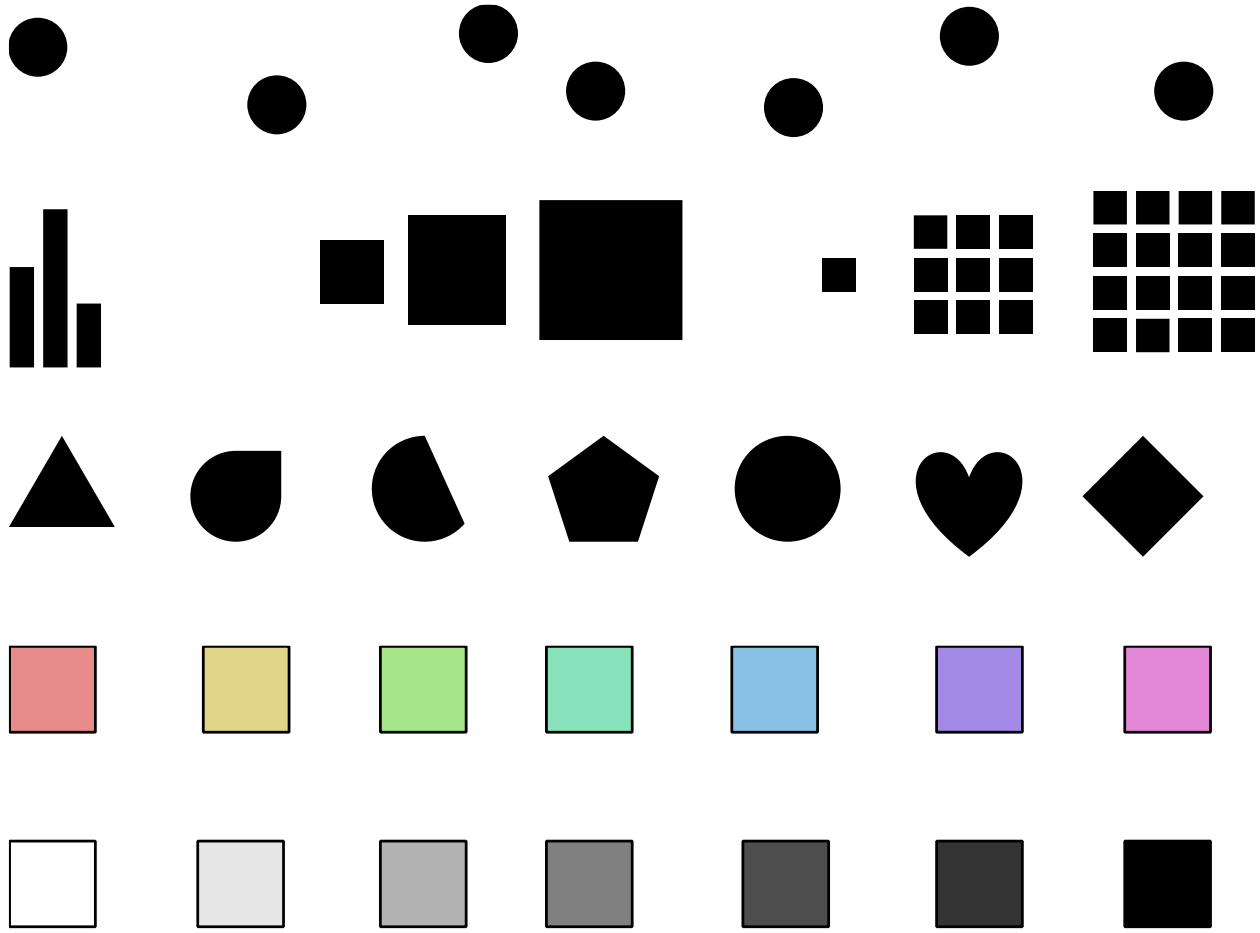
- Basic graphical elements visible on paper/on screen
- Have a dimension (1D, 2D, 3D)



- Point
  - Specific position (2D or 3D)
  - No dimensions (according to math) BUT
  - In visualisation: can have a size and shape
- Line
  - One dimension (1D)
  - Variable length/size
  - In visualisation: can have a width
- Area
  - Two dimensions (2D)

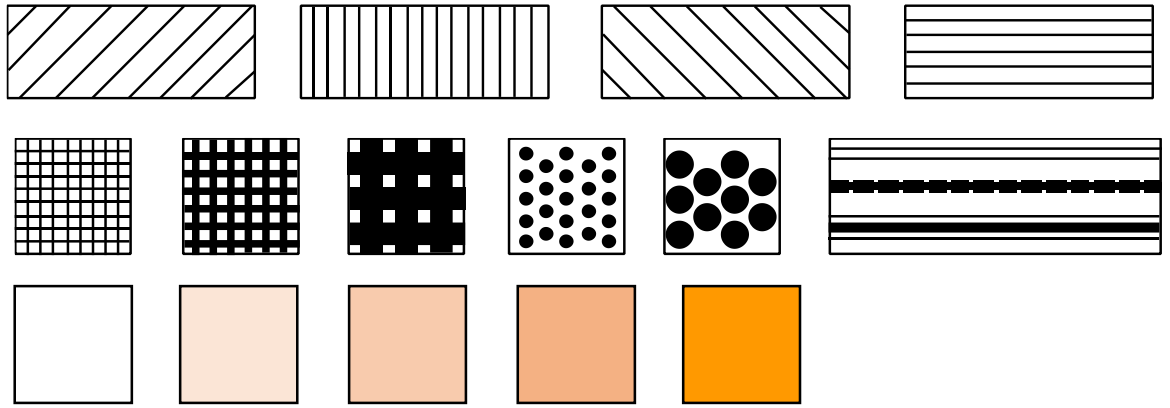
visual variables [or **channels** as they are called by Munzner]

- Position
- Size
- Shape
- Color (hue)
- Color (value)



more visual variables

- Orientation / angle
- Texture
- Saturation
- Transparency
- Curvature
- Movement
- Flickr



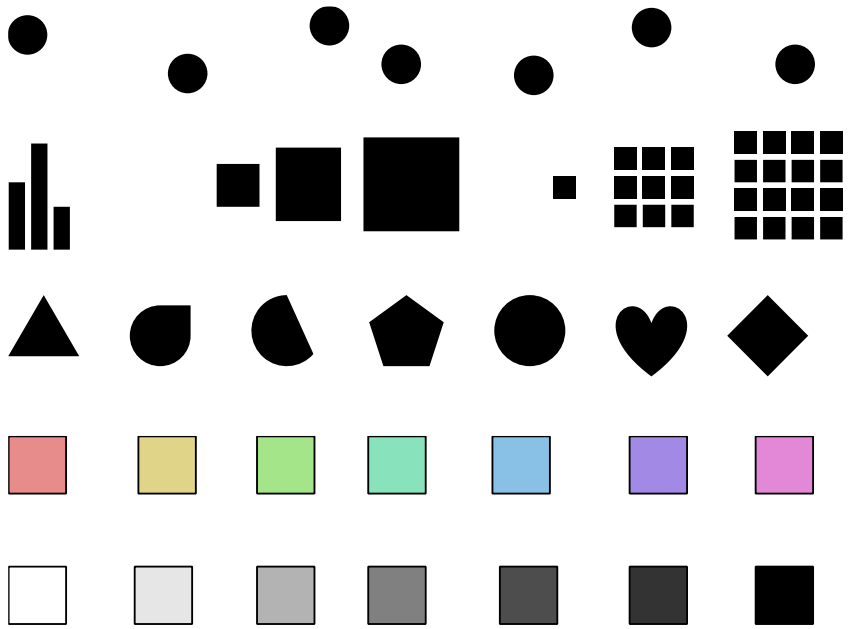
from data to visuals

- Categorical attributes
- Ordered attributes
  - Ordinal
  - Quantitative

Visual Mapping

- Marks

- Visual variables (properties of marks)



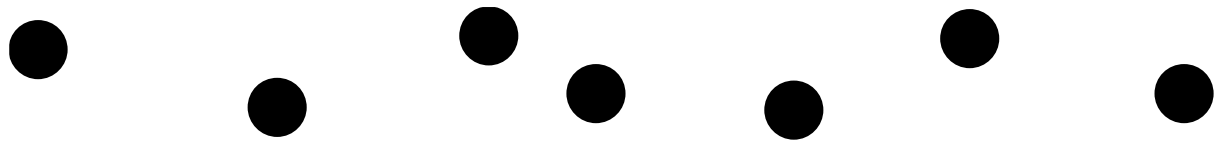
from data to visuals

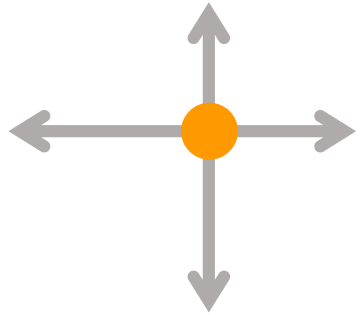
- The choice of visual variables to represent your data attributes **hugely** matters!
- Encoding the same attributes using different visual variables can result in a different perception, experience and interpretation of the represented information
  - Perceptual processing
  - Cognitive processing

perceptual characteristics of visual variables

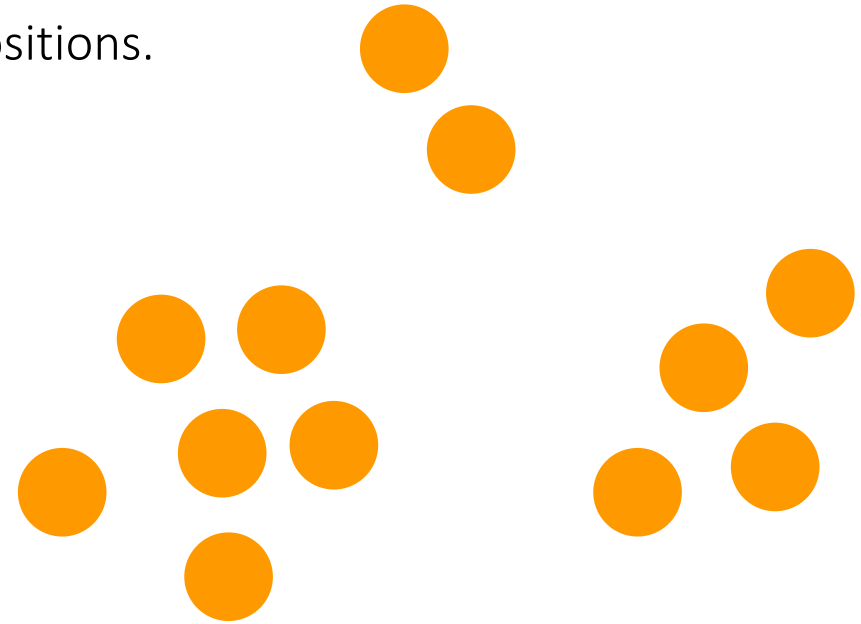
- Order
  - Can an order in the marks be perceived?
- Quantitative
  - Can marks be perceived as proportional to each other?
- Associative & selective
  - Can marks be perceived as similar?
  - Can marks be perceived as different?
- How many distinctions within one variable can we perceive?

position

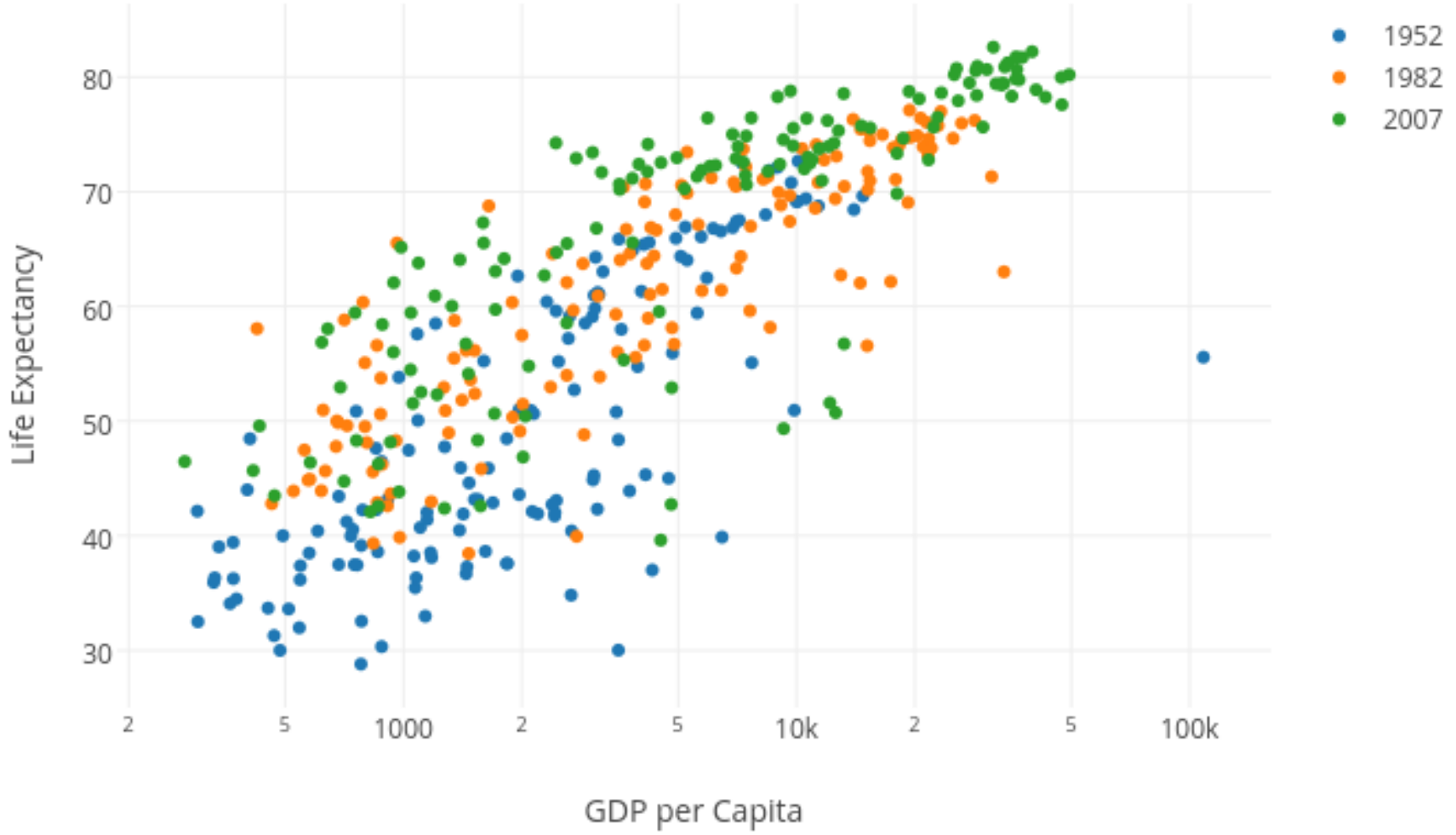




- Changes in horizontal & vertical position (2D)
- Order
  - Yes, we can perceive an order of positions, e.g., from left to right.
- Quantitative
  - Yes, we can estimate the quantitative difference between positions.
- Associative & Selective
  - Yes, similar positions can be easily picked out.
  - Yes, different positions can be easily picked out.
- How many distinctions are possible?
  - Many – scales well to large data sets.



position



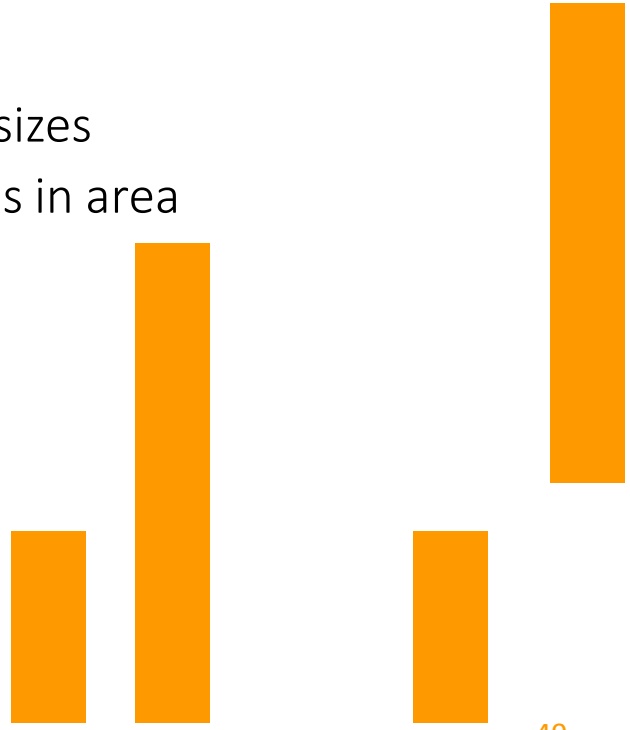
https://plot.ly/pandas/line-and-scatter/

size

- Changes in length, area, repetition
- Order
  - Yes, we can order marks based on their size.
- Quantitative
  - Yes, we can estimate the quantitative differences between different sizes
  - However, differences in length are easier to estimate than differences in area

size

- Changes in length, area, repetition
- Order
  - Yes, we can order marks based on their size.
- Quantitative
  - Yes, we can estimate the quantitative differences between different sizes
  - However, differences in length are easier to estimate than differences in area
  - Also, alignment matters

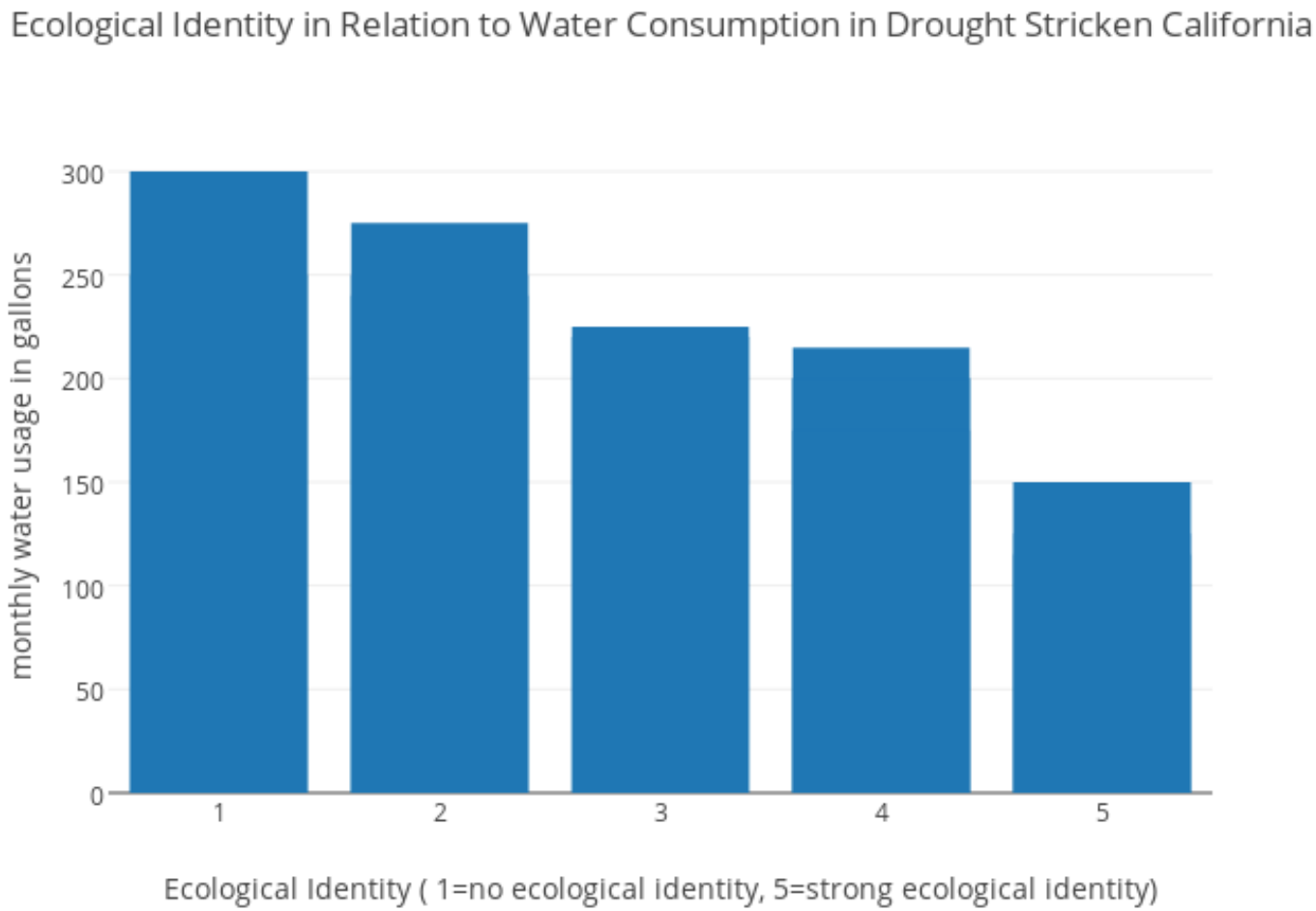


size

- Selective & Associative
  - Yes, but limited: approx. 5 different sizes can be easily distinguished.
  - Also depends on the spatial distribution.

- How many distinctions are possible?
  - Many, but with limitations (see above)
  - Alignment matters
  - The way we apply size matters: areas are less easy to distinguish and estimate than length

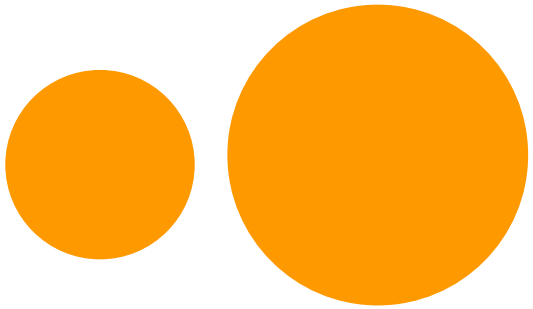
size



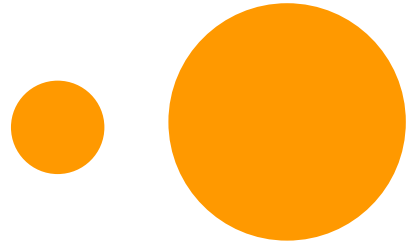
https://plot.ly/~renae.farris611/5/ecological-identity-in-relation-to-water-consumption-in-drought-stricken-califor/

size

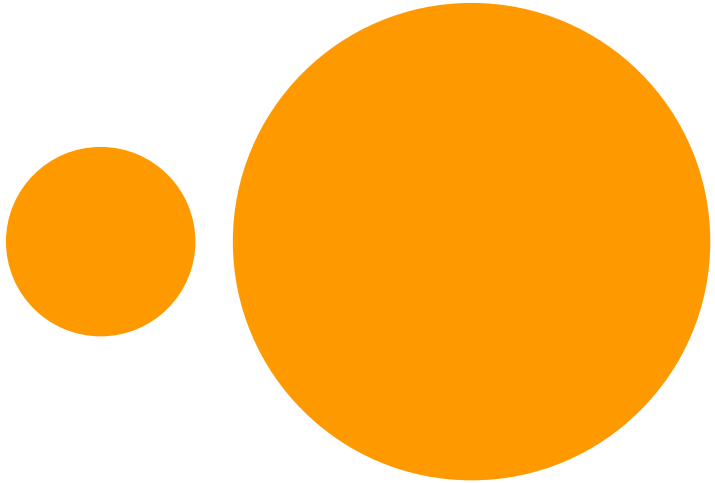
- Area?
- Radius?
- Diameter?



area = 1 area = 2.5

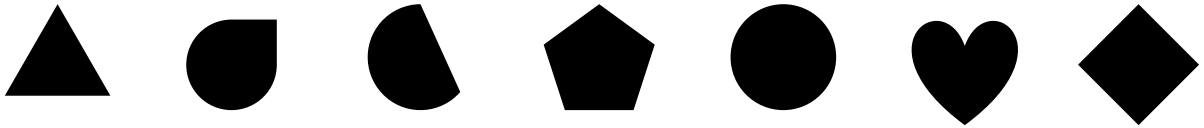


radius = 1 radius = 2.5



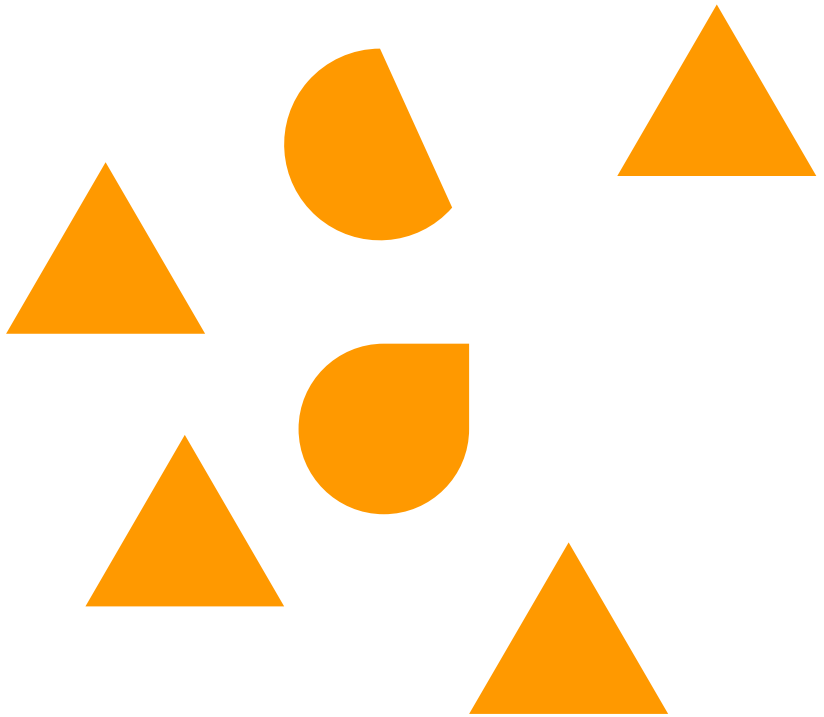
diameter = 1 diameter = 2.5

shape

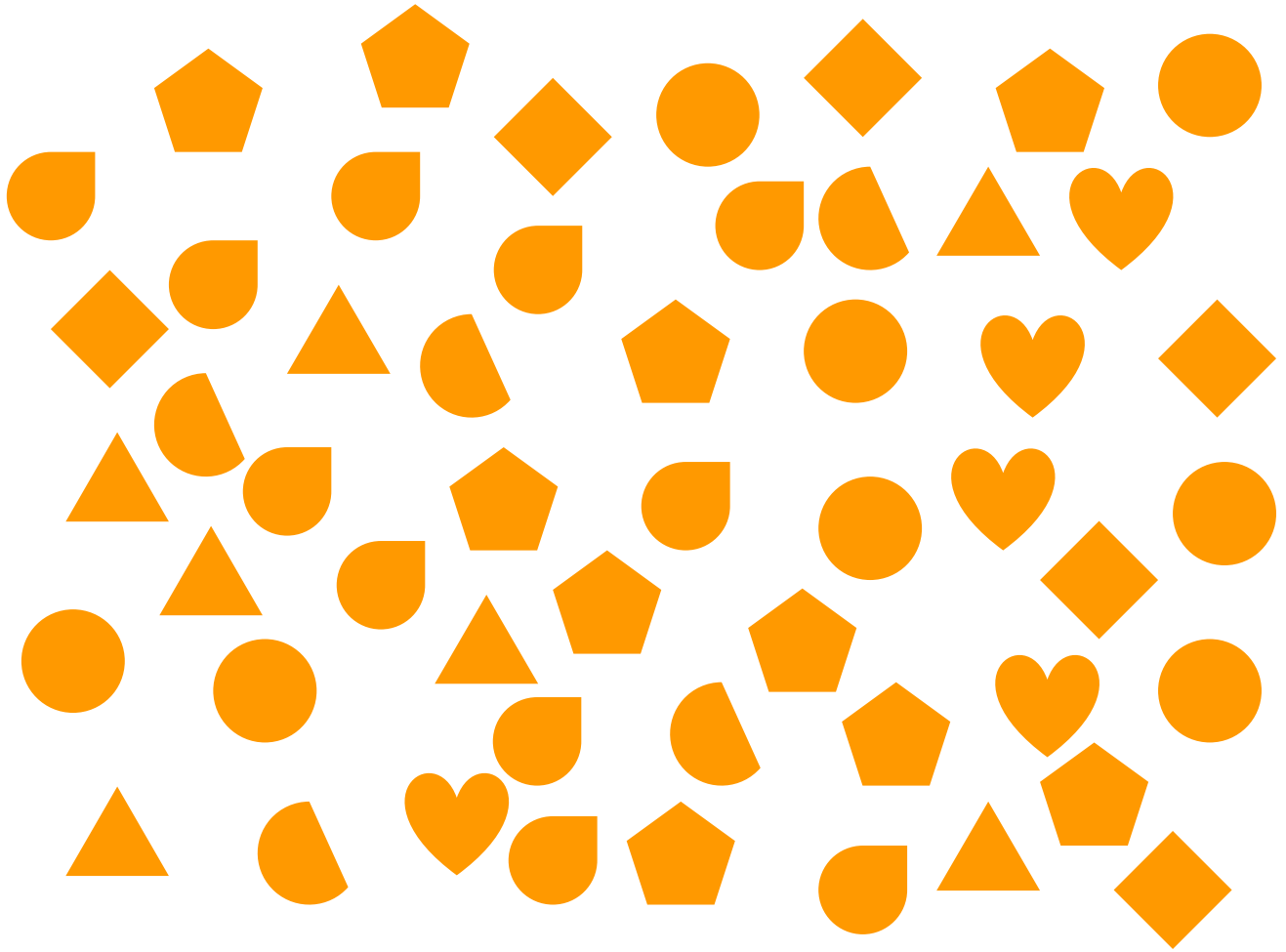


- Order
  - No.
- Quantitative
  - No.
- Associative & Selective
  - Yes, but depends on the number of other shapes (distractors) and distribution.
- How many distinctions are possible?
  - Infinite possibilities, perceptually limited.

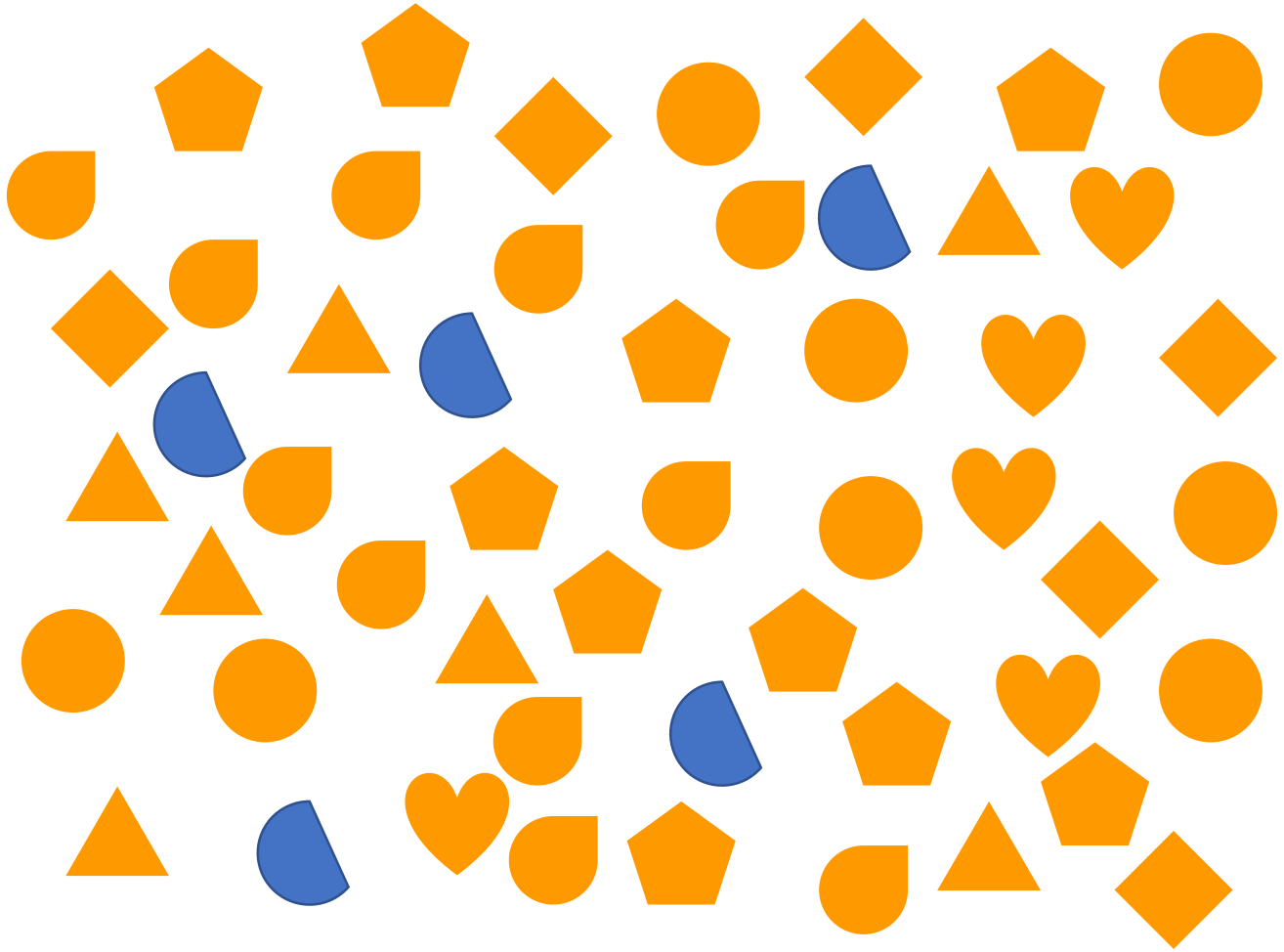
shape



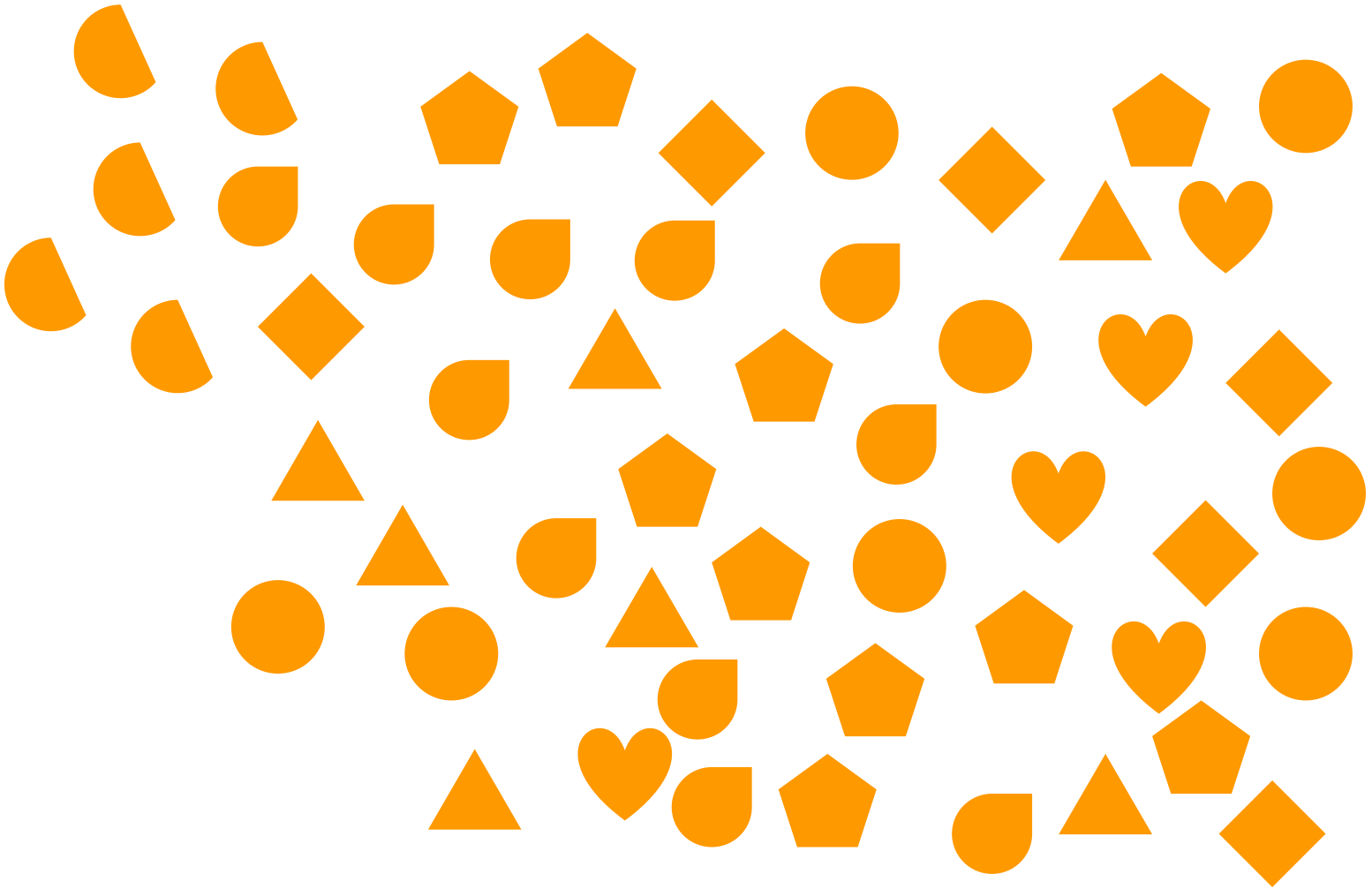
shape



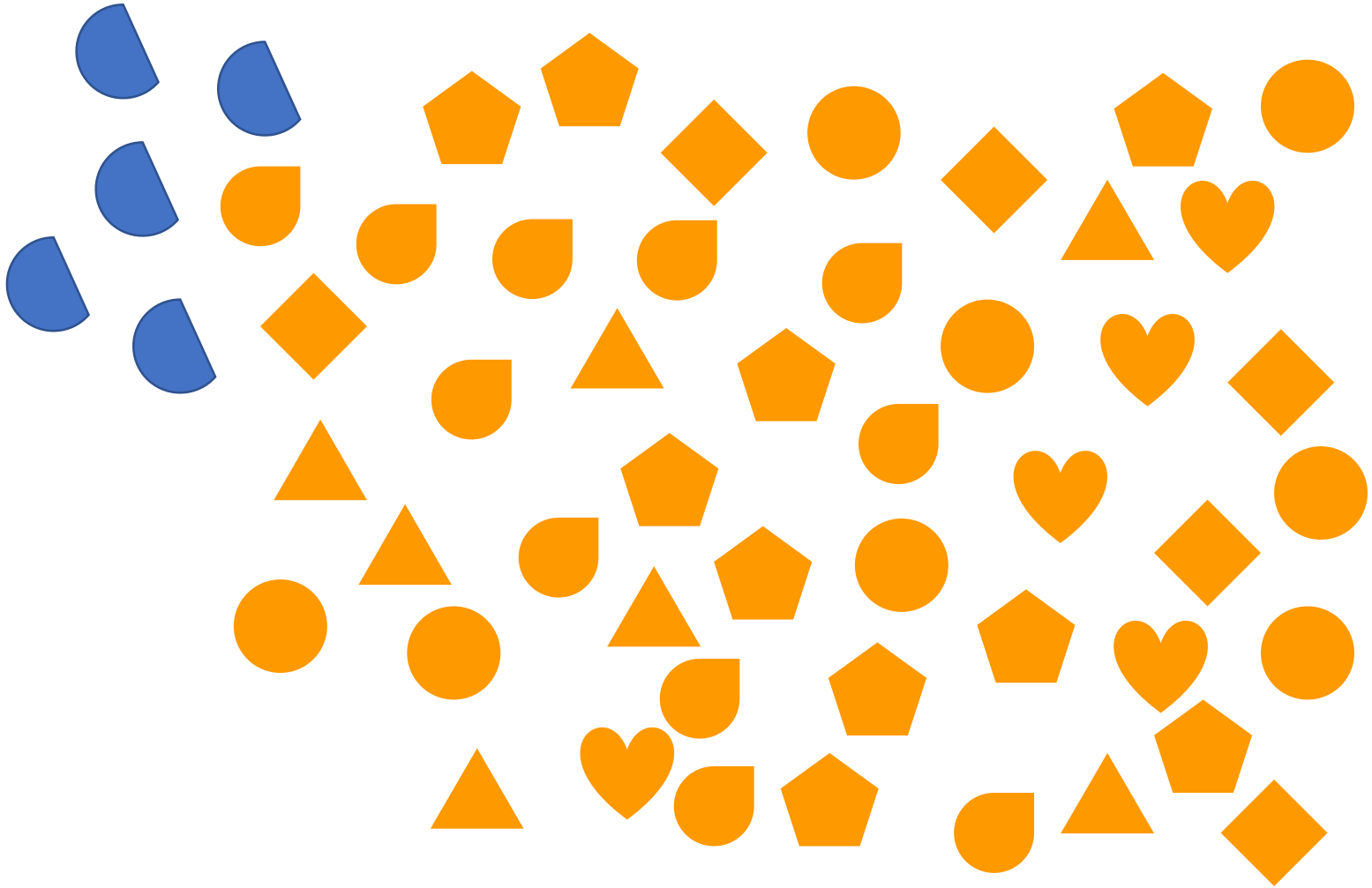
shape



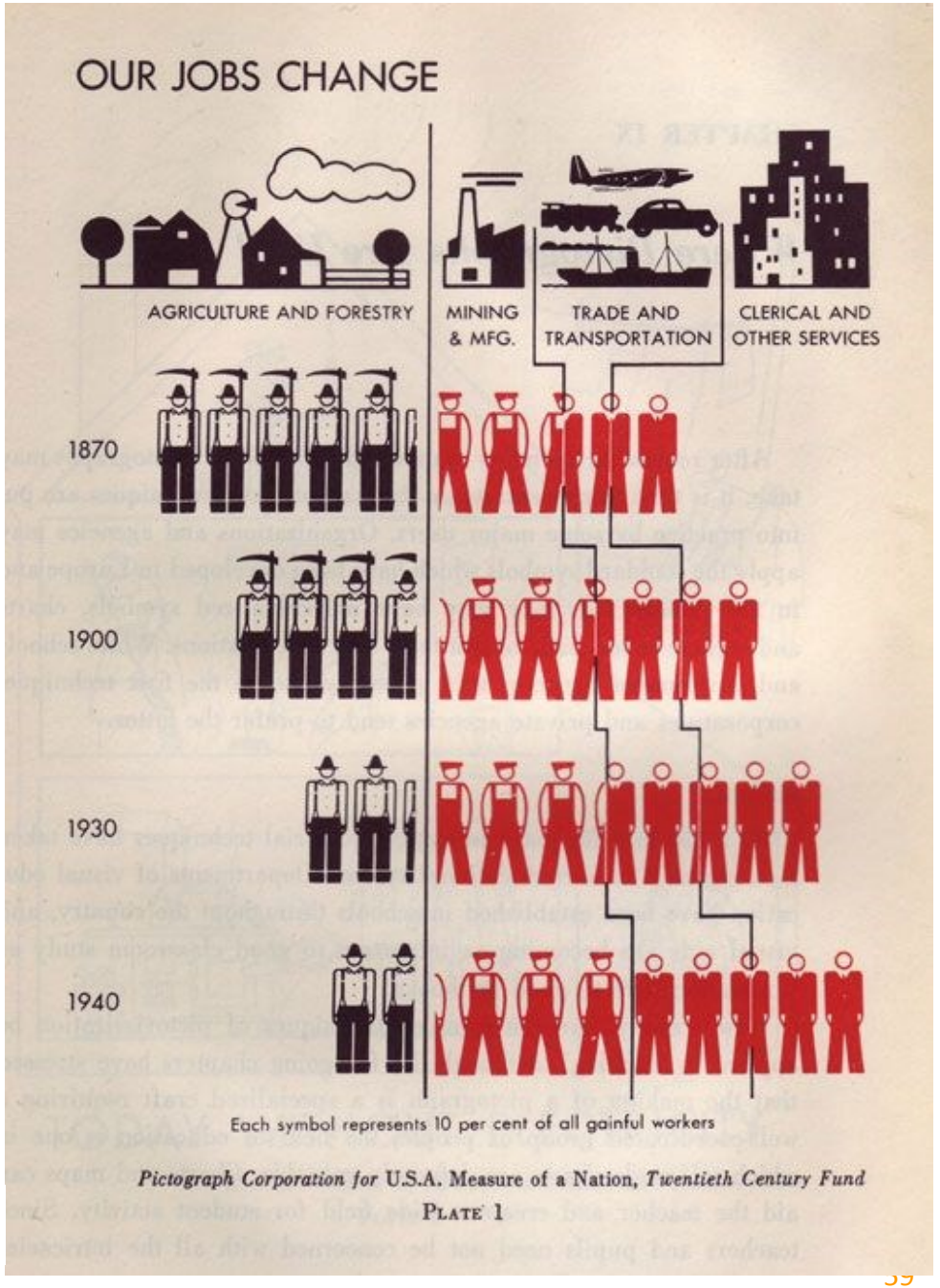
shape



shape



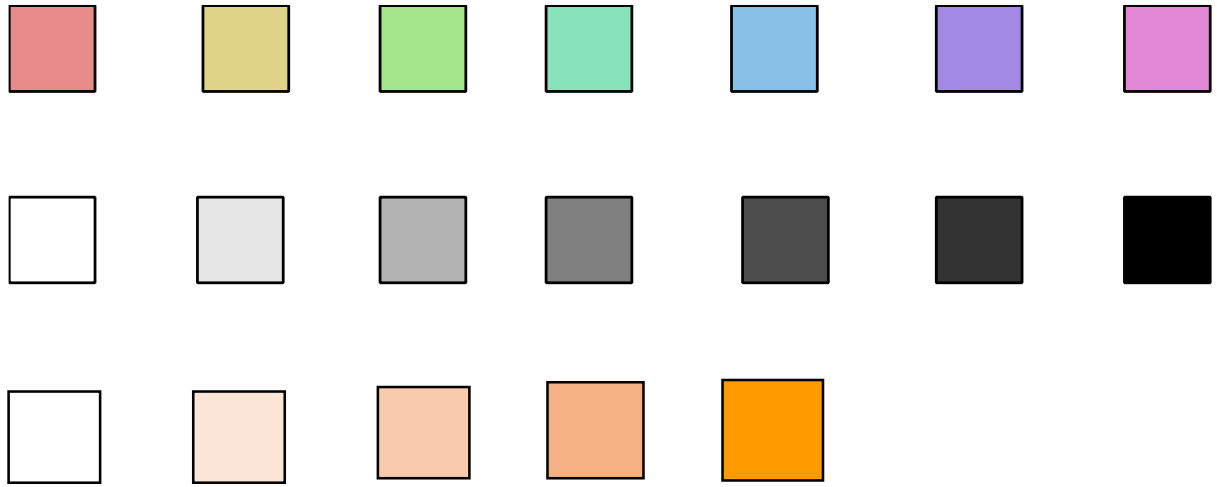
shape



U.S. Job Change, Otto Neurath

colour

- Hue
- Brightness (value)
- Saturation



colour (hue)

- Ordered
  - No.
- Quantitative
  - No. No. No. No!!!
- Associative & Selective
  - Yes, but it depends on the number of distractors and spatial distribution.
- How many distinctions are possible?
  - Infinite in theory, but highly limited perceptually to approx. 7 colours.

colour (hue) ordered or quantitative?

colour (value) ordered or quantitative?

colour (value/brightness) 

- Ordered
  - Yes, we can perceive an order in marks that vary in brightness/value, e.g., light to dark.
- Quantitative
  - No. We cannot estimate the quantitative difference between different colour values.
- Associative & Selective
  - Yes, but depends on number of shades (typically not more than 7)
- How many distinctions are possible?
  - Nearly infinite but perceptually limited.

colour (value/brightness)

http://erikalovesmaps.blogspot.co.uk/2013/10/lab-5-choropleth-maps-in-grayscale.html

visualization process

- Consider the types of attributes in your dataset
- Identify the most important attributes to be represented
  - Driven by the tasks and questions you want to facilitate with your visualization
- Consider their attribute type and choose accordingly

lecture 2 – summary

- the process of visualization – what is involved?
  - Visualization pipeline
- Data and attribute types
  - Categorical & ordered (including ordinal and quantitative)
- Visual marks and variables

(tentative) release of Practical 1

- Check the “practical” folder on StudRes
- Read the practical → questions on Monday
- Find the visualization postcard assigned to you here (look for your student ID)
  - https://uh.host.cs.st-andrews.ac.uk/CS5044/P1_visAssignments.htm

- If you cannot find your ID in the list, email uh3@st-andrews.ac.uk

Next week

- Monday exercise class
  - Sketching visualisations
  - Info on Practical 1
- Thursday lecture
  - Expressiveness and effectiveness principles
  - Basic visualization techniques I

- Readings
  - Munzner, Chpt. 5: Marks and Channels
  - Munzner, Chpt. 7: Arrange Tables
  - Card et al. Readings in information visualization (Chapter 1) (optional)

catching up on Java Script

- For those of you, who would like to catch up on Java Script, you can check out the lecture notes and exercises from CS5002 – Programming Principles and Practice
  - https://studres.cs.st-andrews.ac.uk/CS5002/
- If you want to take a peak at D3.js before we start working with it in-class, you can check out the following book and webpage. Note that this is optional for the students completely new to JS and/or web programming. We will get started with D3.js from scratch in Week 4 or 5.
  - Scott Murray – Interactive Data Visualization for the Web (2nd Ed.)
  - https://alignedleft.com/work/d3-book-2e