# 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/

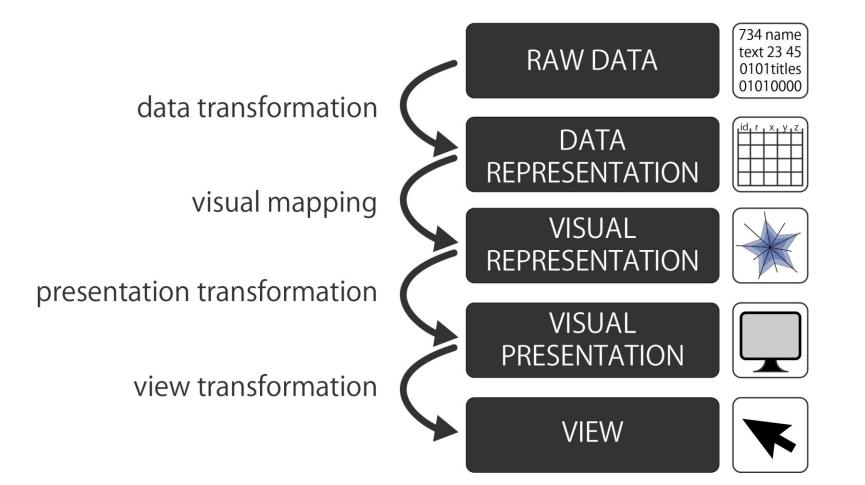
#### outline for today

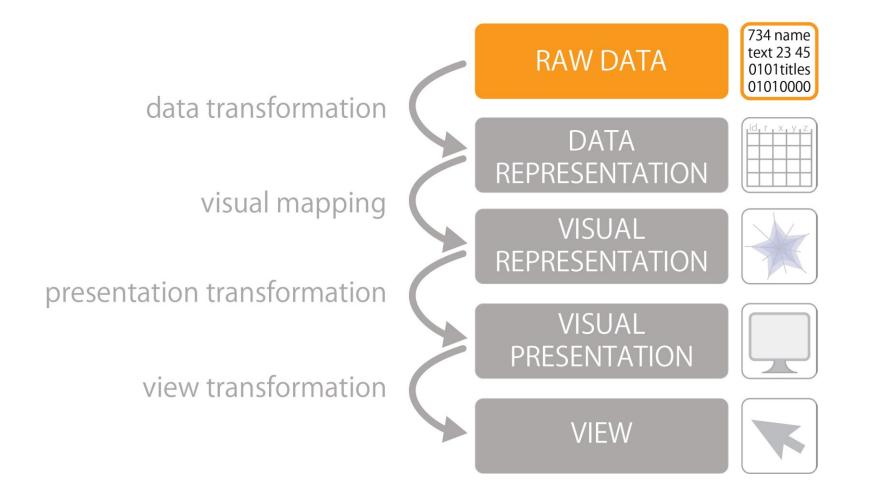
- Process of creating visualizations
- What? data types
- How? visual marks and variables

#### 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

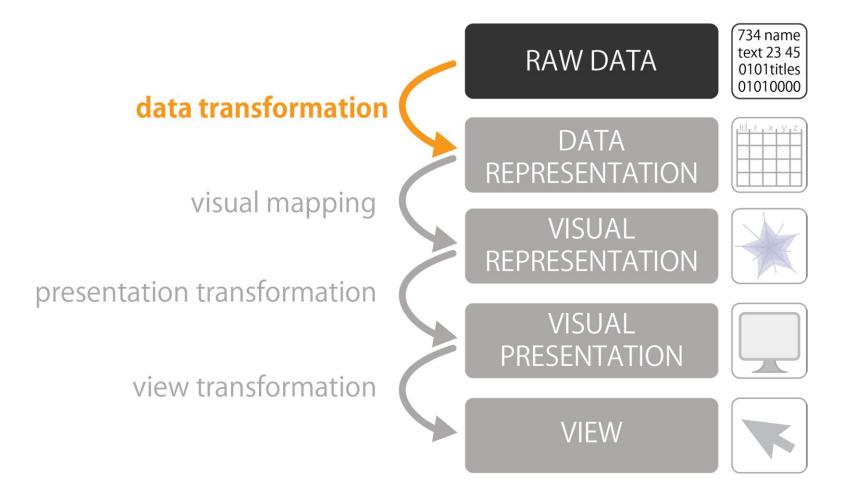
the visualization process



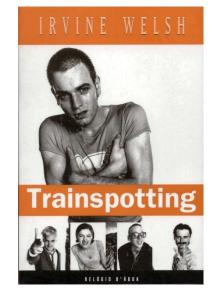


### example: raw data

|                    | 269 Philippines                           | APAC                       | Southeast As    | OFF-PA-1000 Office Suppli Paper  | Enermax Con                   | 49.302        | 3                     | 0.45        | -18.828                     |   |
|--------------------|---|----------------------------|-----------------|--|-------------------------------|---------------|-----------------------|-------------|-----------------------------|---|
|                    | 270                                       | •                          |                 | -1000 Office Suppli Paper  | Wirebound N                   | 30.44         | 4                     | 0           | 14.3068                     |   |
|                    |   | erimen <sup>.</sup>        | t dat           | A -1000 Office Suppli Paper  | Xerox 230                     | 12.96         |                       |             |                             | photographs                             |
| 512<br>512         |   | cinici                     | c aut           |  | Computer Pri                  | 1.344<br>5.34 | 1                     |             |                             |   |
|                    | 273<br>274 Cuba                           | LATAM                      | Caribbean       | AR-10C Office Suppli Binders<br>OFF-AR-1000 Office Suppli Art  | Cardinal Hole<br>Boston Pens, | 5.34<br>19.32 |                       |             |                             |   |
|                    | 274 Cuba<br>275 Mexico                    | LATAM                      | North           | OFF-BI-1000 Office Suppli Binders  | Ibico Index Ta                | 41.64         |                       |             |                             |   |
|                    | 276 Brazil                                | LATAM                      |                 | OFF-LA-1000 Office Suppli Labels   | Novimex Ship                  | 5.92          | -                     |             |                             |   |
|                    | 277 Germany                               | EU                         |                 | OFF-BI-10004 Office Suppli Binders   | Acco Binding                  | 52.59         |                       | A COMPANY   |                             |   |
|                    | 278 United States                         | 77506 US                   |                 | OFF-AP-1000 Office Suppli Appliances   | Eureka Dispo                  | 1.624         |                       |             |                             |   |
|                    | 279 United States                         | 92037 US                   |                 | OFF-PA-1000 Office Suppli Paper  | Great White                   | 17.94         |                       |             |                             |   |
|                    | 280 Nigeria                               | Africa                     |                 | OFF-STA-100 Office Suppli Art  | Stanley Highl                 | 5.364         |                       | Ne          | you                         |   |
| 512                | 281 Zimbabwe                              | Africa                     | Africa          | OFF-SAN-100 Office Suppli Art  | Sanford High                  | 9.612         |                       |             |                             |   |
| 512                | 282 Turkey                                | EMEA                       | EMEA            | OFF-BOS-100 Office Suppli Art  | Boston Pens,                  | 34.128        | THAT                  |             | mh                          | arad yours list and the second          |
| 512                | 283 Brazil                                | LATAM                      | South           | OFF-PA-1000 Office Suppli Paper  | Green Bar Me                  | 84            | <b>H</b> E            | TE          |                             | ered your                               |
| -8a I              | EEE VisWeek @viswe                        | eek - 16 Oct 2009          |                 | 0 Office Suppli Paper  | SanDisk Mes                   | 18.64         |                       |             |                             |   |
| 60 k 60 (          | Coming up: InfoVis                        | Papers: Text Visualization | n at Blenheim # |  | Eaton Parchn                  | 26.94         | P                     | ~ 1         |                             | tfo?                                    |
| ÷ .                | 6 t3                                      | • •••                      |                 | Coffice Suppli Binders   | Avery Binder,                 | 58.05         | PG                    | 9           |                             | Life?                                   |
|                    |   |                            | -               | Contraction of the second s  | -                             | 1             |                       |             |                             |   |
|                    | EEE VisWeek @viswe                        |                            |                 |  |                               | 4             |                       |             |                             |   |
|                    | coming up: Vis Pap<br>at Dennis A/B #visw | ers: Advanced Volume V     | risualization   |  |                               | 6             | 3                     | 0           | 11.2308                     | 0.01 High                               |
| c                  | at Definits AVD #VISW                     | CCK                        |                 | 100  | ~ C                           | 6             |                       | 0           | 2.4                         | 0.00 Medium                             |
|                    | 4 £7                                      | • •••                      |                 | Itt h  |                               | -             | 1                     |             | = 42                        |   |
| -2                 | EEE VisWeek @viswe                        | eek - 16 Oct 2009          |                 | Digitiz  | nd/h                          | andw          | /ritta                | n           |                             |   |
| ، مەلۇرە<br>مەلۇرە | Coming up: Vis Pap                        | ers: Visual Encoding at N  | Marlborough     | Digitiz  | .cu/II                        | anuv          | iiii                  | 11          | parager.                    | barg dugar, bintan                      |
| - ů -              | 6 t3                                      | •                          |                 | all bu   |                               |               |                       |             | 0 0                         | ir sourcehi                             |
|                    |   |                            |                 | <i>D</i> notes   | or tex                        | KT I          |                       |             |                             | tion selve ye                           |
|                    | EEE VisWeek @viswe                        |                            |                 | DA   | 0. 00/                        |               |                       |             | eablo. On                   | ada da hel                              |
|                    | Coming up: InfoVis                        | Papers: Systems at Blen    | heim #viswe     | 711  | 1º HO .                       | plach         | in teles-             | reter you   | Jouera perch                | tight planting                          |
|                    | 4 t7                                      | ••••                       |                 | 14   | 5 8 8 8 02                    | A " 6.10      | Cité ja de            | Ana blpc    | : bir alalet                | " Give durgent                          |
| 0                  | EEE VisWeek @viswe                        | ook - 15 Oct 2009          |                 | area als we the 1940   | 1 12 3/ 2 2                   | C             | liger. Bu             | ana his     | we love ada                 | a atmyer                                |
|                    |   | at Marlborough B/C #vis    | sweek           | 14   | - 11th Fel 2                  | F ? e oyá     | lanyar, ca- 6         | ) Muanne    |                             | Zarye alli                              |
|                    | <b>6 6</b>                                |                            |                 | of my they w   | 15 3                          | 2 2 - Aria    | weathin               | i de allei, |                             | principanti di                          |
|                    | The Let                                   | •                          |                 |  | 1 mil                         |               | Anadowin              |             | e. En arindo                | a herdni b                              |
|                    | EEE VisWeek @viswe                        |                            |                 | to anot  | X A                           | yapu          | national for          |             | Ana o da i<br>indisigle son |   |
|                    |   | ers: Neurobiological and   | Molecular Ir    | E  | 11                            | eiter         |                       |             | a laistan diran             | par. o da a                             |
| -                  | Dennis A/B #viswee                        | ĸ                          |                 | green and run  | C. M. M                       | enon t        |                       | sintileniza | nip binalim                 |   |
|                    | 4 t7                                      | • •••                      |                 | The later 1  |                               | hal           | myar. &               | bolom ille  | gintigo olare               | h galding il                            |
| 0                  | EEE VisWeek @viswe                        | ack - 15 Oct 2009          |                 | The betwee A.  | D. chan                       | 5 Lini J      | lemini .              | Muammer     | once endehile               | so so muluza                            |
|                    |   | el: Challenges in Large E  | Data Visualiz   | in tuir  |                               | to dej        | pismos. ida .         | vereh kendi | vorligini anto              | mlandinuga e                            |
|                    | Marlborough B/C #v                        |                            |                 | they maken.  | L + 75. 12                    | - halm        | us. Sindi             |             | bile nefet et               |   |
|                    | 6 t3                                      | • •••                      |                 | finat production ,   | S marine Cl                   | Junin         | un anda-              | arton anen  | nenezi, sahij               | lennegi 60                              |
|                    |   | *                          |                 | June 1 de mar  | 800                           | Solut         | bir estel             | ichs and    | hatten japa                 | setun historia                          |
|                    |   |                            |                 | ATTON A  | 0.100                         | - Jon 2       | Comment               | portage (   | hatlmaga h                  | sice veriger. I                         |
| 53                 | witte                                     | r strean                   |                 | rather greater   | n in ch                       | m             | and the second second |             |                             |   |
| · •                | vvill                                     | Jucun                      | · · · 📕 📕       | 01   |                               | agent - E     |                       |             |                             |   |
|                    |   |                            | <u> </u>        | Then genne to  | nes ha                        | -             |                       |             |                             | and a second and a second as the second |
|                    |   |                            |                 | from bie   | Sur telet                     | in            |                       | <b>_</b>    | ~                           |   |
|                    |   |                            |                 | former de  | any man                       |               |                       |             | Perci                       | onal log data                           |
|                    |   |                            |                 | a state of the sta |                               |               |                       |             |                             |   |
|                    |   |                            | 1               |  |                               | A STATE       |                       |             |                             |   |
|                    |   |                            |                 | and the second se  |                               |               |                       |             |                             |   |



#### example: data transformation

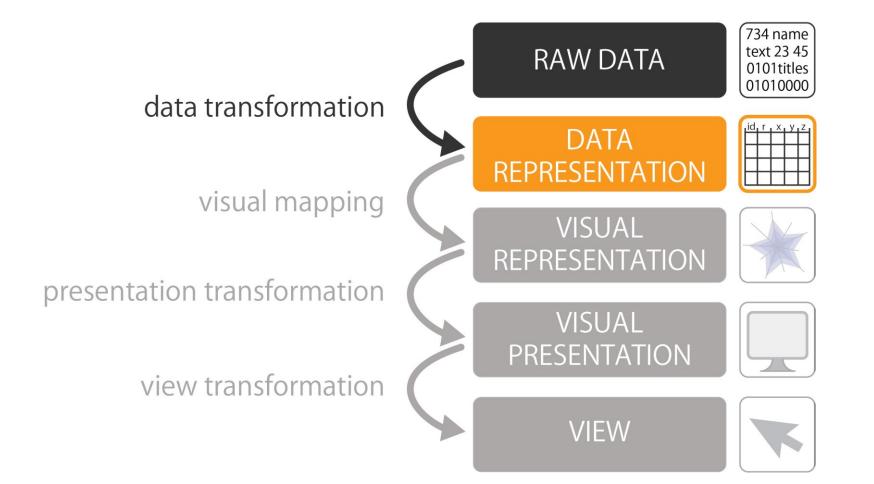


'Mother superior' wis Johnny swan; also kent as the white swan, a dealer whae wis based in Tollcross and covered the Sighthill and Wester Hailes schemes. ah preferred tae score fi swanney, or his sidekick raymie, rather than seeker n the muirhoose-leith mob, if ah could.

#### In the Maleh 1002



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



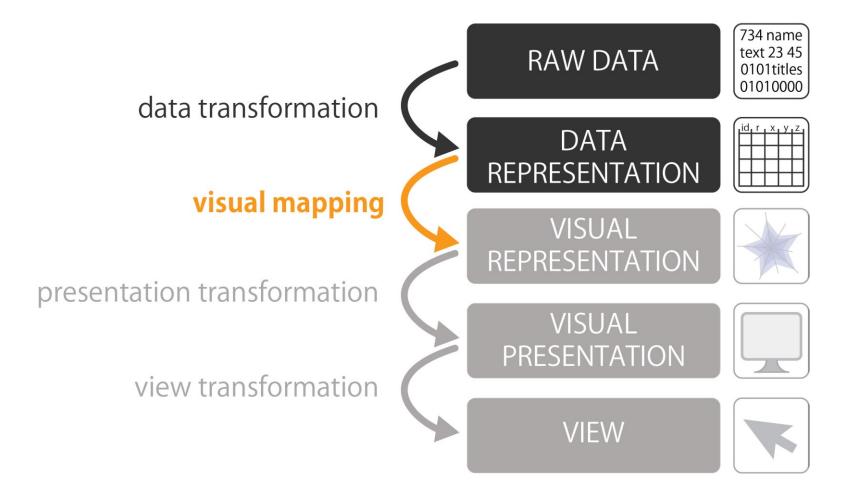
## data representation

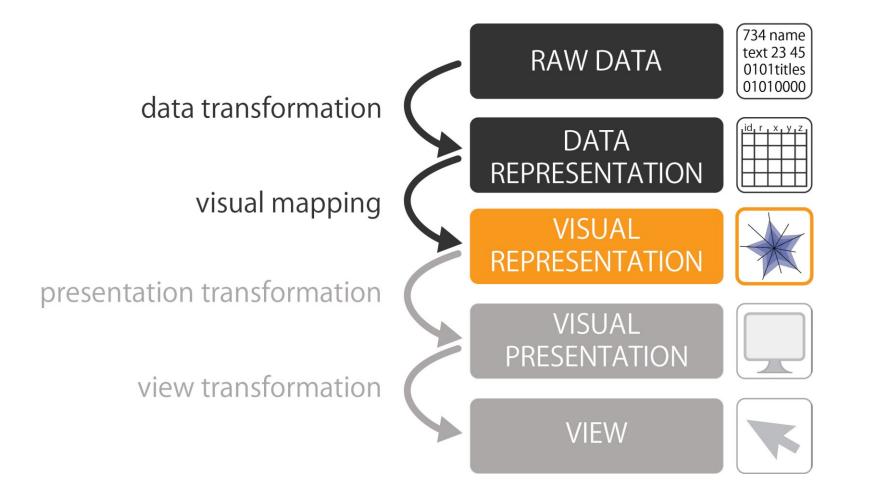
analytical abstraction

- 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

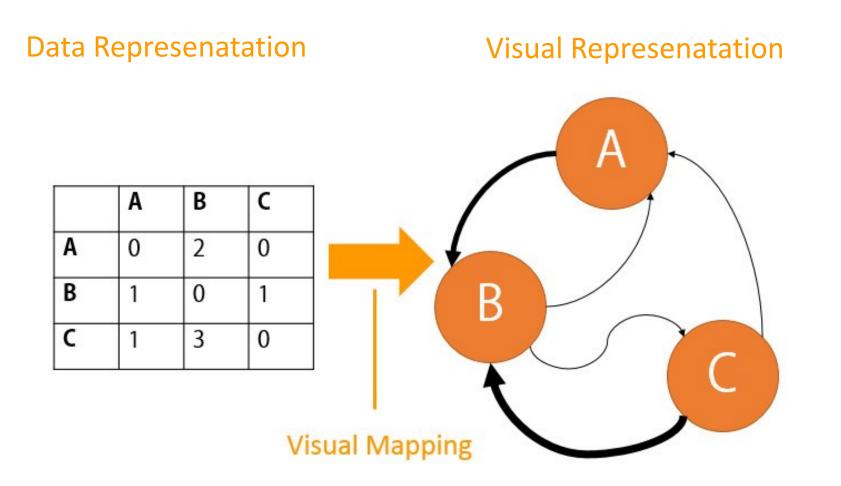
— ...

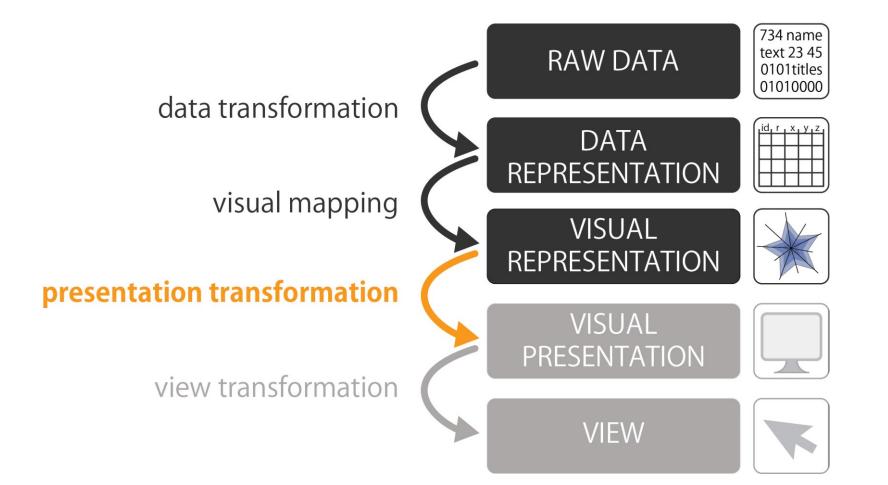
| Author   | Title         | Publication<br>year | Locations                                 | Text  |
|----------|---------------|---------------------|---|---|
| I. Welsh | Trainspotting | 1993                | Tollcross,<br>Sighthill,<br>Wester Hailes | Mother superior' wis<br>Johnny swan; also<br>kent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





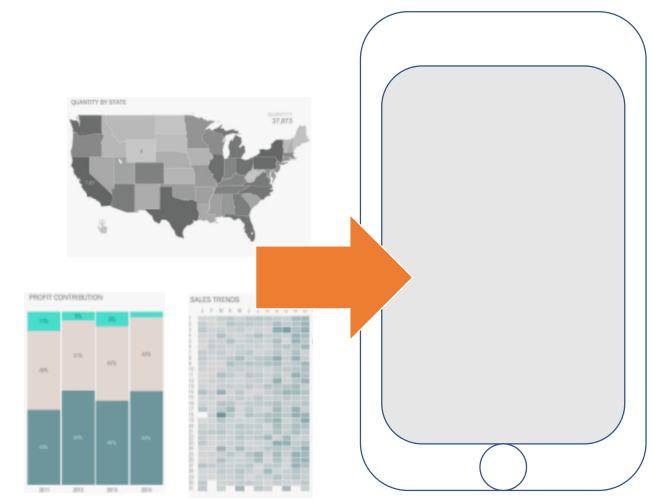
example: data representation

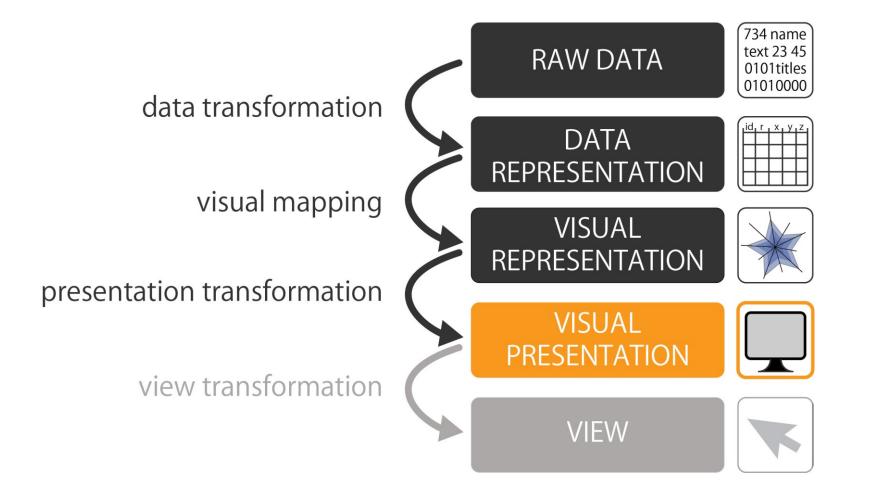


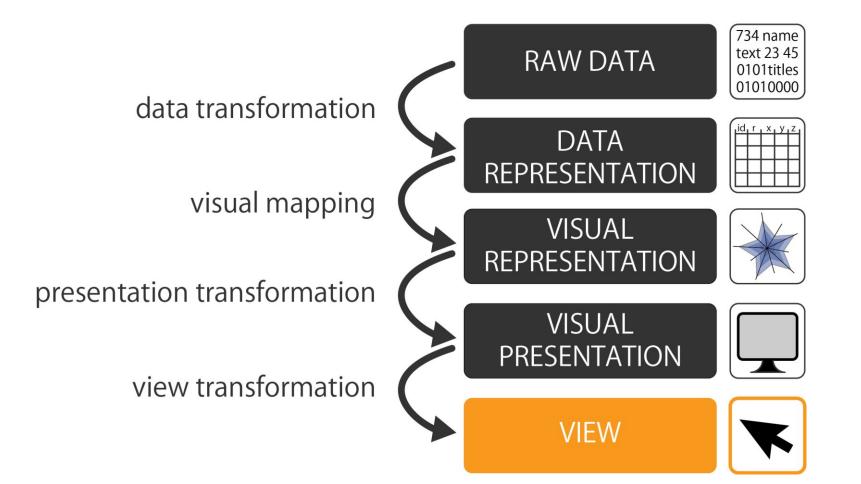


#### presentation transformation

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







#### view

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

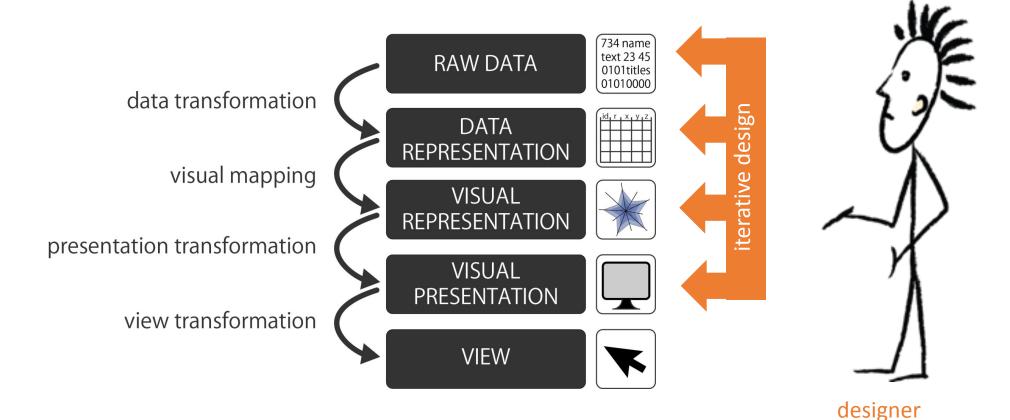
— ...





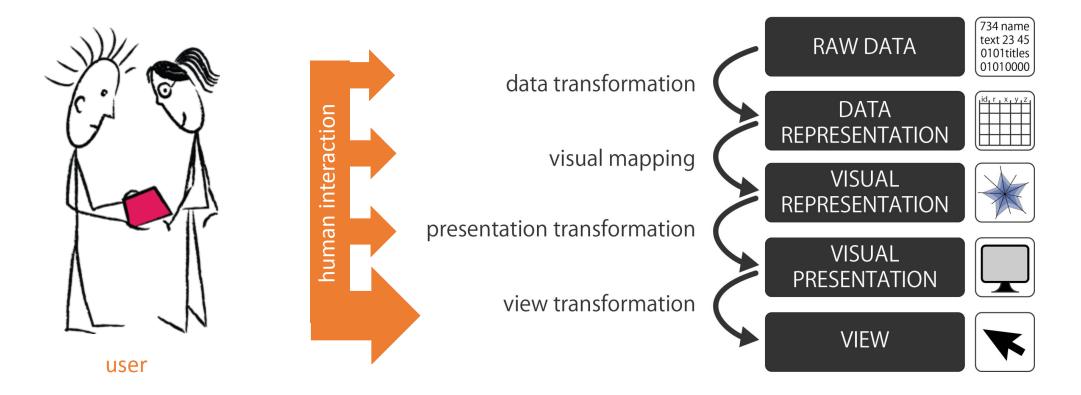
#### visualization pipeline – design perspective

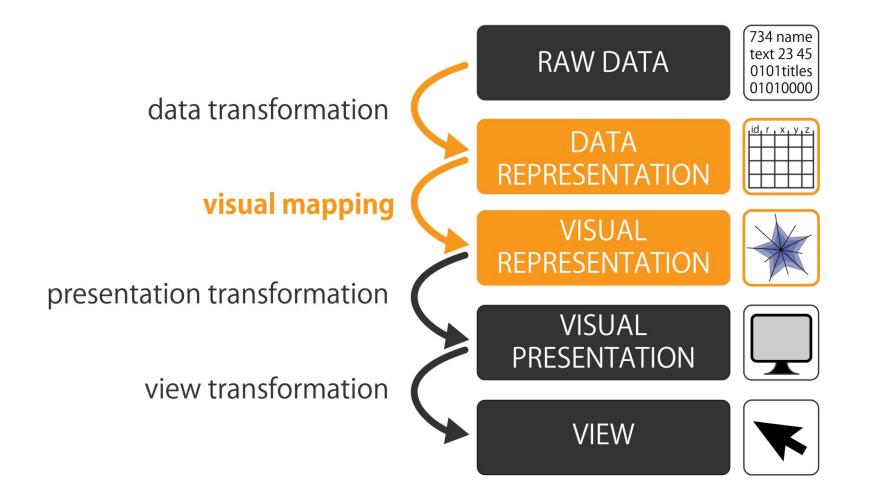
• Guides the visualization design process



#### visualization pipeline – user perspective

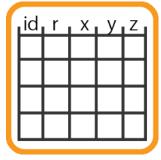
Guides the exploration process





## 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   |

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

• Data attributes

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

• Items or *data points* 

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

• 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   | Μ          | Durian          |
| 4  | Desmond | 13  | L          | Elderberry      |
| 5  | Fanny   | 10  | S          | Lychee          |

• 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   | Μ          | 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...)

- 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

| ID | Name    | Age | Shirt Size | Favourite Fruit |
|----|---------|-----|------------|-----------------|
| 1  | Apple   | 8   | S          | Banana          |
| 2  | Basil   | 7   | S          | Pear            |
| 3  | Clara   | 9   | Μ          | 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     |

#### 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   | М          | 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  $(\neq)$
- Examples

- ...

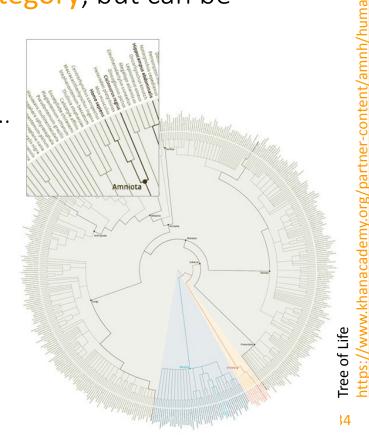
- Movie genre
- City names
- Student names
- Types of pets



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...



## 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

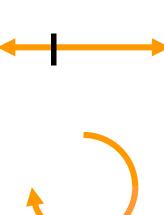


- Age categories [18-25], [26-35], [36-50], [>50]
- Order sizes [>=0 and < 1000], [>=1000 and < 2000], [>=2000 and < 3000]
- Units of time: Months, years, decades...

#### Different types of ordering ordered attributes

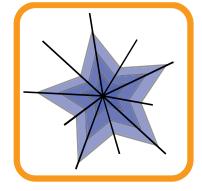
#### Sequential

- 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

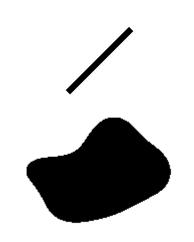


#### marks

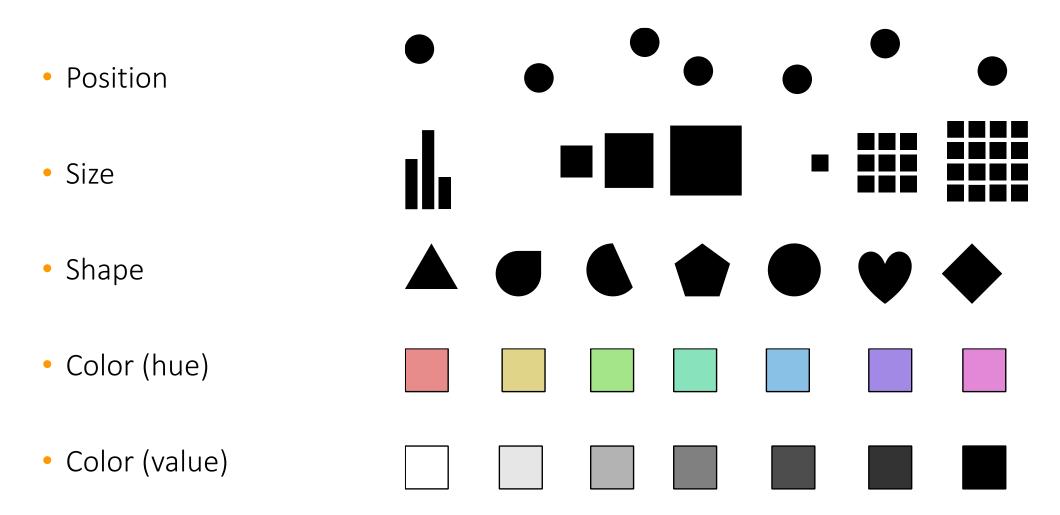
- Basic graphical elements visible on paper/on screen
- Have a dimension (1D, 2D, 3D)

• 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)

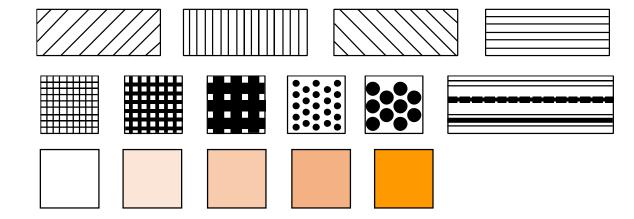


#### visual variables [or channels as they are called by Munzner]



#### more visual variables

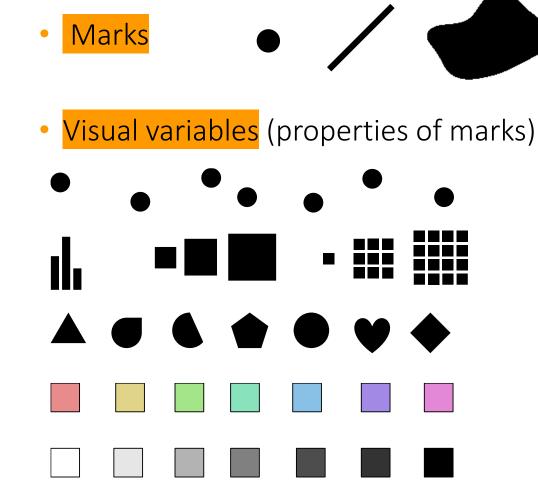
- Orientation / angle
- Texture
- Saturation
- Transparency
- Curvature
- Movement
- Flickr



#### from data to visuals

- Categorical attributes
   Ordered attributes

   Ordinal
   Quantitative



#### 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

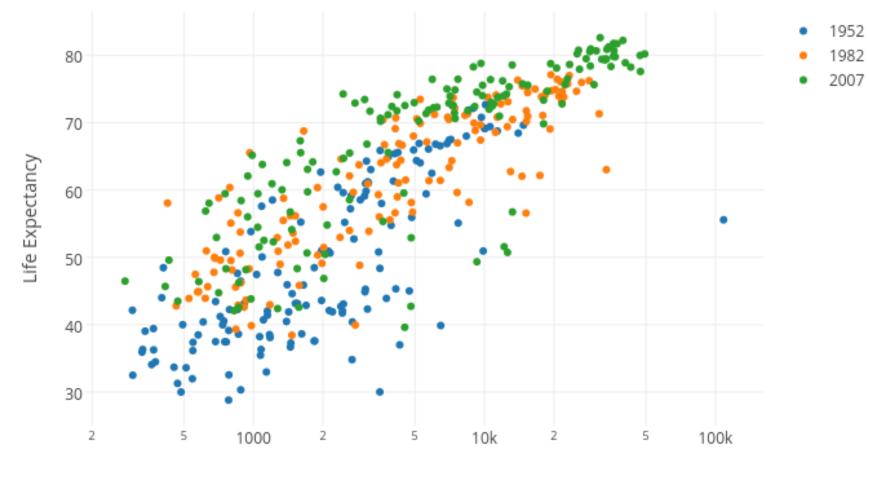
• Order

- 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



GDP per Capita

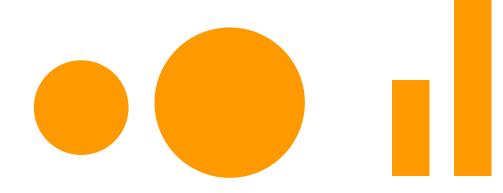
https://plot.ly/pandas/line-and-scatter/



- Changes in length, area, repetition
- Order

size

- 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





- Changes in length, area, repetition
- Order

size

- 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

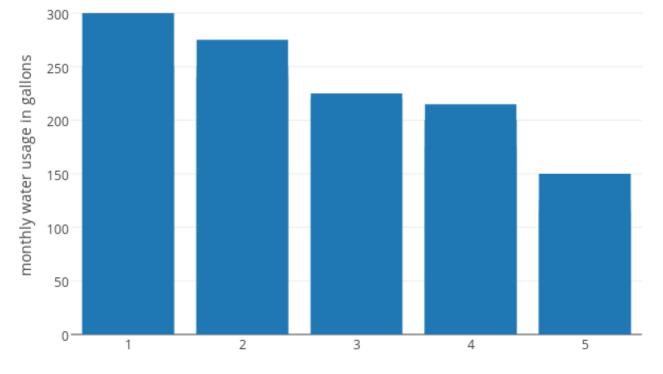
# 

#### 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

#### Ecological Identity in Relation to Water Consumption in Drought Stricken California

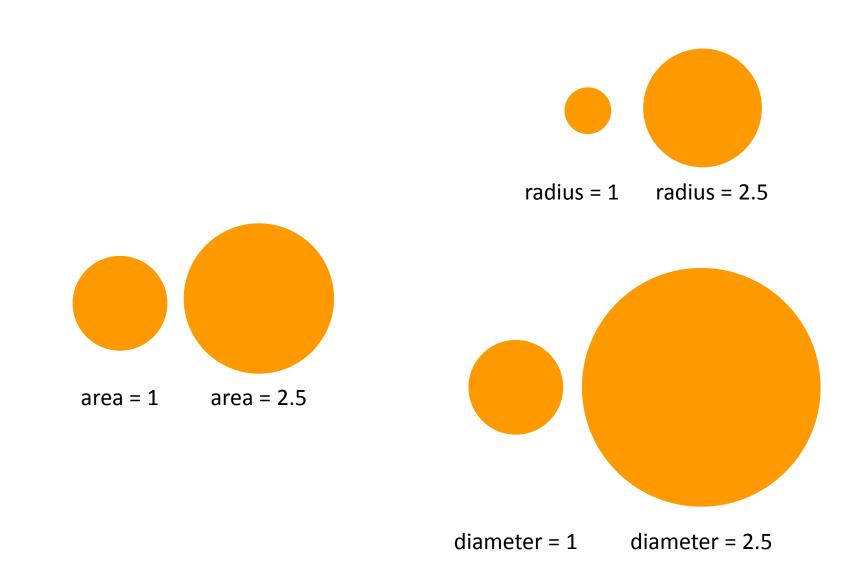


Ecological Identity (1=no ecological identity, 5=strong ecological identity)

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



- Area?
- Radius?
- Diameter?

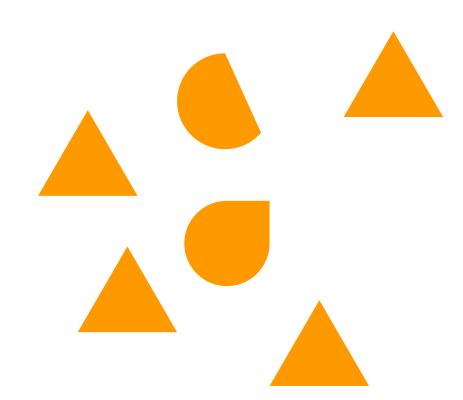


#### shape

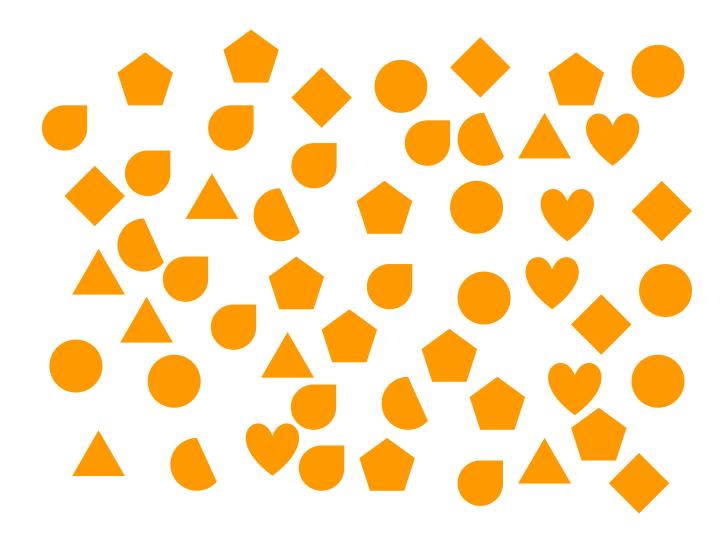
# 

- 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.

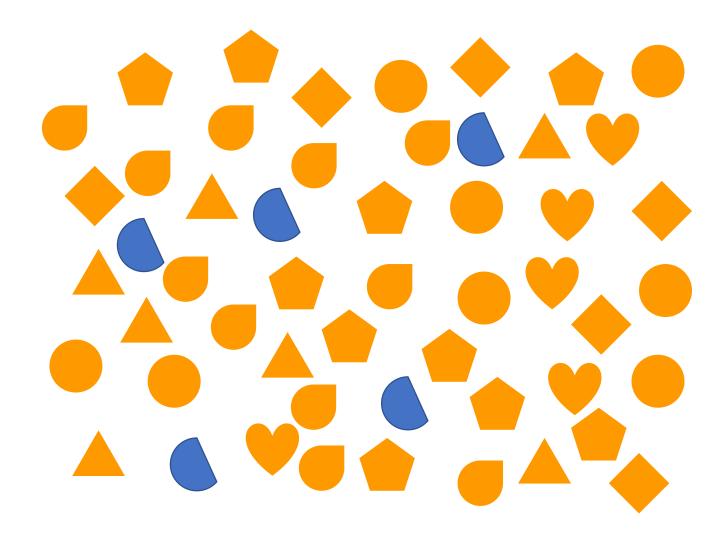


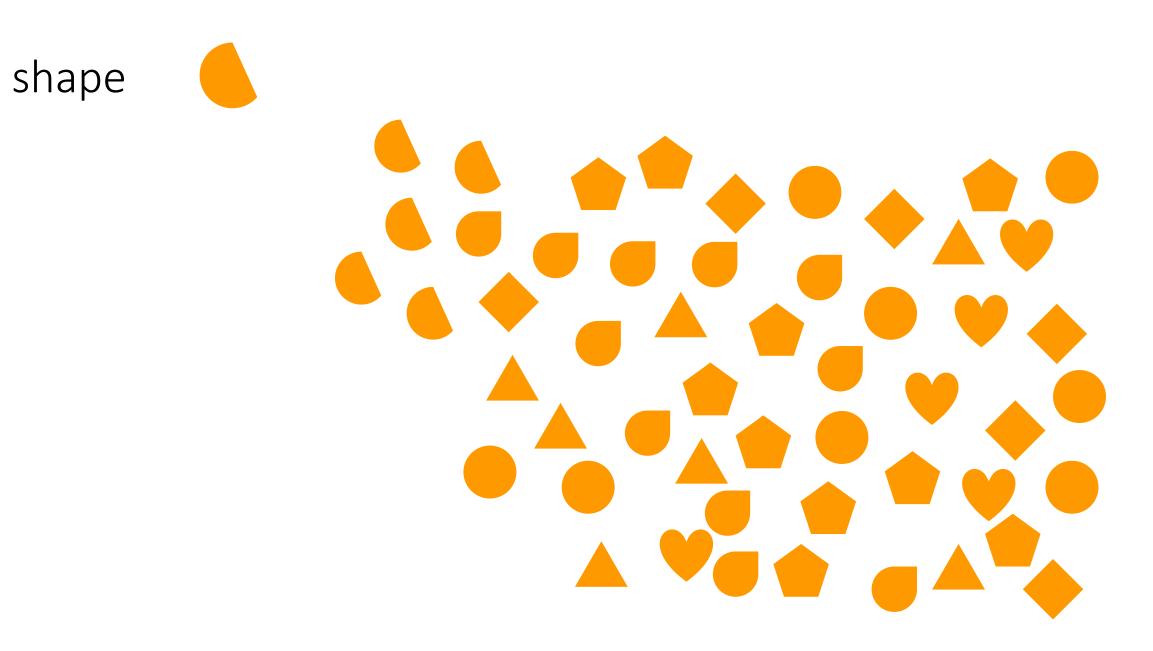


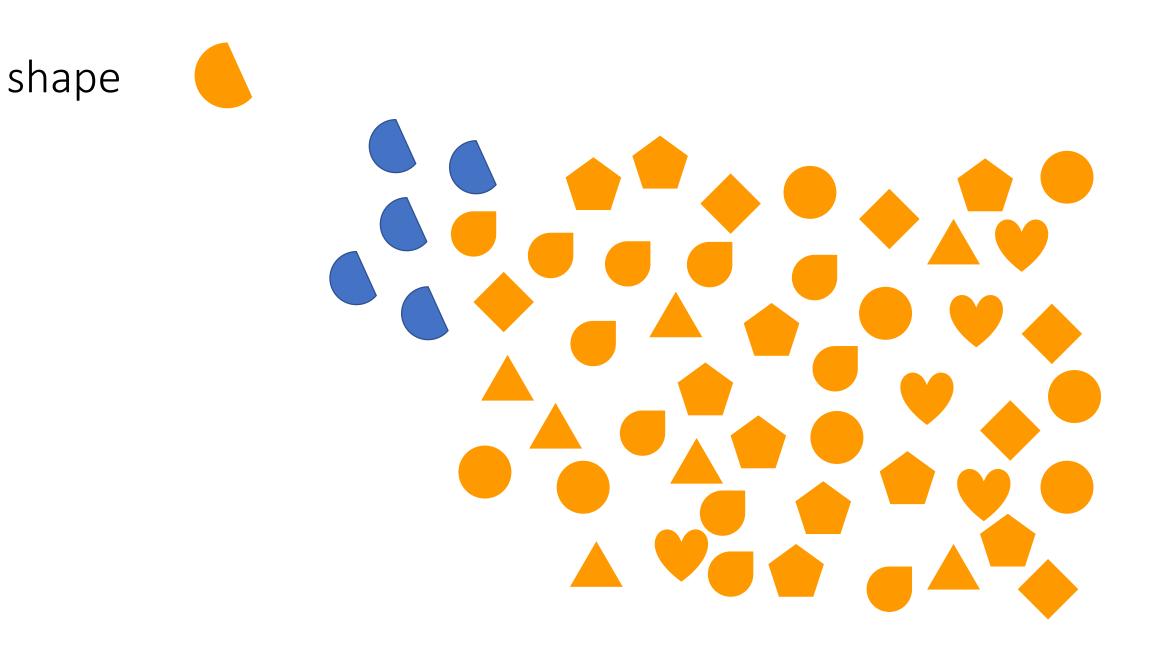




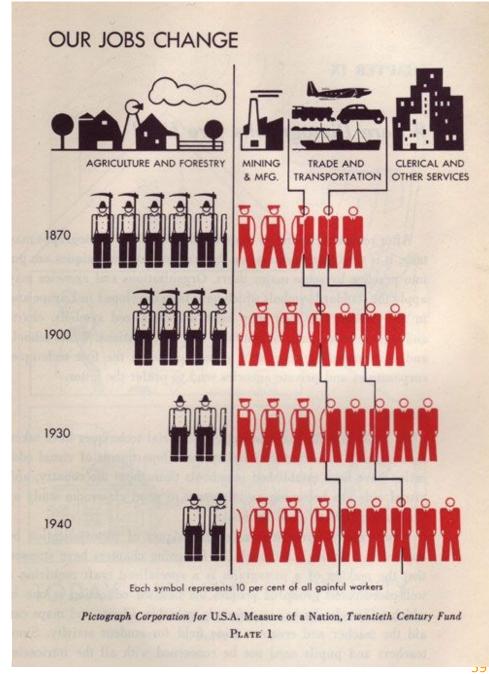








#### shape

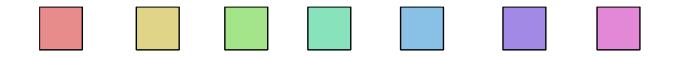


U.S. Job Change, Otto Neurath

#### 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?



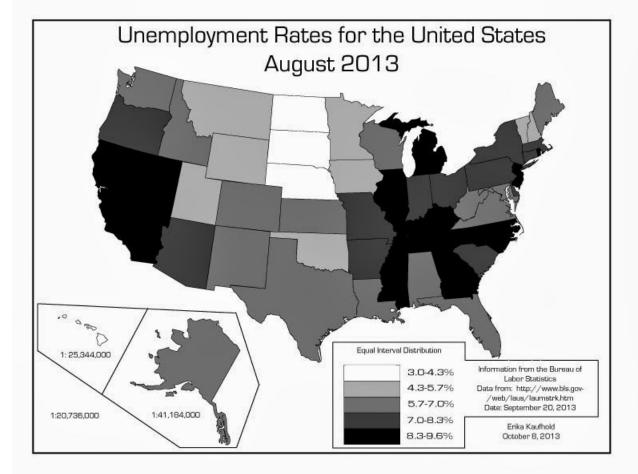
#### colour (value) ordered or quantitative?



# 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-choroplethmaps-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  $\rightarrow$  questions on Monday
- Find the visualization postcard assigned to you here (look for your student ID)

   <u>https://uh.host.cs.st-andrews.ac.uk/CS5044/P1\_visAssignments.htm</u>
- 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 – <u>https://studres.cs.st-andrews.ac.uk/CS5002/</u>
- 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 (2<sup>nd</sup> Ed.)
  - <u>https://alignedleft.com/work/d3-book-2e</u>

