Handling Data

SCC252/5836 Visualizaão Computacional

Profa. Maria Cristina

2º/2023

Content

- Data abstraction
- Data set types
- Attribute type
- From data to visual encoding

What is Data?

- Data as a fundamental component of visualization

- But what is data?
- How is data generated?
- Why do we use it? What is it for?

- Google `data´?

Definition

- https://www.merriam-webster.com/dictionary/data
- Data – Wikipedia

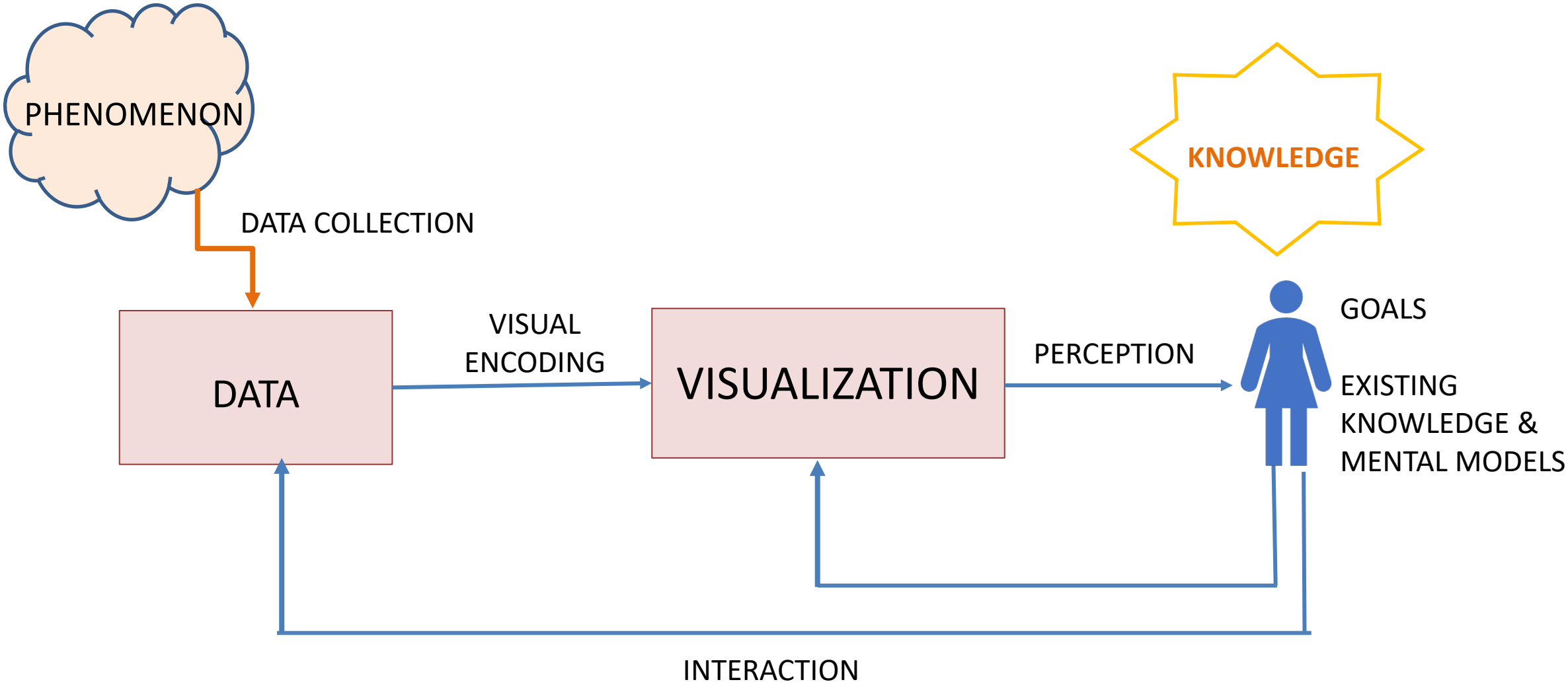
Definition

- https://www.merriam-webster.com/dictionary/data

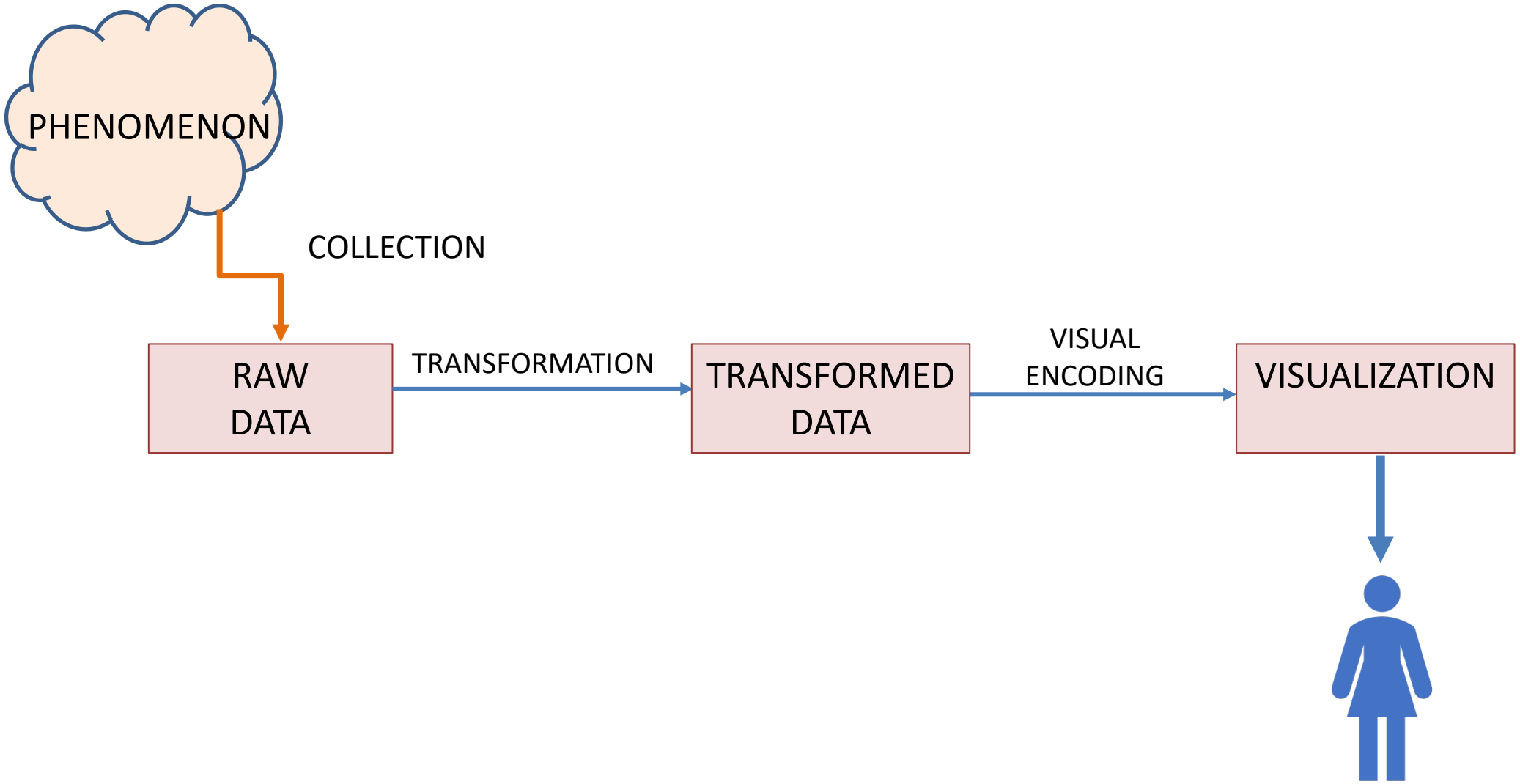
- Data – Wikipedia

- Data as a basis for reasoning, discussion and calculation
- Information recorded in a digital format and is useful for something...

Data in the visualization pipeline



Data in the visualization pipeline



Data in the visualization pipeline

- Data collection
- Data transformation
- Encoding: requires a decision on how to encode

- Ex.
  - Bar charts good for comparing numerical values relative to different categories
  - Colored map good for comparing density (proportion) per spatial area

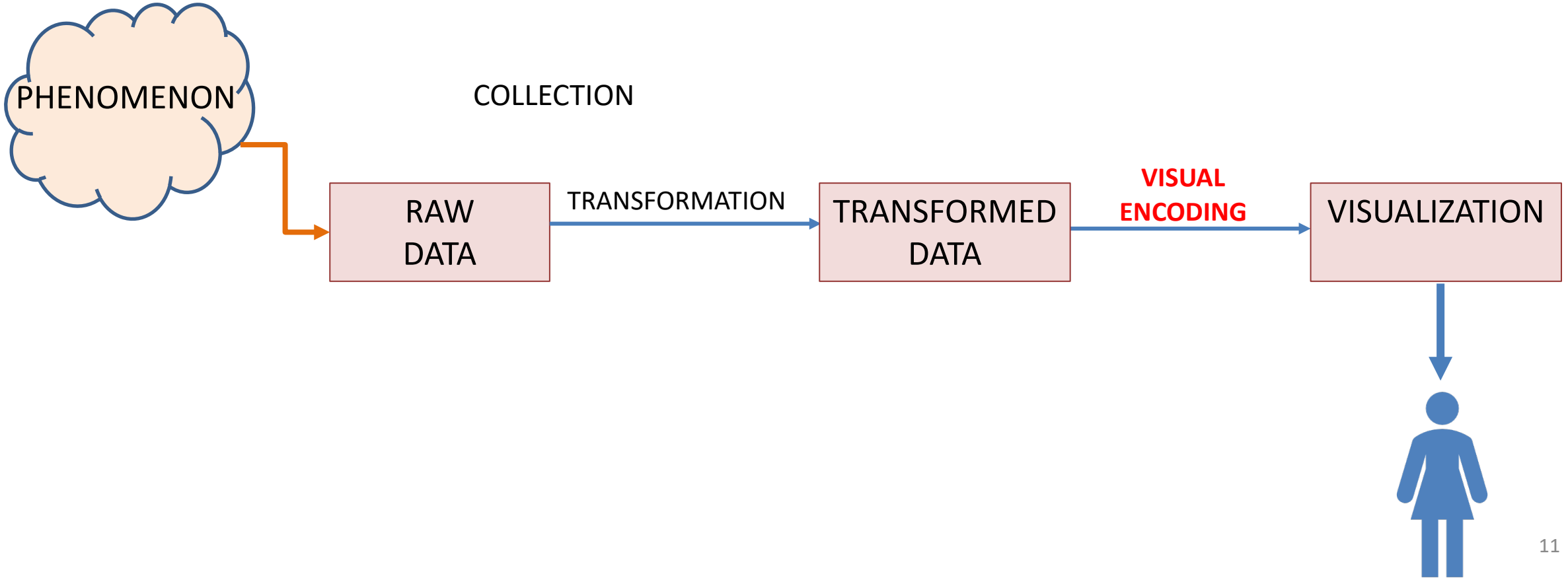
Data abstraction

- Ex.
  - Bar charts good for comparing numerical values relative to different categories
  - Colored map good for comparing density (proportion) per spatial area

- Knowing what type of of data you have determines the choices of what visual representations are available and appropriate.

Data abstraction

- Describe data in ways that help to decide what operations and encoding methods are **available** and **appropriate**



Data abstraction

… A way to recognize common structures/properties in data coming from very different domains

Hence the name “data abstraction”

abstract away from the domain…

Examples?

Example: Networks

- Friendship in Facebook
- Interactions between proteins
- Connections between criminals

- ➔ all networks!

Example: Spatial distribution

- Animal movements
- Election results by counties
- Simulation of airflow in na aircraft

- ➔ all spatial phenomena

Example: Spatial distribution

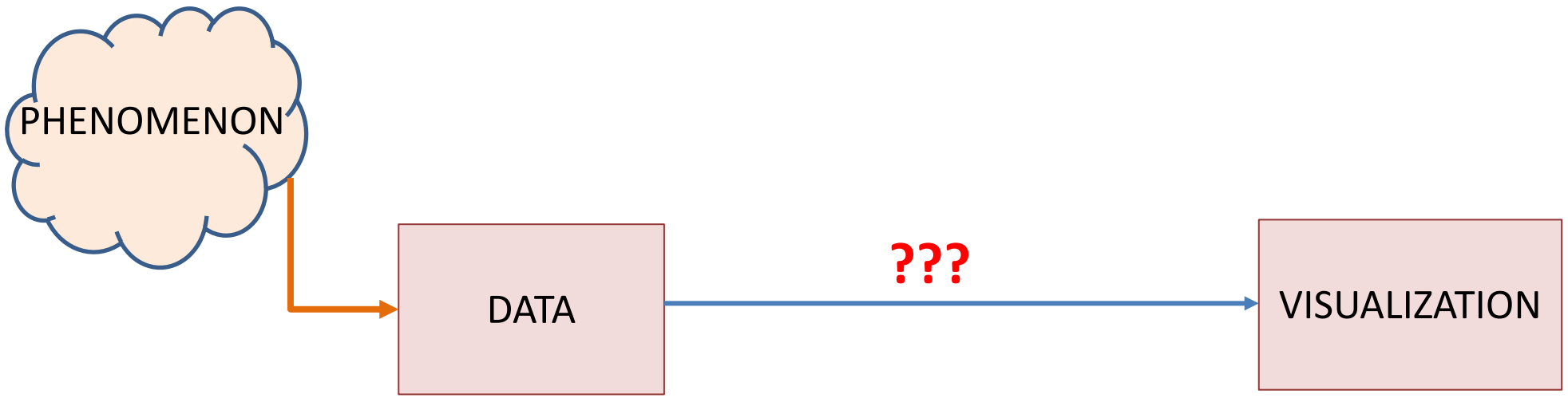
- Animal movements
- Election results by counties
- Simulation of airflow in na aircraft

- ➔ all spatial phenomena

Data abstraction: abstracting away from the domain the characteristics of the data that are useful to decide what visual representations are available and appropriate

Data abstraction

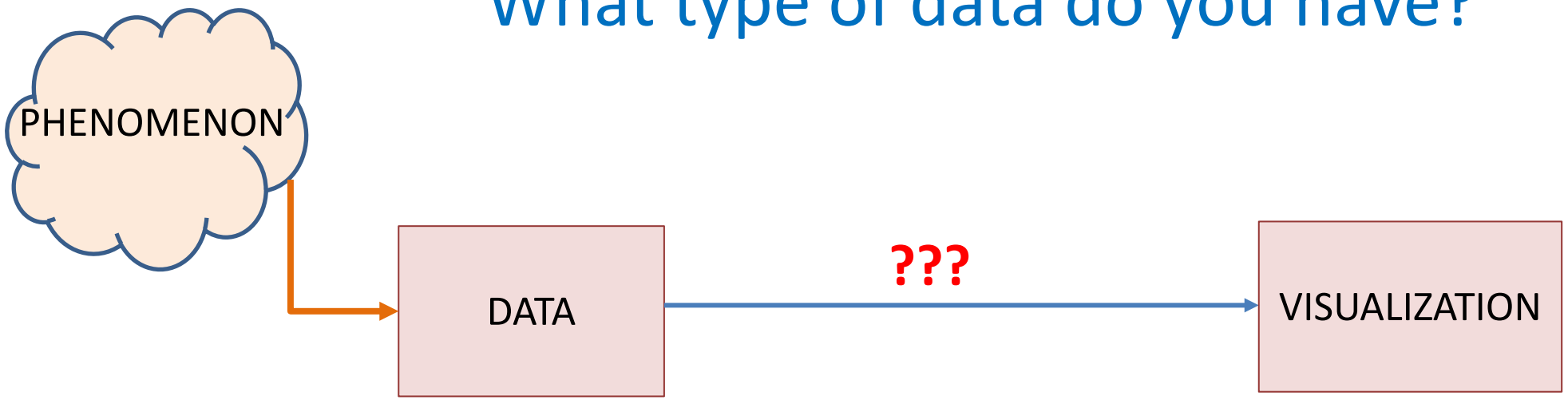
How can I visualize this data?



Data abstraction

How can I visualize this data?

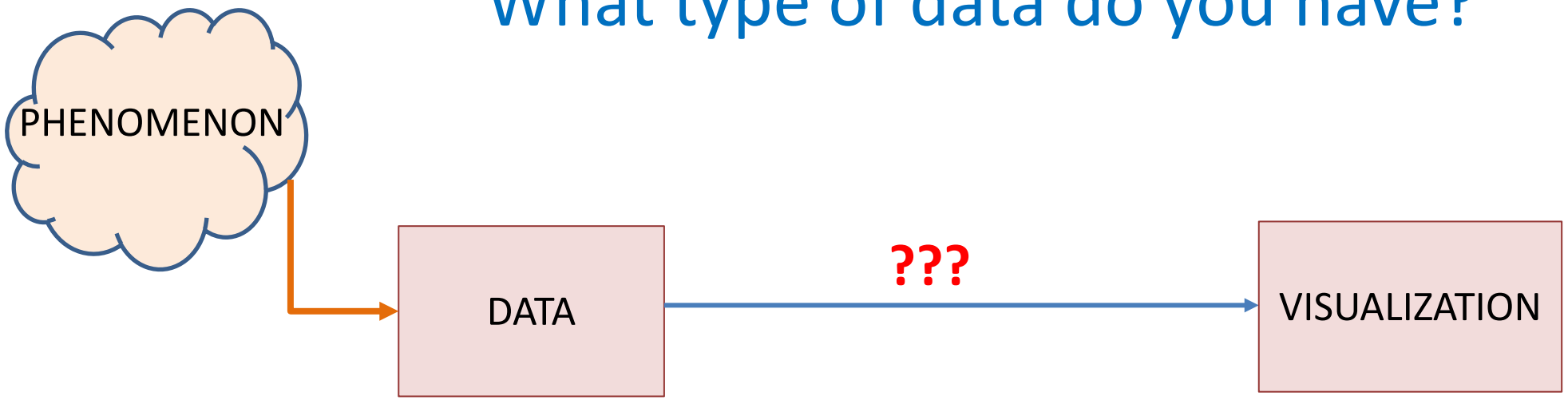
What type of data do you have?



Data abstraction

How can I visualize this data?

What type of data do you have?



Data abstraction first!

Data sets

- Data set = collection of items, described by multiple attributes

- Items = the objects/entities we want to visualize
- Attributes = properties of these objects/entities

- Items ~ records ~ instances ~ examples ~samples
- Attributes ~fields ~ features ~characteristics ~dimension

Examples

- 1 item describes a sales transaction... Attributes?

- 1 item describes a person...

Data set types

- Two major classes (in InfoVis , or in handling abstract data)
  - Tables
  - Networks

Tables

M	N	O	P	Q	R	S	T	U	V
City	Zip Code	State	Region	Customer Segment	Product Category	Product Sub-Category	Product Name	Product Container	Product Base Margin
Bowie	20715	Maryland	East	Home Office	Office Supplies	Storage & Organization	Safco Industrial Wire Shelving	Large Box	
McKeesport	15131	Pennsylvania	East	Small Business	Office Supplies	Storage & Organization	Perma STOR-ALL Hanging File Box, 13 1/8"W x 12 1/4"D x 10 1/2"H	Small Box	0,68
Prior Lake	55372	Minnesota	Central	Consumer	Office Supplies	Binders and Binder Accessories	Avery Trapezoid Ring Binder, 3" Capacity, Black, 1040 sheets	Small Box	0,36
Napa	94559	California	West	Corporate	Furniture	Chairs & Chairmats	Hon 4070 Series Pagoda Armless Upholstered Stacking Chairs	Jumbo Drum	0,56
Napa	94559	California	West	Corporate	Furniture	Chairs & Chairmats	Hon Valutask Swivel Chairs	Jumbo Drum	0,69
Napa	94559	California	West	Corporate	Office Supplies	Storage & Organization	Dual Level, Single-Width Filing Carts	Small Box	0,59
Montebello	90640	California	West	Consumer	Office Supplies	Paper	Black Print Carbonless Snap-Off® Rapid Letter, 8 1/2" x 7"	Wrap Bag	0,4
Montebello	90640	California	West	Consumer	Office Supplies	Paper	White GlueTop Scratch Pads	Wrap Bag	0,39
Lake Oswego	97035	Oregon	West	Home Office	Office Supplies	Pens & Art Supplies	Dixon Prang® Watercolor Pencils, 10-Color Set with Brush	Wrap Bag	0,44
Baton Rouge	70802	Louisiana	South	Corporate	Technology	Computer Peripherals	Fellowes Mobile Numeric Keypad, Graphite	Small Box	0,66
Baton Rouge	70802	Louisiana	South	Corporate	Technology	Office Machines	Polycom ViaVideo Desktop Video Communications Unit	Large Box	0,37
Baton Rouge	70802	Louisiana	South	Corporate	Office Supplies	Paper	Staples Wirebound Steno Books, 6" x 9", 12/Pack	Wrap Bag	0,36
Draper	84020	Utah	West	Home Office	Furniture	Office Furnishings	36X48 HARDFLOOR CHAIRMAT	Medium Box	0,78
Layton	84041	Utah	West	Corporate	Technology	Office Machines	Panasonic KX-P2130 Dot Matrix Printer	Jumbo Drum	0,59
Kearney	68847	Nebraska	Central	Corporate	Office Supplies	Paper	Xerox 4200 Series MultiUse Premium Copy Paper (20Lb. and 84 Bright)	Small Box	0,4
Kearney	68847	Nebraska	Central	Corporate	Technology	Telephones and Communication	T18	Small Box	0,57
Phenix City	36869	Alabama	South	Home Office	Office Supplies	Binders and Binder Accessories	Avery Heavy-Duty EZD I Binder With Locking Rings	Small Box	0,37
Phenix City	36869	Alabama	South	Home Office	Furniture	Tables	Bush Advantage Collection® Round Conference Table	Jumbo Box	0,64
Pharr	78577	Texas	Central	Corporate	Office Supplies	Binders and Binder Accessories	Cardinal Poly Pocket Divider Pockets for Ring Binders	Small Box	0,4
Pharr	78577	Texas	Central	Corporate	Office Supplies	Paper	"While you Were Out" Message Book, One Form per Page	Wrap Bag	0,35
Bryan	77803	Texas	Central	Home Office	Furniture	Chairs & Chairmats	Office Star - Professional Matrix Back Chair with 2-to-1 Synchro Tilt and Mes	Jumbo Drum	0,61
Deer Park	11729	New York	East	Home Office	Office Supplies	Paper	Rediform Wirebound "Phone Memo" Message Book, 11 x 5-3/4	Wrap Bag	0,36
Fort Dodge	50501	Iowa	Central	Consumer	Technology	Computer Peripherals	Memorex 4.7GB DVD+RW, 3/Pack	Small Pack	0,4
Fort Dodge	50501	Iowa	Central	Consumer	Technology	Telephones and Communication	3285	Small Box	0,59
Ormond Beach	32174	Florida	South	Corporate	Office Supplies	Storage & Organization	Perma STOR-ALL Hanging File Box, 13 1/8"W x 12 1/4"D x 10 1/2"H	Small Box	0,68
Norman	73071	Oklahoma	Central	Small Business	Furniture	Office Furnishings	Electrix Halogen Magnifier Lamp	Large Box	0,59
Norman	73071	Oklahoma	Central	Small Business	Furniture	Office Furnishings	Luxo Professional Fluorescent Magnifier Lamp with Clamp-Mount Base	Large Box	0,59
Norman	73071	Oklahoma	Central	Small Business	Technology	Office Machines	Panasonic KX-P1150 Dot Matrix Printer	Jumbo Drum	0,56
Dubuque	52001	Iowa	Central	Corporate	Furniture	Chairs & Chairmats	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	Jumbo Drum	0,55
Mount Prospect	60056	Illinois	Central	Home Office	Technology	Office Machines	Canon S750 Color Inkjet Printer	Jumbo Drum	0,38
Bedford	76021	Texas	Central	Consumer	Furniture	Chairs & Chairmats	Metal Folding Chairs, Beige, 4/Carton	Jumbo Drum	0,58
Bedford	76021	Texas	Central	Consumer	Furniture	Chairs & Chairmats	Hon 4700 Series Mobuis Mid-Back Task Chairs with Adjustable Arms	Jumbo Drum	0,64
Bedford	76021	Texas	Central	Consumer	Furniture	Office Furnishings	Telescoping Adjustable Floor Lamp	Large Box	0,6
Bedford	76021	Texas	Central	Consumer	Furniture	Tables	Iceberg OfficeWorks 42" Round Tables	Jumbo Box	0,7
Bedford	76021	Texas	Central	Consumer	Technology	Telephones and Communication	232	Small Box	0,55
Monroe	71203	Louisiana	South	Home Office	Office Supplies	Paper	IBM Multi-Purpose Copy Paper, 8 1/2 x 11", Case	Small Box	0,4
Royal Palm Beach	33411	Florida	South	Consumer	Office Supplies	Binders and Binder Accessories	GBC VeloBinder Strips	Small Box	0,35
Royal Palm Beach	33411	Florida	South	Consumer	Furniture	Office Furnishings	G.E. Longer-Life Indoor Recessed Floodlight Bulbs	Small Pack	0,37
Hilton Head Island	29915	South Carolina	South	Consumer	Furniture	Chairs & Chairmats	Novimex Swivel Fabric Task Chair	Jumbo Drum	0,74

Items x Atributes: items in rows, attributes in columns

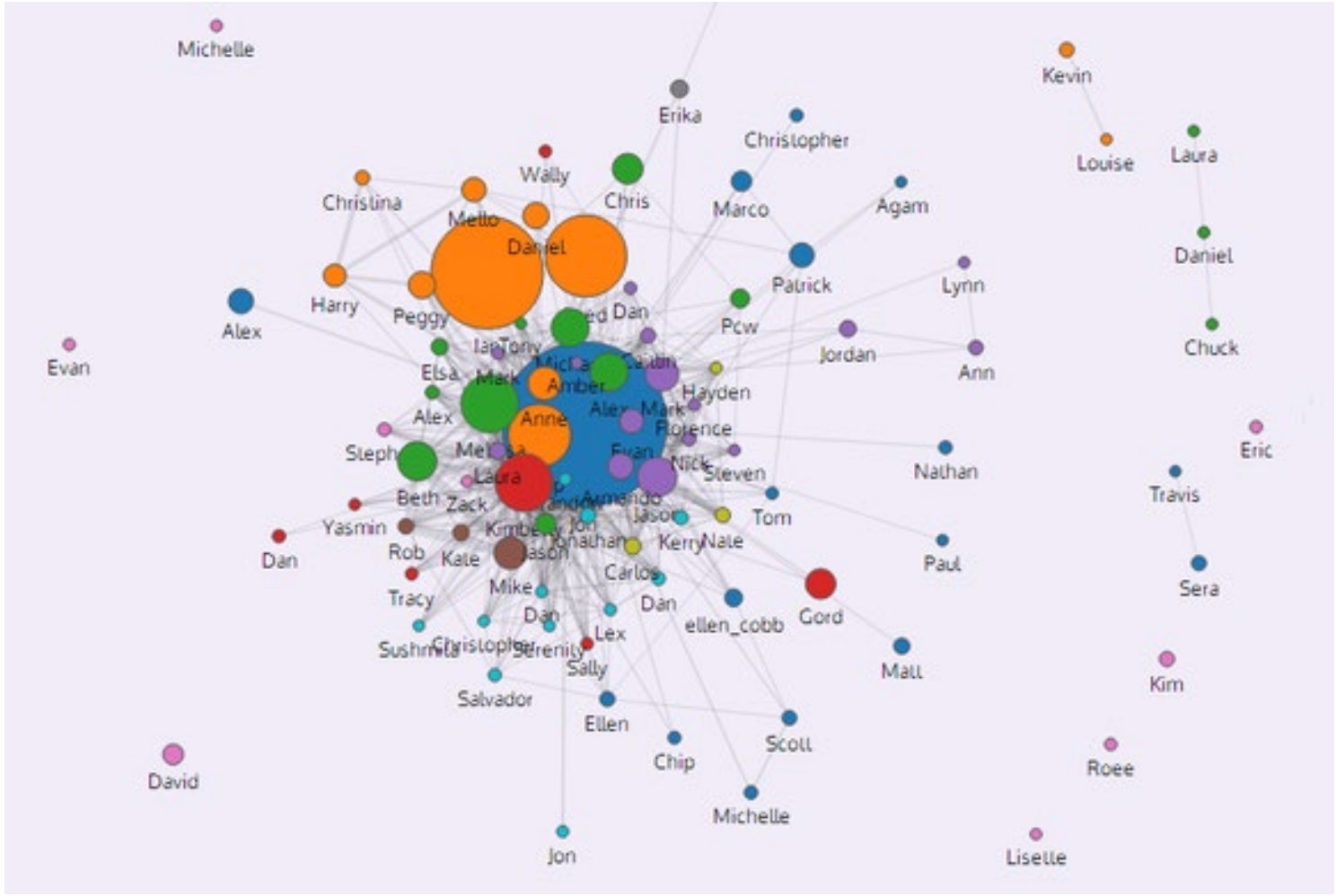
Tables

- Grid of items x attributes

- 1 row = 1 item
- 1 column = 1 attribute

- All items described by the same set of attributes
- (apart from missing values)

Networks & Trees



Nodes and links

Image from MIT´s Immersion project

https://www.media.mit.edu/projects/immersion

Networks & Trees

- Objects/entities are connected by links
- Objects/entities called nodes, connections called links

- Can also have atributes:
  - Items (nodes) can be described by multiple attributes
  - Links can also be described by multiple attributes

Exercise

- Try to think of two example datasets, one that is configured a table and one that is configured as a network

Data set types

- Others

- See Fig. 2.2 book by Munzner

Data abstraction

- First step: identify the data set type
- Next step: identify the attributes´ types
  - What type of attributes we have?

Attribute types

- Categorical: values correspond to categories, no intrinsic order
- Ordinal: values correspond to categories, with an intrinsic order
  - `distance´ between the categories not known
  - arithmetic operations between categories are not applicable
- Quantitative

Attribute types

- Categorical: values correspond to categories, no intrinsic order
- Ordinal: values correspond to categories, with an intrinsic order
- Quantitative: values represent some measured quantity
  - `Distance´ between values can be computed
  - Can perform arithmetic operations with values

Attribute types

- Categorical: values correspond to categories, no intrinsic order
- Ordinal: values correspond to categories, with an intrinsic order
- Quantitative: values represent some measured quantity
- See Fig. 2.4 book by Munzner

Example

- Product Sales data set

- Row = item = 1 order
- Columns = attributes of the order

Example

- Product category
- Order priority
- Sales

- Order ID
- Ship mode

Attribute semantics

- Refers to the `meaning´ of some attributes
- There are some predefined semantics in the data that it is useful to identify early, in particular

- Spatial semantics: attributes with spatial properties (identify a place, or a region)

- Temporal semantics: attributes that describe information related with time

Example

- Product Sales data set

- Region
- Latitude & Longitude

- Order data

Attribute semantics

- Sequential, Diverging, Cyclic
- See Fig. 2.4 book by Munzner
- Examples
  - Month of order
  - Profit

Attribute semantics

- Attributes may be Hierarchical

- Ex. produc category and product sub-category

Exercise

- Let us identify the type and characteristics of other attributes
  - Customer segment
  - City, State, Region
  - Order ID (tricky!)
  - Product container (also tricky)

Data abstraction in summary

- Data set types: Tables, Networks, Trees
- Attribute types: categorical, ordinal, quantitative
- Additional attribute characteristics:
  - Spatial, temporal
  - Sequential, diverging, cyclic
  - Hierarchical

From data abstraction to visualization

- Why it is useful to attribute types?

- Knowing the nature of the data attribute gives you guidance on selecting the appropriate visual representations

From data abstraction to visualization

- Example: it's inappropriate to use a line chart to represent unordered attributes.

- Ex. Understanding and using Line Charts | Tableau

From data abstraction to visualization

- Example: bar charts are appropriate to display information about categories and frequencies or statistics related with these categories.

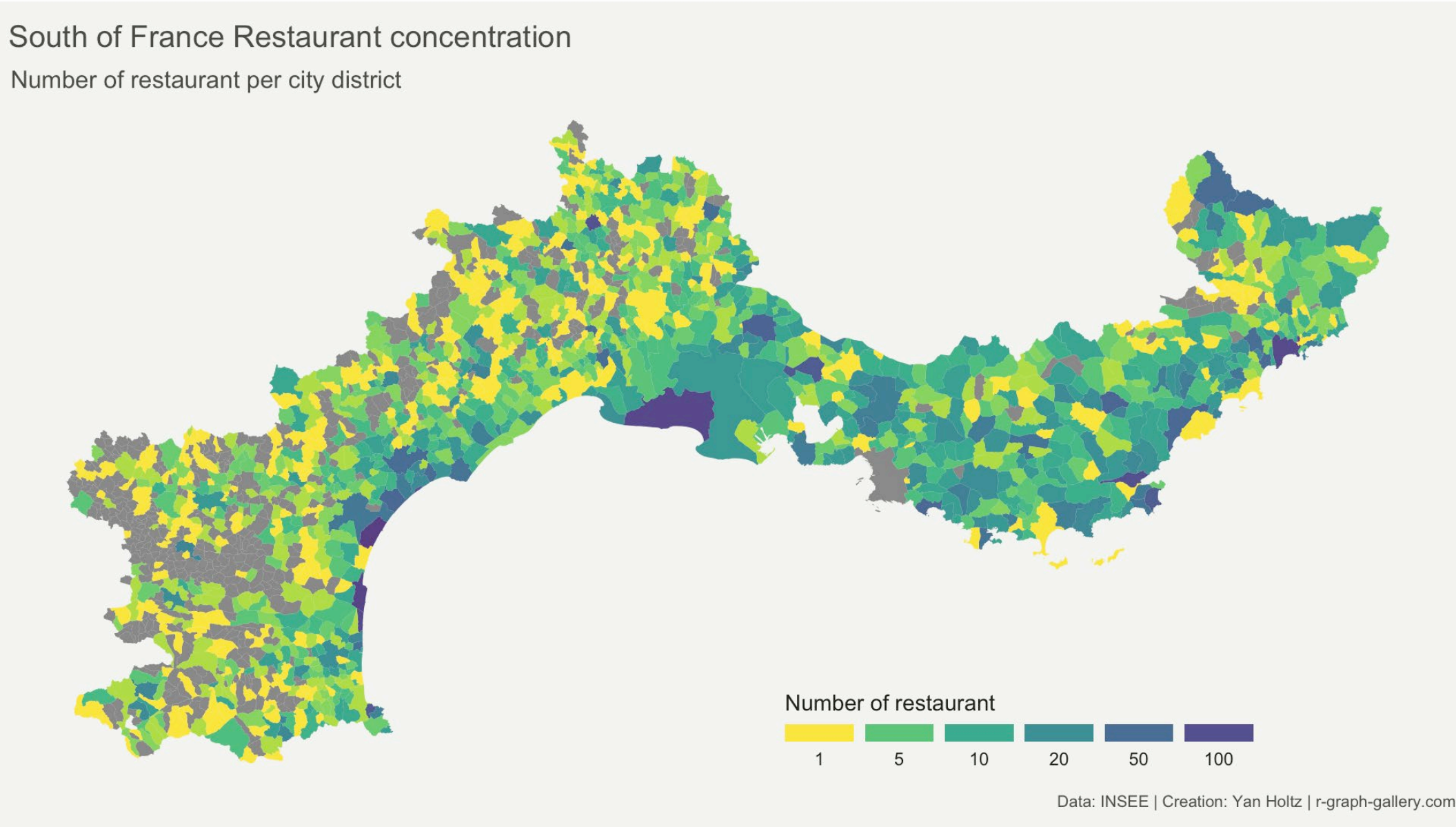
- Ex. Understanding and Using Bar Charts | Tableau

- Bar charts can be reordered if they show categorical data
- But not if they show ordinal data

From data abstraction to visualization

- Example: spatial attributes can be visualized with spatial metaphors

- E.g., maps



[Choropleth map – from Data to Viz (data-to-viz.com)]

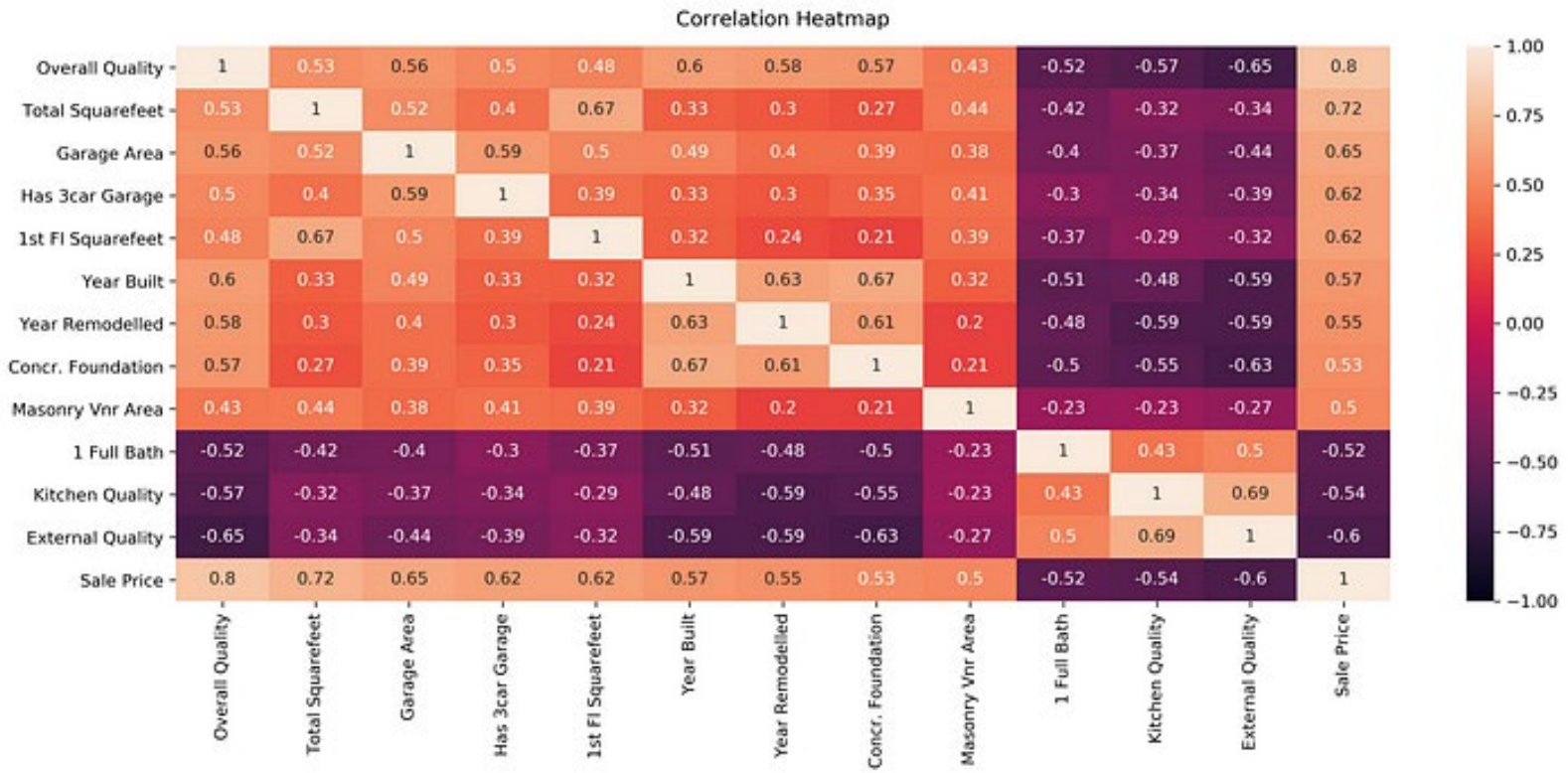
From data abstraction to visualization

- Example: quantitative diverging attribute

- Color mapping to convey quantitative attributes
- Choice of color scale to emphasize sequential or diverging nature

From data abstraction to visualization

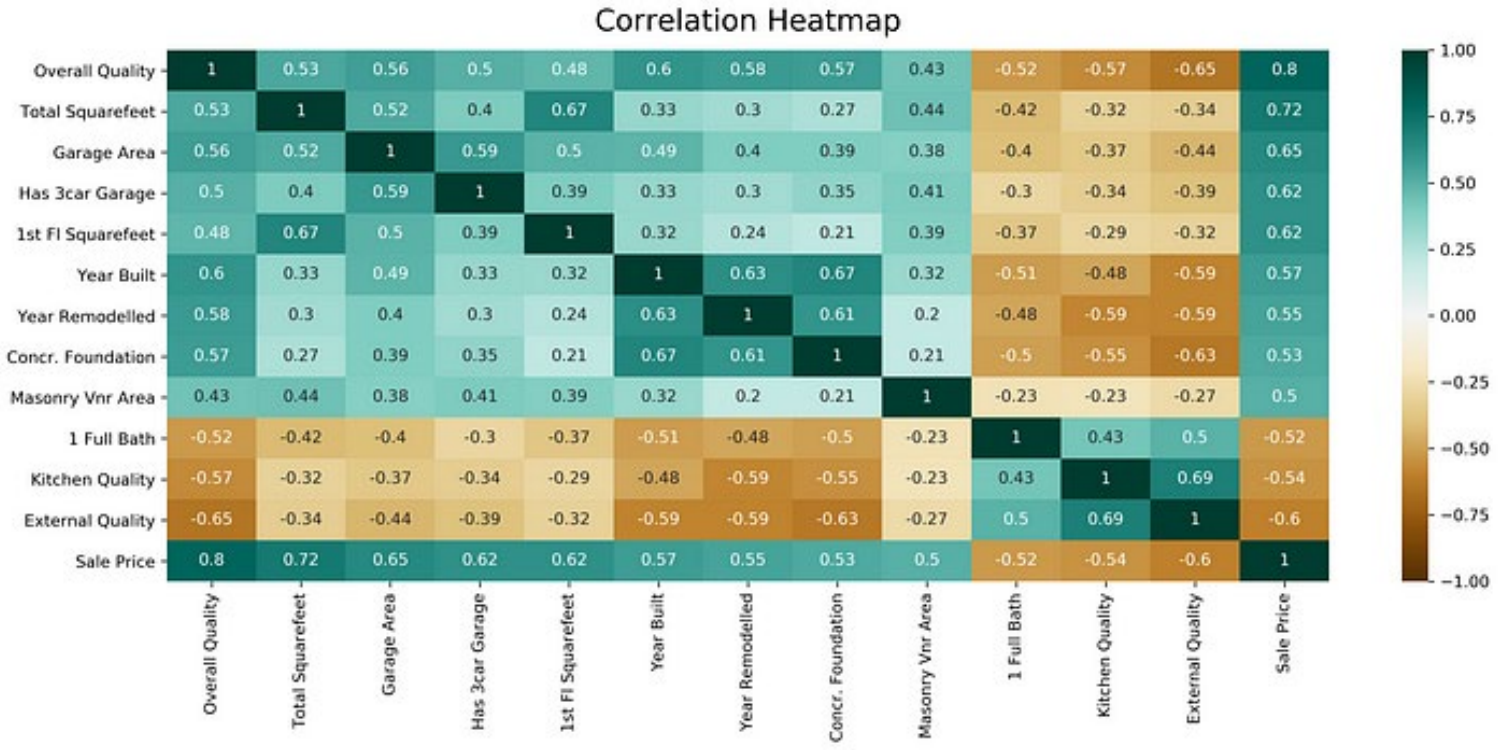
- Example: quantitative diverging attribute, sequential color map



[How to Create a Seaborn Correlation Heatmap in Python? | by Bibor Szabo | Medium]

From data abstraction to visualization

- Example: quantitative diverging attribute, divergent color map



[How to Create a Seaborn Correlation Heatmap in Python? | by Bibor Szabo | Medium]

From data abstraction to visualization

- Example: quantitative diverging attribute

When to use sequential and when to use diverging color scales - Datawrapper Blog

Data profiling

- It is essential to know the data, familiarize yourself with the data
  - Data coming from different sources
  - Meaning of data often not evident
  - Talk with your `client´/user of the visualization
  - Effort and interaction required to familiarize with the data

Data profiling

- Performing data abstraction is a very useful exercise to make appropriate decisions when deciding on how to visually encode the data!

- It is not sufficient, though...

Data quality

- Checking data quality: very important and time consuming

- Data `wrangling´
  - Data may not be ready for analysis
  - Errors, missing values
  - Need to fix problems, perhaps transform data (data manipulation)

  - Not covered in this course

Sources/material

T. Munzner, Visualization Analysis & Design

Information Visualization Fundamentals, online course in Coursera