Data Visualization & Design

Week 4

Announcements

Columbia's Group for Experimental Methods in the Humanities & Columbia University Libraries ask you to join us for

Puerto Rico Mapathon

for Hurricane Relief

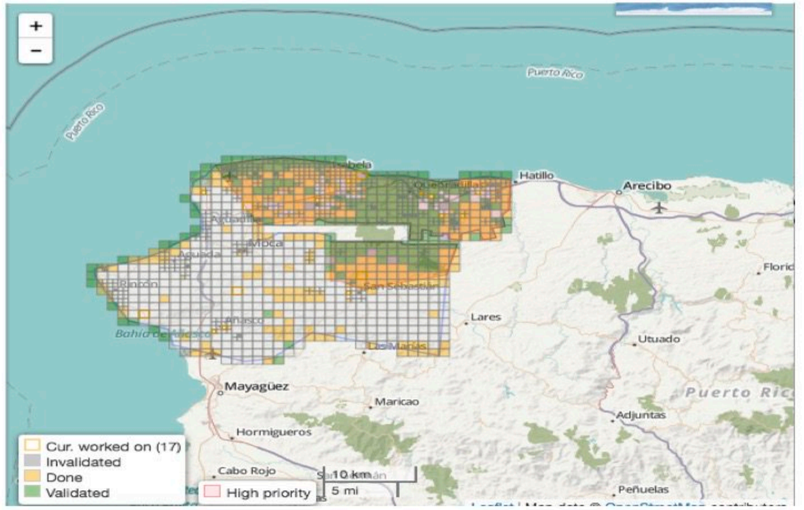
September 29, 2017 | 2PM-5PM | Studio@Butler (208B)

Come help with relief efforts on the ground in Puerto Rico by contributing your time to open-source mapping.

Following the recent hurricane, people around the world are using the OpenStreetMap platform to give their time to hurricane relief efforts. The Red Cross in Puerto Rico has requested two tasks we can help with to help with their relief efforts. During the mapathon, we will teach people how to help with these efforts through mapping, and we will map together.

No mapping experience necessary. No knowledge of local terrain is necessary. Come any time during the afternoon.

Please bring your laptop & RSVP to agil@columbia.edu

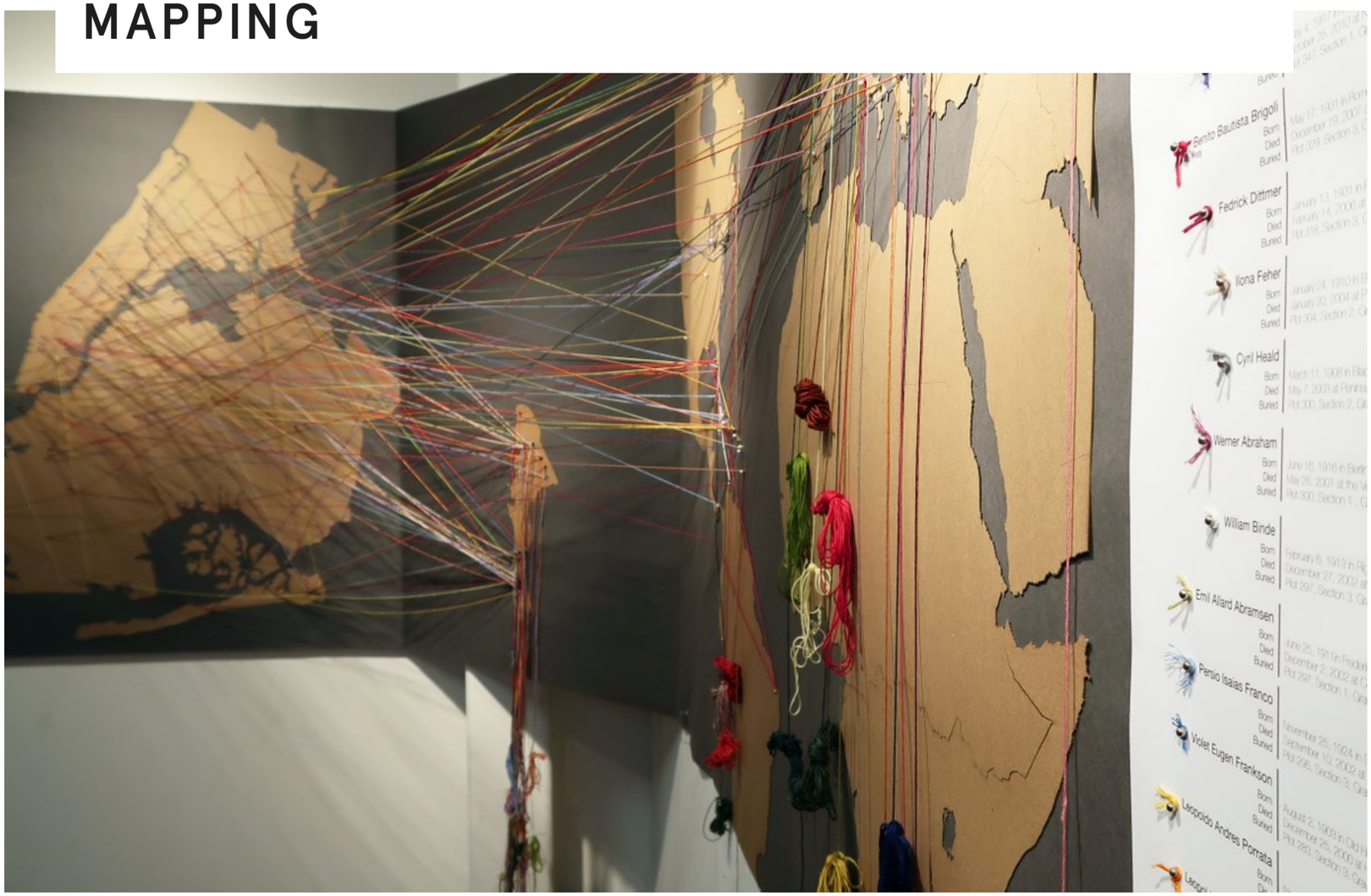


This week in **visualization —**


Searches for fever
in the United States
Symptom
I'm not
feeling well
Temperature
To everything there is a season, and diseases are no different. Google search patterns show us what health issues people worry about, when they worry about them, and how epidemics spread through the world.
Gabriel Gianordoli for Google News Lab
Toddler

Source

YOU ARE HERE NYC: ART, INFORMATION, AND MAPPING



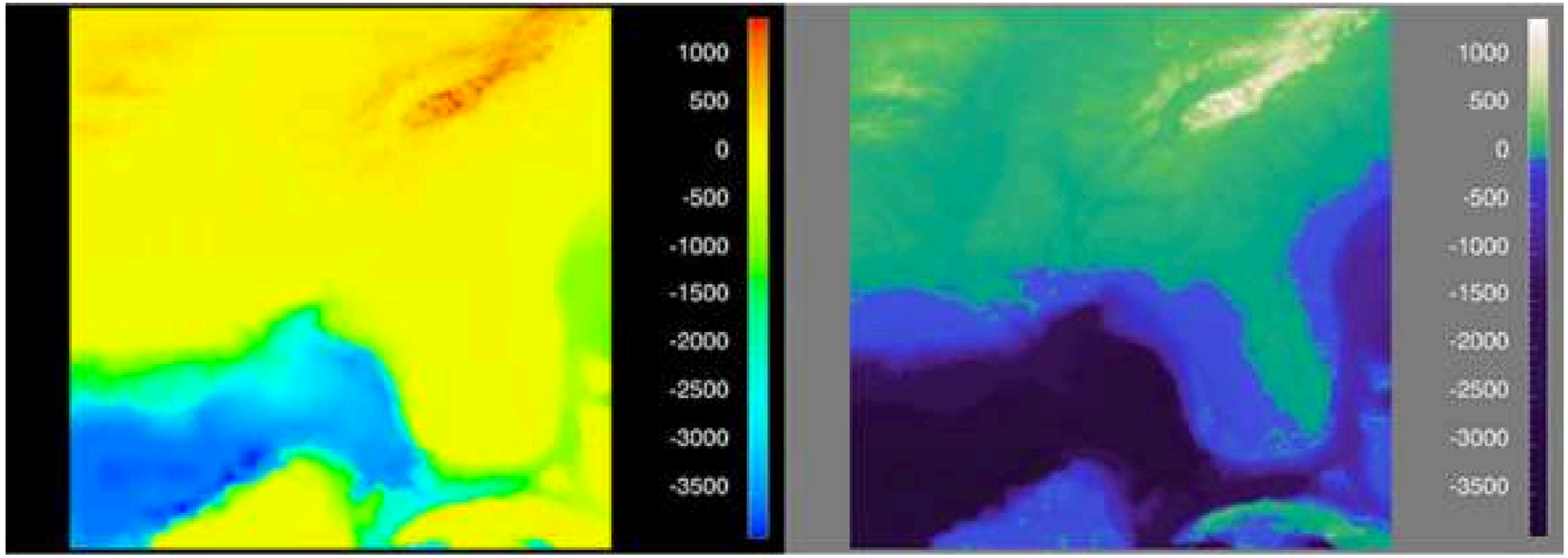
[Source]

Integrity

A violation of integrity may be **accidental** or **intentional**.

Accidental

- Sloppy, lazy (not careful about data cleaning)
- Un-knowing (blindly trusting a default or template)



The choice of color scale can inadvertently distort the structure in the data. *(Rogowitz and Treinish, 1996)*

Intentional

- Removing or selecting data
- Using others’ work as your own

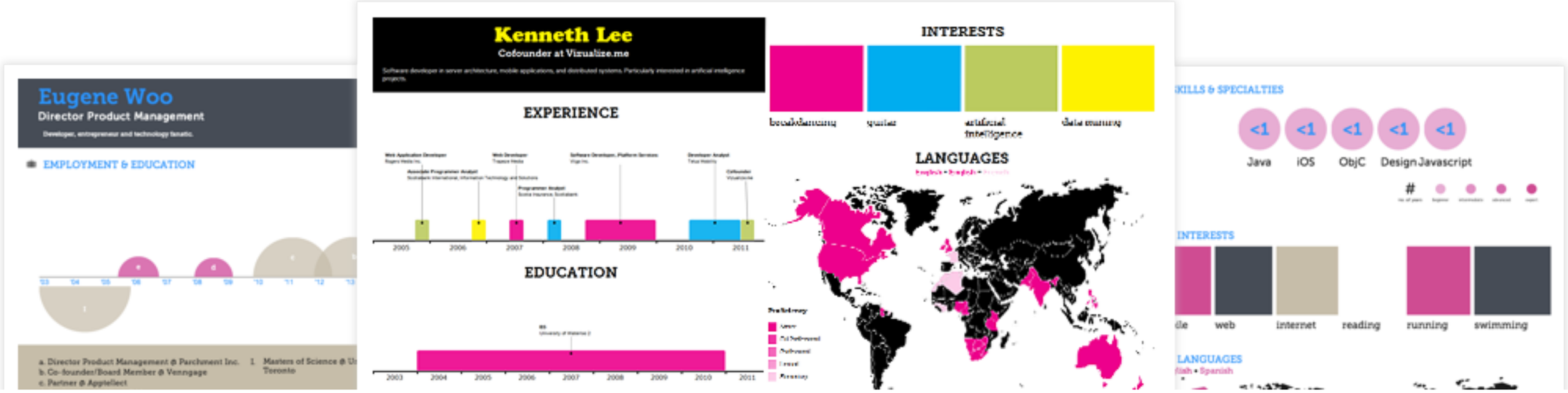
What is the line between **storytelling** and **lying**?

What is the line between **storytelling** and **lying**?

…what about **advocacy** and **propaganda**?

Using templates — **ResumUP.com**

Using templates — **Visualize.me**



If you use a component of a visualization generator, **credit the source** and use it thoughtfully in your context.

Guidelines

- ***Do*** credit all resources
- ***Do*** use open-source libraries and tools
- ***Do*** gather inspiration from others' projects and studies
- ***Do not*** use a prepackaged template wholesale, as though it was your own creation
- ***Do not*** copy and paste titles or charts into your projects
- ***Do not*** pass off others' work as your own

“Is this 100% my work?”

— When in doubt, ***cite*****.**

Munzner —

Data Abstraction

Datasets

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Type refers to the ***structural or mathematical*** interpretation of the data.

What kind of thing is it?

Datasets

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Datasets

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

How do these combine to form a larger structure?

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

What kinds of mathematical operations are meaningful for it?

Datasets

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Example — **“Quantity” vs. “Code”**

11222

If this value is a ***count of boxes of detergent*** **—**

- Type = **quantity**
- Adding two quantities together makes sense
- Informs appropriate visualization method (bar chart, line chart, etc.)

Example — **“Quantity” vs. “Code”**

11222

If this value is a ***zip code*** **—**

- Type = **code**
- Adding two codes together *does not* make sense
- Informs appropriate visualization method (point map, etc.)

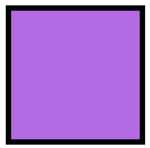
Semantics refers to the ***real-world meaning*** of the data.

- Human name?
- Company name?
- Abbreviated company name?
- Fruit?
- Age?
- Day of the month?

Apple

Orange

Pear



Strawberry

Melon

Semantic meaning — ***Human name***

	A	B	C	D	E	F	G	H
1	**Name**	**University**	**Major**	**Minor**	**Work Experience**	**Internship**	**#Internships**	**Visualization Experience**
2	Jiaqi Wang	Penn State University	Human Resources	Economics		Yes	3	Yes
3	Justin Yoon	University of Washington	Economics			Yes	1	Yes
4	Ain UI Mardhiah Rosli	Univerisyt of Sydney	Economics		4.5			Yes
5	Michael Li	City University of Hong Kong	Quantitative Finance & Risk Management			Yes	1	Yes
6	Leo Prashant Shelley		Doctor of Medicine		15			No
7	Danlu Pan	Hang Zhou, China	Applied Mathematics			Yes	1	Yes
8	Jingyun Xu	University of Miama	Finance and Mathematics					Yes
9	Pavlos Giannakis	Columbia University			7			No
10	Weiton Zhou	Shanghai Jiaotong University	Finance			Yes	2	Yes
11	Lu Qiao	Wuhan Univeristy	Financial Engineering			Yes	2	Yes
12	Yajie Yu	University of Oregon	Business Administration			Yes	1	No
13	Zhou Wang	UC San Diego	Management Science	Mathematics		Yes	3	No
14	Akashna Anand	Nirma University	MBA		2			Yes
15	Sifan Du	University of Miami	Information systems			Yes	2	No
16	Yifan Ni	UC San Diego	Management Science	Business		Yes	2	Yes
17	Emma Du	College of William and Mary	Economics		1			Yes
18	Sui Ping Suen	Rensselaer Polytechnic Institue	Business & Mathematics			Yes	2	Yes
19								

In a table, the **column name** typically provides the semantic meaning of an attribute.

Attribute types

Physical vs. **Abstract**

Attribute type — **Physical**

- Characterized by storage format and machine operations
- Ex. bool, int, float, string…

Attribute type — **Abstract**

- Provide descriptions of data
- Characterized by methods and attributes
- May be organized into a *hierarchy*

Qualitative

- Nominal
- Ordinal

Quantitative

- Interval
- Ratio

…an aside on **hierarchies**

- Possible for any abstract data type
- Often, abstract data types carry strong implicit hierarchies

Implicit hierarchy example — **Geography**

- Postal code
- City district
- City
- State
- Country
- Continent

Implicit hierarchy example — **Time**

- Minute
- Hour
- Day
- Week
- Month

Implicit hierarchy example — **Time**

- Can be seen as **ordinal** (entries in a diary)
- Can be seen as **quantitative** (timings in a race)
    - **Interval** vs. **Ratio** — **Timestamp** vs. **Duration**

Data types

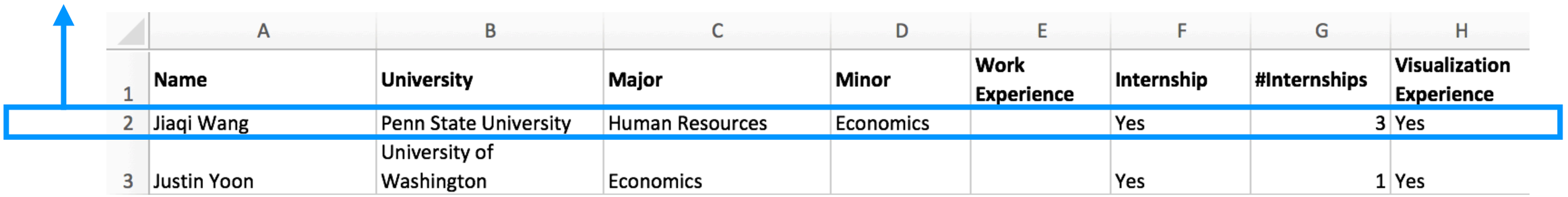
Data types

- Item
- Attribute
- Link
- Position

Data type — **Item**

- An individual entity that is discrete, such as a row in a table or a node in a network.

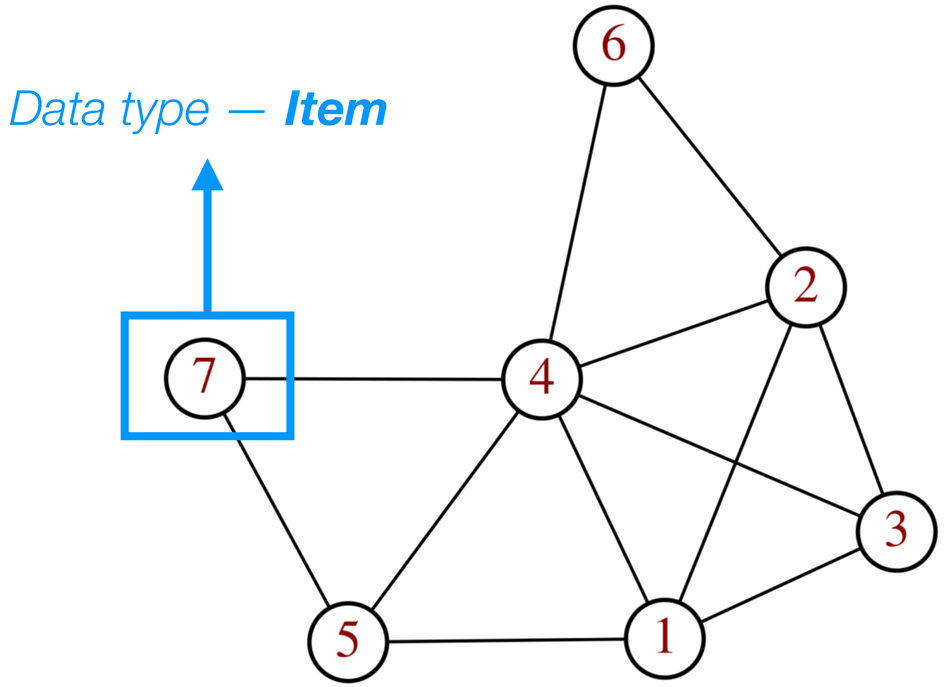
*Data type — **Item***



	A	B	C	D	E	F	G	H
1	**Name**	**University**	**Major**	**Minor**	**Work Experience**	**Internship**	**#Internships**	**Visualization Experience**
2	Jiaqi Wang	Penn State University	Human Resources	Economics		Yes	3	Yes
3	Justin Yoon	University of Washington	Economics			Yes	1	Yes

Data type — **Item**

- An individual entity that is discrete, such as a row in a table or a node in a network.



Data type — **Attribute (variable)**

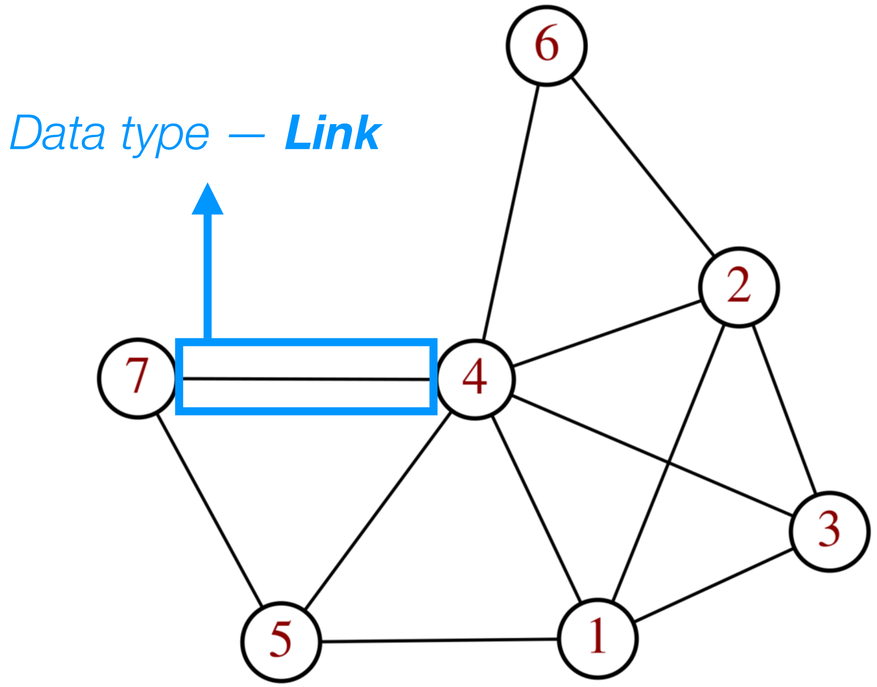
- Some specific property that can be measured, observed, or logged

*Data type — **Attribute***

	A	B	C	D	E	F	G	H
1	**Name**	**University**	**Major**	**Minor**	**Work Experience**	**Internship**	**#Internships**	**Visualization Experience**
2	Jiaqi Wang	Penn State University	Human Resources	Economics		Yes	3	Yes
3	Justin Yoon	University of Washington	Economics			Yes	1	Yes
4	Ain UI Mardhiah Rosli	Univerisyt of Sydney	Economics		4.5			Yes
5	Michael Li	City University of Hong Kong	Quantitative Finance & Risk Management			Yes	1	Yes
6	Leo Prashant Shelley		Doctor of Medicine		15			No
7	Danlu Pan	Hang Zhou, China	Applied Mathematics			Yes	1	Yes
8	Jingyun Xu	University of Miama	Finance and Mathematics					Yes
9	Pavlos Giannakis	Columbia University			7			No

Data type — **Link**

- A relationship between items, typically between a network



Data type — **Position**

- Spatial data, providing a location in 2D or 3D space

Dataset types

Dataset types

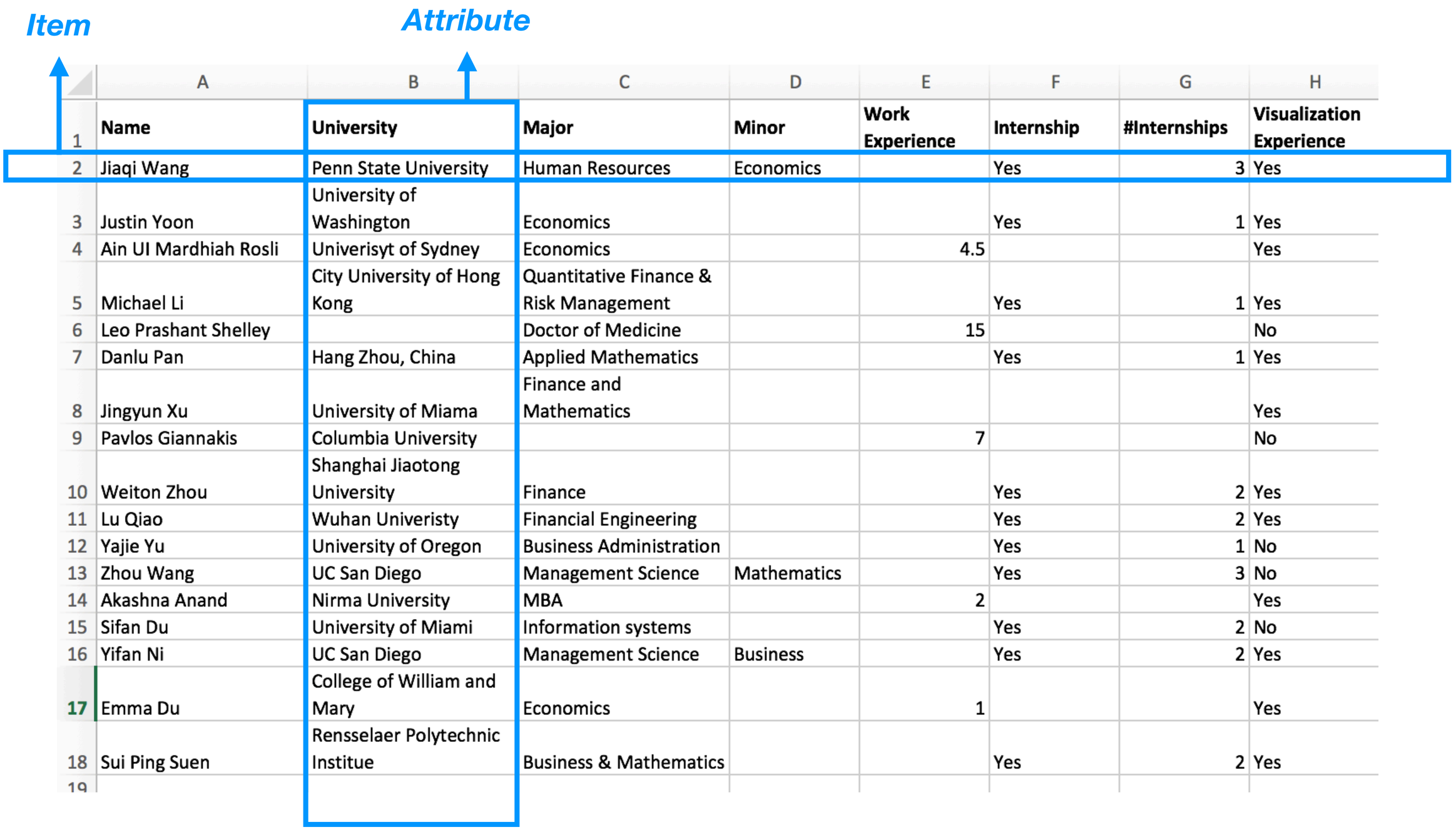
- Table
- Network, tree
- Field
- Geometry

Dataset type — **Table**

- Each data **item** in a new row
- Each column contains an **attribute**

Dataset type — ***Table***

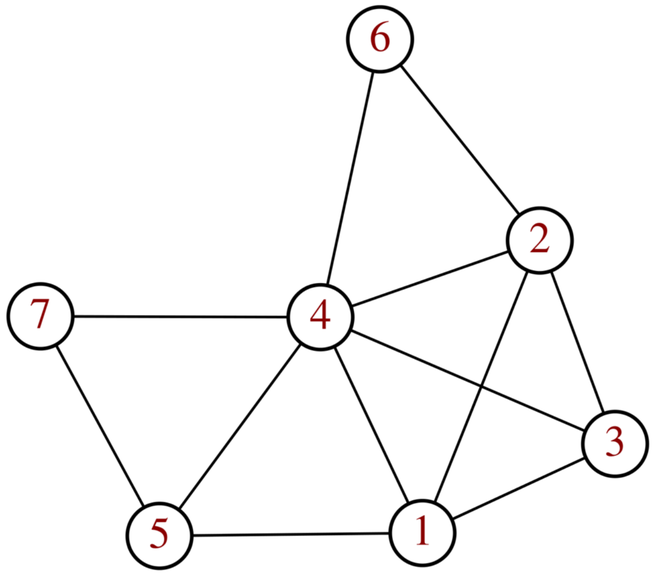
	A	B	C	D	E	F	G	H
1	**Name**	**University**	**Major**	**Minor**	**Work Experience**	**Internship**	**#Internships**	**Visualization Experience**
2	Jiaqi Wang	Penn State University	Human Resources	Economics		Yes	3	Yes
3	Justin Yoon	University of Washington	Economics			Yes	1	Yes
4	Ain UI Mardhiah Rosli	Univerisyt of Sydney	Economics		4.5			Yes
5	Michael Li	City University of Hong Kong	Quantitative Finance & Risk Management			Yes	1	Yes
6	Leo Prashant Shelley		Doctor of Medicine		15			No
7	Danlu Pan	Hang Zhou, China	Applied Mathematics			Yes	1	Yes
8	Jingyun Xu	University of Miama	Finance and Mathematics					Yes
9	Pavlos Giannakis	Columbia University			7			No
10	Weiton Zhou	Shanghai Jiaotong University	Finance			Yes	2	Yes
11	Lu Qiao	Wuhan Univeristy	Financial Engineering			Yes	2	Yes
12	Yajie Yu	University of Oregon	Business Administration			Yes	1	No
13	Zhou Wang	UC San Diego	Management Science	Mathematics		Yes	3	No
14	Akashna Anand	Nirma University	MBA		2			Yes
15	Sifan Du	University of Miami	Information systems			Yes	2	No
16	Yifan Ni	UC San Diego	Management Science	Business		Yes	2	Yes
17	Emma Du	College of William and Mary	Economics		1			Yes
18	Sui Ping Suen	Rensselaer Polytechnic Institue	Business & Mathematics			Yes	2	Yes
19								



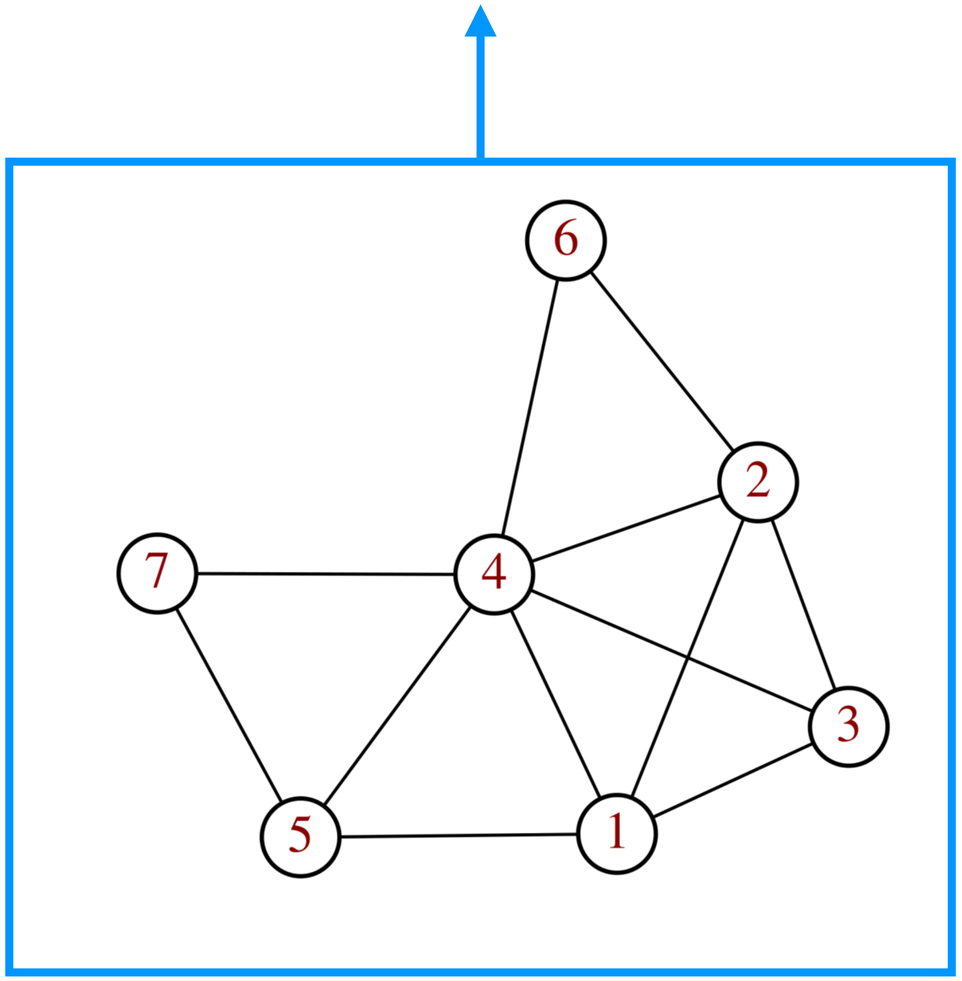
	A	B	C	D	E	F	G	H
1	**Name**	**University**	**Major**	**Minor**	**Work Experience**	**Internship**	**#Internships**	**Visualization Experience**
2	Jiaqi Wang	Penn State University	Human Resources	Economics		Yes	3	Yes
3	Justin Yoon	University of Washington	Economics			Yes	1	Yes
4	Ain UI Mardhiah Rosli	Univerisyt of Sydney	Economics		4.5			Yes
5	Michael Li	City University of Hong Kong	Quantitative Finance & Risk Management			Yes	1	Yes
6	Leo Prashant Shelley		Doctor of Medicine		15			No
7	Danlu Pan	Hang Zhou, China	Applied Mathematics			Yes	1	Yes
8	Jingyun Xu	University of Miama	Finance and Mathematics					Yes
9	Pavlos Giannakis	Columbia University			7			No
10	Weiton Zhou	Shanghai Jiaotong University	Finance			Yes	2	Yes
11	Lu Qiao	Wuhan Univeristy	Financial Engineering			Yes	2	Yes
12	Yajie Yu	University of Oregon	Business Administration			Yes	1	No
13	Zhou Wang	UC San Diego	Management Science	Mathematics		Yes	3	No
14	Akashna Anand	Nirma University	MBA		2			Yes
15	Sifan Du	University of Miami	Information systems			Yes	2	No
16	Yifan Ni	UC San Diego	Management Science	Business		Yes	2	Yes
17	Emma Du	College of William and Mary	Economics		1			Yes
18	Sui Ping Suen	Rensselaer Polytechnic Institue	Business & Mathematics			Yes	2	Yes
19								

Dataset type — **Network**

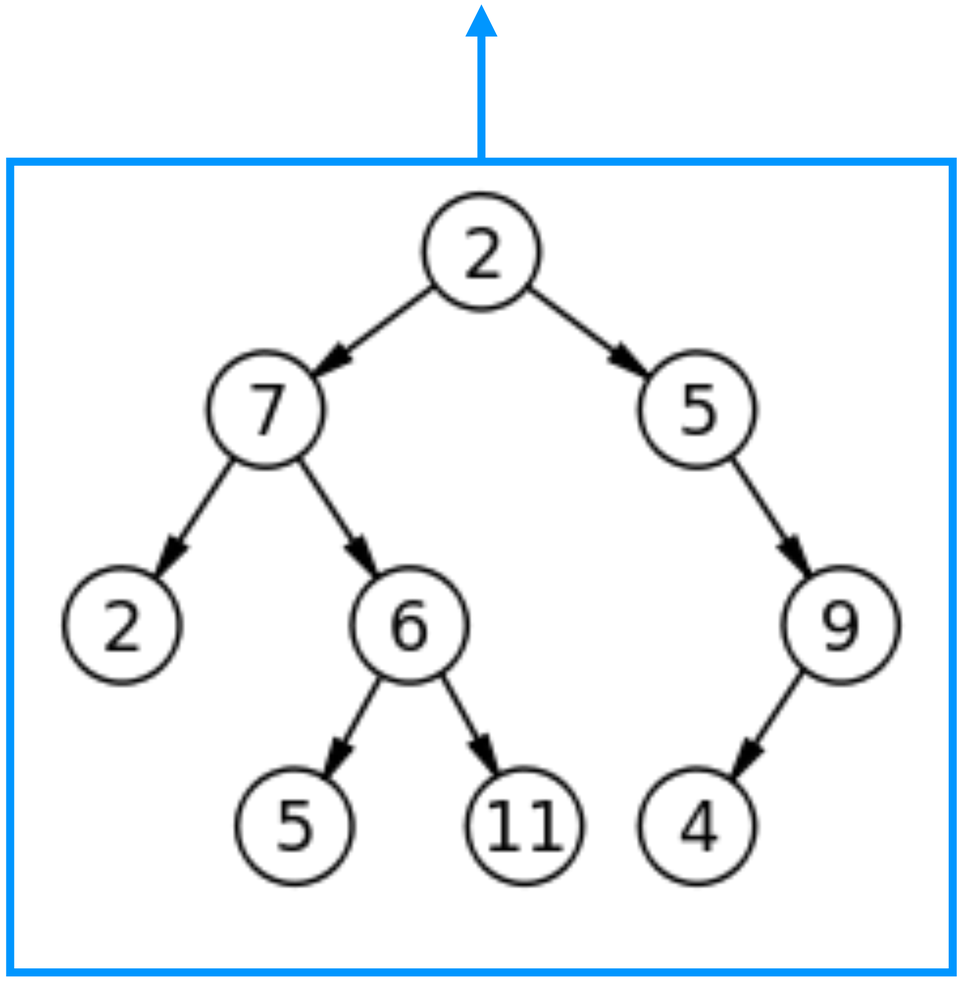
- Each item in a network is a **node**
- The relation between two items is a **link**
- Specifies that there is some kind of relationship between two or more items

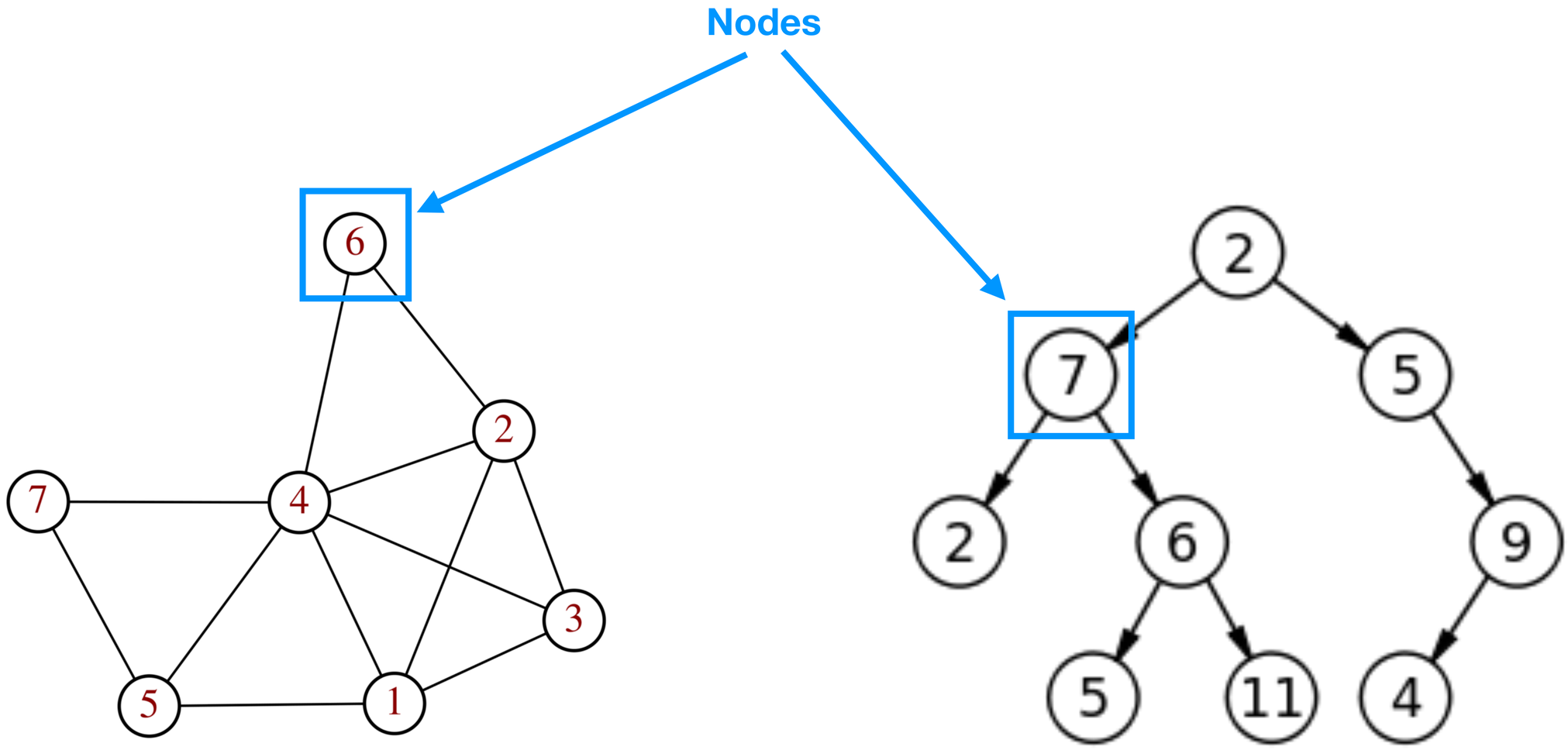


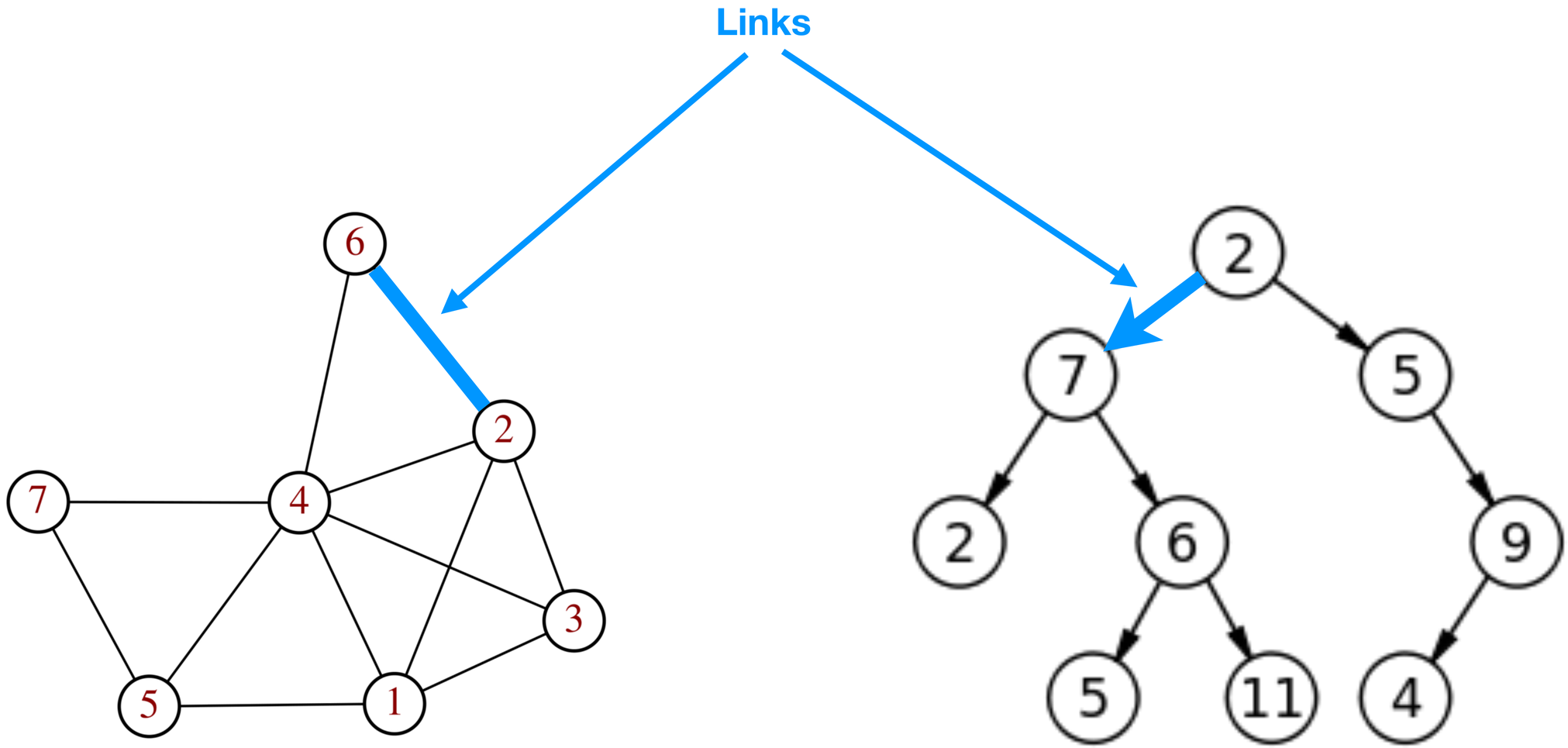
*Dataset type — **Network***



*Dataset type — **Tree***



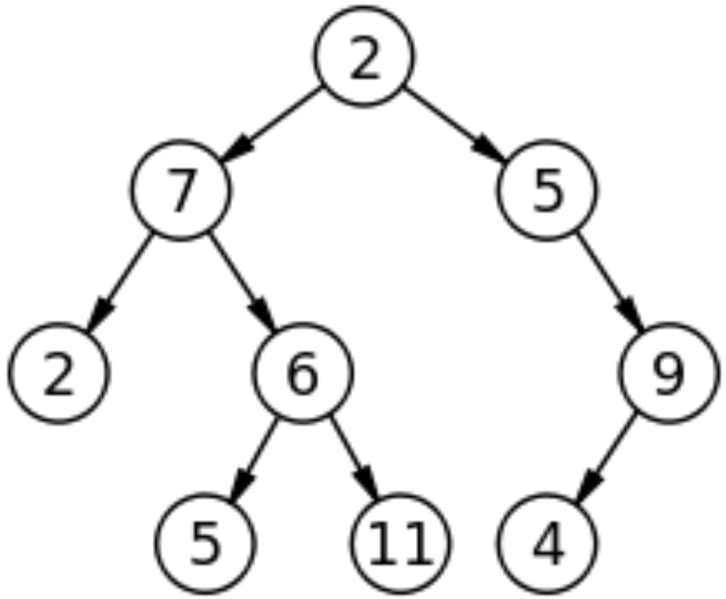

Nodes
6
2
7
4
3
5
1
2
7
5
2
6
9
5
11
4

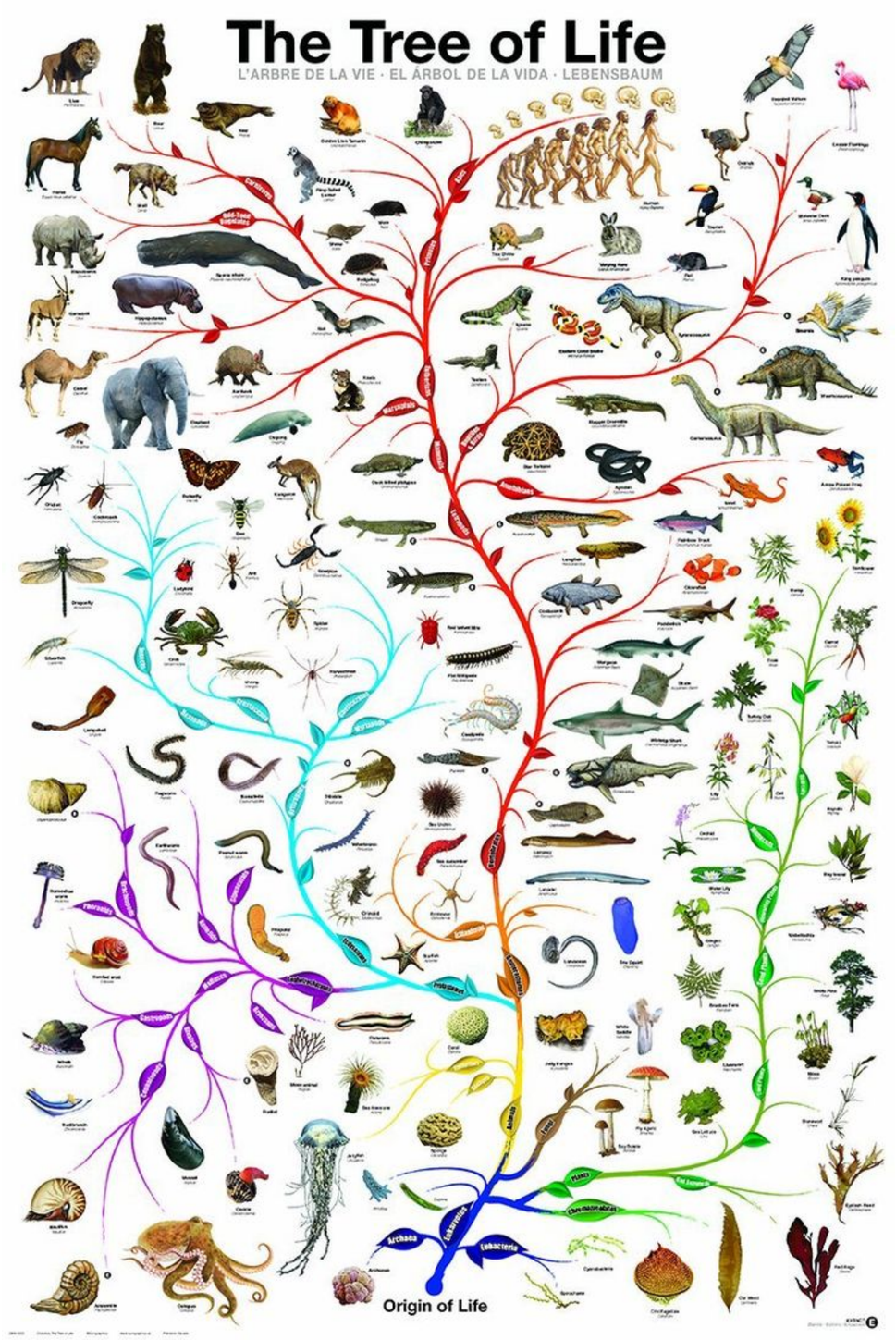

Links
6
2
7
4
3
5
1
2
7
5
2
6
9
5
11
4

Social network

Dataset type — **Network (tree)**

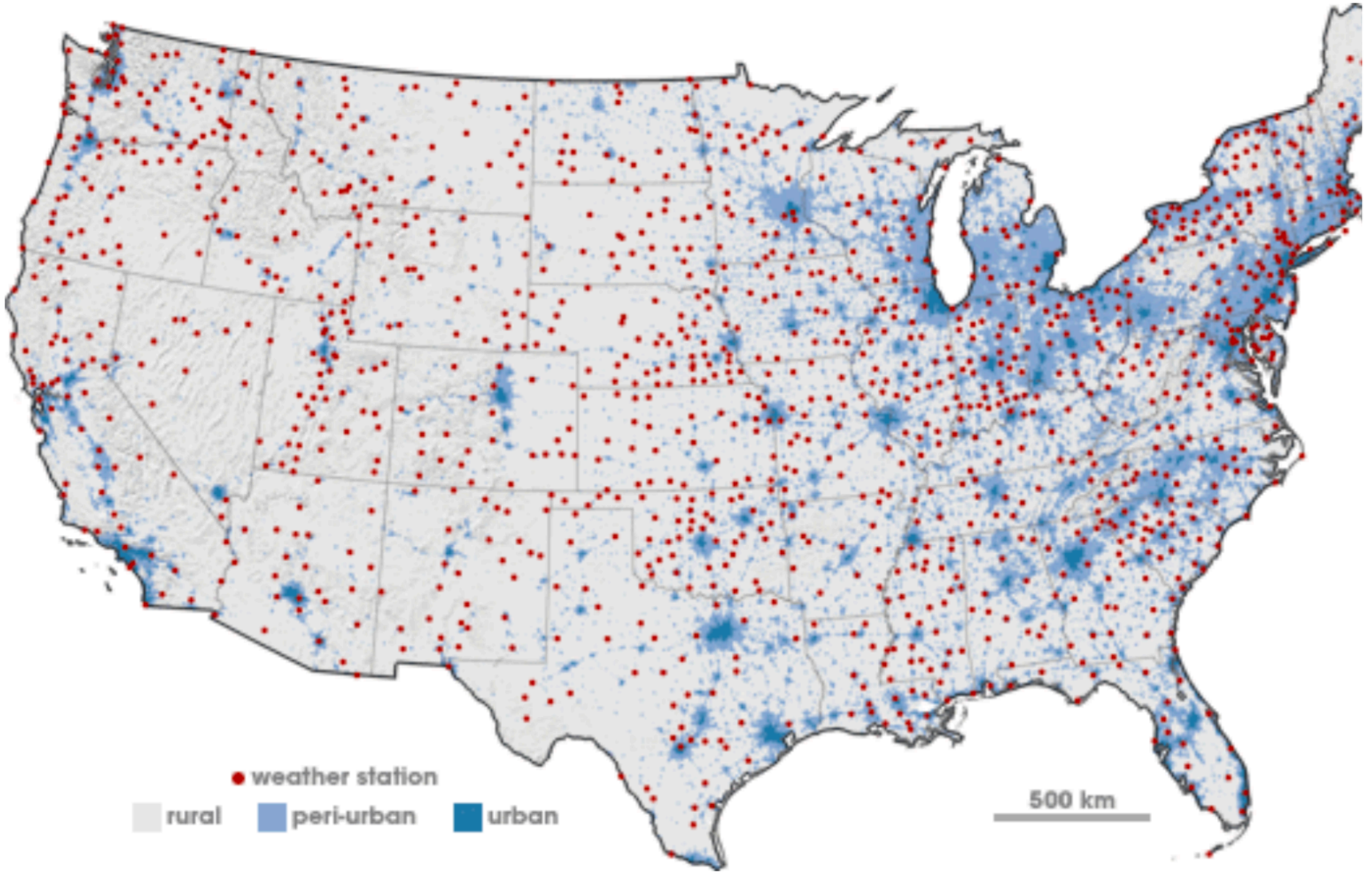
- Networks with a hierarchical structure are **trees**
- Each child node has only one parent node pointing to it




The Tree of Life
L'ARBRE DE LA VIE · EL ÁRBOL DE LA VIDA · LEBENSBAUM
Origin of Life

Type — **Field**

- The **field** dataset type also contains attribute values associated with cells
- Each **cell** in a field contains measurements or calculations from a continuous domain (conceptually infinite values).
- Considerations
    - **Sampling** — How frequently to take measurements
    - **Interpolation** — How to show values between points in a way that does not mislead
- Ex. (in the physical world) temperature, pressure, wind velocity


weather station
rural
peri-urban
urban
500 km

Type — **Geometry**

- Specifies information about the shape of items with specific spatial positions

Marks & Channels

Mark — Represents **items** or **links**

Channel — Changes appearance based on an attribute (visual variable)

Marks

Points

0D

Lines

1D

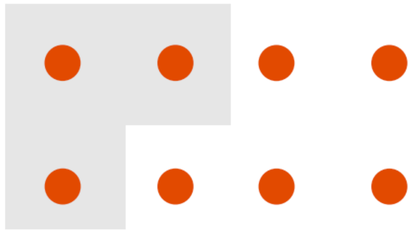
Areas

2D

3D — Volume (but rarely used)

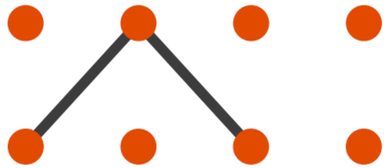
Marks (for links)

Containment

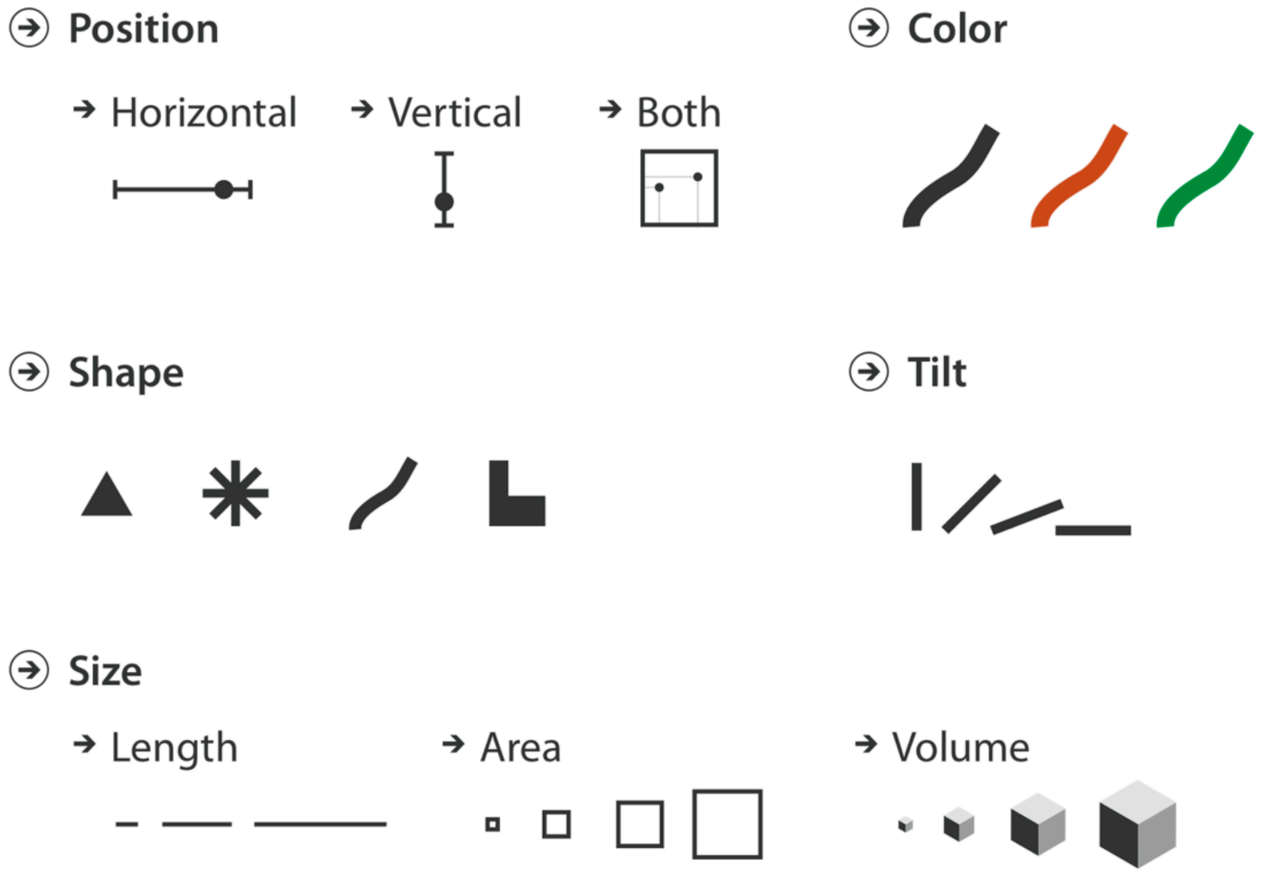


(can be nested)

Connection

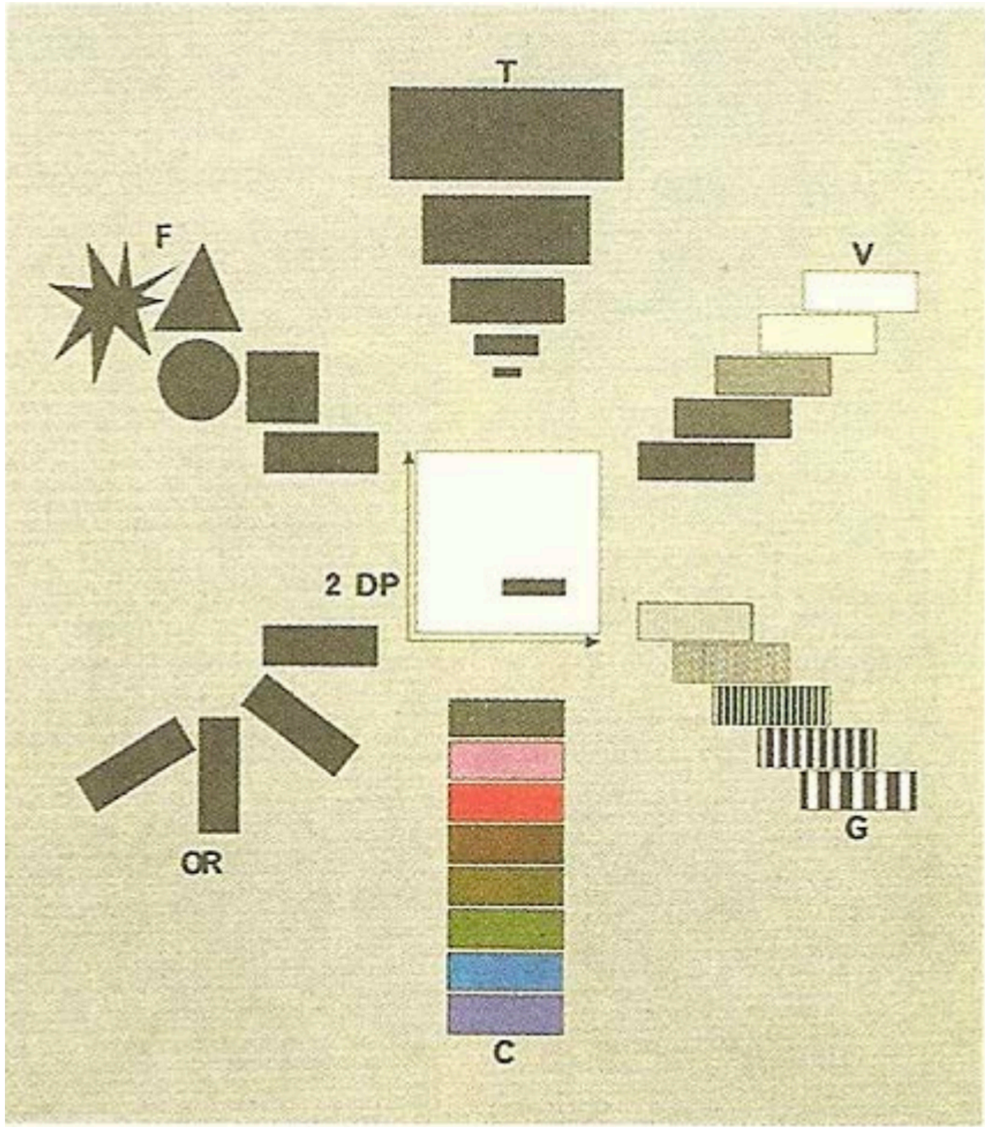


Channels (a.k.a. visual variables) — via Jock Mackinlay



Size

Shape



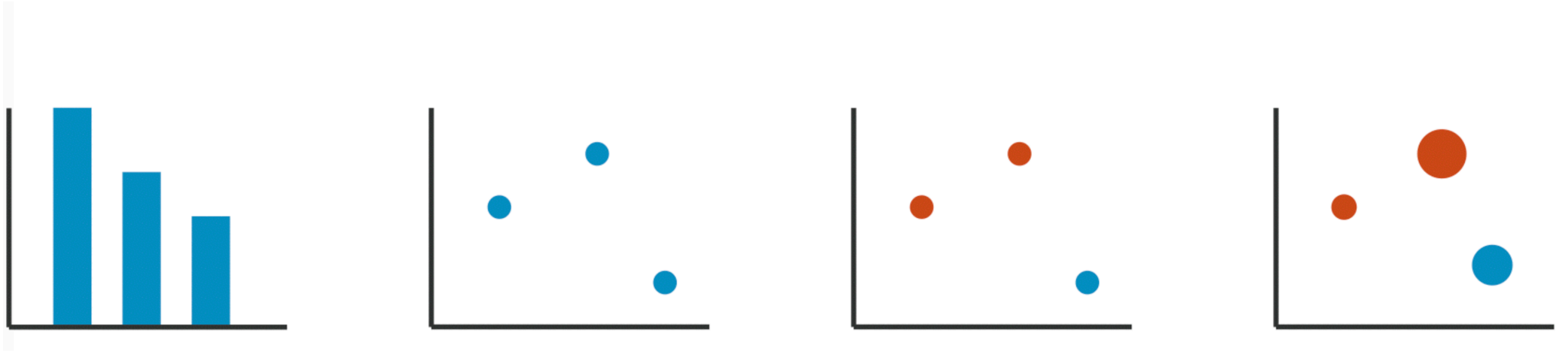
Value

Orientation

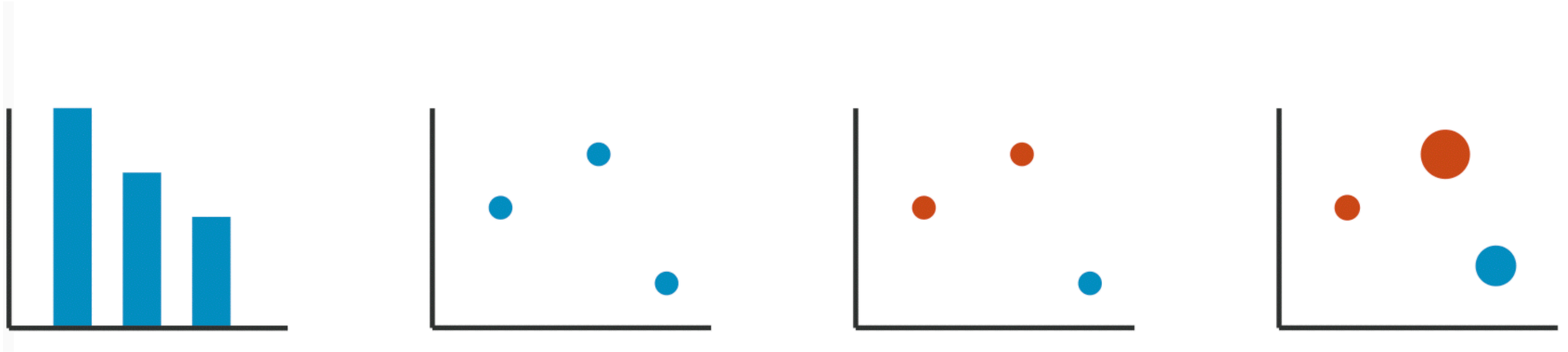
Texture

Color

Using marks and channels



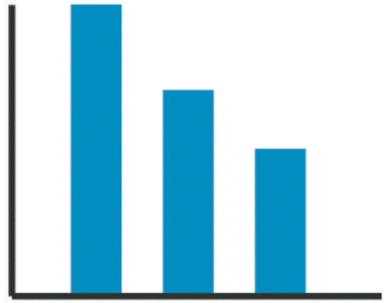
Using marks and channels



Mark: Line

Channel: Length,
position (vertical),
position (horizontal)

Using marks and channels

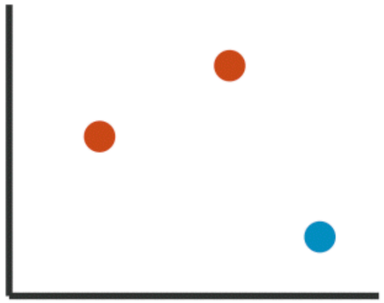


Mark: Line

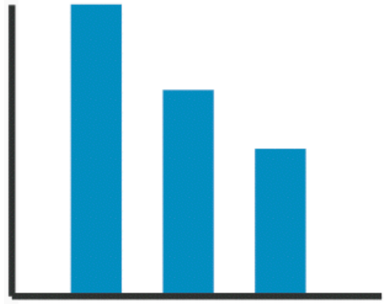
Channel: Length, position (vertical), position (horizontal)

Mark: Point

Channel: Position (vertical), position (horizontal)



Using marks and channels



Mark: Line

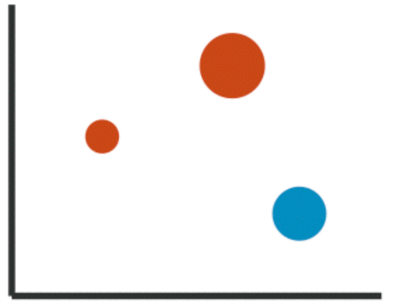
Channel: Length, position (vertical), position (horizontal)

Mark: Point

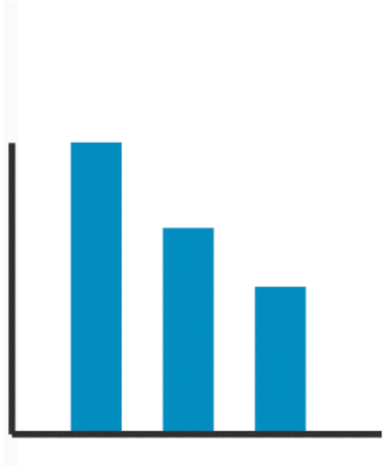
Channel: Position (vertical), position (horizontal)

Mark: Point

Channel: Position (vertical), position (horizontal), color



Using marks and channels



Mark: Line

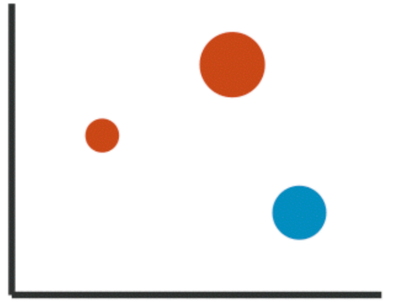
Channel: Length, position (vertical), position (horizontal)

Mark: Point

Channel: Position (vertical), position (horizontal)

Mark: Point

Channel: Position (vertical), position (horizontal), color

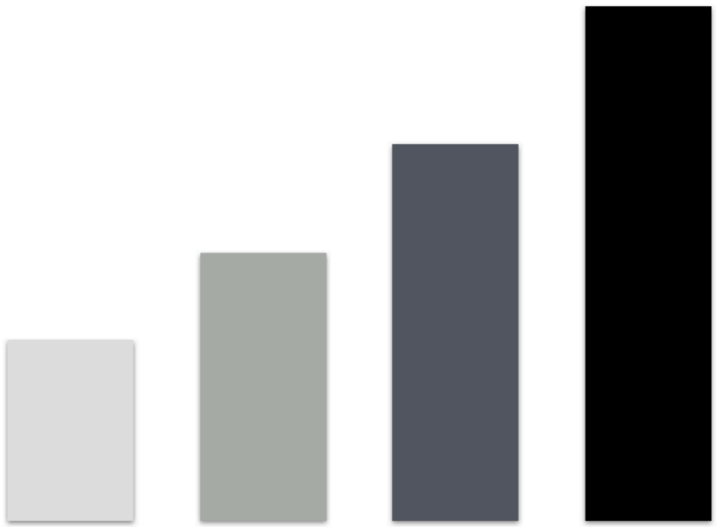


Mark: Point

Channel: Position (vertical), position (horizontal), color, size

Analyze graphics as a combination of **marks and channels** showing **abstract data dimensions**.

Redundant encoding



Length, position, and value

Magnitude Channels

— How much?

- Position
- Length
- Saturation…

Quantitative data

Identity Channels

— What? Where?

- Shape
- Color (hue)
- Spatial region…

Qualitative data

Break — **Homework review**

Mackinlay ranks channels in terms of **effectiveness**

- His **effectiveness principle** states that the most important attributes in a visualization should be encoded with the highest-ranked channels
- Where do the rankings come from? Mackinlay considers the following:
  - Selectivity
  - Quantifiability
  - Discriminability
  - Popout

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Most

Effectiveness

Least

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

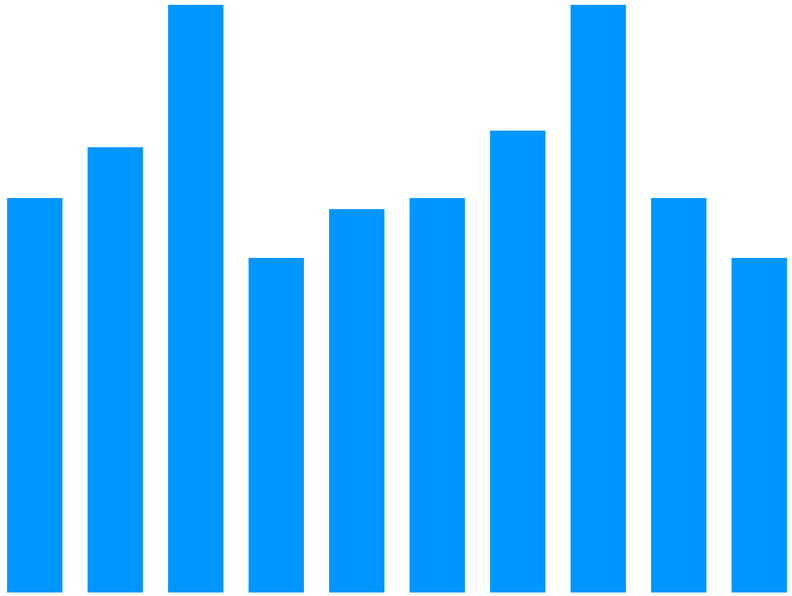
Shape

Selectivity

- Is a mark distinct from other marks?
- Can we make out the difference between two marks?

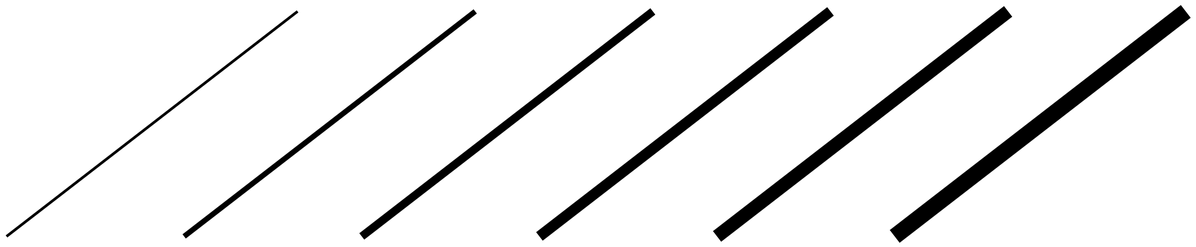
Quantifiability

- Can we reliably compute the difference between marks?



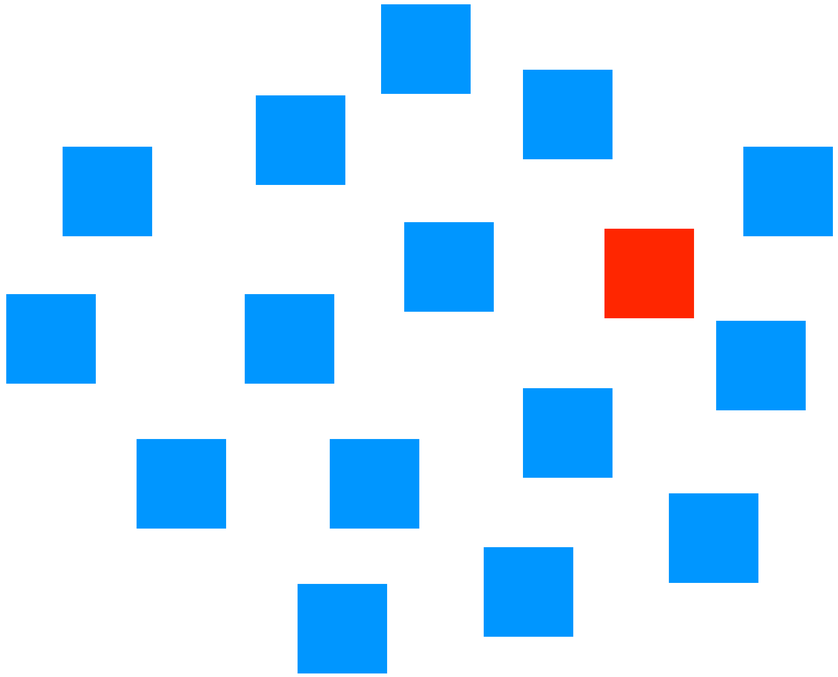
Discriminability

- How many usable steps? (Few for line width)



Popout

- Can a different item be noticed immediately?

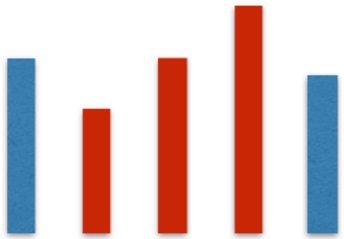


Position is the strongest channel (visual variable), and is suitable for **all data types**.

Other factors affecting accuracy

- Alignment

- Distractors

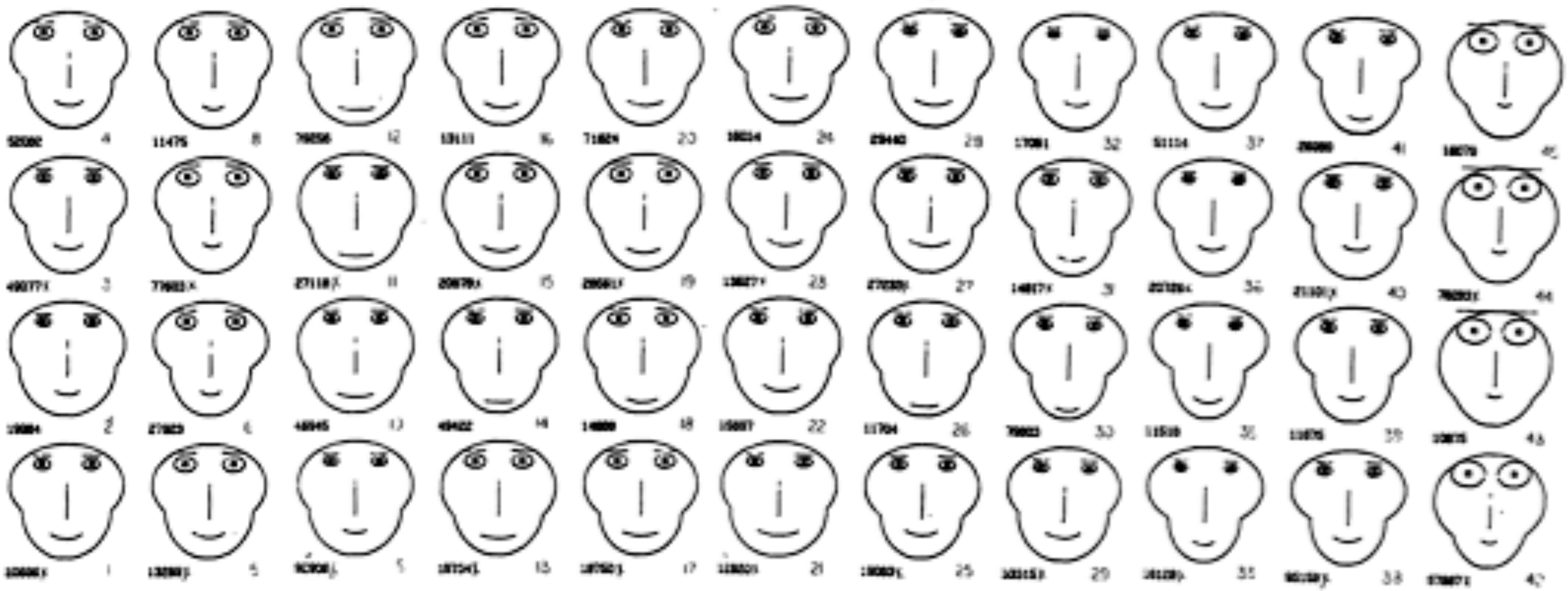


- Distance

- Common scale

Bonus — **Chernoff Faces**

- Idea: use facial parameters to map quantitative data



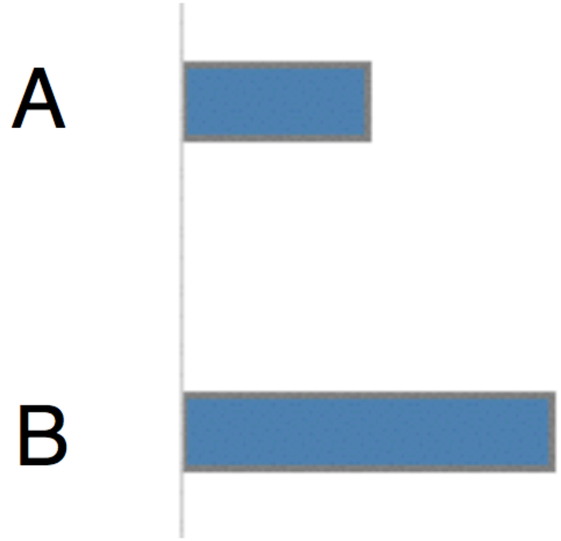
Bonus — **Chernoff Faces**

- Idea: use facial parameters to map quantitative data

Perception in practice

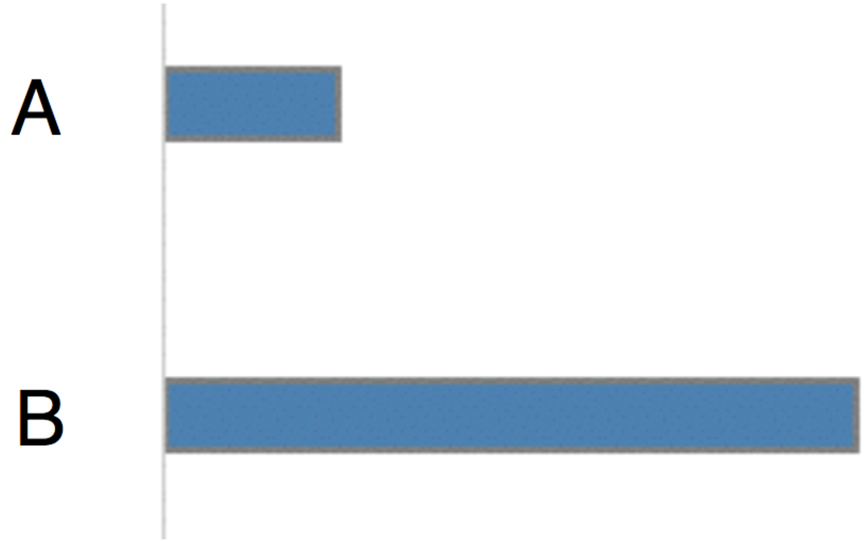
Test — **How much longer?**

Test — **How much longer?**

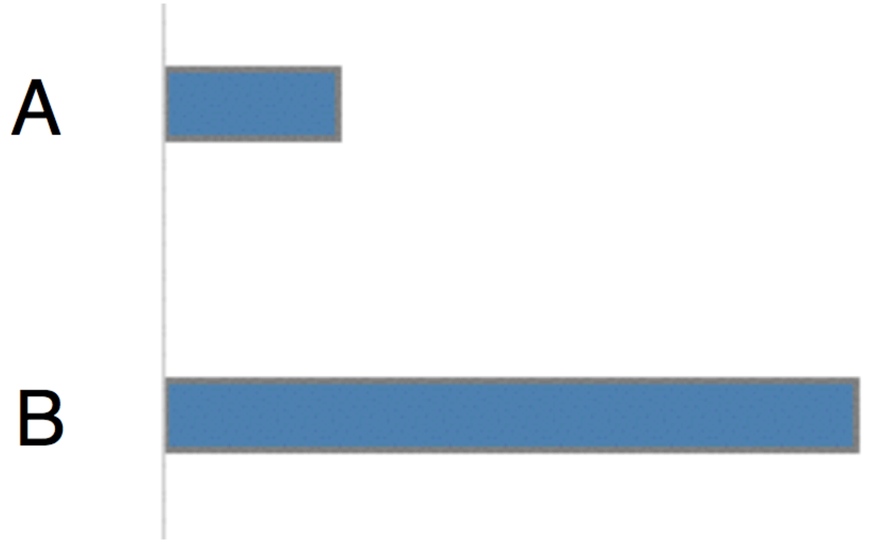


—> **2X**

Test — **How much longer?**

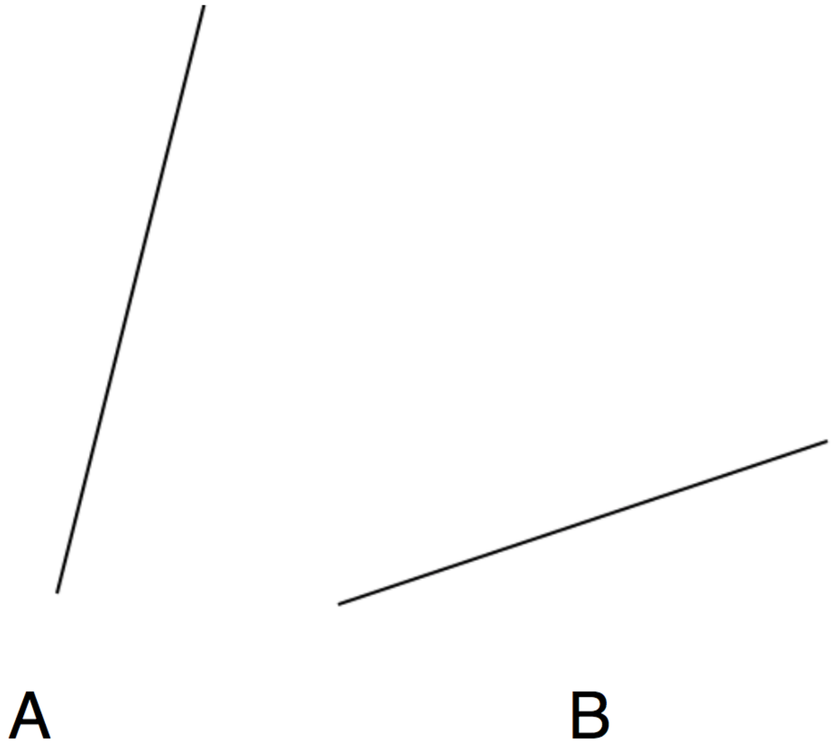


Test — **How much longer?**

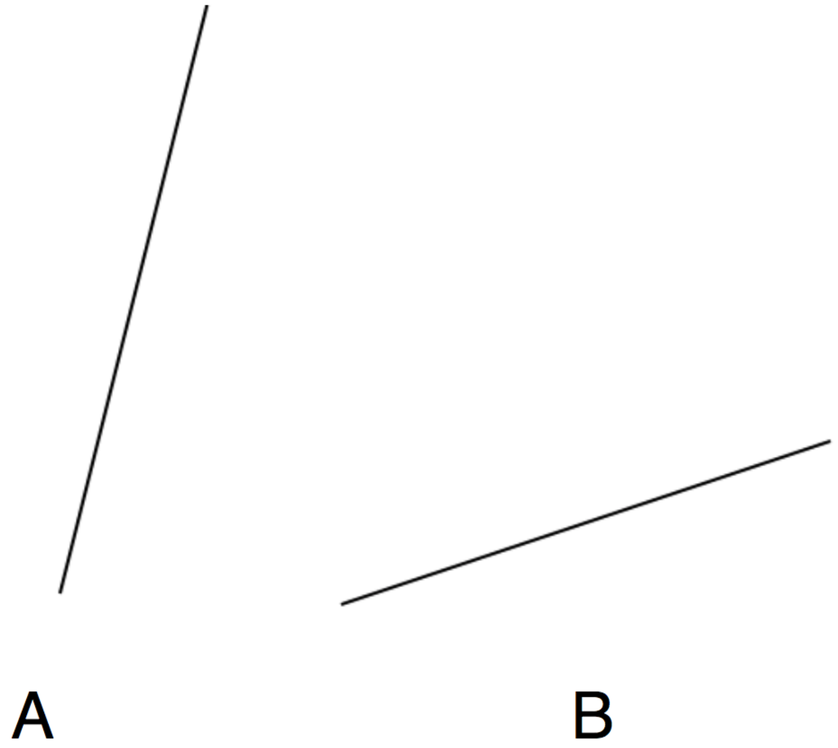


—> **4X**

Test — **How much steeper?**



Test — **How much steeper?**



—> **~4X**

Test — **How much larger (area)?**

A B

Test — **How much larger (area)?**

A B

—> **5X**

Test — **How much larger (area)?**

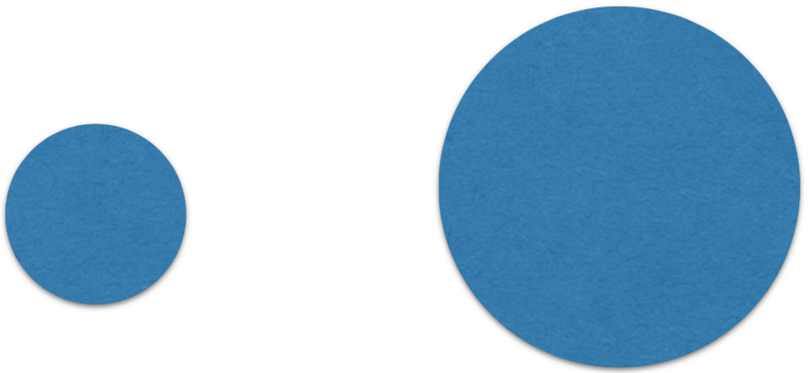
A B

Test — **How much larger (area)?**

A B

—> **3X**

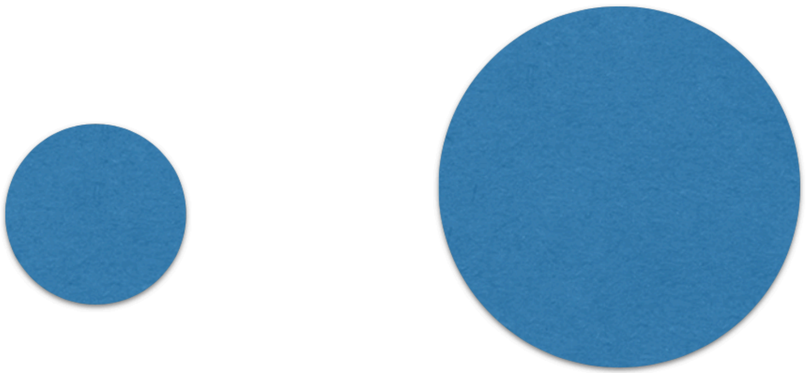
Test — **How much larger (diameter)?**



A

B

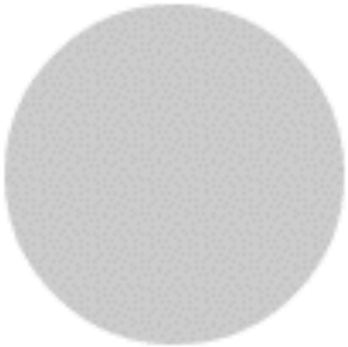
Test — **How much larger (diameter)?**



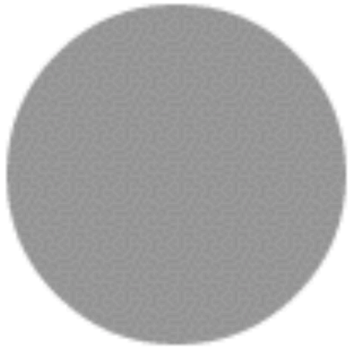
A B

—> **2X**

Test — **How much darker?**

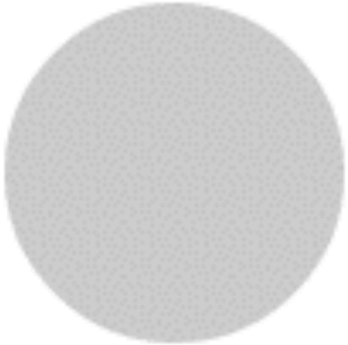


A

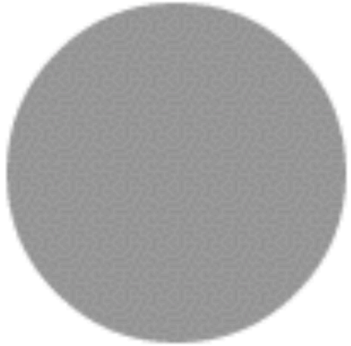


B

Test — **How much darker?**



A



B

—> **2X**

Questions?