

COS30045

---

# DATA VISUALISATION

TOPIC 03: DATA ABSTRACTION

Commonwealth of Australia  
*Copyright Act 1968*

**Notice for paragraph 135ZXA (a) of the *Copyright Act 1968***

**Warning**

This material has been reproduced and communicated to you by or on behalf of Swinburne University of Technology under Part VB of the *Copyright Act 1968* (the *Act*).

The material in this communication may be subject to copyright under the *Act*. Any further reproduction or communication of this material by you may be the subject of copyright protection under the *Act*.

Do not remove this notice.

IMAGES USED IN THIS PRESENTATION: FROM MICROSOFT CLIPART (UNLESS OTHERWISE SPECIFIED)

Acknowledgment: This unit is partially based on the Visualisation unit developed by Alex Lex and Hanspeter Hfister (with permission).

Source of images used in this lecture available as links in the PDF version.



# DATA ABSTRACTION

---

What:

- ▶ understand the different types data

Why:

- ▶ the type of data available will drive the design

How:

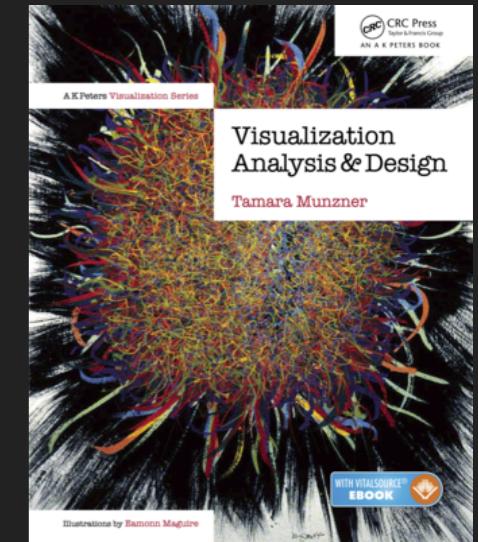
- ▶ now: understand the properties of different data types
- ▶ later: look at how to best represent different data types in visualisation



# DATA ABSTRACTION

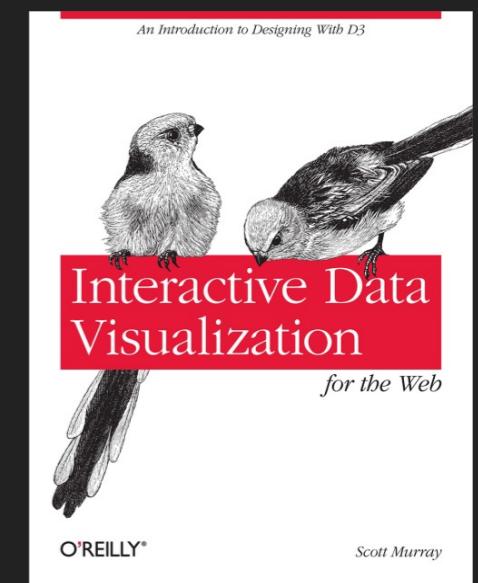
## Lecture Readings

- ▶ VAD Ch 2



## Tutorials

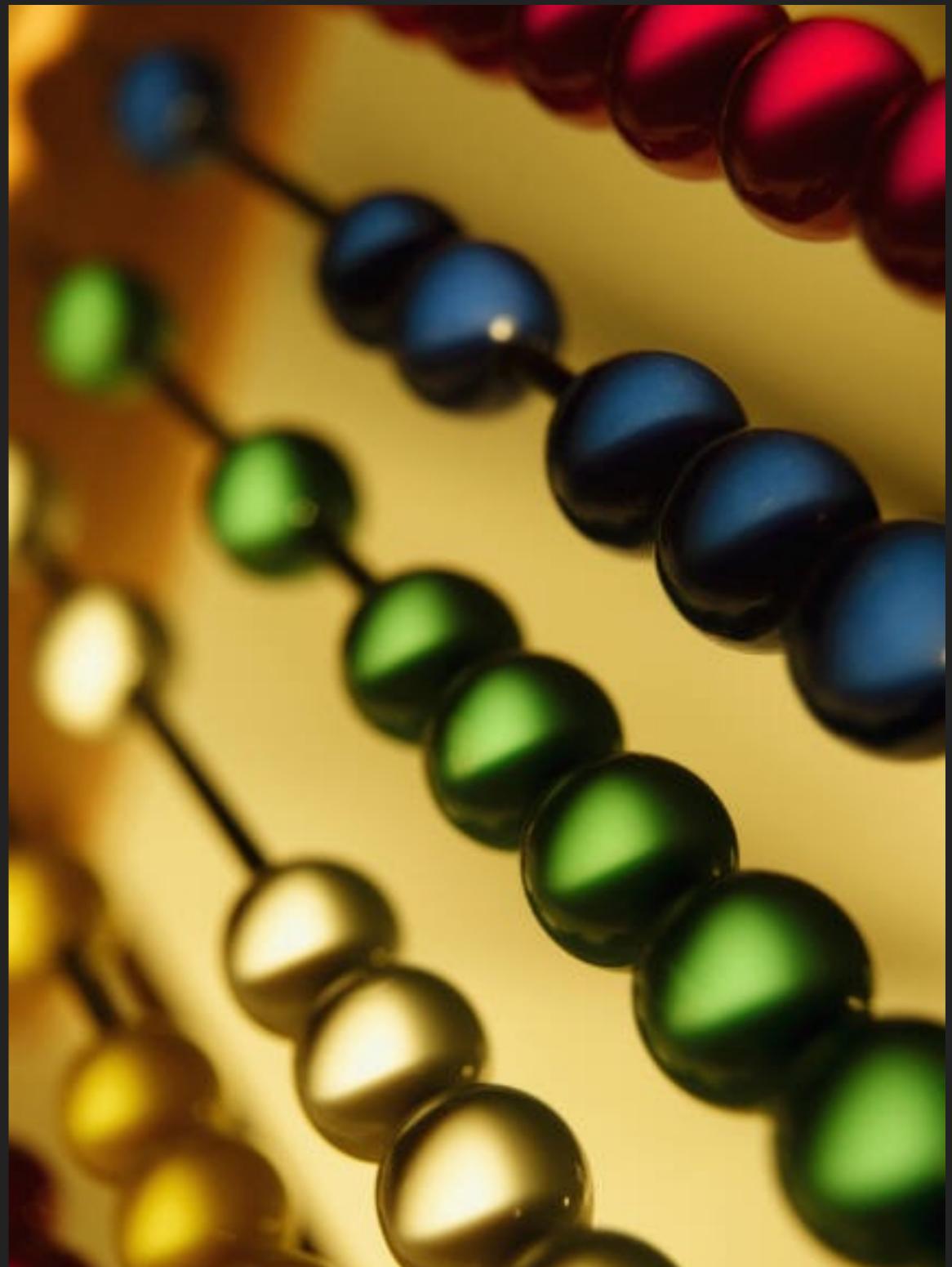
- ▶ IDV Ch



---

# DATA SEMANTICS AND TYPES

- ▶ What kind of data can be visualised?
  - ▶ dataset types (collection)
    - ▶ tables, networks, fields, geometry, clusters
  - ▶ data types (instance)
    - ▶ items, attributes, links, positions, grids
  - ▶ dataset availability
    - ▶ static, dynamic



---

# DATA SEMANTICS AND TYPE

Basil, 7, S, Pear

- ▶ Semantics: Real-world meaning
  - ▶ what does Basil represent?
- ▶ Data type:
  - ▶ what kind of data is '7'?
  - ▶ a number: number minutes to complete a task, height of a box, items in an order?
  - ▶ a name/code: name of shipping container, name of shop, dish number on restaurant menu, reference number for a transaction?

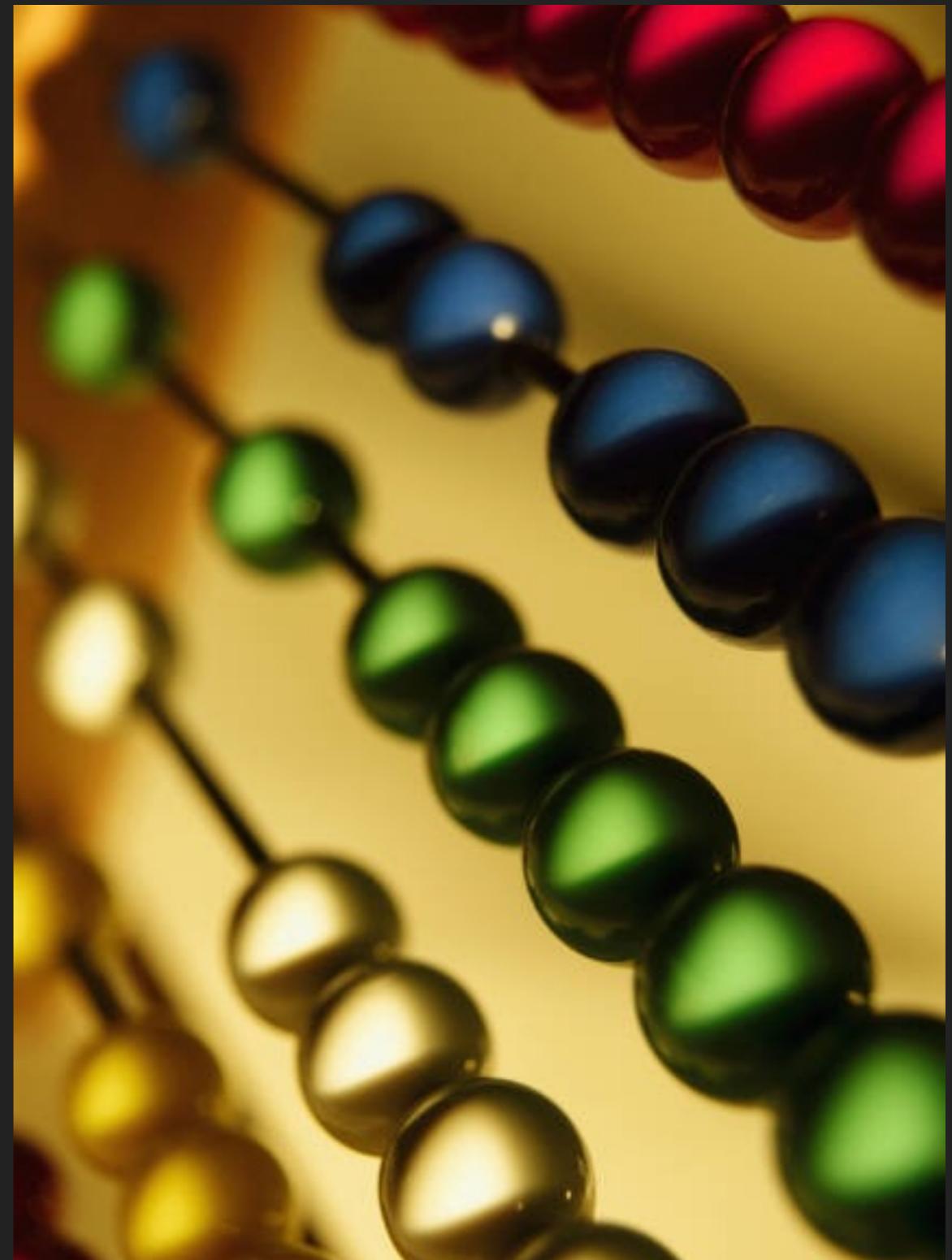
## DATA SEMANTICS: METADATA

1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

---

# DATA TYPES

- ▶ items
- ▶ attributes
- ▶ links
- ▶ positions
- ▶ grids



## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

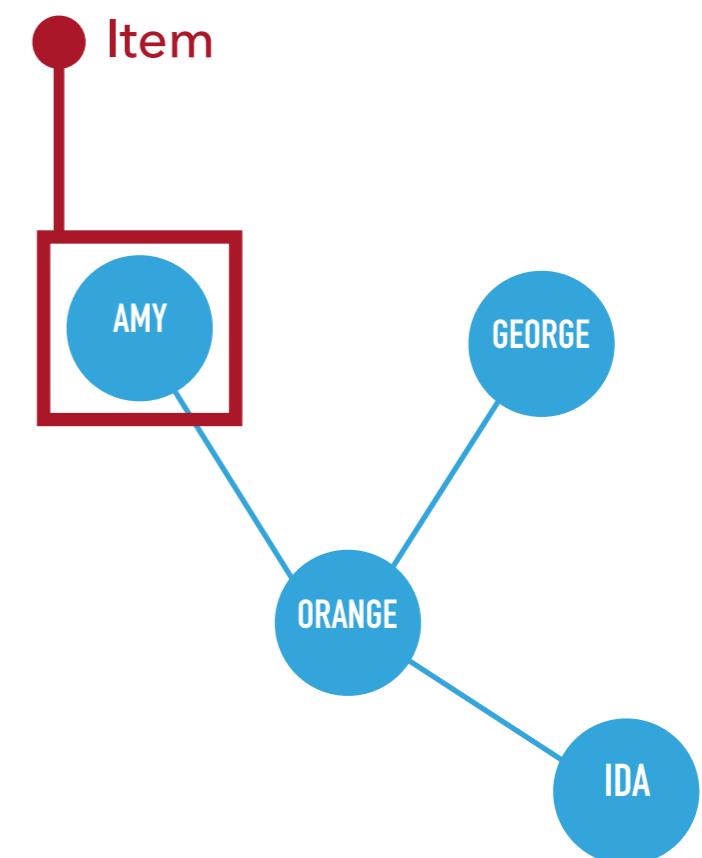
## GRIDS

- ▶ a discrete entity



ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Table



Network

## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

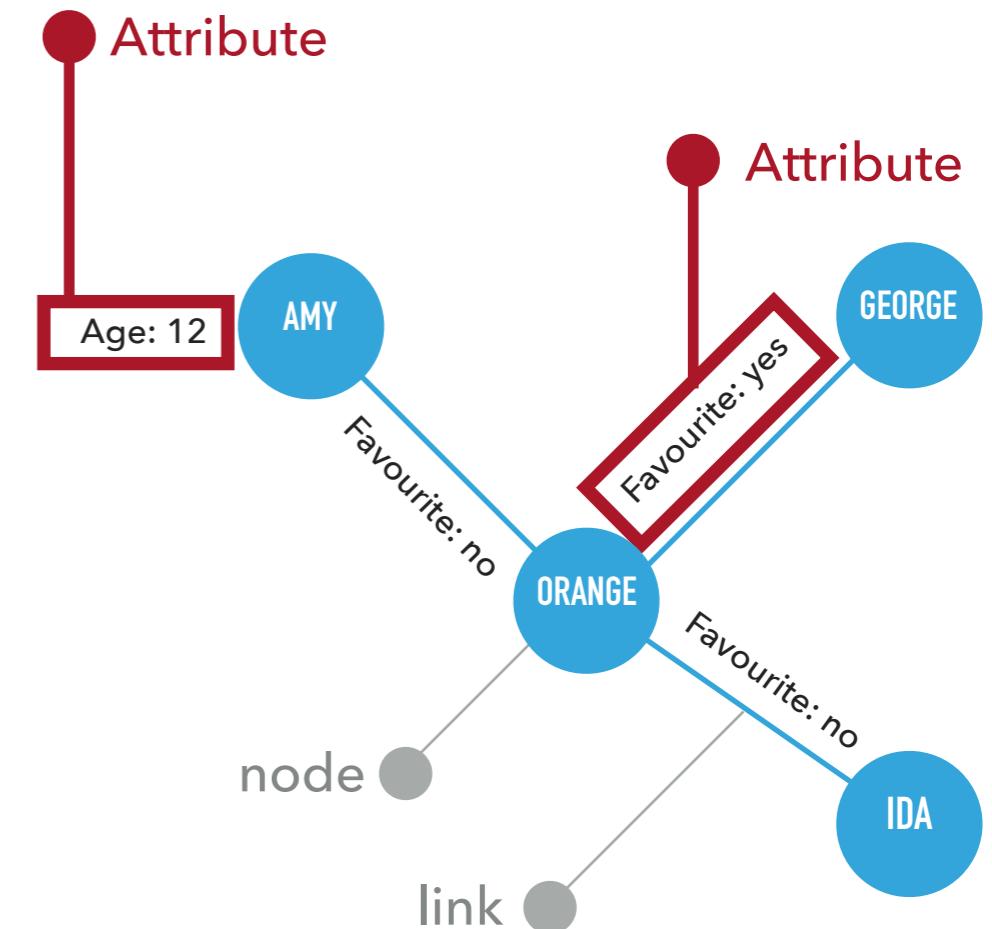
## GRIDS

- ▶ a property that can be measured observed or logged

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Table

Attributes are associated columns



Network

Attributes can be associated with nodes and links

## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

- there are different types of attributes (scale of measurement)

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

- measurement type determines options for visualisation

### Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

- there are different types of attributes (scale of measurement)

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Table

### Attribute Types

→ Categorical



→ Ordered



→ Ordinal



→ Quantitative



## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

- there are different types of attributes (scale of measurement)

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

### Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



# SCALE OF MEASUREMENT

- ▶ nominal (categorical)
- ▶ ordinal (ordered)
- ▶ interval (ordered with equal distances between points on scale)
- ▶ ratio (ordered with a fixed zero point)



## ITEMS

## ATTRIBUTES

## LINKS

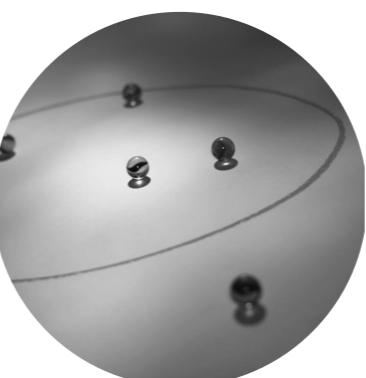
## POSITIONS

## GRIDS

Scales of measurement

- ▶ Nominal
- ▶ categorical
- ▶ mathematical operations
- ▶ equal to
- ▶ examples:
  - ▶ label/name, gender, file type, true/false, shape

Qualitative



## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

Scales of measurement

- ▶ Ordinal
- ▶ ordered data
- ▶ mathematical operations
- ▶ equal to
- ▶ less/greater than
- ▶ examples:
- ▶ places in a race, t-shirt size,

Qualitative



## ITEMS

## ATTRIBUTES

## LINKS

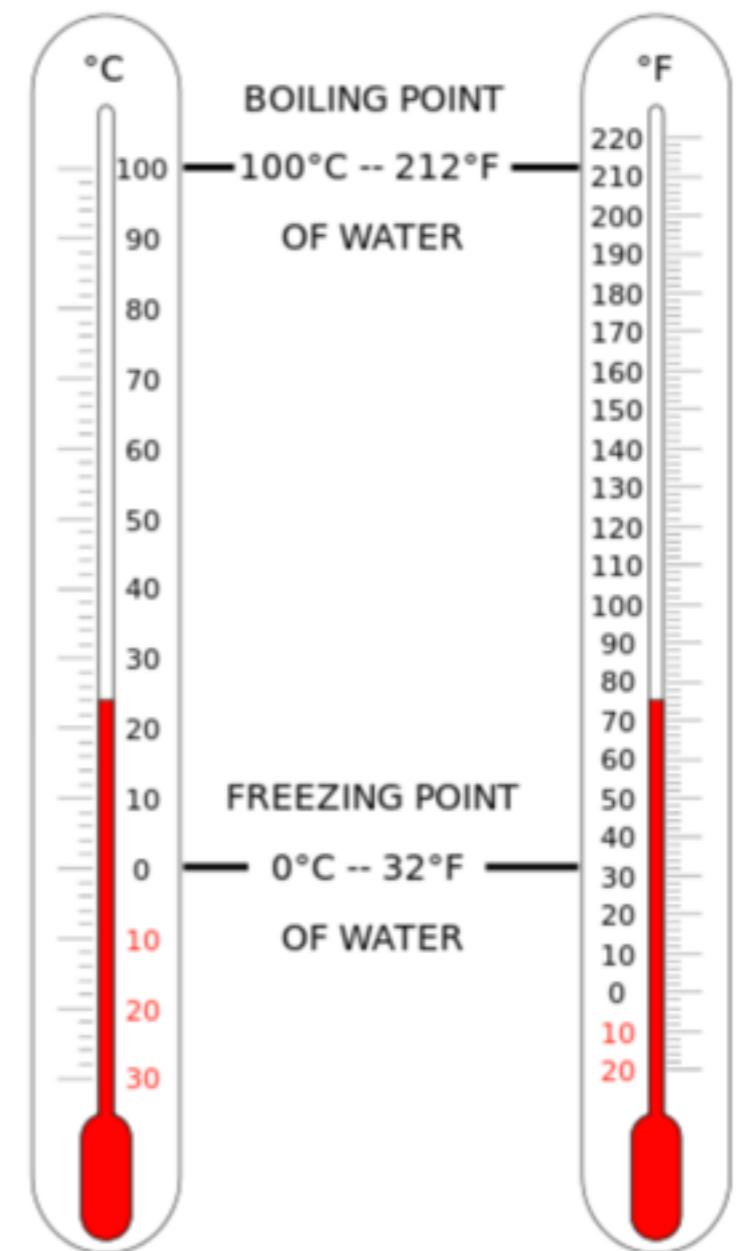
## POSITIONS

## GRIDS

Scales of measurement

- ▶ Interval
  - ▶ ordered data
  - ▶ equal distance between intervals
  - ▶ no true zero point
  - ▶ mathematical operations
    - ▶ equal to, less/greater than, add, subtract

Quantitative



## ITEMS

## ATTRIBUTES

## LINKS

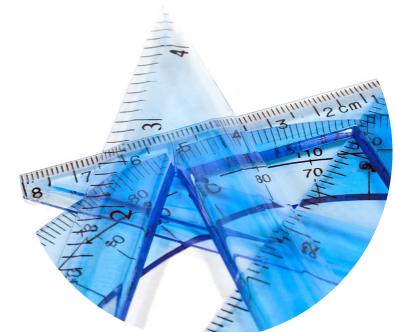
## POSITIONS

## GRIDS

Scales of measurement

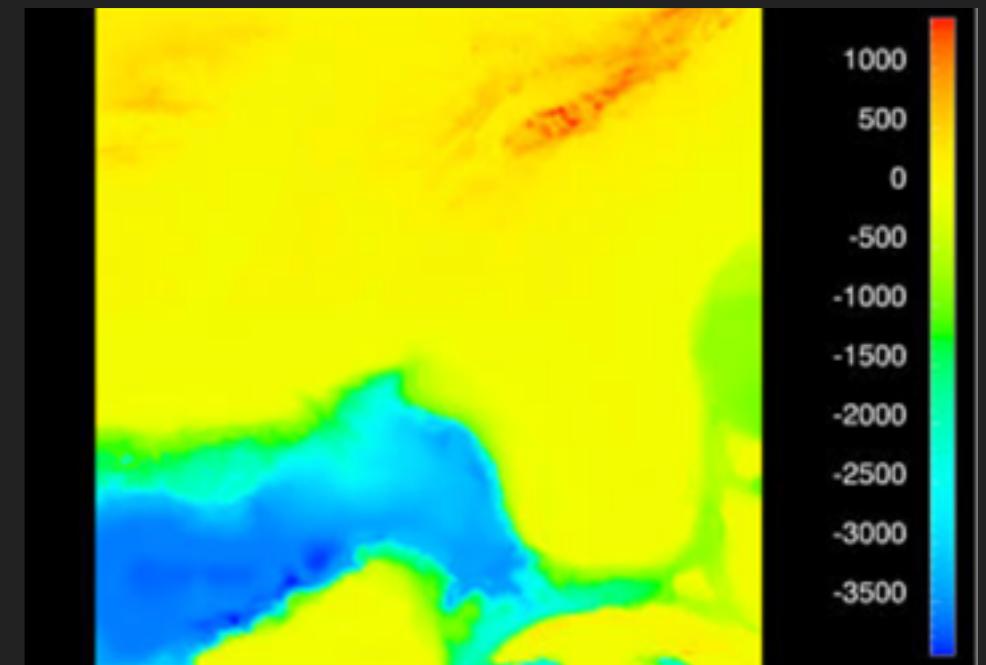
- ▶ Ratio
  - ▶ ordered data
  - ▶ equal distance between intervals
  - ▶ true zero point
- ▶ mathematical operations
  - ▶ equal to, less/greater than, add, subtract, multiply, divide

Quantitative

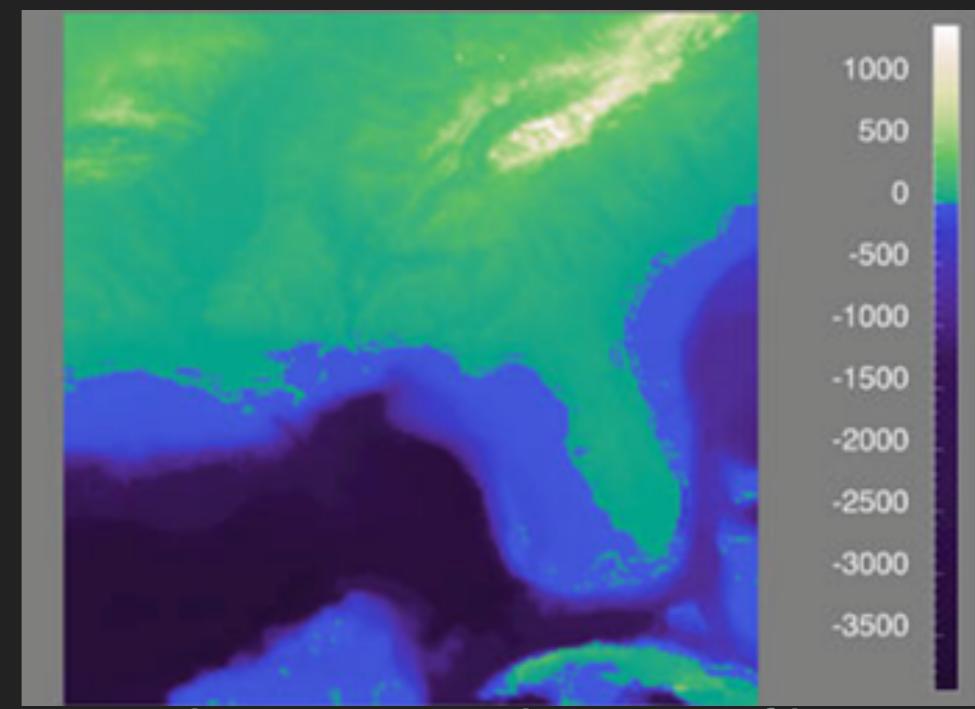


# DATA SET SEQUENCE

- ▶ Sequential data
  - ▶ homogenous range from min to max (e.g., temp, length, weight)
- ▶ Divergent data
  - ▶ two sequences of data pointing in opposite direction (e.g. Sea Level (above and below))
  - ▶ common zero point
  - ▶ important to show zero crossing



Using continuous graduations of hue (rainbow colour map) to encode sea level



Using divergent graduations of hue to encode sea level (green vs blue)

## ITEMS

## ATTRIBUTES

## LINKS

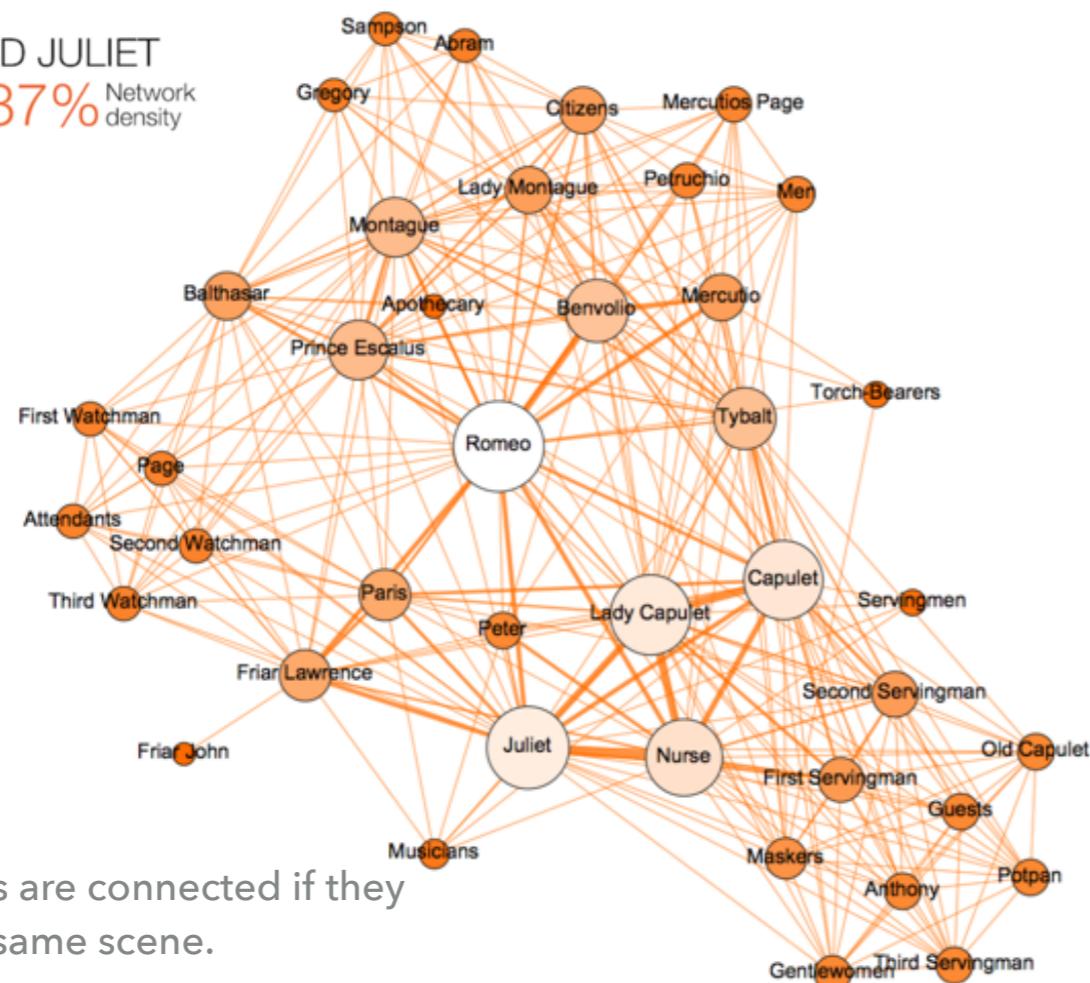
## POSITIONS

## GRIDS

- ▶ a relationship between items within a network

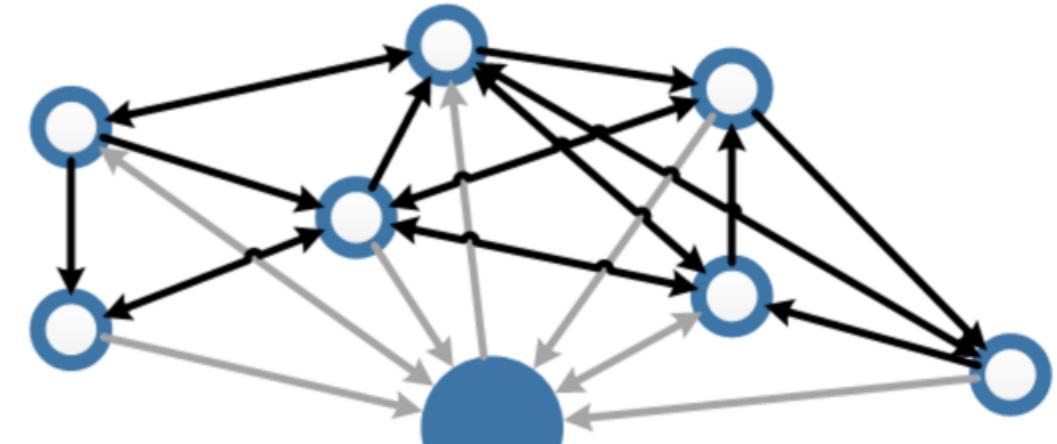
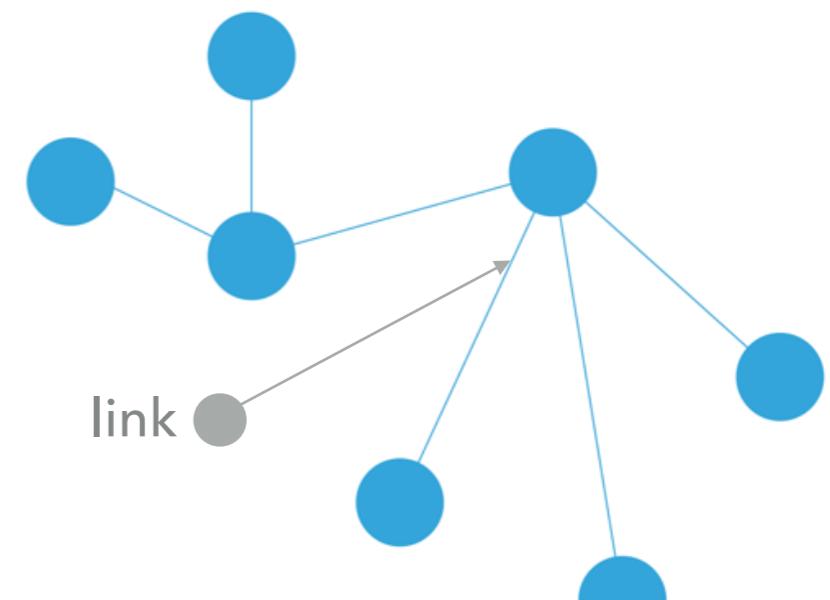
### ROMEO AND JULIET

Number of characters 41 | 37% Network density



Two characters are connected if they appear in the same scene.

Their size and color intensity are proportional to their weighted degree. M Grandjean



@fsharporg

Links can go one-way or both ways  
(follower links on fsharporg Twitter account) E Gabasova

## ITEMS

## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

- ▶ spatial data providing a location in 2D or 3D space

Map

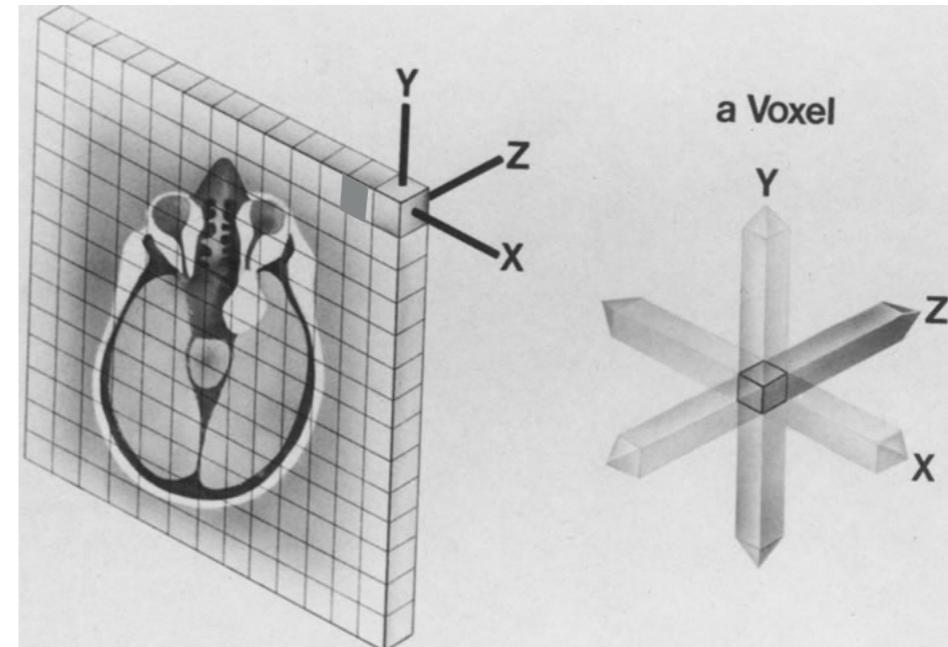


Latitude =  
 $23^{\circ} 41' 60''$  S

Longitude =  
 $133^{\circ} 52' 0.12''$  E

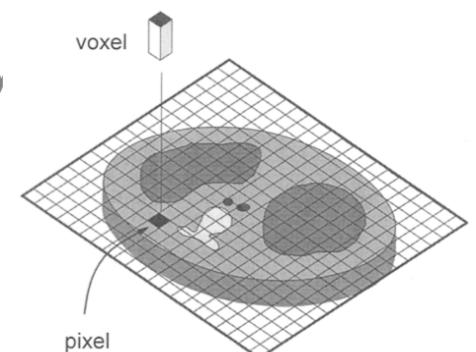
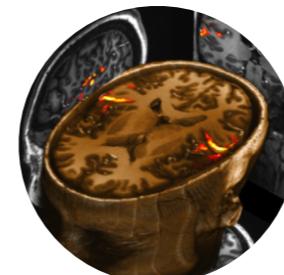
Position specified is relative

MRI scan



Voxel: 3D location - x, y and z

Pixel: 2D location - x and y



## ITEMS

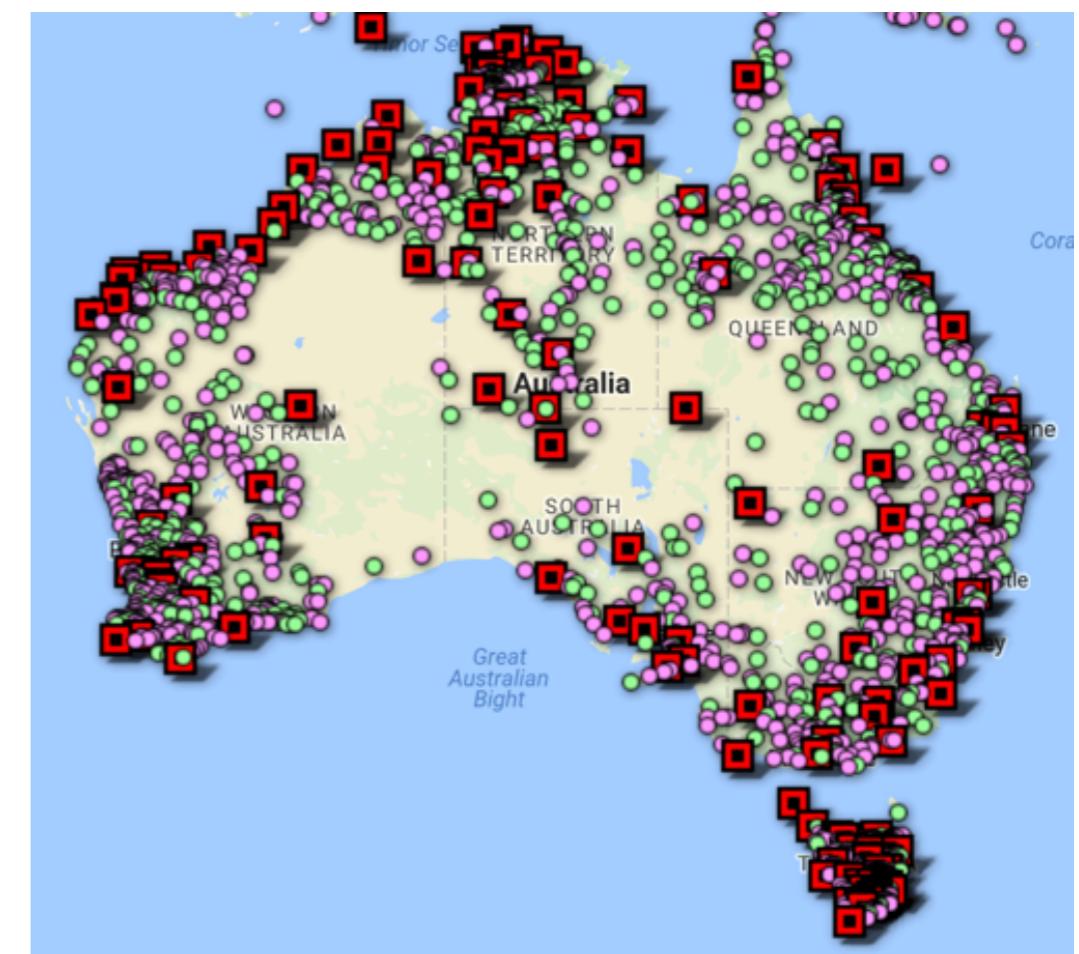
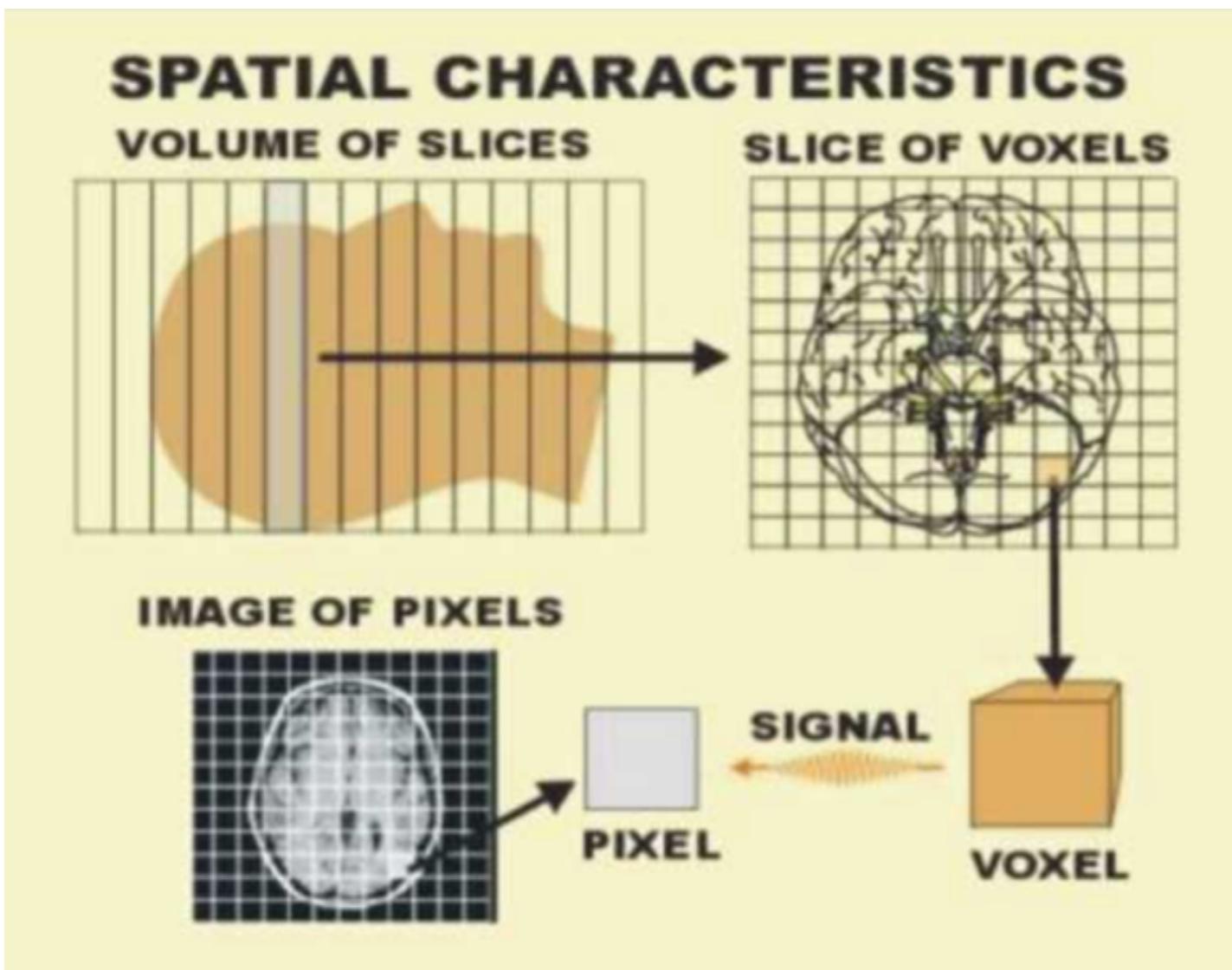
## ATTRIBUTES

## LINKS

## POSITIONS

## GRIDS

- ▶ strategy for sampling continuous data

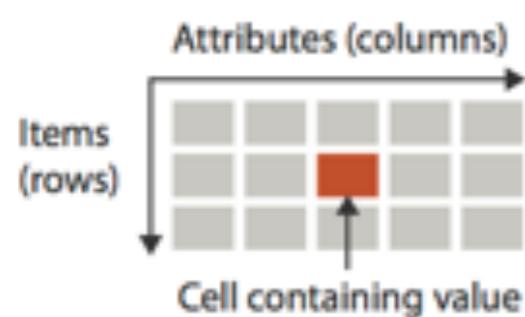


We can not measure weather in every location

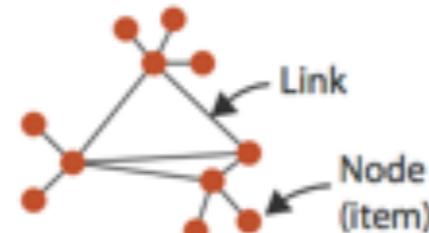
# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

## → Dataset Types

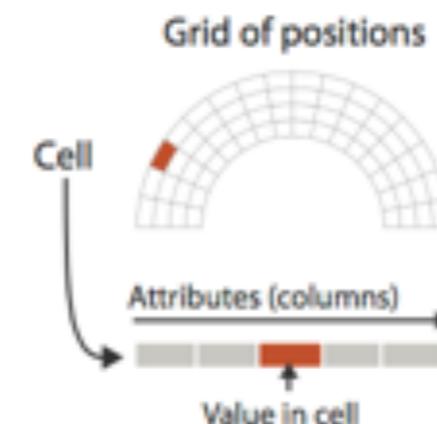
→ Tables



→ Networks



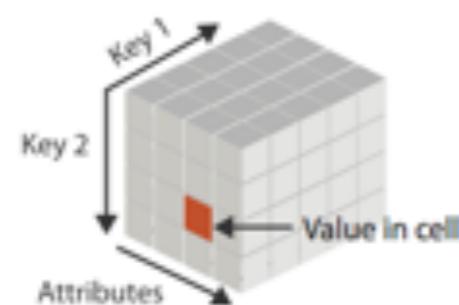
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



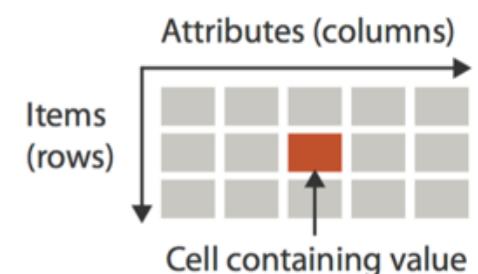
→ Trees



---

## DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

- ▶ a data set that is in the form of a table made up of rows and columns
- ▶ flat table:
  - ▶ row => item of data
  - ▶ column => attribute of data set
  - ▶ cell => value specified by item and attribute



Tables

Items

Attributes

# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

Attributes (columns)					
ID	Name	Age	Shirt Size	Favorite Fruit	
1	Amy	8	S	Apple	
2	Basil	7	S	Pear	
3	Clara	9	M	Durian	
4	Desmond	13	L	Elderberry	
5	Ernest	12	L	Peach	
6	Fanny	10	S	Lychee	
7	George	9	M	Orange	
8	Hector	8	L	Loquat	
9	Ida	10	M	Pear	
10	Amy	12	M	Orange	

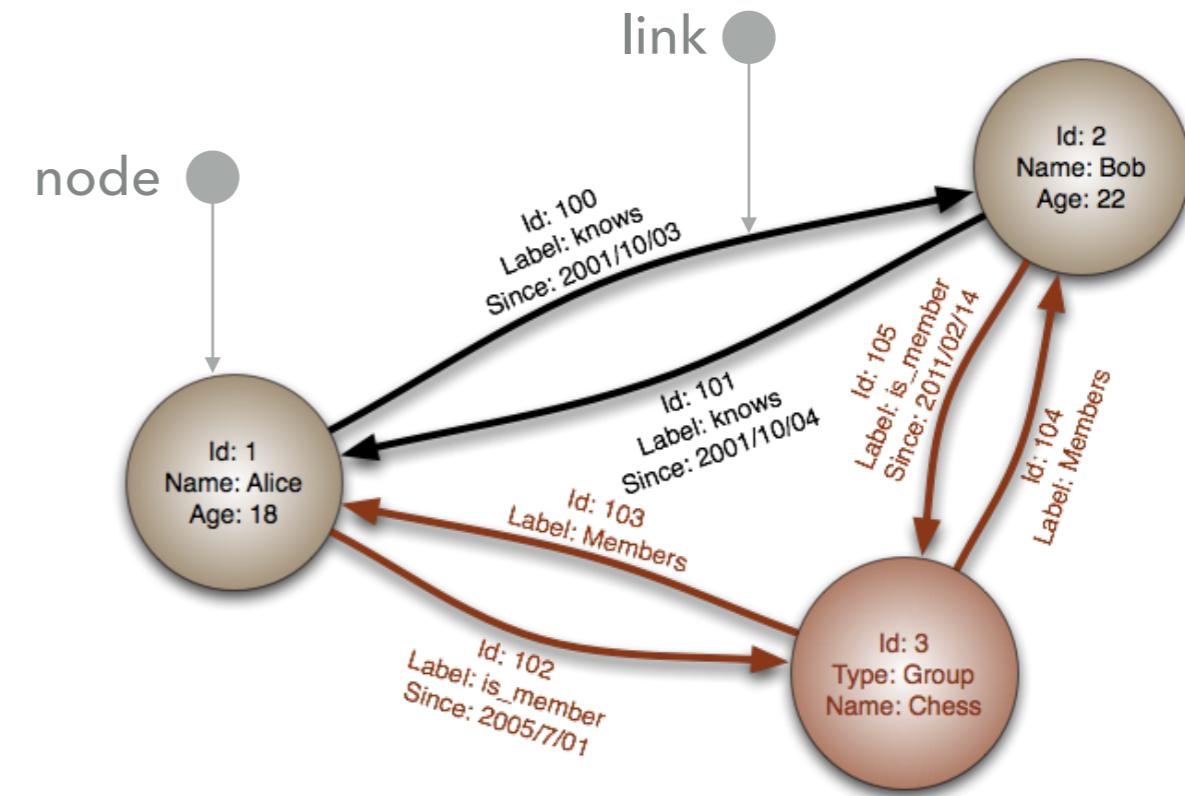
Items (rows)

Cell

More on tables later in semester

# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

- ▶ a data set in the form of items (nodes) and the relationship between them (links)
- ▶ links can go one or both ways
- ▶ nodes can also be associated with attributes



More on networks and trees (a special type of network) later in semester

Networks are also sometimes called 'graphs' (not the same meaning as in statistics - bar graph, line graph are different kind of thing)

Networks & Trees

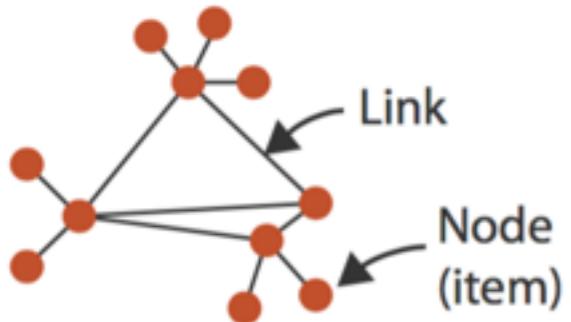
Items (nodes)

Links

Attributes

# NETWORK VISUALISATION

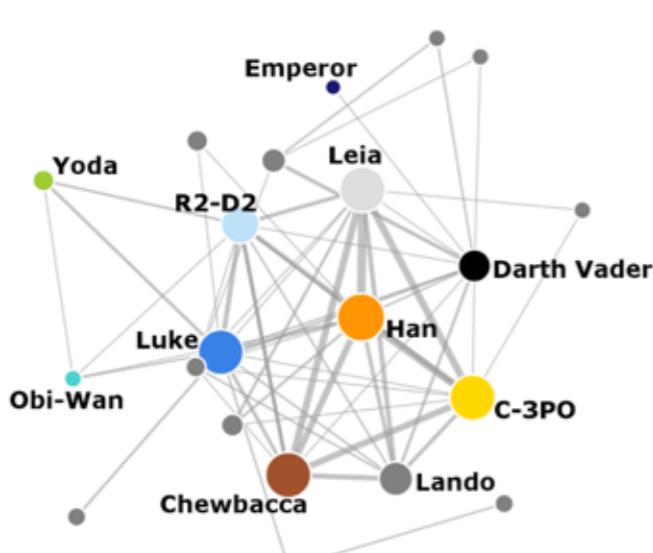
Munzner (2014)



→ Trees

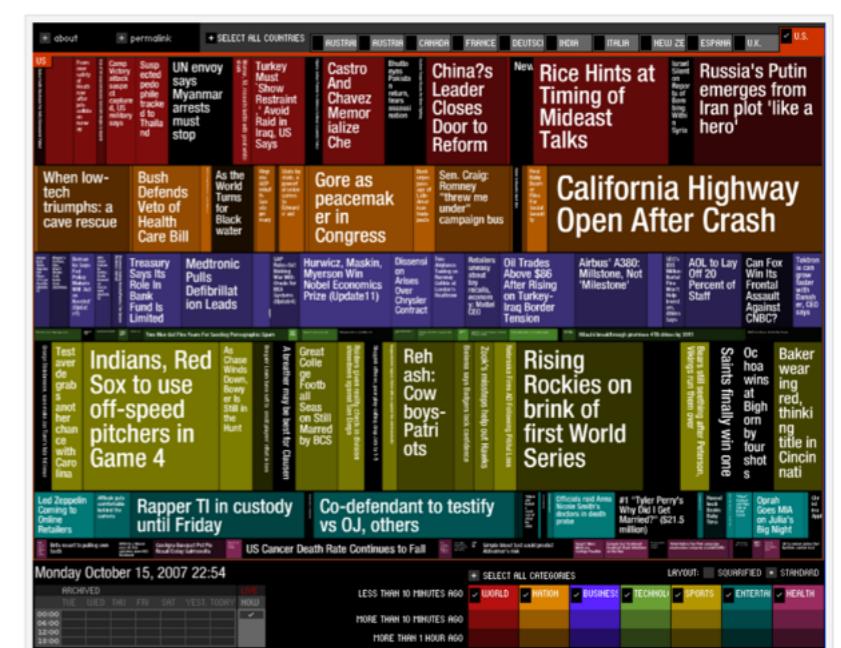
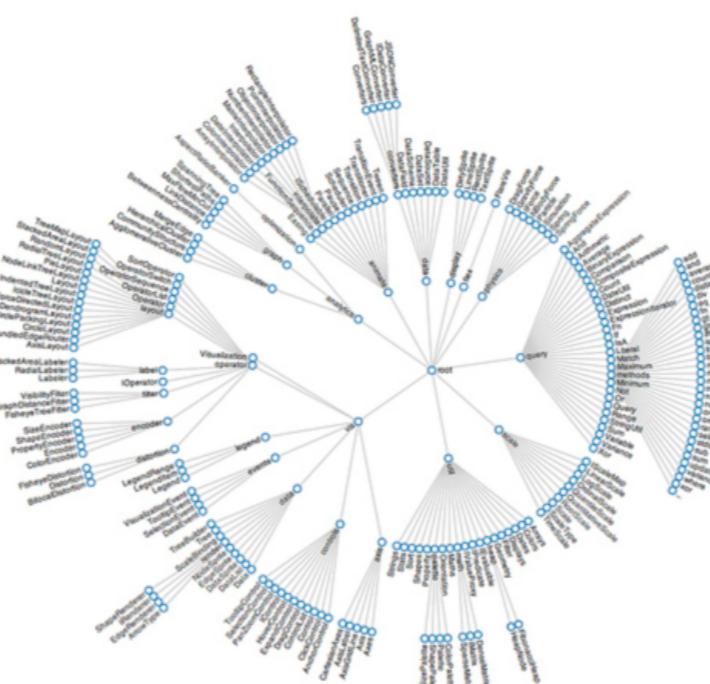


## Episode V: The Empire Strikes Back



Open network

Characters are connected by a link if they speak in the same scene.  
The more they speak together the thicker the line. E. Gabasova



# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

- ▶ Measurements from a continuous domain

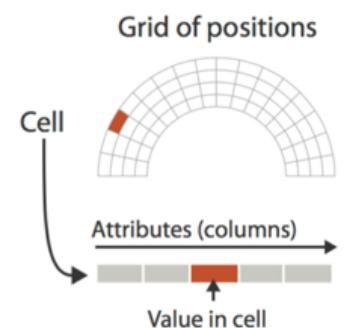
- ▶ For example,

→ Fields (Continuous)

- ▶ temperature, pressure, speed, force

- ▶ Requires sampling

- ▶ Interpolation (to show values between sampled points)



Fields

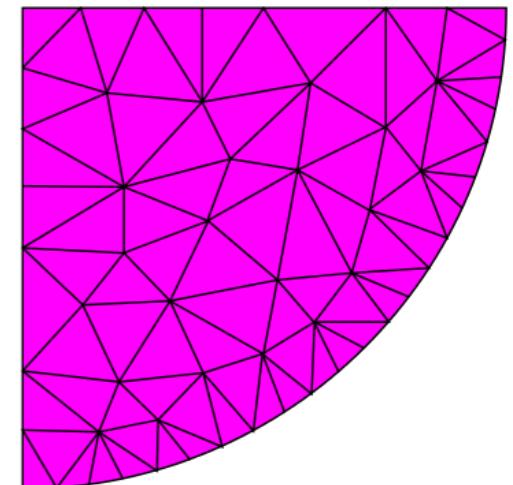
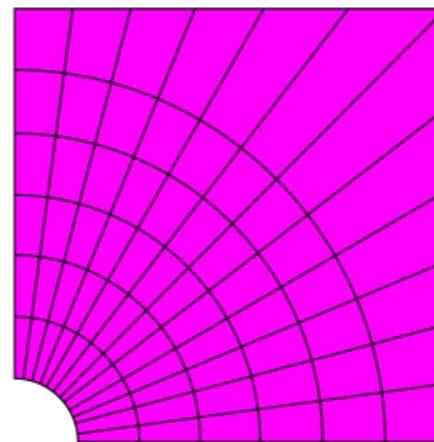
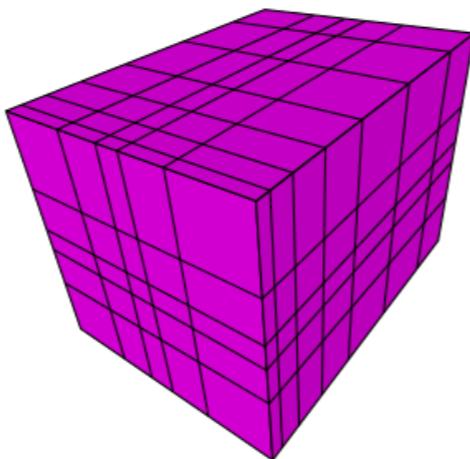
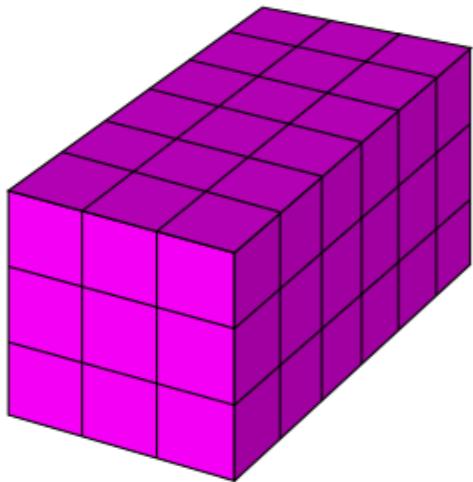
Grids

Positions

Attributes

# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

## Sampling grids



- ▶ Uniform grid
  - ▶ regular sampling
  - ▶ geometry and topology does not have to be stored explicitly

- ▶ Rectilinear grid
  - ▶ non-uniform sampling
  - ▶ good where you need to focus on some areas more than others

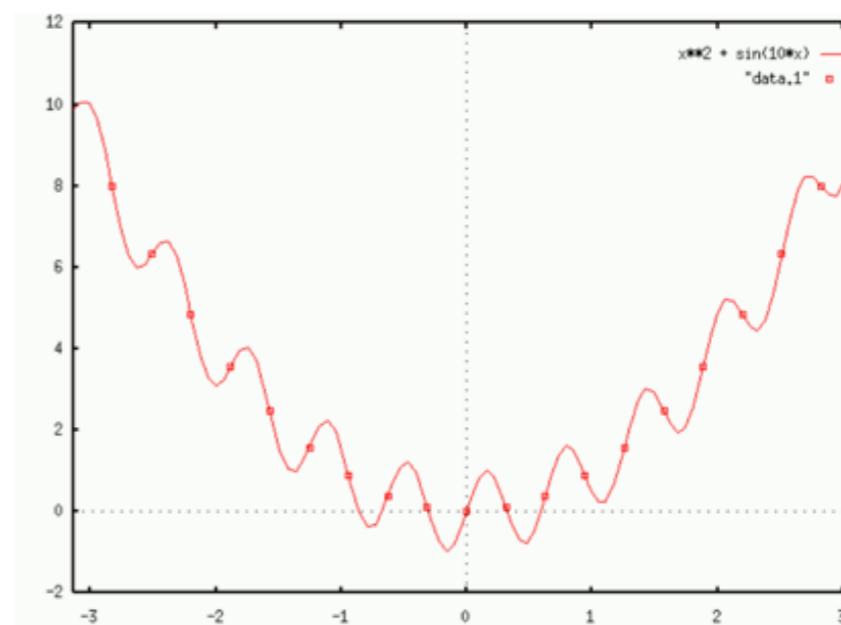
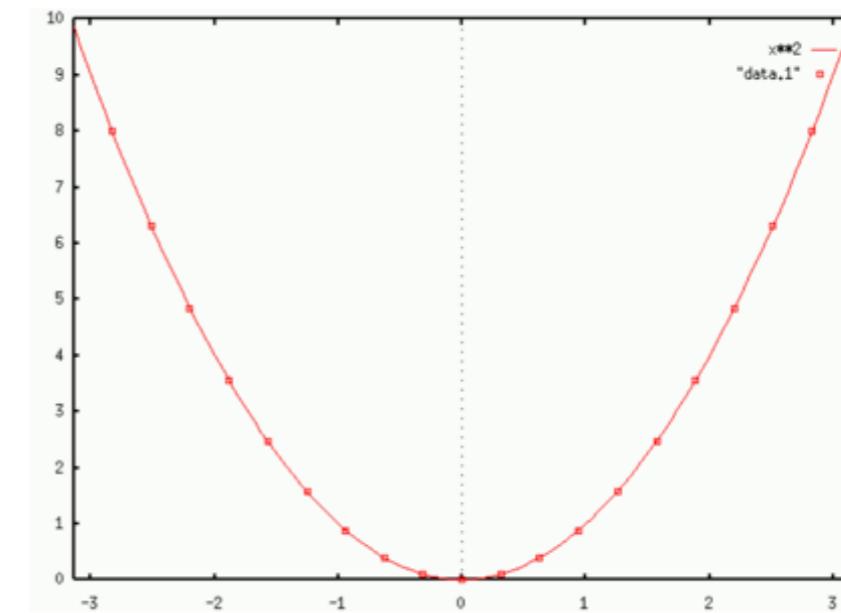
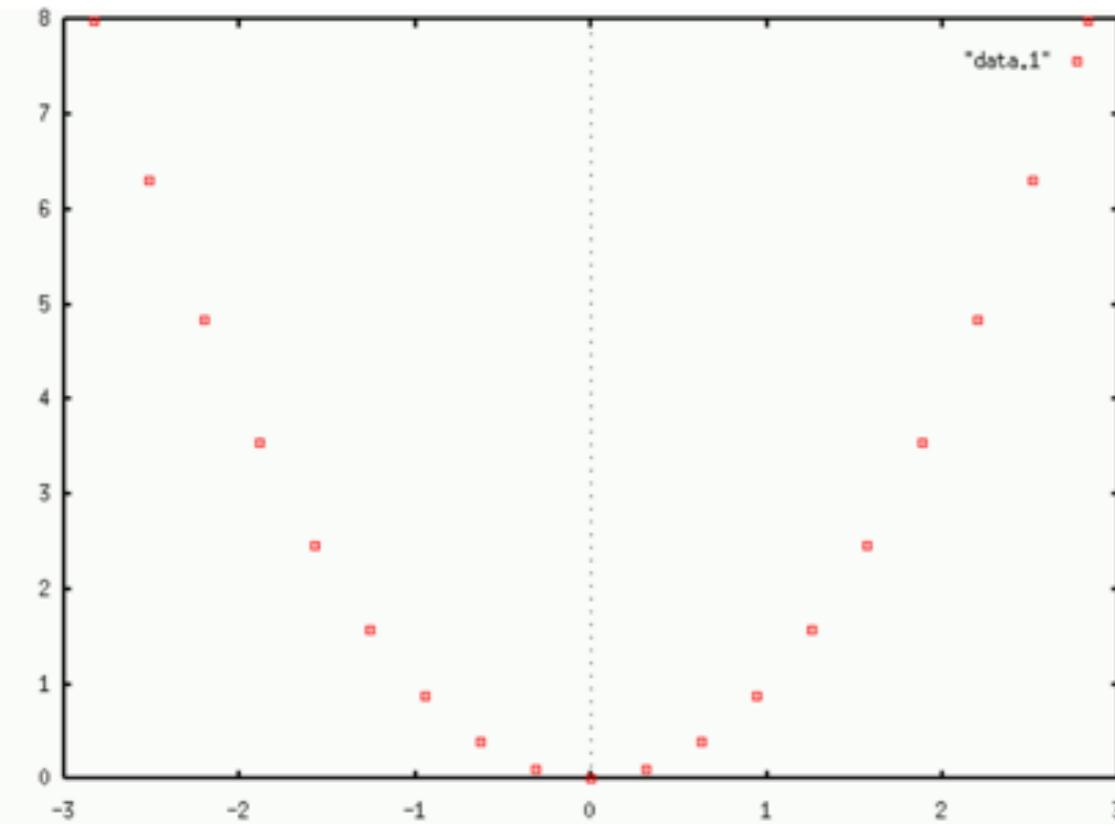
- ▶ Structured grid
  - ▶ allows curvilinear grids
  - ▶ location of each grid need to be specified

- ▶ Unstructured grid
  - ▶ flexible, stores position and connection
  - ▶ social networks
  - ▶ computationally expensive

---

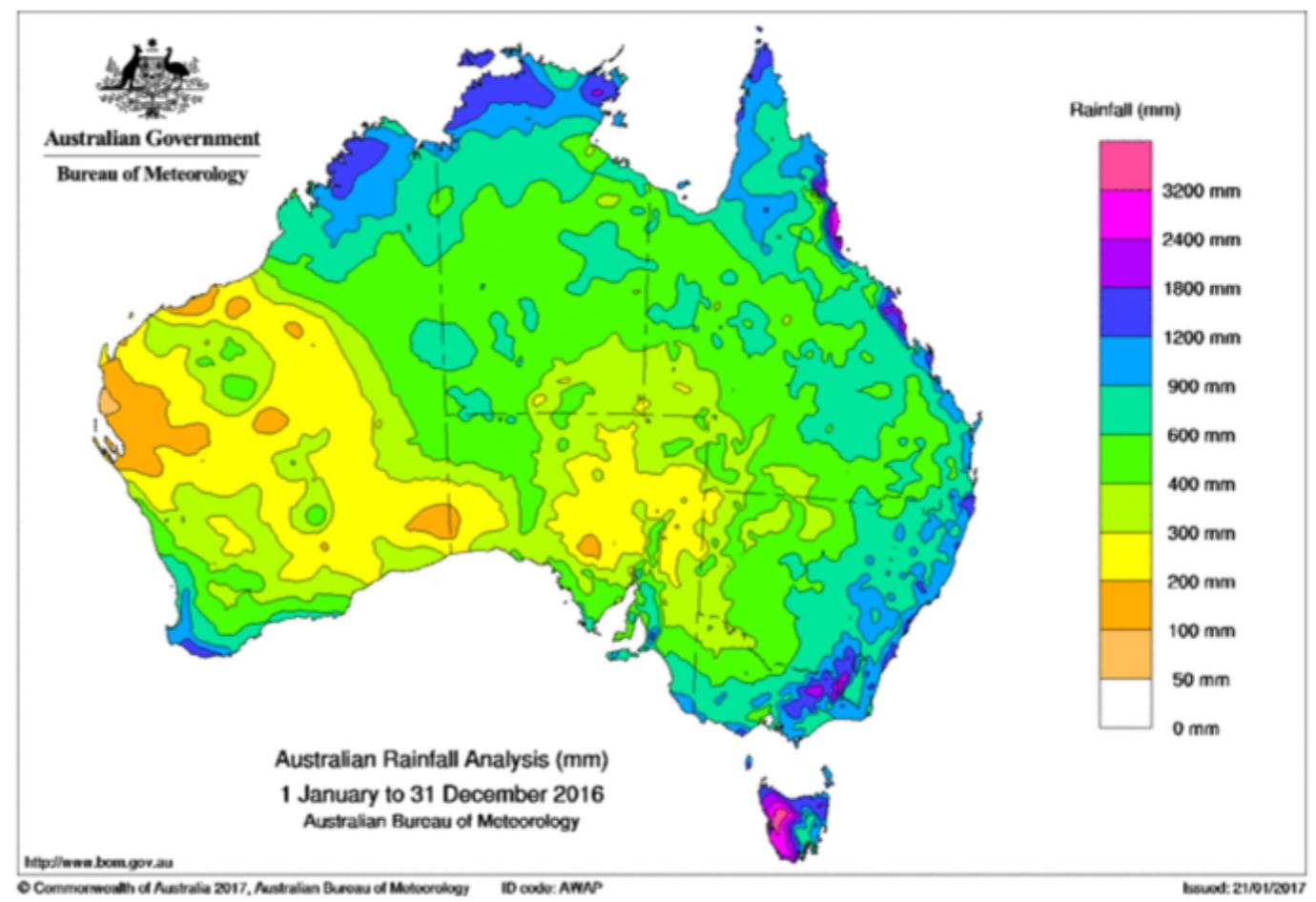
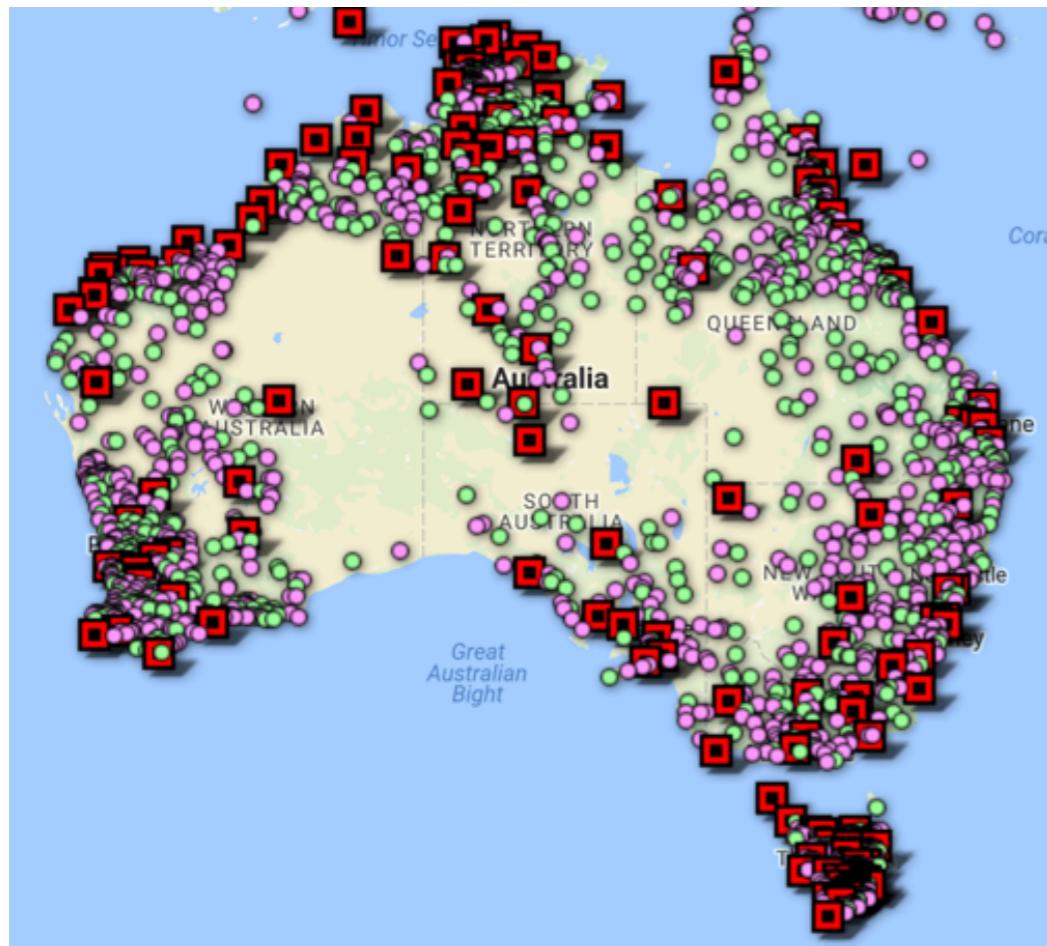
# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

## ► Interpolation



# DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

- ▶ Measurements from a continuous domain

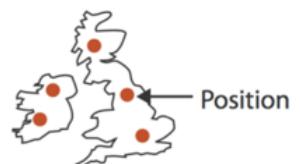


---

## DATASET TYPES: TABLES NETWORKS FIELDS GEOMETRY

- ▶ Specification of the shapes of items in a spatial position
- ▶ Used where shape is important feature of data
- ▶ Explicit spatial positions
  - ▶ points, lines, curves, surface, regions, volumes
- ▶ important in computer graphics (not core data vis)

→ Geometry (Spatial)



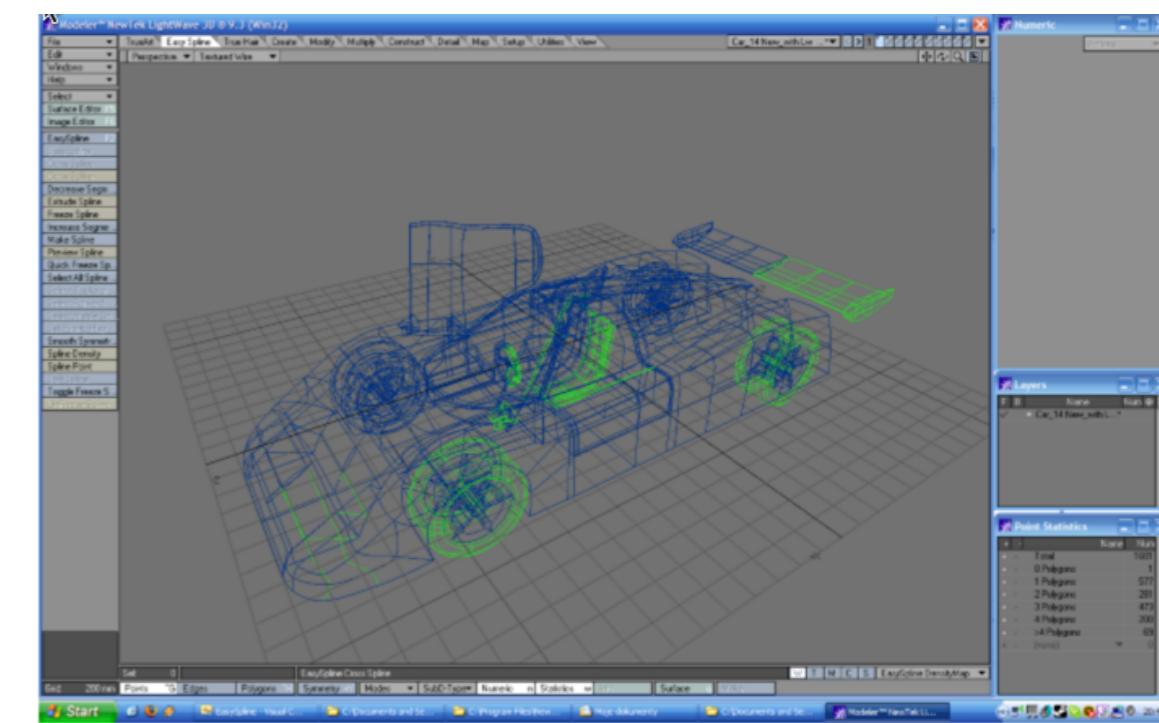
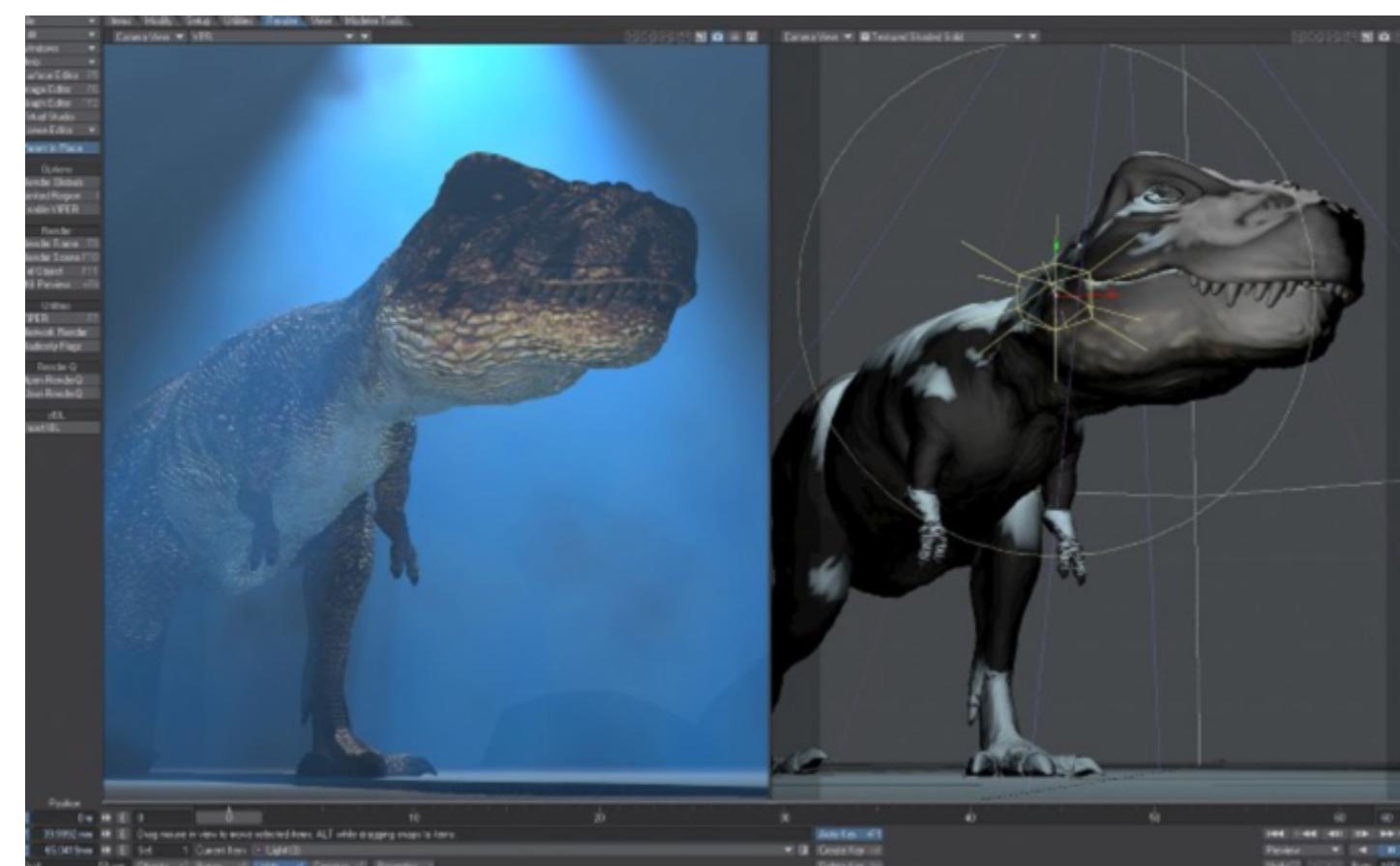
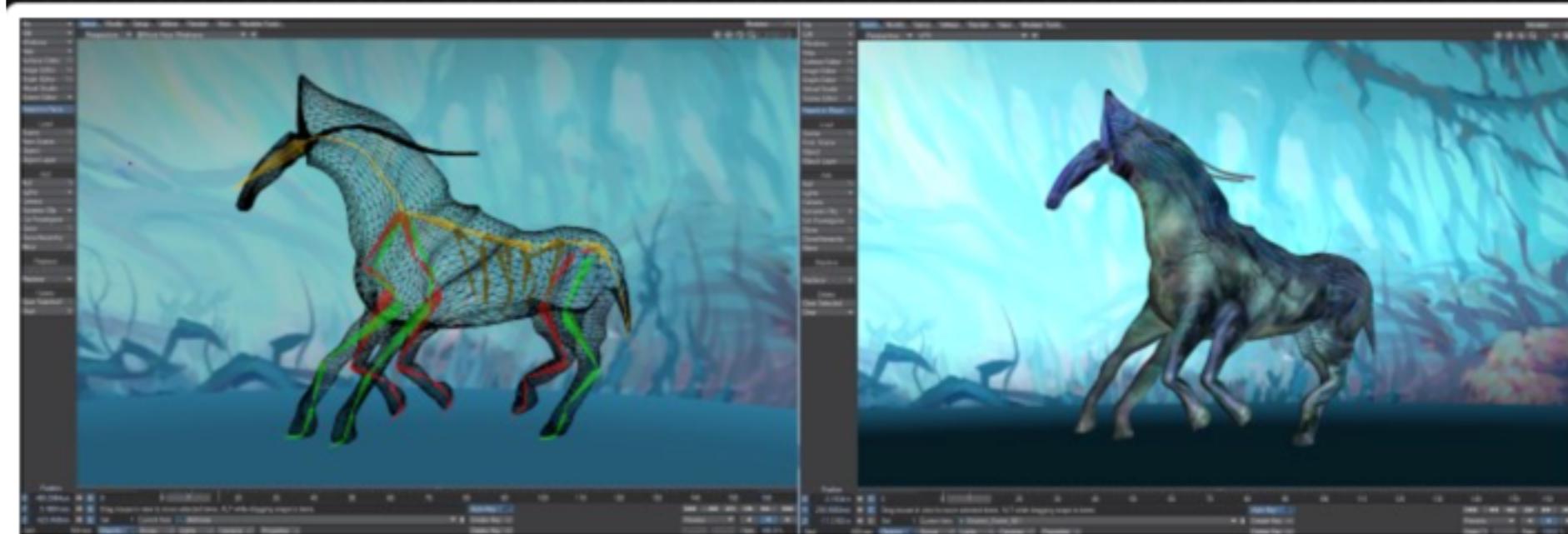
Geometry

Items

Positions

# GEOMETRY

- ▶ More computer graphics domain, than data vis

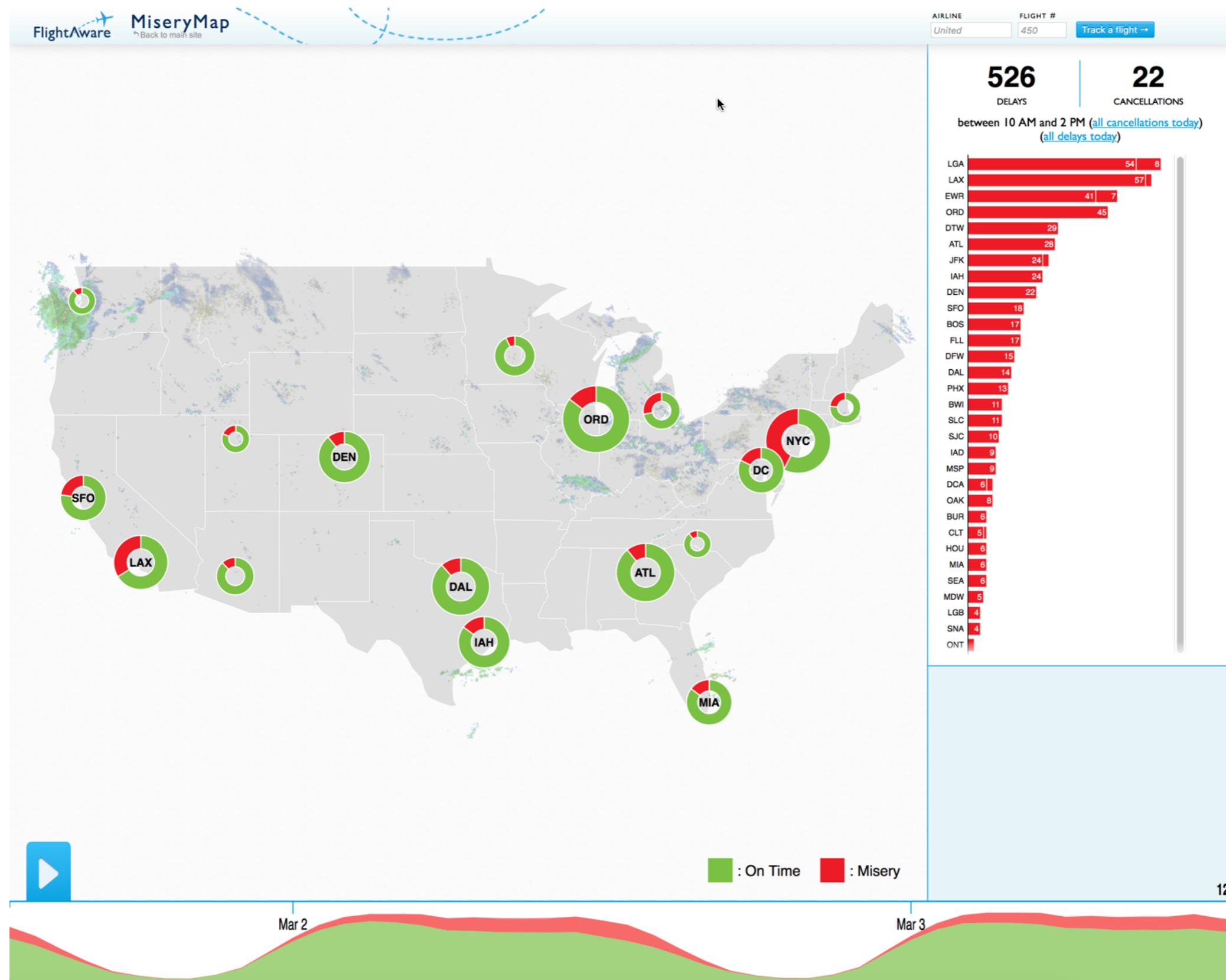


# DATA SET AVAILABILITY

- ▶ static/off line
- ▶ dynamic/online (e.g., updated web page)
- ▶ more complicated to visualise when data set is streaming

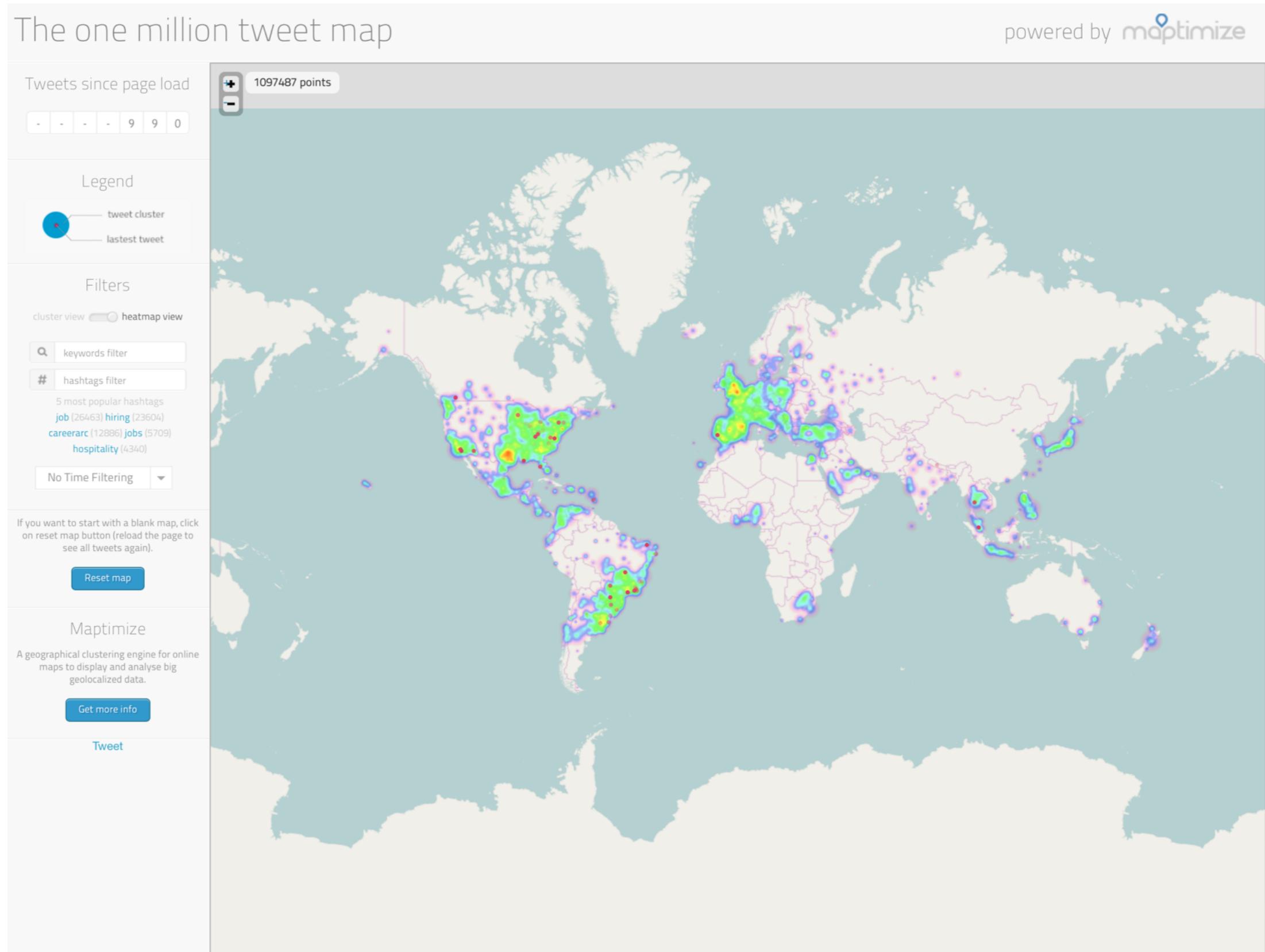


TEXT



# TEXT

---



# KEYS

## ▶ Key

- ▶ 'an attribute that acts as index that is used to look up value attributes' (Munzner, 2014)

Explicit  
Key

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

- ▶ can be explicit (i.e., contained in table as an attribute)
- ▶ no duplicates
- ▶ categorial or ordinal data type
- ▶ usually stored in first column

Implicit  
Key

A	B	C
Order ID	Order Date	Order Priority
1		
2	3 10/14/06	5-Low
3	6 2/21/08	4-Not Specified
4	32 7/16/07	2-High
5	32 7/16/07	2-High
6	32 7/16/07	2-High
7	32 7/16/07	2-High
8	35 10/23/07	4-Not Specified
9	35 10/23/07	4-Not Specified
10	36 11/3/07	1-Urgent
11	65 3/18/07	1-Urgent
12	66 1/20/05	5-Low
13	69 6/4/05	4-Not Specified
14	69 6/4/05	4-Not Specified
15	70 12/18/06	5-Low
16	70 12/18/06	5-Low
17	96 4/17/05	2-High
18	97 1/29/06	3-Medium
19	129 11/19/08	5-Low
20	130 5/8/08	2-High
21	130 5/8/08	2-High
22	130 5/8/08	2-High
23	132 6/11/06	3-Medium
24	132 6/11/06	3-Medium
25	134 5/1/08	4-Not Specified
26	135 10/21/07	4-Not Specified
27	166 9/12/07	2-High
28	193 8/8/06	1-Urgent
29	194 4/5/08	3-Medium

# DATA ABSTRACTION ACTIVITY

- ▶ Learn how to recognise data set and attribute types
- ▶ Learn how to ask questions about the data and transform it in a way that enables you to answer them...



Aid Data taken from here: <a href="http://aiddata.org/country-level-research-datasets">http://aiddata.org/country-level-research-datasets</a>						
aiddata_id	year	donor	recipient	commitment	coalesced_pc	coalesced_purpose_name
12191891	1998	Australia	Indonesia	103323	24040	Informal/semi-formal fin. interm
27339565	2003	Australia	Singapore	297.969	15105	Government and civil society, pu
34378730	2007	Australia	Colombia	139556	15130	Legal and judicial development
13308187	1999	Australia	Timor-Leste	22715.3	12220	Basic health care
38835178	2008	Australia	Bilateral, unspecified	44685.5	15110	Public sector policy and adm. ma
23209843	2004	Australia	Solomon Islands	62942.8	31282	Forestry research
13311786	1999	Australia	Papua New Guinea	35199.6	31110	Agricultural policy & admin. mgn
94862076	2010	Australia	Papua New Guinea	86084.5	31192	Plant/post-harvest prot. & pest c
34451610	2007	Australia	Solomon Islands	30313.2	43010	Multisector aid
23170622	2004	Australia	Cambodia	858311	25010	Business support services & insti
31064132	2007	Austria	Uganda	9836.32	12191	Medical services
50005778	2009	Austria	Africa, regional	5974.79	11110	Education policy & admin. manag
17016824	2002	Austria	Nicaragua	22220.3	14030	Basic drinking water supply and i
15316543	2001	Austria	Algeria	155768	11420	Higher education
11293258	1998	Austria	Bilateral, unspecified	30579.6	24040	Informal/semi-formal fin. interm
35164208	2008	Austria	Malawi	21105.3	13040	Std control including hiv/aids
23611101	2005	Austria	Nicaragua	689009	11420	Higher education
50005703	2009	Austria	Uganda	20888.5	11330	Vocational training
15333611	2001	Austria	Africa, regional	730622	24040	Informal/semi-formal fin. interm
15328132	2001	Austria	Mozambique	637610	15153	Media and free flow of informati
21235774	2004	Belgium	El Salvador	315565	92010	Support to national ngos
35341959	2008	Belgium	Bilateral, unspecified	211247	99810	Sectors not specified
50006855	2009	Belgium	Congo, Dem. Rep.	115955	31182	Agricultural research
94910733	2010	Belgium	Senegal	3431.15	16050	Multisector aid for basic soc. ser
94659679	2010	Belgium	Peru	34686.8	13010	Population policy and admin. mg
35279084	2008	Belgium	Uganda	97175.4	31110	Agricultural policy & admin. mgn
21253341	2004	Belgium	Peru	18504.7	99810	Sectors not specified
27662196	2006	Belgium	Indonesia	21302.8	72010	Material relief assistance and ser
19009279	2003	Belgium	Bilateral, unspecified	68852.7	99820	Promotion of development awar
15372708	2001	Belgium	Burundi	9023.14	31194	Agricultural co-operatives
35285348	2008	Belgium	Africa, regional	2816.62	99820	Promotion of development awar
18996118	2003	Belgium	Nepal	978.04	16010	Social/welfare services
35240450	2008	Belgium	Benin	86604	16050	Multisector aid for basic soc. ser
17066156	2002	Belgium	Senegal	1710.86	11130	Teacher training
94940851	2010	Belgium	Bilateral, unspecified	745976	43082	Research/scientific institutions
95022218	2010	Belgium	Guatemala	139997	74010	Disaster prevention and prepare
21203222	2004	Belgium	Congo, Dem. Rep.	6265.15	43081	Multisector education/training
15396456	2001	Belgium	Zambia	22974.1	92030	Support to local and regional ngc
50070709	2009	Belgium	Bilateral, unspecified	4177.69	99820	Promotion of development awar
27589920	2006	Belgium	Guinea	13046.1	16061	Culture and recreation



# REVIEW

---

## What:

- ▶ understand the different types of data

## Why:

- ▶ the type of data available will drive the design

## How:

- ▶ now: understand the properties of different data types
- ▶ later: look at how to best represent different data types in visualisation