Data Types

Learning Objectives

- What are data types?
- What are the kinds of data we can identify?

Quantitative vs Qualitative

Quantitative data

- measures of values or counts
- expressed as numbers
- data about numeric variables
- e.g. how many; how much; or how often

Quantitative data

Examples:

Data unit	Numeric variable = Quantitative data	
A person	How many hours do you work?	37.5 hours per week
A house	How many square metres is the house?	200 square metres
A business	How much was the last year's profit?	$300, 000
A farm	How many ducks are located on the farm?	100 ducks
A school	How many students are currently enrolled?	5,000 students

Australian Bureau of Statistics. *Statistical Language - Quantitative and Qualitative Data*. https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Statistical+Language+-+quantitative+and+qualitative+data

Qualitative data

- measures of “types”
- may be represented by a name, symbol, or a number code student ID
- data about categorical variables
- e.g. what type

Qualitative data

Examples:

Data unit	Categorical variable = Qualitative data	
A person	Do you work part-time or full-time?	Full-time
A house	In which city is the house located?	Sydney
A business	What type of structure is the business?	Joint-venture
A farm	What is the main activity of the farm?	Poultry
A school	Is it a public or private school?	Public

Australian Bureau of Statistics. *Statistical Language - Quantitative and Qualitative Data*. https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Statistical+Language+-+quantitative+and+qualitative+data

Structured vs Unstructured

Structured data

- Organised information in a database
- Can be fit into a spreadsheet
- Easier to handle

Structured data

Examples:

- Microsoft Excel files (xls, xlsx, xlsm)
- Text files (csv, txt, tab, tsv)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	price	date_sold	suburb	num_bath	num_bed	num_parki	property_t	type	suburb_po	suburb_mv	suburb_sq	suburb_lat	suburb_lng	suburb_ele	cash_rate	property_i	km_from_cbd						
2	530000	########	Kincumber	4	4	2	1351	House	7093	29432	9.914	-33.4725	151.4021	24	2	150.9	47.05						
3	525000	########	Halekulani	2	4	2	594	House	2538	24752	1.397	-33.2177	151.5524	23	2	150.9	78.54						
4	480000	########	Chittaway	2	4	2	468	House	2028	31668	1.116	-33.3268	151.4456	3	2	150.9	63.59						
5	452000	########	Leumeah	1	3	1	344	House	9835	32292	4.055	-34.0538	150.8396	81	2	150.9	40.12						
6	365500	########	North Avo	0	0	0	1850	Vacant lan	2200	45084	1.497	-33.4561	151.436	18	2	150.9	49.98						
7	550000	########	Kincumber	1	3	1	626	House	7093	29432	9.914	-33.4725	151.4021	24	2	150.9	47.05						
8	535000	########	Bensville	1	3	1	556	House	2545	36764	4.925	-33.4991	151.3905	27	2	150.9	43.91						
9	495000	########	Leumeah	1	3	2	582	House	9835	32292	4.055	-34.0538	150.8396	81	2	150.9	40.12						
10	410000	########	Toukley	1	3	3	493	House	4550	25844	3.683	-33.258	151.5432	4	2	150.9	74.11						
11	242500	########	Winmalee	0	0	0	1248	Vacant lan	6202	38740	9.058	-33.6797	150.6112	263	2	150.9	59.15						
12	6500000	########	Point Clare	1	3	2	742	House	3731	31772	3.336	-33.4402	151.3172	62	2	150.9	48.17						
13	890000	########	Picnic Poir	2	4	3	715	House	6160	40560	3.859	-33.973	151.0063	33	2	150.9	22.31						
14	533000	########	Whalan	3	4	2	695	House	5973	24180	2.429	-33.7557	150.8036	37	2	150.9	39.53						
15	1120500	########	North Roc	2	4	2	904	House	7965	40092	5.462	-33.7757	151.0147	92	2	150.9	20.61						
16	830000	########	Winmalee	3	6	2	2109	House	6202	38740	9.058	-33.6797	150.6112	263	2	150.9	59.15						
17	675000	########	Bass Hill	3	3	2	263	Townhous	9069	24388	2.929	-33.9003	150.9931	40	2	150.9	20.43						
18	500000	########	Kincumber	1	3	1	791	House	7093	29432	9.914	-33.4725	151.4021	24	2	150.9	47.05						
19	473000	########	Leumeah	1	3	3	581	House	9835	32292	4.055	-34.0538	150.8396	81	2	150.9	40.12						
20	885000	########	Picnic Poir	1	3	2	557	Vacant lan	6160	40560	3.859	-33.973	151.0063	33	2	150.9	22.31						
21	625000	########	Chittaway	2	4	2	555	House	2028	31668	1.116	-33.3268	151.4456	3	2	150.9	63.59						
22	520000	########	Leumeah	1	3	1	651	House	9835	32292	4.055	-34.0538	150.8396	81	2	150.9	40.12						
23	510000	########	Winmalee	1	3	1	993	House	6202	38740	9.058	-33.6797	150.6112	263	2	150.9	59.15						
24	610000	########	Chittaway	3	5	8	862	House	2028	31668	1.116	-33.3268	151.4456	3	2	150.9	63.59						
25	570000	########	Winmalee	1	3	2	828	House	6202	38740	9.058	-33.6797	150.6112	263	2	150.9	59.15						
26	555000	########	Bensville	2	4	0	748	House	2545	36764	4.925	-33.4991	151.3905	27	2	150.9	43.91						
			domain_properties																				

Unstructured data

- Freeform information
- Cannot be fit into a spreadsheet
- More difficult to handle

Unstructured data

Examples:

- Videos
- Audios
- Images
- Textual, e.g. emails, text messages
- Webpages
- pdfs

Can we visualize unstructured data?

- Machine Learning
- Natural Language Processing

Metadata

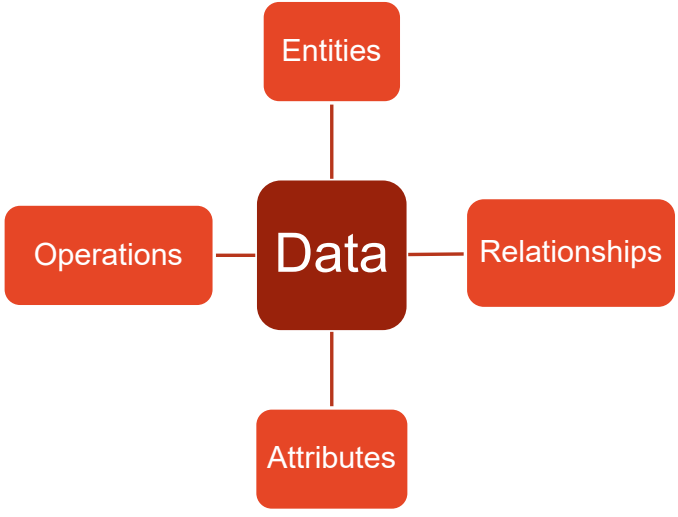
- Derived data
- Data that describes other data

Examples

- A photo – file name, author, date captured, file size etc
- A book – version, author, publisher details, table of content etc
- An email – subject, from, to, date and time sent, sending and receiving server names and IPs etc
- A spreadsheet – tab names, column names, user comments etc

Data Model

Model to describe data



Ware, C. (2013). *Information Visualization: Perception for Design*. Elsevier Science. https://books.google.com.au/books?id=qFmS95vf6H8C

Data model: Entities and Relationships

Entities

Objects of interest /Values

- Can be single
  - people, hurricanes, fish etc
- Can be a group
  - a school of fish

Relationships

Structures that relate entities

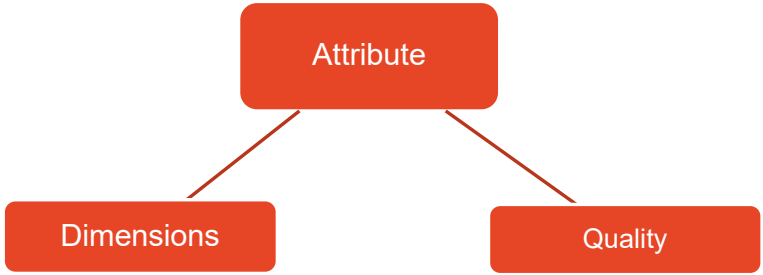
- Can be structural & physical
- Can be conceptual
- Can be causal
- Can be temporal

Data model: Attributes

Attributes of Entities or Relationships

A property of some entity – not independent

- For examples:
  - Colour of a lychee is an attribute of lychee
  - Temperature of water
  - Duration of a trip



Dimensions of Attribute

- Scalar
  - Weight of person
- Vector
  - Direction of travel
- Tensor
  - Direction + Shear force
  - Stressed object
- Field of scalars /vectors /tensors
  - Gravitational field of earth: 3D vector
  - Gravity strength at earth’s surface: 2D scalar

Quality of Attribute (Statistics)

Four Levels of measurement:

- Nominal
  - Labelling purpose
  - Examples?
- Ordinal
  - Can be ordered in a sequence
  - Examples?
- Interval
  - Able to derive the gap between values
  - Examples?
- Ratio
  - Full expression of a real number
  - Examples?

cannot swap eg low, medium, high

eg. temperature F ✓.
low → medium ≠ medium → high

can find absolute zero
eg. 0 kelvin → absolute 0
0 F ≠ absolute 0, can be negative

Quality of Attribute (Computer Programming)

- Category data
  - Nominal scale
- Integer data
  - Ordinal scale
- Real-number data
  - Interval & ratio scale

ENUM

Data model: Operations

Operations

What about processes that are performed on entities & relationships? They are also considered as data!

- Mathematical operation on numbers (+ - x /)
- Merging lists (2 or more into 1)
- Inverting a value (creating opposite)
- Bringing an entity /relationship into existence (mean of the set)
- Deleting an entity / relationship (breakups)
- Transforming (froglet to adult frog)
- Forming (pie from apple & pastry)
- Splitting a single entry (disassemble of machine)

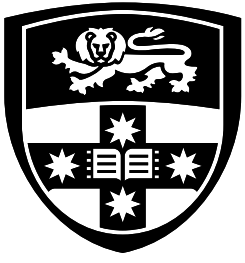
Knowledge Check

What types of data are they?

Imagine that you host a massive multiplayer online game system (such as Fortnite, Minecraft, etc). The game is a team(clan)-based game, and each team (clan) has different numbers of players. There are also different ranks/classes in the game and each rank/class provide a player with access to different game features. Each player also has his/her own friend list, where you can register your friend regardless of the team/clan. Each player has an in-game monetary balance, which allow you to purchase different game items. Each player has its game statistics (such as the number of kill, death)

Summary

- What are data types?
  - Forms of data for us to visualise
- What are kinds of data we can identify?
  - *Quantitative vs Qualitative*
  - *Structured vs Unstructured*
  - *Attributes vs Relationships*
  - *Nominal, Ordinal, Interval and Ratio*
  - *Operation*



THE UNIVERSITY OF
SYDNEY