

# What is Data

Wednesday, November 10, 2021

10:36 PM

Example of a table of data (dataset).

size of house (sq ft)	# of bathrooms	price (000\$)
523	1	115
645	1	150

how to choose A & B  
what is A & B are determined  
by usage.

image	label
cat	cat
bike	not cat
cat	cat
car	not cat

- Acquiring data
- manual labeling
  - from observing behaviors.

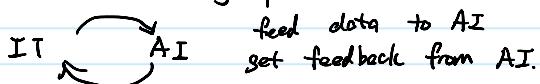
user	time	price	purchased
XY	1	1	✓
YY	2	2	✓
DD	3	3	✗

machine	temp (°C)	pressure (psi)	machine fault

- Download from websites / partnerships

Use and mis-use of data

- don't wait for building perfect data



- Just have more data, it doesn't mean it'll be valuable

Data is messy

- garbage in, garbage out
- data problems
  - o incorrect labels
  - o missing values
- multiple types of data
  - images, audio, text

unstructured

Size of house	# bdrm	price
523	1	113
634	1	001
unknown	unknown	
unknown		

structured