Week 2:
data types and central tendency

What is data?

What is data?

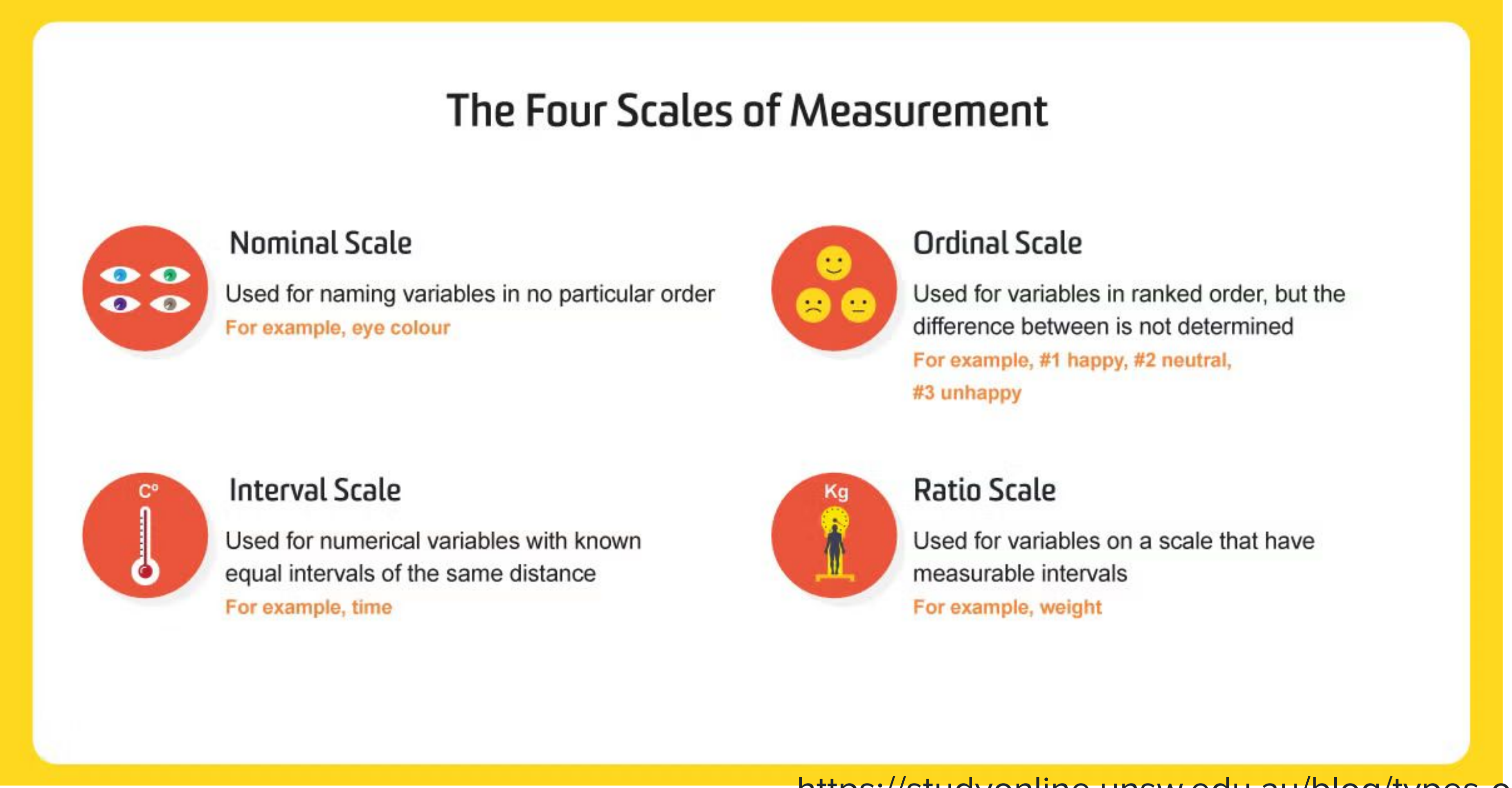
- Abstracted representation of reality
- Limited by what *can be* and what *is* measured

What is your data?

- Discrete or continuous?
- Samples or population?
- Records or summary?

Data types

Data types



The Four Scales of Measurement

Nominal Scale

Used for naming variables in no particular order

For example, eye colour

Ordinal Scale

Used for variables in ranked order, but the difference between is not determined

For example, #1 happy, #2 neutral, #3 unhappy

Interval Scale

Used for numerical variables with known equal intervals of the same distance

For example, time

Ratio Scale

Used for variables on a scale that have measurable intervals

For example, weight

https://studyonline.unsw.edu.au/blog/types-of-data

Data types

Numeric:

- Interval scale
  - : only intervals are meaningful
- Ratio scale
  - : zero is meaningful

Data types

Categorical:

- Ordinal
  - : rank or order is meaningful
- Nominal
  - : unordered categories

Data types

Dates, times:

- Special case of ordinal or interval

Data types in Python

Python data types

`int`

- `42`

`float`

- `3.54`

Python data types

`str`

- `'Hello'`
- `"World"`
- `"What's your name?"`

-
  ```
  '''
  The quick brown fox
  jumps over the lazy dog
  '''
  ```

  (single or double quotes)

Python data types

`bool`

- `True, False`

`None`

- `None`

Python data structures

`list`

- `[1,2,3,4]`
- `['red','yellow','blue']`
- `[1,'blue',3]`

Python data structures

`tuple`

- `(1,2,3,4)`

  (immutable)

`set`

- `{1,4,2,3}`

  (unordered)

Python data structures

dict

-
  ```
  {
      'count':10,
      'color':'red',
      'flavor':'cherry'
  }
  ```

  key-value pairs

Python objects

Functions: Tools to do something.

- Take inputs
- Can return outputs

- *Called* with ( )
- *Arguments* included in the parentheses are passed to the function

Python objects

Built-in functions:

- `len()`
- `type()`

etc.

Python objects

Built-in functions:

`type()`

- Takes object
- Returns its type

Python data types

- use `type()` to find type

Data types in pandas

Data types in pandas

- Each column is a single data type

Data types in `pandas`

- Numeric
  - `int`
  - `float`

`int` cannot be null: if null values, pandas will convert to `float`

Data types in `pandas`

- `object`
  - for `str` or mixed types

Data types in `pandas`

- `category`
  - for explicit categorical variables
  - useful for ordinal variables, but often not necessary
- `bool`
- `datetime`

Data types in `pandas`

- `.dtypes` shows types

Error handling

Error are your friends!

What is typical?

Central tendency

: a single number that describes the typical, middle value in a set of values

- *summary statistics*
- *mean*, *median*, *mode*

Mean

(Average)

: sum of values divided by count of values

e.g.

- average height
- average temperature
- average number of times a coin comes up heads

Applicable especially for natural phenomena

Mean

Mean is subject to skew by outliers

- Many social phenomena show this skew
- e.g. incomes, home prices, population density, subway delays

Median

: the middle value when all values are sorted in order

May be a better representation of the “typical” value in skewed data

e.g.

- median income
- median height of buildings

Mode

: most common value in a series

Additional useful statistics

Useful statistics

- Maximum
  - `.max()`
- Minimum
  - `.min()`
- Number of unique values
  - `.nunique()`