

## Chapter 1 Section 2: Data Basics

OpenIntro Statistics, 3rd Edition

### Some definitions

- ▶ **Data:** Observations, measurements, and information that is analyzed.
- ▶ **Summary statistic:** A single number summarizing a large amount of data.
- ▶ **Data matrix:** A collection of data with each row a case and each column a variable.
- ▶ **Case:** An observational unit.
- ▶ **Variable:** A characteristic (usually one of many) that is measured from each case.
- ▶ For example, if you are testing the effects of a drug, the cases are the patients and the variables would be things like gender, height, weight, preconditions, blood pressure, death rate, happiness, amount of sleep, weight change, etc...

1/1

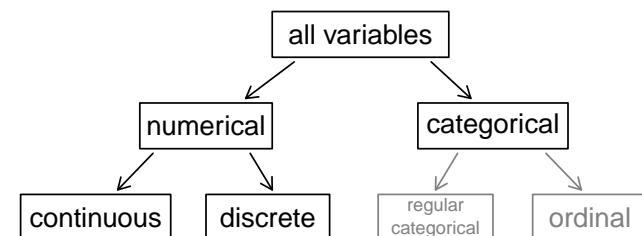
### Data matrix

Data collected on students in a statistics class on a variety of variables:

Stu.	gender	intro_extra	...	dread
1	male	extravert	...	3
2	female	extravert	...	2
3	female	introvert	...	4
4	female	extravert	...	2
⋮	⋮	⋮	⋮	⋮
86	male	extravert	...	3

← case

### Types of variables



- ▶ Numerical variables take values that can be added, subtracted, and averaged in a sensible way.
- ▶ Discrete numerical variables take on values with jumps e.g. counts, "how many".
- ▶ Continuous numerical variables take on values without jumps e.g. weights, heights, "how much".