Why do we need statistics?

Statistics is the analysis of data (and randomization)

We gather data about individuals (Observational units, or case)

Individuals → People, places, things, events

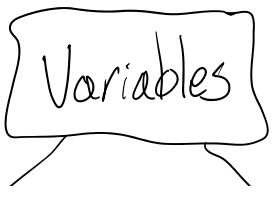
Birds

Name	Species	Weight	flight
Carl	Blue jay	.2kg	N
John	Sparrow	.1kg	Y
Jane	Penguin	10kg	N
Anne	Crow	1kg	Y
ID	Categ	Num	categ

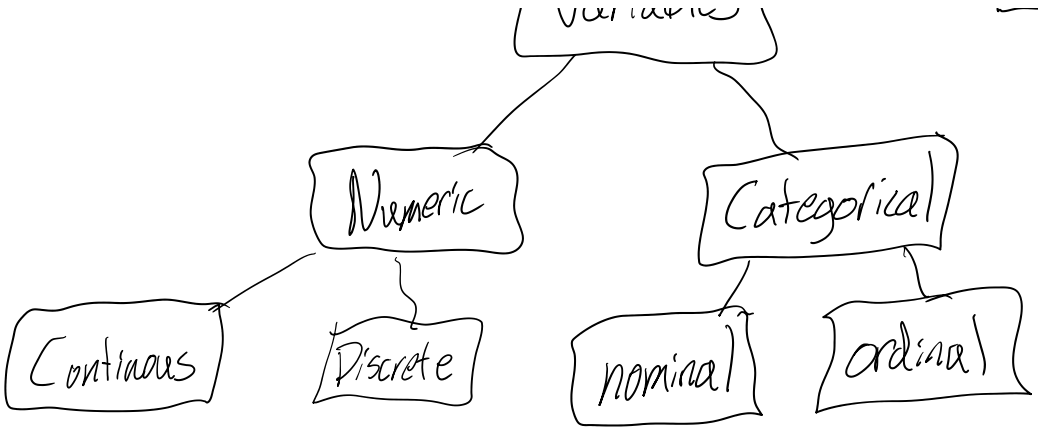
States (US)

Name	Code	Budget	Population	Vote
Wisconsin	WI	$29B	2.1M	D
Washington	WA	$45B	3.1M	D
Wyoming	WY	$31B	.9M	R
ID	ID	Num	Num	Categ

Variables → An aspect of an individual in a population that varies from individual to individual.



ID



Numeric (Quantitative) - Has inherent numeric value (ie. it is a number)
If a variable has units (ie. $, "people" kg, m/s) it is likely numerical
If we can do math (meaningfully) with the values, they are likely numerical.

Categorical (Qualitative) - break the individuals down into categories or describe some Quality

Nominal — There is no order to the categories that is relevant to the categories Ex: See above
Ordinal - There is specific ordering to the categories
Ex: Freshman, Sophomore, Junior, Sr
Ex: Some HS, GED, Associates, Bachelors, Masters, Doctoral degree
Grey area: Movie ratings ☆ 1-10 ☆

Movie A 10☆ Movie B 5☆

"Is movie A twice as good as movie B?"

Continuous → Values can take any numerical value and are infinitely divisible.

Discrete → values can on whole number numeric values

Interval vs Ratio

Interval data does not have a meaningful 0 (Zero)
Ex: Farenheit or Celcius (Can be negative)

Ex: Farenheit or Celcius (Can be negative)

Ratio data does have a meaningful zero, cannot be negative

Ex: Internet connection speed
# of Siblings

Kelvin → 0°K is no thermal energy