More Stats! (now our first Quiz is looooooming ☹)

Unit 1 – Basic Ideas

Your Seasoned Veteran Professor Colton

Unit 1 – Part 2

Data Classification

- Individuals and Data
- Types of Data

Where do Data come from?

Individuals

- The objects described by a set of data.
- Individuals may be people, but they may also be animals or things.

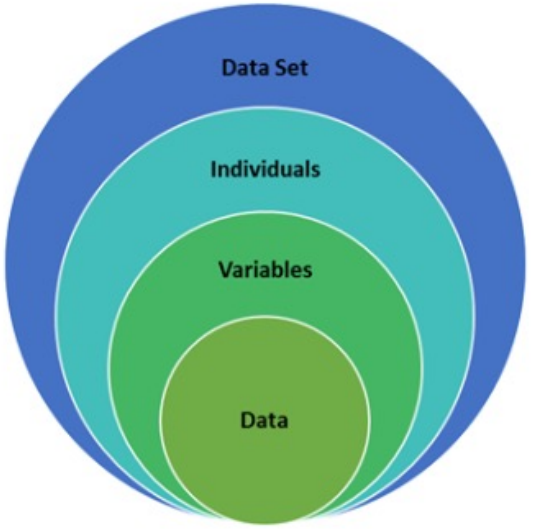
Variable

- Any characteristic of an individual.
- A variable can take different values for different individuals.

Data

- The actual measurements recorded for individuals are called data.

Structure of a dataset



Where does Data Come from?

Name	Major	Points	Grade
Advani, Sura	Comm.	397	B
Barton, David	Hist.	323	C
Brown, Annette	Lit.	446	A
Chiu, Sun	Psyc.	405	B
Cortez, Maria	Psyc.	461	A

- Who are the Individuals? *The students*
- What are the Variables? *Major, Points, Grade*

Types of Data

Qualitative (Categorical) Data

- Non-Numerical data with different categories.
- Ex) States, letter grades, class standing, etc.

Quantitative Data

- Numerical data, counts or measurements
- Arithmetic operations such as adding and averaging make sense
- Ex) Income, GPA, Height, Weight, etc.

Example

Researchers spent lots of time and money weighing the stuff put out for recycling in two neighborhoods.

Who are the Individuals?

Households (recycling pickup is done for residences, not for people, one at a time)

What is the Variable? What type of data is it?

Weight in pounds of recycling

Weight is a numerical (quantitative) variable.

LCQ: Data Types

On Tuesday, April 19, 2016, in the Boston Globe, a report on the 2016 Boston Marathon listed each runner's gender, country, age, and time.

a) Who are the individuals for this study?

b) Identify the variables as categorical or quantitative, and include units (if applicable).

LCQ: Data Types

On Tuesday, April 19, 2016, in the Boston Globe, a report on the 2016 Boston Marathon listed each runner's gender, country, age, and time.

a) Who are the individuals for this study? *Boston Marathon runners*

b) Identify the variables as categorical or quantitative, and include units (if applicable).

- *Gender: categorical*
- *Country: categorical*
- *Age: quantitative (years)*
- *Time (hours, minutes, seconds): quantitative*

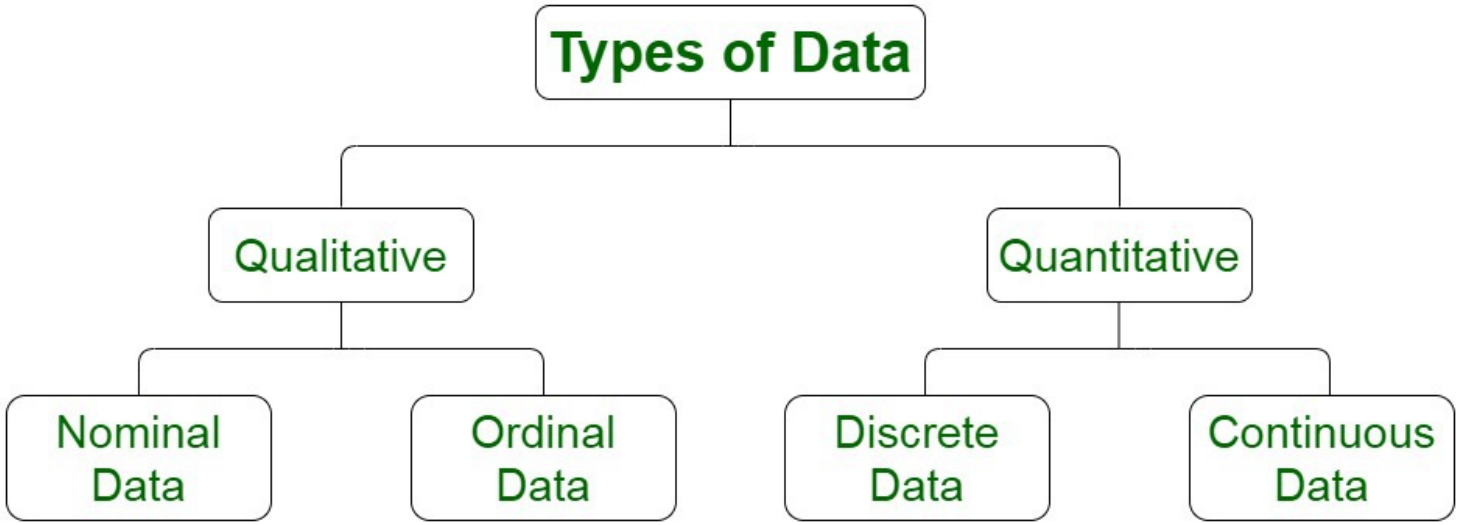
More Types of Data

Random Variable (RV)

- Outcomes of random experiments
- Interested in studying these, so we have to classify them as best as possible

Further Classification

- Each data type (Qualitative and Quantitative) can further be divided into two distinct types
- Important for determining what we can do with the data, how to describe, how to display



https://www.geeksforgeeks.org/explain-different-types-of-data-in-statistics/

Qualitative Data Classification

The two below deal with **Qualitative Random Variables**.

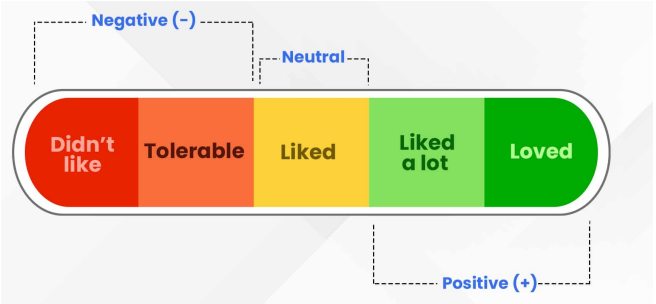
Nominal Random Variable

- Just categories, **no order**
- Ex) Favorite color, yellow isn't greater than blue
- Lowest level of information

Ordinal Random Variable

- Still categories, but there is an **inherent (natural) order**
- Think rankings: 1st, 2nd, 3rd, …

Another example of an ordinal scale: Happiness - 1, 2, 3, 4, 5



https://www.voxco.com/blog/ordinal-data/

Quantitative Data Classification

These two deal with **Quantitative Random Variables.**

Discrete Random Variables

- Have a finite number of possible values.
- When thinking about Discrete RVs, think of numbers that are countable.
- Examples
  - Number of doors on a car
  - Goals scored in a soccer season (24, 15, 3, etc)

Continuous Random Variables

- Can take any number that is in a continuous interval.
- When thinking about Continuous RVs, think of numbers that are measurable.
- Examples
  - Time Spent on Quiz
  - Temperature Outside
  - Football Player Weight

https://prepnuggets.com/glossary/continuous-random-variable/

LCQ: Data Classification

- Classify the following variables as either Qualitative or Quantitative.
  - If Quantitative, is it Discrete or Continuous?
  - If Qualitative, is it Nominal or Ordinal?

a) Time you woke up this morning:

b) Types of fruit:

c) Points scored in a single basketball game:

d) Age:

e) Class standing (freshman, sophomore, …):

f) Average points per game in basketball for a whole season:

LCQ: Data Classification

- Classify the following variables as either Qualitative or Quantitative.
  - If Quantitative, is it Discrete or Continuous?
  - If Qualitative, is it Nominal or Ordinal?

a) Time you woke up this morning: *Quantitative, continuous*

b) Types of fruit: *Qualitative, nominal*

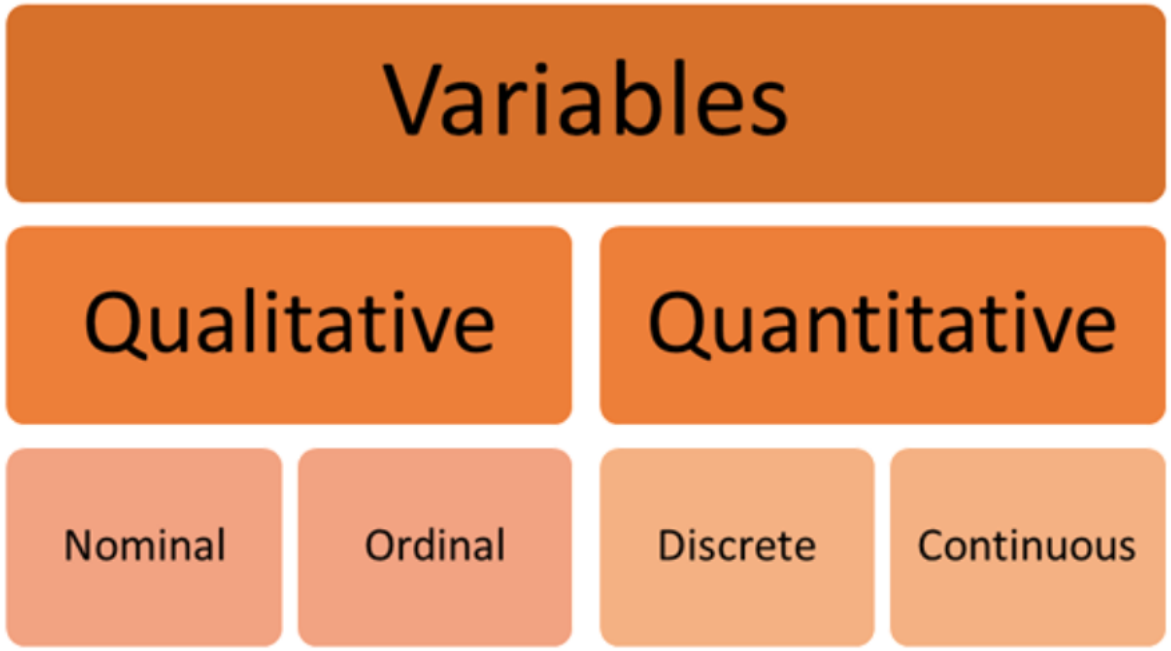
c) Points scored in a single basketball game: *Quantitative, discrete*

d) Age: *Quantitative, continuous (but often reported as discrete)*

e) Class standing (freshman, sophomore, …): *Qualitative, ordinal*

f) Average points per game in basketball for a whole season: *Quantitative, continuous*

Summary of Data Types



Another LCQ: Data Classification

- Classify the following variables as either Qualitative or Quantitative.
  - If Quantitative, is it Discrete or Continuous?
  - If Qualitative, is is Nominal or Ordinal?

a) Quality of a museum tour on scale 1 (poor) to 5 (great):
b) Population size of turtles in a country:
c) Temperature (Fahrenheit):
d) Car manufacturer (Toyato, Honda, etc.):
e) Starting time of class:
f) Cooking times of different meats:

Another LCQ: Data Classification

- Classify the following variables as either Qualitative or Quantitative.
  - If Quantitative, is it Discrete or Continuous?
  - If Qualitative, is is Nominal or Ordinal?

a) Quality of a museum tour on scale 1 (poor) to 5 (great): *Qualitative, ordinal*
b) Population size of turtles in a country: *Quantitative, discrete*
c) Temperature (Fahrenheit): *Quantitative, continuous*
d) Car manufacturer (Toyato, Honda, etc.): *Qualitative, nominal*
e) Starting time of class: *Quantitative, continuous*
f) Cooking times of different meats: *Quantitative, continuous*

PROBLEM SESSION!!!!!!!!!!

Problem #9

An automobile manufacturer wants to know what college students think about electric vehicles. They ask you to conduct a survey that asks students, “Do you think there will be more electric or gasoline powered vehicles on the road in 2025?” and “How likely are you to buy an electric vehicle in the next 10 years?” (scale of 1 = not at all likely to 5 = very likely).

a) Who are the individuals?
b) Name the variables & identify them as quantitative or categorical
c) Did the data come from a survey or experiment? And why?
d) Do you have any concerns about the data?

#9 Solution

a) Individuals - College students
b) There are 2 variables: Response to “Do you think….?” and Response to “Whether they’d purchase an electric vehicle”. Both are categorical.
c) Survey, just asking questions (observational) and NOT imposing a treatment
d) If the manufacture is trying to gauge market interest for electric vehicles, perhaps the audience of the survey isn’t the best. College students might not be able to afford electric vehicles soon. So maybe their responses aren’t the best indication for the manufacturer.

Problem #3

For each variable, would you describe it as primarily categorical, or quantitative? If quantitative, what are the units? If categorical, it is ordinal or simply nominal?

House_ID	Neighborhood	Mail_ZIP	Acres	Yr_Built	Full_Market_Value	Size
413400536	Greenfield Manor	12859	1.00	1967	$1,00,400	960
4128001474	Fort Amherst	12801	0.09	1961	$1,32,500	906
412800344	Dublin	12309	1.65	1993	$1,40,000	1620
4128001552	Granite Springs	10598	0.33	1969	$67,100	900
412800352	Arcady	10562	2.29	1955	$1,90,000	1224
413400322	Ormsbee	12859	9.13	1997	$1,26,900	1056

#3 solution

a) Describing variables
   - House_ID: categorical, nominal
   - Neighborhood: categorical, nominal
   - Mail_ZIP: categorical, nominal
   - Acres: quantitative, acres
   - Yr_Built: quantitative, years
   - Full_Market_Value: quantitative, dollars
   - Size: quantitative, square feet

Problem #1

Here is a dataset about Diamonds. Classify each variable as Categorical or Quantitative

- If Categorical, is it nominal or ordinal?
- If Quantitative, discrete or continuous?

Carat	Cut	For Sale	Number of Bids	Asking Price
0.23	Ideal	Yes	10	123
0.4	Premium	Yes	3	90
0.12	Good	No	5	400
0.32	Good	No	12	389
0.4	Very Good	Yes	20	500
0.03	Ideal	No	5	89

Problem #1 Solution

Here is a dataset about Diamonds. Classify each variable as Categorical or Quantitative

- If Categorical, is it nominal or ordinal?
- If Quantitative, discrete or continuous? Interval or ratio?

Carat	Cut	For Sale	Number of Bids	Asking Price
0.23	Ideal	Yes	10	123
0.4	Premium	Yes	3	90
0.12	Good	No	5	400
0.32	Good	No	12	389
0.4	Very Good	Yes	20	500
0.03	Ideal	No	5	89

Carat: Quantitative, continuous
Cut: Categorical, ordinal
For Sale: Categorical, nominal
Number of bids: Quantitative, discrete
Asking Price: Quantitative, continuous