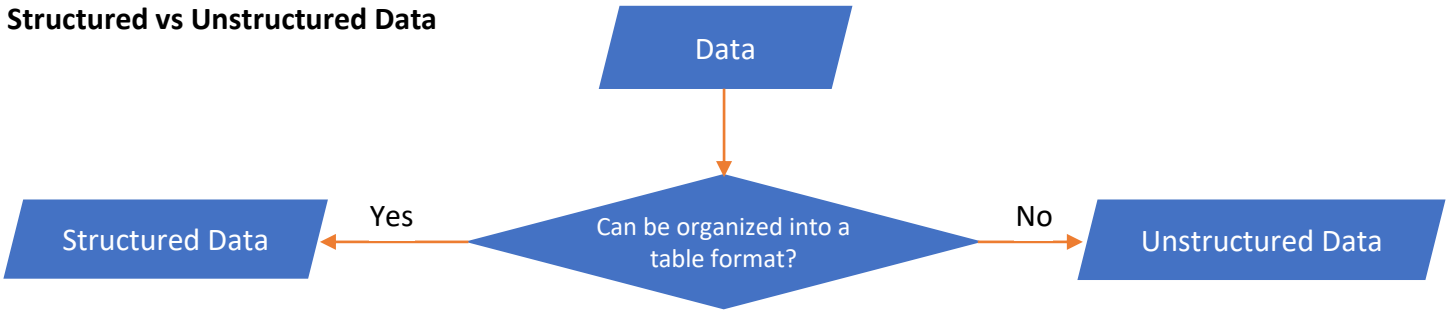
STAT 107 – Fall 2020

[Unit 1] Data & Variables

Ha Khanh Nguyen (hknguyen)

Structured vs Unstructured Data

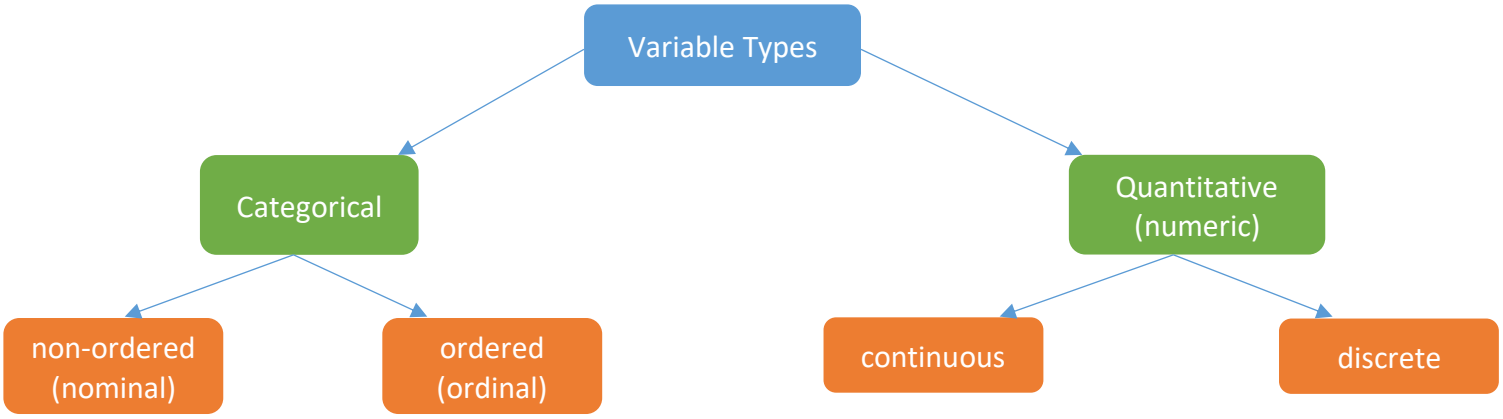


	C	D	E	F	G
1	TRANSACTION DETAILS	VALUE DATE	WITHDRAWAL AMT	DEPOSIT AMT	BALANCE AMT
2	TRF FROM Indiaforensic SERVICES	29-Jun-17		1,000,000.00	1,000,000.00
3	TRF FROM Indiaforensic SERVICES	5-Jul-17		1,000,000.00	2,000,000.00
4	FDRL/INTERNAL FUND TRANSFE	18-Jul-17		500,000.00	2,500,000.00
5	TRF FRM Indiaforensic SERVICES	1-Aug-17		3,000,000.00	5,500,000.00
6	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	6,000,000.00
7	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	6,500,000.00
8	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	7,000,000.00
9	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	7,500,000.00
10	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	8,000,000.00
11	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	8,500,000.00

Source: Bank Transaction Data on Kaggle (https://www.kaggle.com/apoorvwatsky/bank-transaction-data)

Source: chibird.com

Different Types of Variables



Practice: Identify the type of data

	C	D	E	F	G
1	TRANSACTION DETAILS	VALUE DATE	WITHDRAWAL AMT	DEPOSIT AMT	BALANCE AMT
2	TRF FROM Indiaforensic SERVICES	29-Jun-17		1,000,000.00	1,000,000.00
3	TRF FROM Indiaforensic SERVICES	5-Jul-17		1,000,000.00	2,000,000.00
4	FDRL/INTERNAL FUND TRANSFE	18-Jul-17		500,000.00	2,500,000.00
5	TRF FRM Indiaforensic SERVICES	1-Aug-17		3,000,000.00	5,500,000.00
6	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	6,000,000.00
7	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	6,500,000.00
8	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	7,000,000.00
9	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	7,500,000.00
10	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	8,000,000.00
11	FDRL/INTERNAL FUND TRANSFE	16-Aug-17		500,000.00	8,500,000.00

```
{
 "student": [
   {
     "id": 12345,
     "name": "Harry Potter",
     "age": 11
   },
   {
     "id": 12346,
     "name": "Ron Weasley",
     "age": 11
   }
 ]
}
```

Practice: Identify the type of the variable of interest.

20 runners run a mile as fast as they can. Their times are recorded.

Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

50 Students are asked what their major is.

Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

100 Married Couples are asked how many children they have.

Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

20 runners are asked to run a mile as fast as they can. Next to each runner's name, the coach records "yes" or "no" to indicate whether or not they broke the 5-minute mark.

Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

Judges score musicians across a number of different criteria using four choices: "superior," "excellent," "good," or "needs work."

Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

To-do:

- Finish Lab 1, commit and push the lab using git commands!
- Join Piazza if you haven't done so!
- Log in to PrairieLearn using your NetID login and add "STAT/CS/IS 107" to your course list.