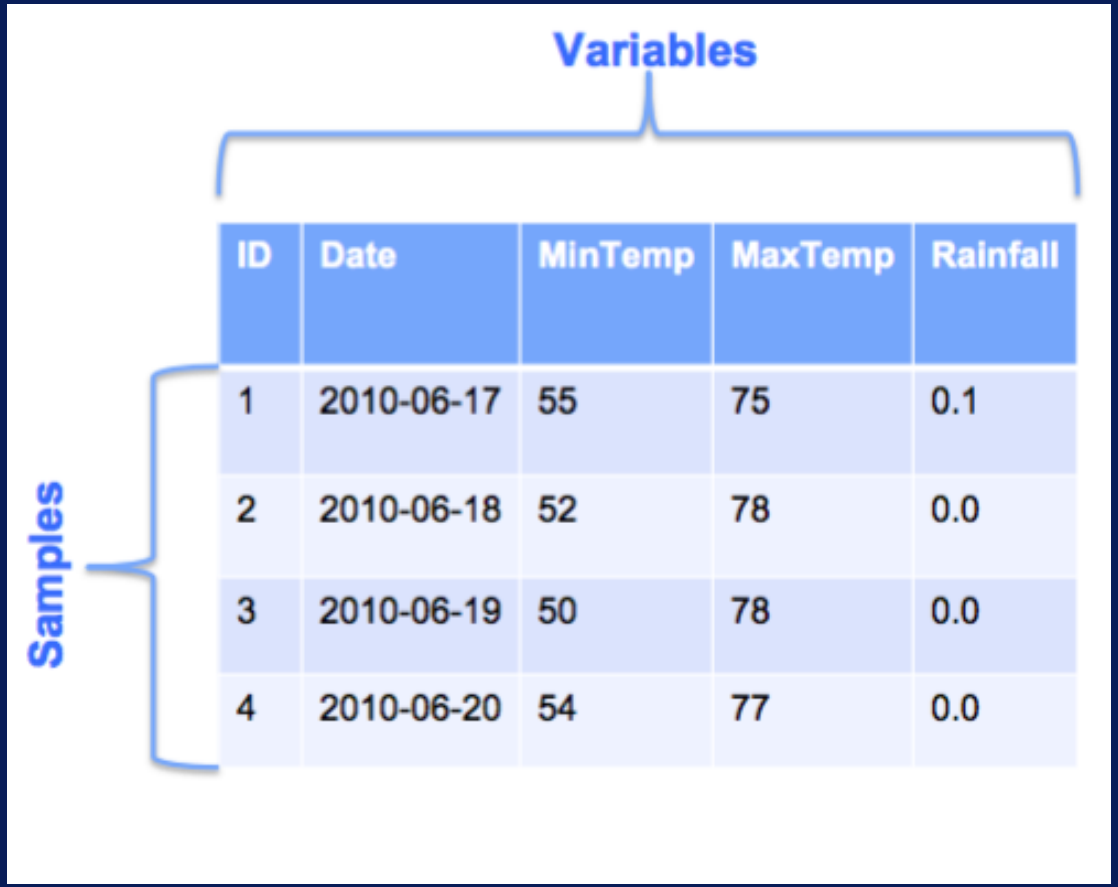
Data Terminology

After this video you will be able to..

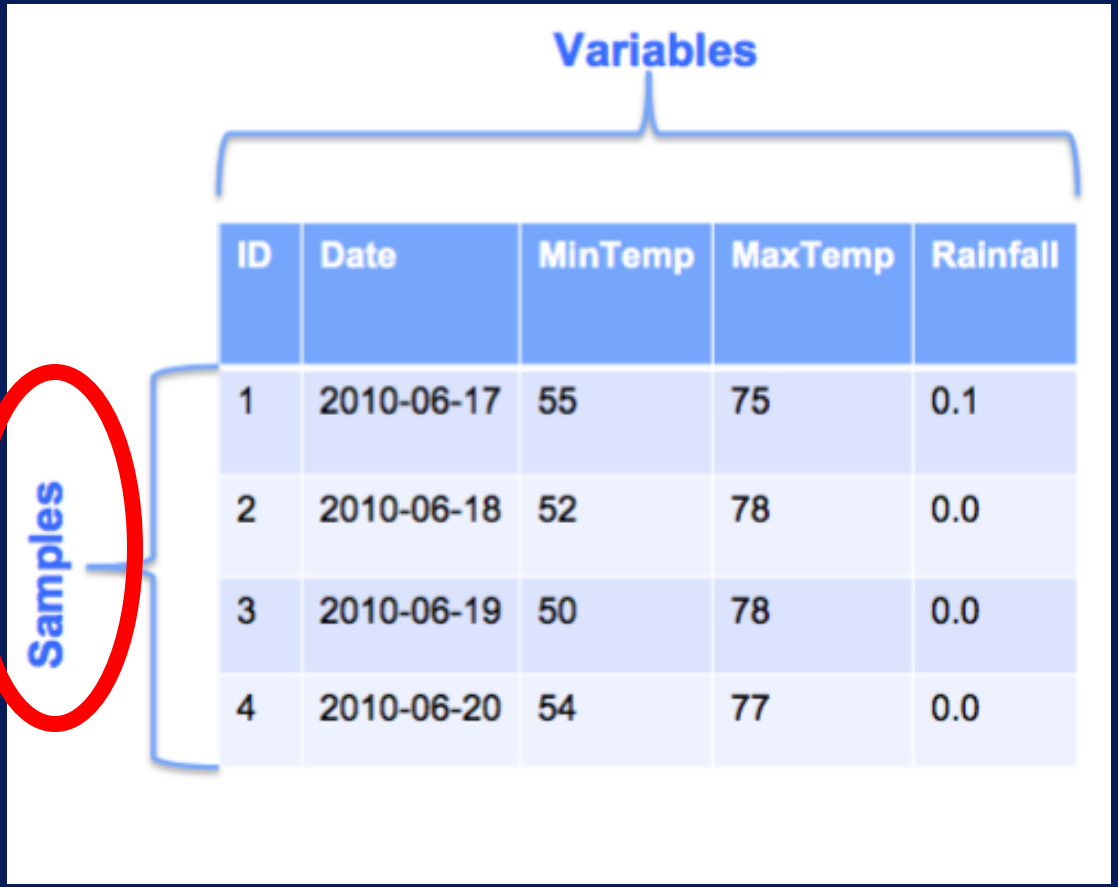
- Describe what a feature is and how it relates to a sample
- Name some alternative terms for ‘feature’
- Summarize how a categorical feature differs from a numerical feature

Terms to Describe Data



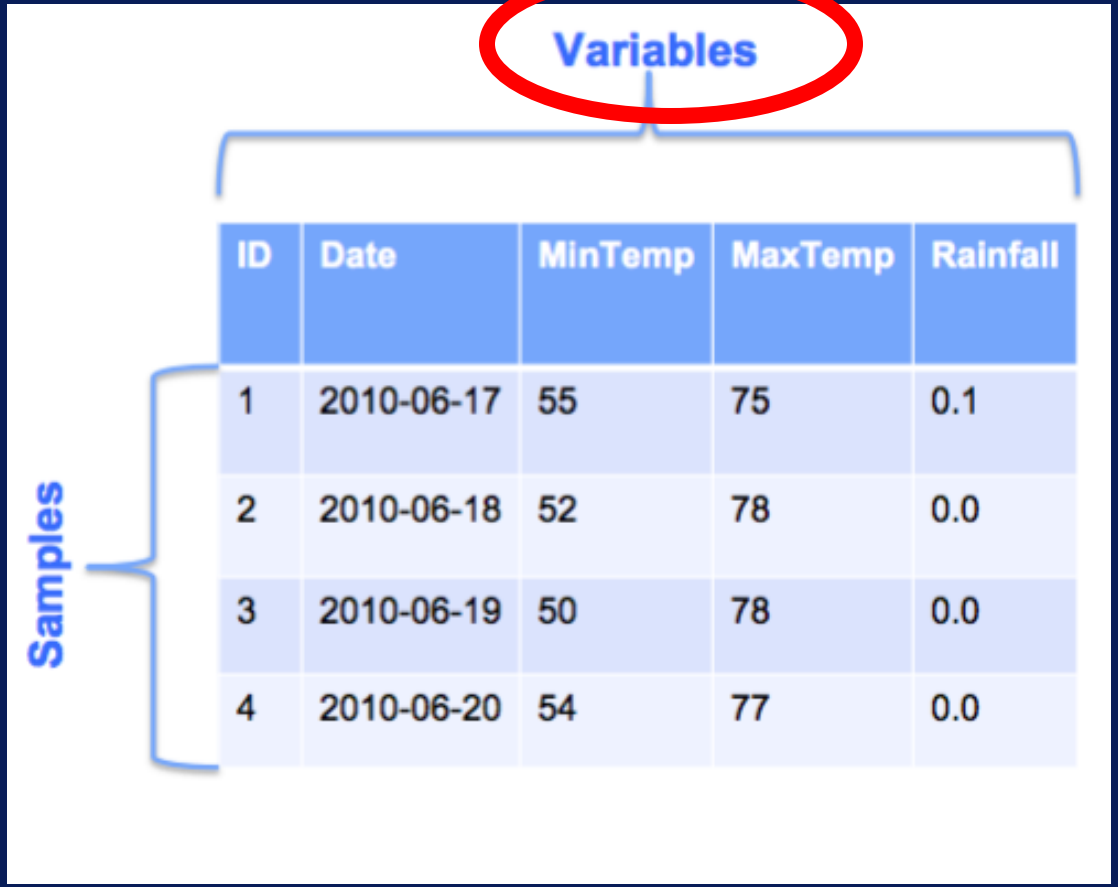
ID	Date	MinTemp	MaxTemp	Rainfall
1	2010-06-17	55	75	0.1
2	2010-06-18	52	78	0.0
3	2010-06-19	50	78	0.0
4	2010-06-20	54	77	0.0

Terms to Describe Data



ID	Date	MinTemp	MaxTemp	Rainfall
1	2010-06-17	55	75	0.1
2	2010-06-18	52	78	0.0
3	2010-06-19	50	78	0.0
4	2010-06-20	54	77	0.0

Terms to Describe Data



ID	Date	MinTemp	MaxTemp	Rainfall
1	2010-06-17	55	75	0.1
2	2010-06-18	52	78	0.0
3	2010-06-19	50	78	0.0
4	2010-06-20	54	77	0.0

Other Names for ‘Sample’

sample

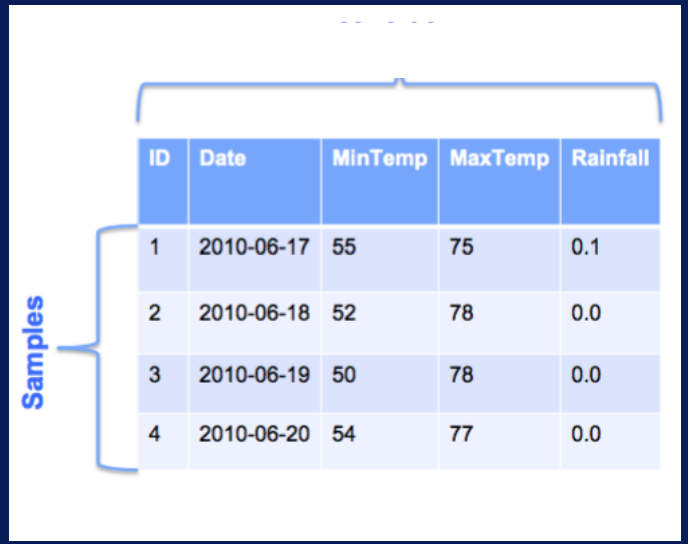
row

instance

observation

record

example



Other Names for ‘Variable’

variable

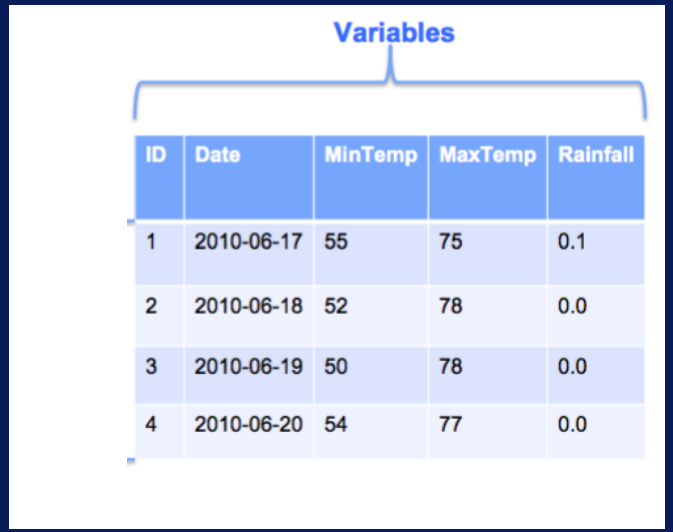
feature

dimension

column

attribute

field



Data Types

- Most common

Numeric **Categorical**

- Others

String **Date** **…**

Numeric Variables

- Values are numbers
- Also called ‘quantitative’

1

$7x10^5$

-0.4902

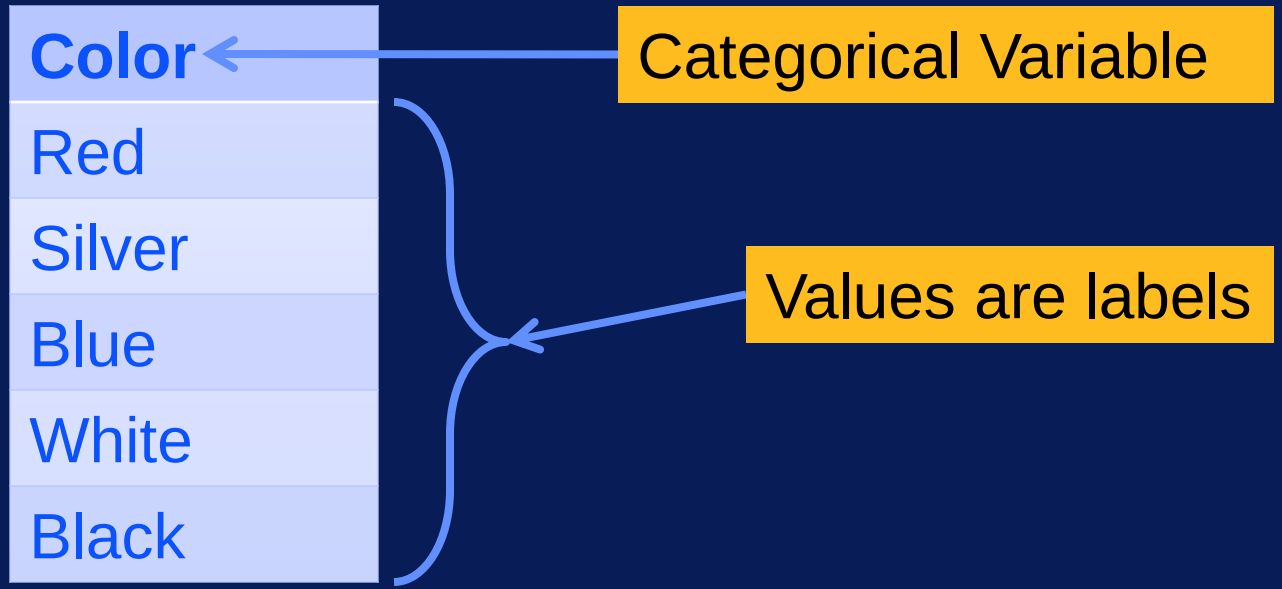
163.92

Examples of Numeric Variables

- Height
- Score on an exam
- Number of transactions per hour
- Change in stock price

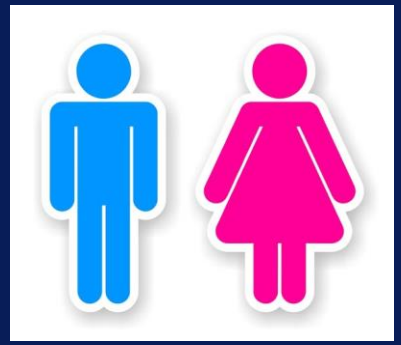
Categorical Variables

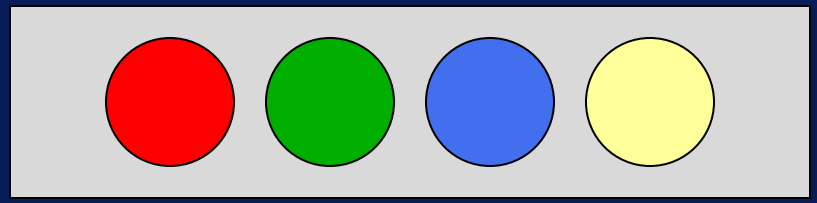
- Values are labels, names, or categories
- Also called ‘qualitative’ or ‘nominal’



Examples of Categorical Variables

- Gender
- Marital status
- Type of customer
- Product categories
- Color of an item



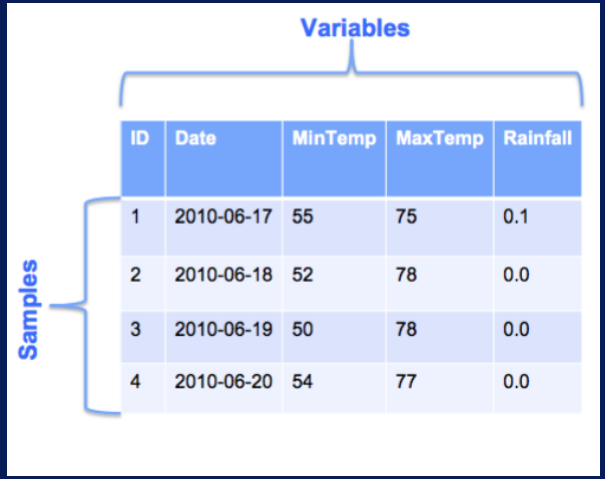


Sample

- Instance
- Record
- Row
- Observation
- …

Variable

- Feature
- Field
- Column
- …



Categorical

Qualitative
Nominal

Numeric

Quantitative