

# 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

The diagram shows a table with 5 columns and 5 rows. The first row is the header, and the following four rows are data. A bracket labeled 'Variables' spans the top of the columns. A bracket labeled 'Samples' spans the left of the rows.

Variables				
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

Samples

# Terms to Describe Data

Variables

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

Samples

# Terms to Describe Data

**Variables**

**Samples**

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

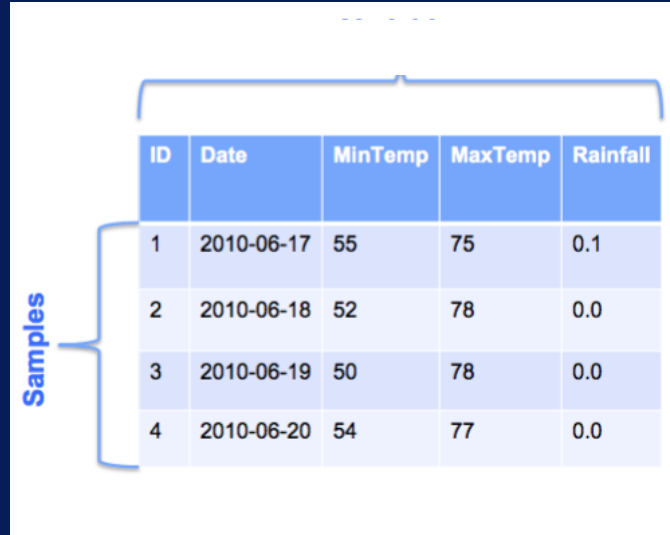
row

instance

observation

record

example



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 'Variable'

variable

feature


dimension

column

attribute

field

Variables



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

# Data Types

- Most common

**Numeric**

**Categorical**

- Others

**String**

**Date**

**...**



# Numeric Variables

- Values are numbers
- Also called 'quantitative'

1

$7 \times 10^5$

163.92

-0.4902

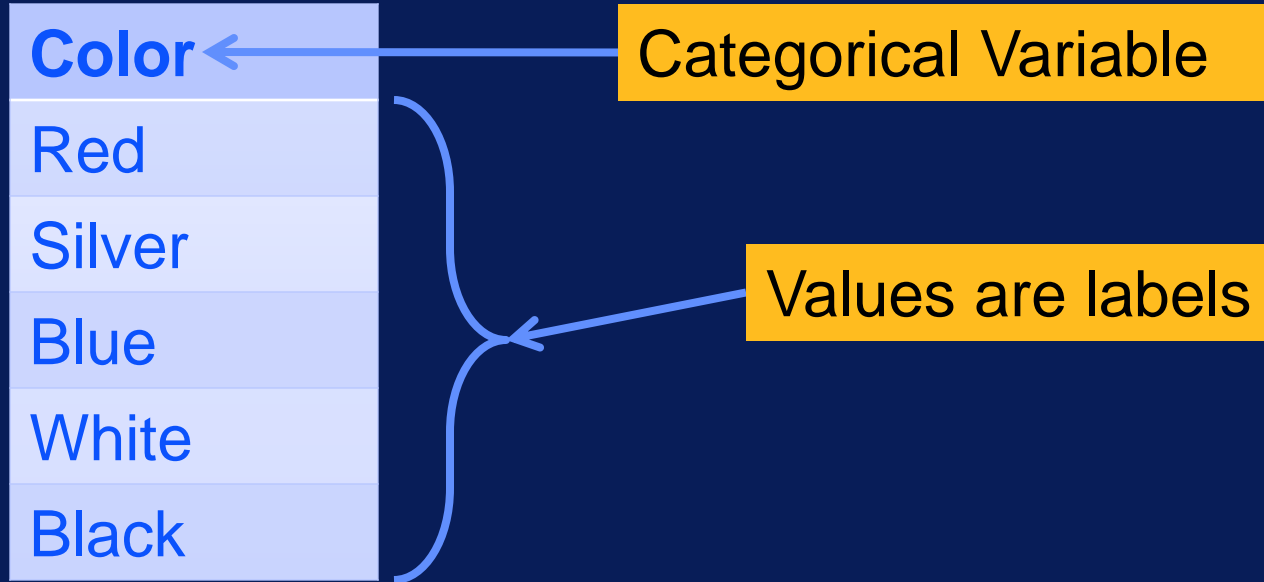
# Examples of Numeric Variables

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



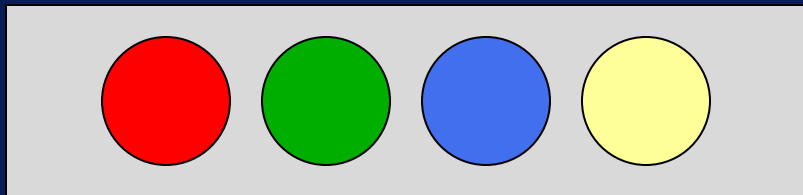
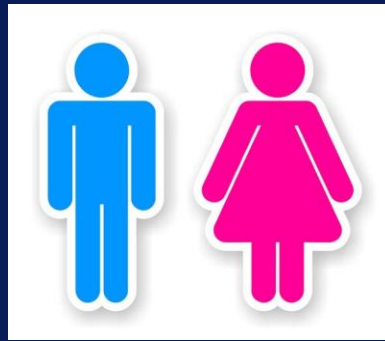
# Categorical Variables

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



# Examples of Categorical Variables

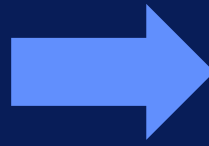
- Gender
- Marital status
- Type of customer
- Product categories
- Color of an item



**Sample**

- Instance
- Record
- Row
- Observation
- ...

**Variable**



- Feature
- Field
- Column
- ...

**Categorical**  
*Qualitative*  
*Nominal*

**Numeric**  
*Quantitative*

Variables				
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

Samples