

MATH 201: Lecture 1b Handout

Section 1.2 Data Basics

Name: _____ Date: _____

Let's take a look at data on our solar system. It is organized into a **data matrix**, where:

- each row is called a _____.
- each column is called a _____.

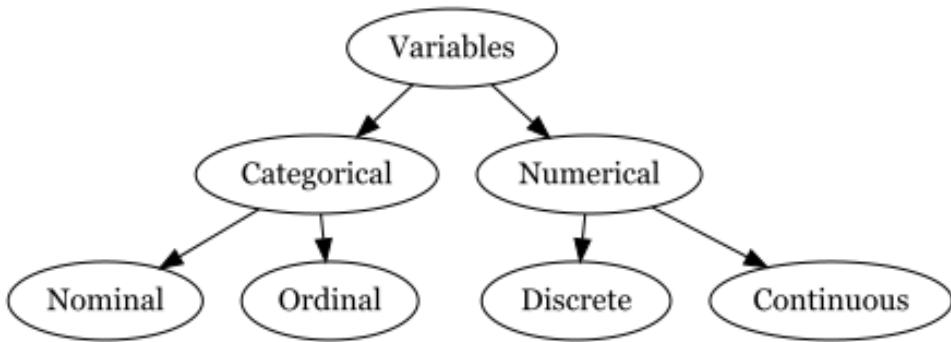
name	mass	length_of_day	mean_temp	n_moons	ring_system	surface_pressure
Mercury	0.330	4222.6	positive	0	FALSE	0.00
Venus	4.870	2802.0	positive	0	FALSE	92.00
Earth	5.970	24.0	positive	1	FALSE	1.00
Mars	0.642	24.7	negative	2	FALSE	0.01
Jupiter	1898.000	9.9	negative	95	TRUE	NA
Saturn	568.000	10.7	negative	146	TRUE	NA
Uranus	86.800	17.2	negative	28	TRUE	NA
Neptune	102.000	16.1	negative	16	TRUE	NA

It is always good to look for a description of your data to better understand how the data was collected and on whom/what.

variable	description
name	Name of planet
mass	Mass in 10^{24} kg
length_of_day	Length of day in hours
mean_temp	Whether mean temperature in Celsius is positive or not
n_moons	Number of moons
ring_system	Whether the planet has a set of rings around it
surface_pressure	Surface pressure in bars

Describe some patterns or features you notice about this data.

Variable types



A _____ variable (sometimes called quantitative) has numeric meaning and we can do mathematical operations with it.

- A _____ variable is a type of _____ variable with a finite or countable number of possibilities.
- A _____ variable is a type of _____ variable with an infinite number of possibilities.

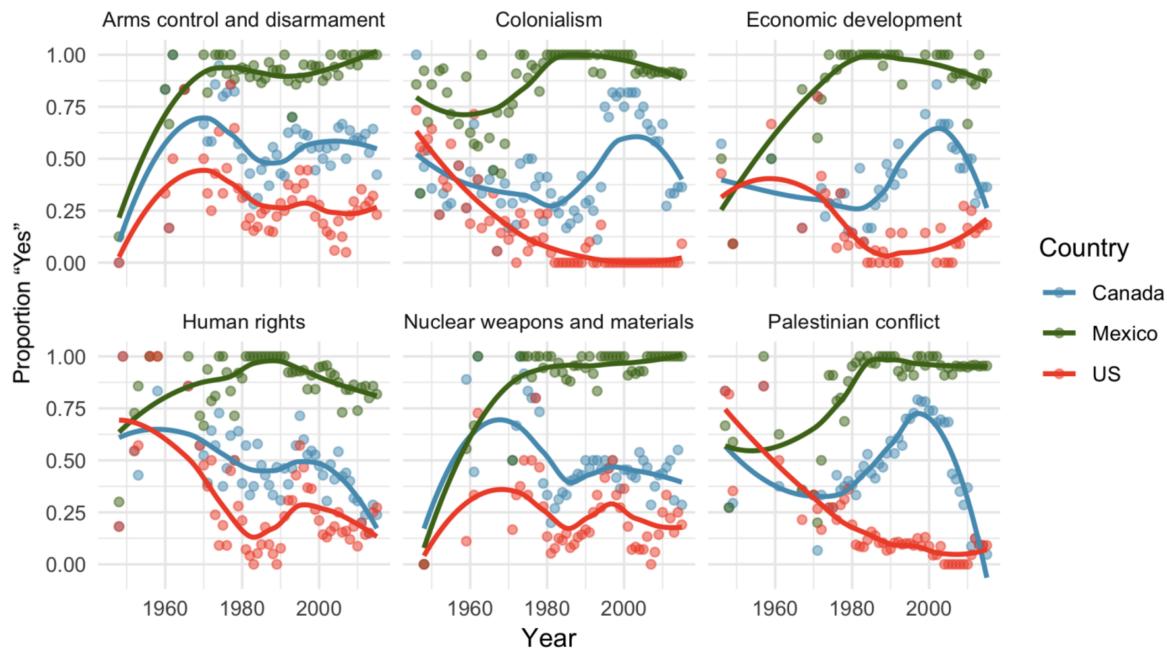
A _____ variable (sometimes called qualitative) does not have numeric meaning, even if it is recorded as a number, and we cannot do mathematical operations with it.

- A _____ variable is a type of _____ variable with a natural ordering to its possible values.
- A _____ variable is a type of _____ variable withouts a natural ordering to its possible values.

For each variable in the `planets` dataset determine its type.

OpenIntro Exercise 1.12 UN Votes.

The visualization below shows voting patterns in the United States, Canada, and Mexico in the United Nations General Assembly on a variety of issues. Specifically, for a given year between 1946 and 2015, it displays the percentage of roll calls in which the country voted yes for each issue.



- (a) Try to draft the data matrix used to produce this visualization, specifically write out the column names and fill in at least four rows.

- (b) Indicate each variable type.

Relationships between variables

If two variables

- have a correspondence we say they are **associated**.
- do not have a correspondence we say they are **not associated**.

Can you think of an unrecorded variable that is likely associated with the mass of a planet or having a ring system?