POLI 30 D: Political Inquiry

Professor Umberto Mignozzetti
(Based on DSS Materials)

Lecture 02 | Observations and Variables

Plan for Today

- What is a theory?
  - definition of theory?
  - political science theories?

- What are data/datasets?
  - what is an observation?
  - what is a variable?

- Types of variables based on content
  - character vs. numeric variables
  - binary vs. non-binary variables

Before we start

Announcements:

- Quizzes and Participation:
  - I will start these at week 03.
- We **do have** Lab this week:
  - Bring your laptop to install and configure R and R Studio
  - Meet and greet your TAs. They are here to help!
- Github page: https://github.com/umbertomig/POLI30Dpublic
- Podcast: I didn't know I need to wear a mic. Will do from now on.

Political Science Theories

What is a theory?

Political science goals:

- Describe and understand political phenomena.
- Identify regularities and patterns in political behavior.
- Build theories that enable us to explain, predict, and sometimes affect political outcomes.
- But what is a theory?

What is a theory?

One definition:

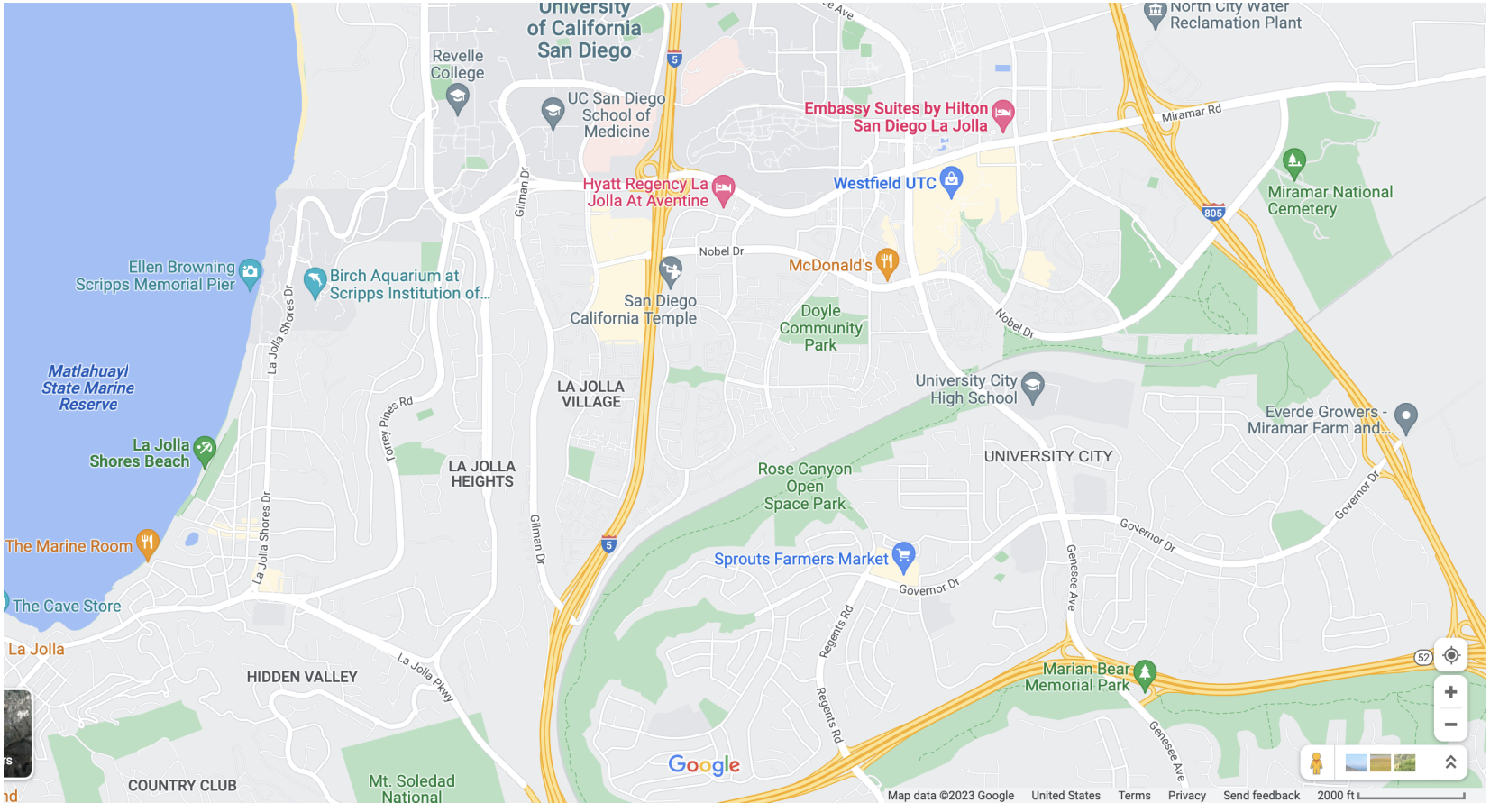
> *A set of statements or principles devised to explain a group of facts or phenomena, especially one that has been repeatedly tested or is widely accepted and can be used to make predictions about natural phenomena.*

Another definition:

> *A general statement about how the world may work. A theory specifies a causal mechanism. We assume that there are underlying laws that we seek to reveal.*

What is a theory?

My preferred definition:



What is a theory?

- Theories don't have to be (completely) right to be helpful.
- Theories are continuously being assessed, updated, and even replaced with new evidence.
- All theories are wrong, but some are useful!
- What do you think is **not** a theory?

Example: What is the Sun?

- One theory: It's a giant chariot being driven by the sun god.
  - The sun god drives across the sky once per day, then takes a break.
- Another theory: It is a star at the center of the Solar system.
  - Giant ball of hot plasma.
- We use the data to see which theories are better.
- But all of them are incomplete by design.

Example PoliSci Theory

- Lyall (2009). "Does Indiscriminate violence Incite Insurgent Attacks? Evidence from Chechnya"?
- Idea: Indiscriminate violence reduces insurgence violence.
  - It creates enormous logistical problems for insurgencies.
  - It may cause the local population to turn against the insurgency.

Anatomy of a simple theory

- Dependent Variable: to be explained

  AKA: "response variable", "regressand", "outcome", "predicted variable", "explained variable", target", or "label"

- Independent Variable: used to explain something

  AKA: "predictor variable", "regressor", "treatment", "manipulated variable", "explanatory variable", "risk factor", "feature", or "input variable"

- Relationship and mechanism
- Units

Lyall (2009) 's Example

- Dependent variable: Insurgence violence
- Independent variable: Indiscriminate violence
- Mechanism:
  - It creates enormous logistical problems for insurgencies.
  - It may cause the local population to turn against the insurgency.
- Units: Insurgence

Corruption Example

Corruption is lower in more developed countries because citizens have more ability to monitor politicians' performance due to increased access to information and education.

- Independent variable?
- Dependent variable?
- Mechanism?
- Unit of analysis?

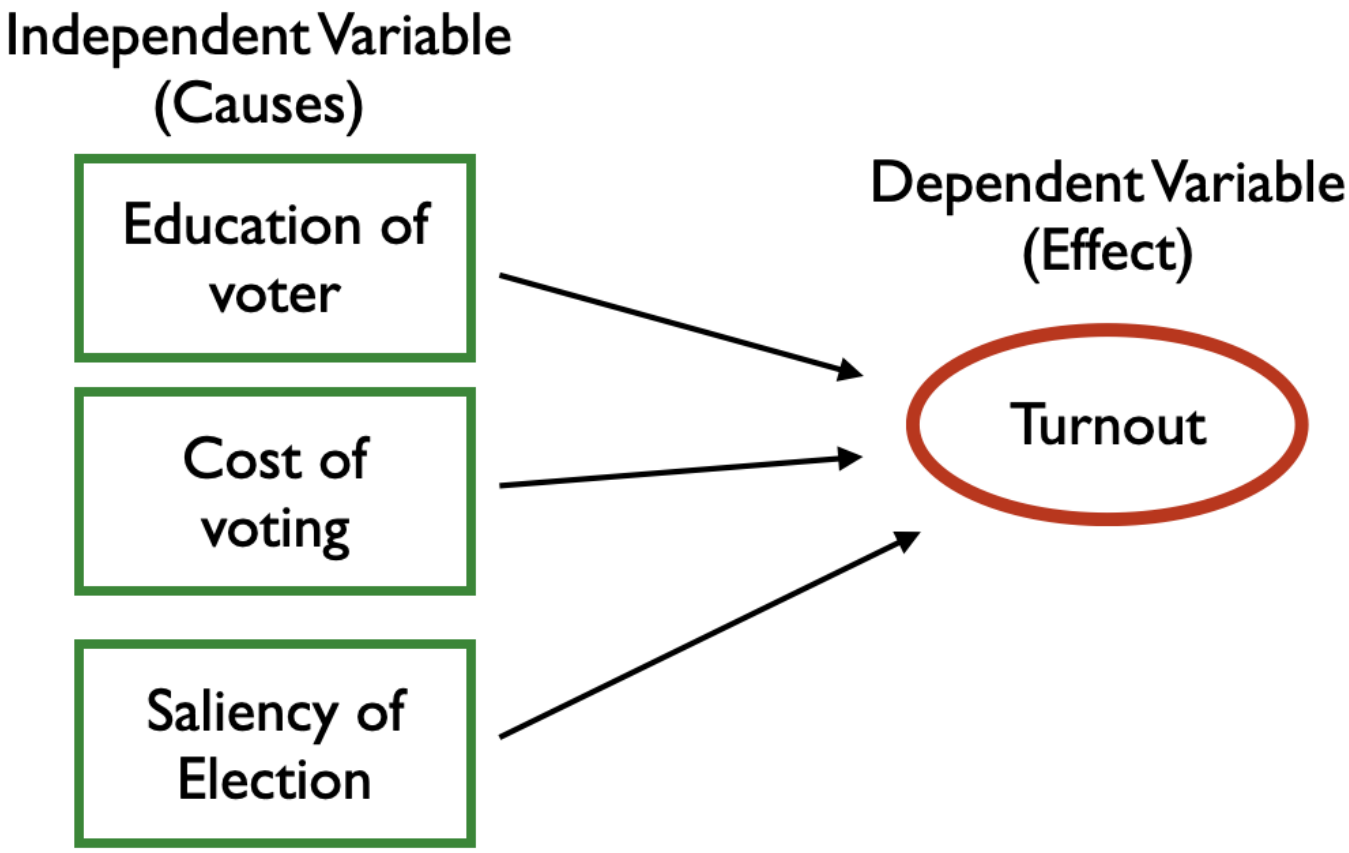
Turnout Example

Citizens with more education are more likely to turn out to vote than those with less because understanding the issues increases the perceived importance of political outcomes.

- Independent variable?
- Dependent variable?
- Mechanism?
- Unit of analysis?

Turnout Example

- Sometimes, a theory has multiple causes:



Good theory

- Falsifiable
- Non-normative
- General
- Explains how and why two or more variables are related (only theory)

Hypotheses

- To test a theory, we should derive hypotheses from it.
- Hypotheses: "...a testable statement about the empirical relationship between an independent and dependent variable."
- Empirical implications of our theory - what we expect to see in data if our theory is correct.
- A good hypothesis:
  - Ties variance in cause to variance in effect
  - Is falsifiable
  - Is general, or "not immediately verifiable''

Example

- **Theory:**

  Citizens that understand the issues at stake in politics are more likely to participate in politics than those that do not because understanding the issues increases the perceived importance of political outcomes.

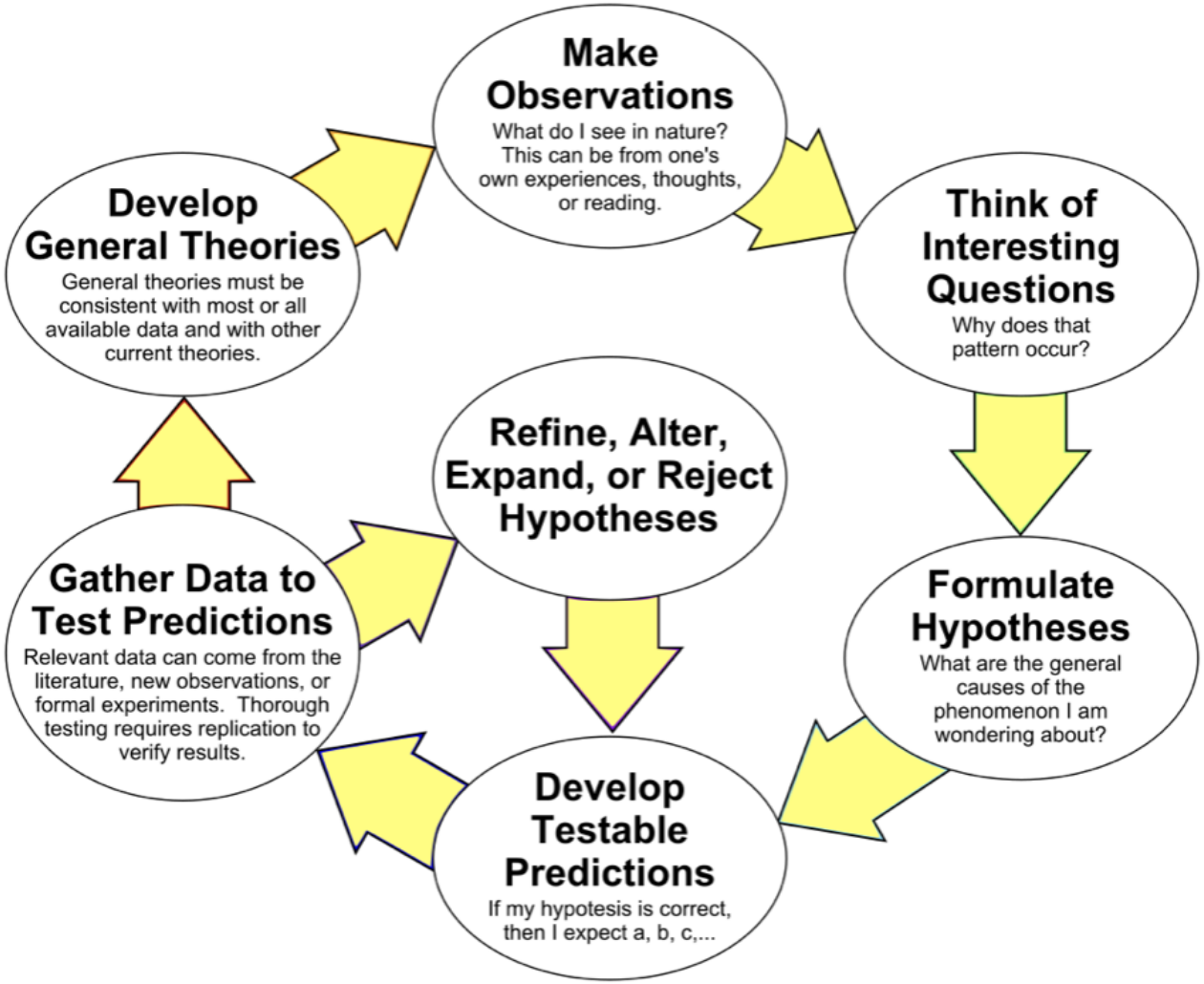
- **Hypothesis:**

  Turnout will be higher among individuals with higher education than those with lower education.

Scientific Method

- A unique way to create and falsify theories. It involves:
  - Create theories
  - Observe
  - Hypothesize
  - Test
  - Repeat many times and eventually revise (improve) the theories we have
  - Not just three steps - a process of knowledge acquisition

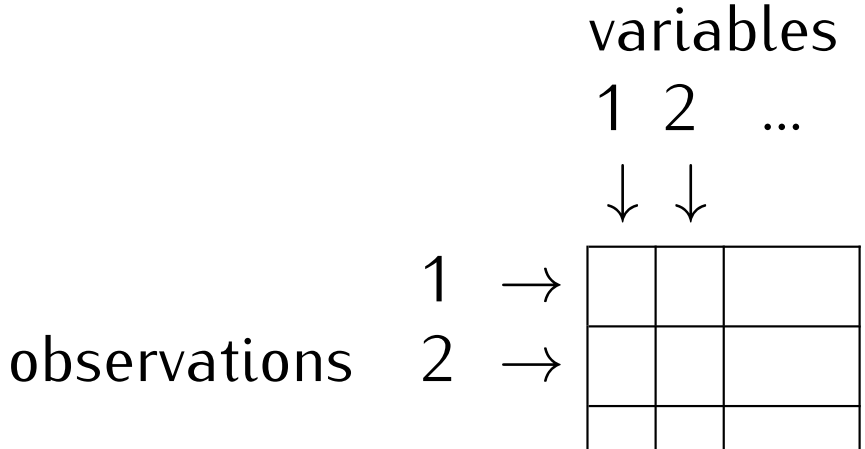
Scientific Cycle



Political Science Data

What are Data/Datasets?

- To test theories, we need data. What is data? What are datasets?
- Datasets capture the characteristics of a particular set of individuals or entities:
  - students, classrooms, schools, etc.
- Datasets are typically organized as **dataframes** where rows are observations and columns are variables



Example of a Dataframe

Each column is a variable

Each row is an observation (i=1,2,3,….10)

i	first_name	test_score
1	ana	80
2	elena	75
3	maria	99
4	juan	67
5	diego	89
6	carlos	80
7	olivia	70
8	jorge	86
9	adolfo	92
10	marta	83

What is an observation?

- It is the information collected from a particular entity or individual in the study
- The **Unit of observation** of the dataset defines the individuals or the entities that each observation in the dataset represents
  - if the Unit of observation is students, each row in the dataset represents a different student
- We usually refer to an observation by the row number in the dataset, which we denote as i
  - what is the first observation (i=1) in the dataframe above?

What is a variable?

- A variable contains the values of a changing characteristic for the various individuals or entities in the study
- Every column of data in a dataset is a variable
  - if the Unit of observation is students, each variable captures a specific characteristic of the students, for all the students in the study
- We usually refer to a variable by its name
  - *first_name*, *test_scores*

Notation

- When defining new variables, we represent a variable and its contents in the following format:

$$X = \{10, 5, 8\}$$

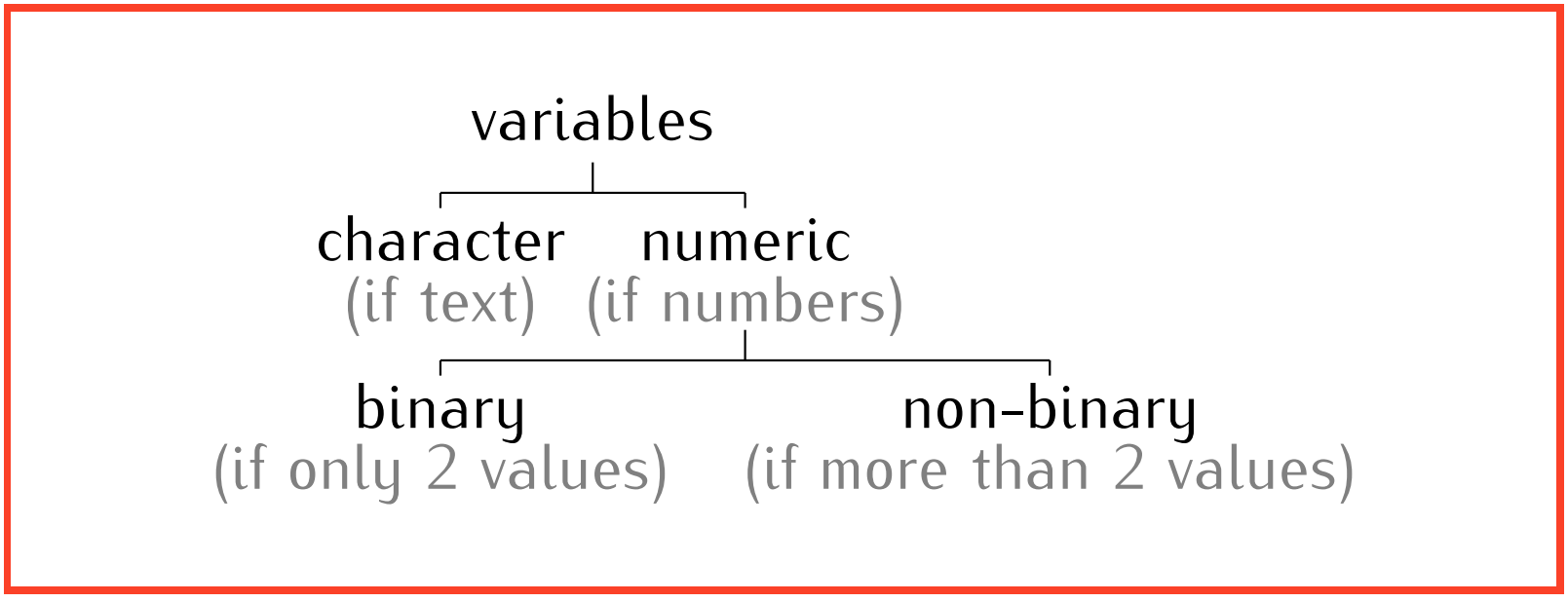
- On the left-hand side of the equal sign, we identify the name of the variable:
  - what is the name of the variable here?
- On the right-hand side of the equal sign and inside curly brackets, we have the content of the variable: multiple observations, separated by commas
  - what are the observations in X?

Notation

$$X = \{10, 5, 8\}$$

- To represent each observation we use X_i
  - where i stands for the observation number
  - the subscript i means that we have a different value of X for each value of i
  - what is X_3?
- The total number of observations is denoted as n
  - what does n equal to here?

Types of Variables Based on Content



Character vs. Numeric

- **Character variables** contain text
  - *first_names*={ana, elena, maria, ...}
- **Numeric variables** contain numbers
  - *test_score*= {80, 75, 99, ...}

Numeric: Binary variables

- **Binary variables** can take only two values: 1s and 0s
- They represent the presence/absence of a trait:
  - 1 if individual i has the trait
  - 0 if individual i does not have the trait
- Example: $voted = \{1, 0, 0, 1, 1, 1, 0\}$ where

$$voted_i = \begin{cases} 1 \text{ if individual } i \text{ voted} \\ 0 \text{ if individual } i \text{ didn't vote} \end{cases}$$

- Can you think of another example?

Numeric: Non-binary variables

- **Non-binary variables** can take more than two values
  - *distance*={1.452, 2.345, 0.298}
  - *dice_roll*={2, 4, 6}
- Can you think of another example?

Summary

- **Today's Class:**
  - Theory and scientific theories
  - Data/datasets
  - Observations and variables
  - Unit of observation
  - Character vs. numeric variables
  - Binary vs. non-binary variables
- Next class:
  - Computing and interpreting means

Questions?

See you in the next class!