

ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

# INTRODUCTION TO DATA

PART 2

UC San Diego

COMPUTER SCIENCE & ENGINEERING  
HALICIOĞLU DATA SCIENCE INSTITUTE

# Explanatory and Response Variables

## Example

- Suppose we have a data set containing observations across many variables in US counties: unemployment rate, population, state, home ownership rates, percent of units in multi-unit buildings, etc.
- Suppose we want to understand the relationship between median household income and population change.
- Specifically, suppose we want to know:
  - If there is an increase in the median household income in a county, does this drive an increase in its population?



# Explanatory and Response Variables


- In this example, *median household income* is the **explanatory variable** and *population change* is the **response variable**.
- If we think/believe/suspect that one variable *might causally affect* another, we call the first one the **explanatory variable** and the second one the **response variable**.



- If there is no hypothesized relationship, these labels would not be applied.

# Observational Studies and Experiments

**There are two ways to collect data:**

1. Observational studies
  2. Experiments
- 
- A decorative teal triangle is located in the bottom right corner of the slide, pointing towards the top right.

# Observational Studies

- Data is collected in a way that does not interfere with how the data comes about.
  - For example: extracting data from surveys, or existing records, or following a cohort of individuals.
- Good for providing evidence of natural occurring association between variables.
- Cannot show a causal relationship between variables.

# Experiments

- Usually experiments are designed to investigate the possibility of a causal connection, and contain an explanatory and response variable
  - For example: You might design an experiment in order to understand the effect of a certain drug (*explanatory variable*) on mortality over the next year (*response variable*)
- Usually, researchers will collect a sample of individuals and split them into groups. The individuals in each group are assigned a treatment.
- When individuals are randomly assigned to each group the experiment is called a **randomized experiment**. In our example you could flip a coin to determine whether an individual receives a placebo, or the actual drug.