Data Basics

DASC 512

Overview

- Defining the Research Question
- Data Matrices
- Types of Variables
- Relationships between Variables

But first, some admin…

Syllabus

Make sure you've read the syllabus and watched the course intro

Defining the Research Question

Data without appropriate context is meaningless.

- Population: The entire group of interest
  - Example: American adults, registered voters, or likely voters
- Observational Unit (Case): A single element of the population set
  - Examples: One person, bomb drop, or weather station
- Variables: Measured characteristics of the observational unit
  - Examples: Poll response, miss distance, or temperature

Data Matrix

We will interact with data in a data matrix

Column = Variable

Row = Case

Case	Var1	Var2	Var3	Var4
1	1.2	2.3	3.4	4.5
2	1.3	2.4	3.5	4.6
3	2.1	3.2	4.3	5.4
4	3.1	4.2	5.3	6.4

Types of Variables

- Numerical / Quantitative
  - Continuous or Discrete
  - Ratio, Interval, or Ordinal
- Categorical / Qualitative
  - Ordinal or Nominal

Ratio and Interval Variables

- Ratio variables have a real zero point
  - Example: Temperature (K)
  - All mathematical operators have meaning

- Interval variables have an arbitrary zero point
  - Example: Temperature (°C)
  - $>$, $<$, $+$, and $-$ have meaning, but not $\times$ or $\div$

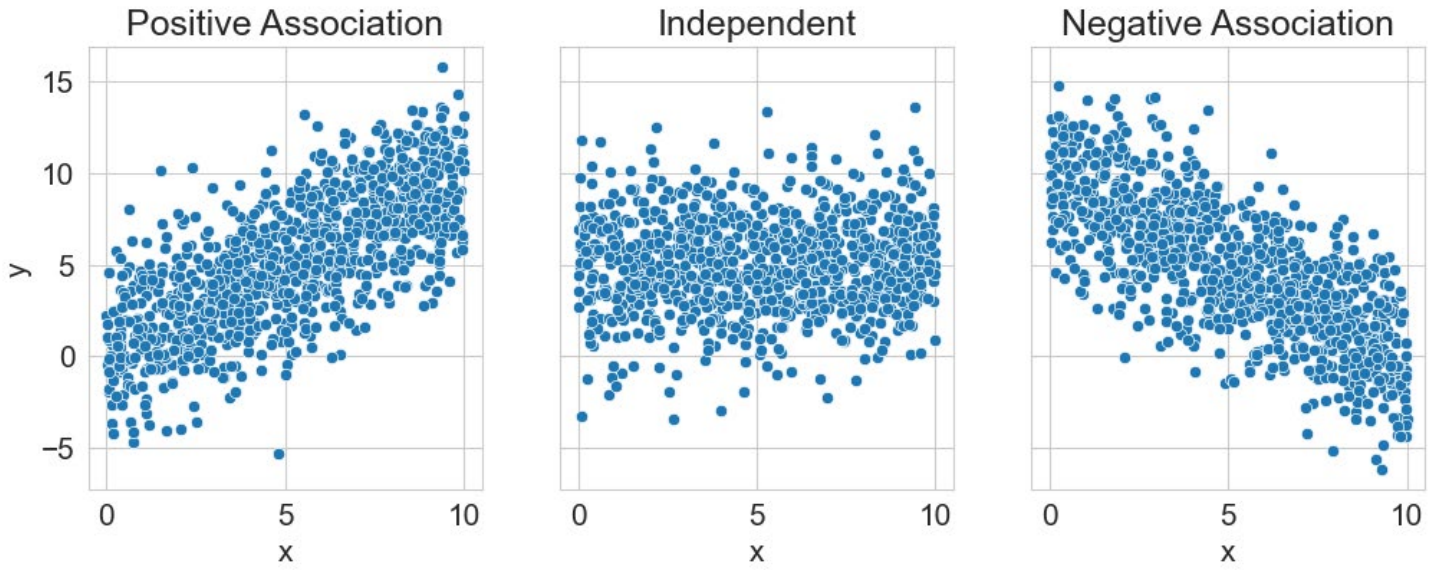
Ordinal and Nominal Variables

- Ordinal variables have some natural ordering
  - Examples: Finishing place, military rank, Likert scale responses
  - $>$ and $<$ have meaning, but not $+$, $-$, $\times$, or $\div$
  - Note: There is research to suggest that well-designed surveys' Likert scale responses can be treated as interval data.

- Nominal variables have no natural ordering
  - Examples: Color, Team, Brand
  - Mathematical operators have no meaning

Relationships between Variables

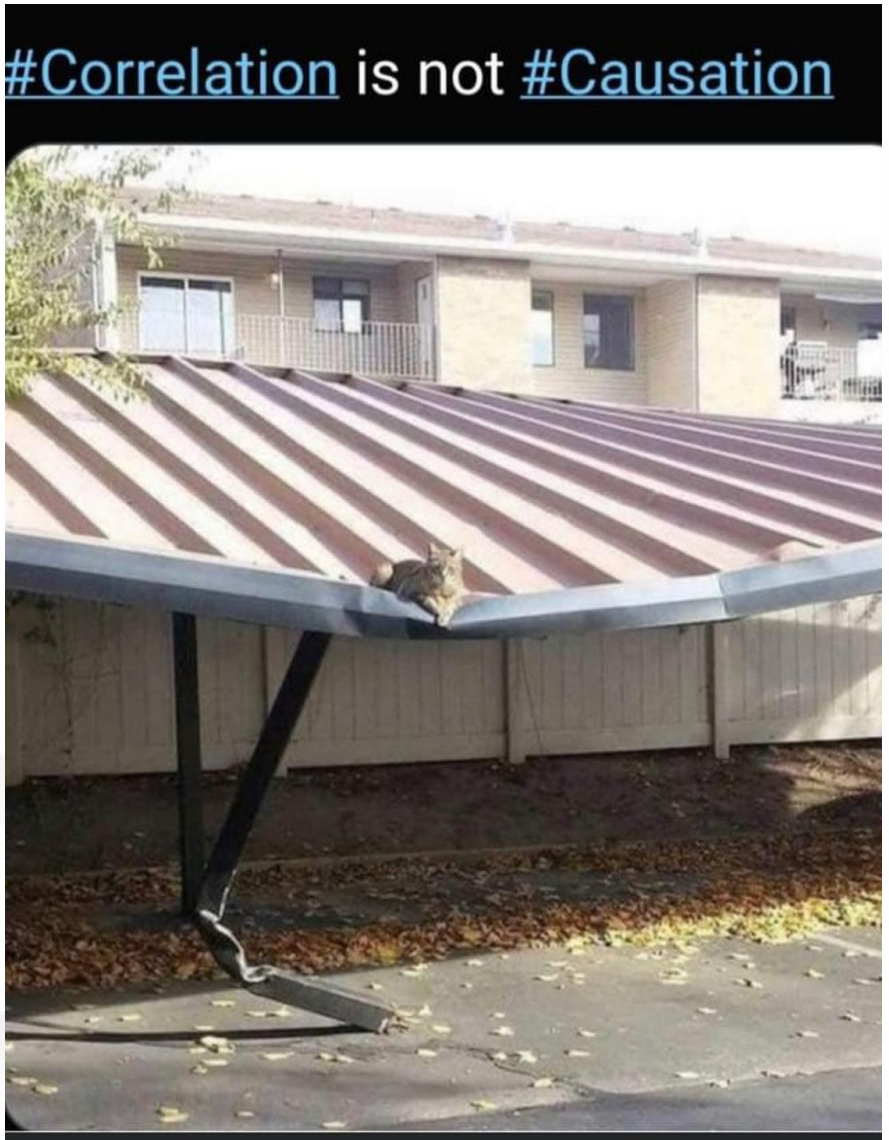
- Positive Association: when one variable increases, the other tends to also increase
- Negative Association: when one variable increases, the other tends to decrease
- Independent: two independent variables have no association



Relationships between Variables

- Independent or explanatory variable: A variable that is **hypothesized** to have some effect upon the dependent or response variable

- **Association does not imply causation!**



Recap

- Defining the Research Question
- Data Matrices
- Types of Variables
- Relationships between Variables

We will see all this again in much more detail later in this course.