

# Defining Data Access

## By Jinseok Kim

# What is Data Access?

- **Data access** refers to the methods, approaches, or tools to retrieve information from various sources.
  - Data access for **designed data**.
  - Data access for **gathered data**.

# Data Access for Designed Data

- The data access in designed data is a sampling frame.
- A **sampling frame** is a set of materials or procedures that are used to select a sample.
  - We prefer frames that give every eligible unit a nonzero probability of being selected (for **probability samples**).
  - A non-probability sampling frame is a frame (but not one under which the sampling is controlled by the sampler).

# Data Access for Gathered Data (1)

- For gathered data, the methods or procedures used for acquiring information of interest that might be stored as:
  - Structured data: rows/columns in a table format.
  - Semi-structured: table like but cells may be lists or text, etc.
  - Unstructured: no rows/columns, like text data.

# Data Access for Gathered Data (2)

- Examples of data access for gathered data:
  - Purchase data from a third party or reseller (e.g., BrandWatch for Twitter or Reddit data).
  - Manually retrieve information (“Cut and Paste”).
  - Scrape webpage contents in HTML using web-scraping tools or code in R or Python.
  - Use an Application Programming Interface (API)

# What's Next?

- In the next lecture, we will take a look at threats concerning data access for gathered data.



**© Faculty Presenter**

**Except where otherwise noted, this  
work is licensed under CC BY-NC 4.0**