

Instructor: Nara Yoon, Ph.D. (yoonnx@jmu.edu)

Spring 2024 | Thursdays 4-7pm, Hart 3008

Data Analytics for Managerial Leadership

WK 6: Data project

RESEARCH WORKFLOW

1. RQ formulation
2. Data collection
3. Data cleaning
4. Data analysis
5. Communication of results

RESEARCH QUESTION FORMULATION

Ask yourself:

- **What is the problem that I want to address in leadership studies?**
- **How does the answer to my RQ contribute to scholarship and/or practice?**

DATA COLLECTION

Ask yourself:

- What would be the *ideal* data structure to answer my RQ?
- Where is my data source?
- What does my data represent?
 - Universe (whole population)
 - Sample (cases)

DIFFERENT WAYS TO COLLECT DATA

- Download from online source
- API package in R (use package to access API)
- API (we need API key – e.g., rtweet, spotify)
- Web scraping

DATA COLLECTION SOURCES (EXAMPLE)

A little stats: Data sources

Google data search engine

U.S. city open data census

U.S. government's open data

Census data API

Data USA

R Twitter package API

DATA COLLECTION SOURCES (EXAMPLE)

[Pew Research Center](#)

[National Center for Charitable Statistics \(NCCS\)](#)

[Propublica](#)

[American Time Use Survey \(ATUS\)](#)

[Kauffman Entrepreneurship](#)

DATA ANALYSIS (EMPIRICAL STRATEGY)

Ask yourself:

- Unit of analysis?
- Am I interested in *explanation* or *prediction*?
(descriptive stats vs. predictive models – implication for generalization)

COMMUNICATION OF RESULTS

Ask yourself:

- **What are the sound conclusion that can be drawn from the findings?**
- **What is the most effective way to communicate findings to a broader audience?**