Data Collection Methods

Pros and Cons of Primary and Secondary Data
Where do data come from?

- We’ve seen our data for this lab, all nice and collated in a database – from:
  - Insurance companies (claims, medications, procedures, diagnoses, etc.)
  - Firms (demographic data, productivity data, etc.)
Where do data come from?

- Take a step back – if we’re starting from scratch, how do we collect / find data?
  - Secondary data
  - Primary data
Secondary Data

- Secondary data – data someone else has collected
  - This is what you were looking for in your assignment.
Secondary Data – Examples of Sources

- County health departments
- Vital Statistics – birth, death certificates
- Hospital, clinic, school nurse records
- Private and foundation databases
- City and county governments
- Surveillance data from state government programs
- Federal agency statistics - Census, NIH, etc.
Secondary Data – Limitations

- What did you find on the frustrating side as you looked for data on the state’s websites?
Secondary Data – Limitations

- When was it collected? For how long?
  - May be out of date for what you want to analyze.
  - May not have been collected long enough for detecting trends.
  - E.g. Have new anticorruption laws impacted Russia’s government accountability ratings?
Secondary Data – Limitations

- Is the data set complete?
  - There may be missing information on some observations
  - Unless such missing information is caught and corrected for, analysis will be biased.
Secondary Data – Limitations

- Are there confounding problems?
  - Sample selection bias?
  - Source choice bias?
  - In time series, did some observations drop out over time?
Secondary Data – Limitations

- Are the data consistent/reliable?
  - Did variables drop out over time?
  - Did variables change in definition over time?
  - E.g. number of years of education versus highest degree obtained.
Secondary Data – Limitations

- Is the information exactly what you need?
  - In some cases, may have to use “proxy variables” – variables that may approximate something you really wanted to measure. Are they reliable? Is there correlation to what you actually want to measure?
  - E.g. gauging student interest in U.W. by their ranking on FAFSA – subject to gamesmanship.
Secondary Data – Advantages

- No need to reinvent the wheel.
  - If someone has already found the data, take advantage of it.
Secondary Data – Advantages

- It will save you money.
  - Even if you have to pay for access, often it is cheaper in terms of money than collecting your own data. (more on this later.)
Secondary Data – Advantages

- It will save you time.
  - Primary data collection is very time consuming. (More on this later, too!)
Secondary Data – Advantages

- It may be very accurate.
  - When especially a government agency has collected the data, incredible amounts of time and money went into it. It’s probably highly accurate.
Secondary Data – Advantages

- It has great exploratory value
  - Exploring research questions and formulating hypothesis to test.
Primary Data

- Primary data – data you collect
Primary Data - Examples

- Surveys
- Focus groups
- Questionnaires
- Personal interviews
- Experiments and observational study
Primary Data - Limitations

- Do you have the time and money for:
  - Designing your collection instrument?
  - Selecting your population or sample?
  - Pretesting/piloting the instrument to work out sources of bias?
  - Administration of the instrument?
  - Entry/collation of data?
Primary Data - Limitations

- Uniqueness
  - May not be able to compare to other populations
Primary Data - Limitations

- Researcher error
  - Sample bias
  - Other confounding factors
Data collection choice

- What you must ask yourself:
  - Will the data answer my research question?
Data collection choice

To answer that

- You must first decide what your research question is
- Then you need to decide what data/variables are needed to scientifically answer the question
Data collection choice

- If that data exist in secondary form, then use them to the extent you can, keeping in mind limitations.
- But if it does not, and you are able to fund primary collection, then it is the method of choice.