

# Where do data come from?

- How do we collect / find data?
  - Secondary data
  - Primary data

# Secondary Data

**Data gathered and recorded by someone else prior to and for a purpose other than the current project.**

**Is often:**

- **Historical**
- **Already assembled**
- **Needs no access to subjects**

# **Common Research Objectives for Secondary Data studies**

**Fact Finding - Identifying success patterns**  
**- Tracking trends**

**Model building - Estimating market potential**  
**- Forecasting enrollment**  
**- Selecting courses, subjects**

**Data Base - Development of**

**Prospect Lists**

**Marketing**

**- Enhancement of**

**Customer Lists**

# Advantages of Secondary Data

- **Inexpensive**
- **Obtained Rapidly**
- **Information is not Otherwise Accessible**
- **Can Provide Insights for more formal research**

# **Disadvantages of Secondary Data**

- **Uncertain Accuracy**
- **Data Not Consistent with Needs**
- **Inappropriate Units of Measurement**
- **Time Period Inappropriate (Dated)**

**What Criteria should be used to evaluate the applicability of Secondary data?**

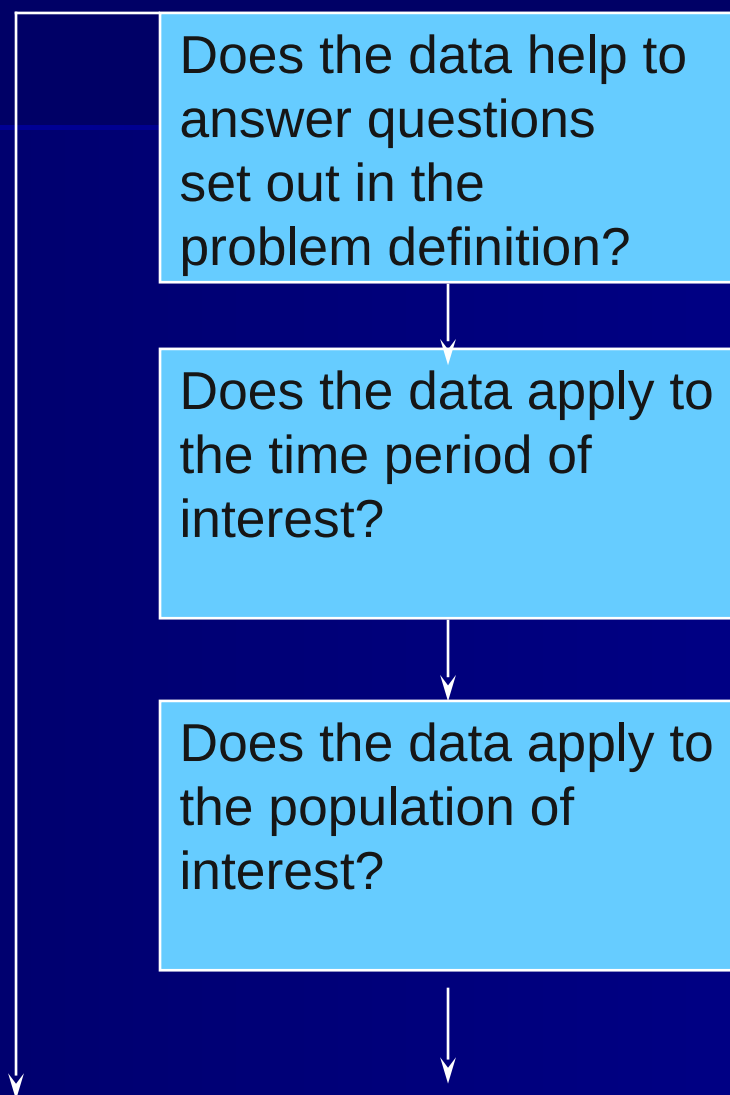
# Evaluating Secondary Data

**Applicability  
to project  
objectives**

Does the data help to answer questions set out in the problem definition?

Does the data apply to the time period of interest?

Does the data apply to the population of interest?



# Evaluating Secondary Data (continued)

**Applicability  
to project  
objectives**

Do the other terms  
and variable  
classifications  
presented apply?

Are the units of  
measurement  
comparable?

**Accuracy  
of the data**

If possible, go to the  
original source of the  
data?





# Evaluating Secondary Data (continued)

**Accuracy  
of the data**

Is the cost of data acquisition worth it?



Is there a possibility of bias?



Can the accuracy of data collection be verified?

# Internal Data

- **Accounting information**
- **Sales information**
- **Backorders**
- **Customer complaints**

# 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

- 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 .



# 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.