CSC 220: Exploring Global Health Crisis Data  
FALL 2022 - Week 11

Research Assistant: Joori Almarzouki

jalmarzouki@uri.edu

horizontal line

# 18 November 2022

Youtube video from ASU: [What are Data and Data Literacy](https://www.youtube.com/watch?v=yhO_t-c3yJY&ab_channel=ArizonaStateUniversity)

Section 1:

What is data literacy?

*"The world’s most valuable resource is no longer oil, but data."*

*--The Economist, 2017*

If the idea of data literacy makes you nervous, think about times you have successfully solved problems. Have you ever wanted to:

* Recreate a favorite meal without having a recipe?
* Figure out whether adopting a certain pet would be a good fit for your lifestyle?
* Understand how to care for a loved one who was feeling ill?

What did you do? You likely started with a question and then used a variety of methods to collect facts to help you answer that question. You may have noticed that answering one question led to additional questions and the need for additional facts. Once you answered all your questions, you probably implemented the answers in a solution and communicated that solution with others.

If any of this sounds familiar, you already have experience with data literacy.

**Data literacy definition:**

Simply stated, data literacy is the ability to explore, understand, and communicate with data.

Section 2:

Asking good questions

*"If I had an hour to solve a problem and my life depended on the solution, I would use the first 55 minutes determining the proper question to ask, for once I know the proper question, I could solve the problem in less than five minutes."*

*--Albert Einstein (1879-1955)*

Every day there are millions of opportunities to improve people's lives by making better use of data. Are you interested in disease research, education patterns, industrial efficiency, patient care, or government spending? The opportunities are endless!

To begin making better use of data, consider an important aspect of data literacy, exploration. The key to successful data exploration is the formulation of good questions.

**For example**, you may love dogs and care about health. You could ask, "*Is owning a dog good for one's health?*" Now, compare that question to this one: "*How do dog owners and non-dog owners compare in health outcomes among people with chronic illnesses in the United States?*"

**The first question is broad; there are no clear criteria for what counts as "good." The second question is much more specific. It uses clearly defined terms and narrows its focus to a specific population. You could more easily explore data to answer the second question than the first.**

Asking why

When you begin to explore the answer to any question, you will likely find yourself asking follow-up questions. Usually, it is not enough to stop after answering your original question. In the example above, if it turns out that health outcomes among people with chronic illnesses in the United States vary with dog ownership, the next question to ask is, "why."

*"The greatest experts in any field are those who never forget how much they have to learn."*

*--Stephen Few, Now You See It*

The prominent information technologist Stephen Few identified a list of traits that help people to work effectively with data, traits that he calls "aptitudes and attitudes." These traits will also help you to ask better questions.

Aptitudes and attitudes that inspire good questions:

* Interest, curiosity, and imagination
* Self-motivation
* Open-mindedness and flexibility
* Awareness of what's worthwhile, pattern spotting, and healthy skepticism
* Methodicalness
* Ability to analyze and ability to synthesize
* Familiarity with the data
* Skills of data analysis

Section 3:

Data fundamentals

*"We believe in the triumph of facts. A collection of data is a collection of facts, and facts are the foundation of human progress."*

*--Tableau Software website*

You know that data literacy is the ability to explore, understand, and communicate with data. But what exactly is "data"?

A collection of data is a collection of facts. Even more specifically, consider this expanded definition. Jeffrey Leek, a data scientist working as a professor at Johns Hopkins Bloomberg School of Public Health, adapted this expanded definition from Wikipedia:

*"Data is comprised of values of* ***qualitative*** *or* ***quantitative variables****, belonging to a* ***set of items****."*

* **Set of items**: Sometimes called the population, this is the group of objects you are interested in.
* **Variable**: A measurement, property, or characteristic of an item that may vary or change.
* **Qualitative**: describes qualities or characteristics, such as gender, name, or hair color.
* **Quantitative**: addresses measurable characteristics, such as height, weight, or temperature.

How is data collected?

Data can be collected in a variety of ways, including questionnaires, interviews, observations, analysis of documents, web scraping, and machine measurements. Received or collected data is called raw data. **Raw data**, which can also be known as source data or primary data, **has not been processed in any way**. This means it has not been run through any software, had any variables manipulated, had any data removed from the data set, and has not been summarized in any way. **Raw data often allows the fullest range possible for data analysis**, since no data has been removed or summarized.

Some examples of raw data include: Unformatted Excel files, or hand-entered numbers collected when looking through a microscope.

Data sources

A data source contains the data used for exploration, understanding, and communication. In Tableau, for example, every chart you see has a connected data source that supplies the data.

**Some data source types:**

**SPREADSHEETS:** such as Microsoft Excel or Google sheets, organize data in a flat structure, which means the records are stored as single rows of data.

**RELATIONAL DATABASES:** store data in multiple tables, with each row assigned a unique identifier. Users pull data from different tables together using Structured Query Language (SQL). The "relational" aspect indicates a logical connection between different tables.

**CLOUD DATA:** Sometimes, users prefer to store their data in the cloud. This includes data stored in such places as Amazon Web Services or Microsoft Azure.

**Class Discussion:**

* Section 1: What is the importance of being data literate?
* Section 2: What are the criteria for asking good questions? Demonstrate how some of the aptitudes and attitudes listed in section 2 help you ask better questions.
* Section 3: What are some pros and cons of using raw data?