How can the representation and analysis of data help us study real world problems:

Statistics is a way of thinking that can lead to better decision making. They are the methods that allow us to work with data effectively.

What is data?

Data are the facts about the world that one seeks to study and explore.
Some data are unsummarized, such as the facts about a single ticket-selling transaction, whereas other facts, such as weekly ticket grosses, are **summarized**, derived from a set of unsummarized data.

Business Statistics
Business statistics focuses on interpreting the results of applying statistical methods. You interpret those results to help you enhance business processes and make better decisions. Specifically, business statistics provides you with a formal basis to summarize and visualize business data, reach conclusions about that data, make reliable predictions about business activities, and improve business processes.

To minimize errors while making predictions using statistical tools, we should use a framework that organizes the set of tasks that we follow to apply statistics properly. The five tasks that comprise the **DCOVA framework** provide one such framework.

DCOVA Framework
• **D**efine the data that you want to study to solve a problem or meet an objective.
• **C**ollect the data from appropriate sources.
• **O**rganize the data collected, by developing tables.
• **V**isualize the data collected, by developing charts.
• **A**nalyze the data collected, to reach conclusions and present those results.

Business analytics

Of the recent changes that have made statistics an important part of our business education, the emergence of the set of methods collectively known as business analytics may be the most significant change of all. **Business analytics** combine traditional statistical methods with methods from management science and information systems to form an interdisciplinary tool that supports fact-based decision making. Business analytics include

• statistical methods to analyze and explore data that can uncover previously unknown or unforeseen relationships.
• information systems methods to collect and process data sets of all sizes, including very large data sets that would otherwise be hard to use efficiently.
• management science methods to develop optimization models that support all levels of management, from strategic planning to daily operations.

Big data are collections of data that cannot be easily browsed or analyzed using traditional methods. *Big data* implies data that are being collected in huge *volumes*, at very fast rates or *velocities* (typically in near real time), and in a *variety* of forms other than the traditional structured forms such as data processing records, files, and tables and worksheets.

Operational definitions

A **variable** defines a characteristic, or property, of an item or individual that can vary among the occurrences of those items or individuals. For example, for the item "book," variables would include title and number of chapters, as these facts can vary from book to book.

Using the definition of variable, you can state the definition of ***data***, in its statistical sense, as the set of values associated with one or more variables

A **statistic** refers to a value that summarizes the data of a particular variable

For the inferential methods that will be discussed later , you must always look for logical causality. **Logical causality** means that you can plausibly claim something directly causes something else. For example, you wear black shoes today and note that the weather is sunny. The next day, you again wear black shoes and notice that the weather continues to be sunny. The third day, you change to brown shoes and note that the weather is rainy. The fourth day, you wear black shoes again and the weather is again sunny. These four days seem to suggest a strong pattern between your shoe color choice and the type of weather you experience. You begin to think if you wear brown shoes on the fifth day, the weather will be rainy. Then you realize that your shoes cannot plausibly influence weather patterns, that your shoe color choicecannot *logically cause* the weather. What you are seeing is mere coincidence.

Trimmed mean

The mean is quite sensitive to a single outlier, whereas the median is impervious to many outliers. Since extreme behavior of either type might be undesirable, we briefly consider alternative measures that are neither as sensitive as nor as insensitive as . To motivate these alternatives, note that and are at opposite extremes of the same "family" of measures. The mean is the average of all the data, whereas the median results from eliminating all but the middle one or two values and then averaging. To paraphrase, the mean involves trimming 0% from each end of the sample, whereas for the median the maximum possible amount is trimmed from each end. A **trimmed mean** is a compromise between and . A 10% trimmed mean, for example, would be computed by eliminating the smallest 10% and the largest 10% of the sample and then averaging what remains.