

MITOCW | MIT15_071S17_Session_7.2.01_300k

In this lecture, we'll discuss the idea of using visualization to better understand data and to provide insights on the problem we're addressing.

Why visualization?

People often say that a picture is like a thousand words.

In the same spirit, John Tukey, a major statistician at Princeton, wrote that "the picture-examining eye is the best finder we have of the wholly unanticipated." Visualizing data allows us to discern relationships, structures, distributions, outliers, patterns, behaviors, dependencies, and outcomes.

Visualization is further useful for initial data exploration, for interpreting models, and for communicating results effectively.

Let us give some examples of different modes of visualization that illustrate these points.

The figure shows the miles per gallon of a car as a function of the car's weight.

The figure clearly illustrates that as the weight of the car increases, the miles per gallon decrease.

The same graph, but now colors of the points signify the number of cylinders in the car: four for red, six for green, and eight in blue.

On the same data, we now plot a regression line that captures the intuition that as the weight of the car increases, the miles per gallon decrease.

In this plot, we'll visualize burglaries in the city of Houston by combining data and geographical location in a map.

This plot illustrates, using a heat map, the usage of rented bicycles from the Hubway company.

The horizontal axis is the hour of the day, and the vertical axis the day of the week, starting on Sunday.

The heat map shows that the usage increases during the morning and night rush hours on weekdays.

The next plot helps us visualize histograms of different categories using the Hubway data.

This plot shows US unemployment by state.

The lighter colors corresponding to smaller unemployment, and the darker colors corresponding to larger unemployment rates.

The plan this week is to create all of these visualizations.

We'll see how visualizations can be used to better understand data, communicate information more effectively, show the results of analytical models.

In the next video, we'll discuss the World Health Organization, and how they use visualizations effectively.