In this video, we'll discuss how we can create visualizations that are used in predictive policing models.

In almost every application, before we even consider a predictive model, we should try to understand the historical data.

Many cities in the United States and around the world, provide logs of reported crimes, usually the time, location, and nature of the event.

In this lecture, we'll use data from the city of Chicago, in the United States, about motor vehicle thefts.

Given this data on crimes, suppose we wanted to communicate crime patterns over the course of an average week.

We could display daily crime averages using a line graph, like the one shown here, but this doesn't seem too useful.

We can see that crime tends to be higher on Saturday, but when on Saturday, and where?

We could replace our x-axis with the hour of the day and have a different line for every day of the week to understand when crime occurs in more detail.

But this would be a jumbled mess with seven lines, and probably very hard to read.

We could instead use no visualization at all, and instead present information in a table, like the one shown here.

For each hour and day, we have the total number of crimes that occurred.

This is a valid representation of the data, but large tables of numbers can be hard to read and understand.

So how can we make the table more interesting and usable?

A great way to visualize information in a two-dimensional table is with a heat map.

Heat maps visualize data using three attributes.

Two of the attributes are on the x and y-axes, typically displayed horizontally and vertically.

The third attribute is represented by shades of color.

In this example, lower values in the third attribute correspond to colors closer to blue, and higher values in the third attribute correspond to colors closer to red.

For example, the x-axis could be hours of the day, the y-axis could be days of the week, and the colors could correspond to the amount of crime.

In a heat map, we can pick different color schemes based on the type of data to convey different messages.

In crime, a yellow to red color scheme might be appropriate because it can highlight some of the more dangerous areas in red.

Your eye is naturally drawn to the red areas of the plot.

In other applications, both high and low values are meaningful, so having a more varied color scheme might be useful.

And in other applications, you might only want to see cells with high values, so you could use a gray scale to make the cells with low values white.

The x and y-axes in a heat map don't need to be continuous.

In our example, we have a categorical or factor variable -- the day of the week.

And we can even combine a heat map with a geographical map, which we'll discuss later in this lecture.

This type of heat map is frequently used in predictive policing to show crime hot spots in a city.

In this lecture, we'll use Chicago motor vehicle theft data to explore patterns of crime, both over days of the week, and over hours of the day.

We're interested in analyzing the total number of car thefts that occur in any particular hour of a day of the week over our whole data set.

In the next video, we'll switch to R, and start by creating a basic line plot counting the total crime by the day of the week.