**Introduction to categorical plots using Seaborn**

Creating and updating categories is only part of using categorical data. Let's start working on building visualizations that use categorical data.

**Our third dataset**

In this chapter, we will use a new dataset, the Las Vegas TripAdvisor reviews dataset. This dataset contains information on 504 reviews from 21 hotels in Las Vegas collected in 2015.

**Las Vegas reviews**

Using the dot-info method on our dataset, we can see that we have information on the hotel guest, such as their home country and traveler type, as well as information on the hotel, such as if it has a pool, gym, tennis court, or other amenities.

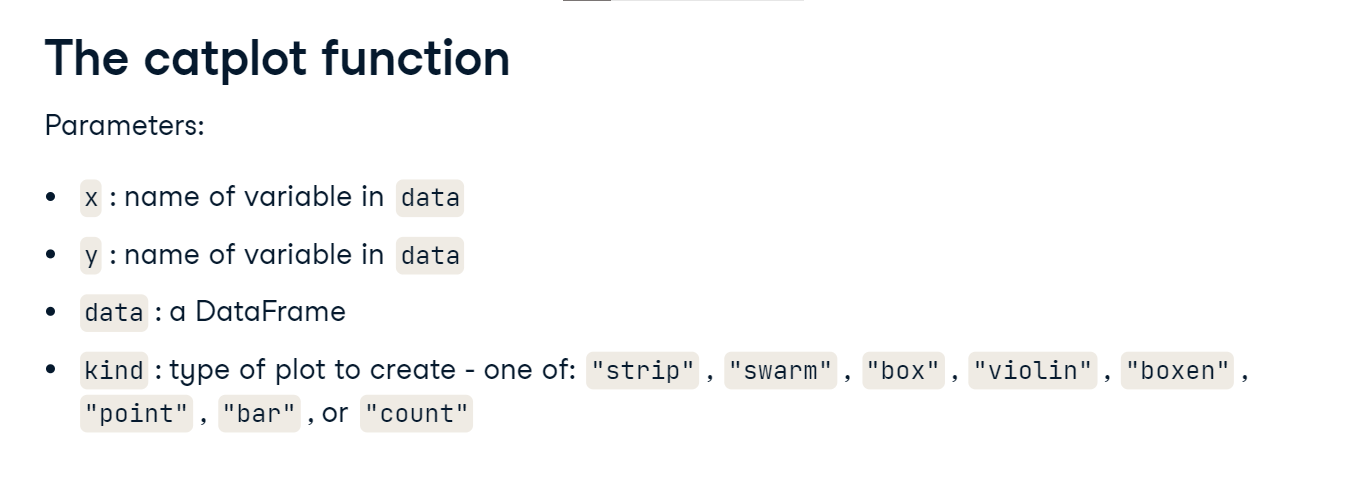
1. 1 https://www.kaggle.com/crawford/las-vegas-tripadvisor-reviews

**Seaborn**

To visualize this data, we are going to use the Python library seaborn, which we have loaded as sns. Datacamp offers two great courses on seaborn - if you'd like additional practice, definitely give those courses a try. For our purposes, we will focus solely on categorical plots using seaborn, and more specifically, the catplot function. Note that Seaborn is based off of the Python library matplotlib, so we have loaded matplotlib's pyplot as plt. Depending on the environment you are coding in, you may need to run plt-dot-show after creating your graphic for it to display.

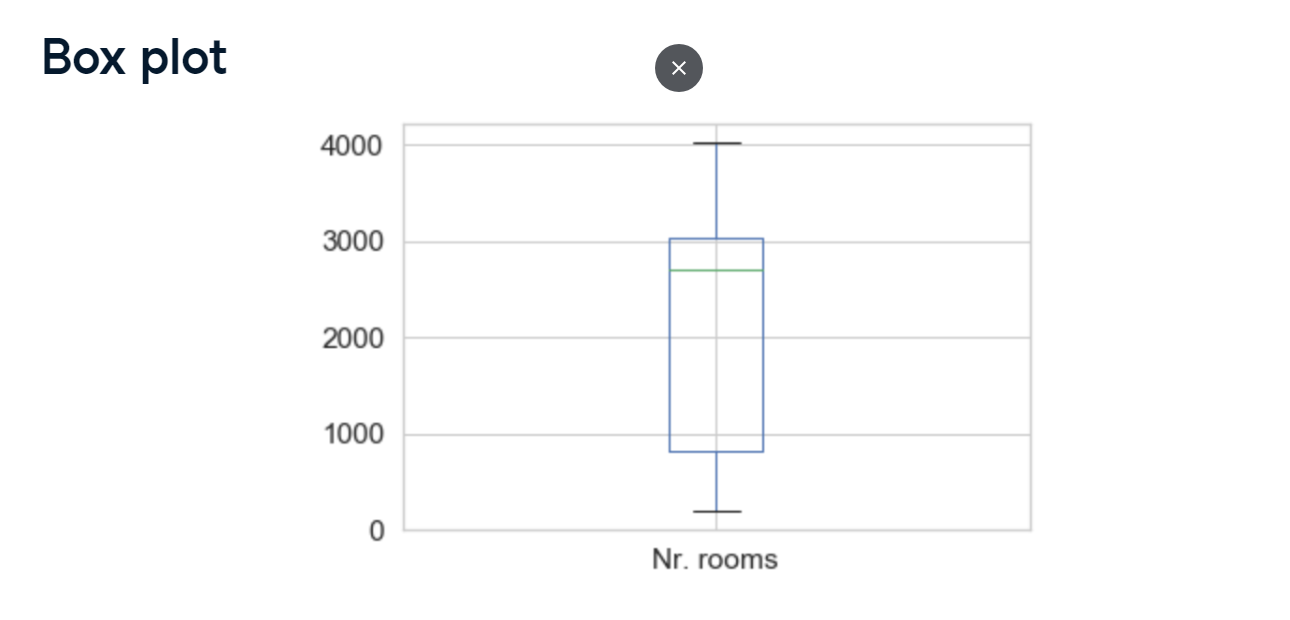
**The catplot function**

Whether we are creating scatterplots, distribution plots, or just counting the number of responses, the catplot function is capable of handling the task. Let's look at the common parameters of a catplot. Both the x and y parameter are names of variables found in the DataFrame being used, while the kind parameter specifies the type of graphic to create. In this chapter we will cover several uses of the kind parameter.



**Box plot**

One type of plot that catplot can create is a box plot. As a reminder, a box plot shows information on the quartiles of numerical data. In this example, we are looking at the number of rooms in hotels. The middle line of the box shows the median of the data, which is hovering around 2,800 beds. The bottom and top of the box show the 25th and 75th quartiles and look to be around 800 and 3000 beds respectively. Consult the linked wikipedia page if you need a refresher on the other elements of a box plot.

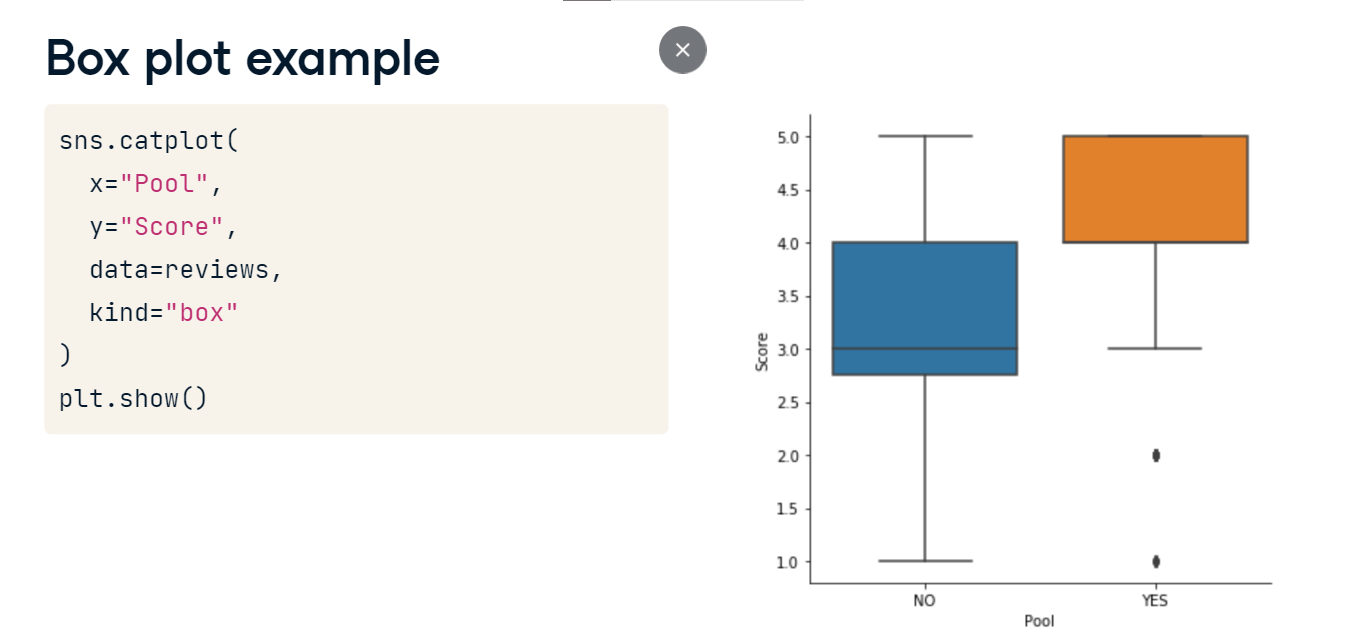


**Review score**

Before we look at an example, let's understand the numerical column we are going to explore. The review score is a value between one and five, and is the rating of the hotel given by the person doing the review. Most scores are four or five, but there are a few that are three and below.

**Box plot example**

Let's look at the review score, across the categorical variable Pool, using box plots. This means that we can check the distribution of score given the hotel has a pool or not. Notice that for each category in Pool, a box plot for responses that match that category has been created. Two other things you may notice are that the text is tiny, and that it's hard to tell where the two outliers are - those tiny black dots under the orange box.



**Two quick options**

We can fix both of these issues using sns-dot-set and sns-dot-set-style. First, we increase the font size using font-scale, and then we add gridlines to the plot by specifying whitegrid for the style. The new graphic is easier to read, and we can tell that there are outliers at 2 and 1. It looks like a couple guests who stayed at a hotel with a pool did not like their experience.

