Cheatsheets / **Data Visualization in Python**

Introduction to Seaborn

Seaborn

Seaborn is a Python data visualization library that builds off the functionalities of Matplotlib and integrates nicely with Pandas DataFrames. It provides a high-level interface to draw statistical graphs, and makes it easier to create complex visualizations.

Estimator argument in barplot

The `estimator` argument of the `barplot()` method in Seaborn can alter how the data is aggregated. By default, each bin of a barplot displays the mean value of a variable. Using the `estimator` argument this behaviour would be different.

The `estimator` argument can receive a function such as `np.sum`, `len`, `np.median` or other statistical function. This function can be used in combination with raw data such as a list of numbers and display in a barplot the desired statistic of this list.

Seaborn barplot

In Seaborn, drawing a barplot is simple using the function `sns.barplot()`. This function takes in the paramaters `data`, `x`, and `y`. It then plots a barplot using `data` as the dataframe, or dataset for the plot. `x` is the column of the dataframe that contains the labels for the x axis, and `y` is the column of the dataframe that contains the data to graph (aka what will end up on the y axis).

Using the Seaborn sample data “tips”, we can draw a barplot having the days of the week be the x axis labels, and the total_bill be the y axis values:

```
sns.barplot(data = tips, x = "day", y = "total_bill")
```

Barplot error bars

By default, Seaborn's `barplot()` function places error bars on the bar plot. Seaborn uses a bootstrapped confidence interval to calculate these error bars.

The confidence interval can be changed to standard deviation by setting the parameter `ci = "sd"`.

Seaborn hue

For the Seaborn function `sns.barplot()`, the `hue` parameter can be used to create a bar plot with more than one dimension, or, in other words, such that the data can be divided into more than one set of columns.

Using the Seaborn sample data "tips", we can draw a barplot with the days of the week as the labels of the columns on the x axis, and the total_bill as the y axis values as follows:

```
sns.barplot(data = tips, x = "day", y = "total_bill", hue = "sex")
```

As you can see, `hue` divides the data into two columns based on the "sex" - male and female.

Seaborn function plots means by default

By default, the seaborn function `sns.barplot()` plots the means of each category on the x axis.

In the example code block, the barplot will show the mean satisfaction for every gender in the dataframe `df`.

```
sns.barplot(data = df, x = "Gender", y = "Satisfaction")
```

Box and Whisker Plots in Seaborn

A box and whisker plot shows a dataset's median value, quartiles, and outliers. The box's central line is the dataset's median, the upper and lower lines marks the 1st and 3rd quartiles, and the "diamonds" shows the dataset's outliers. With Seaborn, multiple data sets can be plotted as adjacent box and whisker plots for easier comparison.

Seaborn Package

Seaborn is a suitable package to plot variables and compare their distributions. With this package users can plot univariate and bivariate distributions among variables. It has superior capabilities than the popular methods of charts such as the barchart. Seaborn can show information about outliers, spread, lowest and highest points that otherwise would not be shown on a traditional barchart.

← **Previous**

Related Courses

Course

Data Visualization in Python

Learn how to present data graphically with Python, Matplotlib, an

Career Path

Data Science