

Visualization with Scatter Plots

This lesson explains what a scatter plot is, why it is used, and how to visualize data with scatter plot using Python libraries.

WE'LL COVER THE FOLLOWING ^

- Introduction to scatter plots
- Scatter plots in Python

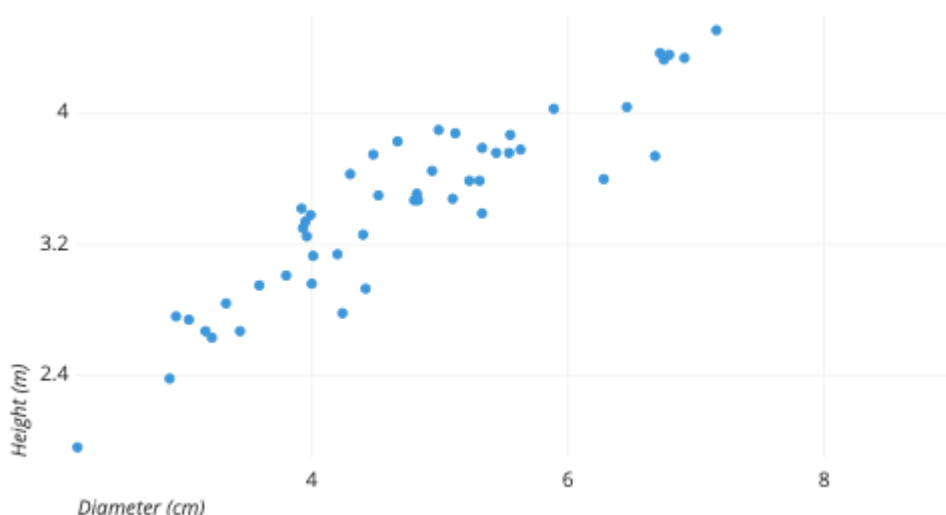
Introduction to scatter plots

First, let's take a look at **scatter plots**.

Scatter plots are great for plotting two variables to visualize how they might correlate and what relationship there might be between the variables.

As an example, look at the scatter plot below.

In this plot, we see a positive relationship between **Height** and **Diameter**. It is easy to see that relationship in a scatter plot by looking for points that are close to the 45-degree line pointing up and to the right.



Scatter plots in Python

Scatter plots in Python

Seaborn makes it easy to create scatter plots using the `lmpplot()` function. The key inputs are:

- `x` which is the column name for your x-axis variable
- `y` which is the column name for your y-axis variable
- `data` which is your Pandas dataframe

Let's look at an example:

```
from sklearn.datasets import load_boston
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Set the palette and style to be more minimal
sns.set(style='ticks', palette='Set2')

# Load data as explained in introductory lesson
boston_data = load_boston()
boston_df = pd.DataFrame(boston_data.data, columns=boston_data.feature_names)

# Create the scatter plot
sns.lmpplot(x="CRIM", y="NOX", data=boston_df)
# Remove excess chart lines and ticks for a nicer looking plot
sns.despine()
```

You can see in our code that we pass the `x` and `y` variables we want to use from our `data`, which is expected to be a Pandas dataframe that has the x and y values we wish to plot.

By default, it plots a regression line so you can see potential linear relationships more easily. Our variables appear to be somewhat positively correlated. The lighter shade around the line is the 95% confidence interval for our regression line and is calculated using bootstraps of our data. You can see all of the options for this plot [here](#).

Also, we used one of seaborn's styles and palettes. This allows us to control at a high-level how our plots will look. You can learn more about controlling aesthetics [here](#). We chose a style and palette which creates a more minimalistic look. We also called `despine()` after creating the plot to remove excess chart lines and ticks which I believe creates a cleaner look.

That's how your data can be visualized through a scatter plot. Next, we'll look at bar plots.