**Visualizing the Data:**

Visualizing the data is a very important step after we get the dataset,

before training the data, we need to have a general understanding of our dataset.

Scatter plots show how much one variable is affected by another.

“Matplotlib” can help us to draw scatter plot.

**Examples 1:**

* **Import the library**

 import matplotlib.pyplot as plt

* **Drawing scatter plot between two variables, and size is 0.1, color is red**

 plt.scatter(housing[“median\_income”], housing["median\_housing\_value”],s=0.1, color=”red”)

* **Add a title and XY labels**

 plt.title(“relationship between median\_income and median\_housing\_value”)

 plt.xlabel(“median income”)

 plt.ylabel(“median housing value”)

* **Change some parameters**

 plt.scatter(housing[“median\_income”], housing["total\_rooms”],s=0.1, color=”blue”)

 plt.title(“relationship between median income and total\_rooms”)

 plt.xlabel(“median income”)

 plt.ylabel(“total\_rooms”)

 :

Histograms are used to graphically summarize and display the distribution of a process data set. “Matplotlib” can help us to draw histograms.

**Example 2:**

* **Import the library**

 Import matplotlib.pyplot as plt

* **Drawing histogram for all 9 variables**

 housing.hist(bins=10, figsize=(10,10), color = (“red”))

More bins means the given range is divided to more parts.

Figure size, first parameter represents length,

second parameter represents width.

**Change some parameters**

housing[“latitude”].hist(bins = 10, figsize(10,10),color=(“red”)). housing[“latitude”].hist(bins=1000, figsize(10,10),color=(“blue”))



**Reference:**

<https://www.tutorialspoint.com/matplotlib/matplotlib_scatter_plot.htm> --Scatter Plot

<https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.scatter.html> -- Scatter Plot

<https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.hist.html> --Histogram

<https://github.com/ageron/handson-ml--> github jupyter notebook