# 1_data_plot_exercise

May 20, 2019

```
In [ ]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np
```

```
In [ ]: data_url = './housing.data' #Data URL
        df_data = pd.read_csv(data_url, sep='\s+', header = None) #csv   , separate  , Column
        df_data.columns = [
            'CRIM','ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX',\
            'PTRATIO' ,'B', 'LSTAT', 'MEDV']
        # Column Header
        df_data.head()
```

```
In [ ]: fig = plt.figure()
        ax = []
        for i in range(1,5):
            ax.append(fig.add_subplot(2,2,i))

        ax[0].scatter(df_data["CRIM"], df_data["MEDV"])
        ax[1].scatter(df_data["PTRATIO"], df_data["MEDV"])
        ax[2].scatter(df_data["AGE"], df_data["MEDV"])
        ax[3].scatter(df_data["NOX"], df_data["MEDV"])
        plt.show()
```

```
In [ ]: fig = plt.figure()
        fig.set_size_inches(10.0, 10.0, forward=True)
        ax = []
        for i in range(1,5):
            ax.append(fig.add_subplot(2,2,i))

        columns = ["CRIM", "PTRATIO", "AGE", "NOX"]
        colors = ["b", "g", "c", "r"]
        for i in range(4):
            ax[i].scatter(df_data[columns[i]], df_data["MEDV"], color=colors[i], \
                          label=columns[i])
            ax[i].legend()
            ax[i].set_title(columns[i])
        plt.subplots_adjust(wspace=0, hspace=0)
        plt.show()
```

```
In [ ]: df_data.plot()
        plt.show()
```

```
In [ ]: fig = plt.figure()
        fig.set_size_inches(10,5)
        ax_1 = fig.add_subplot(1,2,1)
        ax_2 = fig.add_subplot(1,2,2)
        ax_1.plot(df_data["MEDV"])
        ax_2.hist(df_data["MEDV"], bins=50)
        ax_1.set_title("House price MEDV")
        ax_2.set_title("House price MEDV")

        plt.show()
```

```
In [ ]: from sklearn.preprocessing import StandardScaler
        std_scaler = StandardScaler()
        scale_data = std_scaler.fit_transform(df_data)
        scale_data
```

```
In [ ]: fig = plt.figure()
        ax = fig.add_subplot(1,1,1)
        ax.boxplot(scale_data, labels=df_data.columns)
        fig.set_size_inches(10, 5)
        plt.show()
```

```
In [ ]: pd.scatter_matrix(df_data, diagonal="kde", alpha=1,  figsize=(15, 15))
        plt.show()
```

```
In [ ]: corr_data = np.corrcoef(scale_data.T[:5])
        corr_data
```

```
In [ ]: corr_data.shape
```

```
In [ ]: fig = plt.figure()
        ax = fig.add_subplot(111)
        cax = ax.matshow(corr_data, vmin=-1, vmax=1,  interpolation='nearest')
        fig.colorbar(cax)
        fig.set_size_inches(10,10)
        ticks= np.arange(0,5,1)
        ax.set_xticks(ticks)
        ax.set_yticks(ticks)

        ax.set_xticklabels(df_data.columns[:5])
        ax.set_yticklabels(df_data.columns[:5])

        plt.show()
```

```
In [ ]:
```