HW 08. Exploratory Data Analysis in Pandas

We are going to use the SOCR data on the height and weight of 25 thousand teenagers.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Task 1: Import data from the file weights_heights.csv (which you saved to your local drive) to the dataframe

```
#data = #your code here
```

```
data = pd.read_csv("teenagers.csv")
```

Task 2: Plot histogram of the Height. Make the color red, label the x-axis as "Height", add title to the plot "Height distribution (inch)"

```
#your code here
```

Task 3: Output first 10 rows of the dataframe.

```
# your code here
```

Task 4: Plot histogram of the weight distribution. Make the color green. Label the x-axis, add title to the plot.

```
# your code here
```

Task 5: Use the method *pairplot* of the module Seaborn to visualize the dependencies in your dataframe.

```
# your code here
```

Task 6: Bin the weight into 3 bins: 1 - for light weight, 2 - for medium and 3 for large weight. In your dataframe, create a new column *weight_category* and record the weight category for the each row in that column.

```
# your code here
```

Task 7: Use the method *boxplot* from the module Seaborn to visualize height distribution properties for each weight category. Label *y*-axis as "Height", and *y*-axis as "Weight category"

```
# your code here
```

Task 8: Make a scatter plot of weight vs height, label the axes and add a title to the plot.