# LAB 2 - Simple Linear Regression (continued)

In this lab, we will repeat our efforts to implement single linear regression (which was already done in LAB 1), with small extra steps. This time, we will need to split the dataset into two unequal parts. First, we will have to *train* the algorithm using the larger part, i.e. we will use our simple linear regression algorithm on that half. Next, we will see how our estimation fares against the other part, we will *test* our regression line using that one. Here are the instructions:

- (20 pts) Extract the "Age", "Height" and "Salary" columns from the .csv file just like you did in the first lab session. Then, separate the last 20 rows of those columns from the rest. These 20 rows are going to be the *test* data, and the rest is going to be the *train* data.

- (20 pts) Perform the linear regression algorithm on the *train* data, with "Age" being $x$ and "Salary" being $y$ (via the function you implemented in the previous lab, it would be useful if you imported your previous implementation here). Save the coefficients $(b_{age0}, b_{age1})$.

- (10 pts) Repeat the step above with "Height" being $x$ and "Salary" being $y$. Save the coefficients $(b_{hgt0}, b_{hgt1})$.

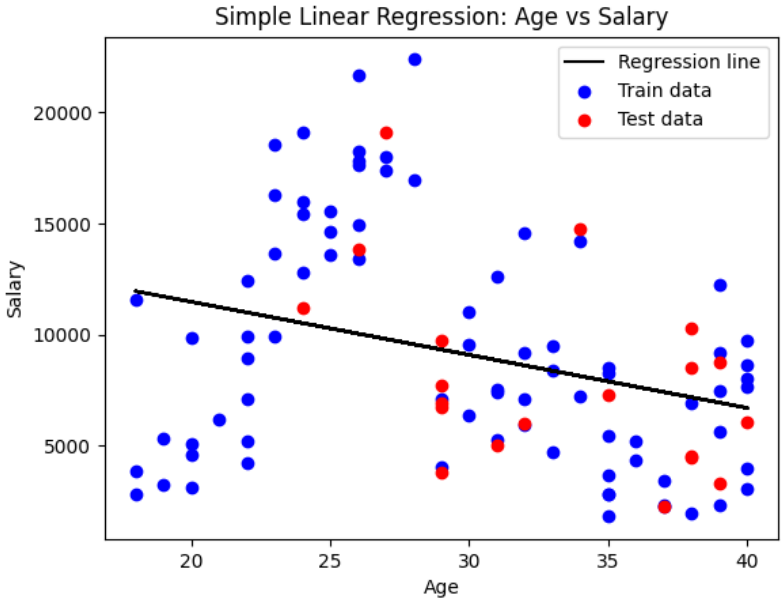
- (50 pts) Test your model by plotting two graphs.

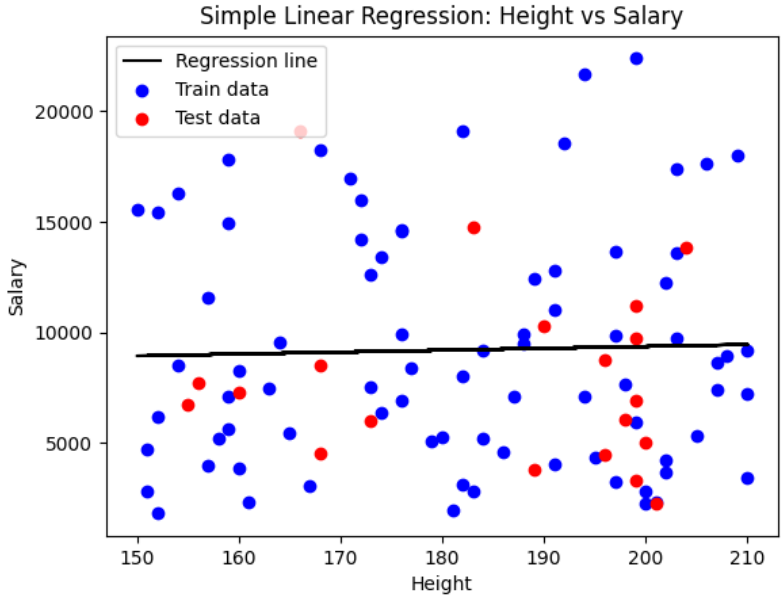
    - Graph 1:
        - Plot the *train* portion of the "Age" and "Salary" columns as a scatter plot, in blue.
        - Plot the *test* portion of the "Age" and "Salary" columns as a scatter plot, in red.
        - Calculate and draw the regression line $\hat{y}$, against "Age" as $x$, in black.
        - Put proper title, axis labels and legend.

    - Graph 2:
        - Same graph as above, except "Age" becomes "Height".

    - Finally, show the plot. It's important to do this at the VERY end of your code, so that we can see both plot windows at the same time without needing to close one of them.

The plots should look like this:




**IMPORTANT NOTE:** The instructions specifically ask you to manually code the calculations. Bypassing these calculations (via using an external library, `sklearn` being the most popular one) will mean that the calculations are not done, therefore corresponding sections will yield 0 points in total.

**IMPORTANT NOTE 2:** Any submissions not running due to a syntax error will receive an automatic 0, so make sure your code is error-free!