Homework 4, My Name.

Please submit the solution on Canvas into the corresponding assignment (e.g. "Homework #1") in the form of R Markdown report, knitted into either of the available formats (HTML, pdf or Word). Provide only code and output. NO NEED TO COPY THE PROBLEM FORMULATION (!)

Problem #1

For the *fl_crime.csv* data, proceed to:

1. Plot the *crime* against *education*. Comment on what you observe. Can this relationship be reasonably described by a straight line?
2. Report the correlation between *crime* and *education*. Explain (just mathematically, not sociologically) why you think that the correlation has this particular value, judging by your plot from part 1.
3. Proceed to fit a *crime* ~ *education* linear regression, and

    a. Write down the fitted equation.
    b. Interpret the intercept. Does it make sense? Why?
    c. Interpret the slope.
    d. Report and interpret the R^2 value.

4. Hard-code the calculation of R^2 value from scratch (explicitly applying formula from the top of slide #38), via only using the $y = fl\_crime\$crime$ and $y.hat = predict(lm.obj)$ as the quantities to work with. Double-check it with R^2 from part 3.
5. Calculate the predicted crime rate for education level of

    - 70,
    - 35.

    Comment on whether we can trust either of the predicted values, and why (name the issue).

Problem #2

Broadband.csv contains data on each country's GDP (measured in billions USD) and the number of broadband subscribers. Proceed to

1. Plot the # of broadband subscribers against GDP and calculate the correlation between these two variables. Describe the relationship.
2. Fit a linear regression for the # of broadband subscribers onto GDP, and

    a. Write down the fitted equation.
    b. Interpret the intercept. Does it make sense? Should we trust it? Why?
    c. Interpret the slope.
    d. Report and interpret the R^2 value.

Problem #3

~~3.26~~, ~~3.70~~, ~~3.83~~