Please download the data set `R03_assignment_dataset.csv`. Load it into your R session as a data frame `df`:

```
> df = read.csv("R03_assignment_dataset.csv")
```

Task 3.1 (Visualization).

In this task, we will visualize simple linear regression analyses in the data set.

(a) From the data set, choose **one** metric variable as a predictor x and **one** metric variable as a criterion y.

(b) Calculate the bivariate correlation r_{xy} between x and y. Is the relation between x and y positive, negative, or near zero?

(c) Calculate the explained variance r_{xy}^2.

(d) Create a publication-ready plot for the simple linear regression analysis[1]:

    (i) Create a scatterplot to visualize the relation between x and y.

    (ii) Set labels for the x-axis, the y-axis, and the title.

    (iii) Adjust the range of the axes with `xlim(lower, upper)` and `ylim(lower, upper)`.

    (iv) For global styling, add a *theme* to your plot[2]:

```
> ggplot(...) + ... + theme_classic()
```

    Common choices are `theme_classic()` or `theme_minimal()`, but you can use any theme you like for this task.

    (v) Add a linear regression line to the scatterplot using the function `geom_smooth`.

    (vi) Lower the `alpha` value of the points in the scatterplot to increase the visibility of the regression line.

    (vii) **Optional:** Use the internet to find out to save the plot as a file (e.g. in `png` or `pdf` format).

(e) **Optional:** Repeat the above tasks with an additional *grouping variable* to get separate analyses for each value of that grouping variable.

---

[1]We won't test the assumptions for the underlying regression in this task. However, if you want to be thorough and test the discussed assumptions, feel free to do it in this task as well!

[2]visit `https://ggplot2.tidyverse.org/reference/ggtheme.html` for more info on themes.

Task 3.2 (Multiple Regression).

In this task, we will perform a multiple regression analysis.

(a) From the data set, choose **one** metric variable as a criterion y and **five** metric variables as potential predictors $x_1, \ldots, x_5$.

(b) Fit a multiple linear regression to the data with the function `lm()`:

```
> model = lm(y ~ x1 + x2 + x3 + x4 + x5, data = df) # replace y ... x5
```

(c) Remove non-significant predictors and repeat the analysis. If you have no predictors left, choose new ones from the data set until you have a model with 3–4 statistically significant predictors $(p < .05)$[3].

(d) Check the assumptions of the multiple linear regression. If assumptions are violated, use the internet to estimate the severity of the violation and find out how to proceed with the analysis.

(e) Report the adjusted explained variance R^2_{adj} for your model.

(f) Report whether the *entire* model is statistically sigificant based on the $F-$statistic.

(g) **Optional:** Use your usual text-processing program (e.g. Word) to create a table of the results that contains the following information from the output of `summary(model)`:

- Predictor names and intercept
- $b-$value (regression weights) for each predictor and intercept
- $t-$value for each predictor and intercept
- $p-$value for each predictor and intercept

Please solve the assignment in the `.Rmd` format and export it in a suitable format (e.g. `pdf` or `html`)

---

[3]Please don't to that in real research applications...