

Please download the data set `R04_assignment_dataset.csv`. Load it into your R session as a data frame `df`:

```
> df = read.csv("R04_assignment_dataset.csv")
```

Task 4.1 (Exploring and visualizing the data set).

- (a) How many observations does the data set contain?
- (b) Which variables (columns) does the data set contain?
- (c) Plot the distribution of the variable `sustainability` as a density plot. What do you observe?
- (d) Plot the distribution of the variable `sustainability` as a density plot again, but this time draw separate density plots for each `diet` type¹. What do you observe?
- (e) Extend the previous plot with proper axes labels, a title, and a suitable theme.
- (f) Based on your insights from the visualization: Briefly describe the data and derive statistical hypotheses for the t -test in the following task.

¹**Hint:** Use the `fill` aesthetic in the initial `ggplot` call

Task 4.2 (Exploring group differences).

In this task, you will test the hypothesis whether the subjects' sustainability scores differ with respect to their diets (meat vs. vegetarian vs. vegan).

- (a) Create three vectors `sustainability_meat`, `sustainability_vegetarian`, and `sustainability_vegan` that contain only the sustainability score of the respective group.

Hint: You can use the following approach to extract the data as separate vectors (replace the ... parts accordingly):

```
> sustainability_... = df %>% filter(diet == "...") %>% select(...)
```

- (b) Choose one pairwise comparison (e.g. meat vs. vegetarian) and perform the complete t -test:
- (i) Formulate the null hypothesis H_0 and the alternative hypothesis H_1 .
 - (ii) Calculate the appropriate t -test for your question numerically with the function `t.test()`.
 - (iii) Interpret the result and decide on one of the hypothesis.
 - (iv) Calculate and interpret² the effect size as Cohen's d .
 - (v) Are your results congruent with the plot from the previous task?
- (c) **Optional:** Repeat the hypothesis test for another pairwise comparison (e.g. vegetarian vs. vegan).
- (d) **Bonus:** The `t.test()` function also supports a formula syntax to work with data in long format:

```
> t.test(outcome ~ group_variable, data = df)
```

Prepare the data frame `df` accordingly to use this syntax instead of extracting the outcome variable data as separate vectors.

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

²you can find the conventions on the internet