

Comments on Ninia's assignment

1a. Format of the assignment: is the assignment well structured?

The work is well structured. Tasks are labeled clearly (Task 4, Task 5, etc.), and each task includes both code and explanation. For example, in Task 7 you show histograms in both base R and ggplot, and then move directly into boxplots. That kind of ordering makes it easy to follow.

1b. Format of the assignment: is the formatting appropriate?

Formatting is clean and consistent. He/she used comments to explain code, and the R Markdown headings (###) are helpful for navigation. One small point: calls like `View(microarray_data)` and `View(magic_guys)` are fine when exploring, but they don't show up in knitted HTML/PDF outputs. It might be better to replace those with `head()` or `str()` or something like these so that readers always see the result.

2a. Completeness of the assignment: are all the tasks addressed?

All requested tasks (1, 4–8) are present. For example, Task 6 correctly includes two custom functions (`outlier_check()` and `mean_trimmed()`), as well as short explanations of pipes and the apply family. Nothing important is missing.

3a. Correctness of the answers: are the given answers correct to the best of your understanding?

The calculations are mostly spot-on: `sqrt(10)` and `log2(32)` return the expected results. Summations with `sum(1:1000)` and `seq(2,1000,by=2)` are correct. For Task 5, the CO2 averages and medians are computed separately for Quebec and Mississippi, which is right. One place to clarify for the question “arrange 100 genes in triples,” he/she used `permutations(100, 3)` which gives ~970,200 rows because order matters. If the task intended “choose 3 genes ignoring order,” then `choose(100, 3)` (~161,700) would be the

correct function. A short note in the report explaining the difference between “arrangements” and “combinations” would make this clearer.

4a. Validity of the drawn conclusions: do the provided analyses address the questions asked?

Yes. For example:

In Task 7, he/she plotted Jedi vs. Sith heights both with histograms and boxplots, which directly answers “compare distributions.”

In Task 8, he/she show chromosome length vs. protein-coding genes, complete with correlation statistics. That addresses the question of whether they are correlated.

4b. Validity of the drawn conclusions: are the provided insights conclusive given the conducted analyses?

The insights are appropriate and conclusive at the assignment level. For example, his/her comment that “Quebec plants always have a higher CO₂ uptake than Mississippi plants” is well supported by the plot. Similarly, pointing out the stronger correlation of miRNAs with chromosome length is meaningful. One way to strengthen these even more would be to report the R^2 values from the regression models (e.g., `summary(lm(...))$r.squared`) so readers know how strong the correlations are numerically.

5a. Style of the written R code: is the provided R code well documented?

The code is well commented. For example, in Task 7 he/she annotates that fill controls inside-bin color and color controls the border in ggplot histograms which is very helpful.