Worksheets: Review a species richness analysis

Suppose a fellow student asked you to review their species richness analysis. **Indeed, it contains a mistake!**

As you review their report, *discover how symbols and text convey formatting and the structure of the document.* On the following pages is first a plain text file, `analysis.qmd` (in a single grey box), and then the same file is converted to a formatted PDF (using *Typst*) on the next page. You will compare the two documents to investigate which parts of the plain text are interpreted as formatting and how natural language and R code are intermixed.

Exercise

1. Take a highlighter and mark all the text in `analysis.qmd` that does not appear in the output document. Then look at the PDF and mark everything not present in `analysis.qmd`.
2. Formatting in this PDF includes different levels of headings and *emphasized* (in italics) and **strong** statements (in bold). What parts of `analysis.qmd` convey the formatting information for these elements?
3. Can you find the R output of the analysis in one of the paragraphs in the PDF? There are two ways in which code is integrated into `analysis.qmd`. What are they, and how do they differ?
4. **Bonus: Did you uncover the error in the analysis?**

   If you have enough time in class, continue with the questions below. Otherwise, answer them at home.

5. The output document has a table referenced in the text. What are the two components necessary for creating the reference?
6. At the beginning of `analysis.qmd` is a section fenced by `---`. What do you believe this section is for, and do you notice anything special about how the information is organized? Now compare with the lines starting with `#|`. Do you notice any similarities or differences?
7. There are two blocks fenced by ```` ```{r} ```` and ```` ``` ```` in `analysis.qmd`, but only one of them appears in the output. How is this controlled?

analysis.qmd

````
---
title: "Species richness analysis"
format: typst
---

# Identifying species in microbial communities

To describe the species found in a microbial community, we isolate the DNA of
a sample and then amplify and sequence a suitable barcoding marker, typically
the 16S rRNA gene. We then group all similar sequences into so-called
**amplicon sequence variants (ASV)**. For our purposes here, we assume that
*each ASV corresponds to a bacterial species*. By counting the number of each
ASV in each sample, we get an ASV counts table (@tbl-asv-counts).

```{r}
#| label: tbl-asv-counts
#| echo: false
#| tbl-cap: "ASV counts of samples collected at different depths of a lake."

suppressMessages(library(dplyr))

asv_counts <- read.csv("asv-counts.csv")
asv_counts
```


## Species richness

A simple way of characterizing a microbial community is to count how many
species/ASVs we have in a sample. This metric is called **species richness**.
For example, to determine the species richness of the water sample at 10 m, we
can select just that column, filter the rows for ASVs with one or more counts,
and then count the number of rows.

```{r}
#| echo: true

species_richness_10_m <- asv_counts |>  # Use the ASV counts table.
  select(depth_10_m) |>                 # Select the sample of interest.
  nrow()                                # Count the number of rows.
```

The water sample at 10 m has a species richness of `r species_richness_10_m`.
````

Species richness analysis

Identifying species in microbial communities

To describe the species found in a microbial community, we isolate the DNA of a sample and then amplify and sequence a suitable barcoding marker, typically the 16S rRNA gene. We then group all similar sequences into so-called **amplicon sequence variants (ASV)**. For our purposes here, we assume that *each ASV corresponds to a bacterial species*. By counting the number of each ASV in each sample, we get an ASV counts table (Table 1).

Table 1: ASV counts of samples collected at different depths of a lake.

```
   name depth_10_m depth_20_m
1 ASV_1         28         14
2 ASV_2          5         12
3 ASV_3          0          3
4 ASV_4         14         41
5 ASV_5         18         28
```

Species richness

A simple way of characterizing a microbial community is to count how many species/ASVs we have in a sample. This metric is called **species richness**. For example, to determine the species richness of the water sample at 10 m, we can select just that column, filter the rows for ASVs with one or more counts, and then count the number of rows.

```
species_richness_10_m <- asv_counts |>  # Use the ASV counts table.
  select(depth_10_m) |>                 # Select the sample of interest.
  nrow()                                # Count the number of rows.
```

The water sample at 10 m has a species richness of 5.