\usepackage{booktabs} \usepackage{longtable} \usepackage{array}
\usepackage{multirow} \usepackage{wrapfig} \usepackage{float}
\usepackage{colortbl} \usepackage{pdflscape} \usepackage{tabu}
\usepackage{threeparttable} \usepackage{threeparttablex}
\usepackage[normalem]{ulem} \usepackage{makecell}
\usepackage{xcolor}

TutoRial - Part 2

Marine Ecosystem Dynamics - 2025

Pipes

Pipes, expressed as `%>%` or `|>`, are very useful and make our code clearer. Using pipes, our data *flow* from one function to another.

Exercises

- Rewrite these chunks of code using the pipes

```
sum(c(2.2,4.1,2,pi))
```

```
round(sum(c(2.2,4.1,2,pi)))
```

```
round(sum(c(2.2,4.1,2,pi)), digits = 3)
```

Tidy the data with `tidyr`

As seen in the slides, a tidy table has:

1. Each variable in its own column
2. Each observation in its own row

To reach this, `tidyr` has 4 key functions:

1. `pivot_longer`
2. `pivot_wider`
3. `unite`
4. `separate`

Exercises

- If this is not done yet, download the dataset `zooplankton_seasonality.csv`

You can download the dataset on GitHub

- Import the dataset in your environment

- Is this dataset a tidy dataset?
- Separate the column `Coordinates` in 2 news columns: `Longitude` and `Latitude`
- Combine the column `Group`and `Taxa` into a new column `Group_Taxa` and save the dataframe as `tidy_df`
- Create a wide table with columns having the `Biomass` values for each `Group_Taxa` and save the dataframe as `wide_df`

Data handling with `dplyr`

After finishing tidying the data, we often use the `dplyr` package to process our data.

Exercises

- What is the class of the `Year` columns of the `tidy_df` dataframe?
  If they are not numeric, `mutate` them as numeric values.
- Then, kepp all `Year` between `2012` and `2015`
- Then, only keep the data from the `Station` `BY31`
- Then, `select` all columns except `Longitude` and `Latitude`
- Then, `rename` `Month_abb` as `Month`
- Then, `group_by`: `Month` and `Group_Taxa` and take the `Biomass` average and standard deviation and save the dataframe as `summarized_df`

Ploting the data with `ggplot2`

In this part, we will build a plot step by step using the grammar of graphic in `ggplot2`

- Load the package and only keep the values for the copepod `Acartia` from the `summarised_df` dataset in a new dataset called `acartia`
- Initiate a ggplot with the dataset `acartia` with the `Month` as the x-axis and the `average` biomass as the y-axis
- Add a `barplot` geometry to the plot
- Arrange the bar from the lowest to the highest values
- Add a color filling in the bars according the Month
- Change the axis as `Biomass` and `Month`, and add a title