

## Week 2: R can be Mess; so be tidy!

You've made it through week 1 of R summer school! Well done!

Now that you know your way around the RStudio neighborhood I'll start to get into the nitty gritty details of setting up a good workflow.

### Learning Objectives

1. Able to efficiently “clean up” data sets so they are easy to wrangle in R using the janitor package.
  - a. Here are some good slides on how to name things. Trust me: it is worth your time to get this one right <https://speakerdeck.com/jennybc/how-to-name-files>
2. Select a subset of data using the select function
3. Become familiar with how to use the pipe function
4. Learn how to use the “summarize/summarise” function to create data summaries
5. Learn the basics of “pivoting” data between long and wide formats.

### Activity 1: Getting Clean Beaches

In this activity we are going to take the syndneybeaches.csv file and clean it up using the tidyverse & janitor packages in R. Overall, our goal is to create a new “tibble” in R called “cleanbeaches” that will contain cleaned up column names of our beaches data set. The enterococci column should now be named “beachbugs” through the rename function. You should also be able to repeat the process using a pipe (%>%).

Step-by-step guide: <https://rliessydney.org/courses/ryouwithme/02-cleanitup-1/>

### Activity 2: Arrange the Beach

Now we can move on to 4 new functions: arrange, filter, group\_by and summarize. These will allow you to organize your data more and create summaries of the data. At the end of this activity you should be able to summarize the data in a way that you can tell which beach has the best/worst bacteria levels.

Step-by-step guide: <https://rliessydney.org/courses/ryouwithme/02-cleanitup-2/>

### Activity 3: Mutations

Sometimes we may want to calculate new variables from the existing data and/or change the format of the data columns. In this activity we will “mutate” the data to create new columns using separate, unite, mutate and other functions from dplyr. This should make data workflows easier because you can avoid a lot of the “point and click” edits in Excel.

Step-by-step guide: <https://rliessydney.org/courses/ryouwithme/02-cleanitup-3/>

### Activity 4: Pivot!

Sometimes the data isn't in the right format for analysis. Maybe you need to condense a few columns into one or need to take column and expand it into many. Many of you may be used to just copy and pasting data in Excel to accomplish this, but it is time consuming and error prone. Therefore, we are going to learn how to systematically do this in R.

Before we do this activity you need to create a new script in your scripts folder called “pivot\_intro.R”. In addition, you need to download two datasets to your data folder: “beachbugs\_wide.csv” and “beachbugs\_long.csv”. These can be downloaded here:

<https://github.com/jenrichmond/RLadiesSydney-blogdown/tree/master/csv>

Now, our job is to move the wide data to long and vice versa.

Step-by-step guide: <https://rliessydney.org/courses/ryouwithme/02-cleanitup-5/>

### Answer Key

The “answers” to this week’s code can be found on the repository. Please remember that there is more than 1 way to accomplish something in R. If you think you have a better solution share it with the class!

[https://github.com/arcaldwell49/STORK\\_summeR\\_2021/tree/main/Week2](https://github.com/arcaldwell49/STORK_summeR_2021/tree/main/Week2)