Mapping Our Steps

2022-06-26

# The data science life cycle

In Hadley Wickam and Garrett Grolemund’s book, [*R for Data Science*](https://r4ds.had.co.nz/index.html), they introduce a model of tools found in the typical data analysis project:



DS Life Cycle

This process consists of:

**importing** your data into R;

**tidying** your data by shaping and storing it in a consistent format (luckily, this has largely already been done for us).

Once the data is tidy and coherent, we will need to:

**transform** the data by narrowing in on observations of interest, creating new variables, and calculating summary statistics;

**visualize** it to gain new perspective and insights;

**model** the data with a well-defined research question.

And this will all be for naught if without effectively **communicating** our findings.

In the lab section for Biostat 705 and 707, we’ll be working in a similar loop. To this toolbox, we might also add **Workflow Management**, which encompasses the bulk of what we do.

The below visualization aims to provide a visualization of this as we encounter it in-action.

