Week 4:

Why do you typically need to change data from wide to long when working with data in R or other programming languages?

Arranging and wrangling data is a good practice to organize data such that it can be transformed and evaluated easily especially for visualization, or simply making it “tidy” [1]. Using a long format also works well and may times require data to be reformatted, typically in R for statistical analysis and creating plots with ggplot2 using different variable types and for visualization [2, 3]. One of the datacamp sites [4] demonstrated an example of data showing the measurement of each baby’s height and weight through age. The author gave examples of different ways to convert data into wide vs long, how to clean up step-by-step, making the data easily followed and understandable. Another key point mentioned here is the importance of deciding which columns from wide-form will be used and combined as a column in long-form before converting. Also, the author mentioned the functions in R from different packages that can be used for conversion, including the usage of melt() for wide-to-long and dcast() for long-to-wide from reshape2 package. More detailed descriptions and usage of reshape2 package can also be found on a site explained by Sean C Anderson [3] as well.

Ref:

[1] <https://argoshare.is.ed.ac.uk/healthyr_book/reshaping-data-long-vs-wide-format.html>

[2] <https://www.statology.org/long-vs-wide-data/>

[3] <https://seananderson.ca/2013/10/19/reshape/>

[4] <https://www.datacamp.com/tutorial/long-wide-data-R>