**Week 4 Discussion**

Why do you typically need to change data from wide to long when working with data in R or other programming languages? Your response should be one paragraph and you need to respond to three other student’s posts.

One large distinction between wide and long data is that in wide data there is no repetition in the first column; however, the long format will have a column for each variable. Long format is also known as “tidy data.” “Tidy data has the following attributes: 1. Each column is its own variable; 2. Each row is one observation; (and) 3. Each cell is one value”[[1]](#endnote-1) Furthermore, in some instances, if we have wide data and we add a new column but not each element of column one has a value then we have to add a 0 or NA. For example, if we consider the who data and had each year as a column with the observed cases as the value (where column one is country); then if we add another year and some countries have 0 cases that year, then 0 or NA would be added. Instead if we use long format, then country, year, and cases observed are separate columns, if the year has no data there is no reason to include a row for that year. Not having extra values will allow our data to be better analyzed by not adding superfluous zeros or NAs. Having long format also enables it to be easier to see all the variables we may want to analyze. Having a column for each variable will allow for calculations to perform on specific variables easier. The functions sum, average, count, etc. all are performed on columns; therefore, the data for an entire variable is needed within the same column.

1. [www.rforecology.com](http://www.rforecology.com)

   https://tavareshugo.github.io/ [↑](#endnote-ref-1)