Jonathan Grabowski – Bioinformatics Lab 4 Debugging Reflection

When debugging, the first problem that we encountered was the gsub function not working properly. This is because we assigned the Paper ID column from the Lab4Data dataset to “LEEs”, and we were using LEEs in the gsub function. Once we understood why gsub wasn’t working, we swapped “LEEs” for “Lab4Data$Paper.ID” and the gsub function worked because the junk characters were removed.

The next problem we ran into was deciding to use the unlist function. We were confused on whether the string split function would automatically change the vector of lists into a vector of characters. We decided to not use the unlist function, and our line of code did not work. That’s when we decided to use unlist function, which fixed our second problem.

We knew we would run into a problem when deleting rows from the dataset. We thought that if we started from the last row of the dataset and worked our way up to the first row that the row numbers would stay the same. We forgot that we had run a line of code that reversed the row numbers. When we tried to run code to find the length of the unique number of papers, R returned multiple “1’s” which we knew was incorrect. Once we realized that going bottom up in rows was causing an issue, we changed our code, and it worked again.

In the embedded “if” loop of our code, we were confused on whether to use “[each]” or “Lab4Data$Paper.ID[each]”. We did not fully understand what “each” was supposed to be, but after consulting our lines of code with Dr. Hansen, she told us that “each” was supposed to represent a number.

After lots of trial and error, we were able to get the correct number of rows that contained more than one Paper ID (83).