

# My Growth as a Data Scientist

Caleb Jensen

At the beginning of this quarter I would call myself a poor data scientist. I was capable of some modeling with from previous classes and looking at certain summaries but my knowledge of data cleaning and using pipelines to accomplish that task was poor. In previous classes I would work with pre-cleaned data to create predictions or look at correlations and causations. Getting to this step takes a significant amount of data cleaning so a primary goal for myself in this course was and is to improve in this area. I believe I have made large steps forward in my ability to filter, select, and pivot data to be in a format I desire. For instance, just looking at lab 4 I can see large strides have been made. Without too much difficulty I was able to select the columns I wanted, group it properly, and use both filter statements and mutating joins to put two related data set together. Six weeks ago, had I been asked to do such a task I would have been utterly clueless. An additional area I have had to work on is with strings. I have always had trouble working with character data types and was clueless about regular expressions. Working through practice activity 5.2 has made me much more confident. I can iterate changes to strings across multiple observations and specify characters to change or locations to change them in. Lab 8 was also helpful in making me comfortable with string manipulation and the use of regular expressions. Doing that previously would have terrified me but after that my confidence is much higher. Previously changing a city name or eliminating white space would not have been possible for me, but now I feel quite confident that I can eliminate unwanted characters or series of characters, white space, and a variety of other tasks. I also really like the ability to output professional looking data tables that can be searched and filtered by the end user. That ability makes my output report look much more professional. I also feel much more confident in writing functions in R that can eliminate repetitions in code and make it easier to understand. This was an area I had little experience in entering the class and am very happy with my improvement in. There are many other areas I have grown as a data scientist, but I feel my ability to clean data, mutate it to be in a usable form, select or filter to the data I am interested in, pivot it to the orientation I want and work with strings to make them simple to use. I hope to continue improving in these skills and working with raw data to move it to the point where it can be visualized or modeled.