Qualtrics, R and Data Visualization: the importance of tidyverse and ggplot2 collaboration to portray meaningful results

As a requirement for my Ed. D. coursework at the University of North Carolina Wilmington (UNCW), I completed my internship during the Spring 2020 semester with Aligned Impact, an evaluation firm based in Greensboro, North Carolina. The internship led me to work on an annual report that the evaluation firm produces for the stakeholders. I was pleasantly surprised to find that the report was generated using R, while the underlying data was collected using the Qualtrics system. I had prior experience with the Qualtrics system from my coursework at UNCW, but this internship was my first opportunity to work with R. The internship gave me the opportunity to learn how to use R, do data importing, data cleanup and formatting, data visualization, creating the report using R Markdown, and working on the final report presentation.

The data cleanup and formatting aspects of the internship revolved around how the Qualtrics data first needed to be imported into R and prepared for the desired statistical analysis and graphing functions of R. In this respect, I learned tidyverse, one of the popular R set of packages. In the beginning of the internship, I had a weekly meeting with the research scientist of my internship team discussing various aspects of cleaning up the data: for example, the data may have some missing information, and/or the datatype of the imported data may need to be changed. Additionally, the R functions that we used required the data to be in a certain format, for example we needed to combine certain columns of data into one column.

On the other hand, the data visualization aspect of R introduced me to ggplot2. This led me to discover the grammar of graphics, and the aesthetics of visualization in ggplot2. In the second half of the internship, my weekly meetings with the evaluation team's research scientist

focused on how to use different ggplot2 functions. Although understanding how R functions work has a steep learning curve, once I understood the concept of inputs for the functions, it became easy to understand how the other similar functions would also work. So, when I started graphing, overtime I became comfortable creating several different kinds of graphs. While it is important to understand how the myriad of ggplot2 functions work, it is even more important to decide which kind of visualization is best suited for the results that matter the most to the stakeholders. We considered many different graphs before ultimately settling for the Violin and the Density graphs in ggplot2 because these were the best fits to display our data equitably.

Once the data is imported and formatted correctly, and the correct format for displaying and graphing data has been selected, the focus shifts to creating the final report for the stakeholders. Instead of doing statistical analysis in one software package, data visualization in another, and then combining it with text in yet another software package, all of this can be done under the R umbrella using R Markdown scripting. The internship also provided me with the experience of creating a professional report using R Markdown. Our final report was in PDF format, and included text, data displayed in tabular format with colored highlighting of important aspects of data, and data-visualizations of different kinds. Additionally, I also worked on editing a presentation for the main stakeholders' group.

The internship was a perfect opportunity for me to experiencing what I had learned in last year's AEA 2019 Conference: that many professional evaluators are using R for their statistical analysis and data visualization. After this internship, I have become convinced of why an open-source and free software package like R is a more relevant tool set for graduate students pursuing evaluation as their career.