**Week 3: Visualizing data using R**

Wonjin Jeong

I found that visualizing data using R has numerous convenient functions. This week, through the reading materials and data simulations in R, I learned how to create tables, histograms, and scatterplots via a command line interface. I liked that this method allows me to see what kinds of stories can be told visually even before starting data analysis. When I used other types of graphical user interface (e.g., MS PowerPoint, Adobe Photoshop), on the other hand, some kinds of data analysis and preliminary results were necessary to generate such graphical data. I used to make tables and charts after finishing my data analysis work for the intended purposes. As a disadvantage of using a command line interface compared to using a graphical user interface, I began to realize that it would be almost impossible for me to remember this syntax and algorithms (and this made me wonder if programmers would memorize all of them), which will make this process of learning R languages more challenging and time-consuming.

Ensuring reproducibility of data analysis is an important issue in collaborative scientific research to promote transparency of research and data disclosure. It will also help other researchers to test the study findings and provide new insights into the study. To enable this reproducibility, we can provide the R code that was used for our analysis work. R and R studio are open-source statistical software, and I believe it is easier to have access to them compared to other costly statistical software options such as IBM SPSS, SAS, and Mplus.

I found the most difficulty in adjusting to the complicated language of R and R Studio. Before this class, I did not have experience directly using R Markdown and functions associated with data visualization. While I have been using R for the past two years, I became more aware of my limited experience with R in terms of visualizing data. Another thing I noticed is that, in my research, I have always used primary data that is collected by our research team and partner agencies directly from the targeting populations through surveys. Using secondary data appears to be somewhat intimidating and inspiring to me considering the breadth and depth of data. In the class, it was interesting to see how R allows us to answer specific questions. Learning these “new” R languages is an exciting challenge for me, and I am thrilled to apply the analytic skills learned in this class to my own study.