A4: Common Analysis
Reflection
Preston Stringham
November 4, 2021

One of the key characteristics of this assignment is the collaborative process which we all participated in(in one way or another). In particular, learning about how to represent data from peers was an excellent exercise that I think can translate well to real-world research and work. From my personal experience with this project, I found myself turning to my peers with questions about visualizations techniques, statistical refreshers, and coding questions. There were a few students in particular that helped me with my visualization that I will cite in this reflection. While I cannot attribute the code specifically, the idea for using a confidence interval for my visualization was proposed to me by my peers Apoorv Sharma and Maya Patwardhan. Their help was extremely useful as I was stuck wondering how I might show "significant" changes in the rate of infection, but using a more simple and statistical approach proved to be beneficial for my visualization as I could easily point to days which did have significant change. A statistical refresher to create the confidence interval was given to me by Stacey Wheeler. With regards to more of the technical tools used, I utilized matplotlib to create my scatter plot visualization as well as Pandas for data cleaning and preprocessing.

One concept that I really learned during this exercise was the idea of misinformation. Particularly, when phrasing a problem from a single particular viewpoint seems to completely disregard all other influential variables. For example, by looking at my visualization alone, it is easy to say that masks may have helped ease the first wave of COVID in Middlesex County. However, there are likely *many* more factors that could be better attributed to the attenuation of the change of the infection rate during this time such as lockdowns, jobs and classes moving to a remote format, etc. Simply saying masks were the sole cause of the attenuation feels untrue and completely disregards so many other important variables. At the same time there were other parts of my visualization that point to the fact that masks may not have helped, especially during times of significant positive change in the rate of infection.

Another lesson I learned during this exercise is the influence of collaboration on our work. I think this clearly brings out more of the human-centeredness of data science. Collaboration is interesting because it can completely change the outcome of a project, for better or for worse. Being able to openly communicate about our findings with our peers completely changed what I thought my final product would look like. Looking through our Slack channels and seeing others' work made me think, "Maybe I should also incorporate that [visualization technique/tool] into my own visualization." I believe that when working alone, we seem to be stuck in our usual routine of creating visualizations in a specific way or using techniques that we are familiar with, but collaboration brings out a different feeling within us. It allows us to see

ideas from a completely different perspective and I think that this is a critical part of learning from others. I certainly appreciated the collaborative nature of this assignment for this reason.

In general, I thought that this assignment certainly ties into some of the ideas we have learned previously. This certainly reminds me of the work of Brian Keegan who used his own data science skills to question the validity of statements made in the press. Certainly, the effectiveness of masks has been brought into question substantially over the past two years due to potential misinformation. It is easy to find data that fits a narrative, despite its complete disregard for other influential variables that are at work. It is important to do our best as data scientists to provide people with the truth of the data to the best of our abilities.