* Statistical/Hypothetical Question
  + What are the relationships between long-term nursing home residents reporting that they are experiencing increased need for help in daily activities, and short-term nursing home residents reporting that they are experiencing increased moderate or severe pain, in states with more than 500 nursing homes?
* Outcome of your EDA
  + The outcome was pretty unclear. This being the first time I’ve worked with this data, as well as my ongoing lack of fluency on these topics, I can see that I can get the tests to work but I suspect that my data preparation is not necessarily adequate to the task. I was however able to observe key differences and measure critical statistics in the area I was trying to understand.
* What do you feel was missed during the analysis?
  + I definitely lost the forest for the trees. I needed to spend a lot of time just learning the basics of the techniques, and when you’re just learning the basics, you just can’t see the whole picture. For example, when I got to the part of my analysis where I knew the resident state variable was critical, I learned that in order to do regression on categorical data, I had to transform the variables into something that could be represented numerically. I learned about a technique called label encoding that represented each state with a unique value in a new column, but this was inadequate to the task because Python took those values to mean a particular order, which is not accurate. One Hot encoding would have been a lot better for that task, where each state would have been a simple Boolean of 0 or 1, rather than numbers 1 – 19.
* Were there any variables you felt could have helped in the analysis?
  + I think having state populations overall would have been helpful for deeper demographic analysis, as well as population details on people who reside in nursing homes generally.
* Were there any assumptions made you felt were incorrect?
  + Certainly. I assumed that I could use pandas and matplotlib for the whole thing and actually found better techniques online using scipy and seaborn. I think a critical takeaway there is to use the right tools for the right jobs, and that there’s almost always a better way to do things. In particular I ran into a lot of information on doing pretty much exactly what I did but in an iterative fashion, which is SO much more elegant. Something to strive for!
* What challenges did you face, what did you not fully understand?
  + I still struggle with my statistics fundamentals. I just don’t think I “get it.” I know it’ll take time but still. I need to spent way, way more time grasping just what the statistical tests are actually doing and what they’re actually conveying.
* Submit a link to your repository to the assignment link during the final week of class.
  + Done!