

The purpose of my EDA was to find how many hours of sleep affect the different stages of sleep. There are 5 variables that I used in my analysis. I used sleep duration, REM sleep percentage, light sleep percentage, deep sleep percentage, and lastly sleep efficiency. I was curious to know how little sleep I could get and still get a good night's rest. I wanted to find how many hours of sleep were associated with high percentages of light sleep, deep sleep, sleep efficiency, and REM. The outcome of my EDA was that none of these variables are affected by how much sleep you get. I found that deep sleep percentage was a hit or miss. Even if you slept 5 hours you can get 70% deep sleep whereas you can sleep 10 hours and get only 20% deep sleep. The results were sporadic and chaotic. I felt that I was missing more variables in my analysis. I used only 5 and more variables were available to use, such as alcohol, and caffeine consumption. These two variables were most likely the cause behind my chaotic results. I assumed that I had no outliers but after looking at my results there may have been an error in the beginning of my analysis. I still stand by the notion that the outliers were valid because they are natural variations of each variable. I was faced with lots of challenges this project. One major challenge that I faced was interpreting the correlations, the hypothesis test, and the least squares model. These 3 aspects of the project were a huge challenge and I still do not know if I analyzed or conducted them properly. In class all of these analysis seemed simple because the book made it simple and made it understandable in the context of the thinkstats book. Overall this term project was difficult and had a learning curve but I appreciated going through the book one more time.