**Gautam: CLEANING**

I initially opened up the GDP Data Set to find that I needed to do some crucial data cleaning which would allow us to use the data in our charting. There was an extra column that needed to be removed because it was of no use to us. This was done using the DEL function. A DROPNA function was also performed on the CSV in case there was incomplete data that needed to be removed. Next I used .DROP and .ILOC functions to remove unwanted rows that would have skewed our final readings of the graphs. I then outputted the CSV with a new name.

For the Chronic Illness dataset, which was far larger and had much more data than the GDP data I started off by importing the CSV into a data frame and debugging the low memory error I received using the LOW\_MEMORY=FALSE Boolean. Next, I deleted unwanted columns using the DEL function and followed that up with a DROPNA function that cleaned up the empty data and corresponding rows. I then used the ASTYPE function to convert data in certain columns to floats so that it would make it easier to chart. I outputted the csv with a new name so that I could pass the data along to Ed who then would clean the Latitude and Longitude columns and finish of the rest of the cleaning.

**ED: CLEANING**

To clean the data, a list comprehension was used. This made it easy to isolate data points like disability rates, overall vs. gender data, etc. Basically the list comprehension filtered out all of the rows not matching the criteria given in the conditional. So this made it easy to keep only the percentages as well.

Another challenge with cleaning the data was the use of a regular expression. Our data had extra parentheses in the latitude and longitude coordinates. So it became necessary to filter these characters out using a simple regular expression, so that Gautam could use them in the Google Maps visualization part of the presentation.

**Gautam: GMAPS**

After Ed handed me the cleaned Disease data as well as the cleaned GDP data, I started to put together heat maps using GMAPS for both sets of data. I imported all my dependencies that I needed to use for the mapping and added my google API key. I then created my locations variable which included the Latitude and Longitude columns from my data files. Next, I created the rating variable. For the Disease data rating I used the 2017 values from the DATAVALUE column and for the GDP Data rating I used the 2017 data column to compare the most recent year of general data we had. Then I plotted the figure and added a heat layer to the figure. Finally, I saved the final figure as a PNG.

**ED: DISABILITY GRAPH**

The hypothesis that richer states are healthier was tested. According to our bar chart, the states with the highest disability rates were Arkansas, Mississippi, Oklahoma, and West Virginia. The states with the lowest disability rates were Connecticut, Delaware, Minnesota, and Wisconsin. This should line up with the GDP data for the richest states vs. poorest states by GDP if our hypothesis is true.

So to do this we take a look at the GDP numbers and find that the poorest states by GDP are Arkansas, Idaho, Mississippi, South Carolina, and West Virginia. This means that 3 of the four poorest states match up with 4 of the states with highest disability rates. Also, the richest states by GDP are Alaska, Washington D.C., New York, and North Dakota. So the converse is not necessarily true, that the richest states are healthiest. From the data, we have found that there is a relationship between GDP and health of a state, though not for both ends of the spectrum.