I have implemented the stacked bar chart of out project.

We have decided to visualize statistics of various diseases in the US and their distribution over many states and over sex.

I have visualized a stacked bar chart that shows the distribution of cancer patients in each state and the number of male and female cancer patients in each state.

I got the data from the CDC website and the link is available in our presentation.

The data was initially in a txt file that I converted into a csv file using Microsoft excel.

Next, the csv file generated was not suitable for a stacked bar chart visualization. It has different columns for the sex and count. What I needed was for there to be a single column for male count in each state and female count in each state.

To do that I modified the dataset using python and it’s NumPy library. Firstly, I dropped the Notes, State code and sex code column from the data table.

Next, I dropped all the na values from the table.

To ensure I don’t encounter problems with spaces in the state’s names, I replaced all spaces in the names of the states with the – symbol.

Finally, I had to get all the integrate the counts in a single column for both male and females. To do so I created a pivot table with the state’s column acting as the index and there were now 2 columns, namely Male and Female counts. Now I had to restore the states column back to the status of column and for that I used the restoreIndex function of NumPy and finally I was able to get the desired table.

Next to create the visualization, I used the general steps of passing the data, renaming the columns, grouping the desired columns, making scales, and then creating rectangles from those groups.

Then I added the axes to the chart and created a transition for the chart. I am working on creating a mouseover event but encountered problems with it and will hopefully implement it in the final project.