The work I did this summer mainly revolved around linear regression, coding linear regression in python, and then applying it to different datasets. The work I did was all remote, and I did not get the opportunity to go to a lab. Linear regression is a statistical technique that involves finding a "line of best fit (simple y=mx+c)" for any given scatterplot of data. This means finding a predicted y-value for each x-value that is closest to the observed y-value for the same x-values. The method I used was called the Least squares method, where the vertical distance between the predicted values for y (on the line of best fit) and the observed values of y (on the scatterplot) is squared and then minimized to find the line of best fit, this is called the square error. The slope and the y-intercept of the line of best fit are found when this square error is minimized. Minimization is done by taking the derivative of the sum of all square errors, both with respect to the slope and y-intercept, and then setting it equal to zero. The derivative gives the slope of a line tangent to any point on a graph, and when this slope is equal to zero (horizontal line), it means there is either a minimum or maximum at that point. I was able to code the Least Squares method into python, a programming language, where I was then able to input any scatterplot of data and it could spit out the slope, y-intercept, and the graph of the line of best fit. To do this, I utilized Numpy, which is a library in python that deals with statistics, and matplotlib, which is a plotting library used for making graphs.

Once I had the one-dimensional linear regression solution in python, we could then input sample data into the code to find the line of best fit. One of these data sets plotted the activation voltages of different colored LEDs against their respective reciprocal wavelengths, and we were able to get a rough value for Planck's constant of 7.16×10^{-34} J/s close to the real value of 6.63×10^{-34} J/s. The specifics of this experiment are not too important, just that the slope of a line can be useful when considering scatterplots of data, as mentioned above.

I also coded in the multi-dimensional solution for linear regression, similar to the one-dimensional solution above, except it takes into account two independent variables instead of one. I used a dataset of age and weight (two independent variables) against blood to test the code, and it worked. The general trend of the data showed that as age and weight increased, so did blood pressure. Overall, even though I didn't get to work in a lab this summer, I gained some

valuable statistics and coding knowledge which I hope to apply in any future research opportunities.

Since I was unable to work in a lab this summer, I worked with my math teacher Dr. K, before I knew he was in the program. I most likely will not be working with him again this summer. I completed the work from last summer, and will not be continuing it next summer. Right now, I am taking the Introduction to Neuroscience course as Burke, and I will hopefully gain a research opportunity there within the next few weeks. One topic that really interested me, that was a lecture in the course, was vision. There is ongoing research about simulating human vision by using neural networks, which particularly interests me because of my previous work, but I will need to look into that in the coming weeks as well as the lab techniques and statistical analysis that comes along with it. At Burke, they are thinking of doing a science writing course soon, where you get a mentor and you attend lab meetings and go into depth on their research, which I may enroll in to further my connections and get into a lab by next summer. Overall, my mentor situation is highly variable on how this science writing course goes as well as how COVID and lab openings happen.

Working with Dr. K over the summer was a great experience, and even though it was completely remote, I feel as though nothing was lost in terms of the quality of the work done. We got on a zoom call a couple times a week - largely dependent upon how independent the work I was currently doing was - and discussed the current objective and future goals. As we had written up a summer timeline before we started the work, we had some pretty clear cut goals, but what I loved most about this work was its flexibility. If I wanted to go into a bit more depth on a certain topic, an example being the proof of multi-dimensional regression, Dr. K was able to make the time to do it. We weren't bound by typical time constraints, and it was the depth we were able to go into these topics that really made the experience unforgettable. Obviously, none of this would have been possible without Dr. K's commitment and expertise, and I'm really thankful for that. Overall, this combination of depth and expertise that I was able to garner from this summer really cemented it as an incredible learning experience, and I'm looking forward to applying these topics in the near future.