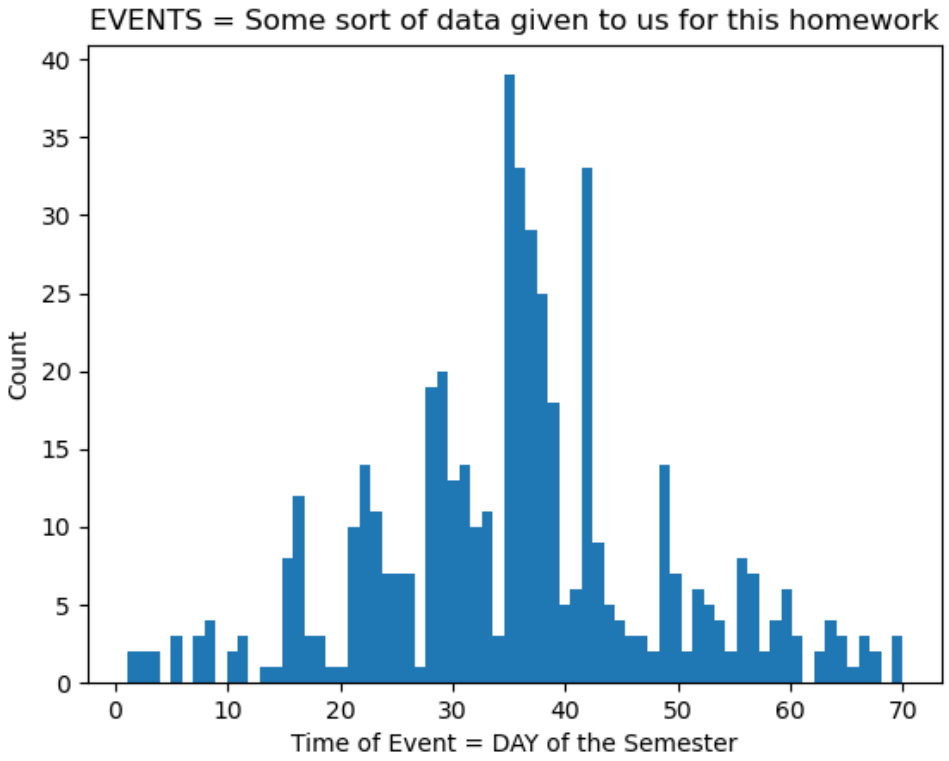
Robert Alemany

For this homework, we were given a csv file of data that represented a frequency table of hours in a semester, and the frequency in which an event happened. Through the course of this homework, we were expected to write a program to create a histogram of the data, with bin intervals of days rather than hours. Then the program had to create a Parzen Density Estimation using a Gaussean filter (really an array used to simulate one), to effectively give a better estimate on continuous intervals.

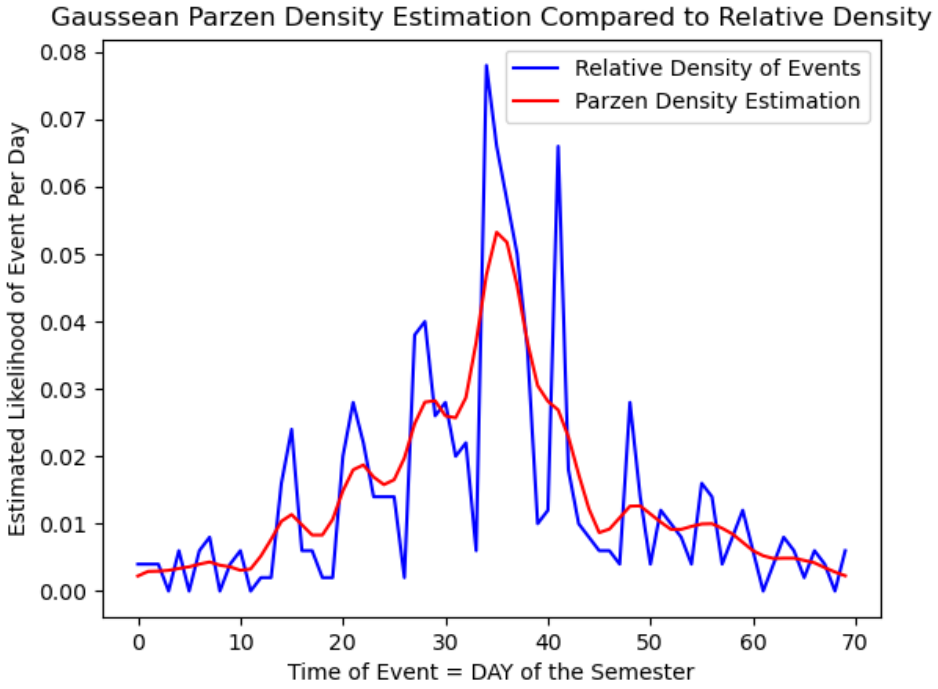
To begin, the program was done in python, using Pandas to read the data, and MatPlotLib to plot and graph the data. The results of the histogram in day intervals are as follows:



With the days going from day 1 to 70, we can see that there seems to be a clear mode at around day 35 being shown here, with nearly 40 events happening on that day. But even then it could be argued that the data might show more than one mode, but around the day 35 mark is the clearest grouping.

However the data does not make it easy to estimate the likelihood of an event at any given day. So through a Parzen Density Estimation, the data will be estimated to a continuous value rather than the discrete values that resulted from our frequency table.

Through normalizing our data, then applying a Gaussean filter on in (provided to us in the form of a array over 7 days), and comparing it to the relative density of the events over day intervals, we receive the following results:



Through our Parzen Density Estimation, we can much more clearly visualize the likelihood of an event happening on an interval of days.

Conclusions:

For this homework assignment, we were expected to take a set of data, and be able to extrapolate meaning from it using techniques learned in class, while at the same time getting familiar with the tools we will be using throughout this course to analyze and model the data we have.

While this small homework might seem simple, by just giving us the instructions on how we should model our data, and our knowledge taken from this class, I've been able to take simple data collected, and turn it to a meaningful model, one with which information and conclusions can be gotten.

But more importantly, I think this assignment demonstrates that with the knowledge we learn in class, we are capable of using the tools given to us, with relatively little guidance, giving us independence and confidence in our abilities for future data analysis tasks tasks.