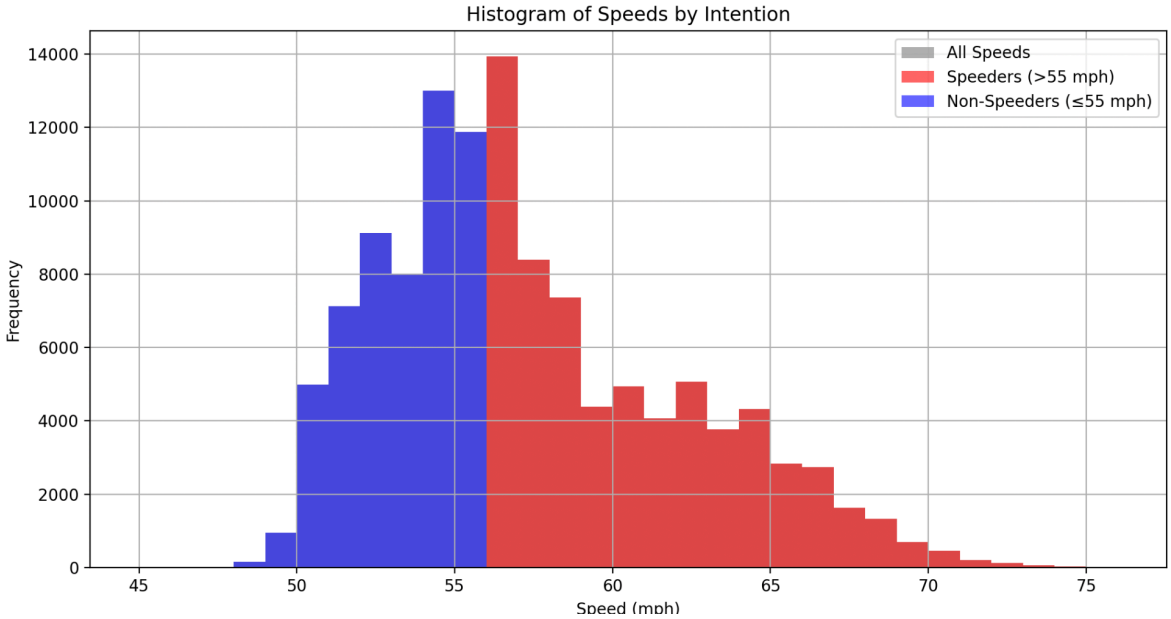
HW_01_ Anna Kurchenko, CSCI 420

D. What was the general trend as the number of elements in each vector went up, what happened to the amount of data within one standard deviation of the origin?
In other words, as the number of attributes used to measure a data point increases, what happens to the density of the records? (2)

> For part A (assuming this is only talking about part A) the fraction within 1 S.D. tended to exponentially decrease as the number of elements went up.
> This means that the density of records decreases the more attributes are used to measure a data point.

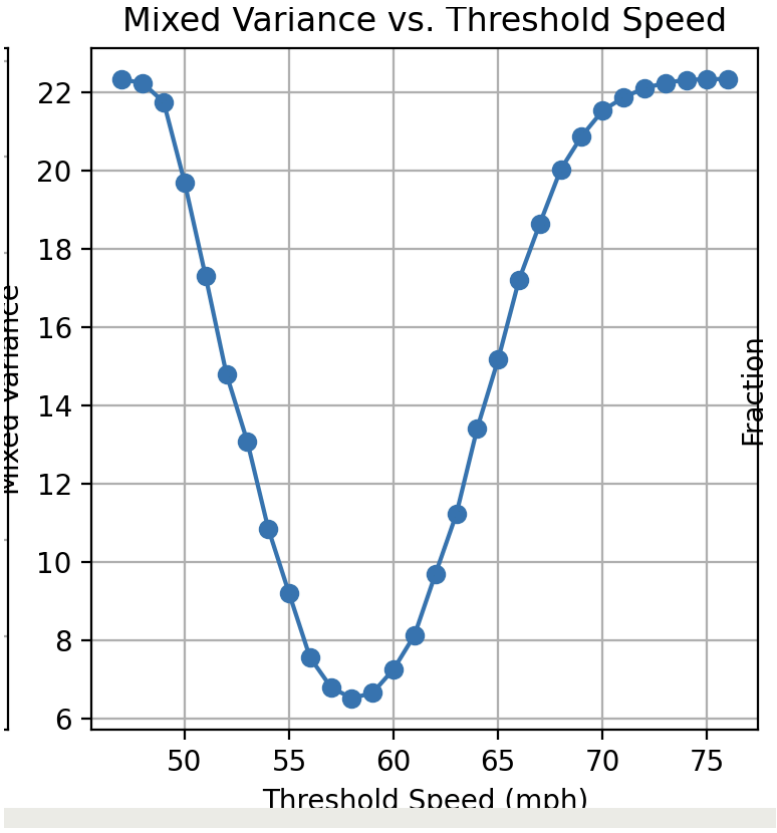
E. Show a bar graph of the entire histogram of speeds, by intention. Remember that the speed limit is 55 mph.
How might we describe this data? Is it a mixture model? What kind of mixture model?
What do you notice about it? What is odd?
Speculate about why various lumps are the way they are? (2)



> This data looks like a normal distribution.
> I googled what a mixture model was - In statistics, a mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belong -

I think this data could be seen as a mixture model since you can extrapolate the distribution of speeds. There are also multiple components in the model - speeds and intent. If I had to categorize the mixture model I would say it's a finite mixture model since it has a fixed number of components.
I notice in this graph that there is an odd number of non-speeders. I would expect most people to go slightly above the speed limit. The even distribution is unexpected. Also, more people in this graph don't speed than those who do. There are more people who go 5 under than who go 5 over. There is also a 'lump' of data for people who do speed going over 60, this makes sense though.
I speculate this is seeking to mimic normal traffic patterns as I think more people speed than those who don't.

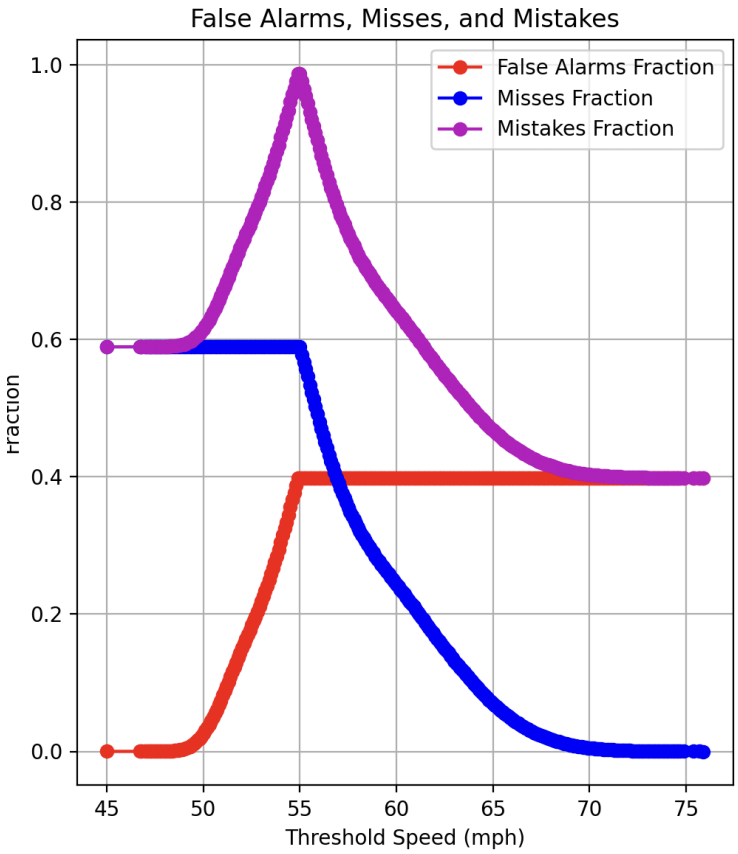
F. In your PDF, show your graph of Mixed Variance versus speed. (2)
This could be used to break the drivers into two groups.
What speed would you use to split the drivers into two groups?



I would use a speed of 55 since it seems to be in the middle of the data, and reflects the speed limit which naturally splits the group into categories of speeders and non-speeders.

G. In your PDF, show your graph of the numbers of false alarms, misses, and mistakes as a function of speed. You want to minimize the number of mistakes.

What one-rule would you use to decide if a driver was trying to speed? (2)



A driver is trying to speed if their fraction is above 0.5.

H. Conclusion: Write up what you learned here using at least three paragraphs. (2)
What did you discover?
Was anything unusual?
What was surprising?
Was there anything particularly challenging? Did anything go wrong?
Provide strong evidence of learning.
Write a conclusion that describes what you learned in this homework.
Points are taken off for writing with bullet points or checkmarks.

I learned how to use the matplot lib graphing library better. I was able to really nicely lay out my findings in graphs and all at the same time in the same window, also modulating if I wanted them to be vertically or horizontally aligned. I discovered in my investigation for this assignment that data mining is kind of a long process and is super meticulous.

No findings were really unusual or surprising to me. But what was challenging was honestly figuring out the matplot lib stuff and plotting things nicely, I don't have a lot of experience with it so it was a nice learning experience. Also before this I had no idea what random gaussian variables were or why they could be useful in code so this was a cool aspect. I had to do some digging online for this as well.

My strong evidence of learning is that I completed this assignment, and I feel as though I have a good flow and organization of code. Everything is modularized and I feel ready to use this in future homeworks as was hinted at by the professor. I also commented throughout the program to make sure I have good documentation for future reference.