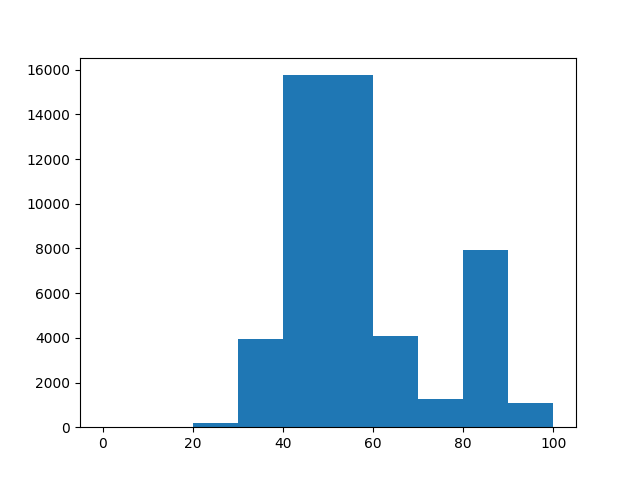
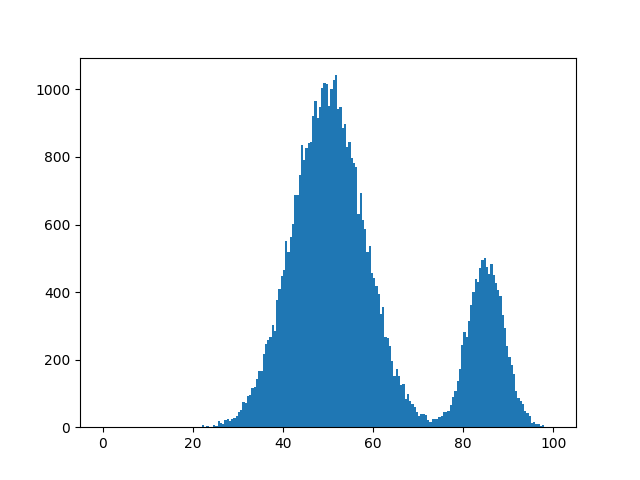
1)

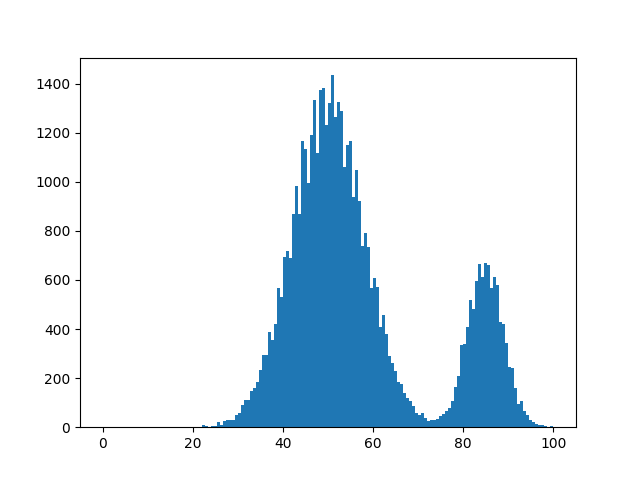


The histogram shows how the data is distributed between numbers 0 and 100. We can see that most the elements of the data are concentrated between numbers 40 and 60. Even though we can see where most of the data is concentrated, increasing the number of bins will help increase the visual clarity of the data.

2) 

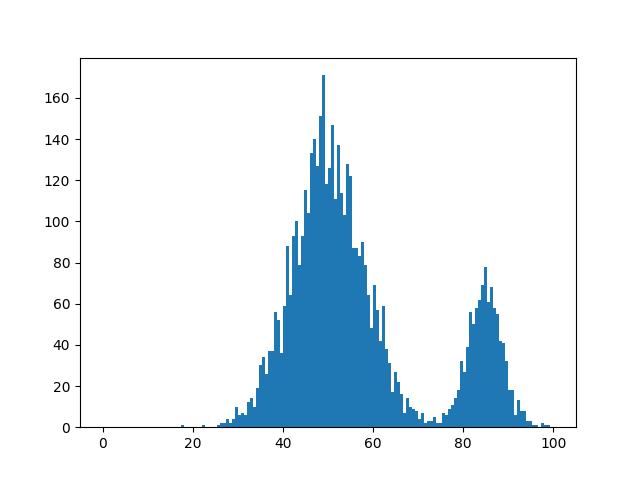
The plot above shows the visual details of the data better than plot 1. We can see that most of the data points are concentrated between numbers 55-60.

3)

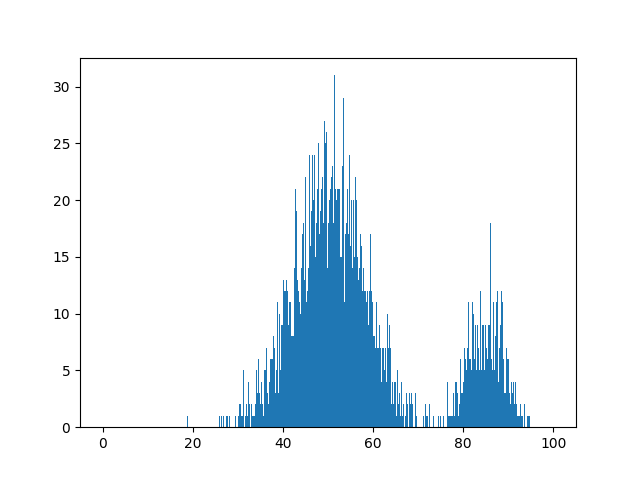


Using the cross-validation formula, I estimated that 150 bins would give the best visual details of the data. The cross-validation score was approximately the least for this score. We can clearly see that the most amount of data is for the number 52 approximately.

4)

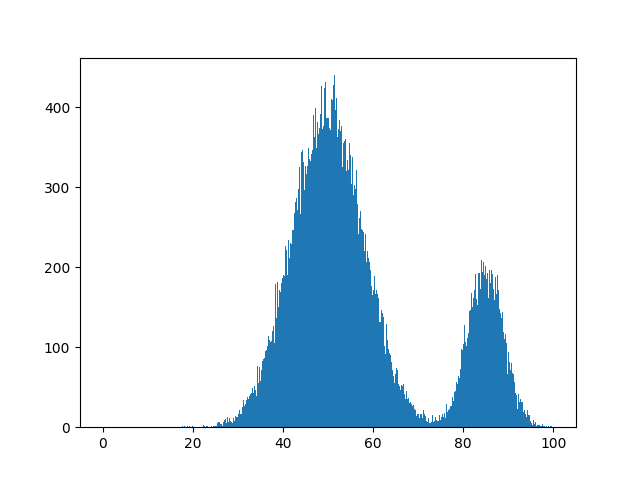


The frequency of the 10% of total data is approximately 1/10 of the normal frequency which is around 160 and the normal highest frequency is 1400. The rest of the data plot looks similar with most of the data concentrated between number 40 and 60.



When 1000 bins are used the graph becomes more unclear and we can see that a lot of the bins are simply empty. The highest frequency still remains around 40-60.

5) There could be instances in which we need different bin widths over uniform bin width in histograms. For example, if there is data which is very sparse in some places and very dense in the rest, different bin widths would be useful.

6) 

The optimal bin number is 500 for this data by using the formula of lecture 2.