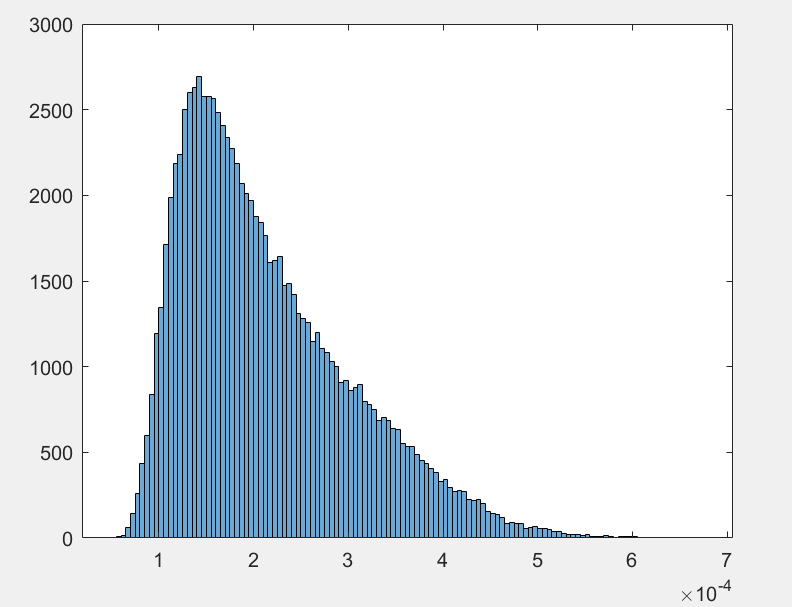
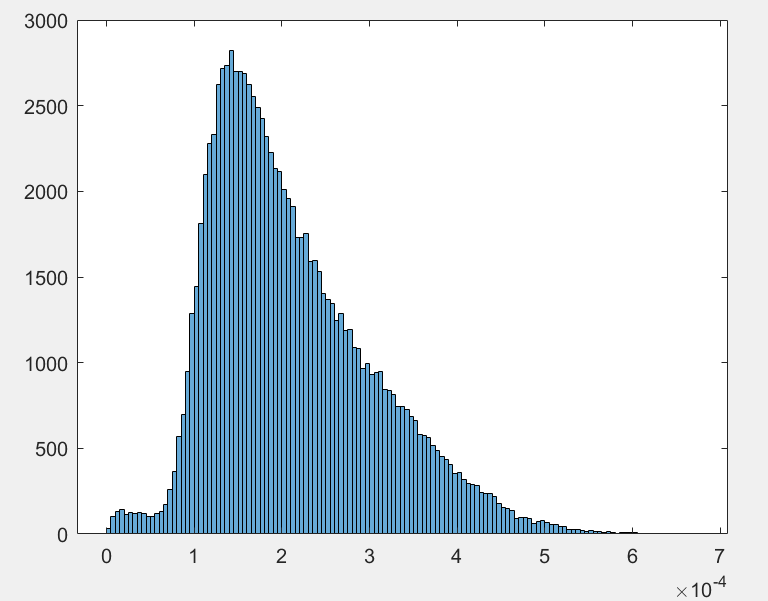
Lab 7

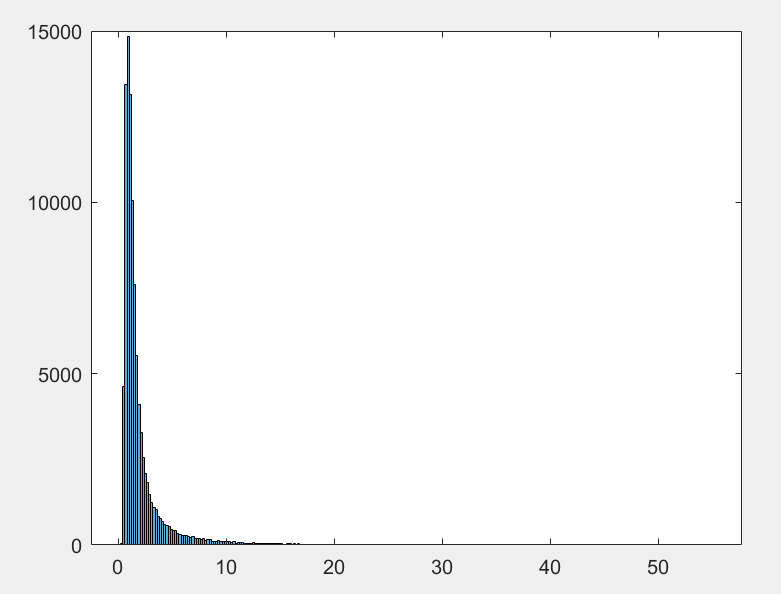
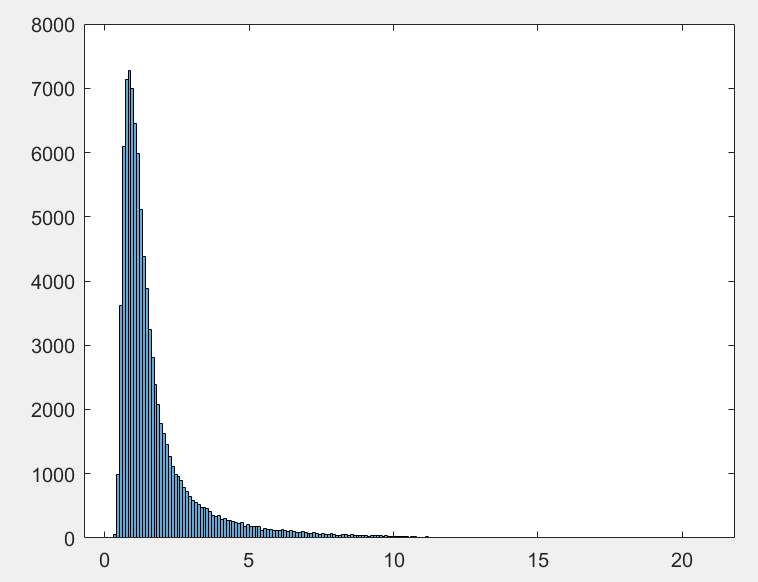
1. First we start by comparing our findings of the Nhiggs/sqrt(NQCD) with the given number of

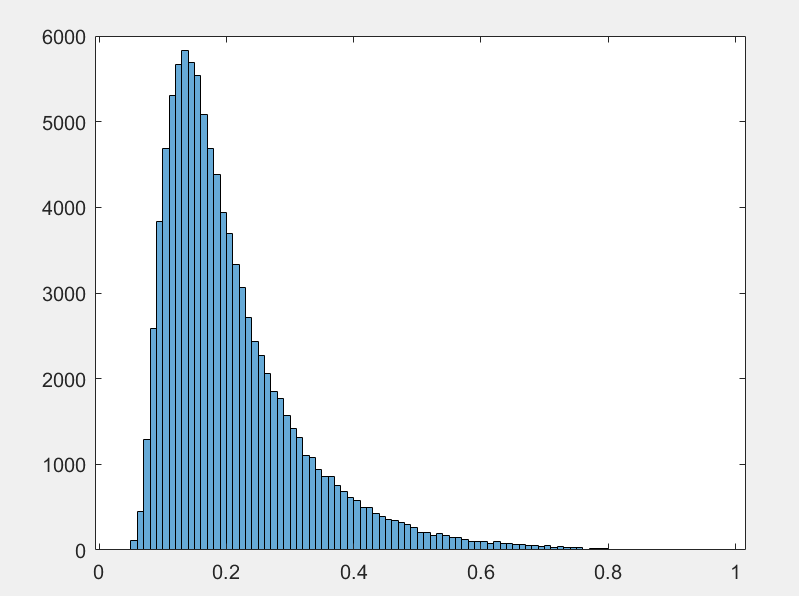
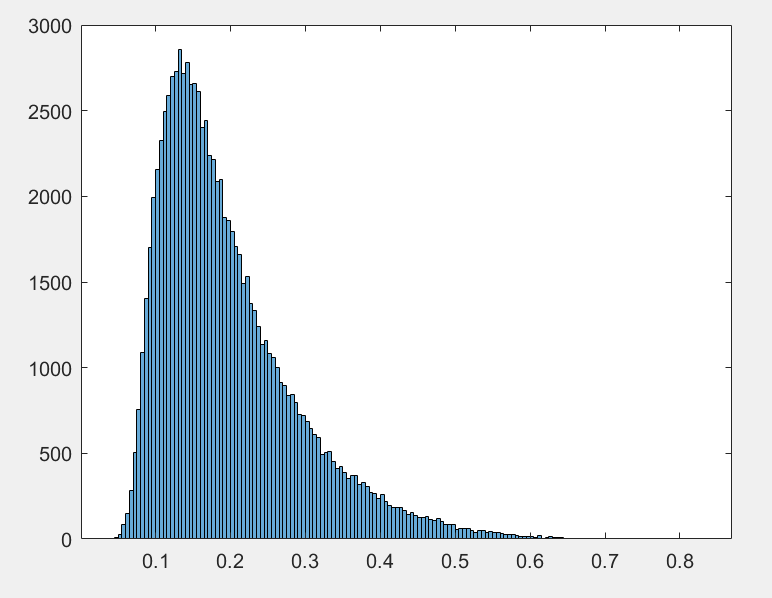
Nhiggs = 50 and NQCD= 2000 which gives us a ratio = 1.12

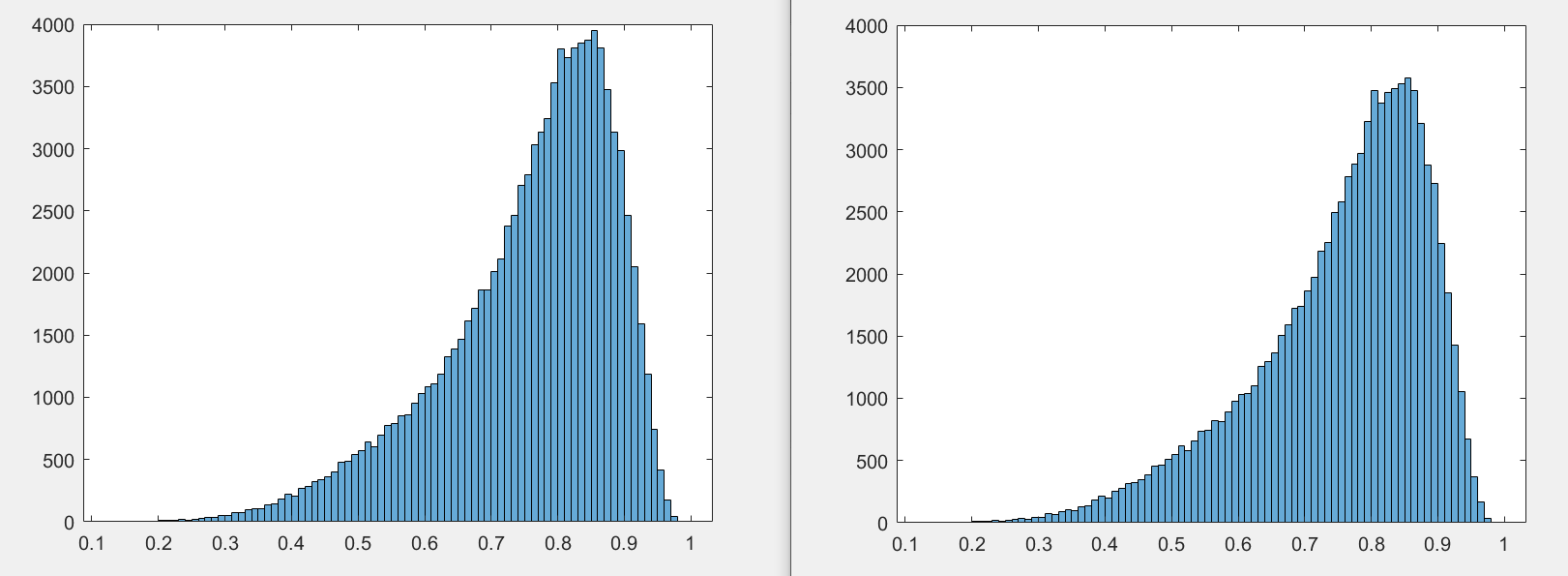
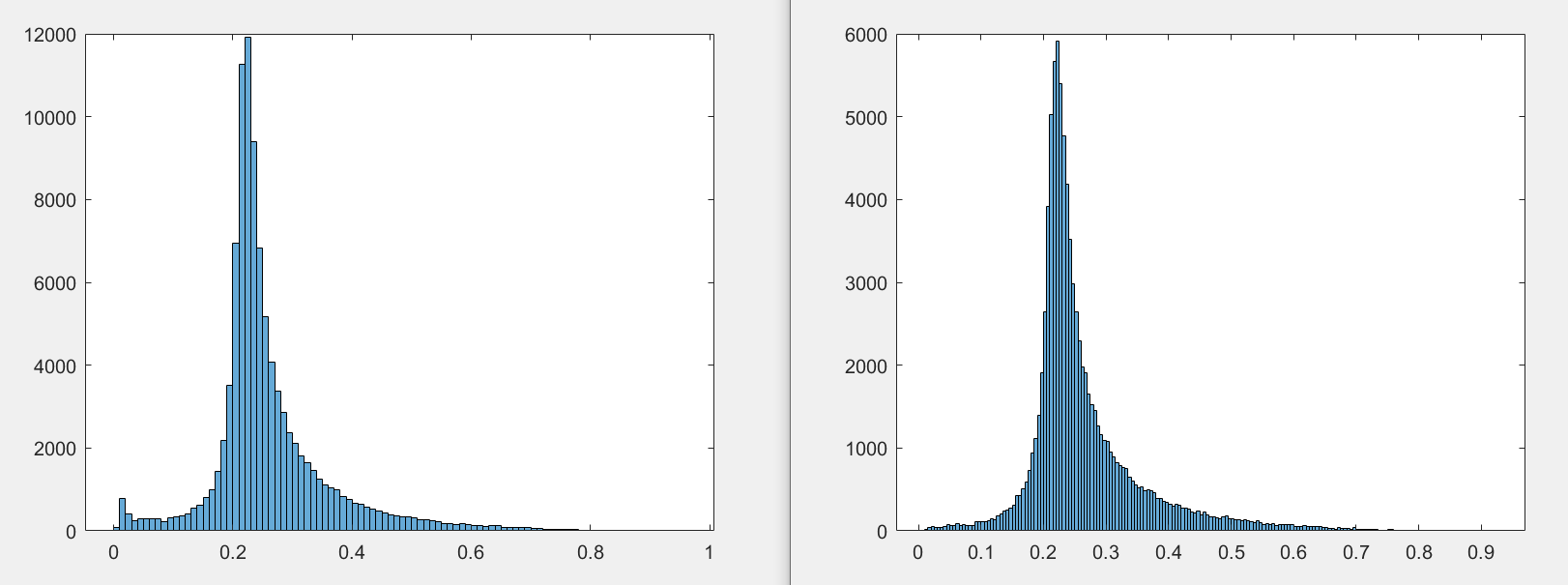
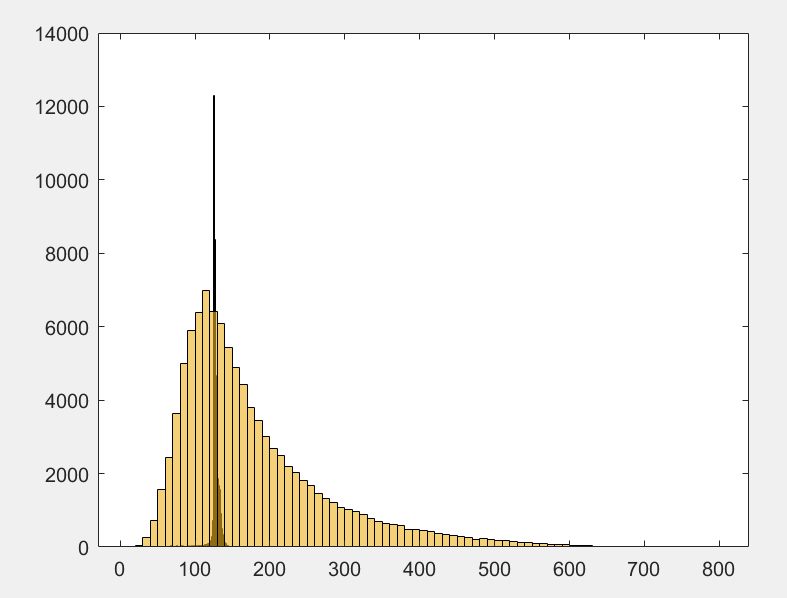
In order to find Nhiggs and NQCD from the given data we use Poisson statistics to normalize the mass data of both Higgs and QCD data. As the ratio Nhiggs/sqrt(NQCD) is the significance of the data set. By our significance calculation we get a significance of 1.12 which is equal to the expected yield value. This means that our data is equally significant and if we normalize it to Nhiggs= 50 counts and NQCD= 2000 counts, will still be the same data. This tells us the minimum number of counts one needs to take data for in order to get a significant data.

1. For different mass cuts I choose sigma values from 1-sigma to 5-sigma. In order to calculate significance after the mass cuts. I use the formula (Nhiggs/105)x50/sqrt(2000x(NQCD/10^5)). As one can intuitively guess that the significance will be highest for 1-sigma mass cut, as it will be the one with least number of NQCD counts. Which can be seen, as significance1 which is the significance for 1-sigma is highest with value = 3.14. (with Nhiggs= 92137 and NQCD=10788)
2. Now for plotting all the 14 features of the data set without mass cut and then with the mass cut which gives us the highest significance i.e. with when we make 1-sigma cuts. Upon observing one can see that other than the mass feature, features like :
   1. Ee3 one can clearly see that due to the mass cut the bump at the start of the graph before the mass cut is removed. Thus helping us get a better Poisson graph for the feature ee3 (left image before mass cut; right image after mass cut).



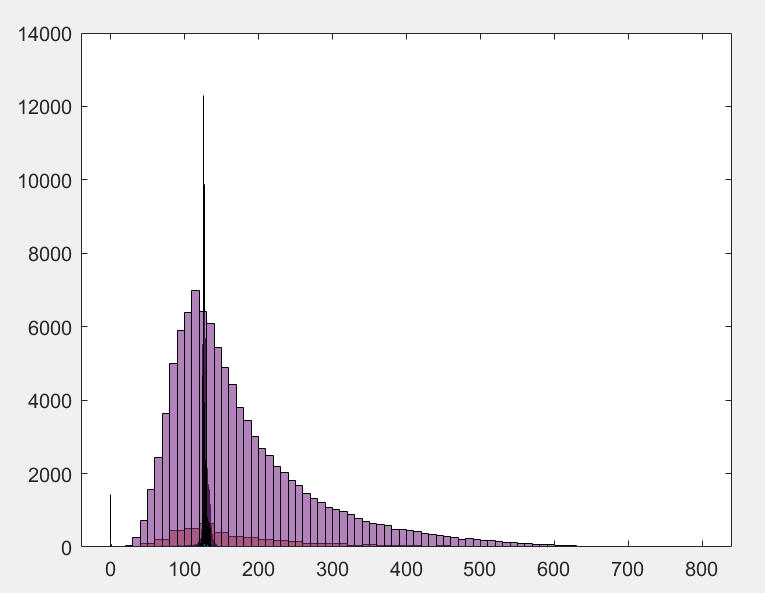
* 1. Another feature is d2, which shows significant decrease in total number of counts per event as shown below by the graphs(one on the left before mass cut and the one on the right after mass cut).
  2. Another feature which sees almost 50% decrease in the total number of counts after the mass cuts is t3(the graph without mass cuts on the left and the graph after mass cuts on the right)



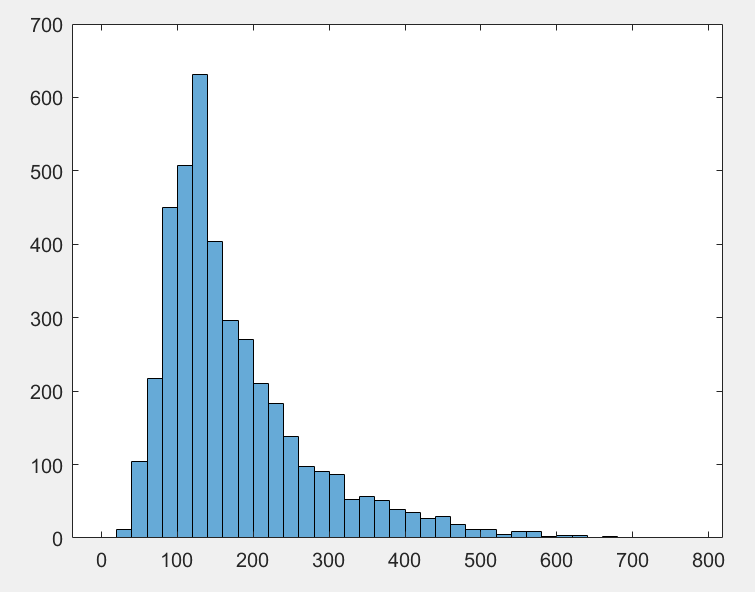
* 1. An interesting feature after the mass cut is t32. Being dependent on t3, the number of counts decreases. But the ratio by which counts in each event decrease is the same, thus preserving the characteristics of the variable t32 with more significance. (one on the left before mass cuts and one on the right after mass cuts)
  2. The mass cut at 1-sigma gives us high significance, as it clears our Higgs data of what may be noise and help us get more smooth curves to be used. As seen with the feature ktdeltaR. The mass cut helps us do the former (one on the left without mass cuts and one on the right after mass cuts).
  3. And as expected the background data has significant changes when before and after mass cut data is compared. As one can see from the graphs in the code. This was expected as the background covers the whole region whereas the signal is only concentrated in the upper half of the background (as shown). Thus an event selection of 1-sigma will give us a very small portion of the whole backround. And as if there’s signal data in our background it would be atleast above 5-sigma(as one won’t be able to claim discovery before 5-sigma). Thus these mass cuts gives us a much better data for our background as the possibility of our data being contaminated after 1-sigma mass cut is decreases which can be seen by the increase in significance.

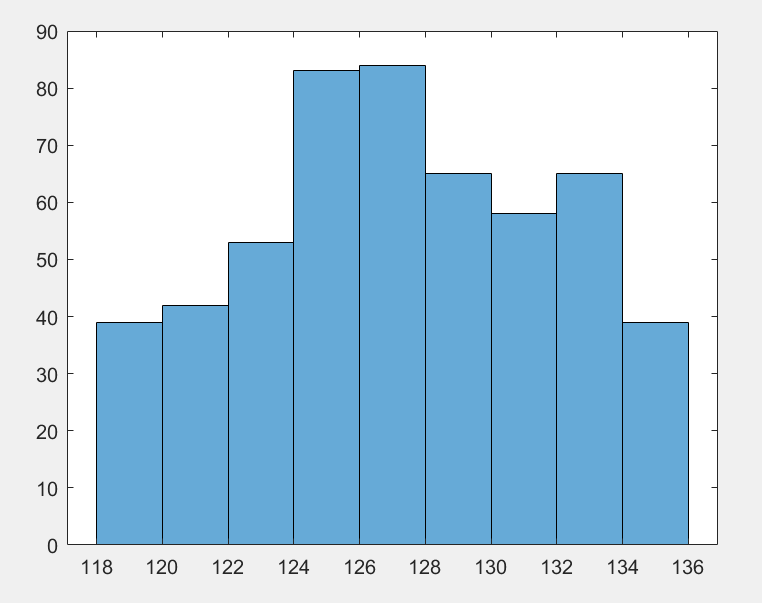
1. Thus, in order to use a feature to help in event selection, I’d choose d2 helps optimizes the ability to separate the N-subjettiness i.e. between different shapes of jets. Thus if one can differentiate well between the type of jet, one can disregard the data of Jets not related to Higgs boson, thus would help in a better event selection than using a 1-sigma selection to maximize significance. This can be done as d2 = eb3 / (eb2)3 with the scaling condition being that e3 ~ (e2)3 . [*Power Counting to Better Jet Observables*. Springer Link, <https://link.springer.com/article/10.1007/JHEP12(2014)009>.]

Lab 8

1. We plot the signal, High luminosity data and the background without any event selection and get this 

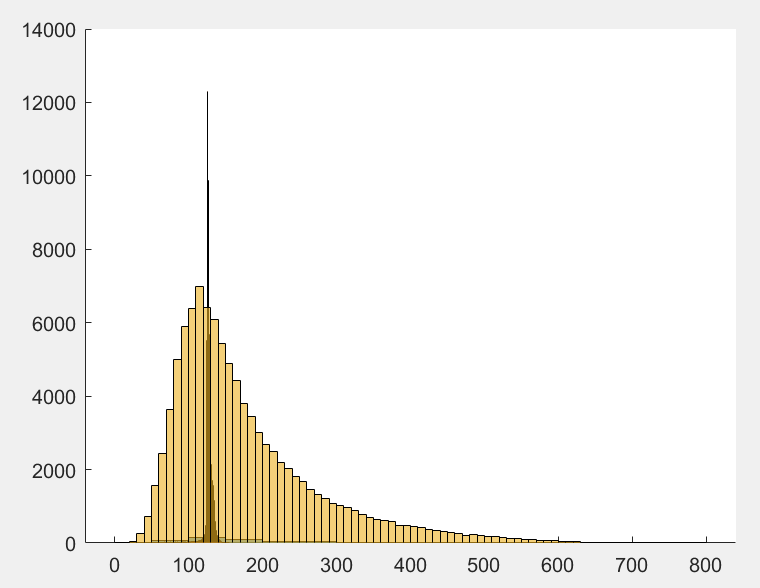
Now with optimal event selection for High luminosity data

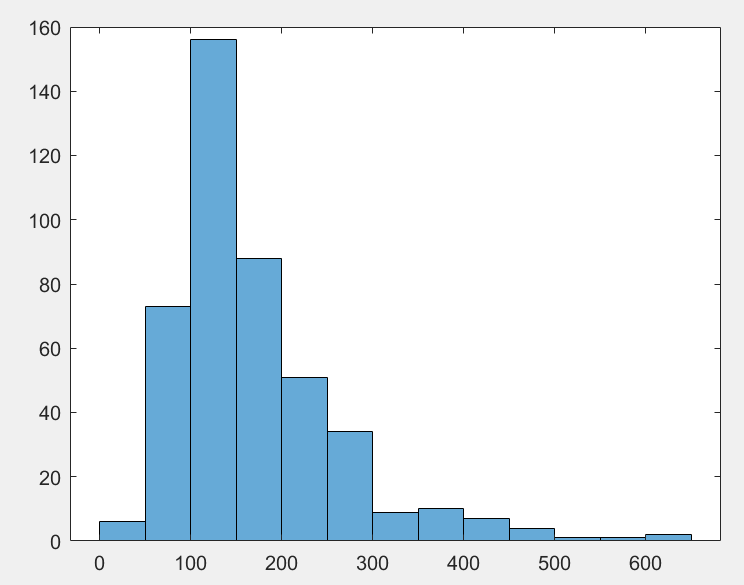
The High luminosity data changes from 

To: 

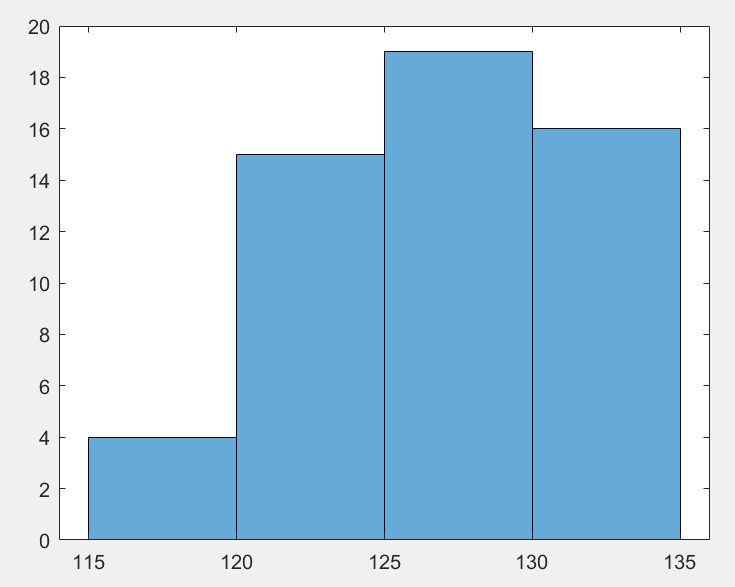
As one can notice that the graph move in towards the mhiggs = 126Gev. This also decreases the number of counts normalizing it to a 100 which is twice the minimum count requirement of Nhiggs= 50.

1. For low luminosity data when plotted without any optimal mass cuts we get (where the one on the bottom is the Low luminosity data)



Now when optimal mass cuts are applied the low luminosity data goes from 

To



As one can see that the values in the graph now have less divergence from the mhiggs=126Gev.

1. When calculating upper 95% confidence significance the expected value ~ 404. But the upon calculation the observed value ~ 111. This shows that the sample size of the data being observed here i.e. the Low luminosity data has a smaller sample size than the Higgs data.
2. Bonus:

the upper 95% confidence interval also = (mean) + (onesigmauncertainity)\*(std.dev)/sqrt(n)

[http://www.stat.yale.edu/Courses/1997-98/101/confint.htm]