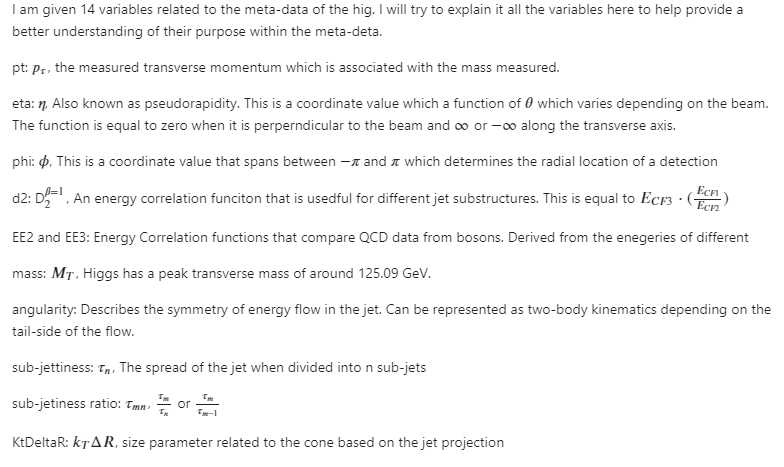
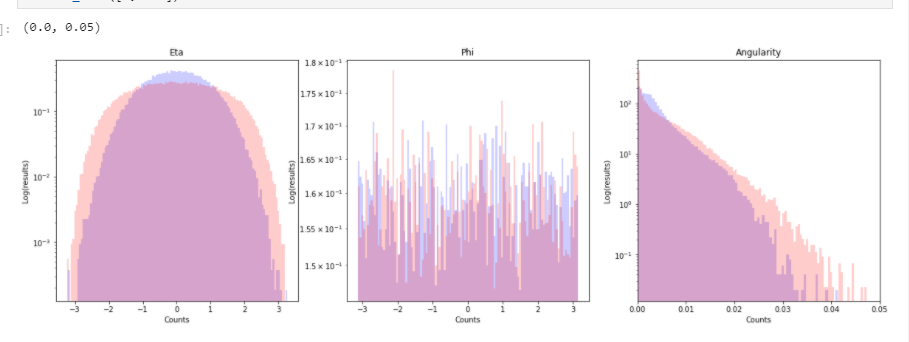
This report includes data taken at the LHC in Geneva. The LHC is responsible for collecting metadata as well as observations related to the collision of two particles, known as the Higgs Boson, at very high speeds. The Meta data is useful in making measurements and determining statistical significance for the Higgs Boson. This report includes suggestions on the different data I choose and how it provides available sensitivity for certain metadata related to the Higgs Boson. The metadata includes values of geometrical attributes, jet size, sub-jettness, calculated mass, transverse momentum, and more.

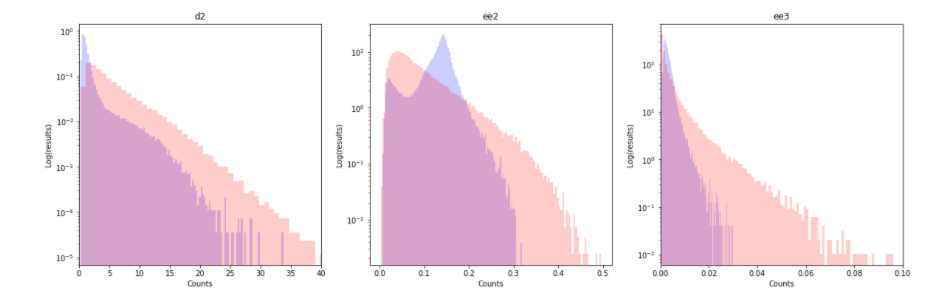
Here is a description of the 14 variables within the metadata. They are necessary to understand for further analysis of varying plots.



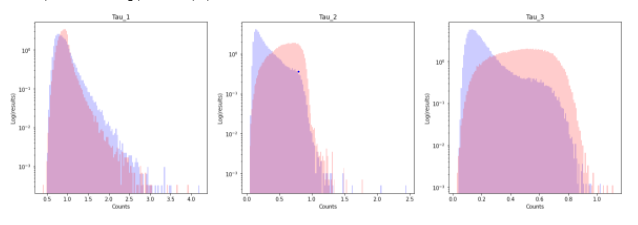
I have listed useful plot below on what data to keep and what data to throw away when taking the log of different plots related to the meta-data. Histograms are labeled as follows: The blue plots represent the Higgs signal data, the red plots represent our QCD background data, and the purple plots represent the overlap between the Higgs and QCD data.



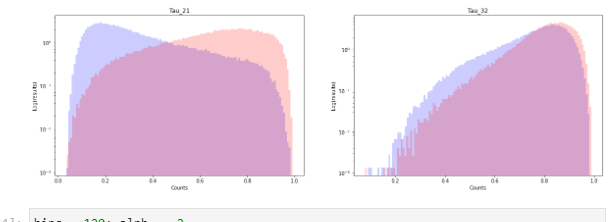
After just some initial plots, there is some discriminatory features found in eta. This looks to be a gaussian distribution between both the signal and background data. It does not look like there is any features that can be used amongst the Phi metadata. The angularity data is interesting, it looks like there could possibly be a Rayleigh or exponential like function being represented here. There is still a lot of overlap and only the tails of the plot reveal information about the angularity metadata.



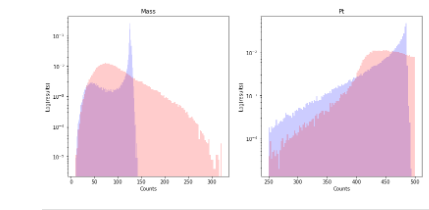
Here is where there is a discernable difference between the signal and the background data. ‘d2’ shows a maximum at a different value than the background. This peak is also more concentrated to 0 compared to the background data. Both ‘ee2’ and ‘ee3’ have very different distributions compared to the background data. The background for ‘ee2’ looks to take on a Poisson or gamma distribution whereas the signal appears to be an ‘extreme value Type 1’ distribution. The signal for ‘ee2’ and ‘d2’ seem to take on a greater peak of concentration. This will be easy to take further cuts when analyzing the data.



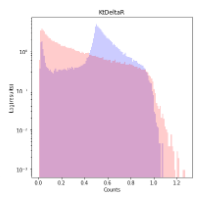
Comparing Tau values, Tau\_1 appears to take on the same form regardless of the signal or background. Tau\_2 and Tau\_3 do have very different distributions, though. After doing some research, Tau\_2 and Tau\_3 seem to take on a chi or chi squared distribution. The background distributions seem to take on a shifted gaussian distribution for Tau\_2 and a gaussian distribution for Tau\_3. I believe a good way to filter the data would be to alter the sensitivity for the certain plots given values for sub-jettiness.



The signal and background data for Tau\_32 seem to take on the same shape, but there could be a way to understand the difference between the two. The signal data, in blue, seems to jut out a little more and is wider. Tau\_21 has very different peaks at different locations. This could be used to change sensitivity of other sets. Tau\_21 seems to follow a ‘Extreme Value Type 1” distribution, but they do not allow for easy cutting parameters as a result. In our following labs, there might be a way to manipulate and cut the data accordingly.

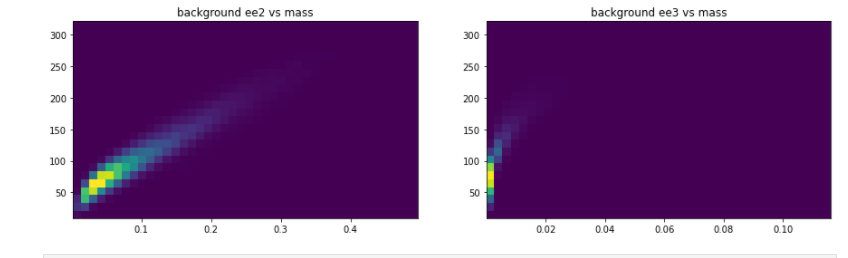


After doing some research the mass’s signal distribution should follow a Gaussian and Landau distribution. I have done more research on a Landau distribution and am able to be able to take cuts so that I can differentiate between the signal and background data. This is because the signal of the mass should act like the combination of the Landau and Gaussian distribution. The ‘pt’ distribution is one that I have never seen before, this is something where more research is needed as well as making specific cuts.

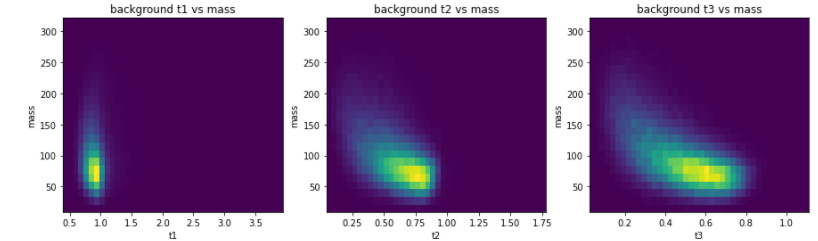


The ‘KtDeltaR’ is very interesting. There is a way to take out some data by using various cuts, but it does not look like it will be very effective. Another interesting note is that this distribution looks like it is a combination of other distributions. More research about how ‘KtDeltaR’ is collected is needed to get a better understanding of the distribution.

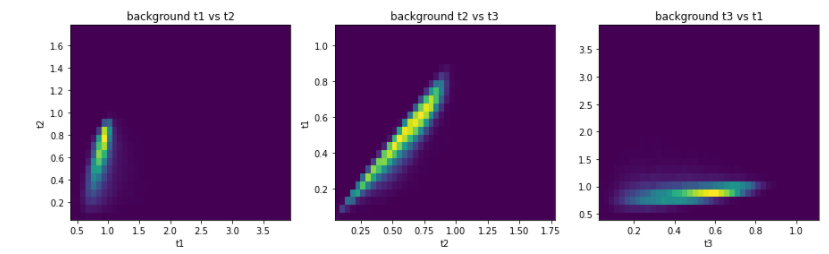
The next part is to understand how different variables are correlated to one another and if there is a trend that can be seen. Most of these plots are based off the mass detection relating to the QCD data. This provides the opportunity to find varying trends related to the metadata which will help to further increase the sensitivity.



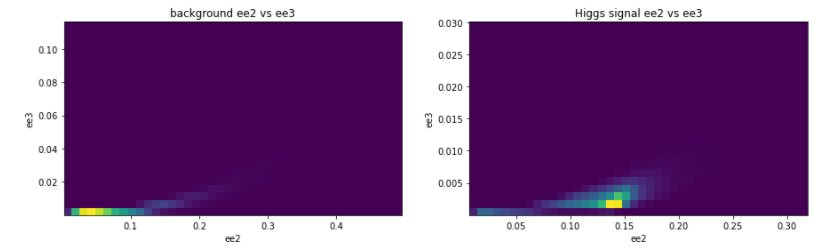
There is an interesting trend when comparing the ‘ee2’ vs ‘mass’ metadata. It shows that as we increase ‘mass’, so does the value of ‘ee2’. ‘ee3’ does not look to vary due to the change in mass. This is very interesting because ‘ee2’ seems to follow that trend.



Above, for varying levels of sub-jetiness, there is a similar distribution when plotted against mass. It seems that the distribution seems to ‘spread out’ but the overall shape is about the same.



There seems to be a linear correlation between the sub-jettiness levels. This is very easy to tell in the ‘background t2 vs t3’ plot and ‘background t1 vs t2’ plot. The ‘t2’ and ‘t3’ trends are at or just below a 1:1 ratio. The ‘t1’ and ‘t2’ plot does follow a linear trend but it is a value that is higher value than 1. The correlation between ‘t1’ and ‘t3’ is a harder trend to distinguish, but it does appear to be linear.



These are two very interesting plots which shows how ‘ee2’ and ‘ee3’ are correlated. The Higgs signal data varies significantly compared to the background data. This needs further research, but I feel there could be some interesting distribution relationships to further understand the data.