CMS Data Analysis 2018 summer project.

Our goal is to search for a new particle which decays to an upsilon and two muons. The upsilon decays to two muons. So our signature will have four muons. We will use the publicly accessible CMS data to do a trial version of this analysis this summer.

This site has a lot of tutorials and guides for learning about CMS.
http://www.i2u2.org/
Go to the CMS e-lab

This page has a lot of information about CMS, including a Slideshow and Event Display demos:
http://www.i2u2.org/elab/cms/library/resources.jsp?options=project

ur VM "validated" — i.e it gets access to and compiles the CMS software, and reads the CMS Open Data files — if you get an output like:
CMS Open Data Analysis

Next go to the CMS open data :http://opendata.cern.ch/ is the main site. Go to CMS (later you can work on other experiments if you like).

If you do not have a cms account, you need to install the CMS virtual environment. It uses VirtualBox, which you can download. Follow the link to Install CMS virtual machine (http://opendata.cern.ch/VM/CMS) Install and validate your installation.

Next you can try out an analysis. Go to this page: http://opendata.cern.ch/collection/CMS-Tools
For example you can choose the Two-lepton / Four-lepton with the 2011 data. Click on the link and it will provide you with all of the information that you need, including the commands to download and run the code (
https://github.com/katilp/OutreachExercise2011/blob/master/README.md

From http://opendata.cern.ch/VM/CMS just follow the directions for setup cms environment and run the DemoAnalyzer

It just runs over the data without doing anything.

Another good tutorial is the CMS HEP tutorial at : http://opendata.cern.ch/record/50

Next try the Z and ZZ to leptons analysis:
https://github.com/katilp/OutreachExercise2011/blob/master/README.md

We can build our analysis off of the Z and Higgs analyses, as far as getting some example code that does something similar to what we want to do.

1. Start with a program to find the Z decaying to two muons. The invariant mass of a particle which decays to two or more objects is calculated by summing the four-vectors (energy, px, `, pz) of each of the daughter particles. Then calculate the mass by m**2 c**4 = E**2 - p**2c**2. There is a function in cmssw that will do this calculation for you once you create an object using the four-vectors of the daughter particles.
2. Next, find the upsilon. You will need to modify the code to change the mass range to that of the upsilon. You will also need to use a different dataset, such as the /MuOnia/Run2012 dataset.
3. Next we want to look at the Higgs to Four-Lepton analysis
4. We want to adapt the code to look for a new resonance at an unknown mass, which will decay into upsilon and two muons. It will be at a lower mass than the Higgs in this example.
    a. The dataset we need is /MuOnia/Run2012 (see http://opendata.cern.ch/record/6042 for how to access this data)
    b. We need to lower the pt cuts on the muons to 3 GeV on each muon. You will need to modifiy the code to do this.
    c. You need to find the mass of each pair of muons, and select one of them to be consistent with the upsilon mass, instead of the Z mass. You will need to modify the code to do this.
    d. Make these plots:
        i. A 1-D plot of the dimuon mass of the pair that is closest in mass to the upsilon
        ii. A 1-D plot of the dimuon mass, where you put in both pair masses
        iii. A 1-D plot of the dimuon mass of the pair which is not closest to the upsilon mass
        iv. A 1-D plot of the four-muon mass
        v. a 2D plot of the mass which is closest to the upsilon on the y axis, and the four-muon mass on the x-axis.
        vi. a 2D plot of the dimuon mass which is NOT closest to the upsilon on the y axis, and the four-muon mass on the x-axis.
    e. When we have isolated our four-muon sample, there are a couple of analyses we can do.
        i. Double Upsilon production: our graduate student, Maksat Haytmuradov, wrote his thesis on this process, and you can read the paper at :

https://arxiv.org/abs/1610.07095 . How many events can you find consistent with double upsilon production?

ii. Upsilon J/Psi production: our graduate student Kamuran Dilsiz studied this. Can you find any events in the 2012 data?

iii. Rare Z decays. Recently CMS came out with a new result for Z→ J/Psi X. The decay of Z → Y(1S) X is also very rare. Look on the pdg website for the predicted branching ratios(. Can we see Z→ Y(1S) mu+ mu-? Probably not in the 2012 data, but we can look, and develop the analysis, and then run on the 2016 and 2017 data with help from a student with a CERN account-- Orgho, Cory, Runxiong may be able to help with this. See this document for using crab : https://docs.google.com/document/d/1PcrdyWlLhO-oDEE83iOsrchOGV8kygzQkFJgC9_8WCQ/edit

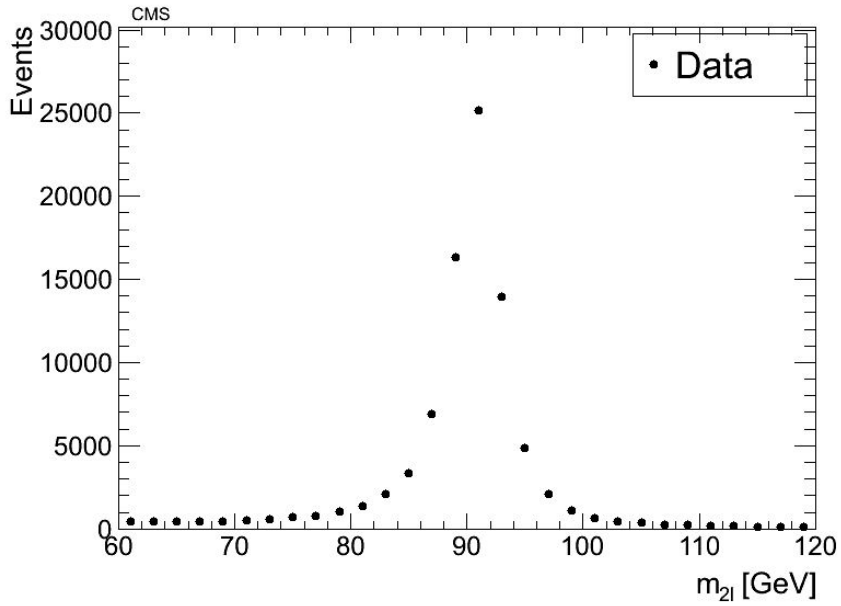
iv. Search for new particles : Here you can search in the four-muon mass spectrum to try to see a new resonance. You will use the fine-bin four-muon mass plots, and use statistical tests to decide if a point which deviates from your expectation is a signal or is a fluctuation. More details to come on this.

1. Here is a simple first way to test for an excess.
    a. Take your four-muon mass plots, and choose your region of interest. Simplest would be a region above the threshold, so for example in the set of plots for Y2<8, you are above the threshold in the 20-40 plot, so fit that entire distribution with a exponential or a polynomial (or exp + poly). You will have to try different functions to get one that fits well (reasonable chi-squared).
        i. If you have a mass distribution with a threshold effect (steep rise on one side, corresponding to the sum of the masses of the daughter particles, for example 18 in the upsilon-upsilon plots) -- then you use a function to fit the threshold and then the exponential fall for masses above the threshold. Or, just ignore the first several bins which will have the threshold effect, and work with the distribution on the exponentially falling side. The threshold region can be important, but also very complicated.
    b. Once you have a functional form that you like, then you go through the histogram, excluding one bin at a time from the fit (your test bin). Fit the histogram (minus the test bin) and then with that function and constants (and their uncertainties) that you get from the fit, calculate the expected value and uncertainty for that test bin.

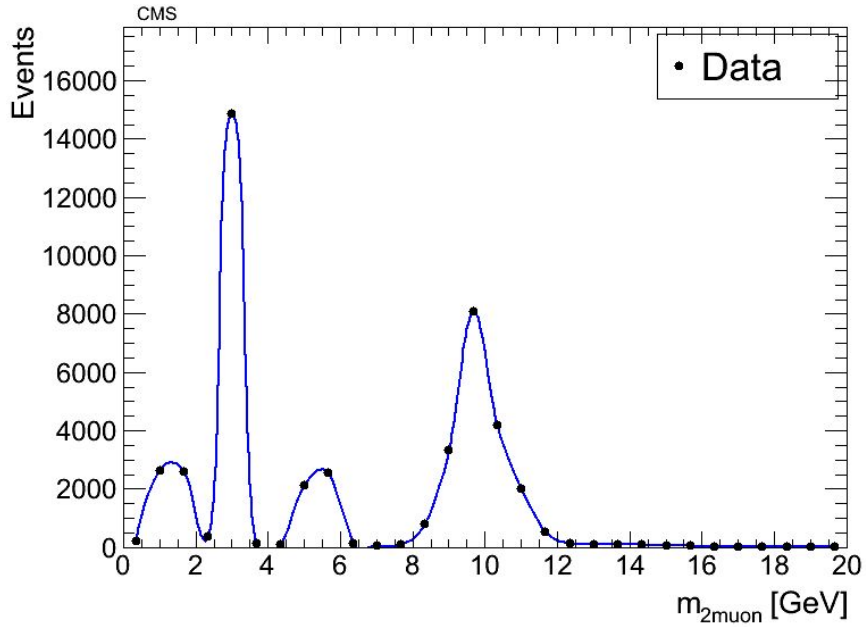
c. Now you have, for the test bin, an expected value +/- error, and an observed value +/- error (the error bar on the point). Combine the errors in quadrature to have one total estimated error (this is not exactly correct but ok for an estimate), and see how many sigma your observed is from your expected. For example, if your expected value from the fit was 3.4 +/- 0.2, and your observed number of events at that point was 6 +/- 0.6, then your total error is sqrt(0.2^2 + 0.6^2) = 0.63 or just round it to 0.6. Then you find the number of sigma you are off by taking the difference (expected - observed)/(total error) -- so in this example (3.4 - 6)/0.6 = -4.3 sigma. Make a new histogram with the same mass binning, and fill the test bin with -4.3. Go through all of the bins doing this same calculation.
d. Look on your histogram for the largest deviation. If you see something above 3 sigma, this is worth further investigation. 3 sigma is the usual threshold for evidence for something new, 5 sigma is considered discovery of something new (of course, much more work would need to be done to claim discovery.

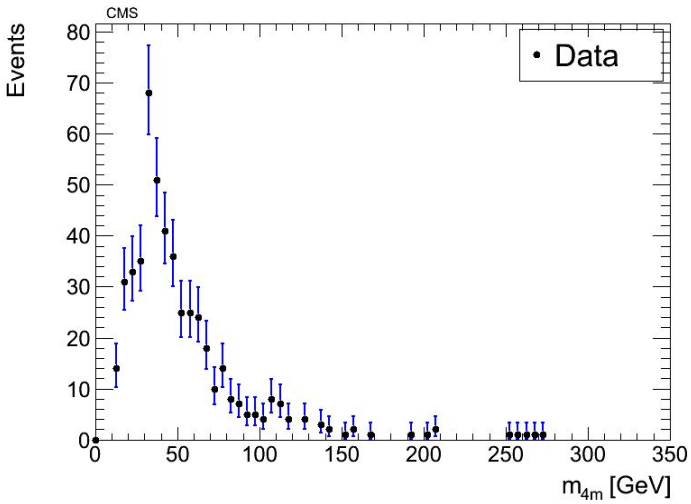
5. More information
    a. To learn more about how to identify physics objects in CMS, read this: http://opendata.cern.ch/about/CMS-Physics-Objects/2010 and http://opendata.cern.ch/about/CMS-Physics-Objects/2011style
    b. CMS HEP tutorial: http://opendata.cern.ch/record/50

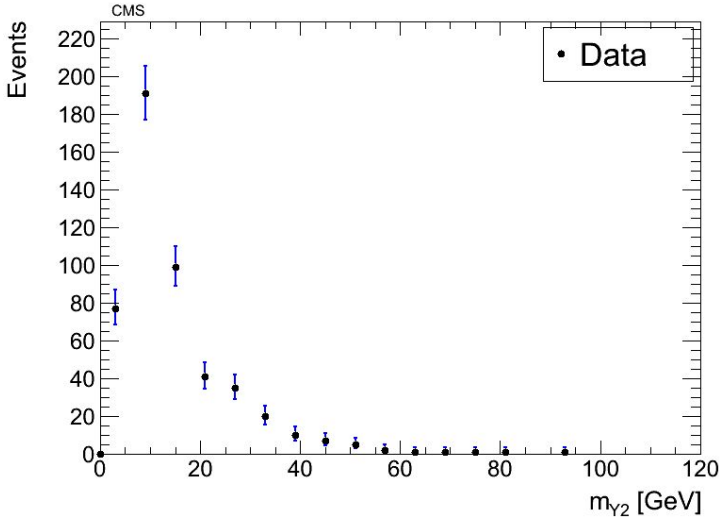
Here is our pretty histogram of Z -> 2 muons. The peak is around 90, with a very normal distribution. Run on ~250k events.



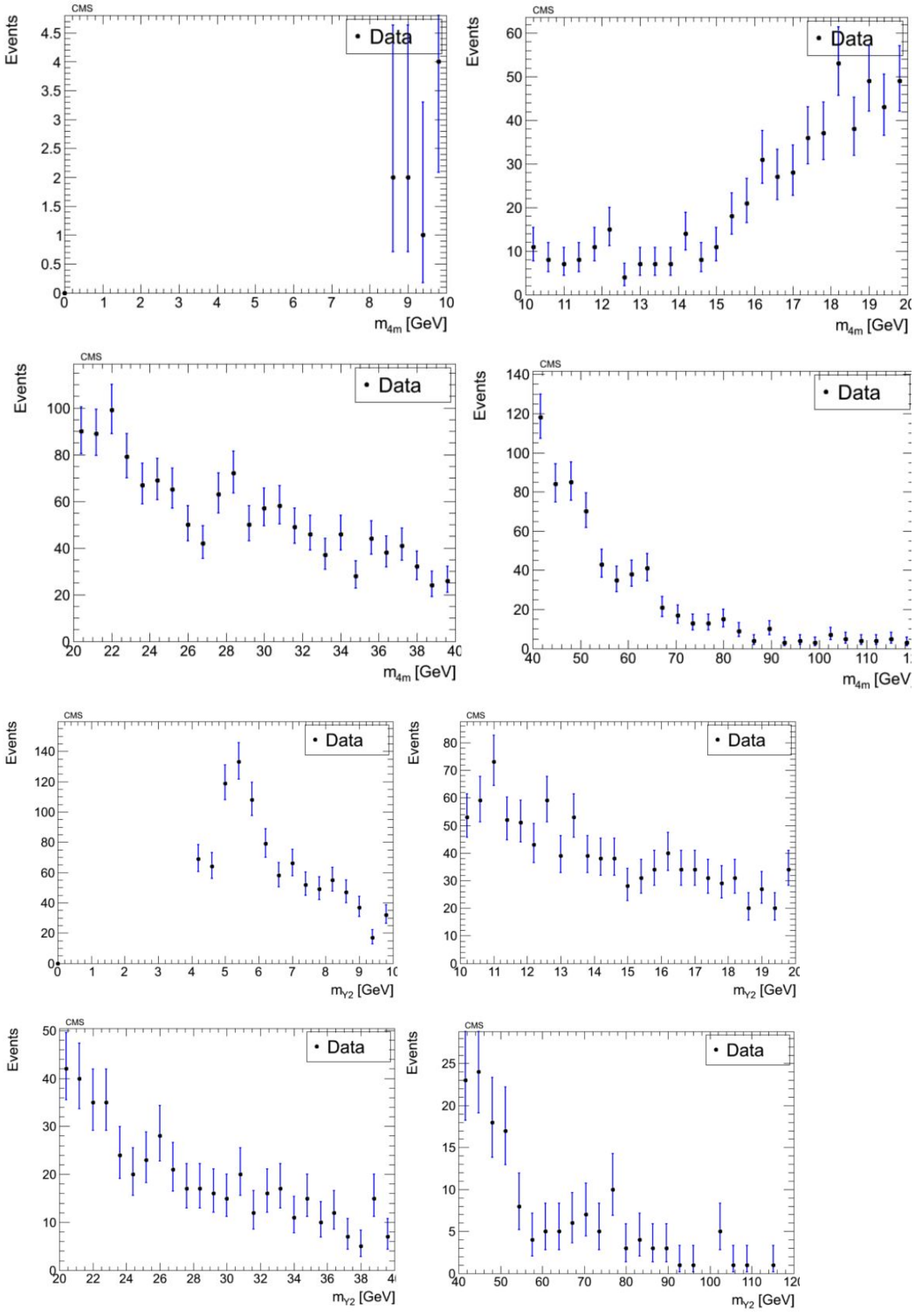
Here is our first graph for upsilon. There is the expected peak around 9.5 GeV, and the peak around 3 GeV is the J/Psi. Run on 100k events.



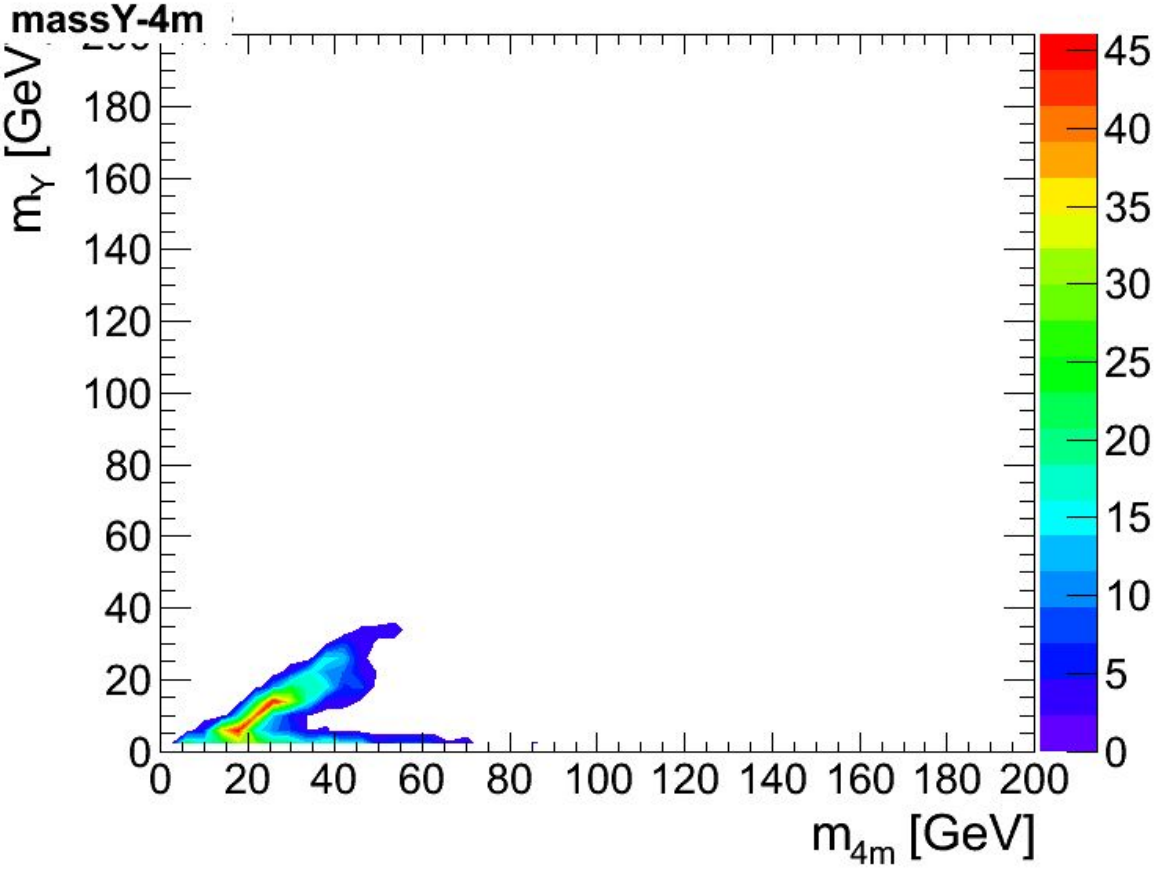

CMS
Events
Data
m_{4m} [GeV]


CMS
Events
Data
m_{Y2} [GeV]

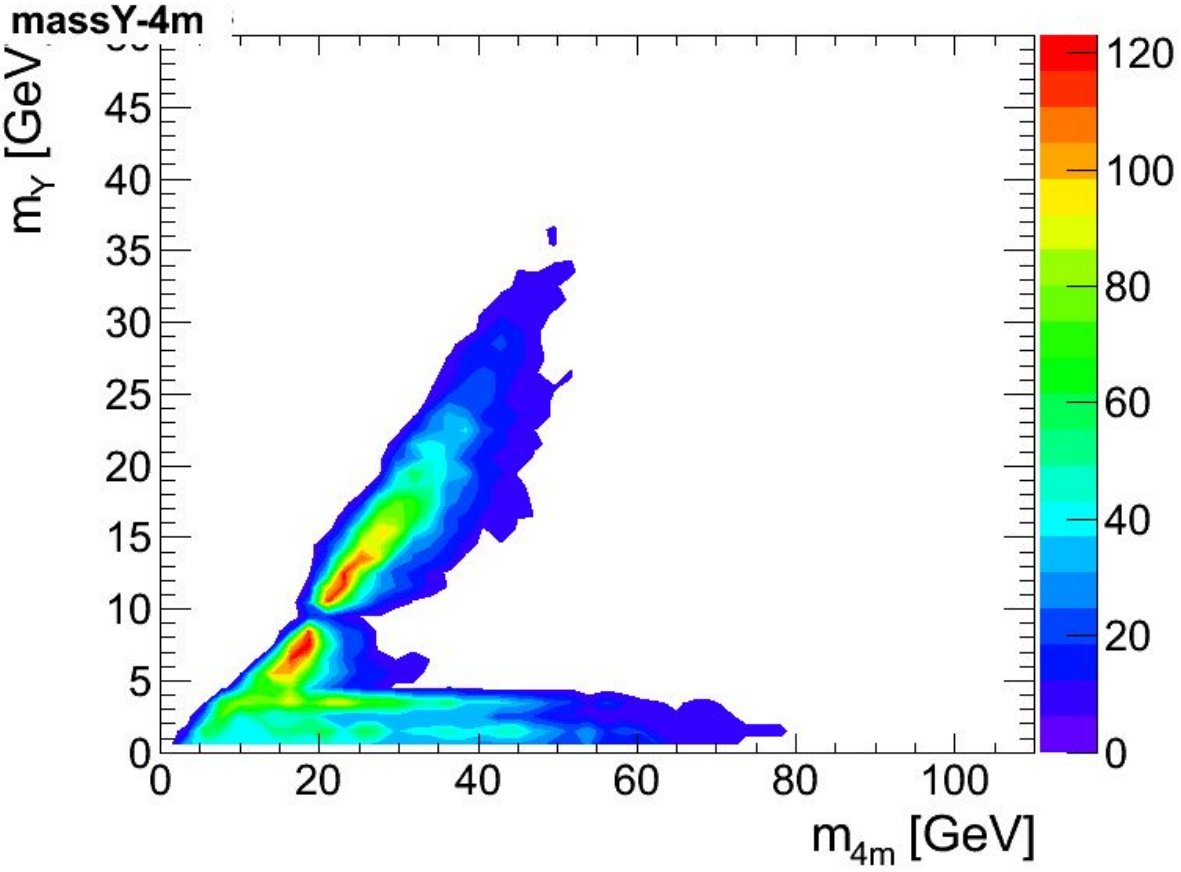
Here are the more precise histograms for 4-muon mass and Y2 mass. Note the different y-axis scales. The absolute peak for the first histogram is in the 40-120 range and the peak for the second is in the 0-10 range. Run on 240,100 events.



Here is plot vi, the 2D histogram of 4-muon mass and the pair furthest from upsilon mass. Run on 1,000,000 events in 5546 seconds.



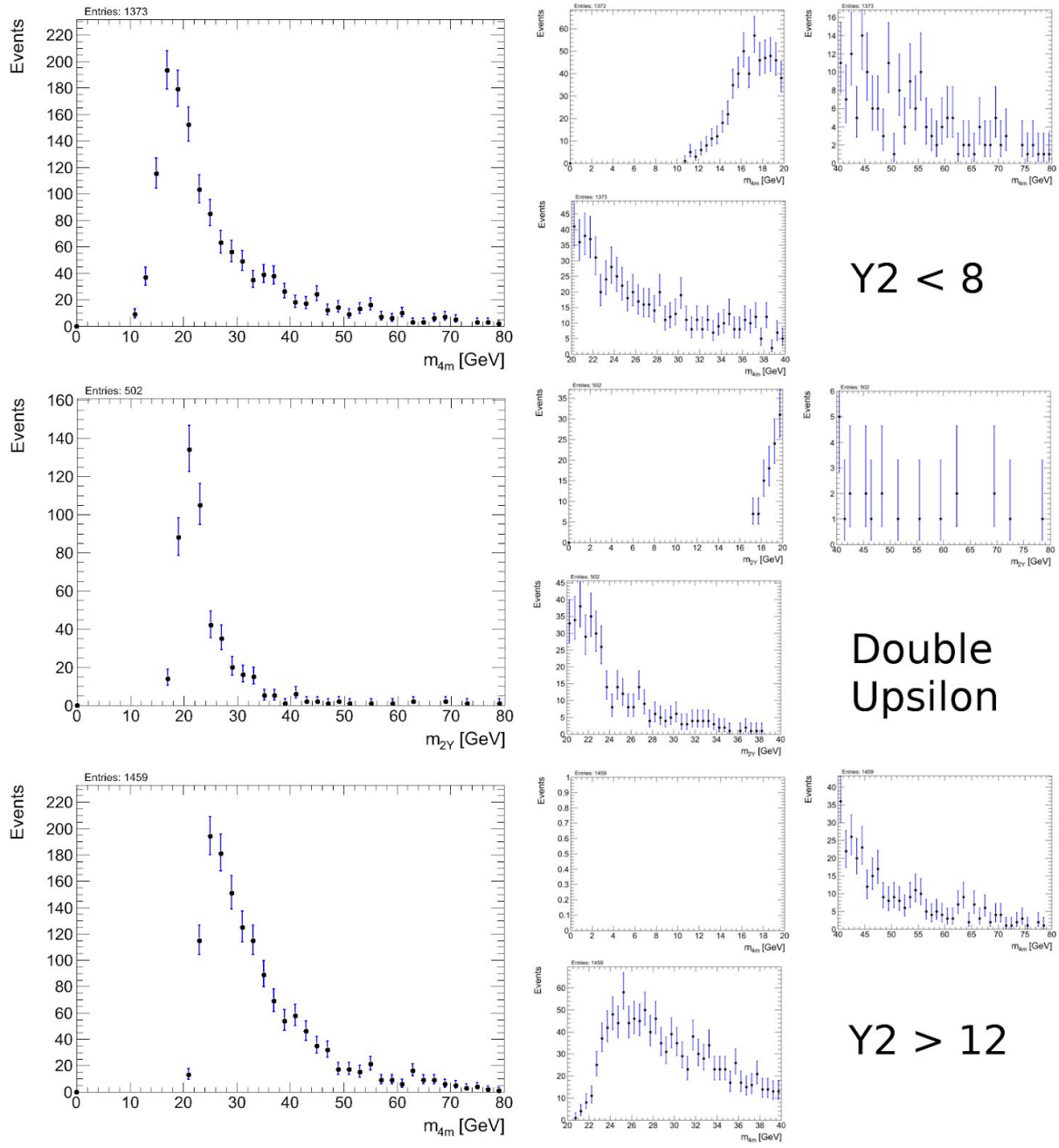
Here is the same histogram with altered scale, run on all of the events from the 2012 upsilon data.

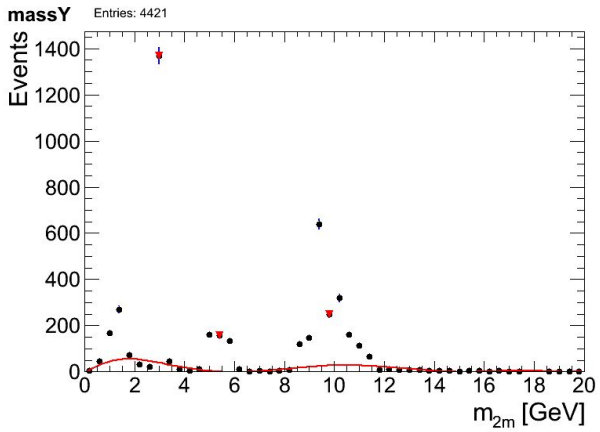


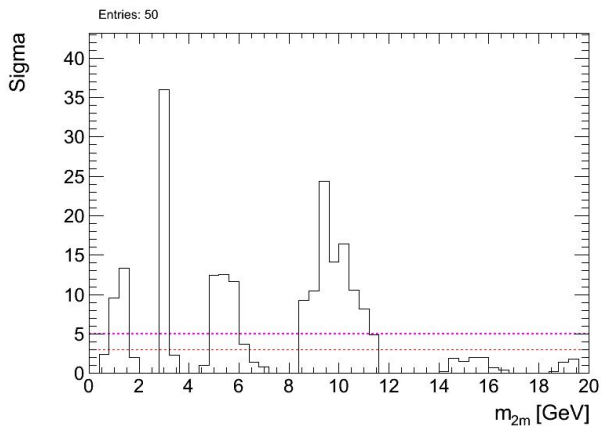
Here are the histograms of upsilon + di-muon mass where:

1. Di-muon mass < 8 GeV
2. Di-muon mass falls within the upsilon mass window (8-12)
3. Di-muon mass > 12 GeV

This analysis was run on all of the events in the 2012 set.
root get value of bin




massY
Entries: 4421
Events
1400
1200
1000
800
600
400
200
0
0 2 4 6 8 10 12 14 16 18 20
m2m [GeV]


Entries: 50
Sigma
40
35
30
25
20
15
10
5
0
0 2 4 6 8 10 12 14 16 18 20
m2m [GeV]