These are instructions for creating an invariant mass histogram using SCILAB and CERN's open data. This is meant to complement the instructions and tutorial for using Excel/OpenOffice to create the same histogram. The data is from dimuon decay and will peak at around 90 GeV for the Z boson. This is actual data from the Compact Muon Solenoid (CMS) which has been acquired, analyzed, filtered and identified as collisions in the Large Hadron Collider (LHC) as those that exhibit two muons. Muons are daughter particles of the Z boson, so when the Z decays, it theoretically exhibits 2 muons.

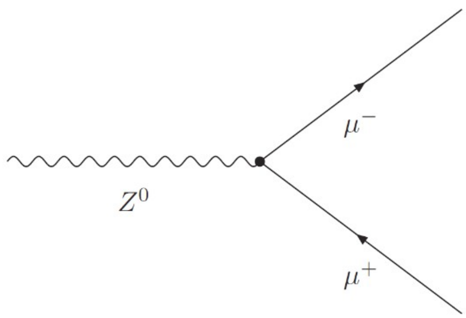


Figure 1: Feynman Diagram of Z-boson decay.

Instructions:

1. Download the comma separated(.csv) file at: http://opendata.cern.ch/record/700 It is the largest file, MuRun2010B.csv, around 15 MB in size. Note: you can have students download the other files if the dataset is too large. The large dataset contains 100,000 collisions, and the other 10 are each 10,000 component chunks. The more data you have in the histogram, the more statistically accurate it will be. I would work with the large data set first.

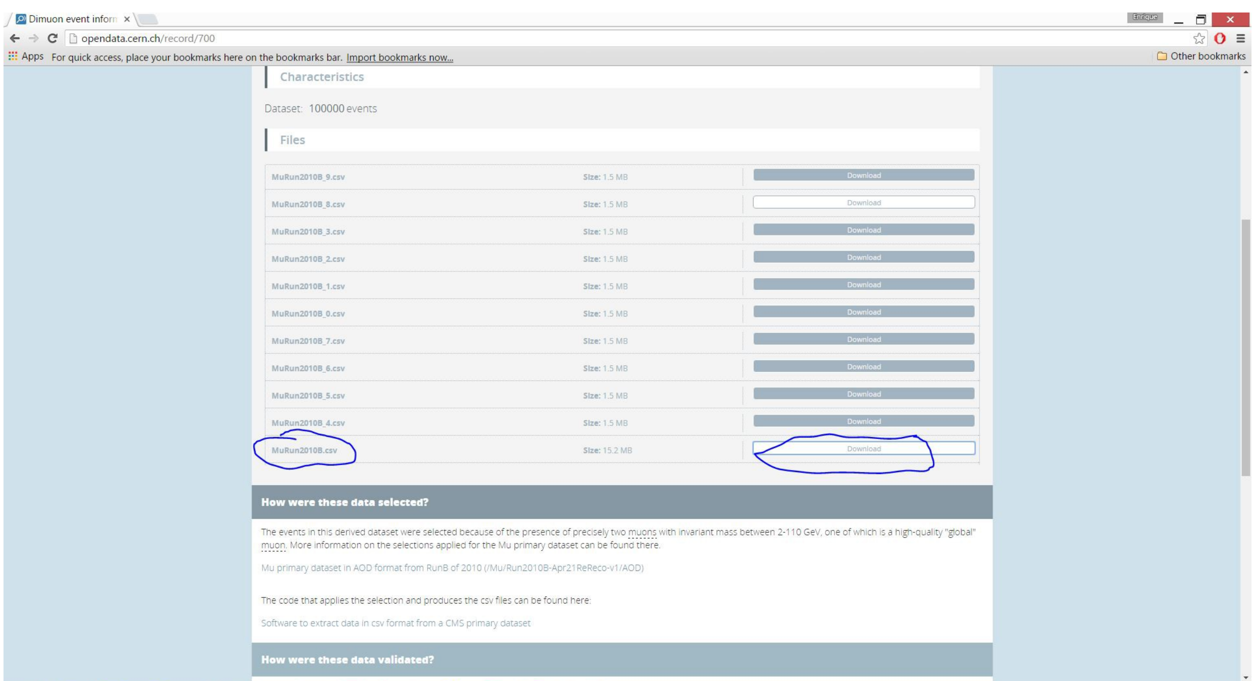


Figure 2: download MuRun2010B.csv (This is data from the first LHC run, taken in 2010)

2. For this activity, you will need to download and install SCILAB. This is the free version of MATLAB, and is available at http://www.scilab.org/

A couple of notes:

- SCILAB has limitations that MATLAB does not, including how much data it can handle.
- I installed this on a PC with Windows 10 and the rest of the instructions match this. You will need 64 bit, 130 MB.
- I have run MATLAB on my mac in the past, and other versions of SCILAB will most likely be compatible with your system.

Figure 3: Download installer, run the file, select language and recommended install.

3. Open/Launch SCILAB and launch scinotes by clicking on the top left icon on the Console.

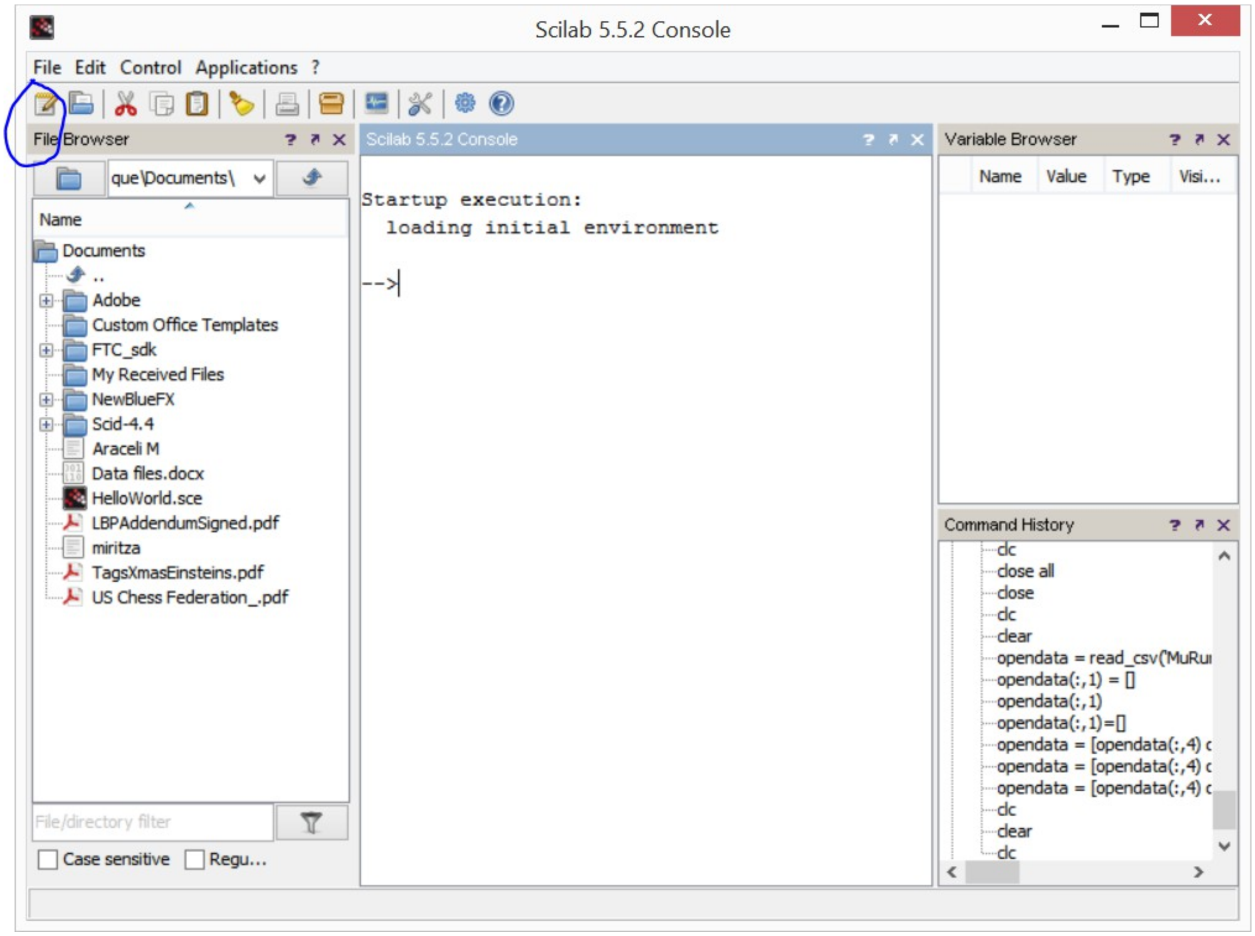


Figure 4: The Console window (Note my documents folder will not look like yours and your Command History will be blank) A blue circle denotes how to launch scinotes.

===

==================IF YOU ARE A SCILAB/MATLAB EXPERT SKIP STEPS 4 AND 5.=================

===

4. Go back to the console window. If you are new to MATLAB/SCILAB, you can enter commands into the console sequentially. This is often helpful when you just starting with SCILAB. For example, you can type:

   --> disp('Hello World')

   And then hit enter. SCILAB will use the disp function to write text.

   You can also try basic mathematical operations, such as:
   --> 1+1
   --> 5*(1+1)

   You can also try assigning variables, such as pi or the golden ratio:
   --> pi = 3.1415
   --> phi = (1+(5)^.5)/2

You will notice that your variable browser will have a new variables, called phi and pi. (by the way, %pi is a built in variable in SCILAB for π to more precision) You can use the golden ratio to solve the equation $x^2 - x - 1 = 0$ For example, type in:

--> phi^2 - phi – 1

Or use pi to calculate the area of a circle of radius 2 meters. Note these commands will only work sequentially. You first have to define variables and only then can you compute the equations.

--> Area = pi*2^2
--> radius = 2
--> Area = pi*radius^2

So you can maybe see where we are going. We will now import the CERN csv file data into SCILAB, define variables, and then process the data to calculate invariant mass, just like we did in the spreadsheet. Finally, we can plot the same graph in SCILAB that we did in the EXCEL tutorial.

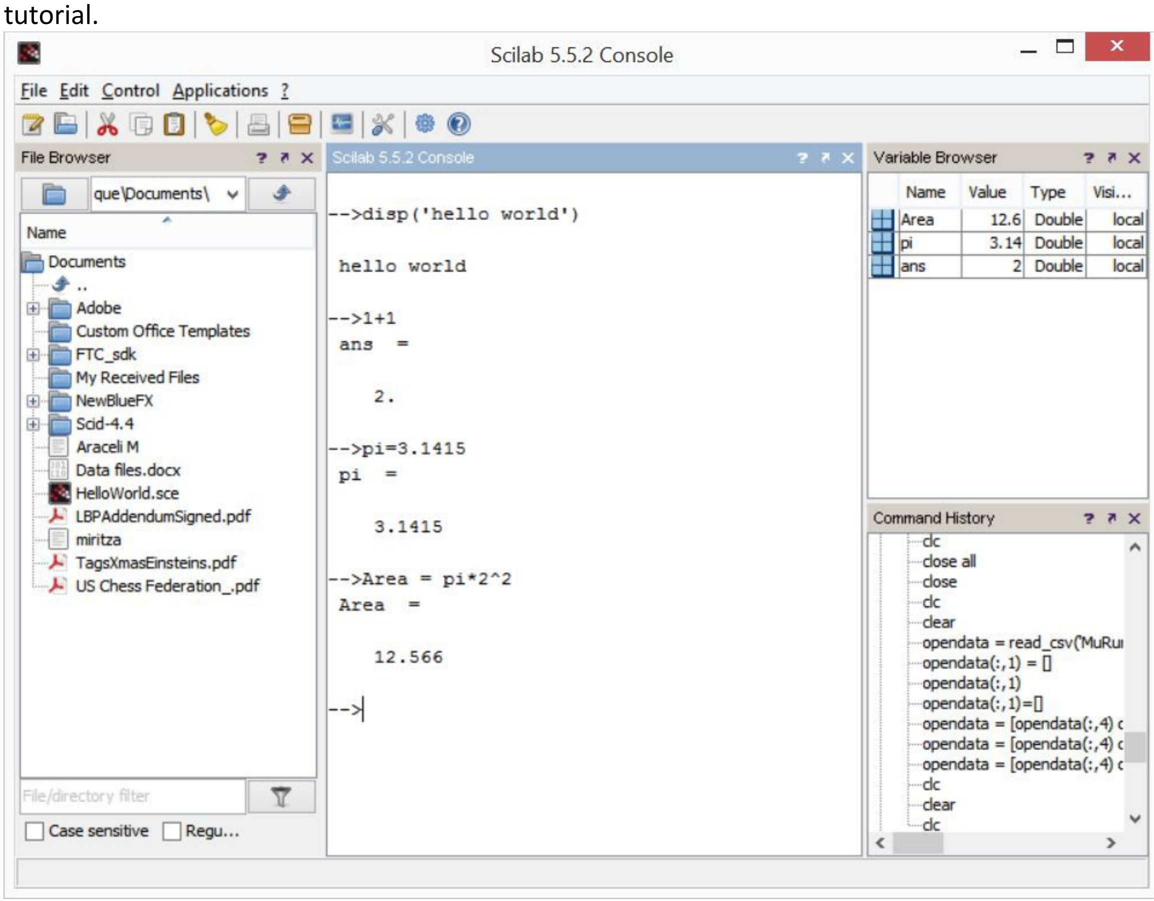


Figure 5: Typing into the console window some initial commands.

A couple of SCILAB tips before we go on:

--> clear

Clear deletes all your variables and resets the console. You will want to use this often when running scripts—our next step… By the way, SCILAB is case sensitive, so don't type in capital letters unless you mean it!

-->clc

This clc command cleans up the console window. Type both of these in before doing the next step.

5.

a. Write you first script: Go to the Scinotes window. Here you will write sequential code like you were typing it into the console, but then you will run the script at the end. Type in:

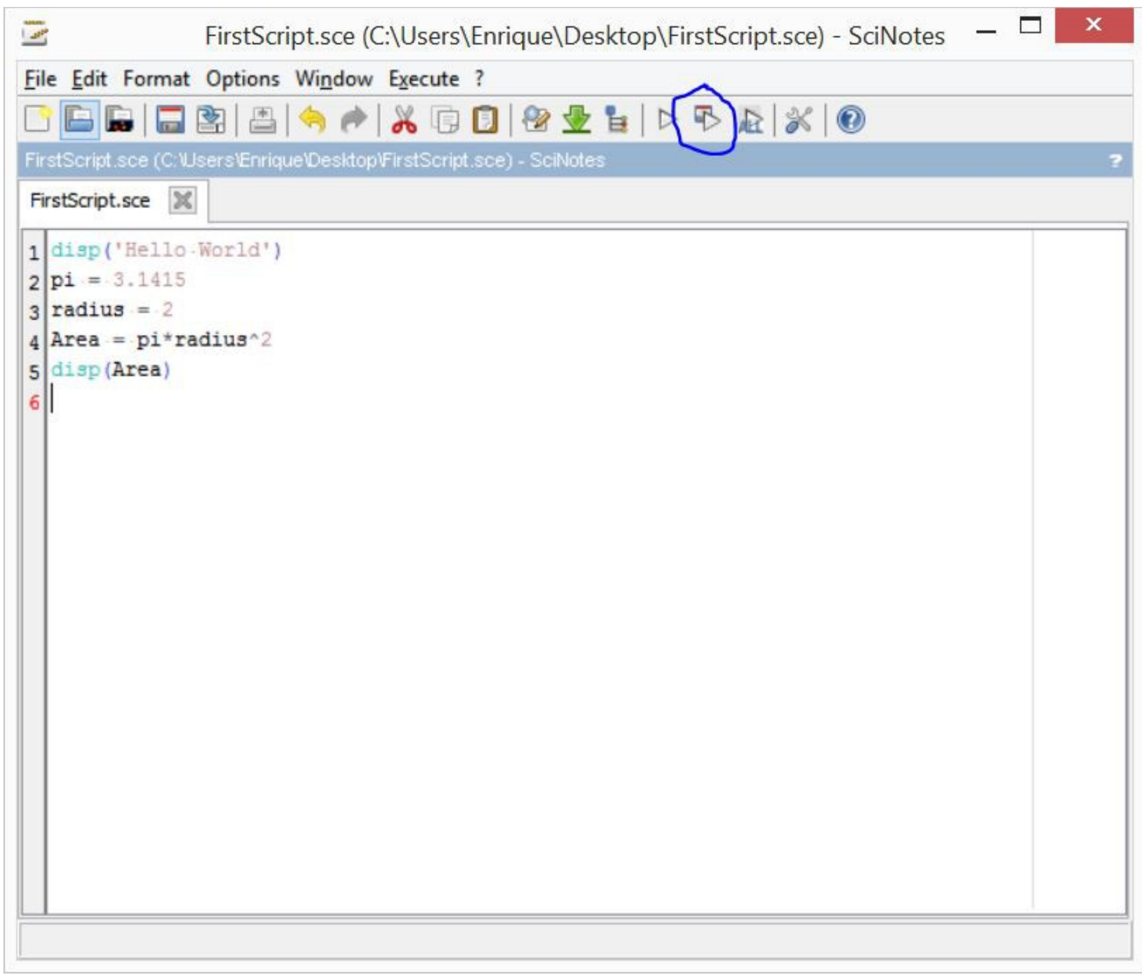


Figure 6: Your first Script: notice the save and run button at the top.

This script when run, should display Hello World, define variables pi and radius, calculate the area and display the Area answer.

b. Now click on the save and execute button on the top. (There are various ways of running and saving scripts, but this button is convenient when you are starting to code.)

Save the file as FirstScript and save it to your favorite directory/folder. I usually use the desktop or a designated SCILAB folder for my projects. This is kind of an important point. Wherever you choose to save it to, that is where you have to move/copy your csv data that you downloaded from CERN.

Now go look at your console. WOW!

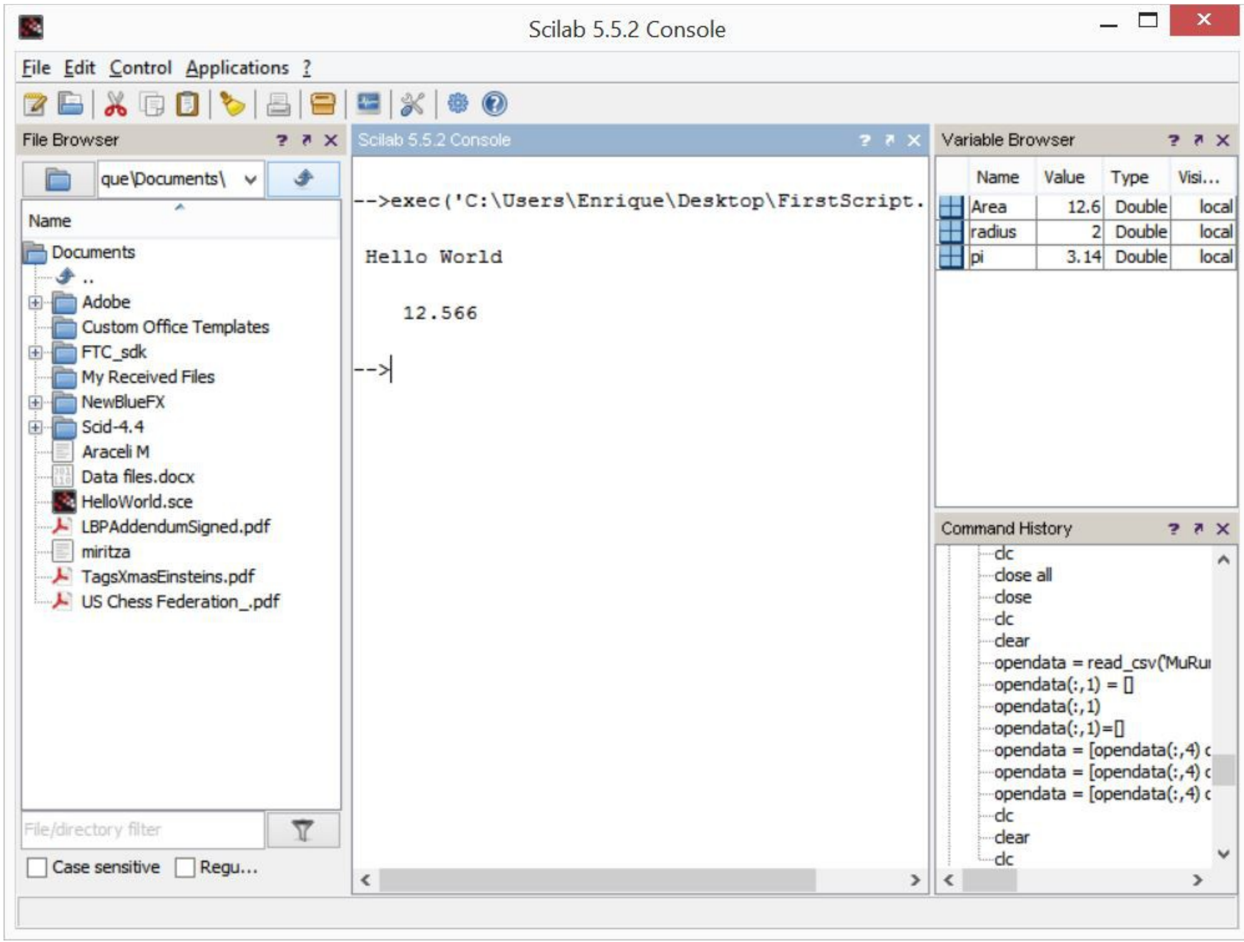


Figure 7: Your first script is run. Area of a 2 meter circle.

There are many resources for learning SCILAB/MATLAB online. I will omit any further study and refer you to google. We will now write a script (sequential list of commands to make a histogram)

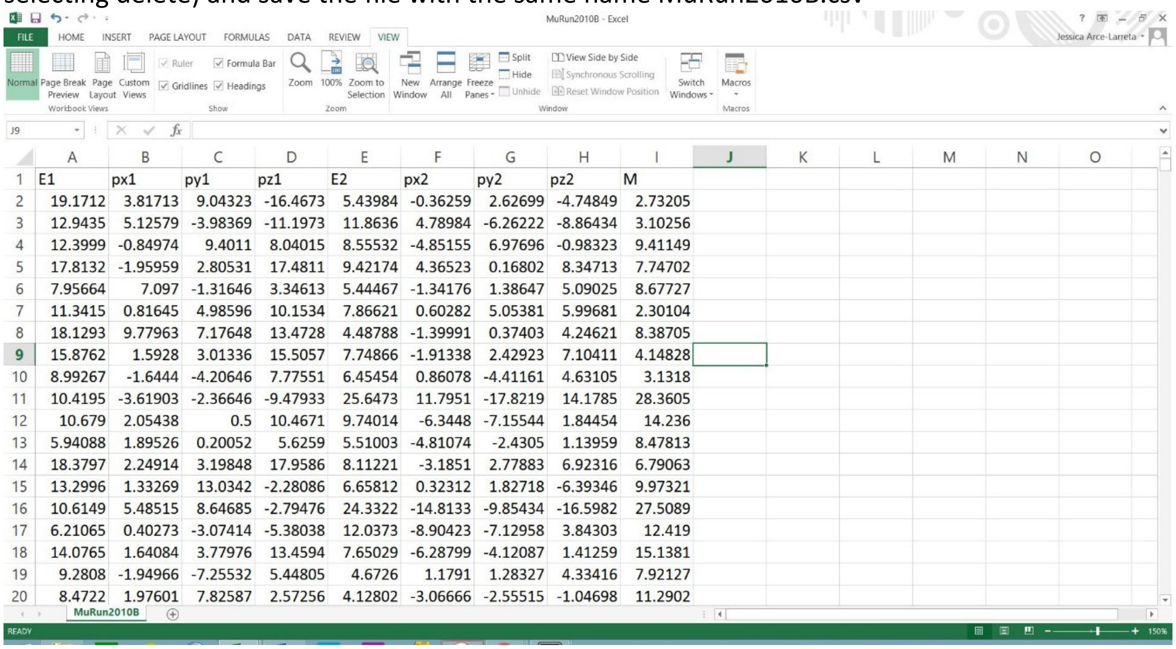
6. Writing the histogram script. Basically, we will do this in four steps:
   - a.i. Import the data
   - a.ii. Define the variables
   - a.iii. Process data/calculate invariant mass
   - a.iv. Graph the data in a histogram

i. Importing Data

1. Move/Copy the MuRun2010B.csv file you downloaded to the directory/folder you are going to save your scripts to – in my case, the desktop.

2. Open the csv file with Excel. Recall from the EXCEL activity, that the only data we need to calculate invariant mass is E1, E2, px1, py1, pz1, and px2, py2, and pz2. I also save the M data, which is the given invariant mass. If you haven't done the excel tutorial, E1 and E2 are the energies of muons 1 and 2 (remember this is data for 2 muon events) measured in GeV. (GeV are giga-electron volts). The remaining data is the momentum of the muons in the x, y, and z direction (3D). Momentum is also measured in GeV since particle physicists define c = 1.

Delete all the unnecessary columns (delete by right clicking on the letter of the Column and selecting delete) and save the file with the same name MuRun2010B.csv



E1	px1	py1	pz1	E2	px2	py2	pz2	M
19.1712	3.81713	9.04323	-16.4673	5.43984	-0.36259	2.62699	-4.74849	2.73205
12.9435	5.12579	-3.98369	-11.1973	11.8636	4.78984	-6.26222	-8.86434	3.10256
12.3999	-0.84974	9.4011	8.04015	8.55532	-4.85155	6.97696	-0.98323	9.41149
17.8132	-1.95959	2.80531	17.4811	9.42174	4.36523	0.16802	8.34713	7.74702
7.95664	7.097	-1.31646	3.34613	5.44467	-1.34176	1.38647	5.09025	8.67727
11.3415	0.81645	4.98596	10.1534	7.86621	0.60282	5.05381	5.99681	2.30104
18.1293	9.77963	7.17648	13.4728	4.48788	-1.39991	0.37403	4.24621	8.38705
15.8762	1.5928	3.01336	15.5057	7.74866	-1.91338	2.42923	7.10411	4.14828
8.99267	-1.6444	-4.20646	7.77551	6.45454	0.86078	-4.41161	4.63105	3.1318
10.4195	-3.61903	-2.36646	-9.47933	25.6473	11.7951	-17.8219	14.1785	28.3605
10.679	2.05438	0.5	10.4671	9.74014	-6.3448	-7.15544	1.84454	14.236
5.94088	1.89526	0.20052	5.6259	5.51003	-4.81074	-2.4305	1.13959	8.47813
18.3797	2.24914	3.19848	17.9586	8.11221	-3.1851	2.77883	6.92316	6.79063
13.2996	1.33269	13.0342	-2.28086	6.65812	0.32312	1.82718	-6.39346	9.97321
10.6149	5.48515	8.64685	-2.79476	24.3322	-14.8133	-9.85434	-16.5982	27.5089
6.21065	0.40273	-3.07414	-5.38038	12.0373	-8.90423	-7.12958	3.84303	12.419
14.0765	1.64084	3.77976	13.4594	7.65029	-6.28799	-4.12087	1.41259	15.1381
9.2808	-1.94966	-7.25532	5.44805	4.6726	1.1791	1.28327	4.33416	7.92127
8.4722	1.97601	7.82587	2.57256	4.12802	-3.06666	-2.55515	-1.04698	11.2902

Figure 8: CERN's DiMuon Data set for 2010. Unnecessary Columns deleted to save space for SCILAB. Without doing this step, you will likely run out of memory on the free version.

3. In the SCINotes window, open a new script file. You can save it as DiMuonHistogram. Type in:

   opendata = read_csv('MuRun2010B.csv')

This command will import the data into an array (matrix) that is 20 columns x 100001 rows into your variable space, called opendata. If you want, you can also type this command into the console to see how it imports. Make sure your file browser is in the same folder/directory as your data, for example, your desktop.

ii. Define Variables. In your script, write:

```
E1=opendata(:,1)
E1=strtod(E1)
E2=opendata(:,5)
```

```
E2=strtod(E2)
px1=opendata(:,2)
px1=strtod(px1)
py1=opendata(:,3)
py1=strtod(py1)
pz1=opendata(:,4)
pz1=strtod(pz1)
px2=opendata(:,6)
px2=strtod(px2)
py2=opendata(:,7)
py2=strtod(py2)
pz2=opendata(:,8)
pz2=strtod(pz2)

opendata = [ ]
```

These commands are defining your variables. For example:

E1 = opendata(: ,1)

E1 is the name of the array (matrix) 1 x 100001 values for energy of the first muon.
opendata is the name of the dataset matrix
(:, 1) the : , 1 is an operator in SCILAB that basically says all values in the first column of opendata matrix

The command "strtod" is converting the incoming data to number format, as they are by default imported as text.

Your script should look something like:

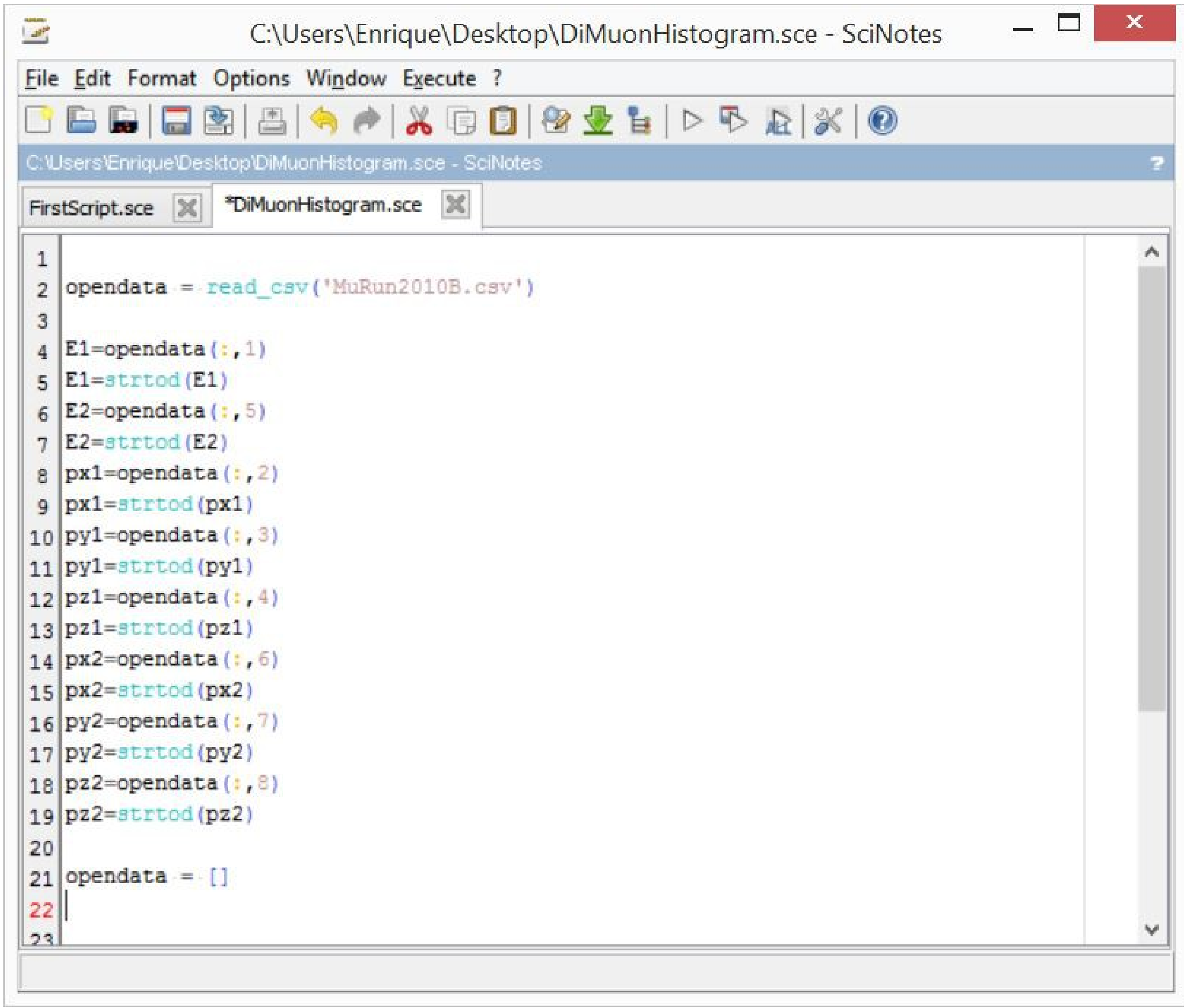


Figure 9: Import and define variables. Note that line 21: opendata =[ ] is the command to delete all of the original data array. This is done to save space because we are a limited, free version.

iii. Process/Calculate Invariant mass.

Type in the following into your script:

```
pxsum= px1+px2
pysum= py1+py2
pzsum= pz1+pz2
presultant=(pxsum.^2+pysum.^2+pzsum.^2).^0.5

mass=((E1+E2).^2-presultant.^2).^.5
```

Should look like this:

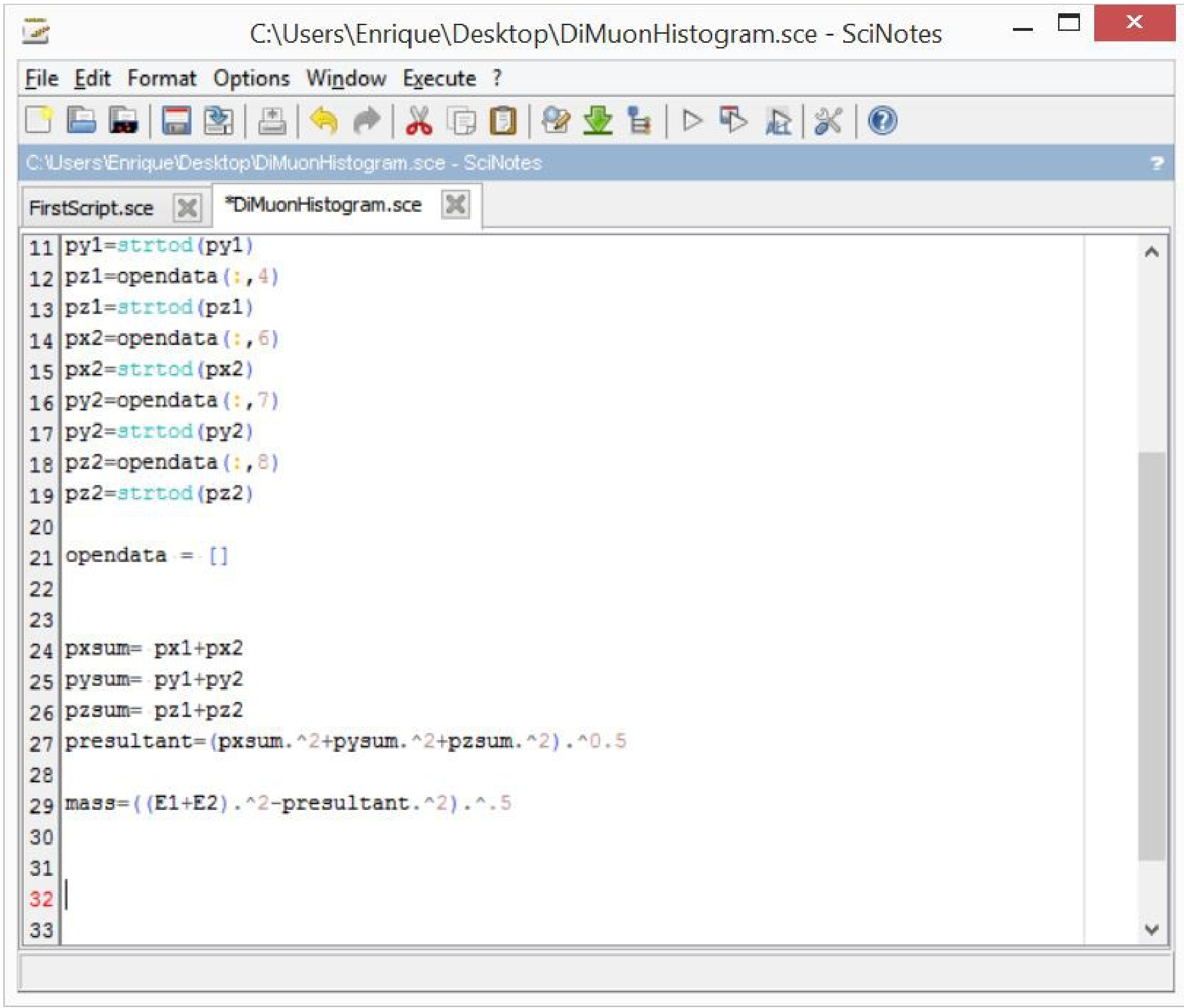
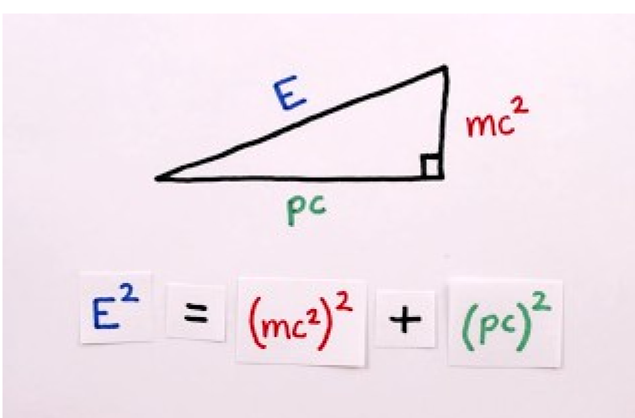


Figure 10: Processing the data, calculating the invariant mass, lines 24-29. 5 lines of code!

If you have already done the EXCEL companion histogram, then you will understand these equations, if not a little physics now:



We can calculate the invariant mass using a right triangle analogy between mass, energy and momentum. This equation simplifies if we set the speed of light, c =1.

We get $E^2 = m^2 + p^2$
Solving for mass m,

$m = (E^2 – p^2)^{1/2}$

When we have 2 muons, we have to add the energies and momenta, so our equation becomes:

$m = [\ (E1+E2)^2 - (P1+P2)^2\ ]^{1/2}$

If you look at line 29: `mass=((E1+E2).^2-presultant.^2).^.5`

That's exactly what we are doing with E1 and E2. But what is going on with momentum? What is presultant? The answer is that momentum is a vector quantity and must be summed in accordance with vector rules. To get a resultant vector, we add the components first, and then get the magnitude of the resultant vector by using the Pythagorean theorem in 3-dimensions.

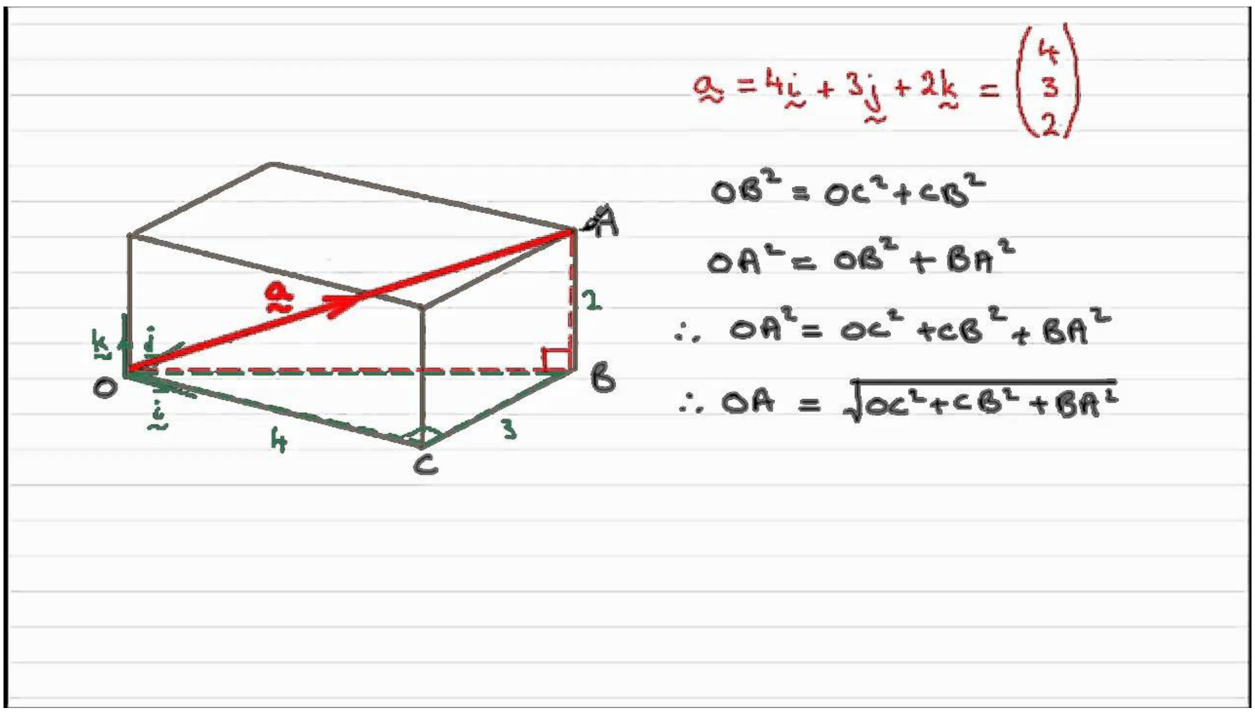


Figure 11: The magnitude of a 3d vector is just the sum of the squares, rooted.

Lines 24-27 Outline the vector sum and resultant

```
pxsum= px1+px2
pysum= py1+py2
pzsum= pz1+pz2
presultant=(pxsum.^2+pysum.^2+pzsum.^2).^0.5
```

==
============SCILAB TIP===
Notice that when you are squaring an array/matrix, you cannot type in x^2, but you need to add a period before the ^, so the correct syntax is x.^2 not x^2

The period tells SCILAB to square each element in the array/matrix term by term. Without the period, SCILAB will square the matrix, a different mathematical operation. The use of the period before multiplication of arrays is also common.

==
==

iv. plotting the Histogram

just type in:

[cf, ind] = histplot(100, mass, normalization=%F);

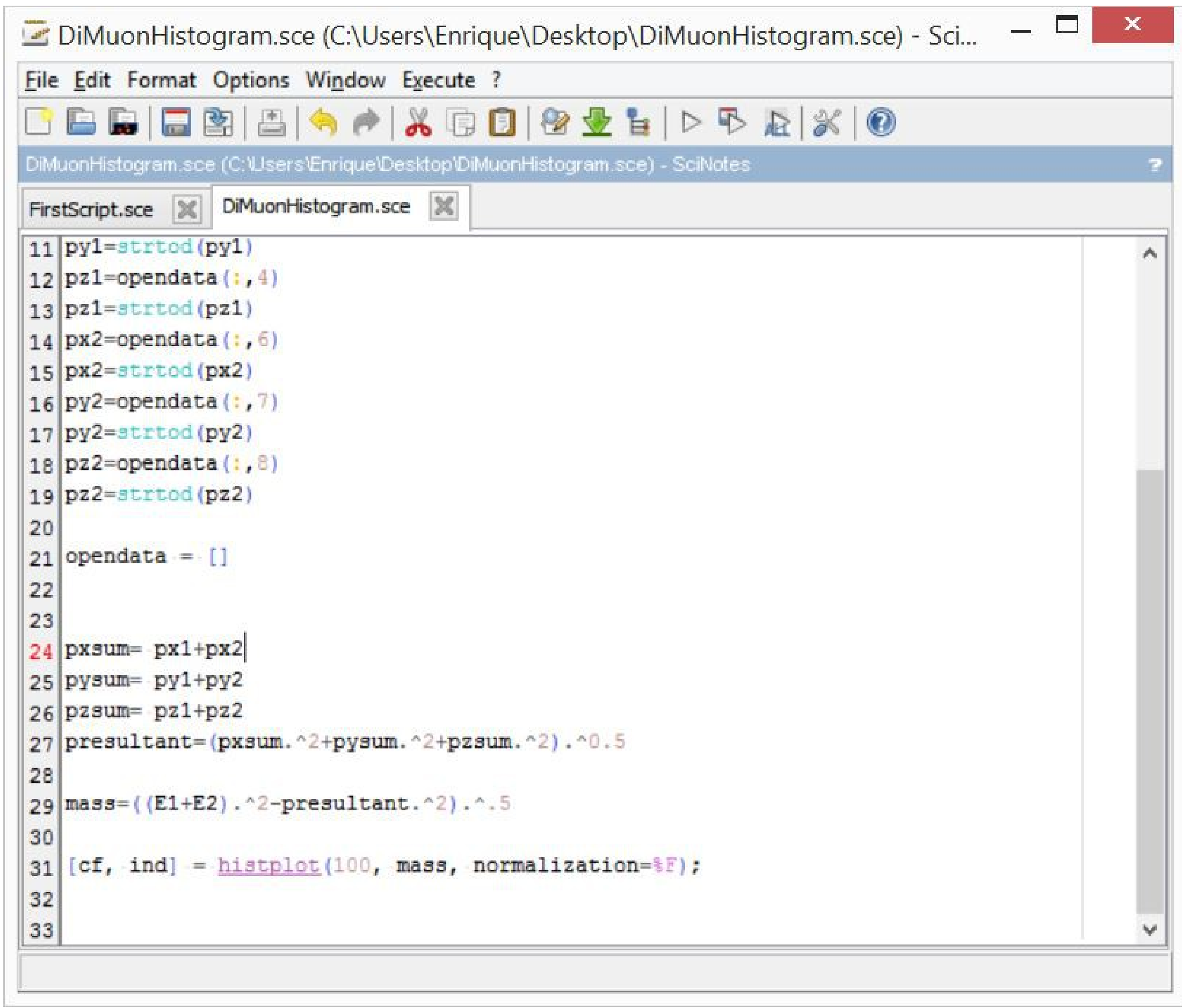


Figure 12: Line 31 uses a pre-built function to plot a histogram. SCILAB has many useful plotting functions. Normalization= false is added because the default plotting option automatically normalizes the data so the area under the curve is 1.0. If you want to see this, just delete the normalization part.

That's it. Time to save and execute the code. Make sure to clear your console first.

This is what happens to the console:

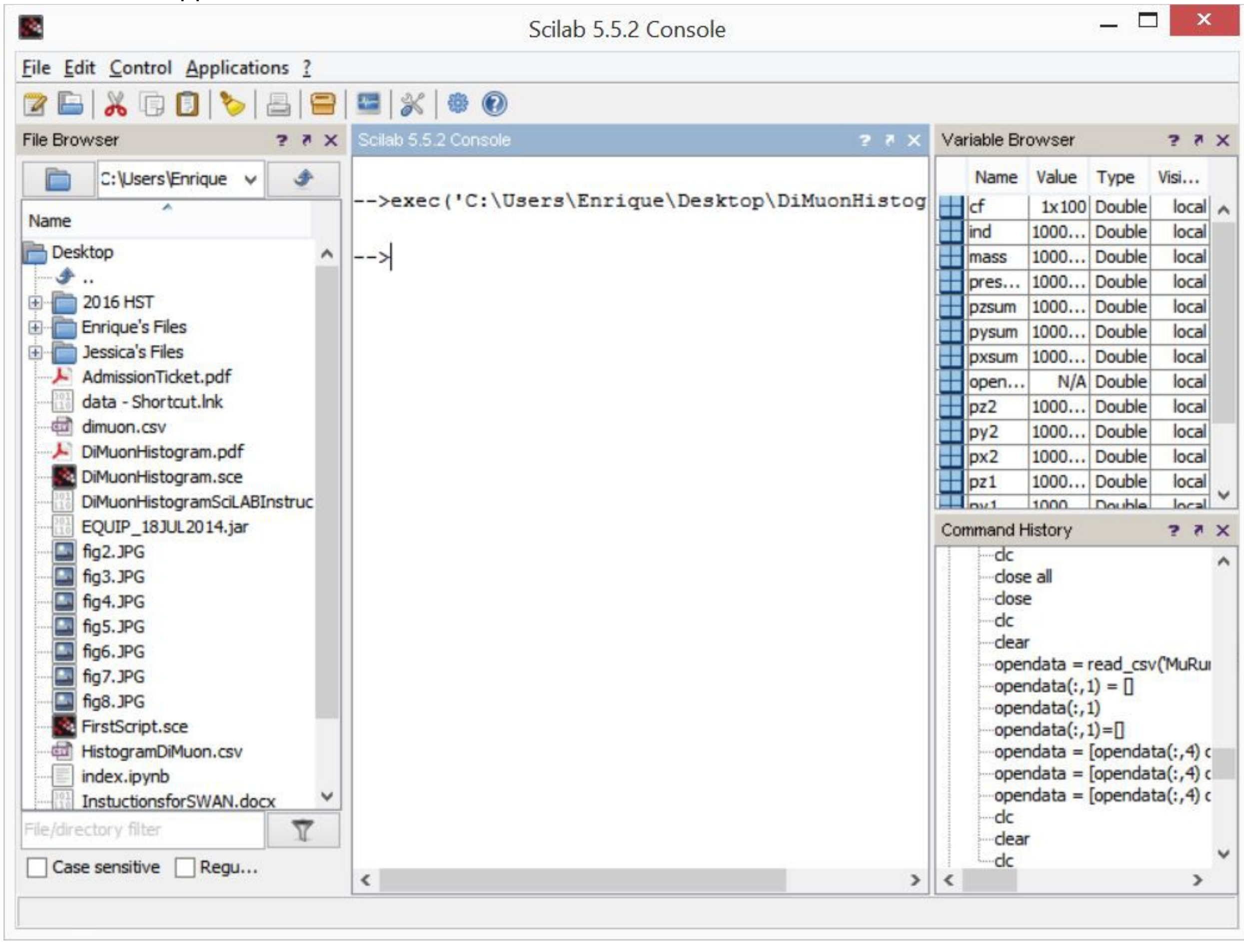


Figure 13: Raw Data: Energy, Momentum, processed data presultant and mass in the variable window.

At this point, you can check to see if you calculated the same invariant mass as was given in the csv document. To view the first few terms of the mass, type:

-->mass(1:20)

*The first value may be a NaN or some other error because it was originally a title.

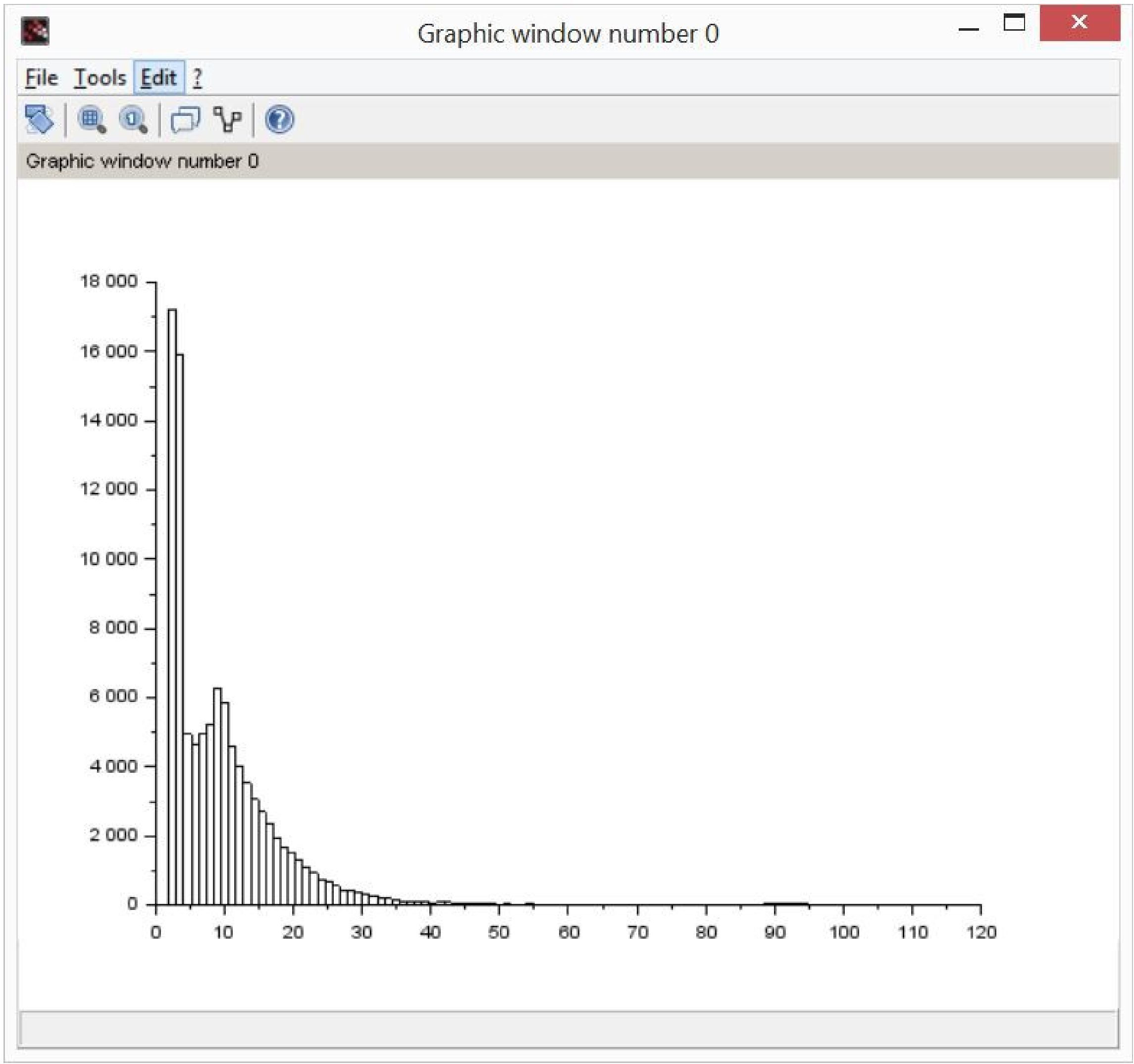


Figure 14: Histogram of invariant mass, not log scale.

At this point, I use the graph menus to add titles and change the y-axis scale to logarithmic. This is so that I get the same graph I made in the EXCEL activity.

Under -> Edit, -> Axis Properties, -> Y tab, change to log scale
In the same window, change the bounds to 1….100000

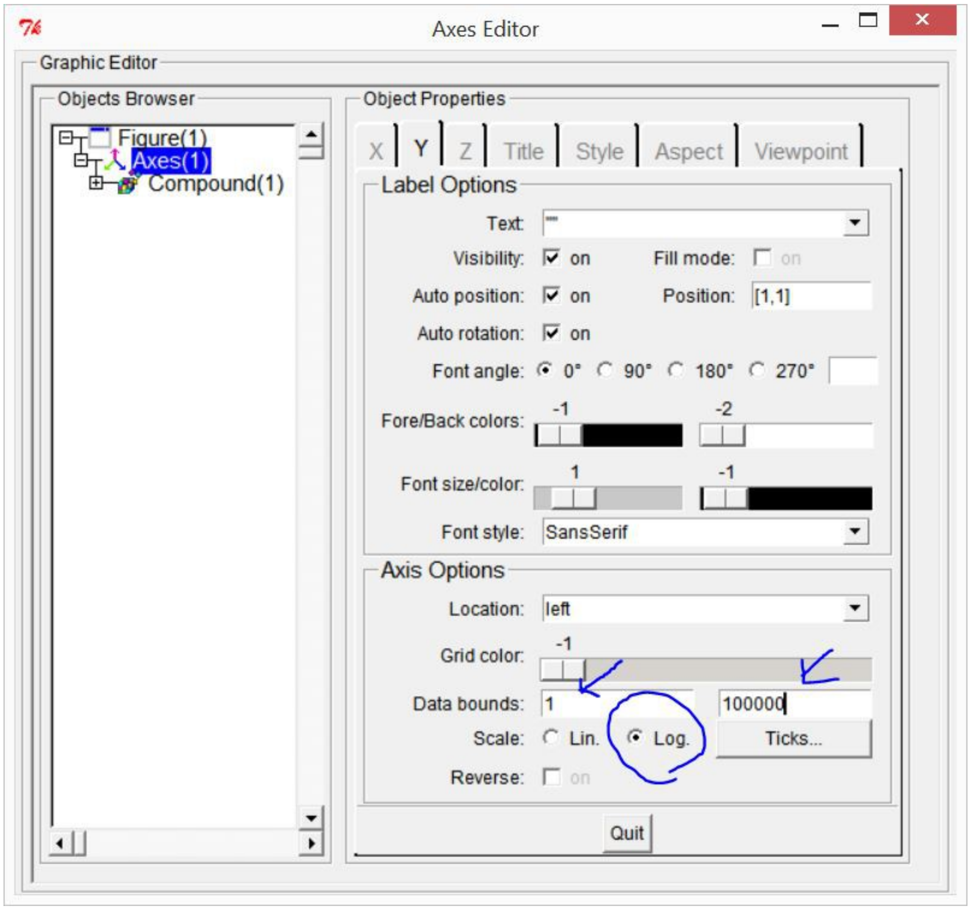


Figure 15: Changing the axis on the Y scale to log, matching the EXCEL Graph

Under ->Edit -> Figure Properties , you can add labels in the Text field. Make sure you type between the quotations.

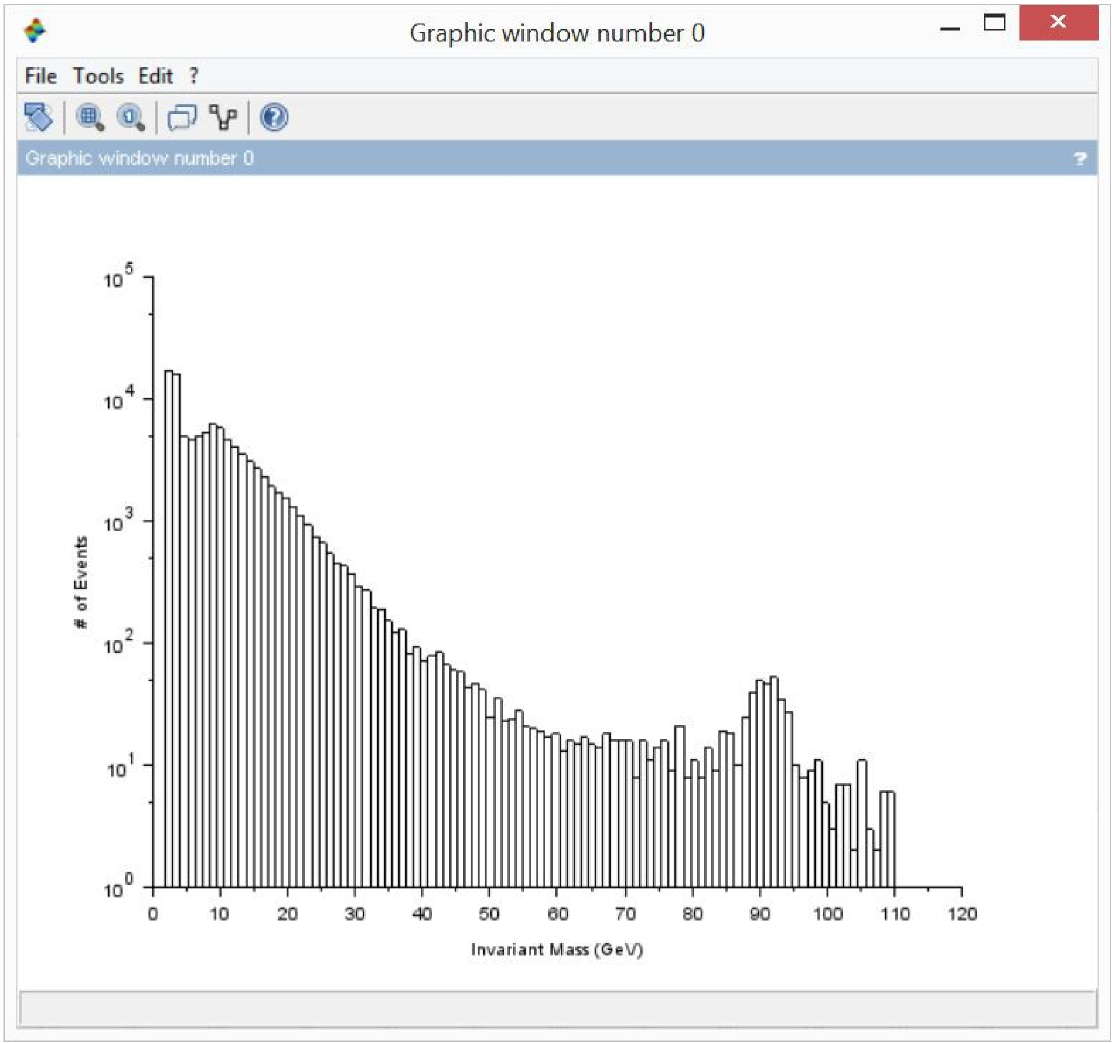


Figure 16: Final Log plot with Labels.

Last step: I export my graph as a pdf since its easy to lose changes. Congrats! You have just made a canonical histogram from actual data taken from CMS at CERN's LHC. What does it mean? Well, there is a bump at about 90 GeV. What is going on there?