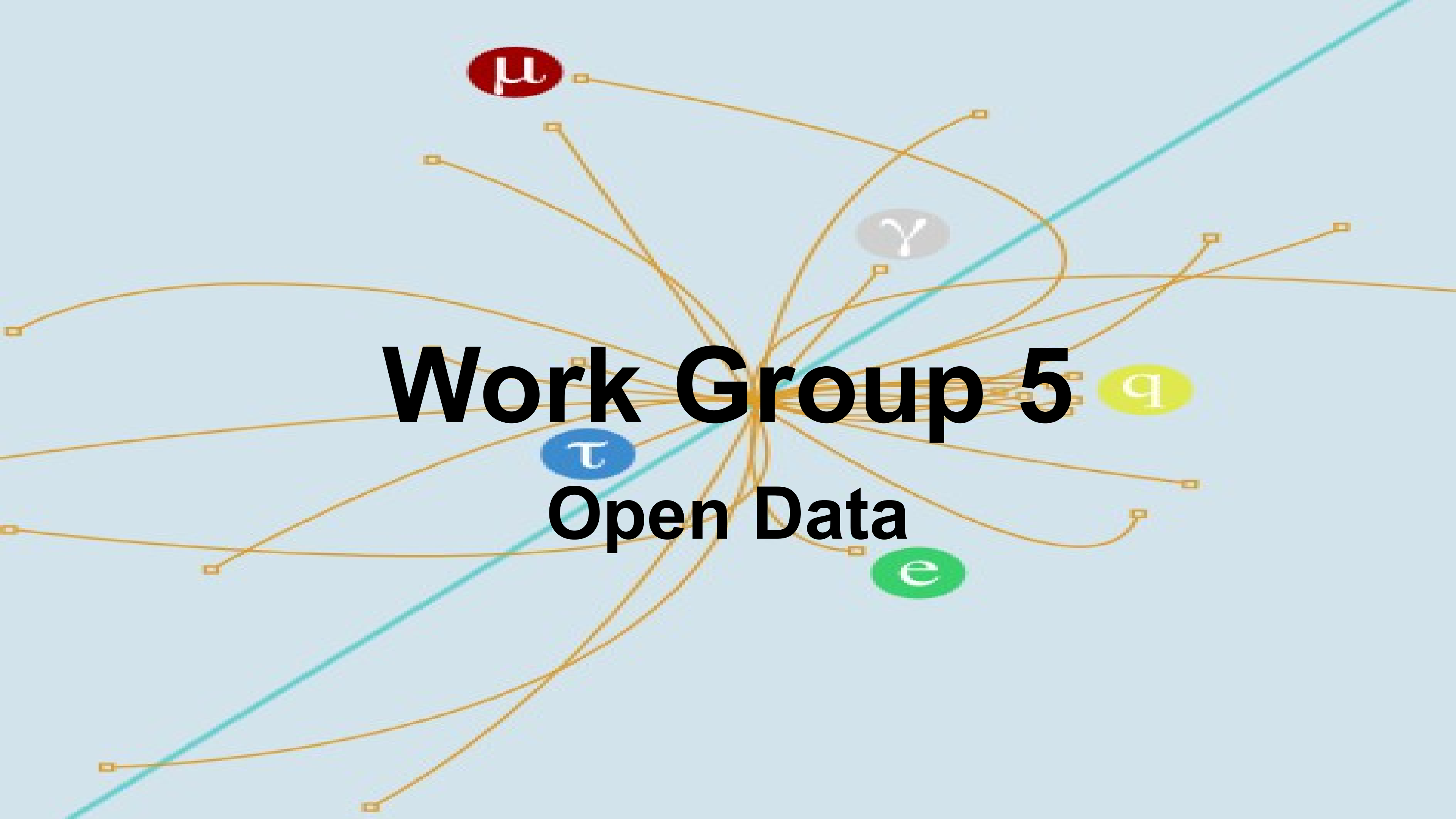
Work Group 5

Open Data



I - Introduction to CERN

1.
What is CERN and what does it do?

2.
Why CERN does what it does?

3.
How do the scientists do research at CERN?
-principles of collisions
-why pt sums up to 0
-detector simulations
-event display
-GEANT4?
-camelia

II - About the Data sets

4.
What is open data? Why does CERN share data?

5.
Where is it used?
(how is it used elsewhere)

6.
An example about the CERN open data file and what it represents

III - Open data exercise

7.
Examples (with dimuon set)(with EXCEL or SciLab)

8.
A more difficult exercise?

IV - For the Teachers:

Instructions for:
-Opening/sharing
/preparing the python(?)
file for the students

Links:

CMS Masterclass:
http://cms.physicsmasterclasses.org/pages/cmswz.html
Info about particles, decays & detection:
http://particleadventure.org/index.html
OtherS:

Outline of Presentation

Maria - Using the Online Visualisation Program

Harri - Using spreadsheets to Analyse CMS Data in the classroom

Enrique - Using Scilab to Analyse CMS Data in the classroom

Enrique - Introduction to Jupyter Notebooks

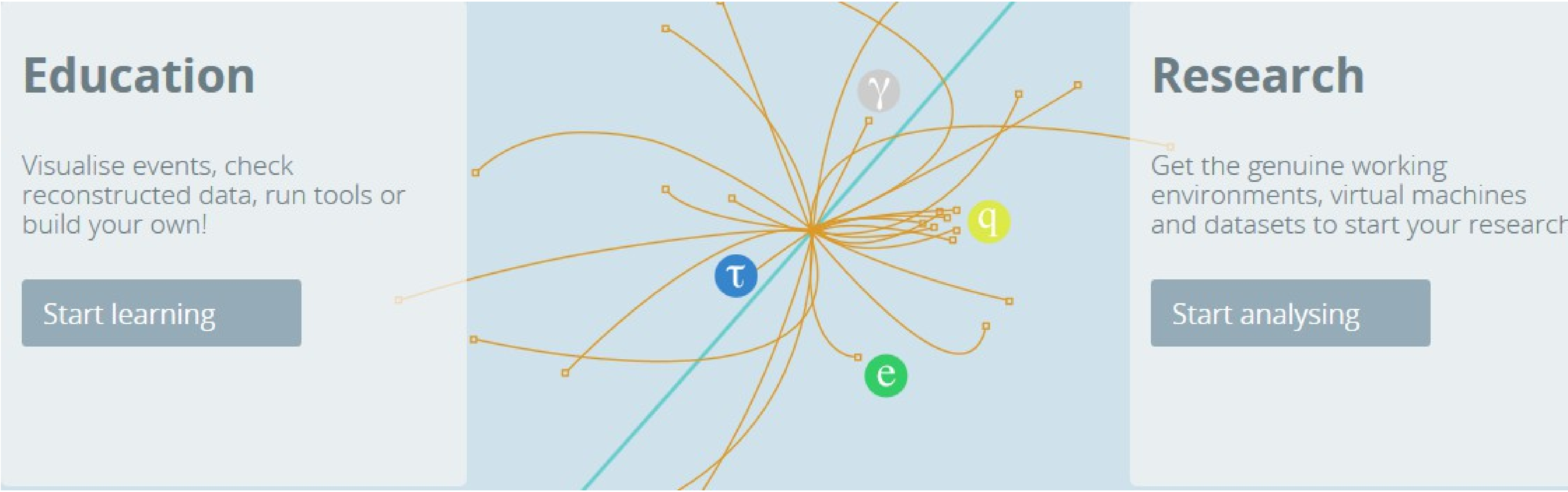
Zorar - Example Analysis of CMS data using Jupyter Notebook (ROOT Framework)

Christoph - Details of using Jupyter Notebooks to Analyse CMS Data

Enrique - How to access and share Jupyter Notebooks

Open Data portal

www.opendata.cern.ch



Open Data portal

Education

The CMS (Compact Muon Solenoid) experiment is one of two large general-purpose detectors built on the Large Hadron Collider (LHC). Its goal is to investigate a wide range of physics such as the characteristics of the Higgs boson, extra dimensions or dark matter.

Explore CMS ›

ALICE (A Large Ion Collider Experiment) is a heavy-ion detector designed to study the physics of strongly interacting matter at extreme energy densities, where a phase of matter called quark-gluon plasma forms. More than 1000 scientists are part of the collaboration.

Explore ALICE ›

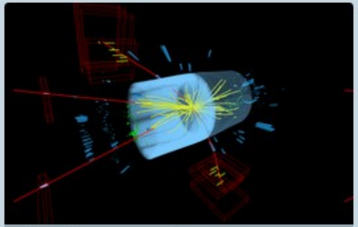
The ATLAS (A Toroidal LHC ApparatuS) experiment is a general-purpose detector exploring topics like the properties of the Higgs-like particle, extra dimensions of space, unification of fundamental forces and evidence for dark matter candidates in the Universe.

Explore ATLAS ›

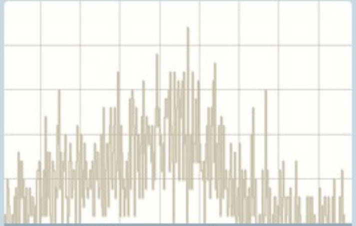
The LHCb (Large Hadron Collider beauty) experiment aims to record the decay of particles containing b and anti-b quarks, known as B mesons. The detector is designed to gather information about the identity, trajectory, momentum and energy of each particle.

Explore LHCb ›

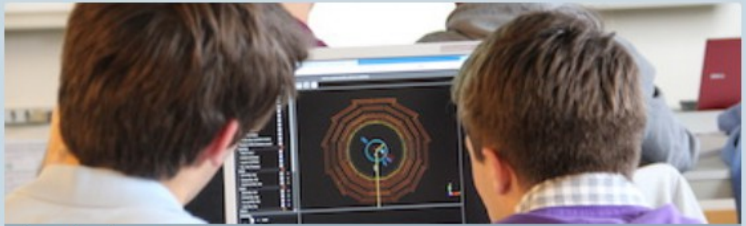
For education purposes, the complex primary data need to be processed into a format (examples below) that is good for simple applications. Get in touch if you wish to build your own applications similar to those shown here



Visualise events ›



Visualise histograms ›



Learning Resources ›

Open Data portal

CMS Derived Datasets

This collection includes data that have been derived from the CMS primary datasets

Years: 2010, 2011

Total records:

59

CMS Tools

This collection includes tools with which the CMS open data can be accessed and used

Years: 2010, 2011

Total records:

17

CMS Learning Resources

This collection includes learning resources that use CMS public data

Total records:

6

CMS Open Data Instructions

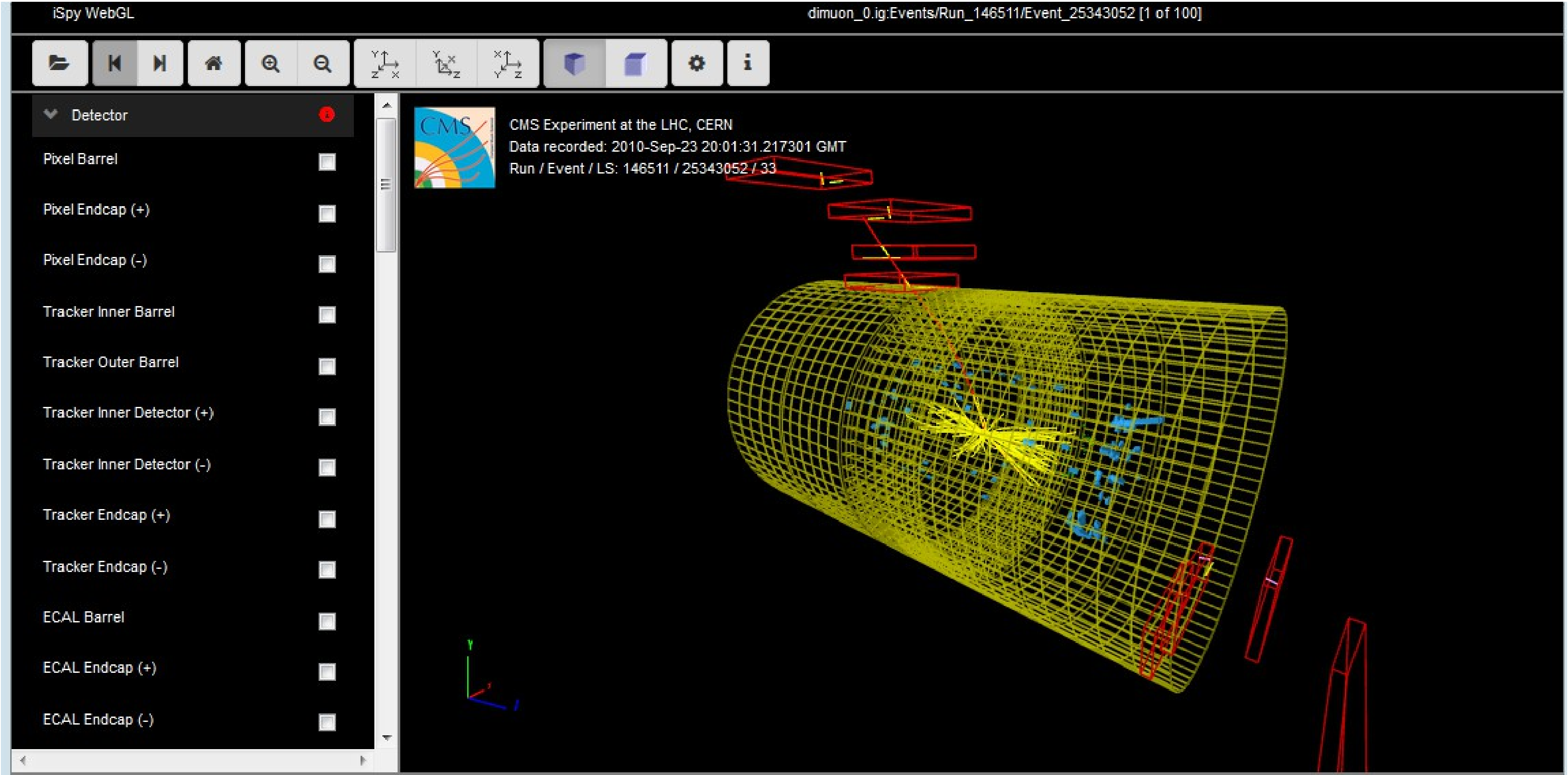
This collection contains CMS open data instructions.

Years: 2010, 2011

Total records:

7

Visualise events



Resources we have created

Document for high School teachers that includes:

Instructions for the use of the visualisation of events, including the description of the CMS detector.

Activity 1: Classroom activity in which students will explore the CMS detector and a dimuon event.

Activity 2: Calculation of invariant mass with the data that can be obtained from the even visualisation display.

Resources: Instructions for the event display

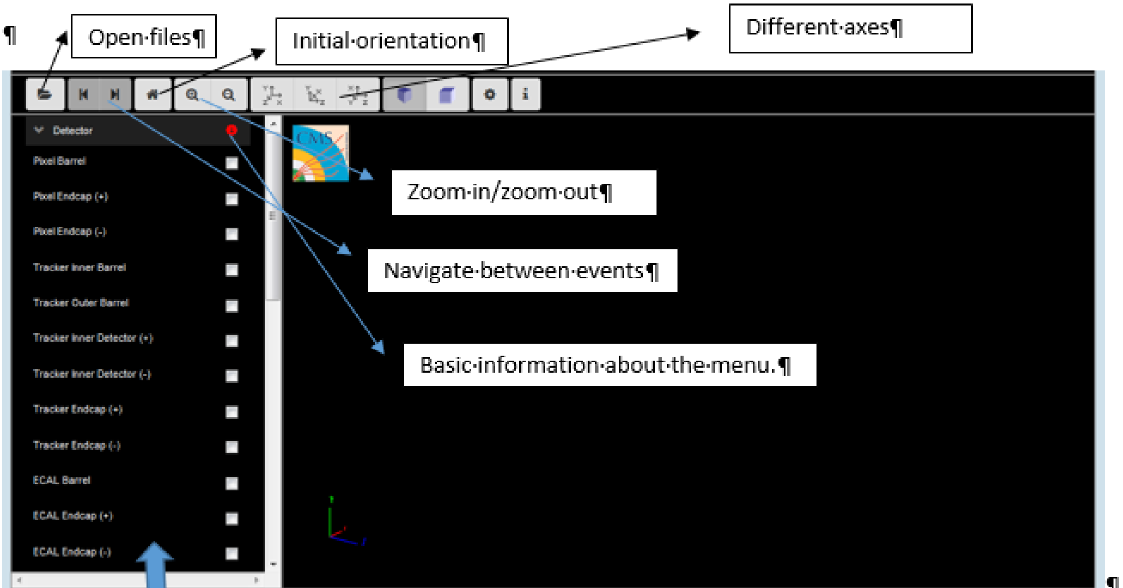
VISUALISATION OF DATA FROM CMS DETECTOR

VISUALISATION DISPLAY

The CMS visual display can be used to visualize data obtained in the CMS detector. It can be found here:

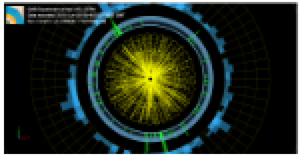
http://opendata.cern.ch/visualise/events/CMS

Some basics about the functionalities of the top up buttons.



Activity 1: Hunting particles after the collision

ACTIVITY ONE: HUNTING PARTICLES AFTER A COLLISION



Description of the activity

In this activity students will explore the particles obtained from a two proton collision in the CMS detector. For this, they will use the even visualizer

Objective

To understand the basics of the CMS detector system and how and where the particles are detected. This activity will also consolidate knowledge about how the properties of the particles can be used to detect them in different ways.

Needed background

Students need to understand the basics about the Large Hadron Collider and the CMS detector.

- Suggested video about the LHC: https://www.youtube.com/watch?v=pQhbhpU9Wrg
- Suggested video about the CMS: https://www.youtube.com/watch?v=S99d9BQmGB0

etc...

Activity 2: Calculation of invariant mass

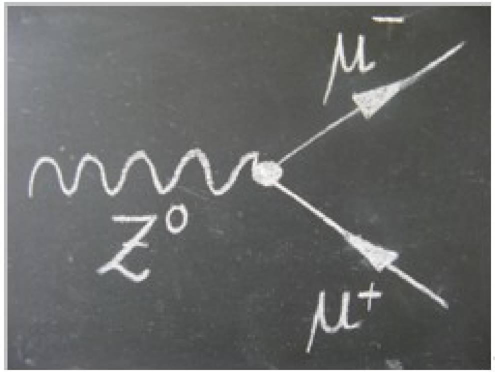
Objective

Learn how to calculate invariant mass in a two-particle decay.

Needed background

The file used for this activity is found in the Education folder of the visualization event: dimuon_0.ig

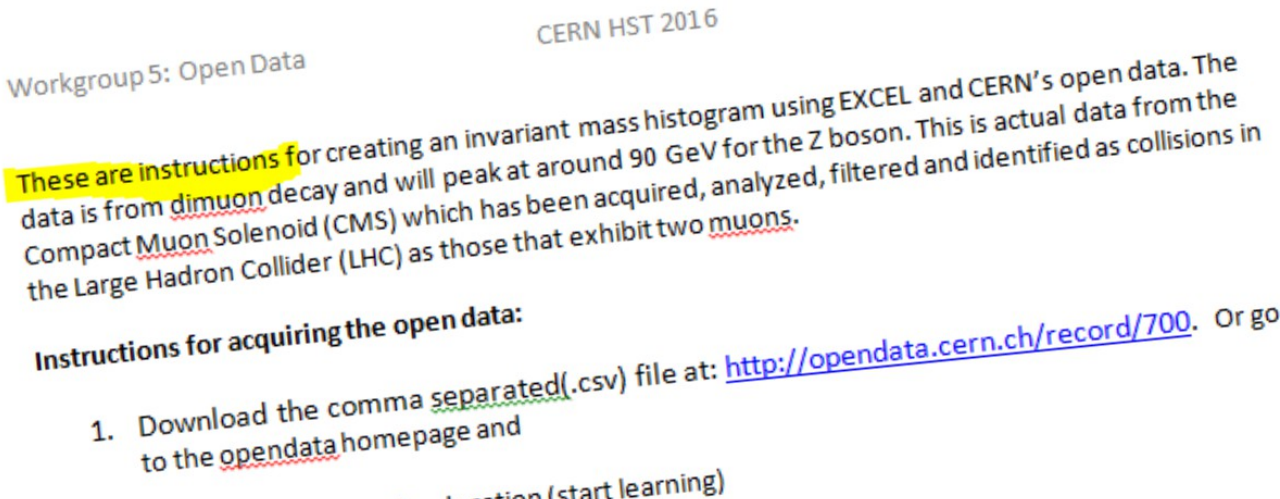
This data corresponds to the decay of a Z boson into two muons (muon-antimuon). The Z boson is an exchange particle mediating weak interactions. Z is neutral, so, the sum of the charges of its decay products must be zero. When the Z boson decays there are different possibilities but we will only focus on the dimuon decay.[1]



Feynman diagram of the decay of the Z boson into two muons.

Using spreadsheets to Analyse CMS Data in the classroom

You can find the instructio



Workgroup 5: Open Data CERN HST 2016

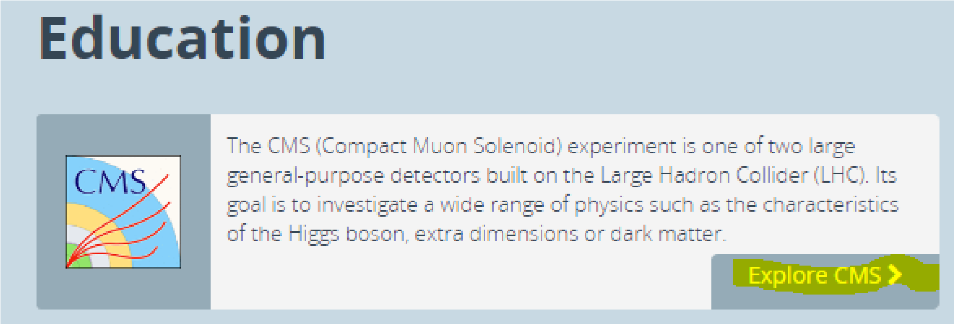
These are instructions for creating an invariant mass histogram using EXCEL and CERN's open data. The data is from dimuon decay and will peak at around 90 GeV for the Z boson. This is actual data from the Compact Muon Solenoid (CMS) which has been acquired, analyzed, filtered and identified as collisions in the Large Hadron Collider (LHC) as those that exhibit two muons.

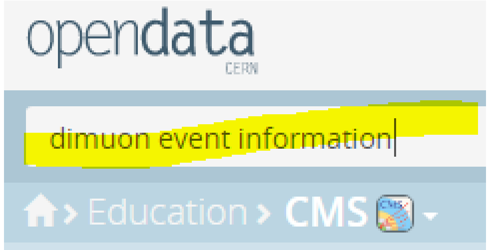
Instructions for acquiring the open data:

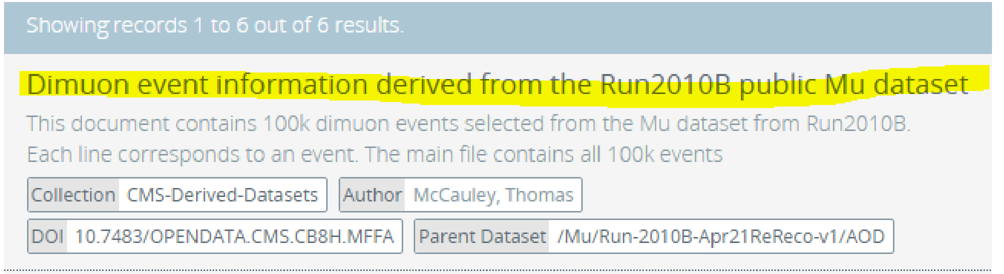
1. Download the comma separated(.csv) file at: http://opendata.cern.ch/record/700. Or go to the opendata homepage and

...ation (start learning)

Step 1: How to get the data



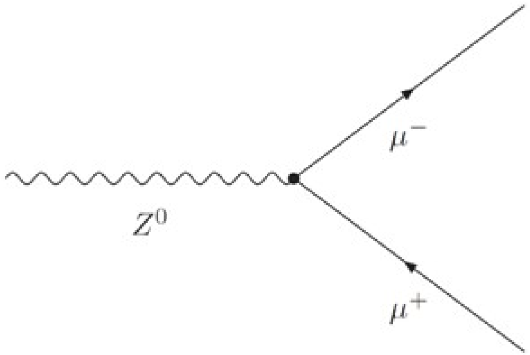


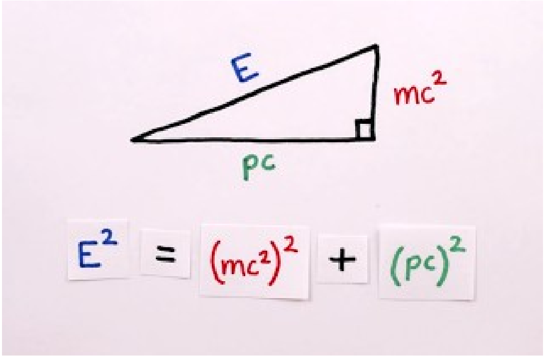


Files

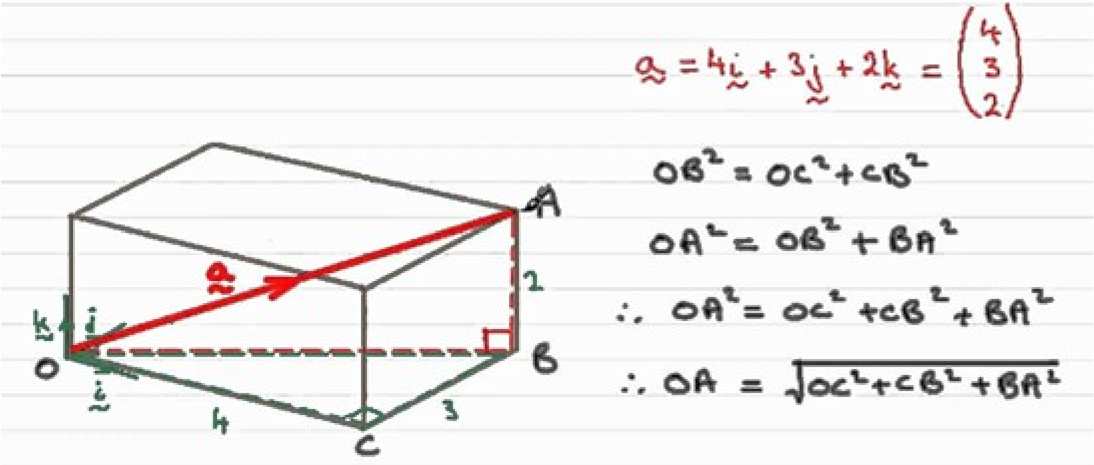
MuRun2010B_9.csv	**Size:** 1.5 MB	Download
MuRun2010B_8.csv	**Size:** 1.5 MB	Download
MuRun2010B_3.csv	**Size:** 1.5 MB	Download
MuRun2010B_2.csv	**Size:** 1.5 MB	Download
MuRun2010B_1.csv	**Size:** 1.5 MB	Download
MuRun2010B_0.csv	**Size:** 1.5 MB	Download
MuRun2010B_7.csv	**Size:** 1.5 MB	Download
MuRun2010B_6.csv	**Size:** 1.5 MB	Download
MuRun2010B_5.csv	**Size:** 1.5 MB	Download
MuRun2010B_4.csv	**Size:** 1.5 MB	Download
MuRun2010B.csv	**Size:** 15.2 MB	Download

Step 2: Doodle around with physics

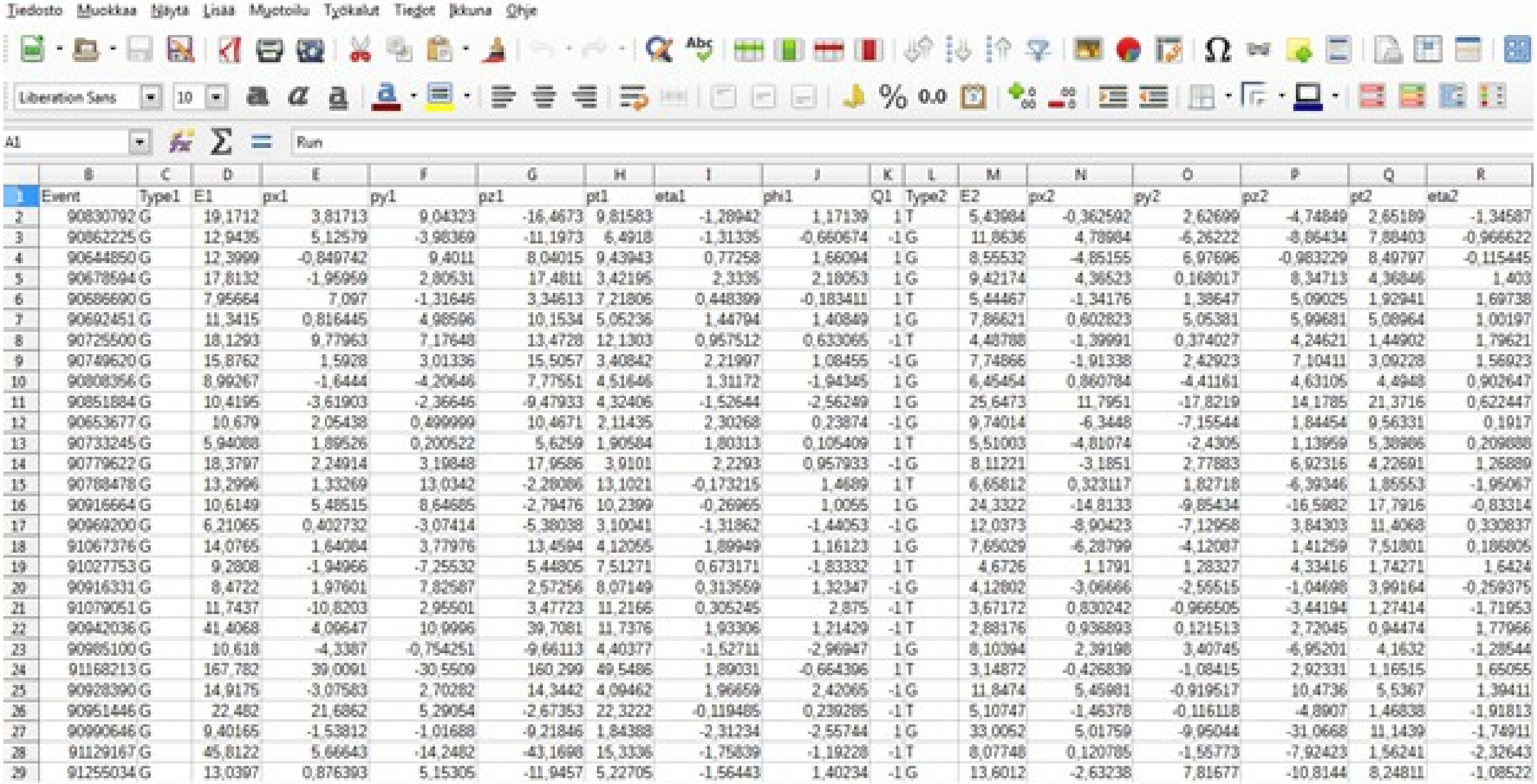




$$m = \sqrt{(E_1 + E_2)^2 - (\bar{p}_1 + \bar{p}_2)^2}$$



Step 3: Play around with the data



1	Event	Type1	E1	px1	py1	pz1	pt1	eta1	phi1	Q1	Type2	E2	px2	py2	pz2	pt2	eta2
2	90830792	G	19,1712	3,81713	9,04323	-16,4673	9,81583	-1,28942	1,17139	1	T	5,43984	-0,362592	2,62699	-4,74849	2,65189	-1,34587
3	90862225	G	12,9435	5,12579	-3,98369	-11,1973	6,4918	-1,31335	-0,660674	-1	G	11,8636	4,78984	-6,26222	-8,86434	7,88403	-0,966622
4	90644850	G	12,3999	-0,849742	9,4011	8,04015	9,43943	0,77258	1,66094	1	G	8,55532	-4,85155	6,97696	-0,983229	8,49797	-0,115445
5	90678594	G	17,8132	-1,95959	2,80531	17,4811	3,42195	2,3335	2,18053	1	G	9,42174	4,38523	0,168017	8,34713	4,38846	1,403
6	90686690	G	7,95664	7,097	-1,31646	3,34613	7,21806	0,448399	-0,183411	1	T	5,44467	-1,34176	1,38647	5,09025	1,92941	1,69738
7	90692451	G	11,3415	0,816445	4,98596	10,1534	5,05236	1,44794	1,40849	1	G	7,86621	0,602823	5,05381	5,99681	5,08964	1,00197
8	90725500	G	18,1293	9,77963	7,17648	13,4728	12,1303	0,957512	0,633065	-1	T	4,48788	-1,39991	0,374027	4,24621	1,44902	1,79621
9	90749620	G	15,8762	1,5928	3,01336	15,5067	3,40842	2,21997	1,08455	-1	G	7,74866	-1,91338	2,42923	7,10411	3,09228	1,56923
10	90808356	G	8,99267	-1,6444	-4,20646	7,77551	4,51646	1,31172	-1,94345	1	G	6,45454	0,860784	-4,41161	4,63106	4,4948	0,902647
11	90851884	G	10,4195	-3,61903	-2,36646	-9,47933	4,32406	-1,52644	-2,56249	1	G	25,6473	11,7951	-17,8219	14,1785	21,3716	0,622447
12	90653677	G	10,679	2,05438	0,499999	10,4671	2,11435	2,30268	0,23874	-1	G	9,74014	-6,3448	-7,15544	1,84454	9,56331	0,1917
13	90733245	G	5,94088	1,89526	0,200522	5,6259	1,90584	1,80313	0,105409	1	T	5,51003	-4,81074	-2,4305	1,13959	5,38986	0,209888
14	90779622	G	18,3797	2,24914	3,19848	17,9586	3,9101	2,2293	0,957933	-1	G	8,11221	-3,1851	2,77883	6,92316	4,22691	1,26889
15	90788478	G	13,2996	1,33269	13,0342	-2,28086	13,1021	-0,173215	1,4689	1	T	6,65812	0,323117	1,82718	-6,39346	1,85553	-1,95067
16	90916664	G	10,6149	5,48515	8,64685	-2,79476	10,2399	-0,26965	1,0055	1	G	24,3322	-14,8133	-9,85434	-16,5982	17,7916	-0,83314
17	90969200	G	6,21065	0,402732	-3,07414	-5,38038	3,10041	-1,31862	-1,44053	-1	G	12,0373	-8,90423	-7,12958	3,84303	11,4068	0,330837
18	91067376	G	14,0765	1,64084	3,77976	13,4594	4,12055	1,89949	1,16123	1	G	7,65029	-6,28799	-4,12087	1,41259	7,51801	0,186806
19	91027753	G	9,2808	-1,94966	-7,25532	5,44805	7,51271	0,673171	-1,83332	1	T	4,6726	1,1791	1,28327	4,33416	1,74271	1,6424
20	90916331	G	8,4722	1,97601	7,82587	2,57256	8,07149	0,313559	1,32347	-1	G	4,12802	-3,06666	-2,55515	-1,04698	3,99164	-0,259375
21	91079051	G	11,7437	-10,8203	2,95501	3,47723	11,2166	0,305245	2,875	-1	T	3,67172	0,830242	-0,966505	-3,44194	1,27414	-1,71953
22	90942036	G	41,4068	4,09647	10,9996	39,7081	11,7376	1,93306	1,21429	-1	T	2,88176	0,936893	0,121513	2,72045	0,94474	1,77966
23	90985100	G	10,618	-4,3387	-0,754251	-9,66113	4,40377	-1,52711	-2,96947	1	G	8,10394	2,39198	3,40745	-6,95201	4,1632	-1,28544
24	91168213	G	167,782	39,0091	-30,5509	160,299	49,5486	1,89031	-0,664396	1	T	3,14872	-0,426839	-1,08415	2,92331	1,16515	1,65055
25	90928390	G	14,9175	-3,07583	2,70282	14,3442	4,09462	1,96659	2,42065	-1	G	11,8474	5,45981	-0,919517	10,4736	5,5367	1,39411
26	90951446	G	22,482	21,6862	5,29054	-2,67353	22,3222	-0,119485	0,239285	-1	T	5,10747	-1,46378	-0,116118	-4,8907	1,46838	-1,91813
27	90990646	G	9,40165	-1,53812	-1,01688	-9,21846	1,84388	-2,31234	-2,55744	1	G	33,0062	5,01759	-9,95044	-31,0668	11,1439	-1,74911
28	91129167	G	45,8122	5,66643	-14,2482	-43,1698	15,3336	-1,75839	-1,19228	-1	T	8,07748	0,120785	-1,55773	-7,92423	1,56241	-2,32643
29	91256034	G	13,0397	0,876393	5,15305	-11,9457	5,22705	-1,56443	1,40234	-1	G	13,6012	-2,63238	7,81677	-10,8144	8,24811	-1,08522

	A	B	C	D	E	F	G	H	I	J	K
1	Run	Event	Type1	E1	px1	py1	pz1	pt1	eta1	phi1	Q1
2	146436	90830792	G	19,1712	3,81713	9,04323	-16,4673	9,81583	-1,28942	1,17139	1
3	146436	90862225	G	12,9435	5,12579	-3,98369	-11,1973	6,4918	-1,31335	-0,660674	-1
4	146436	90644850	G	12,3999	-0,849742	9,4011	8,04015	9,43943	0,77258	1,66094	1
5	146436	90678594	G	17,8132	-1,95959	2,80531	17,4811	3,42195	2,3335	2,18053	1
6	146436	90686690	G	7,95664	7,097	-1,31646	3,34613	7,21806	0,448399	-0,183411	1
7	146436	90692451	G	11,3415	0,816445	4,98596	10,1534	5,05236	1,44794	1,40849	1
8	146436	90725500	G	18,1293	9,77963	7,17648	13,4728	12,1303	0,957512	0,633065	-1

L	M	N	O	P	Q	R	S	T	U
Type2	E2	px2	py2	pz2	pt2	eta2	phi2	Q2	M
T	5,43984	-0,362592	2,62699	-4,74849	2,65189	-1,34587	1,70796	1	2,73205
G	11,8636	4,78984	-6,26222	-8,86434	7,88403	-0,966622	-0,917841	1	3,10256
G	8,55532	-4,85155	6,97696	-0,983229	8,49797	-0,115445	2,17841	-1	9,41149
G	9,42174	4,36523	0,168017	8,34713	4,36846	1,403	0,0384708	1	7,74702
T	5,44467	-1,34176	1,38647	5,09025	1,92941	1,69738	2,33981	-1	8,67727
G	7,86621	0,602823	5,05381	5,99681	5,08964	1,00197	1,45208	-1	2,30104
T	4,48788	-1,39991	0,374027	4,24621	1,44902	1,79621	2,88051	1	8,38705

	A	B	J	K
1	Run	Event	phi1	Q1
2	146436	908	1,17139	1
3	146436	908	-0,660674	-1
4	146436	906	1,66094	1
5	146436	90	2,18053	1
6	146436	90	-0,183411	1
7	146436	90	1,40849	1
8	146436	90	0,633065	-1

Glossary for the columns:

- Run: The data collected by CMS in a given year are divided into sets called "Runs". For example, the data from 2010 were divided into "RunA" and "RunB", the latter being from the second part of the year. Run is a data collection period with a variable length of time
- Event: One proton - proton collision
- Type1: G = global muon, T = tracker muon (observed only in tracker)
- E1: energy of muon number 1
- px1: momentum in the direction of the x-axis of muon number 1
- py1: momentum in the direction of the y-axis of muon number 1
- pz1: momentum in the direction of the z-axis of muon number 1
- pt1: transverse momentum of muon number 1
- eta1 = in experimental particle physics, pseudorapidity, η, is a commonly used spatial coordinate describing the angle of a particle relative to the beam axis. It is defined as

$$\eta = -\ln[\tan\left(\frac{\theta}{2}\right)]$$

- phi1 = is the azimuthal angle, φ, and is measured from the positive x-axis in the x-y-plane.
- Q1: charge of muon number 1
- Type2: G = global muon, T = tracker muon (observed only in tracker)
- E2: energy of muon number 2
- px2: momentum in the direction of the x-axis of muon number 2
- py2: momentum in the direction of the y-axis of muon number 2
- pz2: momentum in the direction of the z-axis of muon number 2
- pt2: transverse momentum of muon number 2
- eta2 = in experimental particle physics, pseudorapidity, η, is a commonly used spatial coordinate describing the angle of a particle relative to the beam axis. It is defined as

$$\eta = -\ln[\tan\left(\frac{\theta}{2}\right)]$$

- phi2 = is the azimuthal angle, φ, and is measured from the positive x-axis in the x-y-plane.
- Q2: charge of muon number 2
- M: Invariant mass

L	M	T	U
Type2	E2	Q2	M
T	5,43984	1	2,73205
G	11,8636	1	3,10256
G	8,5553	-1	9,41149
G	9,4217	1	7,74702
T	5,4446	-1	8,67727
G	7,8662	-1	2,30104
T	4,4878	1	8,38705

Note: Energy and momenta are

Step 4: Overcome cultural differences...

19,1712

19.1712

12.9435

12,9435

5.43984

5,43984

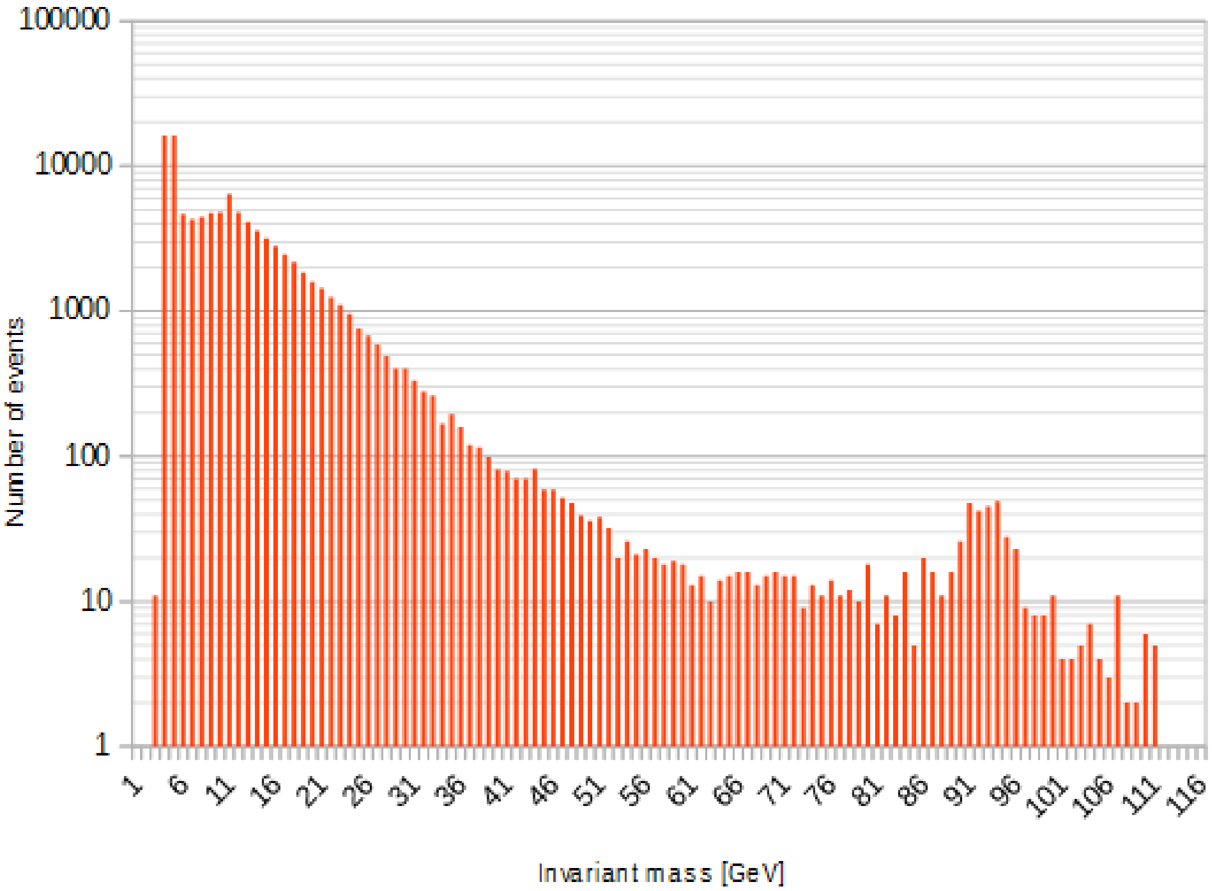
Step 5: Get to work

We can now use our physics skills to

Vector sum Px	Vector sum Py	Vector sum Pz	Resultant momentum	Invariant mass
3,454538	11,67022	-21,21579	24,4588963971	2,7323390924
9,91563	-10,24591	-20,06164	24,6123503452	3,1023250789
-5,701292	16,37806	7,056921	18,7228660689	9,4114574544
2,40564	2,973327	25,82823	26,1098686743	7,7470455409
5,75524	0,07001	8,43638	10,2127467589	8,677264163
1,419268	10,03977	16,15021	19,0693625052	2,3012034007
8,37972	7,550507	17,71901	21,0045989873	8,3871122883
-0,32058	5,44259	22,60981	23,2578602988	4,1480048625

End result

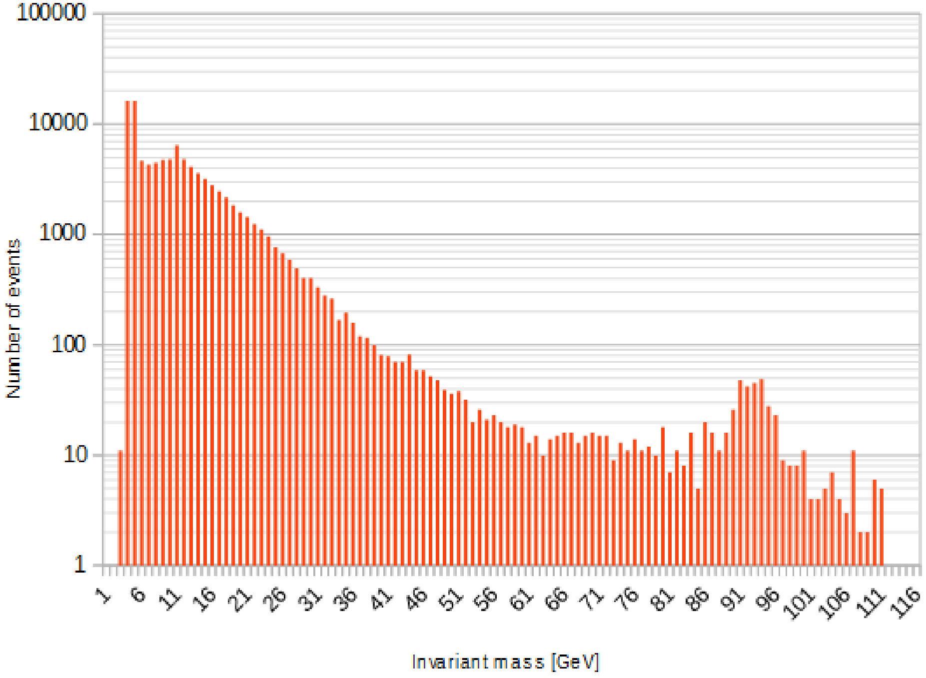
5 σ



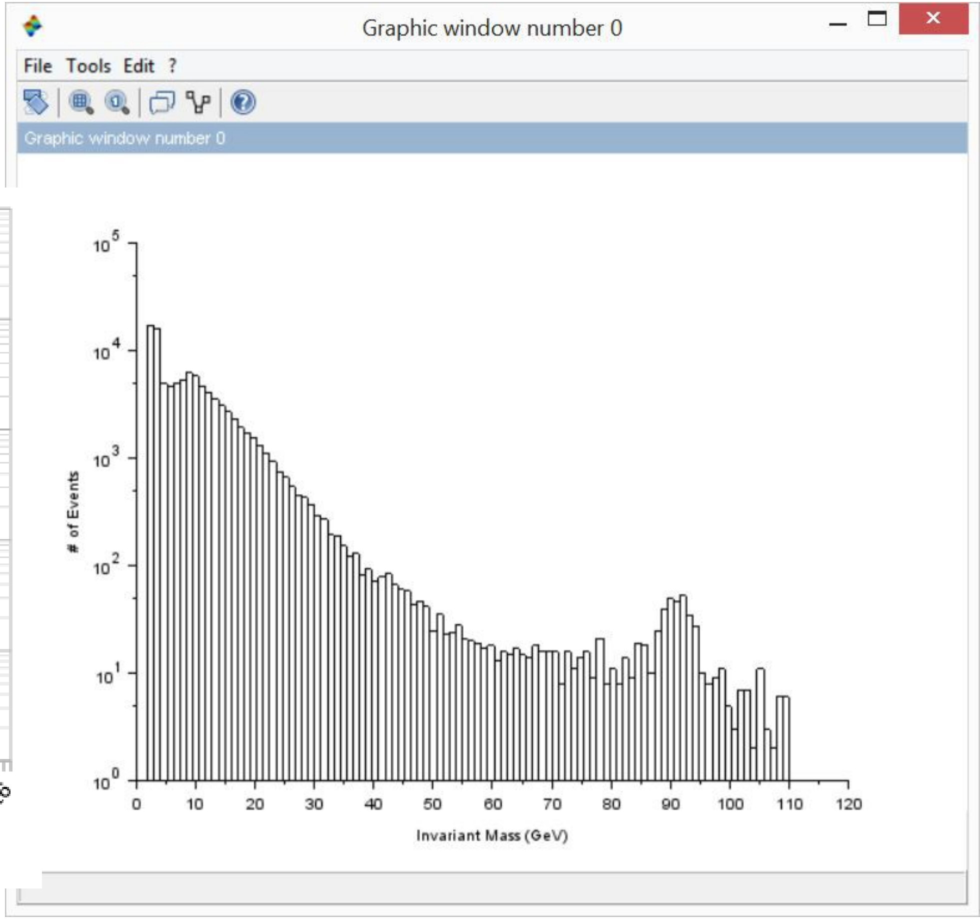
Doing the same thing with SCILAB

Students will:

- create histograms of invariant mass vs # of events using Excel and SCILAB (the same way the Higgs Boson was discovered by CMS and ATLAS experiments at LHC)
- calculate invariant mass using relativistic mechanics (included in the IB Physics option: Relativity)
- use real data from LHC collisions collected by the CMS facility (Covers IB ICT requirement)
- vector sum momentum and calculate a resultant vector



100000
10000
1000
100
10
1
Number of events
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116
Invariant mass [GeV]

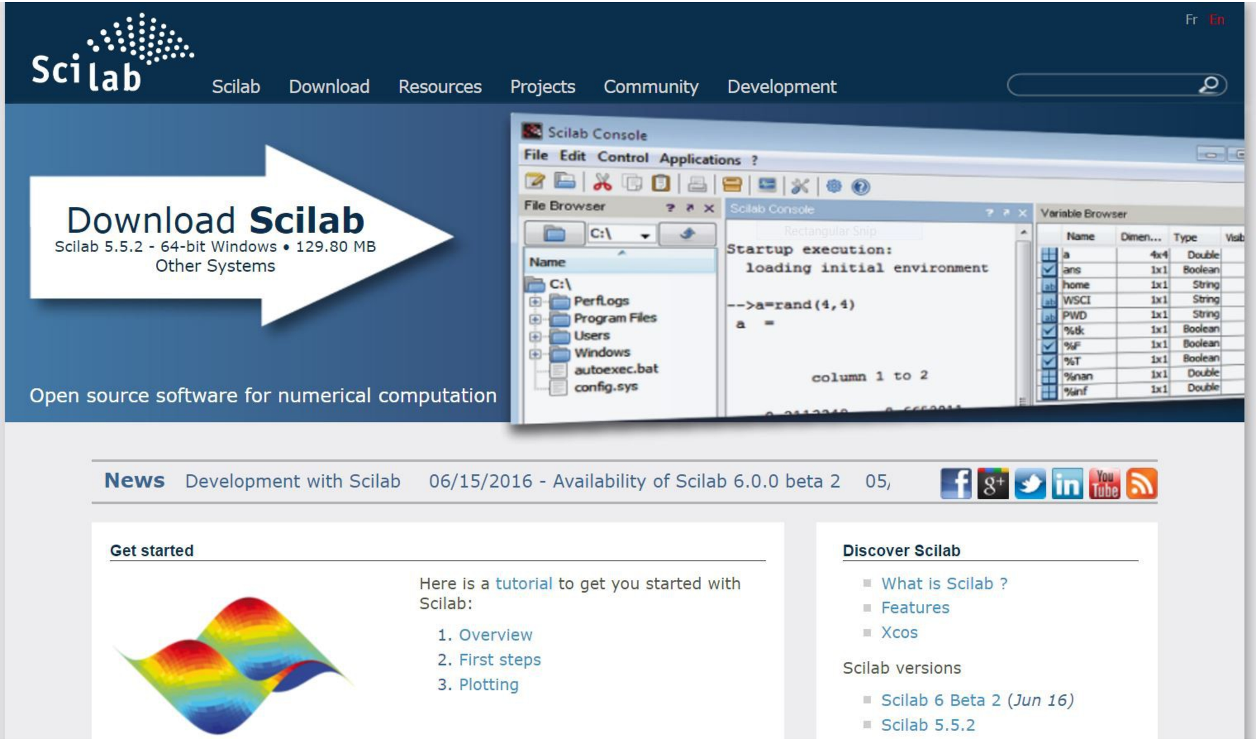


Graphic window number 0
File Tools Edit ?
Graphic window number 0
10^5
10^4
10^3
10^2
10^1
10^0
# of Events
0 10 20 30 40 50 60 70 80 90 100 110 120
Invariant Mass (GeV)

What is SCILAB?

Like MATLAB, just free

- less libraries and
- smaller data sets



What did I develop? - 3 Word Files

Instruction Handout

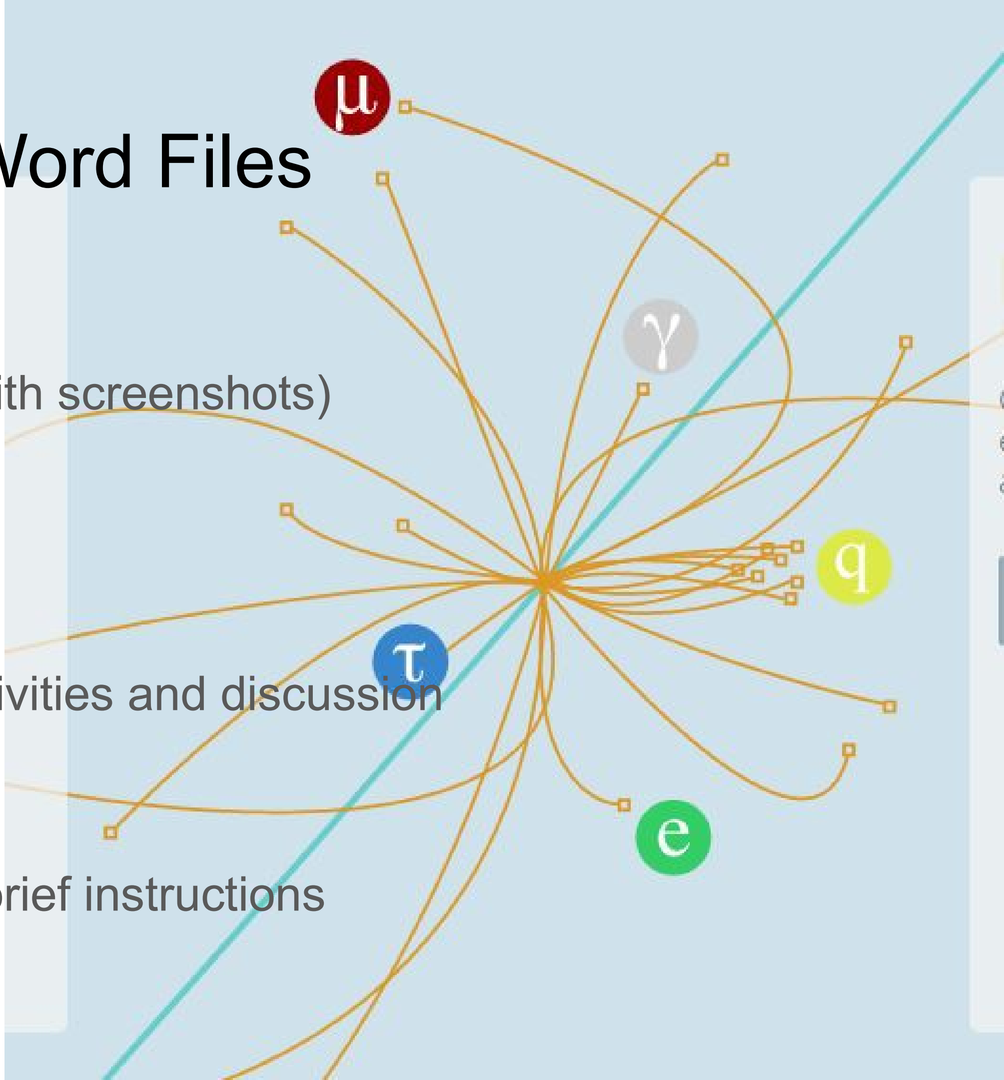
- step by step to make histogram (with screenshots)
- basic tutorial on using SCILAB

Teacher Handout

- suggested pre and post lesson activities and discussion

Student Handout

- assignment and lab handout with brief instructions



SCILAB Script

Approximately 30 lines of code, 4 parts:

- Importing data from .csv
- Defining variables
- Processing Data (Calc Mass)
- Plot Histogram

Students learn:

- how to import .csv files
- process data/run calculations

DiMuonHistogram.sce (C:\Users\Enrique\Desktop\OpenData\DiMuonHistogra...

File Edit Format Options Window Execute ?

DiMuonHistogram.sce (C:\Users\Enrique\Desktop\OpenData\DiMuonHistogram.sce) - SciNotes

DiMuonHistogram.sce

```

opendata = read_csv('MuRun2010B.csv')

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 = []


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

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

```

SCILAB and EXCEL Activity Discussion Questions

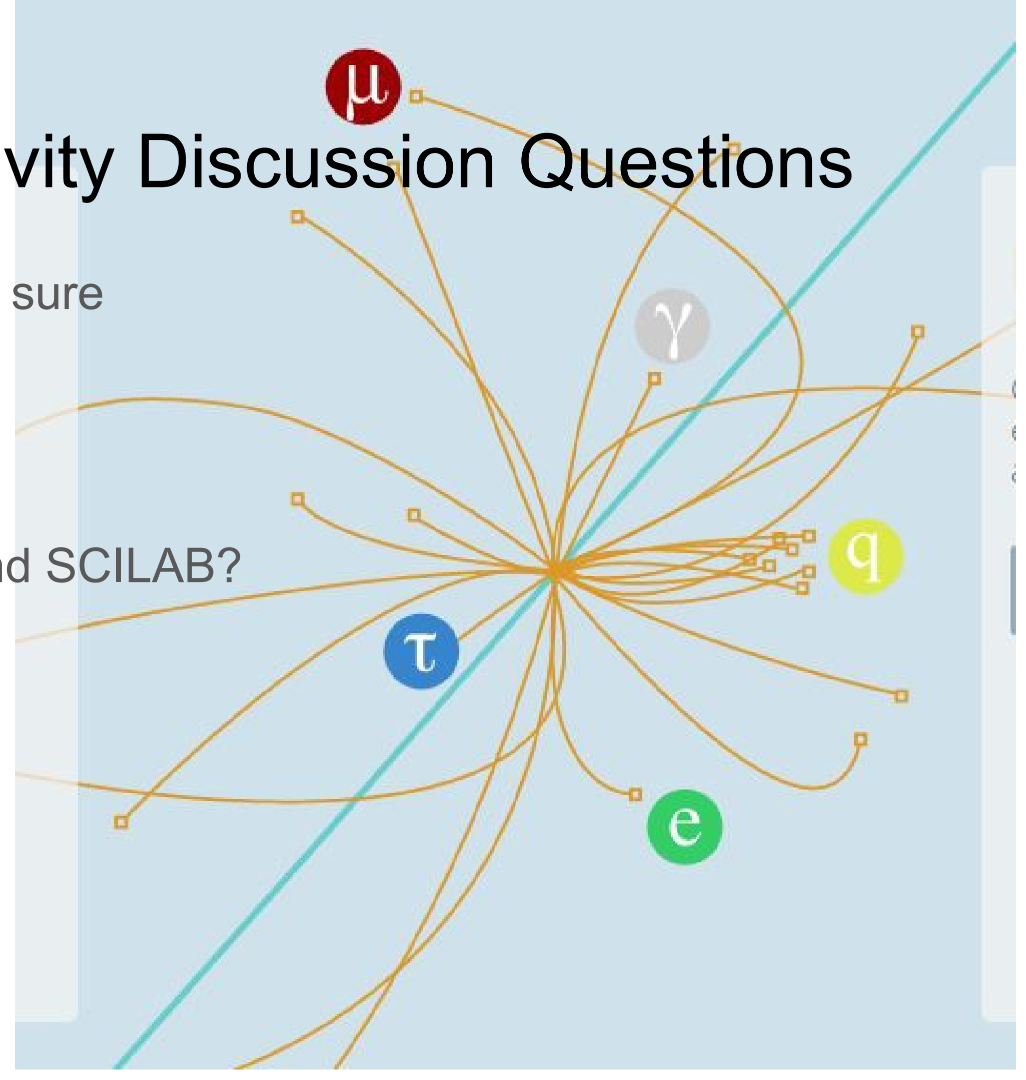
- How much data do we need to be sure there is a statistical peak? (Introduction to CMS by Piotr Traczyk)

- Compare and Contrast EXCEL and SCILAB?

Advantages of each?

- Where did the data come from?

How did we get it? Cost?



What are Jupyter Notebooks?

a web application

- create and share documents that contain

 live code, equations, visualizations .

Uses:

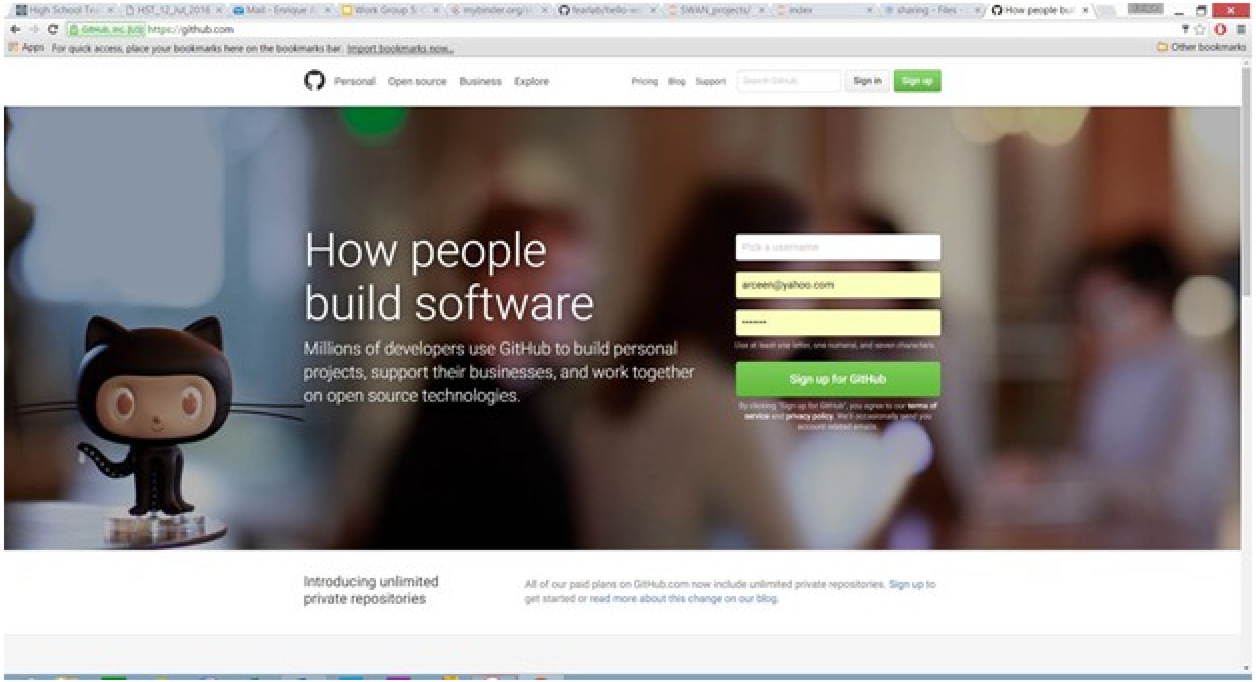
- statistical modeling
- numerical simulation
- machine learning

How to access Jupyter notebooks?

- 3 *Easy* Steps

Step 1

Create Github Account



How to access Jupyter notebooks?

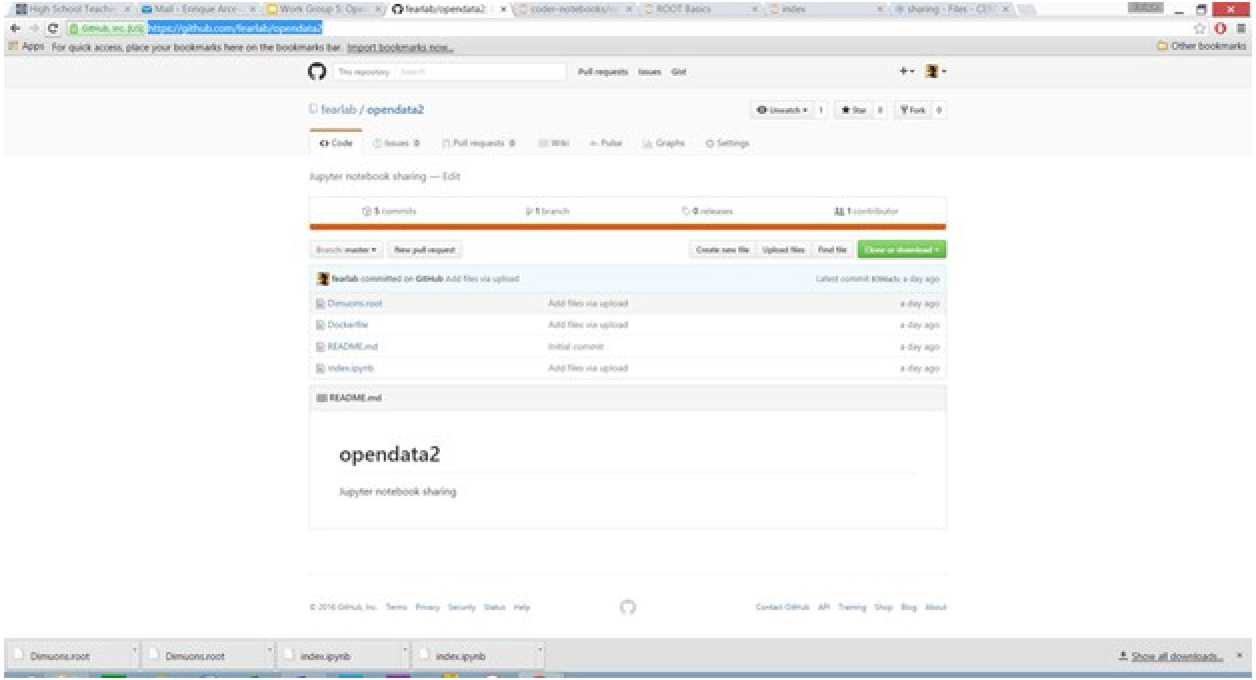
Step 2:

Upload Files

-Use shared data files

-Use shared notebook

files



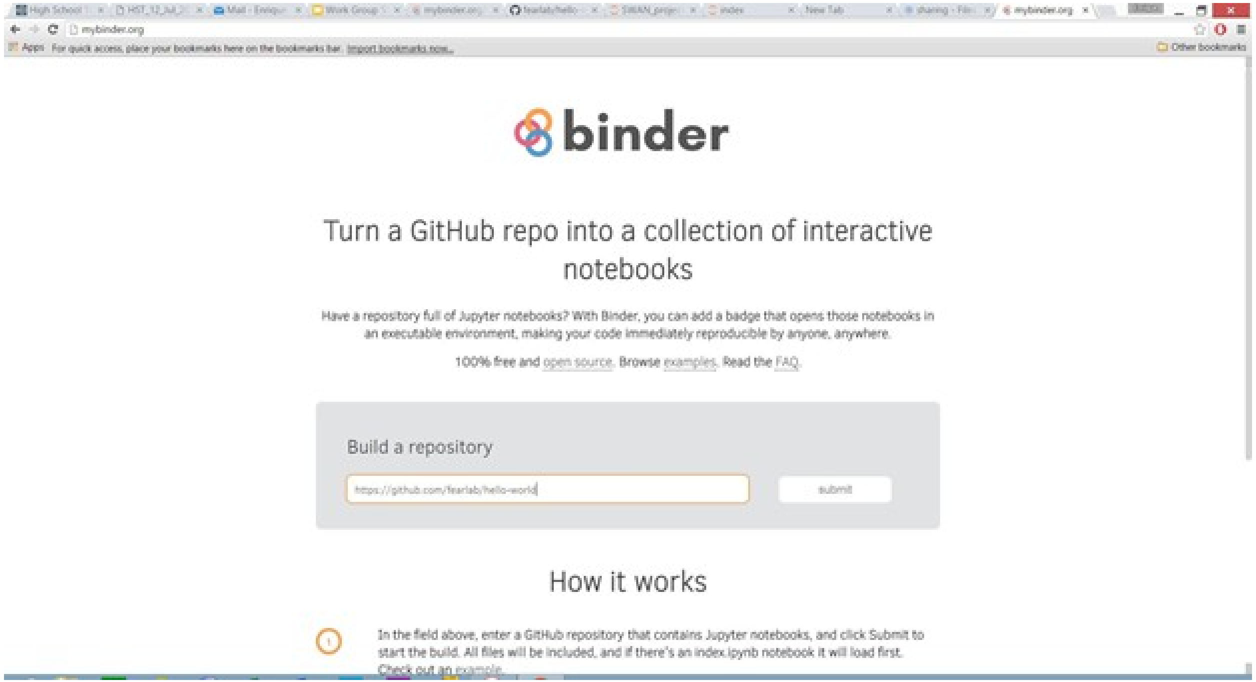
How to access Jupyter notebooks?

Step 3:

- goto mybinder.org

and build a repository

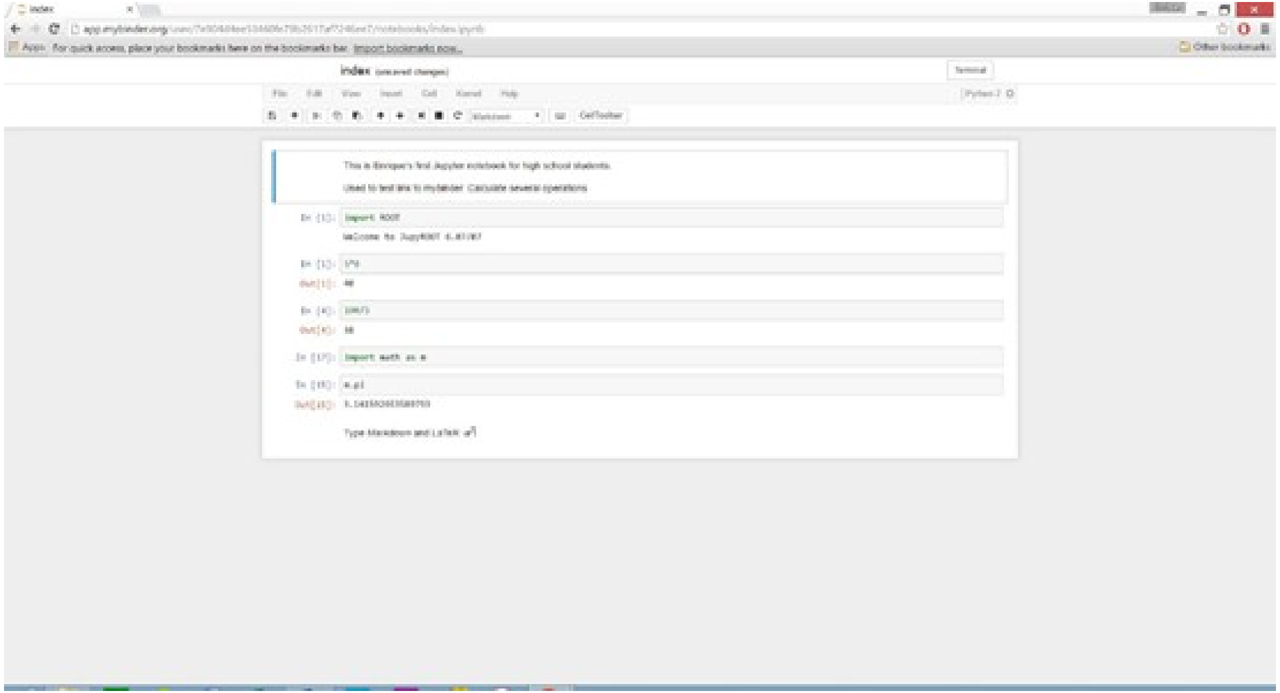
- go get Coffee



How to access Jupyter notebooks?

Students can:

- use notebooks
- no login time
- just open a link



Important links

Google drive link with all the relevant documents:

https://drive.google.com/folderview?id=0B3PRHI4_4xpmZU53a2NHOTg0QjQ&usp=sharing

Mybinder

http://tinyurl.com/gohhfl5

Open Data!