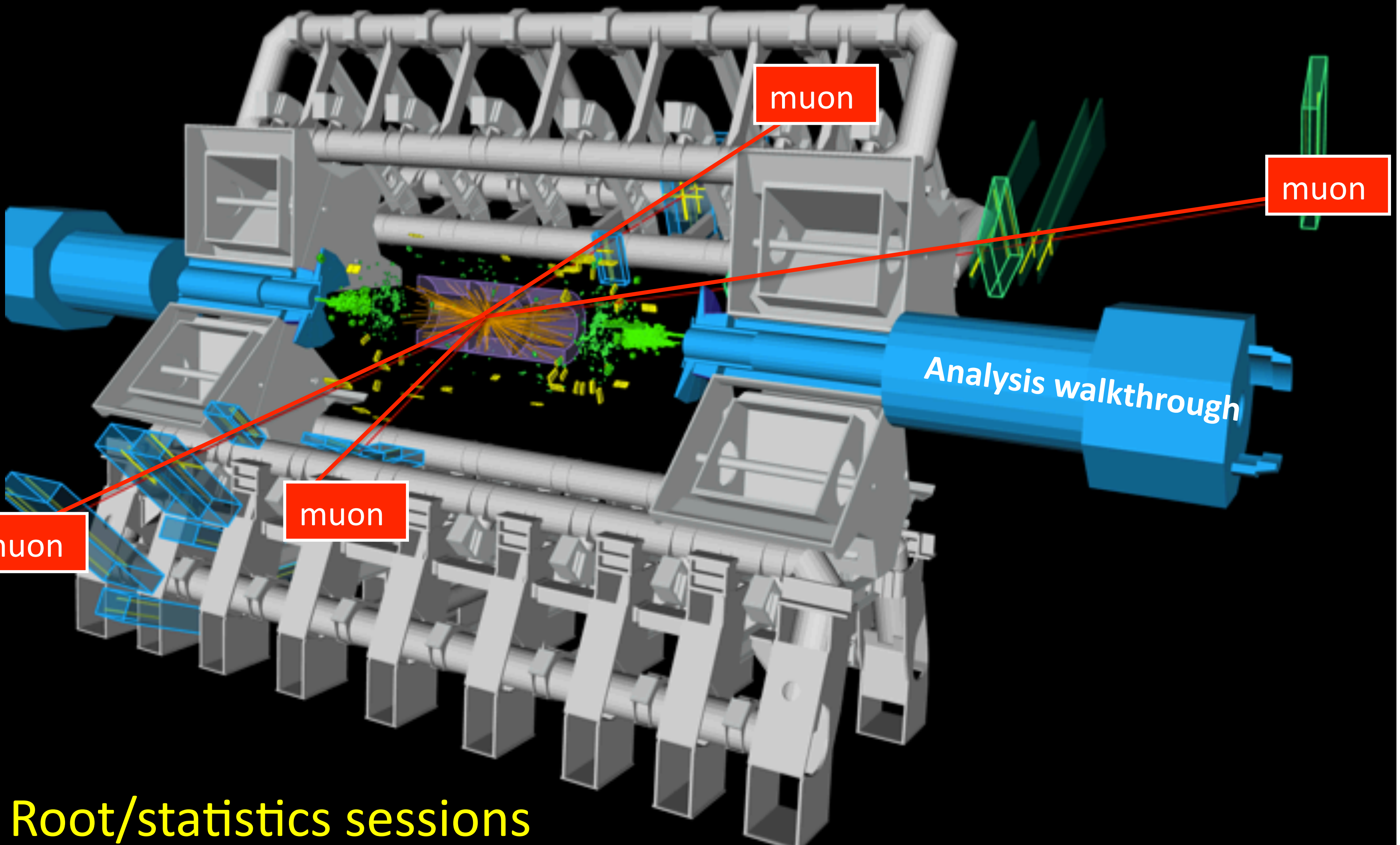
HASCO Summerschool, July 2013 (Göttingen)



Root/statistics sessions

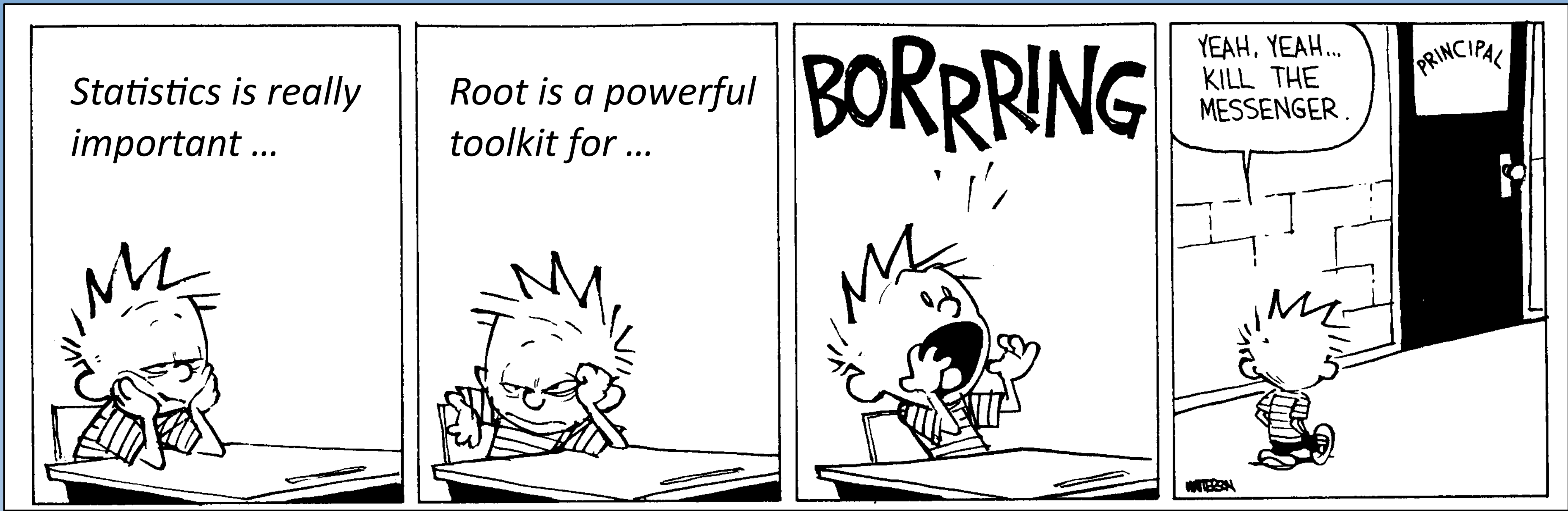
Thursday 1400:-18:15
Friday 14:00:-18:15

Ivo van Vulpen (UvA/Nikhef)
William Bell (Geneva)

Computing (Root) & Statistics

Talking about computing and statistics at a physics school

A lecture on Root and statistics



Will be 95% of your work: master this and you can focus on physics

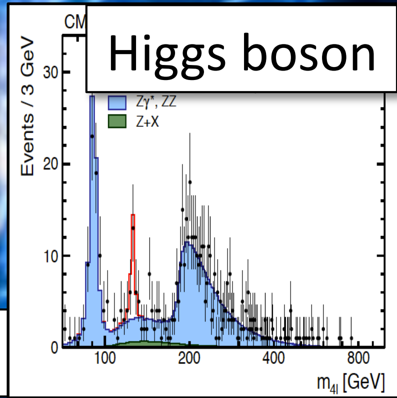
Statistics is everywhere in science and industry

Risk analyses

Banking/consultancy

World champion

Higgs boson



- Many mysteries, folklore, buzz-words, bluffing etc. but **you** need to master it to quantify the results of any study. Do **not** just follow 'what everybody else does' or what your supervisor tells you.

- RooFit, BAT, TMVA, BDT's are excellent and very powerful tools. Make sure **you** understand the basics so you know what you ask it to do.

"Do the basics yourself at least once"

Results from any scientific or business study:

$$\text{Result} = X \pm \Delta X$$

Particle mass, Cross-section, Excluded cross-section, numerical integral, probability of bankruptcy … or becoming a millionair

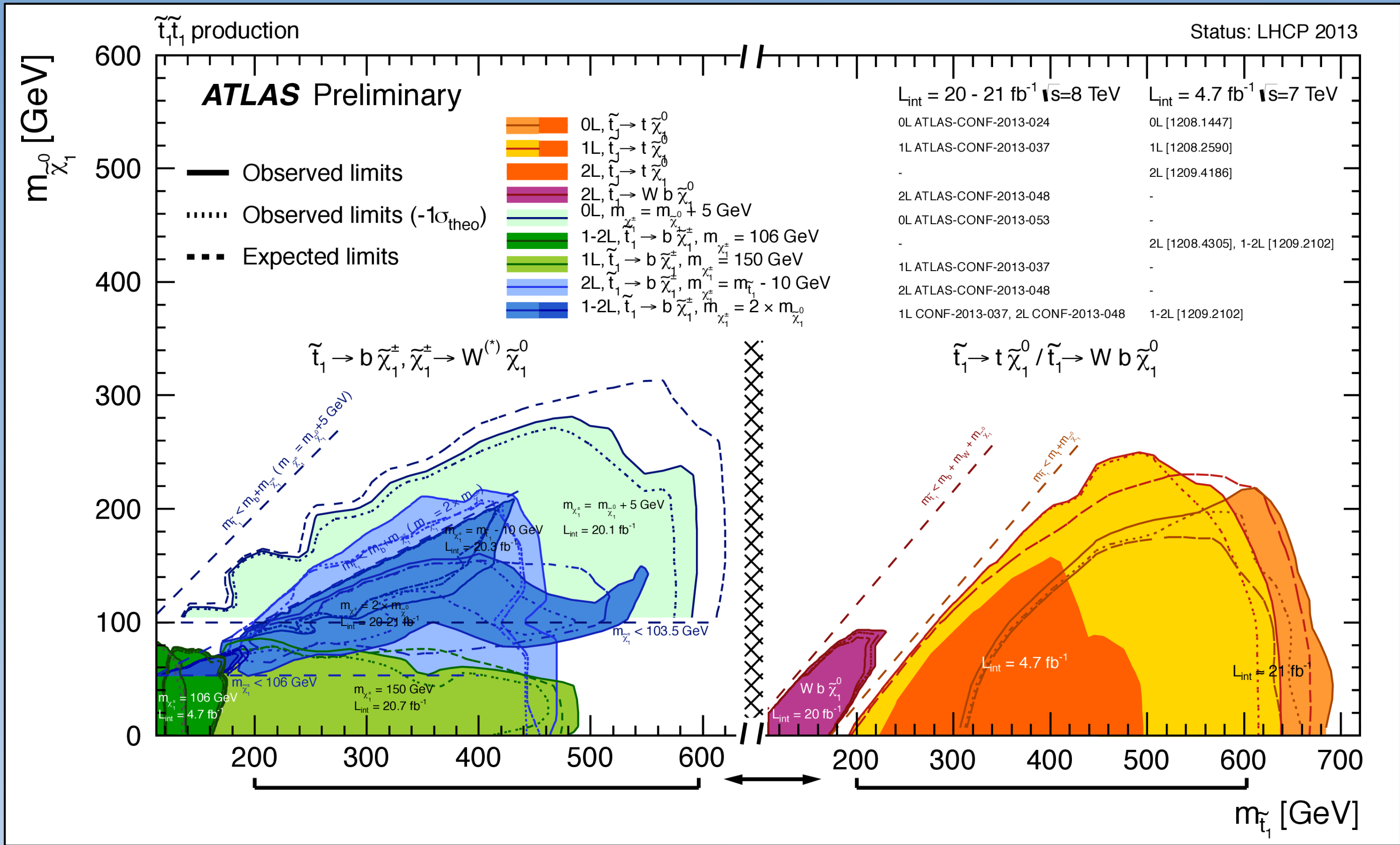
Make sure you know how you extracted this

Note: a statement in a paper/talk "our uncertainty estimat is conservative" often means:

- I have no idea how to quantify the actual uncertainty
- It was too much work to do it properly

Factor 2 means Rudiger Schmidt and the LHC accelerator people would have to produce 4x as much data as now

Summarizing how to interpret your measurement or search is not that easy as it looks. It is important!!



Analysis takes O(1 year). Please do take the additional 1 extra hours to make it crystal clear (and nice).

Root/Statistics sessions

Goal: 1) have you learn Root basics and manipulate histogram data
- do a Likelihood fit and extract a number and it's error
- understand what a significance is

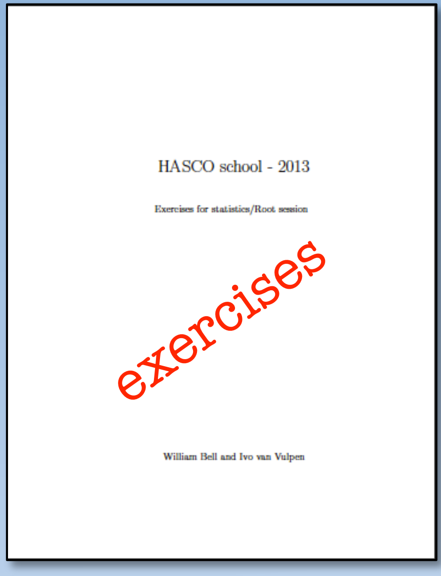
2) give you a feel for what is "out there". No need to re-invent the wheel

Thursday:

Lecture 1: Introduction to Root and statistics-exercises
Exercises: Hands-on computing exercises

Friday:

Lecture 2: Introduction to advanced tools / fits
Exercises: Hands-on computing exercises



Note: - exam questions will cover Root/statistic concepts (i.e. no coding)
- if you do not know any C(++) or are a Root-expert already please let us know

Basic material for the Root examples:

1) Get the data-set and example code: **HascoRootStatisticsCode.tgz**

2) Unpack everything: **tar –vzxf HascoRootStatisticsCode.tgz**

1) Directory /RootExamples/

a) Example0*C (* = 0,1,2,3,4,5) - All examples *.C-files used in this presentation
b) Code for Ntuple production and reading
c) rootlogon.C (some standard Root settings)

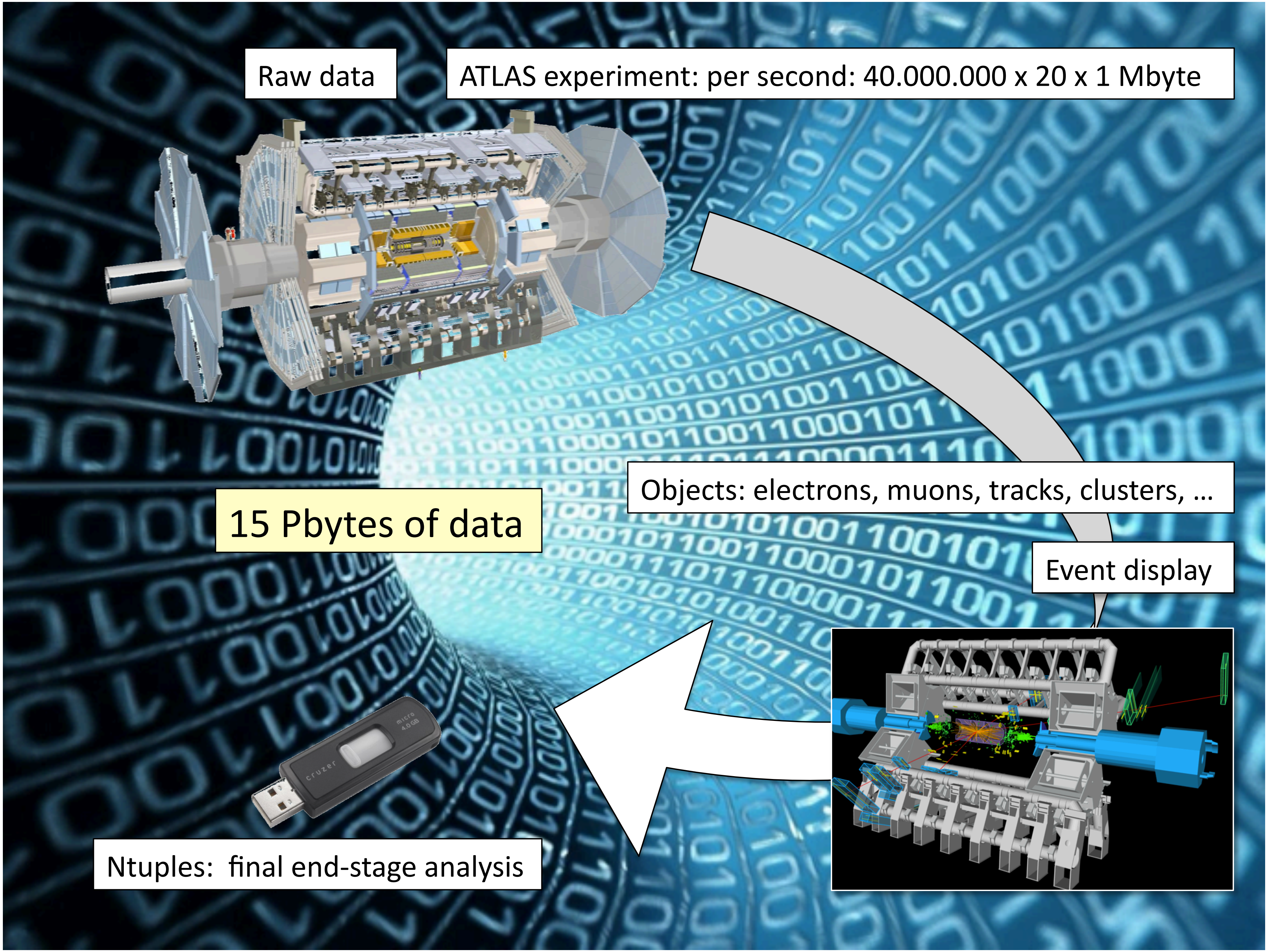
2) Directory /Exercises/

a) Histograms_fake.root
Contains 4 lepton invariant mass histogram distributions (H125, H200, ZZ, data)
b) Hasco_skeleton.C
Some skeleton code (different levels, as minimal as possible). Your code !
c) rootlogon.C (Some standard Root settings)

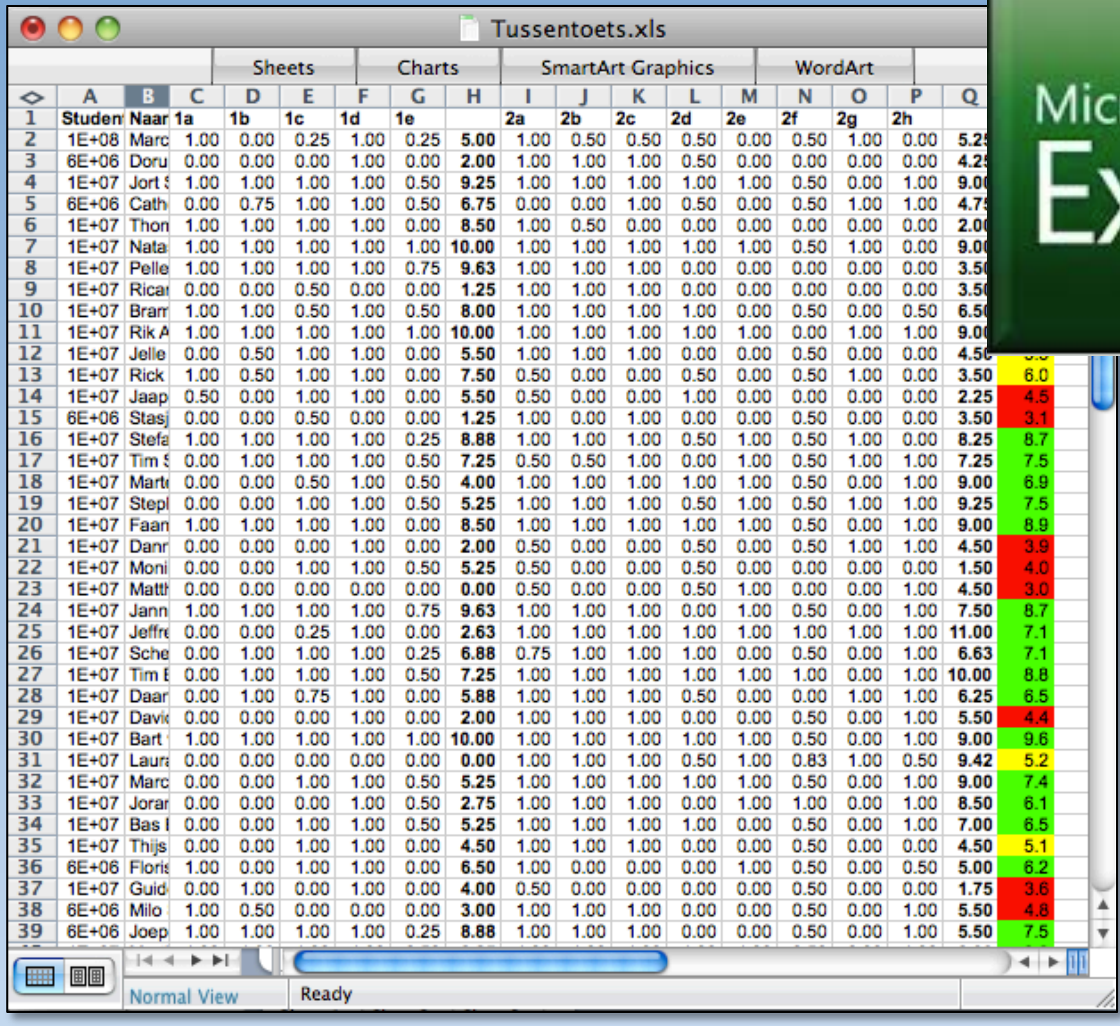
Root

Don’t worry when you do not know C++ or Root already

… well ok, worry a bit, but not too much


Raw data
ATLAS experiment: per second: 40.000.000 x 20 x 1 Mbyte
15 Pbytes of data
Objects: electrons, muons, tracks, clusters, …
Event display
Ntuples: final end-stage analysis

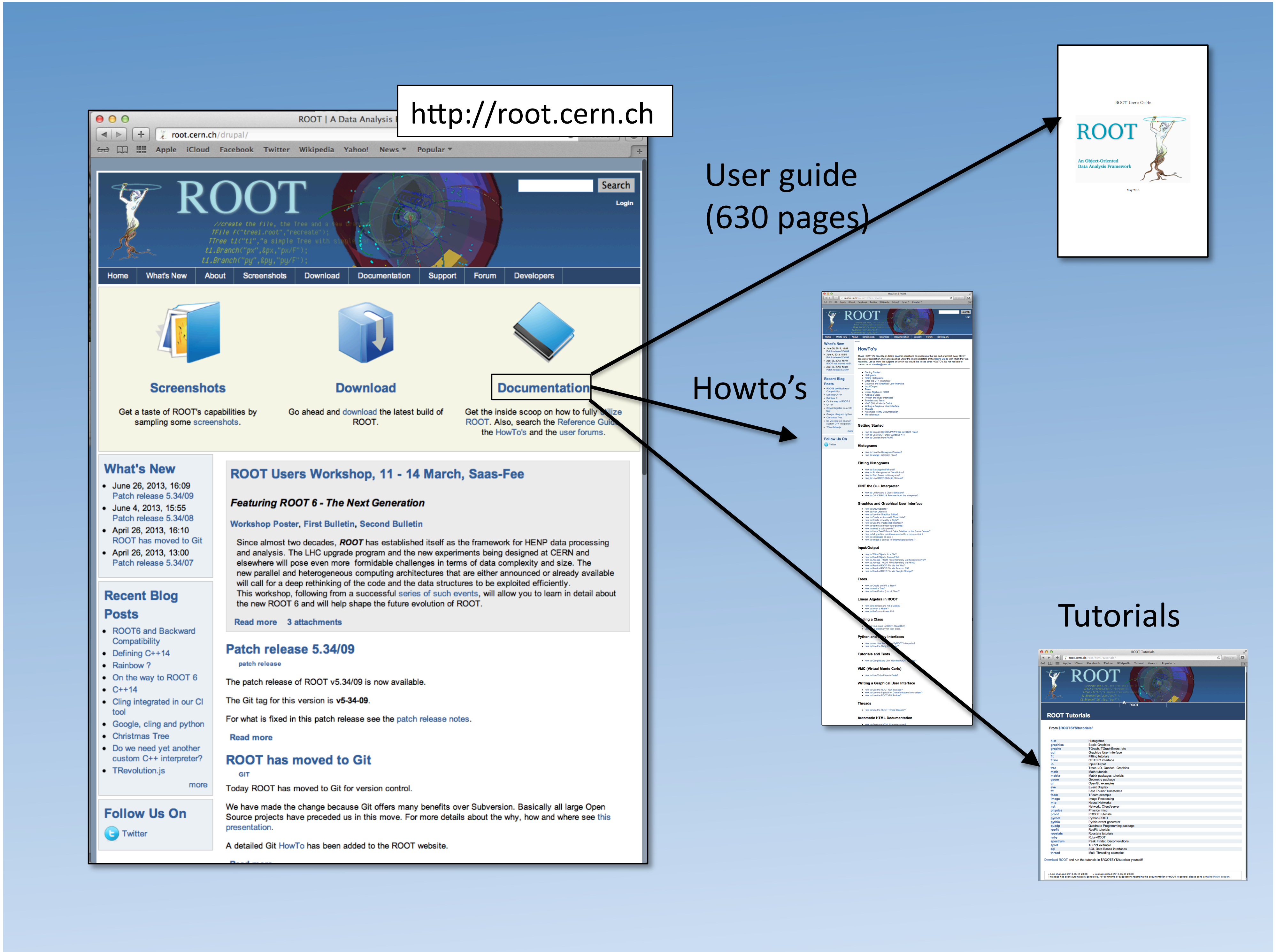
Excell or ascii files will not do



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Studen	Naar	1a	1b	1c	1d	1e		2a	2b	2c	2d	2e	2f	2g	2h		
2	1E+08	Marc	1.00	0.00	0.25	1.00	0.25	5.00	1.00	0.50	0.50	0.50	0.00	0.50	1.00	0.00	5.2	
3	6E+06	Doru	0.00	0.00	0.00	1.00	0.00	2.00	1.00	1.00	1.00	0.50	0.00	0.00	0.00	0.00	4.2	
4	1E+07	Jort	1.00	1.00	1.00	1.00	0.50	9.25	1.00	1.00	1.00	1.00	1.00	0.50	0.00	1.00	9.0	
5	6E+06	Cath	0.00	0.75	1.00	1.00	0.50	6.75	0.00	0.00	1.00	0.50	0.00	0.50	1.00	1.00	4.7	
6	1E+07	Thon	1.00	1.00	1.00	1.00	0.00	8.50	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	2.0	
7	1E+07	Nata	1.00	1.00	1.00	1.00	1.00	10.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0.00	9.0	
8	1E+07	Pelle	1.00	1.00	1.00	1.00	0.75	9.63	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	3.5	
9	1E+07	Ricar	0.00	0.00	0.50	0.00	0.00	1.25	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	3.5	
10	1E+07	Bram	1.00	1.00	0.50	1.00	0.50	8.00	1.00	1.00	1.00	1.00	0.00	0.50	0.00	0.50	6.5	
11	1E+07	Rik A	1.00	1.00	1.00	1.00	1.00	10.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	9.0	
12	1E+07	Jelle	0.00	0.50	1.00	1.00	0.00	5.50	1.00	1.00	1.00	0.00	0.00	0.50	0.00	0.00	4.5	[illegible]
13	1E+07	Rick	1.00	0.50	1.00	1.00	0.00	7.50	0.50	0.00	0.00	0.50	0.00	0.50	1.00	0.00	3.50	6.0
14	1E+07	Jaap	0.50	0.00	1.00	1.00	0.00	5.50	0.50	0.00	0.00	1.00	0.00	0.00	0.00	0.00	2.25	4.5
15	6E+06	Stasj	0.00	0.00	0.50	0.00	0.00	1.25	1.00	0.00	1.00	0.00	0.00	0.50	0.00	0.00	3.50	3.1
16	1E+07	Stefa	1.00	1.00	1.00	1.00	0.25	8.88	1.00	1.00	1.00	0.50	1.00	0.50	1.00	0.00	8.25	8.7
17	1E+07	Tim	0.00	1.00	1.00	1.00	0.50	7.25	0.50	0.50	1.00	0.00	1.00	0.50	1.00	1.00	7.25	7.5
18	1E+07	Mart	0.00	0.00	0.50	1.00	0.50	4.00	1.00	1.00	1.00	1.00	1.00	0.50	0.00	1.00	9.00	6.9
19	1E+07	Stepl	0.00	0.00	1.00	1.00	0.50	5.25	1.00	1.00	1.00	0.50	1.00	0.50	1.00	1.00	9.25	7.5
20	1E+07	Faan	1.00	1.00	1.00	1.00	0.00	8.50	1.00	1.00	1.00	1.00	1.00	0.50	0.00	1.00	9.00	8.9
21	1E+07	Danr	0.00	0.00	0.00	1.00	0.00	2.00	0.50	0.00	0.00	0.50	0.00	0.50	1.00	1.00	4.50	3.9
22	1E+07	Moni	0.00	0.00	1.00	1.00	0.50	5.25	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	1.50	4.0
23	1E+07	Matth	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.50	1.00	0.00	0.00	1.00	4.50	3.0
24	1E+07	Jann	1.00	1.00	1.00	1.00	0.75	9.63	1.00	1.00	1.00	0.00	1.00	0.50	0.00	1.00	7.50	8.7
25	1E+07	Jeffre	0.00	0.00	0.25	1.00	0.00	2.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	11.00	7.1
26	1E+07	Sche	0.00	1.00	1.00	1.00	0.25	6.88	0.75	1.00	1.00	1.00	0.00	0.50	0.00	1.00	6.63	7.1
27	1E+07	Tim E	0.00	1.00	1.00	1.00	0.50	7.25	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	10.00	8.8
28	1E+07	Daar	0.00	1.00	0.75	1.00	0.00	5.88	1.00	1.00	1.00	0.50	0.00	0.00	1.00	1.00	6.25	6.5
29	1E+07	David	0.00	0.00	0.00	1.00	0.00	2.00	1.00	1.00	1.00	0.00	0.00	0.50	0.00	1.00	5.50	4.4
30	1E+07	Bart	1.00	1.00	1.00	1.00	1.00	10.00	1.00	1.00	1.00	1.00	1.00	0.50	0.00	1.00	9.00	9.6
31	1E+07	Laura	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.50	1.00	0.83	1.00	0.50	9.42	5.2
32	1E+07	Marc	0.00	0.00	1.00	1.00	0.50	5.25	1.00	1.00	1.00	1.00	1.00	0.50	0.00	1.00	9.00	7.4
33	1E+07	Joran	0.00	0.00	0.00	1.00	0.50	2.75	1.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	8.50	6.1
34	1E+07	Bas	0.00	0.00	1.00	1.00	0.50	5.25	1.00	1.00	1.00	1.00	0.00	0.50	0.00	1.00	7.00	6.5
35	1E+07	Thijs	0.00	0.00	1.00	1.00	0.00	4.50	1.00	1.00	1.00	0.00	0.00	0.50	0.00	0.00	4.50	5.1
36	6E+06	Floris	1.00	0.00	1.00	1.00	0.00	6.50	1.00	0.00	0.00	0.00	1.00	0.50	0.00	0.50	5.00	6.2
37	1E+07	Guid	0.00	1.00	0.00	1.00	0.00	4.00	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	1.75	3.6
38	6E+06	Milo	1.00	0.50	0.00	0.00	0.00	3.00	1.00	1.00	1.00	0.00	0.00	0.50	0.00	1.00	5.50	4.8
39	6E+06	Joep	1.00	1.00	1.00	1.00	0.25	8.88	1.00	1.00	1.00	0.00	0.00	0.50	0.00	1.00	5.50	7.5

ROOT

An Object-Oriented Data Analysis Framework



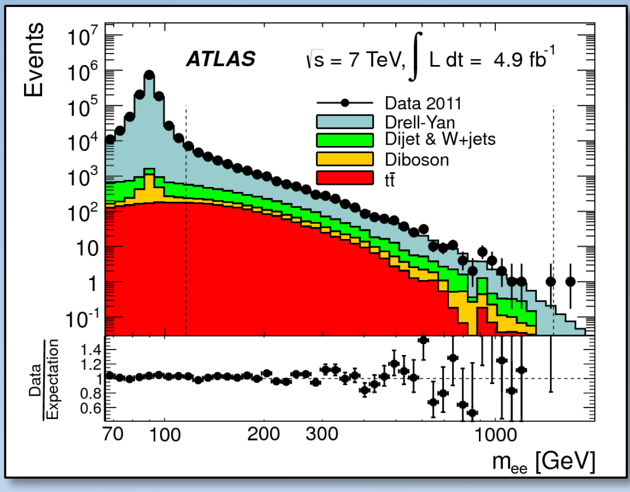
http://root.cern.ch
User guide
(630 pages)
Howto's
Tutorials
ROOT | A Data Analysis
root.cern.ch/drupal/
ROOT
Search
Login
Home
What's New
About
Screenshots
Download
Documentation
Support
Forum
Developers
Screenshots
Get a taste of ROOT's capabilities by sampling some screenshots.
Download
Go ahead and download the latest build of ROOT.
Documentation
Get the inside scoop on how to fully utilize ROOT. Also, search the Reference Guide, the HowTo's and the user forums.
What's New
June 26, 2013, 16:09
Patch release 5.34/09
June 4, 2013, 15:55
Patch release 5.34/08
April 26, 2013, 16:10
ROOT has moved to Git
April 26, 2013, 13:00
Patch release 5.34/07
Recent Blog Posts
ROOT6 and Backward Compatibility
Defining C++14
Rainbow ?
On the way to ROOT 6
C++14
Cling integrated in our CI tool
Google, cling and python
Christmas Tree
Do we need yet another custom C++ interpreter?
TRevolution.js
more
Follow Us On
Twitter
ROOT Users Workshop, 11 - 14 March, Saas-Fee
Featuring ROOT 6 - The Next Generation
Workshop Poster, First Bulletin, Second Bulletin
Since almost two decades, ROOT has established itself as the framework for HENP data processing and analysis. The LHC upgrade program and the new experiments being designed at CERN and elsewhere will pose even more formidable challenges in terms of data complexity and size. The new parallel and heterogeneous computing architectures that are either announced or already available will call for a deep rethinking of the code and the data structures to be exploited efficiently.
This workshop, following from a successful series of such events, will allow you to learn in detail about the new ROOT 6 and will help shape the future evolution of ROOT.
Read more
3 attachments
Patch release 5.34/09
patch release
The patch release of ROOT v5.34/09 is now available.
The Git tag for this version is v5-34-09.
For what is fixed in this patch release see the patch release notes.
Read more
ROOT has moved to Git
GIT
Today ROOT has moved to Git for version control.
We have made the change because Git offers many benefits over Subversion. Basically all large Open Source projects have preceded us in this move. For more details about the why, how and where see this presentation.
A detailed Git HowTo has been added to the ROOT website.
ROOT User's Guide
ROOT
An Object-Oriented
Data Analysis Framework
May 2013
ROOT Tutorials

Root: (HEP) data modeling, manipulation and visualisation

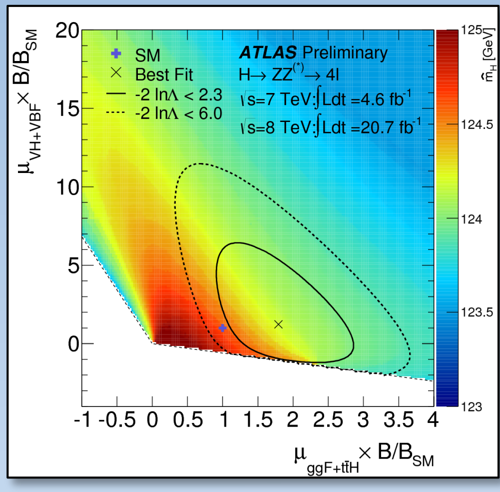
- C++ based analysis framework [High-energy physics toolkit]
  Python interface PyRoot
- Storing, manipulating and analysing data (histograms to complex structures)
- Interpretation of the data: model building, fitting, hypothesis testing, …
- Visualising the data

Root: data visualisation

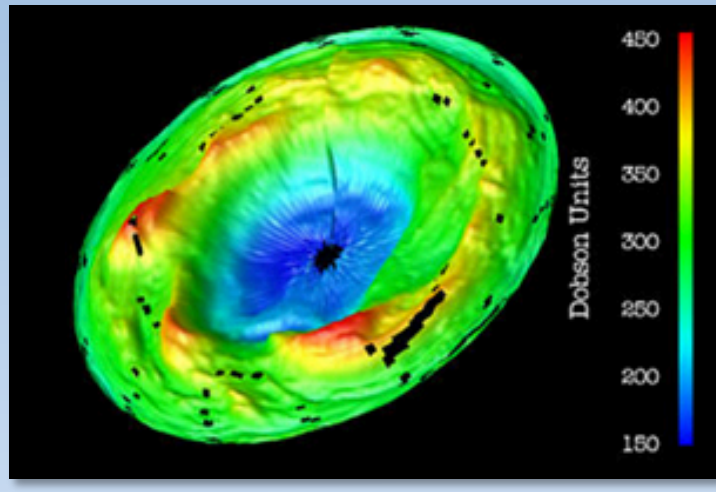
1-dimensional



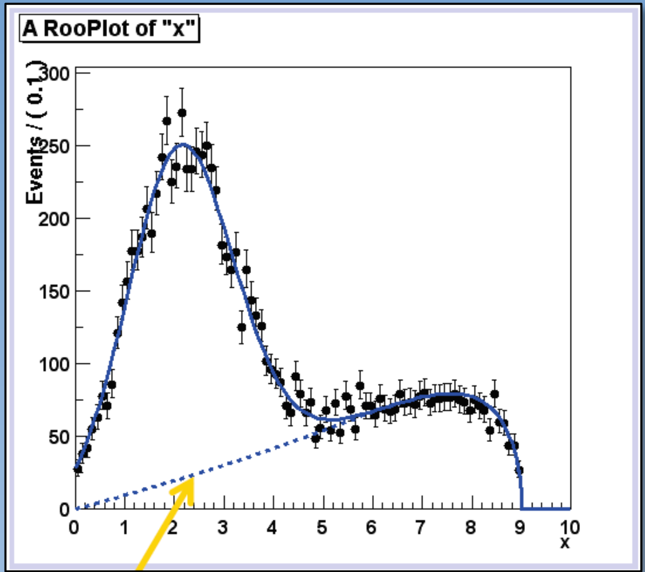
2-dimensional



3-dimensional



Root: data modeling & hypothesis testing



OO Likelihood Modeling & convolutions

Fitting & hypothesis testing

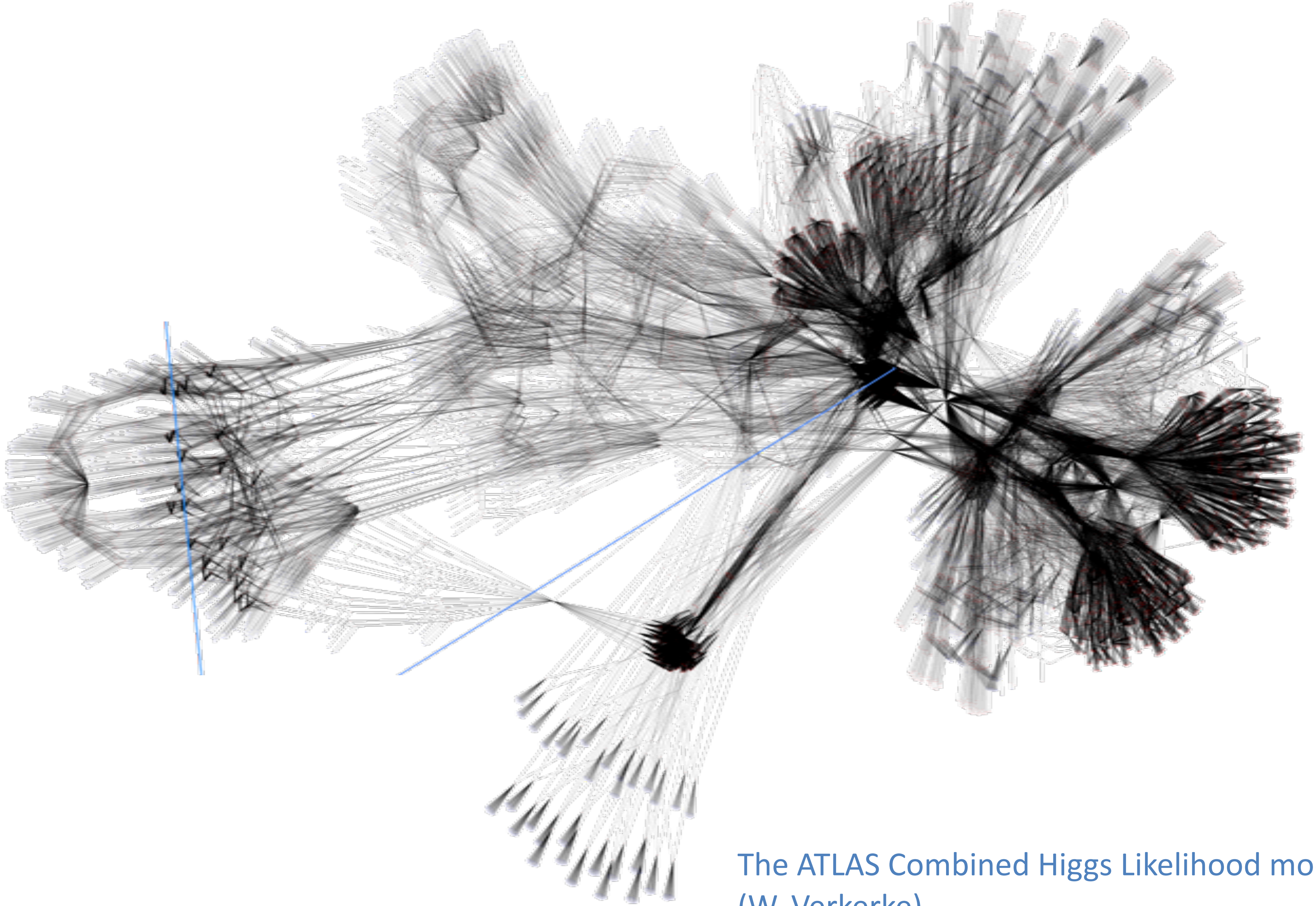
Statistics tools:
(test-statistic, toy-MC, significance, limits, profiling)

→ **We'll do our own fit today in the exercise session**

Multi-dimensional analyses require more complex tools: Neural nets, Boosted decision trees etc. → **tomorrow's lecture**

Root: mathematics & physics tools

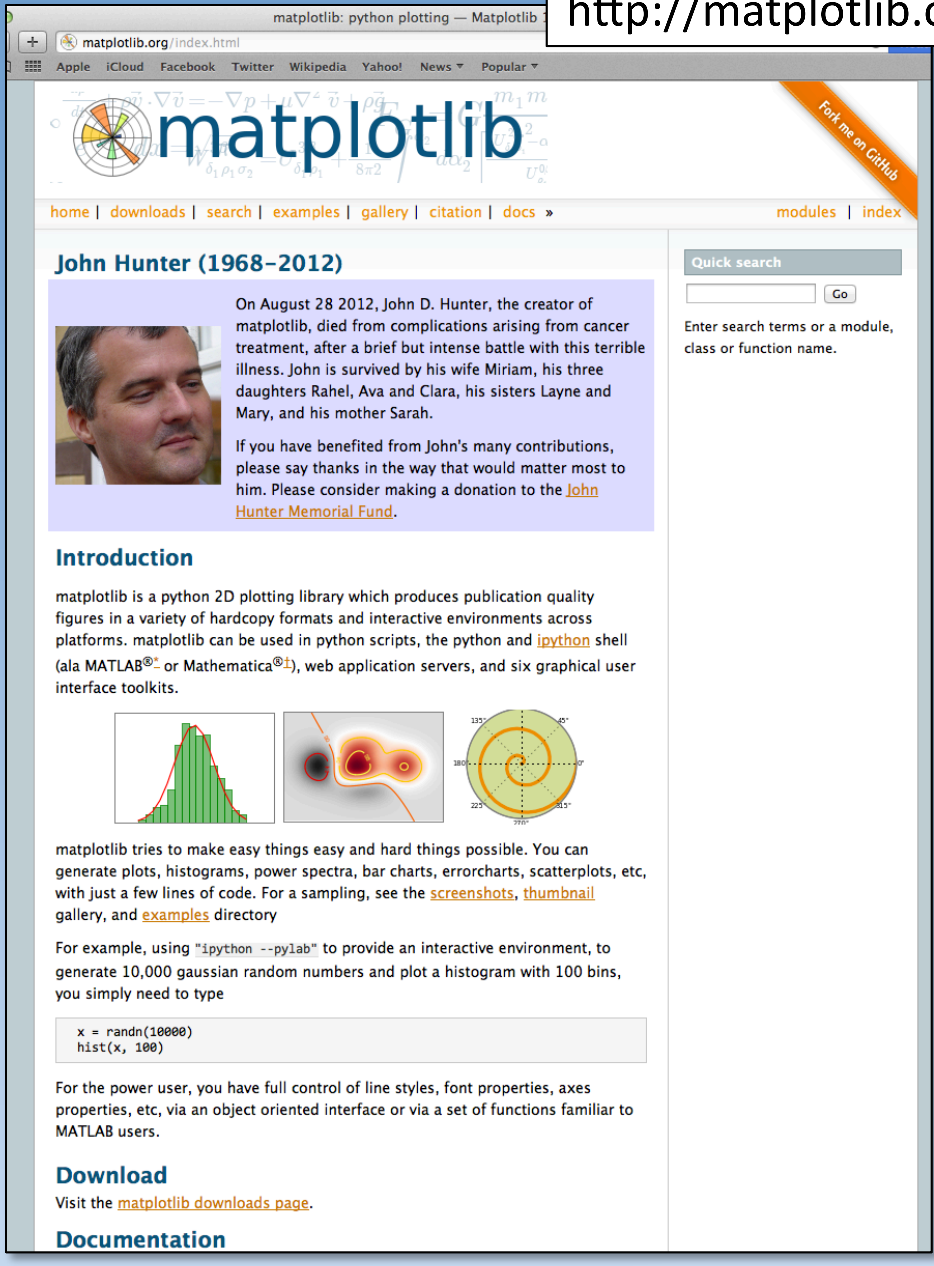
- Types: int, float, double, vectors, matrices, … but also Lorentz vectors etc.
  you can of course build your own class/object/structure
- Random numbers, matrix manipulation, PDG information, boosts
- Geant4, …



The ATLAS Combined Higgs Likelihood model
(W. Verkerke)

Data analysis and visualisation in Python (Matplotlib)

http://matplotlib.org



matplotlib: python plotting — Matplotlib

matplotlib.org/index.html

Apple iCloud Facebook Twitter Wikipedia Yahoo! News Popular

matplotlib

Fork me on GitHub

home | downloads | search | examples | gallery | citation | docs » modules | index

John Hunter (1968–2012)

On August 28 2012, John D. Hunter, the creator of matplotlib, died from complications arising from cancer treatment, after a brief but intense battle with this terrible illness. John is survived by his wife Miriam, his three daughters Rahel, Ava and Clara, his sisters Layne and Mary, and his mother Sarah.

If you have benefited from John's many contributions, please say thanks in the way that would matter most to him. Please consider making a donation to the John Hunter Memorial Fund.

Quick search

Go

Enter search terms or a module, class or function name.

Introduction

matplotlib is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. matplotlib can be used in python scripts, the python and ipython shell (ala MATLAB®* or Mathematica®†), web application servers, and six graphical user interface toolkits.

matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc, with just a few lines of code. For a sampling, see the screenshots, thumbnail gallery, and examples directory

For example, using "ipython --pylab" to provide an interactive environment, to generate 10,000 gaussian random numbers and plot a histogram with 100 bins, you simply need to type

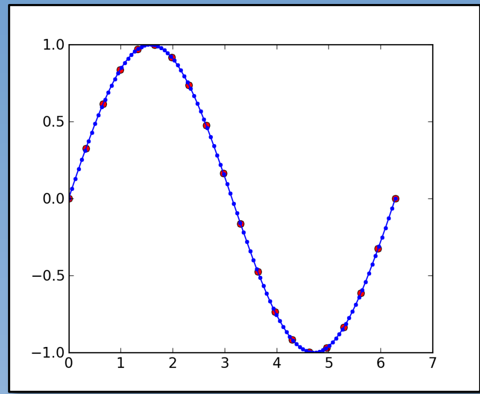
```
x = randn(10000)
hist(x, 100)
```

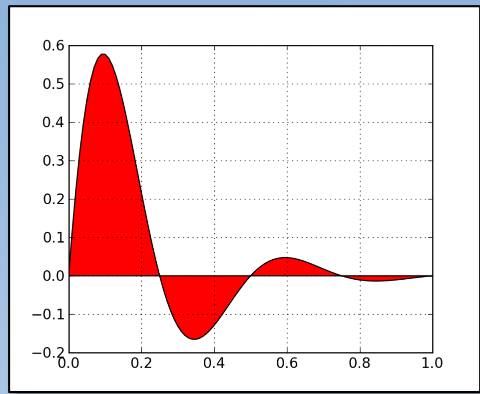
For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

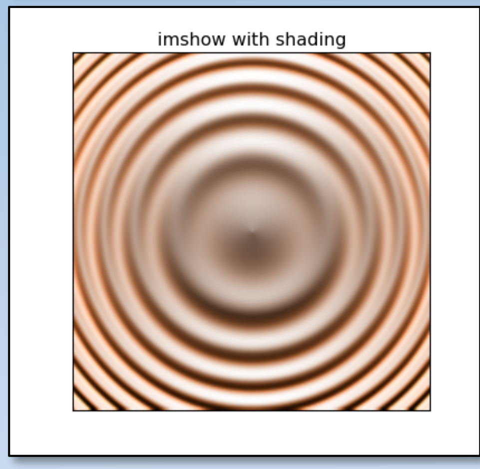
Download

Visit the matplotlib downloads page.

Documentation

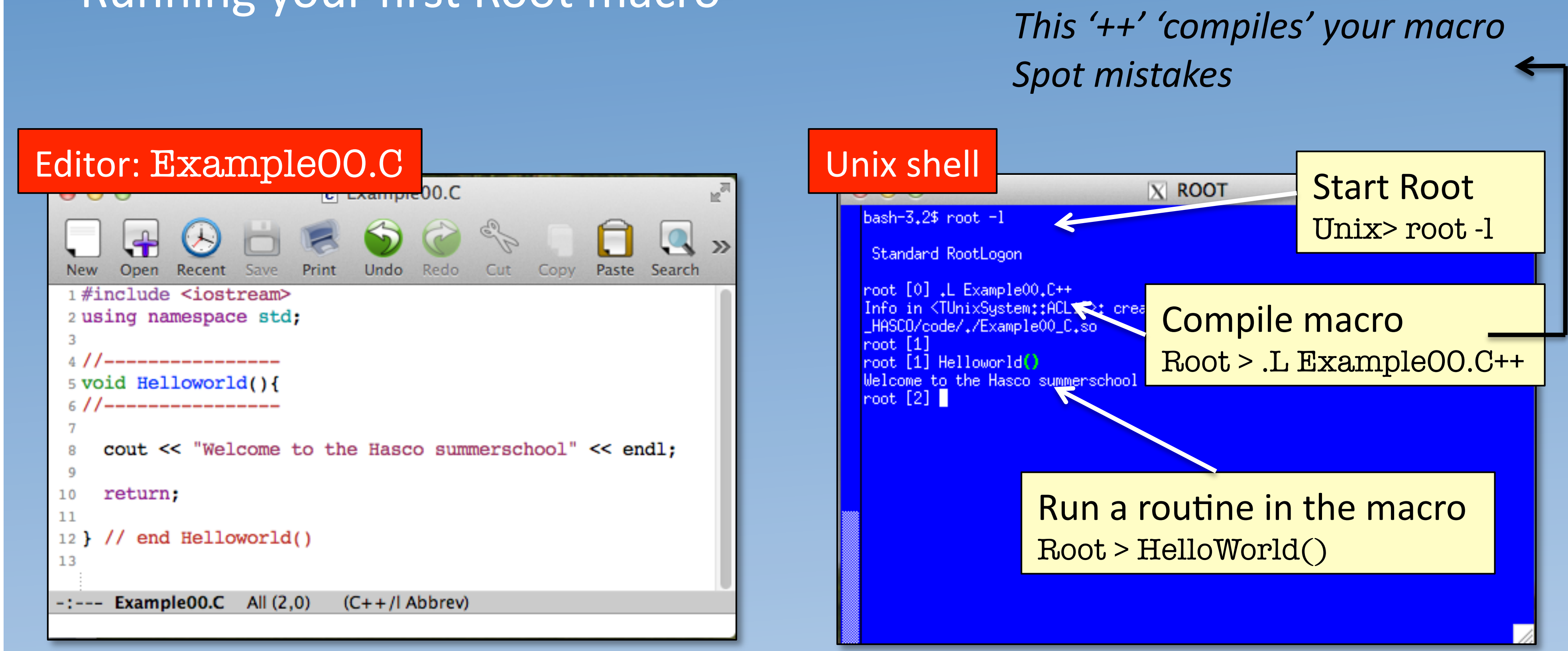






http://epd-free.enthought.com/?Download=Download+EPD+Free+7.3-2

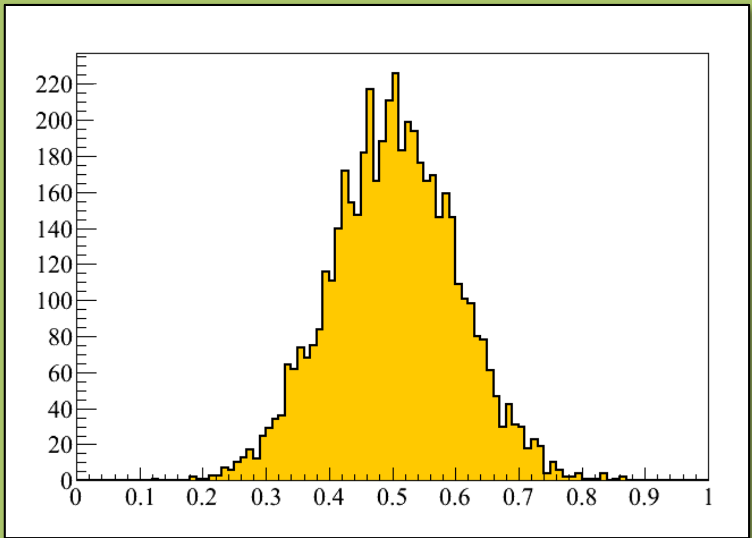
Running your first Root macro

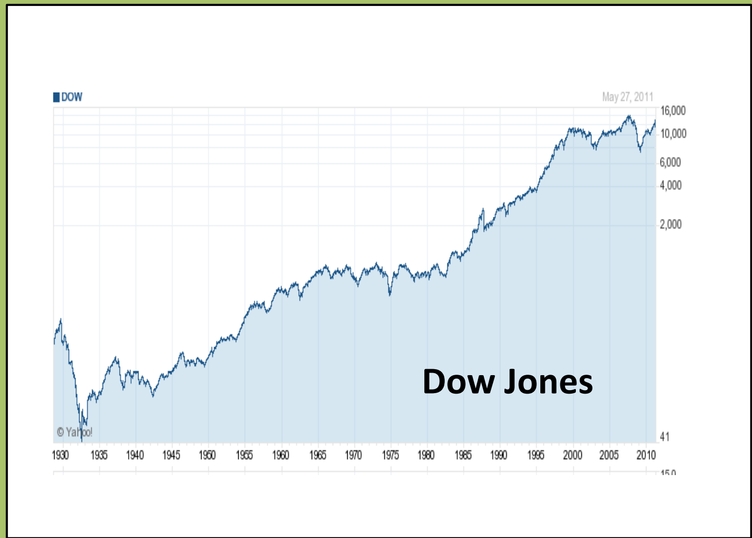


Note: When you start root it automatically processes the file **rootlogon.C**

This is a file where you can set all your default settings

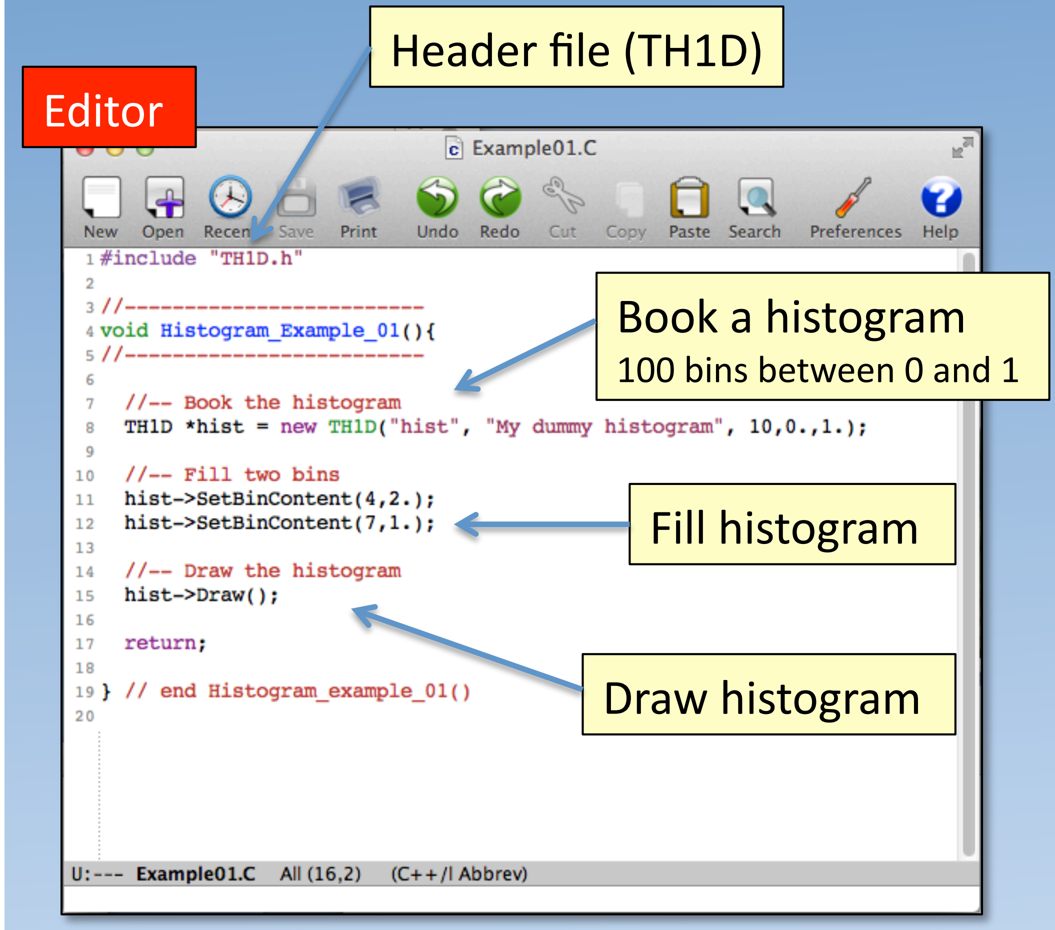
Histograms

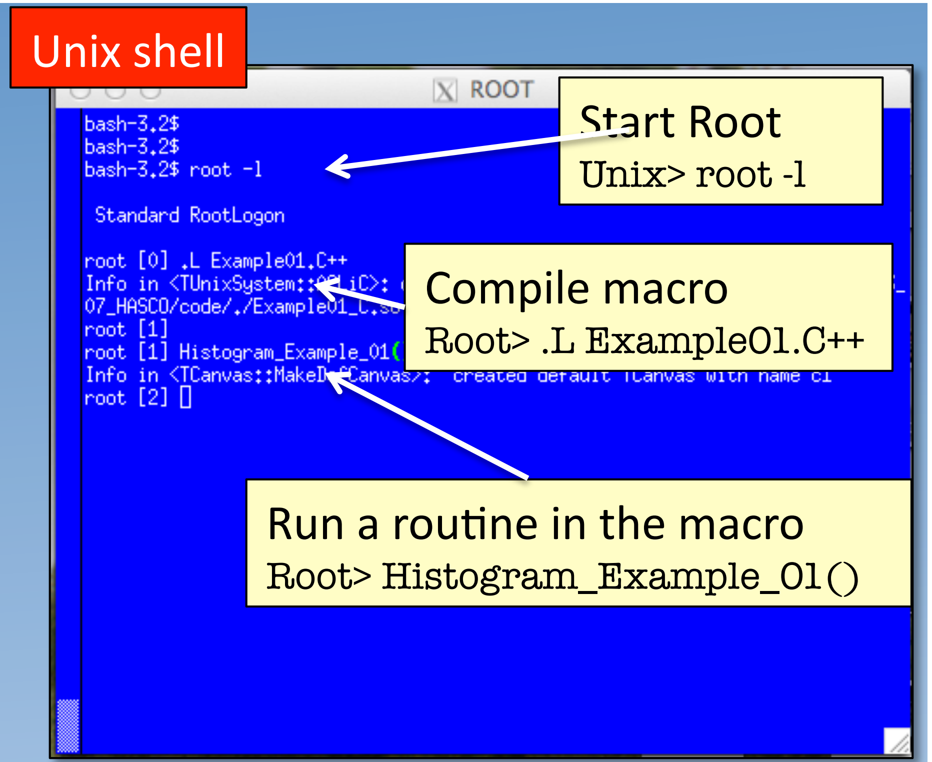


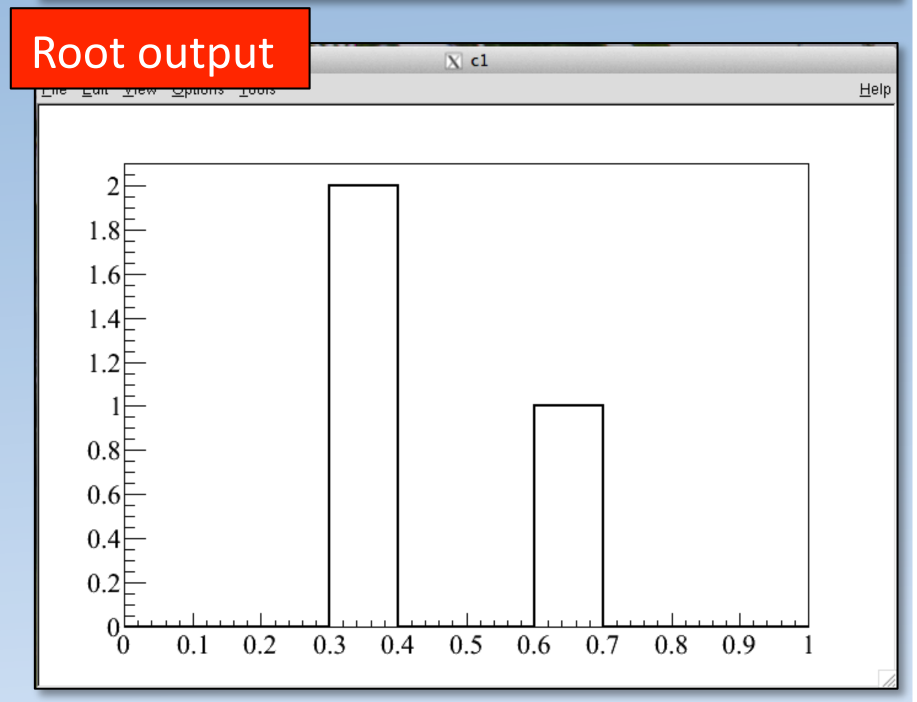

DOW
May 27, 2011
16,000
10,000
6,000
4,000
2,000
41
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Dow Jones

Histogram_Example_01.C

"plot a histogram on screen"

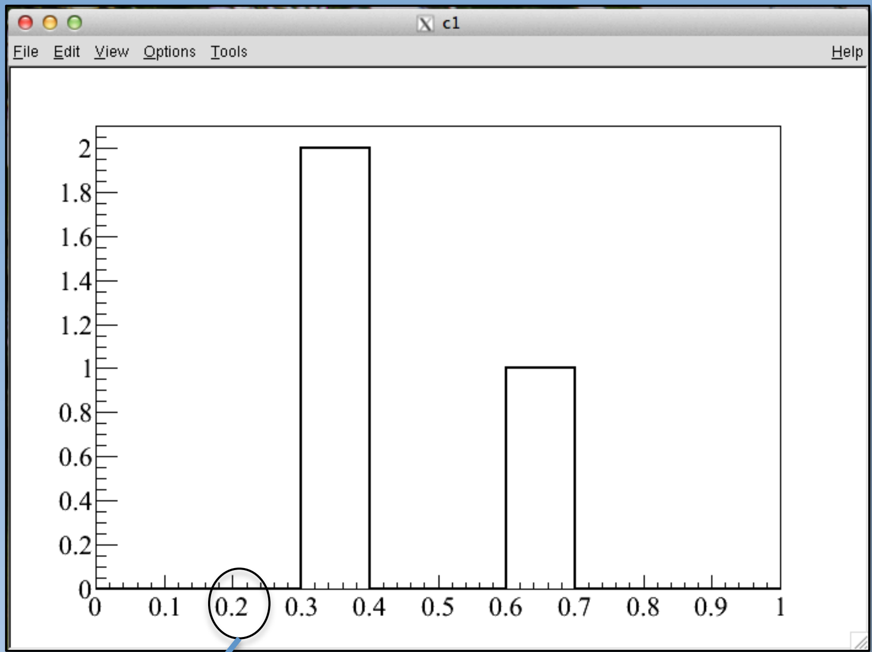






Inspect and style histograms

TH1:
http://root.cern.ch/root/html/TH1.html



Style:

hist->SetFillColor(i_{color})
hist->SetFillStyle()
hist->SetLineColor()
hist->SetLineStyle()
hist->SetLineWidth
…

Inspect:

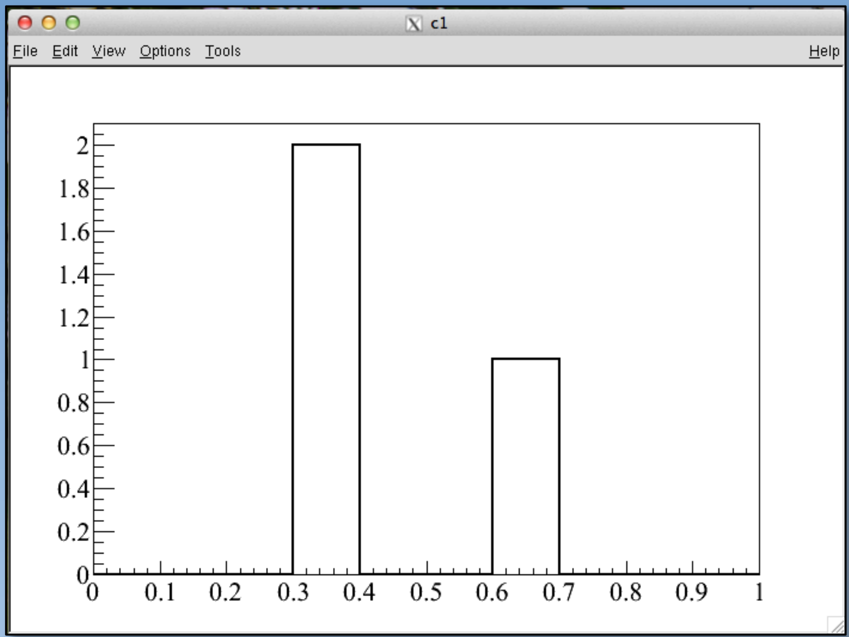
hist->GetRMS()
hist->GetMean()
hist->GetSumOfWeights()
hist->GetMaximumBin()
hist->GetBinContent(i_{bin})
hist->GetBinCenter(i_{bin})
hist->Integrate(…)
hist->Write(…)
…

You'll need these in the exercises

Details:

hist->SetLabelColor()
hist->SetLabelFont()
hist->SetLabelOffset()
hist->SetLabelSize()

Standard



Your style


```
hist->SetFillColor(6);
hist->SetFillStyle(3544);

hist->SetLineColor(4);
hist->SetLineWidth(3);
hist->SetLineStyle(2);

hist->SetLabelColor(2);
hist->SetLabelFont(200);
hist->SetLabelOffset(0.03);
hist->SetLabelSize(0.035);
```

Note: - beauty is in the eye of the beholder
- take some time to make sure you express clearly what you want in a plot
- defailt settings in `rootlogon.C`, but can customize for each plot

Histogram_Example_02.C

"write histogram to file"

TFile

Example02.C

```
#include "TH1D.h"
#include "TFile.h"

#include <iostream>
using namespace std;

//------------------------
void Histogram_example_02(){
//------------------------
//----------------------------------------
//-- Goal: o book histogram and fill 2 bins
//         o write histogram to file
//----------------------------------------

  //-- Book the histogram
  TH1D *hist = new TH1D("hist", "My dummy histogram", 10,0.,1.);

  //-- Fill two bins
  hist->SetBinContent(4,2.);
  hist->SetBinContent(7,1.);

  //-- Draw the histogram
  hist->Draw();

  //-- Save the histogram in a file
  TFile *myfile = new TFile("MySummaryFile.root", "recreate");
  hist->Write();
  myfile->Close();

  cout << " I just wrote the histogram to a file" << endl;

  return;

} // end Histogram_example_02()
```

-:--- Example02.C All (18,20) (C++/l Abbrev)

TFile:
http://root.cern.ch/root/html/TFile.html

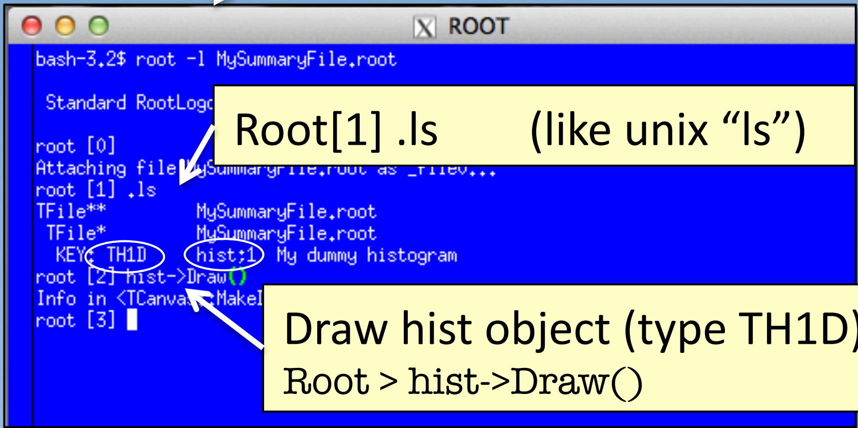
open a file called MySummaryFile.root, write histogram to it and close the file

Histogram_Example_03.C

"read a histogram from a file and plot it"

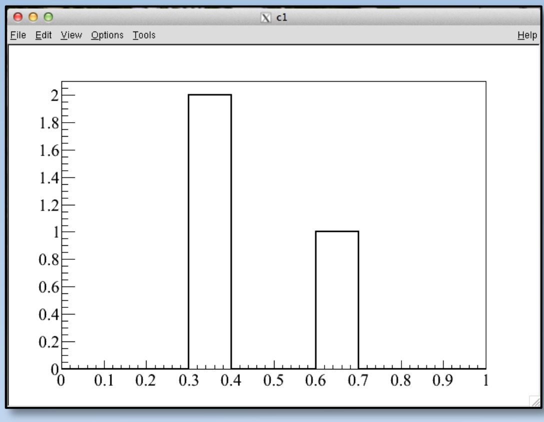
Unix> root –l MySummaryfile.root

option 1: command line

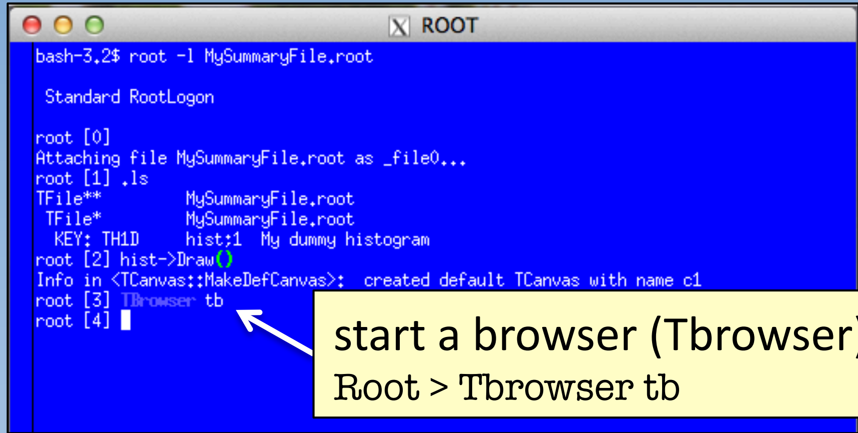


Root[1] .ls   (like unix "ls")

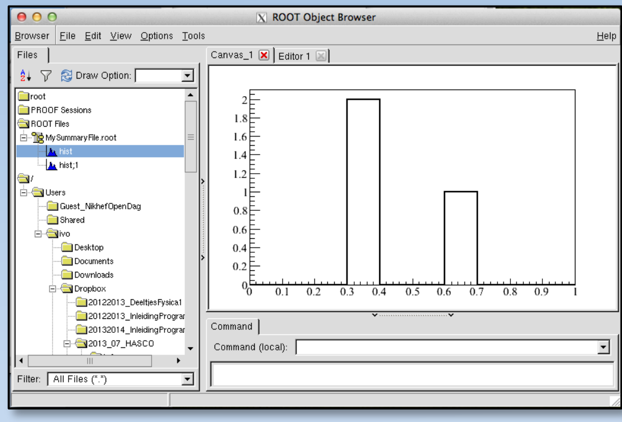
Draw hist object (type TH1D)
Root > hist->Draw()



option 2: browser / clicking



start a browser (Tbrowser)
Root > Tbrowser tb



Histogram_Example_03.C

"read a histogram from a file and plot it"

option 3: using a macro

Example03.C

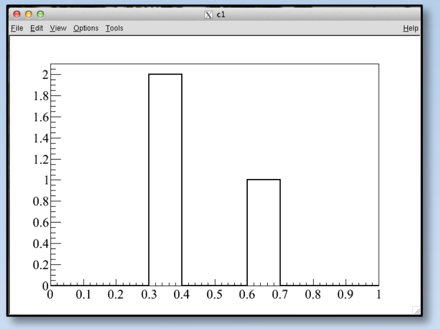
New Open Recent Save Print Undo Redo Cut Copy Paste Search Preferences Help

```
#include "TH1D.h"
#include "TFile.h"

#include <iostream>
using namespace std;



//------------------------
void Histogram_example_03(){
//------------------------
//----------------------------------------
//-- Goal: o read a histogram from a file
//         o make a copy and plot it
//----------------------------------------

  //-- open the file and make a copy of the histogram
  TDirectory* dir = gDirectory;
  TFile *file = new TFile("MySummaryFile.root", "READ");
  dir->cd();
  TH1D *hist_copy = (TH1D*) file->Get("hist")->Clone("hist_copy");
  file->Close();

  //-- Draw the histogram
  hist_copy->Draw();

  return;

} // end Histogram_example_03()
```

-:--- Example03.C All (32,0) (C++/l Abbrev)

→ *open the root file &*
make a copy of the histogram

→ *draw histogram*



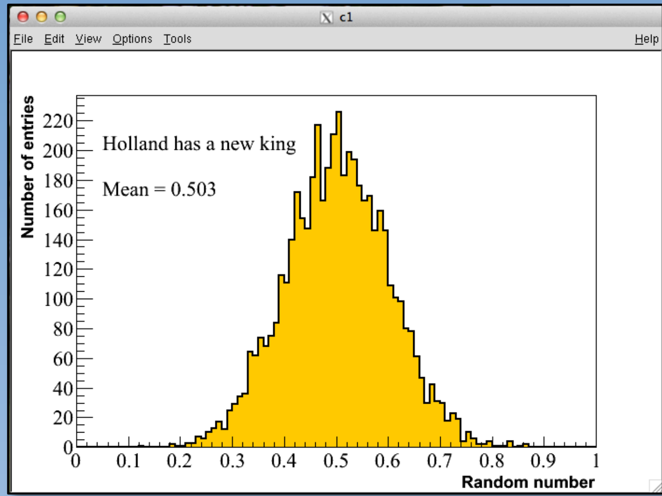
Histogram_Example_04.C

"Fill histogram with random numbers and add some text"

Example04.C

```
#include "TH1D.h"
#include "TColor.h"
#include "TLatex.h"
#include "TRandom3.h"

//------------------------------------------
void Histogram_example_04(int Nrandom = 5000){
//------------------------------------------
//------------------------------------------------------------------------------
//-- Goal: o Fill a histogram with <Nrandom> numbers
//         o plot it and add some information
//------------------------------------------------------------------------------

  //-- book histogram (100 bins between 0. and 1.)
  TH1D *hist = new TH1D("hist", "My dummy histogram", 100,0.,1.);

  //-- fill the histogram with Nrandom random numbers
  TRandom3 *R = new TRandom3();
  for(int i_random = 0; i_random < Nrandom ; i_random++ ){
    double random_number = R->Gaus(0.5,0.1); // random number: center at 0.5 and sigma = 0.1
    hist->Fill(random_number);
  }

  //-- make histogram orange and draw it
  hist->SetFillColor(kOrange);
  hist->Draw();

  //-- add axes
  hist->SetXTitle("Random number");
  hist->SetYTitle("Number of entries");

  //-- add some text
  TLatex *text1 = new TLatex(0.05,200., "Holland has a new king");
  text1->SetTextSize(0.05);
  text1->Draw();

  //-- add some text (advanced)
  //-- plot the mean of the histogram at 75% of the hight of the most populated bin
  double mean = hist->GetMean();
  double most_populated_bin = hist->GetMaximumBin();
  double hight_most_populated_bin = hist->GetBinContent(most_populated_bin);
  TLatex *text2 = new TLatex(0.05,0.75*hight_most_populated_bin,Form("Mean = %5.3f",mean));
  text2->SetTextSize(0.05);
  text2->Draw();

  return;

} // end Histogram_example_04()
```

-:--- Example04.C Top (30,0) (C++/l Abbrev)
Beginning of buffer

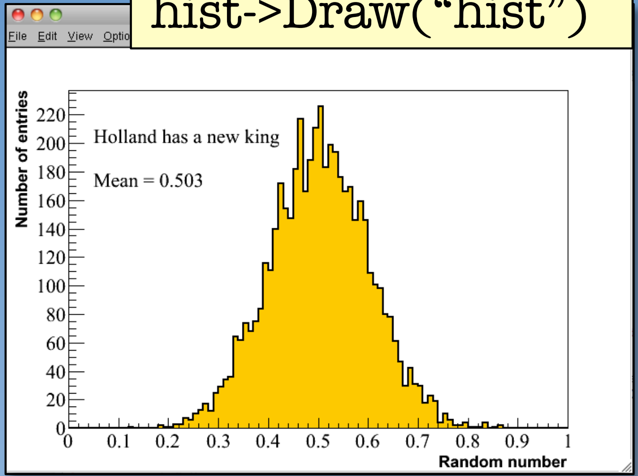


→ *Fill histogram - random numbers*

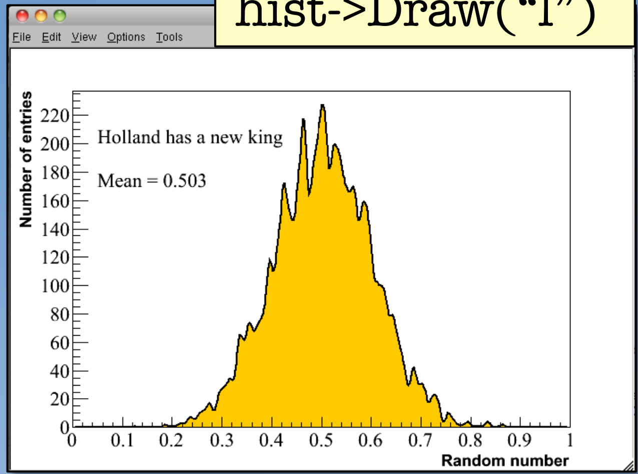
→ *Give it your favorite color and add axes titles*

→ *Add some text in the plot*

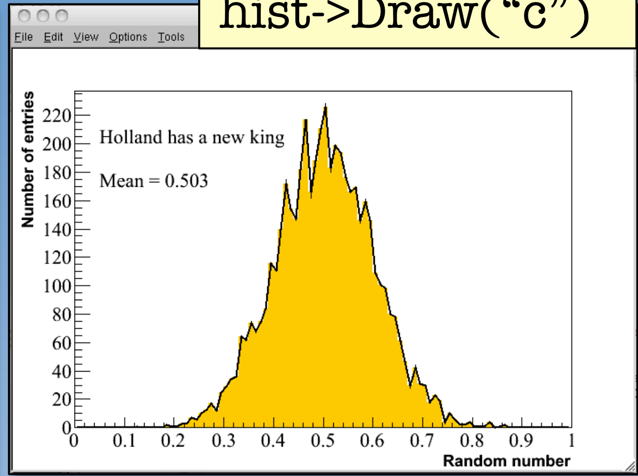
→ *Plot the mean at 75% of the height of the maximum bin* hist->Draw(“hist”)



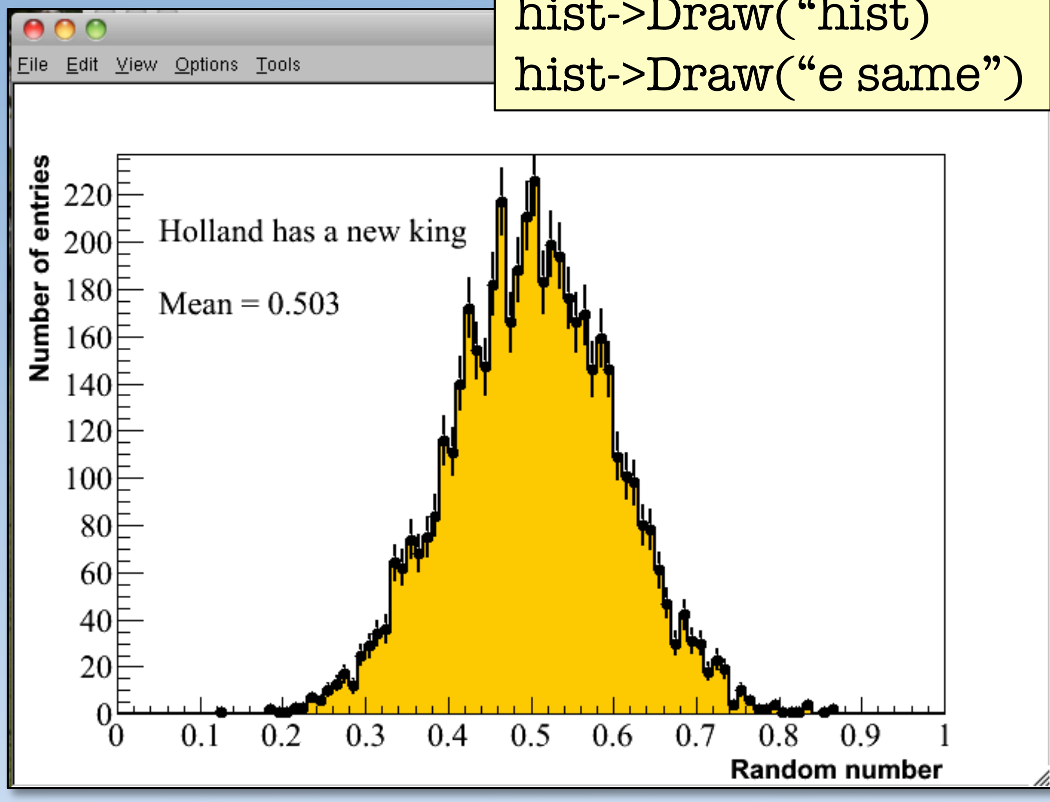
hist->Draw(“l”)



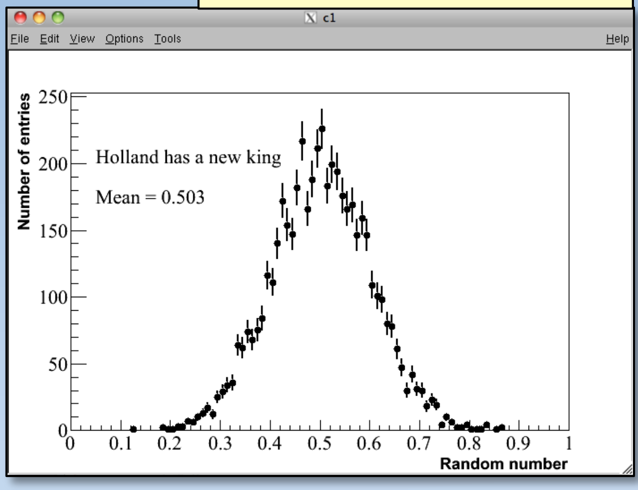
hist->Draw(“c”)



hist->Draw(“hist)
hist->Draw(“e same”)



hist->Draw(“e”)



Histogram_Example_05.C

“Prepare a ‘typical’ Higgs discovery plot”

Example05.C

```cpp
#include "TArrow.h"
#include "TCanvas.h"
#include "TH1D.h"
#include "TLine.h"
#include "TColor.h"
#include "TCanvas.h"
#include "TROOT.h"
#include "TLatex.h"
#include "TFile.h"
#include "TLegend.h"
#include "TLegendEntry.h"
#include "TRandom3.h"

#include <iostream>
using namespace std;

//===========
// Histogram
//===========



//-----------------------------------------
void Histogram_example_05(int Irebin = 20){
//-----------------------------------------
//--------------------------------------------------------------------------
//-- Goal: o Read 2 histograms from a file (signal and background)
//           Rebin them with a factor 20
//         o Prepare stacked histogram (sig+bgr)
//         o print number of events in a region close to 125 GeV
//         o show combined plot
//-------------------------------------------------------------------------

  //-----------------------------------------------------------------------
  //-- [1] Get the Higgs signal, background and data histogram from a file
  //        Prepare sig+bgr distribution and plot them
  //-----------------------------------------------------------------------
  TDirectory* dir = gDirectory;
  TFile *file = new TFile("Histograms_fake.root", "READ");
  dir->cd();
  TH1D *h_sig  = (TH1D*) file->Get("h_m4l_Higgs125_fake")->Clone("h_sig");
  TH1D *h_bgr  = (TH1D*) file->Get("h_m4l_ZZ_fake")->Clone("h_bgr");
  TH1D *h_data = (TH1D*) file->Get("h_m4l_data_fake")->Clone("h_fake");
  file->Close();

  //-- Rebin histograms (only for plotting)
  h_sig->Rebin(Irebin);
  h_bgr->Rebin(Irebin);
  h_data->Rebin(Irebin);
  printf("\n  INFO: Rebinning the histograms with a factor %d. Binwidth is now %5.2f GeV\n\n", Irebin, h_data->GetBinWidth(1));

  //-- Prepare cumulative histogram for signal + background
  TH1D *h_sig_plus_bgr = (TH1D* ) h_bgr->Clone("h_sig_plus_bgr");
  h_sig_plus_bgr->Reset();
  for (int i_bin = 1; i_bin < h_bgr->GetNbinsX(); i_bin++){
       h_sig_plus_bgr->SetBinContent( i_bin, h_sig->GetBinContent(i_bin) + h_bgr->GetBinContent(i_bin));
  }

  //-- print mass and number of events in a 10 GeV mass window around 125 GeV
  double mass_bin = 0.;
  double Nevt_bin_sig = 0.;
  double Nevt_bin_bgr = 0.;
  double Nevt_bin_data = 0.;
  for(int i_bin = 1; i_bin< h_bgr->GetNbinsX(); i_bin++){
    mass_bin = h_data->GetBinCenter(i_bin);
    Nevt_bin_sig  = h_sig->GetBinContent(i_bin);
    Nevt_bin_bgr  = h_bgr->GetBinContent(i_bin);
    Nevt_bin_data = h_data->GetBinContent(i_bin);
    if(fabs(mass_bin-125)<10.){
      printf("  Bin %3d: mass = %5.2f | Nsig = %5.2f  Nbgr = %5.2f and Ndata = %5.2f\n",i_bin, mass_bin, Nevt_bin_sig, Nevt_bin_bgr, Nevt_bin_data);
    }
  }

  //----------------------
  //-- [2] Plot histograms
  //----------------------

  //-- Prepare canvas
  TCanvas * canvas1 = new TCanvas( "canvas1","Standard Canvas",600,400);
  canvas1->cd();

  //-- plot histograms (sig+bgrin higgs color and a restricted mass range on x-axis)
  h_sig->SetFillColor(7);
  h_sig_plus_bgr->SetFillColor(7);
  h_sig_plus_bgr->SetAxisRange(0.,25.,"Y");
  h_sig_plus_bgr->SetAxisRange(50.,400.,"X");
  h_sig_plus_bgr->Draw("hist");

  //-- add the background (plot it over the sig+bgr distribution)
  h_bgr->SetFillColor(2);
  h_bgr->Draw("same");
  h_bgr->Draw("axis same");
  h_data->Draw("e same");

  //-- add axes
  h_sig_plus_bgr->SetXTitle("4-lepton invariant mass [GeV]");
  h_sig_plus_bgr->SetYTitle(Form("Number of events / %3.1f GeV",h_bgr->GetBinWidth(1)));

  //-- create a legend
  TLegend *legend = new TLegend(0.65,0.65,0.90,0.80);
  legend->SetBorderSize(0);
  legend->SetFillColor(0);
  TLegendEntry *legend_entry1 = legend->AddEntry(h_sig," Higgs", "f");
  TLegendEntry *legend_entry2 = legend->AddEntry(h_bgr," ZZ(SM)", "f");
  legend_entry1->SetTextSize(0.05);
  legend_entry2->SetTextSize(0.05);
  legend->Draw();

  //-- prepare gif
  canvas1->Print("./Histogram_example_5.gif");

  return;

} // end Histogram_example_05()
```

ROOT

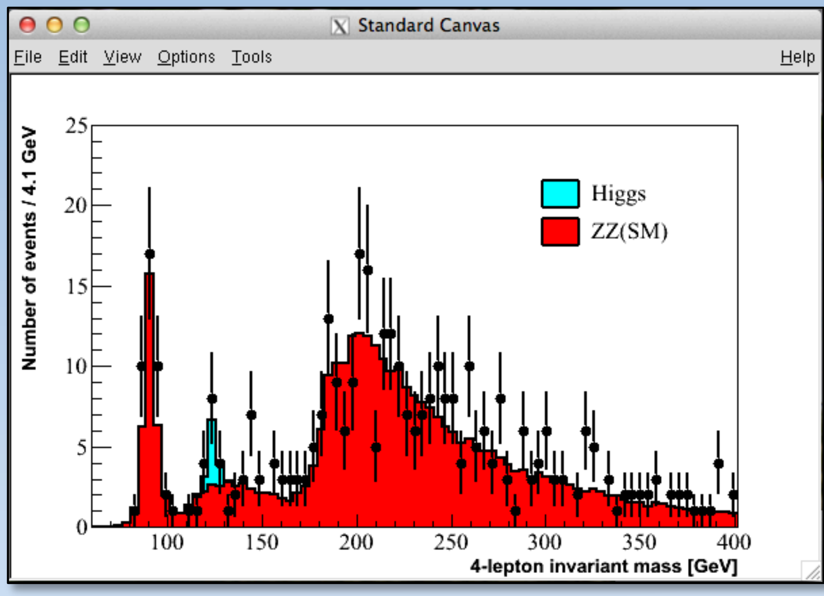
```
bash-3.2$ root -l

 Standard RootLogon

root [0] .L Example05.C++
Info in <TUnixSystem::ACLiC>: creating shared library /Users/ivo/Dropbox/2013_07_HASCO/code/./Example05_C.so
root [1] Histogram_example_05(20)

  INFO: Rebinning the histograms with a factor 20. Binwidth is now  4.11 GeV

  Bin  14: mass = 115.50 | Nsig =  0.23  Nbgr =  1.83 and Ndata =  1.00
  Bin  15: mass = 119.61 | Nsig =  0.94  Nbgr =  2.22 and Ndata =  4.00
  Bin  16: mass = 123.72 | Nsig =  4.00  Nbgr =  2.66 and Ndata =  8.00
  Bin  17: mass = 127.83 | Nsig =  1.34  Nbgr =  2.54 and Ndata =  4.00
  Bin  18: mass = 131.94 | Nsig =  0.07  Nbgr =  2.83 and Ndata =  1.00
Info in <TCanvas::Print>: GIF file ./Histogram_example_5.gif has been created
root [2]
```

→ *Read histograms from file. Also **rebin** them*

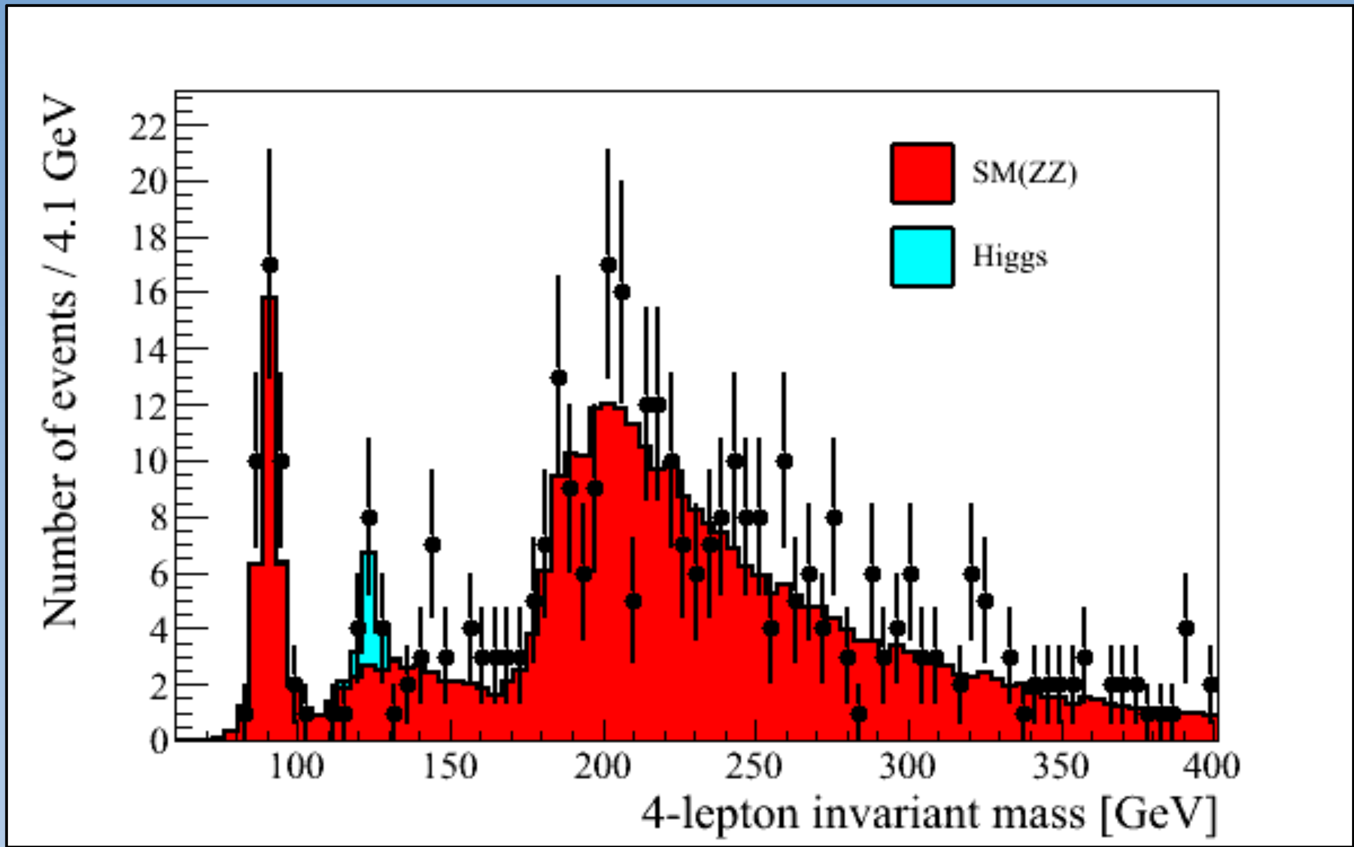
→ *Prepare cumulative histogram*

→ *Determine #events (sig, bgr, data) around 125 GeV*

→ *plot*



Data-set for the exercises: 4 lepton mass



We’ll try to squeeze as much info out of this distribution as we can

Ntuples (Ttree)

Typical "event" consists of several (complex) objects

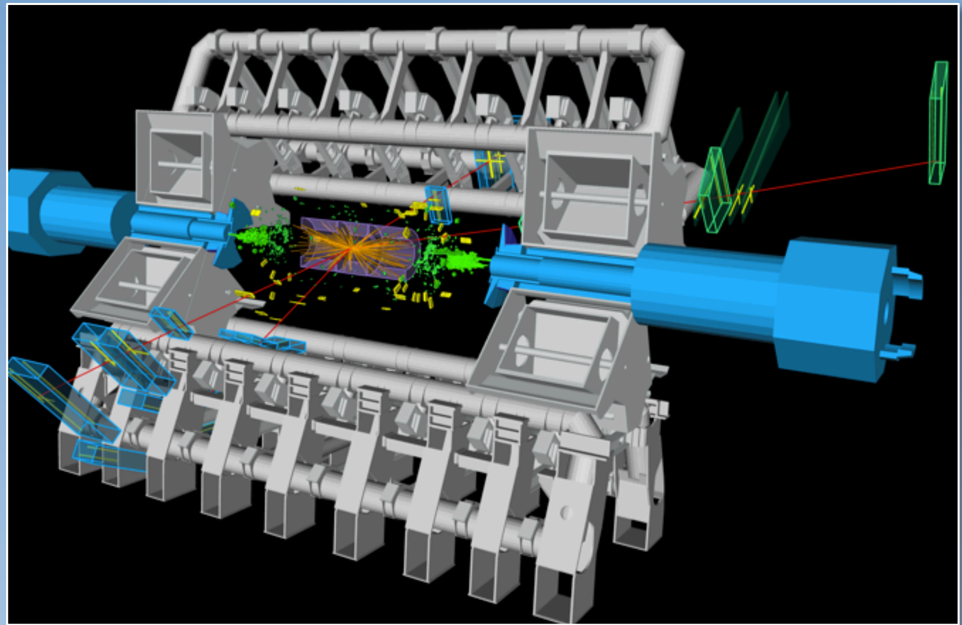
Single variables:
Event number, run number, lumi block, time, date, #muons, #tracks, #clusters, trigger info, …

Tracks:
Vector-like objects: (d_0, z_0, φ_0, θ, q/p), but also isolation, particle type, cluster links, …

Muons, electrons, jets
Muon algorithm types, track, energy, isolation corrections, …

Calorimeter clusters:
Pattern, elements, offsets, …

Jets:
Clusters, tracks, b-tag info, mass, …



Analysis: Complex data structures and librairies

↓

Simplest 'flat' form is an ***Ntuple***

often the endpoint of an physics analysis

Ntuple_create.C

"Prepare a simple Ntuple and save in file"

```
#include "TH1D.h"
#include "TFile.h"
#include "TNtuple.h"
#include "TRandom3.h"

//------------------------
void NtupleCreation(){
//------------------------

  //-- Open a file
  TFile *file = new TFile("MySimpleNtuple.root","RECREATE");

  //-- Create an Ntuple
  TNtuple *ntuple = new TNtuple("ntuple", "my ntuple file", "x:y:z");

  //-- Fill the ntuple with 1000 random variables x,y,z
  double x, y,z;
  TRandom3 *R = new TRandom3();
  int Nevent = 1000;
  for(int i_event = 0; i_event < Nevent ; i_event++ ){
    x = R->Gaus(50.,5.);
    y = R->Gaus(50.,10.);
    z = R->Gaus(50.,15.);
    ntuple->Fill(x,y,z);
  }

  //-- Save the file
  file->Write();

  return;

} // end NtupleCreation()
```

-:--- Ntuple_create.C All (32,0) (C++/l Abbrev)

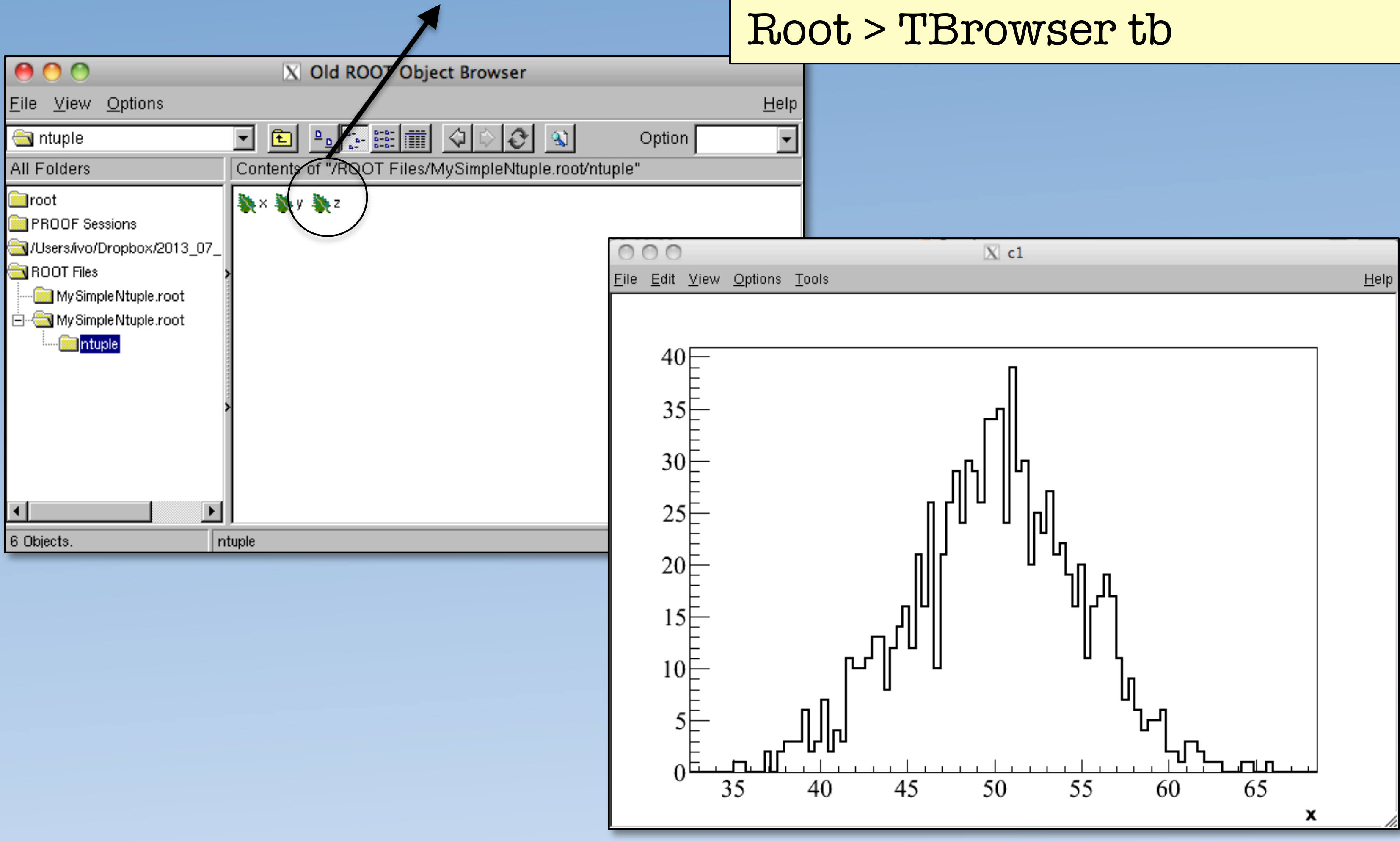
- *Open the fil: MySimpleNtuple.root*
- *Create the Ntuple*
- *Fill the Ntuple*
- *Write the file*

Root > .L Ntuple_create.C++
Root > NtupleCreation()

Looking at an Ntuple: browser

Unix > root –l MySimpleNtuple.root
Root > TBrowser tb

leaf



Looking at an Ntuple: MakeClass

Unix > root –l MySimpleNtuple.root
Root > .ls
Root > ntuple->MakeClass("MyNtupleAnalysis")

Header file: MyNtupleAnalysis.h

```
#ifndef MyNtupleReader_h
#define MyNtupleReader_h

#include <TROOT.h>
#include <TChain.h>
#include <TFile.h>

class MyNtupleReader {
public :
   TTree          *fChain;   //!pointer to the analyzed TTree or TChain
   Int_t           fCurrent; //!current Tree number in a TChain

   // Declaration of leaf types
   Float_t         x;
   Float_t         y;
   Float_t         z;

   // List of branches
   TBranch        *b_x;   //!
   TBranch        *b_y;   //!
   TBranch        *b_z;   //!

   MyNtupleReader(TTree *tree=0);
   virtual ~MyNtupleReader();
   virtual Int_t    Cut(Long64_t entry);
   virtual Int_t    GetEntry(Long64_t entry);
   virtual Long64_t LoadTree(Long64_t entry);
   virtual void     Init(TTree *tree);
   virtual void     Loop();
   virtual Bool_t   Notify();
   virtual void     Show(Long64_t entry = -1);
};

```

-:--- MyNtupleReader.h 8% (31,0) (C/l Abbrev)

→ *Variables & types*

→ *GetEntry()*

→ *Loop()*

C-file: MyNtupleAnalysis.C

```
#define MyNtupleReader_cxx
#include "MyNtupleReader.h"
#include <TH2.h>
#include <TStyle.h>
#include <TCanvas.h>

void MyNtupleReader::Loop()
{
//   In a ROOT session, you can do:
//      Root > .L MyNtupleReader.C
//      Root > MyNtupleReader t
//      Root > t.GetEntry(12); // Fill t data members with entry number 12
//      Root > t.Show();       // Show values of entry 12
//      Root > t.Show(16);     // Read and show values of entry 16
//      Root > t.Loop();       // Loop on all entries
//

   if (fChain == 0) return;

   Long64_t nentries = fChain->GetEntriesFast();

   Long64_t nbytes = 0, nb = 0;
   for (Long64_t jentry=0; jentry<nentries;jentry++) {
      Long64_t ientry = LoadTree(jentry);
      if (ientry < 0) break;
      nb = fChain->GetEntry(jentry);   nbytes += nb;
      // if (Cut(ientry) < 0) continue;
   }
}
```

-:--- MyNtupleReader.C All (29,1) (C++/l Abbrev)

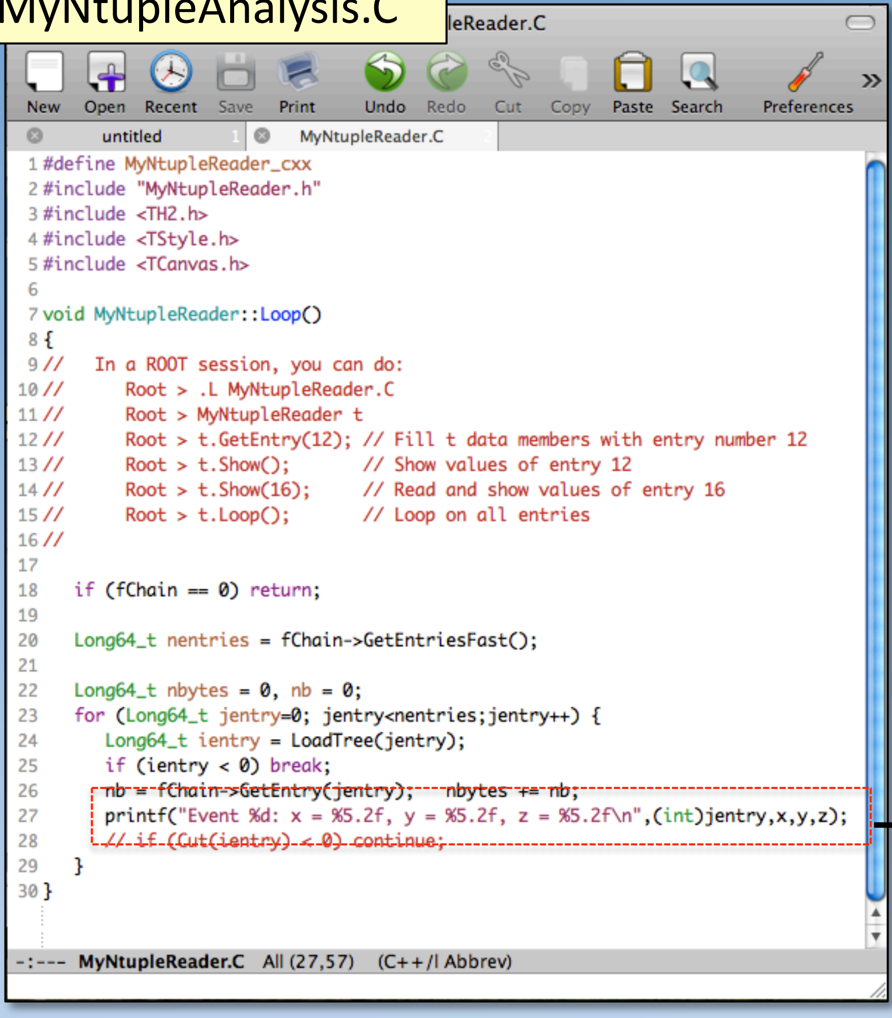
→ *Loop()*

→ *Loop over entries*

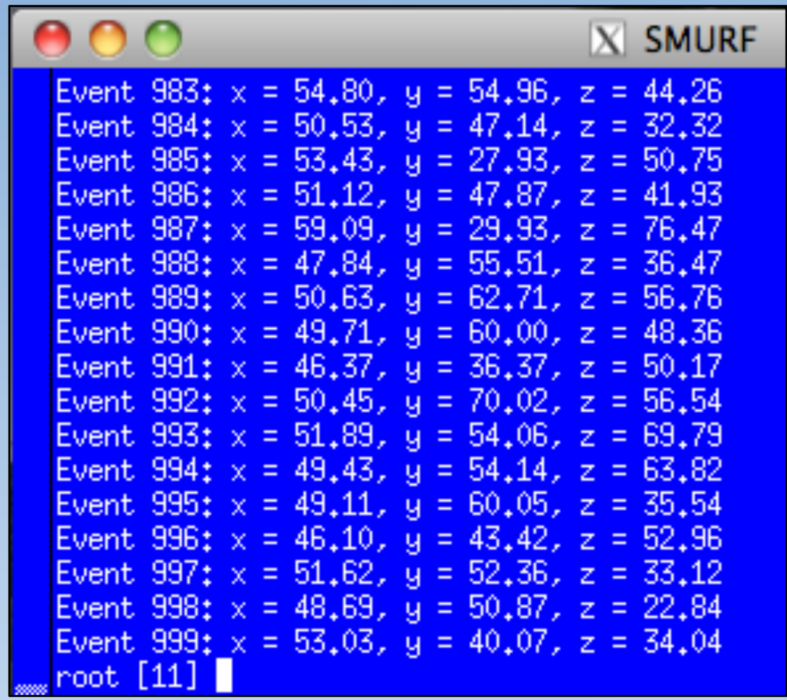
→ *Somewhere a link to MySimpleNtuple.root*

Running your code

MyNtupleAnalysis.C



Unix > root
Root > .L MyNtupleReader.C++
Root > MyNtupleReader mnr
Root > mnr.Loop()



ONE extra line to MyAntupleAnalysis.C:

```
printf("Event %d: x = %5.2f, y = %5.2f, z = %5.2f\n",(int)jentry,x,y,z);
```

Statistics

Statistics

1) Measurement:

- Likelihood fit

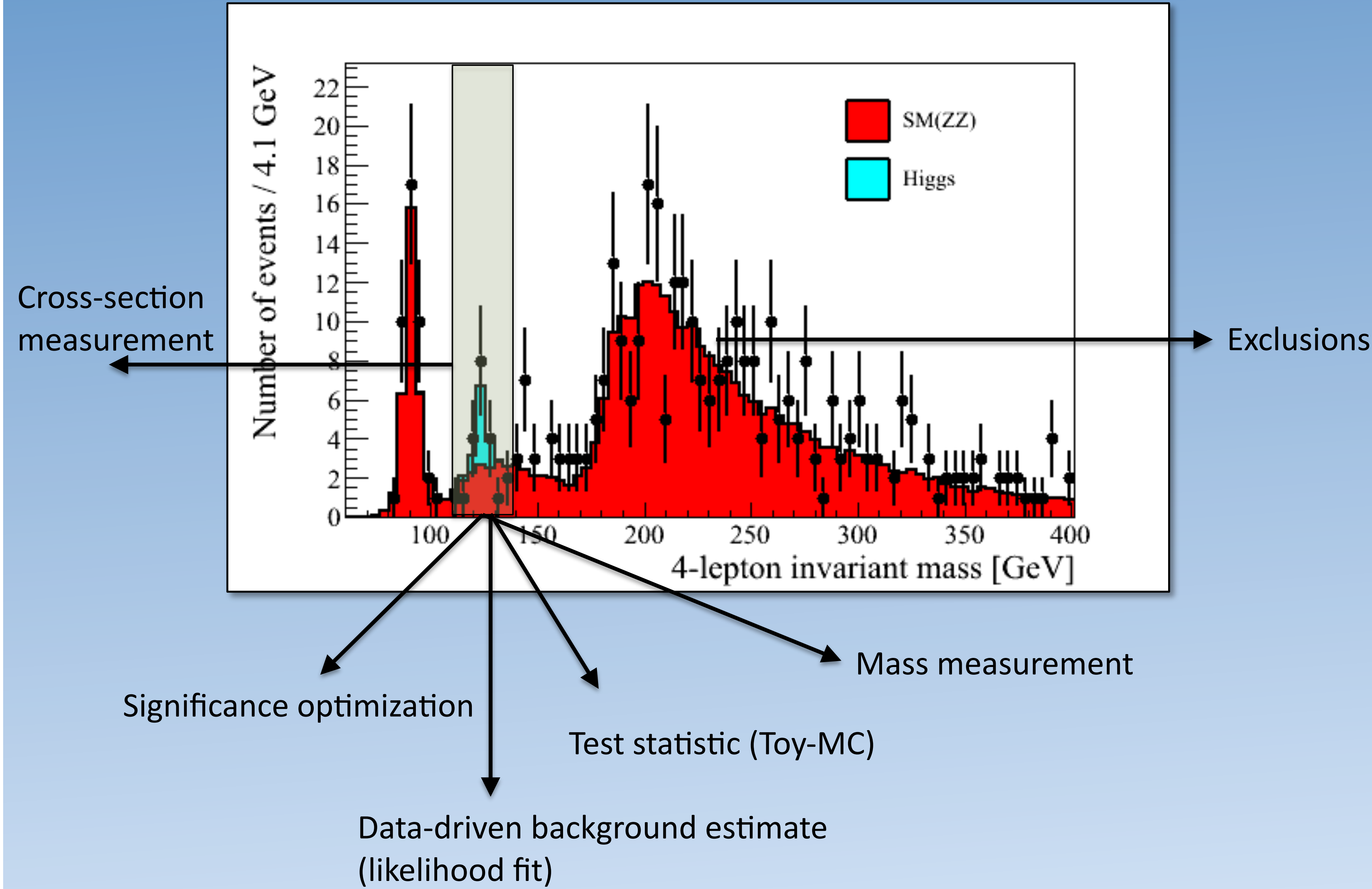
2) Hypotheses testing & search:

- Significance (definition and optimization)

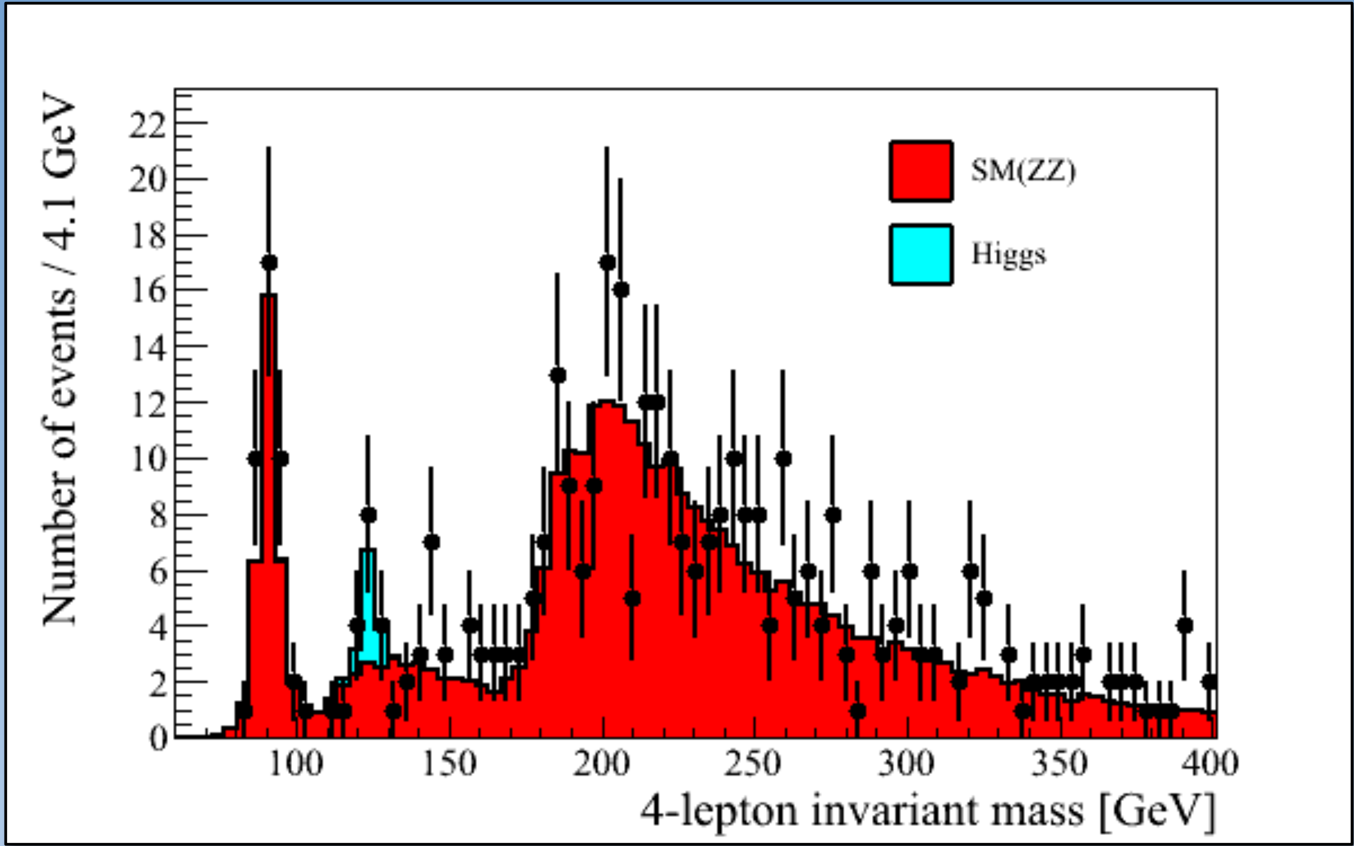
- Test-statistic & pseudo-experiments (toy-MC)

- Exclusion/discovery

Data-set for the exercises: 4 lepton mass



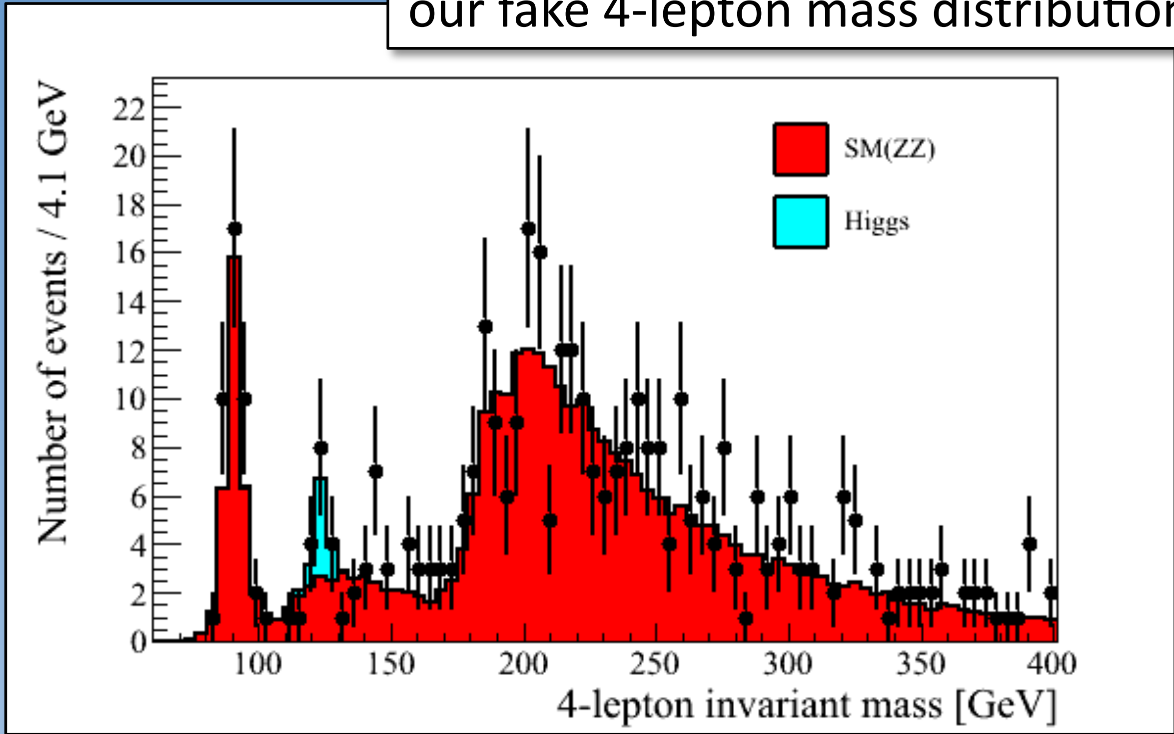
The data-set for the exercises



Note: - Original histograms have 200 MeV bins

- This is fake data (unfortunately)

In the 1st exercise we use re-binned histograms (4 GeV bins) our fake 4-lepton mass distribution



Set 1: fits & measurements

1. Data-driven background estimate (sideband likelihood fit + Poisson distribution)
2. Higgs signal cross-section measurement
3. *Optional: simultaneous mass+ signal cross-section measurement*

Free beer for those who make it

Set 2: hypothesis testing

3. Significance optimization (counting)
4. Compute test statistic (beyond counting)
5. Toy-MC and create test statistic distribution
6. Interpretation: discovery
7. Interpretation: exclusion

Statistics

1) Measurement:

- Likelihood fit

2) Hypotheses testing & search:

- Significance (definition and optimization)

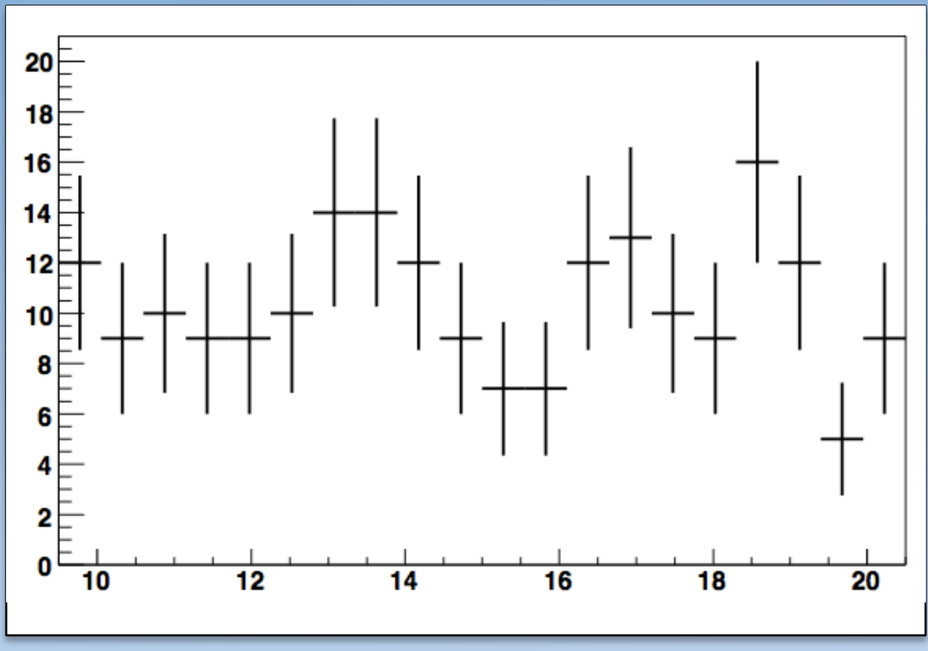
- Test-statistic & pseudo-experiments (toy-MC)

- Exclusion/discovery

Exercise 0

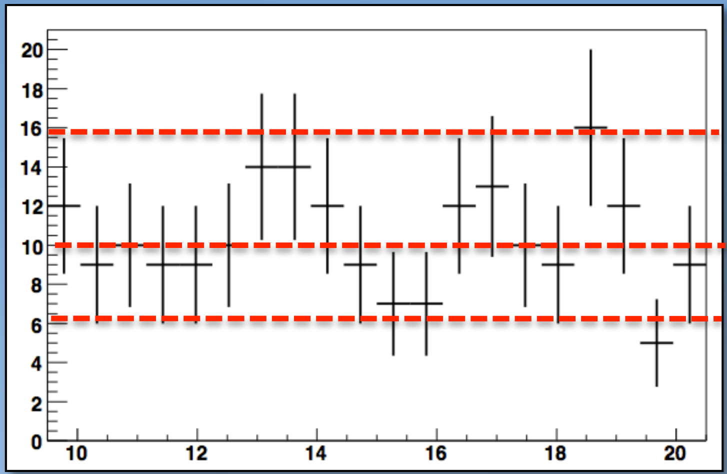
Produce a histogram on the screen

Exercise 0: reproduce this histogram on your screen, compute its mean and create a gif file. Add text if you like.



Note: Look at the example macro's: Histogram_Example_0*.C (*=0,1,2,3,4,5)

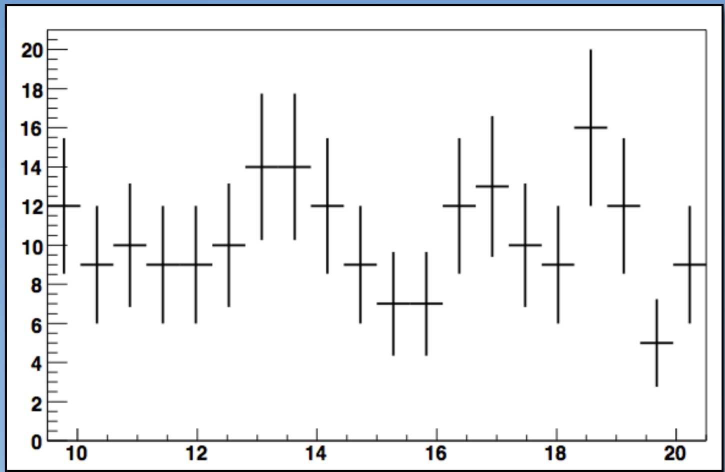
Fitting in 1 slide



You model: $f(x) = \lambda$

Try different values of λ
and for each one compute:

Fitting in 1 slide



You model: $f(x) = \lambda$

Try different values of λ
and for each one compute:

χ^2-fit

Compatibility number :

$$\chi^2 = \sum_{bins} \frac{\left(N_{bin}^{data} - \lambda_{bin}^{expected}\right)^2}{N_{bin}^{data}}$$

Best value:
Value of λ that minimizes χ^2 ($\chi_{min}{}^2$)

Errors:
Values of λ for which $\chi^2 = \chi_{min}{}^2 + 1$

Likelihood-fit

Compatibility number :

$$-2\log(L) = -2 \cdot \sum_{bins} \log(Poisson(N_{bin}^{data} \mid \lambda))$$

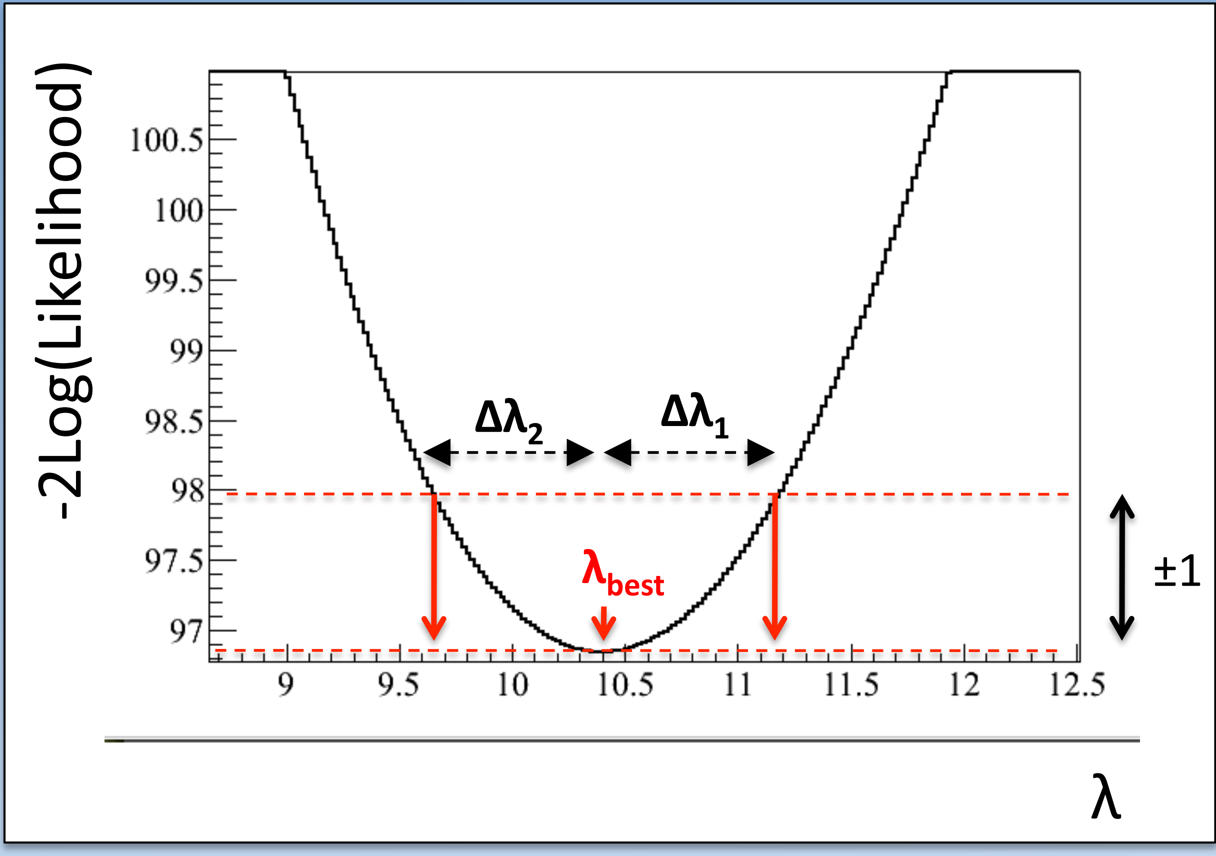
TMath::Poisson(Nevt_bin, λ)

Best value:
Value of λ that minimizes -2Log(L) ($-2\log(L)_{min}$)

Errors:
Values of λ for which 2Log(L) $=$ ($-2\log(L)_{min}$) +1

Result from the fit

$$\text{result}: \quad \lambda = \lambda_{\text{best}\,-\Delta\lambda_2}^{\quad\;+\Delta\lambda_1}$$



Link to Lecture or d'Agostini this morning

Probability (data|α)	≠	Probability(α|data)
What we have computed		What we want
Likelihood		

The Poisson distribution

Binomial with n→∞, p →0 en np=λ

$$P(n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Poisson distribution

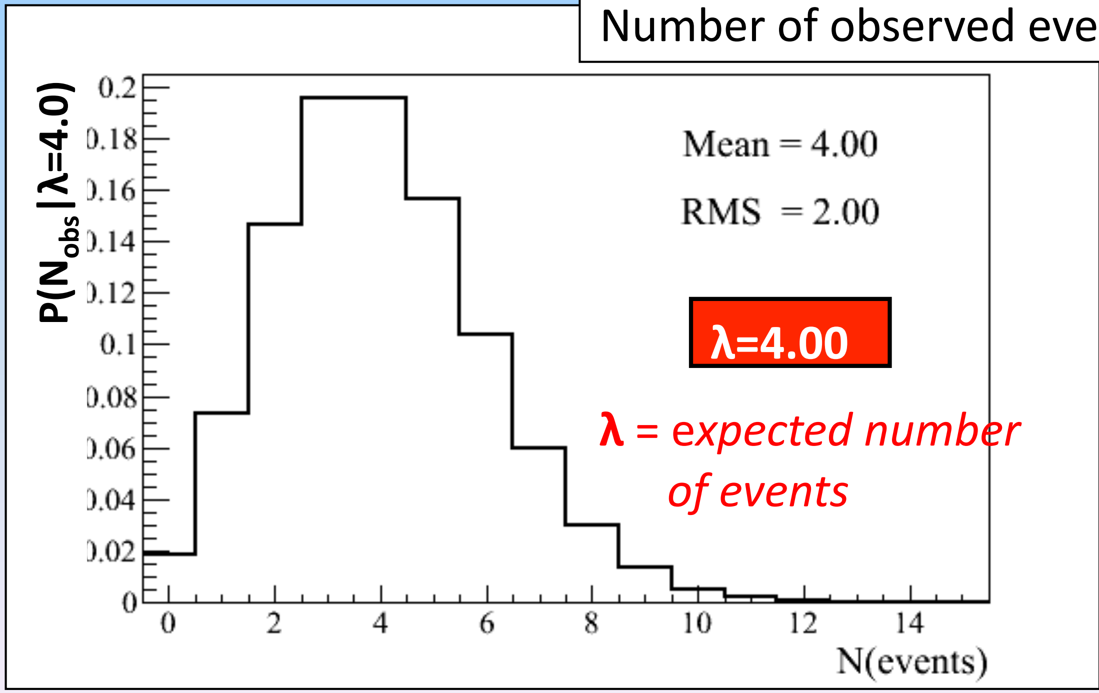
Probability to observe n events when λ are expected

$$P(0 \mid 4.0) = 0.01832$$
$$P(2 \mid 4.0) = 0.14653 \quad !$$
$$P(3 \mid 4.0) = 0.19537$$
$$P(4 \mid 4.0) = 0.19537$$
$$P(6 \mid 4.0) = 0.10420 \quad !$$

#observed → **varying**

λ hypothesis → **fixed**

Number of observed events



λ = *expected number of events*

Known λ (Poisson)

Binomial with n→∞, p →0 en np=λ

$$P(n\,|\,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

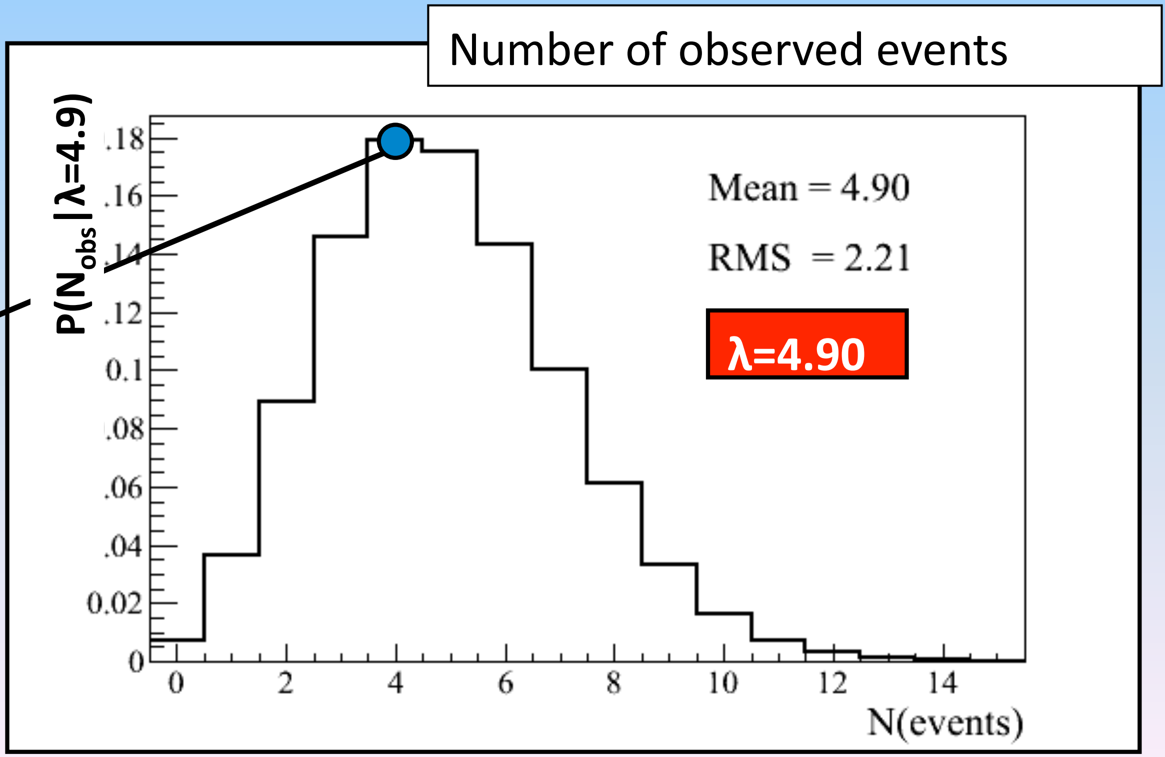
Poisson distribution

Probability to observe n events when λ are expected

$$P(0\,|\,4.9) = 0.00745$$
$$P(2\,|\,4.9) = 0.08940$$
$$P(3\,|\,4.9) = 0.14601$$
$$P(4\,|\,4.9) = 0.17887$$

#observed — **varying**

λ hypothesis — **fixed**



the famous √N

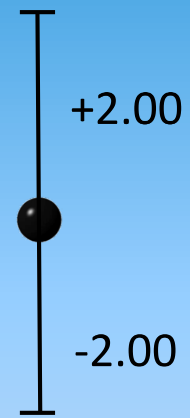
properties

(1) Mean: $\langle n \rangle = \lambda$

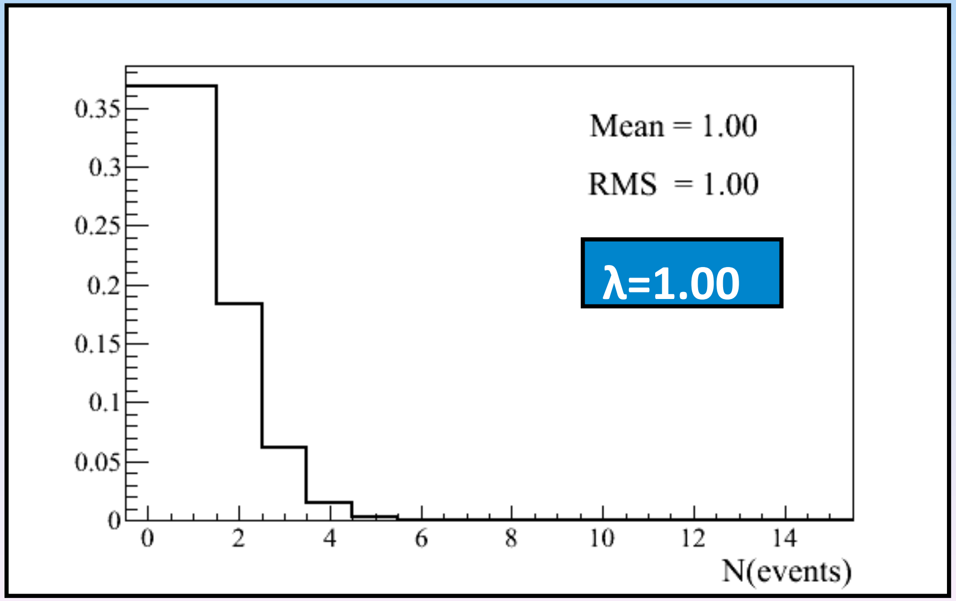
(2) Variance: $\langle (n - \langle n \rangle)^2 \rangle = \lambda$

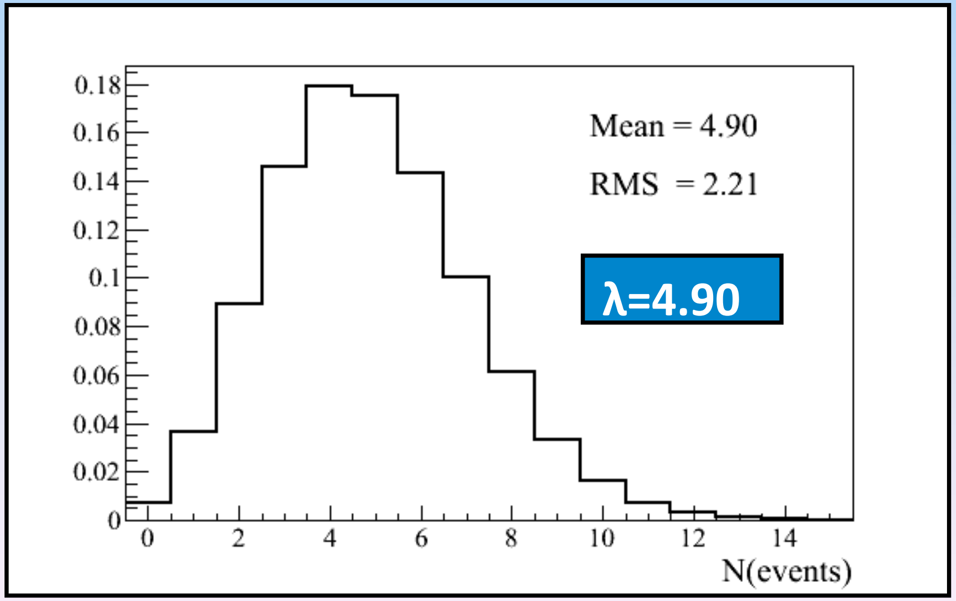
(3) Most likely: first integer ≤ λ

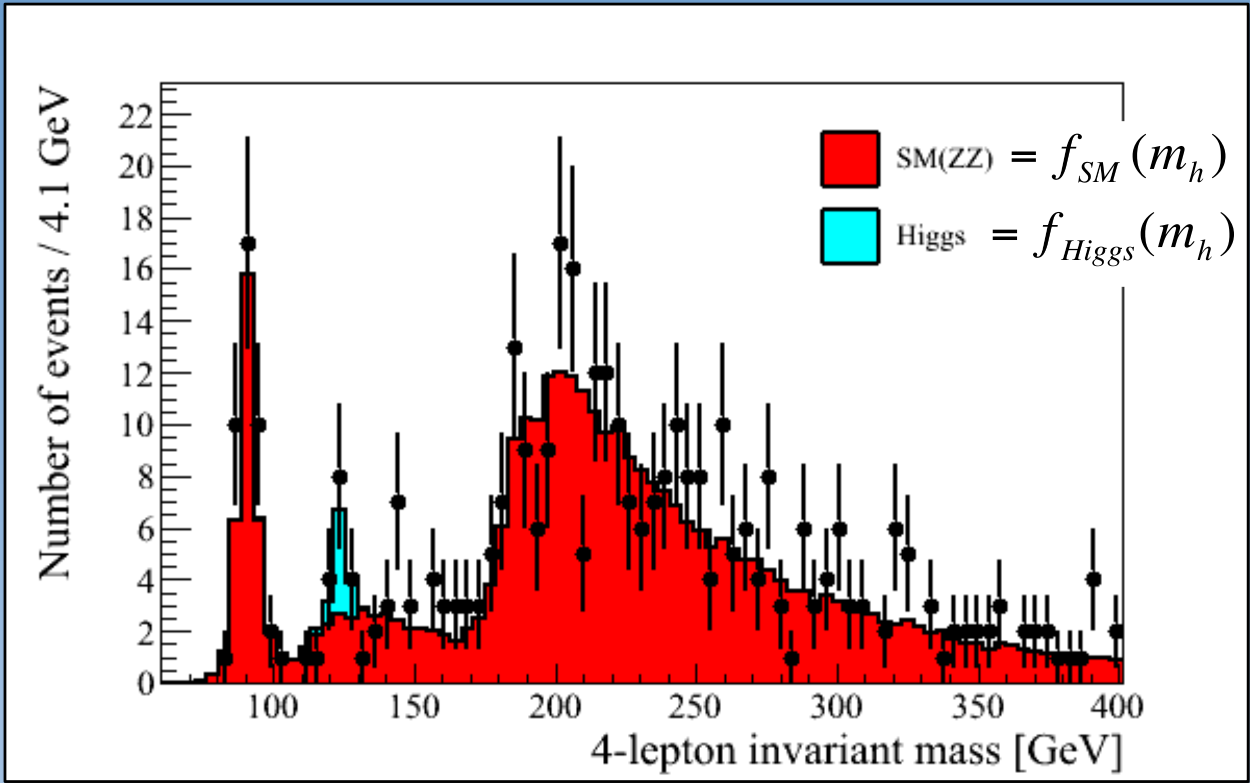
Usual way to represent the error on a data-point



Not ok !







$$f(m_h) = \mu \times f_{Higgs}(m_h) + \alpha \times f_{SM}(m_h)$$

Scale factor for the Higgs

EXERCISE 2

Scale factor for the SM background

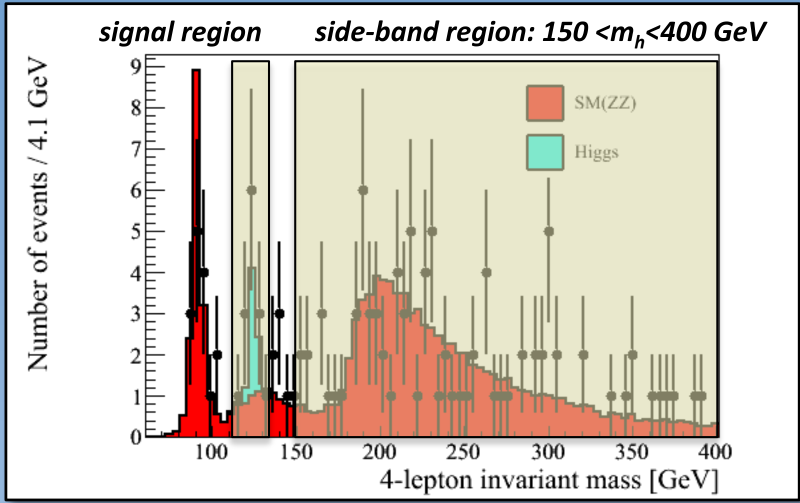
EXERCISE 1

Exercise 1

Data-driven bckground estimate in a
10 GeV mass window around 125 GeV

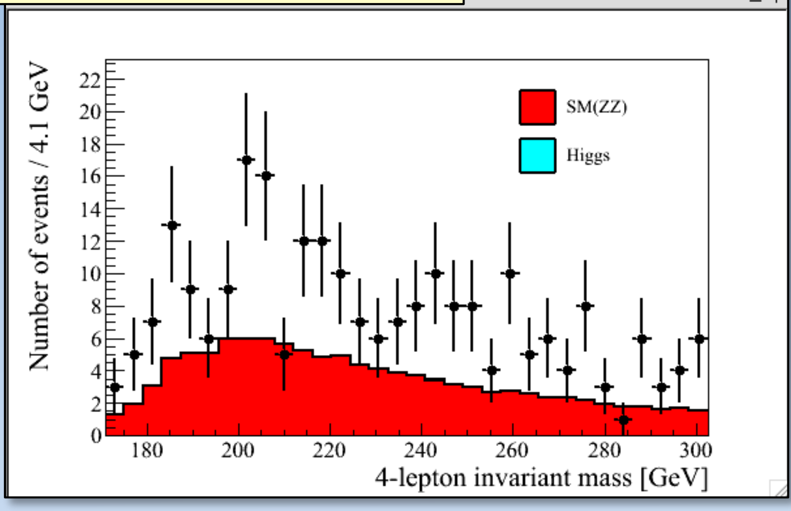
Code you could use:

```
SideBandFit()
```

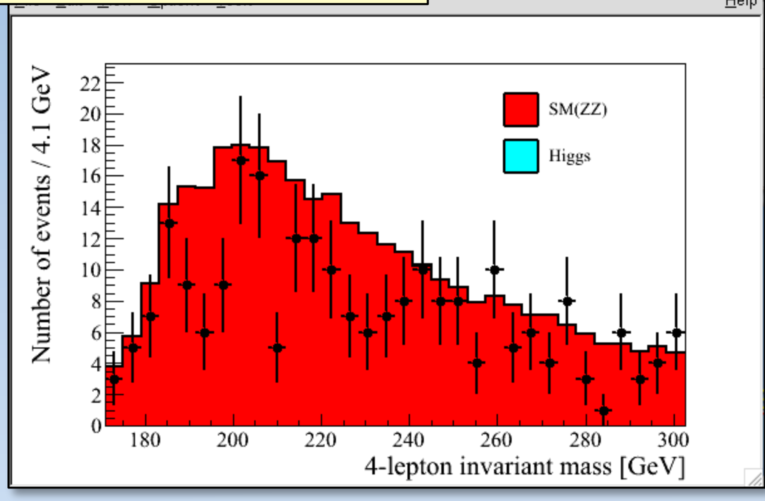


Exercise 1: determine the best estimate for the background scale factor (α) by doing a fit for in the side-band region $175 \leq m_h \leq 300$ GeV (later $150 \leq m_h \leq 400$ GeV)

α = 0.50 (too small)



α = 1.50 (too large)



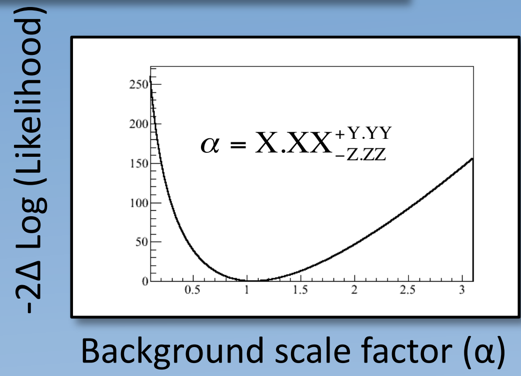
Exercise 1: determine the best estimate for the background scale factor (α) by doing a fit α for in the side-band region 175 ≤ m_h ≤ 300 GeV

1.1 Find the best value α for using a **likelihood fit**

Computing the likelihood:

For each 'guess' of α:

$$-2\log(L) = -2\cdot\sum_{bins}\log(Poisson(N_{bin}^{data} \mid \alpha\cdot f_{bin}^{SM}))$$



1.2 Find the best value α for using a **χ^2 fit**

$$\chi^2 = \sum_{bins}\frac{\left(N_{bin}^{data} - \lambda_{bin}^{expected}\right)^2}{N_{bin}^{data}}$$

Wikipedia: Carl Friedrich Gauss is credited with developing the fundamentals of the basis for least-squares in 1795 at the age of 18. Legendre was the first to publish the method however.

1.3 Discuss the differences between the two estimates

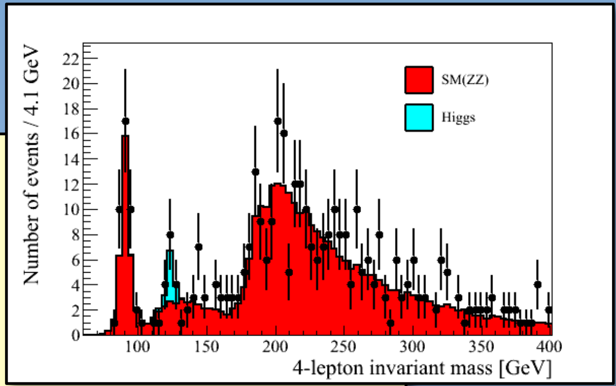
1.4 Redo 1.1 and 1.2 with 150 ≤ m_h ≤ 400 GeV. What happens ?

1.5 Use the likelihood fit, fine binning and 150 ≤ m_h ≤ 400 GeV to determine the best estimate for α ($\alpha_{best} \pm \Delta\alpha_{best}$). Estimate the background level (b ± Δb) in the signal region: 120 ≤ m_h ≤ 130 GeV

Exercise 2

Measurements of Higgs properties

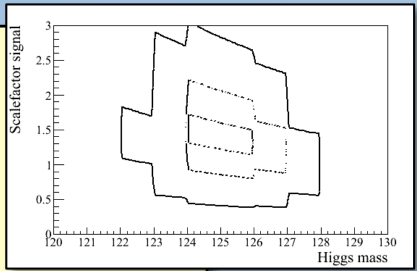
$$-2\cdot\log\left(Likelihood\right) = -2\cdot\sum_{bins}\log\left(Poisson(N_{bin}^{data} \mid \mu\cdot f_{bin}^{Higgs} + \alpha\cdot f_{bin}^{SM})\right)$$



Exercise 2: Measurement of μ (Higgs cross-section w.r.t. SM)

2.1 Do a fit where you leave the signal cross-section and background free. What is the best value for μ and α ?

2.2 What is the uncertainty on μ ? 'Profile' the uncertainty on α.



Bonus for the new Carl Friedrich Gauss'es under us:
Fix α to value from side-band fit and fit for μ and m_h simultaneously

2.3 What is the best value for m_h and μ (and their uncertainties) ?

Note: You'll need to create a template for each value of m_h you test. Assume that the shape of the mass distribution is identical to that at 125 GeV, but correct for changes in cross-section and H→ZZ^* as a function of m_h:
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections

Statistics

1) Measurement:

- Likelihood fit

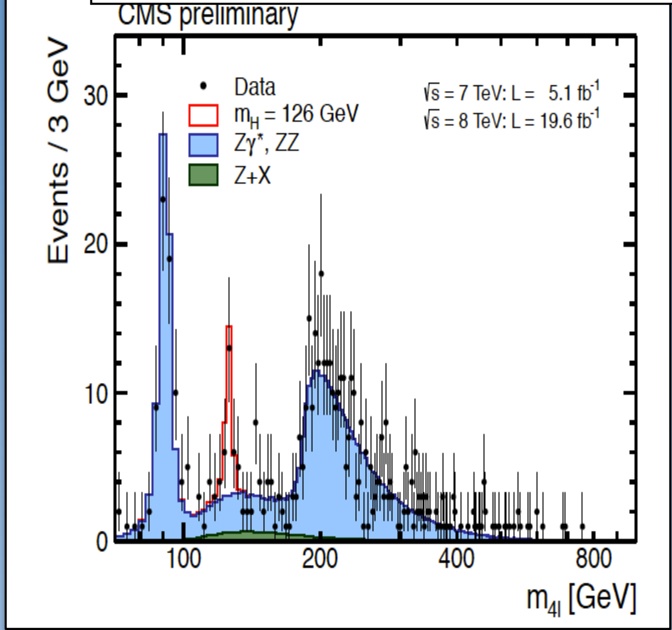
2) Hypotheses testing & search:

- Significance (definition and optimization)

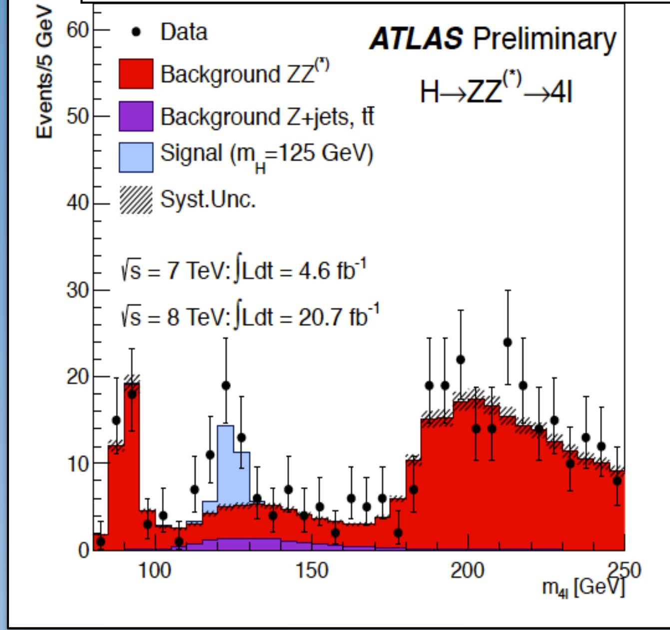
- Test-statistic & pseudo-experiments (toy-MC)

- Exclusion/discovery

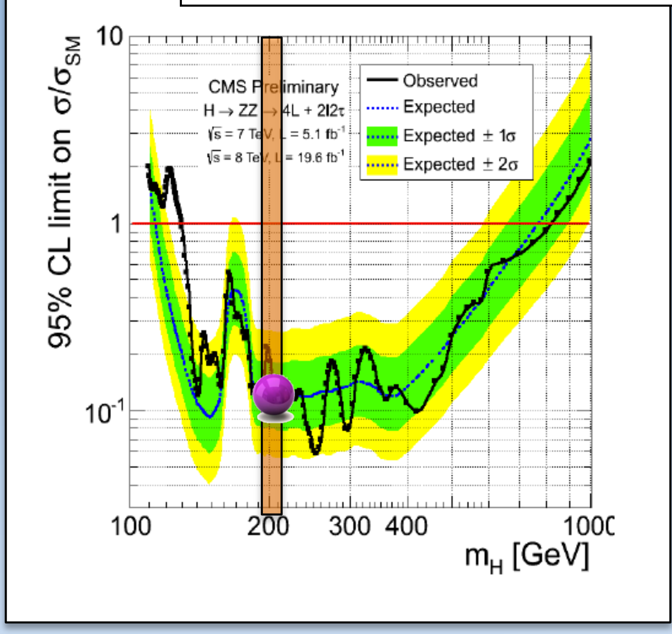
CMS 4 lepton invariant mass



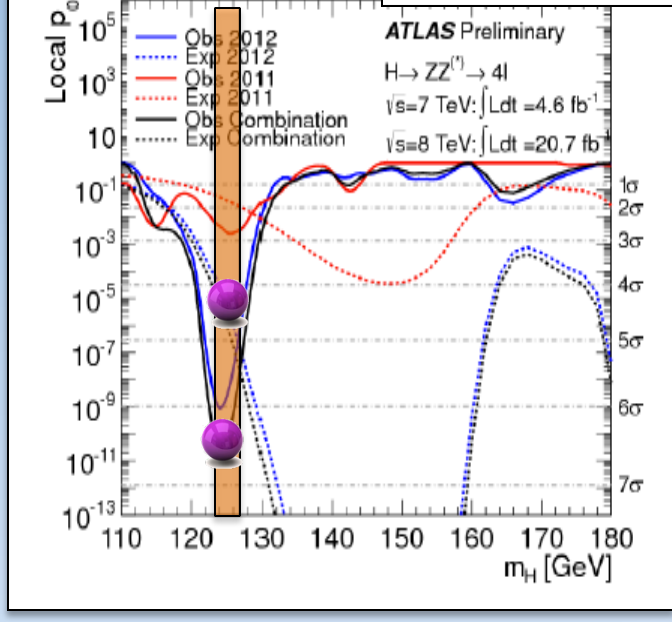
ATLAS 4 lepton invariant mass



Excluded cross-sections

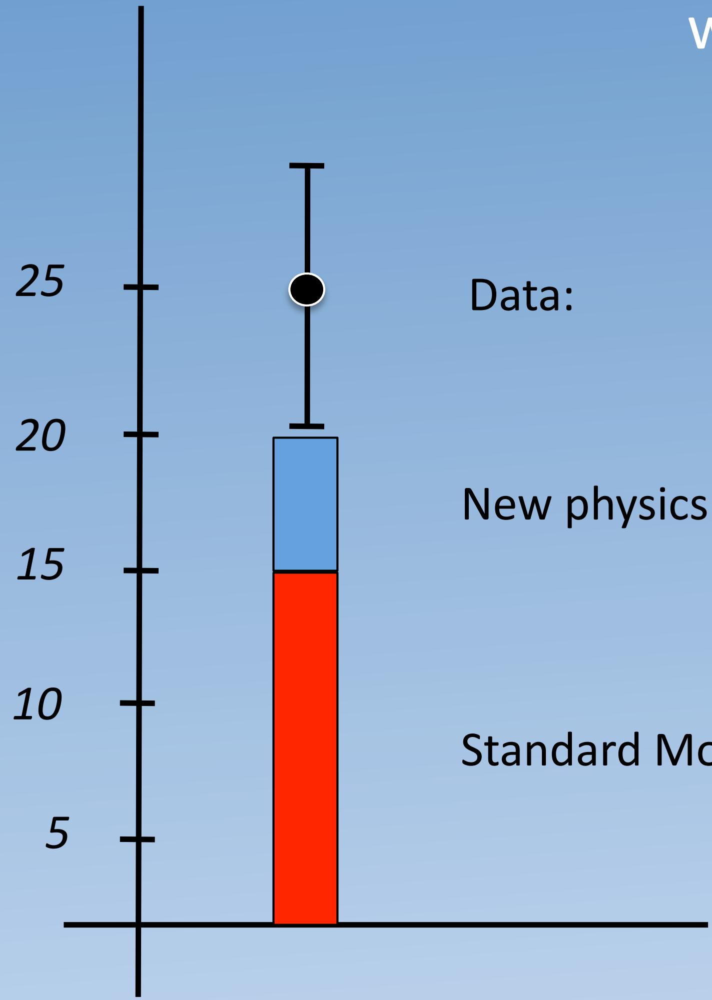


Significances



10 slide mini-lecture on discovery and exclusion

General remark :
what is the significance ?



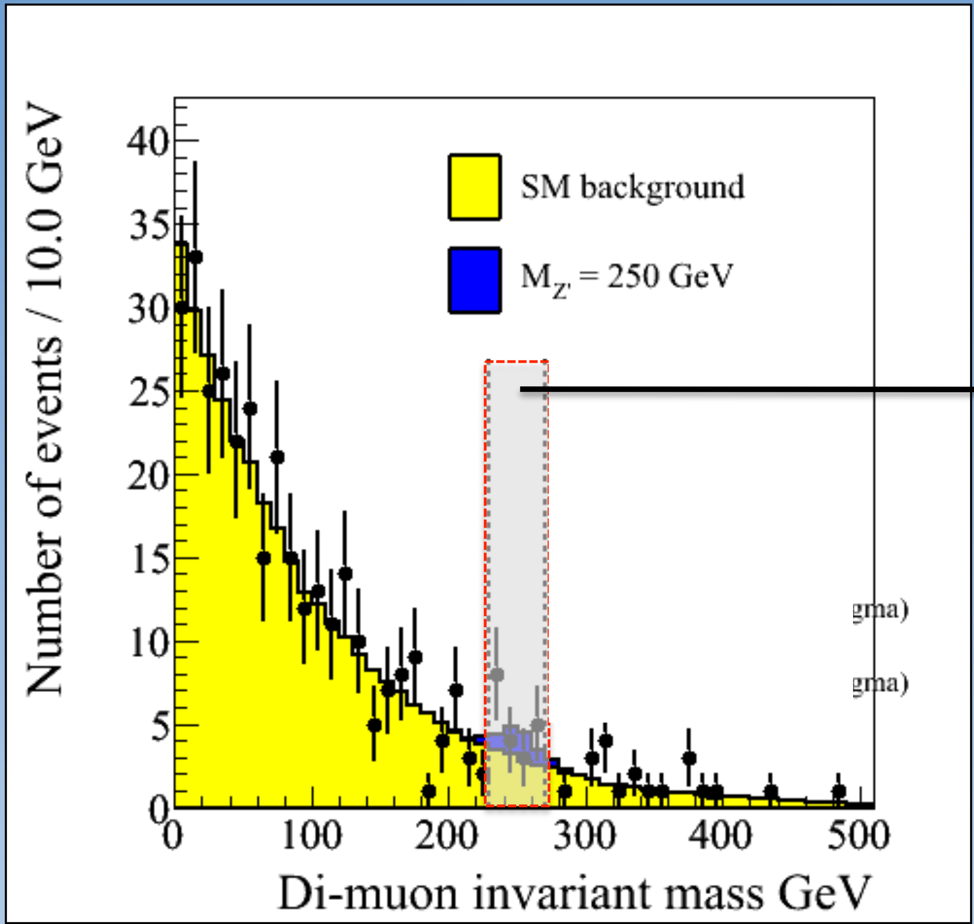
Data: $N_{data} = 25$

New physics: $N_{sig} = 5$

Standard Model: $N_{bgr} = 15$

Significance for N events: probability to observe N events (or even more) under the background-only hypothesis

Counting events in a mass window



Standard Model

SM	10
Higgs	5
Data	12

Ok, now what ?

discovery

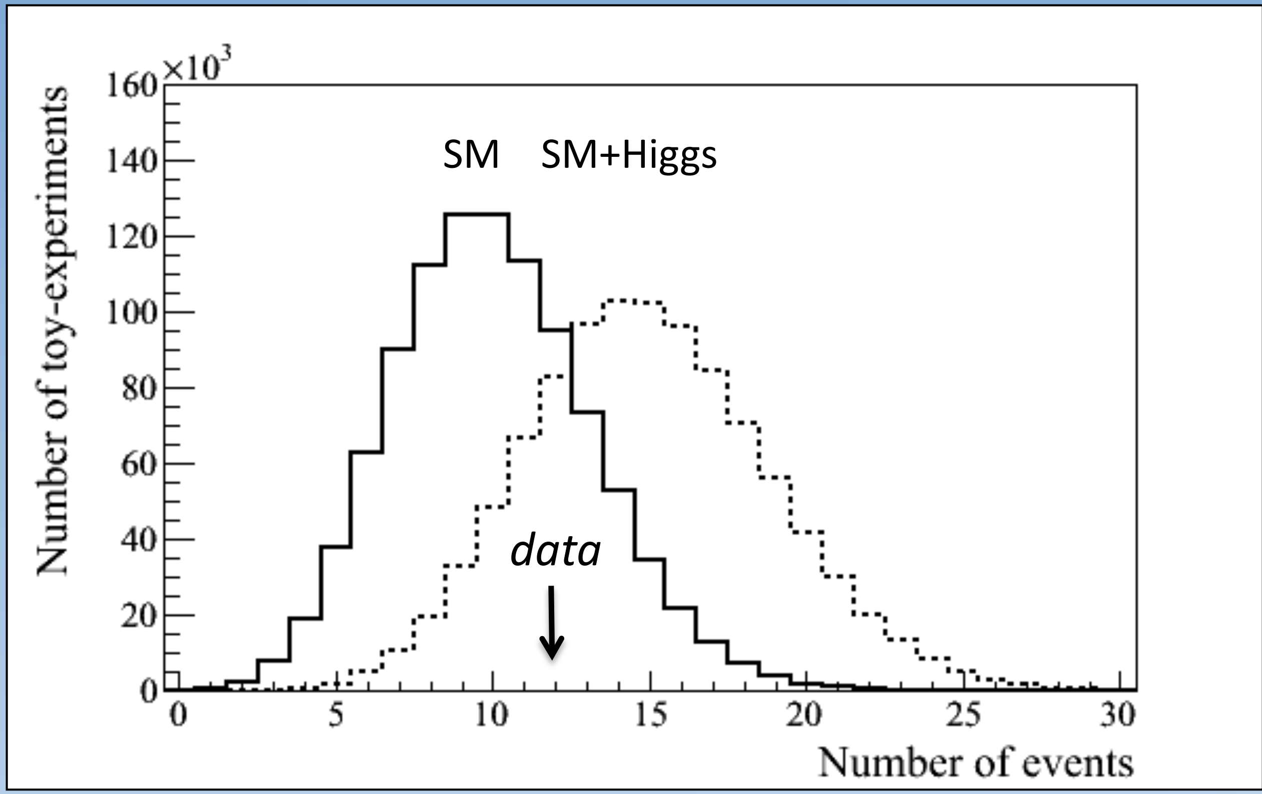
exclusion

Standard Model

SM	10
Higgs	5
Data	12

Ok, now what ?

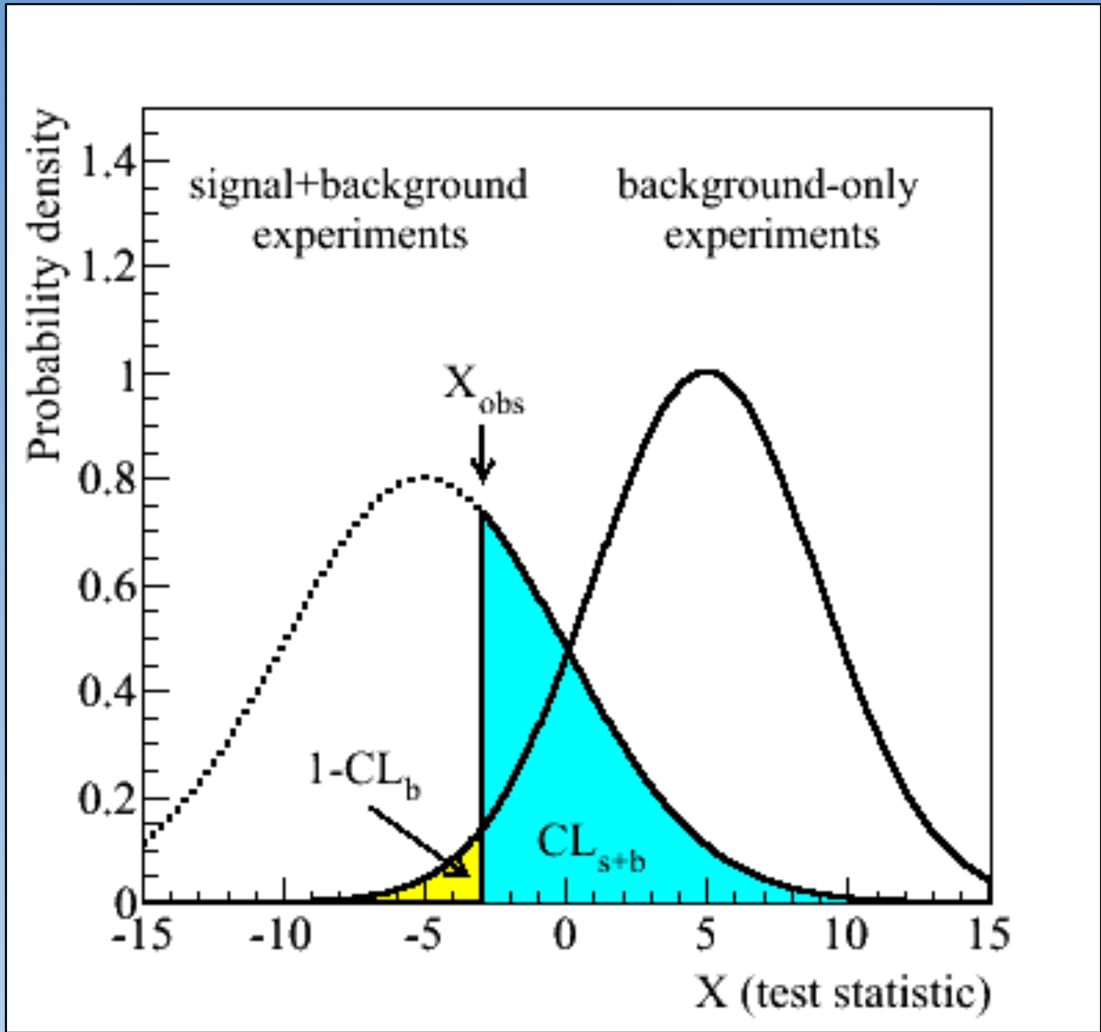
Poisson distribution



Significance for N events: probability to observe N events (or even more) under the background-only hypothesis

X_{obs}: rules for discovery and exclusion

signal like



background like

Discovery: 1-Cl_b <2.87×10^{-7}
Incompatibility with b-only hypothesis

Exclusion: Cl_{s+b} < 0.05
Incompatibility with s+b hypothesis

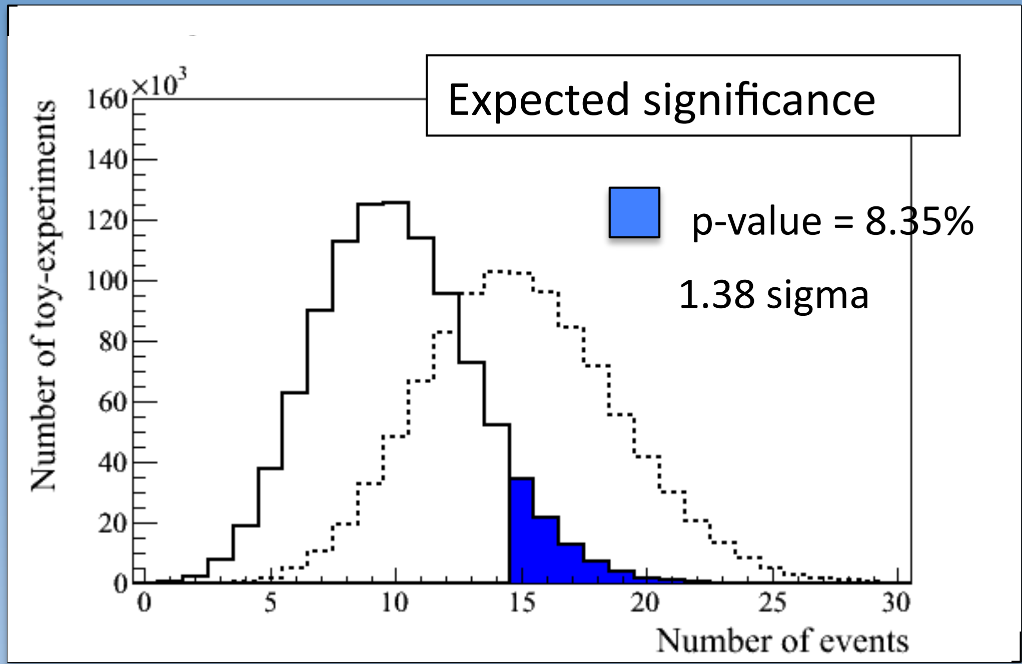
Interpretation

optimistic: discovery

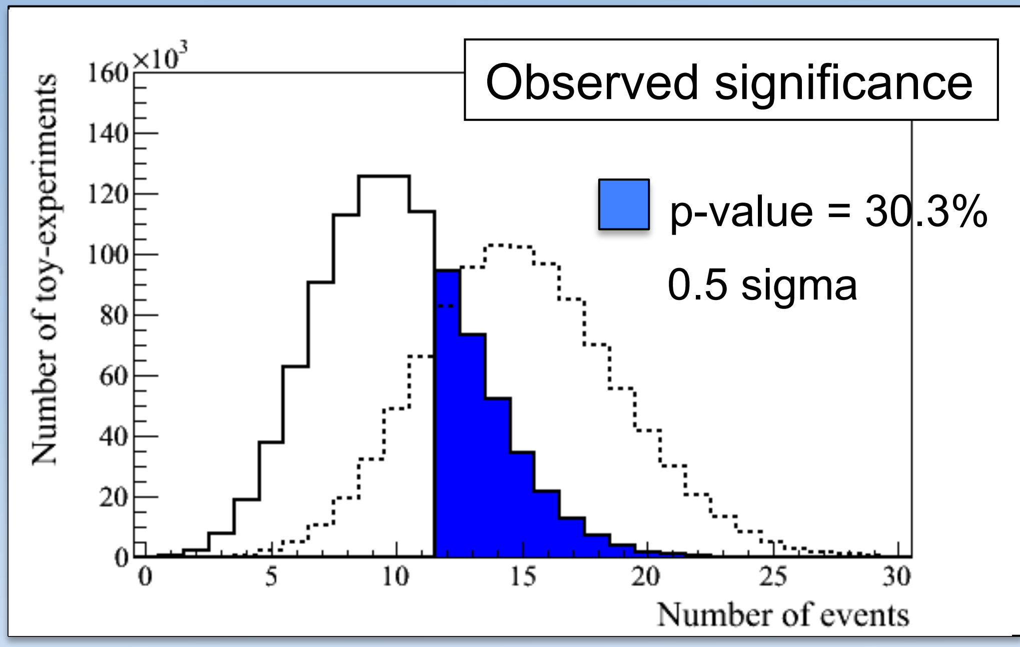
Discovery-aimed: p-value and significance

incompatibiliy with SM-only hypothesis

SM	10
Higgs	5
Data	12



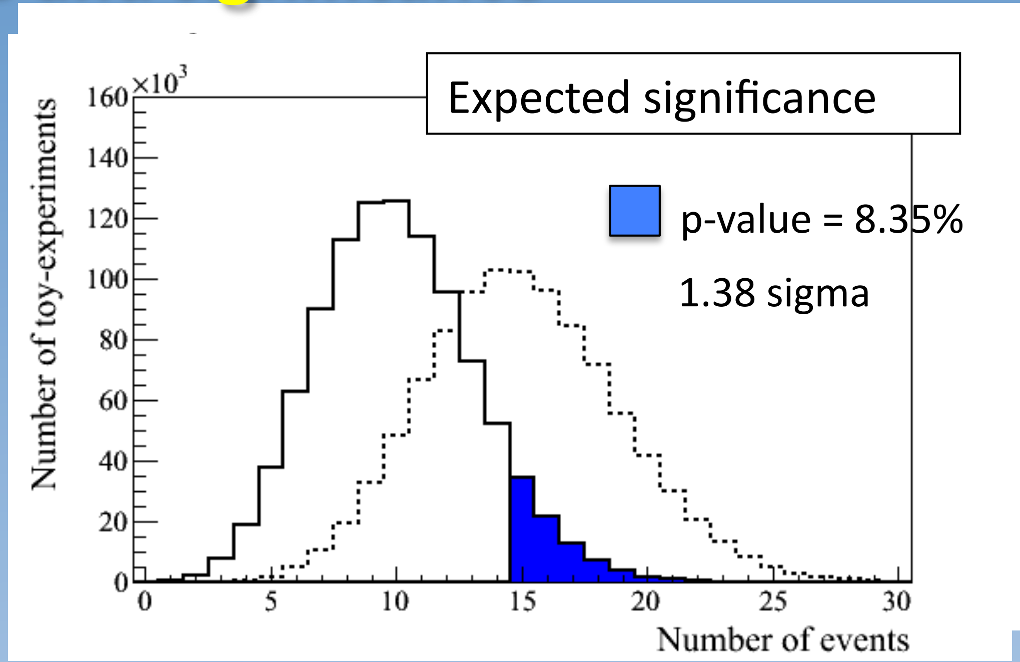
*1) What is the **expected** significance ?*



*2) What is the **observed** significance ?*

Discovery-aimed: p-value and significance

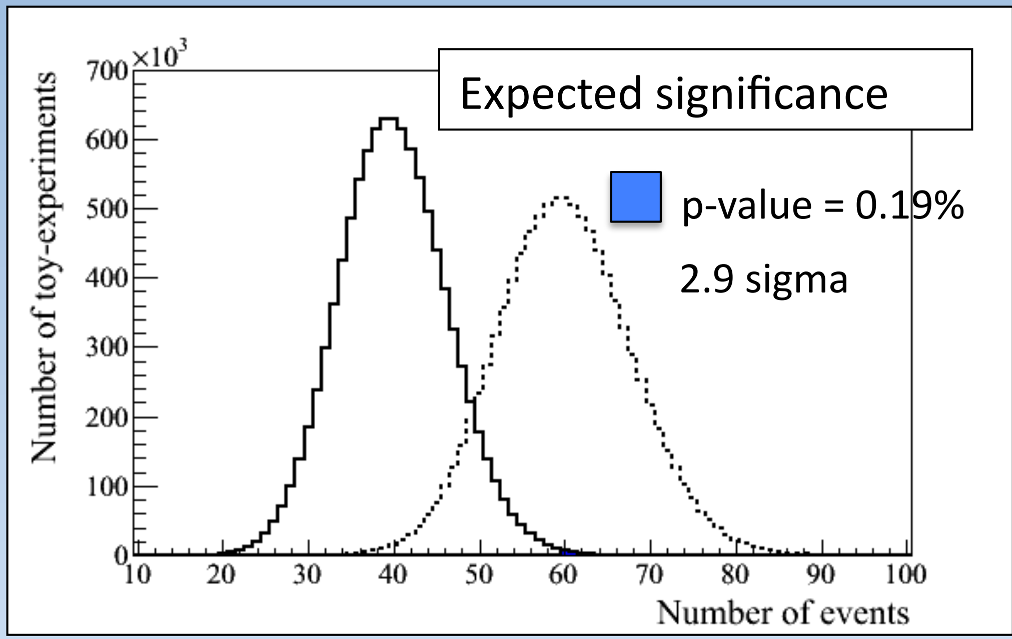
SM	10
Higgs	5



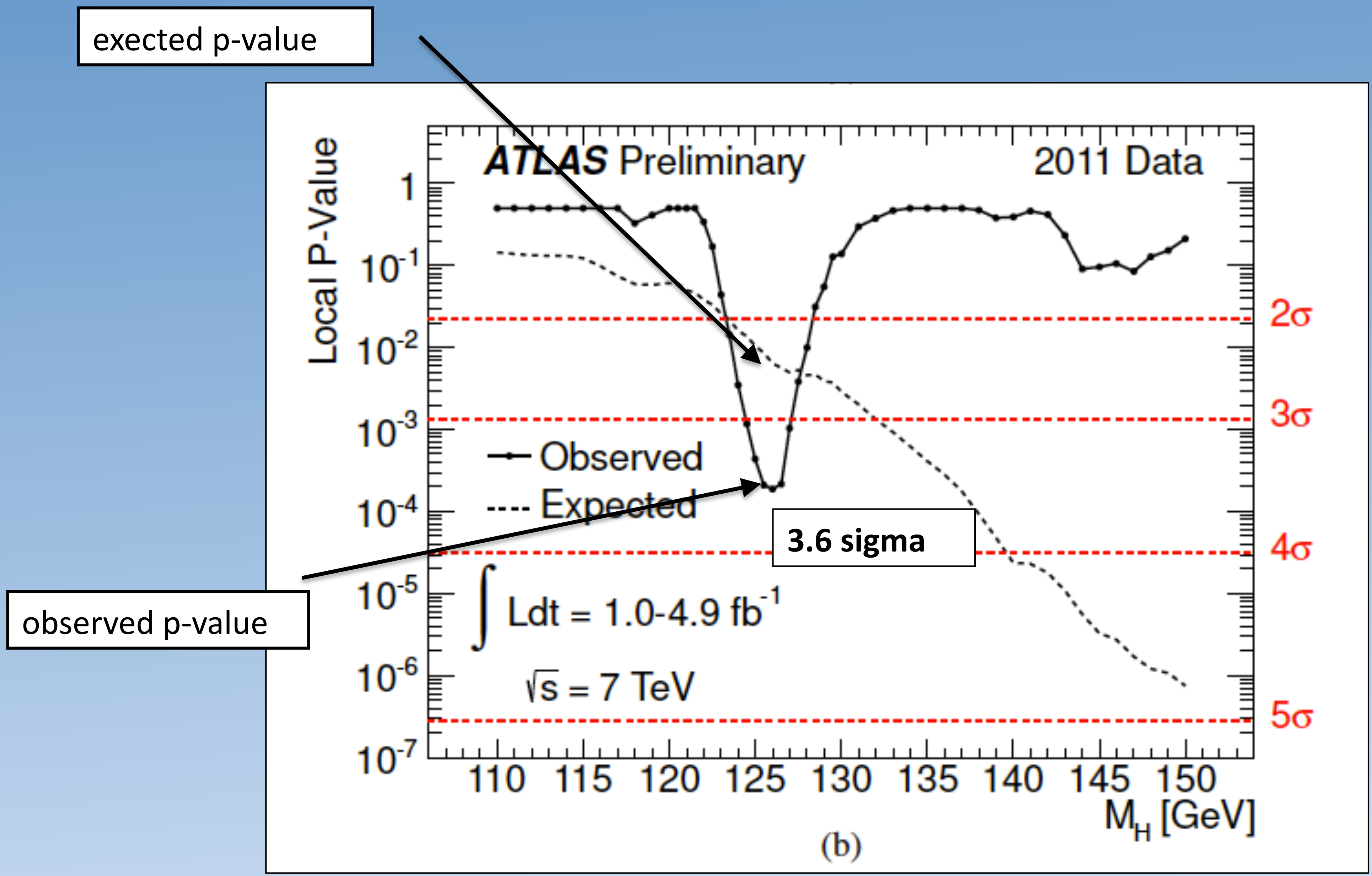
3) At what Lumi do you expect to be able to claim a discovery ?

3 times more LUMINOSITY

SM	30
Higgs	15



Discovery if p-value < 2.87×10^{-7}



exected p-value
observed p-value
3.6 sigma
ATLAS Preliminary
2011 Data
Local P-Value
1
10-1
10-2
10-3
10-4
10-5
10-6
10-7
Observed
Expected
∫ Ldt = 1.0-4.9 fb-1
√s = 7 TeV
2σ
3σ
4σ
5σ
110 115 120 125 130 135 140 145 150
MH [GeV]
(b)

Interpretation

pessimistic: exclusion

When / how do you exclude a signal

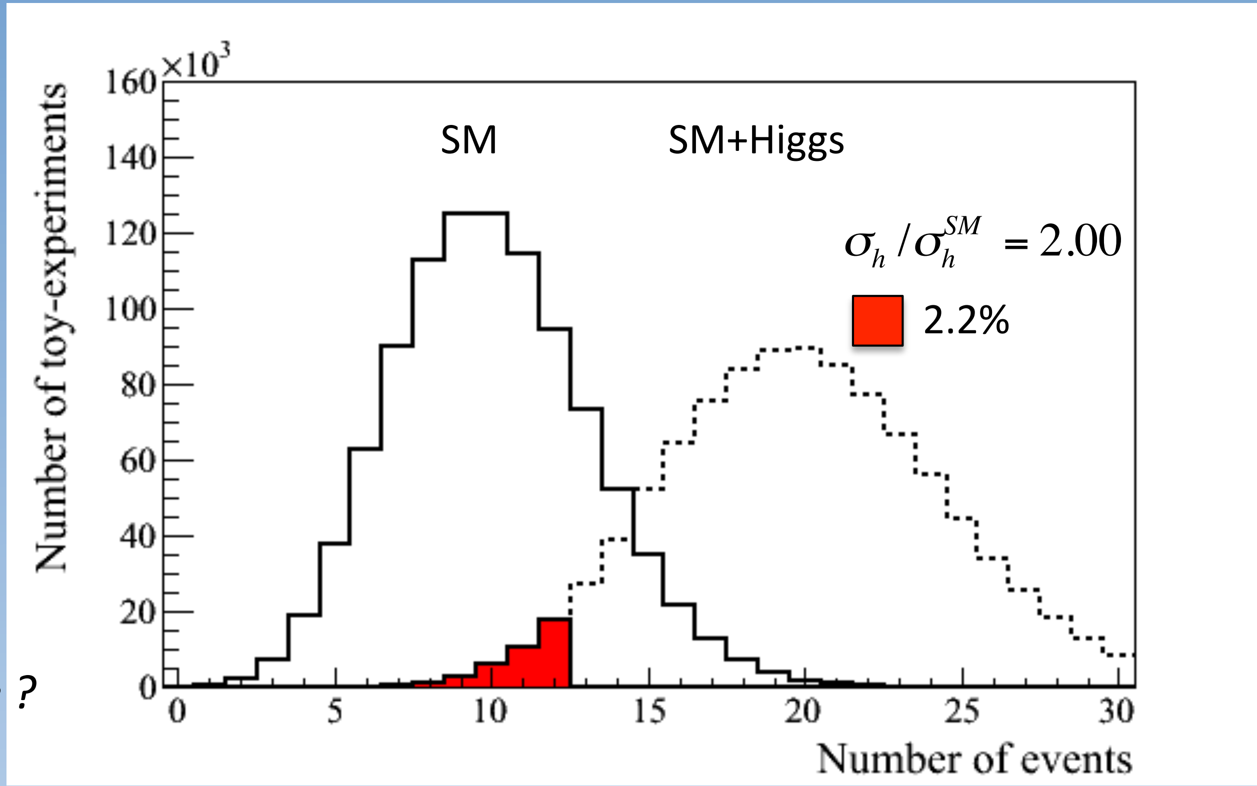
Incompatibility with s+b hypothesis

Standard Model

SM	10
Higgs	5
Data	12

Can we exclude the SM+Higgs hypothesis ?

What σ_h/σ_h^{SM} can we exclude ?



Exclusion: probability to observe N events (or even less) under the signal + background hypothesis

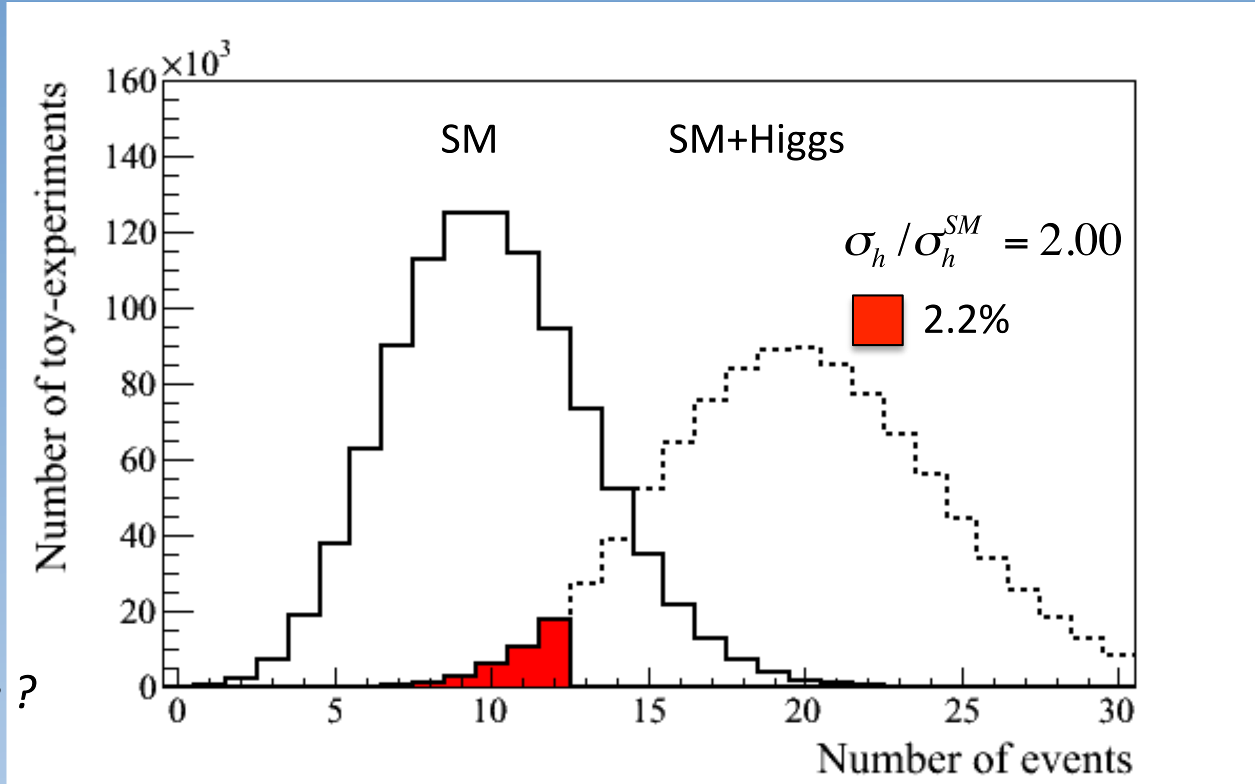
When / how do you exclude a signal

Standard Model

SM	10
Higgs	5
Data	12

Can we exclude the SM+Higgs hypothesis ?

What σ_h/σ_h^{SM} can we exclude ?



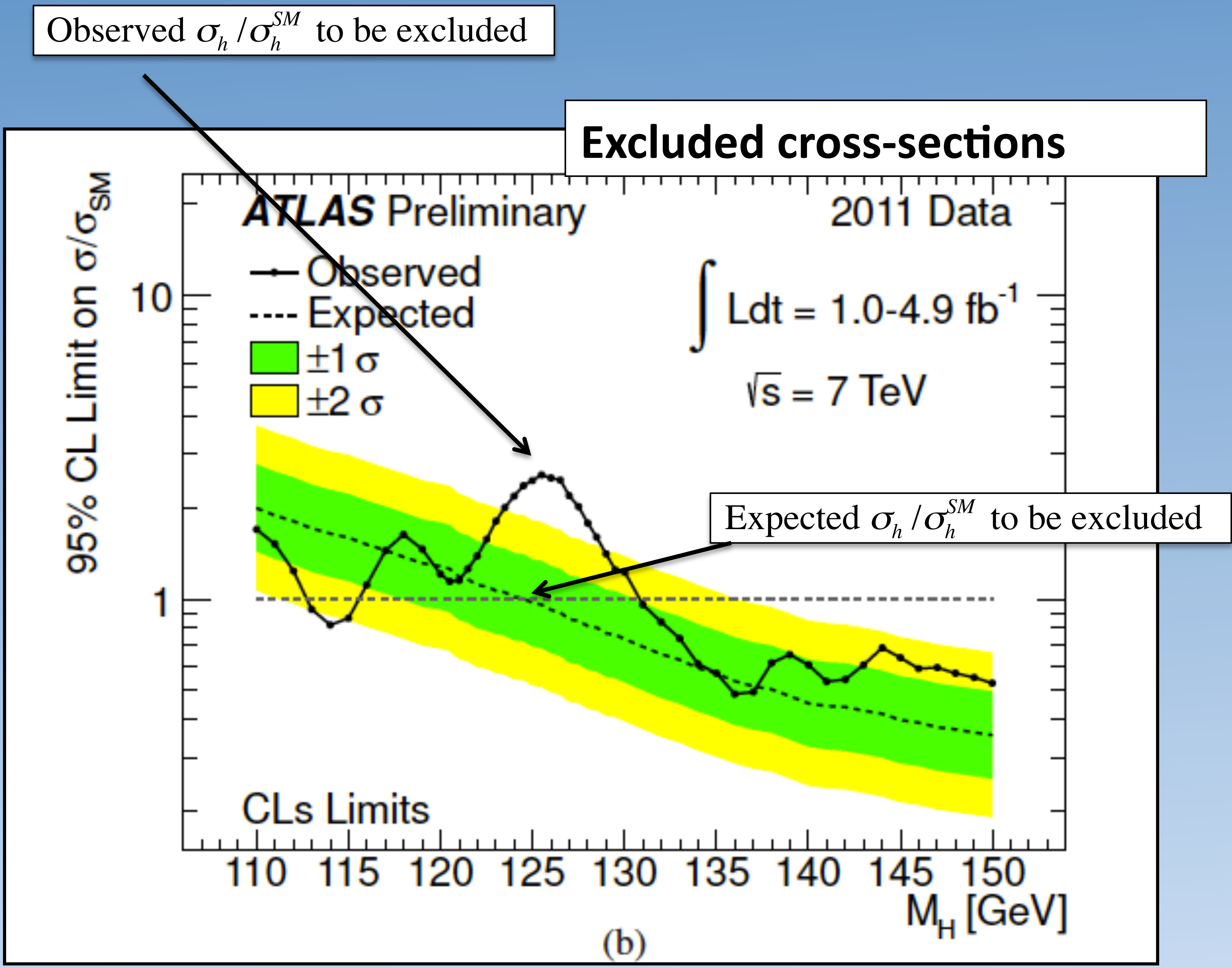
σ/σ_{SM}	SM	# data	SM+Higgs	
1.0	10	12	15.0	18.5 %
1.5	10	12	17.5	6.8%
2.0	10	12	20.0	2.2%

excluded

Expected exclusion ? Use mean SM instead of Ndata

Observed excluded cross-section, σ_h/σ_h^{SM} , = 1.64

Excluded cross-sections



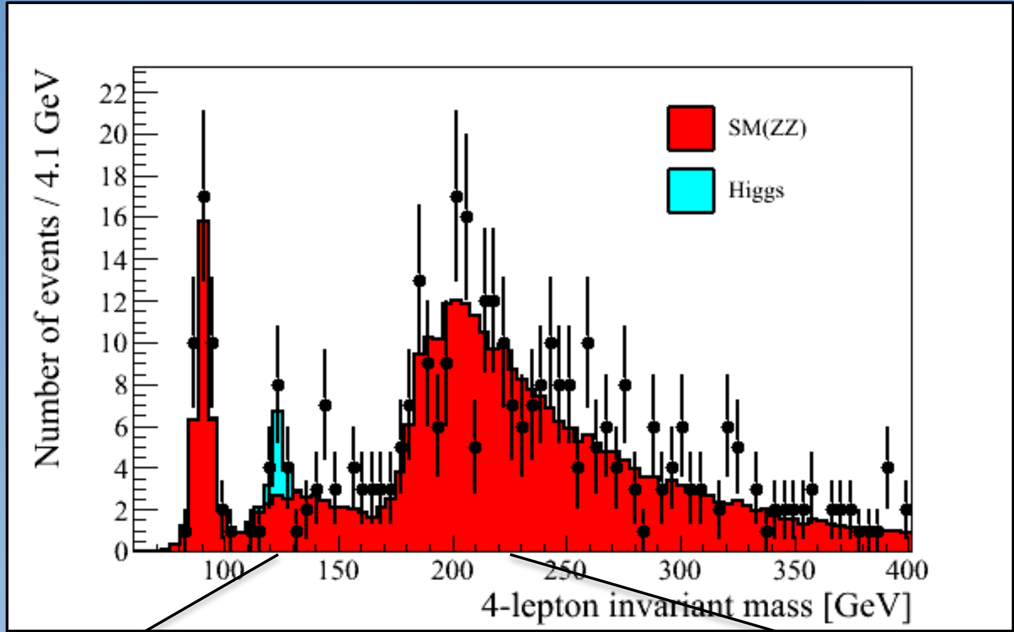
We will try to reproduce a few of these numbers

Create the 4-lepton mass plot

```
root> .L Hasco_skeleton.C++
root> MassPlot(20)
```

↓ Rebin-factor

hist: h_bgr, h_sig, h_data

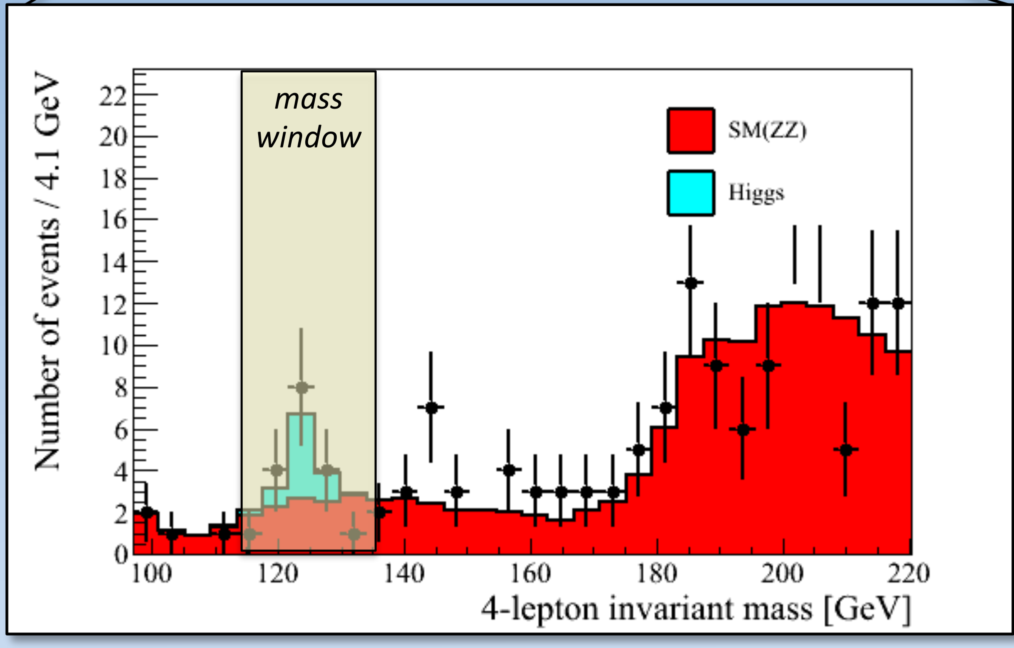


Summary in signal mass region

(using 200 MeV bin and 10 GeV window)

Ndata = 16
Nbgr = 6.42
Nsig = 5.96

Exercises: significance and exclusion

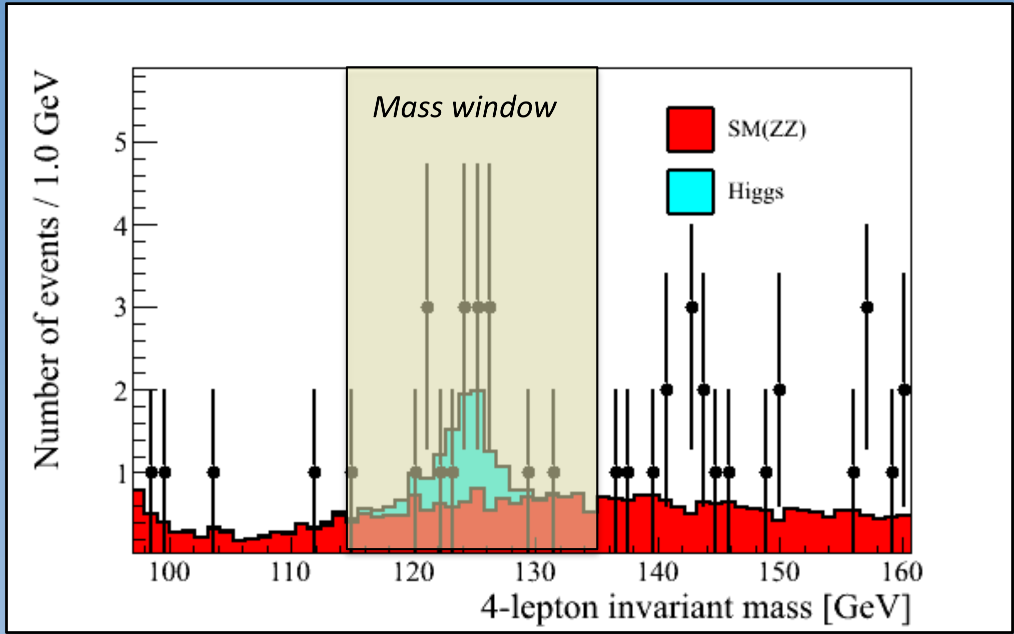


Exercise 3

Optimizing the counting experiment

Code you could use:

```
IntegratePoissonFromRight()
Significance_Optimization()
```



Exercise 1: significance optimization of mass/search window (use Poisson counting)

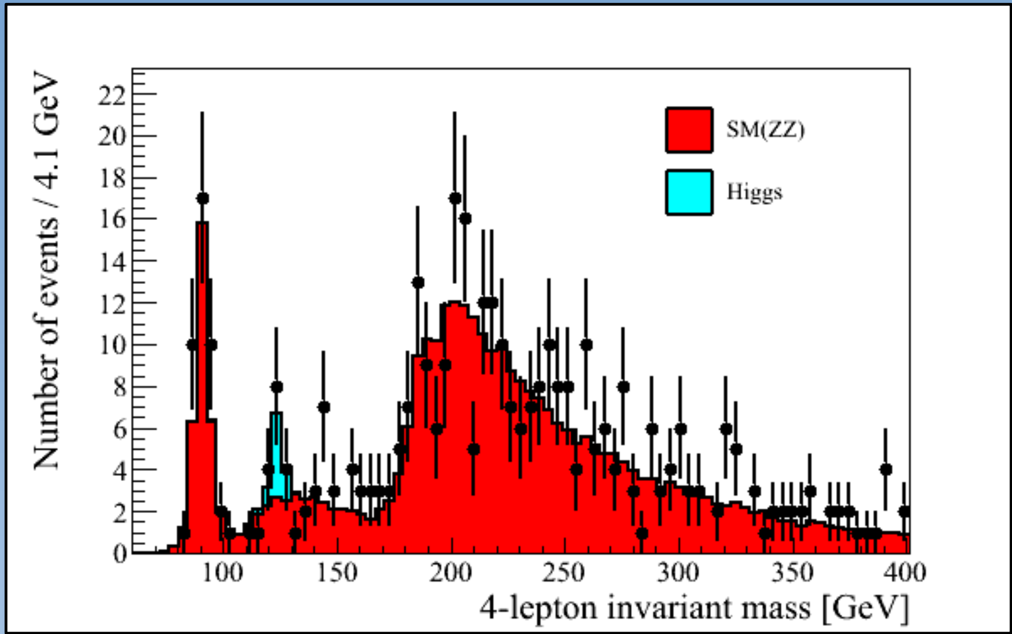
3.1 Find the window that optimizes the expected significance

3.2 Find the window that optimizes the observed significance (and never do it again)

3.3 Find the window that optimizes the expected significance for 5x higher luminosity

3.4 At what luminosity do you expect to be able to make a discovery ?

Beyond simple counting: profile likelihood ratio test-statistic



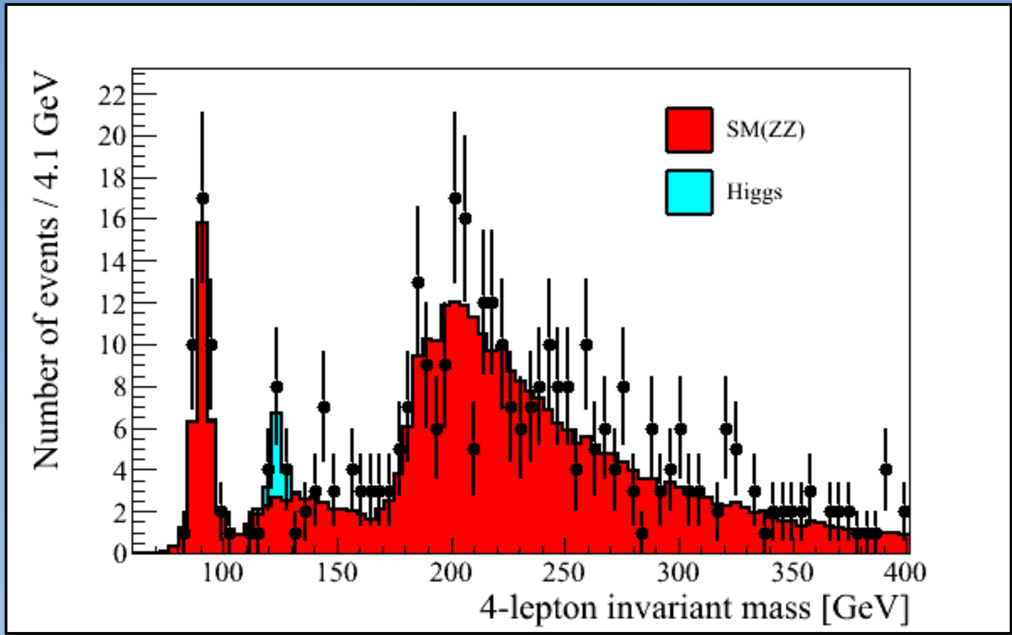
Condense data in **one** number: **X**

LHC experiments:

$$X(\mu) = -2\ln(Q(\mu)), \quad \text{with} \quad Q(\mu) = \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})}$$

We'll use something a bit simpler, but same idea

Beyond simple counnting: likelihood ratio test-statistic



$$-2\cdot \log\left(Likelihood\right) = -2\cdot \sum_{bins} \log\left(Poisson(N_{bin}^{data} \mid \mu\cdot f_{bin}^{Higgs} + \alpha\cdot f_{bin}^{SM})\right)$$

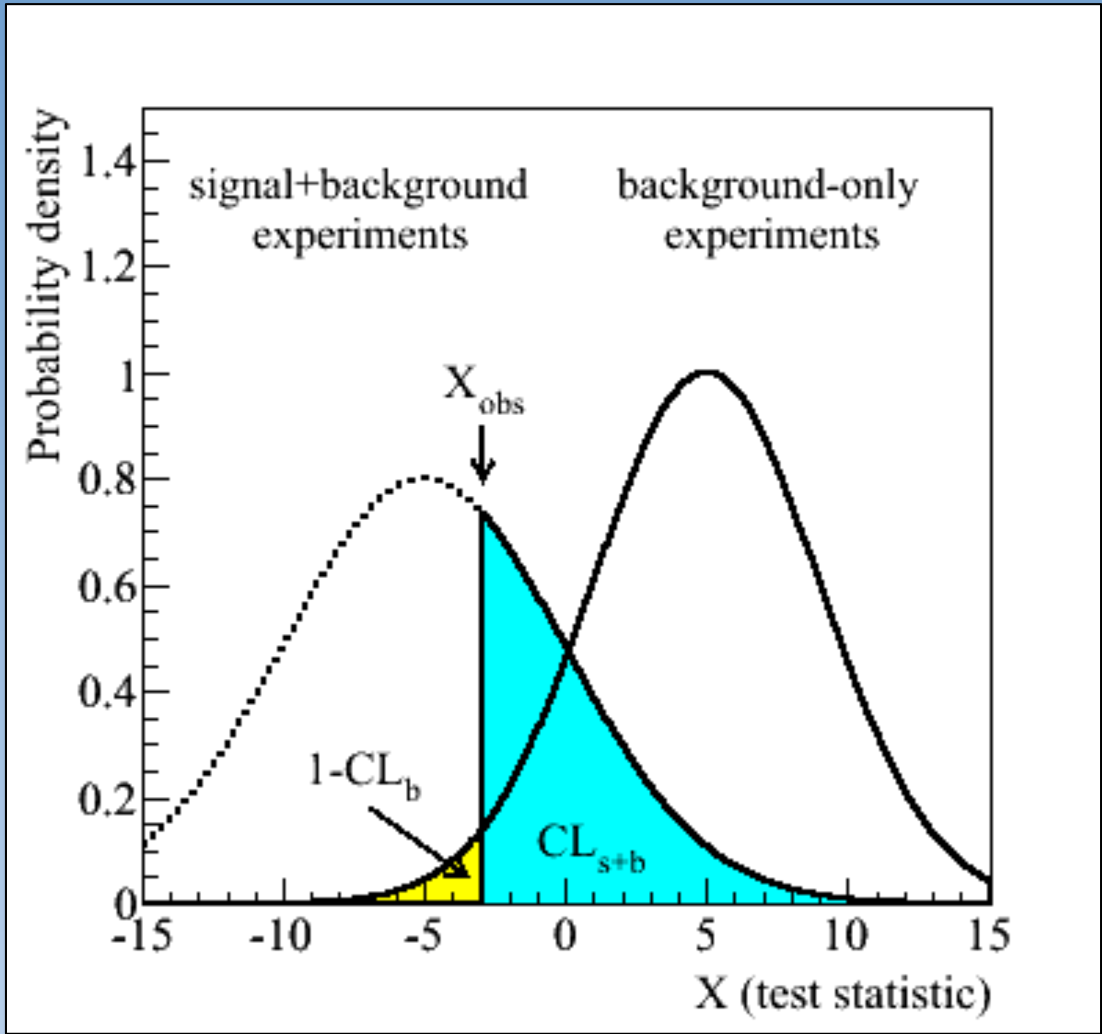
$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$

Likelihood assuming μ_s=1 (signal+background)

Hypothesis 1

Hypothesis 0

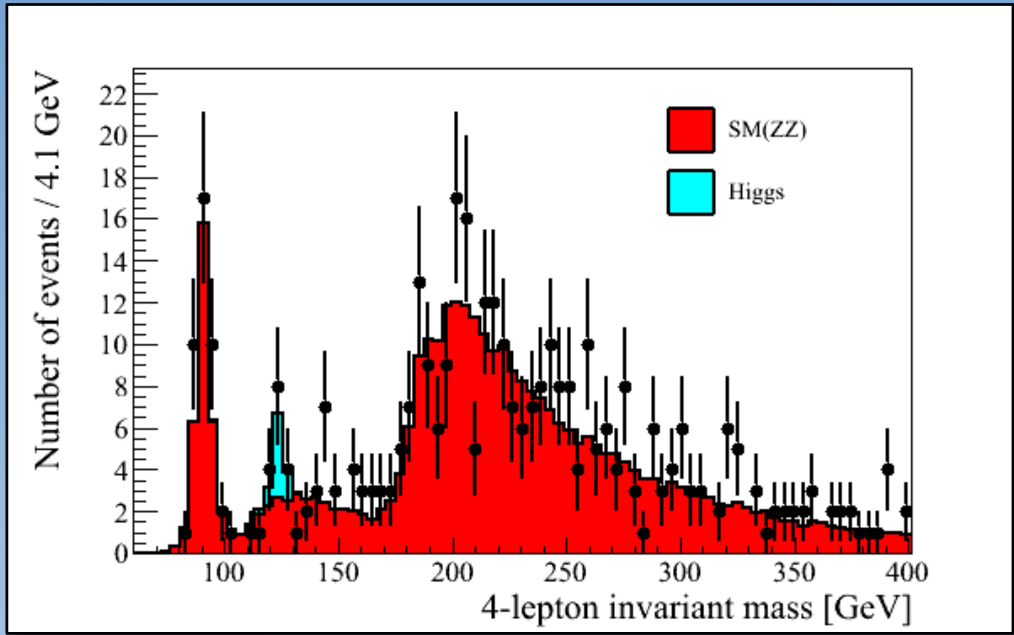
Likelihood assuming μ_s=0 (only background) signal like



background like

Question: does the window not matter ?

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$



X = log(a/b) = log(A) – log(B)

What happenes to if you add a bin at 300 GeV ?
Will I not dilute the channel like in counting ?

In that bin Lik_{bin} = Constant = C

X = log(a/b) = [log(A) + log(C)] – [log(B) + log (C)]
= log(A) – log(B)

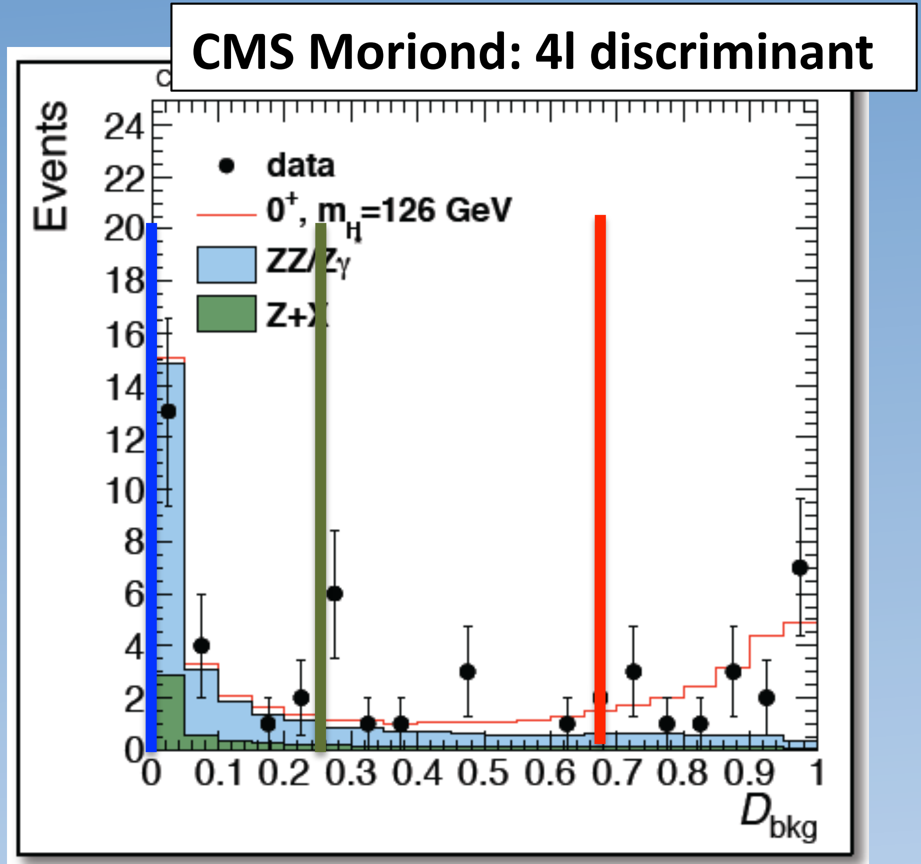
ANY discrimination info is good !

Question: what about more info than mass alone ?

1) Optimal for counting

2) Optimal for LR test stat.

3) Normal procedure



Why: because the ‘information’ you add below D< 0.25 is maybe difficult to verify in terms of correctness: needs signal descripption in very background-like region: systematics. Need to find optimum.

Note: they still evaluate, like you: $X = -2\ln(Q)$, with $Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$

We will use a very simple form for the test statistic

Our exercise (α=1 or from Ex.3):

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)} = $$  / 

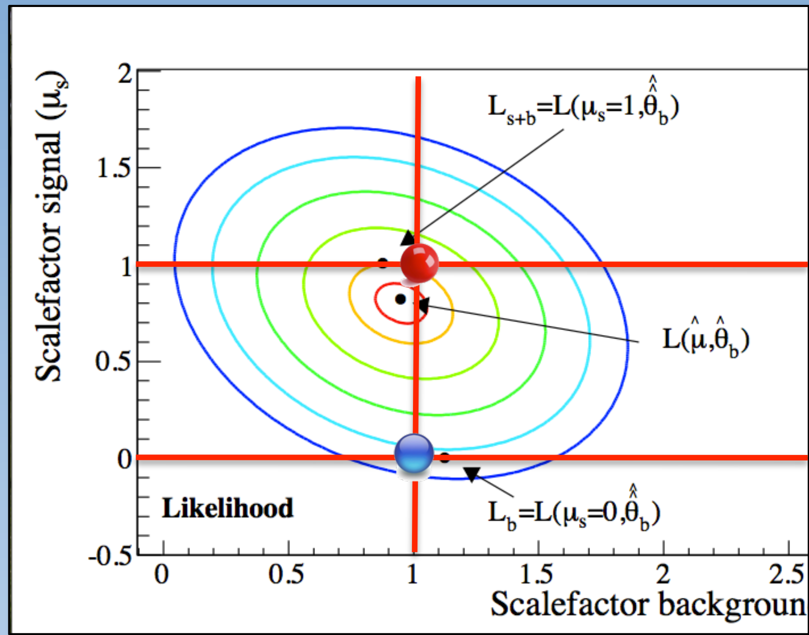
Tevatron-style:

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1, \hat{\hat{\theta}}_{(\mu_s = 1)})}{L(\mu_s = 0, \hat{\hat{\theta}}_{(\mu_s = 0)})}$$

LHC experiments:

$$X(\mu) = -2\ln(Q(\mu)), \quad \text{with} \quad Q(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

2-dimensional fit (α and μ free)



Note:

α_{bgr} is just one of the nuissance parameters θ in a ‘real’ analysis

Exercise 4

compute test-statistic X

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$

→ Likelihood assuming μ_s=1 (signal+background)

→ Likelihood assuming μ_s=0 (only background)

Exercise 4: create the likelihood ration test statistic – beyond simple counting

4.1 Write a routine that computes the likelihood ratio test-statistic for a given data-set

```
double Get_TestStatistic(TH1D *h_mass_dataset, TH1D *h_template_bgr, TH1D *h_template_sig)
```

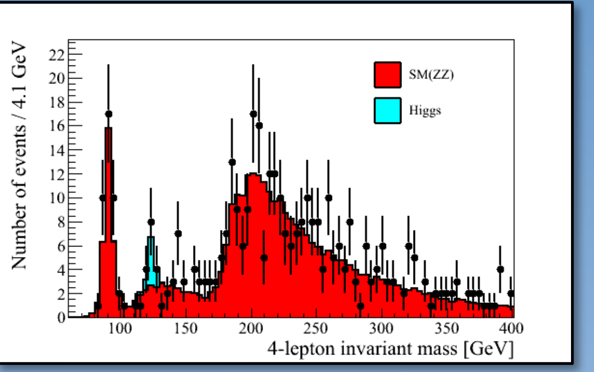
$$-2Log\left(Likelihood_{(\mu,\alpha\,=\,1)}\right) = -2\cdot \sum_{bins} \log\left(Poisson(N_{bin}^{data} \mid \mu\cdot f_{bin}^{Higgs} + \alpha\cdot f_{bin}^{SM})\right)$$

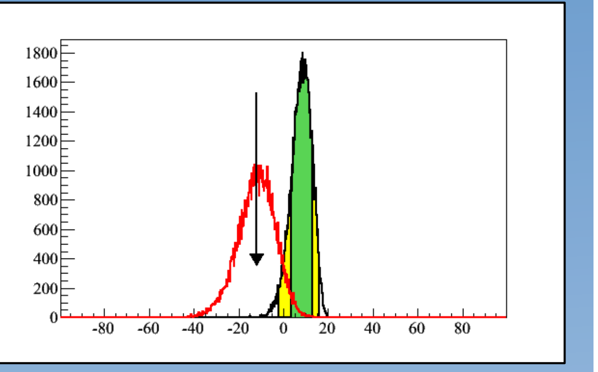
Note: log(a/b) = log(a) – log(b)

4.2 Compute the likelihood ratio test-statistic for the ‘real’ data

Exercise 5

Toy data-sets





Exercise 5: create toy data-sets

5.1 Write a routine that generates a toy data-set from a MC template (b or s+b)

TH1D * GenerateToyDataSet(TH1D *h_mass_template)

How: Take the histogram h_mass_template and draw a Poisson random number in each bin using the bin content in h_mass_template as the central value. Return the new fake data-set.

5.2 Generate 1000 toy data-sets for *background-only* & compute test statistic
Generate 1000 toy data-sets for *signal+background* & compute test statistic

→ plot both in one plot

5.3 Add the test-statistic from the data(exercise 4.2) to the plot

Exercise 6

Summarize separation power

Exercise 6: compute p-value

6.1 Compute the p-value or $1\text{-}CL_b$ (under the background-only hypothesis):

- For the average(median) b-only experiment
- For the average(median) s+b-only experiment [expected significance]
- For the data [observed significance]

6.2 Draw conclusions:

- Can you claim a discovery ?
- Did you expect to make a discovery ?
- At what luminosity did/do you expect to be able to make a discovery ?

Exercise 7:

Exclude a cross-section for a given Higgs boson mass

$$\sigma_h(m_h) = \zeta \cdot \sigma_h^{SM}(m_h)$$

Scale factor wrt SM prediction

Some shortcomings, but we'll use it anyway

Exercise 7: compute CL_{s+b} and exclude Higgs masses or cross-sections

7.1 Compute the CL_{s+b}:

- For the average(median) s+b experiment
- For the average(median) b-only experiment
- For the data

7.2 Draw conclusions:

- Can you exclude the m_h=200 GeV hypothesis ? What ς can you exclude ?
- Did you expect to be able to exclude the m_h=200 GeV hypothesis ? What ς did you expect to be able to exclude ?

Basic material for the Root examples:

1) Get the data-set and example code: **HascoRootStatisticsCode.tgz**

2) Unpack everything: **tar –vzxf HascoRootStatisticsCode.tgz**

1) Directory /RootExamples/

a) Example0*C (* = 0,1,2,3,4,5) - All examples *.C-files used in this presentation
b) Code for Ntuple production and reading
c) rootlogon.C (some standard Root settings)

2) Directory /Exercises/

a) Histograms_fake.root
Contains 4 lepton invariant mass histogram distributions (H125, H200, ZZ, data)
b) Hasco_skeleton.C
Some skeleton code (different levels, as minimal as possible). Your code !
c) rootlogon.C (Some standard Root settings)


Exercises


Good luck