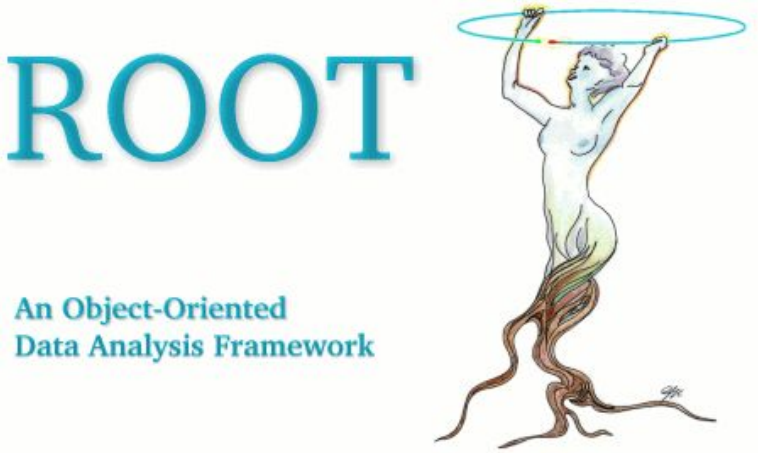
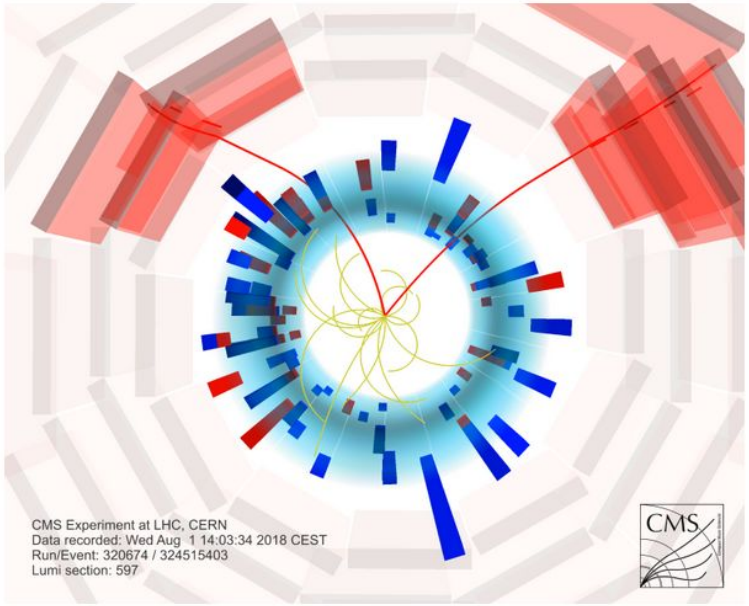
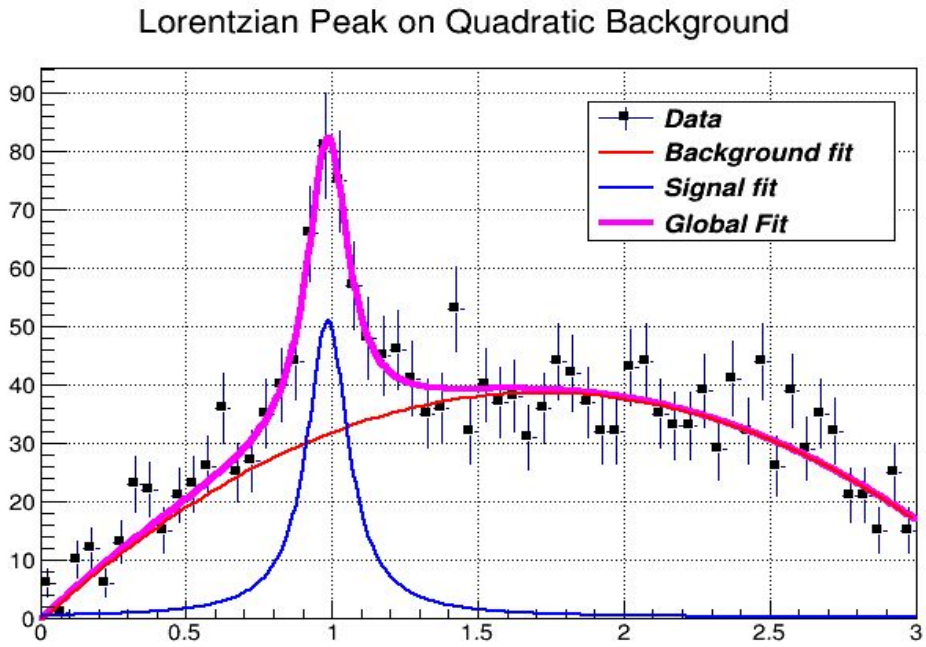






Big data en el CERN y otros contextos

Jhovanny Andres Mejia Guisao
UNIVERSIDAD DE ANTIOQUIA, COLOMBIA

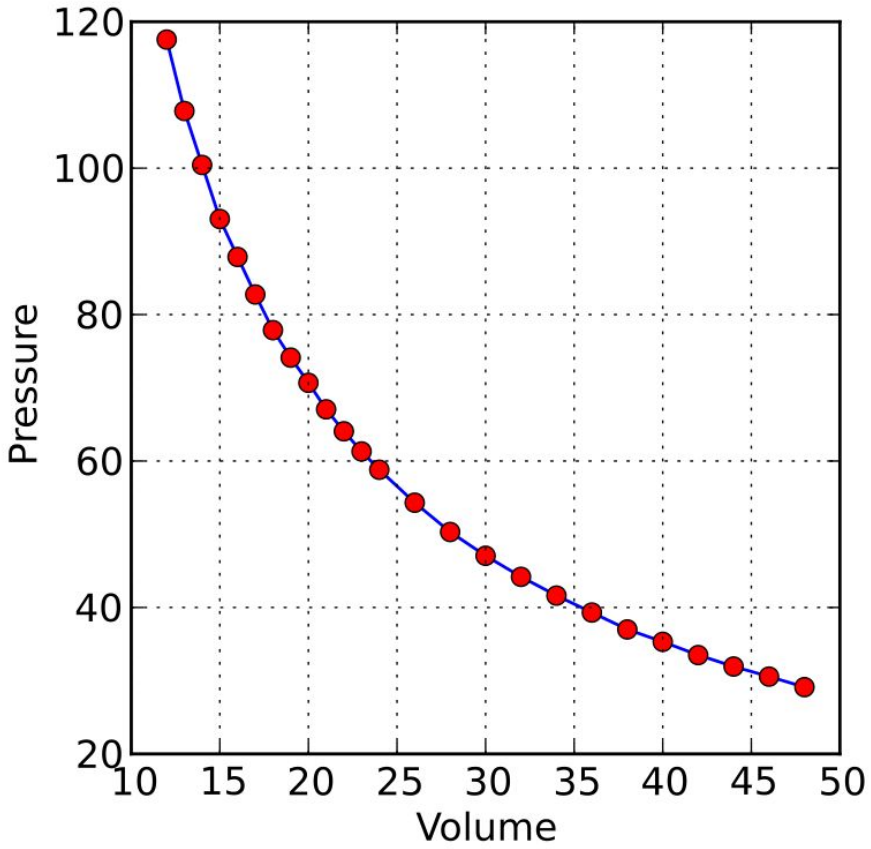
Motivation

What we hope to discuss about scientific data analysis?

- Advanced graphical user interface
- Interpreter for the C++ programming language
- Persistency mechanism for C++ objects
- Used to write every year petabytes of data recorded by the Large Hadron Collider experiments

Input and plotting of data from measurements and fitting of analytical functions.

Motivation

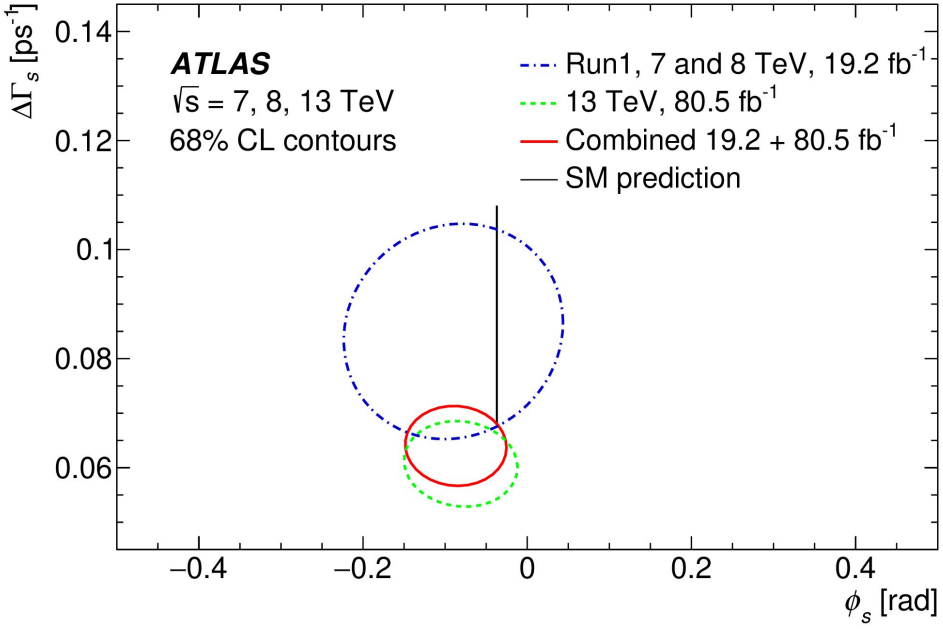


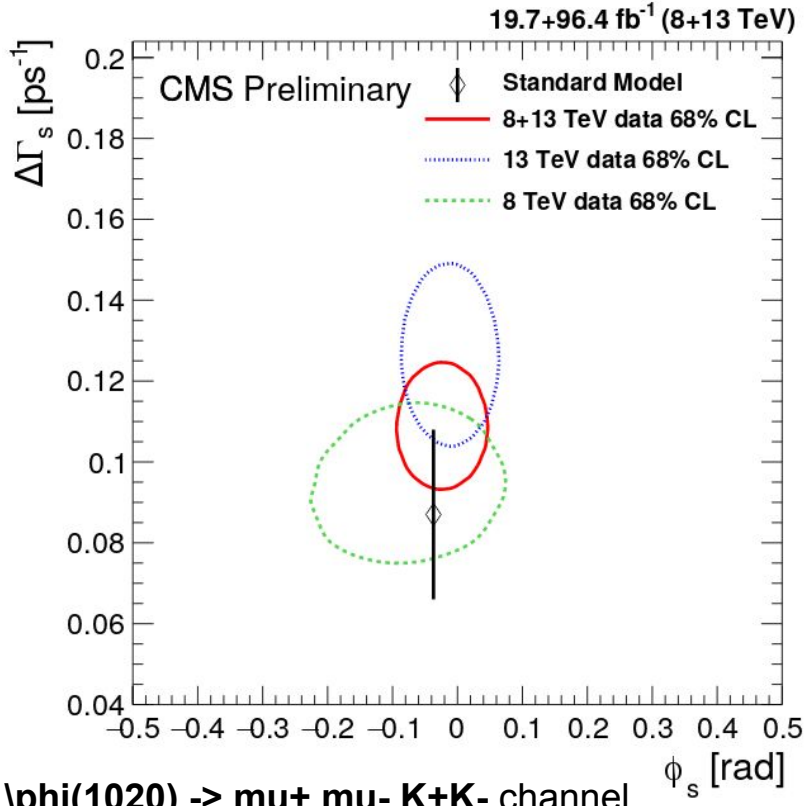
Para un gas a temperatura constante, el volumen es inversamente proporcional a la presión sobre éste[link]

Se puede explicar matemáticamente con:
pV = k

y la tarea del experimentador consiste en determinar la constante, k, a partir de un conjunto de medidas.

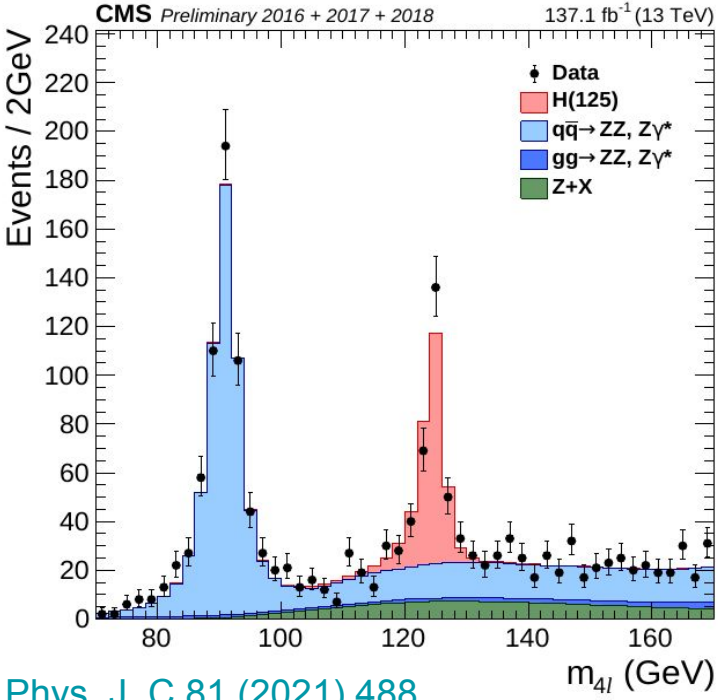
Motivation

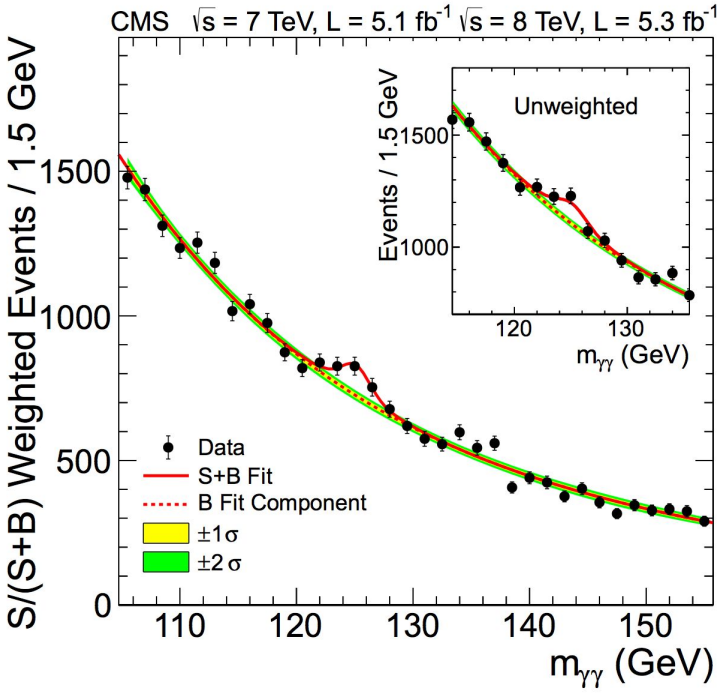




Measurement of the **CP violating phase** in the **Bs-> J/\psi \phi(1020) -> mu+ mu- K+K-** channel in proton-proton collisions at √s = 13 TeV (PLB 816 (2021) 136188)

Motivation



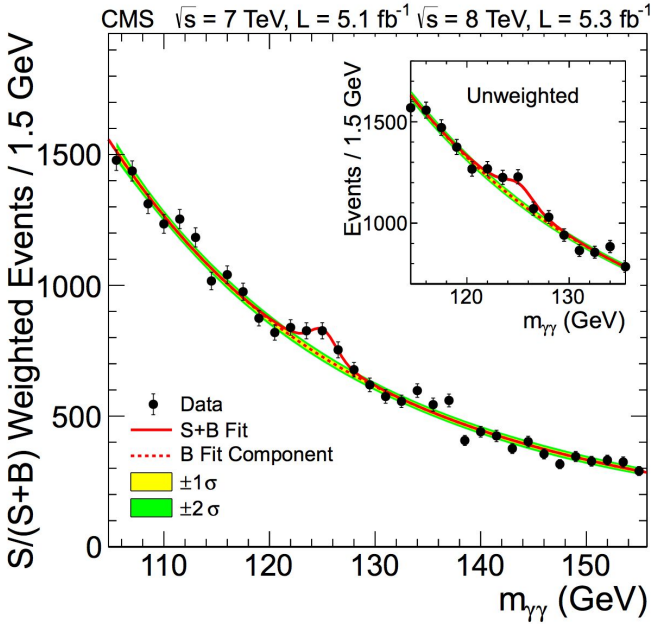


Eur. Phys. J. C 81 (2021) 488

Direct (left) and indirect (right) evidence for the Higgs boson. Left: the invariant mass of four leptons, showing evidence for the Higgs boson at Mh~ 126 GeV. Right: the world average of the W boson mass vs the top-quark mass is shown. Overlayed are the contour plots when the direct Higgs boson mass measurement is used. It is evident that the indirect determination of the Higgs boson mass is quite consistent with the direct measurement.

Motivation

In Quantum mechanics, models typically only predict the **probability density function ("pdf")** of measurements depending on a number of parameters, and the aim of the experimental analysis is to extract the parameters from the **observed distribution** of frequencies at which certain values of the measurement are observed. Measurements of this kind require means to generate and visualize frequency distributions, **so-called histograms,** and **stringent statistical treatment to extract the model parameters from purely statistical distributions.**



Motivation

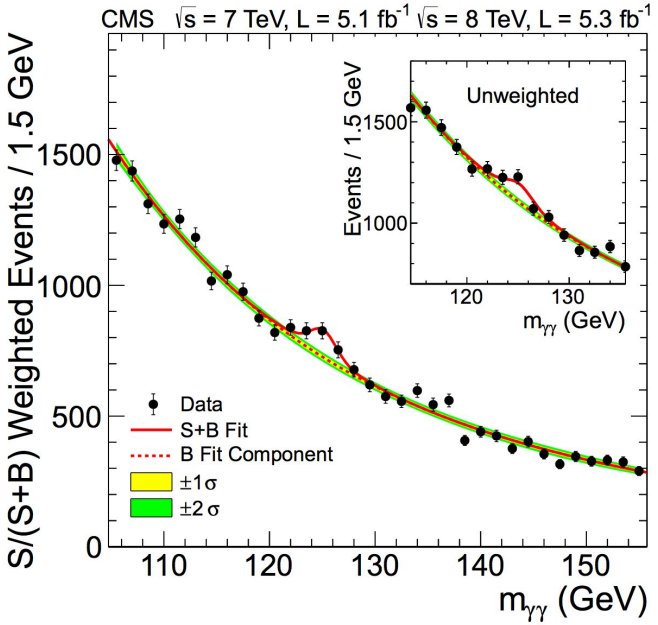
Visualization of the data

Corrections or parameter transformations?

One specialty of experimental physics are the inevitable uncertainties affecting each measurement.

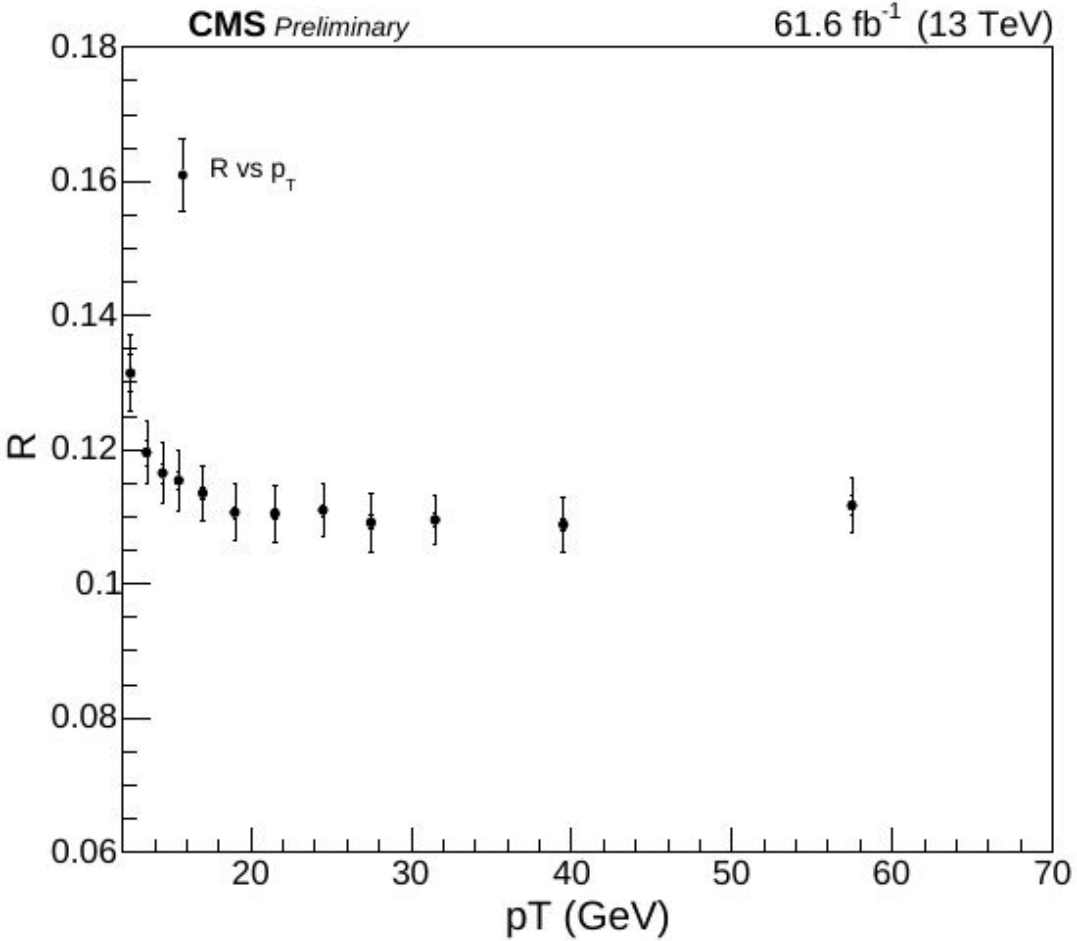
The statistical nature of the errors must be handled properly.

Quite often, the data volume to be analyzed is large - think of fine-granular measurements accumulated with the aid of computers. A usable tool therefore must contain easy-to-use and efficient methods for storing and handling data.



Simulation of expected data is another important aspect in data analysis. By repeated generation of "**pseudo-data**", which are analysed in the same manner as intended for the real data, analysis procedures can be validated or compared.

Motivation



what does this distribution tell us?

why there are two errors? or more?

how can we understand/handle these errors?

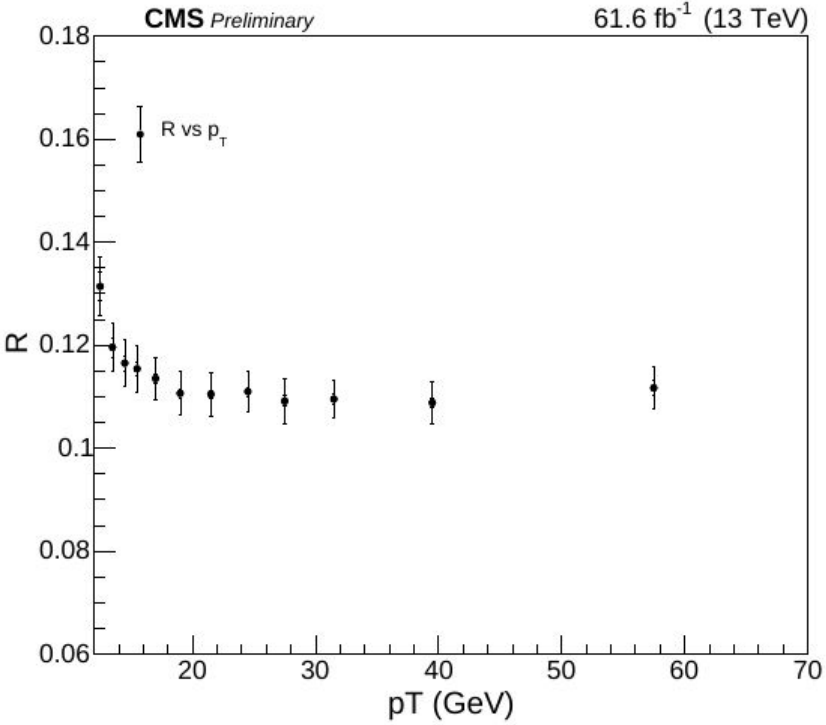
BPH-21-001

Motivation

Table 11: $\mathcal{R}$ and uncertainty (σ_{tot}), including the statistical (σ_{stat}) uncertainty and the fully correlated and uncorrelated systematic uncertainties among the samples (σ_{sys}^{uncor}, σ_{sys}^{cor}). Correlations stem from the common tracking and fit model uncertainties.

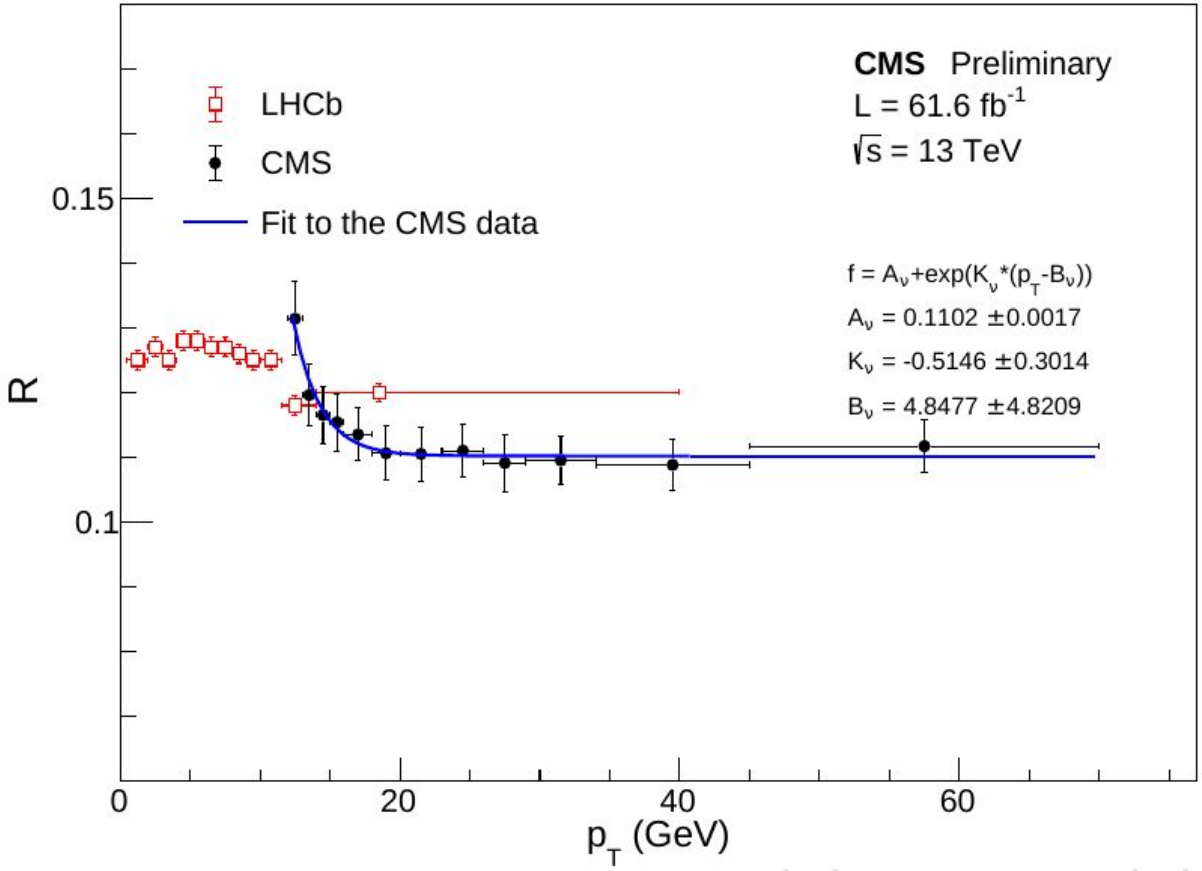
p_T (GeV)	$\mathcal{R}$	σ_{tot}	σ_{stat}	σ_{sys}^{uncor}	σ_{sys}^{cor}
12 – 13	0.1314	0.0058	0.0028	0.0033	0.0038
13 – 14	0.1196	0.0046	0.0019	0.0015	0.0039
14 – 15	0.1165	0.0041	0.0015	0.0016	0.0035
15 – 16	0.1154	0.0042	0.0014	0.0018	0.0036
16 – 18	0.1135	0.0040	0.0009	0.0022	0.0032
18 – 20	0.1106	0.0041	0.0009	0.0022	0.0033
20 – 23	0.1105	0.0042	0.0008	0.0024	0.0034
23 – 26	0.1110	0.0040	0.0009	0.0023	0.0031
26 – 29	0.1091	0.0044	0.0010	0.0020	0.0038
29 – 34	0.1095	0.0037	0.0010	0.0022	0.0028
34 – 45	0.1088	0.0040	0.0009	0.0023	0.0032
45 – 70	0.1117	0.0041	0.0014	0.0021	0.0033

$$\begin{pmatrix}
3.214 & 0.663 & 0.545 & 0.571 & 0.450 & 0.513 & 0.520 & 0.421 & 0.677 & 0.313 & 0.464 & 0.473 \\
0.663 & 2.242 & 0.645 & 0.676 & 0.533 & 0.608 & 0.615 & 0.498 & 0.801 & 0.370 & 0.549 & 0.560 \\
0.545 & 0.645 & 2.005 & 0.556 & 0.438 & 0.499 & 0.506 & 0.410 & 0.658 & 0.304 & 0.451 & 0.460 \\
0.571 & 0.676 & 0.556 & 1.973 & 0.459 & 0.523 & 0.530 & 0.429 & 0.690 & 0.319 & 0.473 & 0.482 \\
0.450 & 0.533 & 0.438 & 0.459 & 1.721 & 0.413 & 0.418 & 0.338 & 0.544 & 0.251 & 0.373 & 0.380 \\
0.513 & 0.608 & 0.499 & 0.523 & 0.413 & 1.766 & 0.476 & 0.386 & 0.620 & 0.287 & 0.425 & 0.433 \\
0.520 & 0.615 & 0.506 & 0.530 & 0.418 & 0.476 & 1.781 & 0.391 & 0.627 & 0.290 & 0.430 & 0.439 \\
0.421 & 0.498 & 0.410 & 0.429 & 0.338 & 0.386 & 0.391 & 1.585 & 0.508 & 0.235 & 0.349 & 0.355 \\
0.677 & 0.801 & 0.658 & 0.690 & 0.544 & 0.620 & 0.627 & 0.508 & 1.966 & 0.378 & 0.560 & 0.571 \\
0.313 & 0.370 & 0.304 & 0.319 & 0.251 & 0.287 & 0.290 & 0.235 & 0.378 & 1.380 & 0.259 & 0.264 \\
0.464 & 0.549 & 0.451 & 0.473 & 0.373 & 0.425 & 0.430 & 0.349 & 0.560 & 0.259 & 1.611 & 0.392 \\
0.473 & 0.560 & 0.460 & 0.482 & 0.380 & 0.433 & 0.439 & 0.355 & 0.571 & 0.264 & 0.392 & 1.691
\end{pmatrix} \cdot 10^{-05}$$



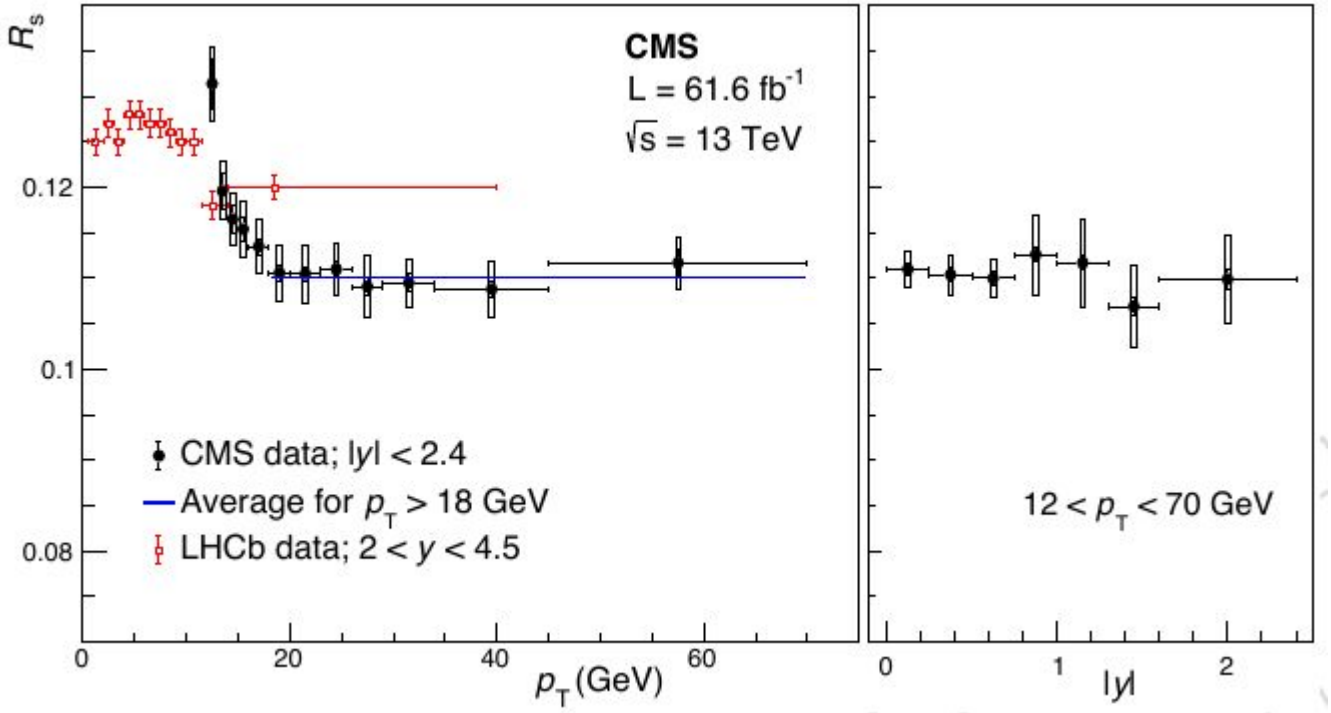
BPH-21-001

Motivation



[BPH-21-001]

Motivation



BPH-21-001

ROOT es muy flexible y proporciona una interfaz de programación para usar en aplicaciones propias y una interfaz gráfica de usuario para el análisis interactivo de datos.

Descripción general y justificación del curso:

Este curso pretende aportar a la formación de los estudiantes de los programas de pregrado de física y astronomía, aportando a sus conocimientos las técnicas y métodos computacionales necesarios para un mejor entendimiento de los procesos involucrados en el tratamiento de los datos procesados en el CERN en colisiones de partículas. Dada la magnitud de los datos tomados, se han desarrollado técnicas para el manejo de cantidades masivas de información. Se pretende que los estudiantes adquieran habilidades computacionales que le permitan desenvolverse en este medio, ya que es un requisito esencial para poder participar de proyectos relacionados con el área de física de partículas experimental. Adicional a esa idea, sin embargo, es importante decir que también de esta forma se garantiza que el estudiante tenga una formación integral tanto en las bases conceptuales aprendidas en la carrera como en las habilidades computacionales que le serán de gran utilidad enfrentando problemas tanto de tipo académico asi como en el ámbito de la industria en el manejo de grandes cantidades de datos.

Objetivo general:
Adquirir conocimientos básicos y experiencia en el uso del marco de análisis de datos usando el software ROOT del CERN y adquirir habilidades para manejo de grandes volúmenes de datos con el mismo.

la clase de hoy, tomada directamente de:
https://root.cern/get_started/courses/

ROOT in a Nutshell

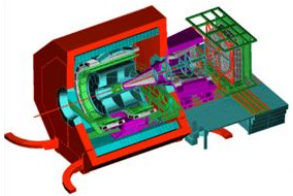
- ROOT is a software framework with building blocks for:
  - Data processing
  - Data analysis
  - Data visualisation
  - Data storage
- ROOT is written mainly in C++ (C++11/17 standard)
  - Bindings for Python available as well
- Adopted in High Energy Physics and other sciences (but also industry)
  - 1 EB of data in ROOT format
  - Fits and parameters' estimations for discoveries (e.g. the Higgs)
  - Thousands of ROOT plots in scientific publications

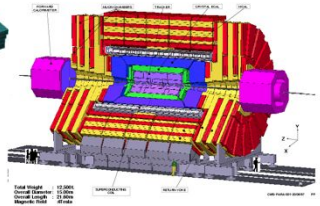
ROOT in a Nutshell

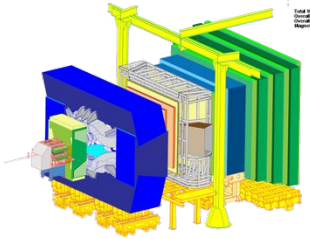
- ROOT can be seen as a collection of building blocks for various activities, like:
  - **Data analysis: histograms, graphs, functions**
  - **I/O: row-wise, column-wise** storage of any C++ object
  - **Statistical tools** (RooFit/RooStats): rich modeling and statistical inference
  - Math: **non trivial functions** (e.g. Erf, Bessel), optimised math functions
  - **C++ interpretation**: full language compliance
  - **Multivariate Analysis** (TMVA): e.g. Boosted decision trees, NN
  - **Advanced graphics** (2D, 3D, event display)
  - **Declarative Analysis**: RDataFrame

ROOT Application Domains

A selection of the experiments adopting ROOT







Event Filtering

Data

Analysis

Offline Processing

Reconstruction

Further processing, skimming

Event Selection, statistical treatment …

Raw

Reco

…

Analysis Formats

Images

Data Storage: Local, Network

Interpreter

- ROOT has a built-in interpreter : CLING
  - C++ interpretation: highly non trivial and not foreseen by the language!
  - One of its kind: Just In Time (JIT) compilation
  - A C++ interactive shell

```
$ root
root[0] 3 * 3
(const int) 9
```

- Can interpret “macros” (non compiled programs)
  - Rapid prototyping possible
- ROOT provides also Python bindings
  - Can use Python interpreter directly after a simple *import ROOT*
  - Possible to “mix” the two languages (????)

Persistency or Input/Output (I/O)

- ROOT offers the possibility to write C++ objects into files
  - This is impossible with C++ alone
  - Used the LHC detectors to write several petabytes per year
- Achieved with serialization of the objects using the reflection capabilities, ultimately provided by the interpreter
  - Raw and column-wise streaming
- As simple as this for ROOT objects: one method - *TObject::Write*

Cornerstone for storage of experimental data

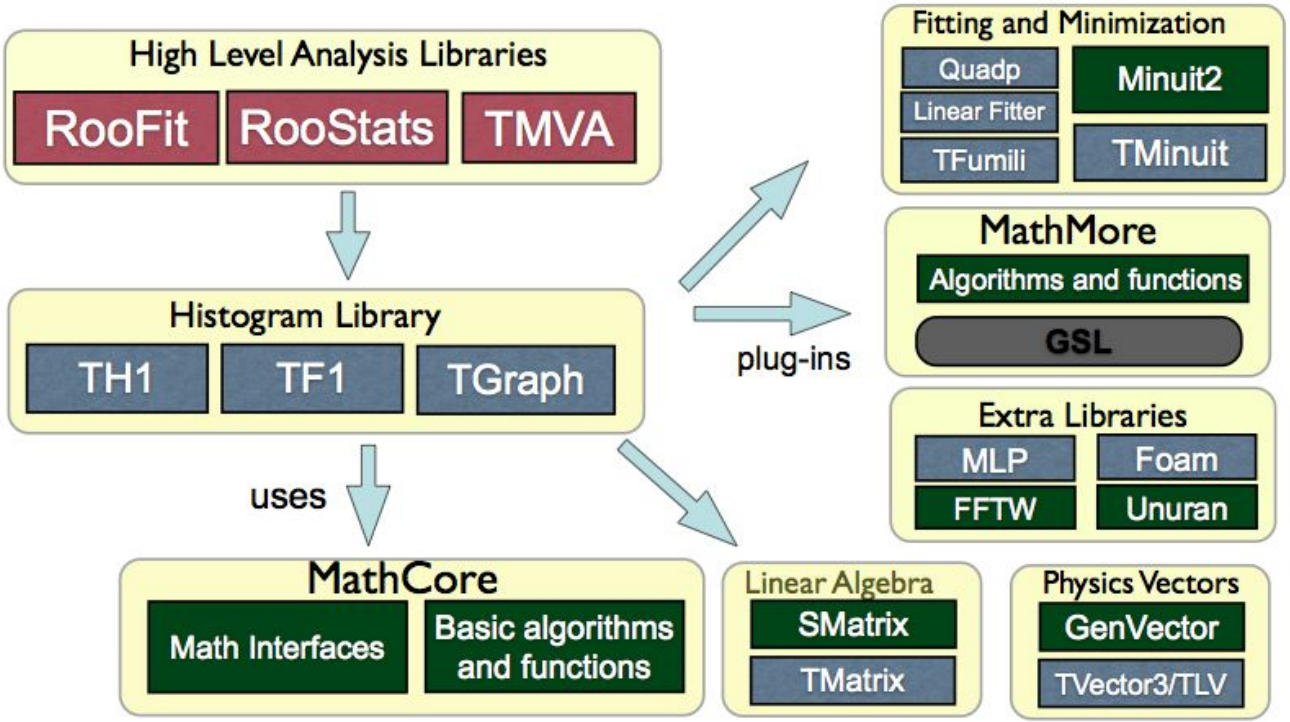
LHC Data in ROOT Format

1 EB

as of 2017

Mathematics and Statistics

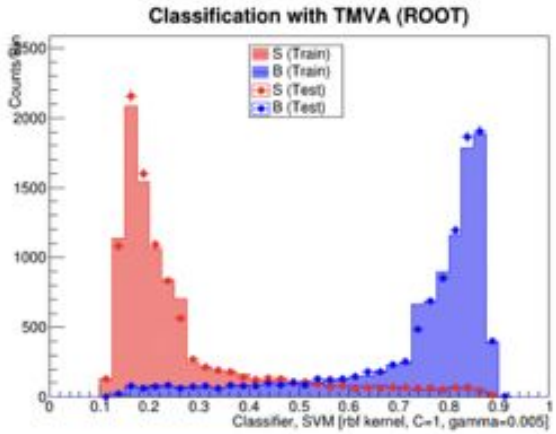
- ROOT provides a rich set of mathematical libraries and tools for sophisticated statistical data analysis



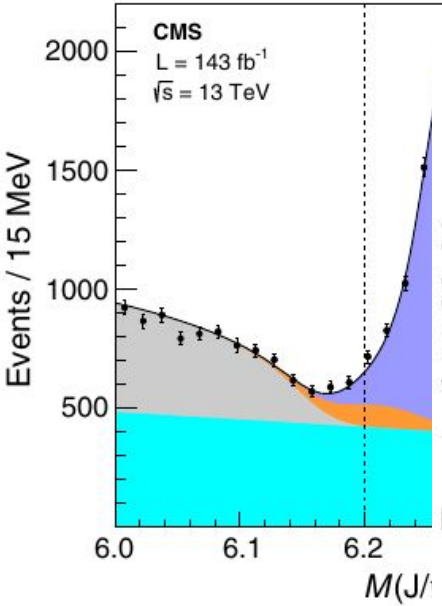
Machine Learning: TMVA

TMVA : Toolkit for Multi-Variate data Analysis in ROOT

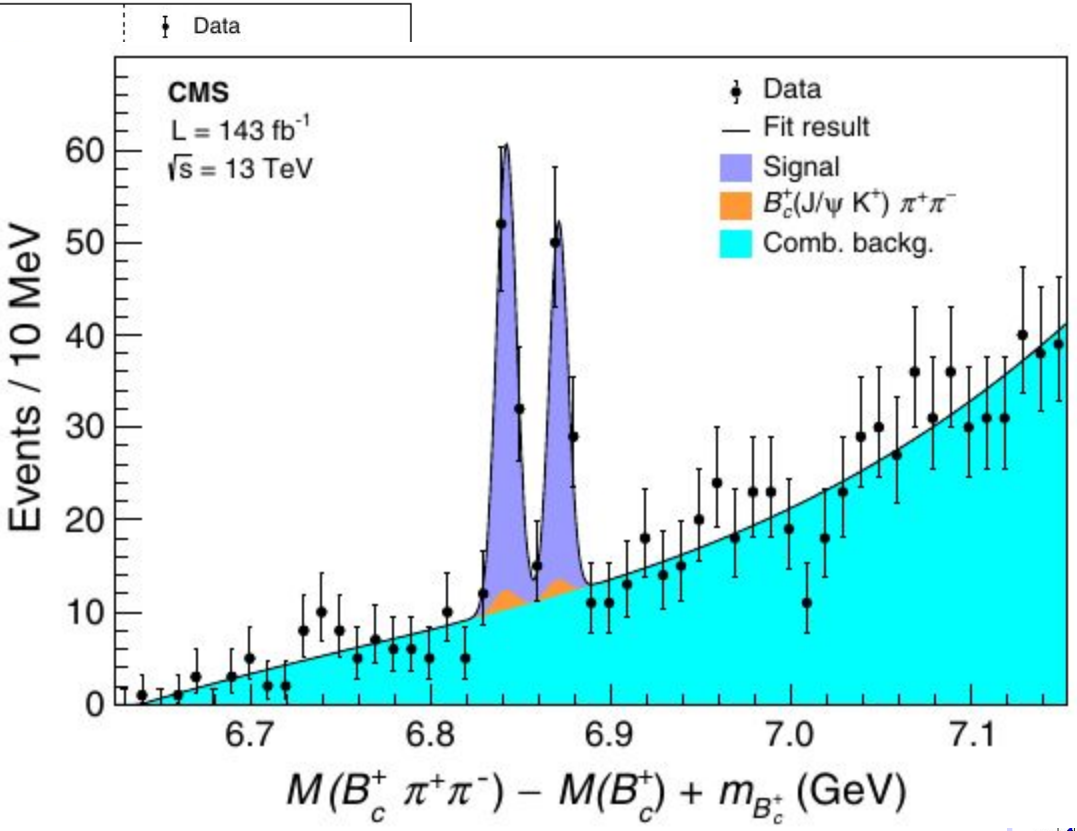
- provides several built-in ML methods including:
  - Boosted Decision Trees
  - Deep Neural Networks
  - Support Vector Machines
- and interfaces to external ML tools
  - scikit-learn, Keras (Theano/Tensorflow), R



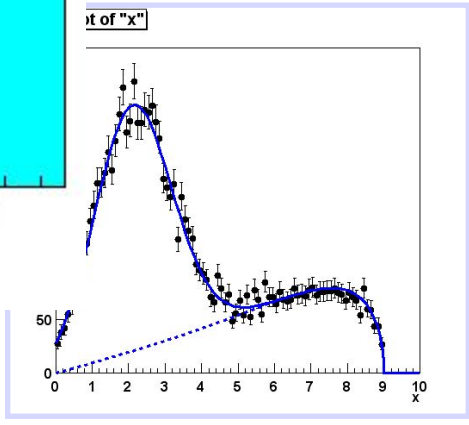
RooFit



- visualizatic
- sharing an



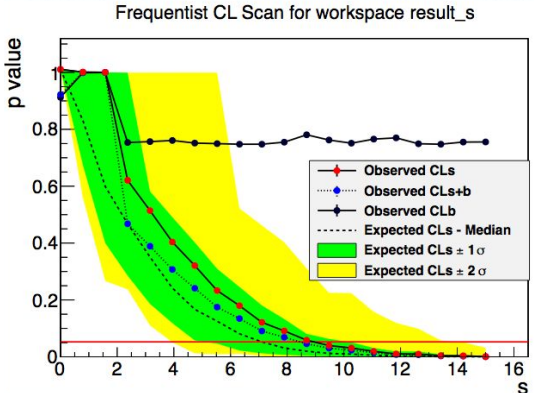
d.f.)

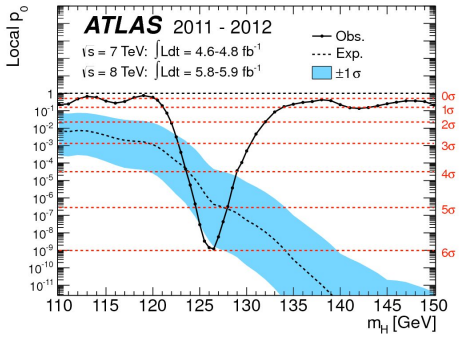


BPH-18-007, Phys. Rev. Lett. 122, 132001

RooStats

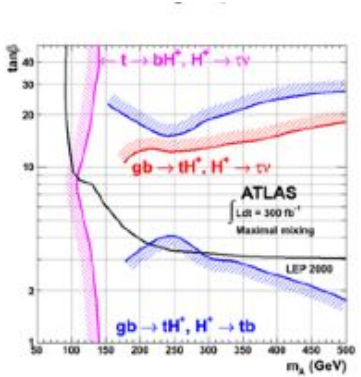
- Advanced Statistical Tools for HEP analysis. Used for :
  - estimation of Confidence/Credible intervals
  - hypotheses Tests
    - e.g. Estimation of Discovery significance
- Provides both Frequentist and Bayesian tools
- Facilitate combination of results

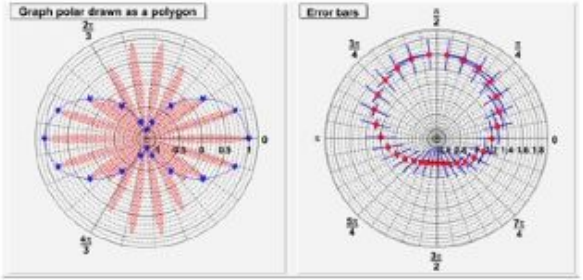


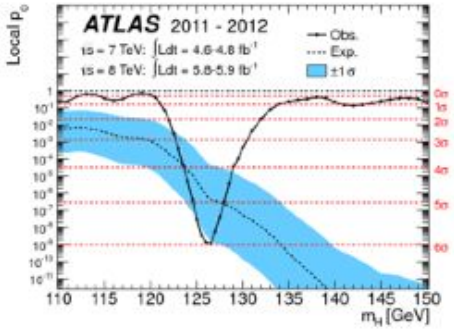


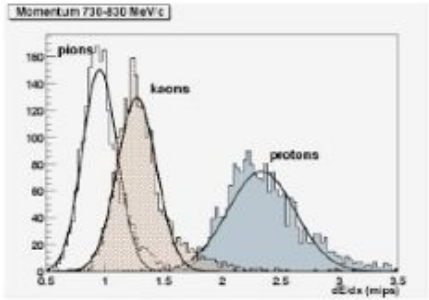
Graphics in ROOT

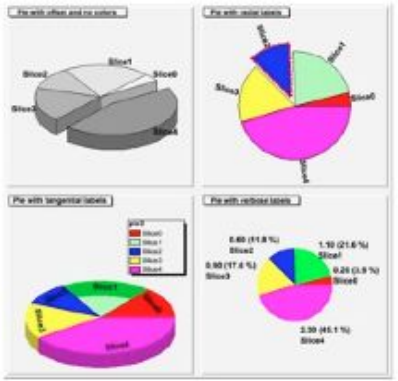
- Many formats for data analysis, and not only, plots

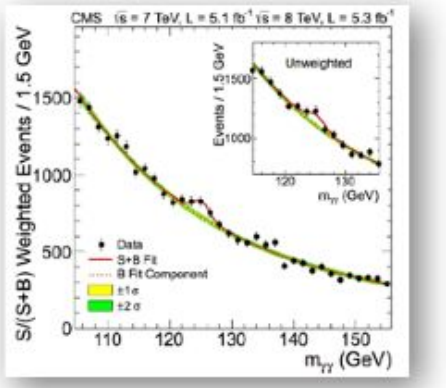




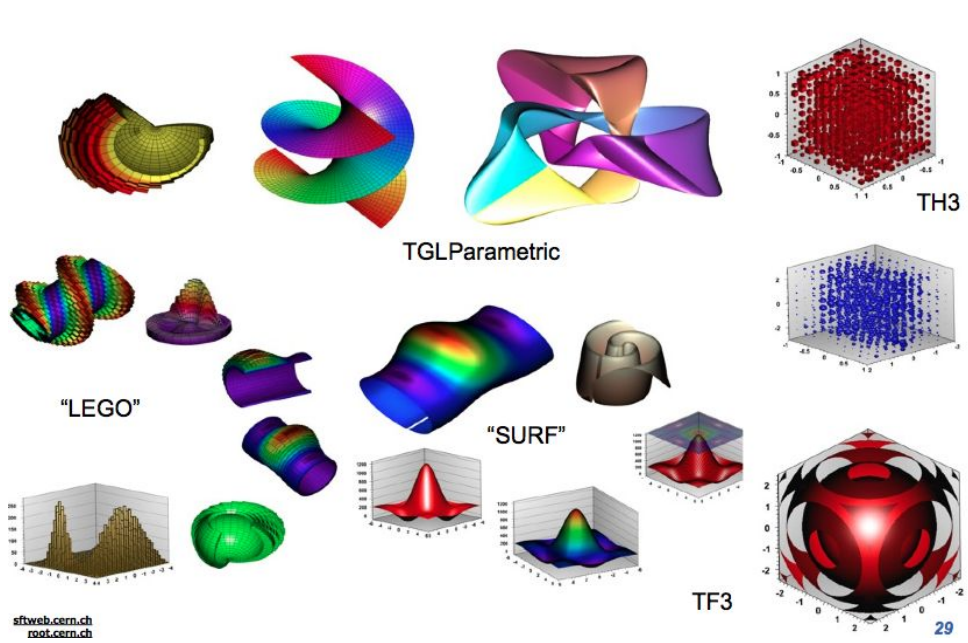


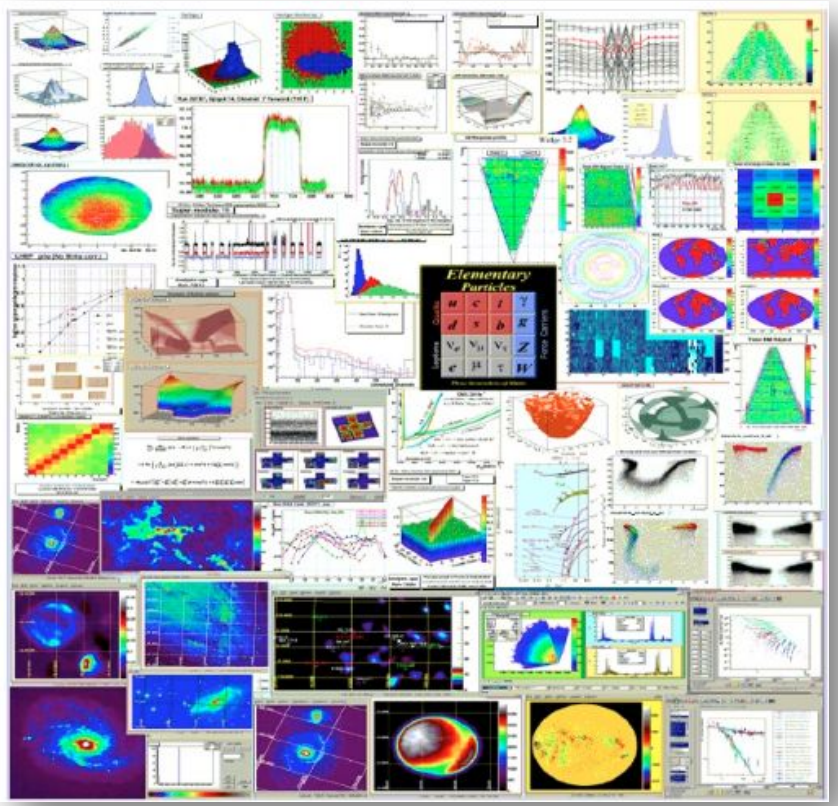






2D and 3D Graphics



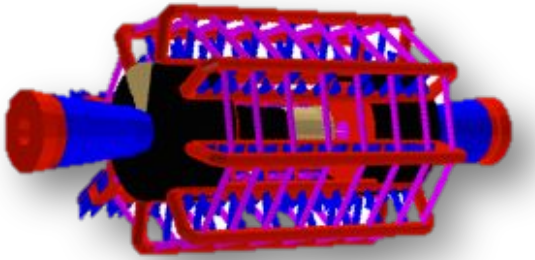


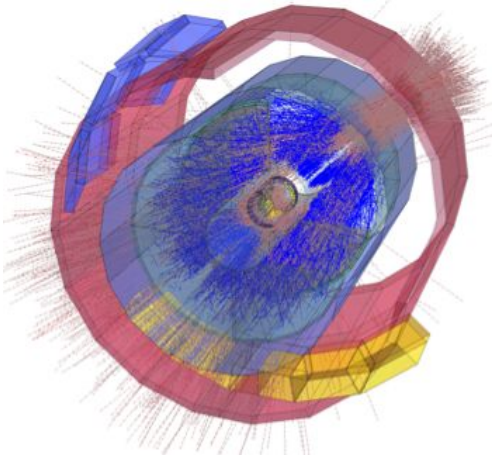
Can save graphics in many formats:
ps, pdf, svg, jpeg, LaTex, png, c, root …

Parallelism

- Many ongoing efforts to provide means for parallelisation in ROOT
- Explicit parallelism
  - **TThreadExecutor** and **TProcessExecutor**
  - Protection of resources
- Implicit parallelism
  - **TDataFrame**: functional chains
  - TTreeProcessor: process tree events in parallel
  - TTree::GetEntry: process of tree branches in parallel
- Parallelism is a fundamental element for tackling data analysis during LHC Run III and HL-LHC

Many More Features!





- Geometry Toolkit
  - Represent geometries as complex as LHC detectors
- Event Display (EVE)
  - Visualise particle collisions within detectors

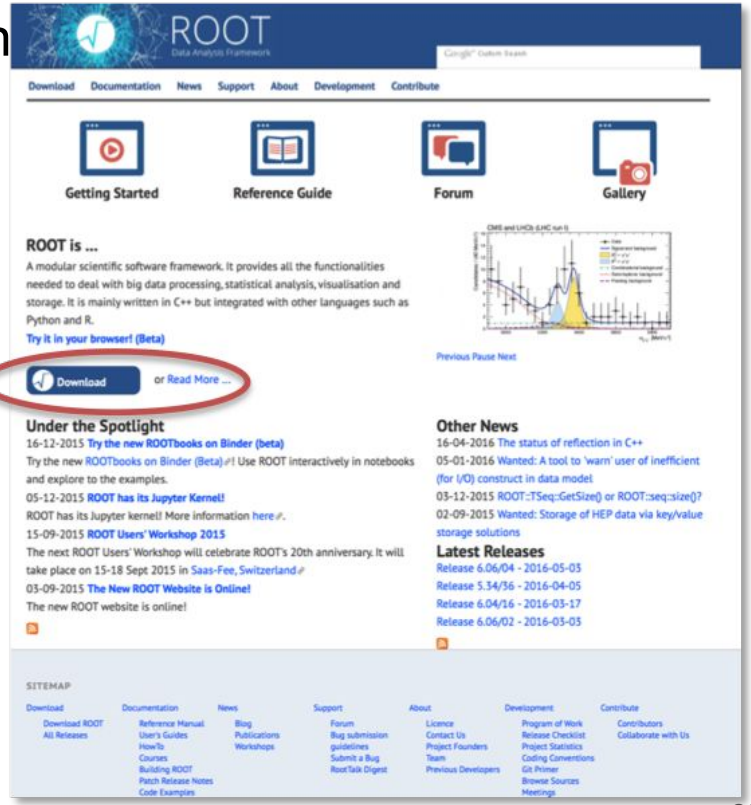
The SWAN Service

- **Data analysis with ROOT “as a service”**
- *Interface:* Jupyter Notebooks
- *Goals:*
  - Use ROOT only with a web browser
    - Platform independent ROOT-based data analysis
    - Calculations, input and results “in the Cloud”
  - Allow easy sharing of scientific results: plots, data, code
    - Through your CERNBox
  - Simplify teaching of data processing and programming

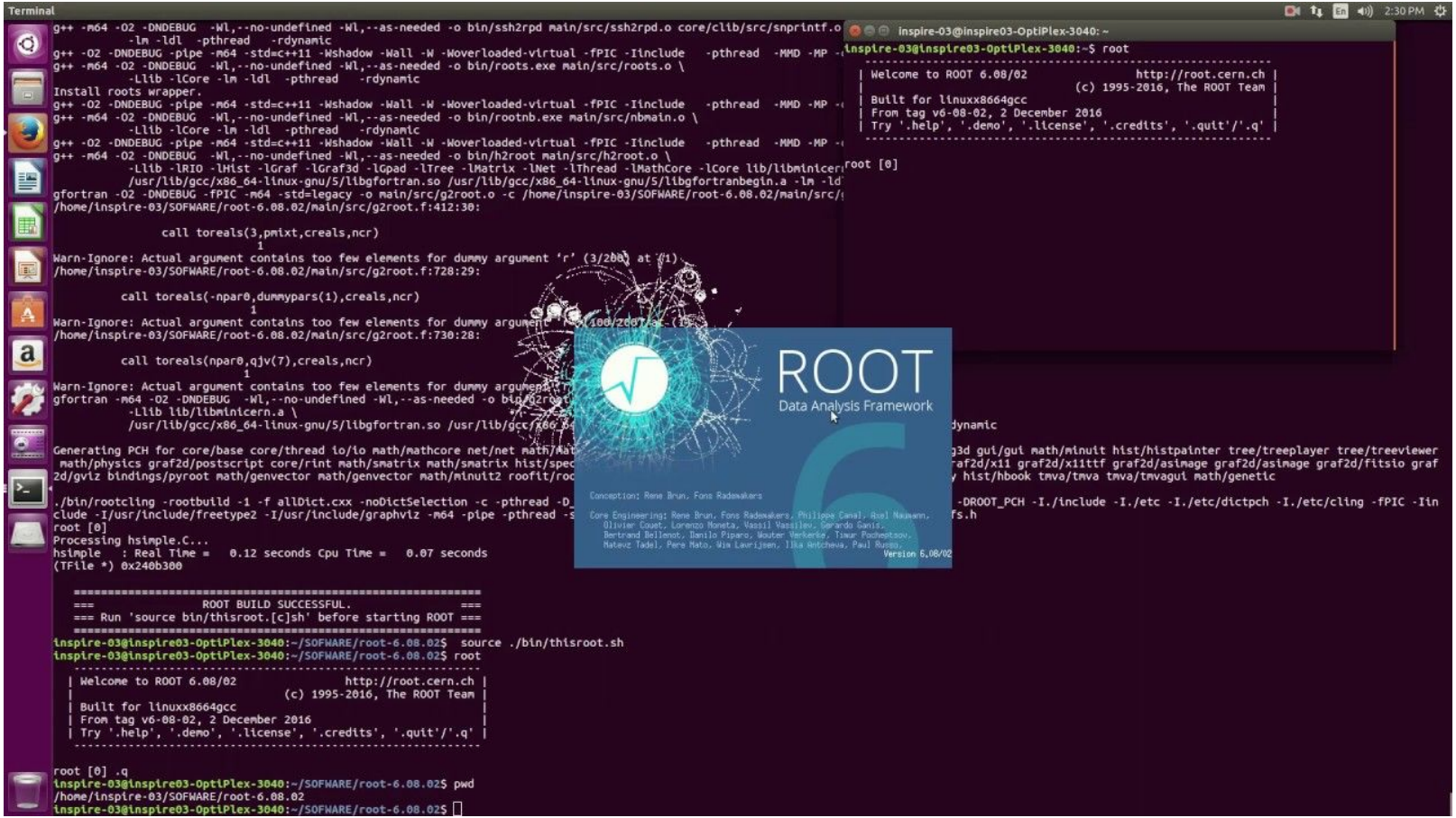
http://swan.web.cern.ch

https://root.cern

- ROOT web site: **the** source of information and help for ROOT users
  - For beginners and experts
  - Downloads, installation instructions
  - Documentation of all ROOT classes
  - Manuals, tutorials, presentations
  - Forum
  - ...



ROOT Basics


Terminal
inspire-03@inspire03-OptiPlex-3040: ~
ROOT
Data Analysis Framework
Conception: Rene Brun, Fons Rademakers
Version 6.08/02
inspire-03@inspire03-OptiPlex-3040:~/SOFWARE/root-6.08.02$ source ./bin/thisroot.sh
inspire-03@inspire03-OptiPlex-3040:~/SOFWARE/root-6.08.02$ root
Welcome to ROOT 6.08/02 http://root.cern.ch
(c) 1995-2016, The ROOT Team
Built for linuxx8664gcc
From tag v6-08-02, 2 December 2016
Try '.help', '.demo', '.license', '.credits', '.quit'/'.q'
root [0] .q
inspire-03@inspire03-OptiPlex-3040:~/SOFWARE/root-6.08.02$ pwd
/home/inspire-03/SOFWARE/root-6.08.02
inspire-03@inspire03-OptiPlex-3040:~/SOFWARE/root-6.08.02$

ROOT Basics: The ROOT Prompt

- C++ is a compiled language
  - A compiler is used to translate source code into machine instructions

- ROOT provides a C++ **interpreter**
  - Interactive C++, without the need of a compiler, like Python, Ruby, Haskell …
    - Code is **Just-in-Time compiled!**
  - Allows reflection (inspect at runtime layout of classes)
  - Is started with the command:

```
root
```

  - **The interactive shell is also called “ROOT prompt” or “ROOT interactive prompt”**

ROOT As a Calculator

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots$$

$$= \sum_{n=0}^{\infty} x^n$$

Here we make a step forward.
We declare **variables** and use a ***for*** control structure.

```
root [0] double x=.5
(double) 0.5
root [1] int N=30
(int) 30
root [2] double gs=0;
```

```
root [3] for (int i=0;i<N;++i) gs += pow(x,i)
root [4] std::abs(gs - (1/(1-x)))
(Double_t) 1.86265e-09
```

Controlling ROOT

- Special commands which are not C++ can be typed at the prompt, they start with a “.”

```
root [1] .<command>
```

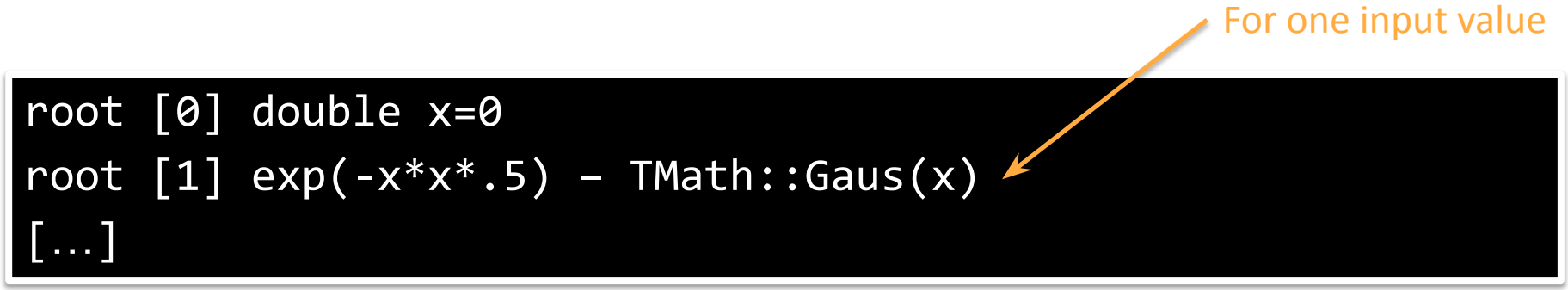
- For example:
  - To quit root use **.q**
  - To issue a shell command use **.! <OS_command>**
  - To load a macro use **.L <file_name>** (see following slides about macros)
  - **.help** or **.?** gives the full list

Ex Tempore Exercise

- Fire up ROOT
- Verify it works as a calculator
- Inspect the help
- Quit

Exercise

- ROOT provides mathematical functions, for example the widely known and adopted Gaussian
- For x values of 0,1,10 and 20 check the difference of the value of a hand-made non-normalised Gaussian and the TMath::Gaus routine



Solution

- For x values of 0,1,10 and 20 check the difference of the value of a hand-made non-normalised Gaussian and the TMath::Gaus routine

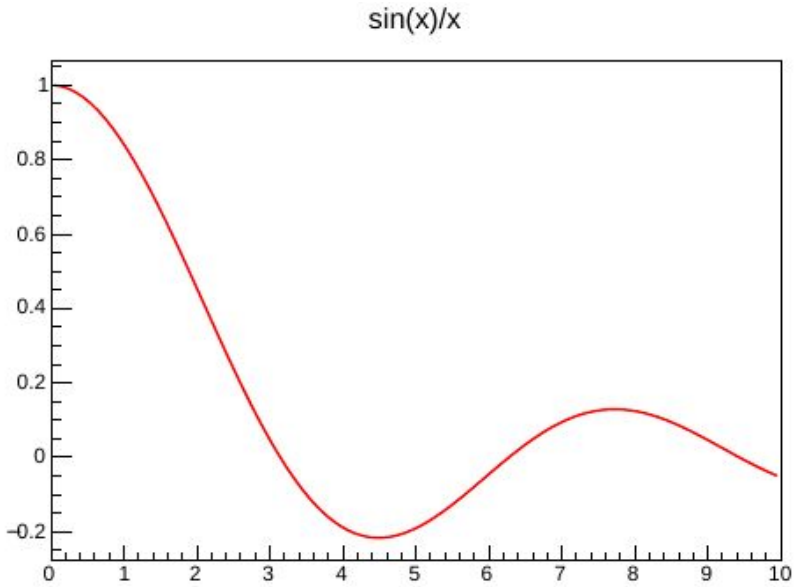
```
root [0] double x=0
root [1] exp(-x*x*.5) - TMath::Gaus(x)
[...]
```

- Many possible ways of solving this! E.g:

```
root [0] for (auto v : {0.,1.,10.,20.}) cout << v << " " <<
exp(-v*v*.5) - TMath::Gaus(v) << endl
```

my first function

```
root [0] auto fa1 = new TF1("fa1","sin(x)/x",0,10);
root [1] fa1->Draw();
[...]
```



Please, take a look at the "toolbar"

1) save as png,pdf,root,.....
2) change the line color, style
3) change the tf1 range
4) change the tf1 function (cos(x)+x)?
4) change the x-label name
6) continue to explore the options

my Error function

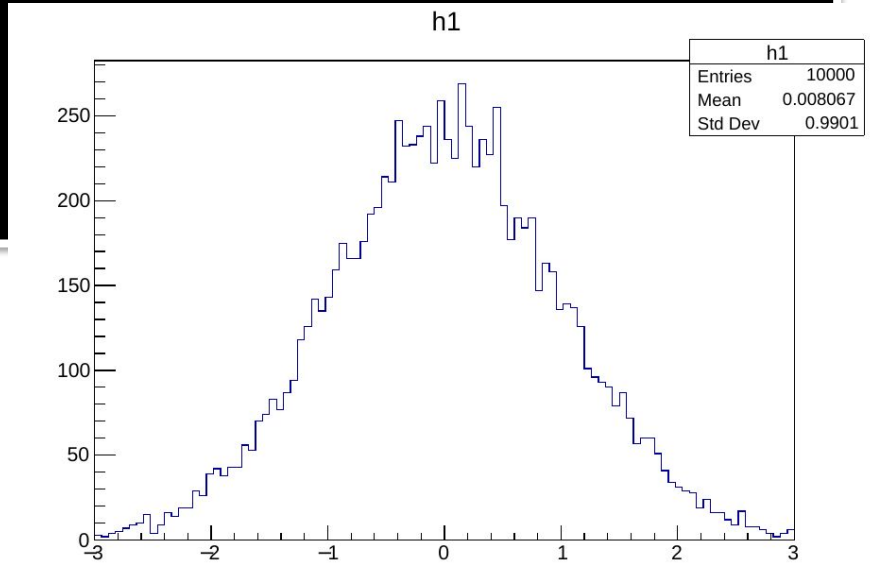
```
root [0] auto fa2 = new TF1("fa2","TMath::Erf(x)",0,2);
root [1] fa2->Draw();
[…]
```

Trigger Effy, https://en.wikipedia.org/wiki/Error_function,

```
root [13] auto fa2 = new TF1("fa2","TMath::Erf(x)",0,2);
root [14] fa2->Draw();
auto fa2 = new TF1("fa2","TMath::Erf(-x)",0,2);
root [15] fa2->Draw();
root [16] auto fa2 = new TF1("fa2","(TMath::Erf((-x+1.0)))",0,2);
root [17] fa2->Draw();
root [18] auto fa2 = new
TF1("fa2","(TMath::Erf((-x+1.0)/0.035))",0,2);
root [19] fa2->Draw();
```

my first histogram

```
root [0] TH1F h1("h1","h1",100,-3,3);
root [1] h1.FillRandom("gaus",10000);
root [2] h1->Draw();
ROOT_prompt_2:1:3: error: member reference type 'TH1F' is not a
pointer; did you mean to use '.'?
h1->Draw();
~~^~
  .
root [3] h1.Draw();
```



Please, take a look at the "toolbar", again.

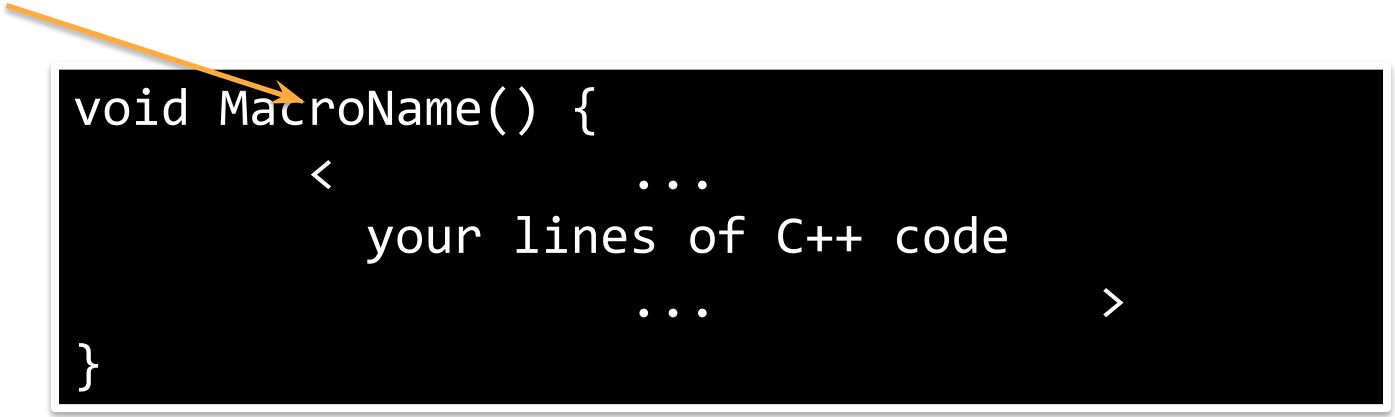
User function

[a]*exp(-0.5*((x-[b])/[c])*((x-[b])/[c]))

ROOT Macros

- We have seen how to interactively type lines at the prompt
- The next step is to write “ROOT Macros” – lightweight programs
- The general structure for a macro stored in file *MacroName.C* is:

Function, no main, same name as the file



Unnamed ROOT Macros

- Macros can also be defined with no name
- Cannot be called as functions!
  - See next slide :)

```
{
        <           ...
          your lines of C++ code
                    ...             >
}
```

Running a Macro

- A macro is executed at the system prompt by typing:

```
> root MacroName.C
```

- or executed at the ROOT prompt using .x:

```
> root
root [0] .x MacroName.C
```

- or it can be loaded into a ROOT session and then be run by typing:

```
root [0] .L MacroName.C
root [1] MacroName();
```

Time for Exercises

- Go back to the geometric series example we executed at the prompt
- Make a macro out of it and run it

Examples:
geometrica_macro.C
geometrica_macroName.C
¿Como volver esto un codigo puramente cpp?
¿geometrica_macroMain.C?

Examples:
myGaussfun.C
¿Como volver esto un codigo puramente cpp?
¿myGaussfunAlaC.C?

Interpretation and Compilation

- We have seen how ROOT interprets and “just in time compiles” code. ROOT also allows to compile code “traditionally”. At the ROOT prompt:

```
root [1] .L macro1.C+
root [2] macro1()
```

Generate shared library and execute function

- ROOT libraries can also be used to produce standalone, compiled applications:

```
int main() {
  ExampleMacro();
  return 0;
}
```

```
> g++ -o ExampleMacro ExampleMacro.C `root-config --cflags --libs`
> ./ExampleMacro
```

Un paréntesis útil: Argumentos de línea de comandos

Por suerte, la interfaz para **transmitir argumentos a una función main() se ha estandarizado en C++**, así que la emisión y recepción de argumentos puede hacerse de manera casi mecánica.

Los argumentos transmitidos a main(), como todos los argumentos de función, **deben declararse como parte de la definición de la función**.

int main(int argc, char *argv[])

Sin importar cuántos argumentos se mecanografíen en la línea de comandos, main() sólo necesita las dos piezas de información estándares proporcionadas por **argc y argv**; el número de elementos en la línea de comandos y la lista de direcciones iniciales que indican dónde se almacena en la actualidad cada argumento.
argc (abreviatura para contador de argumento)
argv (abreviatura para valores de argumentos)

Example: Parentesis_ARG

Cualquier argumento mecanografiado en una línea de comandos se considera una cadena en C. Si se desea transmitir datos numéricos a main(), depende de usted convertir la cadena transmitida en su contraparte numérica

Macro myGaussfun

- Make a macro out of it and run it

Examples:
myGaussfun.C
¿Como volver esto un codigo puramente cpp?
¿myGaussfunAlaC_2.C?
g++ -o miname myGaussfunAlaC_2.C `root-config --cflags --libs`
./miname

My first Canvas

```
root[0] new TCanvas          ← "quick" creation of graphical window with generic properties
(class TCanvas*)0x2c1e5e0    ← Notice: automatic name assignment ("c1")
root[1] c1->Set
```

```
root[1] c1->SetTitle ("HelloCanvas")
root[2] c1->GetTitle ()
(const char* 0x1557339)"HelloCanvas"
root[3] c1->ls ()
root[4] c1->Close()
root[5] TCanvas c2           ← We create a new window. Before we used pointer. Now – object.
root[6] c2.GetName()
(const char* 0x16632b1)"c1_n2"   ← Title is different than variable name!
root[7] TCanvas c3 (
```

Multitude of constructors

```
TCanvas TCanvas(Bool_t build = kTRUE)
TCanvas TCanvas(const char* name, const char* title = "", Int_t form = 1)
TCanvas TCanvas(const char* name, const char* title, Int_t ww, Int_t wh)
TCanvas TCanvas(const char* name, const char* title, Int_t wtopx, Int_t wtopy,
Int_t ww, Int_t wh)
TCanvas TCanvas(const char* name, Int_t ww, Int_t wh, Int_t winid)
root[7] TCanvas c3 ("c3canvas", "My canvas", 600, 400);
```

Proper name of object (within C++).	**"Name" (identifier)** (within ROOT).	Displayed title (just a c-string)

```
root[8] c3Canvas
(class TCanvas*)0x16964d0
root[9] TCanvas* c4 = new TCanvas ("c4canv", "2nd Canvas", 600, 400);
```

Name as identifier or replacement of C++ name

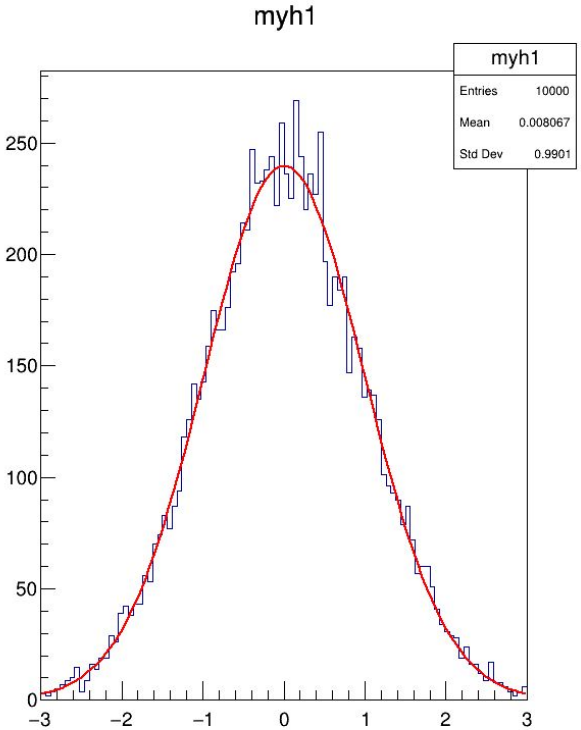
Dynamic allocation (we then use a pointer to an object)

TCanvas

My first Canvas

https://root.cern.ch/doc/master/classTCanvas.html

```
auto myc1 = new
TCanvas("myc1","myc1",600,800);
myc1->cd();
myc1->SetLeftMargin(0.15);
myc1->SetRightMargin(0.06);
myc1->SetTopMargin(0.09);
myc1->SetBottomMargin(0.14);
```



Please, take a look to this example:

https://root.cern/doc/master/canvas_8C.html

Example: myfirtsCanvas.C

Graphs

A graph is a graphics object made of two arrays X and Y, holding the x, y coordinates of n points. There are several graph classes, they are: TGraph, TGraphErrors, TGraphAsymmErrors, TMultiGraph, and TGraphMultiErrors.

```
Int_t n = 20;
Double_t x[n], y[n];
for (Int_t i=0; i<n; i++) {
x[i] = i*0.1;
y[i] = 10*sin(x[i]+0.2);
}
TGraph *gr1 = new TGraph (n, x, y);
TCanvas *c1 = new TCanvas("c1","Grap
// draw the graph with axis, contineous lir
gr->Draw("AC*");
```

Graph Draw Options

The various draw options for a graph are explained in **TGraph::PaintGraph**. They are:

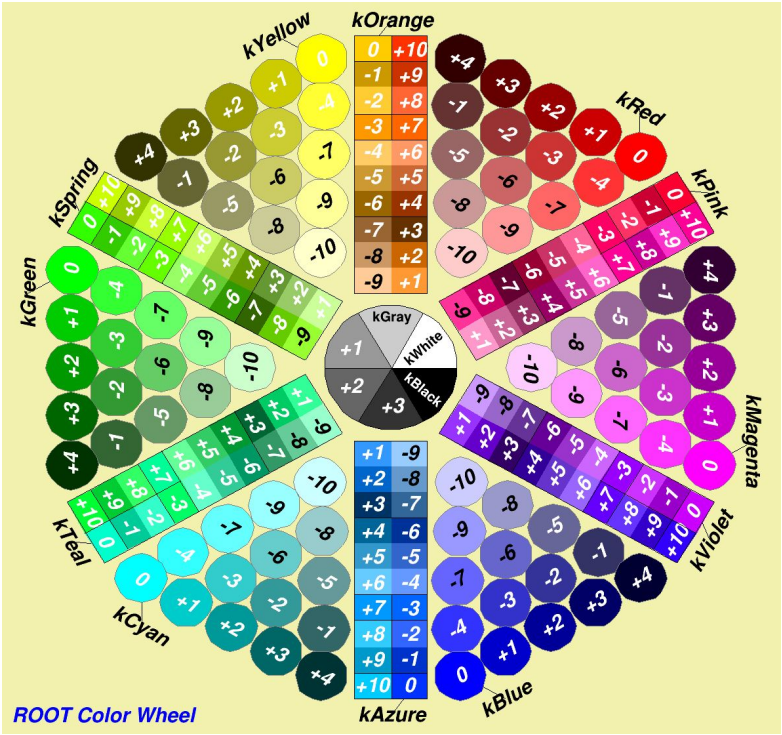
- "L" A simple poly-line between every points is drawn
- "F" A fill area is drawn
- "F1" Idem as "F" but fill area is no more repartee around X=0 or Y=0
- "F2" draw a fill area poly line connecting the center of bins
- "A" Axis are drawn around the graph
- "C" A smooth curve is drawn
- "*" A star is plotted at each point
- "P" The current marker of the graph is plotted at each point
- "B" A bar chart is drawn at each point
- "[]" Only the end vertical/horizontal lines of the error bars are drawn. This option applies to the **TGraphAsymmErrors**.
- "1" ylow = rwymin

```
TGraph *gr2 = new TGraph (n, x, y);
gr2->SetFillColor(40);
gr2->Draw("AB");
```

```
TGraph *gr3 = new TGraph(n,x,y);
gr3->SetFillColor(45);
gr3->Draw("AF")
```

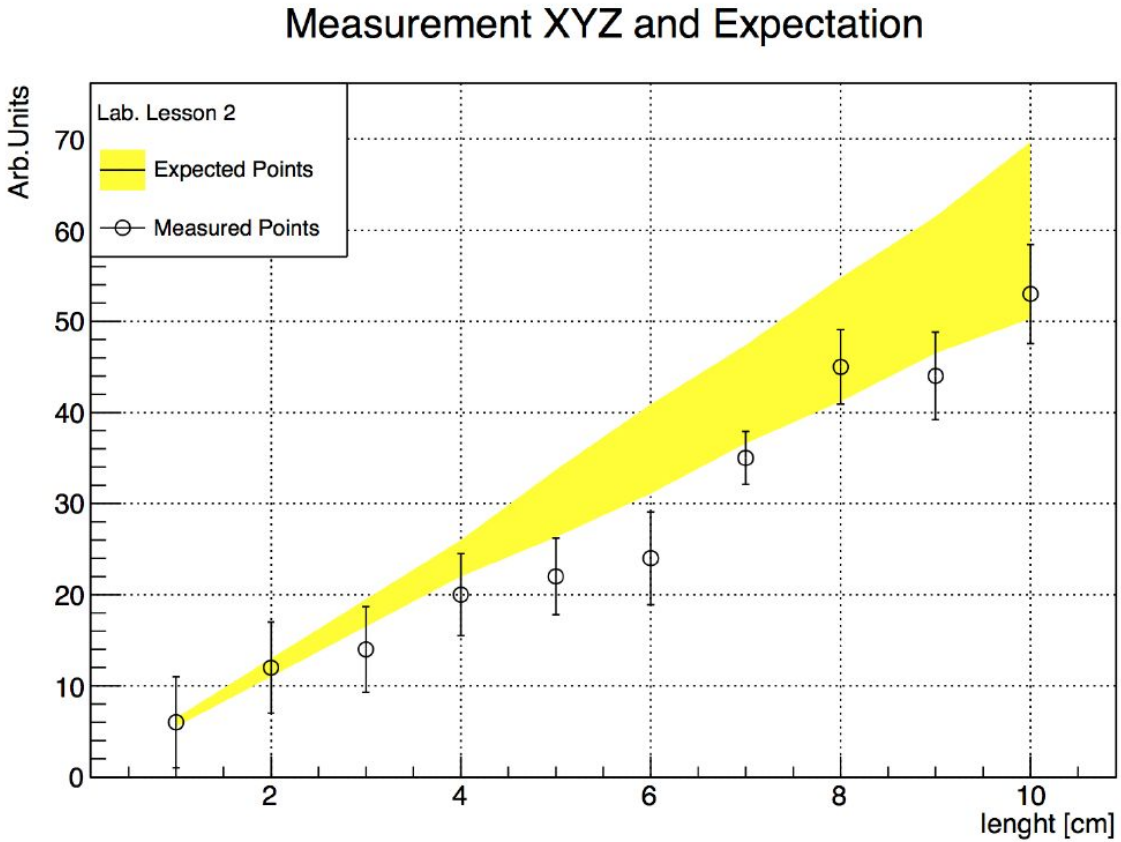
Graphs

TAttMarker Class



Example: TwoGraph.C

"Good Plot" example



Points from file

```
//Example: MyFirtsTgraph.C
void MyFirtsTgraph() {
  TGraphErrors myTgr2("ExampleData.txt");
  myTgr2.Draw("AP");; // shows histogram
}
```

ROOT doesn’t show my histogram!

Points from file

```
//Example: MyFirtsTgraph.C
void MyFirtsTgraph() {
  TGraphErrors *myTgr = new TGraphErrors("ExampleData.txt");
  //auto myTgr = new TGraphErrors("ExampleData.txt");
  myTgr->Draw("AP");
}
```

`new` puts object on “heap”, escapes scope

Points from file

```
//Example: MyTgraph2.C
{
  TGraphErrors myTgr2("ExampleData.txt");
  myTgr2.Draw("AP");; // shows histogram
}
```

if "anonymous" macro, this problem disappear?

Points from file

Check different constructors

```
//On the window
{
root [0]  TGraphErrors myTgr2("ExampleData.txt");
root [1] myTgr2.Draw("AP");
root [2] myTgr2.SetMarkerSize(0.8);
root [3] myTgr2.SetMarkerStyle(20);
root [4] myTgr2.Draw("AP");
}
```

```
root[1] TGraph g ("tgraph2.dat", "%*s %lg %*lg %lg");
root[2] g.Draw ("AP");

root[3] TGraph gr2( <Tab>
(...)
```

`%lg` :	read column as double precision
`%*lg`:	column type is double; omit it.
`%*s` :	column type is string; omit it.

let's move step by step

The data set

```
void Grap_StepbyStep(){
   // The number of points in the data set
   const int n_points = 10;
   // The values along X and Y axis
   double x_vals[n_points] = {1,2,3,4,5,6,7,8,9,10};
   double y_vals[n_points] = {6,12,14,20,22,24,35,45,44,53};
   // The errors on the Y axis
   double y_errs[n_points] = {5,5,4.7,4.5,4.2,5.1,2.9,4.1,4.8,5.43};

   // Instance of the graph
   auto graph = new TGraphErrors(n_points,x_vals,y_vals,nullptr,y_errs);
}
```

This code creates the **data set**.

⇒ Create a macro called "Grap_StepbyStep.C" and execute it.

The data drawing

As this graph has error bars along the Y axis an obvious choice will be to draw it as an **error bars plot**. The command to do it is:

```
graph->Draw("APE");
```

The Draw() method is invoked with three options:

- "A" the **axis** coordinates are automatically computed to fit the graph data
- "P" **points** are drawn as marker
- "E" the **error bars** are drawn

⇒ Add this command to the macro and execute it again.

⇒ Try to change the options (e.g. “APEL”, “APEC”, “APE4”).

Customizing the data drawing

Graphical attributes can be set to customize the visual aspect of the plot. Here we change the **marker style** and **color**.

```
// Make the plot esthetically better
graph->SetMarkerStyle(kOpenCircle);
graph->SetMarkerColor(kBlue);
graph->SetLineColor(kBlue);
```

⇒ Add this code to the macro. Notice the visual changes.

⇒ Play a bit with the possible values of the attributes.

Add a Function

The data set we built up looks very linear. It could be interesting to **compare it with a line** to see what the linear law behind this data set could be.

```
// Define a linear function
auto f = new TF1("Linear law","[0]+x*[1]",.5,10.5);
// Let's make the function line nicer
f->SetLineColor(kRed);
f->SetLineStyle(2);
// Set parameters
f->SetParameters(-1,5);
f->Draw("Same")
```

The function "**f**" graphical aspect is customized the same way the graph was before.

⇒ Add this code to the macro. Execute it again.

⇒ Play with the graphical attributes for the "**f**" function.

Plot Titles

The graph was created without any titles. It is time to define them. The following command **define the three titles** (separated by a ";") in one go. The format is:

```
"Main title ; x axis title ; y axis title"
```

```
graph->SetTitle("Measurement XYZ;length [cm];Arb.Units");
```

The axis titles can be also set individually:

```
graph->GetXaxis()->SetTitle("length [cm]");
graph->GetYaxis()->SetTitle("Arb.Units");
```

⇒ Add this code to the macro. You can choose one or the other way.

⇒ Play with the text. You can even try to add TLatex special characters.

Axis labelling (1)

The axis labels must be very **clear**, **unambiguous**, and **adapted to the data**.

- ROOT by default provides a powerful mechanism of **labelling optimisation.**
- Axis have three levels of divisions: Primary, Secondary, Tertiary.
- The axis labels are on the Primary divisions.
- The method `SetNdivisions` change the number of axis divisions

```
graph->GetXaxis()->SetNdivisions(10, 5, 0);
```

⇒ Add this command. nothing changes because they are the default values :-)

⇒ Change the number of primary divisions to 20. Do you get 20 divisions ? Why ?

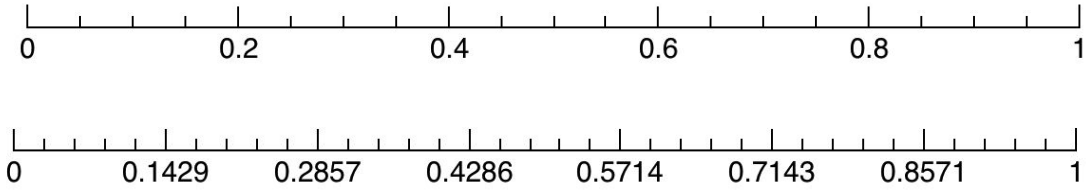
Axis labelling (2)

The number of divisions passed to the `SetNdivisions` method are **by default optimized** to get a comprehensive labelling. Which means that the value passed to `SetNDivisions` is a maximum number not the exact value we will get.

To turn off the optimisation the fourth parameter of `SetNDivisions` to `kFALSE`

⇒ try it

As an example, the two following axis both have been made with **seven** primary divisions. The first one is optimized and the second one is not. The second one is not



Nevertheless, in some cases, it is useful to have non optimized labelling.

Other kinds of axis

In addition to the normal linear numeric axis axis ROOT provides also:

- **Alphanumeric labelled axis**. They are produced when a histogram with alphanumeric labels is plotted.
- **Logarithmic axis**. Set with `gPad->SetLogx()` ⇒ try it (`gPad->SetLogy()` for the Y axis)
- **Time axis**.

Fine tuning of axis labels

When a specific **fine tuning** of an individual label is required, for instance changing its color, its font, its angle or even changing the label itself the method **`ChangeLabel`** can be used on any axis.

⇒ For instance, in our example, imagine we would like to highlight more the X label with the maximum deviation by putting it in red. It is enough to do:

```
graph->GetXaxis()->ChangeLabel(3,-1,-1,-1,kRed);
```

See here: https://root.cern/doc/master/gaxis3_8C_source.html

Legend (1)

ROOT provides a powerful class to build a legend: **`TLegend`**. In our example the legend is build as follow:

```
auto legend = new TLegend(.1,.7,.3,.9,"Lab. Lesson 1");
legend->AddEntry(graph,"Exp. Points","PE");
legend->AddEntry(f,"Th. Law","L");
legend->Draw();
```

⇒ Add this code, try it and play with the options

- The `TLegend` constructor defines the **legend position and its title**.
- Each legend item is added with method **`AddEntry`** which specify
  - **an object to be part of the legend** (by name or by pointer),
  - the **text of the legend**
  - the **graphics attributes to be legended**. "L" for TAttLine, "P" for TAttMarker, "E" for error bars etc …

Legend (2)

ROOT also provides an **automatic way to produce a legend** from the graphics objects present in the pad. The method **`TPad::BuildLegend()`**.

⇒ In our example try **`gPad->BuildLegend()`**

This is a quick way to proceed but for a plot ready for publication it is recommended to use the method previously described.

⇒ Keep only the legend defined in the previous slide

Annotations (1)

Often the various parts of a plot we already saw are not enough to convey the complete message this plot should. Additional **annotations** elements are needed **to complete and reinforce the message** the plot should give.

ROOT provides a collection of basic graphics primitives allowing to draw such annotations. Here a non exhaustive list:

- `TText` and `TLatex` to draw text.
- `TArrow` to draw all kinds of arrows
- `TBox`, `TEllipse` to draw boxes and ellipses.
- Etc ...

Annotations (2)

In our example we added an annotation pointing the "maximum deviation". It consists of a simple arrow and a text:

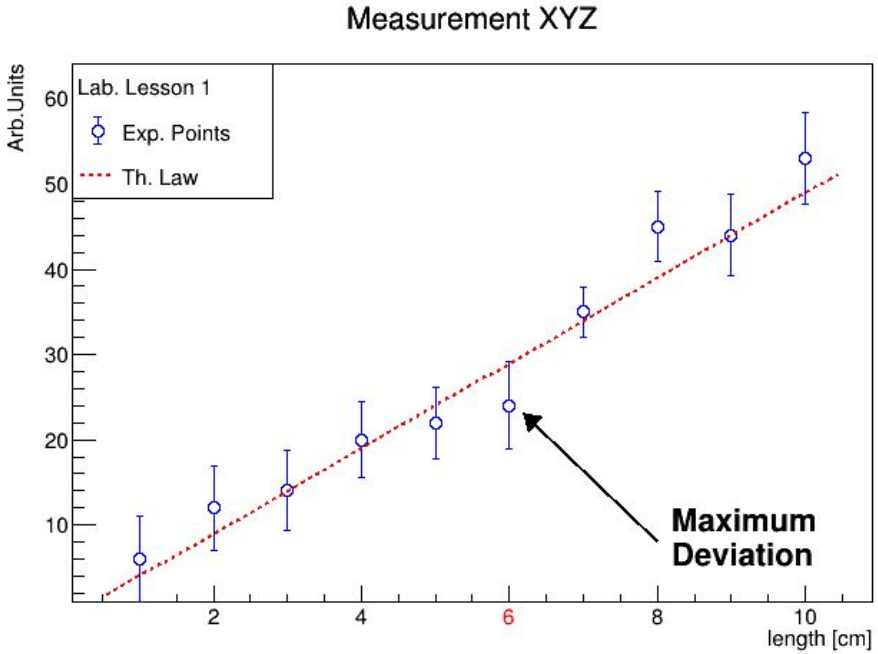
```
// Draw an arrow on the canvas
auto arrow = new TArrow(8,8,6.2,23,0.02,"|>");
arrow->SetLineWidth(2);
arrow->Draw();

// Add some text to the plot
auto text = new TLatex(8.2,7.5,"#splitline{Maximum}{Deviation}");
text->Draw();
```

⇒ Try this code adding it in the macro.

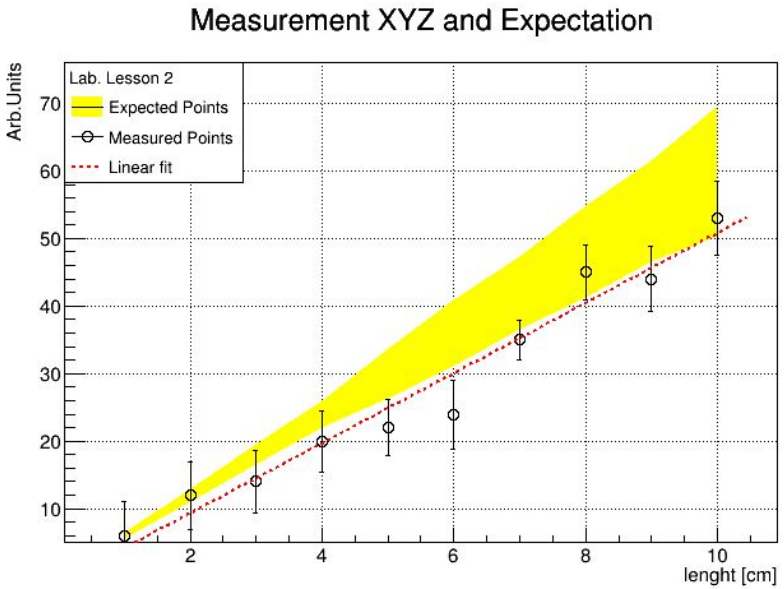
⇒ Notice changing the canvas size does not affect the pointed position.

Seguimiento1

1)



2)



Teniendo en cuenta los archivos: "macro2_input_expected.txt" y "macro2_input_mio.txt", y el ejemplo realizado para el plot "1)", haga el plot "2)" . Además, el conjunto de datos que construimos parece muy lineal. Podría ser interesante ajustarlo con una línea para ver cuál podría ser la ley lineal detrás de este conjunto de datos.
¿Que valores de los parametros del fit obtiene?