ROOT Summer Student Course

M. Czurylo, F. De Geus, L. Moneta, A. Naumann
for the ROOT Team

ROOT
Data Analysis Framework

https://root.cern

Agenda and exercises

- https://indico.cern.ch/event/1298470/

Make Sure One of These Works for You!

- On Lxplus

```
$ ssh -XY <username>@lxplus.cern.ch

[you@lxplus7xx ~]$ source /cvmfs/sft.cern.ch/lcg/views/LCG_103/x86_64-centos7-gcc12-opt/setup.sh
```

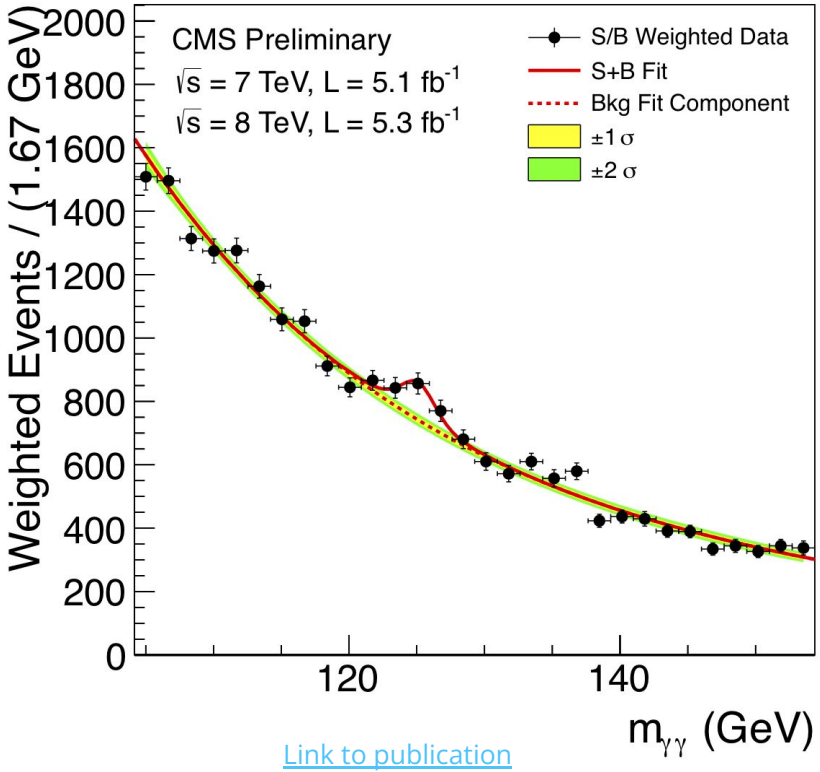
  - To be able to use graphics over SSH on Windows follow these instructions.
- On SWAN: https://swan.cern.ch
  - The Jupyter Notebook service of CERN
  - IMPORTANT: first visit https://cernbox.cern.ch
- On your machine (Linux or Mac)
  - For experts
  - See https://root.cern/install

Note: ROOT on Windows might not have support for some features.

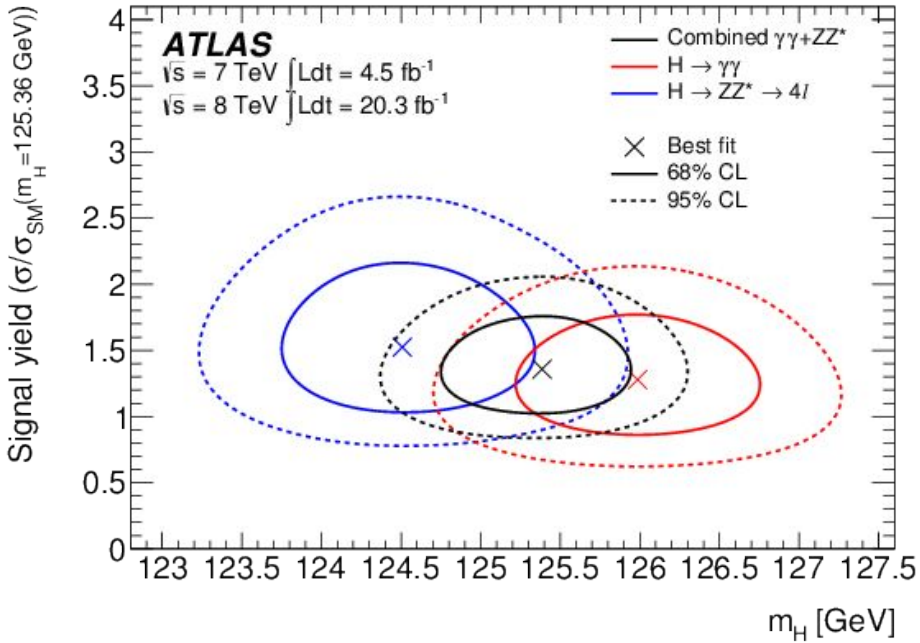
Introduction

A Quick Tour of ROOT

What can you do with ROOT?



Link to publication



Link to publication

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, Neural Nets
- **Advanced graphics** (2D, 3D, event display)
- **Declarative Analysis**: RDataFrame
- And more: HTTP servering, JavaScript visualisation

LGPL Free Software

An Open Source Project

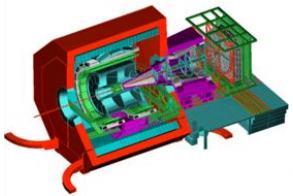
Unstar 595 Fork 433

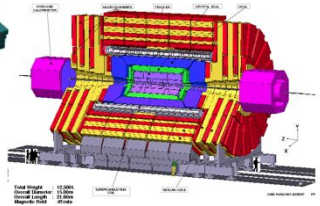
176 contributors

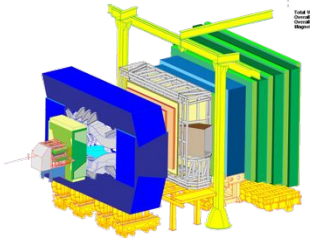
https://github.com/root-project/root

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

Plots

Data Storage: Local, Network

LHC Data in ROOT Format

~1 EB

(exa = 10^{18})

as of 2023

https://root.cern

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

Resources

- ROOT Website: https://root.cern
- Training: https://github.com/root-project/training
- More material: https://root.cern/get_started
  - Includes a booklet for beginners: **the "ROOT Primer"**
- Reference Guide: https://root.cern/doc/master/index.html
- Forum: https://root-forum.cern.ch

From Sources

Expert Level

- Get the ROOT sources:
  - *git clone http://github.com/root-project/root*
  - *Or visit https://root.cern.ch/content/release-62406*
- Create a build directory and configure ROOT:
  - *mkdir rootBuild; cd rootBuild*
  - *cmake ../root*
  - *https://root.cern.ch/building-root* for all the config options
- Start compilation
  - *make -j*
- Prepare environment:
  - *. bin/thisroot.sh*

The ROOT Prompt and Macros

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!**
  - Is started with the command:

```
root
```

  - The interactive shell is also called "ROOT prompt" or "ROOT interactive prompt"

ROOT As a Calculator

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

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

ROOT can be used as a simple calculator, but we let's make a step forward: 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>**
  - **.help** or **.?** gives the full list

Ex Tempore Exercise

- Fire up ROOT
- Verify it works as a calculator
- List the files in /etc from within the ROOT prompt
- Inspect the help
- Quit

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

```
void MacroName() {
         <           ...
           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();
```

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:

Advanced Users

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

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

Time For Exercises

- Exercises

  https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/C%2B%2BInterpreter

PyROOT: The ROOT Python Bindings

PyROOT

- Python bindings for ROOT
- Access all the ROOT C++ functionality from Python
  - Benefit from C++ performance
- Dynamic, automatic
- "Pythonisations" for specific cases

Using PyROOT

- Entry point to use ROOT from within Python

```
import ROOT
```

- All ROOT classes can be accessed from Python (we will present some of them soon!)

```
ROOT.TH1F
ROOT.TGraph
...
```

The ROOTBooks

The Jupyter Notebook

A web-based interactive computing platform that combines code, equations, text and visualisations.

Many supported languages: C++, Python, Haskell, Julia…
One generally speaks about a "kernel" for a specific language

In a nutshell: an "interactive shell opened within the browser"

http://www.jupyter.org

How It Looks Like

Text

Code

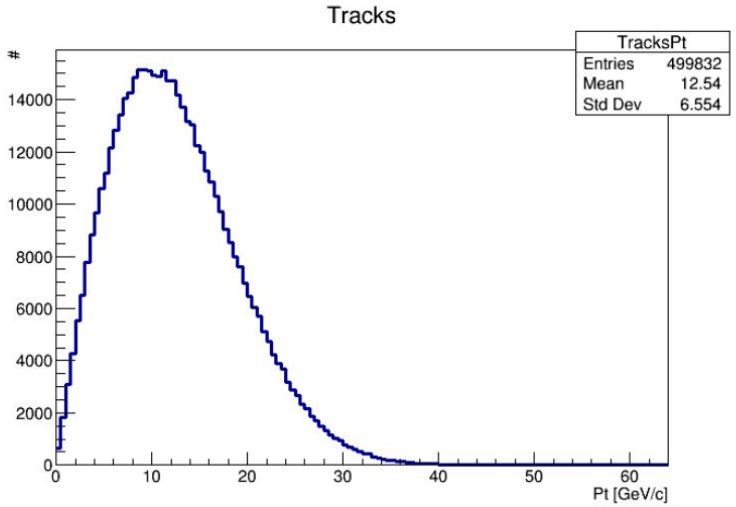
Graphics

Access TTree in Python using PyROOT and fill a histogram

Loop over the TTree called "events" in a file located on the web. The tree is accessed with the dot operator. Same holds for the access to the branches: no need to set them up - they are just accessed by name, again with the dot operator.

In [1]:
```
import ROOT

f = ROOT.TFile.Open("http://indico.cern.ch/event/395198/material/0/0.root");
h = ROOT.TH1F("TracksPt","Tracks;Pt [GeV/c];#",128,0,64)
for event in f.events:
    for track in event.tracks:
        h.Fill(track.Pt())
c = ROOT.TCanvas()
h.Draw()
c.Draw()
```



Use Notebooks at CERN

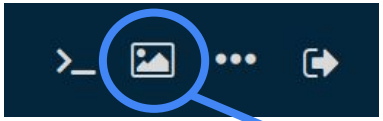
- **SWAN**: **S**ervice for **W**eb based **AN**alysis

- Get a CERNBox (if you don't have one)
  - Visit https://cernbox.cern.ch
- Log in to https://swan.cern.ch
- Create a project and a C++/Python notebook
  - Type in some code
  - Run it
  - Create markdown cells
  - Rerun cells with definitions

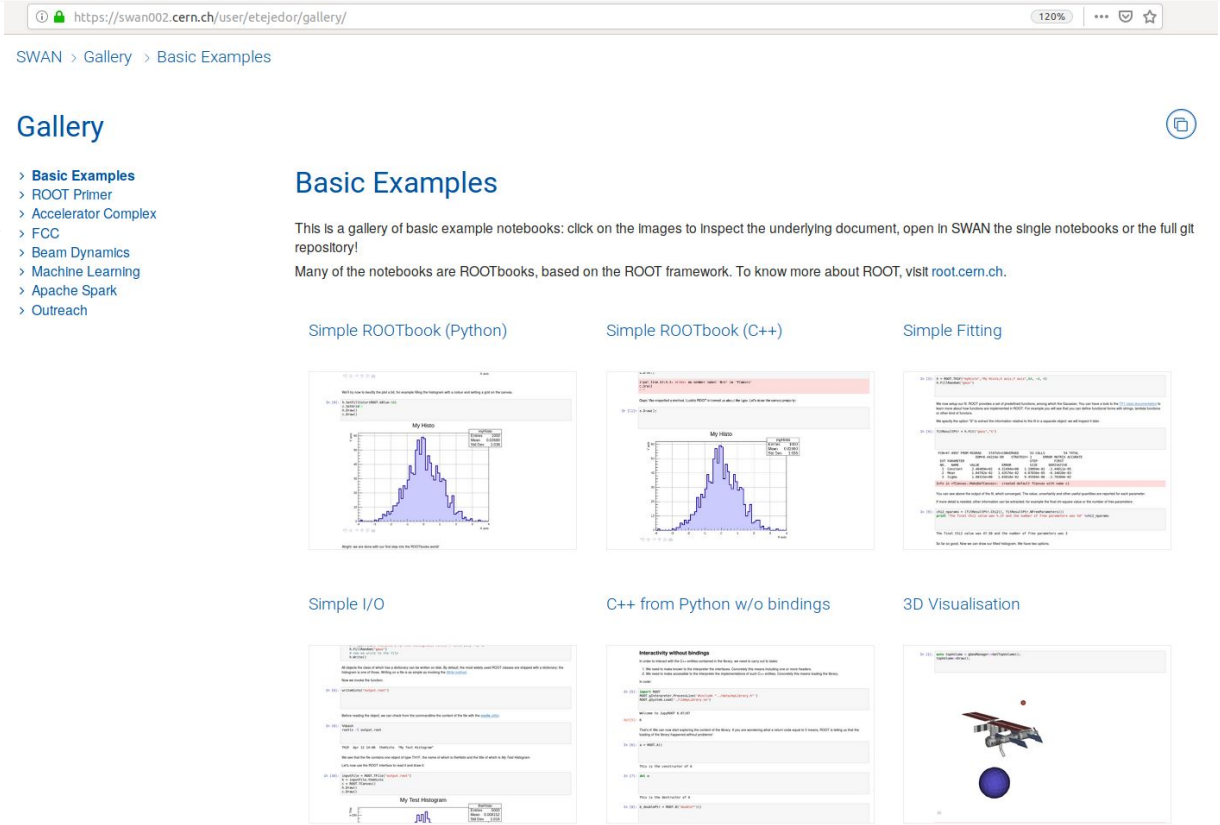
Notebooks On Your Machine

- Possible to install Jupyter as a package
- Fire up with the `root --notebook` command

Examples

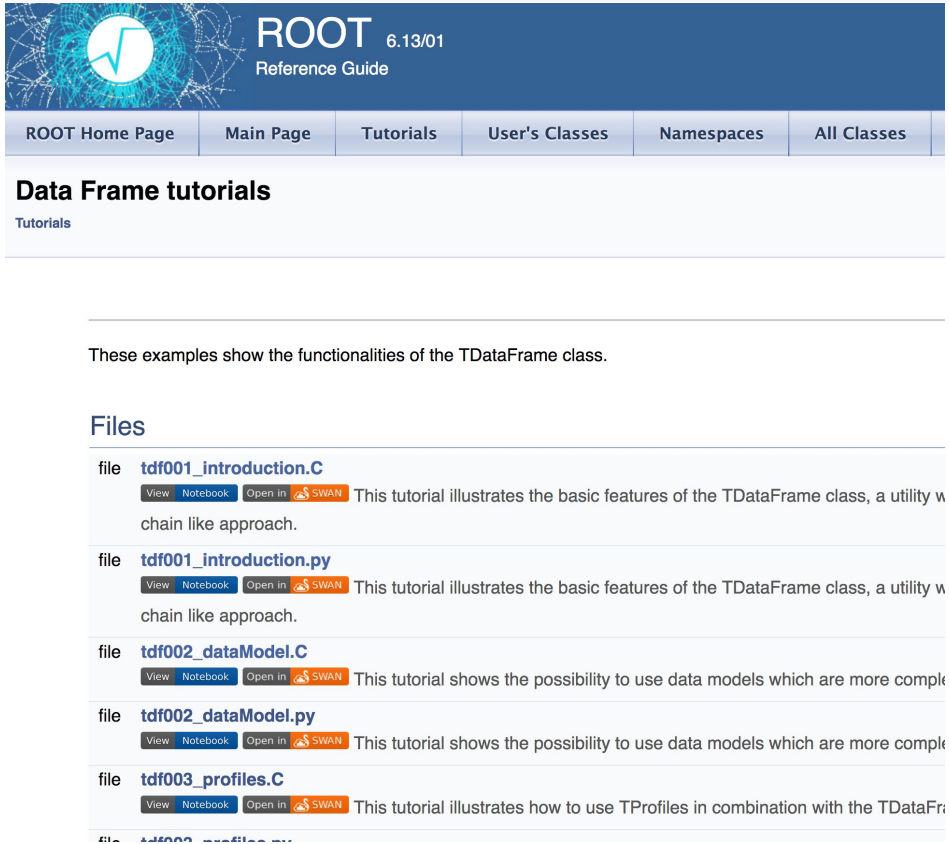


Open the gallery section in SWAN!



More Examples

[ROOT Tutorials]



ROOT 6.13/01
Reference Guide

ROOT Home Page | Main Page | Tutorials | User's Classes | Namespaces | All Classes

Data Frame tutorials

Tutorials

These examples show the functionalities of the TDataFrame class.

Files

file **tdf001_introduction.C**
View Notebook Open in SWAN This tutorial illustrates the basic features of the TDataFrame class, a utility w
chain like approach.

file **tdf001_introduction.py**
View Notebook Open in SWAN This tutorial illustrates the basic features of the TDataFrame class, a utility w
chain like approach.

file **tdf002_dataModel.C**
View Notebook Open in SWAN This tutorial shows the possibility to use data models which are more comple

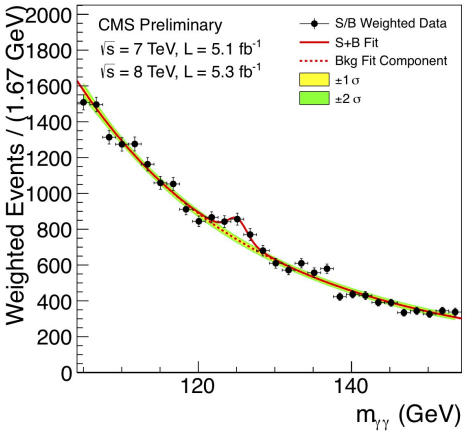
file **tdf002_dataModel.py**
View Notebook Open in SWAN This tutorial shows the possibility to use data models which are more comple

file **tdf003_profiles.C**
View Notebook Open in SWAN This tutorial illustrates how to use TProfiles in combination with the TDataFr

Histograms, Graphs and Functions

Histograms

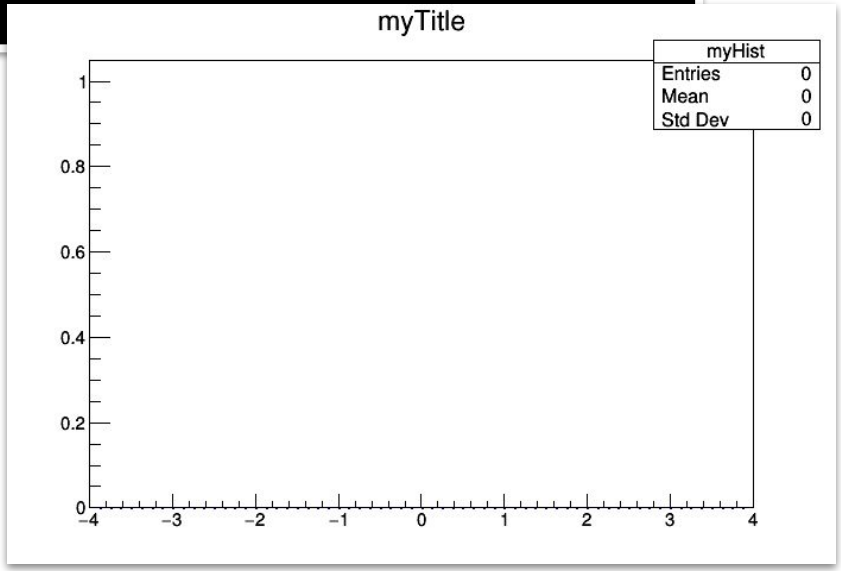
- A simple form of data reduction
  - Can have billions of collisions, the Physics displayed in a few histograms
  - It is like an empirical density estimator
  - Possible to calculate momenta: mean, rms, skewness, kurtosis …
- Collect quantities in discrete categories, the bins
- ROOT Provides a rich set of histograms
  - In multiple dimensions: *TH{1,2,3}* classes + *THN*
  - Holding different precision types
    - *TH1D* is a one-dim histogram holding doubles
    - Have also *TH{1,2,3}F* (float), *I* (int32), *S* (int16), *C* (int8)



My First Histogram

```
root [0] TH1D h("myHist", "myTitle", 64, -4, 4)
root [1] h.Draw()
```

C++



Note that in **Jupyter notebooks** (also in **SWAN**), the figure is not shown directly. You have to:

1. Either call `gPad->Draw()` at the end:

```
In [1]: TH1D h("myHist", "myTitle", 64, -4, 4);
        h.Draw();
        gPad->Draw();
```

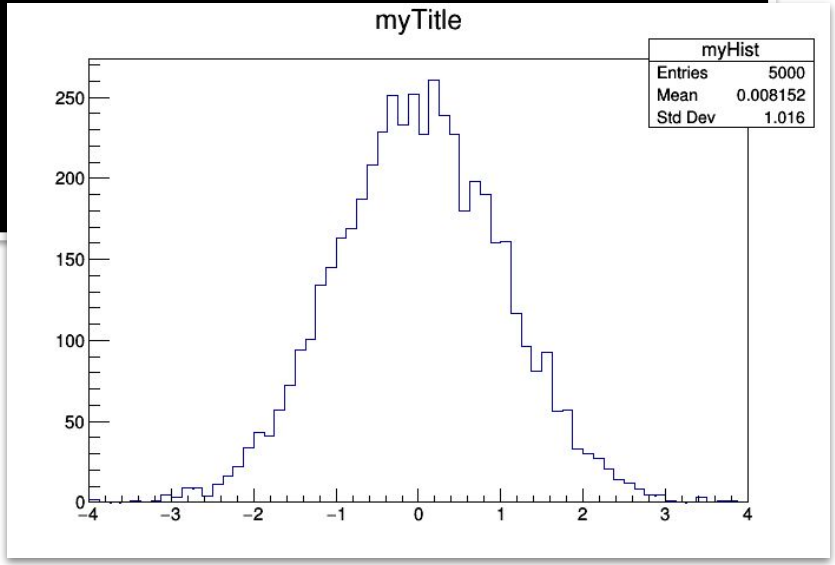
2. Or you can create a `TCanvas` and draw it:

```
In [2]: TCanvas c1;
        TH1D h("myHist", "myTitle", 64, -4, 4);
        h.Draw();
        c1.Draw();
```

My First Histogram

C++

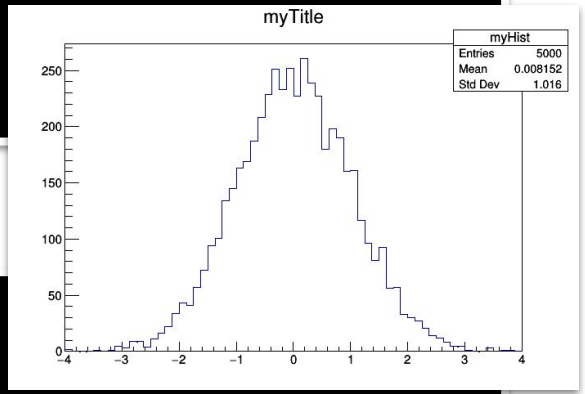
```
root [0] TH1D h("myHist", "myTitle", 64, -4, 4)
root [1] h.FillRandom("gaus")
root [2] h.Draw()
```



And now in Python!

C++

```
> root
root [0] TH1F h("myHist", "myTitle", 64, -4, 4)
root [1] h.FillRandom("gaus")
root [2] h.Draw()
```



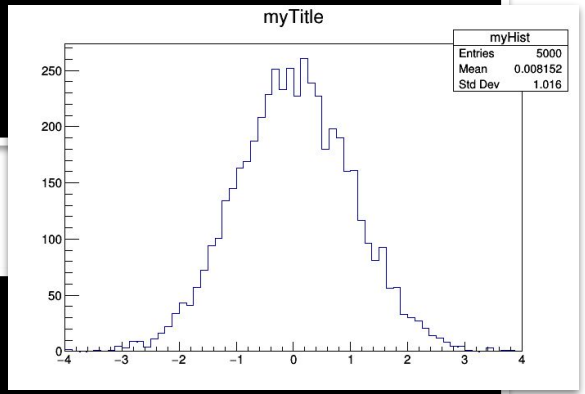

```
> python
>>> import ROOT
>>> h = ROOT.TH1F("myHist", "myTitle", 64, -4, 4)
>>> h.FillRandom("gaus")
>>> h.Draw()
```

Python

And now in Python!

C++

```
> root
root [0] TH1F h("myHist", "myTitle", 64, -4, 4)
root [1] h.FillRandom("gaus")
root [2] h.Draw()
```



also with
individual import

```
> python
>>> from ROOT import TH1F
>>> h = TH1F("myHist", "myTitle", 64, -4, 4)
>>> h.FillRandom("gaus")
>>> h.Draw()
```

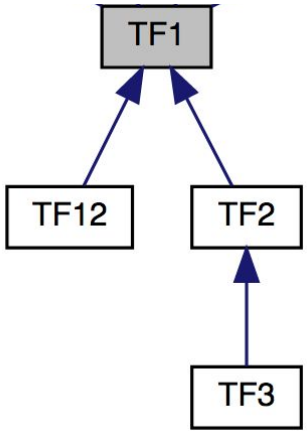
Python

Functions

- Mathematical functions are represented by the **TF1** class
- They have names, formulas, line properties, can be evaluated as well as their integrals and derivatives
  - Numerical techniques for generic cases
  - Automatic differentiation can be used for derivatives

option	description
"SAME"	superimpose on top of existing picture
"L"	connect all computed points with a straight line
"C"	connect all computed points with a smooth curve
"FC"	draw a fill area below a smooth curve

From the TGraphPainter documentation:
https://root.cern.ch/doc/master/classTGraphPainter.html



Functions

Can describe functions as:

- Mathematical formulas (written as strings)
- C++ functions/functors/lambdas
  - Implement your highly performant custom function
- Python functions
  - Fast prototyping on the Python side
- With and without parameters
  - Crucial for fits and parameter estimation

ROOT as a Function Plotter

- The class TF1 represents one-dimensional functions (e.g. *f(x)*):

```
root [0] TF1 f1("f1","sin(x)/x",0.,10.); //name,formula,min,max
root [1] f1.Draw();
```

C++

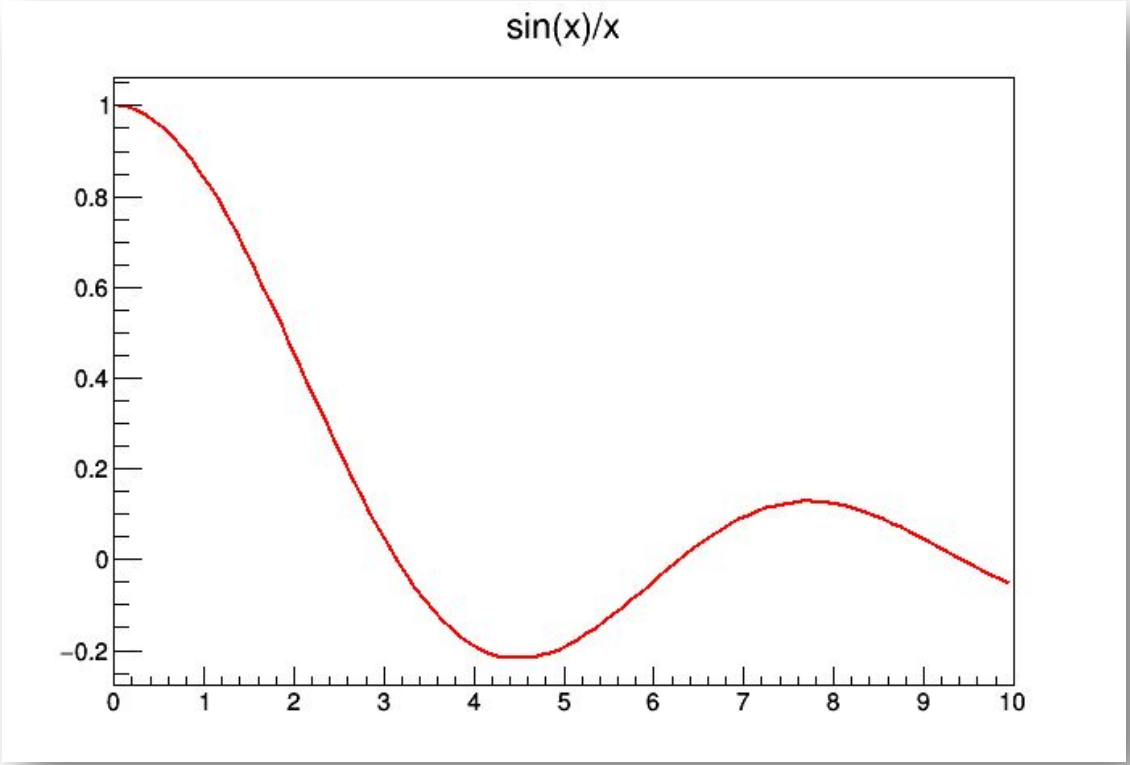
- An extended version of this example is the definition of a function with parameters:

```
>>> f2 = ROOT.TF1("f2","[0]*sin([1]*x)/x",0.,10.)
>>> f2.SetParameters(1,1)
>>> f2.Draw()
```

Python

Try it!

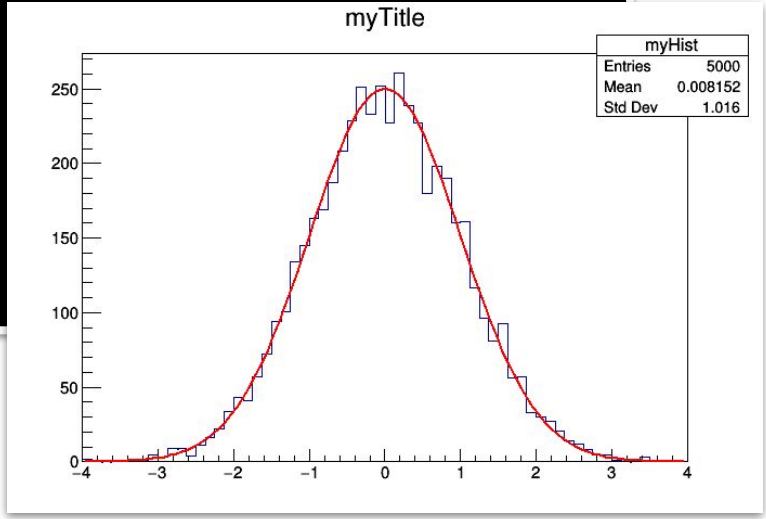
ROOT as a Function Plotter



Another Example

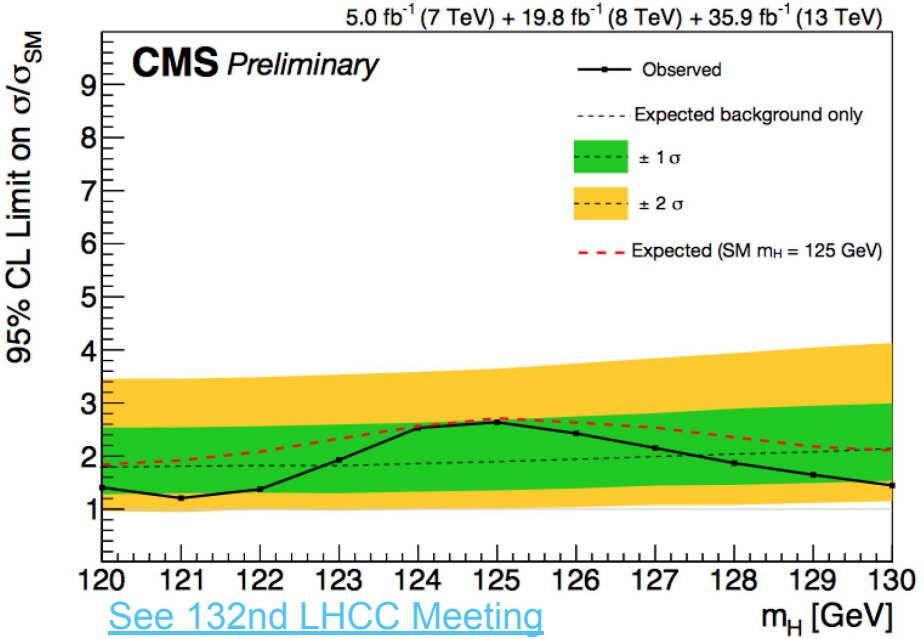
C++

```
root [0] TH1D h("myHist", "myTitle", 64, -4, 4)
root [1] h.FillRandom("gaus")
root [2] h.Draw()
root [3] TF1 f("g", "gaus", -8, 8)
root [4] f.SetParameters(250, 0, 1)
root [5] f.Draw("Same")
```



Graphs

- Display points and errors
- Not possible to calculate momenta
- Not a data reduction mechanism
- **Fundamental to display trends**

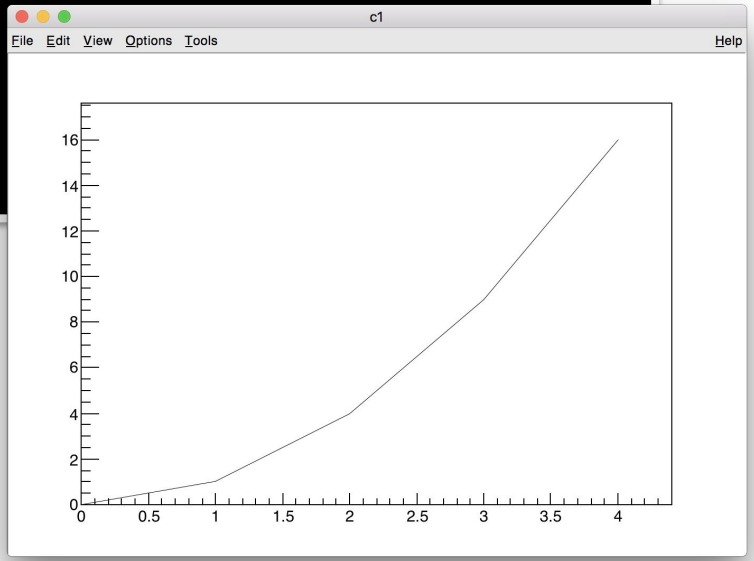


See 132nd LHCC Meeting

My First Graph

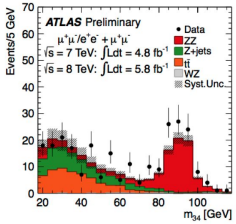
```
>>> g = ROOT.TGraph()
>>> for i in range(5): g.SetPoint(i,i,i*i)
>>> g.Draw("APL")
```

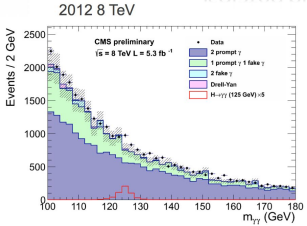
Python

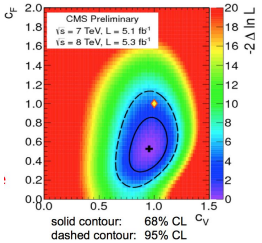


Drawing Options Documentation

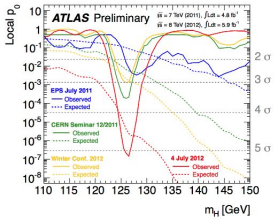
- See the documentation of the THistPainter class for all possible options on drawing histograms

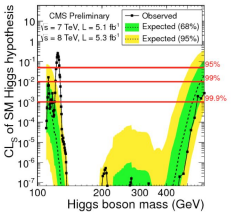


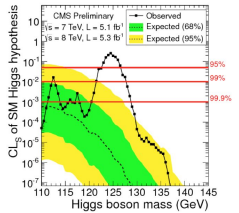




- See the documentation of TGraphPainter for the Graph drawing options







Time for Exercises - C++

https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/HistogramsGraphsFunctions

Time For Exercises - Python

https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/PythonInterface

- In order to run the exercises:
  - Use the Python prompt

```
> python
>>>
```

  - Run a Python script

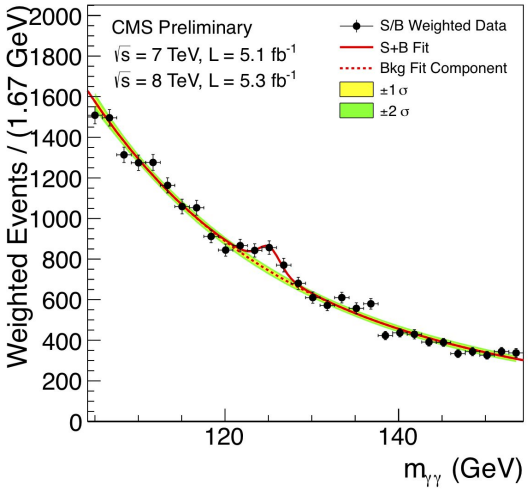
```
> python -i myscript.py
```

  - Use SWAN
    - Create a canvas before drawing: `c = ROOT.TCanvas()`
    - Run `c.Draw()` at the end to see the plot

Parameter Estimation and Fitting

What is Fitting ?

- Estimate parameters of a hypothetical distribution from the observed data distribution
  - y = f (x | θ) is the fit model function
- Find the best estimate of the parameters θ assuming f (x | θ)
- Both Likelihood and Chi2 fitting are supported in ROOT



Example

Higgs → γγ spectrum

We can fit for:

- the expected number of Higgs events
- the Higgs mass

Fitting in ROOT:

- Create first a parametric function object, **`TF1`**, which represents our model
  - need to set the initial values of the function parameters.
- Fit the data object (Histogram or Graph):
  - Call the **`Fit`** method passing the function object
  - various options are possible (see the **`TH1::Fit`** documentation)
- Examine result:
  - get parameter values, uncertainties, correlation
  - get fit quality estimation
- The resulting fit function is also drawn automatically on top of the Histogram or the Graph when calling **`TH1::Fit`** or **`TGraph::Fit`**

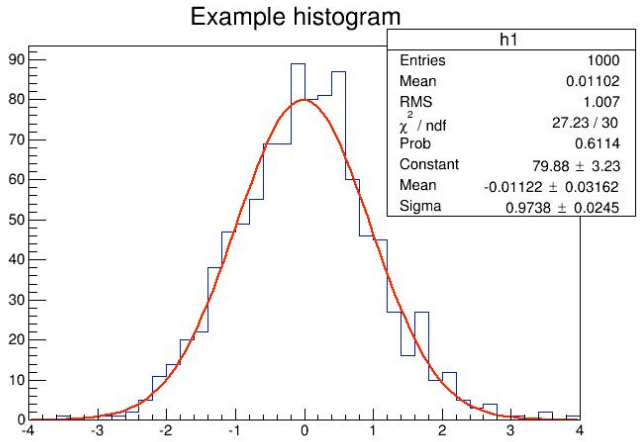
Fitting Histograms

- We have a histogram, h1, and we want to fit it:

```
root [0] TF1 f1("f1","gaus");
root [1] h1.Fit(&f1);
 FCN=27.2252 FROM MIGRAD    STATUS=CONVERGED      60 CALLS          61 TOTAL
                     EDM=1.12393e-07    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                                   STEP         FIRST
  NO.   NAME      VALUE            ERROR          SIZE      DERIVATIVE
   1  Constant     7.98760e+01   3.22882e+00   6.64363e-03  -1.55477e-05
   2  Mean        -1.12183e-02   3.16223e-02   8.18642e-05  -1.49026e-02
   3  Sigma        9.73840e-01   2.44738e-02   1.69250e-05  -5.41154e-03
```

For displaying the fit parameters:

```
gStyle->SetOptFit(1111);
```



Creating the Fit Function

- How to create the parametric function object (**TF1**) :
  - we can write formula expressions using functions:

```
TF1 f1("f1","[0]*TMath::Gaus(x,[1],[2])");
```

    - we can use the available functions in ROOT library and stl
    - **[0],[1],[2] indicate the parameters.**
    - We could also use meaningful names, like [a],[mean],[sigma]
  - There are pre-defined functions

```
TF1("f1","gaus");
```

  - pre-defined functions available: *gaus, expo, landau, breitwigner,crystal_ball,pol{0,1..,N}, cheb{0,1,...10},xygaus,,bigaus*
    - see full list in the documentation of TH1::Fit(), and also in the ROOT reference doc for TFormula

Building More Complex Functions

- Any C++ object (functor) implementing

```
double operator() (double *x, double *p)
```

```
struct Function {
     double operator() (double *x, double *p){
       return p[0]*TMath::Gaus(x[0],p[1],p[2]);
     }
};

Function f;
TF1 f1("f1",f,xmin,xmax,npar);
```

- also a lambda function (with Cling and C++-11)

```
TF1 f1("f1",[](double *x, double *p){return p[0]*x[0];},0,10,1);
```

a lambda can be used also as a string expression, which will be JIT'ed by CLING

```
TF1 f1("f1","[](double *x, double *p){return p[0]*x[0];}",0,10,1);
```

Functionality provided by TFormula

TFormula is based on Cling. Additional functionality provided:

- better parameter definition
  - `TF1("f1","gaus(x, [Constant],[Mean],[Sigma])");`
- function composition by concatenating expressions
  - `TF1 fs("sigma","[0]*x+[1]");`
  - `TF1 f1("f1","gaus(x,[C],[Mean],sigma(x,[A],[B]))");`
- normalized sum for component fitting
  - `TF1 model("model","NSUM(expo, gaus)"`
- convolutions
  - `TF1 voigt("voigt", "CONV(breitwiegner, gaus)", xmin, xmax);`
- can define vectorized functions for faster fitting and evaluation
  - see vectorizedFit tutorial
- support for auto-differentiation (automatic generation of gradient and Hessian)

Fitting Options

- Likelihood fit for histograms
  - option "L" for count histograms;
  - option "WL" in case of weighted counts.
- Default is chi-square with observed errors (and skipping empty bins)
  - option "P" for Pearson chi-square

    expected errors, and including empty bins
- Use integral function of the function in bin
- Compute MINOS errors : option "E"

```
h1->Fit("gaus","L");
```

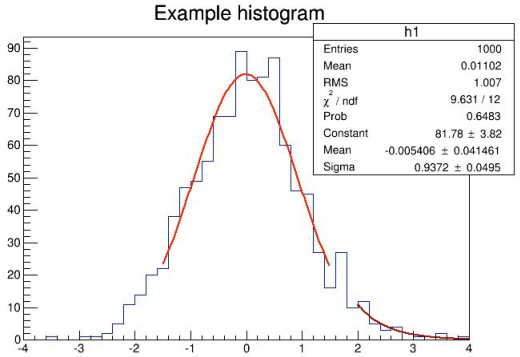
```
h1->Fit("gaus","LW");
```

```
h1->Fit("gaus","P");
```

```
h1->Fit("gaus","L I");
```

```
h1->Fit("gaus","L E");
```

Some More Fitting Options



- Fitting in a Range
  - `h1->Fit("gaus","","",-1.5,1.5);`
- For doing several fits
  - `h1->Fit("expo","+","",2.,4);`
- Quiet / Verbose: option "Q"/"V"
  - `h1->Fit("gaus","V");`
- Avoid storing and drawing fit function (useful when fitting many times)
  - `h1->Fit("gaus","L N 0");`
- Save result of the fit, option "S"
  - ```
    auto result = h1->Fit("gaus","L S");
    result->Print("V");
    ```

All fitting options documented in reference guide or User Guide (Fitting Histogram chapter)
    ```

Minimization

- The fit is done by minimizing the least-square or likelihood function.
- A direct solution exists only in case of linear fitting
  - it is done automatically in such cases (e.g fitting polynomials).
- Otherwise an iterative algorithm is used:
  - Minuit is the minimization algorithm used by default
    - ROOT provides two implementations: Minuit and Minuit2
    - other algorithms exists: Fumili, or minimizers based on GSL, genetic and simulated annealing algorithms
  - To change the minimizer:

```
ROOT::Math::MinimizerOptions::SetDefaultMinimizer("Minuit2");
```

  - Other commands are also available to control the minimization:

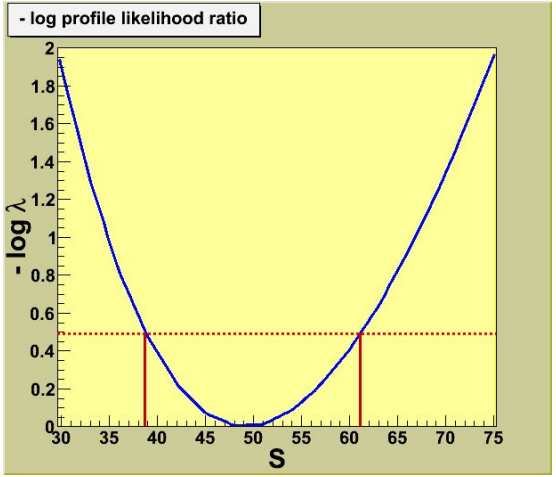
```
ROOT::Math::MinimizerOptions::SetDefaultTolerance(1.E-6);
```

Parameter Errors

Errors returned by the fit are computed from the second derivatives of the log-likelihood function

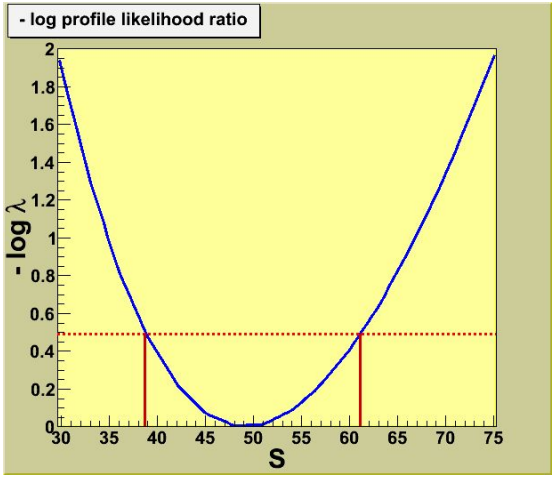
- Assume the negative log-likelihood function is a parabola around minimum
- This is true asymptotically and in this case the parameter estimates are also normally distributed.
- The estimated correlation matrix is then:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) = \left[\left(-\frac{\partial^2 \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1} = \mathbf{H}^{-1}$$



Parameter Errors

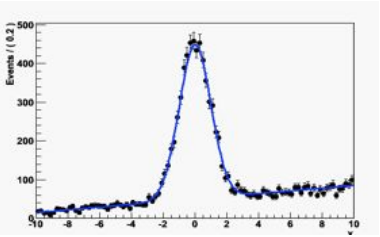
- A better approximation to estimate the confidence level of the parameter is to use directly the log-likelihood function and look at the difference from the minimum.
  - Method of Minuit/Minos (Fit option "E")
    - obtain a confidence interval which is in general not symmetric around the best parameter estimate




```
auto r = h1->Fit(f1,"E S");
r->LowerError(par_number);
r->UpperError(par_number);
```

RooFit: ROOT toolkit for complex fitting

- ROOT fitting can handle complicated functions…
  - …but requires writing much code when fitting using complex models
- RooFit provides functionality for building fitting models
  - complex model building from standard components
    - composition with addition product and convolution
- Fitting requires often **Normalization** of pdfs
  - not always trivial to perform and RooFit does it automatically
- RooFit provides also
  - **MC data generation** from model
  - advance **visualization** of fitting results
  - **simultaneous fit** to different data samples
  - full model description for **reusability**
  - **built-in optimization** for optimal computational performances
    - need to optimize code for acceptable performance in complex fits



- *For more info see the manual or the RooFit courses*

TMVA: Machine Learning in ROOT

- ROOT ML tools are provided in **TMVA** (*Toolkit for MultiVariate Analysis*)
- Provides a set of algorithms for standard HEP usage
  - Common interface to different algorithms with consistent evaluation and comparison
  - Capability for classification and regression
  - Embedded in ROOT: direct connection to input data (ROOT I/O)
  - Most popular algorithms are BDT and ANN (supporting recently some DL tools)
- Interfaces to external ML library :
  - e.g. to Python tools: scikit-learn, Tensorflow/Keras, PyTorch
- Fast inference system for Deep Learning models (SOFIE) and BDT
  - new tool togenerate code and easly evaluate ML models in ROOT that can be trained with other tools (e.g Keras, PyTorch) or xgboost
- *For more info see the manual*

Time for Exercises!

https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/Fitting

Reading and Writing Data

The ROOT File

- In ROOT, objects are written in files*, represented by **TFile** instances
- TFiles are *binary* and can be compressed (transparently for the user)
- **TFiles are self-descriptive**:
  - The information how to retrieve objects from a file is stored with the objects

* this is an understatement - we'll not go into the details in this course!

TFile in Action

```
TFile f("myfile.root", "RECREATE");
```

Option	Description
NEW or CREATE	Create a new file and open it for writing, if the file already exists the file is not opened.
RECREATE	Create a new file, if the file already exists it will be overwritten.
UPDATE	Open an existing file for writing. If no file exists, it is created.
READ	Open an existing file for reading (default).

TFile in Action: Writing

```
TFile f("file.root", "RECREATE");
TH1F h("h", "h", 64, 0, 8);
h.Write("h");
f.Close();
```

C++

- Write to a file
- Close the file and make sure the operation succeeded

```
> rootls -l file.root
TH1F  Jun 24 15:02 2022 h  "h"
```

TFile in Action: Reading

```
TFile f("file.root");
TH1F* h = f.Get<TH1F>("h");
h->Draw();
```

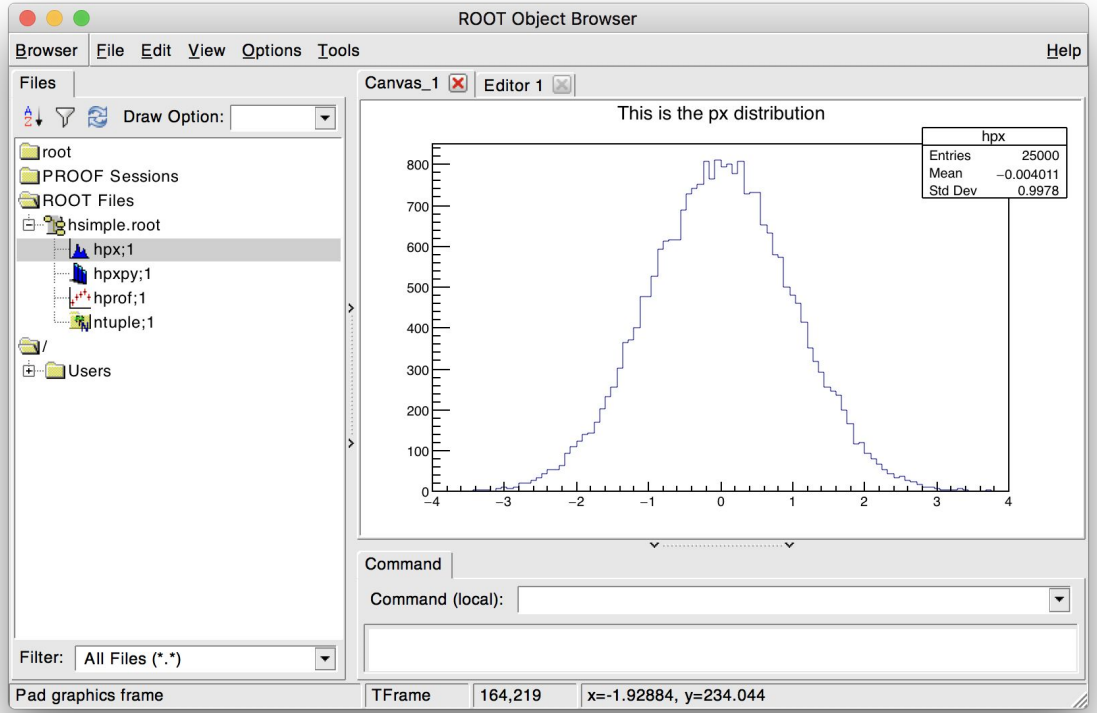
C++

Python

```
import ROOT
f = ROOT.TFile("file.root")
f.h.Draw()
```

Get the histogram by name!
Possible only in Python

Listing TFile Content



- ***TBrowser*** interactive tool

```
> root [0] TBrowser t
```

- ***rootls*** tool: list content
- ***TFile::ls()***: prints content
  - Great for interactive usage

Time For Exercises

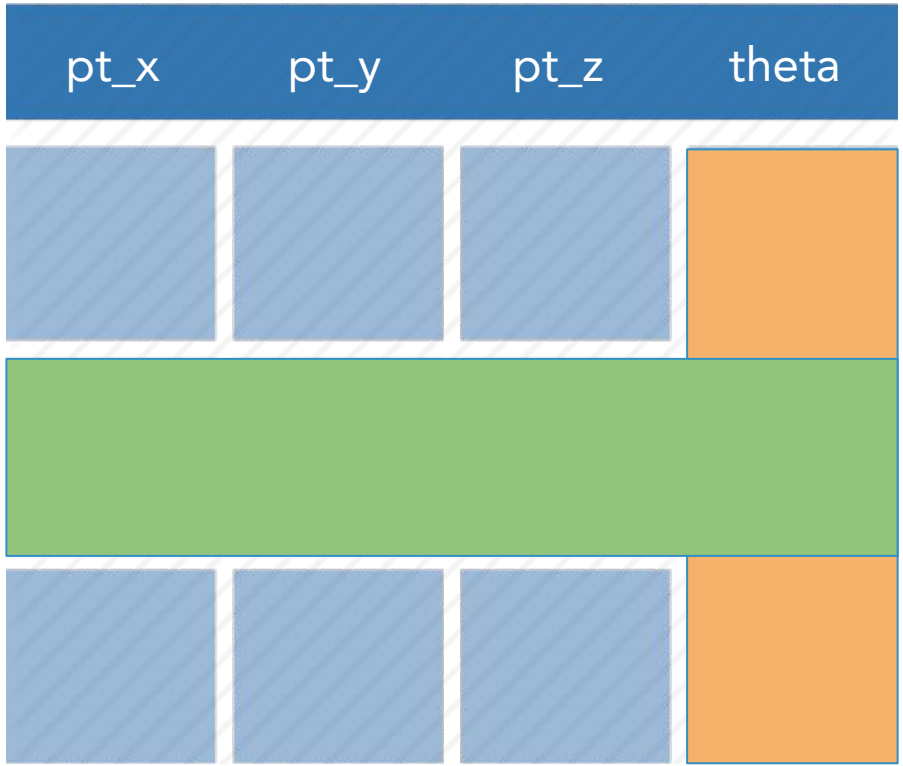
https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/WorkingWithFiles

The ROOT Columnar Format

Columns and Rows

- High Energy Physics: many statistically independent *collision events*
- Create an event class, serialise and write out N instances into a file?
  → No. Very inefficient!
- Organise the dataset in **columns**

Columnar Representation

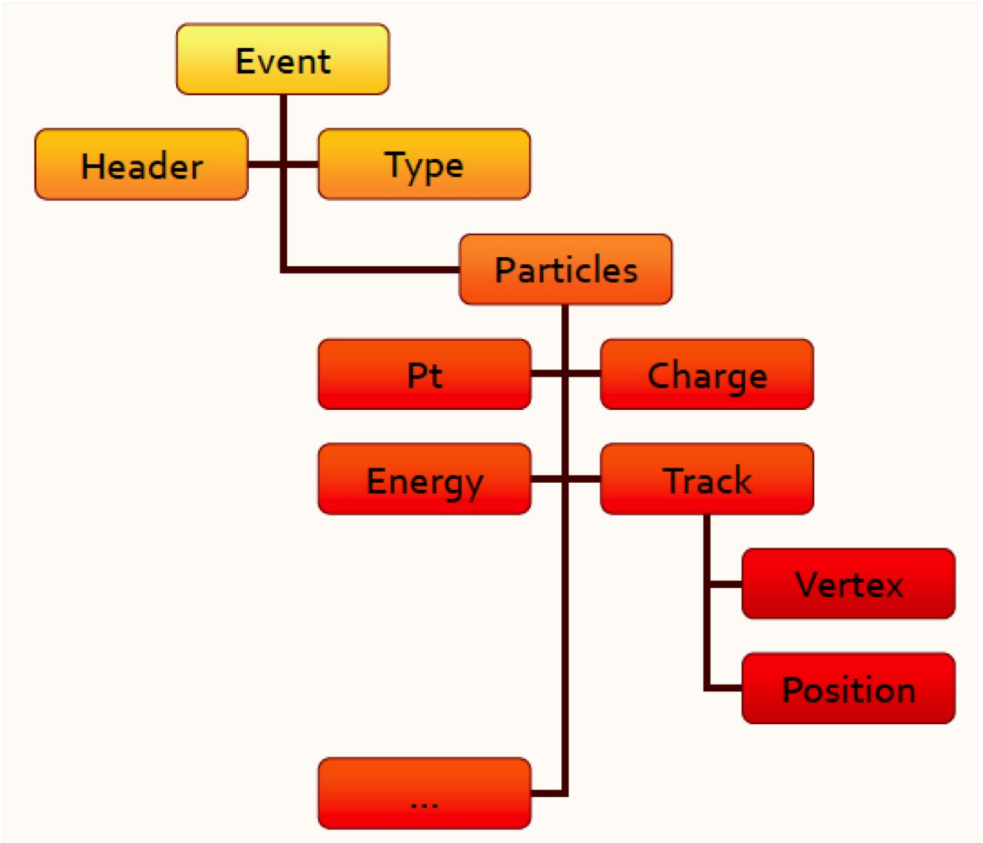


columns
or "branches"
can contain any kind
of c++ object

entries
or events
or rows

Relations Among Columns

x	y	z
-1.10228	-1.79939	4.452822
1.867178	-0.59662	3.842313
-0.52418	1.868521	3.766139
-0.38061	0.969128	1.084074
0.55	-0.21231	50281
-0.184	1.187305	.443902
0.20564	-0.770	0.635417
1.079222	0.32	1.271904
-0.27492	-	43 3.038899
2.047779	-	268 4.197329
-0.45868	4	2 2.293266
0.304731	0.884	0.875442
-0.712	-0.222	0.556881
-0.27	1.181767	470484
0.88	-0.65411	3209
-2.03555	0.527648	4.421883
-1.45905	-0.464	2.344113
1.230661	-0.00565	1.514559
		3.562347



The TTree

A columnar dataset in ROOT is represented by the class **TTree**:

- Also called *tree*, columns also called *branches*
- Columns can contain different types.
- **Support any type of object**
- One row per *entry* (or, in collider physics, *event*)

If just a **single number** per column is required, the simpler **TNtuple** *can* be used.

A modern and simple way to interact with ROOT datasets is to use `RDataFrame`

- Low-level interfaces to deal with datasets do exist but are beyond the scope of this course

RDataFrame: quick how-to

1. build a data-frame object by specifying your data-set
2. apply a series of transformations to your data
   - filter (e.g. apply some cuts) or
   - define new columns
3. apply actions to the transformed data to produce results (e.g. fill a histogram)

Simple Code Example

1. Build RDataFrame

```
ROOT::RDataFrame d("t", "f.root");
auto h = d.Filter("theta > 0").Histo1D("pt");
h->Draw();
```

2. Cut on theta

3. Fill histogram with pt

Filling multiple histograms

```
auto h1 = d.Filter("theta > 0").Histo1D("pt");
auto h2 = d.Filter("theta < 0").Histo1D("pt");
h1->Draw();          // event loop is run lazily once here
h2->Draw("SAME"); // no need to run loop again here
```

Book all your actions upfront. The first time a result is accessed, RDataFrame will fill all booked results.

More on histograms

```
auto h = d.Histo1D({"myName","Title;x",10,0.,1.},
                   "x");
```

You can specify a model histogram with

- a name and a title
- a predefined axis range

Here, the histogram is created with 10 bins ranging from 0 to 1, and the axis is labelled "x".

Define a new column

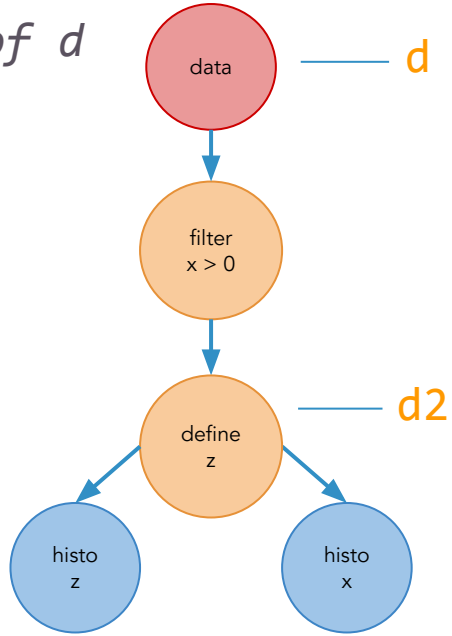
```
double m = d.Filter("x > y")
            .Define("z", "sqrt(x*x + y*y)")
            .Mean("z");
```

`Define` takes the name of the new column and its expression. Later you can use the new column as if it was present in your data.

Think of your analysis as data-flow

```
// d2 is a new data-frame, a transformed version of d
auto d2 = d.Filter("x > 0")
           .Define("z", "x*x + y*y");

// make multiple histograms out of it
auto hz = d2.Histo1D("z");
auto hx = d2.Histo1D("x");
```



You can store transformed data-frames in variables, then use them as you would use an RDataFrame.

Cutflow reports

```
d.Filter("x > 0", "xcut")
 .Filter("y < 2", "ycut");
d.Report()->Print();
```

```
// output
xcut       : pass=49         all=100        --   49.000 %
ycut       : pass=22         all=49         --   44.898 %
```

When called on the main RDF object, `Report` prints statistics for all filters *with a name*

Saving data to file

```
auto new_df = df.Filter("x > 0")
                .Define("z", "sqrt(x*x + y*y)")
                .Snapshot("tree", "newfile.root");
```

We filter the data, add a new column, and then save everything to file. No boilerplate code at all.

Using callables instead of strings

Expert Feature

```
// define a c++11 lambda - an inline function - that checks "x>0"
auto IsPos = [](double x) { return x > 0.; };
// pass it to the filter together with a list of branch names
auto h = d.Filter(IsPos, {"theta"}).Histo1D("pt");
h->Draw();
```

any callable (function, lambda, functor class) can be used as a filter, as long as it returns a boolean

RDataFrame: declarative analyses

```
ROOT::RDataFrame d("treename", "file.root");
auto h = d.Filter(IsGoodEntry, {"x","y"})
          .Histo1D("x");
```

A function taking 2 values in input, returns a boolean

- full control over *the analysis*
- no boilerplate
- common tasks are already implemented?
- parallelization is not trivial?

RDataFrame: parallelism

```
ROOT::EnableImplicitMT();
ROOT::RDataFrame d("treename", "file.root");
auto h = d.Filter(IsGoodEntry, {"x","y"})
          .Histo1D("x");
```

- full control over *the analysis*
- no boilerplate
- common tasks are already implemented
- parallelization is automatic

C++ / JIT / PyROOT

C++ and just-in-time compiled code

```
d.Filter("th > 0").Snapshot("t","f.root","pt*");
```

PyROOT -- just leave out the `;`

```
d.Filter("th > 0").Snapshot("t","f.root","pt*")
```

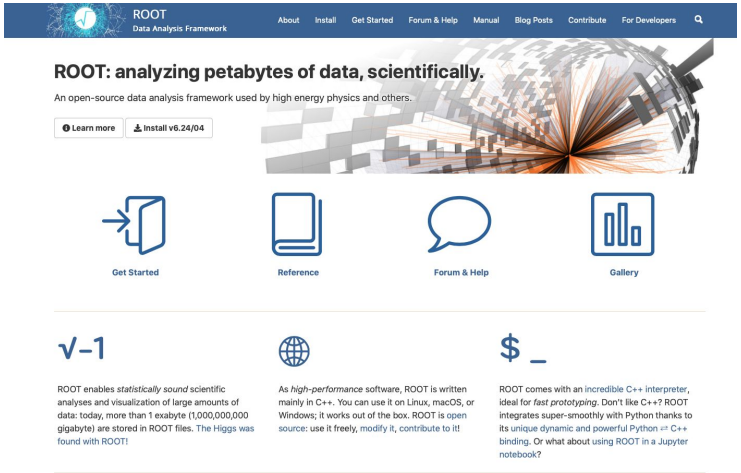
Time For Exercises

https://github.com/root-project/training/tree/master/SummerStudentCourse/2023/Exercises/WorkingWithColumnarData

Wrap up

ROOT Links

- Web page: https://root.cern
- Documentation: https://root.cern.ch/doc/master/index.html
- Forum: https://root-forum.cern.ch
- github: https://github.com/root-project
- Email: root-dev@cern.ch
- @root-project
- https://www.linkedin.com/groups/1826455



Backup

X11 forwarding on Windows

- You will need to install Xming to be able to use ROOT graphics over a SSH connection.
  You can get it here.
- After install, click on the XLaunch shortcut on your Start menu. Be sure to use the "Multiple windows" configuration, with "Display number" set to **0** and that "No Access Control" is checked.
- `ssh` needs to see a value for the `DISPLAY` environment variable. On a `cmd.exe` session:

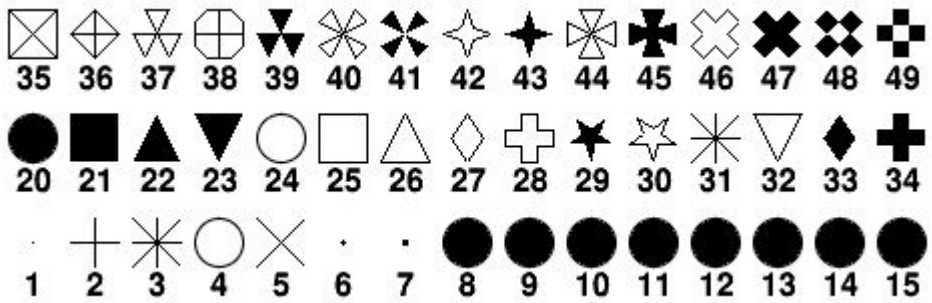
```
C:\...\> set DISPLAY=:0
C:\...\> ssh -XY <username>@lxplus.cern.ch
```

Creating a Nice Plot: Survival Kit

Credit: https://www.swissknifeshop.com

The Markers

From the TAttMarker documentation:
https://root.cern.ch/doc/master/classTAttMarker.html

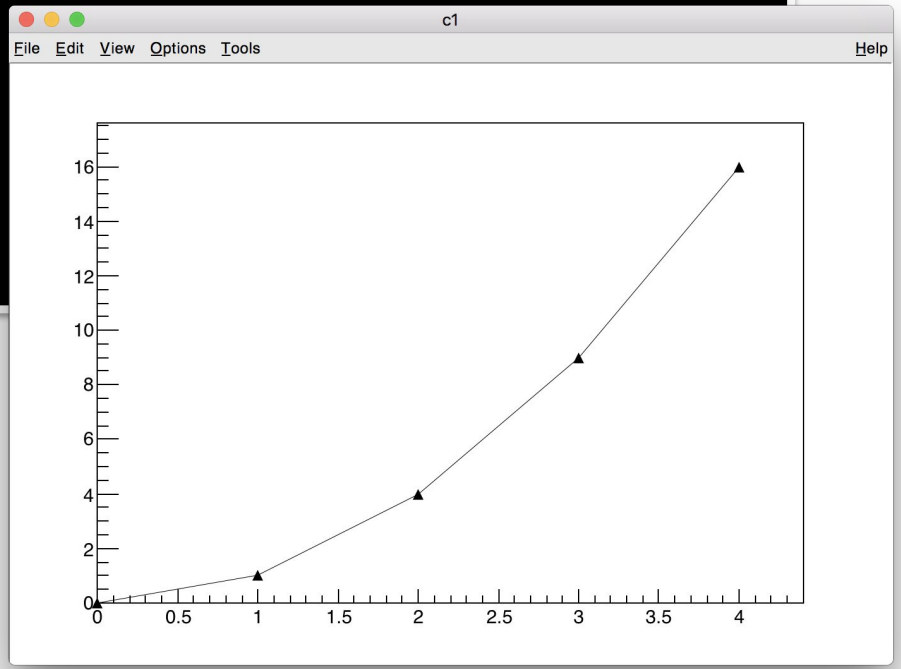



```
kDot=1, kPlus, kStar, kCircle=4, kMultiply=5,
kFullDotSmall=6, kFullDotMedium=7, kFullDotLarge=8,
kFullCircle=20, kFullSquare=21, kFullTriangleUp=22,
kFullTriangleDown=23, kOpenCircle=24, kOpenSquare=25,
kOpenTriangleUp=26, kOpenDiamond=27, kOpenCross=28,
kFullStar=29, kOpenStar=30, kOpenTriangleDown=32,
kFullDiamond=33, kFullCross=34 etc…
```

Also available through more friendly names ☺

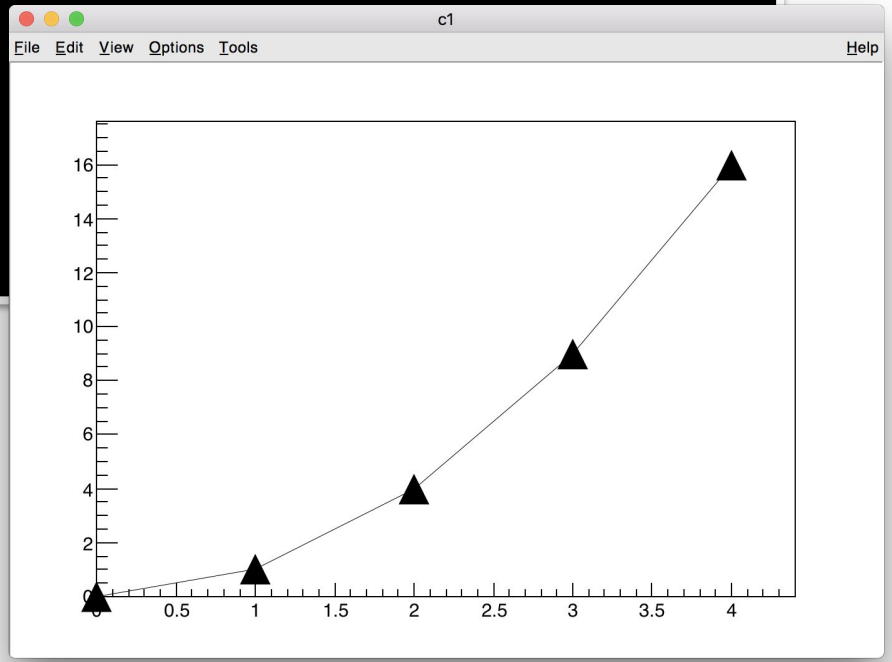
My First Graph

```
root [3] g.SetMarkerStyle(kFullTriangleUp)
```



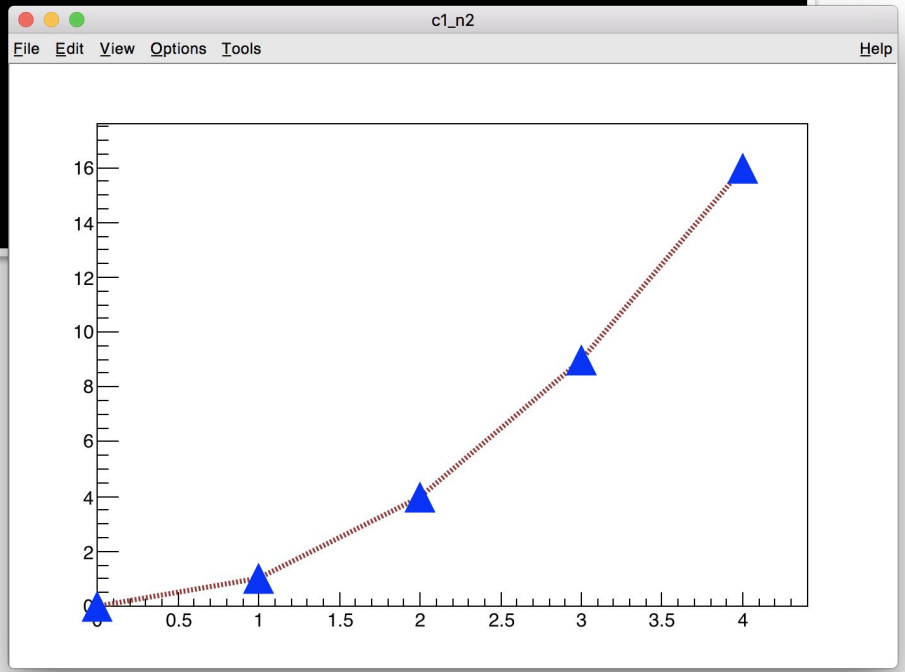
My First Graph

```
root [3] g.SetMarkerStyle(kFullTriangleUp)
root [4] g.SetMarkerSize(3)
```



My First Graph

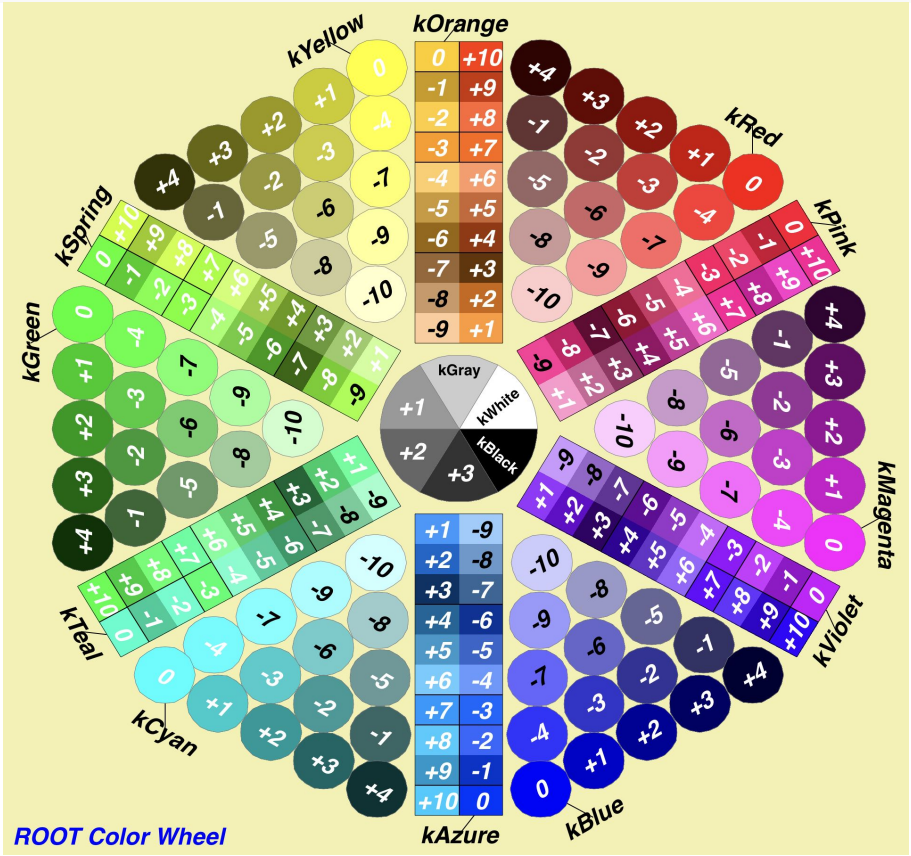
```
root [5] g.SetMarkerColor(kAzure)
root [6] g.SetLineColor(kRed - 2)
root [7] g.SetLineWidth(2)
root [8] g.SetLineStyle(3)
```



Question:
How do you find information on line styles?

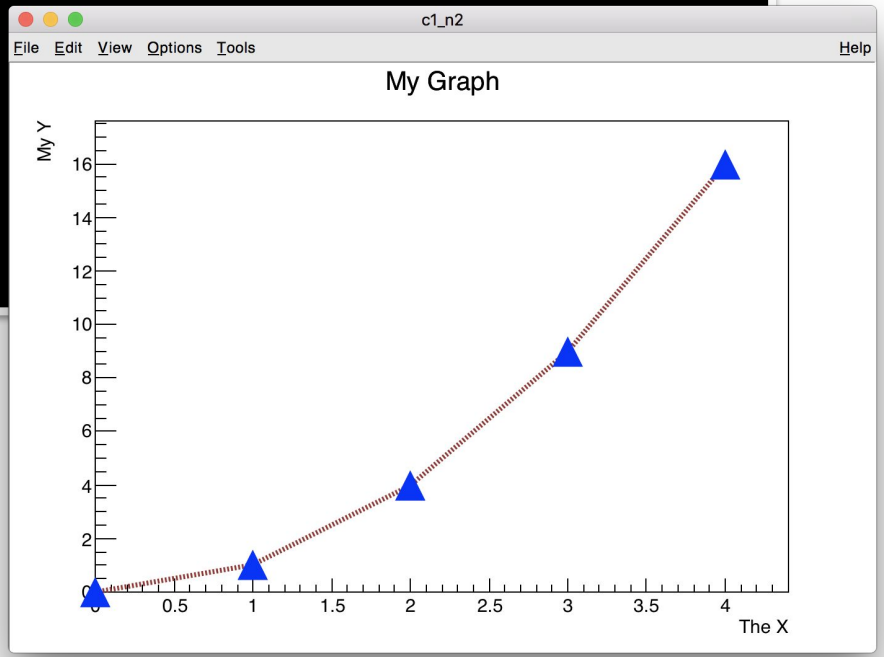
See TAttLine documentation

The Colors (TColorWheel)



My First Graph

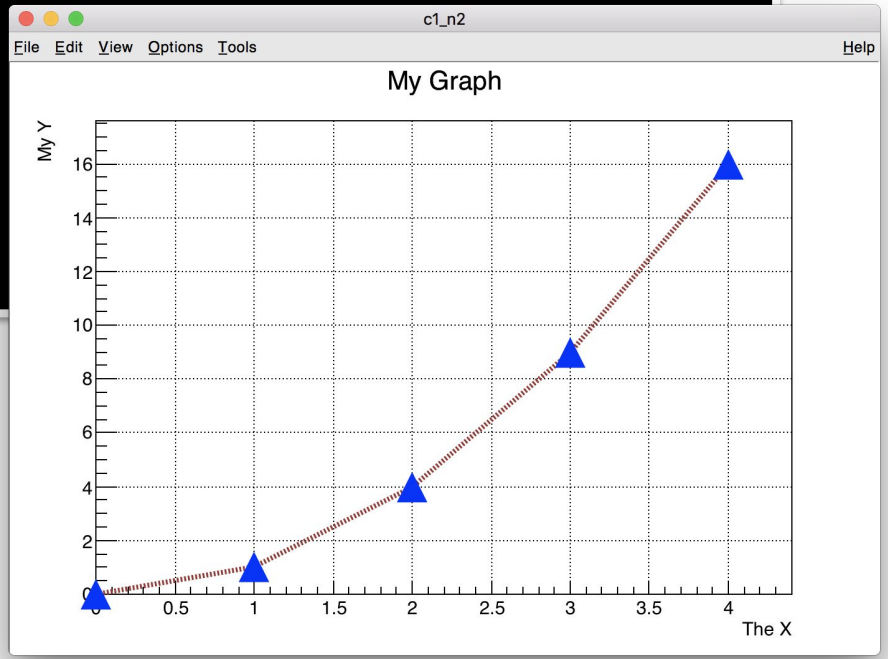
```
root [9] g.SetTitle("My Graph;The X;My Y")
```



Shortcut:
Directly set titles for x and y axis.

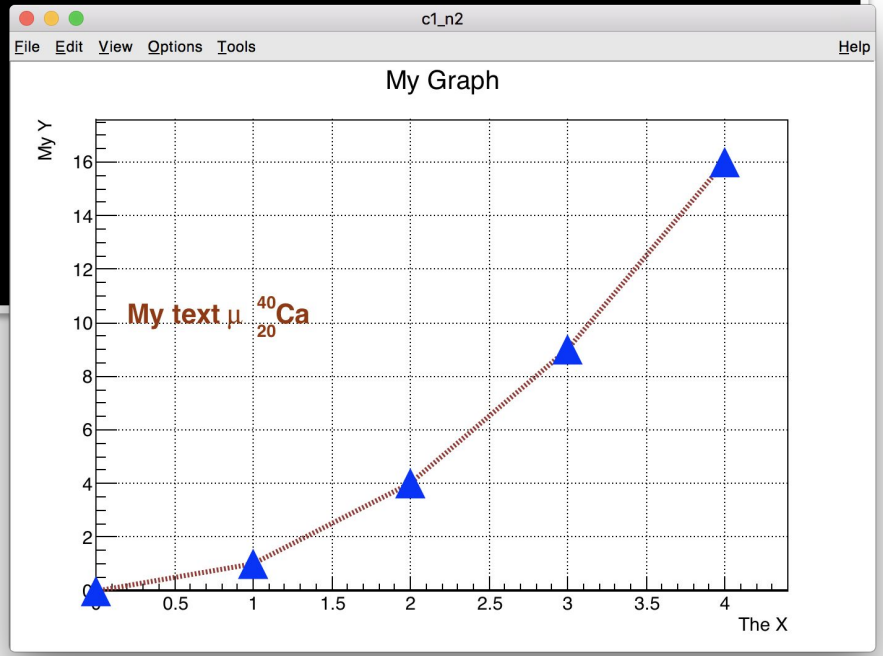
My First Graph

```
root [10] gPad->SetGrid()
```



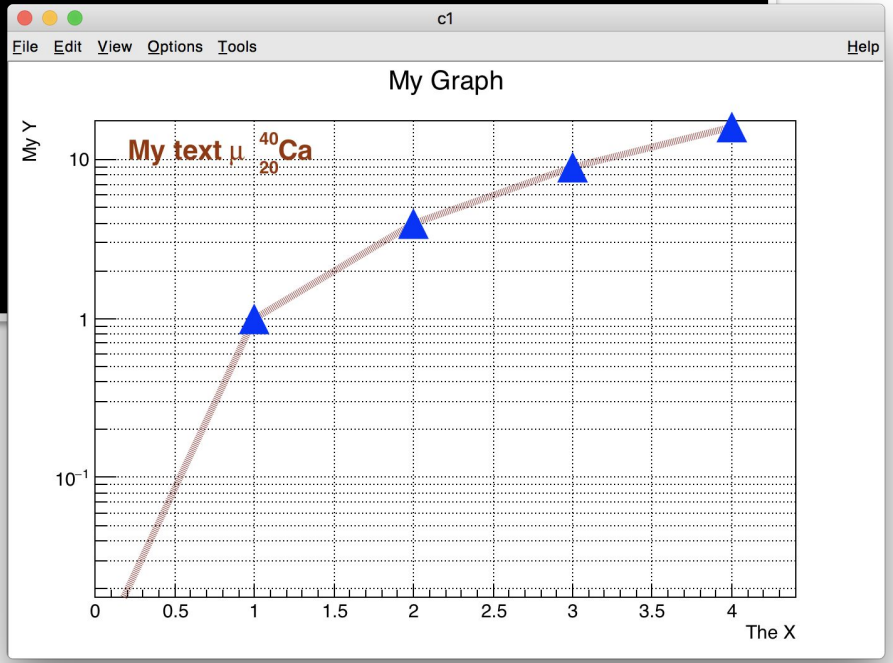
My First Graph

```
root [10] auto txt = "#color[804]{My text #mu {}^{40}_{20}Ca}"
root [11] TLatex l(.2, 10, txt)
root [12] l.Draw()
```



My First Graph

```
root [13] gPad->SetLogy();
```



Interactivity

```
root [0] #include "a.h"
root [1] A o("ThisName"); o.printName()
ThisName
root [1] dummy()
(int) 42
```

a.h

```
# include <iostream>
class A {
public:
  A(const char* n) : m_name(n) {}
  void printName() { std::cout << m_name << std::endl; }
private:
  const std::string m_name;
};

int dummy() { return 42; }
```

Unnamed ROOT Macros

- Macros can also be defined with no name
- Cannot be called as functions!

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