
**Summer Student Project**

# Web-based pre-Analysis Tools

*Student:* Tetiana Moskalets
V. N. Karazin Kharkiv National University

*Supervisors:*
Dr. Arturo Rodolfo Sanchez Pineda
University of Naples "Federico II" and INFN

Dr. Francesco Conventi
University of Naples "Parthenope" and INFN

CERN Summer Student Programme
July-August 2014

**Abstract**

The project consists in the initial development of a web based and cloud computing services to allow students and researches to perform fast and very useful cut-based pre-analysis on a browser, using real data and official Monte-Carlo simulations (MC). Several tools are considered: ROOT files filter, JavaScript Multivariable Cross-Filter, JavaScript ROOT browser and JavaScript Scatter-Matrix Libraries.

Preliminary but satisfactory results have been deployed online for test and future upgrades.

# Contents

# 1 Introduction

The cut-optimization procedure is a very time consuming process during all the performance of a physics analysis. Immersing into the coding procedure, researcher often is not able to think about the physics of the data until a plot is produced. The idea is to simplify the beginning of the analysis and allow researcher to follow the behavior of the variables while making cuts in an "on the flight" regime. Also very important point is to automatize some simple but necessary actions, for example, "filtering" of ROOT [1] files.

Thus, the main objectives of the present project are:

- Creation of scripts to produce the JavaScript [2] input tables in lxplus [3] environment.
- Development of the different Web-based analysis tools to be deployed.
- Online visualisation of the results.

Nowadays, more powerful computers and faster internet connections allow the analysis of large sets of data in ways that were impossible just few years ago. This permits simple but useful analysis with a performance compatible with a ROOT session in your laptop or in lxplus.

In order to show the advantages of our online tools, especially JavaScript Multivariable Cross-Filter [4], and the possibility to use them for a real analysis, we took MC samples for $H \rightarrow ZZ \rightarrow llqq$ and background processes and applied the simplified version of cuts from the $H \rightarrow ZZ \rightarrow llqq$ analysis [5] in ATLAS [6] in the website form. In this exercise, the numbers of remaining evens per each of the cuts are very close to those from the official analysis (see Table.1)

Apart from locating such scripts on single web-pages, it is also possible to include them in TWiki pages [7], that are widely used for sharing updates, analysis strategies, current studies situation, preliminary results, cutflow comparisons, and so forth into the collaboration.

All this, together with commercial analytical services (i.e. provided by Google [8]) can be used not to produce a final paper plot, but to create a faster way to exchange ideas, tutorials, classes and real useful information for the ATLAS analyses for the next data taking at LHC (so called Run II). Where one of the aspects to take in consideration is the huge amount of data, analyses, tunings and techniques to put under review and test for the coming years.

# 2 JavaScript ROOT Browser

Let's start with a consideration of a quite useful tool, that, however, has been already created by ROOT developers. It can be used for browsing the ROOT files: histograms, canvases, graphs. One can open a plot from a file, located on the website (from the menu) or enter full url address of other file, if it is in the same domain.

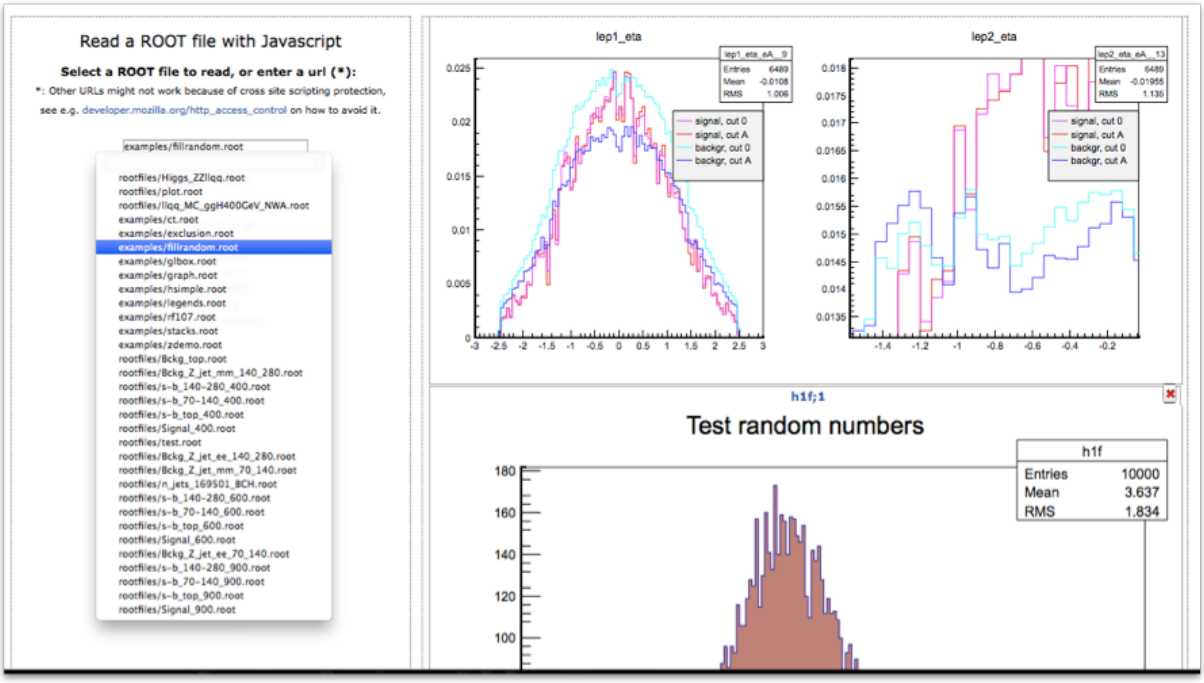


Figure 1: JavaScript ROOT browser with several plots opened, plot on the top and right is zoomed in.

Different plots can be opened at the same time for comparison. Also zooming is allowed by selecting the area inside the pad. For example, see the Fig.1: two plots on the top are normalized plots of the $H \rightarrow ZZ \rightarrow llqq$ signal and Top background MC samples and plot in the bottom is one of the examples created by the ROOT team.

Here you can find our web-page JavaScript ROOT Browser implementation:
http://arturos.web.cern.ch/arturos/html/

# 3 ROOT files filter

For a particular study in some analyses we don't really need the whole ROOT file, containing all saved variables. Usually only part of it is needed. For that reason, we created a simple script to automate the "filtering" process, that produces a table, containing only chosen information in columns. The script needs the input "in.txt" file (see Fig.2), containing the following lines: ROOT files to be filtered (each name in a new line), and variables to be included in the new ROOT file (each variable in a new line).

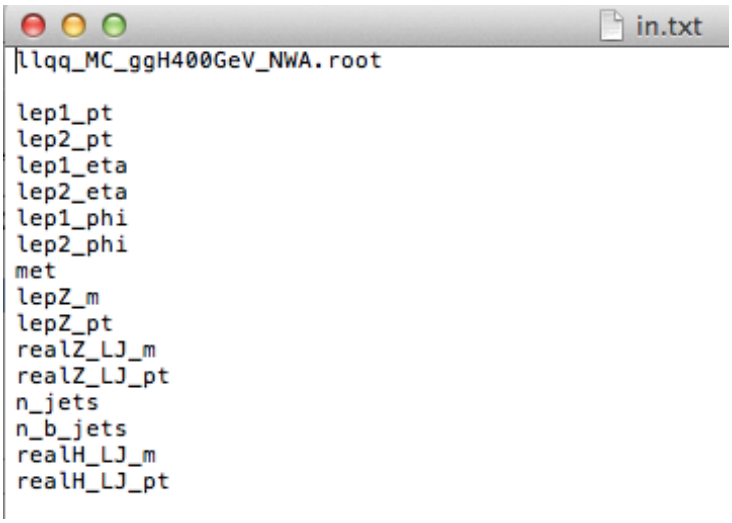


Figure 2: Example of the input for the ROOT files filter.

The output is filtered versions of all the ROOT files, named as "[old_root_file_name]_cut.root". And also the script creates ".csv" files, corresponding to each filtered tree, named as the trees from the input listed ROOT files (see Fig.3).

Here is the script code:
https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/rootcut.c

# 4 JavaScript Scatter-Matrix Library

The custom tables, produced by ROOT files filter, can be used as an input for a JavaScript Scatter-Matrix Library that allows to produce scatter plots of all possible combinations of two variables inside the input table. Different samples can be plotted at the same time (see Fig.4). Besides studying the dependencies between all pairs of variables it is possible to get such plots (for all pairs with one variable) corresponding to all possible values of some additional variable (see Fig.4).

Each time the selection on the left side menu is done, all the plots are rebuilt automatically.

Here, as an example, the MC samples of the $H \rightarrow ZZ \rightarrow llqq$ ATLAS analysis are loaded into the ScatterPlot matrix:
http://arturos.web.cern.ch/arturos/scatter/demo/demo.html

# 5 JavaScript Multivariable Cross-Filter

Cross-Filter is a JavaScript library [4] developed for working with large sets of variables in a web browser. The Cross-Filter works fast enough to analyse data, containing more than a million events. We used this library to produce webpages with a set of histogram-like plots of variables, that are the mostly often used in analysis cutflows. Such web-pages are very convenient for pre-analysing, because cuts can be performed with the cursor and the behaviour of other variables

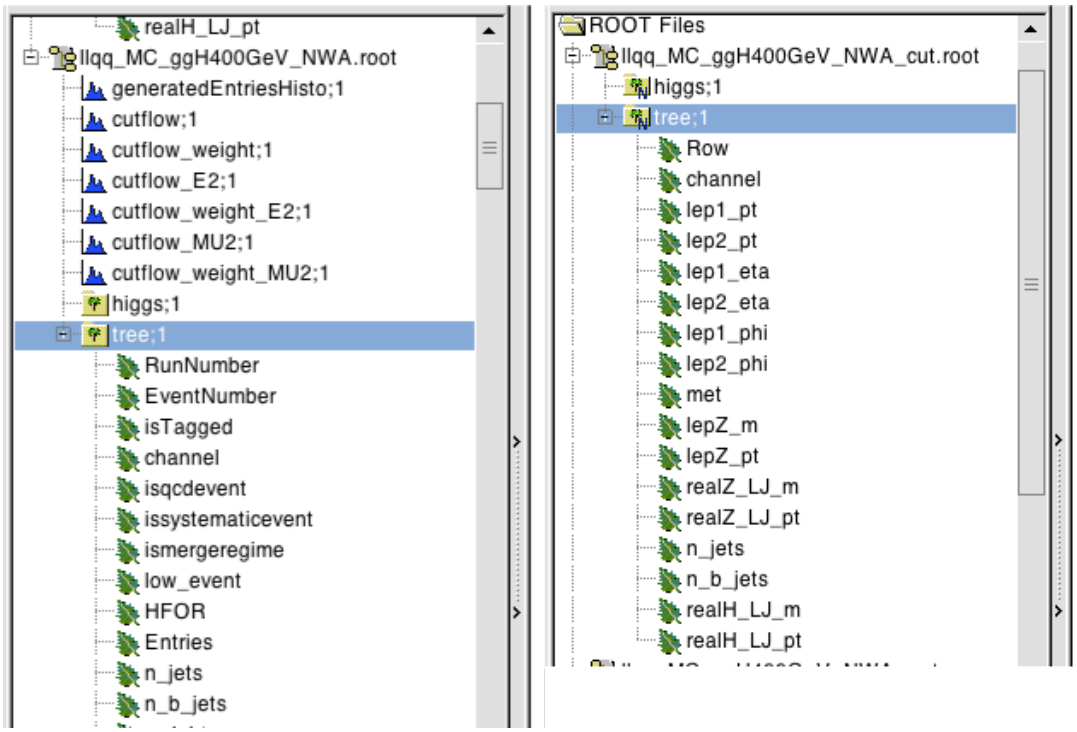


Figure 3: Structure of the initial ROOT file "llqq_MC_ggH400GeV_NWA.root" (left) and the filtered version "llqq_MC_ggH400GeV_NWA_cut.root", corresponding to the input information in Fig.(2) (right).

will be seen immediately, that significantly simplifies the procedure of finding the right range of cuts.

As a basis for our web-page we used the Multivariable Analyser example:
http://arturos.web.cern.ch/arturos/analyzer_v0/

As an example, we considered different signal and background MC samples for $H \rightarrow ZZ \rightarrow llqq$ analysis in order to check how large will be the possible differences in the number of events remaining after some basic cuts from official analysis [5] in the Cross-Filter web-page and those from the official cutflows. For example, for the signal sample llqq_MC_ggH400GeV_NWA (MC simulation for 400GeV Higgs, see Fig.5) we applied the following cuts:

- OppositeSign: The product of the charges of the two selected muons (lep_chargeproduct) should be equal to -1. (Not to be applied to Electron channel)
- TwoJets: The number of jets (n_jets) should be at least 2.
- DileptonMass: The invariant mass of the two leptons (lepZ_m) must be between $83GeV$ and $99GeV$.
- MET: The value of the Missing Transverse Energy (met) should be less than $60GeV$.
- NumTagJets[0-1-2]: Number of b-jets (n_b_jets) is equal to 0 (or 1 or 2).
  *Comment: the analysis is divided in three categories with respect to the number of b-jets,*

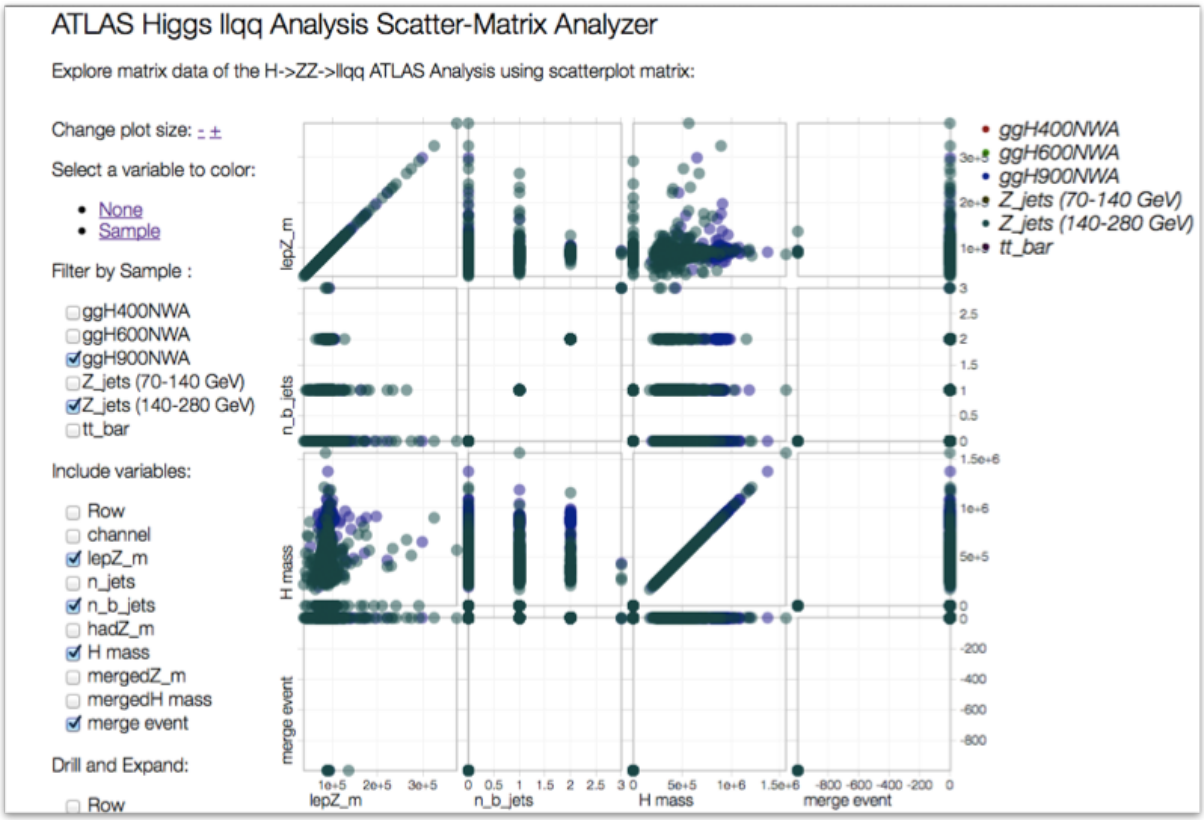


Figure 4: Scatter matrix analyser

*so this and the next 2 cuts are applied for each of the cases: 0 or 1 or 2 b-jets into the event.*

- PtLeadingJet[0-1-2]: The transverse momentum of the leading jet (realJ1_LJ_pt) should be greater than $45GeV$.
- DiJetMass[0-1-2]: The invariant mass of the two selected jets (realZ_LJ_m) should be between $70GeV$ and $105GeV$.

The Table.1 shows how the number of events changes after the cutflow for the electron channel.

The numbers after each of the cuts are rather close, this is quite good for the pre-analysis. Again, our objective is not to do the complete analysis, but to simplify and speed-up the estimation of the right cuts, that have to be clarified in the next stage with the usual coding procedure.

Here you can find the Cross-Filter web-pages for signal and background MC samples for $H \rightarrow ZZ \rightarrow llqq$ analysis processes:

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/llqq_MC_ggH400GeV.html

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/llqq_MC_ggH600GeV.html

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/llqq_MC_ggH900GeV.html

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/Top.html

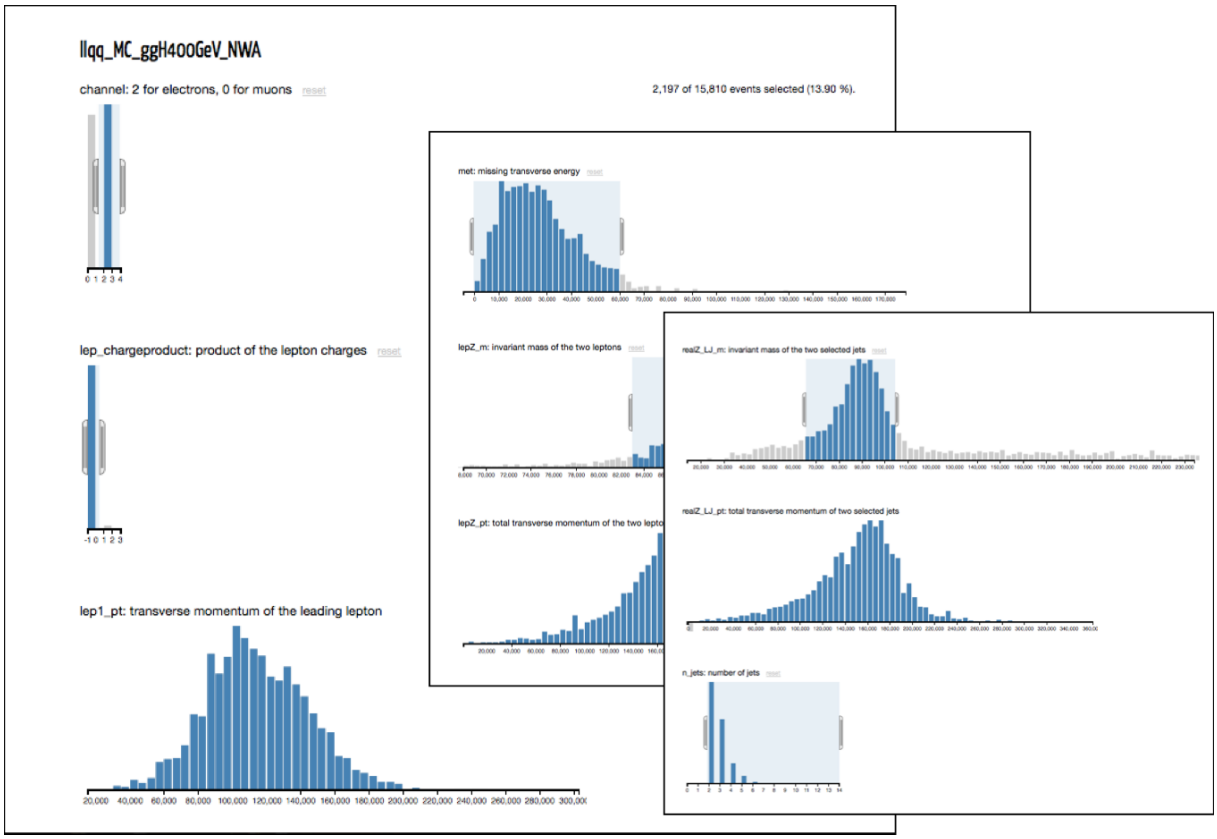


Figure 5: Preview of parts of Cross-Filter web-page with cuts applied.

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/Z_jet_70_140.html

https://twiki.cern.ch/twiki/pub/Main/SummerStudentProject2014Napoli/Z_jet_140_280.html

# 6 Conclusion

In the current project we considered several online tools, that can be used for a preliminary analysis. Part of them, such as JavaScript ROOT Browser and JavaScript Scatter-Matrix Libraries were already developed and we considered them as example tools, mostly, for data visualisation. Another tool, JavaScript Multivariable Cross-Filter, was rewritten from one of the examples using the Java Script Cross-Filter library. Cross-Filter is particularly useful for the pre-analysis since it is visual and allows to see the dynamics of the variables that researcher is working with. As an input dataset for the Cross-Filter we can use the output of the other script, created in this project, ROOT files filter, that takes chosen variables of ROOT tree and creates the new ROOT files and corresponding ".csv" tables.

Thus, the objectives posed in this project: creation of scripts to produce the web-analysis input tables, development and deploy of different Web-based analysis tools and online visualisation of the results are achieved.

| Cut | Official cutflow | Web-based pre-analysis | Difference (%) |
|---|---|---|---|
| OppositeSign | 8056 | 8056 | 0.00 |
| TwoJets | 7282 | 7282 | 0.00 |
| DileptonMass | 6089 | 6088 | 0.02 |
| MET | 5763 | 5758 | 0.09 |
| NumTagJets0, | 4263 | 4261 | 0.05 |
| NumTagJets1, | 973 | 972 | 0.10 |
| NumTagJets2 | 513 | 511 | 0.39 |
| PtLeadingJet0, | 4217 | 4197 | 0.48 |
| PtLeadingJet1, | 957 | 957 | 0.00 |
| PtLeadingJet2 | 512 | 506 | 1.18 |
| DiJetMass0, | 2182 | 2160 | 1.01 |
| DiJetMass1, | 438 | 439 | 0.29 |
| DiJetMass2 | 387 | 383 | 1.04 |

Table 1: Number of events remaining after applying the cuts from the official cutflow and in the Cross-Filter web-page for MC signal sample $H \rightarrow ZZ \rightarrow llqq$ for 400 GeV Higgs mass.

# References


[1] ROOT data analysis framework:
https://root.cern.ch/

[2] JavaScript programming language:
https://developer.mozilla.org/en-US/docs/Web/JavaScript

[3] LXPLUS service:
http://information-technology.web.cern.ch/services/lxplus-service

[4] Cross-Filter JavaScript library repository:
https://github.com/square/crossfilter

[5] $H \rightarrow ZZ \rightarrow llqq$ High Mass Analysis:
https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/
/HiggsllqqWinter2013#High_Mass_Analysis_Selection

[6] ATLAS Collaboration, JINST 3, S08003 (2008)

[7] TWikiSite:
https://twiki.cern.ch/

[8] Google Apps Script:
https://developers.google.com/apps-script/