Interactive Documentation: Statistics that Prove that you are Rooftiing Wrong

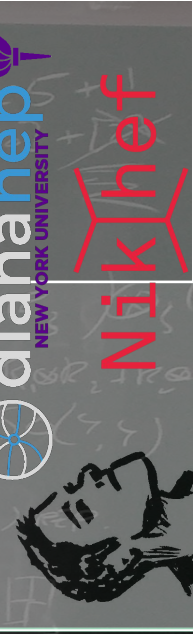
Interactive Documentation: Statistics that Prove that you are Rooftiing Wrong

Modeling with RooFit and RooStats

Vince Croft, Carsten Burgard and Wouter Verkerke

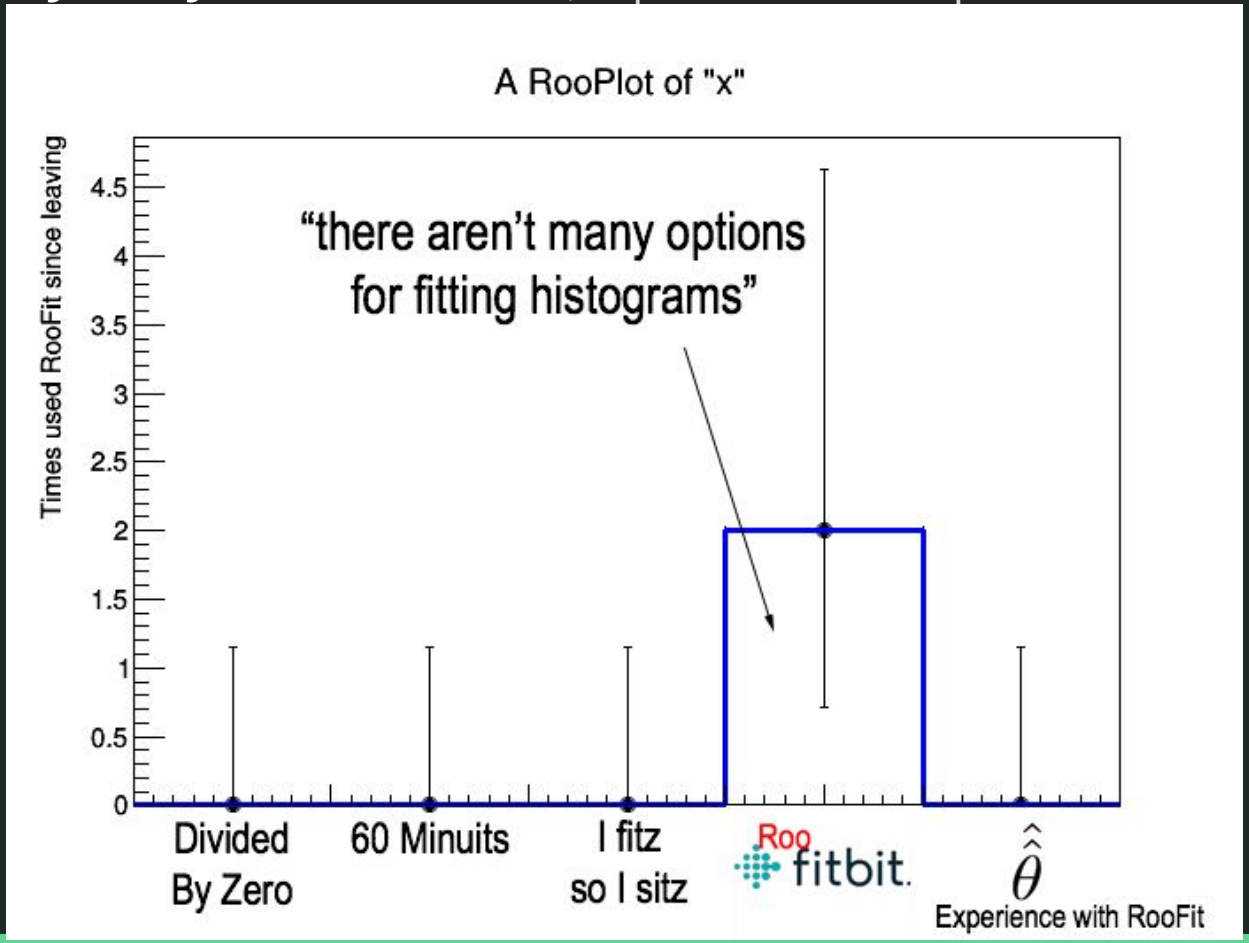
September 11, 2018 - ROOT Users Forum, Sarajevo

Vince Croft



Our Survey Says...

Former RooFit/RooStats users, now working in data science, report on their experiences and needs.



Our Survey Says...

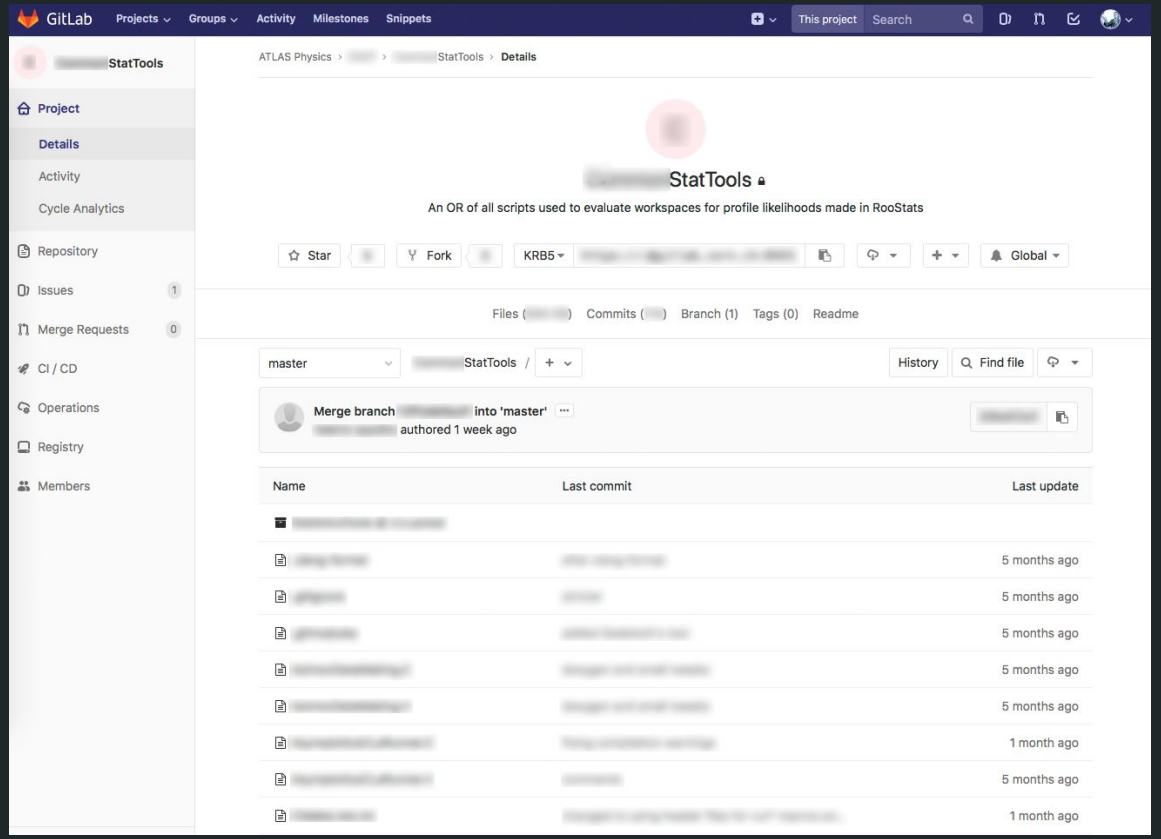
A Summary

Despite a high dependence of the majority of RooFit usage on the ability to build complex likelihoods with multiple regions and hundreds of regions, most users rely on others to provide an interface to the underlying statistical interpretation.

RooFit remains an excellent tool with which to model ‘Semi-parametric’ distributions and provide quantitative inference on these models. Education might be the only barrier stopping people from making use of this.

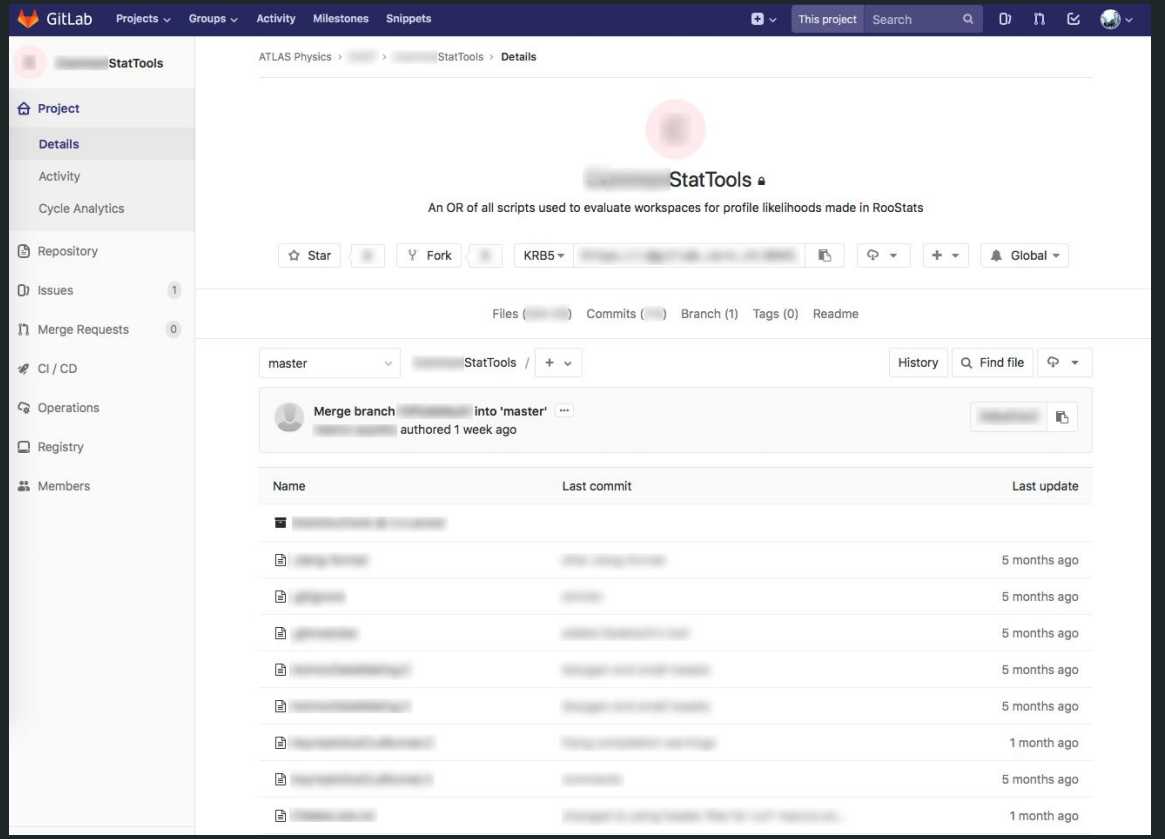
Anatomy of a Stats Framework

- Approx 30 files
- Represents hundreds of researcher hours to compile and thousands to write
- A conglomerations of *ALL* scripts needed for a researcher to go from analysis+data to result.
- Shows how RooStats tools are actually used...



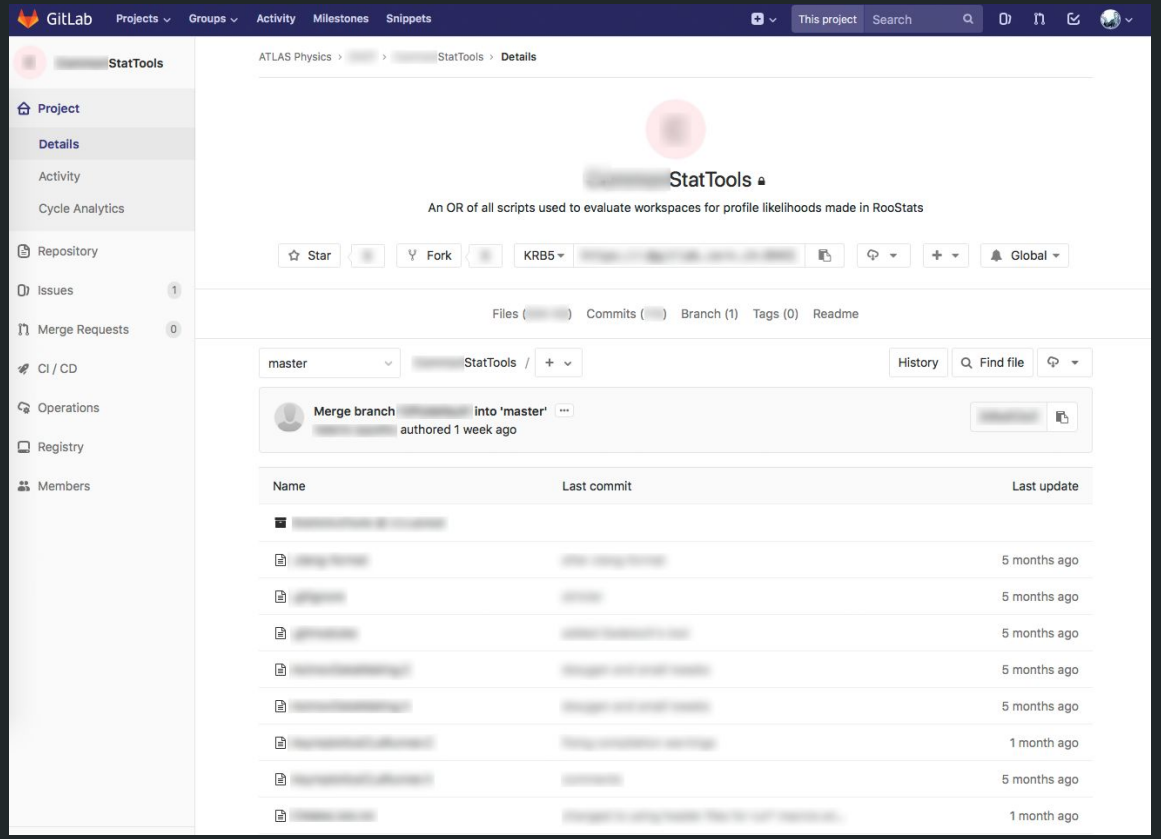
Family Feud/Fortunes

- Plumbing - 20 %
- Procedural macros - 15 %
- Reinventing the Wheel - 35%
- Misc - 10%
- Simple RooStats Calls - 20%



Ideal RooFortunes

- Plumbing
- Procedural macros
  - Diagnostics
  - Inference
  - Analysis/Collaboration specific tests and tools
  - Updates to RooStats API



RooFortunes

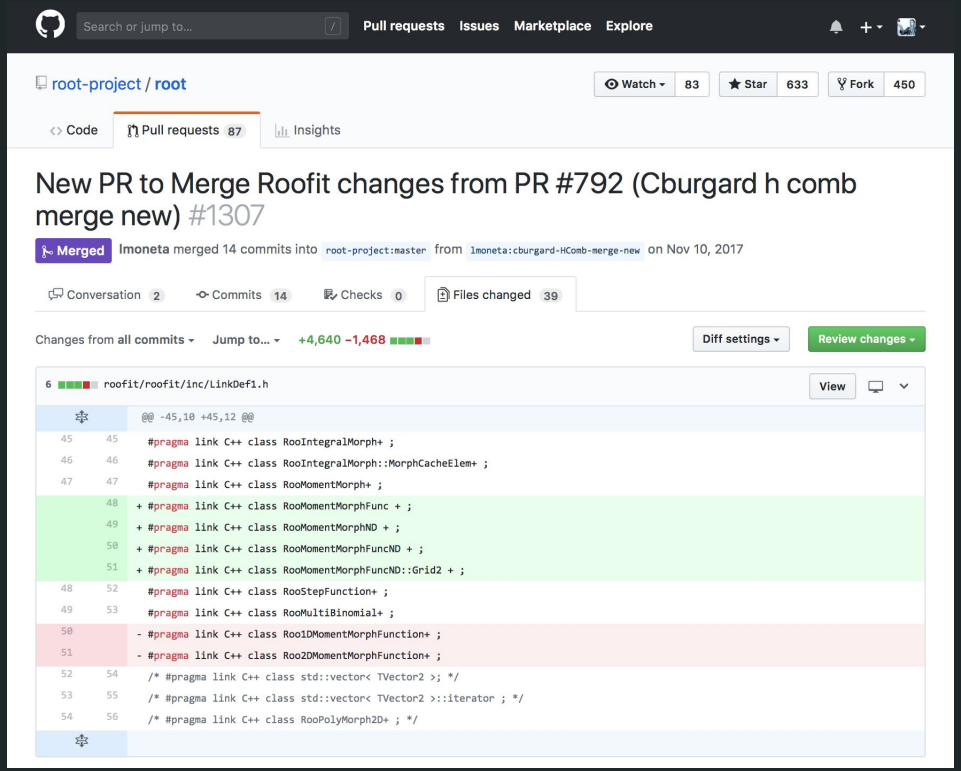
A Summary

Many tools passed into the domain of the ‘fit framework guru’ do not make it into general usage. Efficient computation and precise inference are dependent on this knowledge being passed on.

Education about RooFit usage should not be constrained to beginner development. Advanced topics need to be addressed and shared and a dialog between users and developers started.

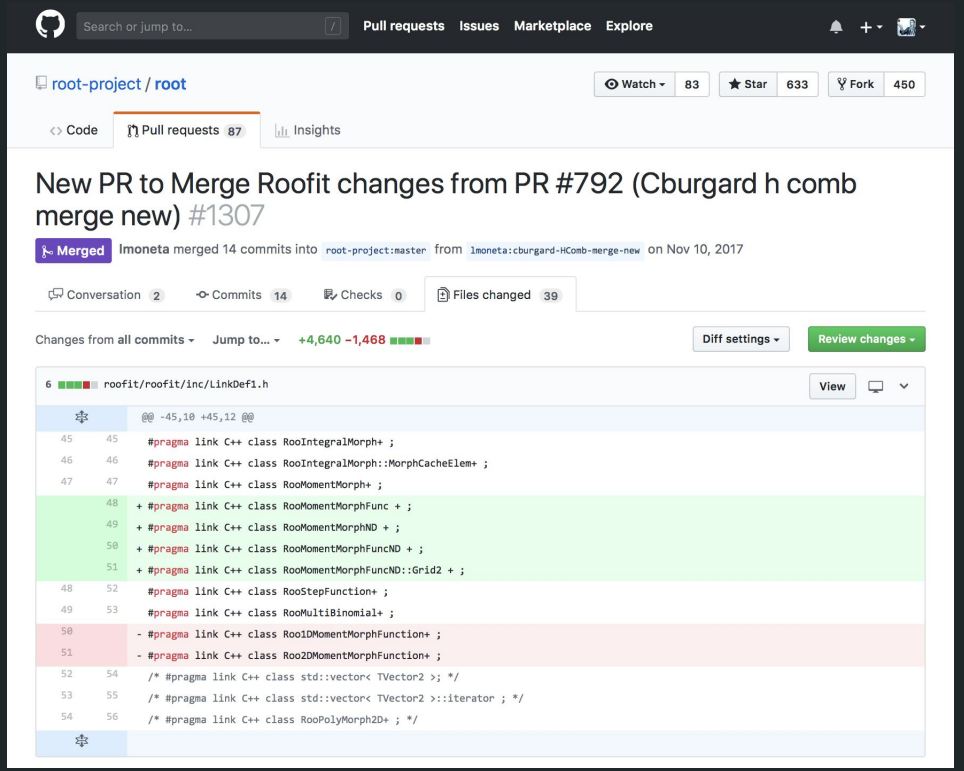
Open Fit

- The Demon PR
  - 6108 lines changed.
  - 39 files changed.
  - 7 years of ATLAS updates and bug fixes.
  - Half a year of PR conflicts and discussion.
  - 0 issue tracking, continuity, version control or accrediting authors.



Open Fit

- Want to make sure that all bugs and updates are reported, and quickly available to all users.
- Some features are not useful to all users and critical to others.
- Help should first come from those in similar situations.
- RooFit/RooStats bugs and developments can be conceptually very different from standard computing issues.



Summary

- Documentation is underway… but it's a large task.
  - Needs accurate representation of both statistical and computational aspects for analyses.
  - Needs to be useful for pedagogical, instructional and reference purposes.
  - Can be used to validate new tools.
- Interactive
  - Hosted on SWAN (and eventually maybe more)
  - See a plot, make a plot.
- Representative
  - For analysts by analysts.
  - Currently lots of ATLAS recommendations, maybe eventually expand.