IRIS-HEP Retreat - SSC

Sudhir Malik
University of Puerto Rico Mayaguez
26 May 2020

https://iris-hep.org/ssc.html

Team

- Training is a voluntary activity where IRIS-HEP supports training events
- There is no dedicated personnel
- The instructors are drawn from HEP LHC community
- Prepares workforce to contribute to intellectual goals of IRIS-HEP (and HL-LHC)

Activities

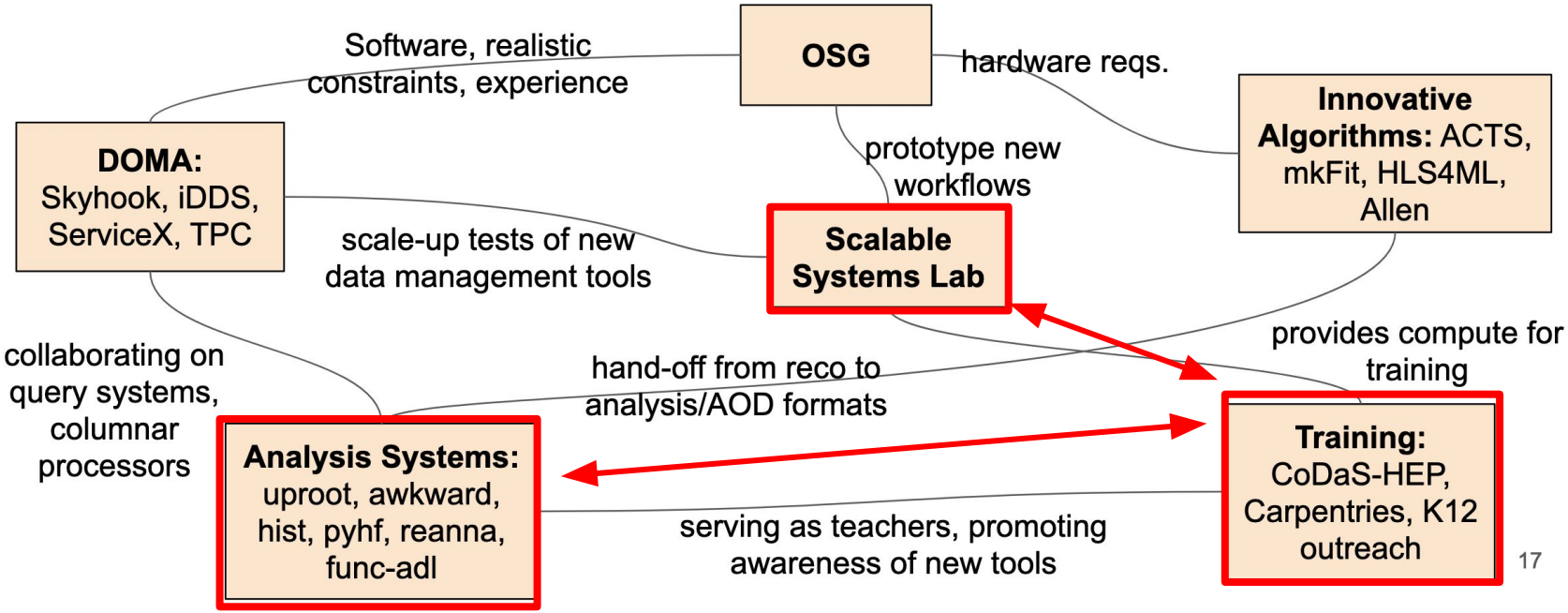
- Software Carpentries
- Developer's Training
- Mentoring
- Outreach and Education
- Blueprint

Training Links

- Training Events
  - https://indico.cern.ch/category/11386/
- Training Material
  - https://github.com/hsf-training
- Blueprint
  - https://docs.google.com/document/d/1uUGOJPp0kH0kTofe5GocfRNhbxm1oPxR9KTSbzvuoIQ/edit

Interactions within IRIS-HEP

IRIS-HEP is made up of teams that interact on an as-needed basis, forming collaborations both long and short-term. This data flow diagram is the result of gathering the teams by topic area and aggregating their lines of communication.



Collaborations outside IRIS-HEP

- Carpentries
  - https://carpentries.org
- High Energy Physics Software Foundation (HSF)
  - https://hepsoftwarefoundation.org/
- Training at National Labs - Fermilab, Argonne, LBNL, CERN
  - Neutrinos, Collider, Nuclear
- Quarknet (Data Camp)
- Latin American HEP
- LAWSCHEP
  - https://indico.cern.ch/event/813325/
  - https://indico.cern.ch/event/813325/contributions/3784509/attachments/2002241/3345728/LAWSCHEP2019-SummaryConclusions.pdf
- Snowmass 2021

Challenges

- Training on software across different experiment is grand challenge and IRIS-HEP is mitigating in that direction
- Build up community (from HEP) of instructors
- Success in training depends on voluntary participation from experts and developers in IRIS-HEP community as well as experiments

Year 3+ (HEP training)

- Produce Software Training curriculum
  - https://demo.codimd.org/x_XjjowgSDacKRNaeTS-yg?view#
  - Standard basic carpentry curriculum
  - Produce a HEP specific training curriculum
- Scale up the activities basing off the above
- Integrate Sustainability
  - Training & best practices beyond just coding
  - Project-level guidance - how to start and maintain projects, how to build community
  - Policy, including how to provide credit to software developers and maintainers
- Blueprint Meeting - experts from outside HEP community (Computing, Carpentries etc.)

Years 3+ (Education/Outreach)

- Expanding IRIS-HEP Outreach program
- Partnering with Quarknet
  - Reach the established K-12 teachers community
  - Integrate Software training with established programs
    - Data Camps, Masterclasses, e-labs
- Expand the STEM programming outreach via universities other than UPRM (e.g. other IRIS-HEP universities)

Summary

- Training has picked up momentum with **multiple training** and **outreach** events, 270 people trained (April 2019-Feb 2020)
- **Synergy** with Software Carpentries (https://software-carpentry.org/), HSF, FIRST-HEP and other projects
- Outreach activities in **Hispanic and underrepresented communities**, programming for STEM teachers, ML Hackathon among undergraduates from physics, math, engineering
- Plan in works to **synergise outreach with Quarknet** (Data Camp)
- Training for **non-US based HEP community** (CERN events)
- **Training Blueprint** - scaling up the activities, standardize basic carpentry curriculum, HEP specific basic curriculum
- Training **work** and **vision is communicated** at several **international workshops** and **conferences**
- Training is **building community of software developers, mentors around its users** that is essential for sustainability and support
- *Training is making a difference in HEP community and beyond*

Back up

Events (upto July 2020)

Upcoming events (Impacted by COVID-19 and post earthquake in Puerto Rico) :

- 27-31 July 2020 - *Computational and Data Science for High Energy Physics (CoDaS-HEP) 2020 School* - Princeton University - Webpage (undecided)
- ~~21-23 April 2020~~ 2-4 June 2020 - Virtual Pipelines Training (all online) - Indico page
- ~~21-22 March 2020~~ - Programming for STEM at UPRM - Indico page (postponed to Virtual style in June)

Past events:

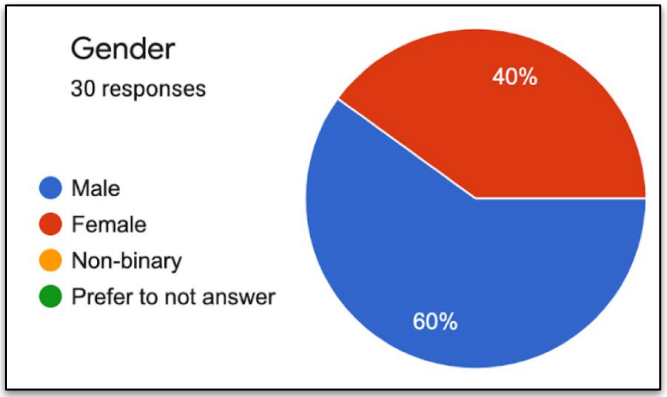
- 20 February 2020 - Training Blueprint Meeting (Virtual) - Indico Page
- 17-19 February 2020 - Analysis Preservation Bootcamp - Indico Page
- 27-29 November 2019 - Software Carpentry at CERN - Indico page
- 9-21 August 2019 - ATLAS Software Carpentries Training - LBNL - Indico page
- 22-26 July 2019 - *Computational and Data Science for High Energy Physics (CoDaS-HEP) 2019 School* - Princeton University - Webpage
- 10 June 2019 - FIRST-HEP/ATLAS Software Training - Argonne National Laboratory - Indico page
- 3-4 June 2019 - An introduction to programming for STEM teachers - UPRM- Indico page
- 24-26 April 2019 - Machine Learning Hackathon for UPRM Students - UPRM - Indico page
- 1-2 April 2019 - Software Carpentry Workshop - Fermilab - Indico page

See backup slides for specific details

Software Carpentry at Fermilab

- 1-2 April 2019
- Agenda: https://indico.fnal.gov/event/20233/
- Version Control with Git
- Python Foundations
- Building Programs with Python
- Data analysis - Numpy, Pandas
- Data analysis Cont. and Graphs
- Advanced Python and PyROOT, uproot
- Post-workshop Survey

Software Carpentry at CERN



- 27-29 November 2019
- **Unix** (e.g. shell, bash and scripting)
- **Git and Github** – how to version control your code
- **Python** – fundamentals of using the Python language
- Jupyter Notebooks and **SWAN**
- **Python for analysis** – how to combine Python with ROOT to start analysing data (i.e. pyroot and uproot)

Developers Training

CoDaS-HEP School (Princeton)

The CoDaS-HEP school aims to provide a broad introduction to these critical skills as well as an overview of applications High Energy Physics. Specific topics to be covered at the school include:

- Parallel Programming
- Big Data Tools and Techniques
- Machine Learning - Technology and Methods
- Practical skills: performance evaluation, use of git for version control

The program includes both lectures and practical hands-on exercises.

http://codas-hep.org

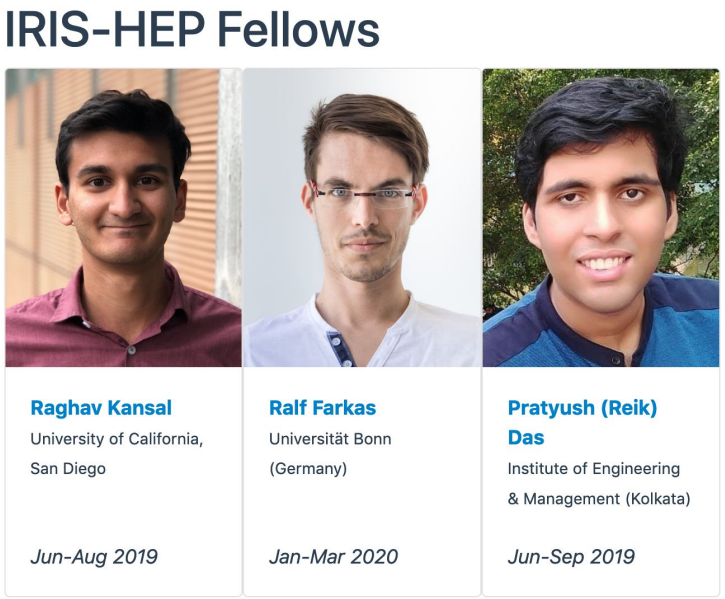
Mentoring

- **IRIS-HEP Fellows Program**

Key Insight: we need to provide incentivized and explicit paths forward for enthusiastic students from the more advanced training schools (ESC/Bertinoro, CoDaS-HEP, MLHEP, etc.) or for people who become engaged with our software projects in other ways.

Project focused: bring students into contact with “mentors” to work on a specific, pre-defined project, allowing them to grow their software skills and project experience. The fellow supports, when possible, travel and subsistence for a 3 month extended stays in the mentor’s institution.

We began this activity as part of DIANA/HEP (an NSF SI2-SSI project) and are continuing it in IRIS-HEP.

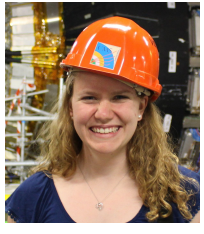


We have had 3 fellows thus far during the ramp-up of IRIS-HEP, and have just announced a call for Fellow proposals for summer, 2020 and beyond.

Mentors/Developers

Sebastian Macaluso (CoDaS-HEP 2017) has taken a postdoc position with NYU in IRIS-HEP and is focusing primarily on ML for jet physics within IRIS-HEP.

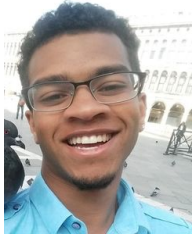
Michael Hedges (CoDaS-HEP 2017) took a postdoctoral position with Purdue (no funding connection to IRIS-HEP), but has become an active contributor to scikit-hep/awkward-array. In particular he is working to adapt it for use with Pandas with a goal of using it in the Mu2e experiment at FNAL



Allison Reinsvold Hall (CoDaS-HEP 2018) took a postdoc position with FNAL and is now an active contributor to the Parallel Kalman Filter tracking project in IRIS-HEP (funded by FNAL and a DOE SciDAC project).

Oksana Shadura (CoDaS-HEP 2019) is a computer science graduate student working on DIANA/HEP (and transitioning soon to IRIS-HEP) in the Analysis Tools/Systems area.

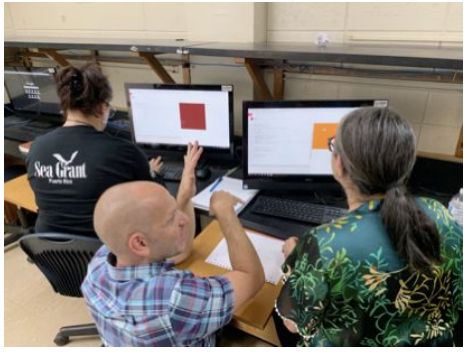
Mason Proffitt (CoDaS-HEP 2019) is a U.Washington graduate student working in the Analysis Systems area and contributing to uproot, awkward-array, etc.



Tres Reid (CoDaS-HEP 2019) is a Cornell graduate student working on mkFit (tracking) in the Innovative Algorithms area.

Savannah Thais (CoDaS-HEP 2019) is a Princeton postdoc working in the Innovative Algorithms area, looking at applying GraphNN's for tracking

Programming for K-12 teachers



- 4-5 June, 2019
- Agenda: https://indico.cern.ch/event/817539
- Introduction to Processing and p5js
- Python and Colab
- Basics of python, Jupyter notebooks, and
- Colab (hands-on)
- Data analysis with python

ML Hackathon

- 24-26 April 2019
- Agenda: https://indico.cern.ch/event/809812
- Talks ML/Data science Applications - Physics,
- Computer Science, Math, Engineering undergard students
- Machine Learning, Deep Learning, ANN
- Hands-on exercises
- Hackathon on LHC Physics

Scientific Software Club

Our Software Carpentries material resides in GitHub and inspires SSC at for its related activities

Quote - "The group has derived its initial material and motivation from activities hosted by the NSF funded advance cyberinfrastructure projects like IRIS-HEP and FIRST-HEP at UPRM." (http://www.sscuprm.com/about/overview)
)

Scientific Software Club

The **Scientific Software Club (SSC-UPRM)** is a student organization at the **University of Puerto Rico at Mayagüez** that is dedicated for students with an interest in software development. We want to help students learn and use different tools and techniques that wil help them develop better software and contribute to research and technology, no matter their field of study.

http://www.sscuprm.com/

Recent Posts

Bash Workshop

What's Bash? What do I use it for? Our members had the opportunity of learning the basics of this powerful tool!

Git Workshop

What's Git? Here our members learn the basics of this powerful tool!

Python Workshop

Learn to use Python one of the most powerful programming languages today!

Upcoming Events

Past Events

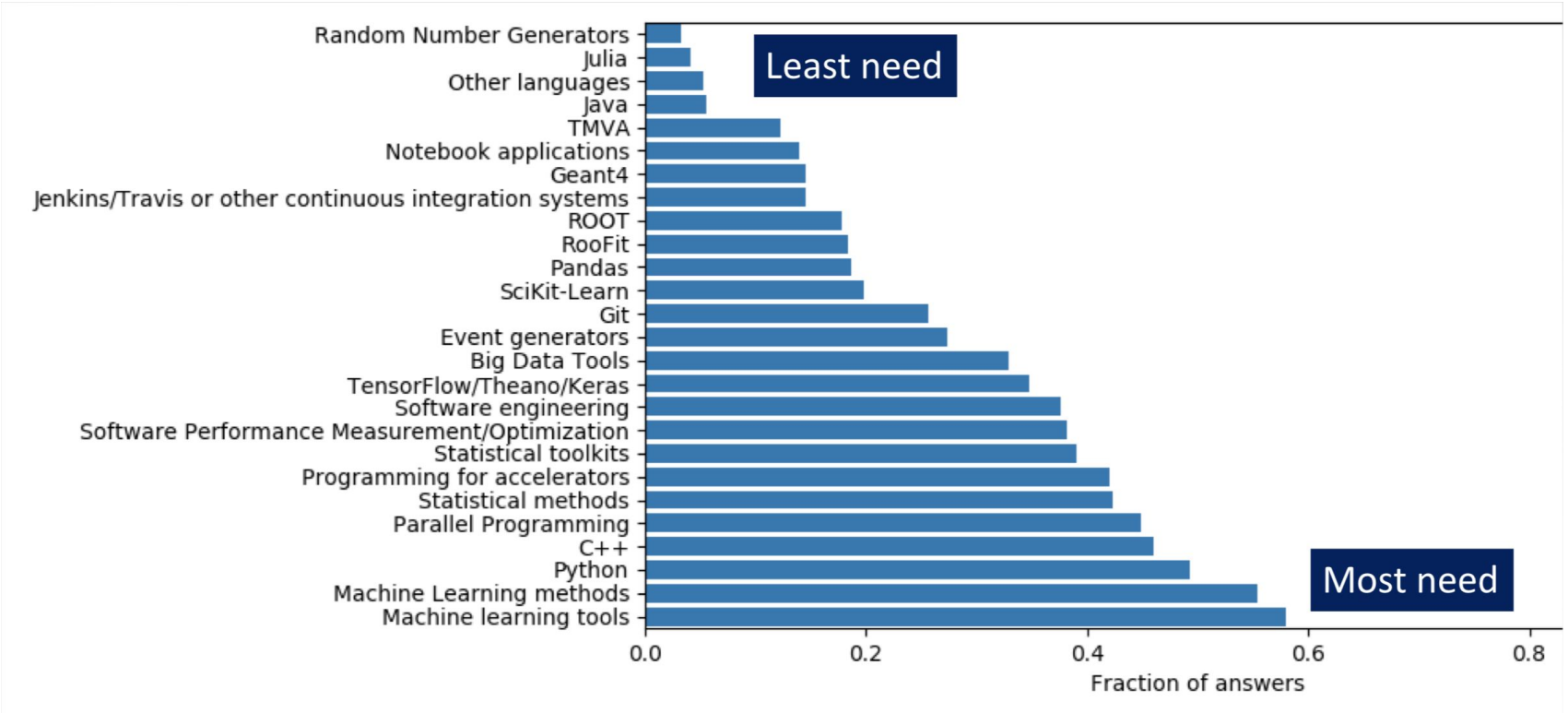
Taller de Bash (Terminal)
When: 9/10/2019, 10:30:00 AM
Where: Centro de Computos

Taller (continuación de Bash)
When: 9/12/2019, 10:30:00 AM
Where: F-437 Salon de Cómputos

Taller de Git/Github
When: 9/26/2019, 10:30:00 AM
Where: TBA

Initial Training Survey (assess needs of the community)

- Survey of training needs (link for results summary), Jan 2019 , 334 people responded!
- Input on training needs, current training practices for various topics in HEP related to software/computing and related software centric areas.
- Information used to seek resources and to organize training activities to meet those training needs.

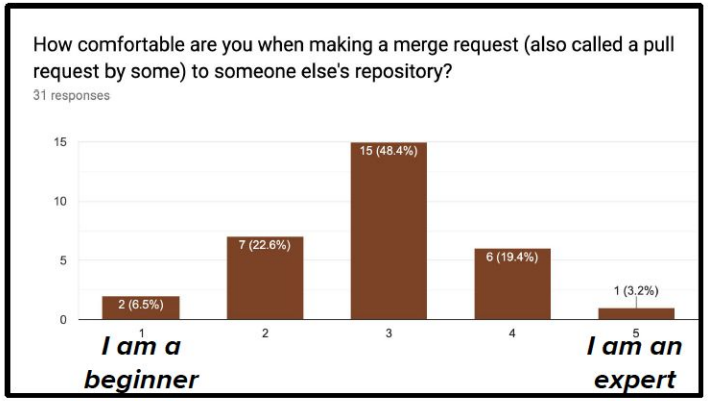


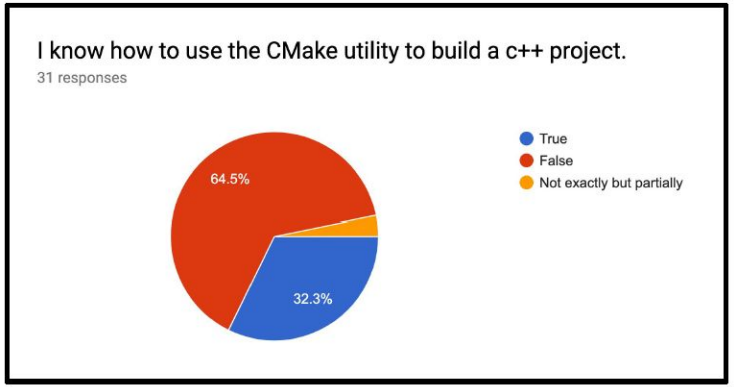
Blueprint

- Tutors involved in training activities are from different experiment communities like ATLAS, CMS, LHVb, Belle-II etc, by design input is present
- Participants in training are from different experiments,
- First blueprint meeting in February 2020
- Future blueprint should involve experts from outside HEP community (Computing, Carpentries etc.) for input

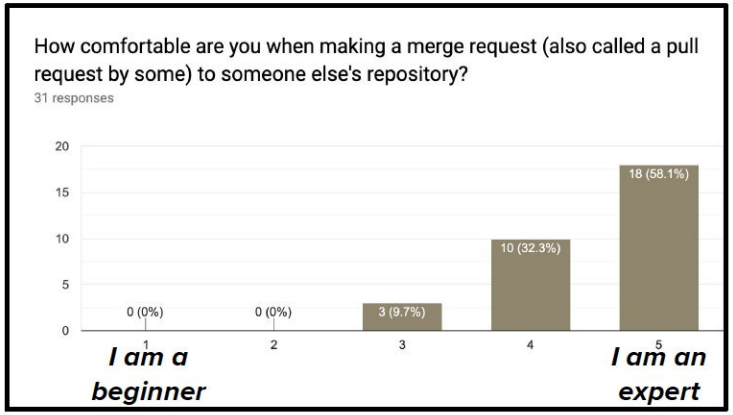
Feedback (training at LBNL)

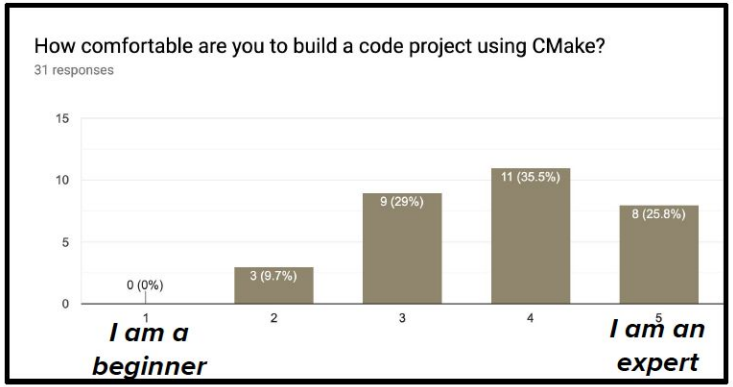
Before





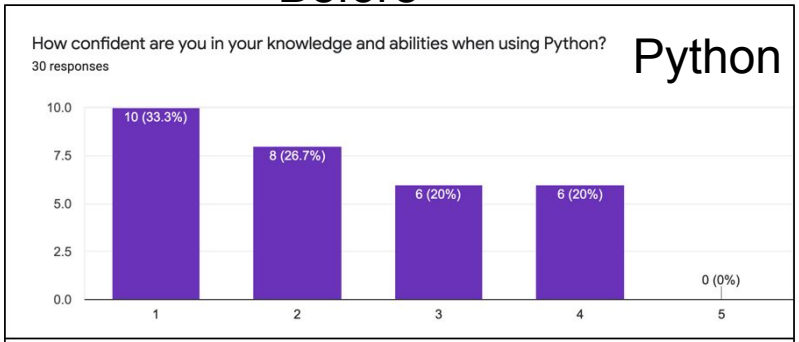
After

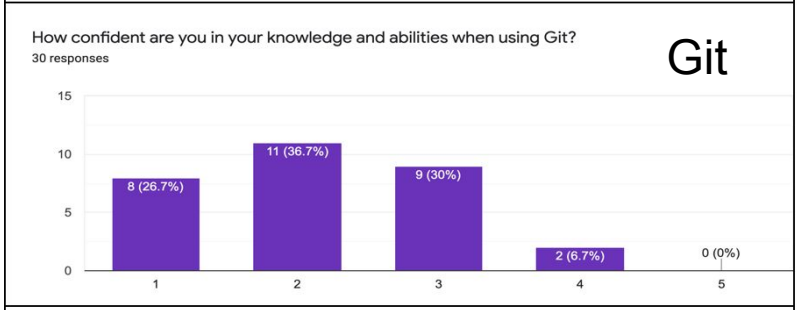


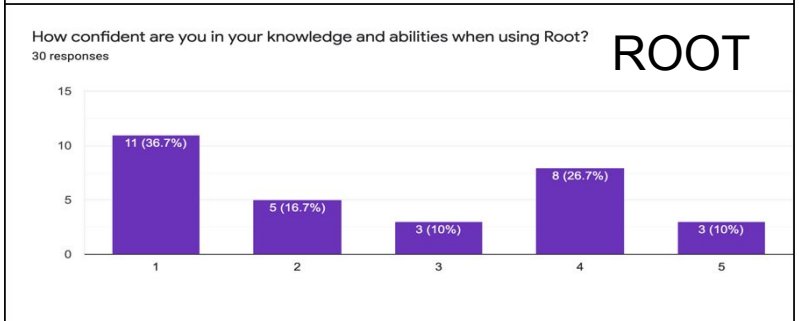


Carpentry Survey (Software Carpentries at CERN)

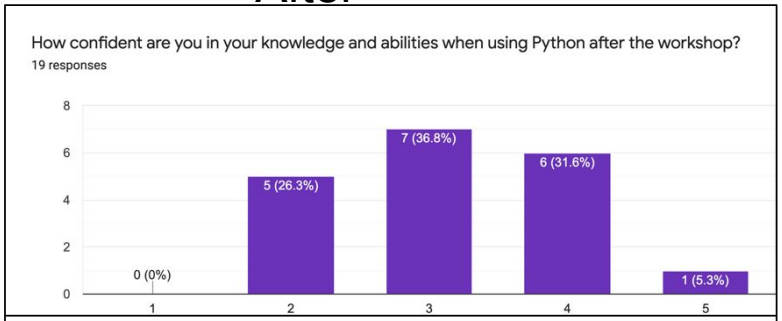
Before

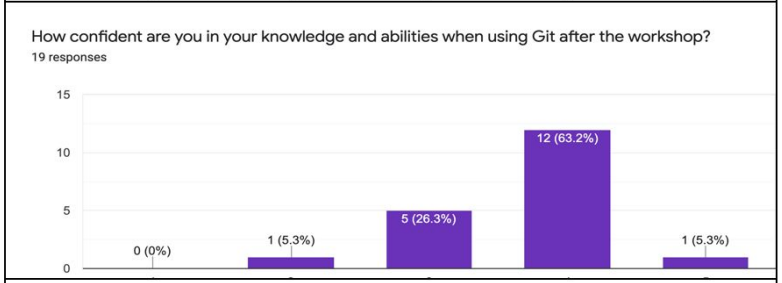


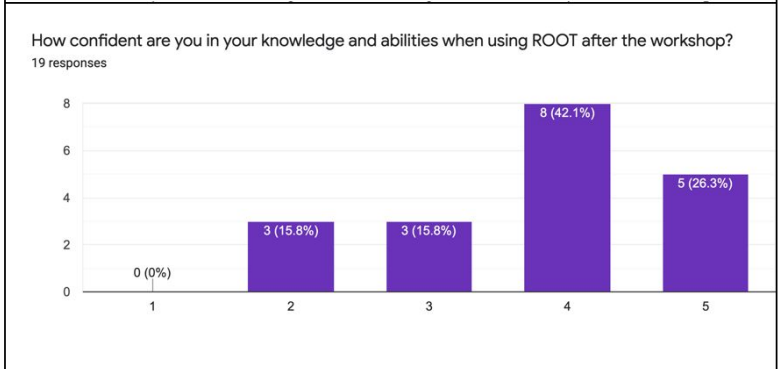




After







Tracking Progress

	B	C	D	E	F	G	H	I	J
4	**Sofrware Sustainabily Core**								Late
5	**Description**	**Y1Q1**	**Y1Q2**	**Y1Q3**	**Y1Q4**	**Y2Q1**	**Y2Q2**	**Y3Q3**	**Y4Q4**
6	Co-organize (with FIRST-HEP) introductory level Software Carpentry (bash/git/python) workshop at Fermilab combined with topical discussion on interoperability of ROOT for advance PhD students			Due					
7	Introductory level Carpentry session for CMS/ATLAS at CERN					Due(shifted)			
8	Invite LHCb/ALICE to showcase Starterkit on pre and first analysis steps		Due						
9	Carpentries curriculum for HEP for newcomers at a LHC/ FNAL Experiments			Due					
10	Post-training co-survey with FIRST-HEP on better assessing training needs			Due					
11	Retune next Carpentry workshop benefiting from above participant survey			Due					
12	Organize and run CoDaS-HEP School (Parallel Programming, ML, Big Data Tools) at Princeton				Due				
13	First brainstorming workshop in conjunction with CoDaS-HEP school to re-visit software and data challenges and correspondingly design future training material and workshops				Due				
14	Summer Outreach Workshop (s) on building software skills at US universities			Due					
15	Second HEP Software Carpentry Workshop at a US university/Lab				Due				
16	Carpentries curriculum for HEP using best practices for HEP				Due				
17	Co-organize (with FIRST-HEP) boot camp on domain specifc topics like advanced ROOT, Geant4 and other HEP-specific tools					Due			
18	First Data Carpentry Workshop catered to HEP needs (expert input from CoDaS-HEP tutors and IRIS-HEP Fellows and field experts)					Due			
19	Software and Data Carpentry Workshops for HEP community by IRIS-HEP program, IRIS-HEP Fellows form a community of tutors for workshops and future domain specific topics					Due			
20	Summer Outreach Workshop(s) on building software and data analytical skills at US universities supported by IRIS-Undergrad Fellows and IRIS Fellows in their own community, "Hack week" as a model for data science education and collaboration						Scheduled for 2020		
21	Organize and run CoDaS-HEP at Princeton						Scheduled for 2020		
22	Second brainstorming workshop in conjunction with CoDaS-HEP school to re-visit software and data challenges preparedness in terms of training material						Scheduled for 2020		

Metrics

M.5.1	Number of introductory level HEP-specific software/computing workshops per year	4 in last year
M.5.2	Number of HEP domain specific trainings/workshops and developer training school CoDaS-HEP School	1
M.5.3	Number of related outreach workshops across participating institution communities	2
M.5.4	Number of IRIS-HEP fellows tied to training and number of trainees becoming mentors	3
M.5.5	Number of people trained each year across HEP community (Number of US ATLAS+CMS PhDs entrants per year is ~ 100)	270 in last year
M.5.6	Introductory HEP curriculum (blueprint)	1
M.5.7	Assessment framework for the ensemble of activities to measure the impact of our activities	In works

- https://docs.google.com/document/d/1gV0bf1lwTZK5Em-yZdJo2PuYjPdu5A09t6xK6N5bKuY/edit#