

Analysis Preservation Workshops

Leonora Vesterbacka

for the Awesome Workshop team/HSF Training Community

Snowmass Computational Frontier Workshop

August 11th 2020



NYU



HEP Software Foundation





- HEP community outputs hundreds of analyses
 - make the most use of every analysis
 - increasing interest for re-interpretability - RECAST/REANA
 - ensure knowledge transfer and code preservation
 - analyses should not disappear when the people who wrote the code move on!
- Workshops introducing software tools for analysis preservation
 - CI/CD, Docker, REANA
 - facilitating preservation of analyses in a robust way
 - small effort now - huge potential benefit in the future



Basics of training

- Interactive training
 - Instructor cover material in shared screen
 - Participants follow on their own laptops
- Mentors play a crucial role
 - help debug issues in real time - sticky notes!
 - 5:1 participant to mentor ratio
 - mentors recruited from previous workshops



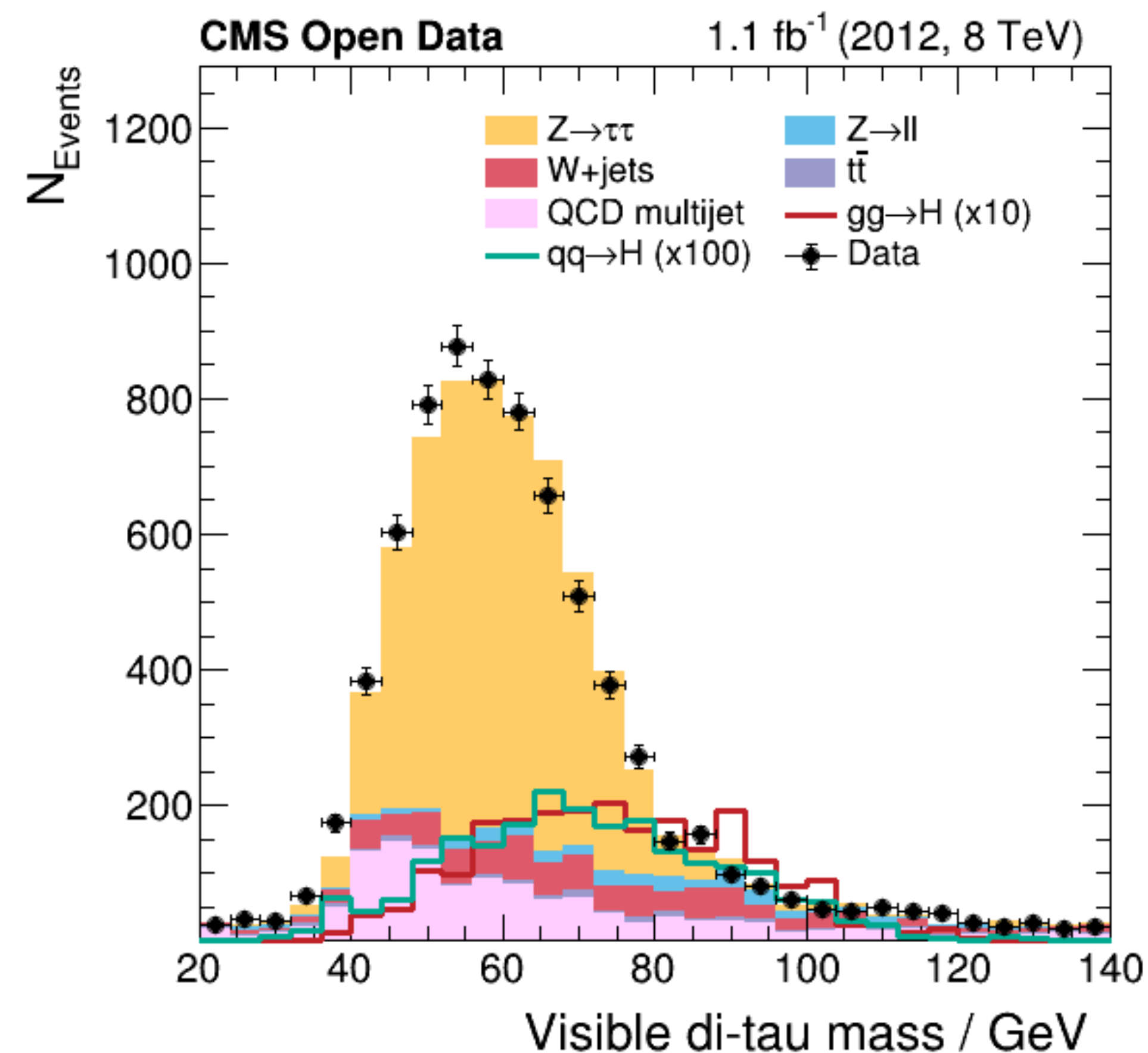
Instructor Giordon Stark introducing Pipelines



Mentor Danika Macdonell helping participant



- Same example material used in all trainings:
 - [CMS open data Higgs to tau tau analysis](#)
 - covers all basic steps of an analysis
 - skim datasets
 - produce histograms
 - make plots
 - make a measurement
 - perfect for automating the workflow!





- Continuous Integration and Development (CI/CD)
 - introduce GitLab's CI/CD, YAML
 - build a CI/CD workflow for example analysis
 - implement CI/CD workflow for participants own analysis





- Continuous Integration and Development (CI/CD)
 - introduce GitLab's CI/CD, YAML
 - build a CI/CD workflow for example analysis
 - implement CI/CD workflow for participants own analysis
- Docker Images and Containers
 - introduce Docker Images and Containers
 - write Gitlab Dockerfile for participants own analysis



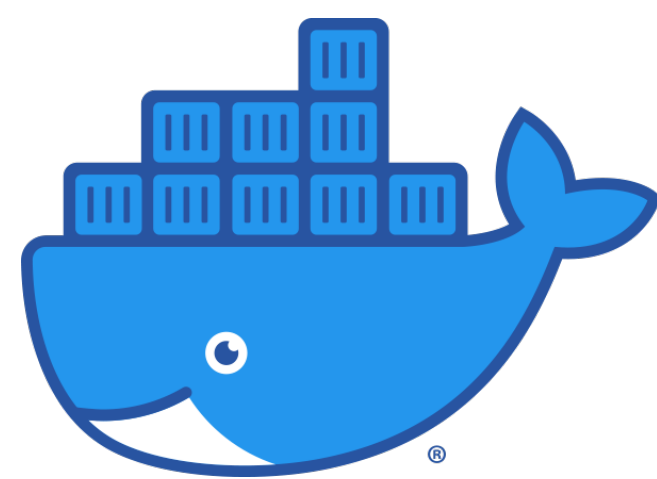


- Continuous Integration and Development (CI/CD)
 - introduce GitLab's CI/CD, YAML
 - build a CI/CD workflow for example analysis
 - implement CI/CD workflow for participants own analysis
- Docker Images and Containers
 - introduce Docker Images and Containers
 - write Gitlab Dockerfile for participants own analysis
- REANA: reusable and reproducible research data analysis platform
 - write a serial workflow for example analysis
 - to be run sequentially on REANA platform





- 1 workshop - 3 topics - 3 days
- ATLAS+CMS Analysis Preservation Bootcamp Feb 19-21 2020 @ CERN, 30 participants



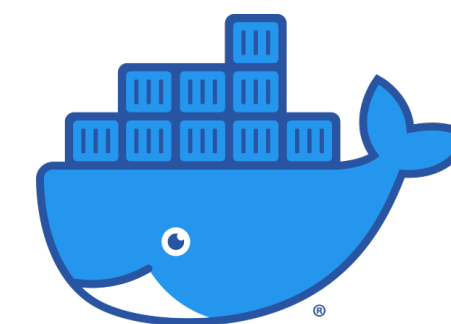
AM:
Interactive lesson +
implementation to example analysis

lunch time

PM:
Experiment specific implementation +
implementation to participants own analysis



- *3 workshops - 1 topics - 3 days*
- Virtual CI/CD training June 2020, 250 participants
- Virtual Docker training July 2020, cap on 100 participants
- Virtual REANA training planned for October 2020



DAY 1:

- Introduction
- Release pre-recorded video lessons
 - captions courtesy of IRIS-HEP!

DAY 2:

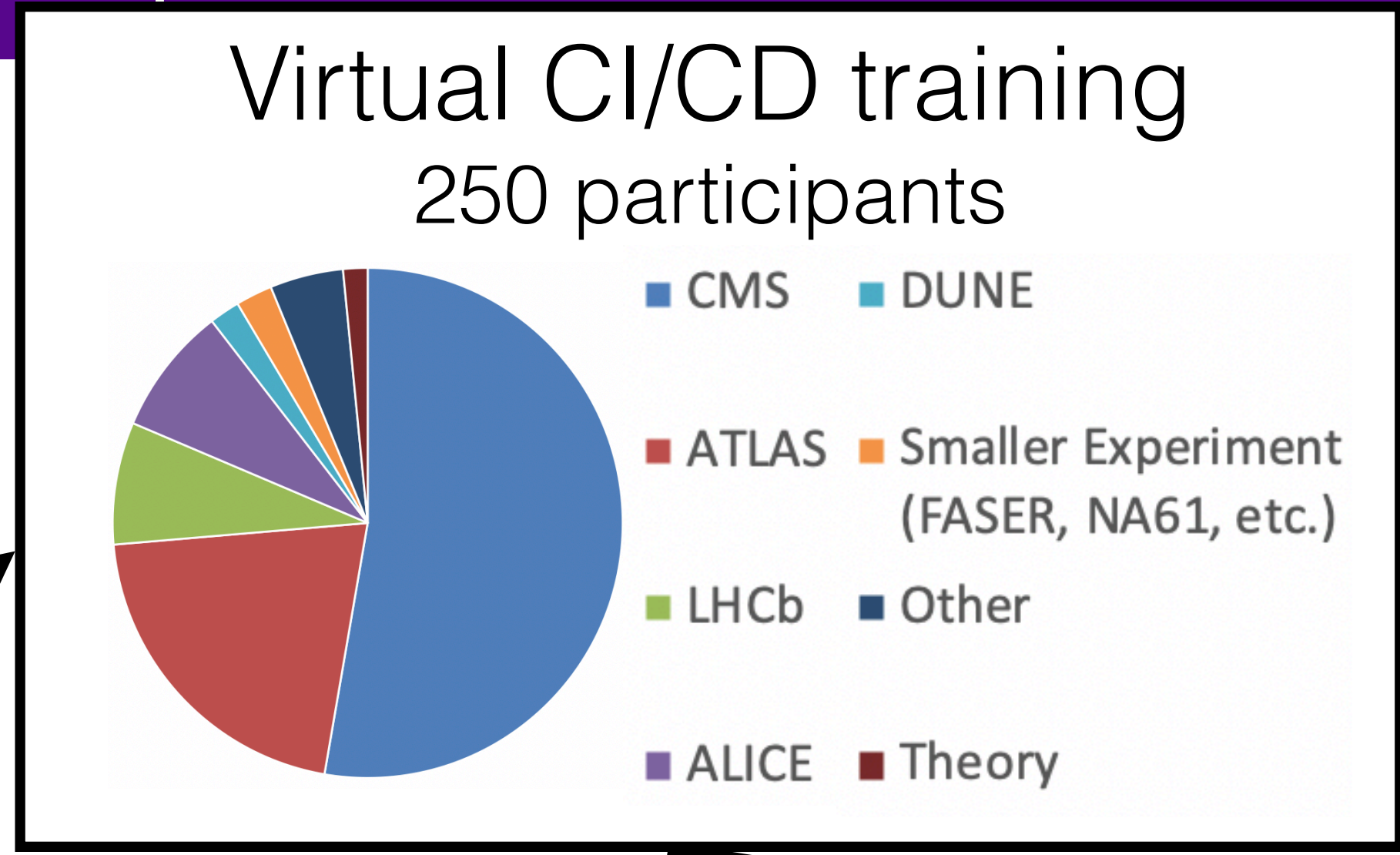
- Watch lessons
- Implement concept on example analysis
- Support available in mattermost channel

DAY 3:

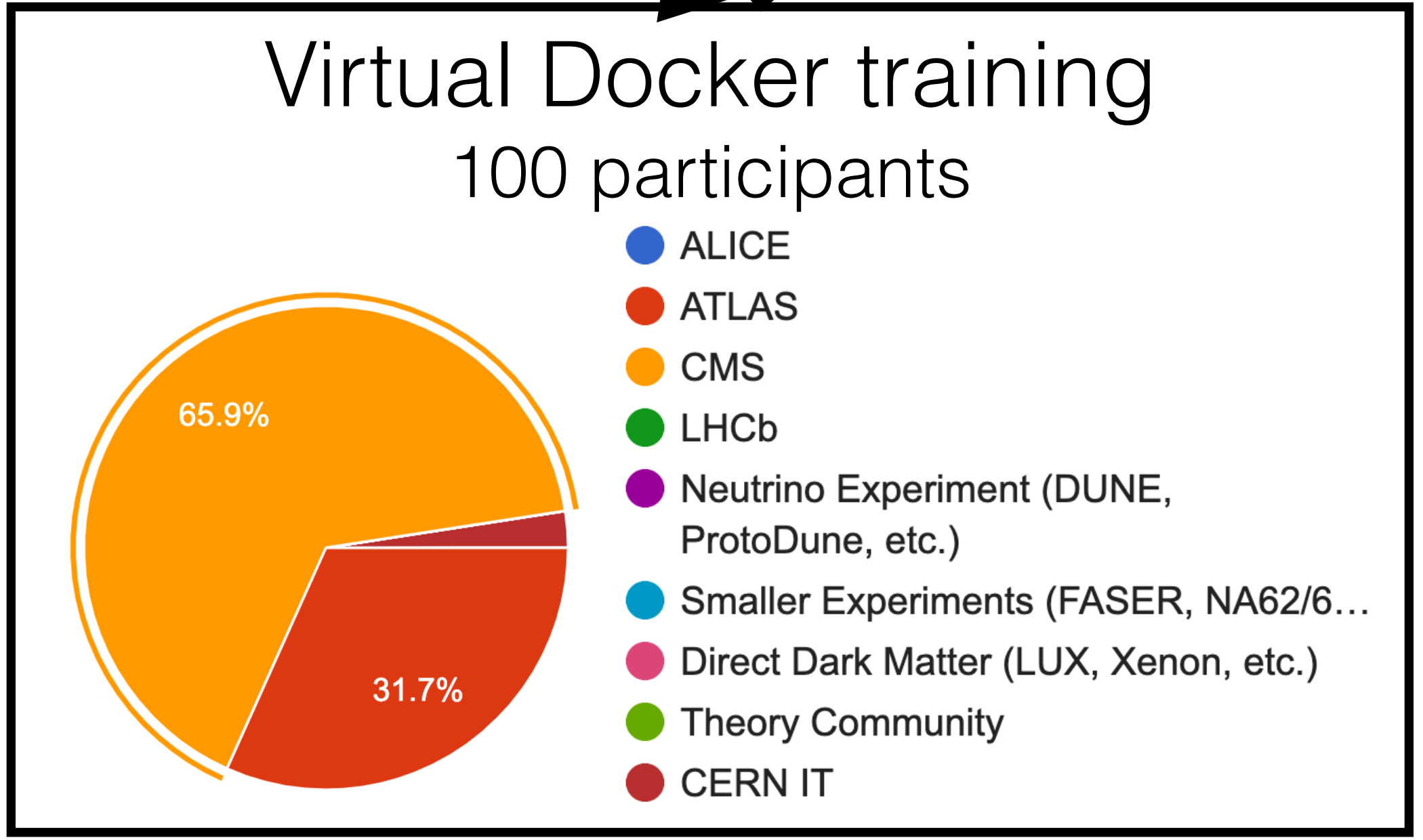
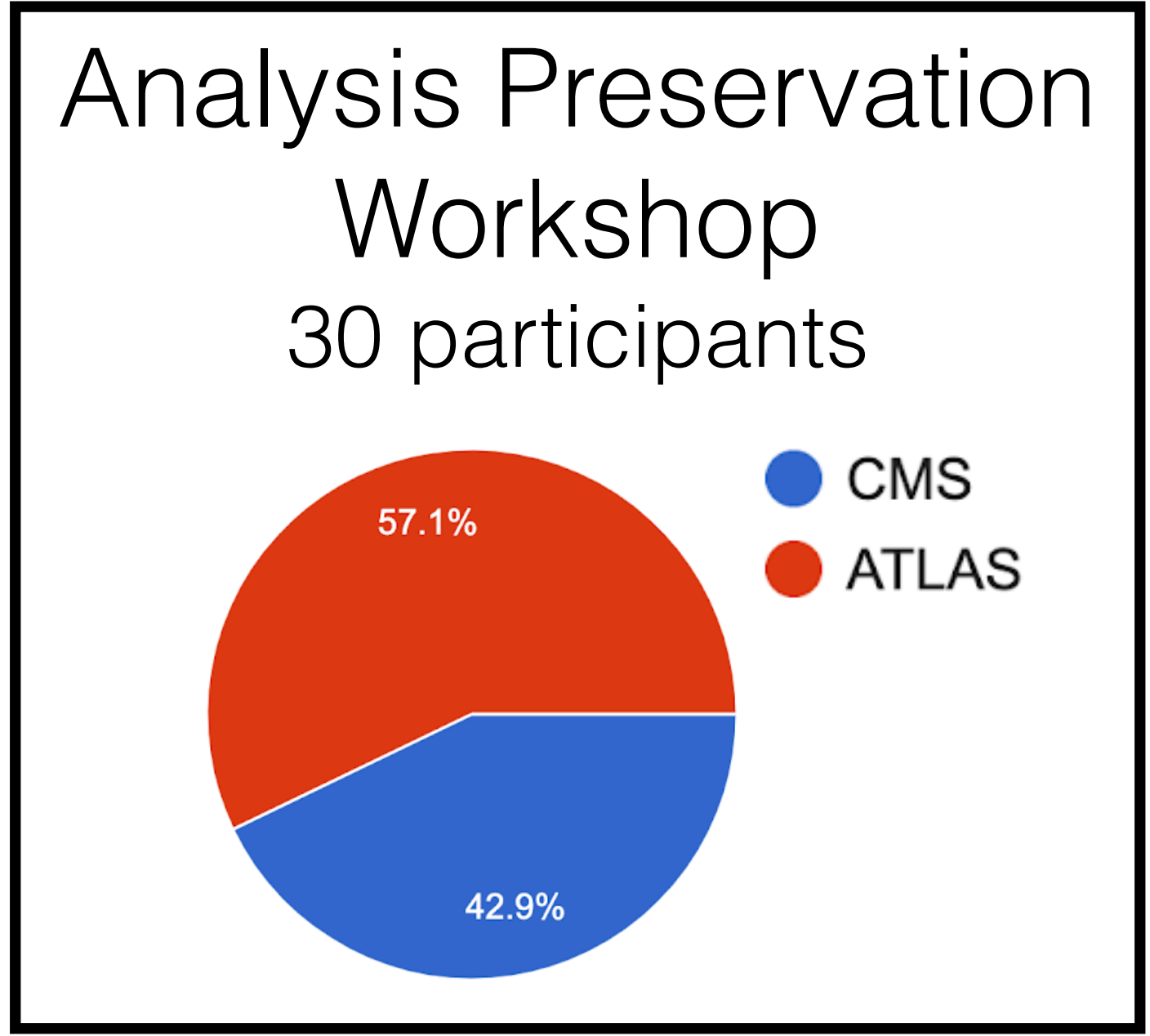
- Hands-on session with mentor on zoom
- Implement concept on participants analysis



US ATLAS Bootcamp
30 participants
US-ATLAS only



Virtual REANA workshop planned

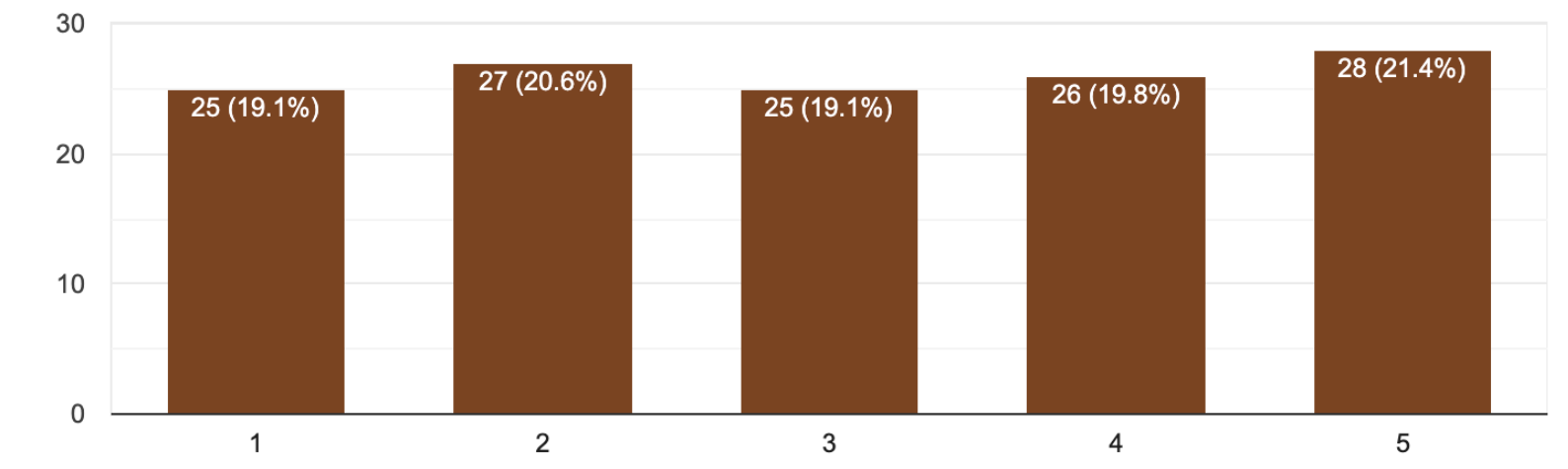




Do they work?

How likely would you have been to attend this bootcamp/workshop had it been held in person at CERN with no external connection?

131 responses



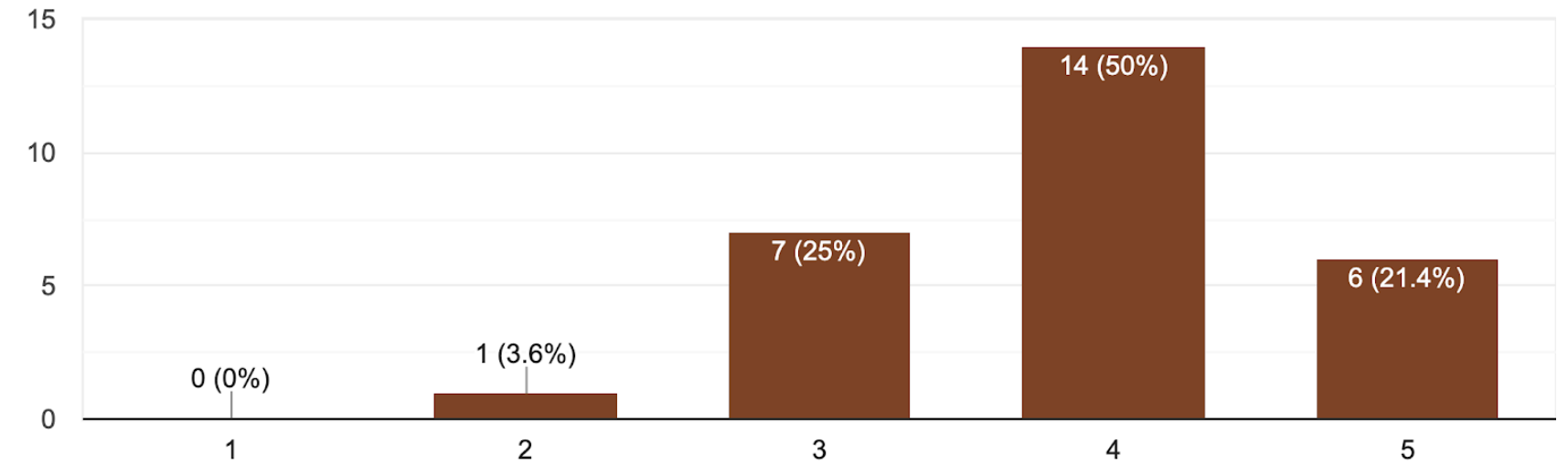
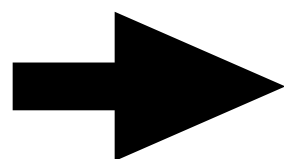
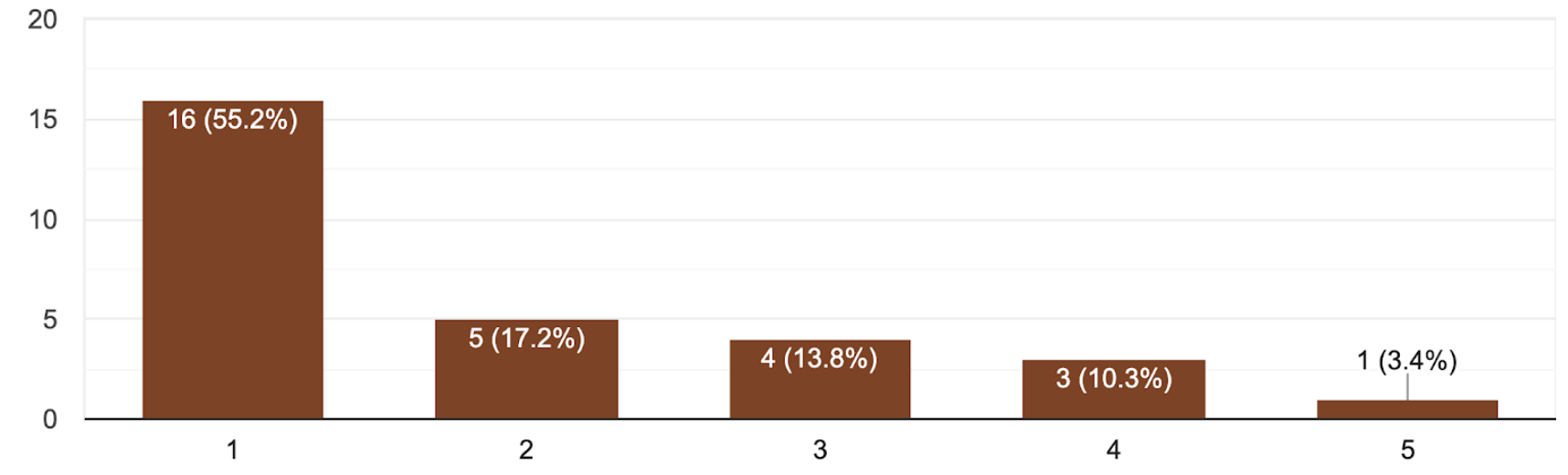
- Virtual complement to - not replacement for - physical workshop
- Virtual trainings allows for wider participation

pre-workshop survey

post-workshop survey

I am confident enough to write Pipeline tests for my analysis.

I am confident enough to write Pipeline tests for my analysis.

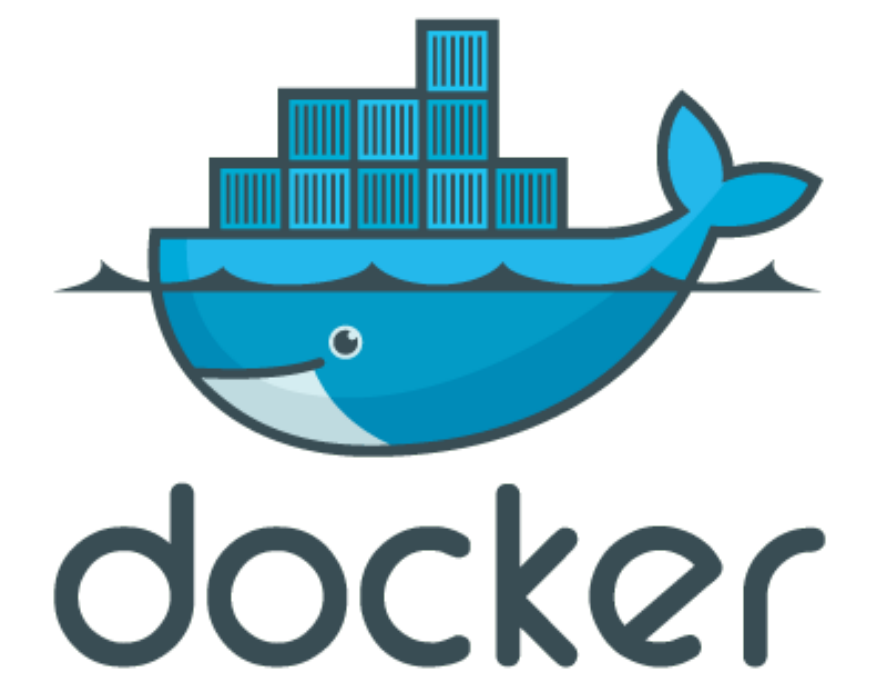
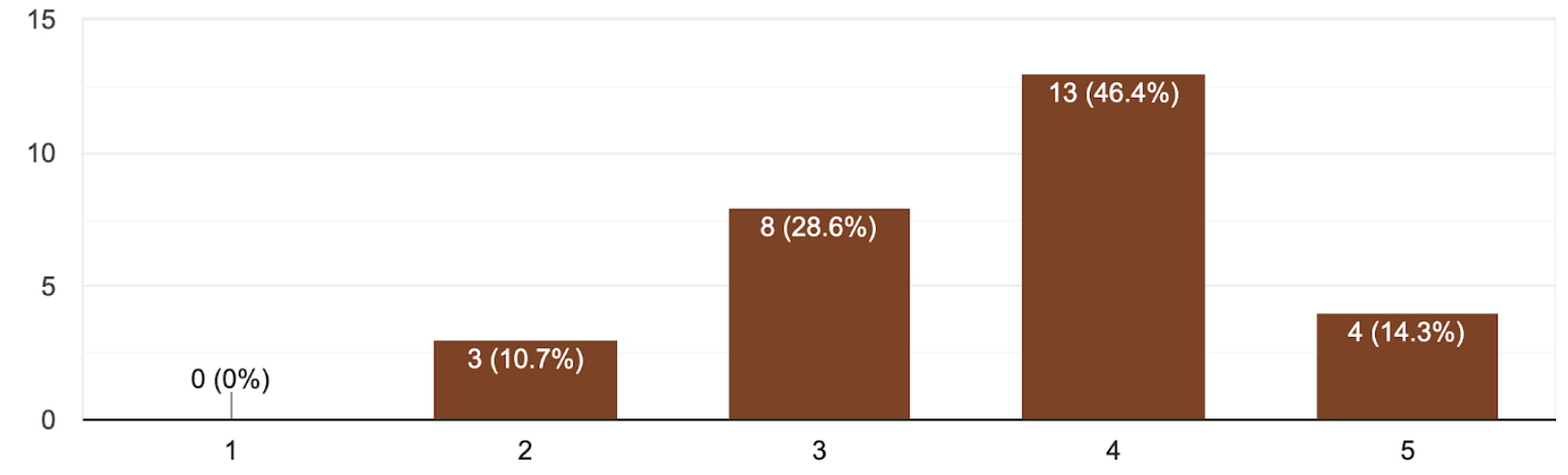
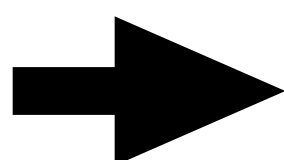
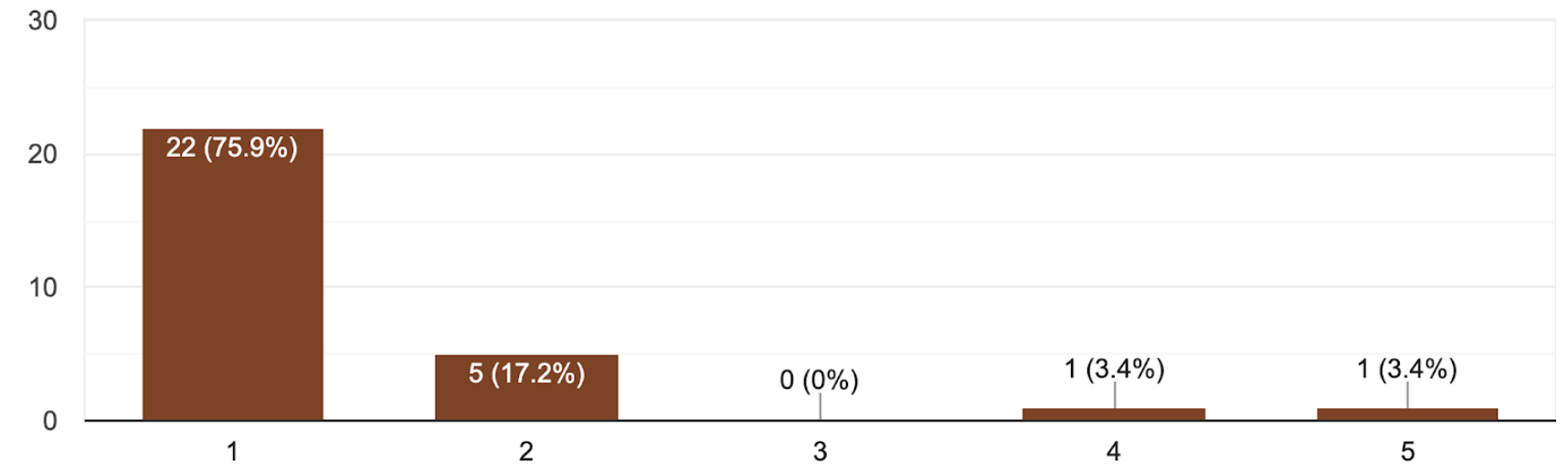


I am confident enough to write a Dockerfile to build an image for my work.

I am confident enough to write a Dockerfile to build an image for my work.

29 responses

28 responses



Novice

Expert

Novice

Expert



- Analysis Preservation Workshops
 - brought to you by the HSF Training Community
 - providing tools for archiving and re-running analyses
 - increase reproducibility - enable re-interpretability
- Physical and Virtual trainings
 - Great interest and participation
 - Physical workshops are great networking opportunities
 - Virtual workshops allow for wider participation
 - video lessons on youtube

