ATLAS Open data conversion for Higgs to 4 lepton analysis.

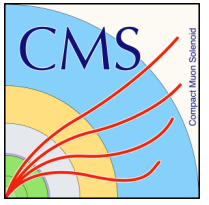
Summer Student Project

Valeriia Pinchuk

Ivan Franko National University of Lviv, Ukraine

Supervised by Achim Geiser,
Lucas Karwatzki, Aviral Akhil

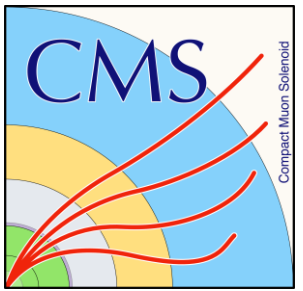
HELMHOLTZ
DESY.



Motivation
and main goals

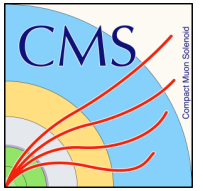
- ATLAS and CMS use completely different data formats
- This locks analysis code into a single experiment
- We want to create a common data layer for both
- Write your code once, run it on ATLAS and CMS data
- Before, only educational ATLAS data were used

Why is this important for PUNCH



This project extends PUNCH's existing analysis of the Higgs to four lepton channel by enabling the same code to process both ATLAS and CMS Open Data on research level. By creating a common data format, we directly support PUNCH's goal of unifying analysis across experiments. Using public data and shared tools makes research more collaborative and accessible.

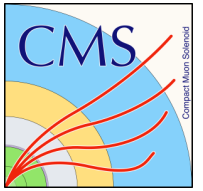
ATLAS PhysLite[1] vs CMS NanoAOD



Columnar (RNTuple)	**format**	Object-based (TTree)
Flat columns	**data**	Structured collections
Direct Python analysis	**use**	ROOT/Uproot analysis

Conversion Purpose: be able to run same code on both

ATLAS Open Data and nTuples



ATLAS Open Data Initiative:

A project to make real LHC collision data publicly available for students, educators, and researchers worldwide.

nTuples:

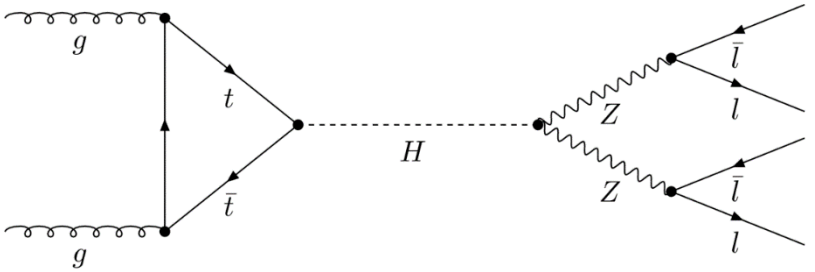
A simple, table-like data structure used in particle physics. Each row is an "event" (a collision), and each column is a measurable property (like electron energy or jet momentum).

Why:

We can start directly with the ATLAS nTuples, which are readily available for download on the ATLAS Open Data portal[2]. These are already in a reduced, analysis-ready format.

H→ZZ→4l “the golden channel”

Validation



Clean Signature: Very few background processes produce four isolated, high-energy leptons

Excellent Mass Resolution: Allows for a sharp, unambiguous reconstruction of the Higgs mass peak at ~125 GeV

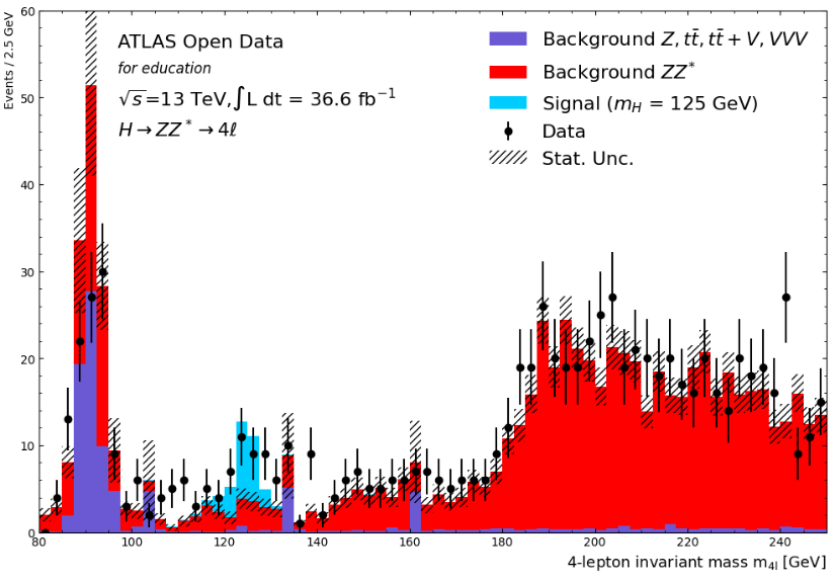
$ggH -> ZZ -> llll$ – main signal
$VV -> 4l$ – main background
$Z/\gamma^* + X$ and $t\bar{t}$

ATLAS tutorial notebook
HZZAnalysis.ipynb[3]

provides a practical introduction to reproducing a simplified Higgs boson analysis (H → ZZ → 4ℓ) using the Python ecosystem — specifically uproot and awkward arrays for ROOT data handling

Since this code is easy to work with — for example, to separate 4μ, 4e, and 2μ2e channels or to use the same datasets — this plot can be taken as a reference to compare the results, as they should be similar

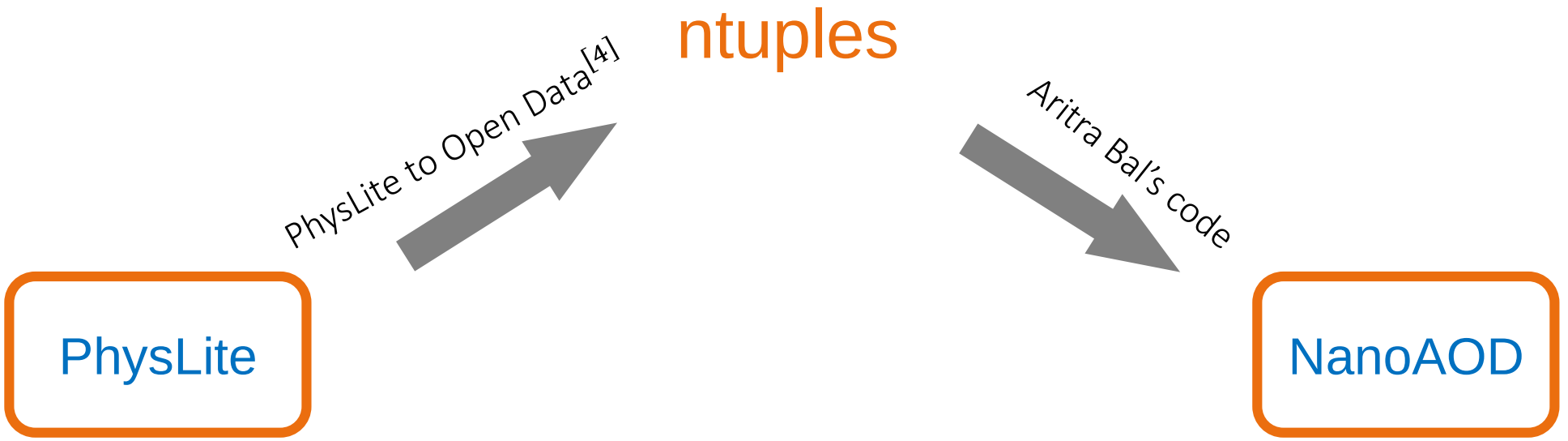
Exactly 4 lepton skim!



Back to conversion

PhysLite → NanoAOD

How to do it?


ntuples
PhysLite to Open Data[4]
Aritra Bal's code
PhysLite
NanoAOD

Analysis itself

Data is already converted

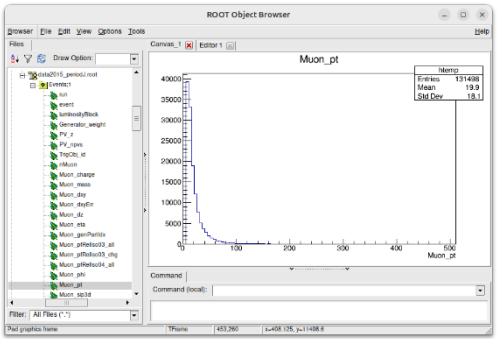
Goal: to validate our conversion and analysis by reproducing the established Higgs boson observation (m_{4l} peak at 125 GeV) using ATLAS Open Data

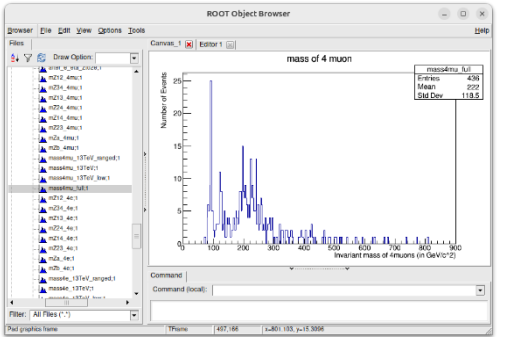
1. Event selection

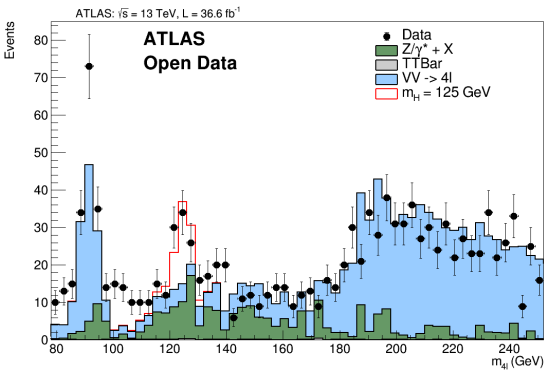
2. Generate histograms

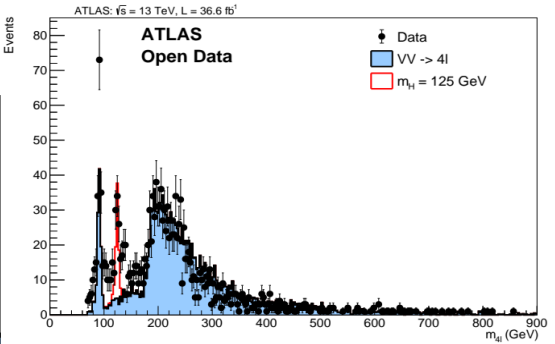
3. Scale MC to Data

4. Interpret the result



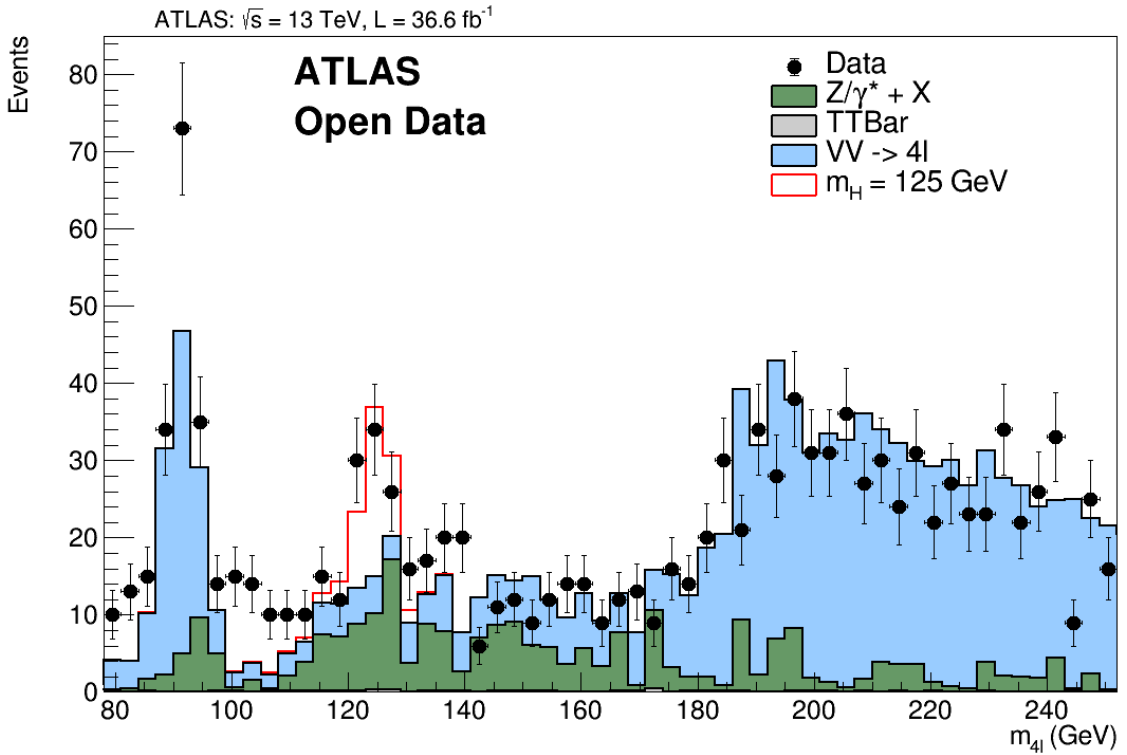






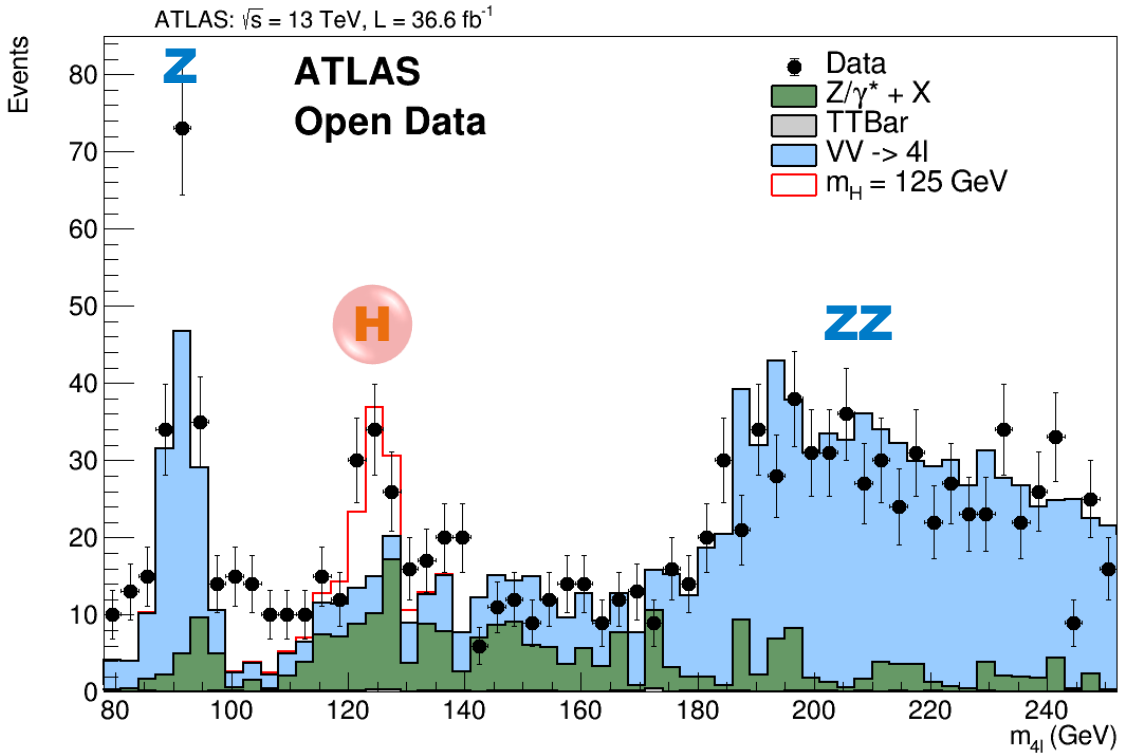
Here comes the final plot:-)

(maybe not really final)



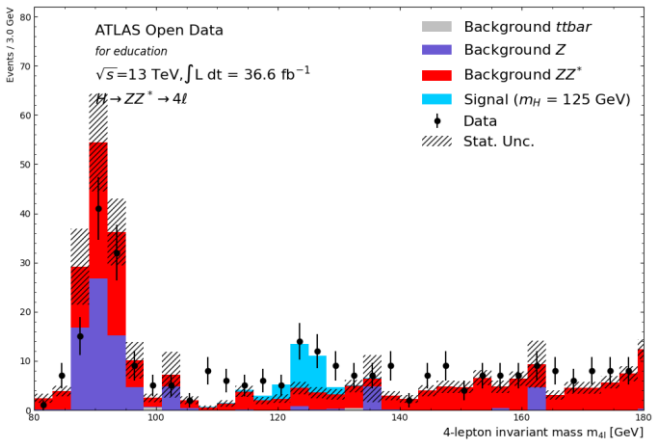
Here comes the final plot:-)

(maybe not really final)



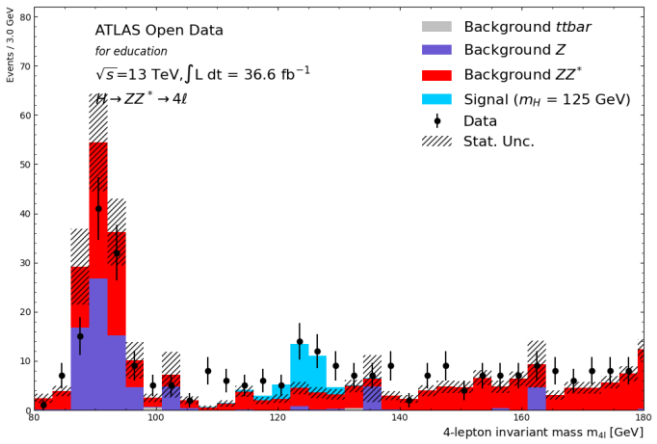
Promised comparison

Atlas tutorial

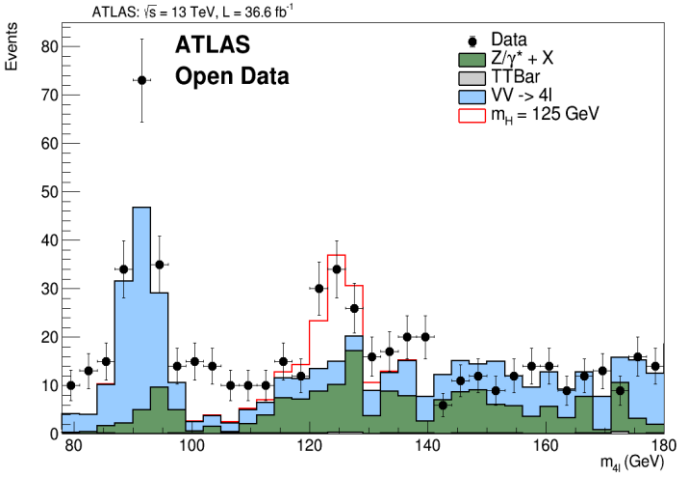


Promised comparison

Atlas tutorial

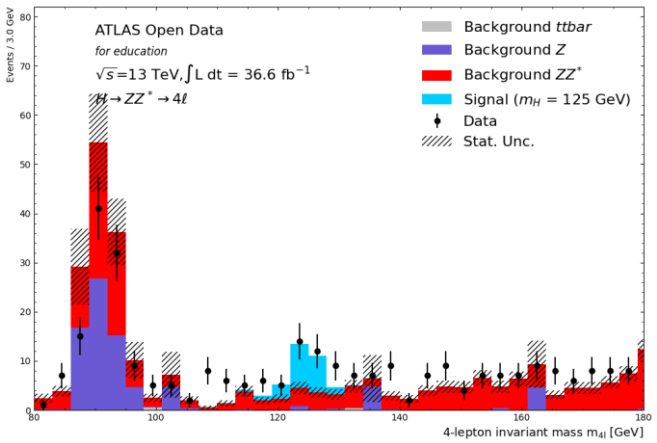


Converted data

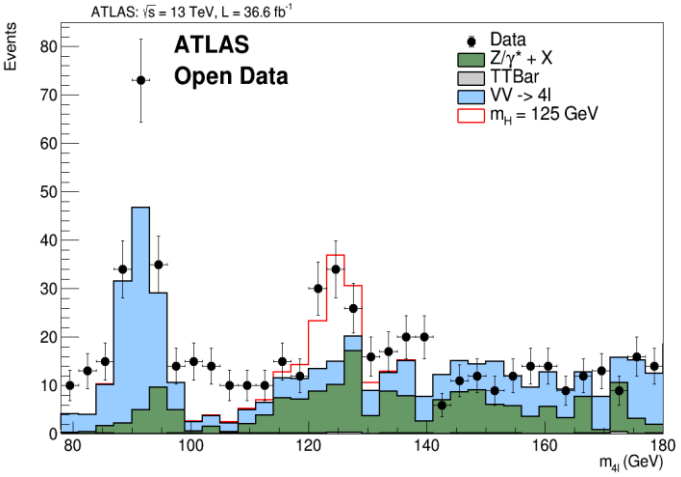


Promised comparison

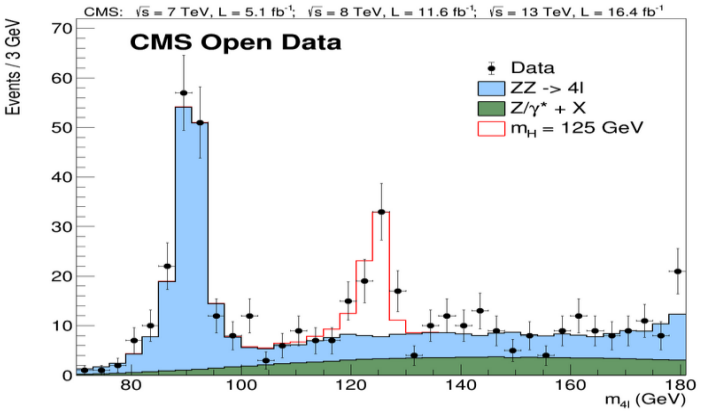
Atlas tutorial



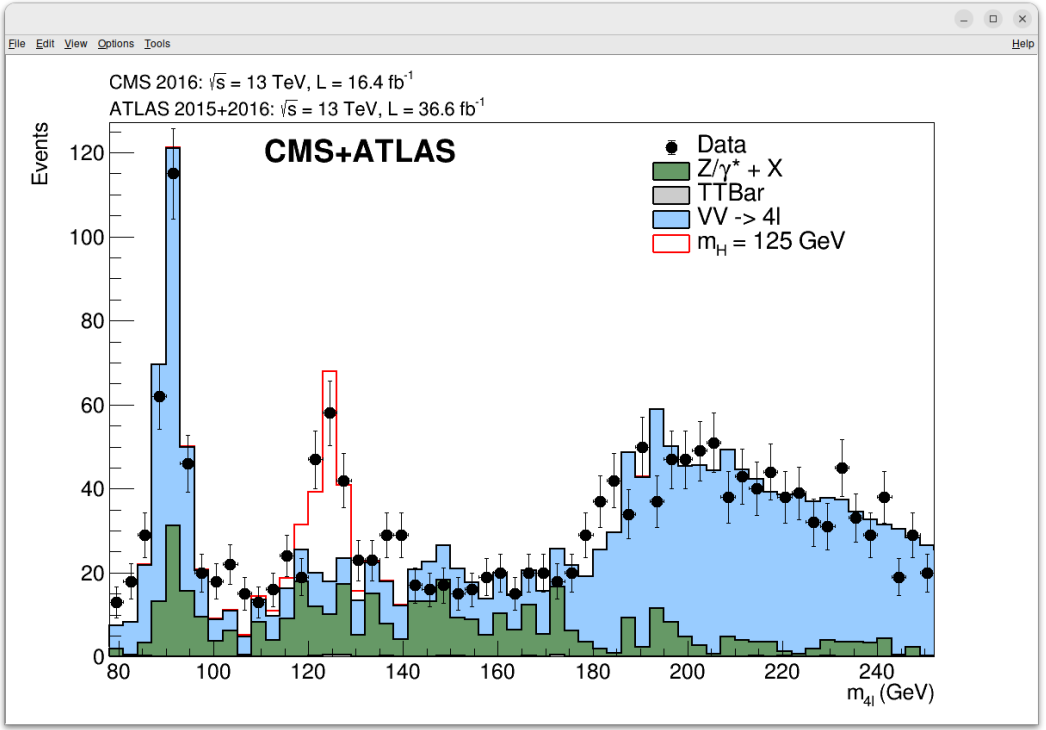
Converted data



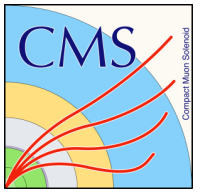
CMS data(from PUNCH)



Results.



- Higgs was found!
- Using converted research-0-level Open Data
- Same code was implemented to both, ATLAS and CMS data



VIELEN DANK

any questions?

HELMHOLTZ
DESY.

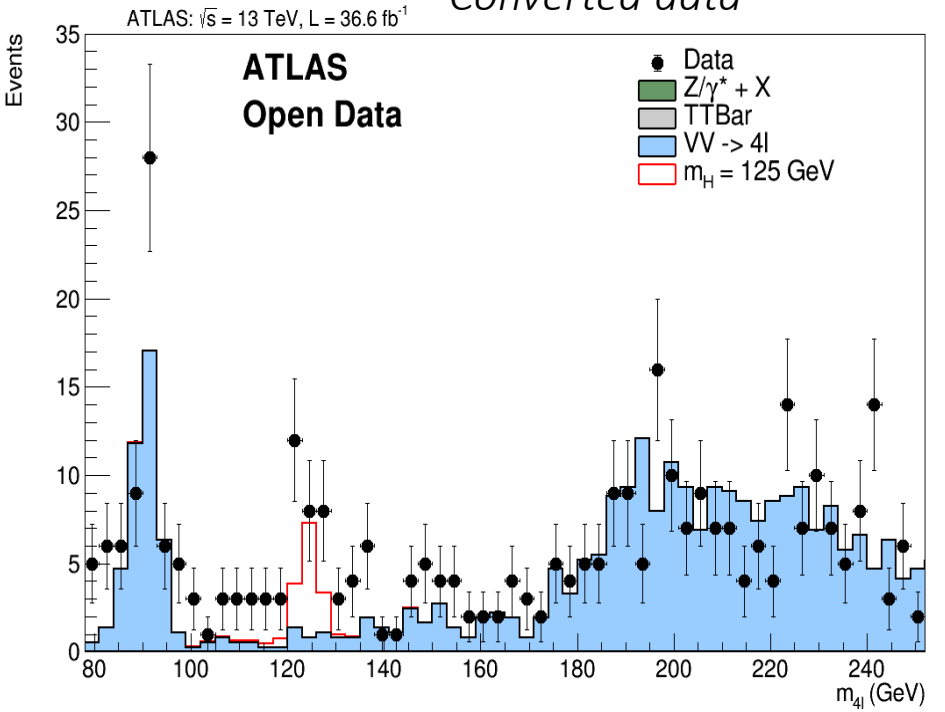
References

[1] - https://opendata.atlas.cern/docs/documentation/data_format/physlite
[2] - https://opendata.cern/record/93910
[3] - https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/HZZAnalysis.ipynb
[4] - https://gitlab.cern.ch/atlas-outreach-data-tools/physlitetoopendata

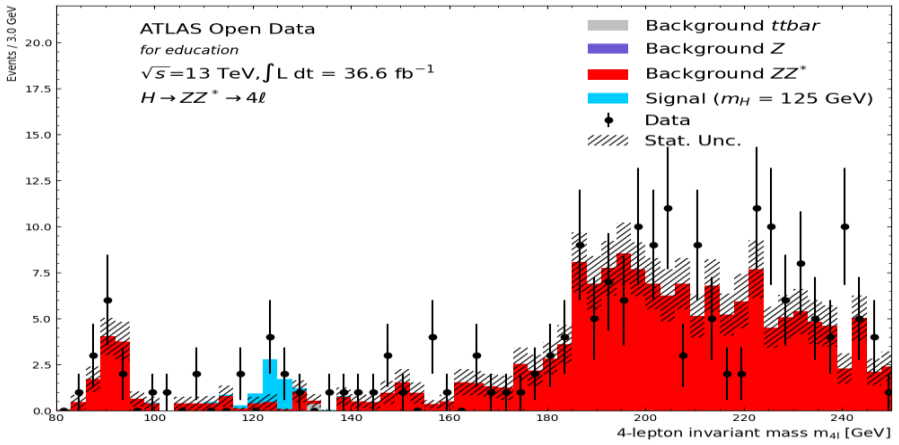
Backup

4μ channel

Converted data



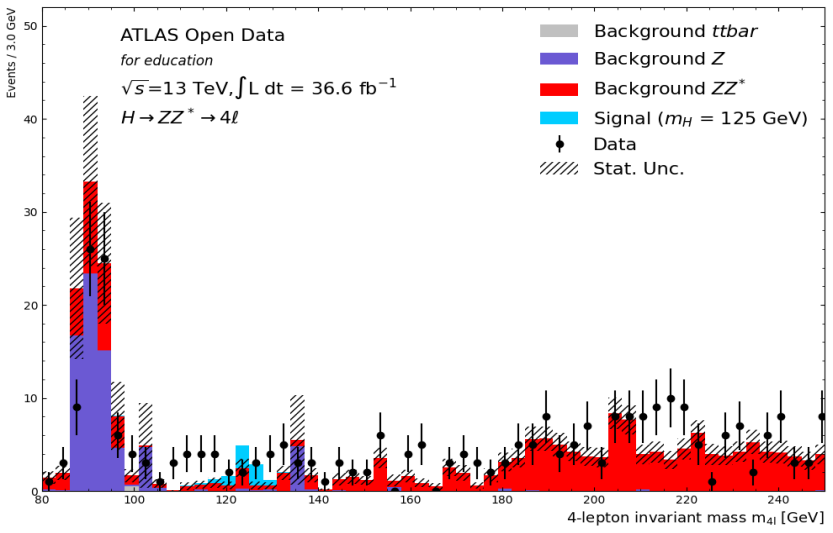
Atlas tutorial



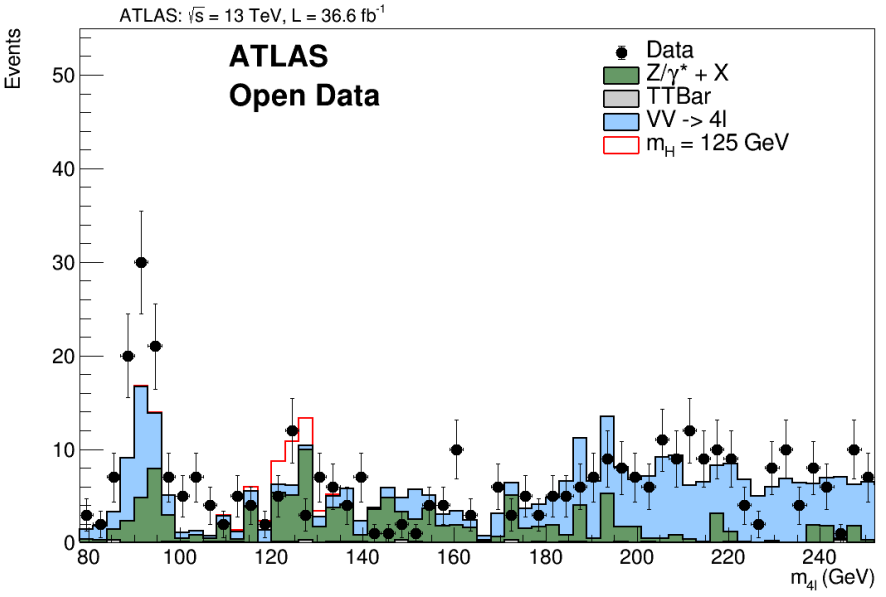
Backup

4e channel

Atlas tutorial



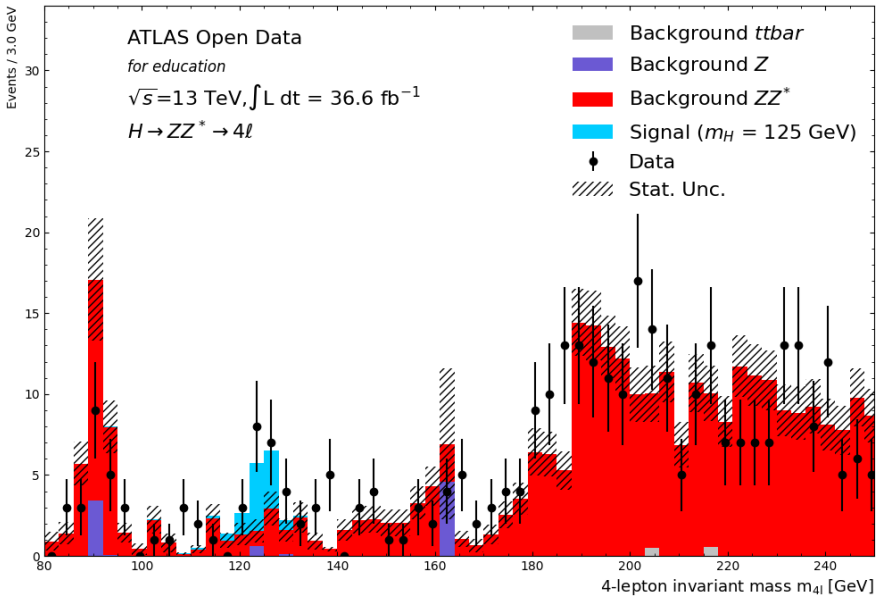
Converted data



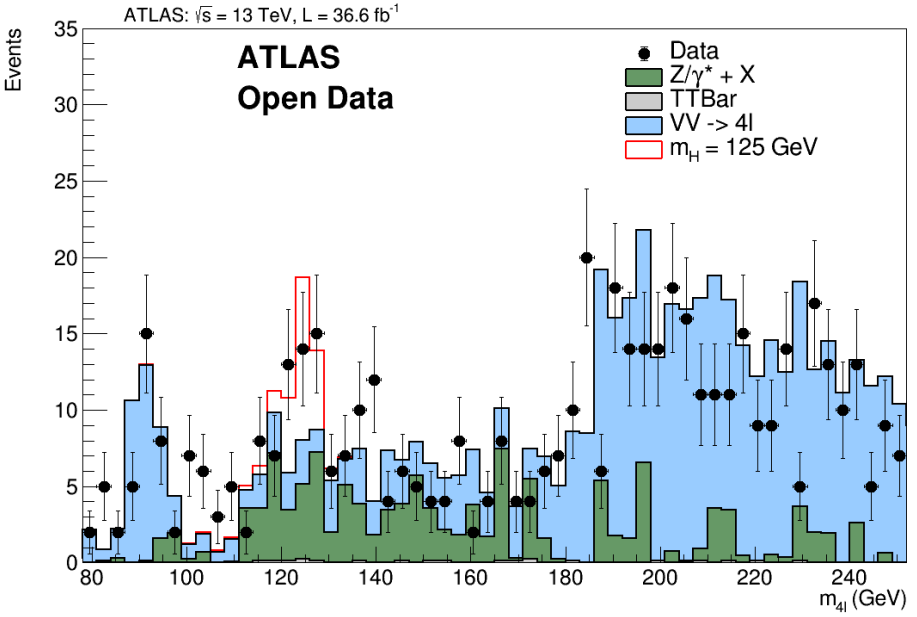
Backup

2μ2e channel

Atlas tutorial



Converted data



Cuts for H→4L analysis

Quality Cuts:
- Muon_isGlobal = true
- Muon_isPFcand = true
- Electron_isPFcand = true

Event Selection:
- QL1 + QL2 + QL3 + QL4 = 0

Selection and Kinematic Cuts:
- p_T > 7 GeV and |η|<2.5 for electrons
- p_T > 5 GeV and |η|<2.4 for muons
- | D_{xy} | < 0.5 and | D_z |<1.0 for electrons+muons
- Rel. Iso < 0.4 for electrons and muons
- Lost Hits <= 1 for electrons

Electron/muon pairs must arise from same primary vertex: | SIP_{3D} |<4.0

For the final event selection two pairs of same flavour, opposite charge leptons are formed.
Possible combination: 4μ and 4e and 2μ2e

For the pair (Z_a → 2L) with invariant mass closest to Z boson mass:
- $Q_{L_i} + Q_j = 0$
- p_{T,L_i} > 20.0 GeV and p_{T,L_j}> 10.0 GeV
- 40.0 GeV < M_{Z_a} < 120.0 GeV

For the other pair (Z_b → 2L), we require:
- $Q_{L_i} + Q_j = 0$
- p_{T,L_i} > 20.0 GeV and p_{T,L_j}> 10.0 GeV
- 12.0 GeV < M_{Z_b} < 120.0 GeV

Backup

Scaling VV to 4l