DSL workshop

TTree::Draw

- Some were not aware of this DSL ☺
- Some not aware of its power
- Some liked the conciseness and feature set (collection handling)
- Some reported that they knew of at least one analysis that was carried entirely in TTree::Draw ..

RDataFrame

- Some read it as ‘R language DataFrame’
- was well received (and at least one DSL implementation is using RDataFrame as the underlying engine – Andrea Rizzi)
- need/interest of having RDataFrame support systematics. For example introduce a node that has a wiggle function that end up duplicating the further nodes and ends with multiple 'results’.
  - Enrico said:

```
df = CreateDataFrame()
for s in sistematics:
      ApplyTheWholeGraph(df.Define("b_with_syst", some_func_of_s(s));
```

- Really need strong collection support …

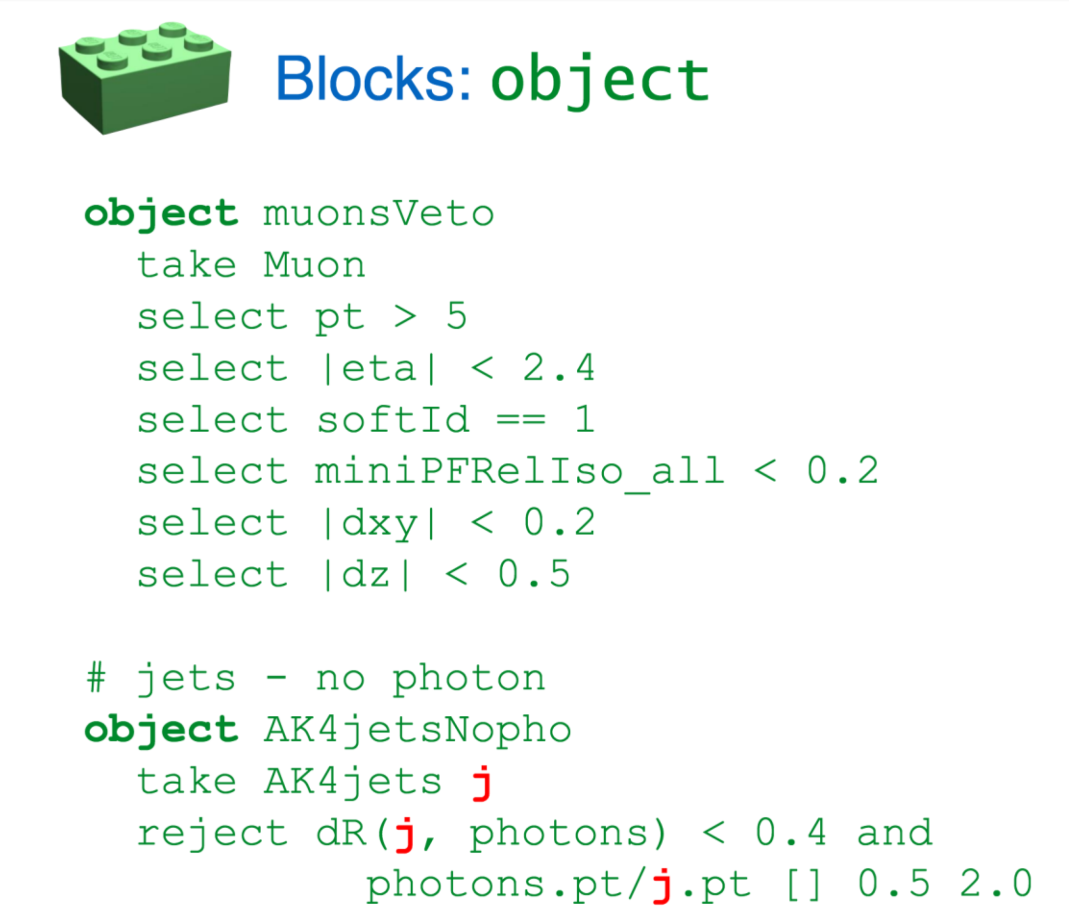
Workshop

- Group of people interested a priori by DSL.
- Hopes for DSL
  Basic requirements:
  - Public
  - Complete
  - Easily learned
  - Demonstrably correct

  Desirable features:
  - Self-contained
  - General programming language-independent
  - Analysis framework-independent

”Options”

- Related to the “Les Houches Analysis Description Accord”:
  - lhada2rivet
  - adl2tnm
  - lhada2checkmate
  - CutLang

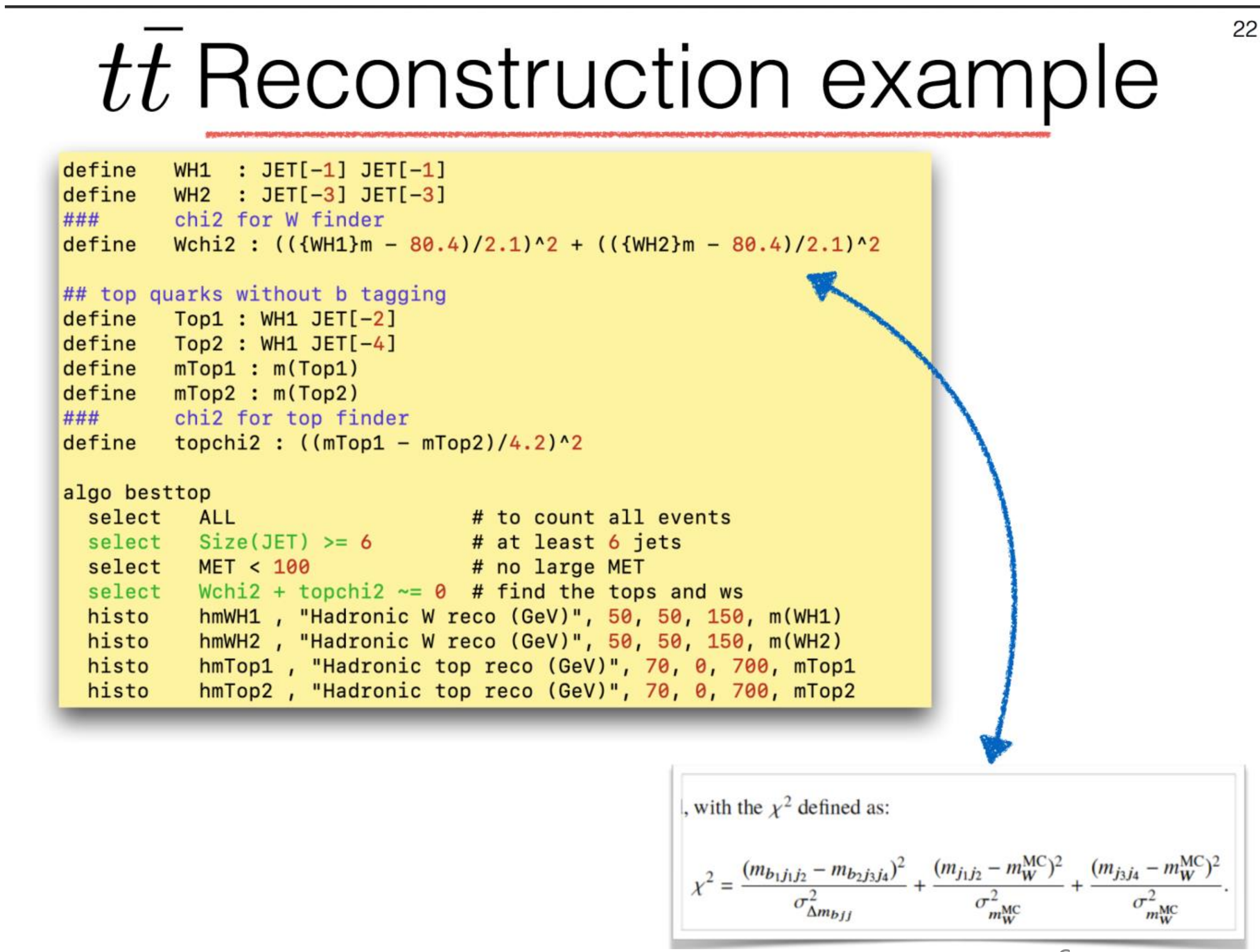


Blocks: object

```
object muonsVeto
  take Muon
  select pt > 5
  select |eta| < 2.4
  select softId == 1
  select miniPFRelIso_all < 0.2
  select |dxy| < 0.2
  select |dz| < 0.5

# jets - no photon
object AK4jetsNopho
  take AK4jets j
  reject dR(j, photons) < 0.4 and
           photons.pt/j.pt [] 0.5 2.0
```

CutLang v2

- yacc, lexx, ROOT LorentzVector and Histograms
- https://indico.cern.ch/event/769263/contributions/3406040/attachments/1838631/3014722/ADLCutLangrazorboost.pdf



$t\bar{t}$ Reconstruction example

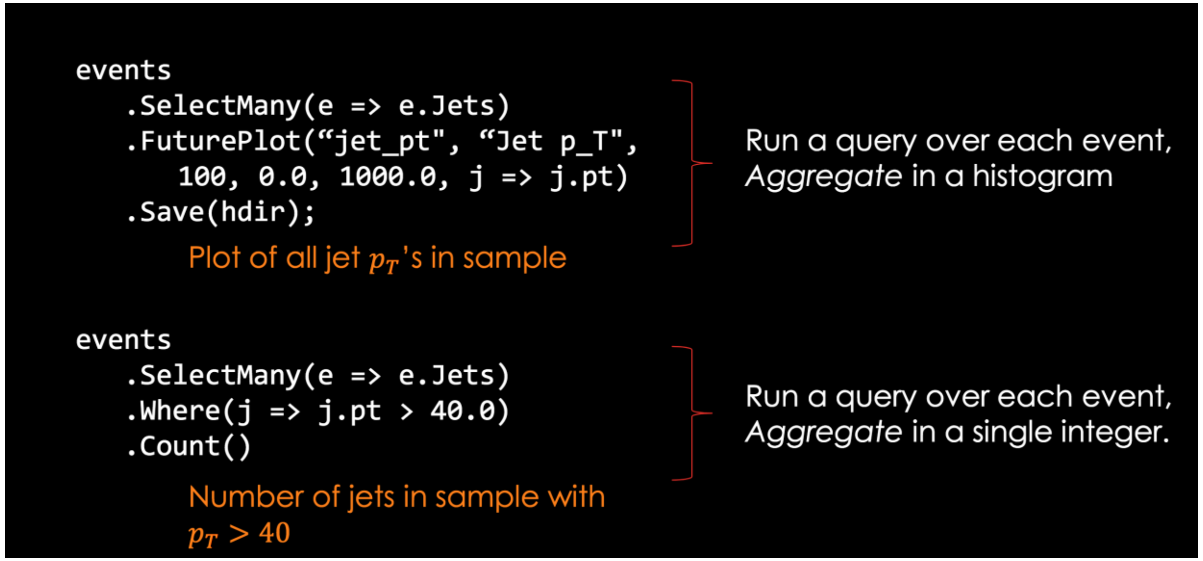
```
define   WH1  : JET[-1] JET[-1]
define   WH2  : JET[-3] JET[-3]
###      chi2 for W finder
define   Wchi2 : (({WH1}m - 80.4)/2.1)^2 + (({WH2}m - 80.4)/2.1)^2

## top quarks without b tagging
define   Top1 : WH1 JET[-2]
define   Top2 : WH1 JET[-4]
define   mTop1 : m(Top1)
define   mTop2 : m(Top2)
###      chi2 for top finder
define   topchi2 : ((mTop1 - mTop2)/4.2)^2

algo besttop
  select   ALL                      # to count all events
  select   Size(JET) >= 6           # at least 6 jets
  select   MET < 100                # no large MET
  select   Wchi2 + topchi2 ~= 0     # find the tops and ws
  histo    hmWH1 , "Hadronic W reco (GeV)", 50, 50, 150, m(WH1)
  histo    hmWH2 , "Hadronic W reco (GeV)", 50, 50, 150, m(WH2)
  histo    hmTop1 , "Hadronic top reco (GeV)", 70, 0, 700, mTop1
  histo    hmTop2 , "Hadronic top reco (GeV)", 70, 0, 700, mTop2
```

, with the χ^2 defined as:

$$\chi^2 = \frac{(m_{b_1j_1j_2} - m_{b_2j_3j_4})^2}{\sigma^2_{\Delta m_{bjj}}} + \frac{(m_{j_1j_2} - m_{\mathrm{W}}^{\mathrm{MC}})^2}{\sigma^2_{m_{\mathrm{W}}^{\mathrm{MC}}}} + \frac{(m_{j_3j_4} - m_{\mathrm{W}}^{\mathrm{MC}})^2}{\sigma^2_{m_{\mathrm{W}}^{\mathrm{MC}}}}.$$

Others

- lhada2rivet
  - Proof of concept for automatic ADL interpretation and code generation
  - Creates ‘rivet’ code.
- LINQ
  - SQL + C#
- YADL / FAST
  - YAML (superset of json with object description, anchors, references, etc..)
  - Python, uproot, numexpr



```
events
    .SelectMany(e => e.Jets)
    .FuturePlot(“jet_pt", “Jet p_T",
        100, 0.0, 1000.0, j => j.pt)
    .Save(hdir);
```

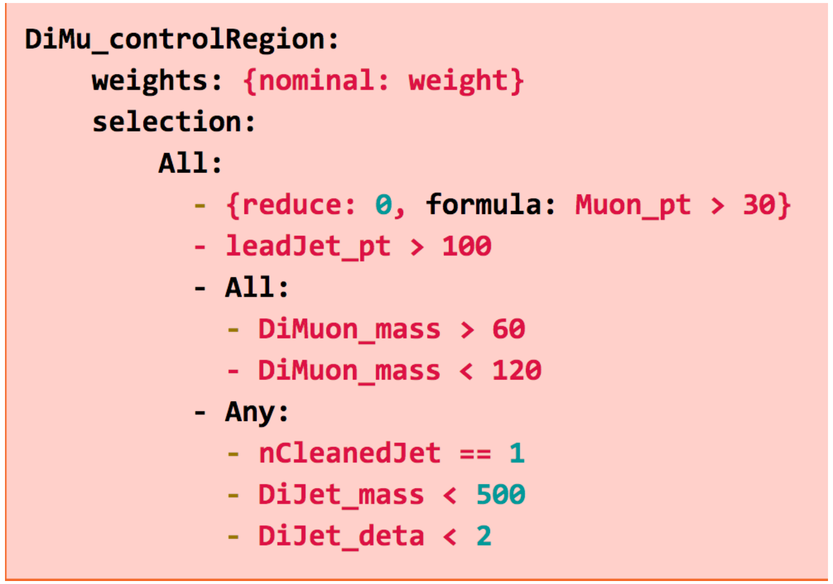
Plot of all jet p_T's in sample

Run a query over each event, *Aggregate* in a histogram

```
events
    .SelectMany(e => e.Jets)
    .Where(j => j.pt > 40.0)
    .Count()
```

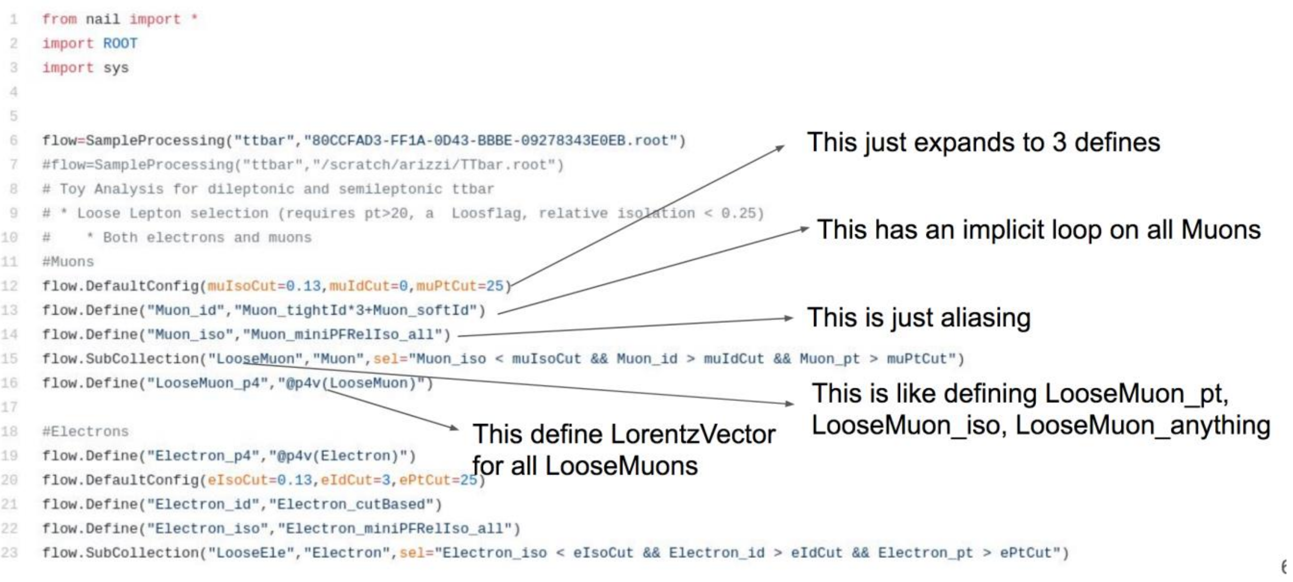
Number of jets in sample with $p_T > 40$

Run a query over each event, *Aggregate* in a single integer.



```
DiMu_controlRegion:
    weights: {nominal: weight}
    selection:
        All:
          - {reduce: 0, formula: Muon_pt > 30}
          - leadJet_pt > 100
          - All:
            - DiMuon_mass > 60
            - DiMuon_mass < 120
          - Any:
            - nCleanedJet == 1
            - DiJet_mass < 500
            - DiJet_deta < 2
```

- NAIL, Andrea Rizzi
  - Python
  - RDataFrame under the hood
  - Plan for ML interface
- AEACuS & RHADAManTHUS



```python
from nail import *
import ROOT
import sys


flow=SampleProcessing("ttbar","80CCFAD3-FF1A-0D43-BBBE-09278343E0EB.root")
#flow=SampleProcessing("ttbar","/scratch/arizzi/TTbar.root")
# Toy Analysis for dileptonic and semileptonic ttbar
# * Loose Lepton selection (requires pt>20, a  Loosflag, relative isolation < 0.25)
#    * Both electrons and muons
#Muons
flow.DefaultConfig(muIsoCut=0.13,muIdCut=0,muPtCut=25)
flow.Define("Muon_id","Muon_tightId*3+Muon_softId")
flow.Define("Muon_iso","Muon_miniPFRelIso_all")
flow.SubCollection("LooseMuon","Muon",sel="Muon_iso < muIsoCut && Muon_id > muIdCut && Muon_pt > muPtCut")
flow.Define("LooseMuon_p4","@p4v(LooseMuon)")

#Electrons
flow.Define("Electron_p4","@p4v(Electron)")
flow.DefaultConfig(eIsoCut=0.13,eIdCut=3,ePtCut=25)
flow.Define("Electron_id","Electron_cutBased")
flow.Define("Electron_iso","Electron_miniPFRelIso_all")
flow.SubCollection("LooseEle","Electron",sel="Electron_iso < eIsoCut && Electron_id > eIdCut && Electron_pt > ePtCut")
```

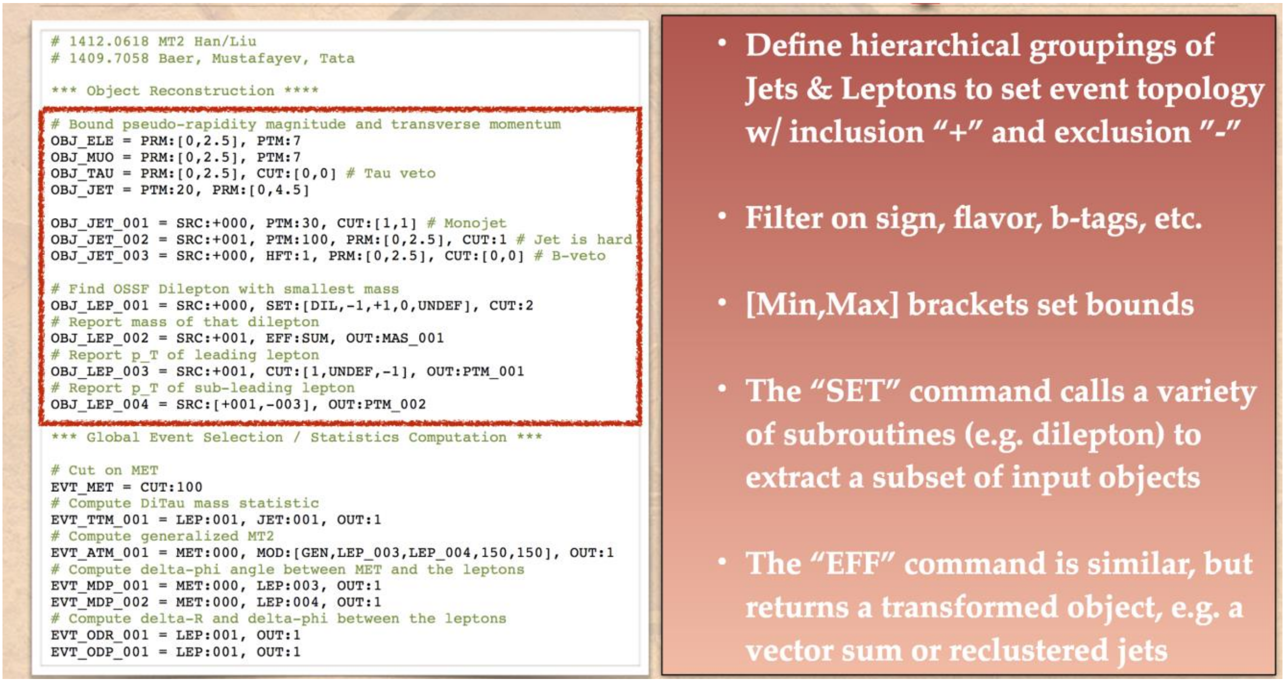
This just expands to 3 defines

This has an implicit loop on all Muons

This is just aliasing

This is like defining LooseMuon_pt, LooseMuon_iso, LooseMuon_anything

This define LorentzVector for all LooseMuons



```
# 1412.0618 MT2 Han/Liu
# 1409.7058 Baer, Mustafayev, Tata

*** Object Reconstruction ****

# Bound pseudo-rapidity magnitude and transverse momentum
OBJ_ELE = PRM:[0,2.5], PTM:7
OBJ_MUO = PRM:[0,2.5], PTM:7
OBJ_TAU = PRM:[0,2.5], CUT:[0,0] # Tau veto
OBJ_JET = PTM:20, PRM:[0,4.5]

OBJ_JET_001 = SRC:+000, PTM:30, CUT:[1,1] # Monojet
OBJ_JET_002 = SRC:+001, PTM:100, PRM:[0,2.5], CUT:1 # Jet is hard
OBJ_JET_003 = SRC:+000, HFT:1, PRM:[0,2.5], CUT:[0,0] # B-veto

# Find OSSF Dilepton with smallest mass
OBJ_LEP_001 = SRC:+000, SET:[DIL,-1,+1,0,UNDEF], CUT:2
# Report mass of that dilepton
OBJ_LEP_002 = SRC:+001, EFF:SUM, OUT:MAS_001
# Report p_T of leading lepton
OBJ_LEP_003 = SRC:+001, CUT:[1,UNDEF,-1], OUT:PTM_001
# Report p_T of sub-leading lepton
OBJ_LEP_004 = SRC:[+001,-003], OUT:PTM_002
*** Global Event Selection / Statistics Computation ***

# Cut on MET
EVT_MET = CUT:100
# Compute DiTau mass statistic
EVT_TTM_001 = LEP:001, JET:001, OUT:1
# Compute generalized MT2
EVT_ATM_001 = MET:000, MOD:[GEN,LEP_003,LEP_004,150,150], OUT:1
# Compute delta-phi angle between MET and the leptons
EVT_MDP_001 = MET:000, LEP:003, OUT:1
EVT_MDP_002 = MET:000, LEP:004, OUT:1
# Compute delta-R and delta-phi between the leptons
EVT_ODR_001 = LEP:001, OUT:1
EVT_ODP_001 = LEP:001, OUT:1
```

- **Define hierarchical groupings of Jets & Leptons to set event topology w/ inclusion "+" and exclusion "-"**
- **Filter on sign, flavor, b-tags, etc.**
- **[Min,Max] brackets set bounds**
- **The "SET" command calls a variety of subroutines (e.g. dilepton) to extract a subset of input objects**
- **The "EFF" command is similar, but returns a transformed object, e.g. a vector sum or reclustered jets**

- How to build your own language tutorial ….

- CERN Analysis Preservation … can be done for ADL based analysis

THE CONVERSATION

- There is a huge amount of activity around Analysis and Query Languages
    - See HSF Data Analysis Forum, CHEP, ACAT, IRIS-HEP
- Think Big
    - The context for Run 3 and Run 4 is much bigger than we are used to
    - Can we do a full analysis with a small team?
    - Scalability?
    - An Analysis System, not just an ADL!