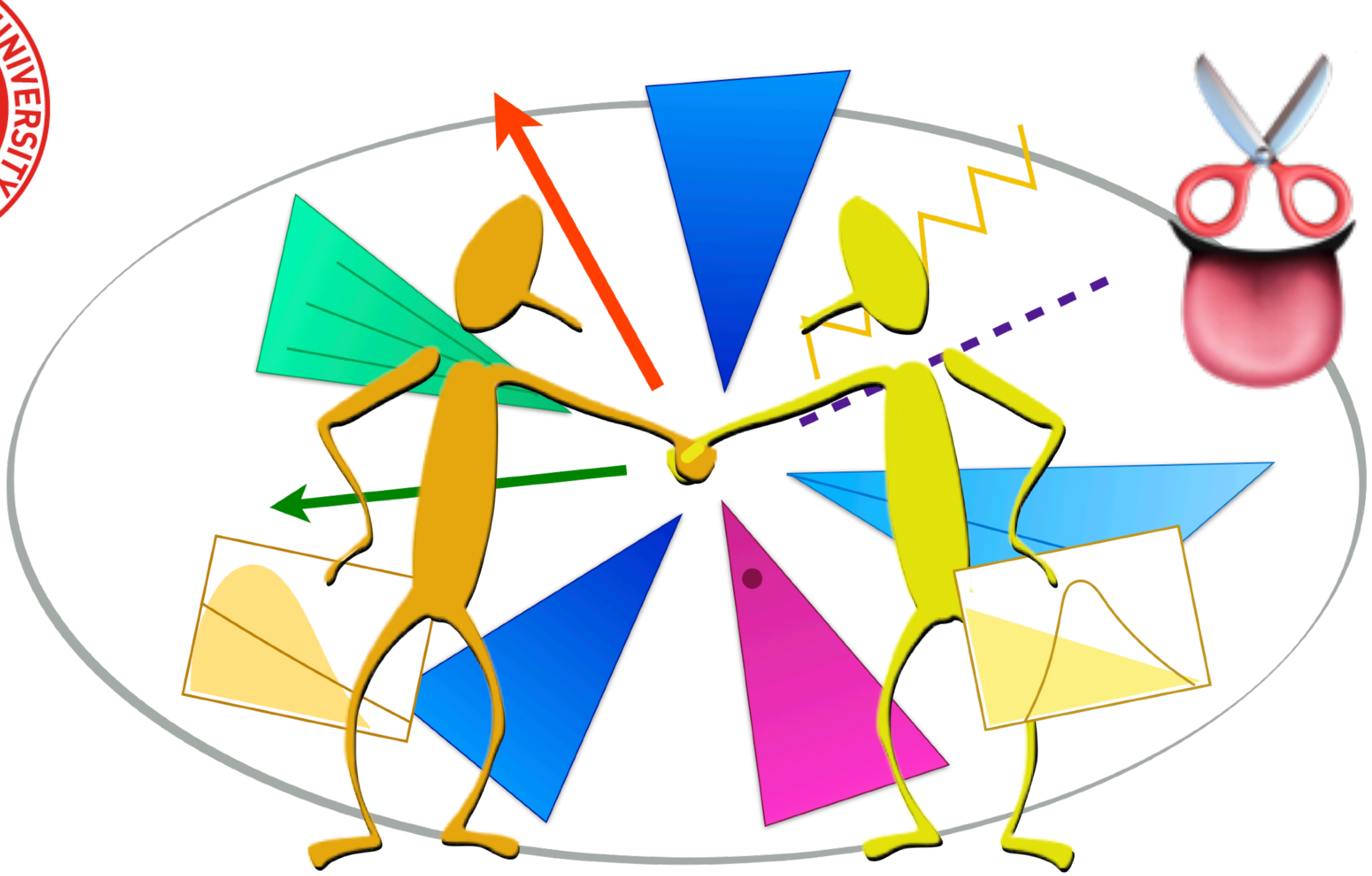


Creative uses of ADL to push boundaries of HEP analysis

Sezen Sekmen (KNU), Gökhan Ünel (UC Irvine)

Analysis Description Language Tutorial & Hackathon
21-22 Nov 2022, Kyungpook National University, Center for HEP

ADL helps to design and document a single analysis in a clear and organized way.

But its distinguishing strength is in navigating and exploring the multi-analysis landscape.

Beyond running on events: Static analysis

ADL organizes analysis information in a well-defined structure using standard syntax rules.

—> We can use this feature for tasks other than running on events.

Static analysis : Parsing and analyzing a code to deduce facts about it without actually executing the code.

Use the ADL files to take a stroll in the multi-analysis landscape.

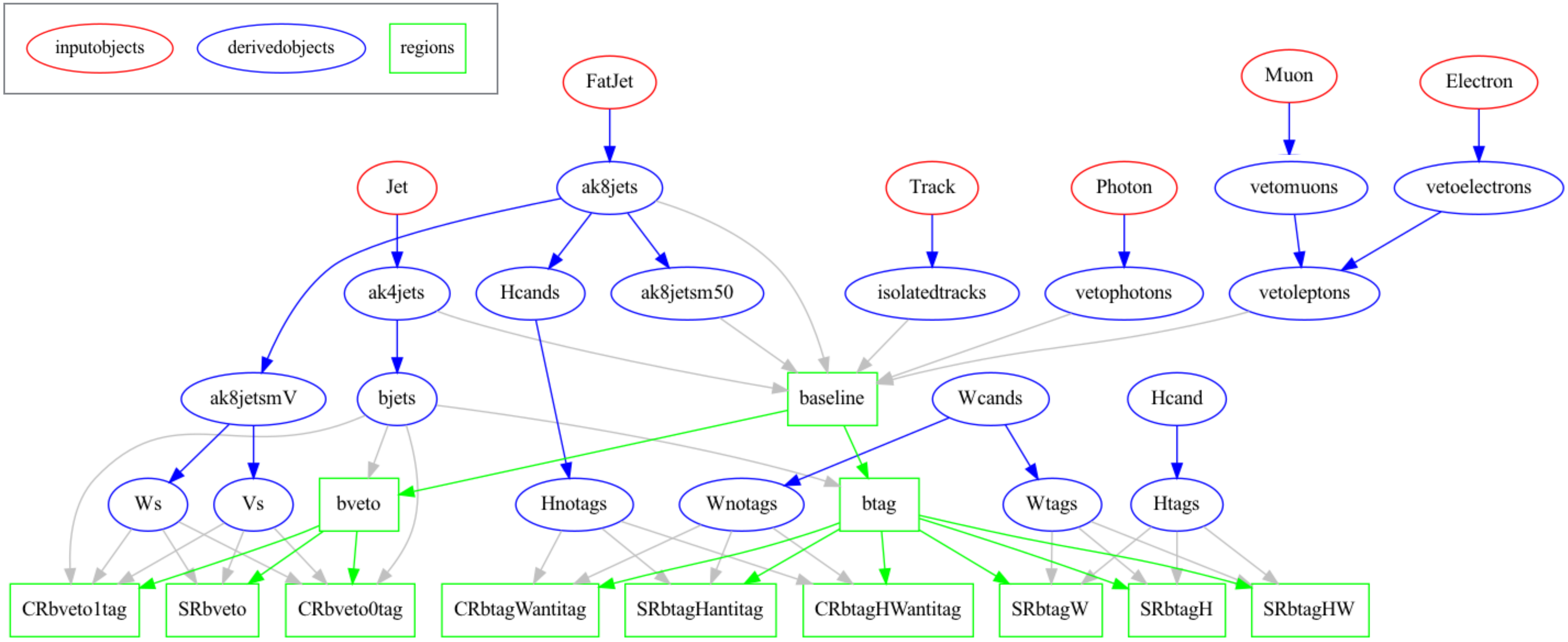
Visualize analyses

We can build tools to convert ADL descriptions into useful visualizations of analyses:

- Automatically build graphs / flowcharts of complete analyses :
    - Graphviz-based Infrastructure ~ready for building such graphs from ADL file input (Burak Şen).
        - objects, regions, bins as nodes —> summary view, full detail view, clickable, …
        - Can edit colors, shapes, …

- Automatically create latex tables for single or multiple analyses
    - Latex tables listing object selections, event selections, cutflows, results , etc.

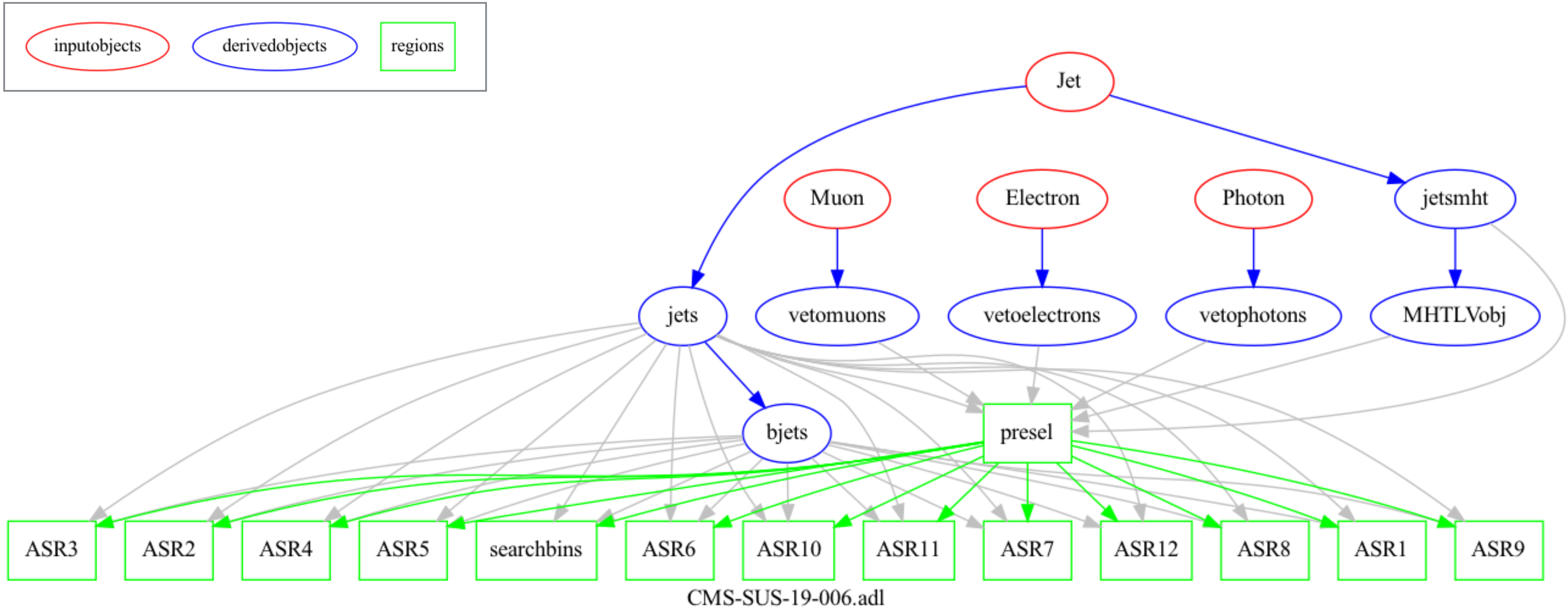
Graph analyses via graphviz



CMS-SUS-21-002.adl

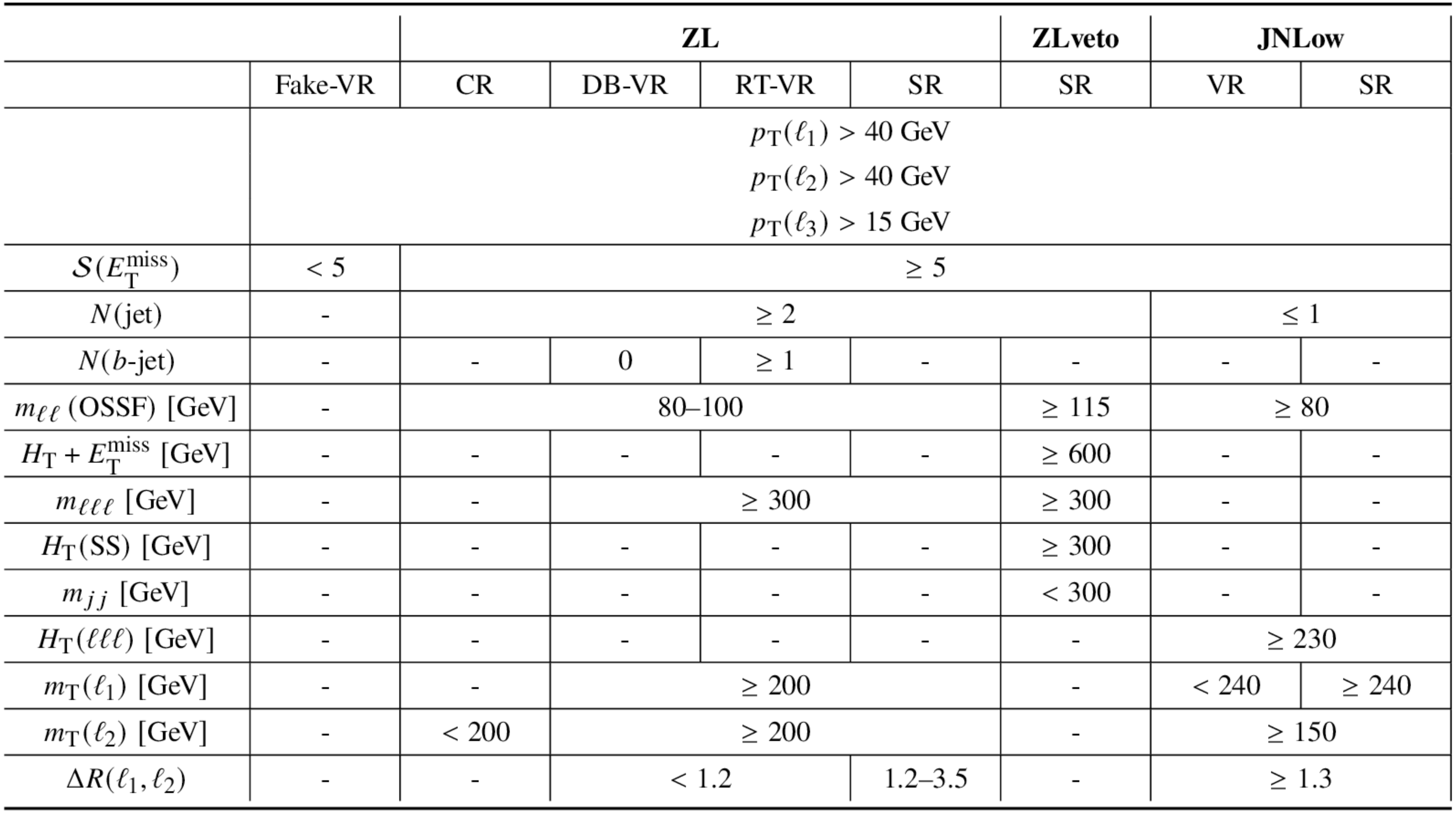
SUS-21-002: Search for EWK SUSY in WW, WZ and WH hadronic final states

Graph analyses via graphviz



SUS-19-006: Search for SUSY with jets and MHT

Automatically convert regions to tables



		ZL				ZLveto	JNLow	
	Fake-VR	CR	DB-VR	RT-VR	SR	SR	VR	SR
	$p_{\mathrm{T}}(\ell_1) > 40$ GeV $p_{\mathrm{T}}(\ell_2) > 40$ GeV $p_{\mathrm{T}}(\ell_3) > 15$ GeV							
$\mathcal{S}(E_{\mathrm{T}}^{\mathrm{miss}})$	< 5	≥ 5						
$N(\mathrm{jet})$	-	≥ 2					≤ 1	
$N(b\text{-jet})$	-	-	0	≥ 1	-	-	-	-
$m_{\ell\ell}$ (OSSF) [GeV]	-	80–100				≥ 115	≥ 80	
$H_{\mathrm{T}} + E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	-	-	-	-	-	≥ 600	-	-
$m_{\ell\ell\ell}$ [GeV]	-	-	≥ 300			≥ 300	-	-
$H_{\mathrm{T}}(\mathrm{SS})$ [GeV]	-	-	-	-	-	≥ 300	-	-
m_{jj} [GeV]	-	-	-	-	-	< 300	-	-
$H_{\mathrm{T}}(\ell\ell\ell)$ [GeV]	-	-	-	-	-	-	≥ 230	
$m_{\mathrm{T}}(\ell_1)$ [GeV]	-	-	≥ 200			-	< 240	≥ 240
$m_{\mathrm{T}}(\ell_2)$ [GeV]	-	< 200	≥ 200			-	≥ 150	
$\Delta R(\ell_1, \ell_2)$	-	-	< 1.2		1.2–3.5	-	≥ 1.3	

Produce such tables automatically from the analysis ADL file.

Will come soon.

ATLAS-EXOT-2020-02: Search for type-III seesaw heavy leptons in leptonic final states

Query analyses

Our analyses carry a huge amount of information.

ADL helps to organize this information into a queryable database:

- Which analyses use electrons with mvaFall17V2noIso_WP90?
  Which require miniPFRelIso_all < 0.1?
- Which analyses use MT? Which use HLT_MyFavoriteTrigger?
- Which analyses use a cut of MET > at least 600 ?
- Which analyses use boosted Ws? How are they defined?
- …

Tools with of varying sophistication can perform such queries: (development in progress)

- even “grep” !
- Python scripts.
- more sophisticated tools based on formal grammar parsing.

Combining analyses

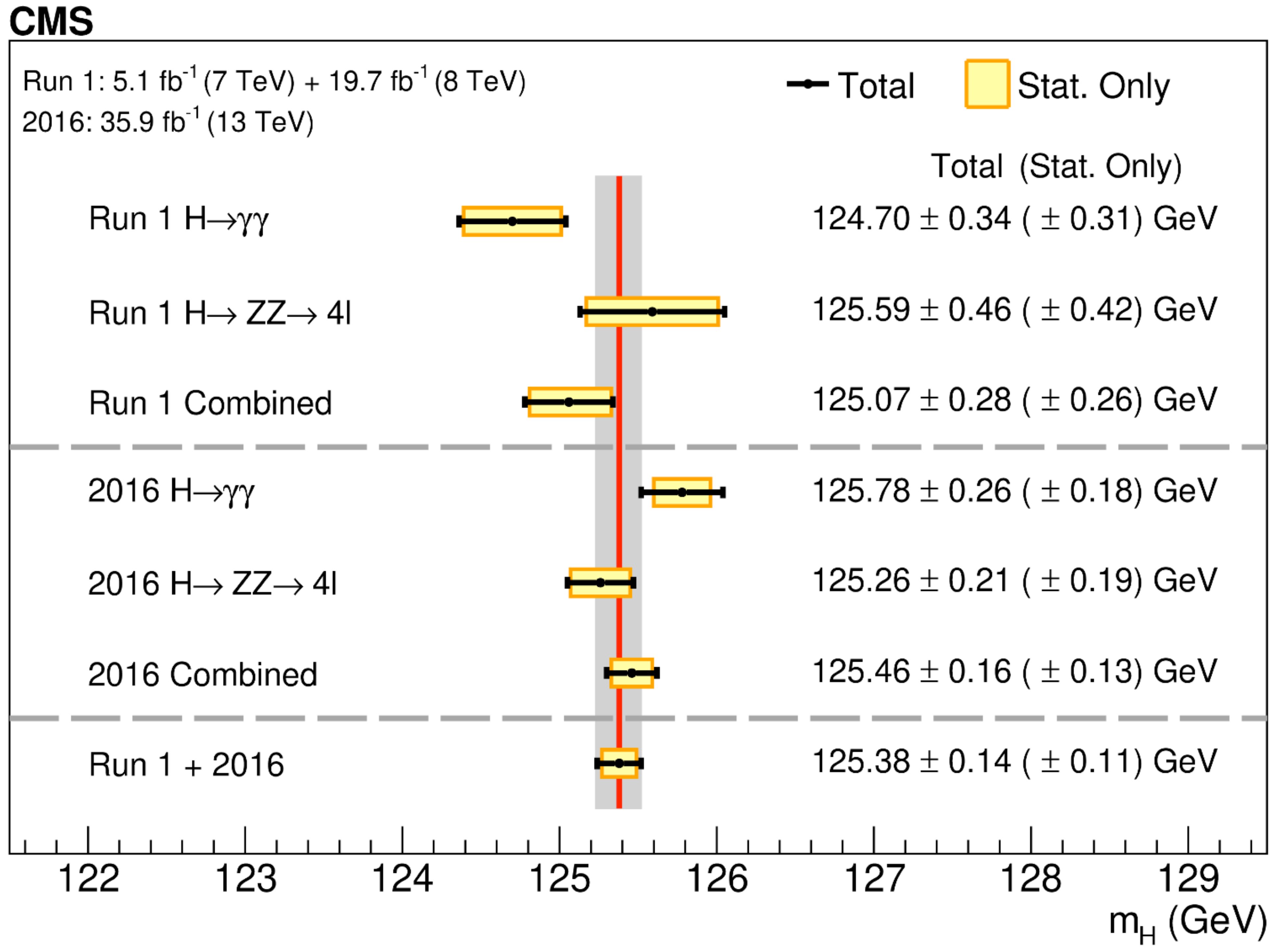
We usually perform analyses in different search channels, then statistically combine them to enhance sensitivity.

One caveat: Analyses selections must be disjoint / non-overlapping / uncorrelated to be combined.

Different Higgs analyses are easily designed to be disjoint, for each production and decay channel: h -> ZZ, h -> γγ , etc.

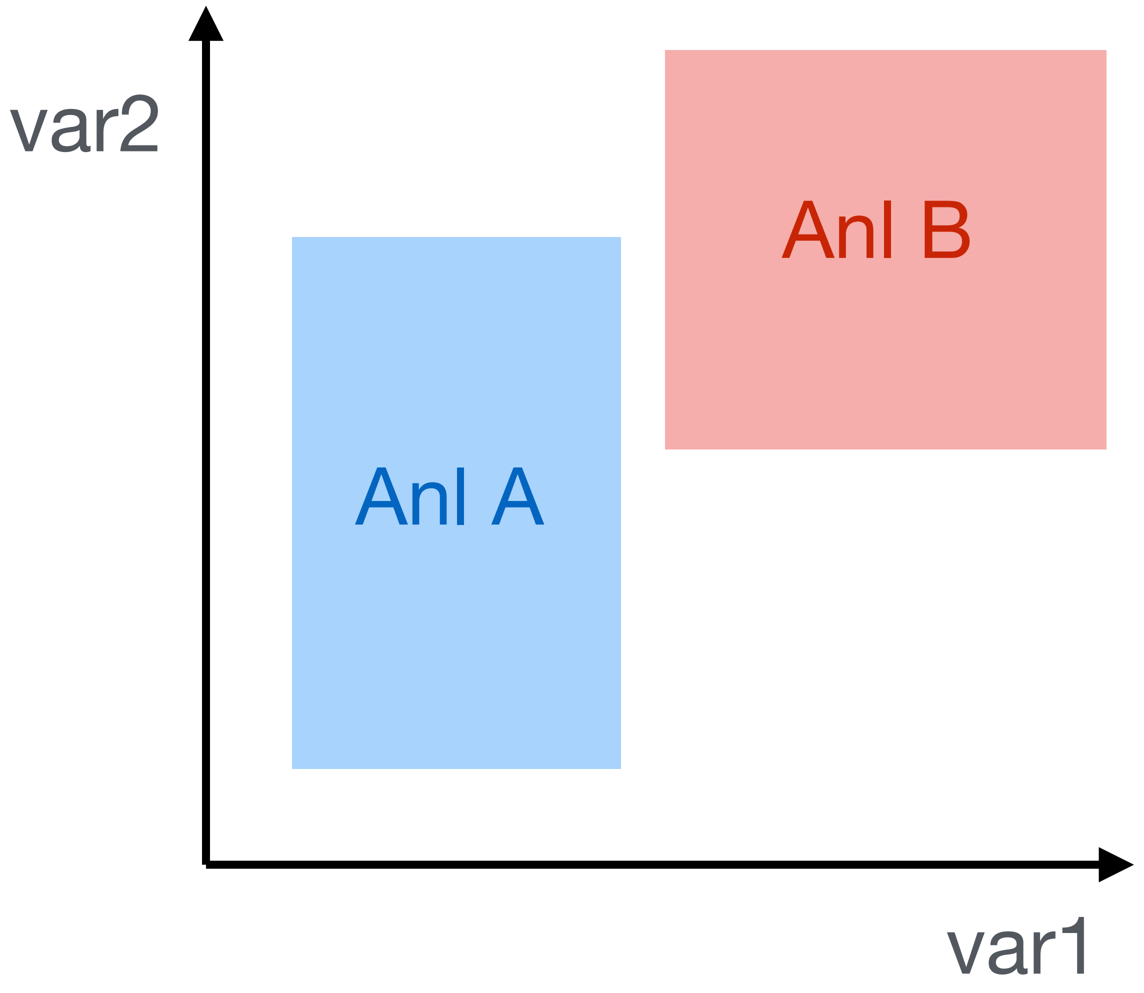
New physics analyses are harder to make disjoint. There are many candidate new particles with many production and decay processes. A lot of analyses overlap with each other.

Either find non-overlapping analyses for combination or design analyses in a non-overlapping way.



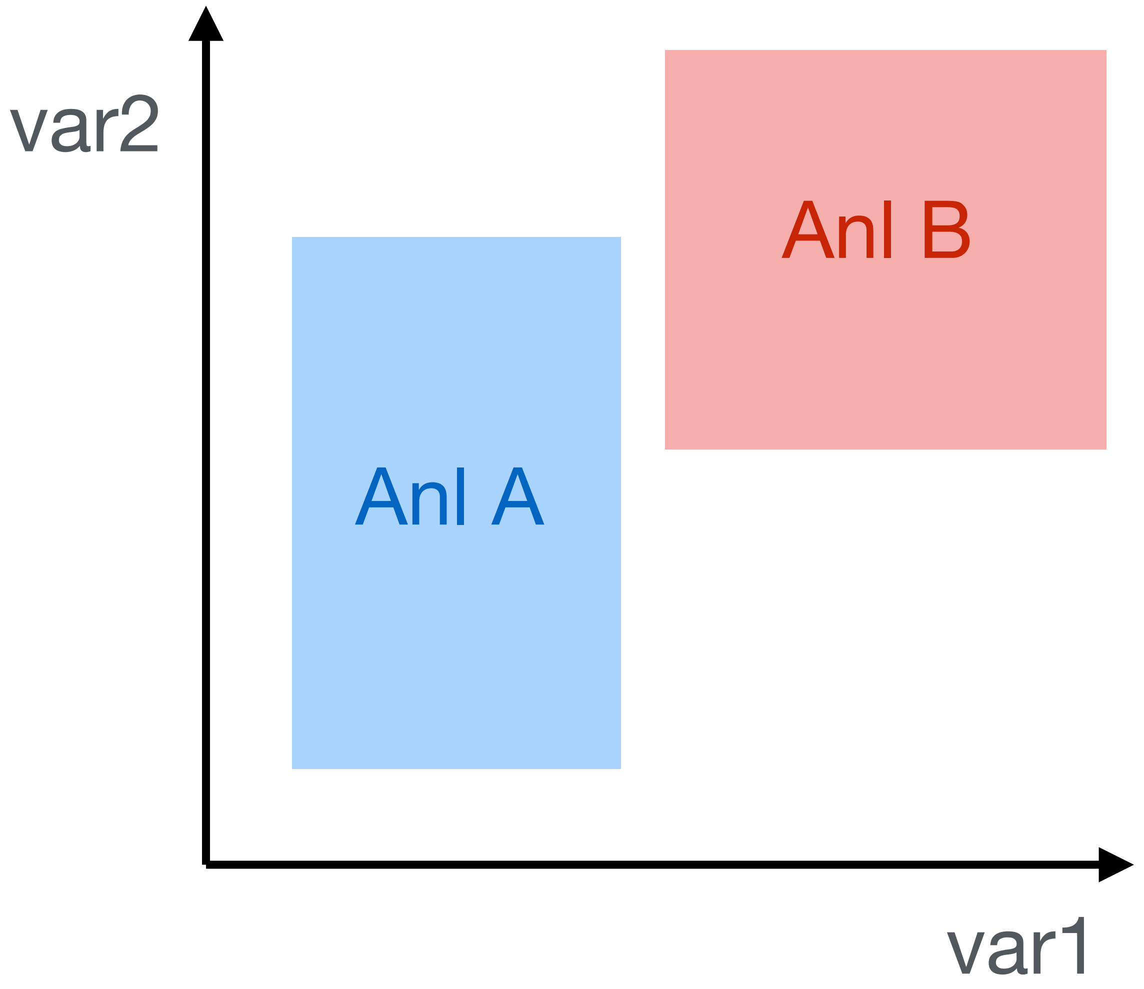
Combined result is better than individual results.

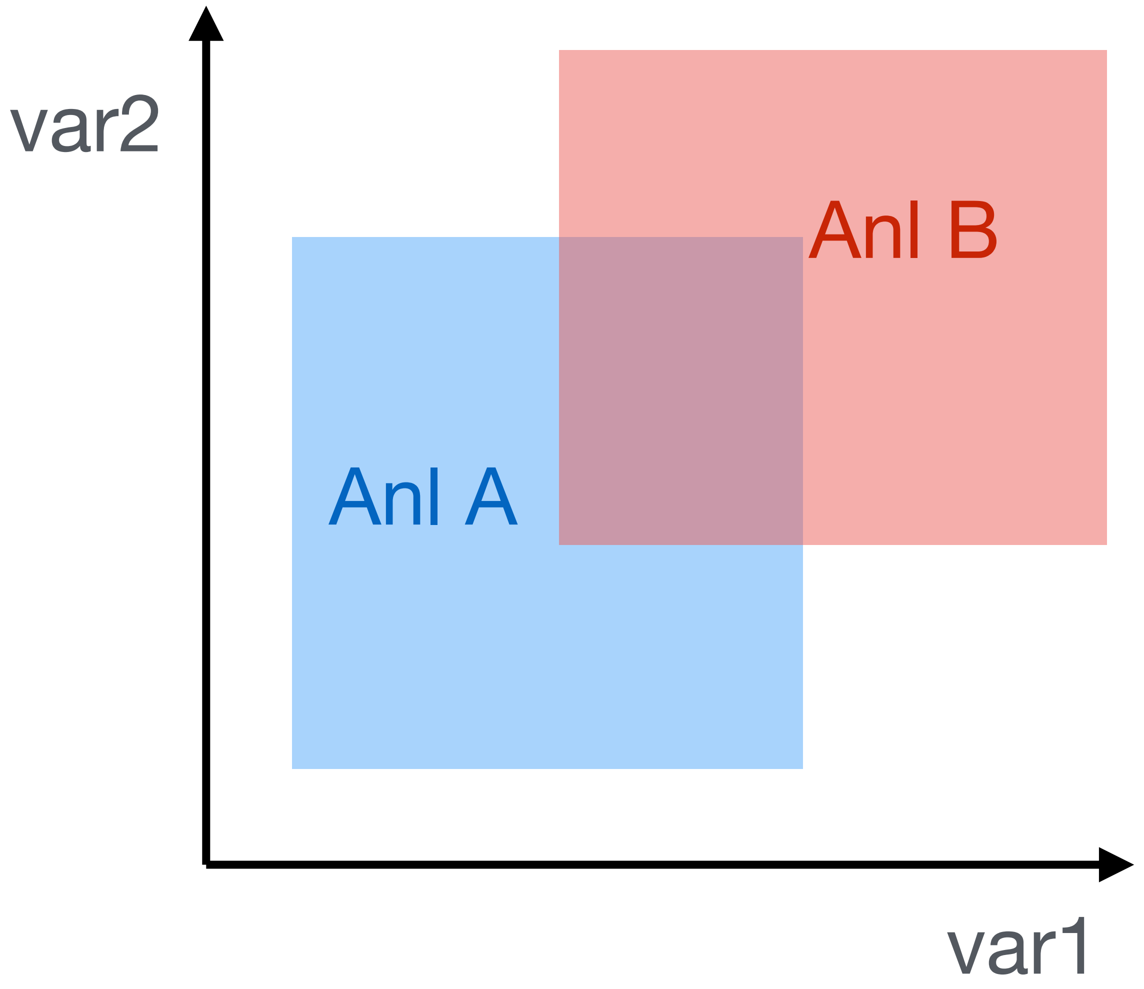
Analysis overlaps



Anl A and Anl B disjoint.
Can combine.

Analysis overlaps

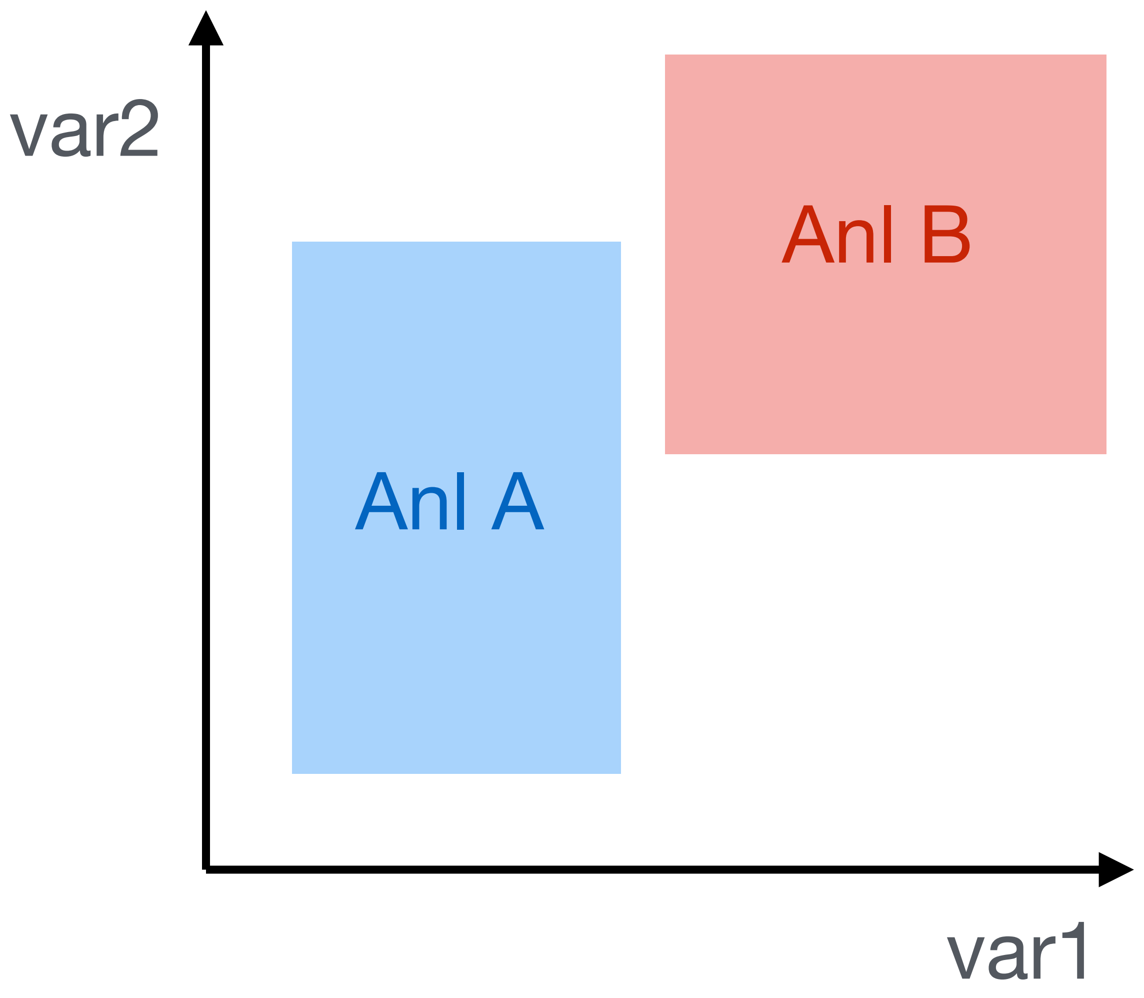




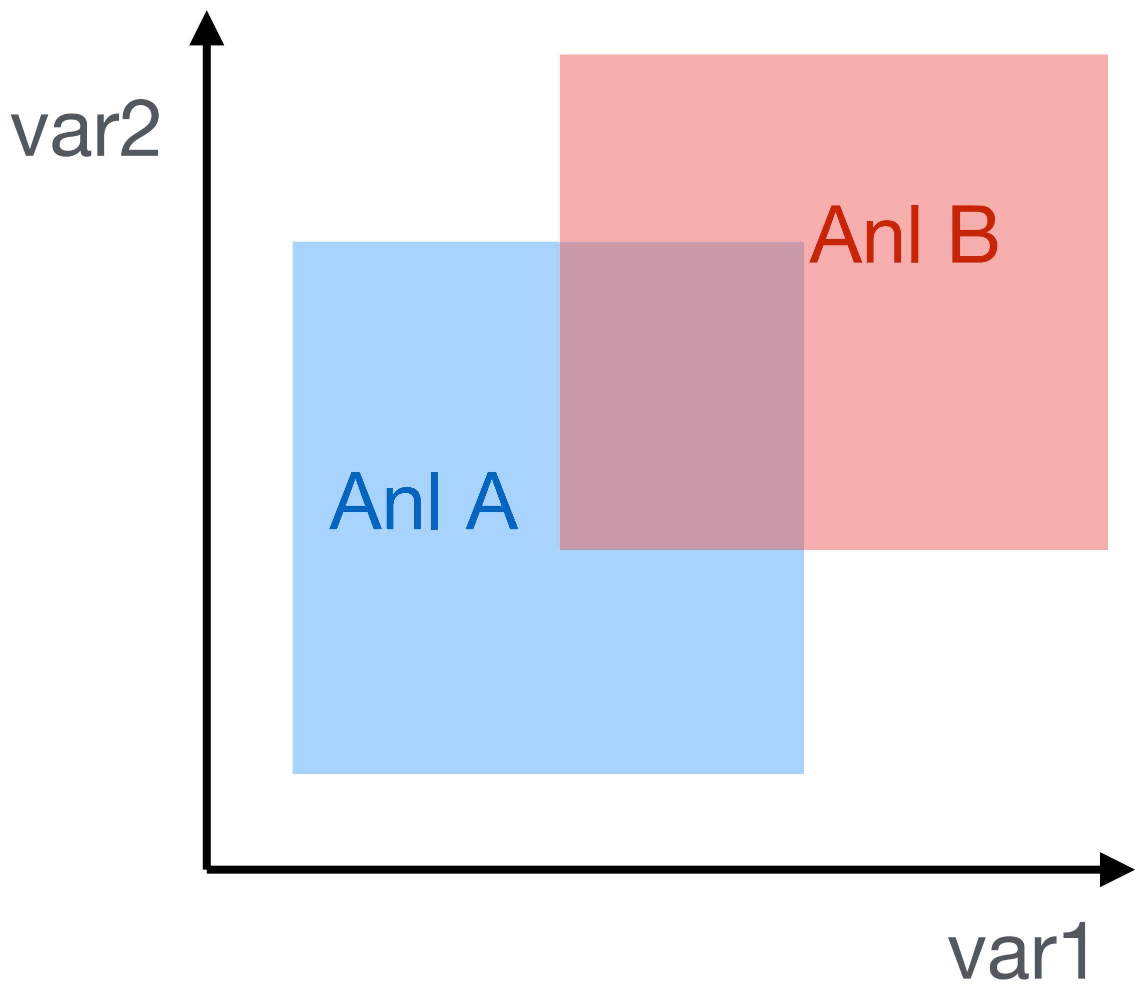
Anl A and Anl B disjoint.
Can combine.

Anl A and Anl B overlap.
Cannot combine.

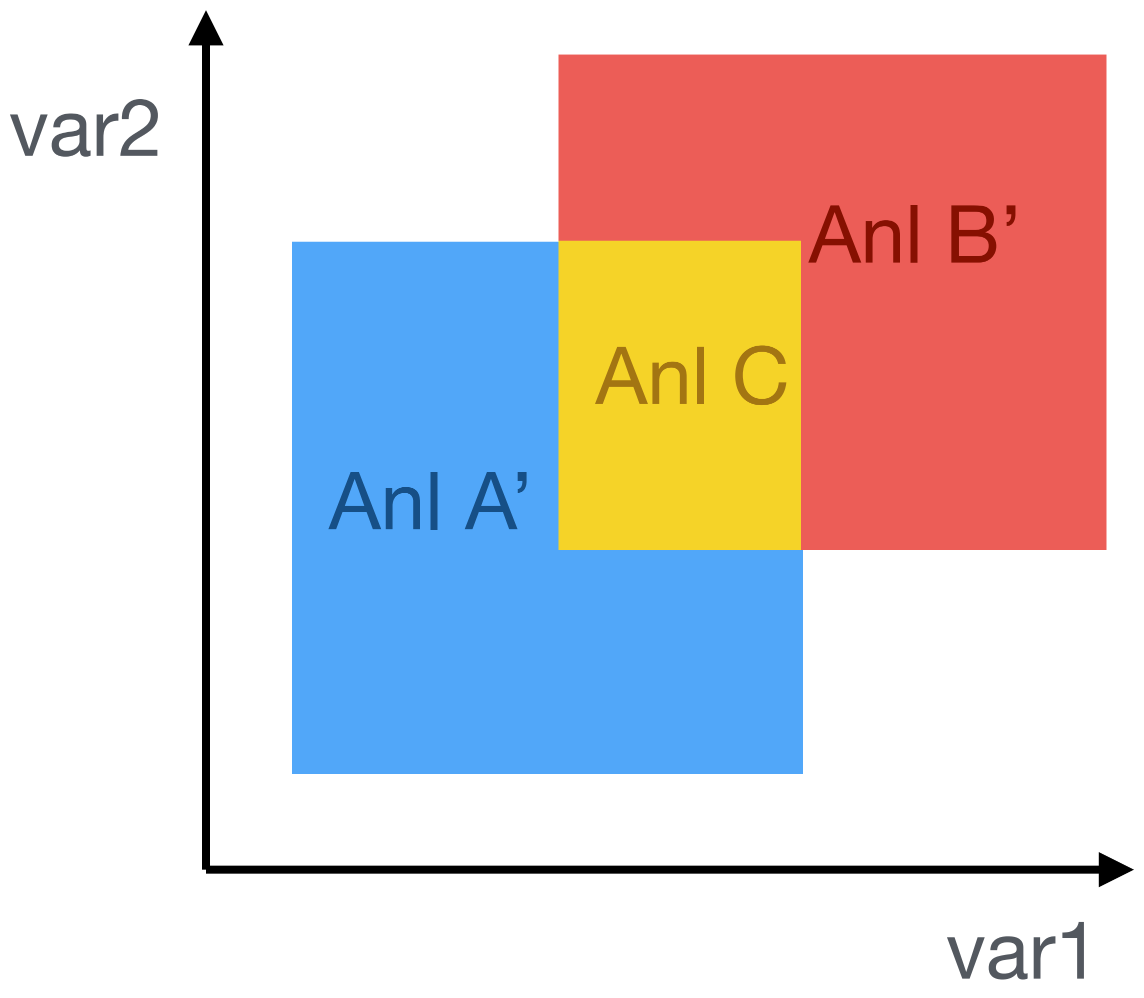
Analysis overlaps



Anl A and Anl B disjoint.
Can combine.



Anl A and Anl B overlap.
Cannot combine.



Use Anl A and B ADLs to determine the boundaries of the overlap region.

Partition phase space into non-overlapping fragments and rerun analysis.

Compare/combine analyses

Standard ADL structure and syntax makes it usable for comparing analysis selections:

- Determine analysis overlaps, identify disjoint analyses or search regions.
- Find the feasible combinations with maximal sensitivity.
- Automate large scale combinations of analyses.

Different levels of comparisons:

- Compare by eye.
- Compare by static analysis of ADL files using scripts.
- Compare by running on events.
- Compare by random sampling of the analysis variables phase space.
- …

Tools can automate advanced comparisons.

- Study started in LH19 (with Harrison Prosper & Wolfgang Waltenberger) (arXiv:2002.12220, contribution 17).

Analyzing excesses in multiple analyses

LHC did not report a big discovery for new physics BUT

ATLAS and CMS are observing various small excesses (i.e. deviations from the SM) in a number of analyses.

These analyses look for different types of new particles, in different final states.

Use ADL to compare object and event selections in these analyses:

- Are the excesses in regions that have overlapping selections?
- Are the excess regions combinable?
- Are there any complementary final states that are not yet explored?

Analysis design

Database of existing analyses can help designing the new ones

- Provide a learning database.
- Answer relevant questions using query, comparison and visualization tools:
  - Which final states did the existing analyses look at?
  - Which final states are unexplored?
  - How much overlap exists between my analysis and the existing ones?

Rapid prototyping, structural thinking, optimization:

- Test and document large numbers of alternative selections in parallel in an organized manner.
- Motivates a more mathematical and organized conception of analyses.
- ADL is a suitable medium for implementing differentiable programming.

How would YOU use ADL?