Machine Enabled Discovery

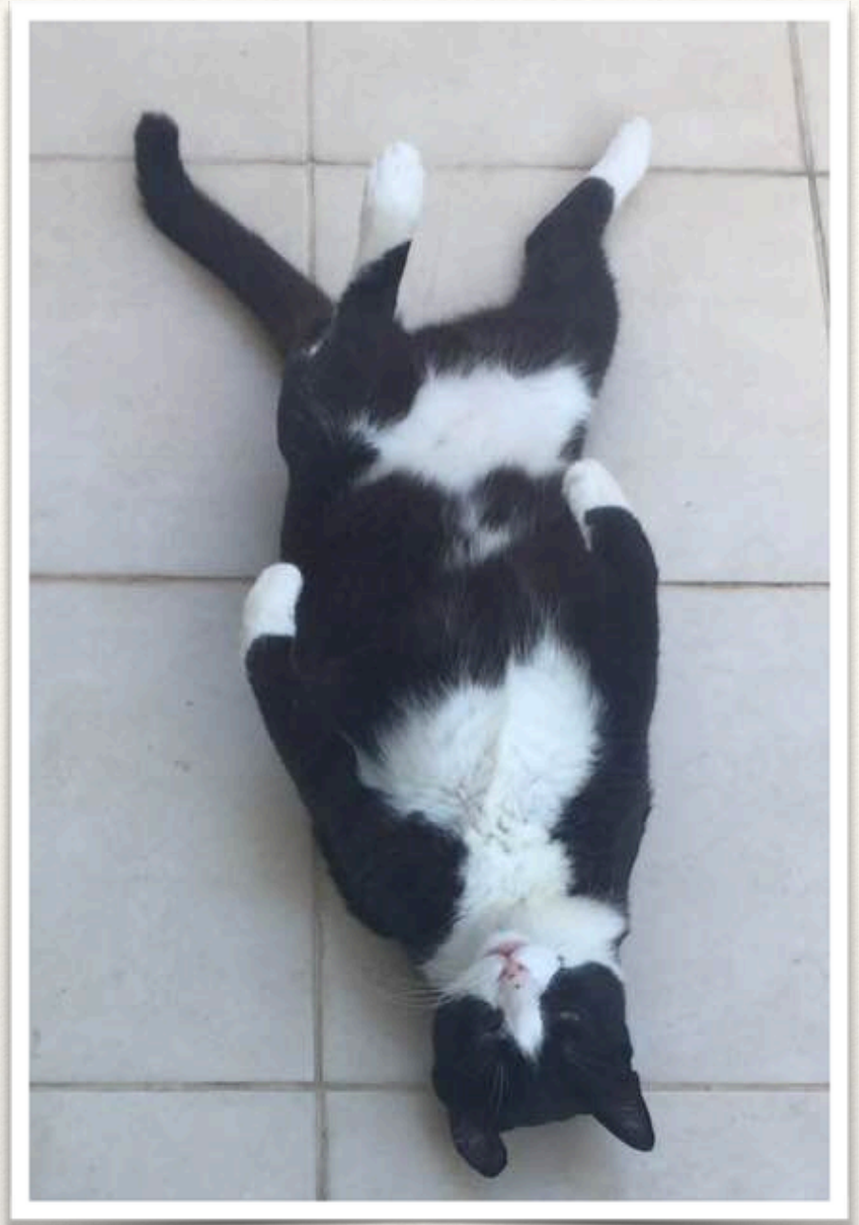
From Spectra ~~to Time Series~~

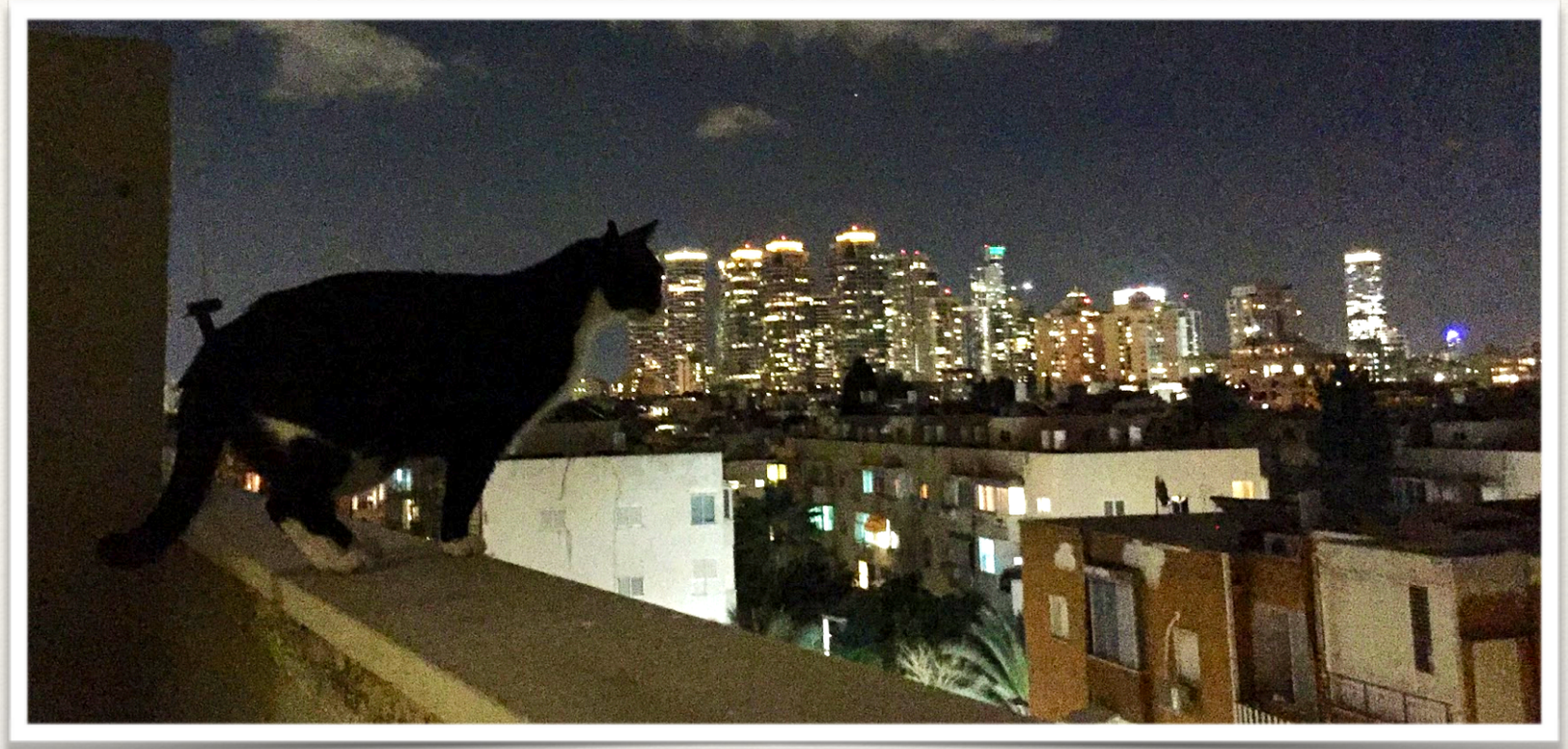
Dovi Poznanski

Tel Aviv University

+ Dalya Baron

+ Itamar Reis





Weirdos in Astronomy

- Quasars were weird stars, pulsars weird radio sources.
- Supernovae, clusters, nebulae.
- Within a class, weirdos stretch the limits of our understanding.
- Also - rare could be common but short lived.

How to find them?

- Eyeball everything.
- Fit model, look at:
  - Extremes in parameter space.
  - Bad fit.
- Learn empirical similarity metric from the data.

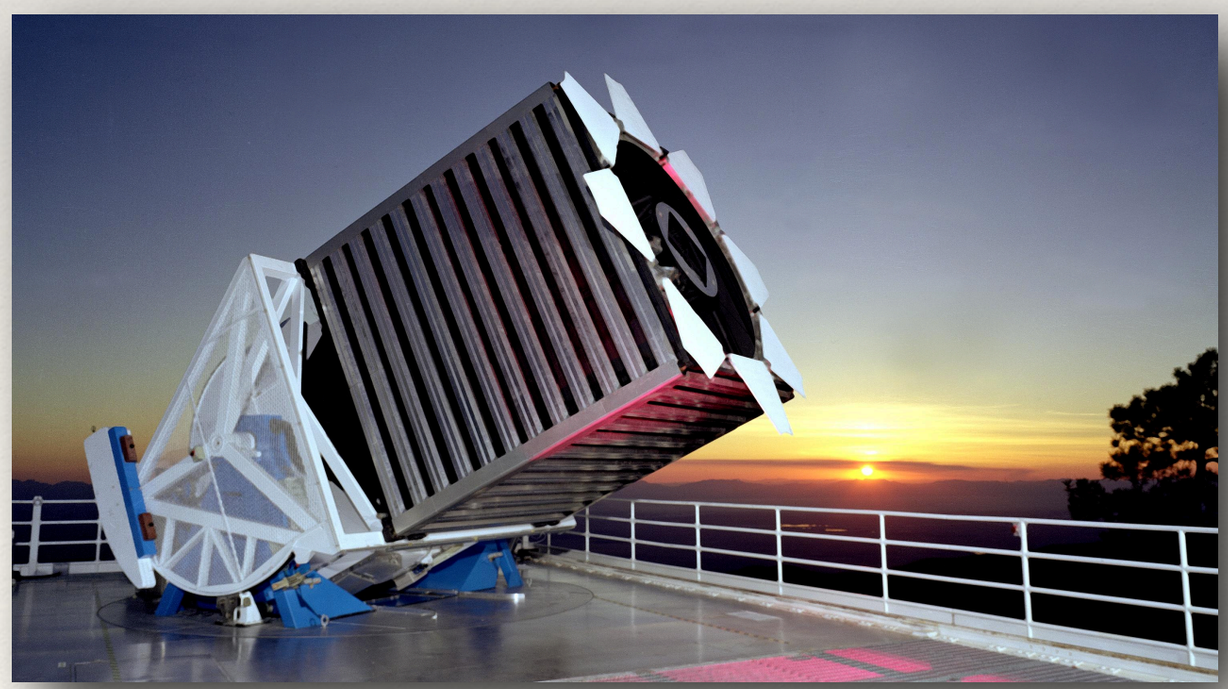
Example: RF-based method

- Using Random-Forest.
- We learn what constitutes a ‘normal’ object.
- By comparing the real sample to a synthetic sample.
- derive from that a similarity / distance metric between the objects.
- Outliers = farthest from most.

Baron & Poznanski 2017

Galaxy Spectra

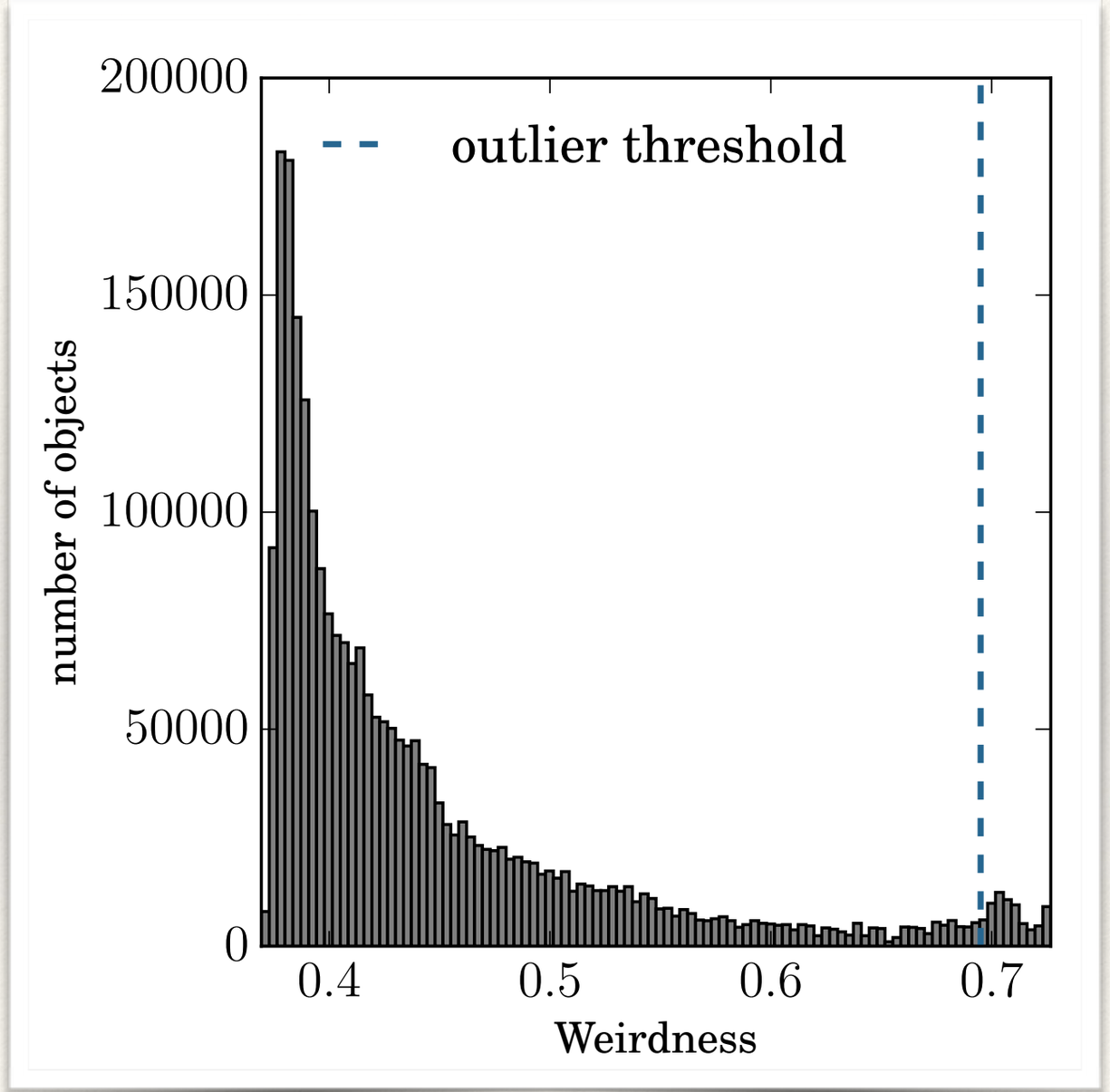
- ~2.5M galaxy spectra from SDSS.
- Most were published and studied years ago.
- Great benchmark.
- Flux(λ) are the features (so ~ 15,000 features).



Baron & Poznanski 2017

Galaxies Results

- Manually examined 400 farthest from others.



So…

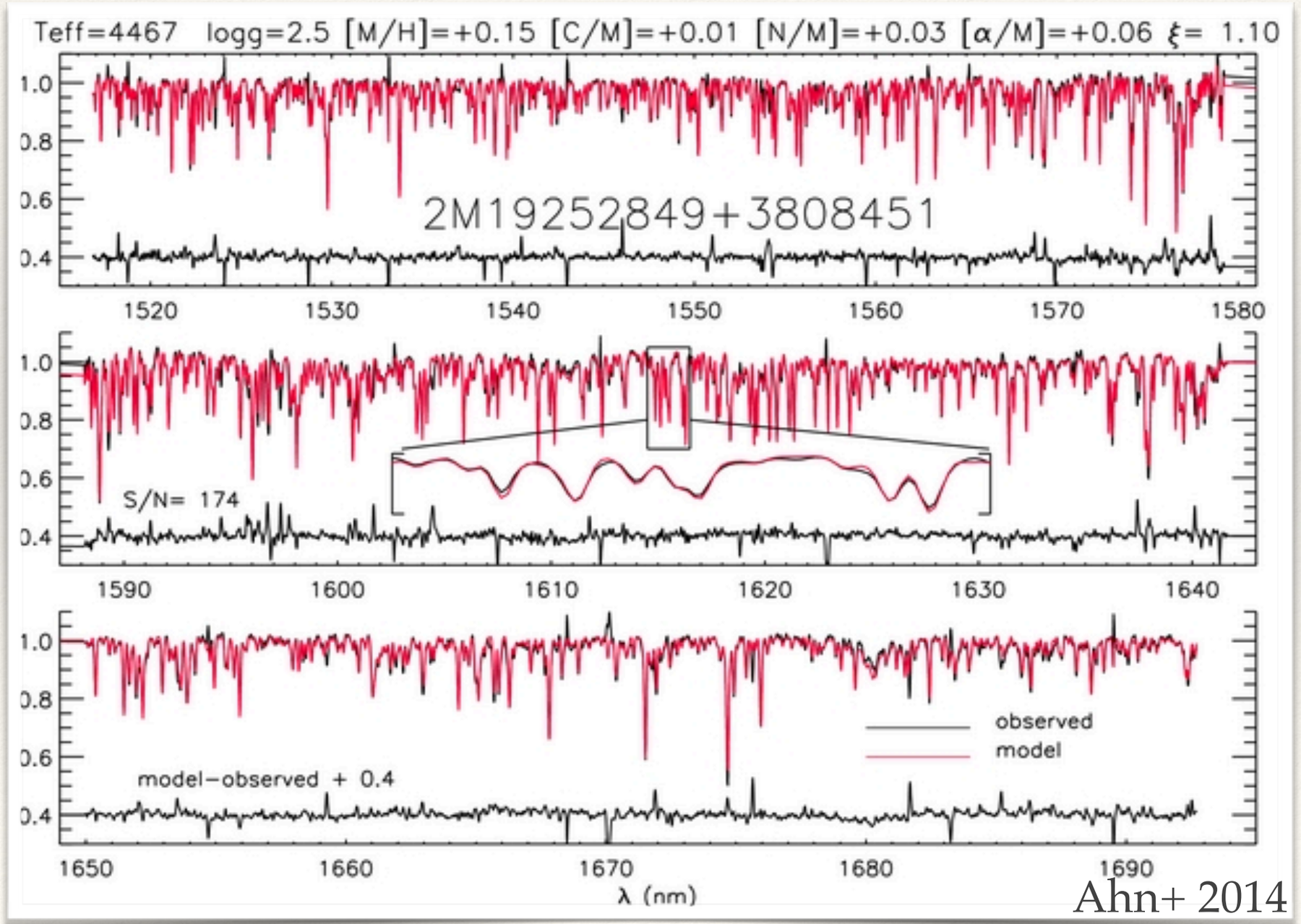
- **All top 400 were indeed outliers.**
- Outliers of many different kinds (emission, absorption, continuum, line ratios).
- Many previously reported (benchmark - ✔), most new (new science - ✔).
- Found multiple new things.
- Can do more than just find weirdos.

Given a Metric

- With 2.5M objects, we (in principle) have a $2.5M^2$ distance matrix.
- It is embedded in too many (~15,000) dimensions…
- Let's reduce them to 2D.

APOGEE + t-SNE

- ~200,000 near-IR stellar spectra from APOGEE.

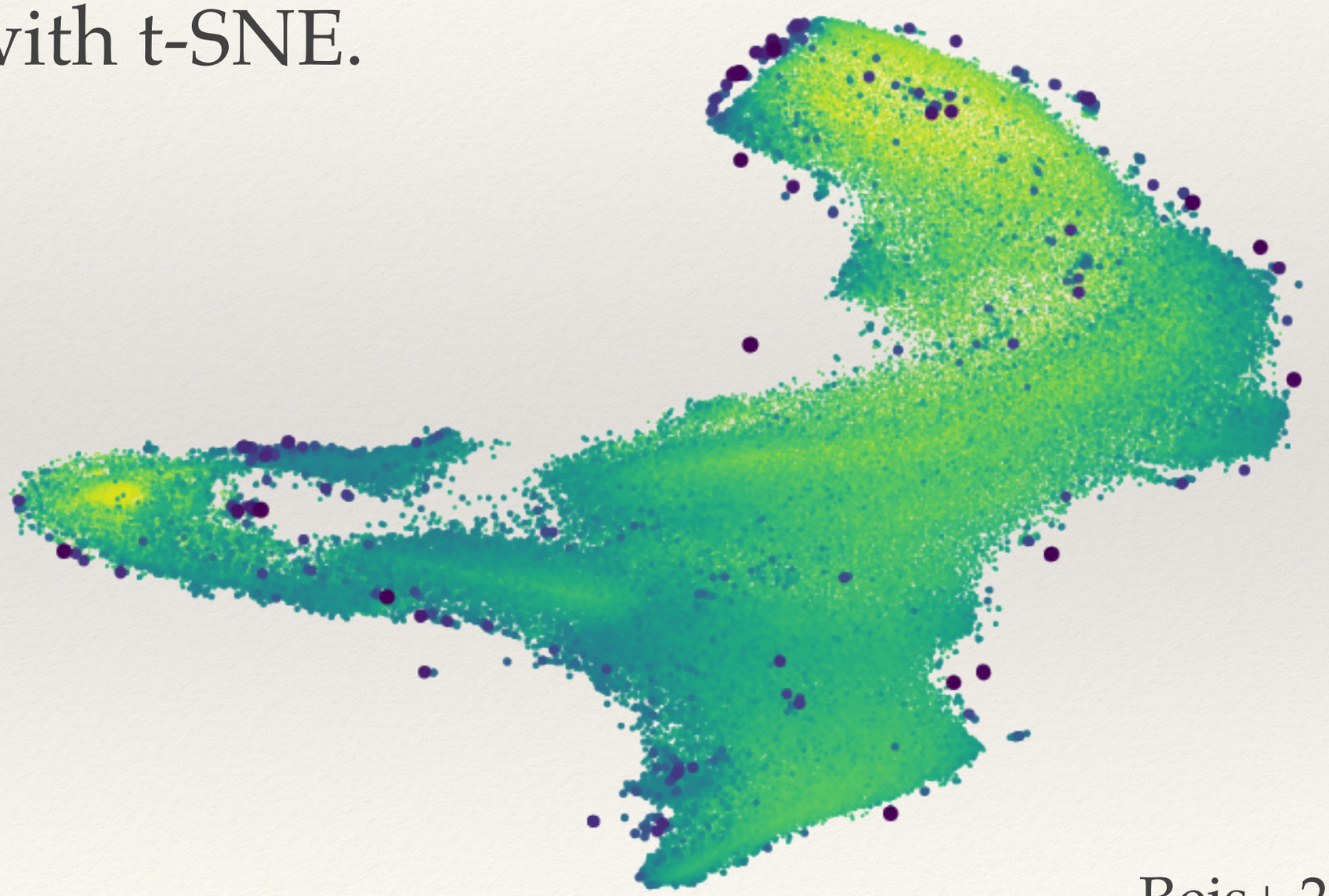


Ahn+ 2014

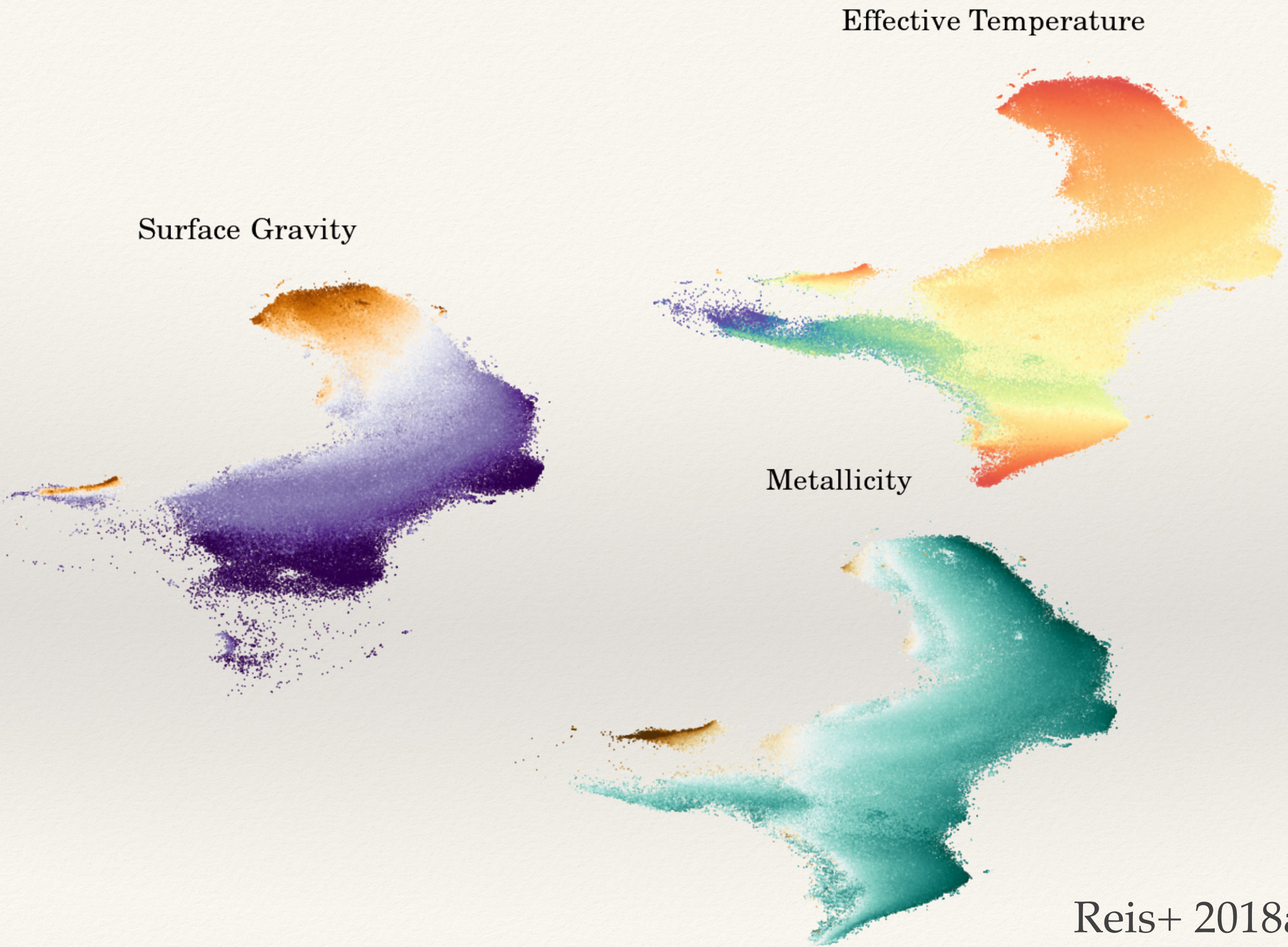
Reis+ 2018a

APOGEE + t-SNE

- RF-based distances.
- Reduced with t-SNE.



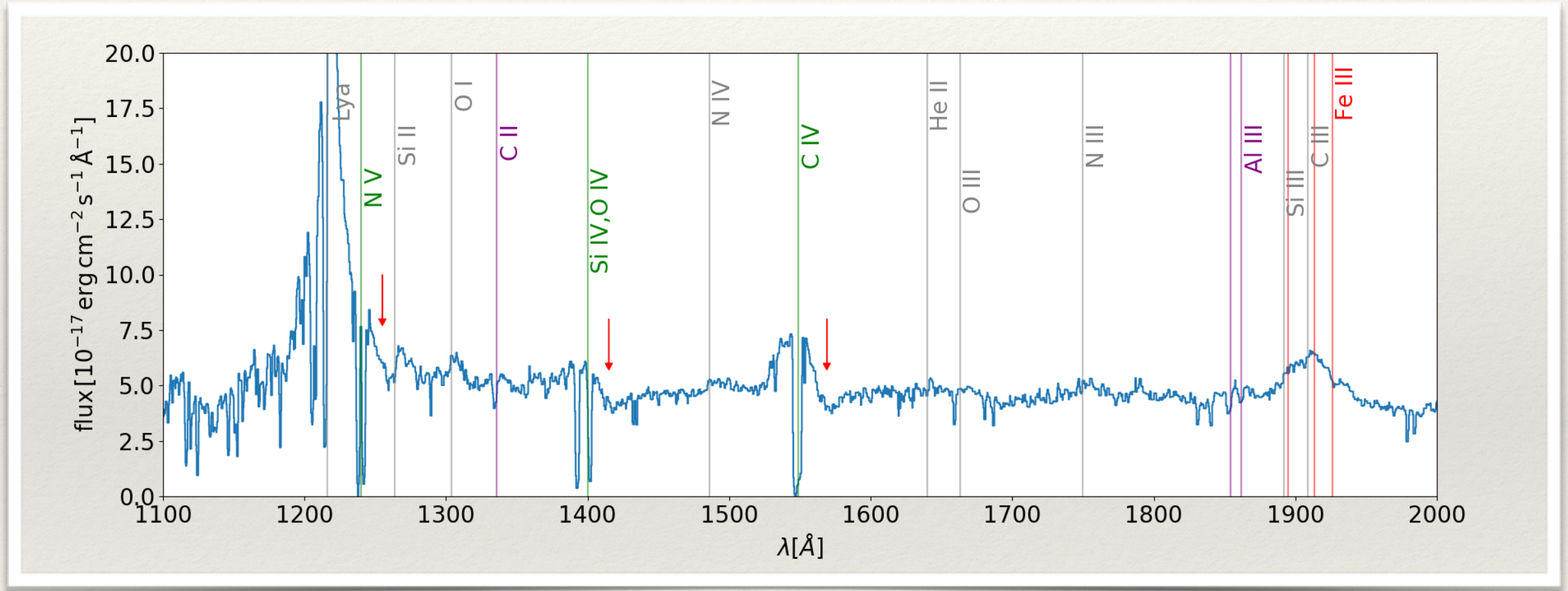
Reis+ 2018a



Reis+ 2018a

Object retrieval

- Find more objects like “this one”:

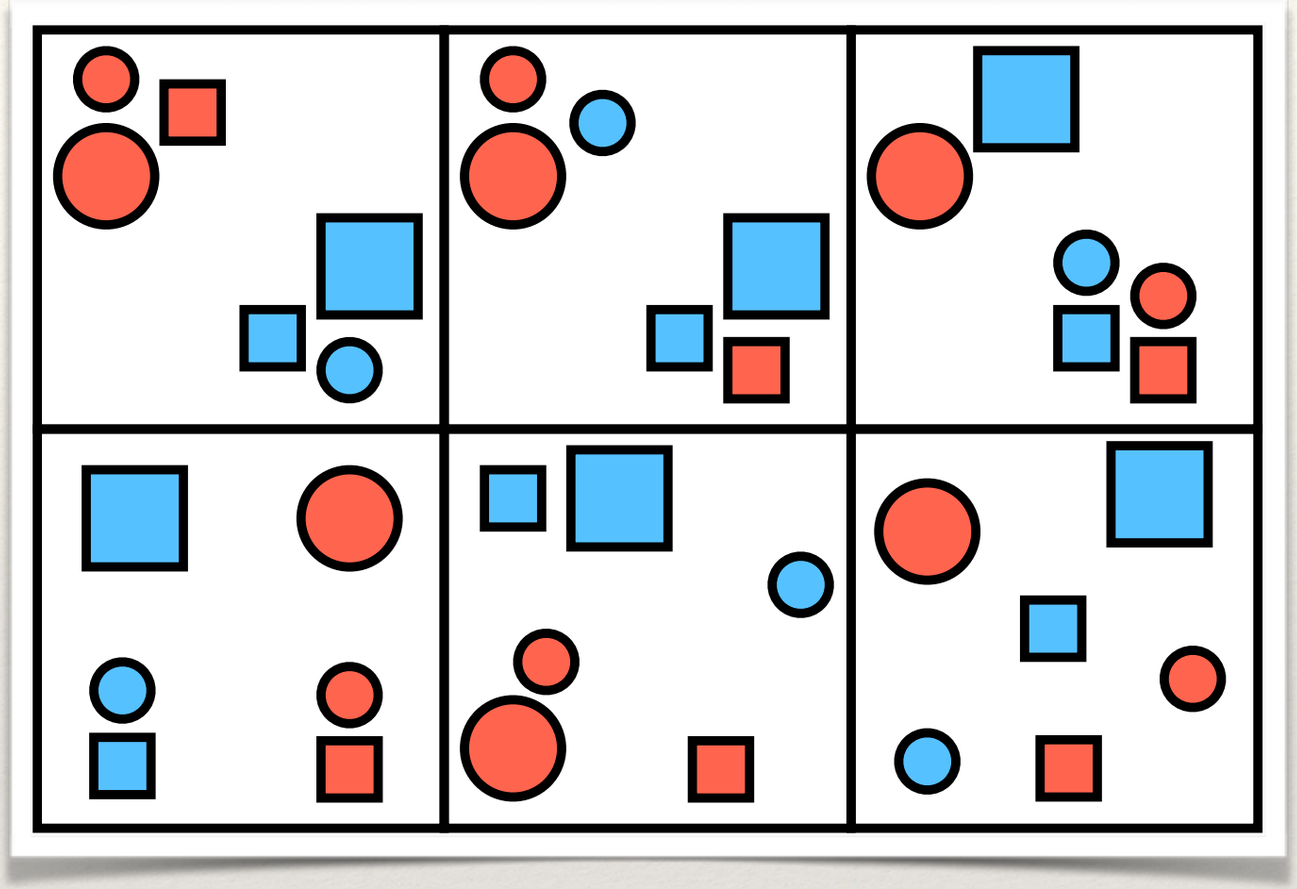


- Tripled known sample of redshifted BAL QSOs.

Hall+ 2013, Reis+ 2018b

Caveat

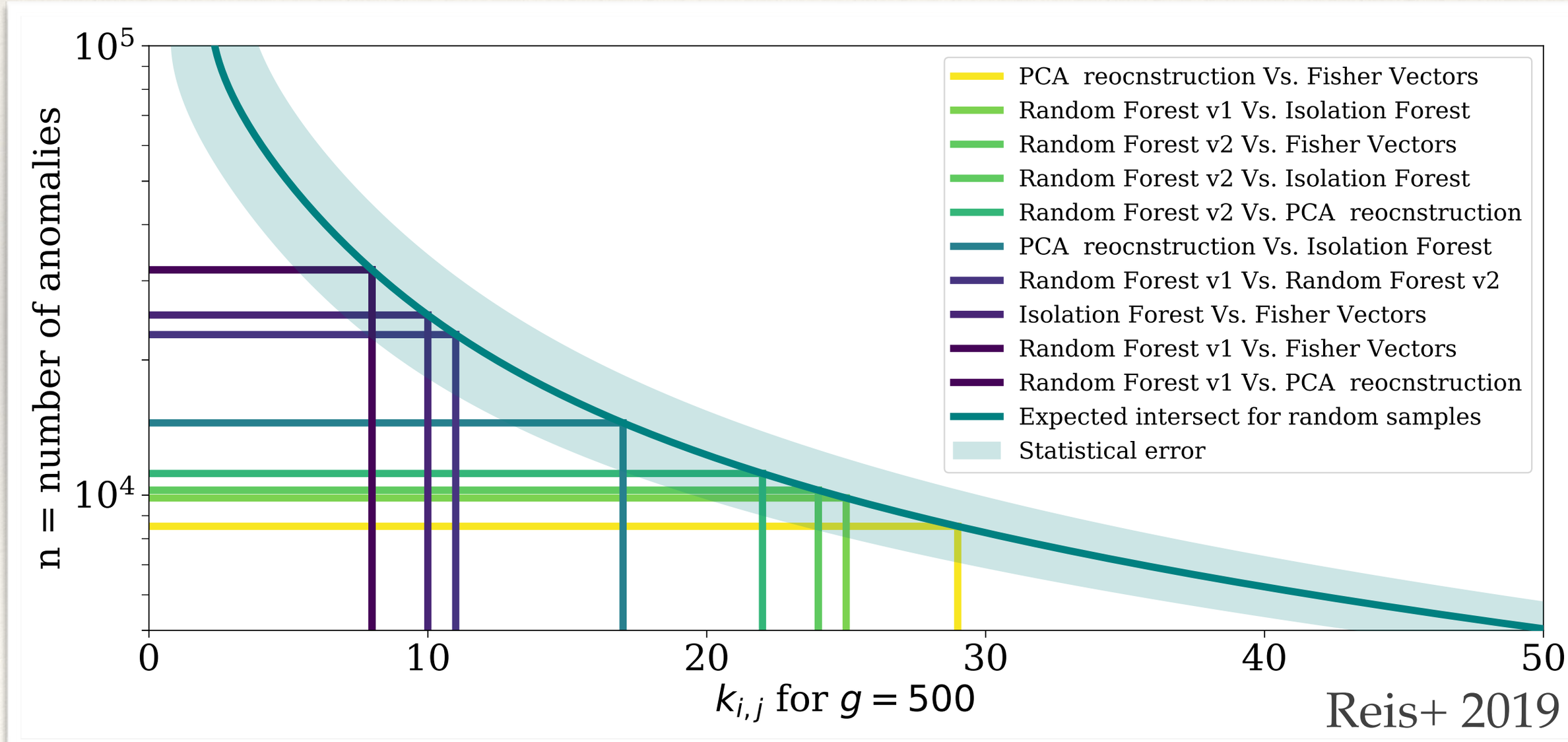
- There are many good algorithms.
- They can each find something.
- Change their hyper-parameters and they will find something else.
- Unsupervised tasks are ambiguous by construction.
- Notion of distance is ambiguous.



Reis+ 2019

Caveat

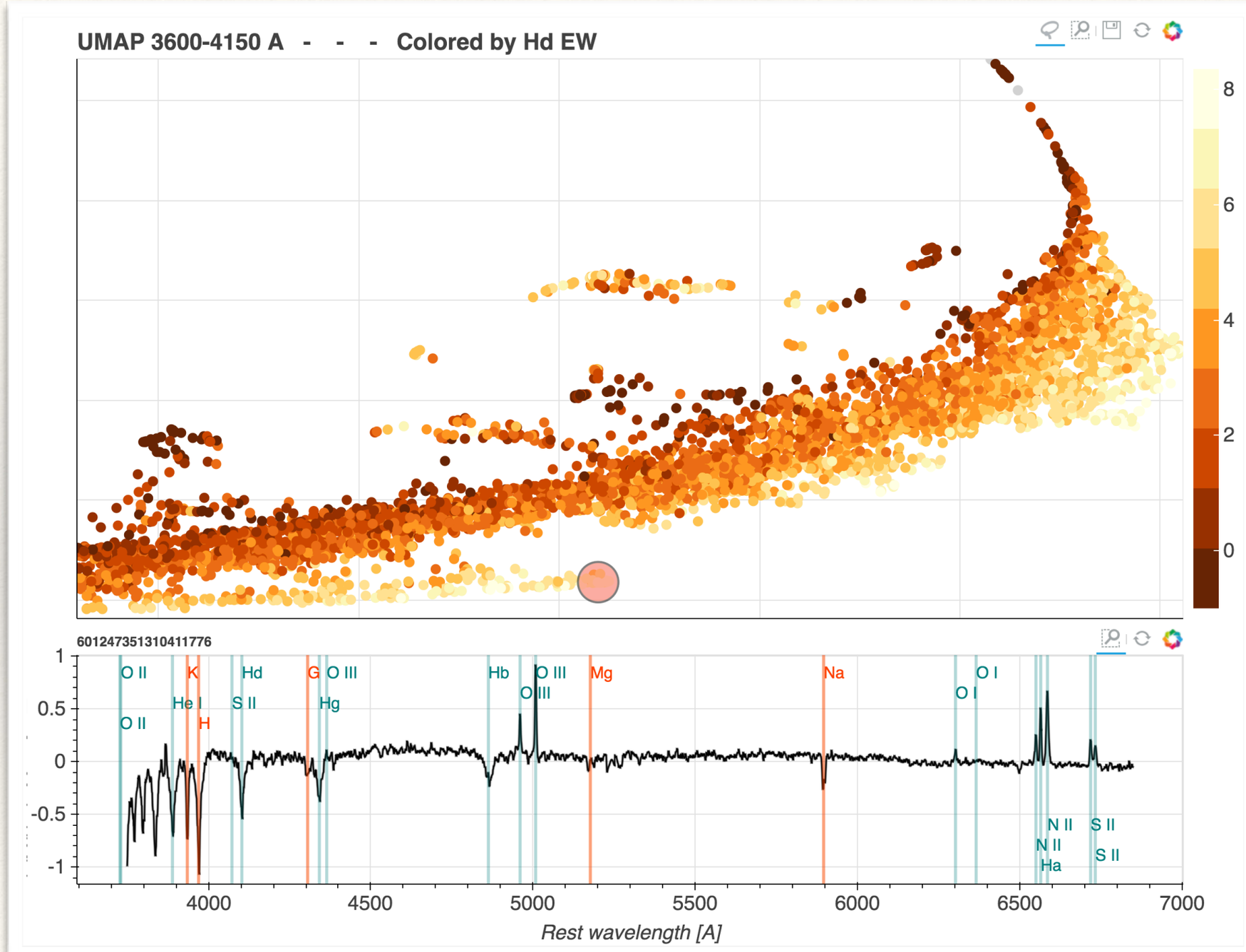
- In fact, overlap between methods is < 10%



Reis+ 2019

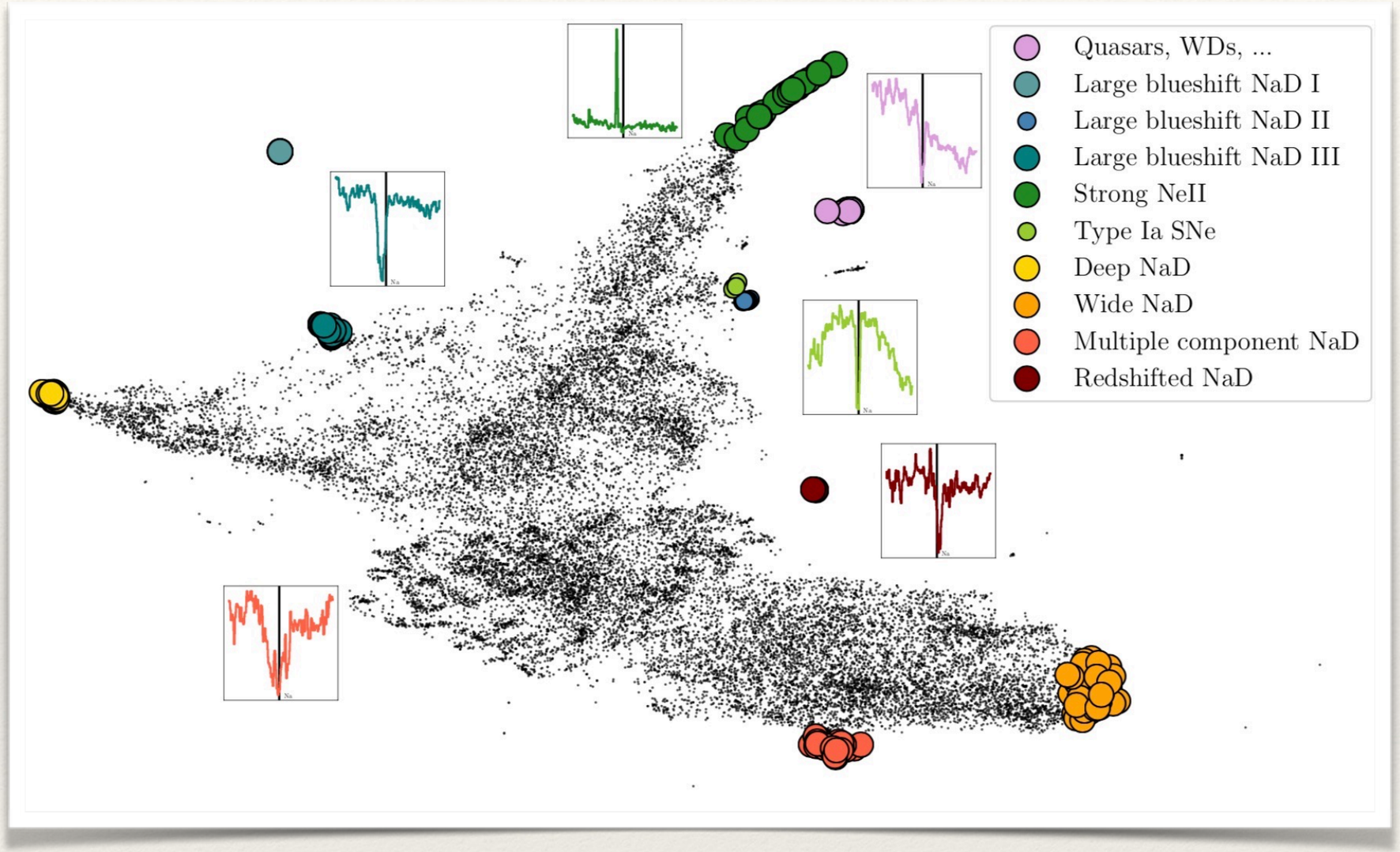
Solution

- Building an interactive tool to scan multiple "views".
- How to interact with thousands of points?
- Can select, stack, color, etc.



Reis+ 2019

Solution



- Quickly vet outliers.
- Try it: https://galaxyportal.space

Reis+ 2019

Machine Enabled Discovery

- We are developing an unsupervised ML toolset that works well for spectra.
- We are evolving to work also with time series.
- There are endless applications with either, and many interesting problems worth solving.

But wait:

But wait:

AAS Job Register
Find and post astronomy related jobs!

Up to Ten Postdoctoral Positions at Tel-Aviv University: GW Followup, High-Energy Transients, SMBH Accretion, Relativistic Jets, Compact Binaries, Chemical Evolution (Observation & Theory)

Below: Tel Aviv on a typical winter day.

Thanks! *dovi@tau.ac.il*