名古屋大学
NAGOYA UNIVERSITY

Understanding Galaxy Evolution through Machine Learning

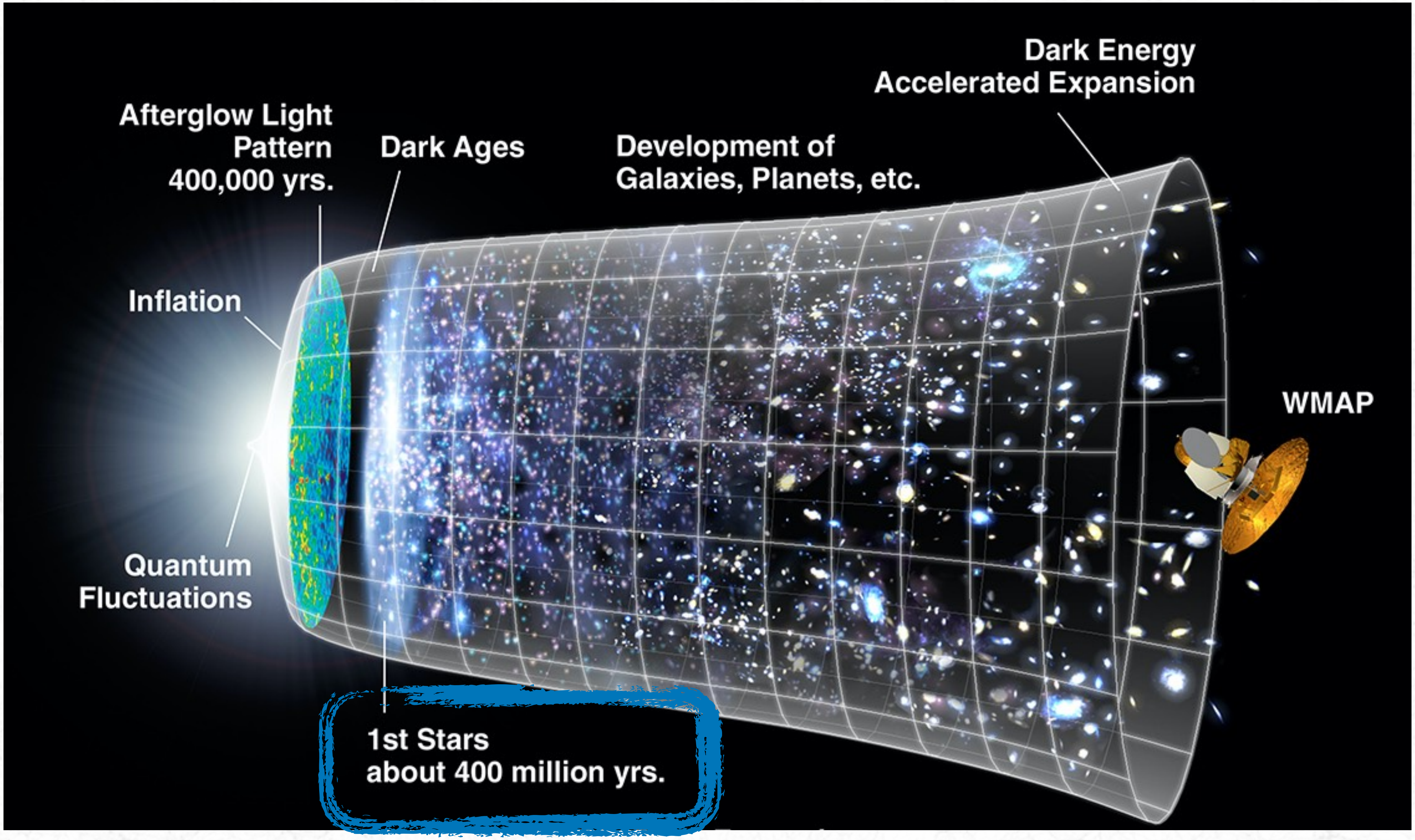
First Stars and First Galaxies Workshop 2019 @ Nagoya U

Suchetha Cooray

Collaborators:

Tsutomu T. Takeuchi (Nagoya U.), Shuntaro Yoshida (Nagoya U.), Kai Kono (Nagoya U.)

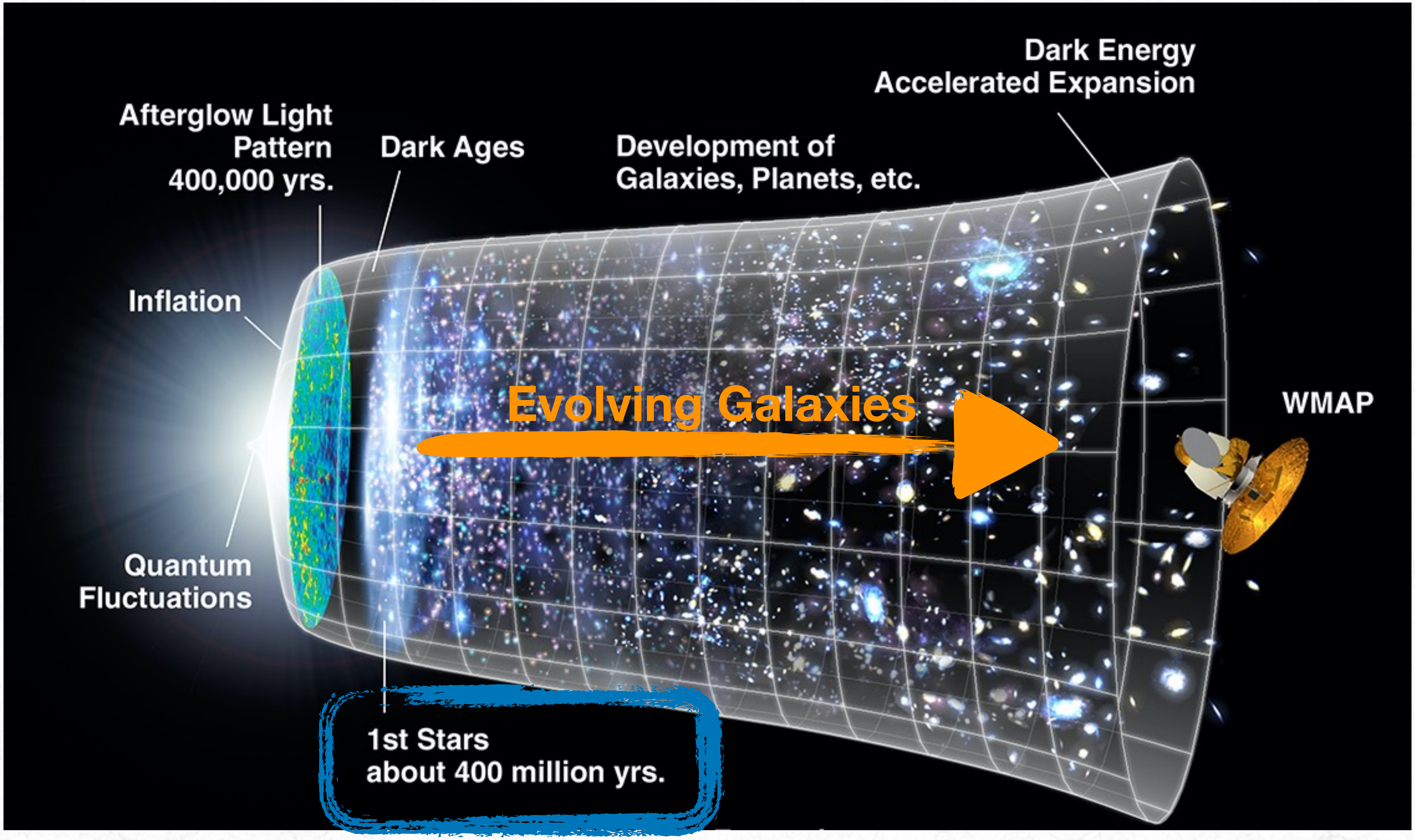
Formation and Evolution of Galaxies



Important!

http://map.gsfc.nasa.gov/m_mm.html

Formation and Evolution of Galaxies



Important!

http://map.gsfc.nasa.gov/m_mm.html

Final Aim of Galaxy Evolution Studies

The theoretical model to explain galaxy evolution would look like this

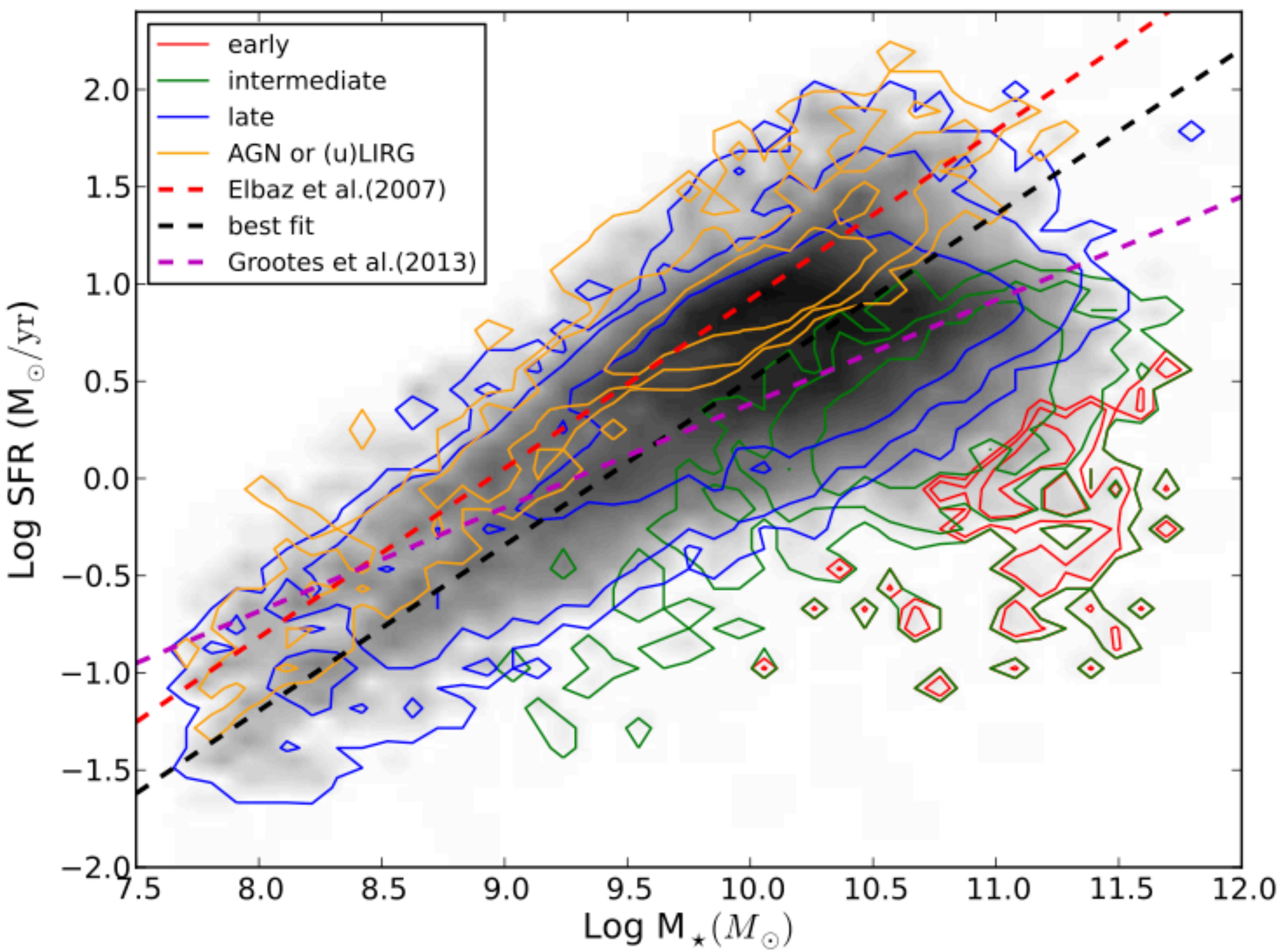
$$
\begin{aligned}
\mathrm{SFR}(t) &= f_1(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_*(t) &= f_2(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{mol}}(t) &= f_3(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{HI}}(t) &= f_4(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{dust}}(t) &= f_5(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{halo}}(t) &= f_6(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
\delta_{\mathrm{gal}}(t) &= f_7(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
&\vdots
\end{aligned}
$$

Clearly, it is a very complicated problem!

It is high time to define the evolution of galaxies with a more objective point of view

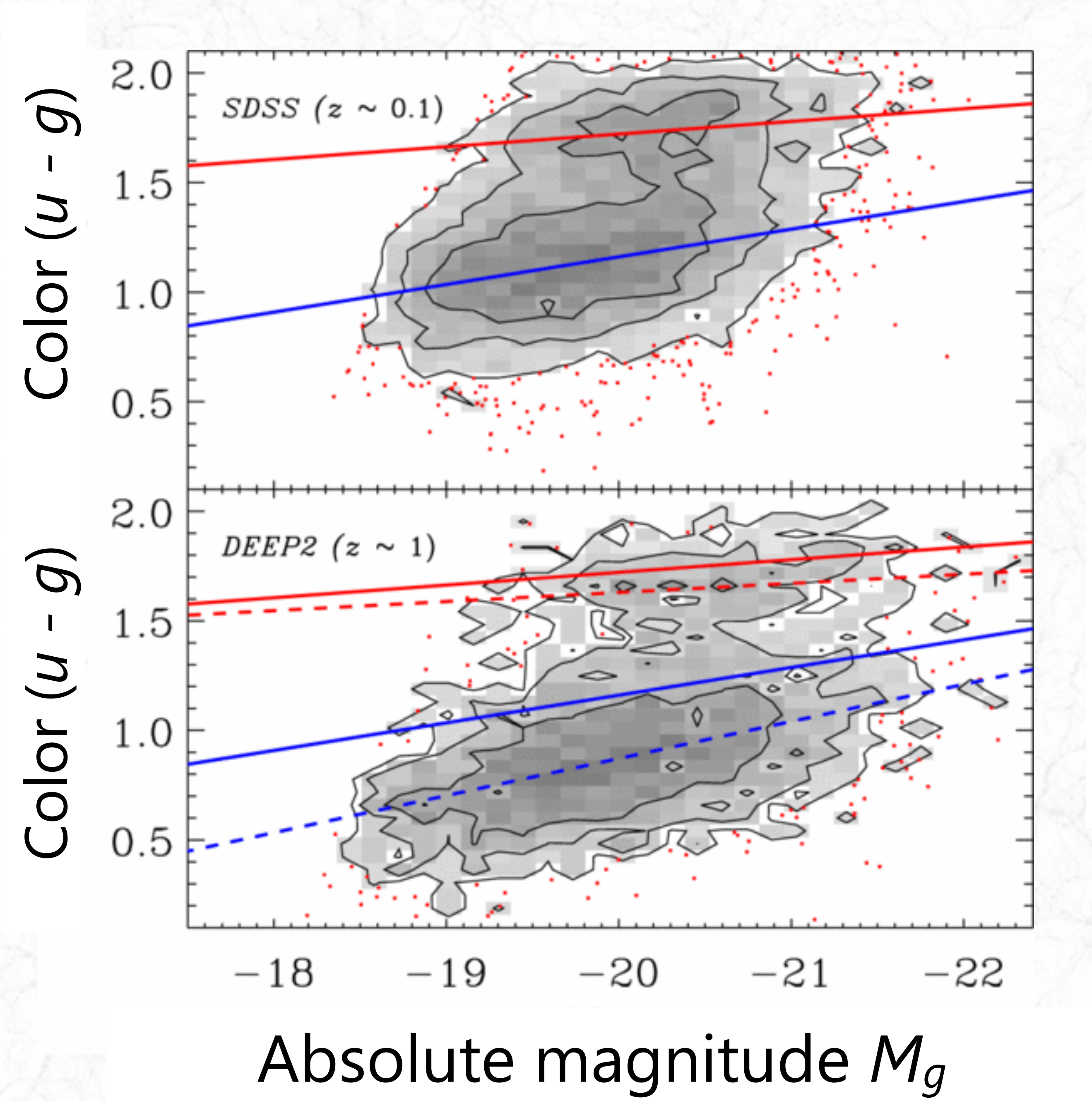
Credit: T. Takeuchi

Conventional Method (Derived Physical Quantities)

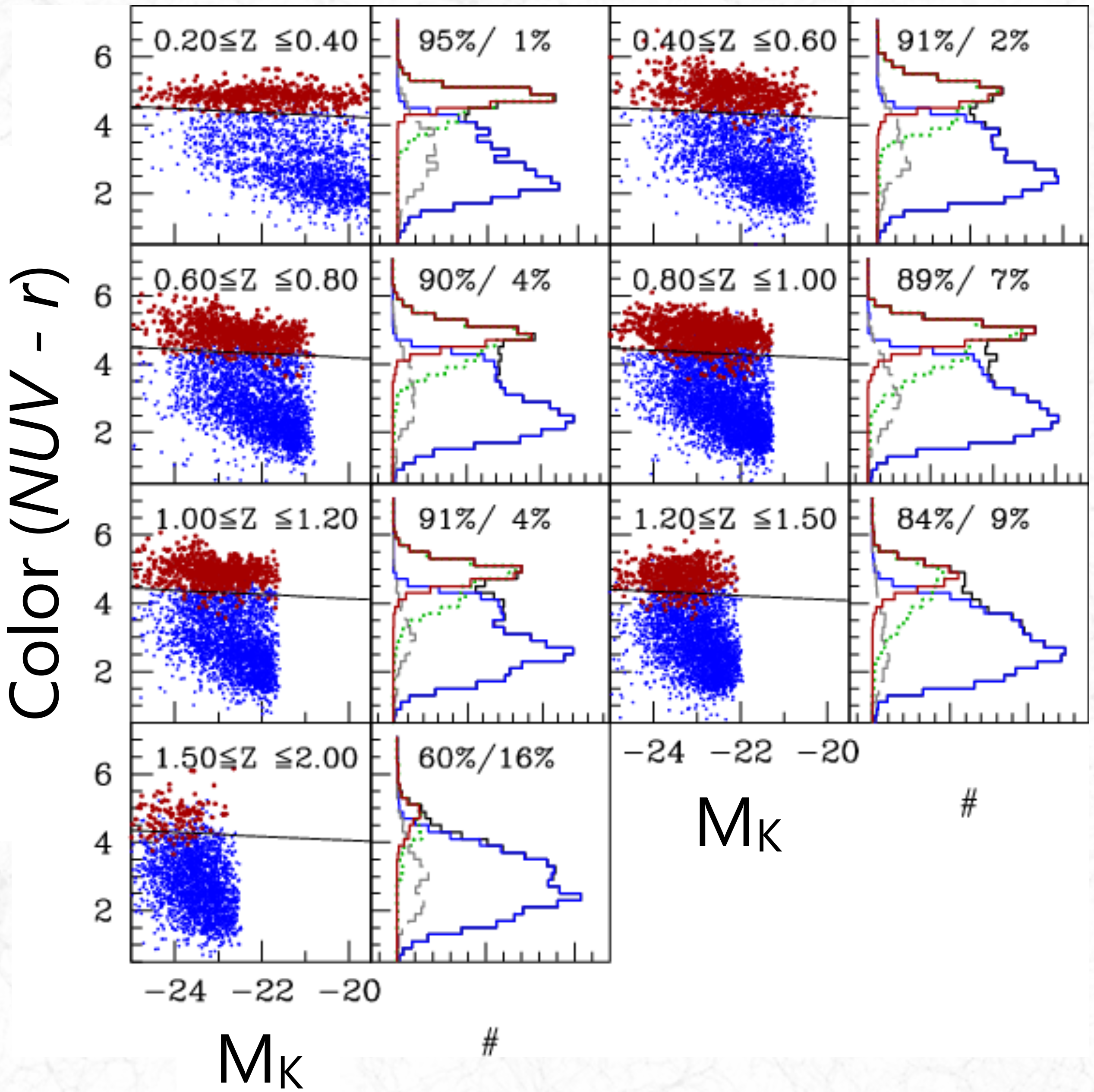


Grootes et al. 2013

Traditional Methods (Color-Magnitude Diagrams)



Blanton (2006)

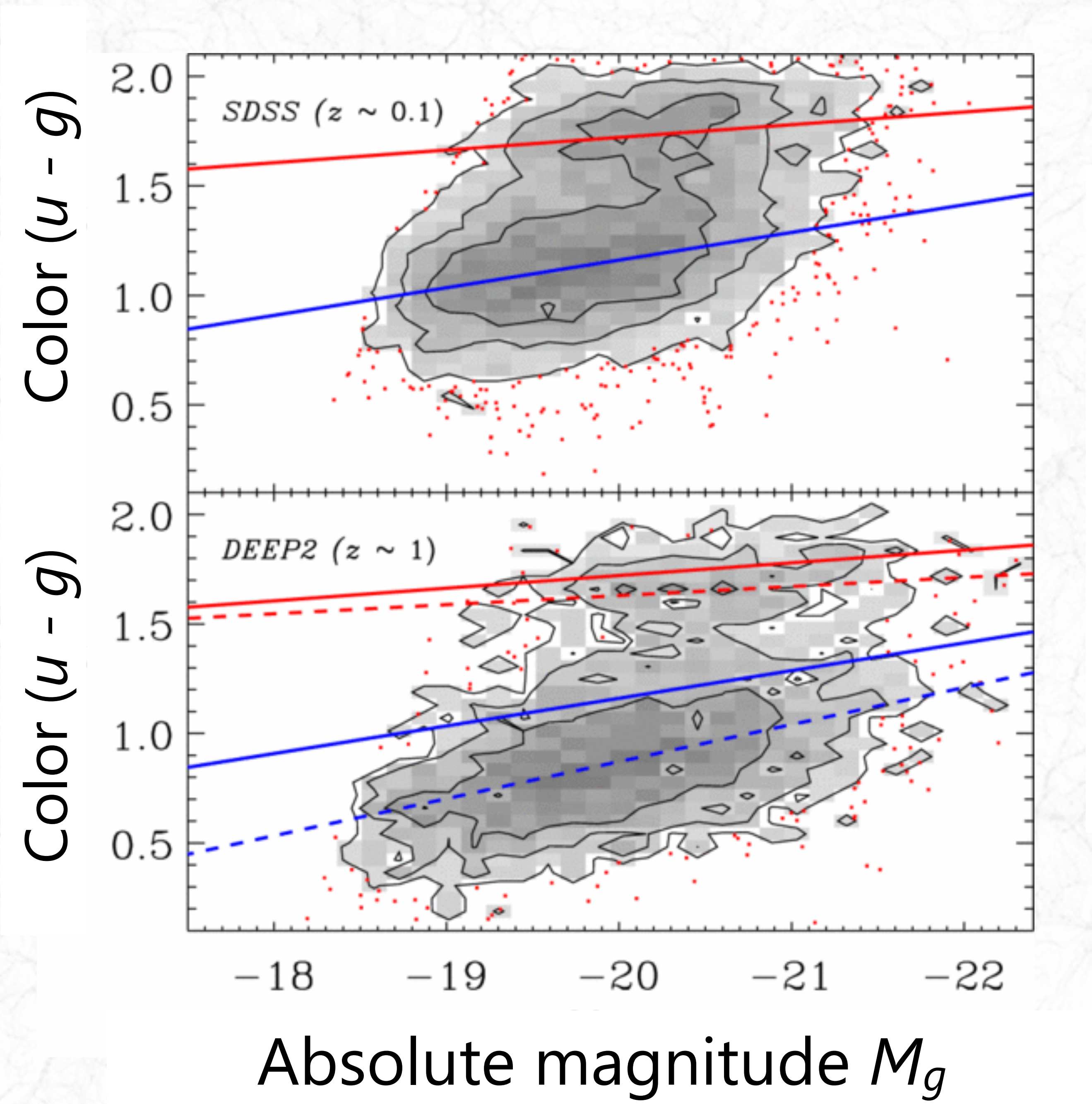


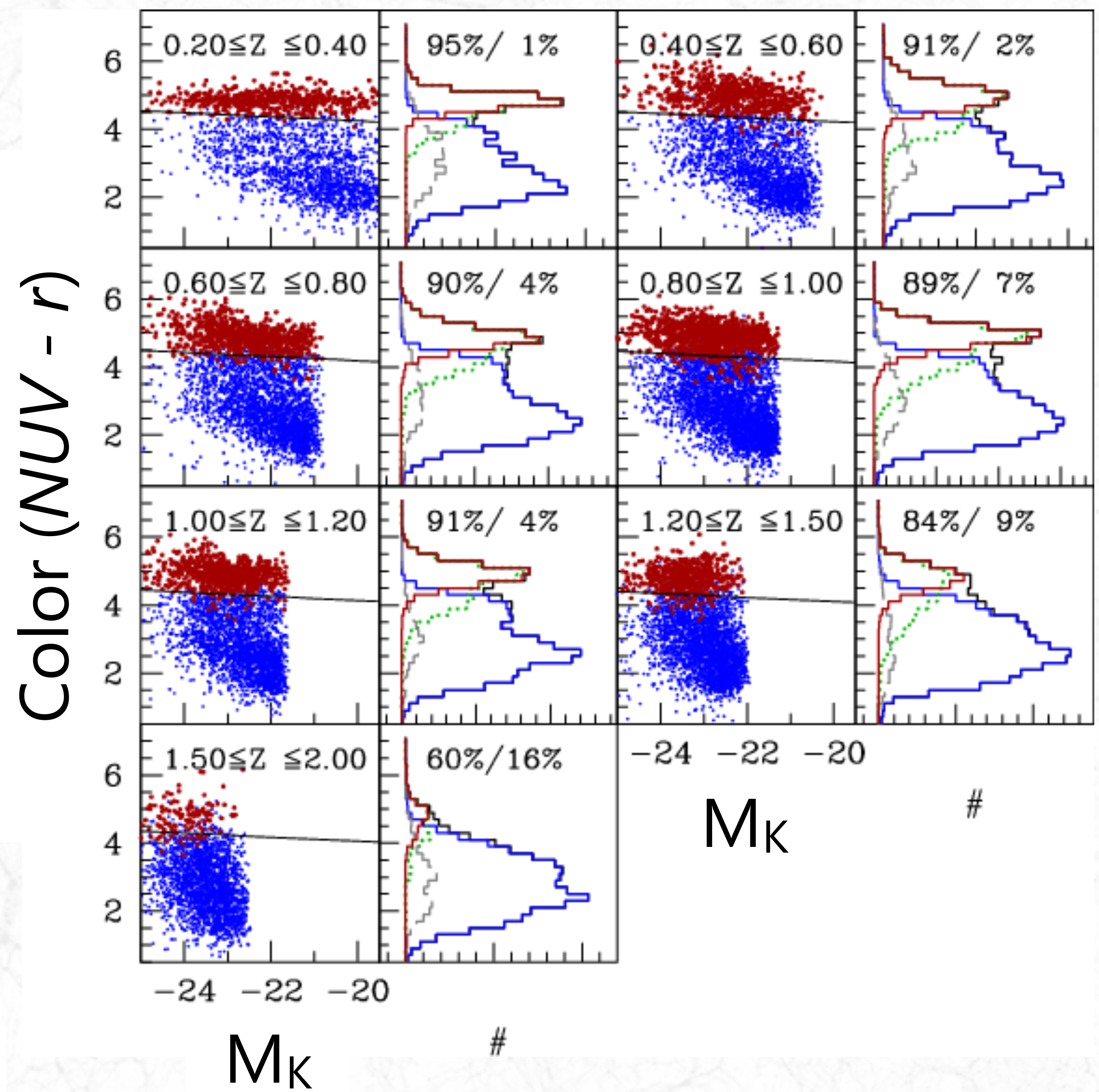
Arnouts et al. (2007)

These plots show clear bimodality / multimodality

Do they have a physical meaning?

Traditional Methods (Color-Magnitude Diagrams)





Redder Galaxies: Red Sequence, Bluer galaxies: Blue Cloud, Boundary: Green Valley

Potential Problems with Color Based Methods

- Colors are basically ratios of two luminosities

  ⇒ Selection effect is always too entangled and messy

  ⇒ Completeness test is almost impossible in a simple way

↓

Suggestion: forget about colors!

Credit: T. Takeuchi

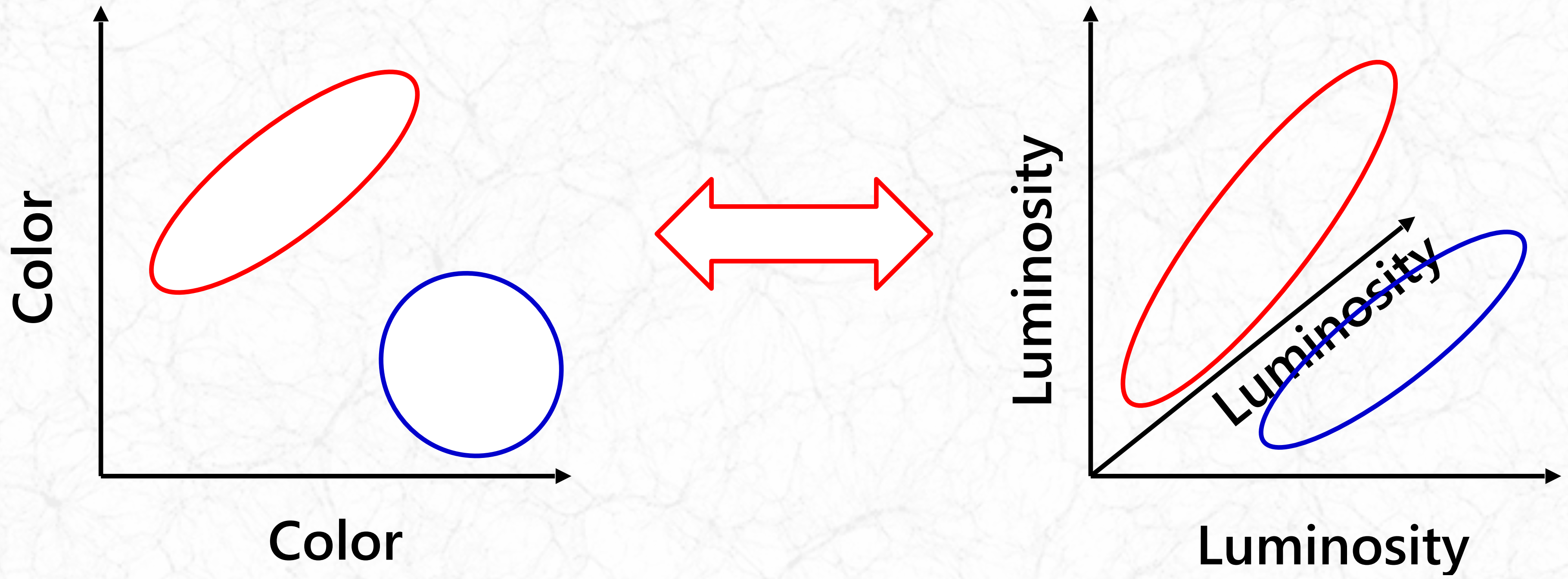
Question 1

Can we find intrinsic groups of galaxies without selection effects?

Multidimensional Luminosity Space

Instead of colors, can we just use the distribution of galaxies in a **multidimensional luminosity (absolute magnitude) space?**

If we have a bimodality in color-color space,
we must have equivalent peaks in
the multidimensional luminosity space



Credit: T. Takeuchi

Multidimensional Luminosity Space

Instead of colors, can we just use the distribution of galaxies in a **multidimensional luminosity (absolute magnitude) space?**

If we have a bimodality in color-color space,
we must have equivalent peaks in
the multidimensional luminosity space

The boundaries can be automatically defined by a **machine-learning type method**

Advantage of this idea is that we simply deal with the selection at each band, not as an entangled multiple selection.

Credit: T. Takeuchi

Data

RCSED

- **R**eference **C**atalog of galaxy **S**pectral **E**nergy **D**istributions (RCSED) (Chilingarian et al. 2016)
- Catalog of galaxies produced as join between GALEX, SDSS, and UKIDSS catalogs, and processed with state-of-the-art spectral analysis methods
- Covers approximately 25% of the sky and contains **k-corrected ultraviolet-to-near-infrared photometry** (11 bands of **FUV, NUV, u, g, r, i, z, Y, J, H, K**) of some **1 million galaxies**, as well as some of their physical properties

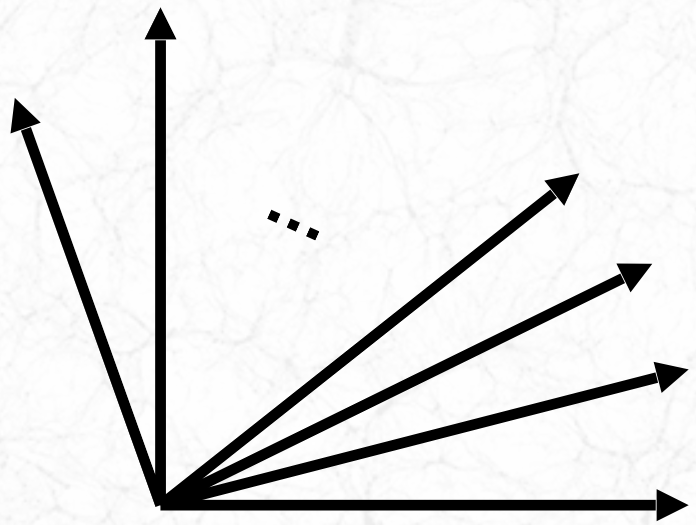
http://rcsed.sai.msu.ru

Classification in Luminosity Space

Generate a subsample with all 11 rest-frame magnitudes (**FUV, NUV, u, g, r, i, z, Y, J, H, K**) -> ~800,000 galaxies

=> Find a volume limited sample that is representative of the whole galaxy sample -> ~30,000 galaxies

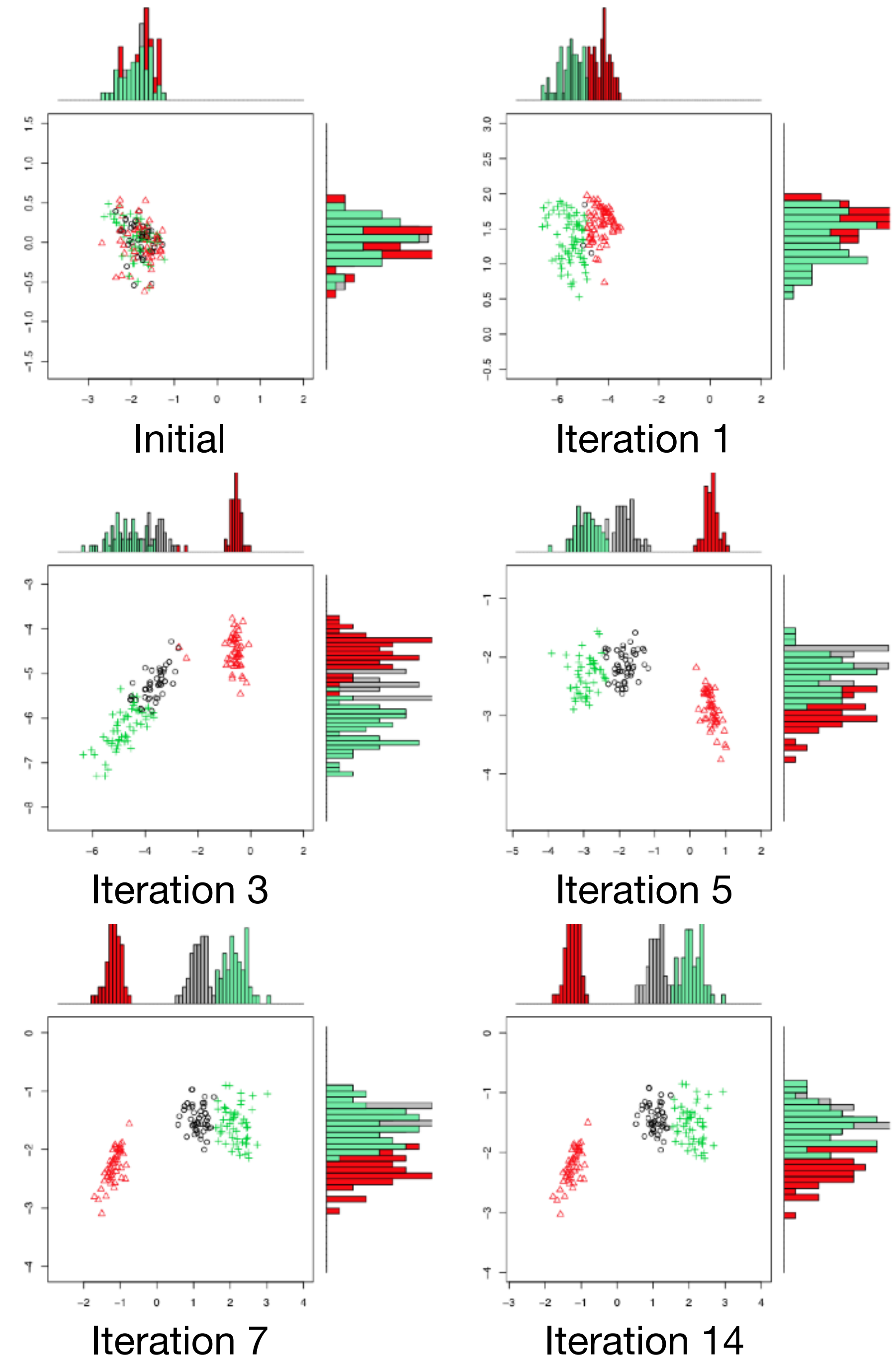
Can you then classify these galaxies by intuition?



Credit: T. Takeuchi

Unsupervised Machine Learning of Galaxy Groups

- **Fisher Expectation-Maximization (FisherEM) algorithm** (Bouveryron & Brunet 2012)
- Subspace clustering method for high-dimensional data.
- Based on the **Gaussian Mixture Model** (GMM) and estimates both the **discriminative subspace** and the **parameters of the mixture model**

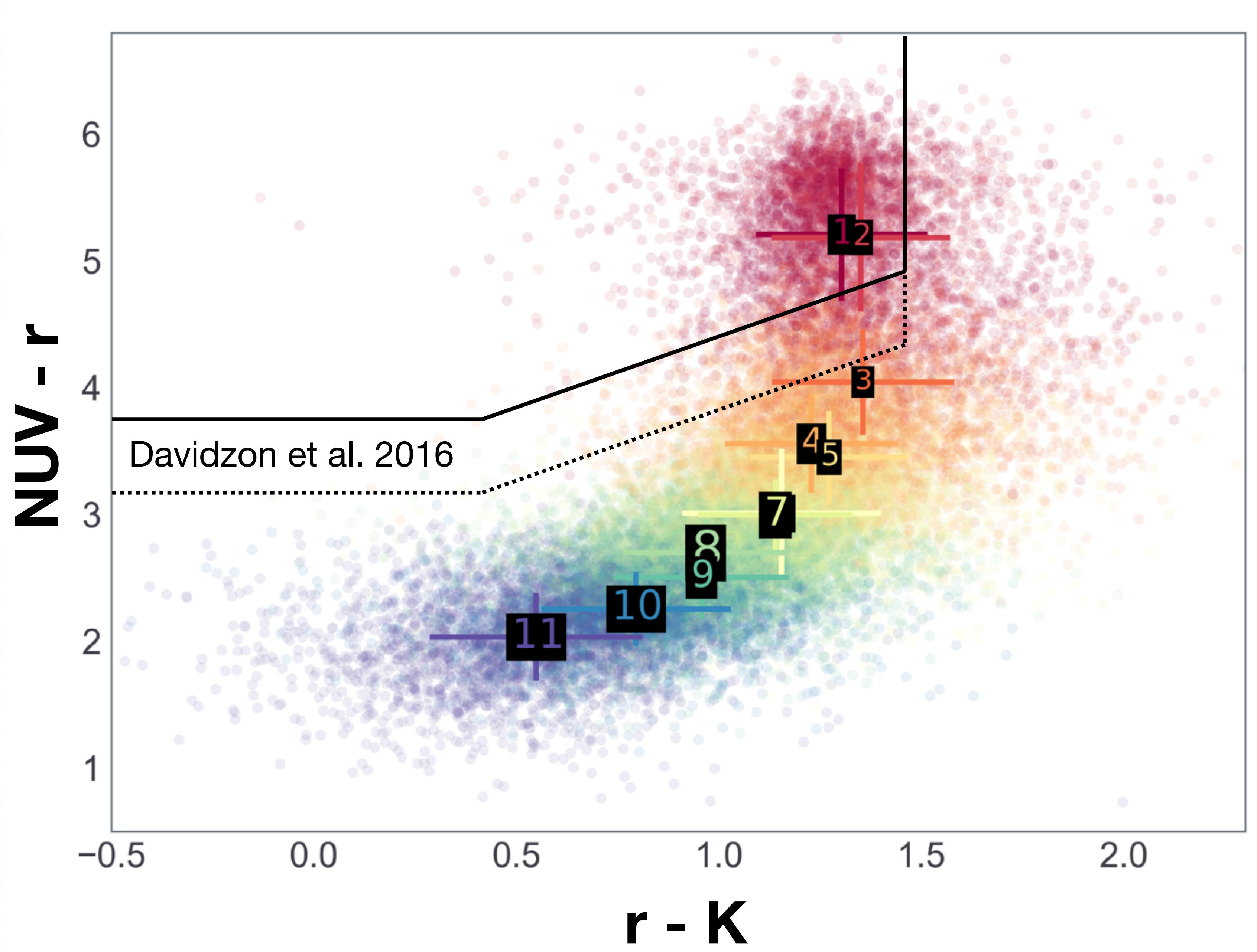


Projected clusters onto estimated latent discriminative sub-space at certain iterations

Machine Learnt Galaxy Classes

We classified the whole galaxy sample into 11 classes

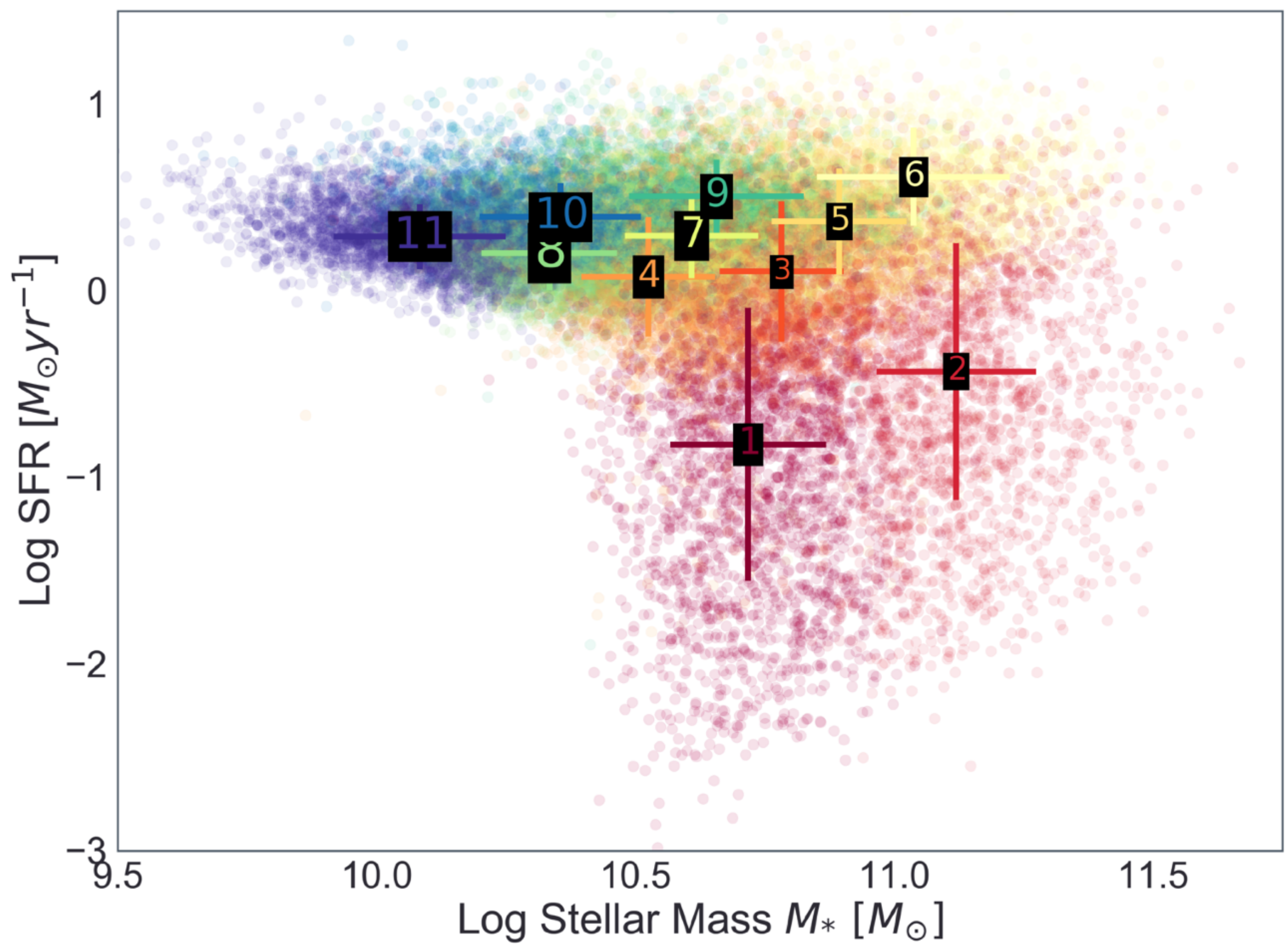
(based on Siudek et al. 2018)



NU*VrK* diagram (Arnouts et al. 2013)

Machine Learnt Galaxy Classes

We classified the whole galaxy sample into 11 classes
(based on Siudek et al. 2018)



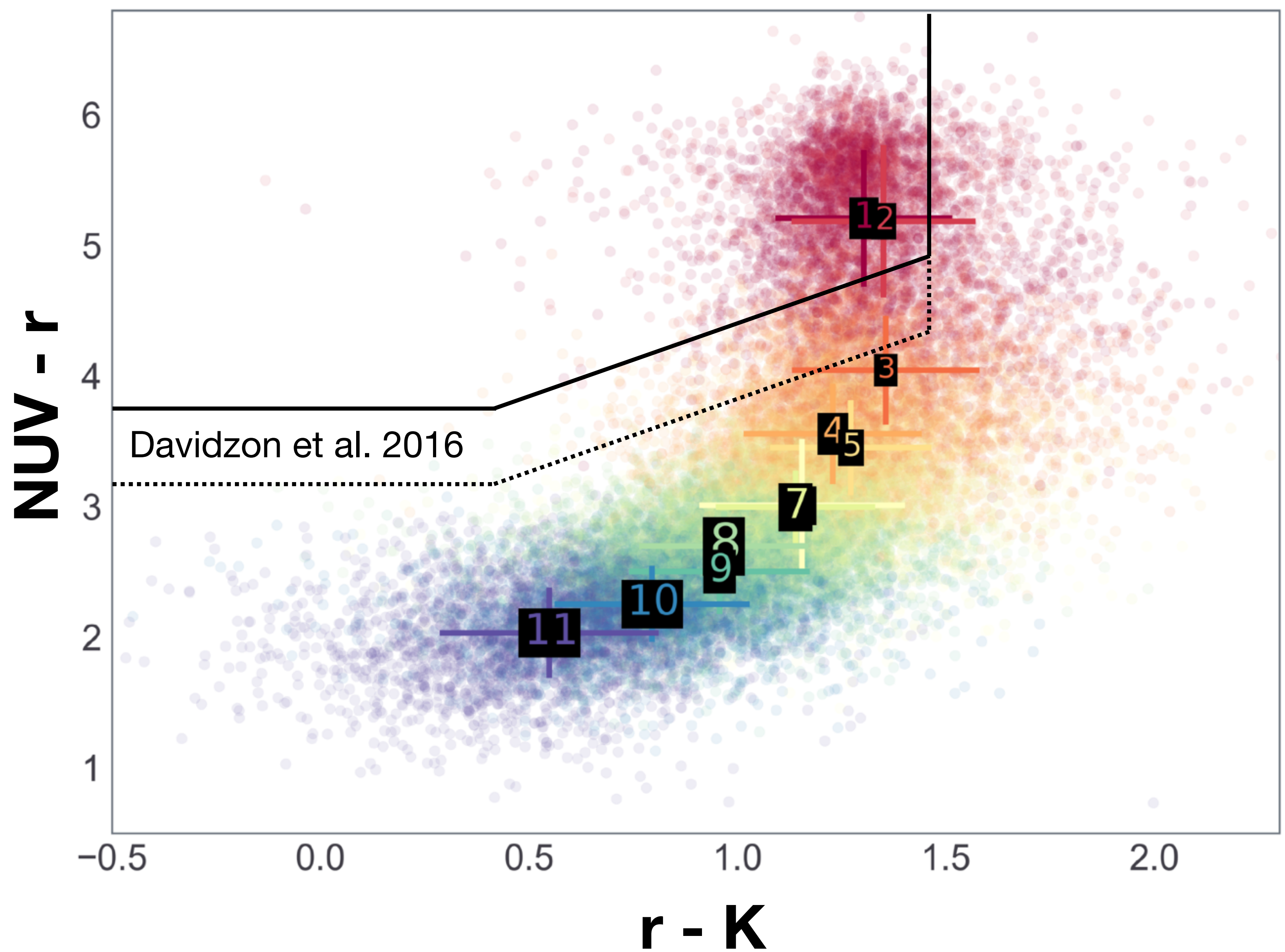
Cross-matched with GSWLC (Salim, Boquien & Lee 2018)

FisherEMくん is so smart that he
learnt about these groups all by himself!

The power of unsupervised machine learning algorithms!

Machine Learnt Galaxy Classes

We classified the whole galaxy sample into 11 classes
(based on Siudek et al. 2018)



NUV*VrK* diagram (Arnouts et al. 2013)

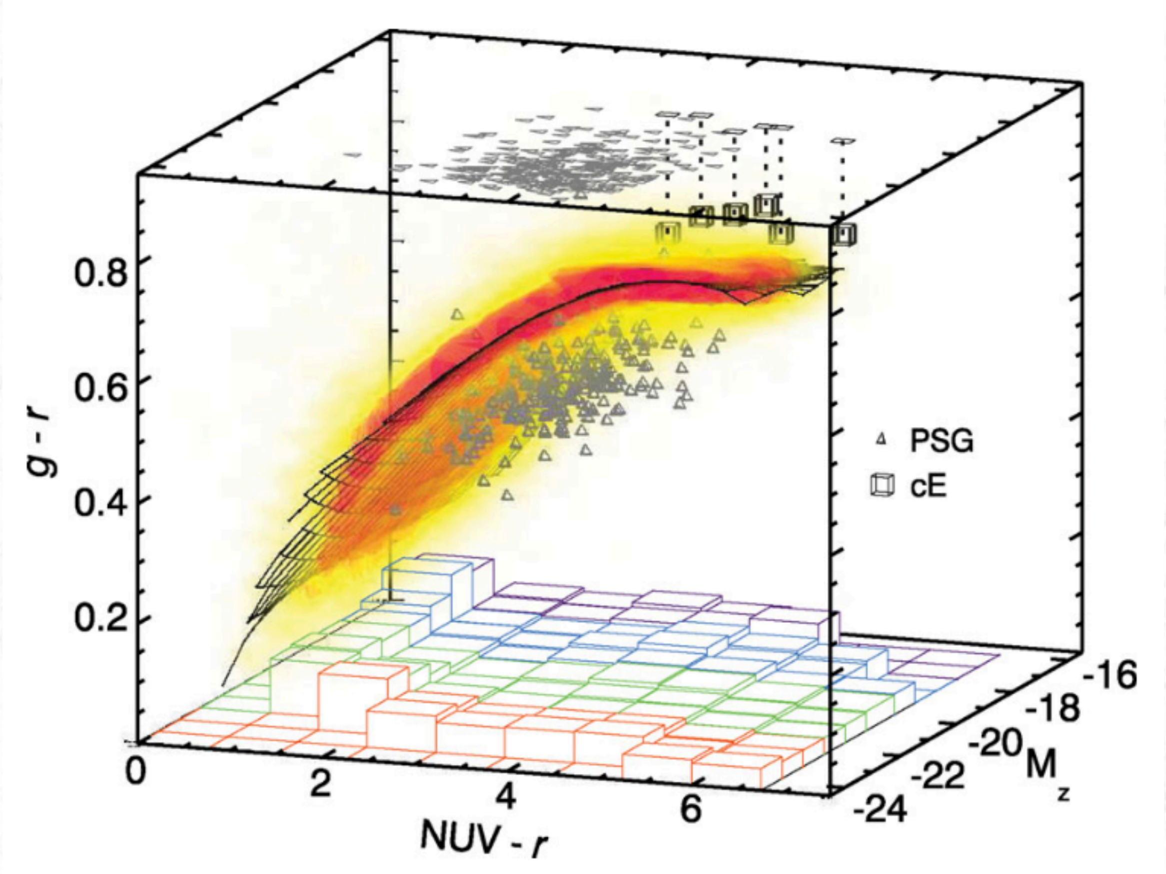
However, the groups don’t seem to be separated in this projection!

Intrinsic smooth distribution?

Possibly a
manifold (多様体)?

Galaxy Manifold?

Studies have claimed a smooth relation of galaxies in the 3D colour–colour–magnitude space smoothly continuing from the 'blue cloud' to the 'red sequence' (e.g. Chilingarian et al. 2012)



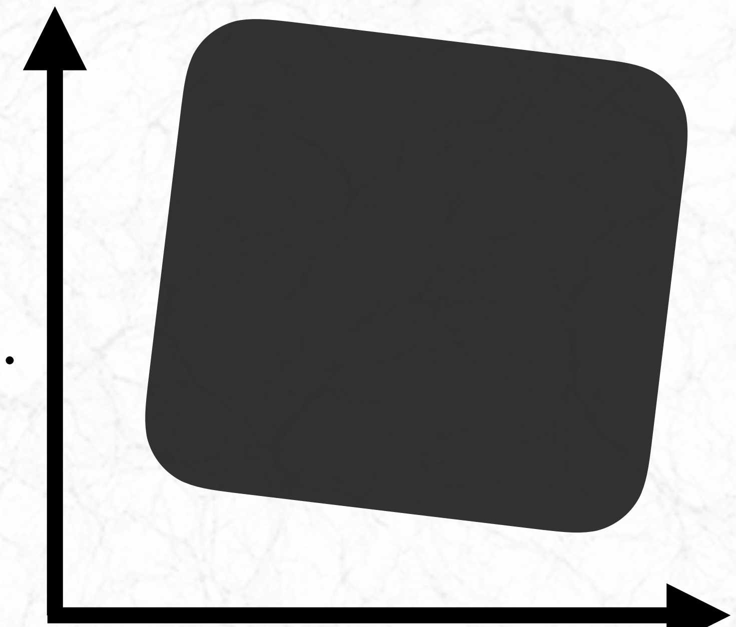
Chilingarian et al. (2012)

Suggests a low dimensional space existing within a higher dimensional space (**Manifold**)

What is the Best Representation of the Manifold?

If the manifold is 2D (surface), only two parameters are sufficient to fully describe it

$$\beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3 + \ldots$$



Linear combination of luminosities

$$\alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3 + \ldots$$

The parameters that best describes the manifold will be a combination of the parameters in the higher dimensional space

However, we mostly only care about magnitudes and colors...

Question 2

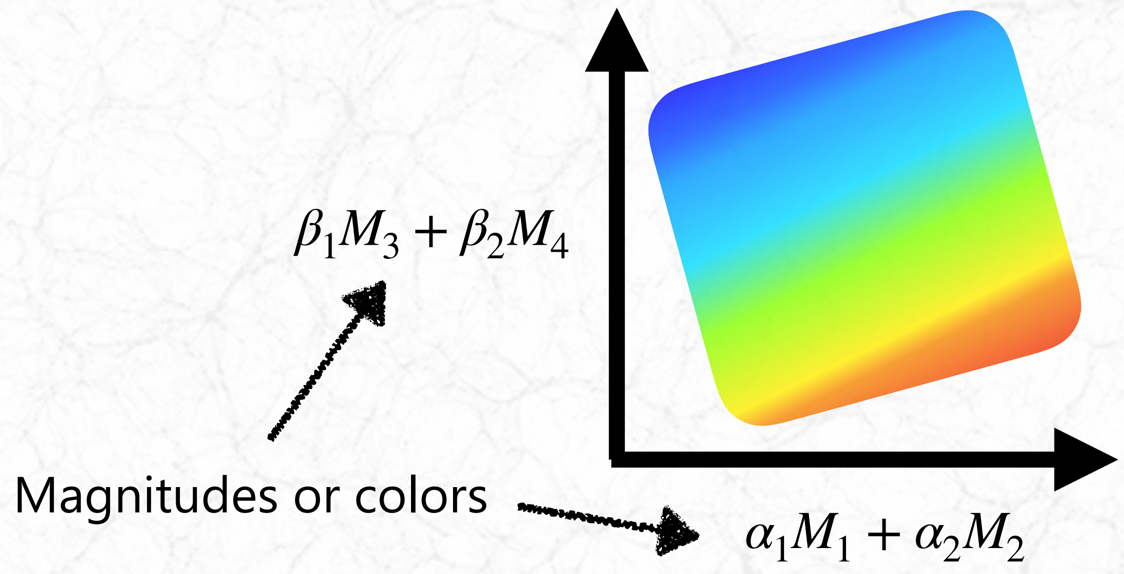
Can we find the magnitudes or colors that best describes a galaxy in the manifold?

Feature Selection

Can we use a machine learning technique to identify which magnitudes/colors (features) best represent the manifold?

Yes! We just need to ask the right question...

What features separates the 11 galaxy classes we identified, the most?

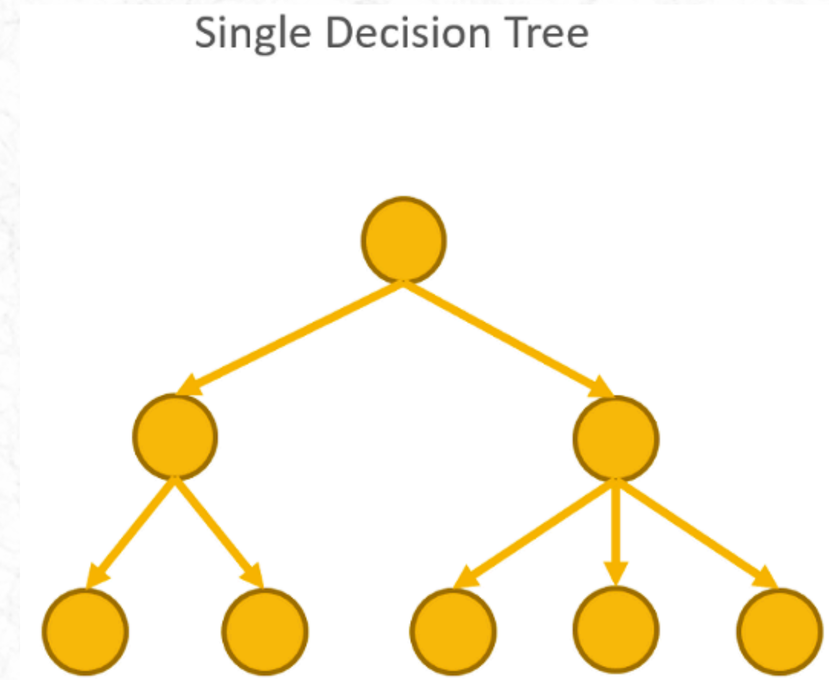


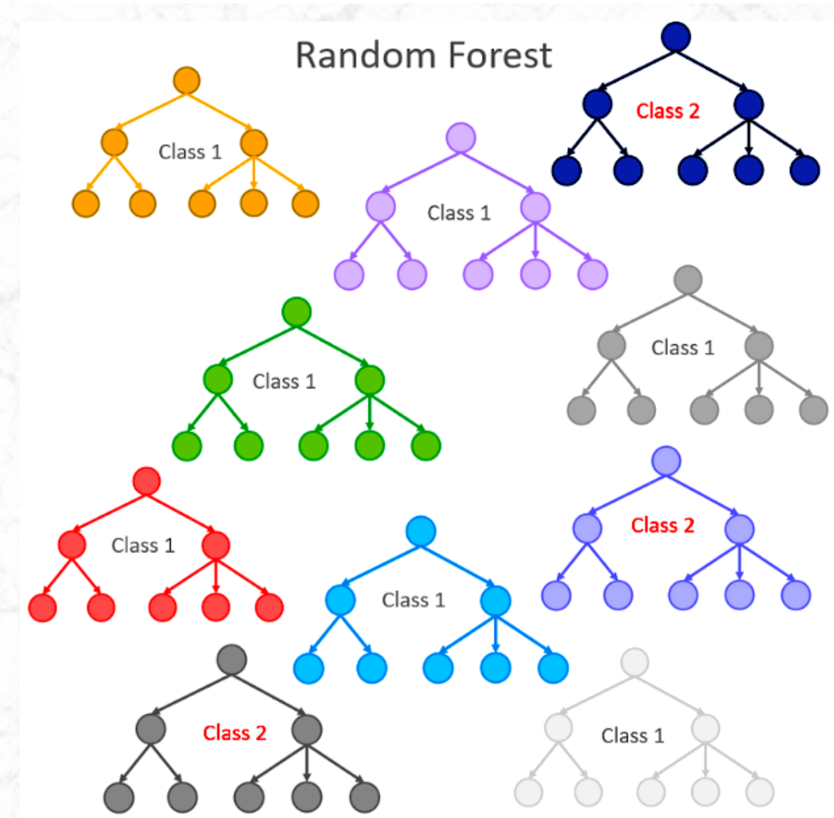
Random Forests for Feature Selection

- **Random Forests or random decision forests** are clustering algorithms that generates random decision trees based on the given features

- Many of such trees produce a random forest

- By providing the answer (class membership), the random forest will **“learn”** the important features



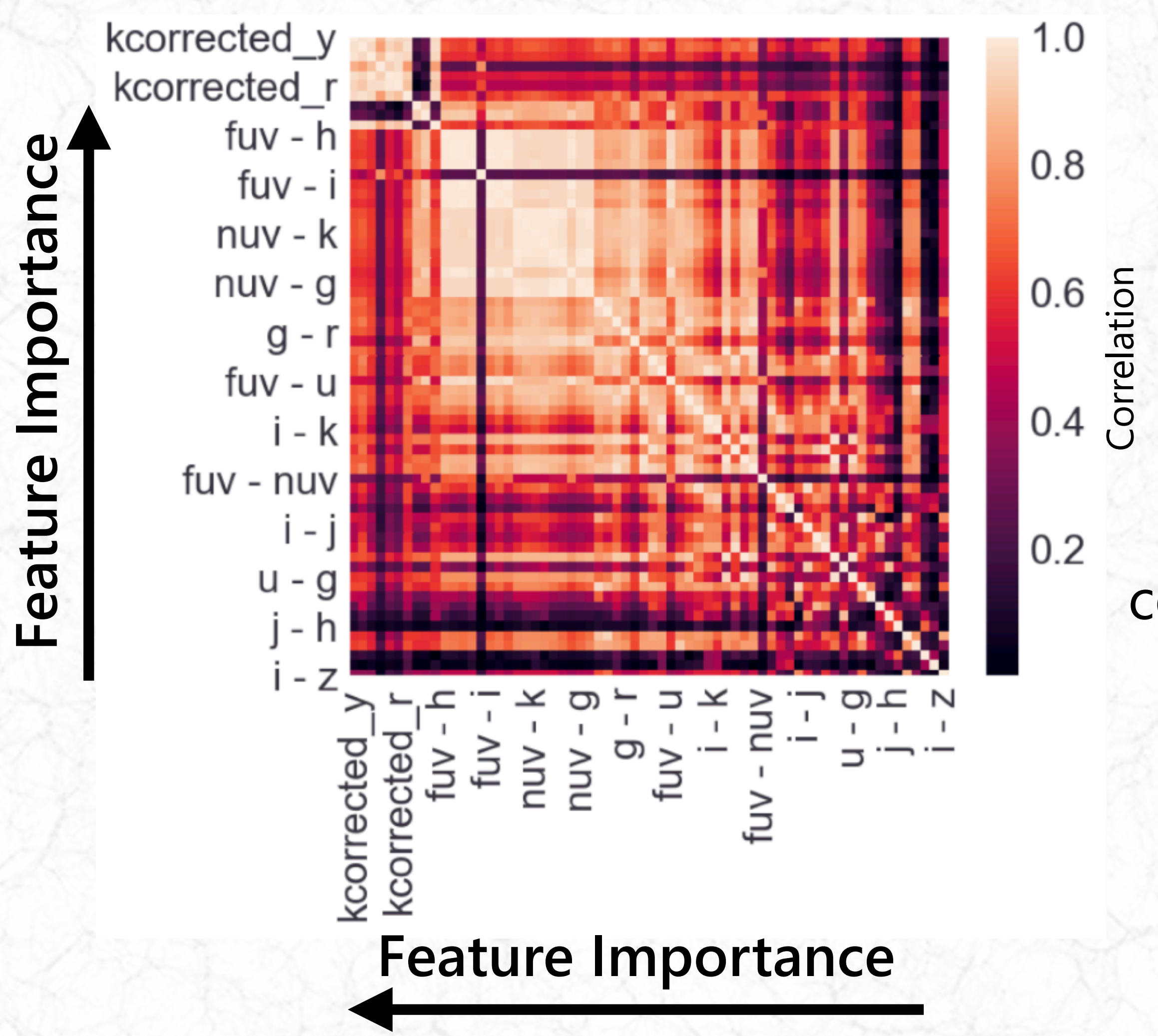


https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147

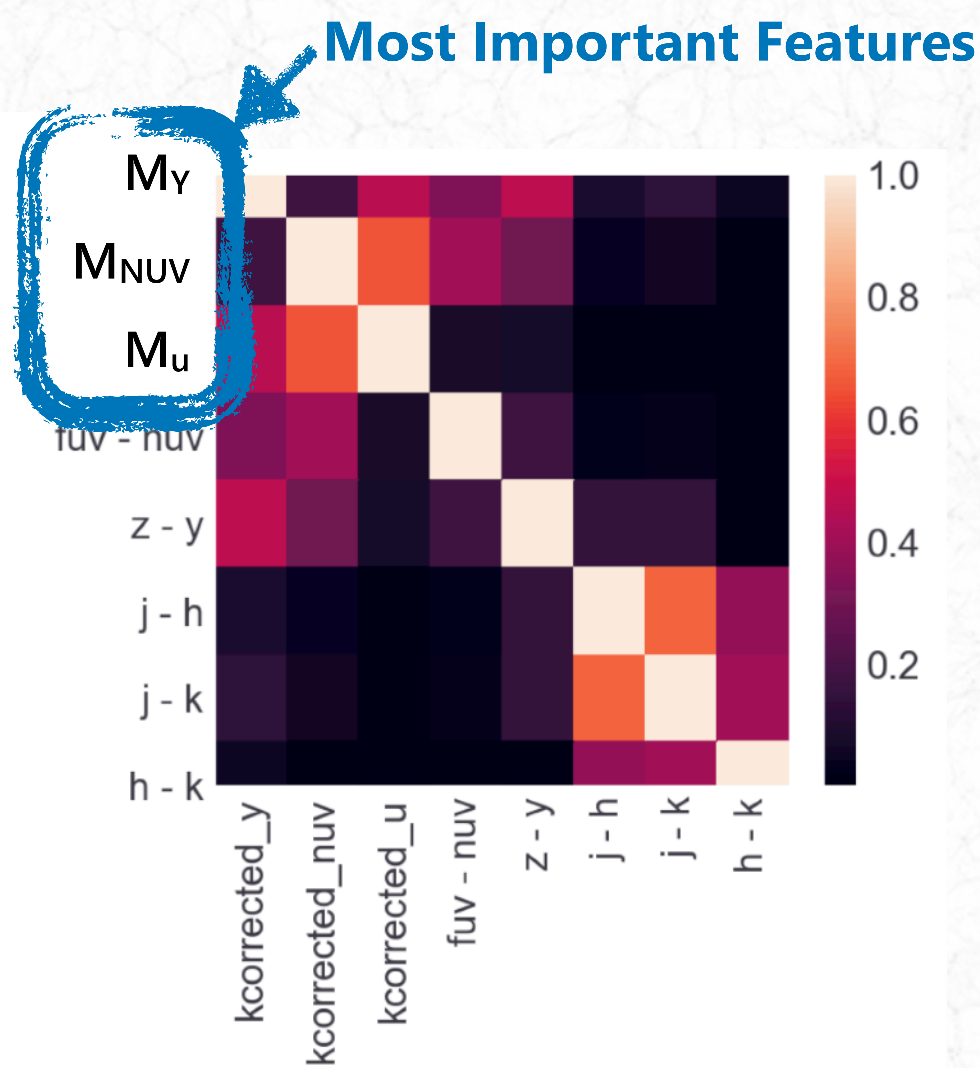
Feature Selection for Galaxy Manifold

Input all the magnitudes and colors with class info → **Random Forest** → Features ranked based on importance

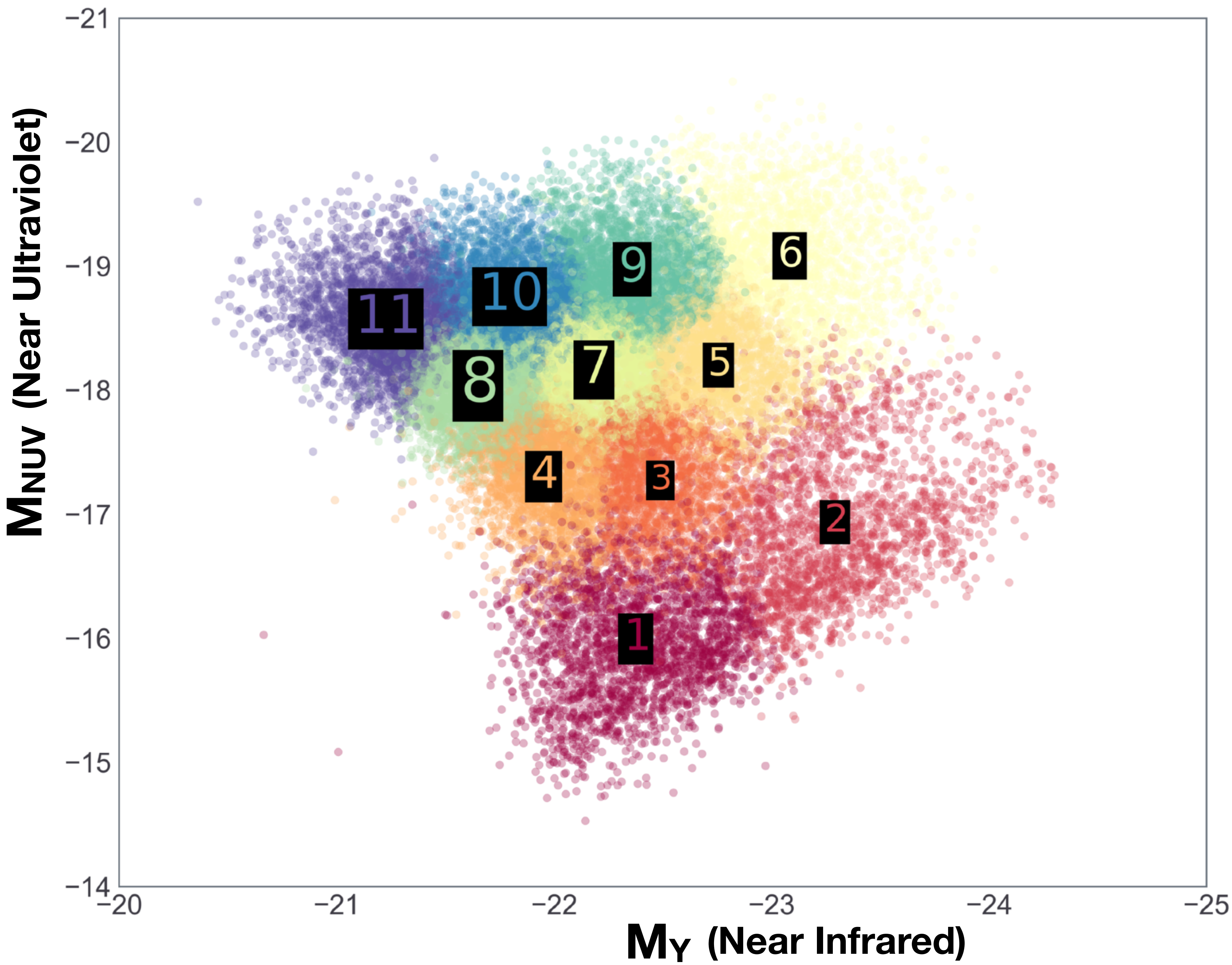
Correlation Heat Map



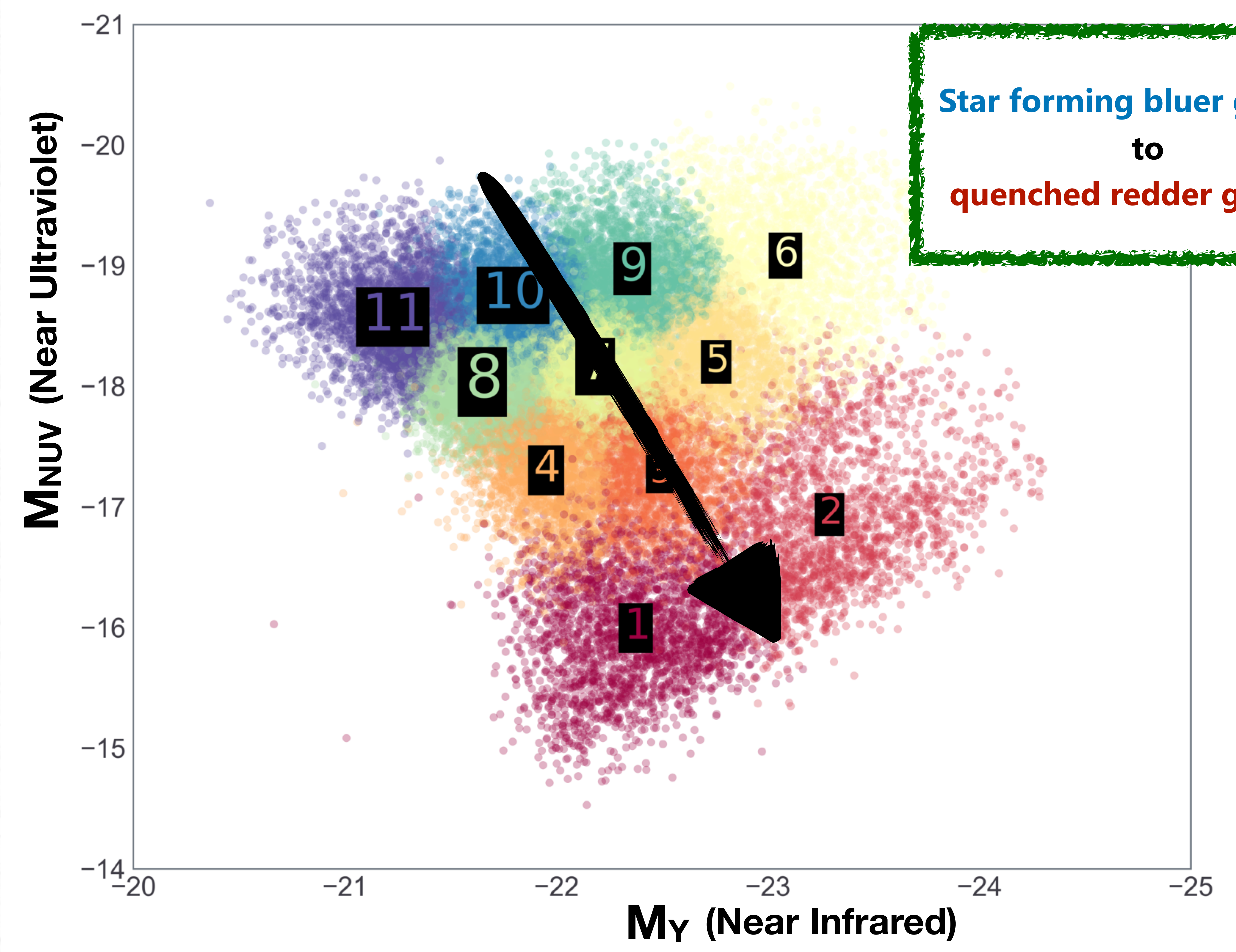
Remove highly correlated features



2 Dimensional Galaxy Manifold



2 Dimensional Galaxy Manifold

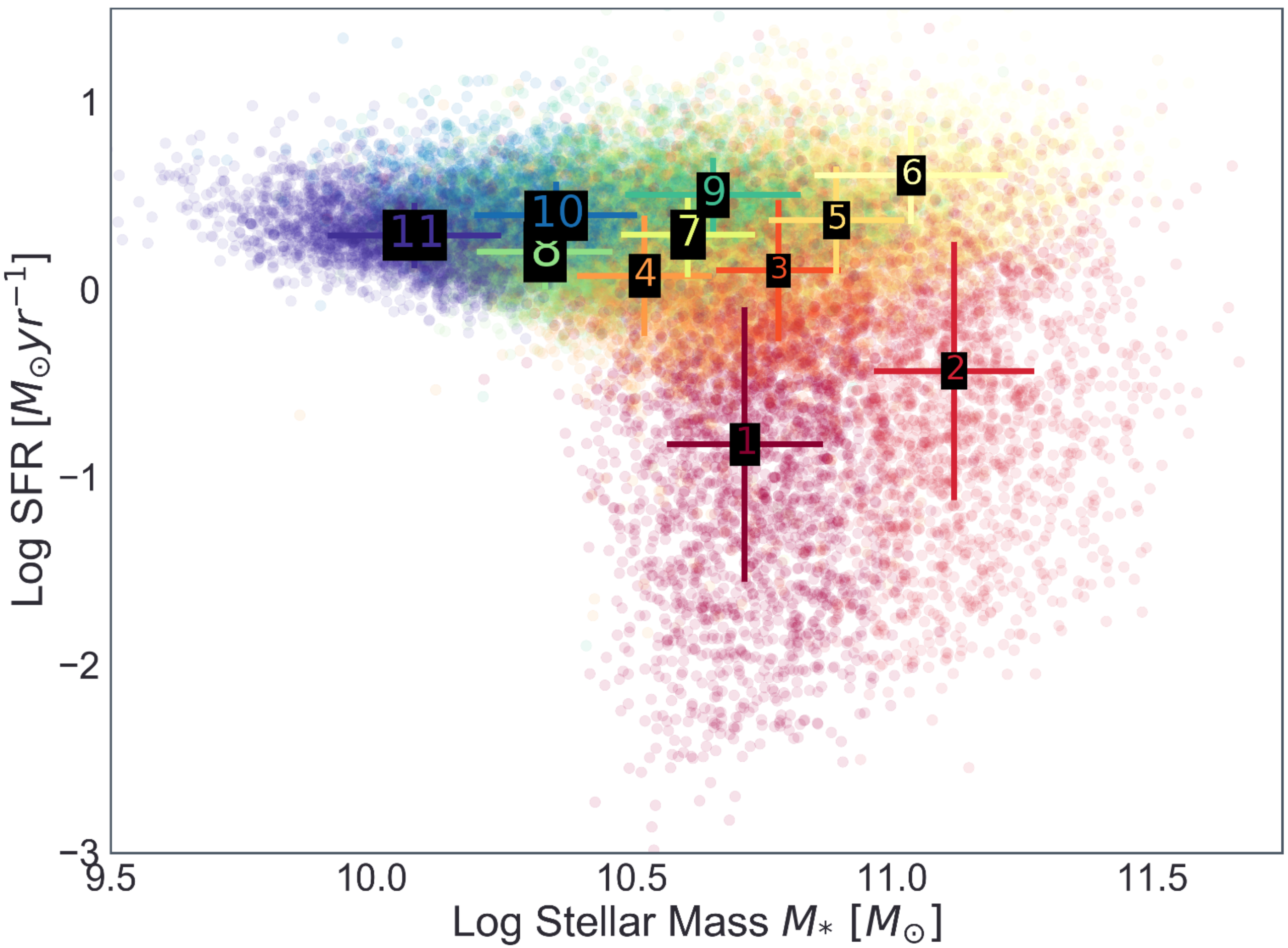


Star forming bluer galaxies

to

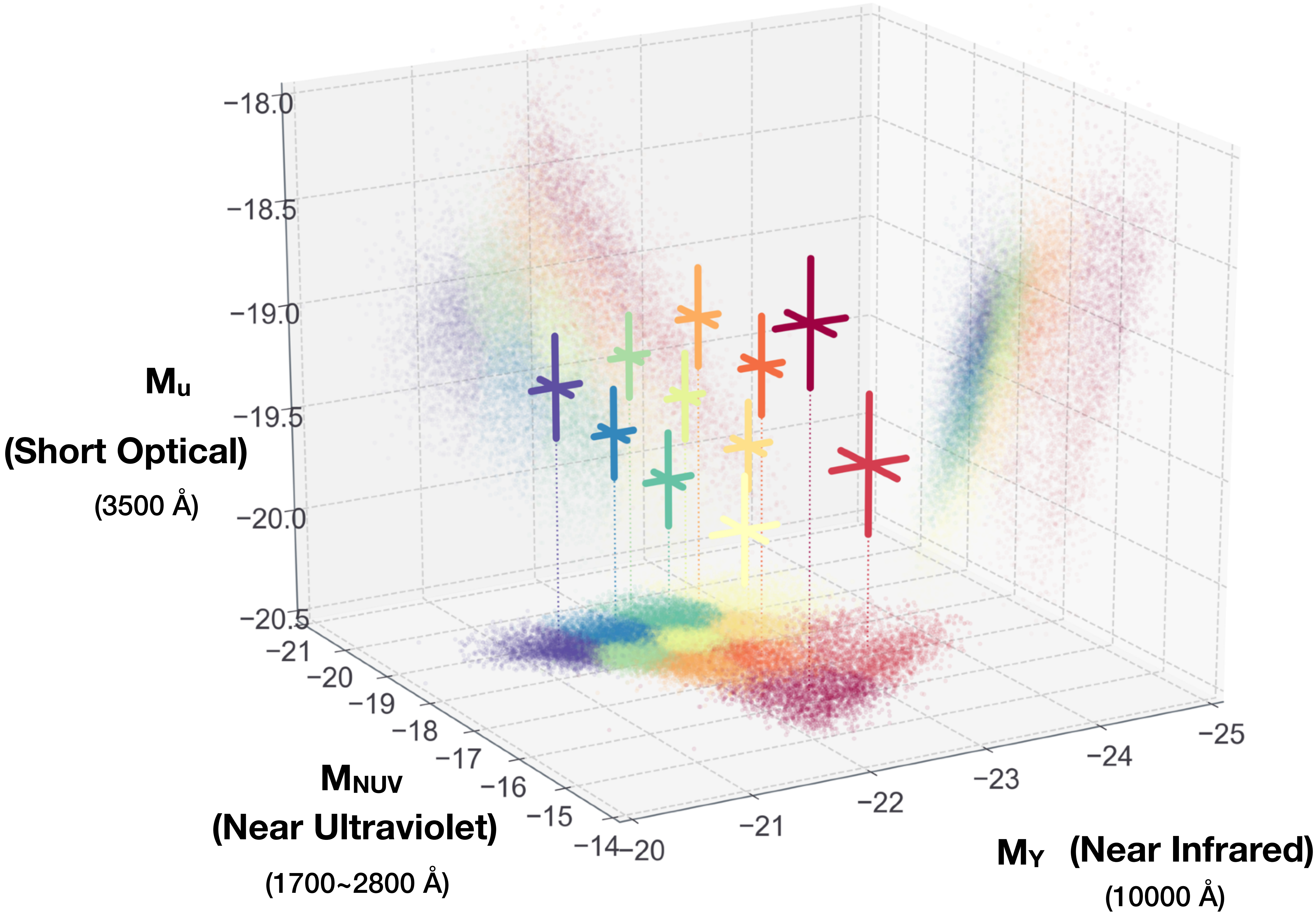
quenched redder galaxies

Comparison to SM-SFR Relation



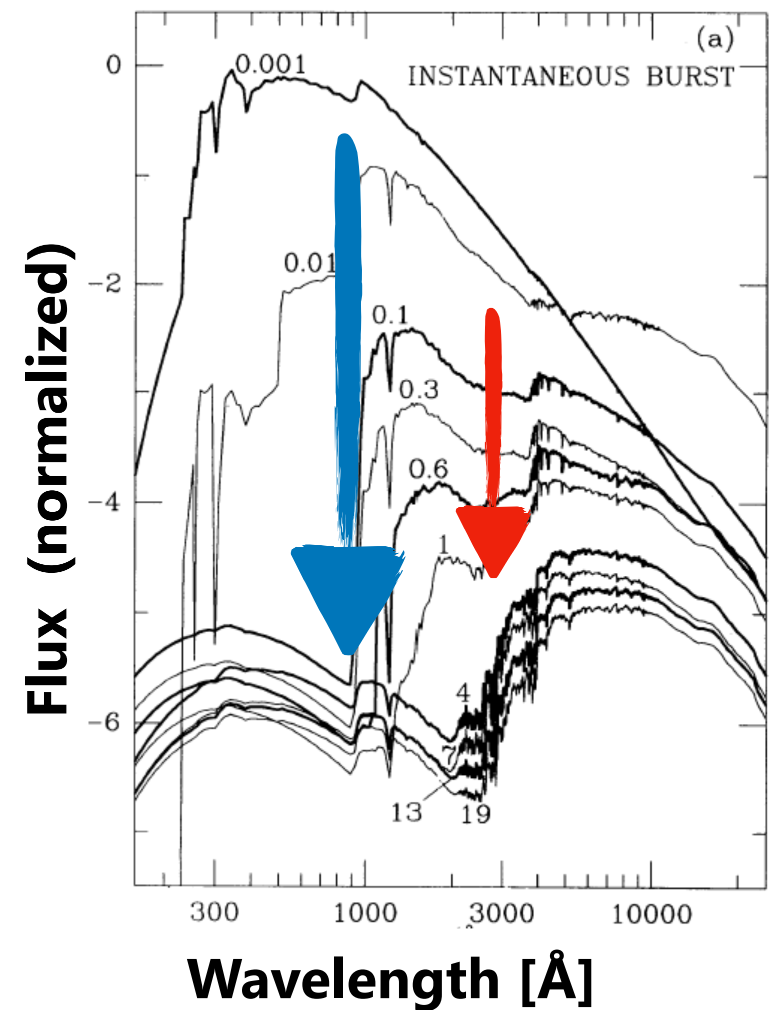
Cross-matched with GSWLC (Salim, Boquien & Lee 2018)

3 Dimensional Galaxy Manifold



Implications of the Results

- The distribution in $M_{NUV} - M_{Y}$ can be explained intuitively by a SSP evolution
- Surprisingly, **luminosities** and **not colors** are the best discriminators
- **Multimodalities** and **dispersions** are effects of suboptimal projections



Bruzual & Charlot (1993)

We should rethink how we classify galaxies. Maybe we should characterize the evolution in terms of a continuous parameter

Summary

- We demonstrated how machine learning methods can aid our understanding of galaxy evolution
  - Finding unobvious patterns in the multidimensional luminosity space
  - Determining the “Galaxy Manifold” and the best projection

Machine Learning algorithms have allowed us to do an unbiased study of galaxy evolution.

The results suggest we fundamentally rethink our galaxy studies

Thanks!

Looking forward to the banquet!