Dimensionality Reduction

Principal Component Analysis (PCA)

t-Stochastic Neighbor Embedding (t-SNE)

Uniform Manifold Approximation and Projection (UMAP)

Quick review

Quick review

We've learned several useful methods already.
What sorts of things are we now good at?

Quick review

We've learned several useful methods already. What sorts of things are we now good at?

Can I always use these? If not, what are the requirements in order to run these methods?

Quick review

We've learned several useful methods already. What sorts of things are we now good at?

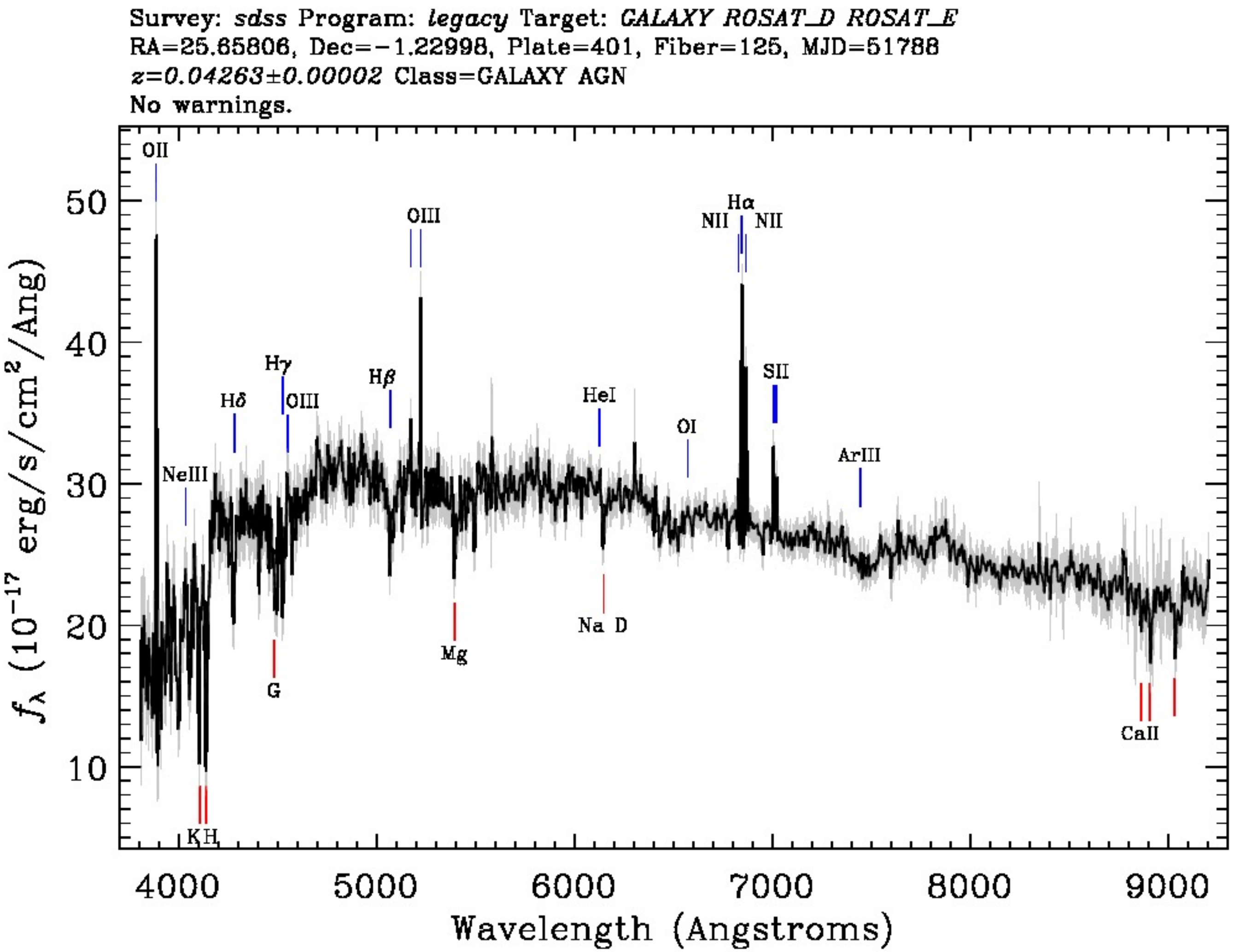
Can I always use these? If not, what are the requirements in order to run these methods?

What should we do if we don't have labels?

What should we do if we don't have labels?

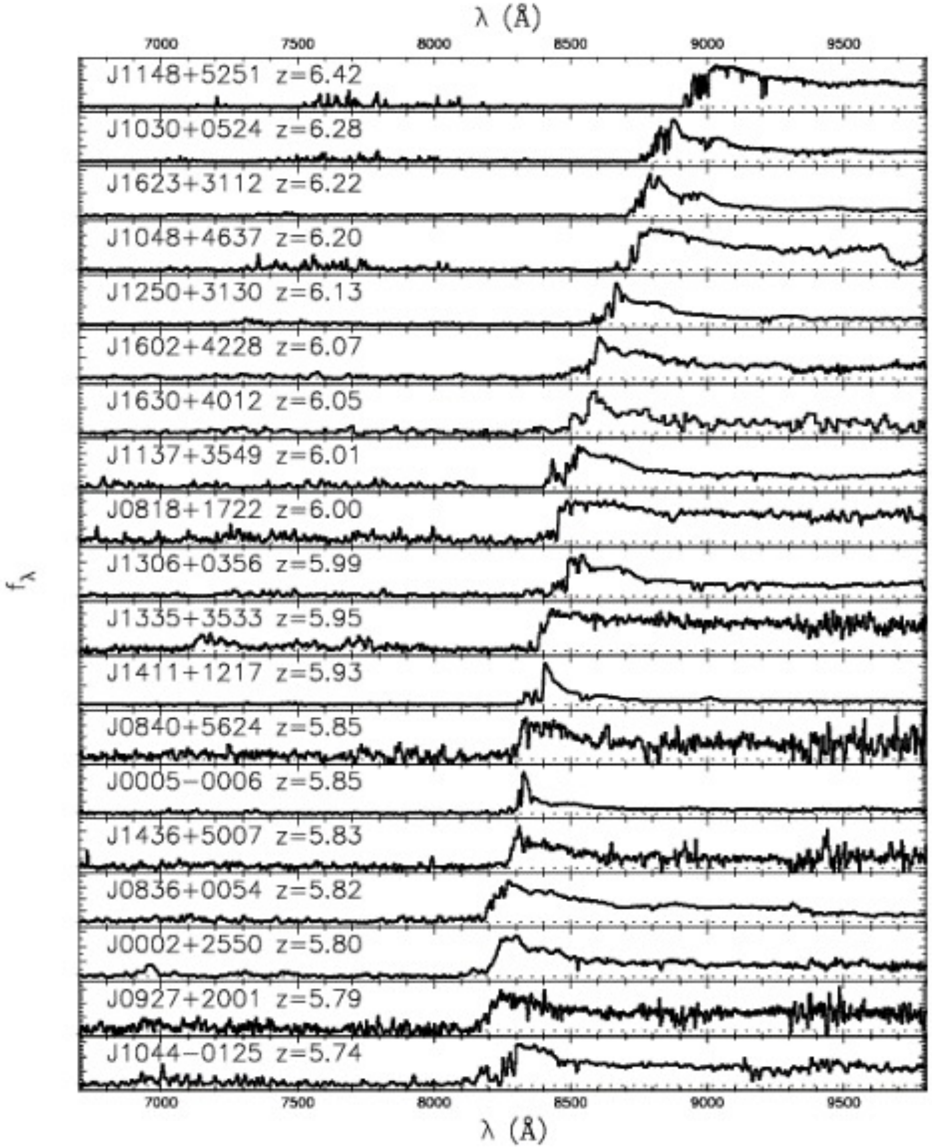
Example: Galaxy spectra

SDSS/Galaxy Zoo



Distant "Lyman Break" Galaxies

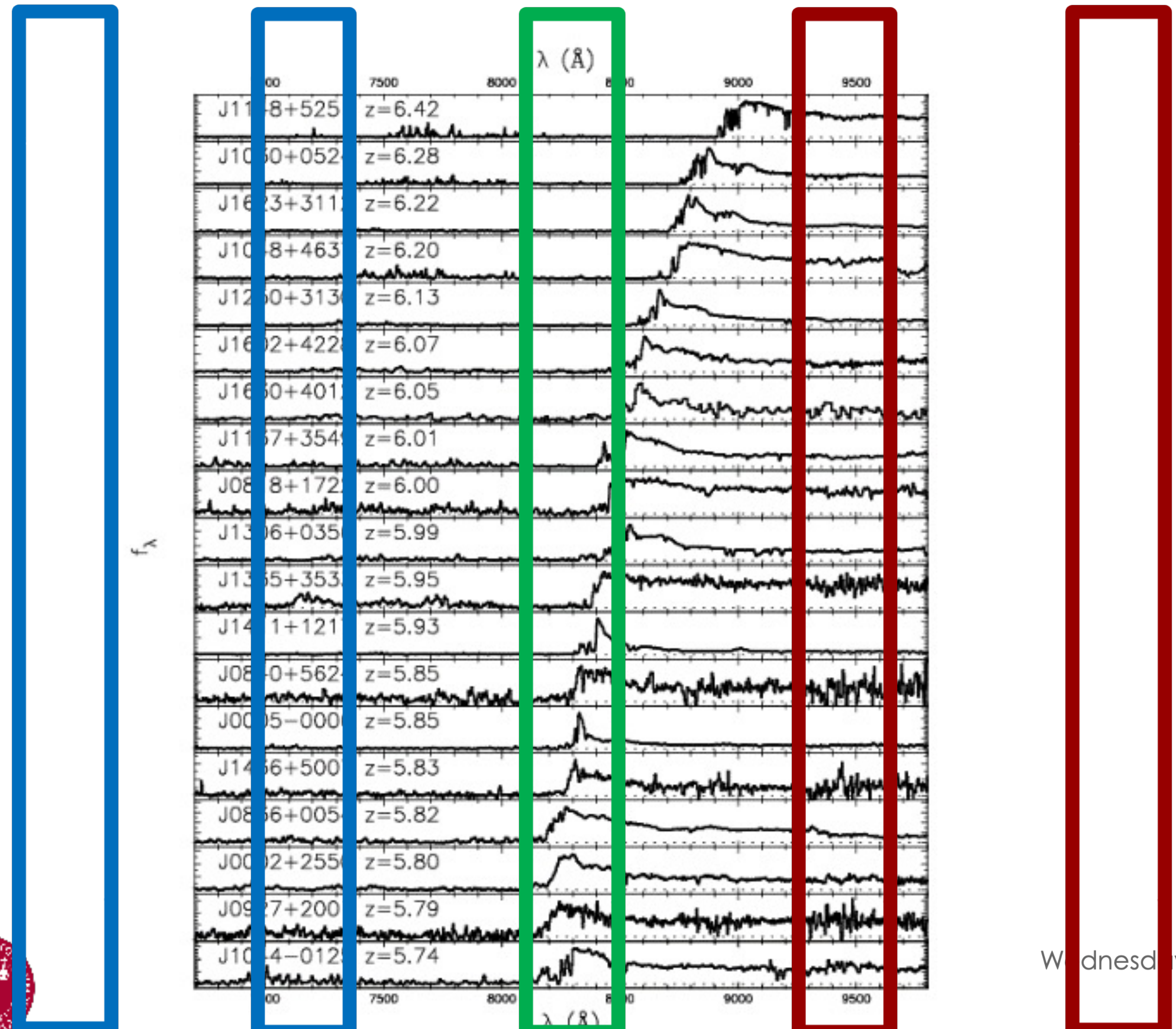
Zaroubi et al. 2013



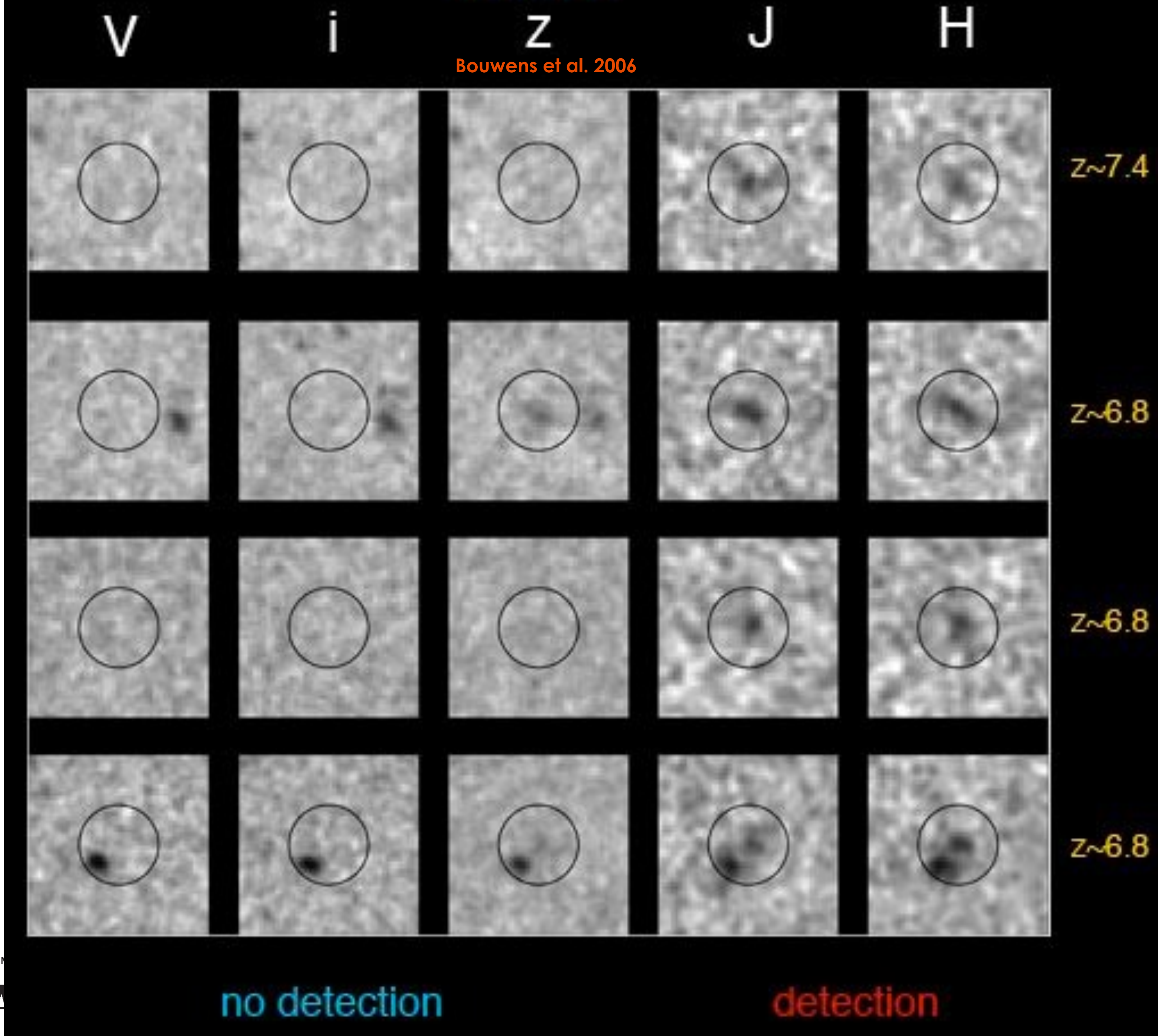
COSMIC DAWN CENTER
DAWN

Distant "Lyman Break" Galaxies

Zaroubi et al. 2013



COSMIC DAWN CENTER
DAWN



Bouwens et al. 2006

Two Fundamental Assumptions for Photometry

COSMIC DAWN CENTER

DAWN

Two Fundamental Assumptions for Photometry

1. If an object is sufficiently well-measured, there is a surjective (one-to-one or many-to-one, but not one-to-many) mapping from photometric fluxes to astrophysical properties

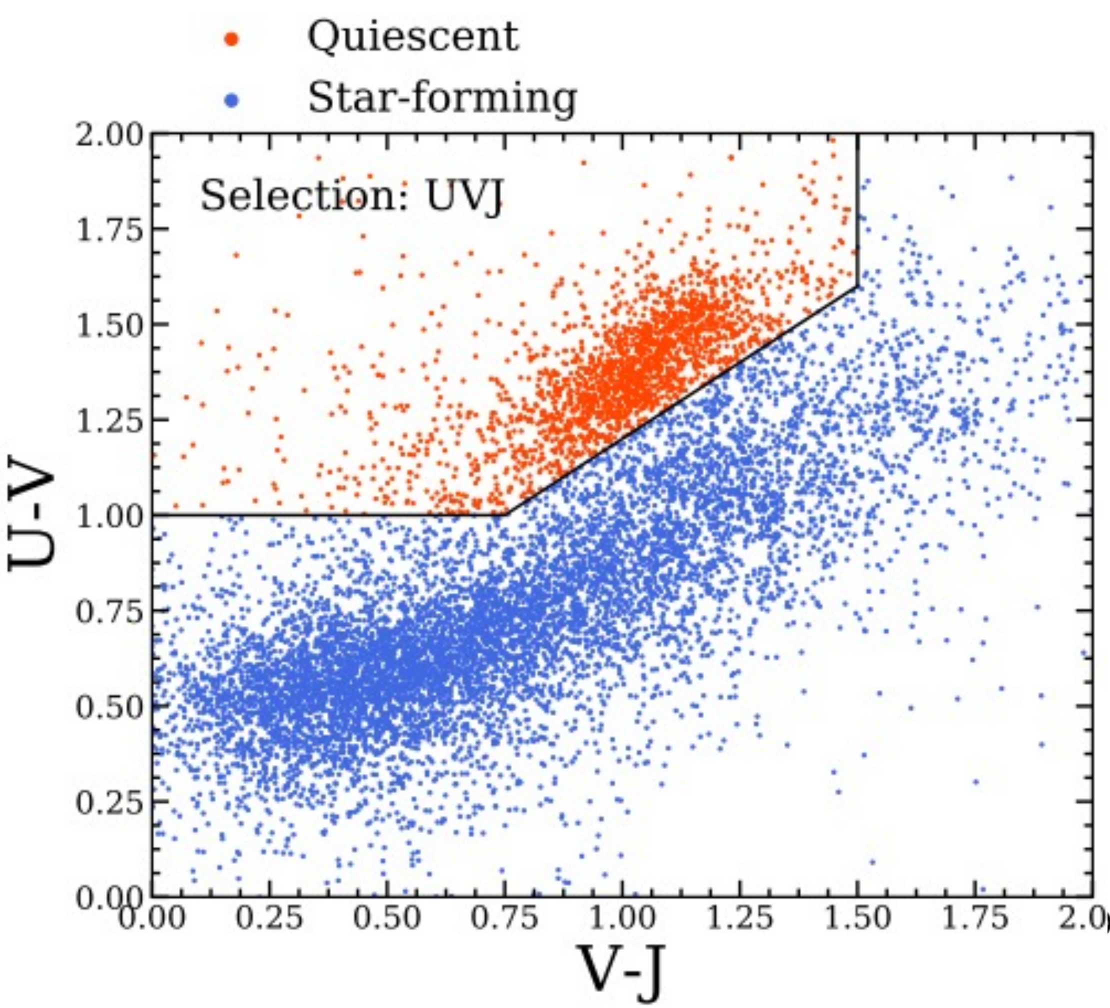
Two Fundamental Assumptions for Photometry

1. If an object is sufficiently well-measured, there is a surjective (one-to-one or many-to-one, but not one-to-many) mapping from photometric fluxes to astrophysical properties.

2. Objects with sufficiently similar photometry should be mapped to similar astrophysical properties.

Approach 1: Color space map using two colors (three bands)

COSMOS2015 catalog, objects at z≈1

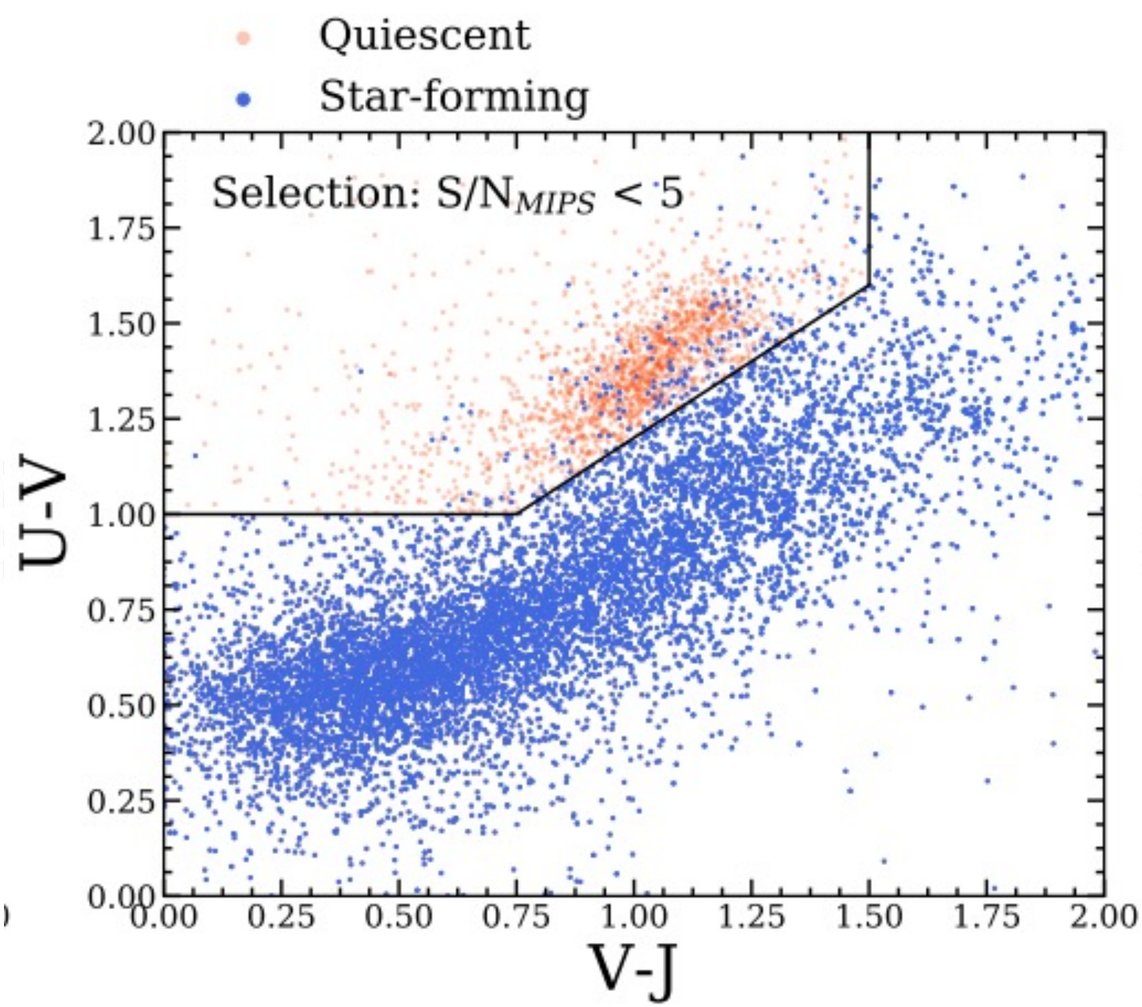


COSMIC DAWN CENTER

DAWN

Approach 1: Color space map using two colors (three bands)

COSMOS2015 catalog, objects at z≈1



COSMIC DAWN CENTER
DAWN

Approach 2: Color space map using all bands

~~Two~~ Three Fundamental Assumptions for Photometry

1. If an object is sufficiently well-measured, there is a surjective (one-to-one or many-to-one, but not one-to-many) mapping from photometric fluxes to astrophysical properties.

2. Objects with sufficiently similar photometry should be mapped to similar astrophysical properties.

3. We can map objects from the full, n-dimensional space with all bands to a smaller one with many neighbors, and the other two assumptions will continue to hold.

~~Two~~ Three Fundamental Assumptions for Photometry

1. If an object is sufficiently well-measured, there is a surjective (one-to-one or many-to-one, but not one-to-many) mapping from photometric fluxes to astrophysical properties.

2. Objects with sufficiently similar photometry should be mapped to similar astrophysical properties.

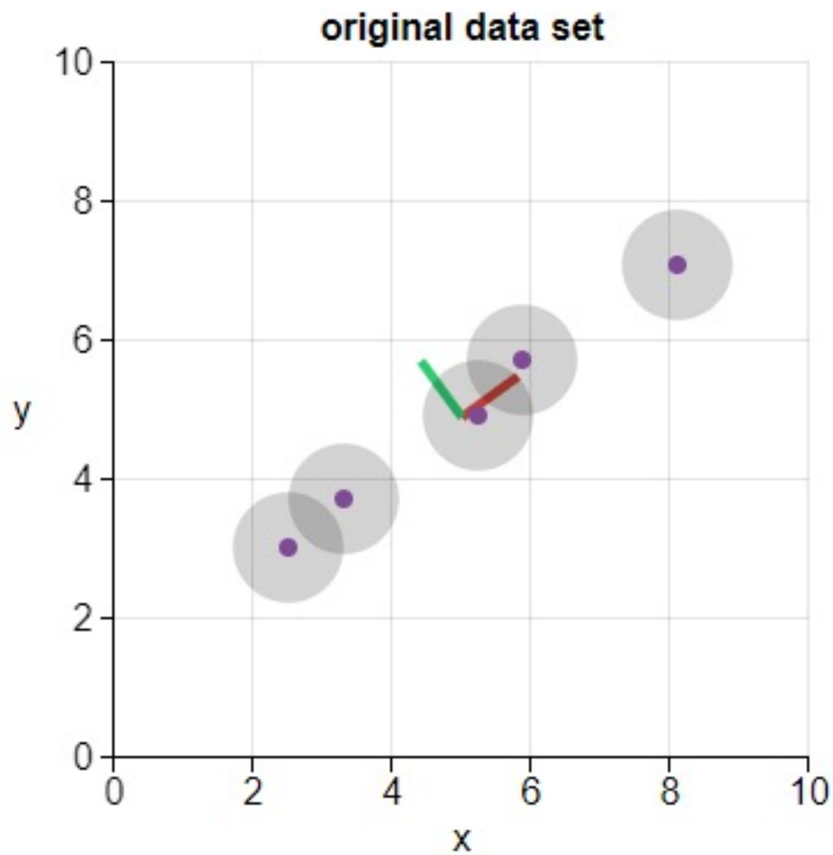
3. We can map objects from the full, n-dimensional space with all bands to a smaller one with many neighbors, and the other two assumptions will continue to hold.

COSMIC DAWN CENTER
DAWN

Can we somehow decide what information is "important" even without labels?

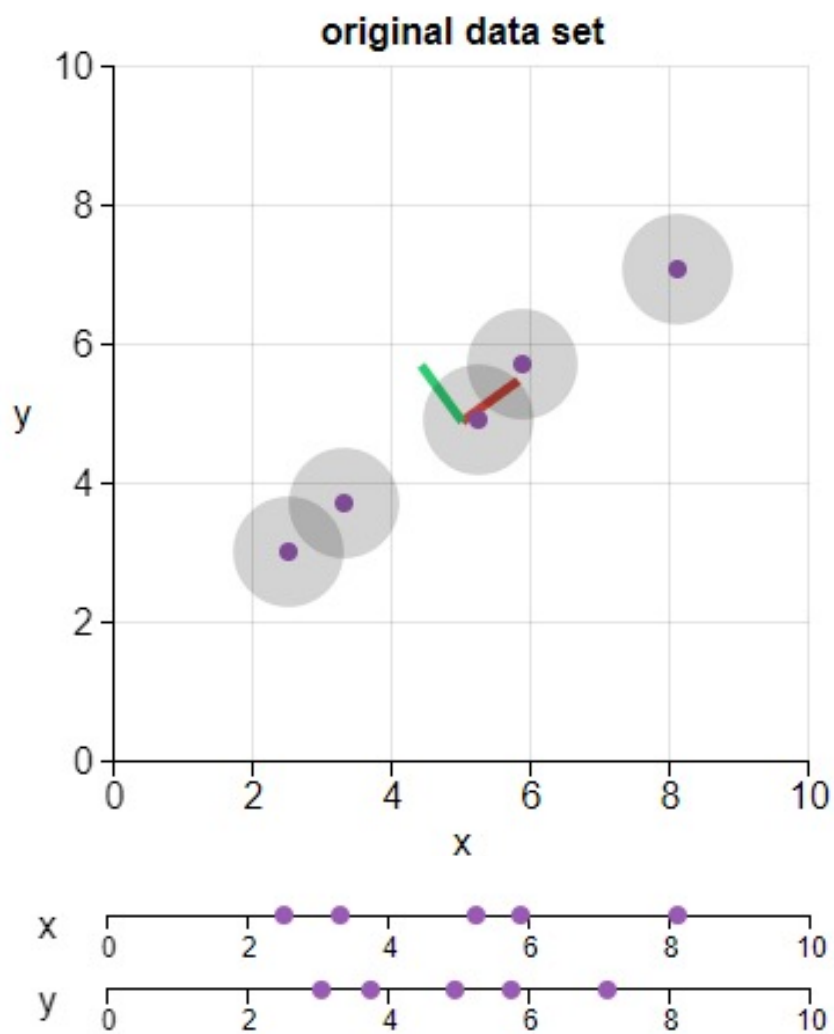
Principal Component Analysis (PCA)

https://setosa.io/ev/principal-component-analysis/



Principal Component Analysis (PCA)

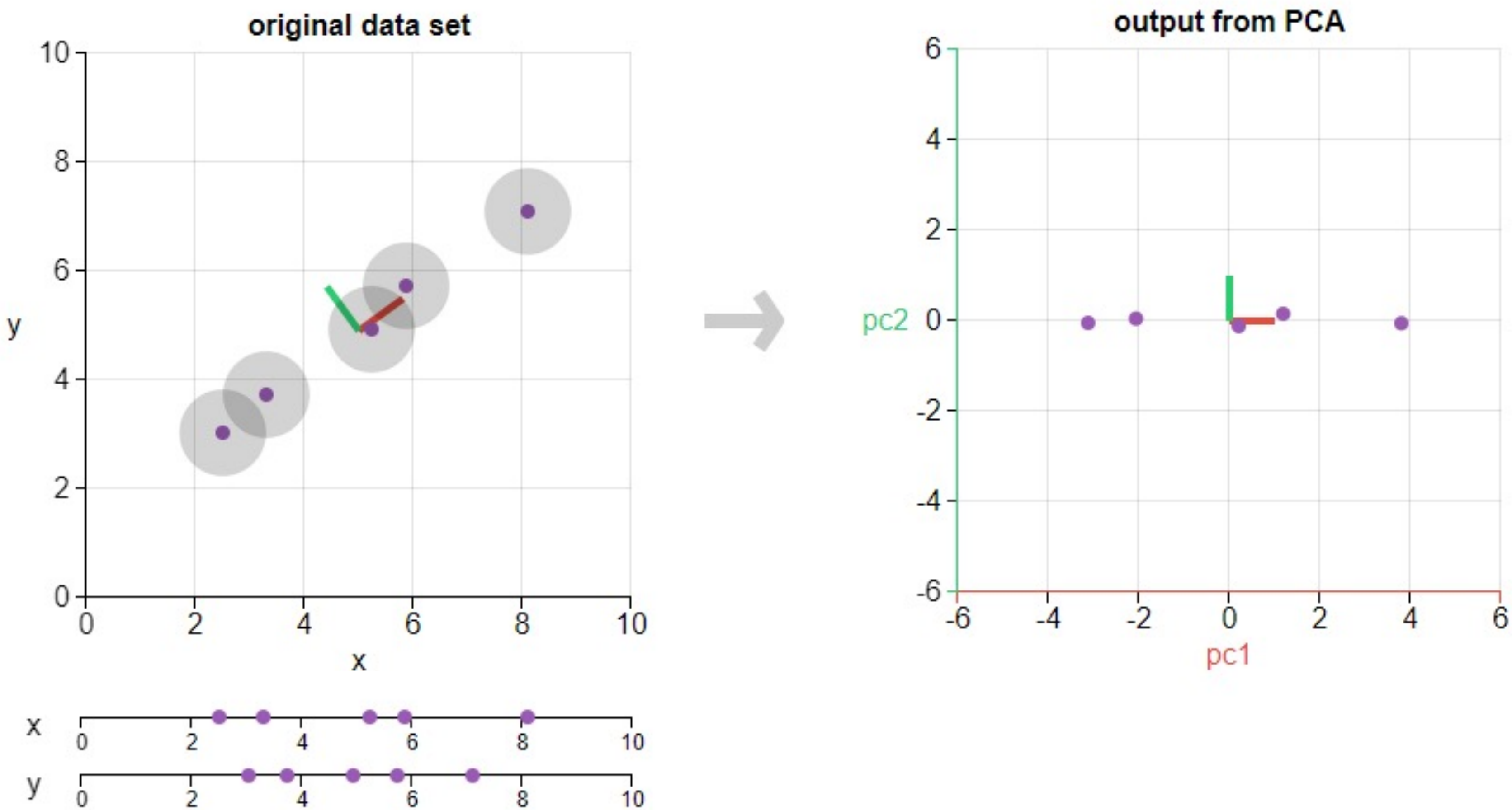
https://setosa.io/ev/principal-component-analysis/



COSMIC DAWN CENTER
DAWN

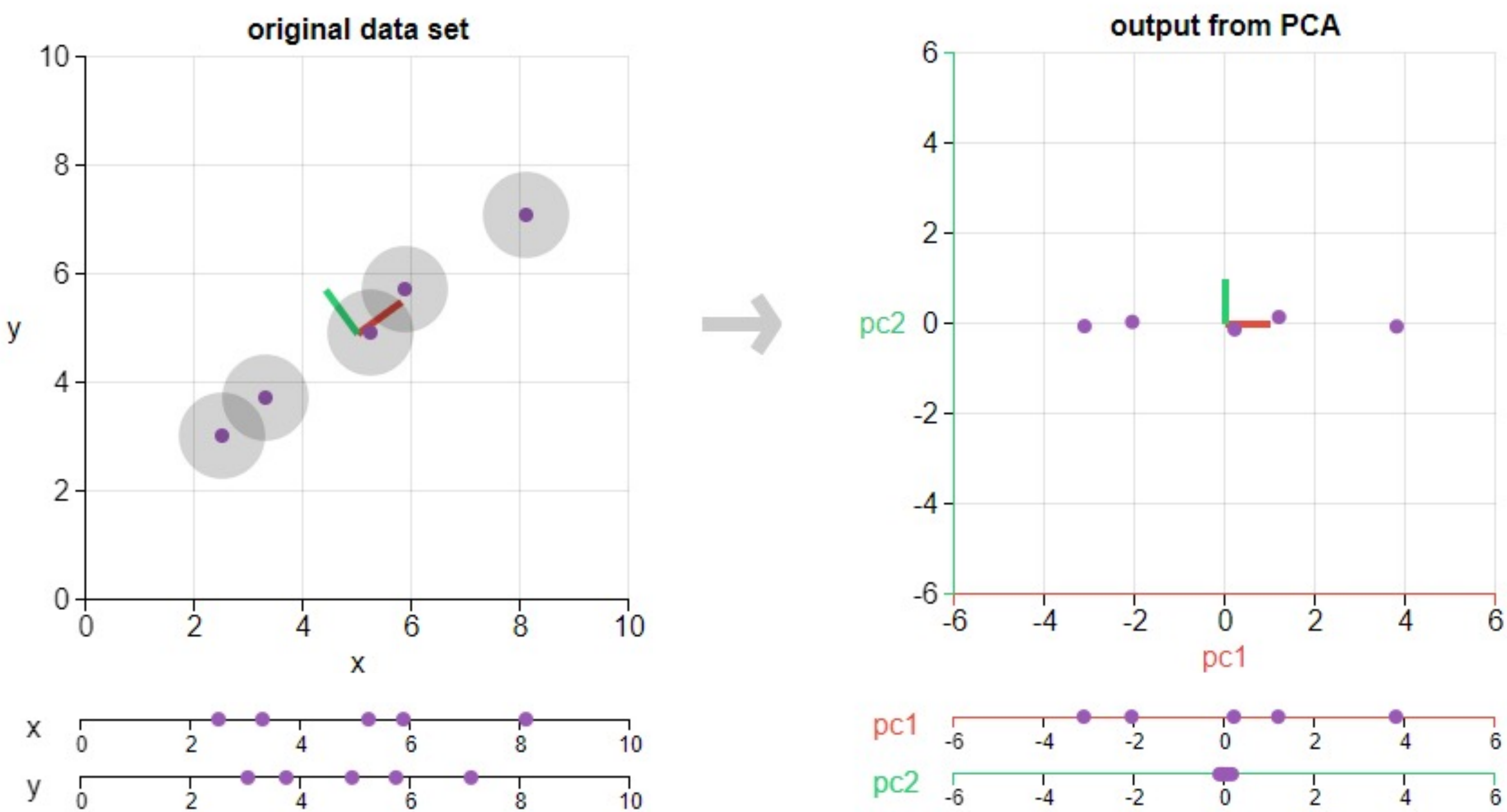
Principal Component Analysis (PCA)

https://setosa.io/ev/principal-component-analysis/



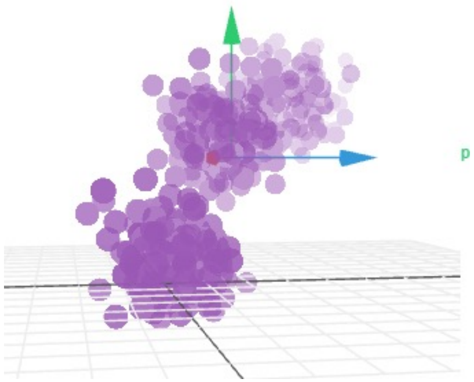
Principal Component Analysis (PCA)

https://setosa.io/ev/principal-component-analysis/



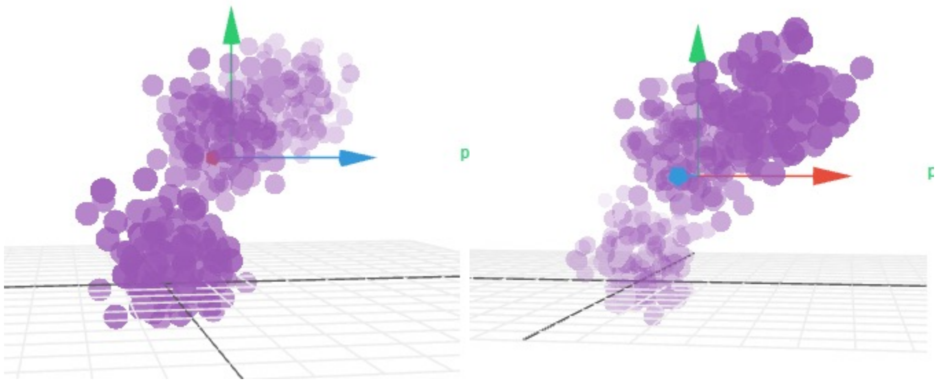
PCA in 3D

https://setosa.io/ev/principal-component-analysis/



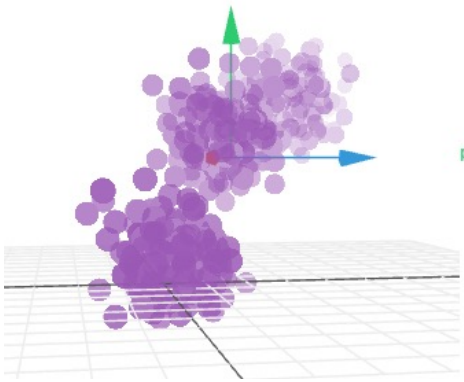
PCA in 3D

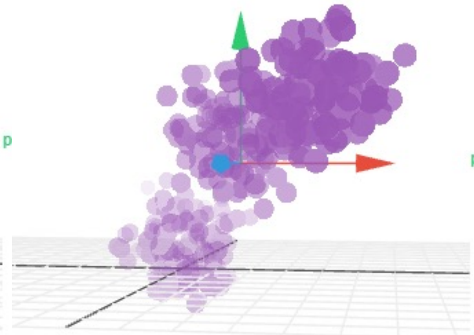
https://setosa.io/ev/principal-component-analysis/

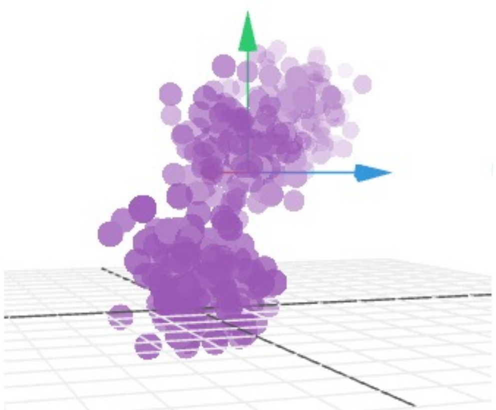


PCA in 3D

https://setosa.io/ev/principal-component-analysis/

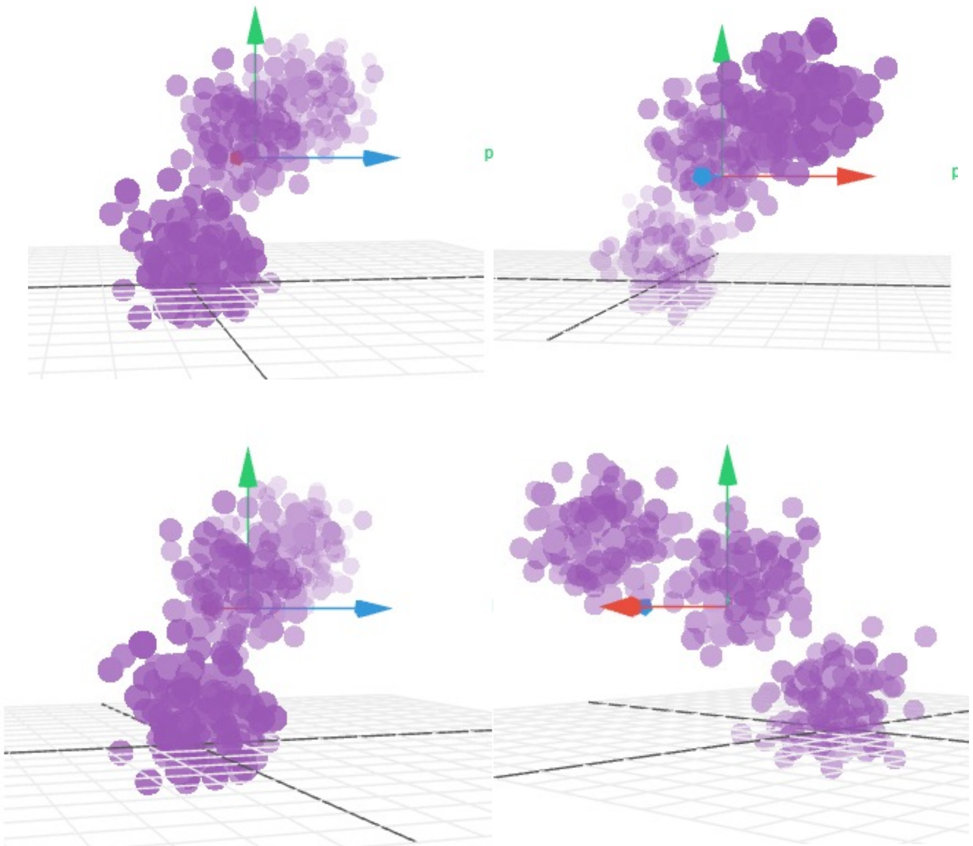






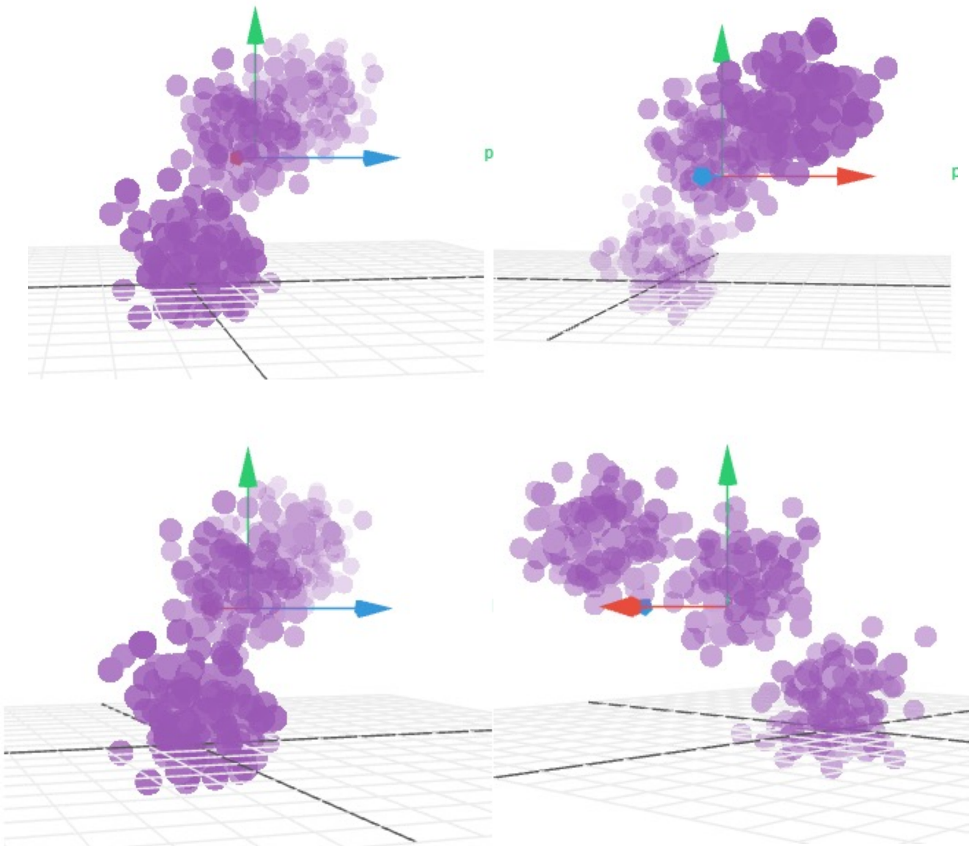
PCA in 3D

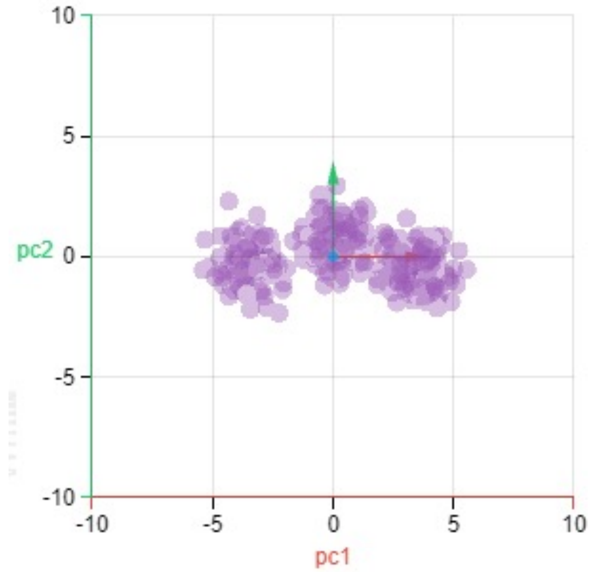
https://setosa.io/ev/principal-component-analysis/



PCA in 3D

https://setosa.io/ev/principal-component-analysis/

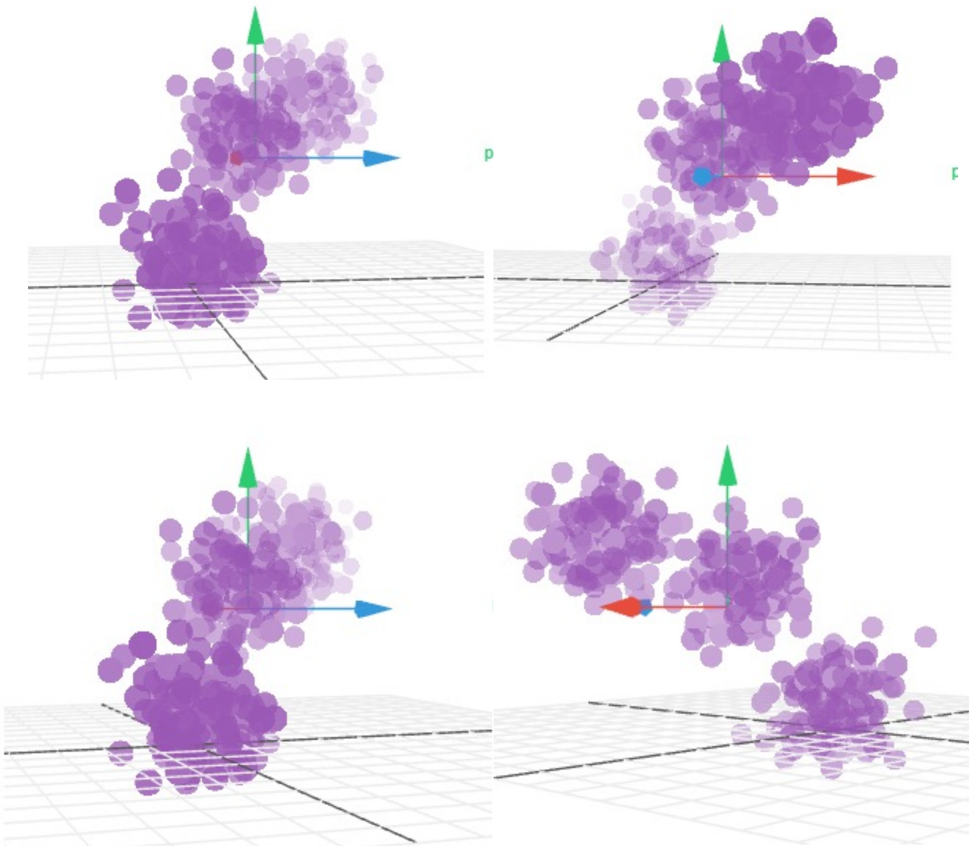


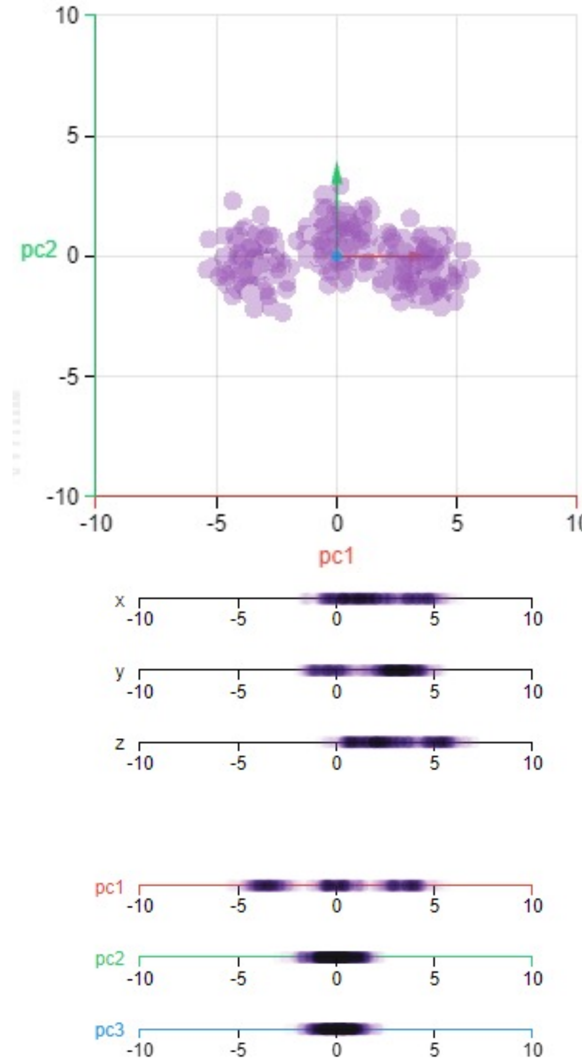


COSMIC DAWN CENTER
DAWN

PCA in 3D

https://setosa.io/ev/principal-component-analysis/





COSMIC DAWN CENTER
DAWN

PCA in 17D

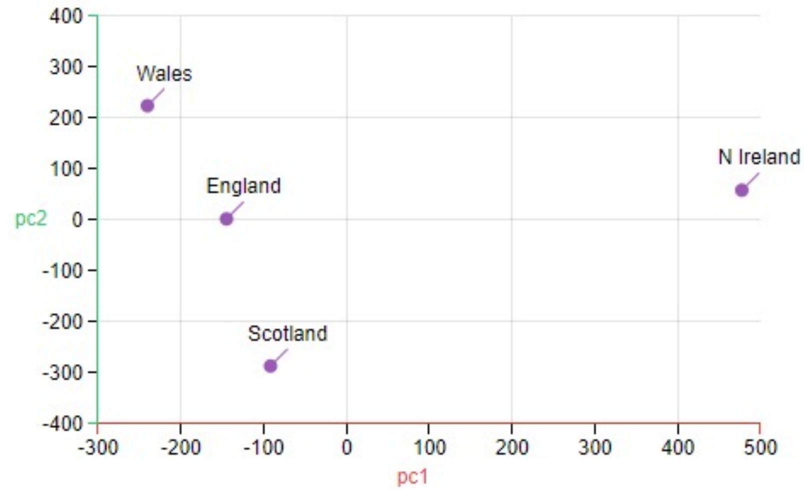
https://setosa.io/ev/principal-component-analysis/

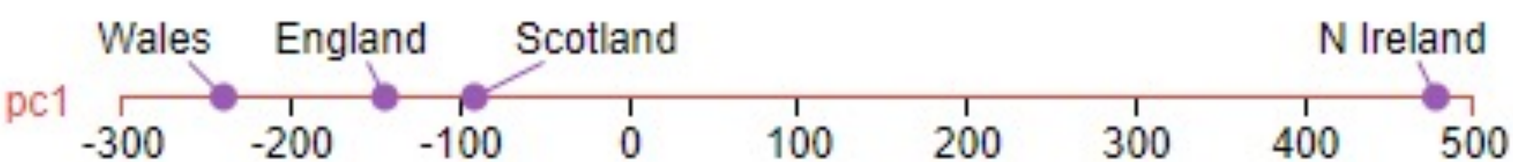
	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

PCA in 17D

https://setosa.io/ev/principal-component-analysis/

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175





COSMIC DAWN CENTER
DAWN

OK, so how can we find the right basis?

1. Standardization

OK, so how can we find the right basis?

1. Standardization

2. Compute covariance matrix

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

OK, so how can we find the right basis?

1. Standardization

2. Compute covariance matrix

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

3. Compute eigenvectors and eigenvalues

OK, so how can we find the right basis?

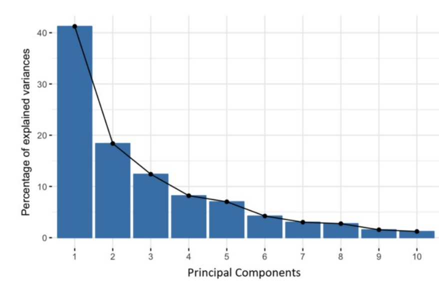
1. Standardization

2. Compute covariance matrix

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

3. Compute eigenvectors and eigenvalues

4. Discard vectors that are not important enough



COSMIC DAWN CENTER
DAWN

Example: Handwritten Digits

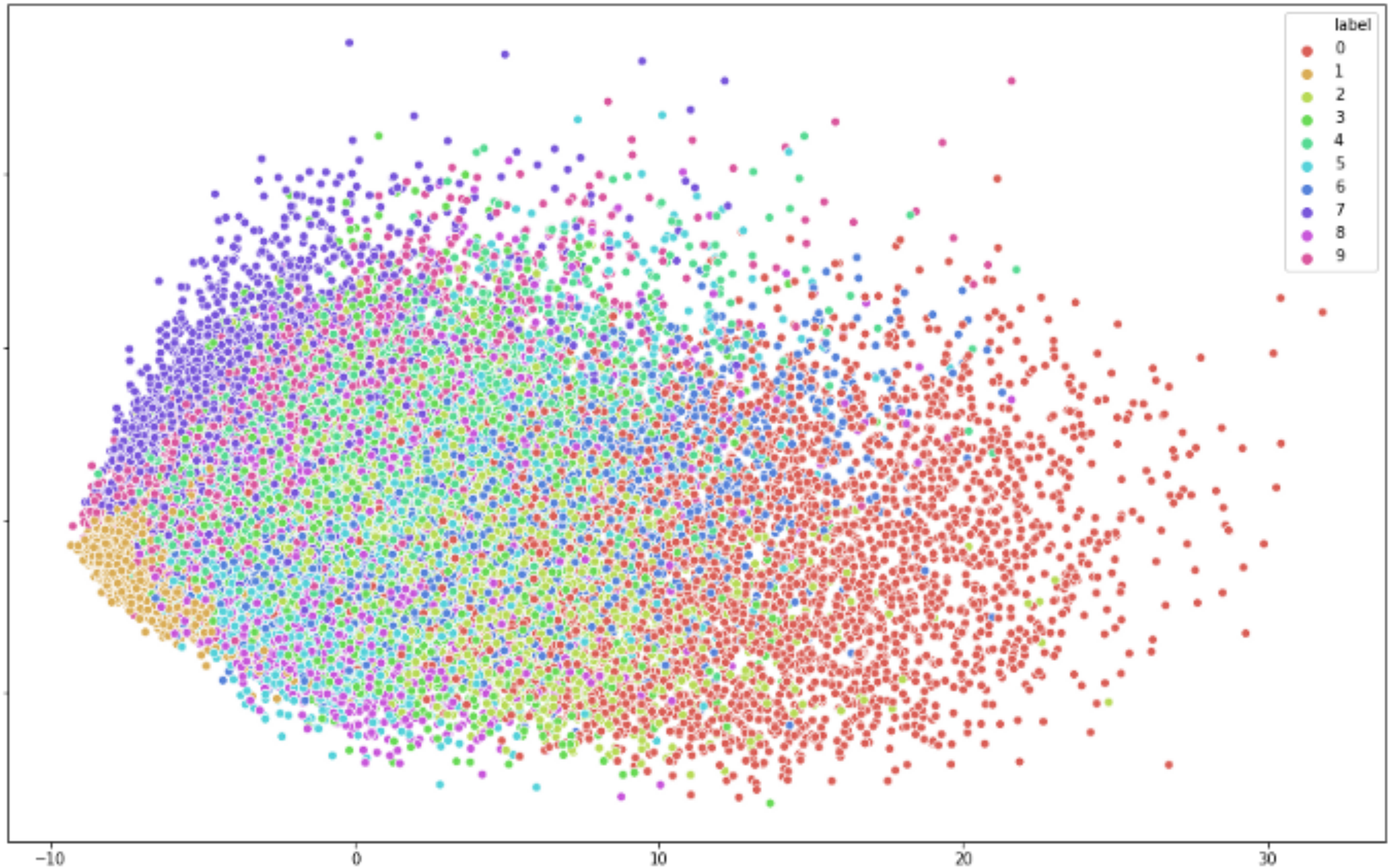
MNIST dataset

COSMIC DAWN CENTER

DAWN

Group "similar" things together

Principal Component Analysis

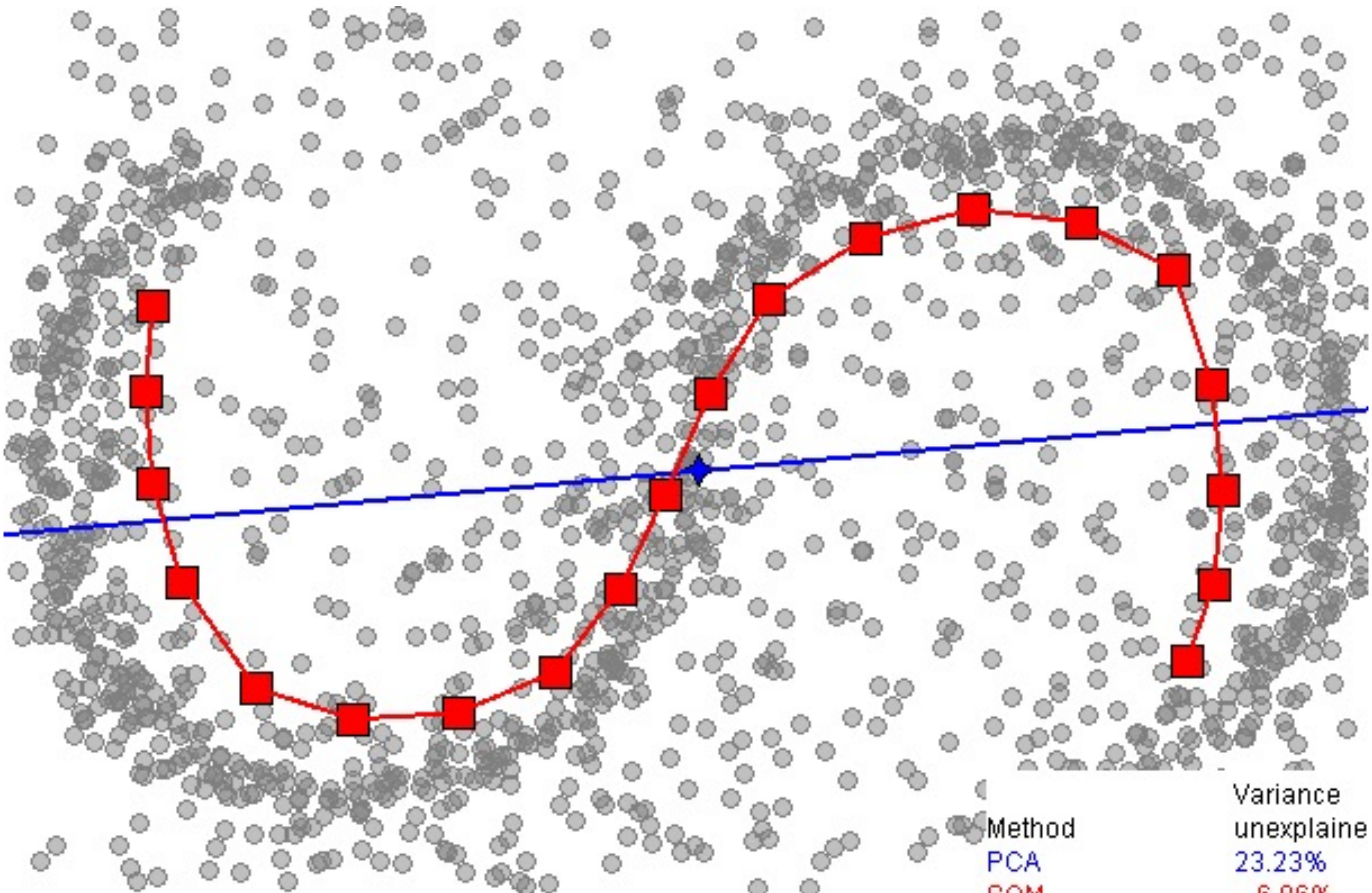


Example: Handwritten Digits

MNIST dataset

Some things aren’t linear!

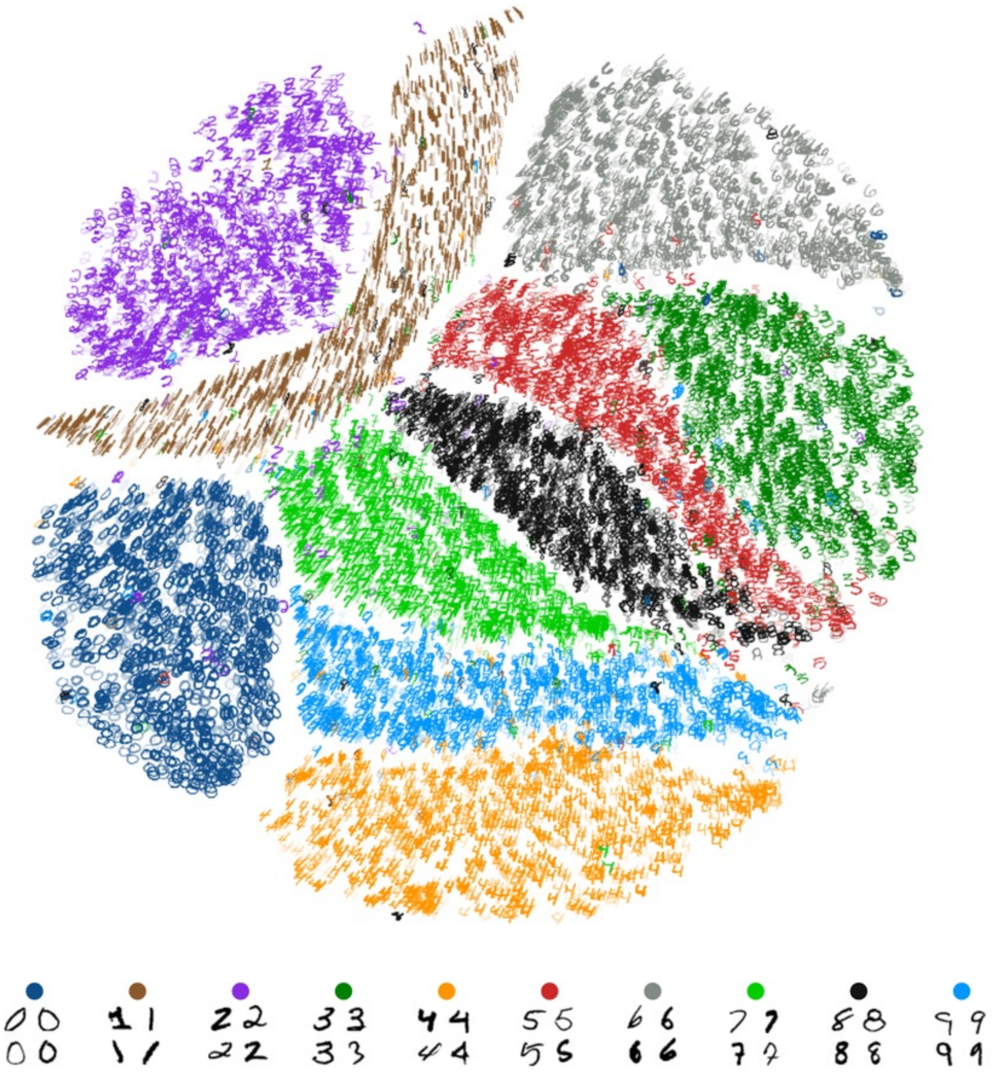
Wikimedia Commons



COSMIC DAWN CENTER
DAWN

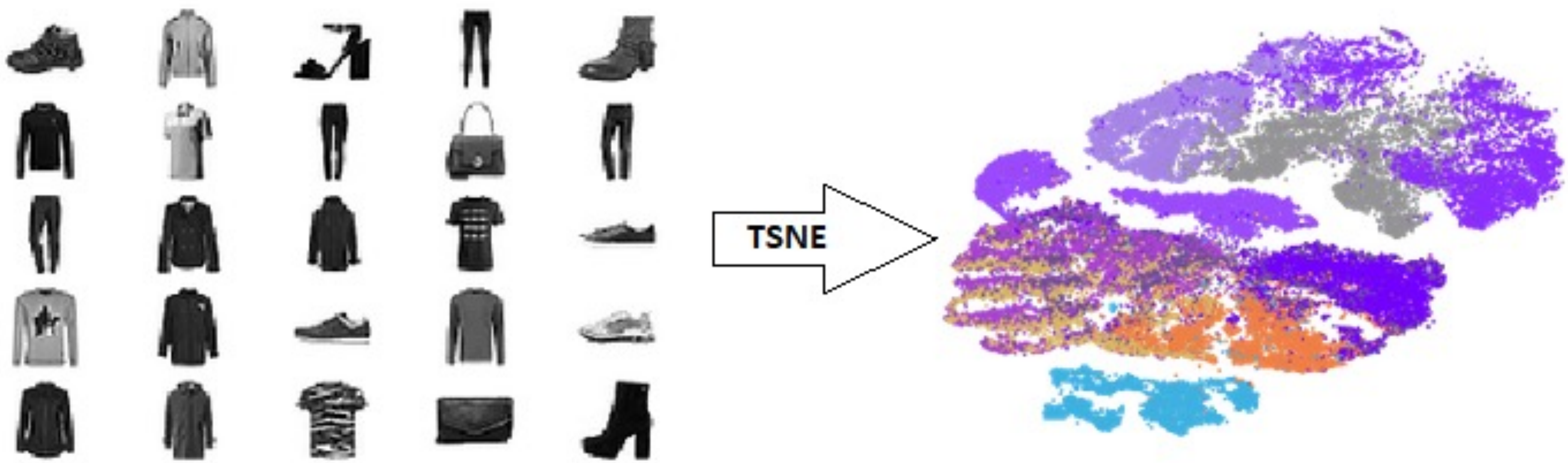
Group “similar” things together

Pezzotti 2019



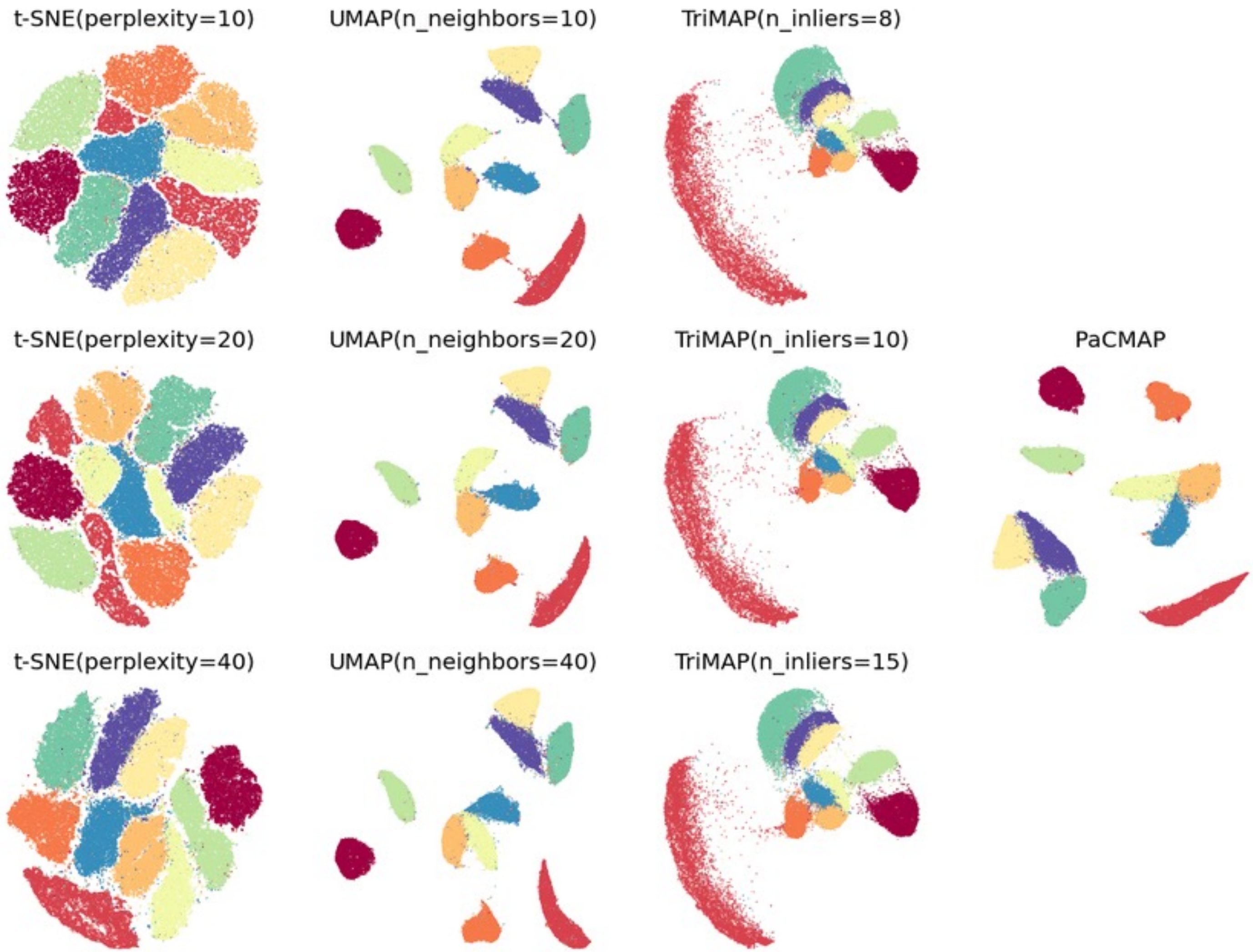
Group “similar” things together

“Fashion MNIST” datasets, t-SNE



Group "similar" things together

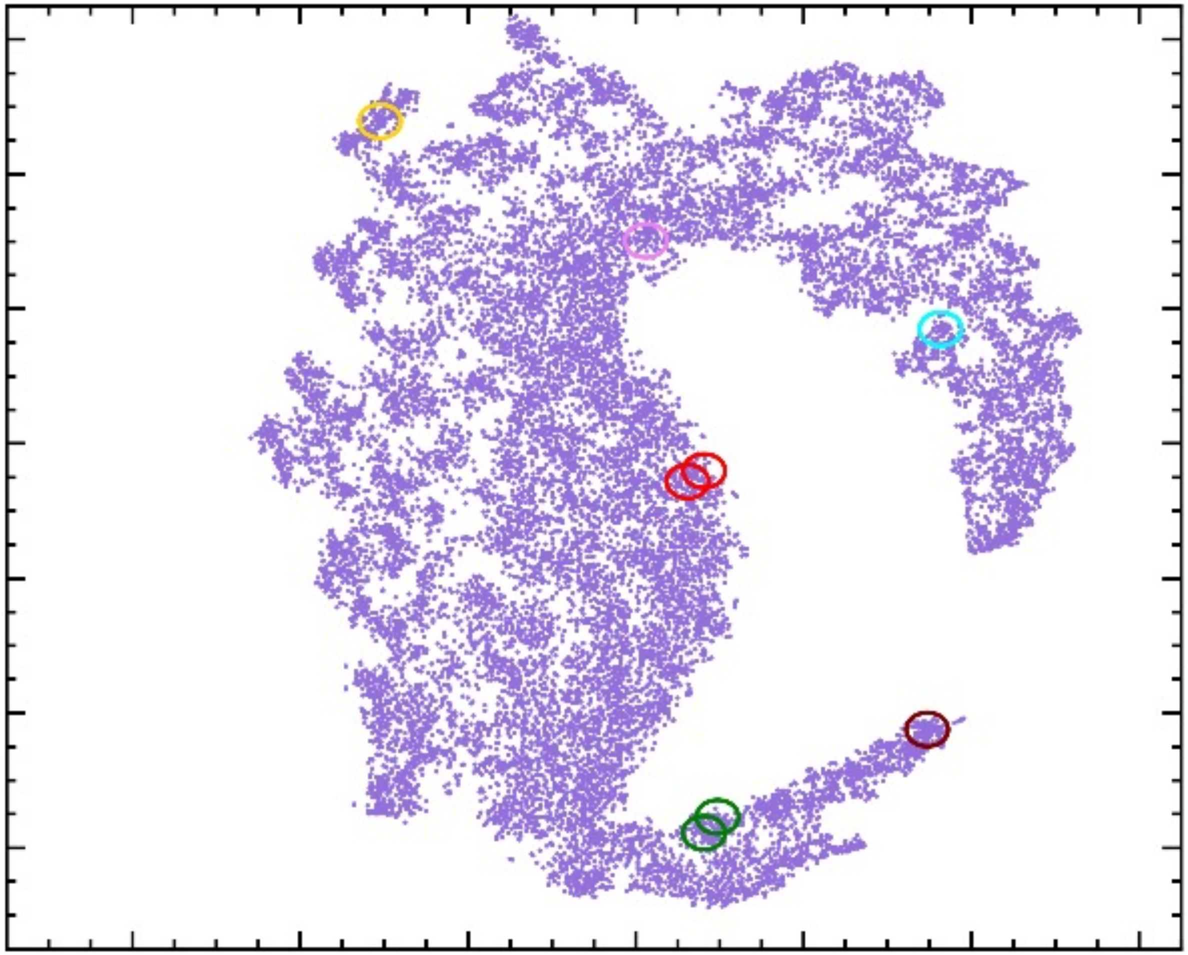
Wang et al. 2020



COSMIC DAWN CENTER
DAWN

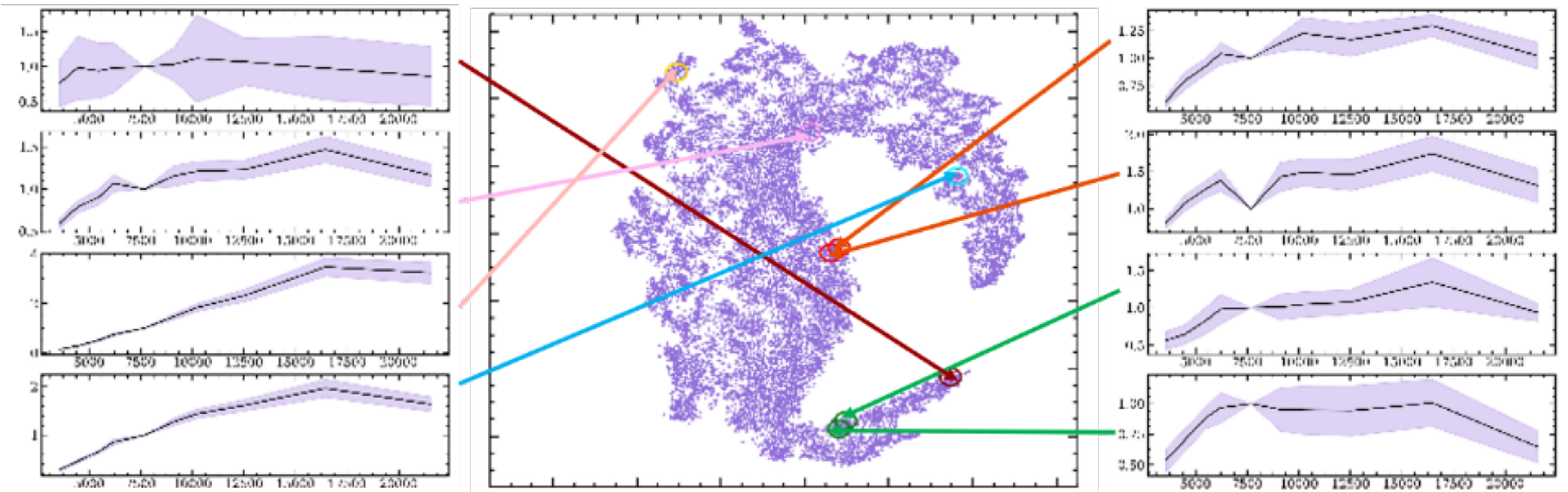
Approach 2: First, make a t-SNE map reducing to two dimensions

COSMOS2015 catalog, objects at z≈1



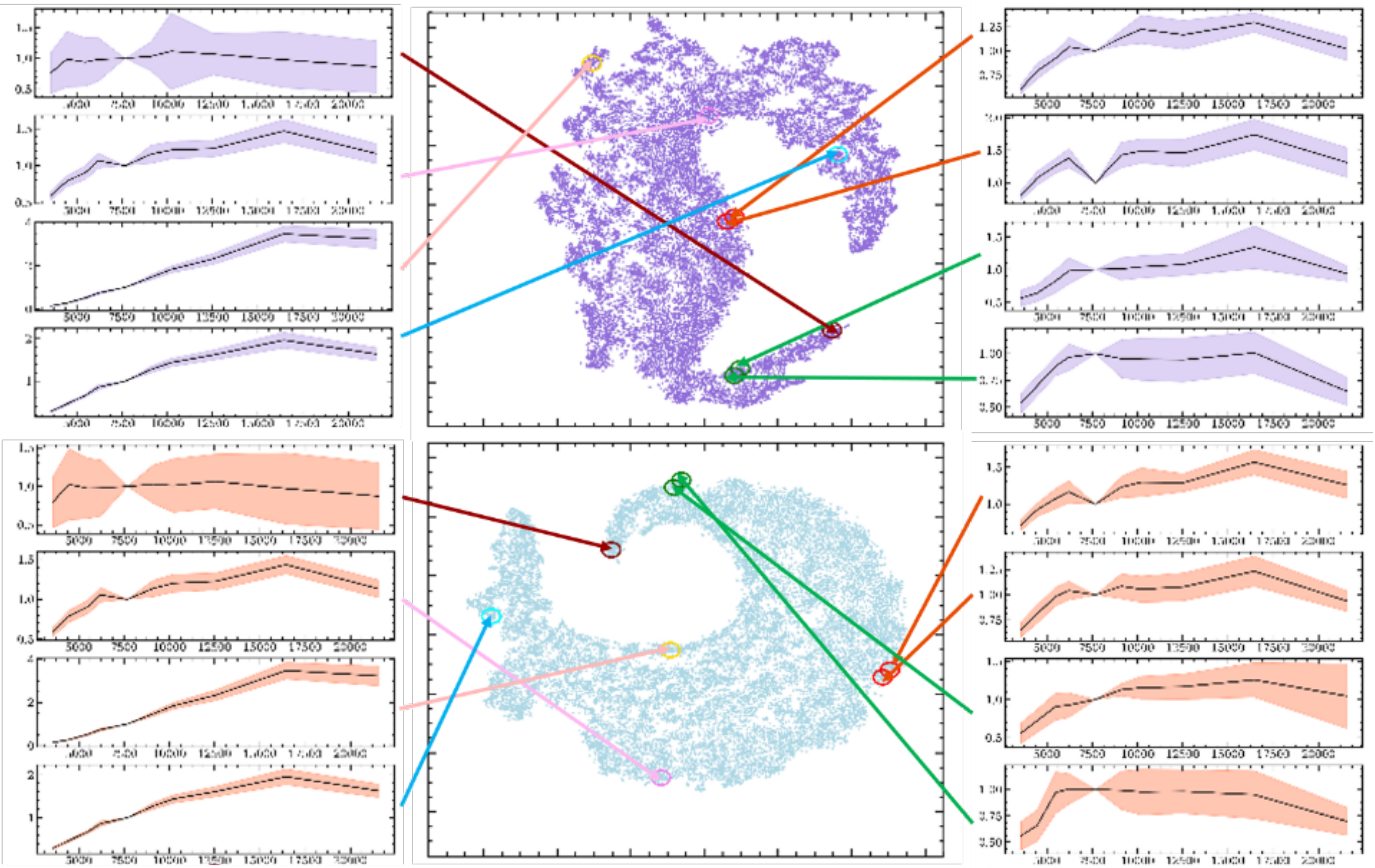
Approach 2: Similar galaxies are nearby

COSMOS2015 catalog, objects at z≈1



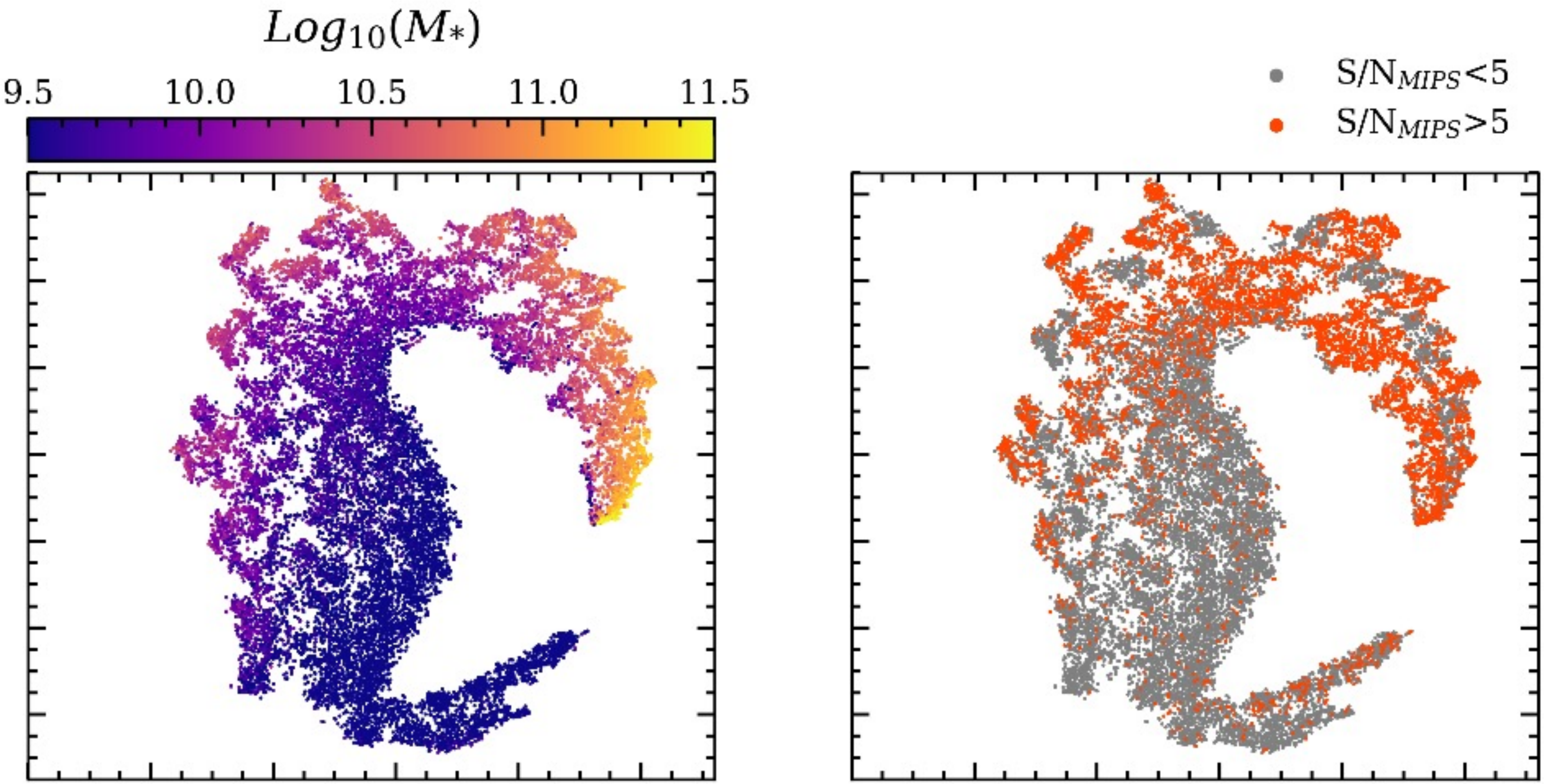
Approach 2: Similar galaxies are nearby

WARNING: positions are neither fixed nor meaningful. Topology is meaningful.



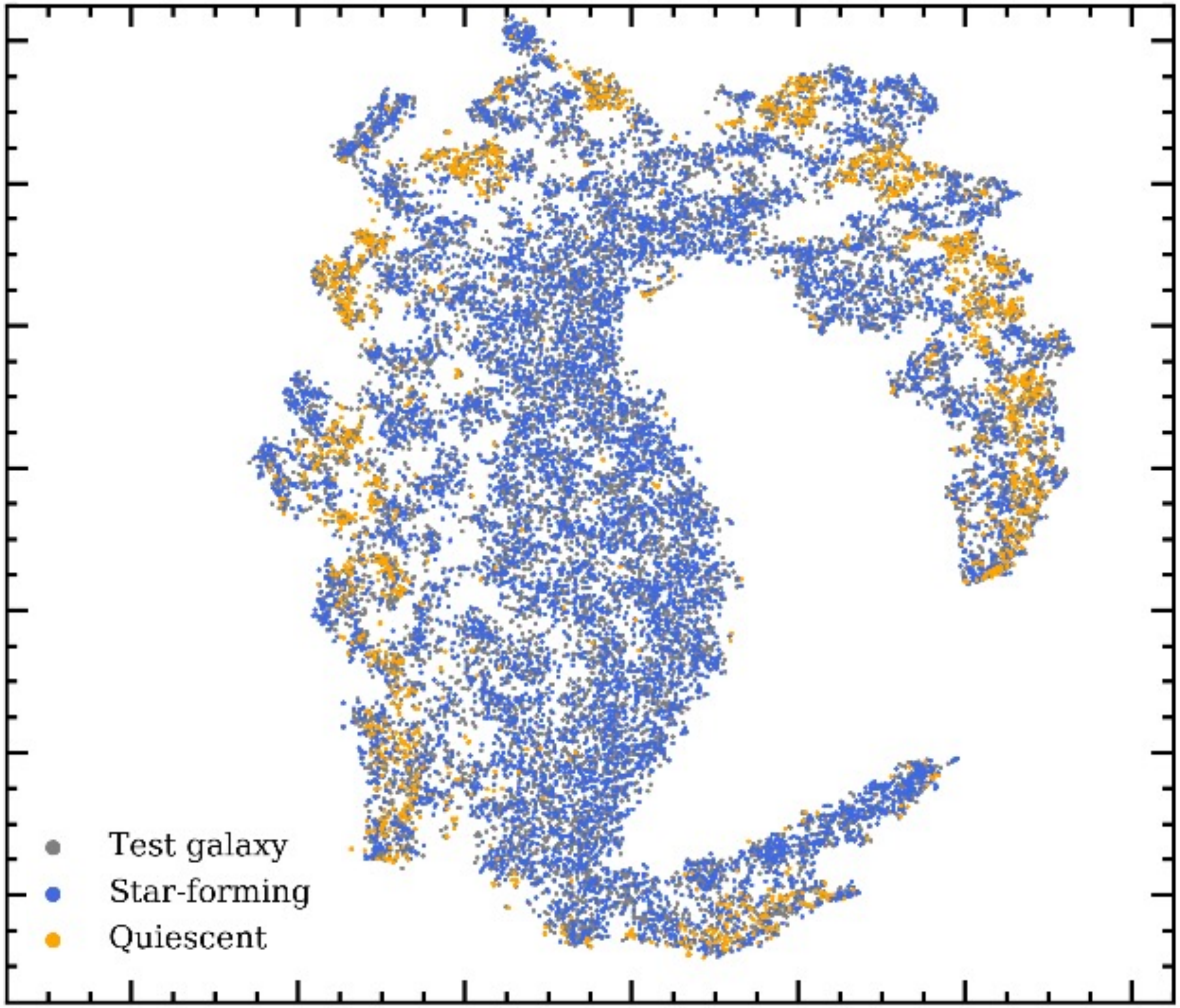
Approach 2: Arranging by photometry also calculates other useful things!

COSMOS2015 catalog, objects at z≈1



Approach 2: Finding "dead" galaxies

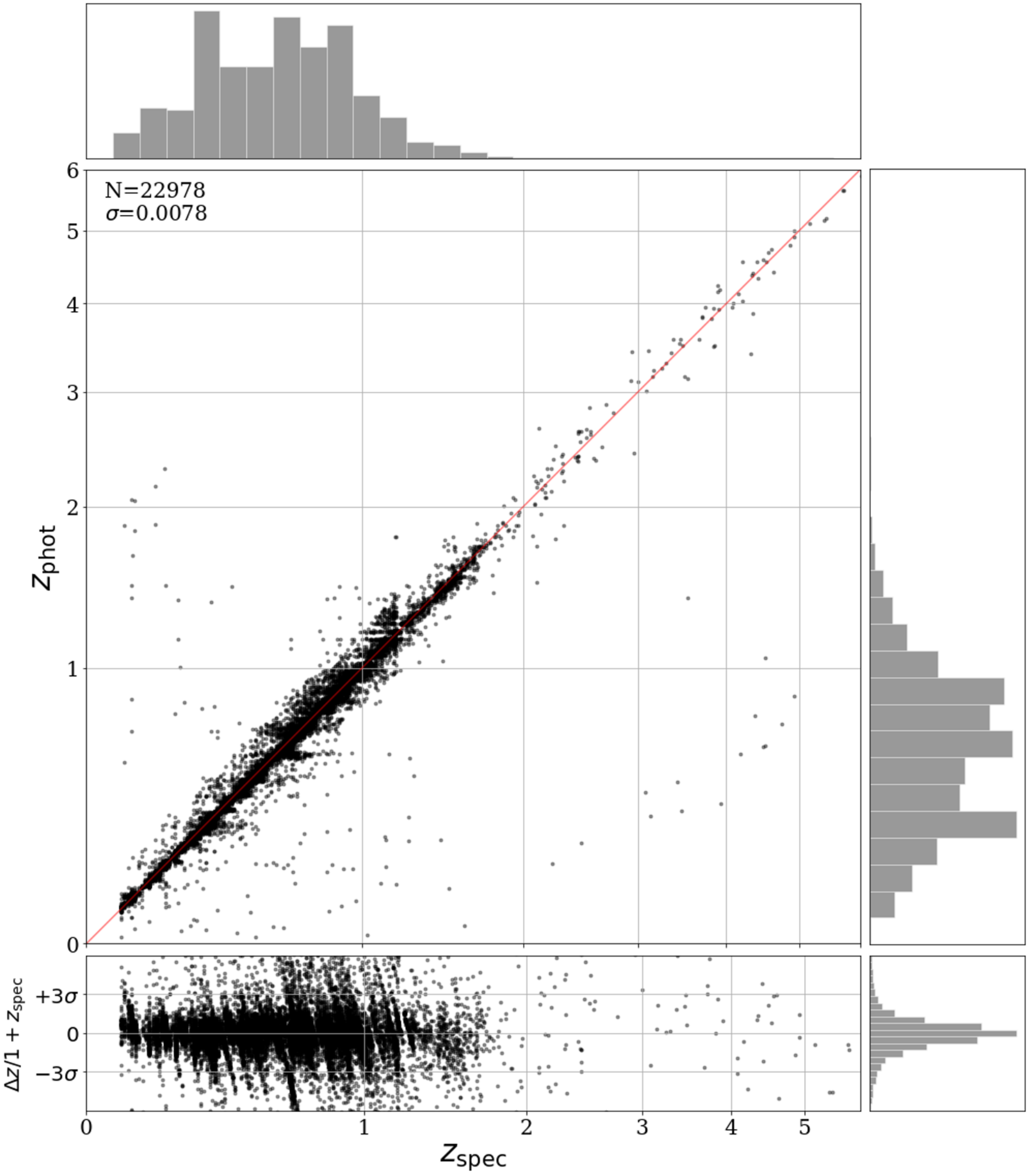
COSMOS2015 catalog, objects at z≈1



COSMIC DAWN CENTER
DAWN

Photometric Redshifts Usually Work…

Hovis-Afflerbach, Steinhardt et al. 2020

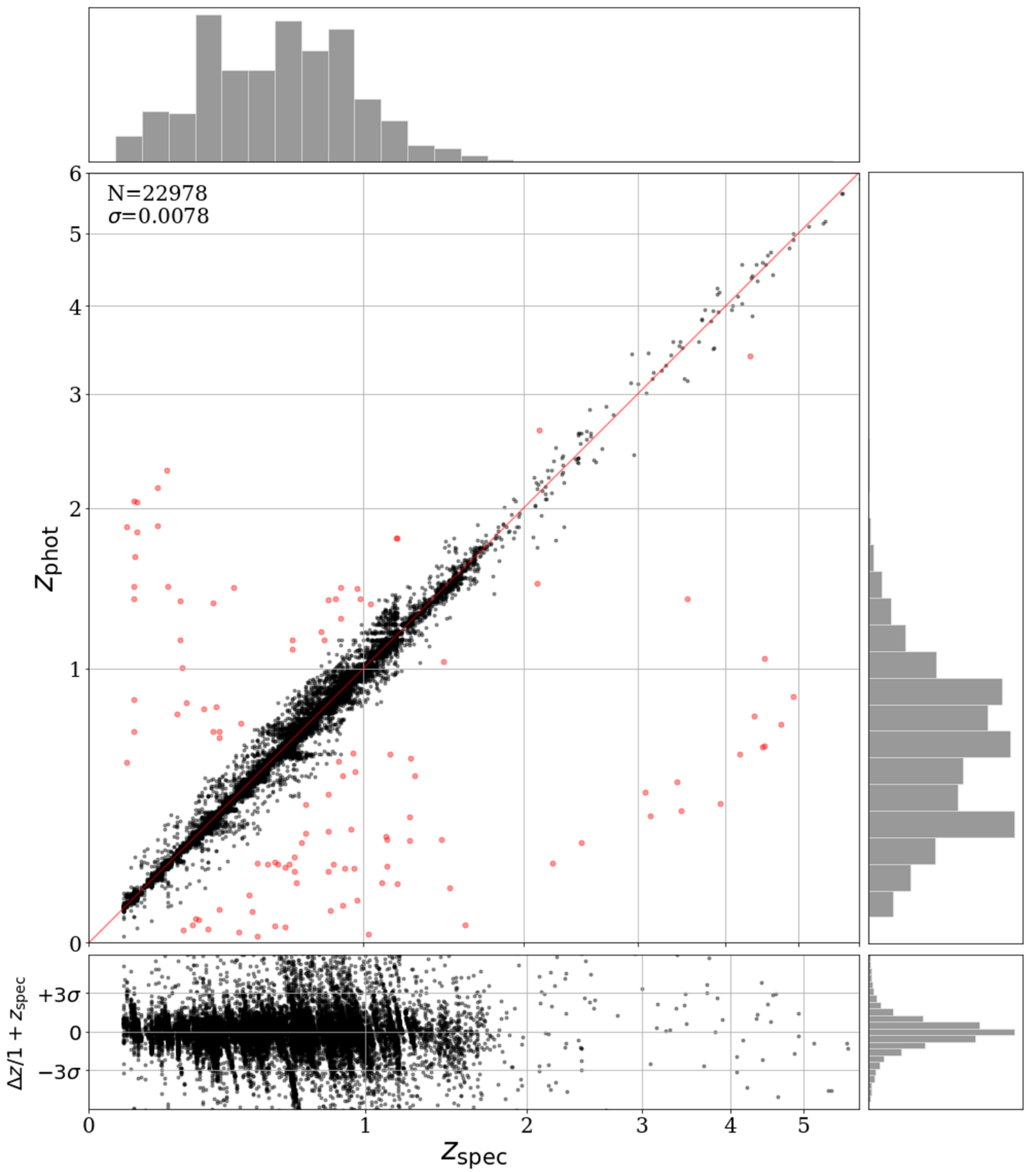


COSMIC DAWN CENTER

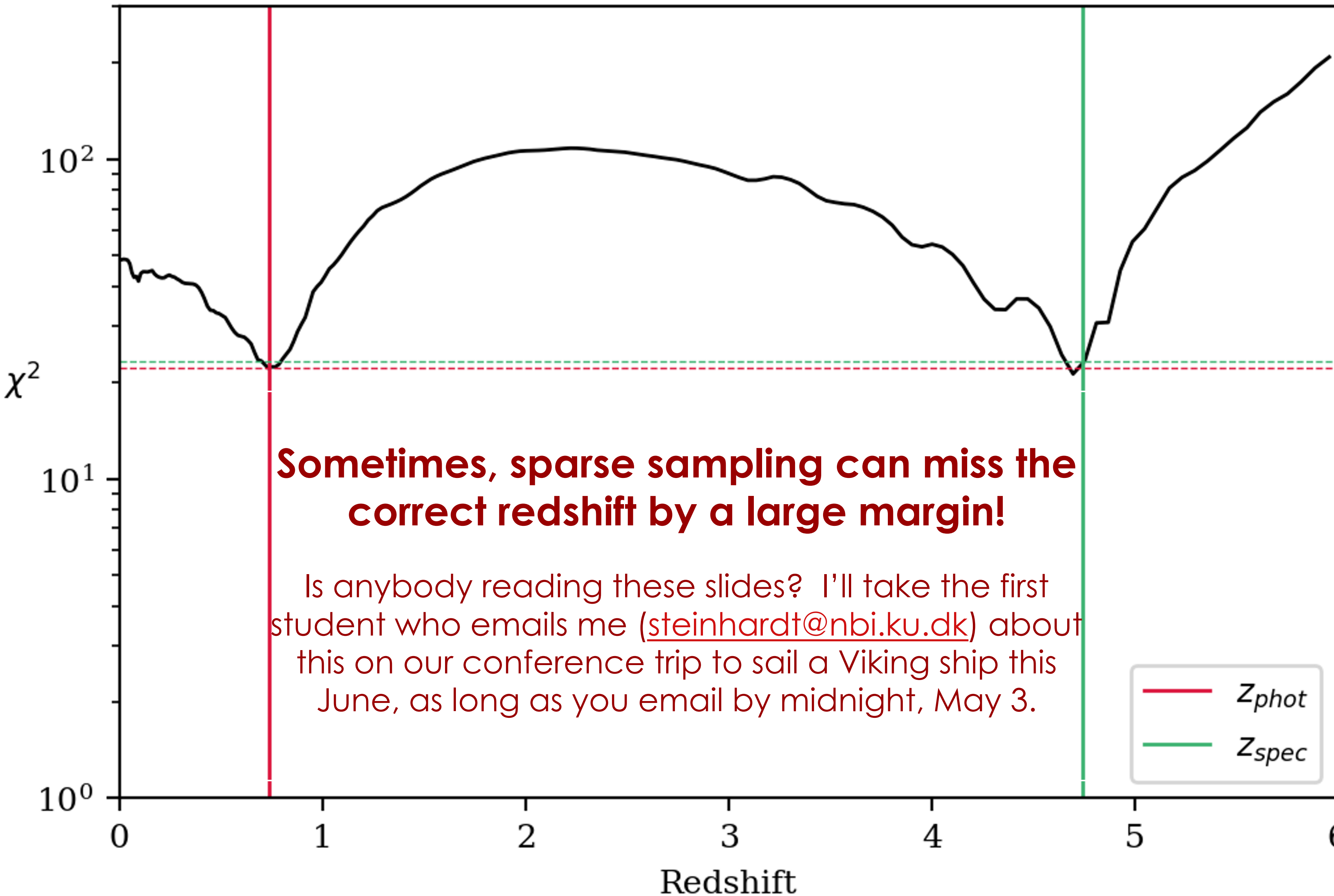
DAWN

Photometric Redshifts Usually Work…

Hovis-Afflerbach, Steinhardt et al. 2020



COSMIC DAWN CENTER
DAWN



Sometimes, sparse sampling can miss the correct redshift by a large margin!

Is anybody reading these slides? I'll take the first student who emails me (steinhardt@nbi.ku.dk) about this on our conference trip to sail a Viking ship this June, as long as you email by midnight, May 3.

…But Not Always!

About 5% of objects fail template fitting[1]

Can we identify these objects without comparing with spectroscopy?

Can we fix these objects and determine the correct properties?

[1]Hildebrandt et al. 2010:
12 photo-z methods tested on objects with 18 bands, between 4.9% and 29% of objects had dz/(1+z) > 0.15