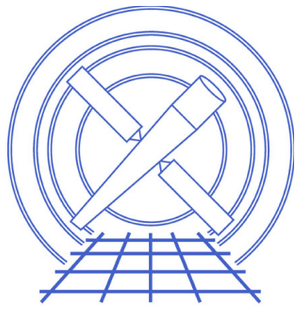
Knowledge Discovery in the era of Rich-Data astronomy: a Chandra perspective

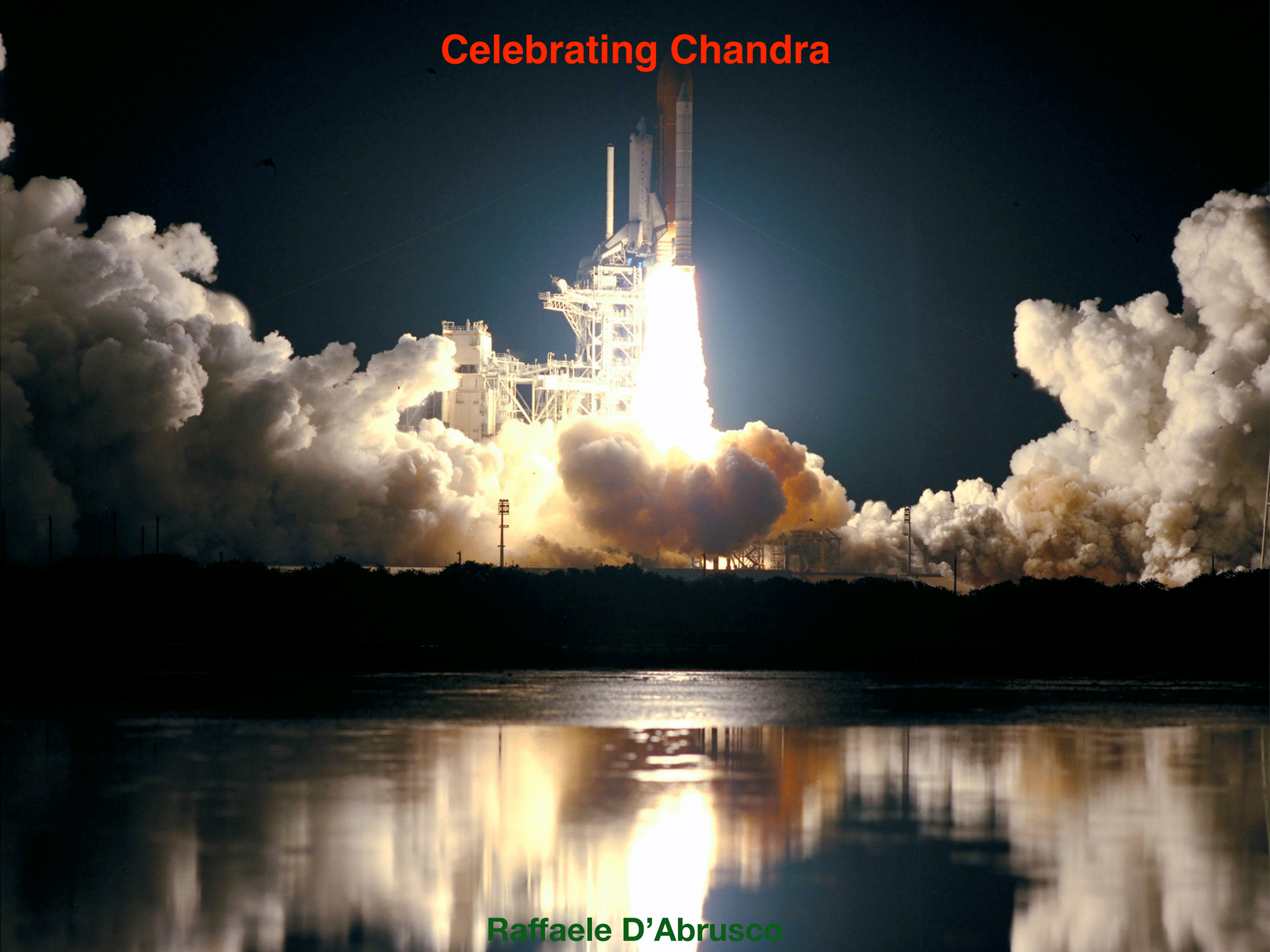
Raffaele D'Abrusco

CXC/SAO



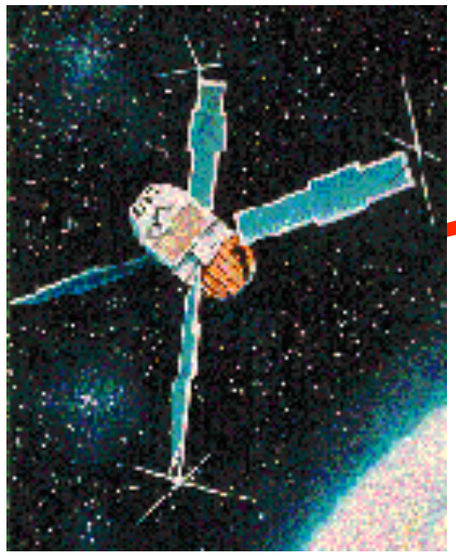
Chandra Science for the Next Decade - Cambridge (MA) - 8/16/16

Celebrating Chandra



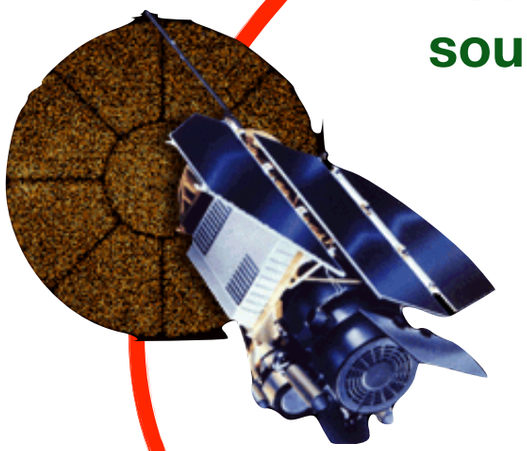
Raffaele D’Abrusco

Giant leaps

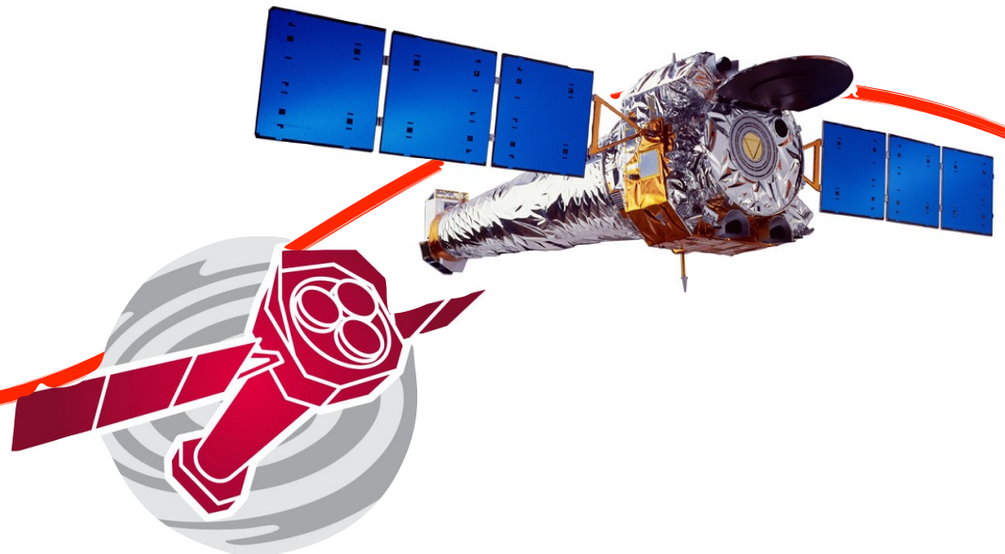


few 100's

few 1000's

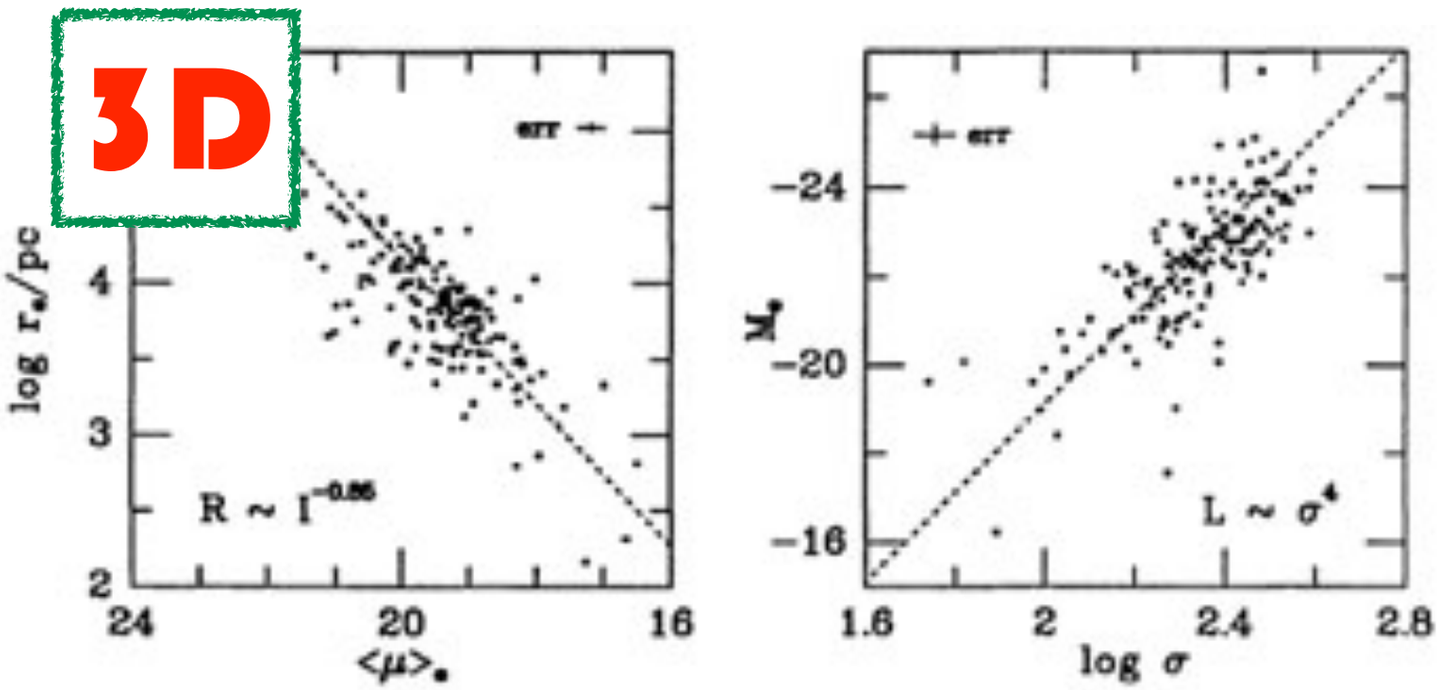


> 100,000 sources



Raffaele D'Abrusco

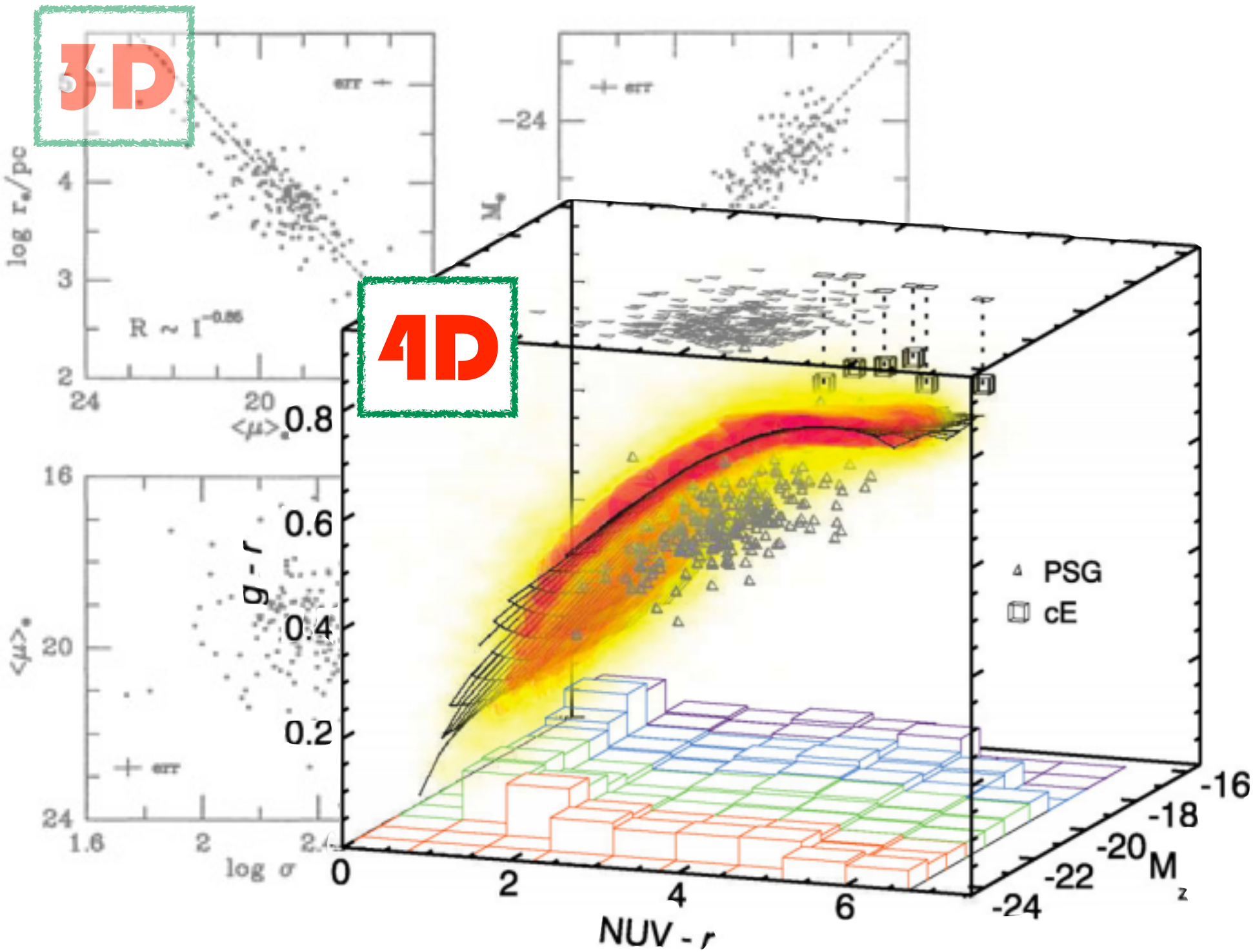
The emergence of multivariate Astronomy



Fundamental plane of elliptical galaxies

(Djorgovski&Davis1987)

The emergence of multivariate Astronomy

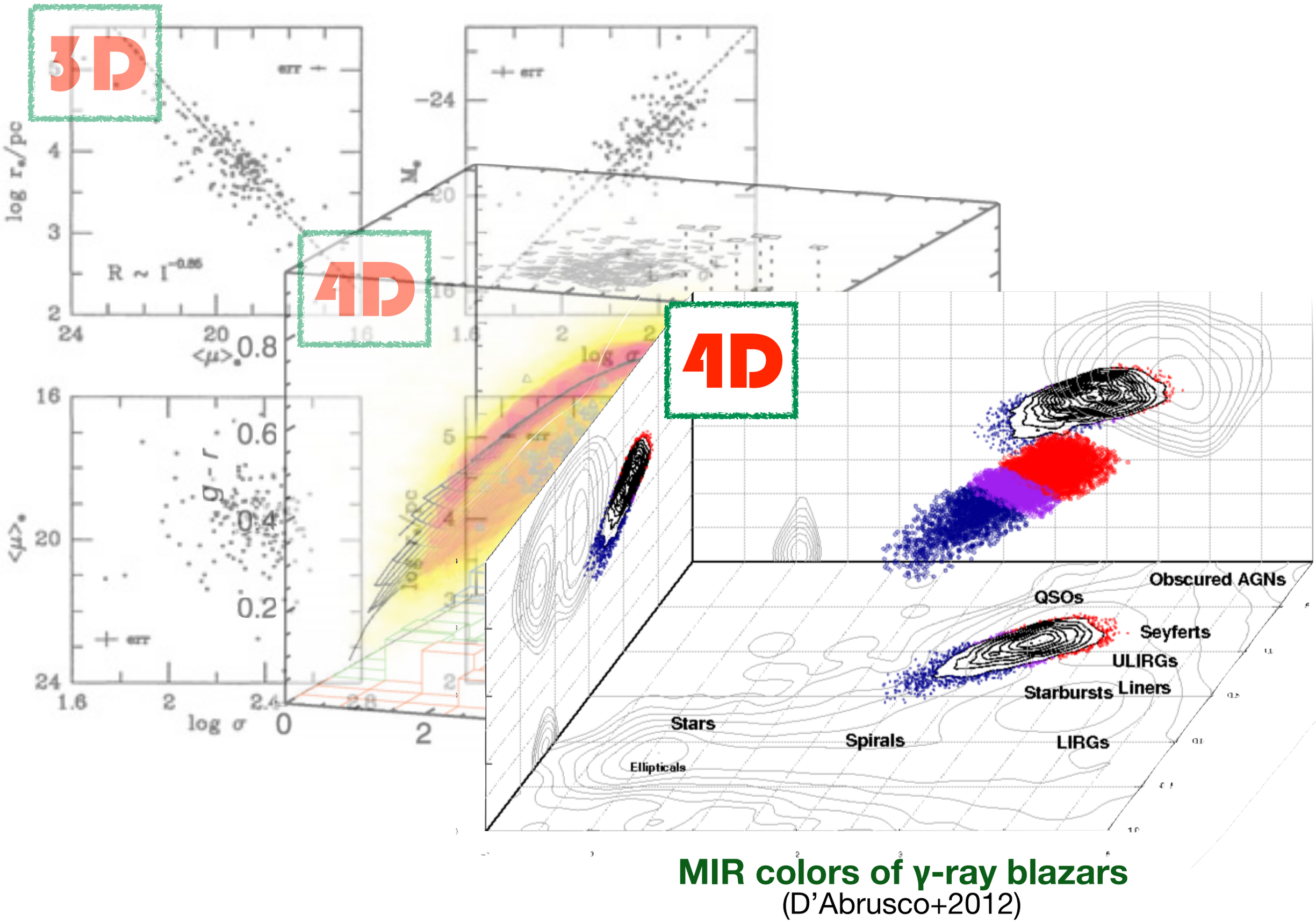


Optical-NUV locus of passive galaxies

(Chilingarian&Zolotukhin2012)

Raffaele D'Abrusco

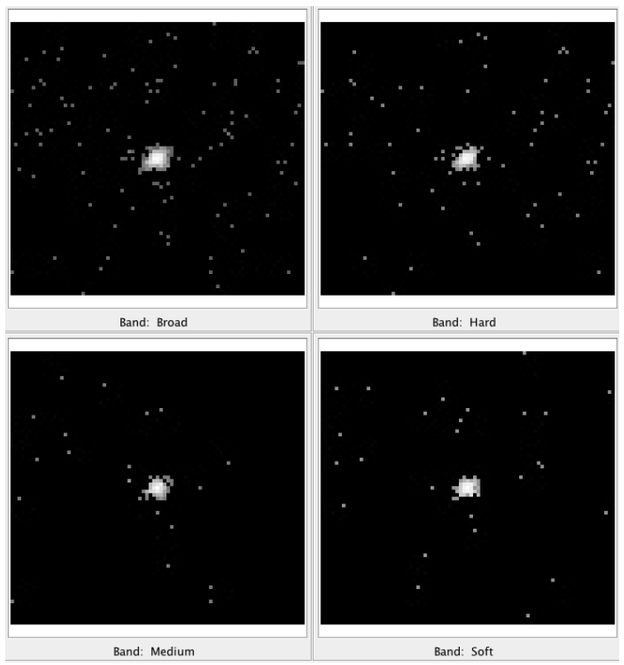
The emergence of multivariate Astronomy

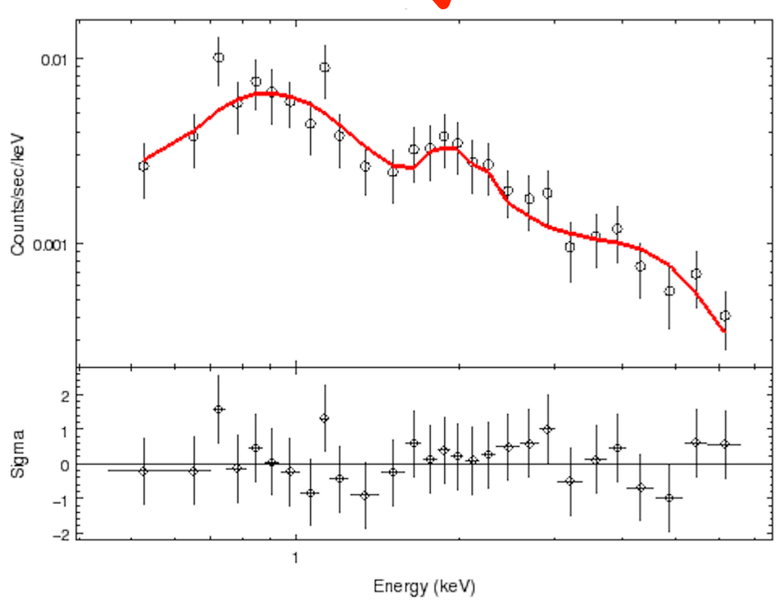


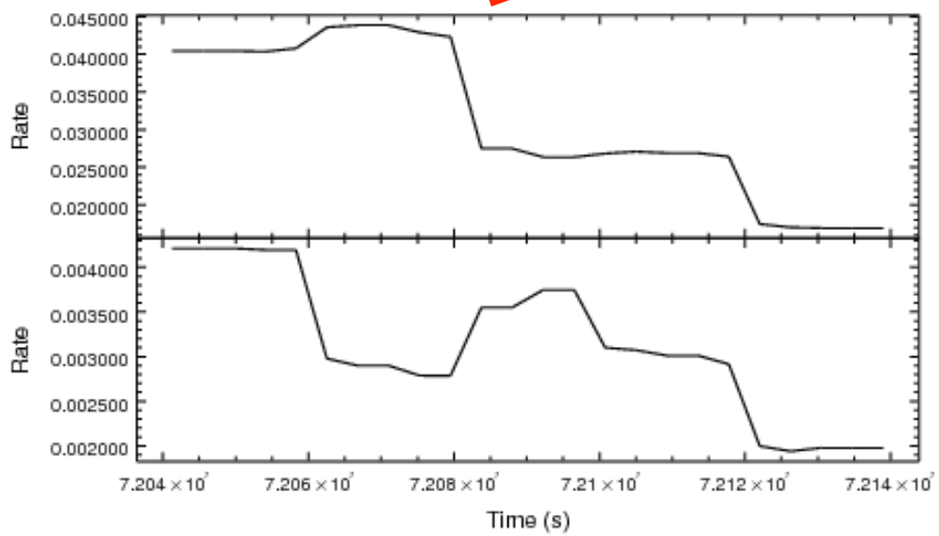
MIR colors of γ-ray blazars
(D'Abrusco+2012)

CHANDRA SOURCE CATALOG

☐	170960	CXO J001300.5-270751	00 13 00.52	-27 07 51.57	0.62953	FALSE	FALSE	26.6051	1.2424E-13	1.19349E-13	1.29181E-1
☑	177343	CXO J001351.5-302500	00 13 51.51	-30 25 00.26	0.320671	FALSE	FALSE	25.6305	2.11018E-13	2.04545E-13	2.17427E-1
☐	82564	CXO J000136.1+130639	00 01 36.10	+13 06 39.50	0.309182	FALSE	FALSE	24.83	1.42236E-13	1.37323E-13	1.47199E-1



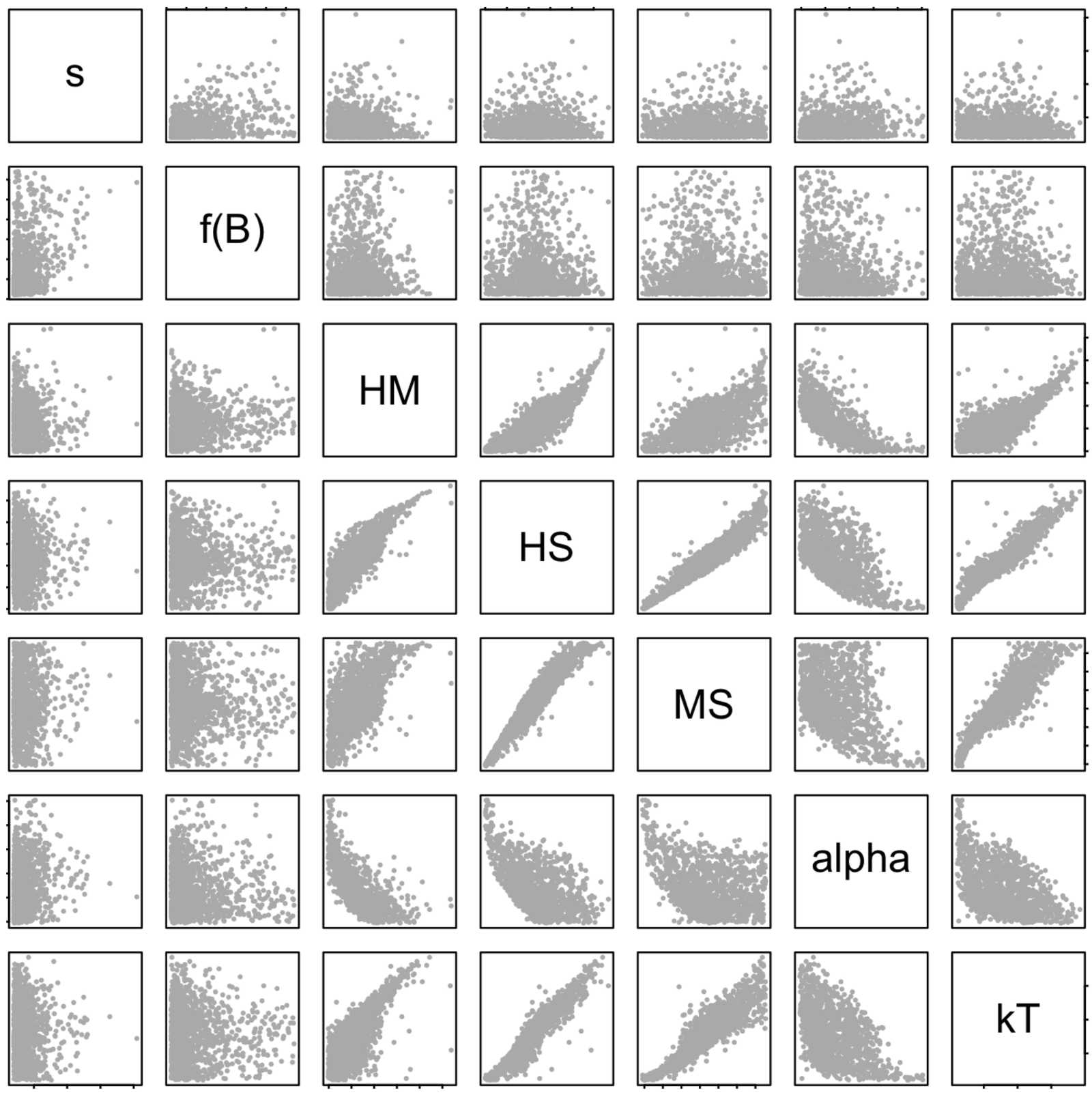




See Ian Evans' poster on CSC status

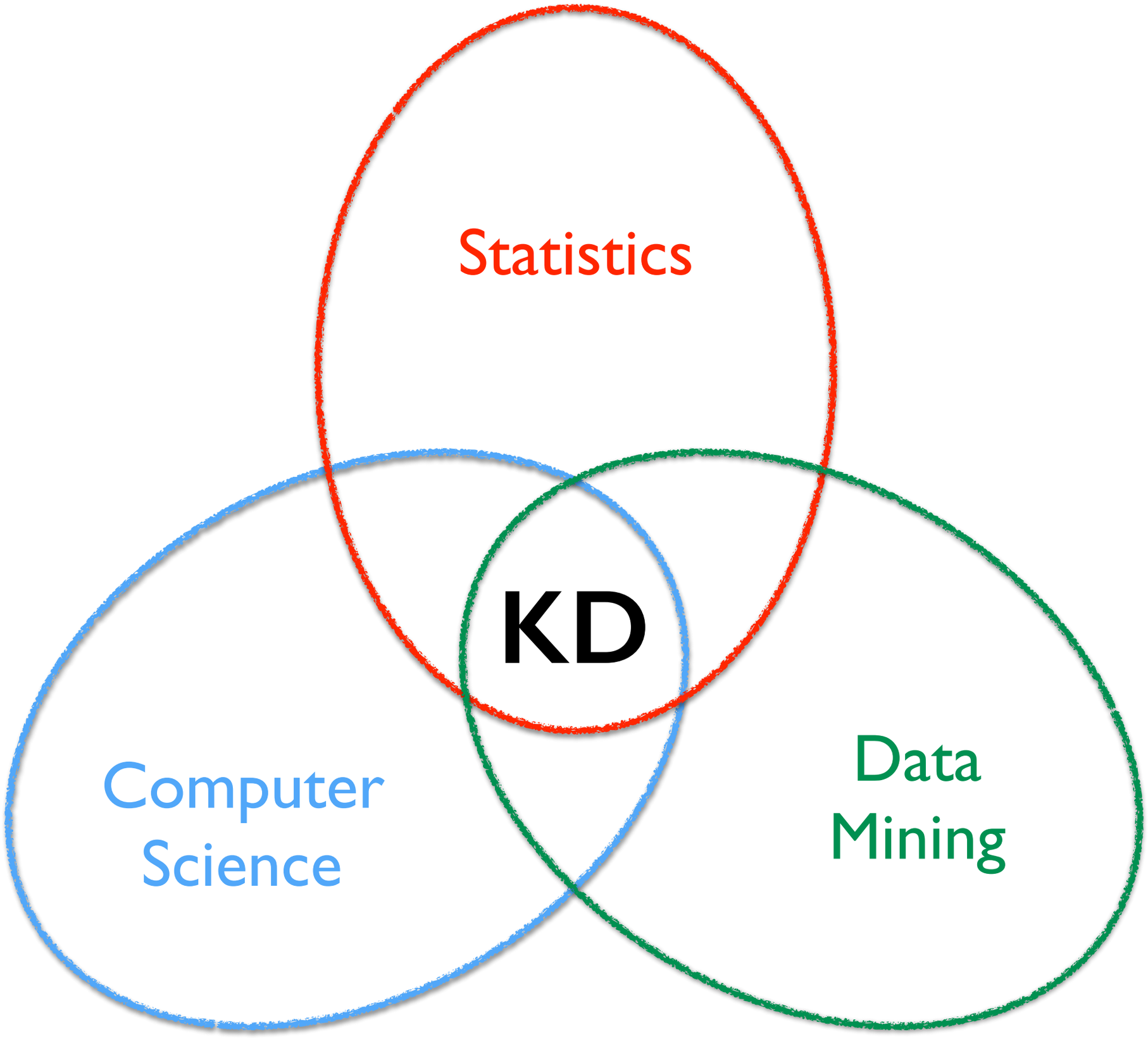
Raffaele D'Abrusco

Exploring High-Dimensional Data



Raffaele D'Abrusco

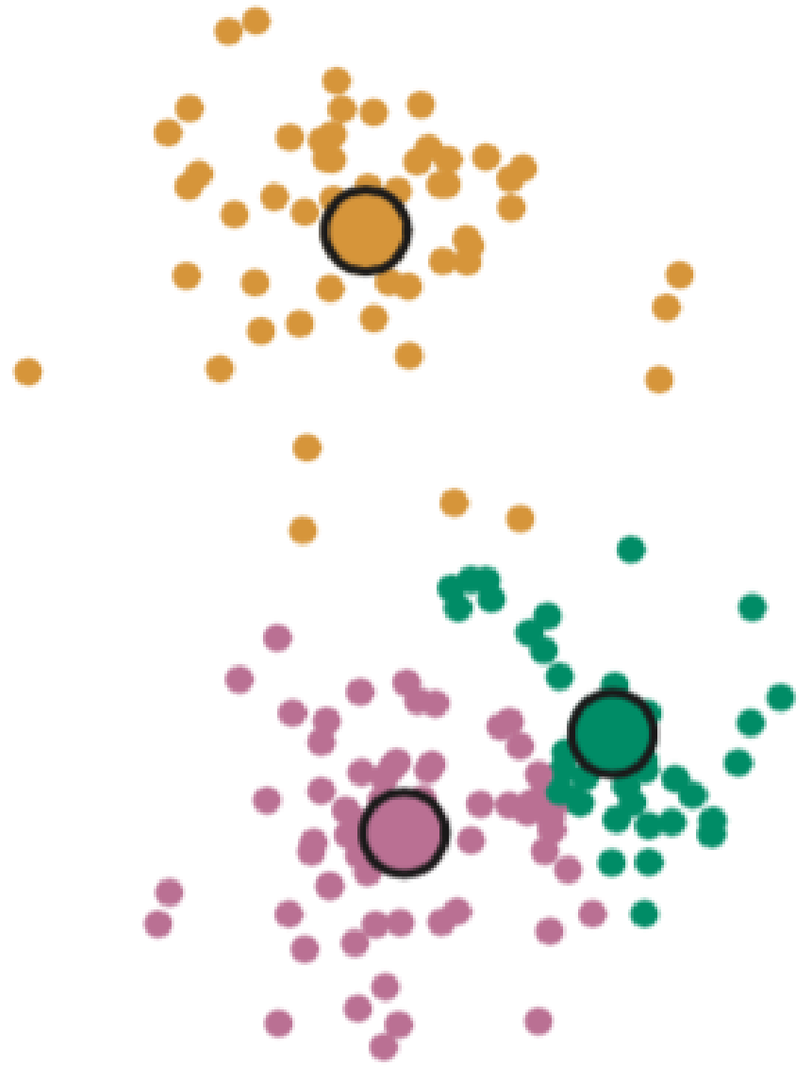
A modern Toolbox: Knowledge Discovery

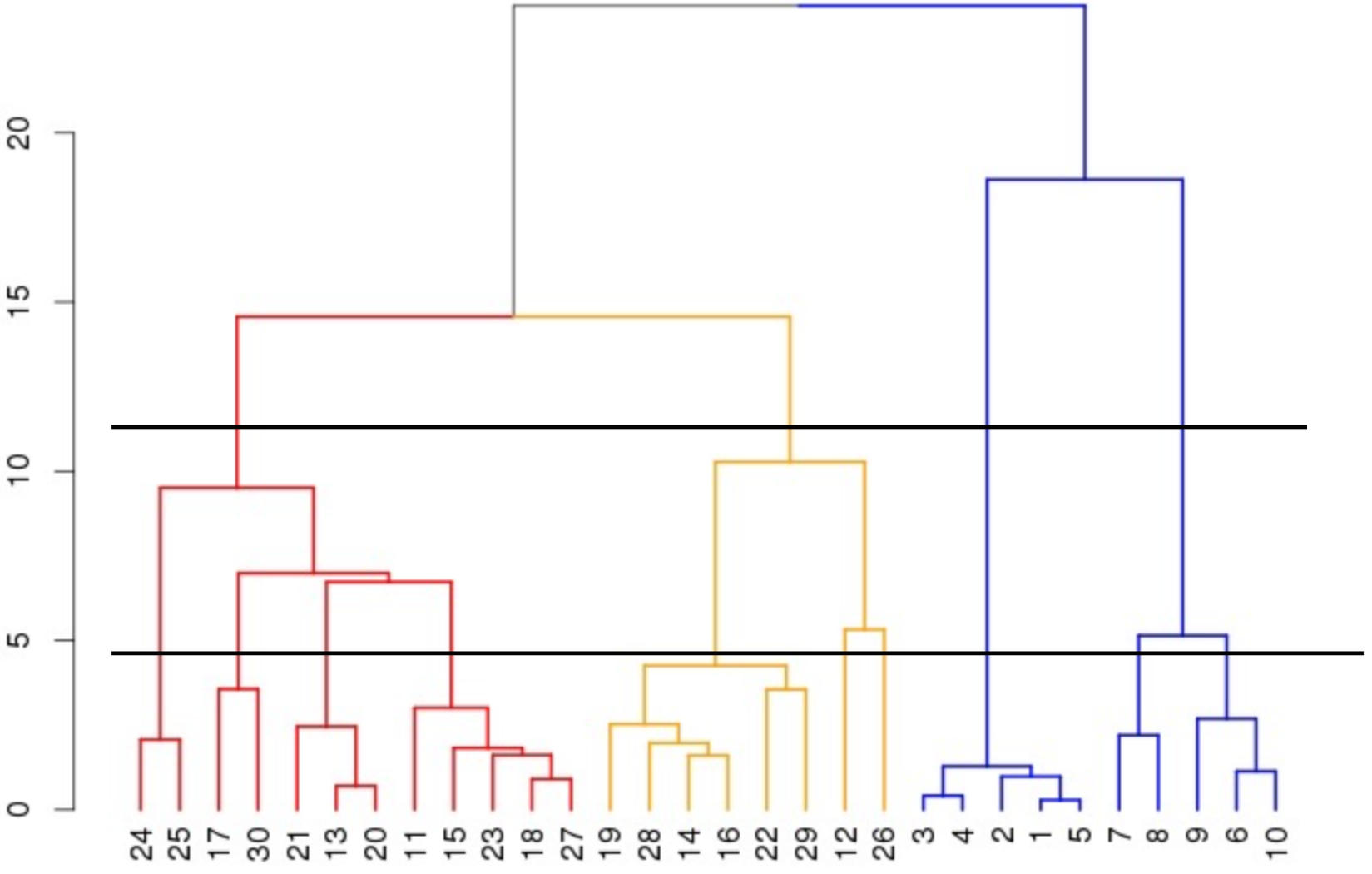


Raffaele D'Abrusco

Learning about the data structure

Finding spontaneous aggregations of source in the *feature* space, the **clusters**





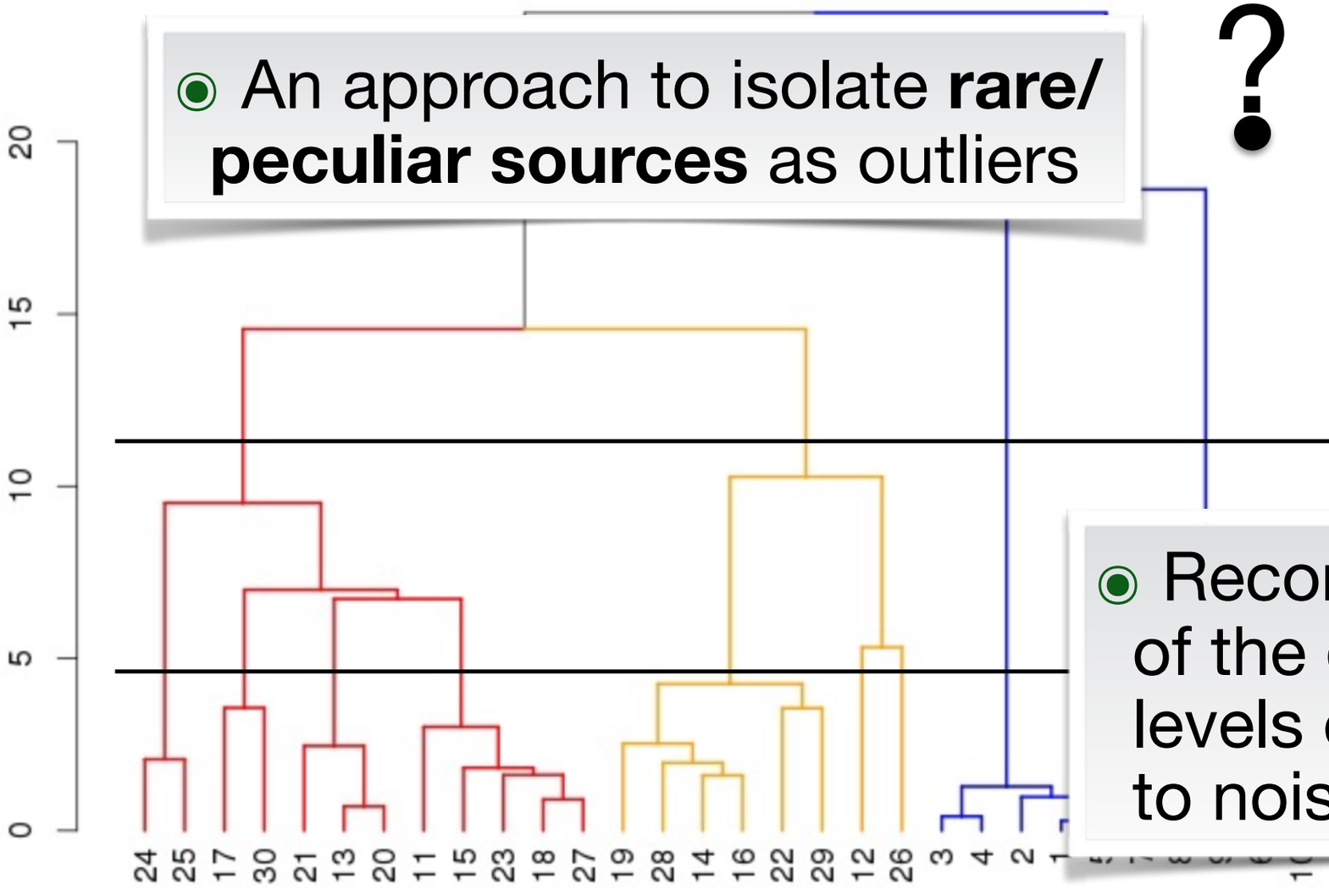
Raffaele D’Abrusco

Learning about the data structure

Finding spontaneous aggregations of source in the *feature* space, the **clusters**

- An approach to isolate **rare/ peculiar sources** as outliers

?



- Reconstructing the **structure** of the dataset at different levels of granularity (i.e. signal to noise)

Raffaele D'Abrusco

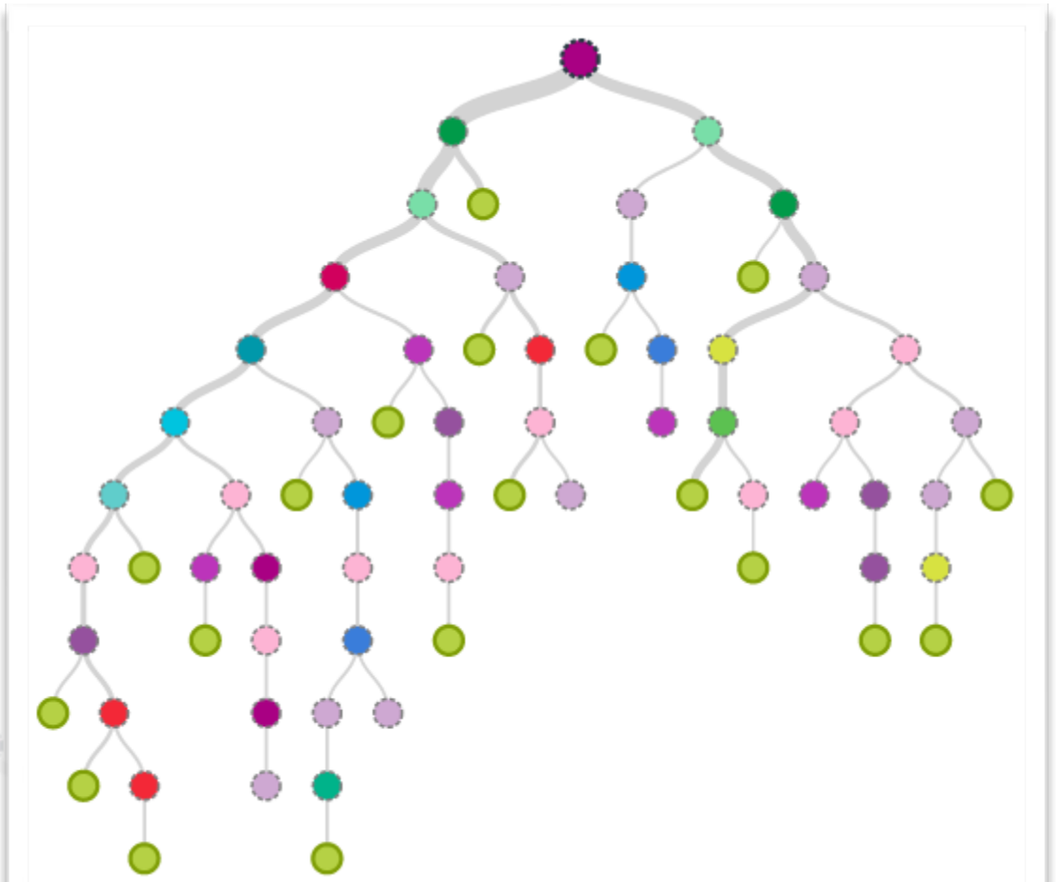
Clustering

Classification

Raffaele D’Abrusco

Searching for known sources

- **Ensembles of classifiers** and **priors** improves accuracy with negligible decrease in flexibility



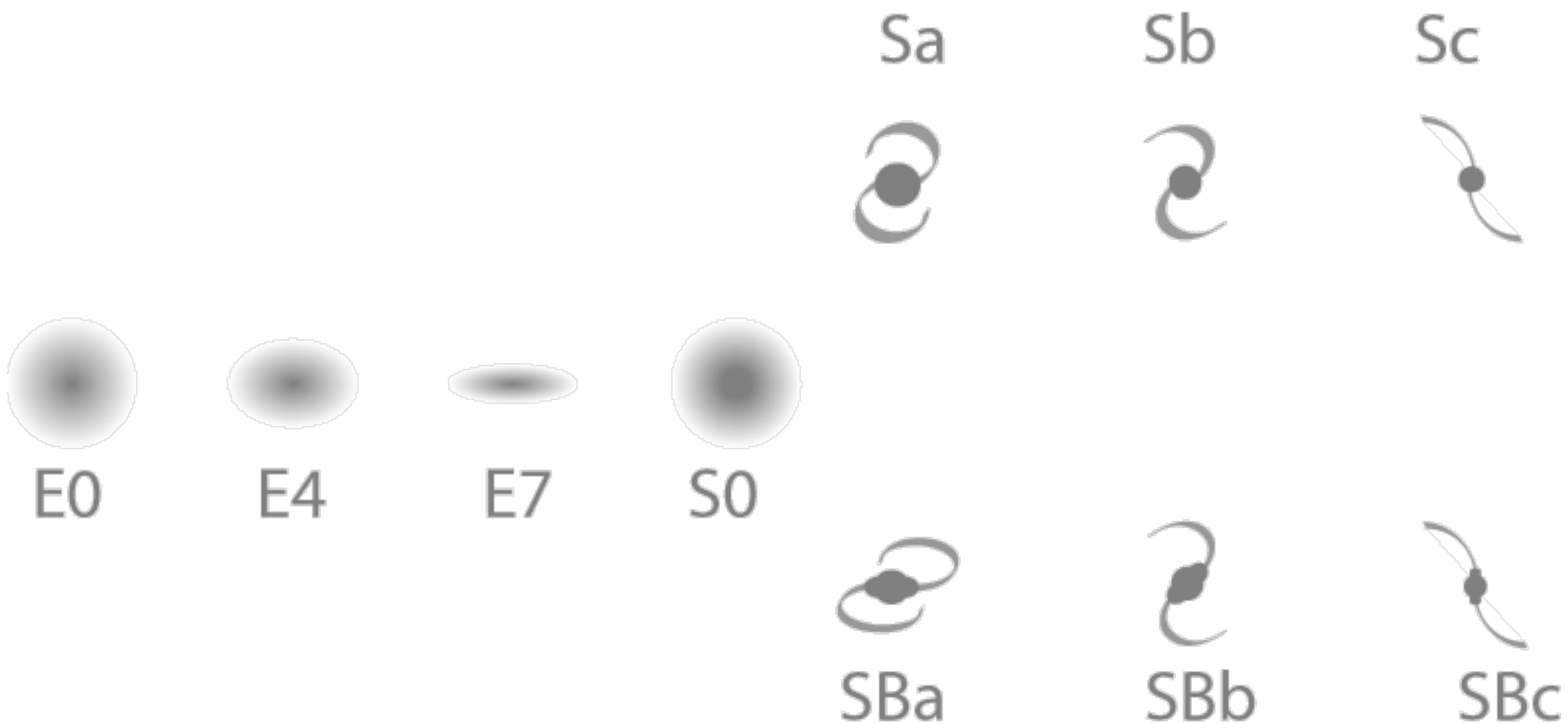
- **Probabilistic classification** can reflect degenerate cases or transitional classes of sources

Raffaele D'Abrusco

Raffaele D’Abrusco

A Taxonomy of the X-ray Sky

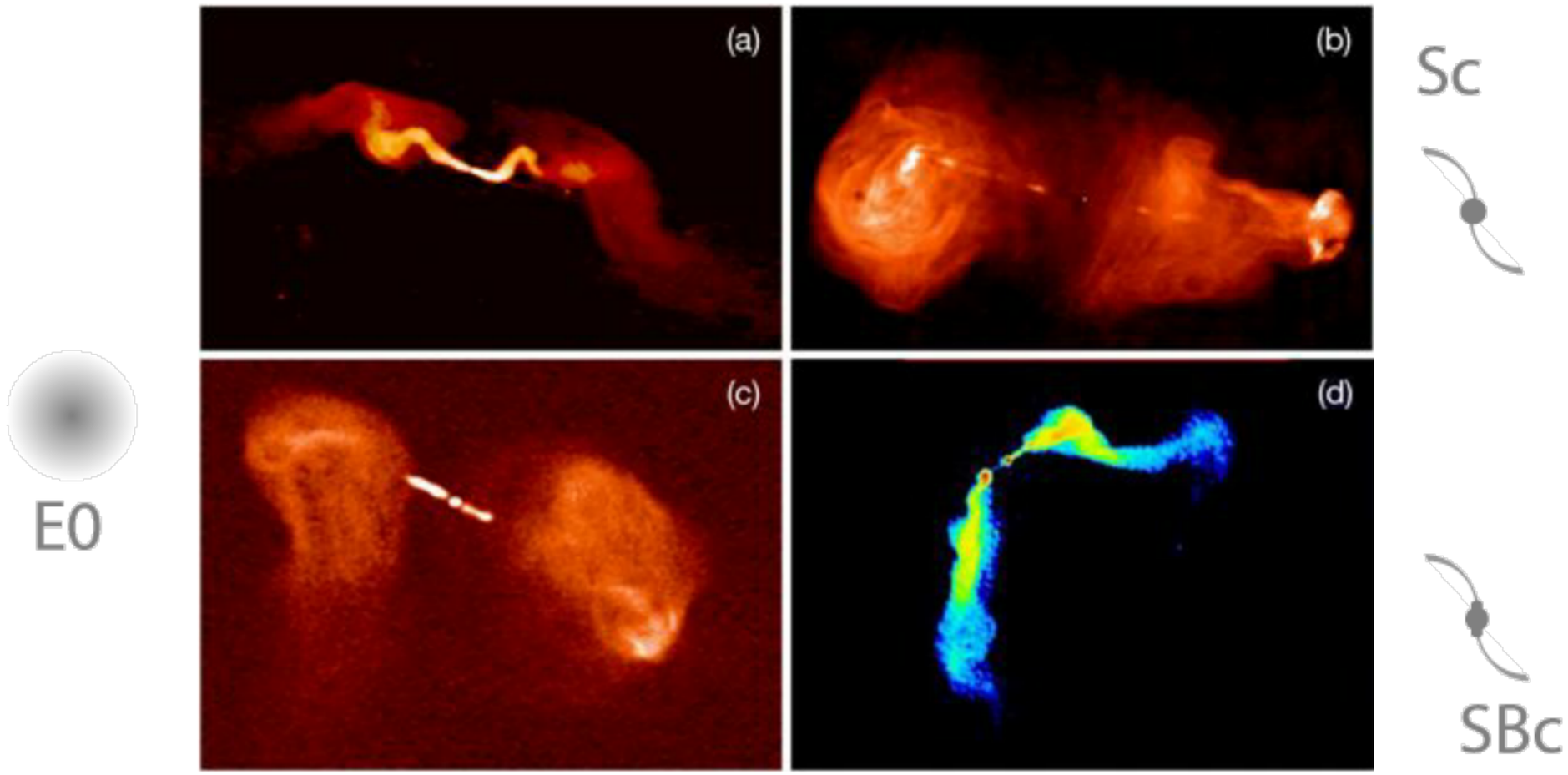
- Opportunity to develop classification framework not based on multi-wavelength knowledge
- Improvement over low-dimensional classifications schemes for specific classes of sources



Raffaele D'Abrusco

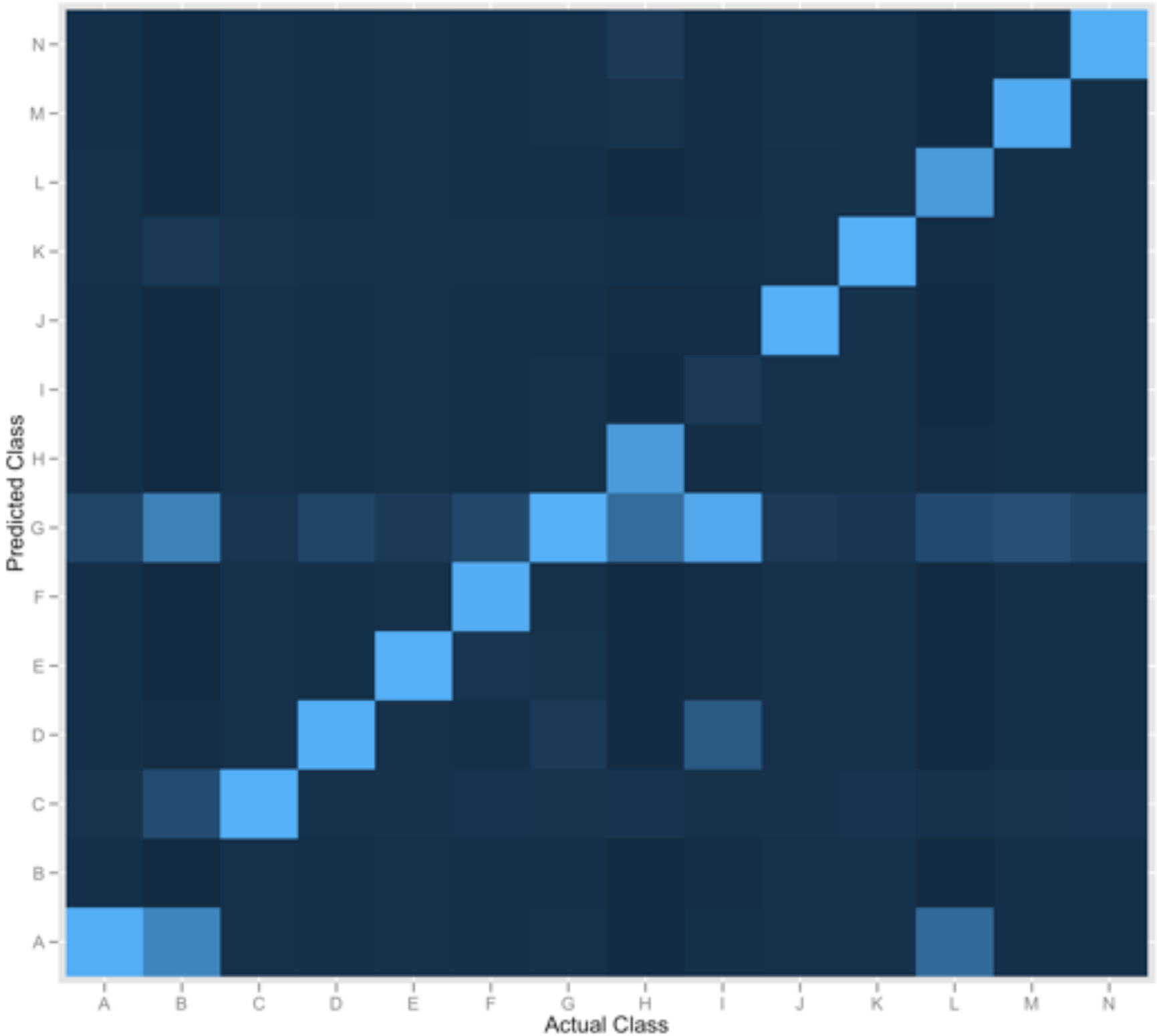
A Taxonomy of the X-ray Sky

- Opportunity to develop classification framework not based on multi-wavelength knowledge
- Improvement over low-dimensional classifications schemes for specific classes of sources



(Banfield+2015)

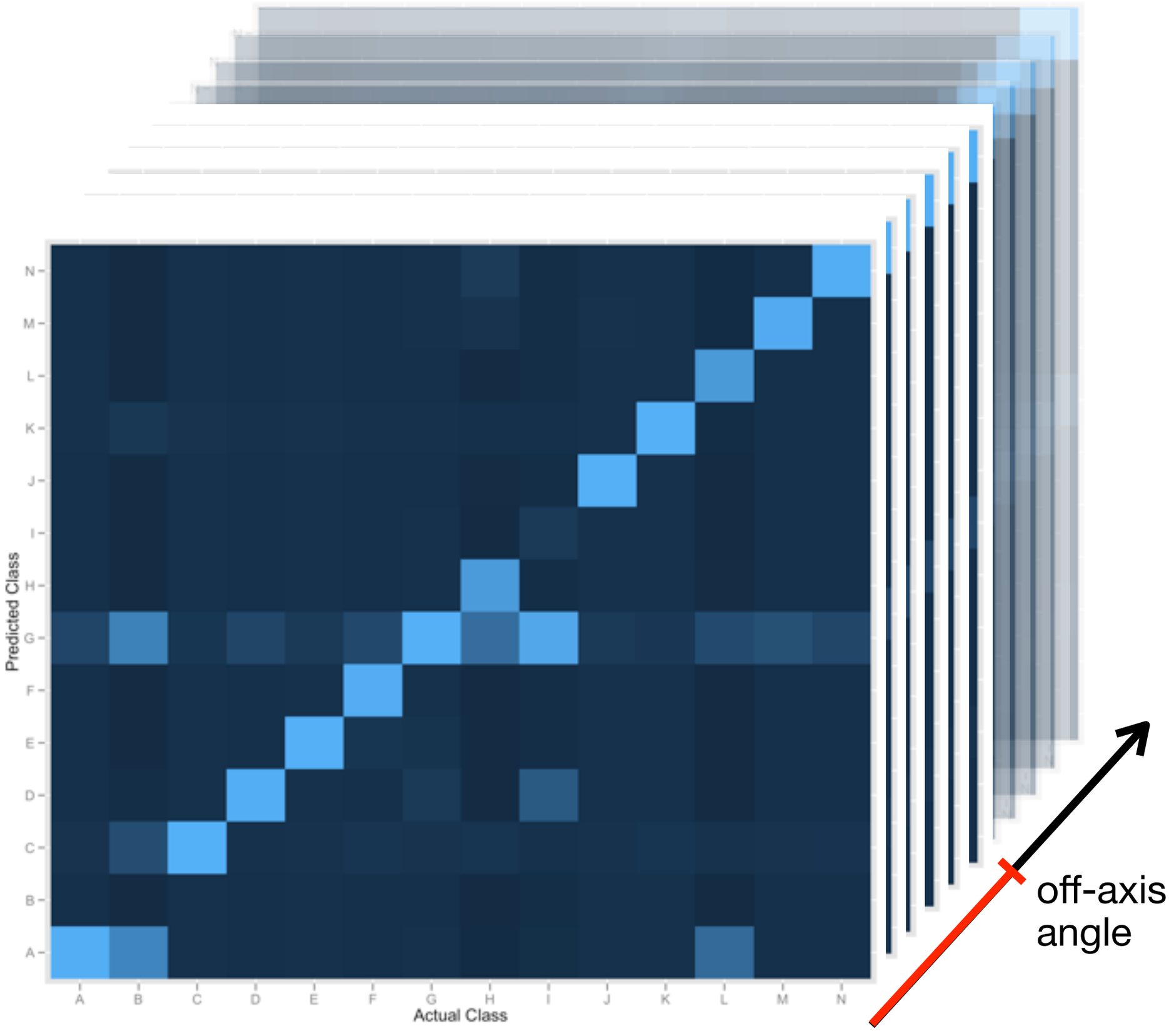
Classification: dealing with instrumental parameters



- number of counts
- off-axis angle
- PSF size, shape
- detection likelihood

Raffaele D’Abrusco

Classification: dealing with instrumental parameters



- number of counts
- off-axis angle
- PSF size, shape
- detection likelihood
- environment type

Raffaele D'Abrusco

Riding the Wave

A KD-powered exploration of the *Chandra* Source Catalog would:

- Maximize the science return of *Chandra* and its legacy value
- Facilitate multi-wavelength research
- **Demonstrate the feasibility of a data-driven** approach to the **exploration** of the X-ray sky in view of future X-ray missions

Raffaele D'Abrusco