

# Evaluando la calidad de los datos de DESI

John Suárez-Pérez  
Jaime Forero-Romero

Universidad de Los Andes  
Astroandes

Febrero de 2021



# Dark Energy Spectroscopic Instrument (DESI).



Fotografía del Observatorio Nacional Kitt Peak. Imagen tomada de DESI HighSchool.

- 70 Instituciones a nivel internacional.
- Laboratorio Lawrence Berkeley (California).
- Observatorio Nacional Kitt Peak, Arizona.
- Inicio su construcción en 2015.
- Inició a tomar datos a finales del 2019.
- Acotar los parámetros cosmológicos que describen la historia de expansión del Universo.

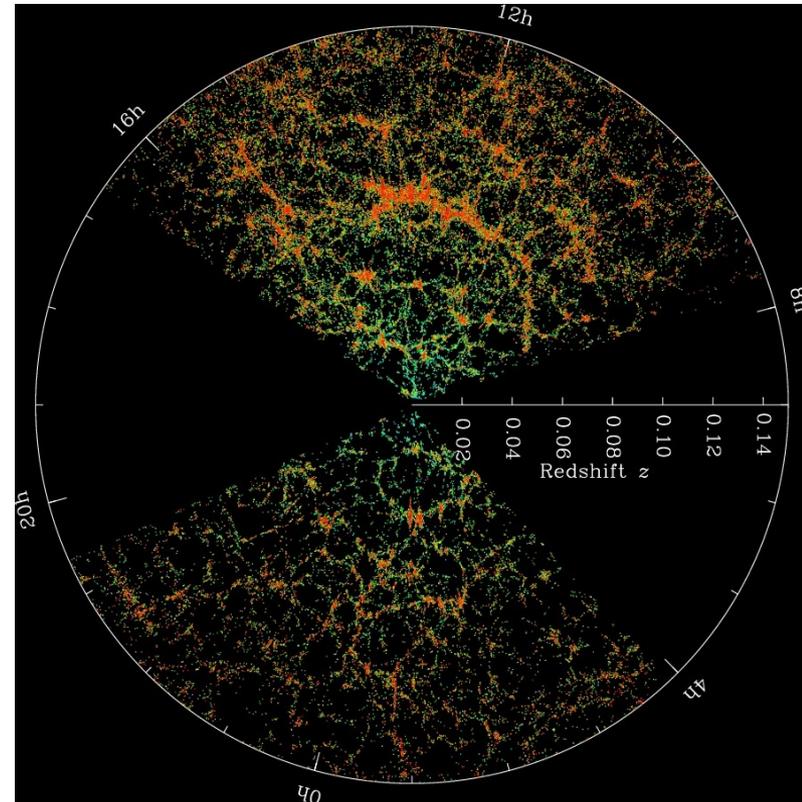
<https://www.desi.lbl.gov/>



# DESI y el mapa 3D más grande del Universo.



- SDSS:  $Z \sim 0.14$
- DESI:  $Z \sim 2$
  
- 30 millones de espectros.
- Galaxias, Quásares, Estrellas.
- Oscilaciones Acústicas de Bariones.



Crédito por la Imagen: M.Blanton and SDSS.

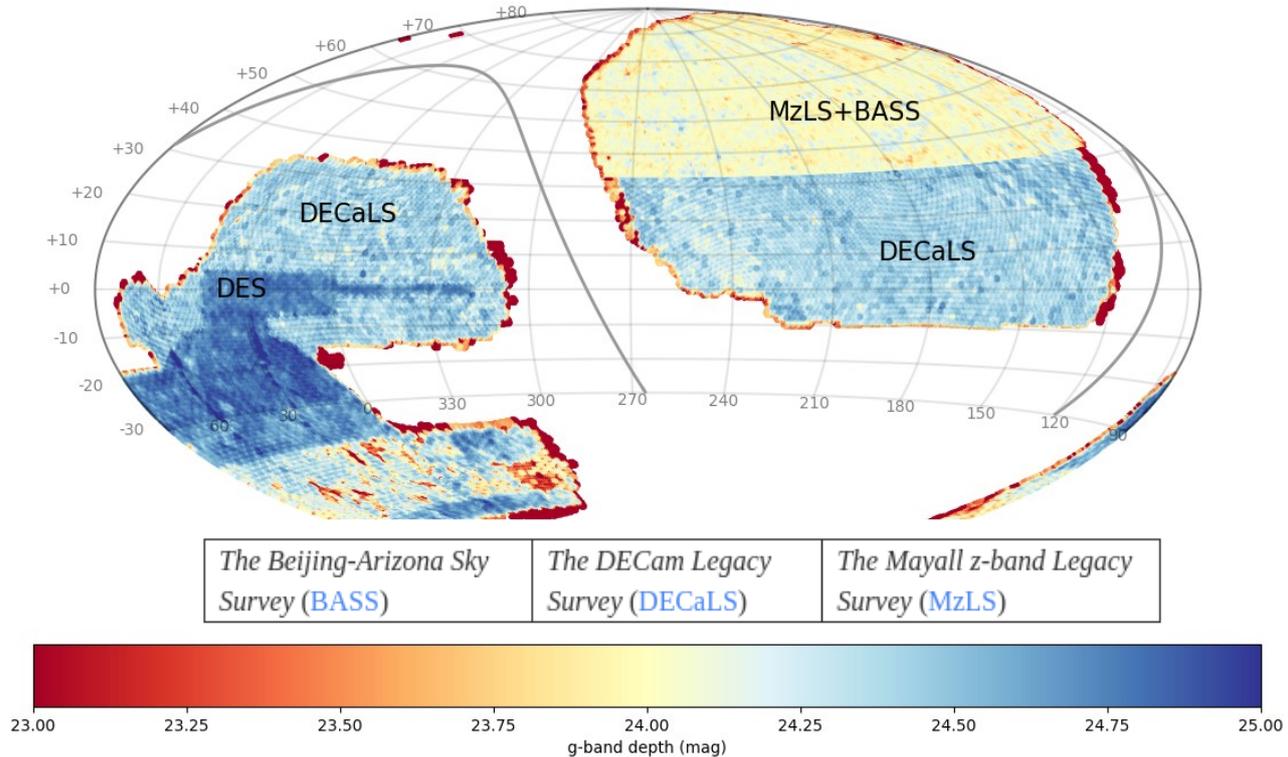


# Observaciones de DESI.



Imagen tomada Legacy Survey.

<https://www.legacysurvey.org/>



- Legacy Survey
- 14,000 deg<sup>2</sup>

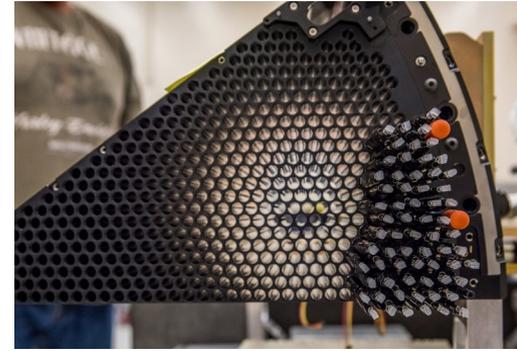
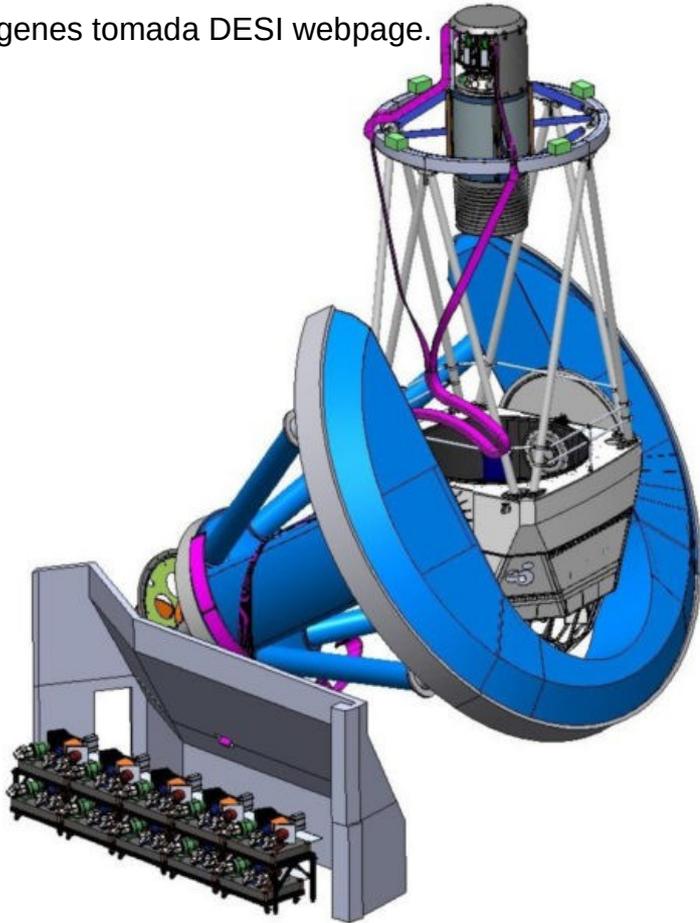
[1] Dey et. al. - Overview of the DESI Legacy Imaging Surveys. 2019



# ¿Cómo se miden los espectros de DESI?



Imágenes tomada DESI webpage.

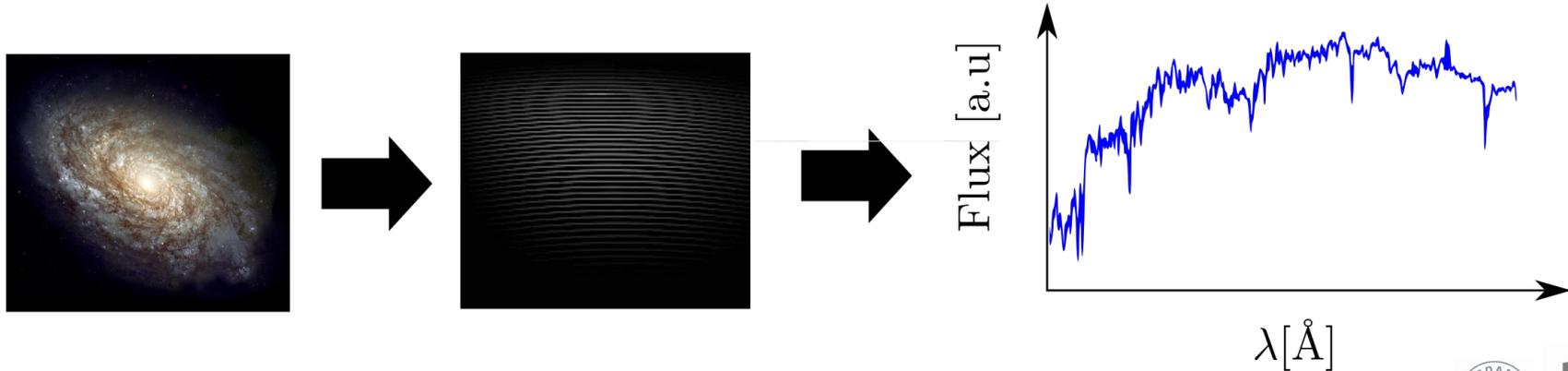


- 8 deg<sup>2</sup> campo de visión
- 5000 Posicionadores controlados computacionalmente
- 5000 espectros tomados en simultáneo.
- Rango de longitud de onda de 360 nm a 980 nm.
- 10 espectrografos en 3 bandas espectrales b, r, z.
- Resolución entre 2000 y 5000.

# Evaluando la calidad de los datos de DESI.



- Observación
- Correcciones (Flat, Bias)
- Calibración (Flujo, longitud de Onda)
- Espectro finalmente calibrado
- **Problema** Encontrar errores instrumentales o de reducción

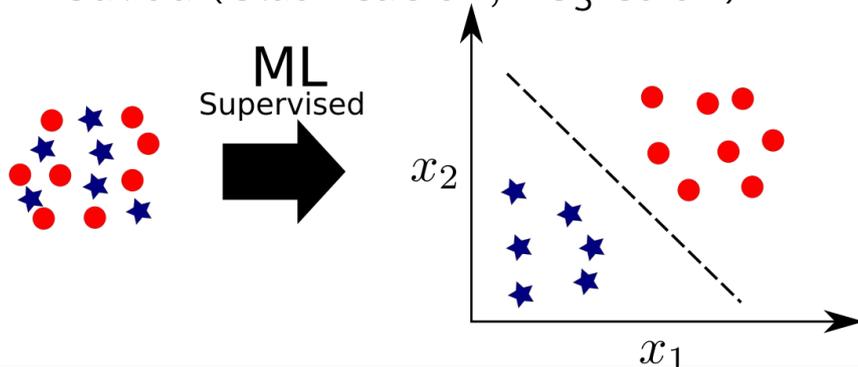


# Machine Learning (Supervised & Unsupervised).



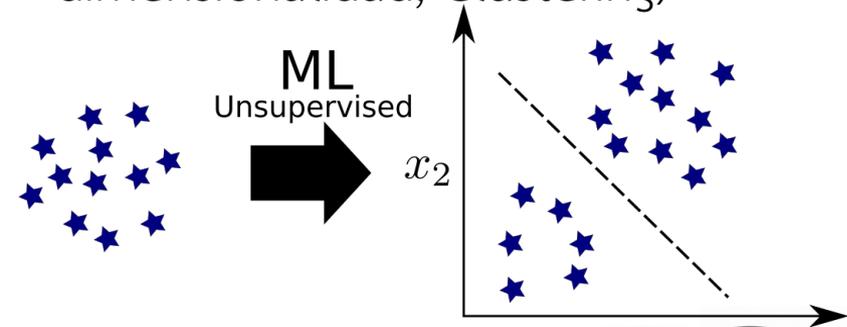
## Supervisado

- Datos crudos
- Preprocesamiento
- Extracción de features
- Train – Test – Validation (Algoritmo)
- Salida (Clasificación, Regresión)



## No Supervisado

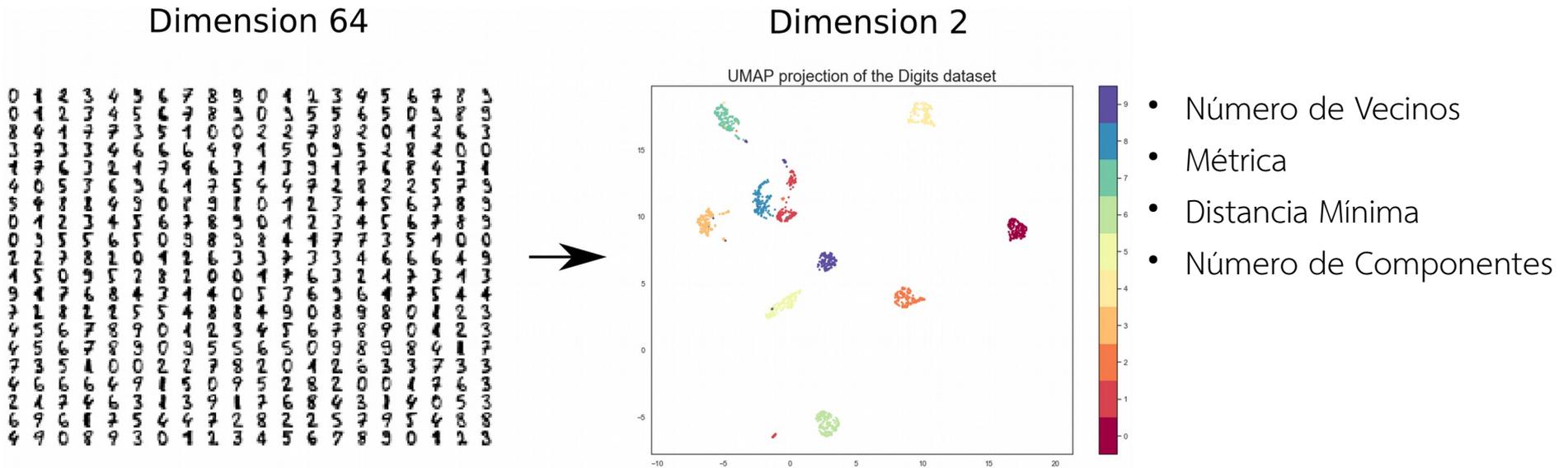
- Datos crudos
- Preprocesamiento
- Extracción de features
- Interpretación (Algoritmo)
- Salida (Reducción de dimensionalidad, Clustering)



# El algoritmo UMAP (Unsupervised).



- Uniform Manifold Approximation and Projection (UMAP) es una técnica de reducción de dimensionalidad que puede ser usada para la visualización pero en general para la reducción de dimensionalidad<sup>1</sup>.
- Ha sido usado en la detección de outliers<sup>2</sup>.



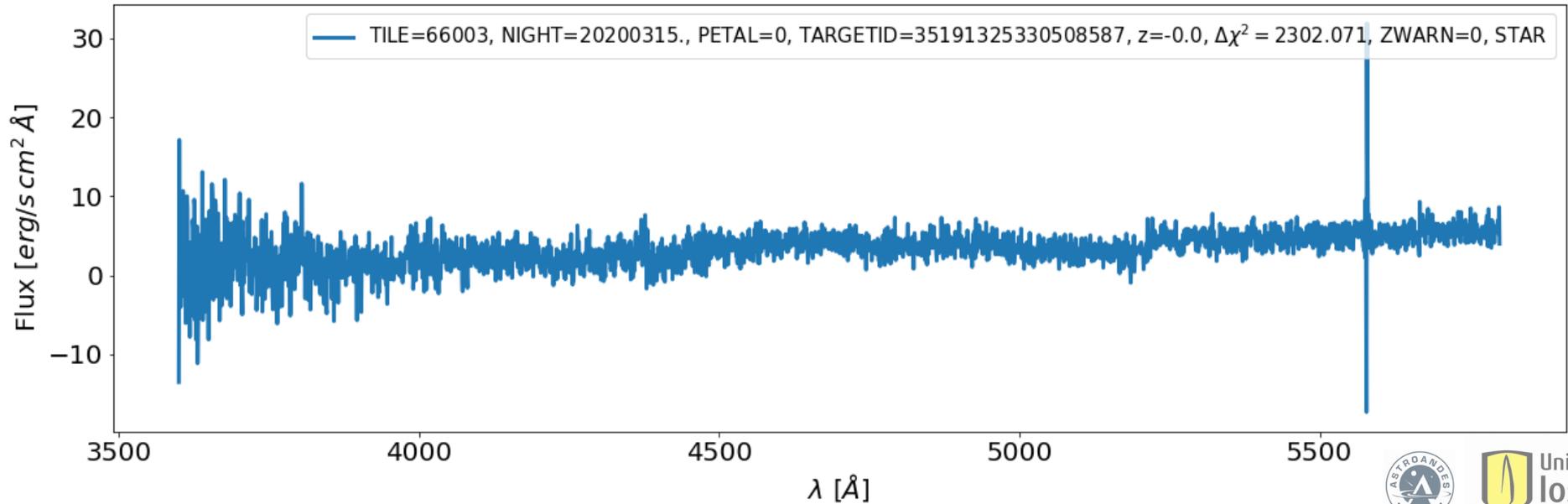
1 <https://umap-learn.readthedocs.io/en/latest>

2 McInnes, L., J. Healy y J. Melville 2018.

# Andes Data Release



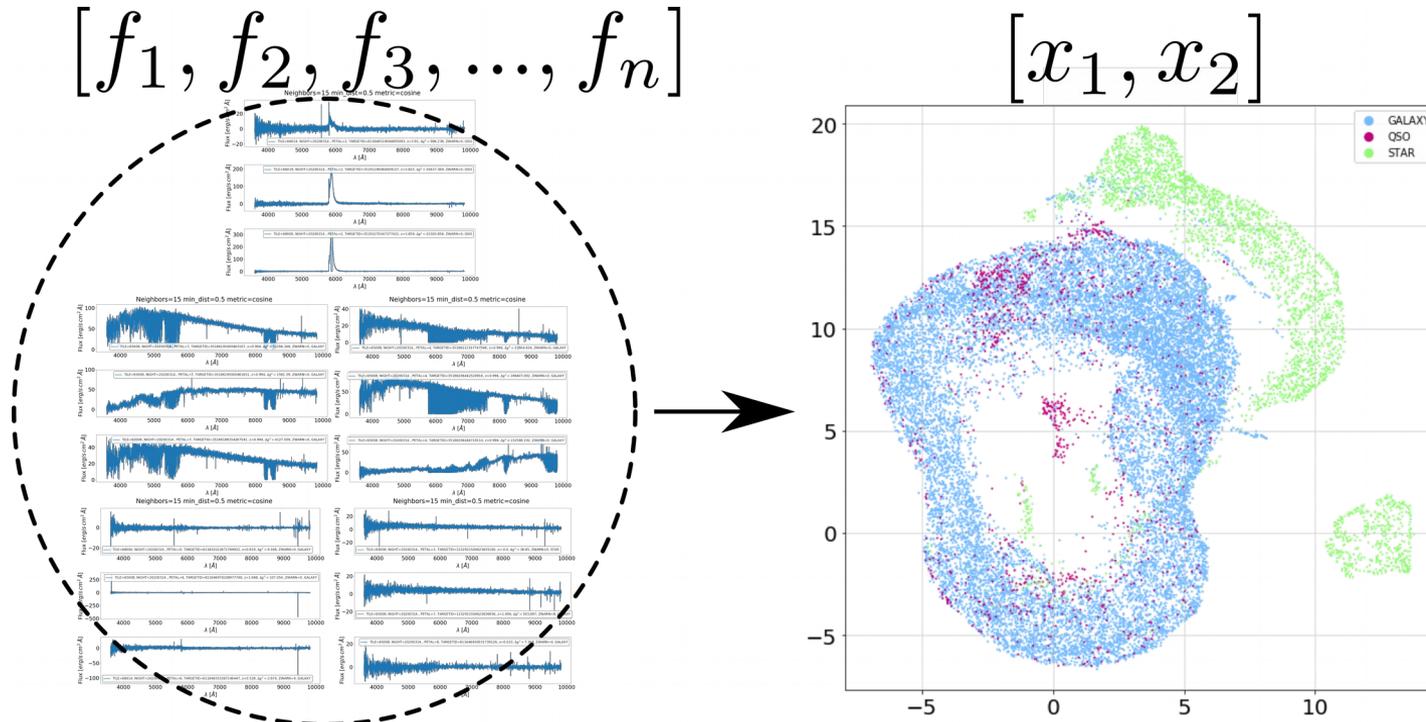
- El Data Release Andes es el primer set de medidas de la colaboración DESI.
- Estas observaciones fueron obtenidas en dos noches de Marzo del 2020.
- El dataset resultante tiene cerca de 22K espectros, cada espectro con medidas de flujo de al menos 2K puntos en cada banda espectral: B, R o Z y aproximadamente 6K en el rango total de longitud de onda de 3600Å a 9600Å.



# Andes + UMAP: Primeros Resultados



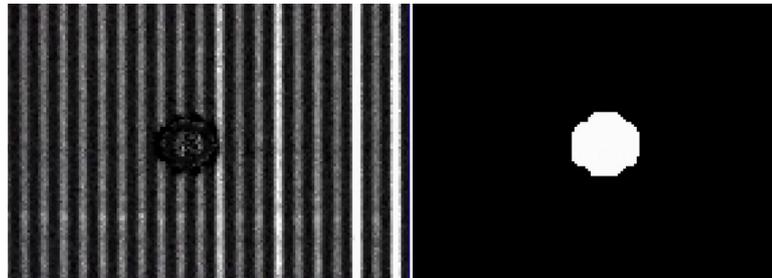
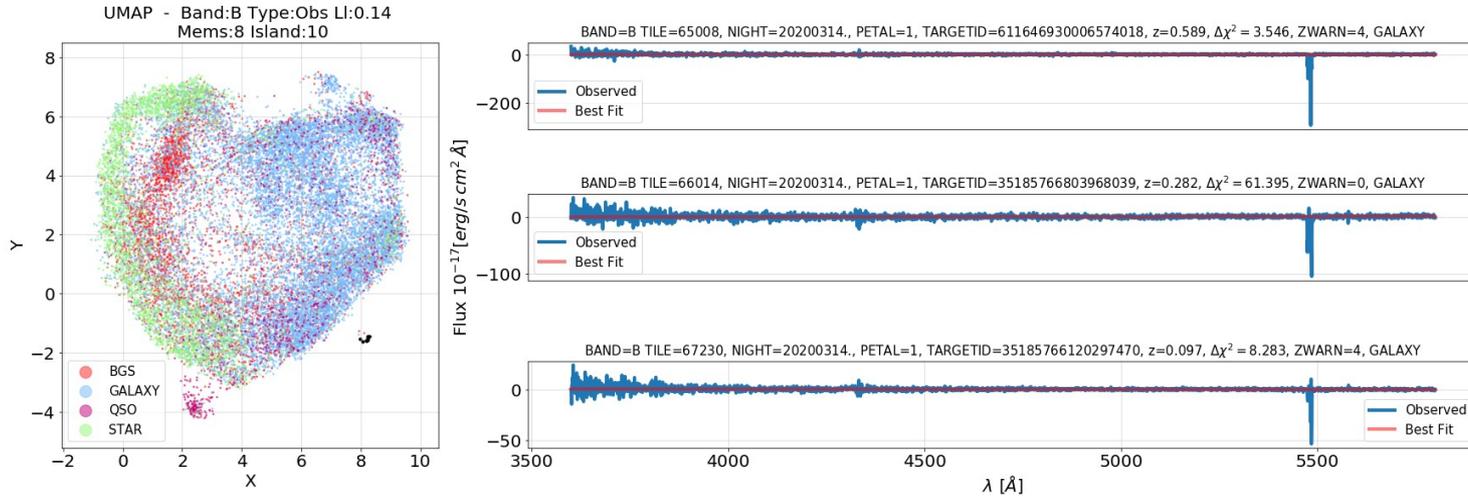
- Usar UMAP para encontrar outliers en el espacio reducido.
- Aplicar UMAP sobre los datos crudos de flujo explorando diferentes metapámetros.



# Andes + UMAP: Primeros Resultados



- Propiedad particular a 5500Å

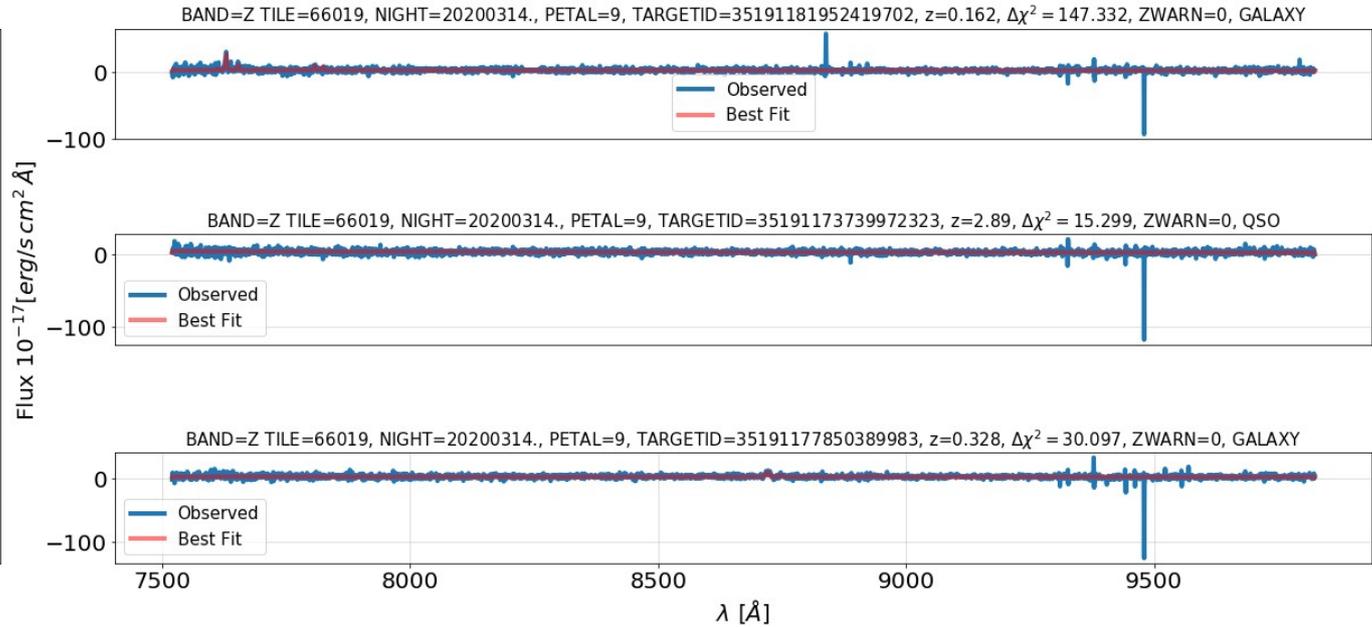
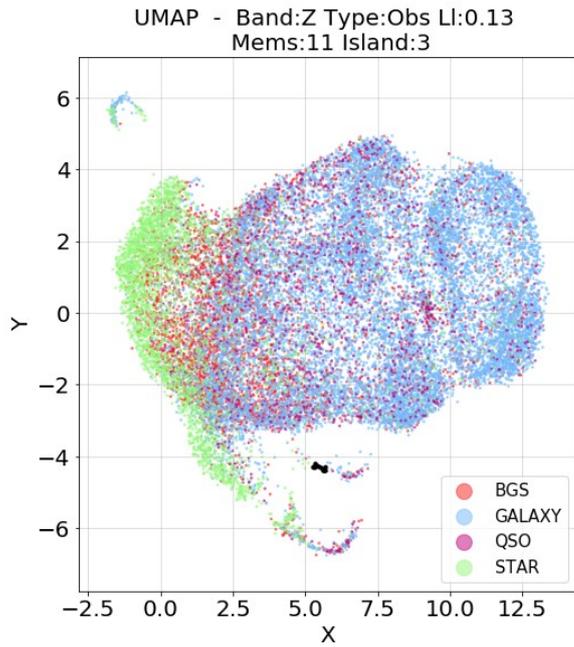


Defecto en la CCD

# Andes + UMAP: Primeros Resultados



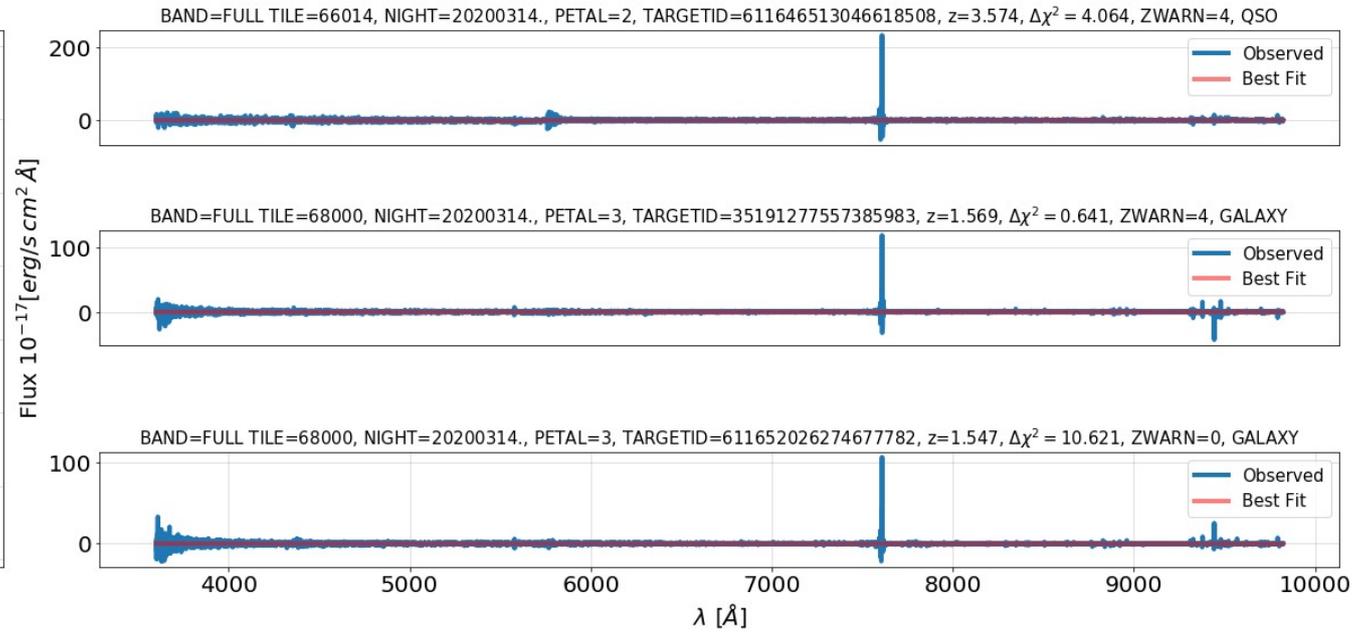
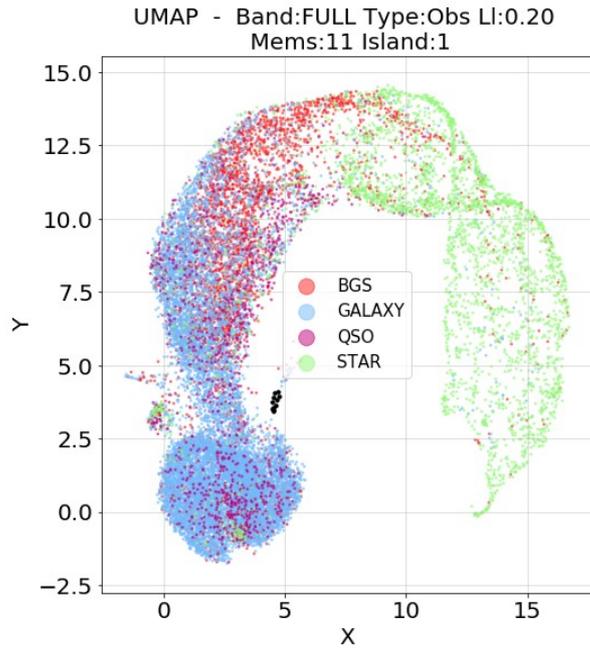
- Propiedad particular a 9500Å



# Andes + UMAP: Primeros Resultados



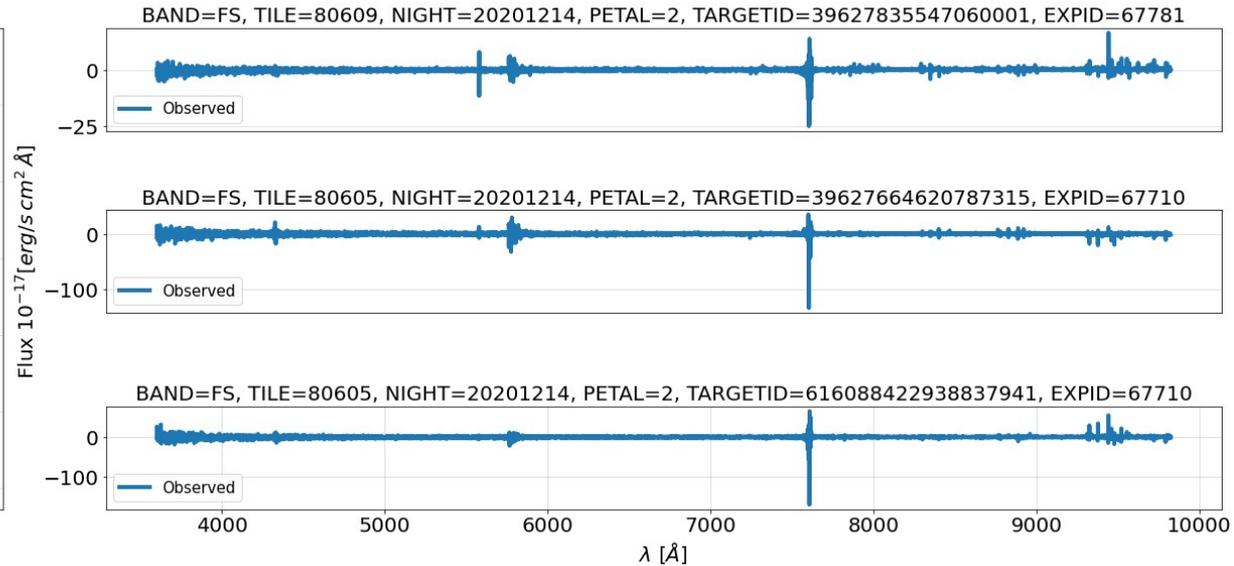
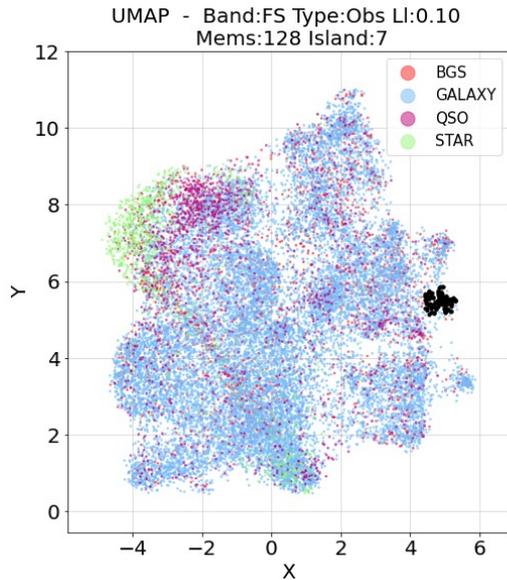
- Propiedad particular a 7600Å



# Blanc + UMAP



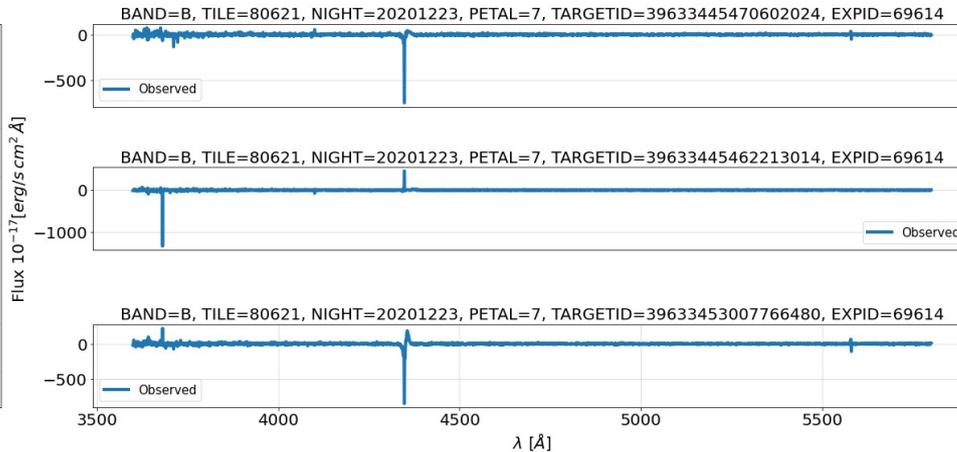
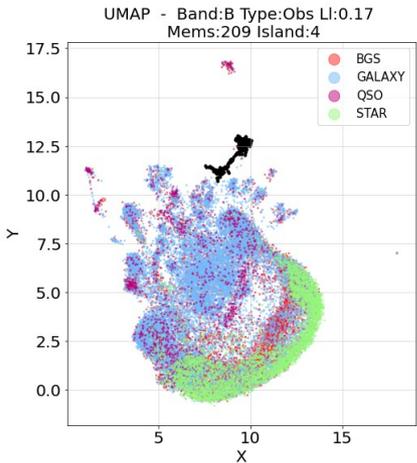
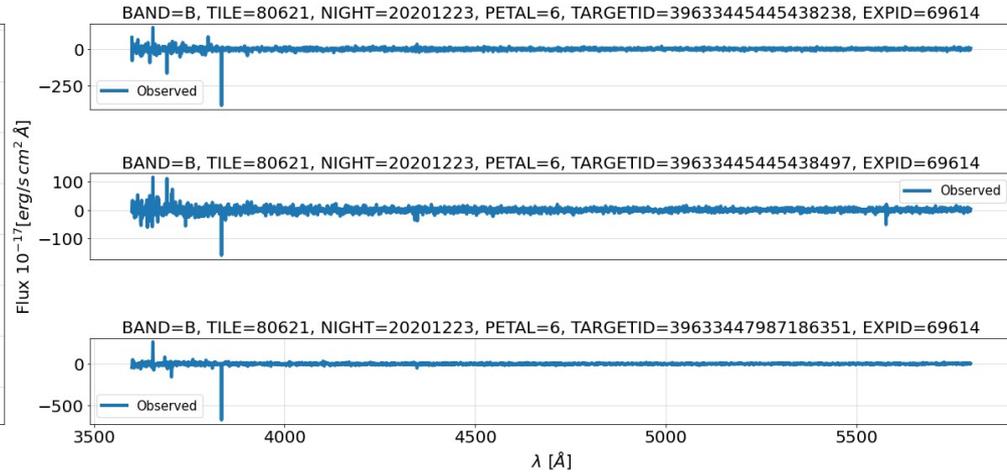
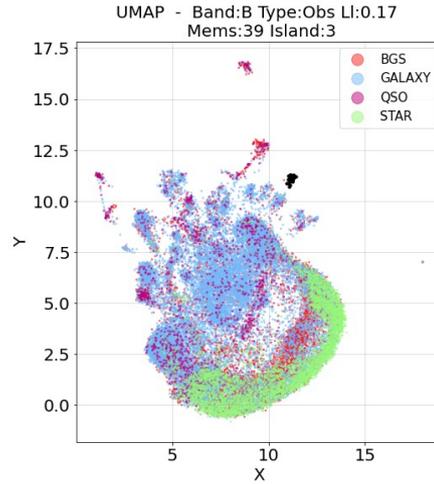
- 10 noches del 14 Dic al 23 Dic del 2020
- ~ 30K espectros por noche
- 300K en total
- Particularidad a 7600Å



# Blanc + UMAP



- Particularidad a 3800Å



- Particularidad a 4300Å



! **Large negative flux around 5500 Angstrom in Andes data** bad spectrum  
#983 opened 19 days ago by jsuarez314

Reporte en Desihub

## UMAP embeddings of DESI Spectra from the Andes Data Release

**Document #:**  
DESI-doc-5843-v1  
**Document type:**  
[Technical Note](#)  
**Submitted by:**  
[Jaime Forero Romero](#)  
**Updated by:**  
[Jaime Forero Romero](#)

**Abstract:**  
We embed DESI spectra into a 2D space using the Uniform Manifold Approximation and Projection (UMAP) algorithm. We use 22105 spectra from the night of March 14, 2020, included in the Andes data release. We compute the embedding separately on the observational data and the best redrock fits. We find that objects with similar classes (QSO, GALAXY, STAR) share similar locations in the embedding, although the class separation is sharper for the redrock fits. However, we find that the observed spectra of ~300 objects are located in outlier islands. In most cases, these islands seem to correspond to calibration/reduction errors from the spectroscopic pipeline. The best-fit spectra have a total of ~1200 objects in outlier islands; we do not have an explanation for those outliers.

**Files in Document:**  

- [desi\\_umap\\_best\\_fit.pdf](#) (5.7 MB)

Get all files as [tar.gz](#), [zip](#).

**Topics:**  

- [Technical Notes](#)

**Authors:**  

- [Jaime Forero Romero](#)
- [John F. Suárez Pérez](#)

**Viewable by:**  

- [DESI](#)

**Modifiable by:**  

- [DESI](#)

**Quick Links:**  
[Latest Version](#)

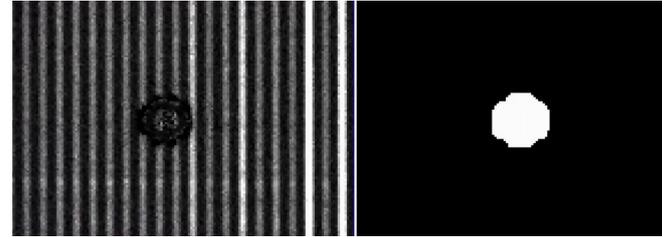
DocDB



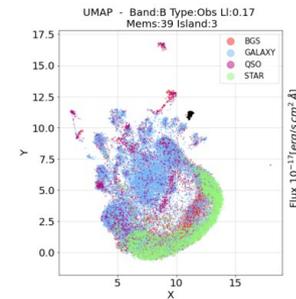
# Conclusiones



UMAP puede ser usado para hacer una primera exploración a modo de filtro para identificar errores instrumentales y/o de calibración.



La representación de baja dimensionalidad puede ser usada en conjunto con otros algoritmos de ML para hacer otro tipo de ejercicios, p.e. clasificación.



Gracias!!!

 [jf.suarez@uniandes.edu.co](mailto:jf.suarez@uniandes.edu.co)

 <https://jsuarez314.gitlab.io>

