

# Challenges and Opportunities for Data Analysis in Large Astronomical Surveys

A.Ederoclite

contacto  
[aederocl@cefca.es](mailto:aederocl@cefca.es)



@aederocl



Ayuda ICTS-MRR-2021-03-CEFCA, financiada por:



Financiado por  
la Unión Europea  
NextGenerationEU

**het•er•o•clite** (*het'ər ə klīt'*),

**adj.** Also, **het•er•o•clit•ic** (*het'ər ə klīt'ik*), **het'er•o•clit'i•cal**.

1. irregular or abnormal;  
anomalous.
2. (Grammar) Irregular in inflection;  
having inflected forms belonging to more than one class of stems.

**n.**

1. a person or thing that deviates from the ordinary rule or form.
2. (Grammar) a heteroclite word.

<https://www.wordreference.com/definition/heteroclite>



# Who is talking to you?

Alessandro Ederoclite

2003 - 2006 PhD in Physics

2006 - 2011 Postdoc at VUB (Belgium), ESO (Chile) and IAA (Spain)

2012 - 2018 Staff astronomer at ceFca (Spain)

2018 - 2021 Professor doutor at IAG/USP (Brazil)

2021 - *now* Tenure-track astronomer at ceFca (Spain)

Scientific interests: Wide field surveys, Time-domain astrophysics, Interacting white dwarfs

Non-scientific interests: Sci-fi, travelling, *gastronomy*



# What is a survey anyway?

*An **astronomical survey** is a general map or image of a region of the sky (or of the whole sky) that lacks a specific observational target.*

*Alternatively, an astronomical survey may comprise a set of images, spectra, or other observations of objects that share a common type or feature.*

[https://en.wikipedia.org/wiki/Astronomical\\_survey](https://en.wikipedia.org/wiki/Astronomical_survey)



# What is data science anyway?

*Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processing, scientific visualization, algorithms and systems to extract or extrapolate knowledge and insights from potentially noisy, structured, or unstructured data.*

*Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession.*

[https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)



## “It depends”

Is one telescope better than another?

Is one data analysis better than another?

“It depends on what you want to do”

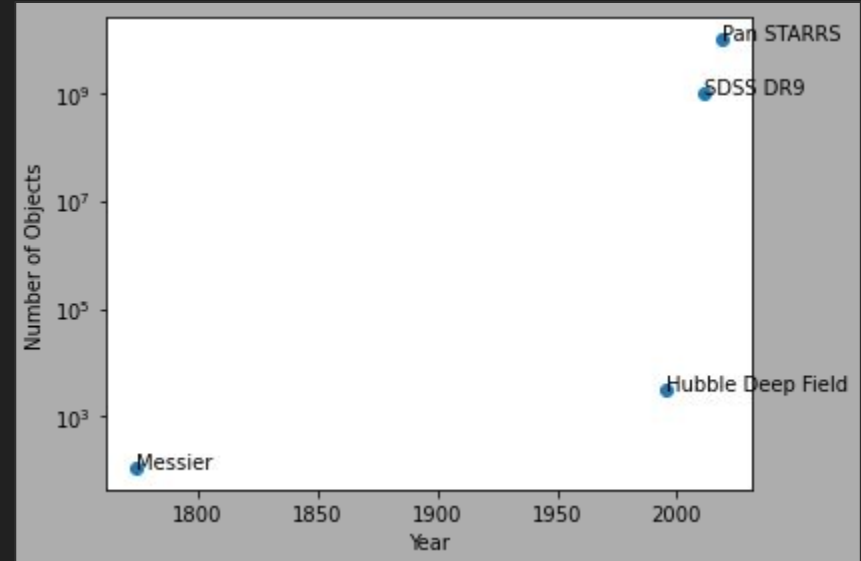
I bet I will repeat this so often that you will be fed up



# Why Data Analysis in Astronomy?

Astronomical dataset are getting larger and larger.

Astronomical data are becoming more complex.





# Change of paradigm in astronomy

- Observations by naked eye
- Observations with the telescope
- Observations with a recording device at the telescope (e.g. photographic plates)
- Observations with digital devices (e.g. CCDs)
- Telescope + computer!





# Astronomy vs Real World

Astronomy

Real world

Easy

Based on  
Physics only

(Mostly)  
uninterrupted

Difficult

Depend on  
telescope  
availability

Depend on  
humans

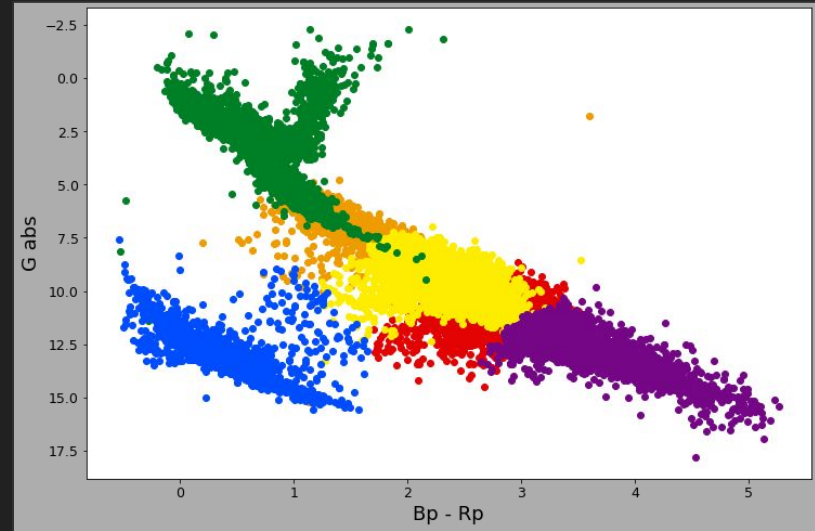
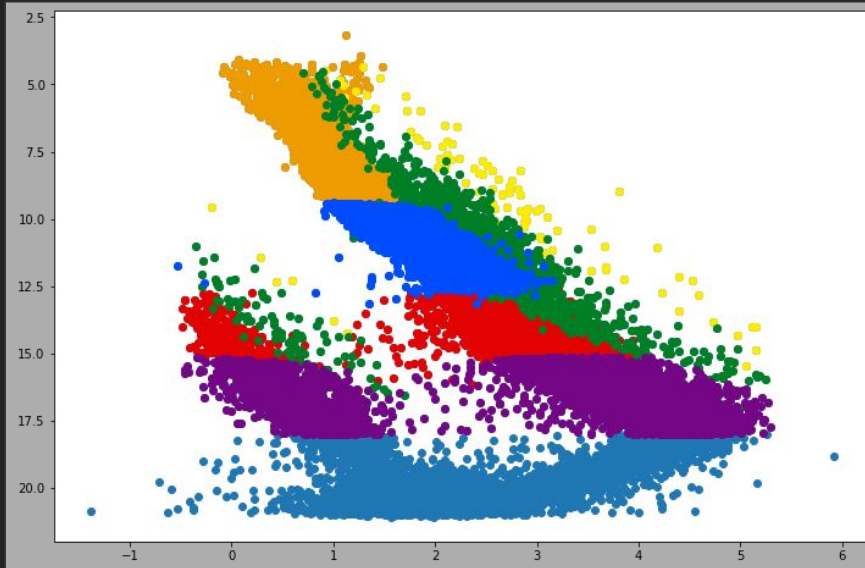
# Correlation is not causation

The advent of the SDSS database was very tempting: people started to correlate everything with everything.



# The problem sits in front of the keyboard

My first attempt to do a ML clustering analysis didn't go as expected





## Golden Rules

Your job is to make hypotheses and verify them.

Data Science is a tool: you are at the driver seat

If you do not ask questions, you do not get answers.

Let's rock'n'roll!