"Machine-learning for analyzing Color-Magnitude diagrams of Galactic Globular Clusters"

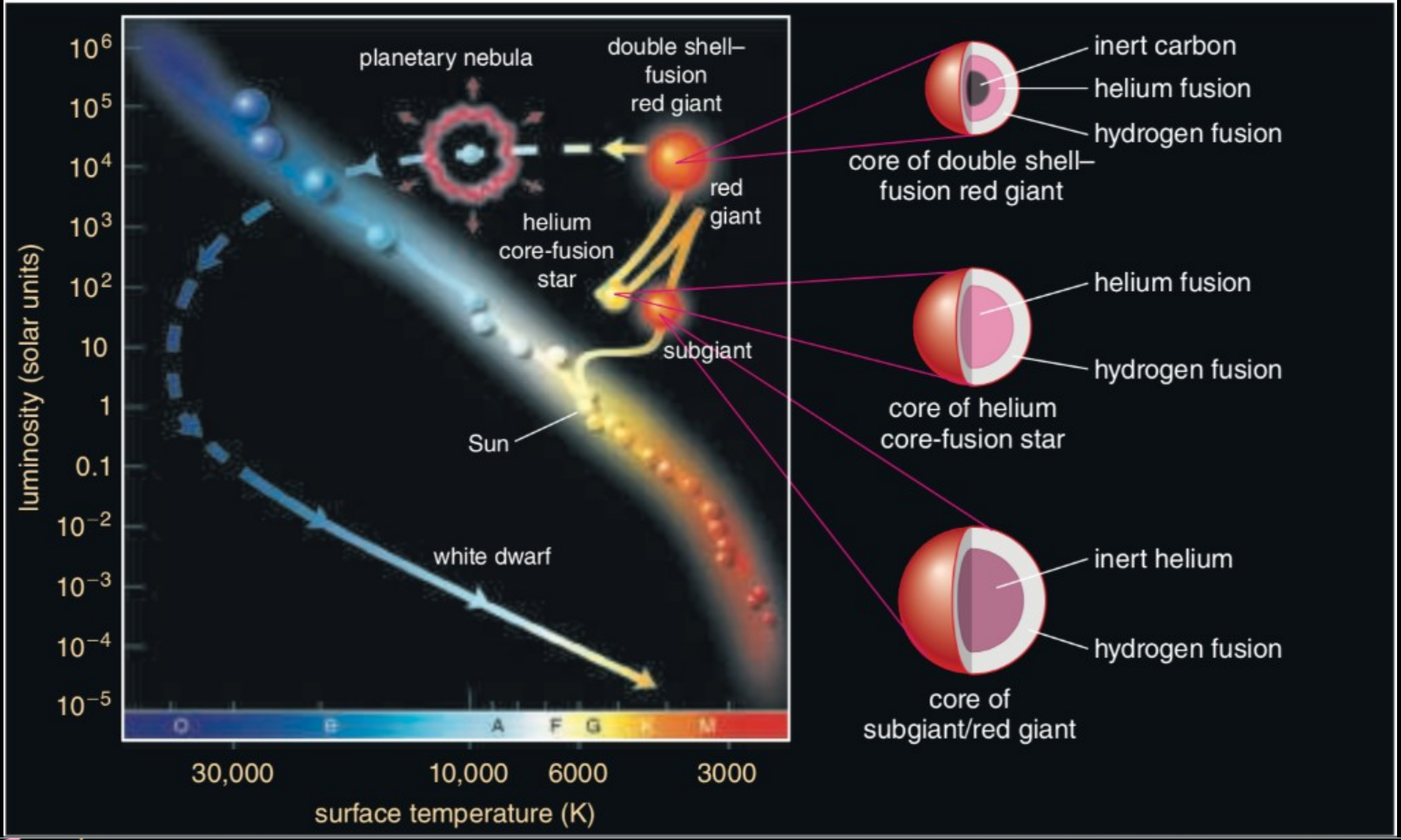
Presenter: Santiago Arceo Díaz

Collaborators: Dr. Kai Zuber (Dreden-TU), Dra Elena Bricio (IT-Col), Dra Xóchitl Trujillo (UCol)

ASP 2024 A VIRTUAL CONFERENCE

ASTRONOMY ACROSS THE SPECTRUM August 22-24

One of the great challenges of Astronomy is to explain the types of stars that we can observe and correlate them with the physical processes occurring in their interior. **This is difficult because it takes a long time for stars to "evolve".**

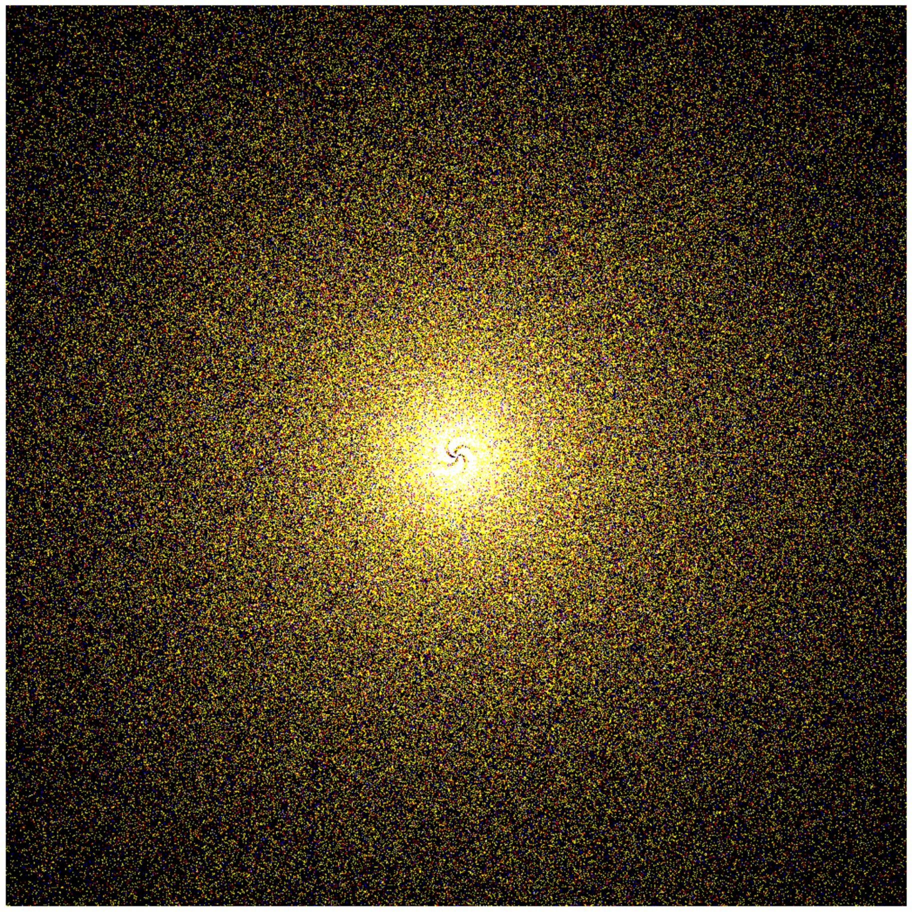


Globular clusters: spherical groups of stars born simultaneously and that share certain characteristics, such as mass and chemical composition.

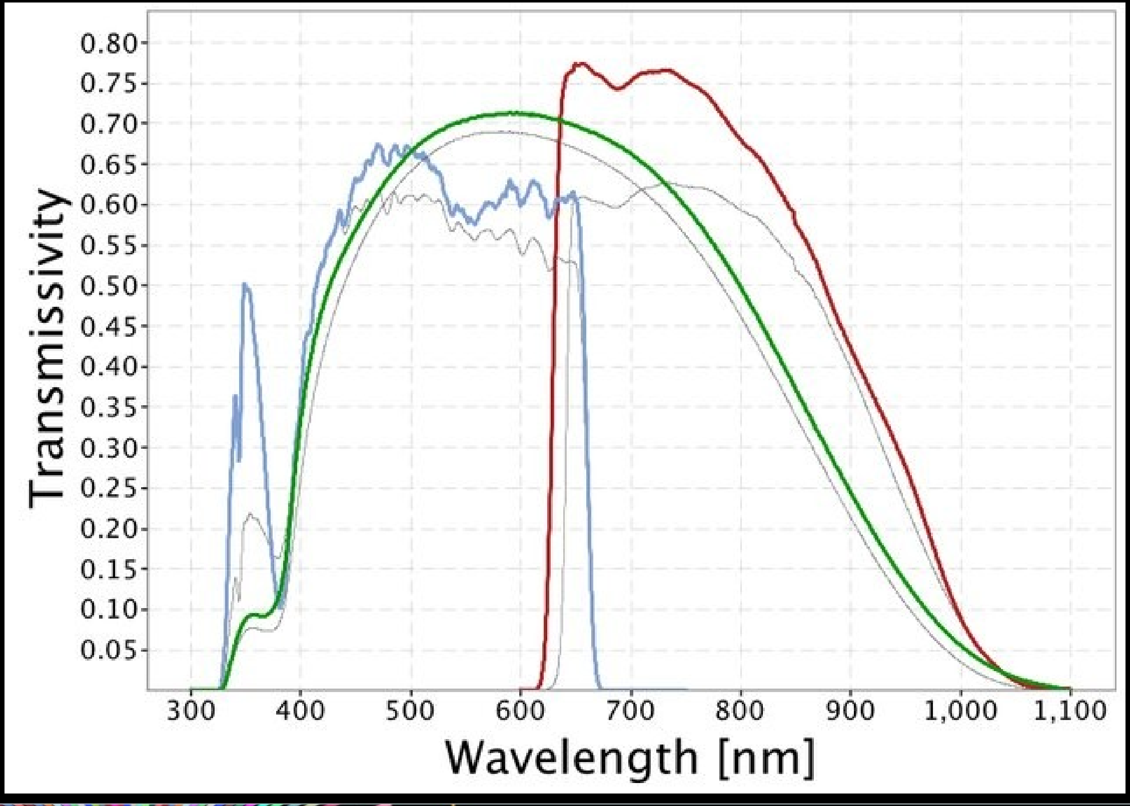
Their nature makes them an **ideal laboratory for studying stellar evolution** **since they allow us to observe stars at different stages of their evolution at the same time.**

What we can learn:

- Age and Evolution of the Universe.
- Galactic Formation and Evolution.
- Cosmic Distance Scale.
- Physics outside the Standard Model of Particles.



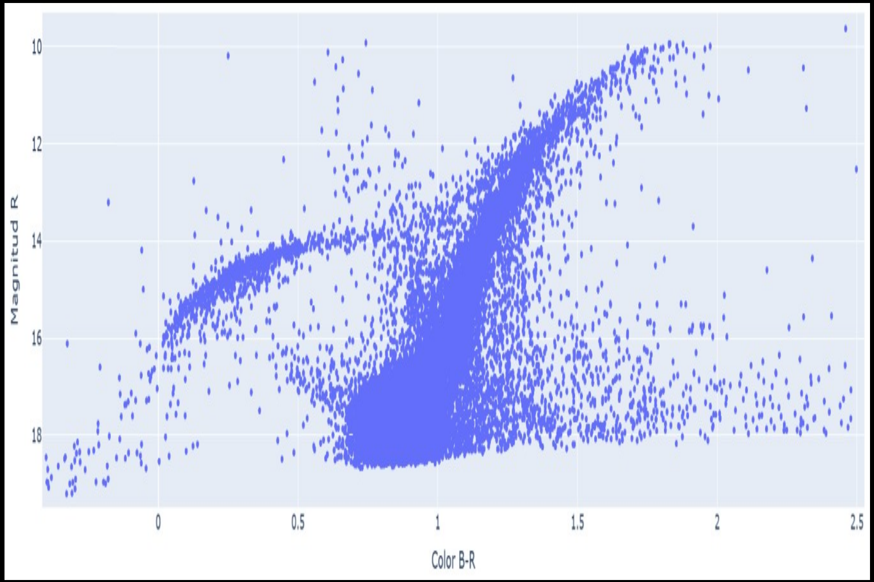
The GAIA mission, launched by the European Space Agency (ESA) in 2013, forms the largest and most precise stellar catalog today.

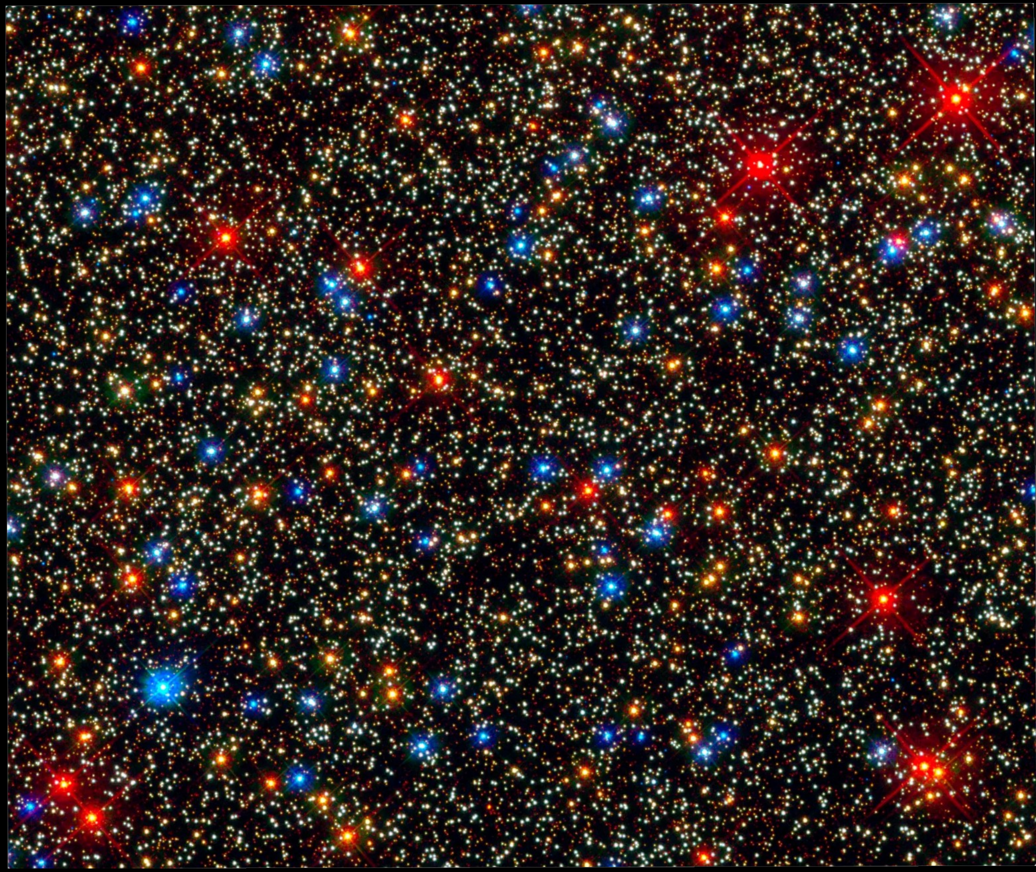


It contains the photometric data of **millions of stars within globular clusters**

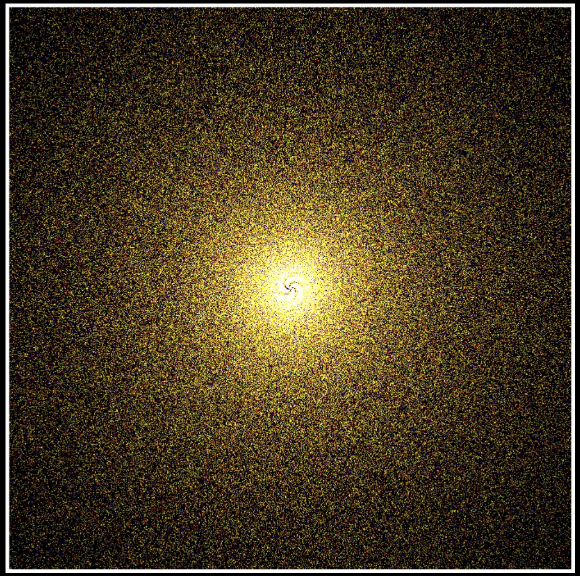
- **G Band (330 - 1050 nm):** This is the primary photometric band of Gaia. It is used for the fundamental astrometric measurements.
- **BP (330 - 680 nm)** Gaia photometric measurement in the blue band **(330 to 680 nanometers).**
- **RP (630 - 1050 nam)** Gaia photometric measurement in the red band **.**

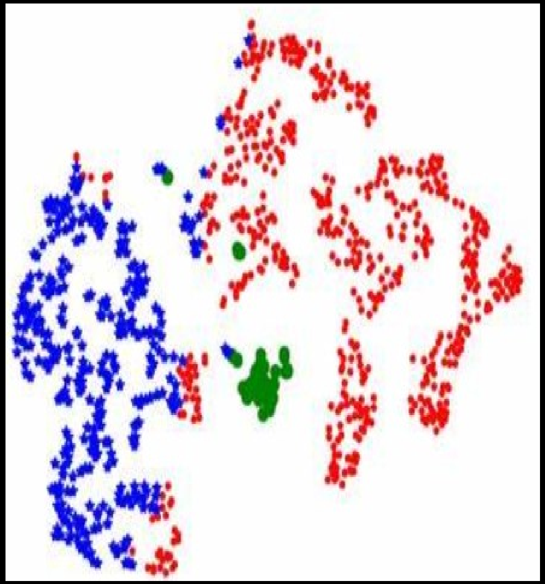
Stars can be classified **according to their color and brightness on a color-magnitude diagram**. Stars in each region of the diagram have distinctive physical properties

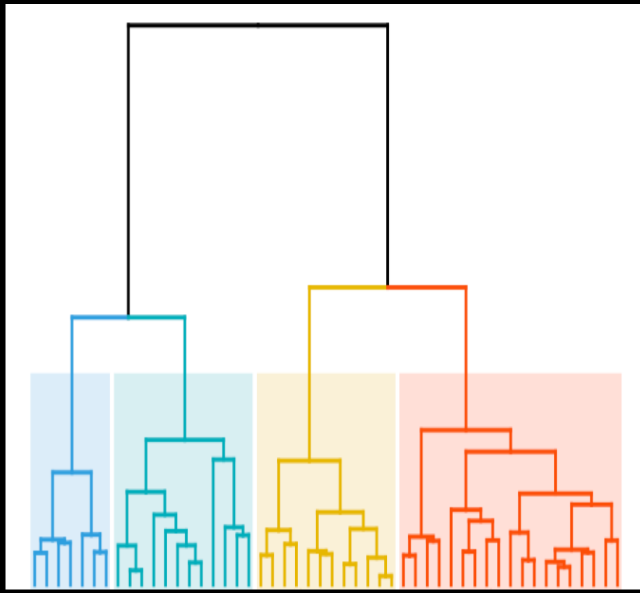




but in a CMD **there are thousands or millions of stars to classify!**





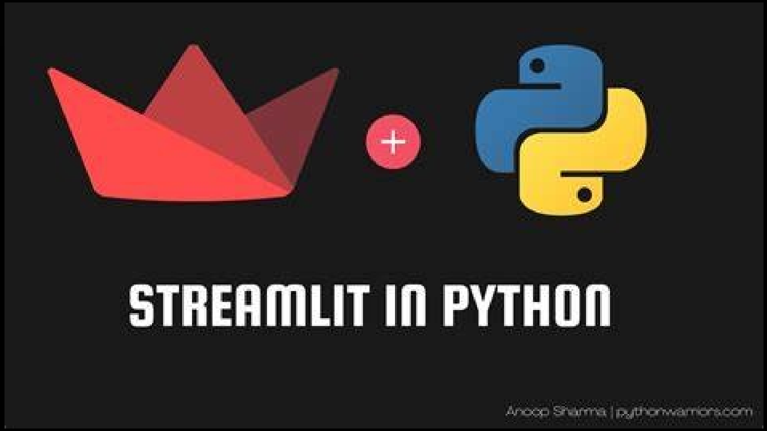


Machine learning enables the automatization of the classification of large volumes of data and objects, significantly reducing the need for human intervention.

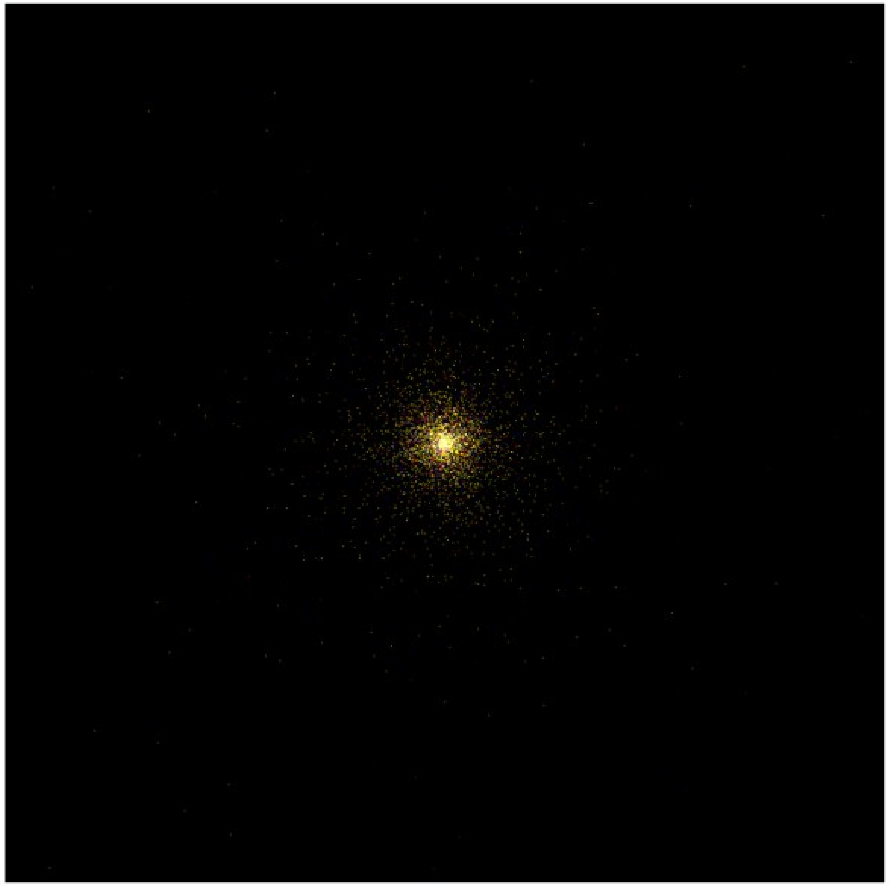
Hierarchical clustering is a data analysis technique that organizes objects into a hierarchical tree based on the similarity or distance between them.

Decision trees select the most relevant features for classification or prediction, helping to identify which attributes have the greatest impact on decision making.

The app was developed in Python 3 and integrated into streamlit cloud



Machine-learning for analyzing Color-Magnitude diagrams of Galactic Globular Clusters



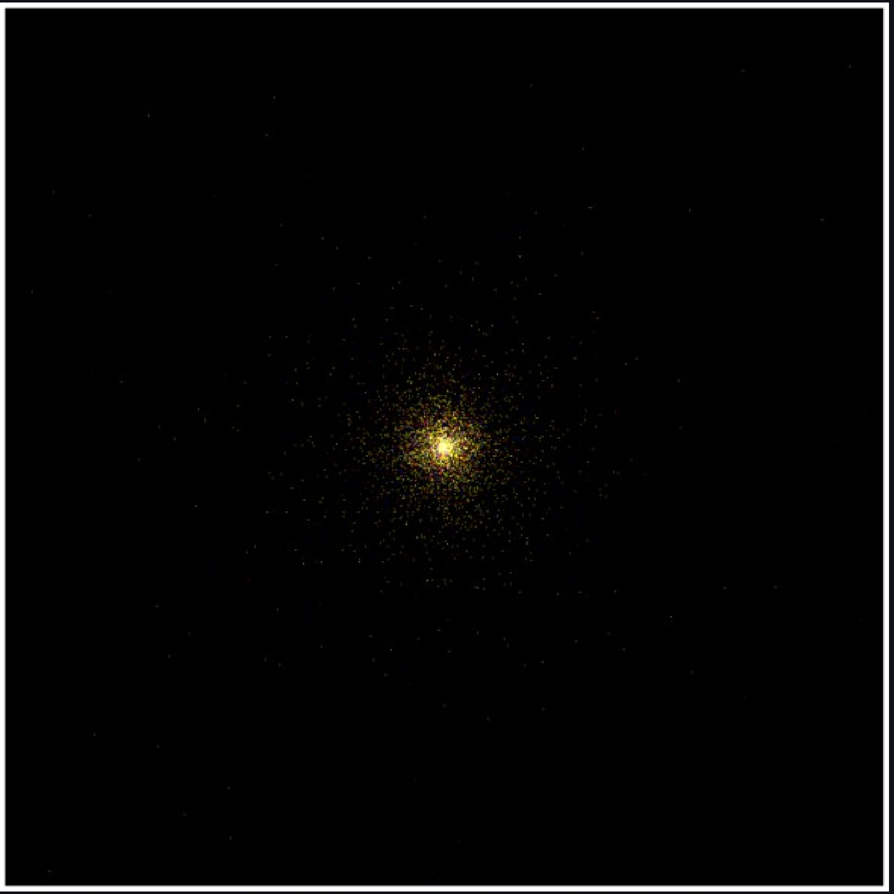
Graphical representation of the appearance of a globular cluster (made from a point distribution that follows the King mass distribution).

Machine-learning for analyzing Color-Magnitude diagrams of Galactic Globular Clusters

Overview:

This application allows the analysis of the color-magnitude diagrams of globular clusters in the Milky Way using machine learning. By applying the **hierarchical clustering algorithm,** the stars in the database are grouped into sets according to their similarity.

In many cases, **these sets correspond to different stages of stellar evolution.** Using a decision tree algorithm, the rules and cut-off points are obtained in the variables of interest that define each set of stars.



Graphical representation of the appearance of a globular cluster (made from a point distribution that follows the King mass distribution).

In the first section you can load data for the central region of any globular cluster from the list. The application displays the database it will use, as well as some details about it.

All generated files can be downloaded by clicking

Color-Magnitude Diagram

Instructions: Please select the files with the **photometry** (Cluster-name_photo.csv) and **observable parameters** (Cluster-name_metal.csv) for any of the globular clusters displayed below to analyze. The third GAIA data release (DR3) obtained data for each globular cluster (https://gea.esac.esa.int/archive/).

Select CSV files to merge:

Tuc47_metal.csv × Tuc47_photo.csv ×

Merged DataFrame:

	source_id	teff_gspphot	logg_gspphot	mh_gspphot	phot_g_mean_mag	phot_bp_me
0	4.689621262329504e+18	None	None	None	None	
1	4.68963263540471e+18	None	None	None	None	
2	4.689632635404711e+18	None	None	None	None	
3	4.689632738492985e+18	None	None	None	None	
4	4.68963280293104e+18	None	None	None	None	
5	4.689632802931043e+18	None	None	None	None	
6	4.689632802931045e+18	None	None	None	None	
7	4.689632807212407e+18	None	None	None	None	
8	4.689632910282381e+18	None	None	None	None	
9	4.68963291028239e+18	None	None	None	None	

Additional information

Number of rows: 174313

Number of columns: 10

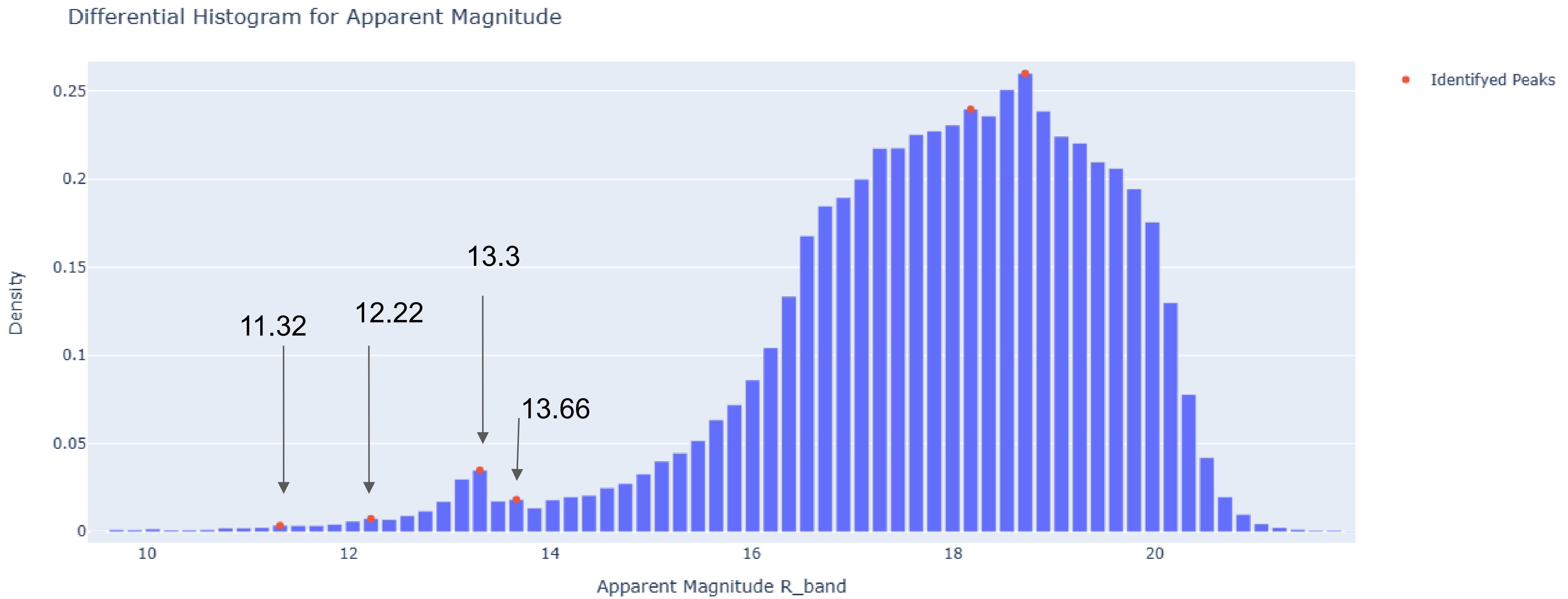
Number of rows with missing data: 173245

Download .xlsx file

Download .csv file

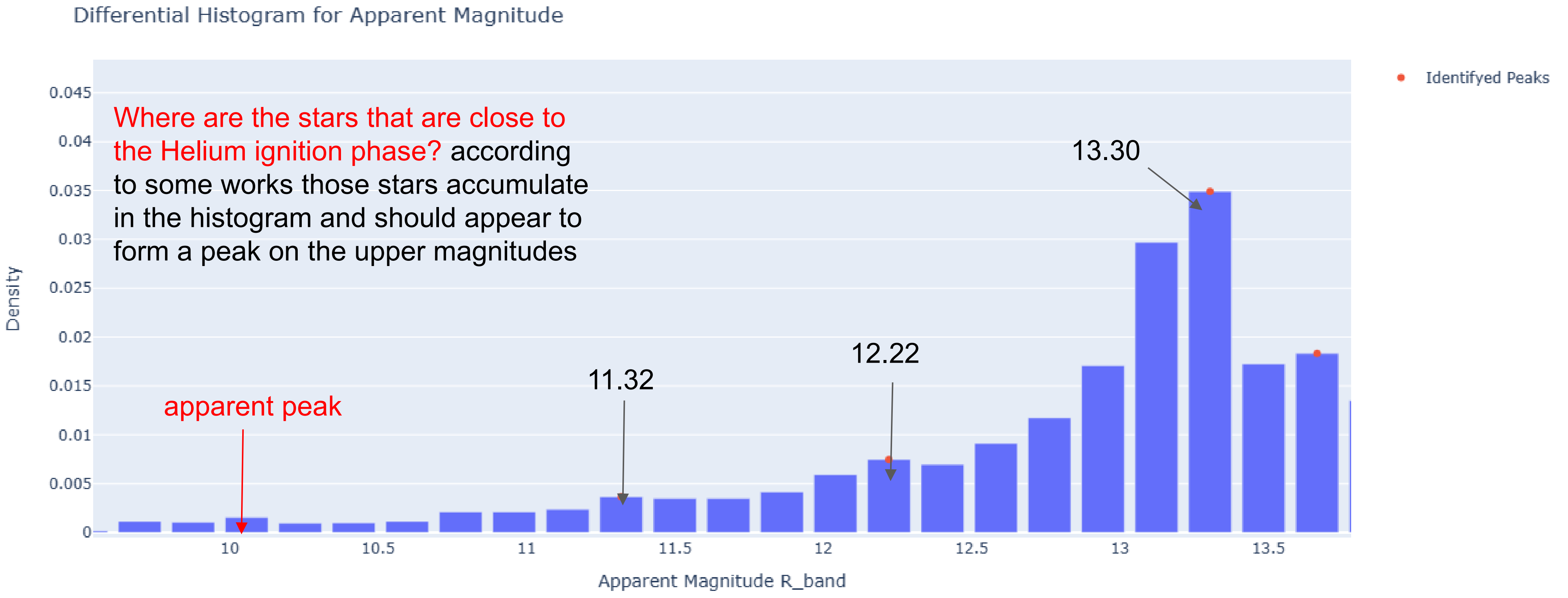
Differential luminosity histogram for the Rp band in 47 Tuc's CMD

Values of phot_R_mean_mag at the peaks of the histogram:
11.32, 12.22, 13.30, 13.66, 18.17, 18.71



Differential luminosity histogram for the Rp band in 47 Tuc's CMD

Values of phot_R_mean_mag at the peaks of the histogram:
11.32, 12.22, 13.30, 13.66, 18.17, 18.71



Two dimensional plots of cluster parameters

In this section, you can visualize the color-magnitude diagrams of the selected globular cluster. Please select the variables to represent the horizontal and vertical axes of the bar. The variables "bp_rp", "bp_g" and "g_rp" correspond to the colors, while "phot_g_mean_mag", "phot_bp_mean_mag" and "phot_rp_mean_mag" correspond to the magnitudes integrated in the G, BP and RP bands. In addition to color-magnitude diagrams, you can create graphs from other variables, such as estimated effective temperature, metallicity, or surface gravity.

Instructions: Select at least two variables to generate a two-dimensional plot. Some of the plot's settings can be manipulated on the menu in its upper right corner. The resulting plot can be saved by clicking on the icon with the shape of a camera.

Plot:

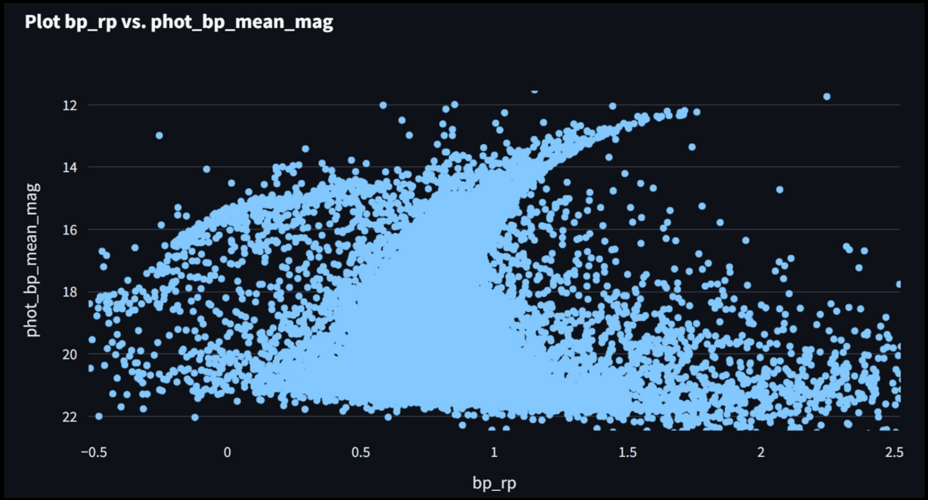
Select the horizontal axis for the plot:

bp_rp

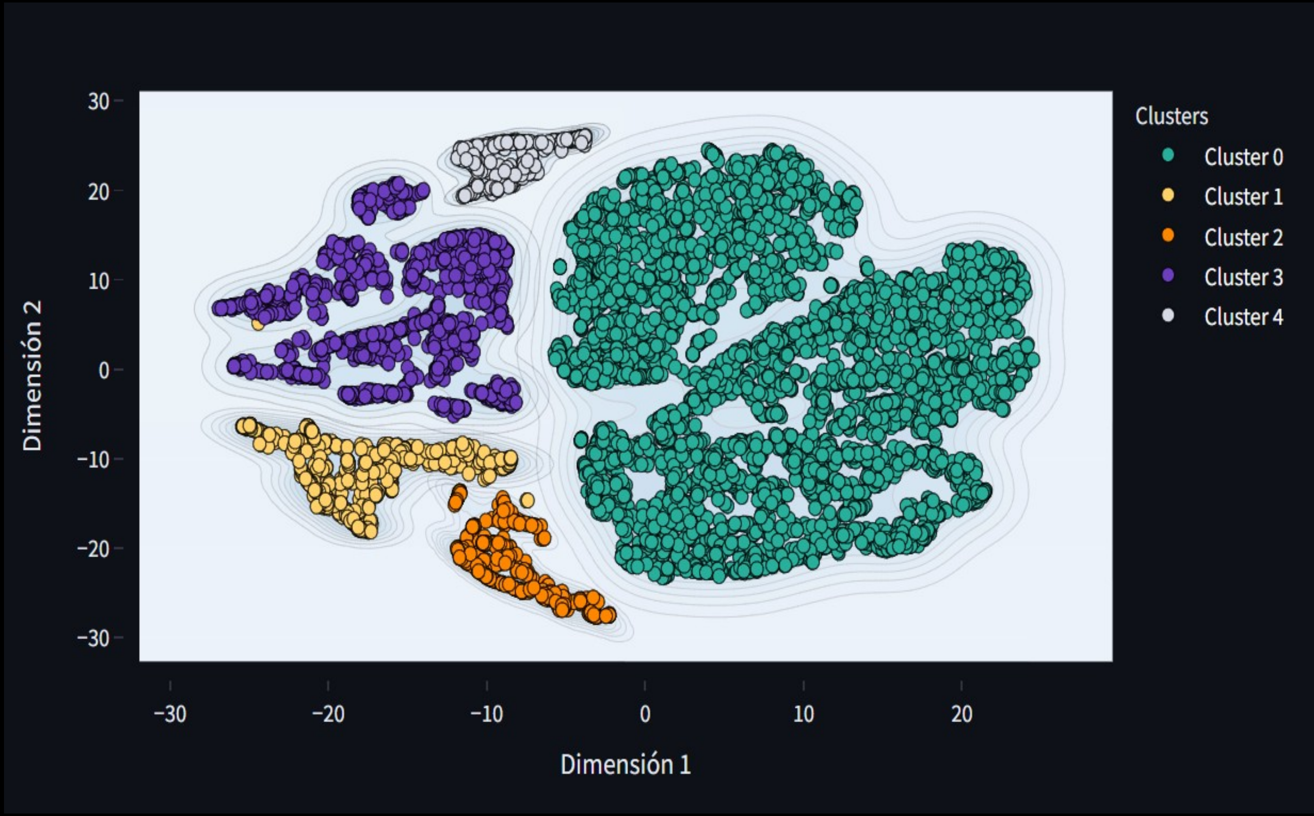
Select the vertical axis for the plot:

phot_bp_mean_mag

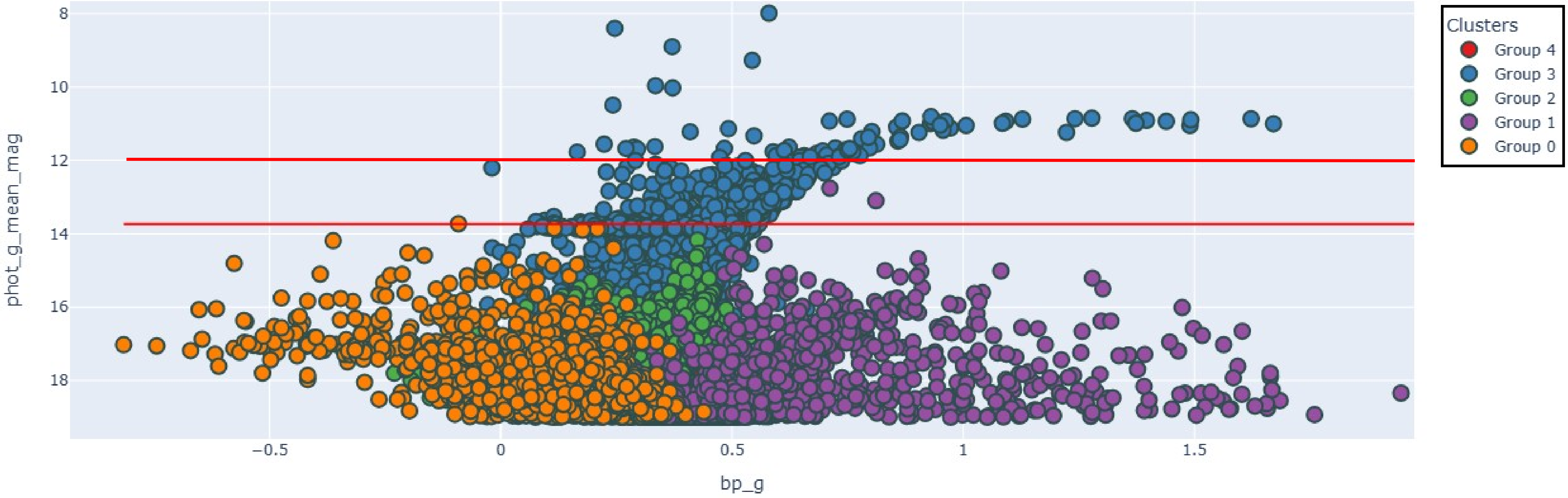
Make Plot



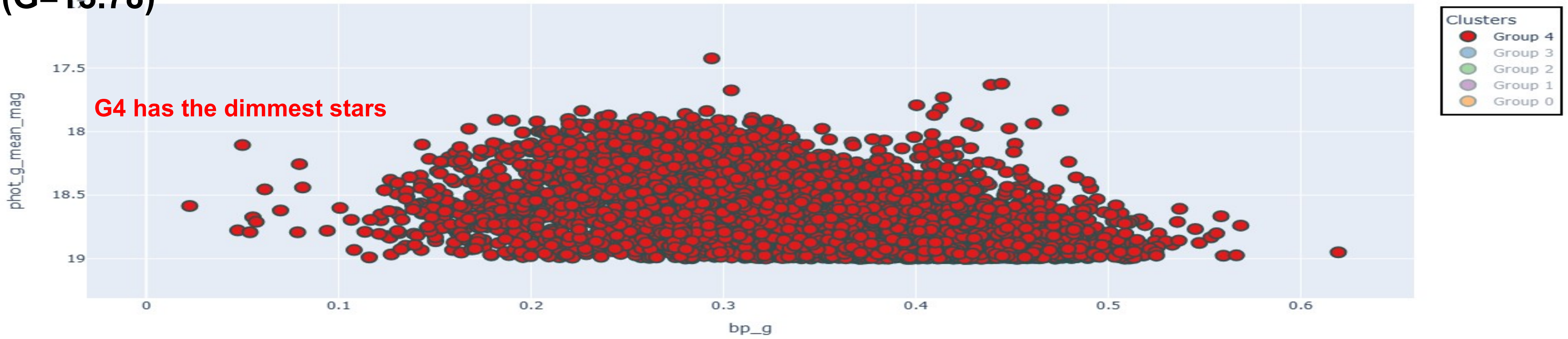
Hierarchical clustering: divides the stars in within the cluster in subsets, in accordance on the degree of similitude they have in some parameters (Gband, teff, logg, g-rp, bp-rp, etc.)



If we give the stars in each group an specific color, this is the resulting CMD (the red line represents the apparent magnitude of the peak in the G band)

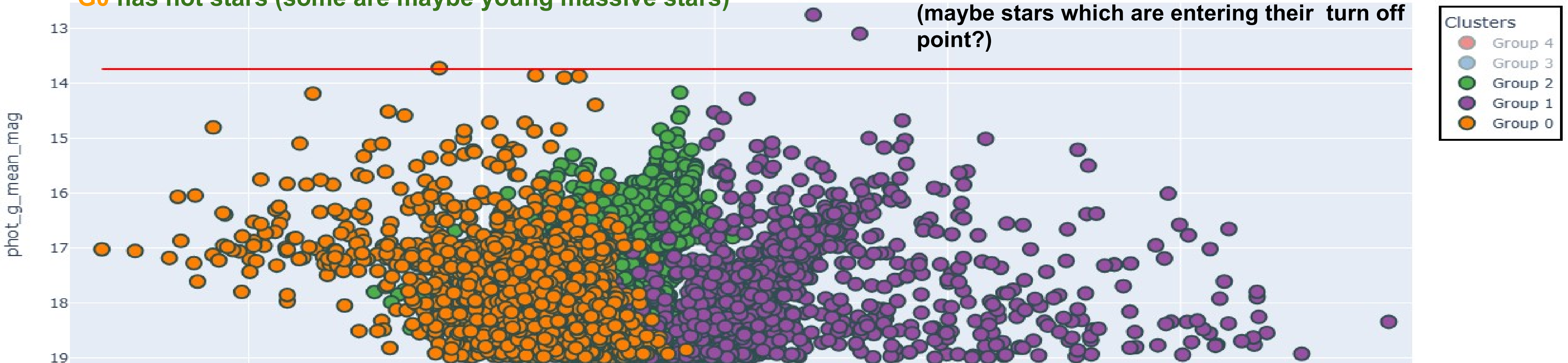


Three of these groups have stars whose luminosity is below the peak shown in the histogram (G=13.78)



G0 has hot stars (some are maybe young massive stars)

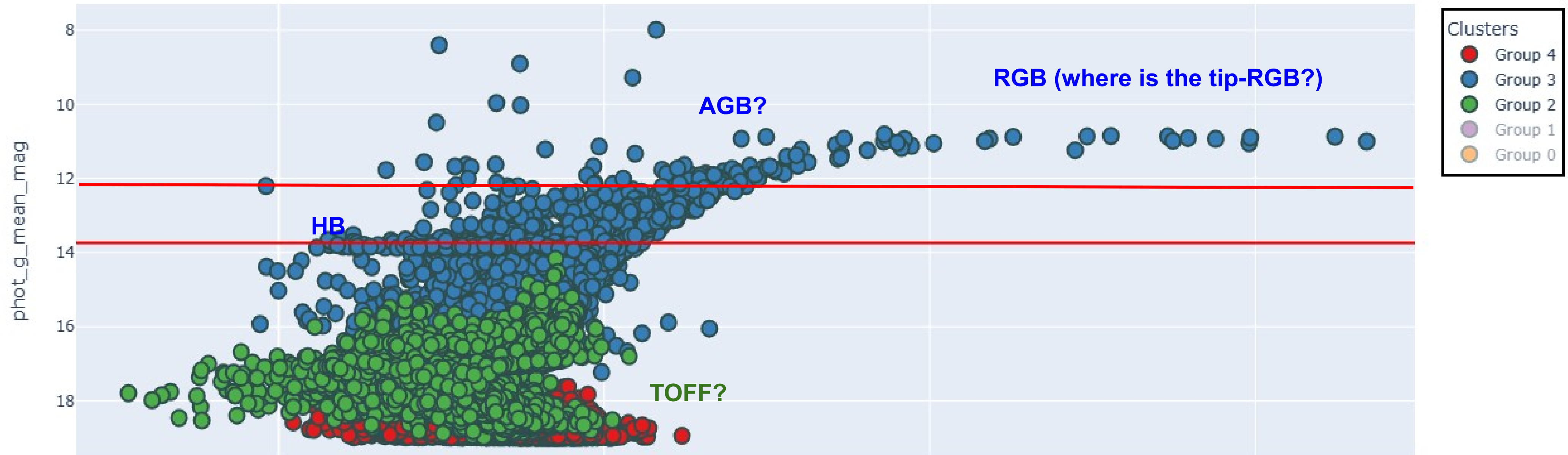
G1 and G2 have somewhat elongated necks (maybe stars which are entering their turn off point?)



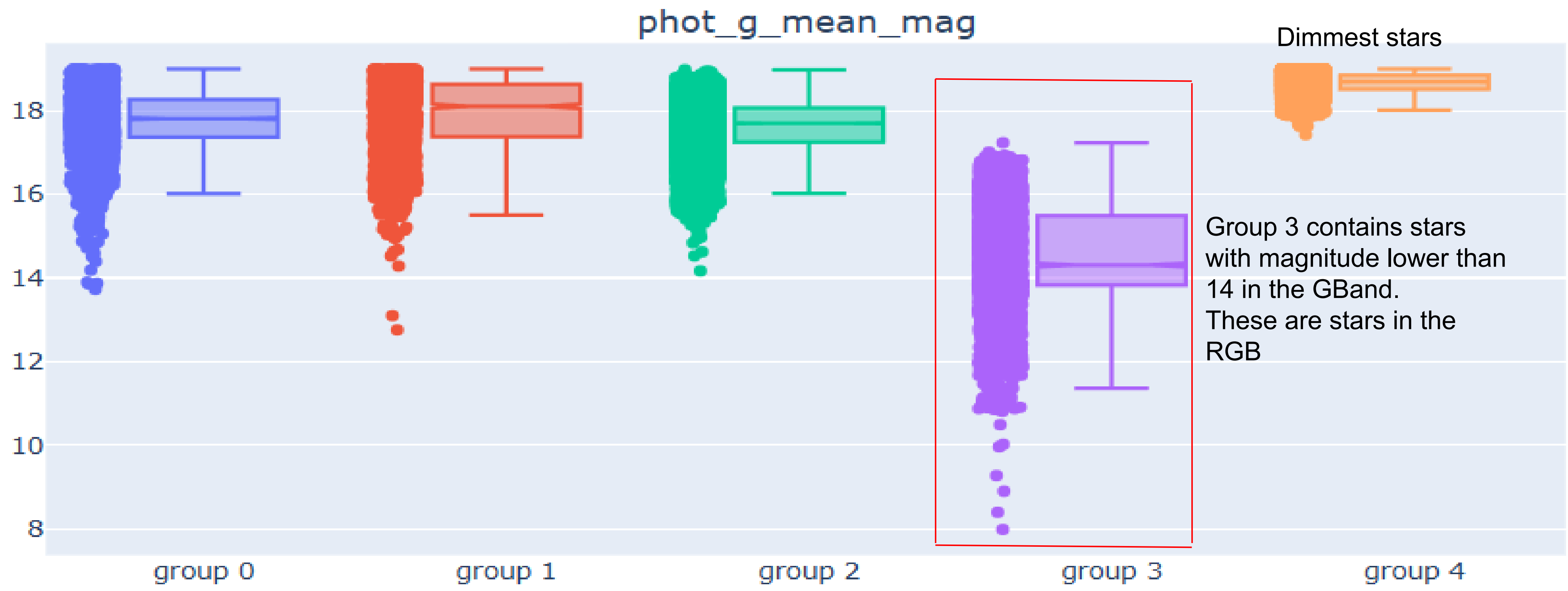
Two groups have stars with luminosity above G=13.78 (see below)

G3 has the brightest stars: sub giants, reg giants (first and second phase) and, also, stars in the horizontal branch (whose position appears to coincide with the, previously, seen peak in the G Band histogram

G2 has stars that are beginning their ascend to the first RGB, also (see the curve) appears to contain TOFF stars

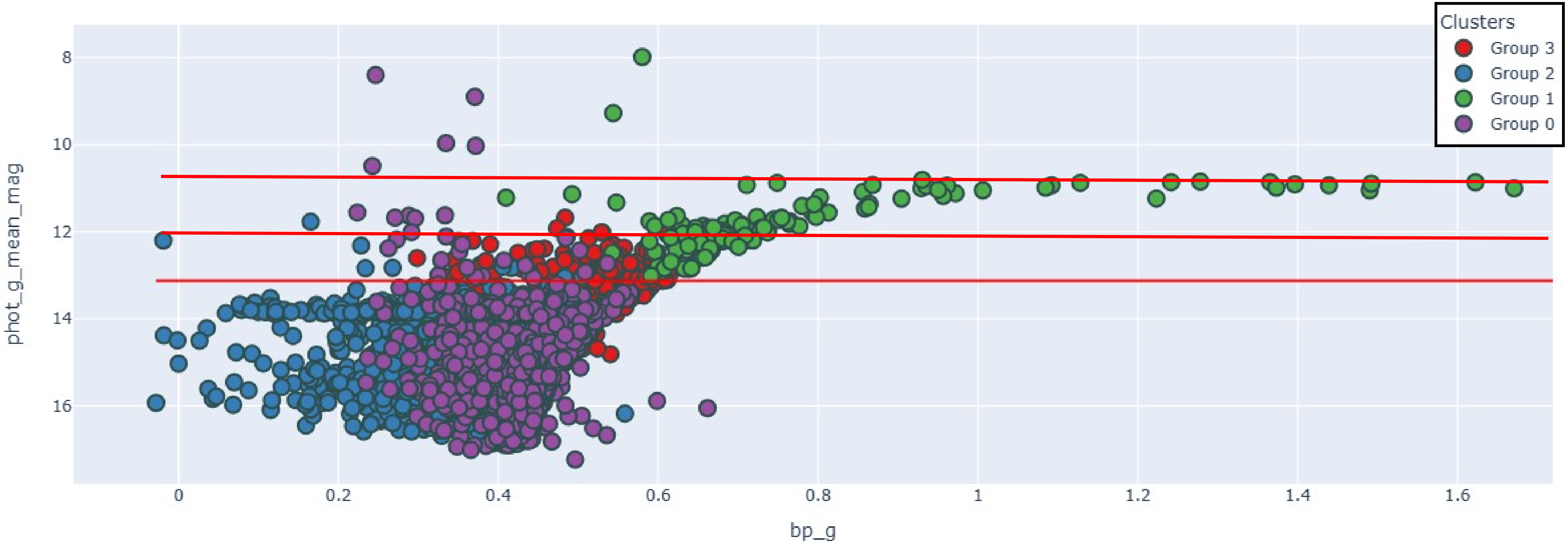


We now present the box plots for each of the star groups in the CMD. The idea is to analyze what makes each group to be different

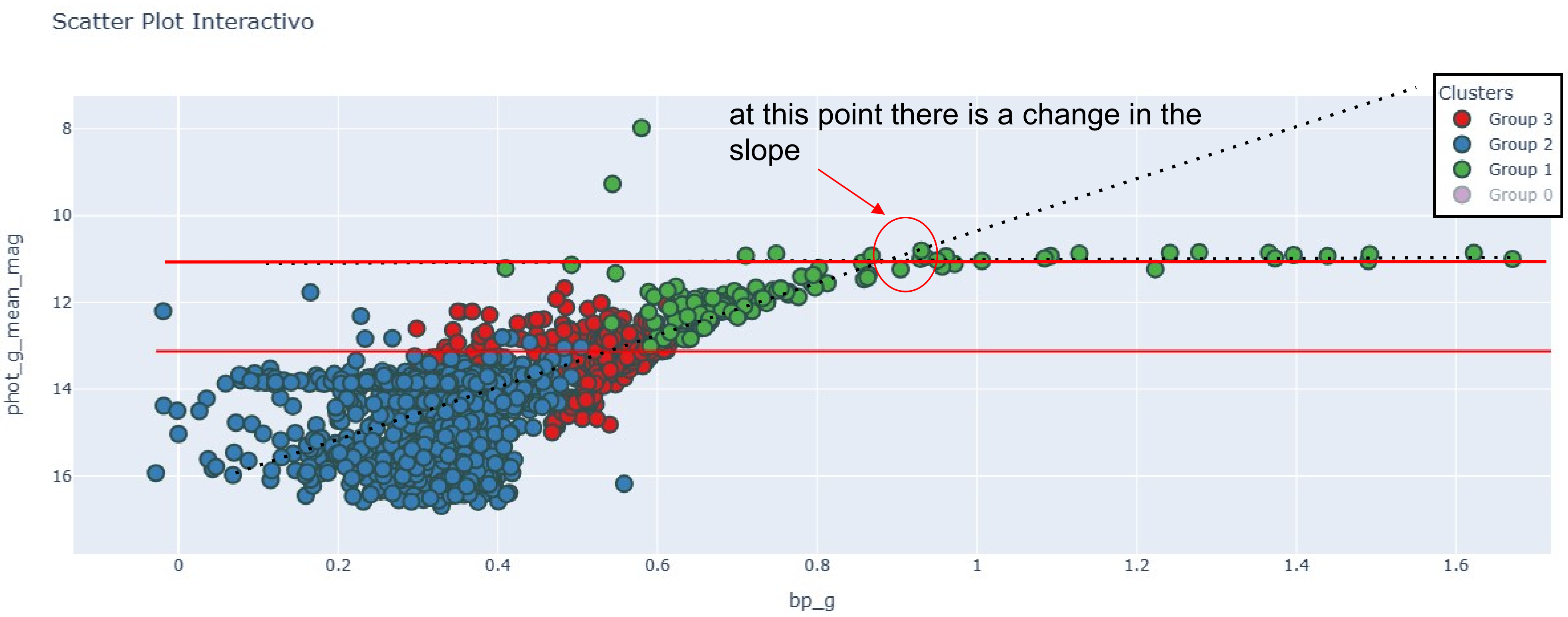


Group 3 contains stars with magnitude lower than 14 in the GBand. These are stars in the RGB

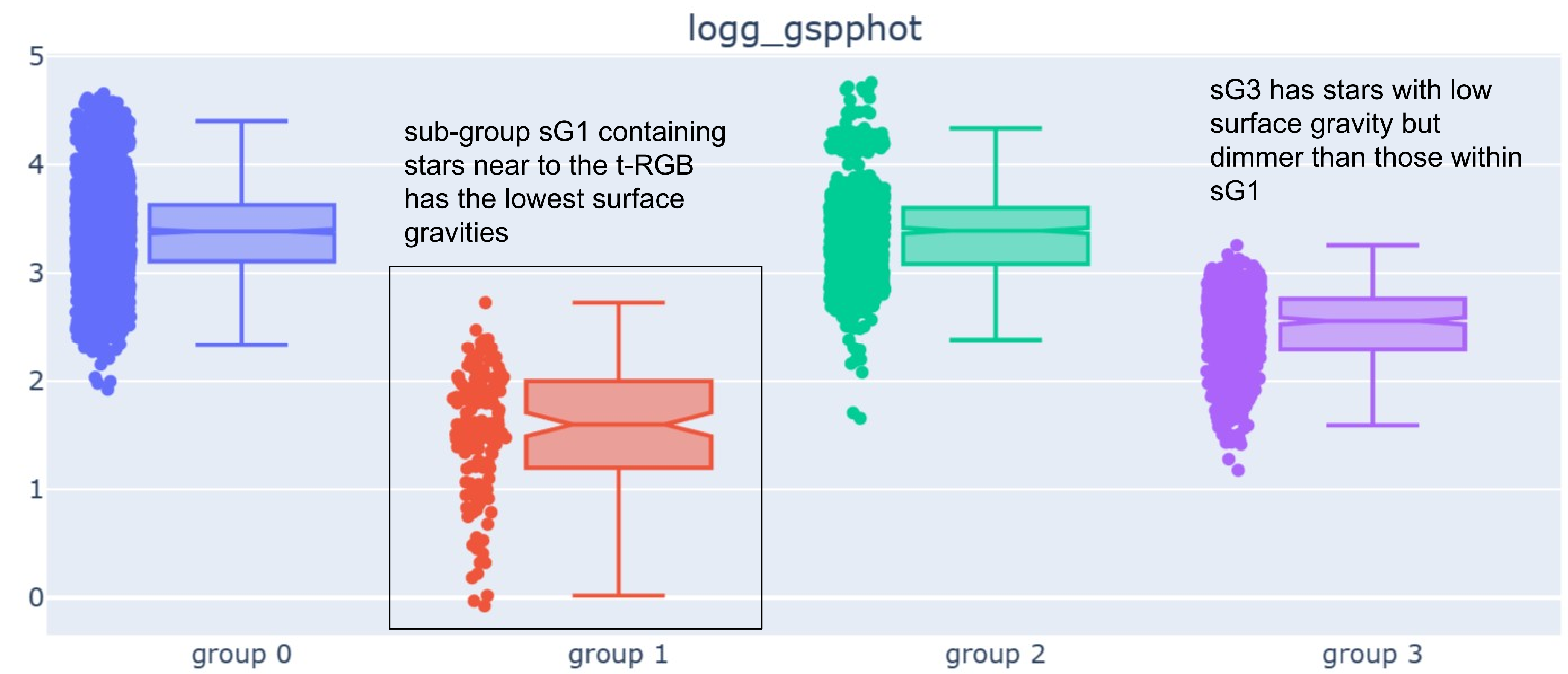
We apply the hierarchical clustering to the stars within Group 3: this time four sub groups are formed.

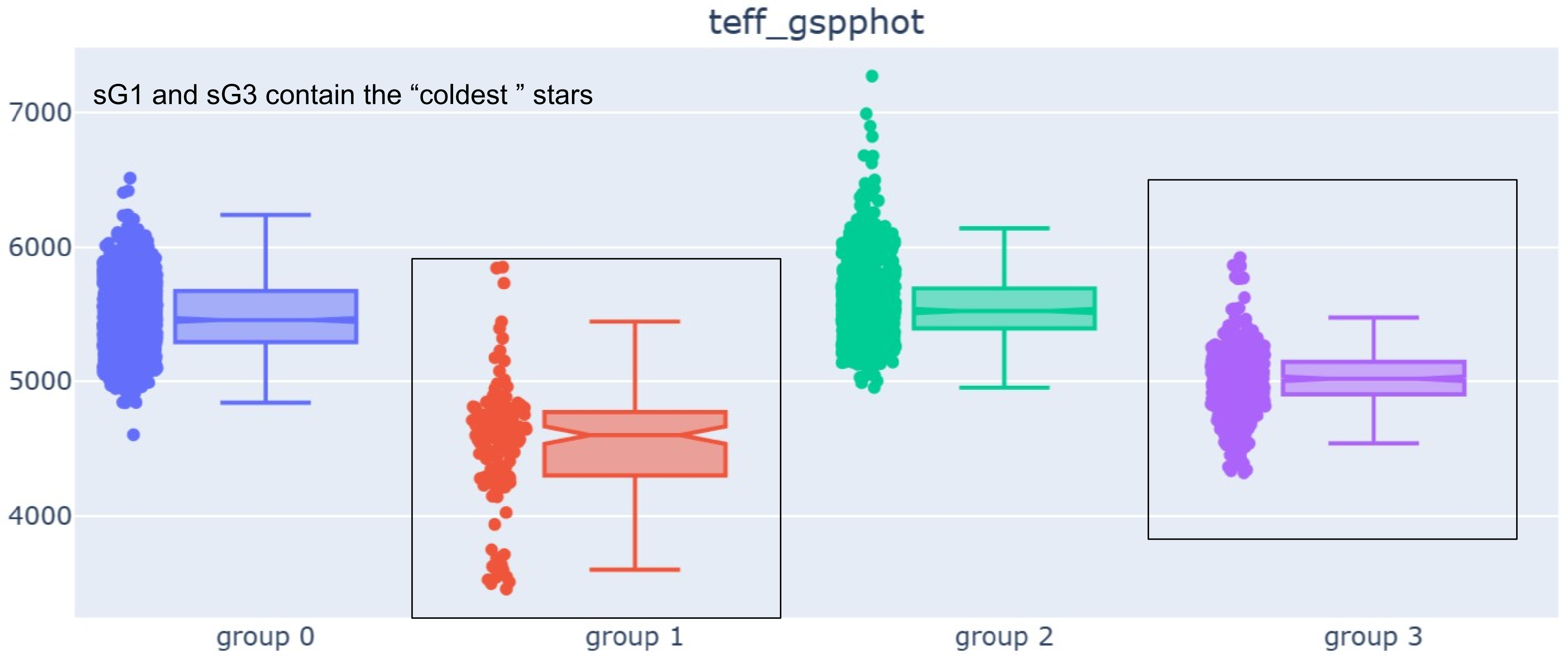


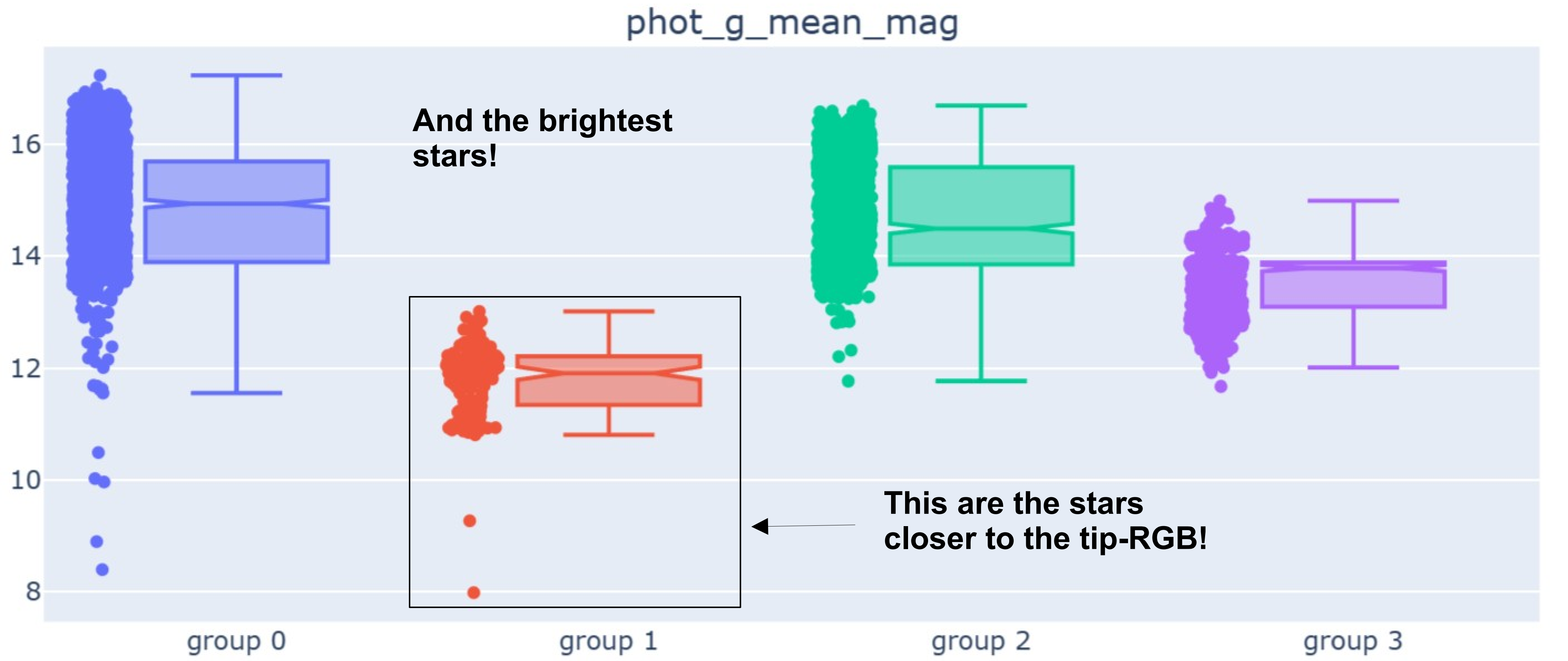
The sub groups 1, 2 and 3 have stars in what could be the sub-giant and giant branch. SG1 has AGB stars



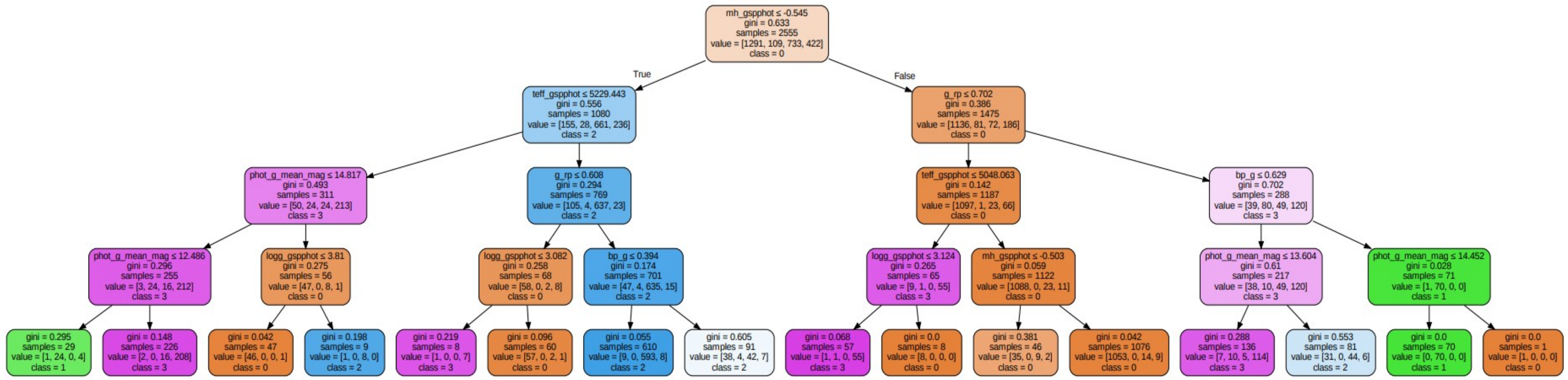
box plots for each of the sub groups within G3




teff_gspphot
sG1 and sG3 contain the “coldest ” stars
7000
6000
5000
4000
group 0
group 1
group 2
group 3


phot_g_mean_mag
16
14
12
10
8
And the brightest
stars!
This are the stars
closer to the tip-RGB!
group 0
group 1
group 2
group 3

Decision Tree explaining the conditions that allow to classify any star as a member of the subgroups within G3



Confusion matrix

```
[[286   0  15   7]
 [  0  22   1   0]
 [  8   0 192   5]
 [  1   1   6  95]]
```

Now, interpreting the values:

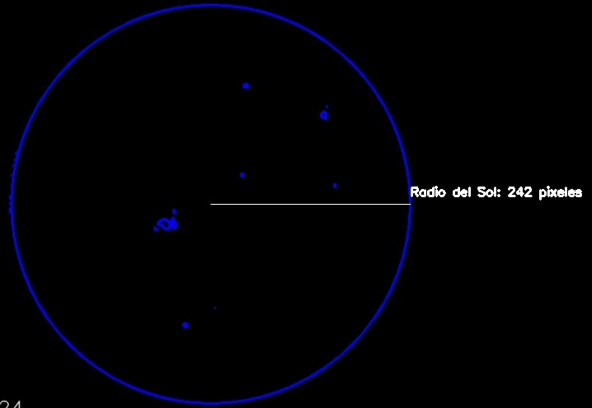
The **diagonal elements** (top-left to bottom-right) represent the number of **correct predictions for each class.**
Off-diagonal elements represent incorrect predictions.
So, looking at some specific values:

> The model predicted **286 instances of class 0 correctly,** but misclassified 15 instances of class 0 as class 2, and 7 instances as class 3.

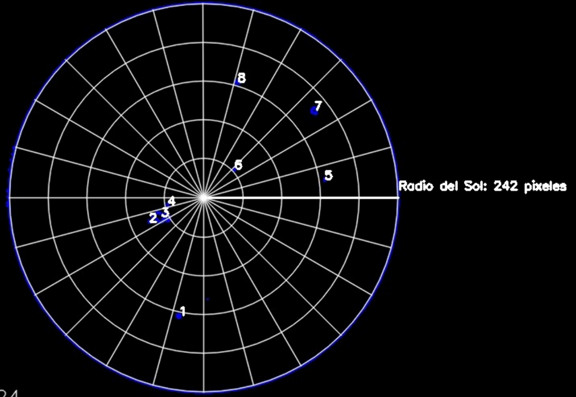
Other work in progress:

Sunspot observation and counting

Fecha: 13–08–2024
Hora: 12:00
Lugar: Colima, Mexico
Autor: Santiago Arceo Diaz



Fecha: 13–08–2024
Hora: 12:00
Lugar: Colima, Mexico
Autor: Santiago Arceo Diaz



Fecha: 13–08–2024
Hora: 12:00
Lugar: Colima, Mexico
Autor: Santiago Arceo Diaz

Thank you for you attention!

Contact: santiagoarceodiaz@gmail.com