

CTA 200 Final Project

Maëlle Magnan

Email: maelle.magnan@mail.utoronto.ca

Student ID: 1008798598

May 20th, 2025

Introduction

For this mini project, we explored galaxy simulations by SQL querying and using APIs to access astrophysical datasets. We began by querying the EAGLE simulation to retrieve key galaxy properties such as stellar mass and star formation rate. First we focused on galaxies at redshift $z = 0$, then expanding to galaxies within $0 < z < 0.5$. Later, we performed a second query which included additional photometric features, such as g - and r -band magnitudes. This allowed us to incorporate galaxy colours into our analysis. We then examined the IllustrisTNG simulation via an API. By analyzing results from both simulations, we compared the star-forming main sequences across different models. Finally, we applied selection criteria to identify galaxies with Milky Way-like properties across the datasets.

1 EAGLE redshift zero query

The first query retrieved the star formation rates and stellar masses of galaxies from the `RefL0100N1504 SubHalo` table of the EAGLE database. We limited our search using the `SnapNum` column, setting this property to 28 which corresponds to a redshift of $z = 0$. The same constraint could have been imposed using the redshift column and setting to zero. Using the obtained characteristics, we plotted the star formation rate versus stellar mass, along with a 2D histogram of the same data, colour-coded by galaxy count, as shown in Fig. 1.

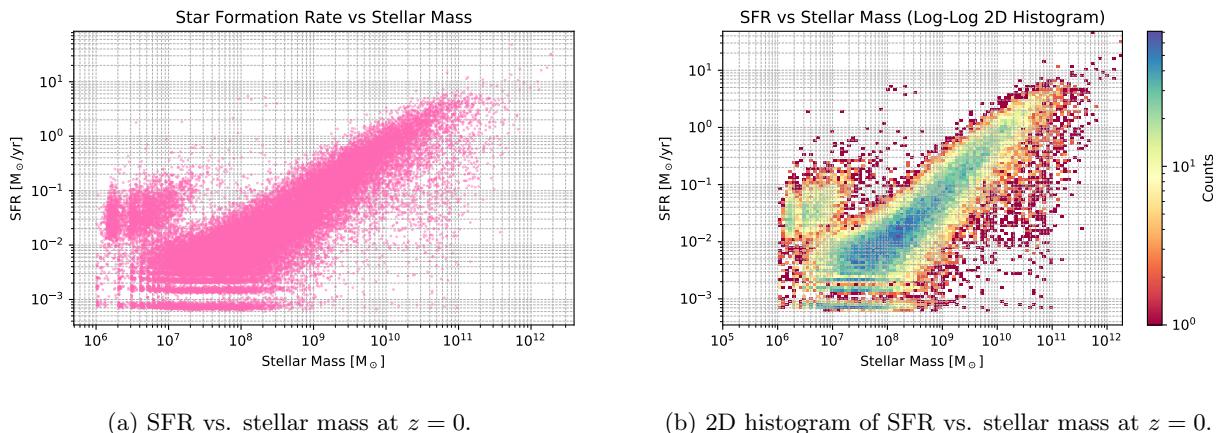


Figure 1: Comparison of scatter plot and 2D histogram of star formation rate vs. stellar mass in the EAGLE simulation at redshift $z = 0$.

Figure 1a reveals a clear star-forming main sequence where galaxies follow a tight, linear correlation between star formation rate and stellar mass. This main sequence closely matches the expected result from real world observations and demonstrates that in general, more massive galaxies have higher star formation rate. However, the figure also displays a drop off at higher stellar masses indicating that this relationship holds up until a certain cut off. We also observe a scattered population of galaxies with high stellar mass but low star formation rates, consistent with the observationally identified ‘red and dead’ galaxies.

The plot also shows some unexpected features which do not agree with current observational data. Firstly, there is an excess of low-mass galaxies with high SFRs. These are most likely caused by limited resolution which can lead to unreliable modelling of small halos. Second, there are sharp horizontal and vertical bands at lower stellar masses and star formation rates. These bands are presumably due to artificial cutoffs or feedback implementations within the simulation code. These discrepancies arise in part because the simulation uses simplified sub-grid models to approximate complex physical processes, allowing it to run efficiently over large cosmological volumes at the cost of some physical accuracy.

1.1 Stretch Goal: Coloured Galaxies

As an extra step, we ran a second query of the same galaxies with redshift $z = 0$ including additionally galactic properties. We extracted both the g and r band magnitudes then plotted the Mass-SFR distribution coloured by the g-r colour. The result is shown in 2.

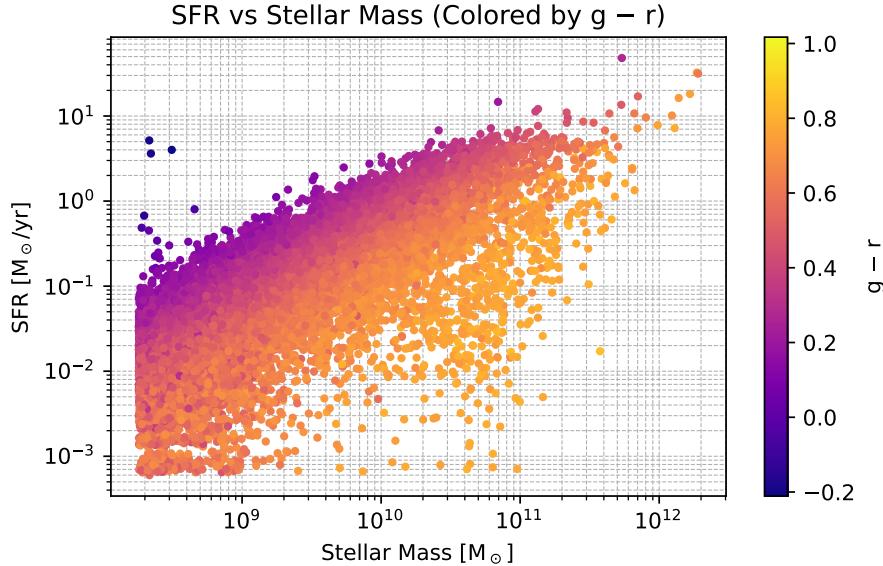


Figure 2: Star formation rate vs. stellar mass for galaxies in the EAGLE simulation, coloured by $g - r$ colour index.

In the coloured distribution, galaxies with bluer colours (lower $g - r$) are located in the star-forming region, while redder galaxies (higher $g - r$) tend to lie on the quiescent sequence. This colour difference reflects the stellar populations. Blue galaxies contain a higher fraction of young, massive stars, which are short-lived and emit strongly in the blue. Red galaxies are dominated by older, lower-mass stars, which are cooler and redder indicating that active star formation has mostly ceased.

2 More Properties and Snapshots

In section one our query only searched for galaxies at redshift $z = 0$. To examine more snapshots, we ran another query, this time searching for galaxies in the redshift range of $0 < z < 0.5$. We then plotted the same Mass-SFR diagram as shown in Fig. 3. As expected increasing the redshift range outputs a larger amount of galaxies. In this case, the first query at $z=0$ gave 2 275 510 galaxies whereas the second query with $0 < z < 0.5$ gave 11 819 412 galaxies. Compared to the subset at $z = 0$, the full galaxy sample shows a broader spread in star formation rates, especially at lower stellar masses, indicating a mix of quenched and actively star-forming galaxies. The star-forming main sequence is also more dispersed in the full sample, while the $z = 0$ galaxies exhibit a tighter correlation between stellar mass and SFR. This suggests that galaxy populations at different redshifts contribute to greater diversity in star formation activity.

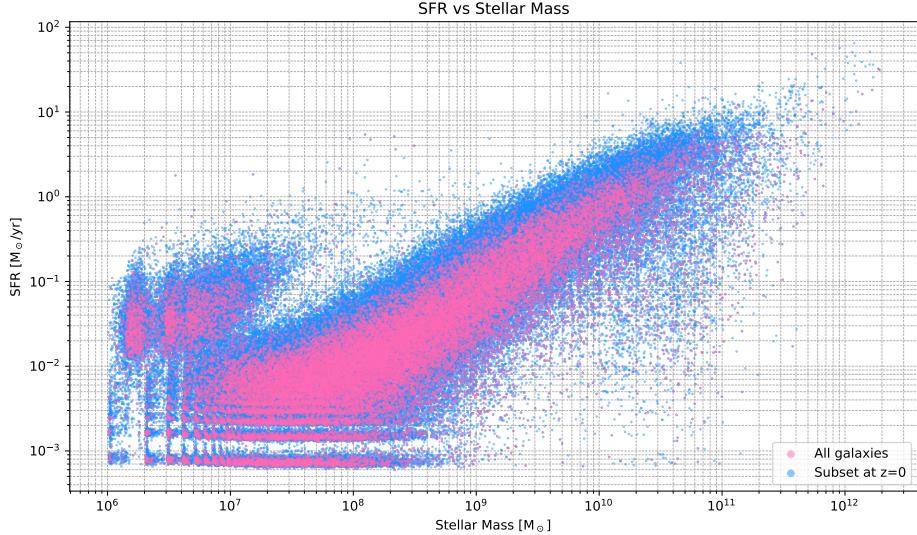


Figure 3: Scatter plot of star formation rate vs. stellar mass in the EAGLE simulation at redshift $z = 0$ and for redshift range $0 < z < 0.5$.

Selecting galaxies across a range of redshifts is more representative of what we observe in galaxy surveys since observational data captures light from galaxies at various distances and therefore times. The maximum redshift a survey can reach is mainly determined by the sensitivity of the telescope (its flux limit) and the brightness of the galaxies being observed. More sensitive instruments can detect fainter, and thus more distant, galaxies.

2.1 Stretch Goal

The EAGLE simulation outputs a tables containing many more elements than just stellar mass and star formation rate. As such, we can determine the most massive galaxy at a certain redshift and subsequently trace its descendants. In order to find the most massive galaxy at a redshift of $z = 0$, we used numpy’s where function. Then using the GalaxyID, TopLeafID, and LastProgID of the most massive galaxy, we queried between the GalaxyId and the LastProgID which gave us the number of total descendants. We then constructed the main branch of a galaxy by iteratively following each DescendantID starting from the top leaf, using a dictionary for efficient lookup, and stopping when no further descendant exists or a loop is detected.

Property	Value
GalaxyID of the most massive galaxy	21379521
Total number of descendant galaxies	194065
Number of main branch descendants	27

Table 1: Summary of the most massive galaxy and its descendant structure.

3 Querying Data from IllustrisTNG

We also examined a second simulation named IllustrisTNG using an API key instead of SQL. We first queried for galaxies at redshift $z = 0$ and found that there were 53 939 galaxies in comparison to EAGLE’s 2 275 510. We then plotted these galaxies over the EAGLE simulation to compare in Fig. 4.

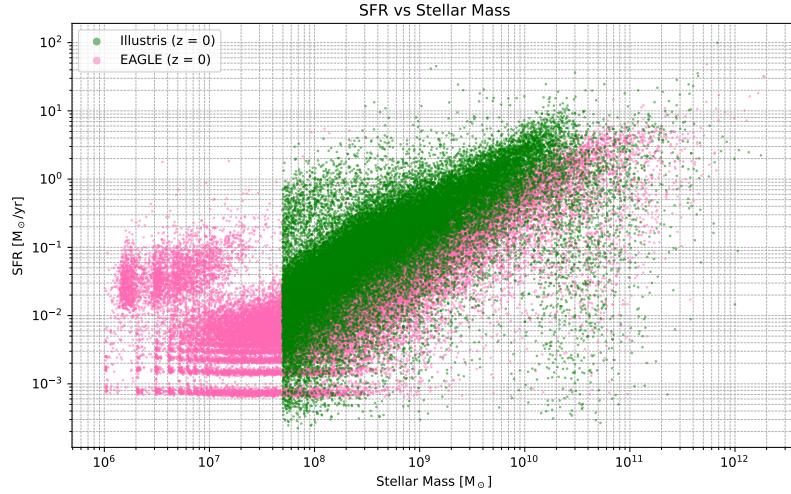


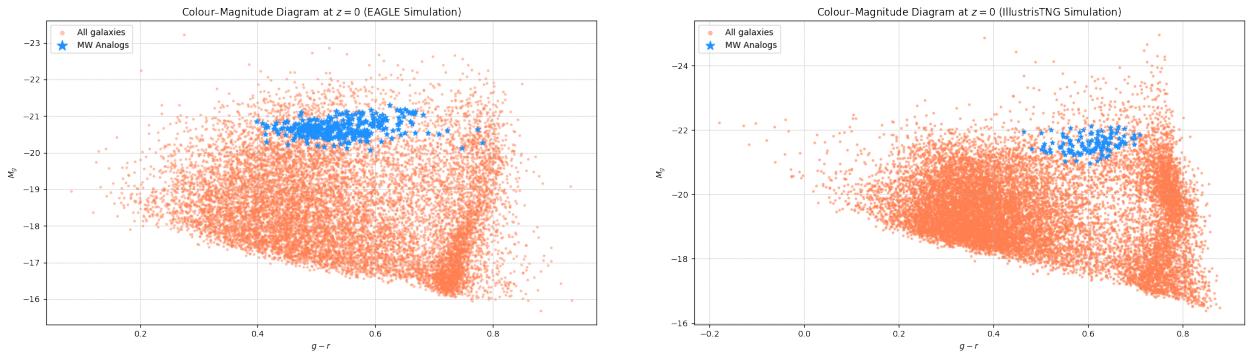
Figure 4: Comparison of the star formation rate versus stellar mass relation from the Illustris and EAGLE simulations. While both simulations show a star-forming main sequence, the EAGLE simulation features a more distinct population of quenched, high-mass galaxies (red and dead).

The plot shows noticeable differences between the EAGLE and Illustris simulations. In Illustris ($z = 0$), there are fewer galaxies with very low star formation rates at high stellar masses, whereas EAGLE contains a more prominent population of these quenched, 'red and dead' galaxies. EAGLE appears to produce a more realistic quenched population consistent with observations of early-type galaxies. Additionally, Illustris does not simulate galaxies with lower stellar mass.

4 Isolating Milky Way-like galaxies in the galaxy colour-magnitude diagram

To identify Milky Way analogs in the simulation, we first selected all central galaxies at redshift $z = 0$ with stellar mass $M_\star > 10^9 M_\odot$. From this sample, we isolated galaxies with stellar masses and star formation rates within three standard deviations of current Milky Way estimates: $M_\star = (6.08 \pm 1.14) \times 10^{10} M_\odot$ and $\dot{M}_\star = 1.65 \pm 0.19 M_\odot \text{ yr}^{-1}$. These selection criteria yielded a subset of galaxies with properties similar to the Milky Way. We then plotted a colour-magnitude diagram in g versus $g - r$, highlighting the Milky Way analogs as shown in Fig. 5

The resulting distributions show that Milky Way analogs occupy a relatively narrow region in colour-magnitude space, suggesting that the Milky Way is not an outlier but rather a typical star-forming galaxy in terms of its stellar mass and colour. This implies that the Milky Way lies along the galaxy main sequence and reinforces the idea that its properties are representative of normal disk galaxies within large-scale simulations.



(a) Scatter plot of star formation rate versus stellar mass at $z = 0$ from the EAGLE simulation. Milky Way analogs are shown in blue.

(b) Scatter plot of star formation rate versus stellar mass at $z = 0$ from the Illustris simulation. Milky Way analogs are shown in blue.

Figure 5: Comparison of Milky Way analogs in the EAGLE and Illustris simulations at redshift $z = 0$. Both panels show star formation rate versus stellar mass, highlighting how the Milky Way occupies a narrow region of the main sequence.