

# 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.

## Section One: EAGLE redshift zero query

Our first query retrieved the star formation rates and stellar masses of galaxies from the EAGLE simulation using the `RefL0100N1504.SubHalo` table. We constrained our query to galaxies with `SnapNum = 28`, which corresponds to a redshift of  $z = 0$ . This constraint could have also been applied using the `Redshift` column. We then created a scatter plot of star formation rate versus stellar mass and generated a 2D histogram of the same data, colour-coded by galaxy counts as seen in Fig. 1.

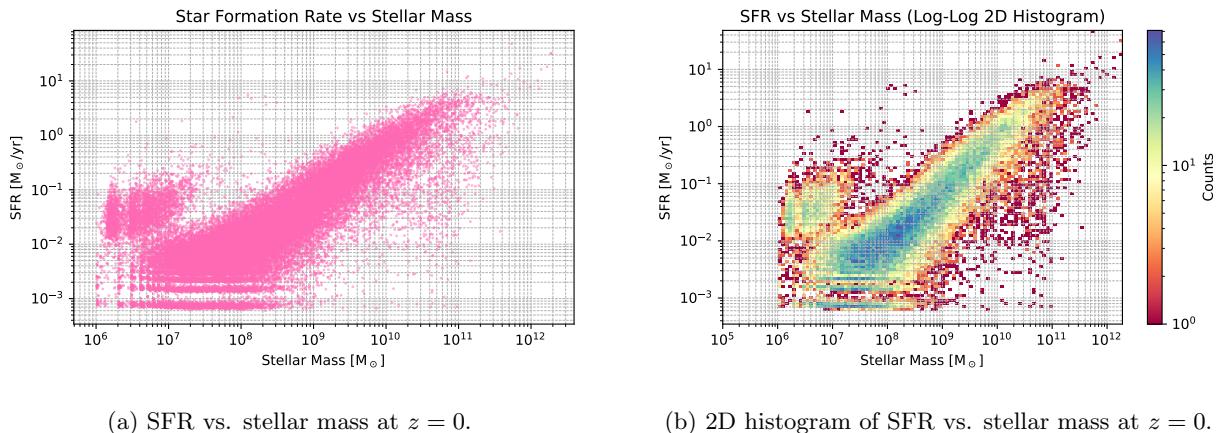


Figure 1: Comparison of scatter plot and 2D histogram of star formation rate vs. stellar mass in the EAGLE simulation at redshift  $z = 0$ .

The figure reveals a clear star-forming main sequence, where galaxies show a tight, almost linear, correlation between stellar mass and star formation rate. This feature reflects the fact that in general more massive galaxies tend to have higher star formation rates. We see a drop off at higher stellar masses which indicates that this relation holds until a certain threshold as seen observationally. The plot also shows some unexpected features such as sharp horizontal and vertical bands at lower stellar masses and star formation rates. These do not appear in observational data and indicate that galaxies below a certain mass or SFR threshold are poorly resolved. Lower resolutions most likely allow the simulations to cover larger ranges of SFR and stellar mass however clearly have their limits. Additionally, the sharp lower bound in SFR may reflect artificial cutoffs or feedback implementations within the simulation code.

## Section Two: More Properties and Snapshots

The first query only searched for redshift  $z=0$  galaxies. To examine more snapshots we changes this to  $0 < z < 0.5$  using the `Redshift` column of the EAGLE table. The new query results are plotted in Fig. 2. We clearly see that adding a larger range of redshifts increases the number of galaxies plotted as expected.

In this case, the first query at  $z=0$  gave 2 275 510 galaxies whereas the second query up to  $z=0.5$  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.

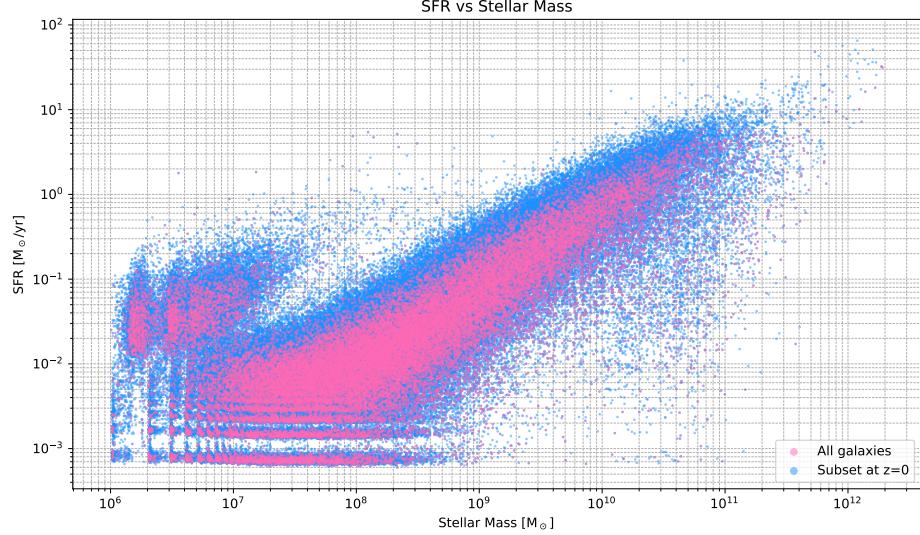


Figure 2: The Lorenz system  $Y(t)$  solution as a function of time for initial condition  $[0, 1, 0]$ .

Selecting galaxies across a range of redshifts, rather than just at  $z = 0$ , is more representative of what we observe in galaxy surveys because such surveys capture light from galaxies at various distances and cosmic 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.

### Section Three: 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. 3.

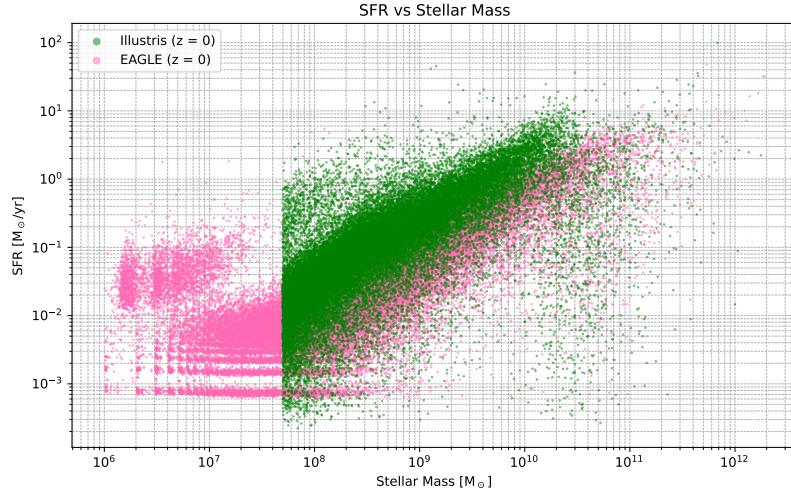


Figure 3: Comparison of the star formation rate versus stellar mass relation from the Illustris and EAGLE simulations. While both simulations show a star-forming main sequence, the EAGLE simulation features a more distinct population of quenched, high-mass galaxies (red and dead).

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.