# Data Science Lab Astrophysics 2024:
# Lab Worksheet 2: Density Estimation

## Lab Project 2: Density Estimators applied to SDSS data

We will again work withe SDSS dataset from worksheet 1.

**Task 1.** Select a somewhat thinner slice in redshift than for the previous exercises, thus reducing the number of galaxies. Make sure you see nice large scale structure in the angular map.

**Task 2.** Use multiple density estimators to compute the angular density of galaxies in RA-DEC space. Show 2D-histograms, kernel density estimates, nearest neighbour density and Voronoi-based estimates as images, each for two different bandwidth/neighbor parameters. [Hint: if one of the estimators runs too long, just reduce the number of galaxies by using a thinner slice. ]

**Task 3.** For the KDE maps, show the one-point statistics $p(\rho)$, i.e. a histogram of the density values. Create one for 'volume weighted statistics', where the test points are uniformly distributed in space, and one for 'mass weighted statistics', where the test points are points drawn from the distribution itself (i.e. the density at the galaxy positions). Determine using plots if/how the degree of Gaussianity of $p(\rho)$ depends on the bandwidth.

**Task 4.** Use again the split into red and blue galaxies from project 1 and show the conditional density estimates for the two populations (i.e. 'mass weighted'). In order to improve the statistics, generate multiple independent slices for which you estimate the density. Are they different or consistent with each other? Use the frequentist Kolmogorov-Smirnov test (read up on it yourself) to test the null hypothesis that the samples are drawn from the same distribution. This test is implemented in `scipy.stats.ks_2samp`.

**Task 5.** Perform the reverse test from 4. Split the galaxies into those in the highest and lowest quartile of the density distribution. Show the distribution of their colors. Again test whether the two samples are consistent with being drawn from the same distribution using the Kolmogorov-Smirnov test.