Comparison of Catalogs

Table of contents

The data

In this script we will compare 2 catalogs Kovlakas et al. (2021) and Karachentsev and Kaisina (2013)

- The data have been joined based on their position in the sky (Ra, Dec, Distance).
- We use TOPCAT to create two joins, an inner and an outer join
- We will use the inner join for 1-1 comparisons
- If we see that the data are similar we can use the outer join

The dataset we are going to use consists of 296 galaxies and 168 columns.

How are we going to compare the data?

Scatter plots and R^2 calculation

1. R^2: Measures the proportion of variance explained by the linear model.
2. Slope of the Fitted Line: Should be close to 1 for a 1-1 correlation.[1]

3. Pearson Correlation ρ: Measures the strength and direction of the linear relationship between two variables, ranging from -1 to 1. [2]

4. Plots: Plots are essential for visually assessing the relationship between two datasets, identifying correlations, trends, and outliers, and evaluating the fit of linear models.

- Histograms: Because not all of our data have the same number of counts, the comparison with histograms between data that are not the same, doesn't help us right now.[3] This is why we will only use histograms for comparing the distribution of same-data columns normalized by their maximum value

[1] Some data seem to have a very good linear correlation but they have many outliers. This is why we will clip the outliers with $\sigma > 3$
[2] In simple linear regression, R^2 is the square of the Pearson correlation coefficient ρ.
[3] When we will use the outer join table we could use histograms due to the large number of counts.

- Correlation Heatmaps: A correlation heatmap is a graphical tool that displays the correlation between multiple variables as a color-coded matrix. It's like a color chart that shows us how closely related different variables are.

  In a correlation heatmap, each variable is represented by a row and a column, and the cells show the correlation between them. The color of each cell represents the strength and direction of the correlation, with darker colors indicating stronger correlations.

- Kernel Density Estimate (KDE) plot: The KDE plot visually represents the distribution of data, providing insights into its shape, central tendency, and spread.

5. Percentage change: We can calculate the percentage change of the data for each galaxy and then we can see if the data are similar, based on minimum, the maximum and the mean value of the difference.

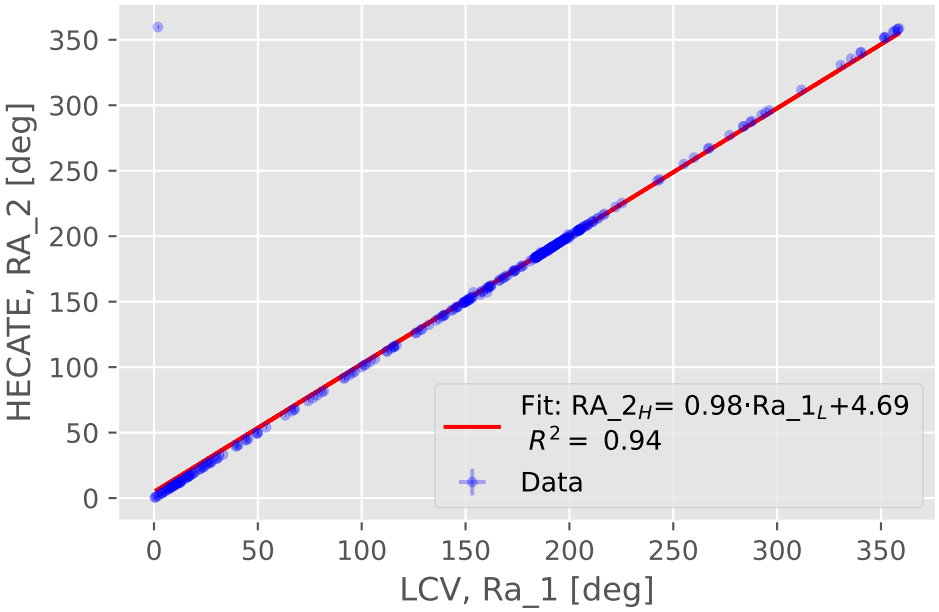
$$\text{Percentage change} = \frac{V_{Hecate} - V_{LCV}}{V_{Hecate}} \cdot 100\%$$

Comparable data

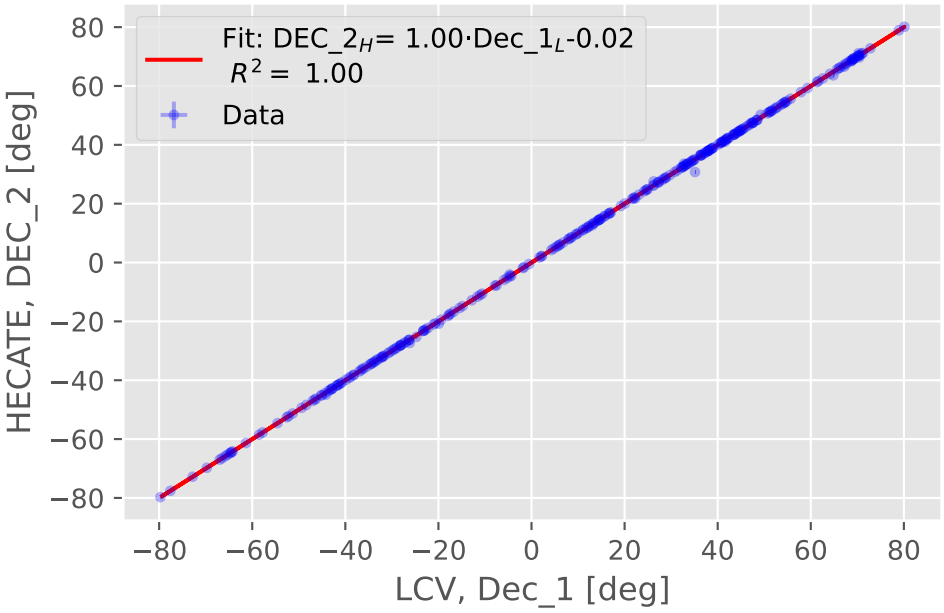
Coordinates

LCV	HECATE	Description	Pearson Correlation [-1,1]
Ra_1	RA_2	Right Ascension	0.972
Dec_1	DEC_2	Declination	1.0
Dis	D	Distance	1.0

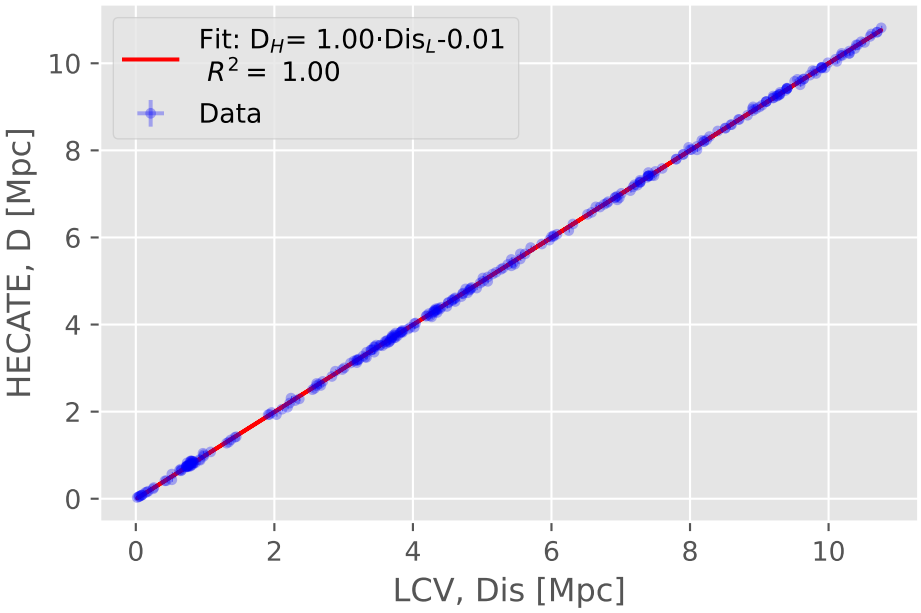
Right Ascension



Declination



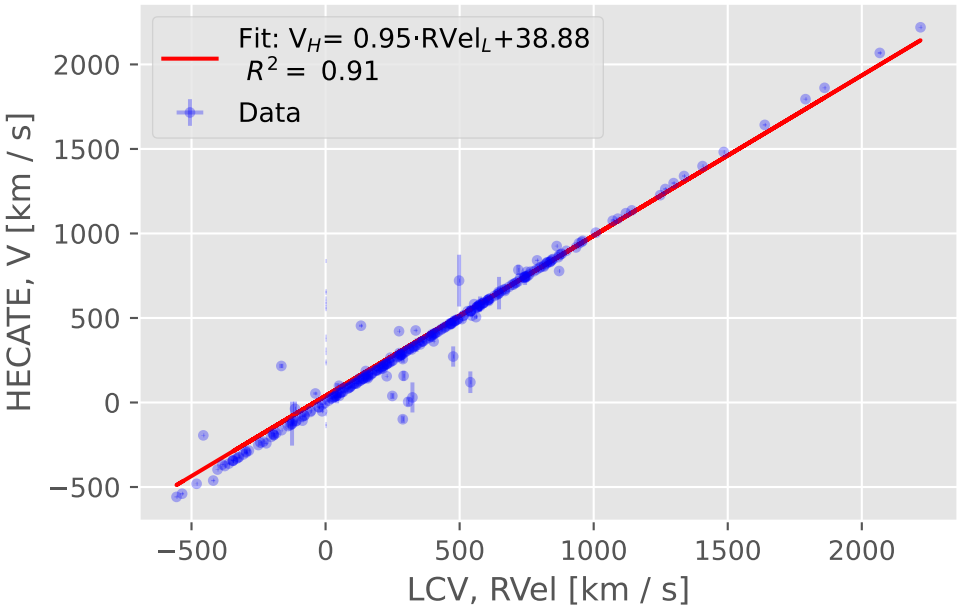
Distance



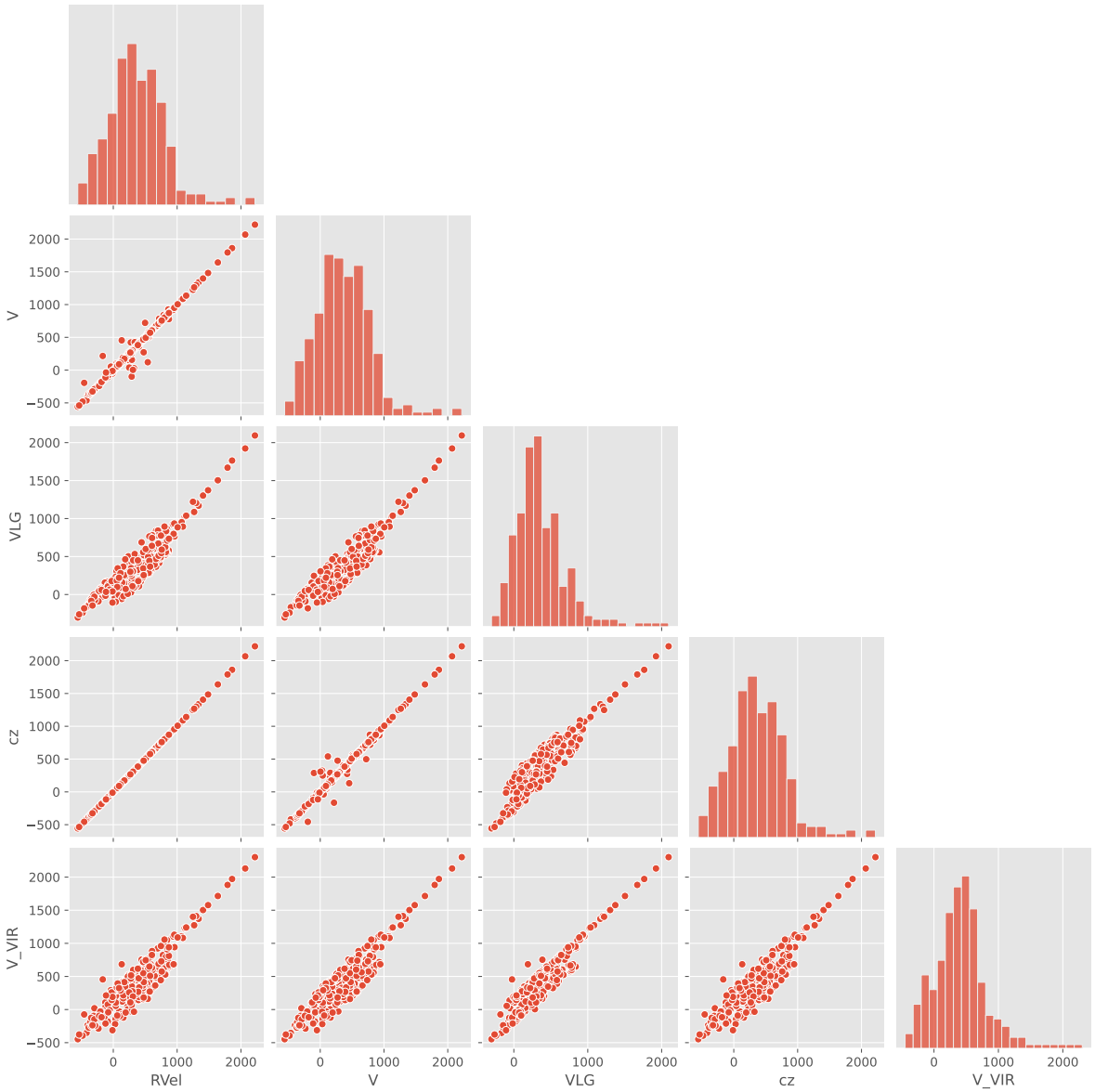
Velocities

LCV	HECATE	Description	Linear Correlation
RVel (km/s)	V (km/s)	Heliocentric radial velocity	0.952
VLG (km/s)		Radial velocity	
cz (km/s)		Heliocentric velocity	
	V_VIR (km/s)	Virgo-infall corrected radial velocity	

Heliocentric radial Velocity

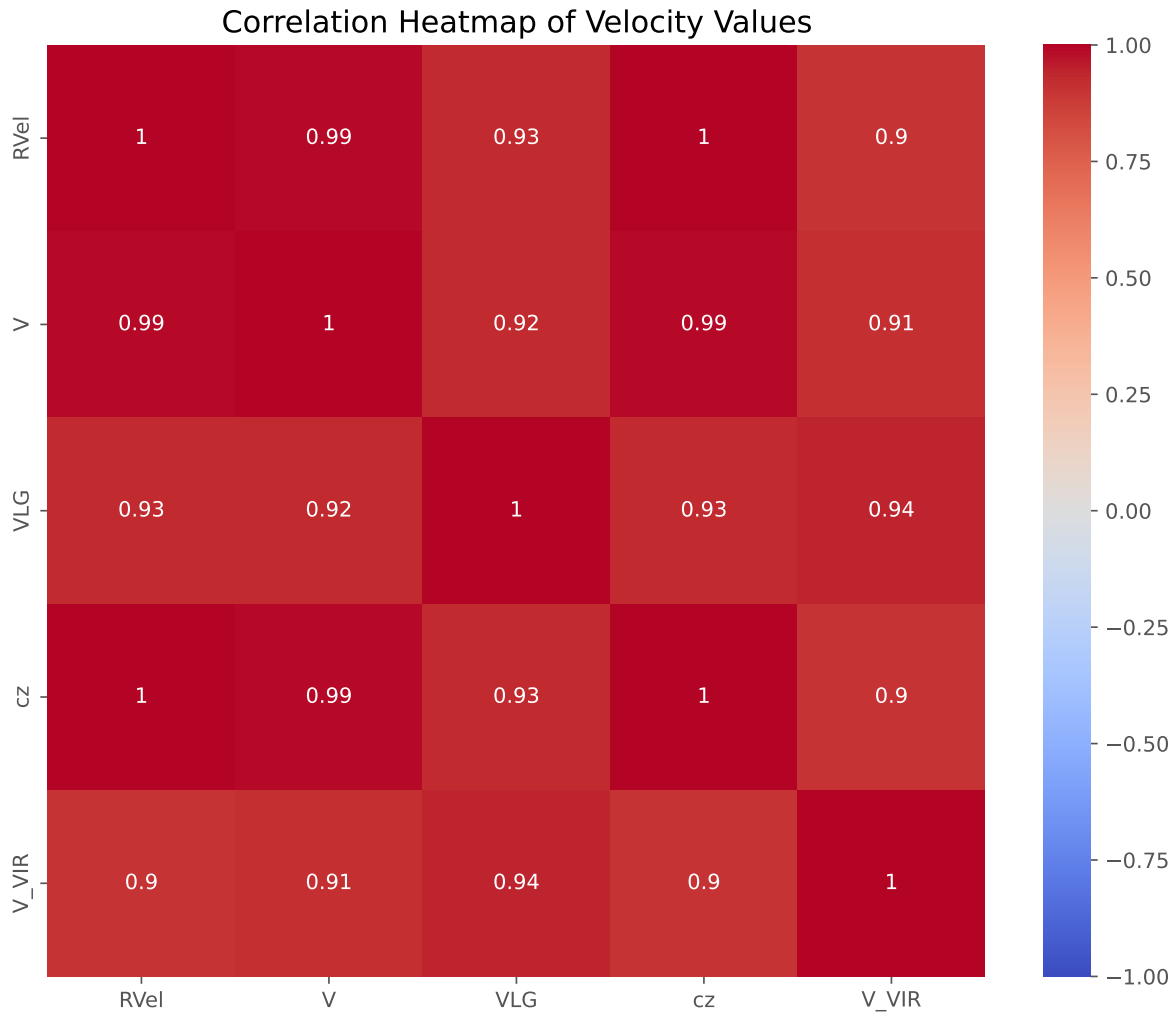


Scatter Grid



[?] The close correlation between all of the velocities, could be due to the fact that all of them measure the velocity of each galaxy, but from a different frame of reference.

Heatmap



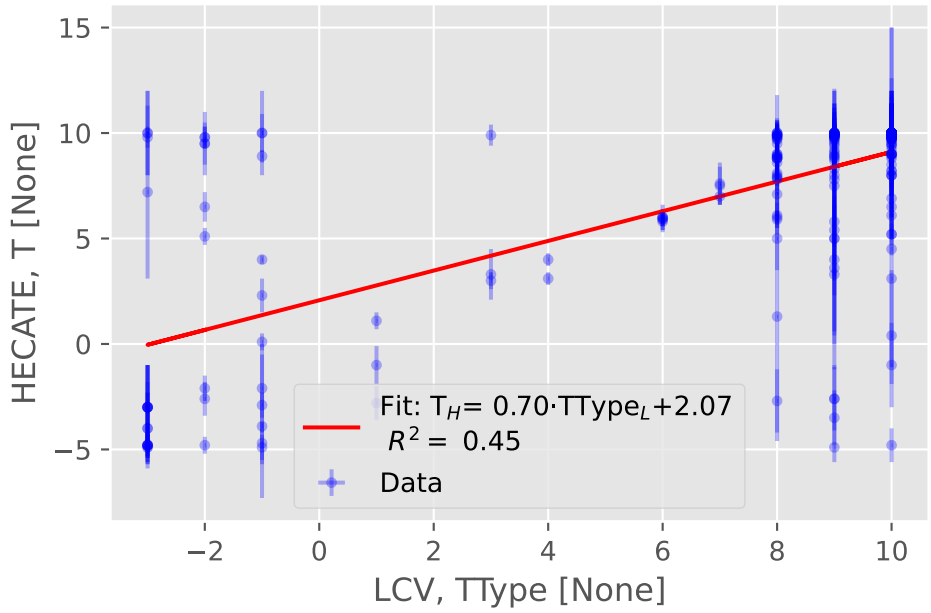
Morphology and Geometry

LCV	HECATE	Description	Pearson Correlation [-1,1]
TType	T (with errors)	Numerical Hubble type following the de Vaucouleurs system	0.6685
inc	INCL	Inclination (deg)	0

LCV	HECATE	Description	Pearson Correlation [-1,1]
a26_1 (Major)	R1 (Semi-major axis)	angular diameter (arcmin)	0.992

Galaxy Types

Hubble stage T	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
de Vaucouleurs class	cE	E	E+	S0−	S00	S0+	S0/a	Sa	Sab	Sb	Sbc	Sc	Scd	Sd
Approximate Hubble class		E			S0		S0/a	Sa	Sa-b	Sb	Sb-c		Sc	



Percentage change:

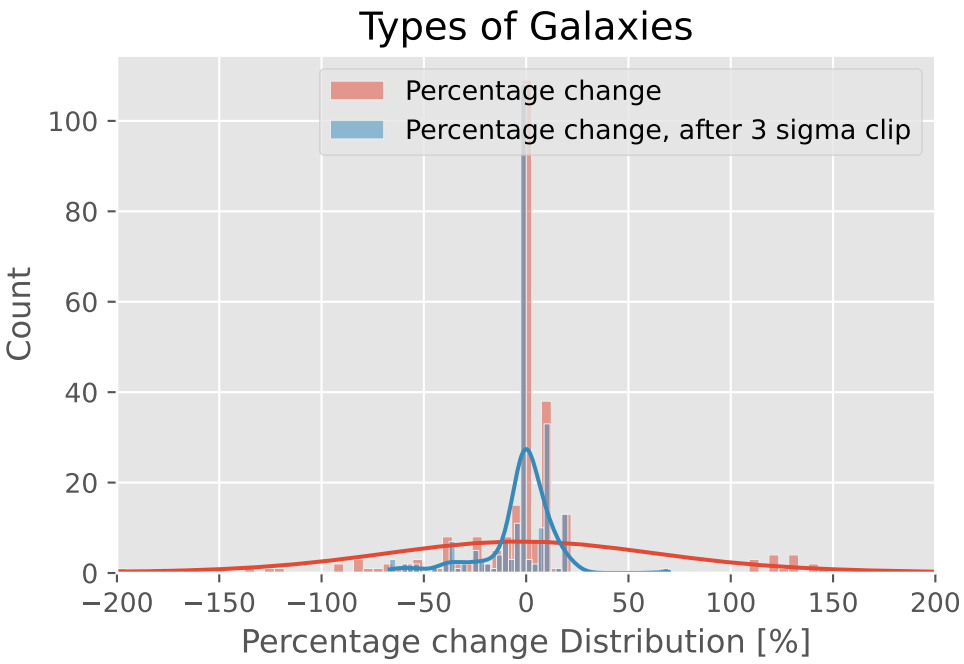
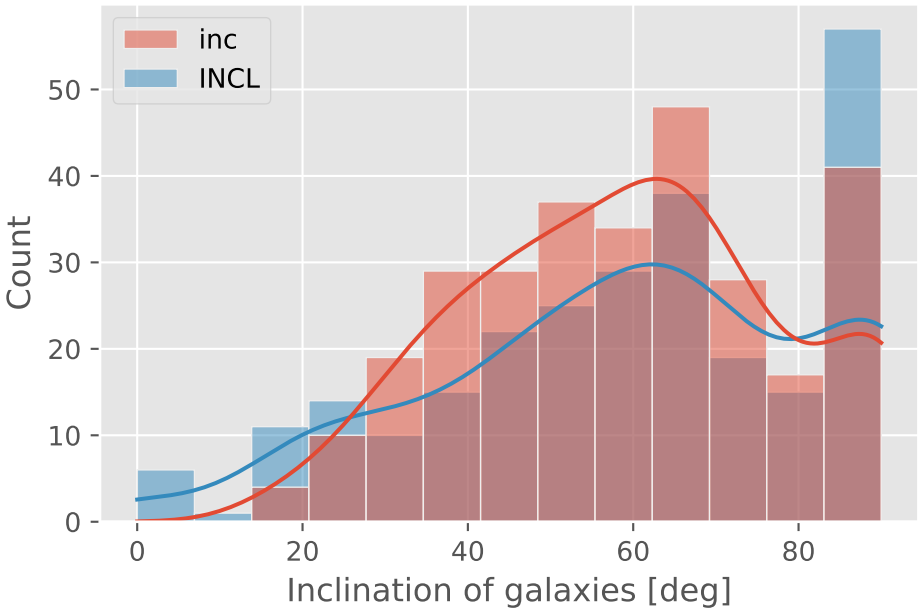
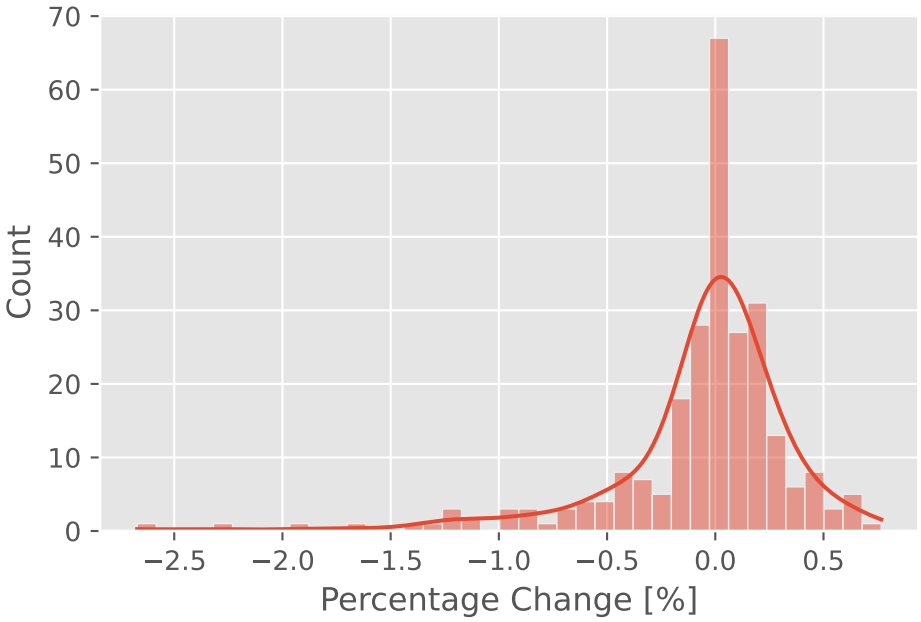


Table 5

	Percentage change	Percentage change, after 3 sigma clip
count	267	228
mean	-21	-3
std	191	17
min	-2,400	-67
50%	0	0
max	1,100	70

Inclination



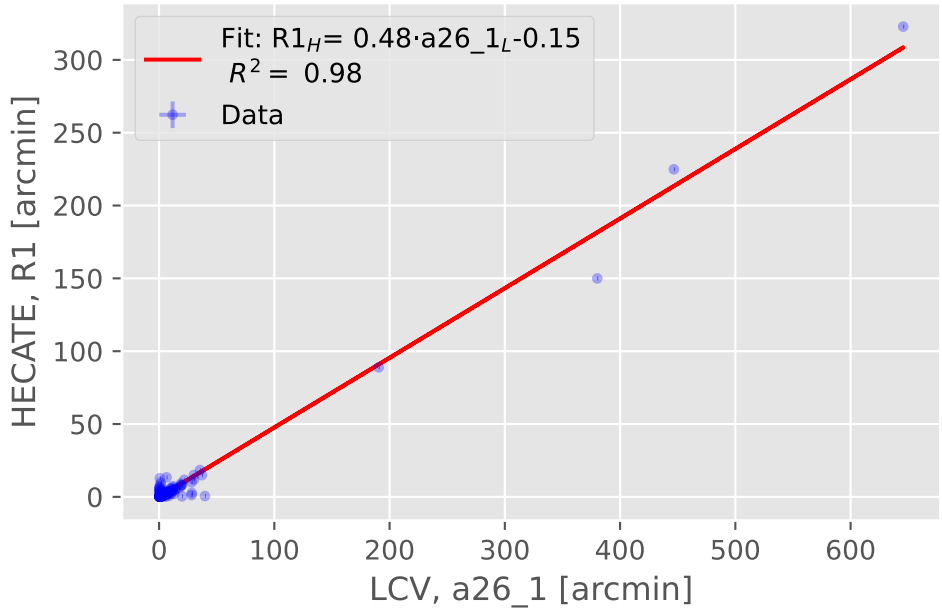


	inc	INCL	Percentage Change [%]
count	296	262	256
mean	59	59	-0

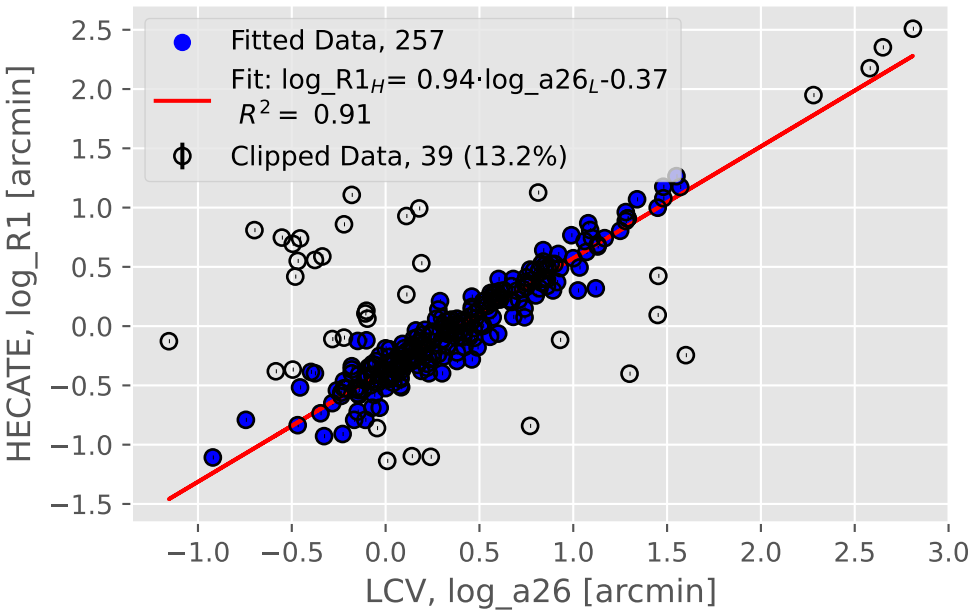
	inc	INCL	Percentage Change [%]
std	19	23	0
min	17	0	-3
25%	45	45	-0
50%	60	62	0
75%	72	77	0
max	90	90	1

We can see that for values in the range $[\sim 30°, \sim 80°]$, the values of the LCV inclination are higher. However, since their means, median, min and maxes are similar and the percentage change is practically 0% (mean, median, $\sigma = 0$ with a range $[-3\%, 1\%]$), we can ignore the differences and assume they are the same values.

Major Axis

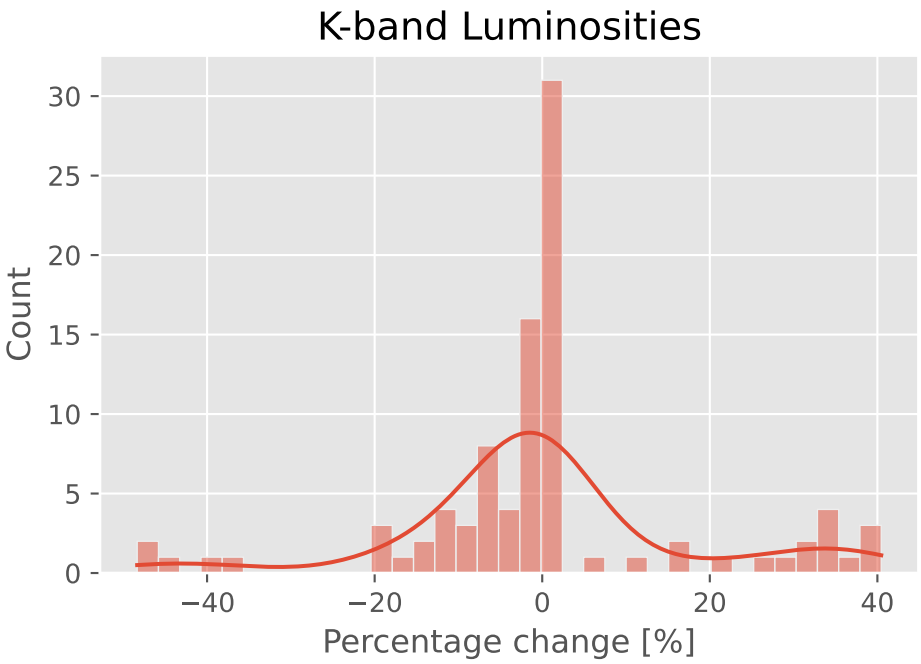


it is not very clear if we truly have a correlation or not. We need to see the linear correlation of the decimal logarithms.



Luminosities

LCV	HECATE	Description	Pearson Correlation [-1,1]
logKLum	logL_K		0.379

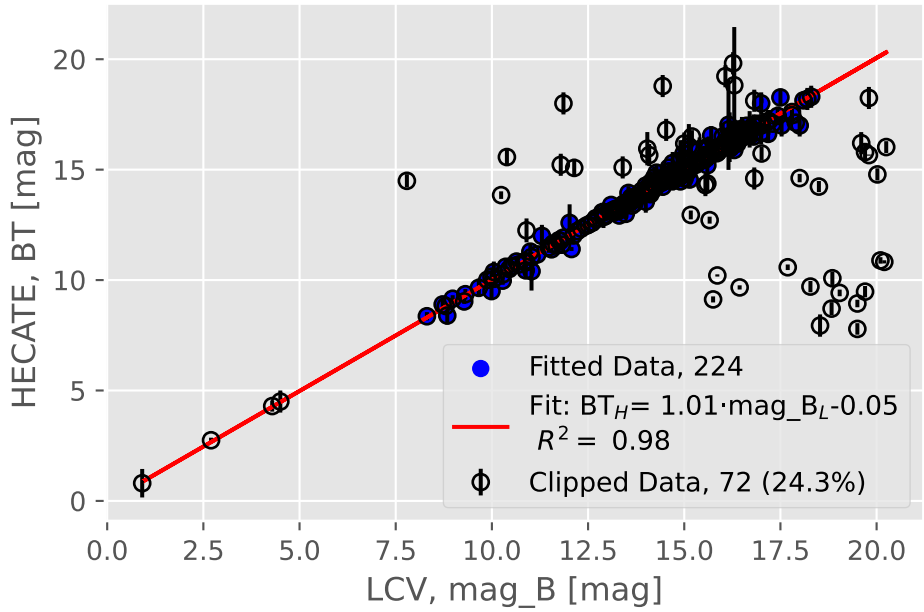


	log(L_K)_{LCV}$	log(L_K)_{HEC}	Percentage Change [%]
count	296	94	94
mean	8	9	0
std	1	1	17
min	4	5	-48
50%	8	9	-0
max	11	11	40

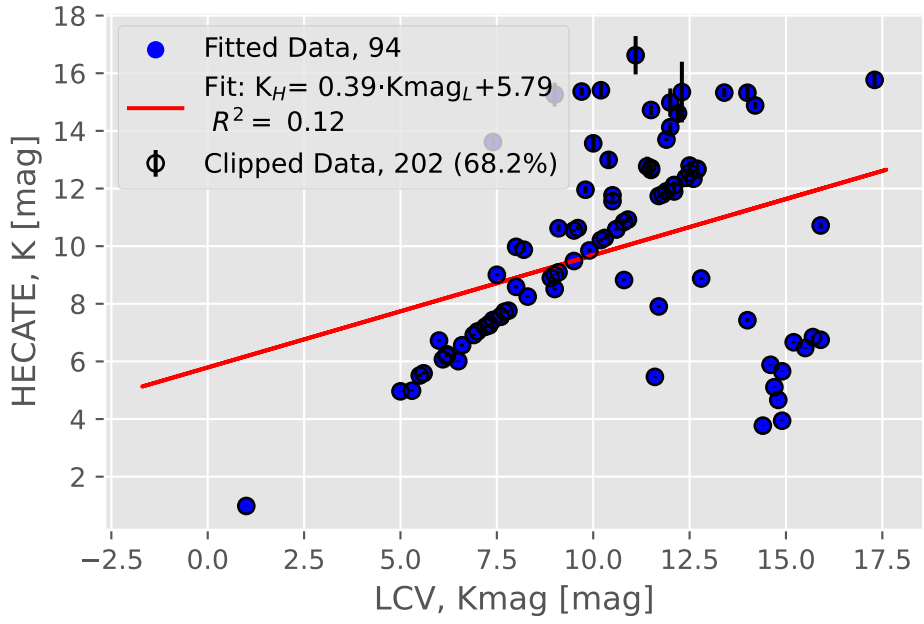
Magnitudes

LCV	HECATE	Description	Pearson Correlation [-1,1]
mag_B (with errors)	BT (with errors)		0.992
Kmag	K	2MASS band magnitude (both)	0.348

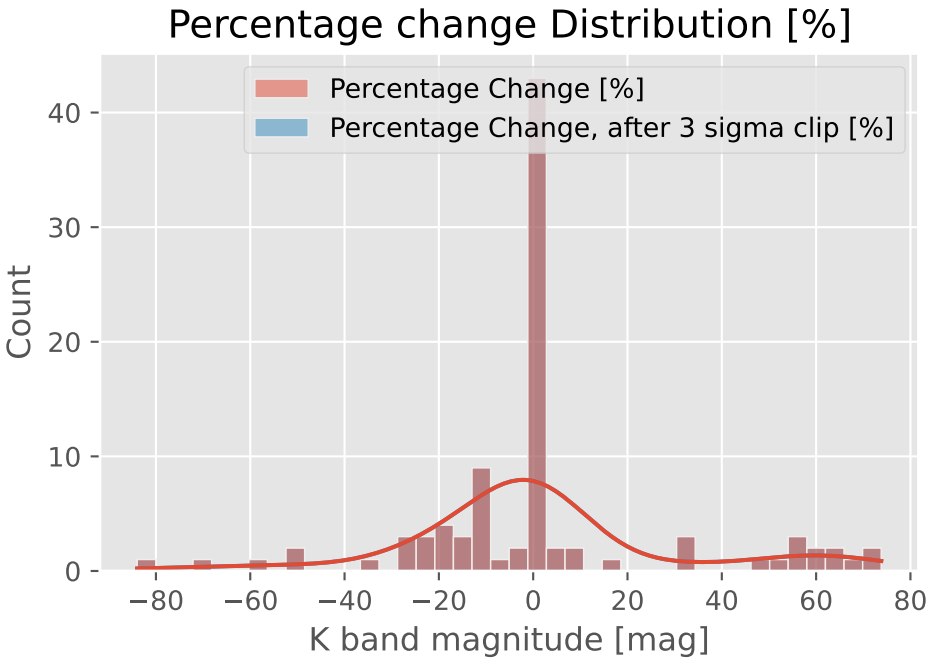
B mag


HECATE, BT [mag]
LCV, mag_B [mag]
Fitted Data, 224
Fit: BT$_H$= 1.01·mag_B$_L$-0.05
R^2 = 0.98
Clipped Data, 72 (24.3%)

K mag


HECATE, K [mag]
LCV, Kmag [mag]
Fitted Data, 94
Fit: K$_H$= 0.39·Kmag$_L$+5.79
R^2 = 0.12
Clipped Data, 202 (68.2%)

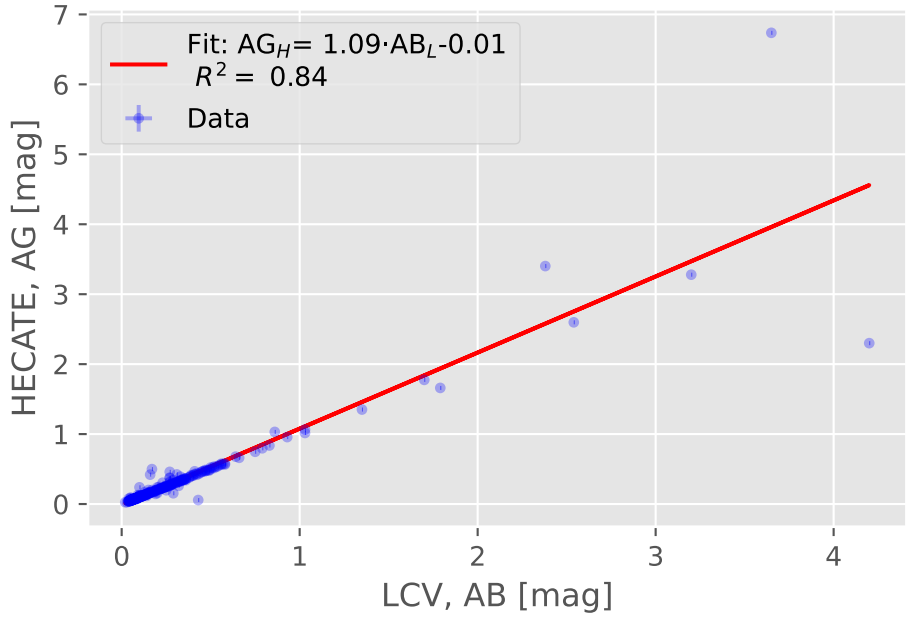
	Percentage Change [%]	Percentage Change, after 3 sigma clip [%]
count	94	94
mean	2	2
std	29	29
min	-84	-84
50%	-0	-0
max	74	74



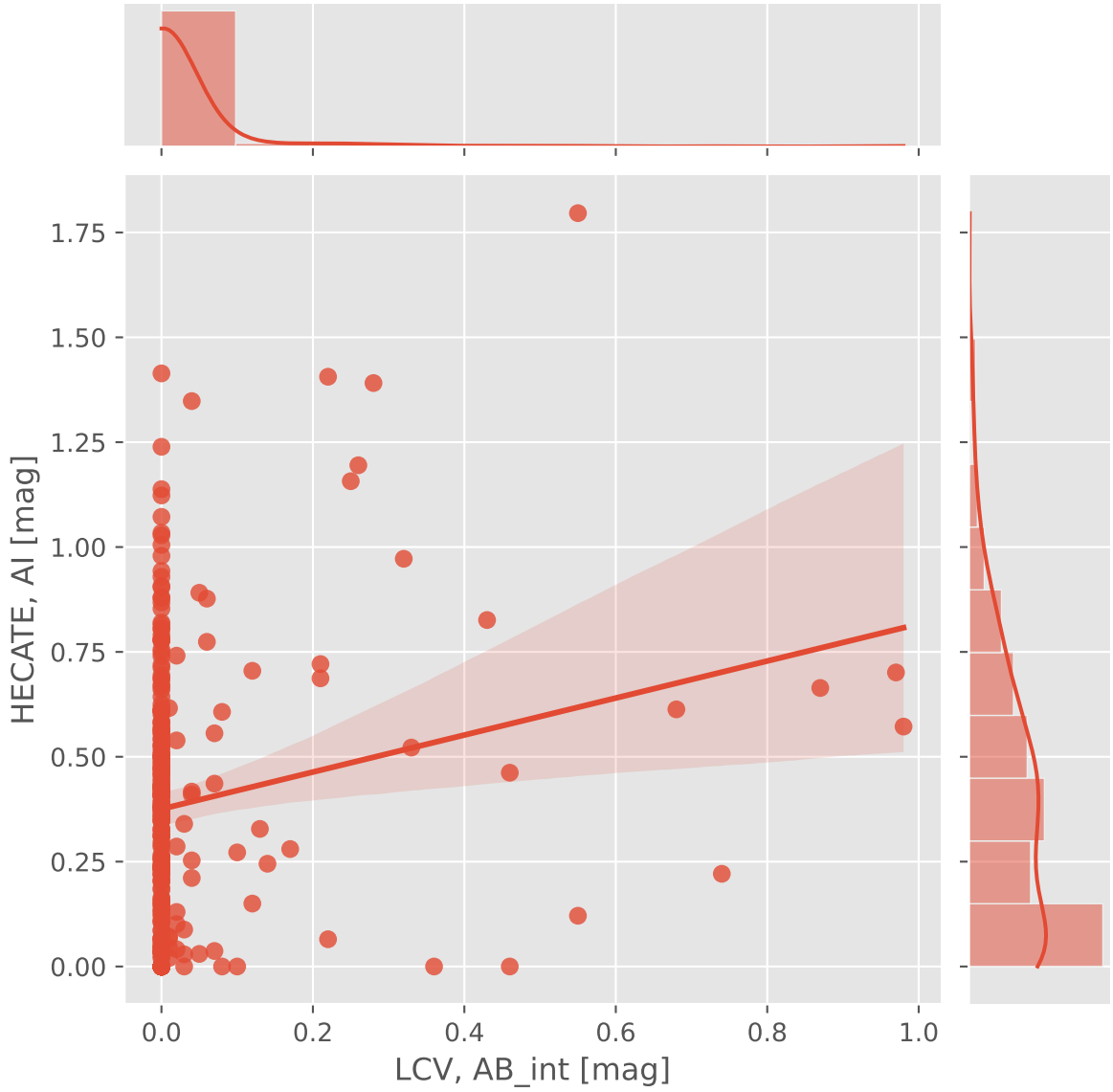
Absorption

LCV	HECATE	Description	Pearson Correlation [-1,1]
AB	AG	Galactic extinction/absorption in B band	0.914
AB_int	AI	Internal/Intrinsic B band extinction/absorption	0

Galactic


Fit: AG$_H$= 1.09·AB$_L$-0.01
R^2 = 0.84
Data
HECATE, AG [mag]
LCV, AB [mag]
0
1
2
3
4
5
6
7

Internal

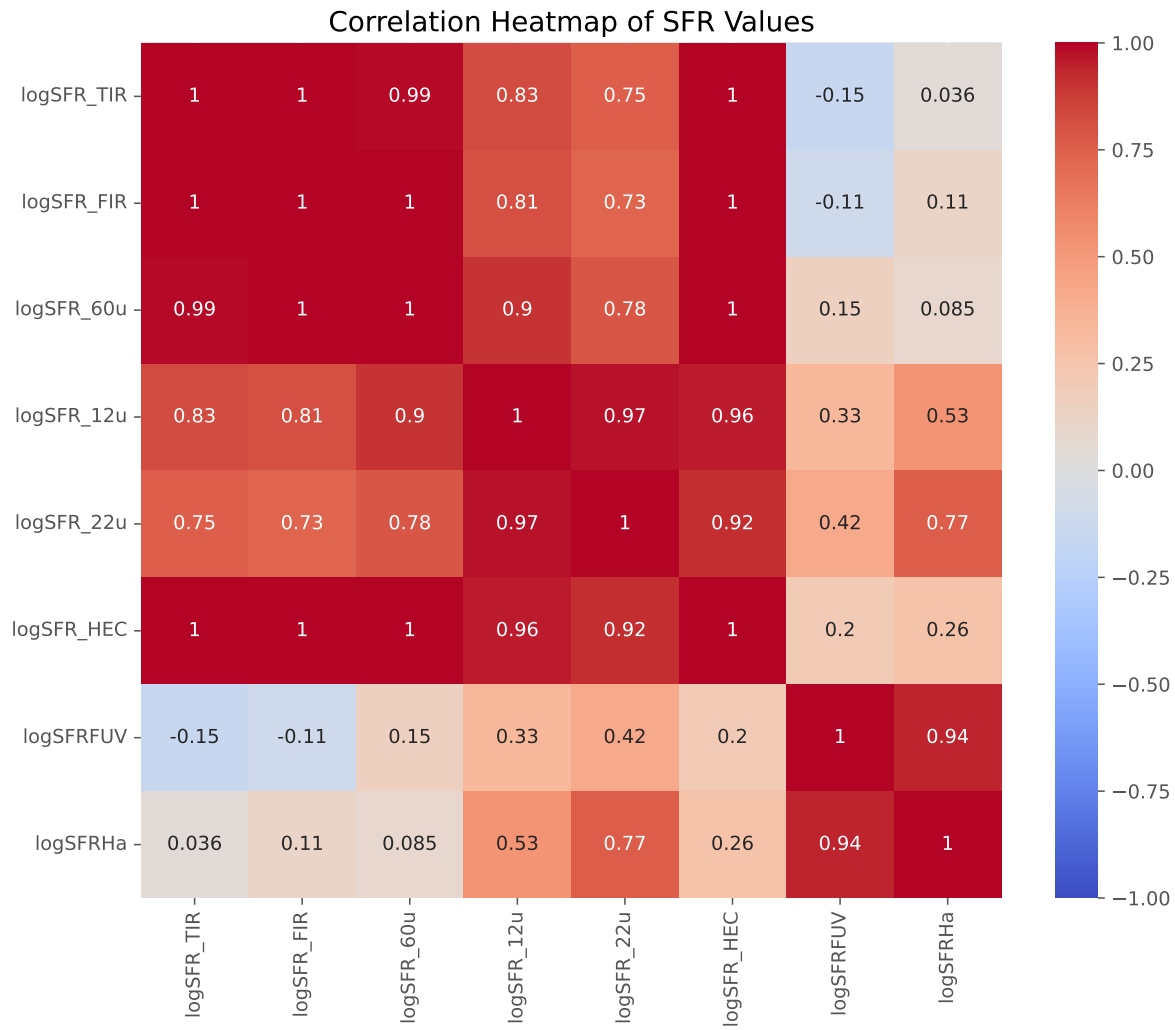


Here we can see an attempt to correlate the internal absorption of the two catalogs, however as it is very clear, there is no correlation, since even their distributions are not similar.

SFR

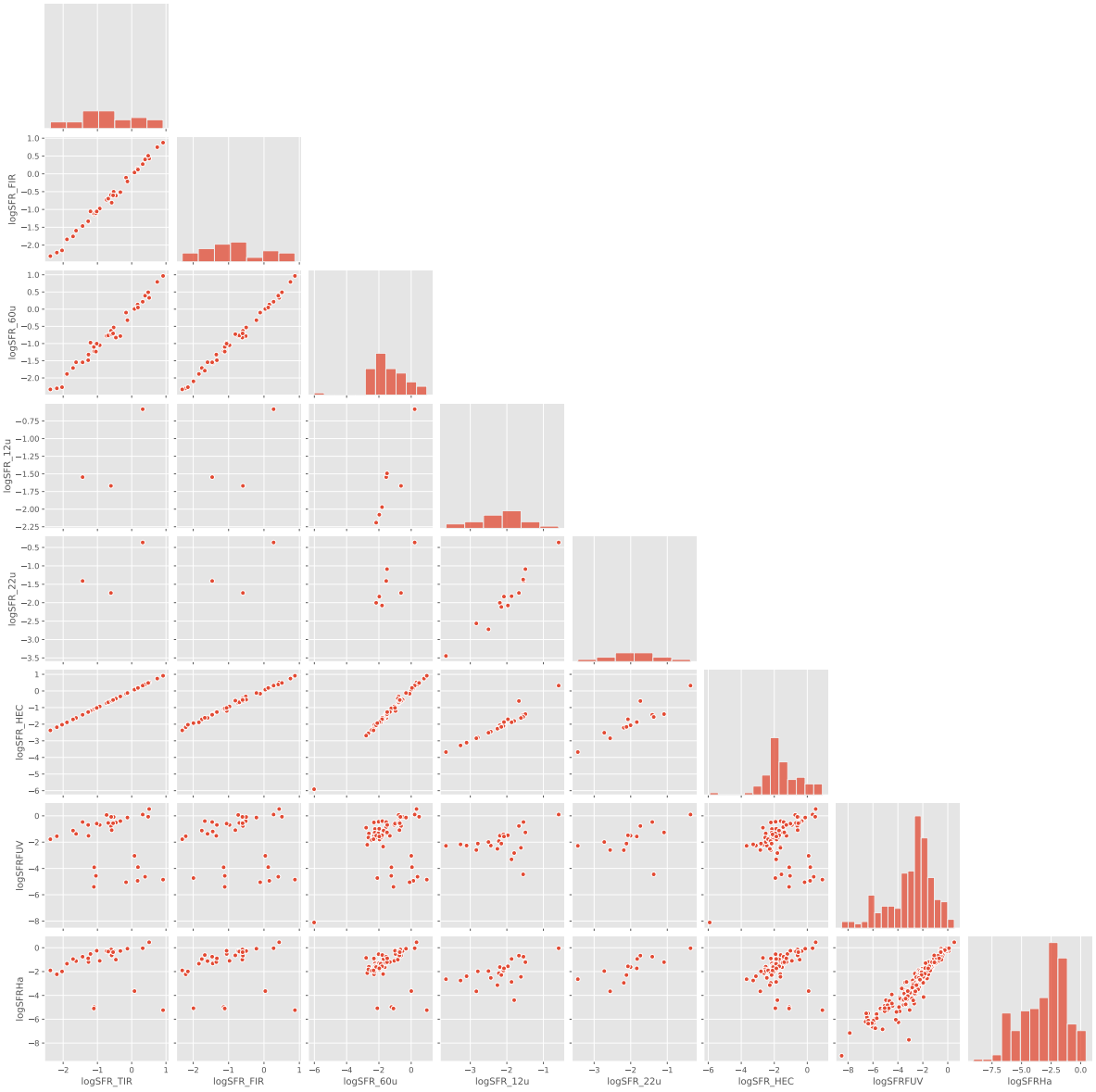
LCV	HECATE	Description	Count
	logSFR_TIR	Decimal logarithm of the total-infrared SFR estimate [Msol/yr]	35
	logSFR_FIR	Decimal logarithm of the far-infrared SFR estimate [Msol/yr]	38
	logSFR_60u	Decimal logarithm of the 60um SFR estimate [Msol/yr]	64
	logSFR_12u	Decimal logarithm of the 12um SFR estimate [Msol/yr]	23
	logSFR_22u	Decimal logarithm of the 22um SFR estimate [Msol/yr]	14
	logSFR_HEC	Decimal logarithm of the homogenised SFR estimate [Msol/yr]	81
	logSFR_GSW	Decimal logarithm of the SFR in GSWLC-2 [Msol/yr]	0
SFRFUV		FUV derived integral star formation rate	257
SFRHa		H{alpha} derived integral star formation rate	249

Heatmap

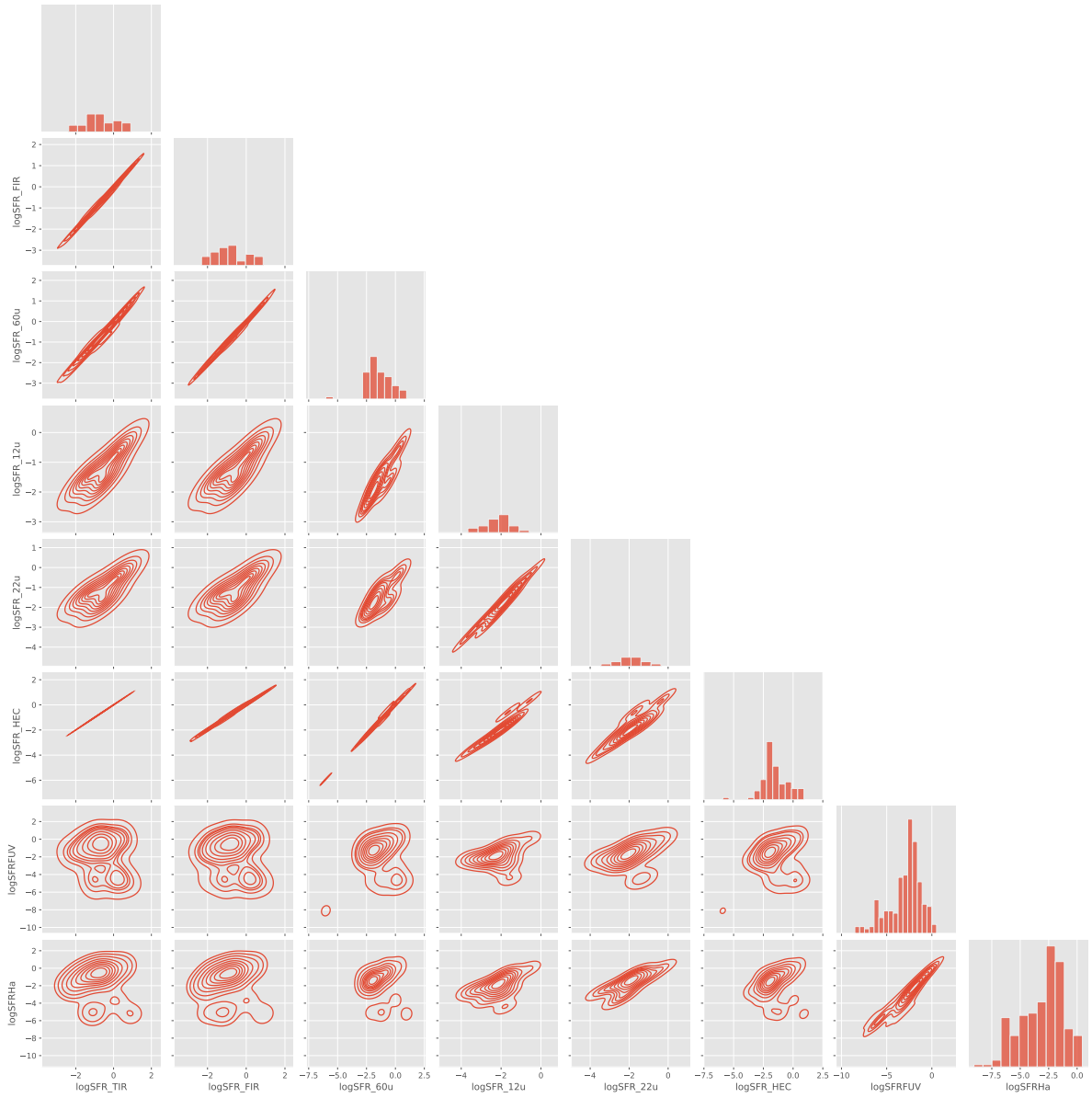


Scatter Grid

```
<Figure size 3000x2400 with 0 Axes>
```


logSFR_TIR
logSFR_FIR
logSFR_60u
logSFR_12u
logSFR_22u
logSFR_HEC
logSFRFUV
logSFRHa

Density Grids

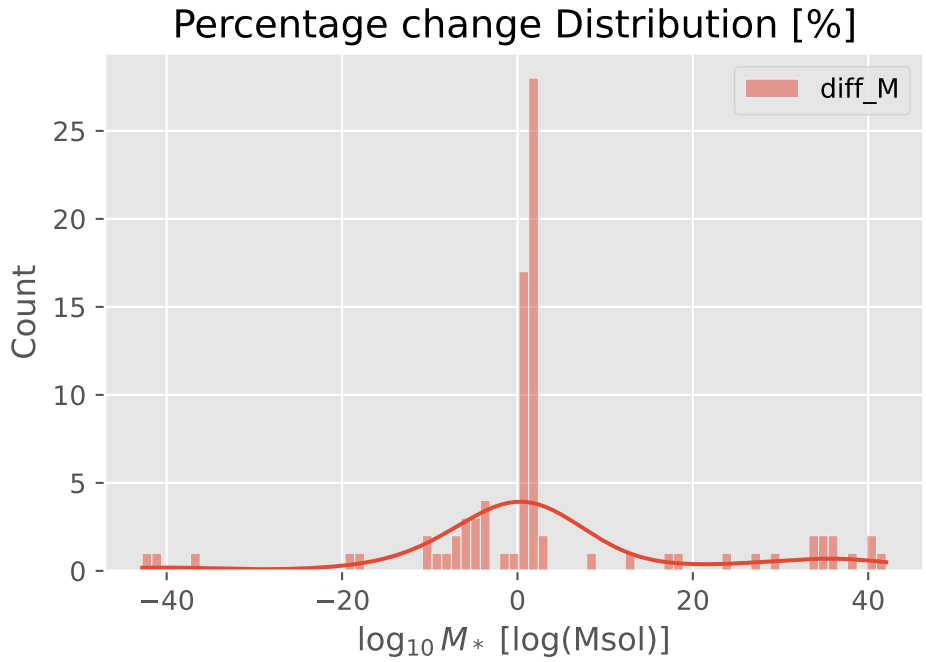


It is possible we dont see a good correlation because we dont have a big enough common sample

Masses

LCV	HECATE	Description	Count
logM26		Log mass within Holmberg radius	263
logMHI		Log mass within Holmberg radius	269
	logM_HEC	Decimal logarithm of the stellar mass [Msol]	87
	logM_GSW	Decimal logarithm of the stellar mass in GSWLC-2 [Msol]	0
logStellarMass		Stellar Mass from M_*/L=0.6	296

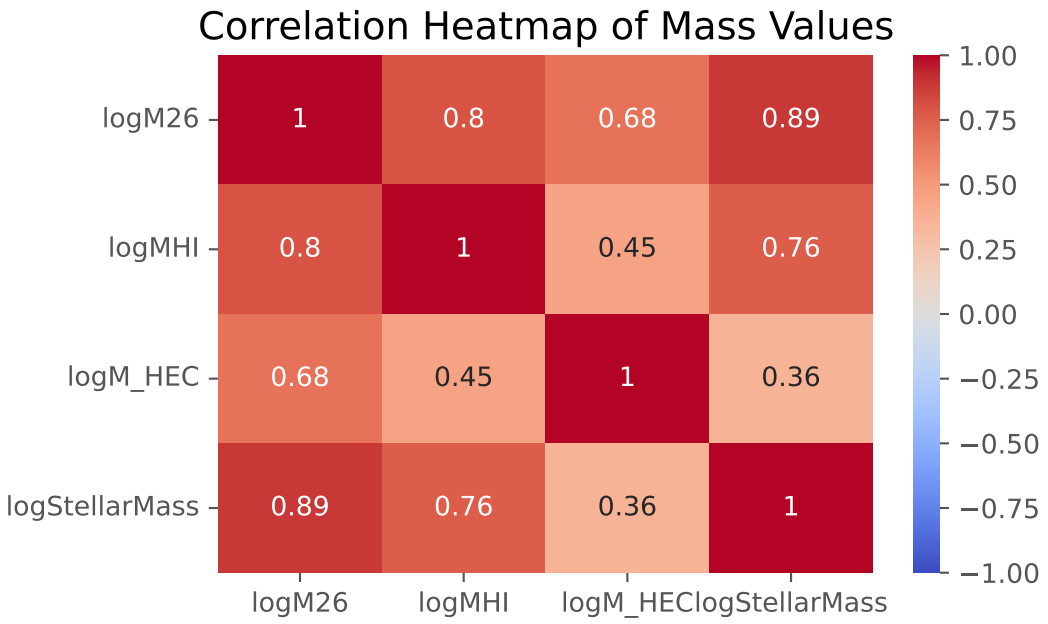
Stellar Masses Comparison



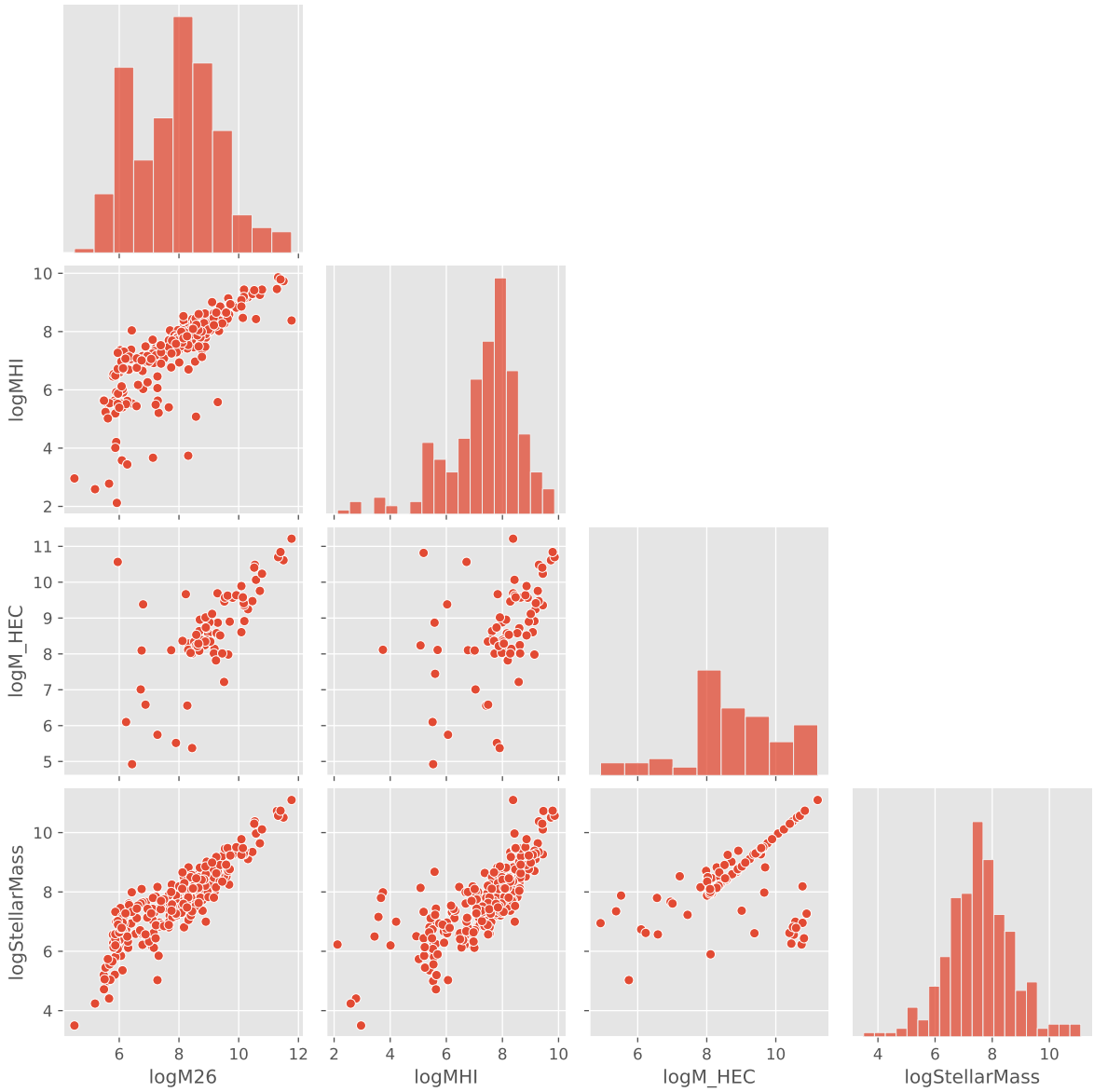
	Percantage Change [%]
count	87
mean	4
std	16

	Percantage Change [%]
min	-43
50%	1
max	42

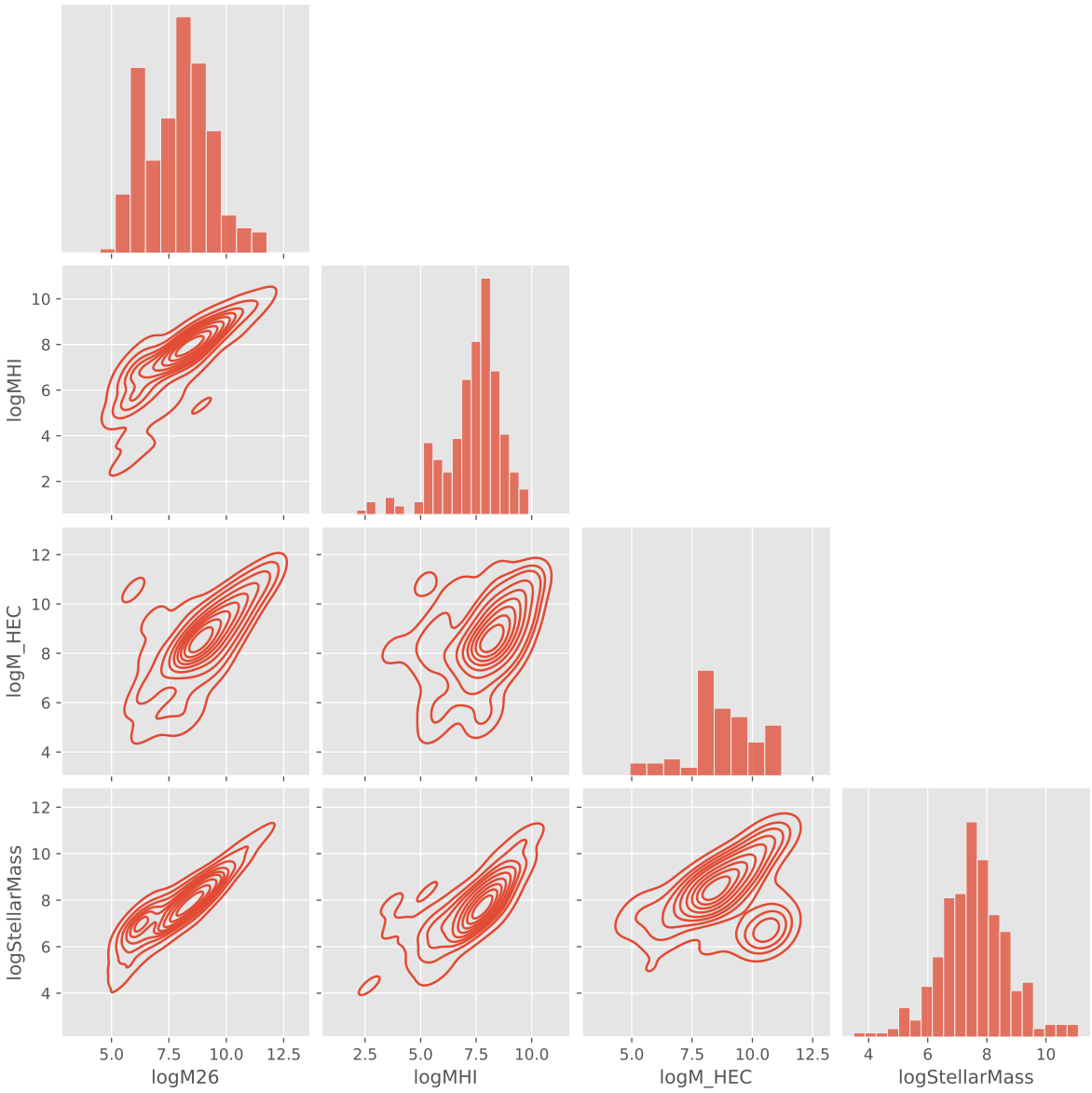
Heatmap



Plot Grid


logMHI
logM_HEC
logStellarMass
logM26
logMHI
logM_HEC
logStellarMass

Density Grid



Karachentsev, Igor D., and Elena I. Kaisina. 2013. "STAR FORMATION PROPERTIES IN THE LOCAL VOLUME GALAXIES VIA H AND FAR-ULTRAVIOLET FLUXES." *AJ* 146 (3): 46. https://doi.org/10.1088/0004-6256/146/3/46.

Karachentsev, Igor D., Dmitry I. Makarov, and Elena I. Kaisina. 2013. "UPDATED NEARBY GALAXY CATALOG." *AJ* 145 (4): 101. https://doi.org/10.1088/0004-6256/145/4/101.

Kovlakas, K., A. Zezas, J. J. Andrews, A. Basu-Zych, T. Fragos, A. Hornschemeier, K. Kouroumpatzakis, B. Lehmer, and A. Ptak. 2021. "The Heraklion Extragalactic Cat-

alogue (HECATE): A Value-Added Galaxy Catalogue for Multimessenger Astrophysics." *Monthly Notices of the Royal Astronomical Society* 506 (September): 1896–1915. https://doi.org/10.1093/mnras/stab1799.