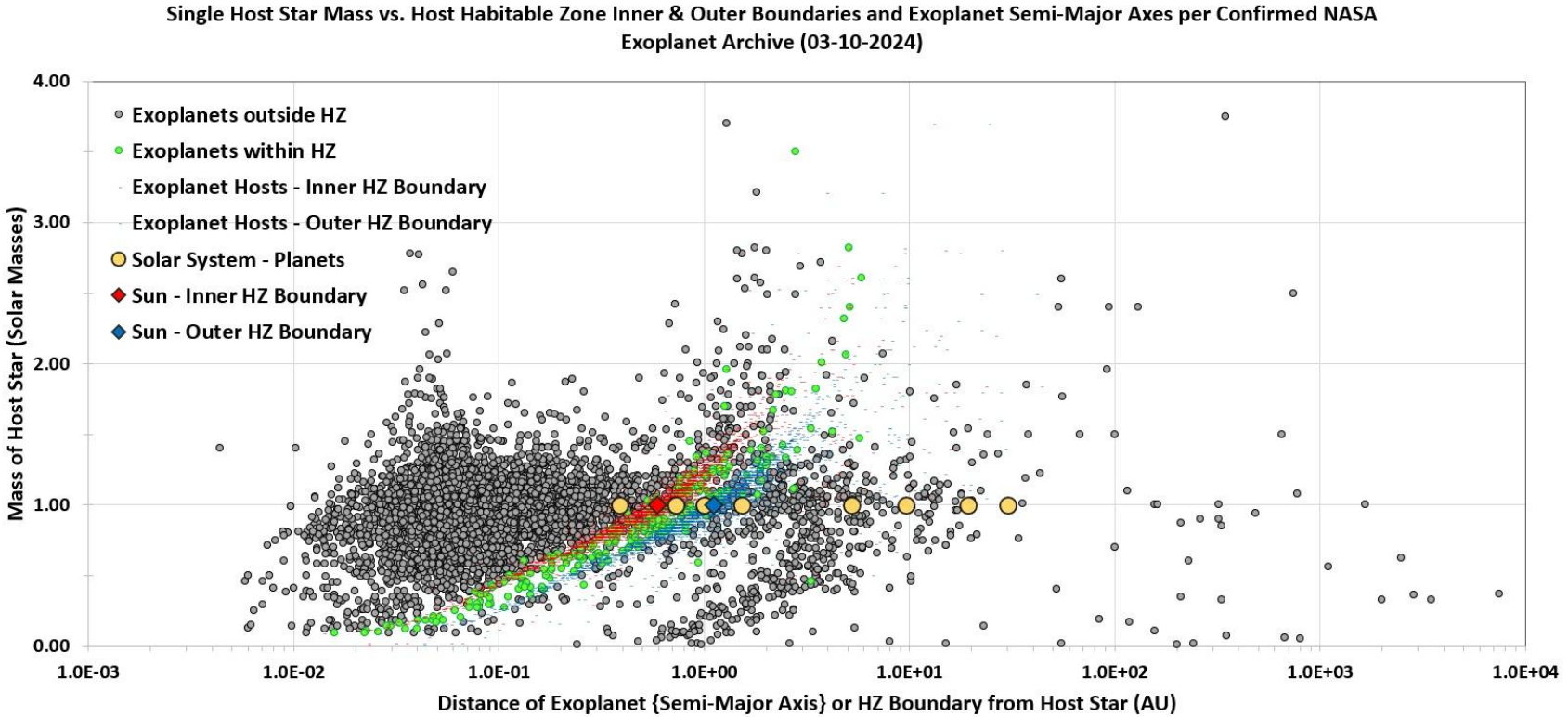
October 7, 2024

exoplanet exploration

galaxies paper - update figure 9

- Differentiate HZ and non-HZ exoplanets using color codes



galaxies paper - HZ limits used in figure 9

- the inner and outer boundaries of HZ in figure 9 are computed based on the equation (1) in the galaxies paper.

$$HZ_{inner} = \frac{R_{\odot} \times T_{\odot}^2 \times k^2 \times (1-A)^{0.5}}{2 \times T_{surf,high}^2}$$

$T_{surf,high}$ = 100 + 273.5

$$HZ_{outer} = \frac{R_{\odot} \times T_{\odot}^2 \times k^2 \times (1-A)^{0.5}}{2 \times T_{surf,low}^2}$$

$T_{surf,low}$ = 0 + 273.5

galaxies paper - try with ML

- join NASA 03-10-2024 data with HWC data from PHL.
- HWC data has a "*P_HABITABLE*" data field that can be used as label
- training data preprocessing:
  - remove data fields that are not relevant to training
  - drop data fields with too much missing values
  - for categorical data fields:
    - filling missing values with mode
    - encode with LabelEncoder
  - for numeric data fields:
    - filling missing values with MICE imputation
  - use SMOTEENN to oversample and downsample to overcome sample imbalances

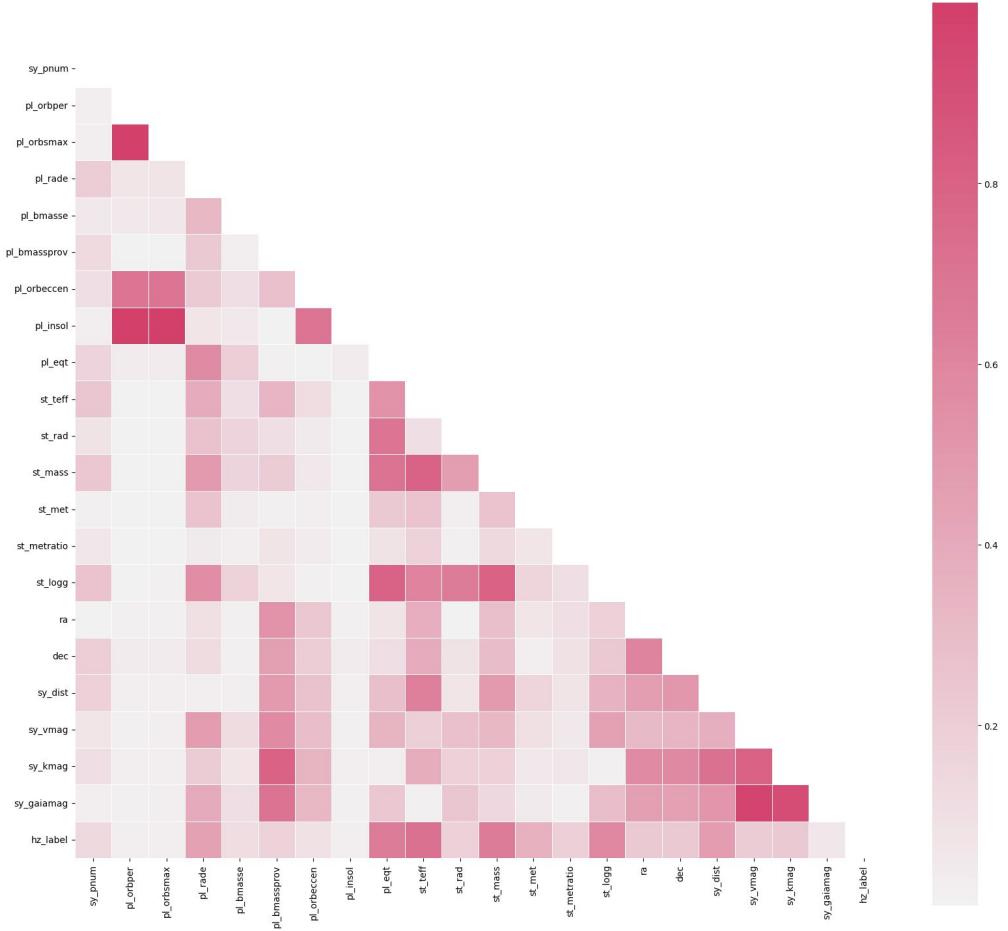
```
hz_label=0, count=4520 (98.798%)
hz_label=1, count=55 (1.202%)
```

galaxies paper - feature correlation analysis

correlation analysis. remove highly correlated features:

- pl_orbeccen
- pl_insol
- sy_gaiamag

end up with 18 features in the training data



galaxies paper - training data features

```
Data columns (total 19 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   sy_pnum        8924 non-null   int64
 1   pl_orbper      8924 non-null   float64
 2   pl_orbsmax     8924 non-null   float64
 3   pl_rade        8924 non-null   float64
 4   pl_bmasse      8924 non-null   float64
 5   pl_bmassprov   8924 non-null   int64
 6   pl_eqt         8924 non-null   float64
 7   st_teff        8924 non-null   float64
 8   st_rad         8924 non-null   float64
 9   st_mass        8924 non-null   float64
 10  st_met         8924 non-null   float64
 11  st_metratio    8924 non-null   int64
 12  st_logg        8924 non-null   float64
 13  ra             8924 non-null   float64
 14  dec            8924 non-null   float64
 15  sy_dist        8924 non-null   float64
 16  sy_vmag        8924 non-null   float64
 17  sy_kmag        8924 non-null   float64
 18  hz_label       8924 non-null   int64
dtypes: float64(15), int64(4)
```

galaxies paper - KNN classifier

KNN Classifier - Classification Report:

	precision	recall	f1-score
0	0.74	0.99	0.84
1	0.98	0.64	0.77
accuracy			0.81