# A bird's-eye view on the habitability of exoplanets via statistical learning techniques

## Project for the exam: Machine learning, statistical learning, deep learning and artificial intelligence - Unsupervised Learning

Marzio De Corato

October 27, 2020

# Overview

- **Final goal**: Survey the performances of different statistical learning algorithm in the prediction of exoplanets habitability
- **Dataset**: Planetary Habitability Laboratory @ UPR Arecibo [1]
- **Algorithms**: Decision Tree, Random Forest, Support Vector Classifier, Logistic Regression, Linear and Quadratic Classifier

# Theoretical background - Exoplanets habitability

- **Habitability**: Rocky planets where water is present in liquid phase
- **Liquid phase**: At first order, if water is present, the liquid phase is controlled by the surface temperature
- **Atmosphere**: The atmosphere ($CO_2$) influences the surface temperature trough the greenhouse effect
- **$H_2$ and $CH_4$**: Other gases such as $H_2$ and $CH_4$ can produce the greenhouse effect, thus the habitable zone can be extended

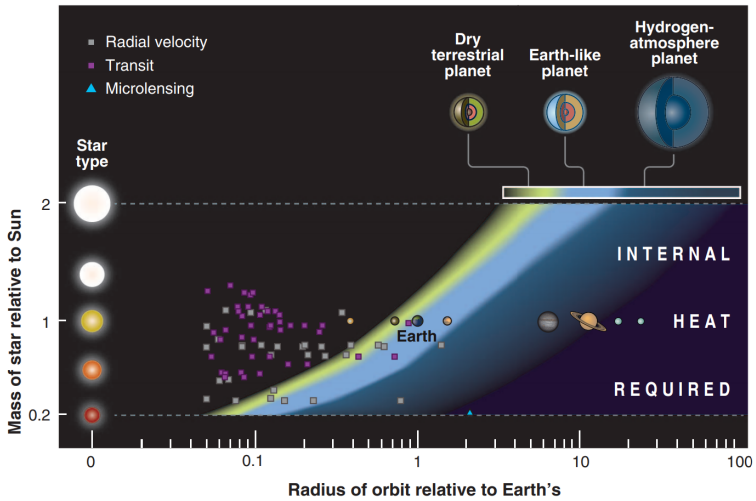



Image taken from [4]

# Theoretical background - Star features

- **Main features**: For this work the main features of star are the stellar luminosity (S_L), its temperature (S_T) and spectral type (S_S_T)
- **H-R diagram**: with these features the Hertzsprung-Russell diagram classify the stars (the temperature and spectral type of a star are two faces of the same medal)

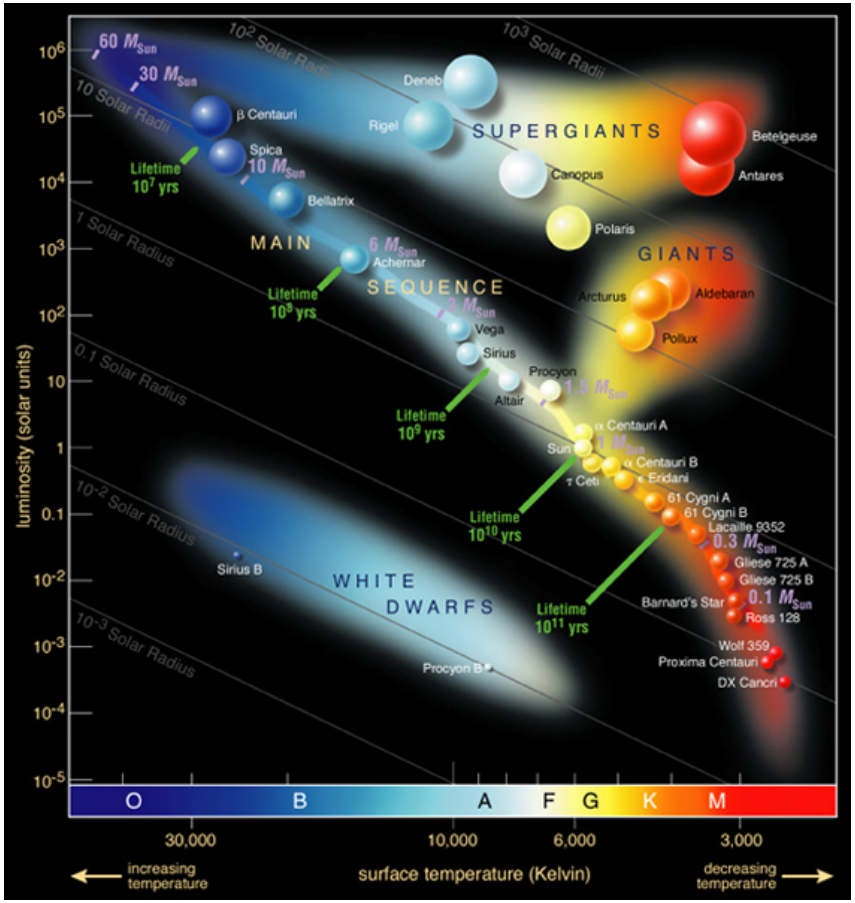



Image taken from [2]

# Theoretical background - Planet features

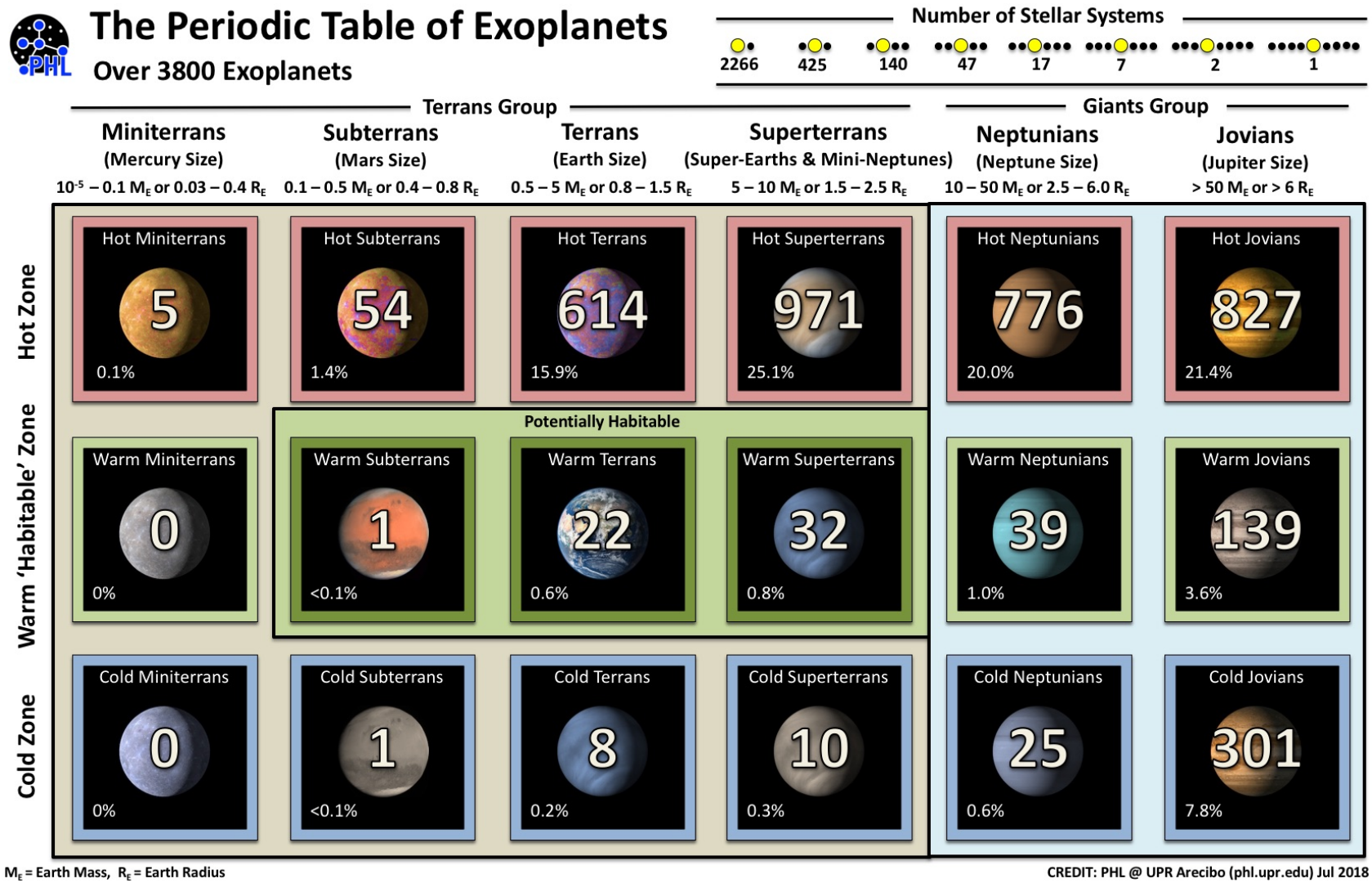
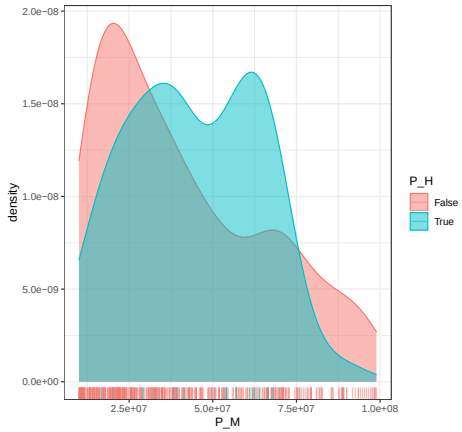



Image taken from [3]

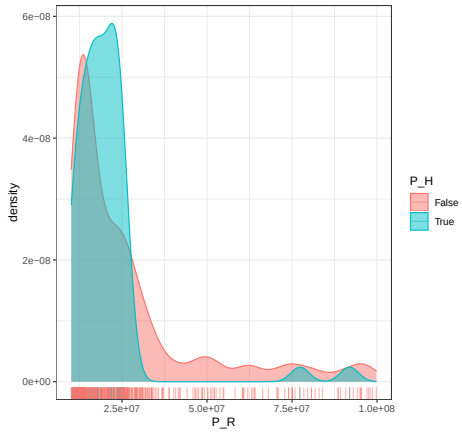
# Theoretical background - Planet features

- **Distance**: in this work the mean planet distance from the host star (P_D), the periastron (P_PN) and the apastron (P_A) as well the termal effective distance (P_D_E) from the host star were considered. These quantities constrain the planet orbital period (P_P) via the $3^{th}$ Kepler law (a corollary of Newton's law of universal gravitation)
- **Mass and Radius**: the (estimated) planet mass (P_M) and its radius (P_R) were considered (these are also useful to distinguish the super-earth planets)
- **Temperature**: the planet equilibrium temperature (P_T_E) defined according to the expression $T_{eq} = T_{star}\sqrt{R/2a}\,(1-A)^{0.25})$ where R is the star radius (S_R), a the planet mean distance (P_D) and A the albedo here considered as 0.3 was considered as well the planet mean stellar flux $P_F$
- **Habitability**: The planet habitability was classified with a boolean variable using the values reported in the dataset [1]

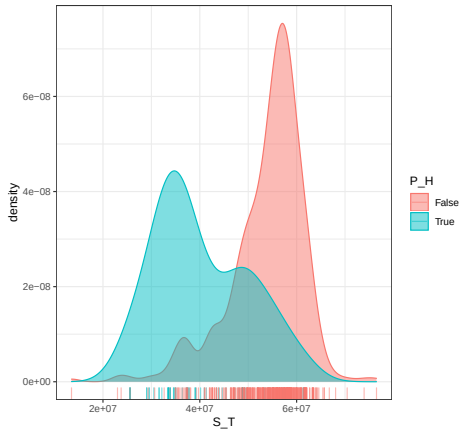
# Data inspection - densities



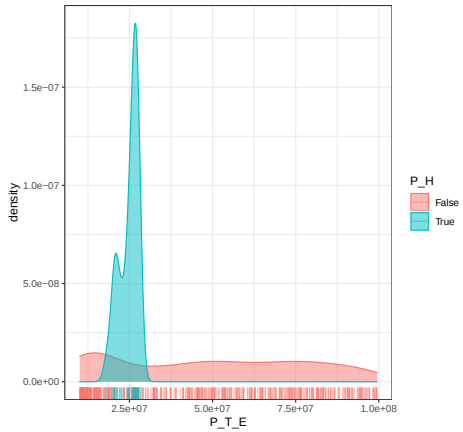

Planet mass



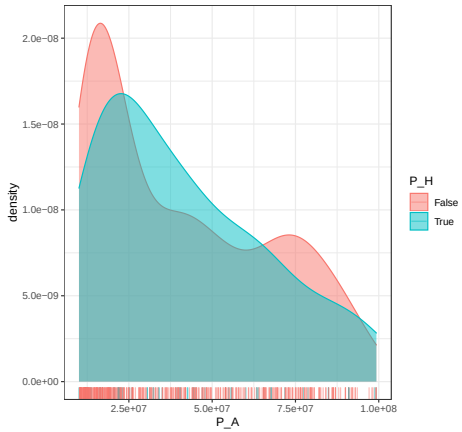

Planet radius



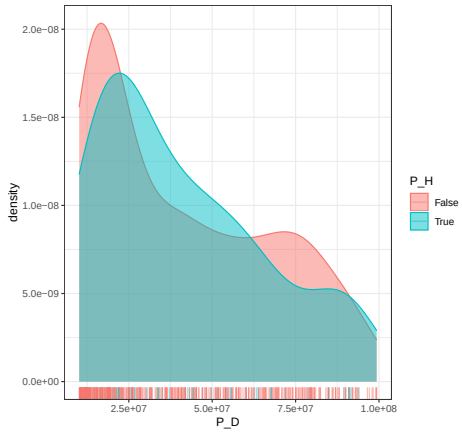

Stellar T



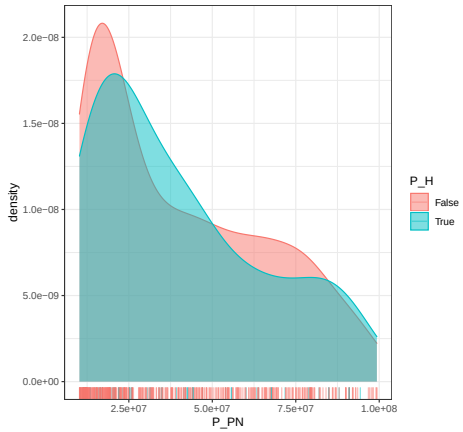

P T E



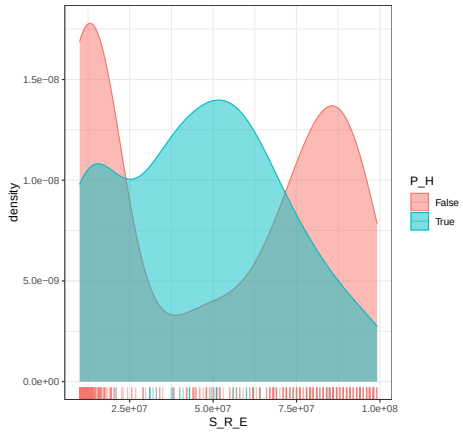

Planet apastron




Planet distance

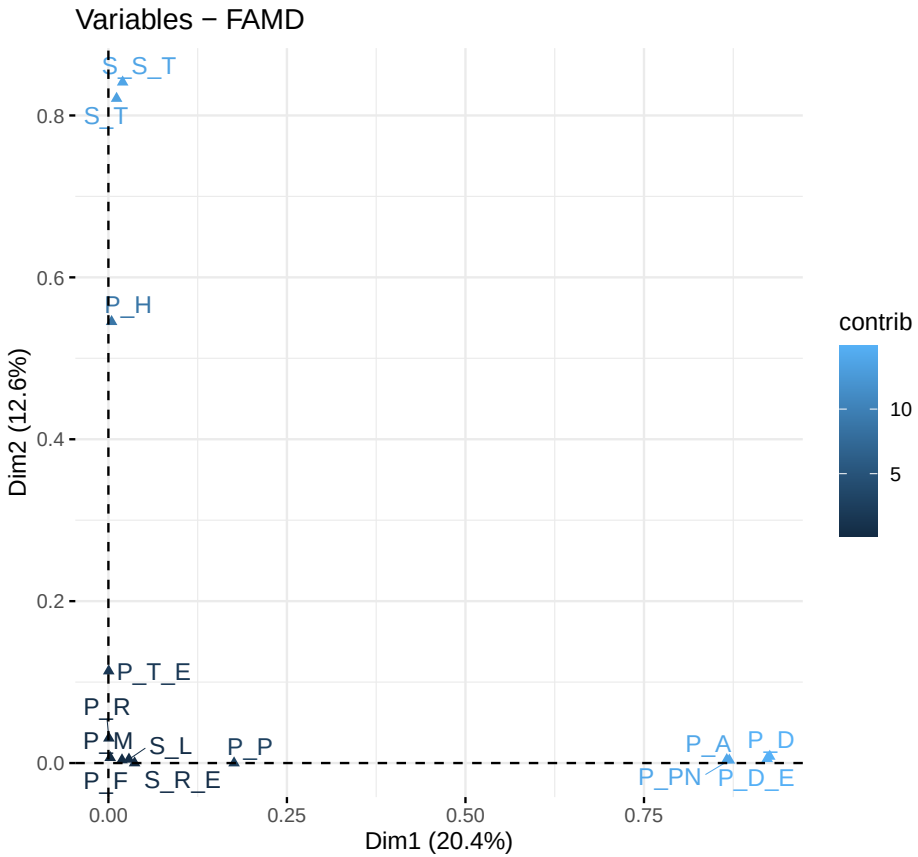


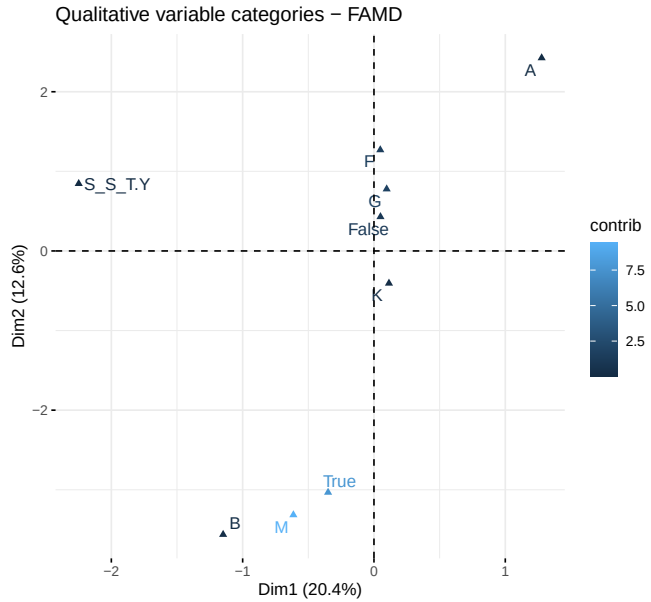

Planet periastron

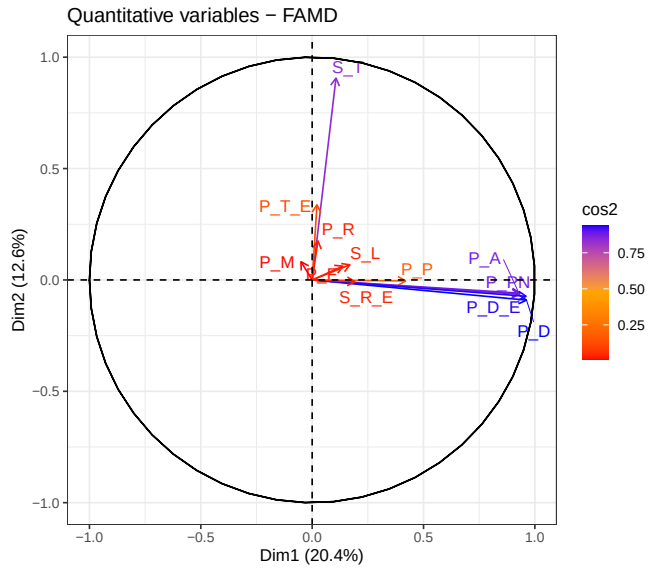



Star radius

# Data inspection - FAMD



# References I


[1] http://phl.upr.edu/projects/habitable-exoplanets-catalog/data/database.

[2] https://www.slideserve.com/ruth-york/chapter-15-surveying-the-stars-powerpoint-ppt-presentation.

[3] http://phl.upr.edu/projects/habitable-exoplanets-catalog/media/pte.

[4] Sara Seager. "Exoplanet habitability". In: *Science* 340.6132 (2013), pp. 577–581.