**Project Title:** Planet Discovery and the Possibility of Extraterrestrial Life

**Team Members:**

1. Gregory Parker
2. Micah West
3. Austin Williams
4. Ryan Brown

**Project Description/Outline:**

Using datasets from NASA’s API on exoplanets and the Planetary Habitability Laboratory’s Habitable Exoplanets Catalog (PHL), we are interested in exploring the trend in planet discovery over the last 10 years, and from what facility the planets were located. We are then interested in looking into how many planets have the potential for life. Using the ranges within each variable (mass, radius, flux, and temperature) from the PHL we plan to collapse the data from NASA to find how many planets fit within these ranges. Next, we would like to pull out which planets fall into most of these ranges. Based on this we plan to reasonably state that there is a possibility of life (and aliens) on the number of planets found. Finally, we are interested in seeing the correlation of the planets above to given PHI variable Earth Similarity Index (ESI). By running correlation tests, we will be able to determine which variable has the most impact on habitability.

**Research Questions to Answer:**

1. How many planets have discovered each year from 2010 to 2020?
2. Which facilities have discovered the most planets?
3. How many planets fall within the ranges given by the PHL data?
4. Which variable (mass, radius, flux, and temperature) has the highest correlation to ESI?

**Datasets to be Used:**

1. NASA’s API: Confirmed Planets API (Exoplanet)
   1. Link to data: https://exoplanetarchive.ipac.caltech.edu/docs/program\_interfaces.html#data
   2. Documentation is found here: <https://exoplanetarchive.ipac.caltech.edu/docs/API_exoplanet_columns.html>
2. Planetary Habitability Laboratory’s Habitable Exoplanets Catalog (PHL)
   1. Located here: <http://phl.upr.edu/projects/habitable-exoplanets-catalog>

**Rough Breakdown of Tasks:**

1. Pull NASA data and keep only required variables and variables of question that will be tested on correlation
2. Rename the variables to be more intuitive
3. Find the number of planets found each year and create graphs
4. Find the top facilities in terms of planets discovered and create graphs
5. Pull in PHL data and calculate upper and lower bounds of variables of interest
6. Merge the two datasets on planet name (to not have duplicate planets)
7. Find the number of planets that fall within bounds of PHI data and create graphs
8. Calculate correlations of variables on the ESI and create linear regression graphs