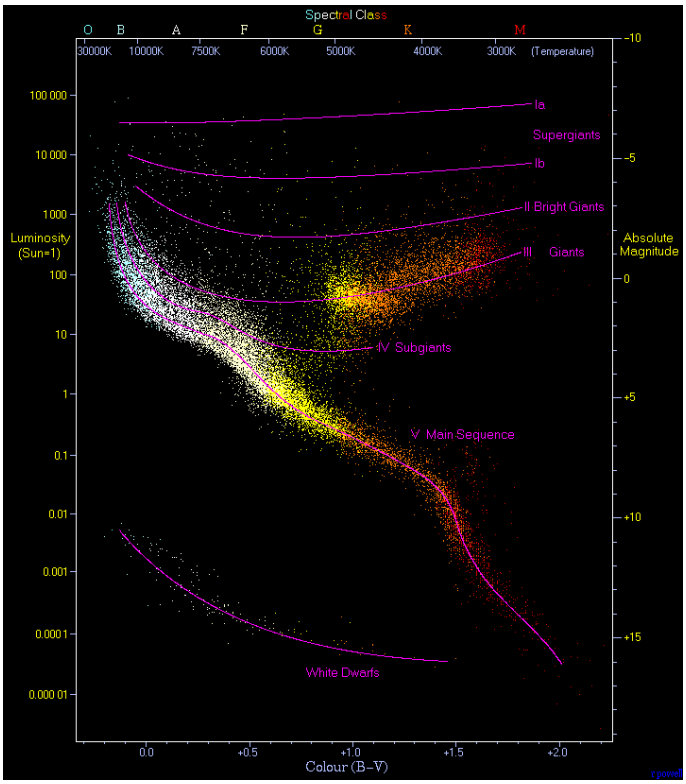
Brayden Matejka and Kavish Loomba
ASTRON 98 Final Project

Attempting to Replicate the Hertzsprung-Russell Diagram



1. Overview

We have chosen to take data from various star clusters available on online databases about the temperature, size, age, spectral class, and other relevant information of stars within those systems. With this data, we will be able to replicate the Hertzsprung-Russell diagram that classifies stars within certain generalized groups. Throughout this proposal we will outline the methods in which we will achieve a striking result.

2. Chosen Phenomenon and Data Source

We will gather information on about 10,000 stars, which is readily available on the website https://vizier.cds.unistra.fr, and attempt to replicate the Hertzsprung-Russell diagram with this data. The website provides a variety of information on an impressively large range of stars (along with the relative certainty of this data), but we will mostly be focusing on the temperature, spectrum, luminosity, absolute magnitude, and color.

3. Equation to Fit Data

We will attempt to visually organize this data in the same way that the Hertzprung-Russel diagram does, which includes playing with the logarithm/exponential of various data points. We will do this in order to achieve a striking visualization that groups the data points into obvious subgroups. With this, we will be able to identify which characteristics of stars lead to placement within certain subgroups.

4. Data Filtering

We have gathered the data on 10,000 stars in a Comma-Separated Values file (.csv), and we will order this data into a list of sublists of float values in order to make gathering certain datapoints easier. We will also take into account the relative certainty of this data (given as a “probability of certainty” percentage as noted by the database) and attempt to filter the data based on how high this value is.

5. Explanation of Model Fit

Historically, just like in the Hertzsprung-Russell diagram, scientists have taken the logarithm of various data (like luminosity and temperature) in order to better the visualization of stars. We will take this method one step further and find the best line(s) of fit under the specific subgroups of stars in order to generate a hypothesis on the classification of star groupings.