Distribution of Stars Within 130 Light Years of Earth
Final Presentation

Revised Hypotheses

Domain of interest: Astronomy Data

Hypthothesis #1: Red dwarf stars [M-class, less than 3,800K] in the local region of Earth [approximately 130 light years] constitute a majority (greater than 51%) of stars

Hypothesis #2: Main sequence stars, similar to our Sun [G-class, 5,200 to 6,000K] are relatively rare at less than 10% of the total number of stars within 130 light years of Earth

Hypothesis #3: Stellar density increases with distance from Earth, with fewer stars within 10 light years compared to the range of 100 to 120 light years

Dataset

GAIA Data Release DR3 from the European Space Agency:
https://www.cosmos.esa.int/web/gaia/dr3
Total dataset: 1.8 billion rows
Used official partner database at Flatiron Institute on 22^{nd} and 5^{th} in NYC called FlatHUB:
https://flathub.flatironinstitute.org/gaiadr3

GAIA DATA RELEASE 3 (GAIA DR3)

This project will be using the following Python libraries:

- pandas: data analysis tool
- numpy: scientific computing
- matplotlib: static, animated, interactive visualizations
- seaborn: statistical graphics based on matplotlib
- scipy: scientific and technical computing
- scikit-learn: machine learning and predictive analysis
- pyvista: 3D visualization and mesh analysis
- mayavi: 3D data visualization and plotting

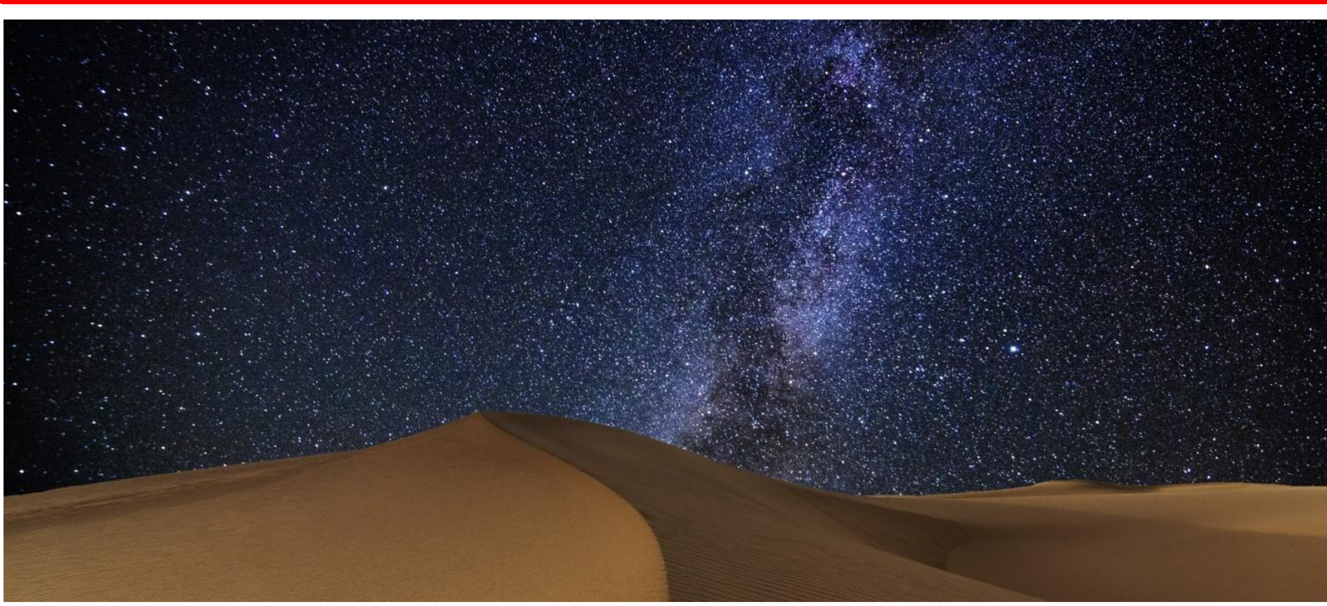
Earth Exists Within a Rare Stellar Configuration

- Our Sun is a G-class main sequence star, which is relatively rare
- Our stellar neighborhood is sparsely populated
- Earth resides in a 'stellar desert'

Two needed conditions appear to be necessary for complex life to develop:
1. G-class [main sequence] star: stable and long-lived
2. Located far from other stars, in relative isolation; has protection from stellar events

Since BOTH conditions are rare, complex, intelligent life does NOT appear to be common in our galaxy

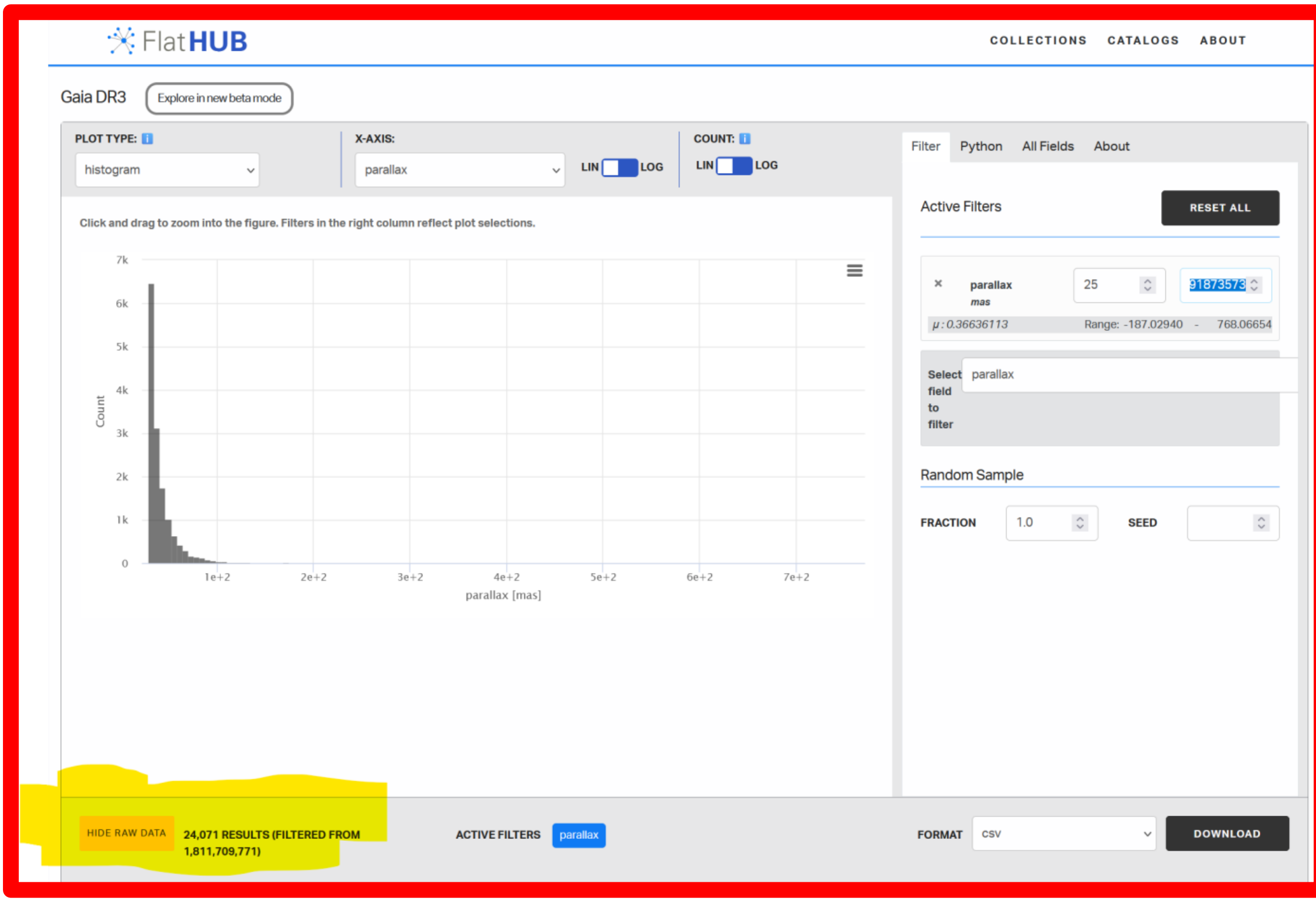
- Distribution graphs of stars indicate M-class [red dwarf] stars dominate the stellar population but do not provide stability nor longevity for life
- Density graphs confirm Earth's isolation in the Galaxy
- This is a rare combination of factors that appears to enable both complex and intelligent life

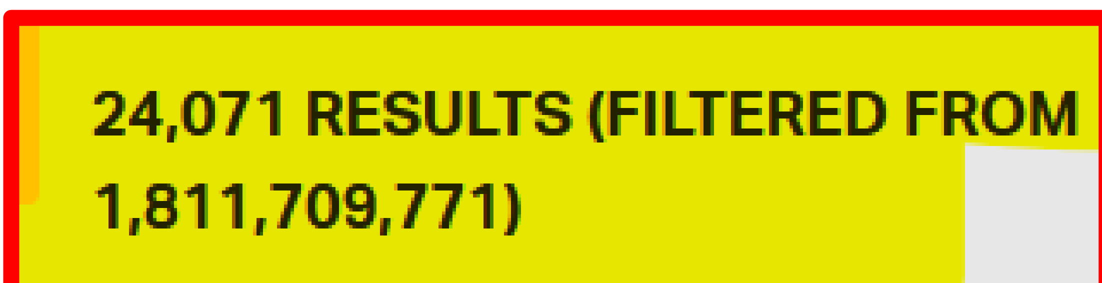


Astronomy Definitions

- Light Year: distance light travels in one year (approximately 5.88 trillion miles)
- Right Ascension & Declination: coordinates used to locate stars in the galaxy
- Kelvin: temperature scale used for stars; Zero K is absolute zero
- Star Class: types of stars; determined by star temperature in Kelvin. Coolest to hottest: M, K, G, F, A, B, O
- Red dwarf (M-class): smallest, coolest, and dimmest star class
- Sun-like (G-class): yellow stars similar in size and temperature to our Sun
- GAIA DR3: satellite named GAIA which produced data for 1.8 billion stars; 1/200th of the total 400 billion stars in the galaxy
- Stellar Neighborhood: the focus of this study; 130 light years

Data Retrieved from Copy of GAIA Database
Stored at Flatiron Institute
Within 25 milliarcseconds of Earth





Raw data fields:
Distance parallax
Color: bp_rp
Position: right ascension, declination

Derived data:
Distance: parallax to light years
Temperature: bp_rp to Kelvin
Stellar class: determined by temperature range

Derived Values

Derivation Paths

1. Distance: convert parallax in milliarcseconds to light years
DistanceLightYears=(1,000/parallax) * 3.26156

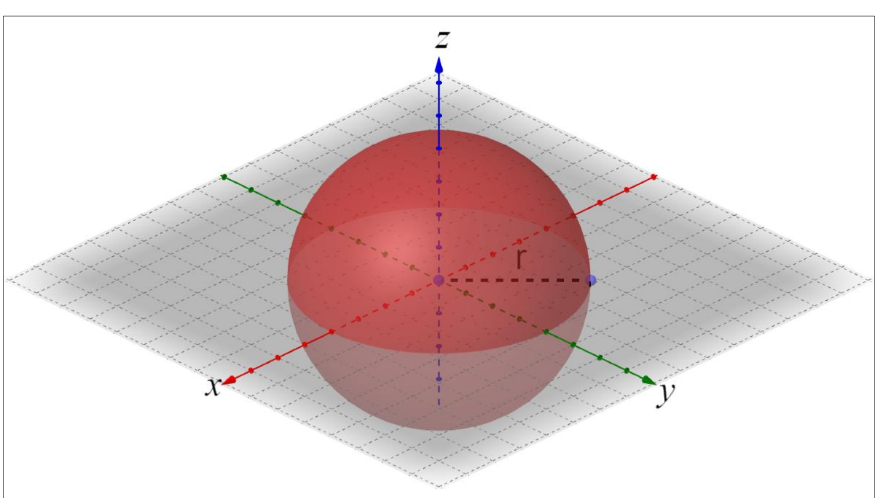
2. Temperature: convert from "bp_rp" to Kelvin
Temperature = 4600 / (0.92 * bp_rp + 1.7) + 4600 / (0.92 * bp_rp + 0.62)

3. Stellar Density: calculate number of stars in shells 10 light years wide
Shell volume from radius1 to radius2: $V = (4/3)\pi([\text{radius2}]^3 - [\text{radius1}]^3)$

4. Cartesian Coordinates: convert Right Ascesnion and Declination to x, y, z
x = DistanceLightYears × cosine(Dec) × cosine(RA)
y = DistanceLightYears × cosine(Dec) × sine(RA)
z = DistanceLightYears × sine(Dec)

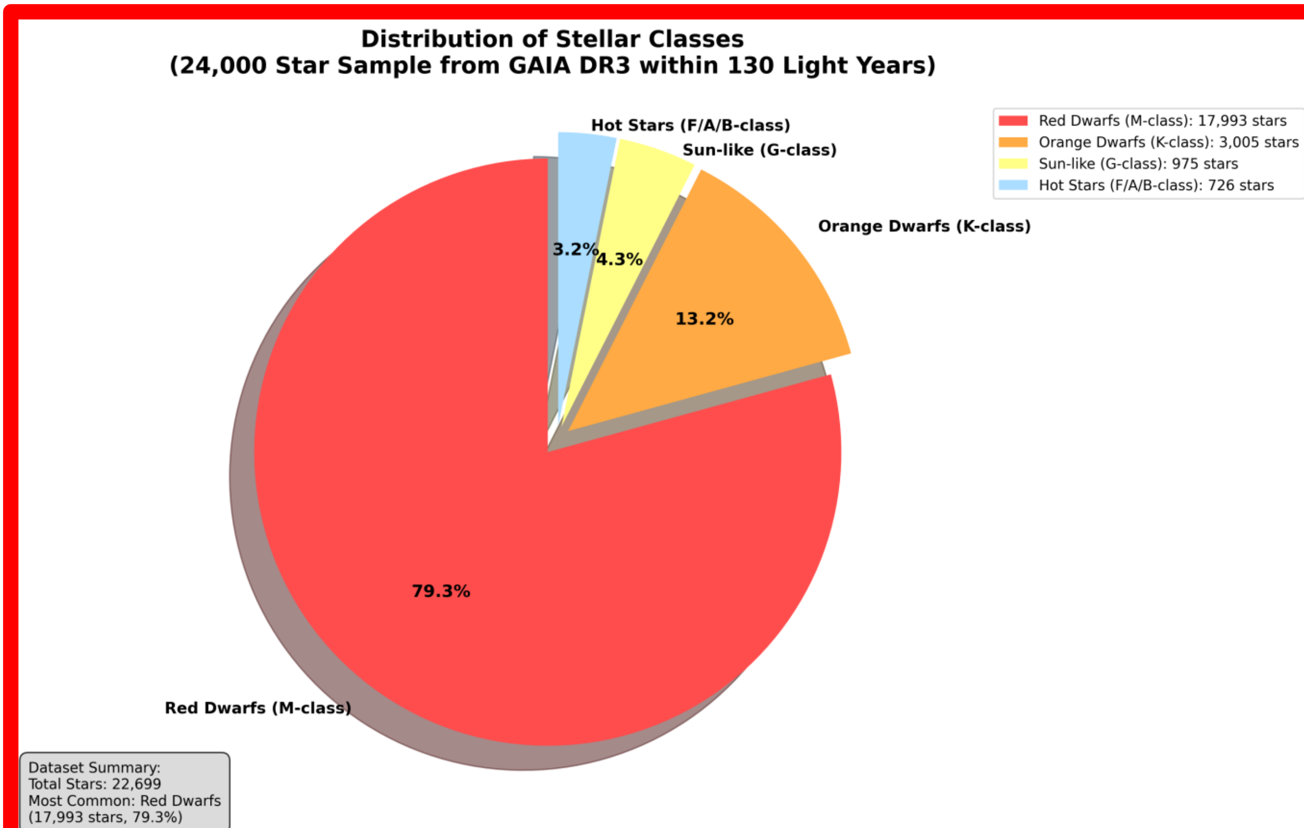
Temperature to Class Table

- 3,800K: Red dwarfs [M-class]
- 3,800K to 5,200K: Orange Dwarfs [K-class]
- 5,200K to 6,000K: Sun-like main sequence stars [G-class]
- Over 6,000K: Hottest stars [F, A, B-class]



Pie Chart – Stellar Classes

Converted parallax to light years
Converted color to temperature in Kelvin
Indicates Red Dwarfs dominate all stars
Sun-like stars are rare



Star class distribution:

Red Dwarfs (M-class)..........17993
Orange Dwarfs (K-class)......3005
Sun-like (G-class)....................975
Hot Stars (F/A/B-class)..........726

Hypothesis #1: Red Dwarf stars constitute more than 79% of all stars within 130 light years of Earth with 17,993 stars

Hypothesis #2: Sun-like star only constitute 4.3% of all stars within 130 light years of Earth with 975 stars

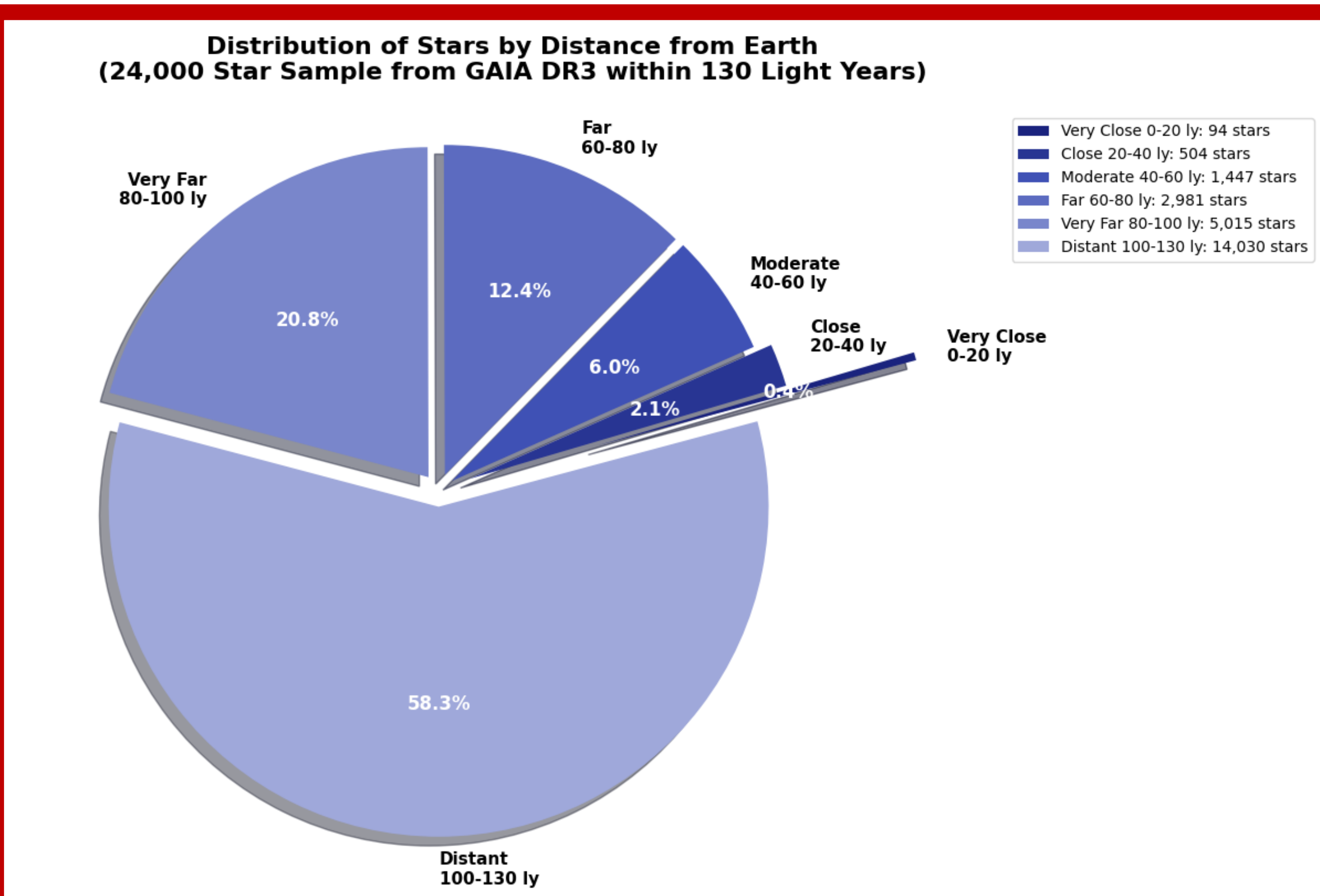
Pie Chart - Distance

Converted parallax to light years

Converted color to temperature in Kelvin

Indicates Red Dwarfs dominate all stars

Sun-like stars are rare



<u>Star distance distribution:</u>

Very close (0-20 ly)....................94

Close (20-40 ly)........................504

Moderate (40-60 ly).............1,477

Far (60-80 ly)........................3,005

Very Far (80-100 ly).............5,015

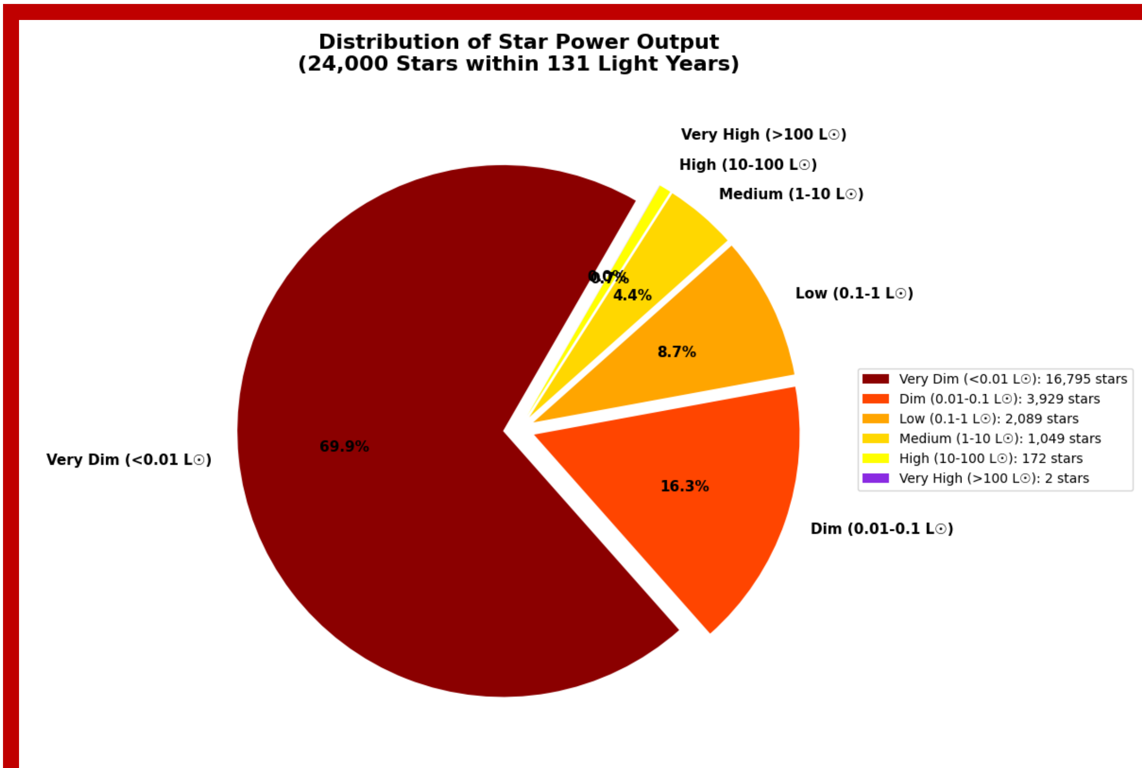
Distant (100-130 ly)...........14,030

Hypothesis #3: sparse population of stars close to Earth

Hypothesis #3: high density of stars 100 to 130 light years

Pie Chart – Power Output

Converted parallax to magnitude
Converted magnitude to luminosity
Indicates very dim stars dominate population
Sun-like stars are relatively rare



<u>Star power distribution:</u>

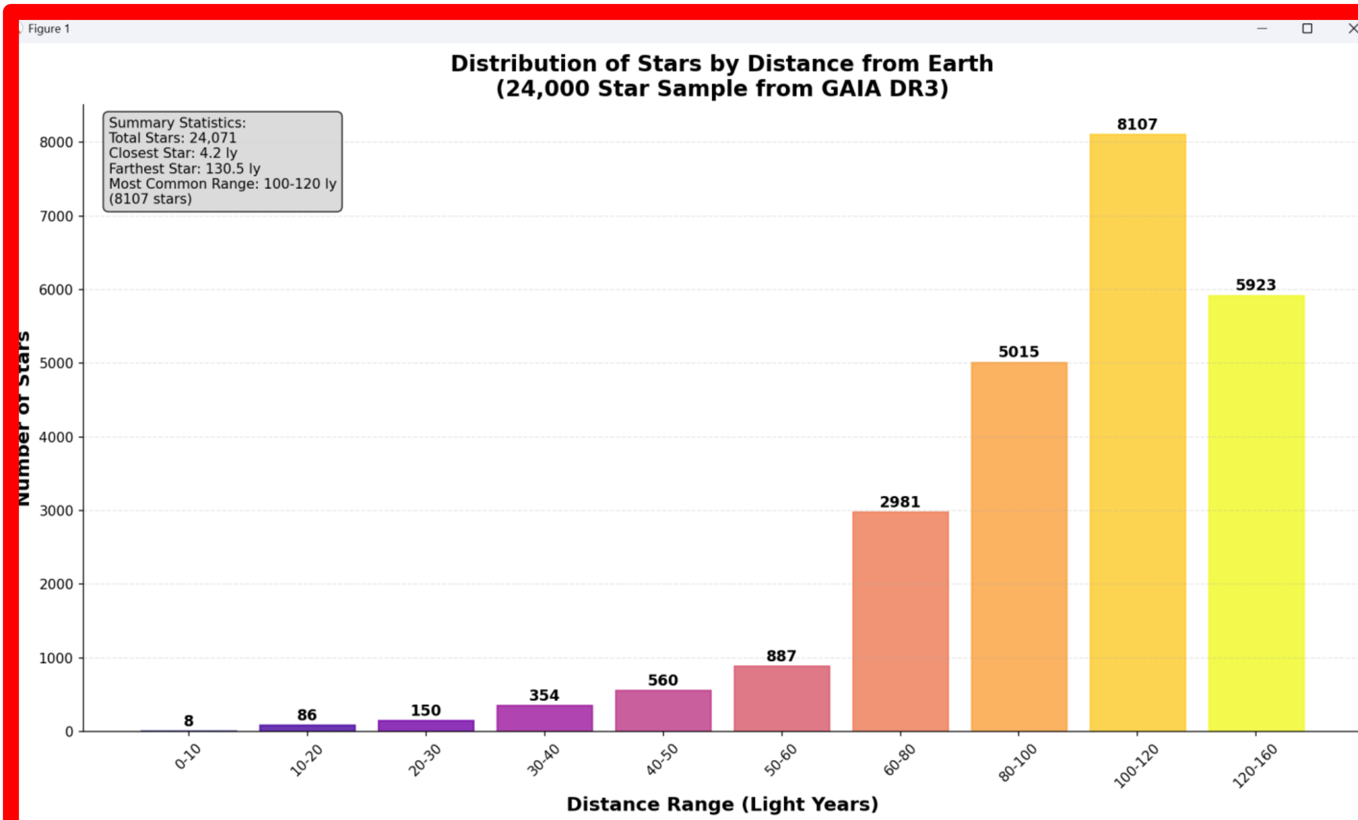
Very dim.............................16,795
Dim.....................................3,929
Low....................................2,089
Medium...............................1,049
High..172
Very High...................................2

Majority of stars are very dim and dim

Bar Chart - Distance

Results:

- Processed 24,000 stars
- Converted parallax to light-years
- Dense cluster of stars at 100 light years to 120 light years
- Sparse population within 10 light years of Earth



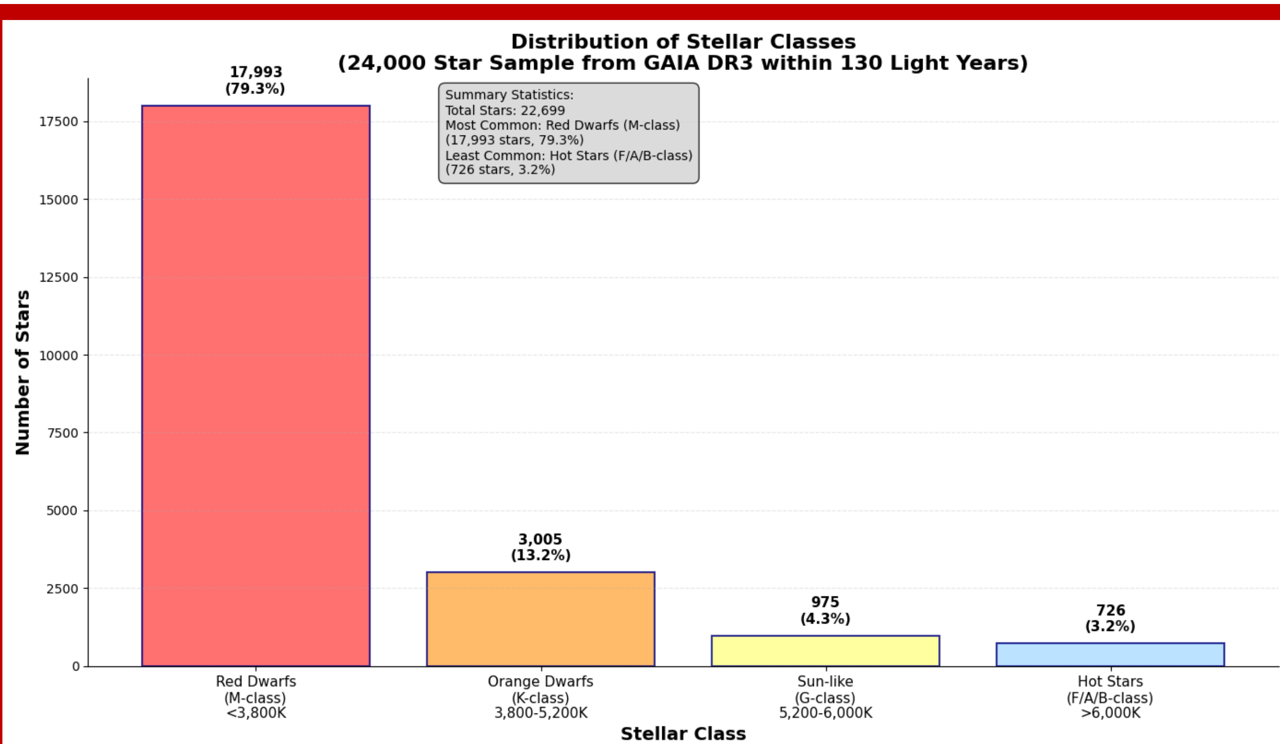
Distance conversion results:
Valid distances calculated: 24071
Min distance: 4.2 light years
Max distance: 130.5 light years

Star counts by distance range:
0-10 ly: 8 stars (0.0%)
10-20 ly: 86 stars (0.4%)
20-30 ly: 150 stars (0.6%)
30-40 ly: 354 stars (1.5%)
40-50 ly: 560 stars (2.3%)
50-60 ly: 887 stars (3.7%)
60-80 ly: 2981 stars (12.4%)
80-100 ly: 5015 stars (20.8%)
100-120 ly: 8107 stars (33.7%)
120-160 ly: 5923 stars (24.6%)

Bar Chart – Types of Stars

Results:

- Processed 24,000 stars
- Converted color to temperature
- Dense cluster of stars at 100 light years to 120 light years
- Sparse population within 10 light years of Earth

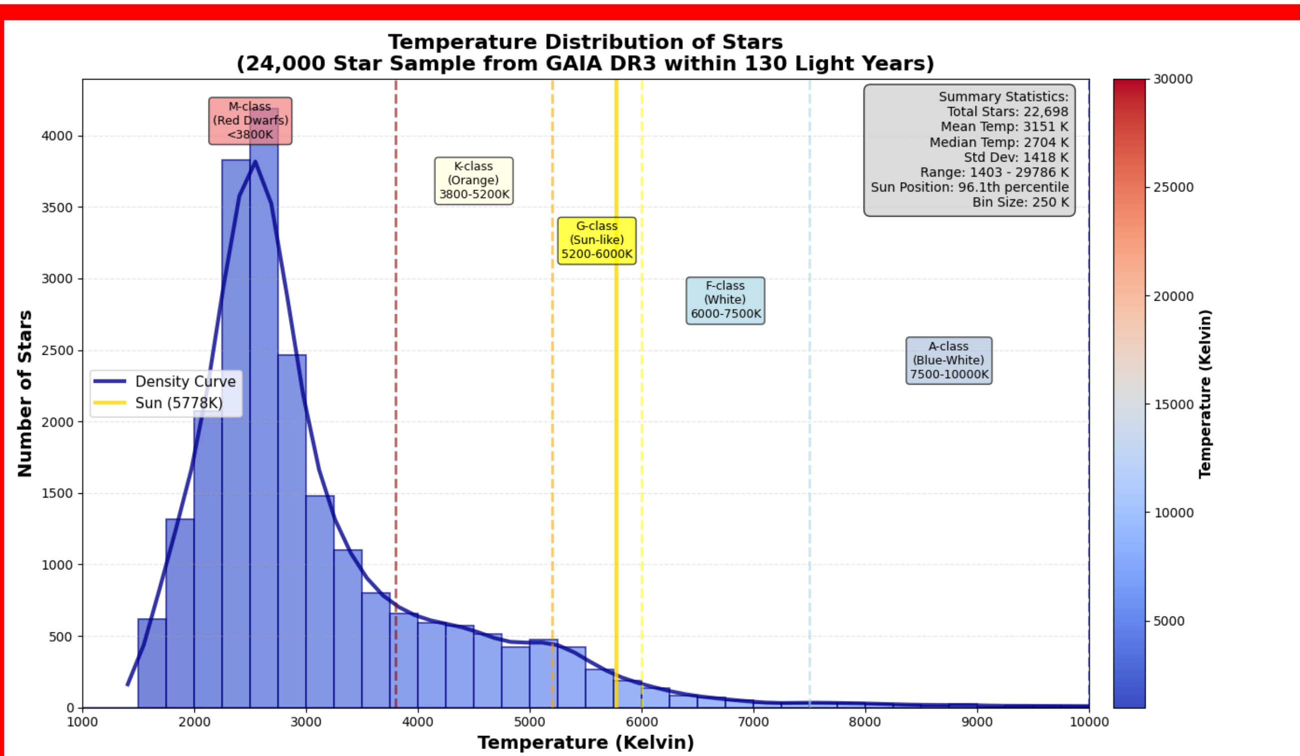


Star counts by class:
Red Dwarfs: 17,993 (79.3%)
Orange Dwarfs: 3,005 stars (13.2%)
Sunlike: 975 stars (4.3%)
Hot Stars: 726 (3.2%)

Distribution Chart – Temperature

Results:

- Processed 24,000 stars
- Converted color to Kelvin
- Vast majority of stars at 3,800K
- Our Sun is hotter than most stars
- Used scipy library in addition to matplotlib

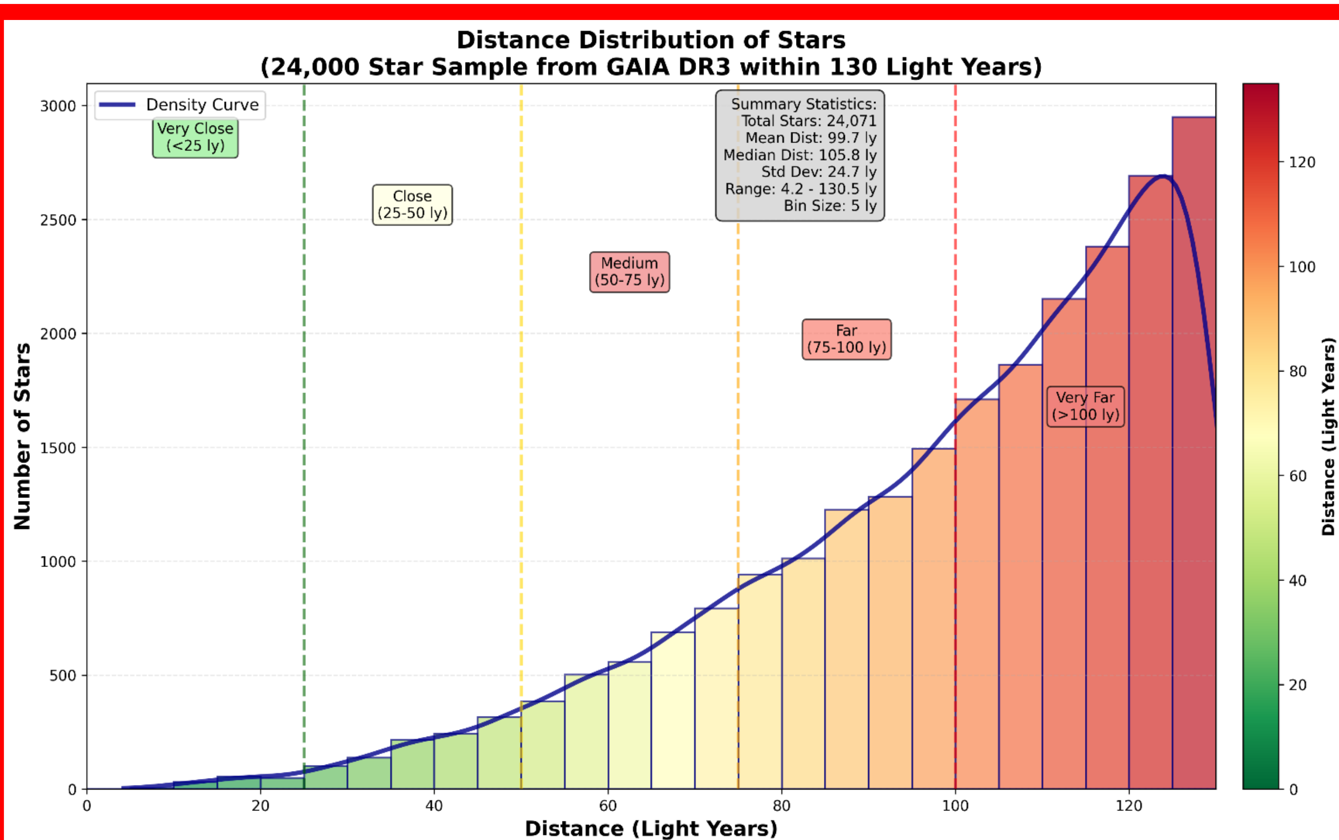


Kernel Density Estimate curve (KDE) in blue

Distribution Chart – Distance from Earth

Results:

- Processed 24,000 stars
- Converted parallax to light-years
- Dense cluster of stars at 100 light years to 130 light years
- Sparse population within 10 light years of Earth
- Used scipy library in addition to matplotlib

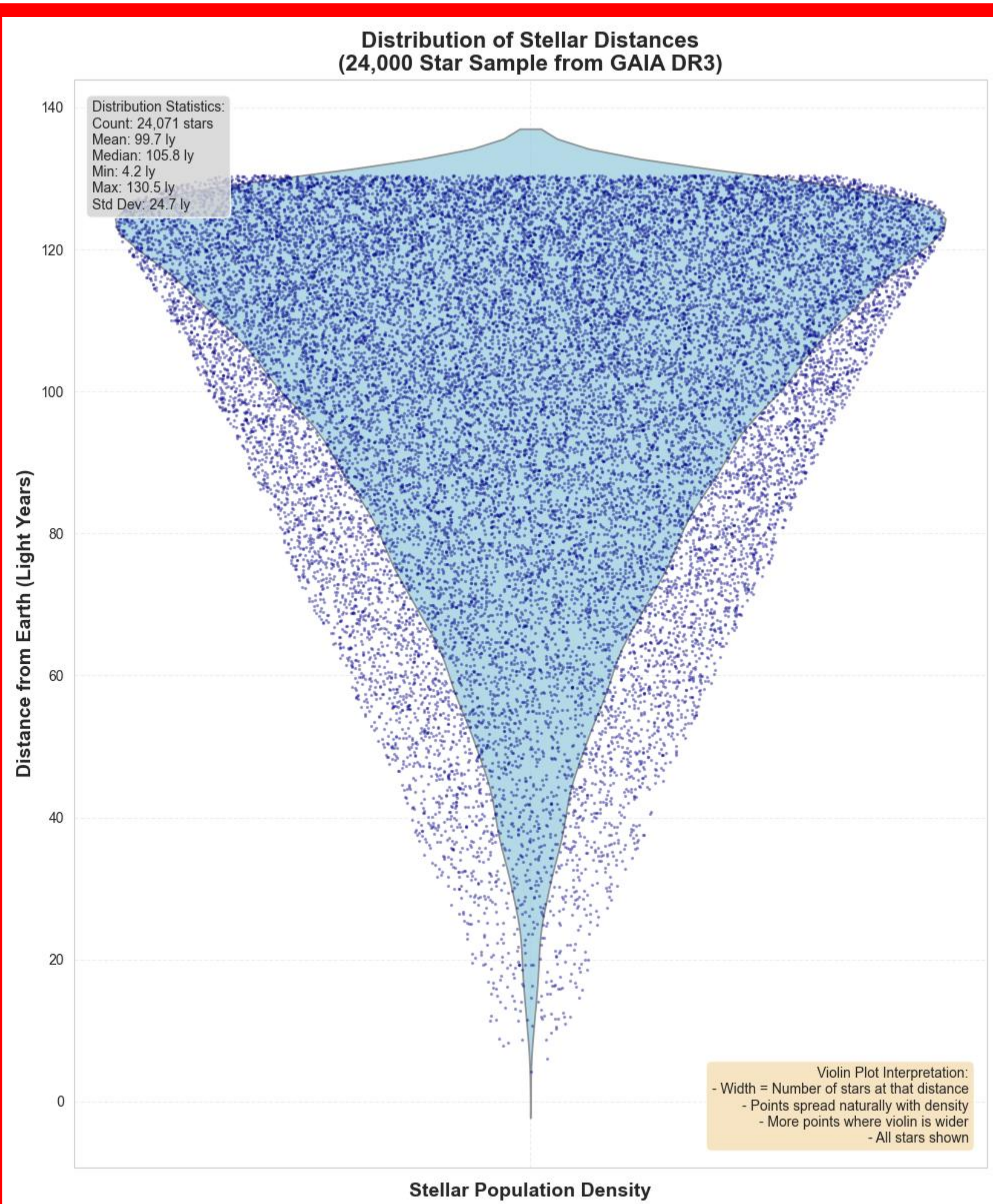


Kernel Density Estimate curve (KDE) in blue

Violin Plot – Distance from Earth

Results:

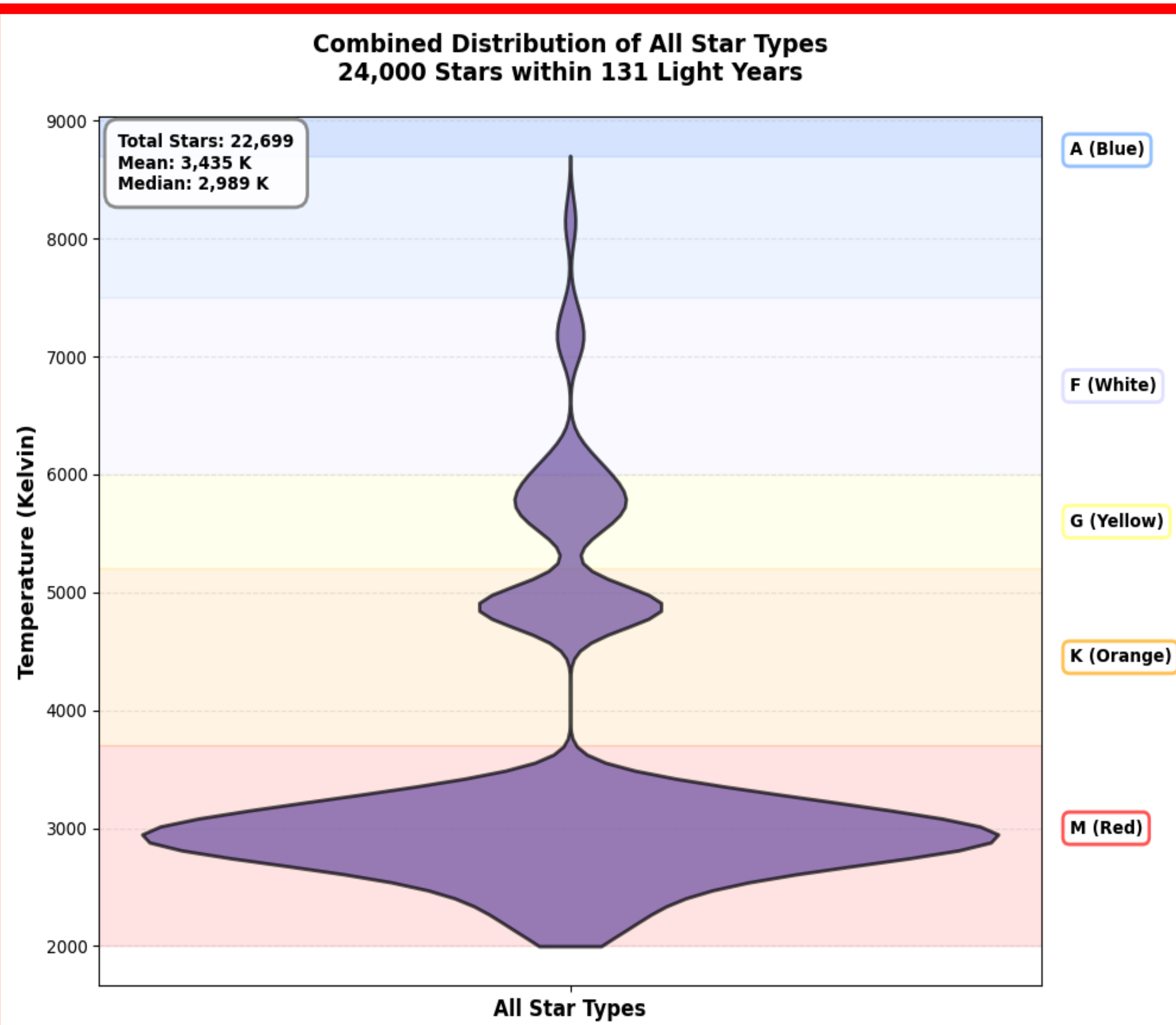
- Processed 24,000 stars
- Converted parallax to light-years
- Dense cluster of stars at 100 light years to 130 light years
- Sparse population within 10 light years of Earth
- Used Seaborn library in addition to matplotlib



Violin Plot – Temperature

Results:

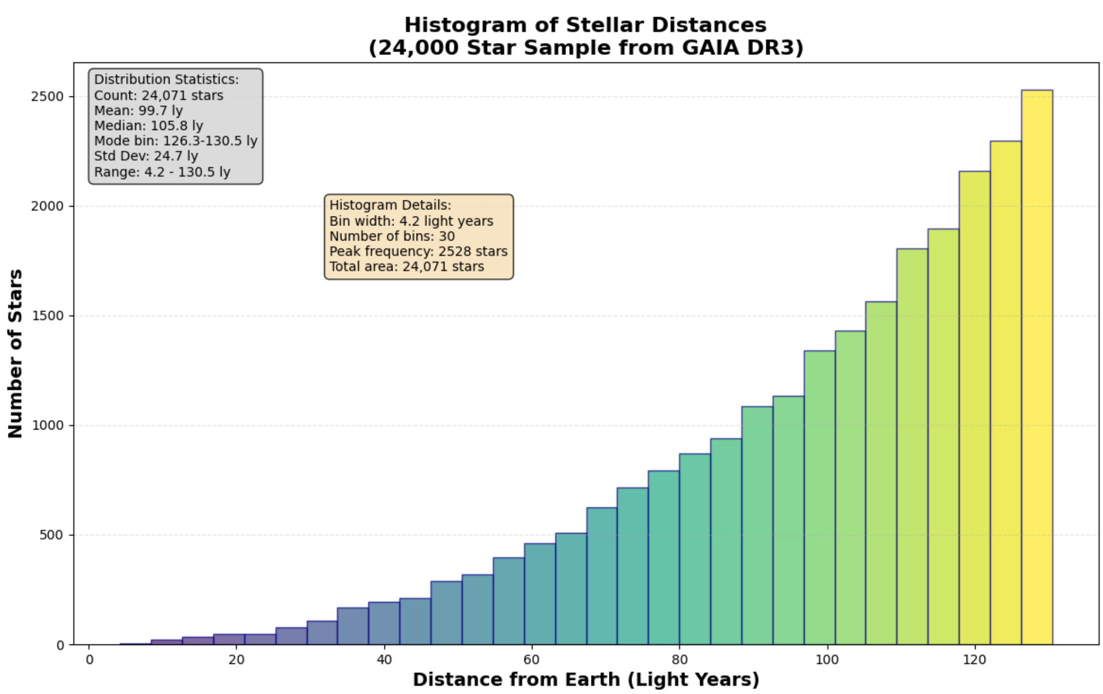
- Processed 24,000 stars
- Converted parallax to light years
- Converted color to temperature in Kelvim
- Dense cluster of Red Dwarf stars at 3,000 Kelvin



Histogram - Distance

Results:

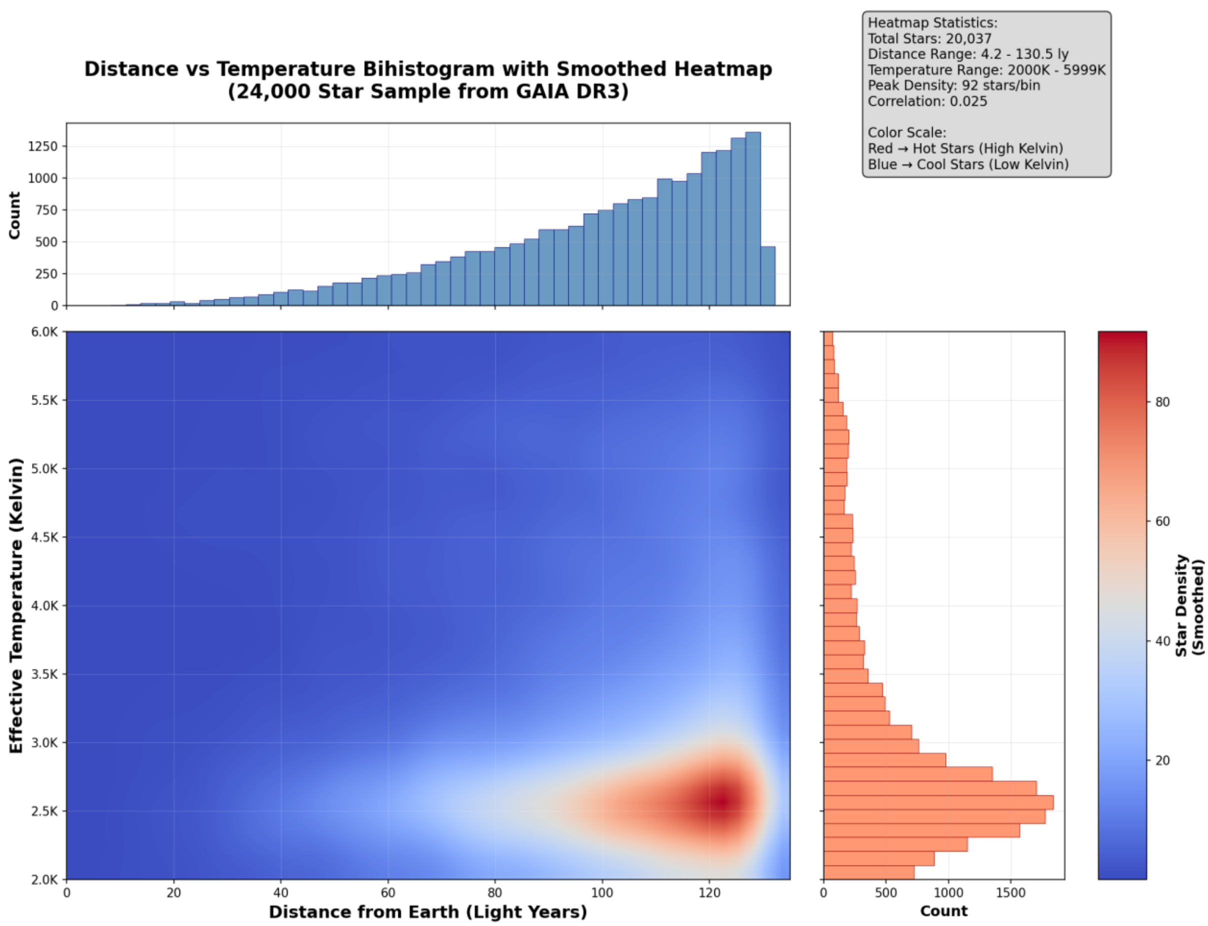
- Processed 24,000 stars
- Converted parallax to light years
- 30 bins at 4.2 light years per bin
- Peak density at 100 to 120 light years
- Very few stars in bins close to Earth
- Exponential increase of stars at about 100 light years



Bi-Histogram with Heat Map

Results:

- Processed 24,000 stars
- Converted parallax to light years
- Converted color to temperature in Kelvin
- Dense cluster of Red Dwarf stars at 2,500 Kelvin
- Peak density at 100 to 120 light years
- Used Scipy library

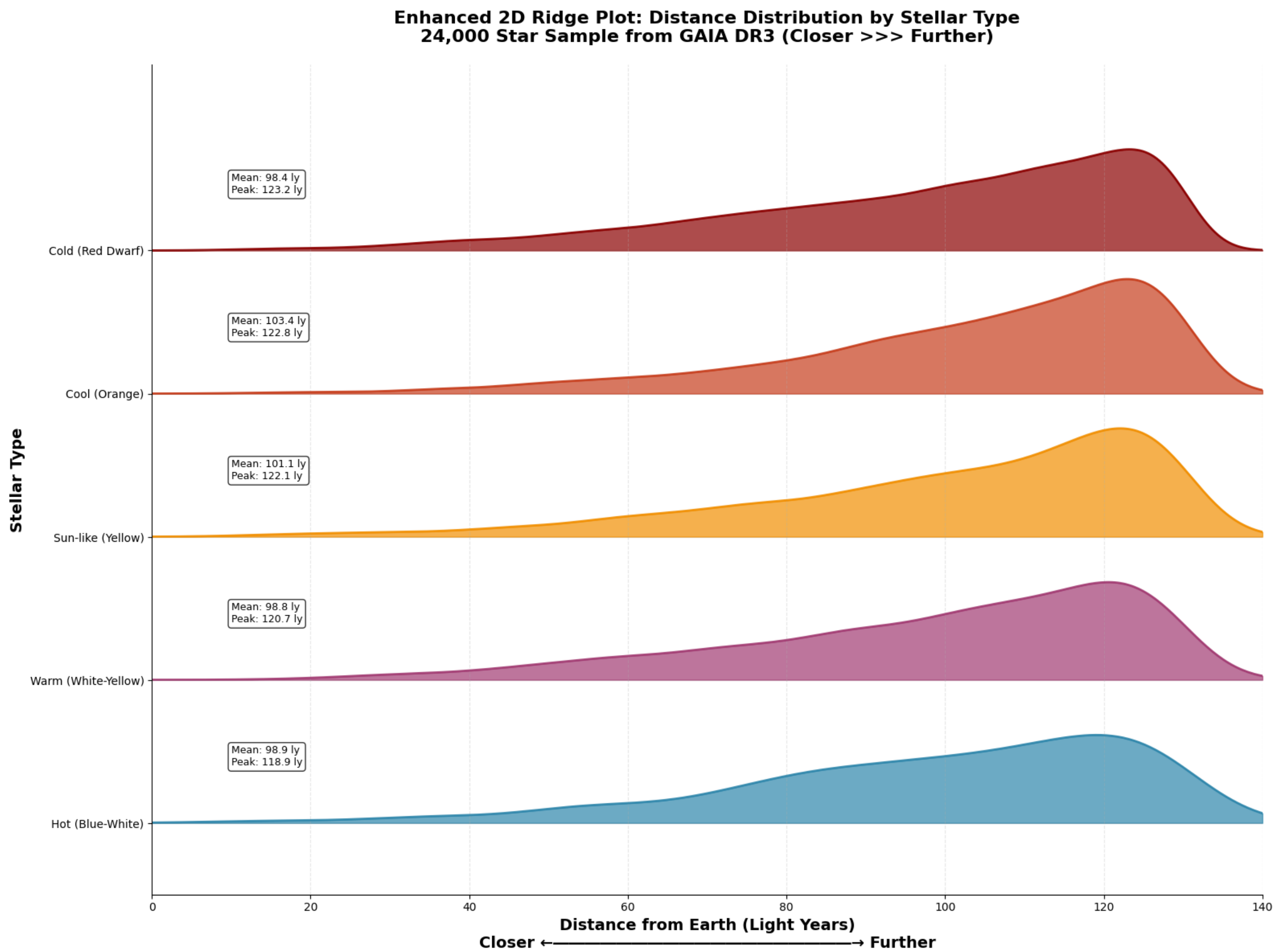


SciPy

2D Ridge Plots

Results:

- Processed 24,000 stars
- Converted color to temperature in Kelvin
- Converted parallax to light years
- Peak density of all star classes at 100 to 120 light years
- Used Scipy library

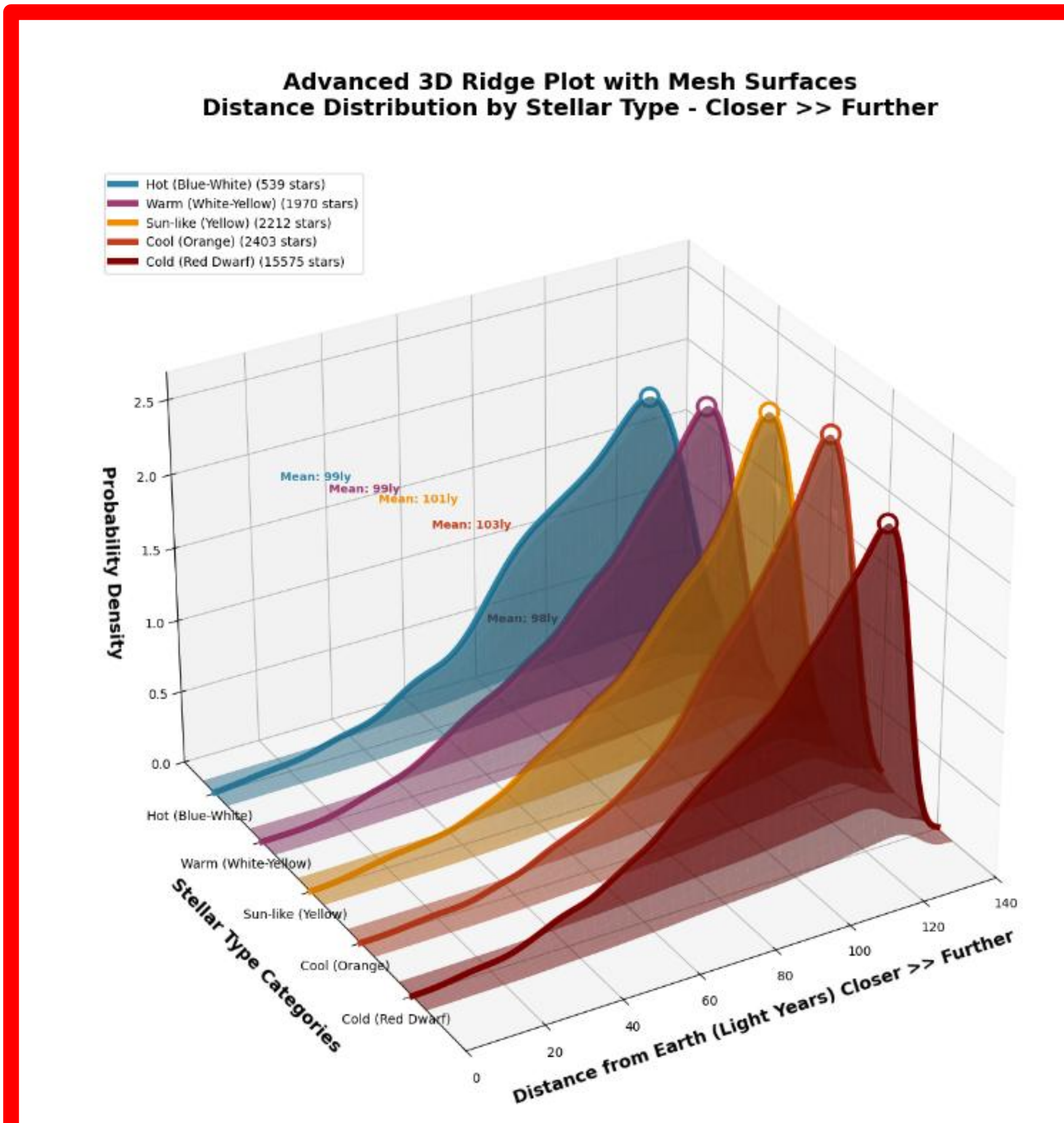


SciPy

3D Ridge Plots

Results:

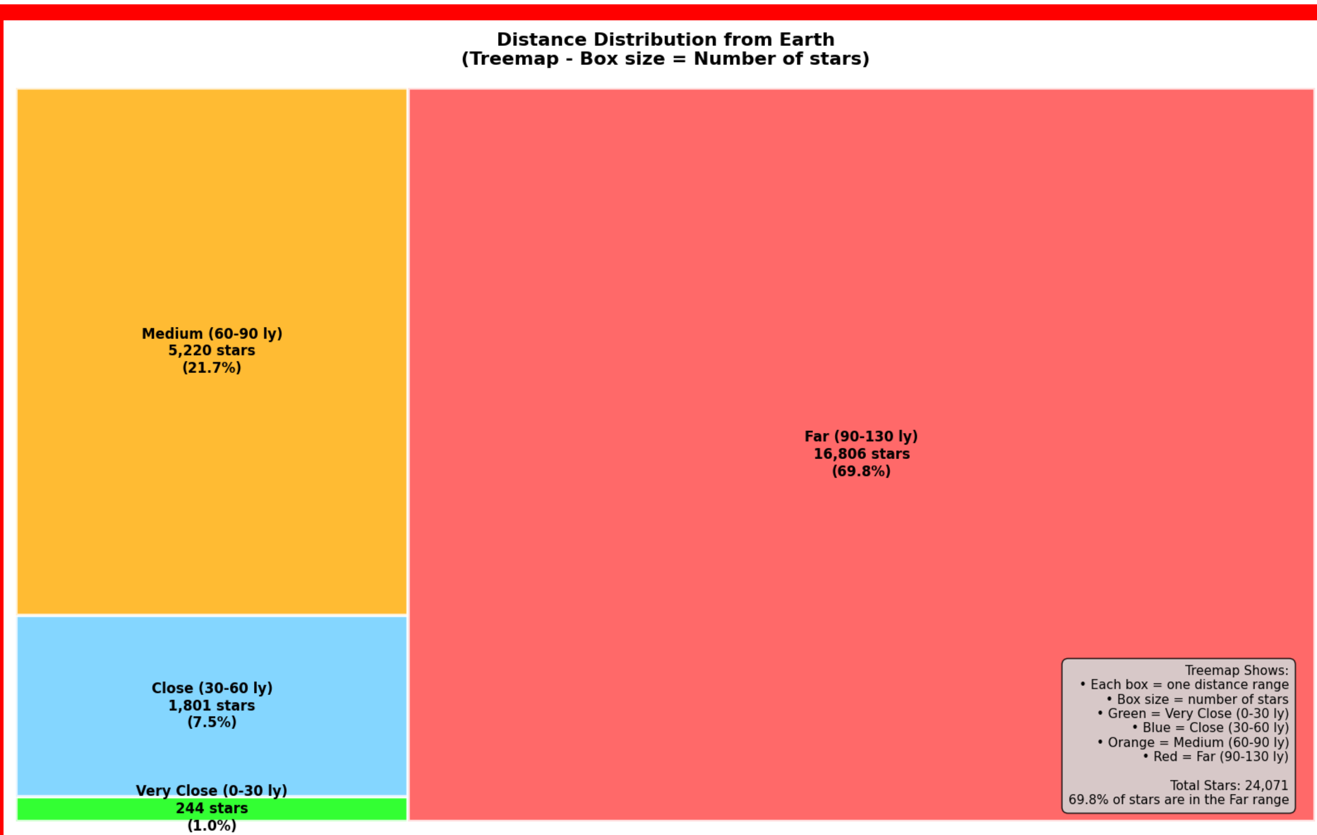
- Processed 24,000 stars
- Converted color to temperature in Kelvin
- Converted parallax to light years
- Peak density of all star classes at 100 to 120 light years
- Used Scipy library



Treemap Distance

Results:

- Processed 24,000 stars
- Converted parallax to light years
- Majority of stars at 90 to 120 light years
- Used Squarify library

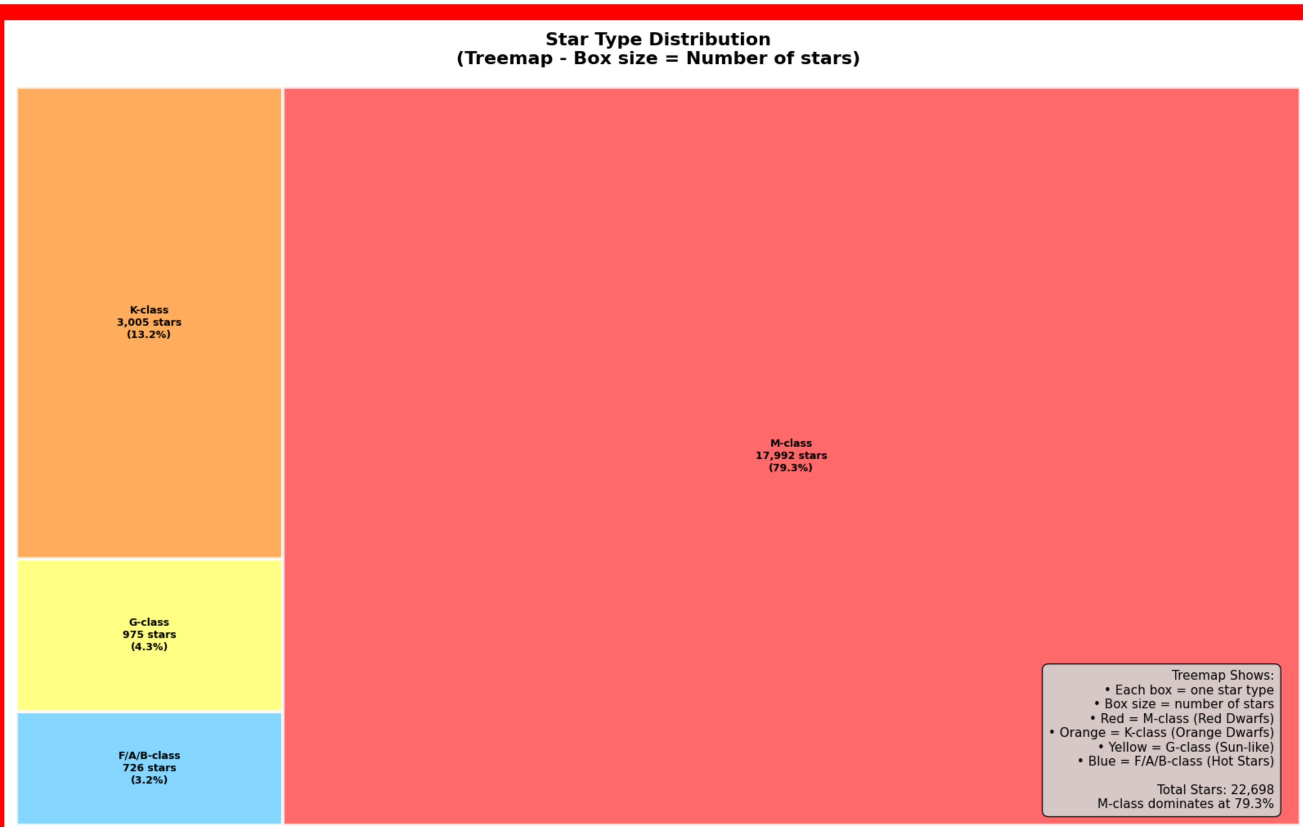


squarify

Treemap Star Class

Results:

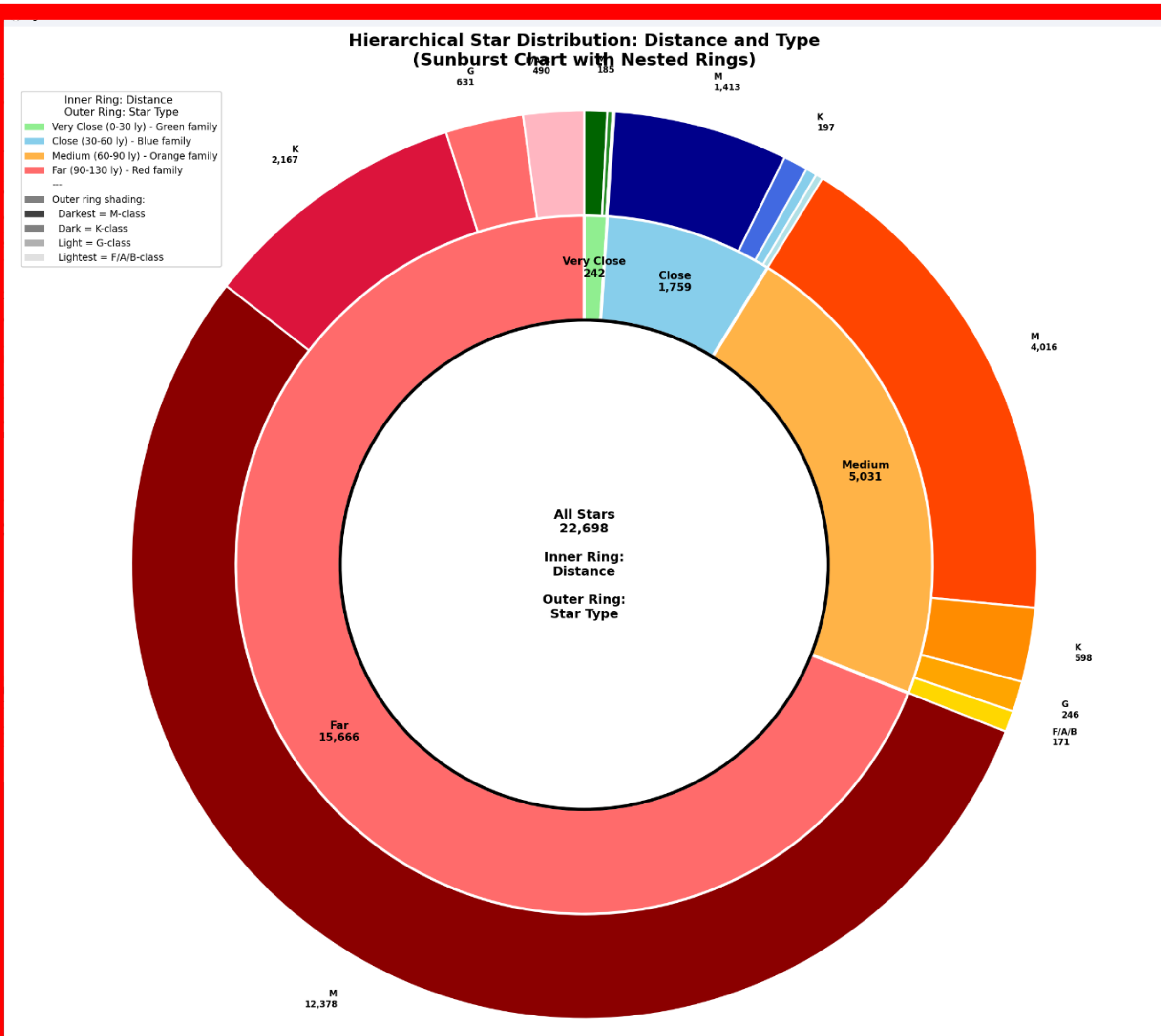
- Processed 24,000 stars
- Converted color to temperature to star class
- Majority of stars are M-class, Red Dwarf
- Used Squarify library



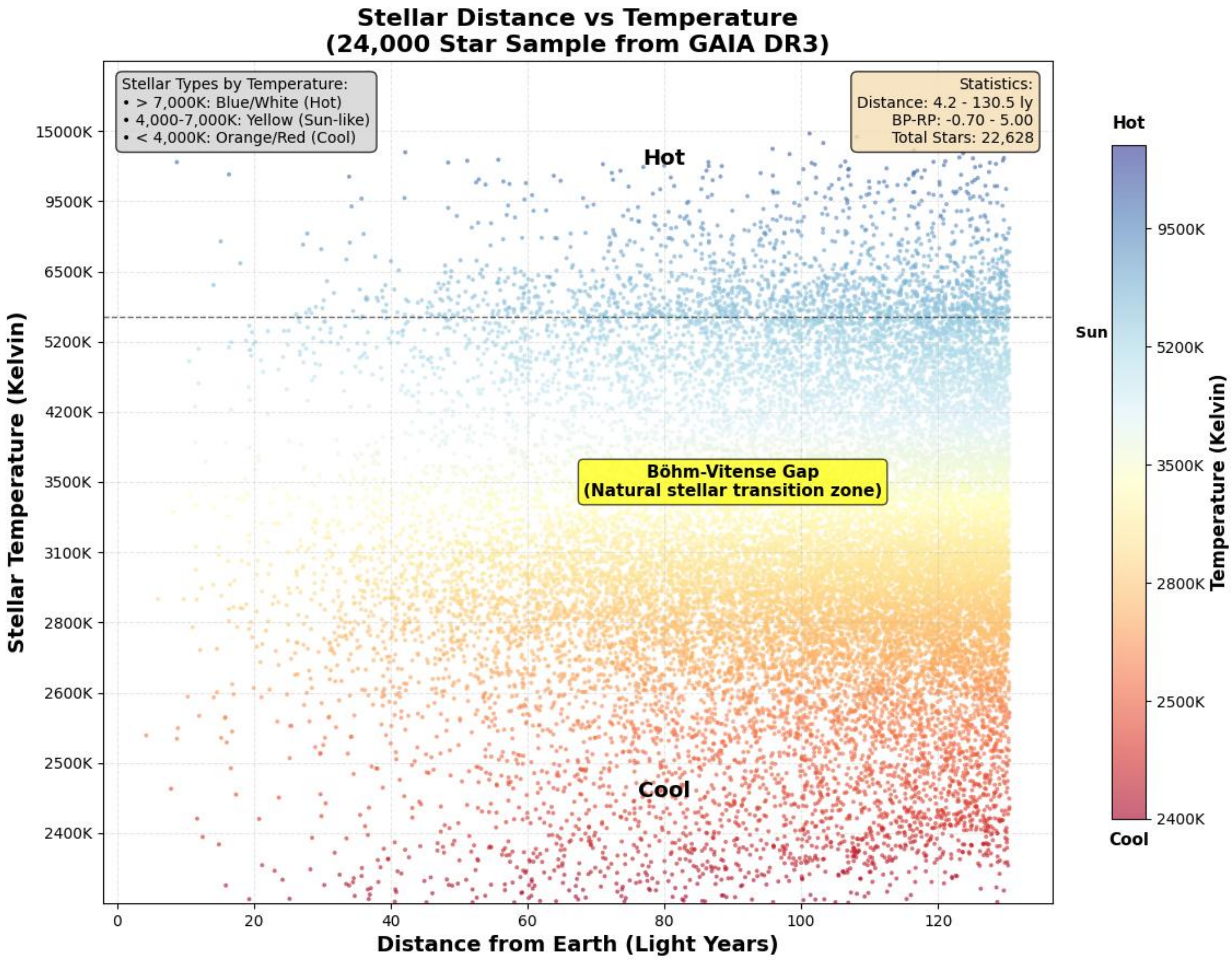
Sunburst – Distance and Star Class

Results:

- Processed 24,000 stars
- Converted color to temperature to star class
- Converted parallax to light years
- Majority of stars are M-class, Red Dwarf
- Large majority are far from Earth



Scatterplot:
Distance vs Temperature



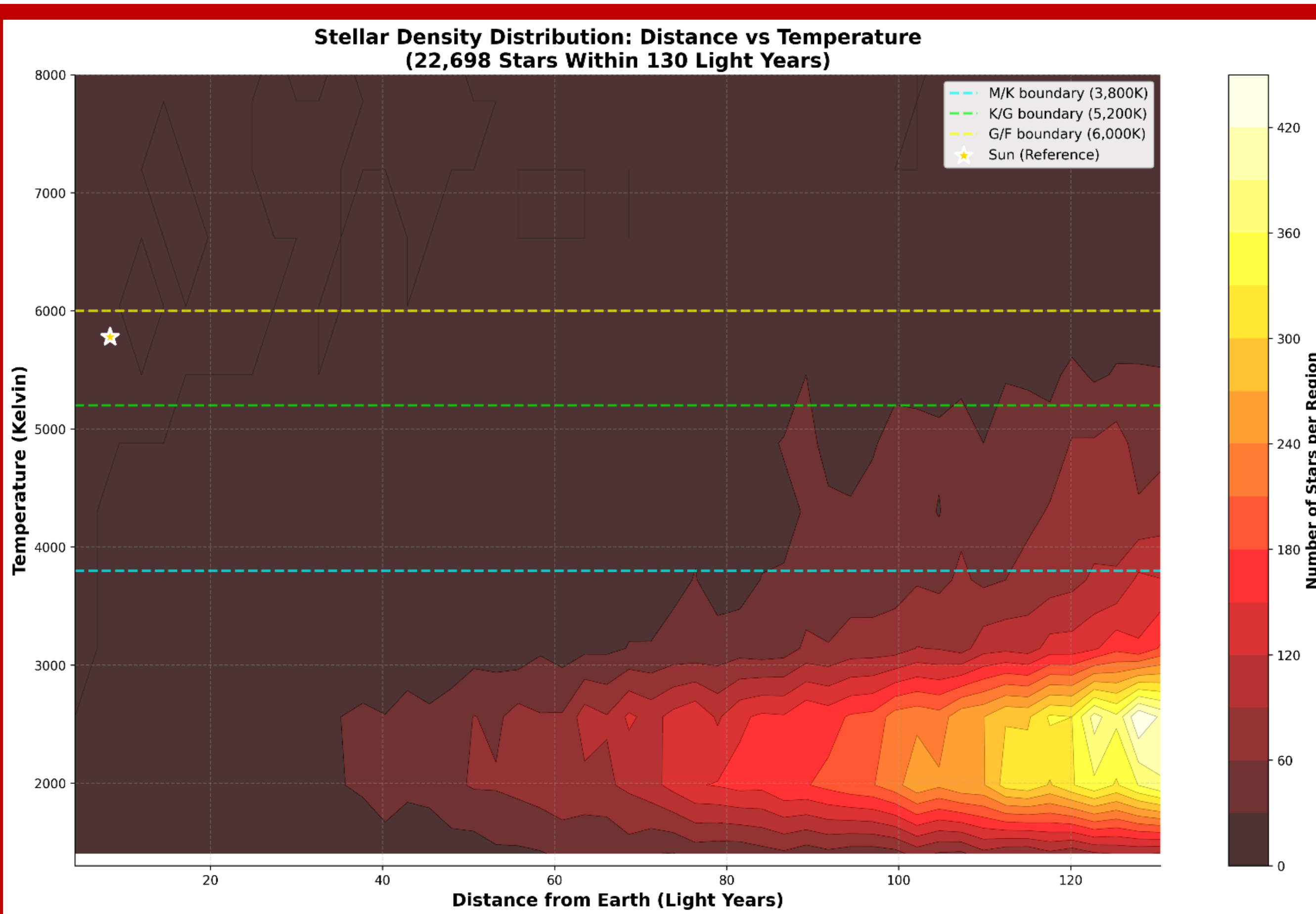
Derived Values:
Distance in Light Years: Derived from parallax
Temperature in Kelvin: Derived from color index

Hypothesis #3: The scatterplot indicates that stellar density increases further away from Earth
There are fewer stars closer to Earth

Contour – Distance vs. Temperature

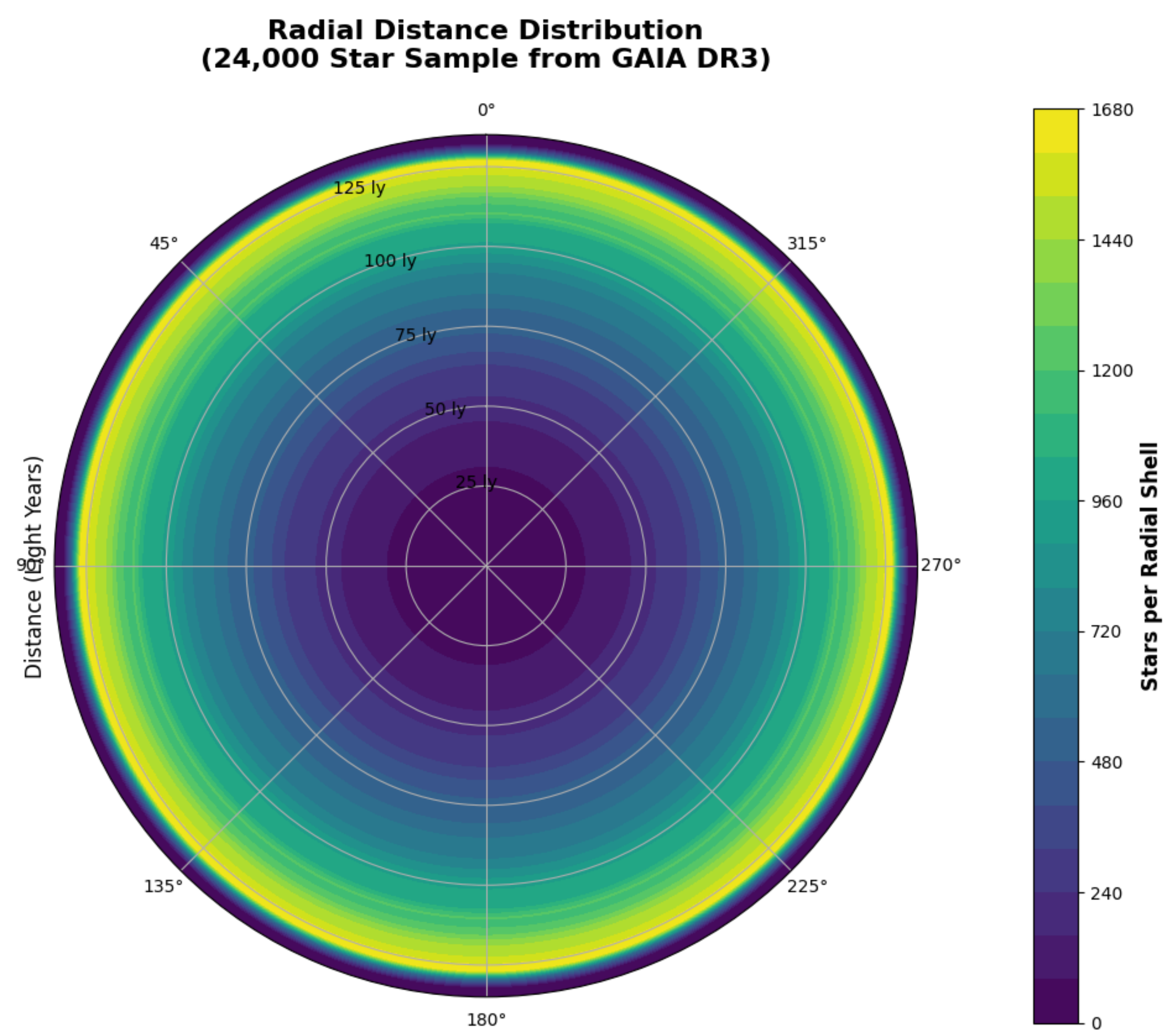
Results:

- Processed 22,000 stars
- Converted parallax to light years
- Large majority are far from Earth
- Density stars gets greater the further away
- Temperatures largely concentrated in M-class



Radial Heatmap – Distance

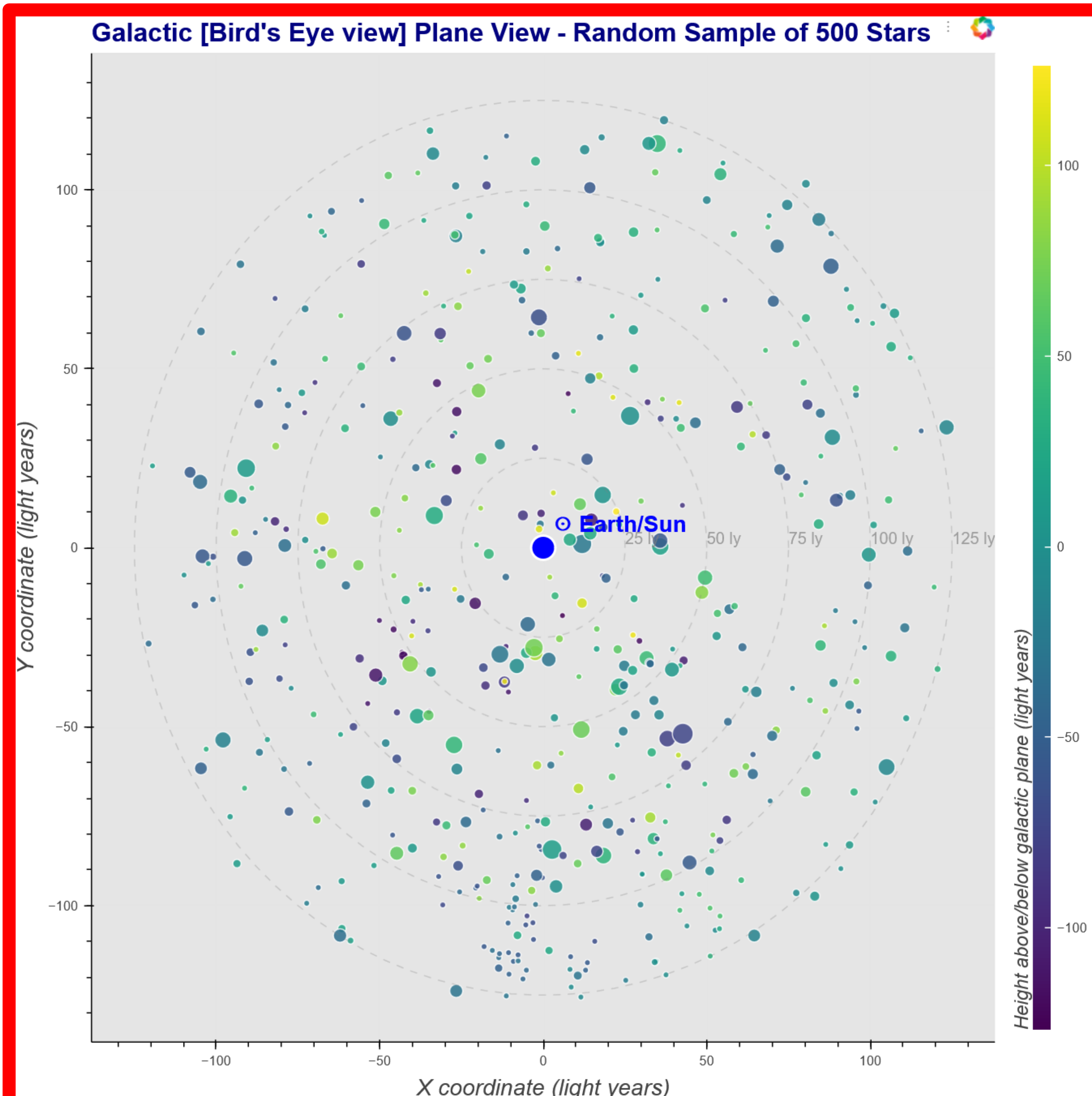
Results:
- Processed 24,000 stars
- Converted parallax to light years
- Large majority are far from Earth
- Density stars gets greater the further away



Bokeh 2D – Bird's Eye View

Results:

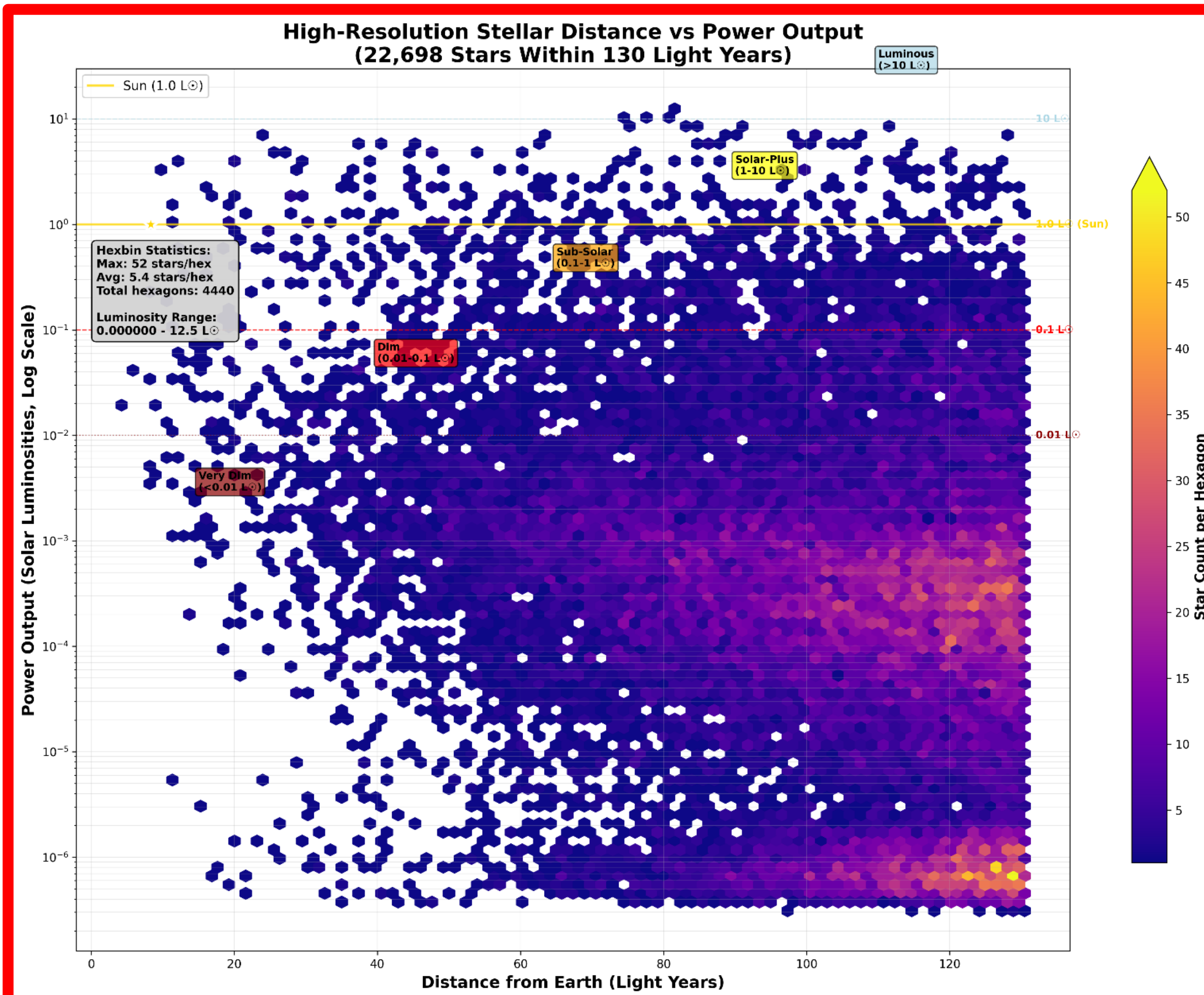
- Random sample of 500 stars
- Converted parallax to light years
- Large majority are far from Earth
- Density stars gets greater the further away



Hexbin – Distance vs. Power Output

Results:

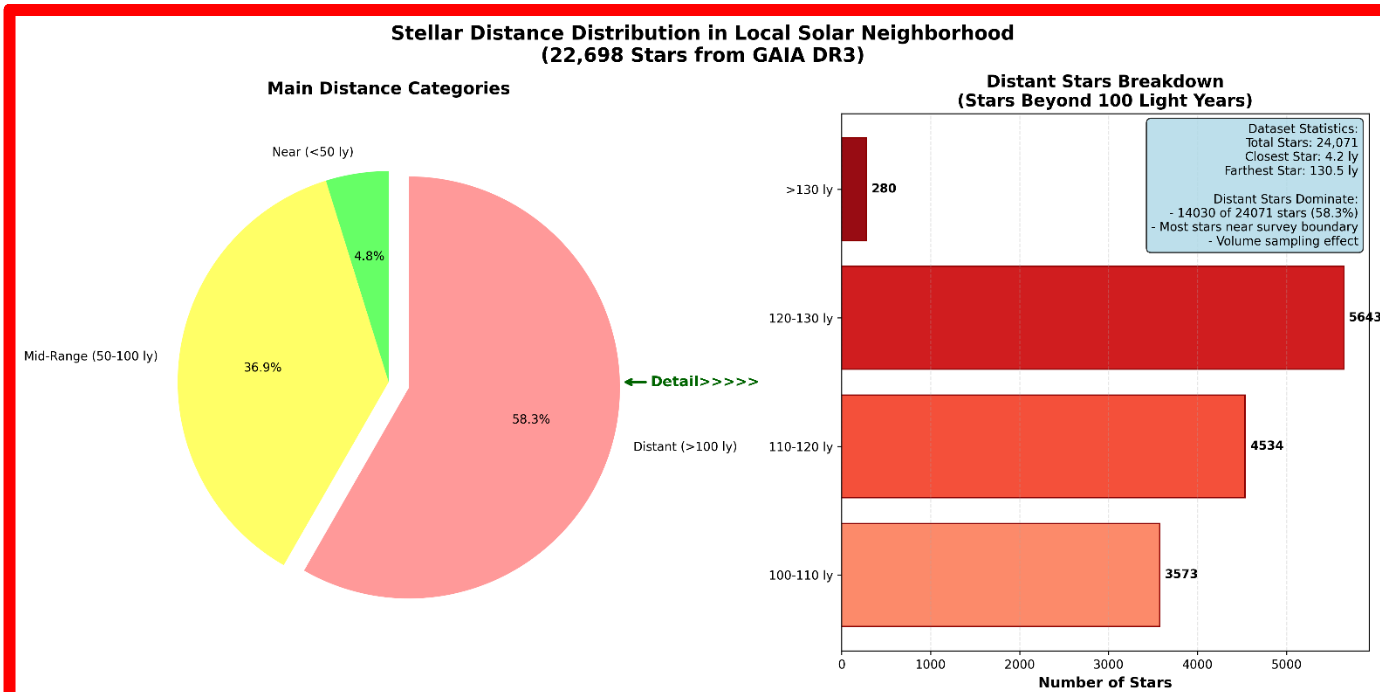
- Processed 22,000 stars
- Converted parallax to light years
- Large majority are far from Earth
- Density stars gets greater the further away
- Majority are Very Dim and Dim M-class stars



Bar of Pie Chart – Distance

Results:

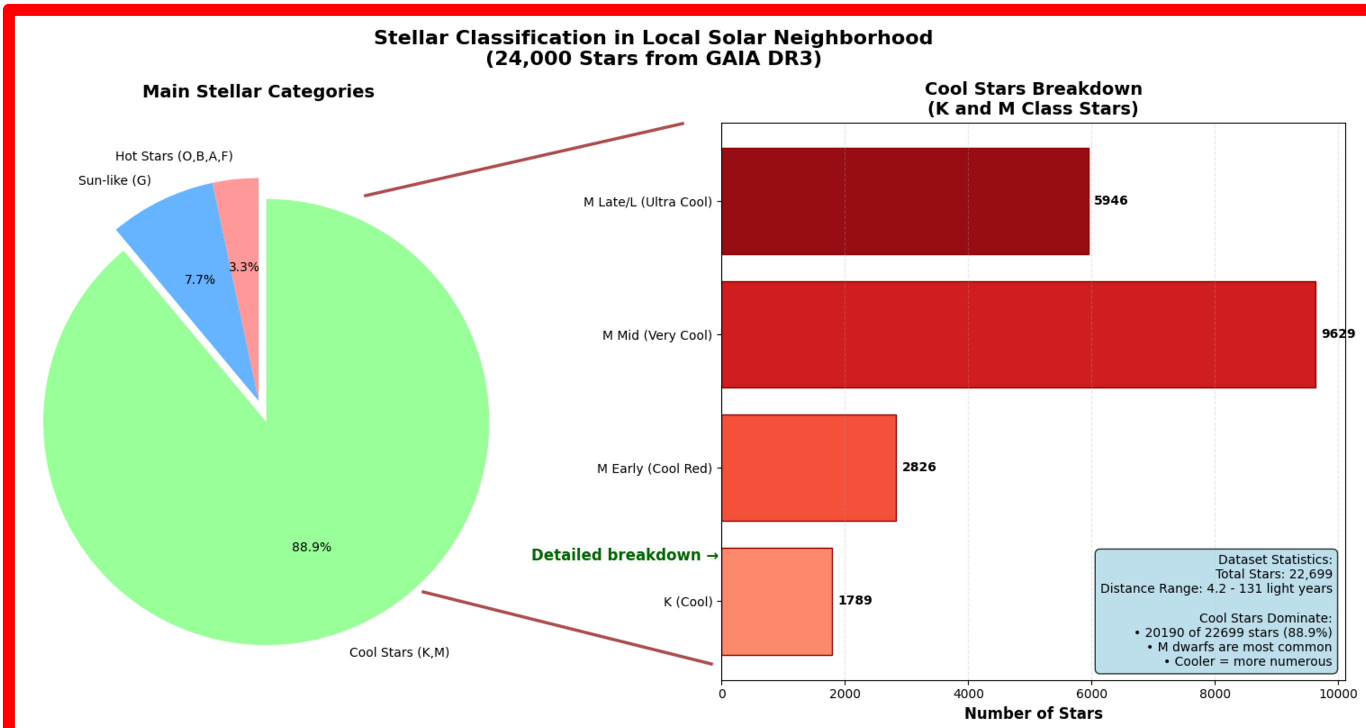
- Sample of 22,000 stars
- Converted parallax to light years
- Large majority are far from Earth
- Density stars gets greater the further away
- Highest density at 120 to 130 light years



Bar of Pie Chart – Class

Results:

- Sample of 24,000 stars
- Converted color to temperature in Kelvin
- Large majority are cool stars
- Within that set, a majority are Very Cool and Ultra Cool



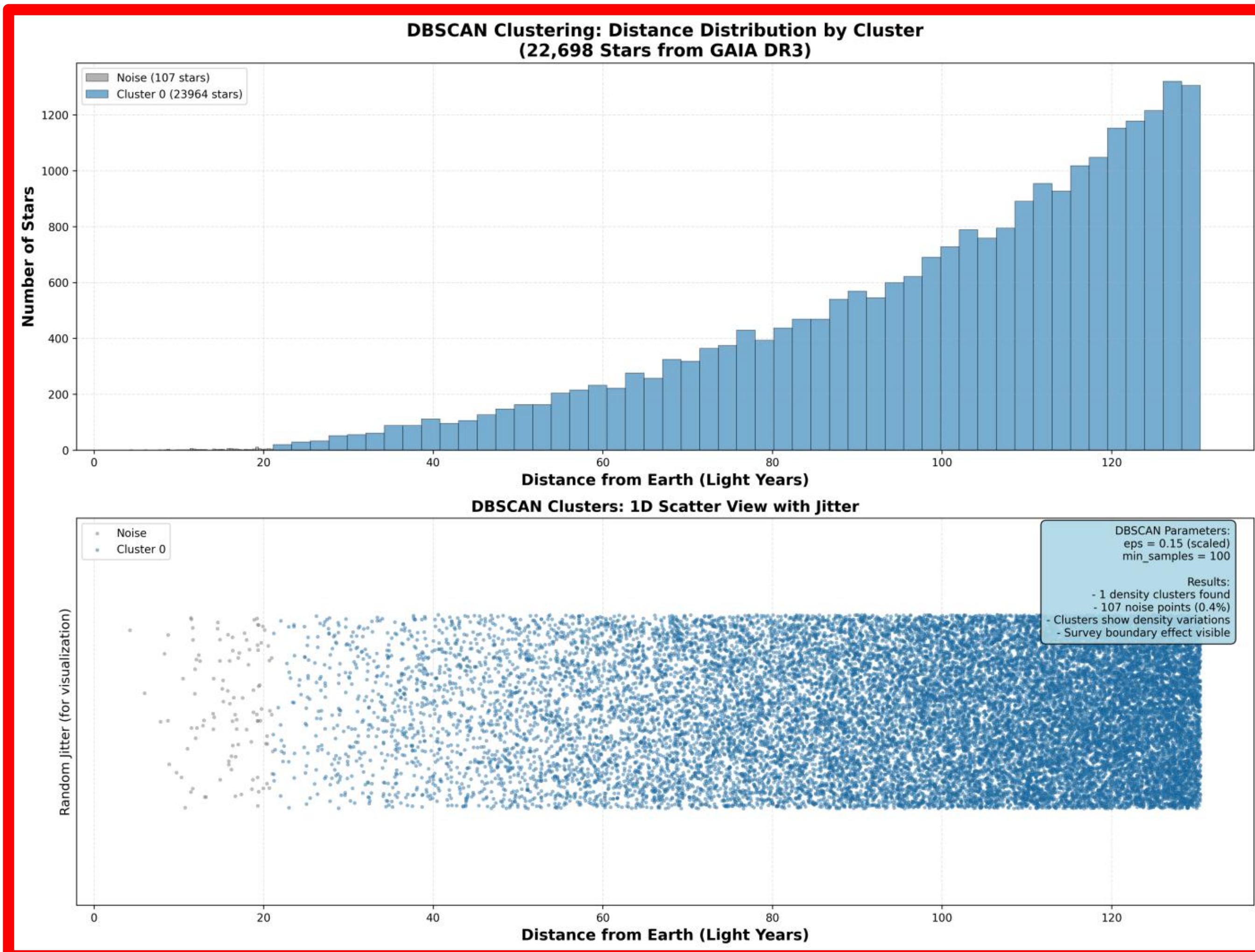
DBSCAN – Distance

Results:

- Sample of 22,000 stars
- Converted parallax to light-years
- Large majority are beyond 100 light years
- Area close to Earth is relatively clear of stars at .5%

Density Based Spatial Clustering of Applications with Noise

“Noise” in this case are isolated stars



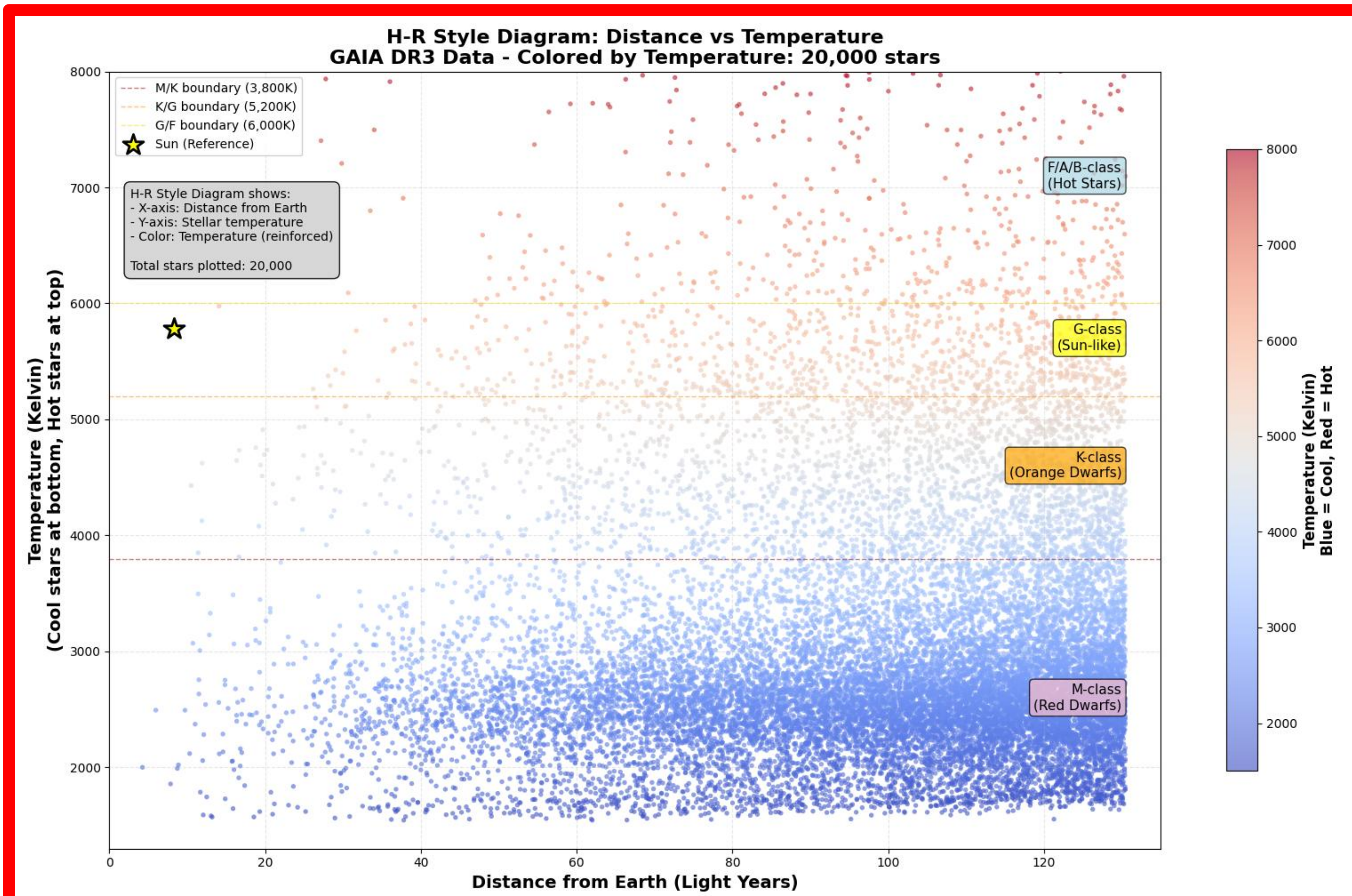
scikit-learn

Machine Learning in Python

H-R Diagram – Distance vs Temperature [Modified]

Results:

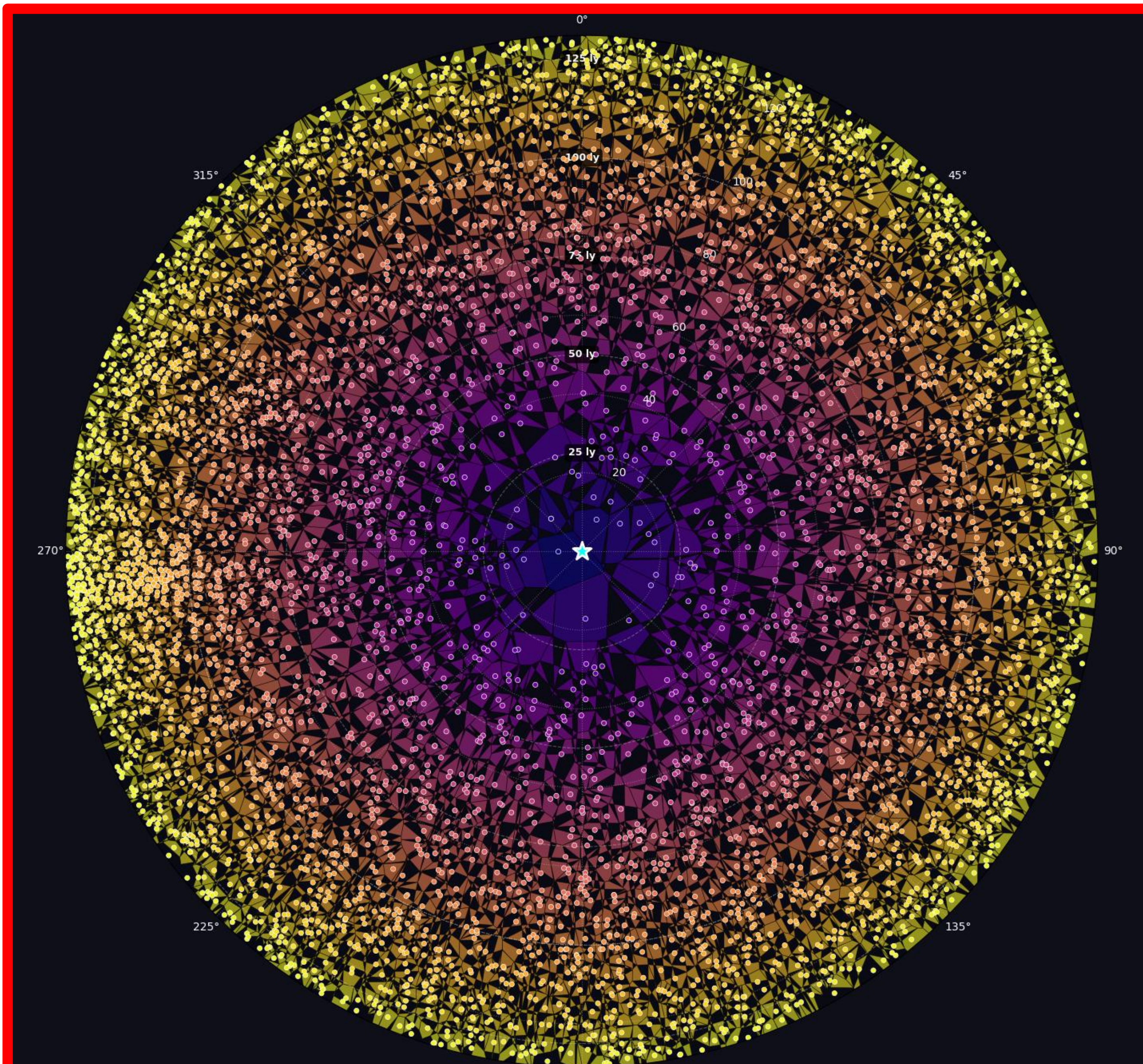
- Sample of 20,000 stars
- Converted parallax to light-years
- Density starts to increase at about 80 to 100 light years
- Area close to Earth has low stellar density
- M-class stars dominate



Radial Voronoi – Distance vs Density

Results:

- Sample of 5,000 stars
- Converted parallax to light-years
- Size of shape indicates density
- Density starts to increase at about 80 to 100 light years
- Area close to Earth has low stellar density



SciPy

Bubble – Stellar Size Comparison

Results:

- No sampling
- 1 R [radius] = our Sun
- Vast majority of stars at M-class [smallest]
- Largest star [O-class] is 15 times the size of our Sun
- Largest star is 50 times larger than smallest

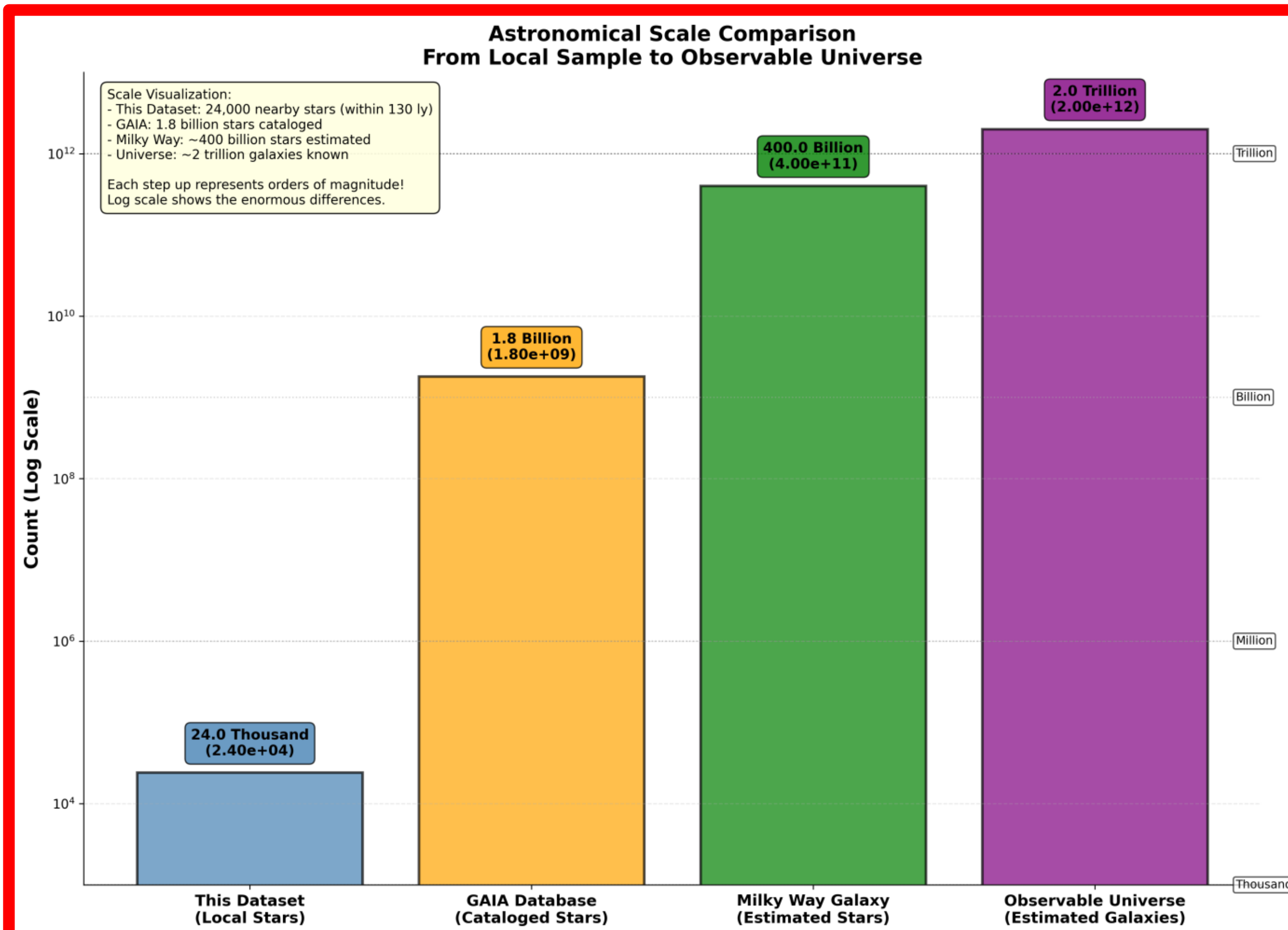
Log Chart – Stellar Scale Comparison

Results:

- Sample size: 24,000, 1.8 billion and 400 billion stars
- Project sample: 24,000 stars
- GAIA database: 1.8 billion stars
- Our galaxy: 400 billion stars
- Current number of galaxies in known universe: 2 trillion

Percentages:

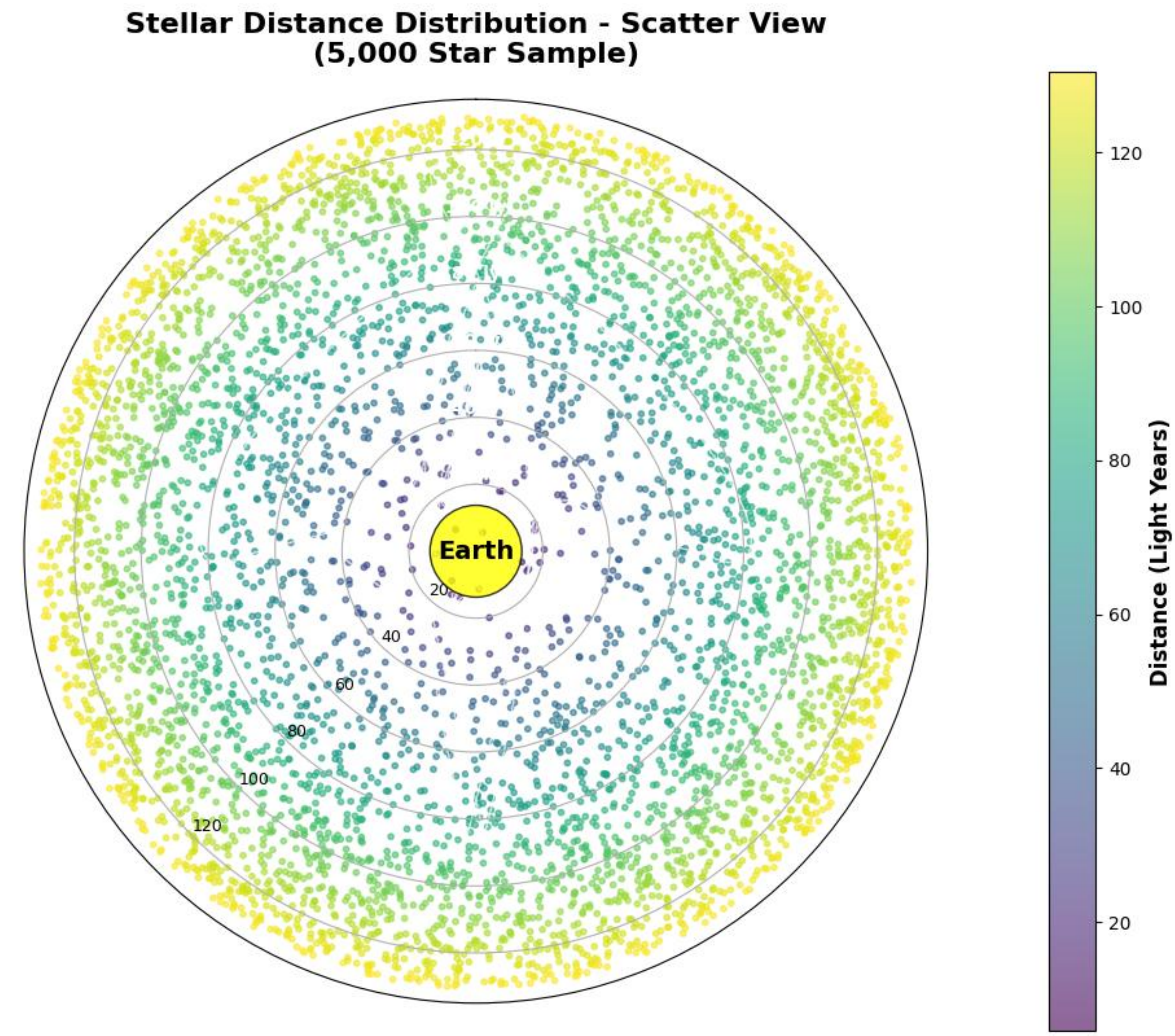
- Sample size vs GAIA total: .0013%
- Sample size vs Galaxy: .000006%
- GAIA total vs Galaxy: .45%
- Our Galaxy vs Total # Galaxies: 1 in 2 trillion



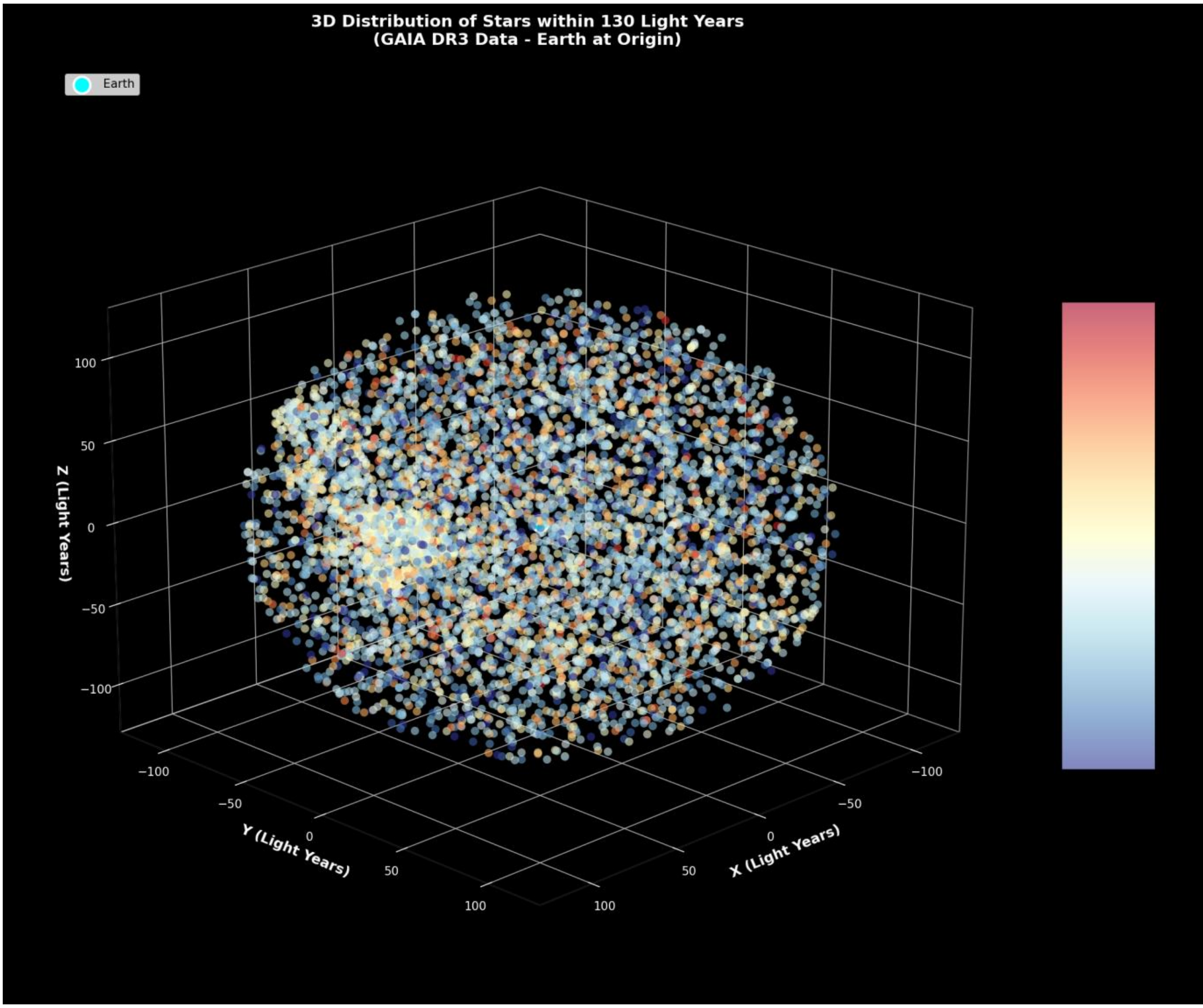
How small do you feel?

Polar Chart – Distance

Results:
- Sample of 5,000 stars
- Converted parallax to light-years
-- Density starts to increase at about 80 to 100 light years
- Area close to Earth has low stellar density

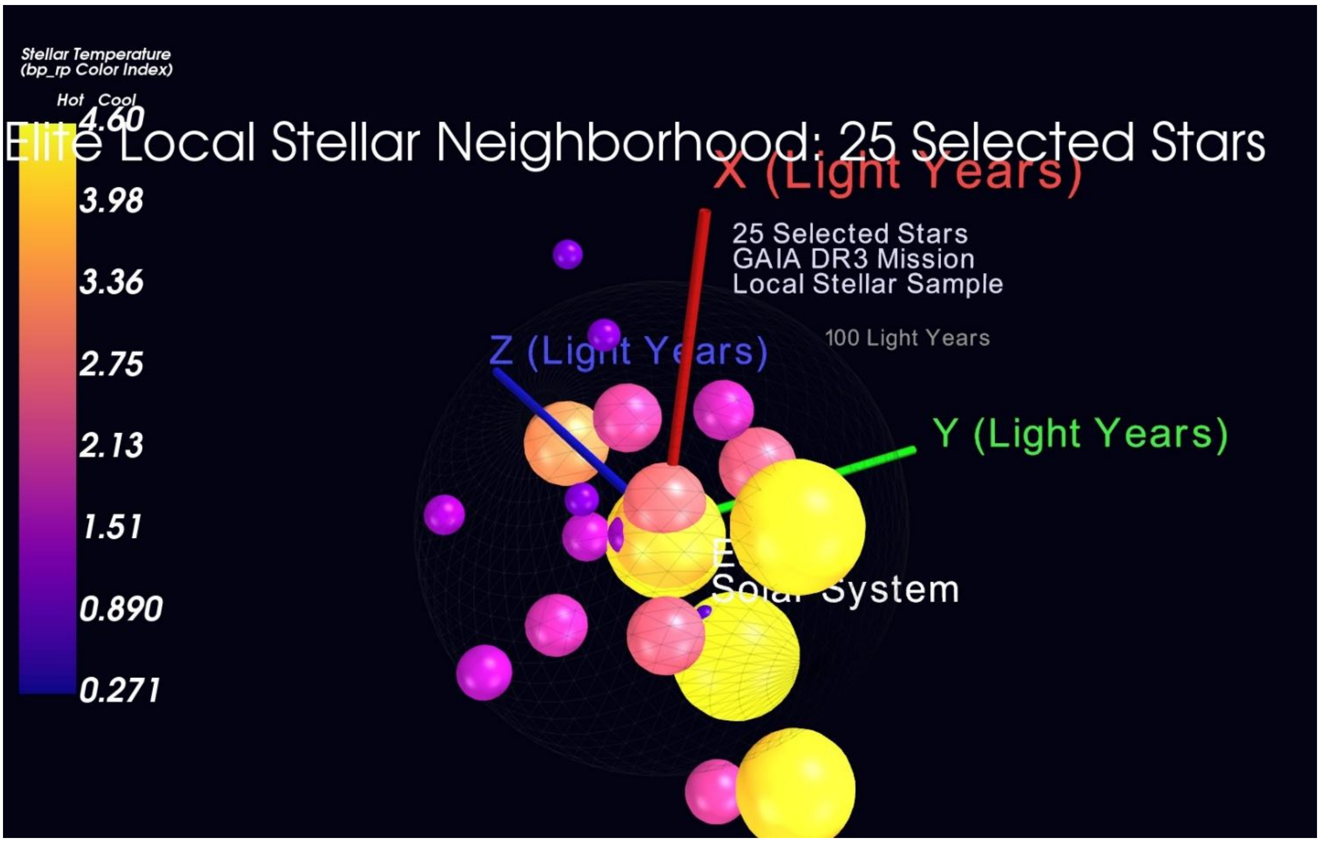


Geospatial
X, Y, Z Coordinates derived from:
Right Ascension, Declination and Light Years



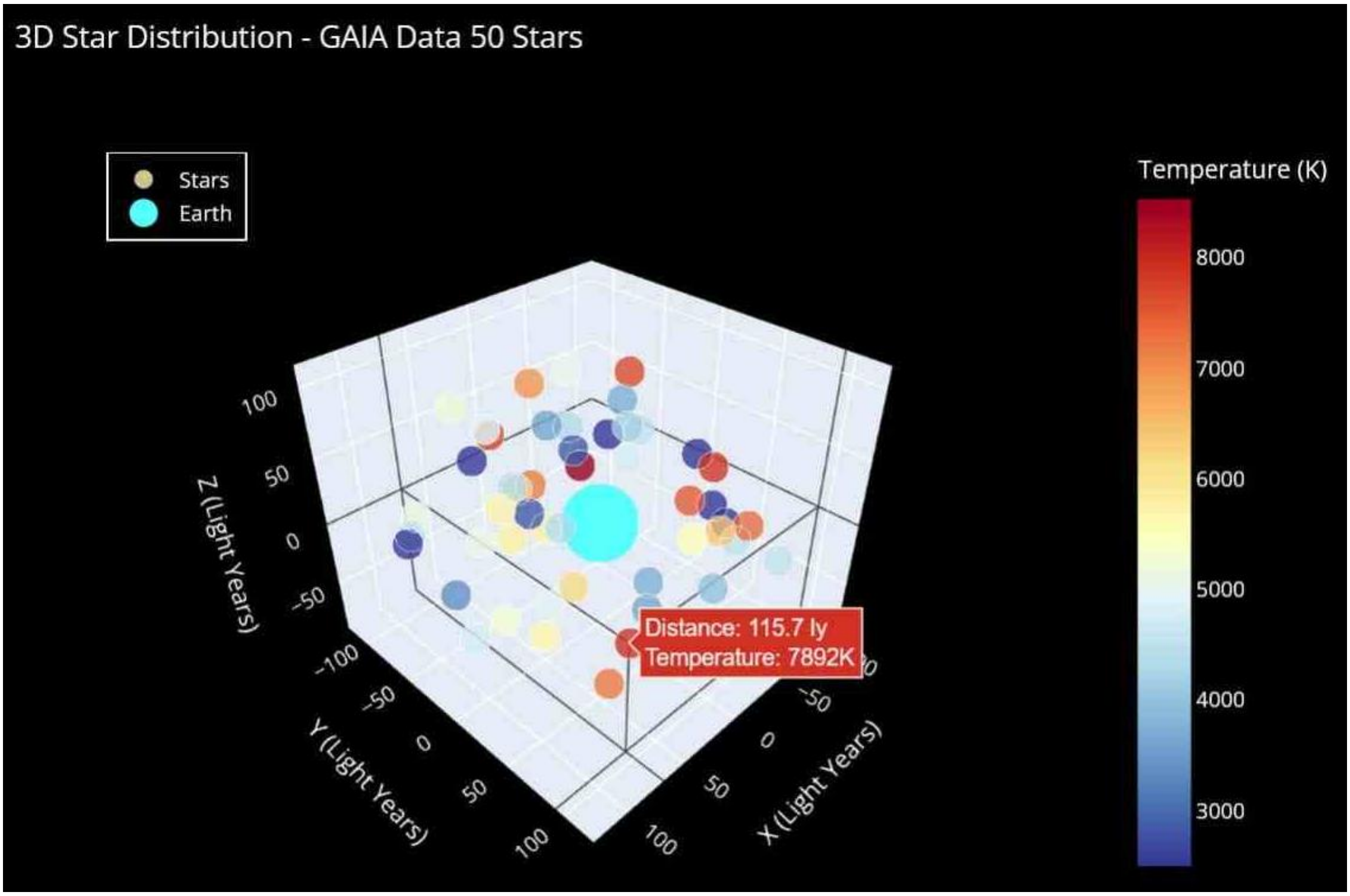
Temperature is color-coded in Kelvin
Sample of 6,000 stars plotted within 130 light years

3D Map of 25 Closest Stars
Using Mayavi Library



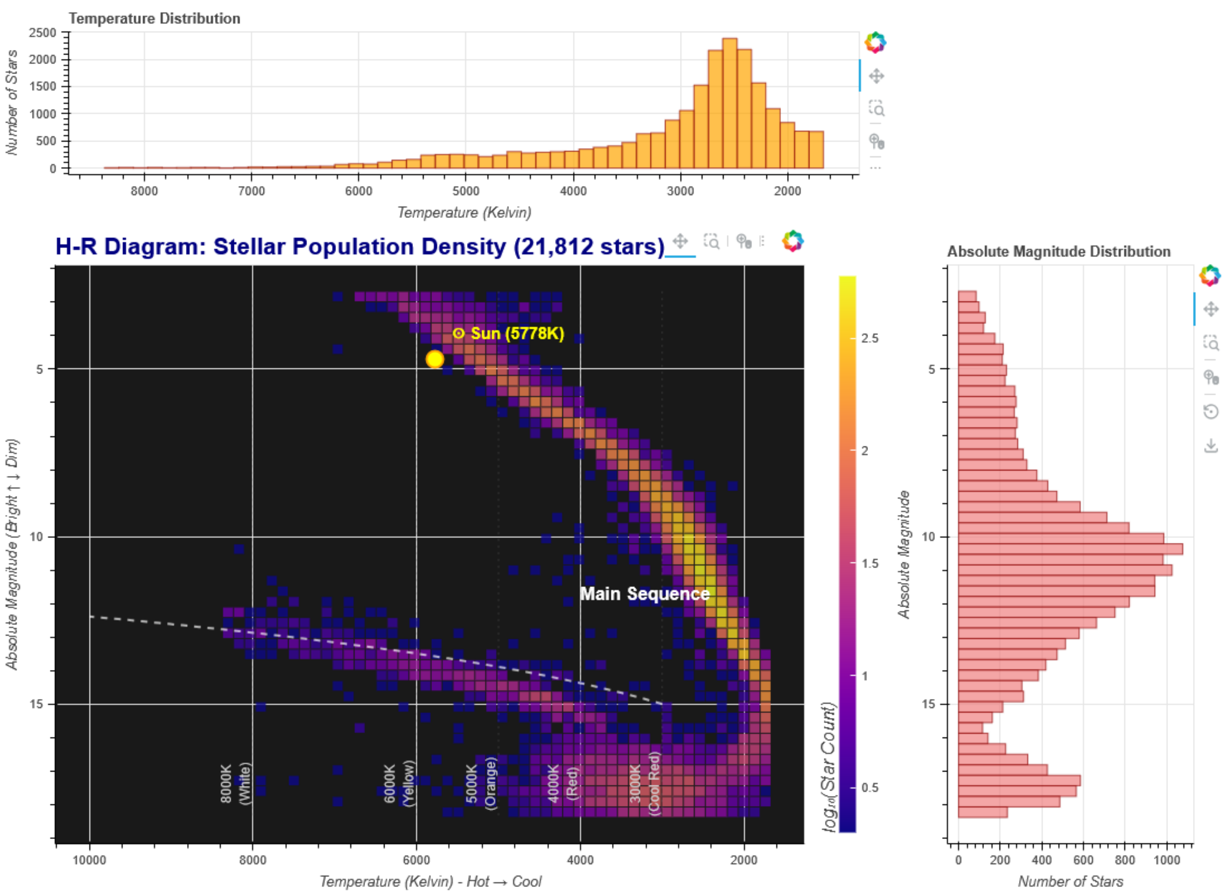
Color Coded by Stellar Temparture

3D Map of 50 Closest Stars
Using Matplotlib Library



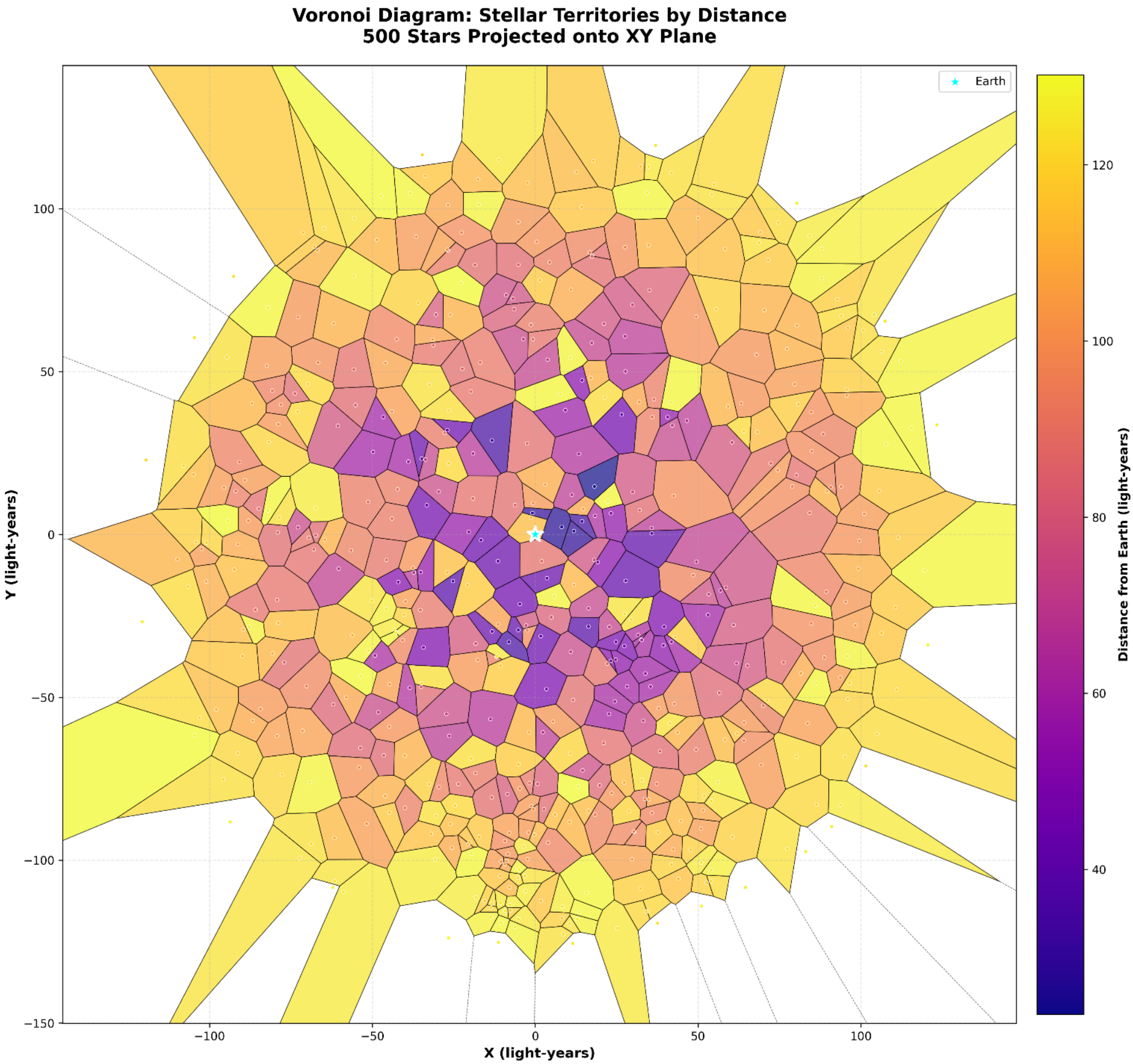
Color Coded by Stellar Temparture
Red is hotter
Blue is cooler

Bihistogram
Temperature vs Absolute Magnitude
24,000 Star Population Density



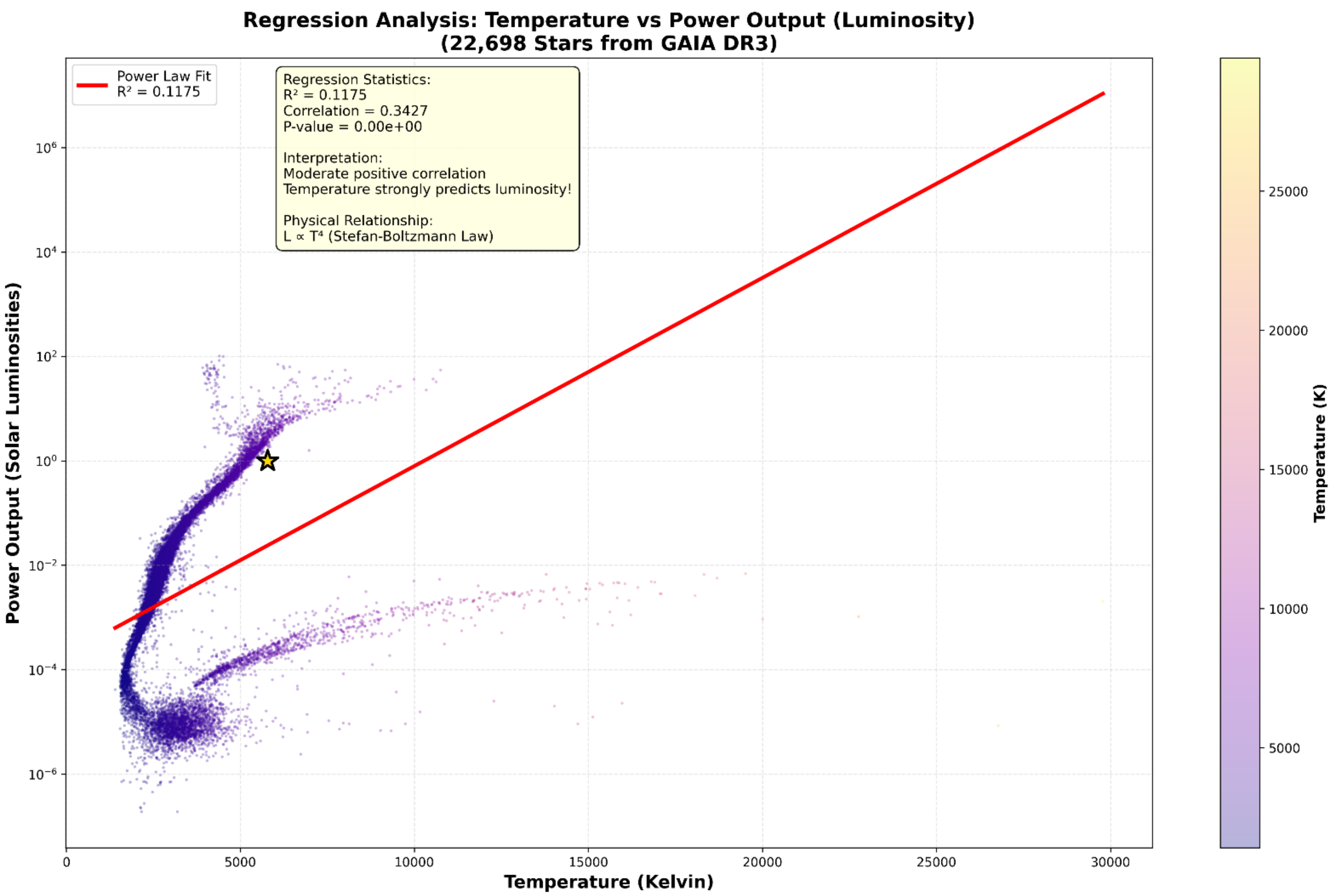
Color Coded by Density
Blue is Higher Density
Yellow is Lower Density

Voronoi Diagram by Distance
500 Stars Projected onto XY Plane



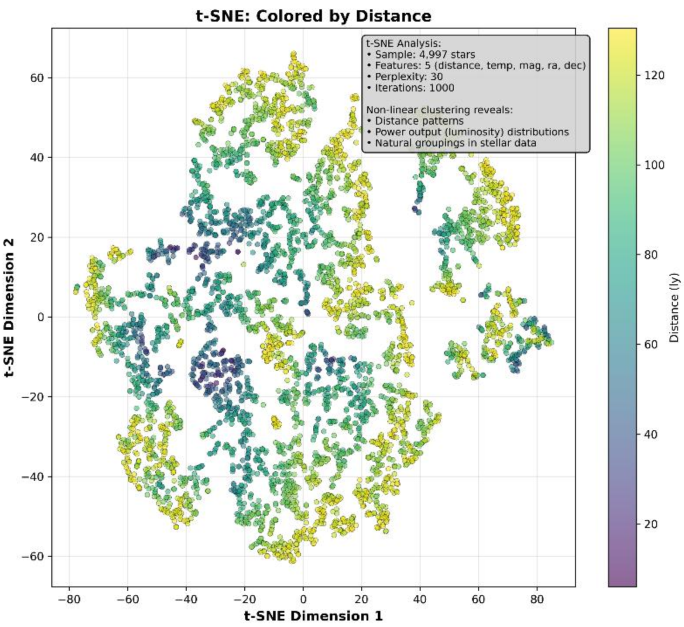
Color Coded by Distance
Blue is Closer to Earth
Yellow is Further from Earth

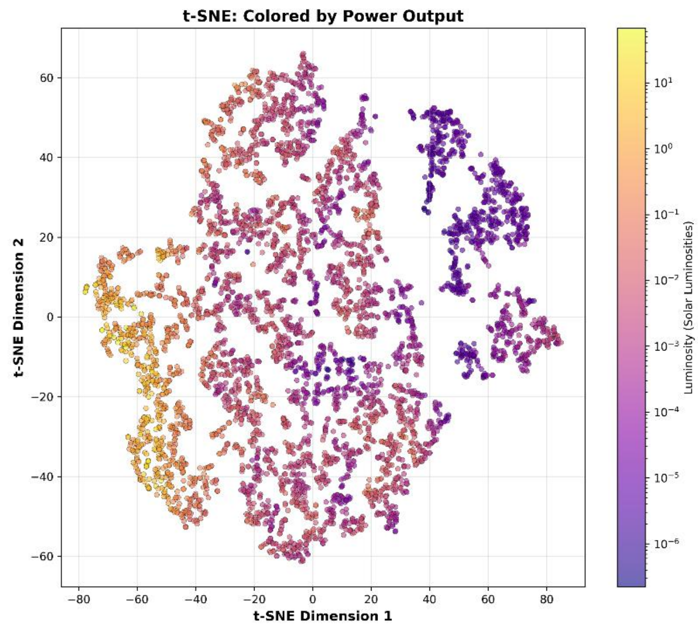
Regression Analysis
Temperature vs Power Output



Color Coded by Temperature
Blue is Cooler

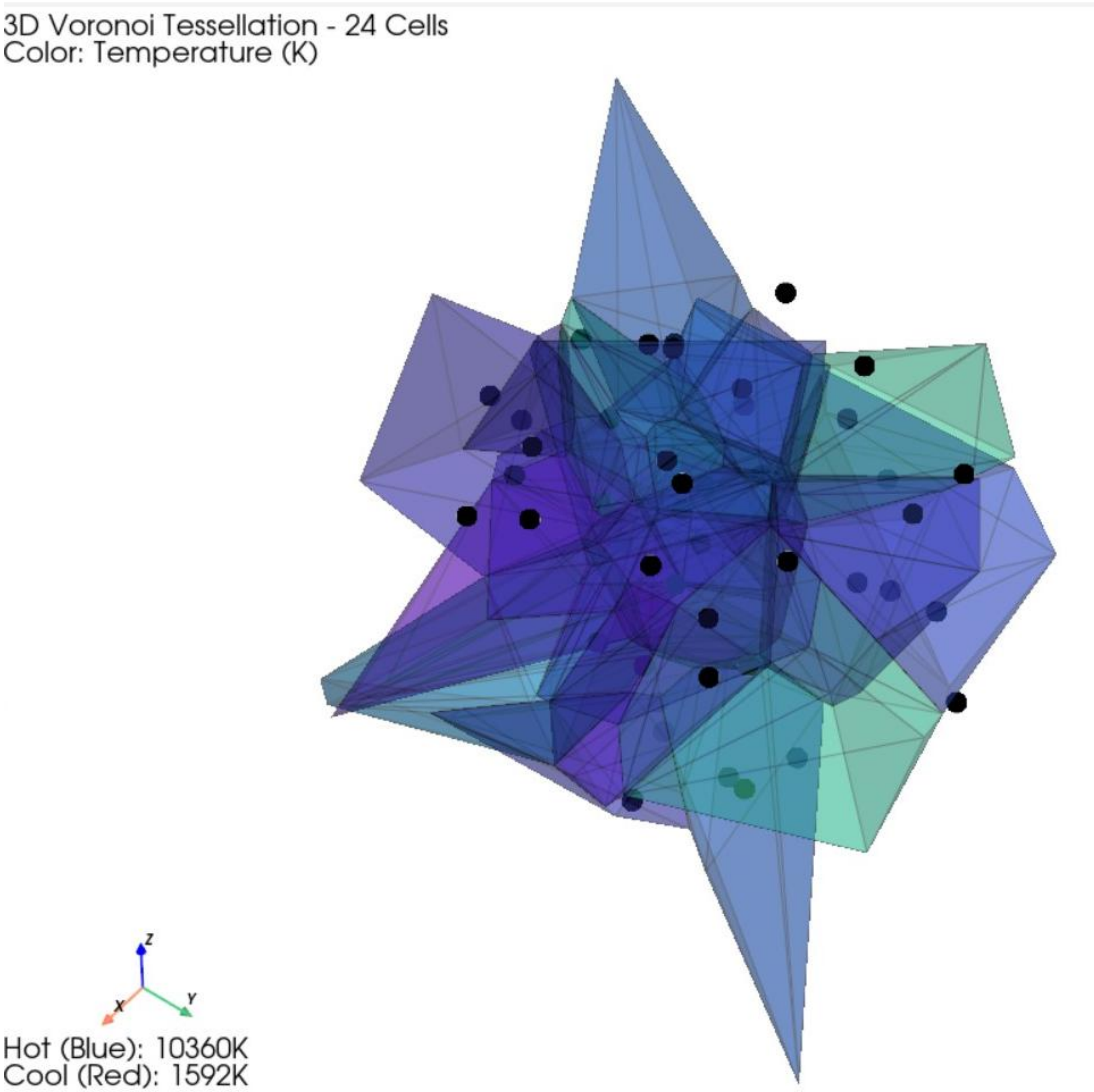
t-SNE
t-distributed Stochasitc Neighbor Embedding
statistical method
Distance and Power Output



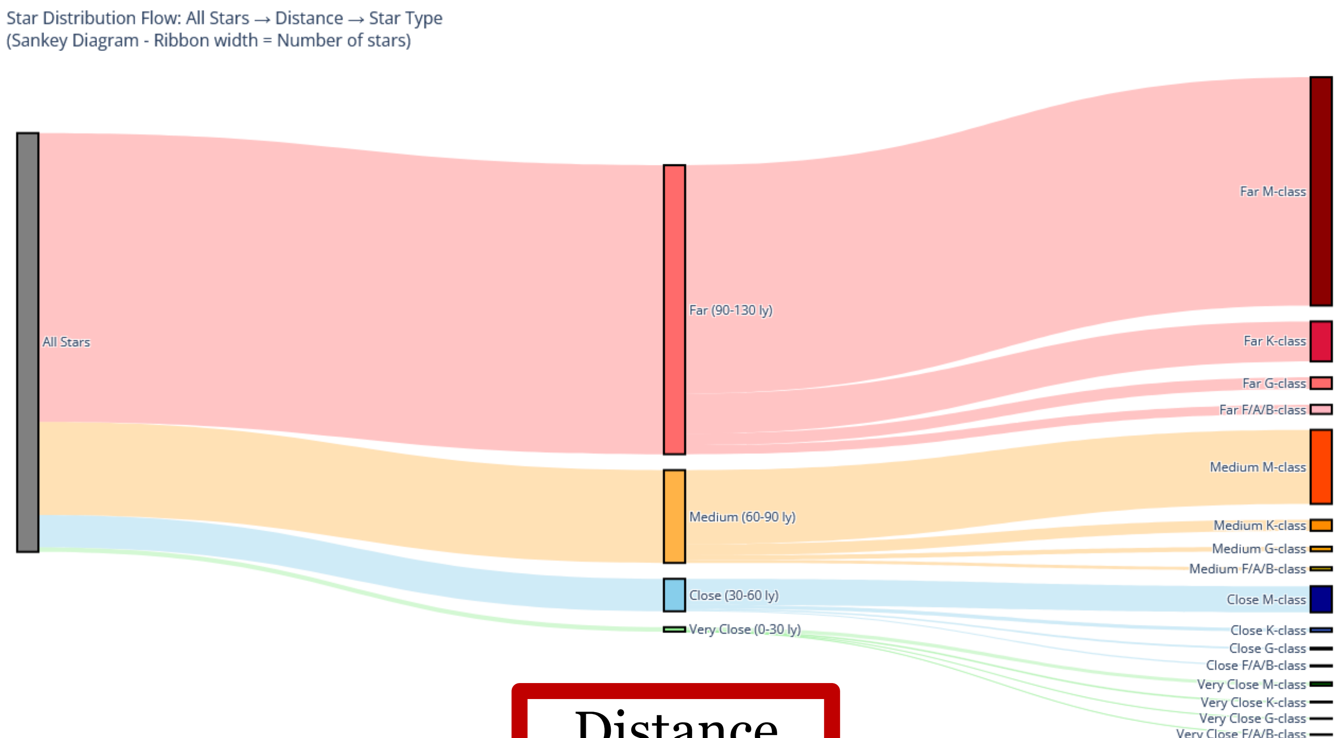


Dimensionality Reduction Technique

3D Voronoi Tessellation
Distance and Power Output



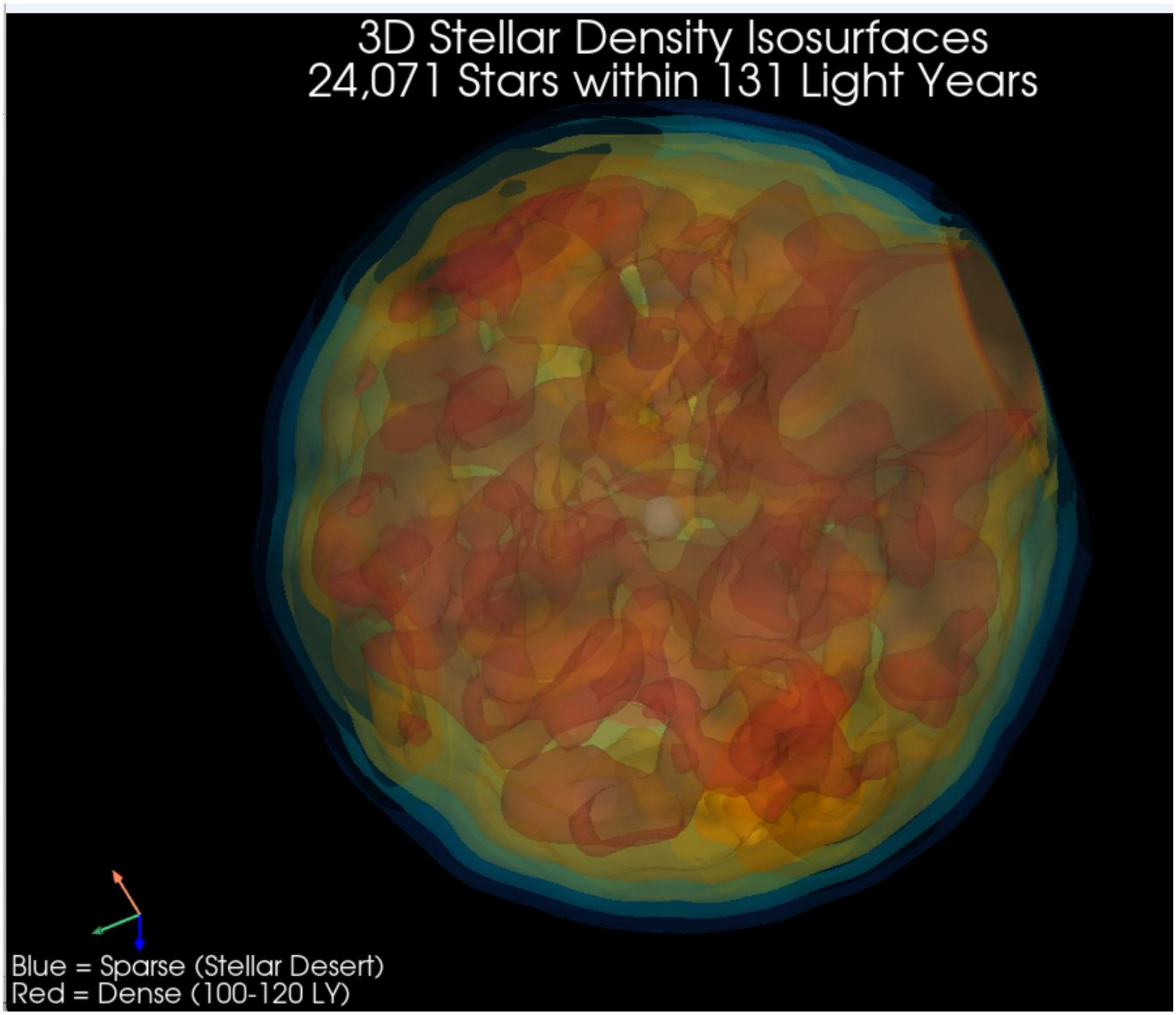
Sankey Diagram
Distance to Star Type



Distance

Star Type

3D Isosurface
Density of 24,000 Stars within 130 Light Years



Color Coded by Density
Blue is Lower Density
Red it Higher Density