Project Report

Parth Ahuja

April 27, 2023

1 DATASET DESCRIPTION

1.1 Description

The data for this project is retrieved from IllustrisTNG Project (`https://www.tng-project.org/`). IllustrisTNG Project is an ongoing series of large-scale cosmological simulations of galaxy formation. The project aims to elucidate the physical mechanisms that drive galaxy formation and evolution and to generate predictions for current and future observational programs. The project comprises three primary runs, TNG50, TNG100, and TNG300, which span a range of volumes and resolutions.

The specific dataset that I will be using for this project is the TNG50 Milky Way+Andromeda (MW/M31) Sample `https://www.tng-project.org/data/milkyway+andromeda/#cat` particularly galaxy 342447 `https://www.tng-project.org/api/TNG50-1/snapshots/99/subhalos/342447/`

This dataset has roughly 5 million points separated by particle type. The particles used in these visualizations are Type0 i.e gaseous particles. The attributes retrieved from this these particles are Coordinates, Density, Internal energy, Velocity (Vector), and Magnetic field (Vector).

Note: You have to sign up for an API key to view some of these pages in detail. The data is available as an `hdf5` extension.

1.2 Download Instructions

Since the data files are large, they are available for download at this link: `https://drive.google.com/file/d/1uSjKHxVC-VkxUxWCKlUWu4pYgbrWovbk/view?usp=sharing`
Note: This includes gas-output.vtk files (large and mini), ctf files, and original 342447.hdf5 file.

2 Problem Objective

The objective of the visualization project is to gain insights into the physical processes that govern the formation and evolution of galaxies and their associated satellite systems. Specifically, the project aims to answer the following questions:

1. How does the density of gas vary across different regions of the central galaxy?
2. Are there any correlations between the velocity vectors of gas in different regions of the central galaxy?
3. What is the relationship between the internal energy of gas and its density within the central galaxy?
4. Are there any correlations between density, velocity, internal energy or magnetic field?

The goal of the project is to use VTK techniques to visualize and analyze the properties of matter, including density, velocity as vector, and internal energy, within a single galaxy from the TNG50 Milky Way+Andromeda sample. Through this visualization, I aim to gain insights into the spatial distribution and correlations between these properties, as well as any potential trends or patterns that may emerge.

3 Methodology

3.1 Boundary Box generation

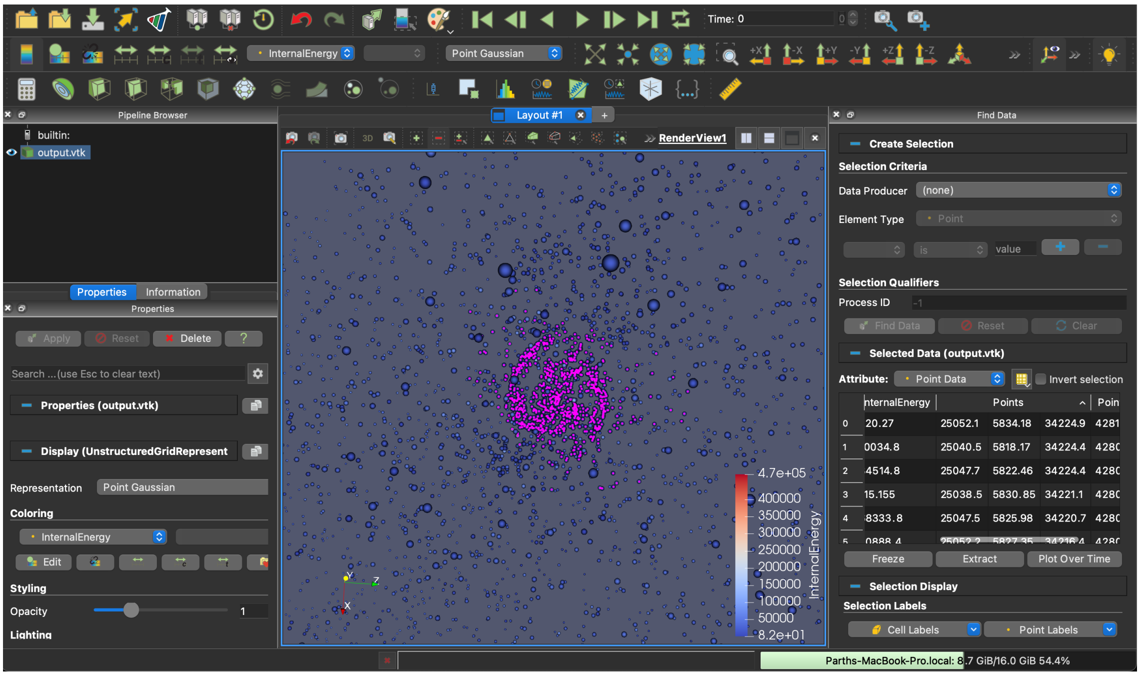

Figure 3.1: Boundary box points

A boundary box that fits the requirements that are discussed in section 3.2 is constructed in Paraview. With the help of Python shell within Paraview, the coordinates for this boundary box are then exported as values listed in table 3.1. Figure 3.1 shows the points selected within the boundary as highlighted in pink.

3.2 Data Processing and conversion

The goal with data sampling is to choose enough points outside the galactic core so it is easy to view particle ejection, especially at the center of the galaxy at the black hole. At the sample time, it's important to not choose too many points that could obstruct the view of the galactic core. Therefore, we can define randomized sampling as the equation:

$$S_n = \{(x_i, y_i, z_i)\}_{i=1}^{n} \tag{3.1}$$

Where the x_i, y_i, and z_i are random variables generated from a uniform distribution defined over the range of x, y, and z within the boundary box B. The sampling script included in the code `hdf2vtk.py` defines the mask mentioned in equation (3.1) as the following values:

Cordinate	**Min**	**Max**
X	25015.3	25071.4
Y	5807.92	5843.36
Z	34159.5	34224.9

Table 3.1: Boundary Box values

The mask is applied using numpy `np.where` and randomized particles are attained from the mask using the `np.random.choice` method.

3.3 VTK Rendering

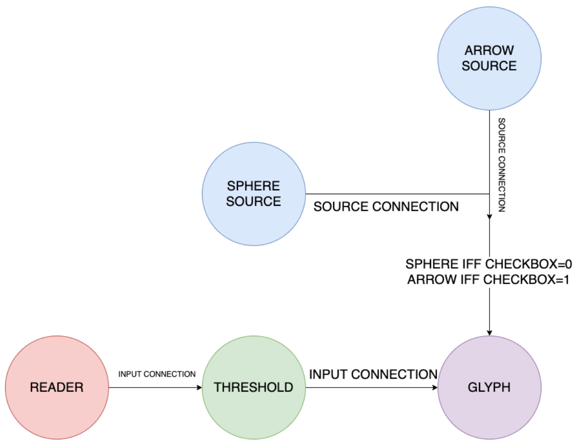

Figure 3.2: VTK Render pipeline

The state diagram in figure 3.2 shows the VTK render pipelines and its stages. The reader is passed through a threshold filter which makes it easy to update the lower floor of the plot in the GUI threshold slider. The output from the threshold is then passed into the Glyph filter. The source for the glyph is switched based on the value of the arrow source checkbox in the GUI. Something to note here is that this checkbox is only available for the vector attributes i.e Velocity and Magnetic Field. The GUI elements of PyQt5 include a dropdown selector for attribute values. an update data statistics button, and two sliders for threshold and scale.

3.4 Statistics

The methodology for the data statistics calculation includes the usage of the following formulas in numpy API

1. The minimum and maximum values for scalars are trivial. For vectors we can define them using magnitude

$$\min_{i=1}^{n} \|\mathbf{v}\|_i, \tag{3.2}$$

$$\max_{i=1}^{n} \|\mathbf{v}\|_i, \tag{3.3}$$

   where $|\mathbf{v}|_i$ is the i-th norm of the vector $\mathbf{v}$. Therefore, a similar notation can be applied to the mean and coefficient of variance which are just listed in scalar formulas.

2. Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.4}$$

   where $\bar{x}$ represents the sample mean, n is the sample size, and x_i represents the i-th observation in the dataset.

3. Standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.5}$$

   where s represents the sample standard deviation, n is the sample size, x_i represents the i-th observation in the dataset, and $\bar{x}$ is the sample mean.

4. Coefficient of variation

$$\mathrm{CV} = \frac{s}{\bar{x}} \times 100\% \tag{3.6}$$

   where CV represents the coefficient of variation, s is the sample standard deviation, and $\bar{x}$ is the sample mean.

The data statistics allow the user to compare different datasets. I found these metrics to be very useful in analyzing certain aspects of the data such as internal energy variation and velocity dispersion.

4 Results

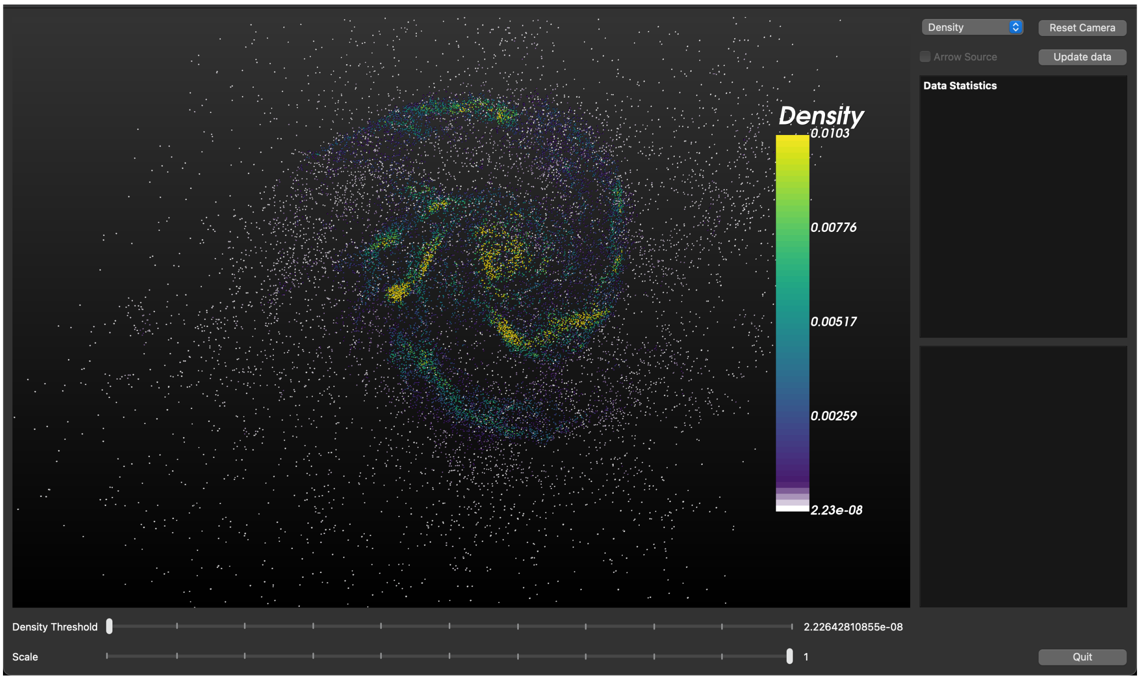

Figure 4.1: Density Plot

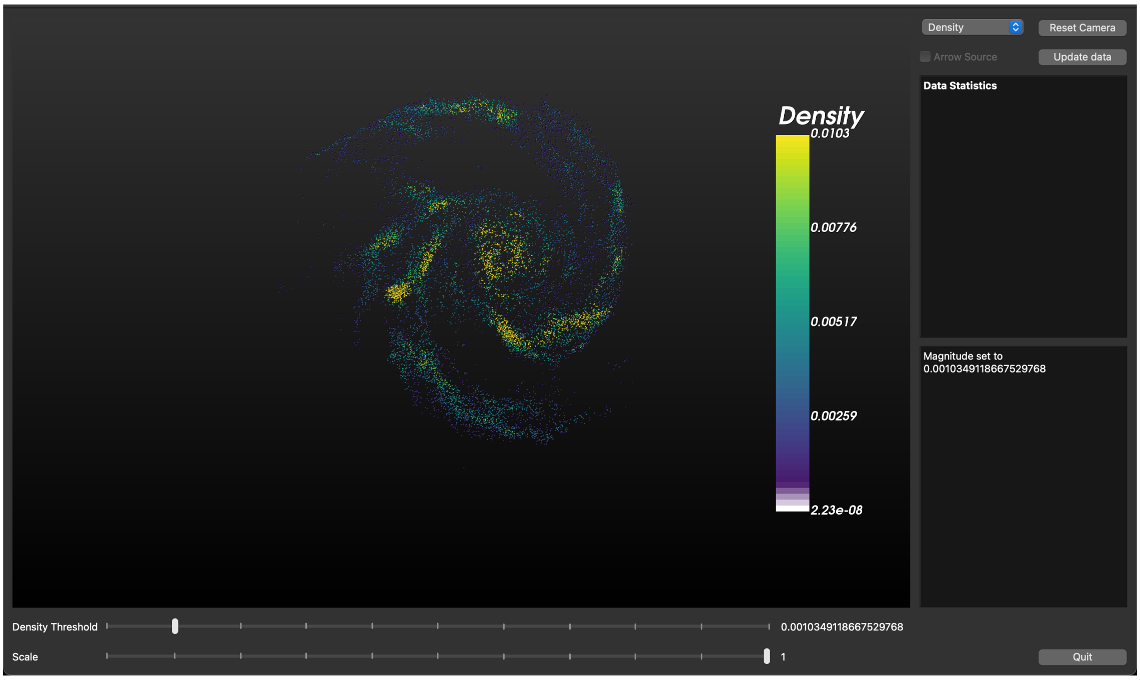

Figure 4.2: Density Threshold at 0.001

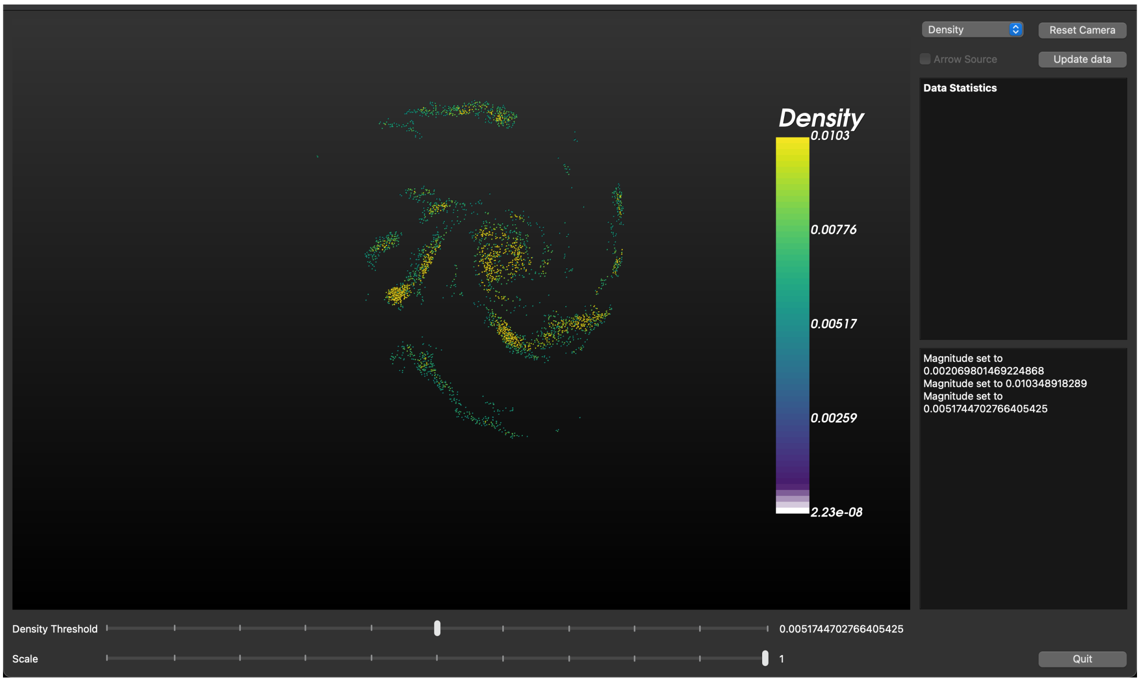

Figure 4.3: Density Threshold at 0.005

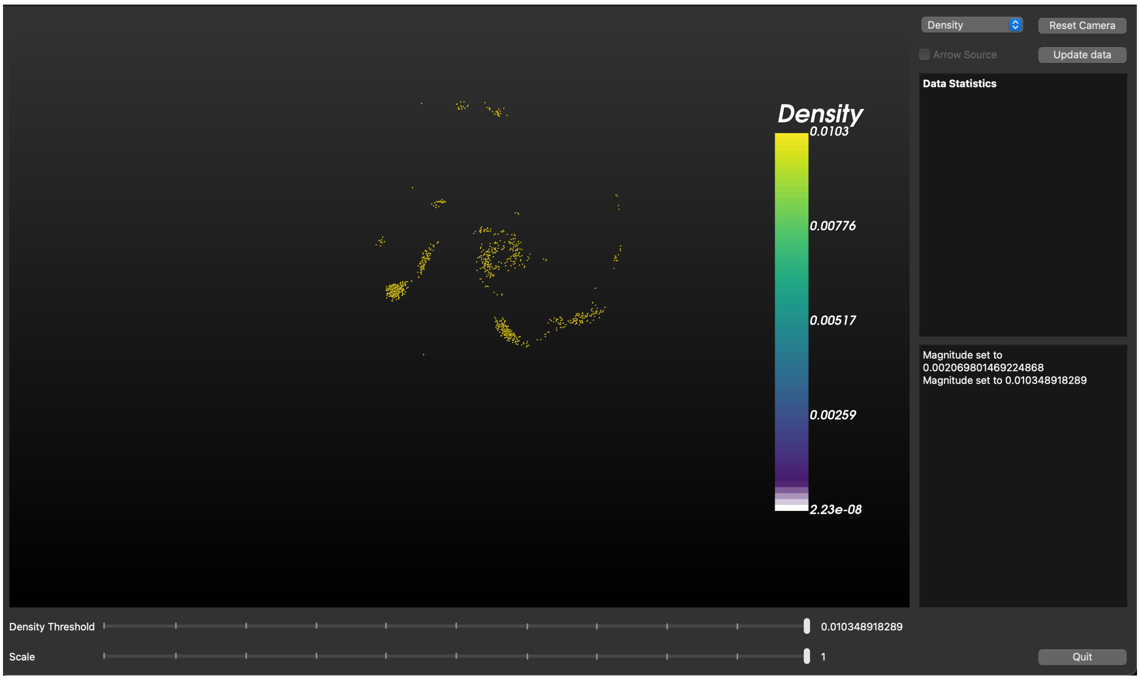

Figure 4.4: Density Threshold at 0.01

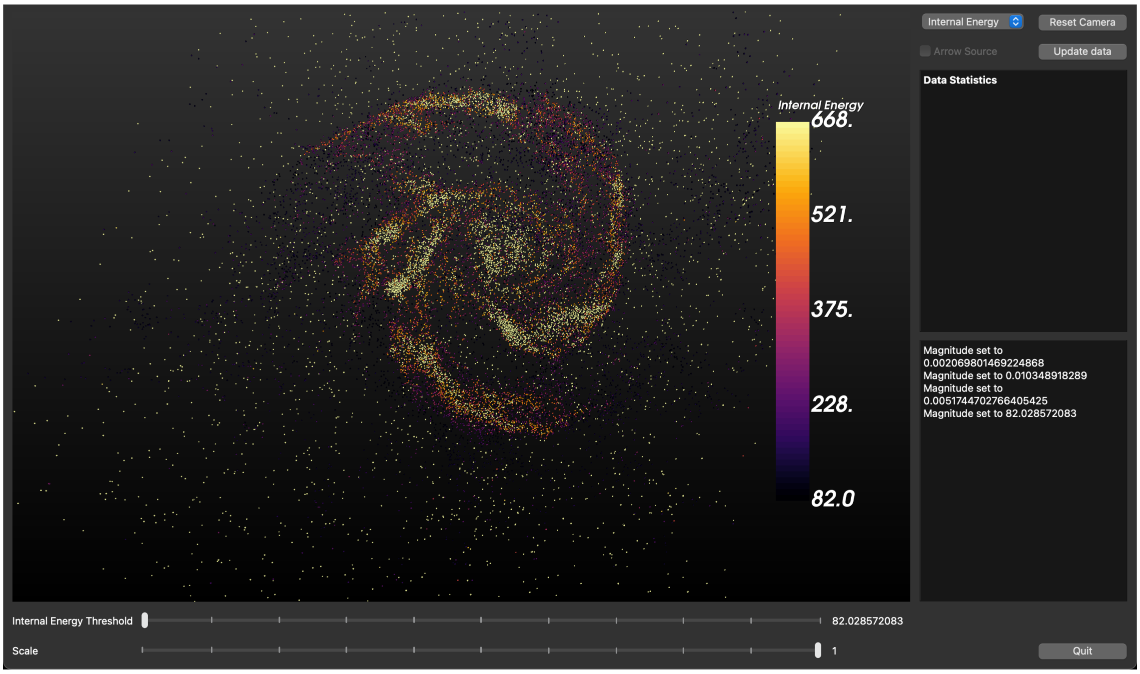

Figure 4.5: Internal Energy Plot

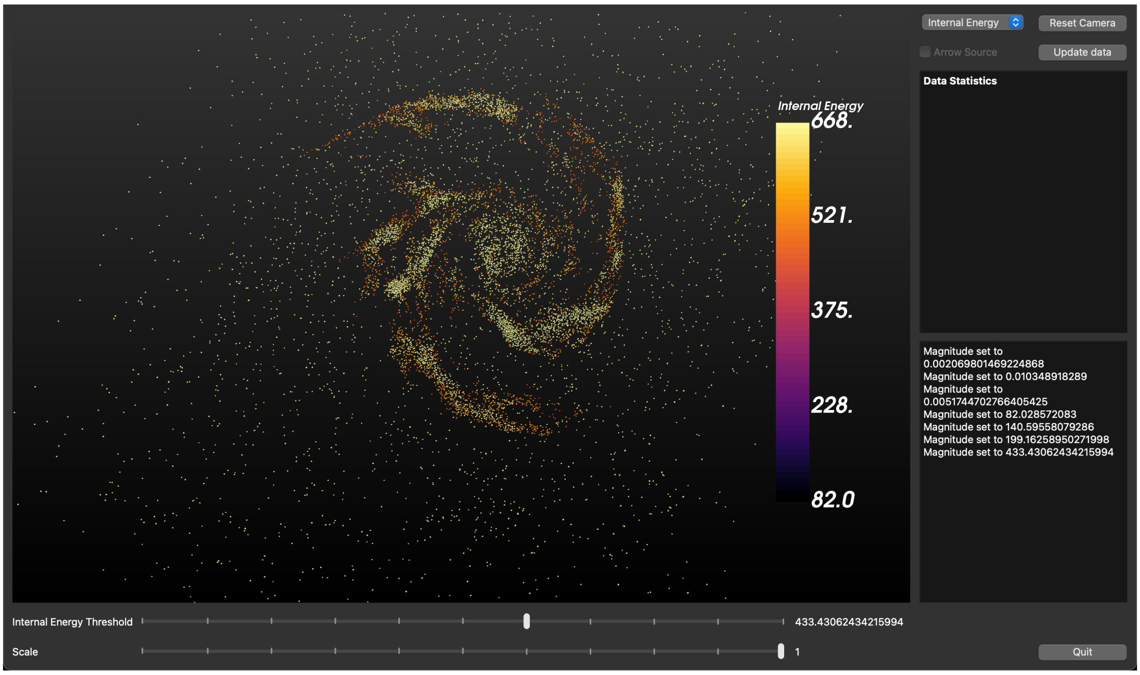

Figure 4.6: Internal Energy Threshold at 433

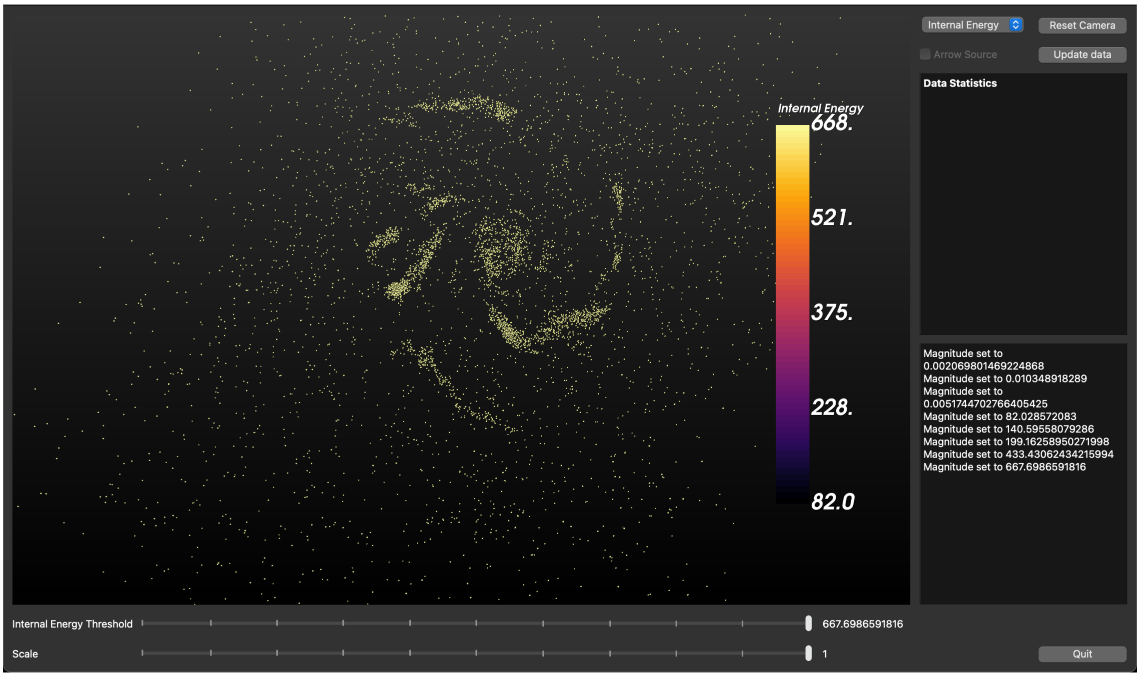

Figure 4.7: Internal Energy Threshold at 667

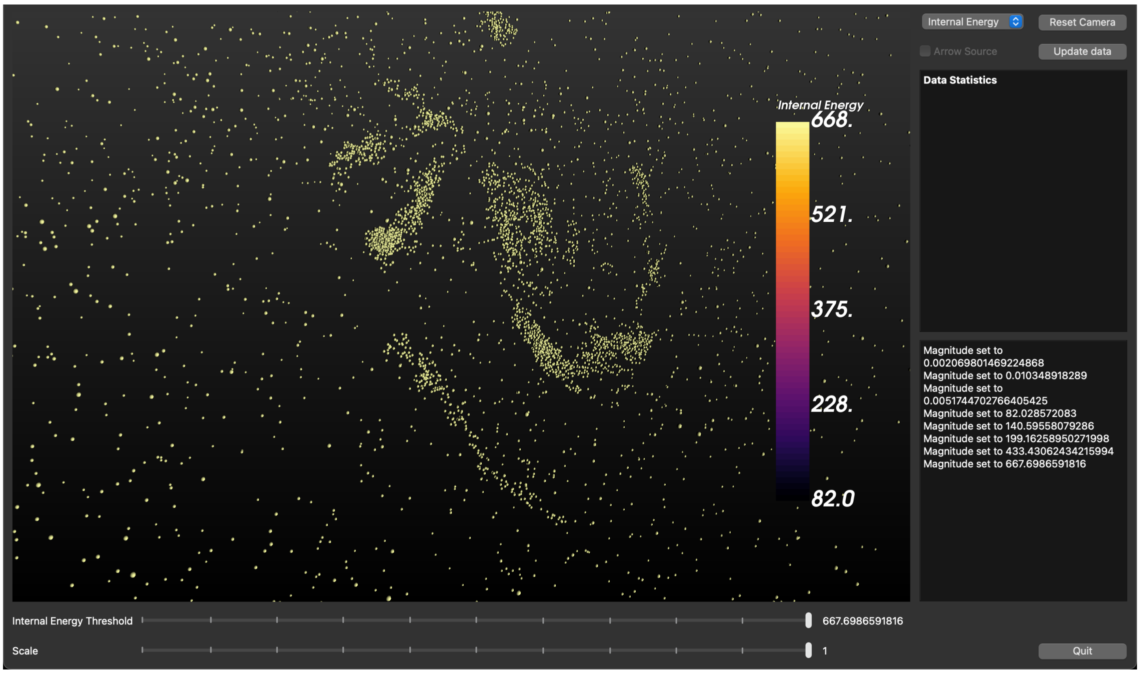

Figure 4.8: Internal Energy Threshold at 667 (Zoomed)

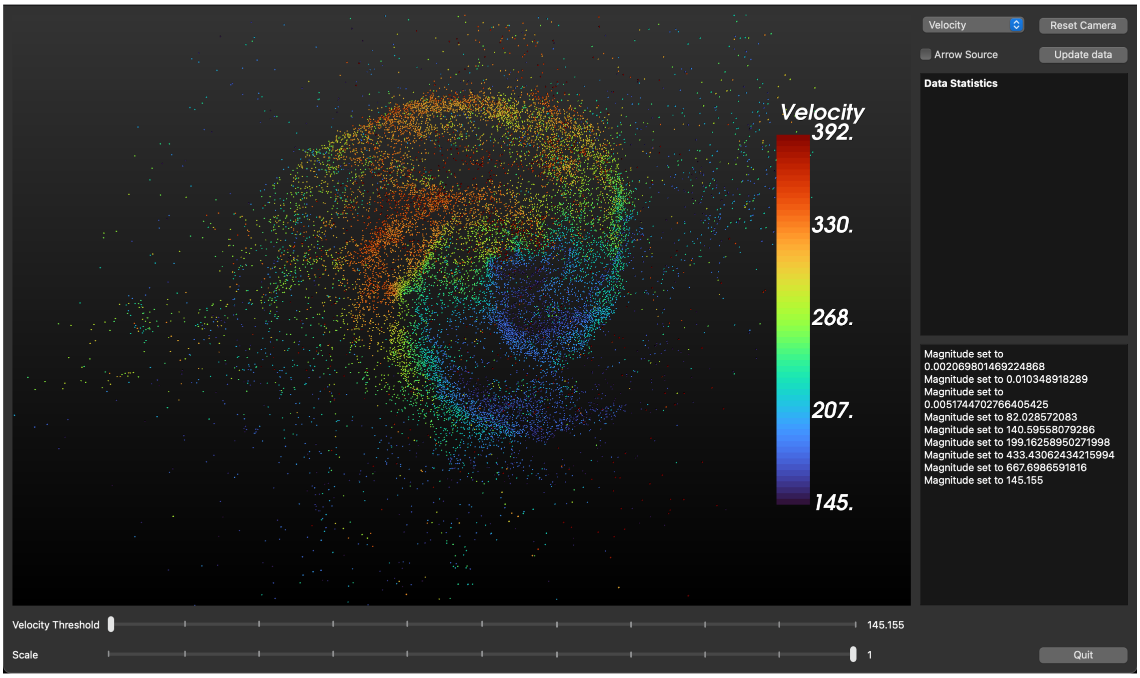

Figure 4.9: Velocity Plot

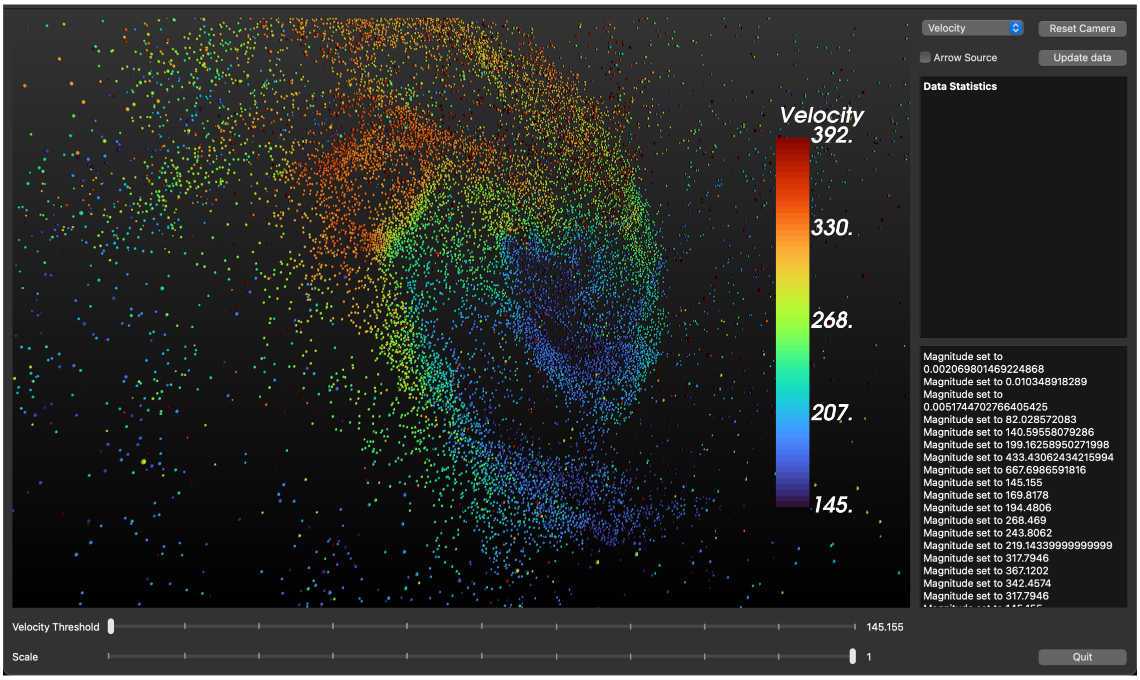

Figure 4.10: Velocity Plot (Zoomed)

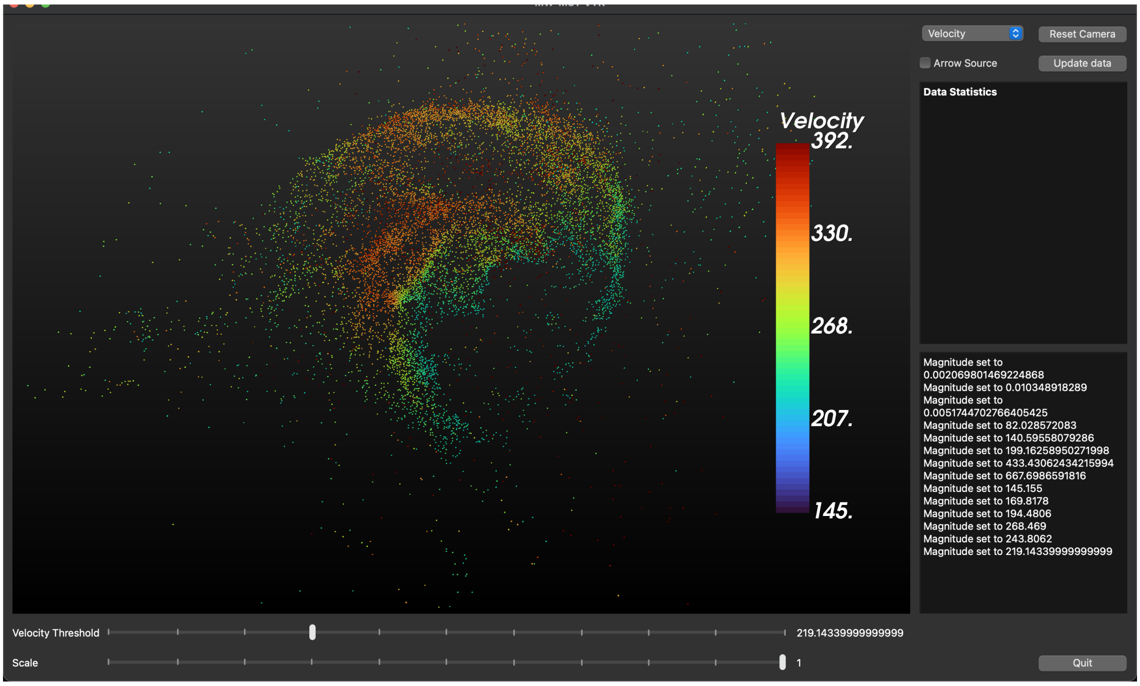

Figure 4.11: Velocity Threshold 219

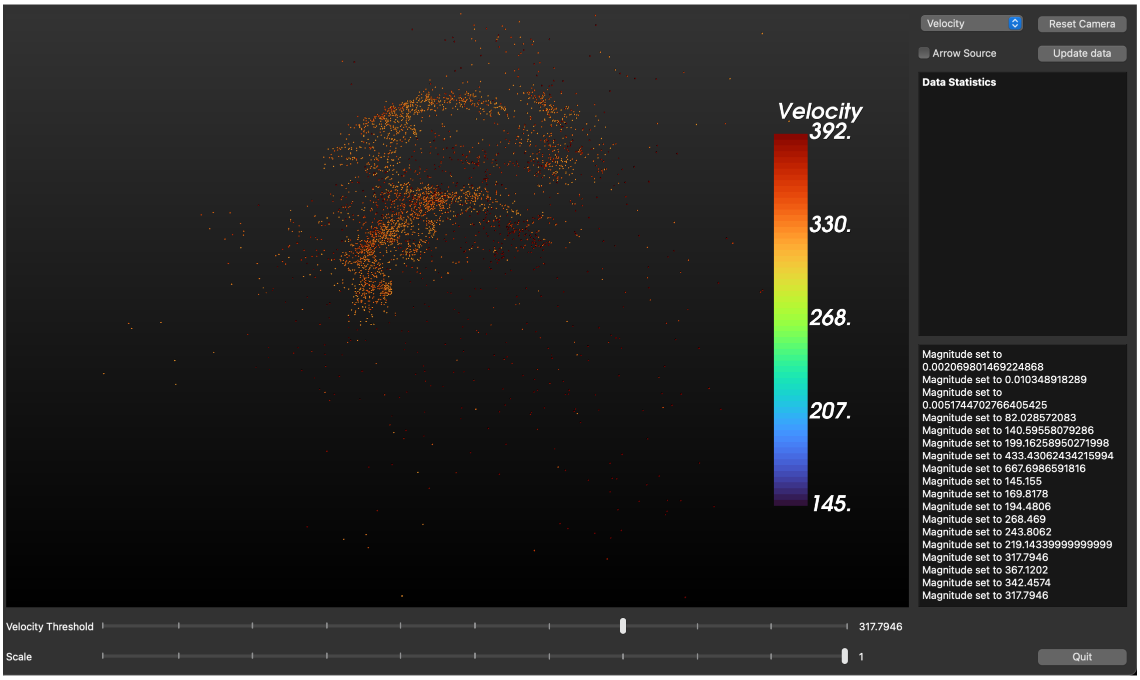

Figure 4.12: Velocity Threshold 317

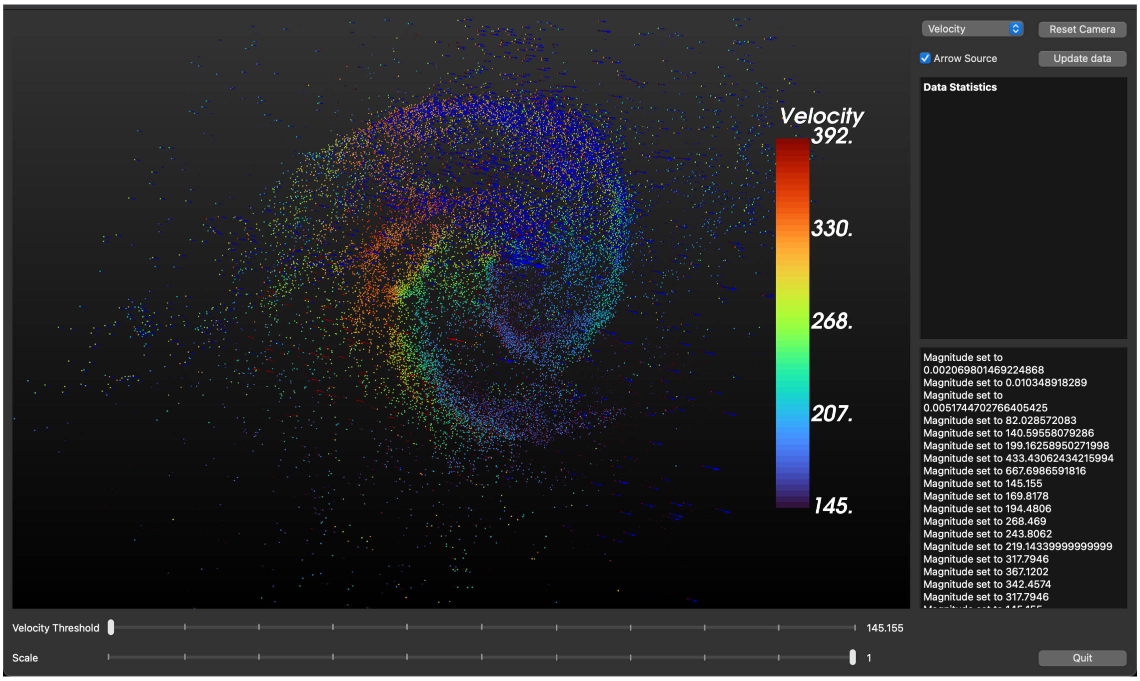

Figure 4.13: Velocity Arrows Enabled

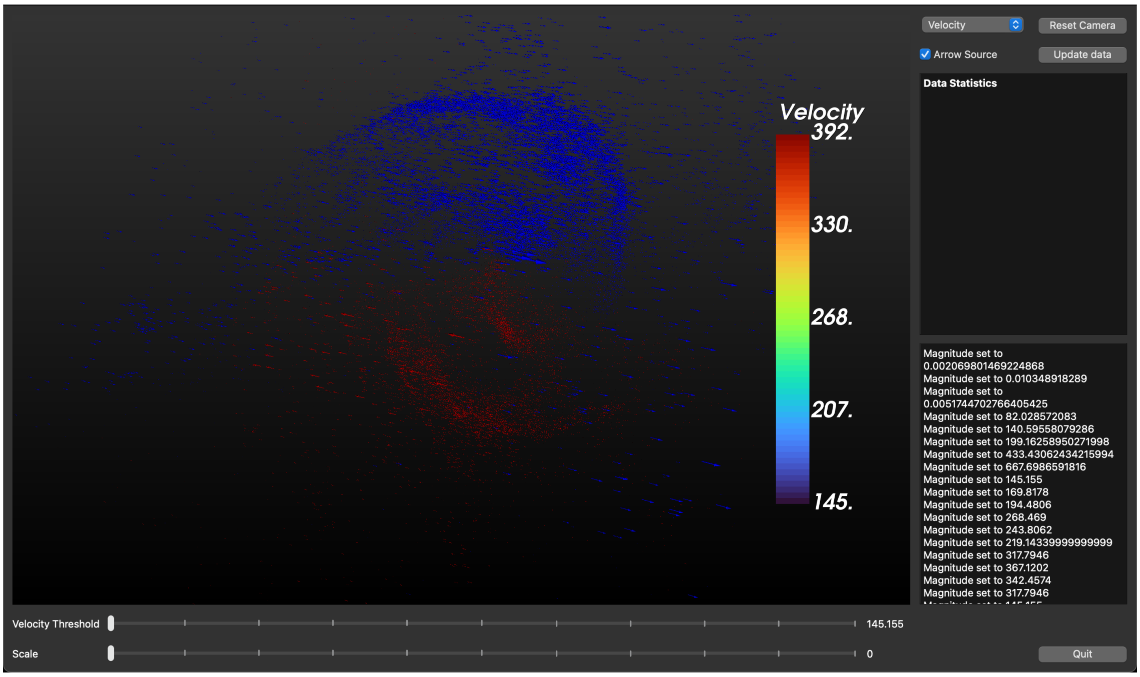

Figure 4.14: Velocity Arrows Scale 0

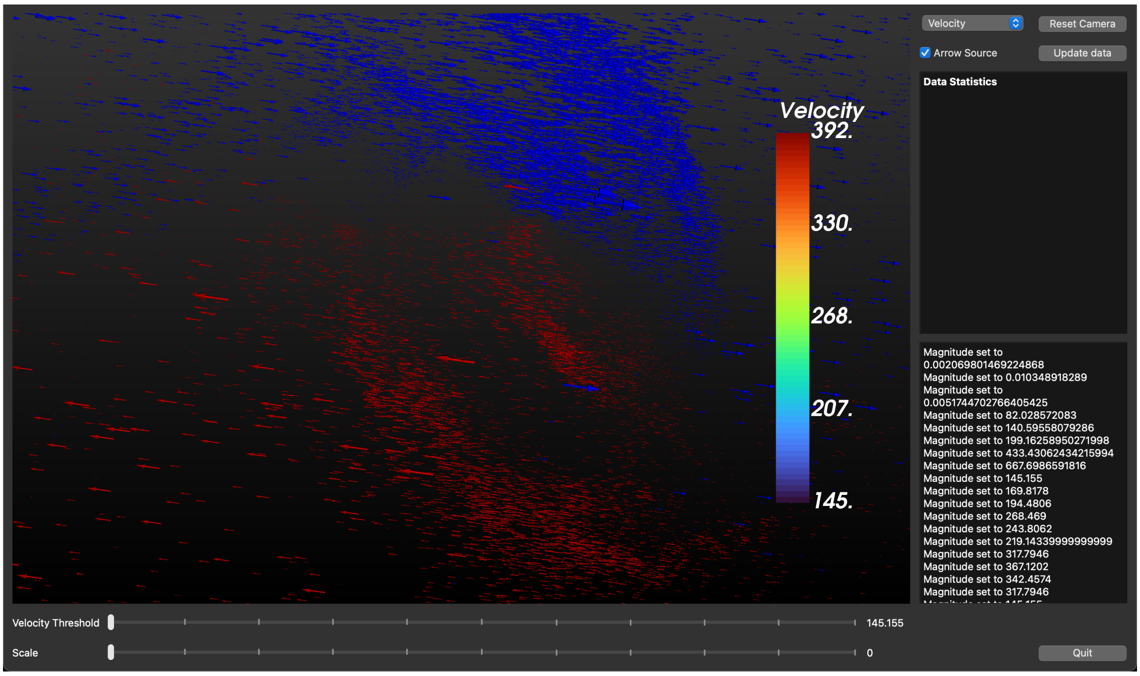

Figure 4.15: Velocity Arrows Scale 0 (Zoomed)

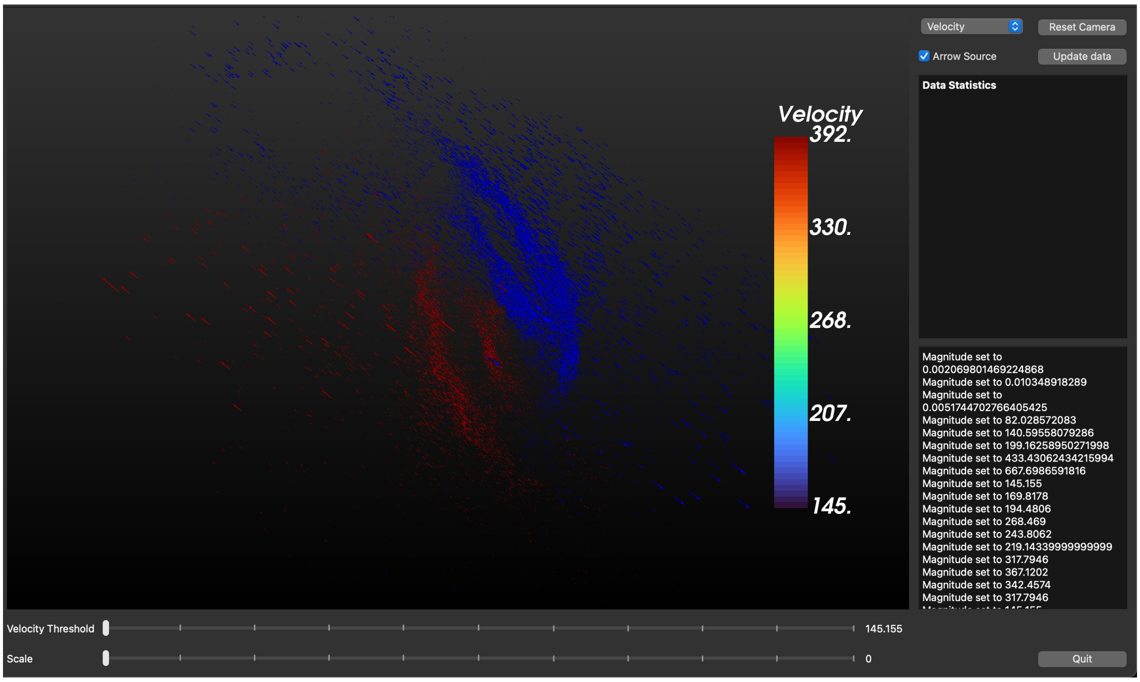

Figure 4.16: Velocity Arrows Scale 0 (Zoomed and Angled)

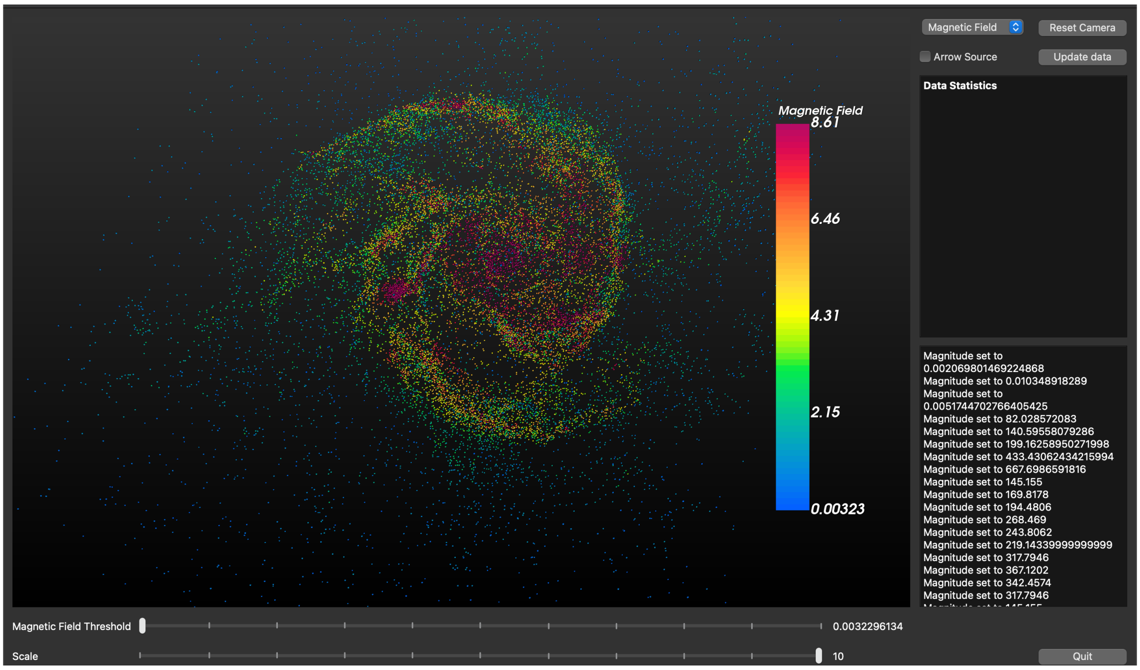

Figure 4.17: Magnetic Field Plot

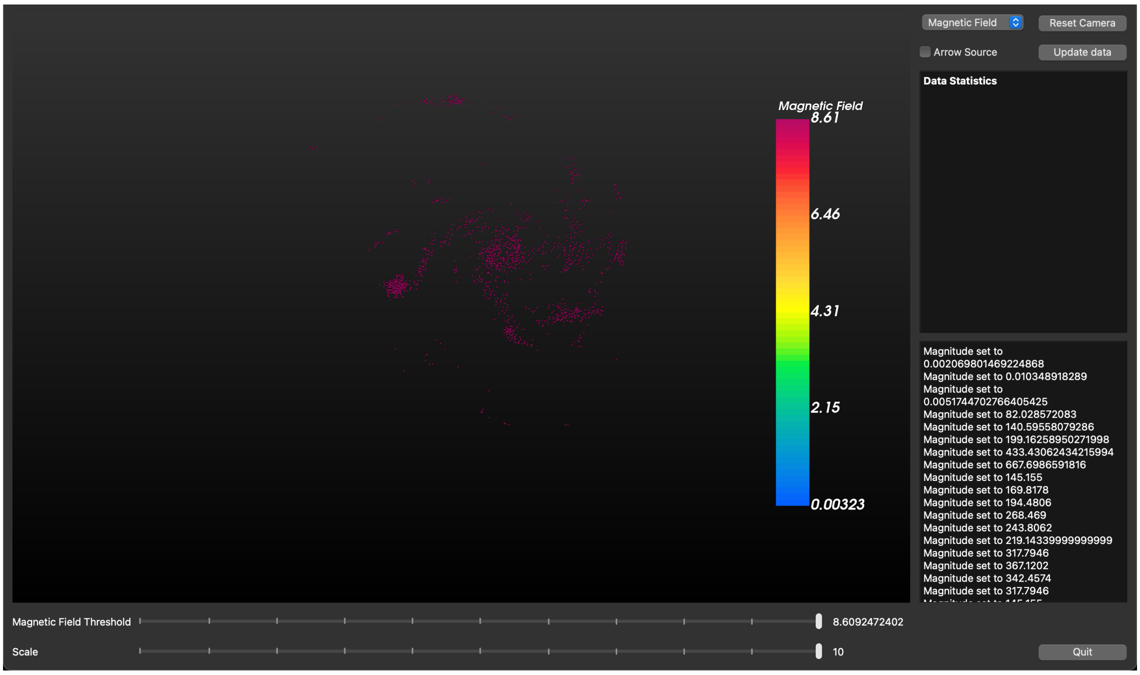

Figure 4.18: Magnetic Field Threshold at 8.6

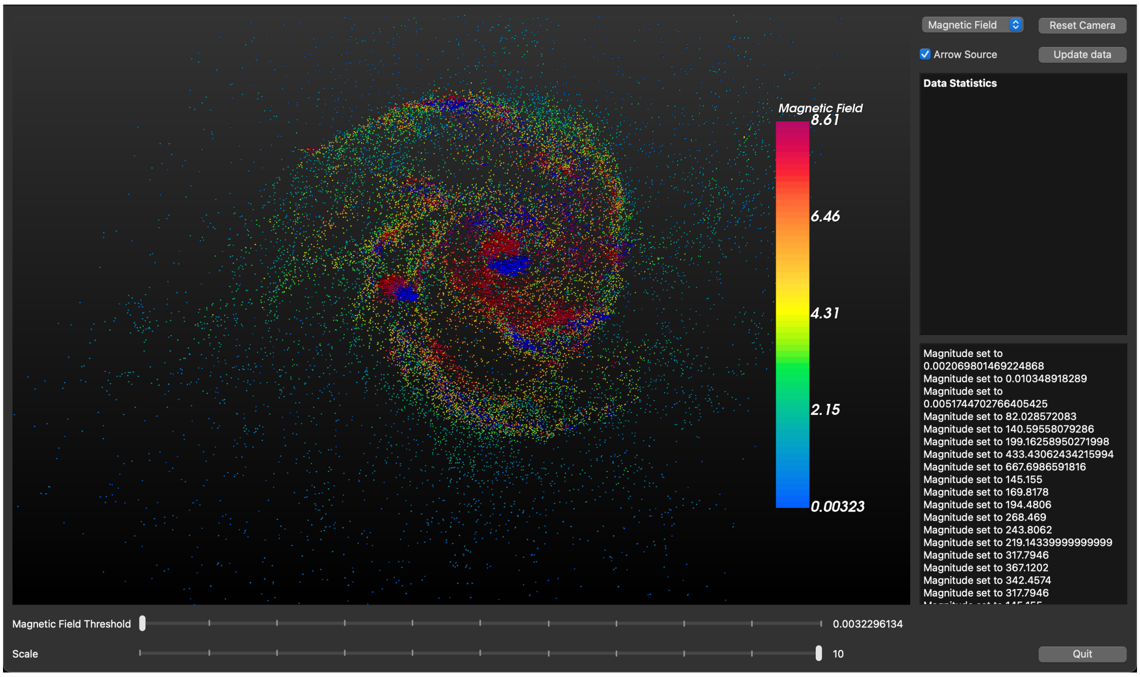

Figure 4.19: Magnetic Field with Arrows

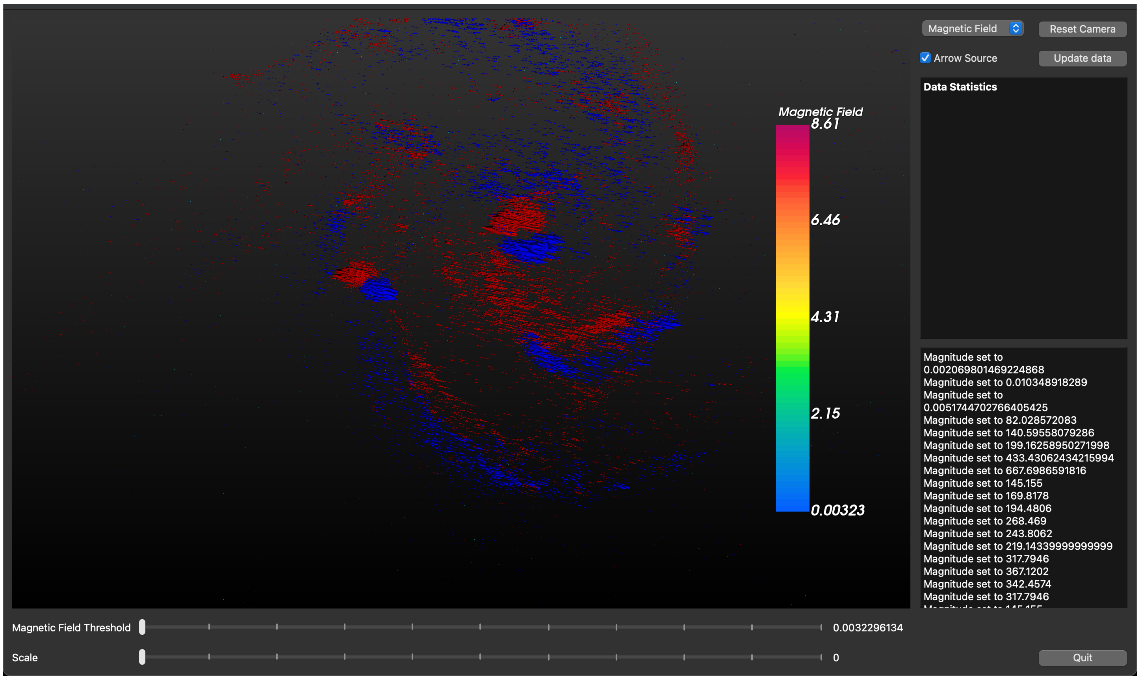

Figure 4.20: Magnetic Field with Arrows and Scale 0

The images provided in the results section include visualizations of the scalar values of density and internal energy. The magnitude of the velocity and magnetic field is also visualized using the CTF and can be converted to an arrow glyph with the checkbox. All the visualizations use different color transfer functions. Namely: Viridis (Density), Inferno (Internal Energy), Rainbow (Velocity), Turbo (Magnetic Field)

5 DISCUSSION

I do believe that I reached all the objectives that I mentioned in the project report. However, I did end up switching the streamtracers to vector arrow glyphs. This was primarily due to the fact that when converting the data from `hdf5` format, the cell connectivity was not defined correctly which caused `SEGFAULT` when I tried to create any meaningful tracers. This may also be due to having many points in the galaxy. The only way I could get this to work was by reducing the points to 1000 which produced facile results. Therefore, I decided to use the arrow glyphs which were satisfactory for the results in these visualizations.

There are a few observations that I made from the data, which I would like to emphasize:

1. Overall all attributes except velocity seem to correlate to each other
2. Internal energy follows a pattern that one would expect within the galaxy; However, As it is clear in image 4.7 the intergalactic space shows very high internal energy. This high energy is a result of various processes, such as shocks from supernovae and gravitational heating caused by the formation of large structures in the universe. Furthermore, In intergalactic regions, gas has high energy due to the process of cosmic reionization. During the early stages of the universe, before stars and galaxies formed, the universe was filled with a highly ionized plasma that emitted and absorbed radiation very efficiently. As the universe expanded and cooled, this plasma became neutral atoms and molecules, a process called recombination.
3. Velocity seems to have the lowest coefficient-variance in the entire dataset. This generally indicates that, except the center of the galaxy, everything is moving at a low dispersion rate. The motion of particles within a galaxy is primarily determined by the gravitational field of the galaxy itself. The mass distribution of the galaxy, including the dark matter, exerts a gravitational force on all objects within it, including stars. This force causes the particles to move in orbits around the center of mass of the galaxy.
   Since all particles within a galaxy are subject to the same gravitational field, they move at roughly the same speed, assuming they are located at similar distances from the center of mass of the galaxy.
4. Finally, I wanted to point out the region next to the center of the galaxy that shows high magnitutude in all atttributes. This is especially clear in figure 4.20. Presumably, this could be another black-hole region within this galaxy.

There are few things that I wish I would've done differently. Firstly, I could've optimized my sampling procedure to perform sampling and perform masking on attributes altogether. This can help decrease the runtime of the `hdf2vtk` script which was 15 mins for mini and 60 mins for large file .vtk file. Second, The UI can be optimized by only re-rendering the points rather than the entire glyph. Due to being short on time I wasn't able to include this optimization.

I think there is a lot more exploration that can be done in this project. I only extracted 4 out of 15 fields available in the dataset. I would also like to inspect that region next to the center of the galaxy with other attributes like X-Ray fields to see if the values are in-line with other black holes.

6 ANNEX

6.1 Statistics for density

```
Min:  4.6358272243e-07
Max:  0.20607112348
Mean:  0.0036686226477292564
Std:  0.01067585982447336
Co-Var:  291.0045771832469
```

6.2 Statistics for internal energy

```
Min:  82.384384155
Max:  721147.625
Mean:  6580.431675342587
Std:  23999.89201892464
Co-Var:  364.7160733976494
```

6.3 Statistics for velocity

```
Min:  13.708792027580605
Max:  2841.6383751106437
Mean:  246.8877384745161
Std:  86.41897765543203
Co-Var:  35.003349372229856
```

6.4 Statistics for magnetic field

```
Min:  0.0141383669209695
Max:  39.17187268198829
Mean:  4.266503642555025
Std:  3.568752669275075
Co-Var:  83.64583669118592
```