Maarten Breddels (@maartenbreddels) - Kapteyn Institute (Groningen, NL)

vaex

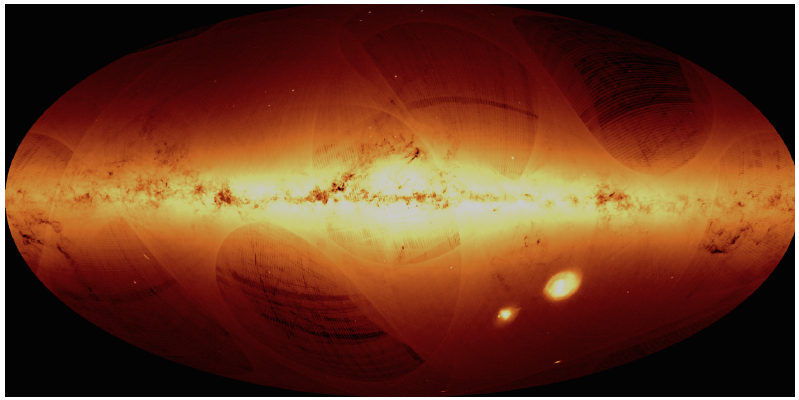
Challenge
With large astronomical catalogues (>1 billion) such as the Gaia catalogue, and Pan-STARRS we need new methods to visualize and explore these large datasets. Scatter plots lead to overplotting, making these often useless and too slow (>> 1 minute).

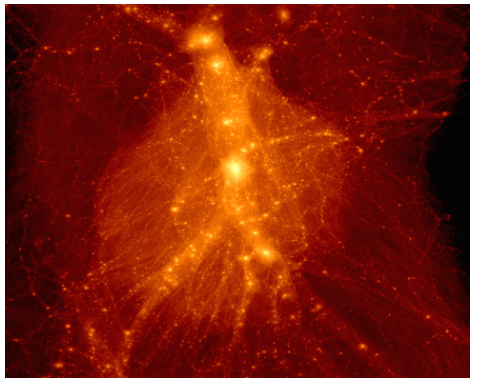
Solution
We solve the performance and visualization issue using binned statistics, e.g. histograms, density maps, and volume rendering in 3d. The Python package `vaex` can process a billion rows per second, and visualize it.

Websites
http://vaex.astro.rug.nl
https://github.com/maartenbreddels/vaex

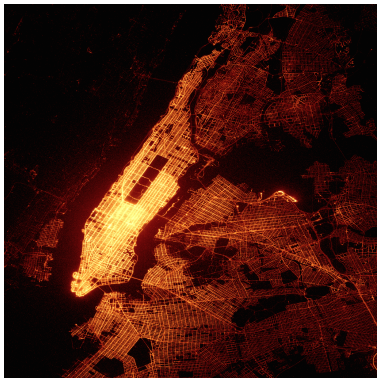
Summary
Yes, you can visualise 1 billion stars (e.g. the whole Gaia catalogue) in ~1 second.

Whole Gaia DR1 catalogue (1e9 stars)





Aquarius-A2 dark matter simulation (6e8 particles)



New York Taxi dataset (1e9 rows)

ipyvolume

The missing 3d plotting library

The Python Jupyter notebook is often the default environment for (data) scientist. However, it is (or was!) lacking a 3d visualisation library that integrates in the notebook.

Solution

ipyvolume:

- easy matplotlib like API
- volume rendering
- scatter + quiver plot
- animations

Upcoming

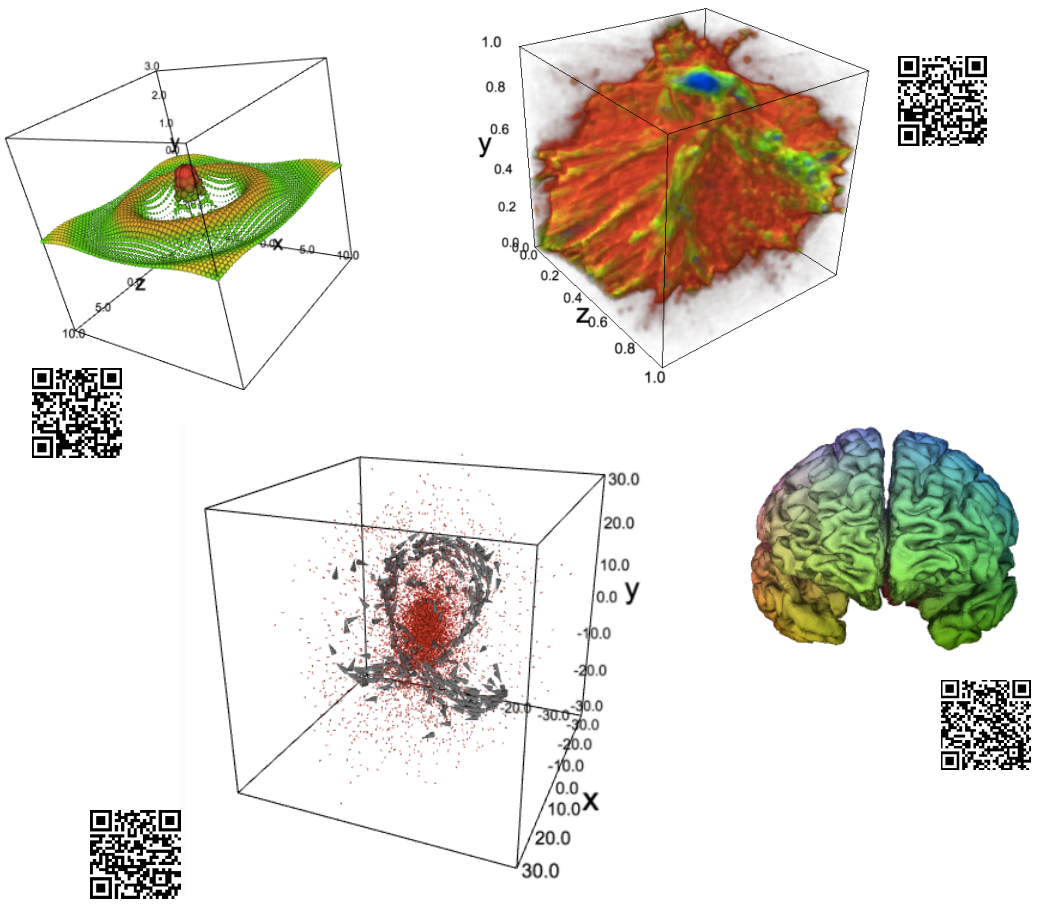
- Lines and mesh plotting
- 100x faster binary data transfer

Websites:

https://ipyvolume.readthedocs.io/
https://github.com/maartenbreddels/ipyvolume

Summary

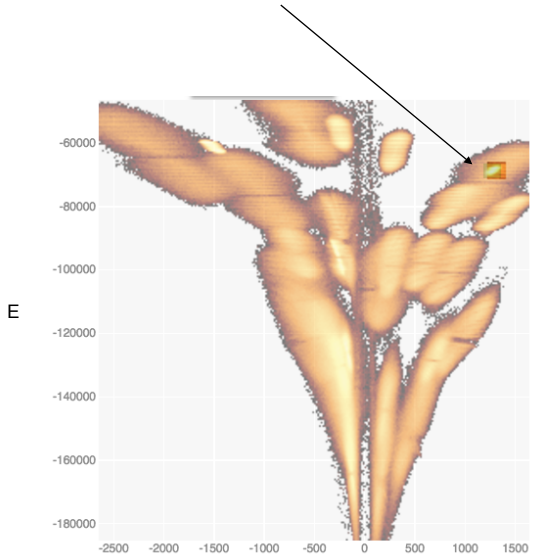
Interactive 3d plotting in the Jupyter notebook



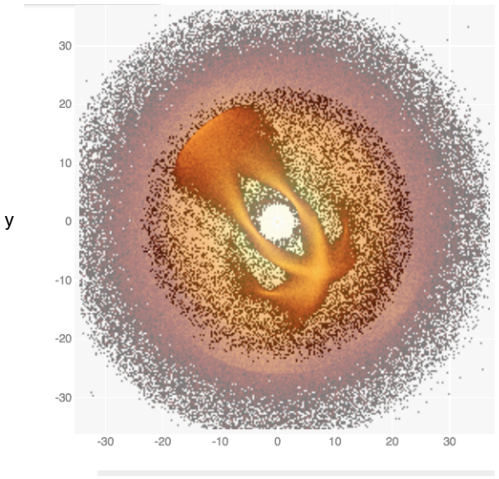
Psst, it also does virtual reality rendering (Google cardboard)

Combing vaex, bqplot and ipyvolume in the Jupyter notebook

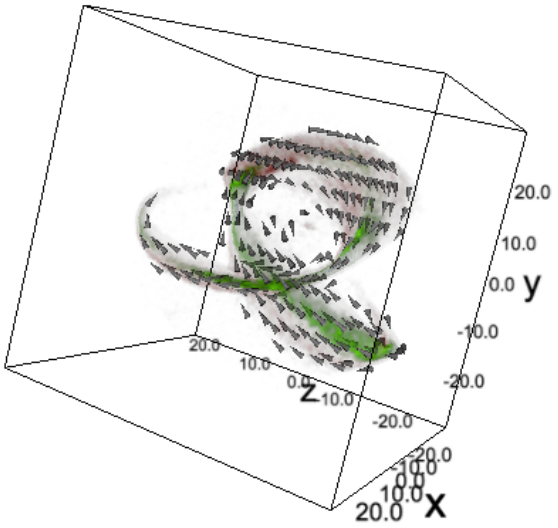
Selecting a clump in energy-angular momentum space



This selection corresponds to a stream, but projected in 2d



Using ipyvolume we van visualize this in 3d



- Poster S14.3 (2nd floor)
- https://github.com/maartenbreddels/ewass-2017
- http://vaex.astro.rug.nl
- https://github.com/maartenbreddels/ipyvolume