

# Realtime visualisation and exploration of the full Gaia catalogue with *vaex* (*on your desktop*)

Maarten Breddels  
&  
Amina Helmi

Part of Gaia DPAC/CU9/WP980



university of  
 groningen

faculty of mathematics  
and natural sciences

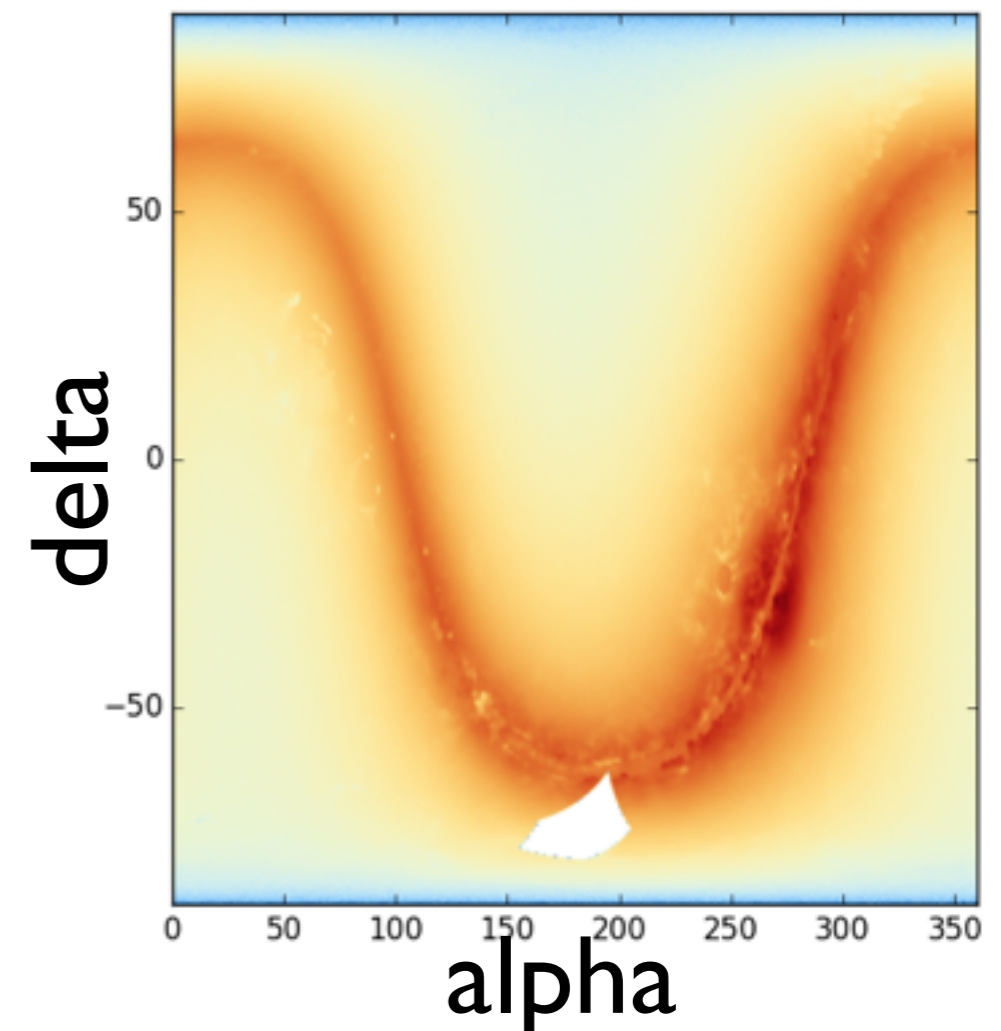
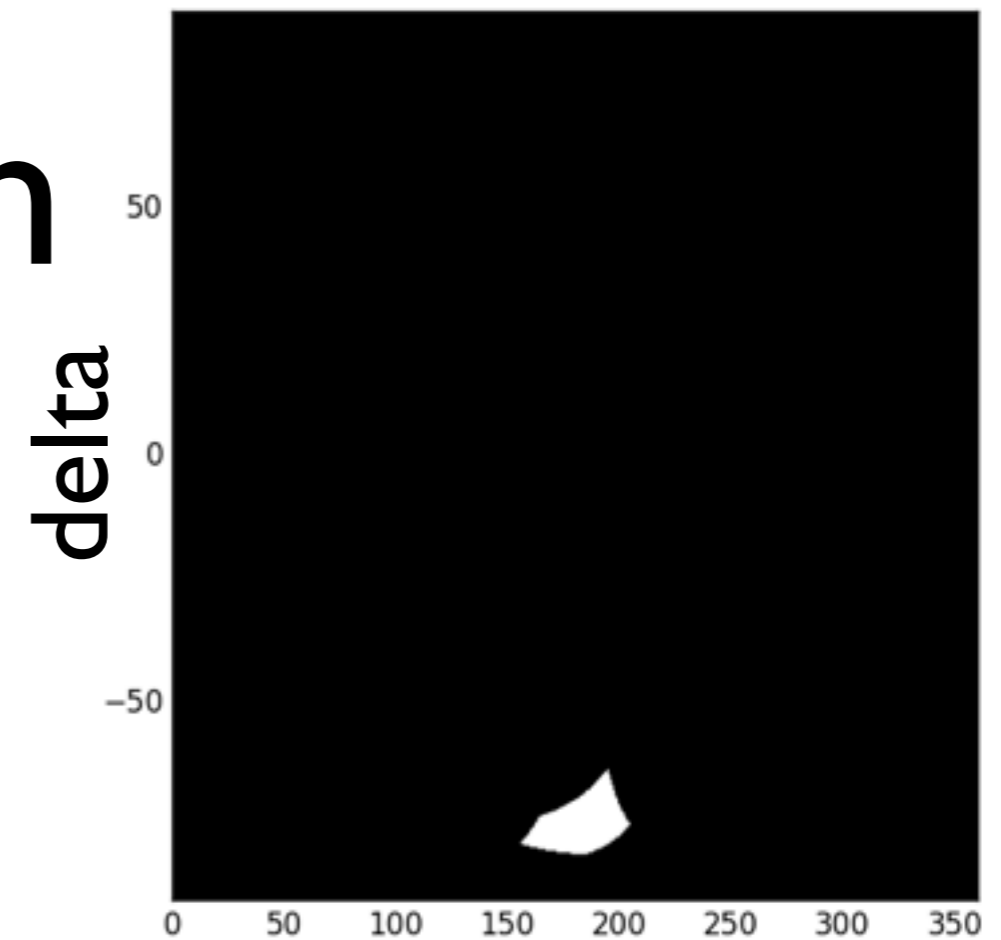
kapteyn astronomical  
institute

# Outline

- Motivation
- Visualisation
- Exploration
  - Selection and linked views
  - Subspace finding
- Visualisation in 3d
- How to get vaex and get your data in?

# Motivation

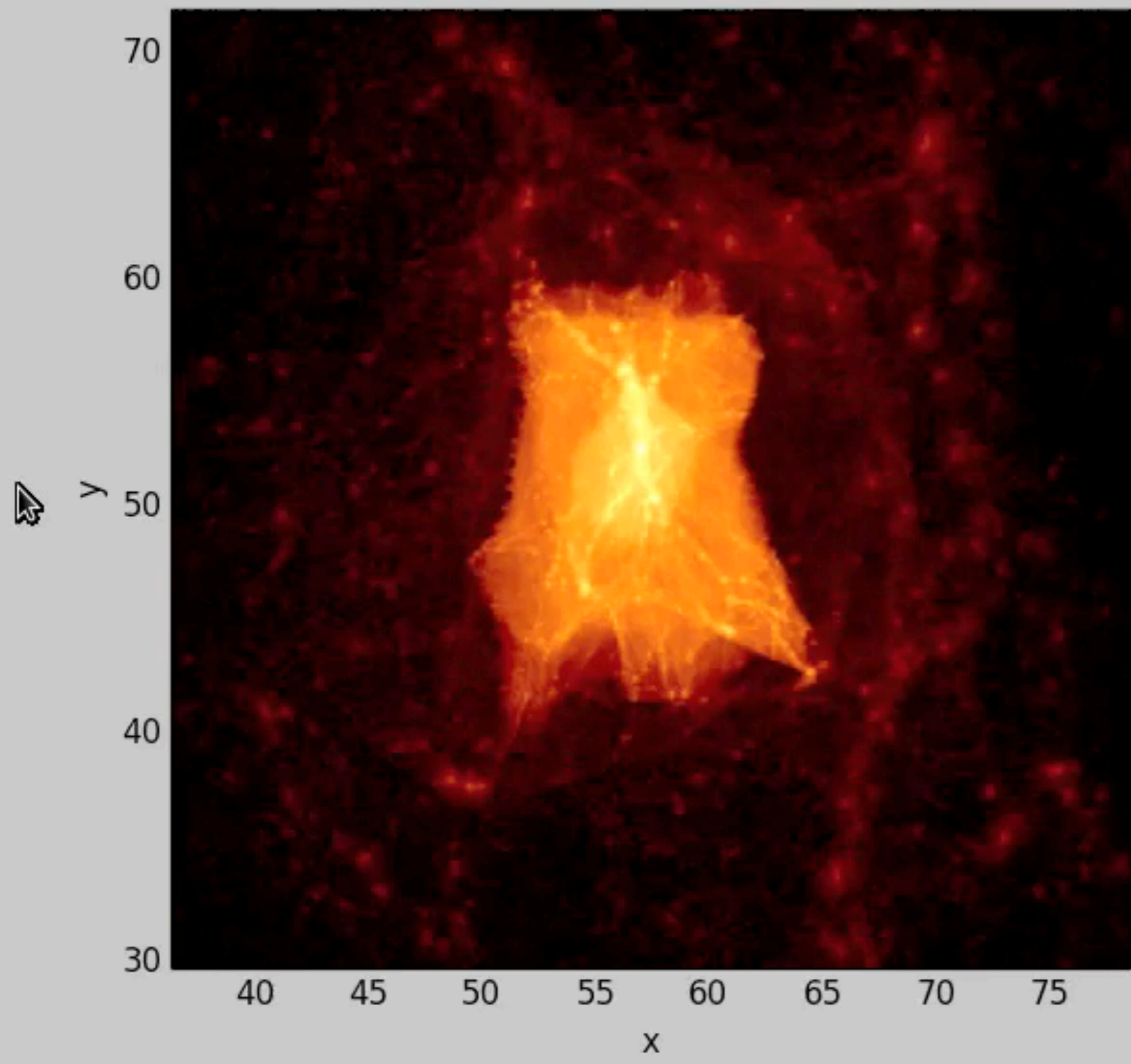
- Visualisation
  - ‘See’ the data
  - Scatter plots do not work well for  $10^9$  rows/objects (like Gaia)
  - Work with densities/averages in 1,2 and 3d
  - Designed to handle  $10^9$  rows/particles interactively
- Integrate with exploration tools
  - Large dimensionality  $\rightarrow$  many subspaces
    - Computer assisted help in finding interesting subspaces (but not fully automated)
  - Directly visualise selections/filtering
- Existing tools
  - TOPCAT comes close, but works with individual rows/particles, no exploratory tools
  - Other tools (not integrated, cumbersome) or small number of rows/particles
  - Your own IDL/Python code: a lot to consider to do it optimal (multicore, efficient storage, efficient algorithms, interactive becomes complex)



# Interactive visualization?

- Interactive?
  - $10^9 * 2 * 8 \text{ bytes} = 15 \text{ GiB}$  (double is 8 bytes)
  - Memory bandwidth: 10-20 GiB/s: ~1 second
  - CPU: 3 Ghz (but multicore, say 4): ~1 second
  - Few cycles per row, simple algorithm
    - Histograms/Density grids
  - Order of 1 second should be possible
    - If it fits/cached in memory, otherwise sdd/hdd speeds (30-100 seconds)
    - column bases storage
- Alternative (laptop)
  - shuffle your data and show 10%, 1% 0.1%....

sel->SAMP compact+ fullscreen toolbars Move Pick Lasso Replace Zoom to rect Zoom out Reset view



**Main**

exchange x and y

exchange colormap

X-axis: x

Y-axis: y

amplitude= log(counts+1)

colormap= afmhot

title=

weight=

---

Vector field

---

Dispersions

---

Display

Export figure Aspect=1 Undo Redo Shuffled Disjoined Lock axis

# The tool: vaex

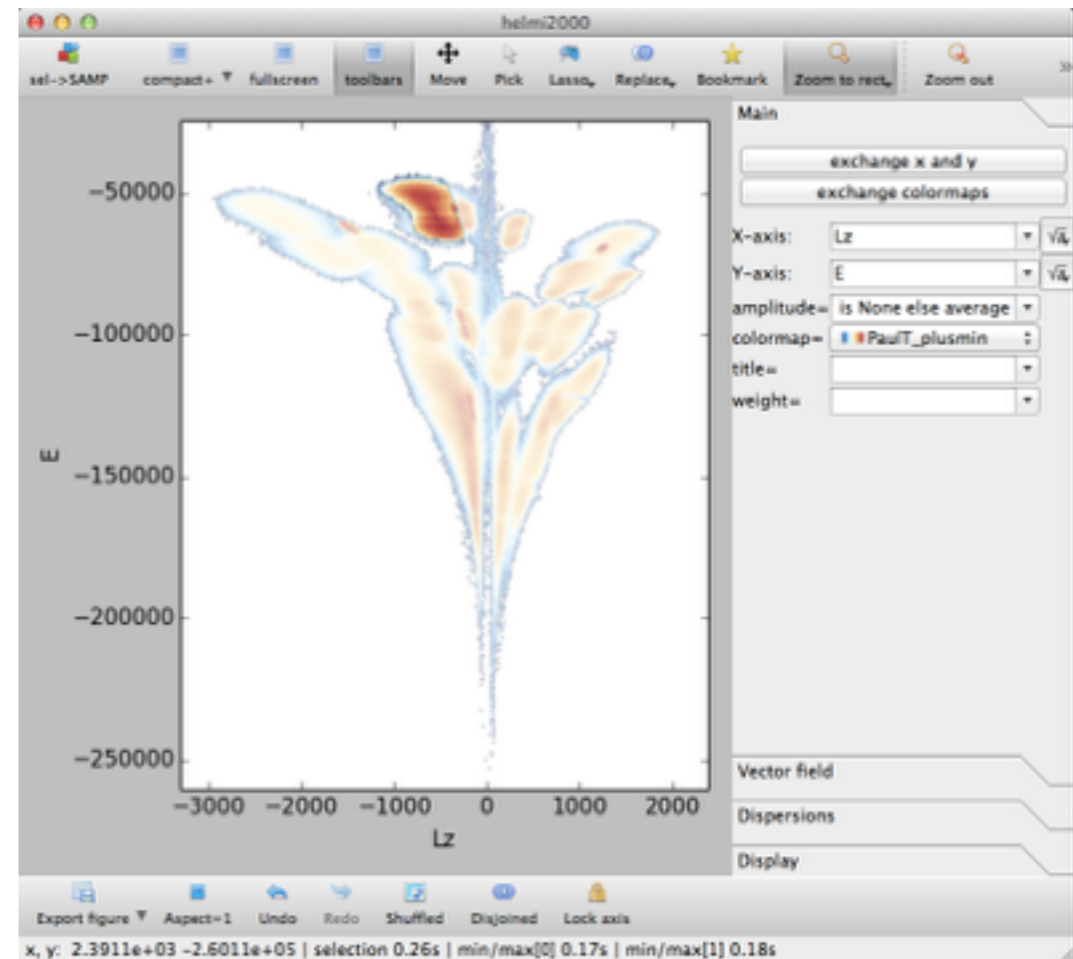
## visualisation and exploration

- what it can do
  - efficiently generate density plots in 1, 2 and 3d
  - interactive/navigation
    - zoom, pan, considers aspect ratio
  - selecting groups / single objects
  - help in finding interesting subspaces / clustering
- Why interesting for you?
  - Visualize (full) Gaia catalogue
  - Any other catalogue: SDSS, KiDS, Pan-STARRS, ...
  - Particle based simulation, Gadget, SPH, Amuse

# The tool: vaex

## visualisation and exploration

- What it can visualise
  - 1d: histograms
  - 2d: density plot/heatmap
  - 3d: volume rendering
  - vector fields, tensor fields
- Custom expression:  
 $\log_{10}(\sqrt{x^2+y^2})$
- selection of groups



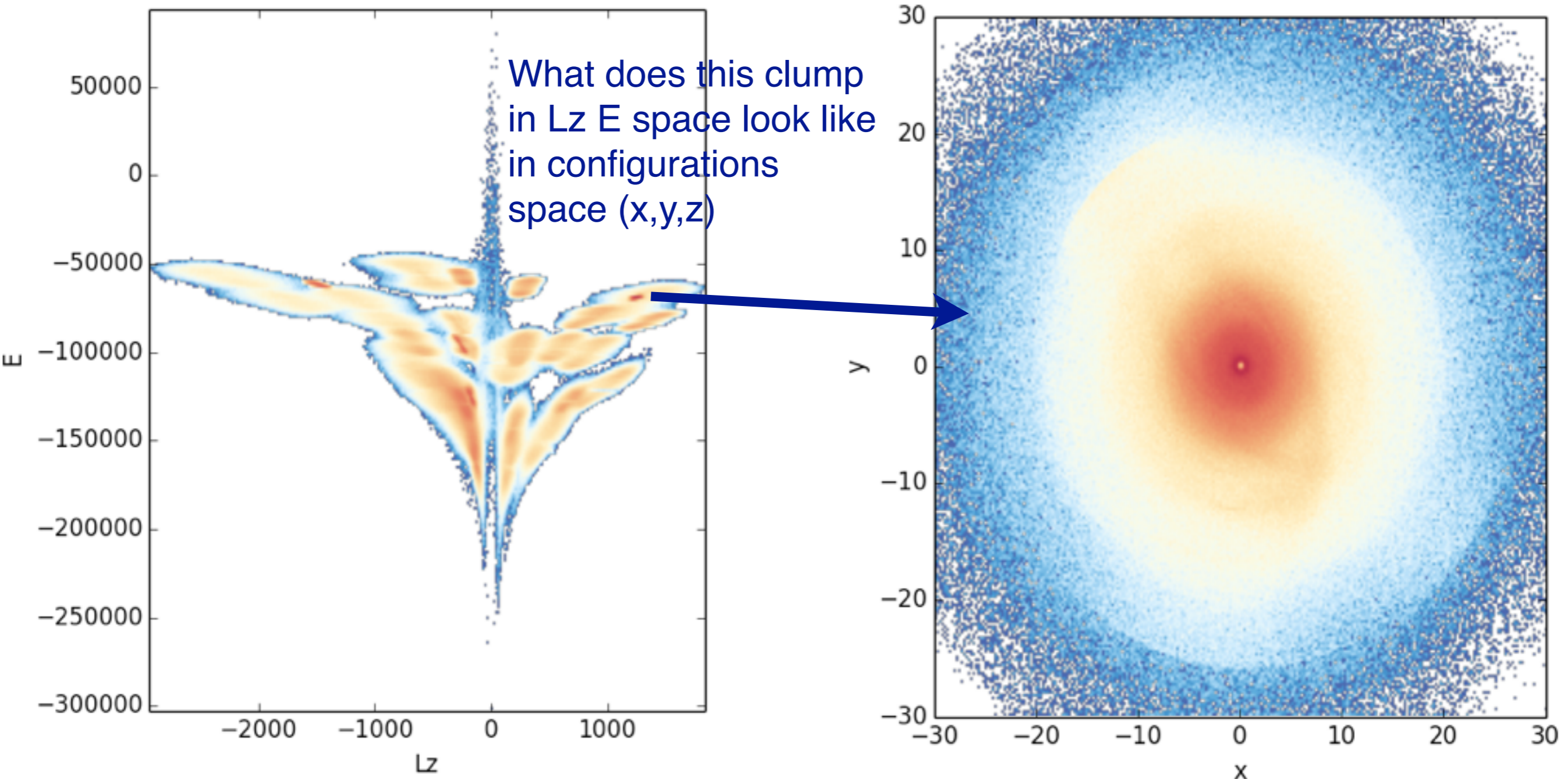
# Outline

- Motivation
- Visualisation
- **Exploration**
  - Selection and linked views
  - Subspace finding
- Visualisation in 3d
- How to get vaex and get your data in?



# Exploration: Selections and linked views

- Helmi and de Zeeuw 2000
  - build up of a MW (stellar) halo from 33 satellites
  - $3.3 * 10^6$  particles
  - Almost smooth in configuration space (x,y,z)
  - Structure visible in E-Lz-L space



VaeX

helmi2000-FeH-s2

Name: helmi2000-FeH-s2  
Columns: 11  
Length: 3,300,000

Plotting:

Data:

Fraction used: 100.0000%

# Exploration: Automated

- High dimensionality  $\rightarrow$  many subspaces
- E-L, E-y, E-x, E-vx, E-vy, E-vz, E-z, E-Lz, L-y, L-x, L-vx, L-vy, L-vz, L-z, L-Lz, y-x, y-vx, y-vy, y-vz, y-z, y-Lz, x-vx, x-vy, x-vz, x-z, x-Lz, vx-vy, vx-vz, vx-z, vx-Lz, vy-vz, vy-z, vy-Lz, vz-z, vz-Lz, z-Lz
- Can we automate this / at least help?
- Ranking subspaces using Mutual Information

# Mutual information

- Measure the information loss between  $p(x,y)$  and  $p(x)p(y)$   $p(x) = \int p(x,y) dy$
- information loss measured using the KL divergence:  $I(X;Y) = \int_Y \int_X p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) dx dy,$
- In short:
  - How much does breaking up correlation (not just linear) change density distribution?

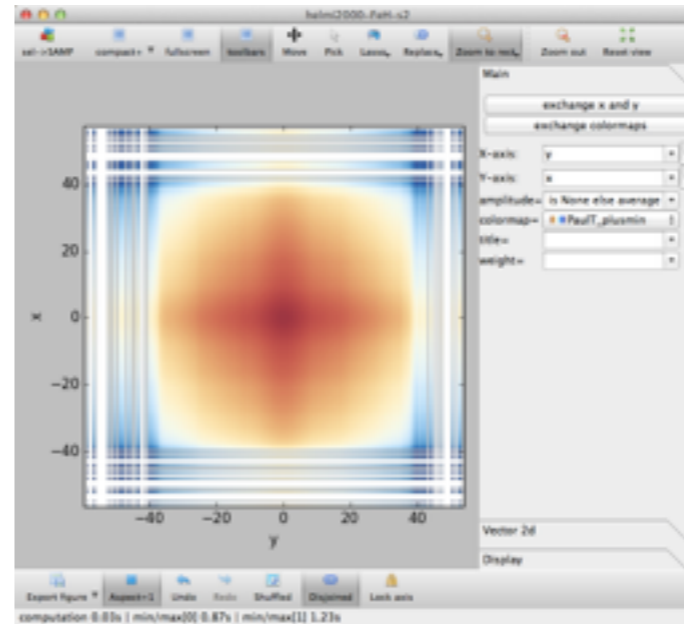
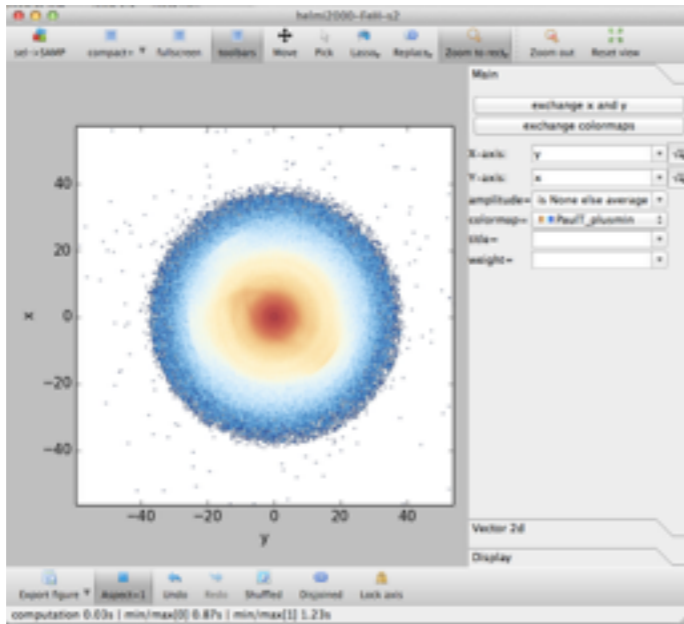
# Ranking by Mutual information

$$p(x,y)$$

$$p(x)p(y)$$

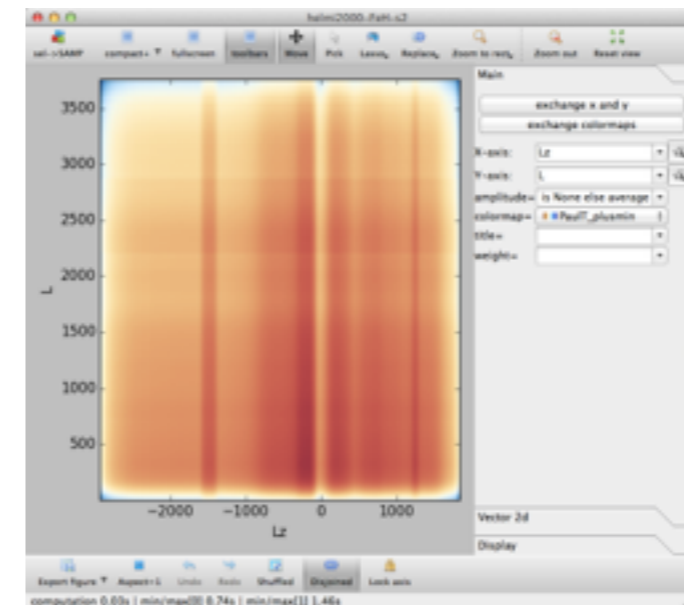
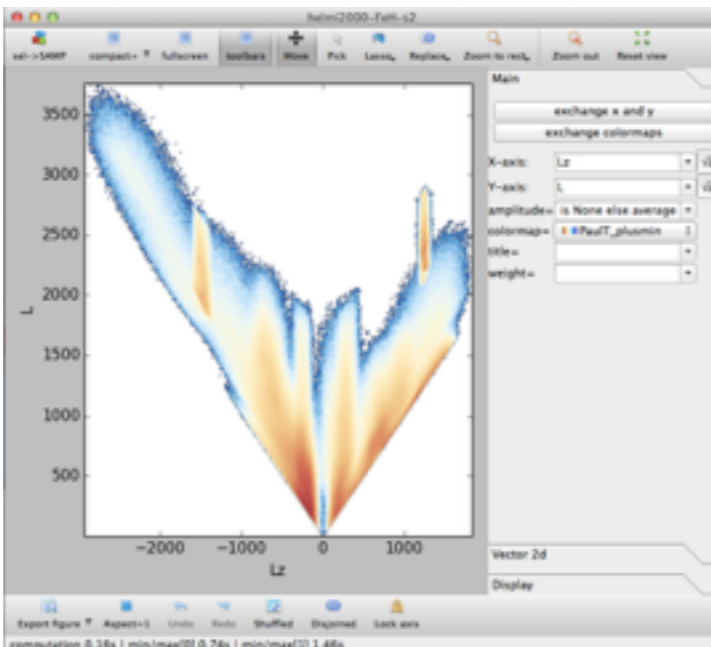
$$I(X;Y)$$

$y,z$



0.003

$Lz,L$



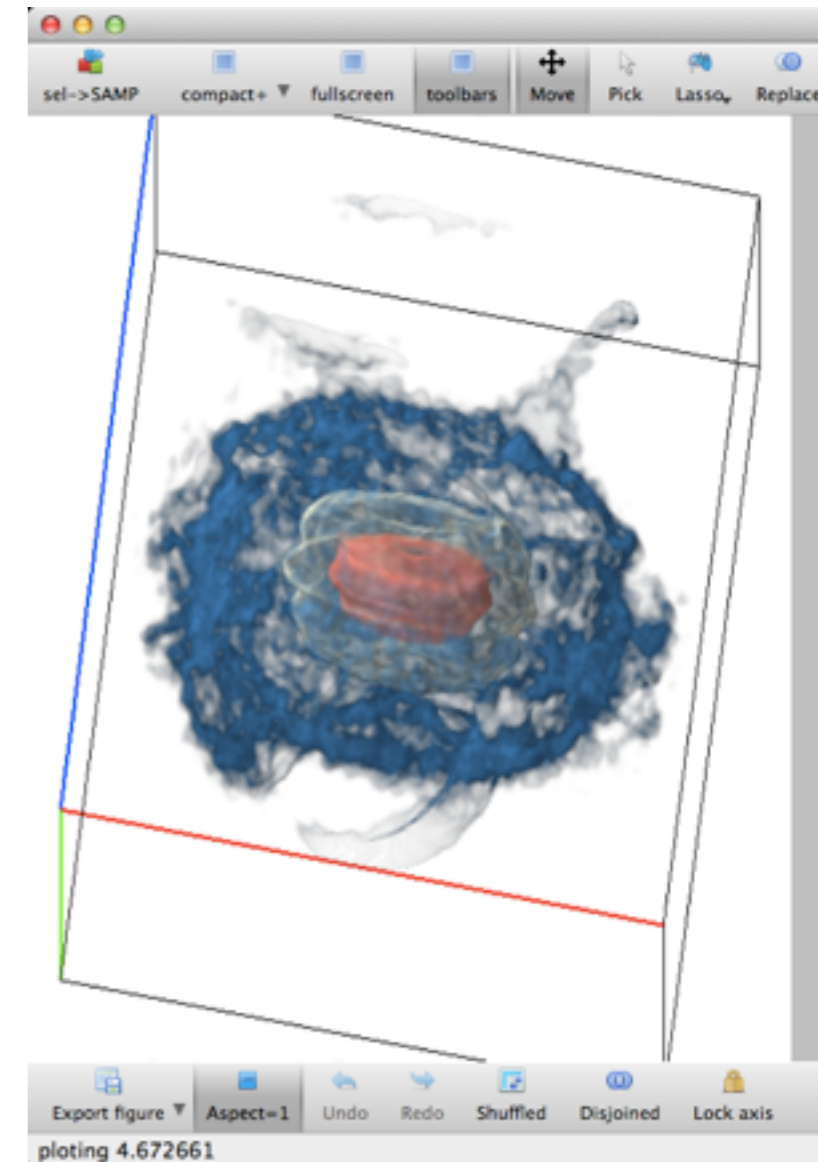
0.837

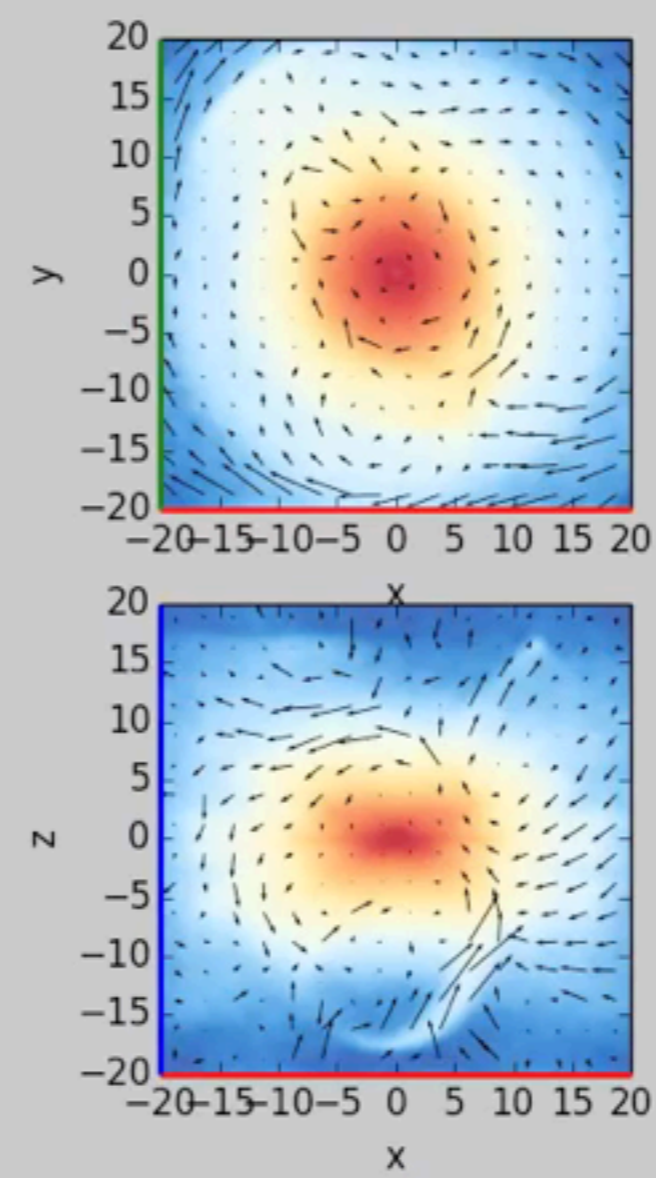
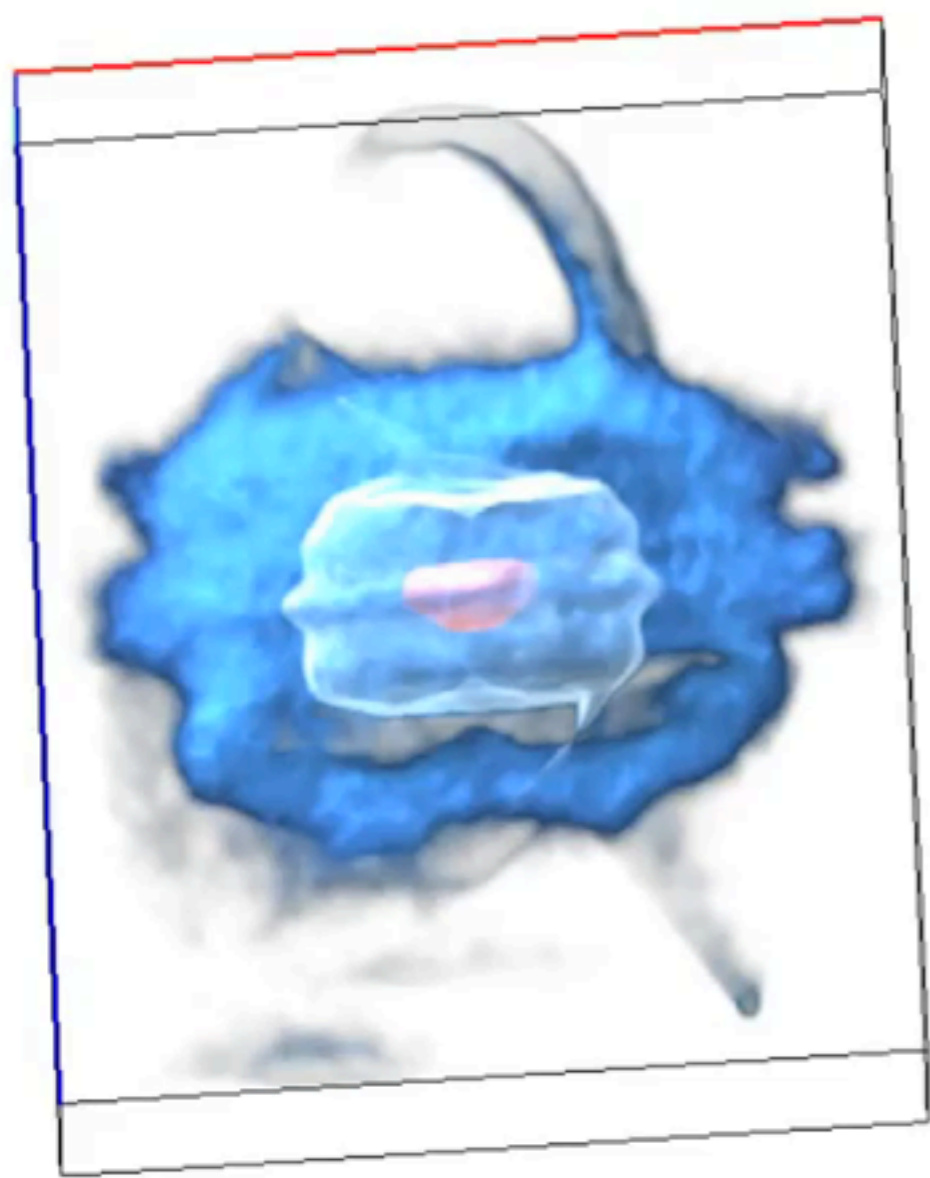
# Outline

- Motivation
- Visualisation
- Exploration
  - Selection and linked views
  - Subspace finding
- **Visualisation in 3d**
- How to get vaex and get your data in?

# 3d visualisation

- Not as good as going from 1d to 2d
- Occlusion, difficult to visualise
- Volume rendering
  - difficult to get a good transfer function
- Gaussian filtering and lighting effects helps





Main

Vector field

Dispersions

Transfer function

orbit

mean: 0.063  
sigma: 0.029  
opacity: 0.029

mean: 0.339  
sigma: 0.025  
opacity: 0.062

mean: 0.900  
sigma: 0.025  
opacity: 0.200

brightness: 3.069

min\_level: 0.024  
max\_level: 1.000

ambient: 0.606  
diffuse: 1.000  
specular: 1.000  
specular\_n: 5.000



# How to get/use

- [www.astro.rug.nl/~breddels/vaex](http://www.astro.rug.nl/~breddels/vaex) (or google 'vaex visualisation')
- standalone binary for OS X and Linux (just download and start)
- Can generate mock datasets, and comes with Helmi and de Zeeuw 2000 dataset
- How to get your data in:
  - hdf5: examples for Python/IDL/c
  - fits: bintable (row based = slower, convert to hdf5)
  - VOTable (using SAMP) (much slower, only for small datasets)
    - If TOPCAT can read it, vaex can
- Gaia data: unknown

# Summary

- vaex: visualisation and exploration
  - of large datasets  $10^{6-9}$  rows
  - 1-2 and 3d visualisation using densities
    - ~6d with vector field overlaid
  - Explore using selections+linked views or automated ranking of subspaces
- Main goal is Gaia catalogue, but tested/suitable for
  - other catalogues
  - simulations (N body)

