

# Visual analytics with the Gaia archive and other Big Data

---

André Moitinho

University of Lisbon - CENTRA



**centra**  
centre for astrophysics and gravitation

# Gaia — Objective: the Milky Way

## → GAIA DATA RELEASE 1



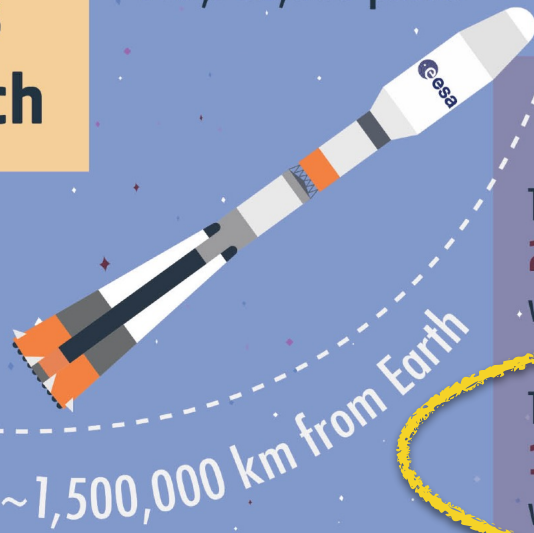
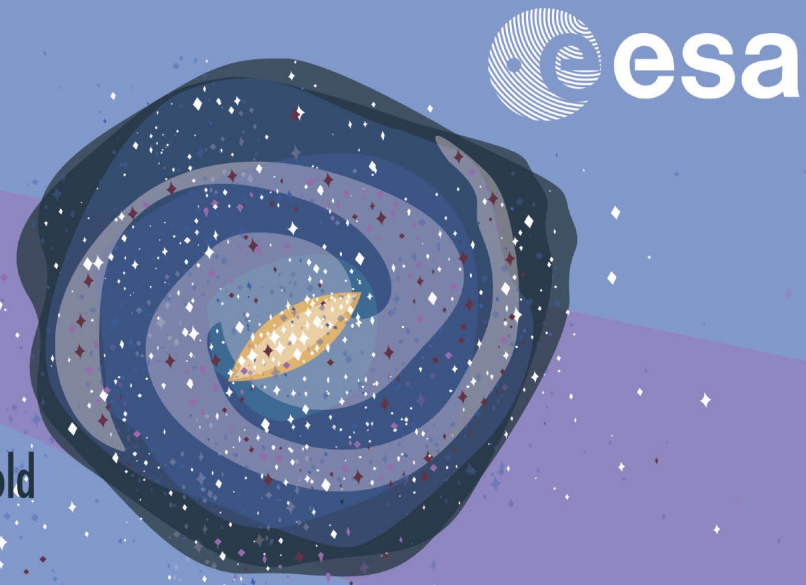
**14**  
September 2016

**1000 days**  
since launch

- 1 spacecraft
- 2 telescopes
- 10 mirrors
- 1 camera
- 106 CCDs
- 937,782,000 pixels



**1 Milky Way**  
>100,000,000,000 stars  
~13,000,000,000 years old



~1,500,000 km from Earth

### Content of the release

Total number of sources in primary astrometric data set:  
**2,057,050**  
with position, magnitude, parallax & proper motion

Total number of sources in secondary astrometric data set:  
**1,140,622,719**  
with position & magnitude

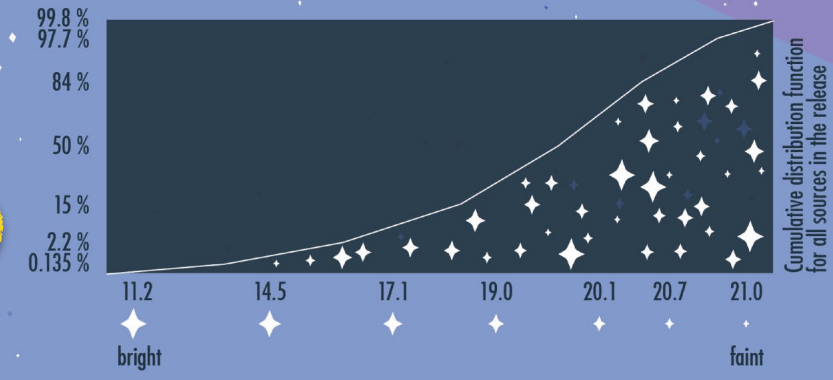
### 3194 Variable stars

- 599 Cepheids (43 new discoveries)
- 2595 RR Lyrae (343 new discoveries)

### 2152 Quasars

with position & magnitude  
Data collected over 14 months

### Magnitude distribution



### Data challenge so far

- >50 billion focal plane transits
- >110 billion photometric observations
- >9.4 billion spectroscopic observations
- ~120,000 hours of computing time to identify stars
- 6 data processing centres

### 1 day on Gaia

- 637,000,000 astrometric measurements
- 155,000,000 photometric measurements
- 13,000,000 spectroscopic measurements
- 70,000,000 celestial objects
- 40 GB of data downlinked to Earth

2022: uarcsec positions

# The growing volume of Astronomical data

---

- Gaia - 1 billion - Spectrophotometry, parallaxes, proper motions, radial velocities, time series
- SDSS - ~2 billions, mostly extra gal. ~750.000 MW spectra. Optical/NIR
- LSST - Future - Optical/NIR
- PanSTARRS - Interesting releases in the future. Optical
- IPHAS - 219 million, R,I, Ha
- VVV - Millions. NIR, Inner MW
- **How can we deal with all these (big) data?**

“Humans are above all visual beings (...)

**Neural** substrates serving the **visual** sense, (...) occupy an astonishing **30 to 40 percent** of the cerebral **cortex**’ total surface area.”

Dr. A. Bartels, MPI Bio. Cyb.

It is thus natural that visual insight is a starting point and even the guiding reference for scientific thought.

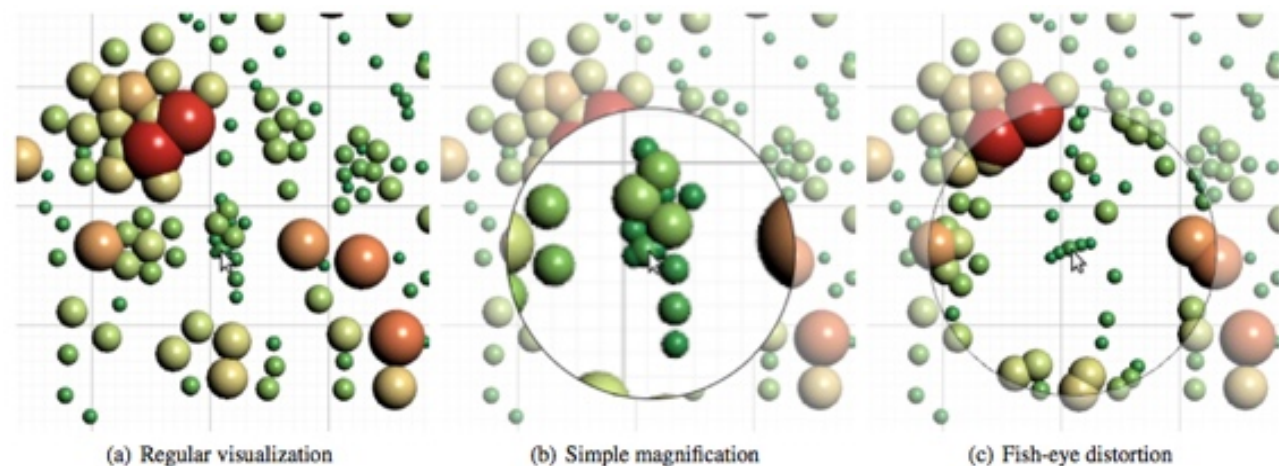
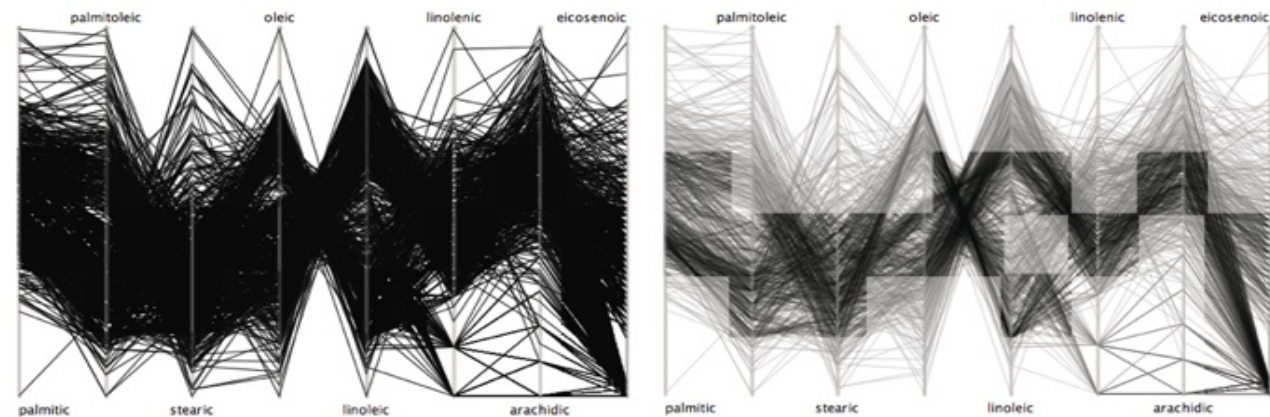
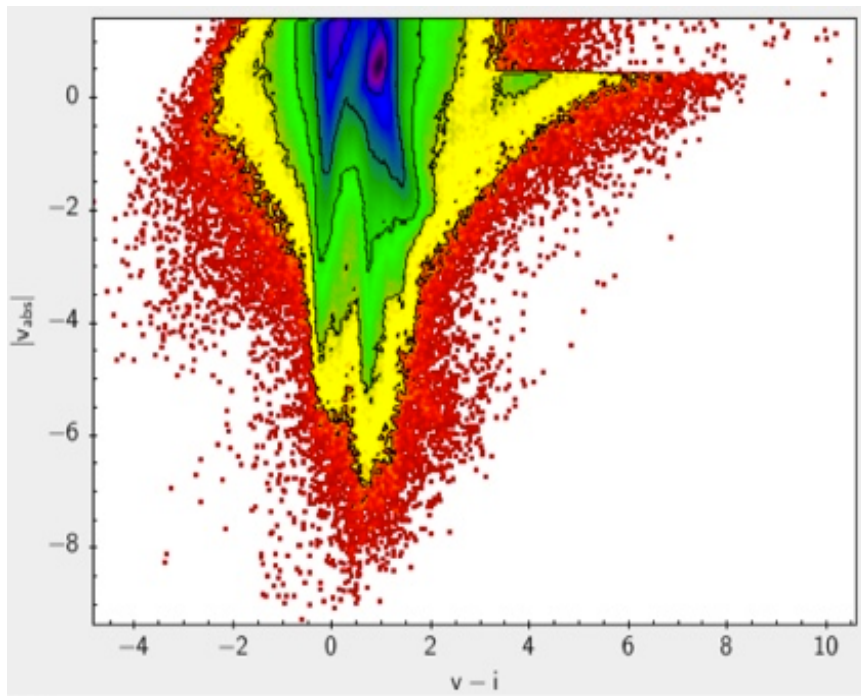


# Challenges in visual exploitation of Big Data

- **Physical size of the archives:** Hardware resources, including bandwidth: data servers [take the programs to the data]
- **Interactivity.** Exploration is interactive -> responsive
- **Analytics:** Too many data to represent and too many high-dimensional\* interrelations: **Data stunning!**

# Challenges in visual exploitation of Big Data

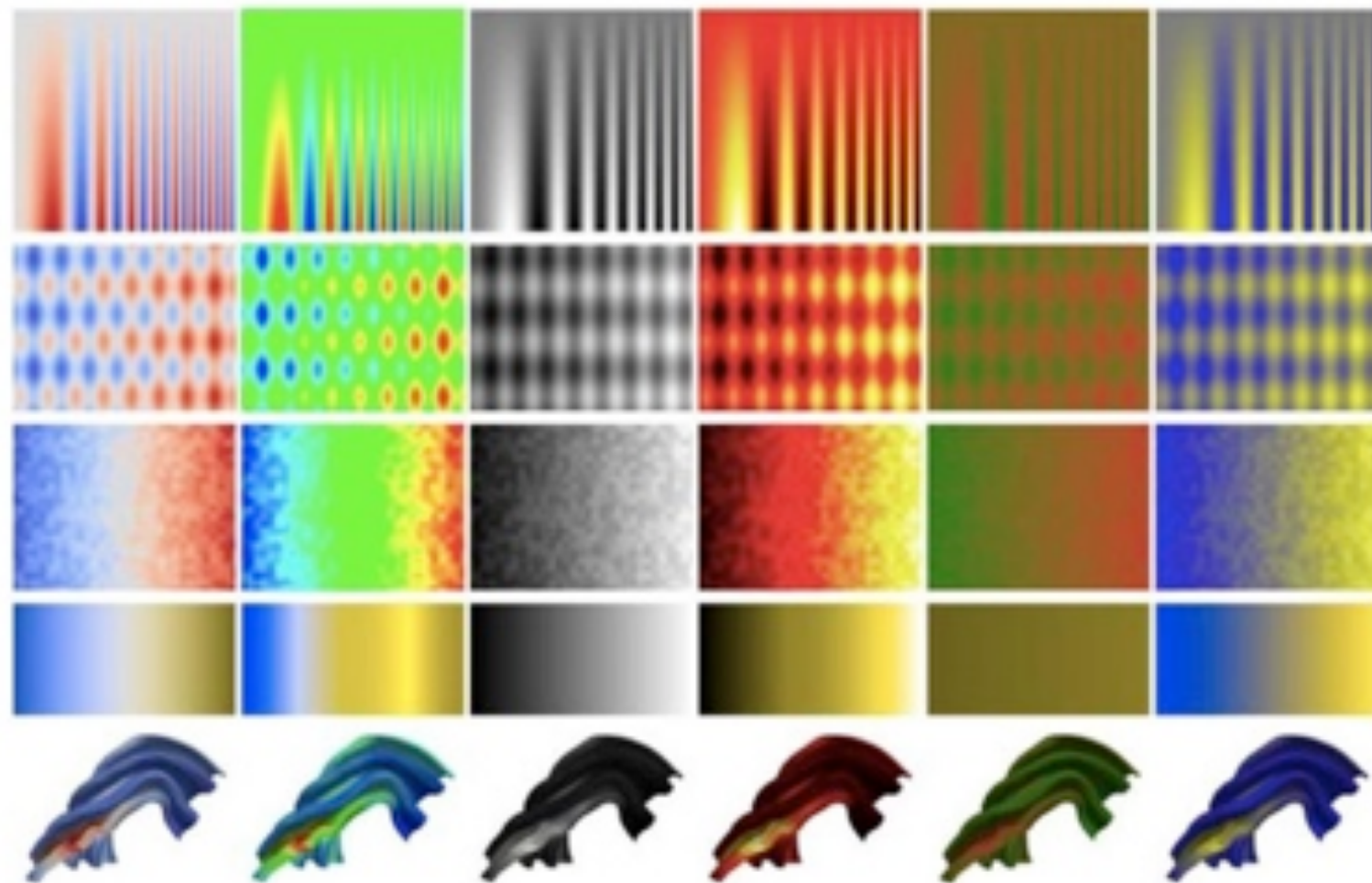
- **Visualisation and analysis challenges** → **Data stunning (Confusion)**



Adopt new habits in data visualisation. Presets.

# Challenges in visual exploitation of Big Data

- Visualisation and analysis challenges
  - **Habits !!**



Comparison of colour maps. From left to right, cool-warm, rainbow, grayscale, heated body, isoluminant, and blue-yellow. And from top to bottom, representations showing spatial contrast, a low-frequency, high-frequency noise, approximation of how the colour map is viewed deuteranope colour-deficient vision and its effect in 3D shading. From Moreland, 2009.

## **So we want facilities that**

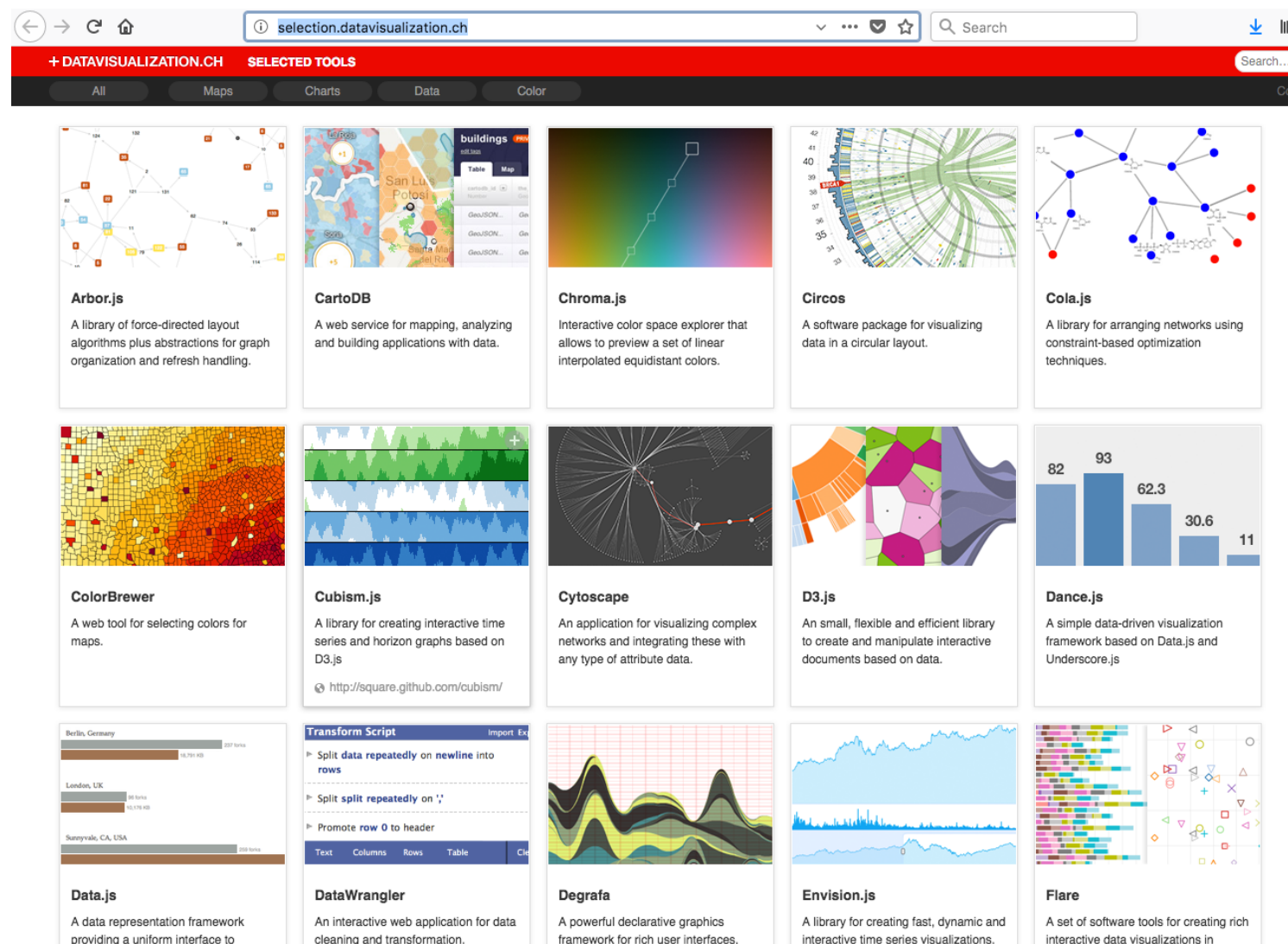
- are up to the technical challenges
- provides the necessary functionalities (for data analysis)
- are preset for Big Data exploration



# What's available

- A lot of visualisation libraries
- It's in fashion!

Check, e.g. <http://selection.datavisualization.ch/>



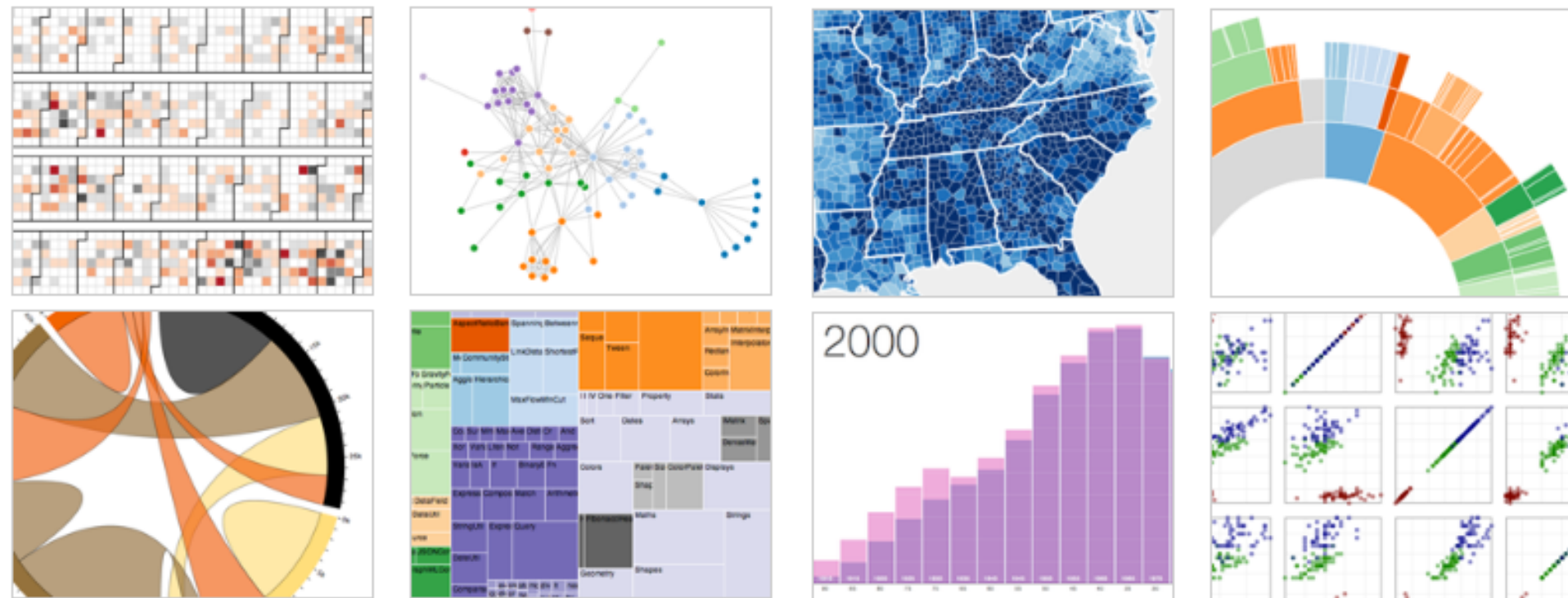


# Visualization Frameworks, Toolkits, Systems

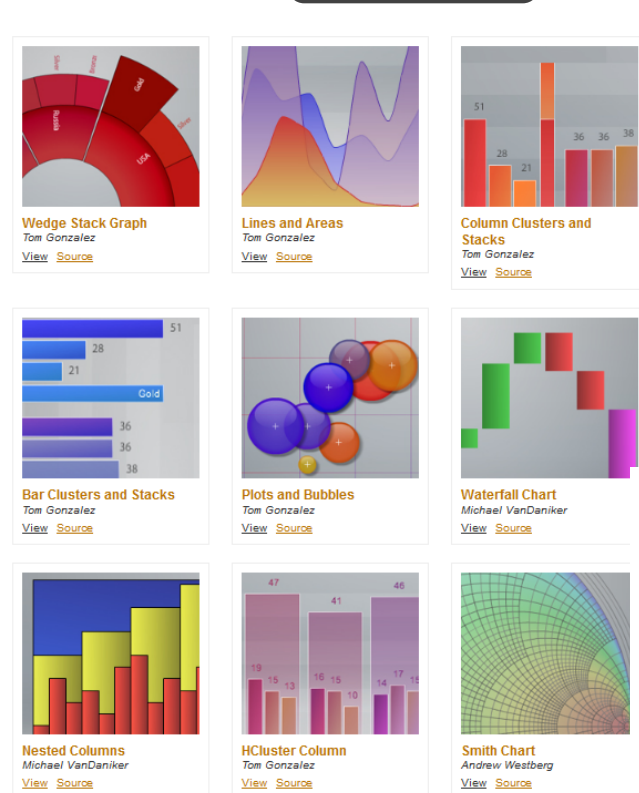
## Google



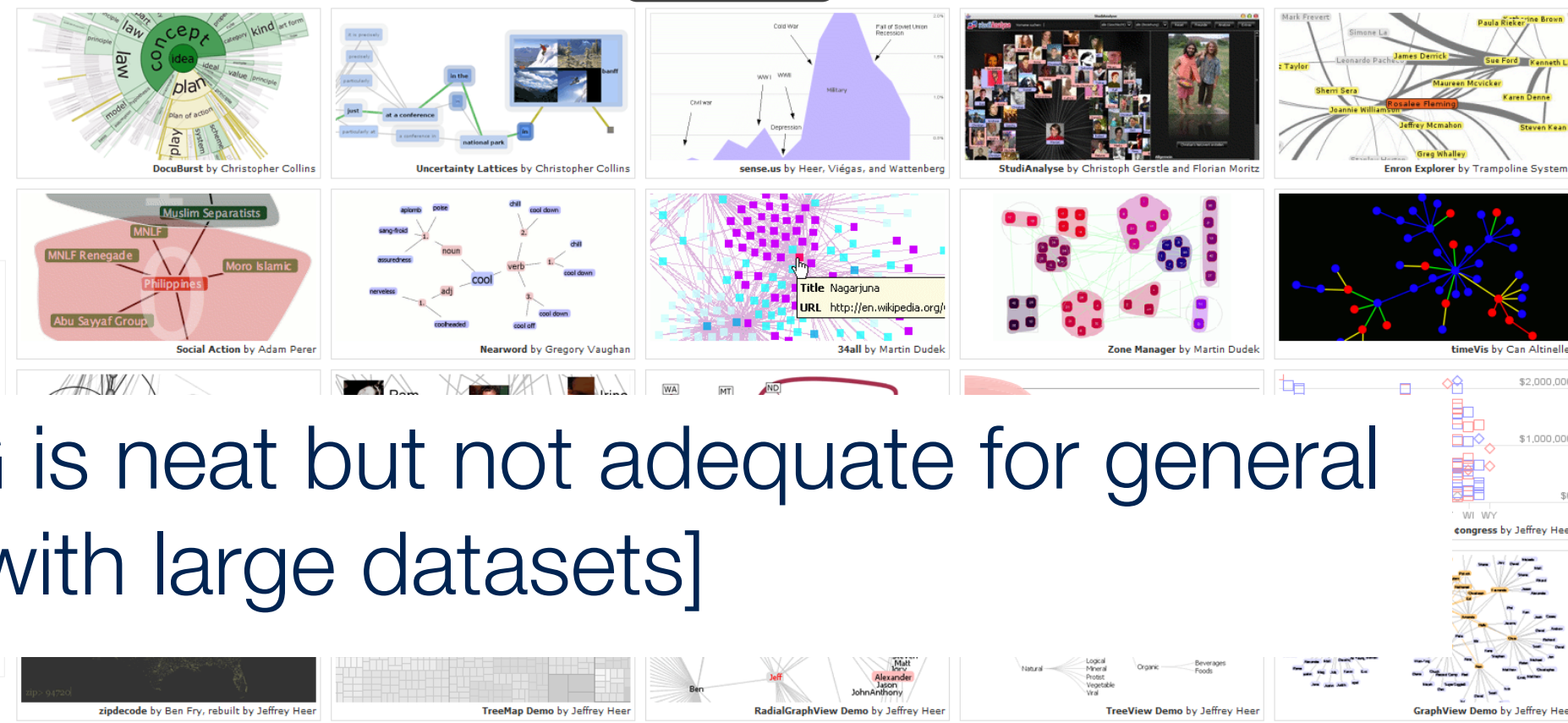
## D3.js



## Axiis



## Prefuse

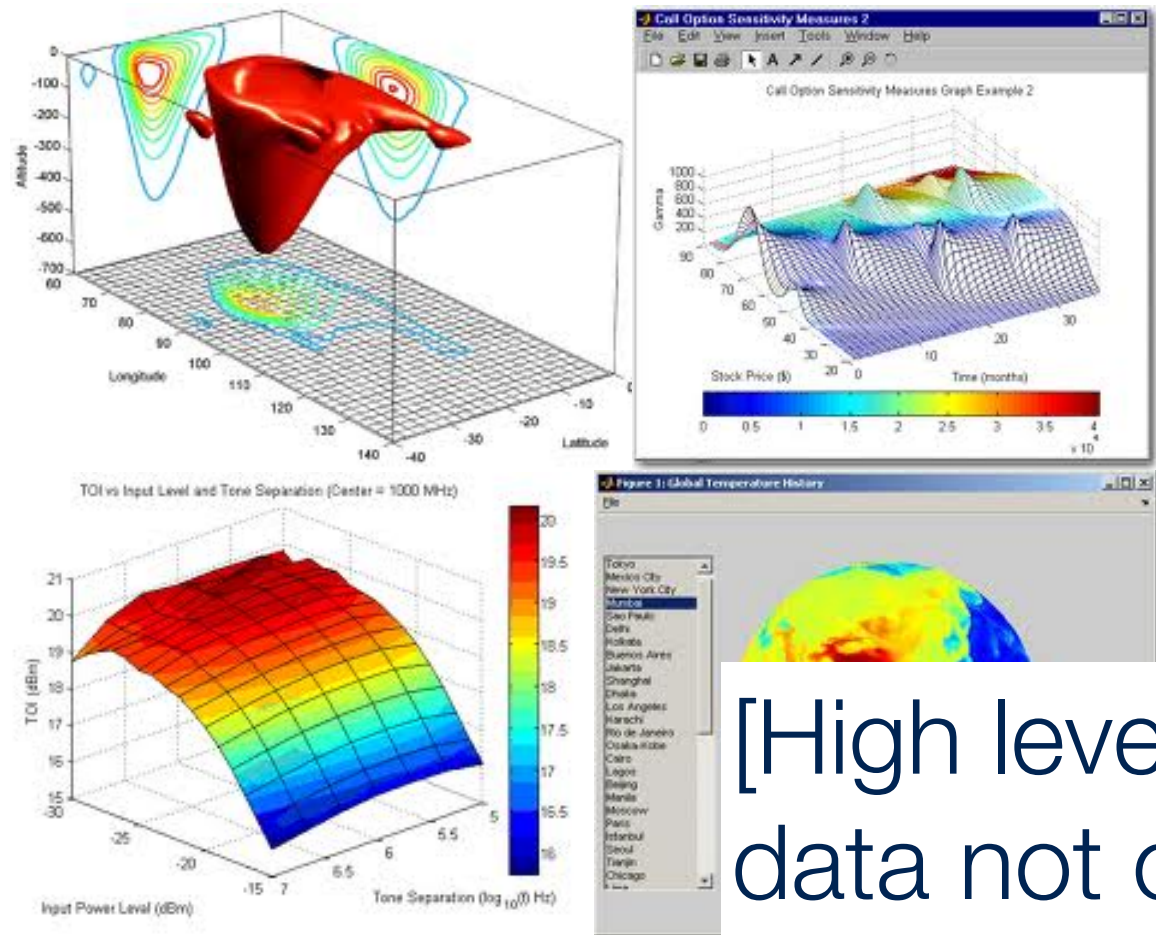


[SVG is neat but not adequate for general use with large datasets]

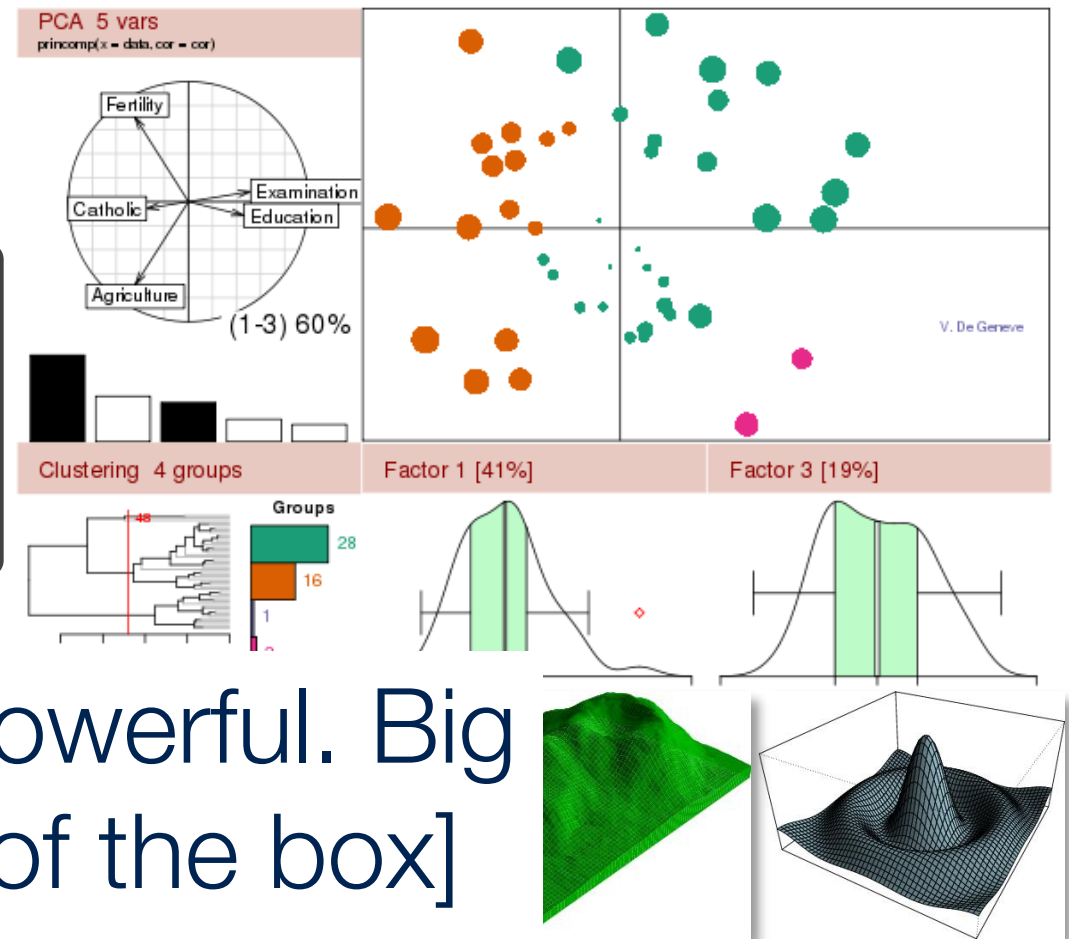


# Visualization Frameworks, Toolkits, Systems (cont.)

Matlab

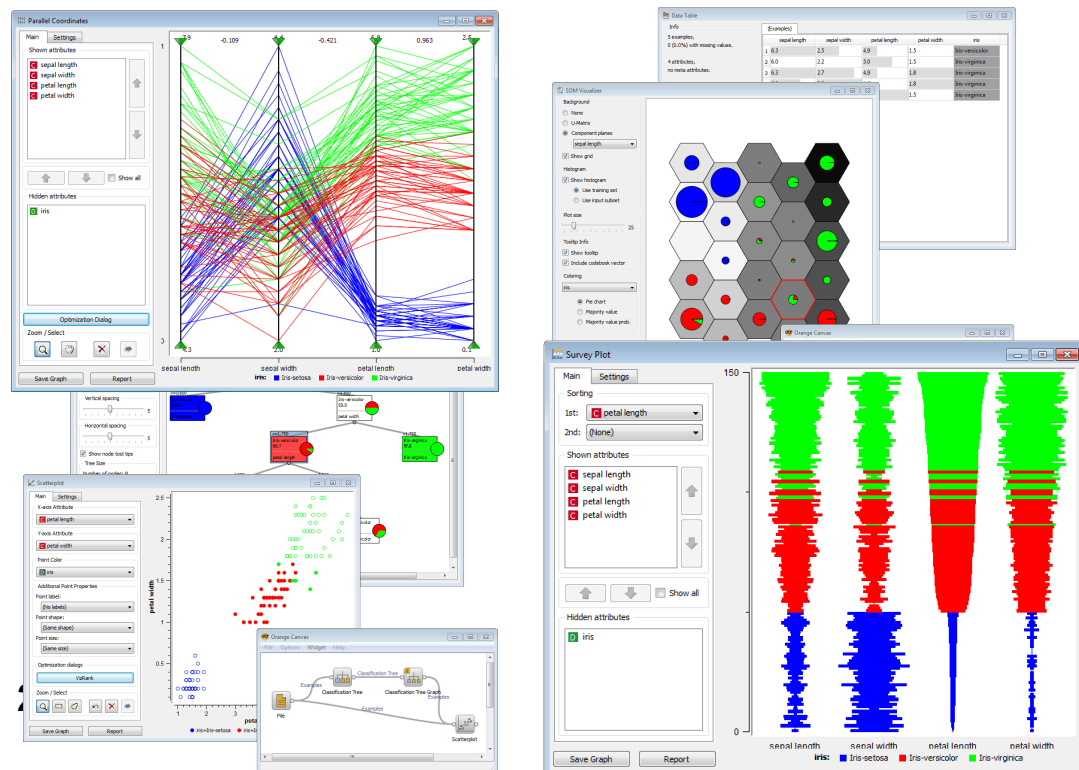


R Project

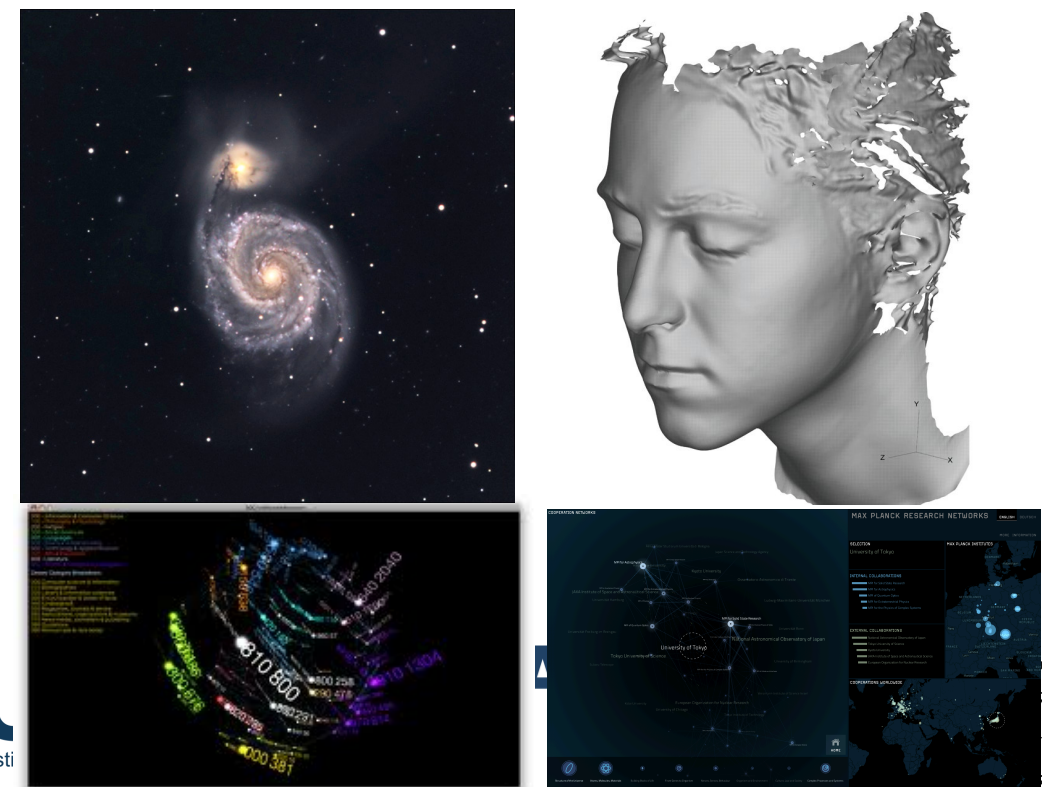


[High level. Powerful. Big data not out of the box]

Orange



Processing





# Visualization Frameworks, Toolkits, Systems (cont.)

## GAV

Brushing

Colour Map

Interaction Managers

Snapshot Manager

Animation Manager

Selection Manager

Visibility Manager

Functional Visualization Component Level

Atomic Visualization Component Level

Analyse

Data Transform

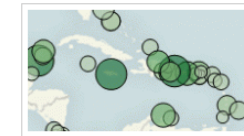
Data Handler

Data & Analysis

Data Providers



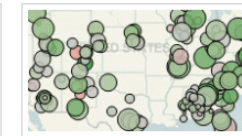
## Tableau



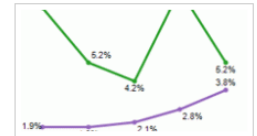
Internet Marketing



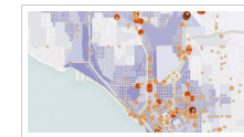
Mobile Business Intelligence



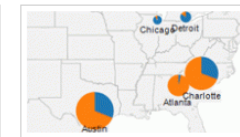
Healthcare



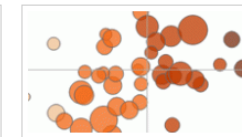
Consumer Packaged Goods



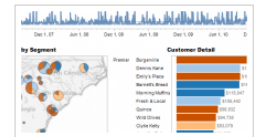
Mapping



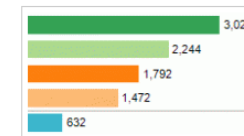
Database Marketing



Game Design



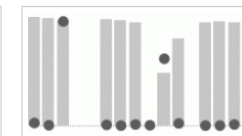
Business Dashboards



Human Resources



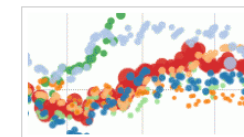
Time Series Analysis



Manufacturing



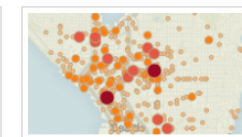
Telecommunications



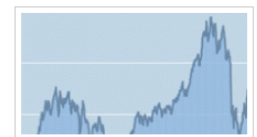
Securities and Investments



Oil and Gas

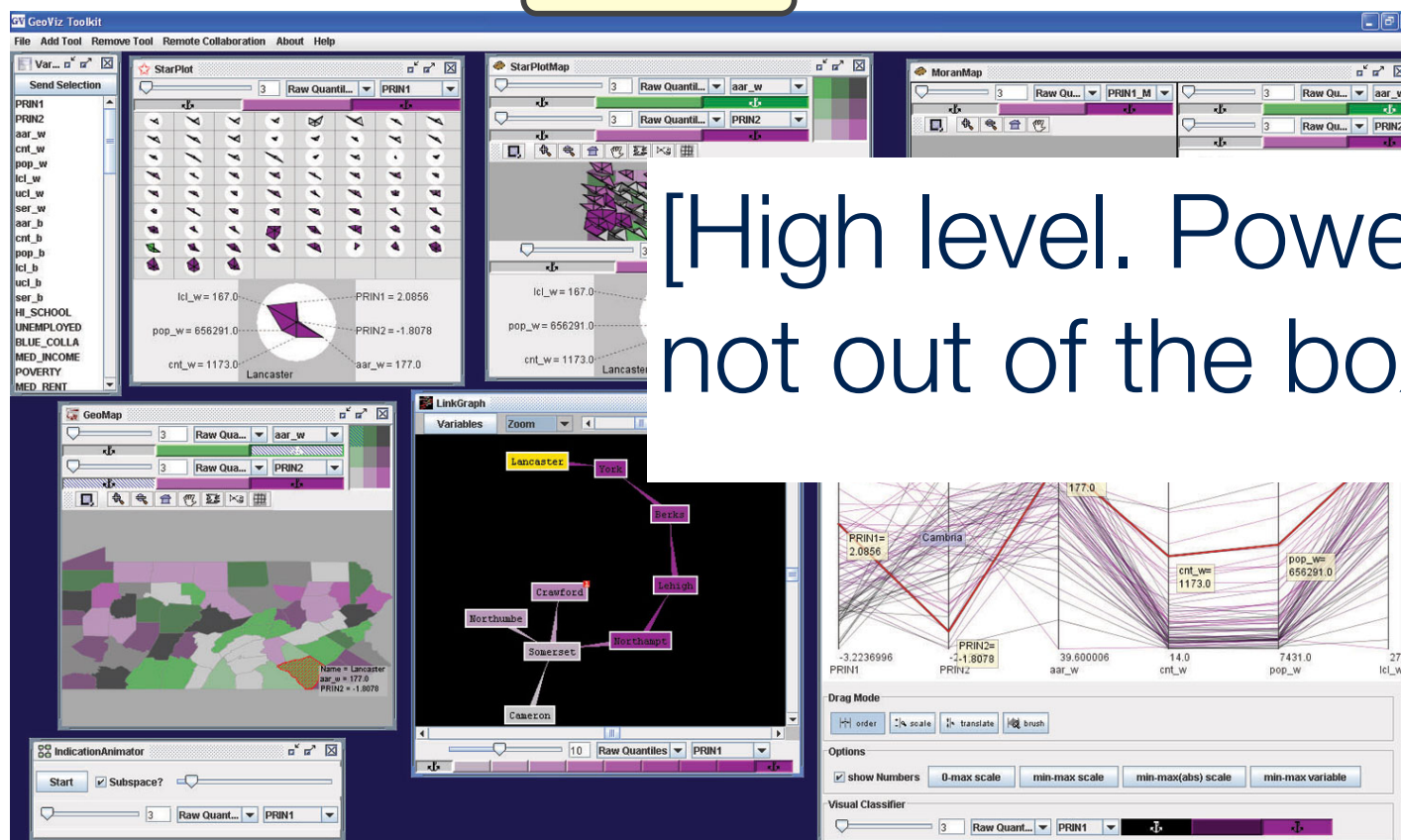


Government

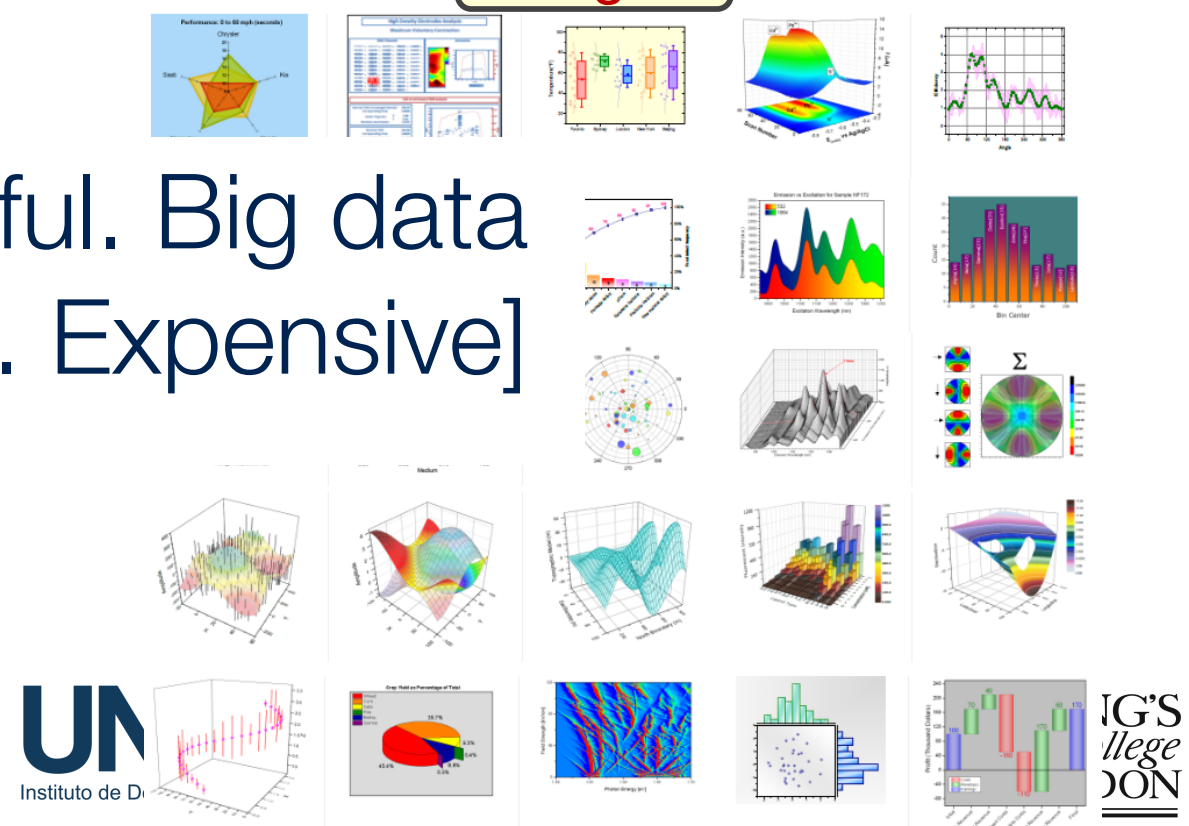


Banking

## GeoVista



## Origin

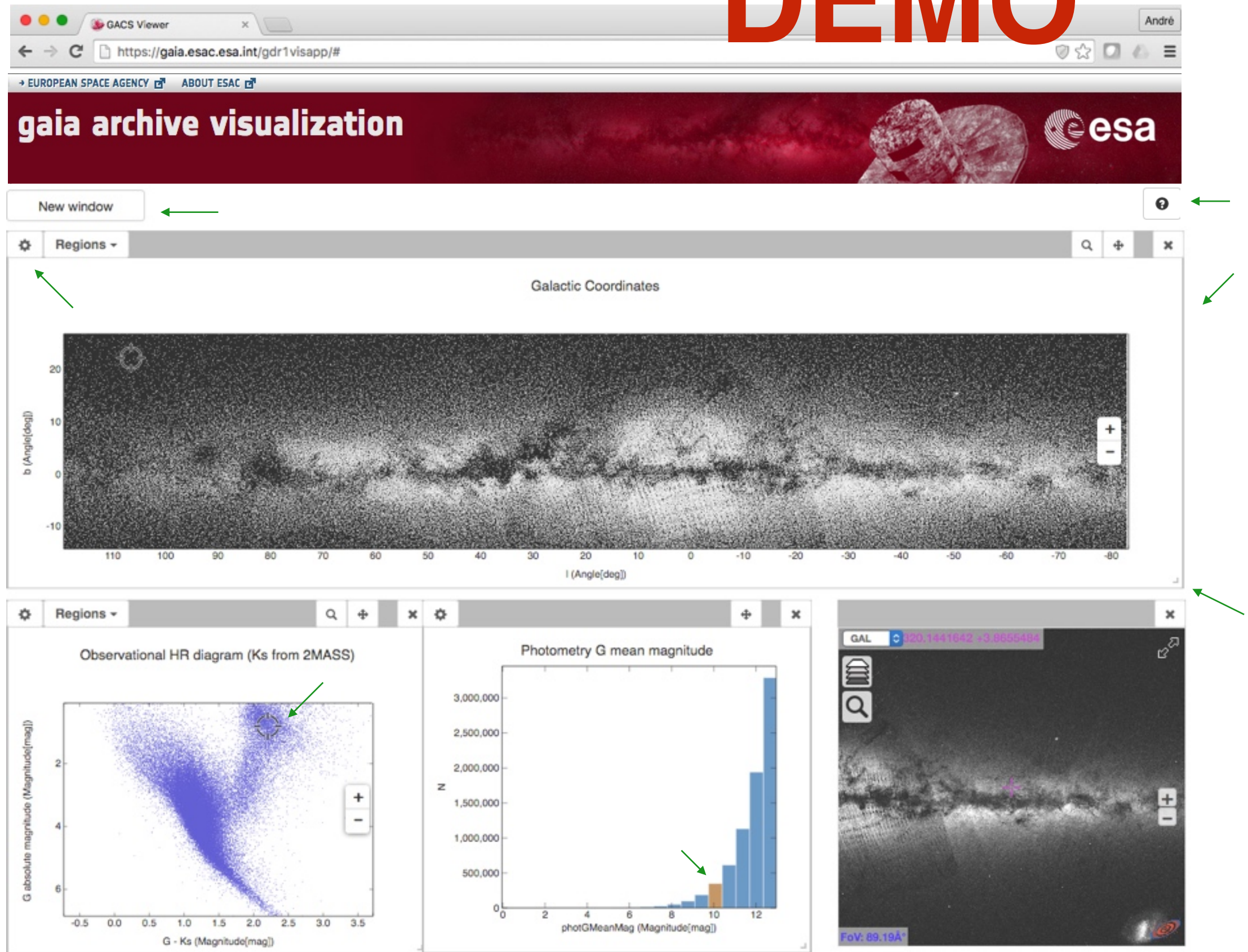


[High level. Powerful. Big data not out of the box. Expensive]



# Gaia interactive visualisation portal

# DEMO



# Gaia interactive visualisation portal -

## Deployment

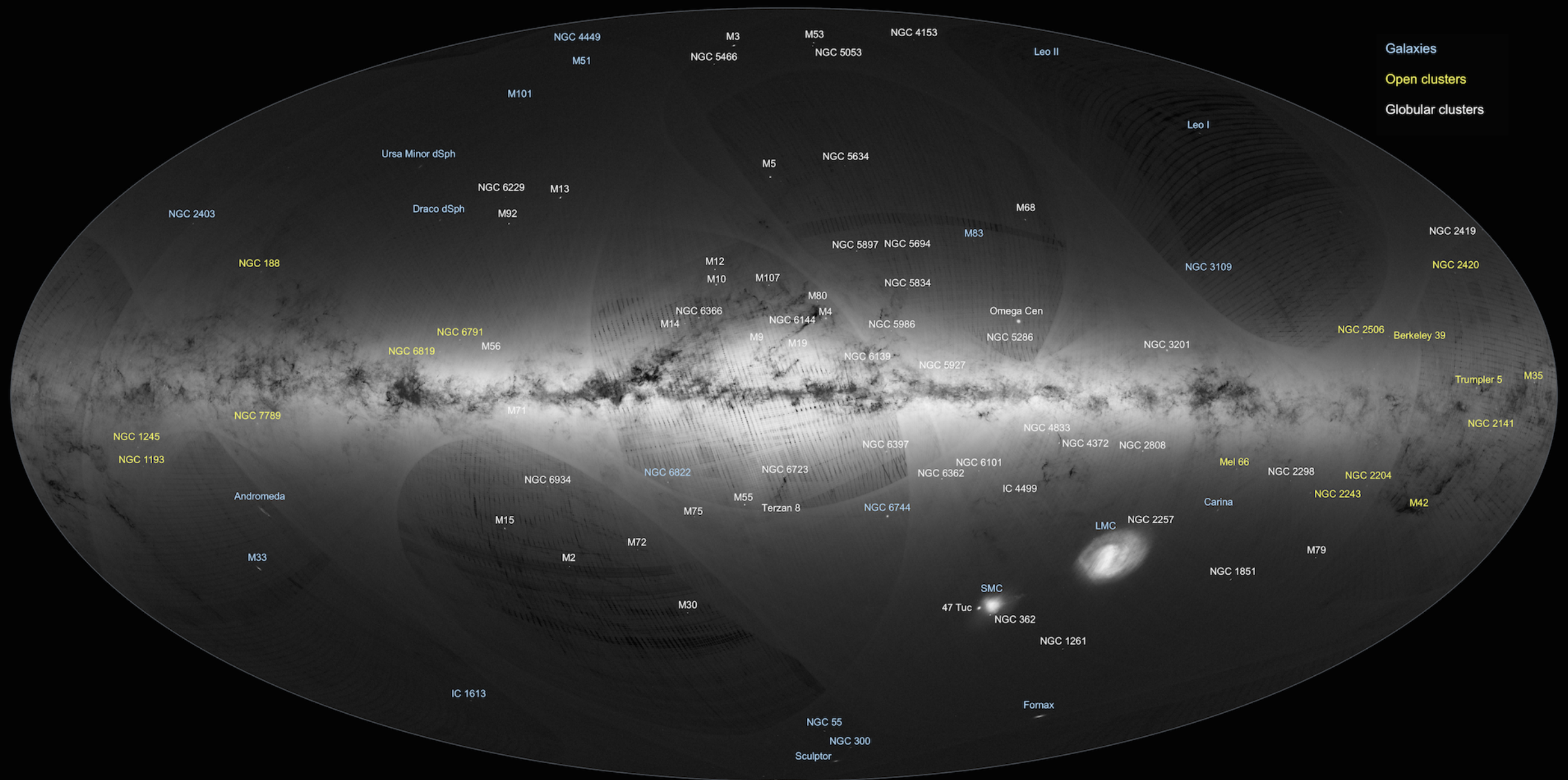
- CPU: Intel(R) Xeon(R) E5-2670 v3 @ 2.30GHz, 16 cores;
- memory: 64 gigabytes;
- storage: 3 TB SSD;
- application server: Apache Tomcat 8;
- Java version: 1.8.

## **Scalable:** (at 19:00 CEST, Sep 14, 2016 - **DR1 day**)

- Single accesses: 4286
- Accesses to help: 173
- Histograms: 145
- Scatter plots: 5650
- Scatter plot tiles: 1557153



# Data Release 1



September 14, 2016 ~1 100 000 000 objects

Race to Document Rare  
Plants Before These Cliffs  
Are Ground to Dust



TRILOBITES  
Long Before Making  
Enigmatic Earthworks,  
People Reshaped Brazil's...

SCIENCETAKE  
How a Little Bit of Hydra Regrows a  
Whole Animal

TRILOBITES  
What Makes a Woman a Good Dancer?  
Watch the Hips, a Study Says

PAID POST: MINH LONG CERAMICS  
What's Next for Vietnamese  
Cuisine?



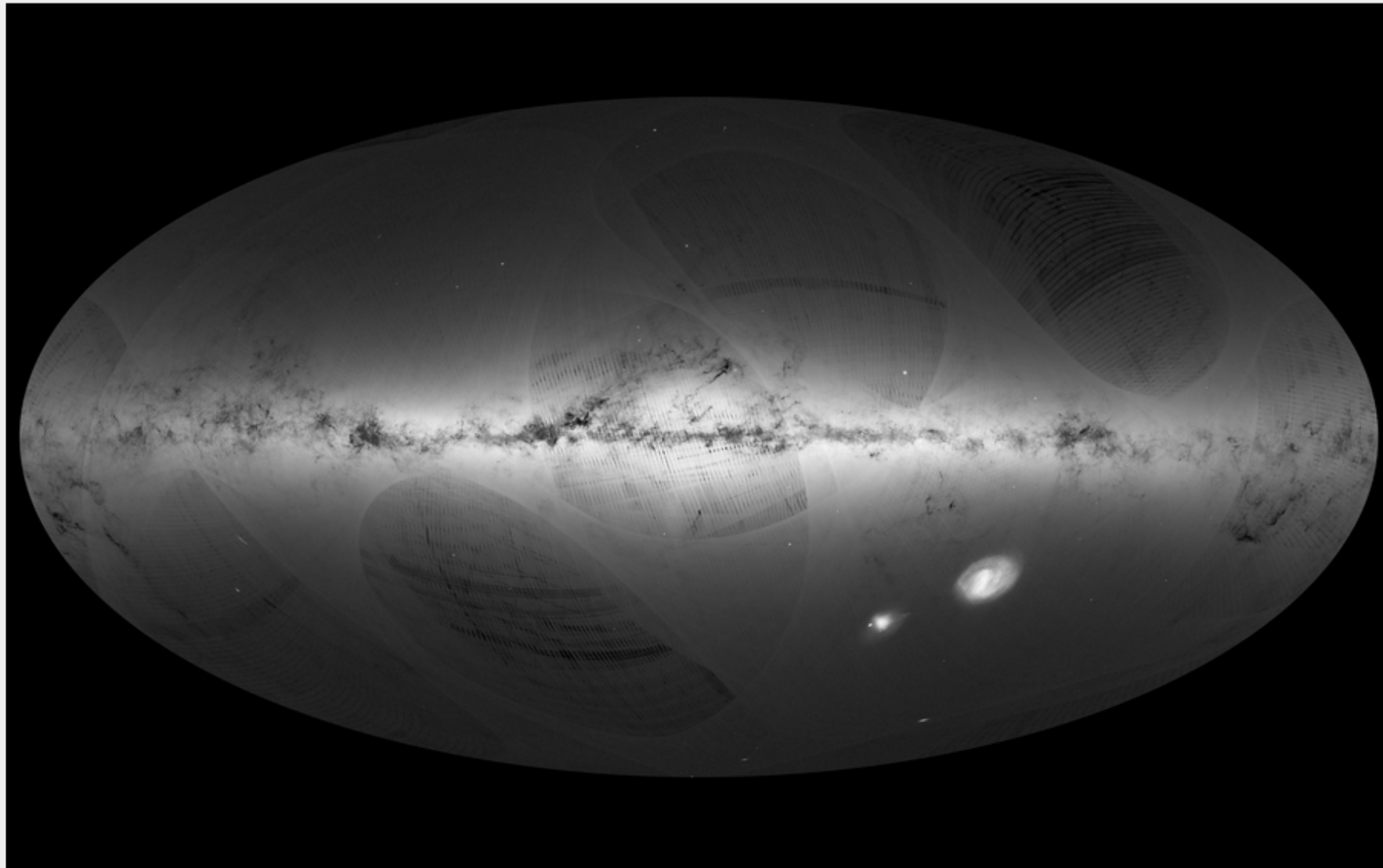
Sa  
St  
in

INNOVATION ▾

PERSPECTIVES ▾

PERFORMANCE ▾

Twitter Facebook LinkedIn Search



## 5 Coolest Things On Earth The Week

Sep 16, 2016 by Tomas Kellner







Com este tipo de exploração mais detalhada poder-se-á abordar as questões que levaram à construção da missão Gaia: Como funciona a nossa galáxia, como se formou e como está a evoluir?

A PARTIR DE  
5. FEIRA 20/11

**XXL 100 XXL**



**FORMIL®**  
Detergente Máquina Lavar  
Roupa 100 Doses

- Emb. 6 kg
- 1 kg = 1,25

**100 doses**

**9,99**

**100**

**FRESCURA AZUL**



**Floralys®** kitchen towels  
with Tuscany design

8 x 64 cm

**8 Rolos**

**3,49**

**Floralys®**  
Papel de Cozinha com Desenhos  
• Emb. 8 rolos

**Cada rolo APENAS 0,44**

**Cleffekt®**  
Lava Louça

- Nas variedades: limão ou maçã
- Cada emb. 1,5 L
- 0,5 L Extra
- 1 L = 0,40

**0,5 L EXTRA**

**0,79**

**0,79**

**0,79**

**Formate Econômico**



**Sonasol**

**Emb. 240 unid.**

**-50%**

Para os preços de mercado

Preço de mercado: 5,37

Preço LIDL: 2,68

Preço unitário: 0,011

**frotto**

100% celulose

**40 unid. GRÁTIS**

**0,90**

**FROTTO®**  
Guardanapo Folha Simples

- Emb. 200 unid.
- 40 unid. Grátis

**SONASOL®**  
Magic

- Nas fragrâncias: rosas e jasmim ou camomila e flor de baunilha
- Emb. 2,8 L
- 1 L = 0,96

**1 LITRO FORMATO ECONÔMICO**

**LISTERINE®**  
Elixir Dentes & Gengivas

- Emb. 1 L

**1 L**

**8,99**

**8,99**

PRINCÍPIO VÁLIDO DE 30/11 ÀS 20H, SALVO POR MOTIVO DE REATURA DE STOCKS OU ERRO TIPOGRÁFICO. ADESIONES NÃO INCLUEM. JOTOS DE SUCESSO DE APRESENTAÇÃO. PROMOÇÕES NÂO ACUMULÁVEIS ENTRE SI.

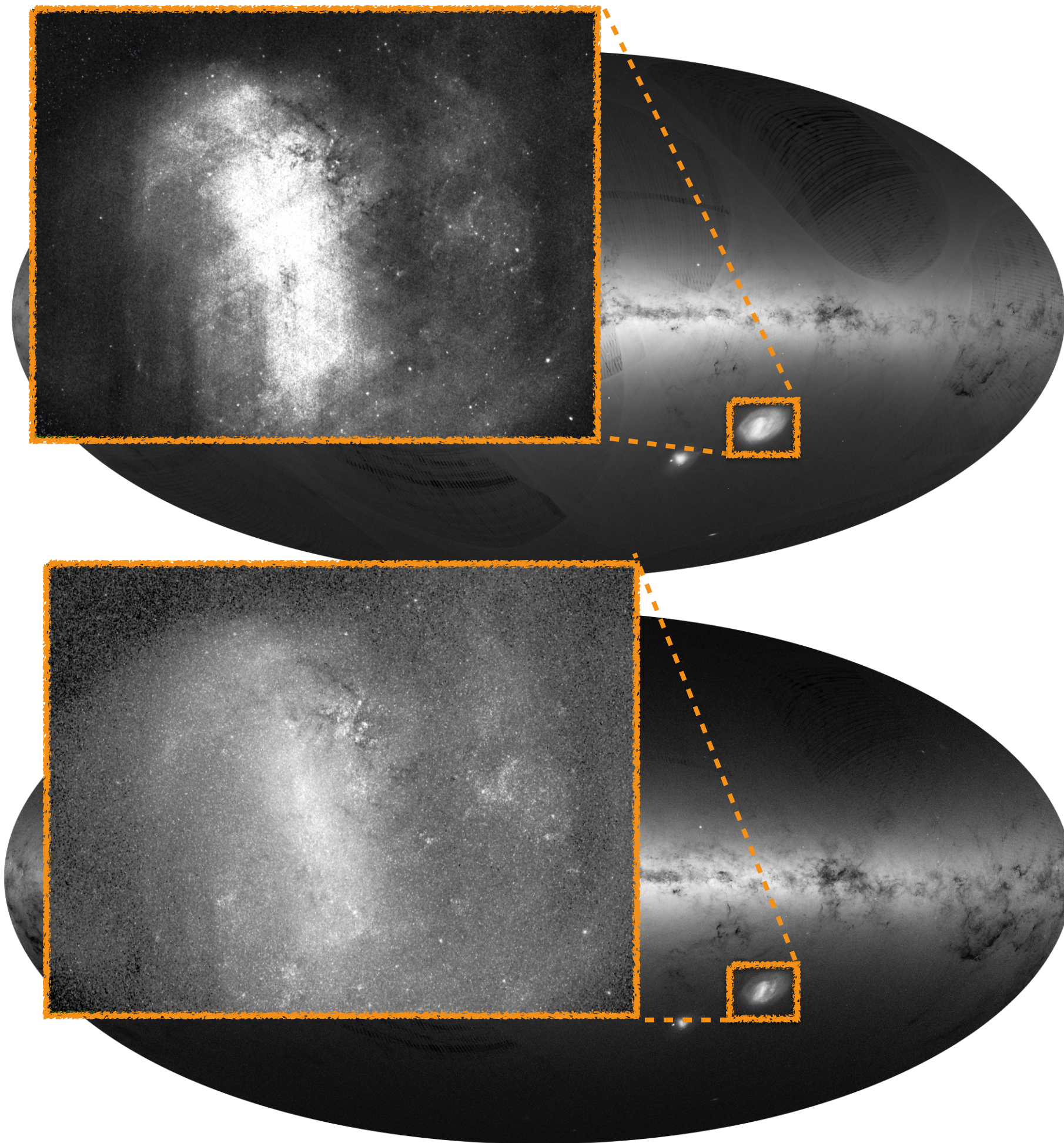
43/2019





more than a pretty face





Gaia source  
density and  
luminous flux  
representations:

**complementary**  
views or **stories**

more stories out  
there

Part of making  
the richness of  
the archive  
**intelligible**



# Data Release 2

**Liberate the data!!**

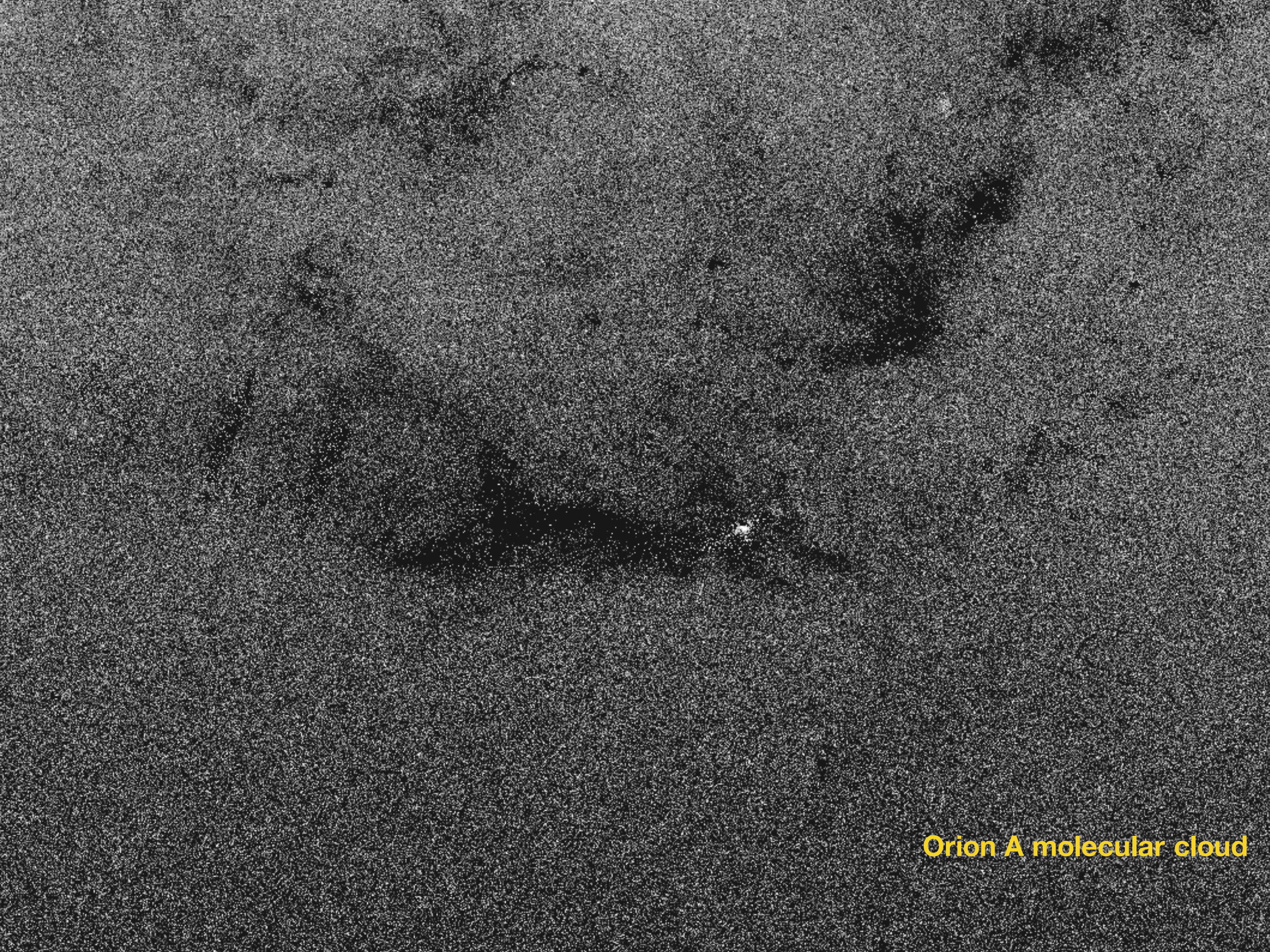
**Explore with  
GAVS**

April 2018

M	T	W	T	F	S	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

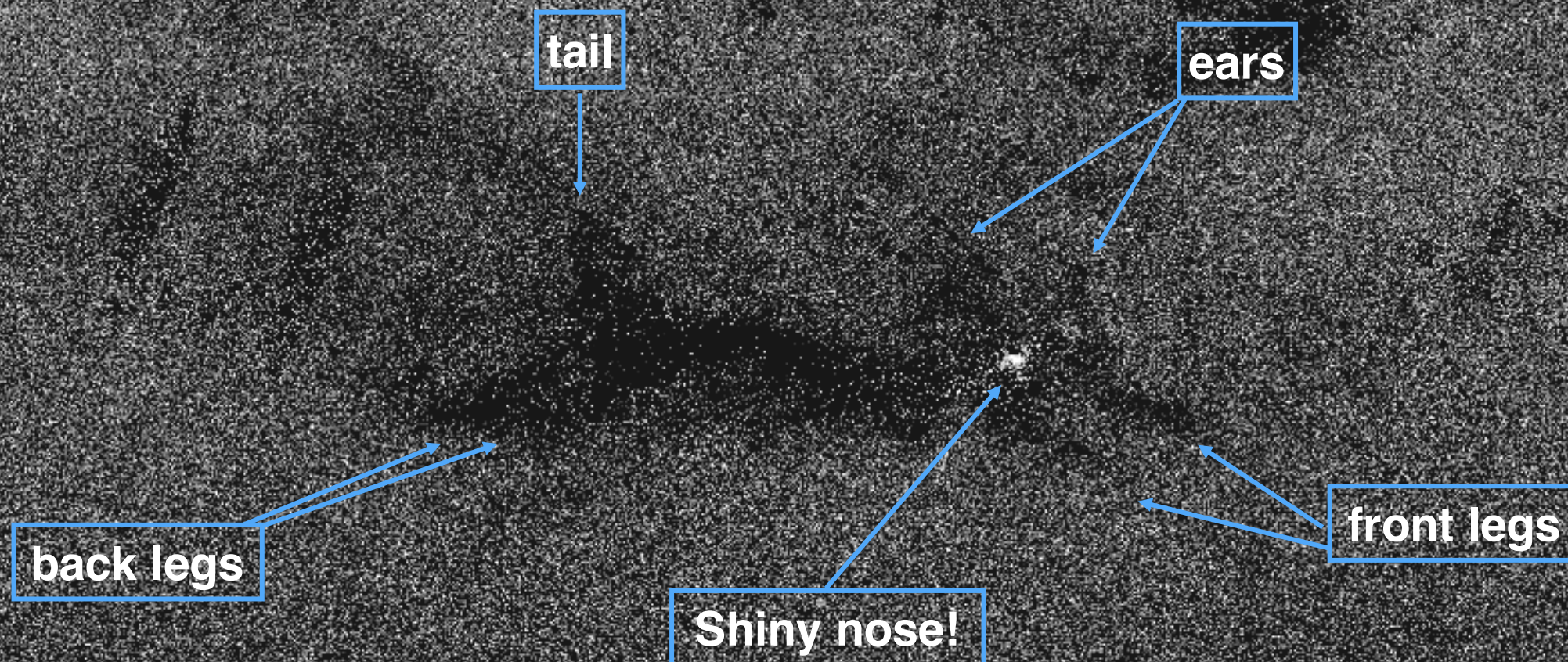
~1 700 000 000 objects





Orion A molecular cloud





**Orion A molecular cloud**



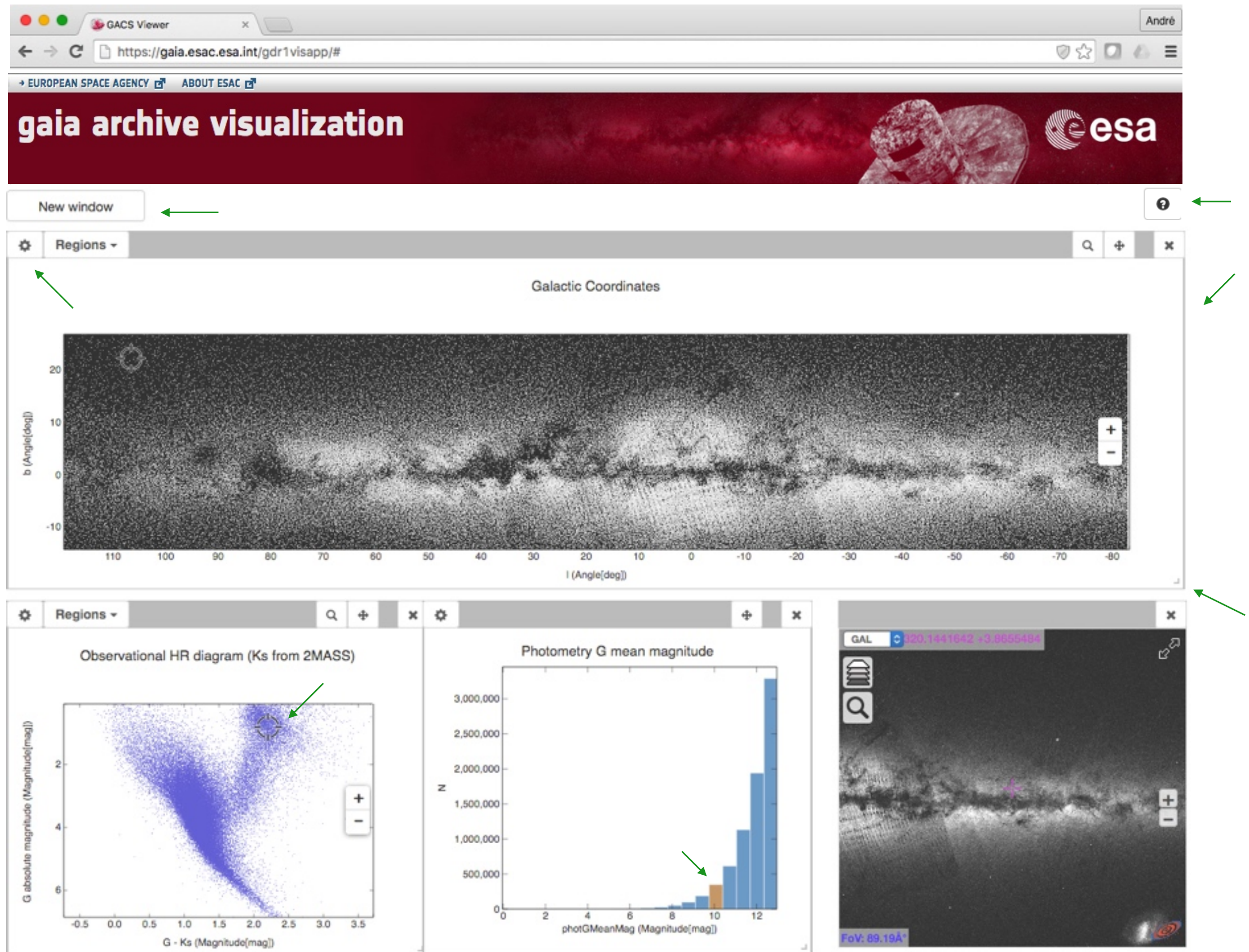
## **During this presentation**

- about 1 million stars were measured by Gaia,
- roughly 10 million astrometric measurements were taken,
- about 300,000 spectra were taken for 100,000 stars

Demo film backup plan

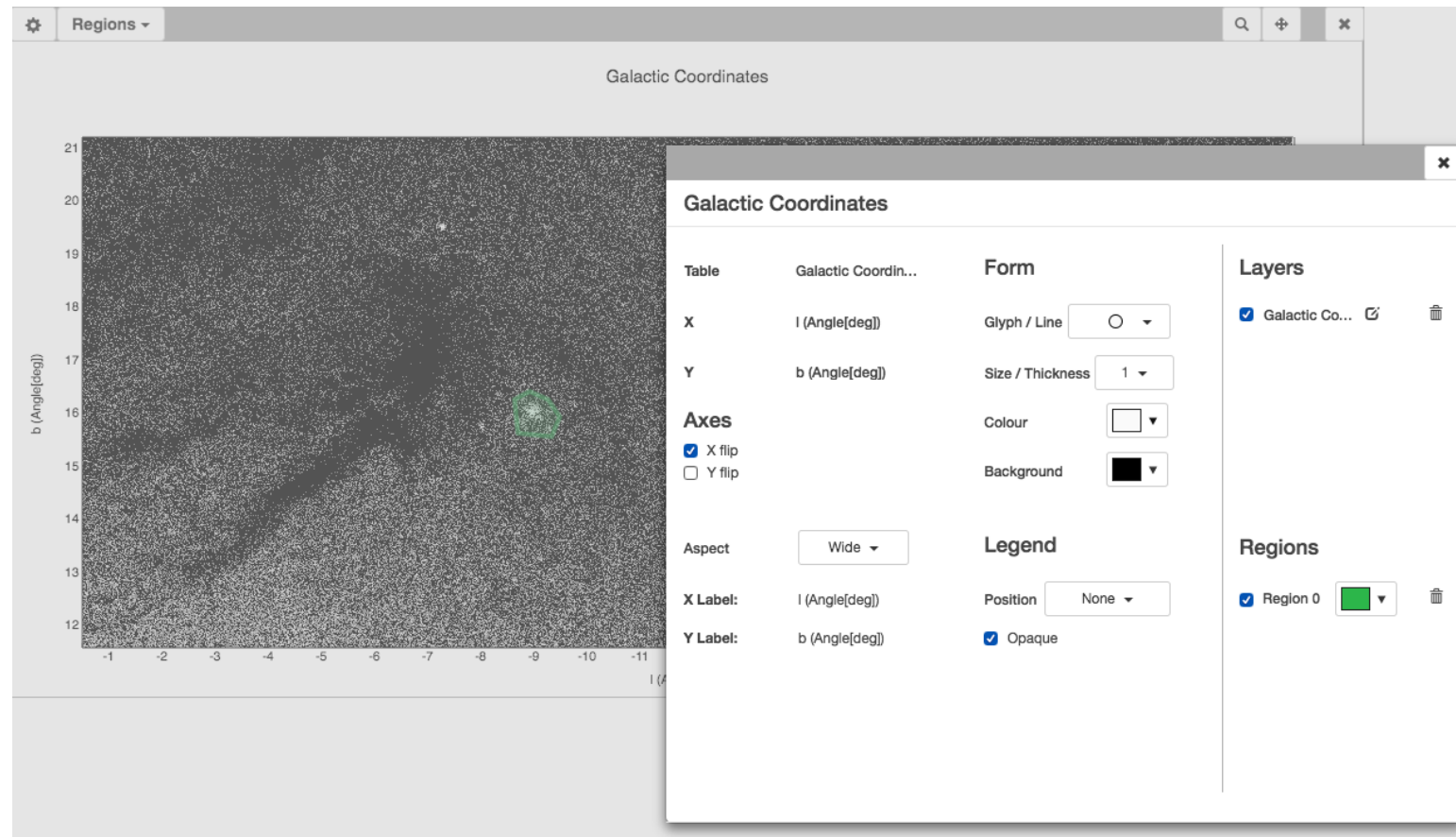


# Gaia interactive visualisation portal



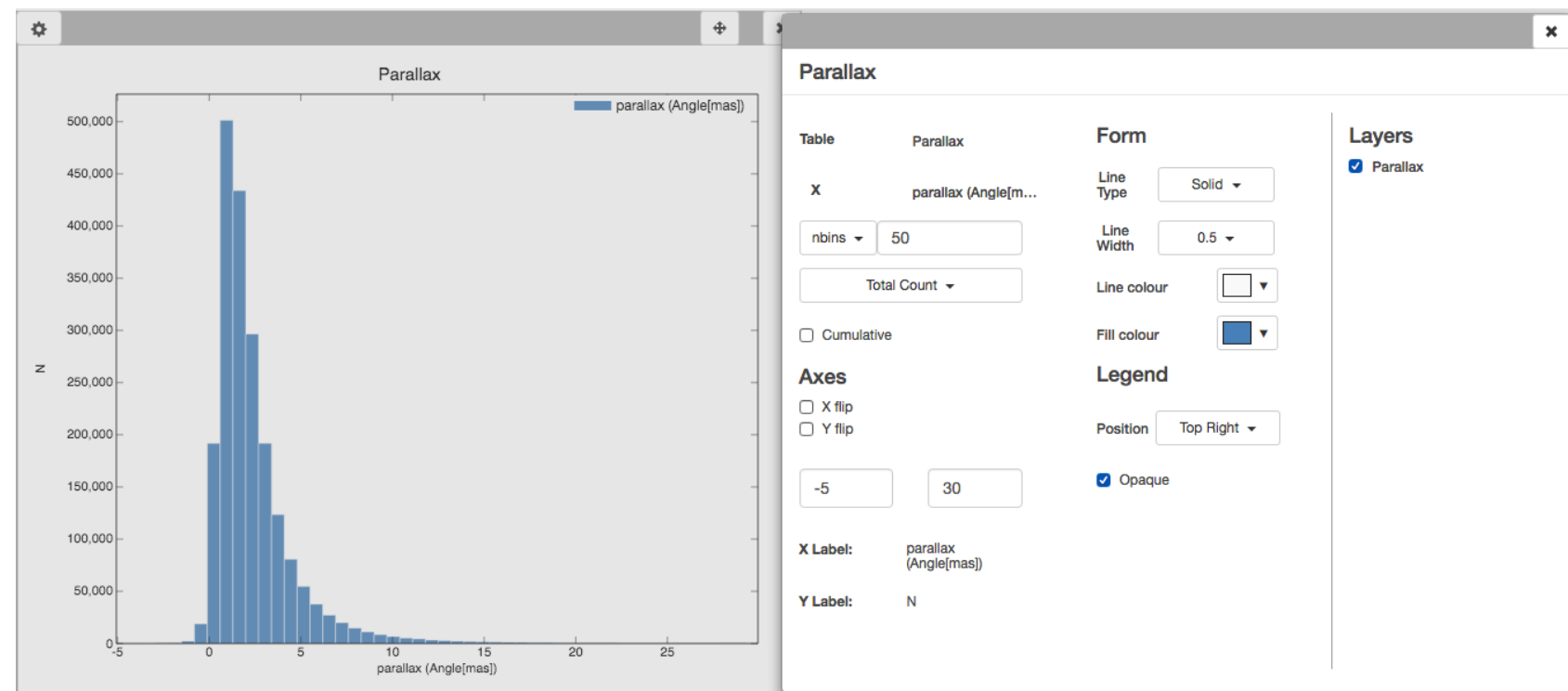
<http://gea.esac.esa.int/visualization/index.html>

# Gaia interactive visualisation portal



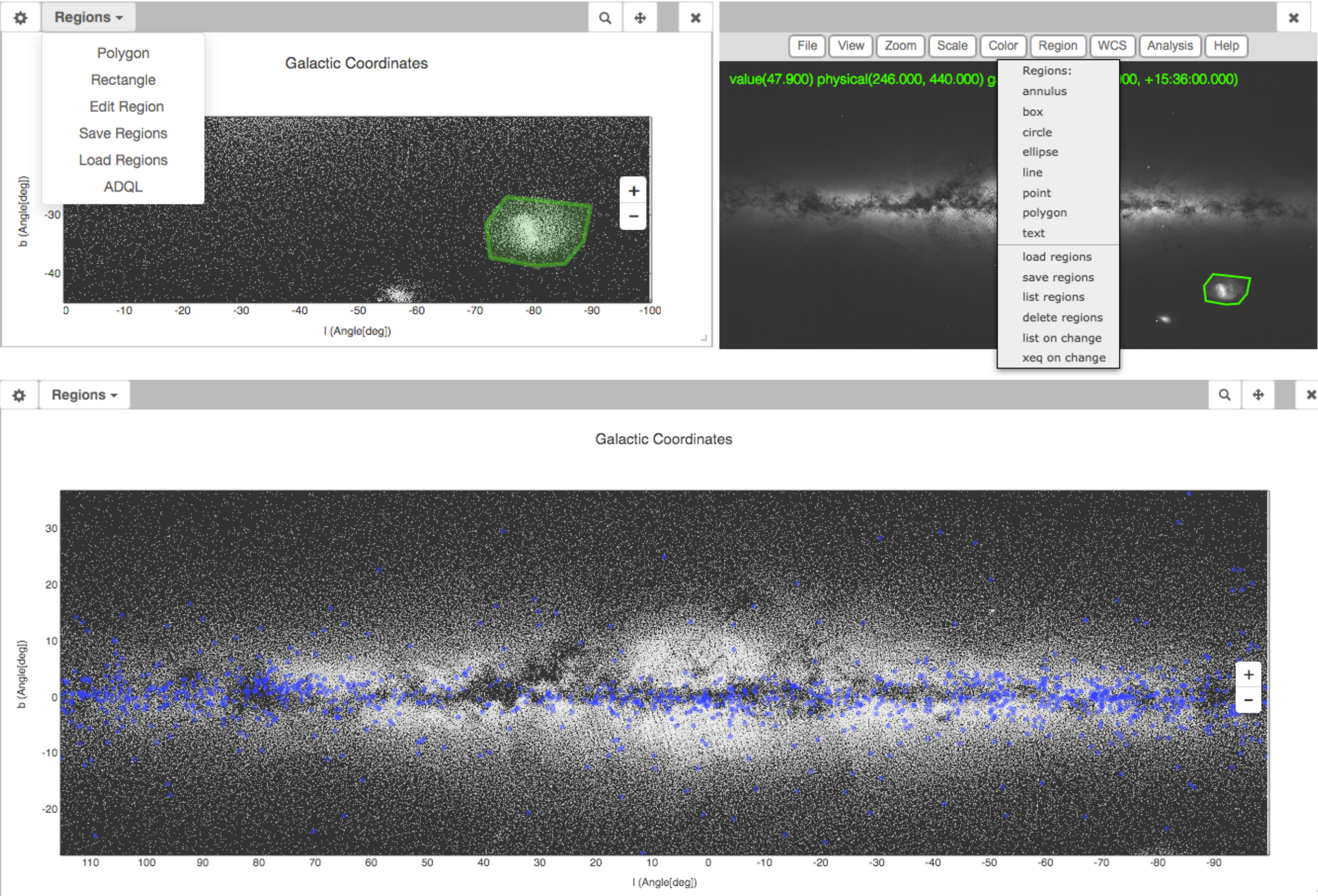
Configuration GUI should be

- intuitive
- minimal
- powerful



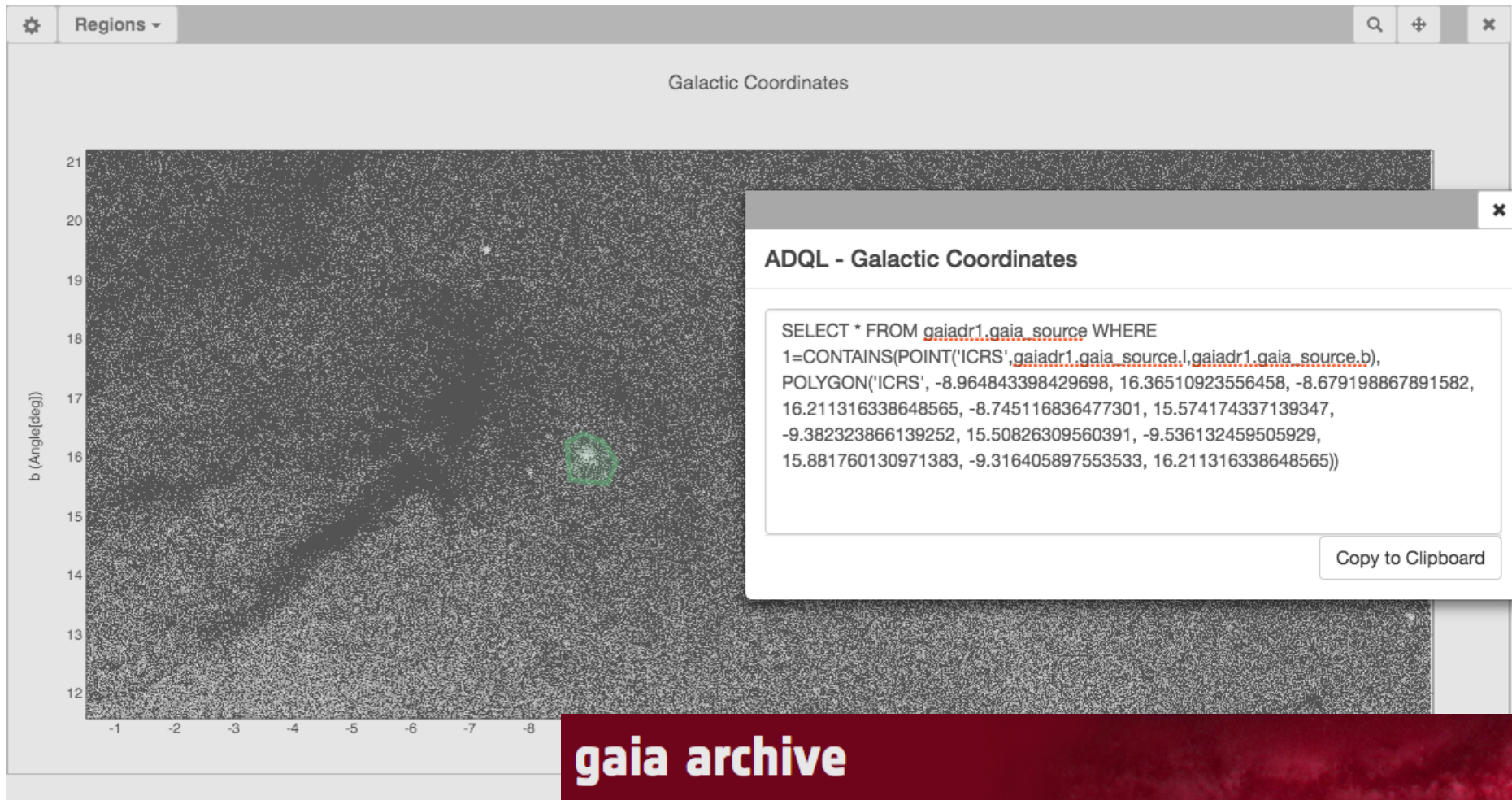


# Gaia interactive visualisation portal - Regions





# Gaia interactive visualisation portal - integrated archive service



Simple ADQL  
**visual queries**

The screenshot shows the Gaia archive search interface. The top navigation bar includes links for HOME, SEARCH, STATISTICS, VISUALIZATION, HELP, and DOCUMENTATION. The main section is titled 'gaia archive' and features a red background with the ESA logo. Below the navigation bar, there are tabs for 'Simple Form', 'ADQL Form', and 'Query Results'. The 'ADQL Form' tab is active, showing a 'Job name' field and a text area for the ADQL query. The query is the same as the one shown in the previous screenshot. Below the query text area are buttons for 'Reset Form' and 'Submit Query'. The 'Query Results' tab is also visible, showing a table with the following columns: Status, Job, Creation date, Num. rows, and Size. The table contains one row of results for job 14780657160940, created on 02-Nov-2016 at 05:48:36, with 50419 rows and a size of 9 MB. The bottom of the page includes a footer with the text 'COPYRIGHT 2000 - 2016 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.' and a version number '(v1.1.0)'.

**gaia archive**

HOME SEARCH STATISTICS VISUALIZATION HELP DOCUMENTATION

Simple Form **ADQL Form** Query Results

Job name:

Query examples

```
1 SELECT * FROM gaiadr1.gaia_source WHERE 1=CONTAINS(POINT('ICRS',gaiadr1.gaia_source.l,gaiadr1.gaia_source.b),
POLYGON('ICRS', -8.964843398429698, 16.36510923556458, -8.679198867891582, 16.211316338648565, -8.745116836477301,
15.574174337139347, -9.382323866139252, 15.50826309560391, -9.536132459505929, 15.881760130971383,
-9.316405897553533, 16.211316338648565))
```

Reset Form Submit Query

Status	Job	Creation date	Num. rows	Size
✓	14780657160940	02-Nov-2016, 05:48:36	50419	9 MB

1-1 of 1

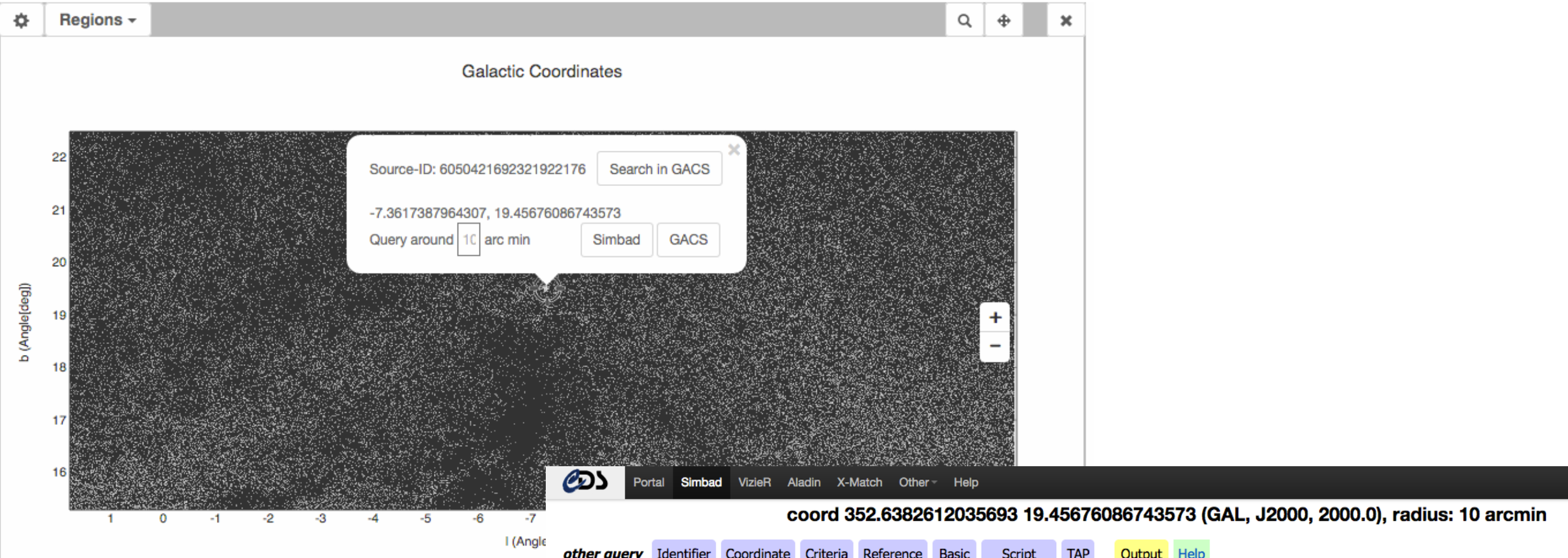
Apply jobs filter Filter this session ☒ Select all jobs ☐ Delete selected jobs

COPYRIGHT 2000 - 2016 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED. (v1.1.0)



# Gaia interactive visualisation portal

- integration with Gaia archive
- CDS services: simbad, sesame name resolver



Query : coord 352.6382612035693 19.45676086743573 (GAL, J2000, 2000.0), radius: 10 arcmin C.D.S. - SIMBAD4 re

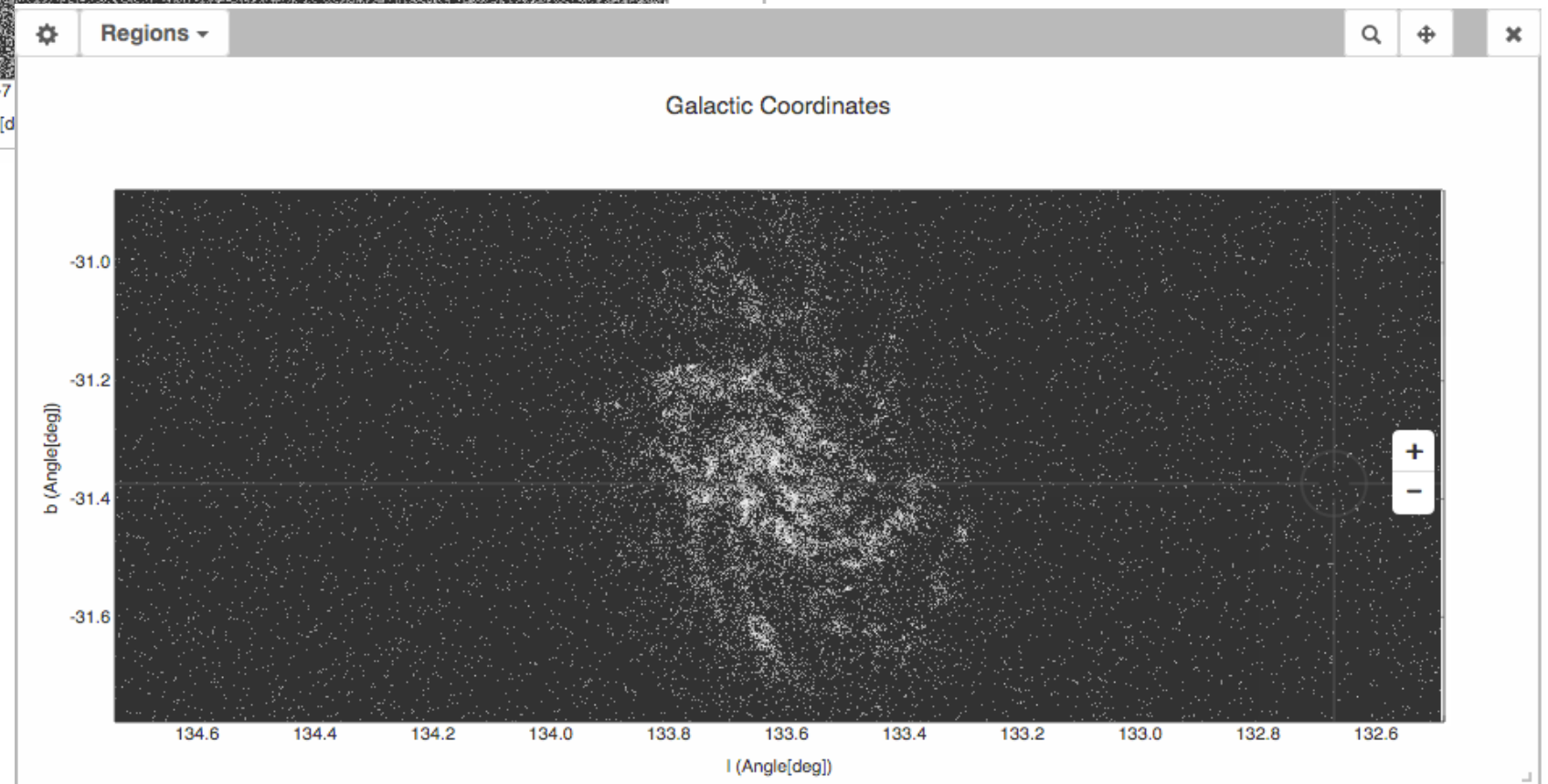
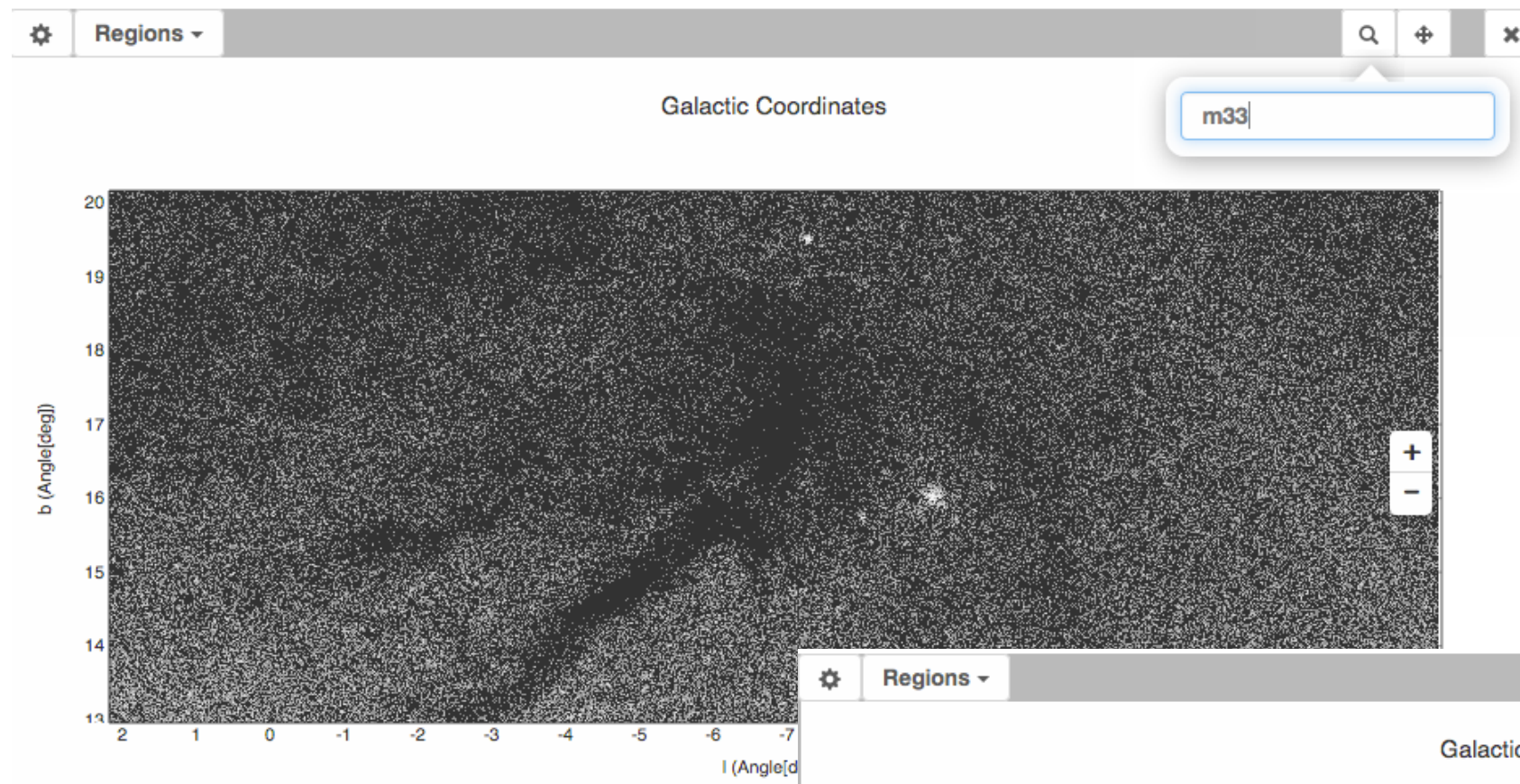
Number of rows : 152

Show 100 entries

N <sup>o</sup>	Identifier	dist(asec)	Otype	ICRS (J2000) RA	ICRS (J2000) DEC	Mag U	Mag B	Mag V	Mag R	Mag I	Spectral type
1	[MMP2009] M80 13787	29.35	HB*	16 16 59.111	-22 59 53.45	18.056		18.261			~
2	[MMP2009] M80 14201	39.54	HB*	16 16 57.762	-22 59 37.14	17.727		17.898			~
3	CXOU J161659.8-225931	53.79	X	16 16 59.88	-22 59 31.1						~
4	CXOU J161655.5-225925	59.56	X	16 16 55.58	-22 59 25.7						~
5	CXOU J161702.0-230033	60.76	X	16 17 02.05	-23 00 33.0						~

# Gaia interactive visualisation portal

- integration with Gaia archive
- CDS services: sesame name resolver - and vice versa!





# Gaia interactive visualisation portal

integration with external applications - DS9/JS9 and Aladin

- provide HiPS and fits maps
- regions
- panel with web versions in visualisation portal

