

# MODERN DATA SCIENCE WITH **VAEX**: A NEW APPROACH TO DATAFRAMES AND PIPELINES

**JOVAN VELJANOSKI**

Sr. Data Scientist @ XebiaLabs



# UNCOMFORTABLY LARGE DATA

---

**Working with %i samples without going to the cloud:**

- ➔  $< 1\_000\_000$  samples
- ➔  $\sim 10\_000\_000$  samples
- ➔  $\sim 100\_000\_000$  samples
- ➔  $\sim 1\_000\_000\_000$  samples
- ➔ larger datasets



# VAEX.IO: WHO ARE WE?



## Jovan Veljanoski

Former astrophysicist

Sr. Data Scientist @ XebiaLabs

Co-founder of [vaex.io](https://vaex.io)

✉ [jovan.veljanoski@gmail.com](mailto:jovan.veljanoski@gmail.com)

🌐 <https://www.linkedin.com/in/jovanvel/>



## Maarten Breddels

Former astrophysicist

Freelancer / consultant / data scientist

Core Jupyter-Widgets developer

Founder of [vaex.io](https://vaex.io)

Principal author of [vaex](https://vaex.io)

✉ [maartenbreddels@gmail.com](mailto:maartenbreddels@gmail.com)

🌐 [www.maartenbreddels.com](http://www.maartenbreddels.com)

🐦 [@maartenbreddels](https://twitter.com/maartenbreddels)

🐙 [github.com/maartenbreddels](https://github.com/maartenbreddels)



## Yonatan Alexander

Head of Data Science at BuiltOn

✉ [jonathan@xdss.io](mailto:jonathan@xdss.io)

🌐 <https://www.linkedin.com/in/xdssio/>



## Mario Buikhuizen

Freelancer / consultant

Front-end / dashboards / widgets specialist

✉ [mbuikhuizen@gmail.com](mailto:mbuikhuizen@gmail.com)



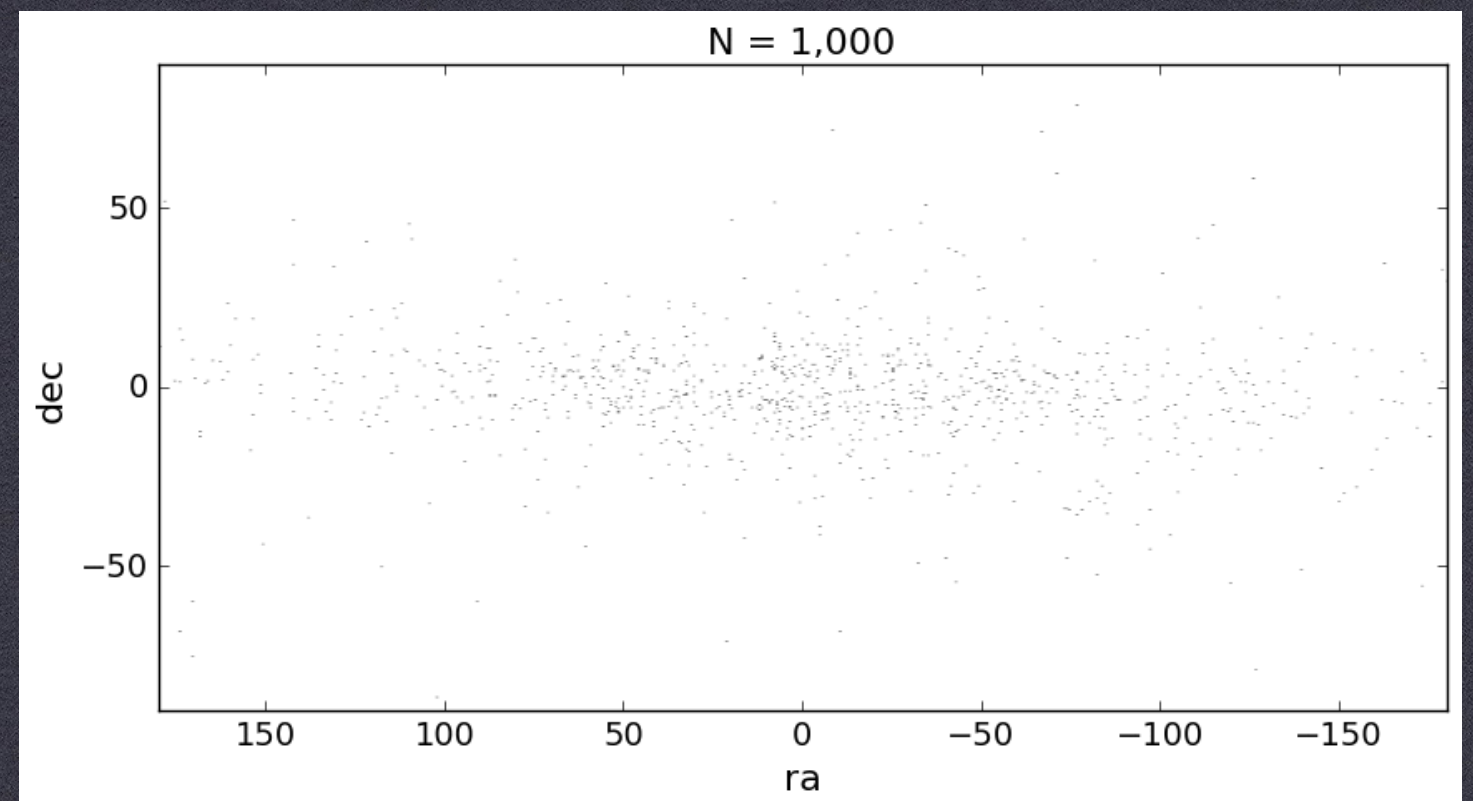
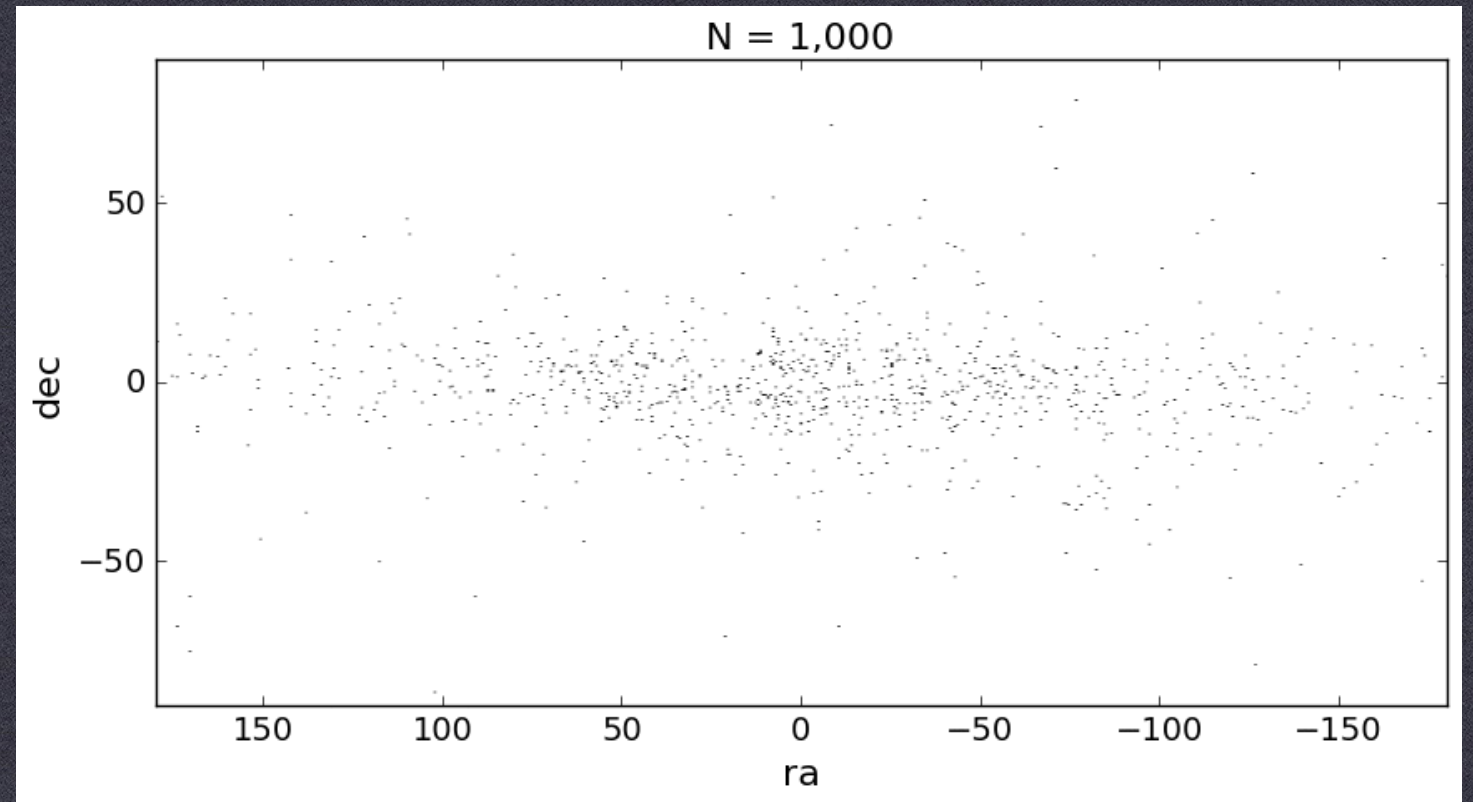
# THE NEED FOR VAEX



The Gaia satellite:

More than 1 billion observations of stars in our Galaxy!

How do we work (explore, filter, visualize, analyze) with such data?





# LIVE DEMO

---

The Jupyter notebooks presented at the live demo can be found at:

<https://github.com/vaexio/vaex-talks>