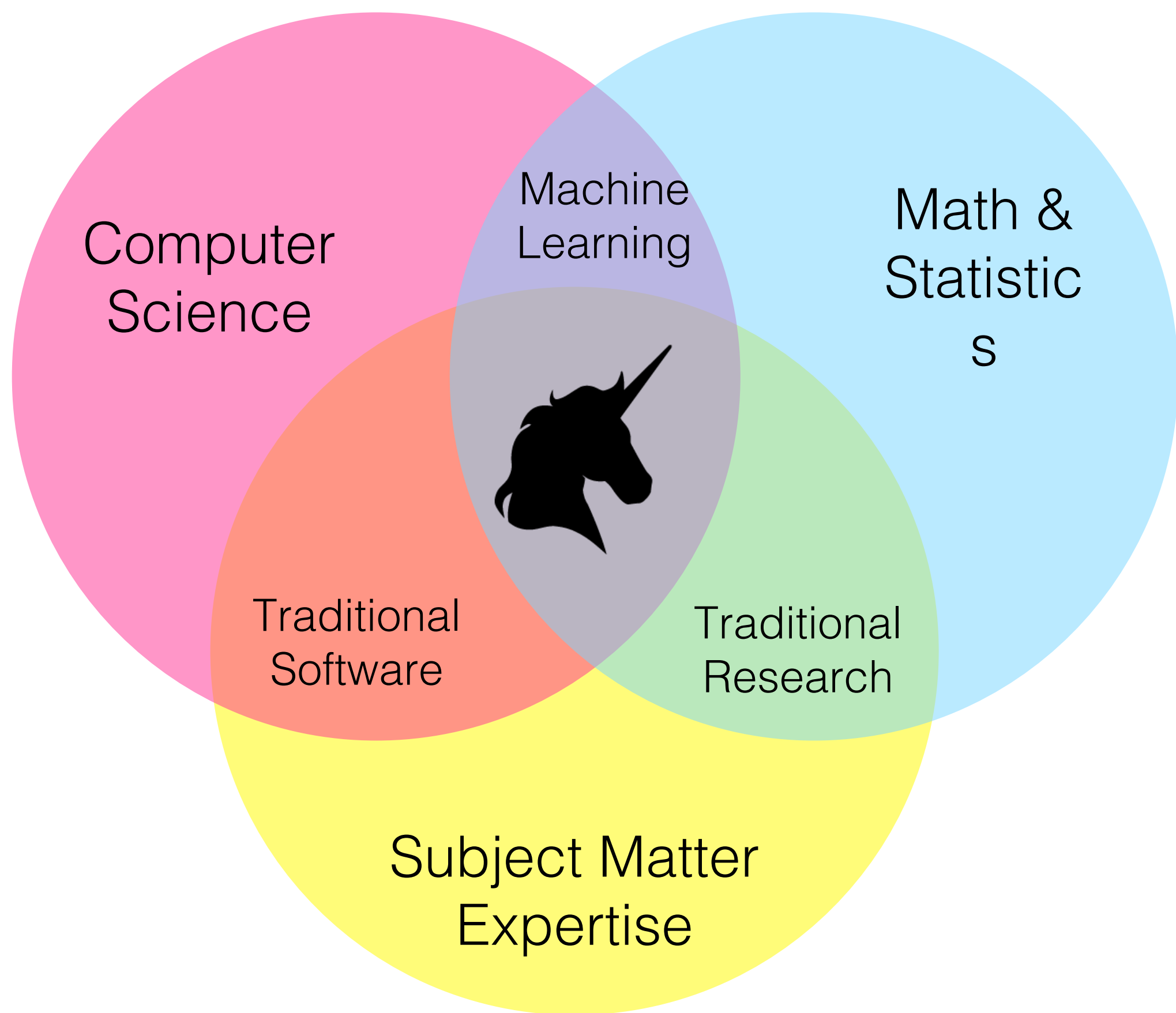




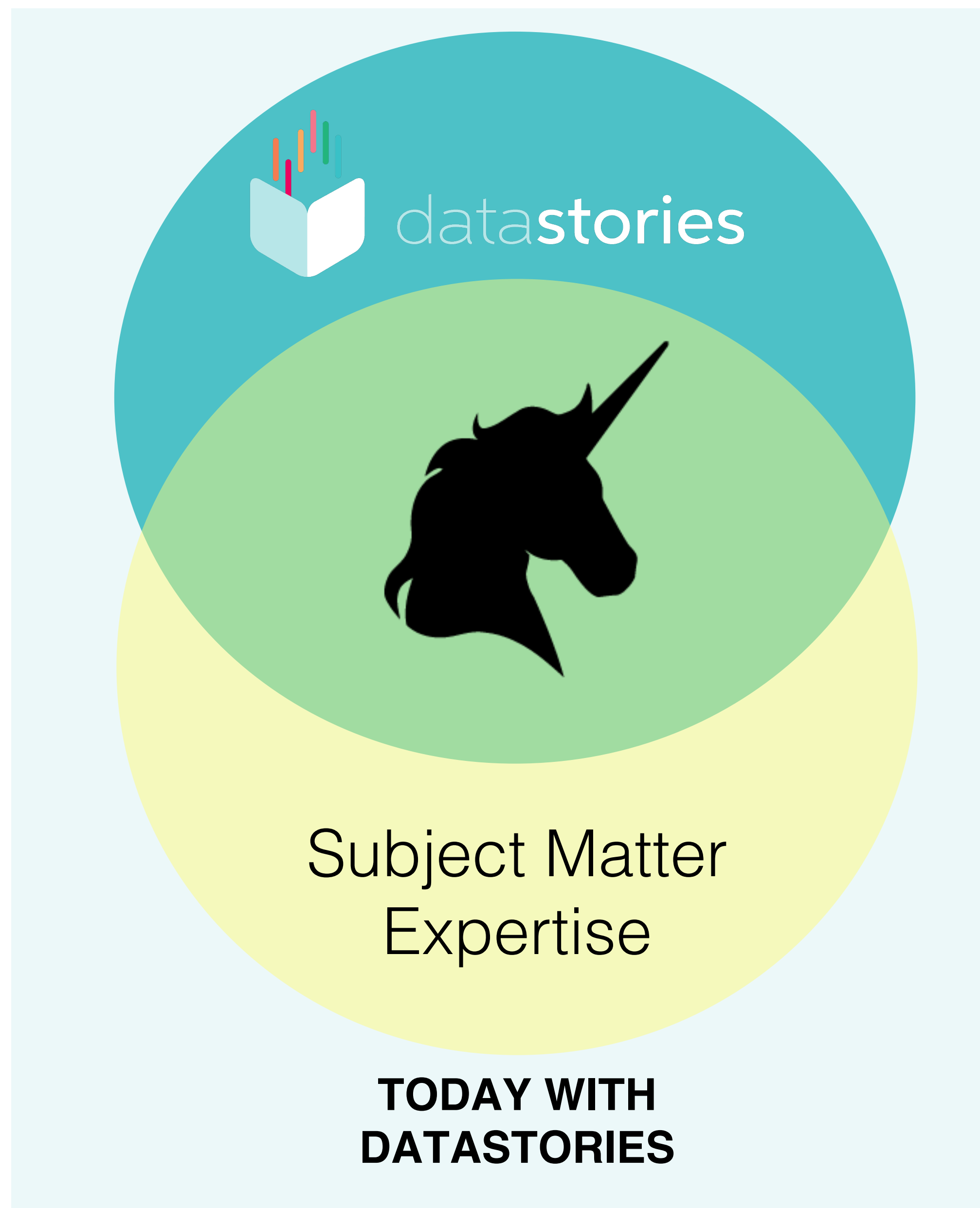
data**stories**

How AI / ML can even help non-data scientists to find the dominant SNPs over a coffee?

SYSTEMATIC INNOVATION ORGANIZATION



BEFORE



**TODAY WITH
DATASTORIES**

AMBITION: #1 in R&D

AI for domain experts...



datastories

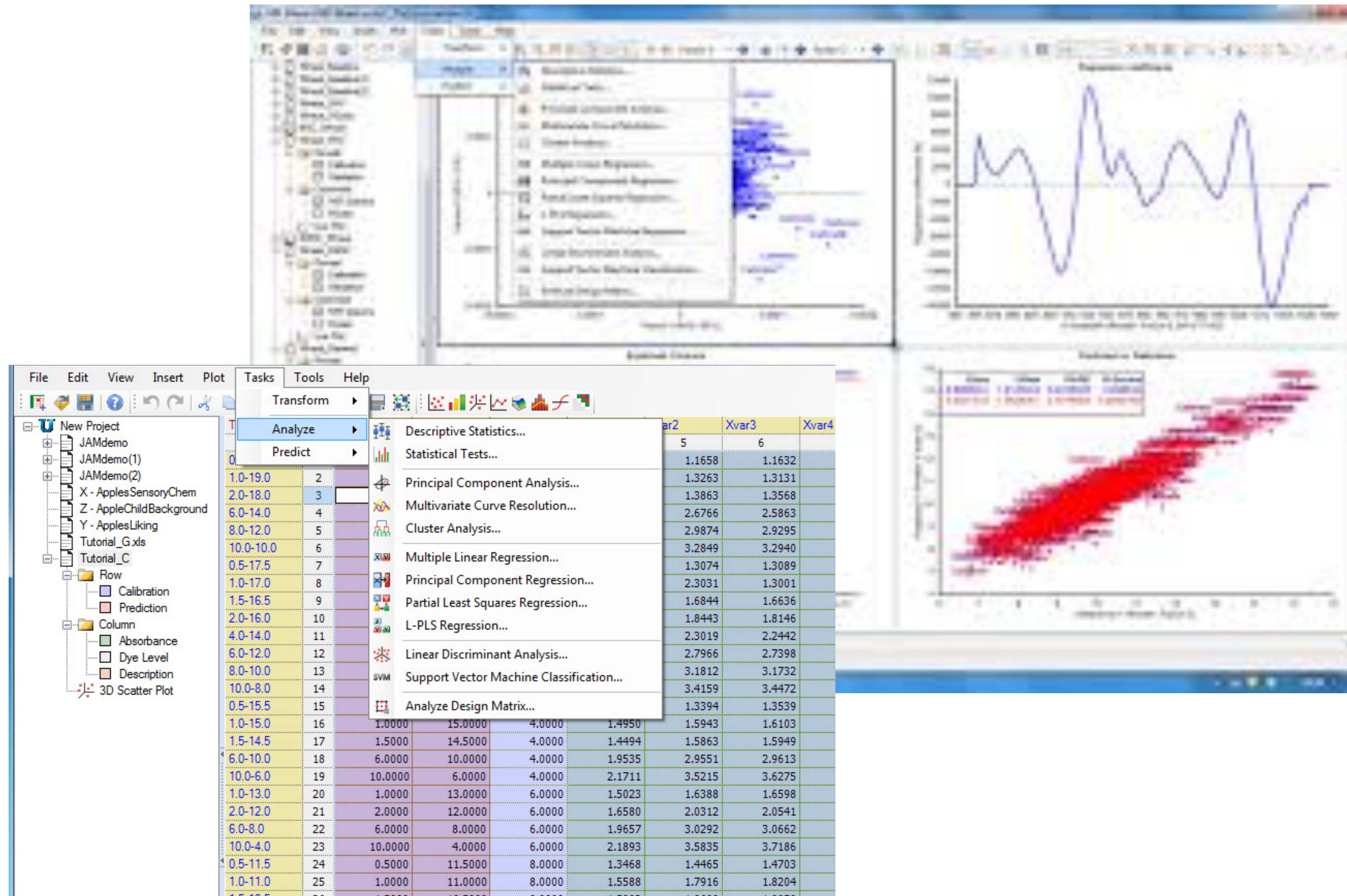


1.000 data scientists in-a-box

Four main challenges

- 1. How does everything interconnects with each other?**
- 2. What are the dominant variables?**
- 3. How can I achieve optimal targets?**
- 4. What are the exceptions to the rule?**

TRADITIONAL APPROACH



FINDING THE RIGHT SNPs

OVER A CUP OF



FINDING THE RIGHT SNPs

The screenshot shows the 'NEW DATASTORY' interface in the Datastories application. The user is logged in as 'ROEL'. The interface is divided into a left sidebar and a main content area. The sidebar contains navigation options: 'UPLOAD START', 'YOUR STORIES 19', 'DATA SOURCES', 'SETTINGS', 'CONTACT SUPPORT', and 'LOGOUT'. The main content area features a progress indicator at the top with four steps: 1 (selected), 2, 3, and a flag icon. Step 1, 'Upload your data and name your story', is active and contains the following fields:

- Your story's title**: A text input field.
- A short description...**: A text area for a brief description.
- The data source to analyse**: A dropdown menu with a 'See all data sources' link and a checkmark.
- Drop your data here or click to upload**: A dashed box for file upload.

A 'Next >' button is located at the bottom of the first step. Below the main content area, steps 2 and 3 are visible but not yet active.

SIMPLE

1. Name

2. Link Dataset

**3. Select Target
eg. phenotype**

RUN

RESULTS

**ACCURACY IMPROVEMENT WITH
20% TO 200%**

**WITH ONLY DOMINANT VARIABLES
> 4 TO 10 SNPs in different cases**



datastories

FINDING THE RIGHT QTL's

OVER A



CHALLENGE

- **+8000 QTLs**
- **ONLY 150 MEASUREMENTS**
- **RESULT OF WEEKS: 45% accuracy**

WHICH ARE THE DOMINANT QTLs?



RESULT AFTER LESS THAN 1 HOUR

The story behind your data.

The screenshot displays the datastories application interface. At the top, the title 'Plant Genomics: KPI' is shown. Below it is a navigation bar with 11 numbered slides: 1. Table of Contents, 2. Data Overview, 3. KPI, 4. Global Data Health, 5. Simple Correlations, 6. Pair-wise plots, 7. Linear vs. Non-linear relations, 8. Predictive Models, 9. What-ifs, 10. Model Accuracy, and 11. Conclusions. The 'Table of Contents' slide is active, showing a grid of 10 items with icons and brief descriptions:

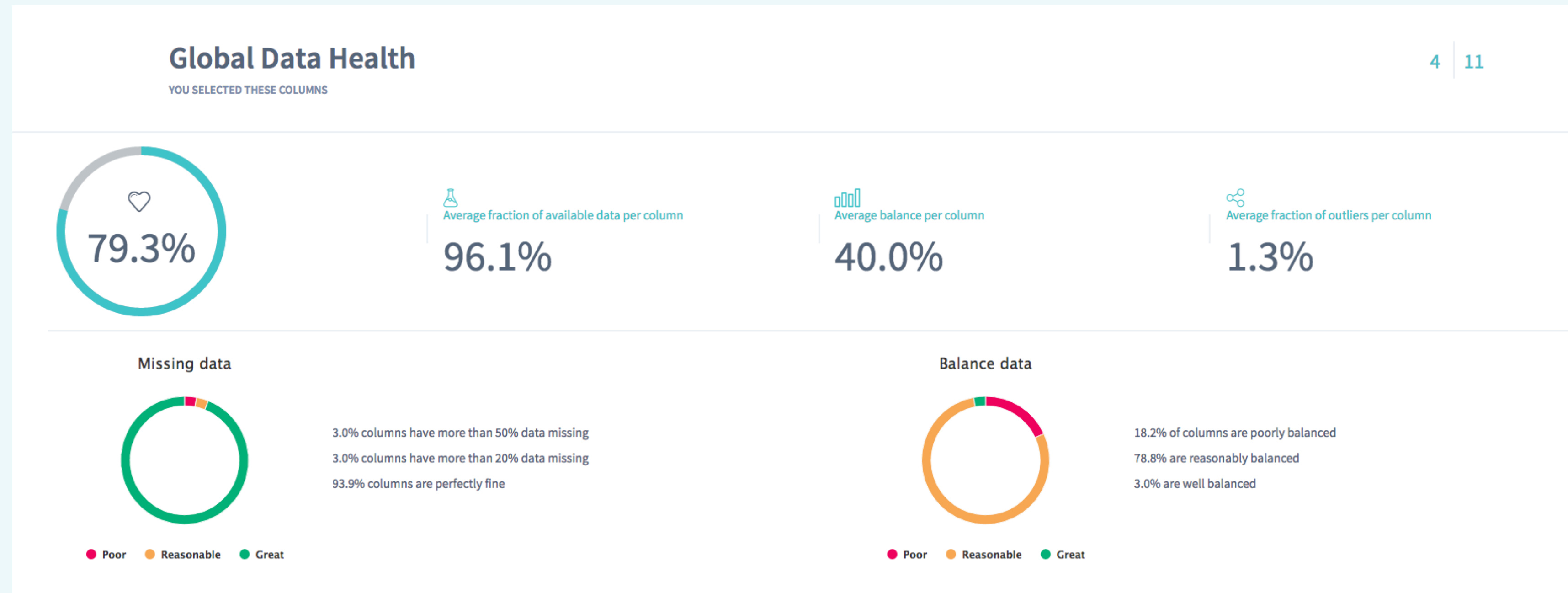
Data Overview Summary of what we got	KPI KPI graph and basic info	Global Data Health Table with selected data columns and basic stats
Simple Correlations Explore relationships between columns	Pair-wise plots Pair-wise data plots with KPI	Linear vs. Non-linear relations Compare pair-wise connections
Predictive Models Summary of modeling and KPI drivers	What-ifs Explore what if scenarios and optimal settings.	Model Accuracy
Conclusions		

The interface also includes a left sidebar with options like 'UPLOAD', 'YOUR STORIES', 'DATA SOURCES', 'SETTINGS', 'CONTACT SUPPORT', and 'LOGOUT'. A top navigation bar includes 'THUMBNAILS', 'PREV SLIDE', 'NEXT SLIDE', and 'EXPORT'.



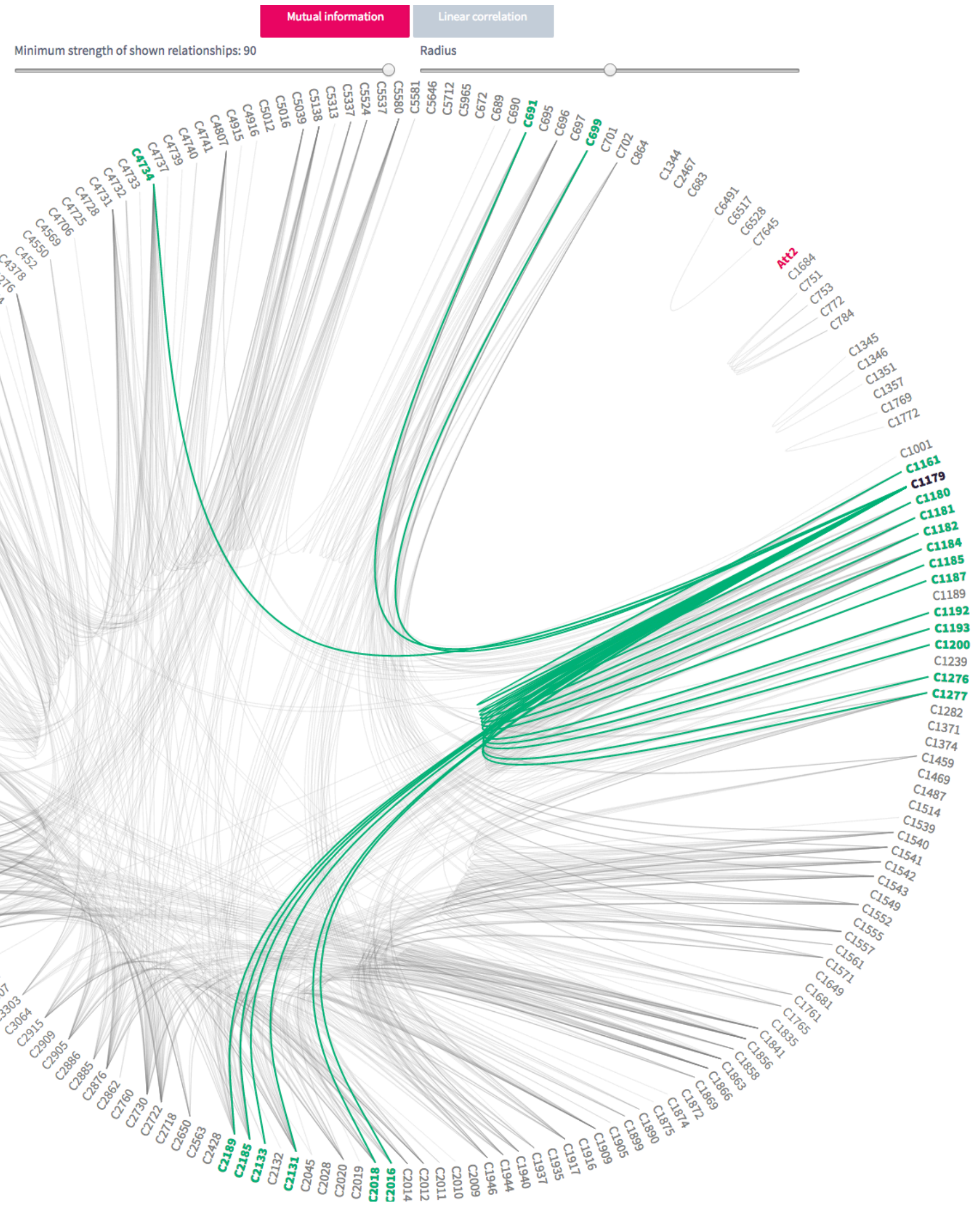
RESULT AFTER LESS THAN 1 HOUR

datastories



Robust & Automatic correction for: missing data, unbalance, binary, categorical, etc...

RESULT AFTER LESS THAN 1 HOUR



**Non-linear
correlation
linkage map**



RESULT AFTER LESS THAN 1 HOUR



PREDICTION ACCURACY

81.3%



NUMBER OF VARIABLES:

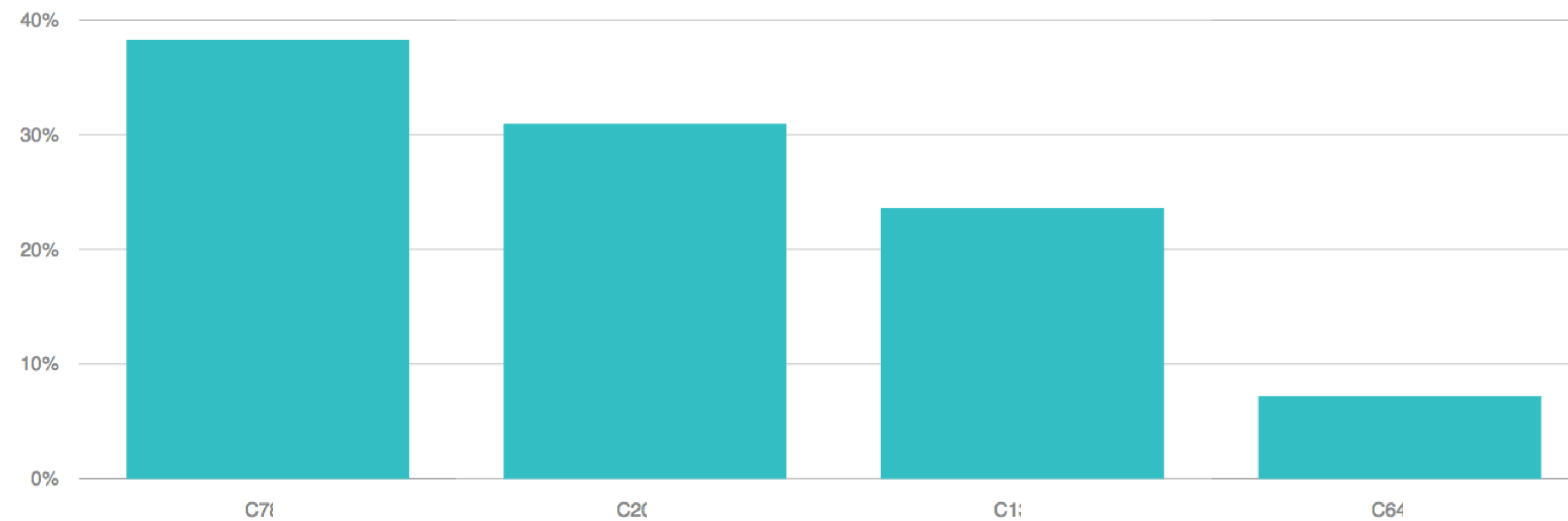
4



NUMBER OF VARIABLES WE STARTED WITH:

8076

Important drivers:



RESULT

AFTER LESS THAN 1 HOUR

- IDENTIFIED 4 DOMINANT QTLs
- ACCURACY IMPROVED FROM 45% TO 81%

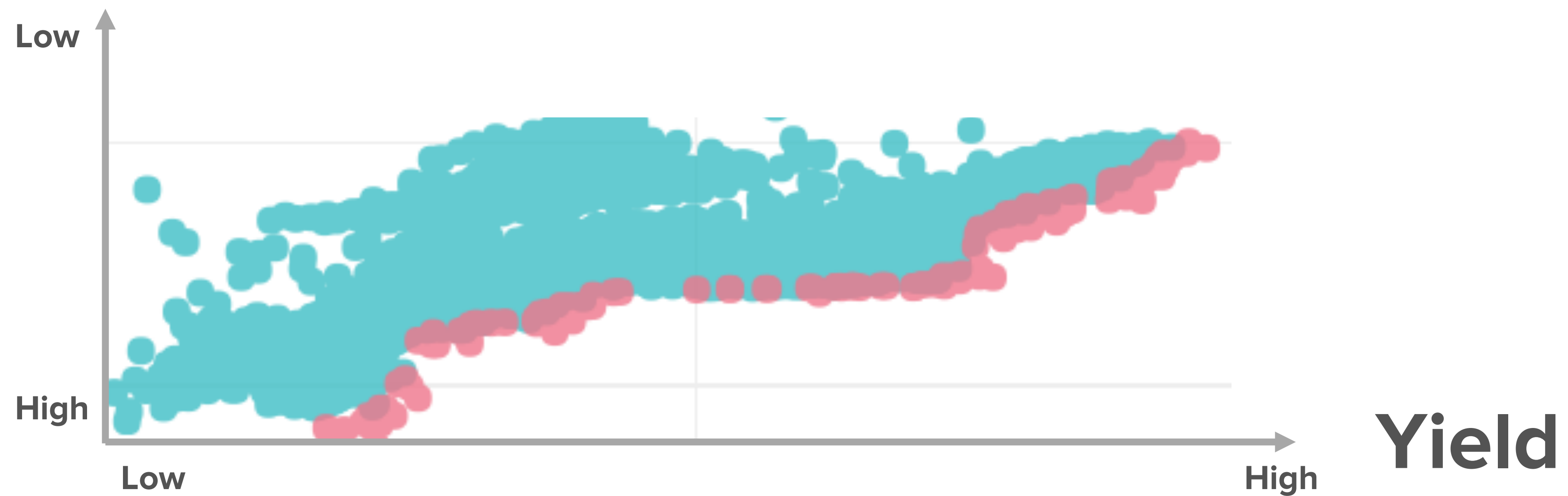


MULTI-OBJECTIVE OPTIMISATION

eg. Yield & Resistance

IN-SILICO EXPERIMENTS

Resistance



WORK IN PROCESS

PROPRIETARY TECHNOLOGY **DATASTORIES CRISPR**

For millions of variables



datastories

APPROACH



- **APPLIED OUR IoT BEST PRACTICES FROM THE INDUSTRY**
- **INSPIRED BY GENOMICS PROPRIETARY TECHNOLOGY FOR VIRTUAL MARKERS**

RESULT DATASTORIES CRISPR

**ACCURACY IMPROVED FROM
45% (PCA) TO +84% (DataStories)**



OFFERING

- **AI SOFTWARE FOR RESEARCHERS**
- **MULTI OBJECTIVE OPTIMISATION**
- **SMART DESIGN OF EXPERIMENTS**
100x to 500x faster
- **UNICORN ACADEMY**



Unicorn Academy

- **Fundamentals of Data Science**
- **Advanced Data Science**
- **Business-driven Data Science**
- **Data science for R&D**
- **Systematic Innovation Organisation**
- **The Data infrastructure and Data engineering behind Data Science / Big Data**



