Interactive Visual Analysis with Galaxy Charts

Aysam Guerler

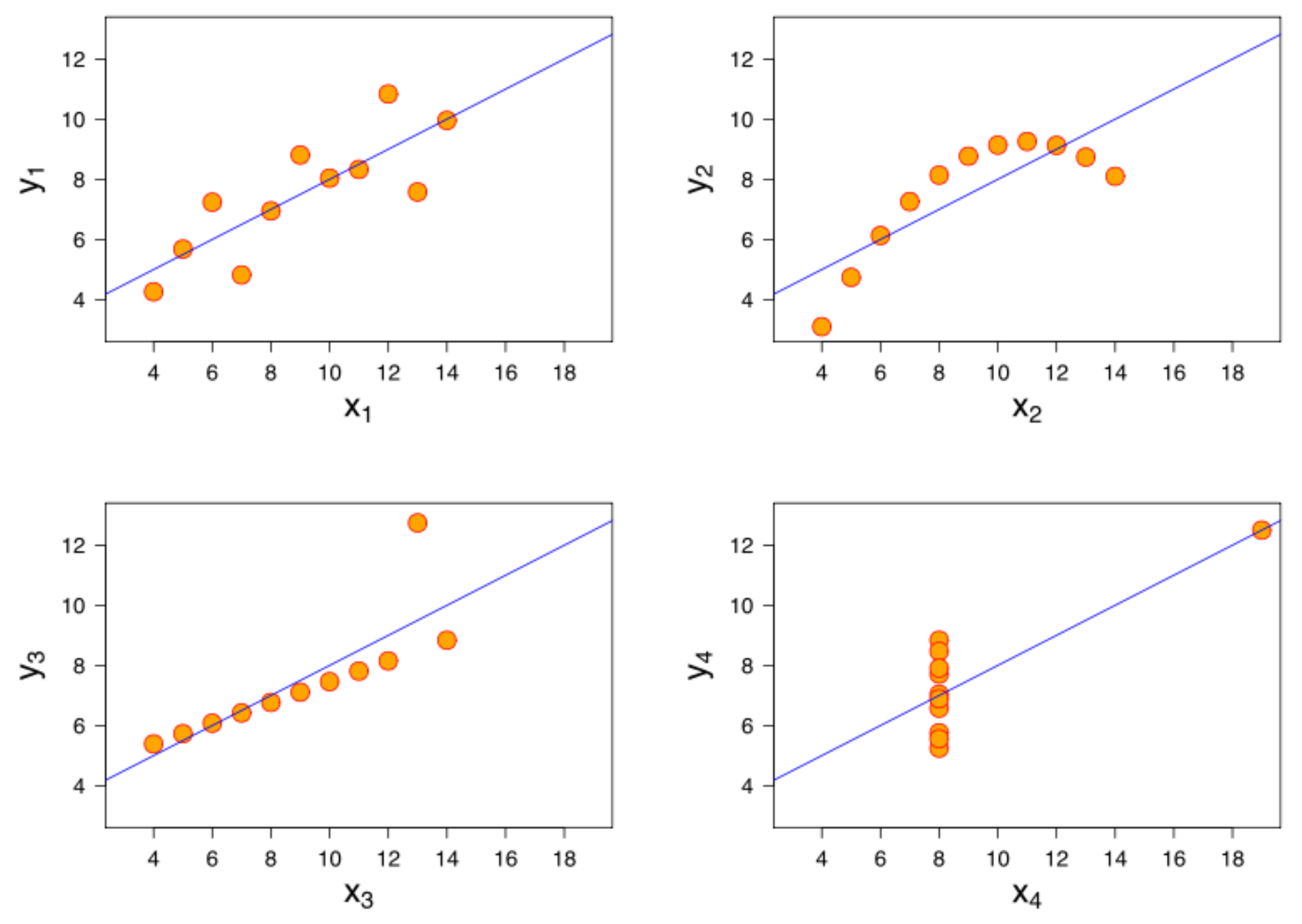
Galaxy Team – Taylor Lab, 2014

JOHNS HOPKINS UNIVERSITY

Why Visualize?

- Quick check: did it work?
- Exploration and hypothesis generation
- Sharing/publishing

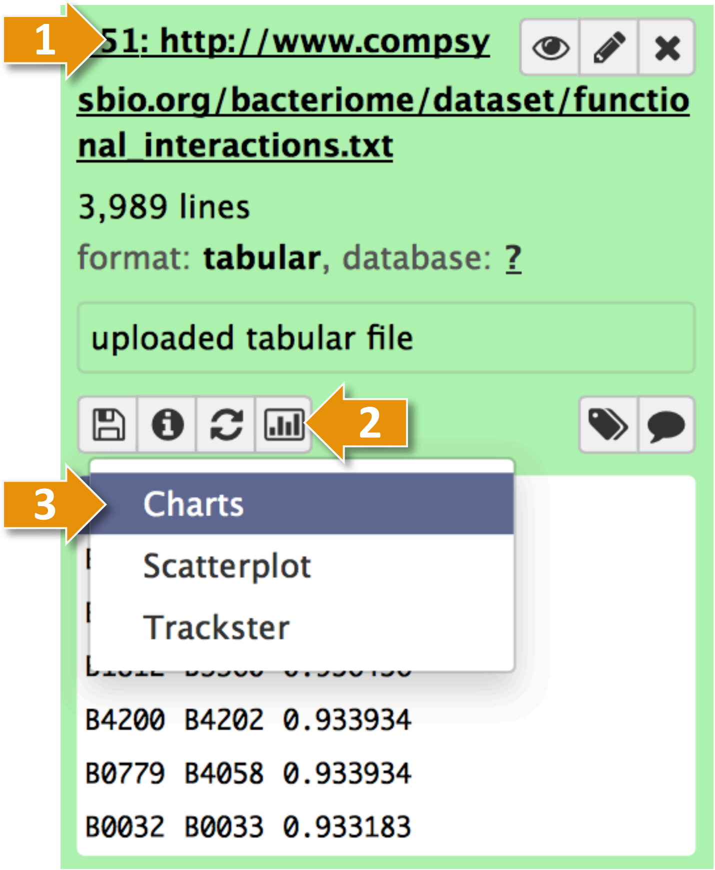
Anscombe’s Quartet



What is Galaxy Charts?

Use Galaxy | **Create Tabular Results** | **Visualize with Galaxy *Charts***

Make a new chart (1 of 4)



51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

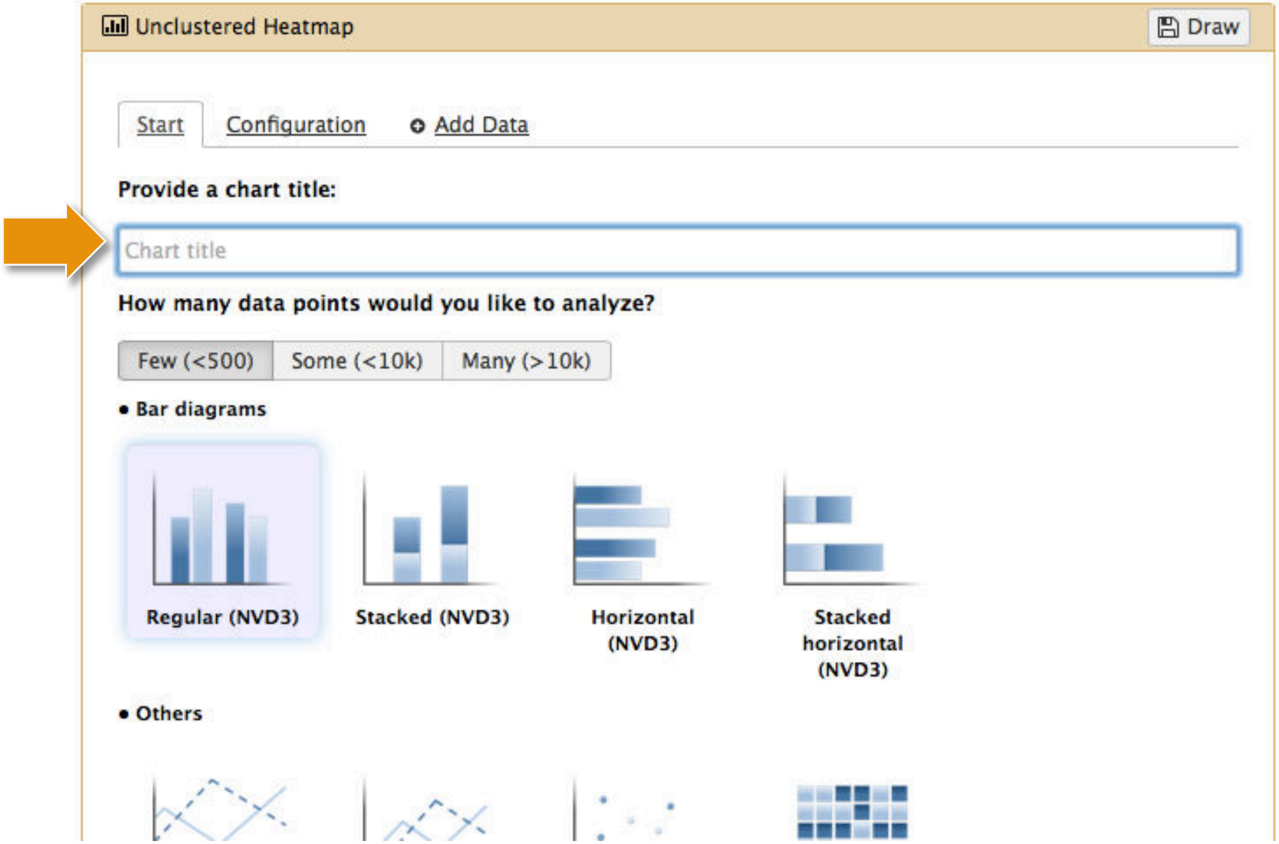
format: **tabular**, database: ?

uploaded tabular file

```
B4200  B4202  0.933934
B0779  B4058  0.933934
B0032  B0033  0.933183
```

Wait for the upload to complete. Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Unclustered Heatmap

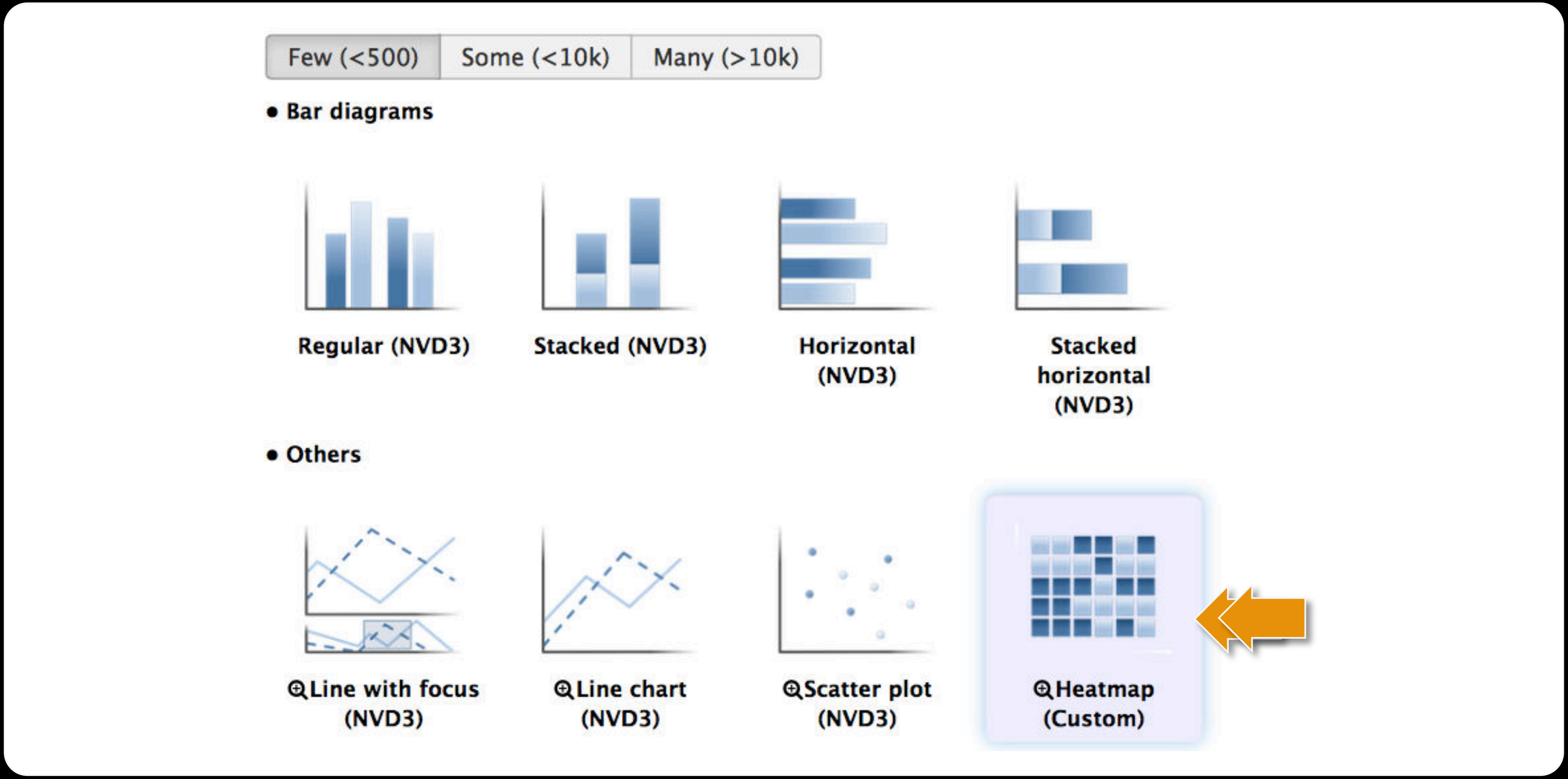
Provide a chart title:

Chart title

How many data points would you like to analyze?

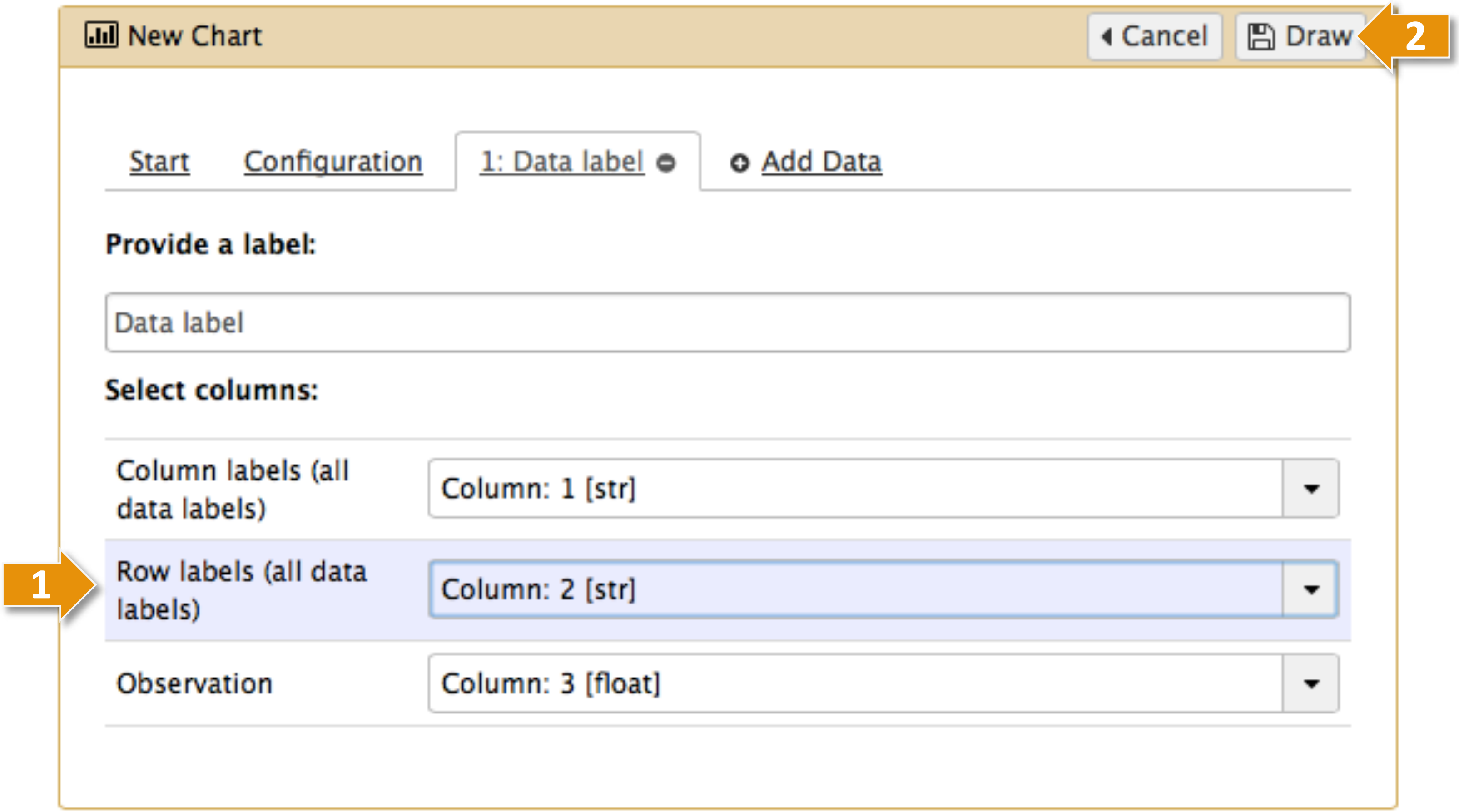
Name your chart **Unclustered Heatmap**.

Select a chart type



Double click on the **Heatmap** icon.

Select data columns



Provide a label:

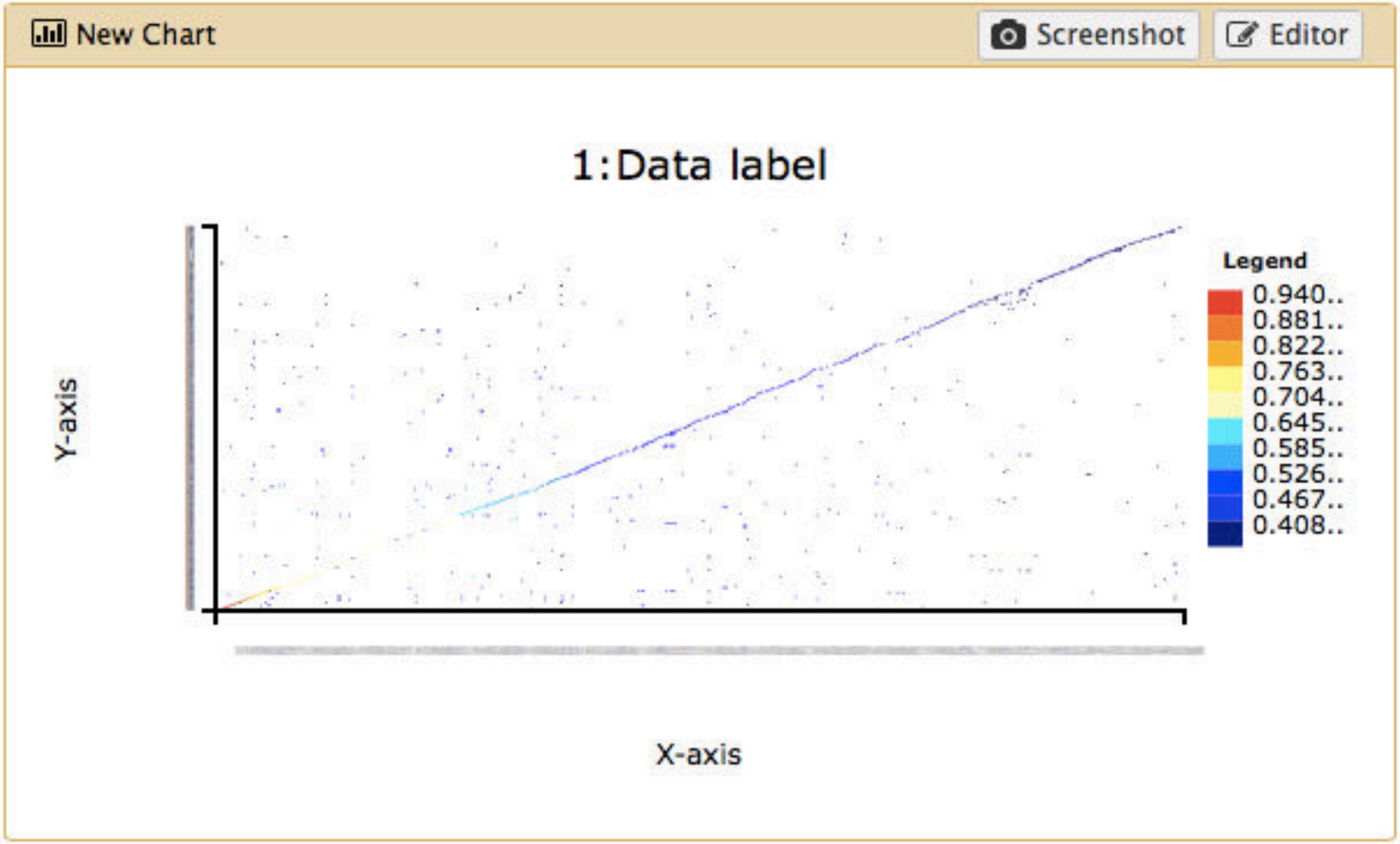
Data label

Select columns:

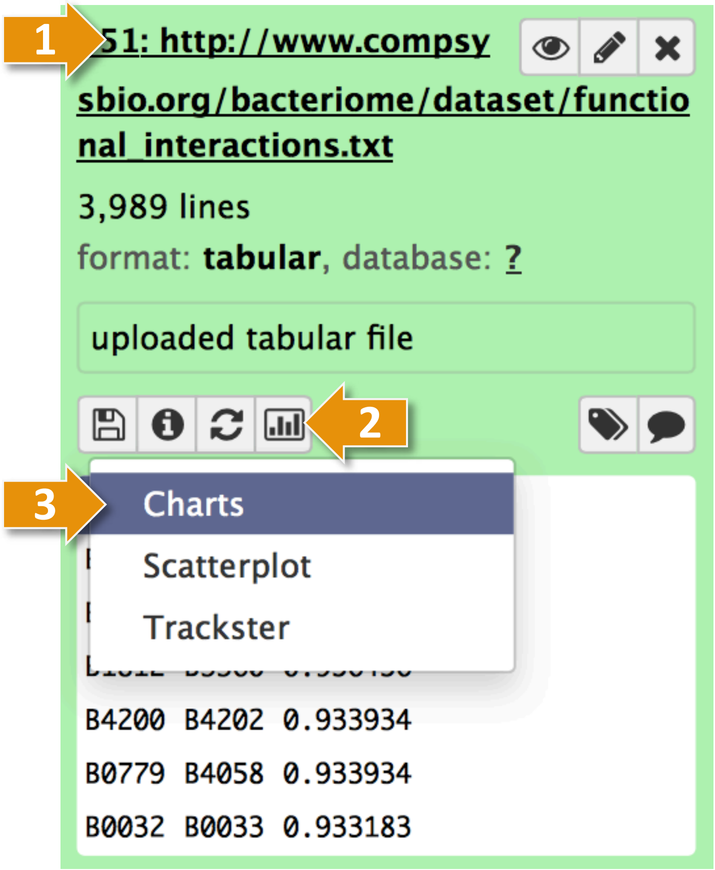
Column labels (all data labels)	Column: 1 [str]
Row labels (all data labels)	Column: 2 [str]
Observation	Column: 3 [float]

At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

Unclustered Heatmap

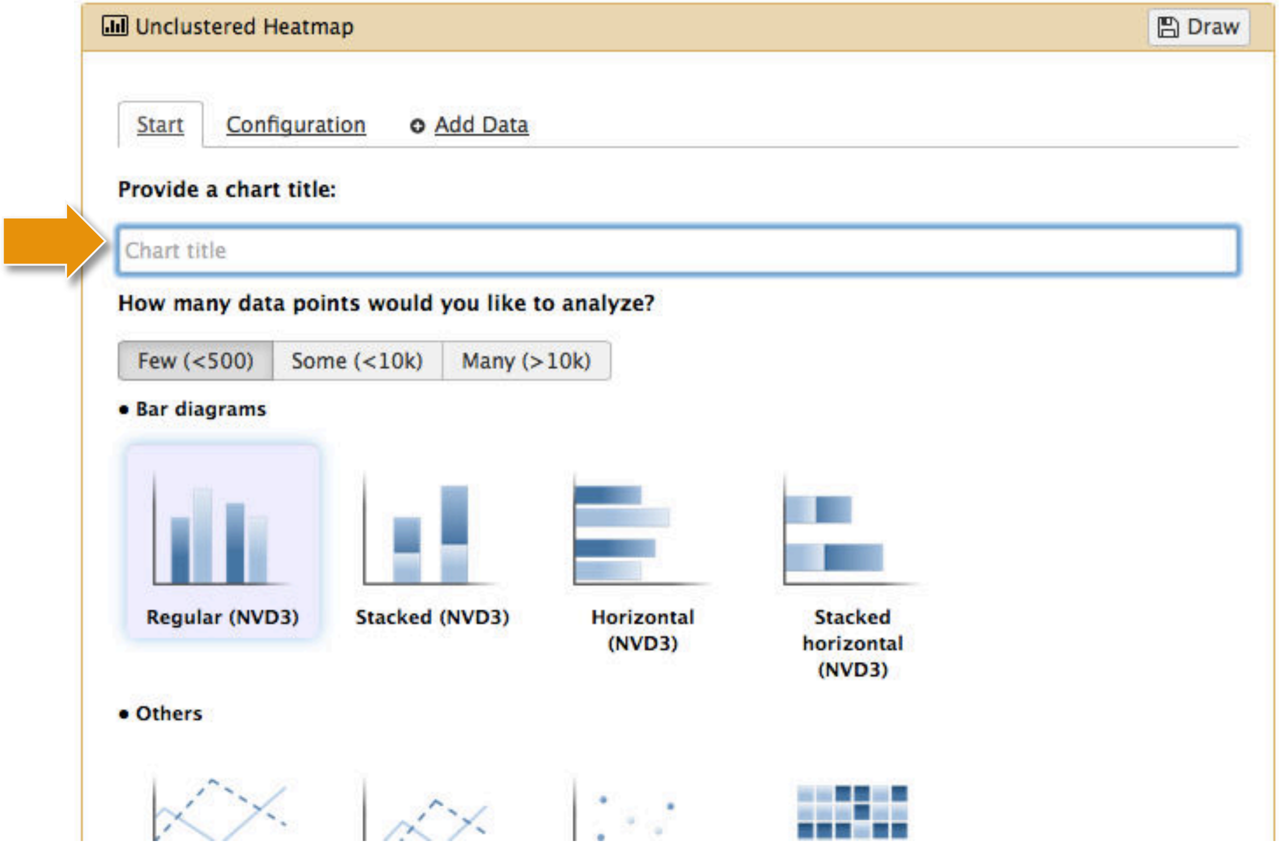


Make a new chart (2 of 4)



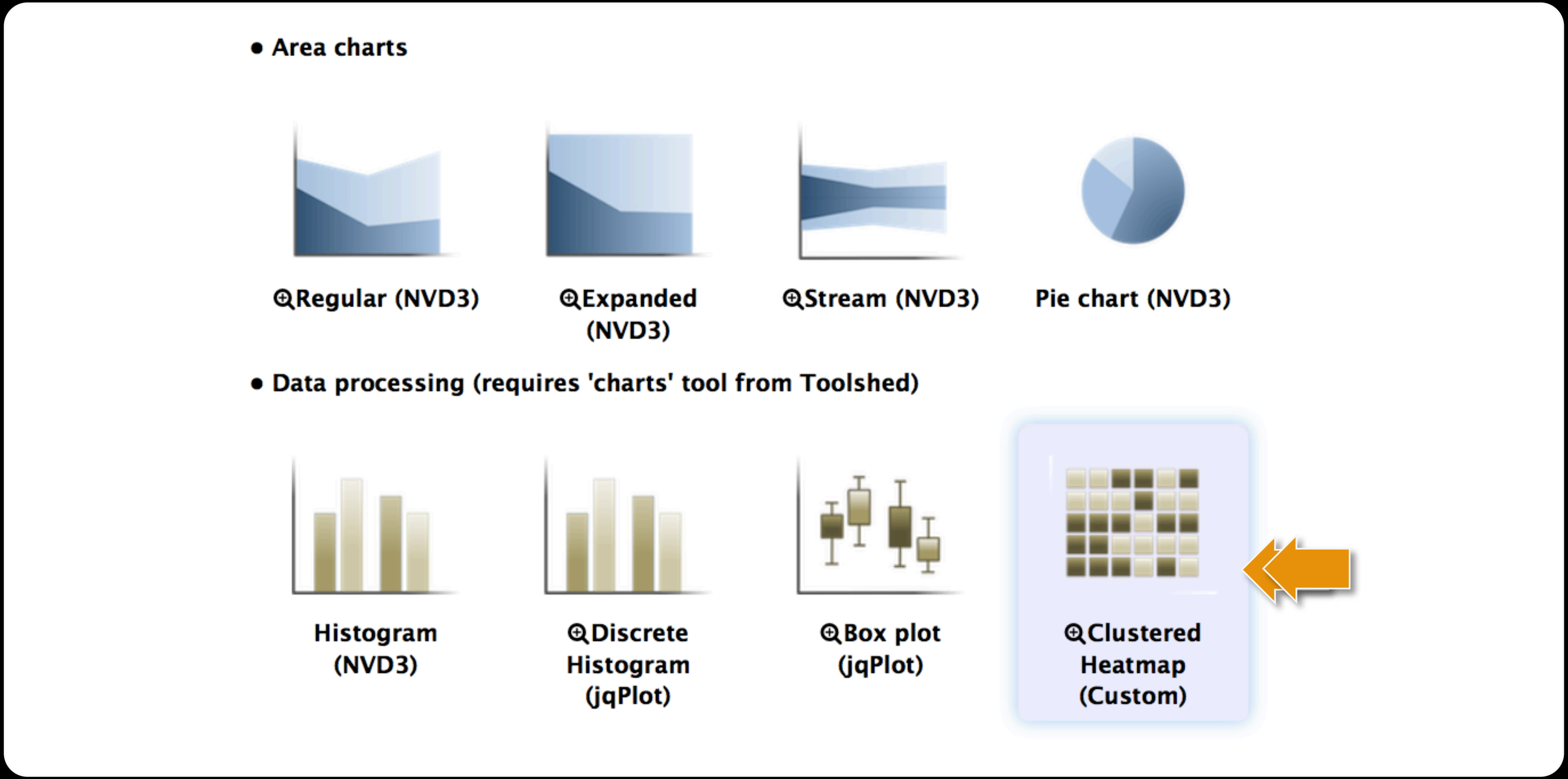
Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Name your chart **Clustered Heatmap**.

Select a new chart type



- Area charts

Regular (NVD3)

Expanded (NVD3)

Stream (NVD3)

Pie chart (NVD3)

- Data processing (requires 'charts' tool from Toolshed)

Histogram (NVD3)

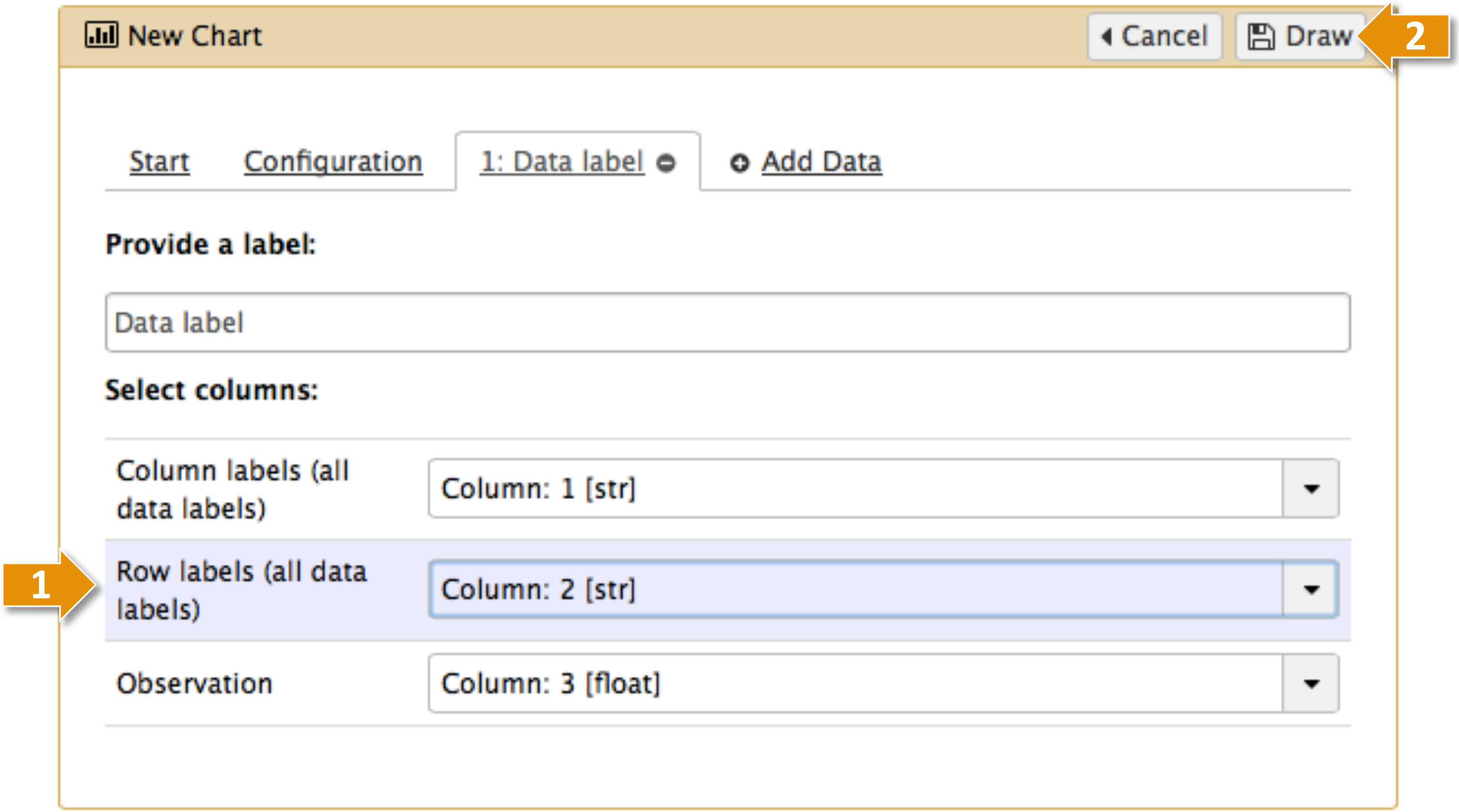
Discrete Histogram (jqPlot)

Box plot (jqPlot)

Clustered Heatmap (Custom)

Double click on the **Clustered Heatmap** icon.

Select data columns



Provide a label:

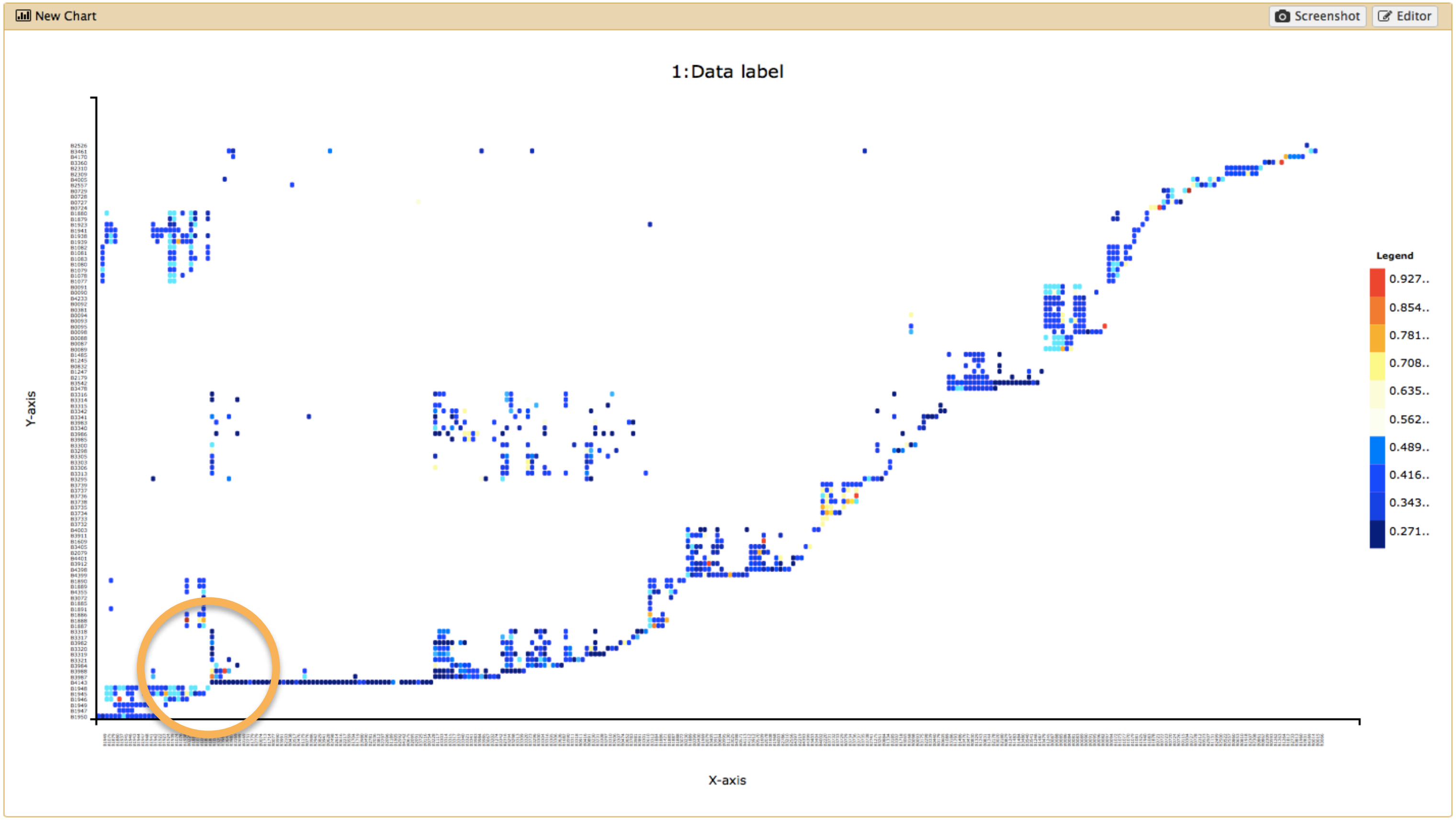
Data label

Select columns:

Column labels (all data labels)	Column: 1 [str]
Row labels (all data labels)	Column: 2 [str]
Observation	Column: 3 [float]

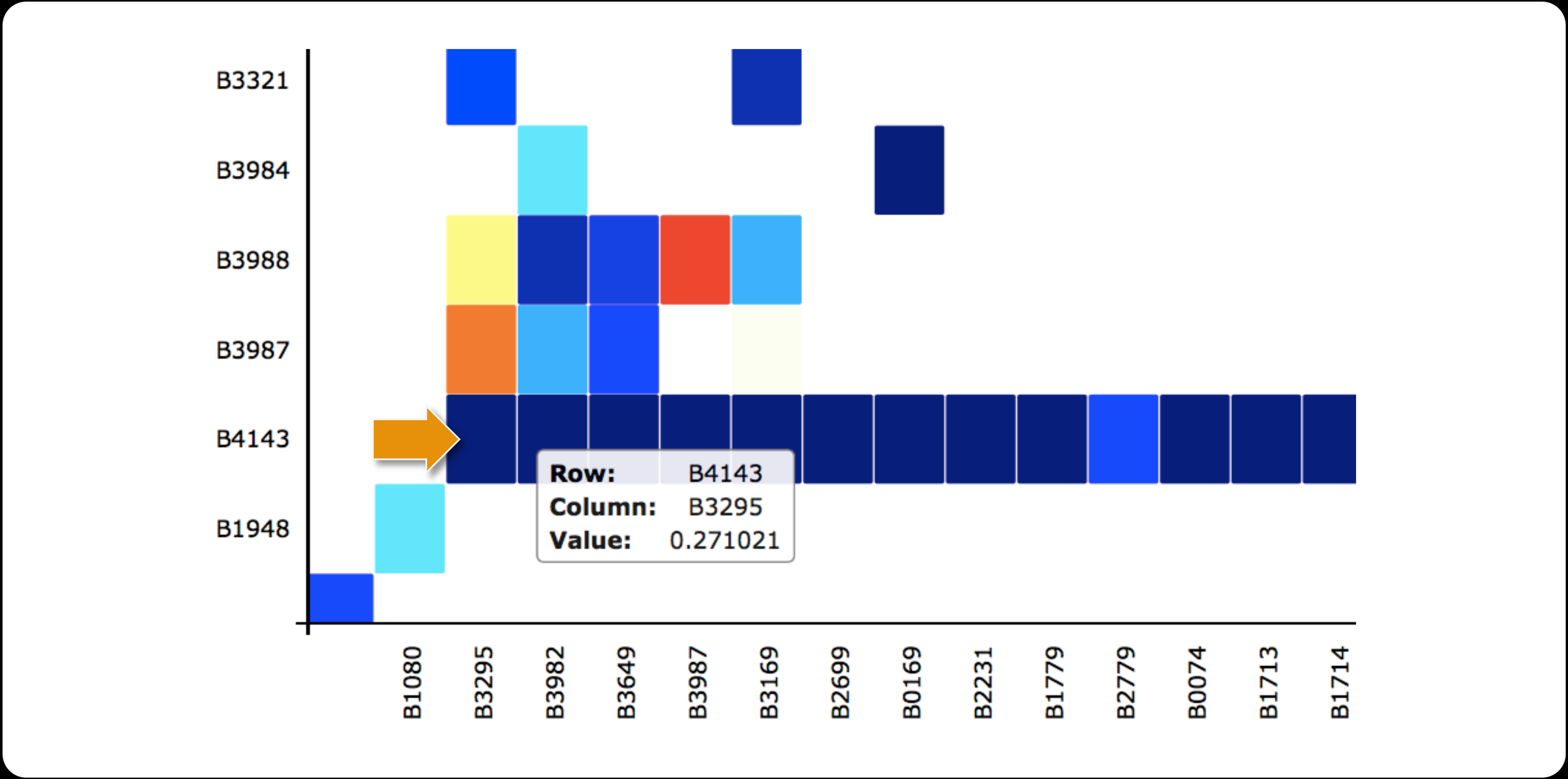
At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

Clustered Heatmap



Use the **mouse wheel or your touch pad** to zoom into the **highlighted area**.

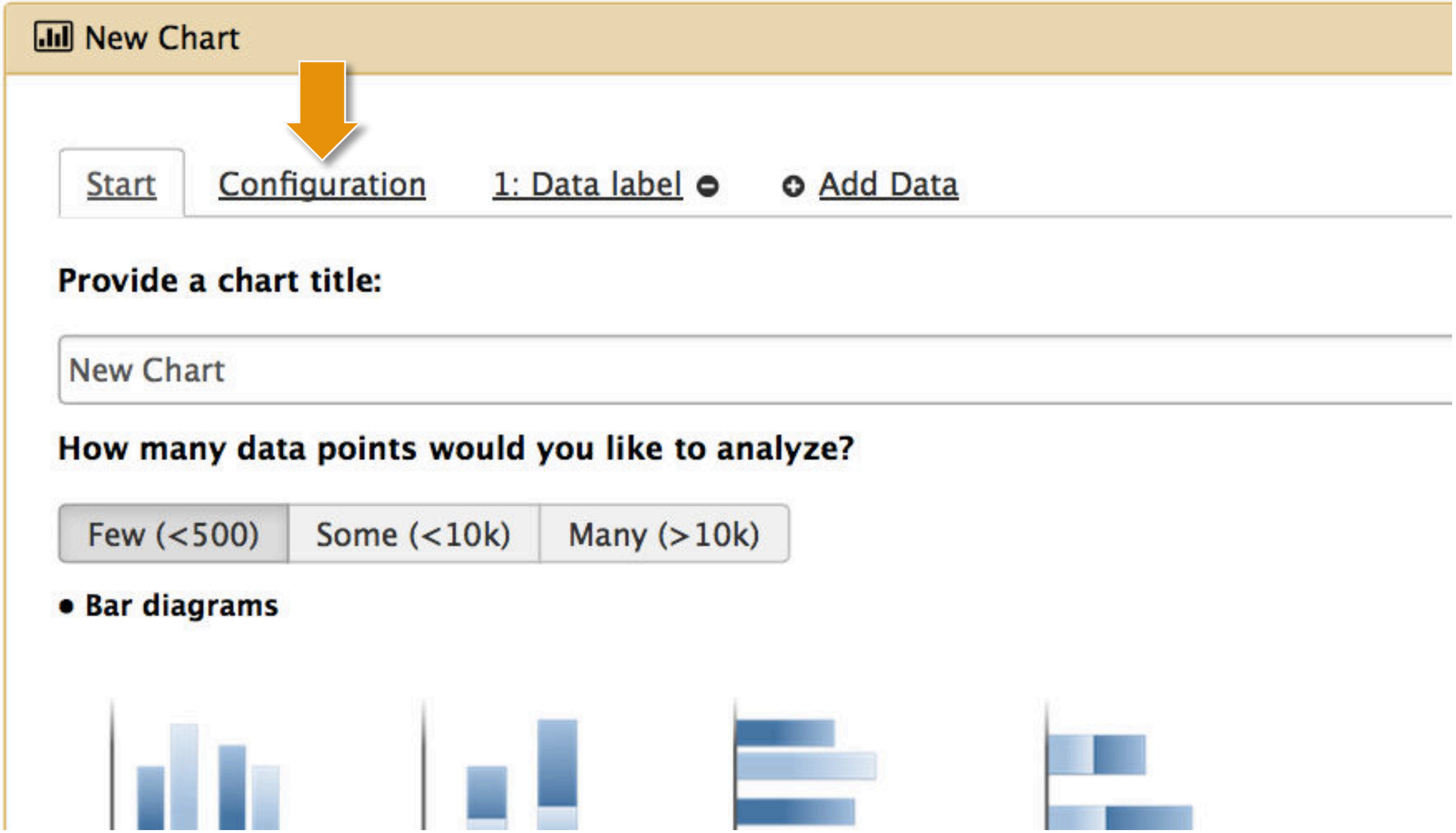
Enlarged view



Row: B4143
Column: B3295
Value: 0.271021

Tooltips popup if you move the mouse pointer over a box. Here the interaction between **B4143** and **B3295** is highlighted. Click on **Editor** again to further customize this chart.

Chart configuration



Provide a chart title:

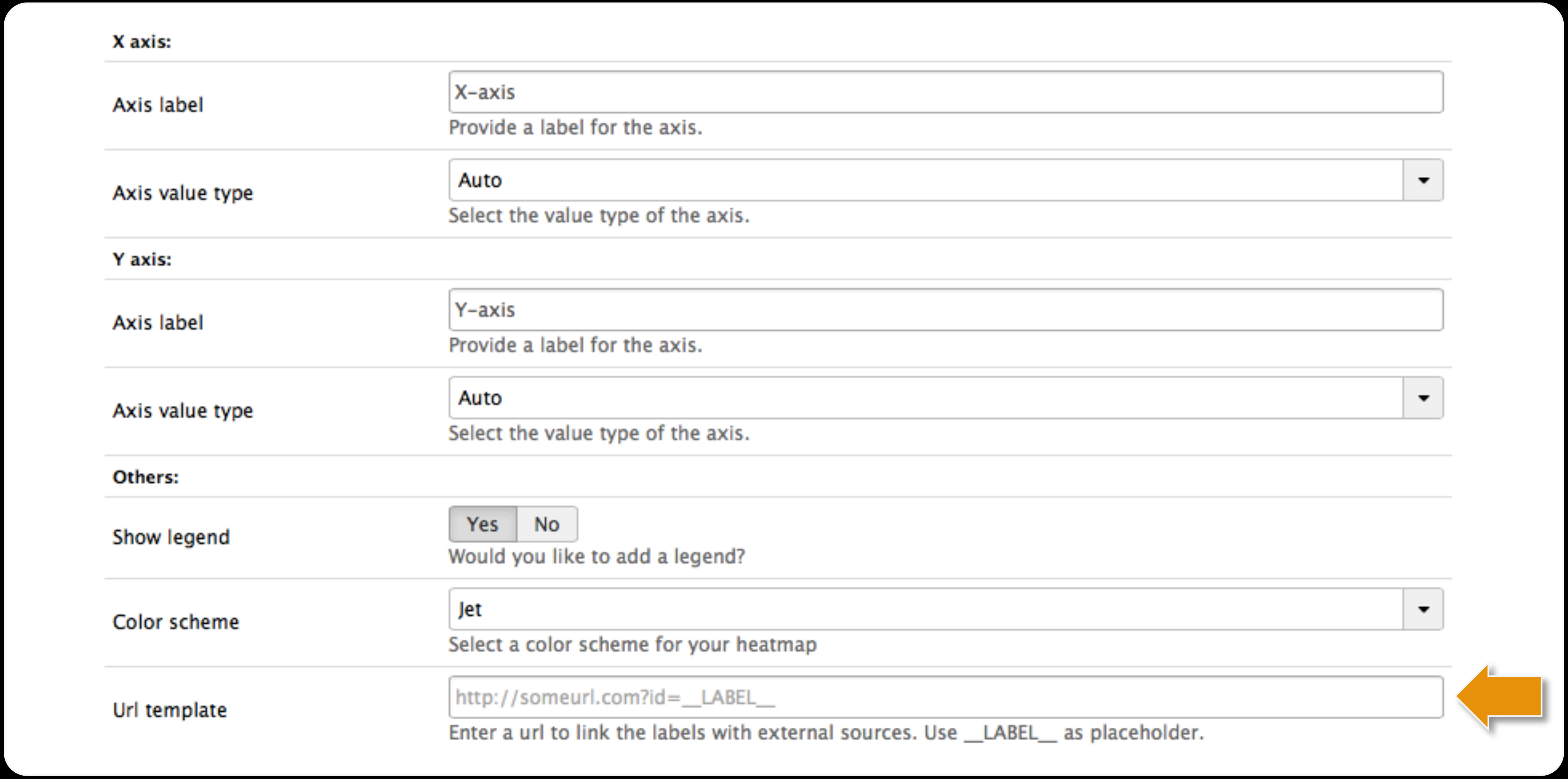
New Chart

How many data points would you like to analyze?

- Bar diagrams

Go to the **Configuration** tab.

Chart settings



X axis:

Axis label: X-axis
Provide a label for the axis.

Axis value type: Auto
Select the value type of the axis.

Y axis:

Axis label: Y-axis
Provide a label for the axis.

Axis value type: Auto
Select the value type of the axis.

Others:

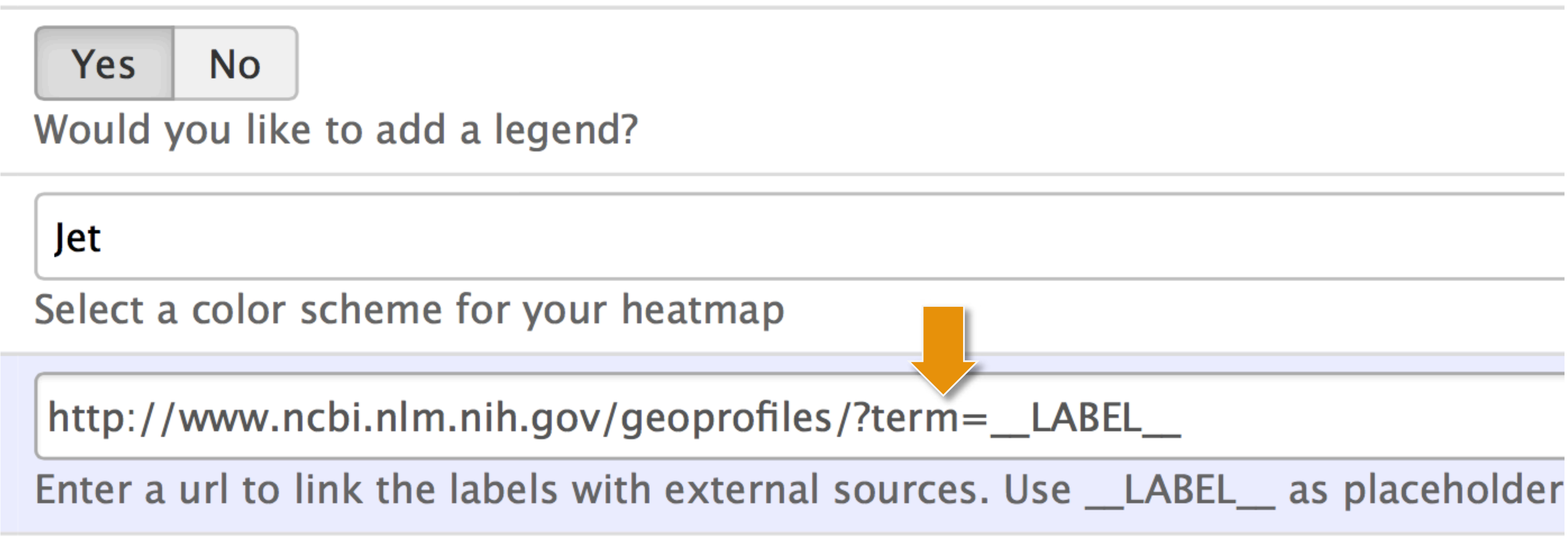
Show legend: Yes No
Would you like to add a legend?

Color scheme: Jet
Select a color scheme for your heatmap

Url template: http://someurl.com?id=__LABEL__
Enter a url to link the labels with external sources. Use __LABEL__ as placeholder.

Heatmap specific options are **highlighted**. Feel free to set **axis labels** or other options.

Define a URL template



Would you like to add a legend?

Jet

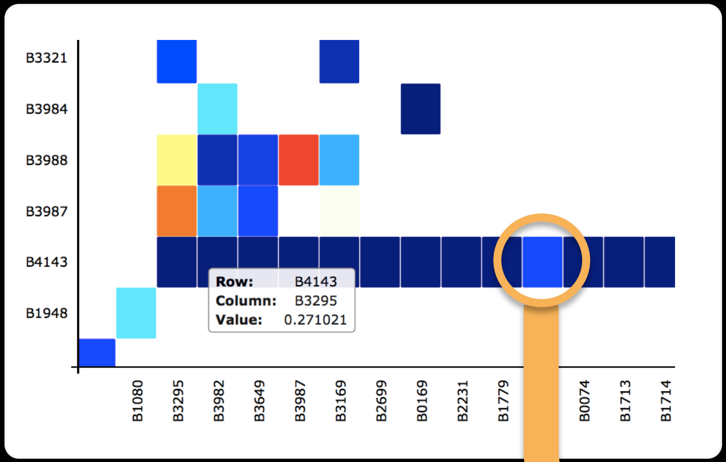
Select a color scheme for your heatmap

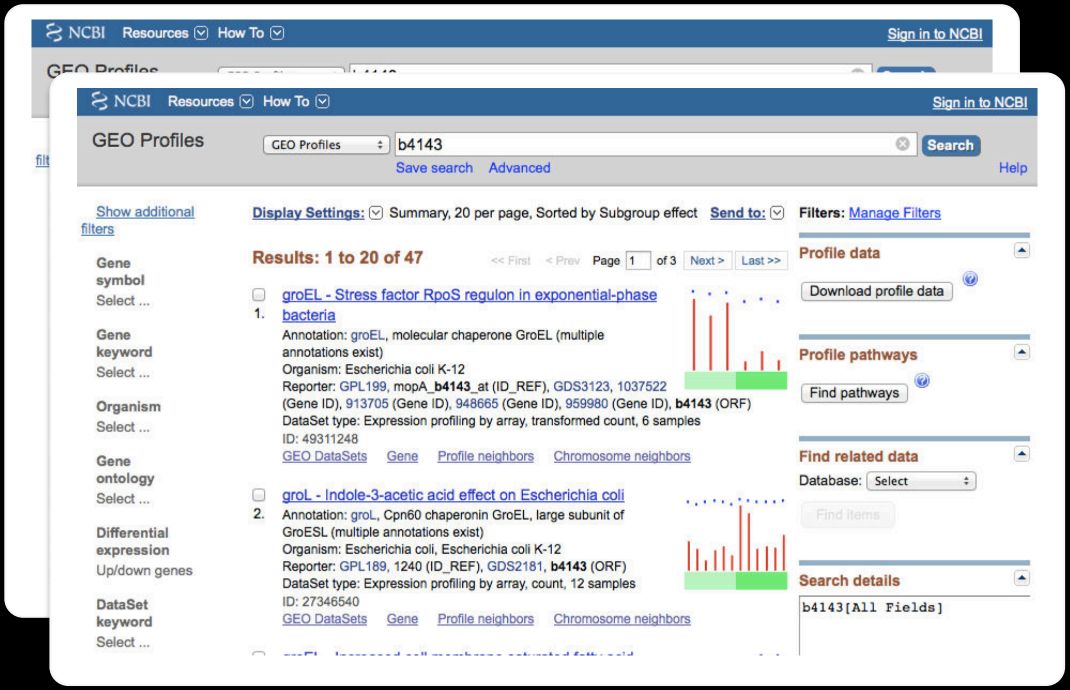
http://www.ncbi.nlm.nih.gov/geoprofiles/?term=__LABEL__

Enter a url to link the labels with external sources. Use __LABEL__ as placeholder

Paste a **database URL** into the template URL field and add the **__LABEL__** tag. You may use **http://www.ncbi.nlm.nih.gov** or any other database. Click on **Draw** to redraw the chart.

Data points linked to web sources





GEO Profiles: b4143

Show additional filters

Gene symbol: Select ...
Gene keyword: Select ...
Organism: Select ...
Gene ontology: Select ...
Differential expression: Up/down genes
DataSet keyword: Select ...

Display Settings: Summary, 20 per page, Sorted by Subgroup effect
Send to:
Filters: Manage Filters

Results: 1 to 20 of 47

<< First < Prev Page 1 of 3 Next > Last >>

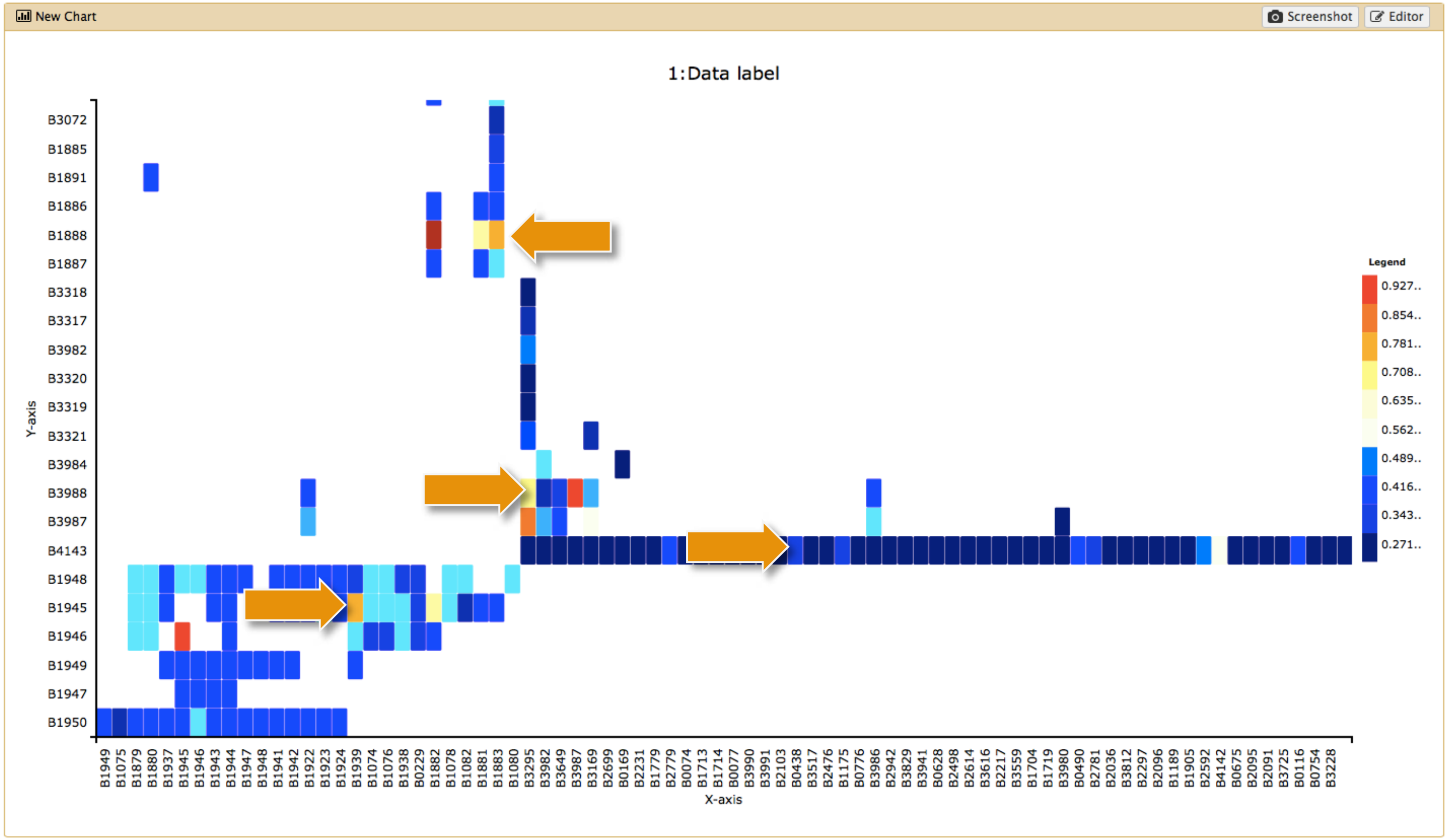
☐ 1. groEL - Stress factor RpoS regulon in exponential-phase bacteria
Annotation: groEL, molecular chaperone GroEL (multiple annotations exist)
Organism: Escherichia coli K-12
Reporter: GPL199, mopA_**b4143**_at (ID_REF), GDS3123, 1037522 (Gene ID), 913705 (Gene ID), 948665 (Gene ID), 959980 (Gene ID), **b4143** (ORF)
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 49311248
GEO DataSets Gene Profile neighbors Chromosome neighbors

☐ 2. groL - Indole-3-acetic acid effect on Escherichia coli
Annotation: groL, Cpn60 chaperonin GroEL, large subunit of GroESL (multiple annotations exist)
Organism: Escherichia coli, Escherichia coli K-12
Reporter: GPL189, 1240 (ID_REF), GDS2181, **b4143** (ORF)
DataSet type: Expression profiling by array, count, 12 samples
ID: 27346540
GEO DataSets Gene Profile neighbors Chromosome neighbors

Profile data: Download profile data
Profile pathways: Find pathways
Find related data: Database: Select; Find items
Search details: b4143[All Fields]

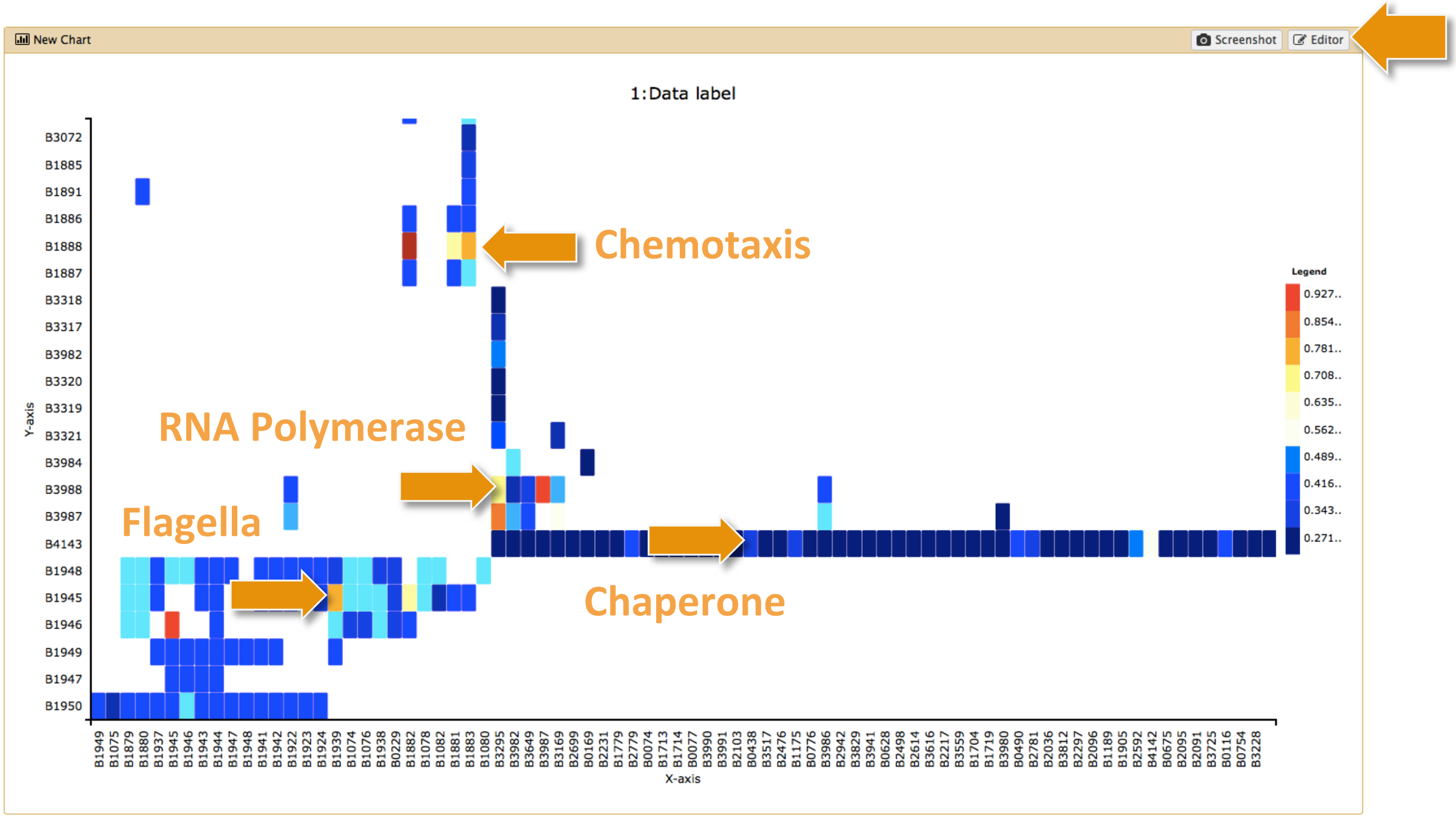
Double click on a **box** and the browser will open two new tabs using the previously defined **URL template**.

Cluster selection and analysis



Select one element from each **highlighted row**. What are the corresponding **protein functions?**

Identified protein categories



Chemotaxis

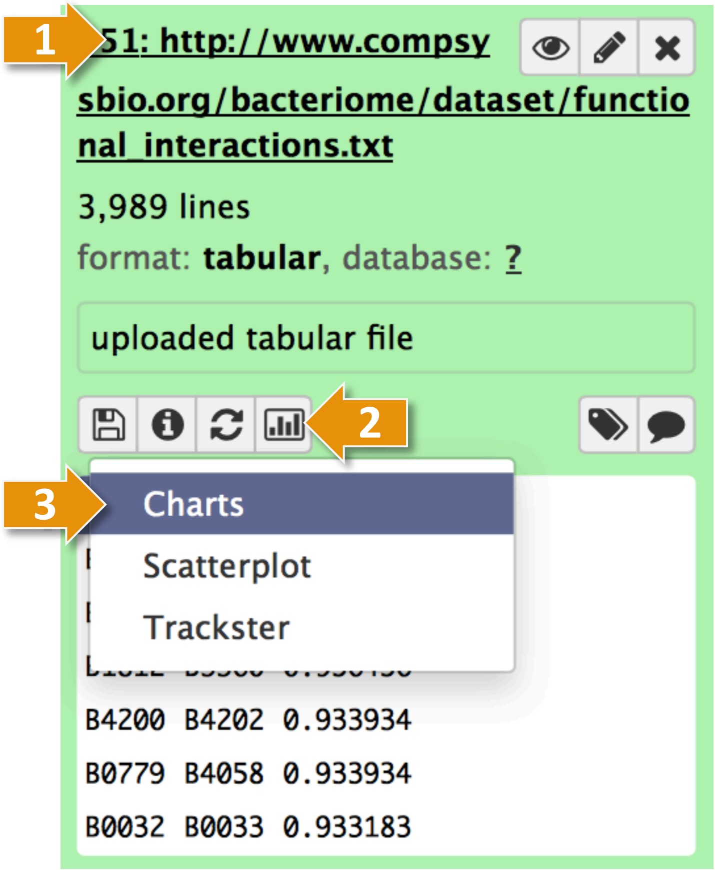
RNA Polymerase

Flagella

Chaperone

Please return to the **Editor**.

Make a new chart (3 of 4)



51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

format: **tabular**, database: ?

uploaded tabular file

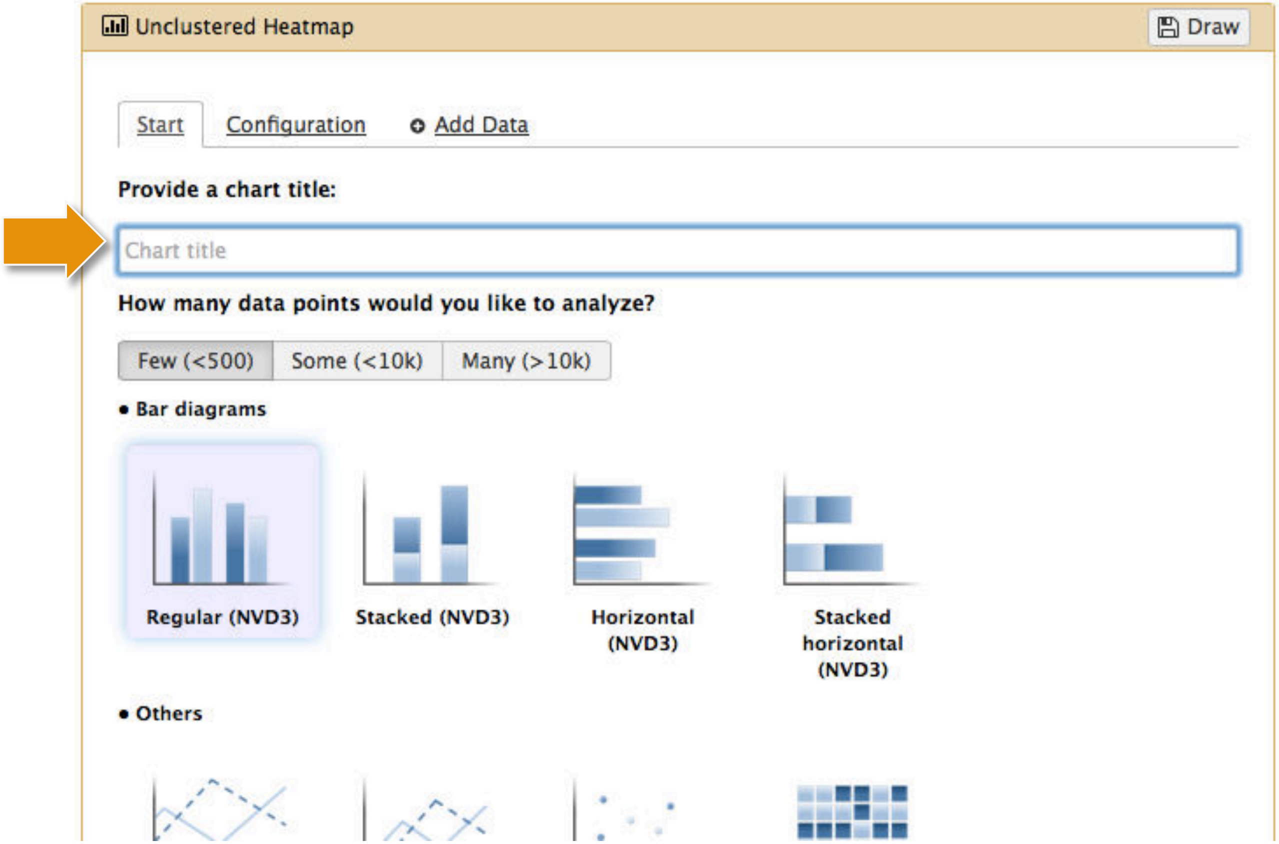
B4200 B4202 0.933934

B0779 B4058 0.933934

B0032 B0033 0.933183

Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Provide a chart title:

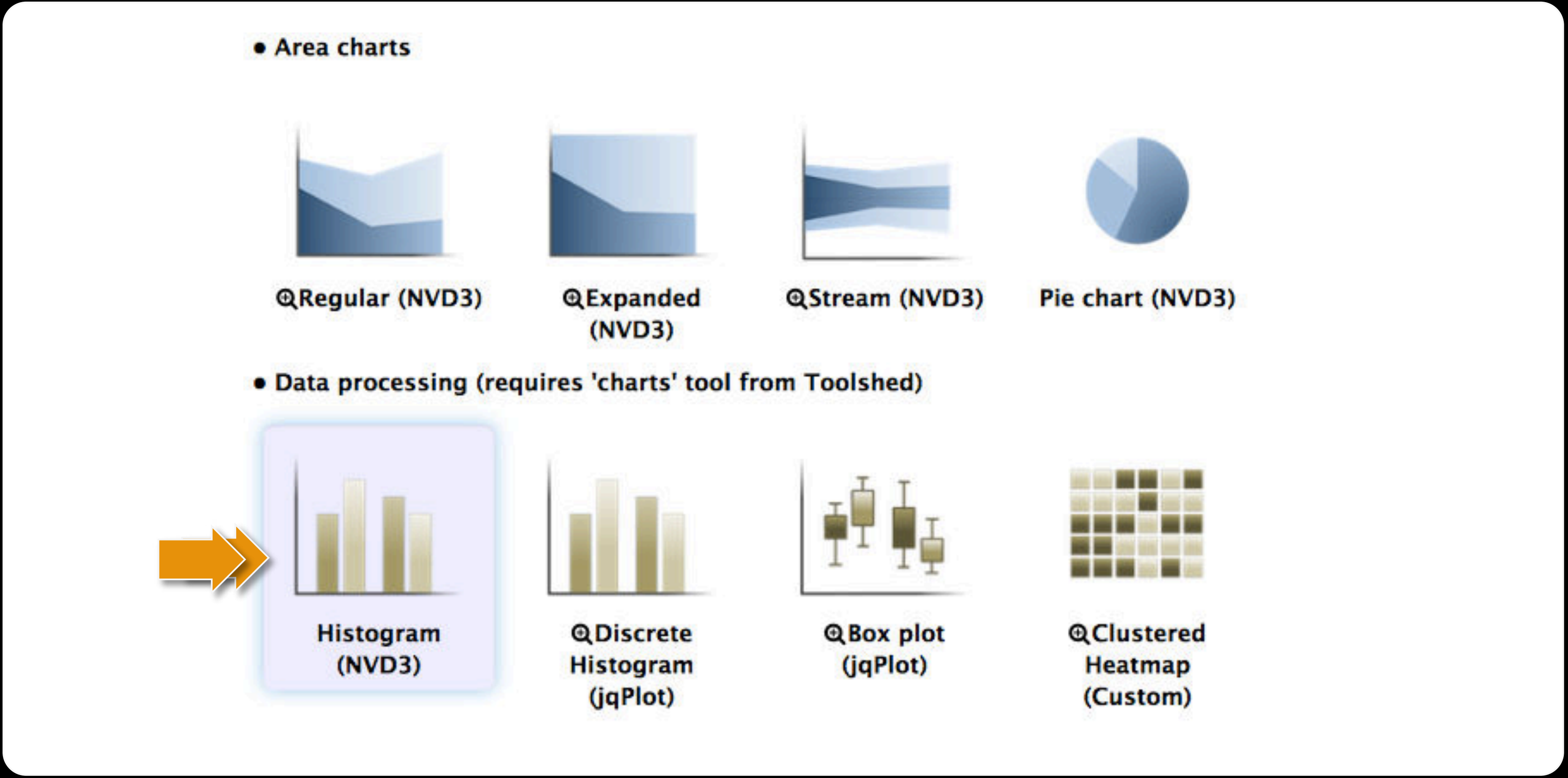
How many data points would you like to analyze?

- Bar diagrams

- Others

Name your chart **Score Histogram**.

Analyze the score distribution

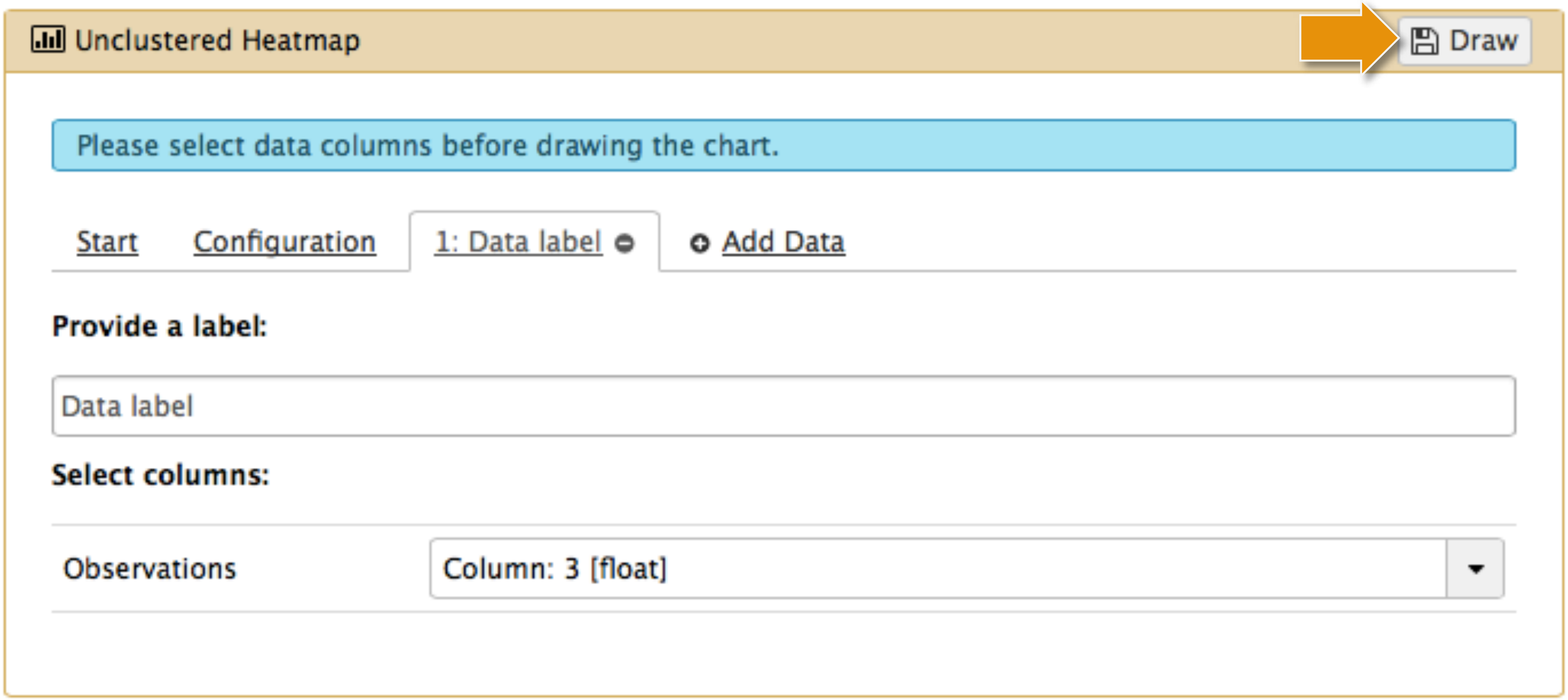


- Area charts

- Data processing (requires 'charts' tool from Toolshed)

Double click on the **Histogram** icon and click on **Draw**.

Give your chart a name



Please select data columns before drawing the chart.

Provide a label:

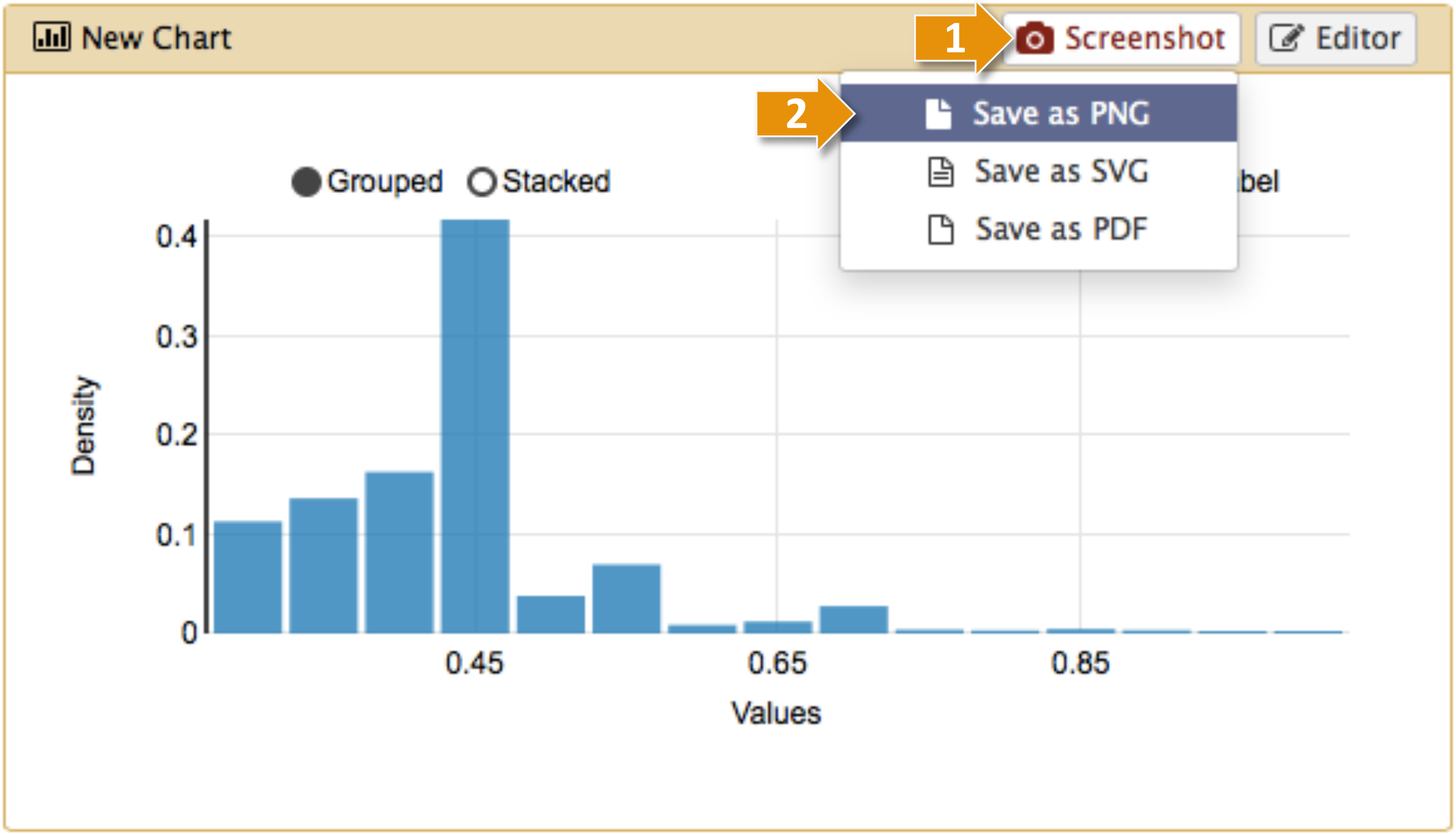
Data label

Select columns:

Observations: Column: 3 [float]

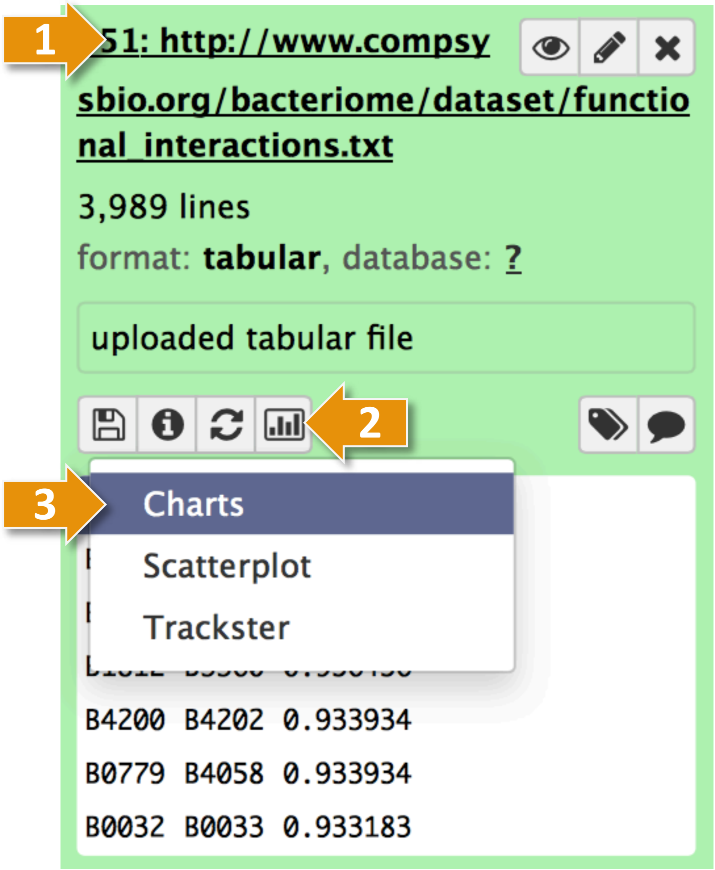
Click on **Draw**.

Export as PNG



Click on **Screenshot** and select **Save as PNG**. Finally, return to the **Editor** again.

Make a new chart (4 of 4)



1

51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

format: **tabular**, database: ?

uploaded tabular file

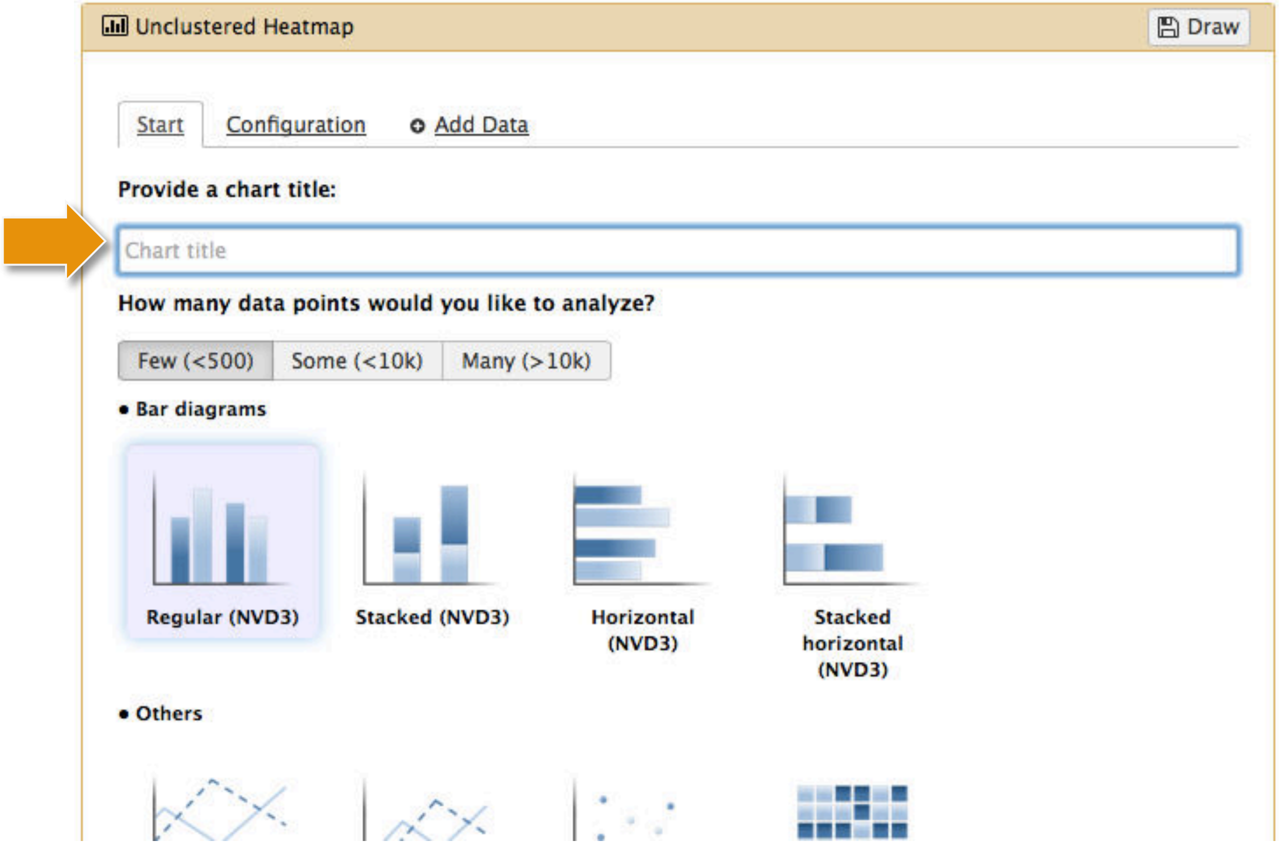
2

3

```
B4200  B4202  0.933934
B0779  B4058  0.933934
B0032  B0033  0.933183
```

Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name

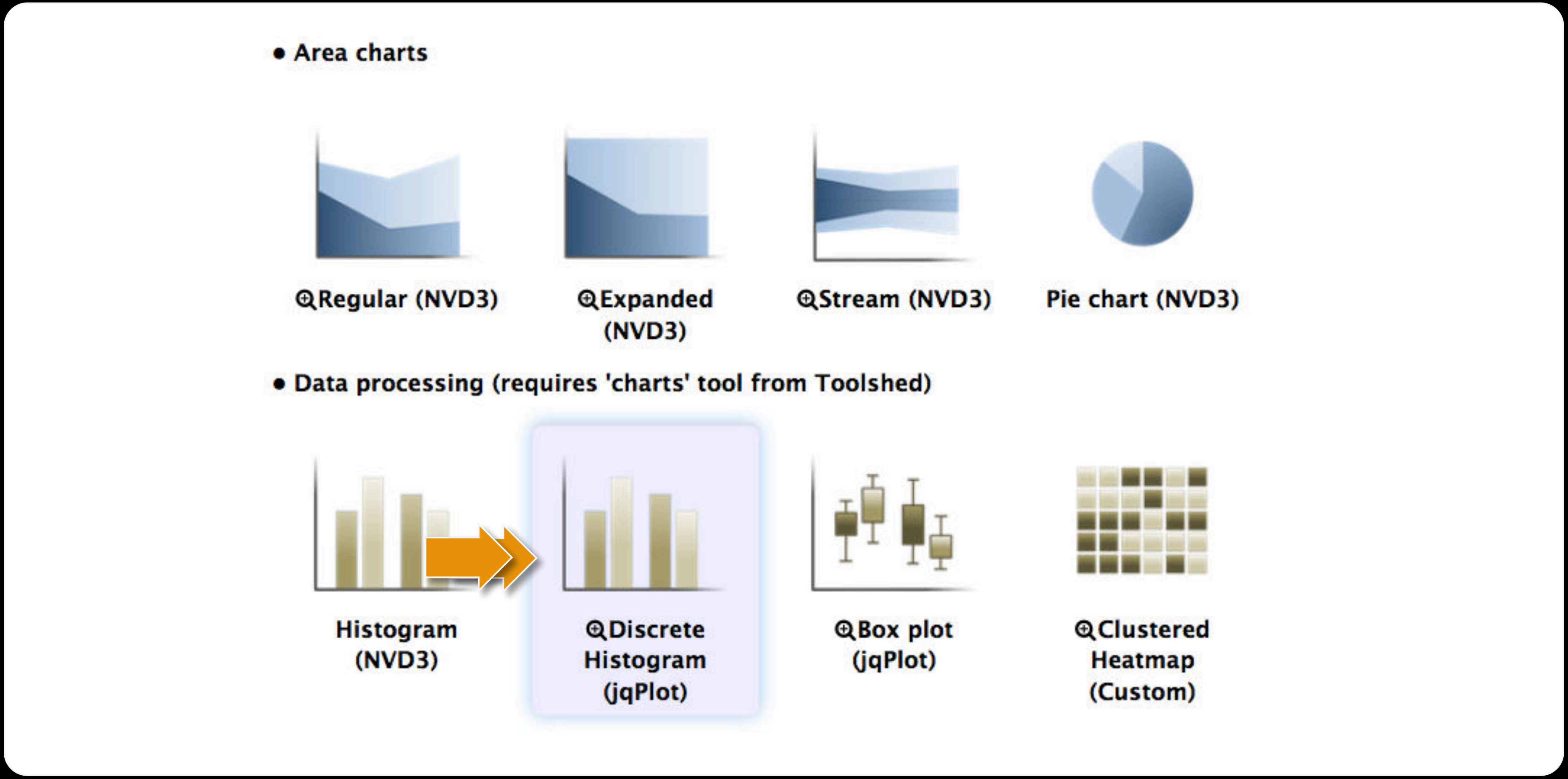


Provide a chart title:

How many data points would you like to analyze?

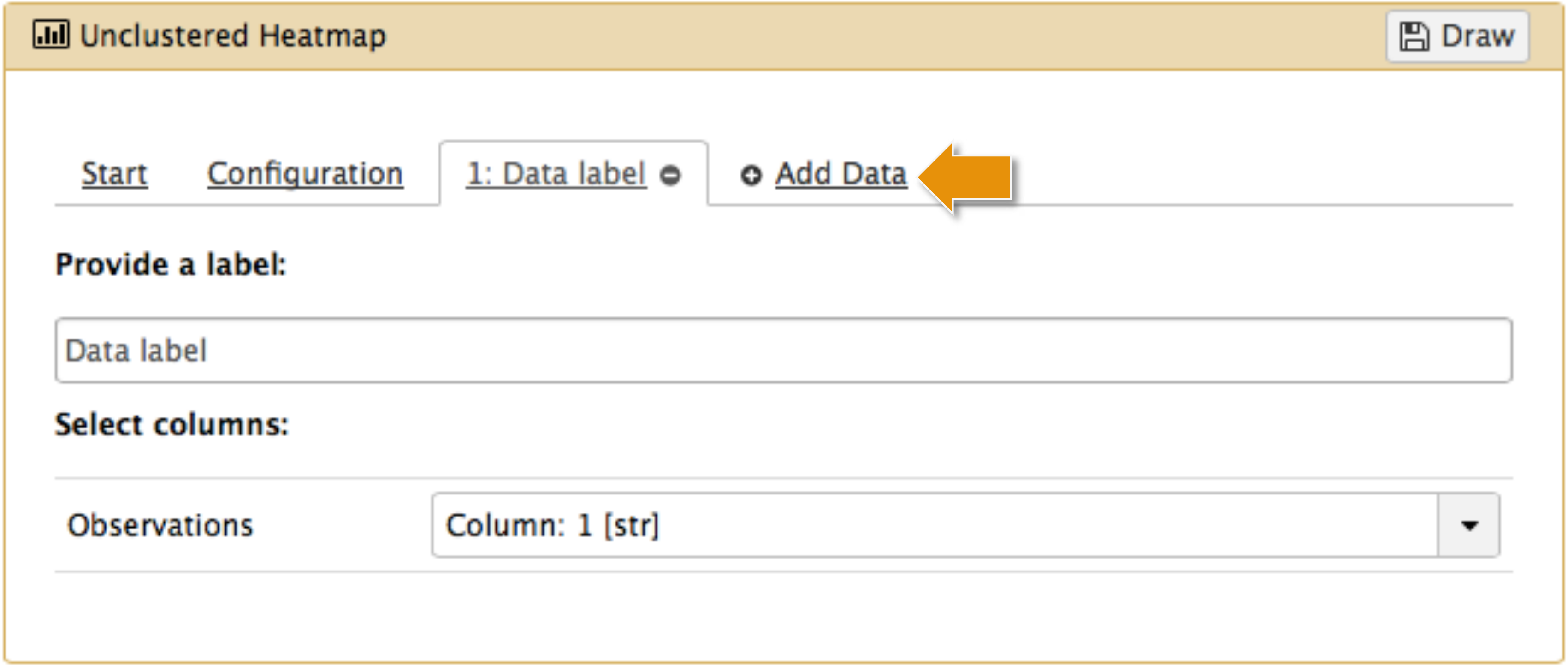
Name your chart **Discrete Histogram**.

Analyze the protein distribution



Double click on the **Discrete Histogram** icon.

Add more data



Unclustered Heatmap

Provide a label:

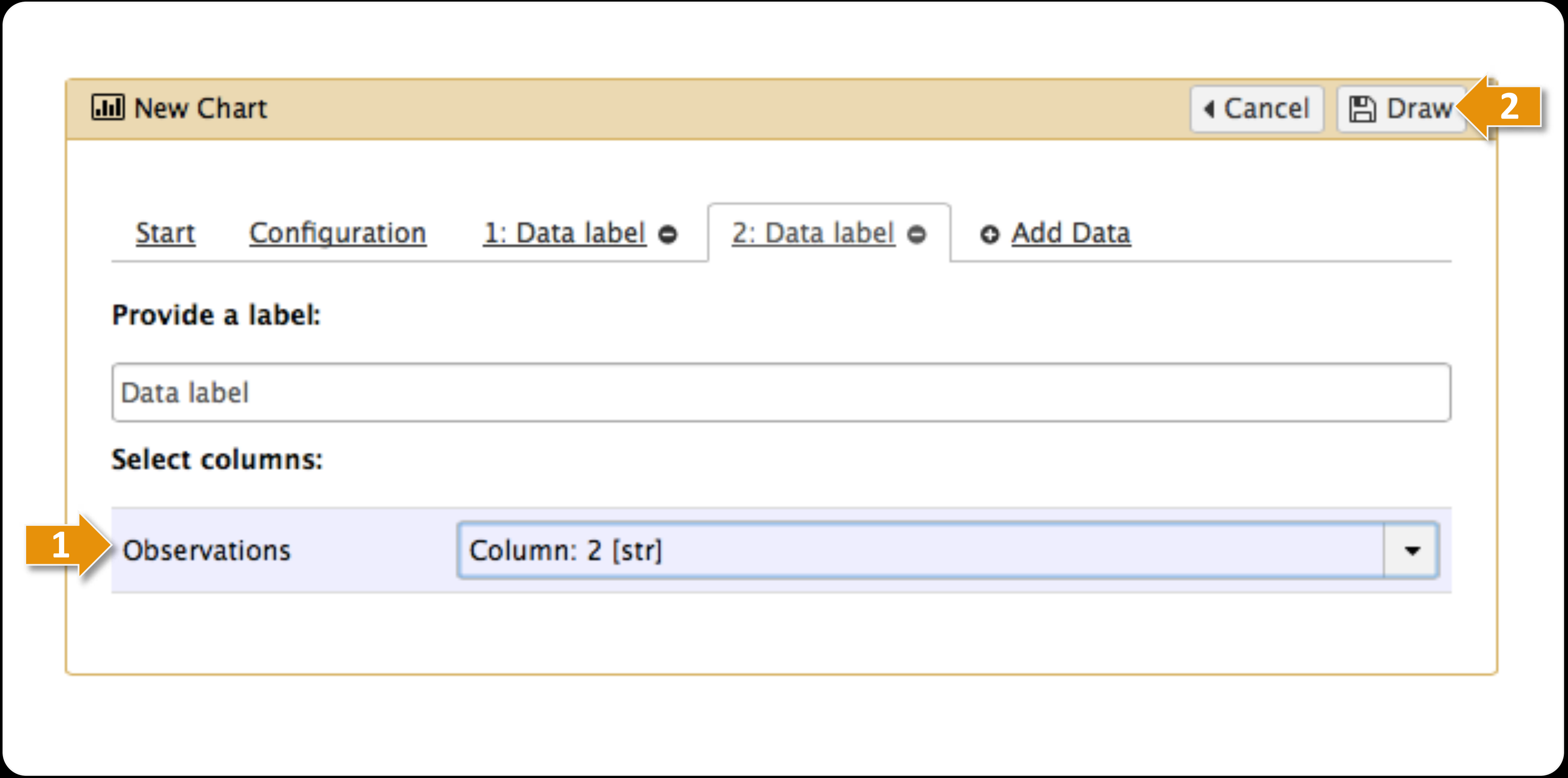
Data label

Select columns:

Observations: Column: 1 [str]

Click on **Add Data**.

Select a second data group



Provide a label:

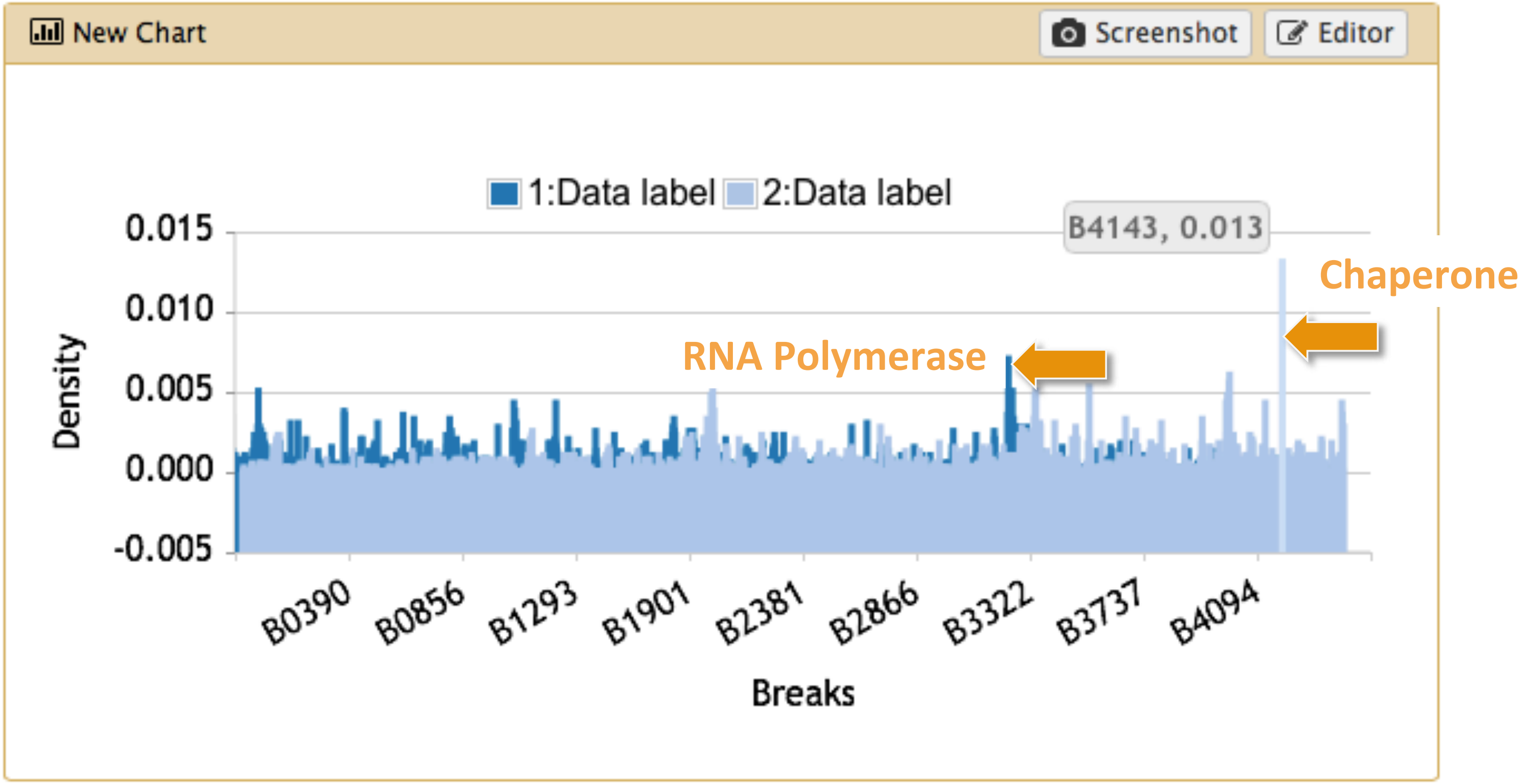
Data label

Select columns:

Observations: Column: 2 [str]

At first click on **Observations** and select **Column 2**. Then, click on **Draw**.

Which proteins have most interactions?



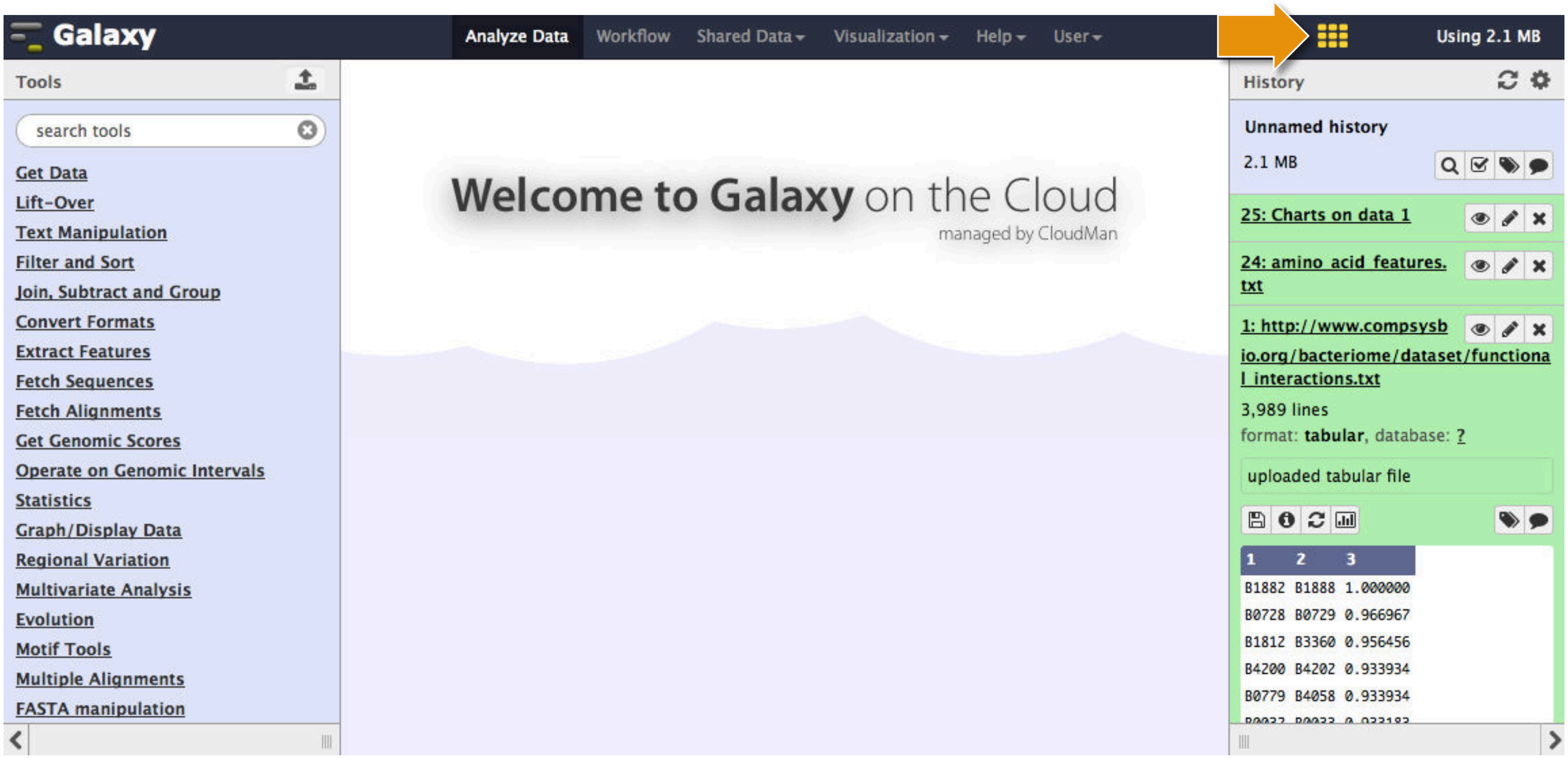
RNA Polymerase

Chaperone

Done with Part I.

Scratchbook

Activate the Scratchbook



Unnamed history

2.1 MB

25: Charts on data 1

24: amino_acid_features.txt

1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

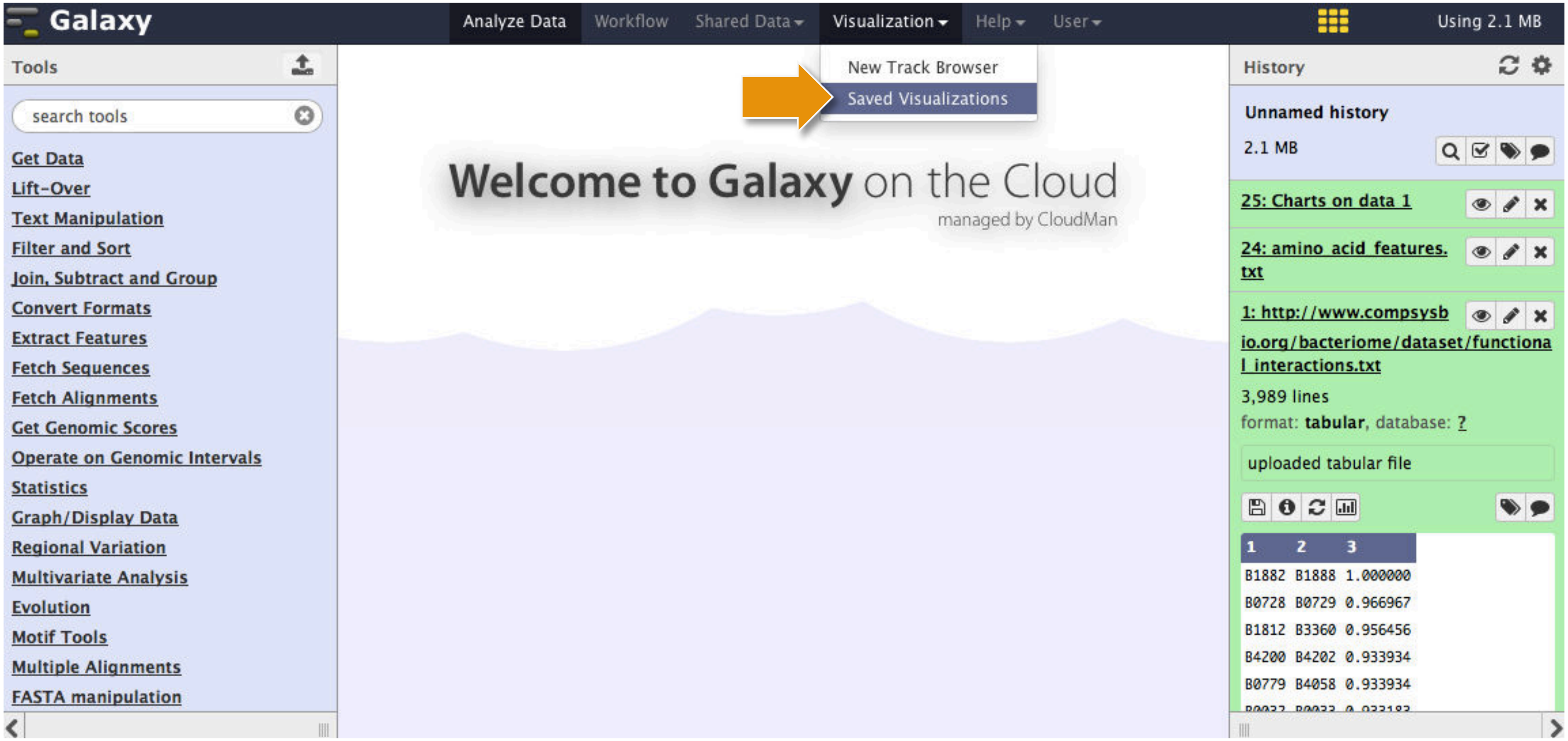
format: **tabular**, database: ?

uploaded tabular file

1	2	3
B1882	B1888	1.000000
B0728	B0729	0.966967
B1812	B3360	0.956456
B4200	B4202	0.933934
B0779	B4058	0.933934

Activate the **Scratchbook** by clicking on the above icon.

Activate the Scratchbook



3,989 lines

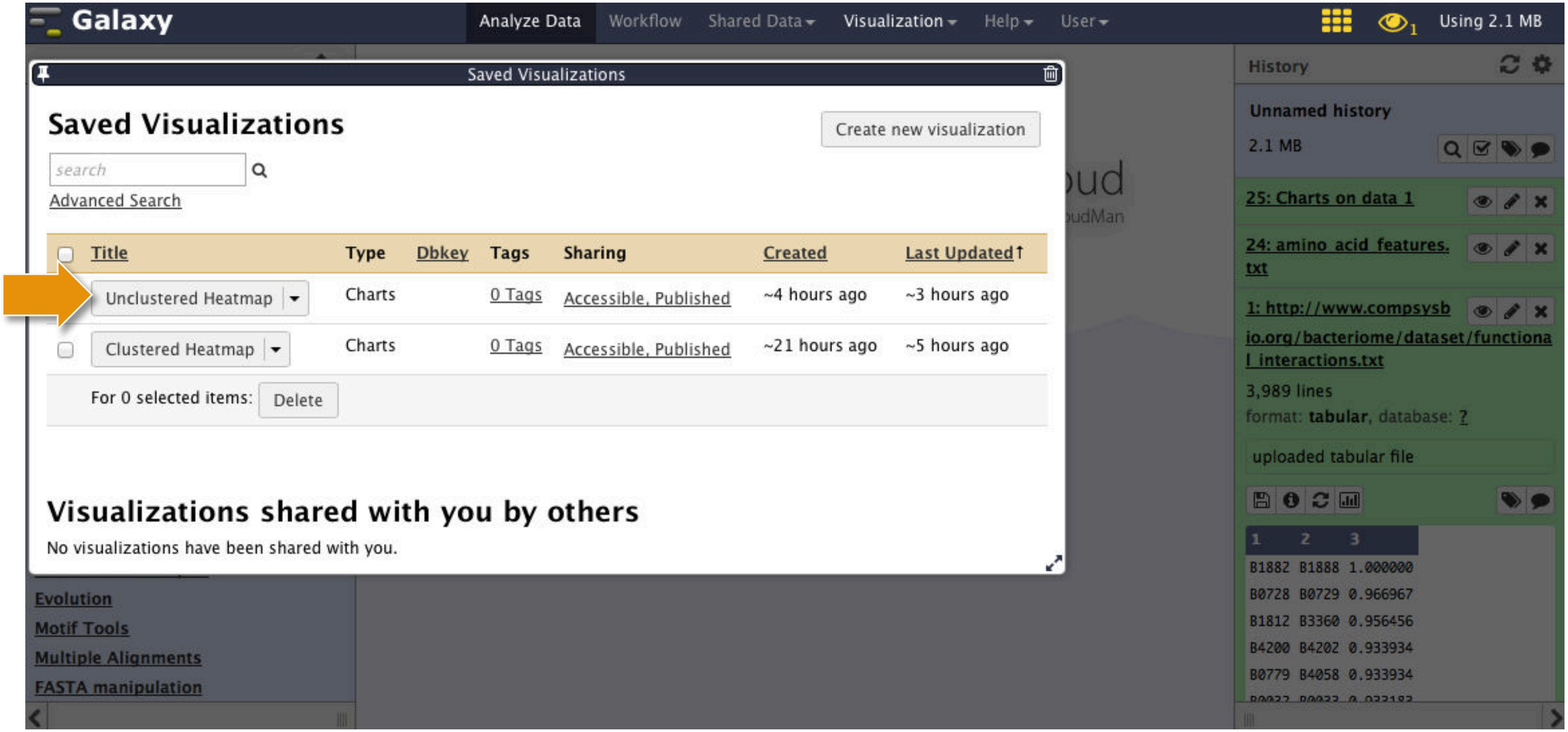
format: **tabular**, database: ?

uploaded tabular file

1	2	3
B1882	B1888	1.000000
B0728	B0729	0.966967
B1812	B3360	0.956456
B4200	B4202	0.933934
B0779	B4058	0.933934

Click on **Saved Visualizations**.

Activate the Scratchbook



Saved Visualizations

Create new visualization

search

Advanced Search

☐	Title	Type	Dbkey	Tags	Sharing	Created	Last Updated↑
☐	Unclustered Heatmap	Charts		0 Tags	Accessible, Published	~4 hours ago	~3 hours ago
☐	Clustered Heatmap	Charts		0 Tags	Accessible, Published	~21 hours ago	~5 hours ago

For 0 selected items: Delete

Visualizations shared with you by others

No visualizations have been shared with you.

Unnamed history

2.1 MB

25: Charts on data 1

24: amino acid features.txt

1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

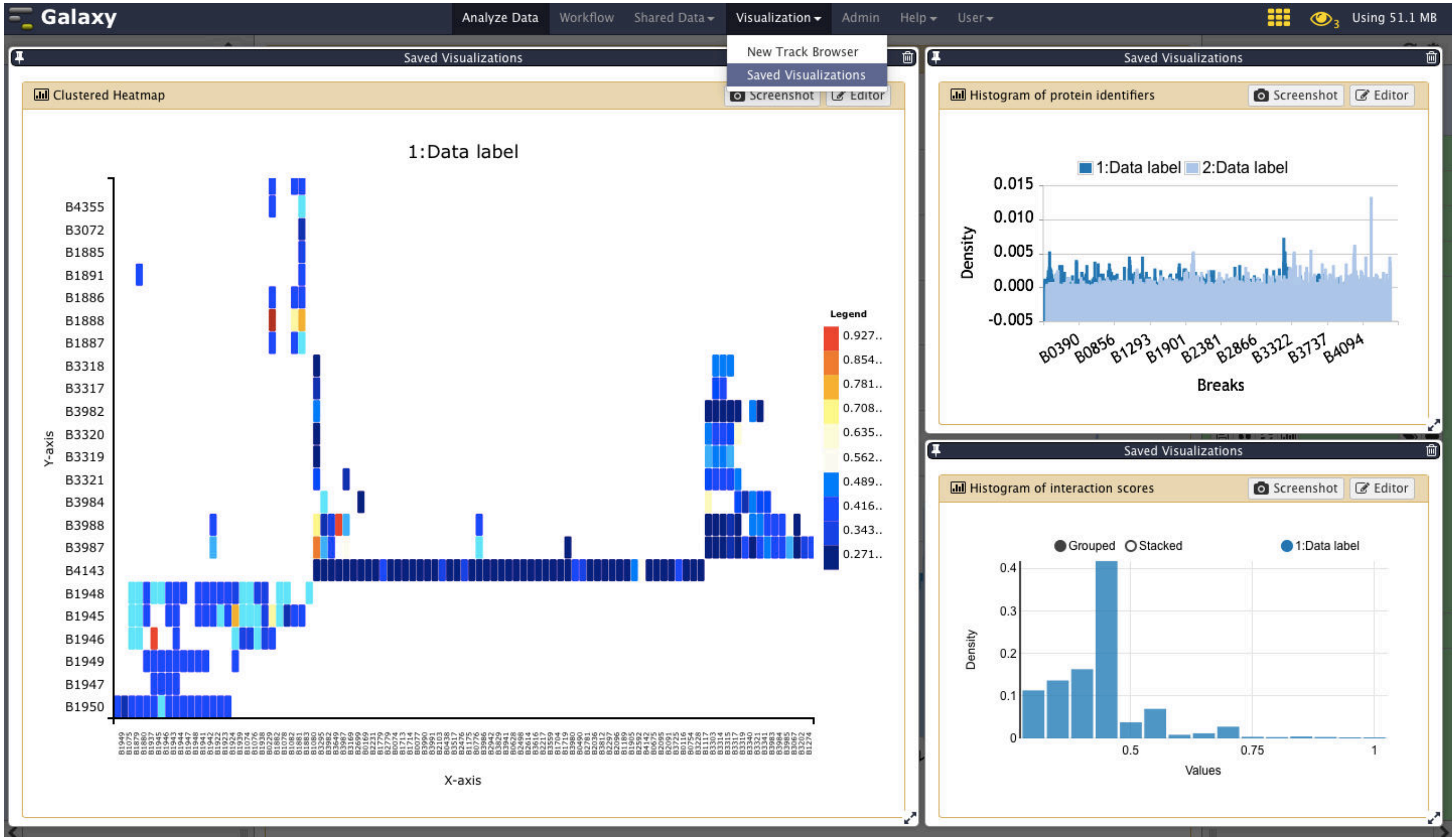
format: **tabular**, database: ?

uploaded tabular file

```
1     2     3
B1882 B1888 1.000000
B0728 B0729 0.966967
B1812 B3360 0.956456
B4200 B4202 0.933934
B0779 B4058 0.933934
```

Select a Visualization and repeat the process by selecting **Saved Visualizations** again.

Scratchbook for multiple charts



Resize all visualizations so they fit into the screen.

Summary

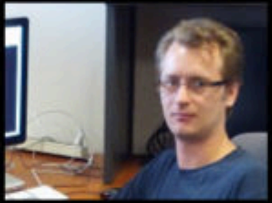
- Galaxy Charts supports **>16 visualization types**
- Integrating **custom visualizations** is easy
- Our **goal** is to handle >10k data points

Thank you.


Enis Afgan


Dannon Baker

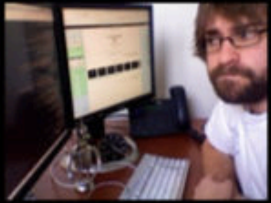

Dan Blankenberg


Dave Bouvier


Martin Čech


John Chilton

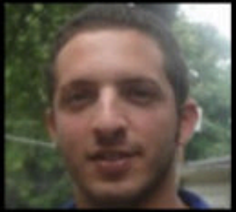

Dave Clements


Nate Coraor


Carl Eberhard


Dorine Francheteau


Jeremy Goecks


Aysam Guerler


Jen Jackson


Greg Von Kuster


Ross Lazarus


Nick Stoler


Anton Nekrutenko


James Taylor