

# Interactive Visual Analysis with Galaxy Charts

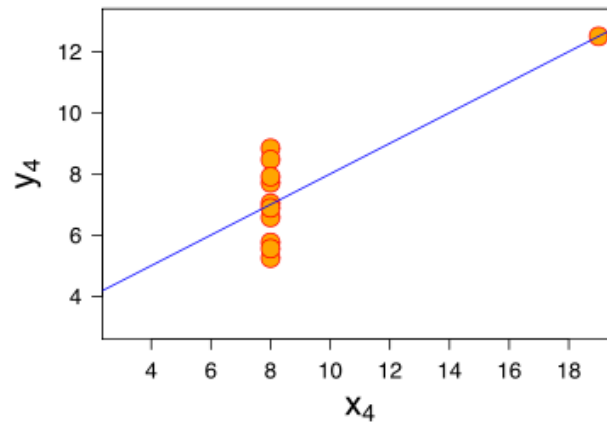
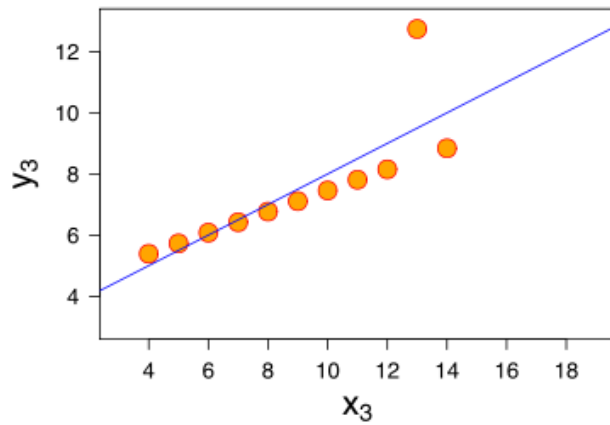
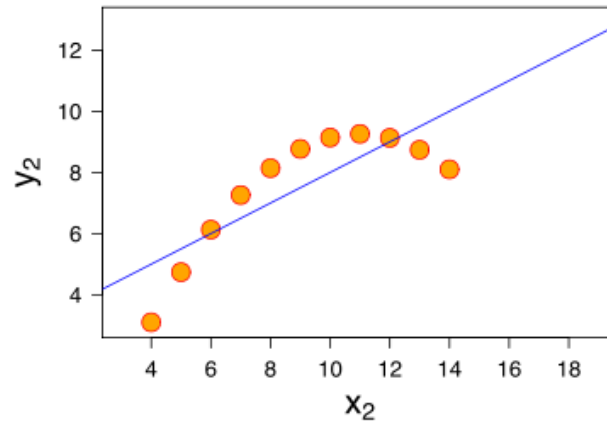
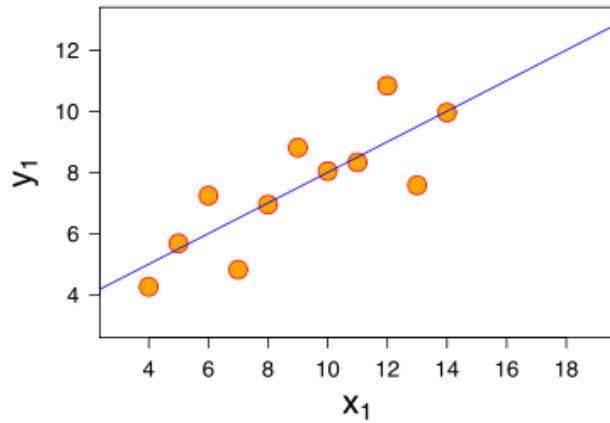
Aysam Guerler

Galaxy Team – Taylor Lab, 2014

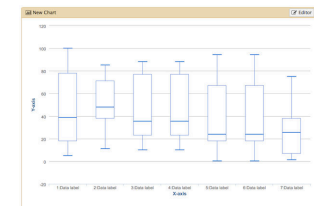
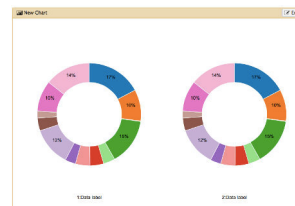
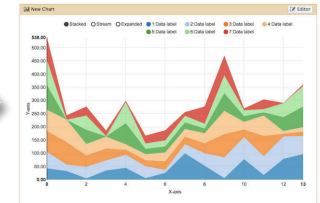
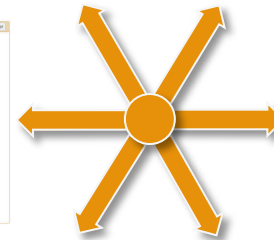
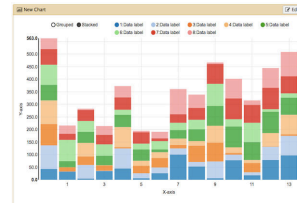
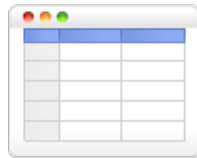
# Why Visualize?

- 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)

1 → S1: [http://www.compsysbio.org/bacteriome/dataset/functional\\_interactions.txt](http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt) [eye] [pencil] [x]

[sbio.org/bacteriome/dataset/functional\\_interactions.txt](http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt)

3,989 lines  
format: **tabular**, database: ?

uploaded tabular file

[save] [info] [refresh] [chart] ← 2 [share] [comment]

3 → Charts

- Scatterplot
- Trackster

|       |       |          |
|-------|-------|----------|
| 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 Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500)  Some (<10k)  Many (>10k)

• Bar diagrams

Regular (NVD3)  Stacked (NVD3)  Horizontal (NVD3)  Stacked horizontal (NVD3)

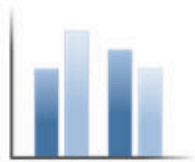
• Others

Name your chart **Unclustered Heatmap**.

# Select a chart type

Few (<500)    Some (<10k)    Many (>10k)

## • Bar diagrams



Regular (NVD3)



Stacked (NVD3)

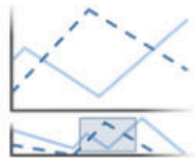


Horizontal (NVD3)



Stacked horizontal (NVD3)

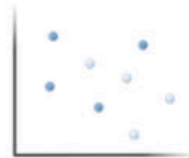
## • Others



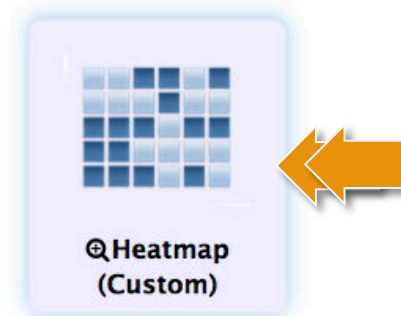
Line with focus (NVD3)



Line chart (NVD3)



Scatter plot (NVD3)



Heatmap (Custom)

Double click on the **Heatmap** icon.

# Select data columns

The screenshot shows a 'New Chart' configuration window with a title bar containing a bar chart icon, the text 'New Chart', and buttons for 'Cancel' and 'Draw'. Below the title bar, there are tabs for 'Start', 'Configuration', and '1: Data label', along with an 'Add Data' button. The 'Provide a label:' section has a text input field containing 'Data label'. The 'Select columns:' section contains three rows, each with a label and a dropdown menu:

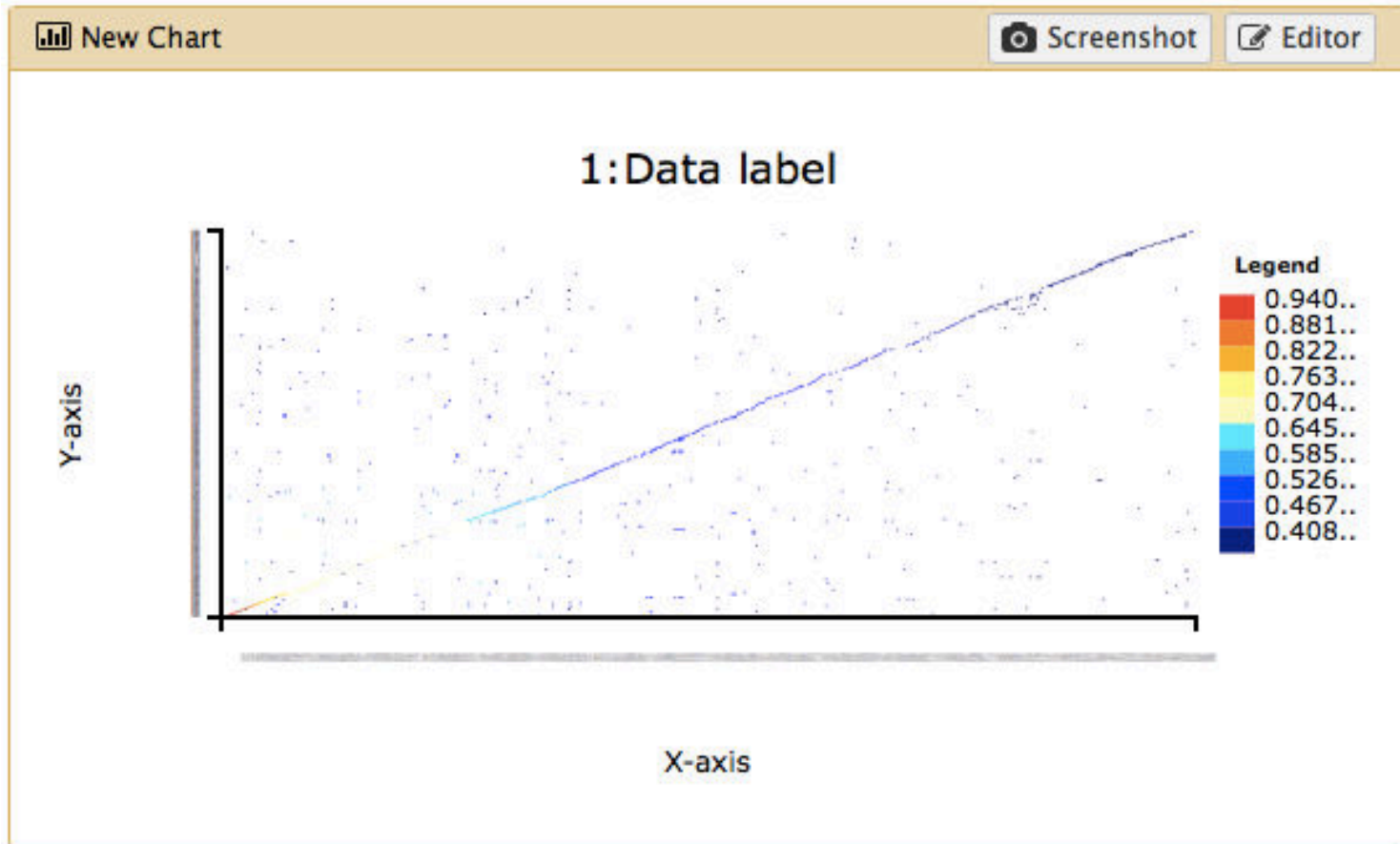
| Label                           | Selected Column   |
|---------------------------------|-------------------|
| Column labels (all data labels) | Column: 1 [str]   |
| Row labels (all data labels)    | Column: 2 [str]   |
| Observation                     | Column: 3 [float] |

Orange arrows with numbers 1 and 2 point to the 'Row labels (all data labels)' dropdown and the 'Draw' button, respectively.

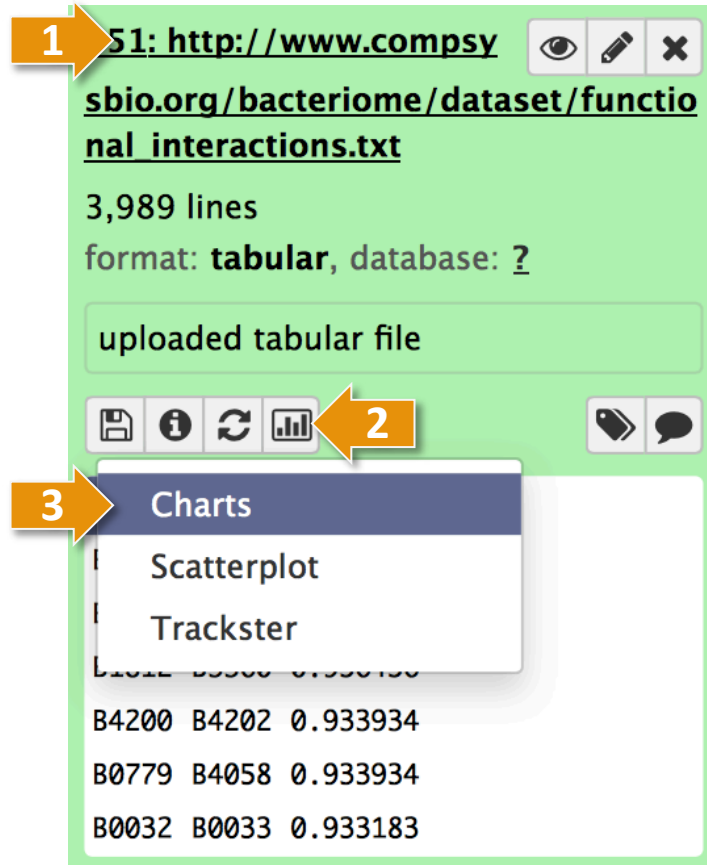
At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.



# Unclustered Heatmap



# Make a new chart (2 of 4)



1 → S1: [http://www.compsysbio.org/bacteriome/dataset/functional\\_interactions.txt](http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt)

3,989 lines  
format: **tabular**, database: ?

uploaded tabular file

2 → [Visualization Icon]

3 → Charts

- Scatterplot
- Trackster

|       |       |          |
|-------|-------|----------|
| B4200 | B4202 | 0.933934 |
| B0779 | B4058 | 0.933934 |
| B0032 | B0033 | 0.933183 |

The screenshot shows a web interface for a dataset. At the top, a green header contains the dataset name and a file icon. Below it, the dataset details are shown: '3,989 lines' and 'format: tabular, database: ?'. A text box below that says 'uploaded tabular file'. A toolbar with icons for save, info, refresh, and visualization is present. An orange arrow labeled '2' points to the visualization icon. A dropdown menu is open, showing 'Charts' selected, with 'Scatterplot' and 'Trackster' as options. An orange arrow labeled '3' points to the 'Charts' option. Below the menu, a table of data is visible, showing columns of IDs and values.

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

# Give your chart a name

Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500)  Some (<10k)  Many (>10k)

• Bar diagrams

Regular (NVD3)  Stacked (NVD3)  Horizontal (NVD3)  Stacked horizontal (NVD3)

• Others

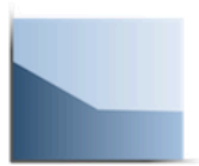
Name your chart **Clustered Heatmap**.

# 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

**New Chart** Cancel Draw

Start Configuration **1: Data label** Add Data

**Provide a label:**

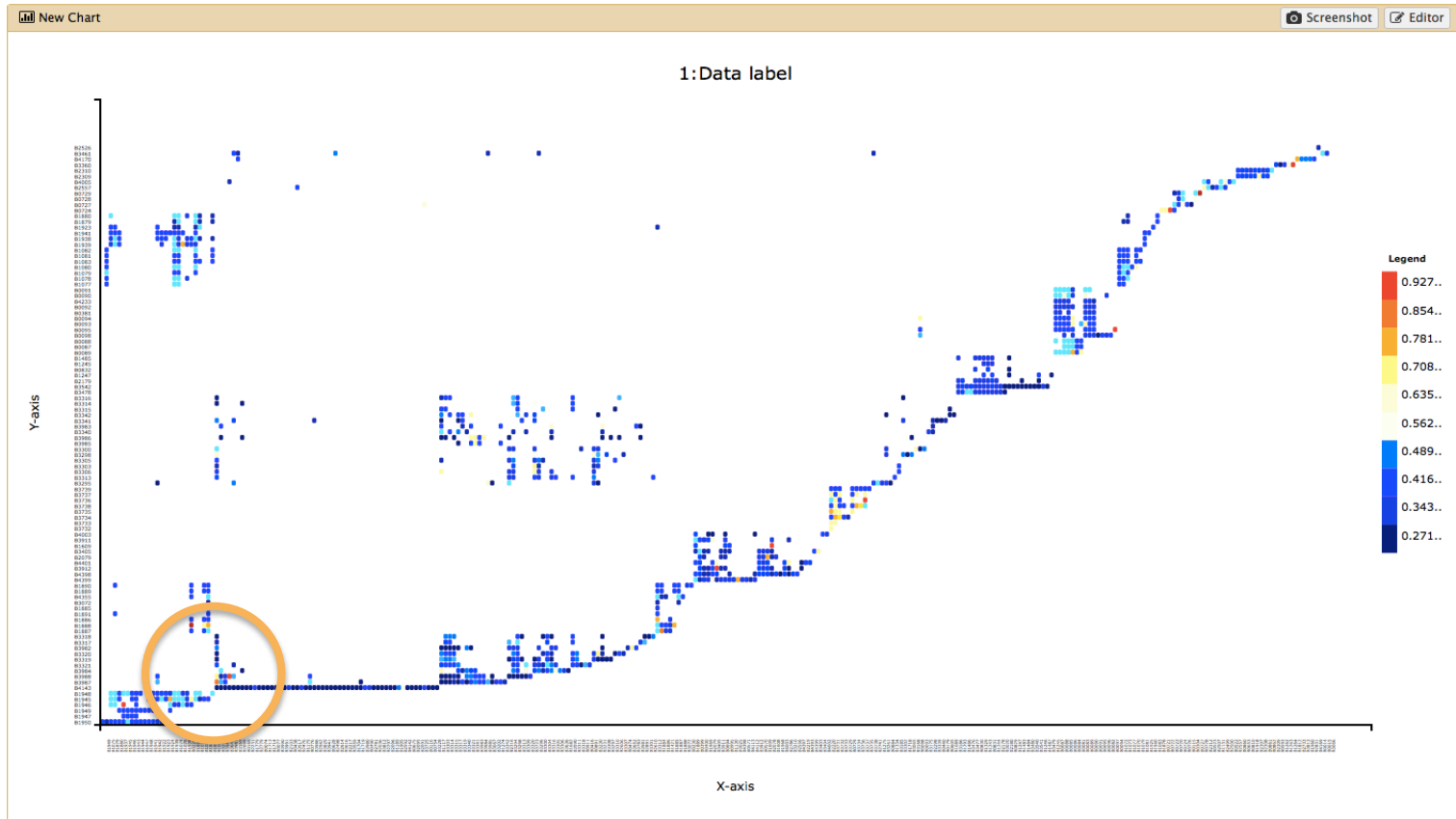
Data label

**Select columns:**

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

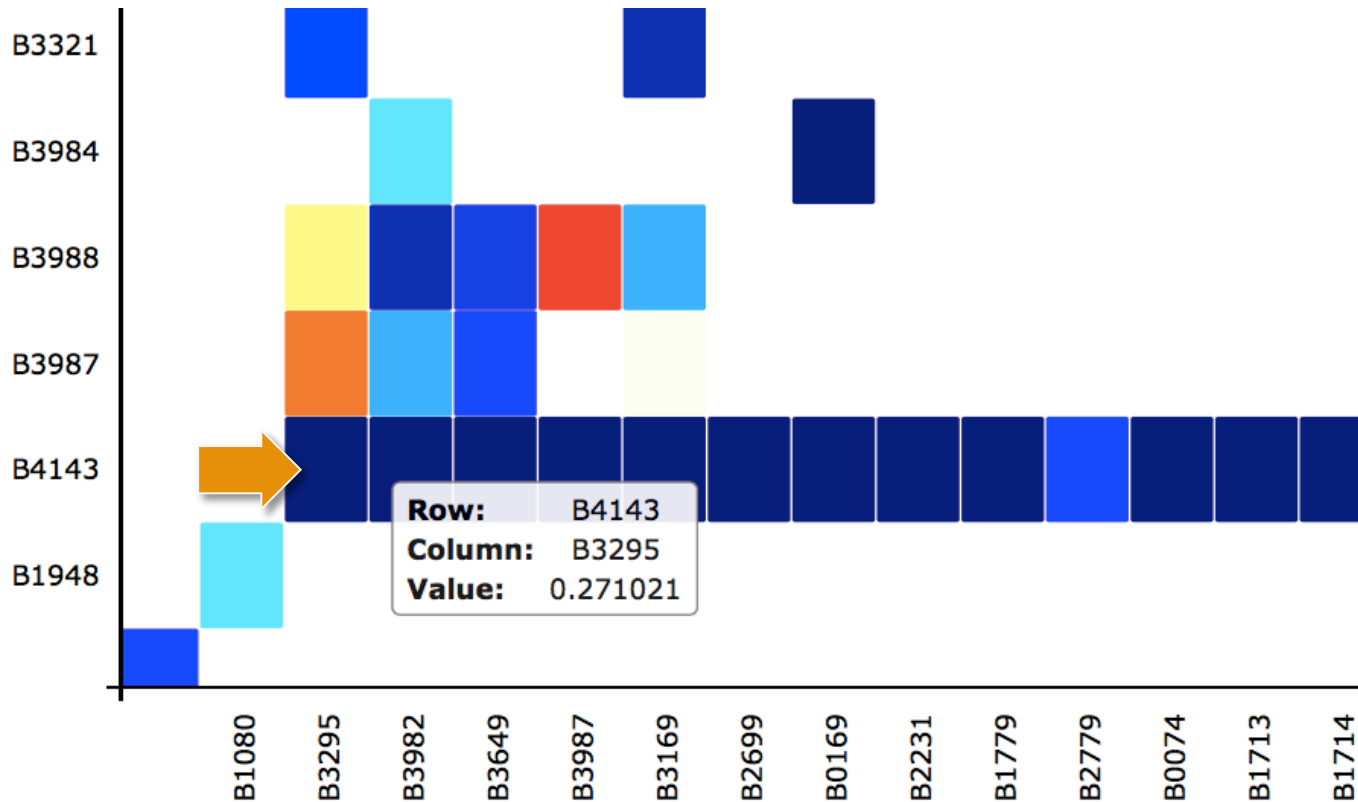
At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

# Clustered Heatmap




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

# Enlarged view



**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

 New Chart

**Start** **Configuration** **1: Data label** **+** Add Data


**Provide a chart title:**

New Chart

**How many data points would you like to analyze?**

Few (<500) Some (<10k) Many (>10k)

• **Bar diagrams**



Go to the **Configuration** tab.



# Chart settings

## X axis:

Axis label

Provide a label for the axis.

Axis value type

Select the value type of the axis.

## Y axis:

Axis label

Provide a label for the axis.

Axis value type

Select the value type of the axis.

## Others:

Show legend

 Yes  No

Would you like to add a legend?

Color scheme

Select a color scheme for your heatmap

Url template

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

Yes No

Would you like to add a legend?

Jet

Select a color scheme for your heatmap

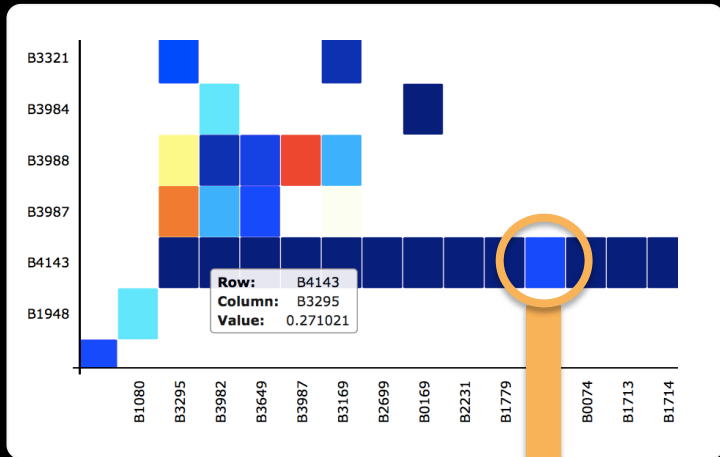


`http://www.ncbi.nlm.nih.gov/geoprofiles/?term=__LABEL__`

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



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

NCBI Resources How To Sign in to NCBI

GEO Profiles

Search GEO Profiles b4143 Save search Advanced Help

Show additional filters

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 >>

Profile data Download profile data

Profile pathways Find pathways

Find related data Database: Select Find items

Search details b4143[All Fields]

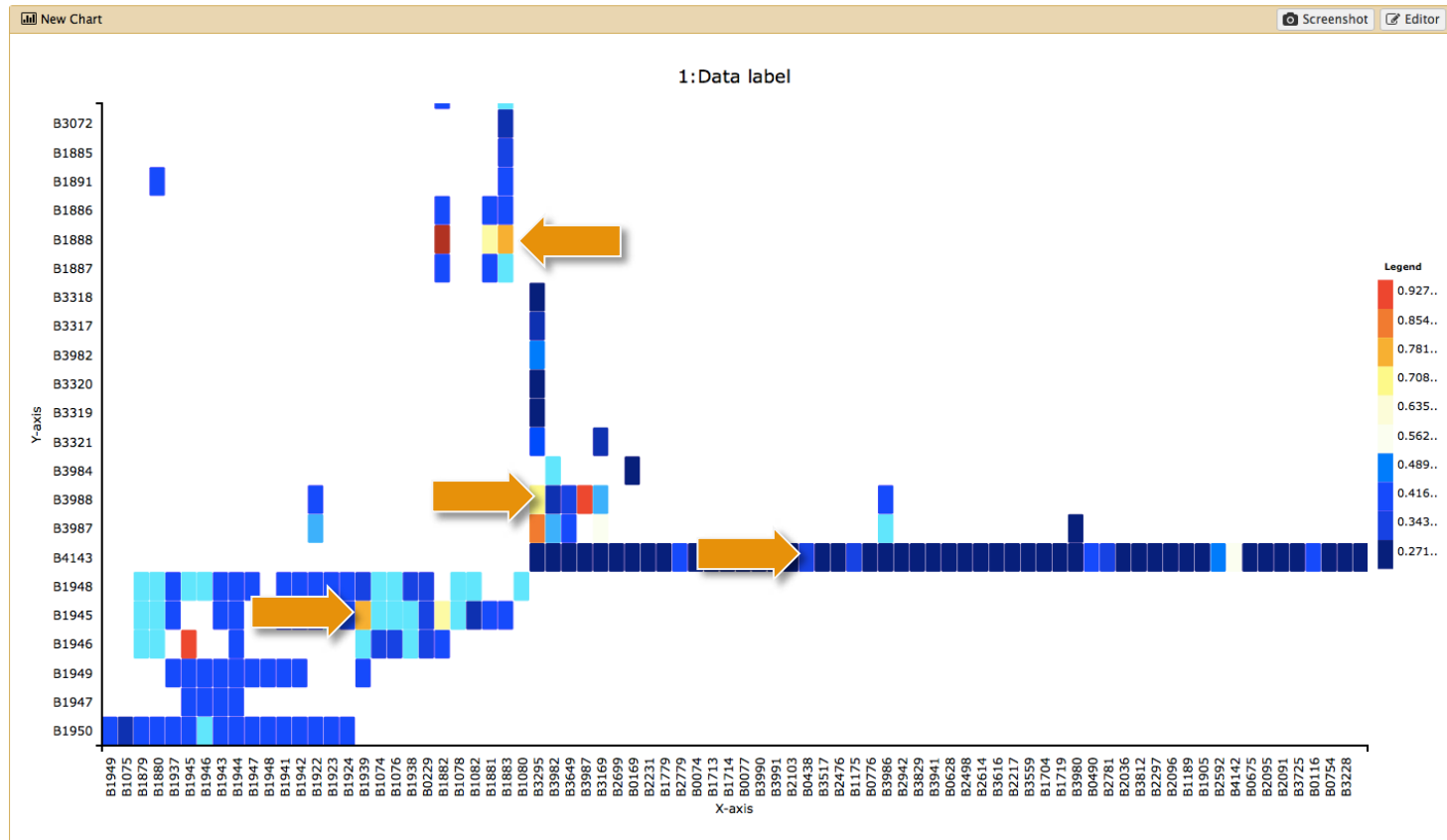
1. [groEL - Stress factor RpoS regulon in exponential-phase bacteria](#)

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](#)

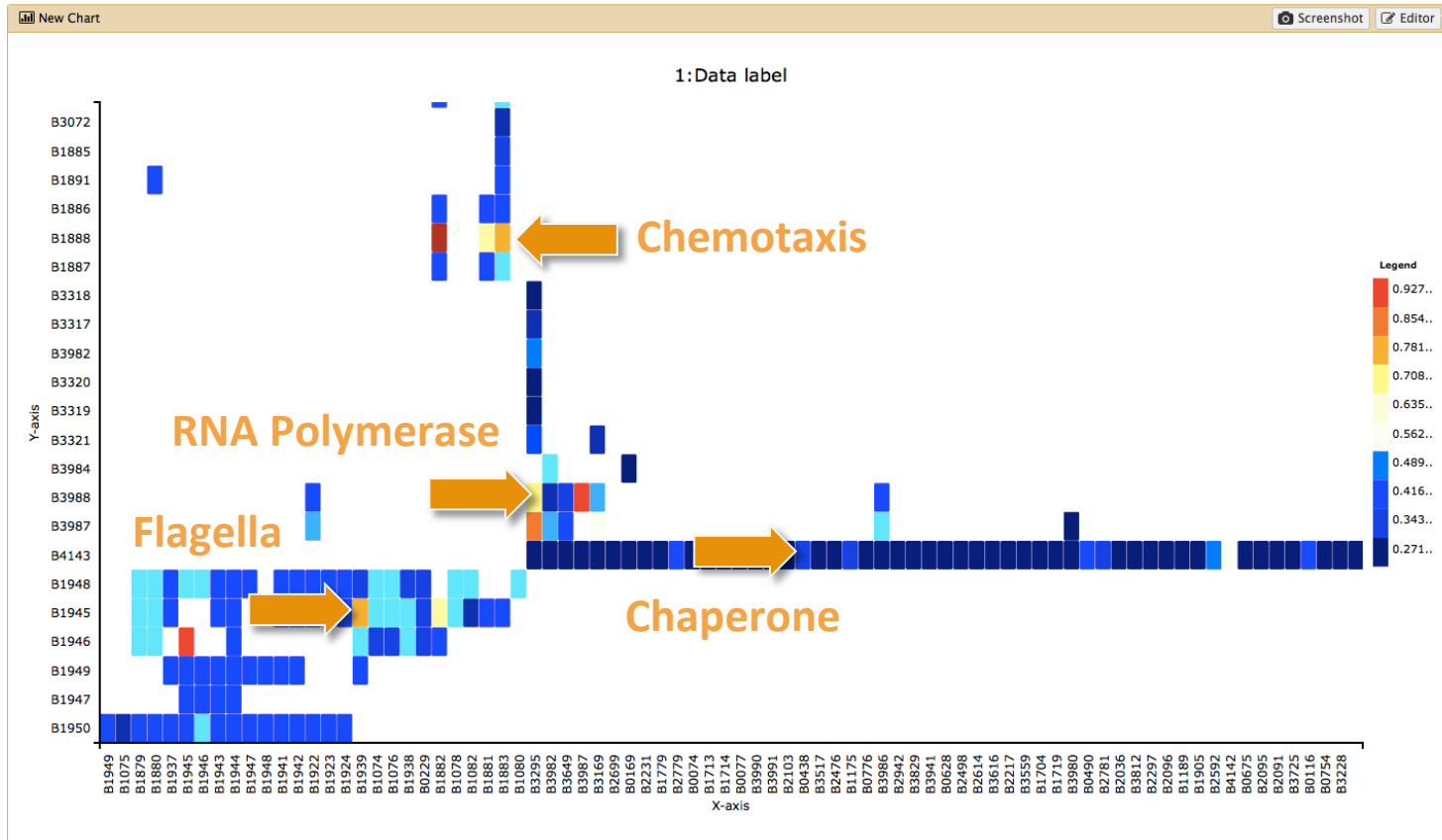
Annotation: groL, Cpn60 chaperonin GroEL, large subunit of GroESL (multiple annotations exist)  
Organism: Escherichia coli, Escherichia coli K-12  
Reporter: GPL189, 1240 (ID\_REF), GDS2181, **b4143** (ORF)  
DataSet type: Expression profiling by array, count, 12 samples  
ID: 27346540  
[GEO DataSets](#) [Gene](#) [Profile neighbors](#) [Chromosome neighbors](#)

# Cluster selection and analysis



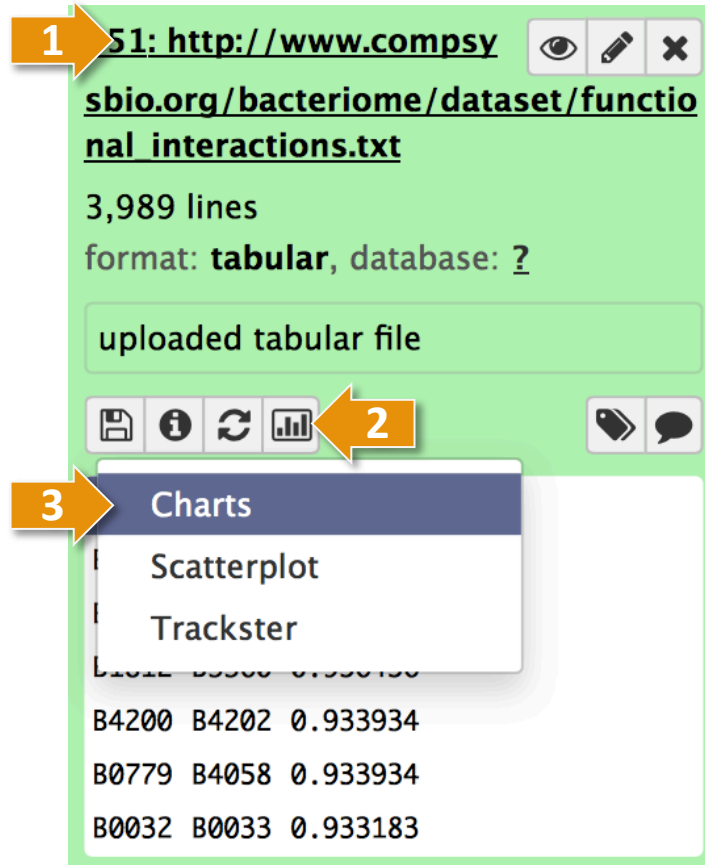
Select one element from each **highlighted row**. What are the corresponding **protein functions**?

# Identified protein categories



Please return to the **Editor**.

# Make a new chart (3 of 4)



1 → S1: [http://www.compsysbio.org/bacteriome/dataset/functional\\_interactions.txt](http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt)

3,989 lines  
format: **tabular**, database: ?

uploaded tabular file

2 → [Visualization Icon]

3 → Charts

- Scatterplot
- Trackster

|       |       |          |
|-------|-------|----------|
| B4200 | B4202 | 0.933934 |
| B0779 | B4058 | 0.933934 |
| B0032 | B0033 | 0.933183 |

The screenshot shows a web interface for a dataset. At the top, a green header bar contains the dataset name and a URL, with an orange arrow labeled '1' pointing to it. Below the header, the dataset is identified as '3,989 lines' and 'format: tabular, database: ?'. A text box below contains 'uploaded tabular file'. A toolbar with icons for save, info, refresh, and visualization is shown, with an orange arrow labeled '2' pointing to the visualization icon. A dropdown menu is open, with an orange arrow labeled '3' pointing to the 'Charts' option. The menu lists 'Charts', 'Scatterplot', and 'Trackster'. Below the menu, a table of data is visible, showing columns of IDs and values.

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

# Give your chart a name

Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500)  Some (<10k)  Many (>10k)

• Bar diagrams

Regular (NVD3)  Stacked (NVD3)  Horizontal (NVD3)  Stacked horizontal (NVD3)

• Others

Name your chart **Score Histogram**.

# Analyze the score distribution

- Area charts



Regular (NVD3)



Expanded (NVD3)

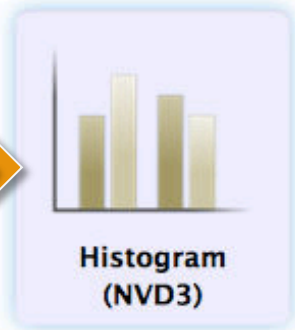


Stream (NVD3)



Pie chart (NVD3)

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



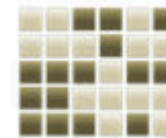
Histogram (NVD3)



Discrete Histogram (jqPlot)



Box plot (jqPlot)




Clustered Heatmap (Custom)

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



# Give your chart a name

Unclustered Heatmap  Draw

Please select data columns before drawing the chart.

[Start](#) [Configuration](#) **1: Data label** [Add Data](#)

**Provide a label:**

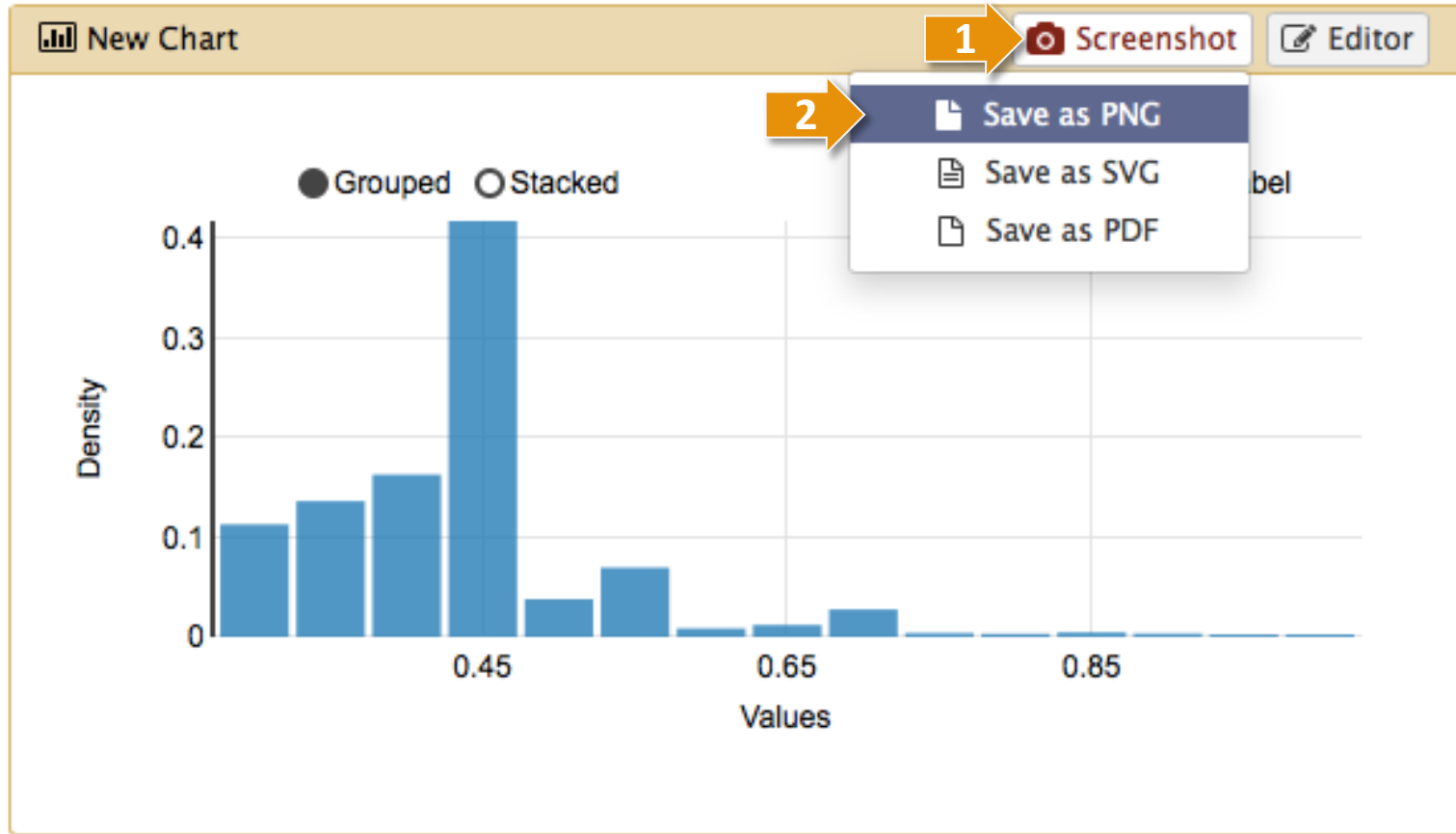
Data label

**Select columns:**

Observations Column: 3 [float]

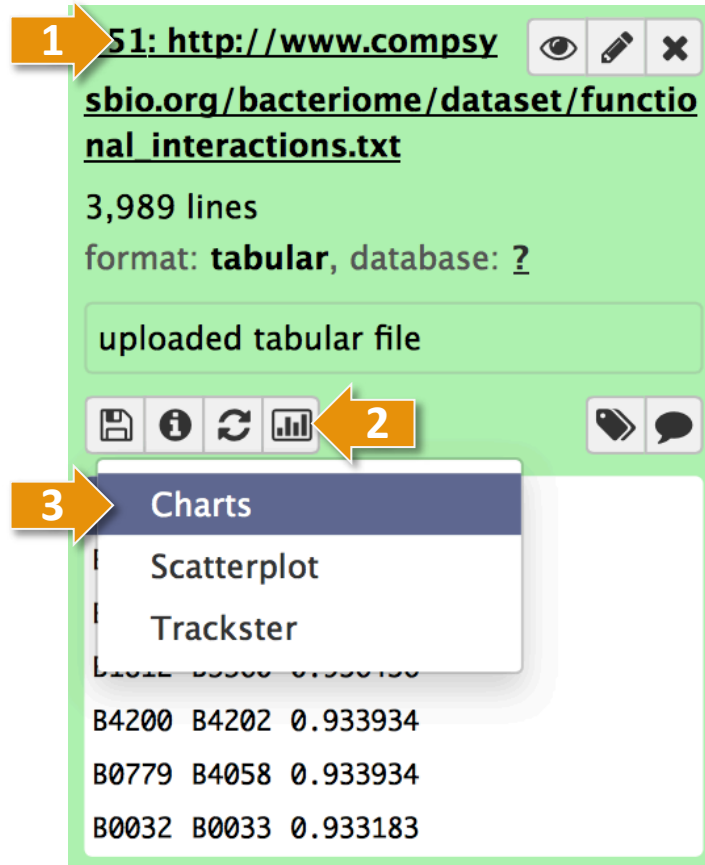
Click on **Draw**.

# 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 → S1: [http://www.compsysbio.org/bacteriome/dataset/functional\\_interactions.txt](http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt)

3,989 lines  
format: **tabular**, database: ?

uploaded tabular file

2 → [Visualization Icon]

3 → Charts

- Scatterplot
- Trackster

|       |       |          |
|-------|-------|----------|
| B4200 | B4202 | 0.933934 |
| B0779 | B4058 | 0.933934 |
| B0032 | B0033 | 0.933183 |

The screenshot shows a web interface for a dataset. At the top, a green header bar contains the dataset name and a URL, with an orange arrow labeled '1' pointing to it. Below the header, the dataset is identified as '3,989 lines' and 'format: tabular, database: ?'. A text box below contains 'uploaded tabular file'. A toolbar with icons for save, info, refresh, and visualization is shown, with an orange arrow labeled '2' pointing to the visualization icon. A dropdown menu is open, with an orange arrow labeled '3' pointing to the 'Charts' option. The menu lists 'Charts', 'Scatterplot', and 'Trackster'. Below the menu, a table of data is visible, showing columns of IDs and values.

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

# Give your chart a name

Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500)  Some (<10k)  Many (>10k)

• Bar diagrams

Regular (NVD3)  Stacked (NVD3)  Horizontal (NVD3)  Stacked horizontal (NVD3)

• Others

Name your chart **Discrete Histogram**.

# Analyze the protein distribution

- Area charts



ⓄRegular (NVD3)



ⓄExpanded (NVD3)



ⓄStream (NVD3)



Pie chart (NVD3)

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



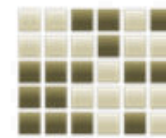
Histogram (NVD3)



ⓄDiscrete Histogram (jqPlot)



ⓄBox plot (jqPlot)



ⓄClustered Heatmap (Custom)

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

# Add more data

Unclustered Heatmap Draw

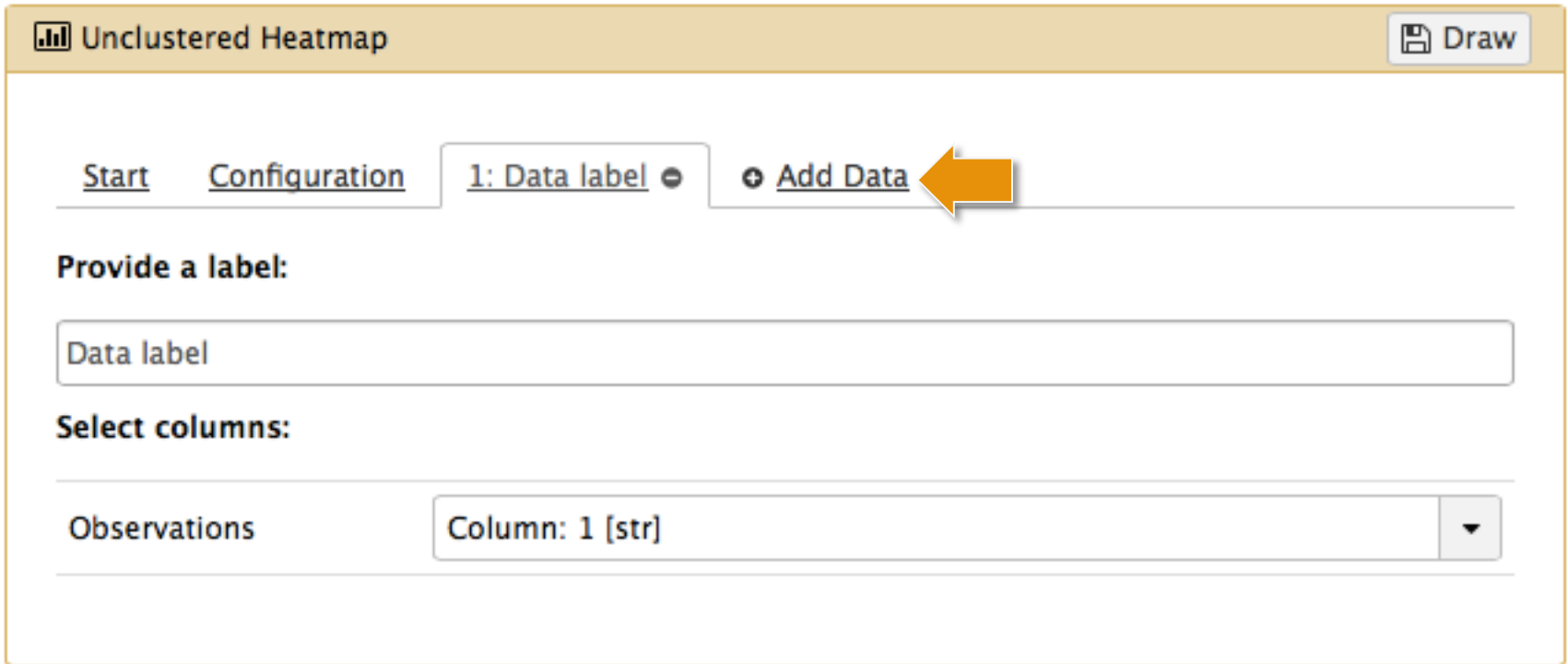
[Start](#) [Configuration](#) [1: Data label](#) [+ Add Data](#)

**Provide a label:**

Data label

**Select columns:**

Observations [Column: 1 \[str\]](#)

The image shows a software interface for an 'Unclustered Heatmap'. At the top, there is a title bar with a heatmap icon and the text 'Unclustered Heatmap', and a 'Draw' button on the right. Below the title bar, there are four tabs: 'Start', 'Configuration', '1: Data label', and '+ Add Data'. An orange arrow points to the '+ Add Data' tab. Under the '1: Data label' tab, there is a section titled 'Provide a label:' with a text input field containing 'Data label'. Below that is a section titled 'Select columns:' with a dropdown menu showing 'Column: 1 [str]' and a small downward arrow.

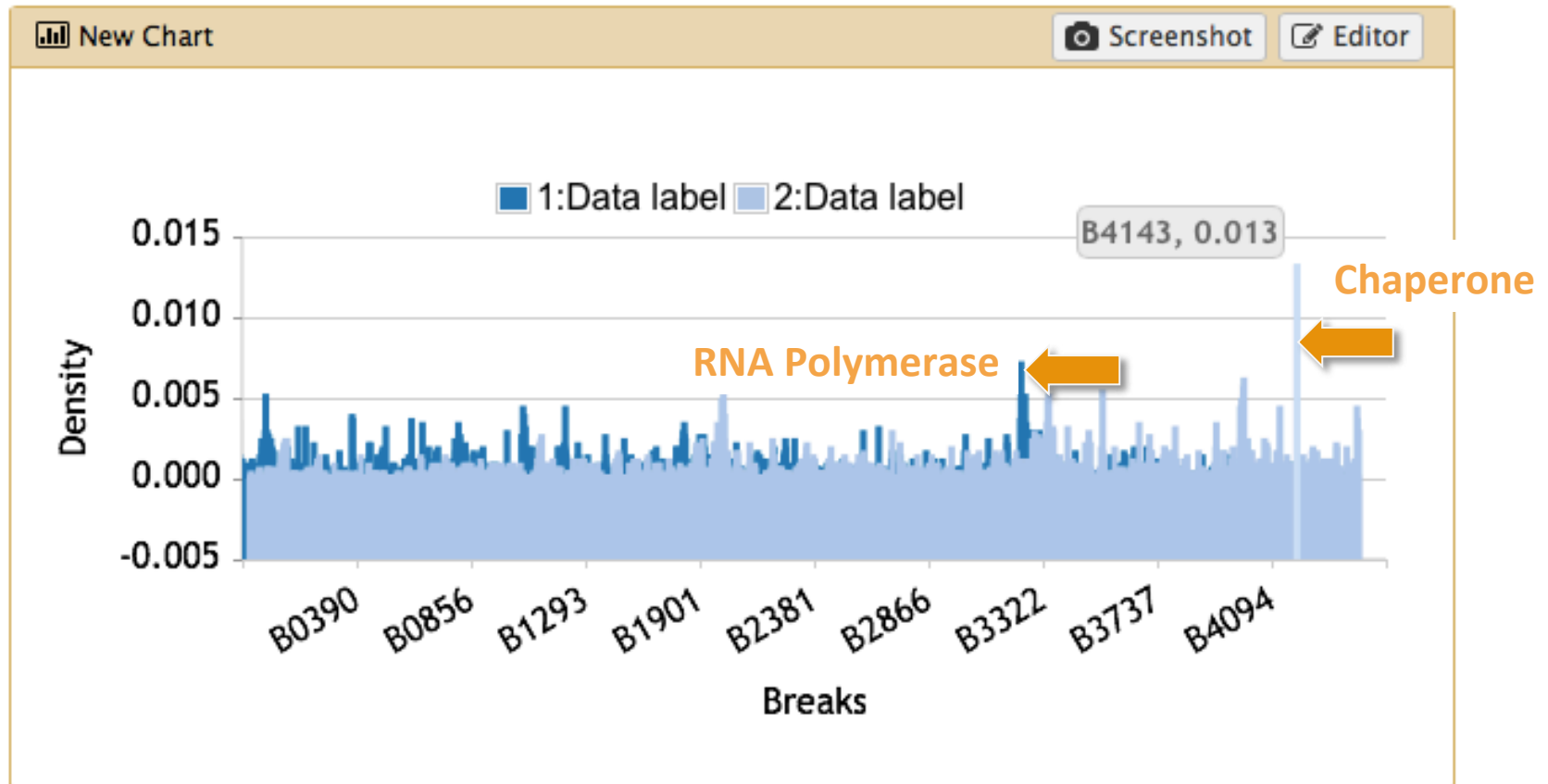
Click on **Add Data**.

# Select a second data group

The screenshot shows a 'New Chart' dialog box with a yellow header bar. The header contains a chart icon, the text 'New Chart', and two buttons: 'Cancel' and 'Draw'. The 'Draw' button is highlighted with an orange arrow labeled '2'. Below the header, there are four tabs: 'Start', 'Configuration', '1: Data label', and '2: Data label'. The '2: Data label' tab is active and highlighted. To the right of the tabs is an 'Add Data' button. Below the tabs, the text 'Provide a label:' is followed by a text input field containing 'Data label'. Below that, the text 'Select columns:' is followed by a list box. The list box has 'Observations' selected and highlighted in blue, with an orange arrow labeled '1' pointing to it. To the right of 'Observations' is a dropdown menu showing 'Column: 2 [str]'.

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

# Which proteins have most interactions?



Done with Part I.



# Scratchbook

# Activate the Scratchbook

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the right side of the navigation bar, there is a grid icon representing the Scratchbook, which is highlighted by an orange arrow. Below the navigation bar, the main content area displays 'Welcome to Galaxy on the Cloud managed by CloudMan'. On the left, there is a 'Tools' sidebar with a search bar and various tool categories. On the right, there is a 'History' sidebar showing a list of jobs, including '25: Charts on data 1', '24: amino acid features.txt', and '1: http://www.compsysb.io.org/bacteriome/dataset/functiona l interactions.txt'. The bottom of the history sidebar shows a table with columns 1, 2, and 3, and rows of data.

| 1     | 2     | 3        |
|-------|-------|----------|
| B1882 | B1888 | 1.000000 |
| B0728 | B0729 | 0.966967 |
| B1812 | B3360 | 0.956456 |
| B4200 | B4202 | 0.933934 |
| B0779 | B4058 | 0.933934 |
| B0022 | B0022 | 0.022182 |

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

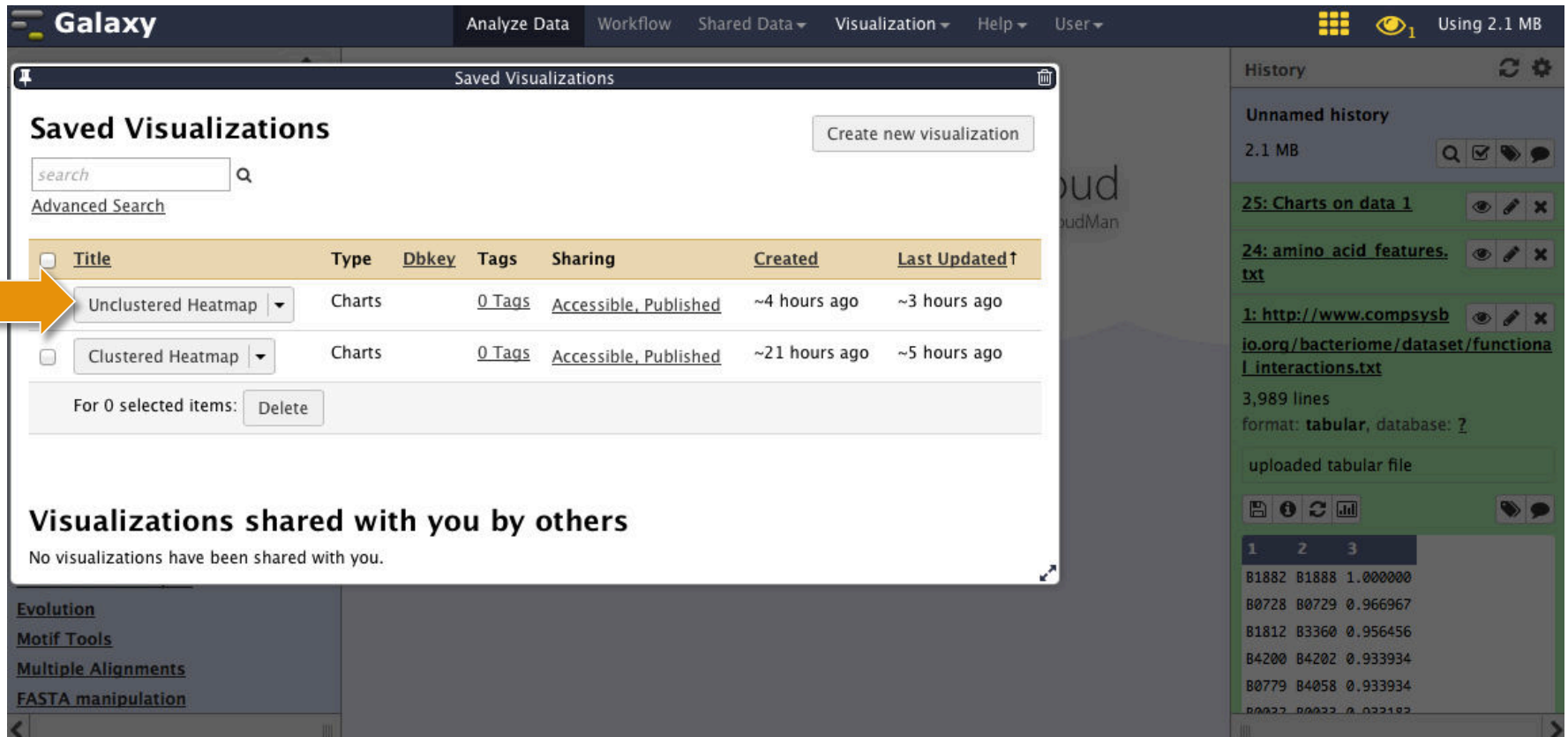
# Activate the Scratchbook

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Visualization' dropdown menu is open, showing 'New Track Browser' and 'Saved Visualizations', with an orange arrow pointing to the latter. The main content area displays 'Welcome to Galaxy on the Cloud' managed by CloudMan. The left sidebar contains a 'Tools' section with a search bar and various tool categories. The right sidebar shows a 'History' section with a list of recent jobs, including '25: Charts on data 1', '24: amino acid features.txt', and '1: http://www.compsysb.io.org/bacteriome/dataset/functiona l interactions.txt'. A table of data is visible at the bottom of the history entry.

| 1     | 2     | 3        |
|-------|-------|----------|
| B1882 | B1888 | 1.000000 |
| B0728 | B0729 | 0.966967 |
| B1812 | B3360 | 0.956456 |
| B4200 | B4202 | 0.933934 |
| B0779 | B4058 | 0.933934 |
| B0022 | B0023 | 0.022182 |

Click on **Saved Visualizations**.

# Activate the Scratchbook

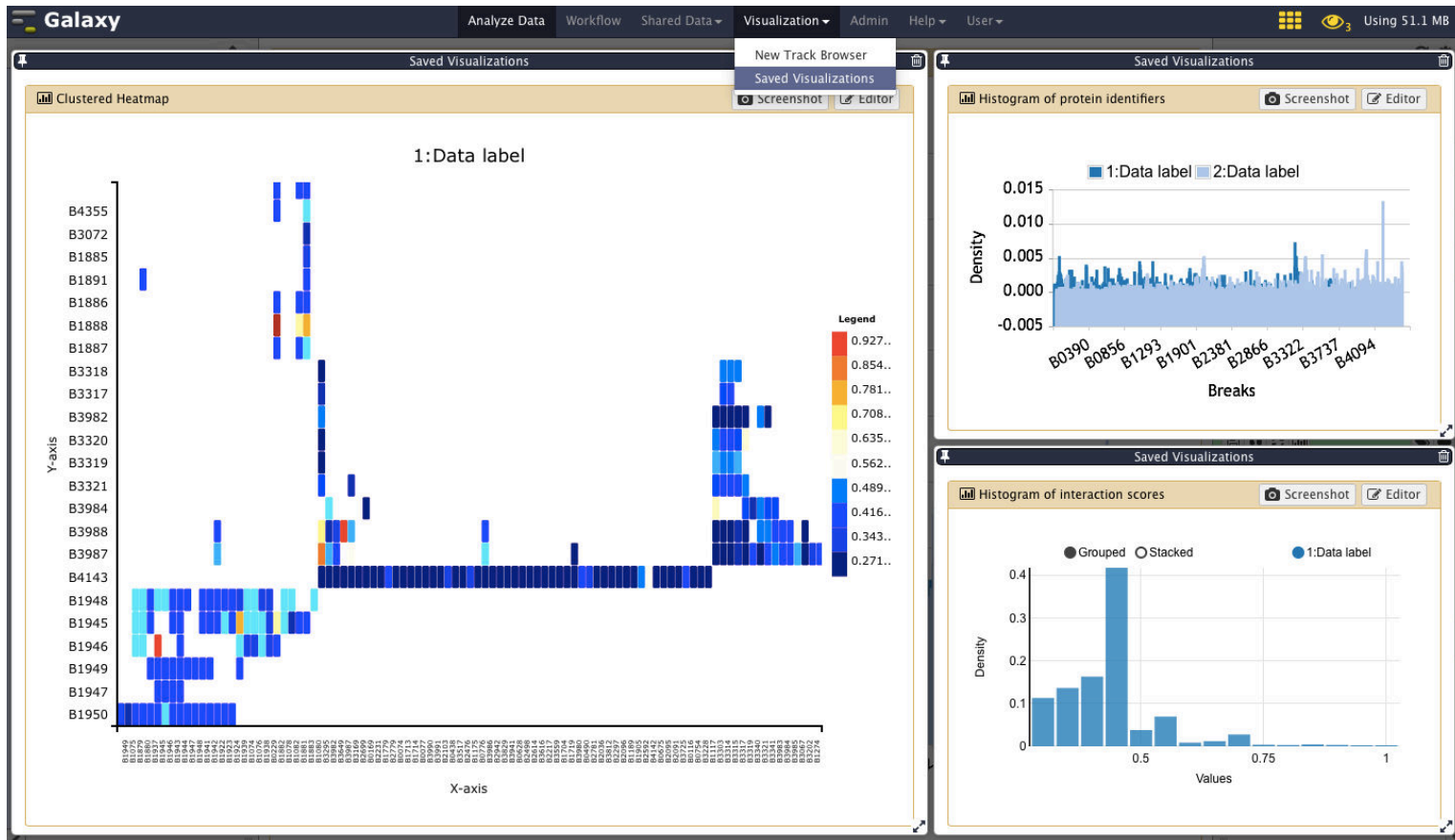


The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Saved Visualizations' panel is open, displaying a table of saved visualizations. An orange arrow points to the 'Unclustered Heatmap' visualization. The table has columns for Title, Type, Dbkey, Tags, Sharing, Created, and Last Updated. Below the table, there is a 'Delete' button for 0 selected items. The right sidebar shows the 'History' panel with a list of recent jobs, including '25: Charts on data 1', '24: amino acid features.txt', and '1: http://www.compsysb.io.org/bacteriome/dataset/functiona l interactions.txt'. The bottom left sidebar shows 'Evolution', 'Motif Tools', 'Multiple Alignments', and 'FASTA manipulation'.

| 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 |

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