

DataViz

User Manual

Version: 08.11.2022

!! download the latest version at:
<https://github.com/barouxlab/DataViz>



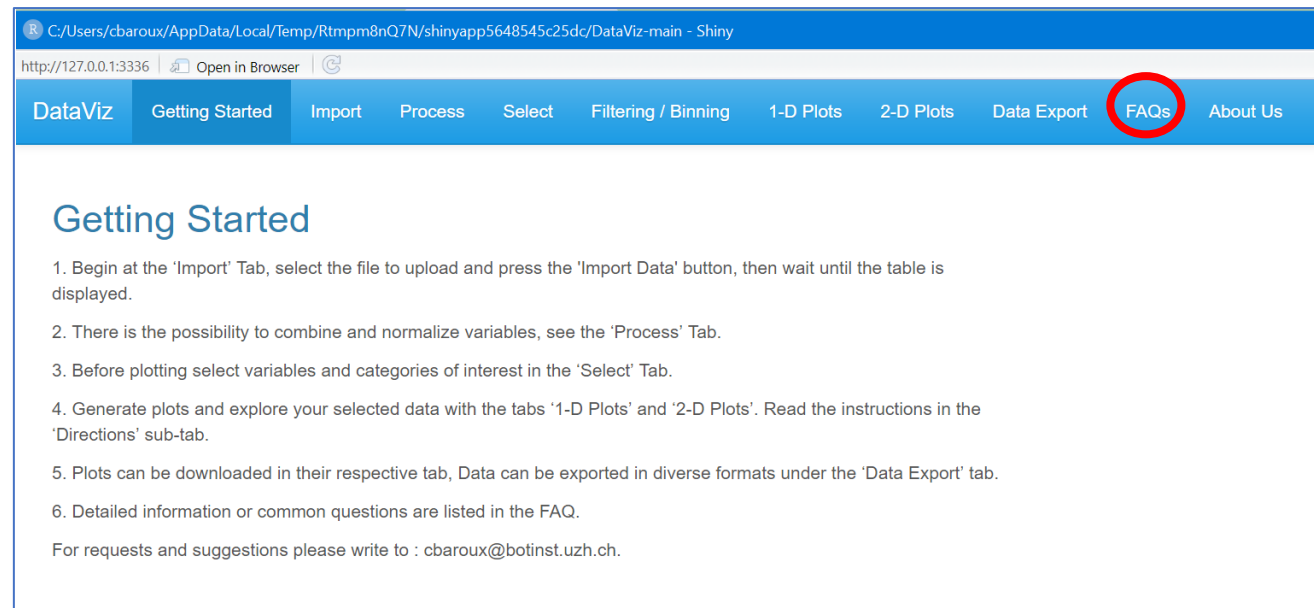
**Universität
Zürich^{UZH}**

Célia Baroux, IPMB
Devin Routh, S3IT

User manual – ‘How to ‘ ?

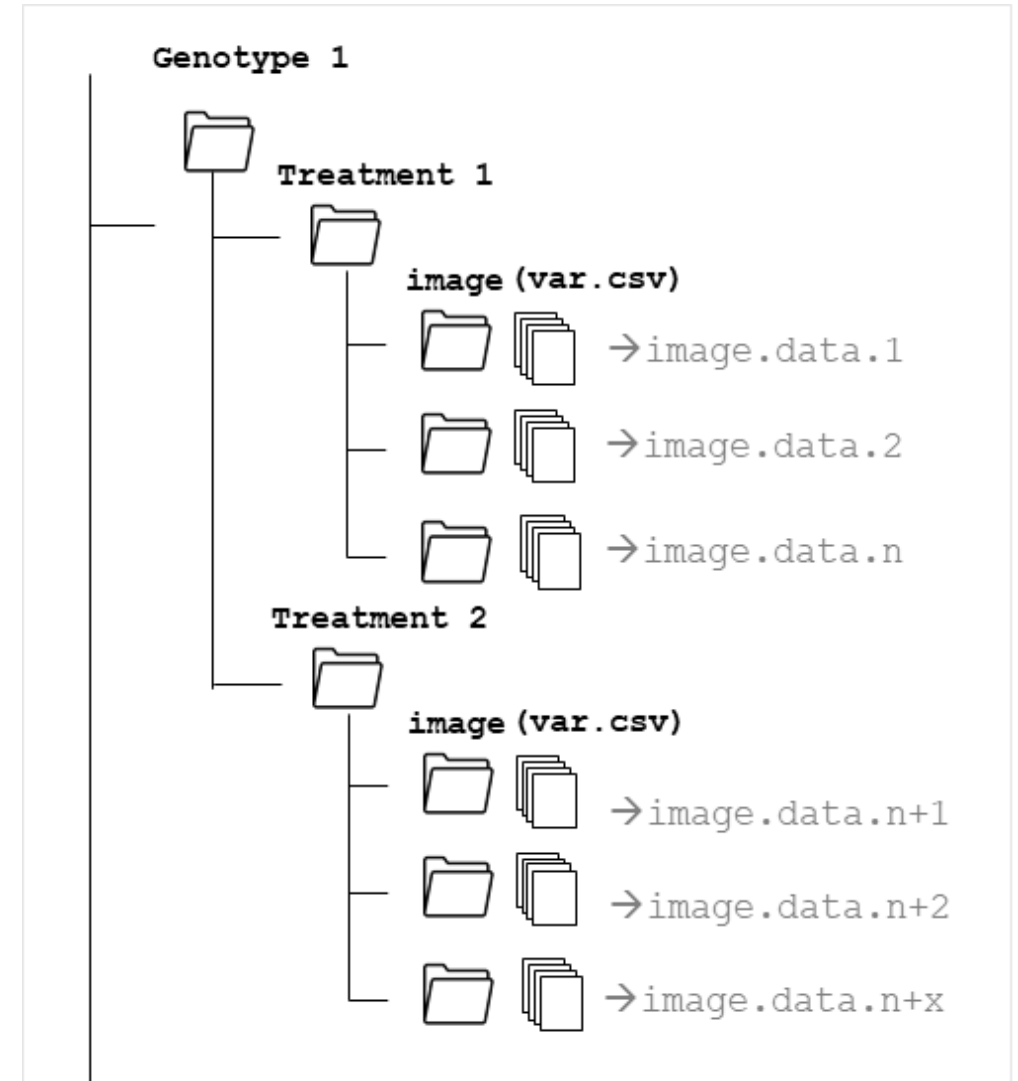


- Browse through specific topic using the ‘Outline View’
- Check for latest update of the user manual on <https://github.com/barouxlab/DataViz>
- Check the FAQ in DataViz for troubleshooting



Data preparation

- **Using data exported from Imaris**
 - DataViz was tested with data exported from Imaris 9.8 and 9.9
 - Image Data must contain a surface called 'Nucleus' as this is a reference object in several calculations
 - Image Data must be organized as follow:
Genotype > Treatment > Image
 - If several genotypes, the treatment folders must be duplicated with identical names
 - The root folder containing the data is converted to a .zip file



Data preparation

- **Using data exported from a different software**
 - In principle DataViz can handle any data organized according to the sample dataframe uploaded on [github/barouxlab/DataViz](https://github.com/barouxlab/DataViz): make sure to organize the data according to this template
 - The requirement of an object called 'Nucleus' is also valid for normalisations. If no normalization necessary, then DataViz will plot the raw data as labelled in the dataframe

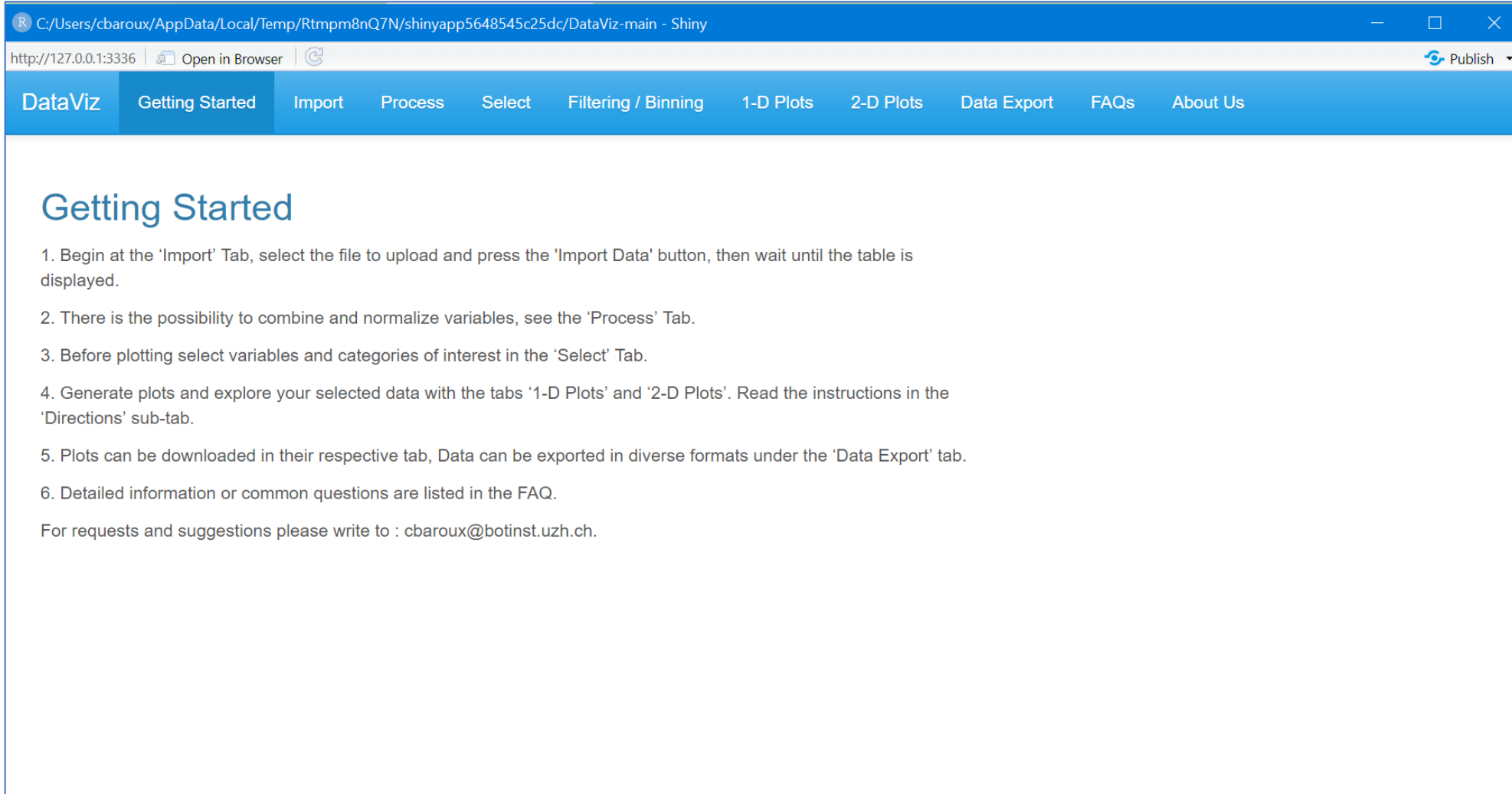
Launch DataViz

- Open your browser, go to:
- <https://github.com/barouxlab/DataViz>
- Open the 'README' file. Read the instructions
- In the section 'running the application' , copy the lines of code *, Paste them In R (or R studio). Run

```
* if(!"shiny" %in% rownames(installed.packages()))  
  install.packages("shiny"); library(shiny)  
  runUrl("https://github.com/barouxlab/DataViz/archive/main.zip")
```

- There is also the possibility to clone the repo locally – see infos in the README file

Launch DataViz



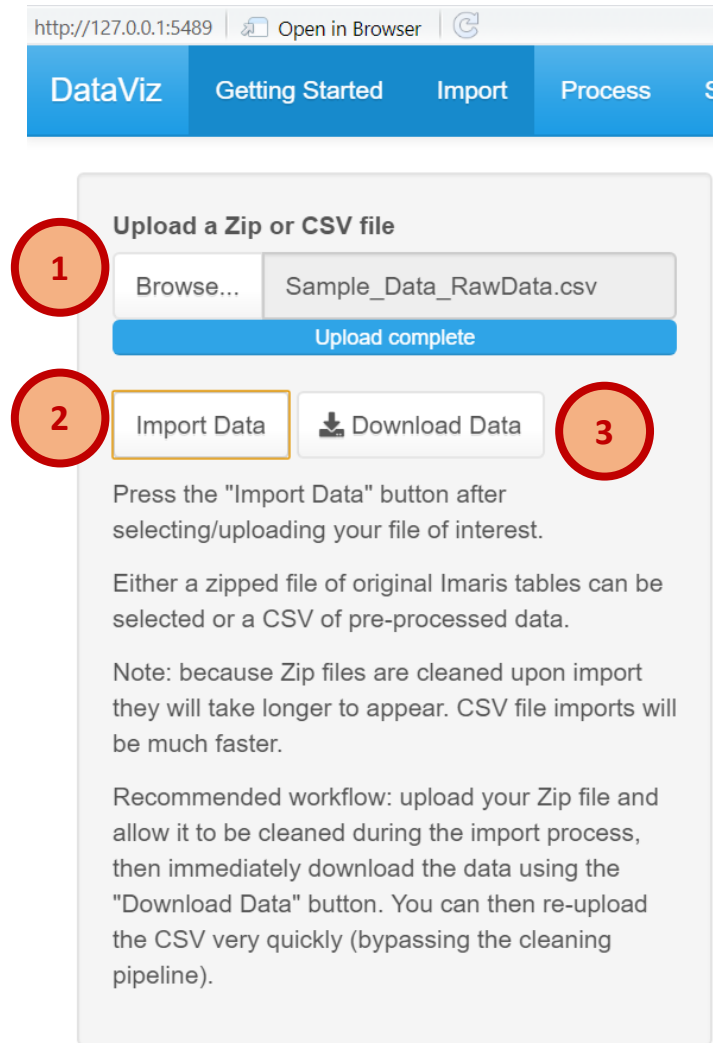
The screenshot shows a web browser window displaying the DataViz Shiny application. The browser's address bar shows the URL `http://127.0.0.1:3336` and includes a 'Publish' button. The application's navigation bar is blue and contains the following tabs: DataViz, Getting Started, Import, Process, Select, Filtering / Binning, 1-D Plots, 2-D Plots, Data Export, FAQs, and About Us. The 'Getting Started' tab is currently selected. The main content area has a white background and features a heading 'Getting Started' in blue. Below the heading is a numbered list of six steps for using the application. At the bottom of the list, there is a line of text providing an email address for requests and suggestions.

Getting Started

1. Begin at the 'Import' Tab, select the file to upload and press the 'Import Data' button, then wait until the table is displayed.
2. There is the possibility to combine and normalize variables, see the 'Process' Tab.
3. Before plotting select variables and categories of interest in the 'Select' Tab.
4. Generate plots and explore your selected data with the tabs '1-D Plots' and '2-D Plots'. Read the instructions in the 'Directions' sub-tab.
5. Plots can be downloaded in their respective tab, Data can be exported in diverse formats under the 'Data Export' tab.
6. Detailed information or common questions are listed in the FAQ.

For requests and suggestions please write to : cbaroux@botinst.uzh.ch.

Upload Data



1. Browse to locate your zip folder. Confirm 'Open'

2. When the bar indicate 'Upload complete', Press 'Import Data'

Wait a few minutes – the import includes a cleaning function, data integrity check and creates a dataframe compiling all the data. This can take a few minutes depending on data size and computational power

3. When the table appears on the right, Press Download data

- If you resume the work at a later stage, with a new session of R, you can upload and import the Raw Data file, the import will be much faster*
- Note: some windows system seem not to recognize the csv format of the file. In that case, the system crashes at opening. Add manually the '.csv' extension to your filename*

Data Integrity Check

Upload a Zip or CSV file

Browse...

Sample_Data_RawData.csv

Upload complete

Import Data

Download Data

Press the "Import Data" button after selecting/uploading your file of interest.

Either a zipped file of original Imaris tables can be selected or a CSV of pre-processed data.

Note: because Zip files are cleaned upon import they will take longer to appear. CSV file imports will be much faster.

Recommended workflow: upload your Zip file and allow it to be cleaned during the import process, then immediately download the data using the "Download Data" button. You can then re-upload the CSV very quickly (bypassing the cleaning pipeline).

Table

Data Integrity

Show 10 entries

Search:

	Genotype	Treatment	Image File	Object ID	Category	Channel	O
1	WT	Treatment B	rsr_568	0	Spot	1	Nu Cei Ma
2	WT	Treatment B	rsr_568	0	Spot	1	Spt
3	WT	Treatment B	rsr_568	0	Spot	2	Nu Cei Ma
4	WT	Treatment B	rsr_568	0	Spot	2	Spt
5	WT	Treatment B	rsr_568	0	Surface	1	Nu
6	WT	Treatment B	rsr_568	0	Surface	1	Nu

Table

Data Integrity

Images with unexpected Objects:

Unexpected Object string:

Images without a Nucleus Center of Mass: rsr_563_rsr_564

Images without a Nucleus: rsr_563_rsr_570, rsr_573_rsr_579, rsr_573_rsr_582

Images with nothing but NA's in a column: rsr_563_rsr_570, rsr_573_rsr_579, rsr_573_rsr_582

Below is a table showing the Image and Object ID's of records that have NA in the Channel

Show 10 entries

Search:

Image File	Object ID
------------	-----------

Once the dataframe is uploaded, check the 'Data Integrity' Tab – some inconsistency in the data may prompt the user to go back to the images and export steps – or can continue with data analysis knowing that not all variables will be present

Data Integrity Check

[Table](#) [Data Integrity](#)

Images with unexpected Objects:

Unexpected Object string:

Images without a Nucleus Center of Mass: rsr_563_rsr_564

Images without a Nucleus: rsr_563_rsr_570, rsr_573_rsr_579, rsr_573_rsr_582

Images with nothing but NA's in a column: rsr_563_rsr_570, rsr_573_rsr_579, rsr_573_rsr_582

Below is a table showing the Image and Object ID's of records that have NA in the Channel

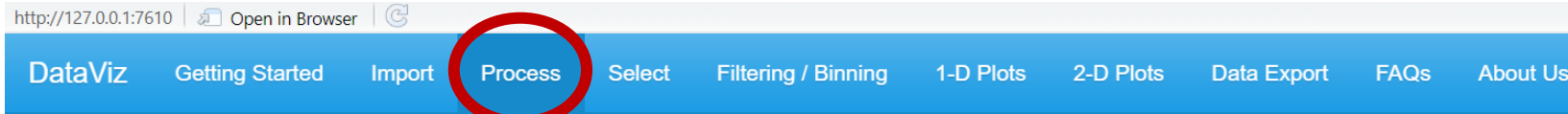
Show entries Search:

Image File	Object ID
------------	-----------

In this example (left):

- some images are missing an object called 'Nucleus' hence intensity normalization will not be performed for these images
- other images are missing a Center of Mass hence the normalization of distances will not be done for the objects in these images
- Some images are missing some statistics, hence NA in one of the column (which statistics can be checked in the data table)

Data Processing



☐ Intensity normalization per image:

- For each intensity variable (sum, standard deviation), the value is divided by the value in the Nucleus, in the corresponding channel (Ch).
- This means the intensity variables for each Nucleus become =1
- This is done per object per image

☐ Distance normalization per image:

- For the distance variable spot-to-surface, for each object, the value is divided by the variable 'Shortest distance' of the Nucleus' Center of Mass (spot) to the Nucleus surface
- Note that this normalization is interesting only if the nucleus is near spherical
- *In future:* this distance normalization will be customizable to other variables (surface-to-surface) and to other metrics (eg Nucleus diameter X,Y,Z ; Area)

Data Processing

❑ Channel Ratio:

- for each object (spot, surface)
- option for Intensity Sum and Intensity Mean
- Customised ratio of signal intensity between channels of choice
- Example: Ch2:Ch1, Ch3:Ch2, Ch4:Ch1 ...

❑ Group Sum:

- For each image, for each intensity variable (sum, standard deviation), per object category, the values of all object are added in the corresponding channel (Ch).
- Example: Group Intensity Sum 'CC' (Ch=1) = Sum of the Intensity Sum of all CC in Channel 1

❑ Signal Density:

- For each object, each channel, the normalized intensity sum is divided by the volume

❑ Object Count:

- For each image, the number of objects in one category is counted
- Example: number of chromocenter (CC) per Nucleus

Data Processing

❑ Normalised distances

- Requirement: spot object called 'Nucleus Center Of Mass' (created during image segmentation)
- For each image, the shortest distance of the spot = 'Nucleus Center Of Mass' to the surface = 'Nucleus' is taken as a reference distance
- For each object, the 'spot-to-surface' shortest distance is divided by the reference distance

❑ Normalised Intensities

- Requirement: surface object called 'Nucleus'
- For each intensity variable (sum, mean, standard deviation), for each object, the value is divided by the value in the 'Nucleus', in the corresponding channel
- As a consequence, the normalized intensity in the 'Nucleus' becomes = 1 , and that of an object is a fraction of it

❑ Relative Intensity Fraction (RIF)

- For each image, the RIF is the Group Intensity Sum divided by the Intensity Sum in the 'Nucleus' in the corresponding channel
- Note: for the chromocenters, in the DNA staining channel, this corresponds to the RHF (Relative Heterochromatin Fraction)

❑ Relative Volume Fraction (RVF)

- For each image, the RVF is the Group Intensity Volume divided by the Volume of the 'Nucleus'

Select Data to be plotted

The screenshot shows the DataViz web application interface. The top navigation bar includes links for Getting Started, Import, Process, Select, Filtering / Binning, 1-D Plots, 2-D Plots, Data Export, FAQs, and About Us. The 'Select' tab is active. On the left sidebar, there are sections for 'Dataset for Selection' (Original Data, Processed Data), 'Channel' (1, 2), 'Object (Type)' (CC, Nucleolus, Nucleus, Nucleus Center of Mass, Speckle), 'Treatment' (Treatment A, Treatment B), 'Genotypes' (WT), and 'Image File' (rsr_568, rsr_571, rsr_572, rsr_574, rsr_579, rsr_582, rsr_590, rsr_600). The main content area shows a table of selected data with columns: Genotype, Treatment, Image File, Object ID, Category, and Channel. The table displays 13 rows of data. Below the table, it says 'Showing 1 to 100 of 178,810 entries' and has pagination controls. The 'Outliers' tab is also visible above the table.

1. Dataset for Selection

2. Channel

3. Select Data

4. Download

5. Outliers

1. Chose the data to be handle: original or processed data
→ If the data imported are already processed (from another session), chose 'original' data
2. Chose the channel, object, treatment, genotype and image file to use for plotting.
→ Think carefully of what needs to be plotted.
→ Selection can be changed dynamically while elaborating plots.
3. Press the 'Select Data' button, a table of selected data appears (right)
4. It is possible to download this data subset for additional analyses in third-party software (eg statistical analyses)
5. Inspect outliers (optional)
→ This step can be performed only after a first round of plotting. In case outliers are visible, it is possible to filter them out .
→ For this, in the 'Outliers' Tab: select the variable, generate outliers, select 'yes' to the 'Remove outliers' option, and press 'Select Data' (3) again

1D-Plot : presentation

The screenshot displays the DataViz application interface. At the top, a blue navigation bar contains links: DataViz, Getting Started, Import, Process, Select, Filtering / Binning, 1-D Plots (highlighted with a red circle), 2-D Plots, Data Export, FAQs, and About Us. Below the navigation bar, the interface is divided into two main sections. On the left, a control panel includes buttons for 'Update Parameters', 'Generate Plots', 'Download Plots', and 'Download Data'. It features three dropdown menus: 'Select Variable' (set to 'Area'), 'Color by' (set to 'Category'), and 'Split by' (set to 'Category'). Below these are tabs for 'Parameters and Formatting', with 'Histo' selected. Further down are input fields for 'X Min' (0) and 'X Max' (1), and a '# of Bins' field (30). On the right, a sub-navigation bar shows tabs for 'Directions', 'Histograms', 'Density Plots', 'Boxplots', 'Boxplot Stats', and 'Violin Plots'. Below this, a message states: '1. Requirements: data to plot have been selected using the Select tab'.

- 1-D plots consist in Histograms, Density Plots, Box Plots and Violin Plots
- Plots parameters and format of the plotting area are adjusted using the Tabs in the section 'Parameters and formatting'
- 'Update parameters' will (re)set the plot parameters based on the selected variable and category for coloring/splitting
- It is important to press 'Update parameters' then 'Generate Plots' each time another variable or category for color/split is chosen.
- If plot parameters are adjusted, press 'Generate Plots' to refresh the plot area
- Box plots statistics deliver the plot's descriptions
- 'Download Data' creates a csv file corresponding to the data plotted (=data subset)

1D-Plot : general directions

The screenshot shows the DataViz application interface. The top navigation bar is blue with white text links: DataViz, Getting Started, Import, Process, Select, Filtering / Binning, 1-D Plots (circled in red), 2-D Plots, Data Export, FAQs, and About Us. Below the navigation bar, on the left, is a sidebar with a light gray background. It contains several sections: 'Update Parameters' and 'Generate Plots' buttons at the top; 'Download Plots' and 'Download Data' buttons below; a 'Select Variable' dropdown menu with 'Area' selected; a 'Color by' dropdown menu with 'Category' selected; a 'Split by' dropdown menu with 'Category' selected; a 'Parameters and Formatting' section with sub-tabs 'Histo', 'Density', and 'Box/Vio' (all in blue); a 'Layout' section with sub-tabs 'Layout', 'Themes', and 'Colors' (all in blue); and input fields for 'X Min' (0), 'X Max' (1), and '# of Bins' (30). On the right, there is a sub-tab area with tabs: 'Directions', 'Histograms', 'Density Plots', 'Boxplots' (circled in red with a red asterisk), 'Boxplot Stats', and 'Violin Plots'. A red vertical line and five red circles with numbers 1 through 5 are overlaid on the interface to indicate the steps: 1 points to the 'Select Variable' dropdown, 2 points to the 'Update Parameters' button, 3 points to the 'Generate Plots' button, 4 points to the 'Parameters and Formatting' section, and 5 points to the 'Download Plots' button.

1. Select the variable to be plotted, the category for coloring and plot splitting in the pull down menus
2. Press 'Update Parameters'
3. Press 'Generate Plots' and select one of the Plot type in the Sub-Tab area (*)
4. Adjust the plot parameters and formatting
 - Specific plot parameters (x, y axes) in the corresponding sub-Tab
 - Layout: set the size of the plots and the number of plots per row
 - Themes: set the background color and presence/absence of gridlines
 - Colors: chose or customize the color palette
5. Once the plots are ready, the plots can be downloaded (Download Plots) as well as the data subset used for the plots (Download Data) ; Boxplot statistics are in the subtab 'Boxplot Stats'

2D-Plot : general directions

1. Select the X and Y variables to be plotted and, eventually, the category for splitting plots, in the pull down menus
2. By default, density contours are active to feature the density of data points on the graph. This option should be deactivated for a small number of datapoints. When deactivated, a field appear where the categorical variable for coloring the datapoints can be chosen.
3. Press 'Update Parameters'
4. Press 'Generate Scatterplot', open the subtab 'Scatter Plot'
5. Adjust the plot parameters and formatting
 - Specific plot parameters (x, y axes) in the corresponding sub-Tab
 - Layout: set the size of the plots and the number of plots per raw
 - Themes: set the background color and presence/absence of gridlines
 - Colors: chose or customize the color palette
6. Once ready, the plots can be resized and downloaded ('Download Plots'), and the data subset used for the plots can be exported as .csv file ('Download Data')
7. plot's statistics are in the subtab 'Scatter Plot Data Table '

Adjust plot theme and colors



Update Parameters

Generate Plots Download Plots

Download Data

Select Variable

Normalized Intensity Mean

Color by

Object

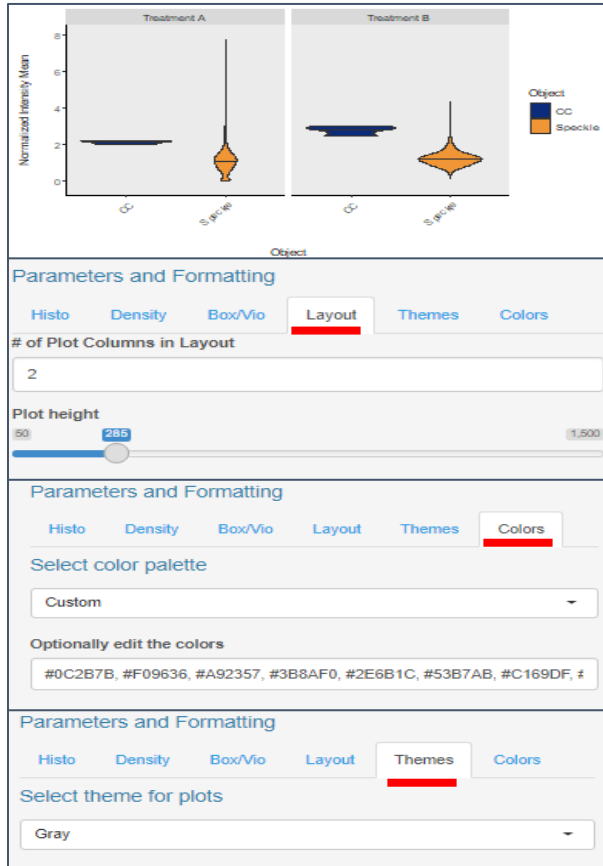
Split by

Treatment

Parameters and Formatting

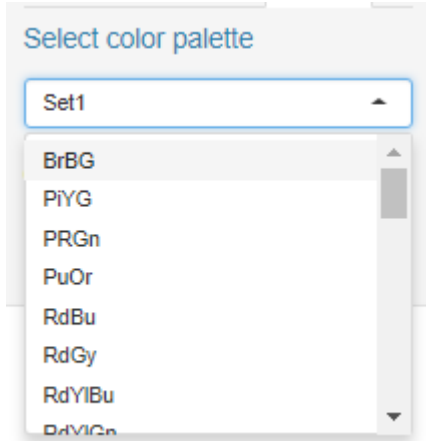
Histo Density Box/Vio

Layout Themes Colors

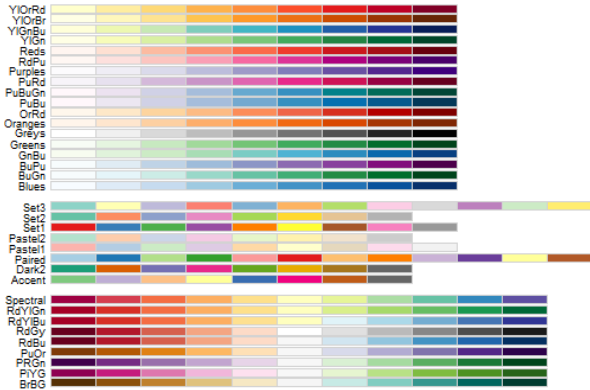


Sample Data, Data selected for CC, Speckles, Violin Plots, Split by: Treatment, Color by: Object type

Define colors manually



Color Brewer Available Palettes



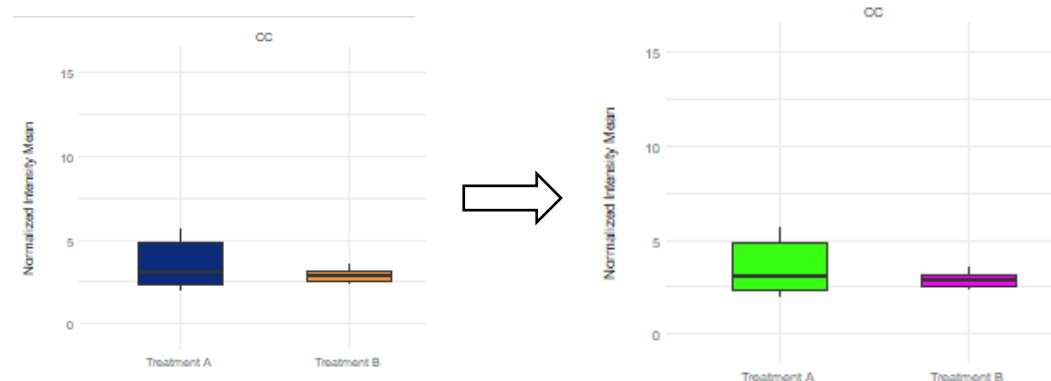
- The palettes in the pull down menu come from the Color Brewer
- The default coloring is 'Custom' corresponding to the following set:



- To change manually the type and order of color, use the window with the heximal code:

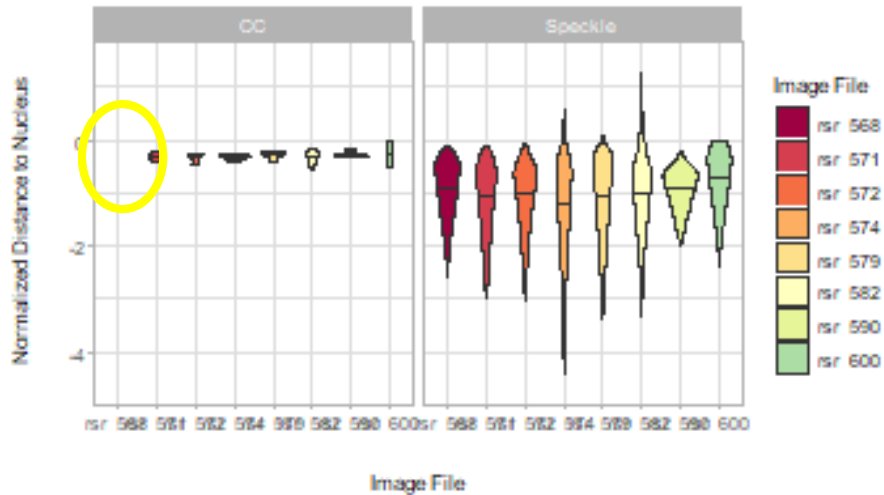


- Use <https://encycolorpedia.com/> to find the heximal code of the desired color
- For instance to have neon green and fuschia copy paste **#39ff14**, **#ff00ff** in the 'optionally edit the colors' window at the beginning of the list (colors must be separated by a comma) – the colors should update automatically



Plot per image file to verify data consistency/variability

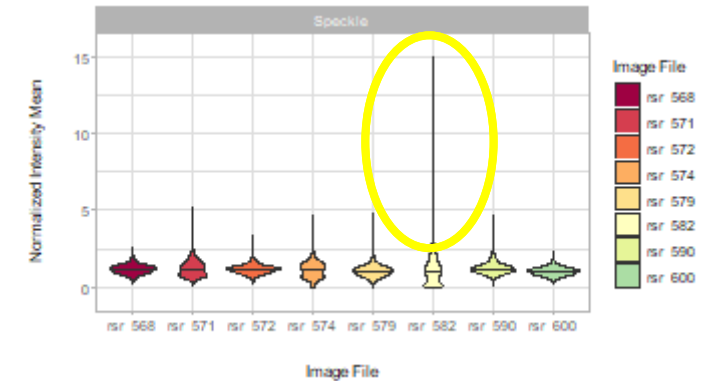
Example 1: Distance of CC and speckles to nucleus periphery



- In this example we can spot one image missing distance measurement for CC – this prompts to return to the image and activate the relevant statistics, then re-export the data

- Note: the color palettes have 9-11 colors. If more image files, the data will not be plotted (example error message – ‘Insufficient values in manual scale. 8 needed but only 7 provided’). To solve this, copy the list of Hex code in Parameters/Colors/ editing window (see slide about colors) and paste it at the end of the same list, after adding a comma.

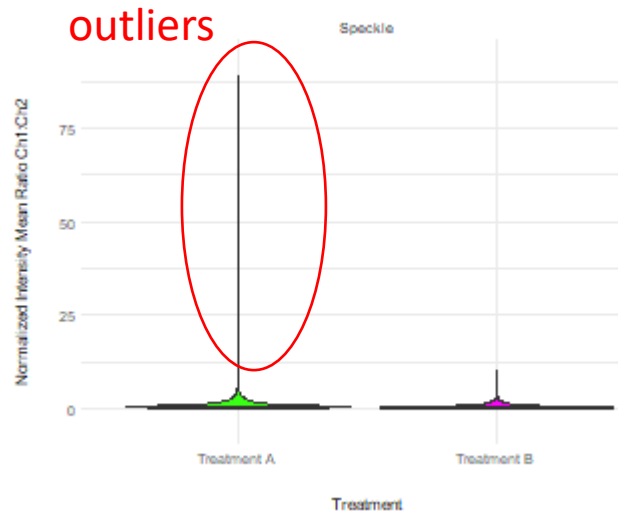
Example 2: Intensity Mean Ch=1 for Speckles



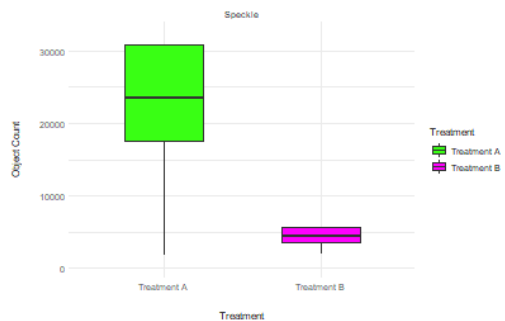
- In this example we can spot one image with highly variable and very high intensities (rsr582)– this prompts to return to the image and verify the spot segmentation, eventually curate and re-export the data

Filter Outliers

Before filtering



Object count (number of speckles before filtering)



Select Data 4

Download

Select the options for filtering then press "Select Data".

Note: there is the possibility to plot data per treatment separately.

Dataset for Selection

☐ Original Data

☒ Processed Data

Channel

☒ 1

☒ 2

Object (Type)

☐ CC

☐ Nucleolus

☐ Nucleus

☐ Nucleus Center of Mass

☒ Speckle

Treatment

☒ Treatment A

☒ Treatment B

Genotypes

☒ WT

Image File

☒ rsr_568

☒ rsr_571

Table Outliers

Outliers for Selected Level

Category Level

Normalized Intensity Mean Ratio Ch1:Ch2 1

Remove outliers on next filter?

☒ Yes 3

☐ No

Generate Outliers for Selected Variable 2

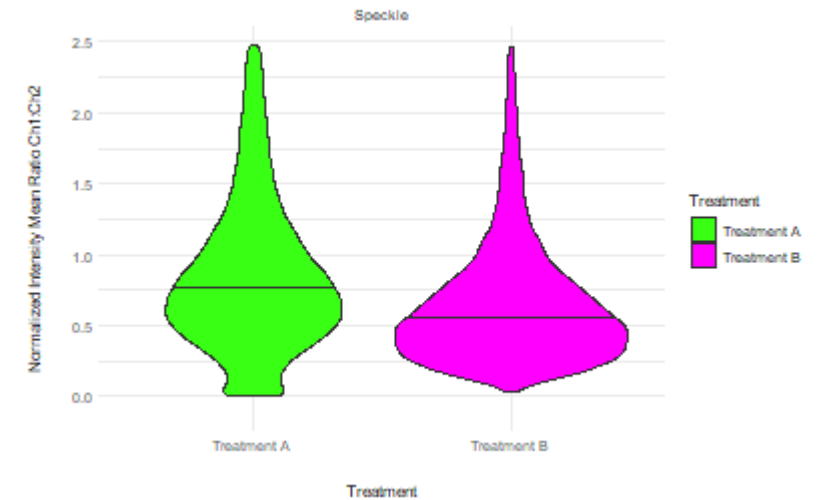
Show 10 entries Search:

	Genotype	Treatment	Image File	Object ID	Category
1	WT	Treatment B	rsr_568	0	Spot
2	WT	Treatment B	rsr_568	0	Spot
3	WT	Treatment B	rsr_568	1	Spot
4	WT	Treatment B	rsr_568	1	Spot
5	WT	Treatment B	rsr_568	2	Spot
6	WT	Treatment B	rsr_568	2	Spot
7	WT	Treatment B	rsr_568	198	Spot
8	WT	Treatment B	rsr_568	198	Spot
9	WT	Treatment B	rsr_568	235	Spot
10	WT	Treatment B	rsr_568	235	Spot

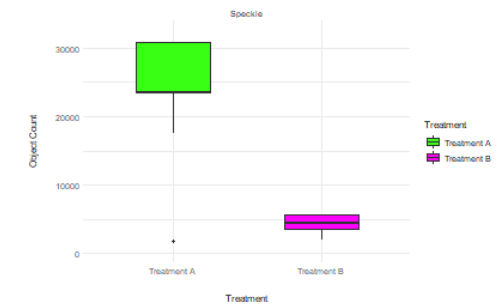
Showing 1 to 10 of 14,392 entries

Previous 1 2 3 4 5 ... 1,440 Next

After filtering

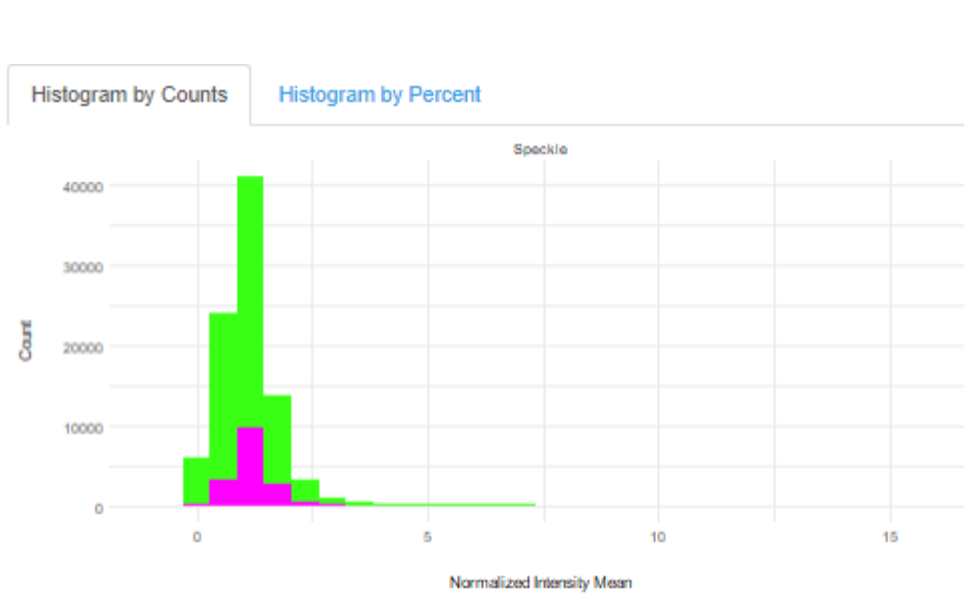


Object count (number of speckles after filtering)



Adjust x and y limits

Example 1:



Parameters and Formatting

Histo [Density](#) [Box/Vio](#) [Layout](#)

[Themes](#) [Colors](#)

X Min

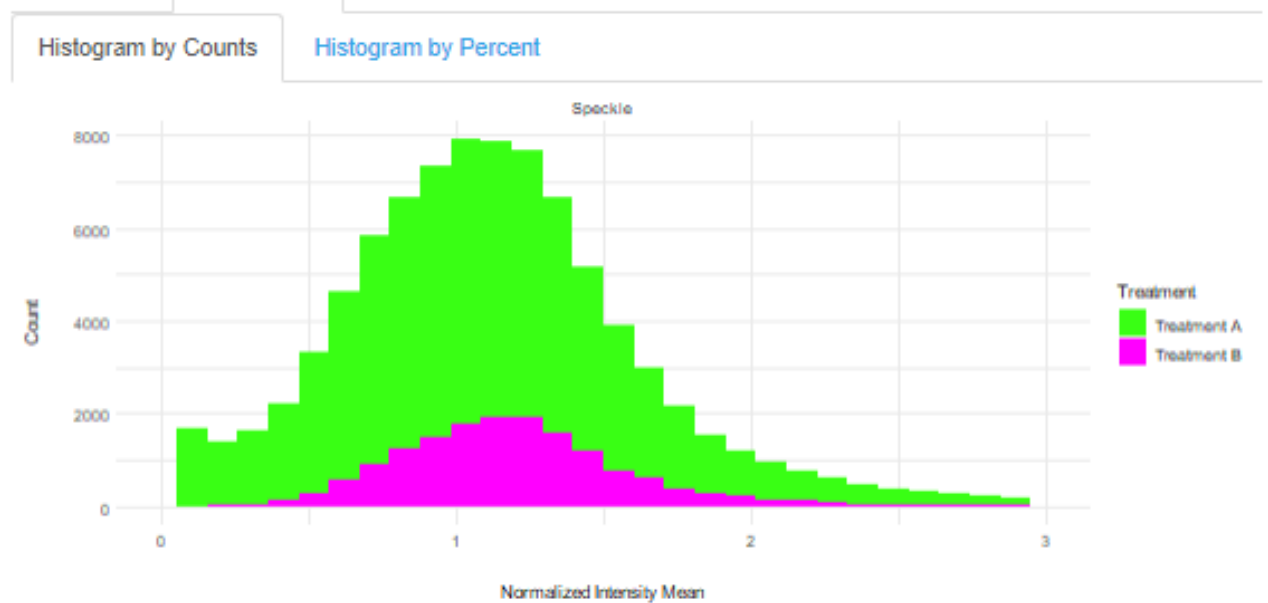
-1,03109020331

X Max

15,98189815135

of Bins

30



Parameters and Formatting

Histo [Density](#) [Box/Vio](#) [Layout](#)

[Themes](#) [Colors](#)

X Min

0

X Max

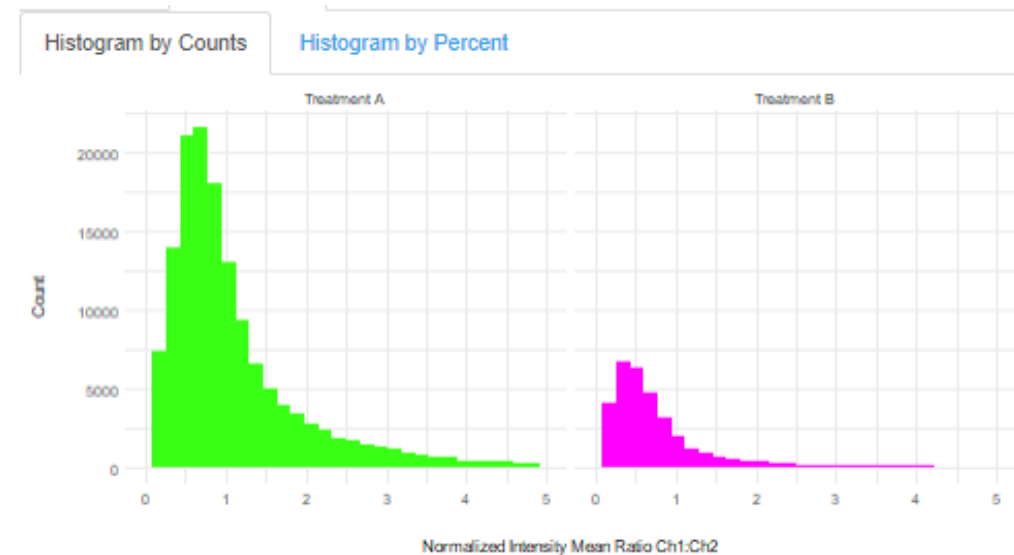
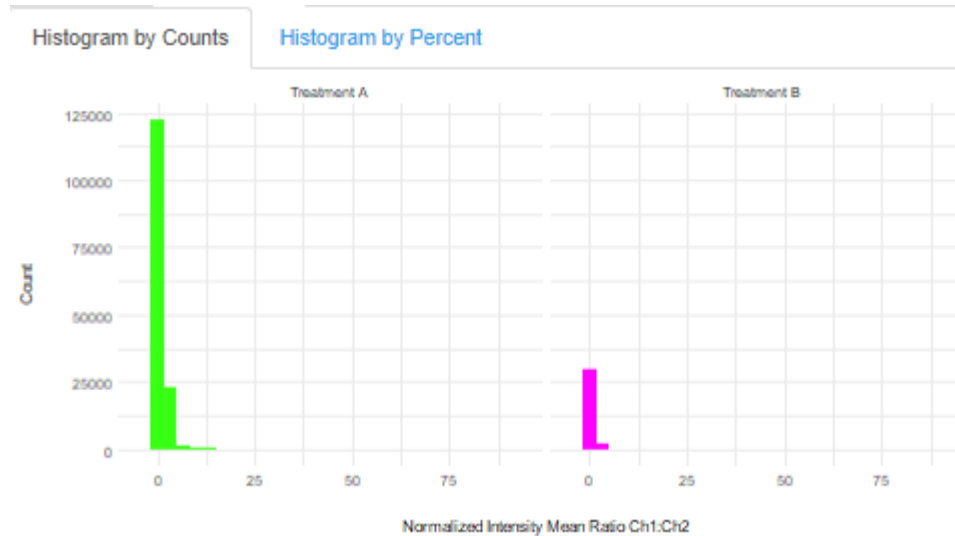
3

of Bins

30

Adjust x and y limits

Example 2: Histogram (or KDE) plot packed on the left side of the graph -> reduce X Max



Note: this is a case where outliers should be removed as they influence the overall shape distribution. The violin plot (left) and the histogram (top) show that most values lay between 0 and 5, with outliers up to 80 : -> see section 'Filter outliers'

Adjust x and y limits

Example 3: Density distribution curve seems broken -> adjust the Y Max limit

Update Parameters Generate Plots

Download Plots Download Data

Select Variable

Distance To 3 Nearest Neighbours

Color by

Treatment

Split by

Treatment

Parameters and Formatting

Histo Density Box/Vio Layout

Themes Colors

X Min X Max

0,039426025 - 0,983280475

Y Min Y Max

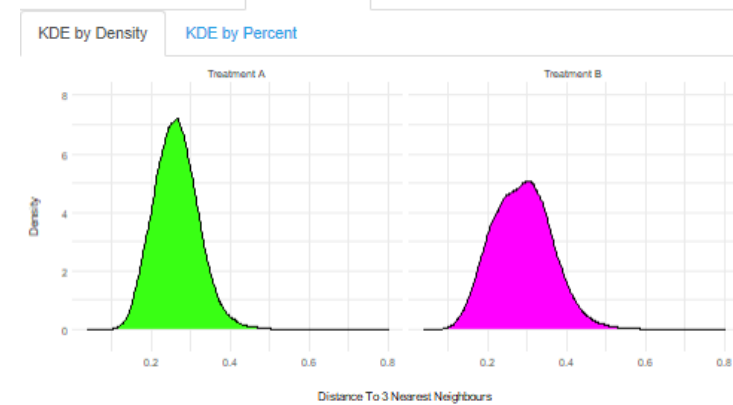
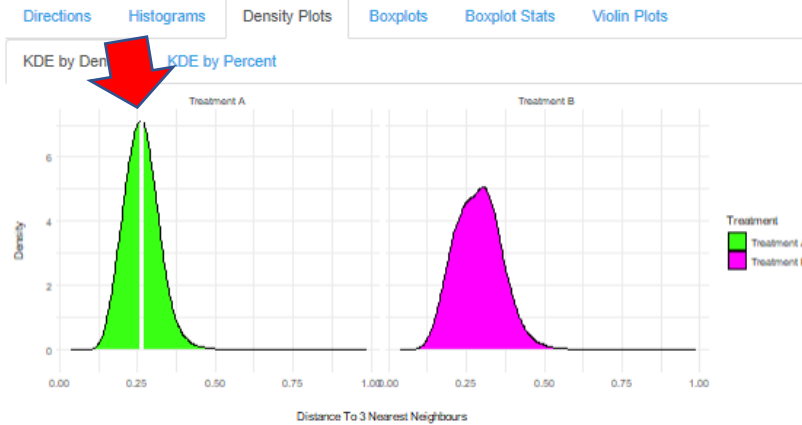
-0,33808707453 - 7,099828565227

KDE Smoothing Adjust

1

of Bins for % KDE

20

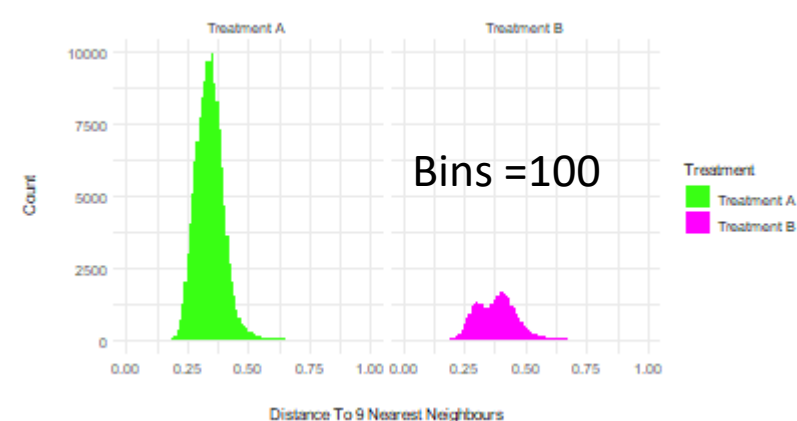
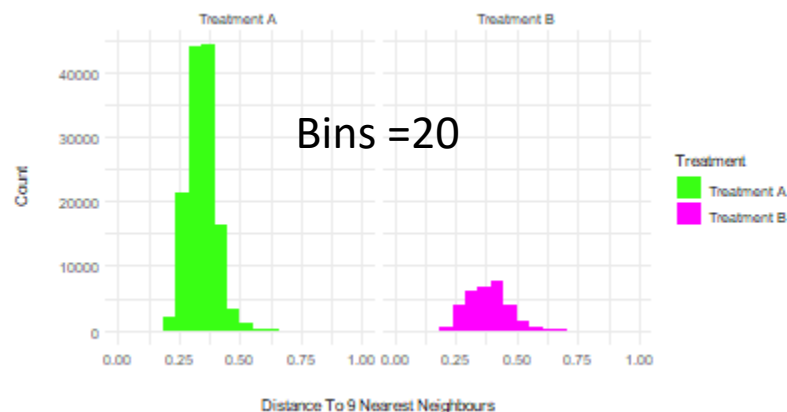
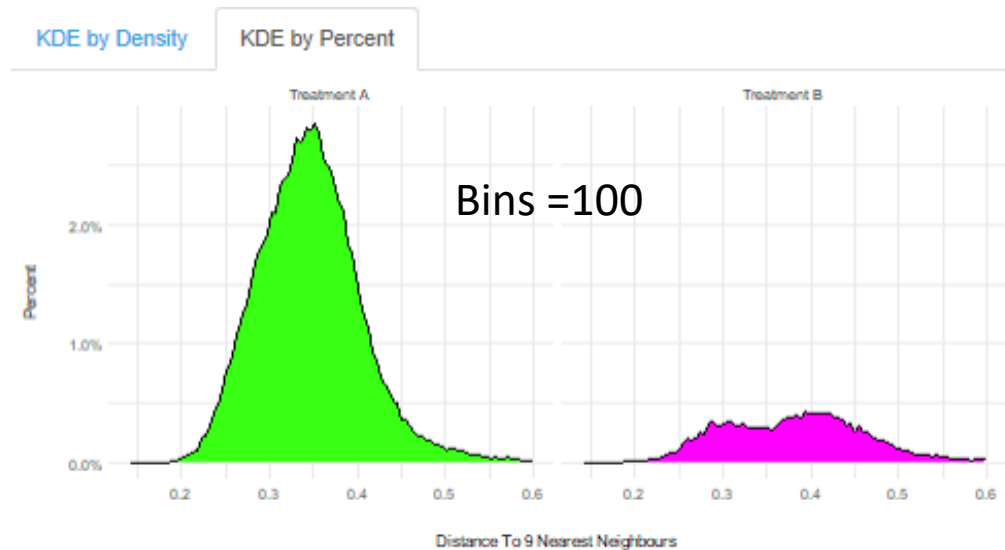
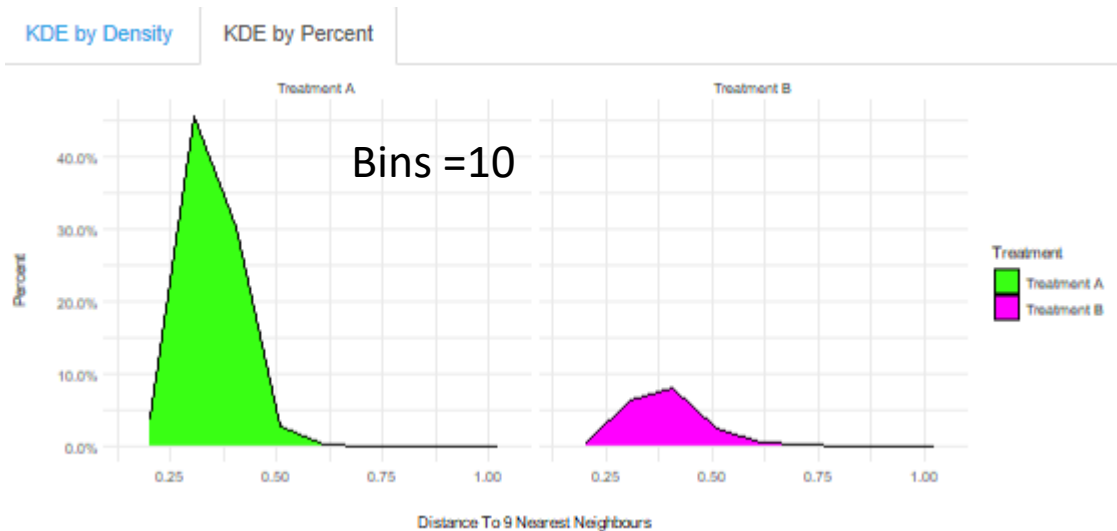


Y Max changed to 8
X Max changed to 0.8

Calculated parameters

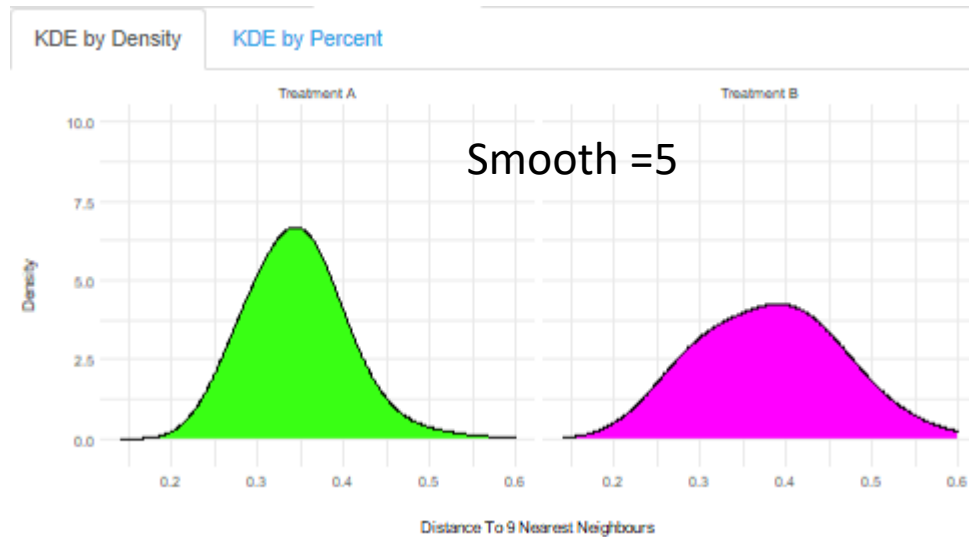
Smooth histograms and KDEs

Example 1 : KDE / Histogram distribution show gross shapes -> adjust the number of bins



Smooth histograms and KDEs

Example 2: effect of smoothing on KDE distribution



Parameters and Formatting

Histo Density Box/Vio Layout

Themes Colors

X Min X Max

0,140915

0.6

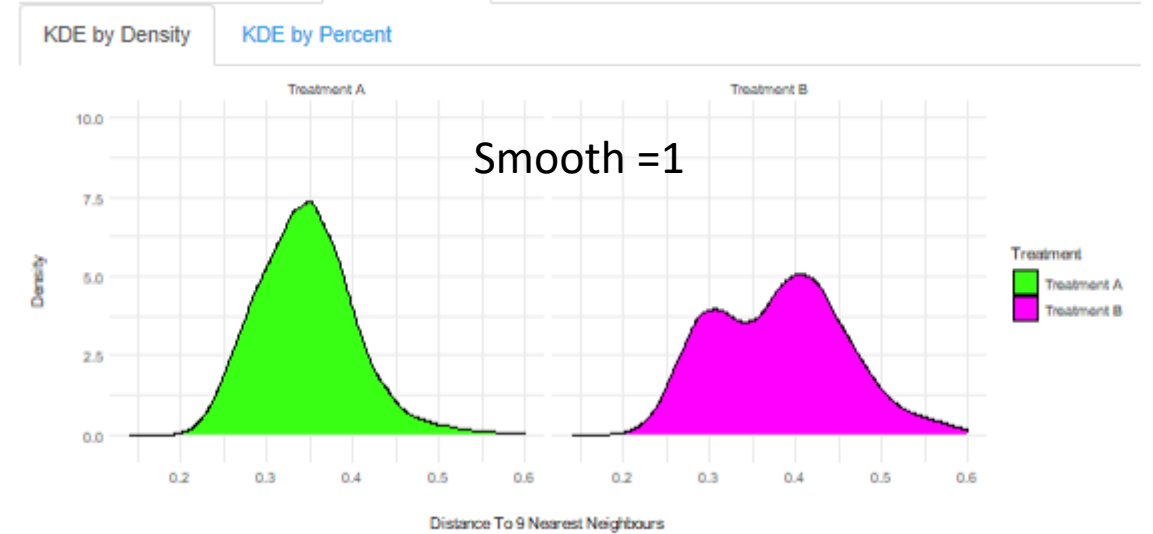
Y Min Y Max

-0,36722564496

10

KDE Smoothing Adjust

5



Parameters and Formatting

Histo Density Box/Vio Layout

Themes Colors

X Min X Max

0,140915

0.6

Y Min Y Max

-0,36722564496

10

KDE Smoothing Adjust

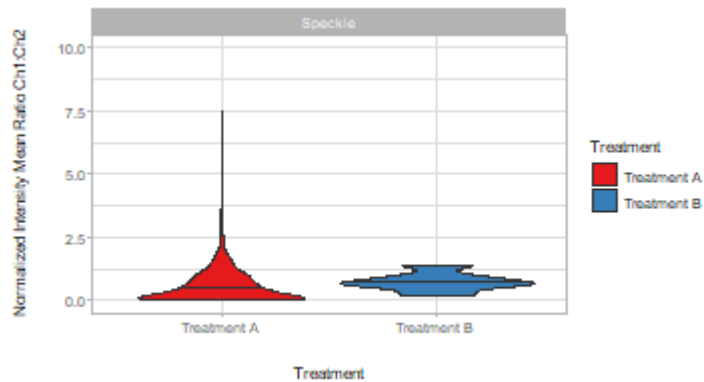
1

Data sub-setting: filter

In this example we select datapoints with value for 'Distance to Nucleus' (=nucleus periphery) between -0.2 and 0 (in μm)

Note that negative distances correspond to datapoints inside the surface (Nucleus). For other applications considering datapoints outside a surface, select a range with positive values

All data



Channel ratio for all speckles

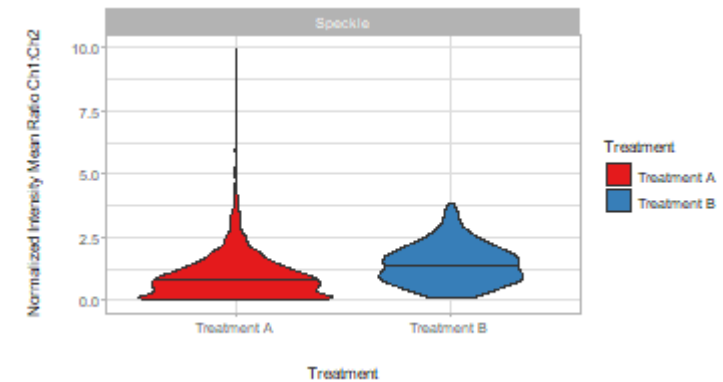
Sample Data, selected for Speckles

The screenshot shows the 'DataViz' software interface. The 'Filtering / Binning' tab is selected and highlighted with a red circle. Within this tab, the 'Filtering' section is highlighted with a red box. It shows a dropdown menu set to 'Distance to Nucleus', with 'Lower Limit (Inclusive)' set to -0.1 and 'Upper Limit (Inclusive)' set to 0. The 'Apply Filter' button is visible. Below the filtering section, the 'Binning' section is partially visible, showing 'Variable for Group Creation' set to 'Area'. To the right of the filtering section, a table titled 'Filtered / Binned Data will be displayed below:' shows a list of 8 entries with columns for Genotype, Treatment, Image File, and Object ID.

	Genotype	Treatment	Image File	Object ID
1	WT	Treatment B	rsr_571	1331
2	WT	Treatment B	rsr_572	73
3	WT	Treatment B	rsr_572	1747
4	WT	Treatment B	rsr_572	3160
5	WT	Treatment B	rsr_572	3162
6	WT	Treatment B	rsr_572	3166
7	WT	Treatment A	rsr_574	53
8	WT	Treatment A	rsr_574	81

- Tab 'Filtering/Binning', Section 'Filtering'
- Enter the range to subset
- Press 'Apply filter'
- Return to the Plot Tab – press 'Generate Plots'
- Data sub-setting can be reversed by applying 'Cancel Filter' and update the plots

Filtered data



Channel ratio for speckles within a 200nm range of the nucleus periphery

This approach allows to focus on a specific set of data points focused on the biological question

Data sub-setting: create groups

In this example we create groups of datapoints located at different distance range from the Nucleus periphery

Note that negative distances correspond to datapoints inside the surface (Nucleus). For other applications considering datapoints outside a surface, select a range with positive values



Filtering

Filter by
Distance to Nucleus

Lower Limit (Inclusive) Upper Limit (Inclusive)
-5 - 0

Apply Filter Cancel Filter

Binning

Subset the data into groups defined by thresholds

Variable for Group Creation
Distance to Nucleus

Create groups by thresholds or subset by a single range

☒ Thresholds (e.g., '0,1,5,10')

☐ Range (e.g., '1,5')

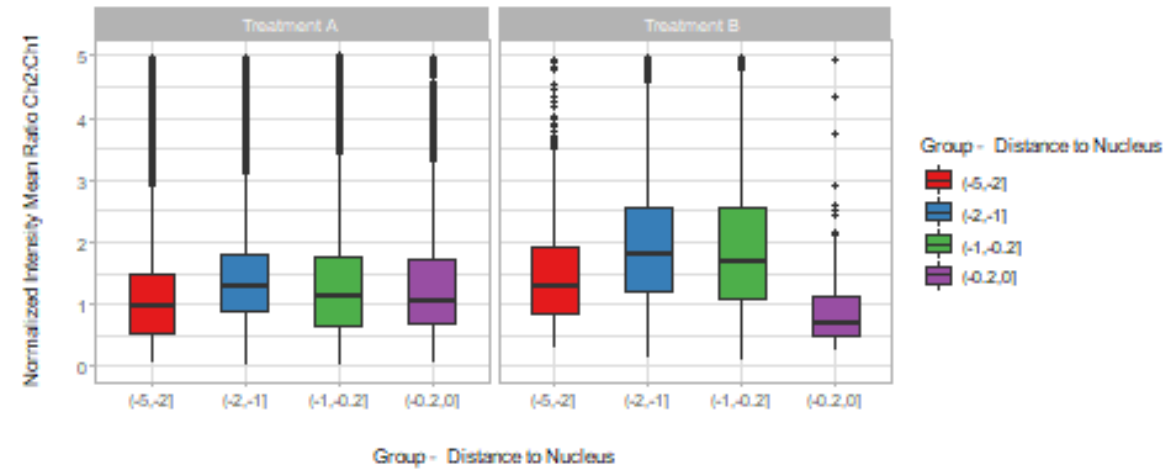
-5, -2, -1, -0.2, 0

Create Bins/Groups

Clear Bins/Groups

- Tab 'Filtering/Binning', Section 'Binning'
- Enter the cut-off values for sub-setting
- Press 'Create Groups/Bins'
- Return to the Plot Tab
- A new category is created 'Group - x' (x = variable used for sub-setting) that can be used for splitting or coloring
- Data sub-setting can be reversed by applying 'Clear Bins/Groups' and update the plots

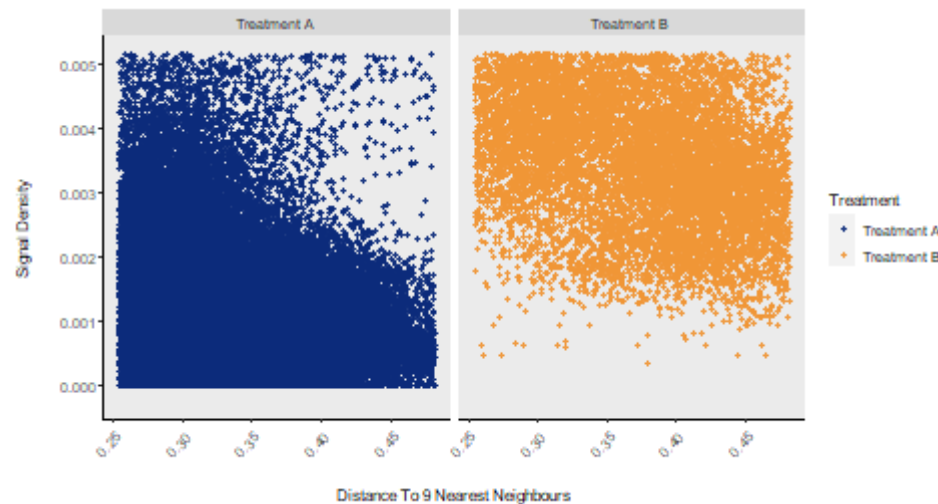
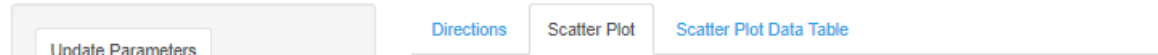
Channel intensity ratio plotted as groups according to their distance to the nucleus periphery



This approach reveals a group showing different intensity ratio between treatment (purple)

The data can be downloaded as csv file in the Plot area for statistical analyses

Scatter plot – datapoints or density contours



X Variable
Distance To 9 Nearest Neighbours

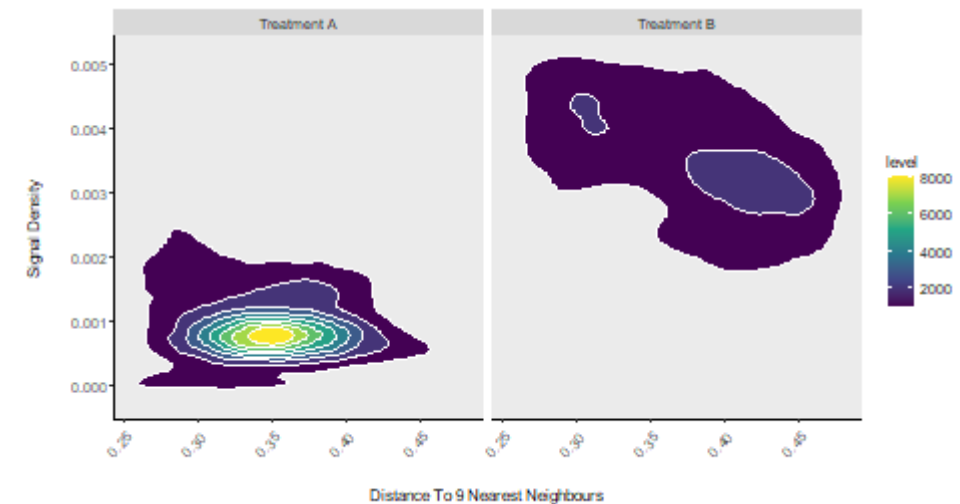
Y Variable
Signal Density

Categorical Variable for Splitting
Treatment

☐ Add contours?

☒ Categorical Variable for Color
Treatment

- In this example, the number of datapoints masks a possible heterogeneous distribution



X Variable
Distance To 9 Nearest Neighbours

Y Variable
Signal Density

Categorical Variable for Splitting
Treatment

☒ Add contours?

- density contours help identifying a different pattern relative to spacing between speckles (distance to 9 nearest neighbours) and signal density (intensity sum:volume ratio)