DataViz

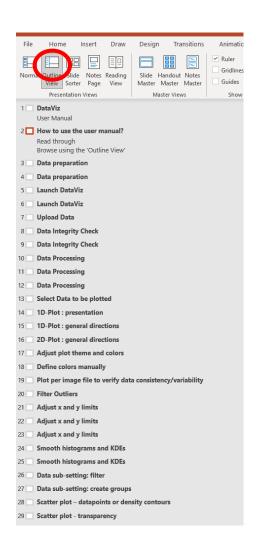
User Manual

Last updated : 30.06.2022

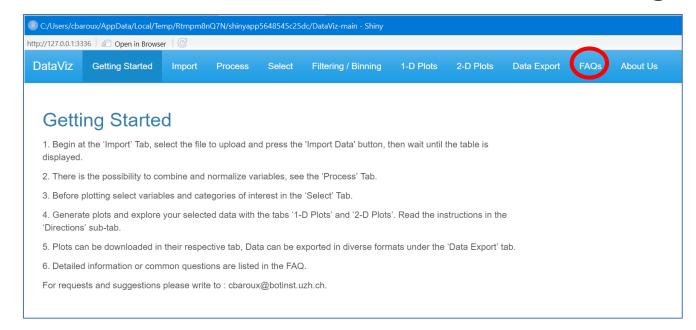


Célia Baroux, IPMB Devin Routh, S3IT

User manual – 'How to '?



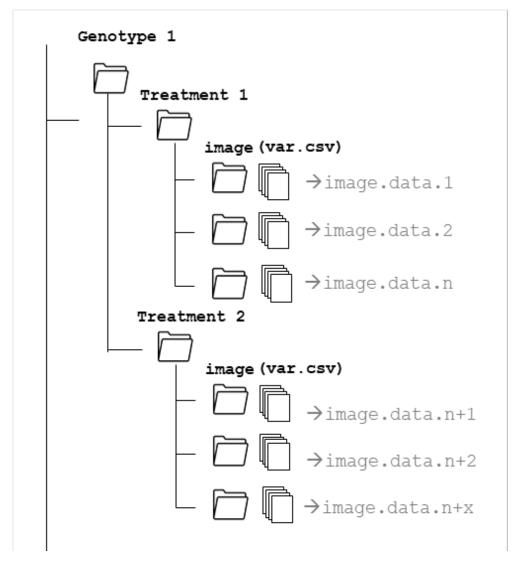
- Browse through specific topic using the 'Outline View'
- Check for latest update of the user manual on https://github.com/barouxlab/DataViz
- Check the FAQ in DataViz for troubleshooting



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: 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

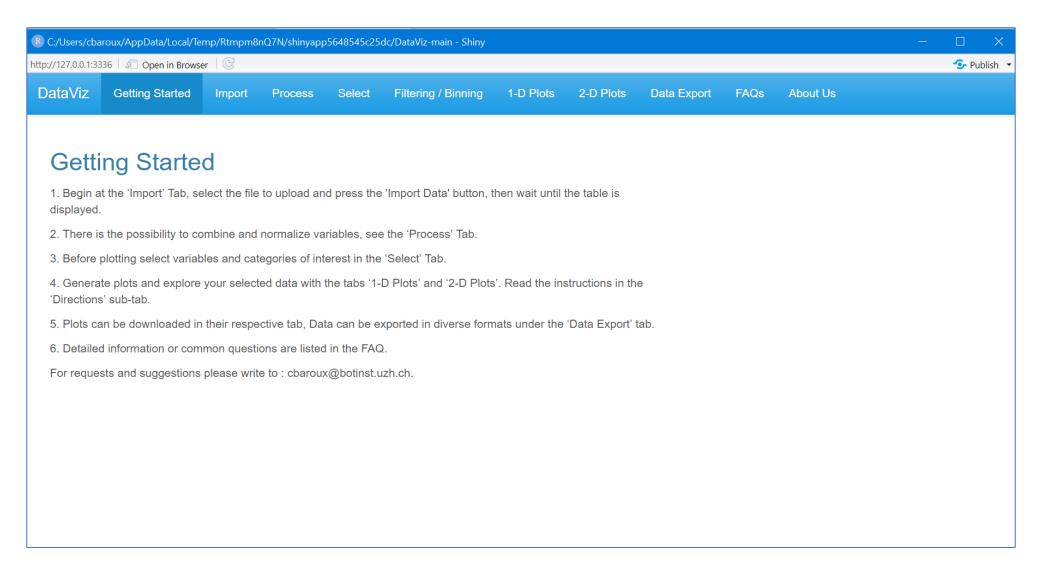
Launch DataViz

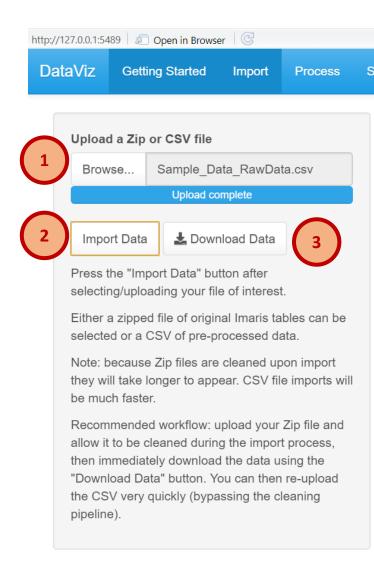
- Open your browser, go to:
- https://github.com/barouxlab/DataViz
- Open the 'README' file. Read the instructions
- In the section 'running the application', copy the lines of code *, Paste them In R (or R studio). Run

```
* 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





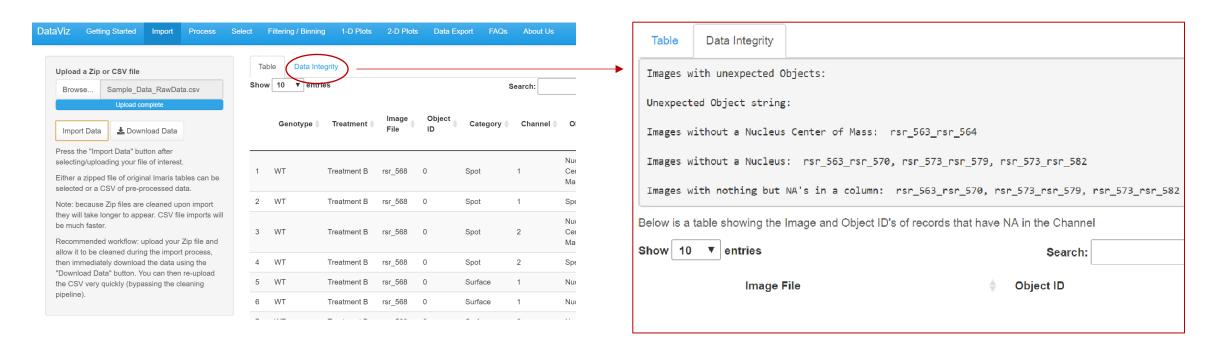
Upload Data

- 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

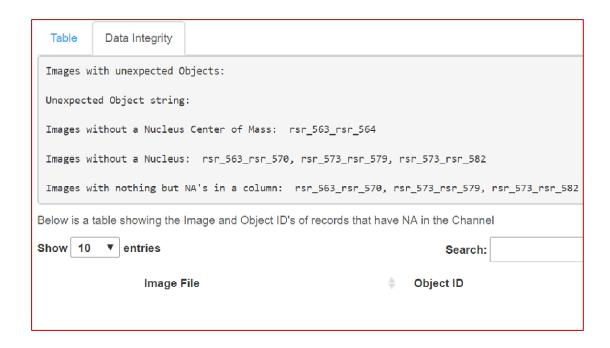
- 3. When the table appears on the right, Press Download data
- If you resume the work at a later stage, with a new session of R, you can upload and import the Raw Data file, the import will be much faster
- Note: some windows system seem not to recognize the csv format of the file. In that case, the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

Data Integrity Check



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

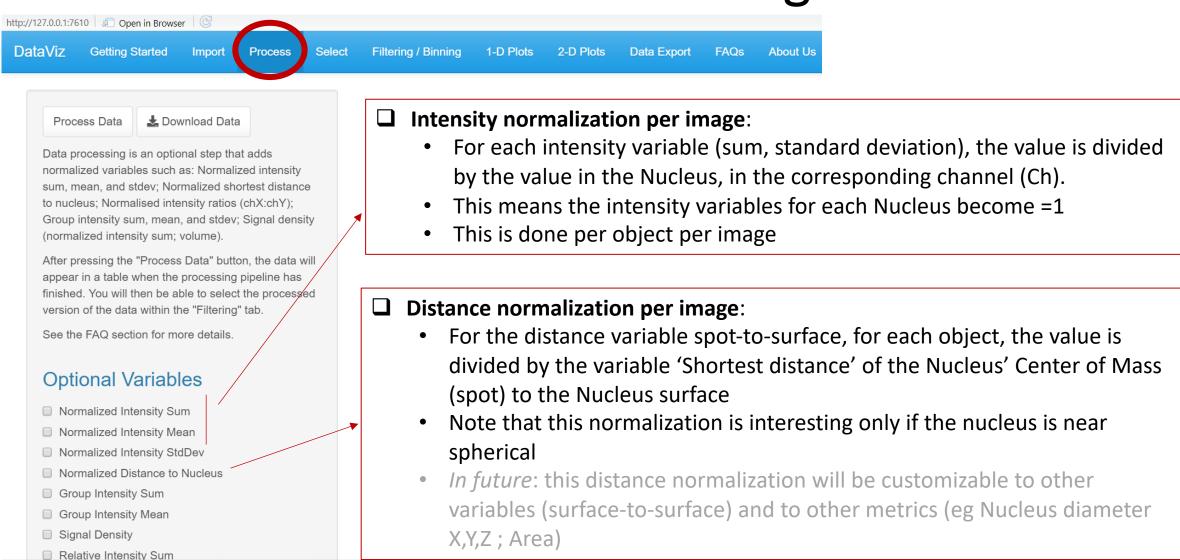
Data Integrity Check



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



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
- Eg Ch2:Ch1, Ch3:Ch2 etc.

☐ 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
 - o 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

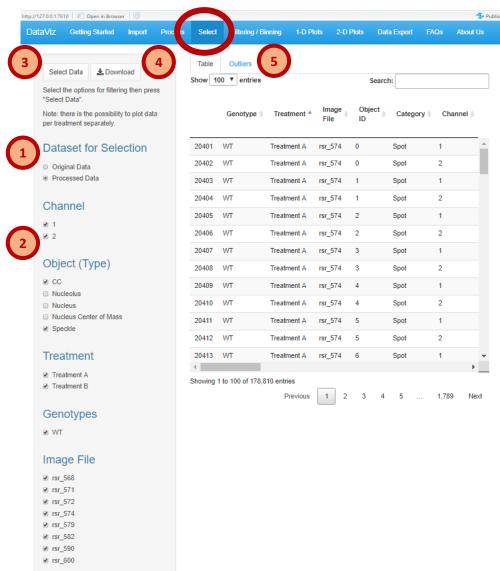
☐ Relative Intensity Fraction

- 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

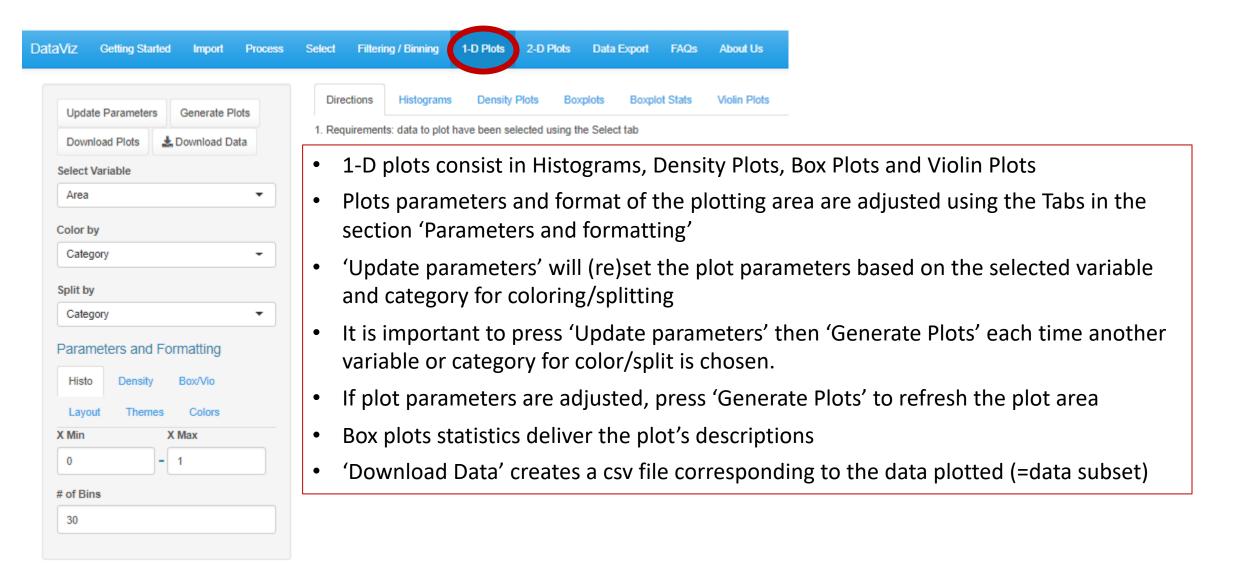
- For each image, the RVF is the Group Intensity Volume divided by the Volume of the 'Nucleus'
- *In future*: will be extended to other metrics

Select Data to be plotted

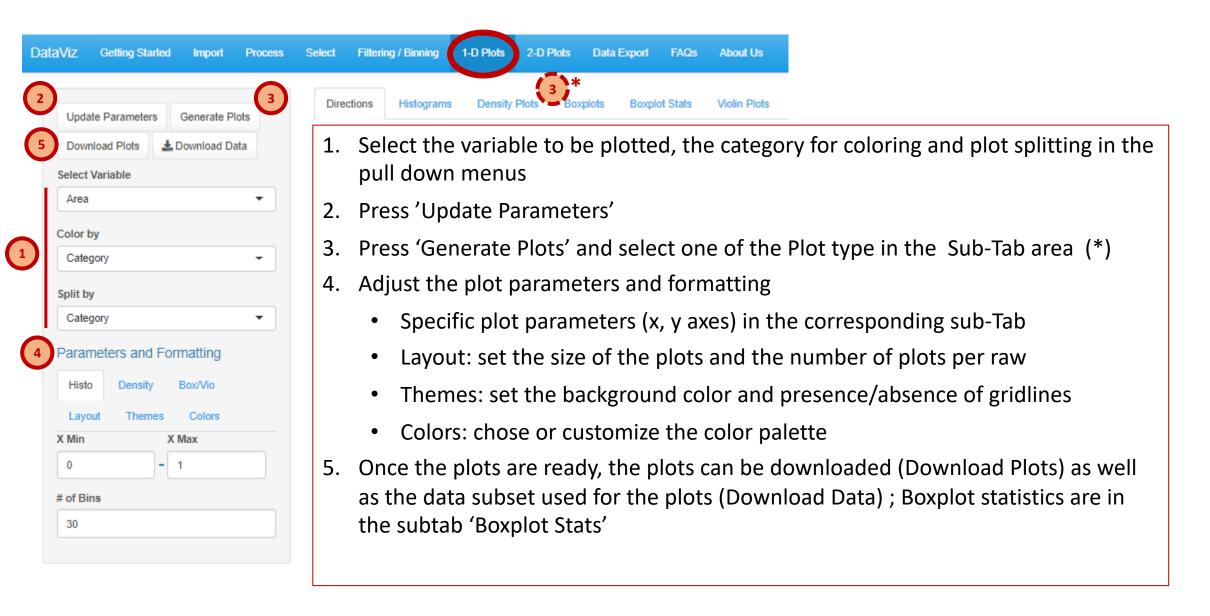


- 1. Chose the data to be handle: original or processed data
- → If the data imported are already processed (from another session), chose 'original' data
- 2. Chose the channel, object, treatment, genotype and image file to use for plotting.
- → Think carefully of what needs to be plotted.
- → Selection 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

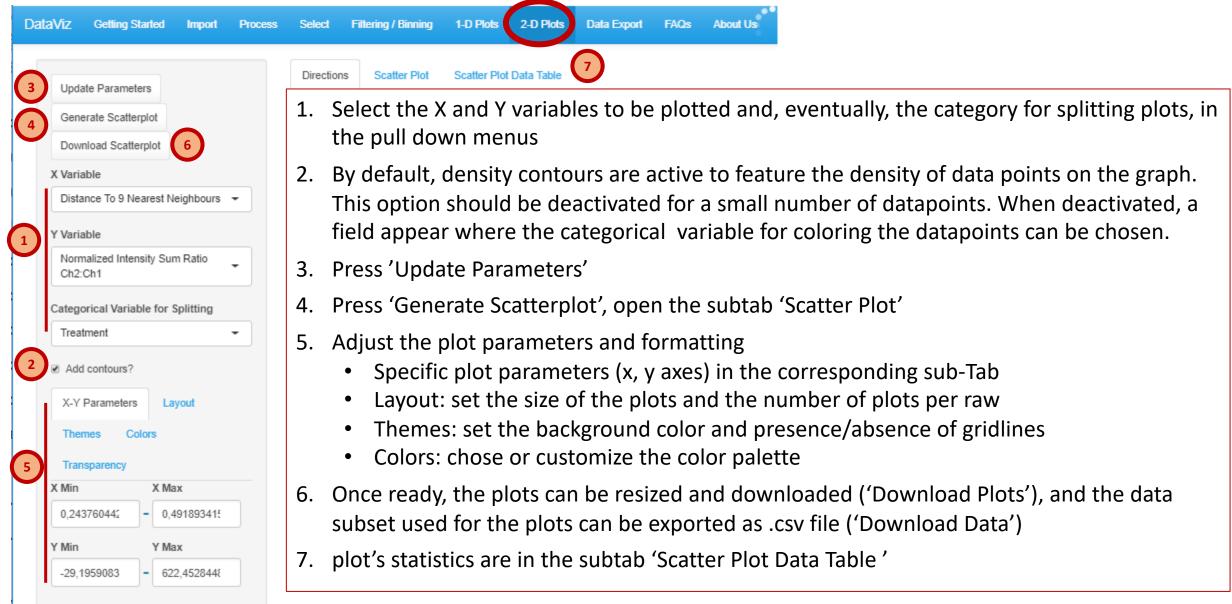
1D-Plot: presentation



1D-Plot: general directions

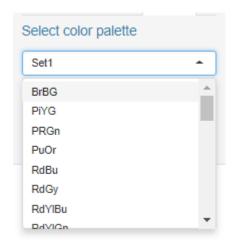


2D-Plot: general directions



Adjust plot theme and colors





Color Brewer Available Palettes

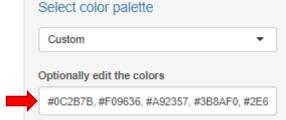


Define colors manually

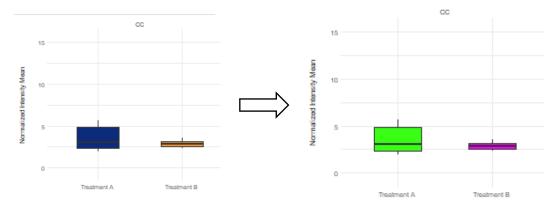
- The palettes in the pull down menu come from the Color Brewer
- The default coloring is 'Custom' corresponding to the following set:



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

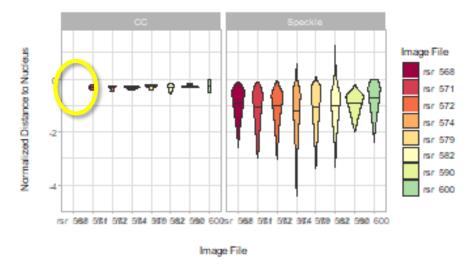


- Use https://encycolorpedia.com/ to find the heximal code of the desired color
- For instance to have neon green and fuschia copy paste #39ff14, #ff00ff in the 'optionally edit the colors' window at the beginning of the list (colors must be separated by a comma) the colors should update automatically



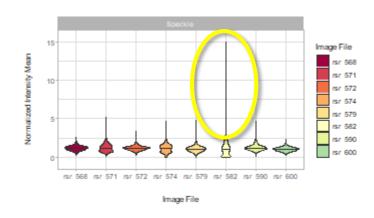
Plot per image file to verify data consistency/variability

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

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
- 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.

Filter Outliers

Before filtering

Select Data

♣ Download

plot data per treatment

Dataset for

Processed Data

Object (Type)

Channel

₽ 2

□ CC

■ Nucleolus

Nucleus

Speckle

Treatment

▼ Treatment A

▼ Treatment B

Genotypes

Image File

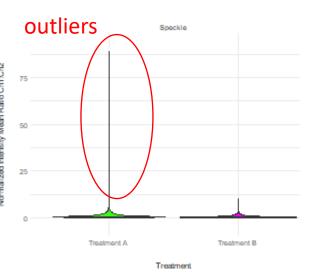
✓ rsr_568

✓ rsr_571

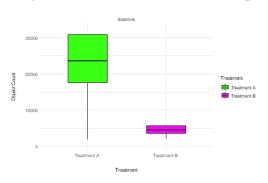
✓ WT

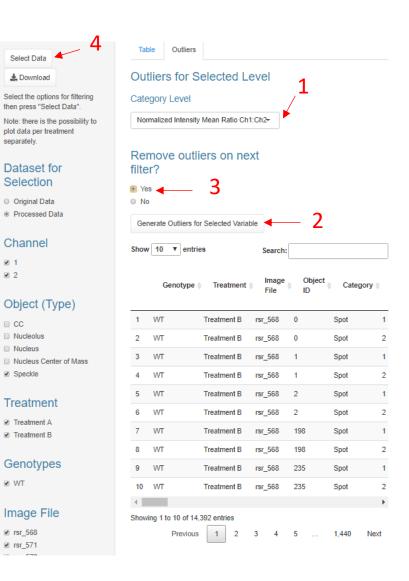
Selection Original Data

separately.

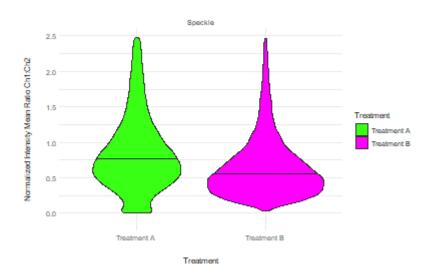


Object count (number of speckles before filtering)

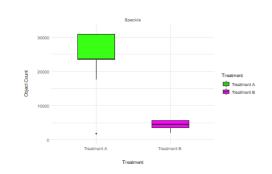




After filtering

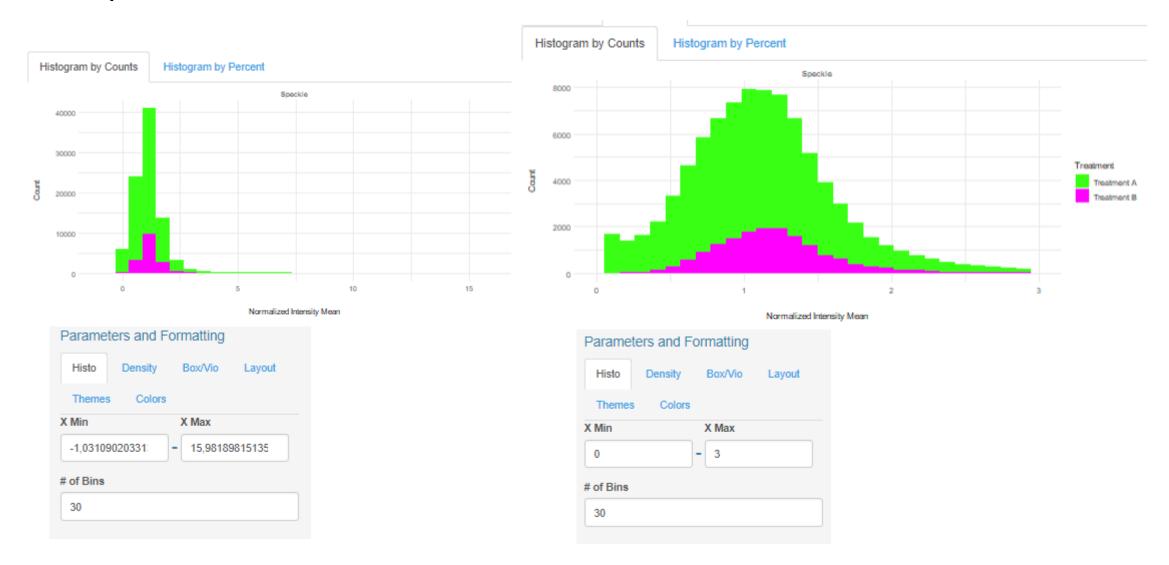


Object count (number of speckles after filtering)



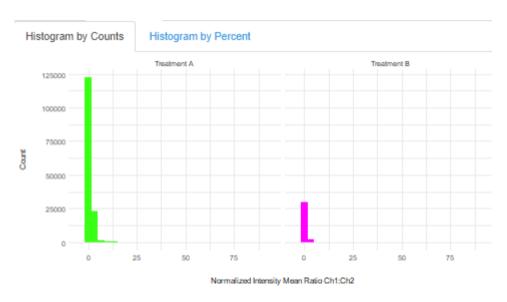
Adjust x and y limits

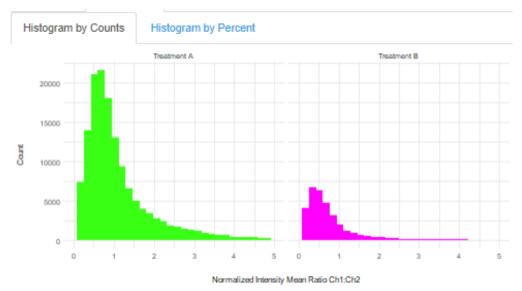
Example 1:



Adjust x and y limits

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

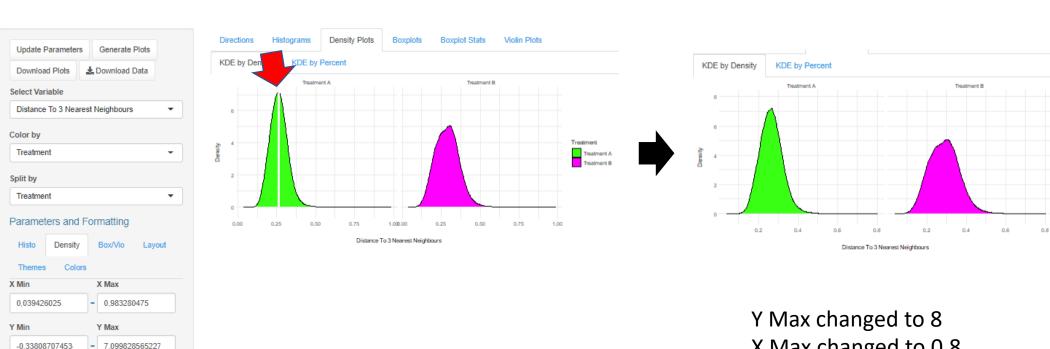




Note: this is a case where outliers should be removed as they influence the overall shape distribution. The violin plot (left) and the histogram (top) show that most values lay between 0 and 5, with outliers up to 80: -> see section 'Filter outliers'

Adjust x and y limits

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



X Max changed to 0.8

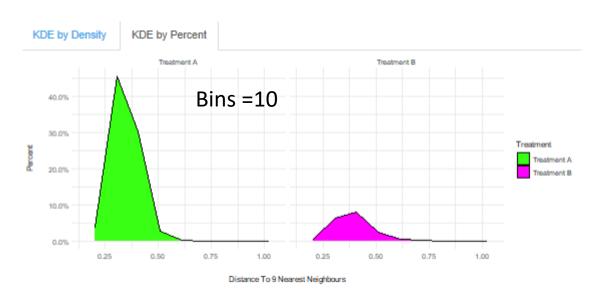
KDE Smoothing Adjust

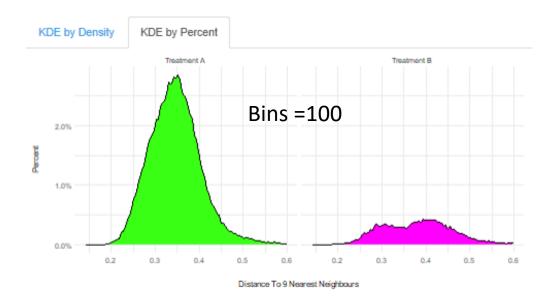
of Bins for % KDE

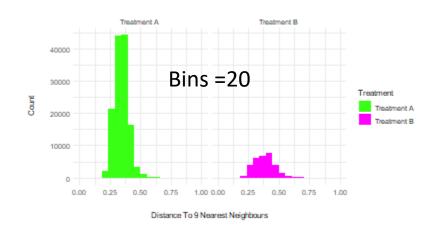
20

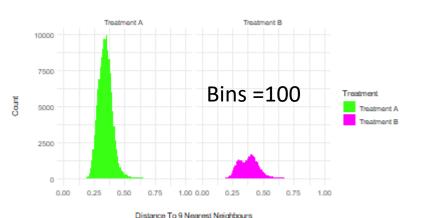
Smooth histograms and KDEs

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



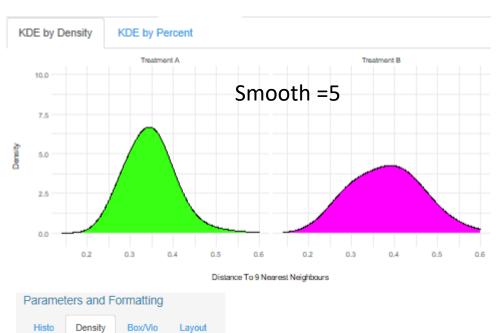


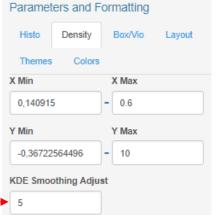




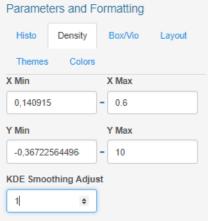
Smooth histograms and KDEs

Example 2: effect of smoothing on KDE distribution





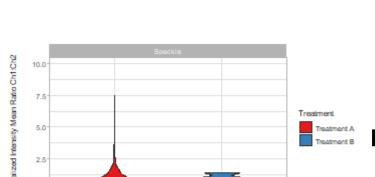




Data sub-setting: filter

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

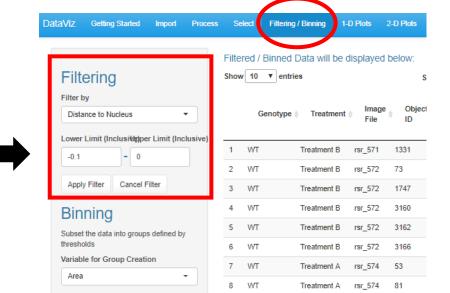


All data

Channel ratio for all speckles

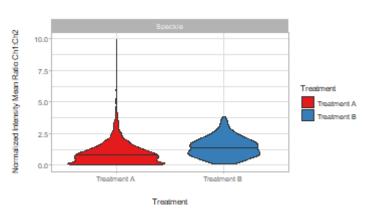
Treatment

Treatment A



- 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

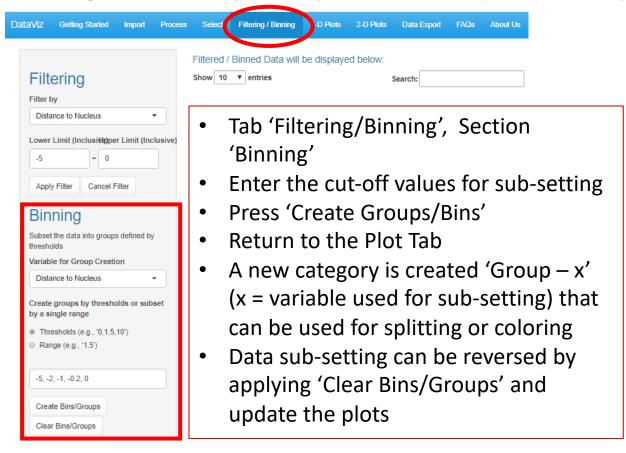
This approach allows to focus on a specific set of data points focused on the biological question

Sample Data, selected for Speckles

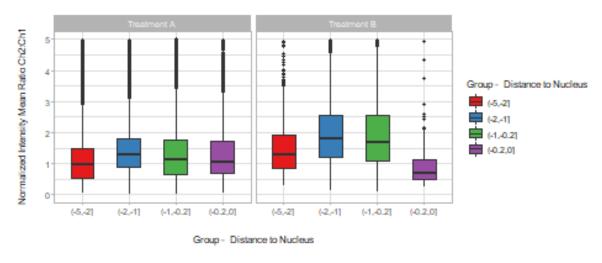
Data sub-setting: create groups

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



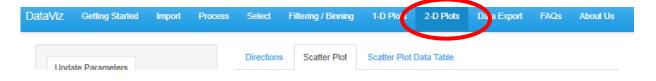
Channel intensity ratio plotted as groups according to their distance to the nucleus periphery

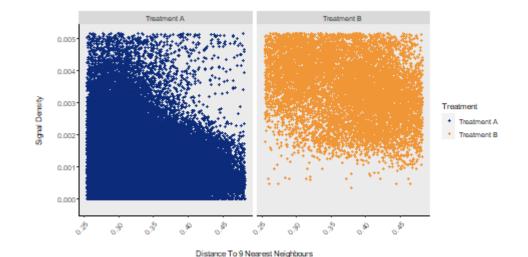


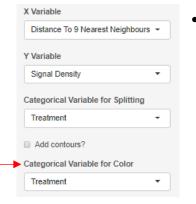
This approach reveals a group showing different intensity ratio between treatment (purple)

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

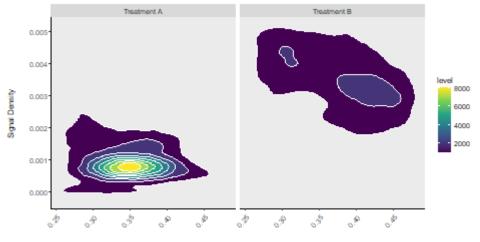
Scatter plot – datapoints or density contours







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



Distance To 9 Nearest Neighbours



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