**User Manual for “data\_augmentation&mass\_binning.r”**

(Version 1.0, Sep 29, 2020)

Shipei Xing, Tao Huan\*

Department of Chemistry, Faculty of Science, University of British Columbia, Vancouver Campus, 2036 Main Mall, Vancouver, V6T 1Z1, BC, Canada

\* Author to whom correspondence should be addressed:

Dr. Tao Huan

Tel: (+1)-604-822-4891

E-mail: thuan@chem.ubc.ca

Internet: <https://huan.chem.ubc.ca/>

* This script performs data augmentation and mass binning using the spectral library in MSP format.
* The program is written in the R language and is publicly available at https://github.com/HuanLab/SteroidXtract.
* Please see below for the code instructions.

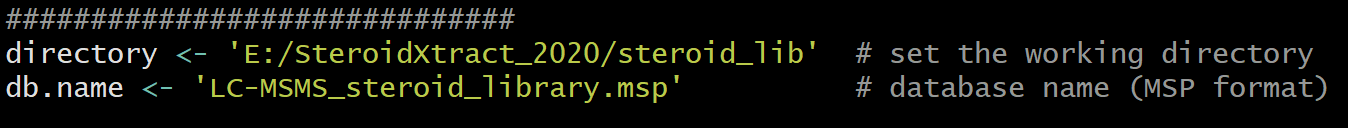
1. Prepare the spectral library in MSP format.
2. Get the R package “metaMS” installed by running the following code in Rstudio.

*if (!requireNamespace("BiocManager", quietly = TRUE))*

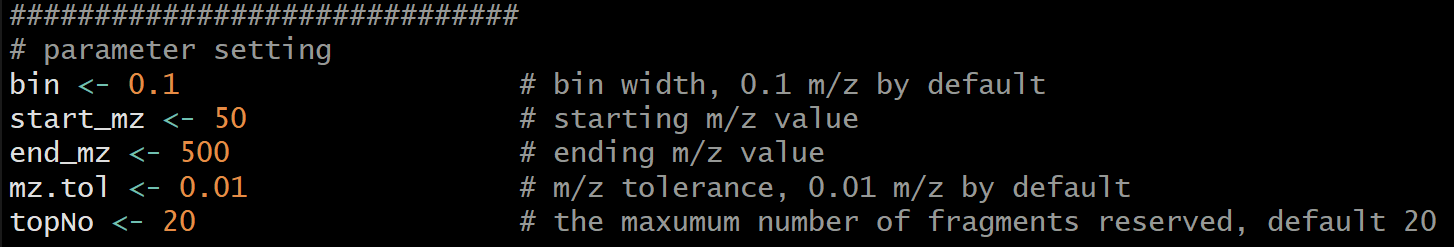
*install.packages("BiocManager")*

*BiocManager::install("metaMS")*

1. Download and open the R script “data\_augmentation&mass\_binning.r” in Rstudio.
2. Change the working directory and database name. Use “/” instead of “\” in the directory.



1. Set the parameters in the following table.



**Table.** Parameter settings.

|  |  |
| --- | --- |
| **Parameter name** | **Function** |
| bin | Bin width of mass binning method, default 0.1 *m/z*. |
| start\_mz | Starting *m/z* value of mass binning, default 50 *m/z*. |
| end\_mz | Ending *m/z* value of mass binning, default 500 *m/z*. |
| mz.tol | The *m/z* tolerance, default 0.01 *m/z*. |
| topNo | The maximum number of fragment ions reserved in each MS2 spectrum, default 20. |

1. Click “Source”, and a CSV file named “data\_matrix.csv” will be created in the working directory.