MyBox: Easy Tools Set

User Guide – Data Tools

Author: Mara

Version: 6.5.9

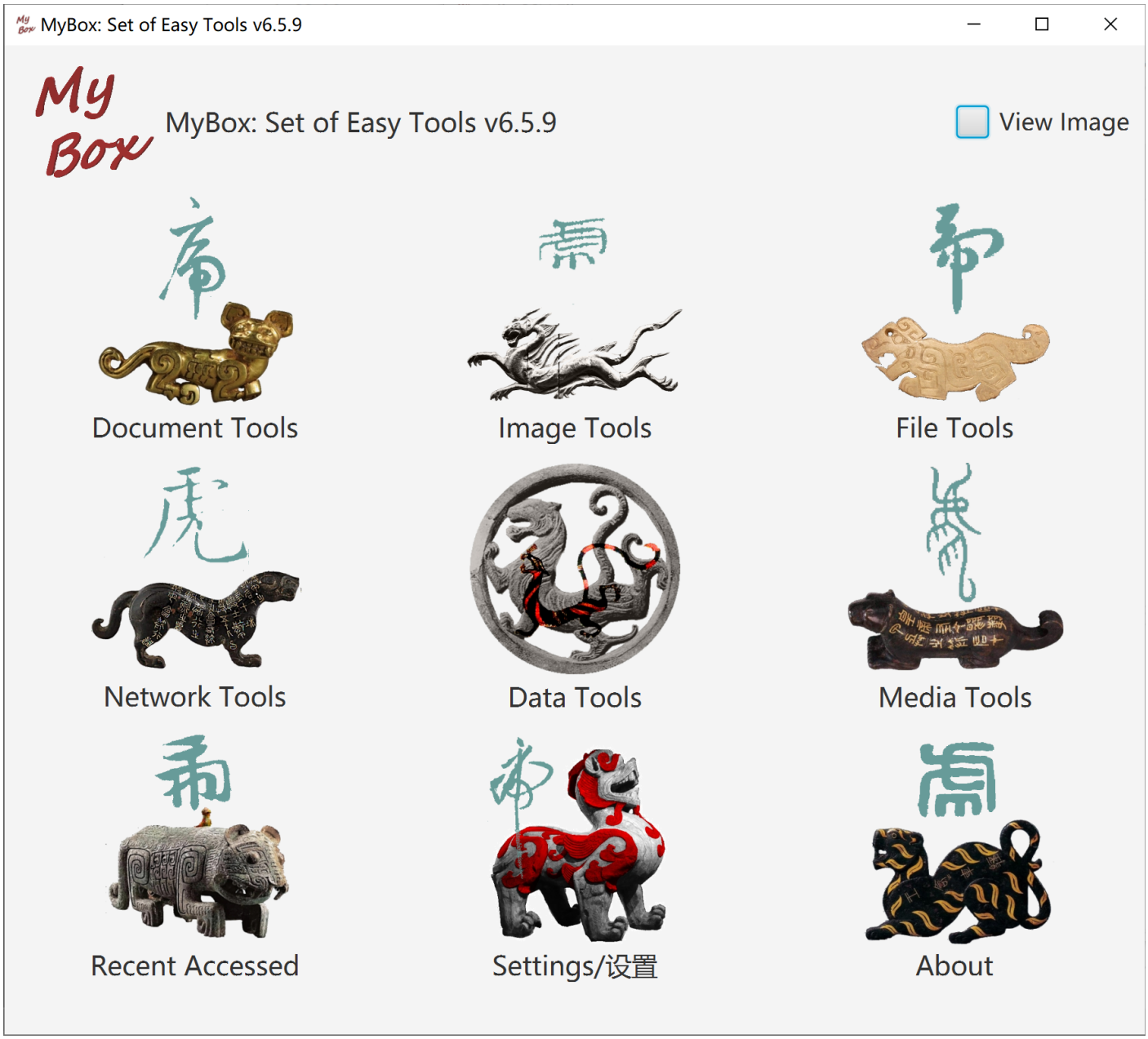
Date: 2022-8-31

Contents

1 Introduction

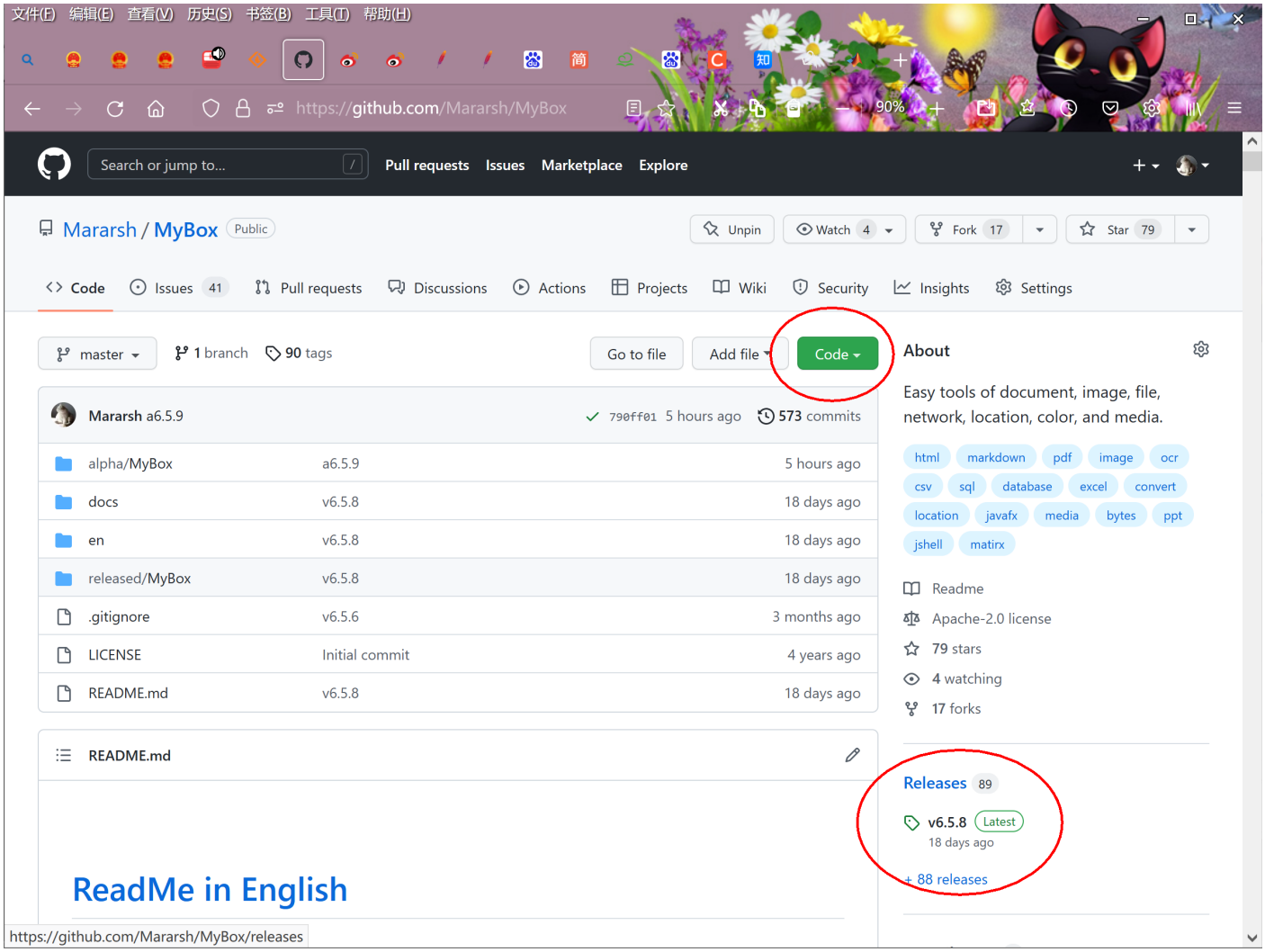
This is desktop application based on JavaFx to provide simple and easy functions. It's free and open sources.

1.1 Main Interface



1.2 Resources Addresses

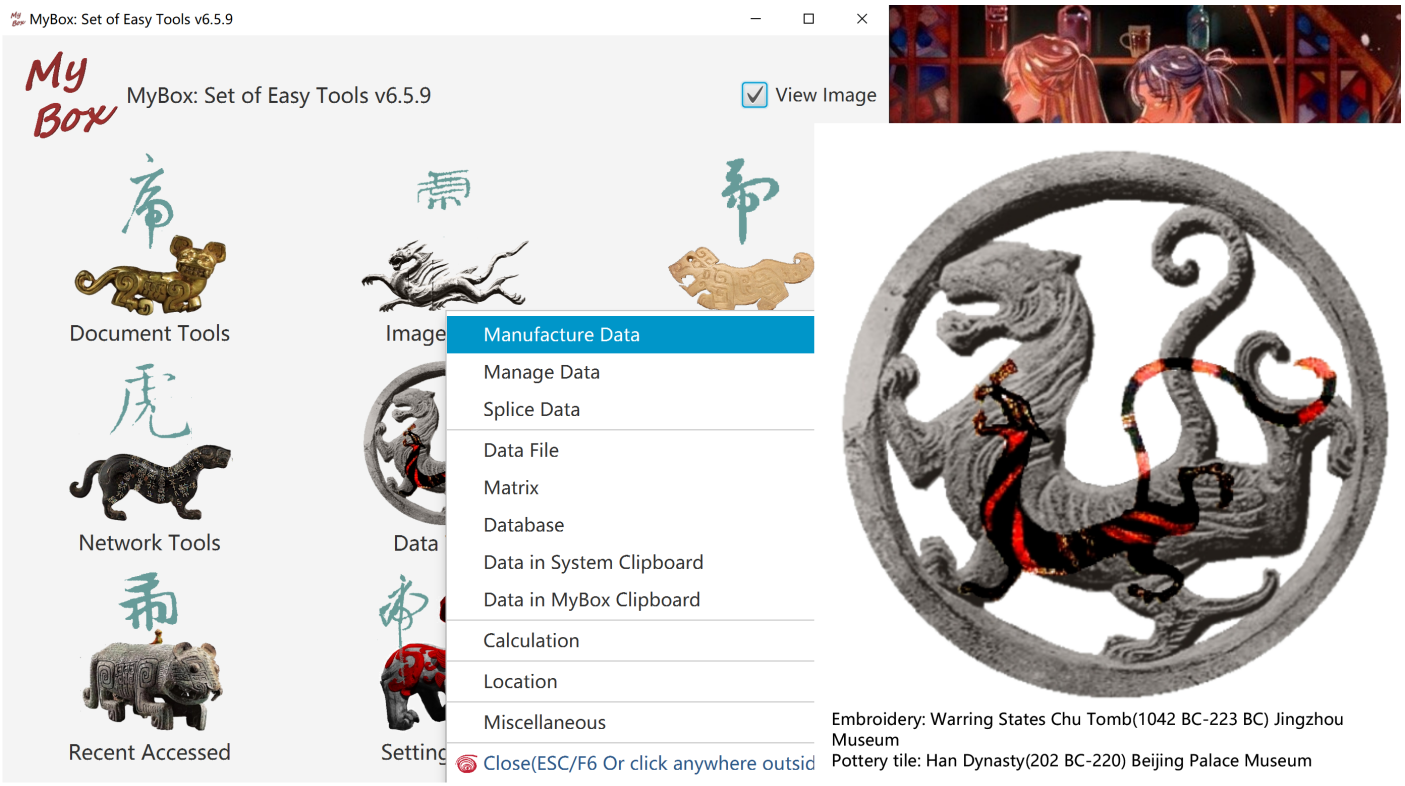
Contents	Link
Project Main Page	https://github.com/Mararsh/MyBox/
Source Codes and Compiled Packages	https://github.com/Mararsh/MyBox/releases
Submit Software Requirements and Problem Reports	https://github.com/Mararsh/MyBox/issues
Data	https://github.com/Mararsh/MyBox_data
Documents	https://github.com/Mararsh/MyBox_documents
Mirror Site	https://sourceforge.net/projects/mara-mybox/files/
Cloud Storage	https://pan.baidu.com/s/1fWMRzym_jh075OCX0D8y8A#list/path=%2F



1.3 Documents

Name	Version	Time	English	Chinese
Development Logs	6.5.9	2022-8-31	html	html
Shortcuts	6.5.6	2022-6-11	html	html
Packing Steps	6.3.3	2020-9-27	html	html
Development Guide	2.1	2020-8-27	PDF	PDF
User Guide - Overview	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide – Document Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Image Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - File Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Network Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Data Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Media Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Development Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt

1.4 Menu of Tools



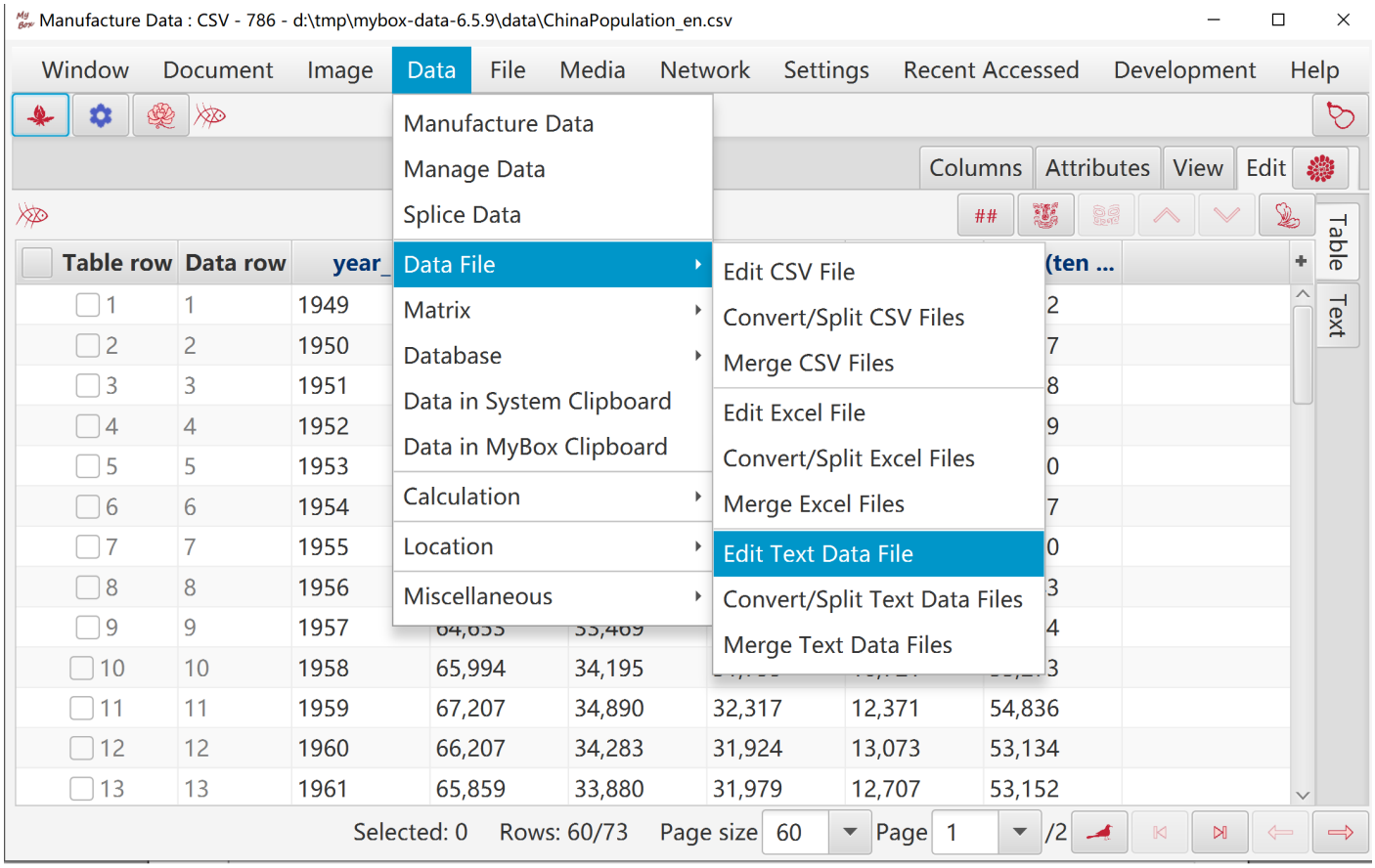
Embroidery: Warring States Chu Tomb(1042 BC-223 BC) Jingzhou Museum
Pottery tile: Han Dynasty(202 BC-220) Beijing Palace Museum

2 Two-dimensional Data

2.1 Edit Data

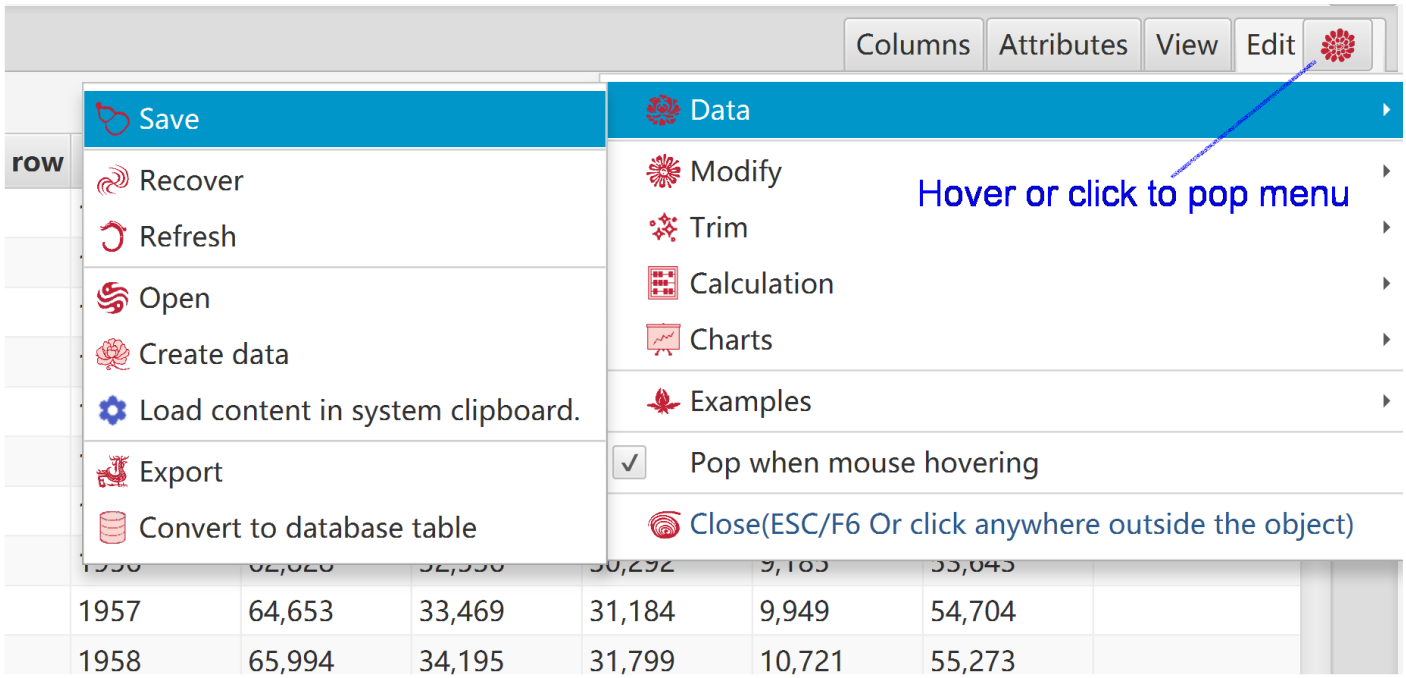
2.1.1 Data Objects

1. Following objects can be edited in consistent way:
    - Data files, including CSV File, Excel file, texts file.
    - Data of MyBox Clipboard
    - Matrices
    - Database tables.
2. Data should be in same width. That is all rows have equal number of columns.
3. Data are paginated. When pages number is larger than 1, changes should be saved before run some functions.
4. When changed, * is displayed in tab header. And ** is displayed when modifications have not applied.
5. Click button "Save" to write modifications to file and database:
    - Changes of rows in "Table", including modify/add/delete/sort, affect rows of current page in file.
    - Changes in "Columns" tab, including modify/add/delete/sort, affect all rows in file.
    - Changes of attributes and columns are saved in database.
6. Click button "Recover" to discard all modifications and load data from file and database.



2.1.2 Functions Menu

Hover or click button “Functions” top pop Functions Menu.



2.1.3 Define Columns

1. Column names should not be null nor duplicated.
2. Data types are used to validate data values:
   - Invalid value is rejected when edit data.
   - Type is ignored when read or calculate data.
   - Data type affects sorting results.
3. Click button "OK" to apply it modifications to "Table".
4. Click button "Cancel" to discard its modifications and pick data from "Table".
5. Can rename all columns with sequence numbers.
6. Can set random colors.

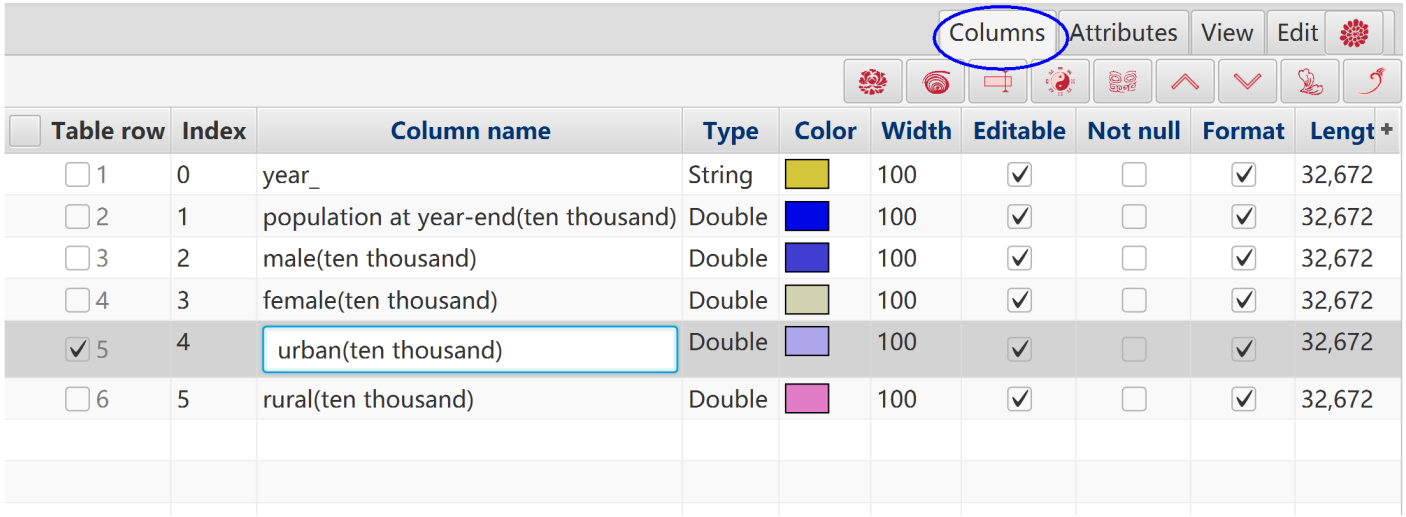
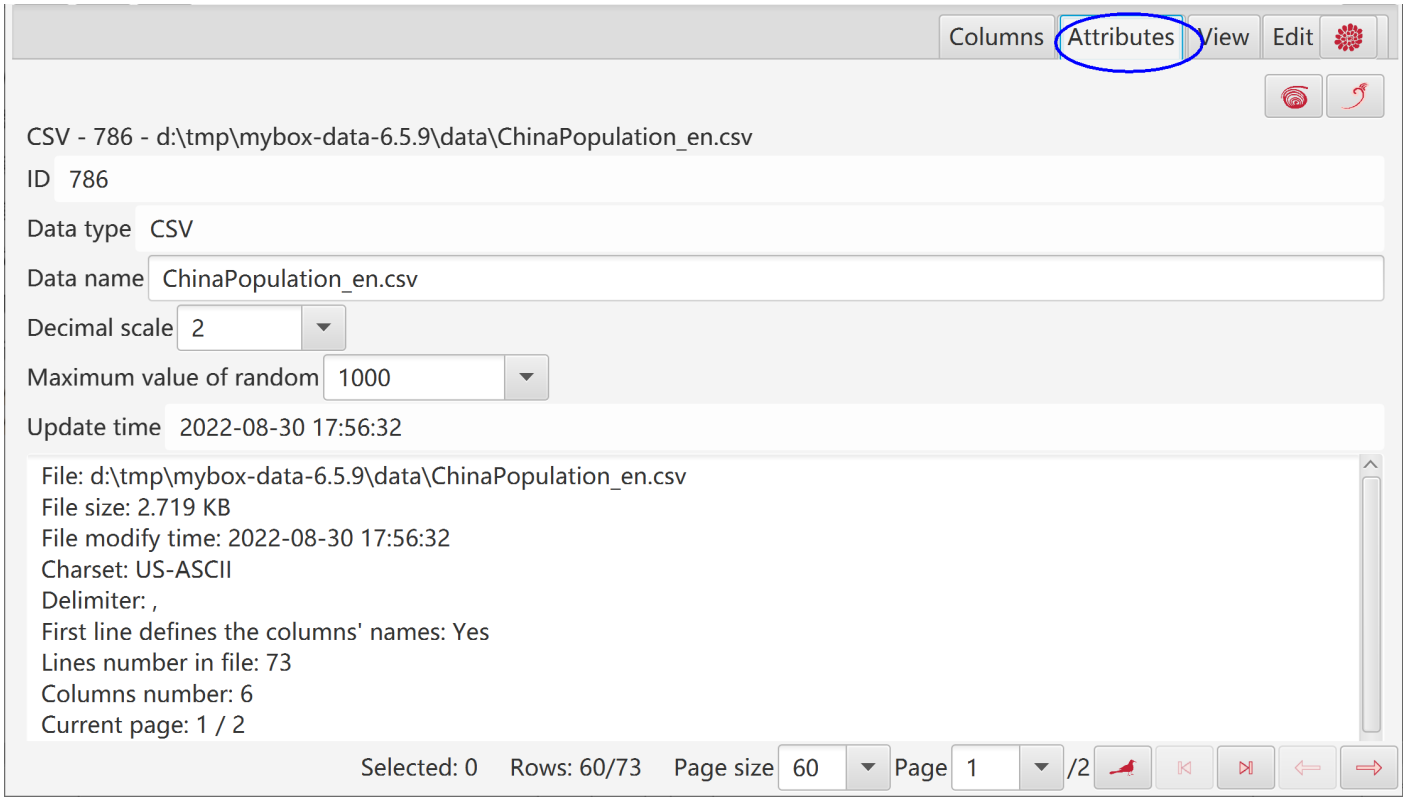


Table row	Index	Column name	Type	Color	Width	Editable	Not null	Format	Lengt
1	0	year_	String		100	☑	☐	☑	32,672
2	1	population at year-end(ten thousand)	Double		100	☑	☐	☑	32,672
3	2	male(ten thousand)	Double		100	☑	☐	☑	32,672
4	3	female(ten thousand)	Double		100	☑	☐	☑	32,672
☑ 5	4	urban(ten thousand)	Double		100	☑	☐	☑	32,672
6	5	rural(ten thousand)	Double		100	☑	☐	☑	32,672

2.1.4 Define Attributes

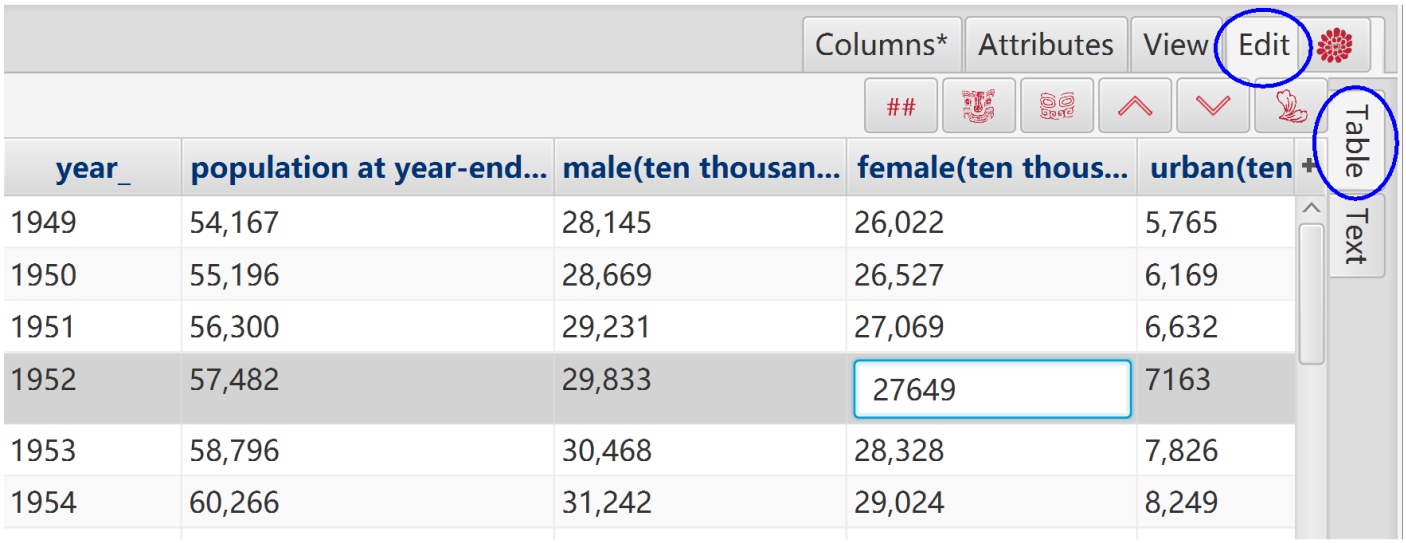
Data name, decimal scale, and maximum value of random.



2.1.5 "Table" Edit Mode

"Table" is the master edit mode:

1. Its modifications are applied to other panes automatically.
2. It is the final data to save.

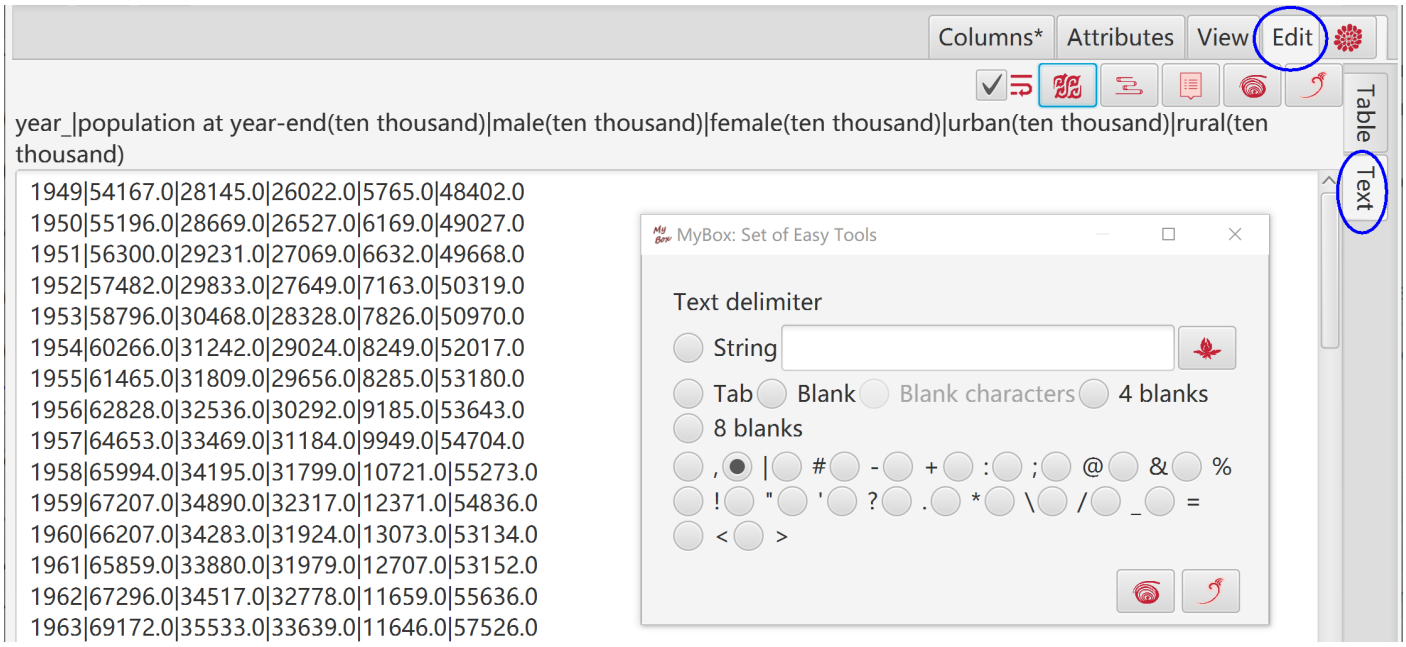


year_	population at year-end...	male(ten thousan...	female(ten thous...	urban(ten
1949	54,167	28,145	26,022	5,765
1950	55,196	28,669	26,527	6,169
1951	56,300	29,231	27,069	6,632
1952	57,482	29,833	27649	7163
1953	58,796	30,468	28,328	7,826
1954	60,266	31,242	29,024	8,249

2.1.6 "Text" Edit Mode

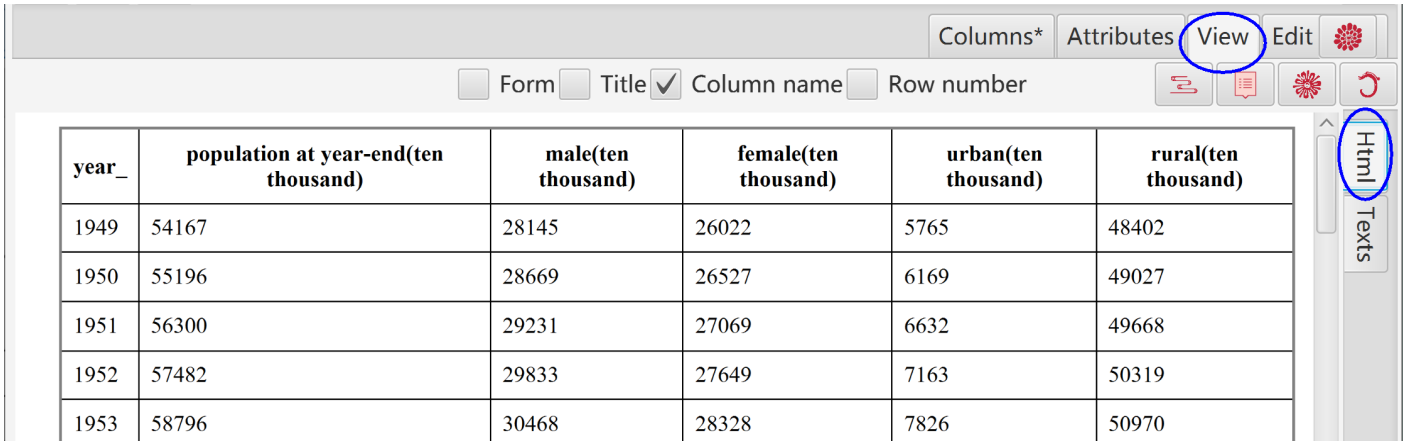
"Text" is the assist edit mode.

1. Click button "OK" to apply its modifications to "Table".
2. Click button "Cancel" to discard its modifications and pick data from "Table".
3. Click button "Delimiter" to pick data from "Table" and apply new delimiter while its modifications are discarded.



2.1.7 View Html

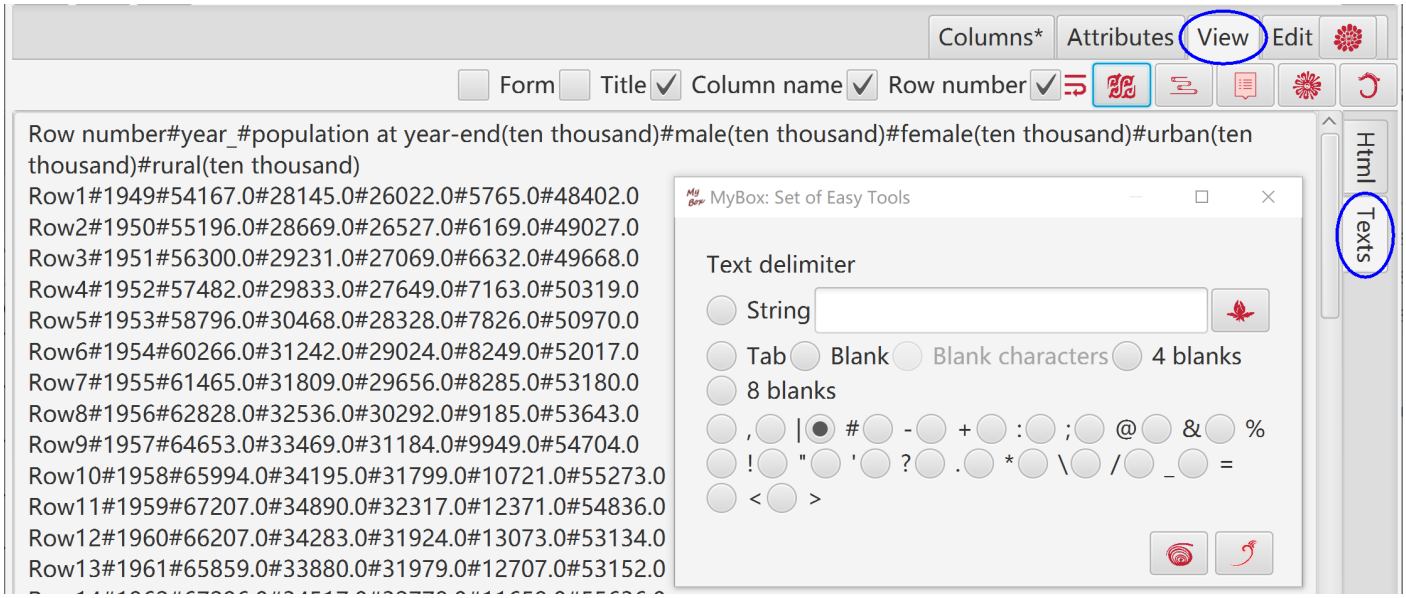
1. View data of current data.
2. Options: Form, Title, Column Names, Row Numbers.



year_	population at year-end(ten thousand)	male(ten thousand)	female(ten thousand)	urban(ten thousand)	rural(ten thousand)
1949	54167	28145	26022	5765	48402
1950	55196	28669	26527	6169	49027
1951	56300	29231	27069	6632	49668
1952	57482	29833	27649	7163	50319
1953	58796	30468	28328	7826	50970

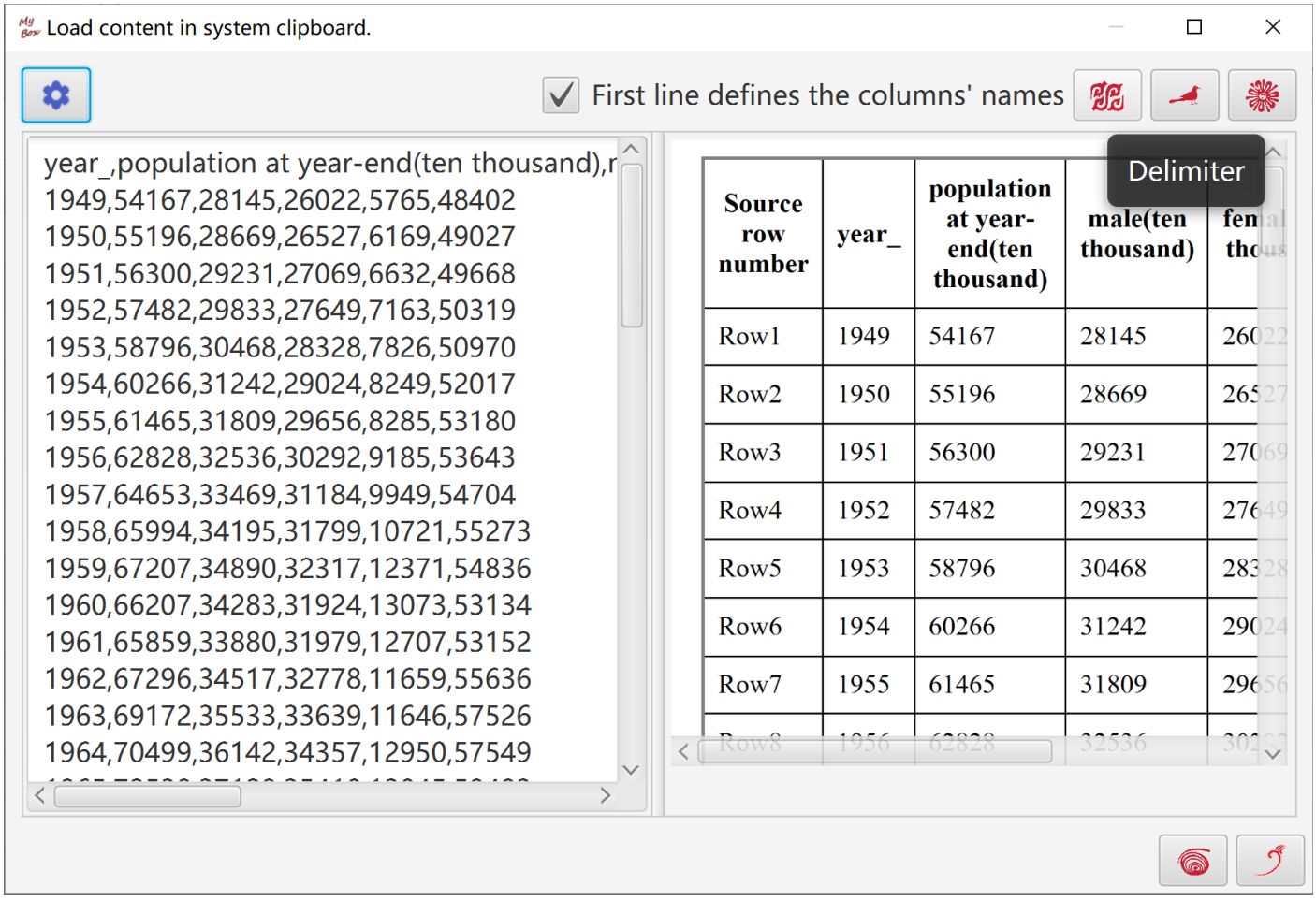
2.1.8 View Text

1. View data of current data.
2. Options: Form, Title, Column Names, Row Numbers.
3. Set delimiter.



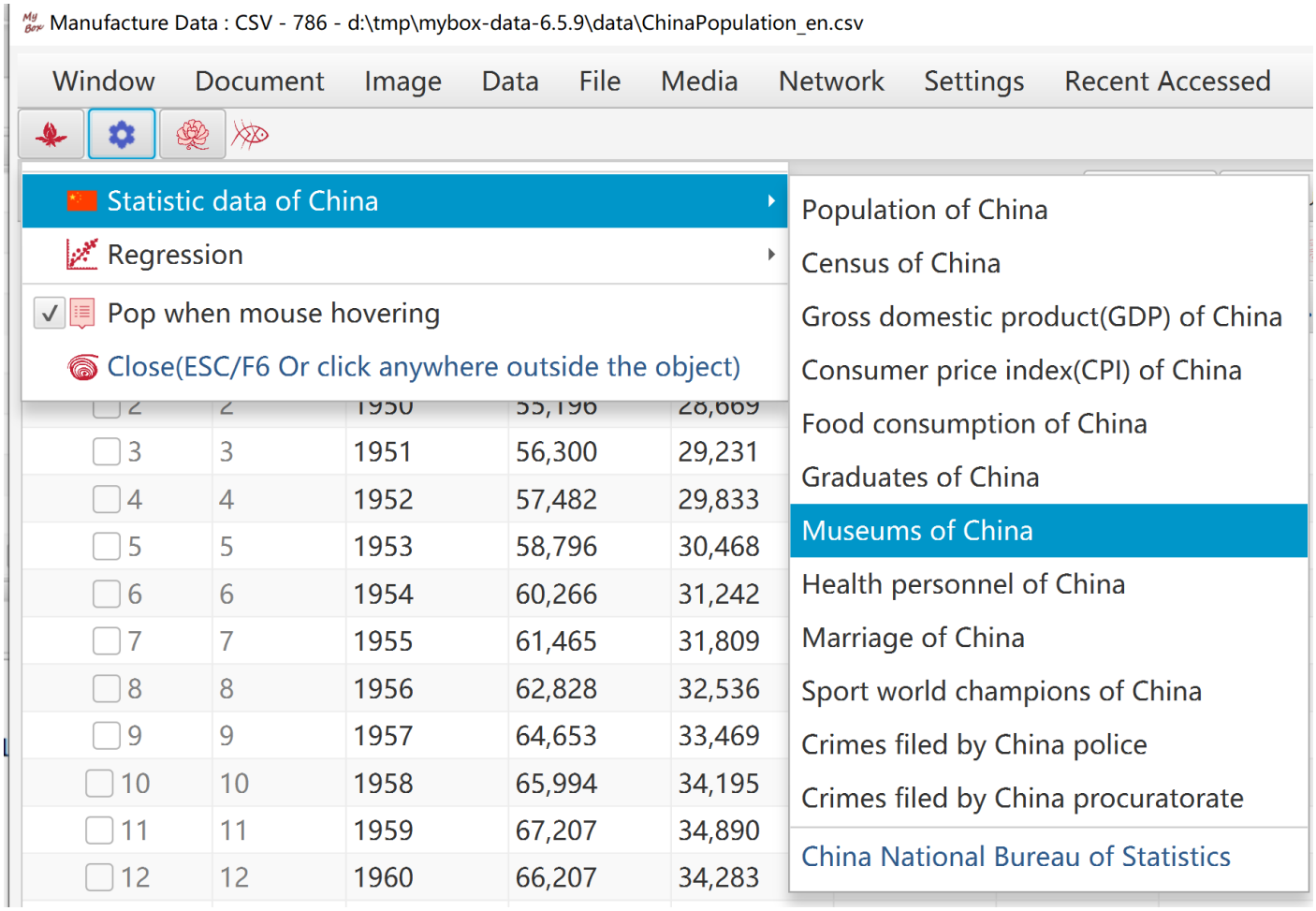
2.1.9 Load Contents in System Clipboard

1. Read and parse contents in system clipboard.
2. Guess delimiter automatically.
3. Delimiter can be chosen from special characters or inputted regular expression.
4. First row can be set as column names.

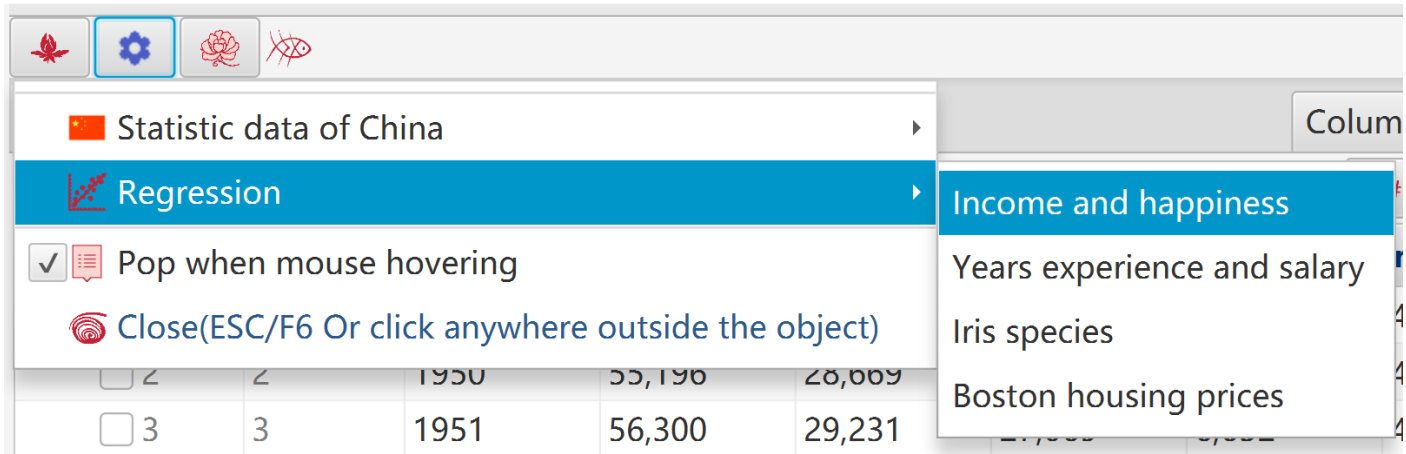


2.1.10 Import Examples

2.1.10.1 Statistic Data of China



2.1.10.2 Data of Regression



2.2 Row Expression

JavaScript expression can be data values when manufacture/trim/calculate data or generate chart:

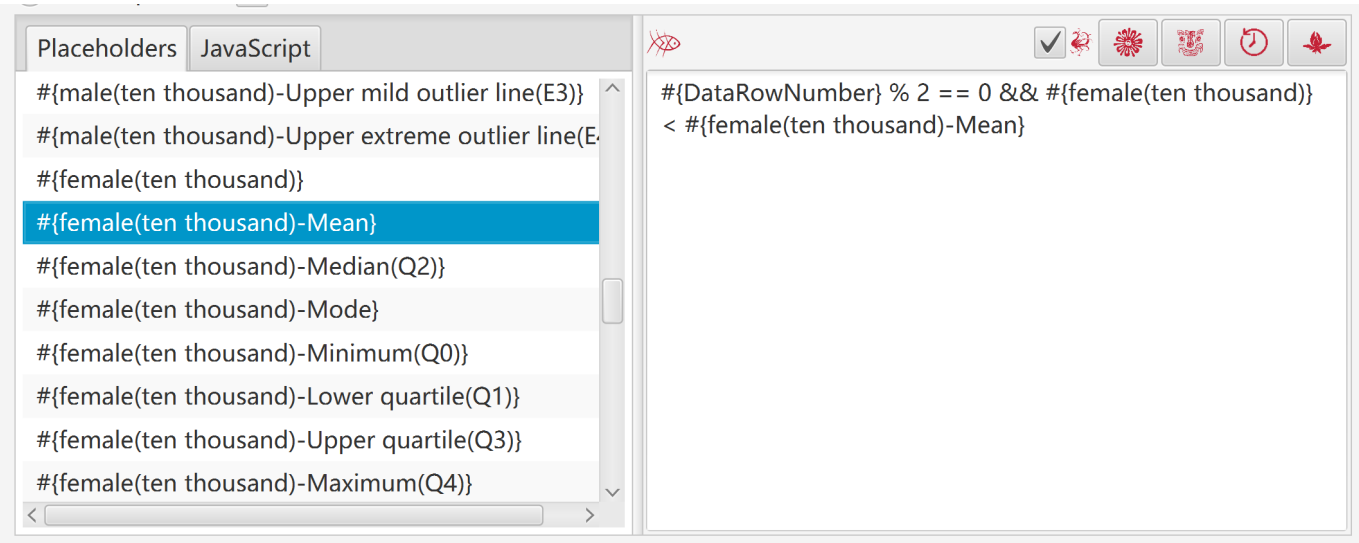
1. If the script is blank, then return empty string.
2. Edit the script:
   - It can include any valid JavaScript elements.
   - It should return a value finally.
   - It can include following placeholders:

     #{TableRowNumber}

     #{DataRowNumber}

     #{<column_name>}

     #{<column_name>- }
3. When MyBox evaluates the expression:
   - Placeholders are replaced with actual values of each data row .
   - '#{xxx}' is handled as string while #{xxx} is handled as number.
   - When handles all pages, script fails when it includes “#{TableRowNumber}” .
4. Hover or click button “Examples” to paste example codes.
5. Hover or click button “Histories” to paste codes in histories.



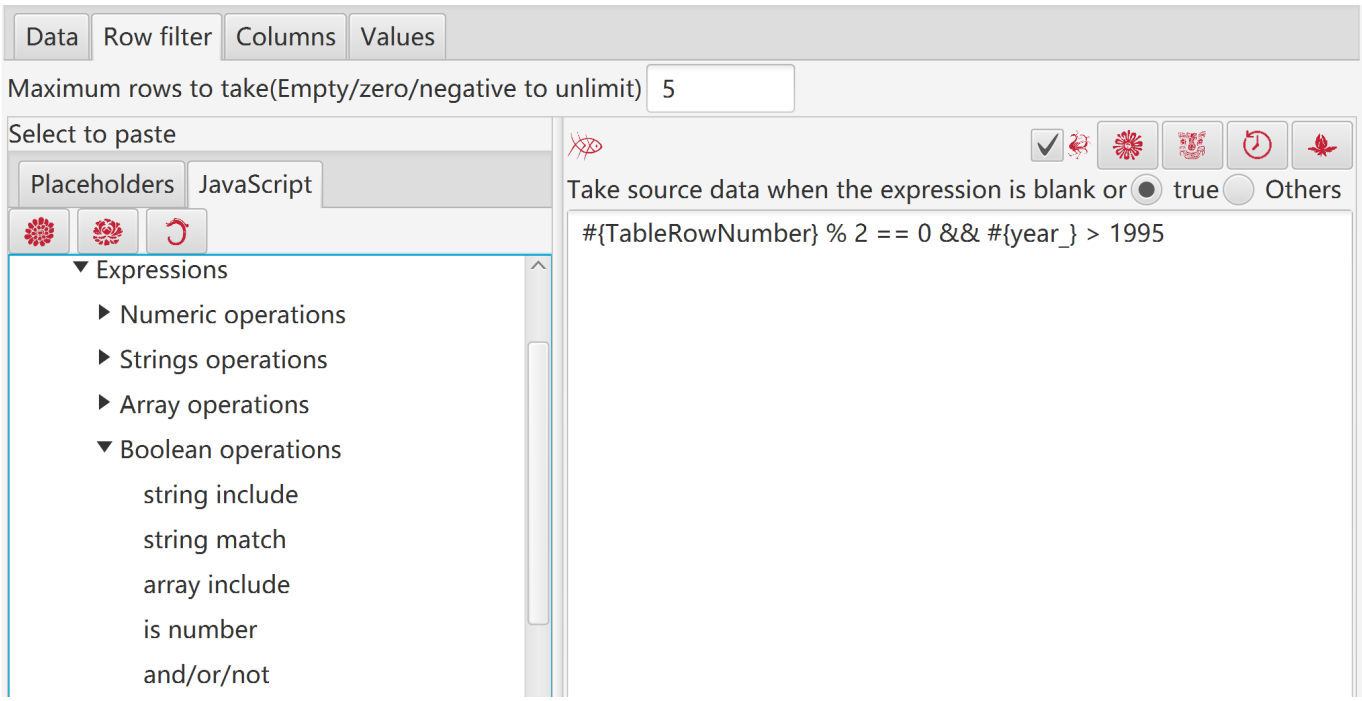
6. Examples:

expression	meaning
#{DataRowNumber} % 2 == 0	data row number is even
#{TableRowNumber} % 2 == 1	odd rows in current page
Math.abs(#{v1}) + Math.PI * Math.sqrt(#{v2})	calculation
'#{v1}'.replace(/hello/ig, 'Hello')	replace all “hello”(case-insensitive) as “Hello” in column “v1”
'#{v1}'.toLowerCase()	lower case of value of column “v1”
'#{v1}'.split(‘,’)	split value of column “v1” by comma
#{v1} - #{v1-Mean}	difference between value of column “v1” and mean of column “v1”

2.3 Row Filter

"Row Filter" is special "Row Expression", and can be condition to filter data rows.

1. It should return boolean value("true" or "false") finally.
2. Can set maximum rows.



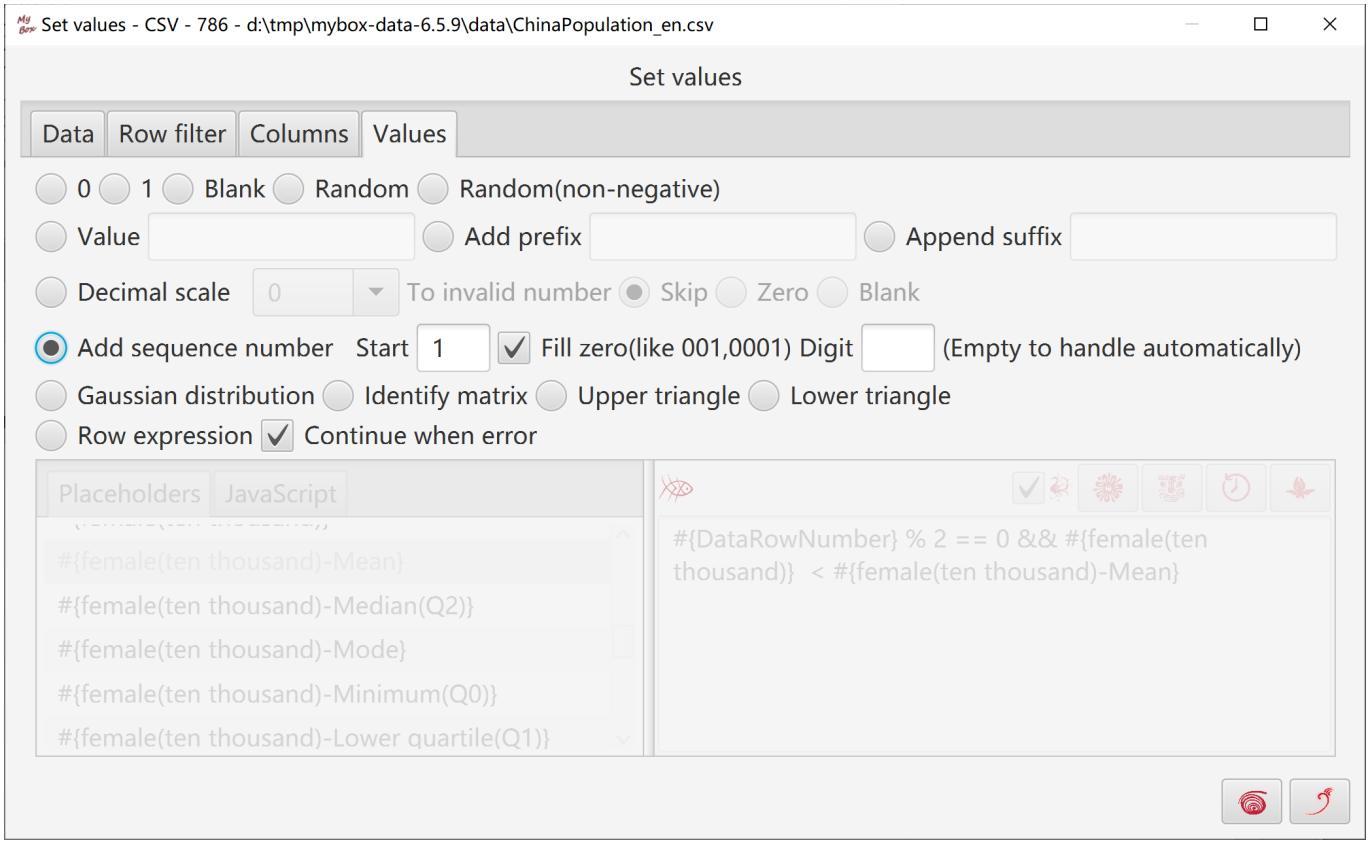
3. Exmaples:

Expression	meaning
#{DataRowNumber} % 2 == 0	data row number is even
#{TableRowNumber} % 2 == 1	odd rows in current page
Math.abs(#{v1}) >= 0	value of column "v1" is number
#{v1}) > 0	value of column "v1" is larger than zero
#{v1} - #{v2} < 100	difference between values of "v1" and "v2" is less than 100
'#{v1}'.length > 0	value of column "v1" is not empty
'#{v1}'.search(/Hello/ig) >= 0	value of column "v1" includes "Hello"(case-insensitive)
'#{v1}'.startsWith('Hello')	value of column "v1" starts with "Hello"
var array = ['A', 'B', 'C'];array.includes('#{v1}')	value of column "v1" is one of "A", "B", "C"
#{v1} < #{v1-Mean}	value of column "v1" is less than mean of column "v1"

2.4 Modify

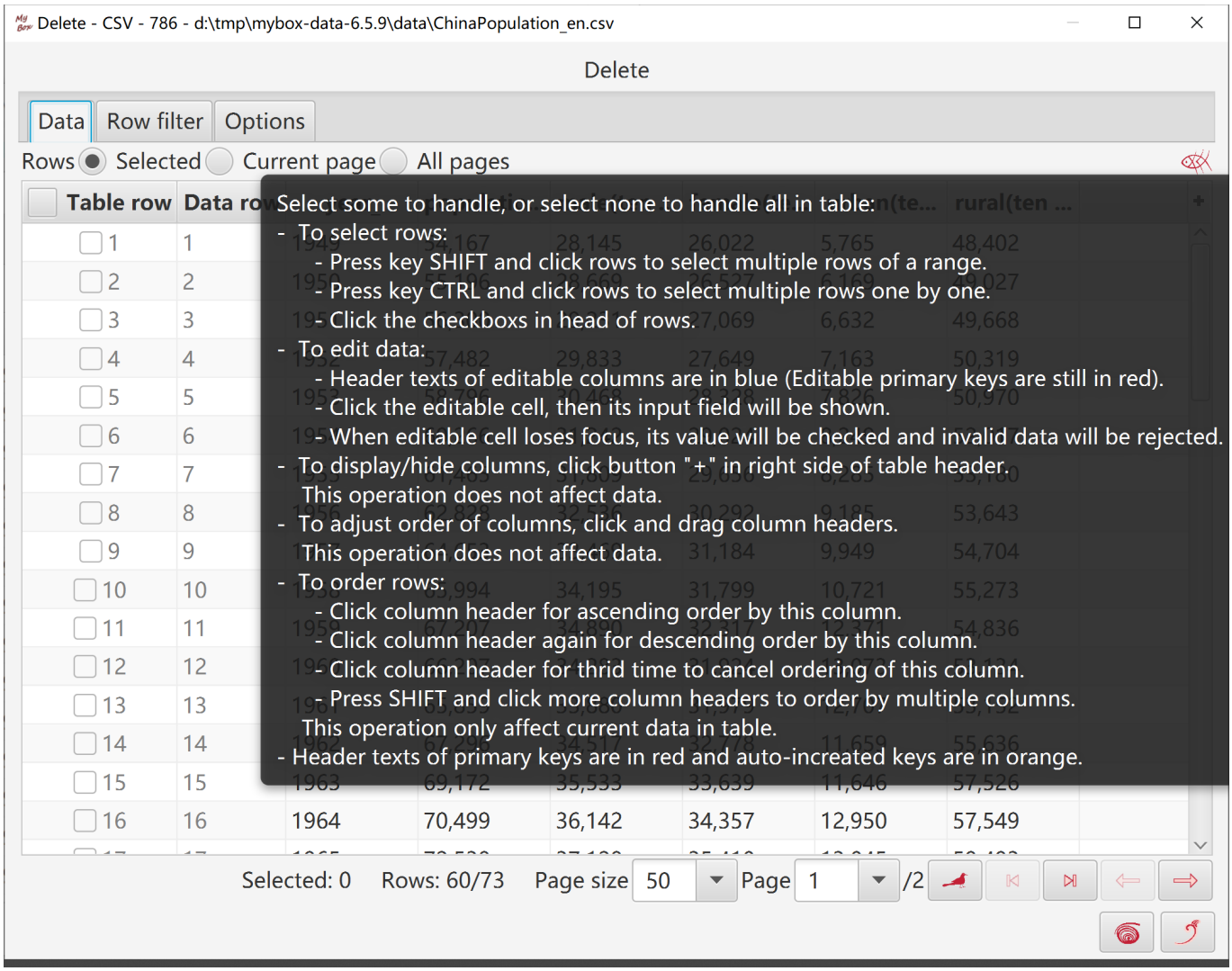
2.4.1 Set Values

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Set values of selected data as following:
   - Constant：0, 1, blank, or inputted value
   - Random, random of non-negative
   - Add prefix, append suffix, set decimal scale, or add sequence numbers
   - When selected data are sqaure, whose rows number equals to columns number, they can be set as following: gaussian distribution, identify matrix, upper triangular matrix, lower triangular matrix.
   - Row expression
3. If handle all pages of data file, then auto-backup before set values.



2.4.2 Delete

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Set row filter.
2. Option: Whether continue when error.
3. If handle all pages of data file, then auto-backup before delete.



2.4.3 Set styles / Mark Abnormal Values

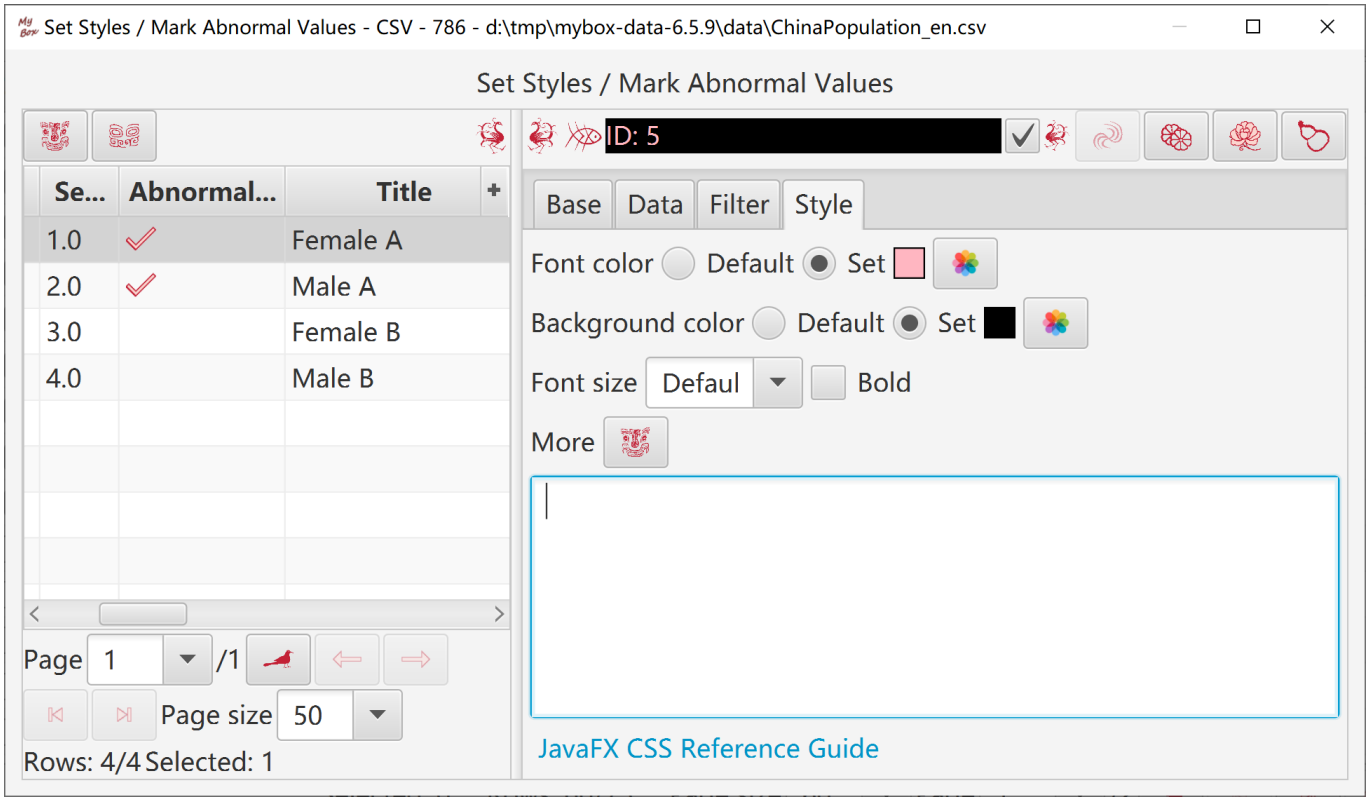
1. Add/Edit/Delete styles.
2. Define conditions to determine which data cells to apply the style:
   - Range of data rows
   - Column names.
   - Row filter.

     Notice, data of a row number may be changed when some rows are added or deleted.

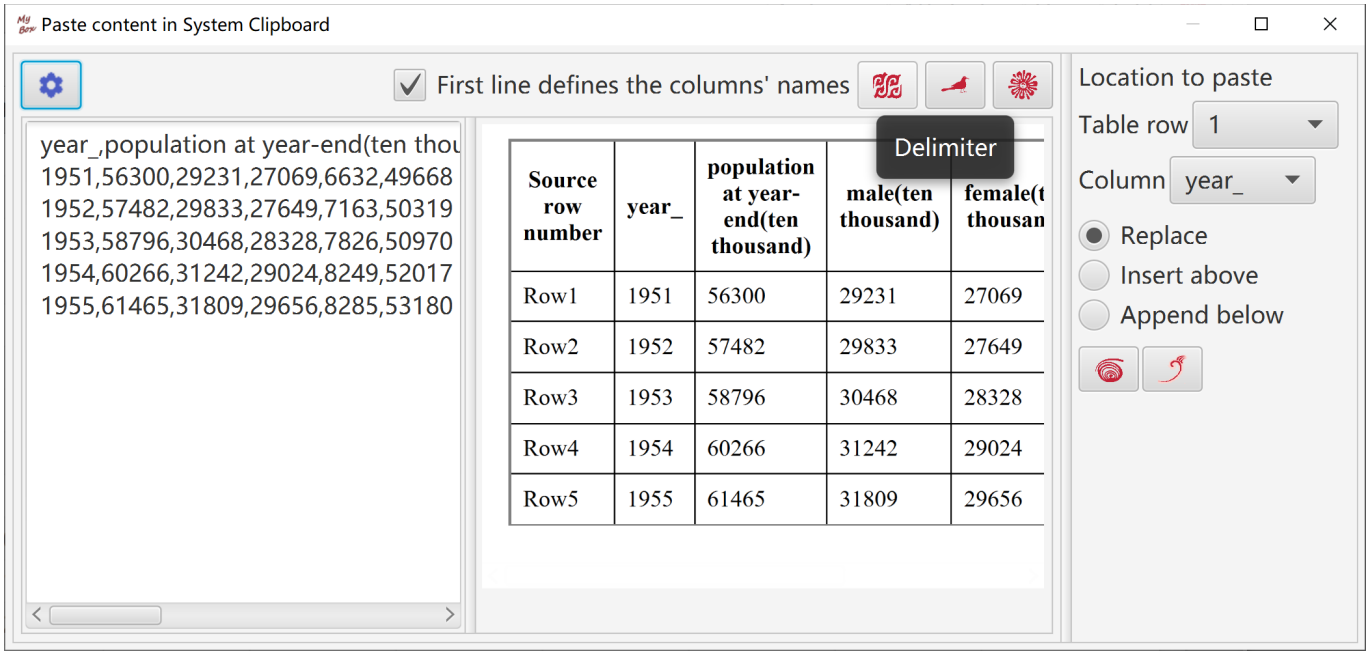
     Example, when insert 2 rows before "row 6", original "row 12" becomes "row 14" while current "row 12" was "row 10".

     So "row number" is not right way to locate a specific data row while rows number is changing.

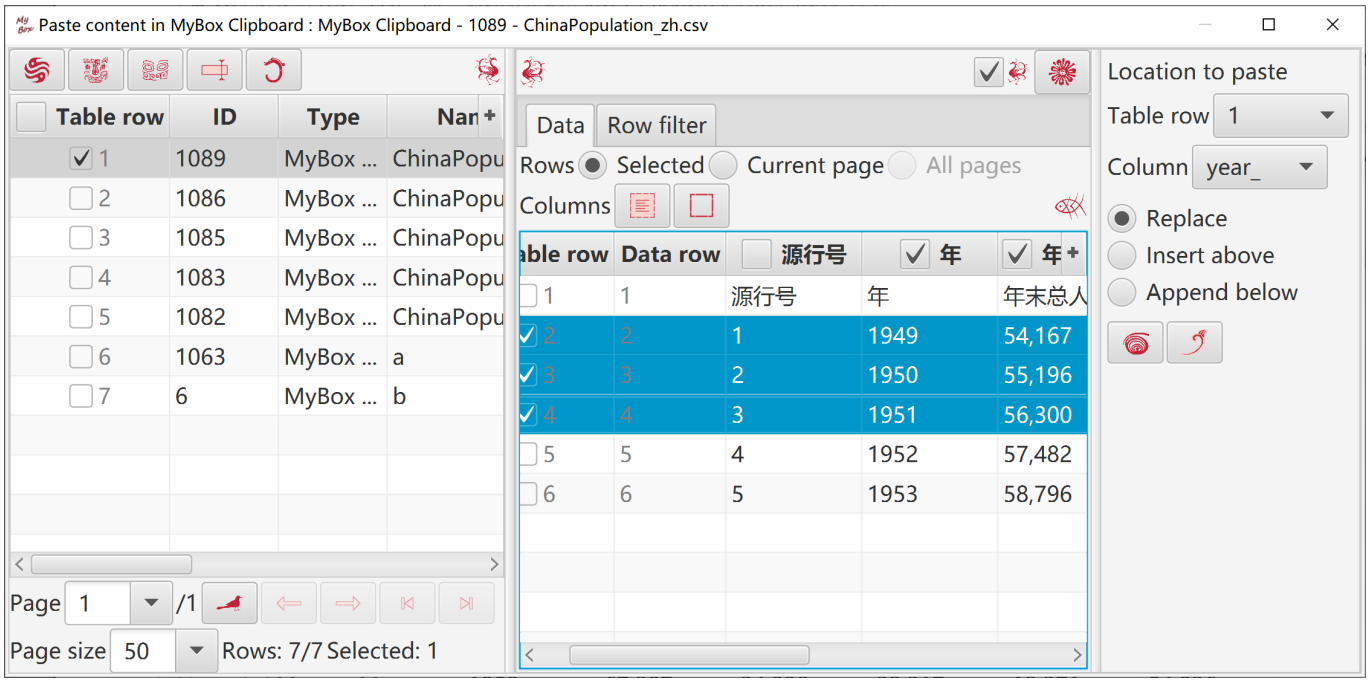
     A way to refer sepcial rows is the expression composed of column values.
3. Define style values:
   - Font color, font size, background color, bold, etc.
   - More values in format of JavaFx CSS.
4. Define title and sequence number of the style.
5. Set whether the style marks abnormal values.
6. All styles are applied to the data one by one in order of their sequence number.



2.4.4 Paste Content in System Clipboard



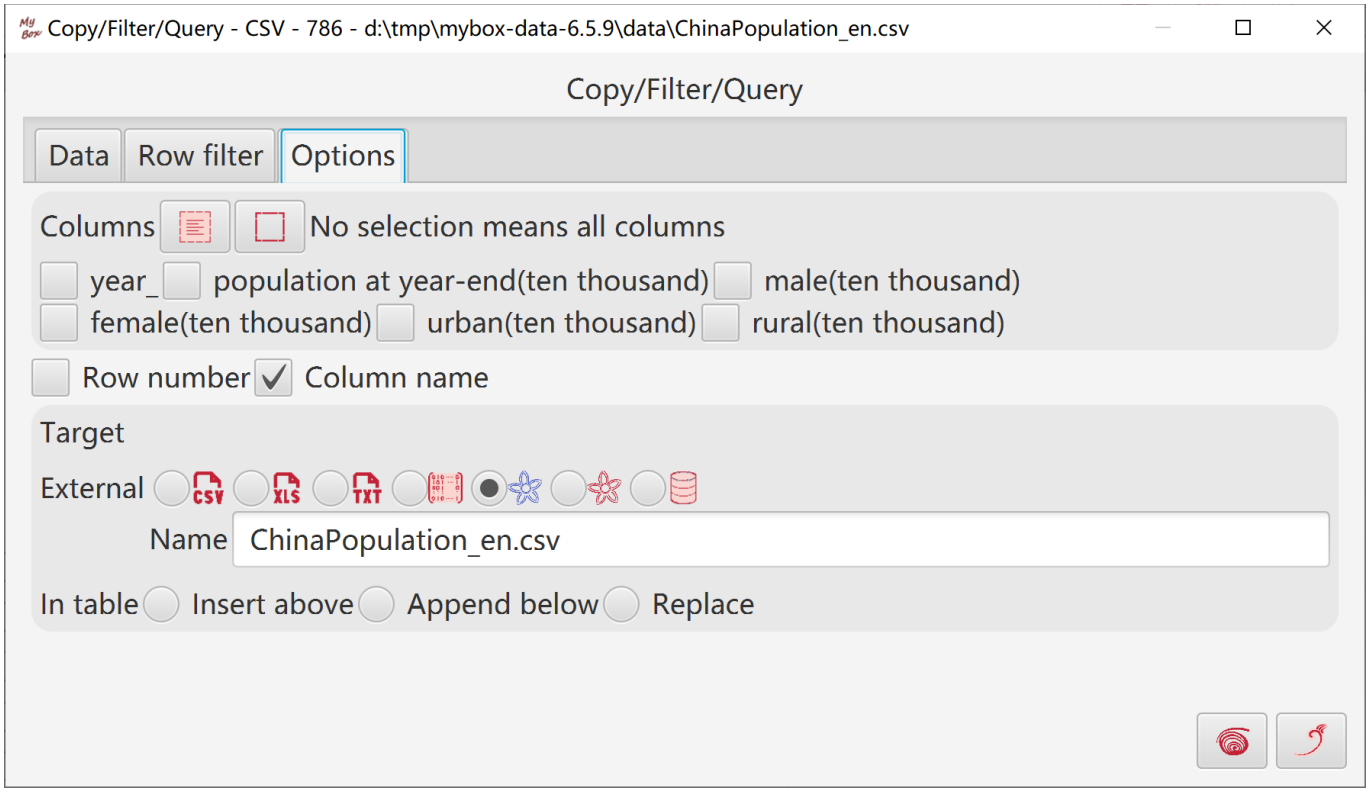
2.4.5 Paste Content in MyBox Clipboard



2.5 Trim

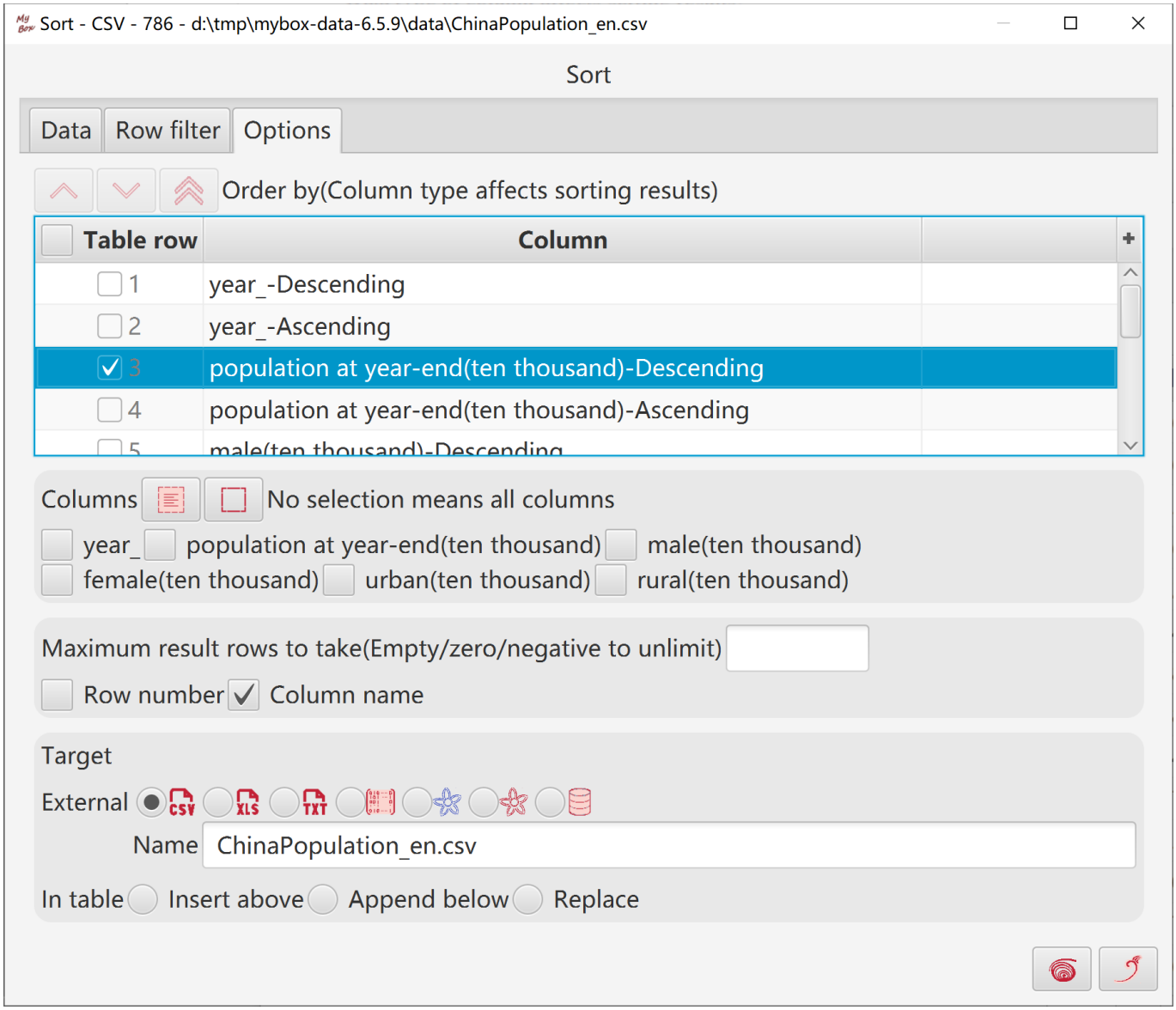
2.5.1 Copy/Filter/Query

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
3. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



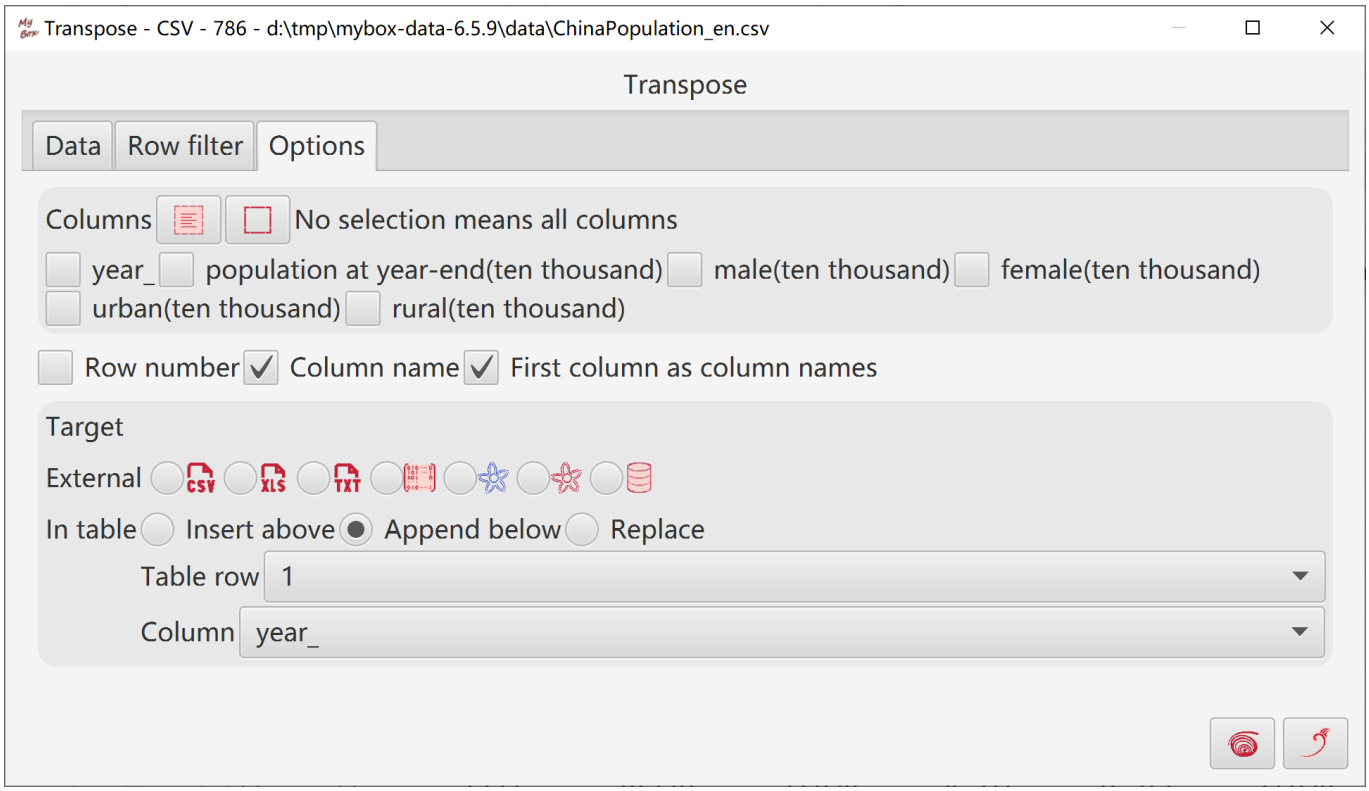
2.5.2 Sort

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Select the column to sort and whether descending.
   - Data type of column affects sorting results.
   - Maximum rows number of results can be set.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



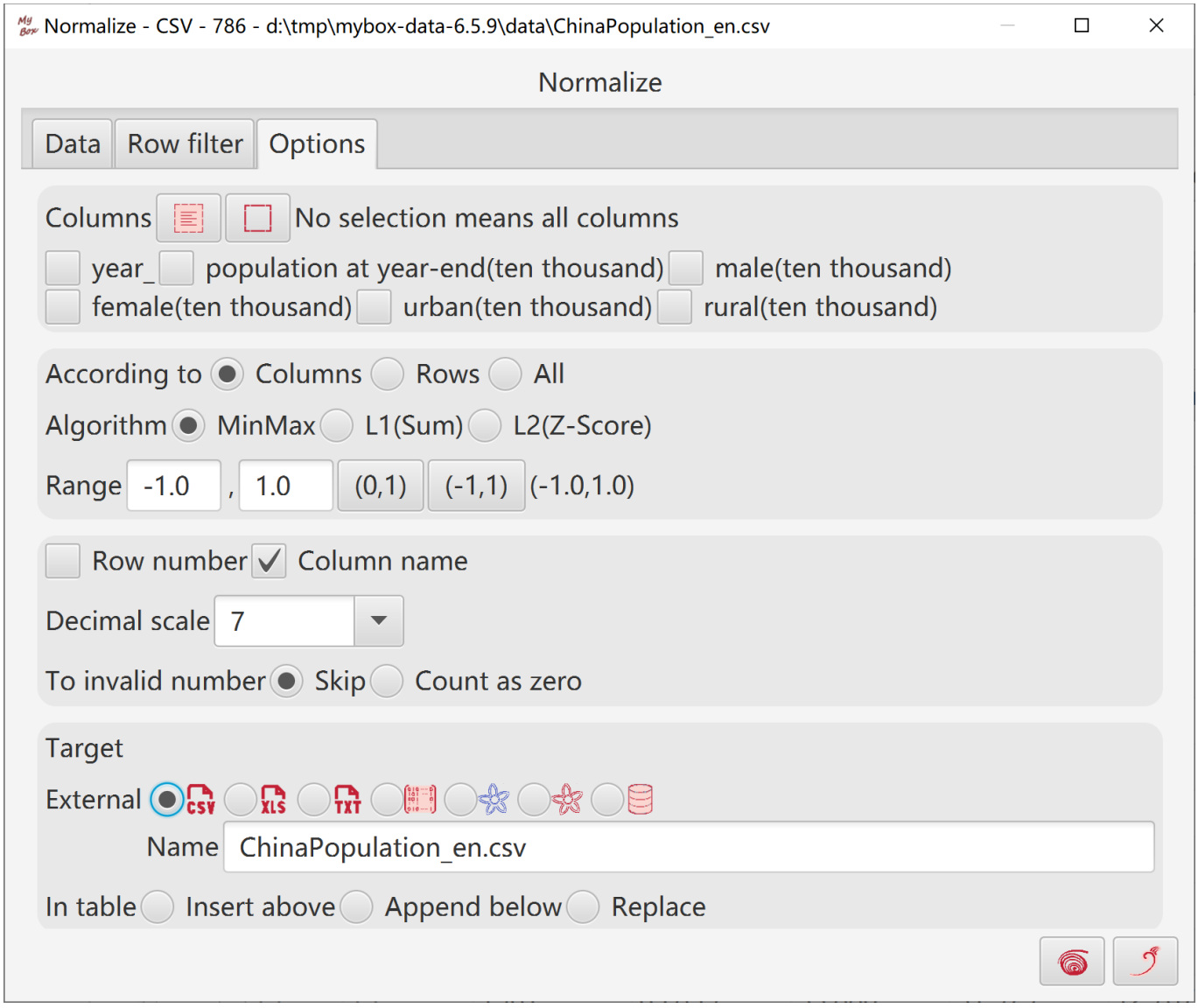
2.5.3 Transpose

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate: Options to set first column as column names.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



2.5.4 Normalization

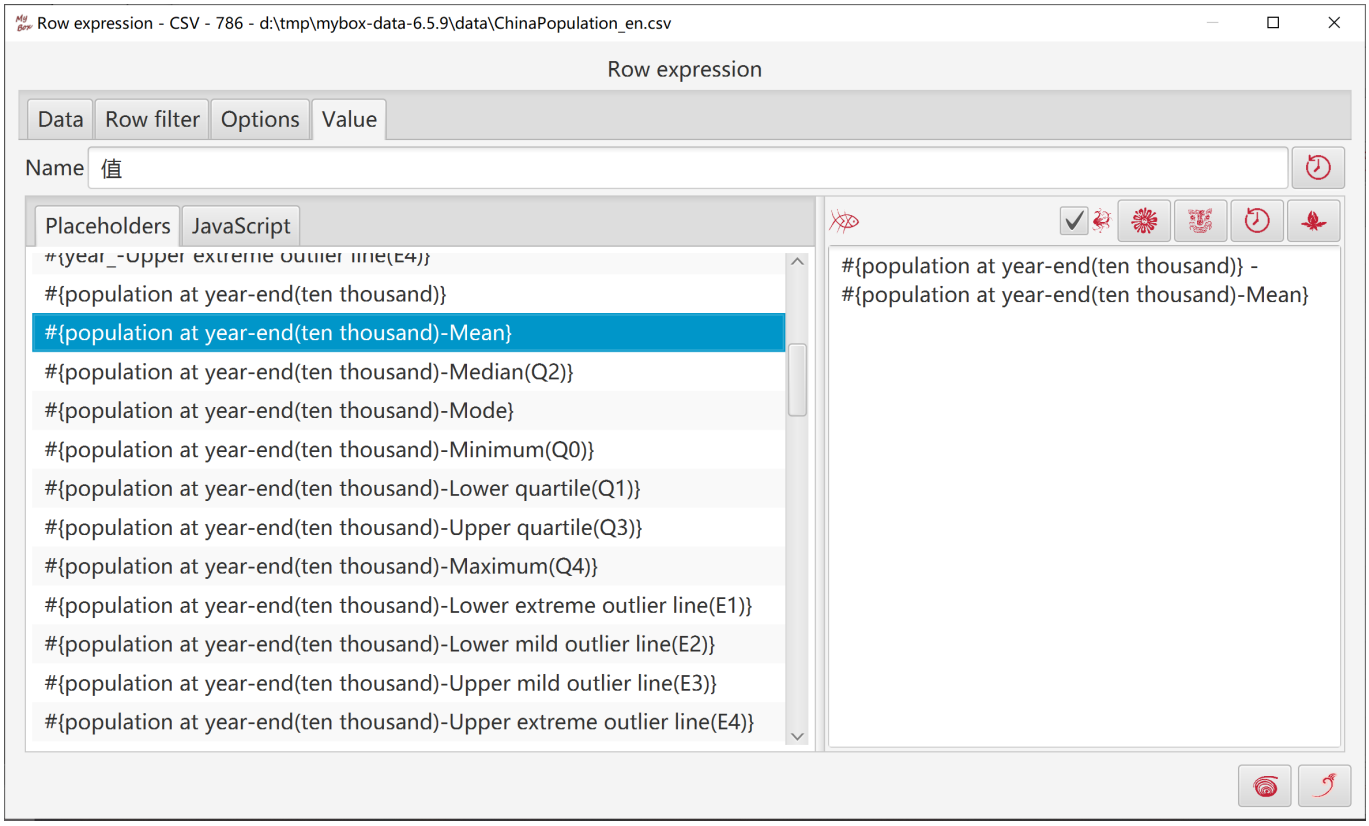
1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Accroding to: Columns/rows/all.
   - Algorithms: MinMax(Range can be set), sum(L1), ZScore(L2).
   - To non-numeric, skip or count as zero.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



2.6 Calculation

2.6.1 Row Expression

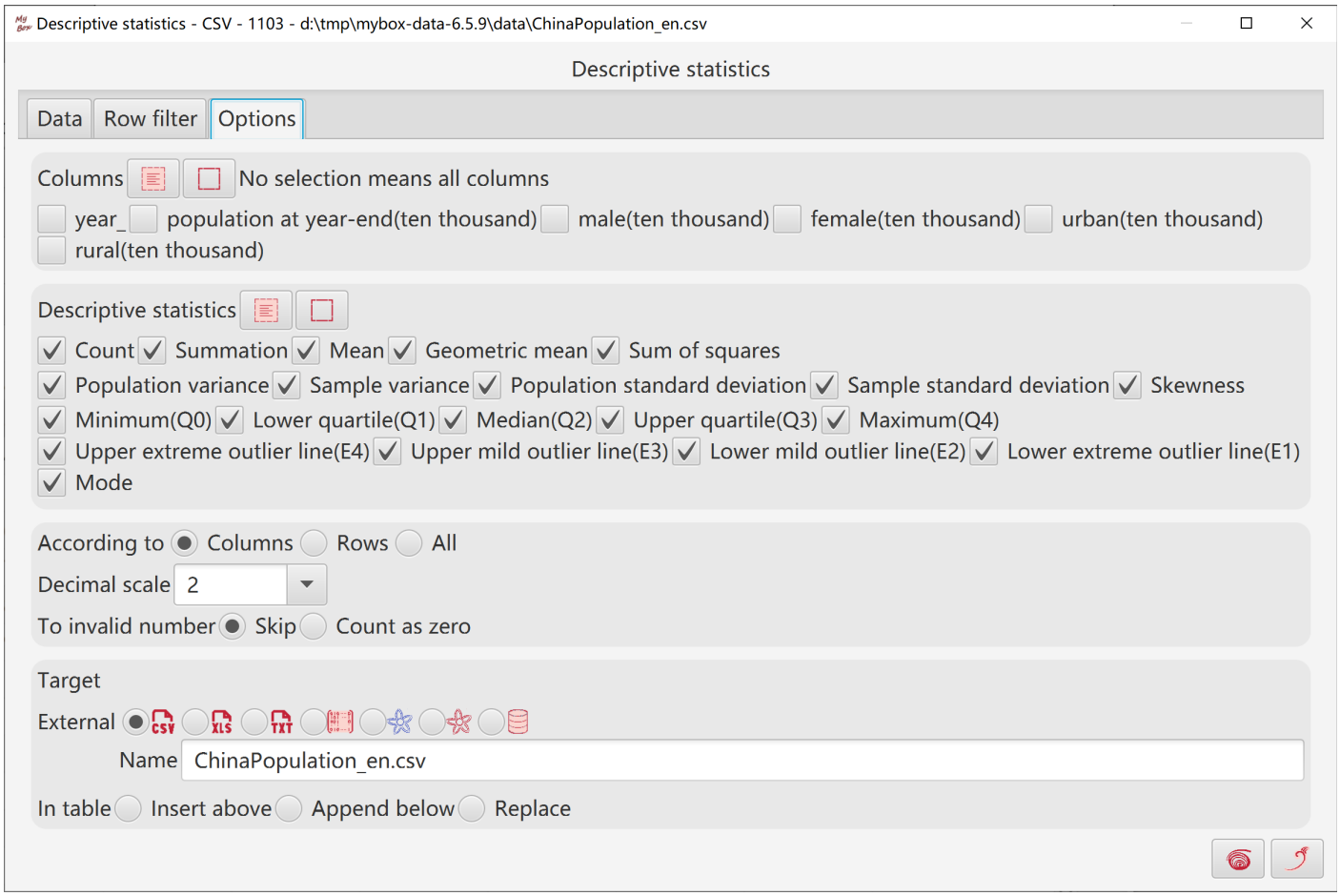
1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Input name of values.
   - Input row expression.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



2.6.2 Descriptive Statistic

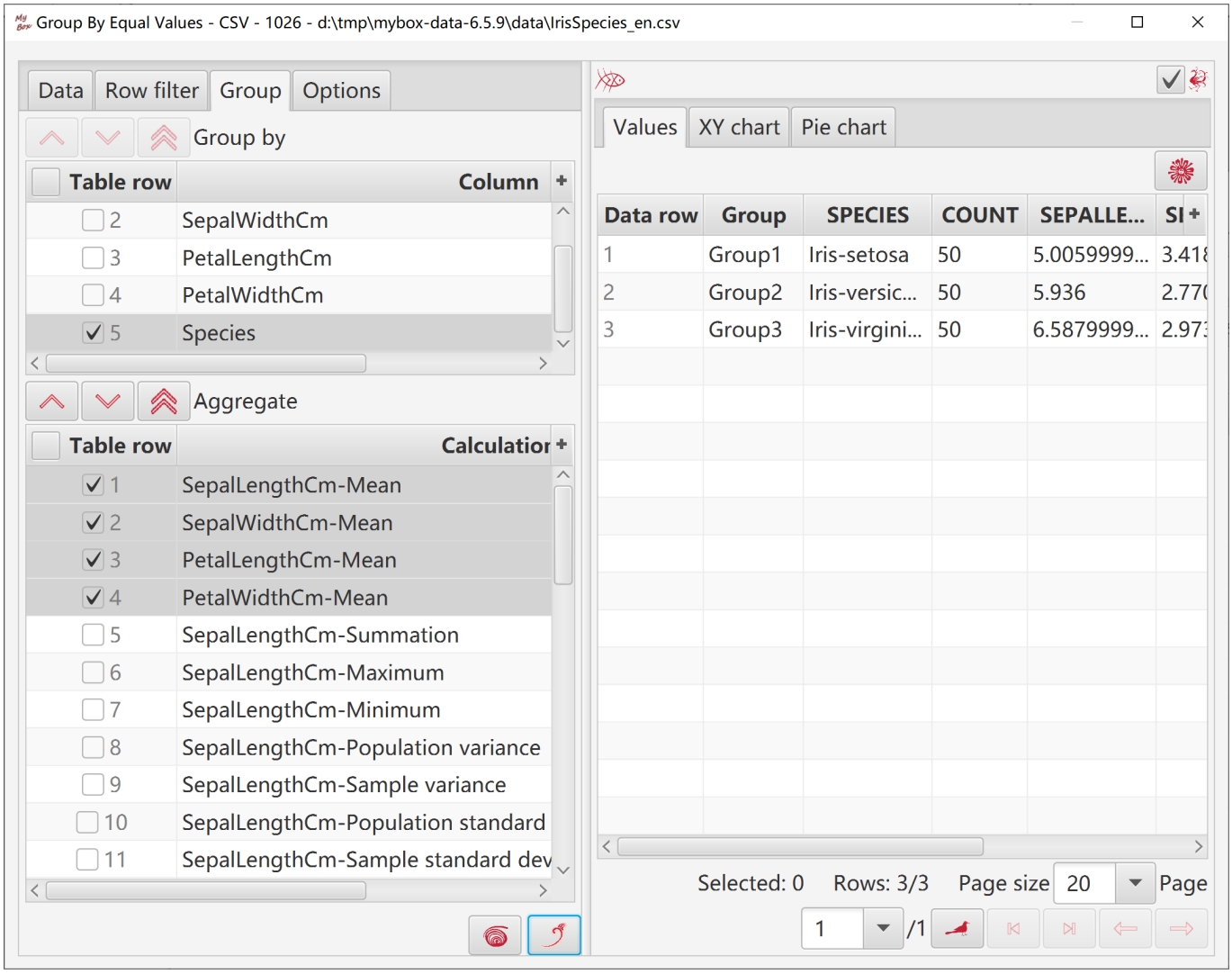
1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Following values can be generated:

     count, sum, mean, geometric mean, sum of squares, population variance, sample variance, population standard deviation, sample standard deviation, skewness, minimum(Q0), lower quartile(Q1), median(Q2), upper quartile(Q3), maximum(Q4), upper extreme outlier line(E4), upper mild outlier line(E3), lower mild outlier line(E2), lower extreme outlier line(E1), mode
   - Accroding to: Columns/rows/all.
   - Set decimal scale.
   - To non-numeric, skip or count as zero.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table, or defined location in table to insert/append/replace.

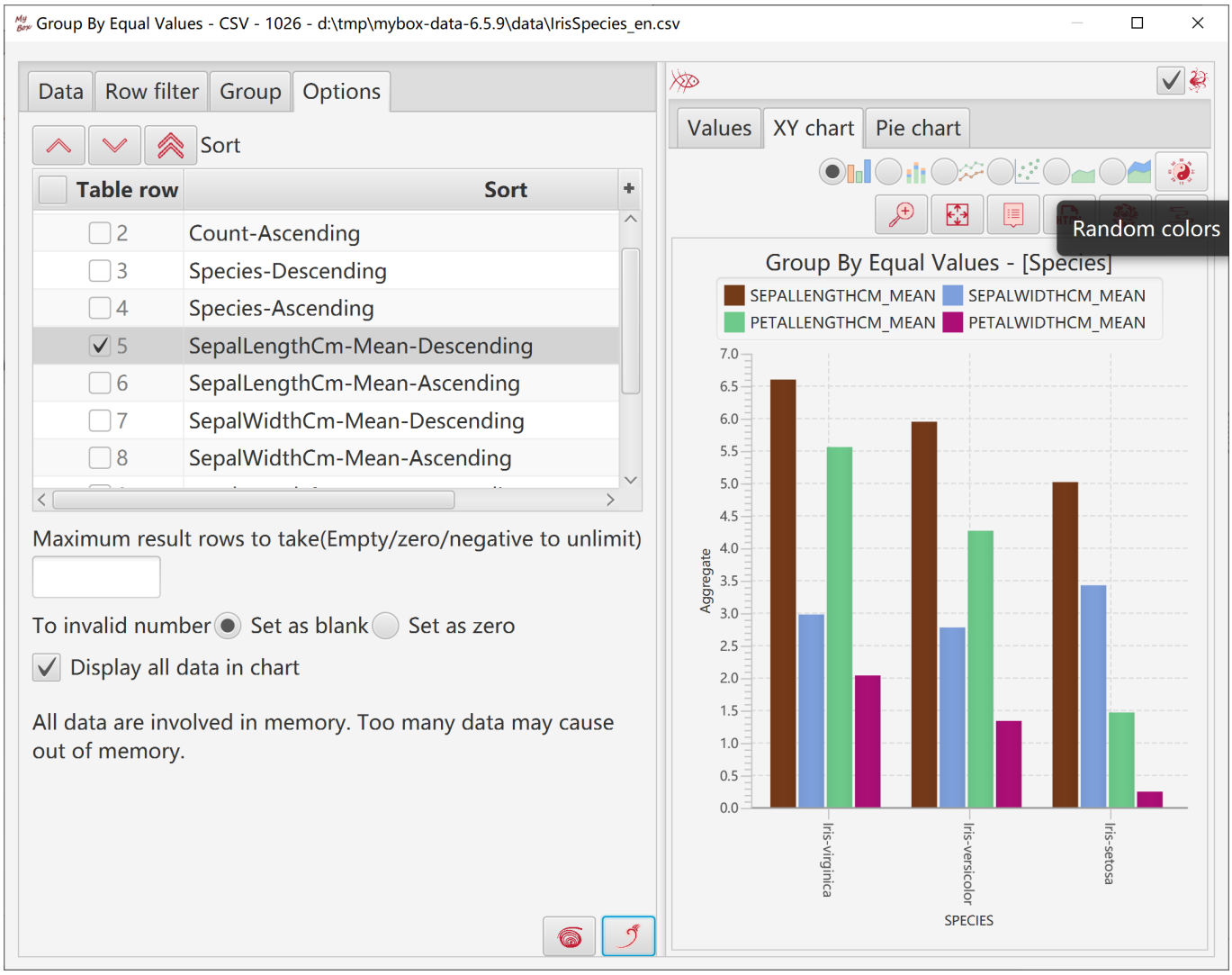


2.6.3 Group b Equal Values

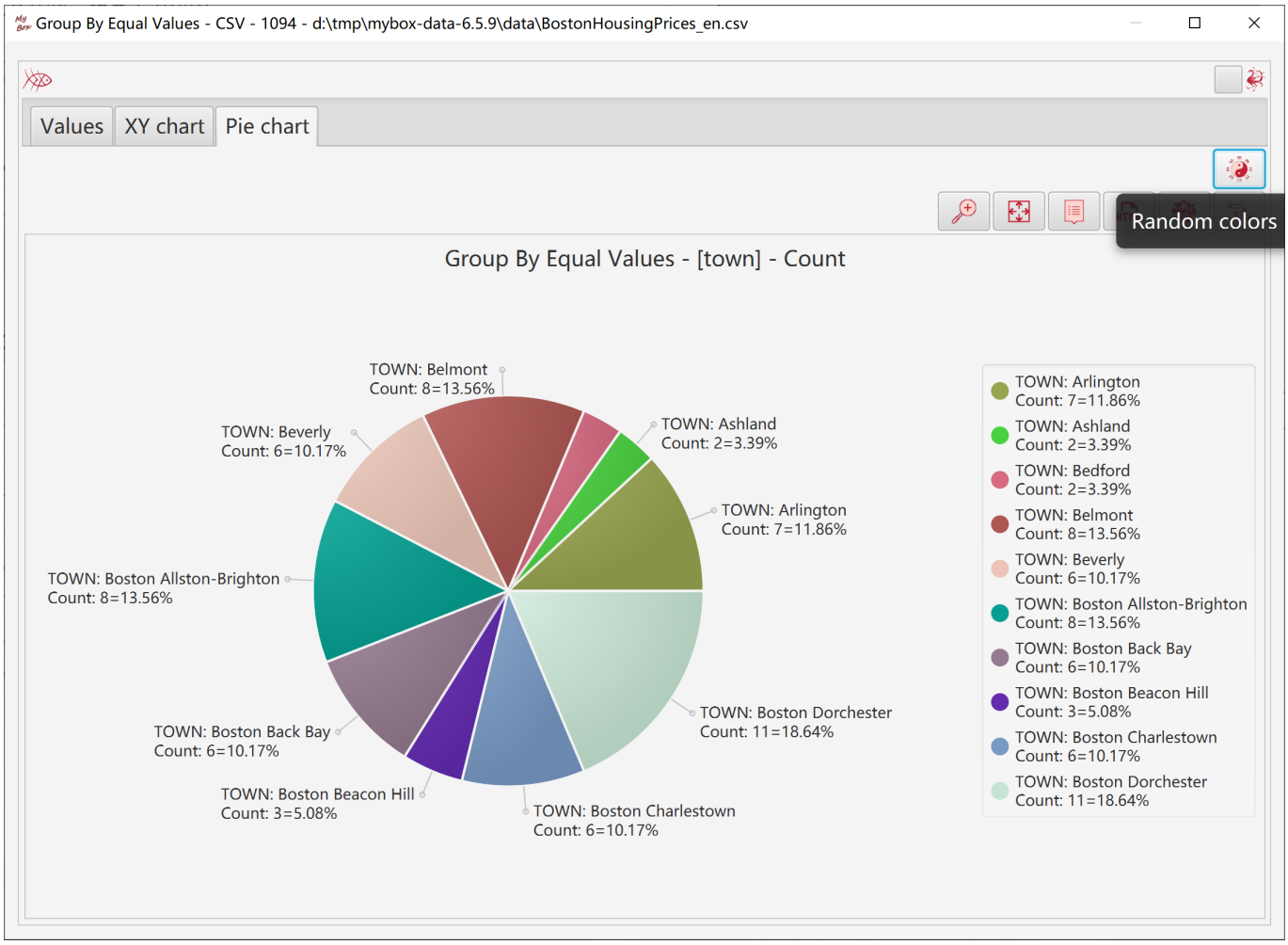
1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Set row filter.
2. Calcuate:
    - Select some columns as "group by". Need not be numbers.
    - Select some aggregate values to be calculated:
        - Like avarage, max, min, variance, standard deviation, etc.
        - "count" is always calcualted, and not in the list.
        - Multiple items can be selected. None selection is permitted.
        - Involved values should be numbers.
    - Select some values as "order by". Multiple items can be selected. None selection is permitted
    - Input maximum number of result rows. Blank/zero/negative means no limitation.
    - To invalid numbers, option to set as blank or zero.



3. Calculated groups and their aggregate values are displayed in XY chart.



4. Calculated groups and their count are displayed in Pie chart.



5. In charts:
   - If only one column in group, category values are values of this column.
   - If multiple columns in group, category values are names of groups.
6. Option to display all data in charts:
   - If yes, lots of data may cause out of memory.
   - If no, only current page of results are display in charts.
7. Data overflow may happen.

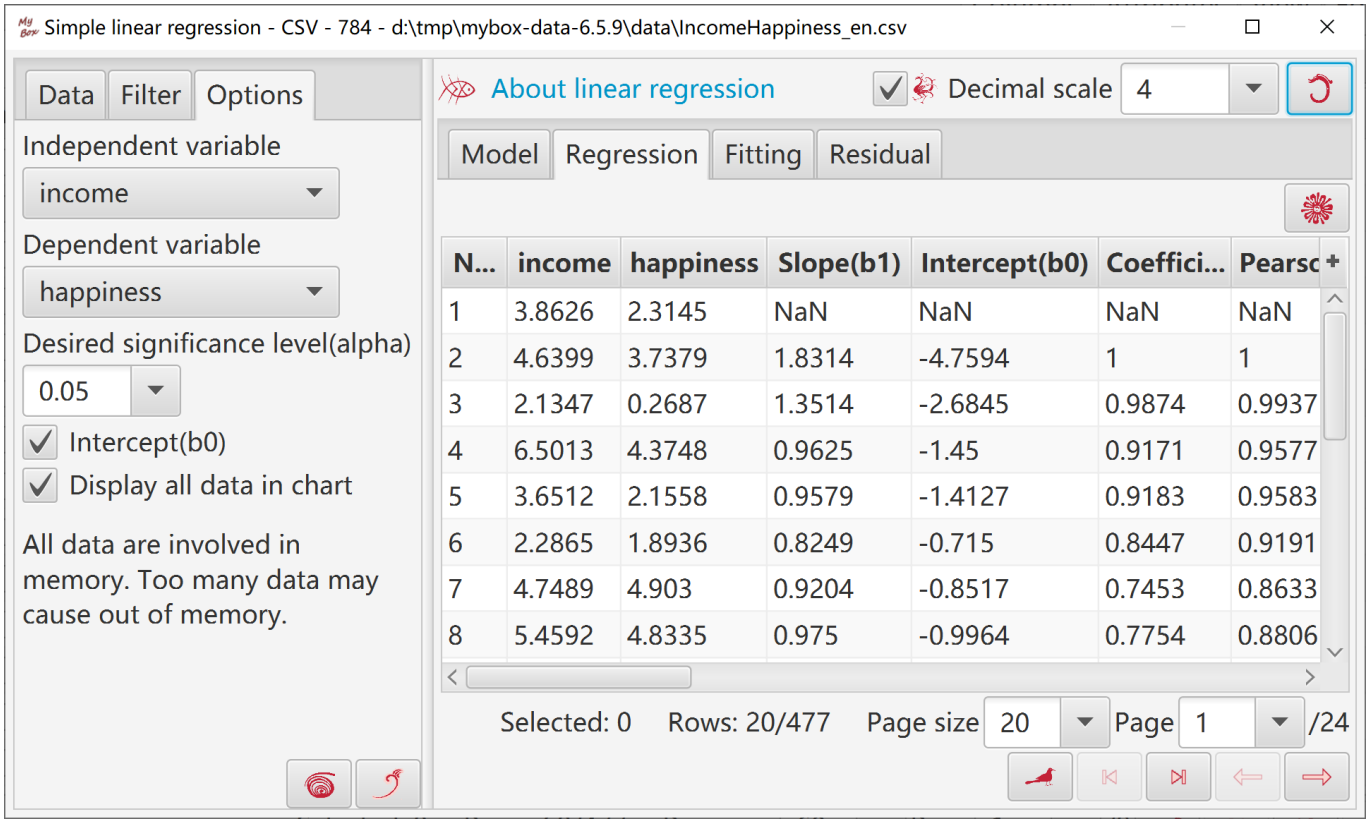
2.6.4 Simple Linear Regression

2.6.4.1 Regression

This tool is based on Apache Commons Math.

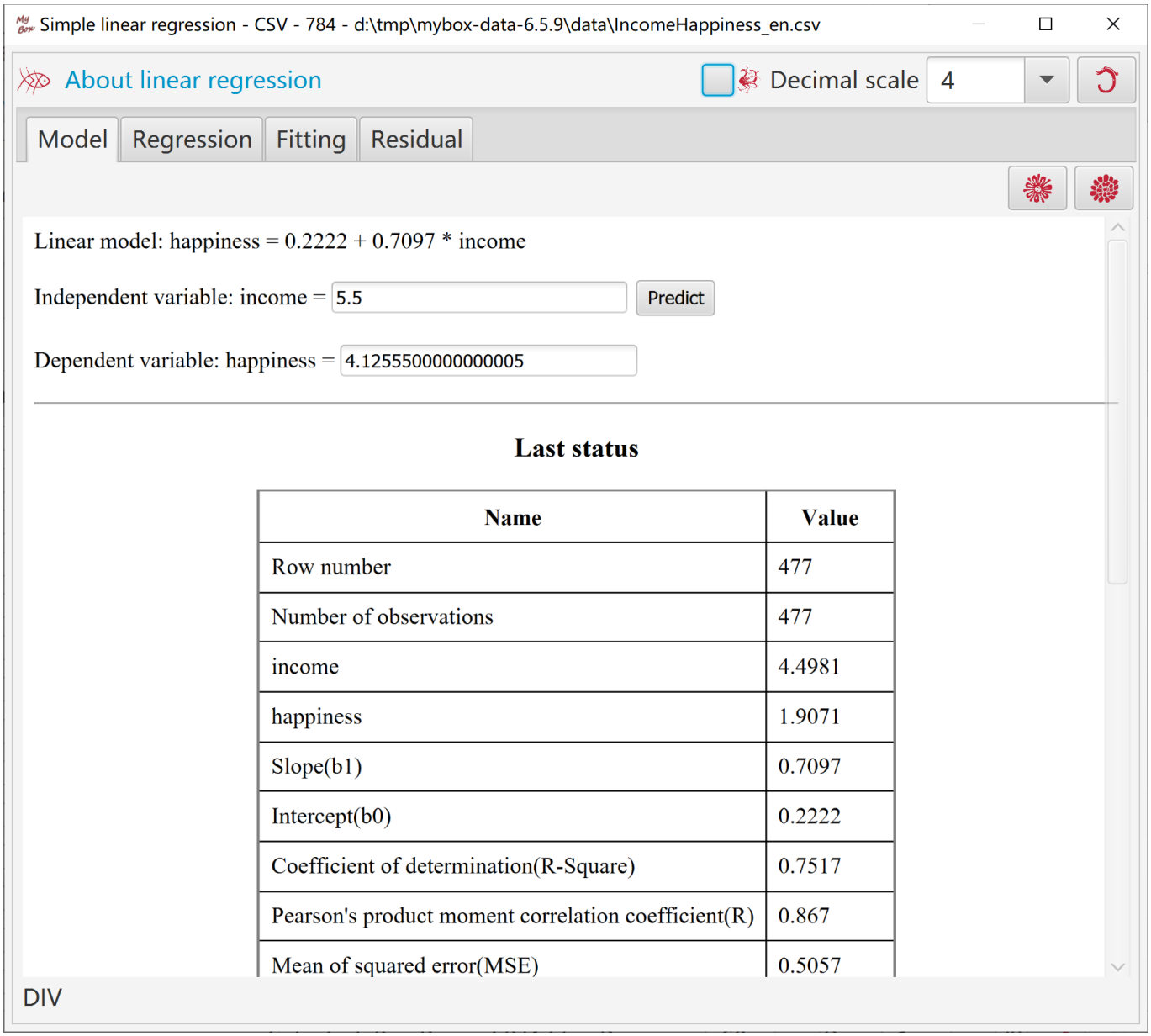
The regression does not store data, so calculation itself has not memory limitation when handle lots of rows.

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Select one column as independent variable.
   - Select another column as dependent variable.
   - Option whether includes intercept.
   - Set decimal scale.
3. Display values status of regression steps in table, including number of observations, slope, intercept, coefficient of determination(R-Square), correlation coefficient(R), mean of squared error(MSE) , sum of squared errors(SSE), total sum of squares(SSTO), sum of squares about regression(SSM/SSR), etc.



2.6.4.2 Model

1. Display fitted liear model.
2. Display data status of last regression step.
3. Input value fo independent variable, and generate predicted value.

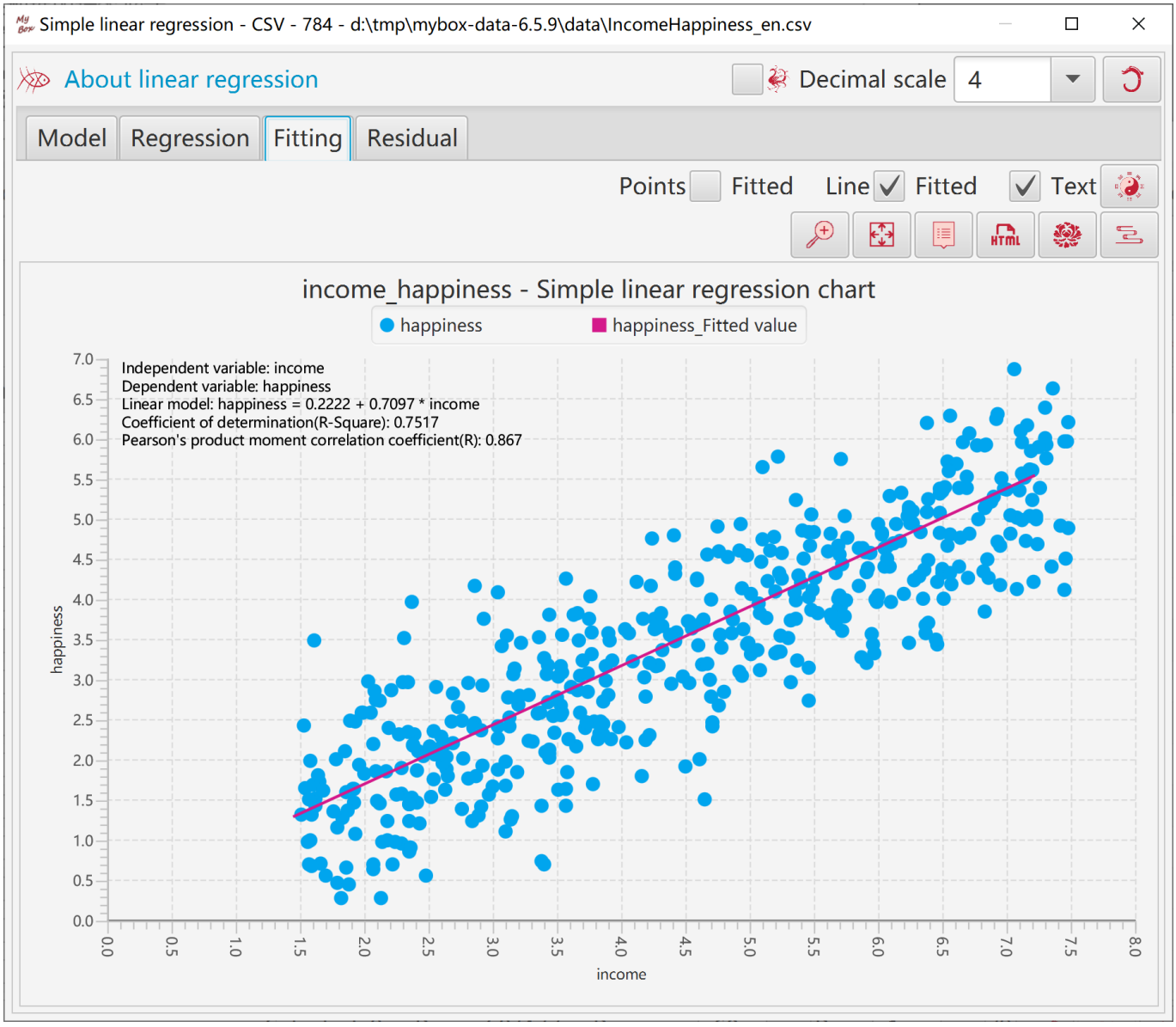


2.6.4.3 Fitted Chart

1. When handle all data rows(all pages), option to display all values in chart.

   When display all values in chart, need concern memory limitation when load lots of data.

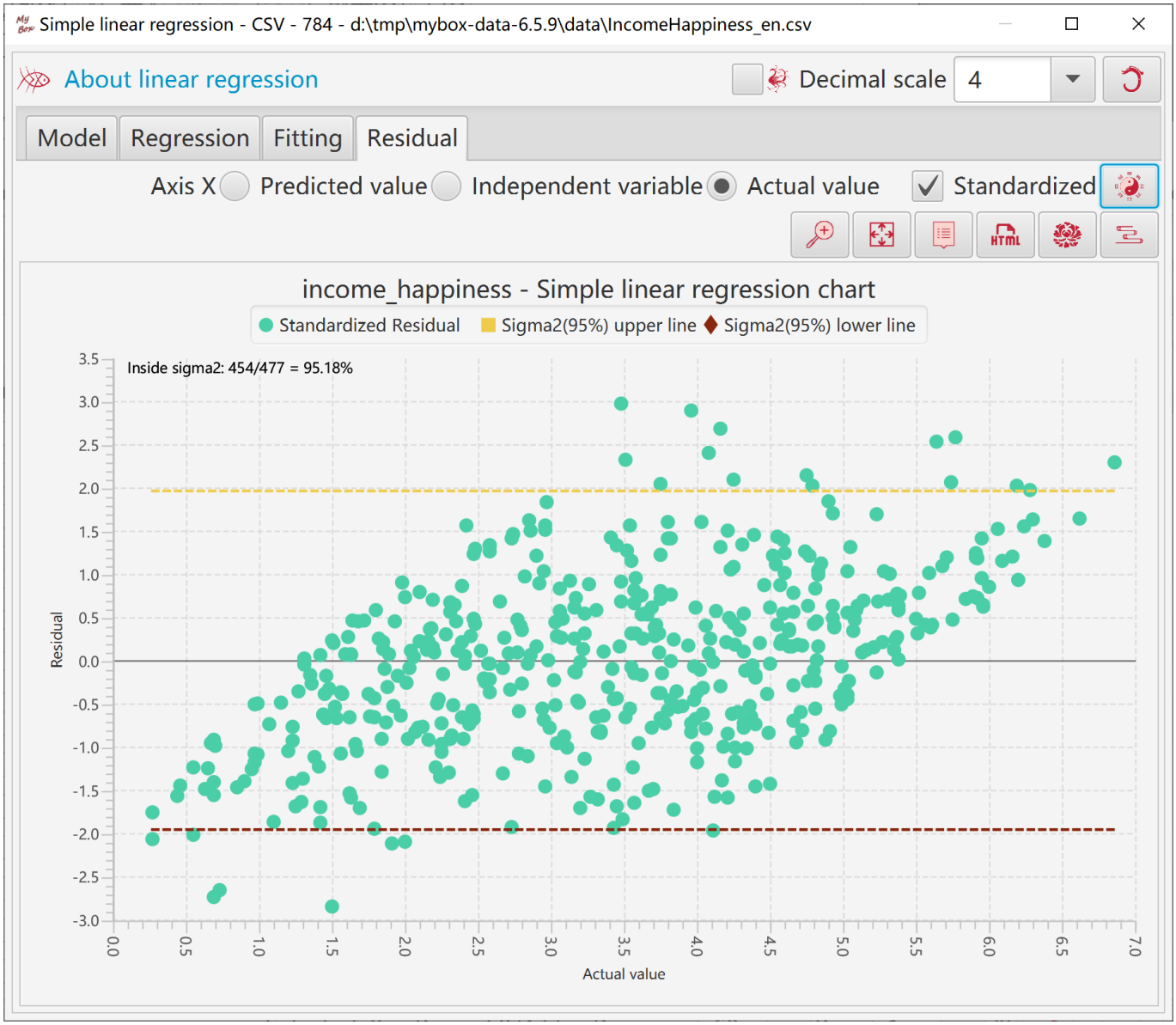
   Or else only values in current page will be displayed in chart while all pages involve regression, and no memory limitation.
2. Set parameters of plot, X axis and Y axis.
3. Options to display fitted points, fitted line, or model description in chart.
4. Options to display data labels.
5. Set random colors to fitted points/line.
6. Fitted chart can be popped.
7. Html including fitted chart and its data can be created.
8. Display fitted chart's data in table.



2.6.4.4 Residual Chart

1. X axis can be set as: predicted value, independent varaible, or actual value.
2. Option whether standardize residuals.

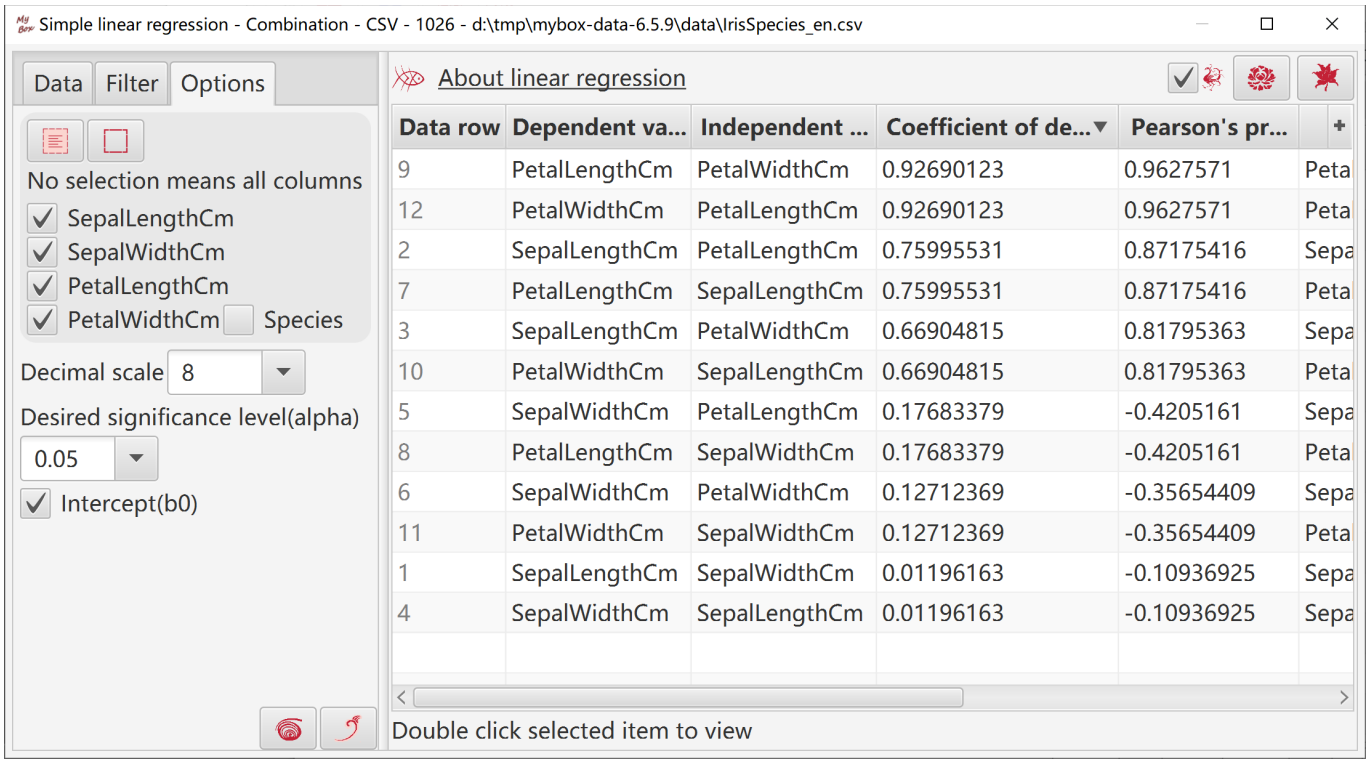
   When standardize residuals, upper line and lower line of Sigma2(95%) will be displayed.
3. Set random color to points and lines.
4. Display residual chart's data in table.



2.6.5 Simple Linear Regression - Combination

This tool helps to generate data of simple linear regression:

1. Select some columns.
2. Select options like decimal scale, alpha, whether include intercept.
3. When click button "OK":
    - Tool make pairs of columns and calculate their regression models.
    - Sort the modes by their coefficient of determination(R-Squre) in descending order.
4. Select one mode and click button "View" to view its regression data, fitting chart, and residual chart.

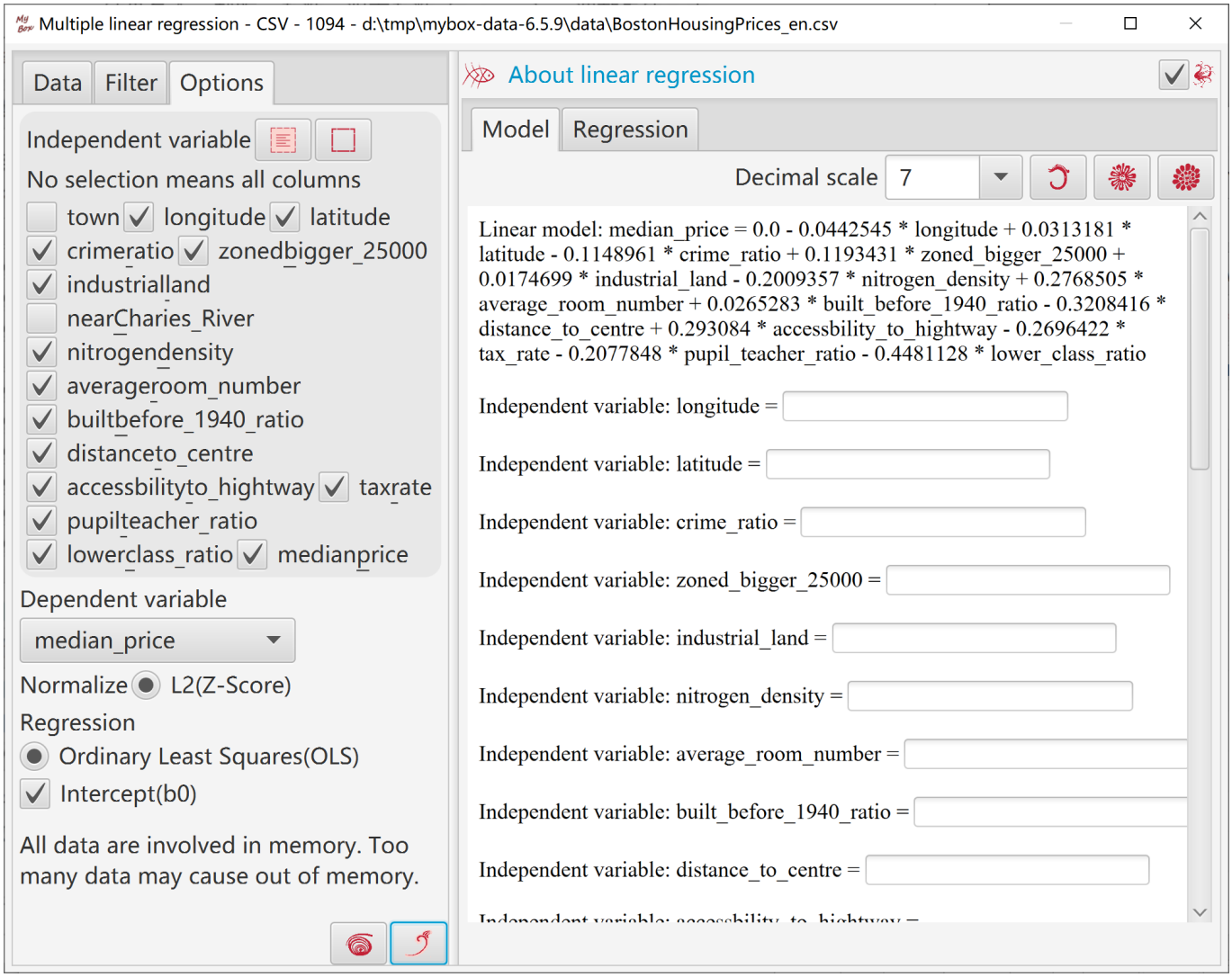


2.6.6 Multiple Linear Regression

2.6.6.1 Regression

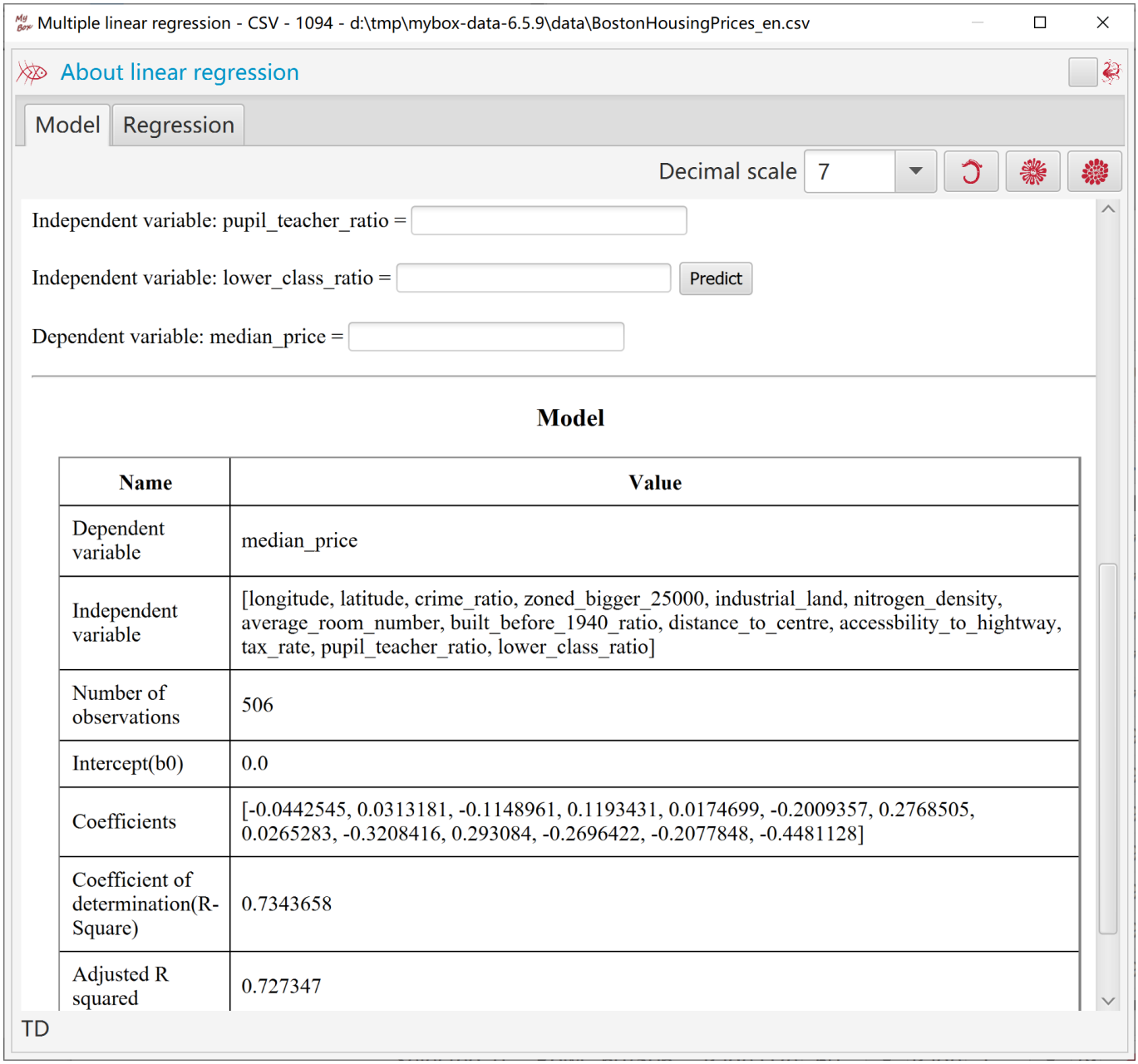
This tool helps to generate data of multiple linear regression based on Apache Commons Math:

1. Select some column as independent variables, whose data should be numbers.
2. Select a column as dependent variable, whose data should be numbers.
3. Select whether include intercept.
4. Click button "OK":
    - Tool normalizes involved data by Z-Score.
    - Tool calculates multiple linear regression by Ordinary Least Squares(OLS).
    - Results include intercept, coefficients, R-Squre, adjusted R-Squre.



2.6.6.2 Model

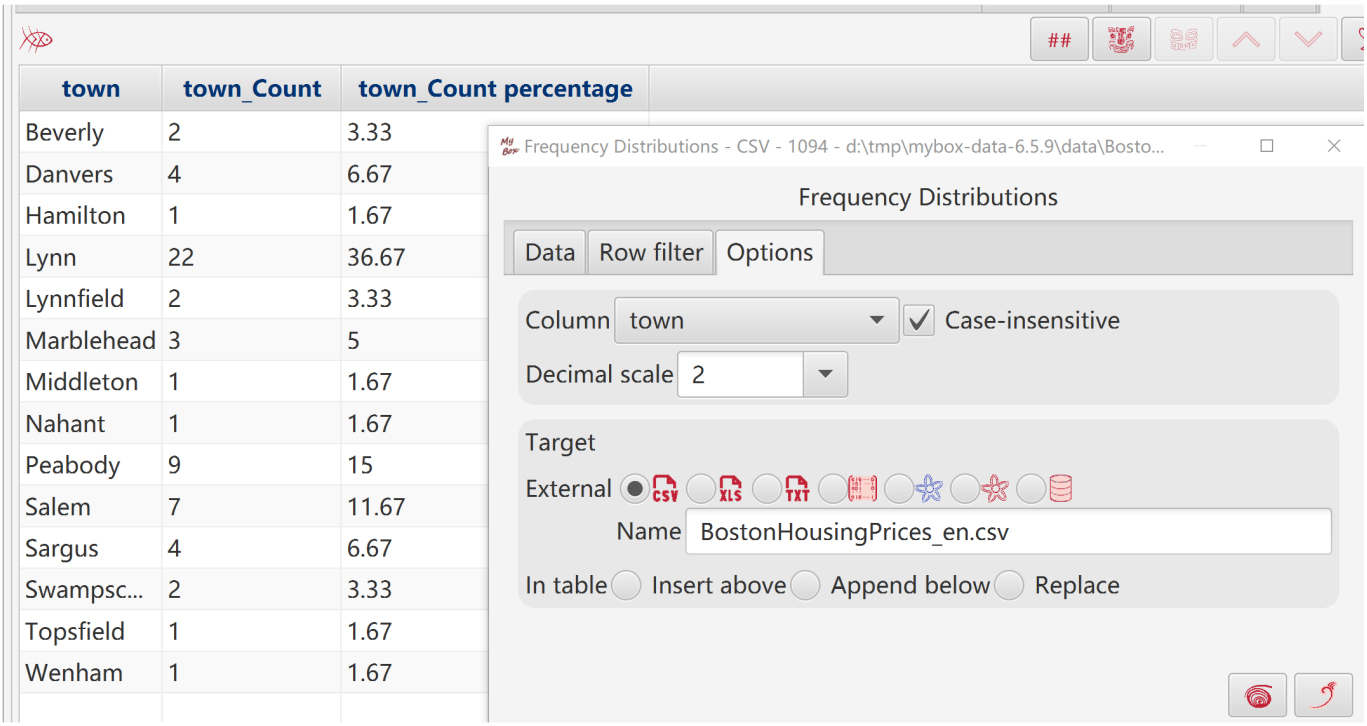
User can input values of independent variables and predict value of dependent variable.



Name	Value
Dependent variable	median_price
Independent variable	[longitude, latitude, crime_ratio, zoned_bigger_25000, industrial_land, nitrogen_density, average_room_number, built_before_1940_ratio, distance_to_centre, accessbility_to_hightway, tax_rate, pupil_teacher_ratio, lower_class_ratio]
Number of observations	506
Intercept(b0)	0.0
Coefficients	[-0.0442545, 0.0313181, -0.1148961, 0.1193431, 0.0174699, -0.2009357, 0.2768505, 0.0265283, -0.3208416, 0.293084, -0.2696422, -0.2077848, -0.4481128]
Coefficient of determination(R-Square)	0.7343658
Adjusted R squared	0.727347

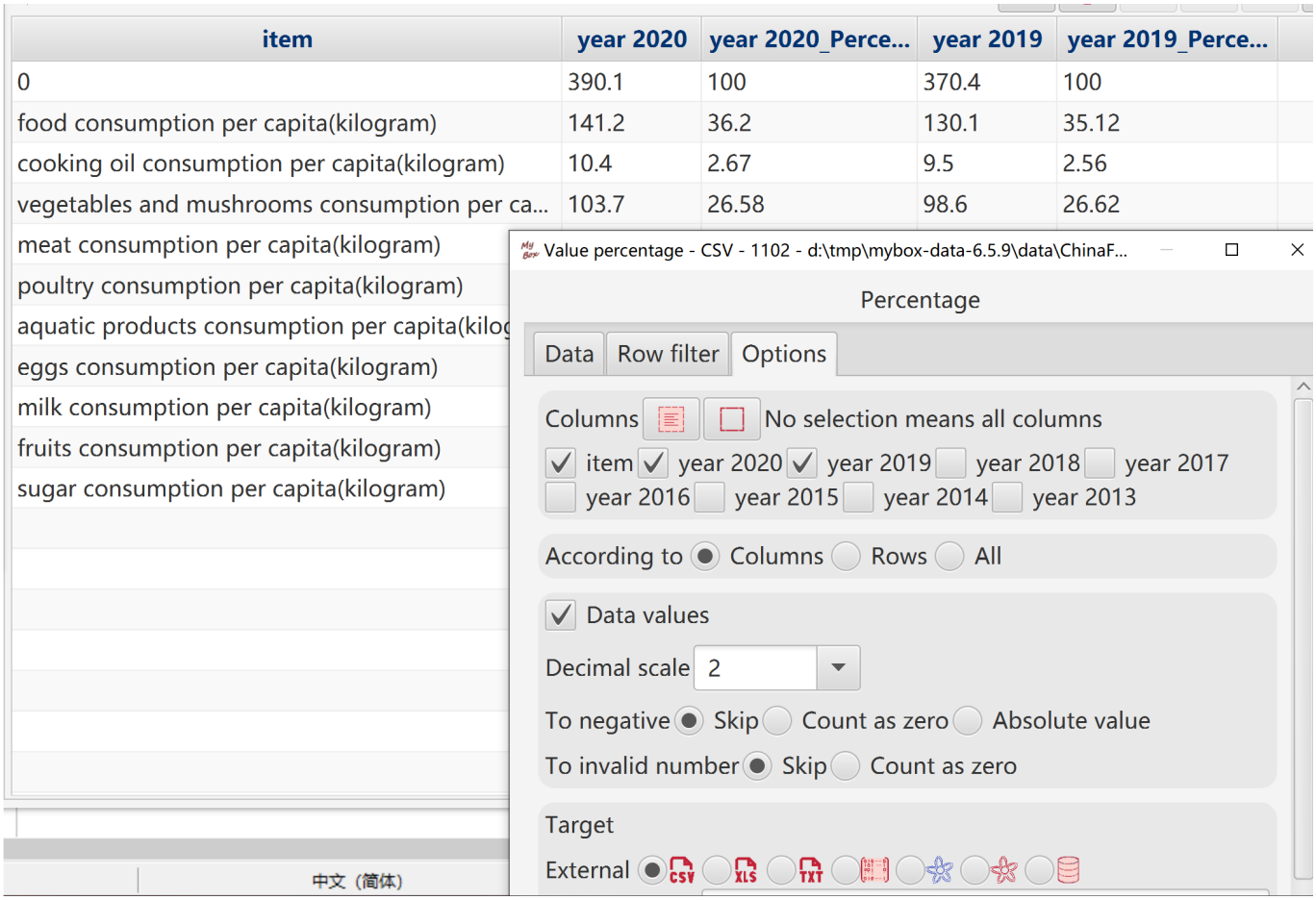
2.6.7 Frequency Distributions

1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Select columns. If no column is selected, then all columns are taken.
    - Set row filter.
2. Calcuate:
    - Select one column to count frequency.
    - Option whether case-insensitive
    - Set decimal scale.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.



2.6.8 Values Percentage

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Accroding to: Columns/rows/all.
   - Select how to treat negative values: zero or absoluate value.
   - Set decimal scale.
   - To non-numeric, skip or count as zero.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.

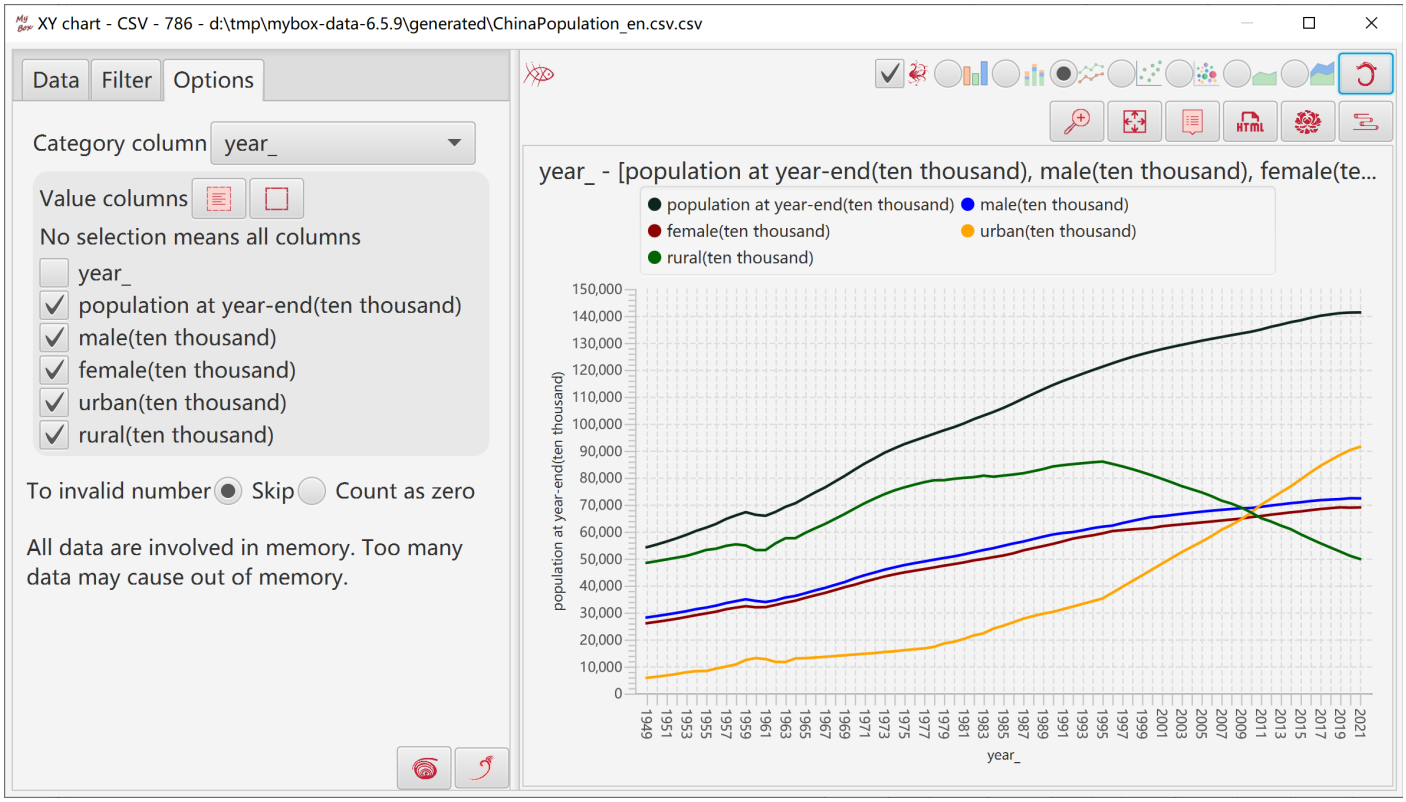


2.7 Chart

2.7.1 XY Chart

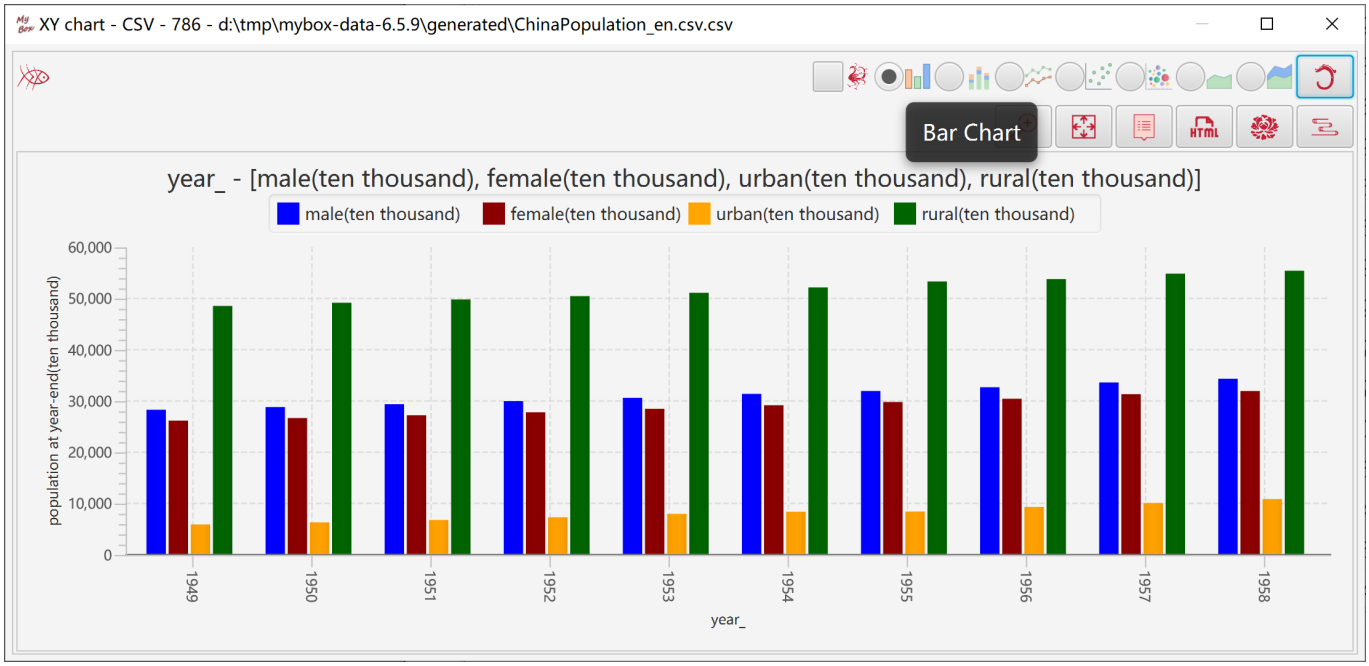
2.7.1.1 Data

1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Select columns. If no column is selected, then all columns are taken.
    - Set row filter.
2. Calcuate:
    - Select one column as "Category Axis", to define data names..
    - Multiple columns can be selected as data in direction of "Number Axis". Different value series are shown in different colors or shapes.
    - To non-numeric, skip or count as zero.
    - When display all data rows(all pages), need concern memory limitation.
    - By default, "Category Axis" is the horizontal axis and "Number Axis" is the vertical axis.
3. Select type of XY Chart.
4. Click button “Menu” to set parameters of chart.
5. Click button “Pop” to display current chart as image in popped window.
6. Click button “Data” to display data of the chart in data table.
7. Click button “Html” to display data of the chart in html page.



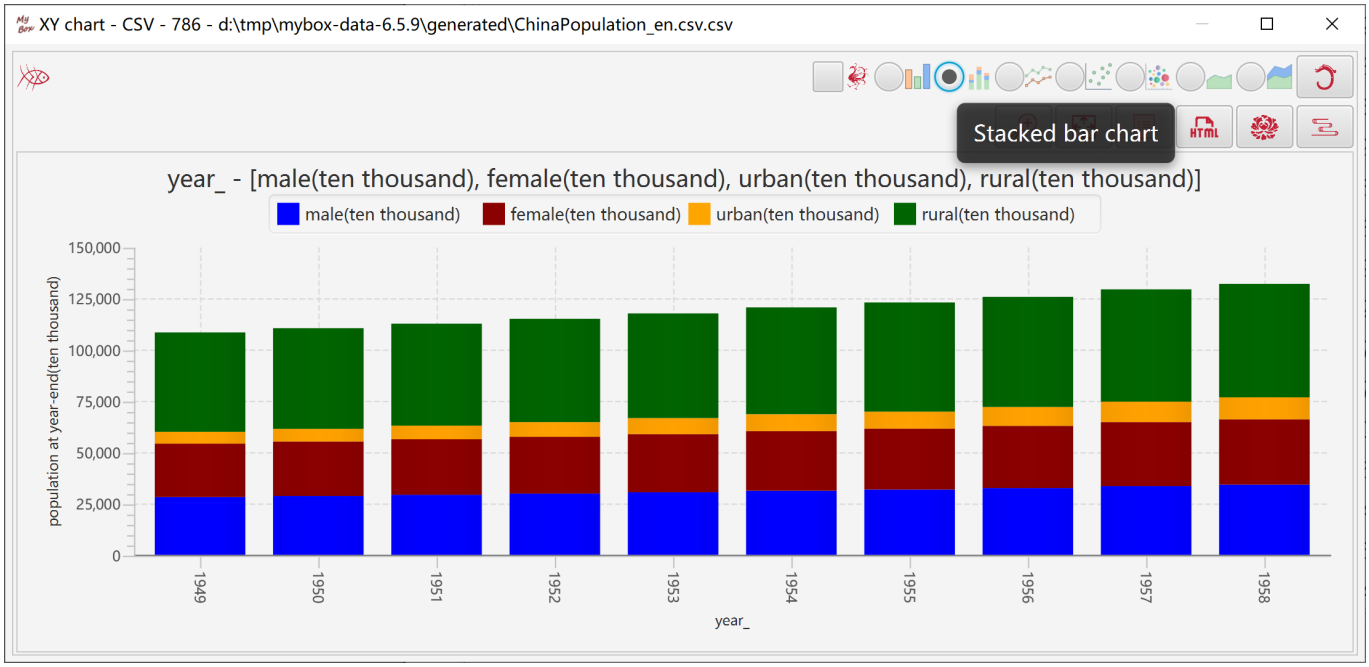
2.7.1.2 Bar Chart

1. Represents data size with bars' heights.
2. "Category" column is always counted as strings.



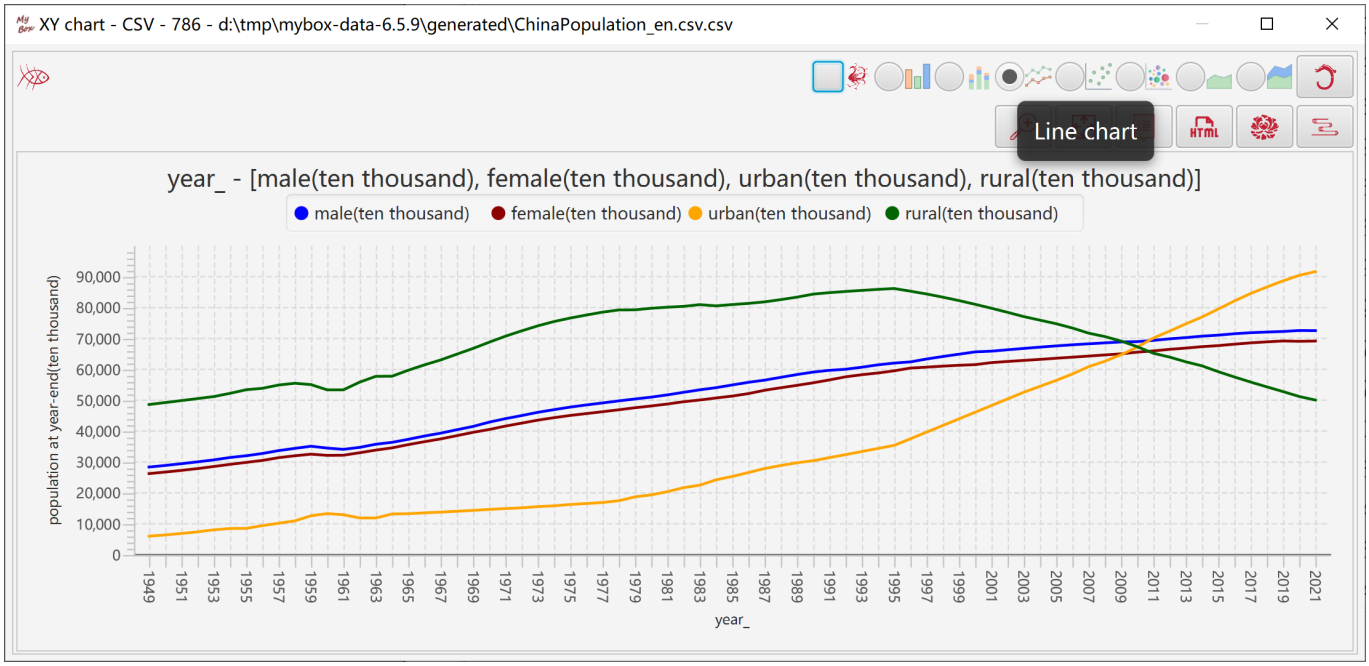
2.7.1.3 Stacked Bar Chart

1. Represents data size with bars' heights.
2. "Category" column is always counted as strings.



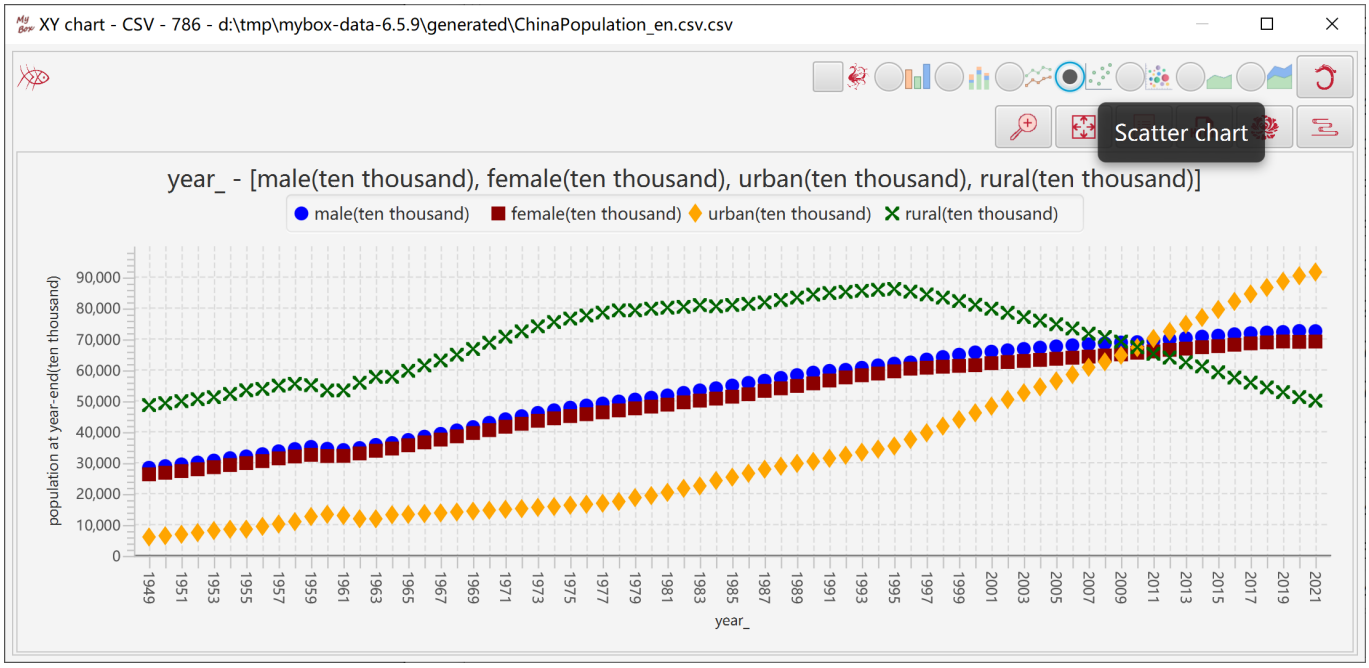
2.7.1.4 Line Chart

1. Represents data trend with lines connecting points.
2. "Category" column can be counted as strings or numbers.



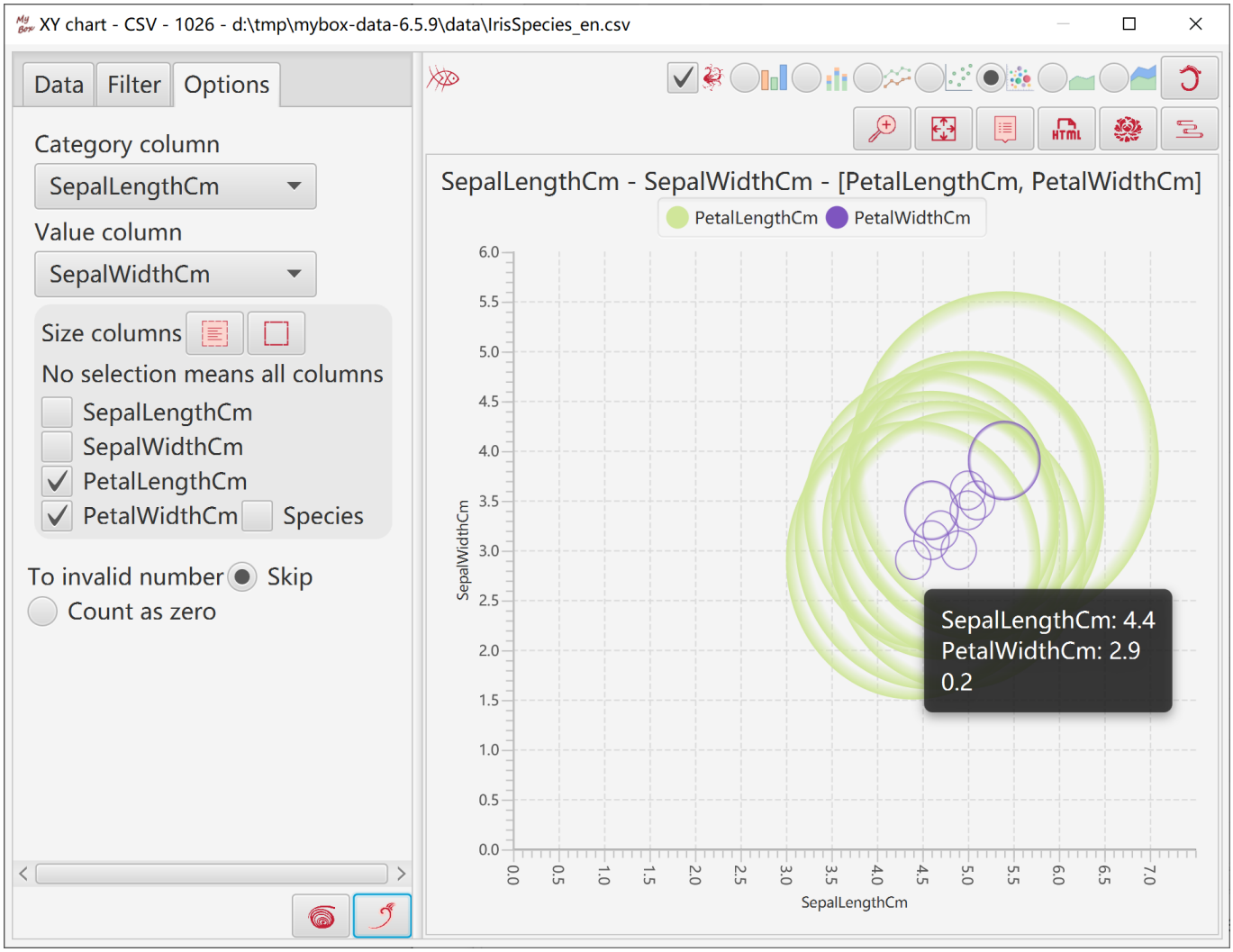
2.7.1.5 Scatter Chart

1. Represents data distribution with symbols.
2. "Category" column can be counted as strings or numbers.



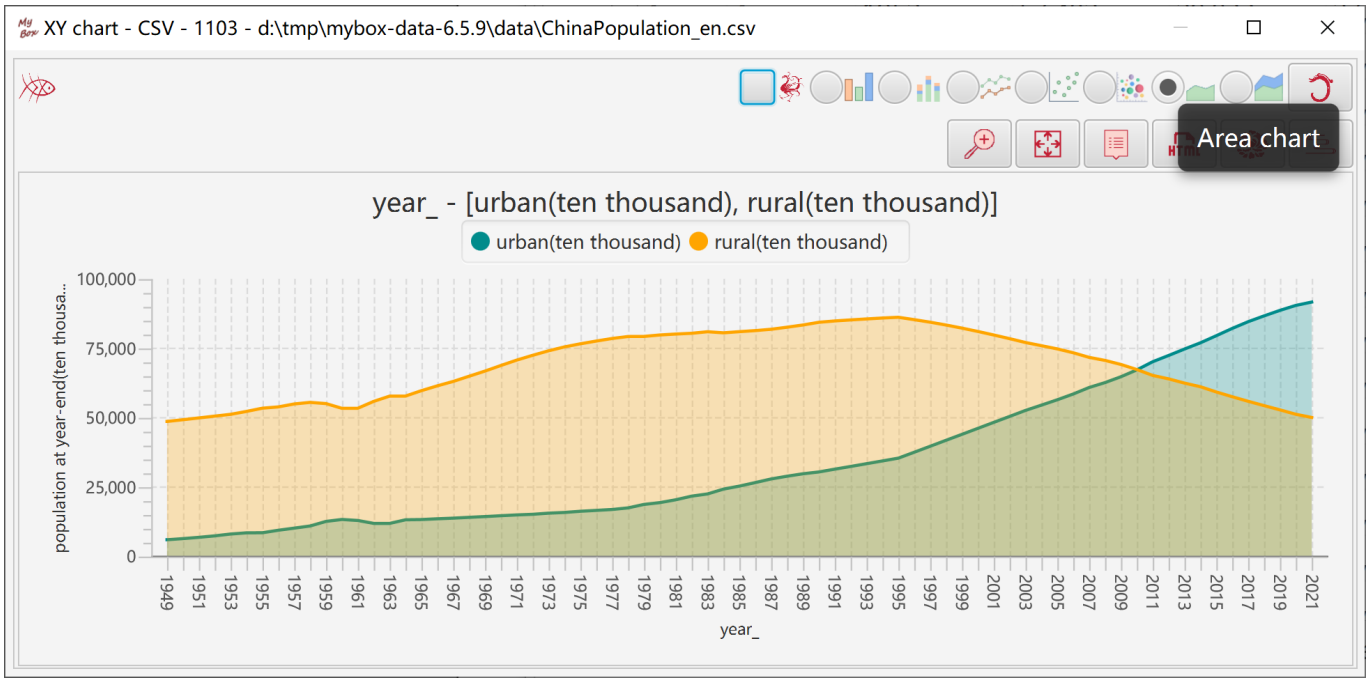
2.7.1.6 Bubble Chart

1. Represents data size with circles of different radius：
2. "Category Column" and "Value Column" define coordinates of data.
3. Select several columns as "Size Columns" to defines data size.
4. All columns should be numbers. Size columns should be non-negative.



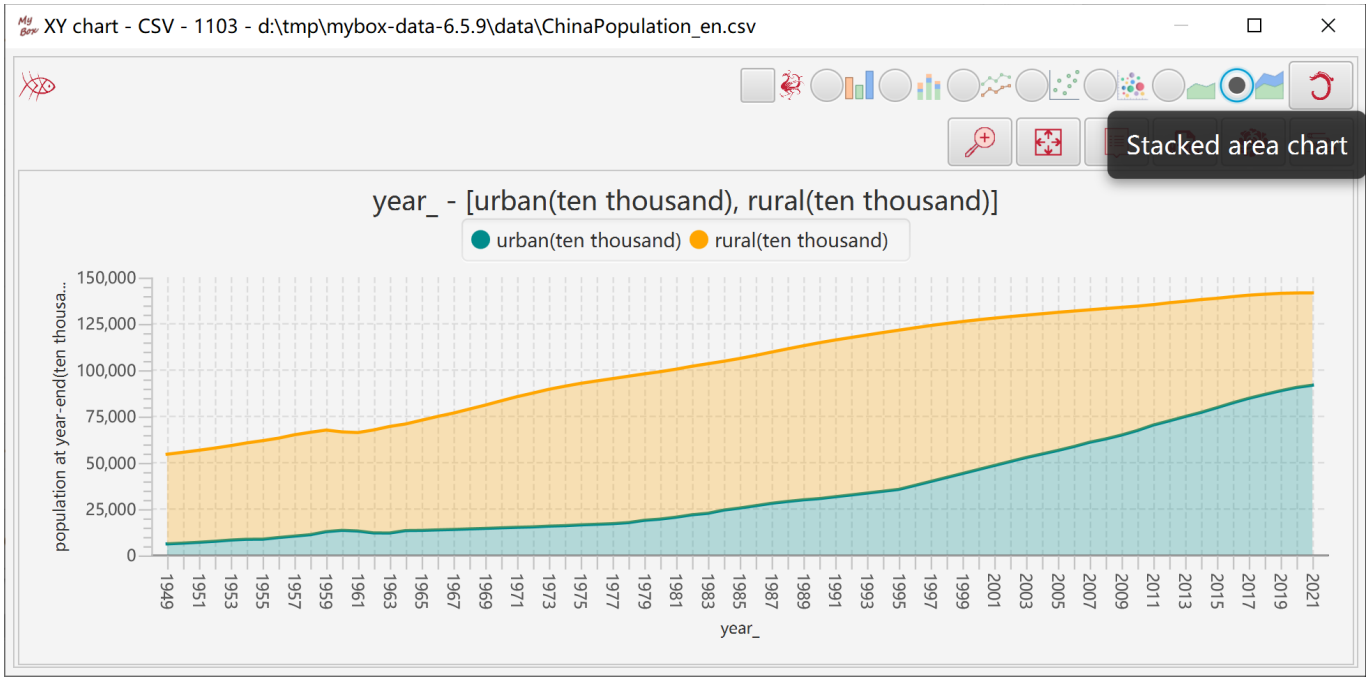
2.7.1.7 Area Chart

1. Represents data size with area size.
2. "Category" column is always counted as strings.



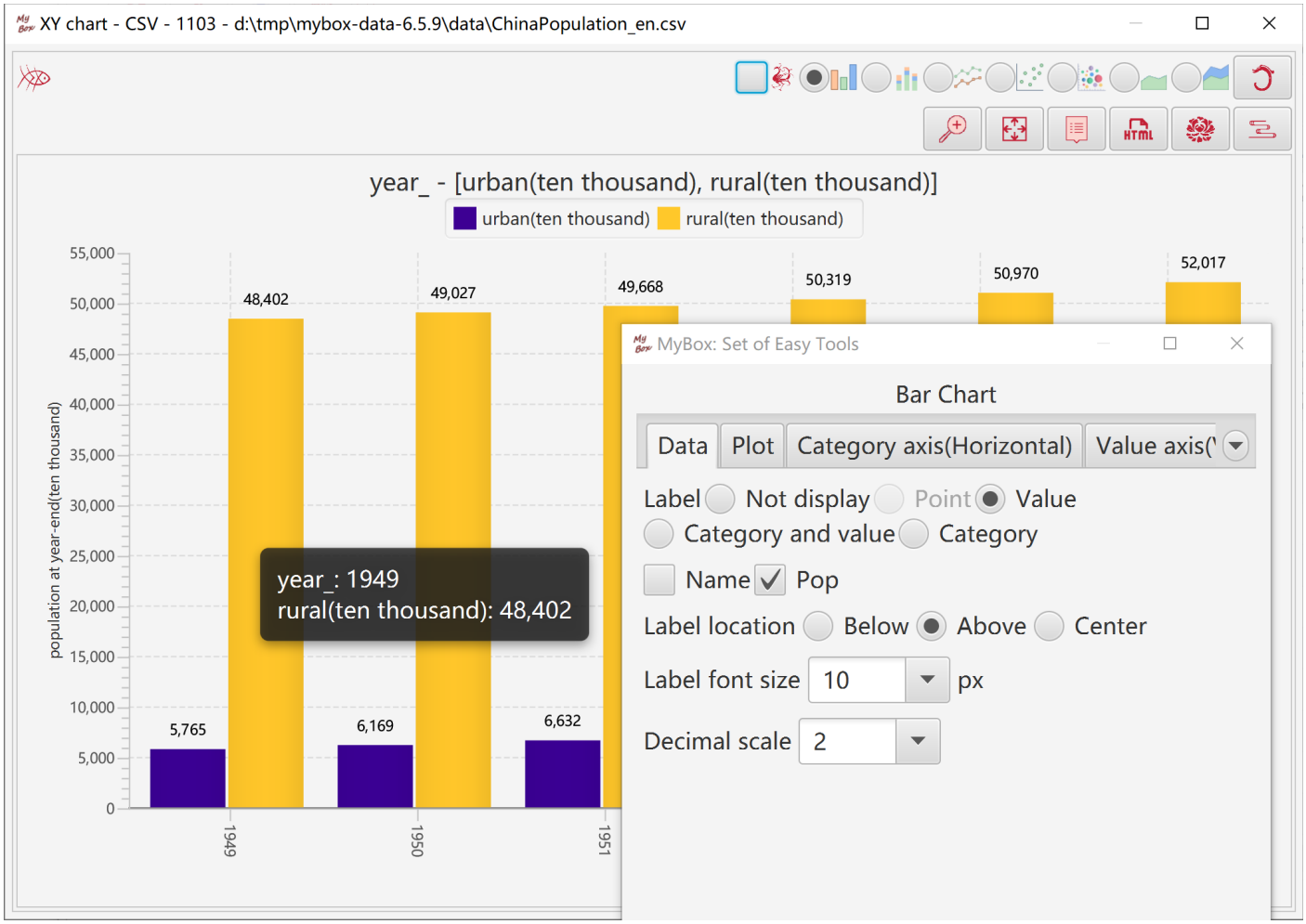
2.7.1.8 Stacked Area Chart

1. Represents data size with area size.
2. "Category" column is always counted as strings.



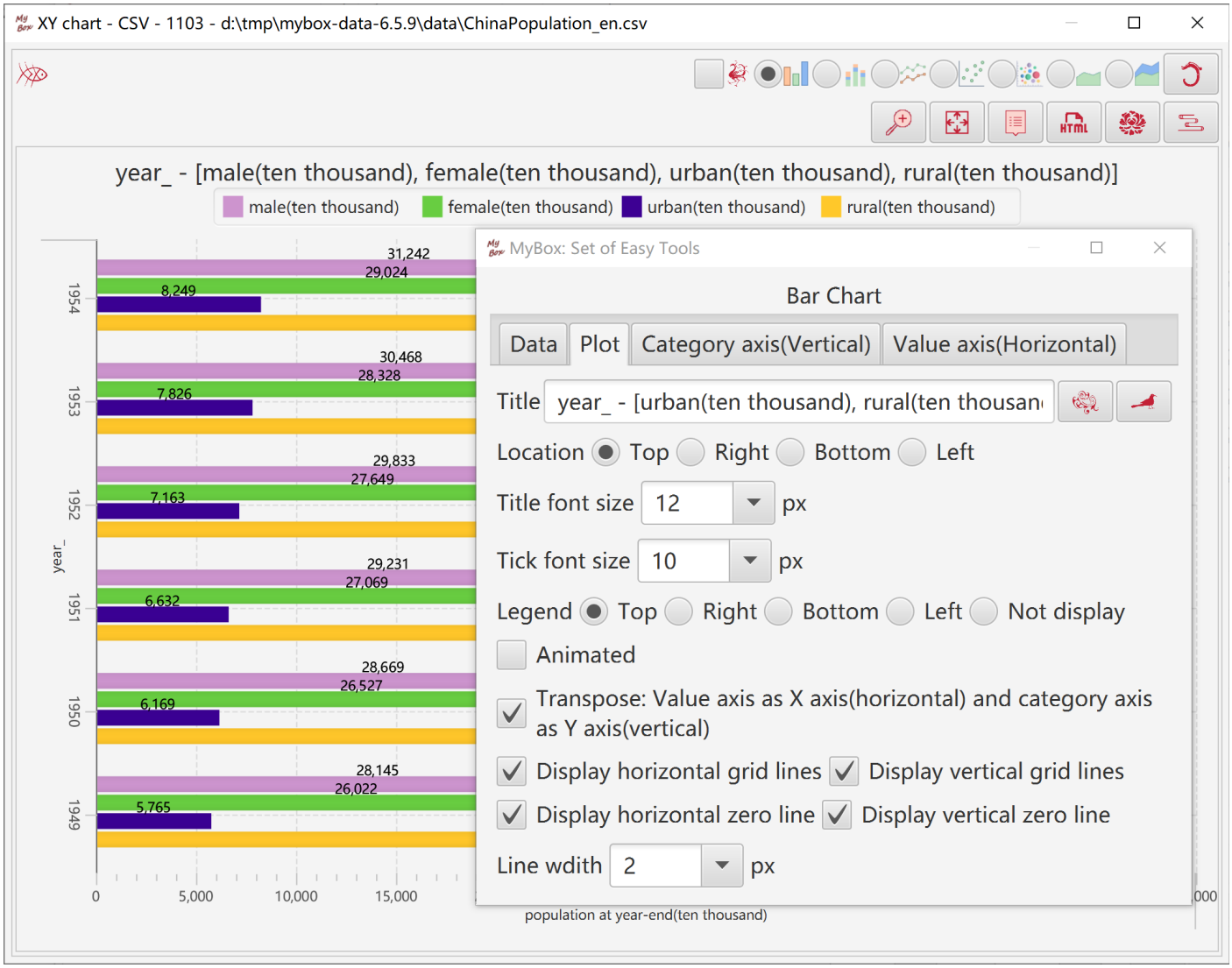
2.7.1.9 Parameters of Data in Chart

1. Options about label: not display, point, value, category, etc.
2. Location of labels.
3. Font size of labels.
4. Decimal scale.



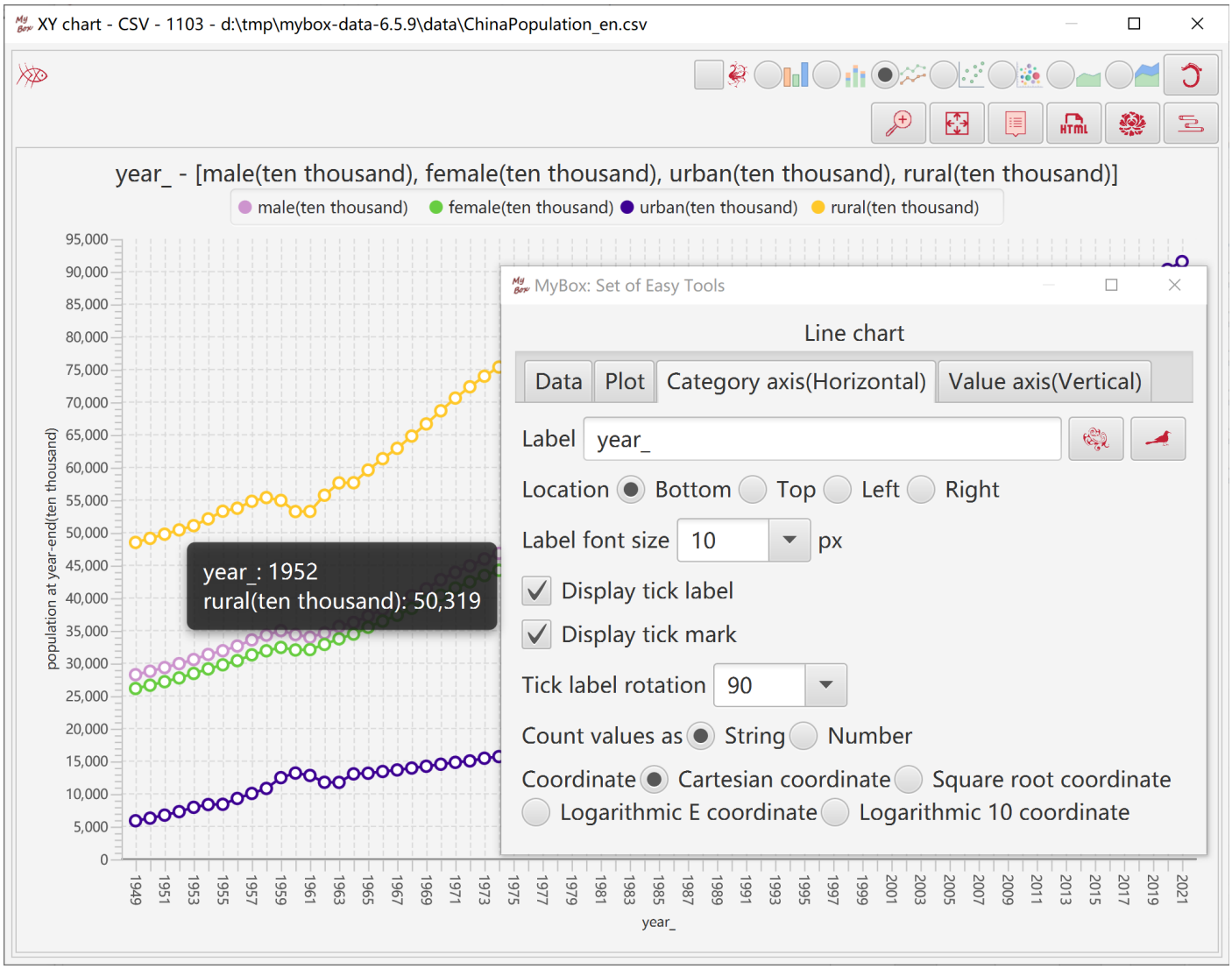
2.7.1.10 Layout

Set parameters of plot: title, font size, location of legend, zero-lines, grid lines, line width, etc.



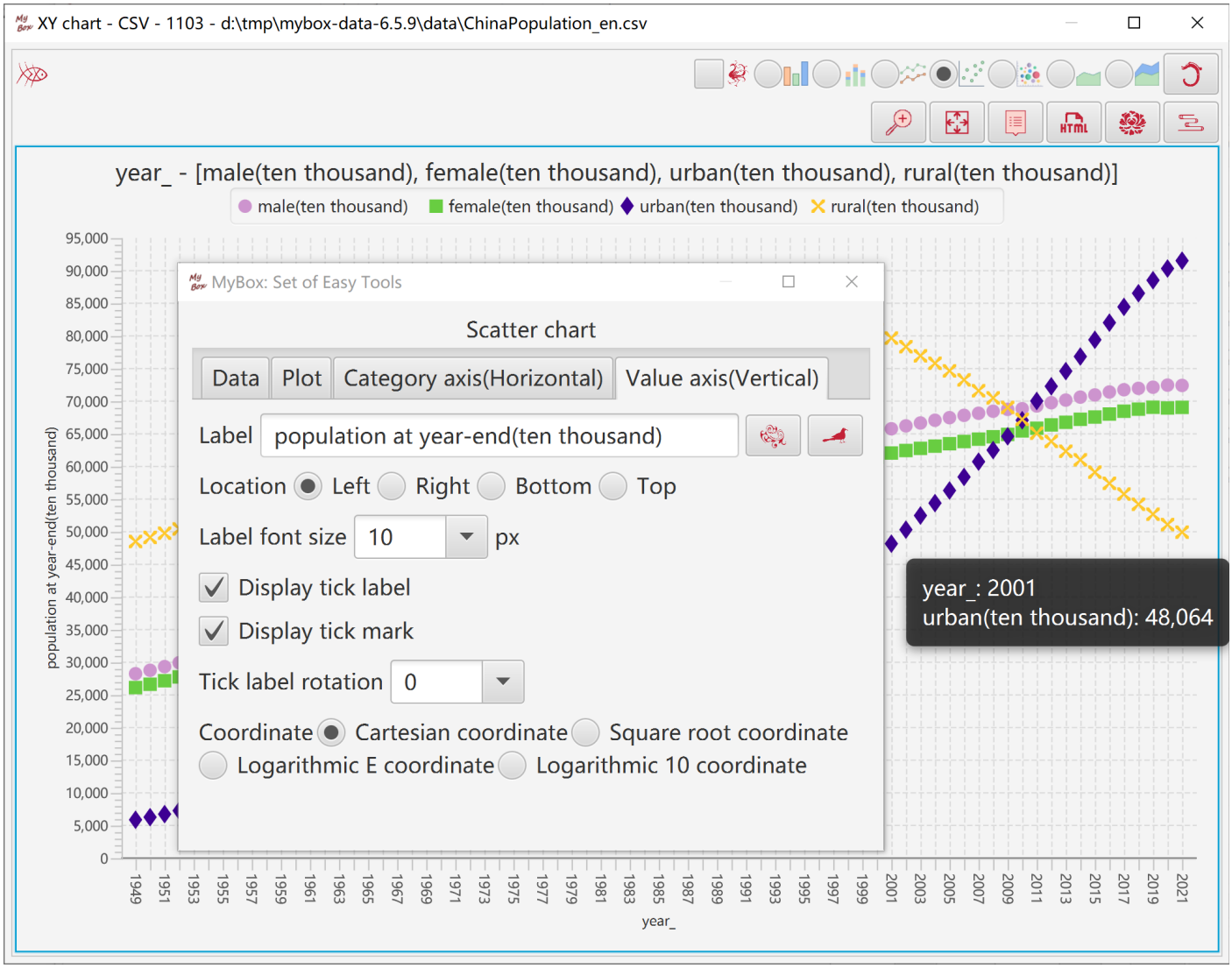
2.7.1.11 Category Axis

Set parameters of category axis: label, font size, location, tick, count values as string or number, coordinate, etc.



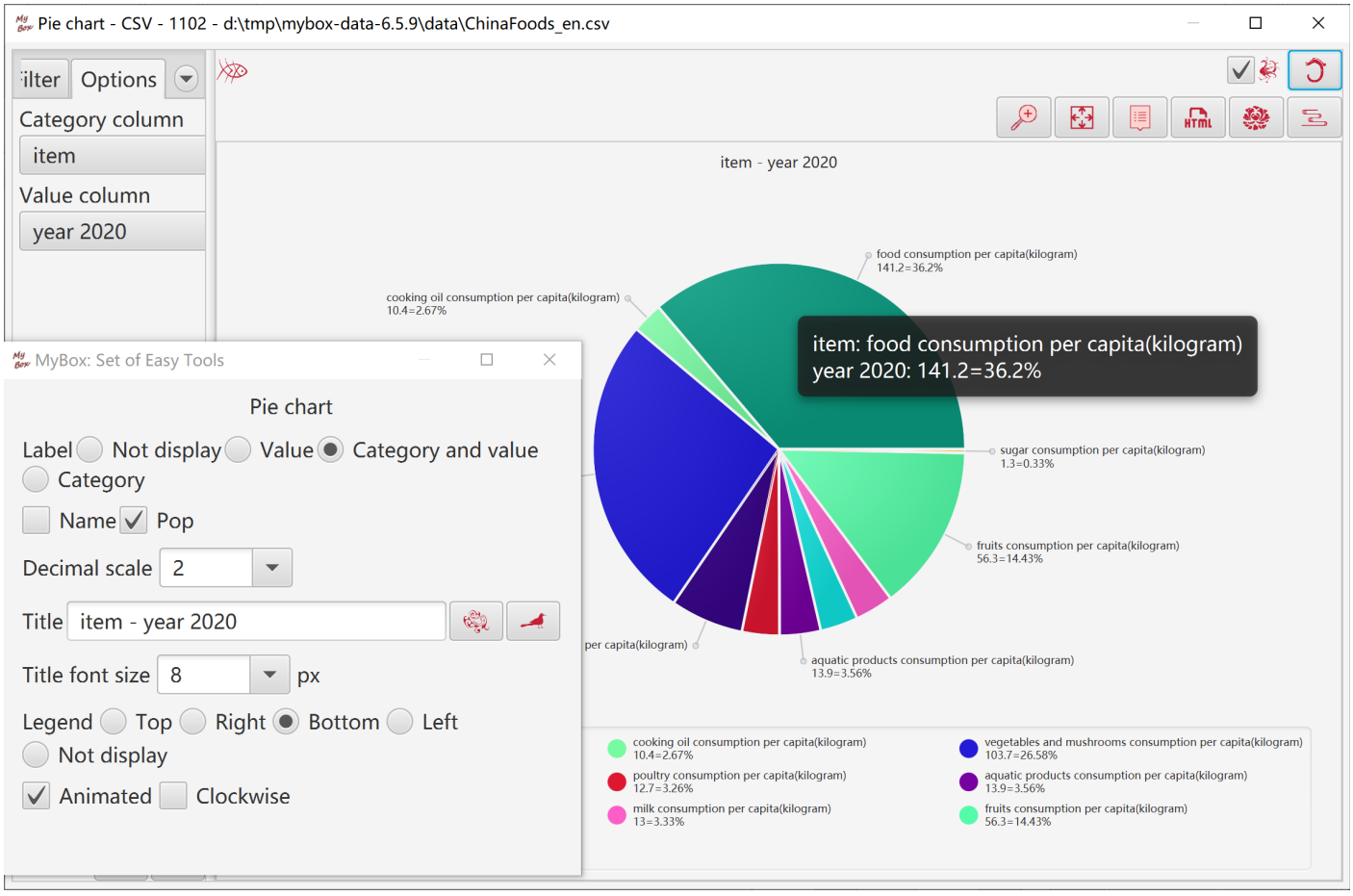
2.7.1.12 Number Axis

Set parameters of number axis: label, font, location, tick, coordinate, etc.



2.7.2 Pie Chart

1. Select data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
2. Calcuate:
   - Select one column as "Category Axis", to define data names..
   - Select another column as "Number Axis".
   - Data numbers are represented as percentages with a circle divided into segments.
   - Value column should be non-negative.
   - When display all data rows(all pages), need concern memory limitation.
3. Click button “Menu” to set parameters of chart.
4. Click button “Pop” to display current chart as image in popped window.
5. Click button “Data” to display data of the chart in data table.
6. Click button “Html” to display data of the chart in html page.



2.7.3 Box-and-whisker Chart

1. Box-and-whisker chart represents data distribution:
   - Sort data according to column/row/all in ascending order.
   - Following items can show aggregation and discreteness of data:

     Minimum Q0 = in 0%(start) of the data list

     Lower quartile Q1 = in 25% of the data list

     Median Q2 = in 50%(middle) of the data list

     Upper quartile Q3 = in 75% of the data list

     Maximum Q4 = in 100%(end) of the data list
   - Following values can be used to mark ourliers of data:

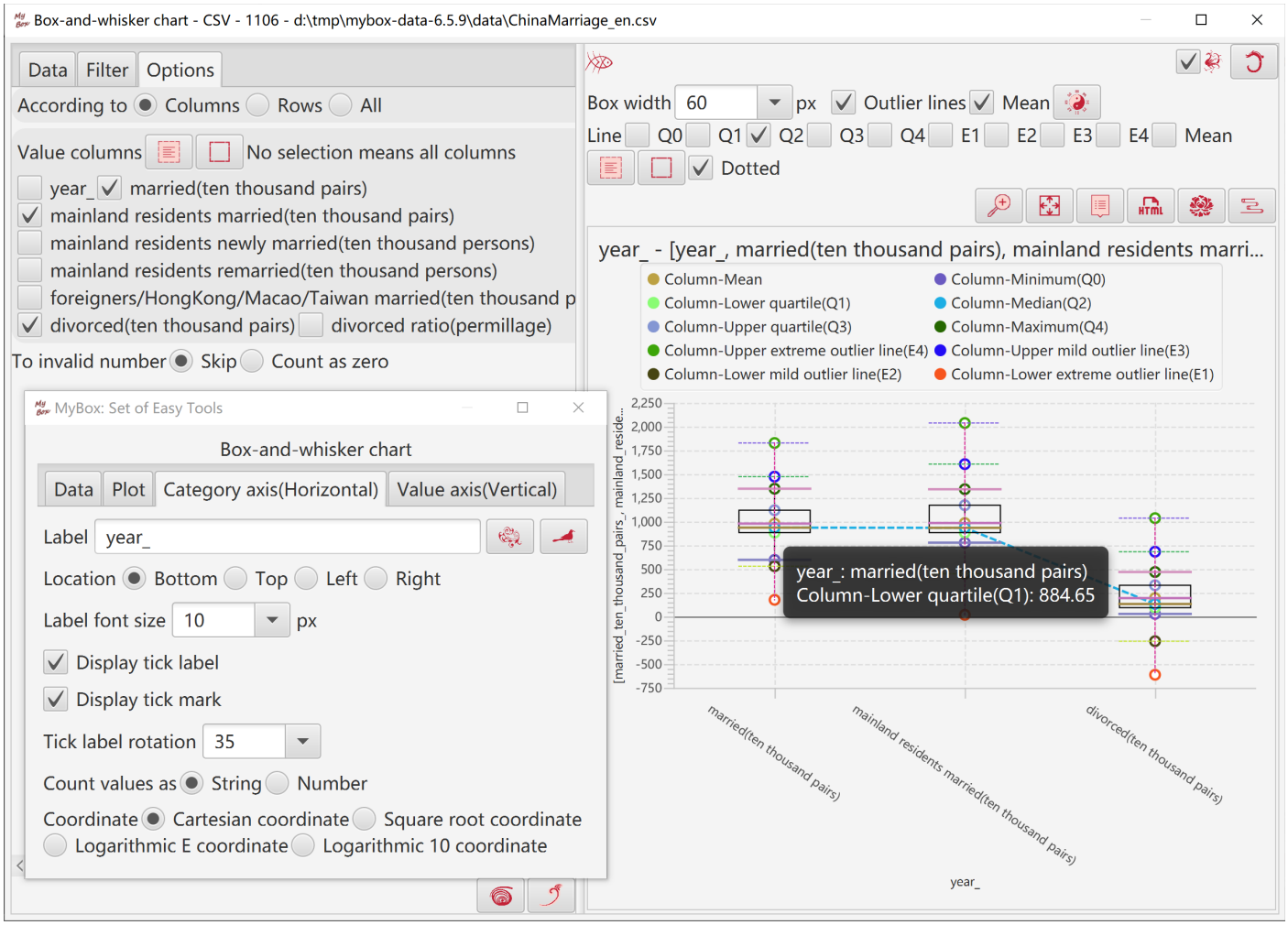
     Lower extreme outlier line E1 = Q1 - 3 * (Q3 - Q1)

     Lower mild outlier line E2 = Q1 - 1.5 * (Q3 - Q1)

     Upper mild outlier line E3 = Q3 + 1.5 * (Q3 - Q1)

     Upper extreme outlier line E4 = Q3 + 3 * (Q3 - Q1)
   - Following values can be referred for discreteness:

     Mean = avarage of the data list
2. Calcuate:
   - Based on line chart.
   - According to: columns, rows(select category column), or all.
3. Set or select:
   - Box width.
   - Whether display outliers lines or mean.
   - Whether display connection lines of values, and whether dotted lines.
   - Random colors.
4. Click button “Menu” to set parameters of chart.
5. Click button “Pop” to display current chart as image in popped window.
6. Click button “Data” to display data of the chart in data table.
7. Click button “Html” to display data of the chart in html page.


Box-and-whisker chart - CSV - 1106 - d:\tmp\mybox-data-6.5.9\data\ChinaMarriage_en.csv
Data
Filter
Options
According to
Columns
Rows
All
Value columns
No selection means all columns
year_
married(ten thousand pairs)
mainland residents married(ten thousand pairs)
mainland residents newly married(ten thousand persons)
mainland residents remarried(ten thousand persons)
foreigners/HongKong/Macao/Taiwan married(ten thousand p
divorced(ten thousand pairs)
divorced ratio(permillage)
To invalid number
Skip
Count as zero
Box width
60
px
Outlier lines
Mean
Line
Q0
Q1
Q2
Q3
Q4
E1
E2
E3
E4
Mean
Dotted
year_ - [year_, married(ten thousand pairs), mainland residents marri...
Column-Mean
Column-Minimum(Q0)
Column-Lower quartile(Q1)
Column-Median(Q2)
Column-Upper quartile(Q3)
Column-Maximum(Q4)
Column-Upper extreme outlier line(E4)
Column-Upper mild outlier line(E3)
Column-Lower mild outlier line(E2)
Column-Lower extreme outlier line(E1)
year_: married(ten thousand pairs)
Column-Lower quartile(Q1): 884.65
[married_ten_thousand_pairs_, mainland_reside...
married(ten thousand pairs)
mainland residents married(ten thousand pairs)
divorced(ten thousand pairs)
year_
MyBox: Set of Easy Tools
Box-and-whisker chart
Data
Plot
Category axis(Horizontal)
Value axis(Vertical)
Label
year_
Location
Bottom
Top
Left
Right
Label font size
10
px
Display tick label
Display tick mark
Tick label rotation
35
Count values as
String
Number
Coordinate
Cartesian coordinate
Square root coordinate
Logarithmic E coordinate
Logarithmic 10 coordinate

2.7.4 Self Comparison Bars Chart

1. Self comparison bars show difference between data and reference values.

   Bars are calculated by following rulers:

   - If value is zero, no bar
   - When compare as absoluate values:

     maximum_value = maximum_absolute_value_of_column/row/all

     precentage = absolute_value / maximum_value

     width = maximum_width * precentage

     color = If value is larger than zero, color_of_column.

     If value is less than zero, inverted_color_of_column
   - When compare as range of minimum and maximum:

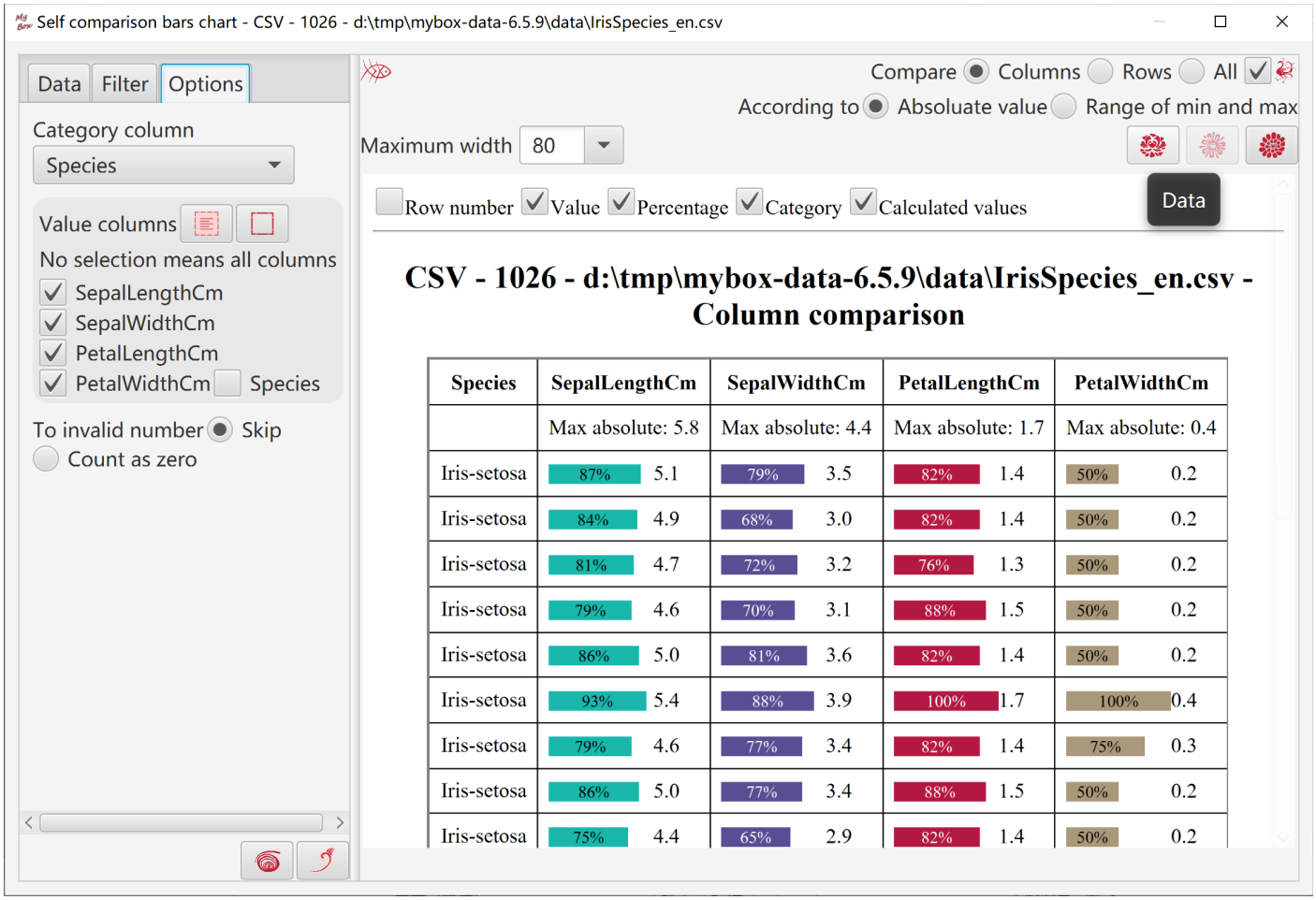
     maximum_value = maximum_value_of_column/row/all

     minimum_value = minimum_value_of_column/row/all

     precentage = (value - minimum_value) / (maximum_value - minimum_value)

     width = maximum_width * precentage

     color = color_of_column
2. Data: Select rows in table or select all data rows(all pages), and select columns.
3. Calcuate:
   - Select one column as category column(unnecessary)
   - Compare: columns/rows/all.
   - According to: absoluate values, or range of minimum and maximum.
   - When display all data rows(all pages), need concern memory limitation.
4. Set or select:
   - Maximum width
   - Whether display row numbers, values, percentages, categories, calculated values.
5. Edit data in chart.
6. Edit html of chart.


Self comparison bars chart - CSV - 1026 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies_en.csv
Data
Filter
Options
Category column
Species
Value columns
No selection means all columns
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
Species
To invalid number
Skip
Count as zero
Compare
Columns
Rows
All
According to
Absoluate value
Range of min and max
Maximum width
80
Row number
Value
Percentage
Category
Calculated values
Data
CSV - 1026 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies_en.csv - Column comparison
Species
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
Max absolute: 5.8
Max absolute: 4.4
Max absolute: 1.7
Max absolute: 0.4
Iris-setosa 87% 5.1 79% 3.5 82% 1.4 50% 0.2
Iris-setosa 84% 4.9 68% 3.0 82% 1.4 50% 0.2
Iris-setosa 81% 4.7 72% 3.2 76% 1.3 50% 0.2
Iris-setosa 79% 4.6 70% 3.1 88% 1.5 50% 0.2
Iris-setosa 86% 5.0 81% 3.6 82% 1.4 50% 0.2
Iris-setosa 93% 5.4 88% 3.9 100% 1.7 100% 0.4
Iris-setosa 79% 4.6 77% 3.4 82% 1.4 75% 0.3
Iris-setosa 86% 5.0 77% 3.4 88% 1.5 50% 0.2
Iris-setosa 75% 4.4 65% 2.9 82% 1.4 50% 0.2

2.7.5 Comparison Bars Chart

1. Comparison bars show difference between two series of data.

   Bars are calculated by following rulers:

   - If value is zero, no bar
   - When compare as absoluate values:

     maximum_value = maximum_absolute_value_of_value_columns

     precentage = absolute_value / maximum_value

     width = maximum_width * precentage

     color = If value is larger than zero, color_of_column.

     If value is less than zero, inverted_color_of_column
   - When compare as range of minimum and maximum:

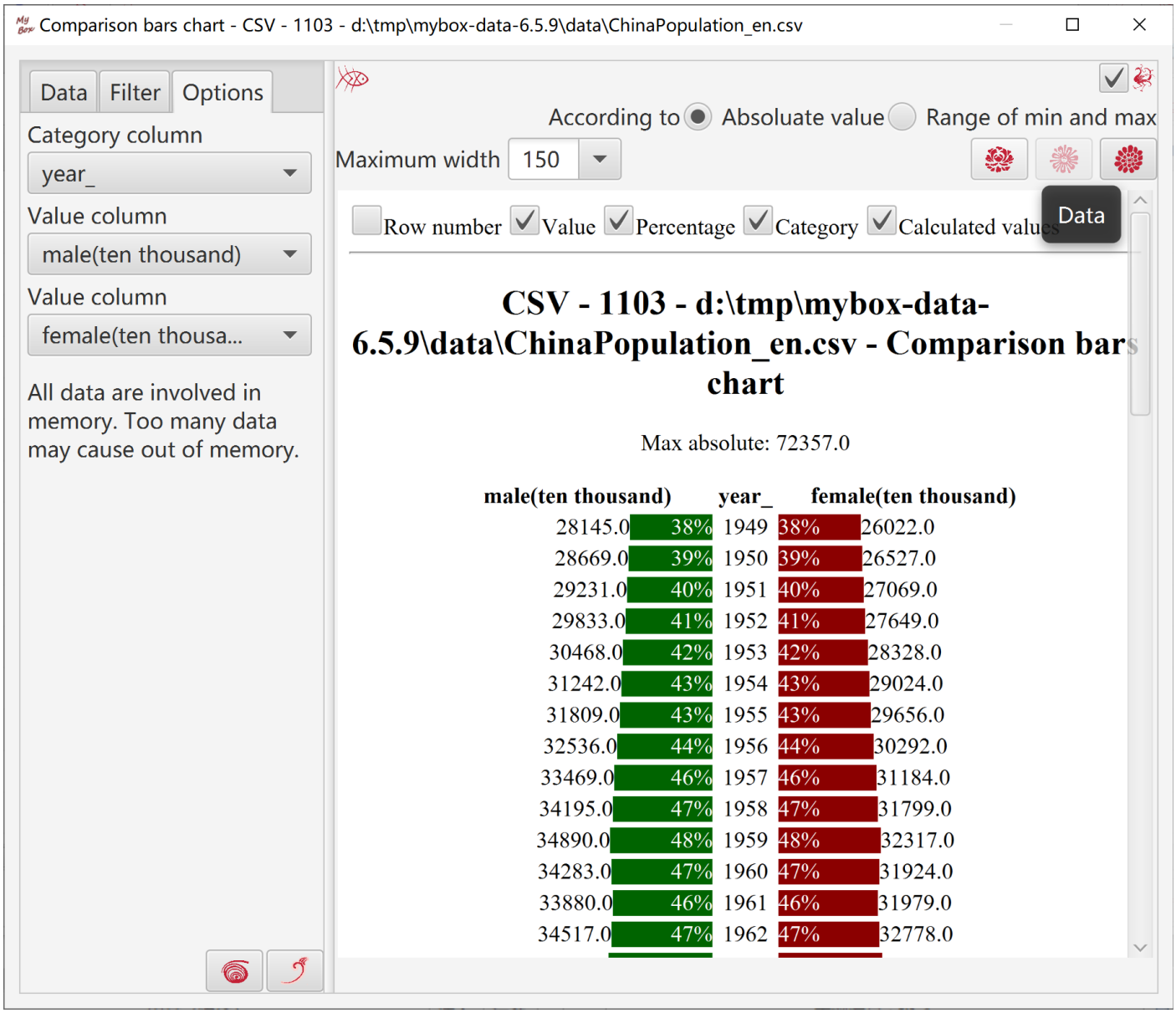
     maximum_value = maximum_value_of_value_columns

     minimum_value = minimum_value_of_value_columns

     precentage = (value - minimum_value) / (maximum_value - minimum_value)

     width = maximum_width * precentage

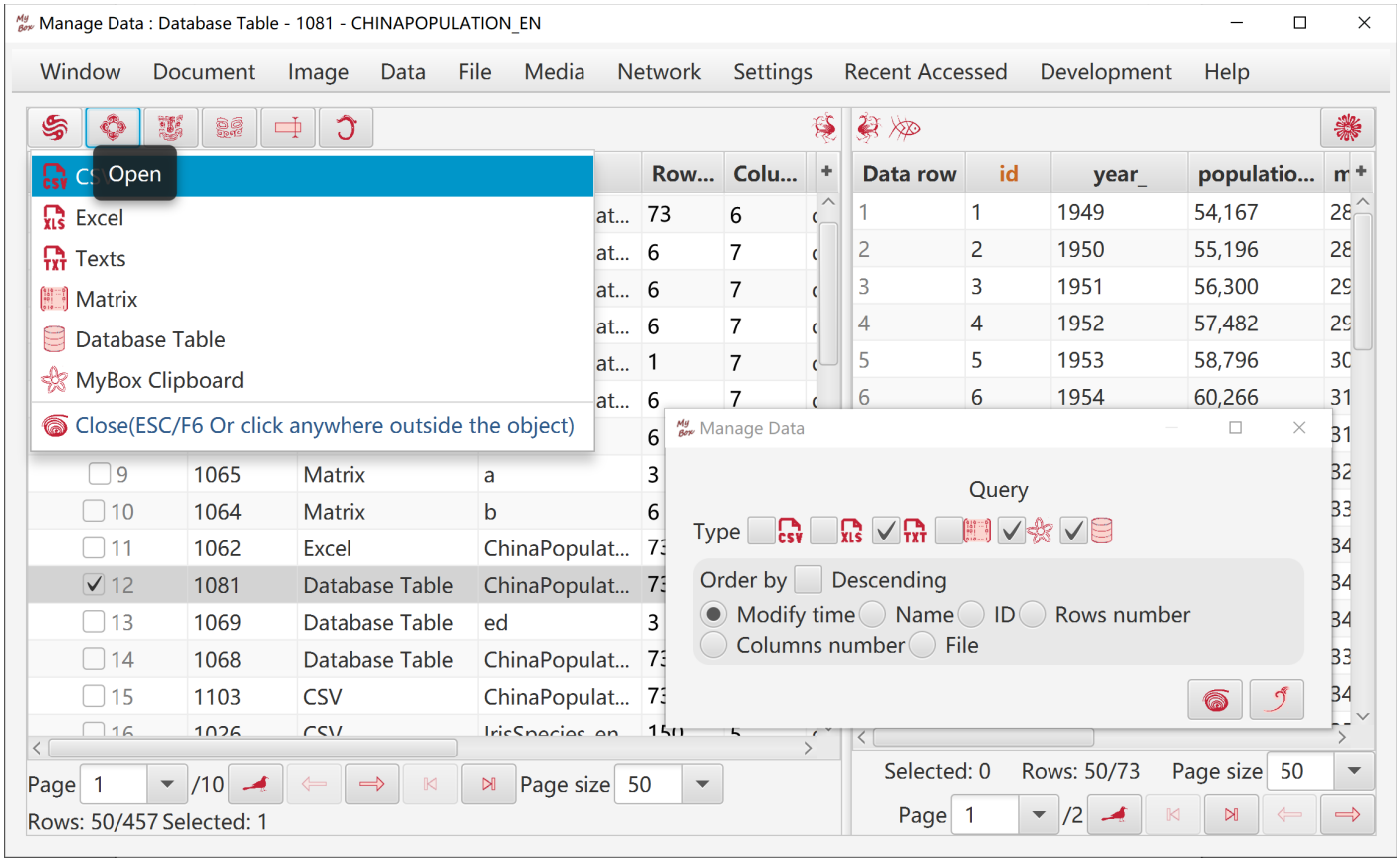
     color = color_of_column
2. Data: Select rows in table or select all data rows(all pages), and select columns.
3. Calcuate:
   - Select one column as category column(unnecessary)
   - Select tow value columns
   - According to: absoluate values, or range of minimum and maximum.
   - When display all data rows(all pages), need concern memory limitation.
4. Set or select:
   - Maximum width
   - Whether display row numbers, values, percentages, categories, calculated values.
5. Edit data in chart.
6. Edit html of chart.


Comparison bars chart - CSV - 1103 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_en.csv
Data
Filter
Options
Category column
year_
Value column
male(ten thousand)
Value column
female(ten thousa...
All data are involved in memory. Too many data may cause out of memory.
According to
Absolute value
Range of min and max
Maximum width
150
Row number
Value
Percentage
Category
Calculated values
Data
CSV - 1103 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_en.csv - Comparison bars chart
Max absolute: 72357.0
male(ten thousand) year_ female(ten thousand)
28145.0 38% 1949 38% 26022.0
28669.0 39% 1950 39% 26527.0
29231.0 40% 1951 40% 27069.0
29833.0 41% 1952 41% 27649.0
30468.0 42% 1953 42% 28328.0
31242.0 43% 1954 43% 29024.0
31809.0 43% 1955 43% 29656.0
32536.0 44% 1956 44% 30292.0
33469.0 46% 1957 46% 31184.0
34195.0 47% 1958 47% 31799.0
34890.0 48% 1959 48% 32317.0
34283.0 47% 1960 47% 31924.0
33880.0 46% 1961 46% 31979.0
34517.0 47% 1962 47% 32778.0

2.8 Manage Data

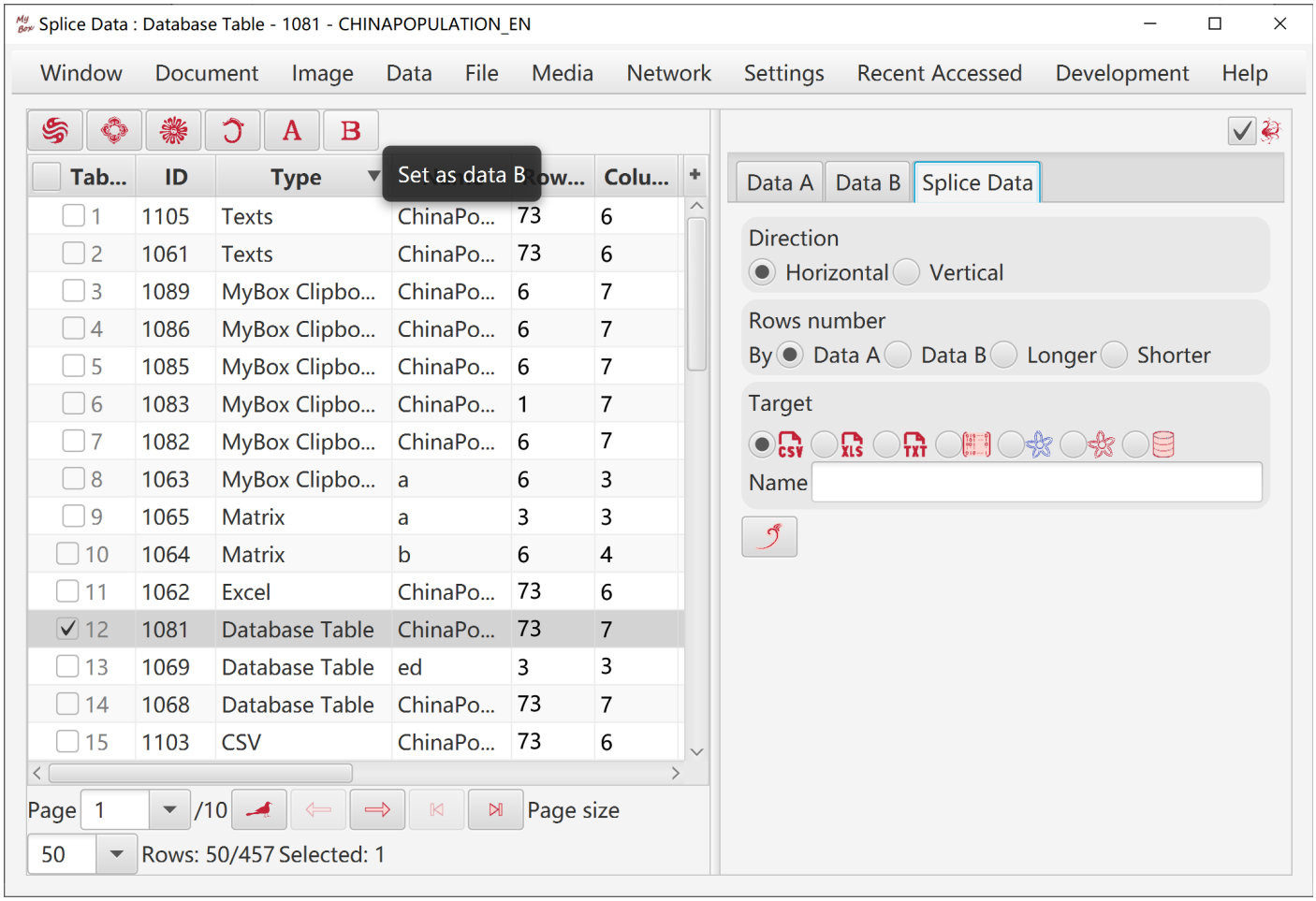
This tool manages following objects:

1. Data Files
   - Record is created/updated when csv/excel/texts data file is opened by its editor.
   - Data are saved in data file.
   - Deleting record of data file will not delete data file itself.
2. Data Clipboards
   - Record is created when data is copied into MyBox Clipboad.
   - Data are saved in file under MyBox internal path.
   - Deleting record of data clipboards will delete its internal file.
3. Matrices
   - Records are maintained by Matrices Manager.
   - Data are saved in MyBox database.
   - Deleting record of matrix will delete data of this matrix.
4. Data Tables
   - Records are maintained by Data Tables Manager.
   - Data are saved in MyBox database tables.
   - Deleting record of data table will delete data of this data table.



2.9 Splice Data

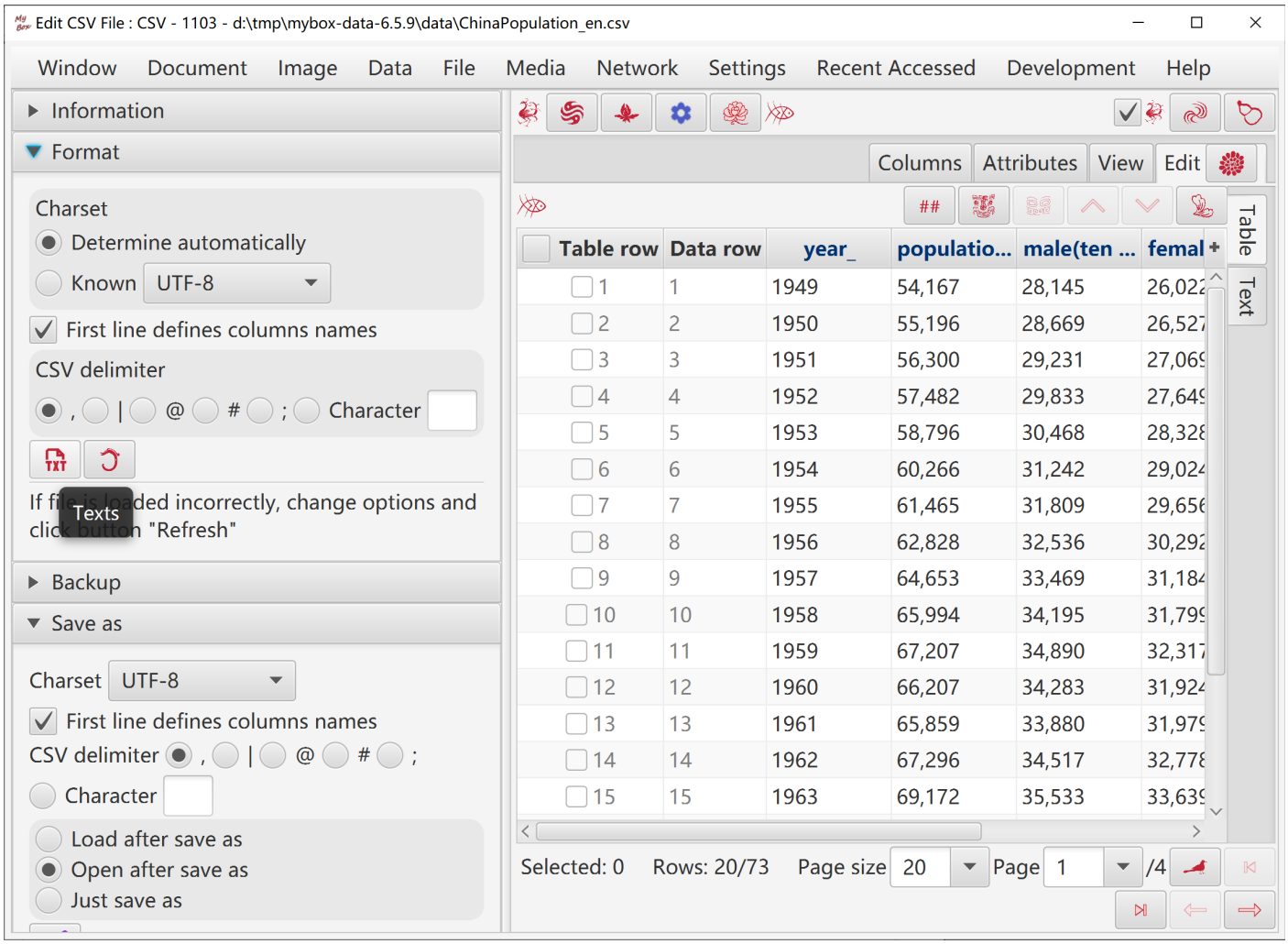
1. Select or open two data.
2. Select rows and columns from the two data:
   - Rows can be: current page, selected rows, or all pages.
   - Select columns. If no column is selected, then all columns are taken.
   - Set row filter.
3. Options:
   - Direction: vertical, horizontal
   - Rows/Columns number by: Data A, Data B, longer, shorter.
4. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.



2.10 Data File

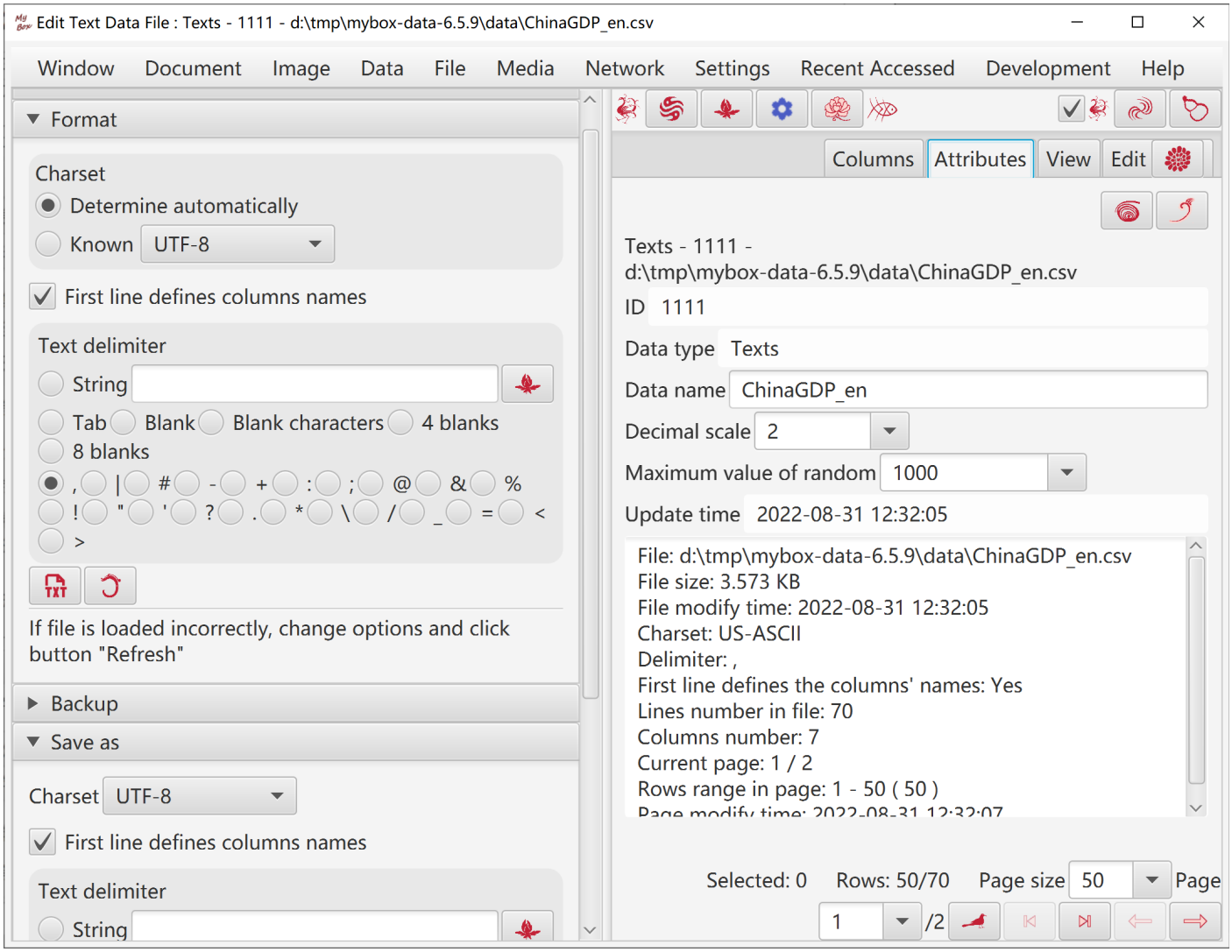
2.10.1 CSV File

1. When file is loaded abnormally, change options and click Refresh button.
2. Options: charset, whether has first line as field names, and delimiter of data.



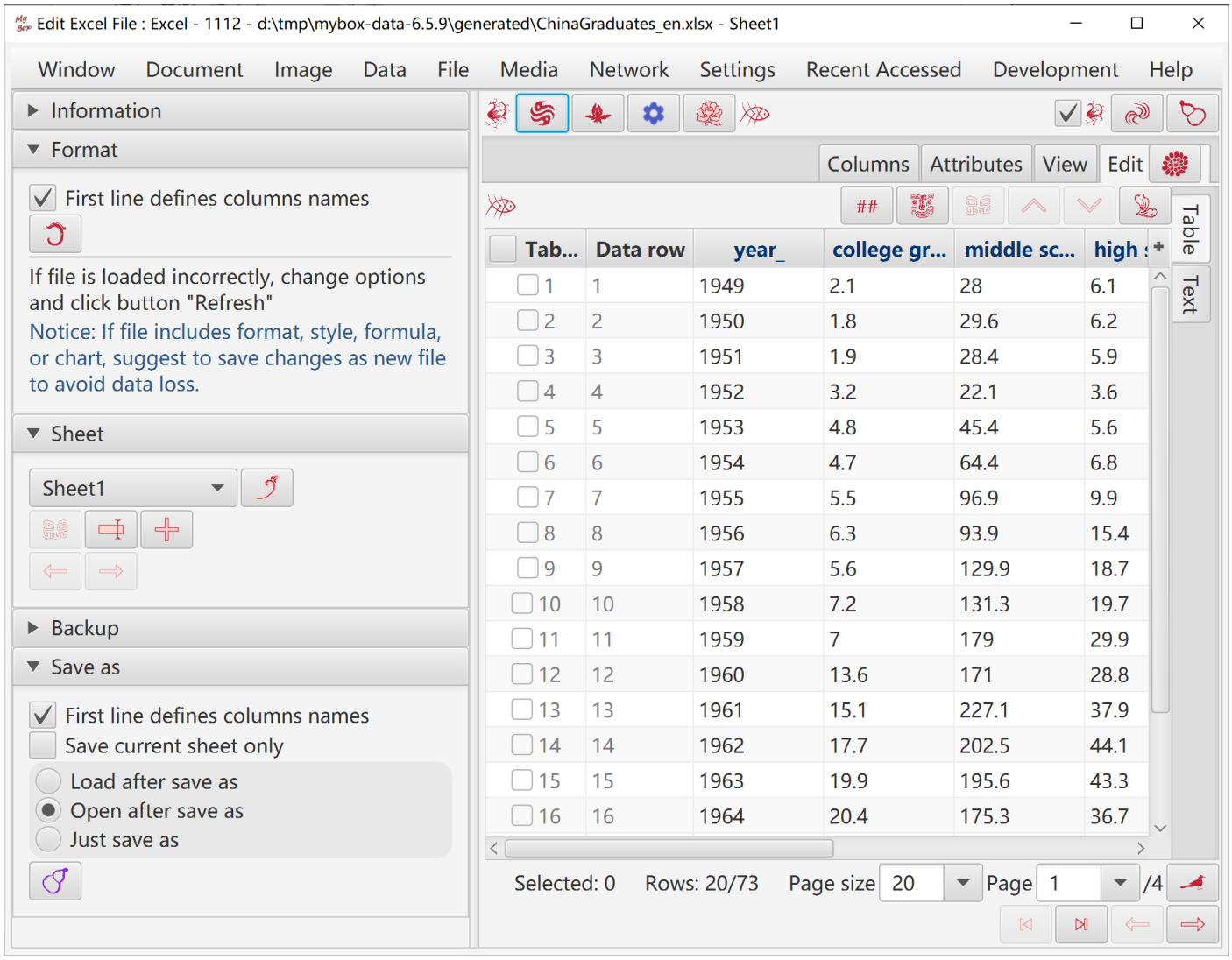
2.10.2 Text File

1. When file is loaded abnormally, change options and click Refresh button.
2. Options: charset, whether has first line as field names, and delimiter of data.



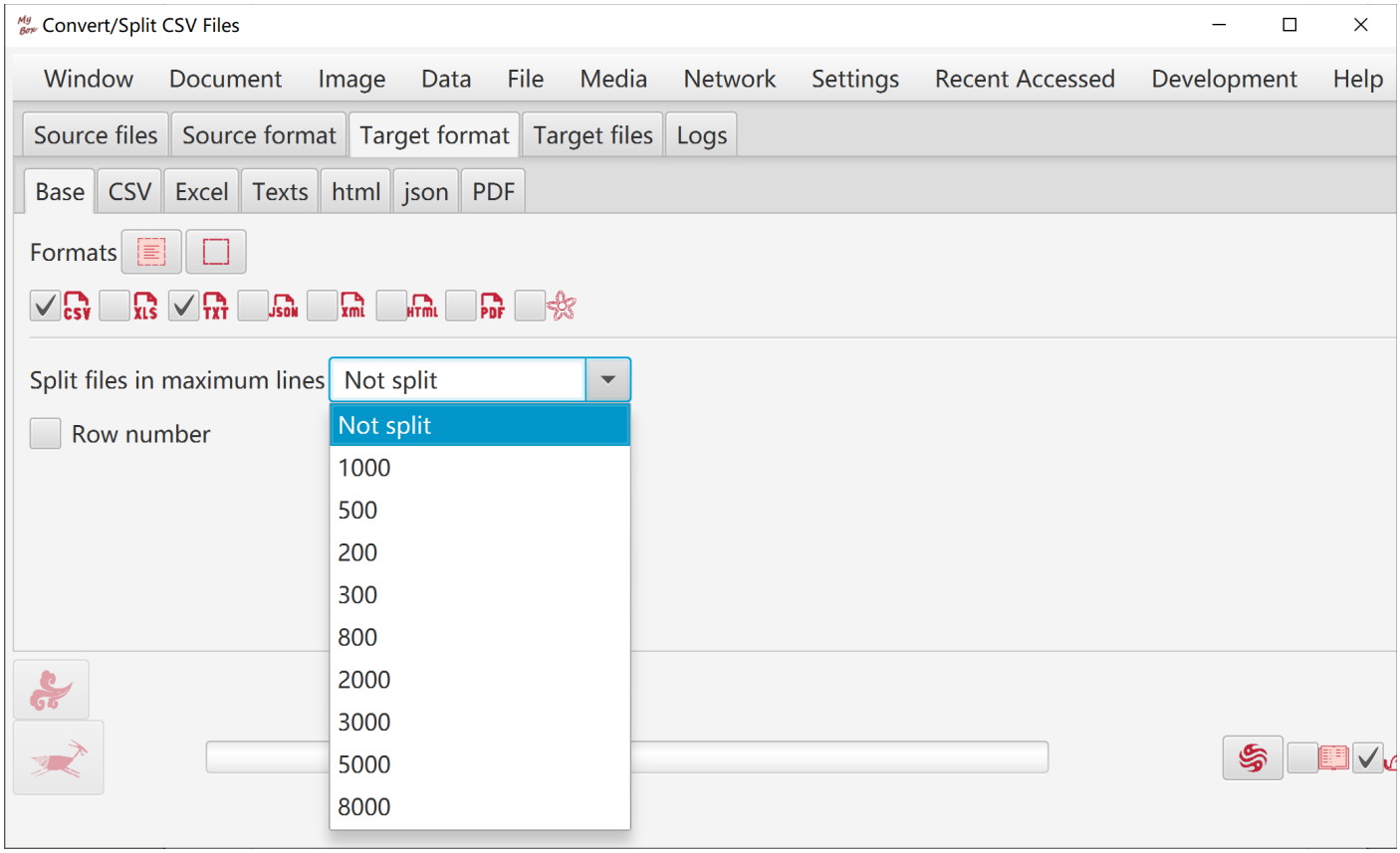
2.10.3 Excel File

1. Options include sheet number and whether has first line as field names.
2. Add/Delete/Rename sheets.
3. Tool can only handle base data in Excel file. If file includes format, style, formula, or chart, suggest to save changes as new file to avoid data loss.



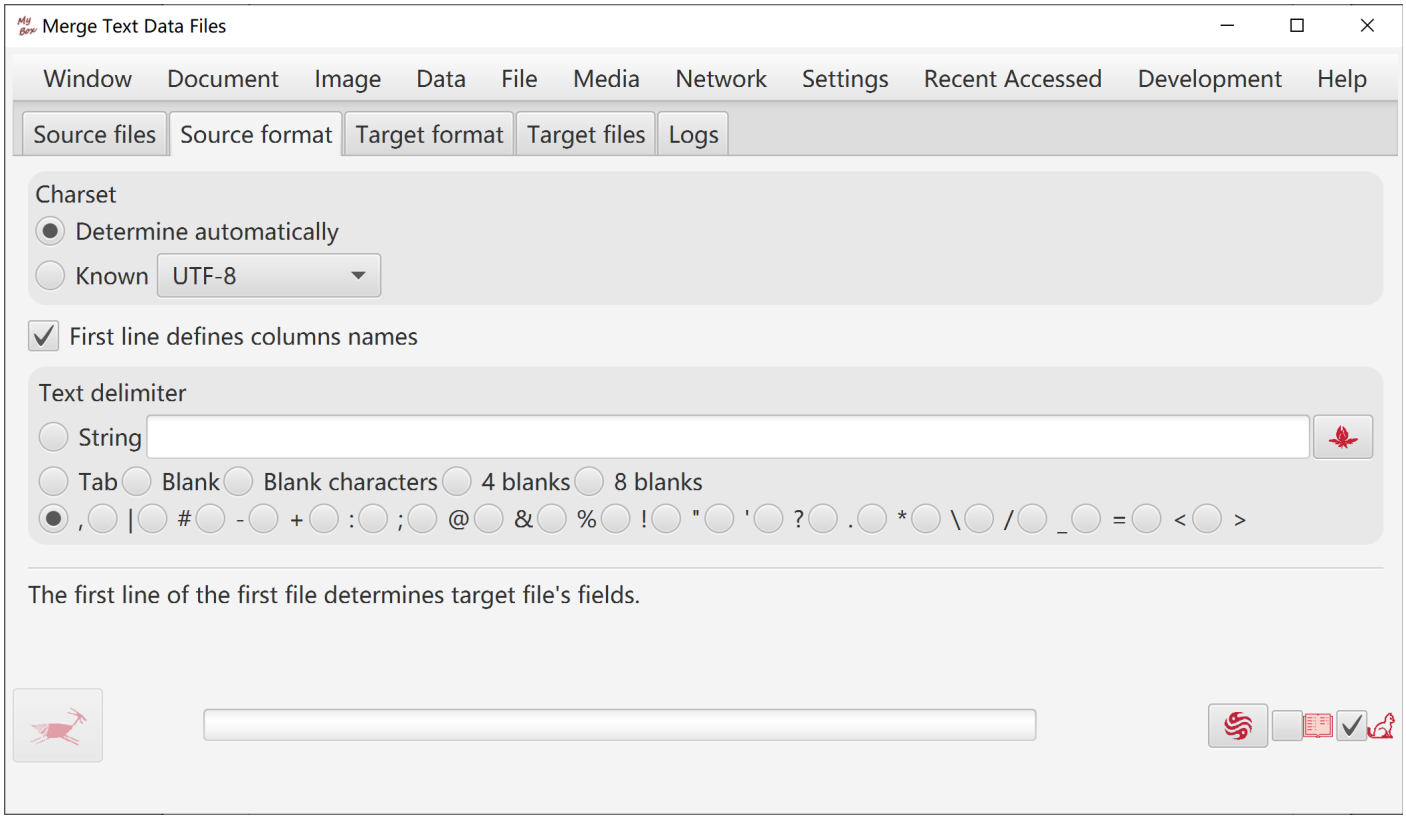
2.10.4 Convert/Split Data Files in Batch

1. Source files' formats can be csv, excel, and text. Options of source files can be set.
2. Target files' formats include csv, text, excel, xml, json, html, pdf. Options of target files can be set.
3. Split files as maximum lines.



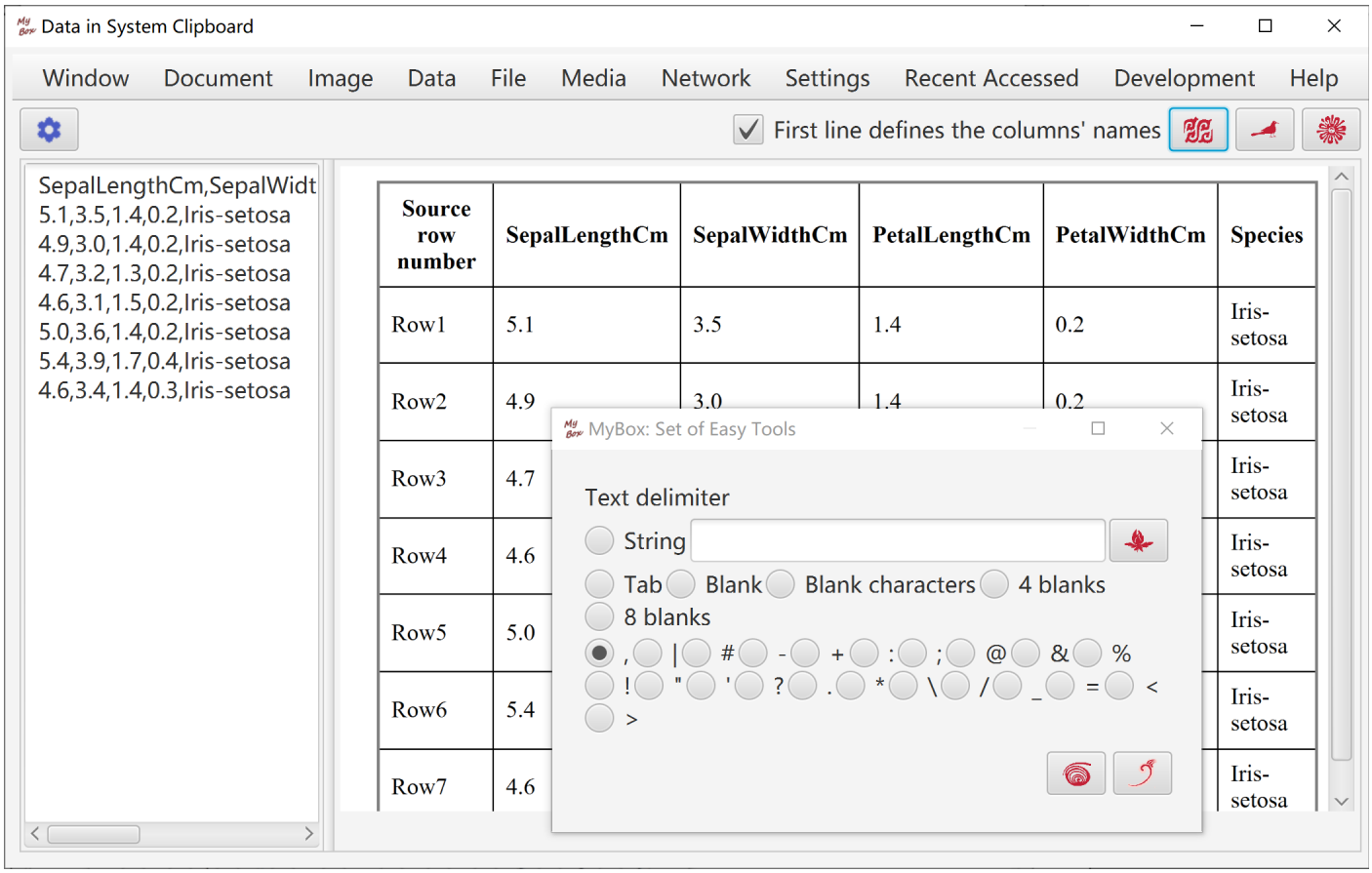
2.10.5 Merge Data Files

1. Set source format.
2. Set target format.

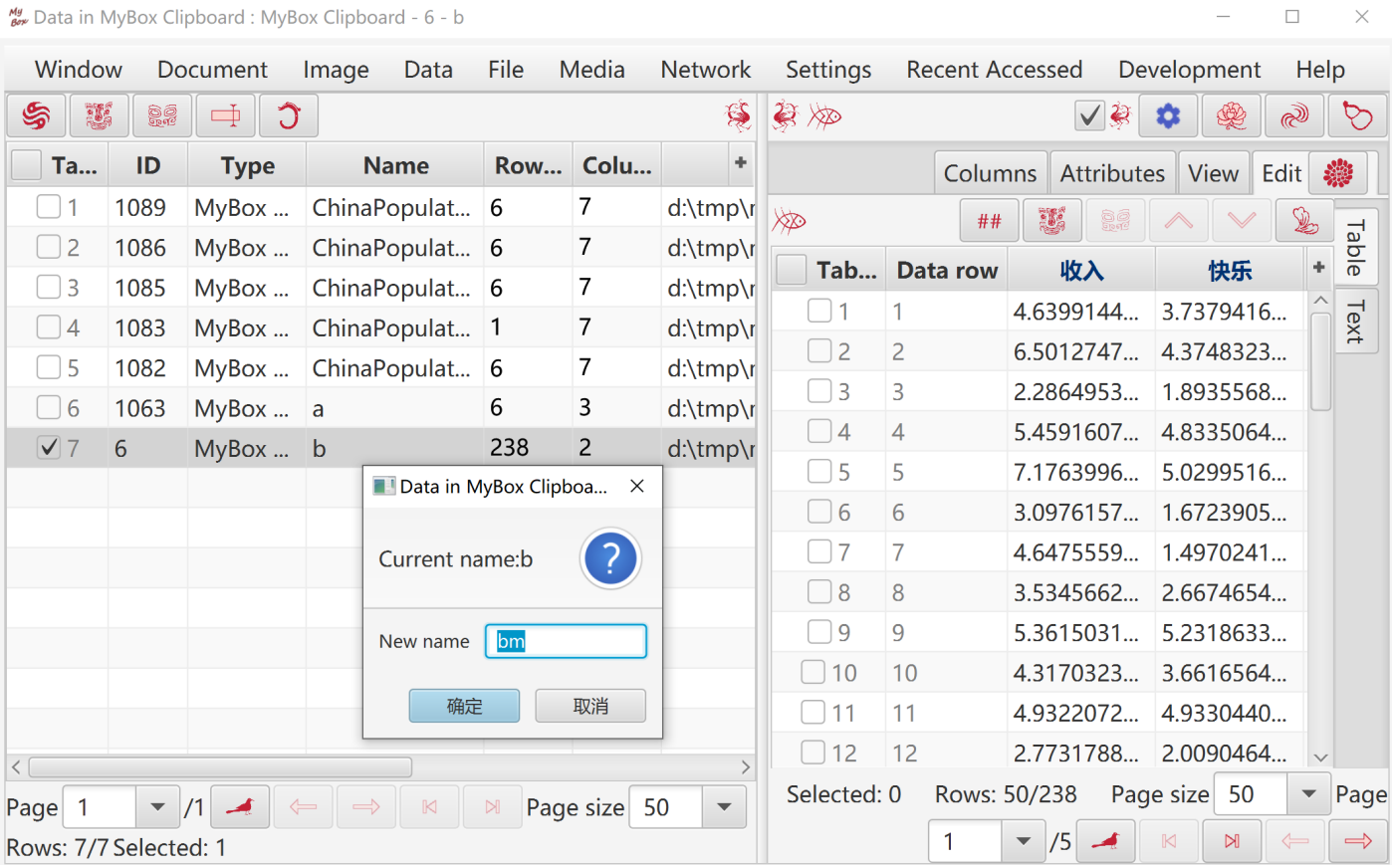


2.11 Data in System Clipboard

1. Read and parse contents in system clipboard.
2. Delimiter can be chosen from special characters or inputted regular expression.
3. First row can be set as column names.



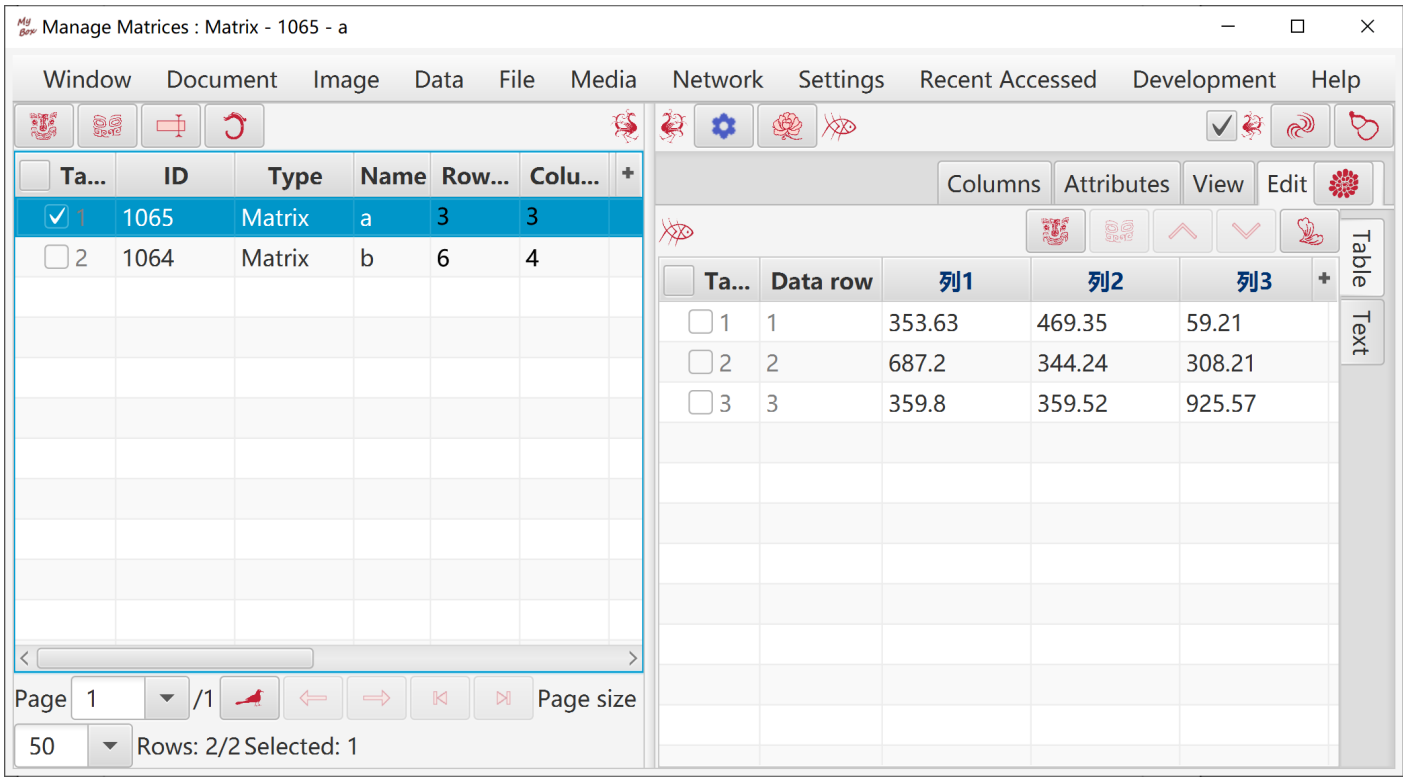
2.12 Data in MyBox Clipboard



2.13 Matrix

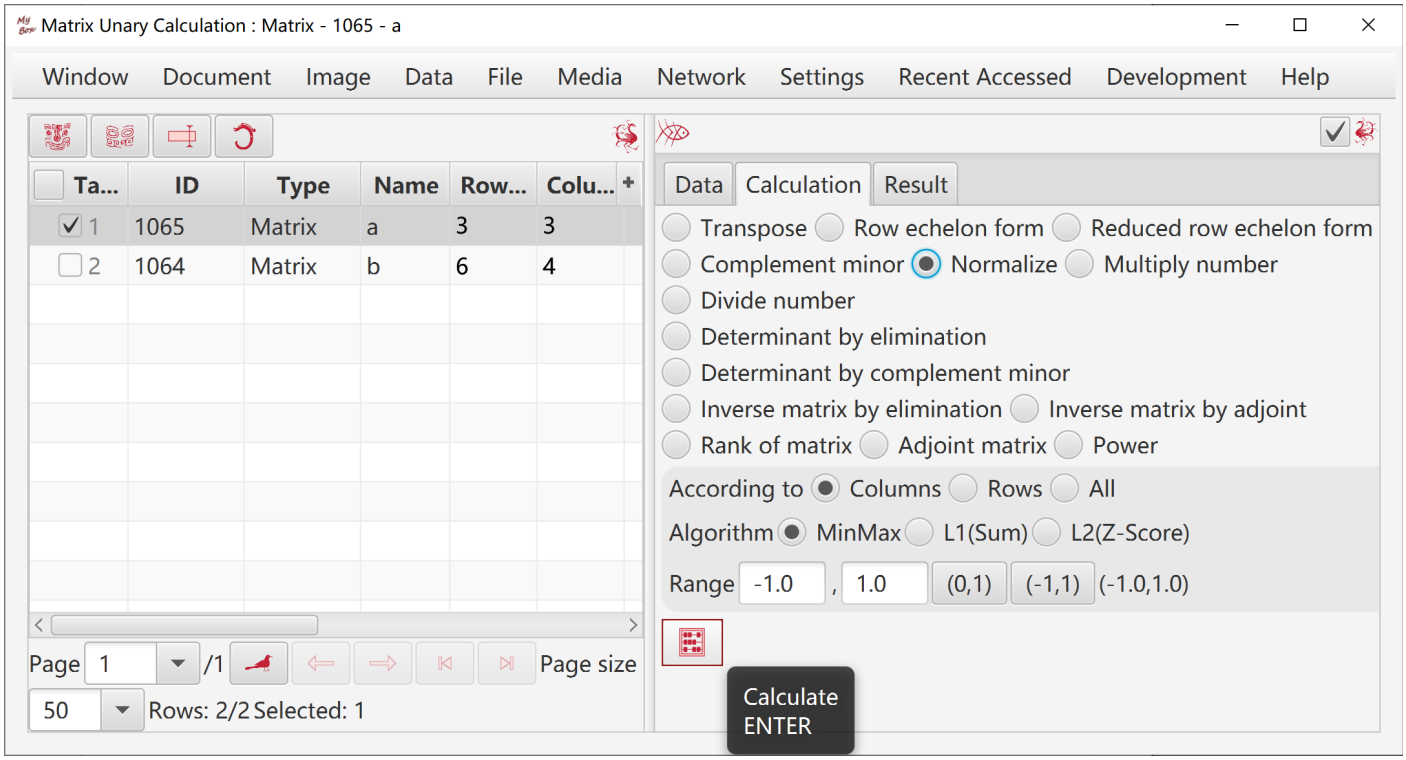
2.13.1 Edit and Manage Matrices

1. Edit matrix.
2. Matrix can be saved and reused.



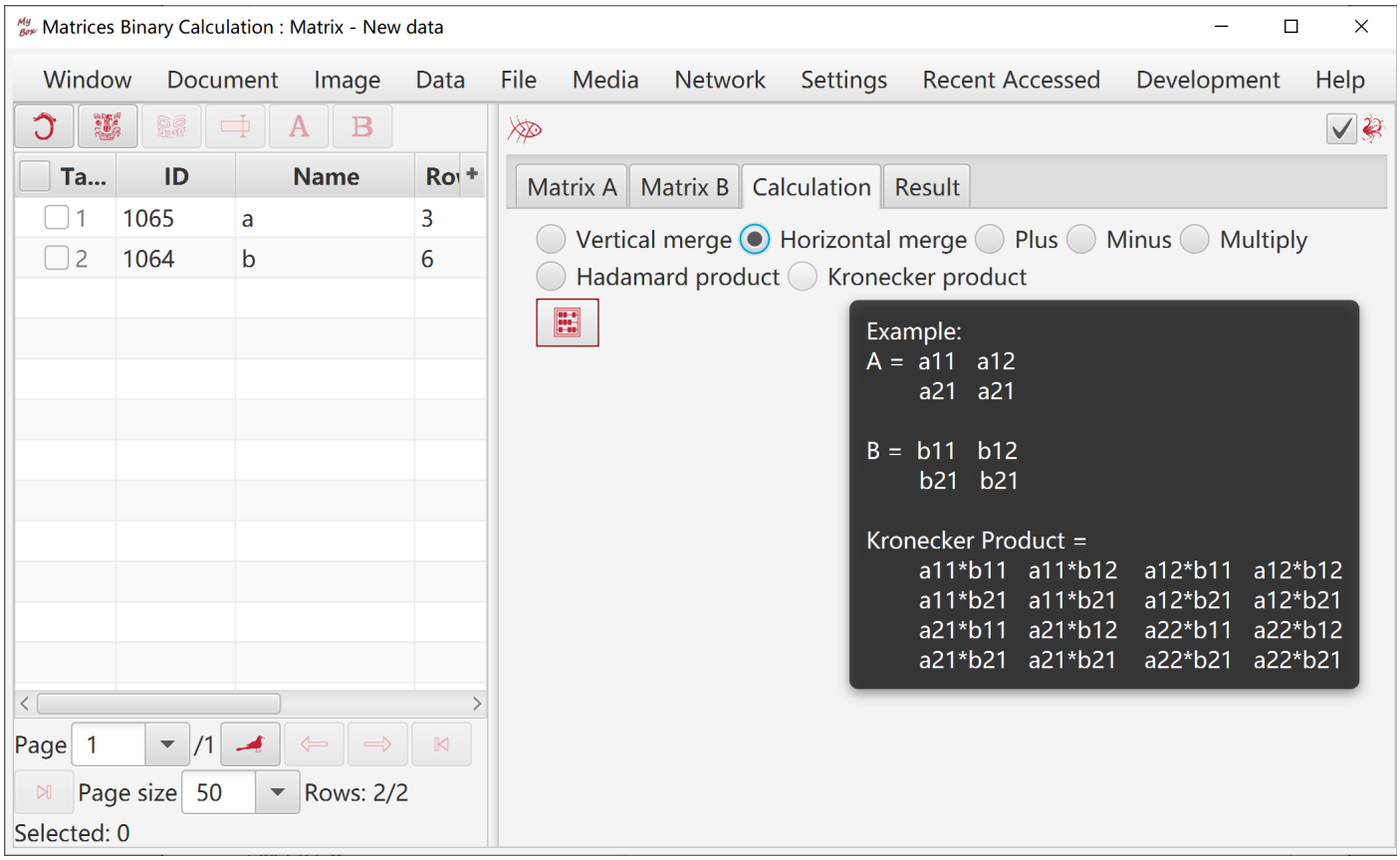
2.13.2 Unary Matrix Calculation

Transpose, Row Echelon Form, Reduced Row Echelon Form, Determinant By Elimination, Determinant By Complement Minor, Inverse Matrix By Elimination, Inverse Matrix By Adjoint, Matrix Rank, Adjoint Matrix, Complement Minor, Normalize, Multiply Number, Divide By Number, Power.



2.13.3 Binary Matrices Calculation

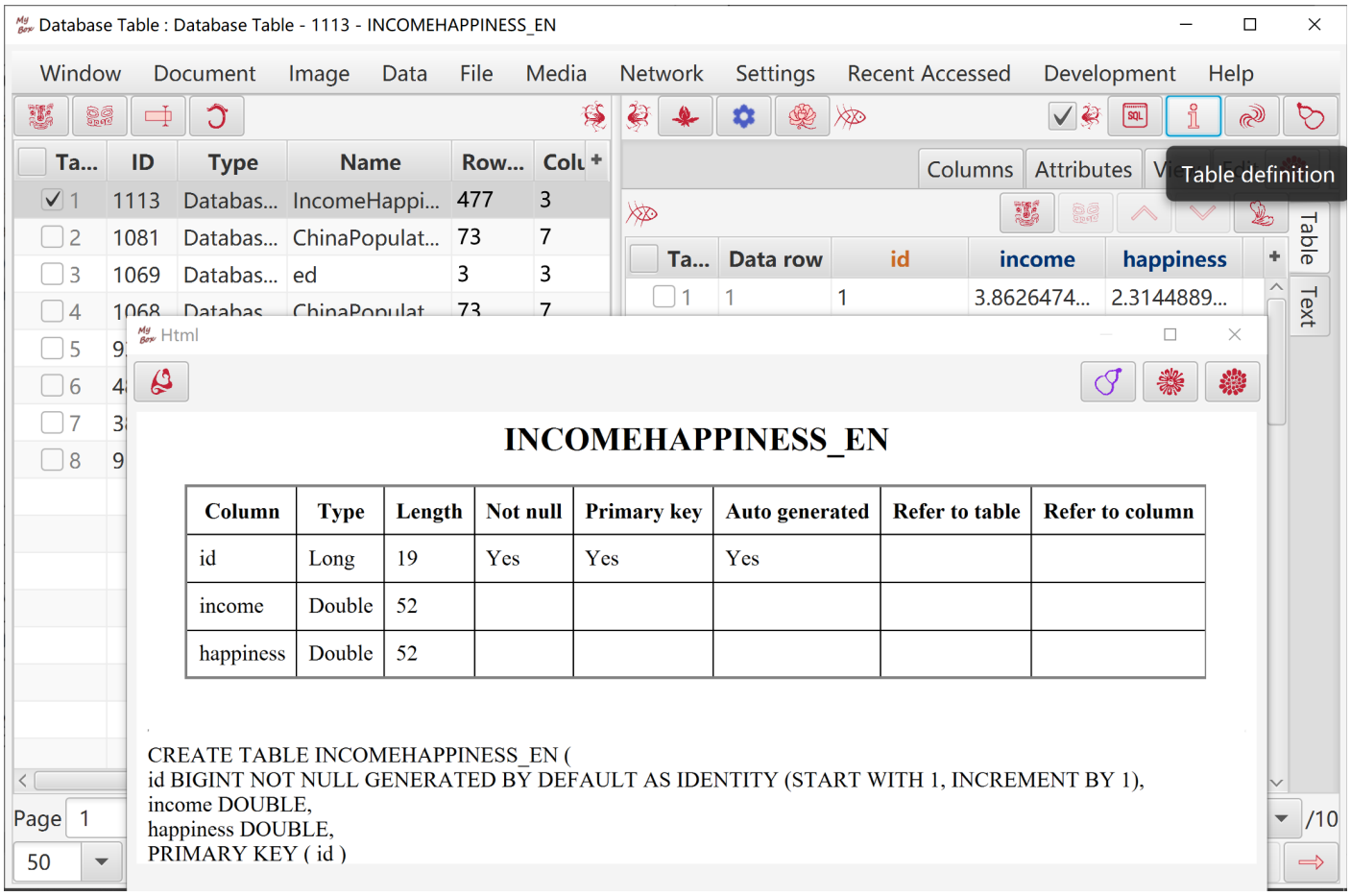
Plus, Minus, Hadamard Product, Kronecker Product, Horizontally Merge, Vertically Merge.



2.14 Database Tables

2.14.1 Manage Database Tables

1. View table definition.
2. Execute SQL.

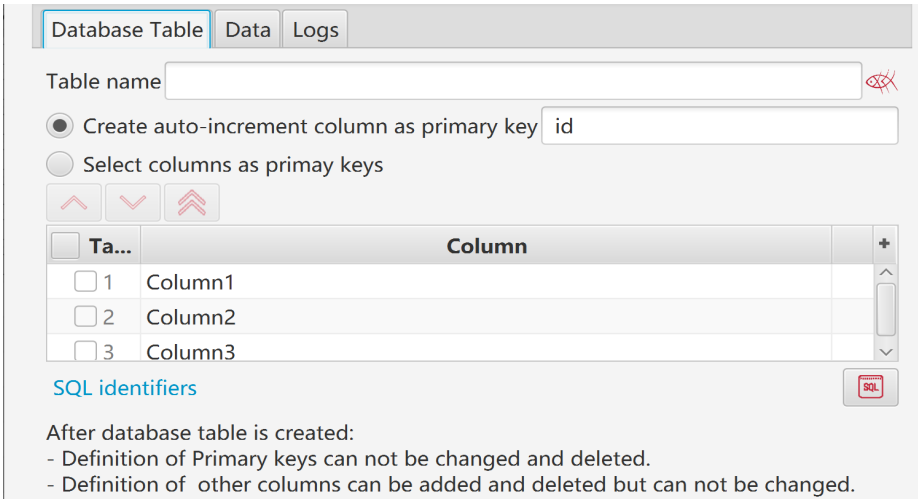


2.14.2 Limitations of SQL Identifier

1. Table name and column names should satisfy "Limitations of SQL identifier":
    - Maximum length is 128.
    - "Ordinary identifier":
        - Not surrounded by double quotation marks.
        - Must begin with a letter.
        - Contains only letters, underscore characters (_), and digits.
        - Permits Unicode letters and digits.
        - Can not be reserved words.
        - It is converted as uppercase when saved in database.
        - It is case-insensitive when referred in SQL statement.

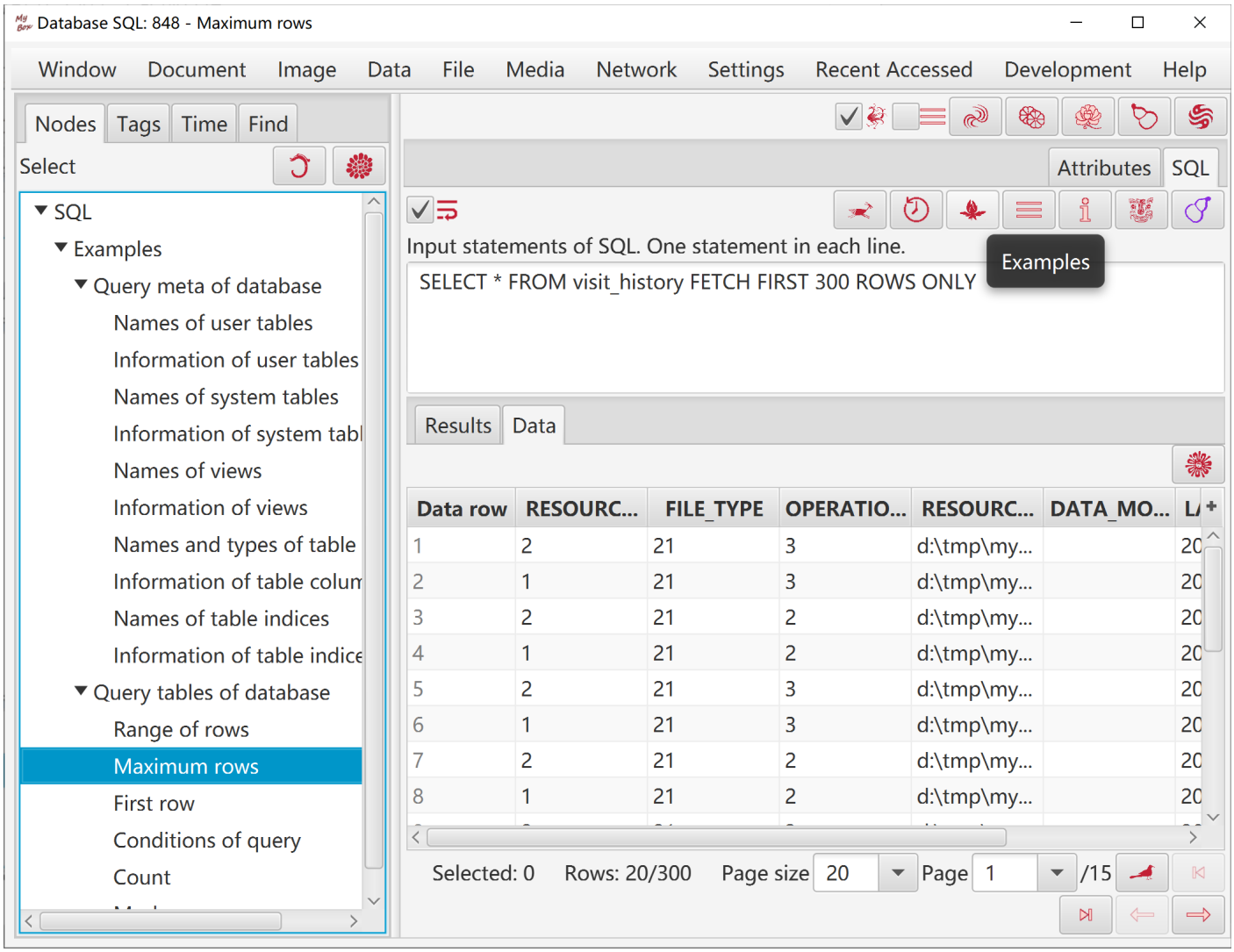
          Example, AbC is same as abc and aBC.
    - "Delimited identifier":
        - Surrounded by double quotation marks.
        - Can contain any characters.
        - It is saved as string inside the double quotations in database.
        - It should be surrounded by double quotations when referred in SQL statement, except for following: It only includes upper case letters and underscores.

          Example, "AbC" is different from AbC or "ABC" while "ABC" is same as ABC and abc.
2. After database table is created:
    - Definition of Primary keys can not be changed and deleted.
    - Definition of other columns can be added and deleted but can not be changed.
3. When MyBox create name of table/column:
    - Invalid characters are converted as underscore characters.
    - If it does not start with a letter, character "a" is added in front of it.



2.14.3 Database SQL

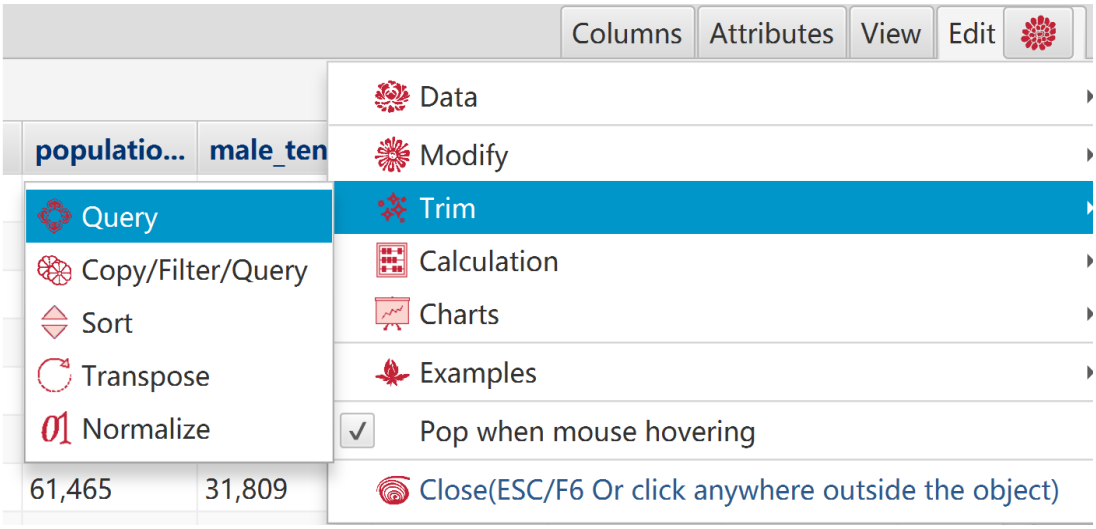
1. Provide examples of SQL statements.
2. List names of all user tables automatically.
3. View table definitions of all user tables.
4. Display outputs of execution and results of query.
5. SQL codes can be organized as information of tree.
6. Can load or save as external files.

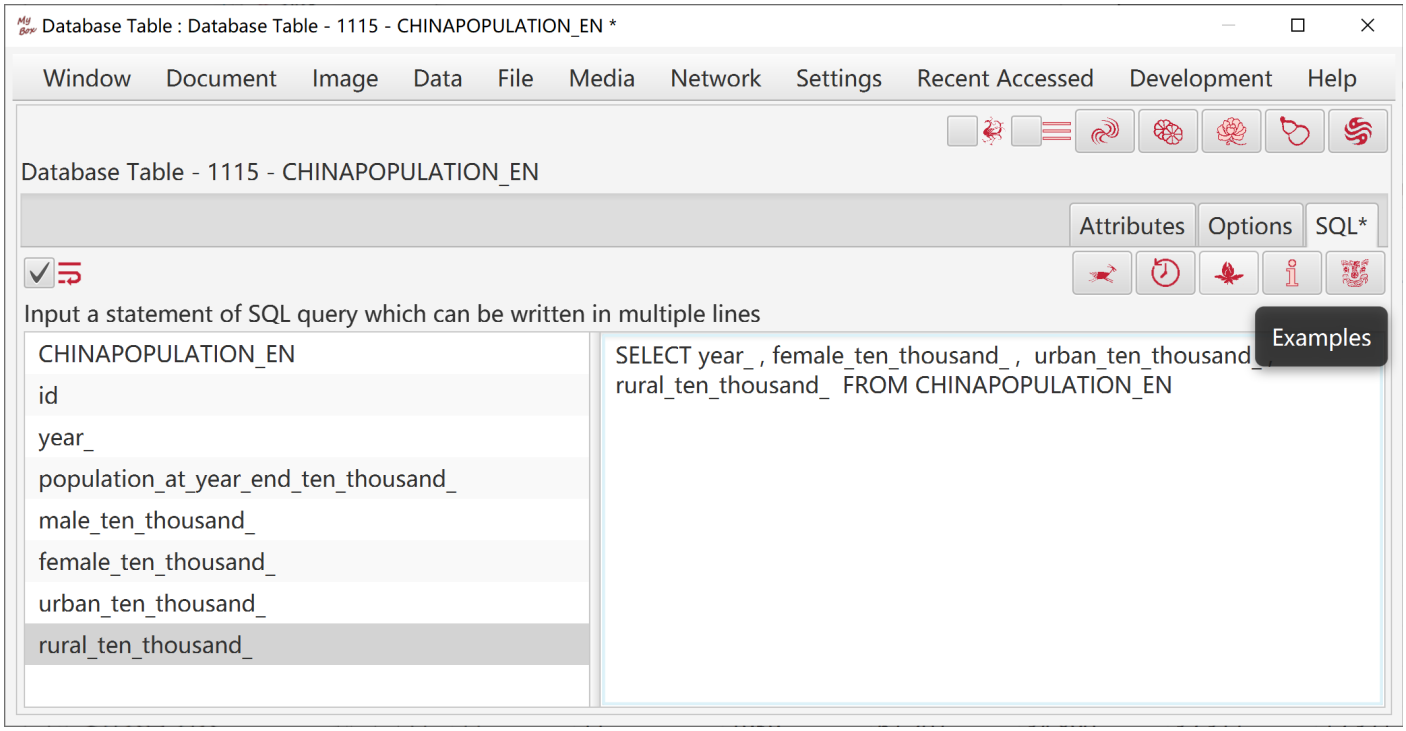


2.14.4 SQL Query

Database table has a special function menu: “Functions” - “Trim” - “Query”, which can help to input and execute SQL query:

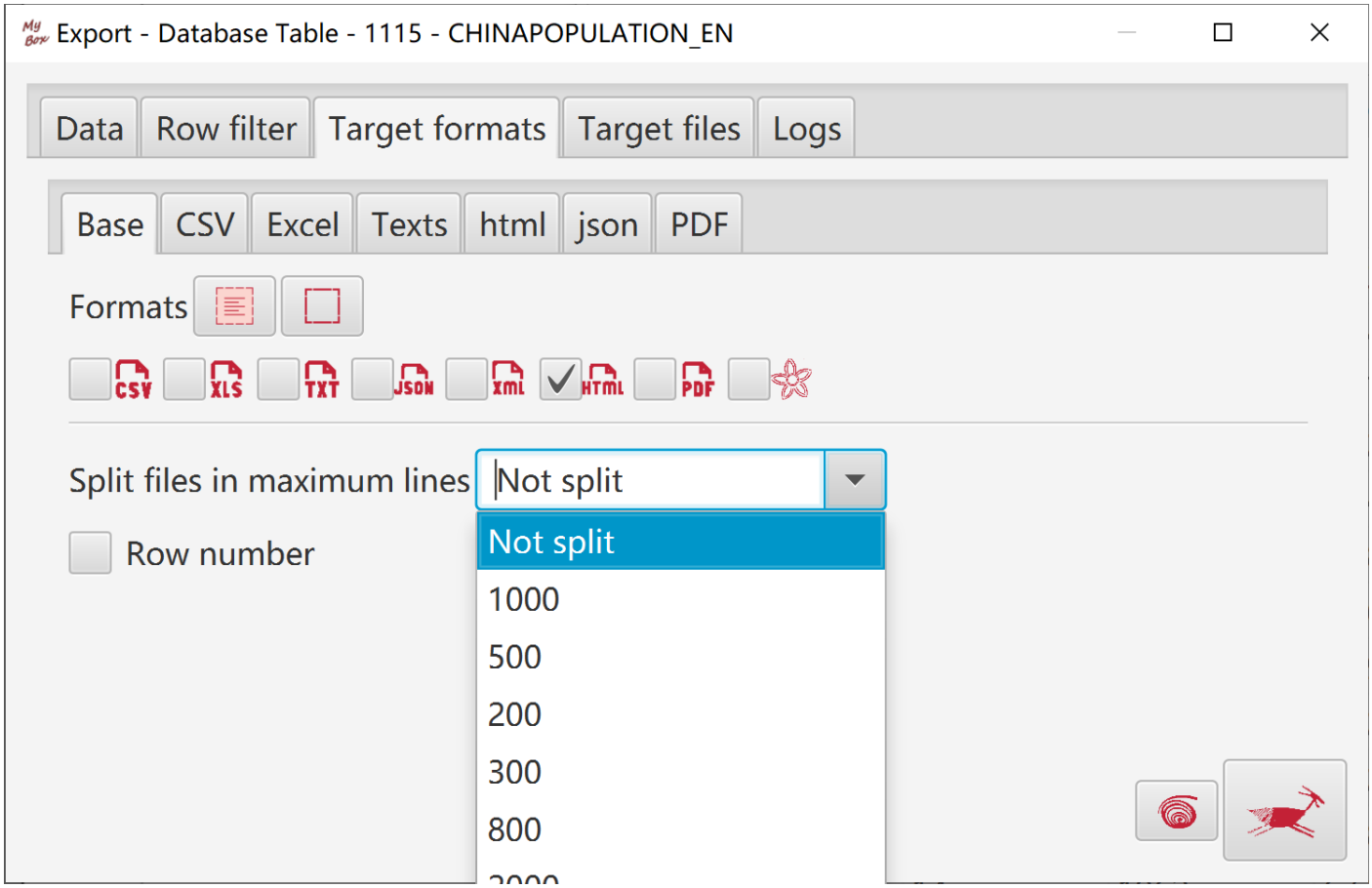
1. Names of table and columns are listed in left.
2. Provide examples and record histories.





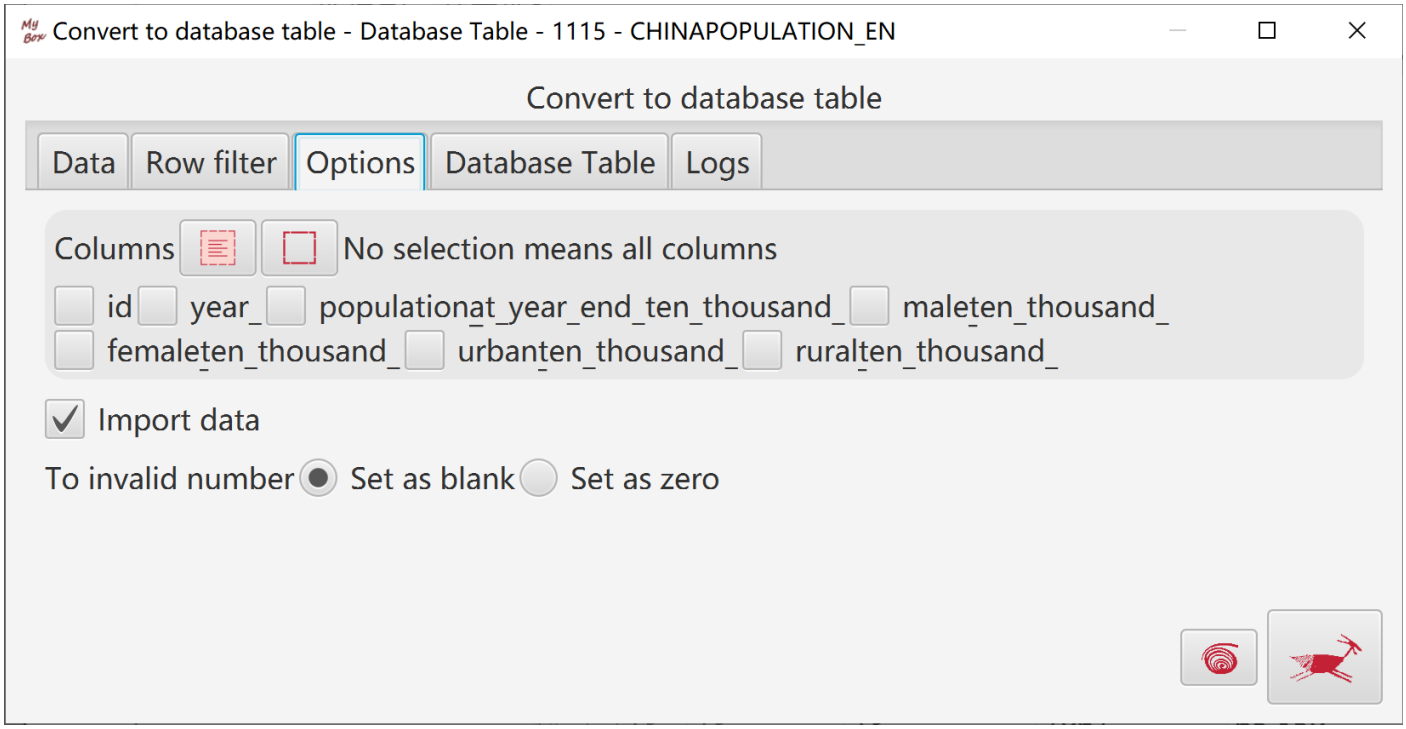
2.15 Export

1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Select columns. If no column is selected, then all columns are taken.
    - Set row filter.
2. Select targets and their formats: csv, texts, excel, xml, json, html, pdf, MyBox Clipboard.
3. Split files in maximum lines.



2.16 Convert to Database Table

1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Select columns. If no column is selected, then all columns are taken.
    - Set row filter.
2. Create auto-increment column, or select some columns as primary key.
3. Option to import data.



3 Calculation Tools

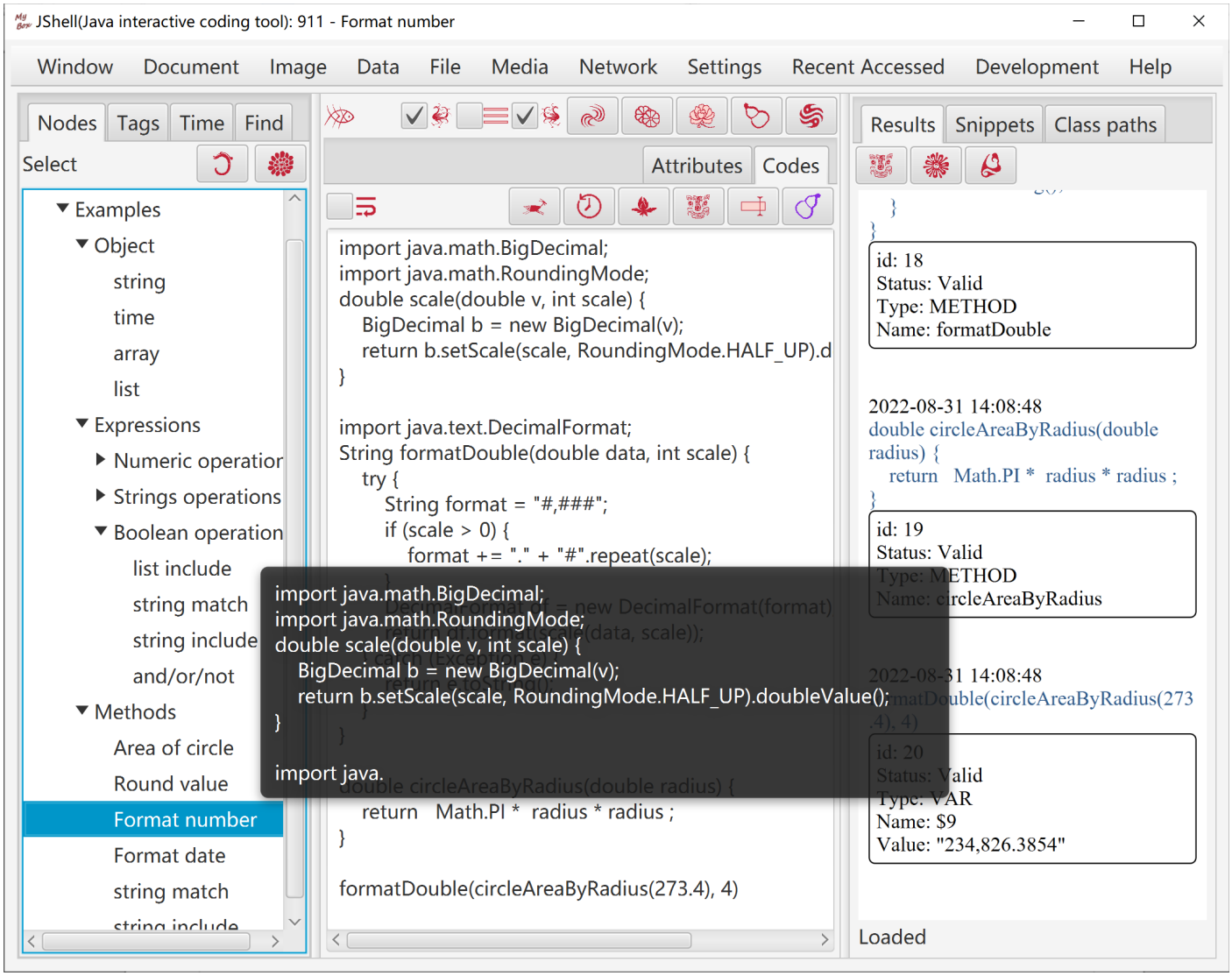
3.1 JShell(Java interactive coding tool)

JShell is one of tools in JDK:

1. JShell provides capability to interactively evaluate "snippets", as Read-Eval-Print Loop (REPL).
2. "Snippet" is a single expression, statement, or declaration of Java programming language code:
   - Semicolons should be in the end of statement while expression need not it.
   - Variables and methods can be defined and called later.
3. Extrenal Java classes should be accessible:
   - JShell picks "CLASSPATH" of system environment.
   - Other jar files or paths can be appended to "CLASSPATH".
   - Except for base classes, most of Java classes should be imported before call them.
4. JShell can be used for scientific computation and Java codes debug.

This tool helps to run JShell in GUI:

1. Input several snippets and click button "Start" to run them:
2. Snippets are evaluated one by one.
3. Results of snippets will affect later snippets, like "an execution environment".
4. Attributes of all evaluated snippets will be shown in a table.
5. Click button "Delete" or "Clear" to drop some or all snippets from current environment.
6. Click button "Reset" to empty JShell and environment becomes blank.
7. Press "CTRL+1" to pop list of code completion suggestions.
8. If added MyBox class paths, all methods of MyBox can be referred.
9. JShell codes are organized in tree. Examples are provided.


JShell(Java interactive coding tool): 911 - Format number
Window
Document
Image
Data
File
Media
Network
Settings
Recent Accessed
Development
Help
Nodes
Tags
Time
Find
Select
Examples
Object
string
time
array
list
Expressions
Numeric operation
Strings operations
Boolean operation
list include
string match
string include
and/or/not
Methods
Area of circle
Round value
Format number
Format date
string match
Attributes
Codes
import java.math.BigDecimal;
import java.math.RoundingMode;
double scale(double v, int scale) {
BigDecimal b = new BigDecimal(v);
return b.setScale(scale, RoundingMode.HALF_UP).d
}
import java.text.DecimalFormat;
String formatDouble(double data, int scale) {
try {
String format = "#,###";
if (scale > 0) {
format += "." + "#".repeat(scale);
}
return Math.PI * radius * radius ;
}
formatDouble(circleAreaByRadius(273.4), 4)
Results
Snippets
Class paths
id: 18
Status: Valid
Type: METHOD
Name: formatDouble
2022-08-31 14:08:48
double circleAreaByRadius(double radius) {
return Math.PI * radius * radius ;
}
id: 19
Status: Valid
Type: METHOD
Name: circleAreaByRadius
id: 20
Status: Valid
Type: VAR
Name: $9
Value: "234,826.3854"
Loaded

3.2 JEXL(Java Expression Language)

JEXL(Java Expression Language) is a library to generate values dynamically with variables and scripts.

1. JEXL has different syntax from Java. It is more like javascript.
2. Before run expression/script, all variables in it should have values held by JexlContext.
3. Refer to Java classes by creating their instances as local variables. Full package name is required.
4. JEXL can be used for scientific computation and data manufacture.

This tool helps to run JEXL in GUI:

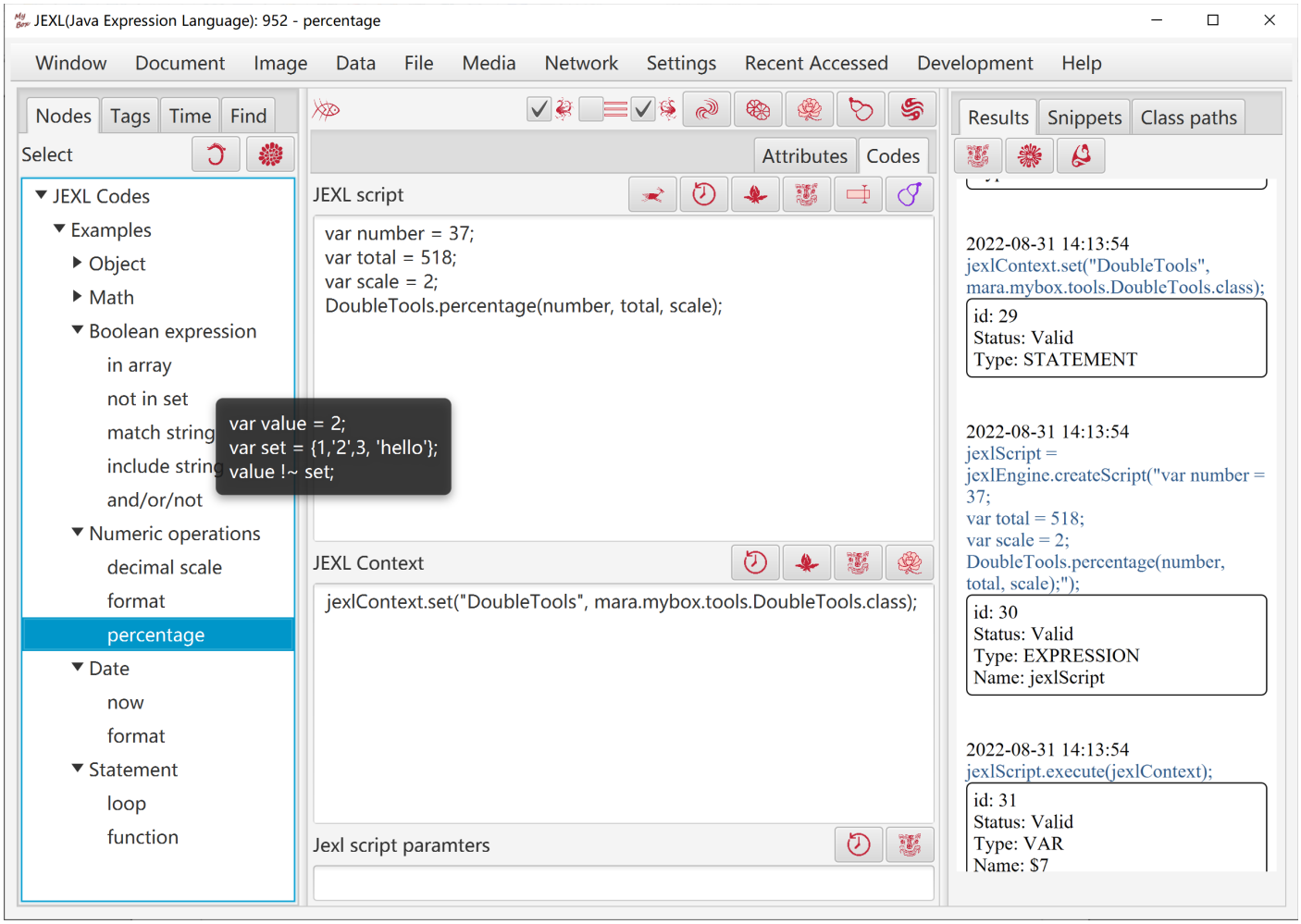
1. Input JEXL expression/script.

   Notice: Use single quotes instead of double quotes to surround strings.
2. Input Java codes of setting JexlContext like following:

   jexlContext.set("name", value);

   Example, set following to use Math.PI in expression/script:

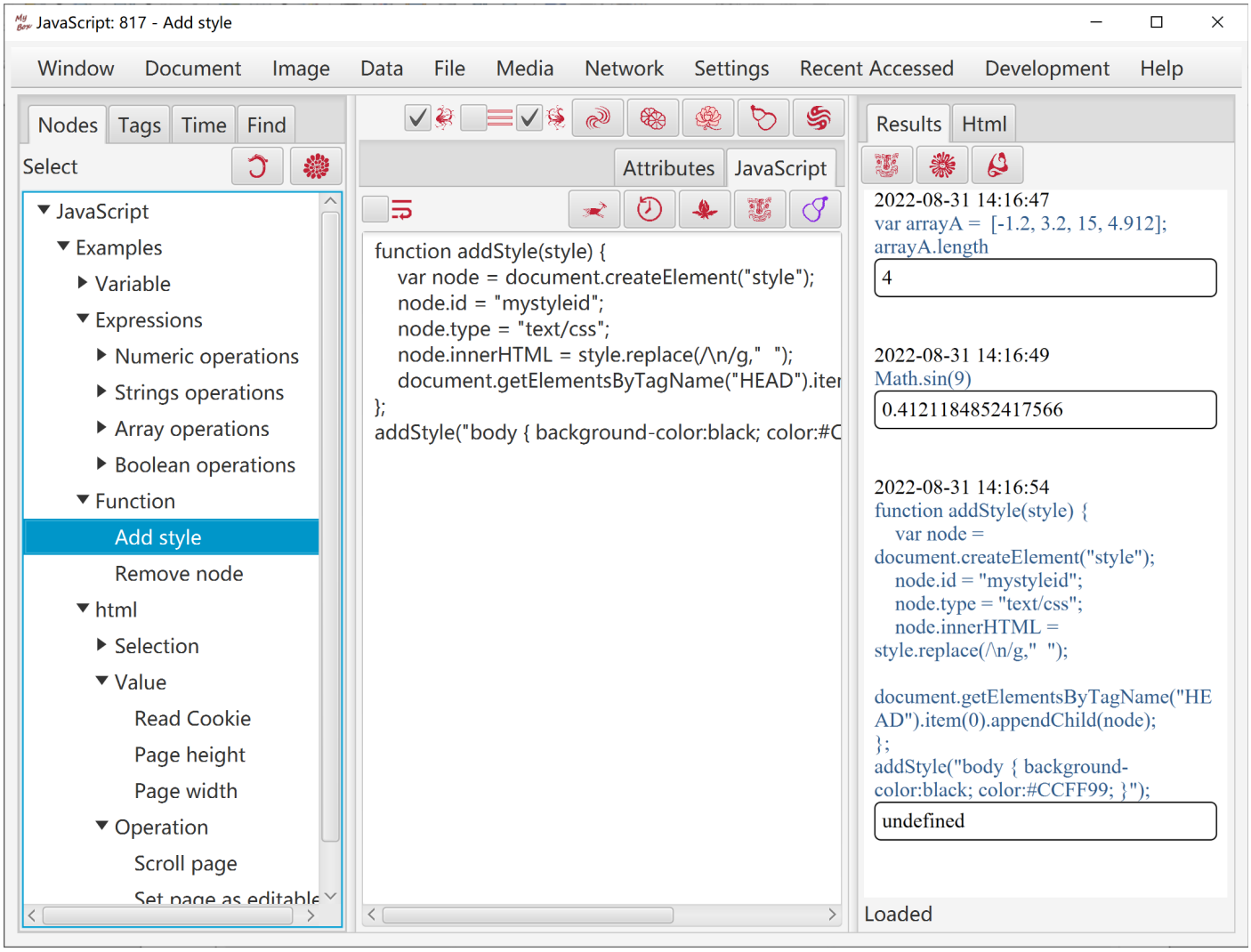
   jexlContext.set("Math", Math.class);
3. Input parameters of JEXL script if any. Separate values by comma.
4. Click button "Start" to evaluate the expression/script.
5. MyBox does following in JShell enrionment automatically:
   - Add MyBox library paths to CLASSPATH.
   - Import necessary JEXL packages.
   - Execute codes of JexlContext.
   - Calculate expression/script with parameters(if any).
6. If all variables and paramters have valid values, result is shown in right pane.
7. JEXL codes are organized in tree. Examples are provided.


JEXL(Java Expression Language): 952 - percentage
Window
Document
Image
Data
File
Media
Network
Settings
Recent Accessed
Development
Help
Nodes
Tags
Time
Find
Select
JEXL Codes
Examples
Object
Math
Boolean expression
in array
not in set
match string
include strin
and/or/not
Numeric operations
decimal scale
format
percentage
Date
now
format
Statement
loop
function
var value = 2;
var set = {1,'2',3, 'hello'};
value !~ set;
Attributes
Codes
JEXL script
var number = 37;
var total = 518;
var scale = 2;
DoubleTools.percentage(number, total, scale);
JEXL Context
jexlContext.set("DoubleTools", mara.mybox.tools.DoubleTools.class);
Jexl script paramters
Results
Snippets
Class paths
2022-08-31 14:13:54
jexlContext.set("DoubleTools", mara.mybox.tools.DoubleTools.class);
id: 29
Status: Valid
Type: STATEMENT
2022-08-31 14:13:54
jexlScript = jexlEngine.createScript("var number = 37;
var total = 518;
var scale = 2;
DoubleTools.percentage(number, total, scale);");
id: 30
Status: Valid
Type: EXPRESSION
Name: jexlScript
2022-08-31 14:13:54
jexlScript.execute(jexlContext);
id: 31
Status: Valid
Type: VAR
Name: $7

3.3 Javascript

This tool helps to edit, run, save Javascript codes.

Javascript codes are organized in tree. Examples are provided.



4 Data of Location

4.1 Data Constraints

4.1.1 Invalid Value

1. Null value of integer/long/short is the minimum value(MIN_VALUE)
2. Null value of double is the maximum value(Double.MAX_VALUE)

4.1.2 Coordinate System:

1. CGCS2000(China Geodetic Coordinate System), real locations and approximate to WGS-84(GPS).
2. GCJ-02(China encrypted coordinate), encrypted data with offsets of real locations.
3. WGS-84(GPS), real locations.
4. BD-09(Baidu encryted coordinate), based on GCJ-02.
5. Mapbar coordinate, based on GCJ-02.
6. When coordinate is unknown or invalid, the default value is CGCS2000.

4.1.3 Coordinate Values

1. Decimal values of longitude and latitude, instead of Degrees Minutes Seconds(DMS), are used when data handled.
2. MyBox provides "Location Tools" to convert coordinate values between decimal and DMS.
3. Valid range of longitude is `-180~180`, and valid range of latitude is `-90~90`.

4.1.4 Time

4.1.4.1 Date Formats

- Date and Time, like: 2014-06-11 13:51:33
- Date, like: 2014-06-11
- Year, like: 2014
- Month, like: 2014-06
- Time, like: 13:51:33
- Time with Milliseconds, like: 13:51:33.261
- Date and Time with Milliseconds, like: 2014-06-11 13:51:33.261
- Date and Time with zone, like: 2020-09-27 12:29:29 +0800
- Date and Time with Milliseconds and zone, like: 2020-09-27 12:29:29.713 +0800
- "T" can be written or omitted between date and time. "2014-06-11T13:51:33" equals to "2014-06-11 13:51:33".

4.1.4.2 Era

"0 AD" = "1 BC" = "0" = "-0" = "0000" = "-0000" = "0001-01-01 00:00:00 BC" = "公元前 1" = "公元前

0001-01-01 00:00:00"

"1 AD" = "1" = "0001" = "0001-01-01 00:00:00" = "0001-01-01 00:00:00 AD" = "公元 1" = "公元 0001-01-01 00:00:00"

"202 BC" = "-203" = "-0203" = "-0203-01-01 00:00:00" = "0202-01-01 00:00:00 BC" = "公元前 202" = "公元前 0202-01-01 00:00:00"

"202 AD" = "202" = "0202" = "0202-01-01 00:00:00" = "0202-01-01 00:00:00 AD" = "公元 202" = "公元 0202-01-01 00:00:00"

4.1.4.3 Examples

2020-07-13 11:30:59

-2020-07-13 11:30:59

-581-01-23

960

公元 960

公元前 770-12-11

公元前 1046-03-10 10:10:10

202 BC

960-01-23 AD

1046-03-10 10:10:10 BC

4.2 Data Operations

1. Add/Delete/Edit/Copy/Clear/Refresh data.
2. Query data:
   - Define and manage query conditions.
   - Current query conditions is displayed on tab "information".
   - Data satisfying current query condition are displayed in tab "Data" in pages.
   - Data rows can be displayed in different colors as values of some column.
3. Import data in csv format:
   - File encoding is UTF-8 or ASCII.
   - The first line defines data headers which are delimited by English commas.
   - Followed each line defines one data row. Data fields are delimited by English commas.
   - The order of fields is not cared.
   - Necessary fields must occupy their locations, but need not have valid values(related to data).
   - Select whether replace existed data. Predefined data or example data always replace existed values.
4. Export data:
   - Define and manage export conditions.

- Export data fields can be selected.
- Export file format can be selected: csv, xml, json, xlsx, html, pdf.
- Select maximum lines to split files.
- Can export current data page.

5. Delete/Clear data:
   - Define and manage delete conditions.
   - Predefined data can not be deleted.
   - Referred data(like foreign keys) can not be deleted.
6. Define, manage, and use "Conditions":
   - "Conditions" are used to execute querying, deleting, or exporting.
   - Set conditions in panes:
     - Data conditions are organized as trees. Multiple nodes can be selected.
     - Multiple data fields can be selected as sorting conditions, and their orders can be changed.
   - Edit condition: Title, where, order by, fetch. They will be merged as final conditon.
   - Manage conditons: Add/delete/edit/copy.
   - Conditions ever executed are saved automatically.
   - Recently visited conditions are listed in pop window of the buttons.

4.3 Map Data

1. Kinds of data can be presented in map, including Geography Codes, Location Data, and Coordinate Querying.
2. Data in map can be:
   - All data which satisfy current query condition. "Maximum number of data" can be set to avoid performance issues.
   - Data in current page.
3. TianDiTu:
   - Accepts coordinates of CGCS2000 and display them at correct locations without offsets.
   - When display other coordinates, MyBox converts them to CGCS2000 to show correct locations.
   - Projection can be selected: EPSG:900913/3857(Web Mercator) or EPSG:4326(Geodetic).
   - Controls can selected: Zoom, Scale, Map Type, Symbols.
   - Map Types: Standard, Satellite, Mixed Satellite, Terrain, Mixed Terrain.
   - Languages in different regions.
   - Range of map levels is 1-18.
4. GaoDe Map:
   - Accepts coordinates of GCJ-02 and display them at correct locations without offsets.

- When display other coordinates, MyBox converts them to GCJ-02 to show correct locations.
- Projection is EPSG:900913/3857(Web Mercator).
- Map layers:
  - Can select multiples: standard, satellite, roadnet, traffic.
  - Roadnet layer and traffic layer are only supported for China.
  - Satellite layer is supported for part of foreign addresses.
  - Opacity can be set for each map layer.
- Map language: Chinese, English, Chinese and English.
- Range of map levels is 3-18
- Can selected "Fit View" to adjust map level and center as best automatically while display all data.

5. Adjust map level by:
   - Scroll mouse wheel.
   - Click map controls.
   - Select "Map Size"
6. Marker image:
   - Selections: point(bubble), circle, or any image.
   - For Location Data, more selections: Data Set Image, Data Image. Point will be used if no valid value.
   - Size can be set(Same size for width and height)
7. Marker text：
   - Selections: Label, Coordinate, Address.
   - For Location Data, more selections: Start Time, End Time, Data Value, etc.
   - Multiples selections can be picked. Each selection will be showns in a line.
   - Size can be set.
   - Can select whether text is bold.
   - Color can be set. For Location Data, "Data Color" can be chosen.
8. Pop information:
   - Detailed information can be popped when mouse is upon marker.
   - Can select whether pop information.
9. Snapshot:
   - DPI can be set.
   - Current map and data in map can be saved and displayed in html.
10. Keys of map can be changed in "Settings". The default keys are free and shared by all MyBox users.

4.4 Geography Code

4.4.1 Data Definition

1. Basical attributes: id, level, longitude, latitude, chinese_name, english_name, 5 codes, 5 aliases,
2. Subordinate: owner, continent, country, province, city, county, town, village, building. ("Ancestors")
3. Auxiliary attributes: altitute, precision, coordinate system, area(square meters), population, comments, isPredefined.

4.4.2 Data Constraints

1. Not null values: id, level, chinese_name or english_name.
2. Values of "level": global(only "Earth"), continent, country, province(state), city, county(district), town, village(neighborhood), building, point of interest.
3. Data is unnecessary to be subordinated level by level. Cross-over can happen. Example, a village is subordinated to Antarctica, and a city belongs to a country without province level.
4. Match data:
    - One of following can determine an address:
        - Match “id"(assigned by MyBox automatically). This is accurate matching.
        - Match "level" + ancestors + "chinese_name"/"english_name"/any one "alias". This is accurate matching.
        - Match "level" + "chinese_name"/"english_name"/any one "alias". This is fuzzy matching. Duplicated names in same level can cause false matching.
    - Matching of name or alias is case-insensitive.
    - Sometimes 5 "code" are useful to match data.

4.4.3 Edit Data

1. "subordinate" of data is set by selecting node in locations tree.
2. "level" of data should be lower than its ancestors.
3. Data must have either chinese_name or english_name.
4. Select or display coordinate in map.
5. Set as "Predefined data" or "Inputted data" against selected rows.

4.4.4 Define Condition

All geograhy codes in MyBox are organized as a Locations Tree by their subordination relationship. Multiple nodes can be selected.

4.4.5 Import Data

4.4.5.1 Predefined Data

Include continents, countries, Chinese provinces /cities /counties.

Countries have values of "area" and "population".

4.4.5.2 CSV Format

- Download address:

  https://github.com/Mararsh/MyBox_data/tree/master/md/GeographyCode/en

- Necessary fields:

  Level,Longitude,Latitude

  And "Chinese Name" or "English Name"

- Optional fields:

  Altitude,Precision,Coordinate System,Square Kilometers,Population,

  Code 1,Code 2,Code 3,Code 4,Code 5,Alias 1,Alias 2,Alias 3,Alias 4,Alias 5,

  Continent,Country,Province,City,County,Town,Village,Building,Comments

4.4.5.3 Data from geoname.org:

- Download address:

  http://download.geonames.org/export/zip/

- Tab-delimited text in UTF8 encoding.
- Data fields：

  countryCode postalCode placeName

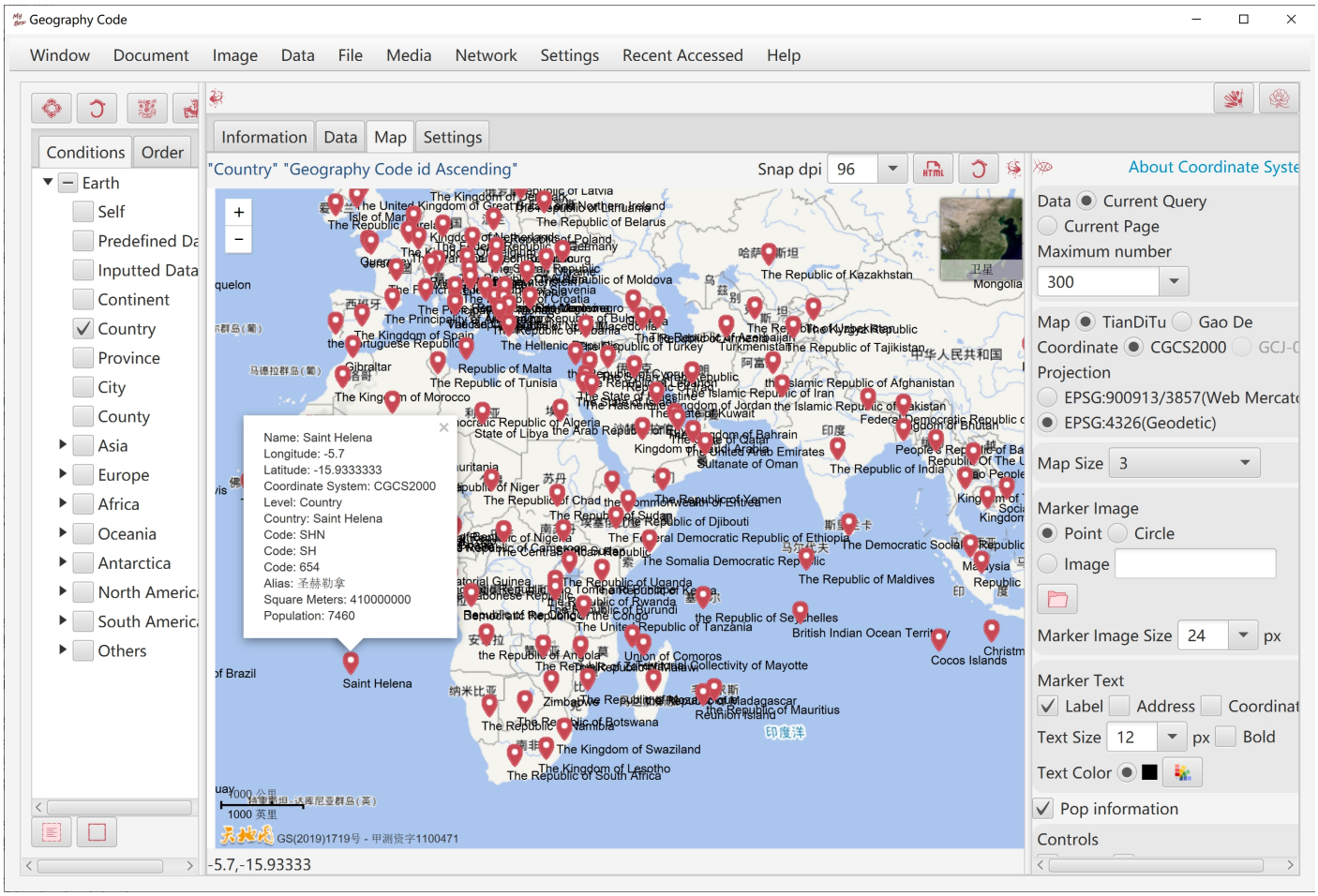
  adminName1 adminCode1 adminName2 adminCode2 adminName3 adminCode3

  latitude longitude accuracy

- Coordinate system is WGS_84.
- Same address is written only once even when it has multiple "postal code" or coordinates.

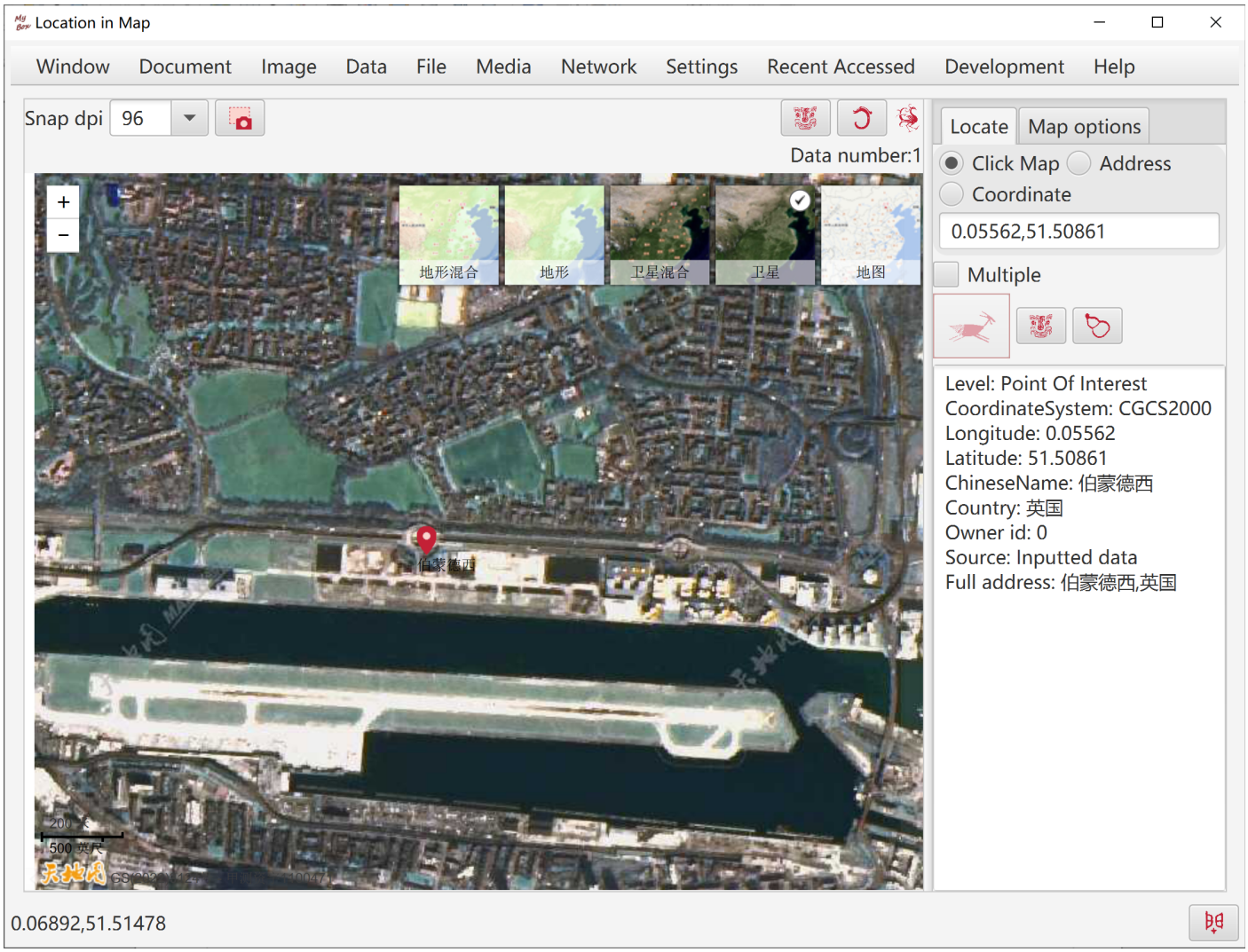
4.4.6 Settings

1. Customize colors of data rows.
2. Provide "Default" and "Random" buttons.


Geography Code
Window
Document
Image
Data
File
Media
Network
Settings
Recent Accessed
Help
Conditions
Order
Earth
Self
Predefined Da
Inputted Data
Continent
Country
Province
City
County
Asia
Europe
Africa
Oceania
Antarctica
North Americ
South Americ
Others
Information
Data
Map
Settings
"Country" "Geography Code id Ascending"
Snap dpi
96
Name: Saint Helena
Longitude: -5.7
Latitude: -15.9333333
Coordinate System: CGCS2000
Level: Country
Country: Saint Helena
Code: SHN
Code: SH
Code: 654
Alias: 圣赫勒拿
Square Meters: 410000000
Population: 7460
Saint Helena
-5.7,-15.93333
About Coordinate Syste
Data
Current Query
Current Page
Maximum number
300
Map
TianDiTu
Gao De
Coordinate
CGCS2000
Projection
EPSG:900913/3857(Web Mercat
EPSG:4326(Geodetic)
Map Size
3
Marker Image
Point
Circle
Image
Marker Image Size
24
px
Marker Text
Label
Address
Coordinat
Text Size
12
px
Bold
Text Color
Pop information
Controls

4.5 Location in Map

1. Query geography code by:
    - Click map.
    - Input address.
        - TianDiTu supports chinese and foreign addresses in Chinese(like "伦敦") or in English(like "Paris")
        - GaoDe map only supports addresses in China.
    - Input longitude and latitude.
2. Query result can be saved in Geography Code table.



4.6 Location Data

4.6.1 Data Definition

1. Basical attributes: data set, label, longitude, latitude, start time, end time.
2. Auxiliary attributes: altitute, precision, coordinate system, speed, direction, data value, data size, image, comments.

4.6.2 Data constraints

1. Each location data belongs to a data set.
2. Data set defines common attributes of some location data, examples:
   - Date format
   - Whether omit "AD" for date AD
   - Text color
   - Image

   These attributes help to distinguish data points in map.

4.6.3 Define Conditions

1. List of data sets. Multiple nodes can be selected.
2. Time tree(Start time). Multiple nodes can be selected.

4.6.4 Map Data

1. At beginning, the first data is made as map center.
2. Location Distribution: All data are displayed in map.
3. Time Sequence:
   - Data are displayed in frames as "Start Time" in ascending order.
   - When "Accumulated" is selected, drawn points will not be erased and points are shown more and more.
   - When "Time Overlay" is selected, all data whose duration(between "start time" and "end time") has intersection with duration of current frame will be treated as valid points of current frame.
     Example, current frame has "start time" of "1044 BC" and "end time" of "221 BC", then all data who or part of who appears in this duration will be displayed in current frame.
   - When "Move Center" is selected, each frame will adjust its map center.
   - When "Link" is selected, lines between adjacent 2 points will be shown.
   - Control frames:
     - Set interval.
     - Select a frame(by start time).
     - Pause/Continue playing.
     - Previous/Next frame.
     - Whether loop.

4.6.5 Snapshots

- For "Location Distribution":
    - html:Data and snapshot of current frame
    - Snapshot of current frame. All supported image formats can be selected.
- For "Time Sequence", more choices:
    - jpg:Snapshots of all frames
    - png:Snapshots of all frames
    - Animated gif:Snapshots of all frames(May out of memory)

4.6.6 Import Data

If data include a dataset which is not in database, the new dataset will be added in database automatically.

4.6.6.1 CSV Format

- Necessary fields: Dataset,Longitude,Latitude
- Optional fields: Label,Address,Altitude,Precision,Speed,Direction,Coordinate System, Data Value,Data Size,Start Time,End Time,Image,Comments

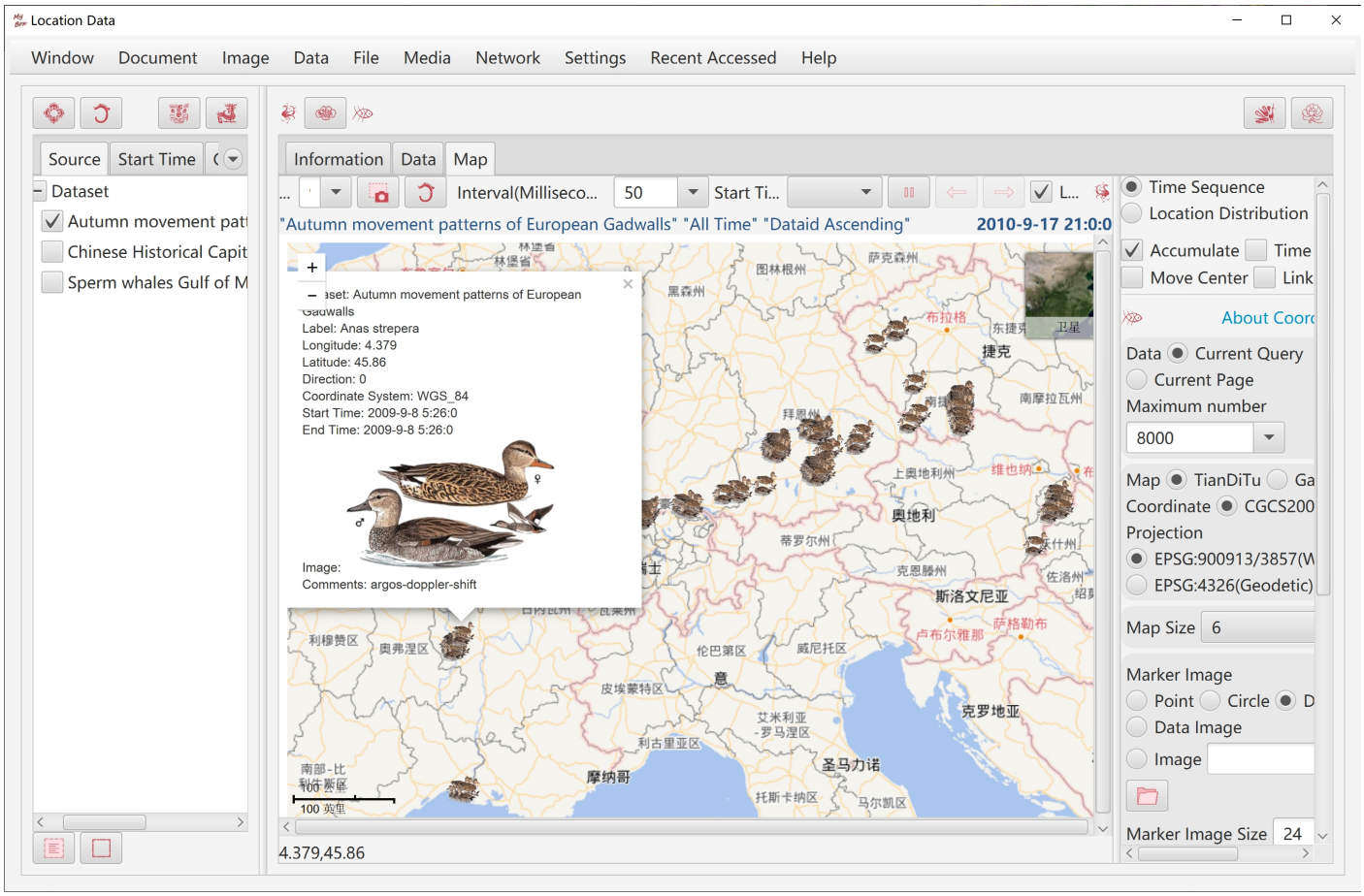
4.6.6.2 Data from movebank.org

- Download address:

  https://www.datarepository.movebank.org/
- Comma-delimited CSV file.
- Necessary fields: timestamp,location-long,location-lat,study-name
- Coordinate system is WGS_84.

4.6.6.3 Examples

1. Chinese Historical Capitals
2. Autumn movement patterns of European Gadwalls
3. Sperm whales Gulf of Mexico


Location Data
Window
Document
Image
Data
File
Media
Network
Settings
Recent Accessed
Help
Source
Start Time
Dataset
Autumn movement pat
Chinese Historical Capit
Sperm whales Gulf of M
Information
Data
Map
Interval(Milliseco...
50
Start Ti...
"Autumn movement patterns of European Gadwalls" "All Time" "Dataid Ascending"
2010-9-17 21:0:0
Label: Anas strepera
Longitude: 4.379
Latitude: 45.86
Direction: 0
Coordinate System: WGS_84
Start Time: 2009-9-8 5:26:0
End Time: 2009-9-8 5:26:0
Image:
Comments: argos-doppler-shift
4.379,45.86
Time Sequence
Location Distribution
Accumulate
Time
Move Center
Link
About Coor
Data
Current Query
Current Page
Maximum number
8000
Map
TianDiTu
Coordinate
CGCS200
Projection
EPSG:900913/3857(W
EPSG:4326(Geodetic)
Map Size
6
Marker Image
Point
Circle
Data Image
Image
Marker Image Size
24

4.7 Location Tools

1. Convert coordinate value between decimal and DMS. Valid examples of DMS:

   48°51'12.28"

   -77° 3' 43.9308"

   48°51'12.28"N

   2°20'55.68"E

   S 34° 36' 13.4028"

   W 58° 22' 53.7348"

   118 度 48 分 54.152 秒

   -32 度 04 分 10.461 秒

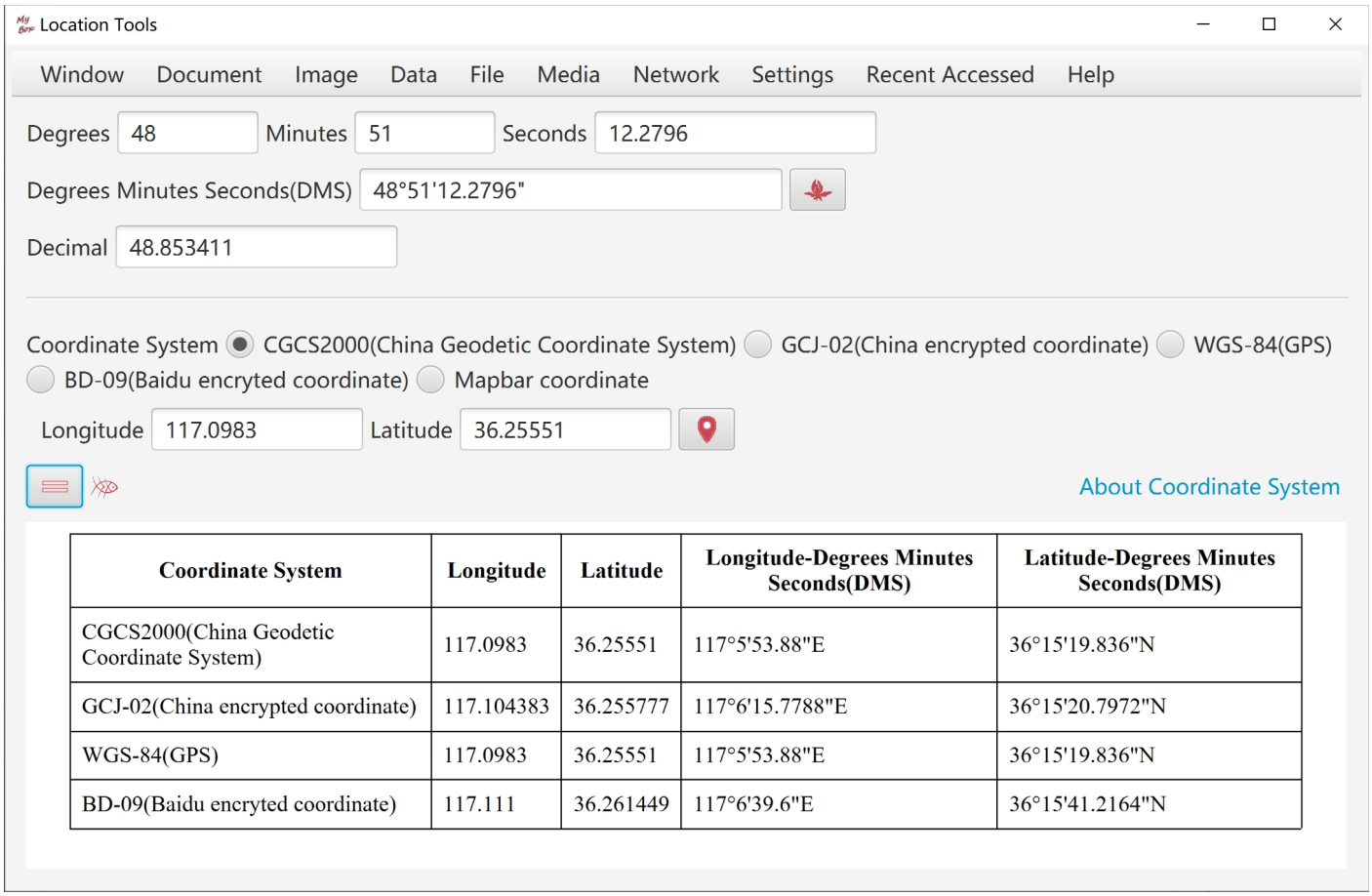
   东经 118 度 48 分 54.152 秒

   北纬 32 度 04 分 10.461 秒

   西经 118 度 48 分 54.152 秒

   南纬 32 度 04 分 10.461 秒

2. Convert coordinate values as other coordinate systems.



Coordinate System	Longitude	Latitude	Longitude-Degrees Minutes Seconds(DMS)	Latitude-Degrees Minutes Seconds(DMS)
CGCS2000(China Geodetic Coordinate System)	117.0983	36.25551	117°5'53.88"E	36°15'19.836"N
GCJ-02(China encrypted coordinate)	117.104383	36.255777	117°6'15.7788"E	36°15'20.7972"N
WGS-84(GPS)	117.0983	36.25551	117°5'53.88"E	36°15'19.836"N
BD-09(Baidu encryted coordinate)	117.111	36.261449	117°6'39.6"E	36°15'41.2164"N

4.8 Epidemic Reports

4.8.1 Data Definition

1. Basical attributes: dataSet, time, location, source.
2. Basical values: confirmed, healed, dead.
3. Subtraction statistic: increased confirmed, increased healed, increased dead.

   Calculated by adjacent rows.
4. Division statistics:
   - healed/confirmed permillage, dead/confirmed permillage
   - confirmed/population permillage, healed/population permillage, dead/population permillage,
   - confirmed/area permillage, healed/area permillage, dead/area permillage.

     When value of "area"/"population" of location is invalid(zero or negitive), corresponding statistics data are meaningless.

     Predefined data "countries" have valid "area"/"population" and they have meaningful statistics values.
5. Accumulation statistics:
   - Values of some countries. Calculated by values of country's provinces.
   - Values of continents. Calculated by values of continent's countries.
   - Values of Earth. Calculated by values of continents.

4.8.2 Data Constraints

1. Not null values: dataSet, time, location
2. Values of "source": "Inputted data", "Predefined data", "Filled data", "Statistics data".
3. "location" is foreign key of "Geography Code", which must have row defined in that table.
4. In "confirm", "healed", "dead", at least one should be larger than zero.
5. One of following can determine a data row:
   - Match id, which is assigned by MyBox automatically. This is accurate matching.
   - Match "dataSet" + "date" + "location". This is accurate matching.
   - This version assumes that only one valid data in each day for same dataSet plus same location.

4.8.3 Edit Data

1. When in single data, location is set by selecting node from locations tree.
2. In interface of "Epidemic Reports of Chinese Provinces" or "Epidemic Reports of Countries", multiple rows can be inputted for same dataSet and time.
3. Modify values of "source" for selected data rows.

4.8.4 Import Data

4.8.4.1 Predefined Data

COVID-19 historical data from Johns Hopkins University.(Till 2020-09-24)

4.8.4.2 CSV format:

- Download address:

  https://github.com/Mararsh/MyBox_data/tree/master/md/COVID19/en

- Necessary fields: Data Set,Time,Confirmed,Healed,Dead,

  And location data which are enough to define a geography code: Longitude,Latitude,Level,Continent,Country,Province,City,County,Town,Village,Building,Point of Interest

- Optional fields: Increased Confirmed,Increased Healed,Increased Dead
- Coordinate system is CGCS2000.

4.8.4.3 COVID-19 historical data from Johns Hopkins University(Global) :

- Download address:

  https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

- Necessary fields: Province/State,Country/Region,Lat,Long.

  And date list like "1/22/20,1/23/20..."

- Coordinate system is WGS_84.
- Australia, Canada and China are reported at the province/state level, and others are at country level.
- Items whose values are all zero will be skipped.

4.8.4.4 COVID-19 daily data from Johns Hopkins University(Global) :

- Download address:

  https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports

- Data fields are change as time flowing...

  Following is format of "01-22-2020.csv":

  Province/State,Country/Region,Last Update,Confirmed,Deaths,Recovered

  Following is format of "05-15-2020.csv":

  FIPS,Admin2, Province_State,Country_Region, Last_Update,Lat,Long_,

  Confirmed, Deaths, Recovered,Active, Combined_Key

- Coordinate system is WGS_84.
- Items whose values are all zero will be skipped.

4.8.4.5 Handle Data of Imported

1. Option: Statistics against dataset.

2. Time of all data are changed as “23:59:00”.
3. If data include a geography code which is not in database, the new geography code will be added in database automatically.

4.8.5 Statistics Data

1. Option to accumulate date
2. Option to calculate subtraction statistic for different location levels.

4.8.6 Define Conditions

1. Data Sources Tree: Data sets and their different sources are organized a tree. Multiple nodes can be selected.
2. Locations Tree: All geograhy codes in MyBox are organized as a tree by their subordination relationship. Multiple nodes can be selected.
3. Times Tree: All times involved in Epidemic Reports of MyBox are organized as a tree. Multiple nodes can be selected.
4. Number of Top Data in Each Day:
   - Unlimit. Charts will not be displayed. Data are queried as condition.
   - Valid value:
     - Data are queried as condition, and then be truncated as top data of each day, by which charts and data are displayed.
     - "Time Descending" is always as the first ordering element automactically.
     - At least one more column should be picked as ordering element.
     - Beside "Time Descending", the first ordering element is called "major quering attribute".
   - Number of Top Data in Each Day" and "Elements of ordering" work for Query and Export, and not for Clear.
5. Edit condition: title, where, order by, fetch, "Number of Top Data in Each Day"(0 or -1 means Unlimit), which are combined together as the final condition.

4.8.7 Display Charts

1. Only when query condition satisfies requirements, charts are displayed. Charts' data are always "Daily top data" and have "Major query attribute".
2. Beside "Major query attribute" , more attribues can be selected, to display multiple dimension data in same chart, or show multiple charts at same time.
3. Chart type: horizontal bars, vertical bars, horizontal lines, vertical lines, pie, map.
4. When there are multiple times in data，charts are animated. Data charts of each time are displayed frame by frame in time ascending.
5. For animated charts, support Pause/Continue, Jump to frame of a time, Last frame, Next frame, setting interval.
6. Common settings, which take effect immediately：
   - Legend location: not display, top, bottom, left, right.

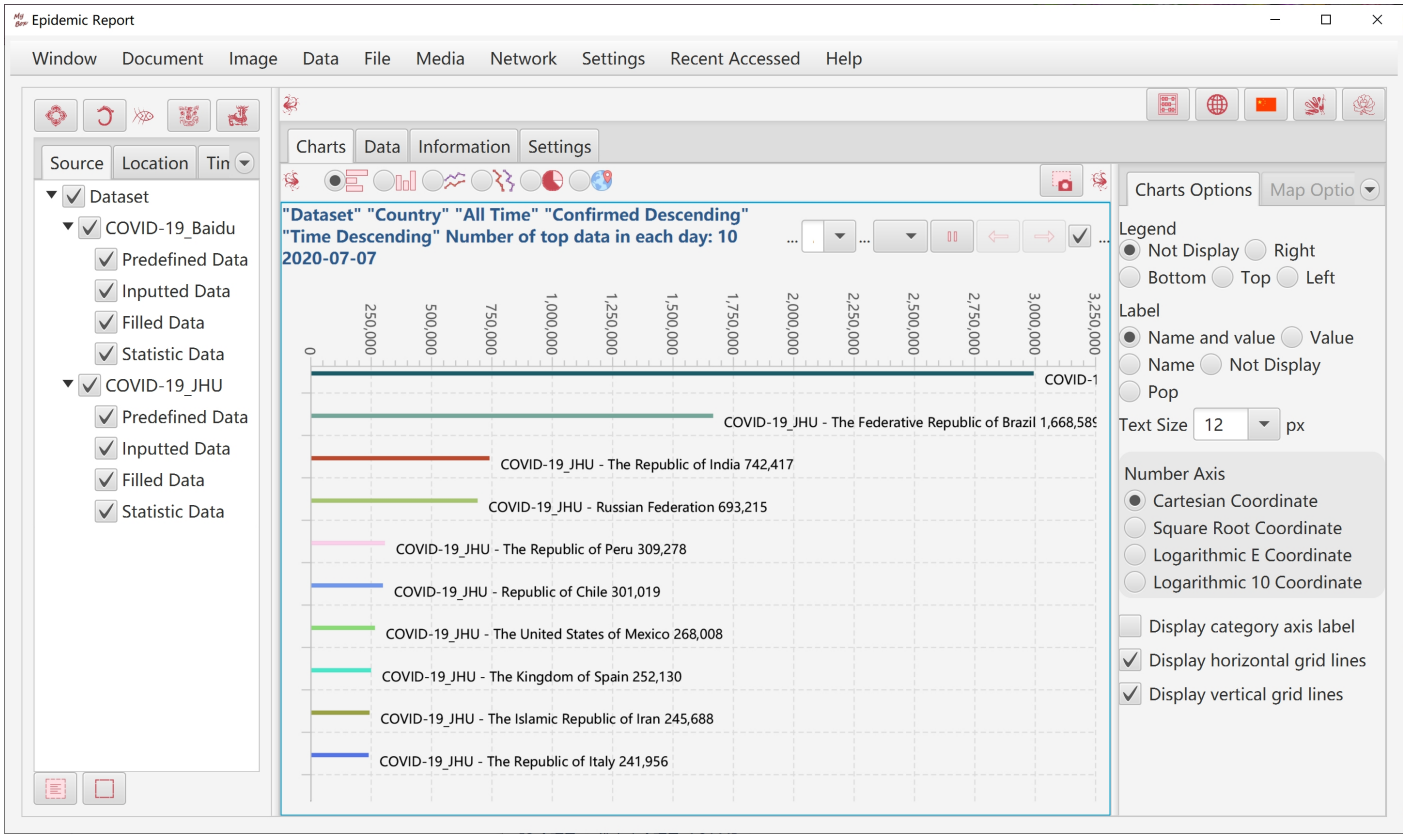
- Value's label: name and value, value, name, not display, pop
- Whether display: category axis, horizontal grid lines, vertical grid lines
- Number axis: Cartesian Cordinate, Square Root Cordinate, Logarithmic E Cordinate, Logarithmic 10 Cordinate.
- Font size
- Parameters of map: level, layers, language

7. Snap chart.
   - Snapshot of current frame. All supported image formats can be selected.
   - jpg:Snapshots of all frames
   - png:Snapshots of all frames
   - Animated gif:Snapshots of all frames(May out of memory)

4.8.8 Settings

1. Snap dpi, maximum width of snapped animated images, time to loading chart's data.

   These parameters are related to memory usage and computer's calculation capacity.
2. Customize colors of data rows as column "source". Provide "Default" and "Random" buttons.
3. Customize colors of data values in charts. Provide "Default" and "Random" buttons.
4. Customize colors of location values in charts. Provide "Random" buttons.



5 Others

5.1 Create Barcodes

1. Supported 1-d barcodes:
    - Types: Code39, Code128, Codabar, Interleaved2Of5, ITF_14, POSTNET, EAN13, EAN8, EAN_128, UPCA, UPCE, Royal_Mail_Customer_Barcode, USPS_Intelligent_Mail
    - Options about 1-d barcodes: Orientation, width/height, dpi, text location, font size, quiet-zone width, etc.
2. Supported 2-d barcodes:
    - Types: QR_Code, PDF_417, DataMatrix
    - Options about 2-d barcodes: Width/height, margin, error correction level, compact mode, etc.
    - A picture can be shown in center of QR_Code. Its size can be adjusted automatically according to error correction level.
3. Examples of parameters and suggested values.
4. Validate generated barcode at once.

5.2 Decode Barcodes

1. Supported 1-d barcodes: Code39, Code128, Interleaved2Of5, ITF_14, EAN13, EAN8, EAN_128, UPCA, UPCE
2. Supported 2-d barcodes: QR_Code, PDF_417, DataMatrix
3. Display contents of barcodes and its meta data including barcode type and error correction level if any.

5.3 Message Digest

1. Create digest for files or inputted texts.
2. Support MD2, MD5, SHA-1, SHA-224, SHA-256, SHA-384, SHA-512/224, SHA-512/256, SHA3-224, SHA3-256, SHA3-384, SHA3-512.
3. Ouput: Base64, Hexadecimal, Formatted hexadecimal.

5.4 Encode/Decode Base64

1. Encode file or texts as Base64.
2. Decode Base64 file or Base64 texts.
3. Set charset for texts.
4. Output as file or texts.

5.5 Extract ttf files from ttc file

<End of Document>