



# **MyBox: Easy Tools Set User Guide – Data Tools**

Author: Mara

Version: 6.6

Date: 2022-9-28

## Contents

1 Introduction.....	6
1.1 Main Interface.....	6
1.2 Resources Addresses.....	7
1.3 Documents.....	8
1.4 Menu of Tools.....	8
2 Data in Two-dimensional Storage Structure.....	9
2.1 Data Objects.....	9
2.2 Functions Menu.....	10
2.3 Edit Data.....	11
2.3.1 Save and Recover.....	11
2.3.2 Define Columns .....	12
2.3.3 Define Data Attributes.....	14
2.3.4 Edit Mode "Table".....	15
2.3.5 Edit Mode "Text".....	16
2.3.6 View Html.....	17
2.3.7 View Text.....	17
2.3.8 Load Contents in System Clipboard .....	18
2.3.9 Import Examples.....	19
2.3.9.1 Statistic Data of China.....	19
2.3.9.2 Data of Regression .....	20
2.4 Row Expression.....	21
2.5 Row Filter.....	23
2.6 Modify.....	24
2.6.1 Set Values .....	25
2.6.2 Delete .....	26
2.6.3 Set styles / Mark Abnormal Values.....	27
2.6.3.1 Manage Styles.....	27
2.6.3.2 Data Scope.....	28
2.6.3.3 Define Styles.....	29
2.6.3.4 Apply Styles.....	29
2.6.4 Paste Content in System Clipboard.....	30
2.6.5 Paste Content in MyBox Clipboard.....	30
2.7 Trim.....	31
2.7.1 Copy/Filter/Query.....	32
2.7.2 Sort .....	33
2.7.3 Transpose .....	34
2.7.4 Normalization .....	35
2.7.5 Split.....	36
2.8 Calculation.....	37
2.8.1 Row Expression.....	38
2.8.2 Descriptive Statistic .....	39
2.8.3 Group Equal Values.....	40
2.8.4 Simple Linear Regression .....	43

2.8.4.1 Regression .....	43
2.8.4.2 Model.....	44
2.8.4.3 Fitted Chart.....	45
2.8.4.4 Residual Chart.....	46
2.8.5 Simple Linear Regression - Combination .....	47
2.8.6 Multiple Linear Regression.....	48
2.8.6.1 Regression.....	48
2.8.6.2 Model.....	49
2.8.7 Multiple Linear Regression - Combination.....	50
2.8.8 Frequency Distributions .....	51
2.8.9 Values Percentage .....	52
2.9 Chart.....	53
2.9.1 XY Chart.....	54
2.9.1.1 Data.....	54
2.9.1.2 Bar Chart.....	54
2.9.1.3 Stacked Bar Chart.....	55
2.9.1.4 Line Chart.....	55
2.9.1.5 Scatter Chart.....	56
2.9.1.6 Bubble Chart.....	56
2.9.1.7 Area Chart.....	57
2.9.1.8 Stacked Area Chart.....	58
2.9.1.9 Parameters of Data in Chart.....	59
2.9.1.10 Layout.....	60
2.9.1.11 Category Axis.....	61
2.9.1.12 Number Axis.....	62
2.9.2 Pie Chart.....	63
2.9.3 Box-and-whisker Chart .....	64
2.9.4 Self Comparison Bars Chart .....	66
2.9.5 Comparison Bars Chart .....	68
2.9.6 XYZ Chart.....	70
2.9.6.1 3-D Scatter.....	70
2.9.6.2 Surface Chart.....	72
2.10 Outputs.....	74
2.10.1 Export.....	75
2.10.2 Convert to Database Table.....	76
2.11 Manage Data.....	77
2.12 Splice Data.....	78
2.13 Data File.....	79
2.13.1 CSV File.....	79
2.13.2 Text File.....	80
2.13.3 Excel File.....	81
2.13.4 Convert/Split Data Files in Batch.....	82
2.13.5 Merge Data Files .....	83
2.14 Data in System Clipboard.....	84
2.15 Data in MyBox Clipboard.....	85
2.16 Matrix.....	86

2.16.1 Edit and Manage Matrices.....	86
2.16.2 Unary Matrix Calculation.....	87
2.16.3 Binary Matrices Calculation.....	88
2.17 Database Tables.....	89
2.17.1 Manage Database Tables.....	89
2.17.2 Limitations of SQL Identifier.....	90
2.17.3 Database SQL.....	91
2.17.4 SQL Query.....	92
3 Script and Expression.....	93
3.1 JShell(Java interactive coding tool) .....	93
3.2 JEXL(Java Expression Language) .....	95
3.3 Javascript .....	97
4 Math Function.....	98
4.1 Define Math Functions.....	98
4.2 Calculate Math Function.....	98
4.3 Data Set.....	100
4.4 XY Chart of Unary Function.....	101
4.5 XYZ Chart of Binary Function.....	102
5 Data of Location.....	103
5.1 Data Constraints.....	103
5.1.1 Invalid Value.....	103
5.1.2 Coordinate System.....	103
5.1.3 Coordinate Values.....	103
5.1.4 Time.....	103
5.1.4.1 Date Formats.....	103
5.1.4.2 Era .....	103
5.1.4.3 Examples .....	104
5.2 Data Operations.....	104
5.3 Map Data.....	105
5.4 Geography Code.....	107
5.4.1 Data Definition.....	107
5.4.2 Data Constraints.....	107
5.4.3 Edit Data.....	107
5.4.4 Define Condition.....	107
5.4.5 Import Data.....	107
5.4.5.1 Predefined Data.....	107
5.4.5.2 CSV Format.....	108
5.4.5.3 Data from geoname.org:.....	108
5.4.6 Settings .....	108
5.5 Location in Map.....	110
5.6 Location Data.....	111
5.6.1 Data Definition.....	111
5.6.2 Data constraints.....	111
5.6.3 Define Conditions.....	111
5.6.4 Map Data.....	111
5.6.5 Snapshots.....	112

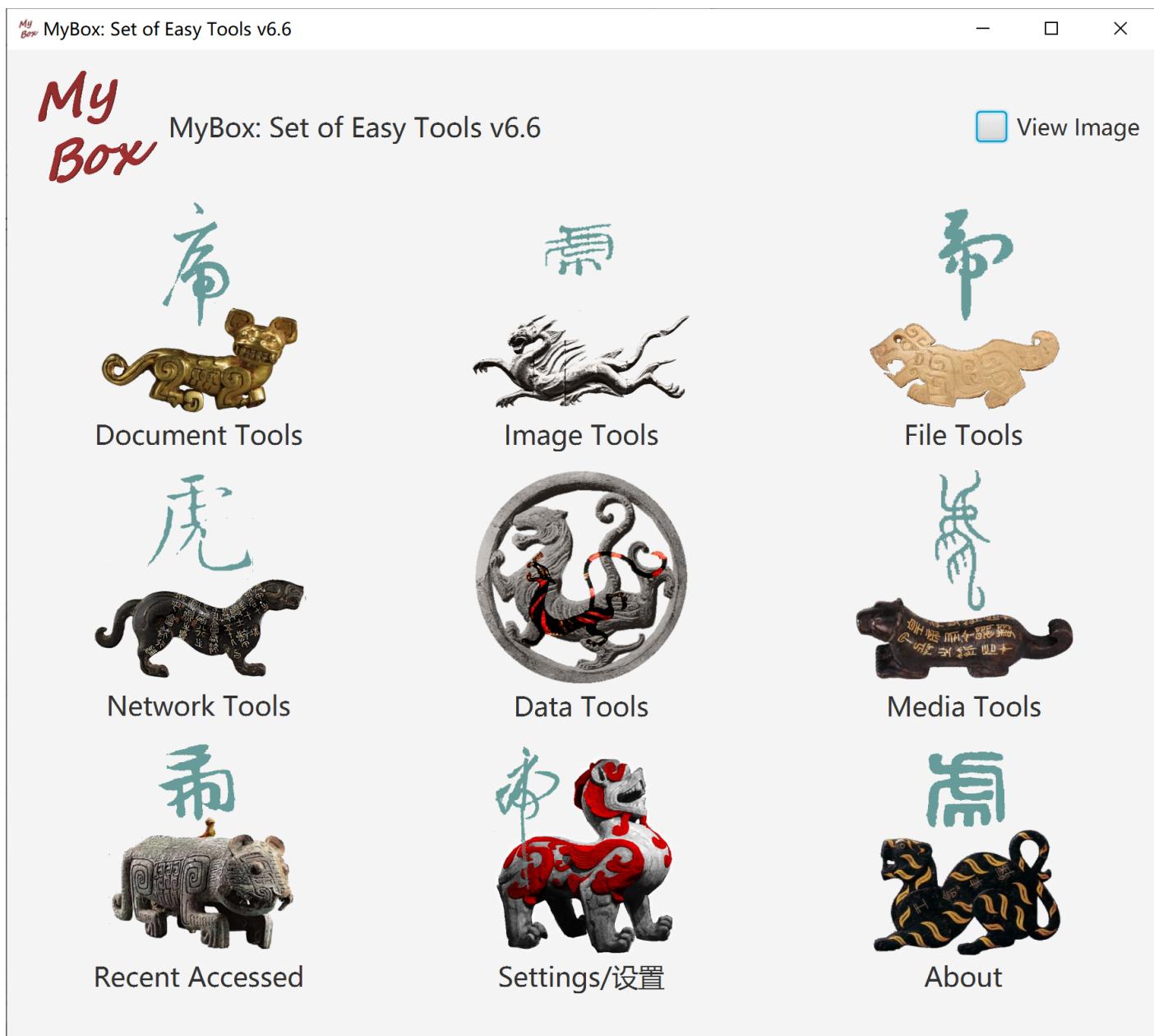
## MyBox User Guide – Data Tools v6.6

5.6.6 Import Data.....	112
5.6.6.1 CSV Format.....	112
5.6.6.2 Data from movebank.org.....	112
5.6.6.3 Examples.....	112
5.7 Location Tools.....	114
5.8 Epidemic Reports.....	115
5.8.1 Data Definition.....	115
5.8.2 Data Constraints.....	116
5.8.3 Edit Data.....	116
5.8.4 Import Data.....	116
5.8.4.1 Predefined Data.....	116
5.8.4.2 CSV format:.....	116
5.8.4.3 COVID-19 historical data from Johns Hopkins University(Global) :.....	116
5.8.4.4 COVID-19 daily data from Johns Hopkins University(Global) :.....	117
5.8.4.5 Handle Data of Imported.....	117
5.8.5 Statistics Data.....	117
5.8.6 Define Conditions.....	117
5.8.7 Display Charts.....	118
5.8.8 Settings.....	118
6 Others.....	119
6.1 Create Barcodes.....	119
6.2 Decode Barcodes.....	119
6.3 Message Digest.....	120
6.4 Encode/Decode Base64.....	120
6.5 Extract ttf files from ttc file.....	120

## 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	<a href="https://github.com/Mararsh/MyBox/">https://github.com/Mararsh/MyBox/</a>
Source Codes and Compiled Packages	<a href="https://github.com/Mararsh/MyBox/releases">https://github.com/Mararsh/MyBox/releases</a>
Submit Software Requirements and Problem Reports	<a href="https://github.com/Mararsh/MyBox/issues">https://github.com/Mararsh/MyBox/issues</a>
Data	<a href="https://github.com/Mararsh/MyBox_data">https://github.com/Mararsh/MyBox_data</a>
Documents	<a href="https://github.com/Mararsh/MyBoxDoc">https://github.com/Mararsh/MyBoxDoc</a>
Mirror Site	<a href="https://sourceforge.net/projects/mara-mybox/files/">https://sourceforge.net/projects/mara-mybox/files/</a>
Cloud Storage	<a href="https://pan.baidu.com/s/1fWMRzym_jh075OCX0D8yA#list/path=%2F">https://pan.baidu.com/s/1fWMRzym_jh075OCX0D8yA#list/path=%2F</a>

The screenshot shows a GitHub repository page for 'Mararsh / MyBox'. The top navigation bar includes links for Pull requests, Issues, Marketplace, and Explore. Below the navigation bar, the repository name 'Mararsh / MyBox' is displayed along with its status as 'Public'. The main content area shows a list of files in the 'master' branch, including 'alpha/MyBox', 'docs', 'en', 'released/MyBox', '.gitignore', 'LICENSE', and 'README.md'. The 'Code' button in the top right of the file list is circled in red. To the right of the file list, there is an 'About' section listing various tools and technologies used, such as html, markdown, pdf, image, ocr, csv, sql, database, excel, convert, location, javafx, media, bytes, ppt, jshell, and matirx. Below the 'About' section, there are links for Readme, Apache-2.0 license, 79 stars, 4 watching, and 17 forks. At the bottom right of the page, a 'Releases' section is shown, featuring a link to 'v6.5.8' (labeled as 'Latest') which was released 18 days ago, and a note indicating '+ 88 releases'. The URL 'https://github.com/Mararsh/MyBox/releases' is visible at the bottom of the browser window.

## 1.3 Documents

Name	Version	Time	English	Chinese
Development Logs	6.6	2022-9-28	<a href="#">html</a>	<a href="#">html</a>
Shortcuts	6.5.6	2022-6-11	<a href="#">html</a>	<a href="#">html</a>
Packing Steps	6.3.3	2020-9-27	<a href="#">html</a>	<a href="#">html</a>
Development Guide	2.1	2020-8-27	<a href="#">PDF</a>	<a href="#">PDF</a>
User Guide - Overview	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Document Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Image Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - File Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Network Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Data Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Media Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>
User Guide - Development Tools	6.6	2022-9-28	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>	<a href="#">html</a> <a href="#">PDF</a> <a href="#">odt</a>

## 1.4 Menu of Tools

The screenshot shows the MyBox application window. At the top left, it says "My Box" and "MyBox: Set of Easy Tools v6.5.9". At the top right, there are window control buttons and a checked checkbox for "View Image". Below the title bar, there are four main categories: "Document Tools" (with a document icon), "Network Tools" (with a tiger icon), "Recent Accessed" (with a bear icon), and a "Data" section which is currently active. The "Data" section has a blue background and contains the following items:

- Manufacture Data** (highlighted)
- Manage Data
- Splice Data
- Separator
- Data File
- Matrix
- Database
- Separator
- Data in System Clipboard
- Data in MyBox Clipboard
- Separator
- Calculation
- Separator
- Location
- Separator
- Miscellaneous
- Setting

At the bottom right of the window, there is a note: "Embroidery: Warring States Chu Tomb(1042 BC-223 BC) Jingzhou Museum" and "Pottery tile: Han Dynasty(202 BC-220) Beijing Palace Museum". A small note at the bottom left says "Close(ESC/F6 Or click anywhere outside)".

## 2 Data in Two-dimensional Storage Structure

### 2.1 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 are represented as two-dimensional storage structure:

- "Columns" define dimensions of data in horizontal direction.
- "Rows" save values of data in vertical direction..
- Data should be in same width. That is all rows have equal number of columns.

The screenshot shows the MyBox Data Tools application window. The title bar reads "Edit CSV File : CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv". The menu bar includes "Window", "Document", "Image", "Data" (which is selected and highlighted in blue), "File", "Media", "Network", "Settings", "Recent Accessed", "Development", and "Help". Below the menu is a toolbar with various icons for file operations like Open, Save, Print, and Export. A large table on the left contains data for 14 rows and 8 columns, with the first column being checkboxes and the second column labeled "age". To the right of the table is a preview pane showing a smaller version of the same data. A context menu is open over the table, listing options: "Edit CSV File", "Convert/Split CSV Files", "Merge CSV Files", "Edit Excel File", "Convert/Split Excel Files", "Merge Excel Files", "Edit Text Data File" (which is also highlighted in blue), "Convert/Split Text Data Files", and "Merge Text Data Files". At the bottom of the interface, there are buttons for "Selected: 0", "Rows: 50/442", "Page size: 50", "Page: 1", and "Page: /9", along with navigation arrows for the table.

## 2.2 Functions Menu

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

The screenshot shows the MyBox Data Tools interface with a CSV file named "DiabetesPrediction\_en.csv" loaded. The "Edit" tab in the top menu bar is selected, and a context menu is open over a data table. The "Edit" tab has a blue border, and the "Edit" icon in the toolbar is circled in red. The context menu is titled "Data" and includes options like Save, Recover, Refresh, Open, Create data, Load content in system clipboard, Export, Convert to database table, Modify, Trim, Calculation, Charts, Examples, and a checked checkbox for "Pop when mouse hovering". Below the table, status information shows "Selected: 0", "Rows: 50/442", "Page size: 50", "Page: 1 / 9", and navigation icons.

Index	Glucose	B.P.	Skin	Insulin	B.F.	Age
1	1	59				
2	2	48				
3	3	72				
4	4	24				
5	5	50				
6	6	23				
7	7	36				
8	8	66	2	26.2	114	
9	9	60	2	32.1	83	
10	10	29	1	30	85	
11	11	22	1	18.6	97	
12	12	56	2	28	85	
13	13	53	1	23.7	92	
14	14	50	2	26.2	97	

Selected: 0 Rows: 50/442 Page size: 50 Page: 1 / 9

## 2.3 Edit Data

### 2.3.1 Save and Recover

1. When changed, \* is displayed in tab header. And \*\* is displayed when modifications have not applied.
2. 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.
3. Click button "Recover" to discard all modifications and load data from file and database.

The screenshot shows the MyBox Data Tools interface. The main window displays a table of data from a CSV file named 'DiabetesPrediction\_en.csv'. The table has 14 rows and 10 columns, with various numerical and categorical values. A context menu is open over the first row, showing options like 'Edit', 'Delete', 'Copy', etc. A modal dialog box is overlaid on the table, prompting the user to 'Need save data before handle'. It contains three buttons: 'Save' (highlighted in blue), 'Not save', and 'Cancel'. The status bar at the bottom indicates 'Selected: 0 Rows: 50/442 Page size 50 / Page 1 / 9'.

Tabl...	Data ...	age	sex	BMI(body ...	BP(averag...	S1(blood ...	S2(blood ...	S3(blood ...	S4(blo...
1	1	59	2	32.1	101	157	93.2	38	4
2	2	48	1	21.6	87	183	103.2	70	3
3	3	72	2	30.5	93	156	93.6	41	4
4	4	24	1						
5	5	50	1						
6	6	23	1						
7	7	36	2						
8	8	66	2						
9	9	60	2						
10	10	29	1						
11	11	22	1	18.6	97	114	57.6	46	2
12	12	56	2	28	85	184	144.8	32	6
13	13	53	1	23.7	92	186	109.2	62	3
14	14	50	2	26.2	97	186	105.4	49	4

### 2.3.2 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.

The screenshot shows the 'MyBox' application interface for editing a CSV file. The main window title is 'Edit CSV File : CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv \*'. The menu bar includes 'Window', 'Document', 'Image', 'Data', 'File', 'Media', 'Network', 'Settings', 'Recent Accessed', 'Development', and 'Help'. The toolbar features icons for document, image, data, file, media, network, settings, recent accessed, development, and help. A color palette dialog is open over the main table, specifically for the 'Color' column of row 4, which contains the value 'BP(average blood pressure)'. The color palette lists various color models and their hex values, such as Click to edit (hex 100), 0x9E556AFF, #9E556A, -6400662, sRGB: 158 85 106 100%, HSB: 343 46% 62%, Adobe RGB: 141 86 105, Apple RGB: 137 64 87, ECI RGB: 148 93 115, sRGB Linear: 87 23 37, Adobe RGB Linear: 69 23 36, Apple RGB Linear: 83 21 37, Calculated CMYK: 0 46 33 38, ECI CMYK: 37 62 45 0, Adobe CMYK Uncoated FOGRA29: 35 78 42 10, XYZ: 0.204702 0.149937 0.116517, CIE-L\*ab: 45.63 32.65 2.11, LCH(ab): 45.63 32.72 3.69, CIE-L\*uv: 45.63 49.19 -3.98, and LCH(uv): 45.63 49.35 355.38. The table has columns for 'Ta...', 'Index', 'Column name', 'Type', 'Color', 'Width', 'Editable', 'Not null', and 'Separat...'. The 'Color' column for index 4 is currently selected and highlighted with a blue border.

Ta...	Index	Column name	Type	Color	Width	Editable	Not null	Separat...
<input type="checkbox"/>	1	age	Double	[Color Box]	61	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	2	sex	Short	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	3	BMI(body mass index)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	4	BP(average blood pressure)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	5	S1(blood serum measurement 1)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	6	S2(blood serum measurement 2)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	7	S3(blood serum measurement 3)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	8	S4(blood serum measurement 4)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	9	S5(blood serum measurement 5)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	10	S6(blood serum measurement 6)	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	11	disease progression one year after ...	Double	[Color Box]		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Selected: 0 Rows: 50/442 Page size 50

7. Edit the cell directly, or select row and click button “Edit” to pop Edit Window.

MyBox: Set of Easy Tools

CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv  
Column 4

Column name

Type  String  Double  Float  Long  Integer  Short  Boolean  Date

Default value

Length  (<= 32672)

Width

Color  

Not null  Editable  Separated by comma

Description

### 2.3.3 Define Data Attributes

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

The screenshot shows the MyBox Data Tools application window. The title bar reads "Edit CSV File : CSV - 7 - d:\tmp\mybox-data-6.5.9\data\SouthGermanCredit\_en.csv". The menu bar includes "Window", "Document", "Image", "Data", "File", "Media", "Network", "Settings", "Recent Accessed", "Development", and "Help". The toolbar features various icons for file operations like Open, Save, Print, and Find. The main area has tabs for "Columns" and "Attributes", with "Attributes" currently selected. A sub-toolbar below the tabs includes icons for "View", "Edit", and "Delete". The main content area displays the following information:

- ID:** 7
- Data type:** CSV
- Data name:** South German credit
- Decimal scale:** 2
- Maximum value of random:** 1000
- Description:**

http://archive.ics.uci.edu/ml/datasets/South+German+Credit  
700 good and 300 bad credits with 20 predictor variables. Data from 1973 to 1975. Stratified sample from actual credits with bad credits heavily oversampled. A cost matrix can be used.
- Update time:** 2022-09-28 11:17:37
- File details:**
  - File: d:\tmp\mybox-data-6.5.9\data\SouthGermanCredit\_en.csv
  - File size: 46.934 KB
  - File modify time: 2022-09-28 11:17:35
  - Charset: UTF-8
  - Delimiter: ,
  - First line defines the columns' names: Yes
  - Lines number in file: 1000
  - Columns number: 21
  - Current page: 1 / 20
  - Rows range in page: 1 - 50 ( 50 )
  - Page modify time: 2022-09-28 11:17:40

At the bottom, there are buttons for "Selected: 0", "Rows: 50/1000", "Page size: 50", "Page: 1 / 20", and navigation arrows (left, right, first, last).

### 2.3.4 Edit Mode "Table"

"Table" is the master edit mode:

1. Its modifications are applied to other panes automatically.
2. It is the final data to save.
3. To string values, except for text file, multiple lines can be edited and saved:
  - When the value is single line(not contain line break):
    - Text field is shown when click the data cell.
    - Write "\n" as line break in the value and commit the change(return or click other place).
  - When the value contains line breaks:
    - Text area is shown when click the data cell.
    - Write the text in multiple lines directly.

Table...	Data ...	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
<input type="checkbox"/> 1	1	5.1	3.5	1.4	0.2	Iris-setosa
<input checked="" type="checkbox"/> 2	2	4.9	3	1.4	0.2	Iris-setosa Tyr multiple lines
<input type="checkbox"/> 3	3	4.7	3.2	1.3	0.2	Iris-setosa
<input type="checkbox"/> 4	4	4.6	3.1	1.5	0.2	Iris-setosa
<input type="checkbox"/> 5	5	5	3			Iris-setosa
<input type="checkbox"/> 6	6	5.4	3			Iris-setosa
<input type="checkbox"/> 7	7	4.6	3			Iris-setosa
<input type="checkbox"/> 8	8	5	3			Iris-setosa
<input type="checkbox"/> 9	9	4.4	2			Iris-setosa
<input type="checkbox"/> 10	10	4.9	3			Iris-setosa
<input type="checkbox"/> 11	11	5.4	3			Iris-setosa
<input type="checkbox"/> 12	12	4.8	3			Iris-setosa
<input type="checkbox"/> 13	13	4.8	3			Iris-setosa
<input type="checkbox"/> 14	14	4.3	3	1.1	0.1	Iris-setosa
<input type="checkbox"/> 15	15	5.8	4	1.2	0.2	Iris-setosa
<input type="checkbox"/> 16	16	5.7	4.4	1.5	0.4	Iris-setosa
<input type="checkbox"/> 17	17	5.4	3.9	1.3	0.4	Iris-setosa
<input type="checkbox"/> 18	18	5.1	3.5	1.4	0.3	Iris-setosa
<input type="checkbox"/> 19	19	5.7	2.8	1.7	0.3	Iris-setosa

Selected: 0 Rows: 50/150 Page size 50 Page 1 /3

### 2.3.5 Edit Mode "Text"

"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.

The screenshot shows the MyBox Data Tools interface. At the top, there is a toolbar with several icons: Columns\*, Attributes, View, and Edit (which is highlighted with a blue circle). Below the toolbar, there is a status bar with some text and icons. The main area contains a table with data and a floating dialog box titled "MyBox: Set of Easy Tools". The dialog box is titled "Text delimiter" and contains various options for defining a delimiter. The "String" option is selected. Other options include "Tab", "Blank", "Blank characters", "4 blanks", and "8 blanks". Below these are several character buttons, each with a specific symbol like '|', '#', '=', etc. At the bottom right of the dialog box are two small icons. The table below the toolbar contains the following data:

year	population at year-end(ten thousand)	male(ten thousand)	female(ten thousand)	urban(ten thousand)	rural(ten thousand)
1949	54167.0	28145.0	26022.0	5765.0	48402.0
1950	55196.0	28669.0	26527.0	6169.0	49027.0
1951	56300.0	29231.0	27069.0	6632.0	49668.0
1952	57482.0	29833.0	27649.0	7163.0	50319.0
1953	58796.0	30468.0	28328.0	7826.0	50970.0
1954	60266.0	31242.0	29024.0	8249.0	52017.0
1955	61465.0	31809.0	29656.0	8285.0	53180.0
1956	62828.0	32536.0	30292.0	9185.0	53643.0
1957	64653.0	33469.0	31184.0	9949.0	54704.0
1958	65994.0	34195.0	31799.0	10721.0	55273.0
1959	67207.0	34890.0	32317.0	12371.0	54836.0
1960	66207.0	34283.0	31924.0	13073.0	53134.0
1961	65859.0	33880.0	31979.0	12707.0	53152.0
1962	67296.0	34517.0	32778.0	11659.0	55636.0
1963	69172.0	35533.0	33639.0	11646.0	57526.0

### 2.3.6 View Html

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

The screenshot shows a table with the following data:

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

### 2.3.7 View Text

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

The screenshot shows the raw text output of the population data:

```
Row number#year_#population at year-end(ten thousand)#male(ten thousand)#female(ten thousand)#urban(ten thousand)#rural(ten thousand)
Row1#1949#54167.0#28145.0#26022.0#5765.0#48402.0
Row2#1950#55196.0#28669.0#26527.0#6169.0#49027.0
Row3#1951#56300.0#29231.0#27069.0#6632.0#49668.0
Row4#1952#57482.0#29833.0#27649.0#7163.0#50319.0
Row5#1953#58796.0#30468.0#28328.0#7826.0#50970.0
Row6#1954#60266.0#31242.0#29024.0#8249.0#52017.0
Row7#1955#61465.0#31809.0#29656.0#8285.0#53180.0
Row8#1956#62828.0#32536.0#30292.0#9185.0#53643.0
Row9#1957#64653.0#33469.0#31184.0#9949.0#54704.0
Row10#1958#65994.0#34195.0#31799.0#10721.0#55273.0
Row11#1959#67207.0#34890.0#32317.0#12371.0#54836.0
Row12#1960#66207.0#34283.0#31924.0#13073.0#53134.0
Row13#1961#65859.0#33880.0#31979.0#12707.0#53152.0
```

A separate window titled "Text delimiter" is open, showing various options for setting the delimiter, with the "String" option selected.

### 2.3.8 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.

Load content in system clipboard.

First line defines the columns' names   

Source row number	year_	population at year-end(ten thousand),r	male(ten thousand)	female(ten thousand)	Delimiter
Row1	1949	54167	28145	26022	,
Row2	1950	55196	28669	26527	,
Row3	1951	56300	29231	27069	,
Row4	1952	57482	29833	27649	,
Row5	1953	58796	30468	28328	,
Row6	1954	60266	31242	29024	,
Row7	1955	61465	31809	29656	,
Row8	1956	62828	32536	30200	,

## 2.3.9 Import Examples

### 2.3.9.1 Statistic Data of China

Manufacture Data : CSV - 786 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

Window Document Image Data File Media Network Settings Recent Accessed

Statistic data of China

Regression

Pop when mouse hovering

Close(ESC/F6 Or click anywhere outside the object)

		1950	55,190	28,069
<input type="checkbox"/>	3	1951	56,300	29,231
<input type="checkbox"/>	4	1952	57,482	29,833
<input type="checkbox"/>	5	1953	58,796	30,468
<input type="checkbox"/>	6	1954	60,266	31,242
<input type="checkbox"/>	7	1955	61,465	31,809
<input type="checkbox"/>	8	1956	62,828	32,536
<input type="checkbox"/>	9	1957	64,653	33,469
<input type="checkbox"/>	10	1958	65,994	34,195
<input type="checkbox"/>	11	1959	67,207	34,890
<input type="checkbox"/>	12	1960	66,207	34,283

Population of China  
Census of China  
Gross domestic product(GDP) of China  
Consumer price index(CPI) of China  
Food consumption of China  
Graduates of China  
**Museums of China**  
Health personnel of China  
Marriage of China  
Sport world champions of China  
Crimes filed by China police  
Crimes filed by China procuratorate  
China National Bureau of Statistics

### 2.3.9.2 Data of Regression

Edit CSV File : CSV - 8 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies\_en.csv \*

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Statistic data of China  
Regression  
Pop when mouse hovering  
Close(ESC/F6 Or click anywhere outside the object)

	2	4.9	3	1.4
1				
2	2	4.9	3	1.4
3	3	4.7	3.2	1.3
4	4	4.6	3.1	1.5
5	5	5	3.6	1.4
6	6	5.4	3.9	1.7
7	7	4.6	3.4	1.4
8	8	5	3.4	1.5
9	9	4.4	2.9	1.4
10	10	4.9	3.1	1.5
11	11	5.4	3.7	1.5
12	12	4.8	3.4	1.6
13	13	4.8	3	1.4
14	14	4.3	3	1.1

Columns\* Attributes View Edit\*

Income and happiness  
Years experience and salary  
Iris species  
Progression Prediction of Diabetes  
Progression Prediction of Diabetes - standardized  
Heart failure  
**Concrete compressive strength**  
Dog's radiographs after surgery  
Salaries of baseball players  
South German credit  
Boston housing prices

0.1 Iris-setosa  
0.2 Iris-setosa  
0.2 Iris-setosa  
0.1 Iris-setosa  
0.1 Iris-setosa

Selected: 0 Rows: 50/150 Page size 50 Page 1 /3

## 2.4 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:
  - o It can include any valid elements which Nashorn can parse(ECMAScript 5.1).
  - o It should return a value finally.
  - o It can include following placeholders:

`#{{TableRowNumber}}`

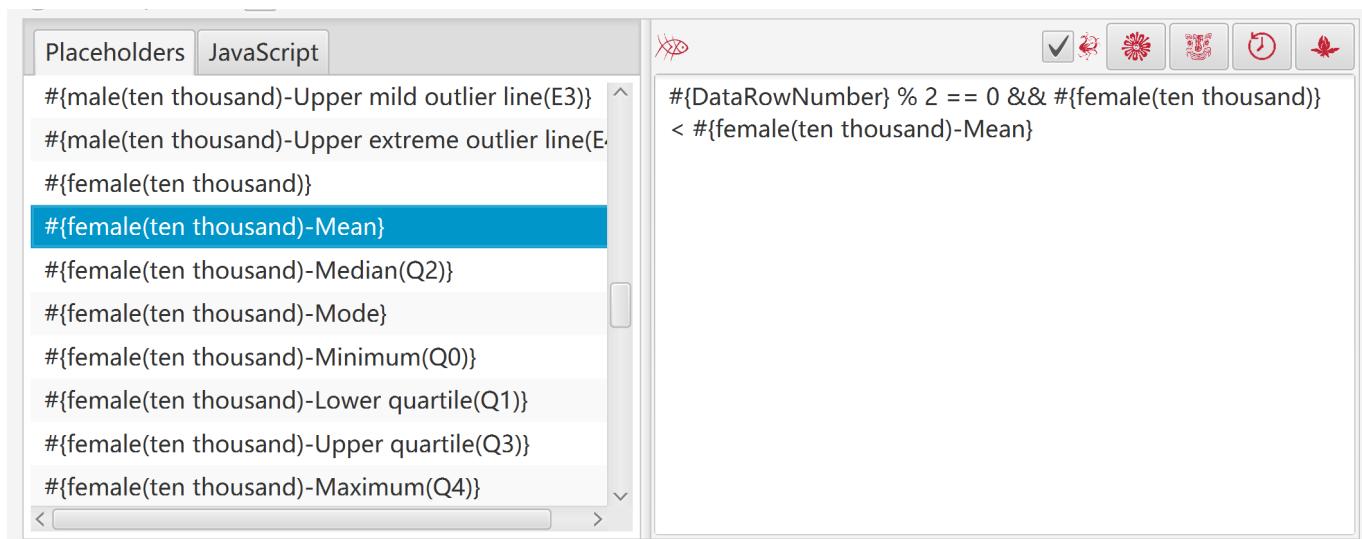
`#{{DataRowNumber}}`

`#{{<column_name>}}`

`#{{<column_name>- }}`

All valid placeholders are listed in left.

- o To save the script, click "Edit" button.
3. When MyBox evaluates the expression:
    - o Placeholders are replaced with actual values of each data row .
    - o Statistic values are calculated by all data.
    - o '`#{{xxx}}`' is handled as string while `#{xxx}` is handled as number.
    - o 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.



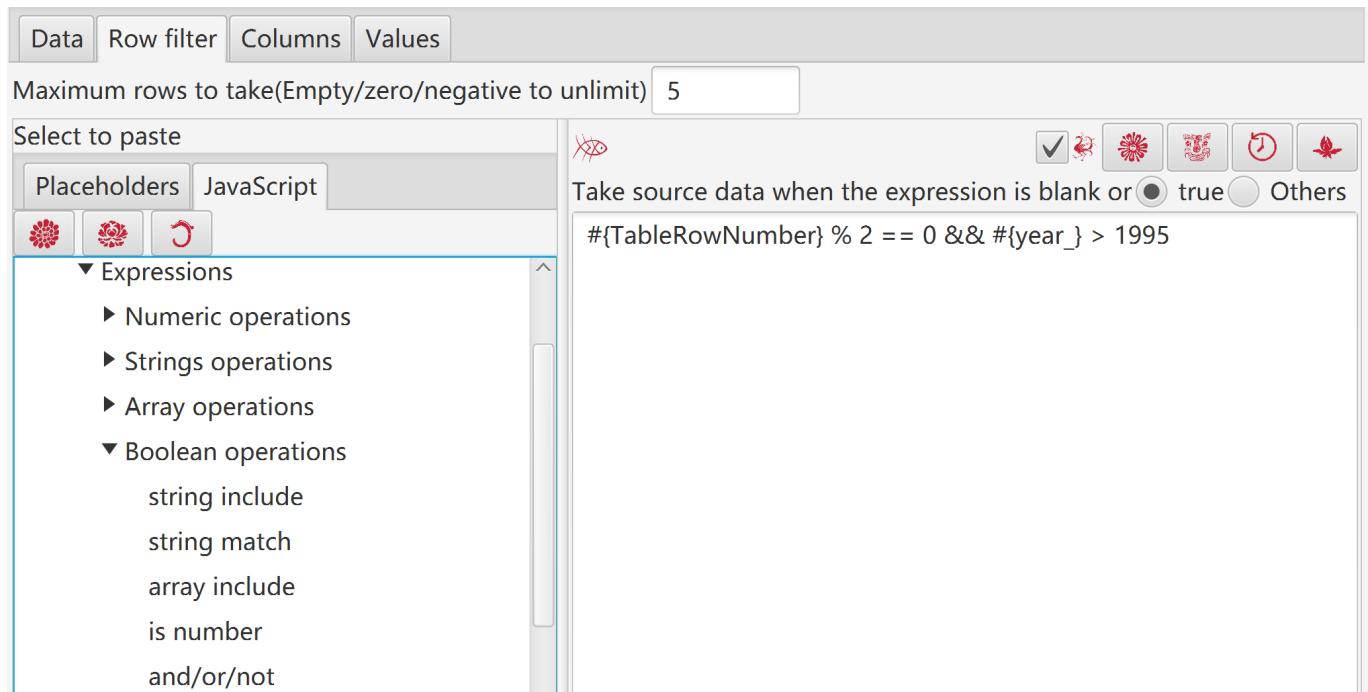
## 6. Examples:

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

## 2.5 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. Examples:

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”

## 2.6 Modify

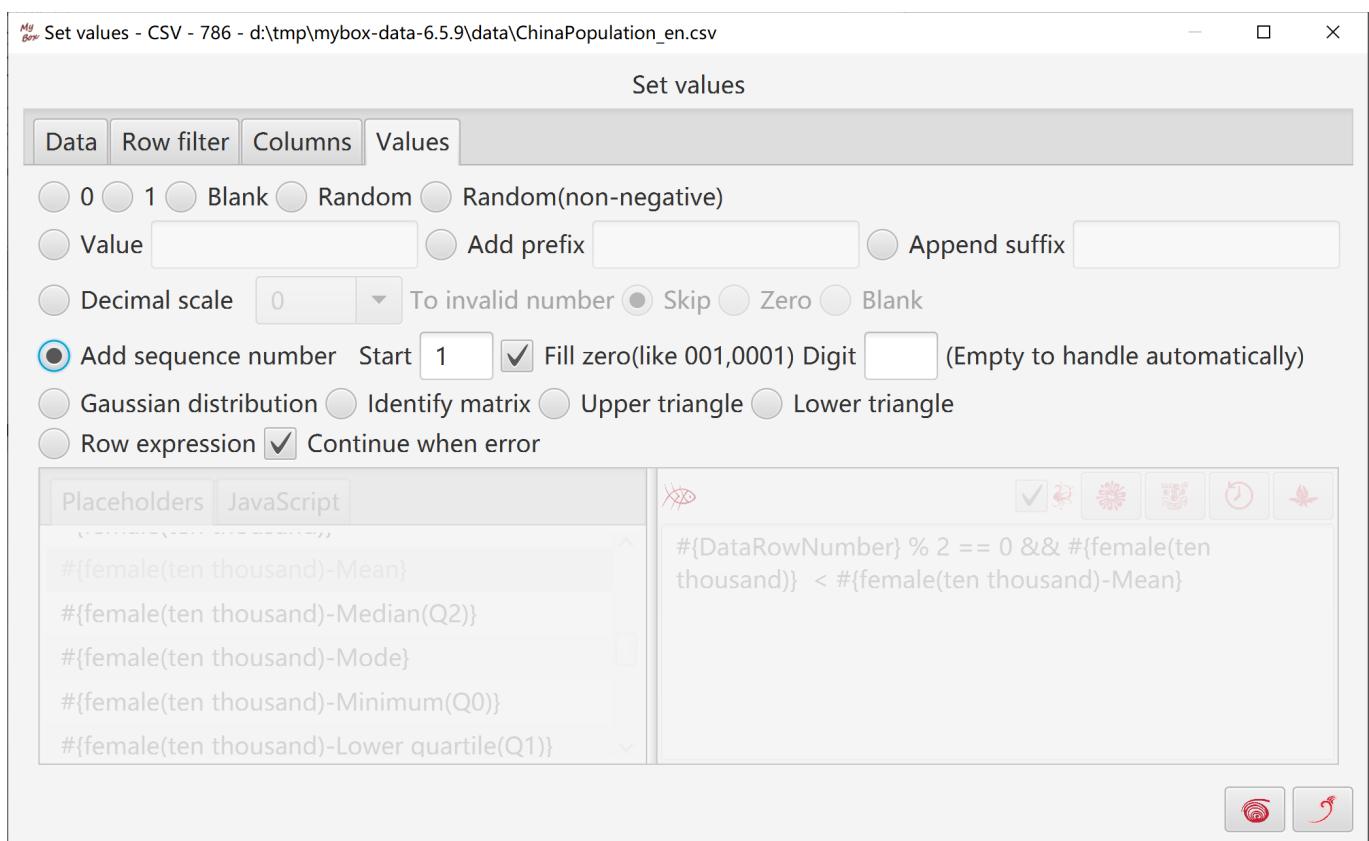
Hover or click button “Function” to select functions under menu item “Modify”.

The screenshot shows the MyBox Data Tools interface with a CSV file named 'IrisSpecies\_en.csv' open. The 'Modify' function menu is displayed as a dropdown on the right side of the screen, containing options like Set values, Delete, Trim, Calculation, Charts, and Examples. A checkmark is present next to 'Pop when mouse hovering'. The main window displays a table of Iris flower data with columns for Sepal Length, Sepal Width, Petal Length, Petal Width, and Species.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	5.1	4.9	4.7	4.6	5	5.4	4.6	5	4.4	4.9	5.4	4.8	4.8	4.3
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	3.6	3.9	3.4	3.4	3.6	3.9	3.4	3.4	2.9	3.1	3.7	3.4	3	3
	1.4	1.7	1.4	1.5	1.4	1.7	1.4	1.5	1.4	1.5	1.5	1.6	1.4	1.1
	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1
	Iris-setosa													

## 2.6.1 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 square, 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.



## 2.6.2 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.

MyBox Delete - CSV - 786 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

Delete

Data Row filter Options

Rows  Selected  Current page  All pages 

<input type="checkbox"/> Table row	<input checked="" type="checkbox"/> Data row	Select some to handle, or select none to handle all in table: n(te... rural(ten ...					
<input type="checkbox"/> 1	1	1952	167	28,145	26,022	5,765	48,402
<input type="checkbox"/> 2	2	1953	186	23,669	16,527	6,169	40,027
<input type="checkbox"/> 3	3	1954	196	20,458	18,229	8,236	49,668
<input type="checkbox"/> 4	4	1955	205	57,482	29,833	7,163	50,319
<input type="checkbox"/> 5	5	1956	214	59,796	30,458	8,236	50,970
<input type="checkbox"/> 6	6	1957	223	61,463	31,809	8,285	53,160
<input type="checkbox"/> 7	7	1958	232	61,463	31,809	8,285	53,160
<input type="checkbox"/> 8	8	1959	241	61,463	31,809	8,285	53,643
<input type="checkbox"/> 9	9	1960	250	61,463	31,809	8,285	54,704
<input type="checkbox"/> 10	10	1961	259	61,463	31,809	10,721	55,273
<input type="checkbox"/> 11	11	1962	268	61,463	31,809	10,721	54,836
<input type="checkbox"/> 12	12	1963	277	61,463	31,809	11,658	55,636
<input type="checkbox"/> 13	13	1964	286	61,463	31,809	11,658	57,526
<input type="checkbox"/> 14	14	1965	295	69,172	35,533	11,646	57,549
<input type="checkbox"/> 15	15	1966	304	69,172	35,533	11,646	
<input type="checkbox"/> 16	16	1967	313	70,499	36,142	12,950	
		1968	322	70,499	36,142	12,950	
		1969	331	70,499	36,142	12,950	
		1970	340	70,499	36,142	12,950	
		1971	349	70,499	36,142	12,950	
		1972	358	70,499	36,142	12,950	
		1973	367	70,499	36,142	12,950	

Selected: 0 Rows: 60/73 Page size 50  Page 1  /2      

### **2.6.3 Set styles / Mark Abnormal Values**

### 2.6.3.1 Manage Styles

1. Add/Edit/Delete/Rename styles.
  2. Define title and sequence number of the style.
  3. Set whether the style marks abnormal values.

My Box Set Styles / Mark Abnormal Values - CSV - 63 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

## Set Styles / Mark Abnormal Values

Seq...	Abnormal ...	Title	From	To
1.0		Analyse of Female	1	-1
2.0		Analyse of Male	1	-1
3.0	✓	Abnormal of Fem...	25	-1
4.0	✓	Abnormal of Male	25	-1

< >

Page 1 /1

Page size 50 Rows: 4/4 Selected: 1

ID: 7

Base Data Filter Style

Sequence 3.0

Title Abnormal of Female

This rule defines abnormal values

### 2.6.3.2 Data Scope

Define conditions to determine which data cells to apply the style:

1. Range of data rows
2. Column names.
3. 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 special rows is the expression composed of column values.

Seq...	Abn...	Title	From
1.0		Analyse of Female	1
2.0		Analyse of Male	1
3.0	<input checked="" type="checkbox"/>	Abnormal of Fem...	25
4.0	<input checked="" type="checkbox"/>	Abnormal of Male	25

Set Styles / Mark Abnormal Values

ID: 7

Base Data Filter Style

DataRowNumber

From 25

Blank, zero, or negative means no limitation to row number

To Blank, zero, or negative means the last row

Columns to be calculated (No selection means all)

year\_population at year-end(ten thousand)

male(ten thousand)  female(ten thousand)

urban(ten thousand)  rural(ten thousand)

Page 1 /1  Page size 50 Rows: 4/4 Selected: 1

### 2.6.3.3 Define Styles

1. Font color, font size, background color, bold, etc.
2. More values in format of JavaFx CSS.

Set Styles / Mark Abnormal Values - CSV - 63 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

Set Styles / Mark Abnormal Values

Seq...	Abn...	Title	From	
1.0		Analyse of Female	1	-1
2.0		Analyse of Male	1	-1
3.0	✓	Abnormal of Fem...	25	-1
4.0	✓	Abnormal of Male	25	-1

Page 1 /1    Page size 50 Rows: 4/4 Selected: 1

ID: 7

Base Data Filter Style

Font color  Default  Set    

Background color  Default  Set    

Font size 1.2em   Bold

More  

[JavaFX CSS Reference Guide](#)

### 2.6.3.4 Apply Styles

When load data page, all styles of the data are applied to the rows one by one in order of their sequence numbers.

Edit CSV File : CSV - 63 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

Window Document Image Data File Media Network Settings Recent Accessed Development Help

#	##	Columns	Attributes	View	Edit	Table				
1	Tabl...	Data ...	year_	populatio...	male(ten t...	female(te...	urban(ten ...	rural(ten t...		+ <input type="button" value=""/>
2	24	24	1972	87,177	44,813	42,364	14,935	72,242		
3	25	25	1973	89,211	45,876	43,335	15,345	73,866		
4	26	26	1974	90,859	46,727	44,132	15,595	75,264		
5	27	27	1975	92,420	47,564	44,856	16,030	76,390		
6	28	28	1976	93,717	48,257	45,460	16,341	77,376		
7	29	29	1977	94,974	48,908	46,066	16,669	78,305		
8	30	30	1978	96,259	49,567	46,692	17,245	79,014		
9	31	31	1979	97,542	50,192	47,350	18,495	79,047		
10	32	32	1980	98,705	50,785	47,920	19,140	79,565		
11	33	33	1981	100,072	51,519	48,553	20,171	79,901		
12	34	34	1982	101,654	52,352	49,302	21,480	80,174		
13	35	35	1983	103,008	53,152	49,856	22,274	80,734		
14	36	36	1984	104,357	53,848	50,509	24,017	80,340		

Selected: 0 Rows: 50/73 Page size 50 Page 1 /2

## 2.6.4 Paste Content in System Clipboard

The screenshot shows the 'Paste content in System Clipboard' dialog. On the left, there is a text area containing CSV data:

```
year_,population at year-end(ten thousand)
1951,56300,29231,27069,6632,49668
1952,57482,29833,27649,7163,50319
1953,58796,30468,28328,7826,50970
1954,60266,31242,29024,8249,52017
1955,61465,31809,29656,8285,53180
```

A checkbox labeled 'First line defines the columns' names' is checked. Below it is a table with the following structure:

Source row number	year_	population at year-end(ten thousand)	male(ten thousand)	female(t thousand)
Row1	1951	56300	29231	27069
Row2	1952	57482	29833	27649
Row3	1953	58796	30468	28328
Row4	1954	60266	31242	29024
Row5	1955	61465	31809	29656

On the right, there is a panel for 'Location to paste' with 'Table row 1' selected. Below it are buttons for 'Replace', 'Insert above', and 'Append below'. At the bottom are two small icons.

## 2.6.5 Paste Content in MyBox Clipboard

The screenshot shows the 'Paste content in MyBox Clipboard' dialog. On the left, there is a table viewer showing rows 1 through 7 of a table. The first row is selected. The table has columns: Table row, ID, Type, and Name. The data is as follows:

Table row	ID	Type	Name
1	1089	MyBox ...	ChinaPopu
2	1086	MyBox ...	ChinaPopu
3	1085	MyBox ...	ChinaPopu
4	1083	MyBox ...	ChinaPopu
5	1082	MyBox ...	ChinaPopu
6	1063	MyBox ...	a
7	6	MyBox ...	b

At the bottom, there are page navigation buttons and a message: 'Page 1 /1 Rows: 7/7 Selected: 1'.

In the center, there is a table with columns: Table row, Data row, 源行号 (Source Row), 年 (Year), and 年+ (Year+). The data is as follows:

Table row	Data row	源行号	年	年+
1	1	1	1949	54,167
2	2	1	1950	55,196
3	3	2	1951	56,300
4	4	3	1952	57,482
5	5	4	1953	58,796
6	6	5		

On the right, there is a panel for 'Location to paste' with 'Table row 1' selected. Below it are buttons for 'Replace', 'Insert above', and 'Append below'. At the bottom are two small icons.

## 2.7 Trim

Hover or click button “Function” and select functions under menu item “Trim”.

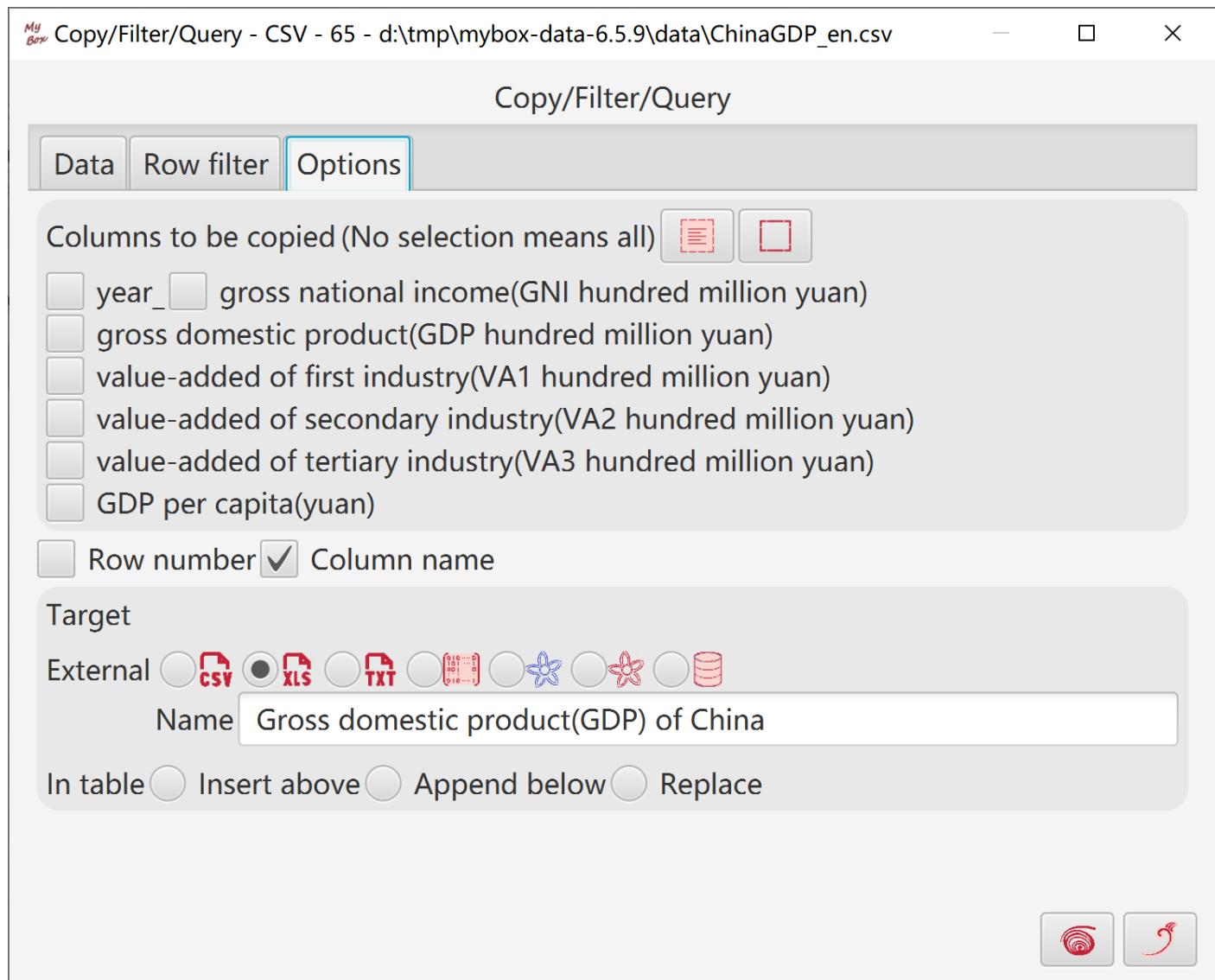
The screenshot shows the MyBox Data Tools interface with a CSV file open. A context menu is displayed over a table of consumption data. The 'item' column of the menu lists various functions: Copy/Filter/Query, Sort, Transpose (which is highlighted), Normalize, Split, Data, Modify, Trim (which is also highlighted), Calculation, Charts, Examples, and a checkbox for 'Pop when mouse hovering'. Below the menu, the table contains data about food consumption per capita in kilogram.

Table...	Data ...	item
1	1	food consumption per capita(kilogram)
2	2	cooking oil consumption per capita(kilogram)
3	3	vegetables and mushrooms consumption per capita(kilogram)
4	4	meat consumption per capita(kilogram)
5	5	poultry consumption per capita(kilogram)
6	6	aquatic products consumption per capita(kilogram)
7	7	eggs consumption per capita(kilogram)
8	8	milk consumption per capita(kilogram)
9	9	fruits consumption per capita(kilogram)
10	10	sugar consumption per capita(kilogram)

Selected: 0 Rows: 10/10 Page size 50 Page 1 /1

## 2.7.1 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.



## 2.7.2 Sort

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - Select the column to sort and whether descending.
  - Data type of column affects sorting results.
  - Maximum rows number of results can be set.
  - Select columns to be copied.
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.

Sort - CSV - 65 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP\_en.csv

Sort

**Data** **Row filter** **Options**

Order by(Column type affects sorting results)

Ta...	Column	
<input type="checkbox"/> 1	year_-Descending	
<input type="checkbox"/> 2	year_-Ascending	
<input checked="" type="checkbox"/> 3	gross national income(GNI hundred million yuan)-Descending	
<input type="checkbox"/> 4	gross national income(GNI hundred million yuan)-Ascending	
<input type="checkbox"/> 5	gross domestic product(GDP hundred million yuan)-Descending	

Columns to be copied (No selection means all)

year  gross national income(GNI hundred million yuan)  
 gross domestic product(GDP hundred million yuan)  
 value-added of first industry(VA1 hundred million yuan)  
 value-added of secondary industry(VA2 hundred million yuan)  
 value-added of tertiary industry(VA3 hundred million yuan)  GDP per capita(yuan)

Maximum result rows to take(Empty/zero/negative to unlimit)

Row number  Column name

Target

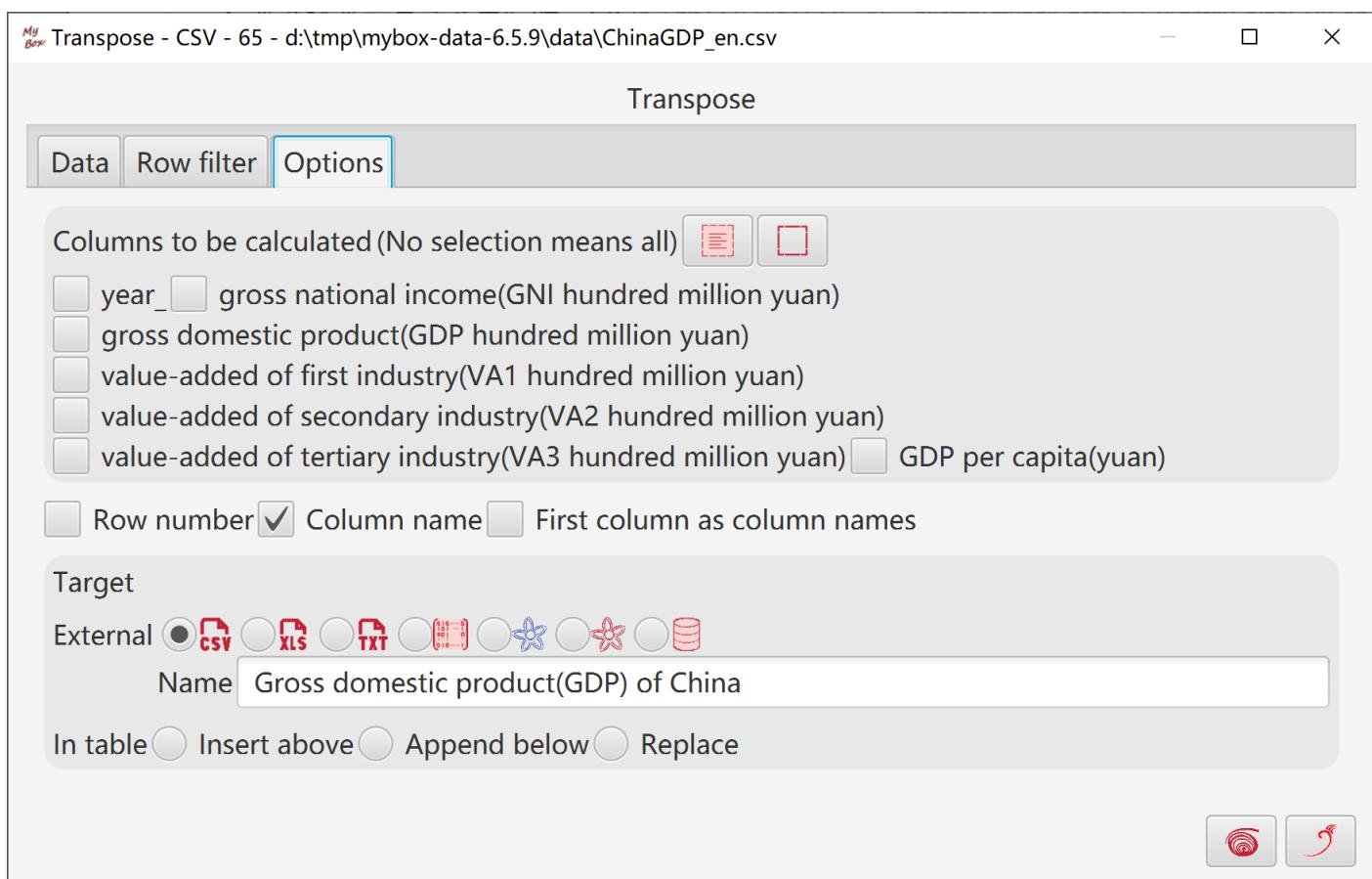
External

Name

In table  Insert above  Append below  Replace

### 2.7.3 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. Calculate: 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.



## 2.7.4 Normalization

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - Select columns to be calculated.
  - According to: Columns/rows/all.
  - Algorithms: MinMax(Range can be set), sum(L1), ZScore(L2).
  - To non-numeric, skip or count as zero.
  - Select columns to be copied.
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.

Normalize - CSV - 8 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies\_en.csv

Normalize

**Data** **Row filter** **Options**

Columns to be calculated (No selection means all)

SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species

According to  Columns  Rows  All

Algorithm  MinMax  L1(Sum)  L2(Z-Score)

Range  ,

Row number  Column name

Decimal scale

To invalid number  Skip  Count as zero

Columns to be copied

SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species

**Target**

External

Name

In table  Insert above  Append below  Replace

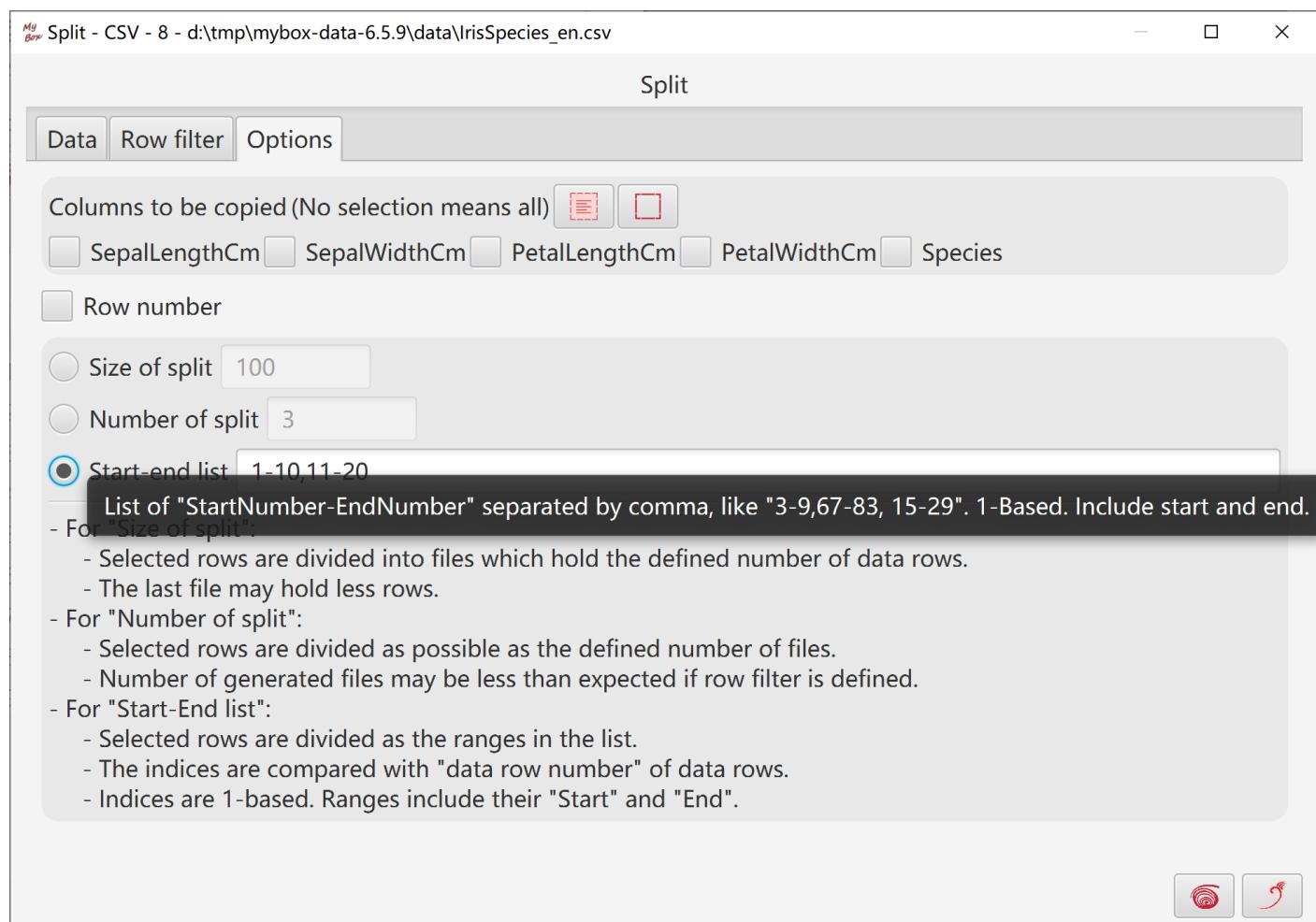
## 2.7.5 Split

### 1. Select data:

- Rows can be: current page, selected rows, or all pages.
- Set row filter.

### 2. Options:

- Select columns to be copied.
- By "Size of split":
  - Selected rows are divided into files which hold the defined number of data rows.
  - The last file may hold less rows.
- By "Number of split":
  - Selected rows are divided as possible as the defined number of files.
  - Number of generated files may be less than expected if row filter is defined.
- By "Start-End list":
  - Selected rows are divided as the ranges in the list.
  - The indices are compared with "data row number" of data rows.
  - Indices are 1-based. Ranges include their "Start" and "End".



## 2.8 Calculation

Hover or click button “Function” to select functions under menu item “Calculation”.

The screenshot shows the MyBox Data Tools interface with a CSV file named "IrisSpecies\_en.csv" loaded. The "Calculation" menu is open, displaying various statistical and analytical functions. The "Descriptive statistics" option is currently selected. The main window displays the first 15 rows of the Iris dataset, which includes columns for Sepal Length, Sepal Width, Petal Length, and Petal Width.

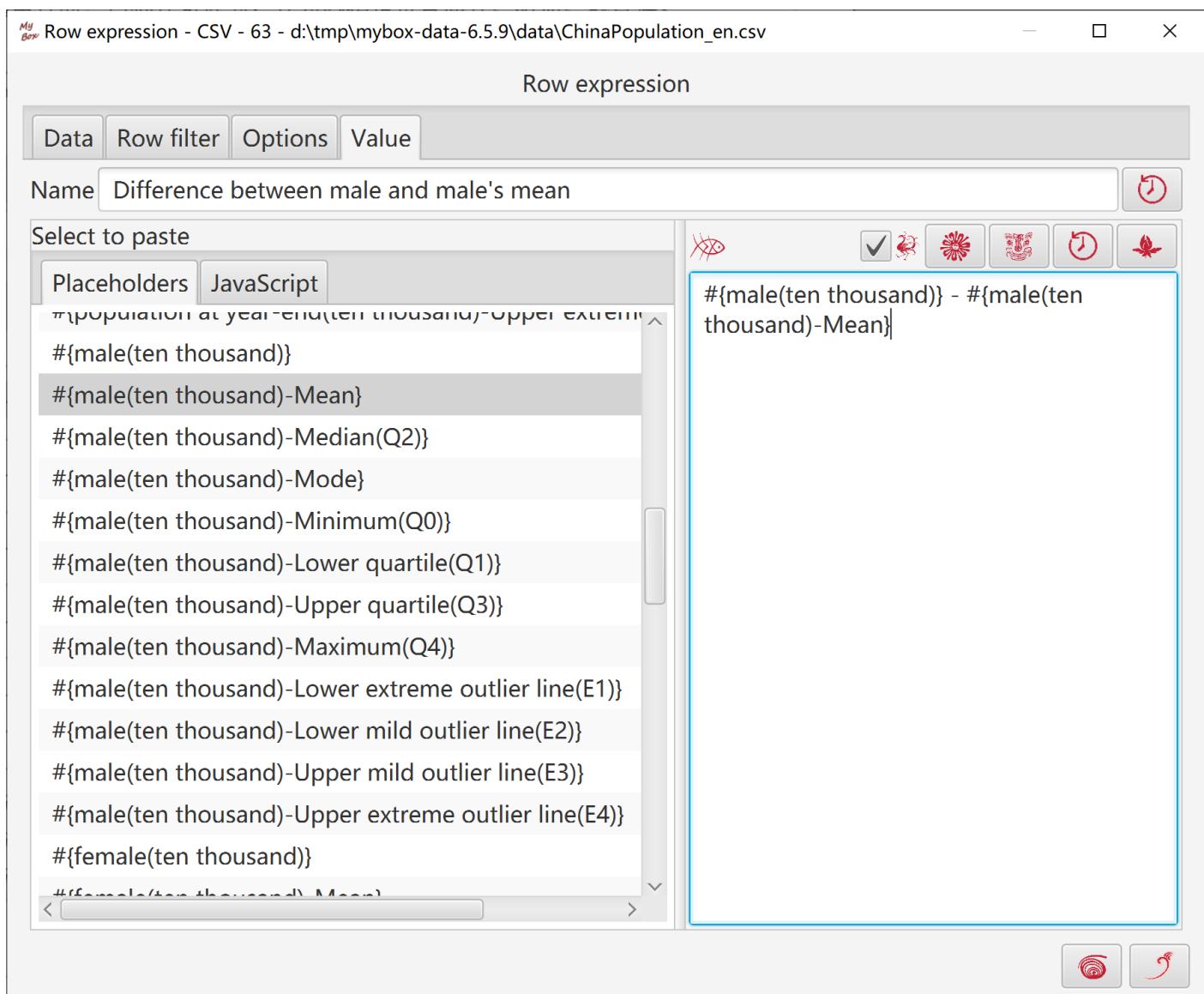
Ta...	Data ...	SepalLeng...	SepalWidt...	PetalLeng...	Peta...
1	1	5.1	3.5	1.4	0.2
2	2	4.9	3.0	1.4	0.2
3	3	4.7	3.2	1.3	0.2
4	4	4.6	3.1	1.5	0.2
5	5	4.5	2.3	1.3	0.1
6	6	4.5	2.3	1.3	0.1
7	7	4.5	2.3	1.3	0.1
8	8	4.4	2.9	1.4	0.2
9	9	4.3	3.0	1.3	0.1
10	10	4.3	3.0	1.3	0.1
11	11	4.3	3.0	1.3	0.1
12	12	4.3	3.0	1.3	0.1
13	13	4.3	3.0	1.3	0.1

Selected: 0 Rows: 50/150 Page size 50 Page 1 /3

- ✖ Data
- ✖ Modify
- ✖ Trim
- ✖ Calculation**
- ✖ Charts
- ✖ Examples
- Pop when mouse hovering
- ✖ Close(ESC/F6 Or click anywhere outside the object)

## 2.8.1 Row Expression

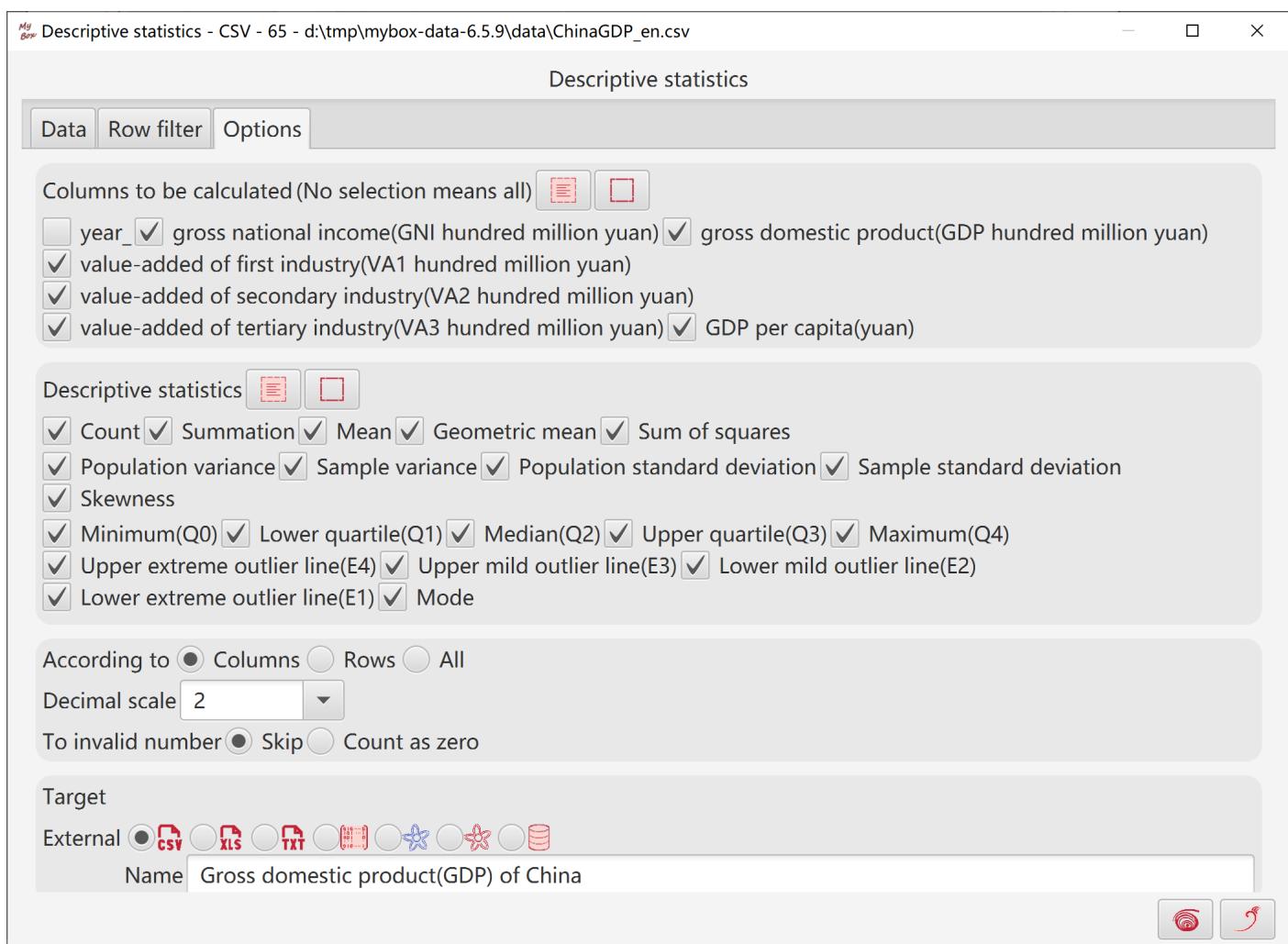
1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - Input name of values.
  - Input row expression.
  - Select columns to be copied.
3. Target can be following: new csv/excel/text file, matrix, system clipboard, MyBox clipboard, database table.
4. When rows are current page or selected ones, target can be defined location in table to insert/append/replace.



## 2.8.2 Descriptive Statistic

1. Select data:
    - Rows can be: current page, selected rows, or all pages.
    - Set row filter.
  2. Calculate:
    - Select columns to be calculated.
    - 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
    - According 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.



### 2.8.3 Group Equal Values

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - 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.

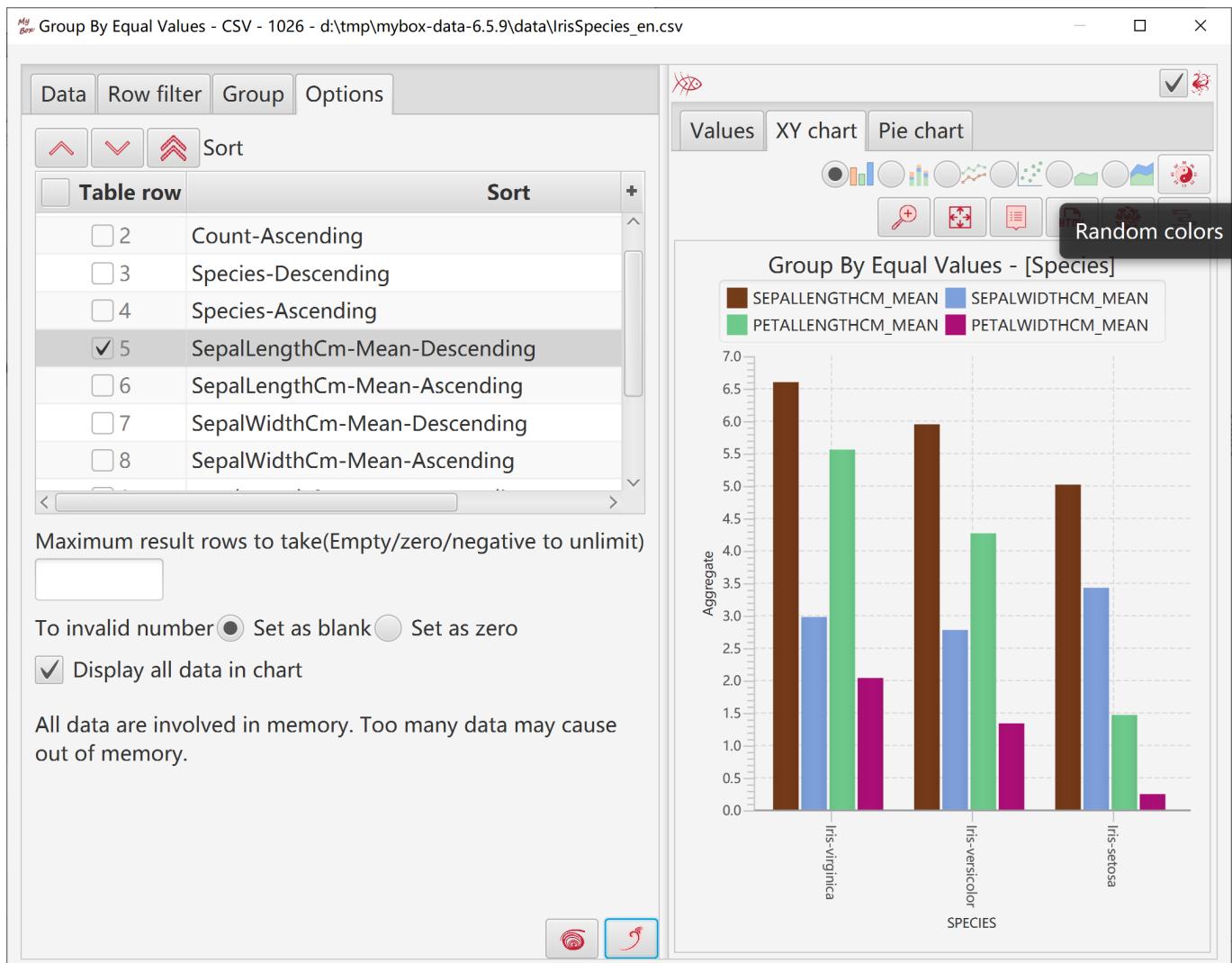
MyBox Group By Equal Values - CSV - 1026 - d:\tmp\mybox-data-6.5.9\data\irisSpecies\_en.csv

The screenshot shows the 'Values' tab of the MyBox Group By Equal Values tool. The interface is divided into several sections:

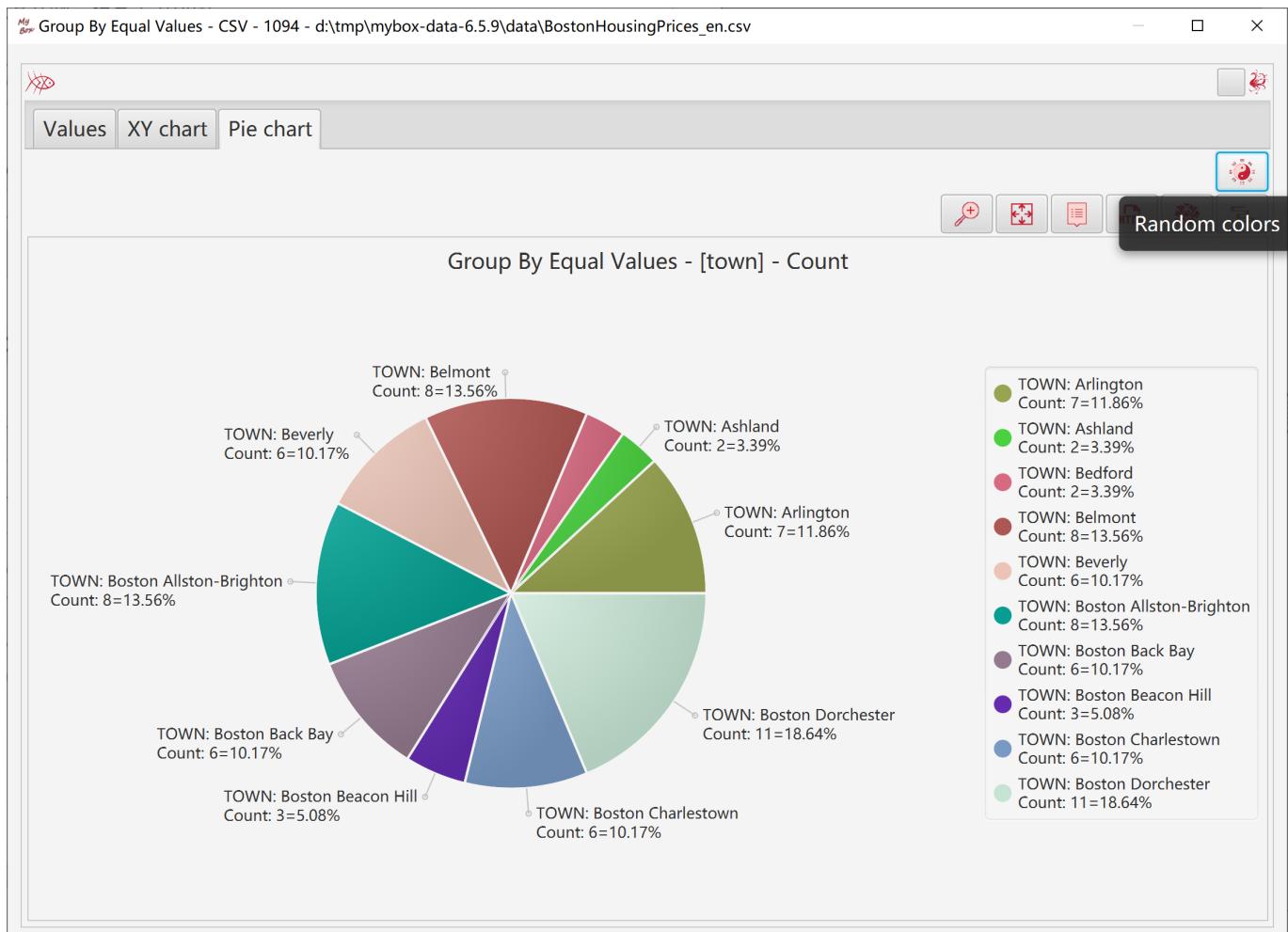
- Top Bar:** Buttons for Data, Row filter, Group, Options, and a search bar.
- Group by Section:** A table where 'Species' is selected as the group column.
- Aggregate Section:** A table showing various statistical calculations for SepalLengthCm.
- Result Table:** A main table displaying grouped data with columns: Data row, Group, SPECIES, COUNT, SEPALLE..., and SI+.
- Bottom Bar:** Buttons for navigation, search, and page settings (Selected: 0, Rows: 3/3, Page size: 20, Page).

Data row	Group	SPECIES	COUNT	SEPALLE...	SI+
1	Group1	Iris-setosa	50	5.0059999...	3.418
2	Group2	Iris-versic...	50	5.936	2.770
3	Group3	Iris-virgini...	50	6.5879999...	2.973

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



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



- 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.
- 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.
- Data overflow may happen.

## 2.8.4 Simple Linear Regression

### 2.8.4.1 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.
- Set row filter.

#### 2. Calculate:

- 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.

N...	income	happiness	Slope(b1)	Intercept(b0)	Coeffici...	Pearson +
1	3.8626	2.3145	NaN	NaN	NaN	NaN
2	4.6399	3.7379	1.8314	-4.7594	1	1
3	2.1347	0.2687	1.3514	-2.6845	0.9874	0.9937
4	6.5013	4.3748	0.9625	-1.45	0.9171	0.9577
5	3.6512	2.1558	0.9579	-1.4127	0.9183	0.9583
6	2.2865	1.8936	0.8249	-0.715	0.8447	0.9191
7	4.7489	4.903	0.9204	-0.8517	0.7453	0.8633
8	5.4592	4.8335	0.975	-0.9964	0.7754	0.8806

### 2.8.4.2 Model

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

Simple linear regression - CSV - 784 - d:\tmp\mybox-data-6.5.9\data\IncomeHappiness\_en.csv

**About linear regression**  Decimal scale 4

Model Regression Fitting Residual

Linear model: happiness =  $0.2222 + 0.7097 * \text{income}$

Independent variable: income =  Predict

Dependent variable: happiness =

Last status	
Name	Value
Row number	477
Number of observations	477
income	4.4981
happiness	1.9071
Slope(b1)	0.7097
Intercept(b0)	0.2222
Coefficient of determination(R-Square)	0.7517
Pearson's product moment correlation coefficient(R)	0.867
Mean of squared error(MSE)	0.5057

DIV

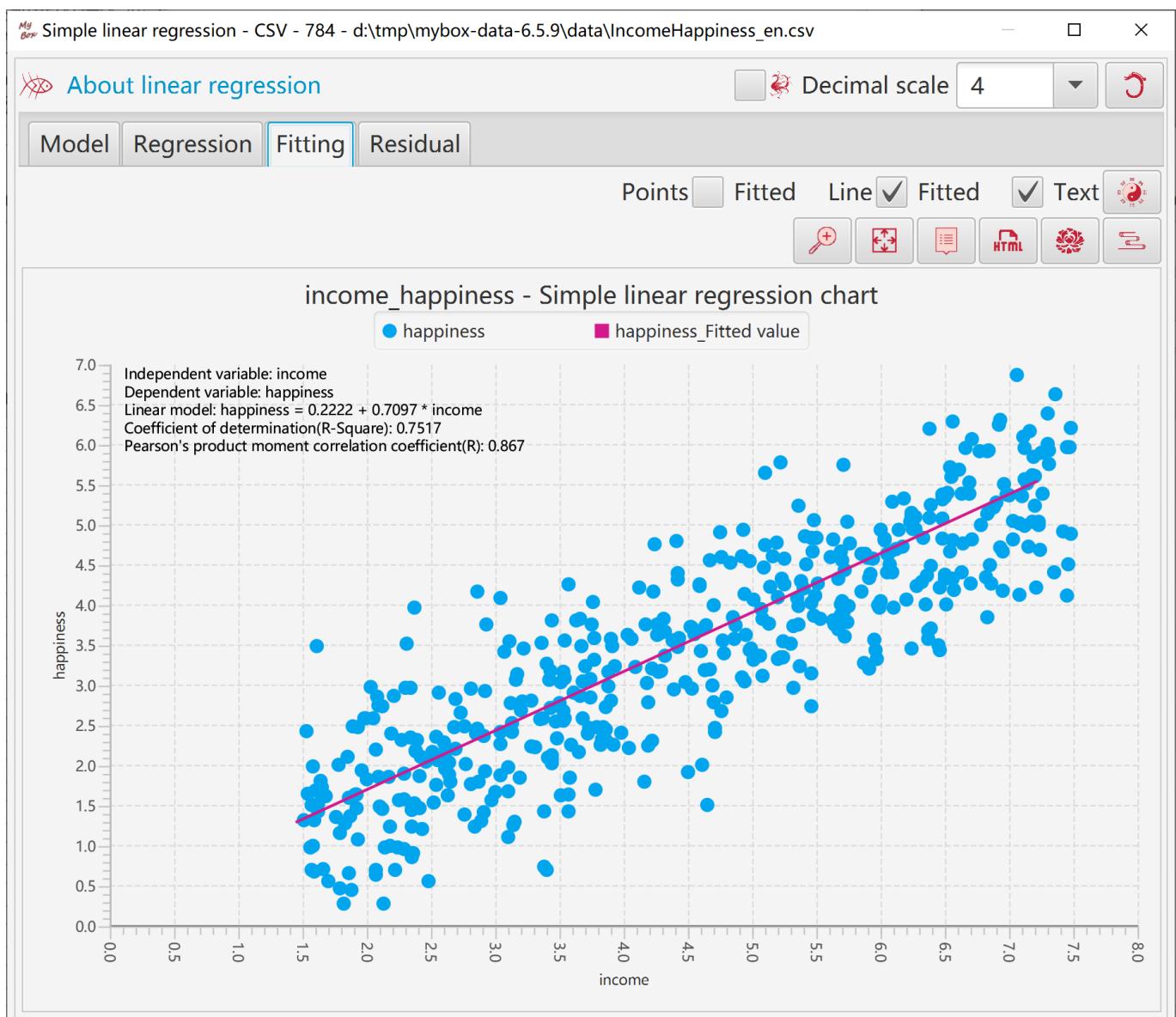
### 2.8.4.3 Fitted Chart

- 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.

- Set parameters of plot, X axis and Y axis.
- Options to display fitted points, fitted line, or model description in chart.
- Options to display data labels.
- Set random colors to fitted points/line.
- Fitted chart can be popped.
- Html including fitted chart and its data can be created.
- Display fitted chart's data in table.



#### 2.8.4.4 Residual Chart

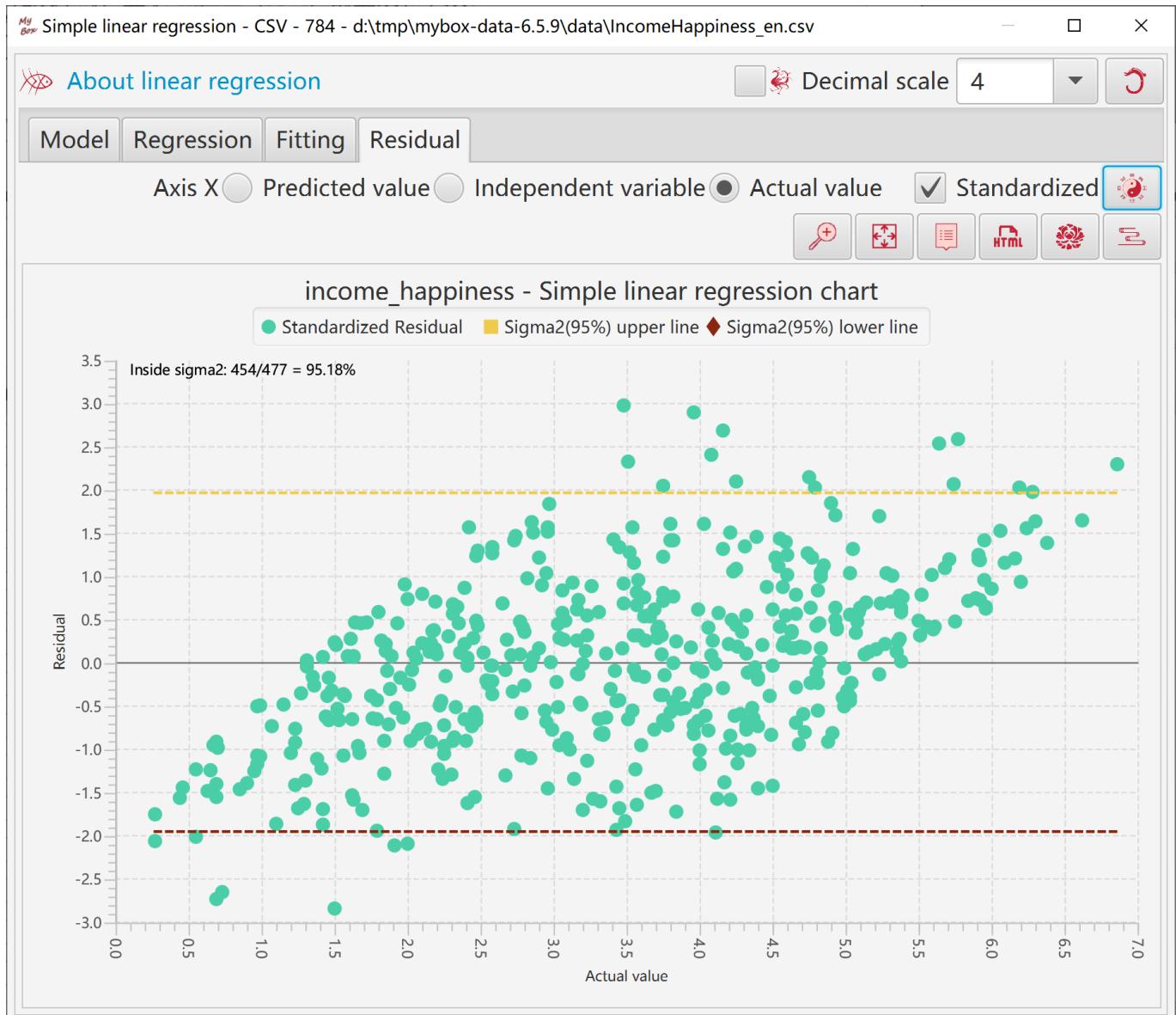
1. X axis can be set as: predicted value, independent variable, or actual value.

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.



## 2.8.5 Simple Linear Regression - Combination

This tool helps to generate data of simple linear regression:

1. Select some columns as candidates of independent variables.
2. Select some columns as candidates of dependent variables.
3. Select options like decimal scale, alpha, whether include intercept.
4. When click button "OK":
  - Make pairs from candidates : one as independent variable and another as dependent variable.
  - Calculate the regression models.
  - Sort the modes by their coefficient of determination(R-Square) in descending order.
5. Select one model and click button "View" to view its regression data, fitting chart, and residual chart.

Simple linear regression - Combination - CSV - 60 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_standardized\_en.csv

	Data ...	Dependen...	Independ...	Coeffi...▼	Pearson'...	
3	disease pr...	BMI(body ...	0.34392...	0.58645013	disease	
9	disease pr...	S5(blood s...	0.32022...	0.56588343	disease	
4	disease pr...	BP(averag...	0.19490...	0.44148385	disease	
8	disease pr...	S4(blood s...	0.18528...	0.43045288	disease	
7	disease pr...	S3(blood s...	0.15585...	0.39478925	disease	
10	disease pr...	S6(blood s...	0.14629...	0.38248349	disease	
5	disease pr...	S1(blood s...	0.04495...	0.21202248	disease	
1	disease pr...	age	0.03530...	0.18788875	disease	
6	disease pr...	S2(blood s...	0.03029...	0.17405358	disease	
2	disease pr...	sex	0.00185...	0.043062	disease	

**Data** **Filter** **Options**

Independent variable (No selection means all)

age  sex  BMI(body mass index)  
 BP(average blood pressure)  
 S1(blood serum measurement 1)  
 S2(blood serum measurement 2)  
 S3(blood serum measurement 3)  
 S4(blood serum measurement 4)  
 S5(blood serum measurement 5)  
 S6(blood serum measurement 6)  
 disease progression one year after baseline

Dependent variable (No selection means all)

age  sex  BMI(body mass index)  
 BP(average blood pressure)  
 S1(blood serum measurement 1)  
 S2(blood serum measurement 2)  
 S3(blood serum measurement 3)  
 S4(blood serum measurement 4)  
 S5(blood serum measurement 5)  
 S6(blood serum measurement 6)  
 disease progression one year after baseline

Decimal scale

Desired significance level(alpha)

Intercept( $b_0$ )

Double click selected item to view

## 2.8.6 Multiple Linear Regression

### 2.8.6.1 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-Square, adjusted R-Square.

Name	Value
Dependent variable	disease progression one year after baseline
Independent variable	[age, sex, BMI(body mass index), BP(average blood pressure), S1(blood serum measurement 1), S2(blood serum measurement 2), S3(blood serum measurement 3), S4(blood serum measurement 4), S5(blood serum measurement 5), S6(blood serum measurement 6)]
Number of observations	442
Intercept( $b_0$ )	0.0
Coefficients	$[-0.006, -0.148, 0.321, 0.2, -0.489, 0.294, -0.062, 0.109, 0.464, 0.042]$
Coefficient of determination(R-Square)	0.518
Adjusted R squared	0.507
Standard Error	0.702
Variance	1.0

## 2.8.6.2 Model

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

MyBox Multiple linear regression - CSV - 60 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_standardized\_en.csv

**Data Filter Options**

**Independent variable (No selection)**

- age  sex  BMI(body mass index)
- BP(average blood pressure)
- S1(blood serum measurement 1)
- S2(blood serum measurement 2)
- S3(blood serum measurement 3)
- S4(blood serum measurement 4)
- S5(blood serum measurement 5)
- S6(blood serum measurement 6)
- disease progression one year after baseline

**Dependent variable** disease progression one year after baseline

Normalize  L2(Z-Score)  Ordinary Least Squares

Intercept(b0)

All data are involved in memory. This may cause out of memory.

**About data analysis**

**Model Regression**

Decimal scale 3

Linear model: disease progression one year after baseline =  $0.0 - 0.006 * \text{age} - 0.148 * \text{sex} + 0.321 * \text{BMI}(\text{body mass index}) + 0.2 * \text{BP}(\text{average blood pressure}) - 0.489 * \text{S1}(\text{blood serum measurement 1}) + 0.294 * \text{S2}(\text{blood serum measurement 2}) - 0.062 * \text{S3}(\text{blood serum measurement 3}) + 0.109 * \text{S4}(\text{blood serum measurement 4}) + 0.464 * \text{S5}(\text{blood serum measurement 5}) + 0.042 * \text{S6}(\text{blood serum measurement 6})$

Independent variable: age =

Independent variable: sex =

Independent variable: BMI(body mass index) =

Independent variable: BP(average blood pressure) =

Independent variable: S1(blood serum measurement 1) =

Independent variable: S2(blood serum measurement 2) =

Independent variable: S3(blood serum measurement 3) =

Independent variable: S4(blood serum measurement 4) =

Independent variable: S5(blood serum measurement 5) =

Independent variable: S6(blood serum measurement 6) =

Dependent variable: disease progression one year after baseline =

**Buttons:**  

## 2.8.7 Multiple Linear Regression - Combination

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

1. Select some columns as candidates of independent variables.
2. Select some columns as candidates of dependent variables.
3. Select whether include intercept.
4. When click button "OK", the tool does following:
  - Make combination of candidates as independent variables and dependent variable .
  - Calculate the regression models.
  - Sort the modes by their adjusted coefficient of determination(R-Square) in descending order.
5. Select one model and click button "View" to view its regression data

Multipe linear regression - Combination - CSV - 60 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_standardized\_en.csv

Data ...	Dependen...	Independent variable	Adjus...	Coeffici...	Coeffici...
19	disease pr...	[sex, BMI(body mass index)]	0.50767...	0.51771...	[-0.1487...
18	disease pr...	[sex, BMI(body mass index)]	0.50764...	0.51657...	[-0.1454...
10	disease pr...	[age, sex, BMI(body mass index)]	0.50656...	0.51774...	[-0.0061...
9	disease pr...	[age, sex, BMI(body mass index)]	0.50650...	0.51657...	[-0.0012...
26	disease pr...	[BMI(body mass index), sex]	0.49165...	0.499727...	[0.3483...
27	disease pr...	[BMI(body mass index), sex]	0.49099...	0.50023...	[0.3445...
16	disease pr...	[sex, BMI(body mass index)]	0.48634...	0.49332...	[-0.1417...
17	disease pr...	[sex, BMI(body mass index)]	0.48585...	0.49401...	[-0.1442...
7	disease pr...	[age, sex, BMI(body mass index)]	0.48525...	0.49342...	[0.0106...
8	disease pr...	[age, sex, BMI(body mass index)]	0.48477...	0.49411...	[0.0109...
24	disease pr...	[BMI(body mass index), sex]	0.471218...	0.47721...	[0.3761...
25	disease pr...	[BMI(body mass index), sex]	0.47023...	0.47744...	[0.3768...
34	disease pr...	[BP(average blood pressure), sex]	0.41056...	0.41991...	[0.2384...
33	disease pr...	[BP(average blood pressure), sex]	0.40788...	0.41594...	[0.2563...
15	disease pr...	[sex, BMI(body mass index)]	0.39582...	0.40267...	[-0.0494...
13	disease pr...	[sex, BMI(body mass index)]	0.39568...	0.39979...	[-0.0635...
14	disease pr...	[sex, BMI(body mass index)]	0.39519...	0.40068...	[-0.0628...
23	disease pr...	[BMI(body mass index), sex]	0.39507...	0.40056...	[0.4903...
4	disease pr...	[age, sex, BMI(body mass index)]	0.39477...	0.40026...	[0.0230...
6	disease pr...	[age, sex, BMI(body mass index)]	0.39463...	0.40287...	[0.0152...
5	disease pr...	[age, sex, BMI(body mass index)]	0.39407...	0.40094...	[0.0175...
21	disease pr...	[BMI(body mass index), sex]	0.39324...	0.39599...	[0.4882...
22	disease pr...	[BMI(body mass index), sex]	0.39283...	0.39696...	[0.4822...
31	disease pr...	[BP(average blood pressure), sex]	0.37119...	0.37689...	[0.2868...

Double click selected item to view

## 2.8.8 Frequency Distributions

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - 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.

The screenshot shows the 'Frequency Distributions' tool window. On the left, there is a table with three columns: 'town', 'town\_Count', and 'town\_Count percentage'. The data is as follows:

town	town_Count	town_Count percentage
Beverly	2	3.33
Danvers	4	6.67
Hamilton	1	1.67
Lynn	22	36.67
Lynnfield	2	3.33
Marblehead	3	5
Middleton	1	1.67
Nahant	1	1.67
Peabody	9	15
Salem	7	11.67
Sargus	4	6.67
Swampsc...	2	3.33
Topsfield	1	1.67
Wenham	1	1.67

The right side of the window contains the configuration interface:

- Frequency Distributions** title bar.
- Data**, **Row filter**, and **Options** tabs. The **Options** tab is selected.
- Column**: **town** dropdown.
- Case-insensitive**: checked checkbox.
- Decimal scale**: **2** dropdown.
- Target** section:
  - External**: **CSV** radio button is selected.
  - Name**: **BostonHousingPrices\_en.csv** text input field.
- In table**, **Insert above**, **Append below**, and **Replace** radio buttons.

## 2.8.9 Values Percentage

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - Select columns to be calculated.
  - According to: Columns/rows/all.
  - Select how to treat negative values: zero or absolute value.
  - Set decimal scale.
  - To non-numeric, skip or count as zero.
  - Select columns to be copied.
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.

.6-DataTools-zh.odt - OpenOffice Writer

编辑(E) 视图(V) 插入(I) 格式(O) 表格(A) 工具(T) 窗口(W) 帮助(H)

File Document Image Data File Media Network Settings Recent Accessed Development Help

Columns\* Attributes View Edit\* Table\* Text\*

Source row number item year 2020 year 2020\_Percent... year 2019 year 2019\_Percent...

Source row number	item	year 2020	year 2020_Percent...	year 2019	year 2019_Percent...
Row1	food consumption per capita(kil...	141.2	36.2	130.1	35.12
Row2	cooking oil consumption per ca...	10.4	2.67	9.5	2.56
Row3	vegetables				
Row4	meat consu				
Row5	poultry con				
Row6	aquatic pro				
Row7	eggs consu				
Row8	milk consu				
Row9	fruits consu				
Row10	sugar consu				
Column-Summation					

Value percentage - CSV - 64 - d:\tmp\mybox-data-6.5.9\data\ChinaFoods\_en\_en.csv

Percentage

Data Row filter Options

Columns to be calculated (No selection means all)  item  year 2020  year 2019  year 2018  year 2017  year 2016  
 year 2015  year 2014  year 2013

According to  Columns  Rows  All

Decimal scale 2

To negative  Skip  Count as zero  Absolute value

To invalid number  Skip  Count as zero

Columns to be copied  item  year 2020  year 2019  year 2018  year 2017  year 2016  
 year 2015  year 2014  year 2013

## 2.9 Chart

Hover or click button “Function” to select functions under menu item “Chart”.

The screenshot shows a CSV file titled "Edit CSV File : CSV - 64 - d:\tmp\mybox-data-6.5.9\data\ChinaFoods\_en\_en.csv" in the MyBox interface. The file contains data about food consumption per capita in kilograms. A context menu is open over the 7th row, which is highlighted in blue. The menu path "Data Tools > Charts" is visible, and the "Self comparison bars chart" option is selected, highlighted with a blue background.

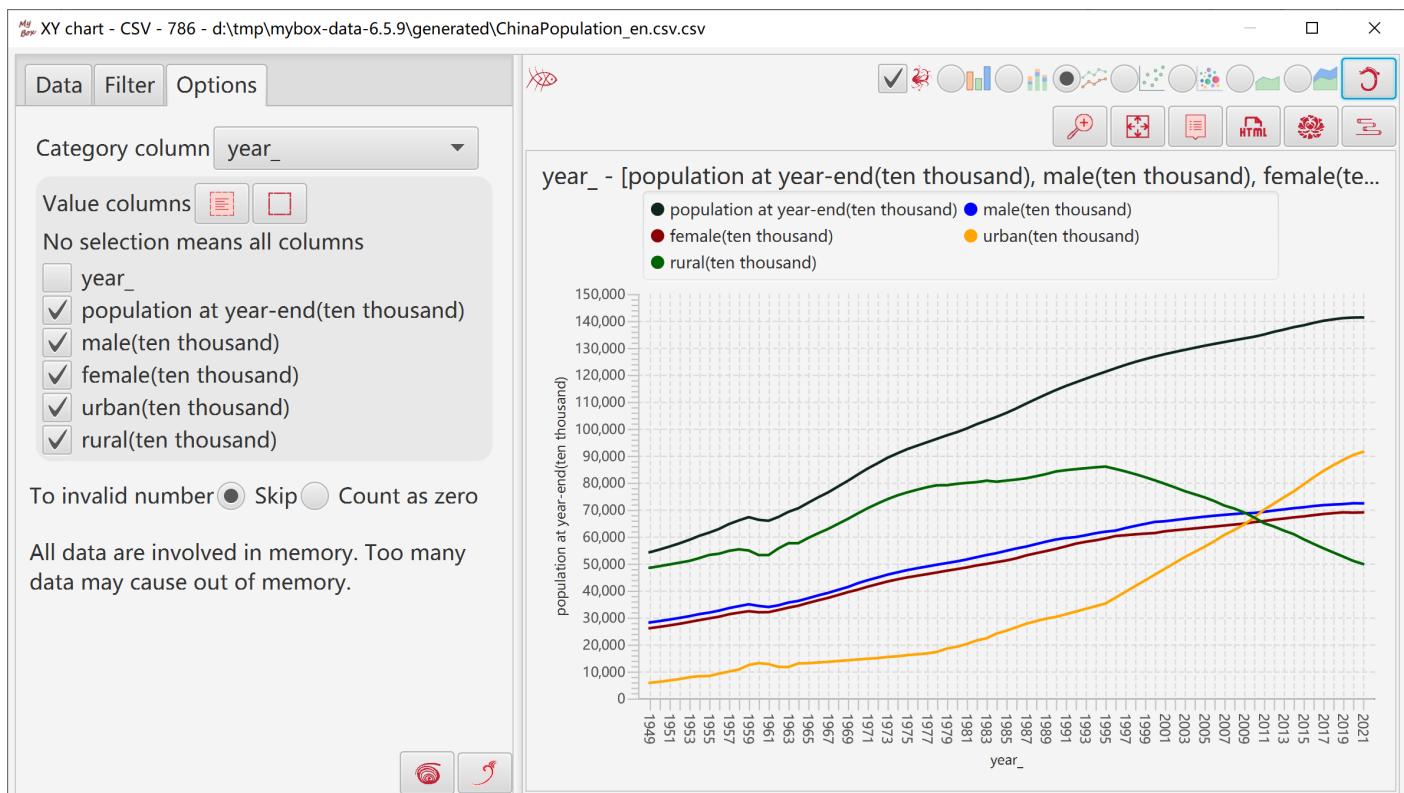
Ta...	Data ...	item	y
1	1	food consumption per capita(kilogram)	14
2	2	cooking oil consumption per capita(kilogram)	10
3	3	vegetables and tubers consumption per capita(kilogram)	10
4	4	meat consumption per capita(kilogram)	10
5	5	poultry consumption per capita(kilogram)	10
6	6	aquatic products consumption per capita(kilogram)	10
7	7	eggs consumption per capita(kilogram)	10
8	8	milk consumption per capita(kilogram)	10
9	9	fruits consumption per capita(kilogram)	10
10	10	sugar consumption per capita(kilogram)	10

Selected: 0 Rows: 10/10 Page size: 50 / Page: 1 / 1

## 2.9.1 XY Chart

### 2.9.1.1 Data

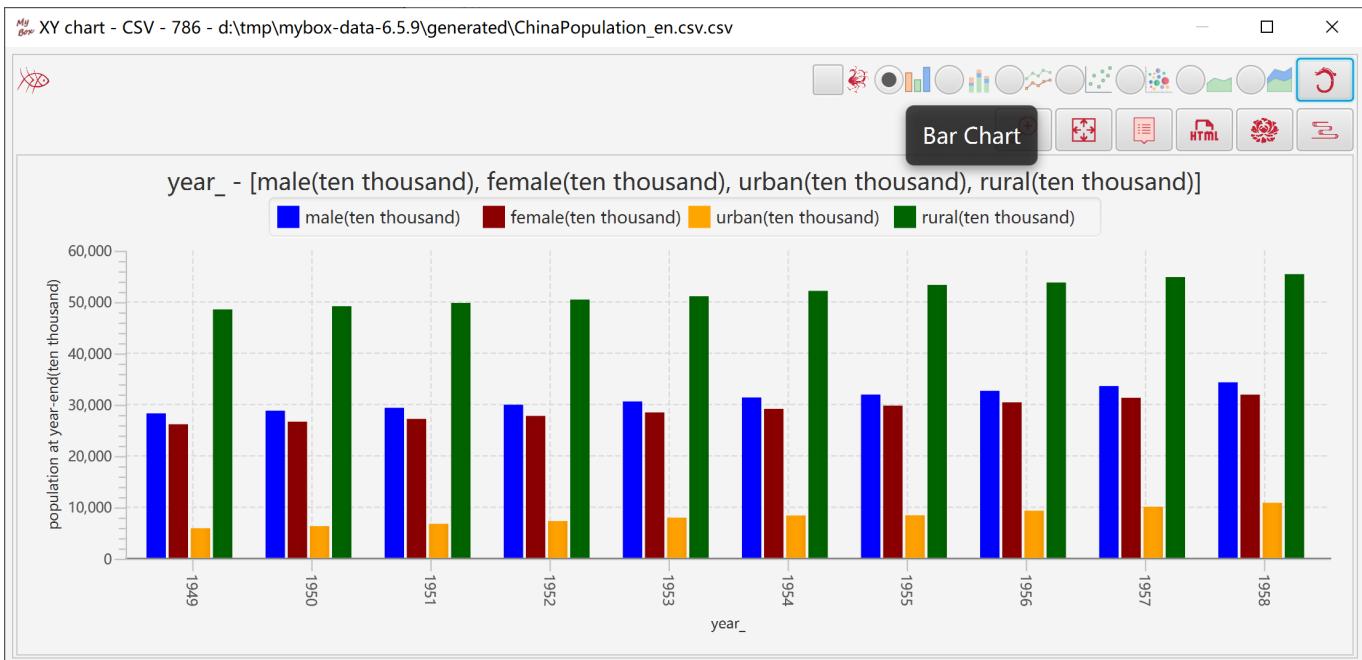
1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - Select one column as "Category Axis", to define data names..
  - Multiple columns can be selected as data in direction of "Number Axis". Different value series are shown in different colors 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.



### 2.9.1.2 Bar Chart

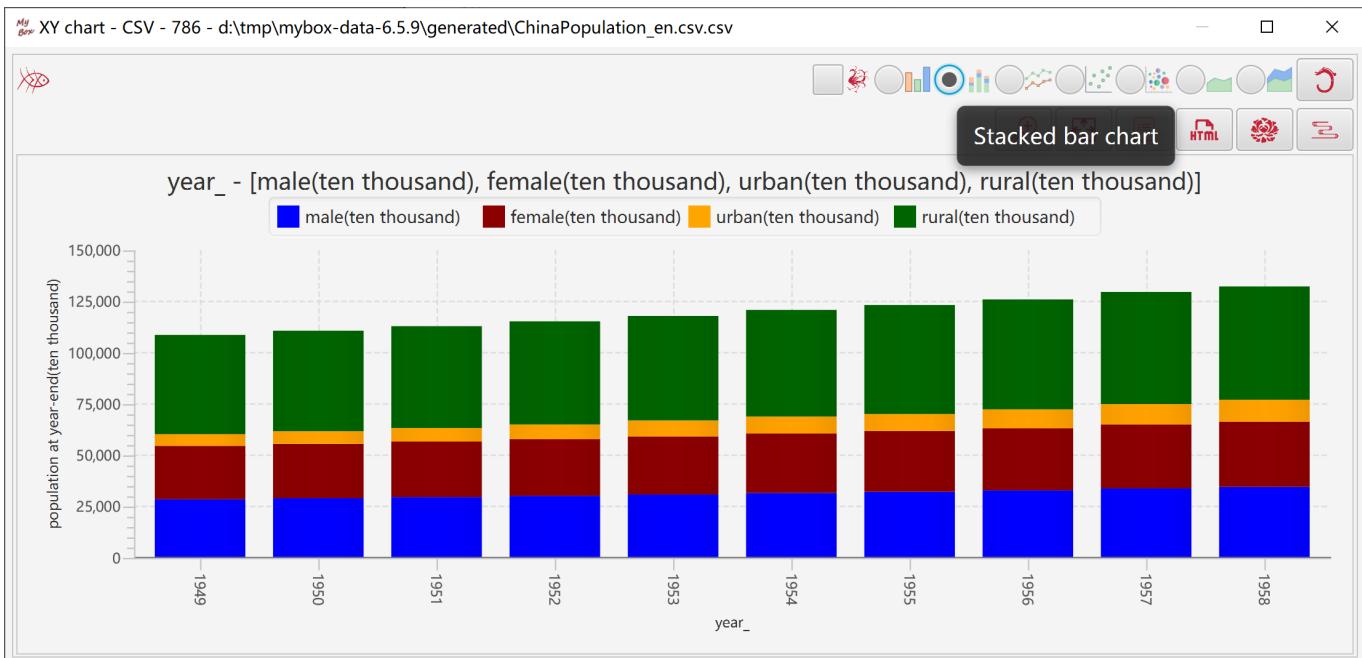
1. Represents data size with bars' heights.

- "Category" column is always counted as strings.



### 2.9.1.3 Stacked Bar Chart

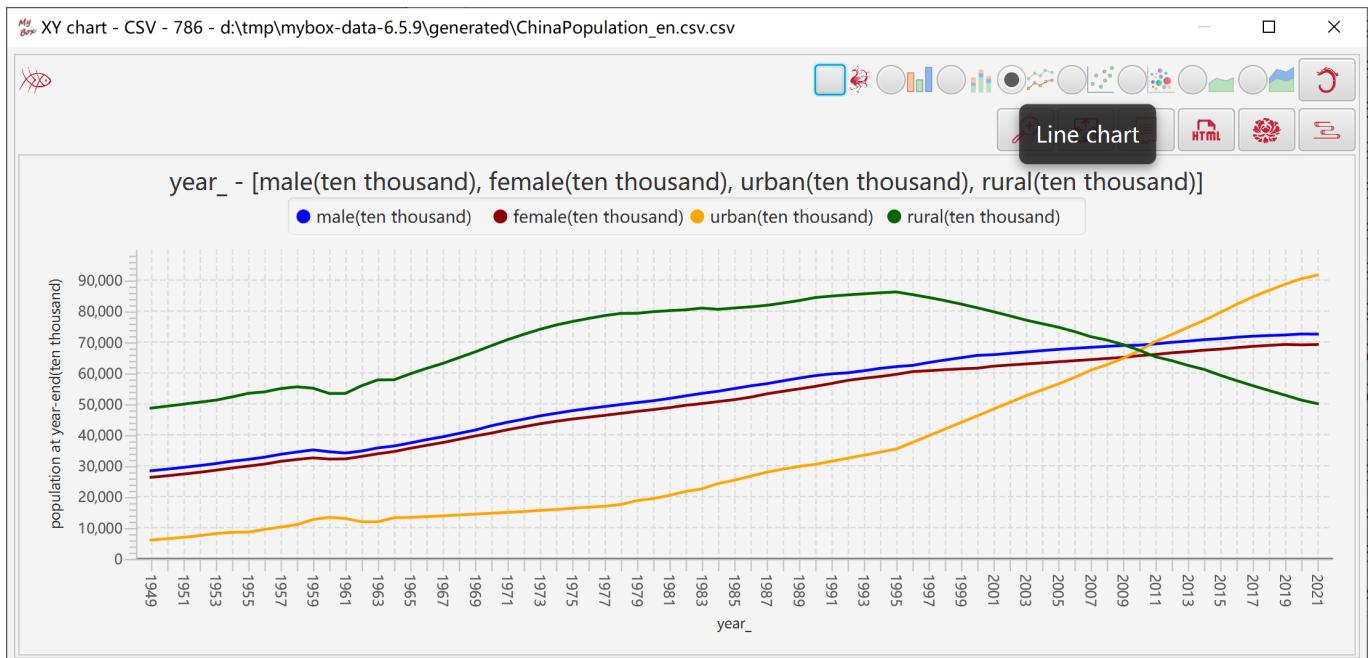
- Represents data size with bars' heights.
- "Category" column is always counted as strings.



### 2.9.1.4 Line Chart

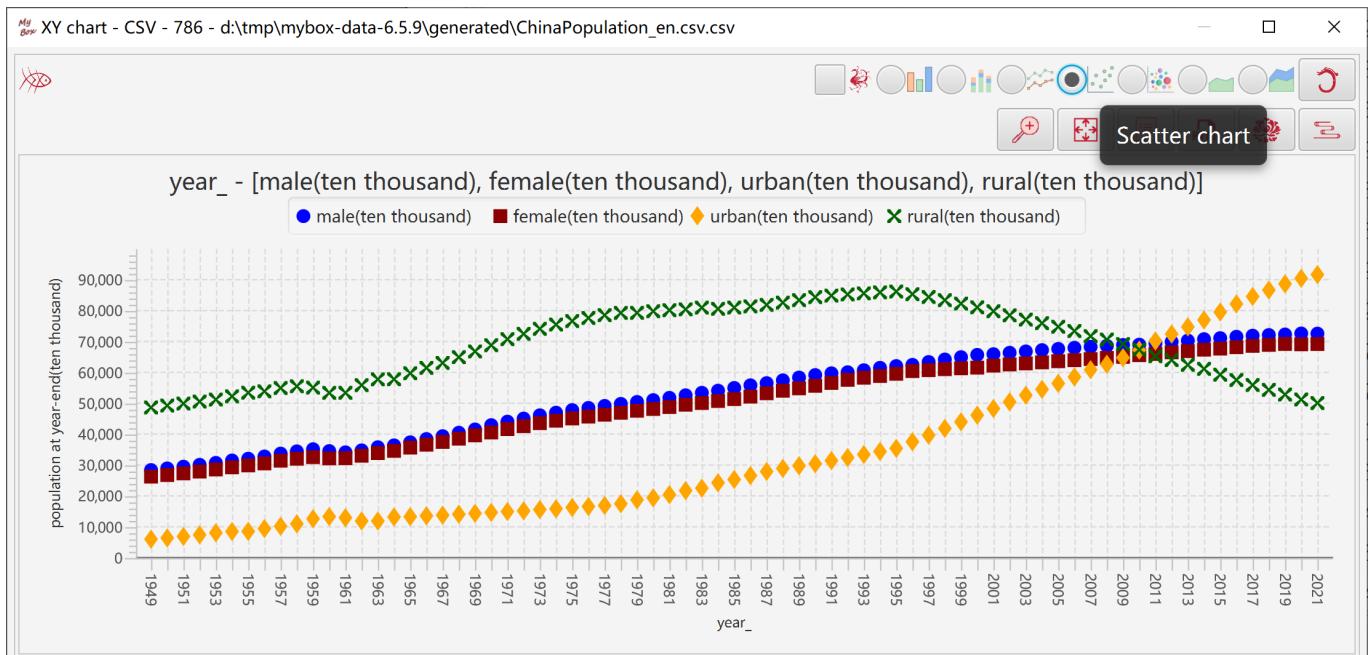
- Represents data trend with lines connecting points.

2. "Category" column can be counted as strings or numbers.



### 2.9.1.5 Scatter Chart

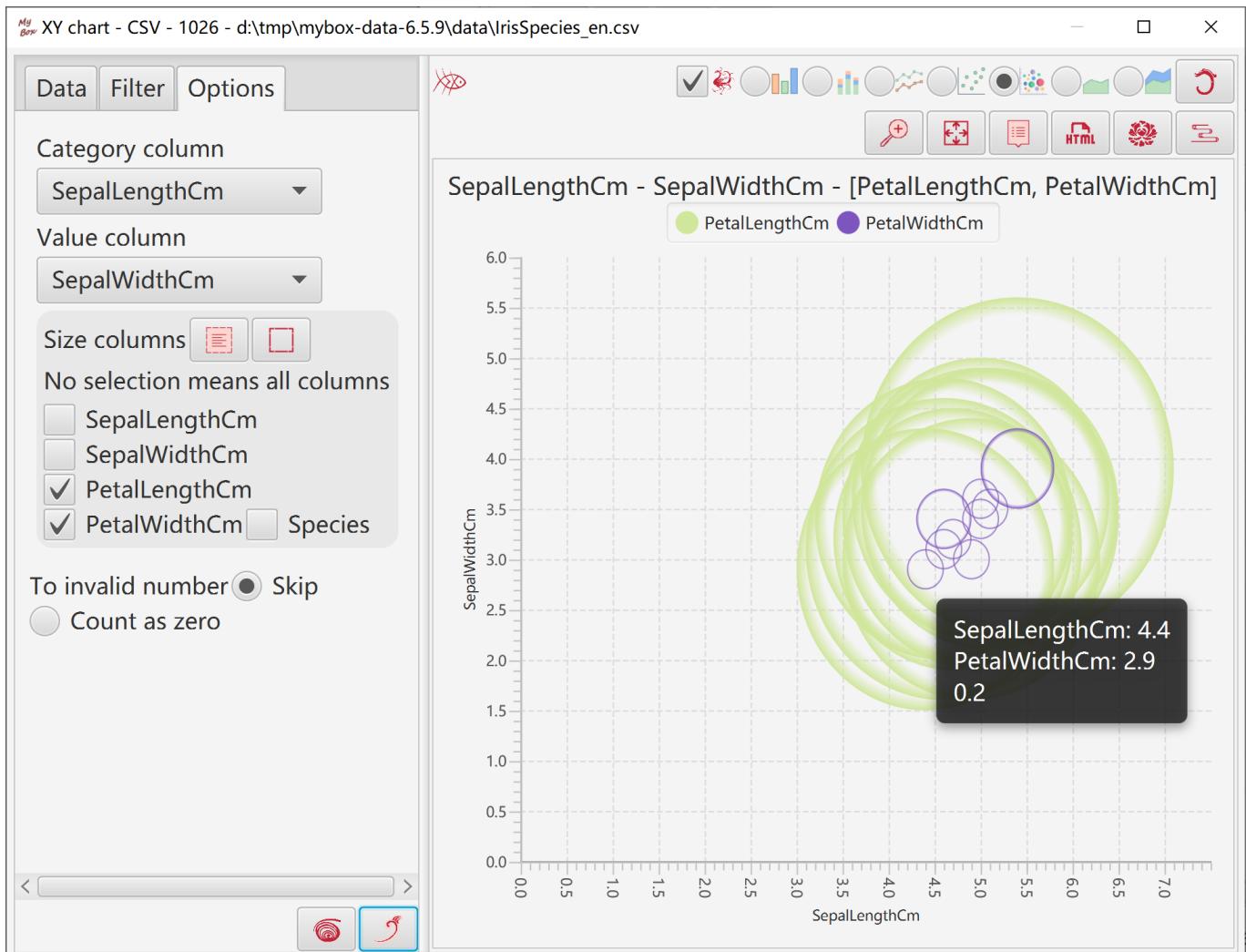
1. Represents data distribution with symbols.
2. "Category" column can be counted as strings or numbers.



### 2.9.1.6 Bubble Chart

1. Represents data size with circles of different radius:

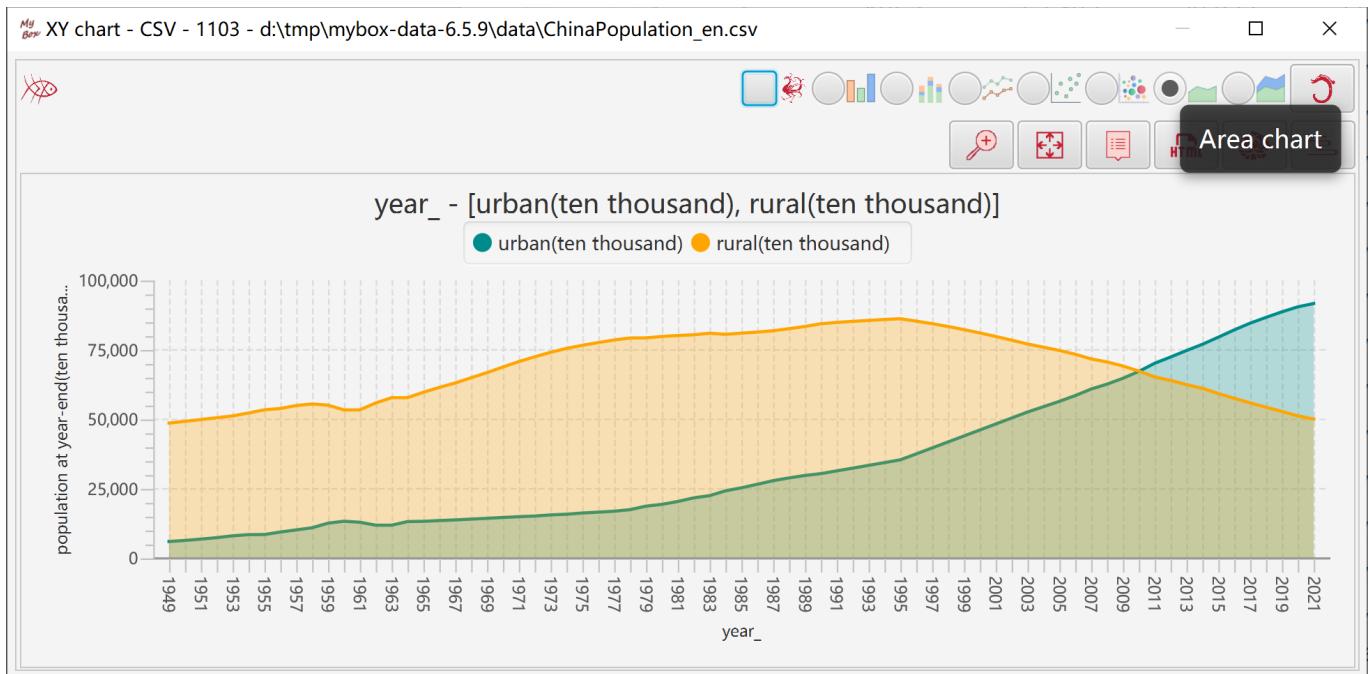
2. "Category Column" and "Value Column" define coordinates of data.
3. Select several columns as "Size Columns" to defines data size.
4. All columns should be numbers. Size columns should be non-negative.



### 2.9.1.7 Area Chart

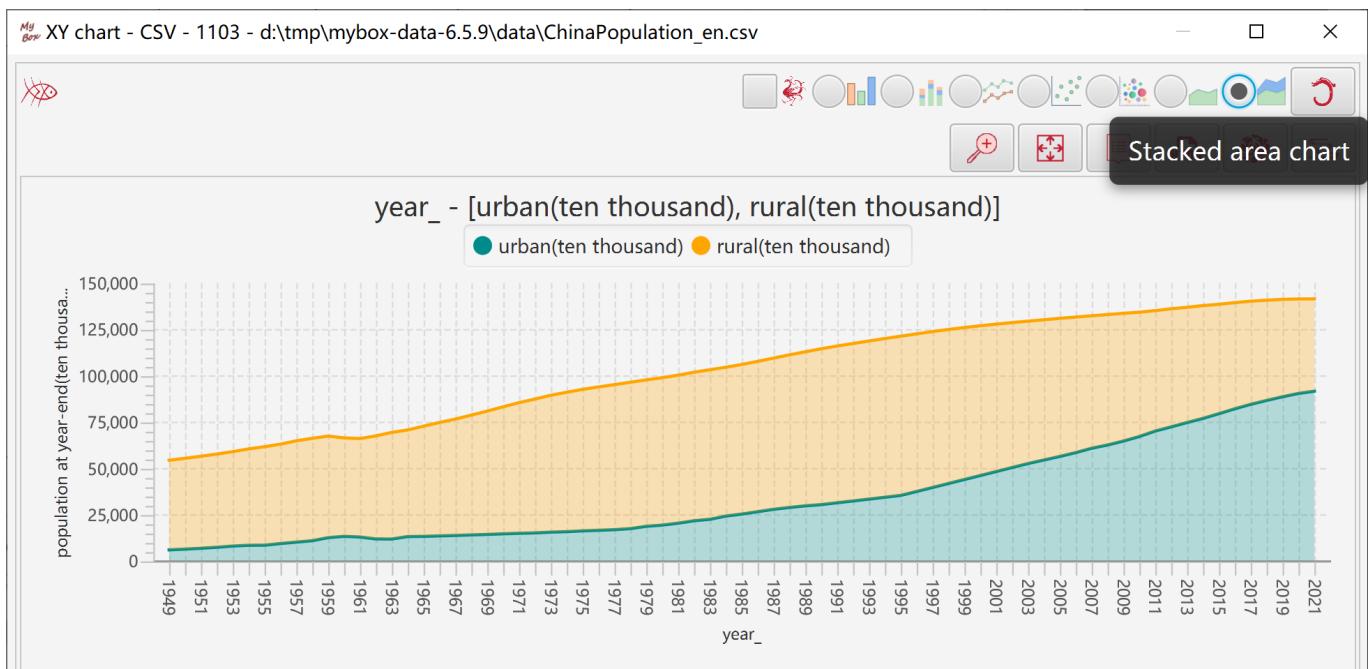
1. Represents data size with area size.

2. "Category" column is always counted as strings.



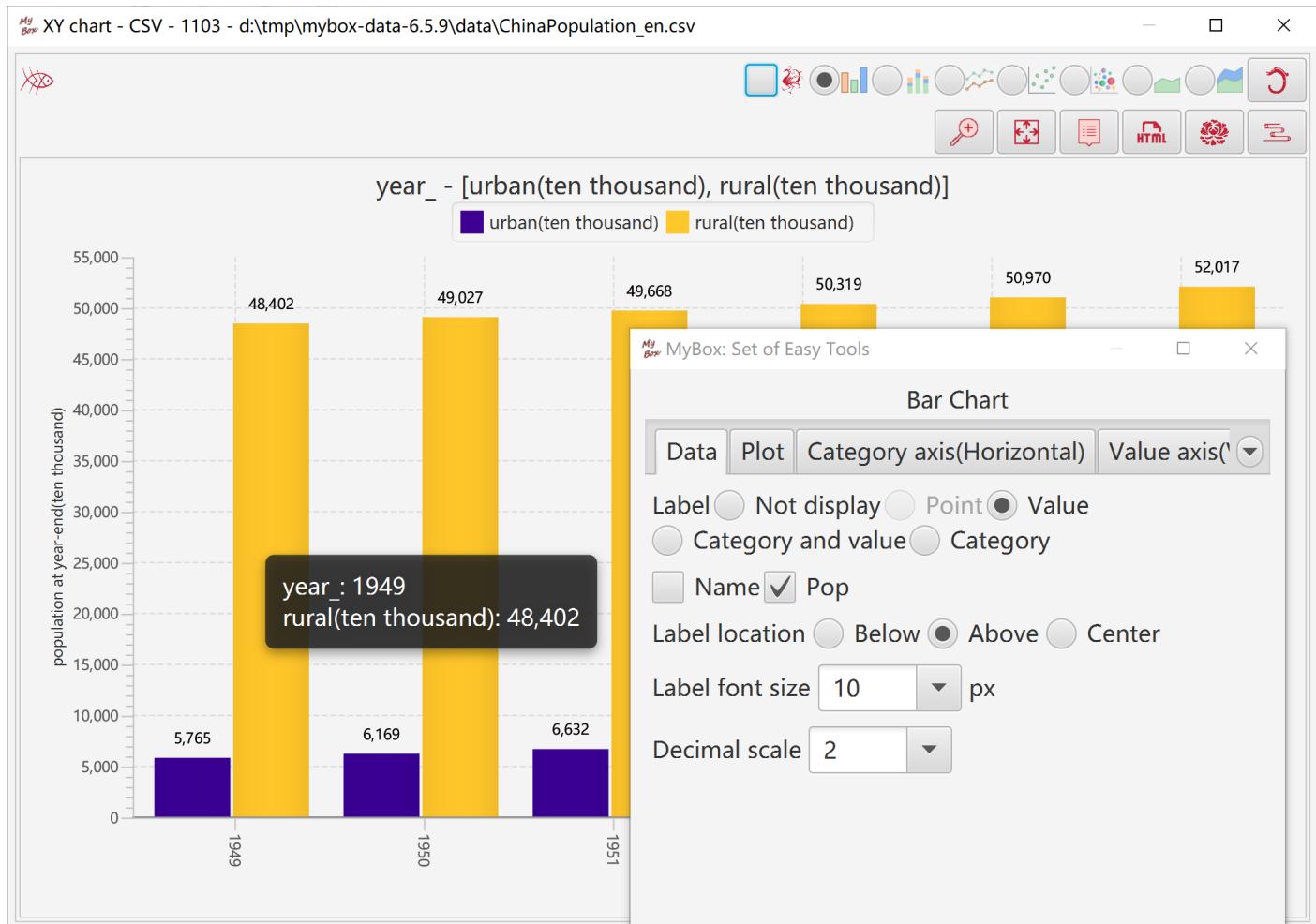
### 2.9.1.8 Stacked Area Chart

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



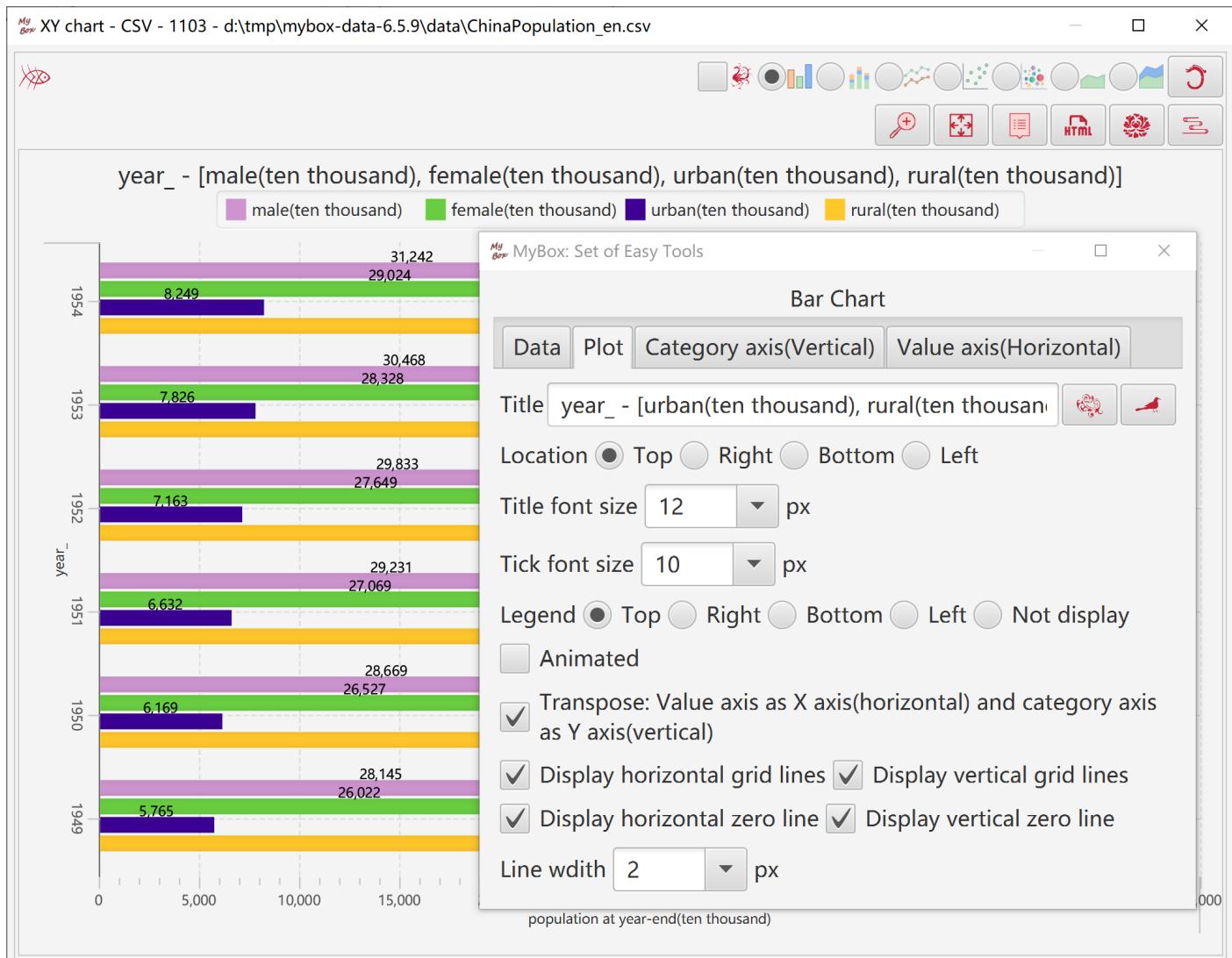
### 2.9.1.9 Parameters of Data in Chart

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



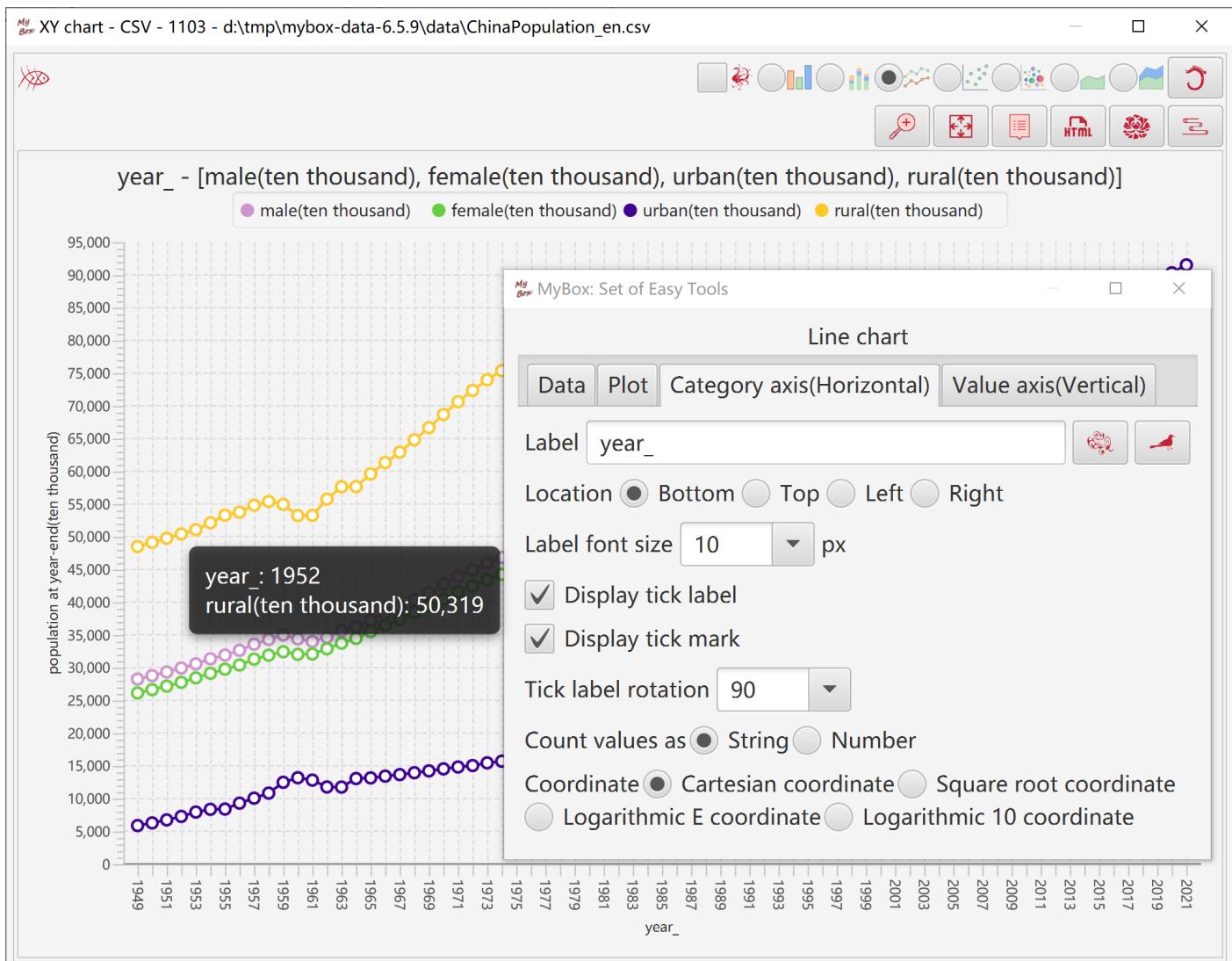
### 2.9.1.10 Layout

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



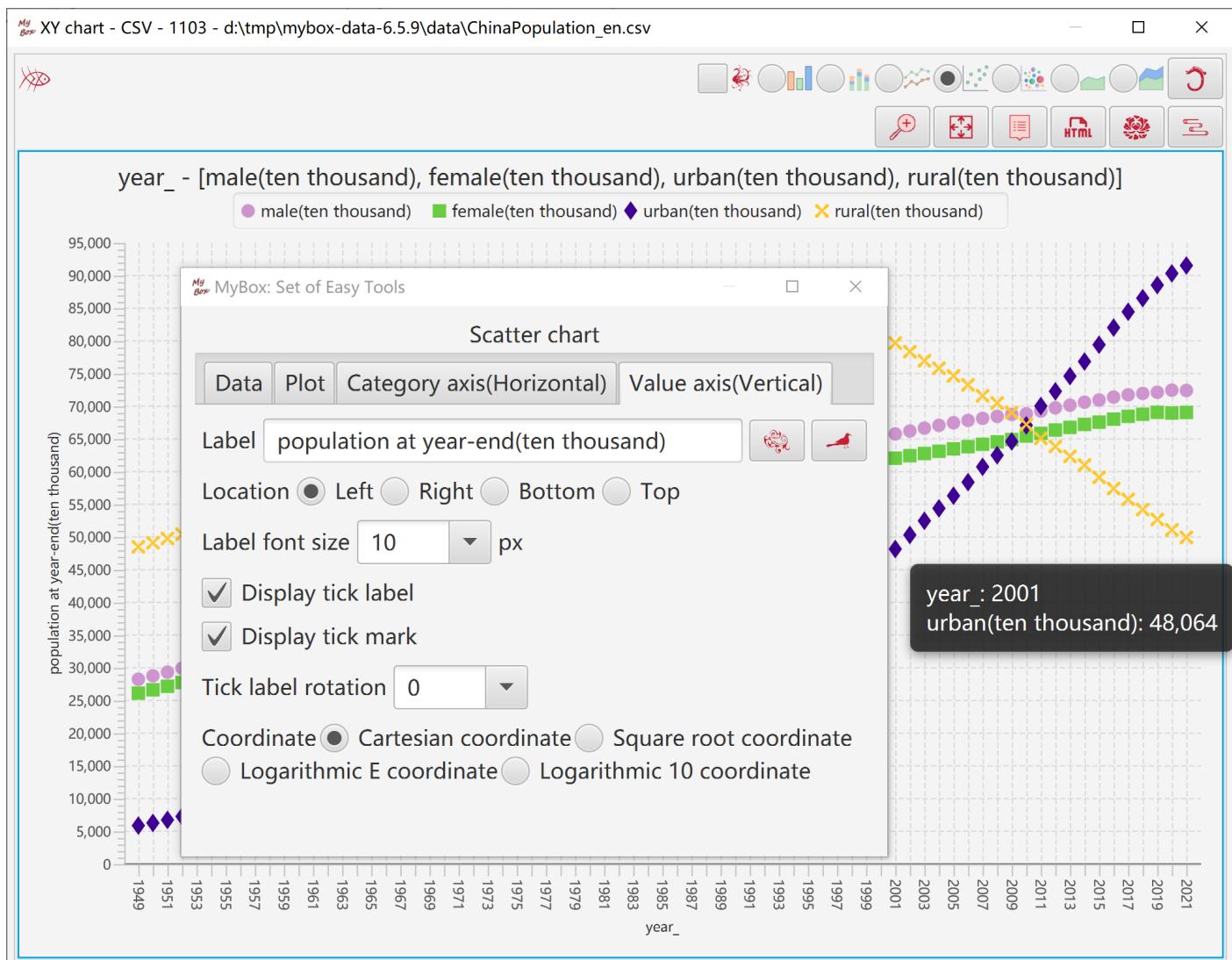
### 2.9.1.11 Category Axis

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



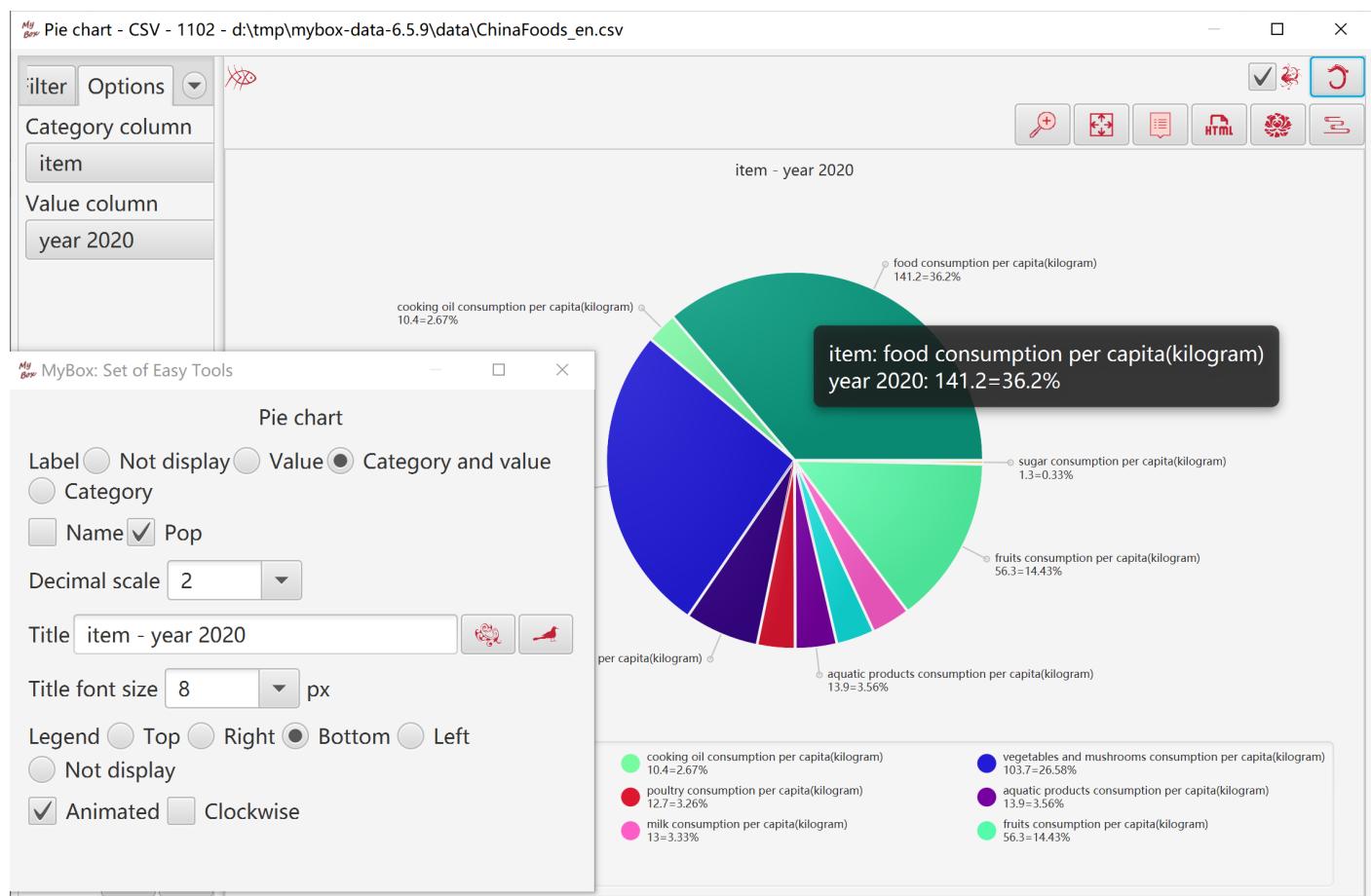
### 2.9.1.12 Number Axis

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



## 2.9.2 Pie Chart

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Calculate:
  - 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.



### 2.9.3 Box-and-whisker Chart

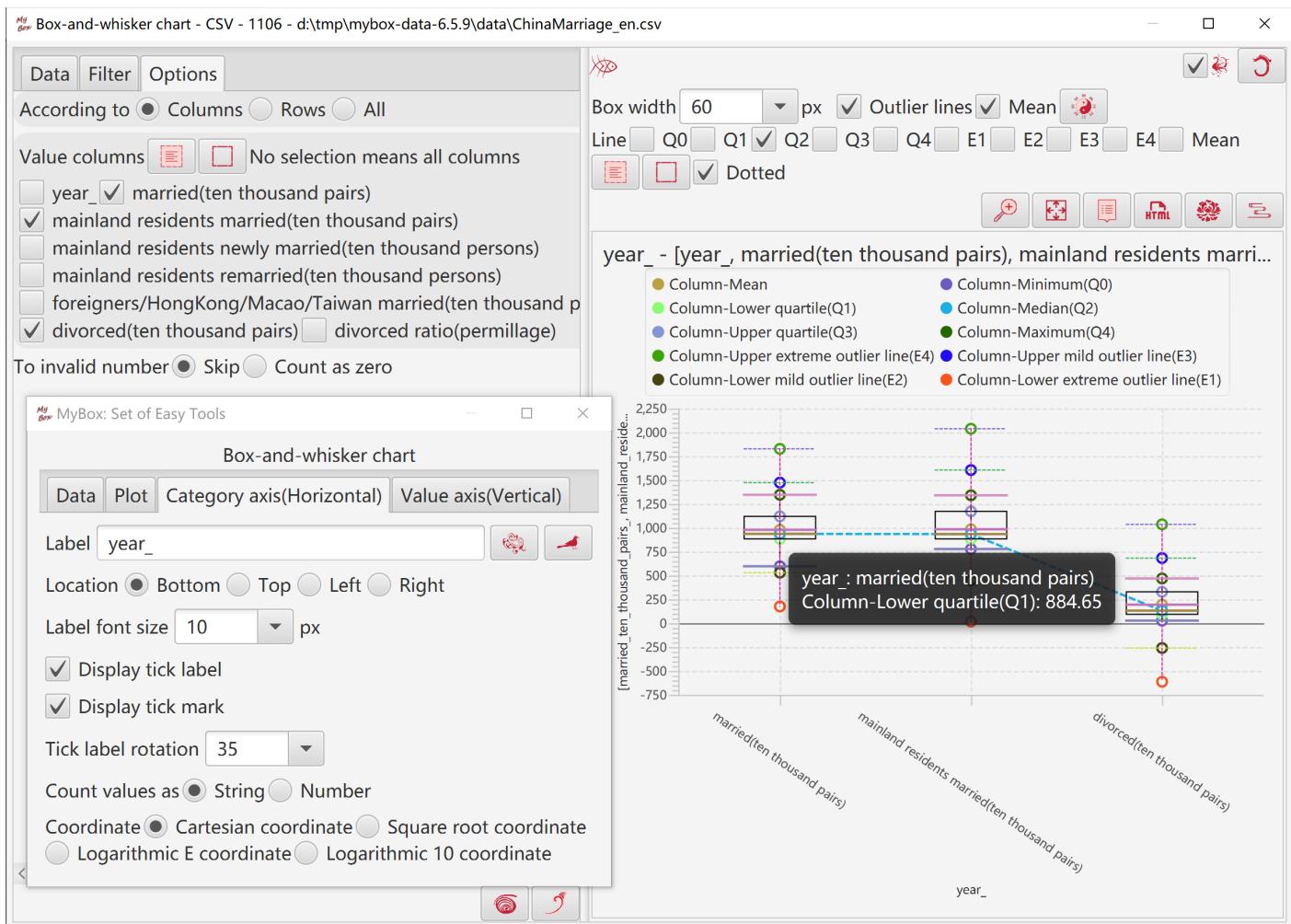
1. Box-and-whisker chart represents data distribution:
  - Sort data according to column/row/all in ascending order.
  - Following items can show aggregation and discreteness of data:

Minimum             $Q_0 = \text{in } 0\%(\text{start}) \text{ of the data list}$   
 Lower quartile     $Q_1 = \text{in } 25\% \text{ of the data list}$   
 Median             $Q_2 = \text{in } 50\%(\text{middle}) \text{ of the data list}$   
 Upper quartile     $Q_3 = \text{in } 75\% \text{ of the data list}$   
 Maximum            $Q_4 = \text{in } 100\%(\text{end}) \text{ of the data list}$

- Following values can be used to mark outliers of data:
  - Lower extreme outlier line     $E_1 = Q_1 - 3 * (Q_3 - Q_1)$
  - Lower mild outlier line        $E_2 = Q_1 - 1.5 * (Q_3 - Q_1)$
  - Upper mild outlier line       $E_3 = Q_3 + 1.5 * (Q_3 - Q_1)$
  - Upper extreme outlier line    $E_4 = Q_3 + 3 * (Q_3 - Q_1)$
- Following values can be referred for discreteness:
  - Mean = average of the data list

2. Calculate:
  - 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.

## MyBox User Guide – Data Tools v6.6



## 2.9.4 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 absolute values:

maximum\_value = maximum\_absolute\_value\_of\_column/row/all

percentage = absolute\_value / maximum\_value

width = maximum\_width \* percentage

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

percentage = (value - minimum\_value) / (maximum\_value - minimum\_value)

width = maximum\_width \* percentage

color = color\_of\_column

2. Data: Select rows in table or select all data rows(all pages), and select columns.

3. Calculate:

- Select one columns to calculated and columns to be copied.
- Compare: columns/rows/all.
- According to: absolute 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.

My Box Self comparison bars chart - CSV - 63 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv

**Data Filter Options**

Columns to be calculated  
(No selection means all)  

- year\_
- population at year-end(ten thousand)
- male(ten thousand)
- female(ten thousand)
- urban(ten thousand)
- rural(ten thousand)

Columns to be copied  

- year\_
- population at year-end(ten thousand)
- male(ten thousand)
- female(ten thousand)
- urban(ten thousand)
- rural(ten thousand)

To invalid number  Skip  Count as zero

All data are involved in memory. Too many data may cause out of memory.

Compare  Columns  Rows  All  

According to  Absolute value  Range of min and max

Maximum width 150 

Row number  Value  Percentage  Others  Calculated values

### CSV - 63 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation\_en.csv - Column comparison

male(ten thousand)	female(ten thousand)	year_	population at year-end(ten thousand)
Max absolute: 72357.0	Max absolute: 68969.0		
38%	37%	1949	54167
39%	38%	1950	55196
40%	39%	1951	56300
41%	40%	1952	57482
42%	41%	1953	58796
43%	42%	1954	60266
43%	42%	1955	61465
44%	43%	1956	62828

## 2.9.5 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 absolute values:

`maximum_value = maximum_absolute_value_of_value_columns`

`percentage = absolute_value / maximum_value`

`width = maximum_width * percentage`

`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`

`percentage = (value - minimum_value) / (maximum_value - minimum_value)`

`width = maximum_width * percentage`

`color = color_of_column`

2. Data: Select rows in table or select all data rows(all pages), and select columns.

3. Calculate:

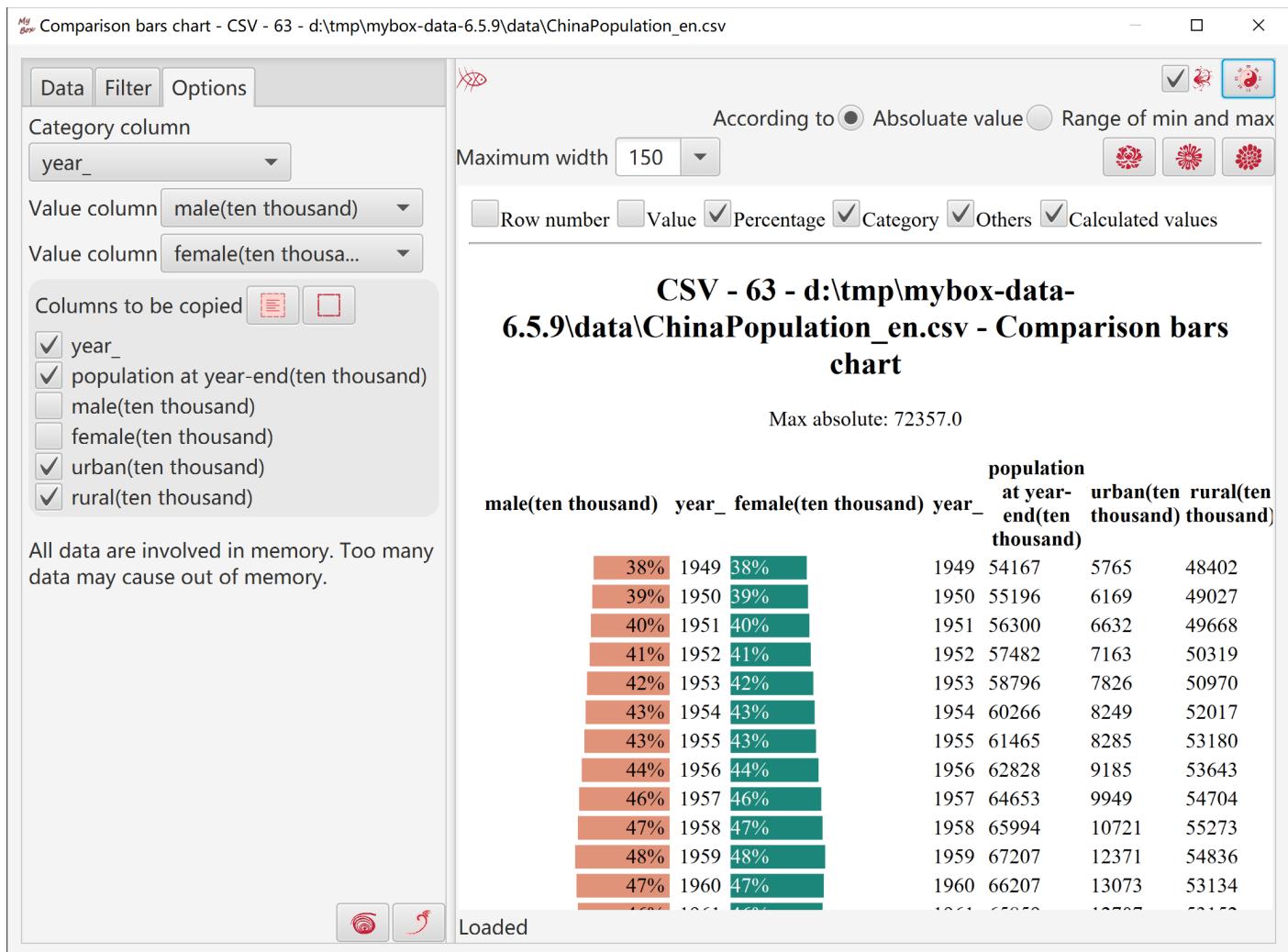
- Select one column as category column(unnecessary)
- Select tow value columns
- According to: absolute values, or range of minimum and maximum.
- When display all data rows(all pages), need concern memory limitation.

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.



## 2.9.6 XYZ Chart

### 2.9.6.1 3-D Scatter

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Select axes:
  - Select one column as axis X. Select as string or number.
  - Select one column as axis Y. Select as string or number.
  - Select several columns as axis Z. Select as string or number.
  - Select other columns to be popped in label.
3. Set parameters of the chart: projection, color, width, height, whether dark, point size.

XYZ chart - CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv

XYZ chart

Data Filter Options

Axis X age  Count values as strings

Axis Y BMI(body mass in...  Count values as strings

Axis Z (No selection means all)    Count values as strings

sex  BP(average blood pressure)  S1(blood serum measurement 1)  
 S2(blood serum measurement 2)  S3(blood serum measurement 3)  S4(blood serum measurement 4)  
 S5(blood serum measurement 5)  S6(blood serum measurement 6)  
 disease progression one year after baseline

Other values to be popped    
 age  sex  BMI(body mass index)  BP(average blood pressure)  S1(blood serum measurement 1)  
 S2(blood serum measurement 2)  S3(blood serum measurement 3)  S4(blood serum measurement 4)  
 S5(blood serum measurement 5)  S6(blood serum measurement 6)  
 disease progression one year after baseline

Type  Scatter chart  Surface chart

Projection  Perspective  Orthographic

Color  Column  Gradient  Random

Width 800.0 px

Height 600.0 px

Dark mode

Point size 10.0

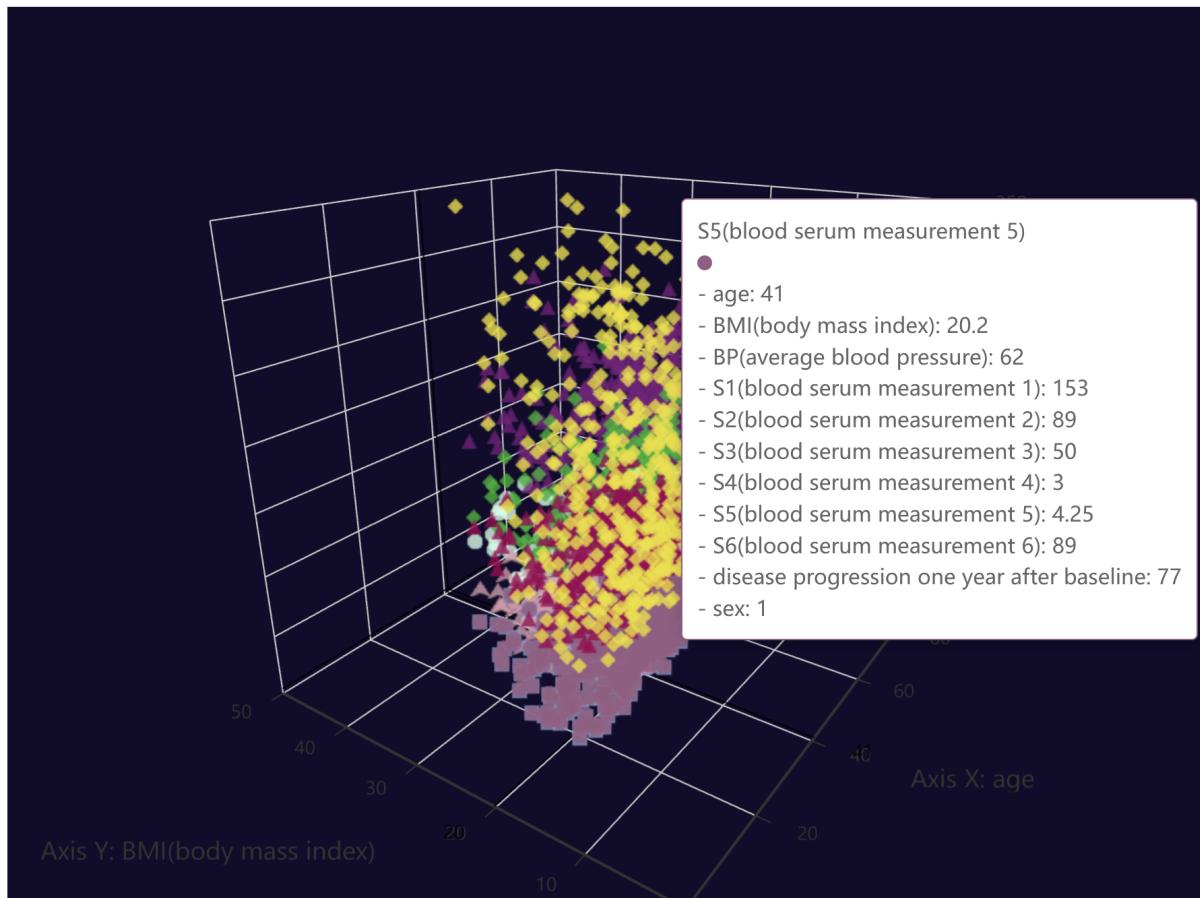




### Progression Prediction of Diabetes - Scatter chart

Axis X: "age" Axis Y: "BMI(body mass index)"

Axis Z: ● BP(average blood pressure) ▲ S1(blood serum measurement 1) ◆ S2(blood serum measurement 2)  
▲ S3(blood serum measurement 3) ■ S4(blood serum measurement 4) ● S5(blood serum measurement 5)  
▲ S6(blood serum measurement 6) ♦ disease progression one year after baseline



### 2.9.6.2 Surface Chart

1. Select data:
  - Rows can be: current page, selected rows, or all pages.
  - Set row filter.
2. Select axeses:
  - Select one column as axis X. Select as string or number.
  - Select one column as axis Y. Select as string or number.
  - Select one column as axis Z. Select as string or number.
  - Select other columns to be popped in label.
3. Set parameters of the chart: projection, color, width, height, whether dark, wire frame.

**XYZ chart - CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv**

**XYZ chart**

**Data** **Filter** **Options**

Axis X   Count values as strings

Axis Y   Count values as strings

Axis Z   Count values as strings

Other values to be popped

age  sex  BMI(body mass index)  BP(average blood pressure)  
 S1(blood serum measurement 1)  S2(blood serum measurement 2)  
 S3(blood serum measurement 3)  S4(blood serum measurement 4)  
 S5(blood serum measurement 5)  S6(blood serum measurement 6)  
 disease progression one year after baseline

Type  Scatter chart  Surface chart

Projection  Perspective  Orthographic

Color  Column  Gradient  Random

Width

Height

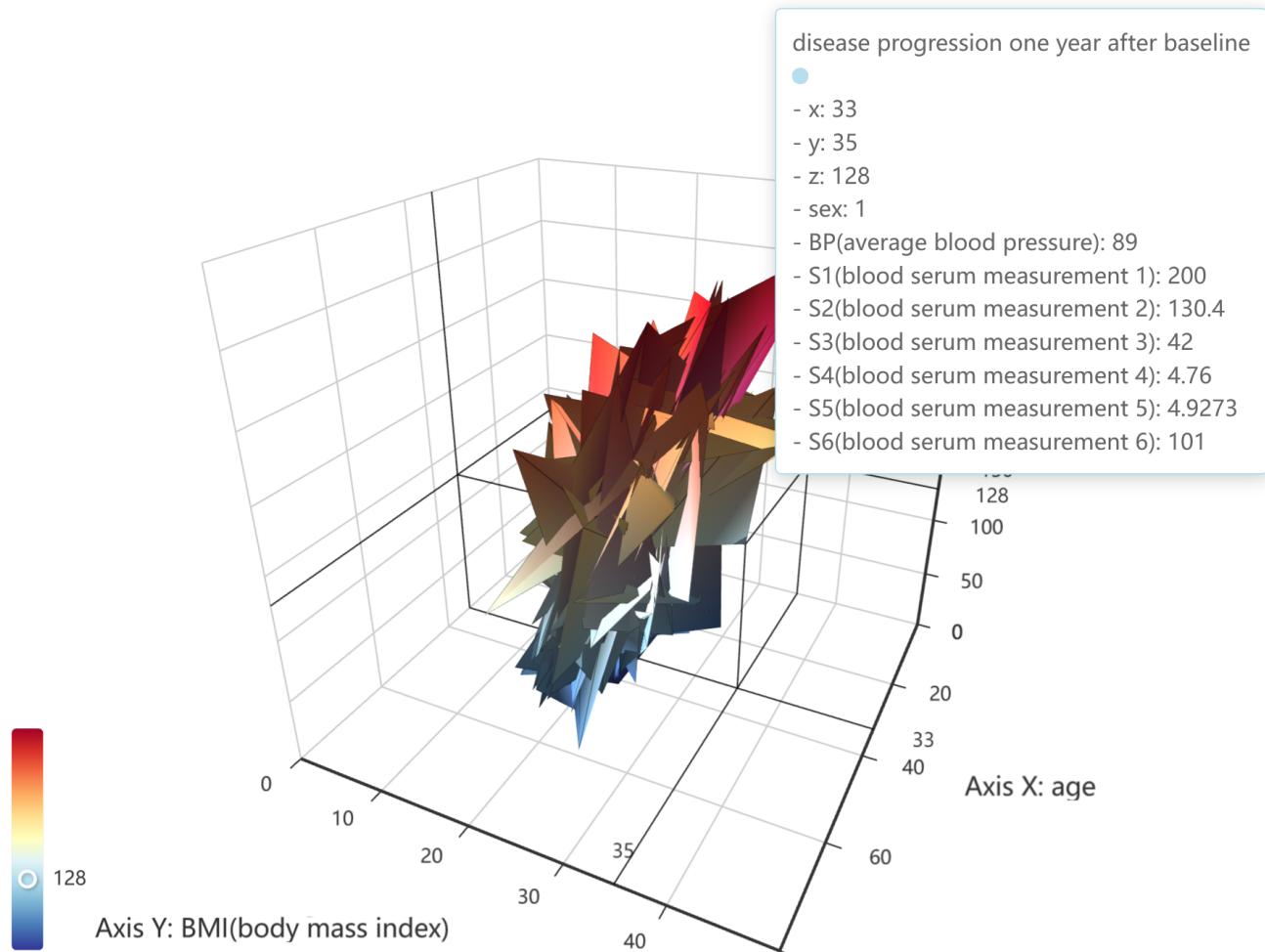
Dark mode  Wire frame





## Progression Prediction of Diabetes - Surface chart

Axis X: "age" Axis Y: "BMI(body mass index)" Axis Z: "disease progression one year after baseline"



## 2.10 Outputs

Hover or click button “Function” to select functions under menu item “Data”.

The screenshot shows a window titled "Edit CSV File : CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv". The menu bar includes "Window", "Document", "Image", "Data", "File", "Media", "Network", "Settings", "Recent Accessed", "Development", and "Help". Below the menu is a toolbar with icons for Save, Recover, Refresh, Open, Create data, Load content in system clipboard, Export, and Convert to database table. A context menu is open over a table, with the "Data" option highlighted. The "Data" menu contains "Modify", "Trim", "Calculation", "Charts", "Examples", and a checked "Pop when mouse hovering" option. At the bottom, there are buttons for Selected: 0, Rows: 50/442, Page size: 50, Page: 1, and navigation arrows.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	59	48	72	24	50	23	36	66	60	29	22	56	53	50
	2	2	1	1	2	2	1	2	32.1	1	1	28	1	2
	26.2	83	85	18.6	92	97	23.7	26.2	179	85	97	184	92	97
	114	180	114	97	186	186	109.2	186	119.4	180	114	144.8	109.2	186
	255	179	185	57.6	144.8	144.8	62	105.4	42	93.4	57.6	32	62	49
	185	119.4	46	43	144.8	144.8	3	49	42	43	46	6	62	49
	56	42	4	4	32	32	2	4	4	4	4	6	3	4
	4.5	4			6	6								

Selected: 0 Rows: 50/442 Page size: 50 Page: 1 /9

## 2.10.1 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.

Export - CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv

Data Row filter Target formats Target files Logs

Rows  Selected  Current page  All pages

Columns     

<input type="checkbox"/> Ta...	<input checked="" type="checkbox"/> Data ...	<input checked="" type="checkbox"/> age	<input checked="" type="checkbox"/> sex	<input checked="" type="checkbox"/> BMI(b...)	<input checked="" type="checkbox"/> BP(ave...)	<input type="checkbox"/> S1(blo...	<input type="checkbox"/> S2(blo...	<input type="checkbox"/> +
<input type="checkbox"/> 1	1	59	2	32.1	101	157	93.2	
<input type="checkbox"/> 2	2	48	1	21.6	87	183	103.2	
<input type="checkbox"/> 3	3	72	2	30.5	93	156	93.6	
<input type="checkbox"/> 4	4	24	1	25.3	84	198	131.4	
<input type="checkbox"/> 5	5	50	1	23	101	192	125.4	
<input type="checkbox"/> 6	6	23	1	22.6	89	139	64.8	

Selected: 0 Rows: 50/442 Page size 50 Page 1 /9       

### 2. Select targets and their formats: csv, texts, excel, xml, json, html, pdf, MyBox Clipboard.

### 3. Split files in maximum lines.

Export - CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv

Data Row filter Target formats Target files Logs

Base CSV Excel Texts html json PDF

Formats           

Split files in maximum lines Not split 

Row number   

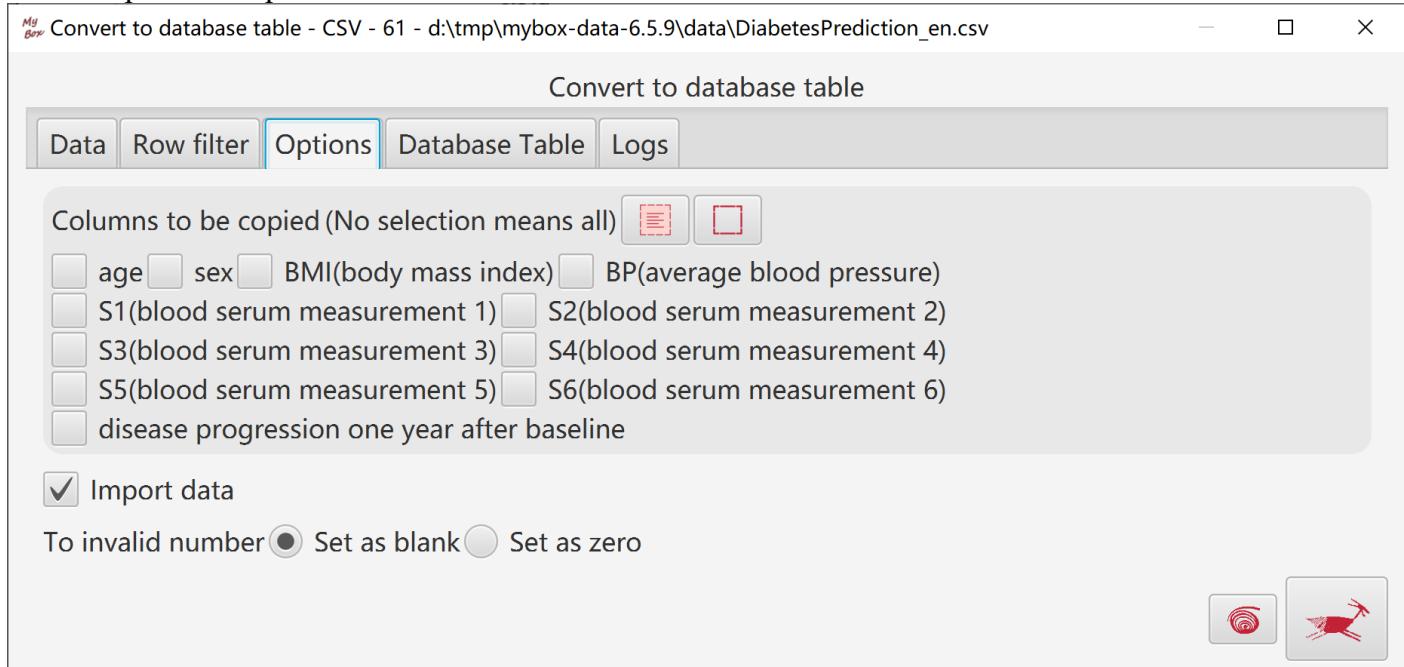
## 2.10.2 Convert to Database Table

### 1. Select data:

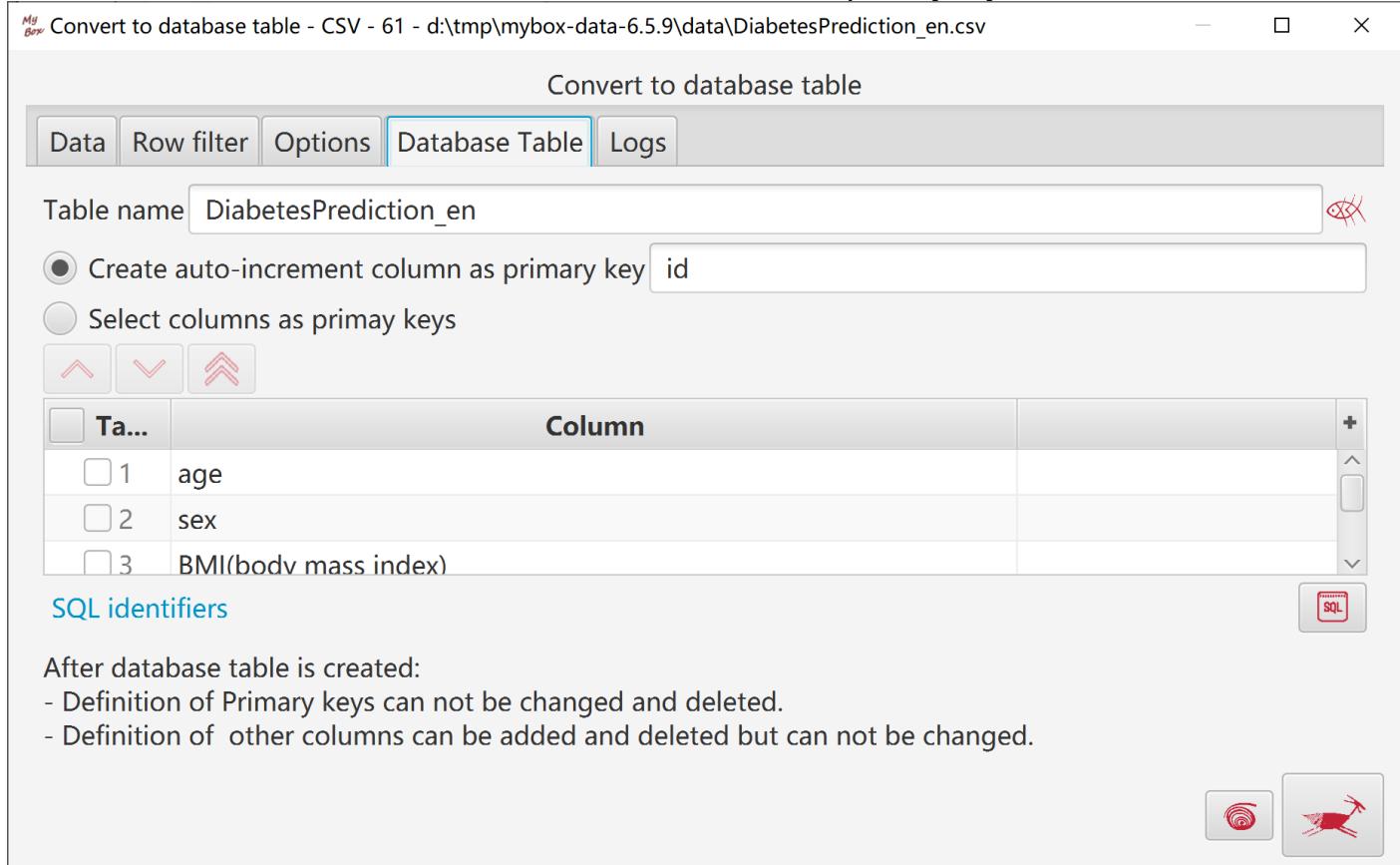
- Rows can be: current page, selected rows, or all pages.
- Set row filter.

### 2. Select columns to be copied.

### 3. Option to import data.



### 4. Create auto-increment column, or select some columns as primary key.



## 2.11 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 Clipboard.
- 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.

The screenshot shows the 'Manage Data' application window. The title bar reads 'Manage Data : Database Table - 1081 - CHINAPOPULATION\_EN'. The menu bar includes 'Window', 'Document', 'Image', 'Data', 'File', 'Media', 'Network', 'Settings', 'Recent Accessed', 'Development', and 'Help'. The toolbar has icons for Open, Save, Print, and Refresh.

The left sidebar lists data objects: 'Open' (selected), 'Excel', 'Texts', 'Matrix', 'Database Table', 'MyBox Clipboard', and 'Close(ESC/F6 Or click anywhere outside the object)'. A tooltip for 'Open' says 'MyBox CSV Manager'.

The main area displays a database table with columns: Row..., id, year\_, populatio..., and m+. The table data is as follows:

Data row	id	year_	populatio...	m+
1	1	1949	54,167	28
2	2	1950	55,196	28
3	3	1951	56,300	29
4	4	1952	57,482	29
5	5	1953	58,796	30
6	6	1954	60,266	31

Below the table, there's a 'Query' section with 'Type' checkboxes for CSV, XLS, TXT, Matrix, Database Table, CSV, and Database Table. It also includes 'Order by' options: Descending, Modify time, Name, ID, Rows number, Columns number, and File.

At the bottom, there are page navigation controls: 'Page 1 /10', 'Page size 50', 'Rows: 50/457 Selected: 1', 'Page 1 /2', and another set of navigation icons.

## 2.12 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.

Splice Data : Database Table - 1081 - CHINAPOPULATION\_EN

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Tab...	ID	Type	Row...	Colu...	+
<input type="checkbox"/> 1	1105	Texts	ChinaPo...	73	6
<input type="checkbox"/> 2	1061	Texts	ChinaPo...	73	6
<input type="checkbox"/> 3	1089	MyBox Clipbo...	ChinaPo...	6	7
<input type="checkbox"/> 4	1086	MyBox Clipbo...	ChinaPo...	6	7
<input type="checkbox"/> 5	1085	MyBox Clipbo...	ChinaPo...	6	7
<input type="checkbox"/> 6	1083	MyBox Clipbo...	ChinaPo...	1	7
<input type="checkbox"/> 7	1082	MyBox Clipbo...	ChinaPo...	6	7
<input type="checkbox"/> 8	1063	MyBox Clipbo...	a	6	3
<input type="checkbox"/> 9	1065	Matrix	a	3	3
<input type="checkbox"/> 10	1064	Matrix	b	6	4
<input type="checkbox"/> 11	1062	Excel	ChinaPo...	73	6
<input checked="" type="checkbox"/> 12	1081	Database Table	ChinaPo...	73	7
<input type="checkbox"/> 13	1069	Database Table	ed	3	3
<input type="checkbox"/> 14	1068	Database Table	ChinaPo...	73	7
<input type="checkbox"/> 15	1103	CSV	ChinaPo...	73	6

Page 1 /10 Page size  
50 Rows: 50/457 Selected: 1

Set as data B

Data A Data B Splice Data

Direction

Horizontal  Vertical

Rows number

By  Data A  Data B  Longer  Shorter

Target

CSV  XLS  TXT  Matrix  Clipboard  Database  Text  Image

Name

## 2.13 Data File

### 2.13.1 CSV File

1. In the file:
  - In general, the first line(header) defines column names, and each of followed lines defines a row of data.
  - Values are separated by "delimiter" which can be string.
  - If a value contains delimiter, it will be surrounded by quotes.
  - If a line starts with "#", it will be skipped.
2. When the file is opened for the first time, the tool guesses its delimiter and charset.
3. When file is loaded abnormally, change options and click Refresh button.
4. Data are paginated. When pages number is larger than 1, changes should be saved before run some functions.
5. Options: charset, whether has first line as field names, and delimiter of data.
6. To string values, multiple lines can be edited and saved.
7. Data can be saved as different charsets and delimiters.

The screenshot shows the MyBox Data Tools interface with the following details:

- Top Bar:** Edit CSV File : CSV - 61 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv
- Menu Bar:** Window, Document, Image, Data, File, Media, Network, Settings, Recent Accessed, Development, Help
- Left Sidebar:**
  - Information:** Shows the file path: d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction\_en.csv
  - Format:**
    - Charset:** Determine automatically (radio button selected), Known: UTF-8
    - First line defines columns names:** Checked
    - Text delimiter:** String (radio button selected), followed by a list of other delimiter options (Tab, Blank, Blank characters, 4 blanks, 8 blanks, etc.)
    - Note:** If file is loaded incorrectly, change options and click button "Refresh".
  - Backup:**
  - Save as:**
- Right Panel:**
  - Toolbar:** Includes icons for Columns, Attributes, View, Edit, and Table.
  - Data View:** A table showing data from the DiabetesPrediction\_en.csv file. The columns are labeled 'Ta...', 'Data ...', 'age', and 'sex'. The data consists of 14 rows of patient information.
  - Page Control:** Shows 'Selected: 0', 'Rows: 50/442', 'Page size: 50', 'Page: 1 / 9'.

	1	2	3	4
age	59	48	72	24
sex	2	1	2	1
age	50	23	36	66
sex	1	1	2	2
age	53	50	11	12
sex	1	2	1	2
age	56	22	13	14
sex	2	1	1	2
age	50	29	11	10
sex	2	1	1	1
age	60	29	10	9
sex	2	1	1	2
age	36	23	7	8
sex	2	1	1	2
age	48	3	2	4
sex	1	2	1	1
age	24	4	4	5
sex	1	1	1	1
age	72	3	3	6
sex	2	1	1	1
age	59	1	1	1
sex	2	1	1	1

## 2.13.2 Text File

1. In the file:
  - In general, the first line(header) defines column names, and each of followed lines defines a row of data.
  - Values are separated by "delimiter" which can be string.  
Regular expression is supported when parse the file.
  - Values should not contain delimiter.
  - If a line starts with "#", it will be skipped.
2. When the file is opened for the first time, the tool guesses its delimiter and charset.
3. When file is loaded abnormally, change options and click Refresh button.
4. Data are paginated. When pages number is larger than 1, changes should be saved before run some functions.
5. Options: charset, whether has first line as field names, and delimiter of data.
6. Not support multiple lines in values.
7. Data can be saved as different charsets and delimiters.

MyBox Edit Text Data File : Texts - 69 - d:\tmp\mybox-data-6.5.9\data\ChinaCPI\_en.csv

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Information Format Charset

Determine automatically  
 Known UTF-8

First line defines columns names

Text delimiter

String

Tab  Blank  Blank characters  4 blanks  
 8 blanks

,  |  #  -  +  :  ;  @  &  %  !  
 "  '  ?  .  \*  \  /  \_  =  <  >

If file is loaded incorrectly, change options and click button "Refresh"

Backup Save as Charset UTF-8

First line defines columns names

Text delimiter

String

Columns Attributes View Edit Table Text

##

Ta...	Data ...	year_	consumer ...	urba ...
1	1	1951	112.5	
2	2	1952	102.7	
3	3	1953	105.1	
4	4	1954	101.4	
5	5	1955	100.3	
6	6	1956	99.9	
7	7	1957	102.6	
8	8	1958	98.9	
9	9	1959	100.3	
10	10	1960	102.5	
11	11	1961	116.1	
12	12	1962	103.8	
13	13	1963	94.1	
14	14	1964	96.3	
15	15	1965	98.8	

Selected: 0 Rows: 50/71 Page size 50 Page

1 /2

### 2.13.3 Excel File

1. Choose one worksheet to handle.
2. Worksheets can be added/renamed/deleted.
3. In general, the first line(header) defines column names, and each of followed lines defines a row of data.
4. If file is read abnormally, change options and click button "Refresh".
5. Data are paginated. When pages number is larger than 1, changes should be saved before run some functions.
6. To string values, multiple lines can be edited and saved.
7. Data can be saved with current sheet only or all worksheets.

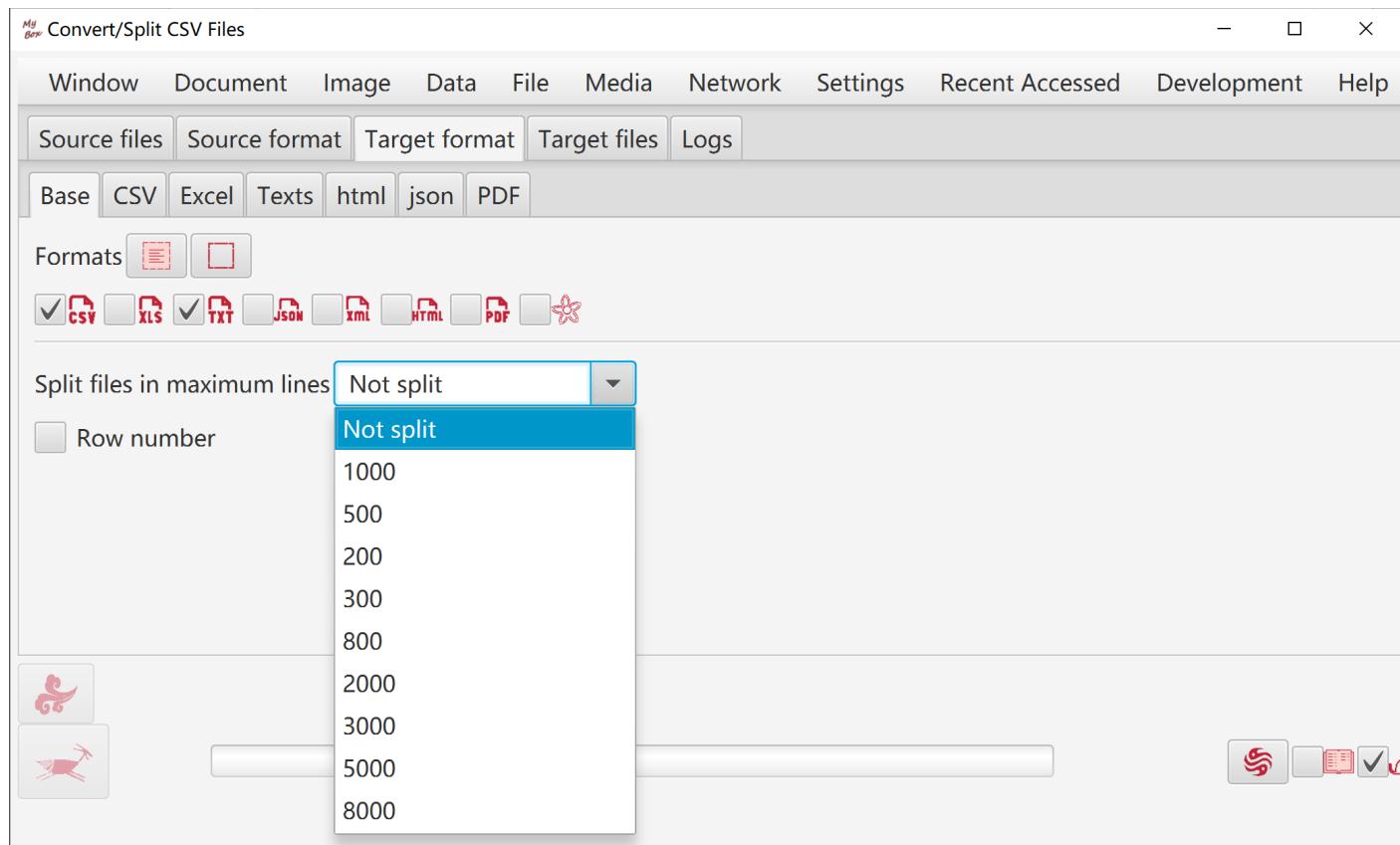
Notice: Tool can only handle base data in Excel file.

If file includes format, style, formula, or chart, suggest to save changes as new file to avoid data loss.

Tab...	Data row	year_	college gr...	middle sc...	high : .
1	1	1949	2.1	28	6.1
2	2	1950	1.8	29.6	6.2
3	3	1951	1.9	28.4	5.9
4	4	1952	3.2	22.1	3.6
5	5	1953	4.8	45.4	5.6
6	6	1954	4.7	64.4	6.8
7	7	1955	5.5	96.9	9.9
8	8	1956	6.3	93.9	15.4
9	9	1957	5.6	129.9	18.7
10	10	1958	7.2	131.3	19.7
11	11	1959	7	179	29.9
12	12	1960	13.6	171	28.8
13	13	1961	15.1	227.1	37.9
14	14	1962	17.7	202.5	44.1
15	15	1963	19.9	195.6	43.3
16	16	1964	20.4	175.3	36.7

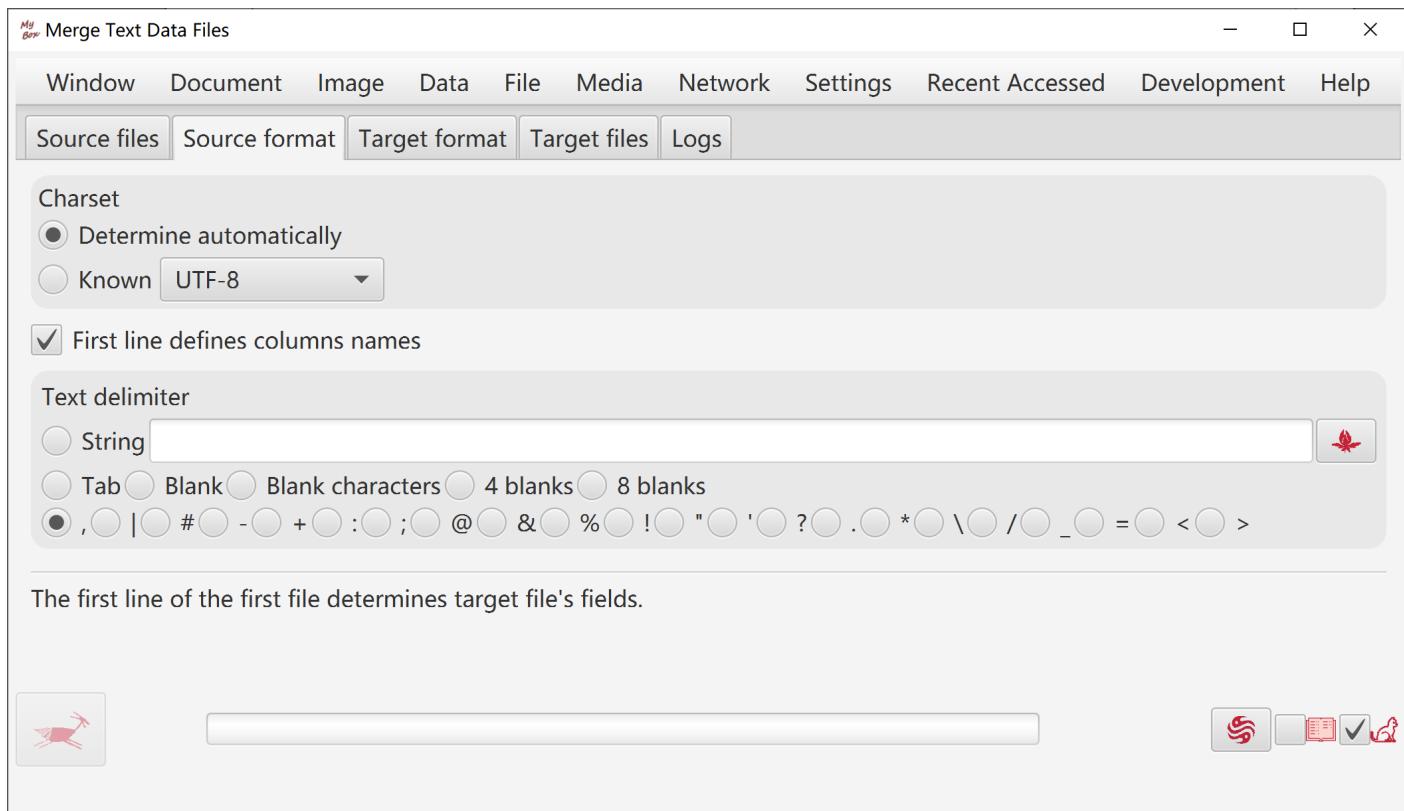
## 2.13.4 Convert/Split Data Files in Batch

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.



## 2.13.5 Merge Data Files

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



## 2.14 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.

MyBox Data in System Clipboard

Window Document Image Data File Media Network Settings Recent Accessed Development Help

First line defines the columns' names   

Source row number	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
Row1	5.1	3.5	1.4	0.2	Iris-setosa
Row2	4.9	3.0	1.4	0.2	Iris-setosa
Row3	4.7				Iris-setosa
Row4	4.6				Iris-setosa
Row5	5.0				Iris-setosa
Row6	5.4				Iris-setosa
Row7	4.6				Iris-setosa

MyBox: Set of Easy Tools

Text delimiter

String   

Tab  Blank  Blank characters  4 blanks

8 blanks

,  |  #  -  +  :  ;  @  &  %

!  "  '  ?  .  \*  \  /  \_  =  <

>

## 2.15 Data in MyBox Clipboard

The screenshot shows the MyBox Data Tools application window titled "Data in MyBox Clipboard : MyBox Clipboard - 6 - b". The window has a menu bar with options: Window, Document, Image, Data, File, Media, Network, Settings, Recent Accessed, Development, and Help. Below the menu is a toolbar with various icons.

The main area contains two tables. On the left, a table lists rows from 1 to 238. Row 7 is selected, indicated by a checked checkbox in the first column. A modal dialog box is open over this table, prompting for a new name, with "bm" entered in the "New name" field. The dialog has buttons for "确定" (Confirm) and "取消" (Cancel).

On the right, there is another table with columns labeled "#", "Tab...", "Data row", "收入" (Income), and "快乐" (Happiness). Rows 1 through 12 are listed, each with a checkbox in the first column and numerical values in the other columns.

At the bottom, there are page navigation controls (Page 1 / 1, Previous, Next, Last), a "Page size" dropdown set to 50, and status information: "Selected: 0", "Rows: 50/238", and "Page: 1 / 5".

## 2.16 Matrix

### 2.16.1 Edit and Manage Matrices

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

The screenshot shows the 'Manage Matrices' window in MyBox Data Tools. The left panel lists matrices with columns for ID, Type, Name, Row..., and Colu... (with a plus sign). Matrix 1065 is selected, showing details for matrix 'a' with 3 rows and 3 columns. The right panel shows the data for matrix 'a' in a table with columns for Data row, 列1, 列2, and 列3. The data is as follows:

Data row	列1	列2	列3
1	353.63	469.35	59.21
2	687.2	344.24	308.21
3	359.8	359.52	925.57

At the bottom, there are page navigation controls (Page 1 /1) and a status bar indicating 50 rows and 2 selected.

## 2.16.2 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.

The screenshot shows the 'Matrix Unary Calculation' window in MyBox. On the left, there is a table with columns: Ta..., ID, Type, Name, Row..., Colu..., and +. Two rows are visible: row 1 (selected) with ID 1065, Type Matrix, Name a, Row 3, Column 3; and row 2 with ID 1064, Type Matrix, Name b, Row 6, Column 4. Below the table are page navigation controls (Page, Page size, Rows: 2/2 Selected: 1) and a zoom control (50%).

On the right, there is a toolbar with icons for transpose, echelon form, determinant, inverse matrix, adjoint matrix, and power. Below the toolbar is a menu bar with items: Window, Document, Image, Data, File, Media, Network, Settings, Recent Accessed, Development, Help.

A large panel on the right contains several tabs: Data, Calculation, and Result. The Calculation tab is active, showing various calculation methods as radio buttons: Transpose, Row echelon form, Reduced row echelon form, Complement minor, Normalize (which is selected), Multiply number, Divide number, Determinant by elimination, Determinant by complement minor, Inverse matrix by elimination, Inverse matrix by adjoint, Rank of matrix, Adjoint matrix, and Power.

Below these buttons are settings for 'According to' (Columns is selected), 'Algorithm' (MinMax is selected), and a 'Range' input field with values -1.0, 1.0, (0,1), (-1,1), and (-1.0,1.0). At the bottom of this panel is a 'Calculate ENTER' button.

### 2.16.3 Binary Matrices Calculation

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

Matrices Binary Calculation : Matrix - New data

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Ta...	ID	Name	Row +
<input type="checkbox"/> 1	1065	a	3
<input type="checkbox"/> 2	1064	b	6

Matrix A Matrix B Calculation Result

Vertical merge  Horizontal merge  Plus  Minus  Multiply  
 Hadamard product  Kronecker product

Example:  
 $A = \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{21} \end{matrix}$   
 $B = \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{21} \end{matrix}$

Kronecker Product =  
$$\begin{matrix} a_{11}*b_{11} & a_{11}*b_{12} & a_{12}*b_{11} & a_{12}*b_{12} \\ a_{11}*b_{21} & a_{11}*b_{21} & a_{12}*b_{21} & a_{12}*b_{21} \\ a_{21}*b_{11} & a_{21}*b_{12} & a_{22}*b_{11} & a_{22}*b_{12} \\ a_{21}*b_{21} & a_{21}*b_{21} & a_{22}*b_{21} & a_{22}*b_{21} \end{matrix}$$

## 2.17 Database Tables

### 2.17.1 Manage Database Tables

1. View table definition.
2. Execute SQL.

Database Table : Database Table - 1113 - INCOMEHAPPINESS\_EN

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Columns Attributes View Table definition

Ta...	ID	Type	Name	Row...	Colu...	
<input checked="" type="checkbox"/>	1	1113	Databas...	IncomeHappi...	477	3
<input type="checkbox"/>	2	1081	Databas...	ChinaPopulat...	73	7
<input type="checkbox"/>	3	1069	Databas...	ed	3	3
<input type="checkbox"/>	4	1068	Databas...	ChinaPopulat...	73	7
<input type="checkbox"/>	5					
<input type="checkbox"/>	6					
<input type="checkbox"/>	7					
<input type="checkbox"/>	8					

INCOMEHAPPINESS\_EN

Column	Type	Length	Not null	Primary key	Auto generated	Refer to table	Refer to column
id	Long	19	Yes	Yes	Yes		
income	Double	52					
happiness	Double	52					

```
CREATE TABLE INCOMEHAPPINESS_EN (
    id BIGINT NOT NULL GENERATED BY DEFAULT AS IDENTITY (START WITH 1, INCREMENT BY 1),
    income DOUBLE,
    happiness DOUBLE,
    PRIMARY KEY ( id )
```

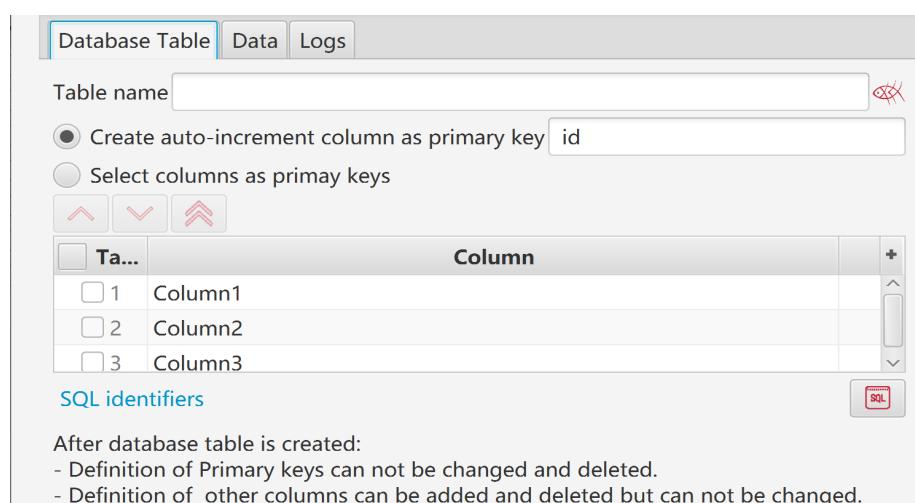
Page 1 /10

## 2.17.2 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.



## 2.17.3 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.

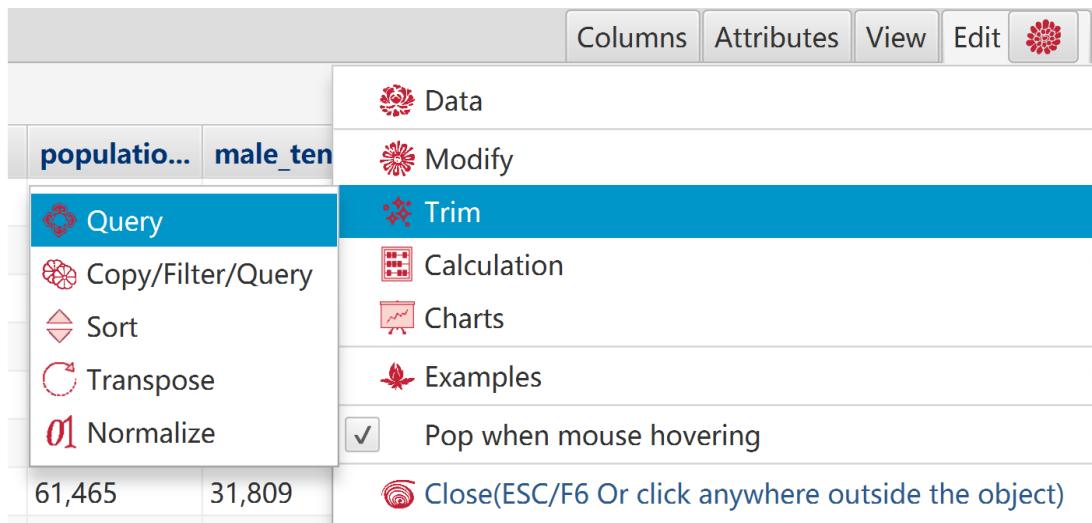
The screenshot shows the 'Database SQL' module of MyBox Data Tools. The left sidebar has a tree view under 'SQL' with 'Maximum rows' selected. The main area has a toolbar with various icons, a query input field containing 'SELECT \* FROM visit\_history FETCH FIRST 300 ROWS ONLY', and a results table showing 8 rows of data from a database.

Data row	RESOURC...	FILE_TYPE	OPERATIO...	RESOURC...	DATA_MO...
1	2	21	3	d:\tmp\my...	20
2	1	21	3	d:\tmp\my...	20
3	2	21	2	d:\tmp\my...	20
4	1	21	2	d:\tmp\my...	20
5	2	21	3	d:\tmp\my...	20
6	1	21	3	d:\tmp\my...	20
7	2	21	2	d:\tmp\my...	20
8	1	21	2	d:\tmp\my...	20

## 2.17.4 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.



Database Table : Database Table - 1115 - CHINAPOPULATION\_EN \*

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Database Table - 1115 - CHINAPOPULATION\_EN

Input a statement of SQL query which can be written in multiple lines

CHINAPOPULATION_EN	Attributes
id	Options
year_	SQL*
population_at_year_end_ten_thousand_	
male_ten_thousand_	
female_ten_thousand_	
urban_ten_thousand_	
rural_ten_thousand_	

SELECT year\_, female\_ten\_thousand\_, urban\_ten\_thousand\_, rural\_ten\_thousand\_ FROM CHINAPOPULATION\_EN

## 3 Script and Expression

### 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. External 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.

## MyBox User Guide – Data Tools v6.6

My 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

Attributes Codes

Results Snippets Class paths

Examples

Object

- string
- time
- array
- list

Expressions

- Numeric operator
- 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
- string include

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);  
 }  
 DecimalFormat df = new DecimalFormat(format);  
 df.setRoundingMode(RoundingMode.HALF\_UP);  
 double result = df.parse(data).doubleValue();  
 return result;  
 } catch (Exception e) {  
 e.printStackTrace();  
 }  
}  
  
import java.  
double circleAreaByRadius(double radius) {  
 return Math.PI \* radius \* radius;  
}  
  
formatDouble(circleAreaByRadius(273.4), 4)

id: 18  
Status: Valid  
Type: METHOD  
Name: formatDouble

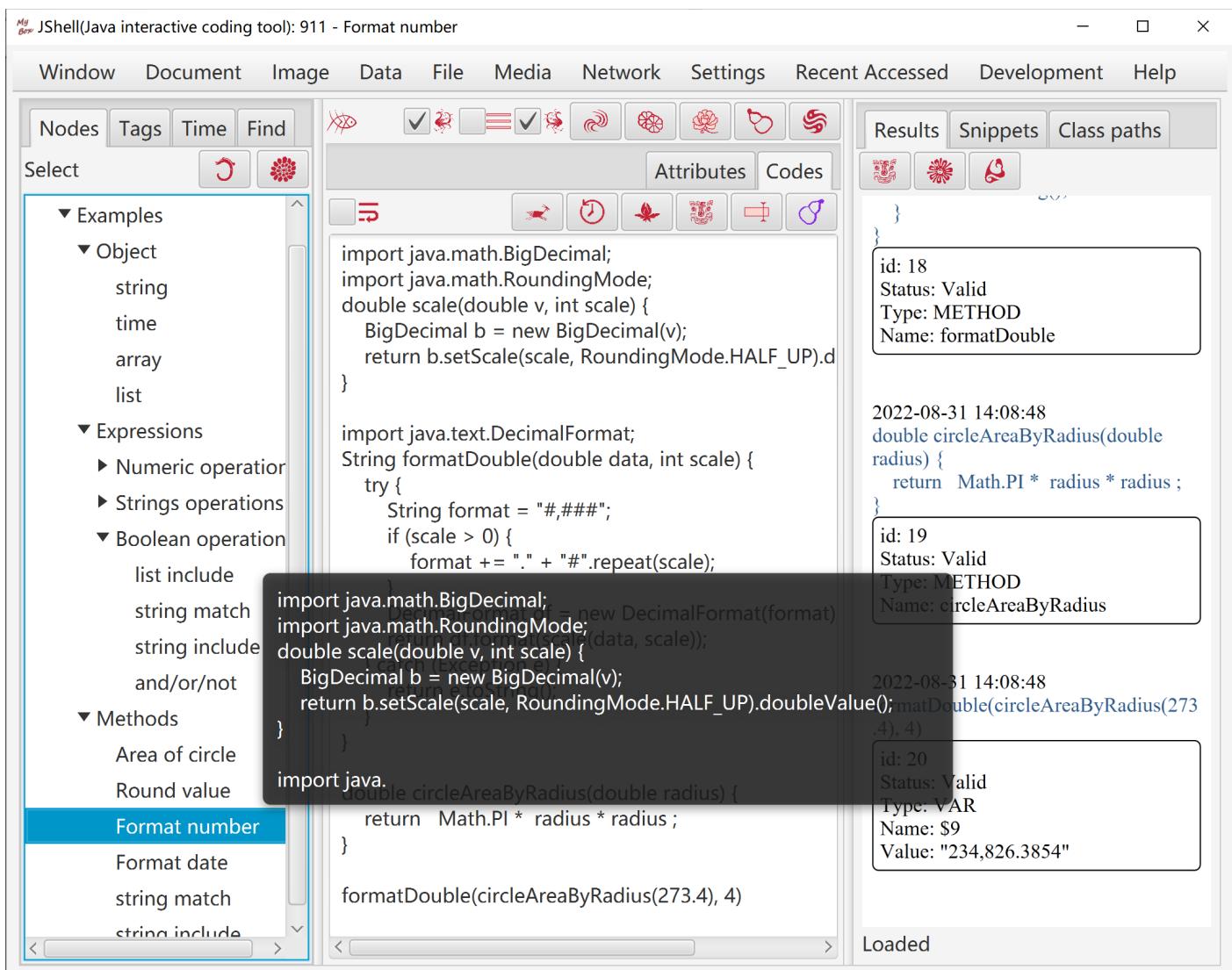
2022-08-31 14:08:48  
double circleAreaByRadius(double  
radius) {  
 return Math.PI \* radius \* radius ;  
}

id: 19  
Status: Valid  
Type: METHOD  
Name: circleAreaByRadius

2022-08-31 14:08:48  
double circleAreaByRadius(double radius) {  
 return Math.PI \* radius \* radius ;  
}  
formatDouble(circleAreaByRadius(273.4), 4)

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 environment 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.

## MyBox User Guide – Data Tools v6.6

The screenshot shows the MyBox Data Tools interface. The left sidebar contains a tree view of JEXL Codes, with the 'percentage' node selected. The main area has tabs for JEXL script, JEXL Context, and Jexl script parameters. The JEXL script tab displays the following code:

```
var number = 37;
var total = 518;
var scale = 2;
DoubleTools.percentage(number, total, scale);
```

A tooltip is shown over the 'percentage' node in the sidebar, containing the same code. The JEXL Context tab shows the context setup:

```
jexlContext.set("DoubleTools", mara.mybox.tools.DoubleTools.class);
```

The right side shows a history panel with three entries:

- 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 manage and run codes in JavaScript:

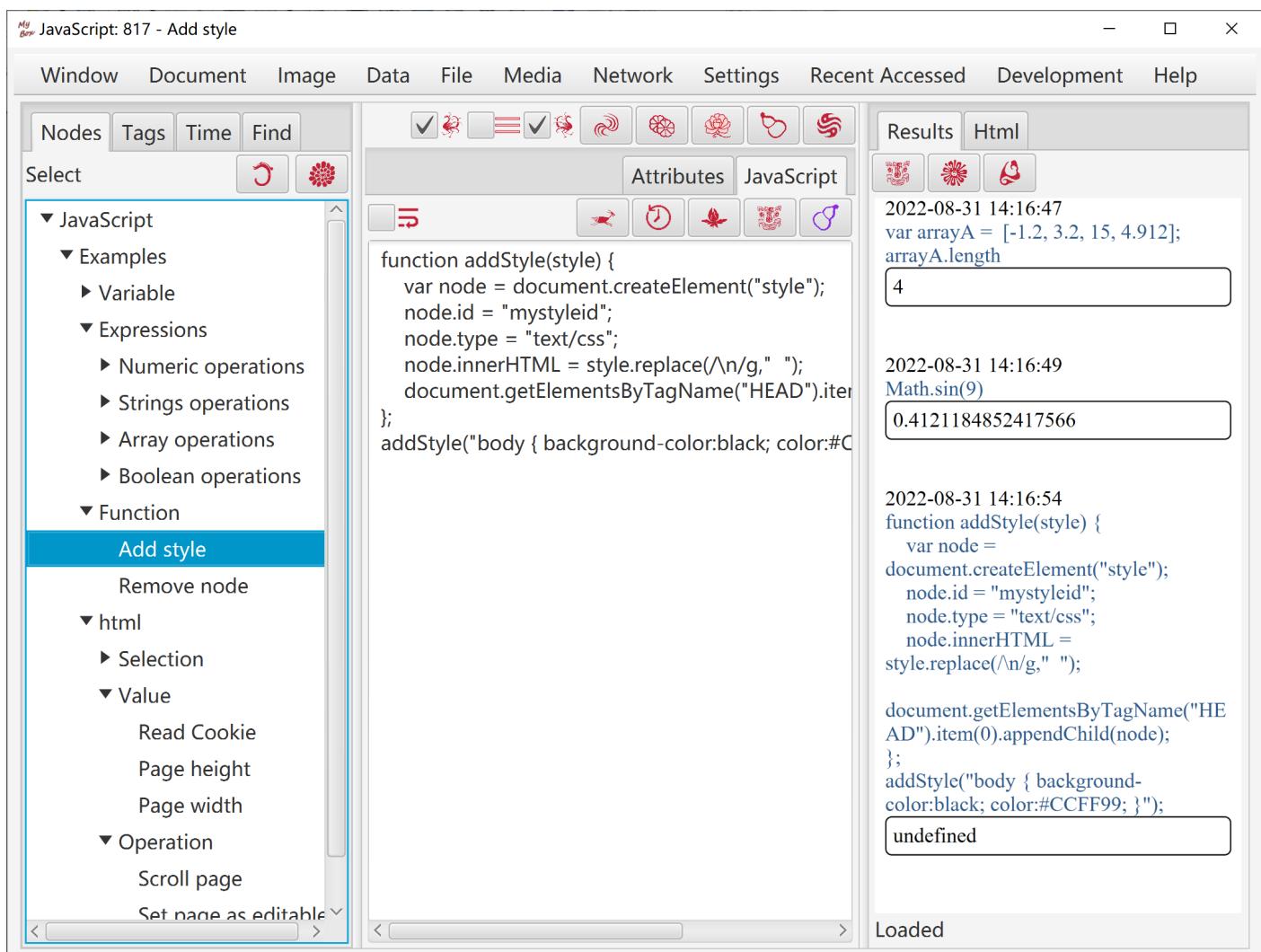
1. Edit codes in JavaScript.

The script can include any valid elements which WebEngine can parse(ECMAScript 6).

2. Run the script.

Its results are displayed in right pane. And it affects the web page in right pane too.

3. Javascript codes are organized in tree. Examples are provided.



## 4 Math Function

This tool helps to manage and calculate math functions:

### 4.1 Define Math Functions

1. List names of variables of the function, separated by English comma.

The names should satisfy rules of JavaScript:

- Can include letters, number, '-', '\_', and unicodes(like Chinese).
- Not start with number.

2. Define function expression as a piece of codes in JavaScript:

- The script can include any valid elements which Nashorn can parse(ECMAScript 5.1).
- It should be a number finally.
- It can refer to the given variables, but should not declare them.

3. Define domain of the function as a piece of codes in JavaScript:

- Blank means the domain is all of real numbers.
- The script can include any valid elements which Nashorn can parse(ECMAScript 5.1).
- It should be a boolean value(true or false) finally.
- It can refer to the given variables, but should not declare them.

4. Give the name of calculation result.

5. Saved in tree.

6. Examples are provided.

### 4.2 Calculate Math Function

1. When the tool calculates a script, 'var <variable\_name>=<variable\_value>;' are inserted in the head of it before the evaluation.
2. If script of domain is not blank, the given values of variables are inserted in it and check the result.
3. If true, continue to calculate the script of function expression. Else this set of values is skipped.

# MyBox User Guide – Data Tools v6.6

My Math Function: 217 - univariate logarithmic

Window Document Image Data File Media Network Settings Recent Accessed Development Help

Nodes Tags Time Find

Select

**Math Function**

- Examples
  - Constants
  - Unary Function
  - Binary Function
    - Univariate Polynomial
    - Geometry
    - Quadric Surface
    - Trigonometry
    - Exponential Function
    - Logarithmic Function
      - univariate logarithmic
  - Numerical Analysis
    - Univariate Normal probability
    - Univariate Standard normal pr
- Ternary Function
  - Trigonometry
- Geometry
  - area of ellipsoid - half\_axie
  - area of ellipsoid - axie
  - volumn of ellipsoid
  - volumn of ellipsoid - axie

Attributes Math Function

Names of variables, separated by English comma

x, y

Expression

```
5 * Math.log(x * y) - 9
```

Function domain

x \* y > 0

Names:::area of ellipsoid,axie\_a,axie\_b,axie\_c  
 $\text{Math.PI} * (\text{axie}_a * \text{axie}_b + \text{axie}_a * \text{axie}_c + \text{axie}_b * \text{axie}_c) / 3$

Result name univariate logarithmic

Calculate Data Set

x y

Decimal scale 8

var x=4.0;  
 var y=5.0;  
 var xMean = 1;  
 var xStd = 2;  
 var yMean = 2;  
 var yStd = 1;  
 var coefficient = 0.2;  
 var dx = x - xMean;  
 var dy = y - yMean;  
 var xyStd = xStd \* yStd;  
 var dco = 1 - coefficient \* coefficient;  
 var px = dx \* dy / (xStd \* yStd);  
 var py = dy \* dx / (yStd \* xStd);  
 var pxy = 2 \* coefficient \* dx \* dy;  
 var expP = -(px + py - pxy) / (2 \* coefficient);  
 var div = 2 \* Math.PI \* xyStd \* Math.abs(expP);  
 Math.exp(expP) / div

Univariate Normal probability density function

2022-09-28 14:13:12  
 var angle=45.0;  
 angle \* Math.PI / 180  
 radian=0.78539816

## 4.3 Data Set

With defined range, interval, and decimal scale, data set of the function can be generated.

The screenshot shows the 'Math Function' configuration window in MyBox Data Tools. The title bar reads "Math Function: 225 - law of sines - for edge(radian)". The menu bar includes Window, Document, Image, Data, File, Media, Network, Settings, Recent Accessed, Development, and Help. The left sidebar has tabs for Nodes, Tags, Time, and Find, with "Select" currently selected. A tree view under "Math Function" shows categories like Examples, Constants, Unary Function, Trigonometry, Geometry, Power Function, Exponential Function, Logarithmic Function, Intercept Function, Unitary Polynomial, piecewise values, Numerical Analysis, Binary Function, Ternary Function, and Trigonometry. Under Trigonometry, "law of sines - for edge(radian)" is selected and highlighted in blue. The main workspace contains several tabs: Attributes, Math Function, Names of variables, separated by English comma (with "angleA\_radian, edge\_a, angleB\_radian" listed), Expression (with the formula "edge\_a \* Math.sin(angleB\_radian) / Math.sin(angleA\_radian)"), Function domain (with the condition "angleA\_radian > 0 && angleB\_radian > 0 && edge\_a > 0"), and Result name (set to "edge\_b"). On the right, there are tabs for Calculate and Data Set, and settings for Range (From -10.0, To 10.0, Number of split 50, Data interval 0.1), Decimal scale (set to 2), and a calculate button.

## 4.4 XY Chart of Unary Function

To unary function, XY charts, including scatter chart and line chart, can be displayed.

The screenshot shows the MyBox Data Tools interface with the following details:

- Title Bar:** MyBox Math Function: 196 - sigmoid
- Menu Bar:** Window, Document, Image, Data, File, Media, Network, Settings, Recent Accessed, Development, Help
- Left Panel (Select):**
  - Math Function:**
    - Examples
    - Constants
    - Unary Function** (selected)
      - absolute
      - Trigonometry
      - Geometry
      - Power Function
      - Exponential Function
      - Logarithmic Function
      - Intercept Function
      - Unitary Polynomial
      - piecewise values** (selected)
        - direct values
        - define functions
    - Numerical Analysis
    - Binary Function
    - Univariate Polynomial
    - Geometry
    - Quadric Surface
- Middle Panel (Attributes/Math Function):**
  - Attributes:** Various icons for file operations.
  - Math Function:**
    - Names of variables, separated by English comma: x
    - Expression:  $1 / (1 + \text{Math.exp}(-x))$
- Right Panel (Calculate/Data Set):**
  - Calculate:** x
  - Range:** From -20.0, To 20.0
  - Number of split:** 100 (radio button selected)
  - Data interval:** 0.1 (checkbox)
  - Decimal scale:** 2
- Bottom Panel (XY chart - sigmoid):**
  - Scatter chart** (radio button selected)
  - Line chart**
  - Functions:** sigmoid
  - Result name:** sigmoid
  - Plot Area:** A scatter plot showing the sigmoid function  $y = 1 / (1 + e^{-x})$  for  $x$  from -2.0 to 2.0. The curve passes through the point (0, 0.5).

## 4.5 XYZ Chart of Binary Function

To binary function, XYZ chart, including 3D scatter chart and surface chart, can be displayed.

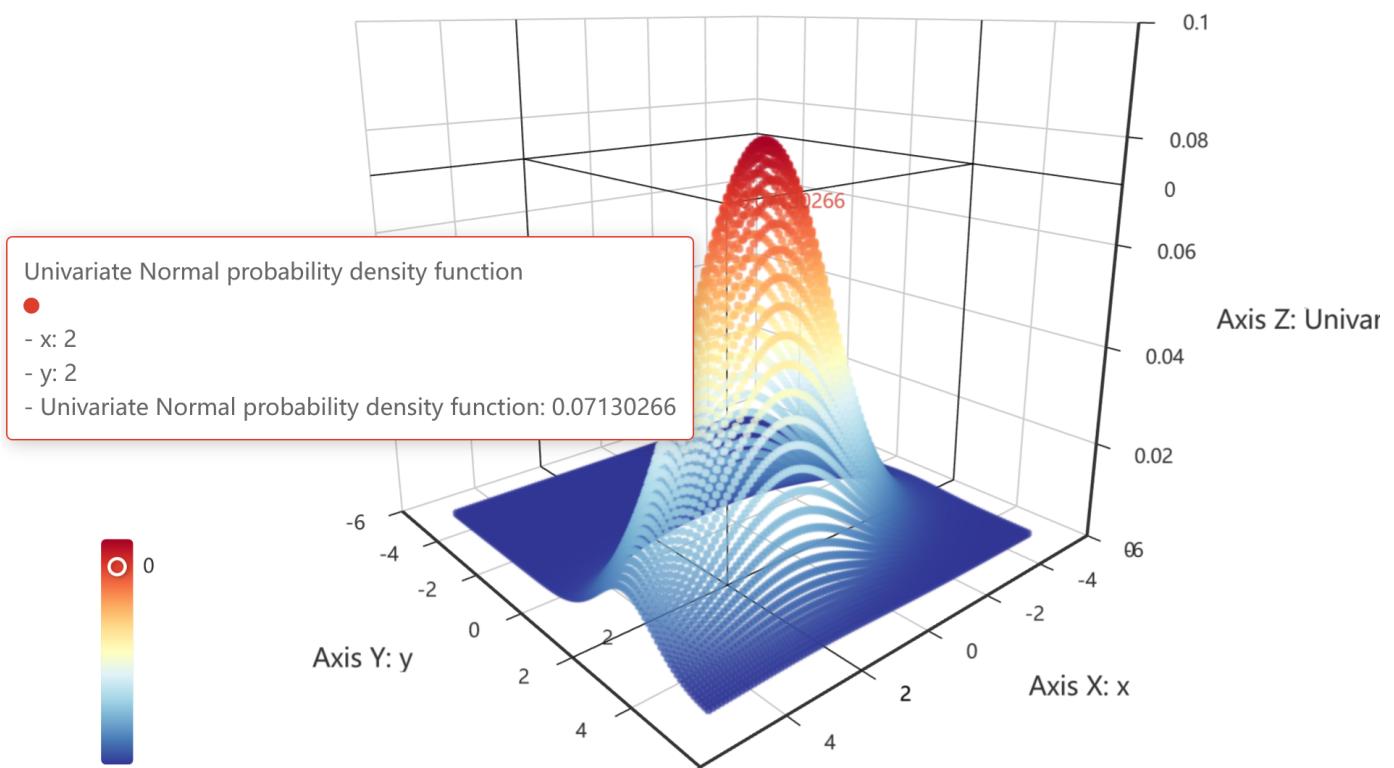
mybox-data-6.5.9/generated/Univariate\_Normal\_probability\_density\_function\_Scatter\_chart\_2022-09-28\_14-2 ☆



### Univariate Normal probability density function - Scatter chart

Axis X: "x" Axis Y: "y"

Axis Z: ● Univariate Normal probability density function



## 5 Data of Location

### 5.1 Data Constraints

#### 5.1.1 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)

#### 5.1.2 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 encrypted coordinate), based on GCJ-02.
5. Mapbar coordinate, based on GCJ-02.
6. When coordinate is unknown or invalid, the default value is CGCS2000.

#### 5.1.3 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'.

#### 5.1.4 Time

##### 5.1.4.1 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".

##### 5.1.4.2 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"

#### 5.1.4.3 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

## 5.2 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.

### 5.3 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.

## 5.4 Geography Code

### 5.4.1 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: altitude, precision, coordinate system, area(square meters), population, comments, isPredefined.

### 5.4.2 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.

### 5.4.3 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.

### 5.4.4 Define Condition

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

### 5.4.5 Import Data

#### 5.4.5.1 Predefined Data

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

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

#### 5.4.5.2 CSV Format

- Download address:  
[https://github.com/Mararsh/MyBox\\_data/tree/master/md/GeographyCode/en](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

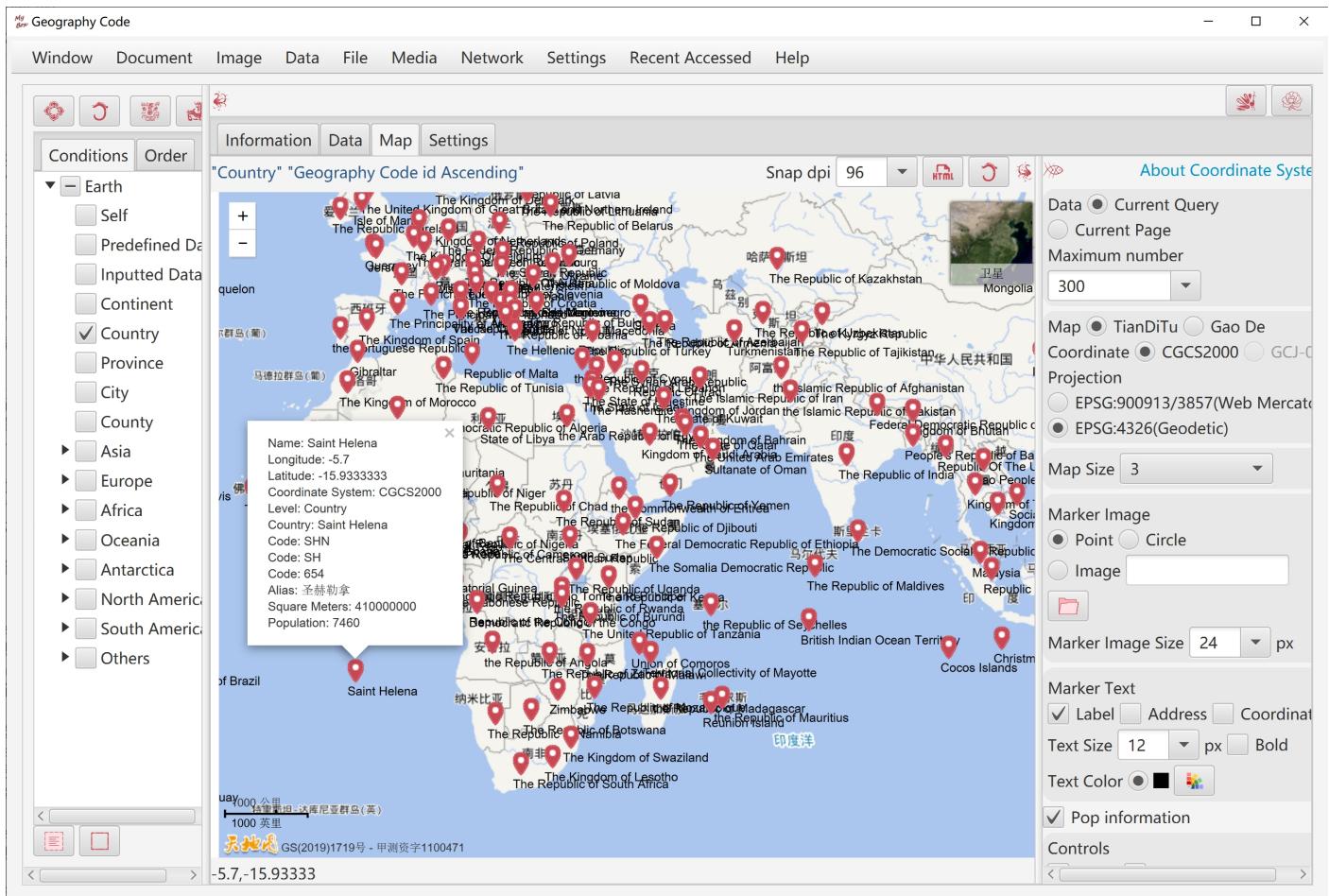
#### 5.4.5.3 Data from geoname.org:

- 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.

#### 5.4.6 Settings

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

# MyBox User Guide – Data Tools v6.6



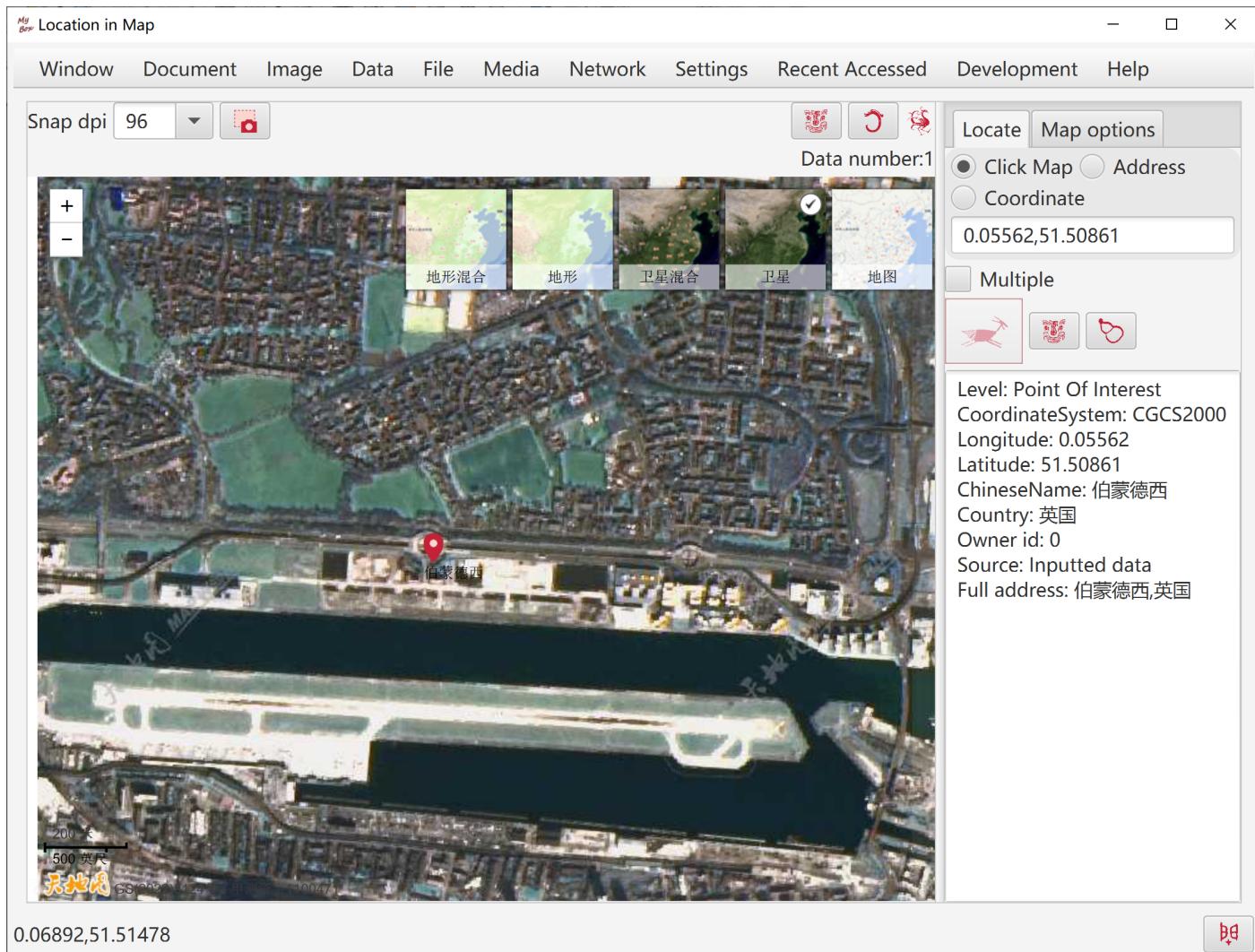
## 5.5 Location in Map

### 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.



## 5.6 Location Data

### 5.6.1 Data Definition

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

### 5.6.2 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.

### 5.6.3 Define Conditions

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

### 5.6.4 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.

## 5.6.5 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)

## 5.6.6 Import Data

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

### 5.6.6.1 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

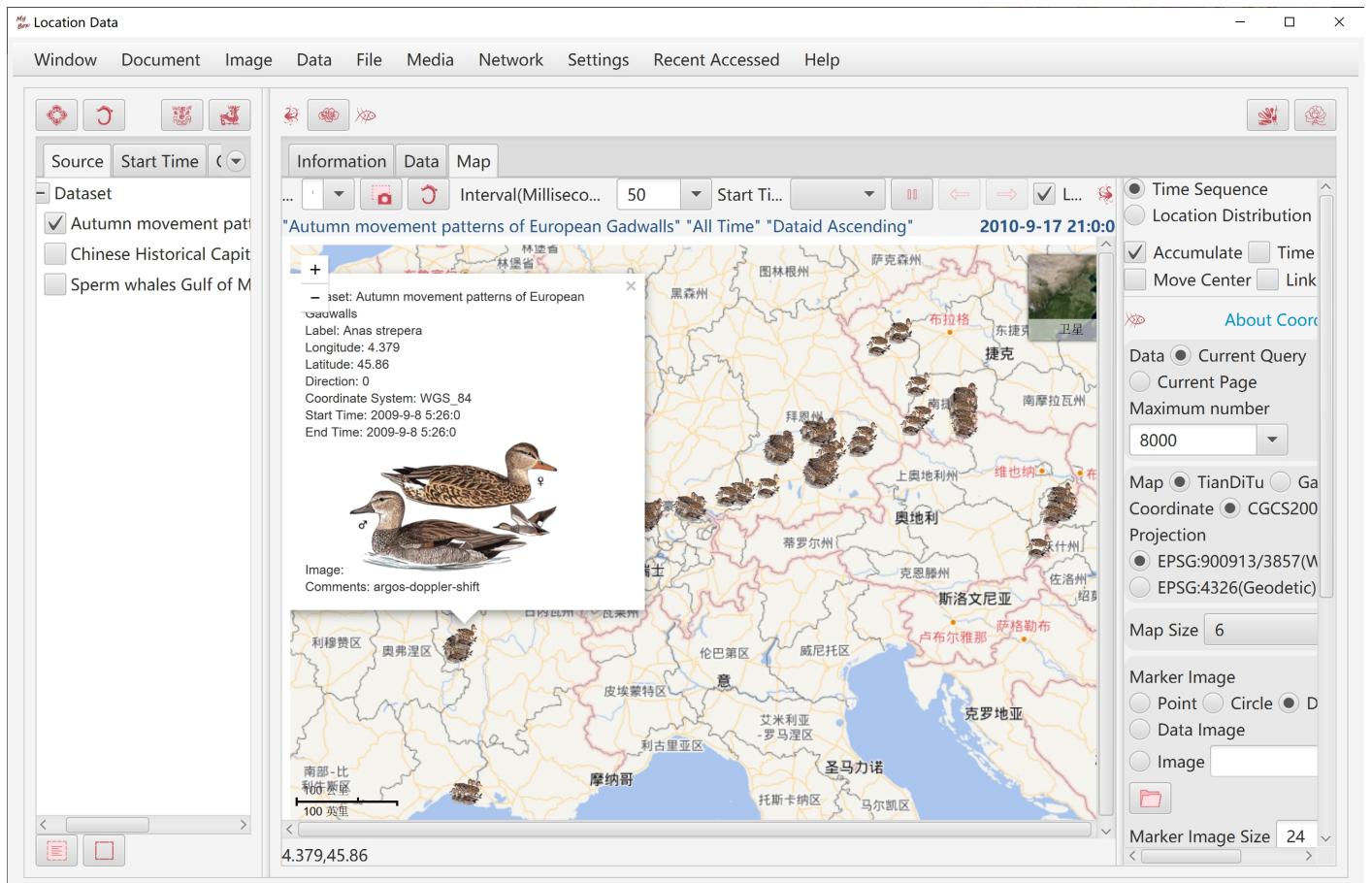
### 5.6.6.2 Data from movebank.org

- 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.

### 5.6.6.3 Examples

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

# MyBox User Guide – Data Tools v6.6



## 5.7 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

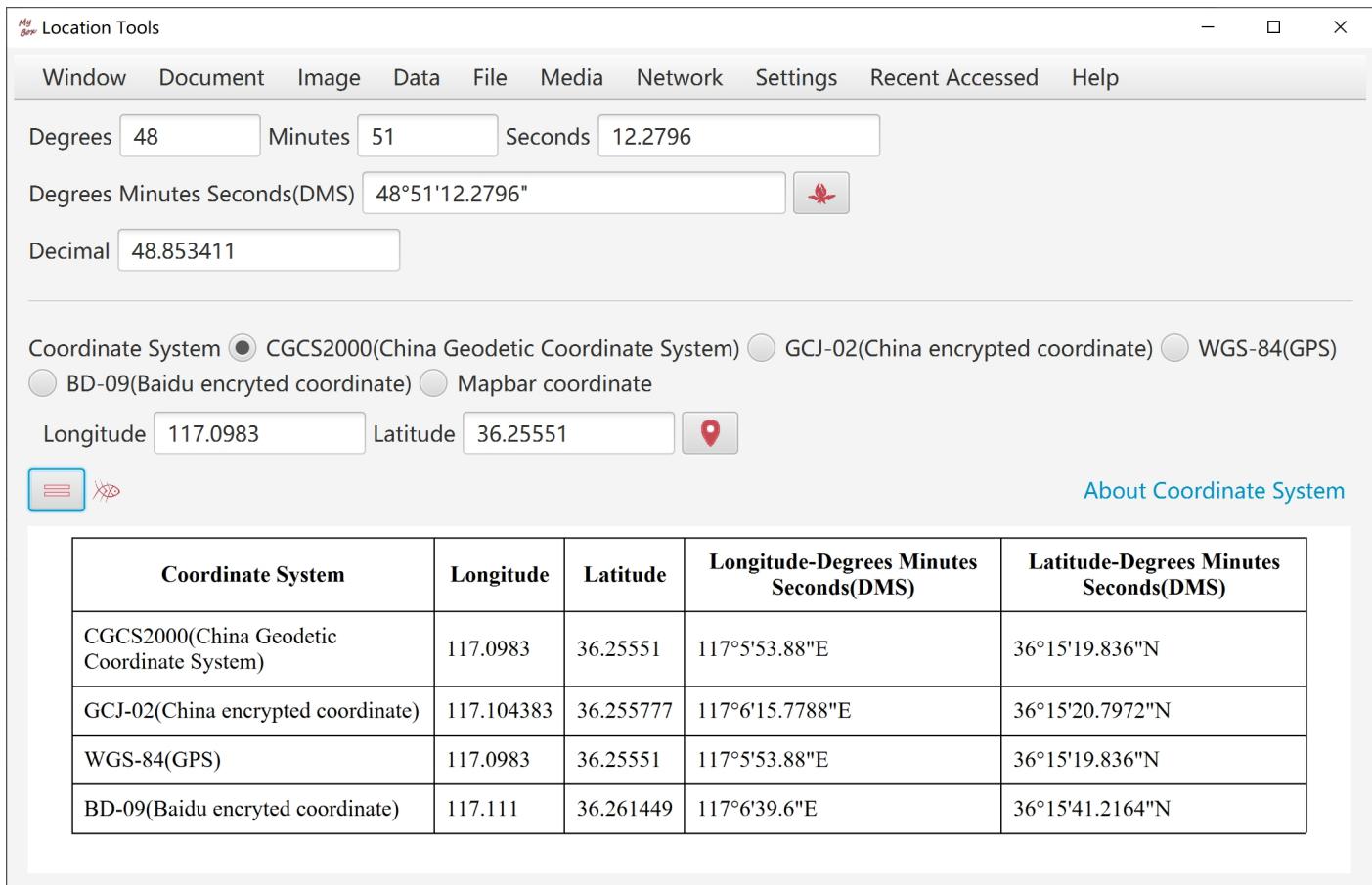
东经 118 度 48 分 54.152 秒

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.



## 5.8 Epidemic Reports

### 5.8.1 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.

## 5.8.2 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.

## 5.8.3 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.

## 5.8.4 Import Data

### 5.8.4.1 Predefined Data

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

### 5.8.4.2 CSV format:

- Download address:  
[https://github.com/Mararsh/MyBox\\_data/tree/master/md/COVID19/en](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.

### 5.8.4.3 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](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.

#### 5.8.4.4 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](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.

#### 5.8.4.5 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.

#### 5.8.5 Statistics Data

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

#### 5.8.6 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 geograhpy 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.

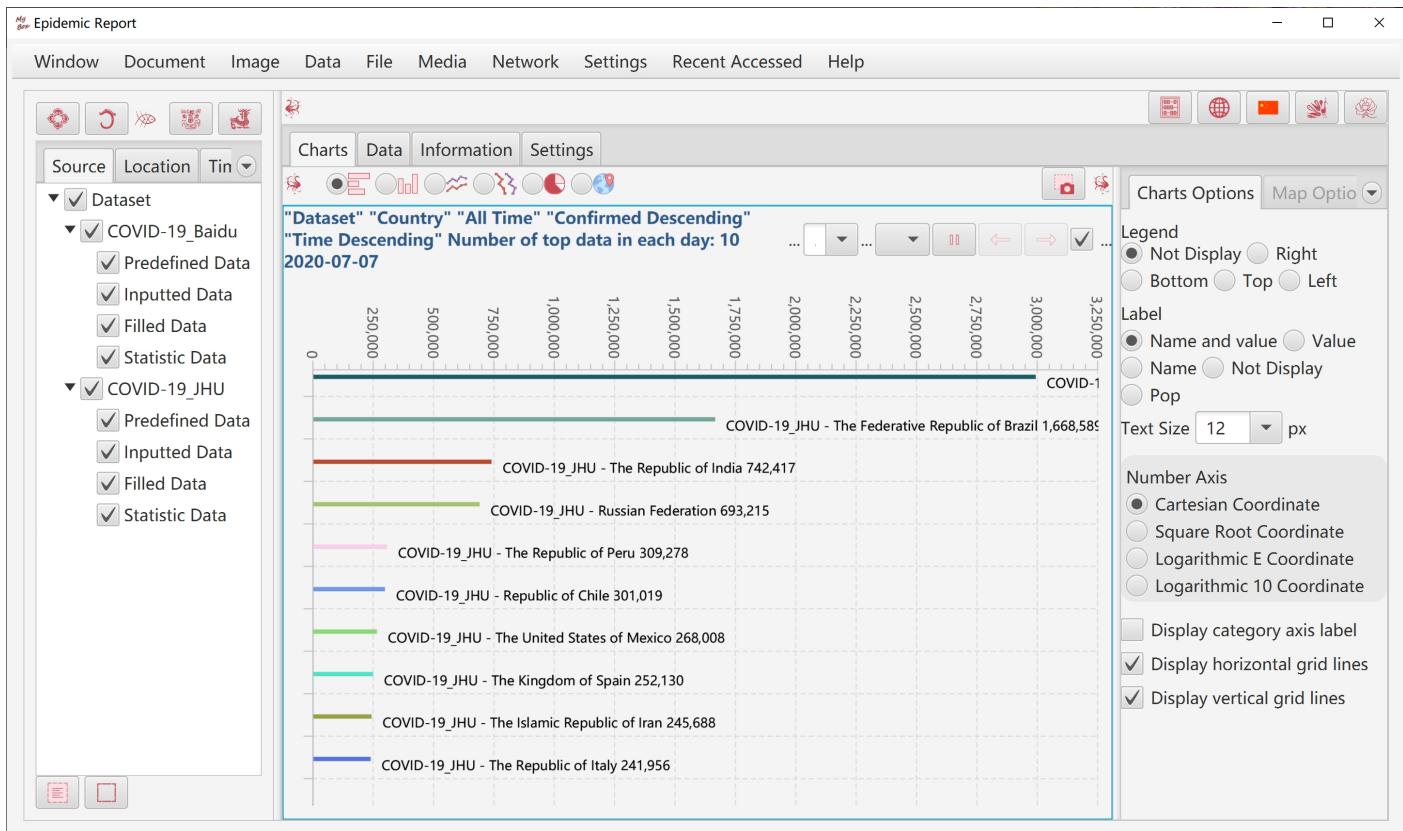
- "Time Descending" is always as the first ordering element automatically.
  - At least one more column should be picked as ordering element.
  - Beside "Time Descending", the first ordering element is called "major quering attribute".
- Number of Top Data in Each Day" and "Elements of ordering" work 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.

### 5.8.7 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)

### 5.8.8 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.



## 6 Others

### 6.1 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.

### 6.2 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.

### **6.3 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.

### **6.4 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.

### **6.5 Extract ttf files from ttc file**

**<End of Document>**