



**MyBox：简易工具箱
用户手册 - 数据工具**

作者：Mara

版本：6.6

日期：2022-9-28

内容目录

1 简介.....	6
1.1 主菜单.....	6
1.2 资源地址.....	7
1.3 文档.....	8
1.4 工具的菜单.....	8
2 二维存储结构的数据.....	9
2.1 数据对象.....	9
2.2 功能菜单.....	10
2.3 编辑数据.....	11
2.3.1 保存与恢复.....	11
2.3.2 定义数据列.....	12
2.3.3 定义数据属性.....	14
2.3.4 表格编辑模式.....	15
2.3.5 文本编辑模式.....	16
2.3.6 查看网页.....	17
2.3.7 查看文本.....	17
2.3.8 加载系统粘贴板的内容	18
2.3.9 导入示例	19
2.3.9.1 中国的统计数据	19
2.3.9.2 回归相关的数据	20
2.4 行表达式.....	21
2.5 行过滤.....	23
2.6 修改.....	24
2.6.1 赋值.....	25
2.6.2 删除.....	26
2.6.3 设置风格/标识异常值.....	27
2.6.3.1 管理风格.....	27
2.6.3.2 数据范围.....	28
2.6.3.3 定义风格.....	29
2.6.3.4 应用.....	29
2.6.4 粘贴系统粘贴板的内容.....	30
2.6.5 粘贴 MyBox 粘贴板的内容.....	30
2.7 整理.....	31
2.7.1 复制/过滤/查询.....	32
2.7.2 排序.....	33
2.7.3 转置.....	34
2.7.4 归一化.....	35
2.7.5 分割.....	36
2.8 计算.....	37
2.8.1 行表达式.....	38
2.8.2 描述性统计.....	39
2.8.3 等值分组.....	40
2.8.4 简单线性回归.....	43

2.8.4.1 回归.....	43
2.8.4.2 模型	44
2.8.4.3 拟合图	45
2.8.4.4 残差图	46
2.8.5 简单线性回归-组合.....	47
2.8.6 多重线性回归.....	48
2.8.6.1 回归.....	48
2.8.6.2 模型.....	49
2.8.7 多重线性回归-组合.....	50
2.8.8 频数分布.....	51
2.8.9 数值百分比.....	52
2.9 数据图.....	53
2.9.1 XY 图.....	54
2.9.1.1 数据.....	54
2.9.1.2 条图.....	54
2.9.1.3 堆叠条图.....	55
2.9.1.4 线图.....	56
2.9.1.5 散点图.....	56
2.9.1.6 气泡图.....	57
2.9.1.7 面积图.....	58
2.9.1.8 堆叠面积图.....	58
2.9.1.9 图数据的参数.....	59
2.9.1.10 布局.....	60
2.9.1.11 类别轴.....	61
2.9.1.12 数值轴.....	62
2.9.2 饼图.....	63
2.9.3 箱线图.....	64
2.9.4 自比较条图.....	66
2.9.5 相比较条图.....	67
2.9.6 XYZ 图.....	68
2.9.6.1 三维散点图.....	68
2.9.6.2 曲面图.....	70
2.10 输出.....	72
2.10.1 导出.....	73
2.10.2 转换为数据库表.....	74
2.11 管理数据.....	75
2.12 拼接数据.....	76
2.13 数据文件.....	77
2.13.1 CSV 文件.....	77
2.13.2 文本文件.....	78
2.13.3 Excel 文件.....	79
2.13.4 批量转换/分割数据文件	80
2.13.5 合并数据文件.....	81
2.14 系统粘贴板中数据.....	82
2.15 MyBox 粘贴板中的数据.....	83
2.16 矩阵.....	84

2.16.1 管理和编辑矩阵.....	84
2.16.2 矩阵的一元计算.....	85
2.16.3 矩阵的二元计算.....	86
2.17 数据库表.....	87
2.17.1 管理数据库表.....	87
2.17.2 SQL 标识符的限制.....	88
2.17.3 数据库 SQL.....	89
2.17.4 SQL 查询.....	90
3 脚本与表达式.....	91
3.1 JShell (Java 交互编程工具)	91
3.2 JEXL (Java 表达式语言)	93
3.3 Javascript.....	95
4 数学函数.....	96
4.1 定义函数:	96
4.2 计算函数.....	96
4.3 数据集.....	98
4.4 一元函数的 XY 图.....	99
4.5 二元函数的 XYZ 图.....	100
5 位置数据.....	101
5.1 数据约束.....	101
5.1.1 无效值	101
5.1.2 坐标系统.....	101
5.1.3 坐标值.....	101
5.1.4 时间.....	101
5.1.4.1 时间格式	101
5.1.4.2 纪元.....	101
5.1.4.3 有效的时间示例.....	102
5.2 数据操作.....	102
5.3 地图数据.....	103
5.4 地理编码.....	105
5.4.1 数据定义	105
5.4.2 数据约束	105
5.4.3 编辑数据	105
5.4.4 定义条件	105
5.4.5 导入数据	105
5.4.5.1 内置的预定义数据.....	105
5.4.5.2 CSV 格式.....	105
5.4.5.3 来自 geoname.org 的位置数据:.....	106
5.4.6 设置	106
5.5 地图上的位置.....	107
5.6 位置数据.....	108
5.6.1 数据定义	108
5.6.2 数据集	108
5.6.3 定义条件	108
5.6.4 地图数据	108
5.6.5 截图	108

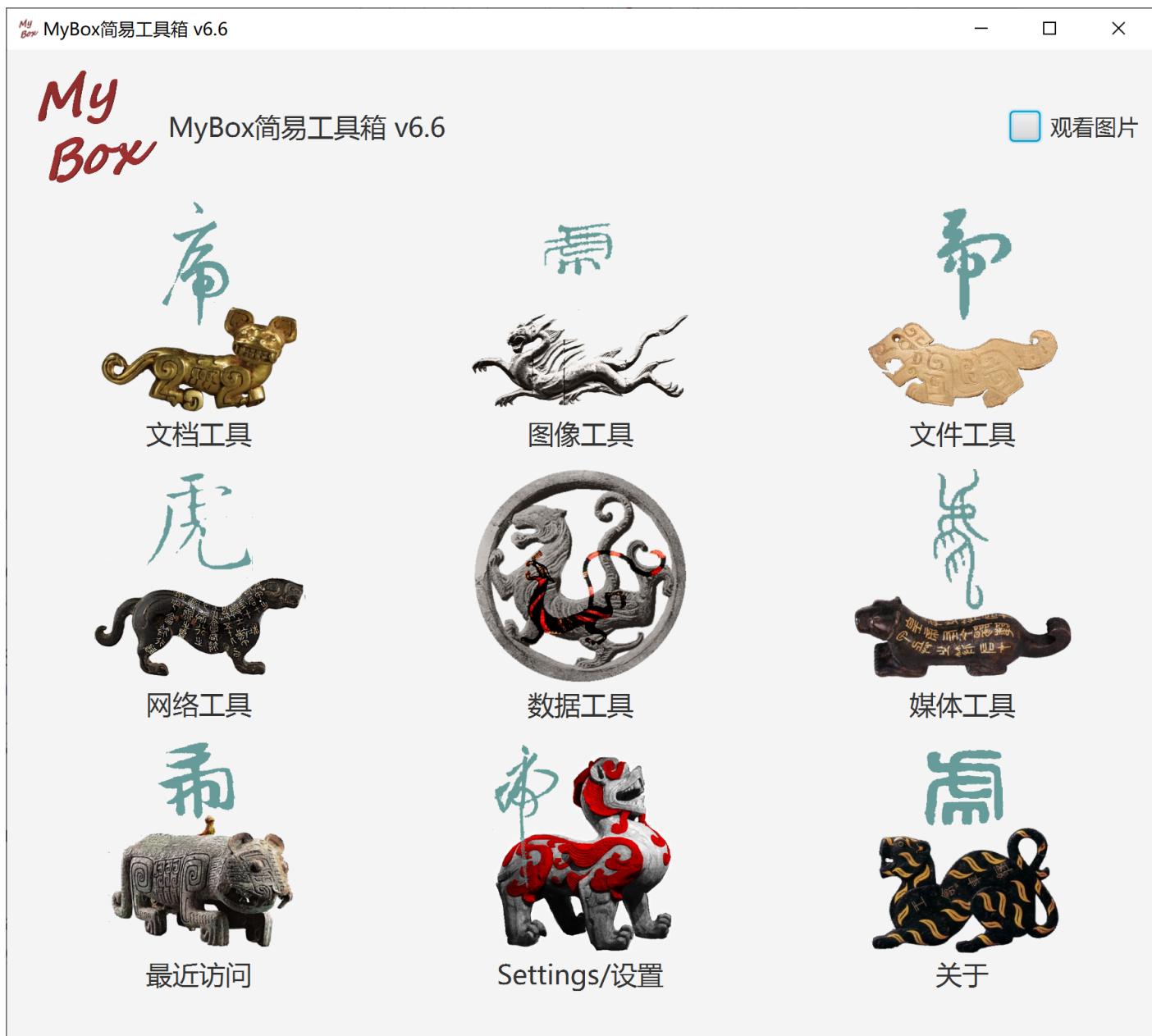
MyBox 用户手册 - 数据工具 v6.6

5.6.6 导入数据	109
5.6.6.1 CSV 格式	109
5.6.6.2 来自 movebank.org 的位置数据。	109
5.6.6.3 示例.....	109
5.7 位置工具.....	111
5.8 疫情报告.....	112
5.8.1 数据定义	112
5.8.2 数据约束	112
5.8.3 编辑数据	112
5.8.4 导入数据	112
5.8.4.1 MyBox 内置预定义数据.....	112
5.8.4.2 CSV 格式	113
5.8.4.3 来自约翰霍普金斯大学的 COVID-19 历史数据（全球）：	113
5.8.4.4 来自约翰霍普金斯大学的 COVID-19 每日数据（全球）。	113
5.8.4.5 导入的数据处理.....	113
5.8.5 统计数据	113
5.8.6 定义条件	114
5.8.7 显示图	114
5.8.8 设置	114
6 其它.....	116
6.1 编码条码	116
6.2 解码条码	116
6.3 消息摘要	116
6.4 编码/解码 Base64	116
6.5 从 ttc 文件中提取 ttf 文件	116

1 简介

这是利用 JavaFx 开发的图形化桌面应用，目标是提供简单易用的功能。免费开源。

1.1 主菜单



1.2 资源地址

内容	链接
项目主页	https://github.com/Mararsh/MyBox/
源代码和编译好的包	https://github.com/Mararsh/MyBox/releases
提交软件需求和问题报告	https://github.com/Mararsh/MyBox/issues
数据	https://github.com/Mararsh/MyBox_data
文档	https://github.com/Mararsh/MyBoxDoc
镜像	https://sourceforge.net/projects/mara-mybox/files/
云盘	https://pan.baidu.com/s/1fWMRzym_jh075OCX0D8y8A#list/path=%2F

文件(F) 编辑(E) 查看(V) 历史(S) 书签(B) 工具(I) 帮助(H)

https://github.com/Mararsh/MyBox

Mararsh / MyBox Public

Code Issues 41 Pull requests Discussions Actions Projects Wiki Security Insights Settings

master 1 branch 90 tags

Mararsh a6.5.9 790ff01 5 hours ago 573 commits

- alpha/MyBox a6.5.9 5 hours ago
- docs v6.5.8 18 days ago
- en v6.5.8 18 days ago
- released/MyBox v6.5.8 18 days ago
- .gitignore v6.5.6 3 months ago
- LICENSE Initial commit 4 years ago
- README.md v6.5.8 18 days ago

ReadMe in English

Releases 89

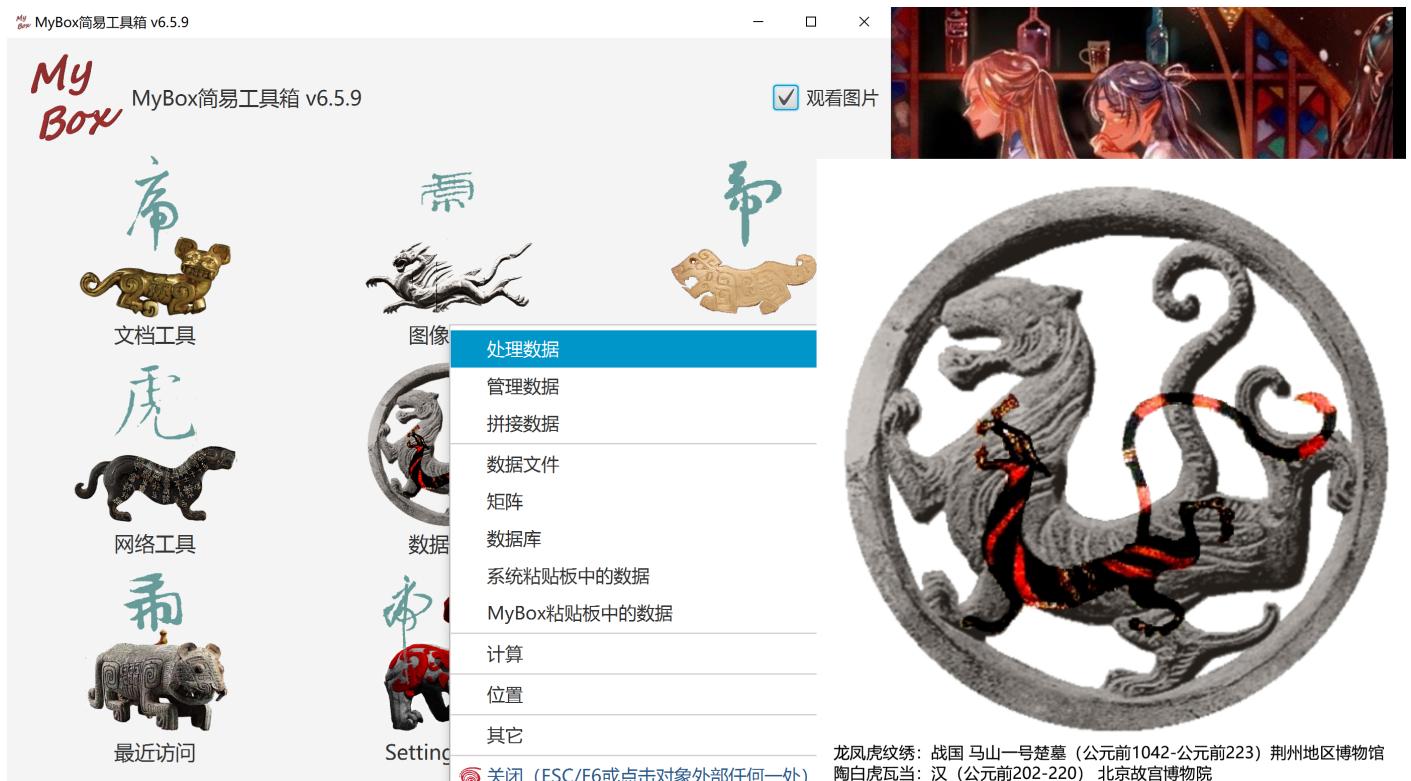
v6.5.8 Latest 18 days ago + 88 releases

https://github.com/Mararsh/MyBox/releases

1.3 文档

文档名	版本	修改时间	英文	中文
开发日志	6.6	2022-9-28	html	html
快捷键	6.5.6	2022-6-11	html	html
打包步骤	6.3.3	2020-9-27	html	html
开发指南	2.1	2020-8-27	PDF	PDF
用户手册-综述	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-文档工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-图像工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-文件工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-网络工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-数据工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-媒体工具	6.6	2022-9-28	html PDF odt	html PDF odt
用户手册-开发工具	6.6	2022-9-28	html PDF odt	html PDF odt

1.4 工具的菜单



2 二维存储结构的数据

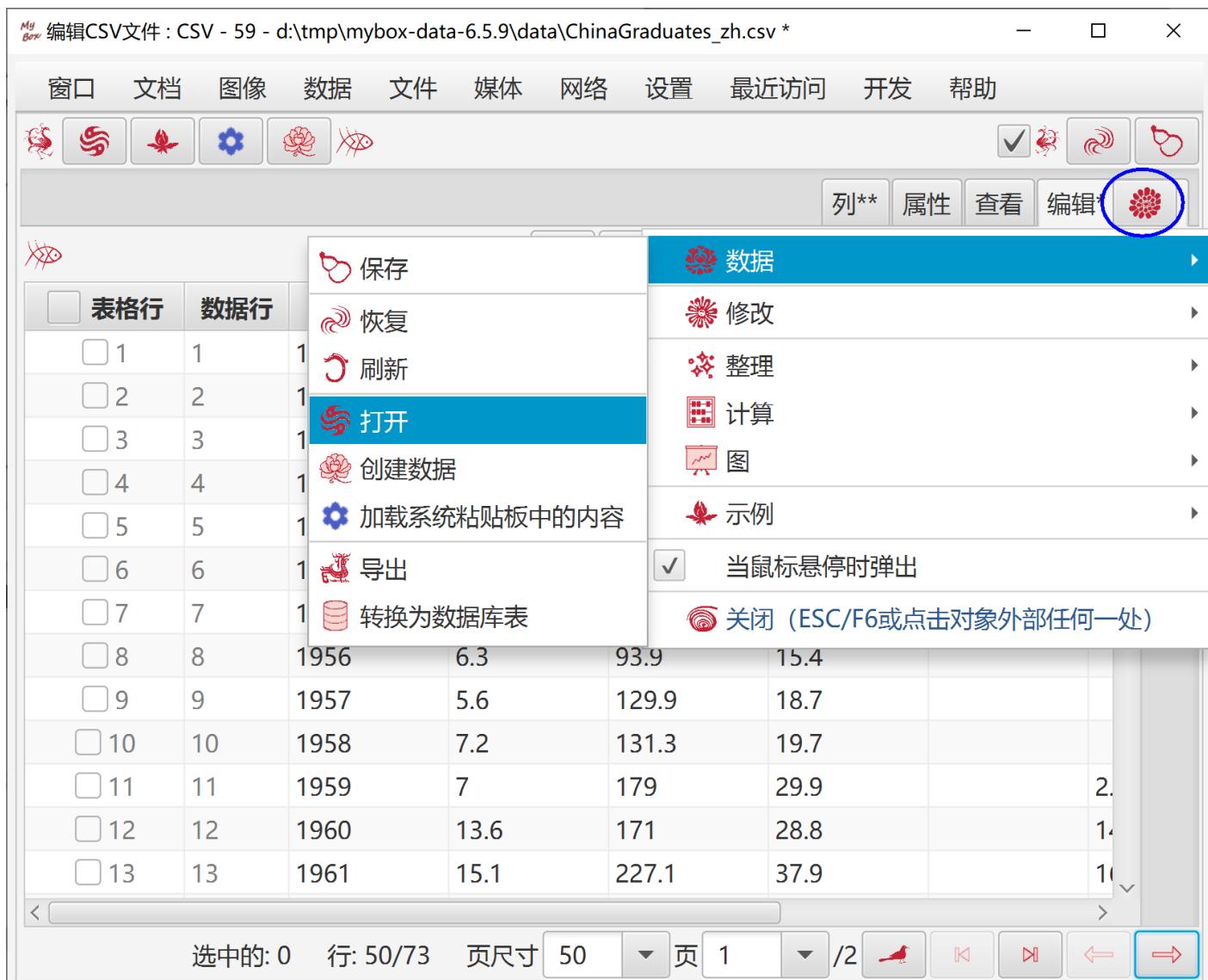
2.1 数据对象

1. 以下对象可以以一致的方式来编辑和计算:
 - 数据文件，包括 CSV 文件、Excel 文件、文本文件
 - MyBox 粘贴板的数据
 - 矩阵
 - 数据库表
2. 数据被表示为二维存储结构：
 - “列”在横向定义数据的维度。
 - “行”在纵向保存数据的值。
 - 数据应当等宽。即所有行的列数都相同。



2.2 功能菜单

悬停或点击“功能”按钮以弹出功能菜单。



2.3 编辑数据

2.3.1 保存与恢复

1. 有修改时，页签头显示*。若修改未应用，则显示 **。
2. 点击按钮“保存”以把修改写入文件和数据库：
 - “表格”中行的变化，包括修改/添加/删除/排序，影响文件中当前页的行。
 - “列”页签中的变化，包括修改/添加/删除/排序，影响文件中所有行。
 - “属性/列的修改被保存到数据库中。
3. 点击按钮“恢复”以丢弃所有修改并从文件和数据库中加载数据。

表格行	数据行	年	普通高等学...	普通中学毕...	普通高中毕...	初中毕业生...
<input type="checkbox"/> 1	1	1949	2.1	28	6.1	
<input type="checkbox"/> 2	2	1950	1.8	29.6	6.2	
<input type="checkbox"/> 3	3	1951	1.9	29.4	5.9	
<input type="checkbox"/> 4	4					
<input type="checkbox"/> 5	5					
<input type="checkbox"/> 6	6					
<input type="checkbox"/> 7	7					
<input type="checkbox"/> 8	8					
<input type="checkbox"/> 9	9					
<input type="checkbox"/> 10	10	1958	7.2	131.3	19.7	
<input type="checkbox"/> 11	11	1959	7	179	29.9	2.
<input type="checkbox"/> 12	12	1960	13.6	171	28.8	1.
<input type="checkbox"/> 13	13	1961	15.1	227.1	37.9	10

选中的: 0 行: 50/73 页尺寸 50 / 2

2.3.2 定义数据列

1. 列名不能为空也不能重复。
2. 列类型用来检验数据值：
 - 在编辑数据时，非法值将被拒绝。
 - 在读取和计算数据时，类型被忽略。
 - 数据类型影响排序的结果。
3. 点击按钮“确定”以把修改应用于“表格”。
4. 点击按钮“取消”以丢弃修改并从“表格”读取数据。
5. 可以重命名所有字段为序列号。
6. 可以随机设置颜色。

表格行	索引	列名	类型	色彩	宽度	可编辑	非空	以 +
<input type="checkbox"/>	1	年	字串		100	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	2	普通高等学校毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input checked="" type="checkbox"/>	3	普通中学毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	4	普通高中毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	5	初中毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	6	职业中学毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	7	普通小学毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>
<input type="checkbox"/>	8	特殊教育学校毕业生数(万人)	双精度		100	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="..."/>

点击以编辑

 0xC1D34BFF
 #C1D34B
 -4074677
 sRGB: 193 211 75 100%
 HSB: 68 64% 83%
 Adobe RGB: 197 210 87
 Apple RGB: 182 203 49
 ECI RGB: 202 218 96
 sRGB Linear: 136 166 18
 Adobe RGB Linear: 145 166 24
 Apple RGB Linear: 140 169 13
 Calculated CMYK: 9 0 64 17
 ECI CMYK: 30 4 82 0
 Adobe CMYK Uncoated FOGRA29: 30 3 80 0
 XYZ: 0.493449 0.5899 0.120933
 CIE-L*ab: 81.29 -19.4 62.29
 LCH(ab): 81.29 65.24 107.3
 CIE-L*uv: 81.29 -6.1 62.33
 L*a*b*: 81.29 65.24 107.3

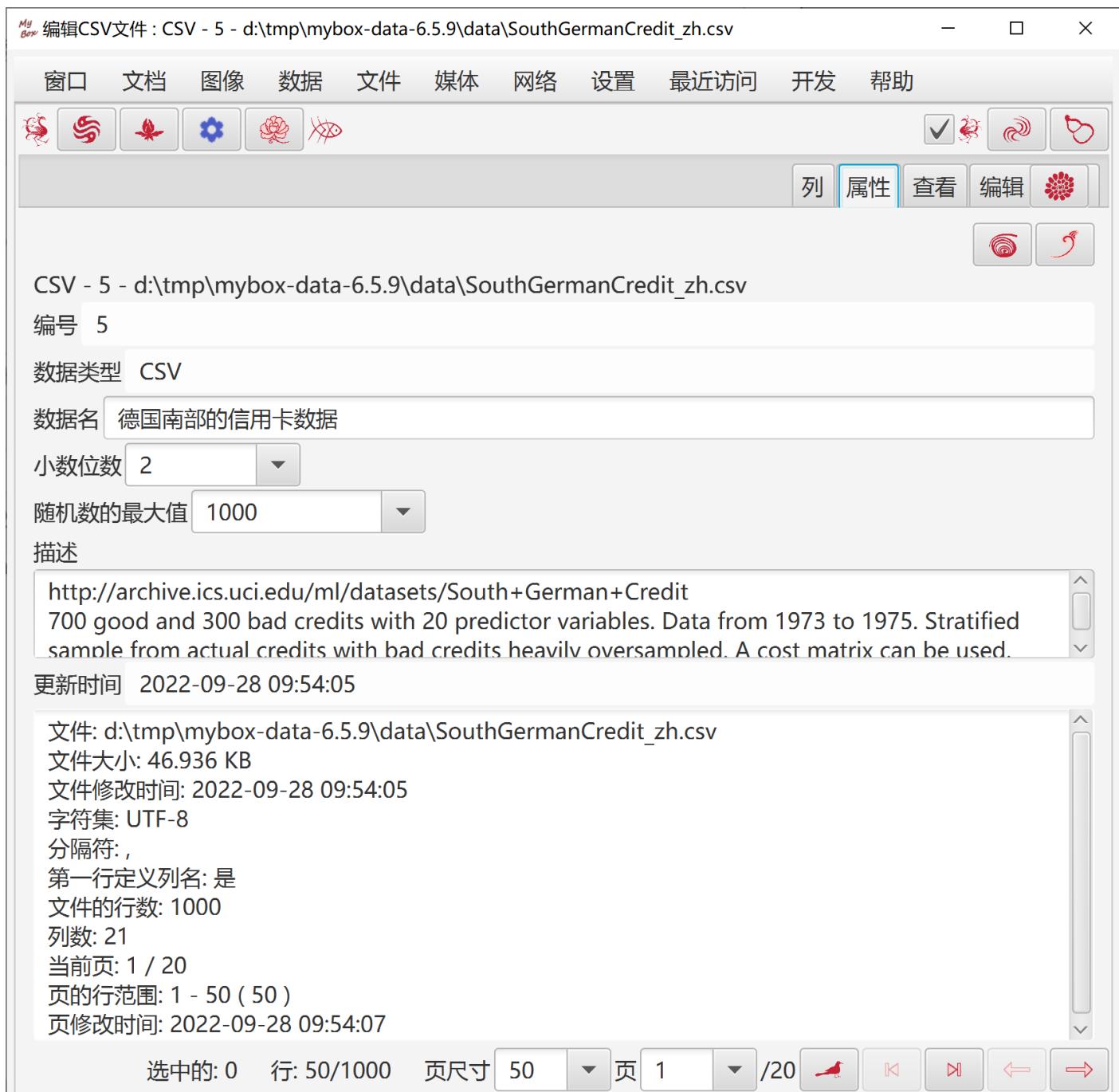
选中的: 0 行: 50/73 页尺寸: 50 页

7. 可以直接点击单元以编辑，也可以选择列并点击“编辑”按钮以弹出列的编辑窗口。



2.3.3 定义数据属性

包括：数据名、小数位数、随机数的最大值。



2.3.4 表格编辑模式

表格是主编辑模式：

1. 它的修改被自动应用其它面板。
2. 它是用于保存的最终数据。
3. 对于字符串类型的数据，除了文本文件，支持编辑和保存多行：
 - 若值是单行的（不包含换行符）：
 - 点击数据单元时显示单行输入框。
 - 在值中写入"\n"作为换行符、然后提交修改（回车或点击其它地方）。
 - 若值中包含换行符：
 - 点击数据单元时显示多行输入框。
 - - 直接编辑多行文本。

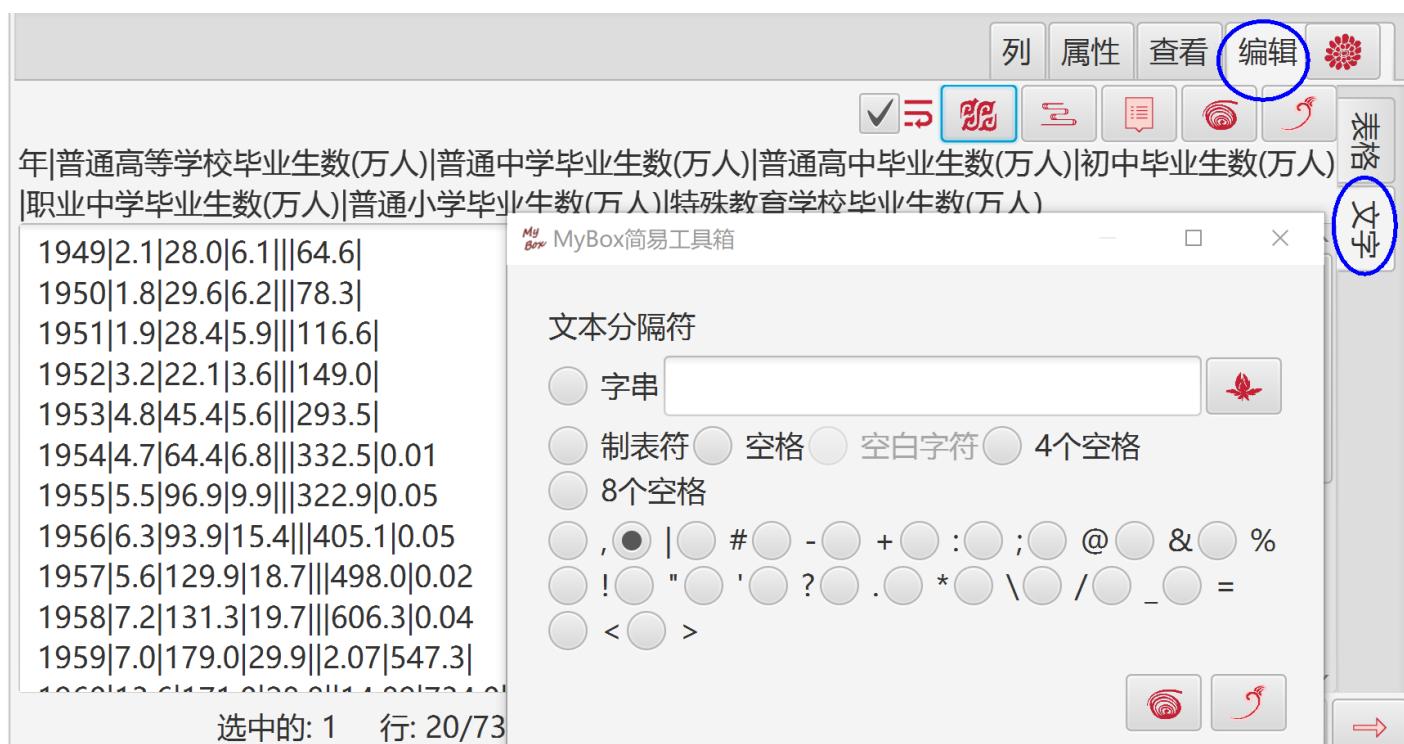
The screenshot shows the MyBox Data Tool interface in 'Table Edit' mode. The window title is 'MyBox 编辑CSV文件 : CSV - 9 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies_zh.csv *'. The menu bar includes '窗口' (Window), '文档' (Document), '图像' (Image), '数据' (Data), '文件' (File), '媒体' (Media), '网络' (Network), '设置' (Settings), '最近访问' (Recent), '开发' (Development), and '帮助' (Help). The toolbar contains icons for various operations like select, filter, sort, and search.

The main area displays a table with 10 rows of data from the Iris dataset. The columns are labeled: 表格行 (Table Row), 数据行 (Data Row), 花萼长度(cm) (Sepal Length cm), 花萼宽度(cm) (Sepal Width cm), 花瓣长度(cm) (Petal Length cm), 花瓣宽度(cm) (Petal Width cm), and 种类 (Species). Row 1 is selected, indicated by a checked checkbox in the first column. A context menu is open over the '种类' cell of Row 1, showing options: 表行号 1 (Row Number 1), 种类 (Species), Iris-setosa (selected), and 多行 (Multi-line). The bottom status bar shows '选中的: 1' (Selected: 1).

2.3.5 文本编辑模式

文本是辅编辑模式：

1. 点击按钮“确定”以把它的修改应用于“表格”。
2. 点击按钮“取消”以丢弃它的修改并从“表格”读取数据。
3. 点击按钮“分隔符”以从“表格”读取数据并应用新的分隔符，它的修改被丢弃。



2.3.6 查看网页

1. 显示当前页的数据。
2. 可选显示：表窗、标题、列名、行号。

年	普通高等学校毕业生数(万人)	普通中学毕业生数(万人)	普通高中毕业生数(万人)	初中毕业生数(万人)	职业中学毕业生数(万人)	普通小学毕业生数(万人)	特殊教育学校毕业生数(万人)
1949	2.1000	28.0000	6.1000			64.6000	
1950	1.8000	29.6000	6.2000			78.3000	
1951	1.9000	28.4000	5.9000			116.6000	
1952	3.2000	22.1000	3.6000			149.0000	

2.3.7 查看文本

1. 显示当前页的数据。
2. 可选显示：表窗、标题、列名、行号。
3. 可设置分隔符。

行号#年#普通高等学校毕业生数(万人)#普通中学毕业生数(万人)#普通高中毕业生数(万人)#初中毕业生数(万人)#职业中学毕业生数(万人)

行1#1949#2.1#28.0#6.1##64.6#
行2#1950#1.8#29.6#6.2##78.3#
行3#1951#1.9#28.4#5.9##116.6#
行4#1952#3.2#22.1#3.6##149.0#
行5#1953#4.8#45.4#5.6##293.5#
行6#1954#4.7#64.4#6.8##332.5#
行7#1955#5.5#96.9#9.9##322.9#
行8#1956#6.3#93.9#15.4##405.1#
行9#1957#5.6#129.9#18.7##498.0#
行10#1958#7.2#131.3#19.7##606.0#

文本分隔符

字串

制表符 空格 空白字符 4个空格
8个空格

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

2.3.8 加载系统粘贴板的内容

1. 读取并解析系统粘贴板的内容。
2. 自动猜测分隔符。
3. 可选数据分隔符：特殊字符、或输入的规则表达式。
4. 可选把第一行定义为列名。

MyBox 加载系统粘贴板中的内容

第一行定义列名   

分隔符  

源行号	年	普通高等学校毕业生数(万人)	普通中学毕业生数(万人)	普通高中毕业生数(万人)	初中毕业生数(万人)	职业中学生数(万人)	普通小学毕业生数(万人)	特殊教育学校毕业生数(万人)
行 1	1949	2.1	28.0	6.1			64.6	
行 2	1950	1.8	29.6	6.2			78.3	
行	1951	1.0	28.4	5.0			116.6	

普通高等学校毕业生数(万人) #
1949#2.1#28.0#6.1##64.6#
1950#1.8#29.6#6.2##78.3#
1951#1.9#28.4#5.9##116.6#
1952#3.2#22.1#3.6##149.0#
1953#4.8#45.4#5.6##293.5#
1954#4.7#64.4#6.8##332.5#0.0
1955#5.5#96.9#9.9##322.9#0.0
1956#6.3#93.9#15.4##405.1#0
1957#5.6#129.9#18.7##498.0#
1958#7.2#131.3#19.7##606.3#
1959#7.0#179.0#29.9##2.07#54
1960#13.6#171.0#28.8##14.89#
1961#15.1#227.1#37.9##16.12#
1962#17.7#202.5#44.1##4.5#55
1963#19.9#195.6#43.3##2.42#4
1964#20.4#175.3#36.7##4.94#5
1965#10.6#200.0#36.0##12.0#6

2.3.9 导入示例

2.3.9.1 中国的统计数据

The screenshot shows the MyBox Data Tool interface. A context menu is open over a table row. The menu items are:

- 中国的统计数据
- 回归
- 当鼠标悬停时弹出
- 关闭 (ESC/F6或点击对象外部任何一处)

The table data is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	
	1	2	1950	1.8	29.6								28.8
	2	3	1951	1.9	28.4								
	3	4	1952	3.2	22.1								
	4	5	1953	4.8	45.4								
	5	6	1954	4.7	64.4								
	6	7	1955	5.5	96.9								
	7	8	1956	6.3	93.9								
	8	9	1957	5.6	129.9								
	9	10	1958	7.2	131.3								
	10	11	1959	7	179								
	11	12	1960	13.6	171								

At the bottom, there is a status bar with the text: 选中的: 0 行: 50/73 页尺寸 50 / 1 / 2.

2.3.9.2 回归相关的数据

MyBox 编辑CSV文件 : CSV - 59 - d:\tmp\mybox-data-6.5.9\data\ChinaGraduates_zh.csv *

窗口 文档 图像 数据 文件 媒体 网络 设置 最近访问 开发 帮助

表格

	1	2	3	4	5	6	7	8	9	10	11	12
<input type="checkbox"/>	1	2	1950	1.8	29.6							
<input type="checkbox"/>	2	3	1951	1.9	28.4							
<input type="checkbox"/>	3	4	1952	3.2	22.1							
<input type="checkbox"/>	4	5	1953	4.8	45.4							
<input type="checkbox"/>	5	6	1954	4.7	64.4							
<input type="checkbox"/>	6	7	1955	5.5	96.9							
<input type="checkbox"/>	7	8	1956	6.3	93.9							
<input type="checkbox"/>	8	9	1957	5.6	129.9							
<input type="checkbox"/>	9	10	1958	7.2	131.3							
<input type="checkbox"/>	10	11	1959	7	179	29.9						
<input type="checkbox"/>	11	12	1960	13.6	171	28.8						

当鼠标悬停时弹出
关闭 (ESC/F6或点击对象外部任何一处)

列** 属性 查看 编辑*

中国的统计数据
回归

收入与快乐
工作经验与薪资
鸢尾花的种类
糖尿病的病情预测
糖尿病的病情预测-标准化
心力衰竭
混凝土抗压强度
狗狗手术后X光片的变化
垒球球员的薪水
德国南部的信用卡数据
波士顿房屋价格

选中的: 0 行: 50/73 页尺寸 50 页 1 /2

2.4 行表达式

在加工/整理/计算数据、生成数据图时，可以输入 JavaScript 表达式作为数据值：

1. 若脚本为空，则返回空字符串。

2. 编辑脚本：

- 脚本可以包含 Nashorn 可以解析的任何合法元素(ECMAScript 5.1)。
- 脚本应当最终是一个数值。
- 脚本可以包含以下占位符：

```
#{表行号}
#{数据行号}
#{<列名>}
#{<列名>-<统计名>}
```

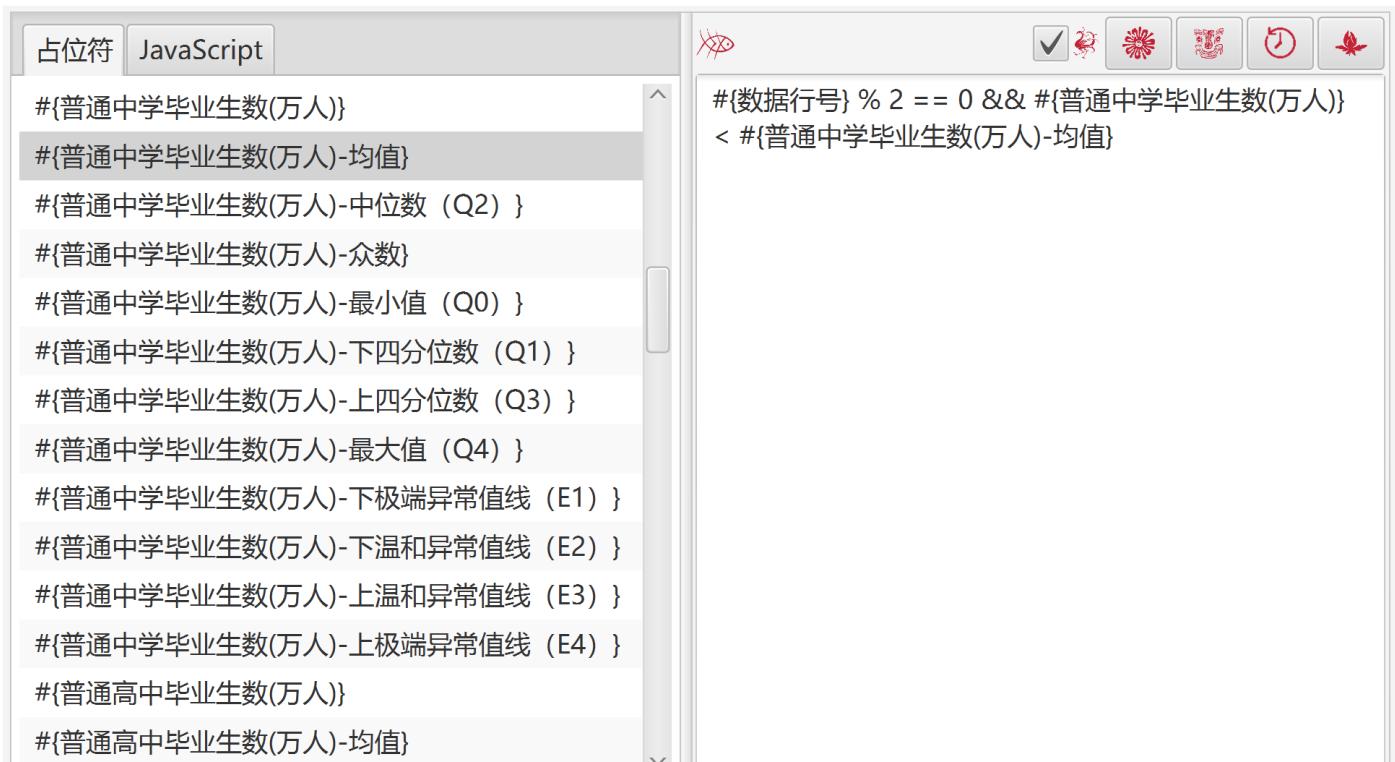
- 左边显示所有合法的占位符。
- 点击“编辑”按钮以保存脚本。

3. 在 MyBox 计算表达式时：

- 占位符被数据行的实际值替换。
- 统计值是用所有数据来计算的。
- '#{xxx}' 被处理为字符串而 #{xxx} 被处理为数字。
- 处理所有页时，脚本若包含"#{表行号}"则会失败。

4. 悬停或点击按钮“示例”，可以粘贴示例代码。

5. 悬停或点击按钮“历史”，可以粘贴历史代码。



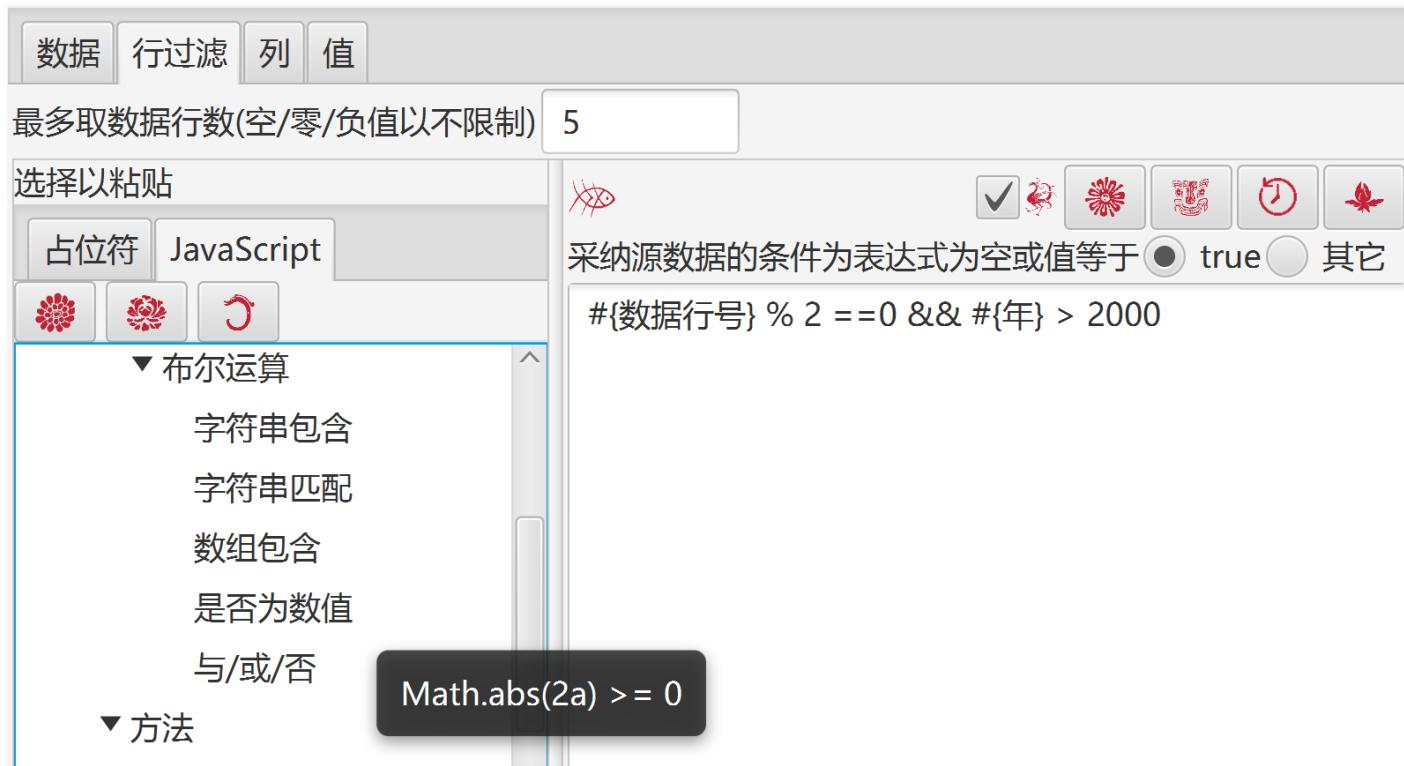
6. 示例:

表达式	含义
#{数据行号} % 2 == 0	数据行号为偶数
#{表行号} % 2 == 1	当前页的奇数行
Math.abs(#{字段 1}) + Math.PI * Math.sqrt(#{字段 2})	数学计算
'#{字段 1}'replace(/hello/ig, 'Hello')	把列"字段 1"的值中所有"hello"(忽略大小写)替换"Hello"
'#{字段 1}'.toLowerCase()	列"字段 1"的值的小写
'#{字段 1}'.split(',')	把列"字段 1"的值按逗号分隔
#{字段 1} - #{字段 1-均值}	列"字段 1"的值与列"字段 1"的平均值之间的差值

2.5 行过滤

“行过滤”是特殊的“行表达式”，用来作为过滤数据行的条件：

1. 脚本应当最终返回布尔值(“true”或“false”)。
2. 可以设置取用数据的最多行数。

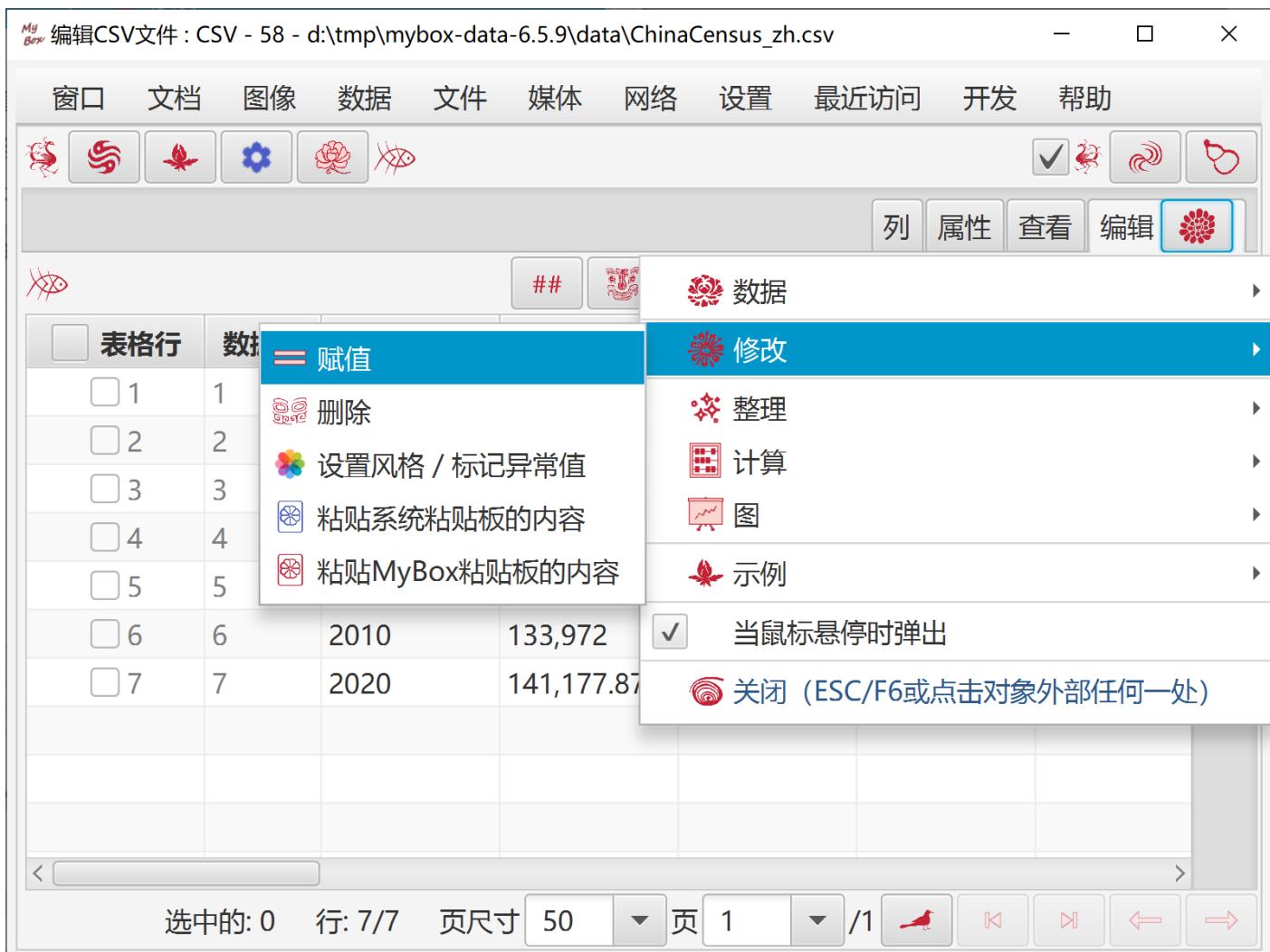


3. 示例：

表达式	含义
<code>#{行号} % 2 == 0</code>	数据行号为偶数
<code>#{表行号} % 2 == 1</code>	当前页的奇数行
<code>Math.abs(#{字段 1}) >= 0</code>	列“字段 1”的值是数值
<code>#{字段 1} > 0</code>	列“字段 1”的值大于零
<code>#{字段 1} - #{字段 2} < 100</code>	列“字段 1”与“字段 2”的值差小于 100
<code>'#{字段 1}'.length > 0</code>	列“字段 1”的值不为空值
<code>'#{字段 1}'.search(/Hello/ig) >= 0</code>	列“字段 1”的值包含字符串“Hello”（忽略大小写）
<code>'#{字段 1}'.startsWith('Hello')</code>	列“字段 1”的值以“Hello”开头
<code>var array = ['A', 'B', 'C']; array.includes('#{字段 1}')</code>	列“字段 1”的值为‘A’或‘B’或‘C’
<code>#{字段 1} < #{字段 1-均值}</code>	列“字段 1”的值小于列“字段 1”的平均值

2.6 修改

悬停或点击功能按钮，选择“修改”下的功能。



2.6.1 赋值

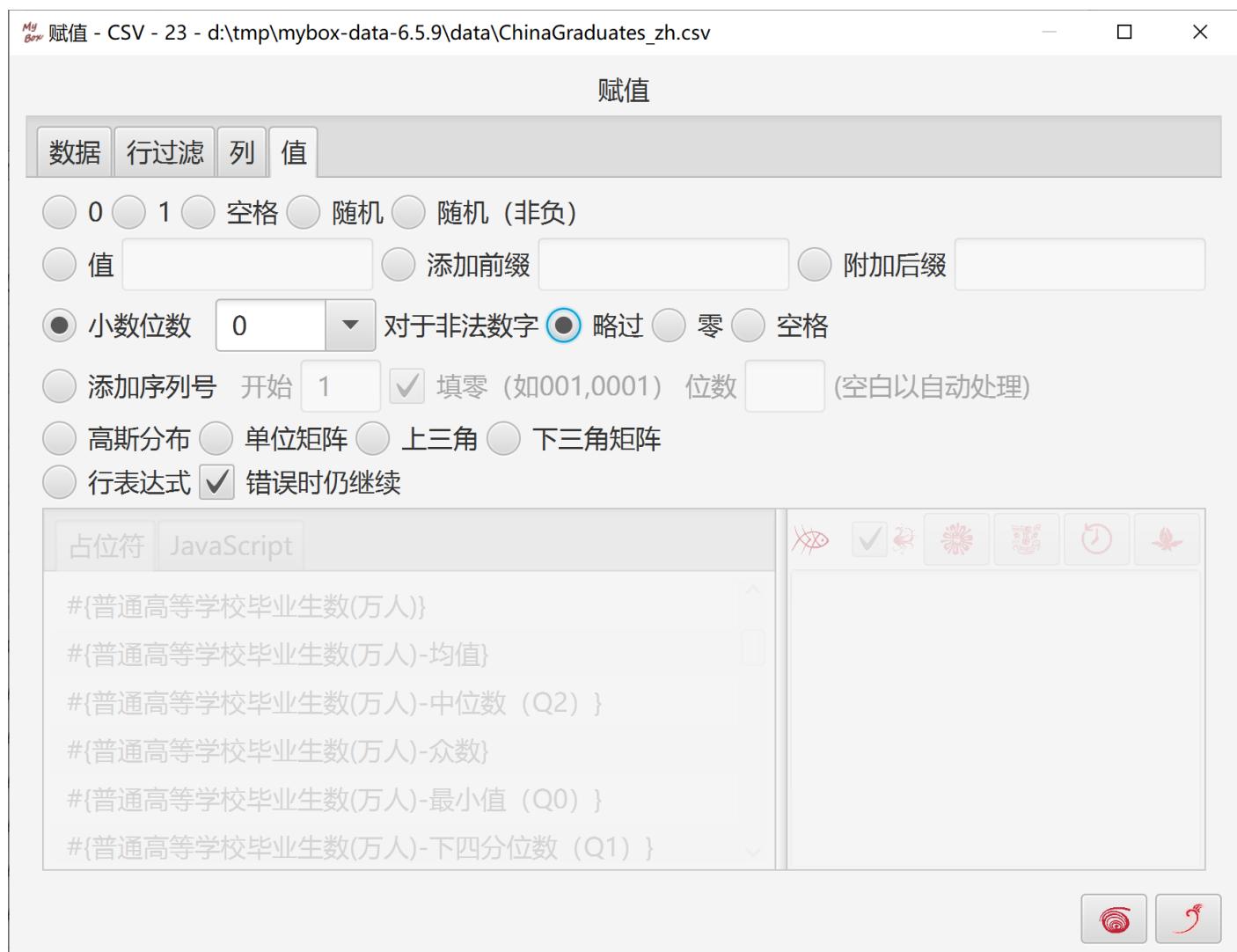
1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 选择列，若不选则取所有列。
- 设置行过滤条件。

2. 对所选数据赋值为：

- 常数：0、1、空格、或输入的值
- 随机数、随机数（非负）
- 添加前缀、附加后缀、设置小数位数、添加序列号
- 当所选数据为方阵（行列数相同），可赋值为：高斯分布、单位矩阵、上三角矩阵、下三角矩阵
- 行表达式

3. 若处理数据文件的所有页，则在赋值之前对数据文件自动备份。



2.6.2 删除

1. 选择数据:
 - 行可以是：当前页、选中的行、或所有页。
 - 设置行过滤条件。
2. 选项：是否在错误时继续处理。
3. 若处理数据文件的所有页，则在删除之前对数据文件自动备份。

MyBox 删除 - CSV - 80 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

删除

数据 行过滤 选项

行 选中的 当前页 所有页

<input type="checkbox"/> 表格行	<input type="checkbox"/> 数据行	年	国民生产总值	第三产业占比
<input type="checkbox"/>	1	1952	679.1	195.1
<input type="checkbox"/>	2	1953	824.4	141.1
<input type="checkbox"/>	3	1954	859.8	210.8
<input type="checkbox"/>	4	1955	911.6	221.5
<input type="checkbox"/>	5	1956	1,030.7	269.1
<input type="checkbox"/>	6	1957	1,071.4	266.5
<input type="checkbox"/>	7	1958	1,312.3	382.8
<input type="checkbox"/>	8	1959	1,447.5	447
<input type="checkbox"/>	9	1960	1,470.1	476.8
<input type="checkbox"/>	10	1961	1,232.3	397.7
<input type="checkbox"/>	11	1962	1,162.2	345.1
<input type="checkbox"/>	12	1963	1,248.3	337.9
<input type="checkbox"/>	13	1964	1,300.0	301.6

选择一些以处理，或者不选任何以处理表中所有:

- 选择数据行:
 - 按住SHIFT键，并点击数据行，以选择多行的范围。
 - 按住CTRL键，并点击数据行，以一行一行地选择多行。
- 点击行开头的复选框。
- 编辑数据:
 - 可编辑列的题头文字为蓝色（可编辑的主键仍为红色）。
 - 点击可编辑的单元格，则它的输入框会被显示出来。
 - 当可编辑单元失去焦点时，它的值将被检查，非法值将被拒绝。
- 要显示/隐藏列，点击表格头行右边的按钮“+”。
- 此操作不影响数据。

要调整列的顺序，点击并拖动列头。此操作不影响数据。

要对表行进行排序:

- 点击列头以按此列的升序排序。
- 再次点击列头以按此列的降序排序。
- 第三次点击列头以取消此列的排序。
- 按住SHIFT键并点击更多的列头以按多列排序。

此操作只影响表中当前数据。

选中的: 0 行: 20/70 - 主键的题头文字为红色。自增主键的题头文字为橙色。

← →

2.6.3 设置风格/标识异常值

2.6.3.1 管理风格

1. 添加、编辑、删除、复制风格。
2. 定义风格的标题和序号。
3. 选择风格是否标识异常值。

设置风格 / 标记异常值 - CSV - 1 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv

设置风格 / 标记异常值

序号	异常值	标题	从	到	+
1.0		女性人口分析A	1	-1	女性
2.0		男性人口分析A	1	-1	男性
3.0	<input checked="" type="checkbox"/>	女性人口分析B	25	-1	女性
4.0	<input checked="" type="checkbox"/>	男性人口分析B	25	-1	男性
< >					
页	1	/1			
页尺寸	50		行: 4/4	选中的: 1	

基础 数据 过滤 风格

编号: 3

序号: 3.0

标题: 女性人口分析B

此规则定义异常值

2.6.3.2 数据范围

定义条件以约束风格应用于哪些数据单元：

1. 数据行的范围
2. 列名
3. 行过滤。

注意：在添加或删除一些数据行以后，数据的行号可能会改变。例如，在第 6 行之前插入两行，则原来的第 12 行变成了第 14 行，而现在的第 12 行是原来的第 10 行。

所以“行号”不是定位特定数据行的正确方式。引用特定的行的一个方法是列值构成的表达式。

The screenshot shows the '设置风格 / 标记异常值' (Style Settings / Marking Abnormal Values) dialog for a CSV file named 'ChinaPopulation_zh.csv'. The dialog has several tabs at the top: 基础 (Basic), 数据 (Data), 过滤 (Filter), and 风格 (Style). The '数据' tab is selected.

基础 (Basic) Tab:

- 数据行号 (Data Row Number):** Set from 25 to the end of the file.
- 要计算的列 (Calculated Columns):** Selected columns include 年 (Year), 女性人口(万人) (Female Population in Ten Thousand), 男性人口(万人) (Male Population in Ten Thousand), and 乡村人口(万人) (Rural Population in Ten Thousand).

Table Data:

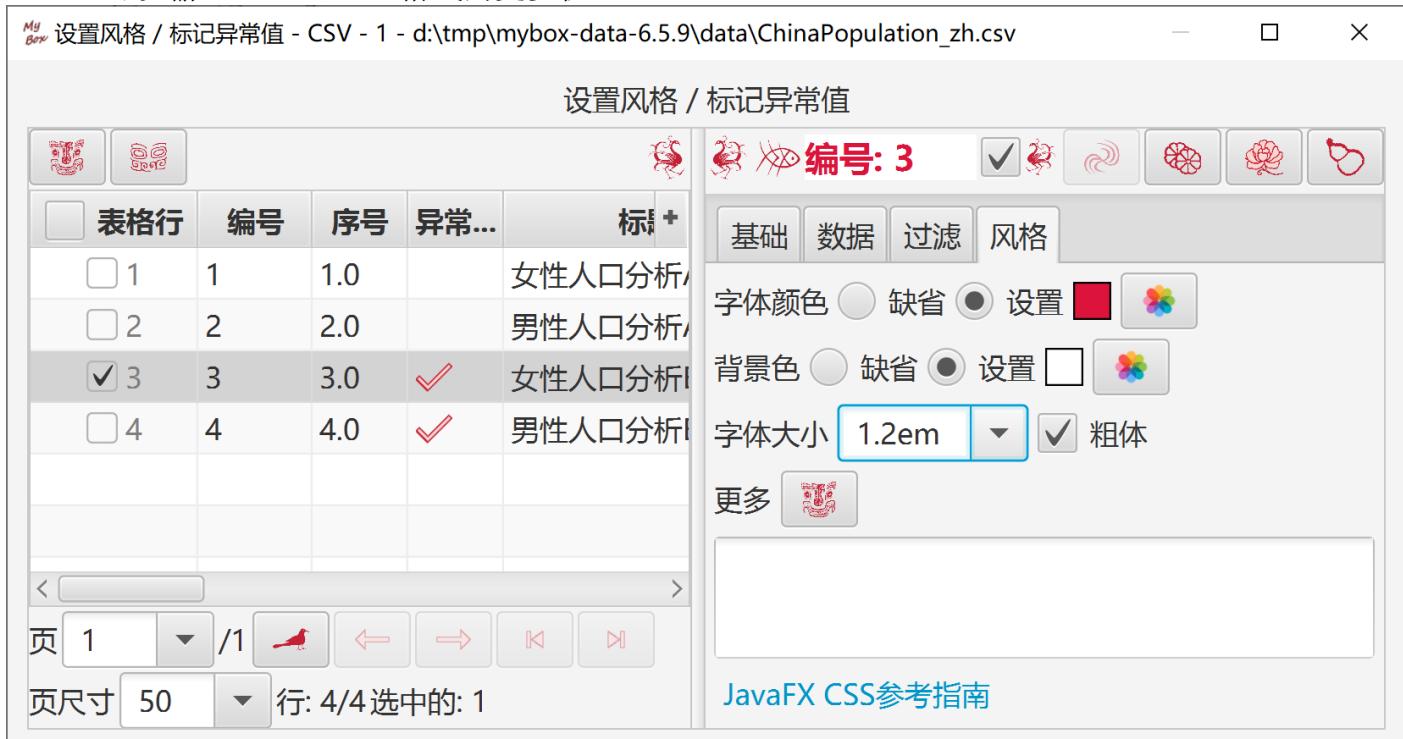
序号	异常值	标题	从	到	操作
1.0		女性人口分析A	1	-1	女性
2.0		男性人口分析A	1	-1	男性
3.0	✓	女性人口分析B	25	-1	女性
4.0	✓	男性人口分析B	25	-1	男性

Page Navigation:

- Page: 1 / 1
- Page Size: 50
- Selected Rows: 1

2.6.3.3 定义风格

1. 字体颜色、字体大小、背景颜色、是否粗体
2. 可以输入 JavaFx CSS 格式的更多值。



2.6.3.4 应用

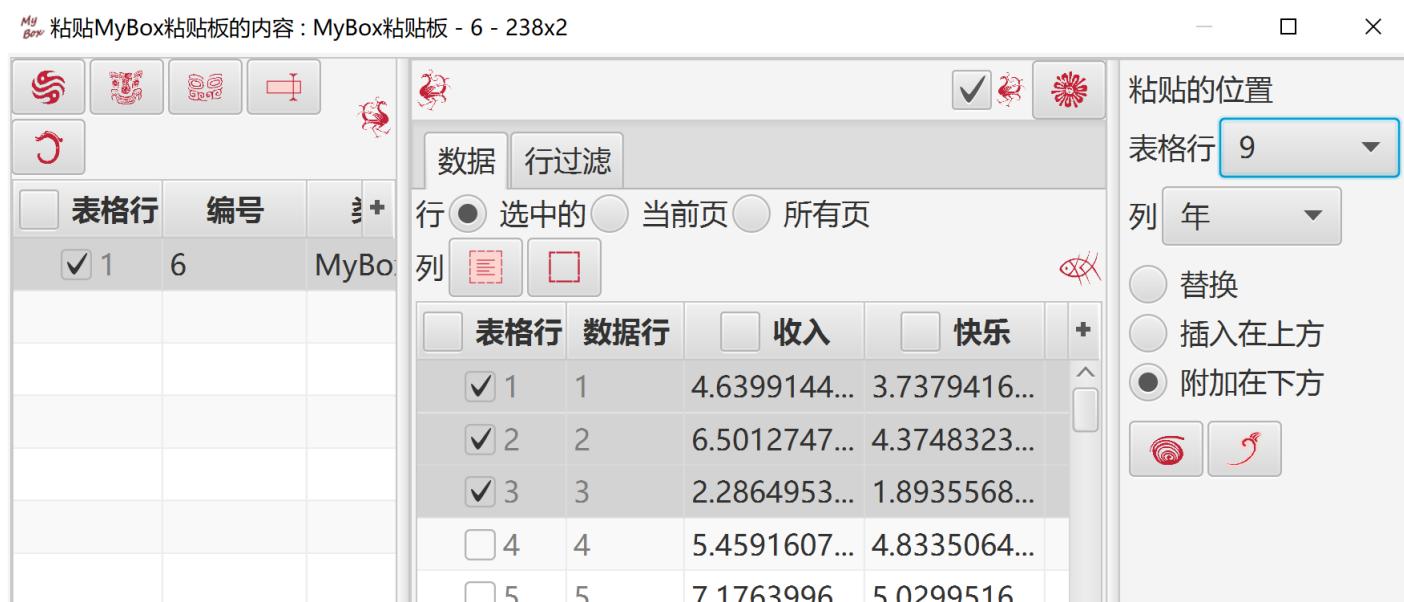
在显示数据页面时，数据的所有风格按序号被逐条应用于数据行。

表格行	数据行	年	年末总人口(万人)	男性人口(万人)	女性人口(万人)	城镇人口(万人)	乡村人口(万人)
<input type="checkbox"/>	24	1972	87,177	44,813	42,364	14,935	72,242
<input type="checkbox"/>	25	1973	89,211	45,876	43,335	15,345	73,866
<input type="checkbox"/>	26	1974	90,859	46,727	44,132	15,595	75,264
<input type="checkbox"/>	27	1975	92,420	47,564	44,856	16,030	76,390
<input type="checkbox"/>	28	1976	93,717	48,257	45,460	16,341	77,376
<input type="checkbox"/>	29	1977	94,974	48,908	46,066	16,669	78,305
<input type="checkbox"/>	30	1978	96,259	49,567	46,692	17,245	79,014
<input type="checkbox"/>	31	1979	97,542	50,192	47,350	18,495	79,047
<input type="checkbox"/>	32	1980	98,705	50,785	47,920	19,140	79,565
<input type="checkbox"/>	33	1981	100,072	51,519	48,553	20,171	79,901
<input type="checkbox"/>	34	1982	101,654	52,352	49,302	21,480	80,174
<input type="checkbox"/>	35	1983	103,008	53,152	49,856	22,274	80,734
<input type="checkbox"/>	36	1984	104,357	53,848	50,509	24,017	80,340

2.6.4 粘贴系统粘贴板的内容



2.6.5 粘贴 MyBox 粘贴板的内容



2.7 整理

悬停或点击功能按钮，选择“整理”下的功能。

The screenshot shows the MyBox Data Tool interface with a CSV file named 'ChinaGDP_zh.csv' loaded. The '整理' (Organize) option is selected from a context menu. The menu also includes options like '计算' (Calculate), '图' (Chart), and '示例' (Example). A checkbox for '当鼠标悬停时弹出' (Show when mouse hovers) is checked.

	表格行	数据行	年	国民总收入	人均GDP	GDP增长率	人均GDP增长率
<input type="checkbox"/>	1	1	1952	1,312.3	383.8	616.7	519.3
<input type="checkbox"/>	2	2	1953	1,447.5	1,447.5	383.8	616.7
<input type="checkbox"/>	3	3	1954	1,470.1	1,470.1	340.7	652.6
<input type="checkbox"/>	4	4	1955	1,232.3	1,232.3	441.1	393.5
<input type="checkbox"/>	5	5	1956	1,162.2	1,162.2	453.1	363.9
<input type="checkbox"/>	6	6	1957	1,248.3	1,248.3	497.5	412.8
<input type="checkbox"/>	7	7	1958	1,469.9	1,469.9	559	519.3
<input type="checkbox"/>	8	8	1959	1,447.5	1,447.5	383.8	616.7
<input type="checkbox"/>	9	9	1960	1,470.1	1,470.1	340.7	652.6
<input type="checkbox"/>	10	10	1961	1,232.3	1,232.3	441.1	393.5
<input type="checkbox"/>	11	11	1962	1,162.2	1,162.2	453.1	363.9
<input type="checkbox"/>	12	12	1963	1,248.3	1,248.3	497.5	412.8
<input type="checkbox"/>	13	13	1964	1,469.9	1,469.9	559	519.3

选中的: 0 行: 50/70 页尺寸: 50 / 1 / 2 < > << >>

2.7.1 复制/过滤/查询

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 选择列，若不选则取所有列。
- 设置行过滤条件。

2. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

3. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。



2.7.2 排序

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择要排序的列、是否降序，可以多选。
- 列的数据类型影响排序的结果。
- 可设置最多取结果的行数。
- 选择要复制的列。

3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。

The screenshot shows the 'Sort' dialog box for a CSV file named 'ChinaGDP_zh.csv'. The main interface includes tabs for 'Data', 'Row Filter', and 'Options'. The 'Sort' tab is active, displaying sorting criteria for column 4 ('National Income (GNP) in billions') set to ascending order.

列
1 年-降序
2 年-升序
3 国民总收入(GNI 亿元)-降序
<input checked="" type="checkbox"/> 4 国民总收入(GNI 亿元)-升序
5 国内生产总值(GDP 亿元)-降序

Below the sort table, there's a section for copied columns (不选表示所有), with checkboxes for 'National Income (GNP)' and 'GDP'. There are also fields for specifying the number of results and choosing row numbers or column names.

The 'Target' section at the bottom allows saving to various formats like CSV, XLS, TXT, or databases, with the current target set to 'Chinese Domestic Product (GDP)'.

2.7.3 转置

1. 选择数据:
 - 行可以是：当前页、选中的行、或所有页。
 - 选择列，若不选则取所有列。
 - 设置行过滤条件。
2. 计算：可选“把第一列当作列名”。
3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。
4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。



2.7.4 归一化

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择要计算的列。
- 可选按照：列、行、所有。
- 算法可选：最大最小值（可设置区间）、和（L1）、ZScore（L2）。
- 可选对非数字：略过、计为零。
- 选择要复制的列。

3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。



2.7.5 分割

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 选项:

- 选择要复制的列。
- 按照"分割的尺寸":
 - 被选择的行被划分为多个文件，每个文件包含定义的数据行数。
 - 最后一个文件可能包含较少的数据行。
- 按照"分割个数":
 - 被选择的行被尽可能划分为定义的文件个数。
 - 若定义了行过滤，则生成的文件数可能会比预期的少。
- 按照"起止列表":
 - 被选择的行被按照列表中的范围进行划分。
 - 这些编号与"数据行号"进行比较。
 - 编号基于1。范围包含"开始"和"结束"。



2.8 计算

悬停或点击功能按钮，选择“计算”下的功能。

The screenshot shows the MyBox Data Tool interface with a CSV file named 'ChinaGDP_zh.csv' loaded. The 'Compute' option is highlighted in the context menu, which also includes other statistical analysis tools like 'Expression', 'Descriptive Statistics', 'Equation Grouping', 'Simple Linear Regression', 'Simple Linear Regression - Combination', 'Multiple Linear Regression', 'Multiple Linear Regression - Combination', 'Frequency Distribution', and 'Value Percentage'.

年	国民总收入 (亿元)
1952	679.1
1953	1,050.0
1954	1,447.5
1955	1,470.1
1956	1,232.3
1957	1,162.2
1958	1,248.3
1959	1,469.9
1960	559
1961	519.3

Bottom status bar: 选中的: 0 行: 50/70 页尺寸: 50 / 2

2.8.1 行表达式

1. 选择数据:

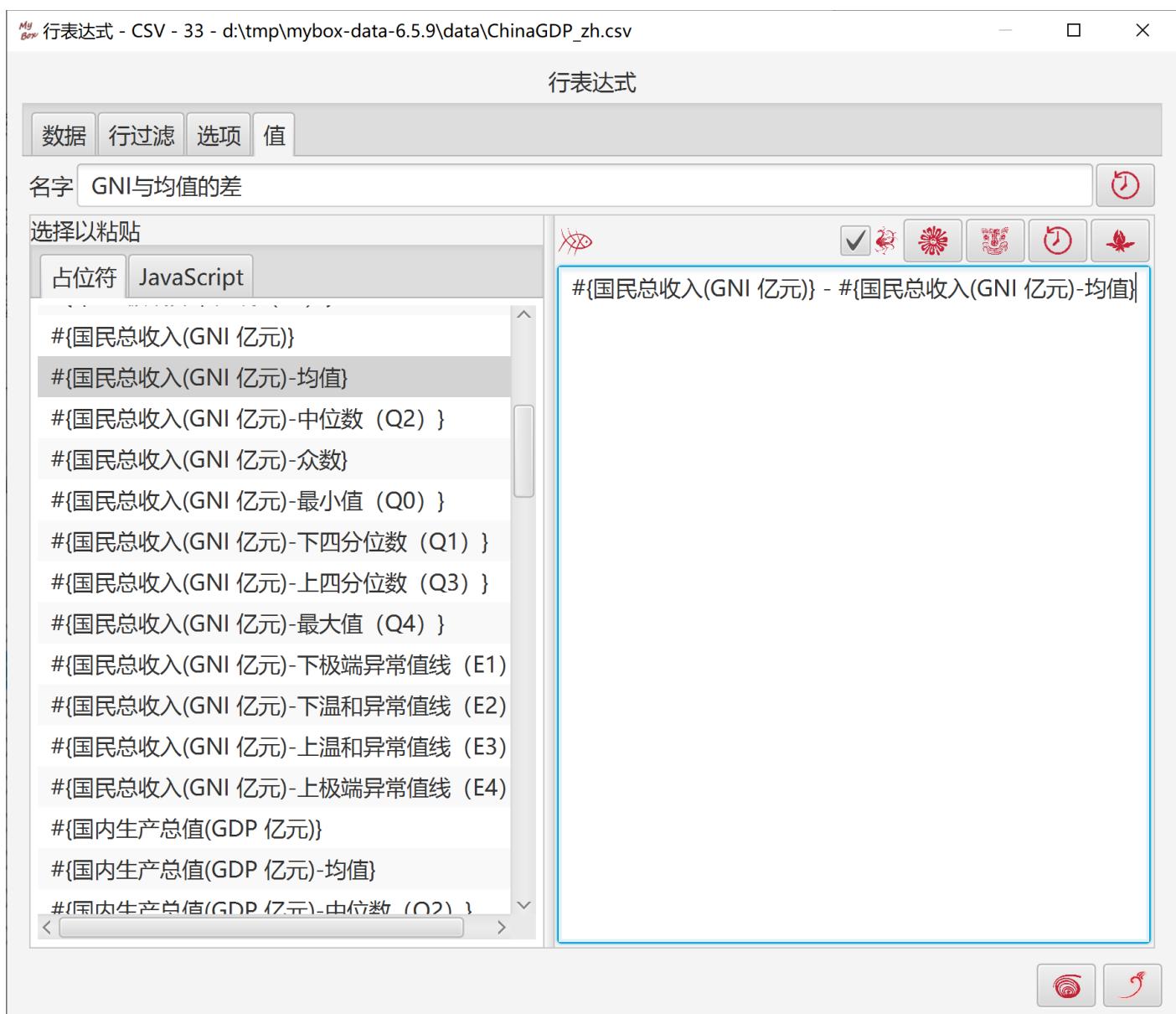
- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 输入数值的名字
- 输入行表达式。
- 选择要复制的列。

3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。



2.8.2 描述性统计

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择要计算的列。
- 选择要统计的值：

计数、合计、均值、几何平均数、平方和、总体方差、样本方差、总体标准差、样本标准差、斜度、最小值 (Q0)、下四分位数 (Q1)、中位数 (Q2)、上四分位数 (Q3)、最大值 (Q4)、上极端异常值线 (E4)、上温和异常值线 (E3)、下温和异常值线 (E2)、下极端异常值线 (E1)、众数

- 选择按照：列、行（可选类比列）、所有。
- 设置小数位数。
- 可选对非数字：略过、计为零。

3. 目标：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表，或者在表内指定位置插入、附加、或替换。



2.8.3 等值分组

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

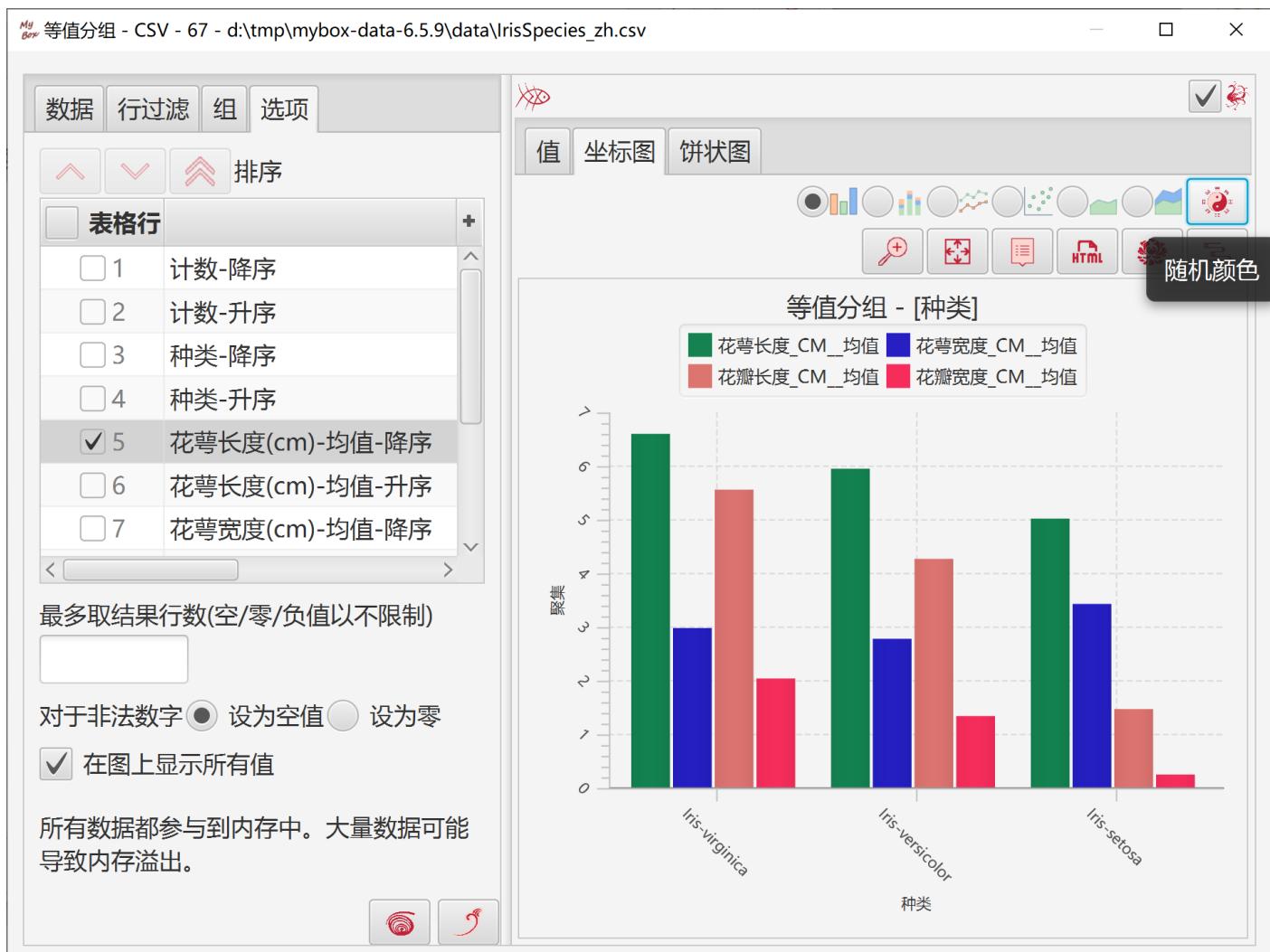
2. 计算:

- 选择要分组的列，可以是多列，不必是数字。
- 选择要计算的聚集值，可以多选，可以不选。
“计数”总是被计算，所以不在选择之列。
涉及的值必须是数字。
- 选择要排序的值，可以多选，可以不选。
- 可以输入最多取结果行数，空/零/负值表示不限制。
- 对于非法数字，可选：设为空值、设为零。

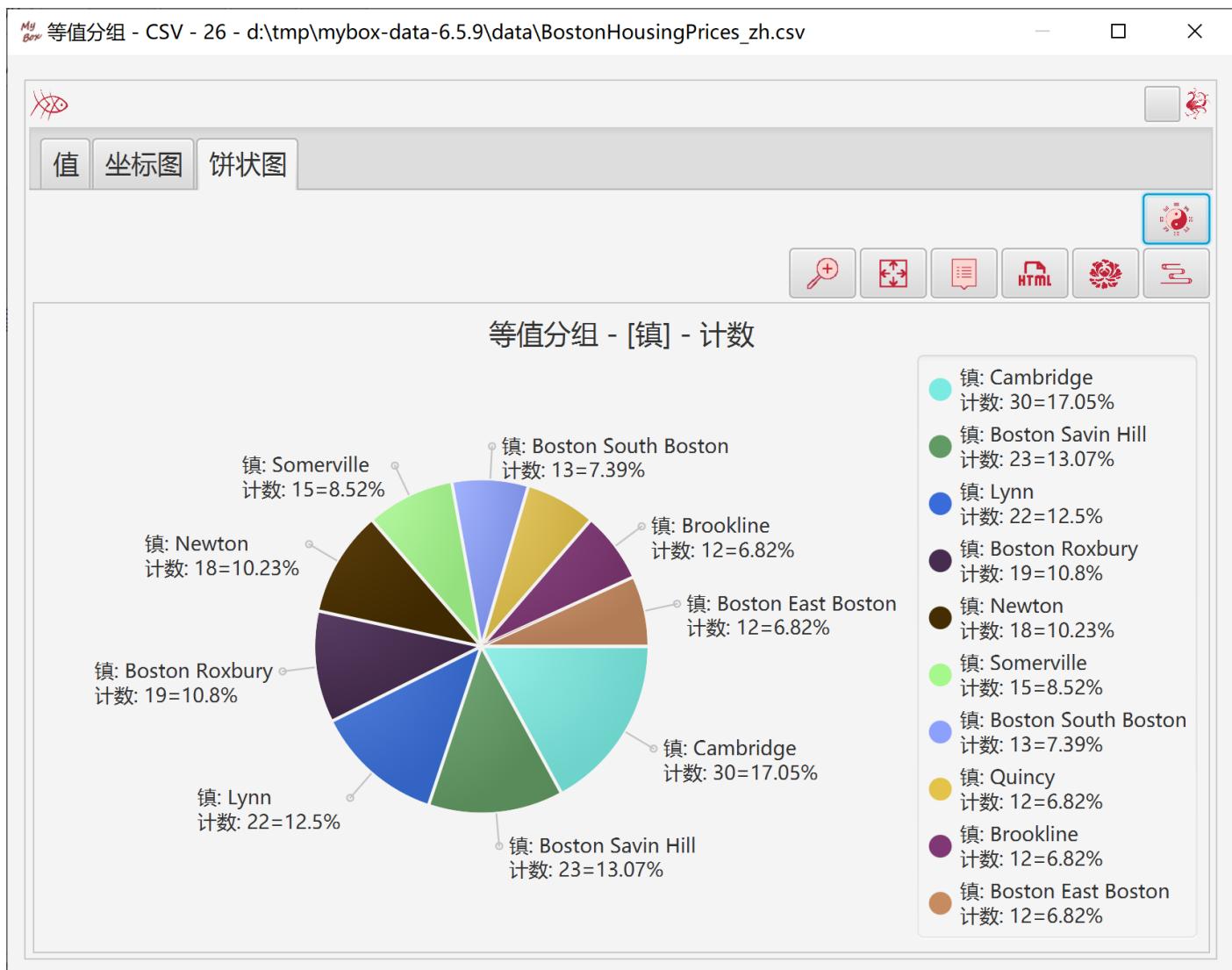
等值分组 - CSV - 67 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies_zh.csv

组	种类	计数	花萼长度_CM_均值	花萼宽度_cm_均值
组1	Iris-setosa	50	5.005999999999999	3.4180000...
组2	Iris-versicolor	50	5.936	2.7700000...
组3	Iris-virginica	50	6.587999999999998	2.9739999...

3. 计算出来的分组及其聚集值，被显示为坐标图。



4. 计算出来的分组及其计数，被显示为饼图。



5. 在图上：

- 若分组只包含一个列，则类别值就是此列的值。
- 若分组包含多个列，则类别值是组名。

6. 可选是否在图上显示所有数据：

- 若选择，则大量数据可能导致内存不够
- 若不选，则只在图上显示结果的当前页。

7. 计算可能数据溢出。

2.8.4 简单线性回归

2.8.4.1 回归

此工具基于 Apache Commons Math。

回归过程不存储数据，所以在处理很多数据时计算本身没有内存限制。

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择一列作为自变量。
- 选择另一列作为因变量。
- 选择是否包含截距
- 设置小数位数

3. 在表格中显示回归每一步的数据值状态，包括：观察数、斜率、截距、判定系数（R 方）、R 值、均方差（MSE）、方差和（SSE）、总体平方和（SSTO）、回归平方和（SSR）等。

观察数	收入	快乐	斜率 (b1)	截距 (b0)	判定系数...	皮尔逊积...	均方 +
1	3.8626	2.3145	NaN	NaN	NaN	NaN	NaN
2	4.6399	3.7379	1.8314	-4.7594	1	1	NaN
3	2.1347	0.2687	1.3514	-2.6845	0.9874	0.9937	0.076
4	6.5013	4.3748	0.9625	-1.45	0.9171	0.9577	0.411
5	3.6512	2.1558	0.9579	-1.4127	0.9183	0.9583	0.276
6	2.2865	1.8936	0.8249	-0.715	0.8447	0.9191	0.409
7	4.7489	4.903	0.9204	-0.8517	0.7453	0.8633	0.797
8	5.4592	4.8335	0.975	-0.9964	0.7754	0.8806	0.720
9	3.4331	3.1723	0.9422	-0.7712	0.7453	0.8633	0.701

2.8.4.2 模型

1. 显示拟合的线性模型。
2. 显示回归最后一步的数据值状态。
3. 输入的自变量，可以生成预测值。

简单线性回归 - CSV - 2 - d:\tmp\mybox-data-6.5.9\data\IncomeHappiness_zh.csv

关于线性回归

小数位数 4

模型 回归 拟合 残差

线性模型: 快乐 = 0.2222 + 0.7097 * 收入

自变量: 收入 = 预测

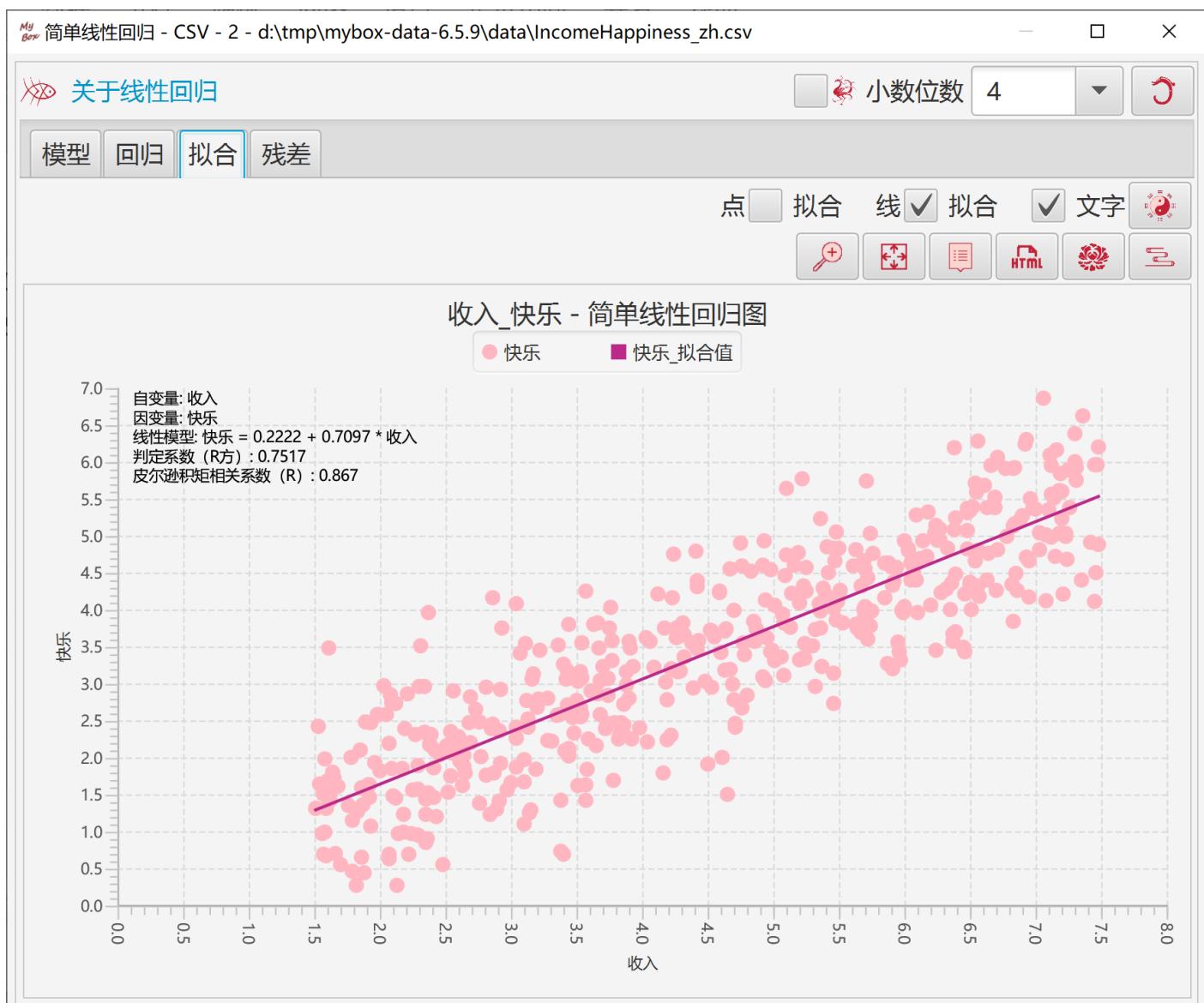
因变量: 快乐 =

最后状态

名字	值
行号	477
观察数	477
收入	4.4981
快乐	1.9071
斜率 (b1)	0.7097
截距 (b0)	0.2222
判定系数 (R方)	0.7517
皮尔逊积矩相关系数 (R)	0.867
均方差 (MSE)	0.5057

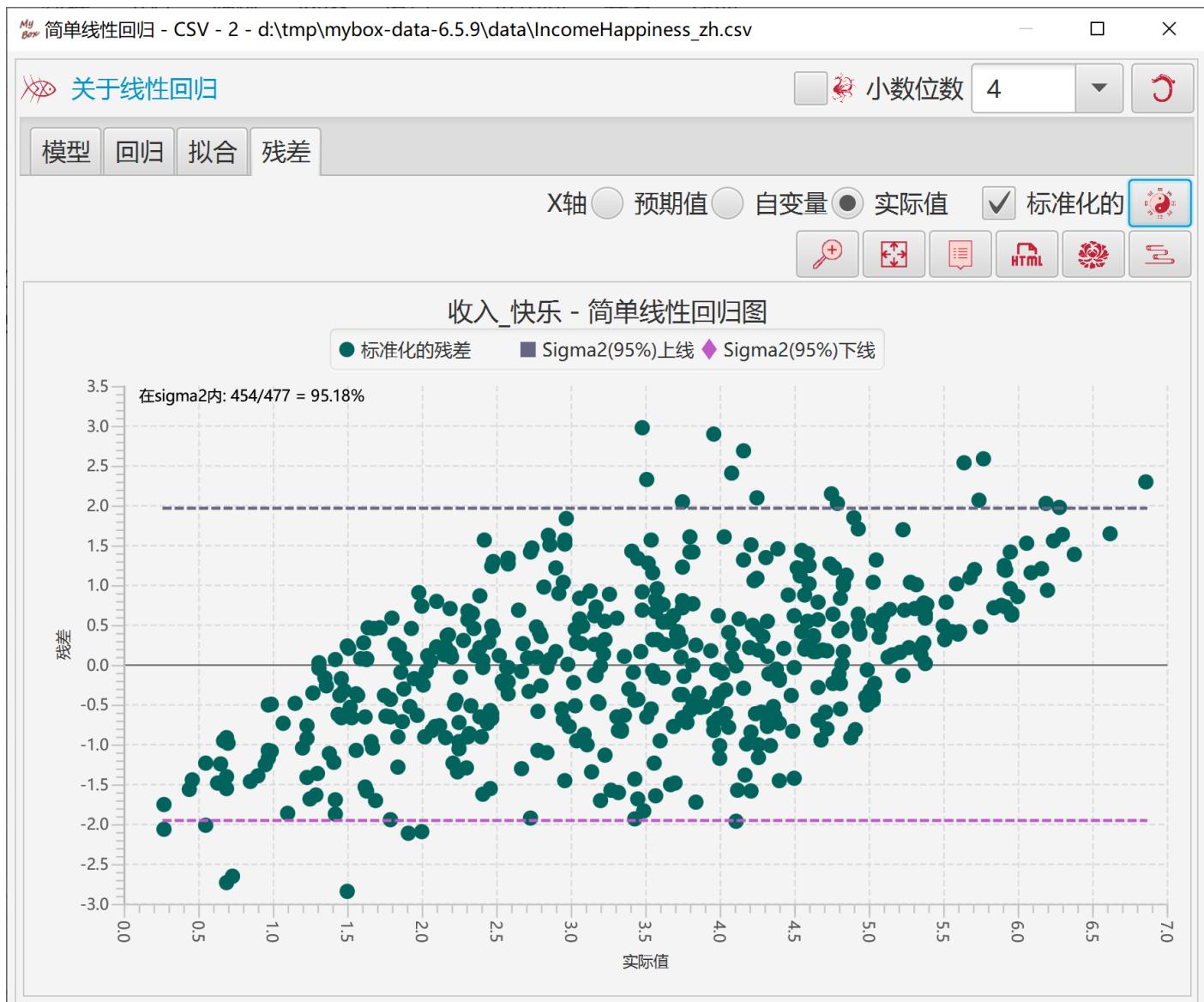
2.8.4.3 拟合图

1. 当处理所有数据行（所有页）时，选择是否在图上显示所有值。
若选择“在图上显示所有值”，则当图中加载很多数据时可能发生内存不够。
否则，图上只显示当前页的数据，同时所有页都参与回归，所以没有内存限制。
2. 可设置布局、横轴、竖轴的参数。
3. 可选择图上是否显示：拟合点、拟合线、模型描述。
4. 可选择数据的标签形式。
5. 可设置拟合点/线的随机颜色。
6. 可弹出拟合图。
7. 可生成包含拟合图及其数据的 html。
8. 在表格中显示拟合图的数据。



2.8.4.4 残差图

1. 可选择 X 轴为：预期值、自变量、实际值。
2. 可选择是否标准化残差。
当标准化残差时，显示 Sigma2 (95%) 的上线和下线。
3. 可设置散点的随机颜色。
4. 在表格中显示残差图的数据。



2.8.5 简单线性回归-组合

此工具帮助生成简单线性回归的数据：

1. 选择一些列作为自变量的候选。
2. 选择一些列作为因变量的候选。
3. 选择小数位数、alpha、是否包含截距。
4. 点击"确定"按钮后，工具自动：
 - 对候选列进行两两组合：一列作为自变量、另一列作为因变量。
 - 计算这些简单线性回归模型。
 - 把模型按判定系数（R-Square）从高到低排序
5. 选择模型，点击"查看"按钮，以进一步查看它的回归数据、拟合图、和残差图。

简单线性回归 - 组合 - CSV - 22 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_standardized_zh.csv

数据行	因变量	自变量	判定系...	皮尔逊...	
3	一年后病情...	BMI(体质指...	0.34392...	0.58645...	一年后病情进...
9	一年后病情...	S5(血清指...	0.32022...	0.56588...	一年后病情进...
4	一年后病情...	BP(平均血压)	0.19490...	0.44148...	一年后病情进...
8	一年后病情...	S4(血清指...	0.18528...	0.43045...	一年后病情进...
7	一年后病情...	S3(血清指...	0.15585...	0.39478...	一年后病情进...
10	一年后病情...	S6(血清指...	0.14629...	0.38248...	一年后病情进...
5	一年后病情...	S1(血清指...	0.04495...	0.21202...	一年后病情进...
1	一年后病情...	年龄	0.03530...	0.18788...	一年后病情进...
6	一年后病情...	S2(血清指...	0.03029...	0.17405...	一年后病情进...
2	一年后病情...	性别	0.00185...	0.043062	一年后病情进...

关于数据分析

数据 过滤 选项

自变量(不选表示所有) 年龄 性别 BMI(体质指数) BP(平均血压) S1(血清指标1) S2(血清指标2) S3(血清指标3) S4(血清指标4) S5(血清指标5) S6(血清指标6) 一年后病情进展

因变量(不选表示所有) 年龄 性别 BMI(体质指数) BP(平均血压) S1(血清指标1) S2(血清指标2) S3(血清指标3) S4(血清指标4) S5(血清指标5) S6(血清指标6) 一年后病情进展

小数位数 8

期望的显著性水平(alpha) 0.05

双击选中项以查看

2.8.6 多重线性回归

2.8.6.1 回归

基于 Apache Commons Math，此工具帮助生成多重线性回归的数据：

1. 选择一些列作为自变量。所有数据必须是数字。
2. 选择一列为因变量。数据必须是数字。
3. 选择是否包含截距。
4. 点击"确定"按钮后：
 - 工具用 Z-Score 算法归一化所选数据。
 - 工具用普通最小二乘法（OLS）计算多重线性回归模型。
 - 结果包含：截距、系数、判定系数（R-Square）、调整后的 R 方。

The screenshot shows the MyBox Data Tool window for 'Multiple Linear Regression - CSV - 22'. The left panel contains settings for variables (Age, Gender, BMI, BP, S1-S6) and regression type (L2(Z-Score), OLS). The right panel displays the 'About Data Analysis' dialog with the 'Regression' tab selected. It shows the regression model results:

名字	值
因变量	一年后病情进展
自变量	[年龄, 性别, BMI(体质指数), BP(平均血压), S1(血清指标1), S2(血清指标2), S3(血清指标3), S4(血清指标4), S5(血清指标5), S6(血清指标6)]
观察数	442
截距 (b0)	0.0
系数	[-0.006, -0.148, 0.321, 0.2, -0.489, 0.294, -0.062, 0.109, 0.464, 0.042]
判定系数 (R 方)	0.518
调整后的 R 方	0.507
标准误差	0.702
方差	1.0

2.8.6.2 模型

用户可以输入自变量的值，利用模型预测因变量的值。

The screenshot shows the MyBox software interface for 'Multiple Linear Regression' (CSV file: d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_standardized_zh.csv). The left panel displays the variable selection and regression settings:

- 自变量 (不选表示所有):** 年龄, 性别, BMI(体质指数), BP(平均血压), S1(血清指标1), S2(血清指标2), S3(血清指标3), S4(血清指标4), S5(血清指标5), S6(血清指标6).
- 因变量:** 一年后病情进展.
- 归一化:** L2(Z-Score) (selected).
- 回归:** 普通最小二乘法(OLS) (selected).
- 截距 (b0):** 选中.
- 提示:** 所有数据都参与到内存中。大量数据可能导致内存溢出。

The right panel shows the regression equation and input fields for prediction:

关于数据分析

模型 回归

小数位数 3

线性模型: 一年后病情进展 = $0.0 - 0.006 * \text{年龄} - 0.148 * \text{性别} + 0.321 * \text{BMI(体质指数)} + 0.2 * \text{BP(平均血压)} - 0.489 * \text{S1(血清指标1)} + 0.294 * \text{S2(血清指标2)} - 0.062 * \text{S3(血清指标3)} + 0.109 * \text{S4(血清指标4)} + 0.464 * \text{S5(血清指标5)} + 0.042 * \text{S6(血清指标6)}$

自变量: 年龄 = []

自变量: 性别 = []

自变量: BMI(体质指数) = []

自变量: BP(平均血压) = []

自变量: S1(血清指标1) = []

自变量: S2(血清指标2) = []

自变量: S3(血清指标3) = []

自变量: S4(血清指标4) = []

自变量: S5(血清指标5) = []

自变量: S6(血清指标6) = [] 预测

因变量: 一年后病情进展 = []

2.8.7 多重线性回归-组合

此工具帮助生成多重线性回归的数据:

1. 选择一些列作为自变量的候选。
2. 选择一些列作为因变量的候选。
3. 选择是否包含截距。
4. 点击"确定"按钮后, 工具自动:
 - 对候选列进行组合作为自变量和因变量。
 - 计算这些多重线性回归模型。
 - 把模型按调整后的判定系数从高到低排序。
5. 选择模型, 点击"查看"按钮, 以进一步查看相应的回归数据。

多重线性回归 - 组合 - CSV - 22 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_standardized_zh.csv

数据
过滤
选项

自变量(不选表示所有)

<input checked="" type="checkbox"/> 年龄	<input checked="" type="checkbox"/> 性别	<input checked="" type="checkbox"/> BMI(体质指数)
<input checked="" type="checkbox"/> BP(平均血压)	<input checked="" type="checkbox"/> S1(血清指标1)	
<input checked="" type="checkbox"/> S2(血清指标2)	<input checked="" type="checkbox"/> S3(血清指标3)	
<input checked="" type="checkbox"/> S4(血清指标4)	<input checked="" type="checkbox"/> S5(血清指标5)	
<input checked="" type="checkbox"/> S6(血清指标6)		
<input type="checkbox"/> 一年后病情进展		

因变量(不选表示所有)

<input type="checkbox"/> 年龄	<input type="checkbox"/> 性别	<input type="checkbox"/> BMI(体质指数)
<input type="checkbox"/> BP(平均血压)	<input type="checkbox"/> S1(血清指标1)	
<input type="checkbox"/> S2(血清指标2)	<input type="checkbox"/> S3(血清指标3)	
<input type="checkbox"/> S4(血清指标4)	<input type="checkbox"/> S5(血清指标5)	
<input type="checkbox"/> S6(血清指标6)		
<input checked="" type="checkbox"/> 一年后病情进展		

归一化 L2(Z-Score)
 回归 普通最小二乘法(OLS)
 截距 (b0)

所有数据都参与到内存中。大量数据可能导致内存溢出。

关于数据分析

数据行	因变量	自变量	调整后...	判定系...	系数...
19	一年后病情...	[性别, BMI(体质指...	0.50767...	0.51771...	[-0.14
18	一年后病情...	[性别, BMI(体质指...	0.50764...	0.51657...	[-0.14
10	一年后病情...	[年龄, 性别, BMI(体质指...	0.50656...	0.51774...	[-0.00
9	一年后病情...	[年龄, 性别, BMI(体质指...	0.50650...	0.51657...	[-0.00
26	一年后病情...	[BMI(体质指...	0.49165...	0.499727	[0.348
27	一年后病情...	[BMI(体质指...	0.49099...	0.50023...	[0.344
16	一年后病情...	[性别, BMI(体质指...	0.48634...	0.49332...	[-0.14
17	一年后病情...	[性别, BMI(体质指...	0.48585...	0.49401...	[-0.14
7	一年后病情...	[年龄, 性别, BMI(体质指...	0.48525...	0.49342...	[0.010
8	一年后病情...	[年龄, 性别, BMI(体质指...	0.48477...	0.49411...	[0.010
24	一年后病情...	[BMI(体质指...	0.471218	0.47721...	[0.376
25	一年后病情...	[BMI(体质指...	0.47023...	0.47744...	[0.376
34	一年后病情...	[BP(平均血压), S1(血清指...	0.41056...	0.41991...	[0.238
33	一年后病情...	[BP(平均血压), S1(血清指...	0.40788...	0.41594...	[0.256
15	一年后病情...	[性别, BMI(体质指...	0.39582...	0.40267...	[-0.04
13	一年后病情...	[性别, BMI(体质指...	0.39568...	0.39979...	[-0.06
14	一年后病情...	[性别, BMI(体质指...	0.39519...	0.40068...	[-0.06
23	一年后病情...	[BMI(体质指...	0.39507...	0.40056...	[0.490
4	一年后病情...	[年龄, 性别, BMI(体质指...	0.39477...	0.40026...	[0.023

双击选中项以查看

2.8.8 频数分布

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择一列计算频数。
- 选择是否忽略大小写
- 设置小数位数

3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。

The screenshot shows the 'Frequency Distribution' feature in MyBox Data Tools. On the left, there is a table with three columns: '镇' (Town), '镇_计数' (Count), and '镇_数目百分比' (Percentage). The data includes:

镇	镇_计数	镇_数目百分比
Lynn	7	35
Marblehead	3	15
Nahant	1	5
Salem	7	35
Swampsc...	2	10

The right pane contains the following configuration options:

- 频数分布** (Frequency Distribution) title.
- 数据** (Data) tab selected.
- 行过滤** (Row Filter) and **选项** (Options) tabs.
- 列** dropdown set to **镇**.
- 忽略大小写** (Ignore Case) checkbox checked.
- 小数位数** (Decimal Places) set to 2.
- 目标** (Target) section:
 - 外部** (External) radio button selected. Sub-options include CSV (selected), XLS, TXT, and others.
 - 名字** (Name) input field: BostonHousingPrices_zh.csv
 - 表内** (In Table) radio button.
 - Positioning options: 插入在上方 (Insert Above), 附加在下方 (Append Below), 替换 (Replace).
- Bottom right icons: a red circular icon with a white swirl and a blue circular icon with a white swirl.

2.8.9 数值百分比

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择要计算的列，若不选则取所有列。
- 选择按照：列、行、所有
- 选择把负数计为：零、绝对值
- 设置小数位数
- 可选对非数字：略过、计为零。
- 选择要复制的列。

3. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。

4. 当选择当前页或选中行时，目标还可以为：在表内指定位置插入、附加、或替换。

The screenshot shows the MyBox application window with a menu bar and toolbar. A data table is displayed in the main area, and a 'Percentage' configuration dialog is open on the right side.

Data Table:

源行号	指标	2020年	2020年_列中百分比	2019年	2019年_列中百分比
行6	居民人均水产品消费量(千克)	13.9	3.56	13.6	3.67
行3	居民人均蔬菜及食用菌消费...	103.7	26.58	98.6	26.62
行2	居民人均食用油消费量(千克)	10.4	2.67		
行10	居民人均食糖消费量(千克)	1.3	0.33		
行4	居民人均肉类消费量(千克)	24.8	6.36		
行5	居民人均禽类消费量(千克)	12.7	3.26		
行8	居民人均奶类消费量(千克)	13	3.33		
行1	居民人均粮食消费量(千克)	141.2	36.2		
行9	居民人均干鲜瓜果类消费量(...	56.3	14.43		
行7	居民人均蛋类消费量(千克)	12.8	3.28		
列-合计			390.1		

Percentage Dialog:

百分比

要计算的列 (不选表示所有) 指标 2020年 2019年 2018年 2017年 2016年
 2015年 2014年 2013年

按照 列 行 所有

小数位数

对于负数 略过 计为零 绝对值

对于非法数字 略过 计为零

要复制的列 指标 2020年 2019年 2018年 2017年 2016年
 2015年 2014年 2013年

目标

2.9 数据图

悬停或点击功能按钮，选择“图”下的功能。

The screenshot shows the MyBox Data Tool interface with a CSV file named "ChinaGDP_zh.csv" open. The main window includes a menu bar with Chinese labels: 窗口 (Window), 文档 (Document), 图像 (Image), 数据 (Data), 文件 (File), 媒体 (Media), 网络 (Network), 设置 (Settings), 最近访问 (Recent), 开发 (Development), and 帮助 (Help). Below the menu is a toolbar with various icons. A context menu is displayed over a row in the data grid, specifically over the fourth column header "国民总收入". This menu lists several chart types: XY图 (XY Chart), 饼状图 (Pie Chart), 箱线图 (Box Plot), 自比较条图 (Self-comparison Bar Chart), 相比较条图 (Comparison Bar Chart), and XYZ图 (XYZ Chart). The "图" (Chart) option is highlighted with a blue background.

表格行	数据行	年	国民总收入
1	1	1952	679.1
2	2	1953	824.4
3	3	1954	950.9
4	4	1955	1,092.3
5	5	1956	1,232.3
6	6	1957	1,372.3
7	7	1958	1,512.3
8	8	1959	1,652.3
9	9	1960	1,792.3
10	10	1961	1,932.3
11	11	1962	2,072.3
12	12	1963	2,212.3
13	13	1964	2,352.3

选中的: 0 行: 50/70 页尺寸: 50 / 1 / 2 < > << >>

2.9.1 XY图

2.9.1.1 数据

1. 选择数据:

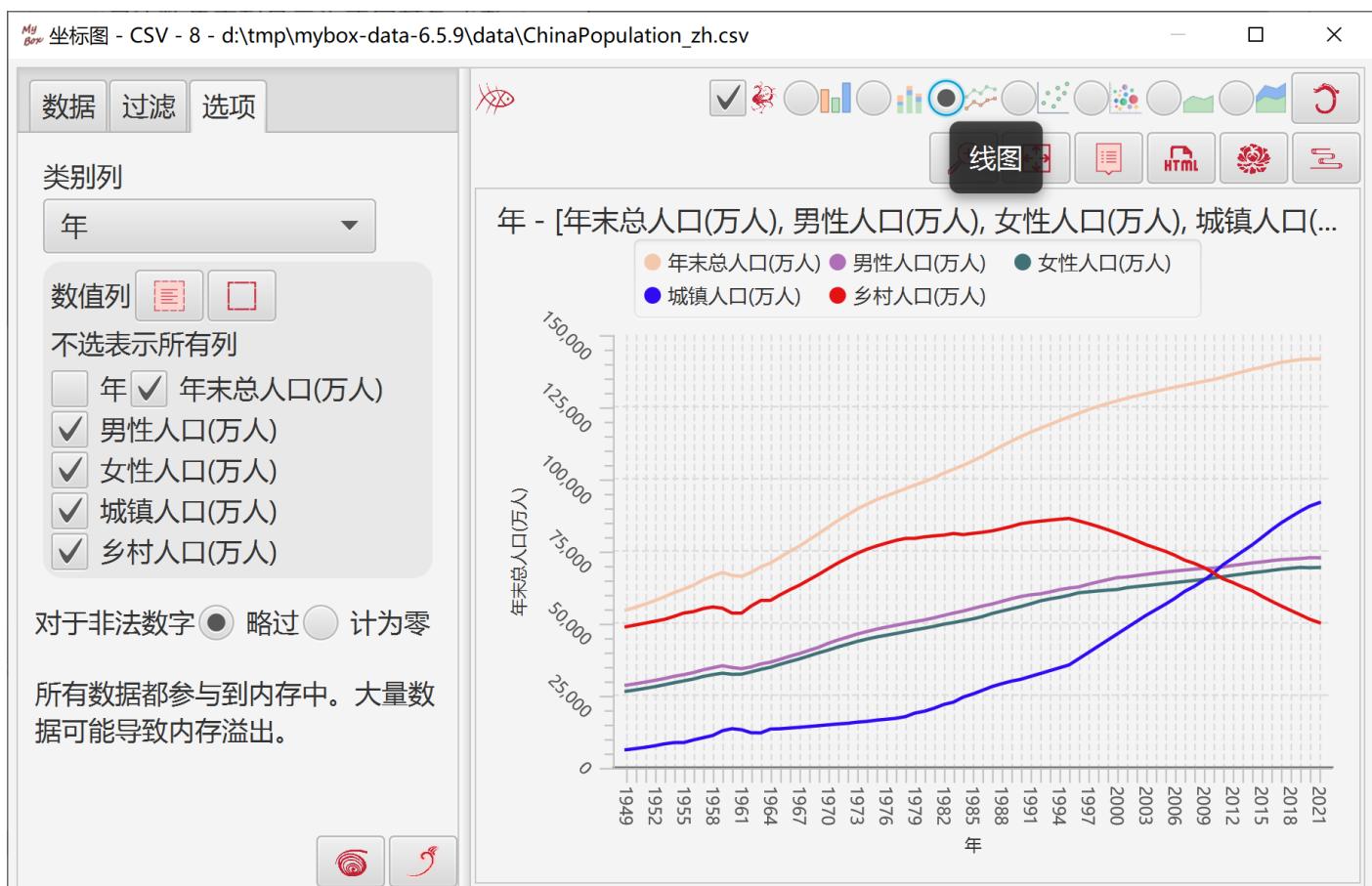
- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

- 选择一列为“类别轴”，以定义数据名字。
- 选择多列为“数值轴”，不同的数值序列显示为不同颜色或者形状。
- 对于非法数字，可选：忽略、或计为零。
- 当处理所有数据行（所有页）时，坐标图需要考虑内存限制。
- 缺省情况下，“类别轴”是横向轴、“数值轴”是竖向轴。

3. 选择 XY 图的类型。

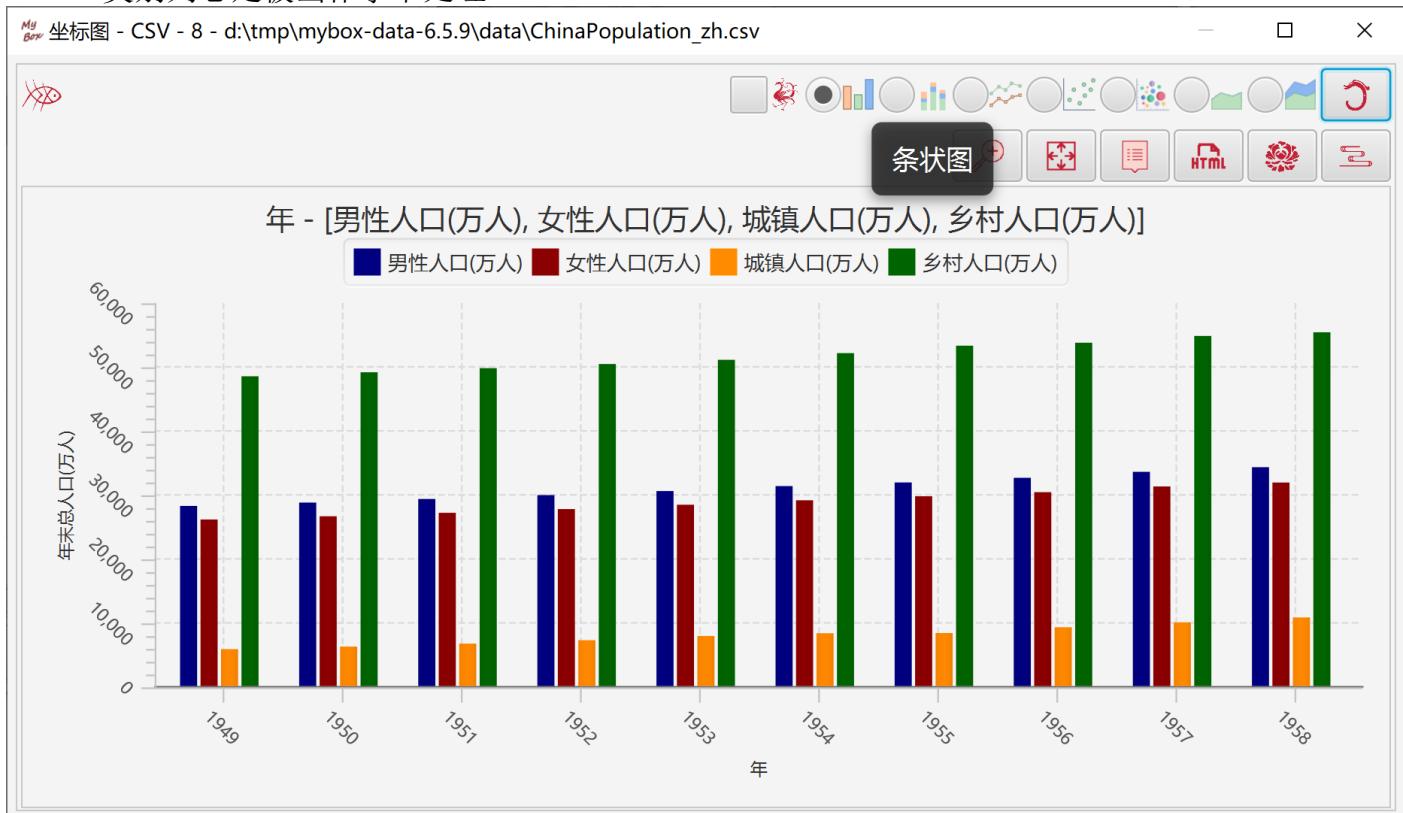
- 点击“菜单”按钮以设置数据图的参数。
- 点击“弹出”按钮以在弹出窗口中显示当前图。
- 点击“数据”按钮以在数据表中显示坐标图的数据。
- 点击“网页”按钮以在网页中显示坐标图的数据。



2.9.1.2 条图

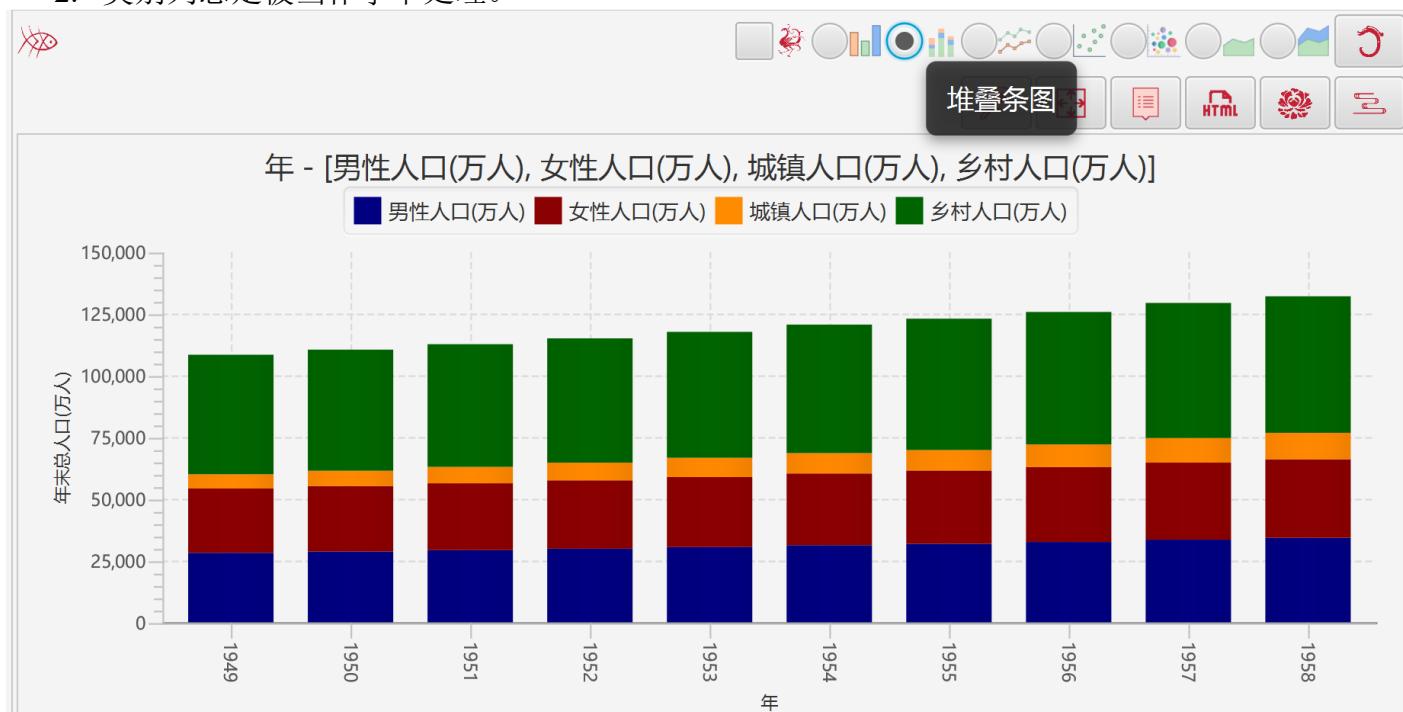
- 以条块的高度表示数据的大小。

2. 类别列总是被当作字串处理。



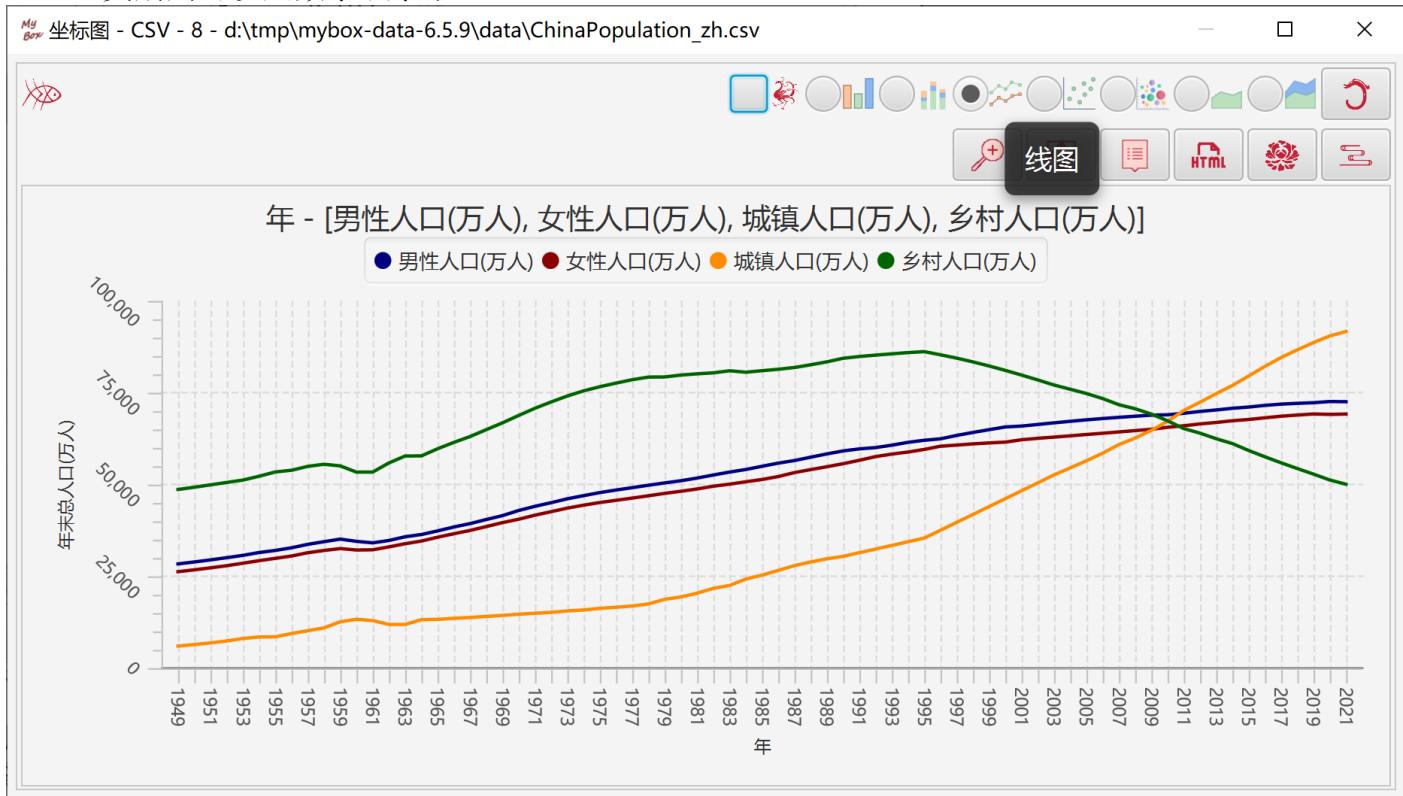
2.9.1.3 堆叠条图

1. 以条块的高度表示数据的大小。
2. 类别列总是被当作字串处理。



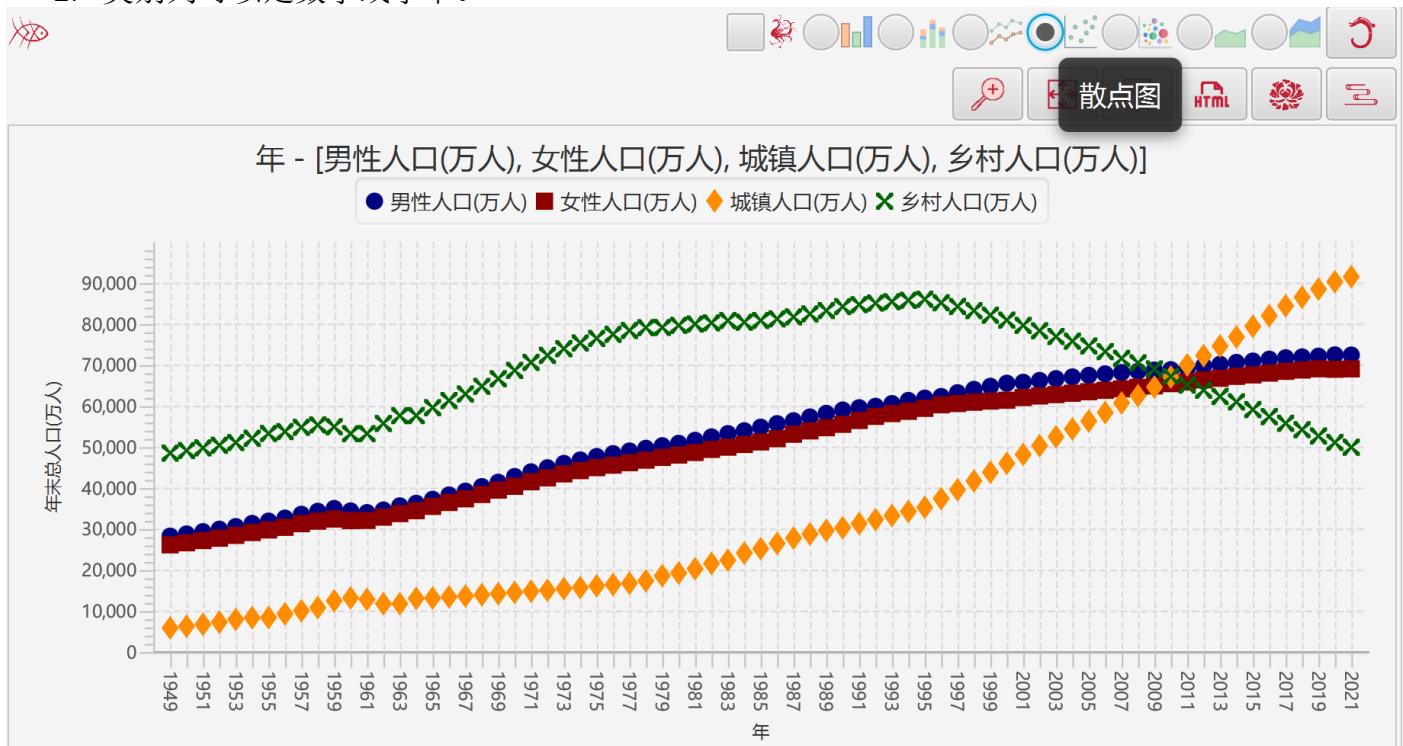
2.9.1.4 线图

- 以连接点的线条表示数据的趋势。
- 类别列可以是数字或字串。



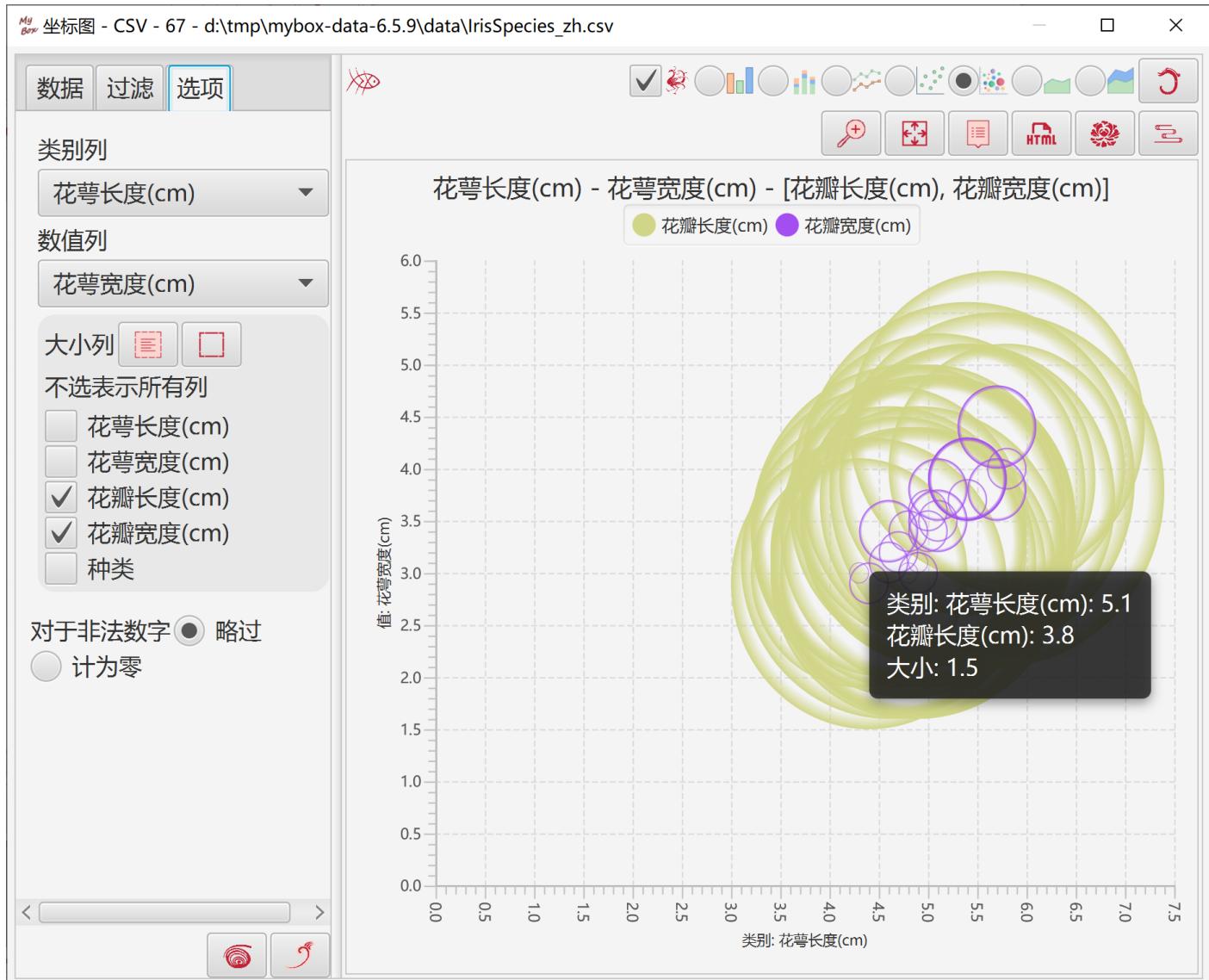
2.9.1.5 散点图

- 以符号表示数据的分布。
- 类别列可以是数字或字串。



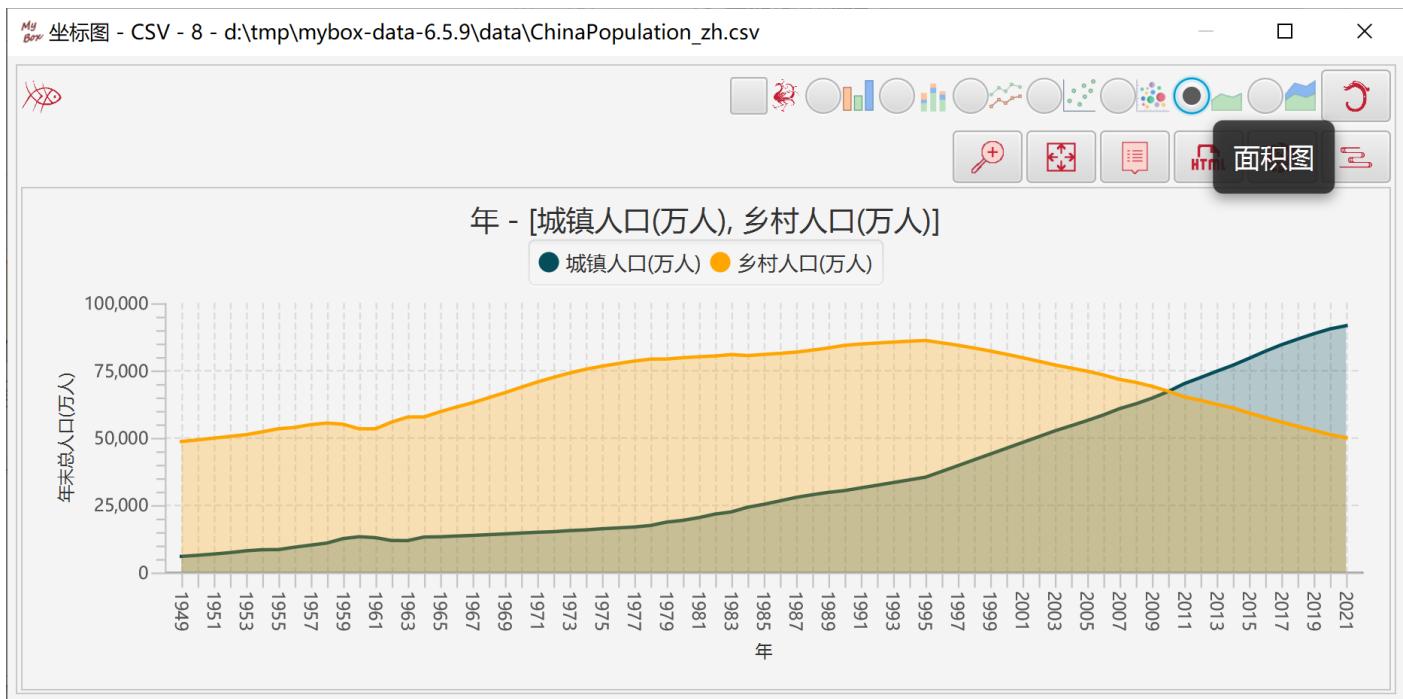
2.9.1.6 气泡图

1. 以不同半径的多个圆表示数据的大小。
2. “类别”列和“数值”列定义数据的坐标。
3. 选择若干“大小”列，定义数据的大小。
4. 所有列都必须是数字，大小列必须是非负数。



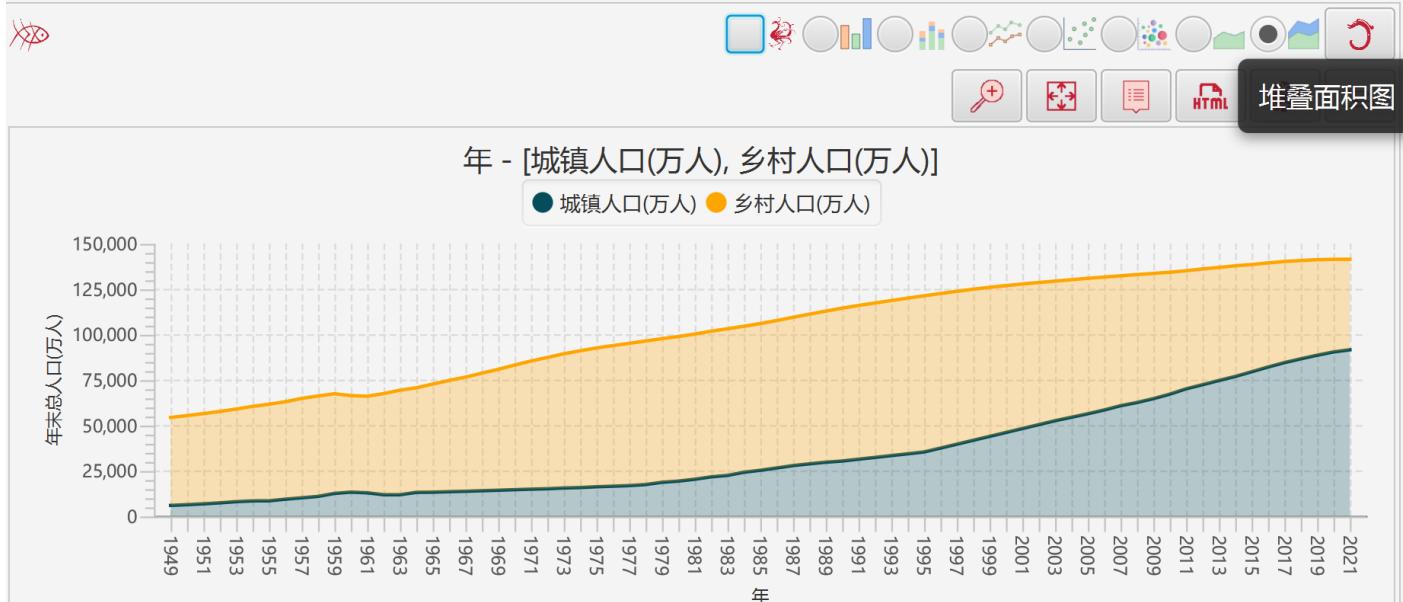
2.9.1.7 面积图

- 以面积表示数据的大小。
- 类别列总是被当作字串处理。



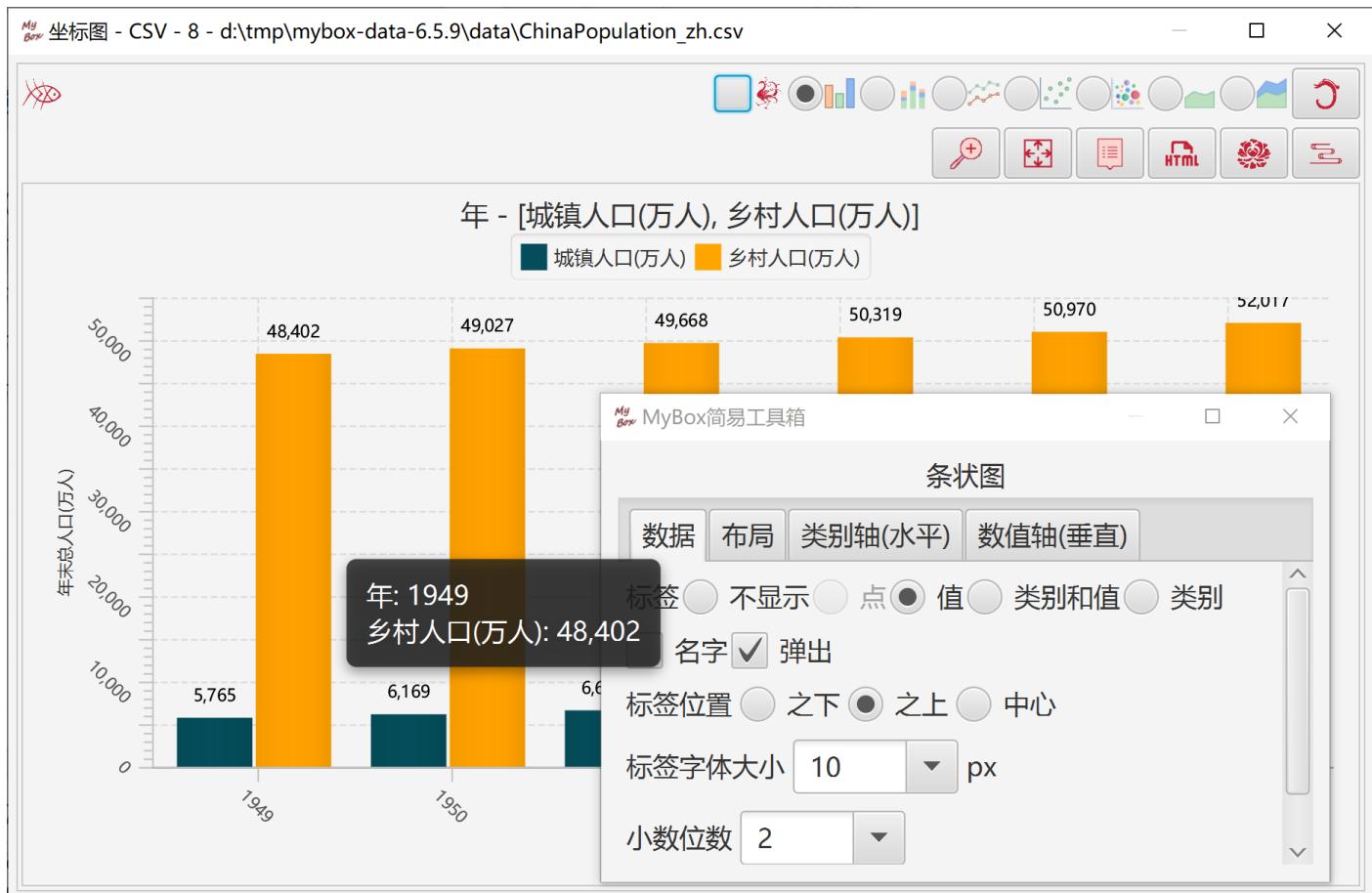
2.9.1.8 堆叠面积图

- 以面积表示数据的大小。
- 类别列总是被当作字串处理。



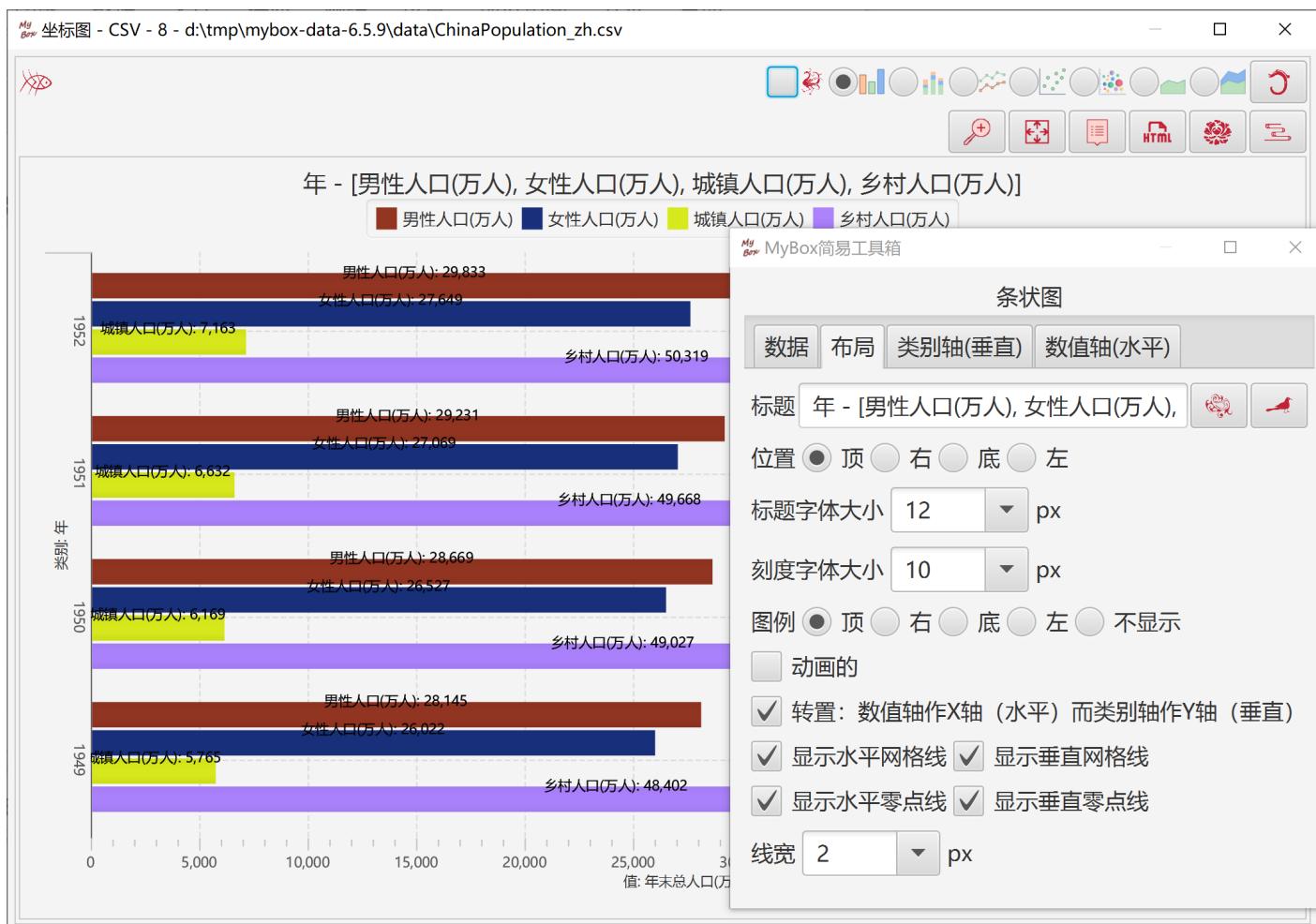
2.9.1.9 图数据的参数

1. 标签, 可选: 不显示、点、值、类别。
2. 标签位置
3. 标签字体大小
4. 小数位数



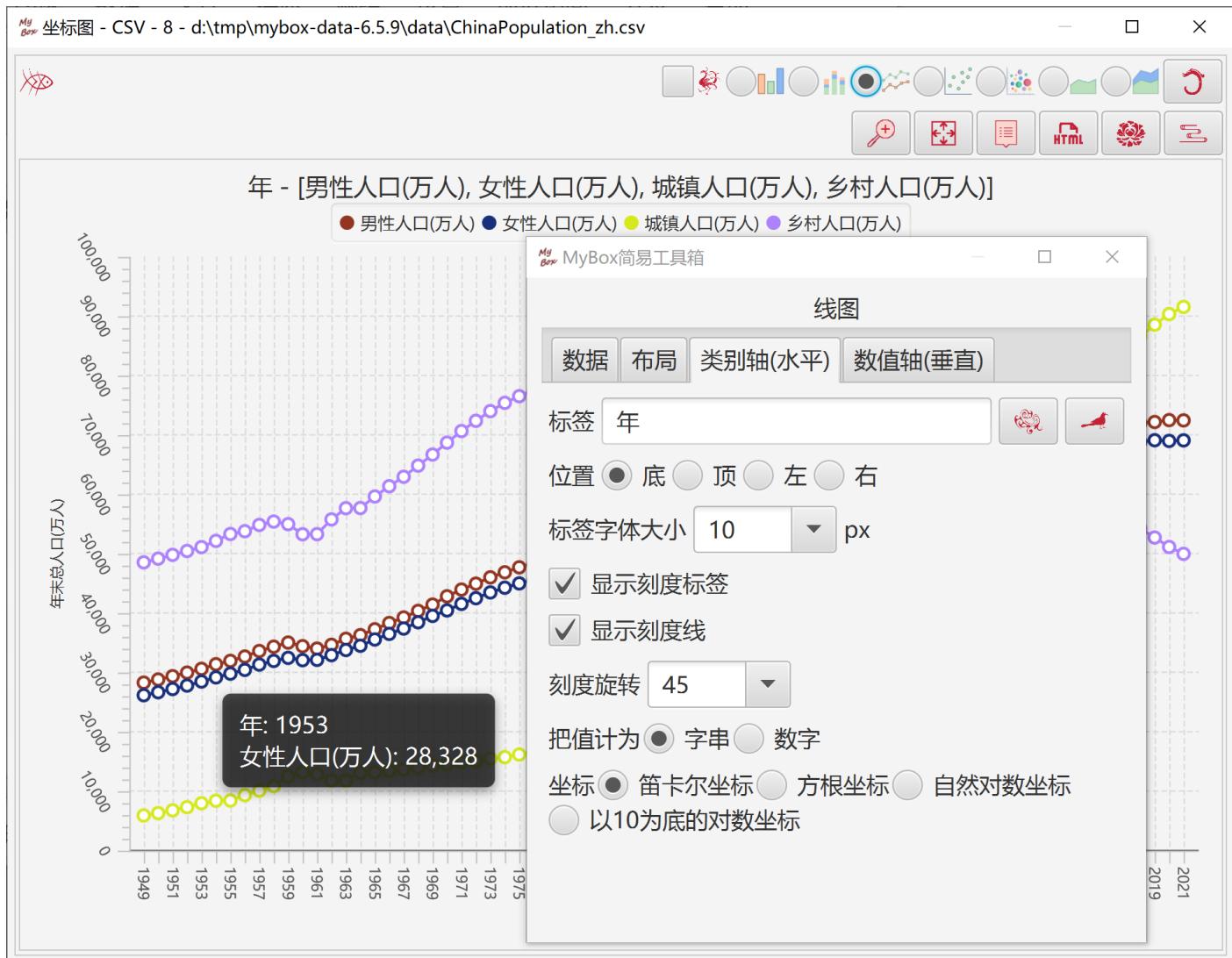
2.9.1.10 布局

设置面板：标题、字体、图例位置、是否显示零度线/网格线、线宽等。



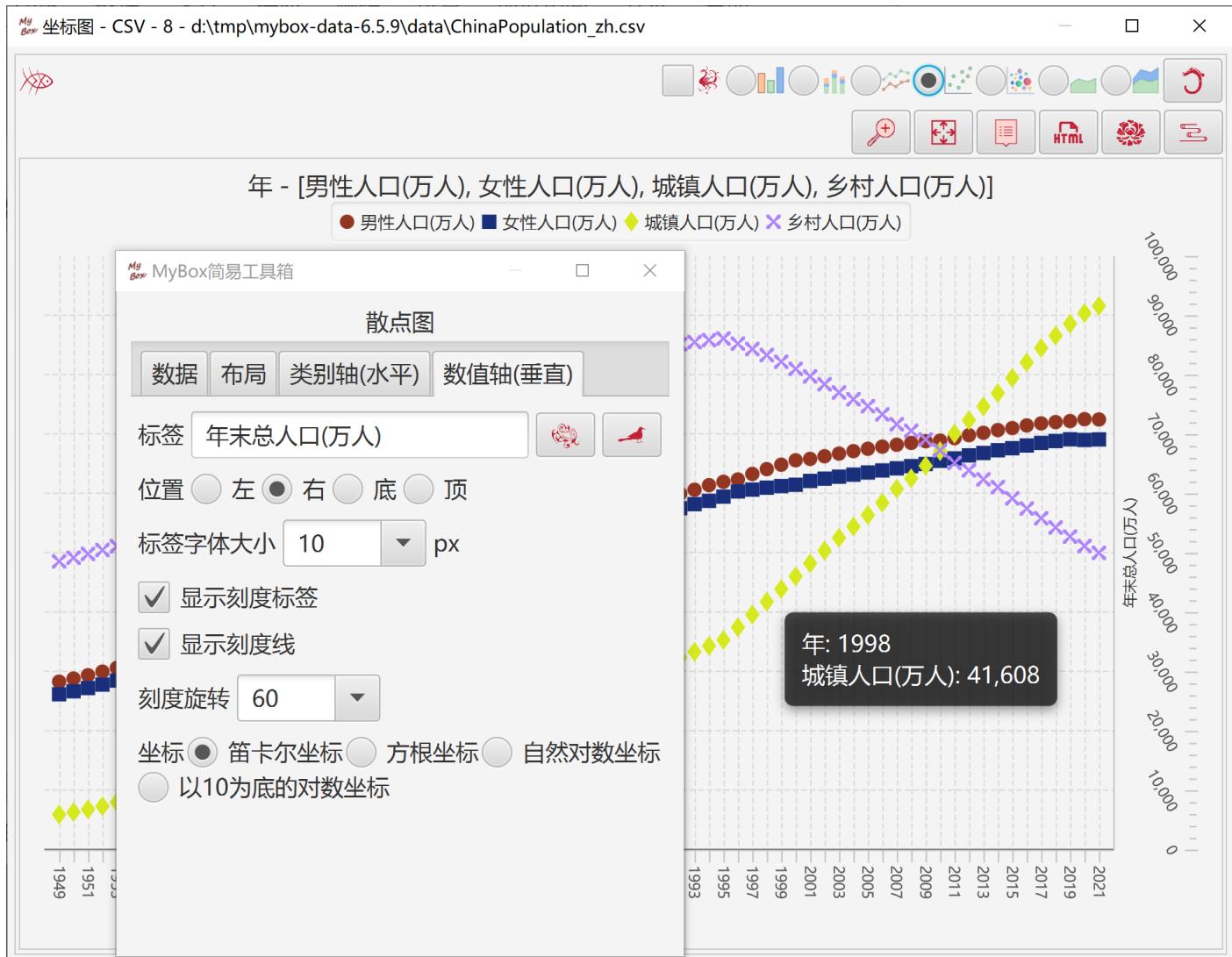
2.9.1.11 类别轴

设置类别轴：标签、字体、位置、刻度、当作字串处理还是数字处理、坐标等。



2.9.1.12 数值轴

设置数值轴：标签、字体、位置、刻度、坐标等。



2.9.2 饼图

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 计算:

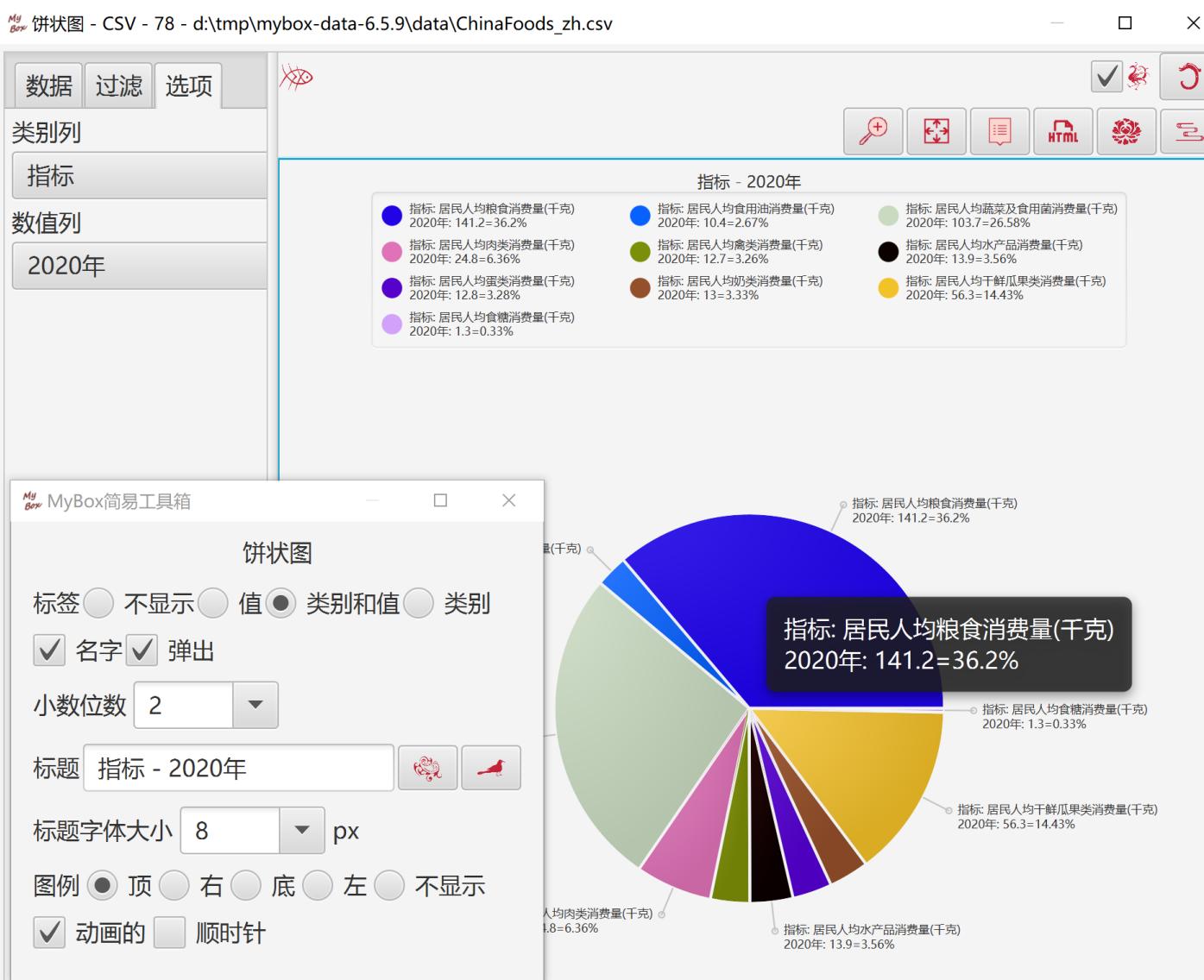
- 选择一列为“类别”。
- 选择一列为“数值”。
- 以分割的圆表示数据的比例。
- 数值列必须都是非负数。
- 当处理所有数据行（所有页）时，饼图需要考虑内存限制。

3. 点击“菜单”按钮以设置数据图的参数。

4. 点击“弹出”按钮以在弹出窗口中显示当前图。

5. 点击“数据”按钮以在数据表中显示坐标图的数据。

6. 点击“网页”按钮以在网页中显示坐标图的数据。



2.9.3 箱线图

1. 箱线图用来展示数据的分布:

- 数据按照列/行/全部来升序排序
- 以下项可以显示出数据的聚集性和离散性:
最小值 $Q_0 =$ 位于数据列的 0% (开头)
下四分位数 $Q_1 =$ 位于数据列的 25%
中位数 $Q_2 =$ 位于数据列的 50% (中部)
上四分位数 $Q_3 =$ 位于数据列的 75%
最大值 $Q_4 =$ 位于数据列的 100% (结尾)
- 以下项可以用于标识数据的异常值:
下极端异常值线 $E_1 = Q_1 - 3 * (Q_3 - Q_1)$
下温和异常值线 $E_2 = Q_1 - 1.5 * (Q_3 - Q_1)$
上温和异常值线 $E_3 = Q_3 + 1.5 * (Q_3 - Q_1)$
上温和异常值线 $E_4 = Q_3 + 3 * (Q_3 - Q_1)$
- 以下项可以用于离散性的参考:
均值 = 数据的平均数

2. 计算:

- 基于线图。
- 选择按照: 列、行 (选择类别列) 、所有.

3. 设置或选择:

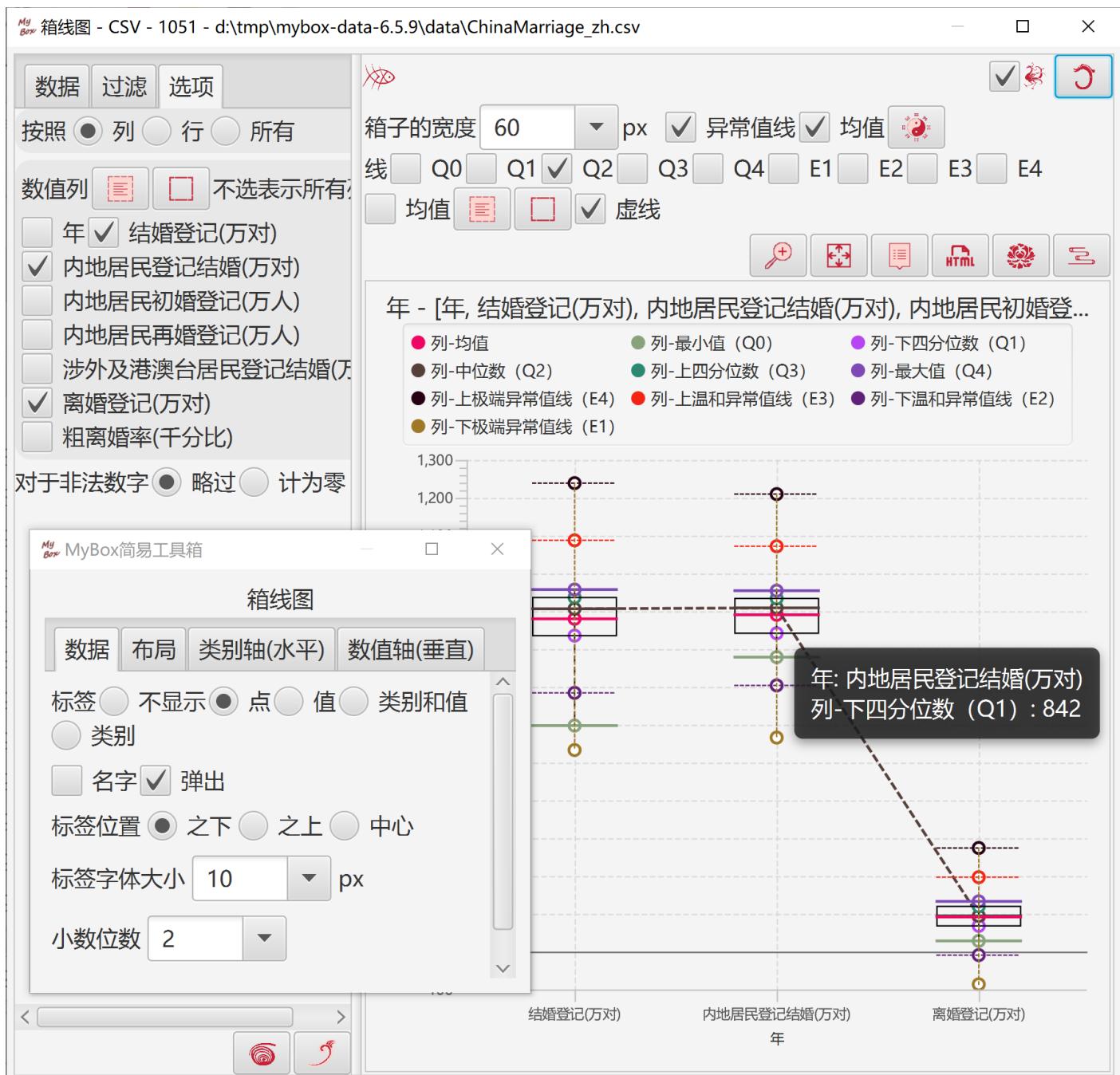
- 箱子的宽度。
- 是否显示异常值线或均值。
- 是否显示各类数值的连线、是否虚线。
- 随机颜色。

4. 点击“菜单”按钮以设置数据图的参数。

5. 点击“弹出”按钮以在弹出窗口中显示当前图。

6. 点击“数据”按钮以在数据表中显示坐标图的数据。

7. 点击“网页”按钮以在网页中显示坐标图的数据。



2.9.4 自比较条图

1. 自比较条图用来对比数据与参考值。以下规则用来计算所选数据值的颜色条：

- 若值等于零，不显示条
- 当按绝对值比较时：

 最大值 = 列/行/所有的最大绝对值

 百分比 = 值的绝对值 / 最大值

 宽度 = 最大宽度 * 百分比

 颜色 = 若值大于零，为列的颜色；若值小于零，为列的颜色的反色

- 当按最大最小值区间比较时：

 最大值 = 列/行/所有的最大值

 最小值 = 列/行/所有的最小值

 百分比 = (值 - 最小值) / (最大值 - 最小值)

 宽度 = 最大宽度 * 百分比

 条的颜色 = 列的颜色

2. 数据：选择表中行、或所有数据行（所有页）。

3. 计算：

- 选择要计算的列、和要复制的列。
- 选择比较：列、行、所有
- 选择按照：绝对值、最大最小值区间
- 当处理所有数据行（所有页）时，自比较条图需要考虑内存限制。

4. 设置或选择：

- 最大宽度。
- 是否显示行号、值、百分比、类别、计算出来的值。

5. 编辑图中数据

6. 编辑图的 html

男性人口(万人)	女性人口(万人)	年	年末总人口(万人)
最大绝对值: 72357.0	最大绝对值: 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

2.9.5 相比较条图

1. 相比较条图用来对比两类数据。以下规则用来计算颜色条:

- 若值等于零, 不显示条
- 当按绝对值比较时:

 最大值 = 两个值列的最大绝对值

 百分比 = 值的绝对值 / 最大值

 宽度 = 最大宽度 * 百分比

 颜色 = 若值大于零, 为列的颜色; 若值小于零, 为列的颜色的反色

- 当按最大最小区间比较时:

 最大值 = 两个值列的最大值

 最小值 = 两个值列的最小值

 百分比 = (值 - 最小值) / (最大值 - 最小值)

 宽度 = 最大宽度 * 百分比

 条的颜色 = 列的颜色

2. 数据: 选择表中行、或所有数据行(所有页)。

3. 计算:

- 选择类别列(非必要)
- 选择两个数值列
- 选择按照: 绝对值、最大最小区间
- 当处理所有数据行(所有页)时, 自比较条图需要考虑内存限制。
- 选择要复制的列。

4. 设置或选择:

- 最大宽度。
- 是否显示行号、值、百分比、类别、计算出来的值。

5. 编辑图中数据

6. 编辑图的 html

男性人口(万人)	年	女性人口(万人)	年末总人口(万人)	城镇人口(万人)	乡村人口(万人)
38%	1949	38%	54167	5765	48402
39%	1950	39%	55196	6169	49027
40%	1951	40%	56300	6632	49668
41%	1952	41%	57482	7163	50319
42%	1953	42%	58796	7826	50970
43%	1954	43%	60266	8249	52017
43%	1955	43%	61465	8285	53180
44%	1956	44%	62828	9185	53643

2.9.6 XYZ图

2.9.6.1 三维散点图

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 选择数据轴:

- 选择一列为“X 轴”，可选是否计为字符串，不选则计为数字。
- 选择一列为“Y 轴”，可选是否计为字符串，不选则计为数字。
- 选择多列为“Z 轴”，可选是否计为字符串，不选则计为数字。
- 选择其它要在标签中弹出的值。

3. 设置图参数：选择投影、颜色、宽度、高度、是否深色模式、点的尺寸。

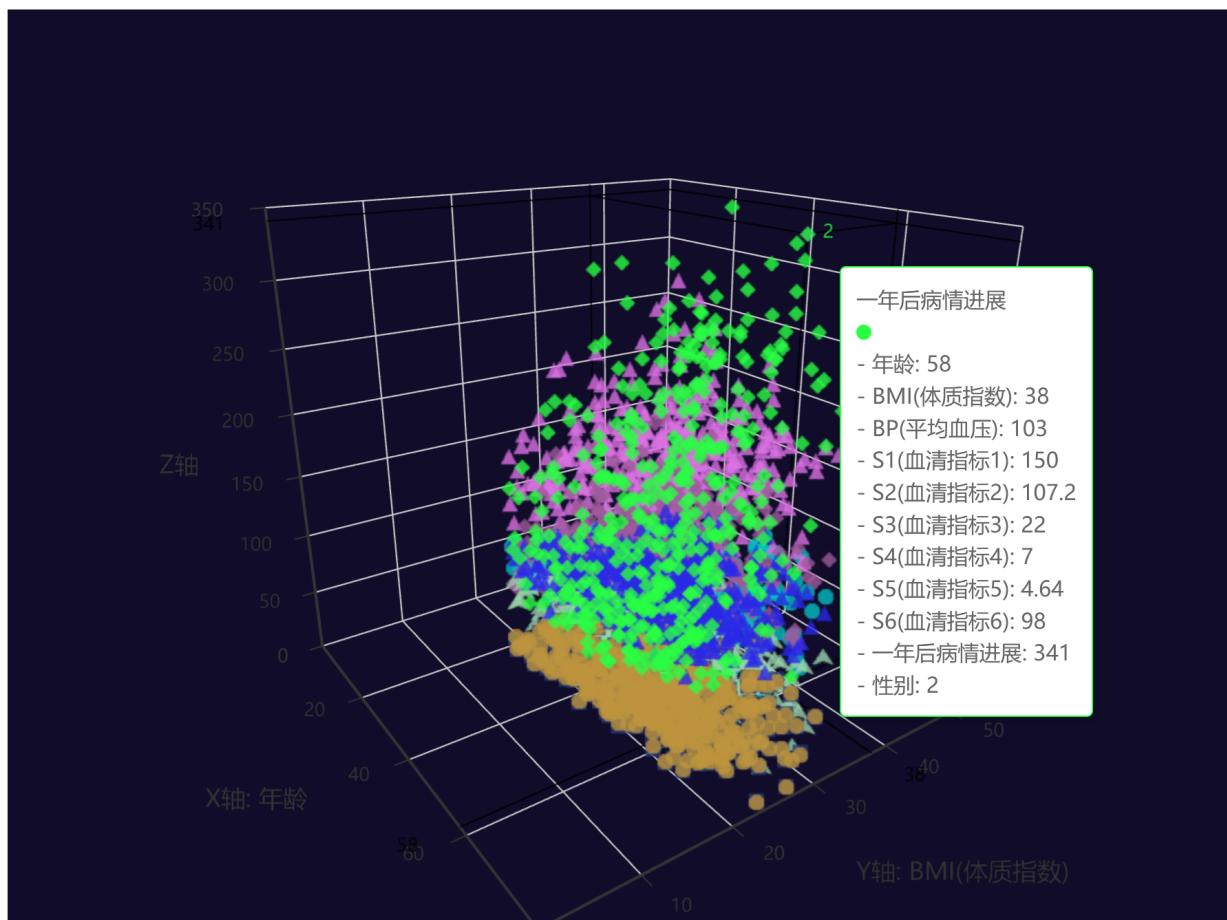




糖尿病的病情预测 - 散点图

X轴: "年龄" Y轴: "BMI(体质指数)"

Z轴: ● BP(平均血压) ▲ S1(血清指标1) ♦ S2(血清指标2) ▲ S3(血清指标3) ■ S4(血清指标4) ○ S5(血清指标5) ▲ S6(血清指标6) ◆ 一年后病情进展



2.9.6.2 曲面图

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 设置行过滤条件。

2. 选择数据轴:

- 选择一列为“X 轴”，可选是否计为字符串，不选则计为数字。
- 选择一列为“Y 轴”，可选是否计为字符串，不选则计为数字。
- 选择一列为“Z 轴”，可选是否计为字符串，不选则计为数字。
- 选择其它要弹出的值。

3. 设置图参数：选择投影、颜色、宽度、高度、是否深色模式、是否显示网格。



mybox-data-6.5.9/generated/糖尿病的病情预测-标准化_曲面图_2022-09-27_19-05-19-913_30.html



糖尿病的病情预测-标准化 - 曲面图

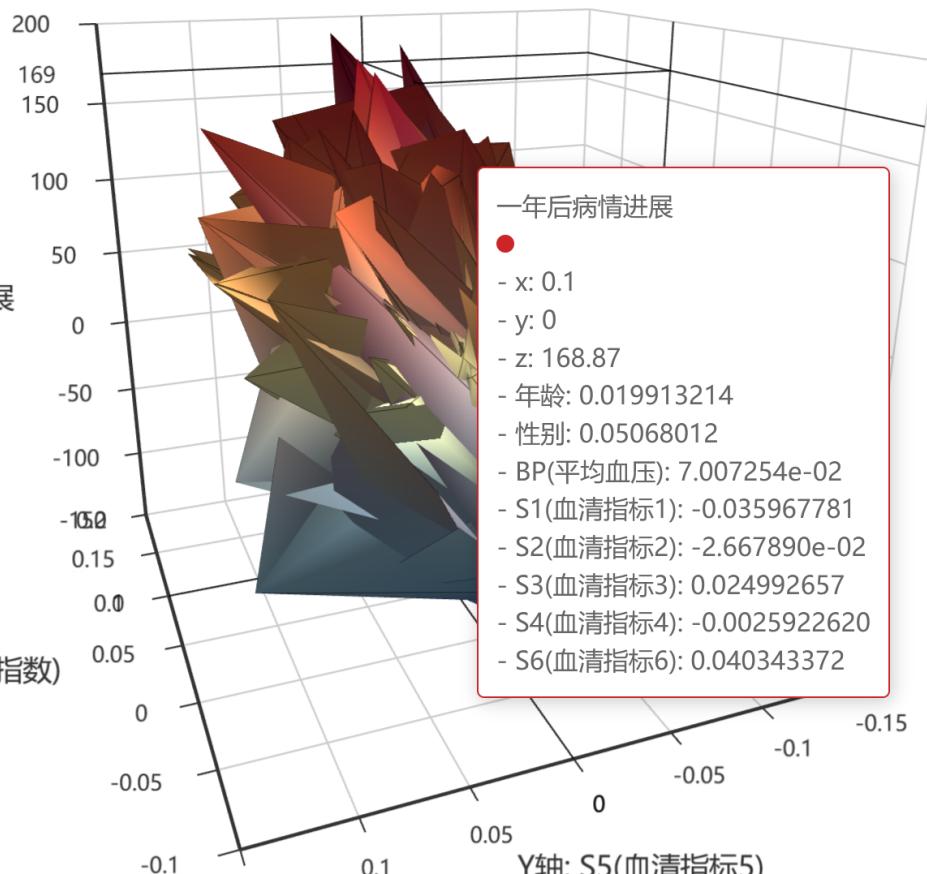
X轴: "BMI(体质指数)" Y轴: "S5(血清指标5)" Z轴: "一年后病情进展"

Z轴: 一年后病情进展



169

X轴: BMI(体质指数)



2.10 输出

悬停或点击功能按钮，选择“数据”下的功能。

The screenshot shows the MyBox Data Tool interface with a CSV file open for editing. A context menu is displayed over a row of data, specifically for '操作' (Operations). The menu items are:

- 保存 (Save)
- 恢复 (Restore)
- 刷新 (Refresh)
- 打开 (Open)
- 创建数据 (Create Data)
- 加载系统粘贴板中的内容 (Load content from system clipboard)**
- 导出 (Export)** (highlighted with a blue background)
- 转换为数据库表 (Convert to database table)
- 修改 (Modify)
- 整理 (Organize)
- 计算 (Calculate)
- 图 (Chart)
- 示例 (Example)
- 当鼠标悬停时弹出 (Show when mouse hovers)
- 关闭 (ESC/F6或点击对象外部任何一处) (Close (ESC/F6 or click anywhere outside the object))

The main window displays a table of GDP data for China from 1959 to 1964. The table has columns for year, value, and growth rate.

年份	值	增长
1959	1,447.5	1,447.5
1960	1,470.1	383.8
1961	1,232.3	441.1
1962	1,162.2	453.1
1963	1,248.3	497.5
1964	1,469.9	559

2.10.1 导出

1. 选择数据:

- 行可以是：当前页、选中的行、或所有页。
- 选择列，若不选则取所有列。
- 设置行过滤条件。

MyBox 导出 - CSV - 33 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

数据 行过滤 目标格式 目标文件 日志

行 选中的 当前页 所有页

列 

<input type="checkbox"/> 表格行	数据行	<input checked="" type="checkbox"/> 年	<input checked="" type="checkbox"/> 国民总...	<input checked="" type="checkbox"/> 国内生...	<input type="checkbox"/> 第一产...	<input type="checkbox"/> 第二产...	<input type="checkbox"/> 第三产...
<input type="checkbox"/> 1	1	1952	679.1	679.1	342.9	141.1	195.1
<input type="checkbox"/> 2	2	1953	824.4	824.4	378	191.6	254.8
<input type="checkbox"/> 3	3	1954	859.8	859.8	392	210.8	257
<input type="checkbox"/> 4	4	1955	911.6	911.6	421	221.5	269.1
<input type="checkbox"/> 5	5	1956	1,030.7	1,030.7	443.9	280.4	306.5
<input type="checkbox"/> 6	6	1957	1,071.4	1,071.4	430	316.6	324.8

选中的: 0 行: 50/70 页尺寸 50 页 1 /2     

2. 目标文件格式: csv、文本、excel、xml、json、html、pdf。

3. 可以设置目标文件的选项。

4. 可按最大行数分割导出后的文件。

MyBox 导出 - CSV - 33 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

数据 行过滤 目标格式 目标文件 日志

基础 CSV Excel 文本 html json PDF

格式 

按最大行数分割文件 不分隔 

行号  

2.10.2 转换为数据库表

1. 选择数据:
 - 行可以是：当前页、选中的行、或所有页。
 - 设置行过滤条件。
2. 选择要复制的列
3. 选择是否导入数据。



4. 生成自增主键、或选择主键列。



2.11 管理数据

此工具管理以下对象：

1. 数据文件

- 当 csv/excel/文本数据文件被相应编辑器打开时，增加/修改此文件的记录。
- 数据保存在数据文件中。
- 删除数据文件的记录不会导致数据文件本身被删除。

2. 数据粘贴板

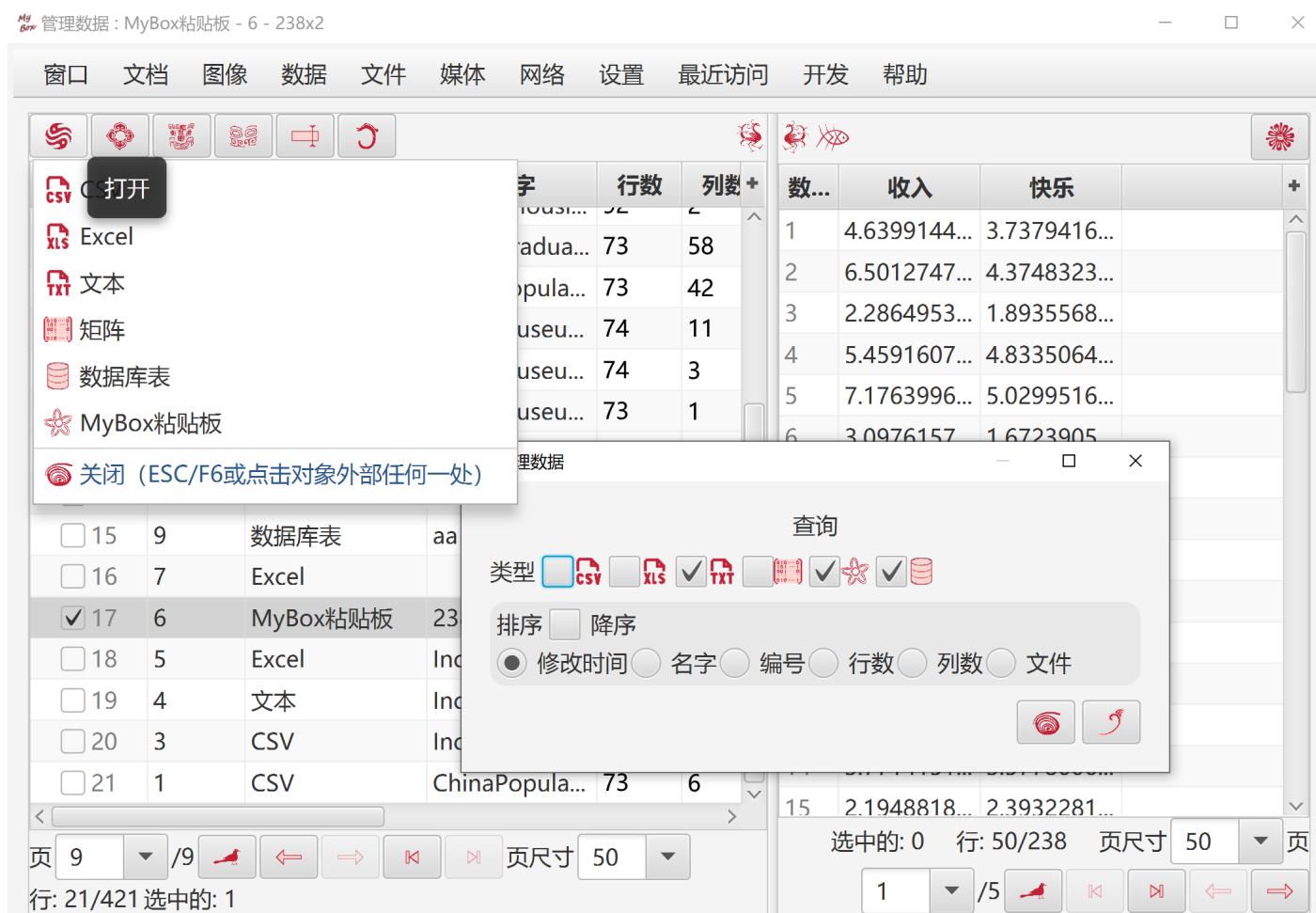
- 当数据复制到 MyBox 粘贴板中时，增加此数据的记录。
- 数据保存在 MyBox 内部目录下的文件中。
- 删除数据粘贴板的记录将会同时删除它的内部文件。

3. 矩阵

- 在矩阵管理器中增删改矩阵。
- 数据保存在 MyBox 数据库中。
- 删除矩阵的记录将会同时删除矩阵包含的数据。

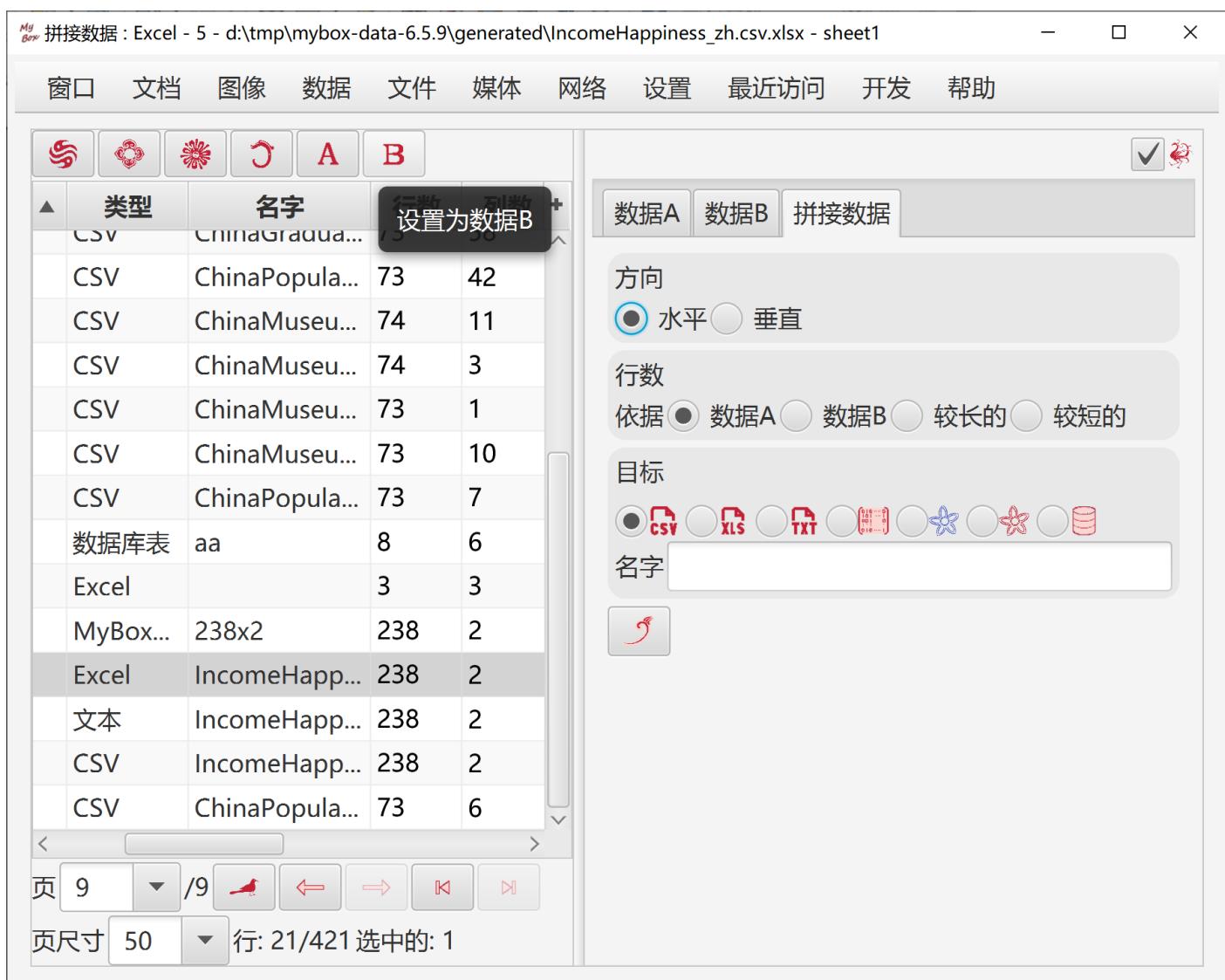
4. 数据库表

- 在数据库表管理器中增删改数据库表。
- 数据保存在 MyBox 数据库的表中。
- 删除数据库表的记录将会同时删除数据库表包含的数据。



2.12 拼接数据

1. 选择或打开两个数据。
2. 分别选择两个数据的行列：
 - 行可以是：当前页、选中的行、或所有页。
 - 选择列，若不选则取所有列。
 - 设置行过滤条件。
3. 拼接选项：
 - 方向：横向、纵向。
 - 行/列数按照：数据 A、数据 B、较长的、较短的。
4. 目标可为：新的 csv/excel/文本文件、矩阵、系统粘贴板、MyBox 粘贴板、数据库表。



2.13 数据文件

2.13.1 CSV 文件

1. 在文件中:
 - 一般地，首行（头）定义列名，其余每行定义一行数据。
 - 值被"分隔符"分离。分隔符可以是字符串。
 - 若值包含分隔符，它将被引号围起来。
 - 若行以"#"开头，则它将被略过。
2. 首次打开文件时，工具猜测它的分隔符和字符集。
3. 当文件读取异常时，改变选项然后点击按钮“刷新”。
4. 数据被分页。当页数大于 1 时，运行一些功能之前必须保存当前页的修改。
5. 选项包括：文件的字符集、是否以第一行作为字段的名字、字段的分隔符。
6. 对于字符串数据支持编辑和保存多行。
7. 数据可以保存为不同的字符集和换行符。

The screenshot shows the MyBox Data Tools interface for editing CSV files. The main window title is "编辑CSV文件 : CSV - 34 - d:\tmp\mybox-data-6.5.9\data\ChinaCPI_zh.csv". The menu bar includes: 窗口 (Window), 文档 (Document), 图像 (Image), 数据 (Data), 文件 (File), 媒体 (Media), 网络 (Network), 设置 (Settings), 最近访问 (Recent), 开发 (Development), and 帮助 (Help). The left sidebar has sections for 信息 (Information) and 格式 (Format). Under Format, there are settings for 字符集 (Character Set) (set to 自动检测 (Automatic detection) or 已知 (Known) with UTF-8 selected), and a checkbox for 第一行定义列名 (First row defines columns). Below that is a section for 文本分隔符 (Text separator) with various options like 制表符 (Tab), 空格 (Space), etc., and a preview area with a red circular icon. A note says: "若文件加载得不正确，改变选项并点击按钮“刷新”" (If the file does not load correctly, change the options and click the "Refresh" button). The right side shows a data grid with columns: 表格行 (Table Row), 数据行 (Data Row), 年 (Year), 居民消费 (Resident consumption), and a date column. The data rows show years from 2001 to 2013 with corresponding CPI values. At the bottom, there are buttons for navigating between pages (2 / 2), and a footer with page navigation icons.

表格行	数据行	年	居民消费
<input type="checkbox"/>	1	51	2001 100.7
<input type="checkbox"/>	2	52	2002 99.2
<input type="checkbox"/>	3	53	2003 101.2
<input type="checkbox"/>	4	54	2004 103.9
<input type="checkbox"/>	5	55	2005 101.8
<input type="checkbox"/>	6	56	2006 101.5
<input type="checkbox"/>	7	57	2007 104.8
<input type="checkbox"/>	8	58	2008 105.9
<input type="checkbox"/>	9	59	2009 99.3
<input type="checkbox"/>	10	60	2010 103.3
<input type="checkbox"/>	11	61	2011 105.4
<input type="checkbox"/>	12	62	2012 102.6
<input type="checkbox"/>	13	63	2013 102.6
<input type="checkbox"/>	14	64	2014 102.6

选中的: 0 行: 21/71 页尺寸: 50 页
2 / 2

2.13.2 文本文件

1. 在文件中:
 - 一般地，首行（头）定义列名，其余每行定义一行数据。
 - 值被"分隔符"分离。分隔符可以是字符串。解析文件时支持规则表达式。
 - 值不能包含分隔符。
 - 若行以"#"开头，则它将被略过。
2. 首次打开文件时，工具猜测它的分隔符和字符集。
3. 当文件读取异常时，改变选项然后点击按钮“刷新”。
4. 数据被分页。当页数大于 1 时，运行一些功能之前必须保存当前页的修改。
5. 选项包括：文件的字符集、是否以第一行作为字段的名字、字段的分隔符。
6. 不支持数据值多行。
7. 数据可以保存为不同的字符集和换行符。

The screenshot shows the MyBox Data Tool interface for editing a CSV file named '文本 - 1061'. The left sidebar has sections for 'Information', 'Format', 'Character Set' (set to 'Automatic detection'), 'Text Separator' (set to comma), and 'Header Row Definition' (checked). The main area displays a table of data with columns: 表格行 (Table Row), 数... (Number), 年 (Year), 年末总人... (Total Population at the end of the year), 男 (Male), and + (Add). The data shows population figures from 1949 to 1963.

表格行	数...	年	年末总人...	男	+
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,	
7	7	1955	61,465	31,	
8	8	1956	62,828	32,	
9	9	1957	64,653	33,	
10	10	1958	65,994	34,	
11	11	1959	67,207	34,	
12	12	1960	66,207	34,	
13	13	1961	65,859	33,	
14	14	1962	67,296	34,	
15	15	1963	69,172	35	

Bottom status bar: 选中的: 0 行: 50/73 页尺寸: 50 页: 1 /2

2.13.3 Excel 文件

1. 选择一个工作表单以处理。
2. 可以添加/重命名/删除工作表单。
3. 在文件中，一般地，首行（头）定义列名，其余每行定义一行数据。
4. 当文件读取异常时，改变选项然后点击按钮“刷新”。
5. 数据被分页。当页数大于 1 时，运行一些功能之前必须保存当前页的修改。
6. 对于字符串数据支持编辑和保存多行。
7. 数据可以只保存当前表单、或者保存全部表单。

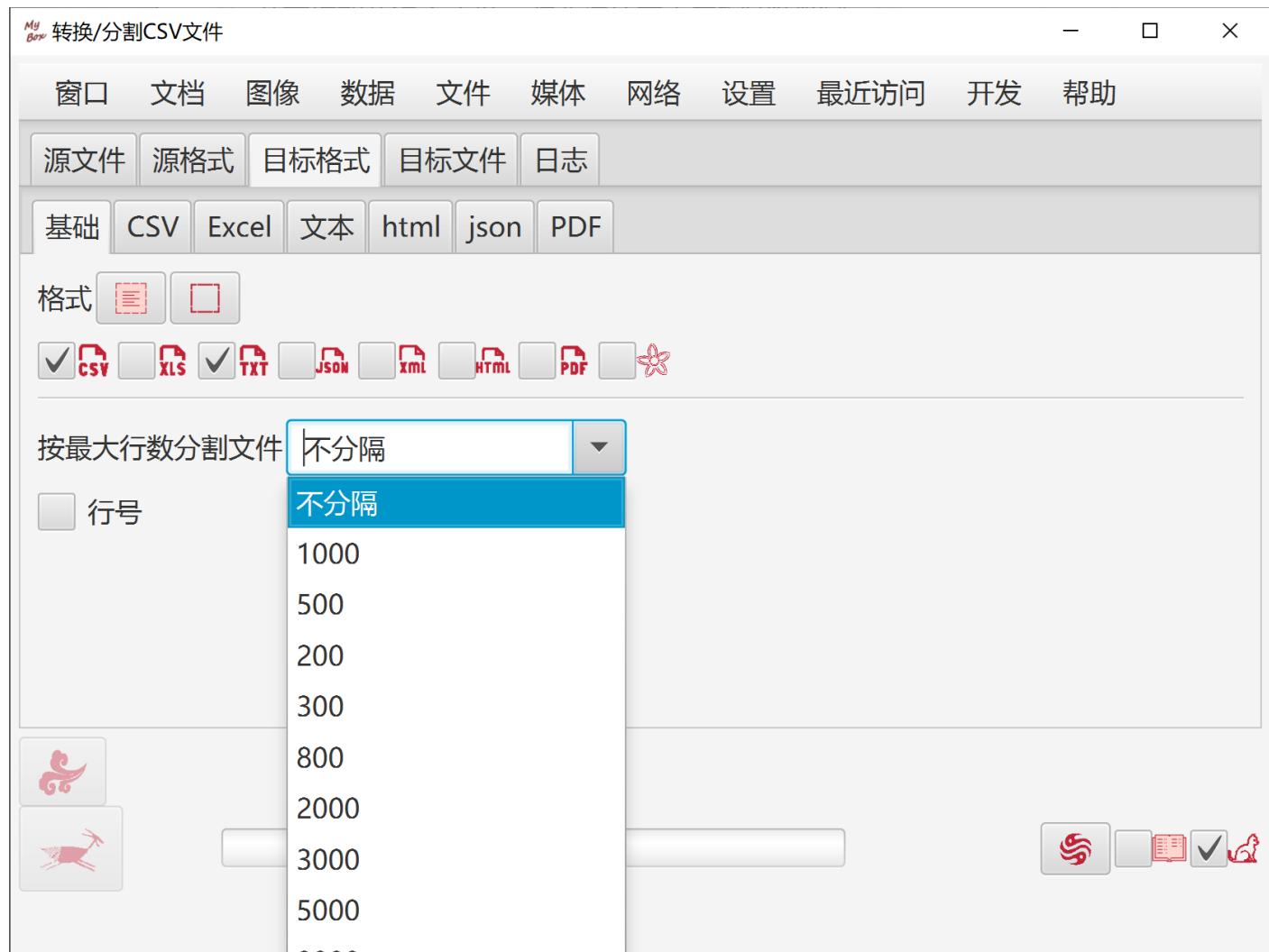
注意：工具只能处理 Excel 文件的基本数据。

如果文件包含格式、风格、公式、或图，建议把修改保存为新文件以免数据丢失。

表格行	数...	年	年末总人...	男性人口(..
1	1	1949	54,167	28,145
2	2	1950	55,196	28,669
3	3	1951	56,300	29,231
4	4	1952	57,482	29,833
5	5	1953	58,796	30,468
6	6	1954	60,266	31,242
7	7	1955	61,465	31,809
8	8	1956	62,828	32,536
9	9	1957	64,653	33,469
10	10	1958	65,994	34,195
11	11	1959	67,207	34,890
12	12	1960	66,207	34,283
13	13	1961	65,859	33,880
14	14	1962	67,296	34,517

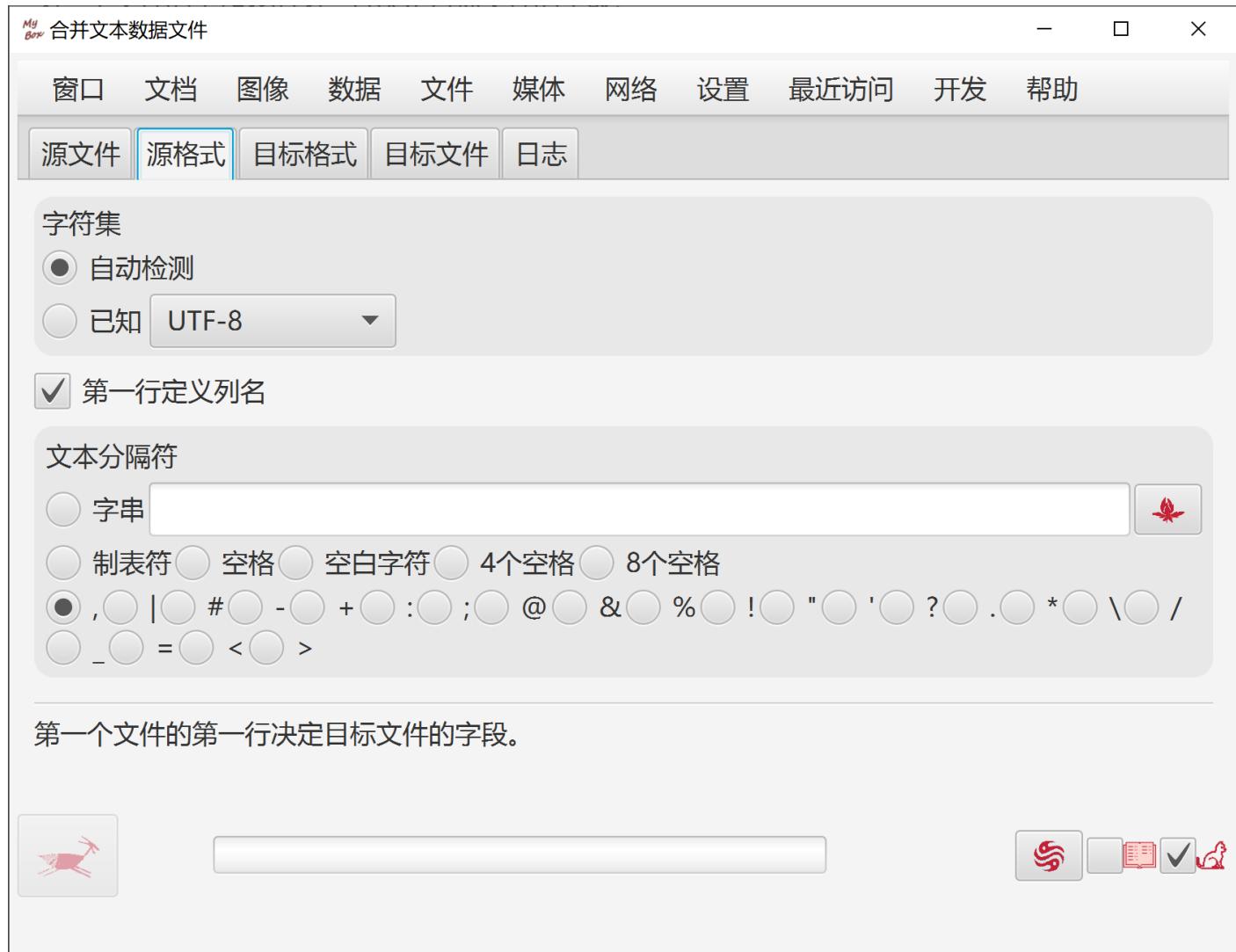
2.13.4 批量转换/分割数据文件

1. 源文件格式: csv、excel、文本。可以设置源文件的选项。
2. 目标文件格式: csv、文本、excel、xml、json、html、pdf。可以设置目标文件的选项。
3. 可按最大行数分割转换后的文件。



2.13.5 合并数据文件

1. 可以设置源文件的选项。
2. 可以设置目标格式。



2.14 系统粘贴板中数据

1. 读取并解析系统粘贴板的内容。
2. 可选数据分隔符：特殊字符、或输入的规则表达式。
3. 可选把第一行定义为列名。

MyBox 系统粘贴板中的数据

窗口 文档 图像 数据 文件 媒体 网络 设置 最近访问 开发 帮助

第一行定义列名   

源行号	花萼长度(cm)	花萼宽度(cm)	花瓣长度(cm)	花瓣宽度(cm)	种类
行1	5.1	3.5	1.4	0.2	Iris-setosa
行2	4.9	3.0	1.4	0.2	Iris-setosa
行3	4.7	3.2	1.3	0.2	Iris-setosa
	1.5	0.2			Iris-setosa
	1.4	0.2			Iris-setosa

MyBox 简易工具箱

文本分隔符

字串

制表符 空格 空白字符 4个空格

8个空格

, | # - + : ; @ & %

! " ' ? . * \ / _ =

< >

< > HTML

2.15 MyBox 粘贴板中的数据

MyBox 粘贴板中的数据 : MyBox粘贴板 - 6 - b

The screenshot shows the MyBox Data Tools application window. The menu bar includes: 窗口 (Window), 文档 (Document), 图像 (Image), 数据 (Data), 文件 (File), 媒体 (Media), 网络 (Network), 设置 (Settings), 最近访问 (Recent), 开发 (Development), and 帮助 (Help). The toolbar features various icons for file operations like copy, paste, and search.

The main area displays a table of clipboard entries:

	表格行	类型	名字	行数	列数	操作
<input type="checkbox"/>	1	MyBox粘贴板	a	6	3	
<input checked="" type="checkbox"/>	2	MyBox粘贴板	b	238	2	

A modal dialog box is open, titled "MyBox粘贴板中的数据 : ...", showing the current name "b" and a new name input field containing "bm". It has "确定" (Confirm) and "取消" (Cancel) buttons.

At the bottom, there are page navigation controls: 页 (Page), 1, /1, 50, 行 (Row), 2/2, 选中的 (Selected), 1, and various search and sort icons.

To the right, a detailed view of the clipboard entry "b" is shown in a table:

表格行	数...	收入	快乐
1	1	4.6399144...	3.737941
2	2	6.5012747...	4.374832
3	3	2.2864953...	1.893556
4	4	5.4591607...	4.833506
5	5	7.1763996...	5.029951
6	6	3.0976157...	1.672390
7	7	4.6475559...	1.497024
8	8	3.5345662...	2.667465
9	9	5.3615031...	5.231863
10	10	4.3170323...	3.661656

Bottom right: 选中的: 0 行: 50/238 页尺寸: 50 页

2.16 矩阵

2.16.1 管理和编辑矩阵

1. 编辑矩阵。
2. 矩阵可保存和取用。

The screenshot shows the MyBox Data Tools application window titled "管理矩阵 : 矩阵 - 1065 - a". The menu bar includes: 窗口 (Window), 文档 (Document), 图像 (Image), 数据 (Data), 文件 (File), 媒体 (Media), 网络 (Network), 设置 (Settings), 最近访问 (Recent), 开发 (Development), and 帮助 (Help). The toolbar features various icons for file operations like Open, Save, Print, and Copy/Paste.

The main area displays two matrices:

表格行	编号	类型	名字	行数	列数
<input checked="" type="checkbox"/>	1065	矩阵	a	3	3
<input type="checkbox"/>	1064	矩阵	b	6	4

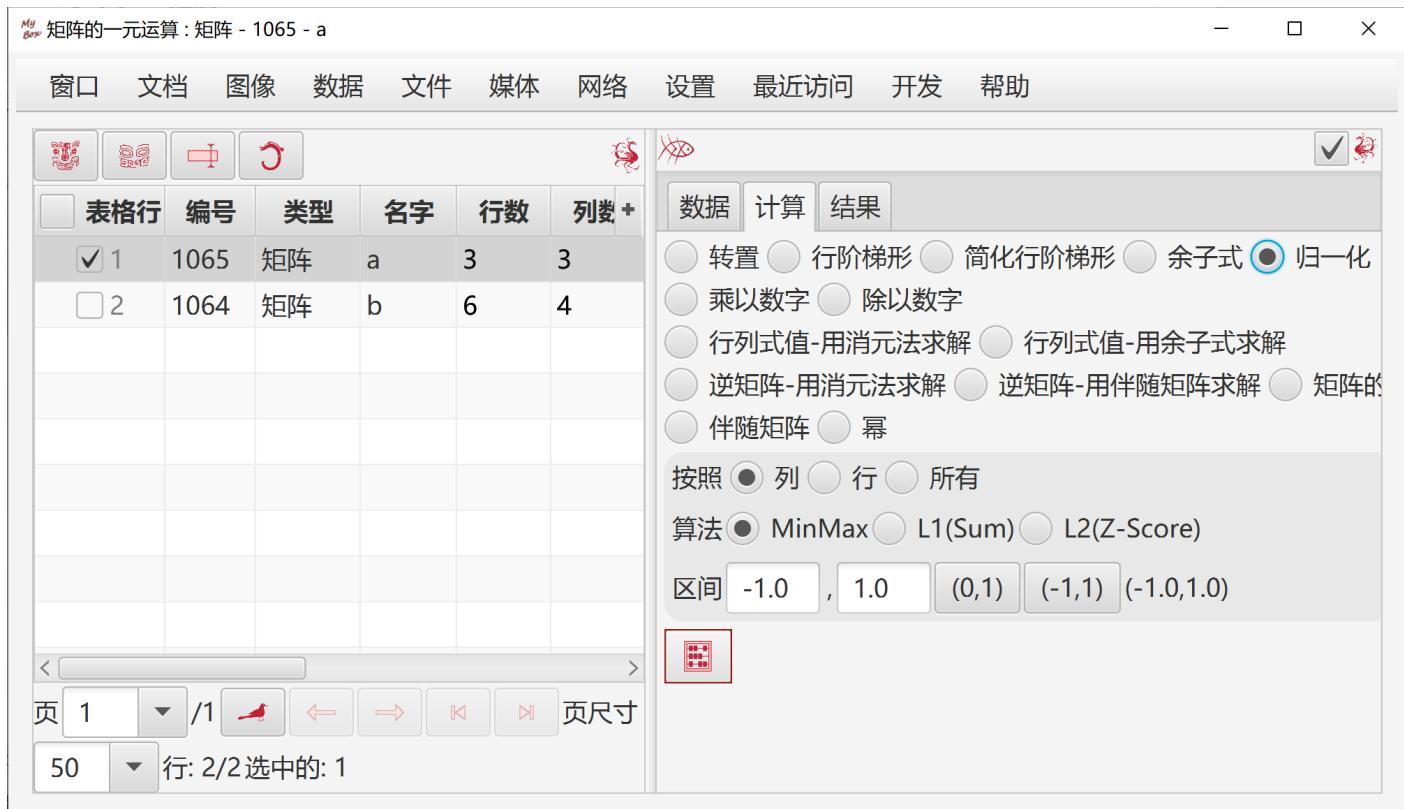
Below the matrices is a table with columns: 表格行, 数..., 列1, 列2, 列+. The data is as follows:

表格行	数...	列1	列2	列+
<input type="checkbox"/>	1	353.63	469.35	59.21
<input type="checkbox"/>	2	687.2	344.24	308.21
<input type="checkbox"/>	3	359.8	359.52	925.57

At the bottom left, there are page navigation buttons (Page 1, /1, back, forward, last) and a page size selector (50). A status bar at the bottom right indicates "行: 2/2 选中的: 1".

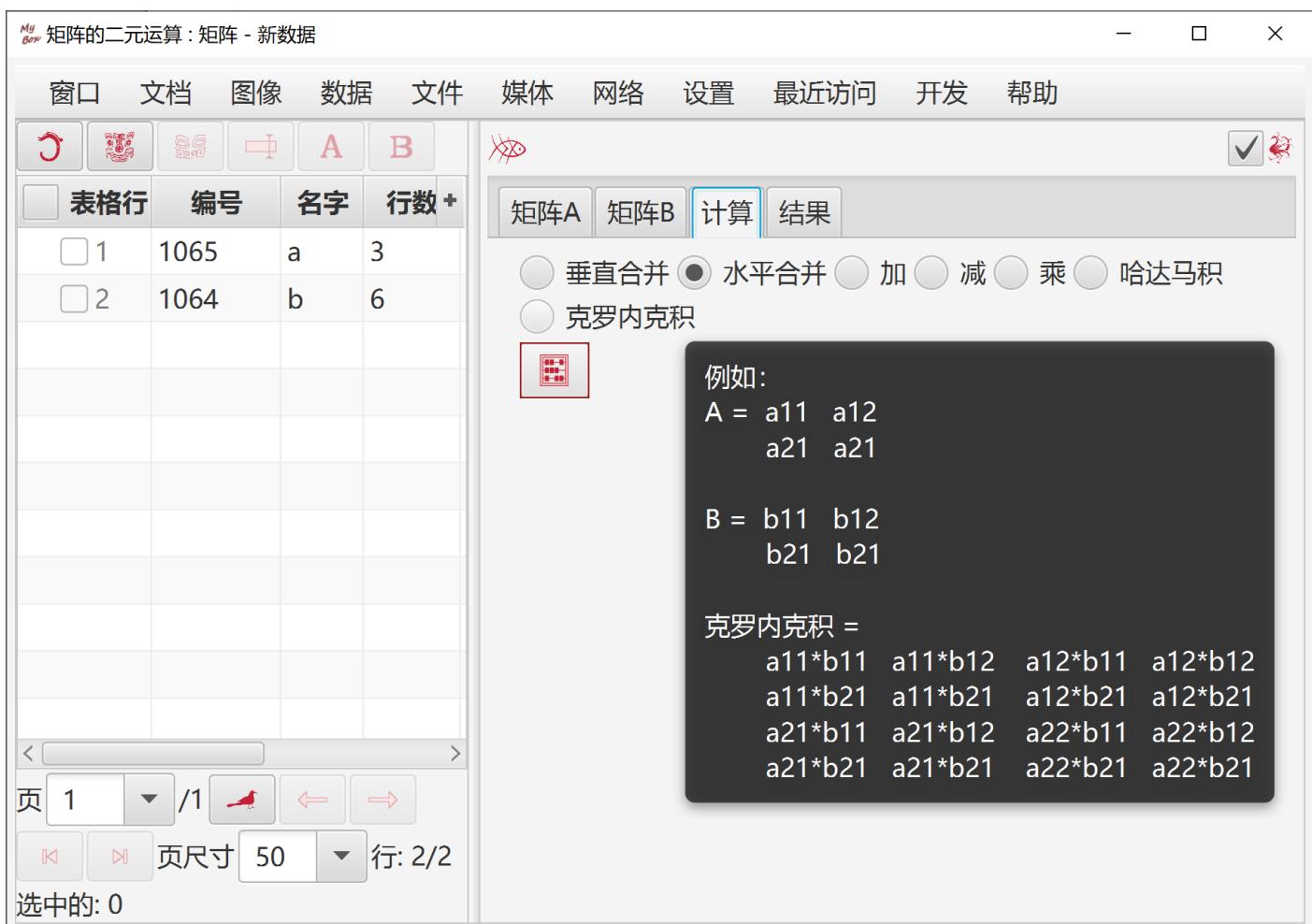
2.16.2 矩阵的一元计算

包括：转置、行阶梯形、简化行阶梯形、行列式值-用消元法求解、行列式值-用余子式求解、逆矩阵-用消元法求解、逆矩阵-用伴随矩阵求解、矩阵的秩、伴随矩阵、余子式、归一化、乘以数值、除以数值、幂。



2.16.3 矩阵的二元计算

包含：加、减、乘、克罗内克积、哈达马积、水平合并、垂直合并。



2.17 数据库表

2.17.1 管理数据库表

1. 可以查看表定义
2. 可以执行 SQL

The screenshot shows the MyBox Data Tools application window for managing database tables. The title bar reads "MyBox 数据库表 : 数据库表 - 387 - INCOMEHAPPINESS_ZH". The menu bar includes: 窗口 (Window), 文档 (Document), 图像 (Image), 数据 (Data), 文件 (File), 媒体 (Media), 网络 (Network), 设置 (Settings), 最近访问 (Recent), 开发 (Development), and 帮助 (Help). The toolbar contains various icons for operations like New, Open, Save, Print, and Database Management.

The main area displays the table structure for "INCOMEHAPPINESS_ZH". On the left is a list of tables with their IDs, names, and types. The table "387 IncomeHapp..." is selected, indicated by a checked checkbox. The table structure is shown in a grid:

表格行	数...	id	收入	快乐
1	1	1	3.8626474...	2.3144889...
2	2	2	4.6399144...	3.7379416...

Below the table structure is the SQL code for creating the table:

```
CREATE TABLE INCOMEHAPPINESS_ZH (
    id BIGINT NOT NULL GENERATED BY DEFAULT AS IDENTITY (START WITH 1, INCREMENT BY 1),
    收入 DOUBLE,
    快乐 DOUBLE,
    PRIMARY KEY (id)
)
```

The bottom status bar shows "页 1" (Page 1) and "页尺" (Page Ruler).

2.17.2 SQL 标识符的限制

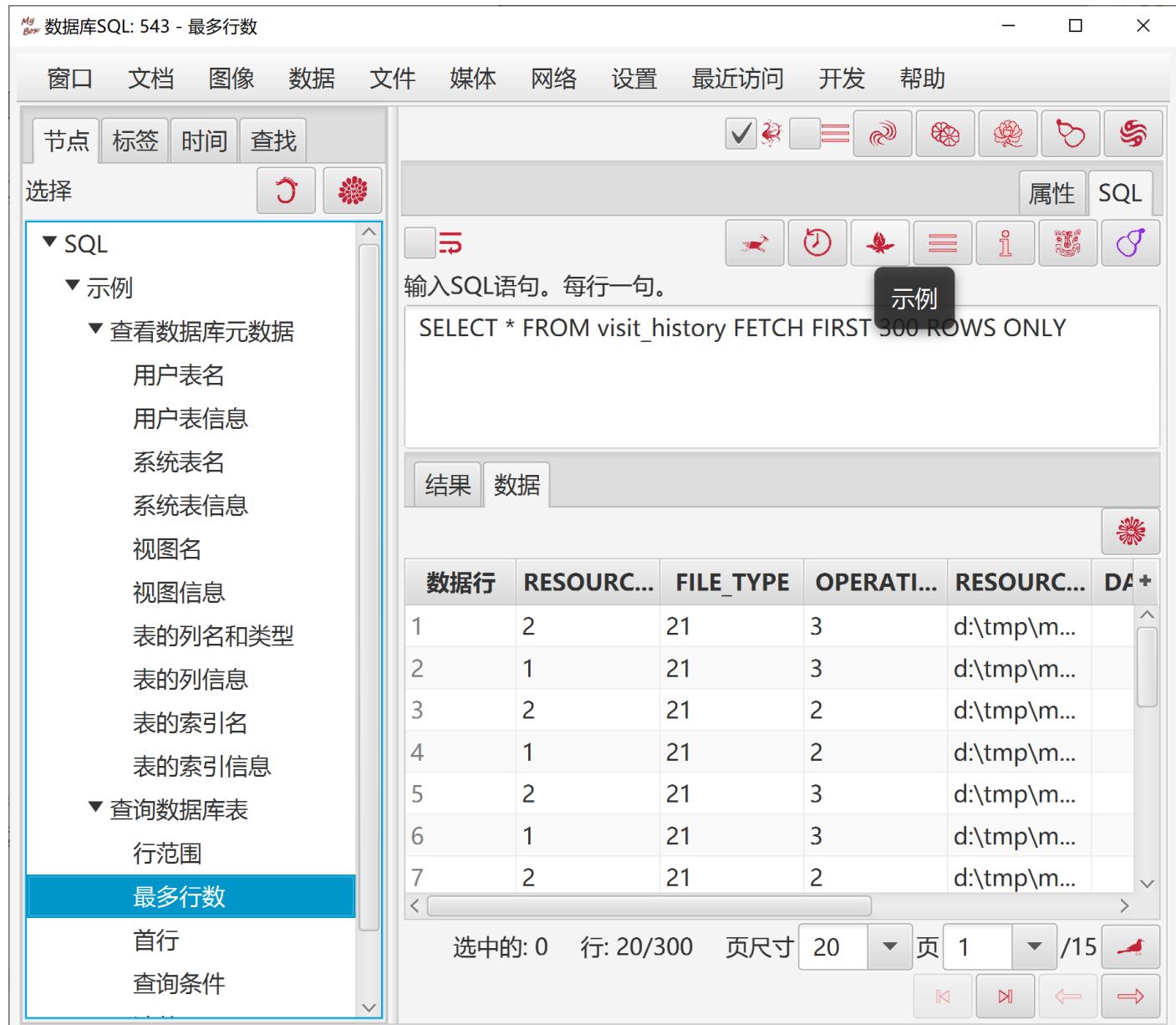
1. 表名和列名应当满足“SQL 标识符的限制”：

- 最大长度为 128。
 - “普通标识符”
 - 指未用双引号包围的标识符：
 - 只能由字母、数字、和下划线(_)构成。
 - 只能由字母开头。
 - 字母和数字可以是 Unicode（可以是中文）。
 - 不能是 SQL 保留字。
 - 当保存在数据库中，它被转换为大写的。
 - 当在 SQL 语句中引用它时，忽略大小写。
例如，AbC 等同于 abc 和 aBC。
 - “定界标识符”
 - 指用双引号包围的标识符。
 - 可以包含任何字符。
 - 当保存在数据库中时，它只取双引号包围的字串。
 - 当在 SQL 语句中引用它时，必须用双引号包围它，除了以下情形：它只包含大写字母和下划线。
例如，“AbC”不同于 AbC 或"ABC"，而"ABC"等同于 ABC 和 aBc.
2. 当 MyBox 创建表名或列名：
- 非法字符将被转换为下划线。
 - 若标识符开头不为字母，则字符“a”将被加到前面。
3. 数据库表被创建以后：
- 不能修改和删除主键的定义。
 - 可以添加和删除列定义，但是不能修改列定义。



2.17.3 数据库 SQL

1. 提供常用 SQL 语句的示例。
2. 自动生成所有用户数据表名的列表。
3. 查看所有用户数据表的定义。
4. 显示执行输出和查询结果。
5. SQL 代码可以被组织为树状信息。
6. 可以打开或保存为外部文件。



2.17.4 SQL 查询

数据库表有一个特殊的菜单项：“功能” - “整理” - “查询”，可以帮助方便地输入和执行 SQL 查询语句：

1. 表名和列名被列在左边
2. 提供示例、记录历史



A screenshot of the MyBox application interface showing a database table named 'CHINAPOPULATION_ZH'. The table has columns: id, 年, 年末总人口_万人_, 男性人口_万人_, 女性人口_万人_, 城镇人口_万人_, and 乡村人口_万人_. On the right, there is a preview pane displaying the following SQL query:

```
SELECT 年, 城镇人口_万人_, 乡村人口_万人_ FROM CHINAPOPULATION_ZH
```

The 'SQL*' tab is selected in the toolbar at the top right. There is also a '示例' (Example) button in the toolbar.

3 脚本与表达式

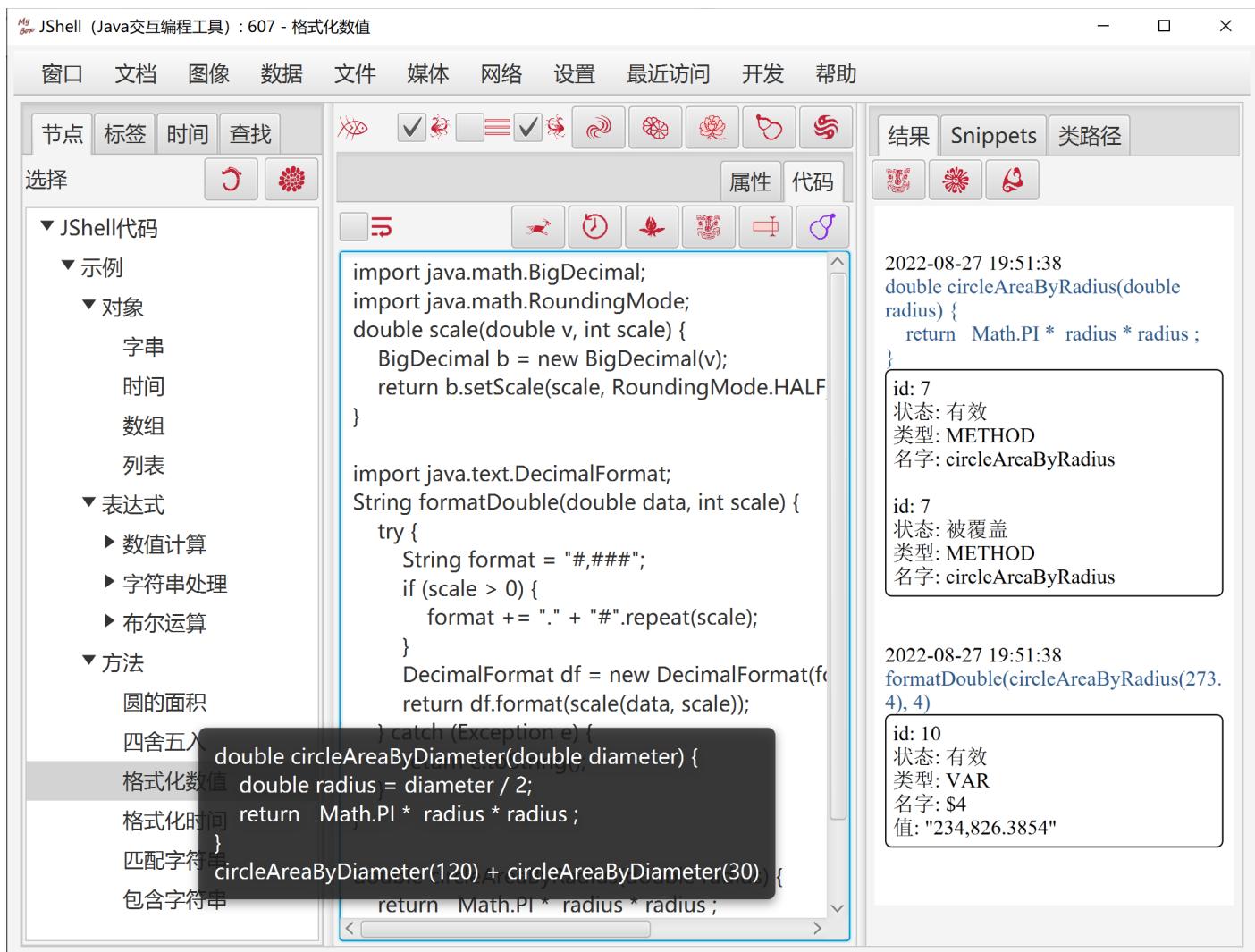
3.1 JShell (Java 交互编程工具)

JShell 是 JDK 包含的工具之一:

1. JShell 提供交互执行"snippets"的能力，即"读取-执行-输出"循环 (REPL)。
2. "Snippet"是 Java 编程语言的单个表达式、语句、或声明的代码：
 - 语句结尾必须有分号，而表达式不需要。
 - 可以定义变量和方法、然后调用它们。
3. 外部 Java 类应当可访问：
 - JShell 取系统环境中 CLASSPATH。
 - 其它 Jar 文件或路径可以附加在 CLASSPATH 后面。
 - 除了基本类，在调用多数 Java 类之前需要把它们导入进来。
4. JShell 可以用于科学计算、和调试 Java 代码。

此工具帮助图形化运行 JShell:

1. 输入若干 snippet，然后点击 "开始"按钮以执行它们：
2. Snippets 被逐个计算。
3. Snippets 的结果会影响后续的 snippets，即如"一个运行环境"。
4. 所有已执行过的 snippets 的属性被显示在一张表中。
5. 点击按钮"删除"或"清除"以把一些或全部 snippets 从当前环境中移除出去。
6. 点击按钮"重置"以清零 JShell，环境将变为空白。
7. 按下"CTRL+1"可以弹出代码的完成建议列表。
8. 若已添加了 MyBox 类路径，则可以引用 MyBox 的所有方法。
9. JShell 代码被组织为树形，并提供示例。



3.2 JEXL (Java 表达式语言)

JEXL (Java 表达式语言) 是一个库文件，以变量和脚本来动态生成值：

1. JEXL 与 Java 的语法有所不同，它更像是 javascript。
2. 在运行表达式/脚本之前，其中所有变量应当在 JexlContext 中有值。
3. 创建 Java 类的实例为本地变量以引用它们。当用 “new” 时注意写全包名。
4. JEXL 可用于科学计算和数据处理。

此工具帮助图形化运行 JEXL：

1. 输入 JEXL 表达式或脚本。

注意：用单引号而不是双引号来围住字符串。

2. 按以下格式输入 Java 代码，以设置 JexlContext：

```
jexlContext.set("name", value);
```

例如，设置以下语句以在表达式/脚本中使用 Math.PI：

```
jexlContext.set("Math", Math.class);
```

3. 输入 JEXL 脚本的参数（如果有）。以逗号分隔。

4. 点击按钮“开始”以计算表达式/脚本。

5. MyBox 在 JShell 环境下自动执行以下步骤：

- 把 MyBox 的库路径加到 CLASSPATH。
- 导入必要的 JEXL 包。
- 运行 JexlContext 代码。
- 带上参数（如果有）计算表达式/脚本。

6. 若所有变量和参数都有合法的值，则结果显示在右边面板中。

7. JEXL 代码被组织为树形，并提供示例。

MyBox 用户手册 - 数据工具 v6.6

The screenshot shows the MyBox Data Tools v6.6 application window. The left sidebar contains a tree view of code snippets categorized by type (JEXL代码, 示例, 布尔表达式, 数组包含, 集合不包含, 字符串匹配, 字符串包含, 与/或/否, 数值处理, 小数位数, 格式, 百分数, 时间, 现在, 格式, 语句, 循环, 函数) and a search bar. The main area has tabs for JEXL脚本 (selected), JEXL环境, and JEXL脚本的参数. The JEXL脚本 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 code editor containing:

```
var value = 3;
var array = [1,'2',3, 'hello'];
value =~ array;
```

The JEXL环境 tab shows the command:

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

The JEXL脚本的参数 tab is empty.

The right side of the interface shows a results panel with tabs for 结果, Snippets, and 类路径. It displays execution logs and variable details:

- Execution log: 2022-08-27 19:54:15, jexlScript = jexlEngine.createScript("var number = 37; var total = 518; var scale = 2; DoubleTools.percentage(number, total, scale);")
id: 32
状态: 有效
类型: STATEMENT
- Variable details: id: 33
状态: 有效
类型: EXPRESSION
名字: jexlScript
- Execution log: 2022-08-27 19:54:15, jexlScript.execute(jexlContext);
id: 34
状态: 有效
类型: VAR
名字: \$8
值: "7.14"

3.3 Javascript

此工具帮助管理 Javascript 代码:

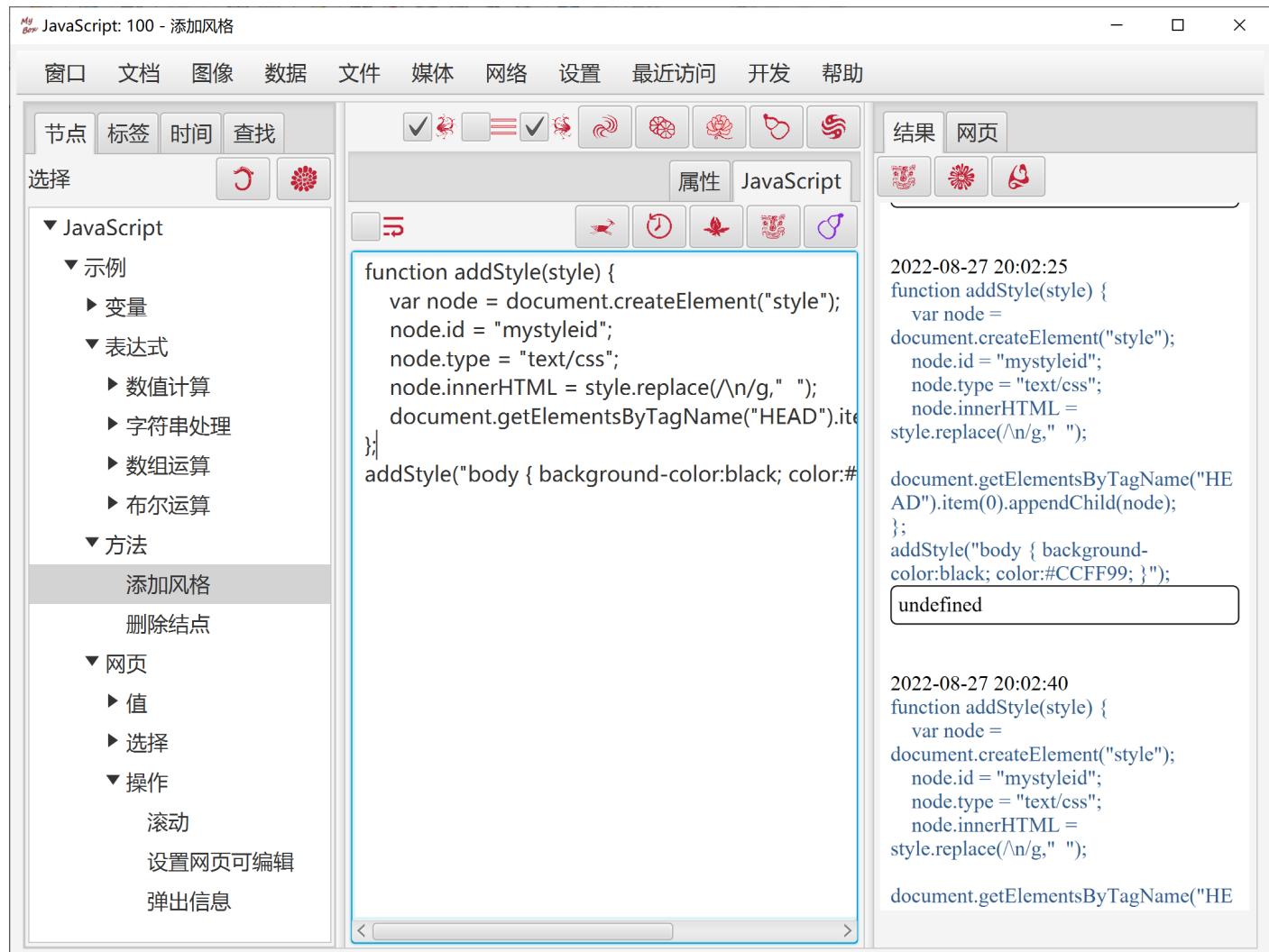
1. 编辑 JavaScript 代码。

脚本可以包含 WebEngine 可以解析的任何合法元素(ECMAScript 6)。

2. 运行脚本。

它的结果显示在右边的面板中。并且它也会影响右边面板中的网页。

3. Javascript 代码被组织为树形，并提供示例。



4 数学函数

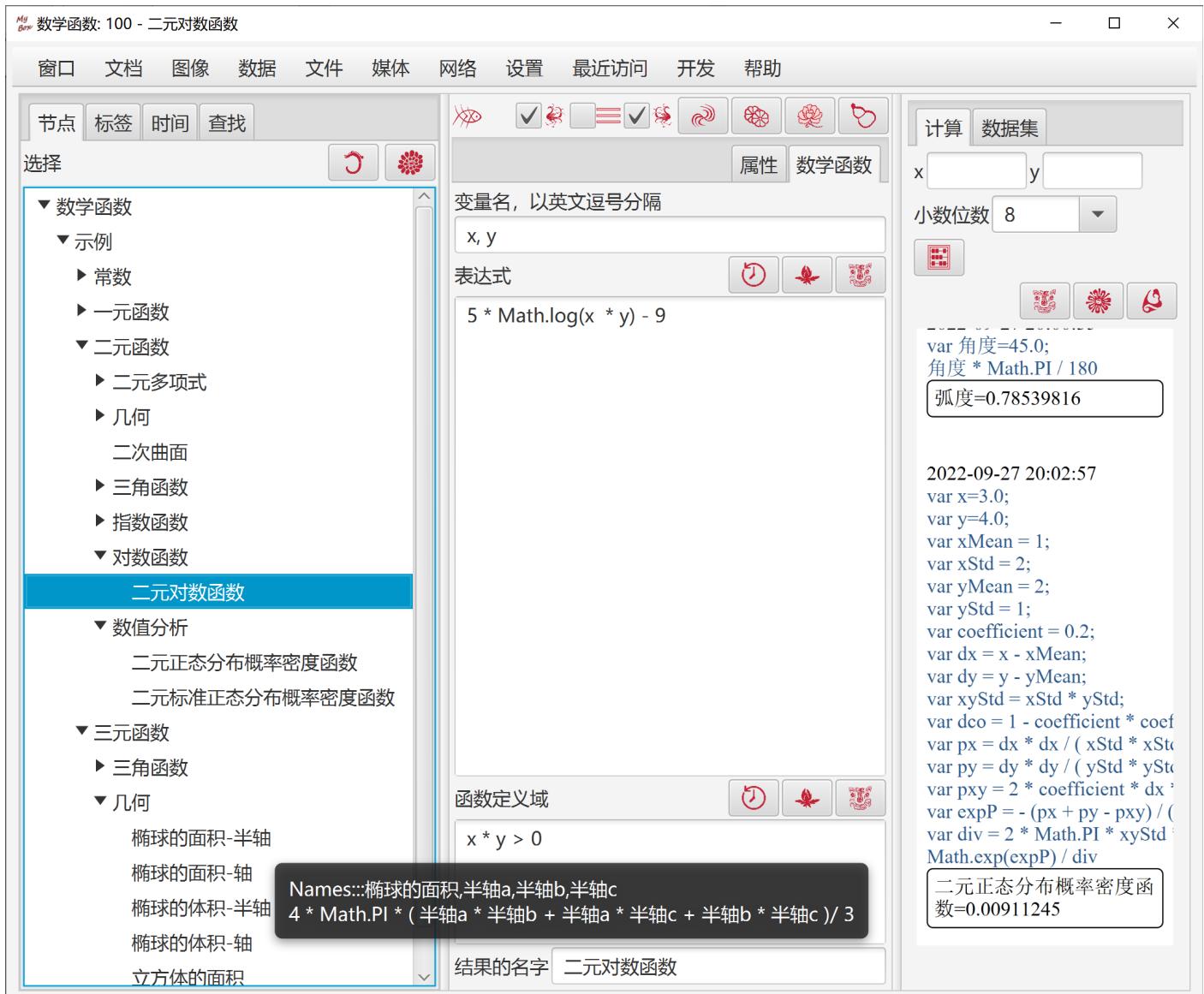
此工具帮助管理和计算数学函数。

4.1 定义函数:

1. 列出函数的变量名，以英文逗号分割。变量名应符合 JavaScript 要求：
 - 可以包含字母、数字、'-'、'_'、和 unicode (如中文)
 - 不能以数字开头。
2. 定义函数表达式为一段 JavaScript 代码：
 - 此脚本可以包含 Nashorn 可以解析的任何合法元素(ECMAScript 5.1)。
 - 它应当最终是一个数字。
 - 它可以引用给出的变量，但是不要申明它们。
3. 定义函数的定义域为一段 JavaScript 代码：
 - 空白表示定义域为所有实数。
 - 此脚本可以包含 Nashorn 可以解析的任何合法元素(ECMAScript 5.1)。
 - 它应当最终是一个布尔值（真或假）。
 - 它可以引用给出的变量，但是不要申明它们。
4. 给出计算结果的名字。
5. 数学函数被保存为树形。
6. 提供示例。

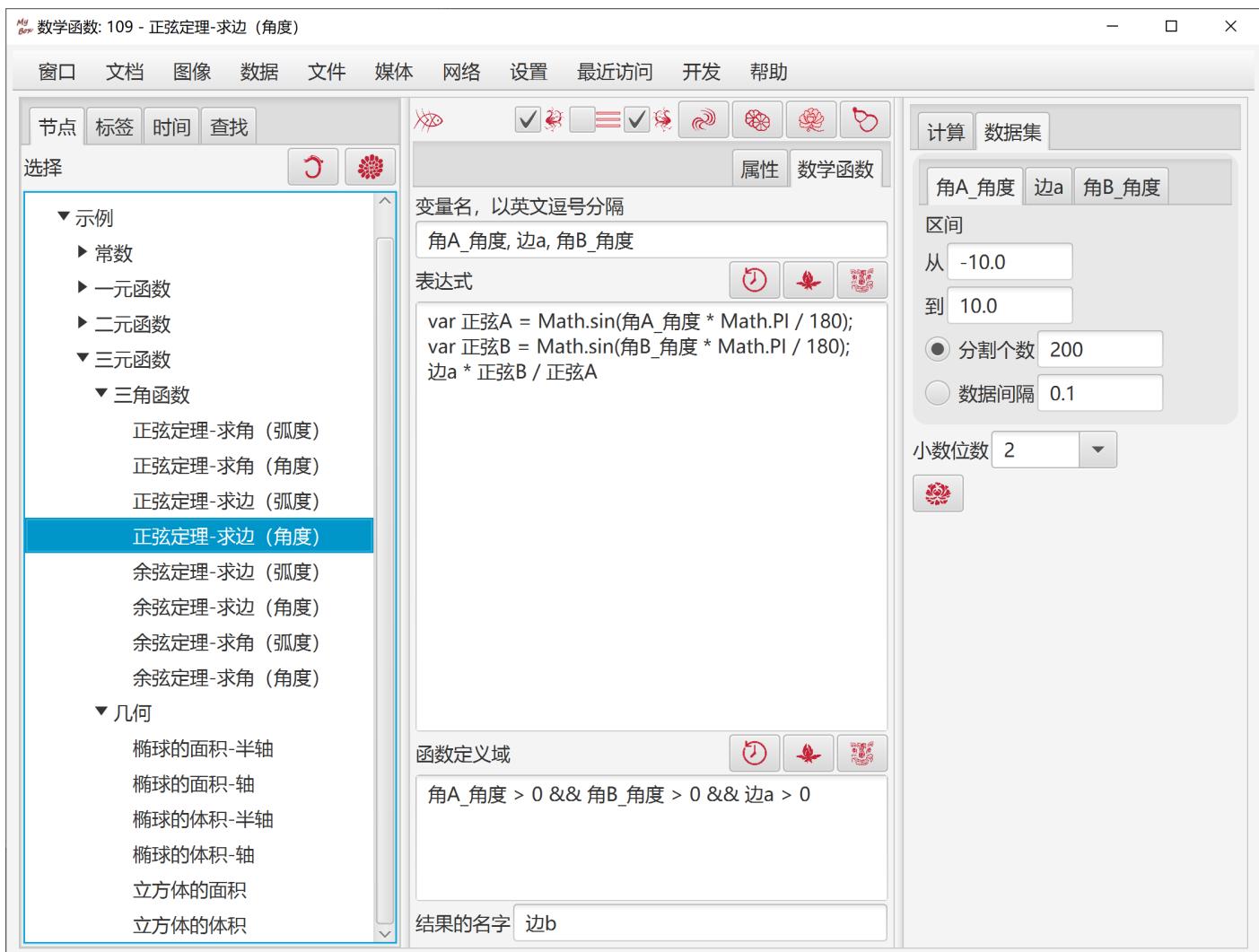
4.2 计算函数

1. 在工具计算脚本时，'var 变量名=<变量值>;' 被插入到脚本的前面，然后开始计算。
2. 若定义域脚本不为空，则插入给定的变量值并且检查结果。
3. 若为真，则继续计算函数表达式的脚本。否则跳过此组变量值。



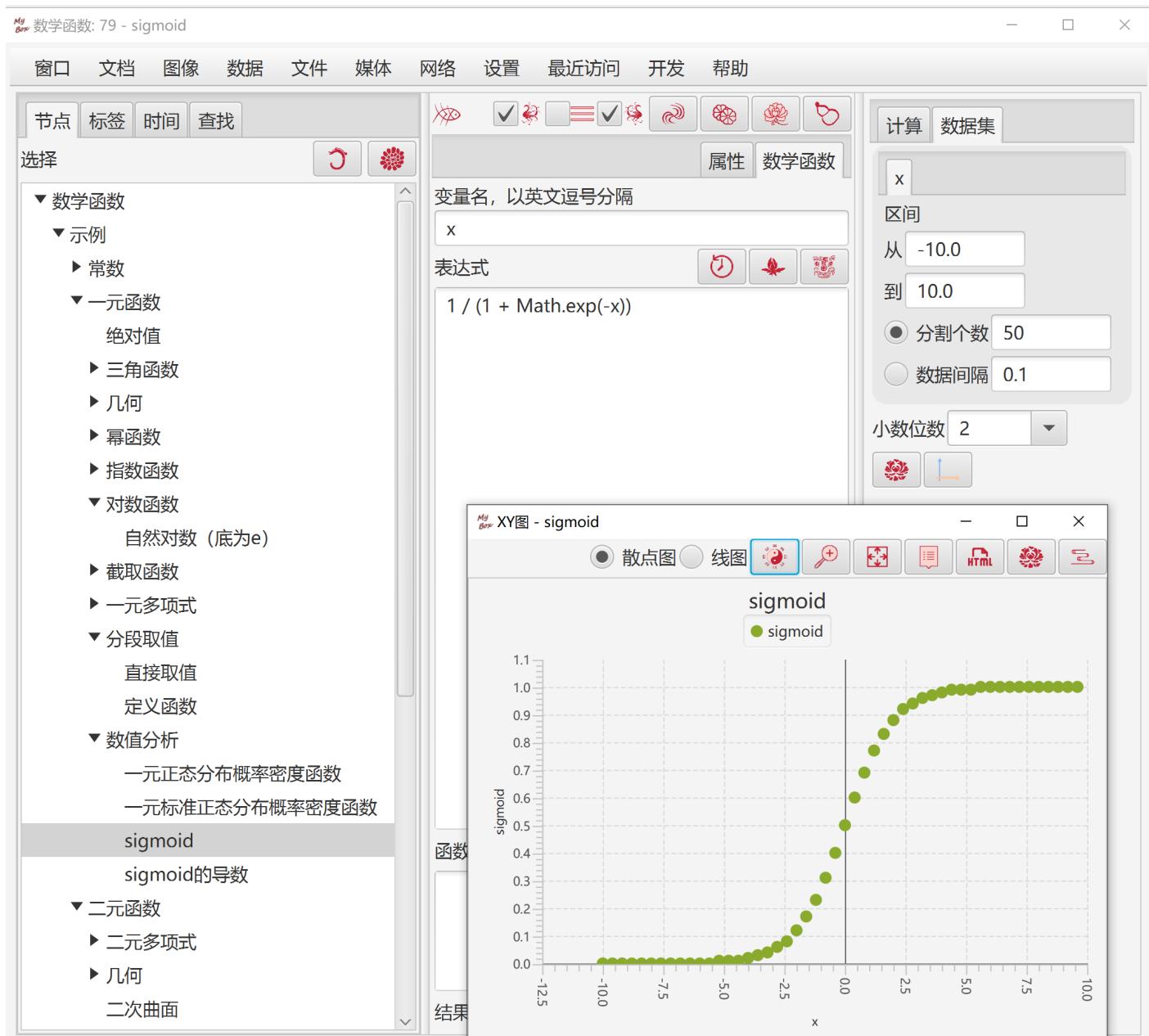
4.3 数据集

给定范围、间隔、和小数位数，可以生成函数的数据集。



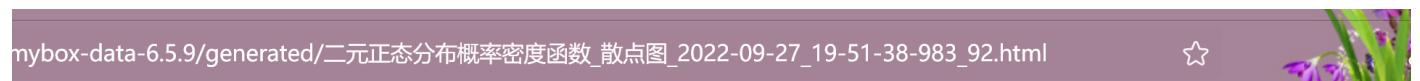
4.4 一元函数的 XY 图

对于一元函数，可以显示函数的 XY 图，包括散点图和线图。



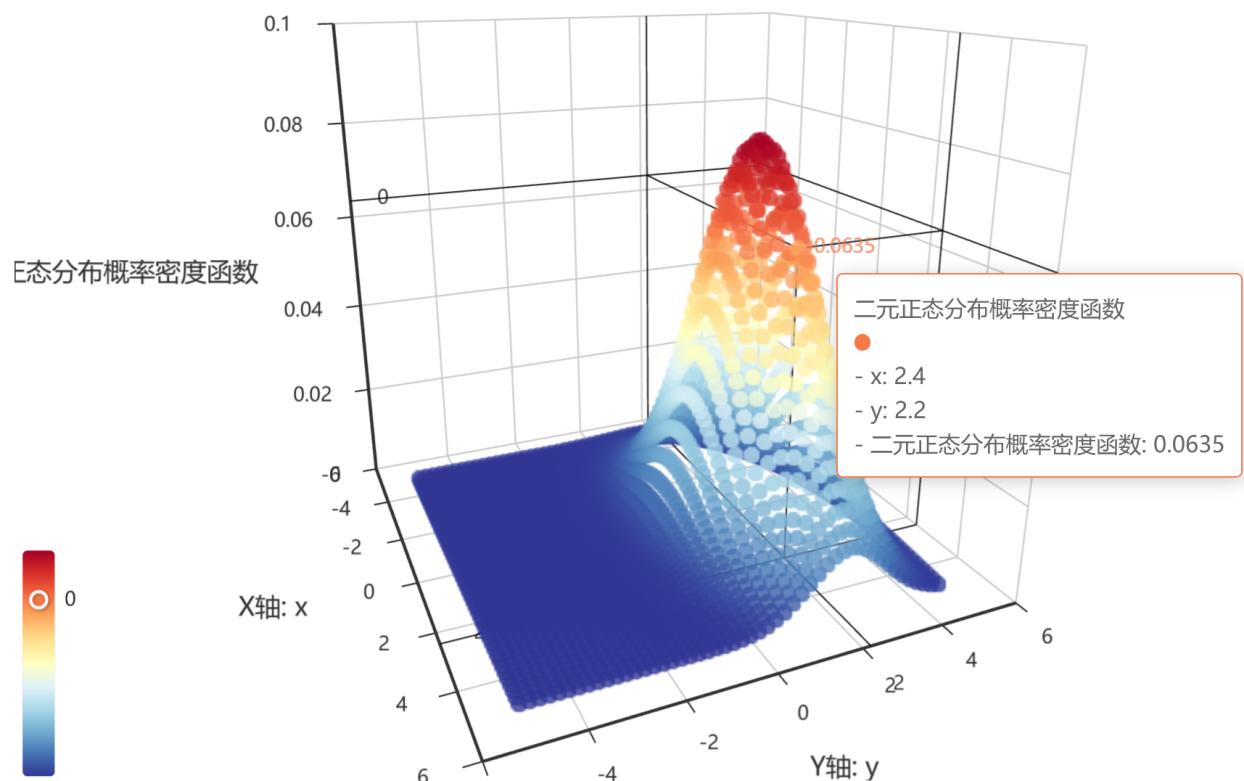
4.5 二元函数的XYZ图

对于二元函数，可以显示函数的XYZ图，包括三维散点图和曲面图。



X轴: "x" Y轴: "y"

Z轴: ● 二元正态分布概率密度函数



5 位置数据

5.1 数据约束

5.1.1 无效值

1. 整型数值 (integer/long/short) 的无效值是最小值 (MIN_VALUE)
2. 双精度数值的无效值是最大值 (Double.MAX_VALUE)

5.1.2 坐标系统

1. CGCS2000 (中国大地坐标), 真实位置, 近似于 WGS_84(GPS)坐标。
2. GCJ-02 (中国加密坐标), 经过加密的数据, 与真实位置有偏差。
3. WGS-84 (GPS), 真实位置
4. BD-09 (百度加密坐标), 基于 GCJ-02
5. Mapbar (图吧坐标), 基于 GCJ-02
6. 当坐标系统未定义或非法时, 缺省值为 CGCS2000。

5.1.3 坐标值

1. 数据处理时均用经纬度的小数而不是“度分秒”(DMS)。
2. MyBox 提供“位置工具”以帮助转换坐标的小数和度分秒。
3. 经度有效值范围: -180~180, 纬度有效值范围: -90~90。

5.1.4 时间

5.1.4.1 时间格式

- 日期时间, 如: 2014-06-11 13:51:33
- 日期, 如: 2014-06-11
- 年, 如: 2014
- 月, 如: 2014-06
- 时间, 如: 13:51:33
- 带毫秒的时间, 如: 13:51:33.261
- 带毫秒的日期时间, 如: 2014-06-11 13:51:33.261
- 带时区的日期时间, 如: 2020-09-27 12:29:29 +0800
- 带毫秒和时区的日期时间, 如: 2020-09-27 12:29:29.713 +0800
- 日期和时间之间可以有或没有“T”。 “2014-06-11T13:51:33”等同于“2014-06-11 13:51:33”。

5.1.4.2 纪元

“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 有效的时间示例

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 数据操作

1. 增/删/改/复制/清除/刷新。
2. 查询数据:
 - 定义和管理查询条件。
 - 当前查询条件被显示在"信息"页签上。
 - 符合条件的数据分页显示在"数据"表中。
 - 数据行可以按数据值显示不同颜色。
3. 导入数据, csv 格式:
 - 文件编码是 UTF-8 或 ASCII。
 - 第一行定义数据头, 以英文逗号分隔。
 - 其余每行定义一条数据, 字段以英文逗号分隔。
 - 字段的顺序随意。
 - 必要字段都应占位, 但不一定必须有值 (与具体数据有关)。
 - 可选是否替换已存在的数据。预定义数据或者示例数据总是替换。
4. 导出数据:
 - 定义和管理导出条件。
 - 导出的数据字段可选。
 - 导出格式可选: csv、xml、json、xlsx、html、pdf。
 - 可选择导出文件的分割行数。
 - 可导出当前数据页。
5. 删除数据:
 - 定义和管理删除条件。
 - 预定义的数据无法被删除。
 - 被引用的数据 (如外键引用) 无法被删除。
6. 定义、管理、和应用"条件":
 - "条件"被用来执行: 查询、删除、导出。
 - 在面板中设置条件:
 - 数据条件形成树结构, 树结点可以多选。
 - 排序的字段可以多选, 顺序可调。

- 编辑条件：标题、where、order by、fetch，它们被拼接成最终条件。
- 管理条件：增/删/改/复制。
- 被执行过的条件被自动保存。
- 最近执行过的条件被列出在按钮的弹出窗口中。

5.3 地图数据

1. 在地图上可以展示多种数据：地理编码、位置数据、或坐标查询。
2. 显示在地图上的数据可以是：
 - 满足当前查询的所有数据。可以设置“最多个数”以免性能问题。
 - 当前数据页。
3. 天地图：
 - 接受 CGCS2000 坐标数据，并把它们显示为无偏差的正确位置。
 - 对于其它坐标数据，MyBox 把它们转换为 CGCS2000 以显示正确位置。
 - 可选投影：EPSG:900913/3857（球面墨卡托）、EPSG:4326（经纬度直投）。
 - 可选控件：缩放、缩放比例、地图类型、符号。
 - 地图类型：地图、卫星、卫星混合、地形、地形混合。
 - 地图自带语言
 - 地图级别为 1-18
4. 高德地图：
 - 接受 GCJ-02 坐标，并把它们显示为无偏差的正确位置。
 - 对于其它坐标数据，MyBox 把它们转换为 GCJ-02 以显示正确位置。
 - 投影为 EPSG:900913/3857（球面墨卡托）。
 - 图层：
 - 可多选：标准图层、卫星图层、路网图层、交通图层。
 - 外国坐标不支持路网图层和交通图层。
 - 部分外国坐标支持卫星图层。
 - 可分别设置每个图层的不透明度。
 - 地图的语言：中文、英文、中英文。
 - 地图级别为 3-18
 - 可选“适应地图”，即将地图自动调整为可以显示所有数据的最佳大小和位置。
5. 调整地图级别：
 - 滑动鼠标滚轮。
 - 点击地图控件。
 - 选择“地图尺寸”。
6. 标注图片：
 - 可选：点（泡泡）、圆形、任意图片
 - 对于位置数据，还可选：数据集图像、数据图像。若无有效值，则为点。
 - 可设置标记图片的尺寸（长宽相同）。
7. 标注文字：
 - 可选：标签、坐标、地址。
 - 对于位置数据，还可选：开始时间、结束时间、数据值等。
 - 可多选。每个选择显示为一行。
 - 可设置标记文字的尺寸。
 - 可选是否粗体。
 - 可设置标记文字的颜色。对于位置数据，还可选：数据颜色。

8. 弹出信息:

- 鼠标放在标注上则可弹出更多信息。
- 可选是否弹出。

9. 截图:

- 可设置截图分辨率。
- 把当前地图和图中数据截取为 html

10. 地图的数据密钥可以在“设置”里修改。缺省的密钥是所有 MyBox 用户共享的免费密钥。

5.4 地理编码

5.4.1 数据定义

1. 基本属性：标识、级别、经度、纬度、中国名、英文名、5个代码、5个别名。
2. 从属属性：属主、洲、国家、省、市、县、镇、村、建筑。（构成“祖先”）
3. 辅助属性：高度、精度、坐标系统、面积（平方米）、人口、注释、是否预定义。

5.4.2 数据约束

1. 必须有值：标识、级别、中文名或英文名
2. “级别”的可取值：全球（只能是“地球”）、洲、国家、省（州）、市、县（区）、镇（乡）、村（居委会）、建筑、兴趣点。
3. 数据不必逐级从属，即可以跨级定义，例如：一个村庄直接属于南极洲；又如：城市直接属于国家，而没有省/州一级。
4. 匹配数据：
 - 以下方式之一可以确定一个地址：
 - 匹配数据标识（由 MyBox 自动赋值）。这是精确匹配。
 - 匹配“级别” + “祖先” + “中文名”或“英文名”或任一别名”。这是精确匹配。
 - 匹配“级别” + “中文名”或“英文名”或任一“别名”。这是模糊匹配，可能有同级重名导致匹配错误的情况。
 - 匹配名字或者别名时，不区分大小写。
 - 有时候“代码”（code1/2/3/4/5）也可以辅助查找。

5.4.3 编辑数据

1. 数据的“从属关系”只能从位置树上选择。
2. 数据的级别必须比祖先低。
3. 数据必须有中文名或者英文名。
4. 可在地图上选择/显示坐标。
5. 对选择的数据项设置为：“预定义的数据”、或“输入的数据”。

5.4.4 定义条件

地理代码按级别和从属关系形成一棵树，可多选

5.4.5 导入数据

5.4.5.1 内置的预定义数据

洲、国家、中国的省/市/县。国家的“面积”和“人口”有有效值。

5.4.5.2 CSV 格式

- 下载地址：
https://github.com/Mararsh/MyBox_data/tree/master/md/GeographyCode
- 以下为必要字段：

Level,Longitude,Latitude
以及"Chinese Name"或 "English Name"

- 以下是可选字段:
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 来自 geoname.org 的位置数据:

- 下载地址:
<http://download.geonames.org/export/zip/>
- 以制表符分隔的文本，编码为 UTF-8。
- 字段:
countryCode postalCode placeName adminName1 adminCode1 adminName2 adminCode2 adminName3 adminCode3 latitude longitude accuracy
- 坐标系统是 WGS_84。
- 同一地址只写一次，即使它有多个邮编或者坐标。

5.4.6 设置

1. 定制数据行颜色。
2. 提供"缺省"和"随机"按钮。



5.5 地图上的位置

1. 查询地理代码:
 - 点击地图
 - 输入地址
 - 天地图支持中外地址的中英文（如“伦敦”、“Paris”）
 - 高德地图只支持中国地址的中文。
 - 输入坐标
2. 可以保存查询出来的地理代码。



5.6 位置数据

5.6.1 数据定义

1. 基本属性：数据集、标签、经度、纬度、开始时间、结束时间。
2. 辅助属性：地址、高度、精度、坐标系统、速度、方向、数据值、数据规模、图像、说明。

5.6.2 数据集

1. 每个位置数据都属于一个数据集。
2. 数据集定义其所包含的位置数据的共同属性，包括：
 - 时间格式
 - 对于公元后的时间是否省略“公元”
 - 文字的颜色
 - 图像

这些属性有助于在地图上显著区分数据点。

5.6.3 定义条件

1. 数据集列表，可多选
2. 时间树（开始时间），可多选

5.6.4 地图数据

1. 开始显示地图数据时，以第一个数据为地图中心。
2. 位置分布：所有数据全部显示在地图上。
3. 时间序列：
 - 按“开始时间”升序逐帧显示数据
 - 若选择“累加”，则已显示帧的数据点不被抹除，即数据点逐帧叠加。
 - 若选择“时间叠加”，则凡是时间区间与当前帧的时间区间有重叠的数据都算作当前帧的有效数据。

例如，当前帧的开始时间是“公元前 1044”、结束时间是“公元前 221”，则所有起止区间与这个时间段有重合的数据都显示在当前帧中。

- 若选择“移到中心”，则每帧都会调整地图中心。
- 若选择“链接”，则显示相邻两帧的连线。

注：只有天地图的国内地址才正常显示连线。

- 帧控制：
 - 可以设置时间间隔
 - 可以选择帧（开始时间）
 - 暂停/播放
 - 前/后帧
 - 是否循环播放

5.6.5 截图

1. 对于“位置分布”，可选

- html: 数据和当前帧的截图
 - 当前帧的截图。格式自选。
2. 对于“时间序列”，还可以选择：
- jpg: 所有帧的截图
 - png: 所有帧的截图
 - 动态 gif: 所有帧的截图（可能内存不够）

5.6.6 导入数据

若数据包含的数据集在数据库中还未定义，则自动添加到数据库中。

5.6.6.1 CSV 格式

- 以下为必要字段：
Dataset,Longitude,Latitude
- 以下是可选字段：
Label,Address,Altitude,Precision,Speed,Direction,Coordinate System,
Data Value,Data Size,Start Time,End Time,Image,Comments

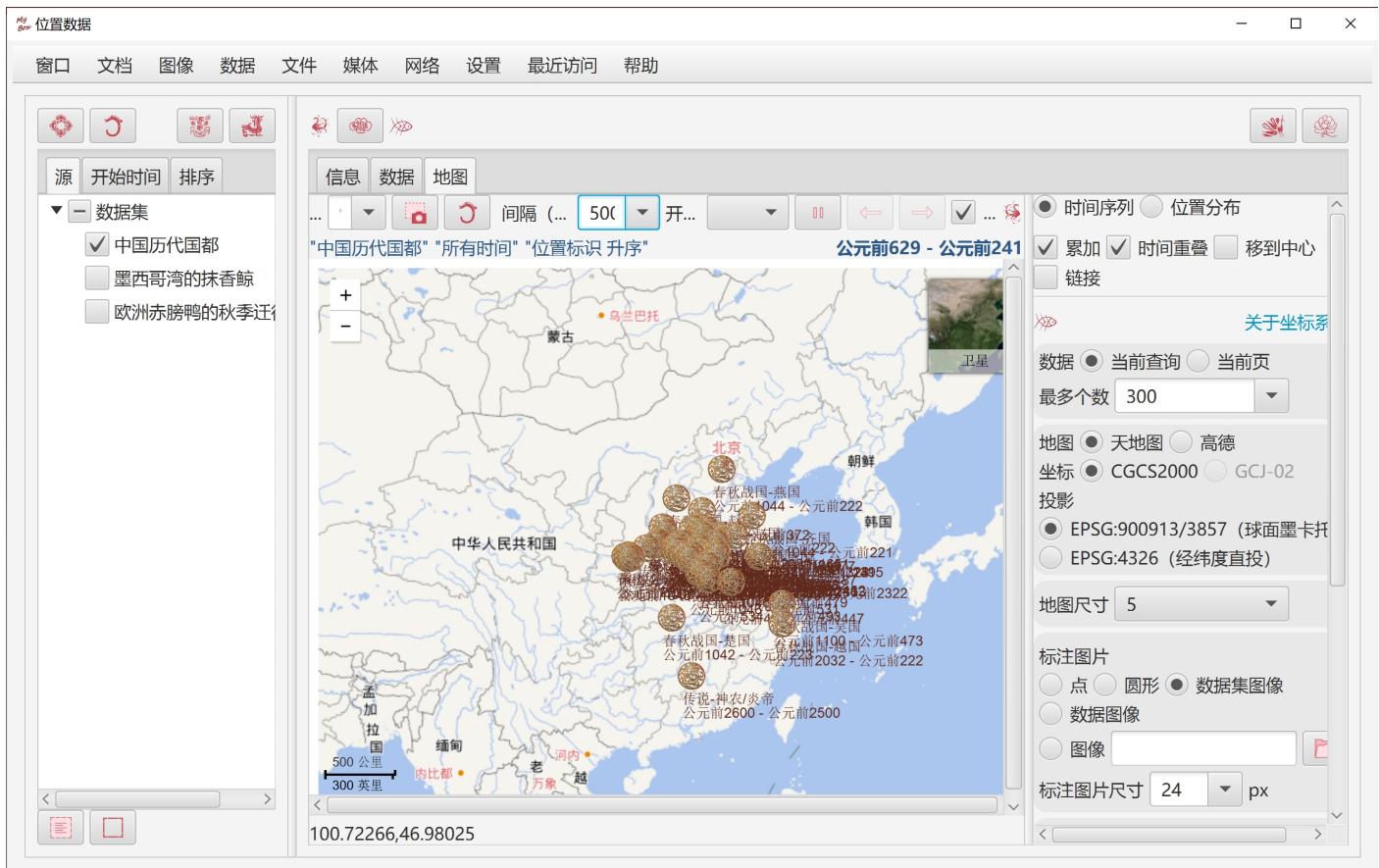
5.6.6.2 来自 movebank.org 的位置数据。

- 下载地址：
<https://www.datarepository.movebank.org/>
- 以逗号分隔 CSV 格式。
- 以下为必要字段：
timestamp,location-long,location-lat,study-name
- 坐标系统是 WGS_84。

5.6.6.3 示例

1. 中国历代都城
2. 欧洲赤膀鸭的秋季迁徙模式
3. 墨西哥湾的抹香鲸

MyBox 用户手册 - 数据工具 v6.6

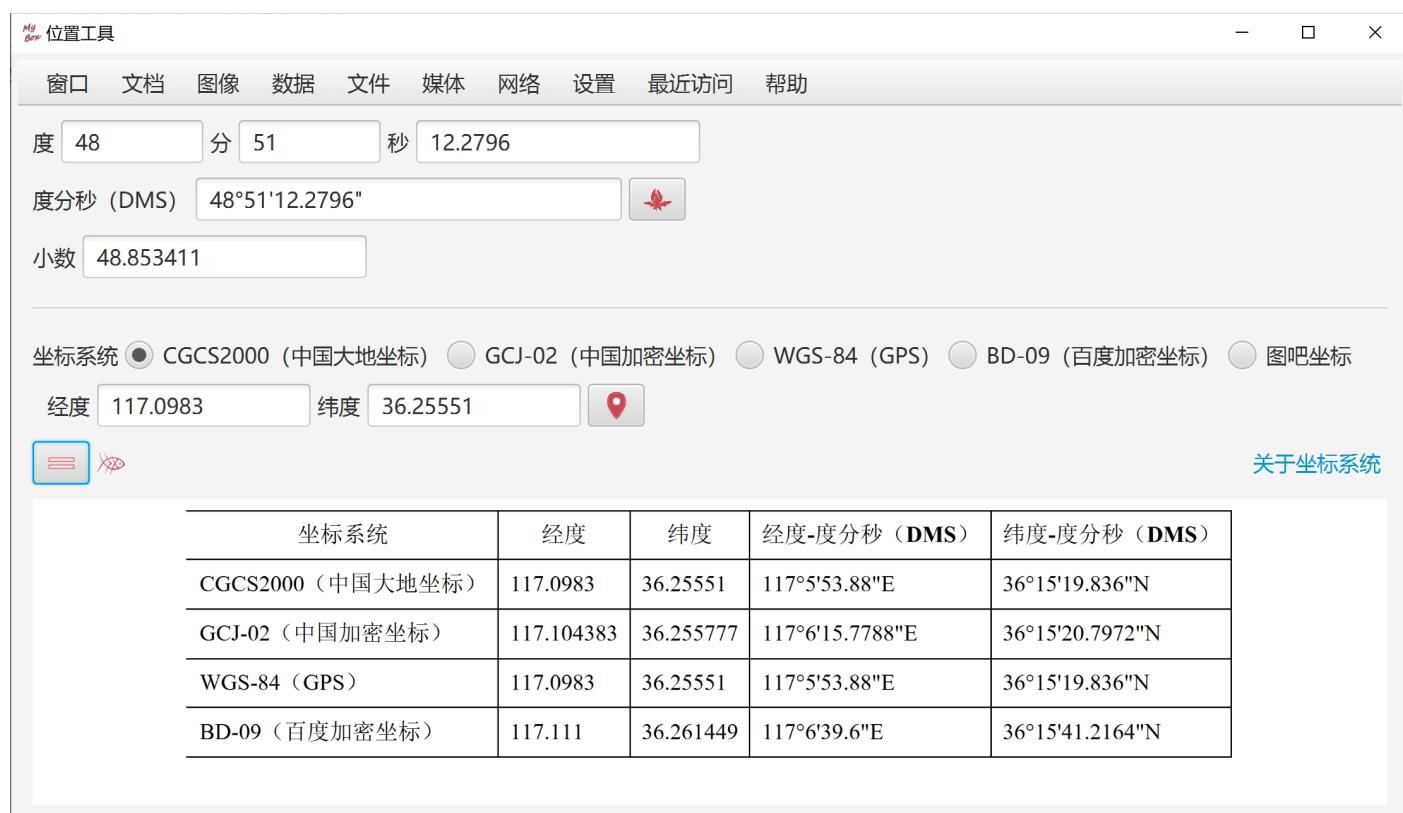


5.7 位置工具

- 转换坐标的小数和度分秒。有效的“度分秒”（DMS）示例：

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

- 把坐标转换为其它坐标系。



5.8 疫情报告

5.8.1 数据定义

1. 基本属性：数据集、时间、位置、来源。
2. 基础数值：确认、治愈、死亡。
3. 差值统计：新增确认、新增治愈、新增死亡。
由前后两天数据行的差值计算出。
4. 除值统计：
 - 治愈/确认千分比、死亡/确认千分比。
 - 确认/平方公里千分比、治愈/平方公里千分比、死亡/平方公里千分比。
 - 确认/人口千分比、治愈/人口千分比、死亡/人口千分比。
当位置数据的"面积"/"人口"值非法（零或负数）时，相应的统计值没有意义（零）。
预定义的"国家"数据有合法的"面积"和"人口"值，因此可以得到有意义的统计值。
5. 累加统计：
 - 一些国家的数值
由此国家的省的数值累加计算出。
 - 洲的数值
由此洲的国家的数值累加计算出。
 - 全球的数值
由各洲的数值累加计算出。

5.8.2 数据约束

1. 必须有值：数据集、时间、位置
2. "来源"的取值：“输入的数据”、“预定义数据”、“填充的数据”、“统计数据”。
3. "位置"是"地理编码"的外键，即所引用的位置必须在"地理编码"中有定义。
4. “确认”、“治愈”、“死亡”这三个值至少有一个大于零。
5. 以下方式之一可以确定一个疫情报告：
 - 匹配数据标识（由 MyBox 自动赋值）。这是精确匹配。
 - 匹配"数据集 +“日期” +“位置””。这是精确匹配。
 - 此版本假设，对于确定的数据集、确定的地址，一天只有一个有效数据。

5.8.3 编辑数据

1. 输入单个数据时，“位置”只能从位置树上选择。
2. 在“中国省会疫情报告”和“全球疫情报告”界面上，可以填写多个地址的同一数据集同一天的数据。
3. 对选择的多个数据项修改“来源”的取值。

5.8.4 导入数据

5.8.4.1 MyBox 内置预定义数据

来自约翰霍普金斯大学的 COVID-19 数据（直到 2020-09-24）。

5.8.4.2 CSV 格式

1. 下载地址：
https://github.com/Mararsh/MyBox_data/tree/master/md/COVID19
2. 以下为必要字段：
数据集,时间,确认,治愈,死亡
以及足以定义一个地理编码的位置数据：
经度,纬度,级别,洲,国家,省,市,区县,乡镇,村庄,建筑物,兴趣点
3. 以下是可选字段：
新增确诊,新增治愈,新增死亡
4. 坐标系统是 CGCS2000。

5.8.4.3 来自约翰霍普金斯大学的 COVID-19 历史数据（全球）：

1. 下载地址：
https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series
2. 以下为必要字段：
Province/State,Country/Region,Lat,Long
以及日期列表 “1/22/20,1/23/20...”
3. 坐标系统是 WGS_84。
4. 澳大利亚、加拿大、和中国的数据是省/州级别的，其它的数据是国家级别。
5. 全部是零的数据项将被跳过。

5.8.4.4 来自约翰霍普金斯大学的 COVID-19 每日数据（全球）。

1. 下载地址：
https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports
2. 字段随着时间在变化。。。
以下是"01-22-2020.csv"的格式：
Province/State,Country/Region,Last Update,Confirmed,Deaths,Recovered
以下是"05-15-2020.csv"的格式：
FIPS,Admin2,Province_State,Country_Region,Last_Update,Lat,Long_,Confirmed,Deaths,Recovered
,Active,Combined_Key
3. 坐标系统是 WGS_84。
4. 全部是零的数据项将被跳过。

5.8.4.5 导入的数据处理

1. 导入时可选是否统计数据集。
2. 导入数据的时间均被改为"23:59:00"。
3. 若数据包含的地理代码在数据库中还未定义，则自动添加到数据库中。

5.8.5 统计数据

1. 可选累加数据：
 - 省/州的数据累加为国家的数据
 - 国家的数据累加为洲的数据
 - 洲的数据累加为地球的数据。
2. 可选差值数据的位置级别。

5.8.6 定义条件

1. “数据源树”：数据集、及其不同数据源形成一棵“数据源树”，树结点可以多选。
2. “位置树”：MyBox 中所有的地理编码按从属关系形成一棵“位置树”，树结点可以多选。
3. “时间树”：MyBox 中所有的疫情报告涉及的时间按年/月/日形成一棵“时间树”，树结点可以多选。
4. “每日首部数据的个数”：
 - 不限制。不显示图，只按条件查询数据。
 - 有效值：
 - 按条件查询数据后、截取每日最前面的数据，以此显示数据和图。
 - “时间降序”自动被设置为最前面的排序条件。
 - 还必须再选择至少一个排序字段
 - 除了“时间降序”，最前面的排序字段被称为“查询的主属性”
 - “每日首部数据的个数”和排序条件只对查询和导出有效，对删除无效。
5. 编辑条件：标题、where、order by、fetch、“每日首部数据的个数”(-1 或 0 表示不限制)，它们被拼接成最终查询条件。

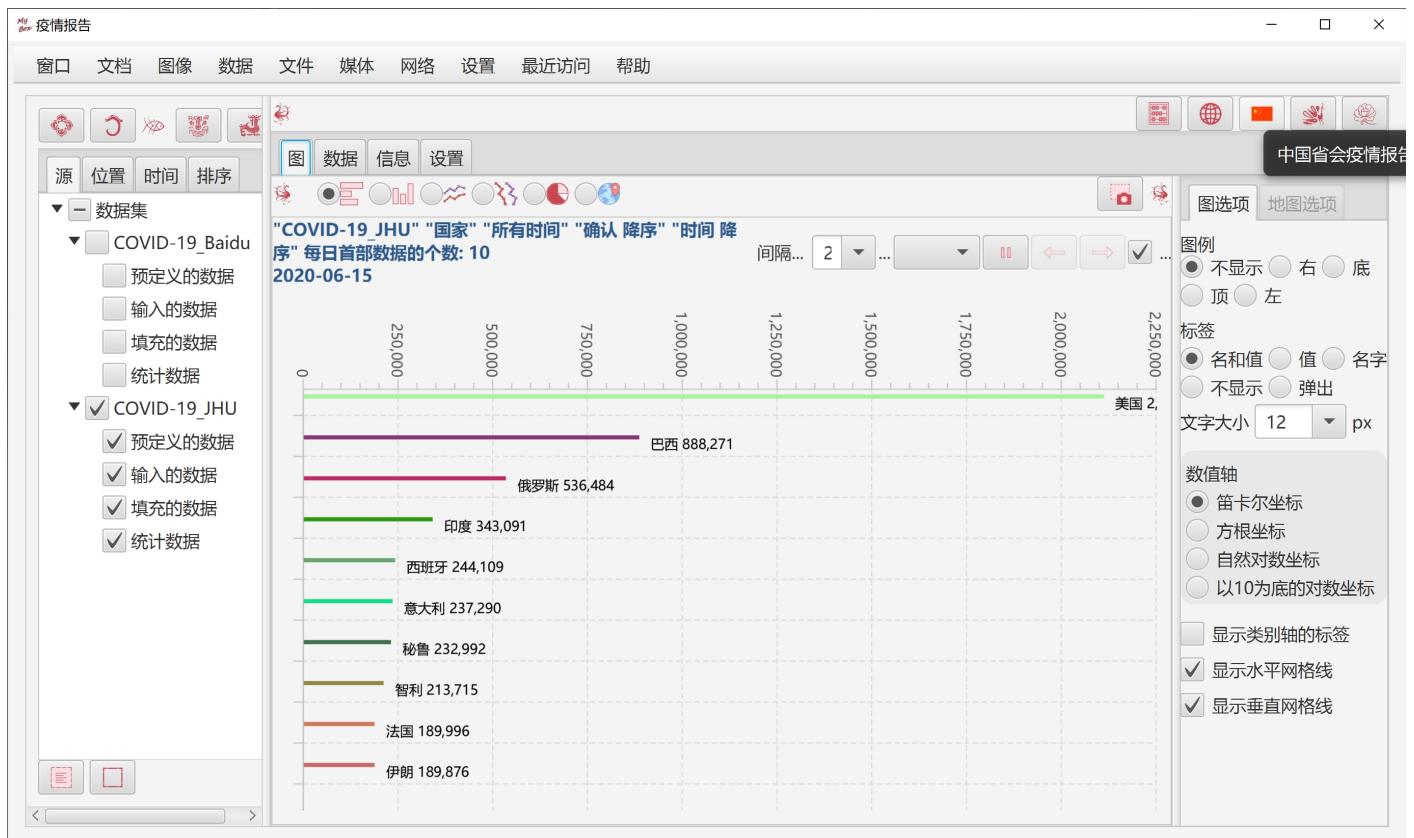
5.8.7 显示图

1. 当查询条件符合要求时，才会显示图：图数据总是“每日首部数据”，并有一个“查询的主属性”
2. 除了“查询的主属性”，可以选择更多数据属性，以在图中显示多维数据、或者同时显示多个数据的图。
3. 可选的图类型：水平条图、垂直条图、水平线图、垂直线图、饼图、地图。
4. 当数据的时间不唯一时，图是动态的：按时间升序逐帧显示每个时间的数据图。
5. 对于动态图，可以暂停/继续、设置间隔、停在指定时间的帧、上一帧、下一帧。
6. 常用选择，即时生效：
 - 图例位置：不显示、顶、底、左、右
 - 数值标签：名和值、值、名、不显示、弹出
 - 显示/不显示：类别轴的标签、水平网格线、垂直网格线
 - 数值轴：笛卡尔坐标、方根坐标、自然对数坐标、以 10 为底的对数坐标
 - 文字大小
 - 地图的参数：级别、图层、语言
7. 截图
 - 当前帧的截图。格式自选。
 - jpg：所有帧的截图
 - png：所有帧的截图
 - 动态 gif：所有帧的截图（可能内存不够）

5.8.8 设置

1. 截图的分辨率、动态截图的最大宽度、加载图数据的时间。
这些参数与内存消耗和计算机计算能力有关。
2. 按“数据源”定制数据行颜色。提供“缺省”和“随机”按钮。
3. 定制图中数据值的颜色。提供“缺省”和“随机”按钮。
4. 定制图中位置值的颜色。提供“随机”按钮。

MyBox 用户手册 - 数据工具 v6.6



6 其它

6.1 编码条码

1. 支持的一维码
 - 类型: Code39, Code128, Codabar, Interleaved2Of5, ITF_14, POSTNET, EAN13, EAN8, EAN_128, UPCA, UPCE, Royal_Mail_Customer_Barcode, USPS_Intelligent_Mail
 - 一维码选项: 朝向、宽高、分辨率、文字位置、字体大小、空白区宽度等。不同类型的选项不同。
2. 支持的二维码
 - 类型: QR_Code, PDF_417, DataMatrix
 - 二维码选项: 宽高、边沿、纠错级别、压缩模式。不同类型的选项不同。
 - 二维码 QR_Code 可以在中心显示一个图片。根据纠错级别自动调整图片大小。
3. 示例参数和建议值。
4. 对生成的条码即时检验。

6.2 解码条码

1. 支持的一维码类型: Code39, Code128, Interleaved2Of5, ITF_14, EAN13, EAN8, EAN_128, UPCA, UPCE
2. 支持的二维码类型: QR_Code, PDF_417, DataMatrix
3. 显示条码内容和元数据 (条码类型、纠错级别等)

6.3 消息摘要

1. 生成文件或者输入文本的消息摘要
2. 支持 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. 输出: Base64、十六进制、格式化的十六进制。

6.4 编码/解码 Base64

1. 把文件或者文本编码为 Base64。
2. 解码 Base64 文件或者 Base64 文本。
3. 对于文本可以设置字符集。
4. 输出为文件或者文本。

6.5 从 ttc 文件中提取 ttf 文件

<文档结束>