



**MyBox：简易工具箱
用户手册 - 数据工具**

作者：Mara

版本：6.6.1

日期：2022-11-16

内容目录

1 简介.....	8
1.1 主菜单.....	8
1.2 资源地址.....	9
1.3 文档.....	10
1.4 工具的菜单.....	11
2 二维存储结构的数据.....	12
2.1 数据对象.....	12
2.2 功能菜单.....	13
2.3 定义数据列.....	14
2.3.1 数据列的管理界面.....	14
2.3.2 列的类型.....	15
2.3.3 列的格式.....	16
2.3.3.1 数字的格式.....	16
2.3.3.2 日期时间/日期的格式.....	17
2.3.3.3 纪元的格式.....	18
2.3.3.4 定义枚举型.....	19
2.3.4 对于非法值的处理.....	19
2.4 列的应用.....	20
2.4.1 加载数据.....	20
2.4.2 显示数据.....	20
2.4.3 编辑控件	21
2.4.4 编辑数据单元.....	22
2.4.5 保存数据.....	24
2.4.6 计算数据.....	24
2.4.7 数据排序.....	25
2.4.8 列的颜色.....	25
2.5 编辑数据.....	26
2.5.1 表格编辑模式.....	26
2.5.2 CSV 编辑模式.....	27
2.5.3 保存与恢复.....	28
2.6 查看数据.....	29
2.6.1 数据的网页.....	29
2.6.2 数据的 CSV.....	30
2.7 定义数据属性.....	31
2.8 行表达式.....	32
2.8.1 行表达式的作用.....	32
2.8.2 编辑行表达式.....	32
2.8.3 计算行表达式.....	33
2.8.4 示例	33
2.9 行过滤.....	34
2.9.1 行过滤的作用.....	34
2.9.2 编辑行过滤.....	34
2.9.3 示例.....	35

2.10 行分组.....	36
2.10.1 分组前后.....	36
2.10.2 分组的作用.....	36
2.10.3 等值分组.....	37
2.10.3.1 分组方式.....	37
2.10.3.2 实现原理.....	37
2.10.3.3 分组的结果.....	38
2.10.4 值区间分组.....	39
2.10.4.1 分组方式.....	39
2.10.4.2 实现原理.....	40
2.10.4.3 分组的结果.....	40
2.10.5 条件分组.....	41
2.10.5.1 分组方式.....	41
2.10.5.2 实现原理.....	41
2.10.5.3 分组结果.....	42
2.10.6 行号分组.....	43
2.10.6.1 分组方式.....	43
2.10.6.2 实现原理.....	43
2.10.6.3 分组的结果.....	44
2.11 修改数据.....	45
2.11.1 添加行.....	46
2.11.2 赋值.....	47
2.11.3 删除.....	48
2.11.4 设置风格/标识异常值.....	49
2.11.4.1 管理风格.....	49
2.11.4.2 数据范围.....	50
2.11.4.3 定义风格.....	51
2.11.4.4 应用风格.....	52
2.11.5 粘贴系统粘贴板的内容.....	53
2.11.6 粘贴 MyBox 粘贴板的内容.....	54
2.12 整理数据.....	55
2.12.1 复制/过滤/查询/转换.....	56
2.12.2 排序.....	57
2.12.3 转置.....	58
2.12.4 归一化.....	59
2.12.5 分割/分组.....	60
2.13 数据计算.....	61
2.13.1 行表达式.....	62
2.13.2 描述性统计.....	63
2.13.3 分组统计.....	64
2.13.3.1 计算的选项.....	64
2.13.3.2 分组数据.....	65
2.13.3.3 统计数据.....	65
2.13.3.4 图数据.....	66
2.13.3.5 XY 图.....	67
2.13.3.6 饼状图.....	68

2.13.4 简单线性回归.....	69
2.13.4.1 实现的基础.....	69
2.13.4.2 计算的选项.....	69
2.13.4.3 回归.....	69
2.13.4.4 模型	70
2.13.4.5 拟合图	71
2.13.4.6 残差图	72
2.13.5 简单线性回归-组合.....	73
2.13.6 多重线性回归.....	74
2.13.6.1 回归.....	74
2.13.6.2 模型.....	75
2.13.7 多重线性回归-组合.....	76
2.13.8 频数分布.....	77
2.13.9 数值百分比.....	78
2.14 数据图.....	79
2.14.1 XY 图.....	80
2.14.1.1 数据.....	80
2.14.1.2 条图.....	81
2.14.1.3 堆叠条图.....	81
2.14.1.4 线图.....	82
2.14.1.5 散点图.....	82
2.14.1.6 气泡图.....	83
2.14.1.7 面积图.....	84
2.14.1.8 堆叠面积图.....	84
2.14.1.9 图数据的参数.....	85
2.14.1.10 布局.....	86
2.14.1.11 类别轴.....	87
2.14.1.12 数值轴.....	88
2.14.2 饼图.....	89
2.14.3 箱线图.....	90
2.14.3.1 计算方式.....	90
2.14.3.2 数据选项	91
2.14.3.3 统计数据.....	91
2.14.3.4 图选项.....	92
2.14.4 自比较条图.....	93
2.14.4.1 计算方式.....	93
2.14.4.2 数据选项.....	93
2.14.4.3 图选项	93
2.14.5 相比较条图.....	95
2.14.5.1 计算方式.....	95
2.14.5.2 数据选项	95
2.14.5.3 图选项	95
2.14.6 XYZ 图.....	97
2.14.6.1 三维散点图.....	98
2.14.6.2 曲面图.....	99
2.14.7 位置分布图.....	101

2.14.7.1 数据选项.....	101
2.14.7.2 地图数据.....	101
2.14.7.3 地图选项.....	102
2.14.7.4 播放选项.....	103
2.15 分组图.....	104
2.15.1 通用处理.....	105
2.15.1.1 数据分组.....	105
2.15.1.2 分组结果.....	106
2.15.1.3 图的显示选项.....	107
2.15.1.4 播放选项.....	107
2.15.2 分组数据 - XY 图.....	108
2.15.2.1 图的特定生成选项.....	108
2.15.2.2 图的特定显示选项.....	108
2.15.3 分组数据 - 饼图.....	109
2.15.3.1 图的特定生成选项.....	109
2.15.4 分组数据 - 箱线图.....	110
2.15.4.1 图的特定生成选项.....	110
2.15.4.2 图的特定显示选项.....	110
2.15.5 分组数据 - 自比较条图.....	111
2.15.5.1 图的特定生成选项.....	111
2.15.5.2 图的特定显示选项.....	111
2.15.6 分组数据 - 相比较条图	112
2.15.6.1 图的特定生成选项.....	112
2.15.6.2 图的特定显示选项.....	112
2.16 输入数据.....	113
2.16.1 加载系统粘贴板的内容	113
2.16.2 导入示例	114
2.16.2.1 中国的统计数据	114
2.16.2.2 回归相关的数据	115
2.16.2.3 位置数据.....	116
2.17 输出数据.....	117
2.17.1 导出.....	118
2.17.1.1 数据源.....	118
2.17.1.2 目标格式.....	119
2.17.1.3 目标文件.....	119
2.17.2 转换为数据库表.....	120
2.18 管理数据.....	121
2.19 拼接数据.....	122
2.20 数据文件.....	123
2.20.1 CSV 文件.....	123
2.20.2 文本文件.....	124
2.20.3 Excel 文件.....	125
2.20.4 批量转换/分割数据文件	126
2.20.5 合并数据文件.....	127
2.21 系统粘贴板中数据.....	128
2.22 MyBox 粘贴板中的数据.....	129

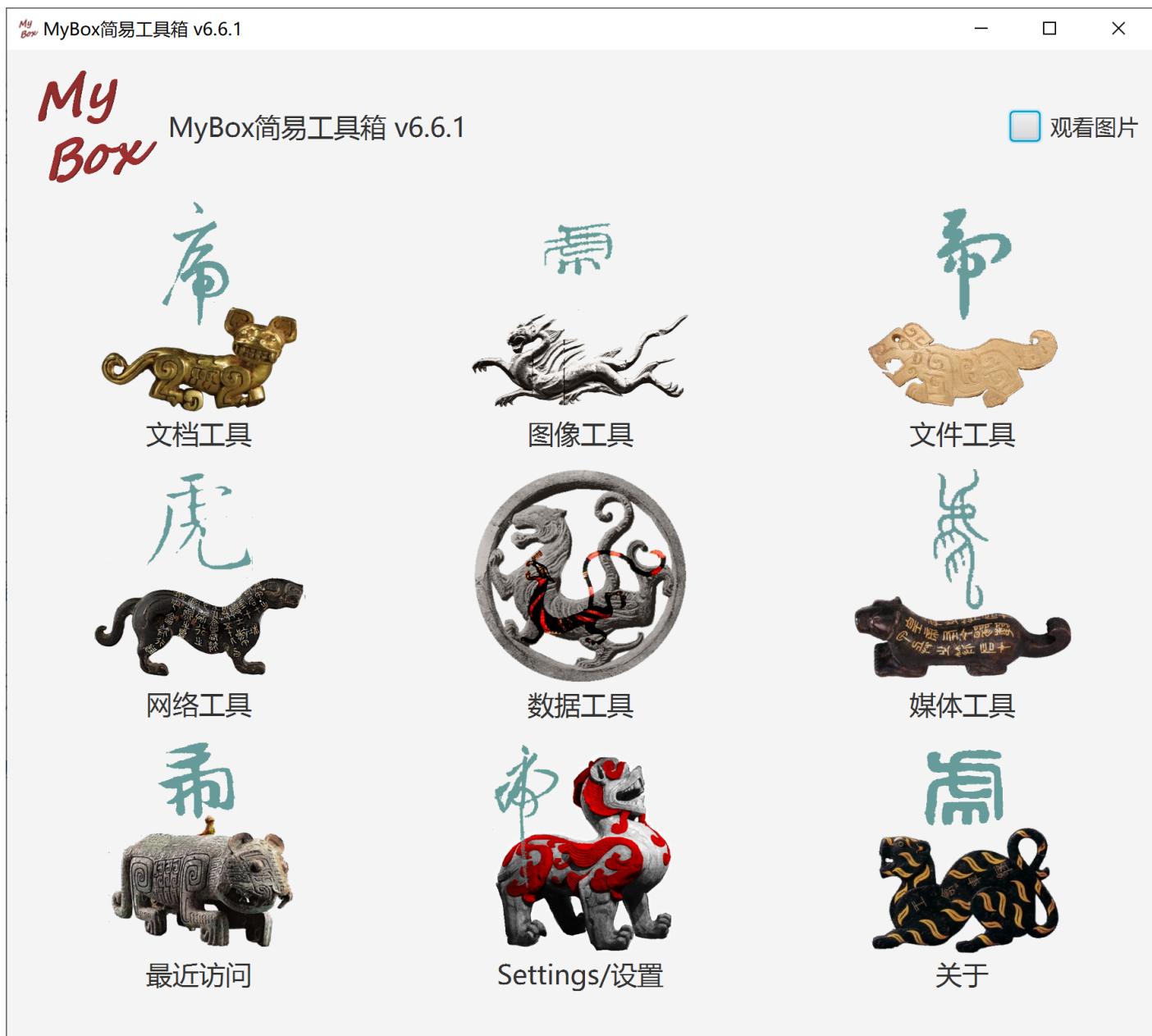
2.23 矩阵.....	130
2.23.1 管理和编辑矩阵.....	130
2.23.2 矩阵的一元计算.....	131
2.23.3 矩阵的二元计算.....	132
2.24 数据库表.....	133
2.24.1 管理数据库表.....	133
2.24.2 SQL 标识符的限制.....	134
2.24.3 数据库 SQL.....	135
2.24.4 SQL 查询.....	136
3 脚本与表达式.....	137
3.1 JShell (Java 交互编程工具)	137
3.1.1 关于 JShell.....	137
3.1.2 图形化运行 JShell.....	137
3.2 JEXL (Java 表达式语言)	139
3.2.1 关于 JEXL.....	139
3.2.2 图形化运行 JEXL.....	139
3.3 Javascript.....	141
4 数学函数.....	142
4.1 定义函数:	142
4.2 计算函数.....	142
4.3 数据集.....	144
4.4 一元函数的 XY 图.....	145
4.5 二元函数的 XYZ 图.....	146
5 位置数据.....	147
5.1 数据约束.....	147
5.1.1 无效值	147
5.1.2 坐标系统.....	147
5.1.3 坐标值.....	147
5.1.4 时间.....	147
5.1.4.1 时间格式	147
5.1.4.2 纪元.....	147
5.1.4.3 有效的时间示例.....	148
5.2 数据操作.....	148
5.3 地图数据.....	149
5.4 地理编码.....	151
5.4.1 数据定义	151
5.4.2 数据约束	151
5.4.3 编辑数据	151
5.4.4 定义条件	151
5.4.5 导入数据	151
5.4.5.1 内置的预定义数据.....	151
5.4.5.2 CSV 格式.....	151
5.4.5.3 来自 geoname.org 的位置数据:.....	152
5.4.6 设置	152
5.5 地图上的位置.....	153
5.6 位置工具.....	154

6 其它.....	155
6.1 编码条码	155
6.2 解码条码	155
6.3 消息摘要	155
6.4 编码/解码 Base64	155
6.5 从 ttc 文件中提取 ttf 文件	155

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

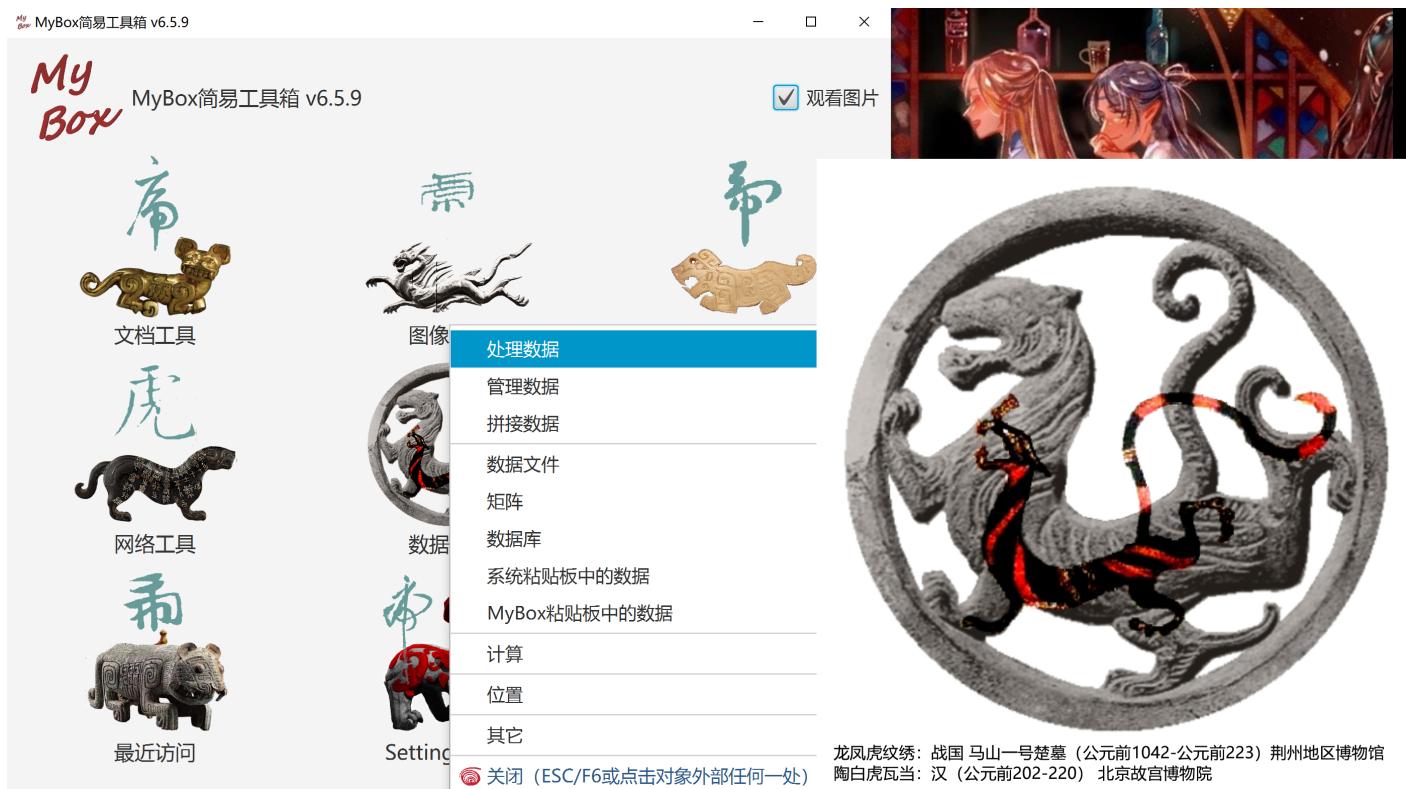
v6.5.8 Latest 18 days ago + 88 releases

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	6.6.1	2022-11-16	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	6.6.1	2022-11-16	html PDF odt	html PDF odt
用户手册-数据工具	6.6.1	2022-11-16	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-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
示例-笔记	6.6.1	2022-11-16	html	html
示例-树形	6.6.1	2022-11-16	html	html
示例-收藏的网址	6.6.1	2022-11-16	html	html
示例-SQL	6.6.1	2022-11-16	html	html
示例-JShell	6.6.1	2022-11-16	html	html
示例-JEXL	6.6.1	2022-11-16	html	html
示例-JavaScript	6.6.1	2022-11-16	html	html
示例-数学函数	6.6.1	2022-11-16	html	html
示例-行过滤	6.6.1	2022-11-16	html	html
关于-颜色	6.6.1	2022-11-16	html	html
关于-坐标系统	6.6.1	2022-11-16	html	html
关于-媒体	6.6.1	2022-11-16	html	html
关于-数据分析	6.6.1	2022-11-16	html	html
调色盘-常用网页颜色	6.6.1	2022-11-16	html	html
调色盘-中国传统颜色	6.6.1	2022-11-16	html	html
调色盘-日本传统颜色	6.6.1	2022-11-16	html	html
调色盘-来自 colorhexa.com 颜色	6.6.1	2022-11-16	html	html
调色盘-MyBox 的颜色	6.6.1	2022-11-16	html	html

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

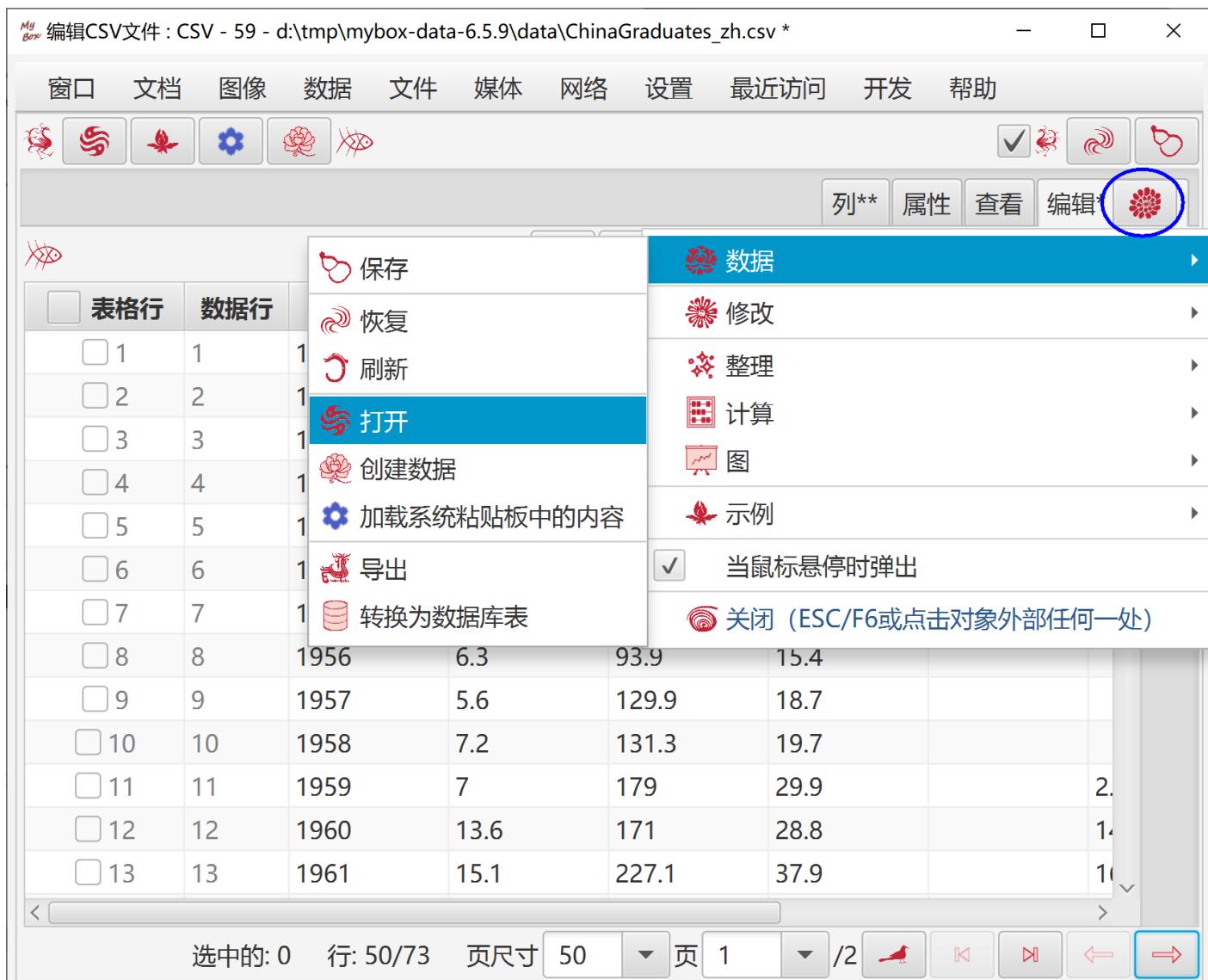
2. 数据被表示为二维存储结构：

- “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 定义数据列

2.3.1 数据列的管理界面

在数据处理的“列”页签下，在表格中增/删/改数据的列：

1. 列名不能为空也不能重复。
2. 直接点击表格单元以编辑。
3. 也可以选择列并点击“编辑”按钮以弹出列的编辑窗。
4. 可以重命名所有字段为序列号。
5. 可以随机设置颜色。
6. 可以调整列的顺序。
7. 点击按钮“确定”以把列的修改应用于当前数据。
8. 点击按钮“取消”以丢弃修改并从当前数据读取列的定义。
9. 修改未保存时，若修改未应用于数据，则页签头显示**，否则页签头显示*。

表格行	索引	列名	类型	色彩	宽度	可编辑	非空	操作
<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

- 16777215
- 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
- LCHuv: 81.29 65.24 107.3

2.3.2 列的类型

1. 列的类型包括：字串、双精度、浮点、长整型、整型、短整型、布尔型、日期时间、纪元、经度、纬度、枚举型、色彩。
2. 此属性用于数据的显示、编辑、计算、和保存。
3. 通常应成对定义经度和纬度。

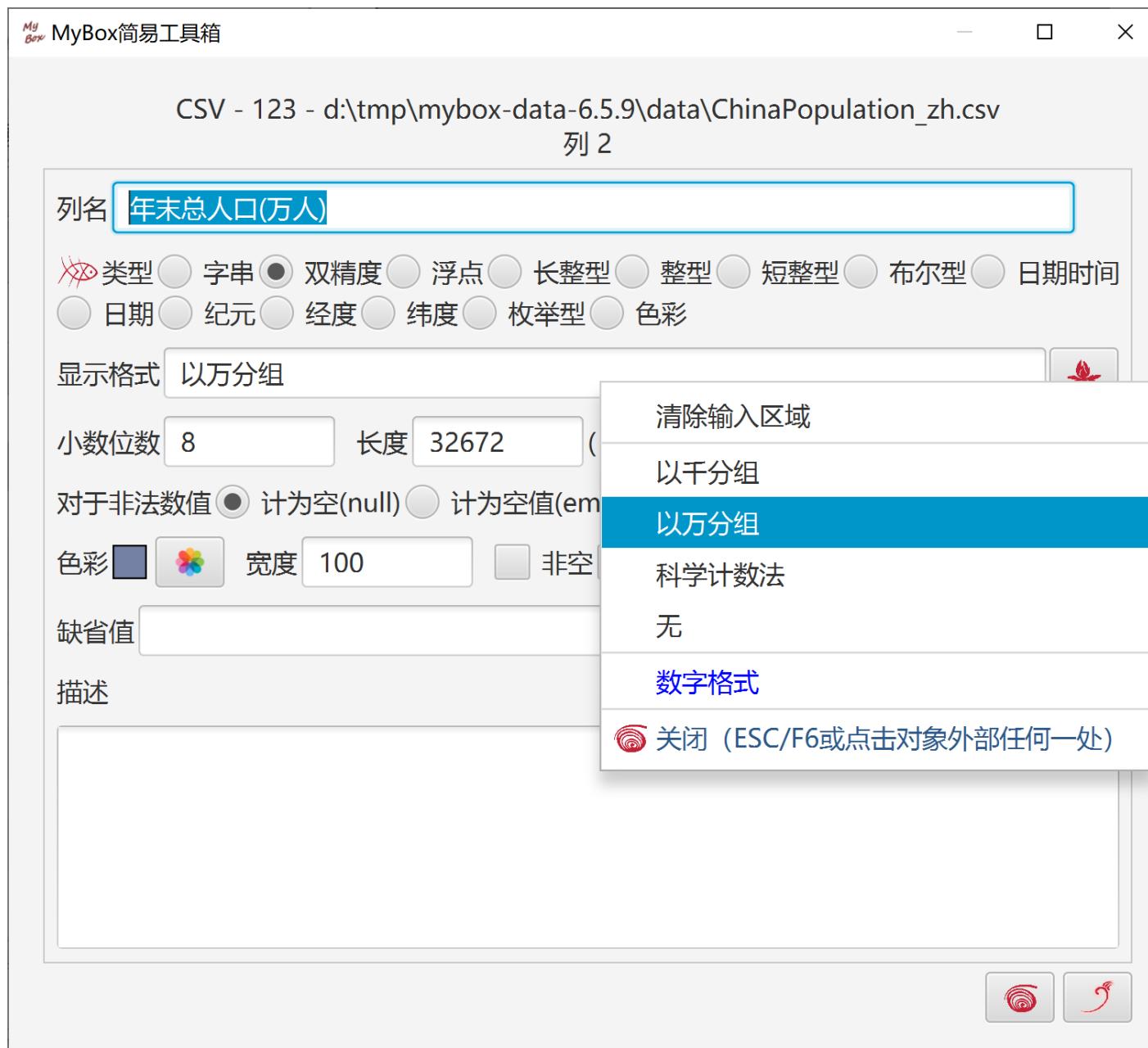


2.3.3 列的格式

- 此属性主要用于显示，在编辑输入数值时不会自动应用格式，而是保持原始的输入。
- 在一些界面上，如“复制”或“导出”，可选“按照列的格式保存日期/时间/纪元和数字”。

2.3.3.1 数字的格式

对于数字类型，可选：以千分组、以万分组、科学计数法、无格式。



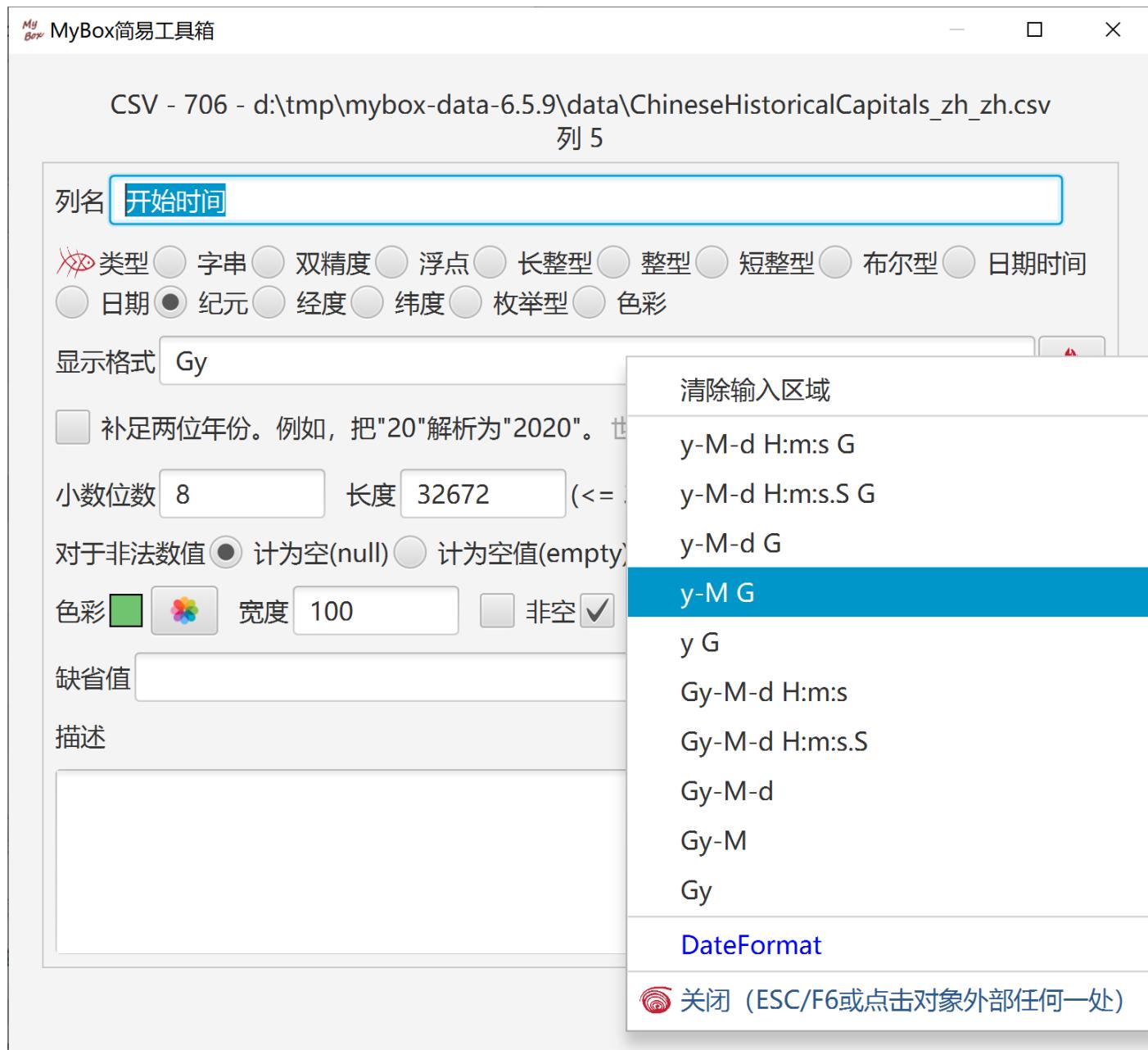
2.3.3.2 日期时间/日期的格式

对于日期时间/日期类型，支持：MM/dd/yy、yy/MM/dd、毫秒、时区、T 分隔、补全世纪。



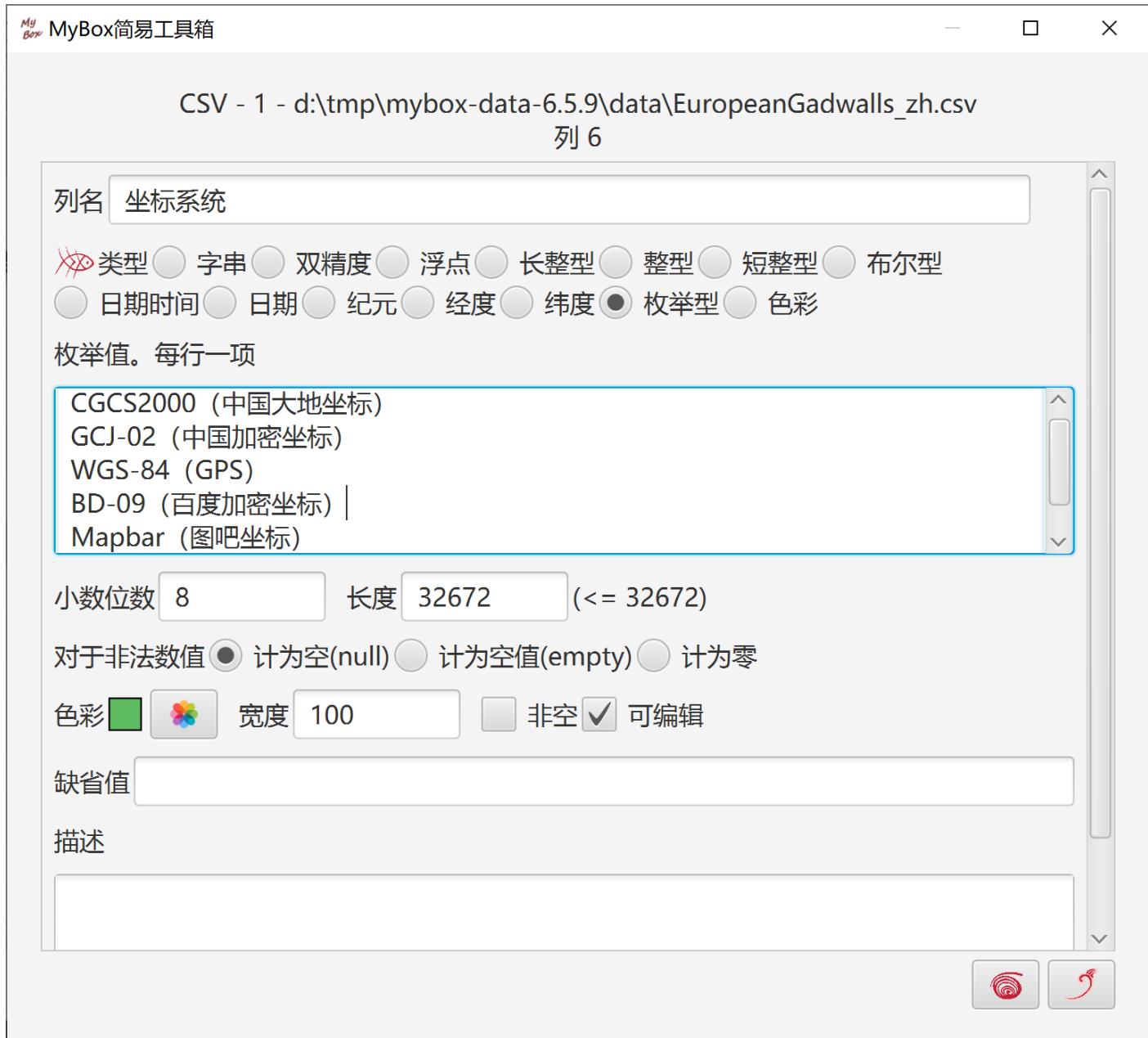
2.3.3.3 纪元的格式

对于纪元类型，支持：M/d/y、y/M/d、毫秒、时区、T分隔、补全世界、中文前后缀（“公元”、“公元前”）、英文前后缀（“AD”、“BC”）。



2.3.3.4 定义枚举型

对于枚举型，可定义数值列表。



2.3.4 对于非法值的处理

1. 数据列对于非法值的处理策略：略过、计为空值、计为零。
2. 在一些上下文环境中，“计为空值”等同于“略过”。
3. 此属性只用于显示或者计算，在编辑输入数值时不会自动处理非法值。

2.4 列的应用

列的应用原则为：“最大包容、最少处理”。

2.4.1 加载数据

加载数据时，不检查列类型，原样读取和导入。

2.4.2 显示数据

1. 按列类型解析数值
2. 按列的定义来处理非法值
3. 按列的格式来重写数值
4. 被显示的数值，可能与当前实际数值并不相同。
5. 例如，列的类型为双精度，读入数值“abc”：
 - 若非法值计为零，则显示为零（实际值仍为“abc”）
 - 若非法值略过或计为空，则显示为“abc”。

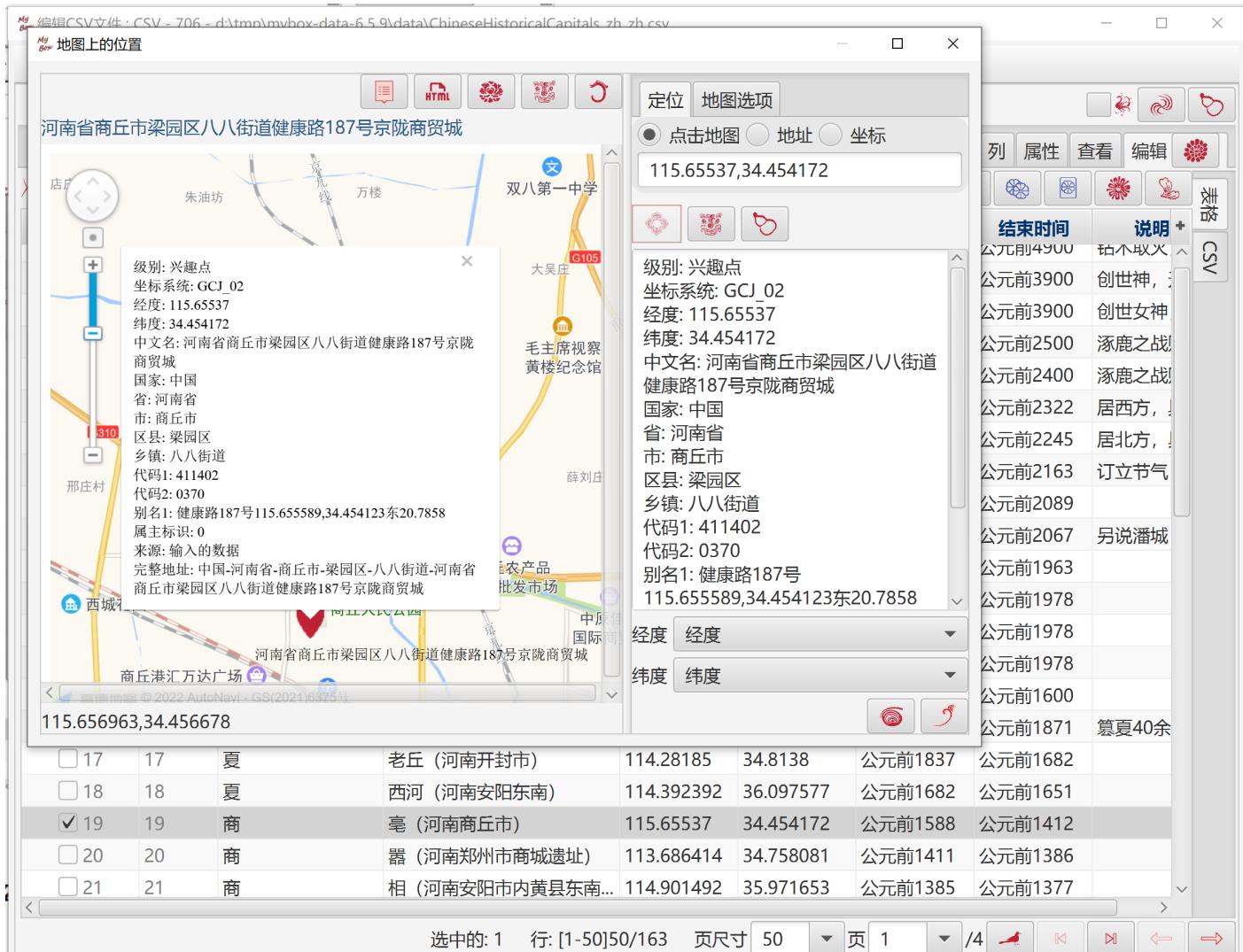
表格行	数据行	年	年末总人口...	男性人口(...	女性人口(...	城镇人口(...	乡村人
<input type="checkbox"/>	1	0	1949	54,167	2,8145	26022	5765
<input type="checkbox"/>	2	1	1950	55,196	2,8669	26527	6169
<input type="checkbox"/>	3	2	1951abc	56,300	2,9231	27069	6632
<input type="checkbox"/>	4	3	1952	57,482	2,9833	27649	7163
<input checked="" type="checkbox"/>	5	4	1953	58796	3,0468	28328	7826
<input type="checkbox"/>	6	5	abc1954	60,266	3,1242	29024	8249
<input type="checkbox"/>	7	6	1955	61,465	3,1809	29656	8285
<input type="checkbox"/>	8	7	1956	62,828	3,2536	30292	9185
<input type="checkbox"/>	9	8	1957	64,653	3,3469	31184	9949
<input type="checkbox"/>	10	9	1958	65,994	3,4195	31799	10721
<input type="checkbox"/>	11	10	1959	67,207	3,4890	32317	12371
<input type="checkbox"/>	12	11	1960	66,207	3,4283	31924	13073
<input type="checkbox"/>	13	12	1961	65,859	3,3880	31979	12707

2.4.3 编辑控件

1. 对于布尔类型，显示选择框。
2. 对于枚举类型，显示下拉选择列表。
3. 对于颜色类型，显示调色盘。
4. 对于经度/纬度，可在地图上选择位置。

The screenshot shows the MyBox Data Tool interface with a CSV file named "HeartFailure_zh.csv" open. The window title bar reads "MyBox 编辑CSV文件 : CSV - 778 - d:\tmp\mybox-data-6.5.9\data\HeartFailure_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 various icons for file operations like Open, Save, Print, and Import/Export. The main area displays a table with columns: 表格行 (Table Row), 数据行 (Data Row), 年龄 (Age), 贫血 (Anemia), 肌酐磷酸激... (Creatinine Phosphokinase), 糖尿病 (Diabetes), 喷血分数 (Hemorrhage Score), and a checkbox column. Row 8 is selected, and a dropdown menu is open over the "年龄" column, listing age ranges: 1-20, 21-40, 41-60, 61-80, and 81-100. The "41-60" option is highlighted. The bottom of the interface shows pagination controls: "选中的: 1 行: [1-50] 50/299" and "页尺寸: 50 页 1 /6".

表格行	数据行	年龄	贫血	肌酐磷酸激...	糖尿病	喷血分数
<input type="checkbox"/> 1	1	61-80	<input type="checkbox"/>	582	<input type="checkbox"/>	20
<input type="checkbox"/> 2	2	61-80	<input type="checkbox"/>	7861	<input type="checkbox"/>	38
<input type="checkbox"/> 3	3	61-80	<input type="checkbox"/>	146	<input type="checkbox"/>	20
<input type="checkbox"/> 4	4	41-60	<input checked="" type="checkbox"/>	111	<input type="checkbox"/>	20
<input type="checkbox"/> 5	5	61-80	<input checked="" type="checkbox"/>	160	<input checked="" type="checkbox"/>	20
<input type="checkbox"/> 6	6	81-100	<input checked="" type="checkbox"/>	47	<input type="checkbox"/>	40
<input type="checkbox"/> 7	7	61-80	<input checked="" type="checkbox"/>	246	<input type="checkbox"/>	15
<input checked="" type="checkbox"/> 8	8	41-...	<input checked="" type="checkbox"/>	315	<input checked="" type="checkbox"/>	60
<input type="checkbox"/> 9	9	1-20	<input type="checkbox"/>	157	<input type="checkbox"/>	65
<input type="checkbox"/> 10	10	21-40	<input checked="" type="checkbox"/>	123	<input type="checkbox"/>	35
<input type="checkbox"/> 11	11	41-60	<input checked="" type="checkbox"/>	81	<input type="checkbox"/>	38
<input type="checkbox"/> 12	12	61-80	<input type="checkbox"/>	231	<input type="checkbox"/>	25
<input type="checkbox"/> 13	13	81-100	<input checked="" type="checkbox"/>	981	<input type="checkbox"/>	30



2.4.4 编辑数据单元

- 在用户点击数据单元开始编辑时，数值原样显示，忽略列类型和列格式。
- 在用户输入和修改数值过程中，按列类型检查输入框中的数值。
 - 若数值非法，则编辑框显示为异常颜色。
 - 若数值合法，则编辑框显示为正常颜色。
 - 数值始终保持用户的输入。
- 当焦点离开数据单元时，检查编辑框中的数值：
 - 若数值未变化，则不检查类型也不提交保存。
 - 若数值有变化，则按列类型检查修改后的数值：
 - 对非法值，拒绝保存，恢复数值、原样显示。
 - 对合法值，提交保存，然后按列类型和列格式显示保存后的数值。

4. 例如，列的类型为双精度，非法值计为空，小数位数为2。读入数值“abc”：
- 数据单元显示为“abc”。
 - 用户修改它的数值为“abc123”：
 - 在用户输入过程中，输入框始终显示为异常颜色。
 - 用户敲击回车或点击其它控件时，此数据单元恢复显示为“abc”并为正常颜色。
 - 用户修改数值为“123.4567”：
 - 在用户输入过程中，输入框始终显示为正常颜色。
 - 用户敲击回车或点击其它控件时，此数据单元被保存为“123.4567”、并显示为“123.46”。

5. 其它数据单元不受影响。即只要不修改数据单元，它们就一直保持原样。

表格行	数据行	年龄	贫血	肌酐磷酸激...	糖尿病	喷血分数
<input type="checkbox"/>	1	75	<input type="checkbox"/>	582	<input type="checkbox"/>	20
<input type="checkbox"/>	2	55	<input type="checkbox"/>	7861	<input type="checkbox"/>	38
<input type="checkbox"/>	3	65	<input type="checkbox"/>	146	<input type="checkbox"/>	20
<input type="checkbox"/>	4	50	<input checked="" type="checkbox"/>	111	<input type="checkbox"/>	20
<input type="checkbox"/>	5	65	<input checked="" type="checkbox"/>	160	<input checked="" type="checkbox"/>	20
<input type="checkbox"/>	6	90	<input checked="" type="checkbox"/>	47	<input type="checkbox"/>	40
<input checked="" type="checkbox"/>	7	75	<input checked="" type="checkbox"/>	246abc	<input type="checkbox"/>	15
<input type="checkbox"/>	8	60	<input checked="" type="checkbox"/>	315	<input checked="" type="checkbox"/>	60
<input type="checkbox"/>	9	65	<input type="checkbox"/>	157	<input type="checkbox"/>	65
<input type="checkbox"/>	10	80	<input checked="" type="checkbox"/>	123	<input type="checkbox"/>	35
<input type="checkbox"/>	11	75	<input checked="" type="checkbox"/>	81	<input type="checkbox"/>	38
<input type="checkbox"/>	12	62	<input type="checkbox"/>	231	<input type="checkbox"/>	25
<input type="checkbox"/>	13	45	<input checked="" type="checkbox"/>	981	<input type="checkbox"/>	30
<input type="checkbox"/>	14	50	<input checked="" type="checkbox"/>	168	<input type="checkbox"/>	22

2.4.5 保存数据

1. CSV/文本/Excel 文件的数值都按字符串写入。
2. 数据库表的数值按“就近类型”写入：

MyBox 的列类型	JDBC 的数据类型
字串(String)	VARCHAR
双精度(Double)	DOUBLE
浮点(Float)	FLOAT
长整型(Long)	BIGINT
整型(Integer)	INT
短整型(Short)	SMALLINT
布尔型(Boolean)	BOOLEAN
日期时间(Datetime)	TIMESTAMP
日期(Date)	DATE
纪元(Era)	BIGINT
经度(Longitude)	DOUBLE
纬度(Latitude)	DOUBLE
枚举型(Enumeration)	VARCHAR
颜色(Color)	VARCHAR

注： derby 不支持负数的时间类型，所以只能用长整型保存纪元。

3. 矩阵的所有数值都按双精度写入。

2.4.6 计算数据

1. 忽略列类型和列格式，用数据原值来处理，
2. 按计算的要求解析数值。例如计算需要双精度数字，则试图将数值转换为双精度。
3. 若数值转换失败，则按列的定义处理非法值。
4. 计算本身也可以定义对非法值的处理策略。计算的定义优先于列的定义。
5. 列和计算都可以定义小数位数。计算的定义优先于列的定义。
6. 例如，列类型为字串，对它进行描述性统计：
 - 每个数值被试图转换为双精度
 - 若非法值计为零，则非法值也被算在“个数”里，同时参与平均值和总和的计算。
 - 若略过非法值，则非法值不参与任何计算。
 - 若非法值计为空，则非法值导致所有计算的结果都非法（Double.NaN）。

2.4.7 数据排序

- 所有涉及到排序的计算，数据都会转换为临时数据库表，然后利用数据库系统来排序。
- 排序结果与列的类型有关。例如，字串“123”小于字串“18”，而数字“123”大于数字“18”。

2.4.8 列的颜色

列的颜色主要用于数据图。

当数据图初始生成时，与列相关的图元素显示为列的颜色。其后用户可以设置为随机颜色。

2.5 编辑数据

2.5.1 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. 除了文本文件，对于字符串类型的数据，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。

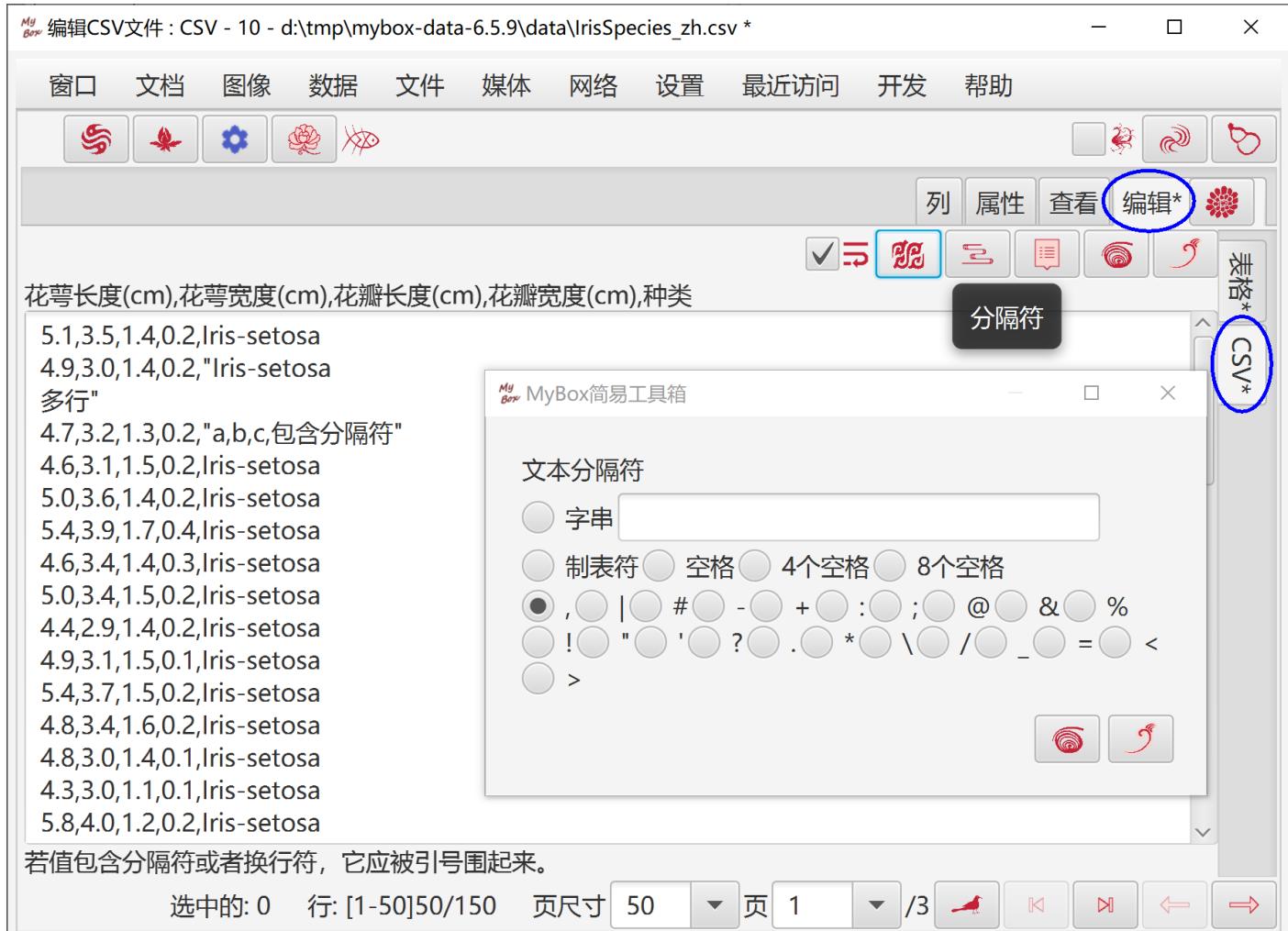
表格行	数据行	花萼长度(cm)	花萼宽度(cm)	花瓣长度(cm)	花瓣宽度(cm)	种类
<input checked="" type="checkbox"/>	1	5.1	3.5	1.4	0.2	Iris-setosa 多行
<input type="checkbox"/>	2	4.9	3	1.4	0.2	Iris-setosa
<input type="checkbox"/>	3	4.7				is-setosa
<input type="checkbox"/>	4	4.6				is-setosa
<input type="checkbox"/>	5	5				is-setosa
<input type="checkbox"/>	6	5.4				is-setosa
<input type="checkbox"/>	7	4.6				is-setosa
<input type="checkbox"/>	8	5				is-setosa
<input type="checkbox"/>	9	4.4				is-setosa
<input type="checkbox"/>	10	4.9				is-setosa

选中的: 1

2.5.2 CSV 编辑模式

CSV 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，当前修改被丢弃。
4. 若数值包含分隔符或换行符，它应被引号包围起来。



2.5.3 保存与恢复

1. 修改未保存时，页签头显示*。
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...
1	1	1949	2.1	28	6.1	
2	2	1950	1.8	29.6	6.2	
3	3	1951	1.9	29.1	5.9	
4	4					
5	5					
6	6					
7	7					
8	8					
9	9					
10	10	1958	7.2	131.3	19.7	
11	11	1959	7	179	29.9	2.
12	12	1960	13.6	171	28.8	1.
13	13	1961	15.1	227.1	37.9	10

选中的: 0 行: 50/73 页尺寸 50 / 2 />

2.6 查看数据

2.6.1 数据的网页

1. 显示当前页的数据。
2. 可选显示：表窗、标题、列名、行号。

编辑CSV文件 : CSV - 10 - d:\tmp\mybox-data-6.5.9\data\IrisSpecies_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

表窗 标题 列名 行号

属性 查看 编辑* 

d:\tmp\mybox-data-6.5.9\data\IrisSpecies_zh.csv

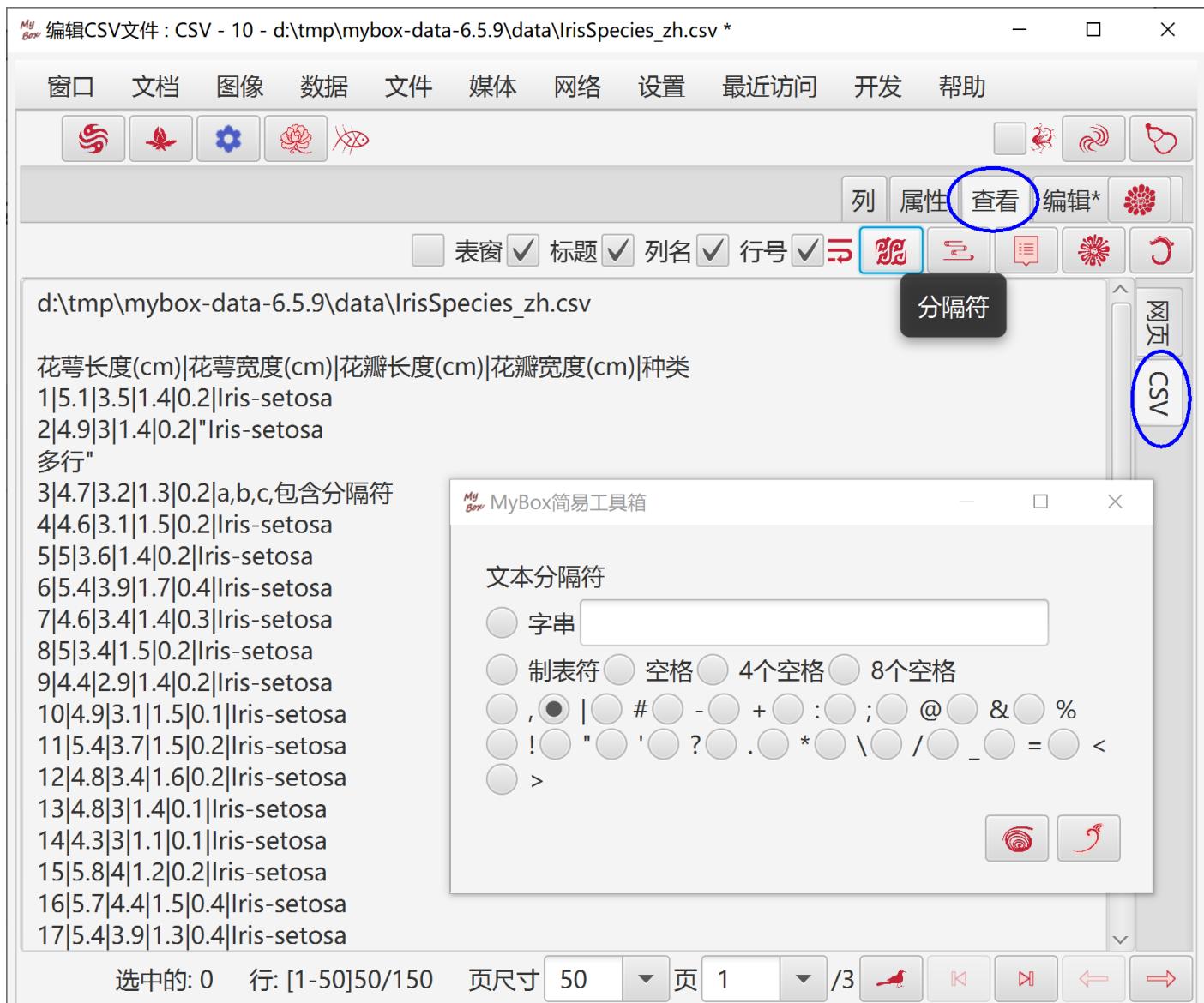
行号	花萼长度(cm)	花萼宽度(cm)	花瓣长度(cm)	花瓣宽度(cm)	种类
行1	5.1	3.5	1.4	0.2	Iris-setosa
行2	4.9	3	1.4	0.2	Iris-setosa 多行
行3	4.7	3.2	1.3	0.2	a,b,c,包含分隔符
行4	4.6	3.1	1.5	0.2	Iris-setosa
行5	5	3.6	1.4	0.2	Iris-setosa
行6	5.4	3.9	1.7	0.4	Iris-setosa
行7	4.6	3.4	1.4	0.3	Iris-setosa
行8	5	3.4	1.5	0.2	Iris-setosa

H2

选中的: 0 行: [1-50]50/150 页尺寸 50 ▾ 页 1 ▾ /3    

2.6.2 数据的 CSV

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。



2.7 定义数据属性

包括：数据名、小数位数、随机数的最大值。

CSV - 5 - d:\tmp\mybox-data-6.5.9\data\SouthGermanCredit_zh.csv

编号 5

数据类型 CSV

数据名 德国南部的信用卡数据

小数位数 2

随机数的最大值 1000

描述

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.

更新时间 2022-09-28 09:54:05

文件: d:\tmp\mybox-data-6.5.9\data\SouthGermanCredit_zh.csv
文件大小: 46.936 KB
文件修改时间: 2022-09-28 09:54:05
字符集: UTF-8
分隔符: ,
第一行定义列名: 是
文件的行数: 1000
列数: 21
当前页: 1 / 20
页的行范围: 1 - 50 (50)
页修改时间: 2022-09-28 09:54:07

选中的: 0 行: 50/1000 页尺寸 50 页 1 /20

2.8 行表达式

2.8.1 行表达式的作用

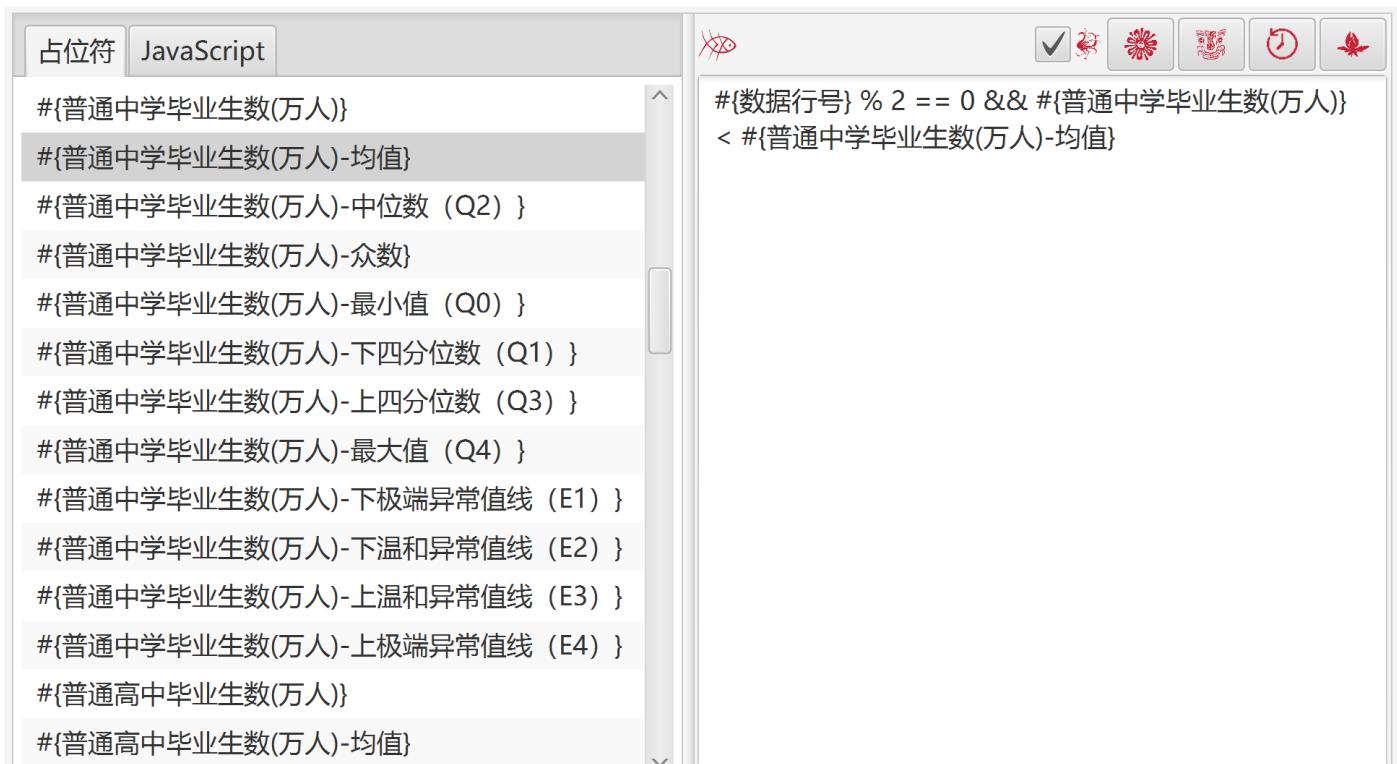
在加工/整理/计算数据、生成数据图时，可以输入 JavaScript 表达式作为数据值。

2.8.2 编辑行表达式

1. 若脚本为空，则返回空字符串。
2. 脚本可以包含 Nashorn 可以解析的任何合法元素(ECMAScript 5.1)。
3. 脚本应当最终是一个数值。
4. 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>}
```

5. 左边显示所有合法的占位符。
6. 点击“编辑”按钮以保存脚本。
7. 悬停或点击按钮“示例”，可以粘贴示例代码。
8. 悬停或点击按钮“历史”，可以粘贴历史代码。



2.8.3 计算行表达式

在 MyBox 计算表达式时：

1. 占位符被数据行的实际值替换。
2. 统计值是用所有数据来计算的。
3. '#{xxx}' 被处理为字符串而 #{xxx} 被处理为数字。
4. 当处理所有数据时，若脚本包含 “#{表行号}” 则它会失败。

2.8.4 示例

表达式	含义
#{数据行号}	整个数据的行号
#{表行号}	当前页的行号
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

2.9 行过滤

2.9.1 行过滤的作用

“行过滤”是特殊的“行表达式”，用来作为过滤数据行的条件。

2.9.2 编辑行过滤

1. 脚本应当最终是一个布尔值("true"或"false")。
2. 可以设置取用数据的最多行数。
3. 可以保存为树形。



2.9.3 示例

表达式	含义
<code>#{数据行号} % 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}' == ''</code>	列"字段 1"的值为 null 或 empty
<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

2.10 行分组

2.10.1 分组前后

1. 在分组之前，可以先过滤数据行。即分组的对象是过滤后的数据行。
2. 为了利用数据库系统实现排序和统计，数据在分组前都被转换为临时数据库表。
3. 对于分组后的数据，可以设置排序条件和每组最多取值个数。

2.10.2 分组的作用

把数据行分组以后，可以：

1. 对源数据添加分组信息，即新增两列：组编号、组参数。
2. 按组将数据分割为多个文件。
3. 按组进行描述性统计。
4. 按组依序显示为动态的 XY 图/饼图/相比较图/自比较图/箱线图。

2.10.3 等值分组

2.10.3.1 分组方式

选择若干列，把这些列的值都相同的数据行分为同一组。

	列
<input type="checkbox"/> 1	年龄
<input checked="" type="checkbox"/> 2	性别
<input type="checkbox"/> 3	BMI(体质指数)
<input type="checkbox"/> 4	BP(平均血压)
<input type="checkbox"/> 5	S1(血清指标1)
<input type="checkbox"/> 6	S2(血清指标2)
<input type="checkbox"/> 7	S3(血清指标3)
<input type="checkbox"/> 8	S4(血清指标4)
<input type="checkbox"/> 9	S5(血清指标5)

对于过滤后的数据行，如下生成组：

- 选择若干列。
- 这些列的值都相同则分为同一组。

2.10.3.2 实现原理

1. 把数据转换为临时数据库表
2. 执行 SQL 语句 “group by”。

2.10.3.3 分组的结果

- 完全：每个数据行都属于一组。
- 不相交：每个数据行都只属于一组。

编辑CSV文件 : CSV - 782 - d:\tmp\mybox-data-6.5.9\generated\糖尿病的病情预测_等值_年龄_性别_2022-11-10_...

The screenshot shows the MyBox Data Tool window with a toolbar at the top containing various icons for file operations, data manipulation, and analysis. Below the toolbar is 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). The main area displays a table with 14 rows of data. The columns are labeled: 表格行 (Table Row), 数据行 (Data Row), 组 (Group), 等值_年龄_性别 (Equivalence_Age_Gender), 年龄 (Age), 性别 (Gender), BMI(体质...) (BMI), and BP(平均血压) (Average Blood Pressure). The '组' column shows group numbers 1 through 6. The '等值_年龄_性别' column contains equivalence classes defined by age and gender. The data shows a range of ages from 19 to 23 and genders 1 and 2. The bottom of the window includes a status bar with '选中的: 0' (Selected: 0), '行: [1-50] 50/442' (Row: [1-50] 50/442), '页尺寸' (Page Size) set to 50, '页' (Page) set to 1, '1/9' (1/9), and navigation icons.

表格行	数据行	组	等值_年龄_性别	年龄	性别	BMI(体质...)	BP(平均血压)
1	1	1	{年龄=19.0, 性别=1}	19	1	23.2	75
2	2	1	{年龄=19.0, 性别=1}	19	1	25.3	83
3	3	1	{年龄=19.0, 性别=1}	19	1	19.2	87
4	4	2	{年龄=20.0, 性别=1}	20	1	22.9	87
5	5	3	{年龄=20.0, 性别=2}	20	2	24.2	88
6	6	3	{年龄=20.0, 性别=2}	20	2	22.1	87
7	7	4	{年龄=21.0, 性别=1}	21	1	24.2	76
8	8	4	{年龄=21.0, 性别=1}	21	1	20.1	63
9	9	5	{年龄=22.0, 性别=1}	22	1	19.9	75
10	10	5	{年龄=22.0, 性别=1}	22	1	19.3	82
11	11	5	{年龄=22.0, 性别=1}	22	1	23	73
12	12	5	{年龄=22.0, 性别=1}	22	1	18.6	97
13	13	6	{年龄=23.0, 性别=1}	23	1	18.8	78
14	14	6	{年龄=23.0, 性别=1}	23	1	29	90

2.10.4 值区间分组

2.10.4.1 分组方式

选择一列，按照它的值区间分割数据行。

分割条件有三种选择：

- 分割的尺寸（值间隔）
- 分割的个数
- "开始值-结束值"的列表



对于过滤后的数据行，如下生成组：

- 选择一列，按照它的值区间分割数据行。
- 值区间可以按以下方式定义：
 - 分割的尺寸（值间隔）
 - 分割的个数
 - "开始值-结束值"的列表
- 忽略所选列的类型，将它的数据解析为双精度数值。

2.10.4.2 实现原理

1. 把数据转换为临时数据库表。
2. 忽略所选列的类型，将它的数据解析为双精度数值。
3. 计算该列的最大值和最小值。
4. 计算各组的区间
5. 构造各组的查询语句，条件类似于“列 \geq 区间开始值 AND 列 $<$ 区间结束值”。

2.10.4.3 分组的结果

- 若按分割尺寸或分割个数，则：
 - 完全：每个数据行都属于一组。
 - 不相交：每个数据行都只属于一组。
- 若按"开始值-结束值"的列表，则：
 - 可能不完全：可能有的数据行不属于任何一组。
 - 可能相交：可能有的数据行属于多个组。

编辑CSV文件 : CSV - 784 - d:\tmp\mybox-data-6.5.9\generated\糖尿病的病情预测_区间_年龄.csv

表格行	数据行	组	区间_年龄	年龄	性别	BMI(体质...)	BP(平均血压)	S1(血清指)
21	21	1	[19,27)	26	1	18.8	83	191
22	22	1	[19,27)	23	2	18	78	171
23	23	1	[19,27)	19	1	25.3	83	225
24	24	1	[19,27)	25	2	22.6	85	130
25	25	1	[19,27)	19	1	23.2	75	143
26	26	1	[19,27)	25	2	23.5	88	143
27	27	2	[27,35)	29	1	30	85	180
28	28	2	[27,35)	34	2	24.7	118	254
29	29	2	[27,35)	31	1	29.7	88	167
30	30	2	[27,35)	30	2	25.2	83	178
31	31	2	[27,35)	32	1	30.5	89	182
32	32	2	[27,35)	32	2	25.4	90.33	153
33	33	2	[27,35)	33	1	25.3	85	155
34	34	2	[27,35)	27	1	19.6	78	128

选中的: 0 行: [1-50]50/442 页尺寸 50 / 1 / 9 < > << >>

2.10.5 条件分组

2.10.5.1 分组方式

定义若干行过滤器，按照这些行过滤器将数据行分割成组。

表达式	反向	最多行数
#{年龄} < 20	false	-1
#{年龄} < 45	false	-1
#{年龄} < 60	false	-1
#{年龄} < 70	false	-1
#{年龄} < 85	false	-1
#{年龄} > 85	false	-1

对于过滤后的数据行，如下生成组：

- 定义若干行过滤器。
- 按照这些行过滤器将数据行分割成组。

2.10.5.2 实现原理

1. 把数据转换为临时数据库表。
2. 对于每个行过滤器，遍历数据行以挑出满足条件的组成员。

2.10.5.3 分组结果

- 可能不完全：可能有的数据行不属于任何一组。
- 可能相交：可能有的数据行属于多个组。

The screenshot shows the MyBox Data Tool interface with a CSV file titled "糖尿病的病情预测_条件_2022-11-10_14-16-10-..." open. The window title bar includes the MyBox logo and the file path. The menu bar contains: 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 data manipulation. The main area displays a grouped data table with the following columns: 表格行 (Table Row), 数据行 (Data Row), 组 (Group), 条件 (Condition), 年龄 (Age), 性别 (Gender), BMI(体质... (BMI (Body Mass Index)), BP(平均血压) (Average Blood Pressure), and S1((S1)). The data rows show age conditions like "{年龄} < 20" and "{年龄} < 45". The bottom of the interface includes a status bar with "选中的: 0 行: [1-50]50/1374 页尺寸 50 / 28" and navigation buttons.

表格行	数据行	组	条件	年龄	性别	BMI(体质...	BP(平均血压)	S1(
1	1	1	{年龄} < 20	19	1	23.2	75	143
2	2	1	{年龄} < 20	19	1	25.3	83	225
3	3	1	{年龄} < 20	19	1	19.2	87	124
4	4	2	{年龄} < 45	19	1	23.2	75	143
5	5	2	{年龄} < 45	19	1	25.3	83	225
6	6	2	{年龄} < 45	19	1	19.2	87	124
7	7	2	{年龄} < 45	20	2	24.2	88	126
8	8	2	{年龄} < 45	20	2	22.1	87	171
9	9	2	{年龄} < 45	20	1	22.9	87	191
10	10	2	{年龄} < 45	21	1	24.2	76	147
11	11	2	{年龄} < 45	21	1	20.1	63	135
12	12	2	{年龄} < 45	22	1	19.9	75	175
13	13	2	{年龄} < 45	22	1	19.3	82	156
14	14	2	{年龄} < 45	22	1	23	73	161

2.10.6 行号分组

2.10.6.1 分组方式

按照数据行号分割数据行。有三种选择：

- 分割的尺寸（数据行号间隔）
- 分割的个数
- "开始行号-结束行号"的列表



2.10.6.2 实现原理

1. 把数据转换为临时数据库表。
2. 计算各组的行号区间
3. 遍历数据行，按行号指定组。

2.10.6.3 分组的结果

- 若按分割尺寸或分割个数，则：
 - 完全：每个数据行都属于一组。
 - 不相交：每个数据行都只属于一组。
- 若按"开始行号-结束行号"的列表，则：
 - 可能不完全：可能有的数据行不属于任何一组。
 - 可能相交：可能有的数据行属于多个组。

MyBox 编辑CSV文件 : CSV - 790 - d:\tmp\mybox-data-6.5.9\generated\糖尿病的病情预测_数据行号.csv

表格行	数据行	组	数据行号	年龄	性别	BMI(体质...	BP(平均血压)	S+
<input type="checkbox"/> 1	51	1	[1,55]	31	1	29.7	88	16
<input type="checkbox"/> 2	52	1	[1,55]	32	1	26.5	86	18
<input type="checkbox"/> 3	53	1	[1,55]	32	2	31	100	17
<input type="checkbox"/> 4	54	1	[1,55]	32	2	22	88	13
<input type="checkbox"/> 5	55	1	[1,55]	32	2	31.4	89	15
<input type="checkbox"/> 6	56	2	[56,110]	32	2	27.8	89	21
<input type="checkbox"/> 7	57	2	[56,110]	32	2	25.4	90.33	19
<input type="checkbox"/> 8	58	2	[56,110]	32	1	30.5	89	18
<input type="checkbox"/> 9	59	2	[56,110]	33	1	19.5	80	17
<input type="checkbox"/> 10	60	2	[56,110]	33	1	18.9	70	16
<input type="checkbox"/> 11	61	2	[56,110]	33	2	20.8	84	12
<input type="checkbox"/> 12	62	2	[56,110]	33	1	35	89	20
<input type="checkbox"/> 13	63	2	[56,110]	33	2	25.4	102	20
<input type="checkbox"/> 14	64	2	[56,110]	33	1	25.3	85	15

选中的: 0 行: [51-100] 50/442 页尺寸 50 页 2 /9 < > << >> ← →

2.11 修改数据

悬停或点击功能按钮，选择“修改”下的功能。

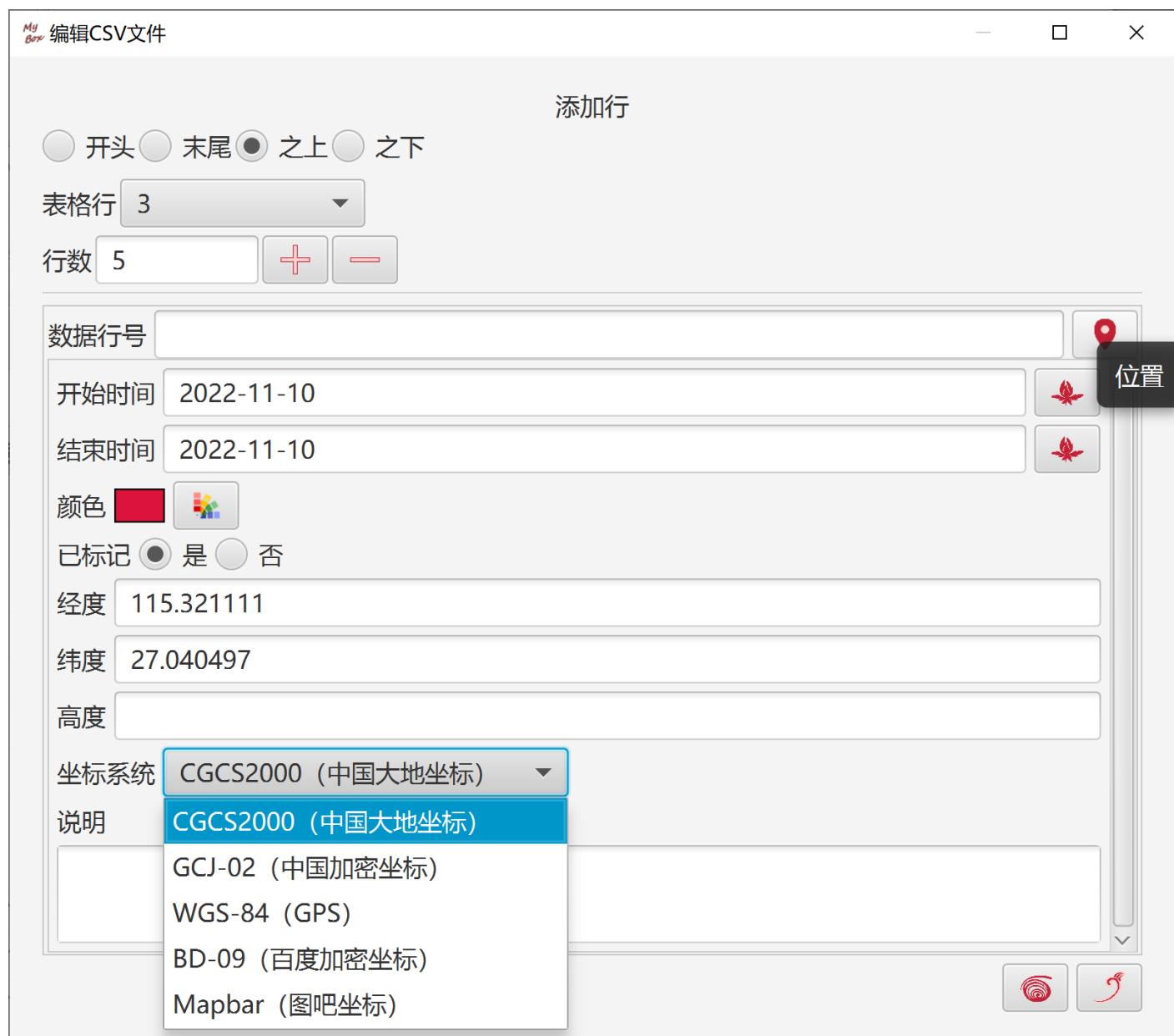
The screenshot shows the MyBox Data Tool interface with a context menu open over a table. The menu is titled "修改" (Modify) and includes options like "整理" (Organize), "计算" (Calculate), "图" (Chart), "分组图" (Group Chart), and "示例" (Example). A checked checkbox indicates "当鼠标悬停时弹出" (Pop up when mouse hovers). The menu also has a "关闭 (ESC/F6或点击对象外部任何一处)" (Close (ESC/F6 or click anywhere outside the object)) option.

	表格行	数据行	年龄
<input type="checkbox"/>	1	75	
<input type="checkbox"/>	2	55	
<input type="checkbox"/>	3	65	
<input type="checkbox"/>	4	50	
<input type="checkbox"/>	5	65	
<input type="checkbox"/>	6	90	
<input type="checkbox"/>	7	75	<input checked="" type="checkbox"/> 246
<input type="checkbox"/>	8	60	<input checked="" type="checkbox"/> 315
<input type="checkbox"/>	9	65	<input type="checkbox"/> 157
<input type="checkbox"/>	10	80	<input checked="" type="checkbox"/> 123
<input type="checkbox"/>	11	75	<input checked="" type="checkbox"/> 81
<input type="checkbox"/>	12	62	<input type="checkbox"/> 231
<input type="checkbox"/>	13	45	<input checked="" type="checkbox"/> 981
<input type="checkbox"/>	14	50	<input checked="" type="checkbox"/> 168

选中的: 0 行: [1-50]50/299 页尺寸 50 / 1 / 6

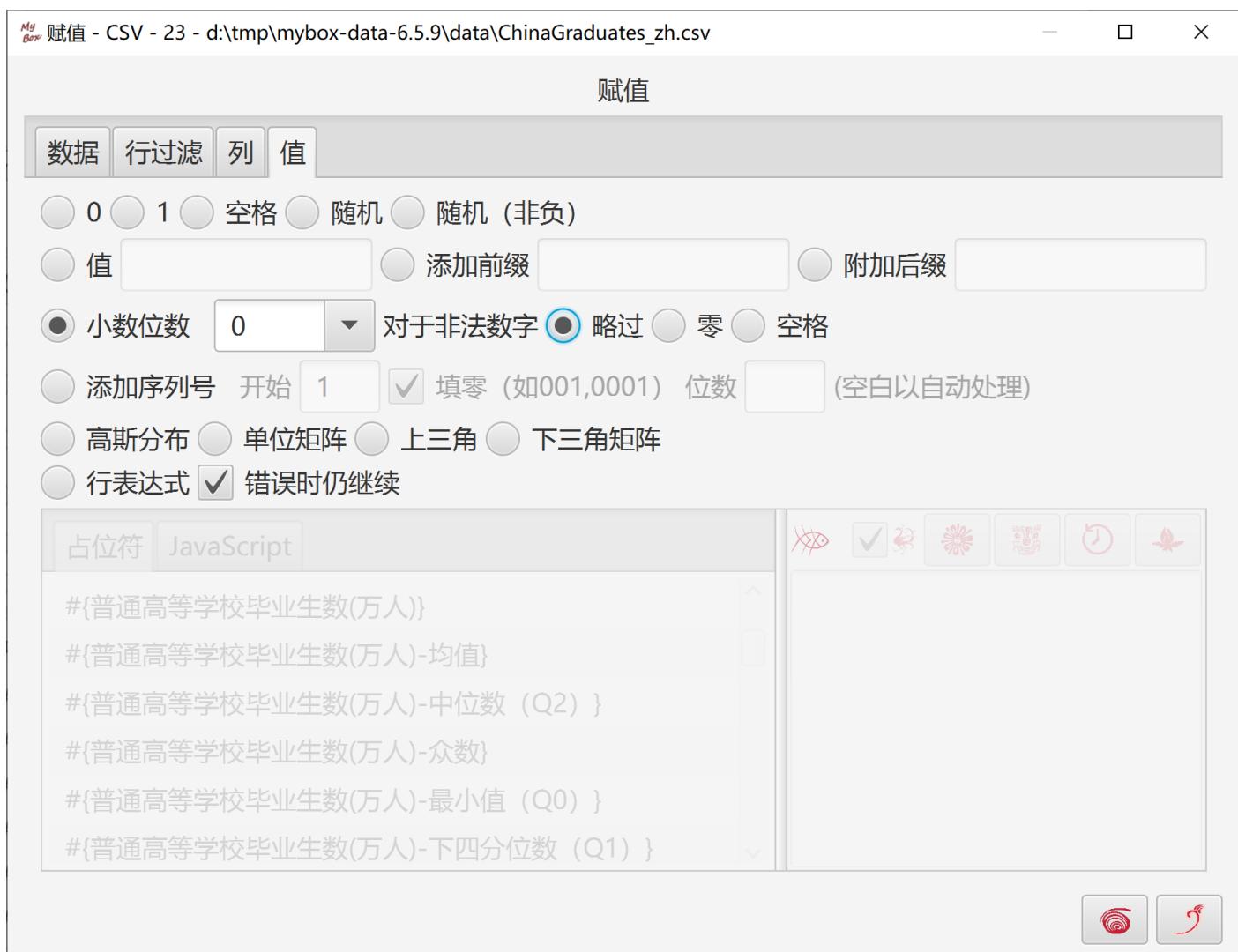
2.11.1 添加行

1. 选择添加的位置：开头、末尾、指定行之上、指定行之下。
2. 设置新增行数。
3. 设置新数据的各个列值。根据列的类型显示数值的编辑控件：
 - 若数据包含经度和纬度，则显示“位置”按钮。
 - 对日期、时间、纪元，显示“示例”按钮。
 - 对布尔值，显示选择框。
 - 对于枚举型，显示数据选择列表。
 - 对于颜色，显示“颜色”按钮。



2.11.2 赋值

1. 选择数据行。可设行过滤。
2. 选择要赋值的列。
3. 选择赋值的方式：
 - 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
4. 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。



2.11.3 删除

1. 选择数据行。可设行过滤。
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	选择一些以处理，或者不选任何以处理表中所有：	第三产业...
<input type="checkbox"/>	1	1952	679.1	- 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。	105.1
<input type="checkbox"/>	2	1953	824.4	- 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，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“+”。	257
<input type="checkbox"/>	3	1954	859.8	- 此操作不影响数据。	210.8
<input type="checkbox"/>	4	1955	911.6	- 要调整列的顺序，点击并拖动列头。	221.5
<input type="checkbox"/>	5	1956	1,030.7	- 此操作不影响数据。	269.1
<input type="checkbox"/>	6	1957	1,071.4	- 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。	365
<input type="checkbox"/>	7	1958	1,312.3	- 此操作只影响表中当前数据。	382.8
<input type="checkbox"/>	8	1959	1,447.5	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。	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"/>

← → ← →

2.11.4 设置风格/标识异常值

2.11.4.1 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

2.11.4.2 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。

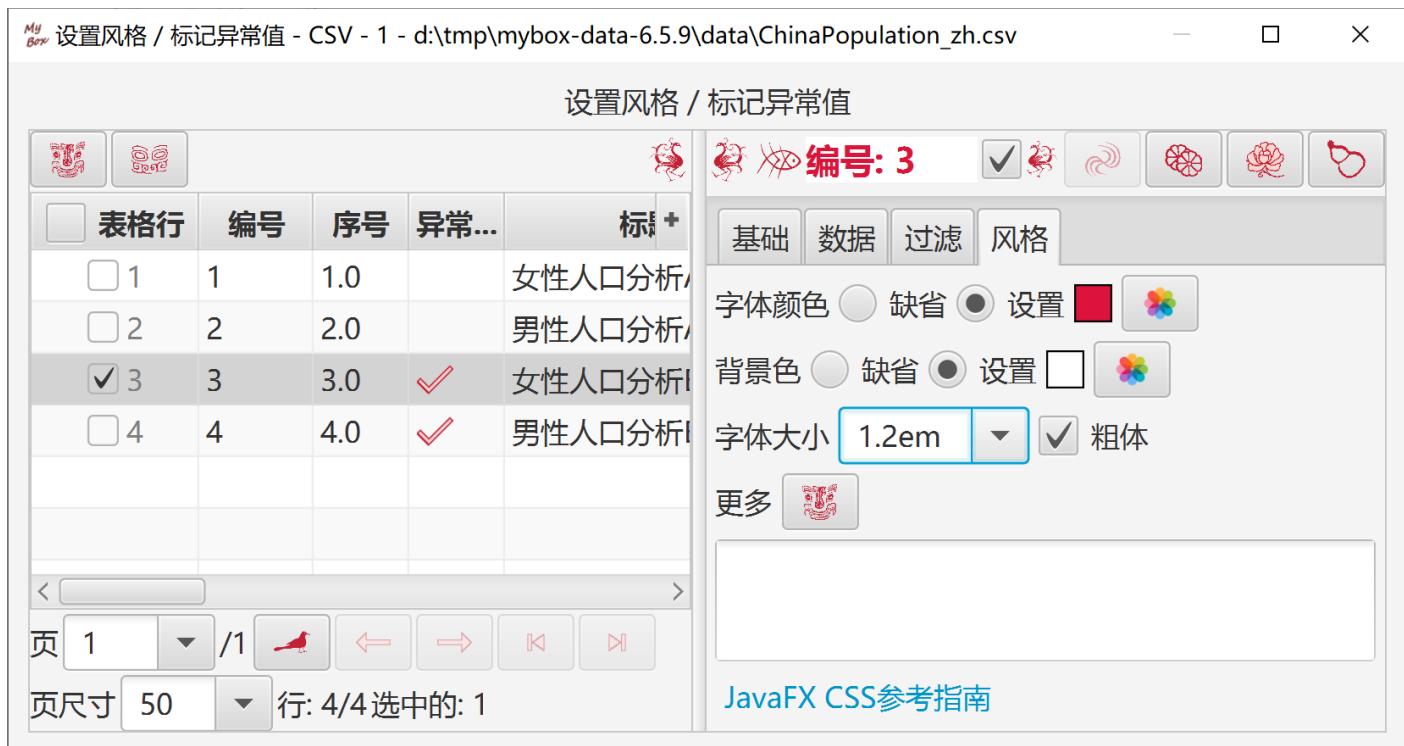
The screenshot shows the '设置风格 / 标记异常值' (Set Style / Mark Abnormal Values) dialog for a CSV file named 'ChinaPopulation_zh.csv'. The dialog has a toolbar at the top with icons for document, search, and other functions. Below the toolbar is a table with columns: 序号 (Row Number), 异常值 (Abnormal Value), 标题 (Title), 从 (From), 到 (To), and + (Add). The table contains four rows of data:

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	✓	女性人口分析B	25	-1	女性
4.0	✓	男性人口分析B	25	-1	男性

To the right of the table is a sidebar with tabs: 基础 (Basic), 数据 (Data) (selected), 过滤 (Filter), 风格 (Style). The '基础' tab shows settings for '数据行号' (Data Row Number) from 25 to the end. The '数据' tab shows checkboxes for columns: 年 (Year), 女性人口(万人) (Female Population in Ten Thousand), 男性人口(万人) (Male Population in Ten Thousand), 城镇人口(万人) (Urban Population in Ten Thousand), and 乡村人口(万人) (Rural Population in Ten Thousand). The '过滤' and '风格' tabs are also present.

2.11.4.3 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。



2.11.4.4 应用风格

- 在显示数据页面时，数据的所有风格按序号从小到大被逐条应用于数据行。
- 若风格有重合的数据范围，则对于重合部分的数据：序号大的风格覆盖序号小的风格。

编辑CSV文件 : CSV - 1 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv

The screenshot shows a data table with the following columns: 表格行 (Table Row), 数据行 (Data Row), 年 (Year), 年末总人口(万人) (Total Population at Year-end (Ten Thousand)), 男性人口(万人) (Male Population (Ten Thousand)), 女性人口(万人) (Female Population (Ten Thousand)), 城镇人口(万人) (Urban Population (Ten Thousand)), and 乡村人口(万人) (Rural Population (Ten Thousand)). The data rows show population figures for each year from 1972 to 1984, with some cells containing bold or colored text. The table is styled with various colors and fonts, demonstrating how styles are applied sequentially.

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

2.11.5 粘贴系统粘贴板的内容

1. 读取系统粘贴板中的内容。
2. 工具猜测分隔符并解析数据。
3. 用户可以指定分隔符来解析。
4. 可选解析方式:
 - CSV 解析：数值可包含分隔符和换行符，不支持空白字符解析和规则表达式解析。
 - 文本解析：数值不可包含分隔符和换行符，支持空白字符解析和规则表达式解析。
5. 可选是否是第一行定义列名。
6. 对解析后的数据，选择要粘贴的行和列。
7. 选择要粘贴的位置：数据行、数据列。
8. 选择粘贴的方式：替换、插入在上方、附加在下方。

MyBox 粘贴系统粘贴板的内容 : CSV - 791 - d:\tmp\mybox-data-6.5.9\AppData\Temp\2022-11-10_18-15-26-534_26

	年龄	贫血	肌酐磷酸激酶(CPK_mcg/L)
1	75	0	582
2	55	0	7861
3	65	0	146
4	50	true	111
5	65	true	160
6	90	true	47
7	75	true	246
8	60	true	315
9	65	0	157
10	80	true	123
11	75	true	81
12	62	0	231

若值包含分隔符或者换行符，它应被引号围起来。

2.11.6 粘贴 MyBox 粘贴板的内容

1. 选择 MyBox 粘贴板中的数据。
2. 选择要粘贴的行和列。
3. 选择要粘贴的位置：数据行、数据列。
4. 选择粘贴的方式：替换、插入在上方、附加在下方。



2.12 整理数据

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' open. A context menu is displayed over a table row, specifically over the '年' (Year) column header. The menu is titled '整理' (Organize) and includes the following options:

- 复制/过滤/查询/转换 (Copy/Filter/Search/Transform)
- 排序 (Sort)
- 转置 (Transpose)
- 01 归一化 (01 Normalization)** (highlighted in blue)
- 分割/分组 (Split/Group)

Other menu items visible in the background include '数据' (Data), '修改' (Modify), '计算' (Calculate), '图' (Chart), '分组图' (Group Chart), '示例' (Example), and '当鼠标悬停时弹出' (Pop up when mouse hovers). There is also a checked checkbox for '关闭 (ESC/F6或点击对象外部任何一处)' (Close (ESC/F6 or click anywhere outside the object)).

表格行	数据行	年	国民总收入...	国内生产...			
<input type="checkbox"/> 1	1	1952					
<input type="checkbox"/> 2	2	1953					
<input type="checkbox"/> 3	3	1954					
<input type="checkbox"/> 4	4	1955					
<input type="checkbox"/> 5	5	1956					
<input type="checkbox"/> 6	6	1957	1312.3	1312.3			
<input type="checkbox"/> 7	7	1958	1447.5	1447.5			
<input type="checkbox"/> 8	8	1959	1470.1	1470.1			
<input type="checkbox"/> 9	9	1960	1232.3	1232.3	441.1	393.5	397.7
<input type="checkbox"/> 10	10	1961	1162.2	1162.2	453.1	363.9	345.1
<input type="checkbox"/> 11	11	1962	1248.3	1248.3	497.5	412.8	337.9
<input type="checkbox"/> 12	12	1963	1469.9	1469.9	559	519.3	391.6
<input type="checkbox"/> 13	13	1964	1734	1734	651.1	608.5	474.4
<input type="checkbox"/> 14	14	1965					

Bottom navigation bar: 选中的: 0 行: [1-50]50/70 页尺寸: 50 页 1 /2

2.12.1 复制/过滤/查询/转换

1. 选择数据行。可设行过滤。
2. 选择要复制的列。
3. 选择是否：
 - 复制行号
 - 复制列名
 - 按照列的格式保存日期/时间/纪元和数字。
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。
5. 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。



2.12.2 排序

1. 选择数据行。可设行过滤。
2. 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。
3. 设置最多取结果的行数。
4. 选择要复制的列。
5. 选择是否复制行号、列名。
6. 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。
7. 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。

MyBox 排序 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

排序

数据 行过滤 选项

按列排序 (列的类型影响排序结果)

表格行	排序
<input type="checkbox"/> 1	年-降序
<input type="checkbox"/> 2	年-升序
<input checked="" type="checkbox"/> 3	国民总收入(GNI 亿元)-降序
<input type="checkbox"/> 4	国民总收入(GNI 亿元)-升序

最多取排序后数据行数(空/零/负值以不限制)

要复制的列 (不选表示所有)

年 国民总收入(GNI 亿元) 国内生产总值(GDP 亿元) 第一产业增加值(VA1 亿元)
 第二产业增加值(VA2 亿元) 第三产业增加值(VA3 亿元) 人均国内生产总值(元)

行号 列名

目标

外部 CSV XLS TXT   

名字 中国的国内生产总值(GDP)

表内 插入在上方 附加在下方 替换

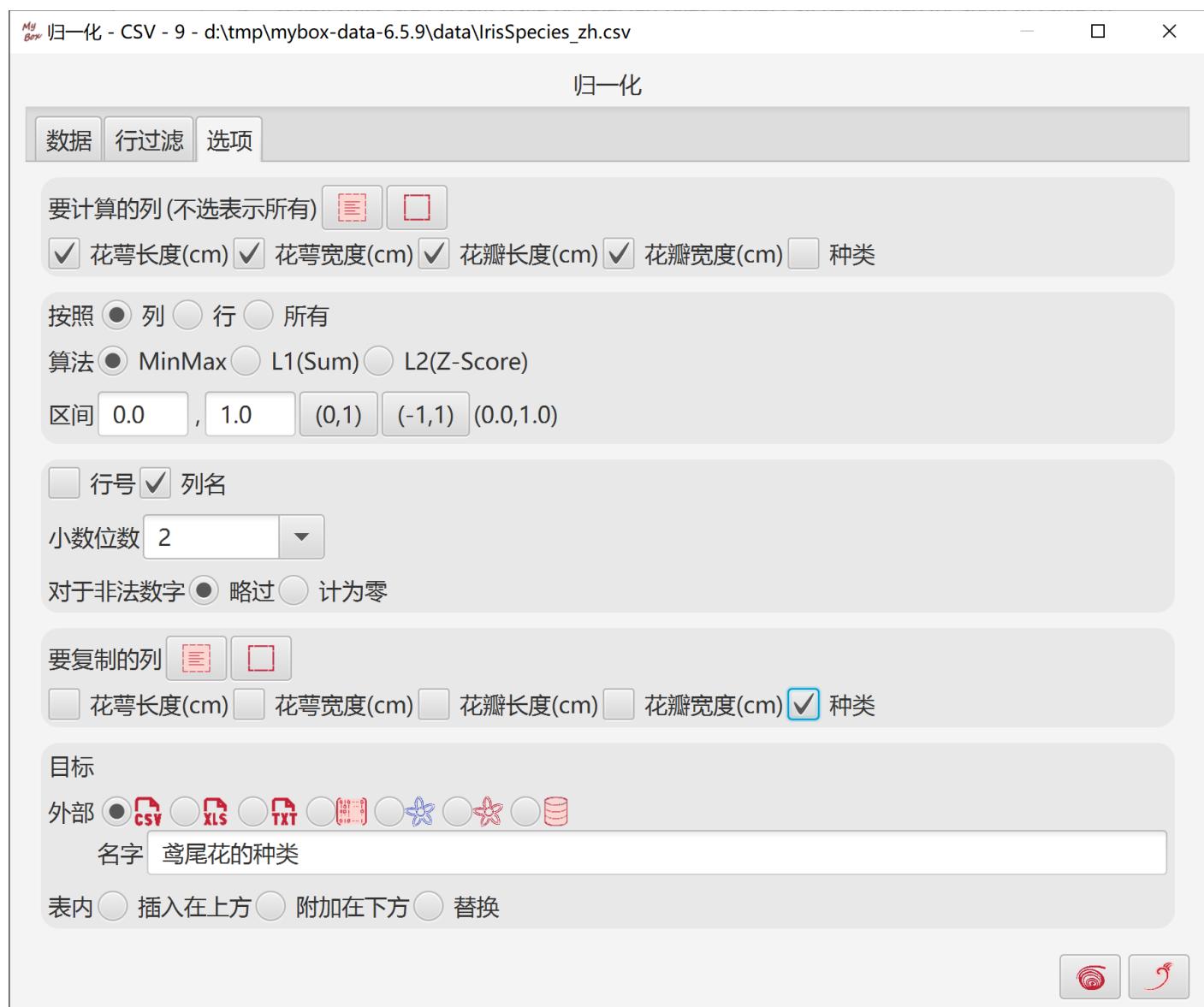
2.12.3 转置

1. 选择数据行。可以设置行过滤。
2. 选择要计算的列。
3. 可选是否:
 - 保存行号
 - 保存列名
 - 把第一列当作列名
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。
5. 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。



2.12.4 归一化

1. 选择数据行。可以设置行过滤。
2. 选择要计算列。
3. 对象可选：按照列、行、所有。
4. 算法可选：最大最小值（可设置区间）、和（L1）、ZScore（L2）。
5. 对非数字可选：略过、计为零。
6. 选择要复制的列。
7. 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。
8. 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。



2.12.5 分割/分组

1. 选择数据行。可以设置行过滤。

2. 设置分组的方式。

3. 选择要复制的列。

4. 设置排序条件、及每组最多行数。

5. 选择是否包含行号

6. 目标可选：

- 包含所有分组的单个文件。

- 每组一个文件。

- 数据库表。

MyBox 分割/分组 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

分割/分组

数据 行过滤 组 选项

要复制的列(不选表示所有)

<input type="checkbox"/> 年	<input type="checkbox"/> 国民总收入(GNI 亿元)	<input type="checkbox"/> 国内生产总值(GDP 亿元)	<input type="checkbox"/> 第一产业增加值(VA1 亿元)
<input type="checkbox"/> 第二产业增加值(VA2 亿元)	<input type="checkbox"/> 第三产业增加值(VA3 亿元)	<input type="checkbox"/> 人均国内生产总值(元)	

按列排序 (列的类型影响排序结果)

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

每组最多行数(空/零/负值以不限制)

行号

目标 包含所有组的单个文件 每组一个文件 数据库表

2.13 数据计算

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' open. A context menu is displayed over a table containing GDP data from 1952 to 1965. The menu is organized into several categories:

- 数据**
- 修改**
- 整理**
- 计算** (highlighted in blue)
- 图**
- 分组图**
- 示例**
- 当鼠标悬停时弹出**
- 关闭 (ESC/F6或点击对象外部任何一处)**

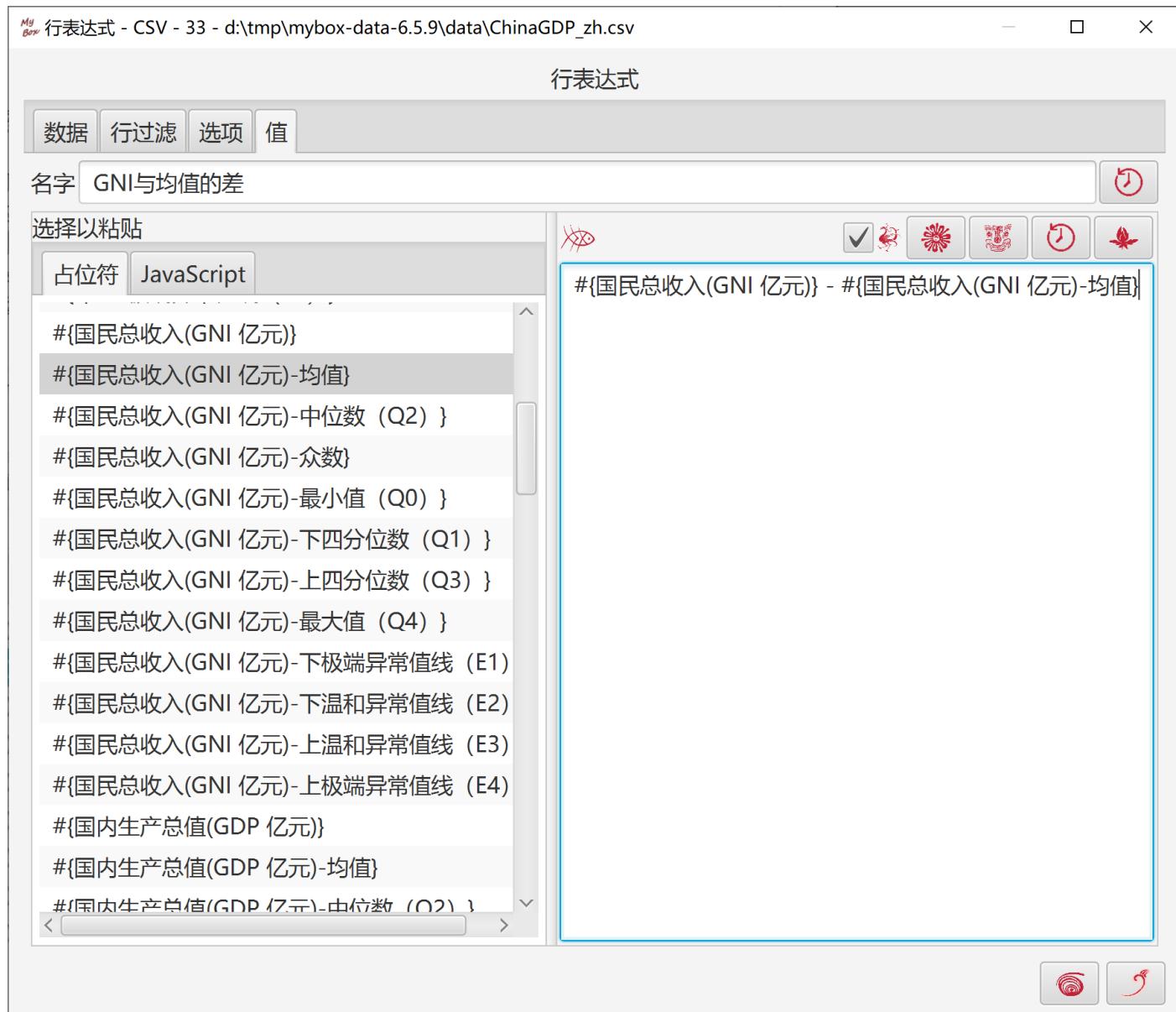
The table data is as follows:

年	国民总收入
1952	679.1
1953	1000.0
1954	1232.3
1955	1162.2
1956	1248.3
1957	1469.9
1958	1469.9
1959	559
1960	519.3
1961	412.8
1962	441.1
1963	393.5
1964	363.9
1965	363.9

At the bottom, there are navigation buttons for pages and arrows.

2.13.1 行表达式

1. 选择数据行。可以设置行过滤器。
2. 选择要复制的列。
3. 输入行表达式及其名字。
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。
5. 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。



2.13.2 描述性统计

1. 选择数据行。可以设置行过滤器。
2. 选择要计算的列
3. 选择要统计的值：
计数、合计、均值、几何平均数、平方和、众数、斜度、
总体方差、样本方差、总体标准差、样本标准差、
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)
4. 选择对象：按照列、行（可选类比列）、所有
5. 设置小数位数。
6. 可选对非数字：略过、计为零。
7. 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。



2.13.3 分组统计

2.13.3.1 计算的选项

1. 选择数据行。可以设置行过滤器。
2. 设置分组条件。
3. 选择要计算的列。
4. 选择统计项。“计数”总是被计算。
5. 设置小数位数。
6. 对于非法值的处理：设为空值、设为零。

MyBox 分组统计 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

数据 行过滤 组 选项
 在图上显示所有值

要计算的列(不选表示所有)
分组数据 统计数据 图数据 XY图 饼状图

数据行	id	组编号	区间_年	年	国民总收入...	国内生产...
1	1	1	[1952,1961)	1952	679.1	679.1
2	2	1	[1952,1961)	1953	824.4	824.4
3	3	1	[1952,1961)	1954	859.8	859.8
4	4	1	[1952,1961)	1955	911.6	911.6
5	5	1	[1952,1961)	1956	1030.7	1030.7
6	6	1	[1952,1961)	1957	1071.4	1071.4
7	7	1	[1952,1961)	1958	1312.3	1312.3
8	8	1	[1952,1961)	1959	1447.5	1447.5
9	9	1	[1952,1961)	1960	1470.1	1470.1
10	10	2	[1961,1970)	1961	1232.3	1232.3
11	11	2	[1961,1970)	1962	1162.2	1162.2
12	12	2	[1961,1970)	1963	1248.3	1248.3
13	13	2	[1961,1970)	1964	1469.9	1469.9
14	14	2	[1961,1970)	1965	1734	1734
15	15	2	[1961,1970)	1966	1888.7	1888.7
16	16	2	[1961,1970)	1967	1794.2	1794.2

小数位数
对于非法数字 设为空值 设为零

2.13.3.2 分组数据

分组后，数据新增 2 列：组编号、组参数。

分组统计 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

数据行	id	组编号	区间_年	年	国民总收入...	国内生产总...	第一产业增...	第二产业增...	第三产 +
1	1	1	[1952,1961)	1952	679.1	679.1	342.9	141.1	195.1
2	2	1	[1952,1961)	1953	824.4	824.4	378	191.6	254.8
3	3	1	[1952,1961)	1954	859.8	859.8	392	210.8	257
4	4	1	[1952,1961)	1955	911.6	911.6	421	221.5	269.1
5	5	1	[1952,1961)	1956	1030.7	1030.7	443.9	280.4	306.5
6	6	1	[1952,1961)	1957	1071.4	1071.4	430	316.6	324.8
7	7	1	[1952,1961)	1958	1312.3	1312.3	445.9	483.6	382.8
8	8	1	[1952,1961)	1959	1447.5	1447.5	383.8	616.7	447
9	9	1	[1952,1961)	1960	1470.1	1470.1	340.7	652.6	476.8

选中的: 0 行: [1-20]20/70 页尺寸 20 / 4 < > << >>

2.13.3.3 统计数据

基于分组数据，对于每组计算指定的统计值。

分组统计 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

数据行	id	组编号	区间_年	列名	组_计数	组_总和	组_均值	组_几何平 +
1	1	1	[1952,1961)	国民总收入(GNI 亿元)	9	9606.9	1067.43	1034.18
2	2	1	[1952,1961)	国内生产总值(GDP 亿元)	9	9606.9	1067.43	1034.18
3	3	1	[1952,1961)	第一产业增加值(VA1 亿元)	9	3578.2	397.58	395.74
4	4	1	[1952,1961)	第二产业增加值(VA2 亿元)	9	3114.9	346.1	303.44
5	5	1	[1952,1961)	第三产业增加值(VA3 亿元)	9	2913.9	323.77	311.89
6	6	1	[1952,1961)	人均国内生产总值(元)	9	1527	169.67	166.37
7	7	2	[1961,1970)	国民总收入(GNI 亿元)	9	14235.9	1581.77	1553.98
8	8	2	[1961,1970)	国内生产总值(GDP 亿元)	9	14235.9	1581.77	1553.98

选中的: 0 行: [1-20]20/48 页尺寸 20 / 3 < > << >>

2.13.3.4 图数据

1. 统计数据被整合为适合生成图的行和列。
 2. 若选择“在图上显示所有值”，则在图上显示此表中的所有页的数据，否则在图上只显示此表中当前页的数据。

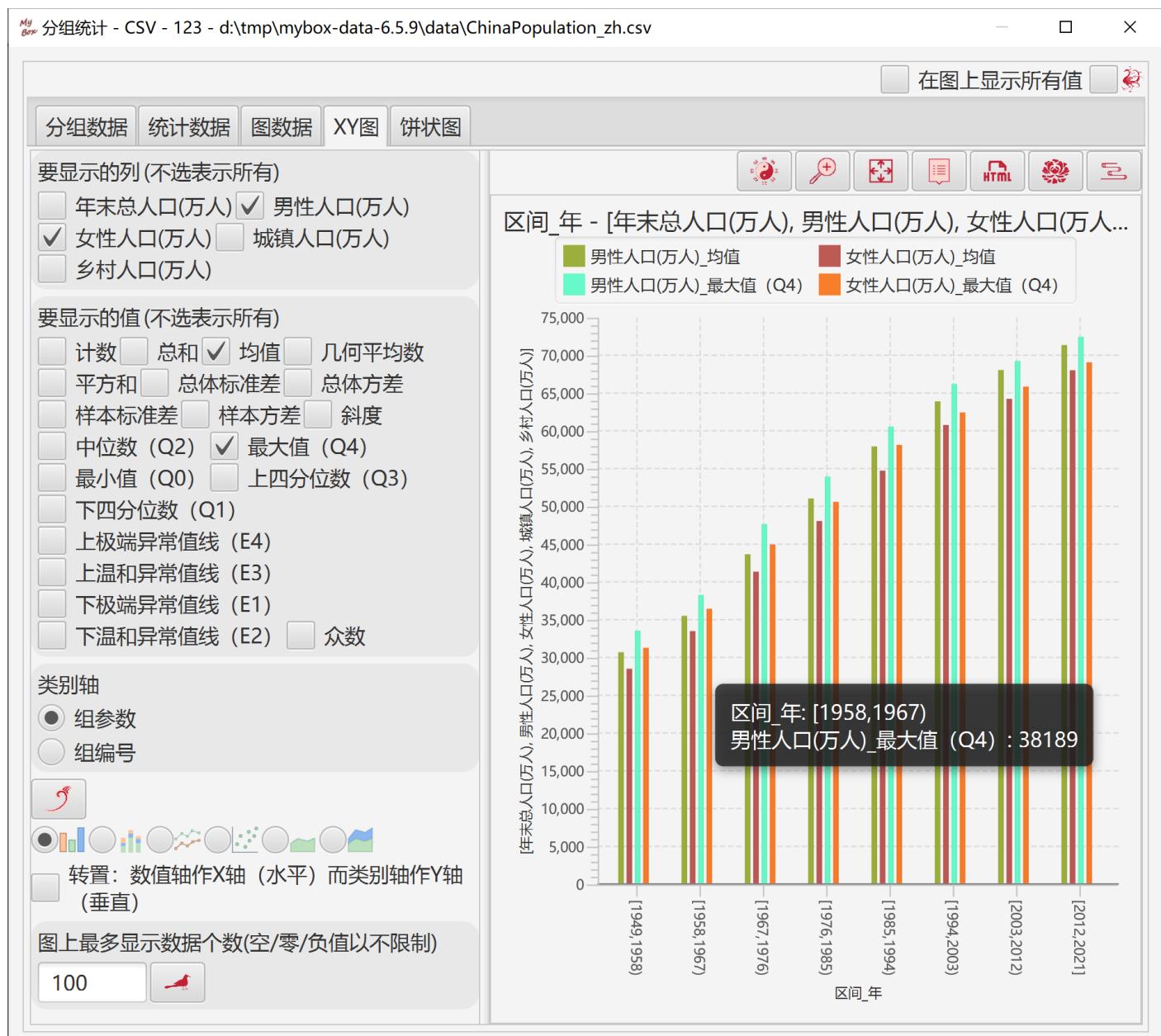
MyBox 分组统计 - CSV - 792 - d:\tmp\mybox-data-6.5.9\data\ChinaGDP_zh.csv

分组数据							<input type="checkbox"/> 在图上显示所有值 
分组数据	统计数据	图数据	XY图	饼状图			
数据行	组编号	区间_年	组_计数	国民总收入(GNI亿元)_总和	国民总收入(GNI亿元)_均值	国民总收入(GNI亿元)_中位数	国民总收入(GNI亿元)_标准差
1	1	[1952,1961)	9	9606.9	1067.43	1034.18	1034.18
2	2	[1961,1970)	9	14235.9	1581.77	1553.98	1553.98
3	3	[1970,1979)	9	25829.7	2869.97	2842.14	2842.14
4	4	[1979,1988)	9	64024.4	7113.82	6649.49	6649.49
5	5	[1988,1997)	9	315828.2	35092.02	30352.14	30352.14
6	6	[1997,2006)	9	1064800.3	118311.14	113624.56	113624.56
7	7	[2006,2015)	9	3822494.2	424721.58	400931.51	400931.51
8	8	[2015,2021]	7	6296896.8	899556.69	887605.52	887605.52

2.13.3.5 XY 图

分组后的统计值可显示为 XY 图:

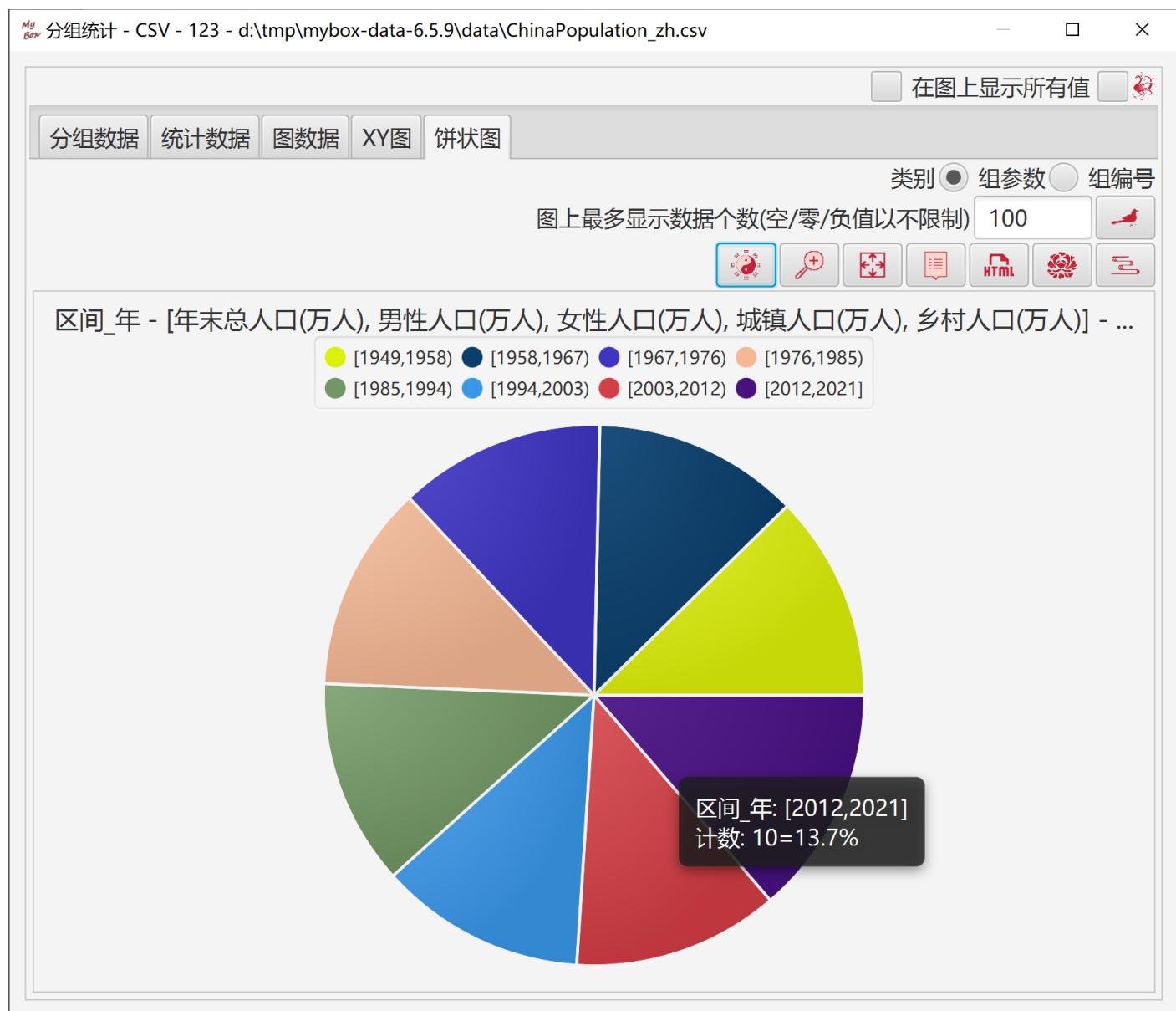
1. 选择要显示的列
2. 选择要显示的统计值。
3. 选择类别轴: 组参数、或组编号。
4. 选择 XY 图类型。
5. 选择是否转置 XY 图
6. 设置图上最多显示的数据个数。



2.13.3.6 饼状图

分组及其“计数”可显示为饼图：

1. 选择类别轴：组参数、或组编号。
2. 设置图上最多显示的数据个数。



2.13.4 简单线性回归

2.13.4.1 实现的基础

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。

2.13.4.2 计算的选项

1. 选择数据行。可设行过滤。
2. 选择一列作为自变量。
3. 选择另一列作为因变量。因变量不能与自变量是相同的列。
4. 设置期望的显著性水平(alpha)。
5. 选择是否包含截距。
6. 选择是否在图上显示所有值。
7. 设置小数位数。

2.13.4.3 回归

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

2.13.4.4 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

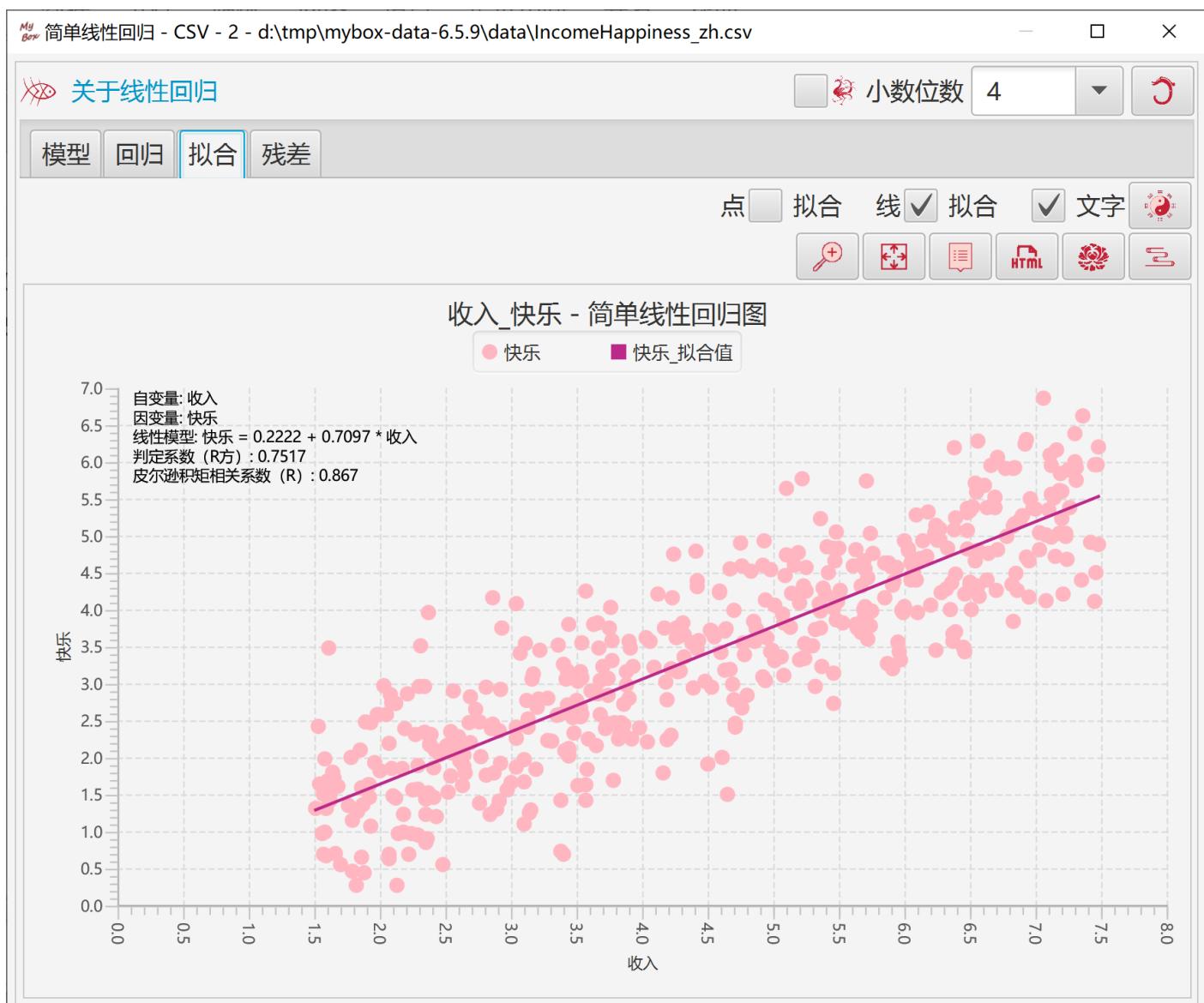
因变量: 快乐 =

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

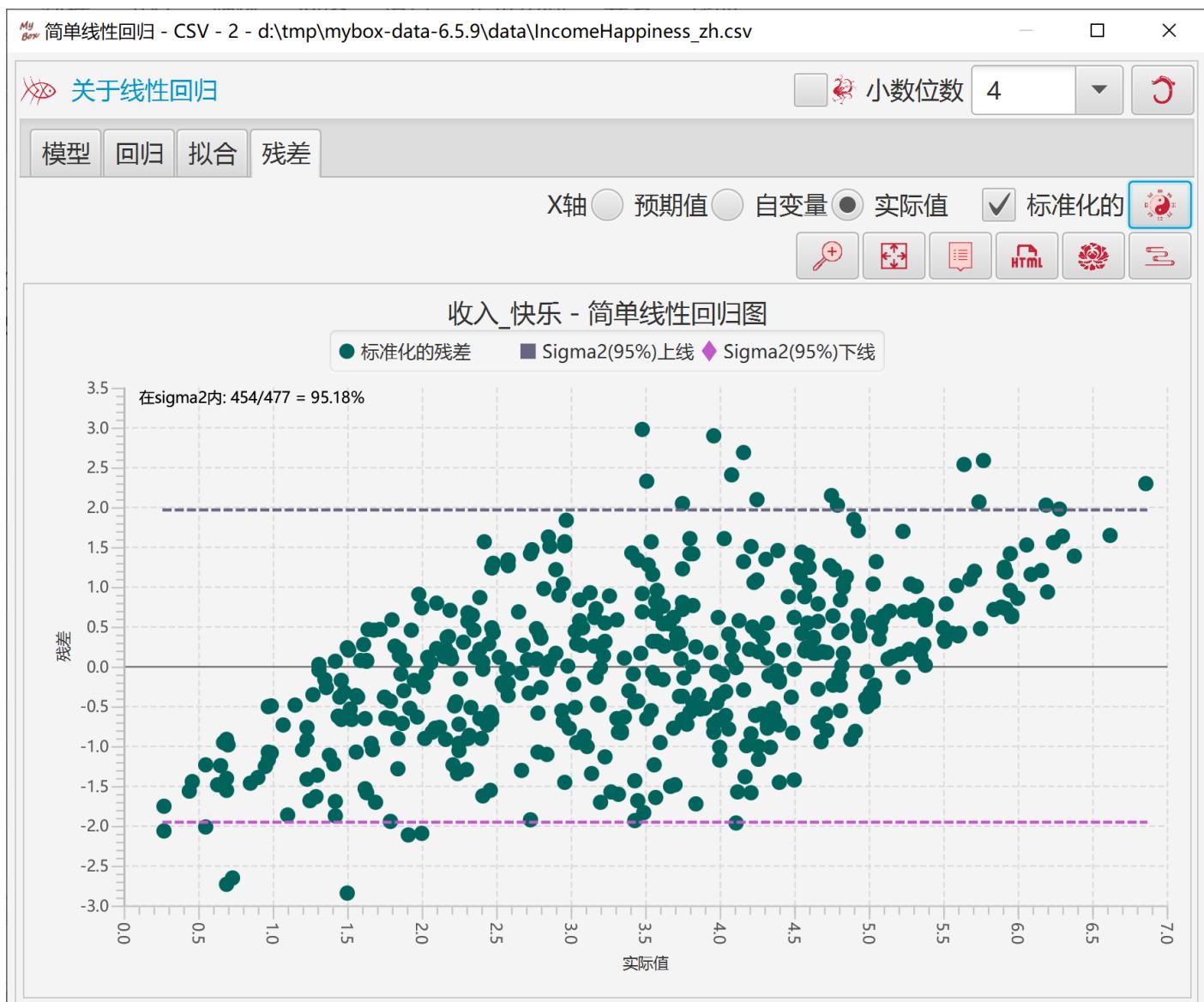
2.13.4.5 拟合图

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。



2.13.4.6 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。



2.13.5 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 - 801 - 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...	一年后病情进展
9	一年后病情...	S5(血清指...	0.32022...	0.56588...	一年后病情进展
4	一年后病情...	BP(平均血压)	0.19490...	0.44148...	一年后病情进展
8	一年后病情...	S4(血清指...	0.18528...	0.43045...	一年后病情进展
7	一年后病情...	S3(血清指...	0.15585...	0.39478...	一年后病情进展
10	一年后病情...	S6(血清指...	0.14629...	0.38248...	一年后病情进展
5	一年后病情...	S1(血清指...	0.04495...	0.21202...	一年后病情进展
1	一年后病情...	年龄	0.03530...	0.18788...	一年后病情进展
6	一年后病情...	S2(血清指...	0.03029...	0.17405...	一年后病情进展
2	一年后病情...	性别	0.00185...	0.043062	一年后病情进展

关于数据分析

数据 过滤 选项

自变量的候选 (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

因变量的候选 (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

期望的显著性水平 (alpha)

截距 (b0)

双击选中项以查看

2.13.6 多重线性回归

2.13.6.1 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 interface with the following details:

- Title Bar:** MyBox 多重线性回归 - CSV - 22 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_standardized_zh.csv
- Left Panel (Data Selection):**
 - Variables (Unselected):** 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).
 - Dependent Variable:** 一年后病情进展
 - Normalization:** L2(Z-Score) (selected)
 - Regression Method:** 普通最小二乘法(OLS) (selected)
 - Intercept:** 截距 (b0) (selected)
 - Note:** 所有数据都参与到内存中。大量数据可能导致内存溢出。
- Right Panel (Analysis Results):**
 - Model Summary:**

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

2.13.6.2 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' interface for a multiple linear regression analysis. The title bar indicates the file is 'DiabetesPrediction_standardized_zh.csv'. The left panel contains a 'Data' tab, a 'Filter' tab, and an 'Options' tab. Under 'Data', there are two sections: 'Independent Variables (Not selected to show all)' and 'Dependent Variable'. The independent variables listed are: 年龄 (Age), 性别 (Gender), BMI (Body Mass Index), BP (Average Blood Pressure), S1 (Blood Glucose Level), S2 (Plasma Insulin Level), S3 (Triceps Skinfold Thickness), S4 (Serum Insulin), S5 (Serum Non-esterified Fatty Acids), S6 (Serum Glucose), and 一年后病情进展 (Progression after one year). The dependent variable is set to '一年后病情进展' (Progression after one year). Below these are buttons for 'Standardize' and 'Filter'. The right panel is titled 'About Data Analysis' and has tabs for 'Model' and 'Regression'. It displays the regression equation: 线性模型: 一年后病情进展 = 0.0 - 0.006 * 年龄 - 0.148 * 性别 + 0.321 * BMI + 0.2 * BP - 0.489 * S1 + 0.294 * S2 - 0.062 * S3 + 0.109 * S4 + 0.464 * S5 + 0.042 * S6. There are input fields for each independent variable and a 'Predict' button. The bottom section shows the predicted value for the dependent variable.

2.13.7 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 - 801 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_standardized_zh.csv

数据行	因变量	自变量	调整后...	判定系...	系...
19	一年后病...	[性别, BMI(体质指数), BP(...	0.50767...	0.51771...	[-0.14
18	一年后病...	[性别, BMI(体质指数), BP(...	0.50764...	0.51657...	[-0.14
10	一年后病...	[年龄, 性别, BMI(体质指数)...	0.50656...	0.51774...	[-0.00
9	一年后病...	[年龄, 性别, BMI(体质指数)...	0.50650...	0.51657...	[-0.00
26	一年后病...	[BMI(体质指数), BP(平均...	0.49165...	0.499727	[0.348
27	一年后病...	[BMI(体质指数), BP(平均...	0.49099...	0.50023...	[0.344
16	一年后病...	[性别, BMI(体质指数), BP(...	0.48634...	0.49332...	[-0.14
17	一年后病...	[性别, BMI(体质指数), BP(...	0.48585...	0.49401...	[-0.14
7	一年后病...	[年龄, 性别, BMI(体质指数)...	0.48525...	0.49342...	[0.010
8	一年后病...	[年龄, 性别, BMI(体质指数)...	0.48477...	0.49411...	[0.010
24	一年后病...	[BMI(体质指数), BP(平均...	0.471218	0.47721...	[0.376
25	一年后病...	[BMI(体质指数), BP(平均...	0.47023...	0.47744...	[0.376
34	一年后病...	[BP(平均血压), S1(血清指...	0.41056...	0.41991...	[0.238
33	一年后病...	[BP(平均血压), S1(血清指...	0.40788...	0.41594...	[0.256
15	一年后病...	[性别, BMI(体质指数), BP(...	0.39582...	0.40267...	[-0.04
13	一年后病...	[性别, BMI(体质指数), BP(...	0.39568...	0.39979...	[-0.06
14	一年后病...	[性别, BMI(体质指数), BP(...	0.39519...	0.40068...	[-0.06
23	一年后病...	[BMI(体质指数), BP(平均...	0.39507...	0.40056...	[0.490
4	一年后病...	[年龄, 性别, BMI(体质指数)...	0.39477...	0.40026...	[0.023
6	一年后病...	[年龄, 性别, BMI(体质指数)...	0.39463...	0.40287...	[0.015
5	一年后病...	[年龄, 性别, BMI(体质指数)...	0.39407...	0.40094...	[0.017
21	一年后病...	[BMI(体质指数), BP(平均...	0.39324...	0.39500...	[0.406

双击选中项以查看

2.13.8 频数分布

1. 选择数据行。可以设置行过滤器。
2. 选择一列计算频数。
3. 选择是否忽略大小写。
4. 设置小数位数。
5. 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。
6. 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。

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

频数分布

数据 行过滤 选项

列 镇 忽略大小写

小数位数 2

目标

外部 CSV XLS TXT 矩阵 星形 曲线 插入

名字 BostonHousingPrices_zh.csv

表内 插入在上方 附加在下方 替换

2.13.9 数值百分比

1. 选择数据行。可以设置行过滤器。
2. 选择要计算的列。
3. 选择对象：按照列、行、所有。
4. 设置小数位数。
5. 对于负数选择：略过、计为零、计为绝对值。
6. 对于非法数字选择：略过、计为零。
7. 选择要复制的列。
8. 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。
9. 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 (窗口, 文档, 图像, 数据, 文件, 媒体, 网络, 设置, 最近访问, 开发, 帮助) and a toolbar with various icons. A data table is displayed in the main area, and a 'Percent' configuration dialog is open on the right.

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		

Percent Configuration Dialog:

- 要计算的列 (不选表示所有):** 指标 (checked), 2020年 (checked), 2019年 (unchecked), 2018年 (unchecked), 2017年 (unchecked), 2016年 (unchecked), 2015年 (unchecked), 2014年 (unchecked), 2013年 (unchecked).
- 按照:** 列 (radio button selected).
- 小数位数:** 2.
- 对于负数:** 略过 (radio button selected).
- 对于非法数字:** 略过 (radio button selected).
- 要复制的列:** 指标 (checked), 2020年 (checked), 2019年 (unchecked), 2018年 (unchecked), 2017年 (unchecked), 2016年 (unchecked), 2015年 (unchecked), 2014年 (unchecked), 2013年 (unchecked).
- 目标:** (empty field).

2.14 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 context menu open over a table of historical capital data. The menu is titled "图" (Chart) and includes options like XY图, 饼状图 (selected), 箱线图, 自比较条图, 相比较条图, XYZ图, and 位置分布.

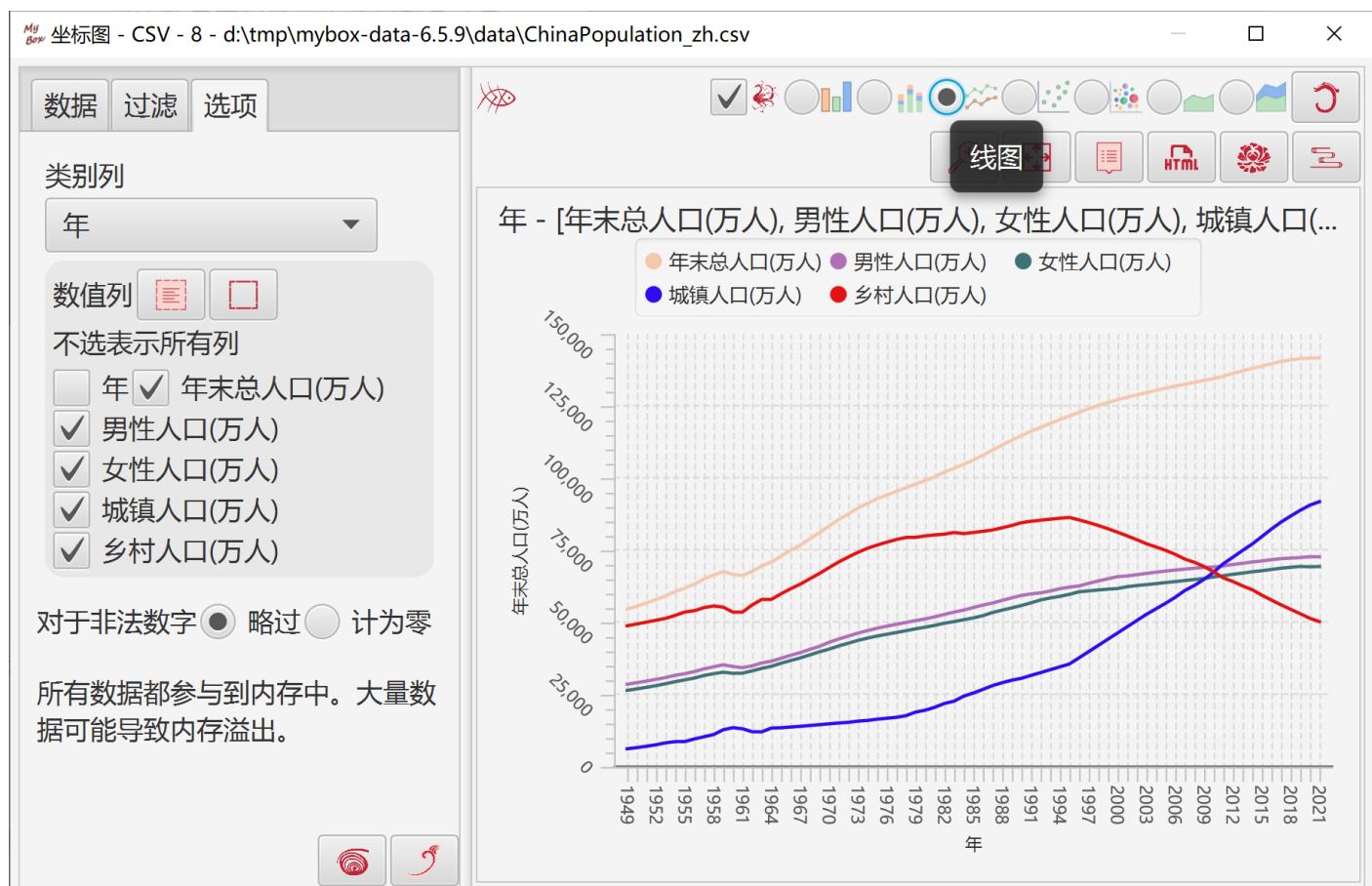
		国家	
<input type="checkbox"/>	1	传说-燧人	未知
<input type="checkbox"/>	2	传说-伏羲	未知
<input type="checkbox"/>	3	传说-女娲	未知
<input type="checkbox"/>	4	传说-神农/炎帝	未知
<input type="checkbox"/>	5	传说-黄帝	未知
<input type="checkbox"/>	6	传说-少昊	未知
<input type="checkbox"/>	7	传说-颛顼	未知
<input type="checkbox"/>	8	传说-帝喾	未知
<input type="checkbox"/>	9	传说-尧	未知
<input type="checkbox"/>	10	传说-舜	未知
<input type="checkbox"/>	11	夏	未知
<input type="checkbox"/>	12	夏	阳城 (河南郑州登封)
<input type="checkbox"/>	13	夏	阳翟 (河南禹州)
<input type="checkbox"/>	14	夏	安邑 (山西运城)

选中的: 0 行: [1-50] 50/163 页尺寸 50 / 页 1 / 4

2.14.1 XY图

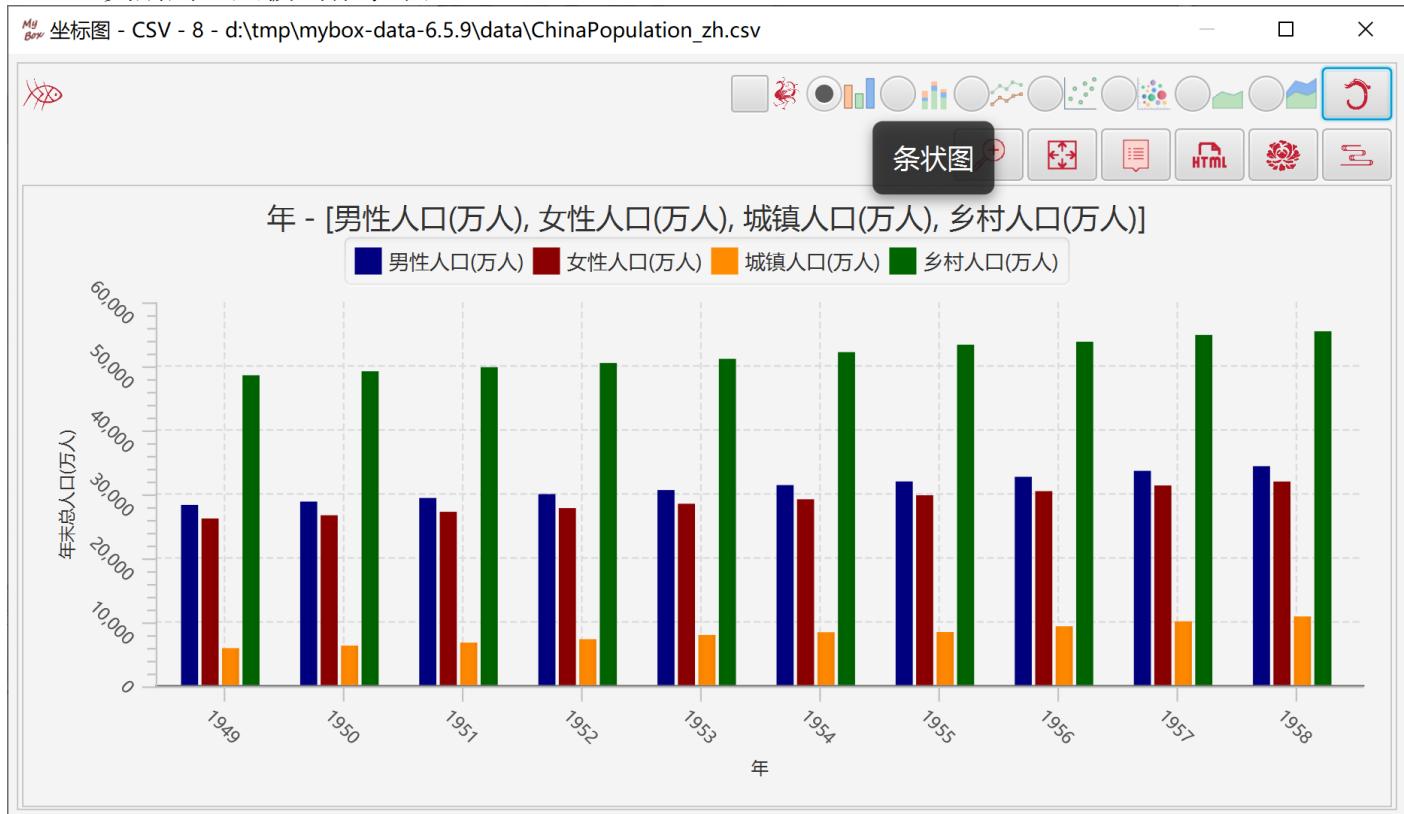
2.14.1.1 数据

1. 选择数据行。可以设置行过滤器。
2. 选择 XY 图的类型。
3. 选择一列为“类别轴”，以定义数据名字。
4. 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。
5. 缺省情况下，“类别轴”是横向轴、“数值轴”是竖向轴。
6. 对于非法数字，可选：忽略、或计为零。
7. 当处理所有数据行（所有页）时，坐标图需要考虑内存限制。
8. 点击“菜单”按钮以设置数据图的参数。
9. 点击“弹出”按钮以在弹出窗口中显示当前图。
10. 点击“数据”按钮以在数据表中显示 XY 图的数据。
11. 点击“网页”按钮以在网页中显示 XY 图的数据。



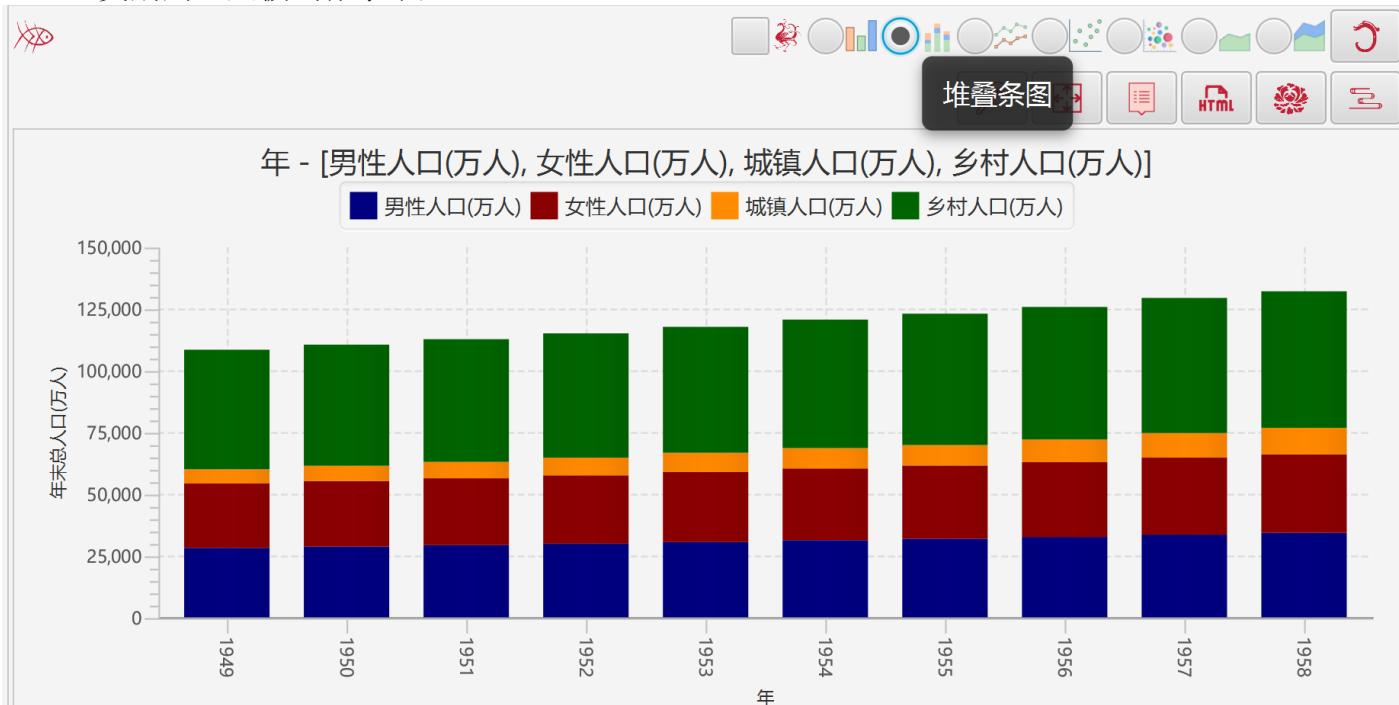
2.14.1.2 条图

- 以条块的高度表示数据的大小。
- 类别列总是被当作字串处理。



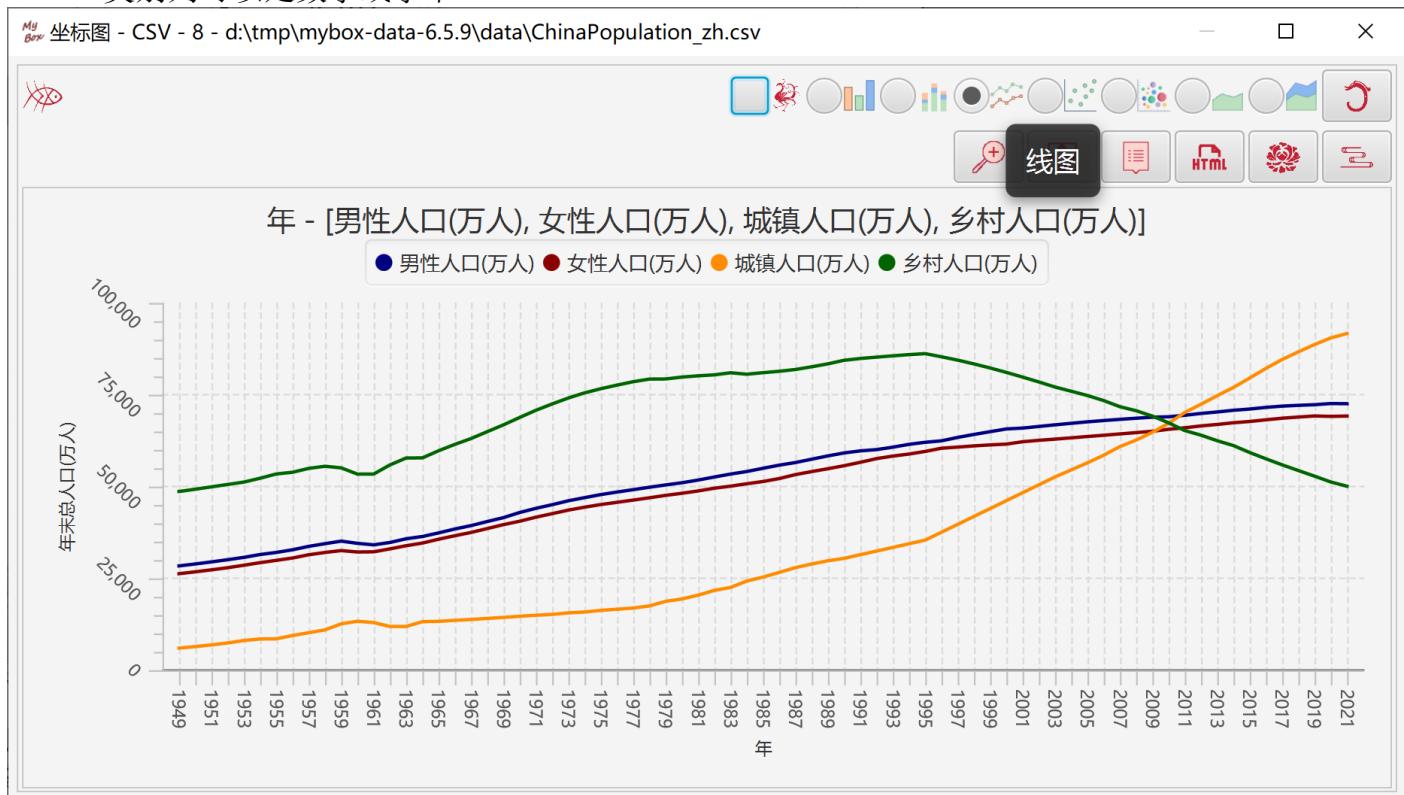
2.14.1.3 堆叠条图

- 以条块的高度表示数据的大小。
- 类别列总是被当作字串处理。



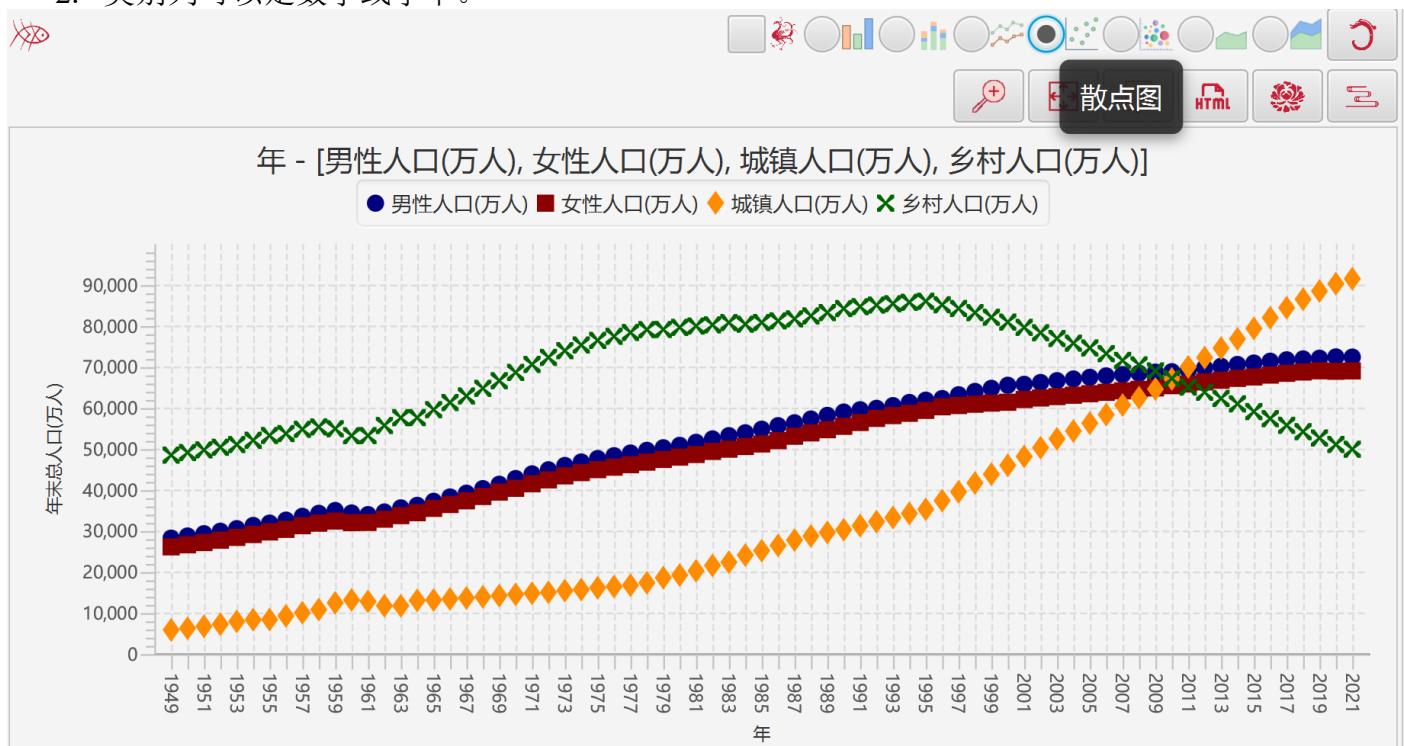
2.14.1.4 线图

- 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。



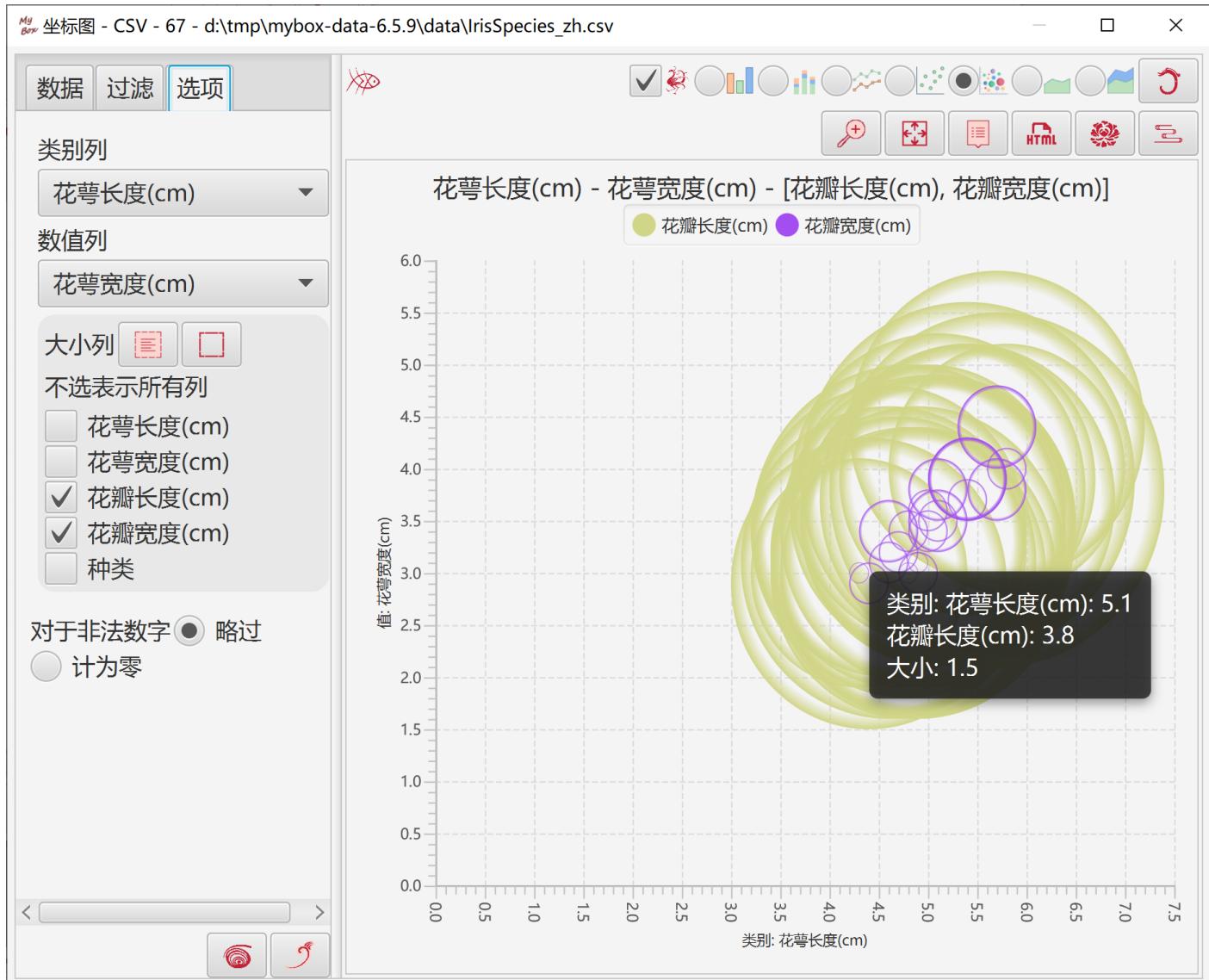
2.14.1.5 散点图

- 以符号表示数据的分布。
- 类别列可以是数字或字串。



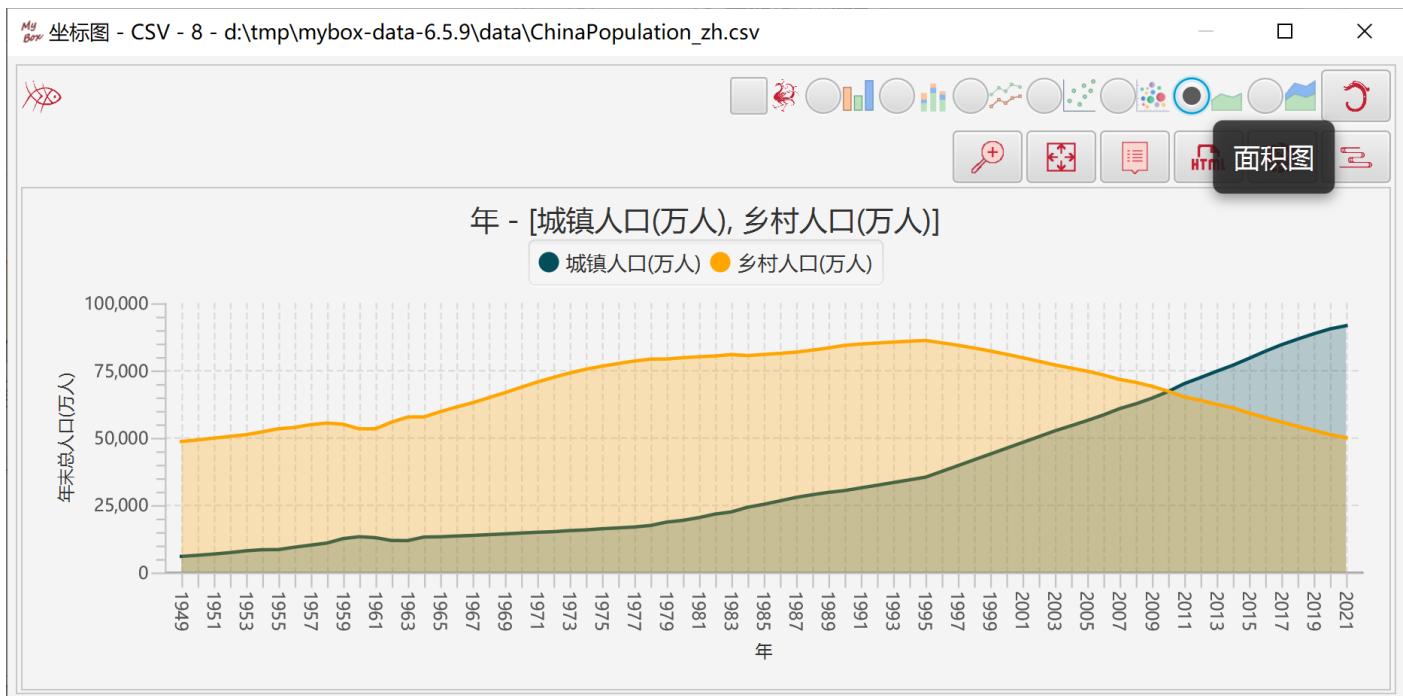
2.14.1.6 气泡图

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。



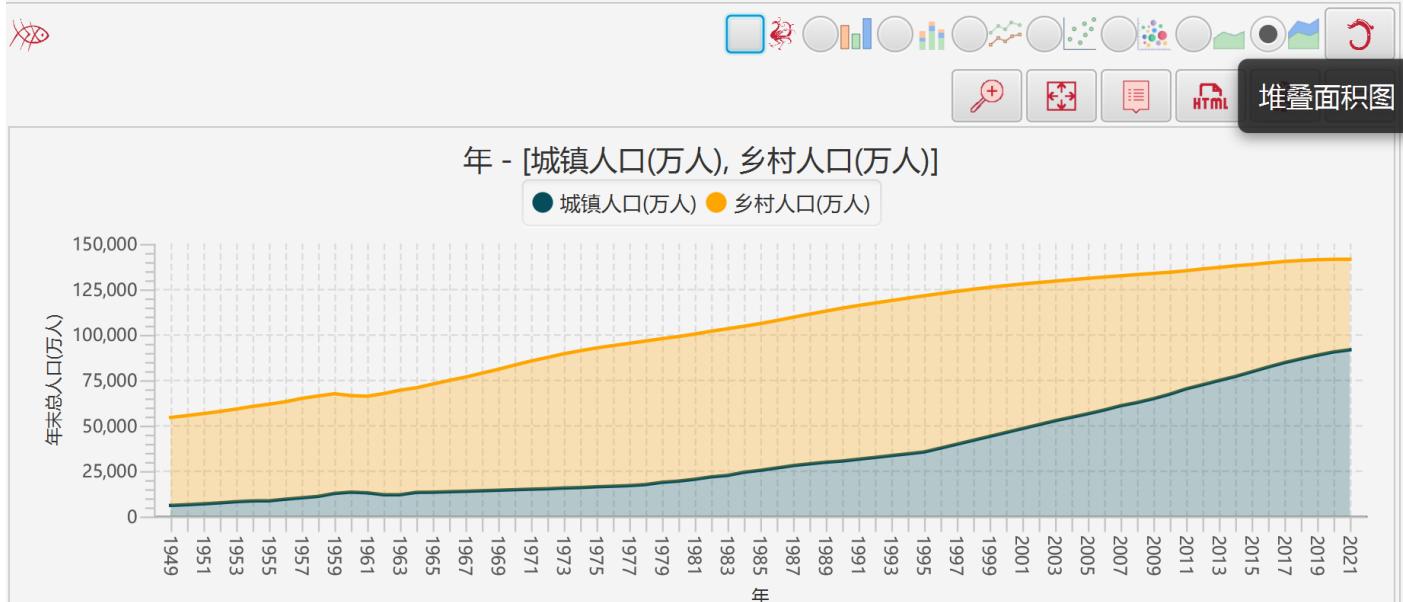
2.14.1.7 面积图

1. 以面积表示数据的大小。
2. 类别列总是被当作字串处理。



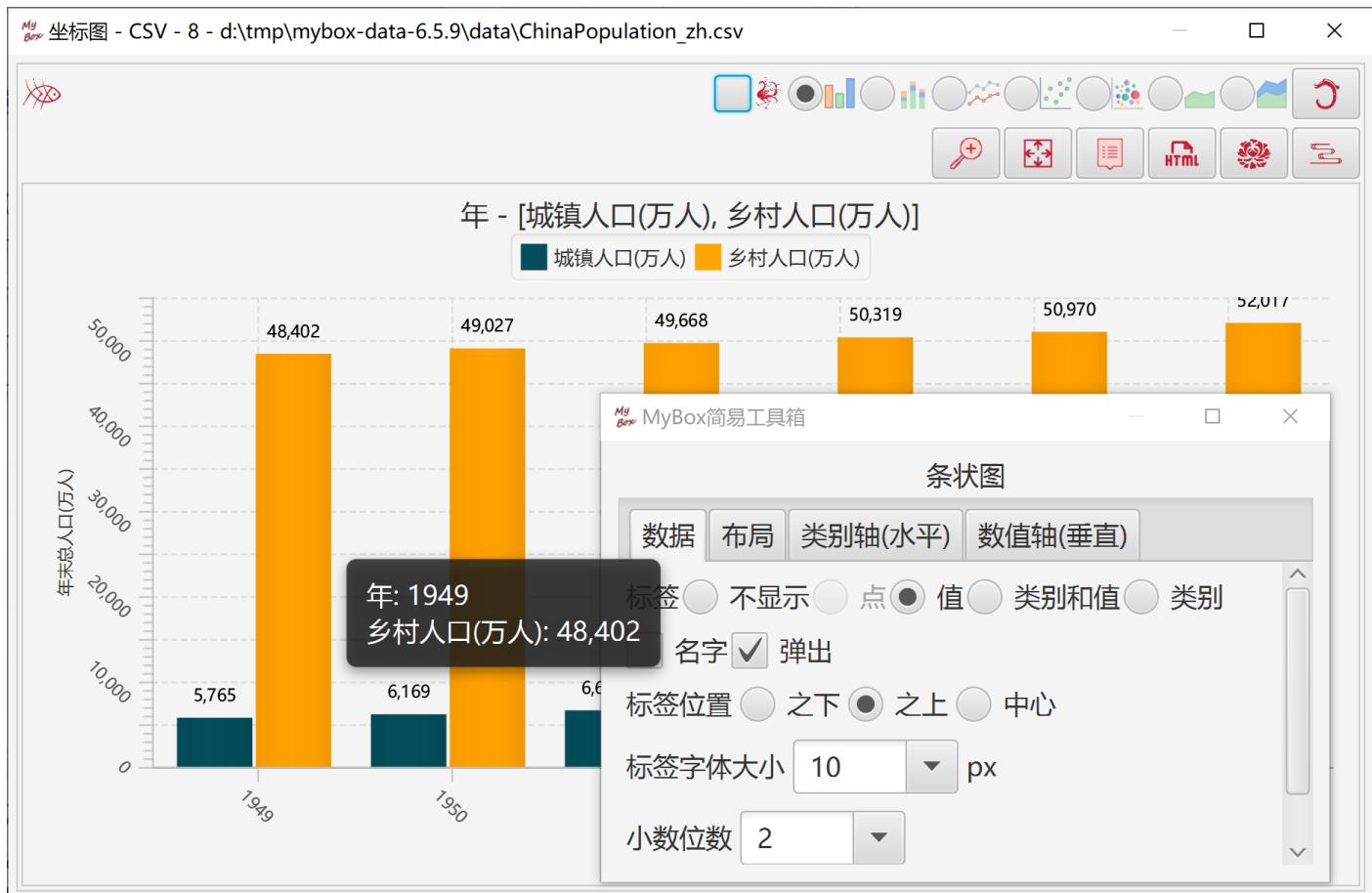
2.14.1.8 堆叠面积图

1. 以面积表示数据的大小。
2. 类别列总是被当作字串处理。



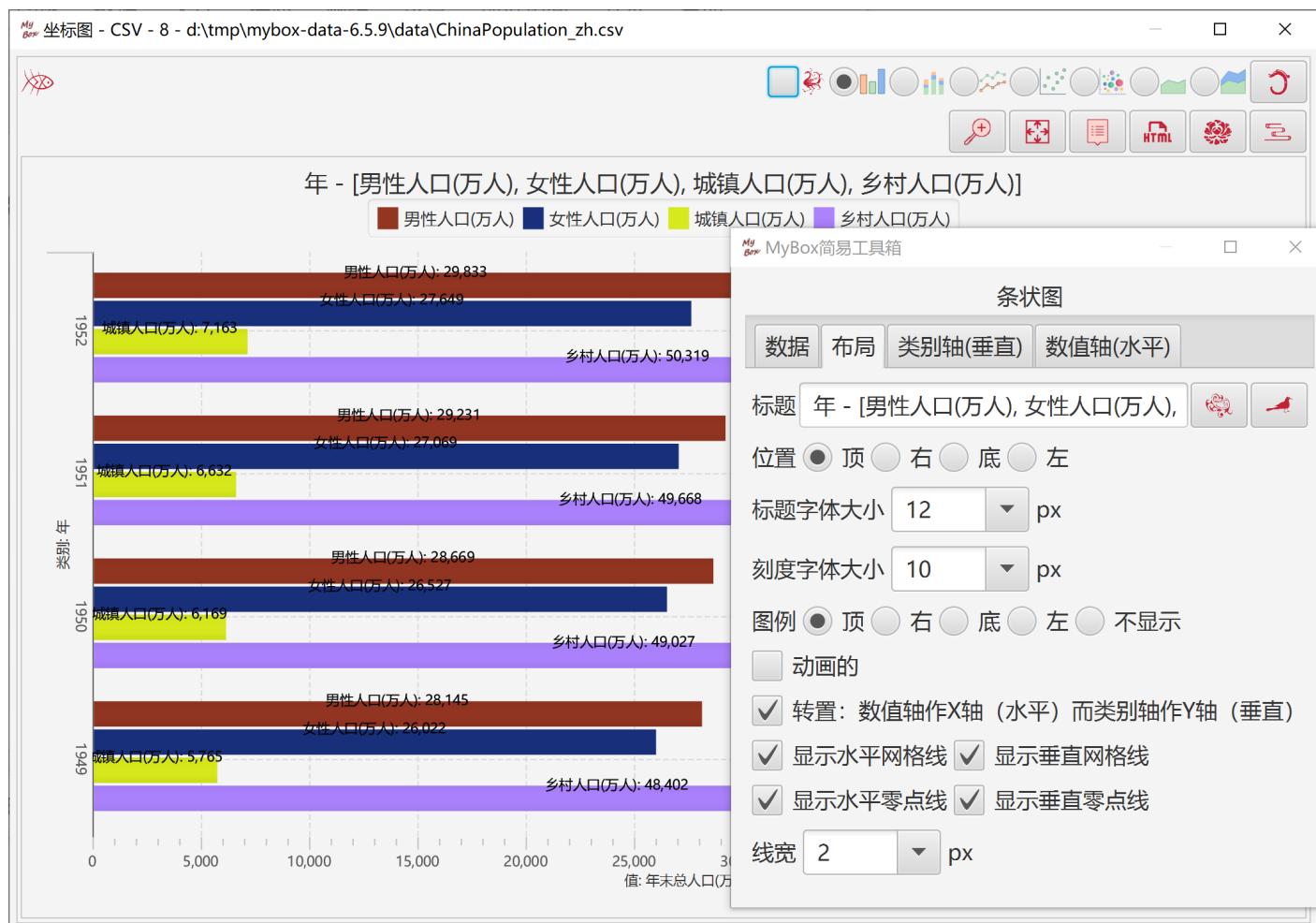
2.14.1.9 图数据的参数

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



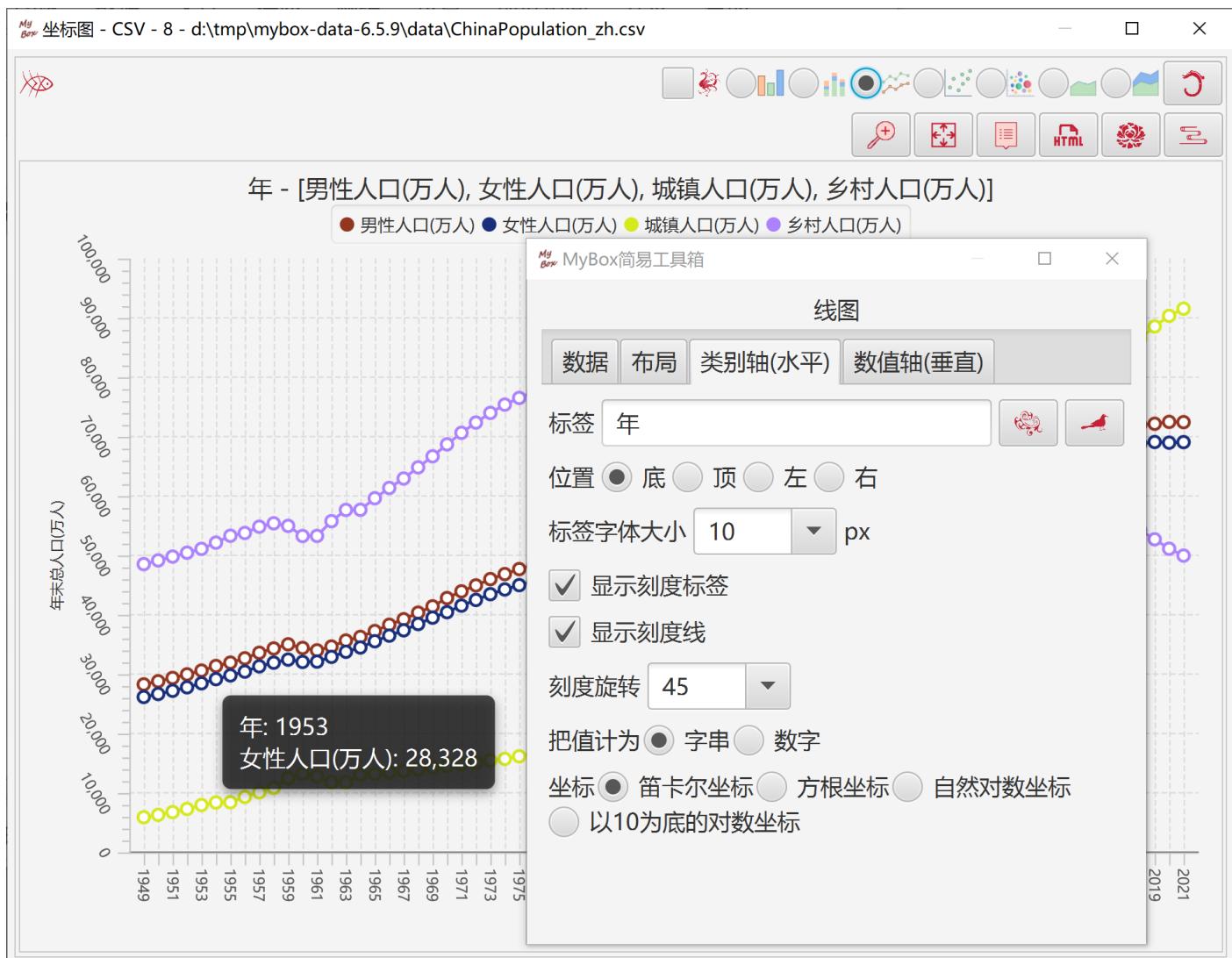
2.14.1.10 布局

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。



2.14.1.11 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。



2.14.1.12 数值轴

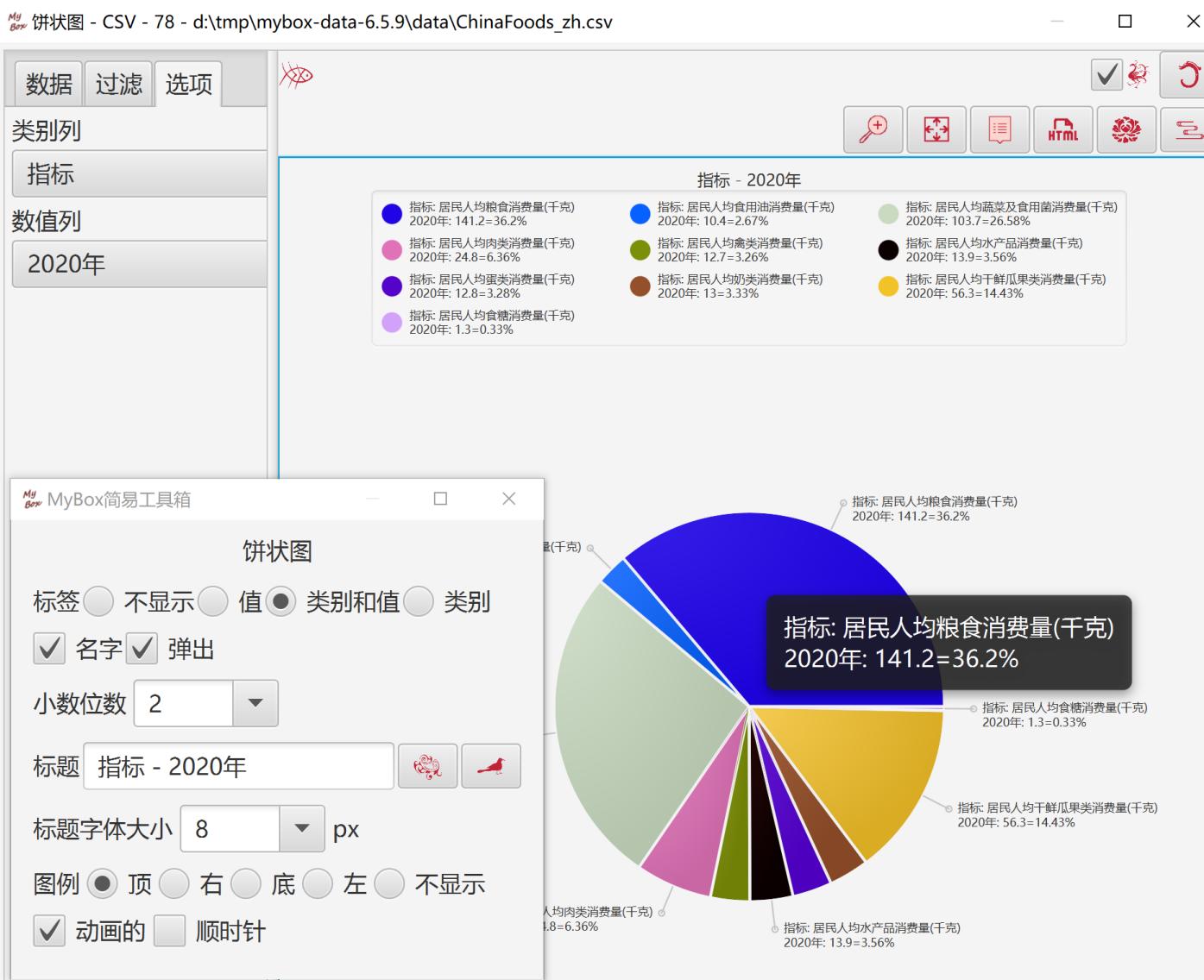
设置数值轴：标签、字体、位置、刻度、坐标等。



2.14.2 饼图

以分割的圆表示数据的比例：

1. 选择数据行。可设置行过滤。
2. 选择一列为“类别”。
3. 选择一列为“数值”，列值必须是非负数。
4. 当处理所有数据行（所有页）时，饼图需要考虑内存限制。
5. 点击“菜单”按钮以设置数据图的参数。
6. 点击“弹出”按钮以在弹出窗口中显示当前图。
7. 点击“数据”按钮以在数据表中显示图的数据。
8. 点击“网页”按钮以在网页中显示图的数据。



2.14.3 箱线图

2.14.3.1 计算方式

箱线图用来展示数据的分布:

1. 数据按照列/行/全部来升序排序。
2. 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)
3. 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 $E_4 = Q_3 + 3 * (Q_3 - Q_1)$
4. 以下项可以用于离散性的参考:
 - 均值 = 数据的平均数

2.14.3.2 数据选项

1. 选择数据、设置行过滤。
2. 选择对象：按照列、行（选择类别列）、所有。
3. 选择要计算的列。
4. 设置排序条件，即最多取排序后的数据个数。
5. 对于非法值：略过、计为零。

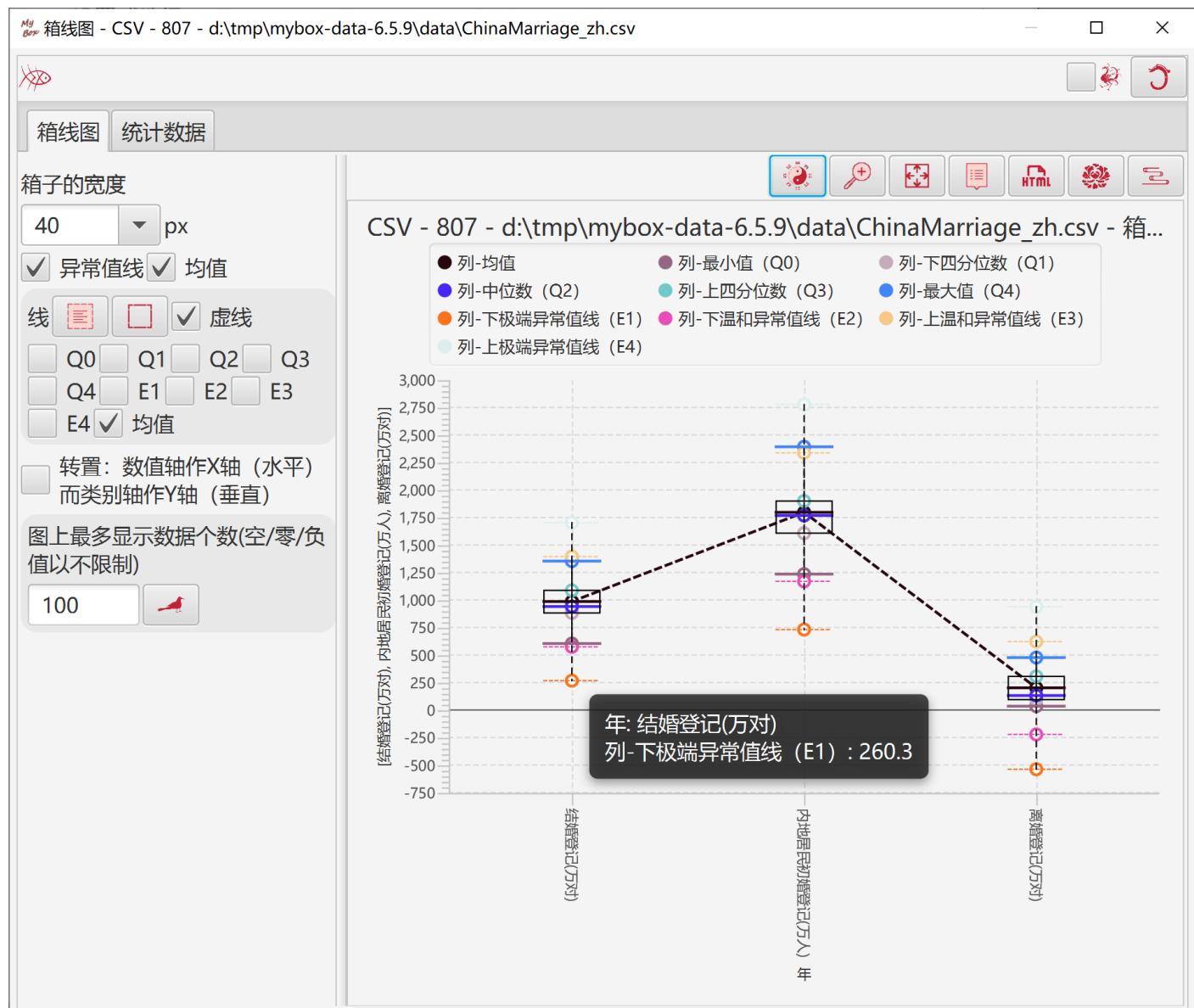
2.14.3.3 统计数据

可以查看和编辑生成的统计数据。

数据行	列名	列-均值	列-最小值...	列-下四分...	列-中位数...	
1	结婚登记(万...	979.6	597.8	875.6	934.1	10
2	内地居民登...	985.69	778.8	880.5	931.84	10
3	离婚登记(万...	195.22	28.5	87.98	125.05	29

2.14.3.4 图选项

1. 箱子的宽度。
2. 是否显示异常值线或均值。
3. 是否显示各类数值的连线、是否虚线。
4. 是否转置
5. 图上最多显示数据个数
6. 随机颜色。
7. 点击“菜单”按钮以设置数据图的参数。
8. 点击“弹出”按钮以在弹出窗口中显示当前图。
9. 点击“数据”按钮以在数据表中显示图的数据。
10. 点击“网页”按钮以在网页中显示图的数据。



2.14.4 自比较条图

2.14.4.1 计算方式

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：

1. 若值等于零，不显示条

2. 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

3. 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

2.14.4.2 数据选项

1. 选择数据、设置行过滤。

2. 选择要计算的列

3. 选择要复制的列。

4. 设置排序条件、及最多取值个数。

5. 设置对非法值的处理。

6. 设置小数位数。

7. 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

2.14.4.3 图选项

1. 选择比较对象：行、列、所有

2. 选择比较方式：绝对值、最大最小区间。

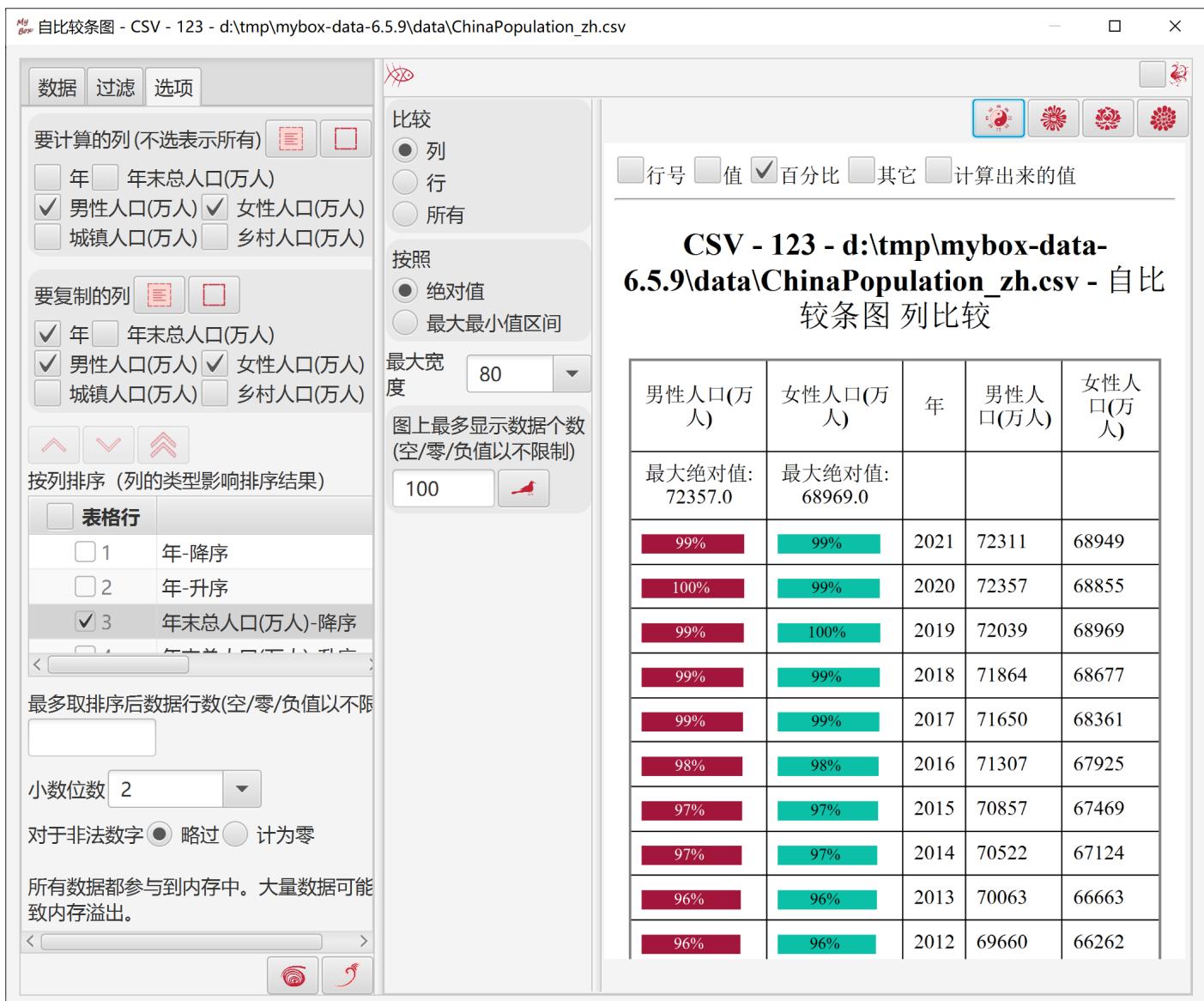
3. 设置条的最大宽度。

4. 设置图上最多显示数据个数。

5. 是否显示行号、值、百分比、类别、计算出来的值。

6. 可以编辑图中数据

7. 可以编辑图的 html



2.14.5 相比较条图

2.14.5.1 计算方式

相比较条图用来对比两类数据。以下规则用来计算颜色条：

1. 若值等于零，不显示条

2. 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

3. 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

2.14.5.2 数据选项

1. 选择数据、设置行过滤。
2. 选择类别列（非必要）。
3. 选择两个要比较的数值列。
4. 选择要复制的列。
5. 设置排序条件、及最多取值个数。
6. 设置小数位数。
7. 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

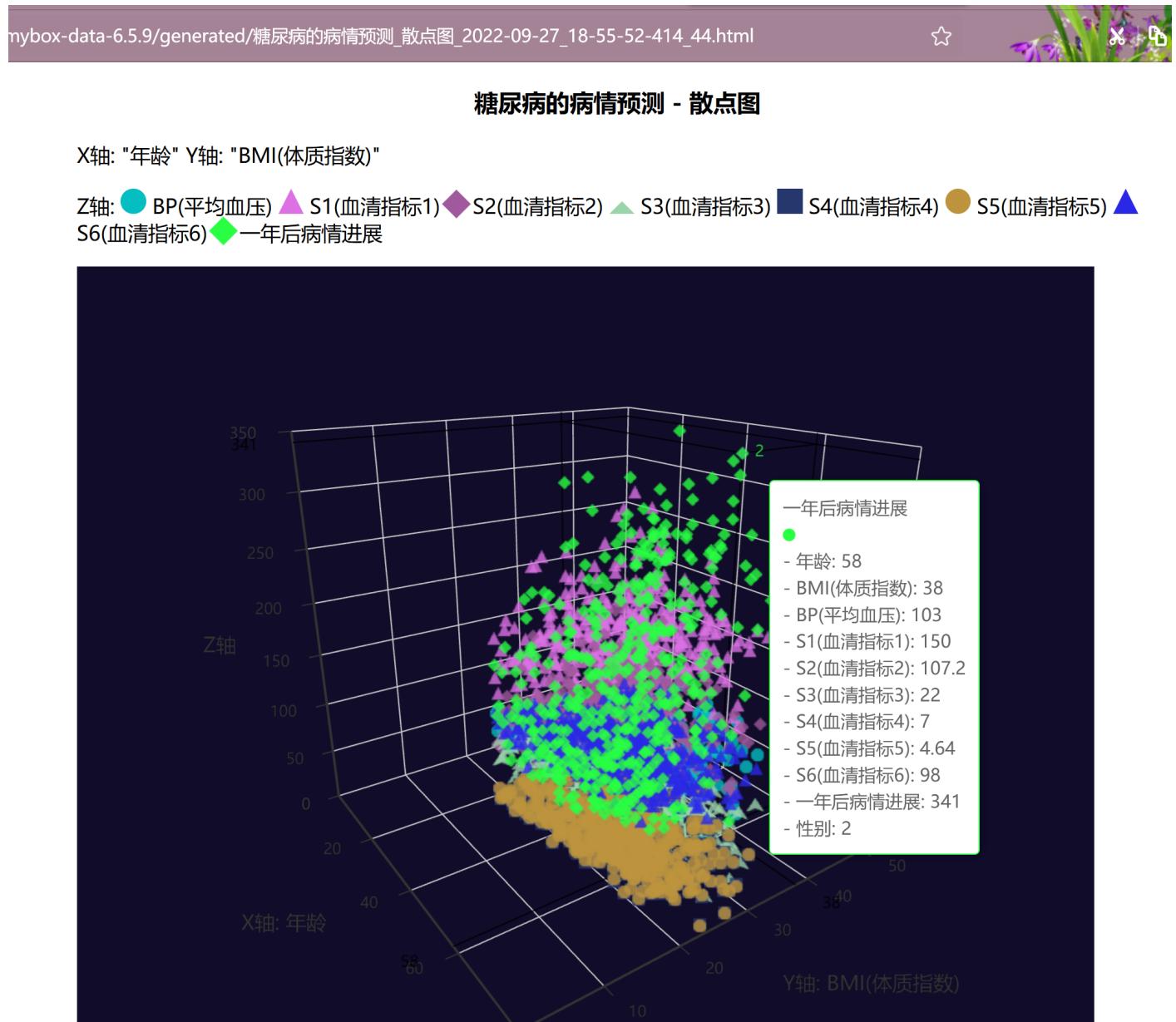
2.14.5.3 图选项

1. 选择比较方式：绝对值、最大最小区间。
2. 设置条的最大宽度。
3. 设置图上最多显示数据个数。
4. 是否显示行号、值、百分比、类别、计算出来的值。
5. 可以编辑图中数据
6. 可以编辑图的 html



2.14.6 XYZ 图

1. 基于 [echarts-gl](#), 借助 WebGL 技术在网页中显示三维图。
2. 由于 JavaFX 基于的 Webview 不支持 WebGL, 所以生成的网页将在系统浏览器中显示。
3. 请确认你的系统浏览器支持 WebGL 且未禁止本地 JavaScript 文件。



2.14.6.1 三维散点图

1. 选择数据行。可以设置行过滤。
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。
3. 选择其它要在标签中弹出的值
4. 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。



2.14.6.2 曲面图

1. 选择数据行，可以设置行过滤。
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。
3. 选择其它要弹出的值
4. 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。

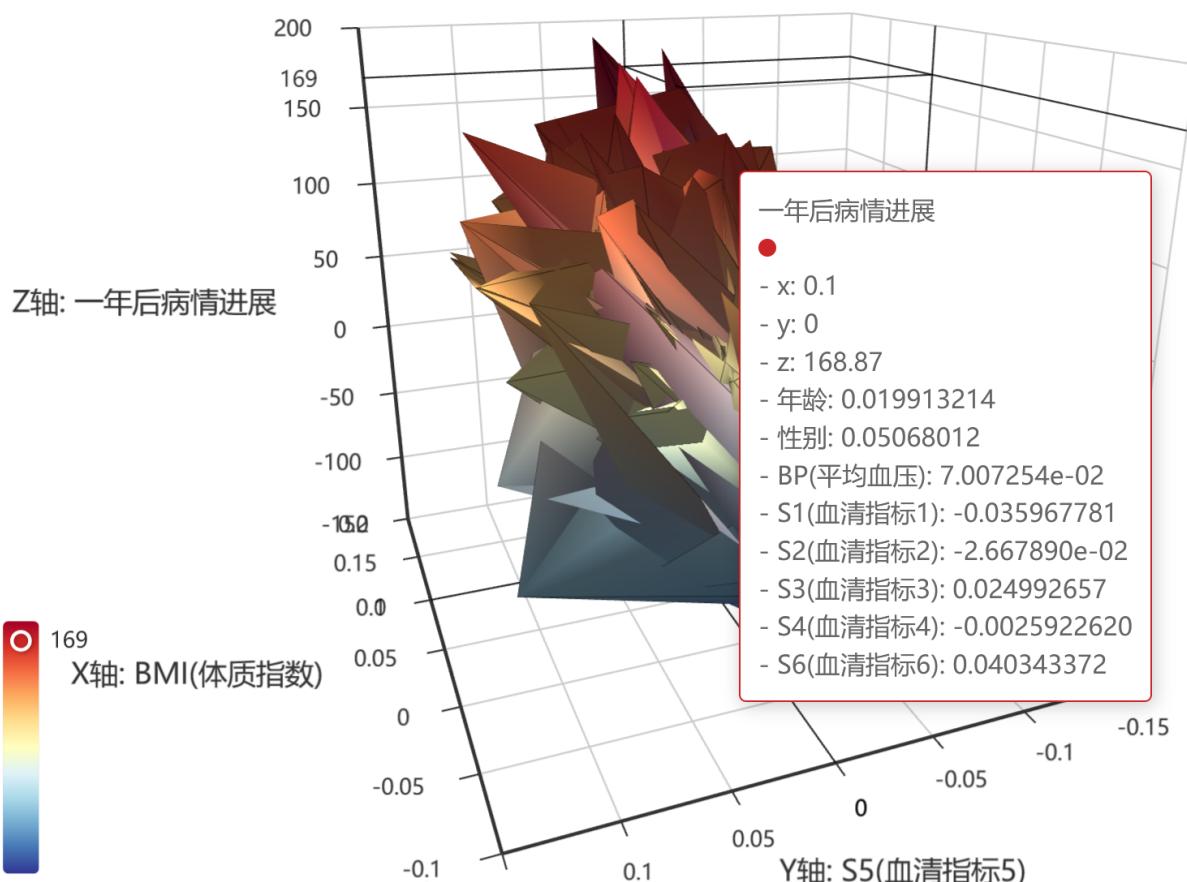


mybox-data-6.5.9/generated/糖尿病的病情预测-标准化_曲面图_2022-09-27_19-05-19-913_30.html



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"



2.14.7 位置分布图

把位置数据动态显示在地图上。

2.14.7.1 数据选项

1. 选择数据行。可以设置行过滤。
2. 各选一列为：标签、经度、纬度、大小。
3. 选择其它要在弹出信息中显示的列。
4. 选择坐标系统。
5. 设置排序条件、最多取值个数。

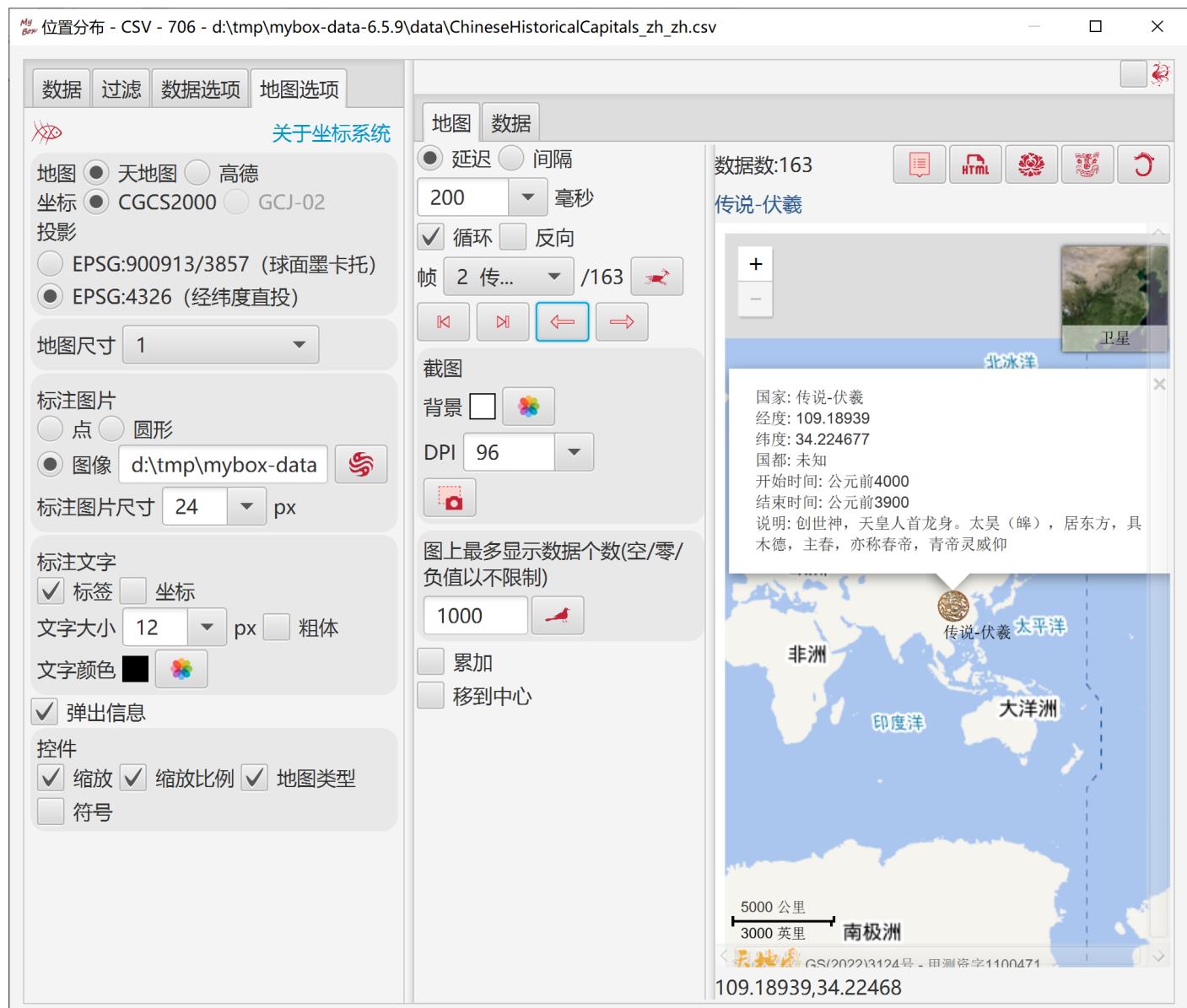
2.14.7.2 地图数据

地图中引用的数据被显示在表中。

数据行	国家	经度	纬度	国者+
1	传说-燧人	112.55699	34.87973	未知
2	传说-伏羲	109.18939	34.224677	未知
3	传说-女娲	111.107528	35.126412	未知
4	传说-神农/炎帝	113.772655	26.489902	未知
5	传说-黄帝	113.730529	34.395562	未知
6	传说-少昊 (皞)	118.822224	34.30994	穷桑 (江苏沐阳)
7	传说-颛顼	115.071879	34.445656	帝丘 (河南商丘)
8	传说-帝喾	115.65637	34.414172	毫 (河南商丘)
9	传说-尧	111.518976	36.088005	唐 (山西临汾)
10	传说-舜	110.329194	34.831075	蒲坂 (山西永济)
11	夏	111.220456	35.141363	安邑 (山西运城)
12	夏	113.050492	34.453667	阳城 (河南郑州)
13	夏	113.488478	34.140701	阳翟 (河南禹州)
14	夏	111.007528	35.026412	安邑 (山西运城)
15	夏	112.693124	34.69775	斟鄩 (河南偃师)
16	夏-后羿、寒浞	112.789612	34.90797	鉏/穷石 (河南唐河)
17	夏	114.28185	34.8138	老丘 (河南开封)
18	夏	114.392392	36.097577	西河 (河南安阳)
19	商	115.65537	34.454172	亳 (河南商丘市)

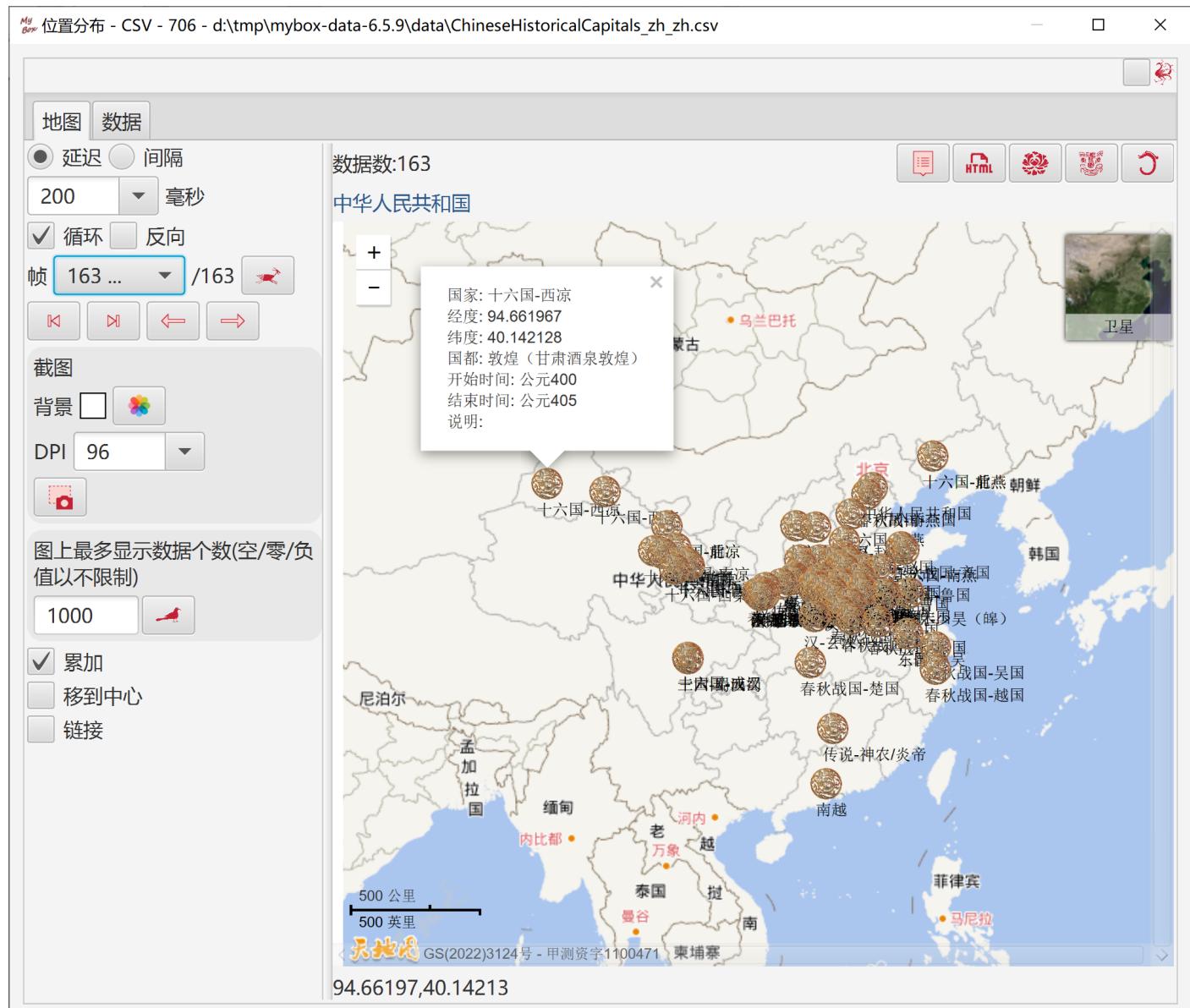
2.14.7.3 地图选项

1. 地图类型：天地图、高德。
2. 选择投影、地图尺寸，
3. 设置标注图片、标注文字。
4. 选择是否弹出信息。
5. 选择地图控件。



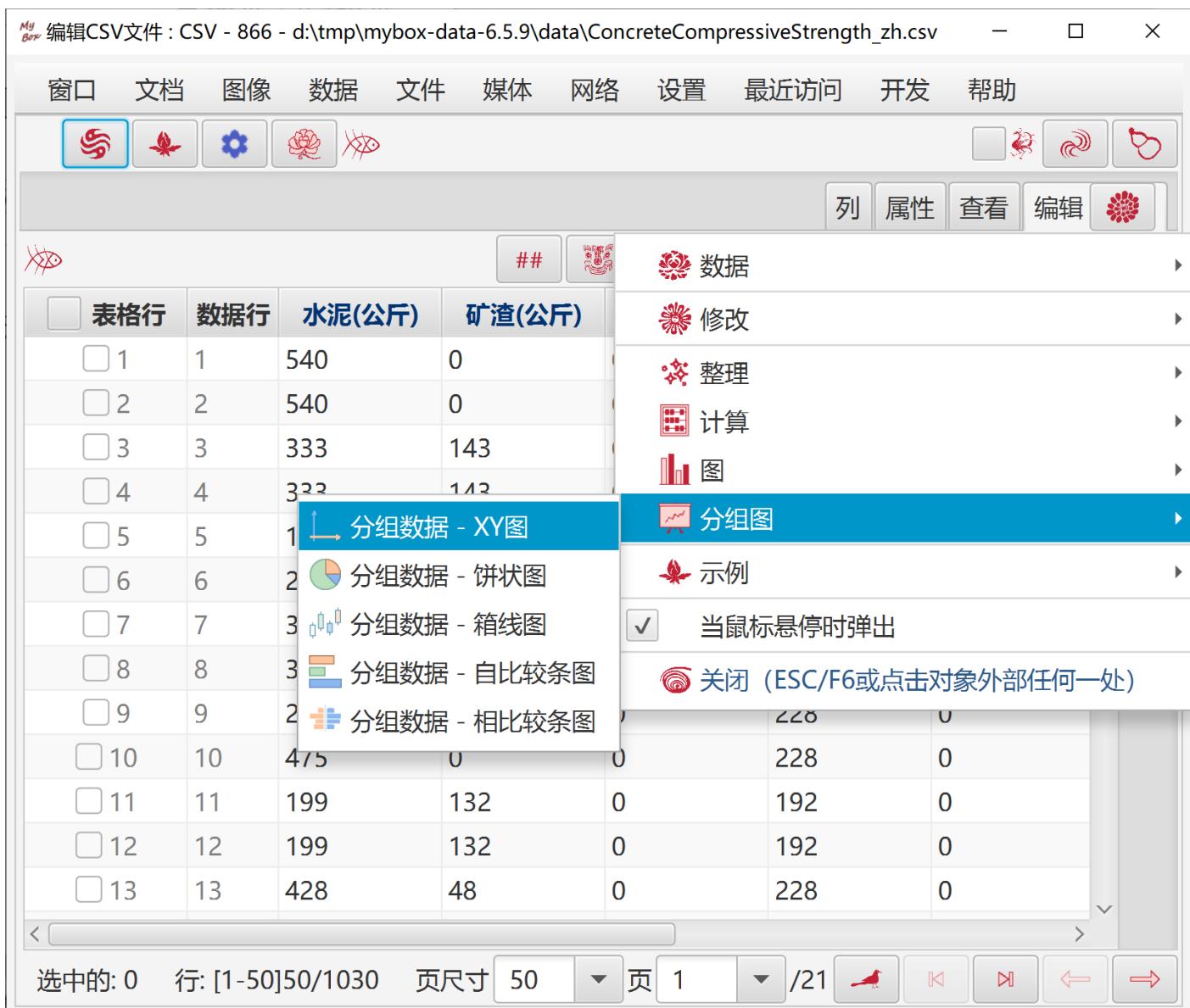
2.14.7.4 播放选项

1. 选择定时类型:
 - 延迟: 显示本帧之后, 等待指定的时间, 再显示下一帧。
 - 间隔: 本帧开始显示的时间与下一帧开始显示的时间之间, 总是相差指定的时间。
2. 选择是否: 循环、反向、累加、移到中心、链接。
3. 自动播放。
4. 指定一帧显示。
5. 导航按钮。
6. 截图:
 - 选择背景色、像素密度
 - 自动逐帧截图, 然后显示在“编辑图像列表”的界面中以进一步处理。
7. 设置图上最多显示数据个数。



2.15 分组图

悬停或点击功能按钮，选择“分组图”下的功能。

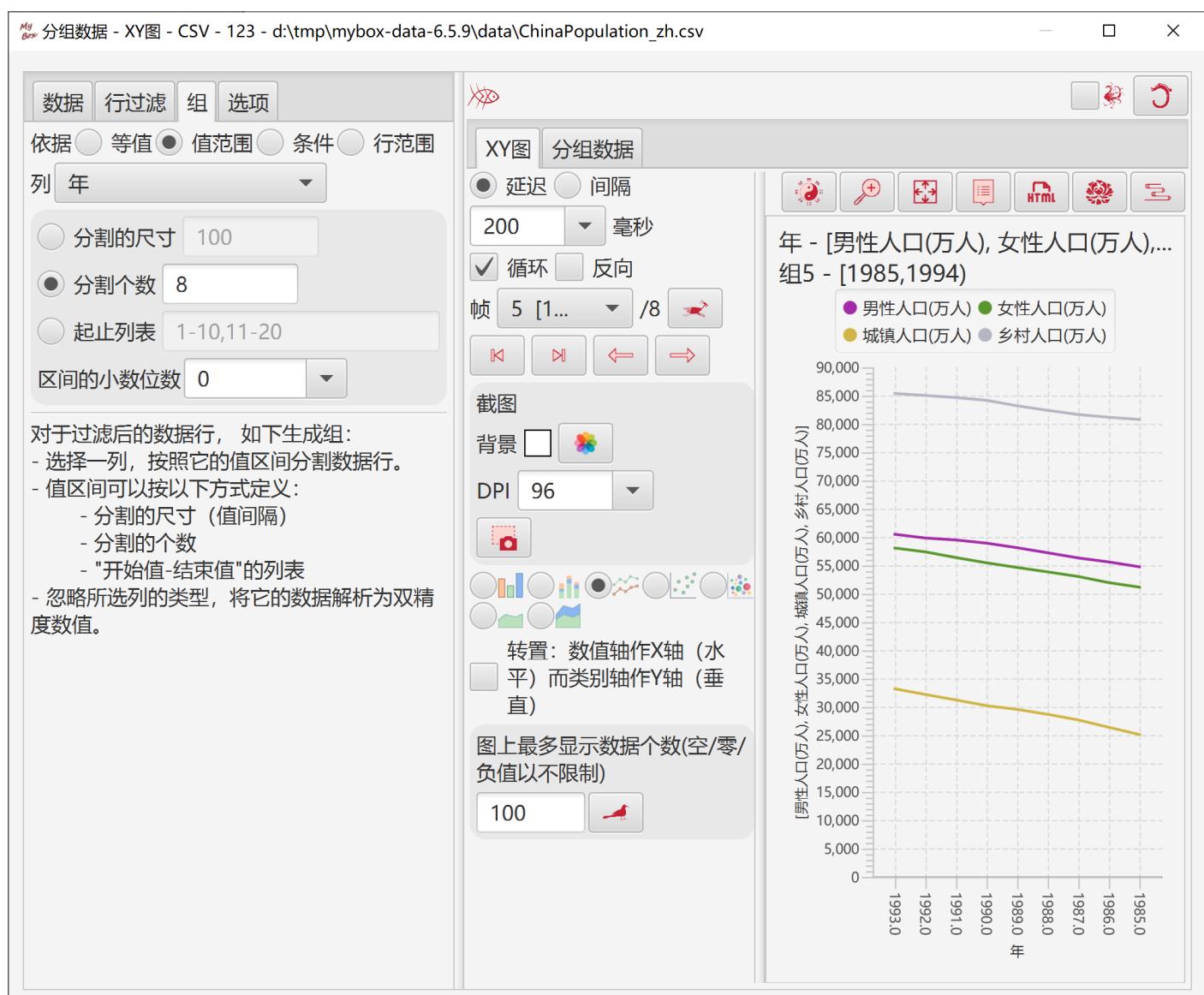


2.15.1 通用处理

数据分组以后，对每组数据都生成一张图，然后依序动态显示这些图。

2.15.1.1 数据分组

1. 选择数据行。可以设置行过滤。
2. 设置分组条件。
3. 设置各种图的特定生成选项。
4. 设置排序条件、最多取值个数。
5. 设置小数位数、对非法值的处理。



2.15.1.2 分组结果

计算后的分组数据被显示在表中。

MyBox 分组数据 - XY图 - CSV - 123 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv

数据 行过滤 组 选项
XY图 分组数据

数据行	id	组	区间_年	年	男性...	女性人...	城镇人...	乡村人口...
41	41	5	[1985,1994)	1989	58099	54605	29540	83164
42	42	5	[1985,1994)	1988	57201	53825	28661	82365
43	43	5	[1985,1994)	1987	56290	53010	27674	81626
44	44	5	[1985,1994)	1986	55581	51926	26366	81141
45	45	5	[1985,1994)	1985	54725	51126	25094	80757
46	46	6	[1994,2003)	2002	66115	62338	50212	78241
47	47	6	[1994,2003)	2001	65672	61955	48064	79563
48	48	6	[1994,2003)	2000	65437	61306	45906	80837
49	49	6	[1994,2003)	1999	64692	61094	43748	82038
50	50	6	[1994,2003)	1998	63940	60821	41608	83153
51	51	6	[1994,2003)	1997	63131	60495	39449	84177
52	52	6	[1994,2003)	1996	62200	60189	37304	85085
53	53	6	[1994,2003)	1995	61808	59313	35174	85947
54	54	6	[1994,2003)	1994	61246	58604	34169	85681
55	55	7	[2003,2012)	2011	69161	65755	69927	64989
56	56	7	[2003,2012)	2010	68748	65343	66978	67113
57	57	7	[2003,2012)	2009	68647	64803	64512	68938

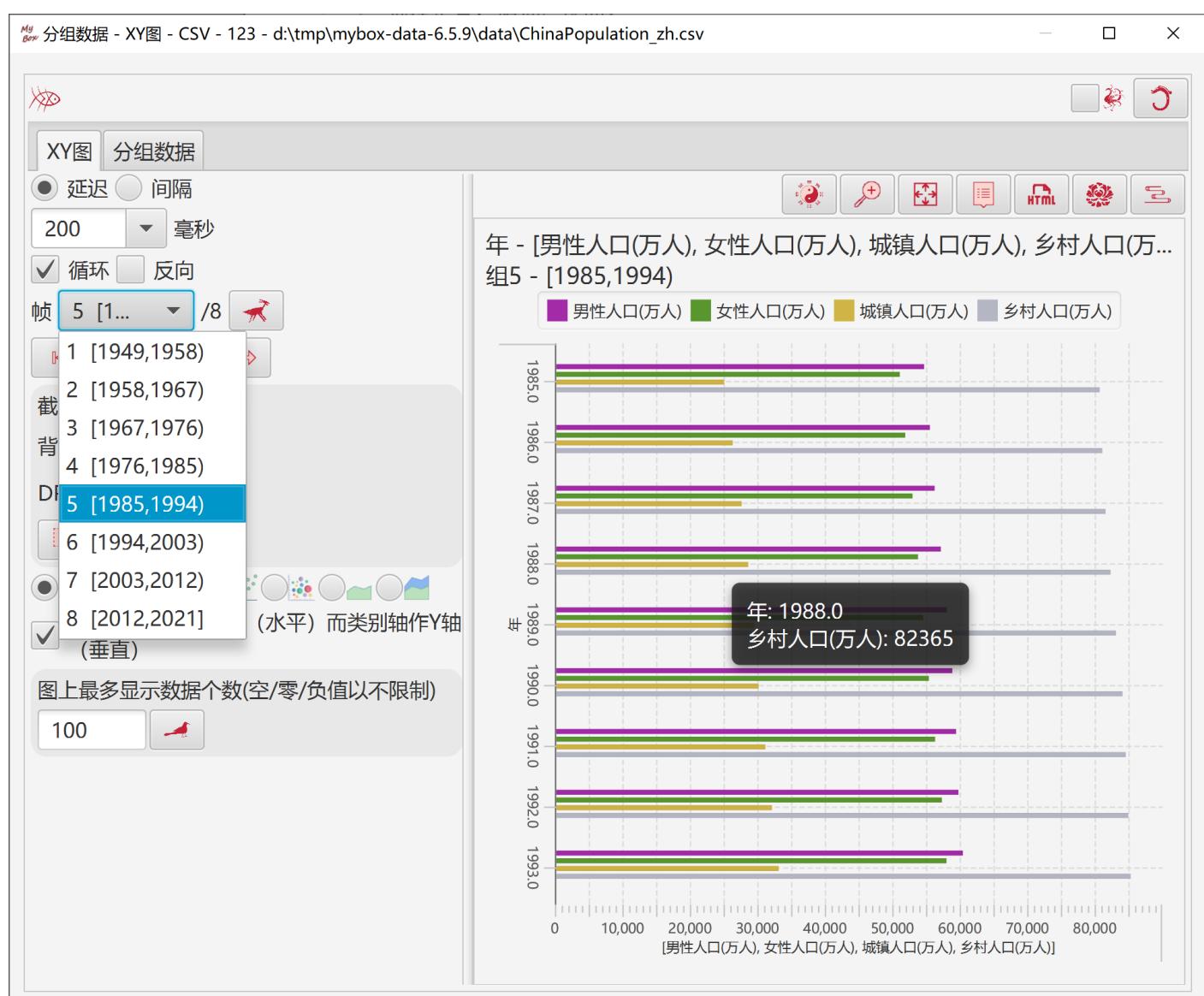
选中的: 0 行: [41-60]20/73 页尺寸 20 页 3 /4

2.15.1.3 图的显示选项

1. 各种图的特定显示选项。
2. 设置图上最多显示个数

2.15.1.4 播放选项

1. 选择定时类型:
 - 延迟：显示本帧之后，等待指定的时间，再显示下一帧。
 - 间隔：本帧开始显示的时间与下一帧开始显示的时间之间，总是相差指定的时间。
2. 选择是否：循环、反向。
3. 自动播放、指定一帧显示、导航按钮。
4. 截图：选择背景色、像素密度，然后自动逐帧截图。



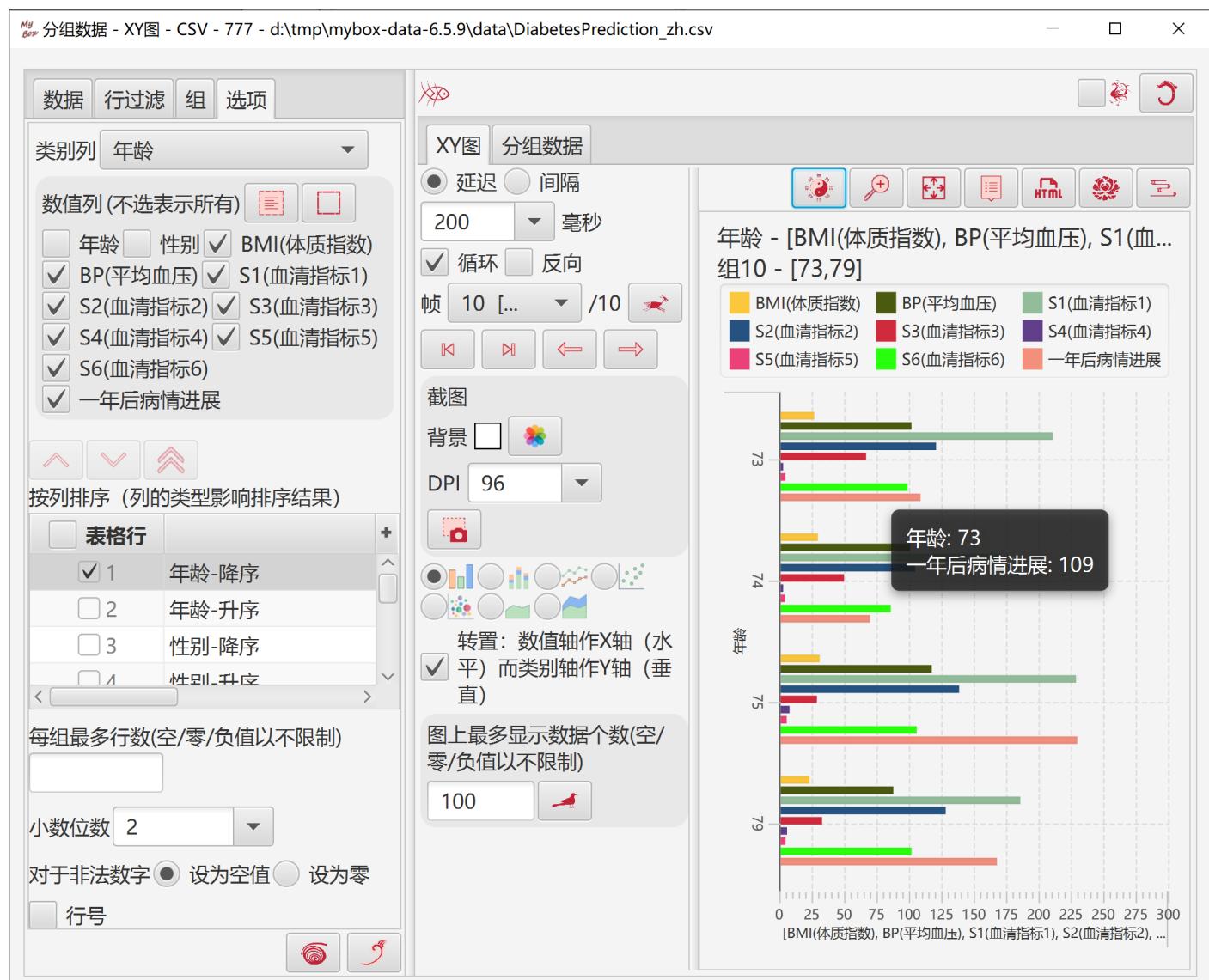
2.15.2 分组数据 - XY 图

2.15.2.1 图的特定生成选项

1. 选择一列为类别轴
2. 选择若干列为数值轴。

2.15.2.2 图的特定显示选项

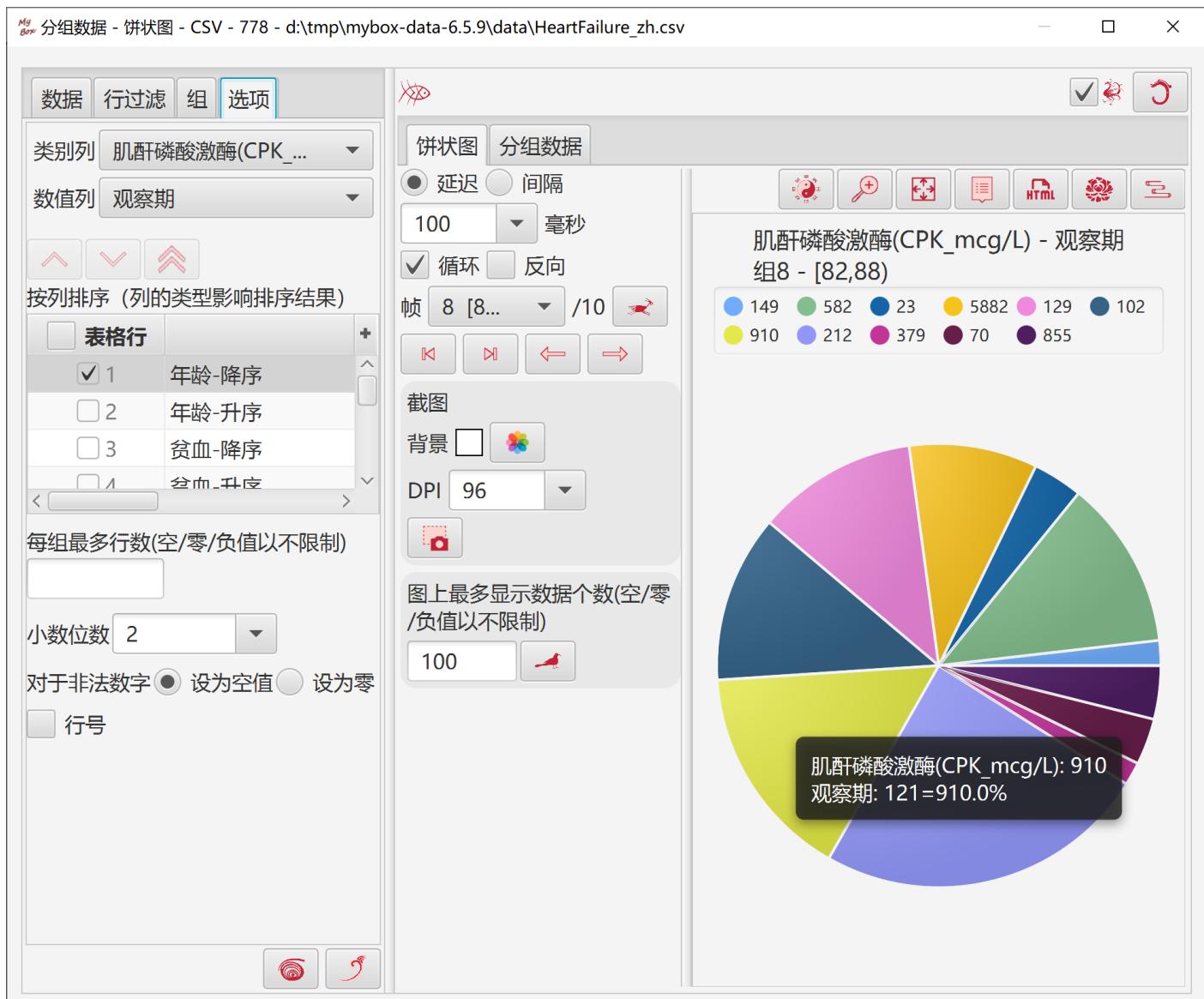
1. XY 图的类型。
2. 是否转置。



2.15.3 分组数据 – 饼图

2.15.3.1 图的特定生成选项

1. 选择一列为类别轴
2. 选择一列为数值轴。



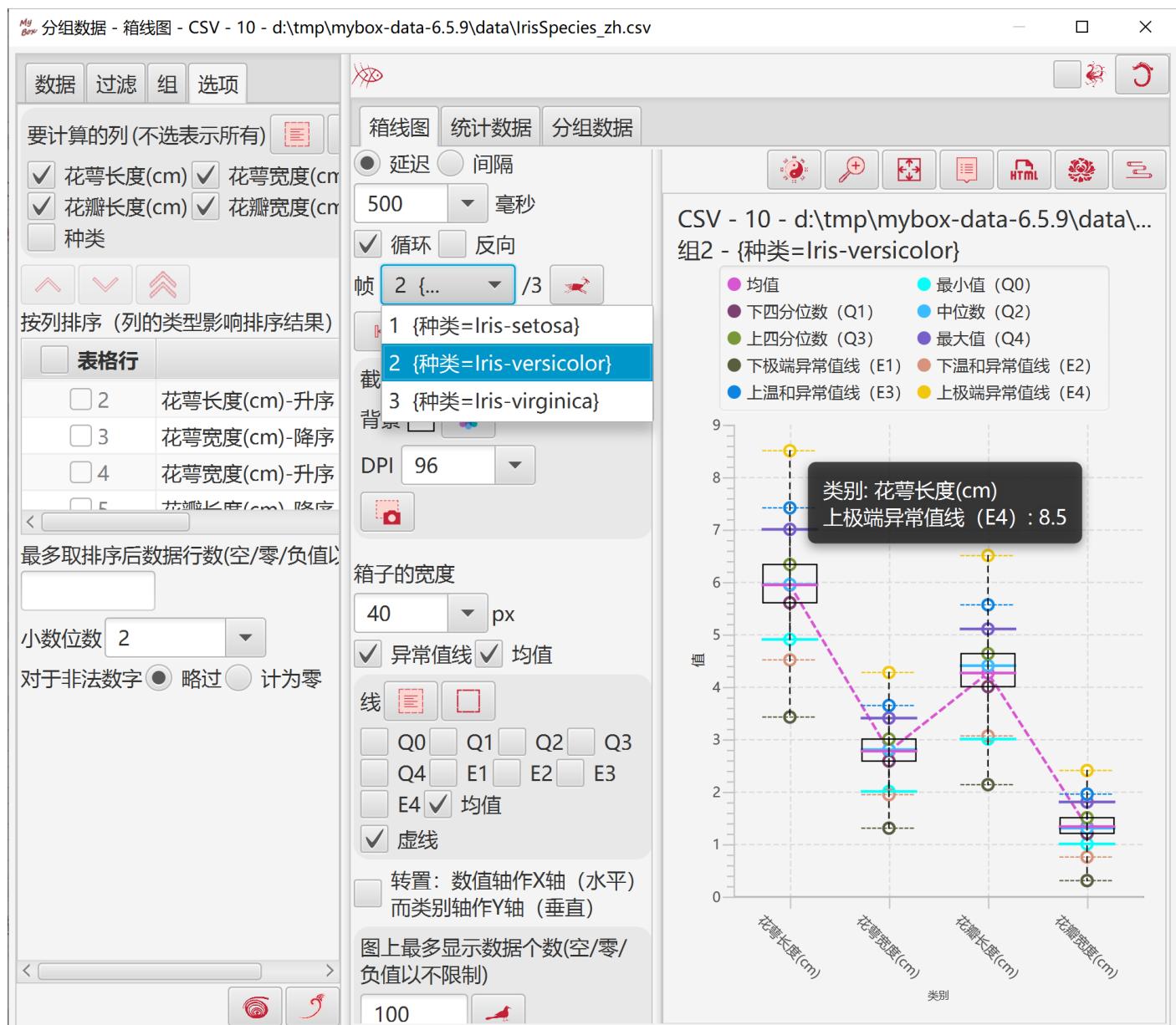
2.15.4 分组数据 - 箱线图

2.15.4.1 图的特定生成选项

- 选择若干要计算的列。

2.15.4.2 图的特定显示选项

- 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。
- 是否转置
- 图上最多显示数据个数



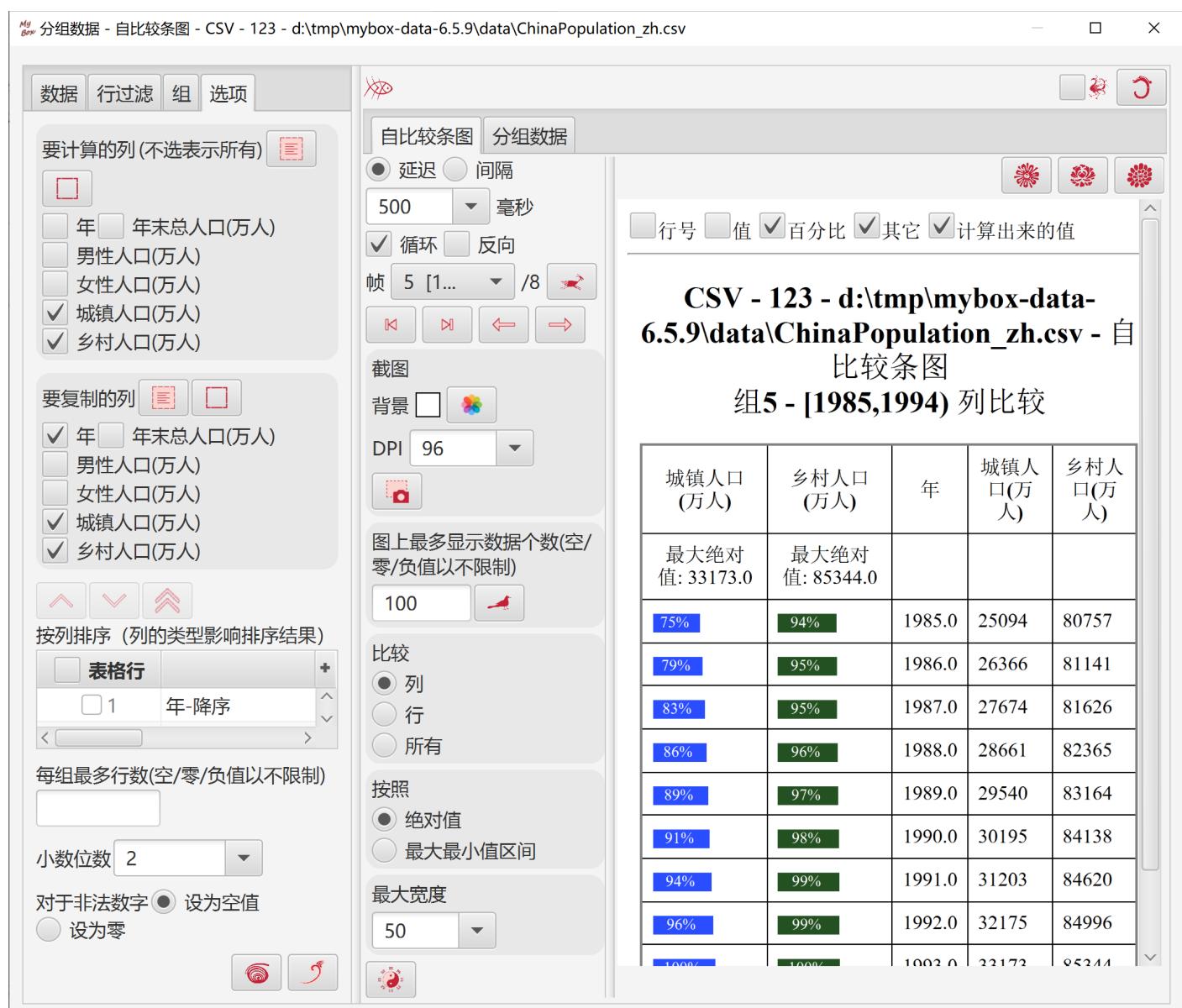
2.15.5 分组数据 - 自比较条图

2.15.5.1 图的特定生成选项

1. 选择若干要计算的列。
1. 选择若干要复制的列。

2.15.5.2 图的特定显示选项

1. 比较对象: 行、列、所有
2. 比较方式: 绝对值、最大最小区间。
3. 条的最大宽度。
4. 是否显示行号、值、百分比、类别、计算出来的值。



2.15.6 分组数据 – 相比较条图

2.15.6.1 图的特定生成选项

1. 选择类别列（非必要）。
2. 选择两个要比较的数值列。
3. 选择要复制的列。

2.15.6.2 图的特定显示选项

1. 比较方式：绝对值、最大最小区间。
2. 条的最大宽度。
3. 是否显示行号、值、百分比、类别、计算出来的值

分组数据 - 相比较条图 - CSV - 123 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv

**CSV - 123 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv - 相比较条图
组5 - [1985,1994]**

最大绝对值: 85344.0

城镇人口(万人)	年	乡村人口(万人)	年	城镇人口(万人)	乡村人口(万人)
29%	1985.0	94%	1985.0	25094	80757
30%	1986.0	95%	1986.0	26366	81141
32%	1987.0	95%	1987.0	27674	81626
33%	1988.0	96%	1988.0	28661	82365
34%	1989.0	97%	1989.0	29540	83164
35%	1990.0	98%	1990.0	30195	84138
36%	1991.0	99%	1991.0	31203	84620
37%	1992.0	99%	1992.0	32175	84996
38%	1993.0	100%	1993.0	33173	85344

2.16 输入数据

2.16.1 加载系统粘贴板的内容

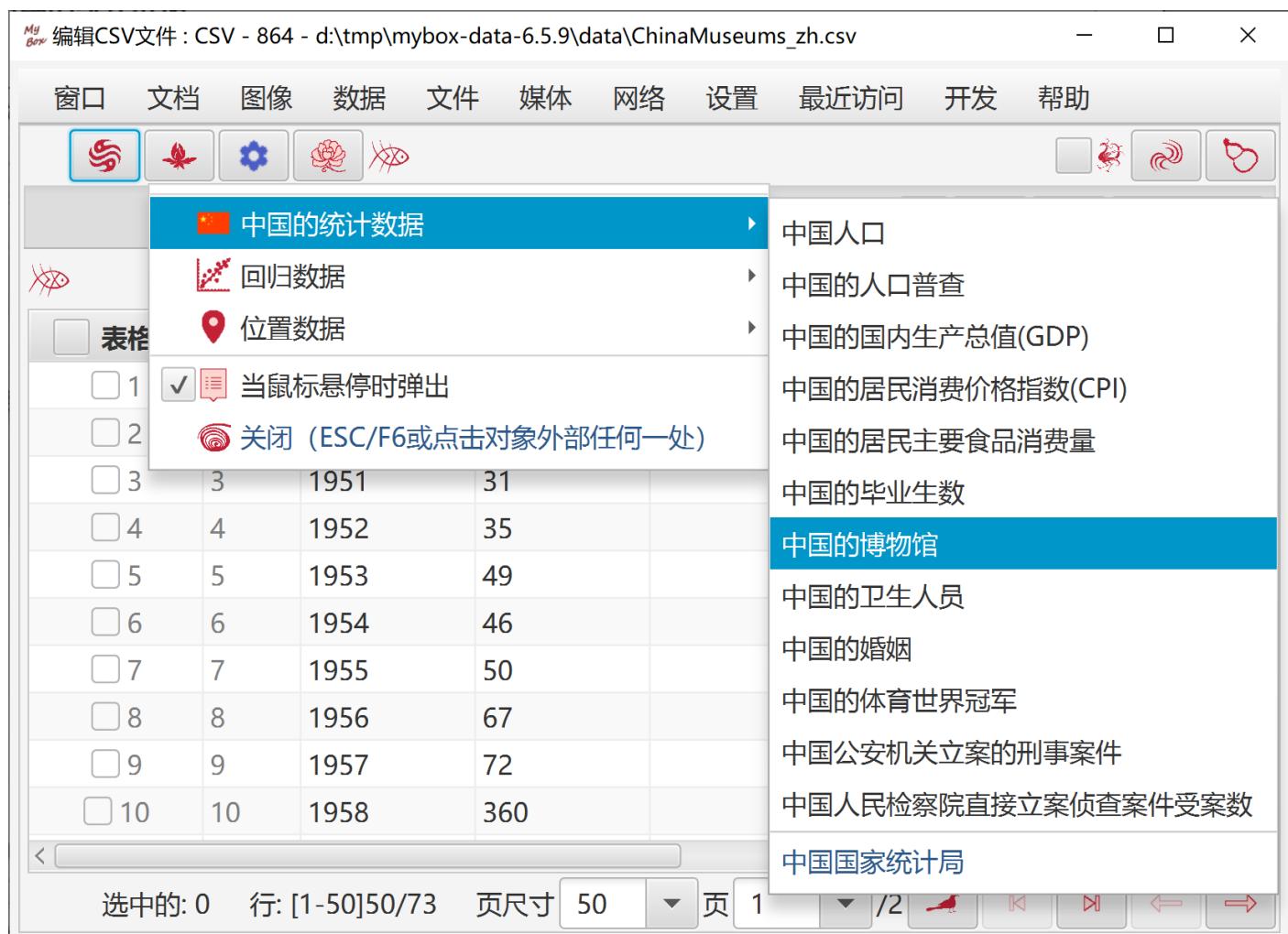
1. 读取系统粘贴板中的内容。
 2. 工具猜测分隔符并解析数据。
 3. 用户可以指定分隔符来解析。
 4. 可选解析方式：
 - CSV 解析：数值可包含分隔符和换行符，不支持空白字符解析和规则表达式解析。
 - 文本解析：数值不可包含分隔符和换行符，支持空白字符解析和规则表达式解析。
 5. 可选是否是第一行定义列名。
 6. 对解析后的数据，选择要加载的内容。可以设置行过滤。

2.16.2 导入示例

悬停或点击按钮“示例”，则弹出示例的菜单。

2.16.2.1 中国的统计数据

来自中国国家统计局的数据。



2.16.2.2 回归相关的数据

可用于回归分析的数据。

The screenshot shows the MyBox Data Tool interface with a menu bar and various tool icons. A context menu is open over a table, listing several regression datasets. The dataset '混凝土抗压强度' (Concrete Compressive Strength) is selected. The table below contains 10 rows of data.

1	3	333	143	0
2	4	333	143	0
3	5	199	132	0
4	6	266	114	0
5	7	380	95	0
6	8	380	95	0
7	9	266	114	0
8	10	475	0	0
9				
10				

选中的: 0 行: [1-50] 50/1030 页尺寸 50 页 1 /21

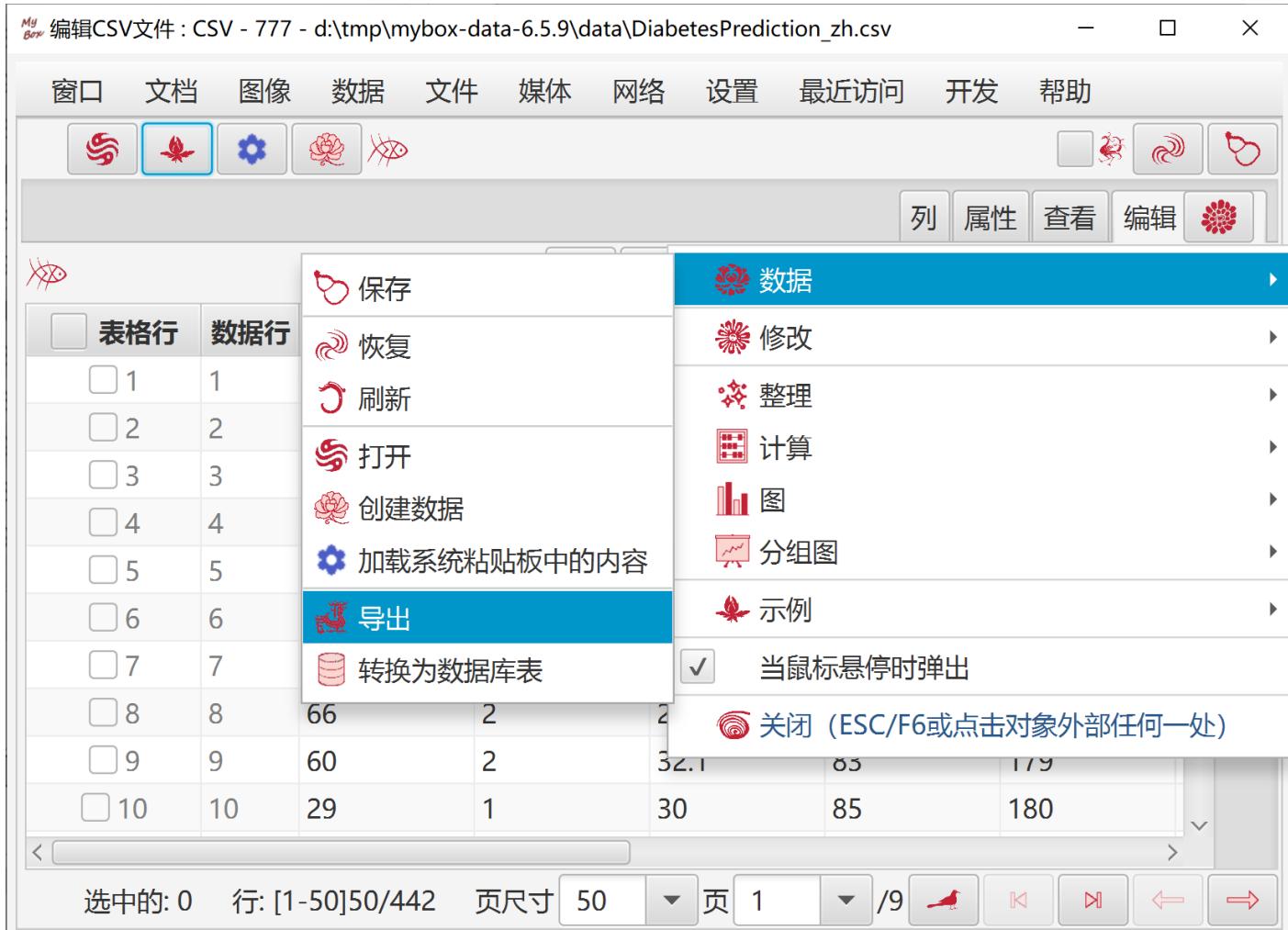
2.16.2.3 位置数据

可用于生成位置分布的数据。

	1	2	3	4	5	6	7	8	9	10	11	12	13	
<input type="checkbox"/>	1	<input checked="" type="checkbox"/> 当鼠标悬停时弹出	当鼠标悬停时弹出	<input type="checkbox"/> 2	<input checked="" type="checkbox"/> 关闭 (ESC/F6或点击对象外部任何...	关闭 (ESC/F6或点击对象外部任何...	<input type="checkbox"/> 3	3	传说-女娲	未知	<input type="checkbox"/> 4	4	传说-神农/炎帝	未知
<input type="checkbox"/>	5	5	传说-黄帝	未知	<input type="checkbox"/> 6	6	传说-少昊 (皞)	穷桑 (江苏沭阳桑墟镇)	<input type="checkbox"/> 7	7	帝丘 (河南商丘睢县帝丘)	113.730529	34.3955	
<input type="checkbox"/>	8	8	传说-颛顼	帝丘 (河南商丘)	<input type="checkbox"/> 9	9	传说-帝喾	毫 (河南商丘)	<input type="checkbox"/> 10	10	传说-尧	111.518976	36.0880	
<input type="checkbox"/>	11	11	传说-舜	唐 (山西临汾)	<input type="checkbox"/> 12	12	夏	蒲坂 (山西永济蒲州)	<input type="checkbox"/> 13	13	安邑 (山西运城市夏县)	110.329194	34.8310	
<input type="checkbox"/>	12	12	夏	阳城 (河南郑州登封)	<input type="checkbox"/> 13	13	夏	阳翟 (河南禹州)	<input type="checkbox"/>		阳翟 (河南禹州)	113.050492	34.4536	
<input type="checkbox"/>	13	13			<input type="checkbox"/>				<input type="checkbox"/>			113.488478	34.1407	

2.17 输出数据

悬停或点击功能按钮，选择“数据”下的功能。



2.17.1 导出

2.17.1.1 数据源

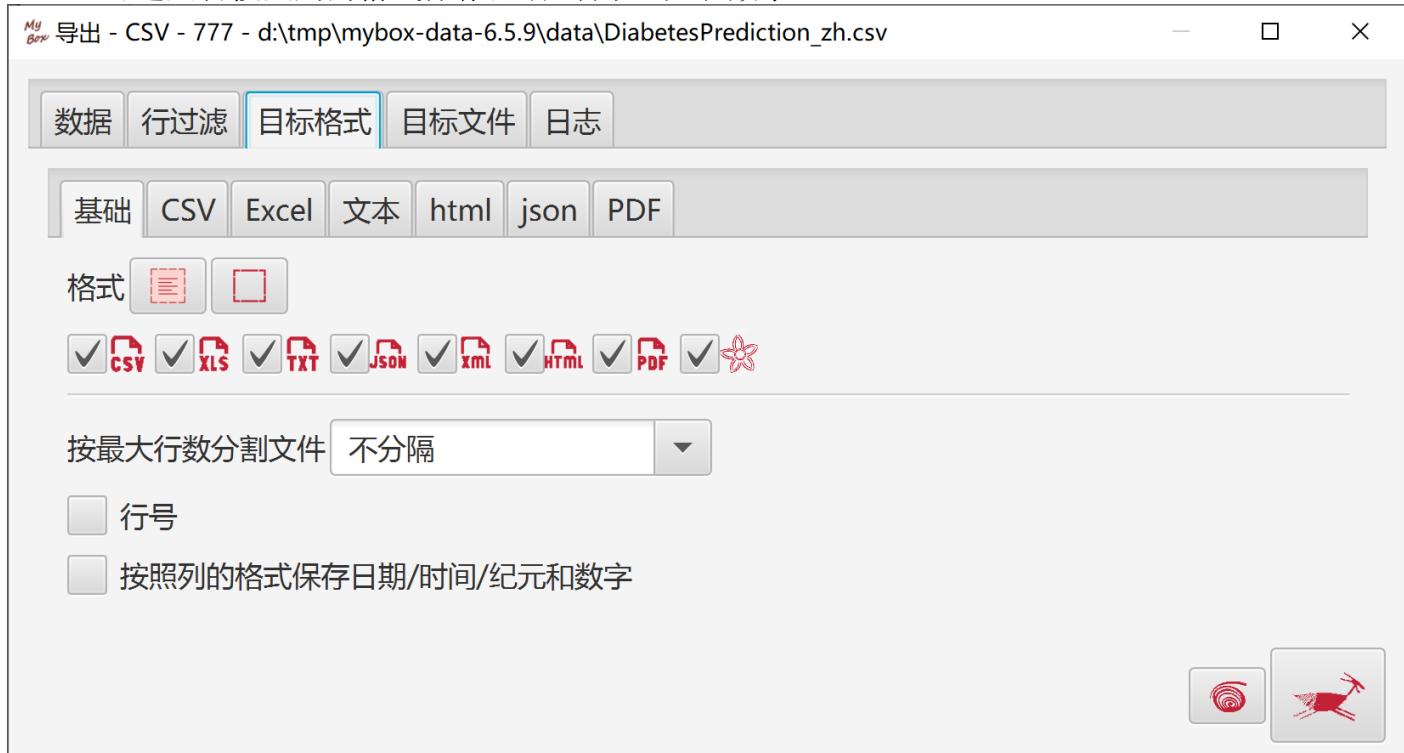
1. 选择数据的行和列。
 2. 可以设置行过滤。

MyBox 导出 - CSV - 777 - d:\tmp\mybox-data-6.5.9\data\DiabetesPrediction_zh.csv

	数据行	数据行	年龄	性别	BMI(体重)	BP(平均收缩压)	S1(血糖)	+
<input type="checkbox"/>	1	1	59	2	32.1	101	157	93.2
<input checked="" type="checkbox"/>	2	2	48	1	21.6	87	183	103.
<input checked="" type="checkbox"/>	3	3	72	2	30.5	93	156	93.6
<input checked="" type="checkbox"/>	4	4	24	1	25.3	84	198	131.
<input checked="" type="checkbox"/>	5	5	50	1	23	101	192	125.
<input type="checkbox"/>	6	6	23	1	22.6	89	139	64.8
<input type="checkbox"/>	7	7	36	2	22	90	160	99.6
<input checked="" type="checkbox"/>	8	8	66	2	26.2	114	255	185
<input checked="" type="checkbox"/>	9	9	60	2	32.1	83	179	119.
<input type="checkbox"/>	10	10	29	1	30	85	180	93.4
<input checked="" type="checkbox"/>	11	11	22	1	18.6	97	114	57.6
<input type="checkbox"/>	12	12	56	2	28	85	184	144.
<input type="checkbox"/>	13	13	53	1	23.7	92	186	109.

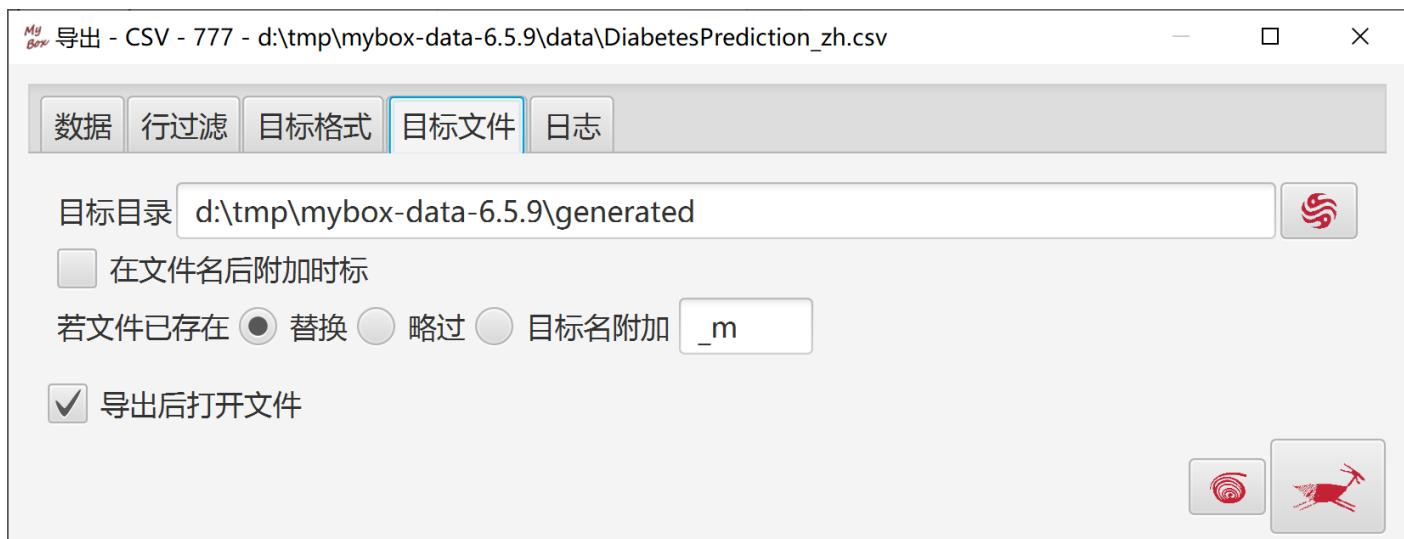
2.17.1.2 目标格式

1. 可选目标文件格式: csv、文本、excel、xml、json、html、pdf、MyBox 粘贴板。
2. 可以设置目标文件的选项。
3. 可按最大行数分割导出后的文件。
4. 可选是否包含行号。
5. 可选是否按照列的格式保存日期/时间/纪元和数字。



2.17.1.3 目标文件

1. 设置目标目录。
2. 选择是否在导出后打开文件。



2.17.2 转换为数据库表

1. 选择数据。可以设置行过滤。
2. 选择要复制的列
3. 选择是否导入数据。



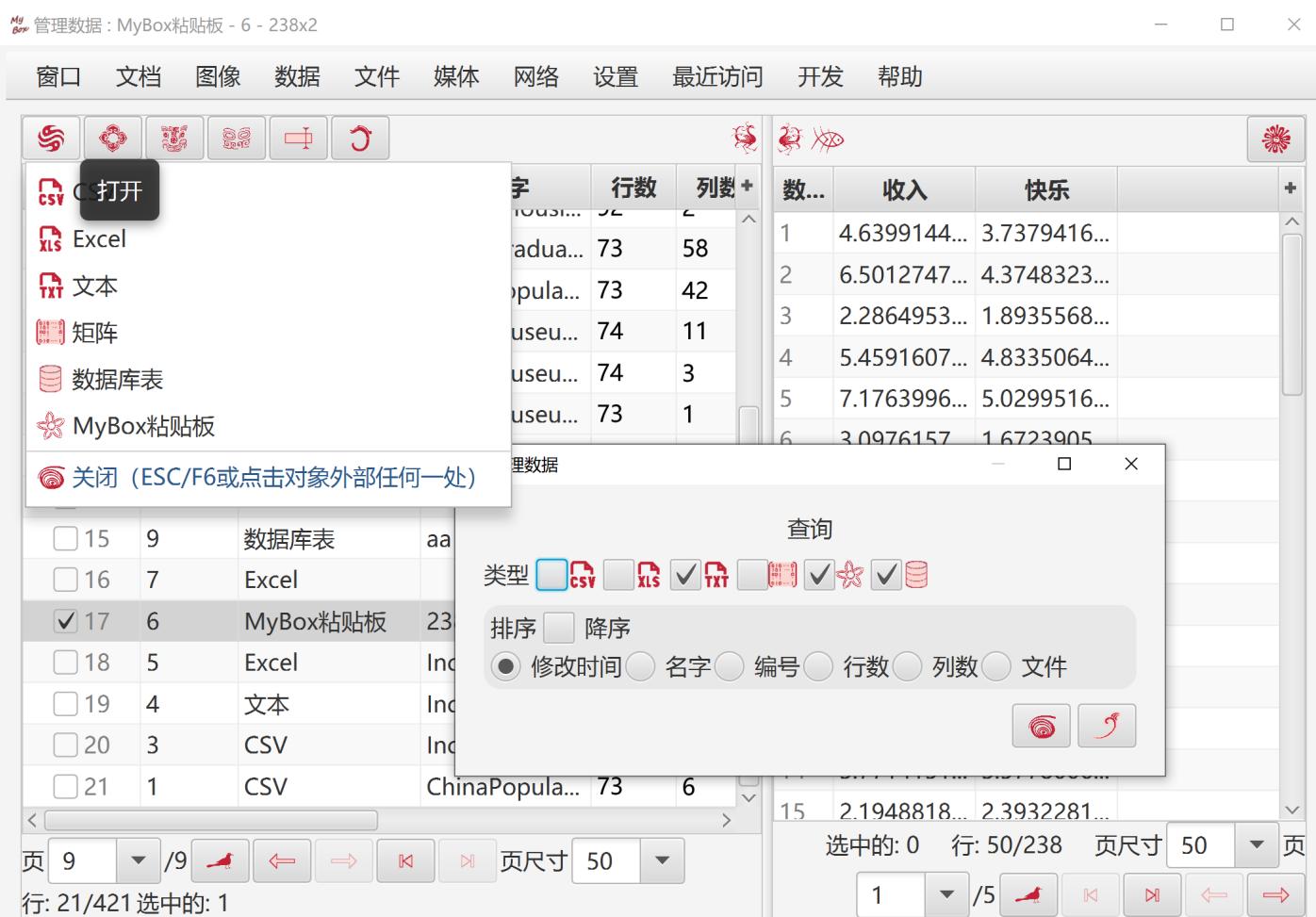
4. 生成自增主键、或选择主键列。



2.18 管理数据

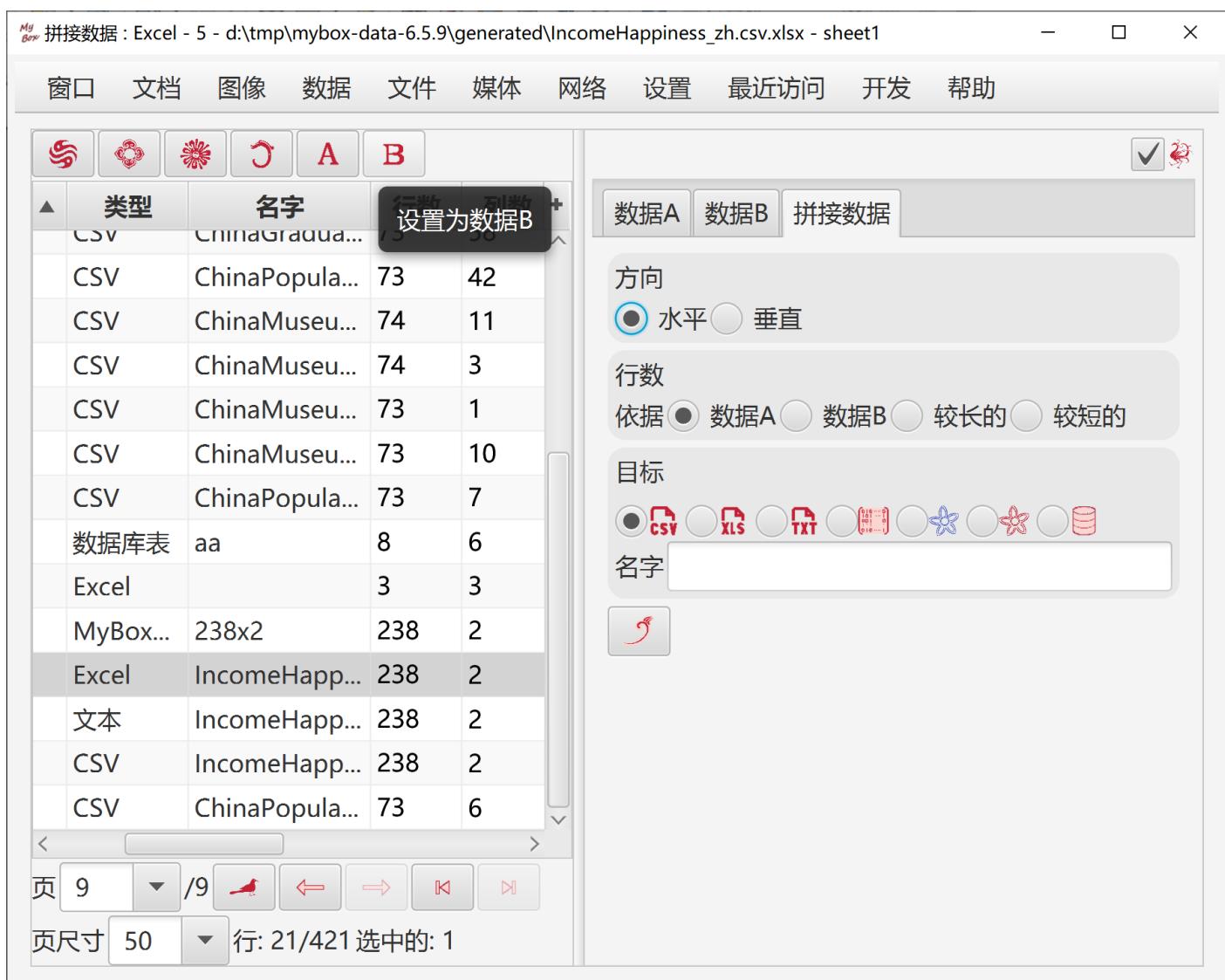
MyBox 在内部数据表中记录数据对象的定义：

	csv/excel/文本	MyBox 粘贴板	矩阵	数据库表
何时增加数据的定义	当数据文件第一次被打开时	当数据复制到 MyBox 粘贴板中	在矩阵管理器中保存新矩阵时	在数据库表管理器中保存新数据库表时
数据的存储位置	数据文件	MyBox 内部目录下的文件	MyBox 数据库表	MyBox 数据库表
当删除数据的定义	数据文件不受影响	删除内部文件	删除矩阵的数据	删除数据库表



2.19 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。



2.20 数据文件

2.20.1 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。
 - 若值包含分隔符或者换行，它应被引号围起来。
 - 若分隔符不是"#", 则以#"开头的行将被略过（当作注释）。
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”。
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。
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. 对于字符串数据支持编辑和保存多行。
6. 数据可以保存为不同的字符集和换行符。

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

2.20.2 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。
 - 值不能包含分隔符和换行符。
 - 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”。

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。
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. 不支持数据值多行。
6. 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 'ChinaPopulation_zh.csv'. The left sidebar contains:

- 窗口**: 编辑文本数据文件 : 文本 - 1061 - d:\tmp\mybox-data-6.5.9\data\ChinaPopulation_zh.csv
- 文档**
- 图像**
- 数据**
- 文件**
- 媒体**
- 网络**
- 设置**
- 最近访问**
- 开发**
- 帮助**
- 信息**
- 格式**

字符集:

- 自动检测
- 已知 **UTF-8**

第一行定义列名 (checked)

文本分隔符:

- 字串 (selected)
- 制表符
- 空格
- 空白字符
- 4个空格
- 8个空格
- , | # - + : ; @ &
- % ! " ' ? . * \ / _ = < >

若文件加载得不正确，改变选项并点击按钮“刷新”

备份

保存为

字符集 **UTF-8**

第一行定义列名 (checked)

The main workspace displays a table of data with the following columns:

	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

2.20.3 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。

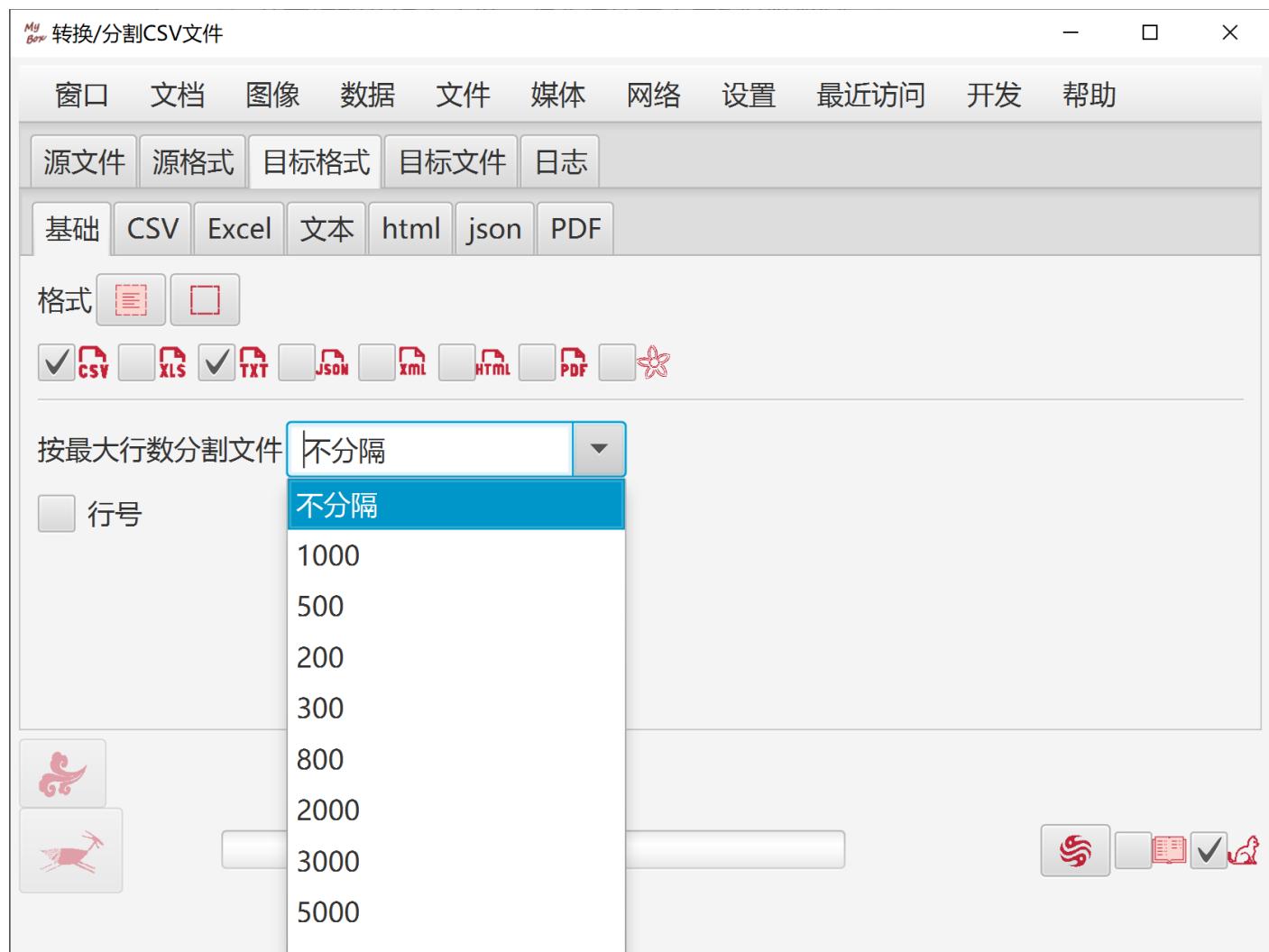
注意：工具只能处理 Excel 文件的基本数据。

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

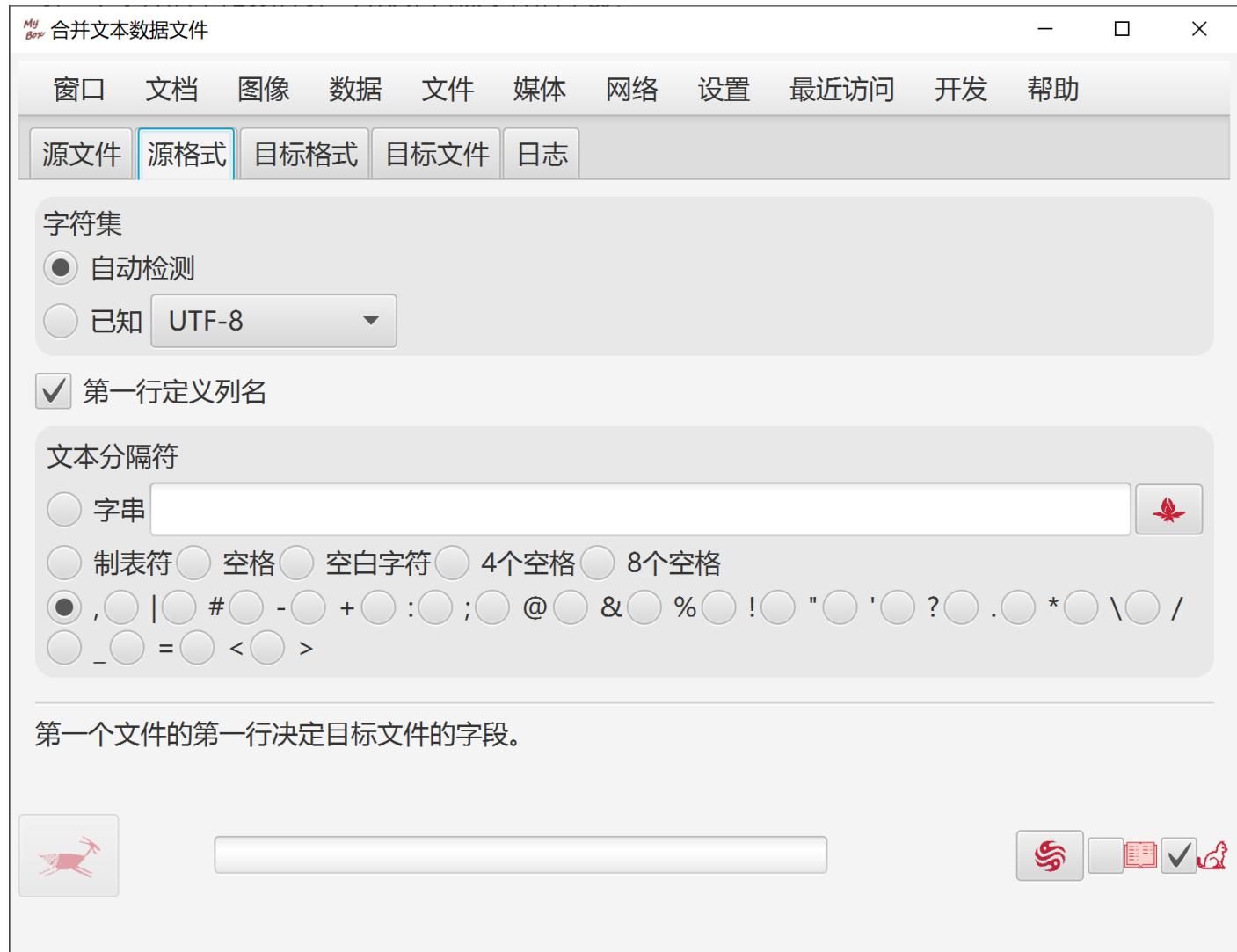
2.20.4 批量转换/分割数据文件

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。



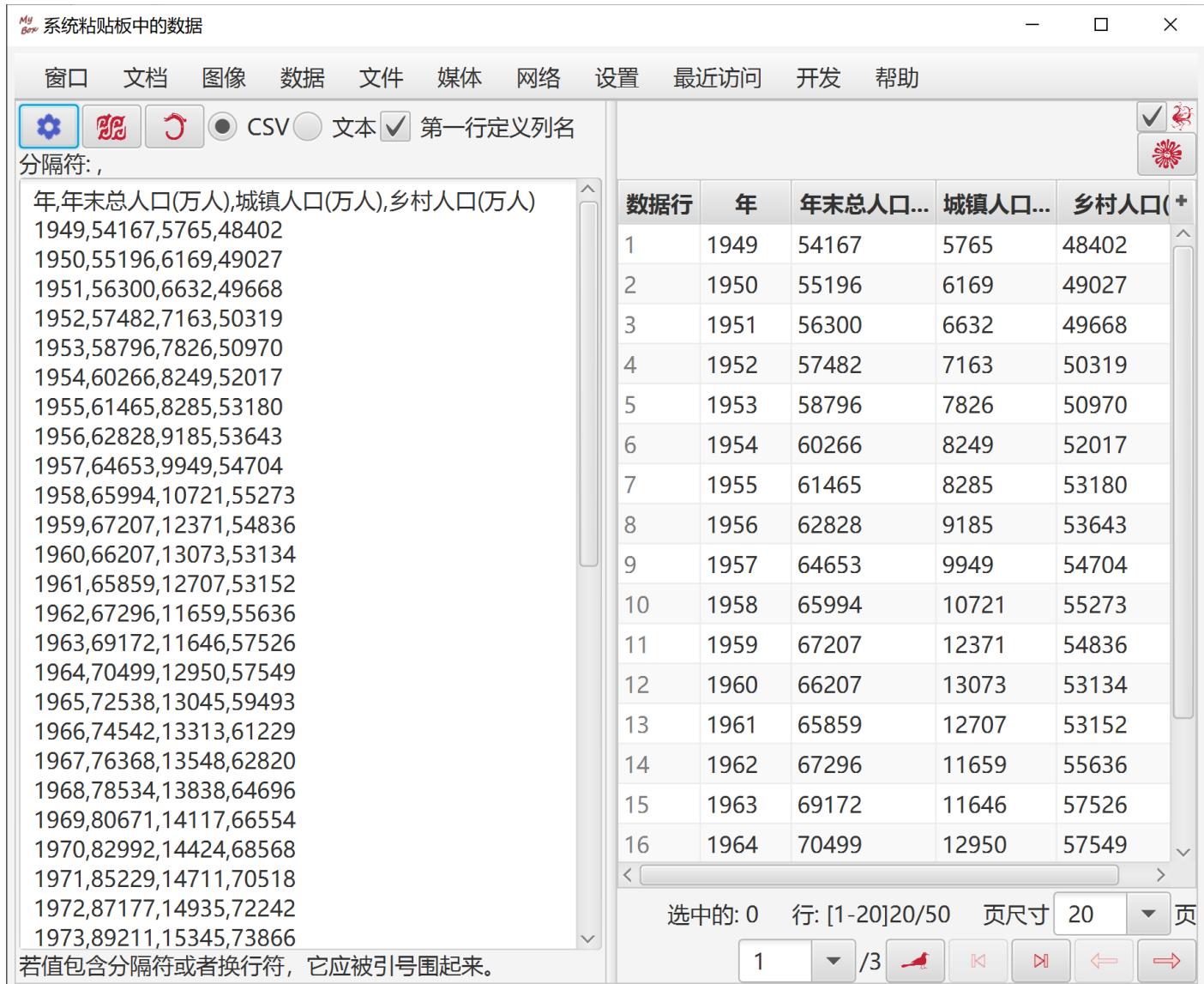
2.20.5 合并数据文件

1. 可以设置源文件的选项。
2. 可以设置目标格式。



2.21 系统粘贴板中数据

1. 读取系统粘贴板中的内容。
 2. 工具猜测分隔符并解析数据。
 3. 用户可以指定分隔符来解析。
 4. 可选解析方式：
 - CSV 解析：数值可包含分隔符和换行符，不支持空白字符解析和规则表达式解析。
 - 文本解析：数值不可包含分隔符和换行符，支持空白字符解析和规则表达式解析。
 5. 可选是否是第一行定义列名。



2.22 MyBox 粘贴板中的数据

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

- Toolbar:** Includes icons for file operations (New, Open, Save, Print, Copy, Paste, Find, Replace), selection (Select All, Select None, Invert Selection), and data manipulation (Copy, Paste, Cut, Delete, Sort Ascending, Sort Descending, Filter, Group, Ungroup).
- Menu Bar:** Windows, Document, Image, Data, File, Media, Network, Settings, Recent, Development, Help.
- Table View:** A table showing clipboard entries. The first entry (row 1) is "MyBox粘贴板" (Type: 表格行, Name: a, Rows: 6, Columns: 3). The second entry (row 2) is also "MyBox粘贴板" (Type: 表格行, Name: b, Rows: 238, Columns: 2). Row 2 is selected.
- Clipboard Editor Dialog:** A modal dialog titled "MyBox粘贴板中的数据 : ..." with the message "当前名:b". It has a text input field "新名字" containing "bm" and two buttons: "确定" (Confirm) and "取消" (Cancel).
- Table Data View:** A large table showing data rows from 1 to 10. The columns are labeled "表格行" (Row), "数..." (Count), "收入" (Income), and "快乐" (Happiness). The data is as follows:

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

- Page Navigation:** Shows page 1 of 50, with 2 rows selected.
- Bottom Status:** Displays "选中的: 0 行: 50/238 页尺寸: 50" and page navigation buttons.

2.23 矩阵

2.23.1 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 Undo/Redo.

The main area displays two matrices:

表格行	编号	类型	名字	行数	列数	
<input checked="" type="checkbox"/>	1	1065	矩阵	a	3	3
<input type="checkbox"/>	2	1064	矩阵	b	6	4

Below the matrices is a detailed view of matrix "a":

表格行	数...	列1	列2	列3
<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 controls (Page 1 of 1) and a row selection counter (50 rows selected). On the right side, there are buttons for column operations (Delete, Insert, Sort, etc.) and a vertical toolbar with icons for matrix operations.

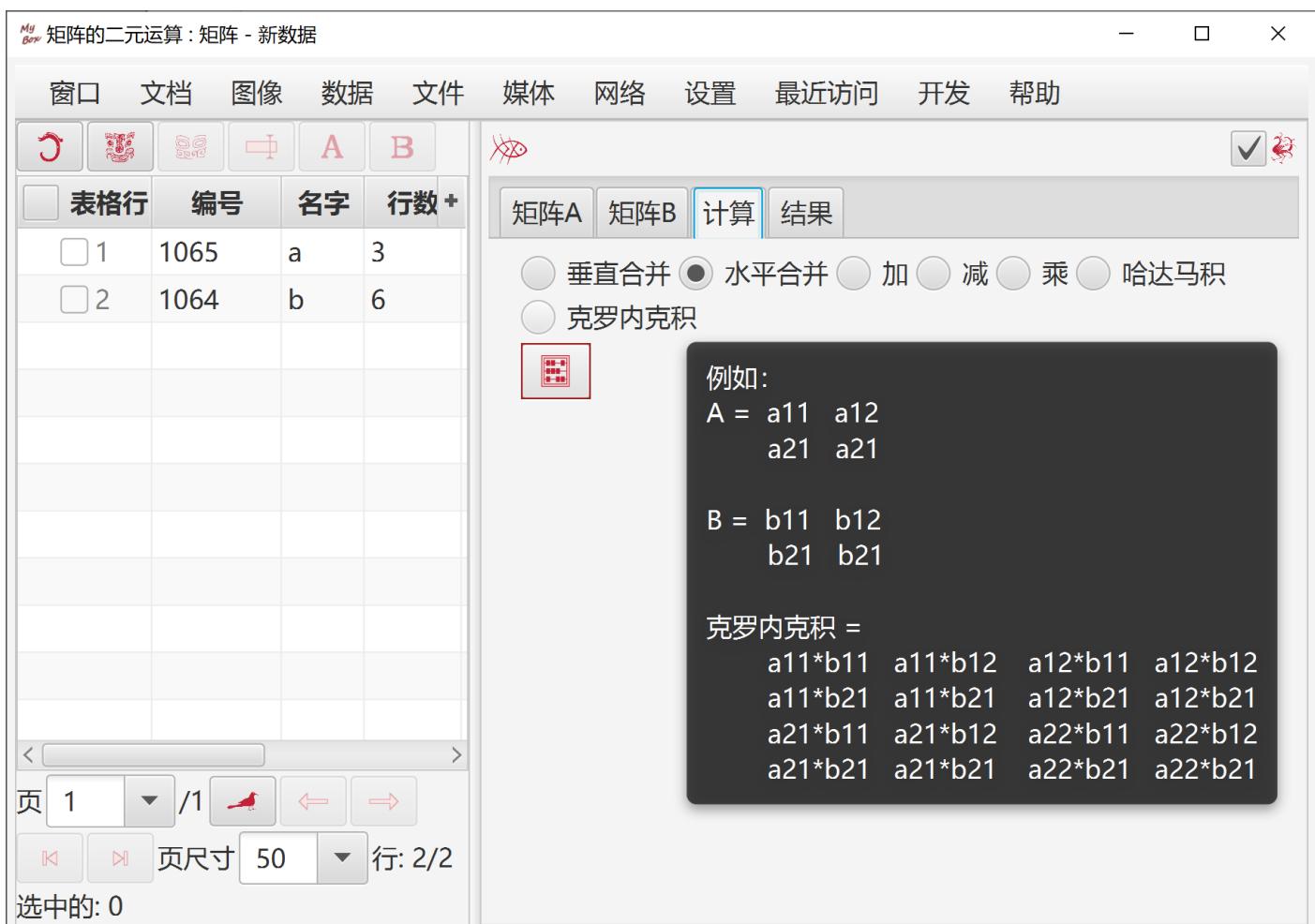
2.23.2 矩阵的一元计算

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。



2.23.3 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。



2.24 数据库表

2.24.1 管理数据库表

1. 可以查看表定义
2. 可以执行 SQL

The screenshot shows the 'Database Tables' window in MyBox Data Tools. The window title is '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 database management tasks.

The main area shows a list of tables:

表格行	编号	类型	名字	
<input type="checkbox"/>	1	930	数据库表	ChinaPopula...
<input type="checkbox"/>	2	488	数据库表	IrisSpecies_zh
<input checked="" type="checkbox"/>	3	387	数据库表	IncomeHapp...
<input type="checkbox"/>	4	9	数据库表	aa

To the right of the table list is a preview of data rows:

表格行	数...	id	收入	快乐
<input type="checkbox"/>	1	1	3.8626474...	2.3144889...
<input type="checkbox"/>	2	2	4.6399144...	3.7379416...

Below the preview is a table definition grid:

列	类型	长度	非空	主键	自动生成	引用表	引用列
id	Long	19	是	是	是		
收入	Double	52					
快乐	Double	52					

At the bottom of the window, there is a SQL script area containing the following code:

```
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 status bar at the bottom left shows '页 1' (Page 1) and '页尺' (Page Ruler).

2.24.2 SQL 标识符的限制

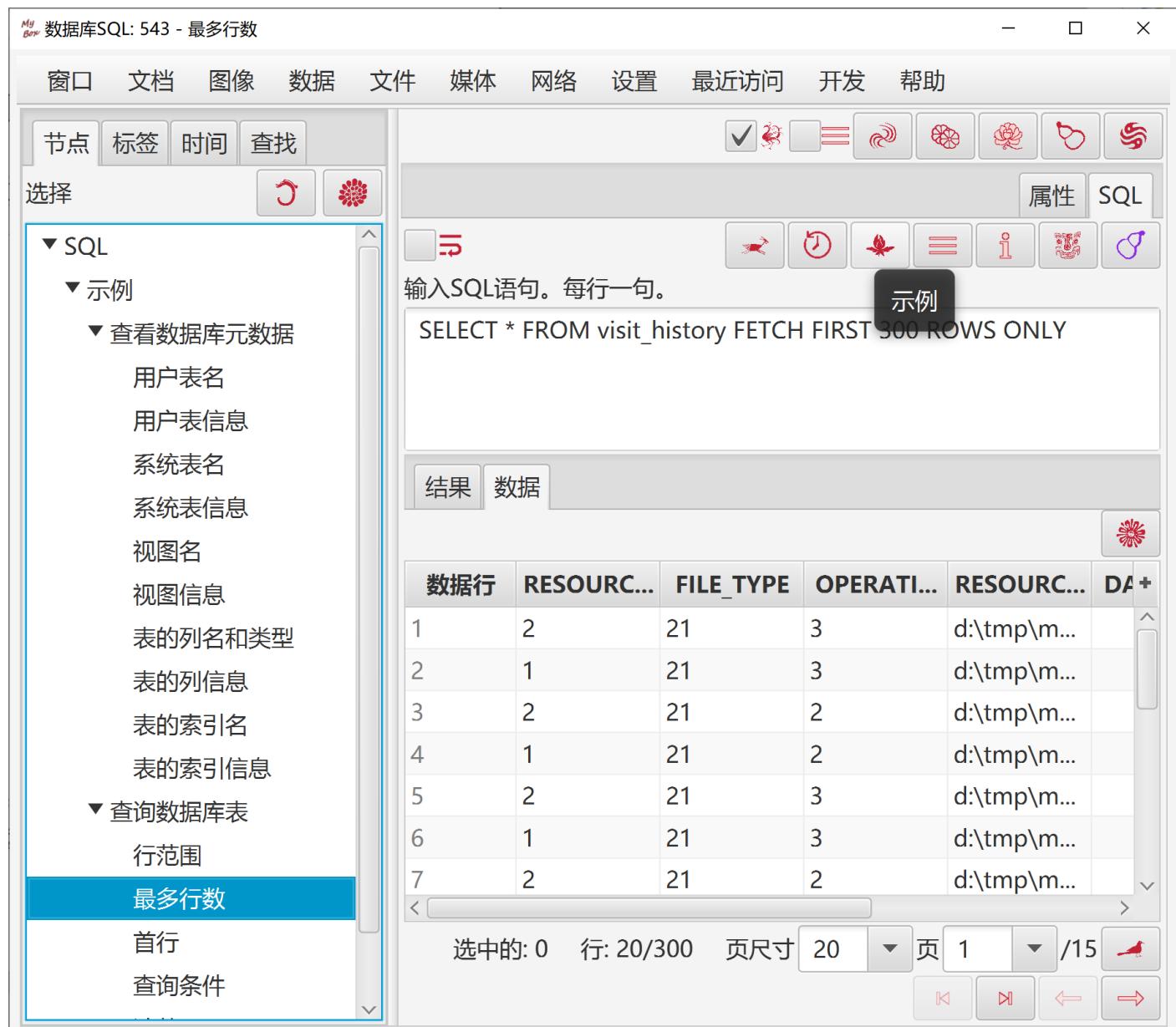
1. 表名和列名应当满足“SQL 标识符的限制”：

- 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。



2.24.3 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。



2.24.4 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



	CHINAPOPULATION_ZH	SELECT 年, 城镇人口_万人_, 乡村人口_万人_ FROM CHINAPOPULATION_ZH
<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	<input type="text" value="id"/>	
	<input type="text" value="年"/>	
	<input type="text" value="年末总人口_万人_"/>	
	<input type="text" value="男性人口_万人_"/>	
	<input type="text" value="女性人口_万人_"/>	
	<input type="text" value="城镇人口_万人_"/>	
	<input type="text" value="乡村人口_万人_"/>	

3 脚本与表达式

3.1 JShell (Java 交互编程工具)

3.1.1 关于 JShell

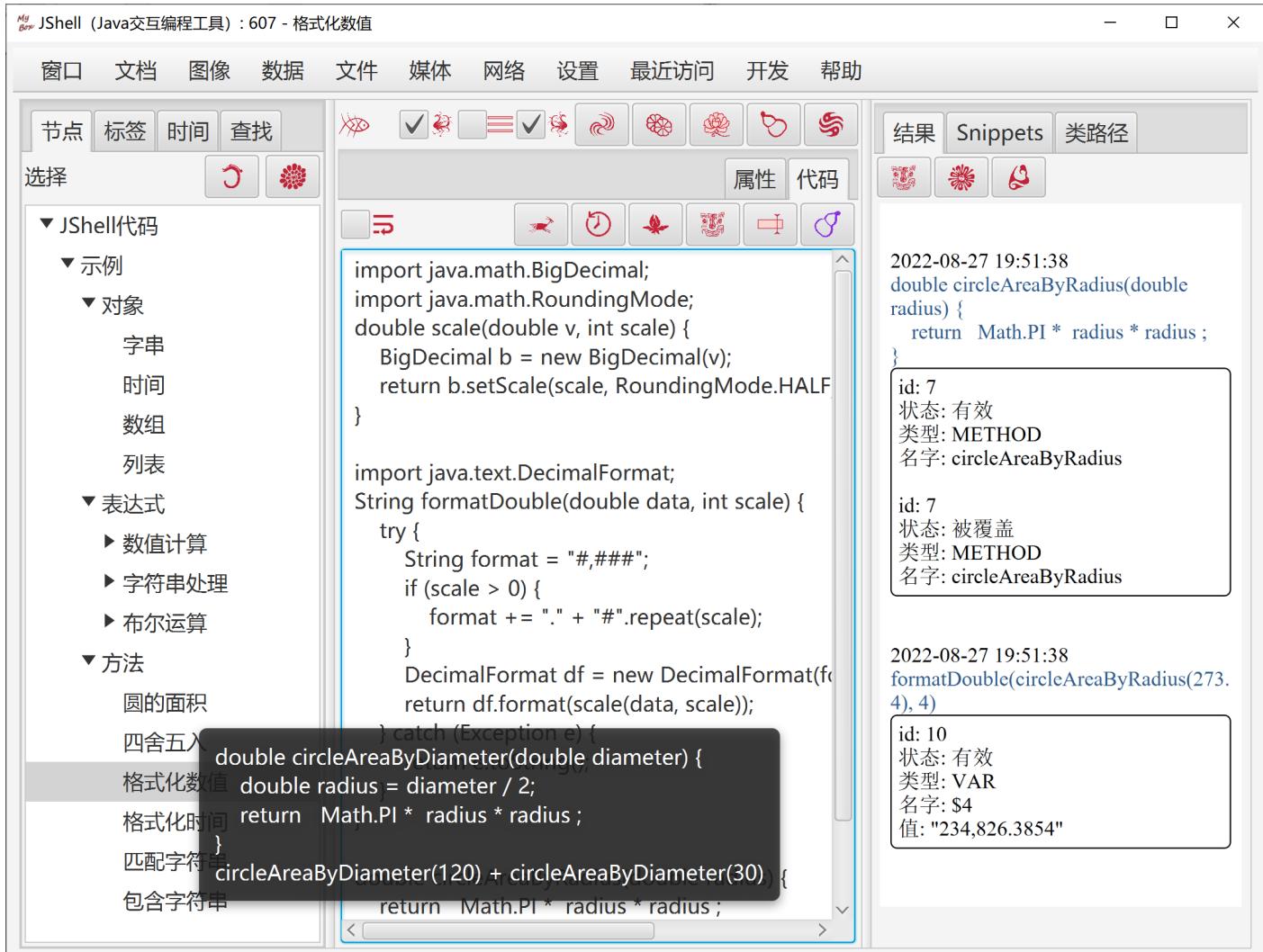
JShell 是 JDK 包含的工具之一:

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。

3.1.2 图形化运行 JShell

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)

3.2.1 关于 JEXL

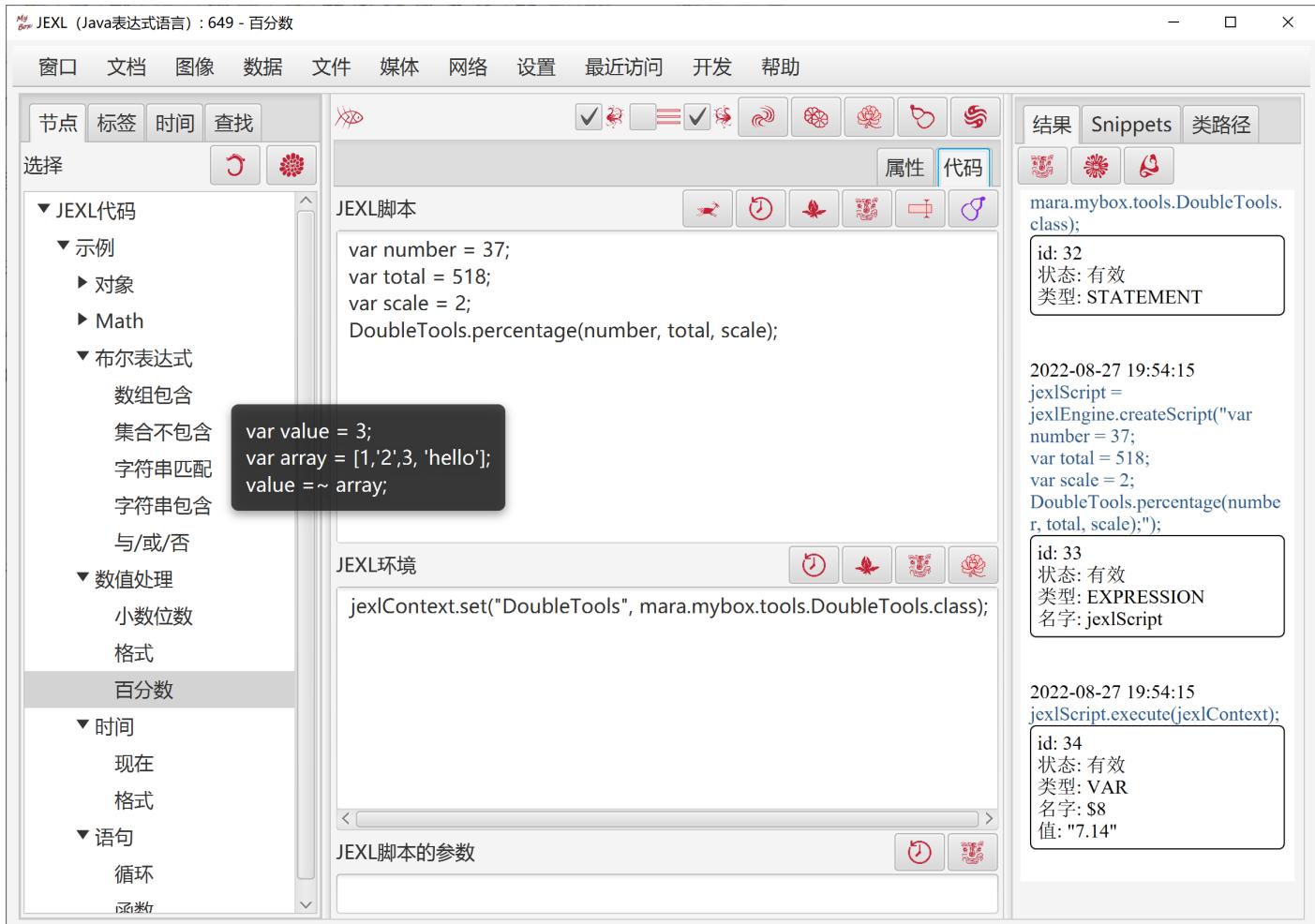
JEXL (Java 表达式语言) 是一个库文件，以变量和脚本来动态生成值：

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。

3.2.2 图形化运行 JEXL

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。



3.3 Javascript

此工具帮助管理 Javascript 代码:

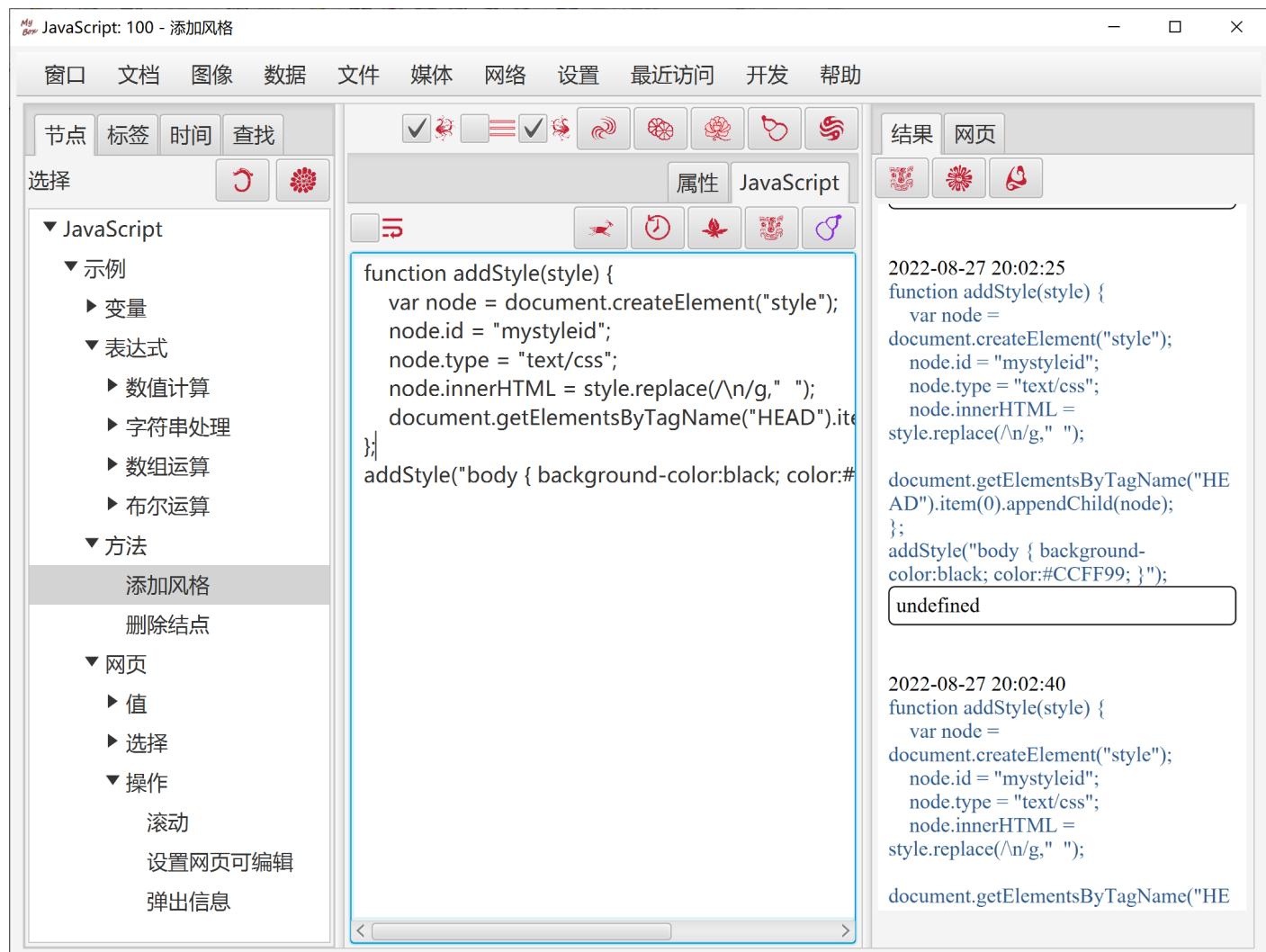
1. 编辑 JavaScript 代码。

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

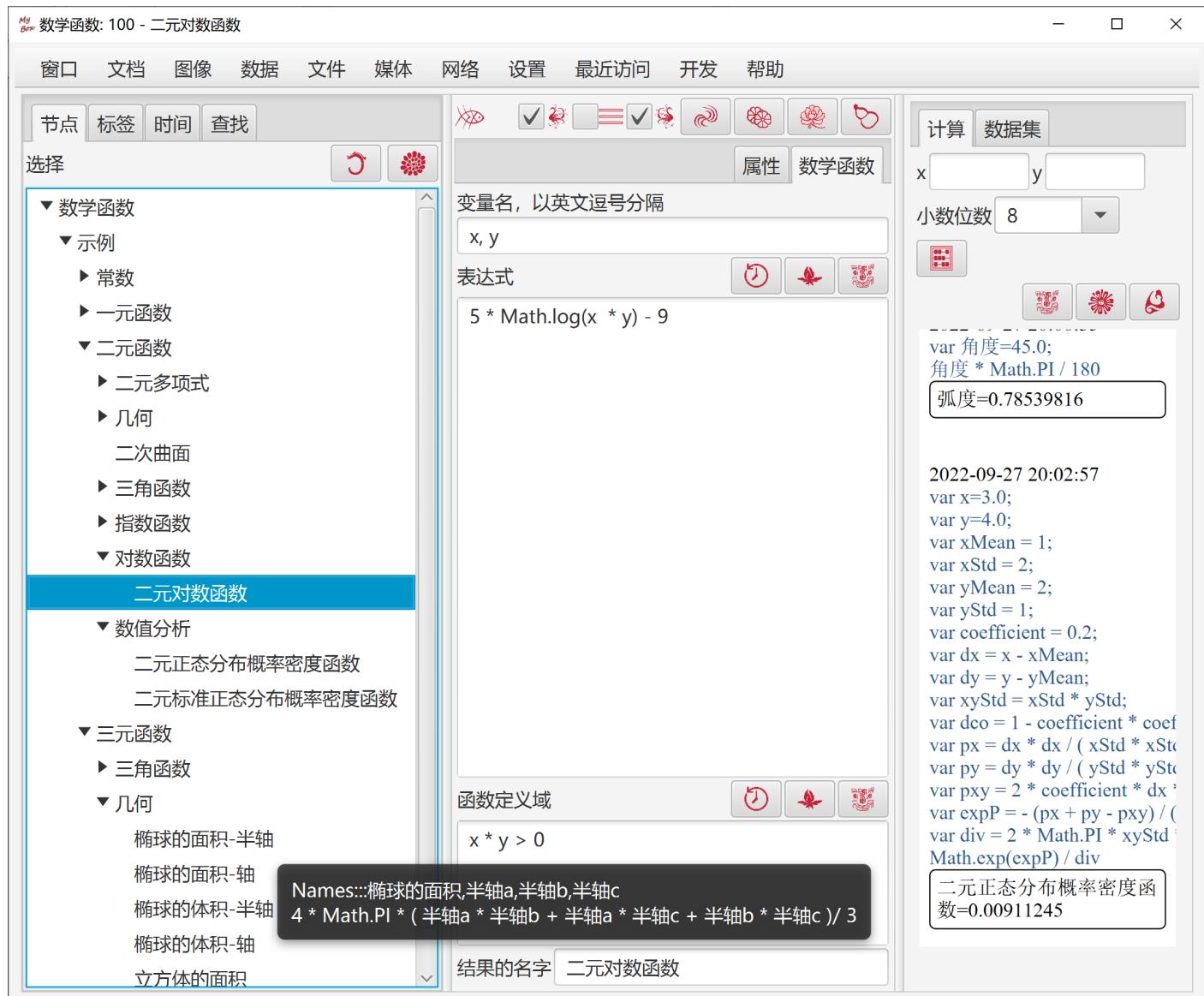
此工具帮助管理和计算数学函数。

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、下划线'_'、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。

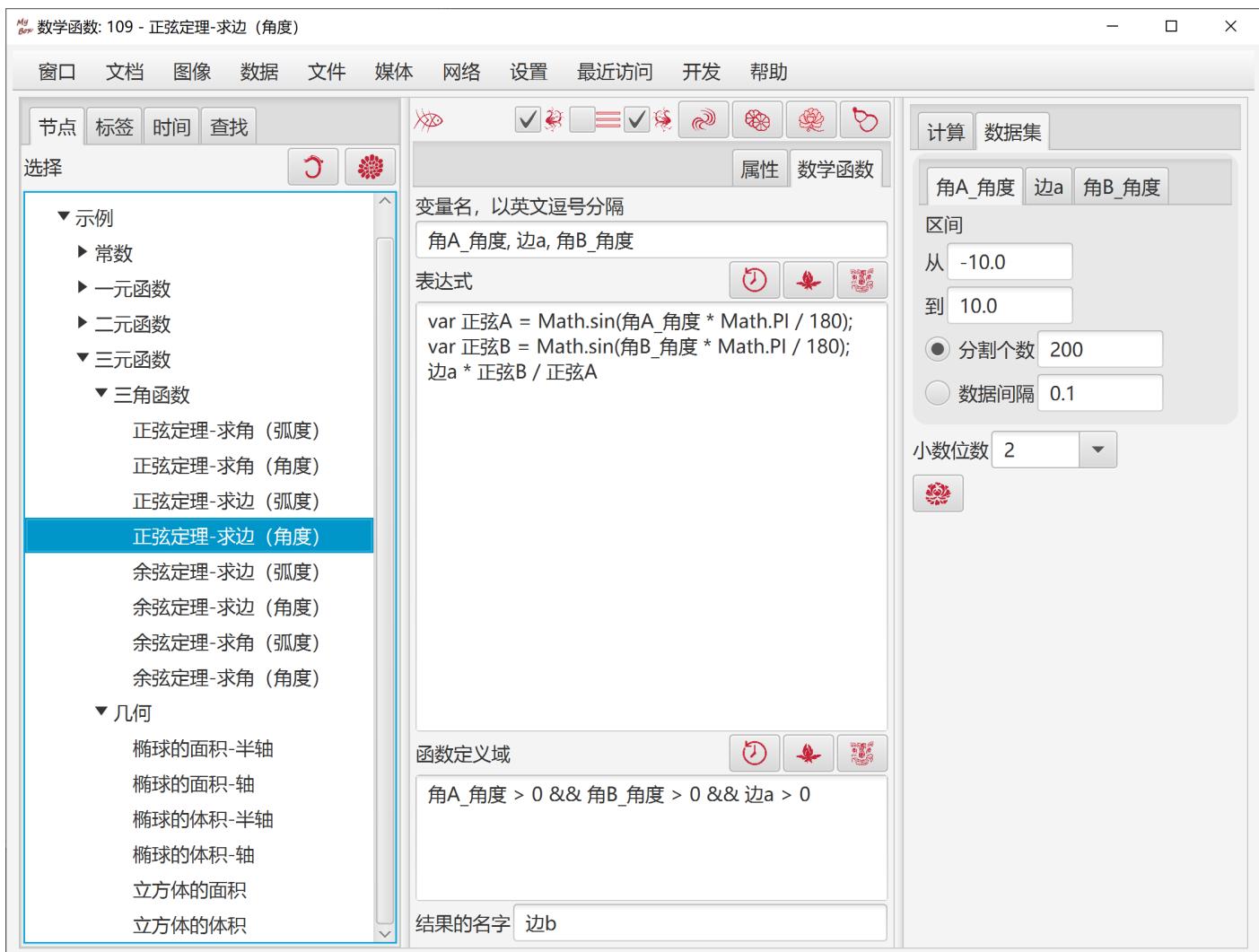
4.2 计算函数

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：
 - 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。



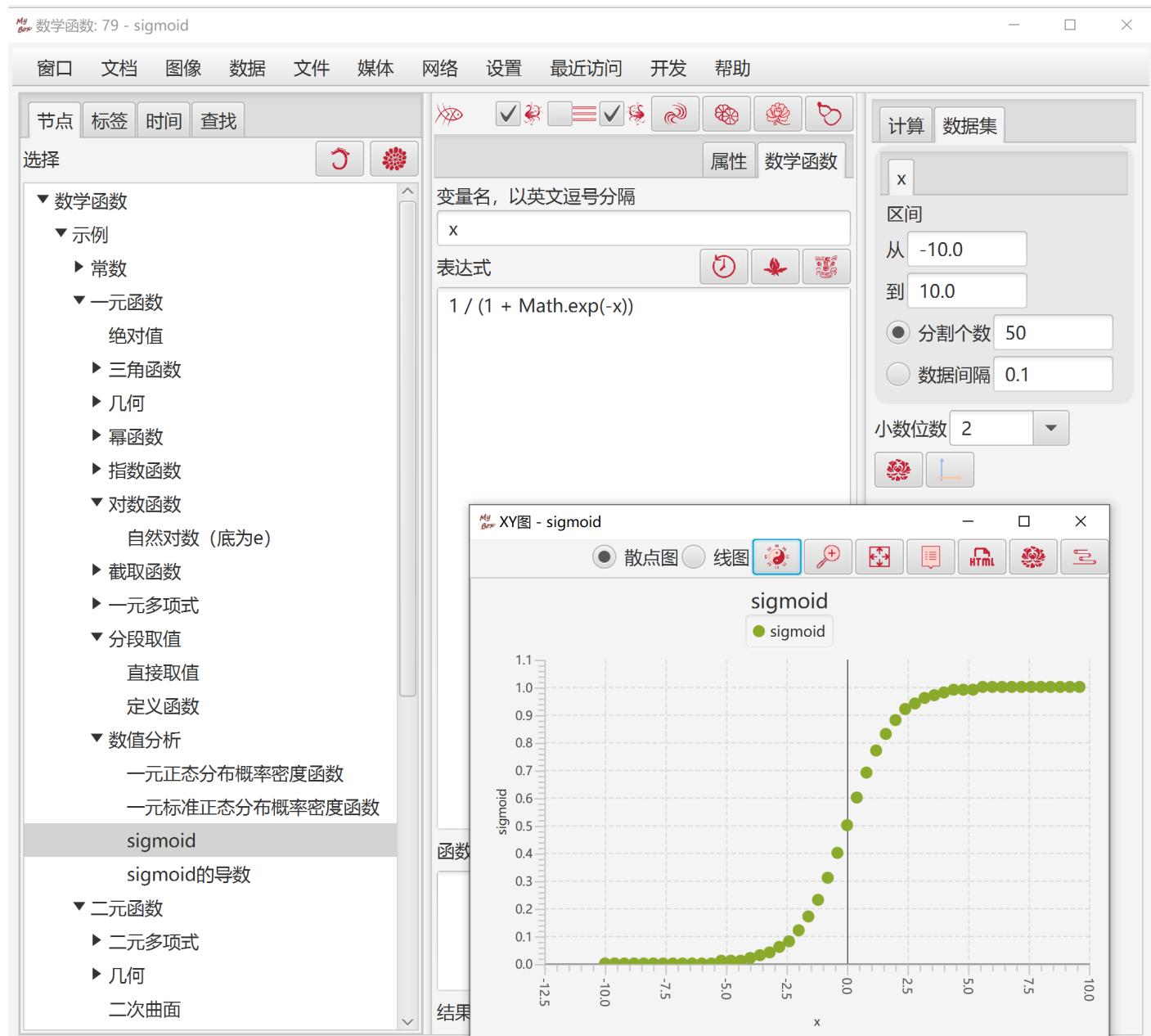
4.3 数据集

给定范围、间隔、和小数位数，可以生成函数的数据集。



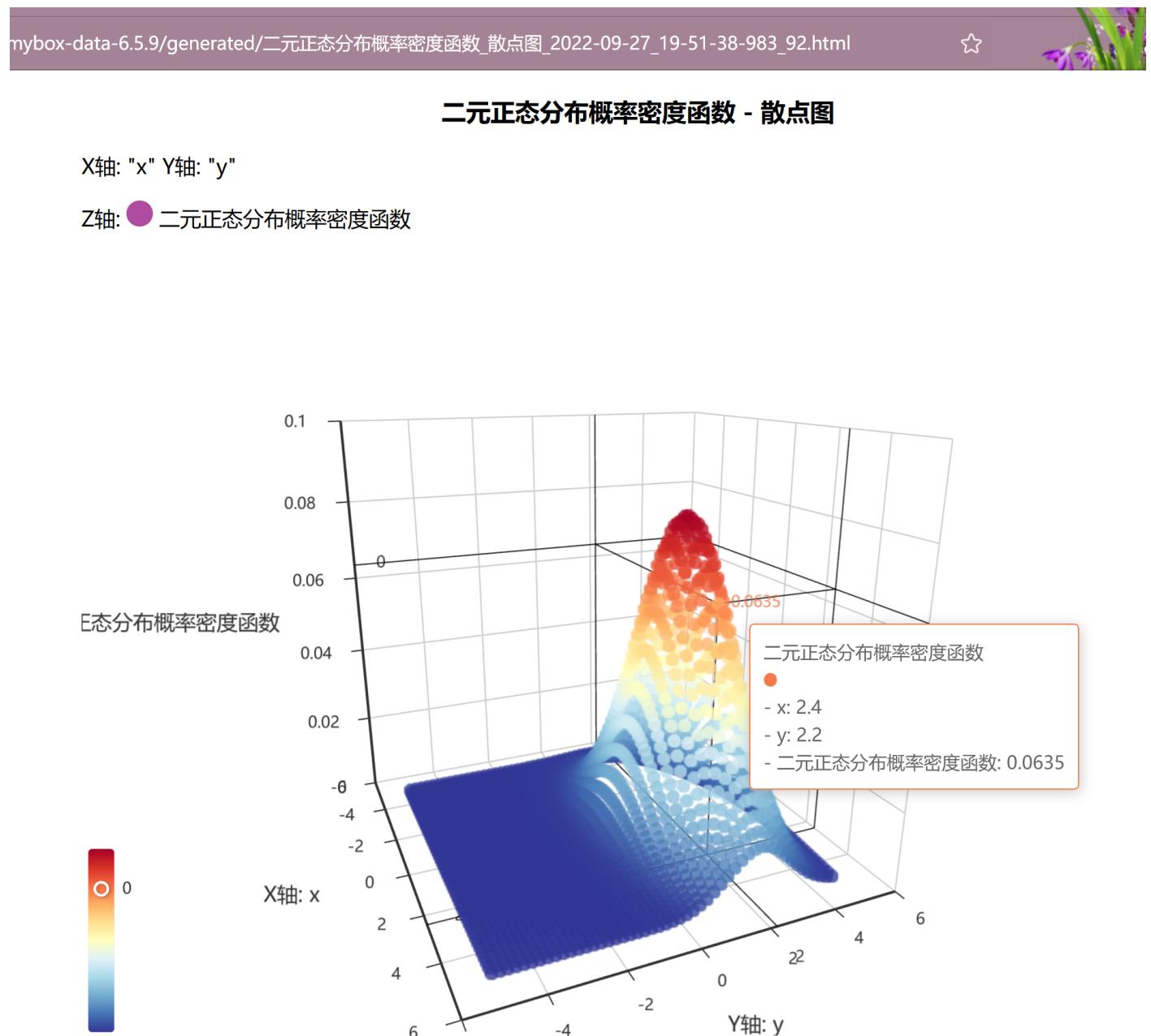
4.4 一元函数的 XY 图

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。



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。

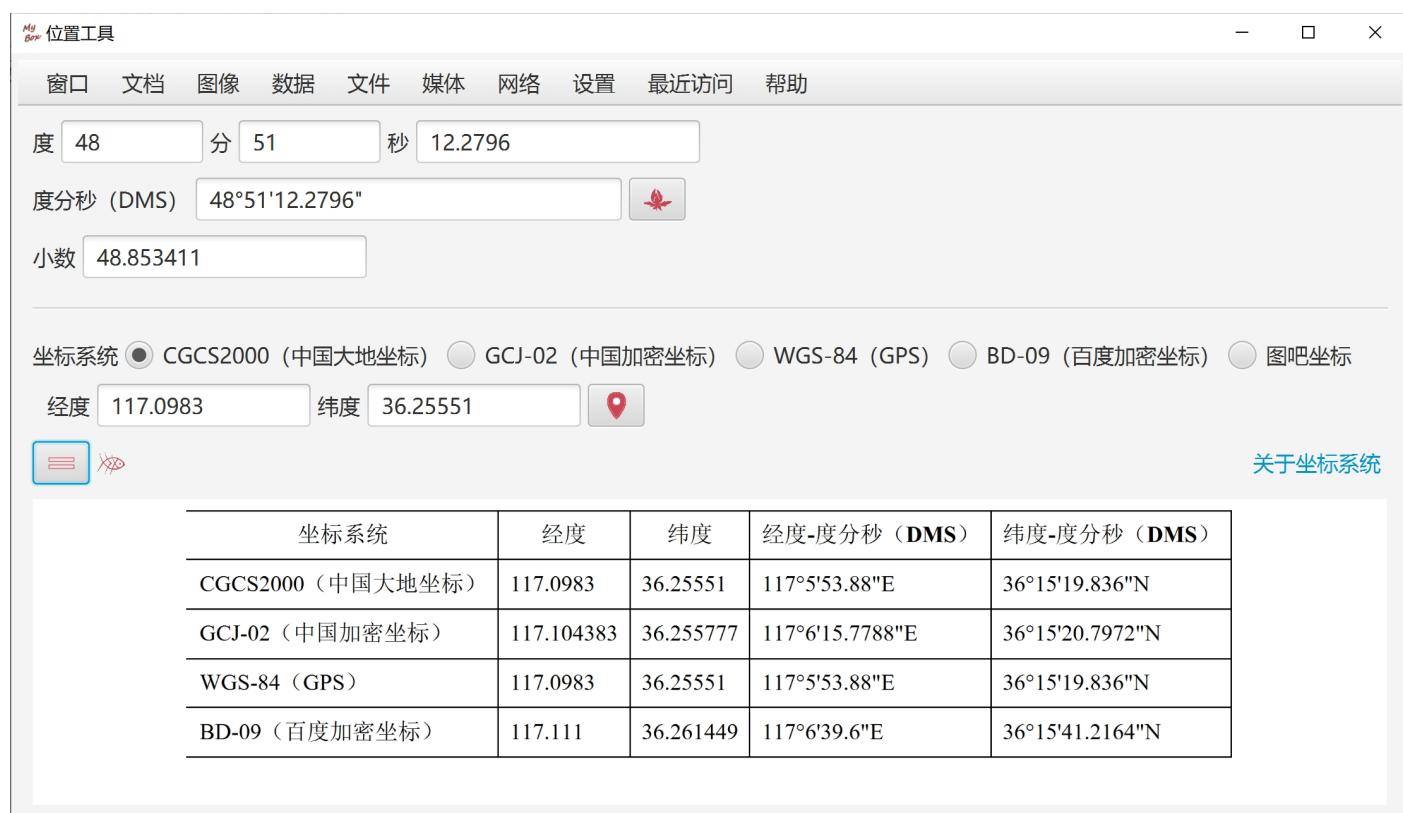


5.6 位置工具

- 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。



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>