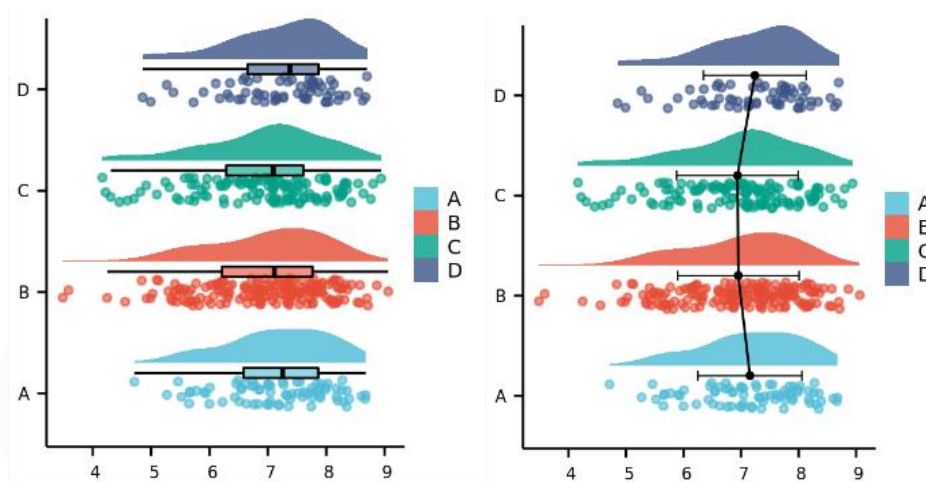


## 基础绘图 - [类别比较] - 云雨图



网址: <https://www.xiantao love>



更新时间: 2023.10.19

## 目录

|                   |    |
|-------------------|----|
| 基本概念 .....        | 3  |
| 应用场景 .....        | 3  |
| 分析过程 .....        | 3  |
| 结果解读 .....        | 5  |
| 数据格式 .....        | 6  |
| 参数说明 .....        | 8  |
| 云(半小提琴) .....     | 8  |
| 雨(点) .....        | 10 |
| 箱 .....           | 11 |
| 误差线 .....         | 12 |
| 连线 .....          | 13 |
| 坐标轴 .....         | 14 |
| 标题文本 .....        | 15 |
| 图注 (Legend) ..... | 16 |
| 风格 .....          | 17 |
| 图片 .....          | 18 |
| 结果说明 .....        | 19 |
| 主要结果 .....        | 19 |
| 方法学 .....         | 20 |
| 如何引用 .....        | 21 |
| 常见问题 .....        | 22 |

## 基本概念

- 云雨图：主要由半小提琴图和散点图组成，形似“云雨”，还提供了箱线图和误差线图。主要用于数据描述，比较不同分类下连续变量的分布特征。还可以用于数据分析，发现数据的异常值、离群值、正态或偏态、双峰或多峰等情况，提高图形的美观性和提供更多信息量。

## 应用场景

- 观察数据的总体分布情况
- 观察和比较不同分组之间的关系
- 添加箱线图和误差线等进一步补充集中趋势

## 分析过程

上传数据 ➡ 数据处理(清洗) ➡ 数据分析 ➡ 可视化

- 数据格式：（具体数据格式要求可以看后面过程的“数据格式”部分）
  - 宽型数据：每列必须为**数值类型**，对应数据点的 x 轴，行名为对应的 y 轴，**最多支持 10 列**
  - 长型数据：第 1 列必须是**字符类型**，第 1 列作为分组信息，**最多支持 10 个**，对应数据点的 x 轴；第 2 列必须为**数值类型**，对应数据点的 y 轴

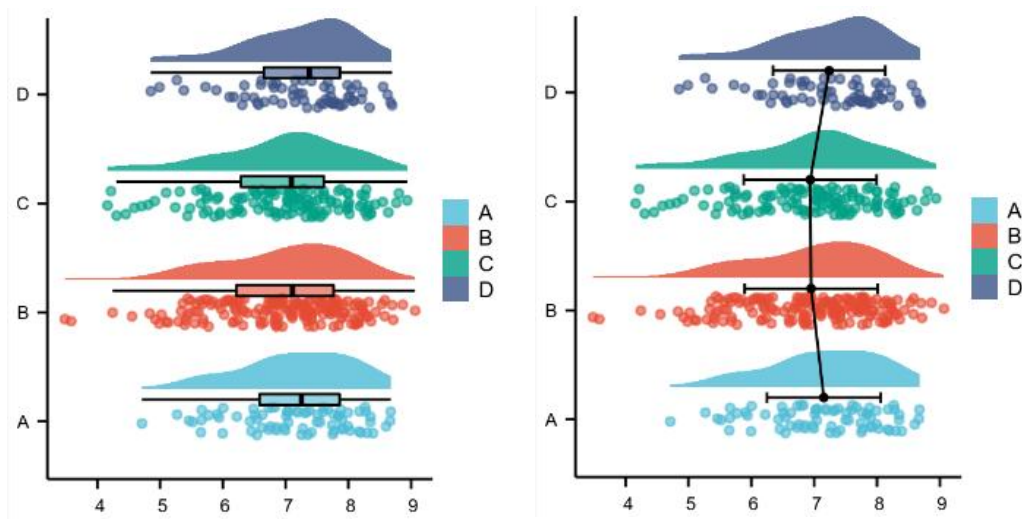
- 宽型数据必须提供至少 1 列，除了第 1 行的行名外，至少 2 行数据，最多 2500 行；长型数据只需提供 2 列，除了第 1 行的行名外，至少 2 行（每一个分组至少 2 行）数据，最多 25000 行。如下：左图为宽型数据，右图为长型数据

|    | A        | B        | C        | D        |
|----|----------|----------|----------|----------|
| 1  | A        | B        | C        | D        |
| 2  | 7.044666 | 6.608041 | 7.227593 | 7.563646 |
| 3  | 6.735964 | 7.407809 | 6.345179 | 4.990993 |
| 4  | 7.88133  | 7.806862 | 7.884849 | 6.664543 |
| 5  | 8.667676 | 6.191397 | 7.120066 | 6.549586 |
| 6  | 7.595751 | 6.840417 | 7.486774 | 7.924405 |
| 7  | 6.223678 | 7.40265  | 8.300704 | 7.267641 |
| 8  | 7.842347 | 8.127054 | 7.050483 | 7.207681 |
| 9  | 7.514485 | 7.144371 | 8.212169 | 7.674203 |
| 10 | 6.510429 | 7.107633 | 6.243006 | 7.761058 |
| 11 | 6.857713 | 6.282318 | 8.319113 | 6.351381 |
| 12 | 7.752567 | 6.178234 | 8.019483 | 6.810427 |
| 13 | 7.992065 | 7.331929 | 5.243225 | 7.988544 |
| 14 | 8.176787 | 7.747443 | 7.202435 | 7.788359 |
| 15 | 5.997809 | 5.562563 | 5.900735 | 7.737385 |
| 16 | 6.828844 | 5.392529 | 4.591073 | 8.697586 |
| 17 | 8.29924  | 8.048609 | 5.864564 | 7.121681 |
| 18 | 8.195816 | 8.095589 | 5.961621 | 5.268285 |
| 19 | 7.024352 | 6.826947 | 6.510165 | 8.675228 |
| 20 | 6.621243 | 7.702264 | 7.010626 | 8.132235 |

|    | A     | B        |
|----|-------|----------|
| 1  | Stage | Value    |
| 2  | A     | 7.044666 |
| 3  | B     | 7.227593 |
| 4  | C     | 6.608041 |
| 5  | D     | 7.563646 |
| 6  | D     | 4.990993 |
| 7  | C     | 7.407809 |
| 8  | D     | 6.664543 |
| 9  | B     | 6.345179 |
| 10 | D     | 6.549586 |
| 11 | C     | 7.806862 |
| 12 | C     | 6.191397 |
| 13 | B     | 7.884849 |
| 14 | B     | 7.120066 |
| 15 | C     | 6.840417 |
| 16 | C     | 7.40265  |
| 17 | C     | 8.127054 |
| 18 | D     | 7.924405 |
| 19 | D     | 7.267641 |
| 20 | A     | 6.735964 |

- 数据处理：对每一列数值类型的数据及其他列数据进行相应处理
  - 数值类型数据只能是纯数值类型数据，不能包含 0，负数、非数值与不规则的值
  - 数据第一、二列不允许有完全一样的值
  - 分类类型数据只能是纯字符类型的数据，不能包含数值，缺失值与无法识别的值
- 可视化：数据清洗后，用 ggplot2 包进行可视化

## 结果解读



- 左图：由半小提琴图和散点图组成云雨图，展示数据分布的密集程度，中间的箱线图提示数据的中值和上下四分位数。不同颜色代表不同分组，纵坐标是对应各个分组，横坐标是对应的值。注意：默认 xy 轴是颠倒的。
- 右图：由半小提琴图和散点图组成云雨图，展示数据分布的密集程度，中间的误差线提示数据的误差范围，连线展示不同组之间的误差变化趋势。不同颜色代表不同分组，纵坐标是对应各个分组，横坐标是对应的值。注意：默认 xy 轴是颠倒的。

## 数据格式

云雨图 - 宽型数据

|    | A        | B        | C        | D        |
|----|----------|----------|----------|----------|
| 1  | A        | B        | C        | D        |
| 2  | 7.044666 | 6.608041 | 7.227593 | 7.563646 |
| 3  | 6.735964 | 7.407809 | 6.345179 | 4.990993 |
| 4  | 7.88133  | 7.806862 | 7.884849 | 6.664543 |
| 5  | 8.667676 | 6.191397 | 7.120066 | 6.549586 |
| 6  | 7.595751 | 6.840417 | 7.486774 | 7.924405 |
| 7  | 6.223678 | 7.40265  | 8.300704 | 7.267641 |
| 8  | 7.842347 | 8.127054 | 7.050483 | 7.207681 |
| 9  | 7.514485 | 7.144371 | 8.212169 | 7.674203 |
| 10 | 6.510429 | 7.107633 | 6.243006 | 7.761058 |
| 11 | 6.857713 | 6.282318 | 8.319113 | 6.351381 |
| 12 | 7.752567 | 6.178234 | 8.019483 | 6.810427 |
| 13 | 7.992065 | 7.331929 | 5.243225 | 7.988544 |
| 14 | 8.176787 | 7.747443 | 7.202435 | 7.788359 |
| 15 | 5.997809 | 5.562563 | 5.900735 | 7.737385 |
| 16 | 6.828844 | 5.392529 | 4.591073 | 8.697586 |
| 17 | 8.29924  | 8.048609 | 5.864564 | 7.121681 |
| 18 | 8.195816 | 8.095589 | 5.961621 | 5.268285 |
| 19 | 7.024352 | 6.826947 | 6.510165 | 8.675228 |
| 20 | 6.621243 | 7.702264 | 7.010626 | 8.132235 |
| 21 | 5.774392 | 6.872803 | 5.929226 | 8.078206 |

数据要求：

- 对于宽型数据：第 1 行会读取为行名，用于绘制云雨图的纵坐标，每一列**必须提供数值类型**，用于绘制云雨图的横坐标。
- 对于宽型数据：**除第一行（第一行作为行名）外，数据至少需要 2 行，1 列**，最多支持 10 列（即绘制出来 10 个云雨图）。
- 对于宽型数据：数据**最多支持 2500 行，10 列**，若验证数据时返回报错，需要在上传数据内进行相应的调整，然后再上传数据。

云雨图 - 长型数据



|    | A     | B        |
|----|-------|----------|
| 1  | Stage | Value    |
| 2  | A     | 7.044666 |
| 3  | B     | 7.227593 |
| 4  | C     | 6.608041 |
| 5  | D     | 7.563646 |
| 6  | D     | 4.990993 |
| 7  | C     | 7.407809 |
| 8  | D     | 6.664543 |
| 9  | B     | 6.345179 |
| 10 | D     | 6.549586 |
| 11 | C     | 7.806862 |
| 12 | C     | 6.191397 |
| 13 | B     | 7.884849 |
| 14 | B     | 7.120066 |
| 15 | C     | 6.840417 |
| 16 | C     | 7.40265  |
| 17 | C     | 8.127054 |
| 18 | D     | 7.924405 |
| 19 | D     | 7.267641 |
| 20 | A     | 6.735964 |

- 对于长型数据：第 1 行会读取为行名，第 1 列**必须提供字符类型**（分类变量中的分组数量**最多支持 10 组**），用于绘制云雨图的纵坐标，第 2 列**必须提供数值类型**，用于绘制云雨图的横坐标。
- 对于长型数据：**除第一行（第一行作为行名）外，数据至少需要 2 行，2 列数据最多支持 25000 行，2 列**，若验证数据时返回报错，需要在上传数据内进行相应的调整，然后再上传数据。
- 数据每一列列名不能重复，不能有空值，不能有不识别的字符。
- 数值类型数据只能是纯数值类型数据，不能包含 0、负数、非数值与不规则的值。

## 参数说明

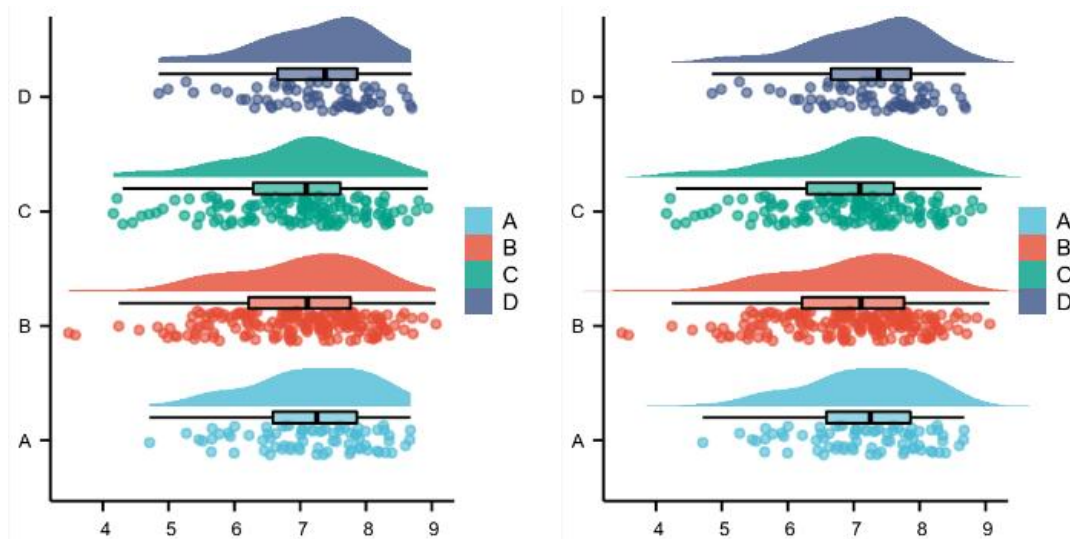
(说明：标注了颜色的为常用参数。)

### 云(半小提琴)

➤ 截断：可选择是否展示截断(数据范围截断)后的数据。默认为截断， 还可以选择不截断，如下：左图为截断，右图不为不截断

- 截断：每个半小提琴的密度都是该组的数据范围内计算的，默认为截断
- 不截断：即每个半小提琴的密度都是在整个数据范围内计算的





- 填充：可以修改图中多边形的填充颜色，最多支持修改 10 个颜色，超出会使用随机颜色。受配色方案全局性修改。
- 描边：可以修改图中多边形描边的颜色，默认白色。
- 描边粗细：可以修改图中多边形描边的线条粗细，默认 0.50pt。
- 不透明度：可以修改图中多边形的不透明度，1 表示完全不透明。

## 雨(点)



- 填充：可以修改图中点的填充颜色，最多支持修改 10 个颜色，超出会使用随机颜色。受配色方案全局性修改。
- 描边：可以修改图中点的描边颜色，最多支持修改 10 个颜色，超出会使用随机颜色。受配色方案全局性修改。
- 样式：可以修改图中点的形状，默认圆形。
- 大小比例：可以修改图中点的大小，默认 0.8。
- 不透明度：可以修改图中点的不透明度，1 表示完全不透明。
- 抖动宽度：可以修改图中点的抖动宽度，0 表示完全不抖动。

## 箱



- 展示箱子：可选择是否展示箱线图。注意：箱线图和误差线图不能同时展示。
- 展示离群点：可选择是否展示离群点。
- 填充：可以修改图中箱子的填充颜色，最多支持修改 10 个颜色，超出会使用随机颜色。受配色方案全局性修改。
- 描边：可以修改图中箱子的描边颜色，默认黑色。
- 描边粗细：可以修改图中箱子的描边线条粗细，默认 0.50pt。
- 不透明度：可以修改图中箱子的不透明度，1 表示完全不透明。
- 宽度：可以修改图中箱子上下宽度（因为默认 xy 轴颠倒，所以是上下的宽度）。

## 误差线

误差线

展示误差线 ☐

类型 均值±标准误

颜色

粗细 0.50pt

宽度 0.1

点大小 0.6

点颜色

- 展示误差线：可选择是否展示误差线。注意：需要同时关掉箱线图的展示。
- 类型：可选择均值±标准误(SEM)、均值±标准差(SD)、中位数~上下四分位，默认均值±标准误(SEM)。
- 颜色：可以选择并修改误差线的颜色。
- 粗细：可以选择并修改误差线的粗细。
- 宽度：可以选择并修改误差线两端的宽度。
- 点大小：可以选择并修改误差线均值/中值（这里是根据误差线的类型选项来的）的点的大小。
- 点颜色：可以选择并修改误差线均值/中值（这里是根据误差线的类型选项来的）的点的颜色。

## 连线



- 展示连线：可选择是否展示误差线中值/均值（这里是根据误差线的类型选项来的）的连线。注意：需要同时关掉箱线图的展示和打开误差线的展示。
- 颜色：可以选择并修改连线的颜色。
- 连线粗细：可以选择并修改连线的粗细。

## 坐标轴

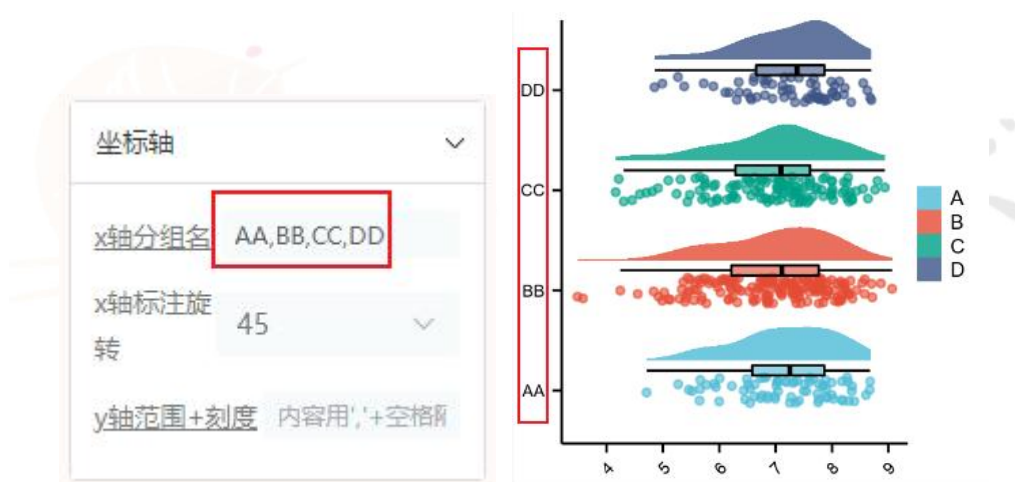
坐标轴

x轴分组名 ,+空格隔开

x轴标注旋转 0

y轴范围+刻度 内容用','+空格隔

- x 轴分组名：可以修改 x 轴分组名字，每个组的名字需要用英文逗号隔开，因为默认颠倒 xy 轴，所以修改了 y 轴刻度对应的分组名，例如：



- x 轴标注旋转：可以选择设置 x 轴标注的倾斜角度。
- y 轴范围+刻度：用于修改 y 轴范围以及刻度，如果需要分割，需要用英文的小括号隔开，数值间需要用英文逗号隔开，例如(1,1,2,5,10,10)。如果调整过大可能会无作用



## 标题文本

| 标题   |        |
|------|--------|
| 大标题  | 大标题内容  |
| x轴标题 | x轴标题内容 |
| y轴标题 | y轴标题内容 |

- 大标题：大标题文本。
- x 轴标题：x 轴标题文本。
- y 轴标题：y 轴标题文本。
- 补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如{{2}}；如果需要下标，可以用两个英文输入法下的中括号括住，比如[[2]]。

## 图注 (Legend)

图注

是否展示

☒

图注标题

图注标题内容

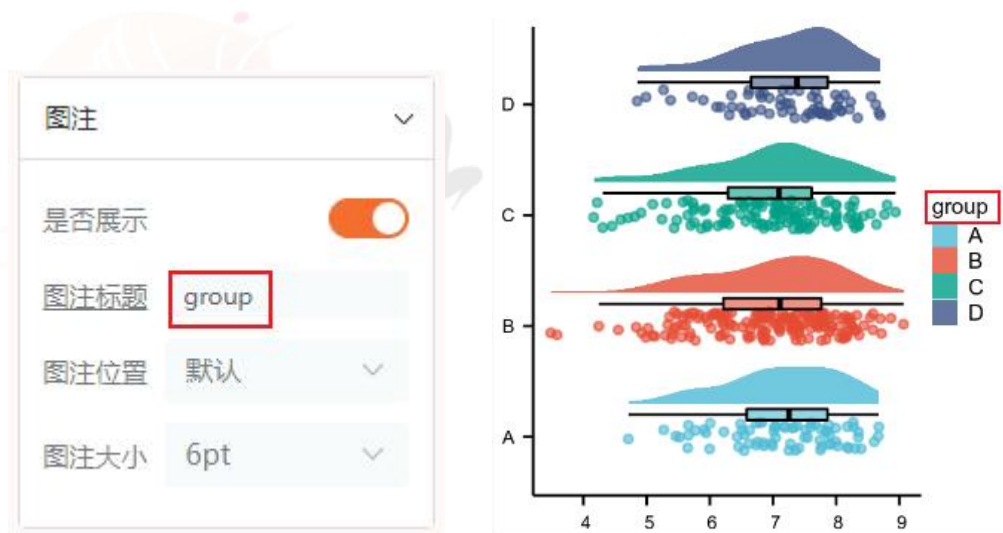
文字大小

6pt

图注位置

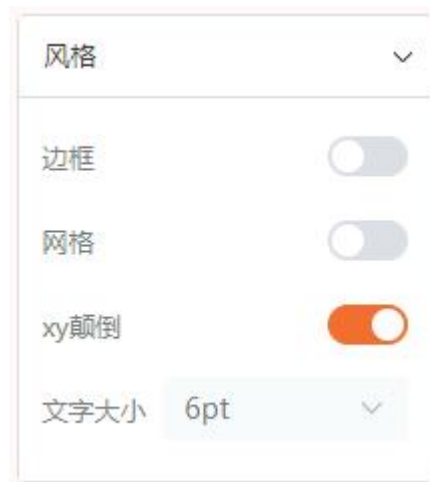
默认

- 是否展示：是否展示图注。
- 图注标题：可以添加图注标题，如：



- 文字大小：图注标题文字的大小，默认为 6pt。
- 图注位置：可选择默认、右、上、下。

## 风格



- 边框：可以选择是否进行添加图形边框的操作。
- 网格：可以选择是否进行添加图形网格的操作。
- xy 颠倒：可以选择是否颠倒 xy 轴，默认颠倒 xy 轴。
- 文字大小：控制整体文字大小，默认为 6pt。

## 图片



- 宽度：图片横向长度，单位为 cm。
- 高度：图片纵向长度，单位为 cm。
- 字体：可以选择图片中文字的字体。

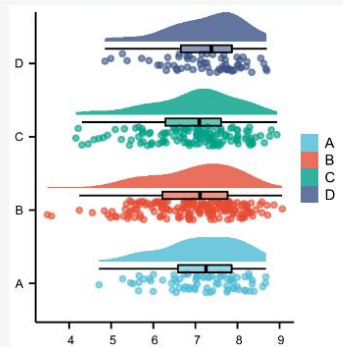


## 结果说明

## 主要结果

### 云雨图

云雨图: 可以用于数据描述; 不同颜色代表不同的分组, 比较不同分类下连续变量的分布特征。



云雨图.pdf

云雨图.tiff

主要结果格式为图片格式, 提供 PDF、TIFF 格式下载

## 方法学

软件：R (4.2.1)版本

R 包：ggplot2 包

处理过程：

(1) 分析不同分类下连续变量的分布情况，用 ggplot2 可视化结果，进而描述各组数据变量的分布特征





## 如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 ([www.xiantao love](http://www.xiantao love))。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



## 常见问题

### 1. 为什么需要使用云雨图?

答: 针对几组数据, 想要观察它们的统计量, 可以通过云雨图来直观的比较。

### 2. x、y 轴可以不颠倒吗?

答: 可以通过设置风格中的 **【xy 颠倒】**, 关闭该选项即可。

