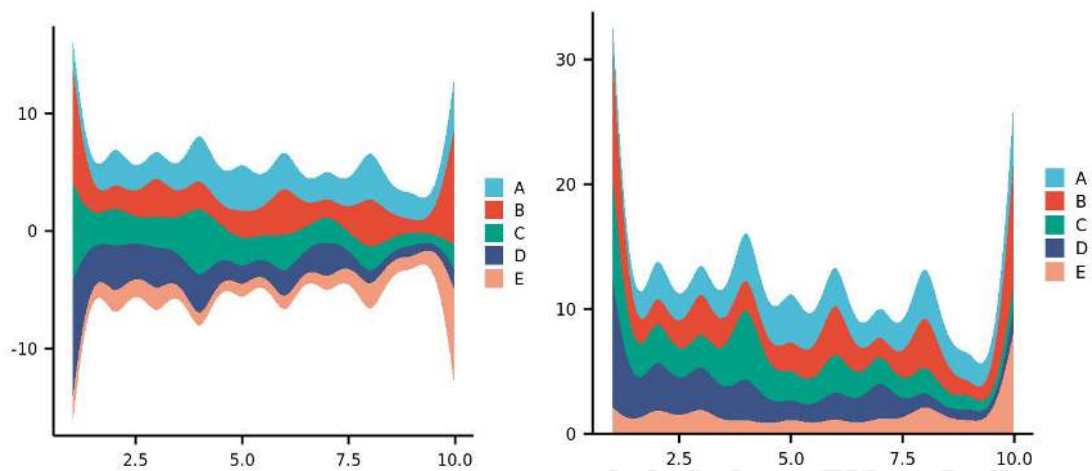


## 基础绘图 - [数据分布] - 量化波形图



网址: <https://www.xiantao.love>



更新时间: 2023.09.08

## 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 .....	7
参数说明 .....	8
波形区域 .....	8
坐标轴 .....	9
标题文本 .....	10
图注 (Legend) .....	11
风格 .....	12
图片 .....	13
结果说明 .....	14
主要结果 .....	14
方法学 .....	15
如何引用 .....	16
常见问题 .....	17

## 基本概念

- 量化波形图：也叫河流图，是堆积面积图的一种变体，在展示多类别及波动幅度大的数据时，可读性更强，外表也更美观，通过表示选择镜像（双向流动）、堆叠（单向流动）和百分比形式的波形，以展示不同组数据的波动和变化趋势。

## 应用场景

- 查看比较不同组的数据随度量梯度波动和变化趋势
- 以及对比不同组的数据随度量梯度变化过程的整体占比
- 其他…

## 分析过程

上传数据 ➡ 数据处理(清洗) ➡ ggstream 计算波形的流动位置和整体形状  
➡ 可视化

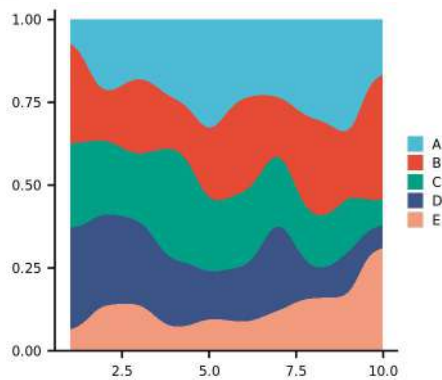
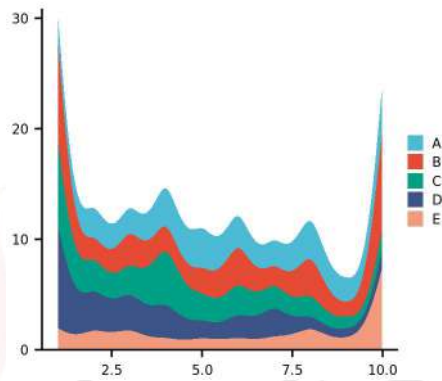
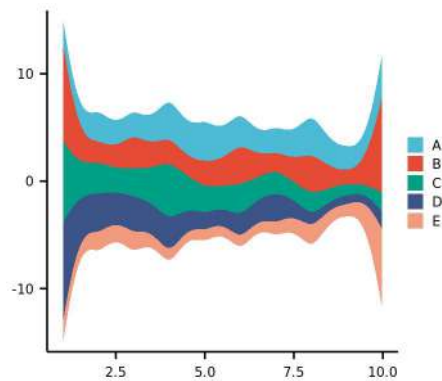
- 数据格式：（具体数据格式要求可以看后面过程的“数据格式”部分）
  - 数据第 1 列必须为数值类型，对应波形的变化度量（可以是时间、温度等）
  - 数据第 2 列必须为数值类型，对应波形的流动大小
  - 数据第 3 列及以后必须是分类类型，对应波形的分组内容

- 必须提供三列；至少两行，至多 1000 行

	A	B	C
1	time	rank_score	group
2	1	1	A
3	1	5	B
4	1	4	C
5	1	5	D
6	1	1	E
7	2	3	A
8	2	2	B
9	2	3	C
10	2	4	D
11	2	2	E
12	3	2	A
13	3	3	B
14	3	2	C
15	3	3	D
16	3	2	E

- 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
  - .....
- 可视化：数据清洗后，使用 ggstream 包和 ggplot2 包进行可视化

## 结果解读



- 上图 1 选择镜像展示，每个不同颜色的波形呈双向流动（上下流动/堆成流动）堆叠成一个整体，注意整体堆叠的总和（即 y 轴上的值）不是数值的总和；上图 2 选择堆叠展示，等同于面积堆叠图，y 轴的数值近似等于变量数据堆叠后的总和；上图 3 选择百分比展示，以百分比的形式显示每一种变量

的整体占比，y 轴对应所占比值。

- 量化波形图波形的颜色表示不同变量/基因/样本，对应数据第 3 列。
- 量化波形图波形的宽窄代表在对应刻度值时观测值的大小，对应数据第 1、2 列，例如：上图中 D 组在 2 这一时刻的观测值是 4，对应波形较宽，在 8 这一时刻的观测值是 1，对应波形较窄，也就是说 D 组随着时间变化是在逐渐减小的。



## 数据格式

	A	B	C
1	time	rank_score	group
2	1	1	A
3	1	5	B
4	1	4	C
5	1	5	D
6	1	1	E
7	2	3	A
8	2	2	B
9	2	3	C
10	2	4	D
11	2	2	E
12	3	2	A
13	3	3	B
14	3	2	C
15	3	3	D
16	3	2	E

	A	B	C
1	time	rank_score	group
2	1	1	A
3	2	3	A
4	3	2	A
5	4	4	A
6	5	4	A
7	6	3	A
8	7	2	A
9	8	4	A
10	9	2	A
11	10	2	A
12	1	5	B
13	2	2	B
14	3	3	B
15	4	2	B
16	5	2	B

数据要求：

- 第一列是用于绘图的横坐标（一个变量代表一个时间度量/其他梯度度量），第二列是波形的变化数值（波形的流动趋势），第三列是波形对应分组，**数据至少需要 3 列，2 行**，第一、二列均需要是数值类型，第三列需要是字符类型的分类变量。
- 波形图的数据第一列是对应度量，第三列是分类变量，第二列是对应的观测值，整理数据时，可以选择根据第一列排序或者第三列排序。
- 数据**最多支持 1000 行，3 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
- 数据每一列列名不能重复，不能有空值，不能有不识别的字符
- 第三列分类变量中的分组数量**最多支持 30 组**

## 参数说明

(说明：标注了颜色的为常用参数。)

## 波形区域

波形区域

展示方式 镜像

波峰 0.75

宽度 0.01


颜色

不透明度 1

- 展示方式：指定波形的流动方式，可选择镜像、堆叠和百分比，默认镜像。
- 波峰：（保持整体形状的前提下）按给定的数值大小调整波形的波峰，范围在 0~1 之间，默认 0.75，值越大，波动越平缓，波峰越少，值越小，波动越强烈，波峰越多。
- 宽度：（保持整体形状的前提下）按给定的数值大小调整波形的宽度，范围在 0~1 之间，默认 0.01，值越大，波形越宽，趋近“圆润”，值越小，波形约窄，趋近“尖锐”。
- 颜色：波形的填充色颜色选项，有多少个分类变量（数据第 3 列的分类变量的数量）就会提取多少个颜色，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。
- 不透明度：波形的透明度。0 为完全透明，1 为完全不透明。



## 坐标轴



A settings panel titled '坐标轴' (Coordinate Axis) with a dropdown arrow. It contains three options: '是否显示x轴' (Whether to display x-axis) with an orange toggle switch turned on; '是否显示y轴' (Whether to display y-axis) with an orange toggle switch turned on; and 'x轴标注旋转' (x-axis label rotation) with a dropdown menu showing '0' and a dropdown arrow.

- 是否显示 x 轴：选择即展示 x 轴。
- 是否显示 y 轴：选择即展示 y 轴。
- x 轴标注旋转：可以选择设置 x 轴标注的倾斜角度。

## 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

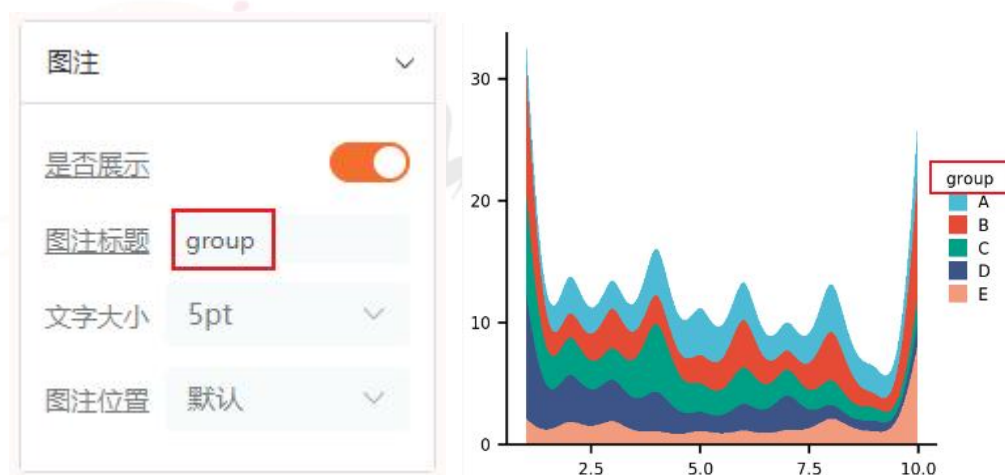
是否展示

图注标题 图注标题内容

文字大小 6pt

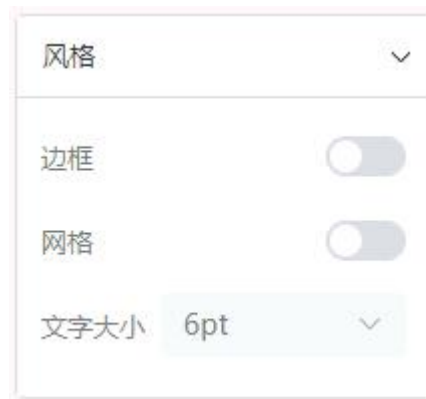
图注位置 默认

- 是否展示：是否展示图注
- 图注标题：可以添加图注标题，如：

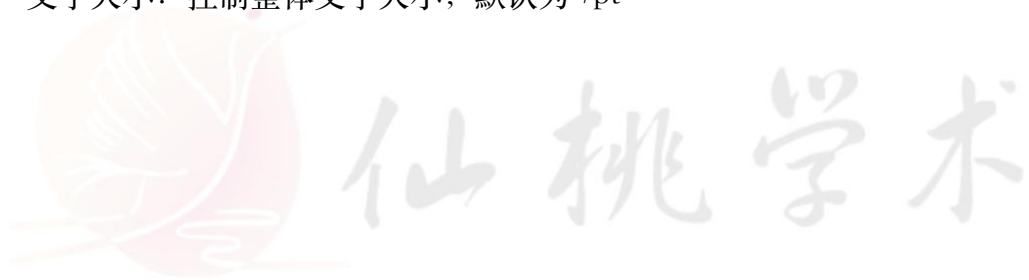


- 文字大小：图注标题文字的大小，默认为 6pt。
- 图注位置：可选择默认、右、上、右上、左上。

## 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内网格的操作
- 文字大小：控制整体文字大小，默认为 7pt



## 图片

图片	▼
宽度 (cm)	6
高度 (cm)	5
字体	Arial ▼

- 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



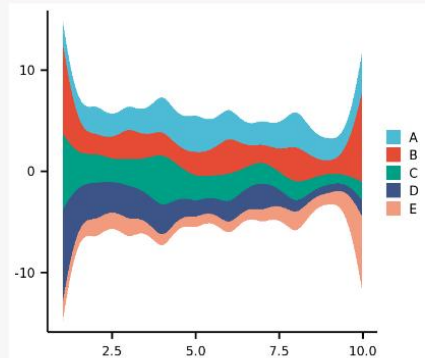
## 结果说明

## 主要结果

### 量化波形图

量化波形图: 围绕着不断变化的中心基线来绘制, 通过使用流动的有机形状 (这些有机形状有点像河流), 量化波形图可用于展示不同类别的数据随着时间(x)的变化。

数据展示: 镜像(mirror)



[量化波形图.pdf](#)

[量化波形图.tiff](#)

[量化波形图.pptx](#)

x轴代表连续变量, 在波形图中显式表示数据点所对应的位置; y轴表示流的高度(对应数据的大小)

在每个x轴上, 波形图将以堆叠的方式显示, 且波形从基线向上或向下延伸(呈镜像); 用于展示波形之间的相对变化, 通过镜像对齐可以更清晰地比较各个波形的趋势。

主要结果格式为图片格式, 提供 PDF、TIFF、PPTX 格式下载

## 方法学

软件：R (4.2.1)版本

R 包：ggplot2 包（用于可视化）、ggstream 包

处理过程：

(1) 将清洗后的数据用 ggplot2 包和 ggstream 包进行量化波形图可视化



## 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. 适用于什么样的数据?

答：例如菌落物种丰度随时间变化的数据，一些有梯度变化的数据，需要观察随度量梯度变化而变化的数据均可。

### 2. 为什么镜像和堆叠的展示方式中，y 轴的值与数据本身的价值不对应?

答：因为这两种方式的计算结果受“波峰”和“宽度”的影响，因此当选择镜像和堆叠这两种展示方式时，y 轴上的数值并不等于数据本身的价值，此时的结果图展示的是变化趋势而非精确变化。

