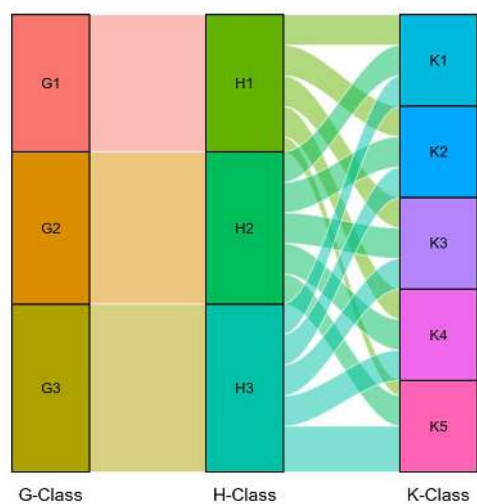


基础绘图 - 桑基图



网址: <https://www.xiantao.love>



更新时间: 2023.03.29

目录	
基本概念	3
基本组成	3
应用场景	3
分析流程	3
主要结果	5
数据格式	6
参数说明	7
桑基图(方块)	7
流向线条	9
标题文本	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

- 桑基图 (Sankey diagram) : 流图 (flow diagram) 的一种, 用来描述流动情况。



基本组成

- 起点: 变量 1
- 终点: 变量 n
- 权重: 数据的大小 (数据行数)
- 节点: 变量中包含的不同的观测值, 可以是数值, 也可以是非数值
- 线条: 变量 1 中的各节点到变量 2 (或到变量 n) 中各节点的连线; 线条粗细表示节点对应数据流量的大小
- 方向/流向: 变量 1 到变量 n 的数据流向 (或变量 n 到变量 1 的数据流向)

应用场景

桑基图: 常应用于生物组成、能源、材料成分、金融等数据的可视化分析

分析流程

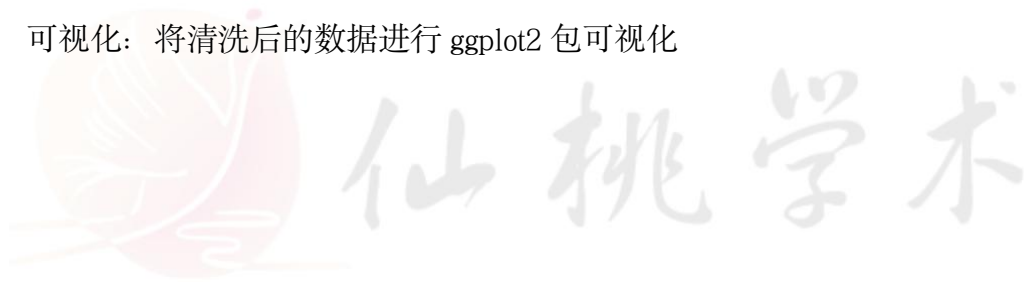
上传数据  数据处理(清洗)  可视化

- 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)

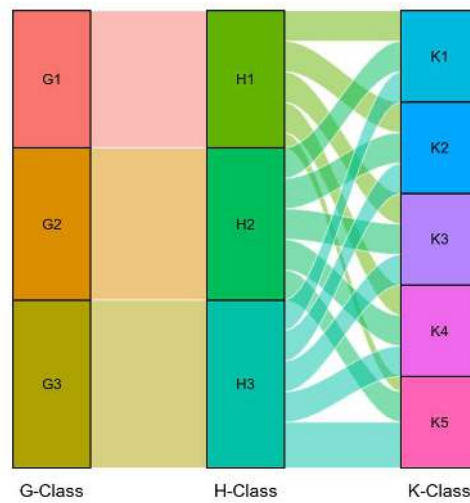
- 上传数据可以是数值类型也可以是分类类型

	A	B	C
1	G-Class	H-Class	K-Class
2	G1	H1	K1
3	G1	H1	K2
4	G1	H1	K3
5	G1	H1	K4
6	G1	H1	K5
7	G1	H1	K1
8	G1	H1	K2
9	G1	H1	K3
10	G1	H1	K4
11	G2	H2	K5
12	G2	H2	K1
13	G2	H2	K2
14	G2	H2	K3
15	G2	H2	K4
16	G2	H2	K5

- 数据处理：对上传数据各列数据进行相应处理
 - 不能包含无法识别与不规则的非字符串类型数据
 -
- 可视化：将清洗后的数据进行 ggplot2 包可视化



主要结果



- 横向坐标（默认从左到右流向）表示变量。
- 纵向坐标（默认从下到上流向）表示权重（数据的大小）。
- 每一根柱子表示一个变量；每根柱子上一个或多个矩形表示不同的节点；节点越宽表示该节点流向下一节点的数据流量越多
- 每一条线对应一个节点。
- 每一种颜色表示将不同起点的节点及其流向下个节点的数据流量分为一类。

数据格式

	A	B	C
1	G-Class	H-Class	K-Class
2	G1	H1	K1
3	G1	H1	K2
4	G1	H1	K3
5	G1	H1	K4
6	G1	H1	K5
7	G1	H1	K1
8	G1	H1	K2
9	G1	H1	K3
10	G1	H1	K4
11	G2	H2	K5
12	G2	H2	K1
13	G2	H2	K2
14	G2	H2	K3
15	G2	H2	K4
16	G2	H2	K5

数据要求：

- 数据至少有 2 列以上，每列至少 2 个观测（2 行，如果不满足这个条件），最多支持 12 列（12 个变量），最多支持 1000 行数据
 - 每一列可以是数值类型，也可以是非数值
 - 单个变量中的观测值小于 100 个(也就是说单个变量不能超过 100 个不同的值)
 - 单个变量不能都没有值(都是缺失)
- 每列数据为一个变量，每一列列名即为桑基图的横向坐标轴（默认从左到右流向）刻度名。图中各变量的顺序与上传数据中各变量的顺序保持一致，若需要调整图中各变量的顺序，需要在上传数据内进行调整，然后再上传数据

参数说明

(说明：标注了颜色的为常用参数。)

桑基图(方块)

柱子(方块) ▾

颜色类型 默认渐变 ▾

填充色 ▾ ▾

描边色 ▾

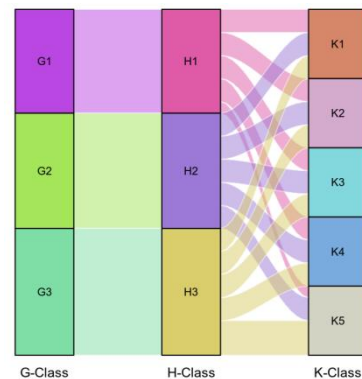
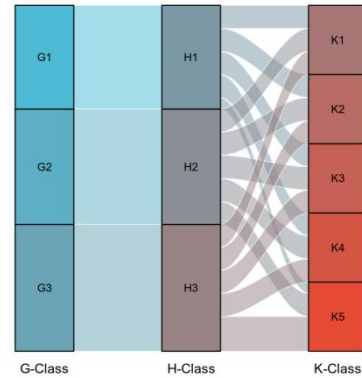
描边粗细 0.50pt ▾

不透明度 1

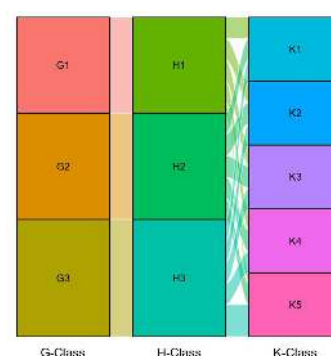
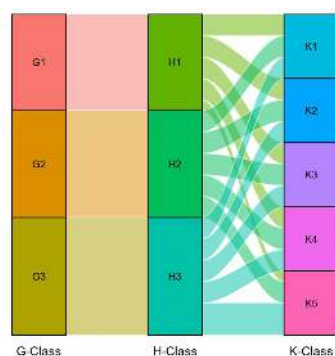
宽度 0.4

- 颜色类型：可以选择修改各柱子(方块)颜色类型，可以通过修改颜色来进行渐变，也可以随机生成不同的颜色；默认渐变：以默认的颜色进行渐变；自定义渐变：可以通过修改填充色的颜色进行渐变；随机颜色：每次随机用不同的颜色进行映射，如下：





- 填充颜色：可以对应修改每个节点的填充颜色（只有在颜色类型选择自定义渐变的时候才会生效）
- 描边颜色：可以对应修改每个节点的描边颜色
- 描边粗细：可以对应修改每个节点的描边粗细
- 不透明度：可以对应修改每个节点的不透明度，1 表示完全不透明，0 表示完全透明
- 宽度：可以对应修改每个节点的宽度（每一个小方块的宽度）0-1 之间，如下：默认为 0.4，右侧为 0.8



流向线条

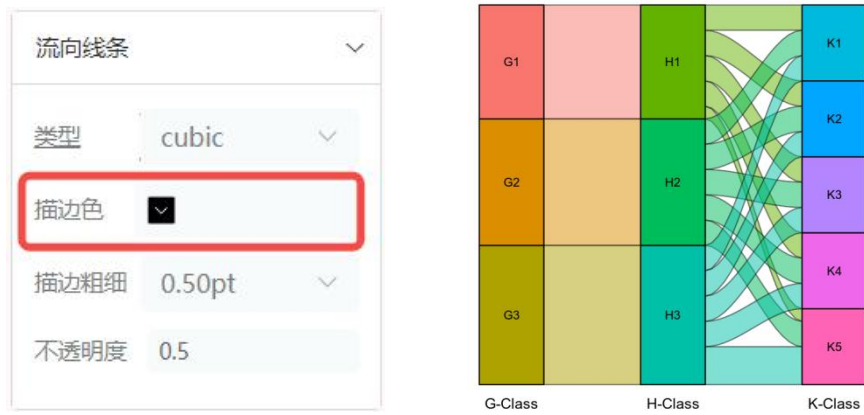
流向线条

类型 cubic
 描边色
 描边粗细 0.50pt
 不透明度 0.5

- **类型**：可以选择绘制不同变量节点与节点之间流向线条的类型，可以选择直线类型、曲线类型等，如下：



- **描边色**：可以修改流向线条的描边颜色，默认为白色，还可以修改为其他颜色，如下：



- 描边粗细：可以修改流向线条的描边粗细，默认为 0.5pt
- 不透明度：可以修改流向线条的不透明度，默认为 0.5，1 表示完全不透明，0 表示完全透明

标题文本

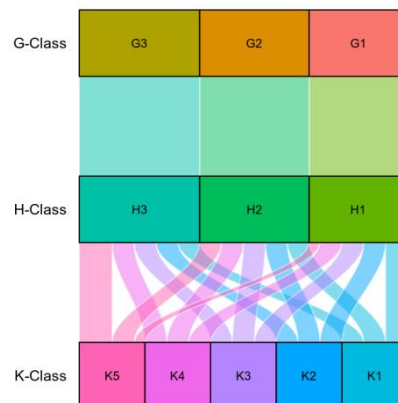
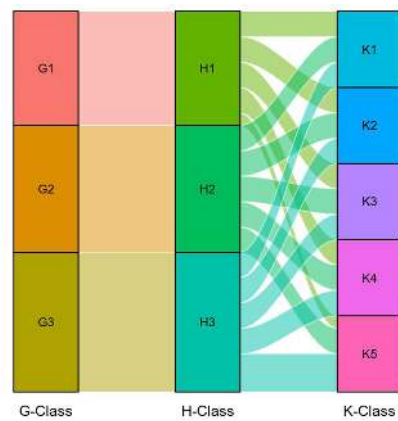
标题	
大标题	大标题内容
x轴标题	x轴标题内容
y轴标题	y轴标题内容

- 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 $[2]$

风格



- xy 颠倒：是否颠倒 xy 坐标，如下：



- 文字大小：针对图中所有文字整体的大小控制
- 字体方向：可以修改每个节点对应文字标志的方向，默认为纵向，还可以选择横向

图片

图片	▼
宽度 (cm)	6
高度 (cm)	6
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

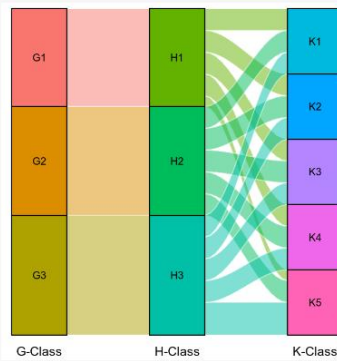


结果说明

主要结果

桑基图

桑基图(Sankey): 用于展示数据在各个分类中组成情况和对应情况



桑基图.pdf

桑基图.tiff

- (1) 横向坐标表示中积层(变量)
- (2) 纵向坐标表示各变量对应的各个值(节点)
- (2) 每个变量对应的一个或多个值(矩形)表示节点 (变量的每一个分类(变量对应的不同的值))
- (3) 连接不同变量之间的条带(线条)表示分支(分流), 分支的宽度对应数据流量的大小

方法学

桑基图可视化在 R 4.2.1 中进行

涉及的 R 包: ggalluvial 包(用于整理数据与可视化), ggplot2 包(用于可视化)

处理过程:

- (1) 数据按照 ggalluvial 包要求整理
- (2) 用 ggplot2 和 ggalluvial 进行可视化



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

