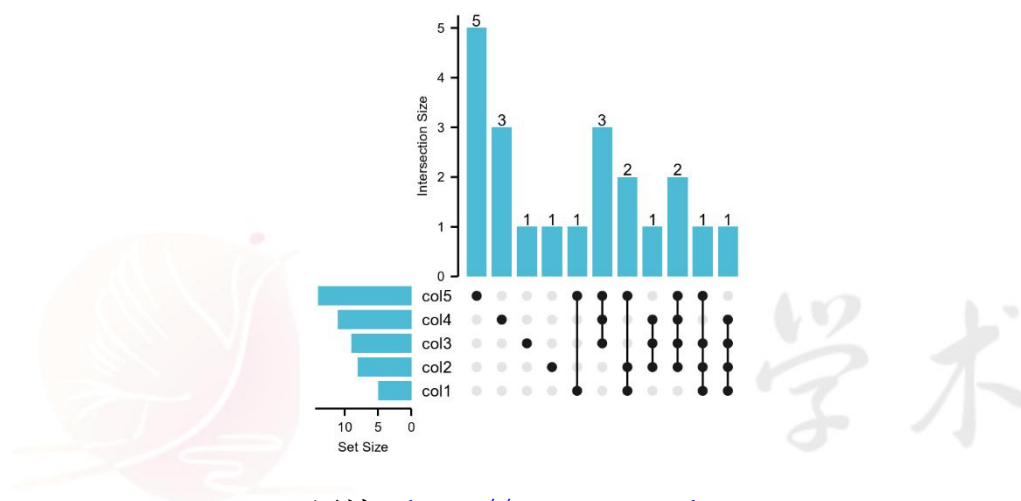


基础绘图 - UpSet 图



网址: <https://www.xiantao.love>



更新时间: 2023.03.15

目录

基本概念	3
应用场景	4
主要结果	5
数据格式	7
参数说明	8
数据处理	8
类型	8
柱(上部分)	9
风格	9
图片	10
结果说明	11
主要结果	11
方法学	12
如何引用	13
常见问题	14

基本概念

- UpSet 图：用于展示不同组之间共有和特有部分的一种可视化方式，是韦恩图的另外一种替代方式。韦恩图比较适合 2-5 组的分析和可视化，而超过 5 组时，韦恩图已经很难完成，此时可以用 UpSet 图来进行分析和可视化。UpSet 图也可以进行少于 5 组的分析和可视化。



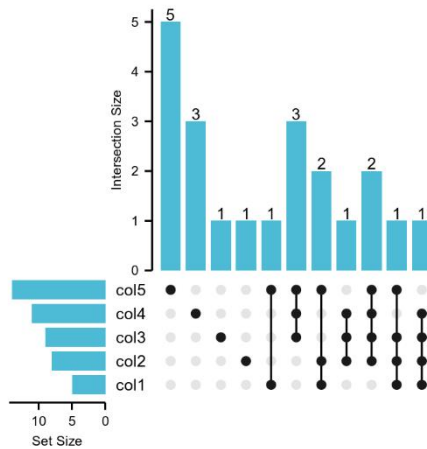
应用场景

UpSet 图主要用于展示 3-7 组（目前设置最大的是 7 组）之间的共有和特有的情况。比如：

- 多个数据集分析后的差异分子列表共有和特有的情况
- miRNA 多个数据库分析后的结果
- 其他 ...

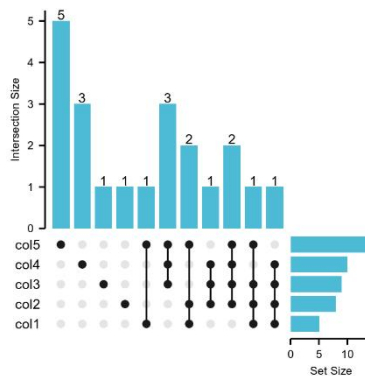


主要结果

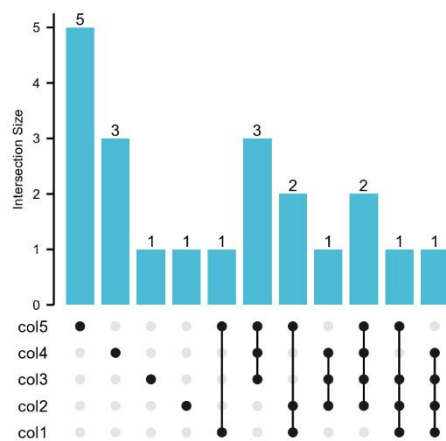


UpSet 图主要包含 3 个部分：

- 上部分为各个分组独有和共有的数量，比如第一个柱子代表在 col5 分组独有 5 个，第 6 个柱子为 col5 分组、col4 分组和 col3 分组共有但是其他组没有的数量，共 3 个。
- 下部分为各个分组独有和共有的分类情况，分别对应上部分的柱子，连线代表共有，比如第 6 列代表 col5 分组、col4 分组和 col3 分组共有但是其他组没有。
- 左部分（或者右部分）代表各个分组中包含的没有重复的元素数量。
- 可视化形式：选择参数【类型】，如下：



类型选择-上下右



类型选择-上下

数据格式

	A	B	C	D	E
1	col1	col2	col3	col4	col5
2	a	b	d	e	a
3	b	c	e	f	b
4	c	d	f	g	c
5	d	e	g	h	d
6	e	f	h	i	ee
7		g	i	j	f
8		h	j	k	g
9		ddd	k	ll	hh
10			aaa	mm	i
11				nl	j
12				k	k
13					l
14					m
15					n

数据要求：

- 至少需要提供 3 - 7 列的数据，且至少 2 行。
- 每一列代表一个分组，列名不能重复。
- 除了第一行的列名，之后的内容都是对应分组含有的内容。
- 工具对空格以及所有的特殊字符都敏感（请仔细检查数据，尤其是最后带有空格），并且分析时会默认剔除掉单个分组下重复的元素。
- 最多 30000 行，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。

参数说明

(说明：标注了颜色的为常用参数。)

数据处理



- 是否删除每一组内的重复元素，默认删除。

类型



- 类型：Upset 图的展示形式，可以选 上下左、上下右、上下。

柱(上部分)



- **填充色**：柱子的填充色颜色选项，只取第一个颜色。受配色方案全局性修改。
- **描边色**：柱子的描边色，只取第一个颜色，默认黑色。不受配色方案全局性修改。
- **描边粗细**：柱子描边的粗细，默认为 0.00pt。
- **不透明度**：柱子的透明度。0 为完全透明，1 为完全不透明。
- **宽度**：柱子的宽度控制，默认 0.8。

风格



- **边框**：是否添加外框

- 网格：是否添加网格
- 文字大小：针对图中所有文字整体的大小控制

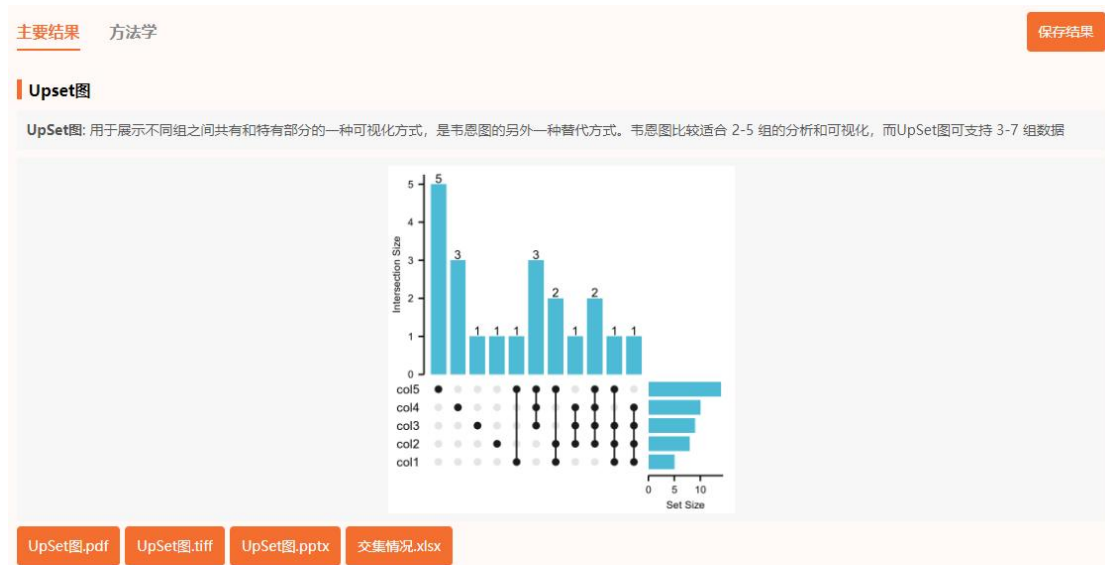
图片



- 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



主要结果格式为图片格式, 提供 PDF、TIFF 格式下载。

	A	B	C	D	E	F
1	elements	col1	col2	col3	col4	col5
2	a	1	0	0	0	1
3	b	1	1	0	0	1
4	c	1	1	0	0	1
5	d	1	1	1	0	1
6	e	1	1	1	1	0
7	f	0	1	1	1	1
8	g	0	1	1	1	1
9	h	0	1	1	1	0
10	ddd	0	1	0	0	0
11	i	0	0	1	1	1
12	j	0	0	1	1	1
13	k	0	0	1	1	1
14	aaa	0	0	1	0	0
15	ll	0	0	0	1	0
16	mm	0	0	0	1	0
17	nl	0	0	0	1	0
18	ee	0	0	0	0	1
19	hh	0	0	0	0	1
20	l	0	0	0	0	1
21	m	0	0	0	0	1
22	n	0	0	0	0	1

➤ 另外, 提供交集情况表格 `xlsx` 下载。

■ 第 1 列为上传数据中各组中包含的元素的总的合集。

- 第 2 列以及以后为各个分组, 各个分组下对应的 0 代表 这个分组没有这个元素, 对应的 1 代表 这个分组含有这个元素。

方法学

所有分析和可视化均在 R 4.2.1 中进行

涉及的 R 包: ggplot2 包

处理过程: 分析数据每个组之间的特有和共有的部分, 并且用 ggplot2 包对结果进行可视化。



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)。

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. 为什么图片中的内容被压缩了?

答:

由于文字不会被压缩,如果左侧的文字很多,就会压缩右侧图的内容而导致坐标轴文字重叠。解决方案可以是:

- ① 增加图片宽度/高度;
- ② 修改上传数据对应的列文字(长度)。

2. 工具的分析结果不准确?

答:

有一个自己检验的方法,在“交集情况.xlsx”表格中对各个分组进行过滤和筛选,然后单独拿一个元素出来,在自己上传的数据中进行搜索(搜索的时候不要选中任一 一个区域,如果选择了这个区域,excel 会只在这个区域搜索),看看是否在这个交集列对应的原始列中存在。

每一列对应的元素数量对不上,很可能这一列含有了重复的内容。工具在进行分析和可视化前,默认都会先对每一列进行去重的处理。

3. 如何修改图中每组的顺序?

答:

可以在上传数据中调整列的顺序,就能调整图中下部分各个分组的顺序。

