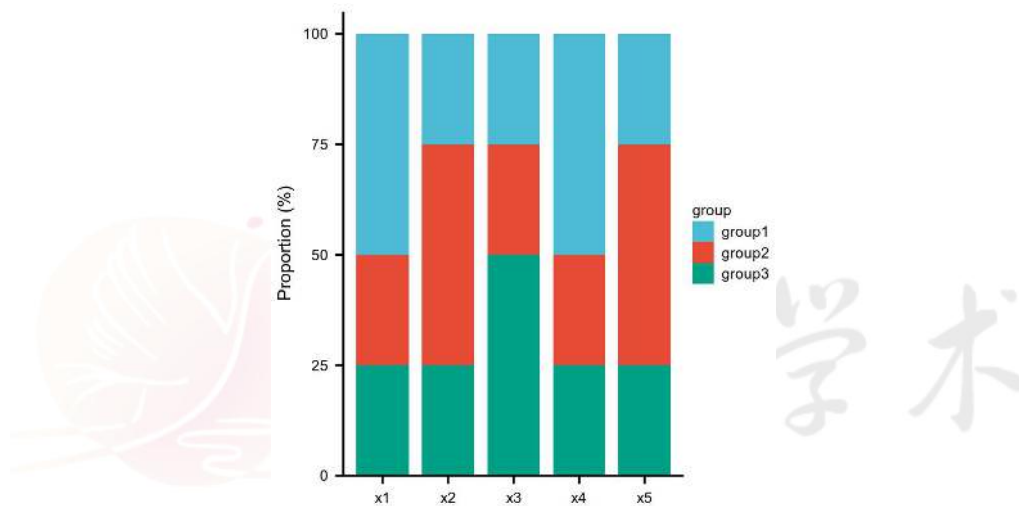


## 基础绘图 - 叠加柱状图-统计



网址: <https://www.xiantao.love>



更新时间: 2023.03.17

## 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 .....	7
统计 .....	7
柱 .....	9
连线 .....	10
标注 .....	11
标题文本 .....	12
图注 (Legend) .....	13
坐标轴 .....	13
风格 .....	14
图片 .....	15
结果说明 .....	16
主要结果 .....	16
补充结果 .....	17
方法学 .....	18
如何引用 .....	19
常见问题 .....	20

## 基本概念

- 柱状图：用柱子的高度或者柱子的相对高度来表示数据的大小情况
- 叠加（叠加）柱状图：
  - **叠加比例柱状图**：用于查看不同分类中 分组的组成比例情况
  - **叠加数值柱状图**：用于查看不同分类中 分组数值的差异。与叠加比例柱状图的差别在于：叠加比例柱状图每组都会计算每个分组的比例情况

## 应用场景

叠加柱状图主要用来展示数据各分类（每一列）中各分组（每一行）的数据分布情况。常应用于数据量对比

## 分析过程

上传数据 ➡ 数据处理(清洗) ➡ 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）
  - 数据**第 1 列**为**分类类型**，对应叠加柱状图 **x 轴**
  - 数据**第 2 列**为**分类类型**，表示**分组/分类**

	A	B
1	x	group
2	x1	group1
3	x2	group2
4	x3	group3
5	x4	group1
6	x5	group2
7	x1	group3
8	x2	group1
9	x3	group2
10	x4	group3
11	x5	group1
12	x1	group2
13	x2	group3
14	x3	group1
15	x4	group2
16	x5	group3

➤ 数据处理：对第 1、2 列分类类型数据进行相关处理

■ 不能含有无法识别的特殊字符或者是非字符

➤ 分析：对第 1、2 列分类类型数据进行相关分析

■ 统计检验

◆ 分组数量统计

分组数量统计表格			
x	group1	group2	group3
x1	2	1	1
x2	1	2	1
x3	1	1	2
x4	2	1	1
x5	1	2	1

◆ 统计检验 p 值

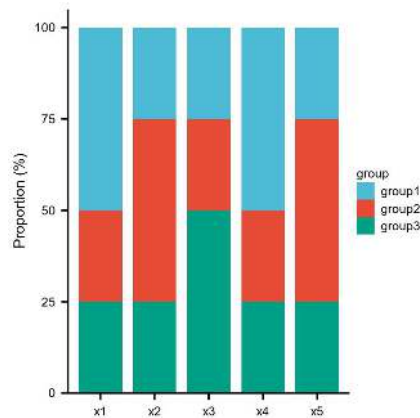
检验p值表格				
分组	变量	统计量	检验p值	检验方法
group	x		1	Fisher test

分组数量统计表格：表中展示了变量（数据第一列）在分组（数据第二列）中的数量统计

检验p值表格：· 数据中不存在level不满足 理论频数>5或者总样本数>40 的条件，建议选用Fisher精确检验(Fisher test)

➤ 可视化：将分析后得到的分组数据进行 ggplot2 包可视化

## 结果解读



### ➤ 叠加柱状图-统计

- 横向坐标表示分类(对应上传数据第 1 列)
- 纵向坐标表示各分组的百分比(默认)(分组(上传数据第 2 列)在各分类(上传数据第 1 列)中的所占比值)、或频数

- 可以直观比较不同分类中不同分组的占比情况。
- 一种颜色表示一个分组。

## 数据格式

	A	B
1	x	group
2	x1	group1
3	x2	group2
4	x3	group3
5	x4	group1
6	x5	group2
7	x1	group3
8	x2	group1
9	x3	group2
10	x4	group3
11	x5	group1
12	x1	group2
13	x2	group3
14	x3	group1
15	x4	group2
16	x5	group3

数据要求：

- 数据需要 2 列（必须是 2 列），每列至少 2 个观测，最多支持 20000 行数据
  - 第 1、2 列均为分类类型
  - 不能含有无法识别的特殊字符或者是非字符
- 第 1 列作为分类变量，其排列的顺序与上传数据中的顺序一致，如果需要调整，则需要自己调整数据再重新上传数据。对应的分类的个数至少 2 个、最多 10 个不同分类
- 第 2 列作为分类变量，其对应分类的个数至少 2 个、最多 10 个不同分类。

## 参数说明

(说明：标注了颜色的为常用参数。)

## 统计

统计 ▼

统计方法 auto ▼

是否展示 p值 不标注 ▼

- 统计方法：可以选择变量之间进行统计学检验的方法，默认为 auto，表示自动选择合适的统计方法(与分析部分统计检验(补充结果部分有关))

检验p值表格				
分组	变量	统计量	检验p值	检验方法
group	x		1	Fisher test

分组数量统计表格：表中展示了变量（数据第一列）在分组（数据第二列）中的数量统计

检验p值表格：· 数据中不存在level不满足 理论频数>5或者总样本数>40 的条件，建议选用Fisher精确检验(Fisher test)

- 还可以选择卡方检验，连续矫正卡方，Fisher 精确检验的方法

检验p值表格				
分组	变量	统计量	检验p值	检验方法
group	x	2.380952	0.9670513	Chisq test

分组数量统计表格：表中展示了变量（数据第一列）在分组（数据第二列）中的数量统计

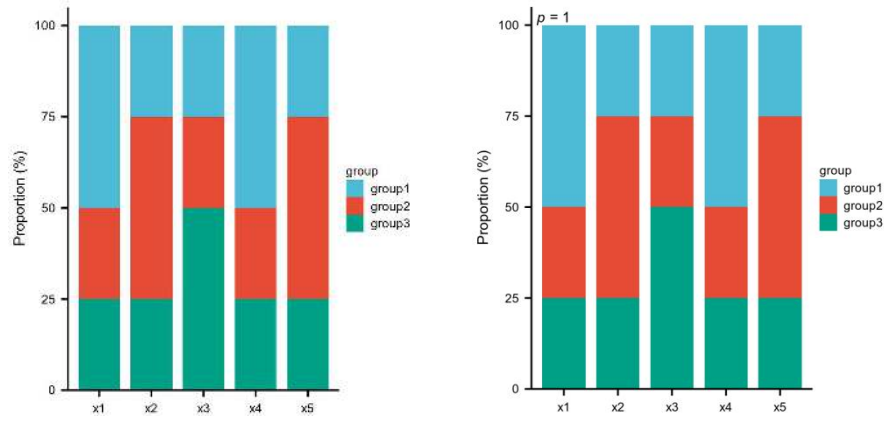
检验p值表格：· 数据中不存在level不满足 理论频数>5或者总样本数>40 的条件，建议选用Fisher精确检验(Fisher test)

检验p值表格				
分组	变量	统计量	检验p值	检验方法
group	x	2.380952	0.9670513	Yates' correction

分组数量统计表格：表中展示了变量（数据第一列）在分组（数据第二列）中的数量统计

检验p值表格：· 数据中不存在level不满足 理论频数>5或者总样本数>40 的条件，建议选用Fisher精确检验(Fisher test)

- 是否展示 p 值：可以选择是否在图形上展示上传数据进行统计学检验的相关结果，如下：左侧为不展示，右侧为展示





## 柱

柱

堆叠类型 百分比

颜色

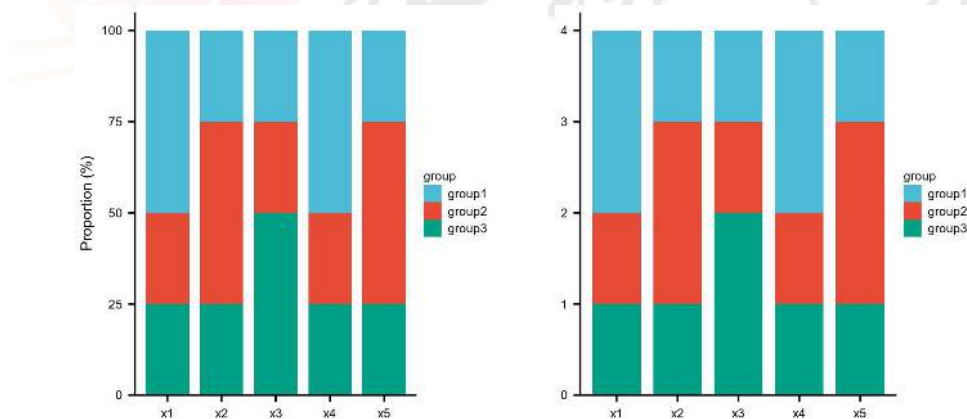
描边色

描边粗细 0.00pt

宽度 0.8

不透明度 1

- 堆叠类型：可以选择修改叠加柱状图堆叠的类型，默认以百分比形式（各分组在各变量中的占比），还可以选择频数（分组(上传数据第2列)在各分类(上传数据第1列)中的个数）的形式，如下左百分比(默认)形式，右为频数：



- 颜色：可以选择并修改叠加柱状图对应柱子的填充颜色
- 描边颜色：可以选择并修改叠加柱状图对应柱子的描边颜色
- 描边粗细：可以选择并修改叠加柱状图对应柱子的描边粗细
- 宽度：可以修改叠加柱状图对应柱子的宽度
- 不透明度：可以修改柱状图每一根柱子的不透明度

## 连线

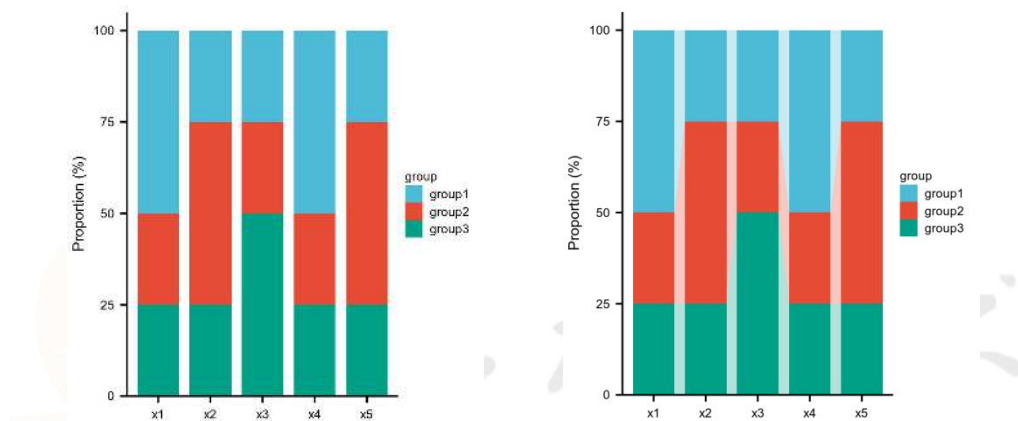
连线
 ⌵

是否展示连线
 ☐

不透明度 0.3

➤ 是否展示连线：可以选择是否对柱状图之间进行连线操作

■ 选择展示：连线的宽度与其连接的柱状图个分组的宽度一致



➤ 不透明度：首先选择展示，则可以修改柱状图间连线的不透明度

## 标注

标注

标注映射
不映射

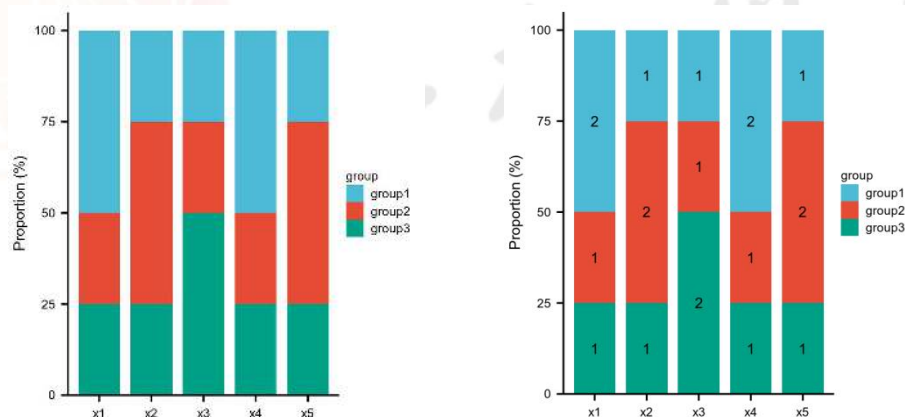
大小
6pt

方向
纵向

标签背景

标签描边
0.75pt

- 展示：可以选择是否对柱状图进行标注映射操作，默认不映射，还可以选择频数映射，如下：



- 大小：可以选择并修改标注文本的大小
- 方向：可以选择并修改标注文本的方向
- 标签背景：当选择对柱状图进行标注映射操作时，可以选择是否对标签进行背景化操作，如下：

标注

是否标注 频数

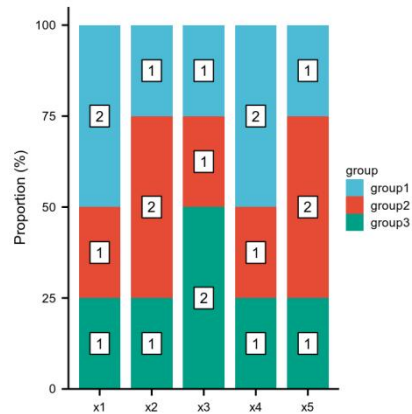
大小 6pt

方向 横向

标签背景

标签描边

0.50pt



- 标签描边：当选择展示标签背景后，可以选择并修改标签的描边粗细

## 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

图注位置

默认

- 展示：可以选择是否展示图注操作
  - 选择展示：将会展示图注
- 图注标题：首先选择展示，则可以修改需要上传的图注标题信息
- 图注位置：首先选择展示，则可以选择展示图注的位置

## 坐标轴

坐标轴

x轴标注旋

转

0

y轴范围+刻度

逗号隔开

- x 轴标注旋转：可以选择 x 轴标注旋转的角度
- y 轴范围+刻度：可以控制 y 轴范围和刻度，可只提供 2 个值来控制范围。  
形如 0.1, 0.2, 0.3 (最小值和最大值不能超过可视化数据范围 20%，如果调整过大可能会无作用)

## 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线的操作
- xy 颠倒：可以选择是否进行 xy 颠倒的操作
- 环形柱状图：可以选择是否进行展示环形柱状图的操作
- 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

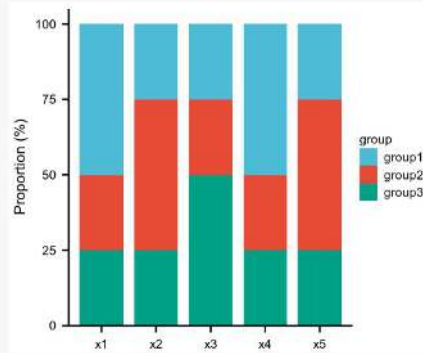


## 结果说明

## 主要结果

### 叠加柱状图-统计

叠加柱状图-统计: 统计数据分类情况后用叠加柱状图来展示结果



叠加柱状图.pdf

叠加柱状图.tiff





## 补充结果

### 分组数量统计与检验

- 统计各个分组中变量的数量
- 评估分类变量的统计检验方法

分组数量统计表格

x	group1	group2	group3
stage1	2	1	1
stage2	1	2	1
stage3	1	1	2
stage4	2	1	1
stage5	1	2	1

检验p值表格

分组	变量	统计量	检验p值	检验方法
group	x		1	Fisher test

分组数量统计表格:

· 表中展示了变量（数据第一列）在分组（数据第二列）中的数量统计

检验p值表格:

· 数据中不存在level不满足理论频数> 5或者总样本数> 40的条件, 建议选用Fisher精确检验(Fisher test)

- 这里提供数据各分类（数据第 1 列）中各分组（数据第 2 列）的数量
- 评估分类变量的统计检验方法

## 方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包：ggplot2、ggalluvial 包（用于可视化）

处理过程：

- (1) 将清洗后的数据进行统计分析，得到各分类中各分组的占比。
- (2) 数据用 ggplot2 包绘制叠加柱状图。



## 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. 参数-统计-统计方法中的“auto”是什么意思?

答：“auto”表示会根据上传数据特点，自动选择适合的统计方法进行分析，如果有需要进行特定统计方法的可以下来选项来选择

