

基于 B 中学学生信息的探索性分析

庄闪闪

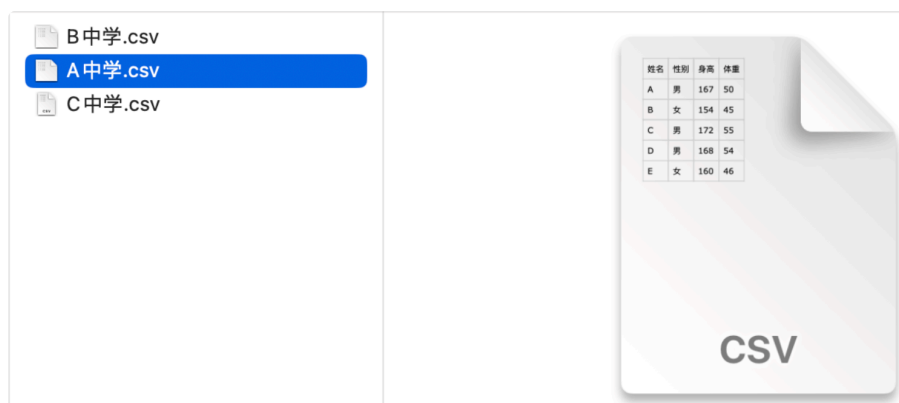
2023/04/12

目录

1	背景	1
1.1	加载数据	2
1.2	探索性分析	3
2	总结	3
3	未来展望	4

1 背景

假设我们得知不同学校学生信息，包括：姓名，身高，体重。如：



* 领导/硕导/博导 * 要求分别给出不同学校的学生信息分析报告。此时，可以使用 R Markdown 开展这项工作。

1.1 加载数据

首先，我们加载 B 中学学生的信息。

```
library(tidyverse)
library(knitr)
df = read.csv(paste0("data/",params$new_title, ".csv", sep=''))
knitr::kable(df, align = "c", caption = '学生信息')
```

表 1: 学生信息

姓名	性别	身高	体重
A	男	156	53
B	女	165	46
C	女	172	51
D	男	178	60
E	男	169	53

1.2 探索性分析

根据该数据，我们对不同性别的身高和体重进行探索性分析。

```
df %>% group_by(`性别`) %>% summarise(" 平均身高" = mean(`身高`),
                                       " 身高标准差" = sd(`身高`),
                                       " 平均体重" = mean(`体重`),
                                       " 体重标准差" = sd(`体重`)) -> df_summary
knitr::kable(df_summary, align = "c",
              caption = " 不同性别的身高和体重情况")
```

表 2: 不同性别的身高和体重情况

性别	平均身高	身高标准差	平均体重	体重标准差
女	168.5000	4.949747	48.50000	3.535534
男	167.6667	11.060440	55.33333	4.041452

```
library(showtext)
library(ggsci)
showtext.auto()
df %>% pivot_longer(`身高`:`体重`,
                    names_to = "Class",
                    values_to = " 数值") %>%

  ggplot() +
  geom_boxplot(aes(性别, 数值, fill=性别)) +
  scale_fill_aaas() +
  facet_wrap(vars(Class), scales = "free")
```

2 总结

这个项目主要测试 RMarkdown 的参数化报告，主要参考：

1. R Markdown 指南第 7 章：使用 R Markdown 开展项目工作

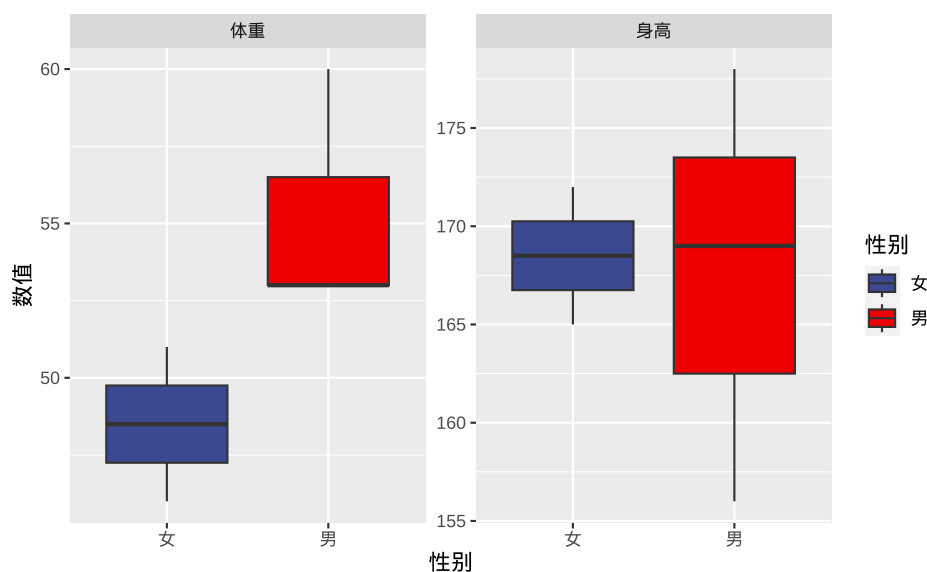


图 1: 不同性别学生的身高和体重分布情况

2. Iterate multiple RMarkdown reports

由于时间有限，这个报告后续没有添加了。

3 未来展望

- 对于不同的数据，虽然可以自动化绘制不同的报表，但是由于数据不同，结果也会不同。不同结果如何达到自动化分析呢？这个问题还需要再商讨。个人觉得如果能把类似 ChatGPT 结合起来，或许真能实现自动给出完整的分析报告了。类似的 ChatGPT 的 R 包有：chatgpt, gptstudio。教程有：1, 2, 供参考。
- 制作该可重复性报告的前提是：不同数据需要一致的数据结构。这点是否可以拓展？是否有专门的包能够自动对数据处理，并给出统一的数据格式？
- 制作这个教程花费较长时间，欢迎大家一键三连呀～教程对应的项目代码已经开源在 GitHub 中，欢迎 Fork 和 star。或者公众号后台回复

[批量制作数据分析报告] 免费获得。