

2151299_ 苏家铭 _hw1

苏家铭

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
# 加载文件
```

```
library(tidyverse)
```

```
## Warning: 程辑包'ggplot2'是用R版本4.3.2 来建造的
```

```
library(mosaic)
```

```
library(ggplot2)
```

```
library(moments)
```

```
library(sjlabelled)
```

```
WVS <- readRDS("0 Visualization_data_WVS.rds")
```

```
WVS$V2A <- as.character(WVS$V2A)
```

```
# 查看共有多少个国家
```

```
num_countries <- length(unique(WVS$V2A))
```

```
print(num_countries)
```

```
## [1] 61
```

```
# 由结果可知，共有 61 个国家
```

```
# 筛选出国家编码变量（V2A）和生活满意度变量（V23）
```

```
WVS_V2A_V23 <- WVS %>% select(V2A, V23)
```

```
# 筛选出的国家编码变量和生活满意度变量保存在变量 WVS_V2A_V23 中
```

```
# 得到□活满意度变量的 median, mean, standard deviation, skewness, kurtosis 统计量。
# 通过这些统计量分析该变量的分布特征。
```

```
# 中位数
```

```
median_value <- median(WVS_V2A_V23$V23)
```

```
# 平均值
```

```
mean_value <- mean(WVS_V2A_V23$V23)
```

```
# 标准差
```

```
sd_value <- sd(WVS_V2A_V23$V23)
```

```
# 偏度
```

```
skewness_value <- skewness(WVS_V2A_V23$V23)
```

```
# 峰度
```

```
kurtosis_value <- kurtosis(WVS_V2A_V23$V23)
```

```
# 输出统计量
```

```
print(paste("Median:", median_value))
```

```
## [1] "Median: 7"
```

```
print(paste("Mean:", mean_value))
```

```
## [1] "Mean: 6.77966839725339"
```

```
print(paste("Standard Deviation:", sd_value))
```

```
## [1] "Standard Deviation: 2.36545405254884"
```

```
# 报错: Error in skewness(WVS_subset$V23) : could not find function "skewness"
```

```
# 偏度需要按装 moment 包
```

```
print(paste("Skewness:", skewness_value))
```

```
## [1] "Skewness: -0.757957138969565"
```

```
print(paste("Kurtosis:", kurtosis_value))
```

```
## [1] "Kurtosis: 3.42200704704269"
```

```
# 通过这些统计量分析该变量的分布特征:
```

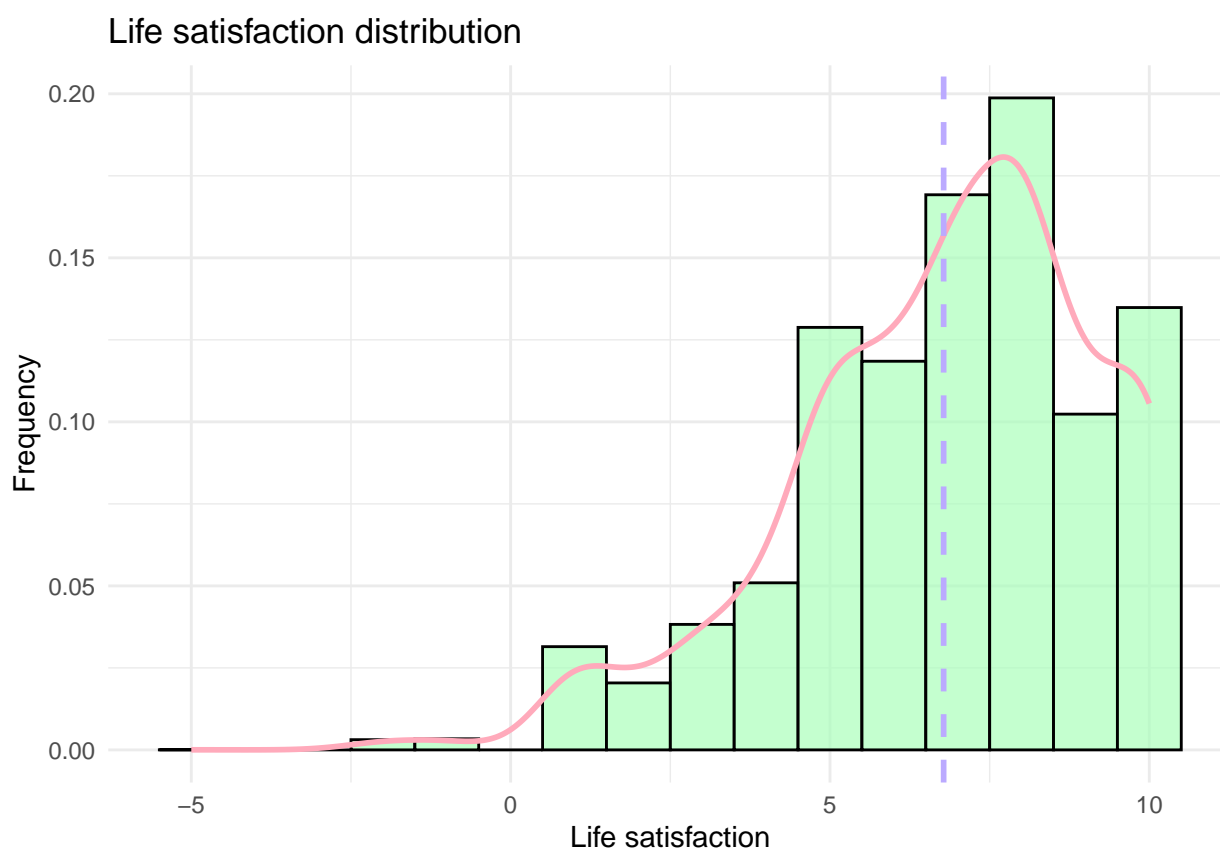
```
# 1. 中位数是 7 表明有一般的观测值小于等于 7, 另一半大于等于 7
```

```
# 2. 平均值是数据集中所有值的总和除以观测值的数量, 是 6.7797 表明数据分布稍向左倾斜
```

3. 标准差是 2.3655 表示数据的变异程度相对较小，大多数观测值集中在均值附近
 # 4. 偏度是-0.758 为负表示数据分布偏向左侧，即分布尾部向左延伸。
 # 5. 峰度为 3.422 表明数据分布相对于正态分布来说更具尖峭性，意味着数据中可能存在一些异常值或尖峭的峰。

绘制生活满意度变量的直方图，并添加密度曲线和均值线
 # 绘制直方图

```
ggplot(data = WVS_V2A_V23, aes(x = as.integer(V23), y = after_stat(density))) +
  geom_histogram(binwidth = 1, fill = "#aaffbb", color = "black", alpha = 0.7) +
  geom_line(stat = "density", color = "#ffaabb", adjust=3, linewidth = 1) +
  geom_vline(xintercept = mean_value, color = "#bbaaff", linetype = "dashed", linewidth = 1) +
  labs(title = "Life satisfaction distribution",
       x = "Life satisfaction",
       y = "Frequency") +
  theme_minimal()
```



通过直方图对变量的分布特征进行文字解读：

1. 数据分布形状：从直方图可以看出，生活满意度数据呈现一种单峰分布，即存在一个

主要的峰值。这意味着大多数人的生活满意度分数集中在某一范围内。

2. 中心趋势：直方图中的垂直线（通常是在均值位置）表示生活满意度的平均值。在这个图表中，可以看到平均值附近有较高的柱状条，这表明平均生活满意度分数约为 6.8 左右。

3. 分散程度：直方图中柱状条的分布宽度反映了生活满意度数据的分散程度。在这里，柱状条相对狭窄，表明大多数人的生活满意度分数相对接近均值，即分散程度较小。

4. 异常值：直方图显示了分布中是否存在异常值或离群值。在这里，可以注意到一些低分值的柱状条，这可能表示一小部分人的生活满意度明显低于平均水平，可能是离群值。

5. 分布特征：总体而言，生活满意度数据呈现了一个中心集中、分散程度相对较小的分布，其中大多数人的满意度分数集中在均值附近。并且满意度满分的人也不少，说明人们对于生活还是较满意的。

统计每个国家的平均生活满意度分数，然后从低到高绘制这些国家

统计每个国家的平均生活满意度分数

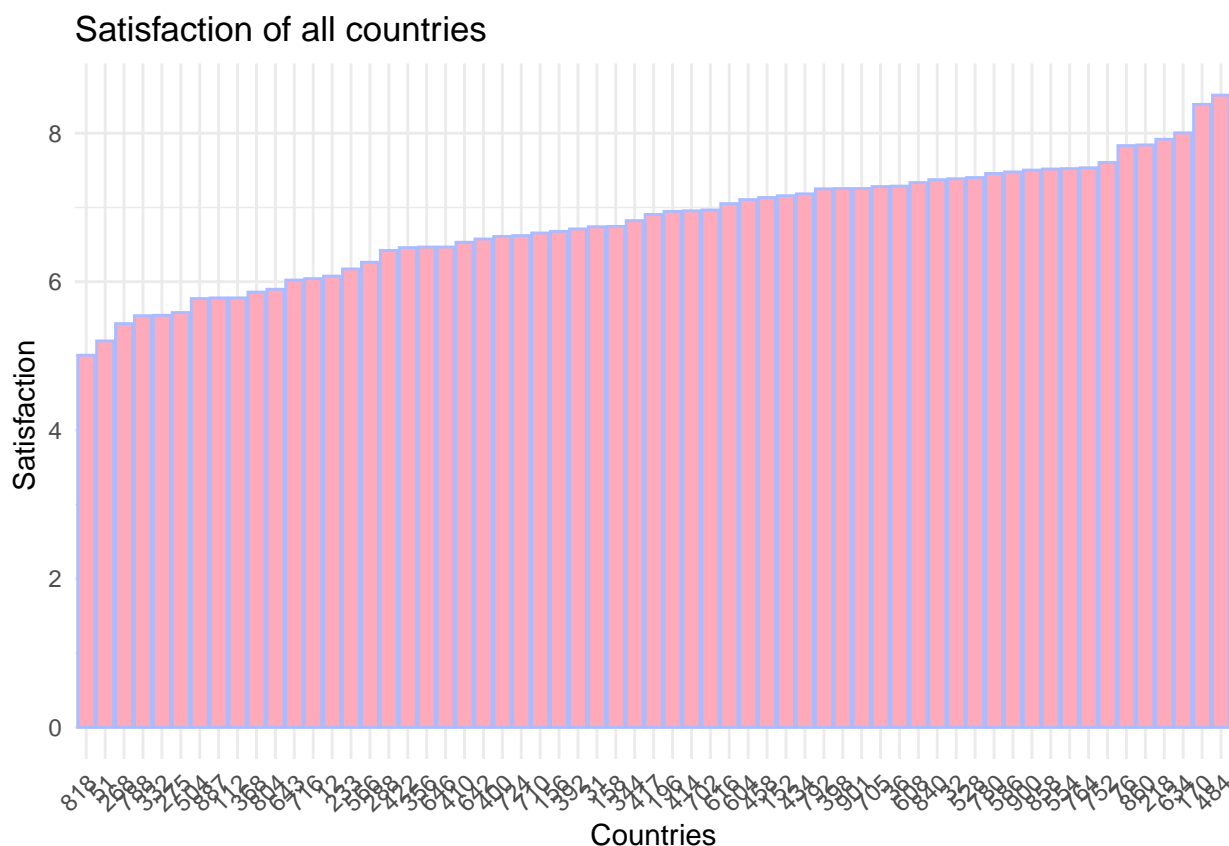
```
country_avg_satisfaction <- WVS_V2A_V23 %>%
  group_by(V2A) %>%
  summarize(avg_satisfaction = mean(V23))
```

从低到高绘制国家

```
country_avg_satisfaction <- country_avg_satisfaction %>%
  arrange(avg_satisfaction)
```

绘制图表

```
ggplot(data = country_avg_satisfaction, aes(x = reorder(V2A, avg_satisfaction),
                                                y = avg_satisfaction)) +
  geom_bar(stat = "identity", fill = "#ffaabb", color = "#aabbff") +
  labs(title = "Satisfaction of all countries",
       x = "Countries",
       y = "Satisfaction ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



由图可知，国家编号为 818 的平均生活满意度最低，国家编号为 484 的平均生活满意度最高

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.