hw1

YangTengchao

导入数据

```
WVS <- readRDS("WVS.rds")
WVS$V2A <- as.character(WVS$V2A)
```

查看一共多少个国家

```
country_list <- unique(WVS$V2A)
country_count <- length(country_list)
print(country_count)
```

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

```
# 共有61个国家
```

筛选国家编码变量（V2A）和生后满意度变量（V23）

```
filter_result <- WVS[, c("V2A", "V23")]
# 这里展示前六行
head(filter_result)
```

```
##   V2A V23
## 1  12   8
## 2  12   5
## 3  12   4
## 4  12   8
## 5  12   8
## 6  12   7
```

□活满意度变量的 median, mean, standard deviation, skewness, kurtosis 统计量

```
if (!require(moments)) {
  install.packages("moments")
}
```

```
## 载入需要的程辑包: moments
```

```
library(moments)

V23 <- WVS$V23

median_V23 <- median(V23)
mean_V23 <- mean(V23)
sd_V23 <- sd(V23)
skewness_V23 <- skewness(V23)
kurtosis_V23 <- kurtosis(V23)

cat("median:", median_V23, "\n")
```

```
## median: 7
```

```
cat("mean:", mean_V23, "\n")
```

```
## mean: 6.779668
```

```
cat("standard deviation:", sd_V23, "\n")
```

```
## standard deviation: 2.365454
```

```
cat("skewness:", skewness_V23, "\n")
```

```
## skewness: -0.7579571
```

```
cat("kurtosis:", kurtosis_V23, "\n")
```

```
## kurtosis: 3.422007
```

中位数为7说明有一半的受统计对象的生活满意度低于7，而另一半人的生活满意度高于7

平均值为6.779668说明在中位数为7的基础上，生活满意度低于7的对象打分较低，而高于7的对象打分也不高。数据分布向左倾斜

标准差为2.365454说明数据差异较小，大多分布在均值附近

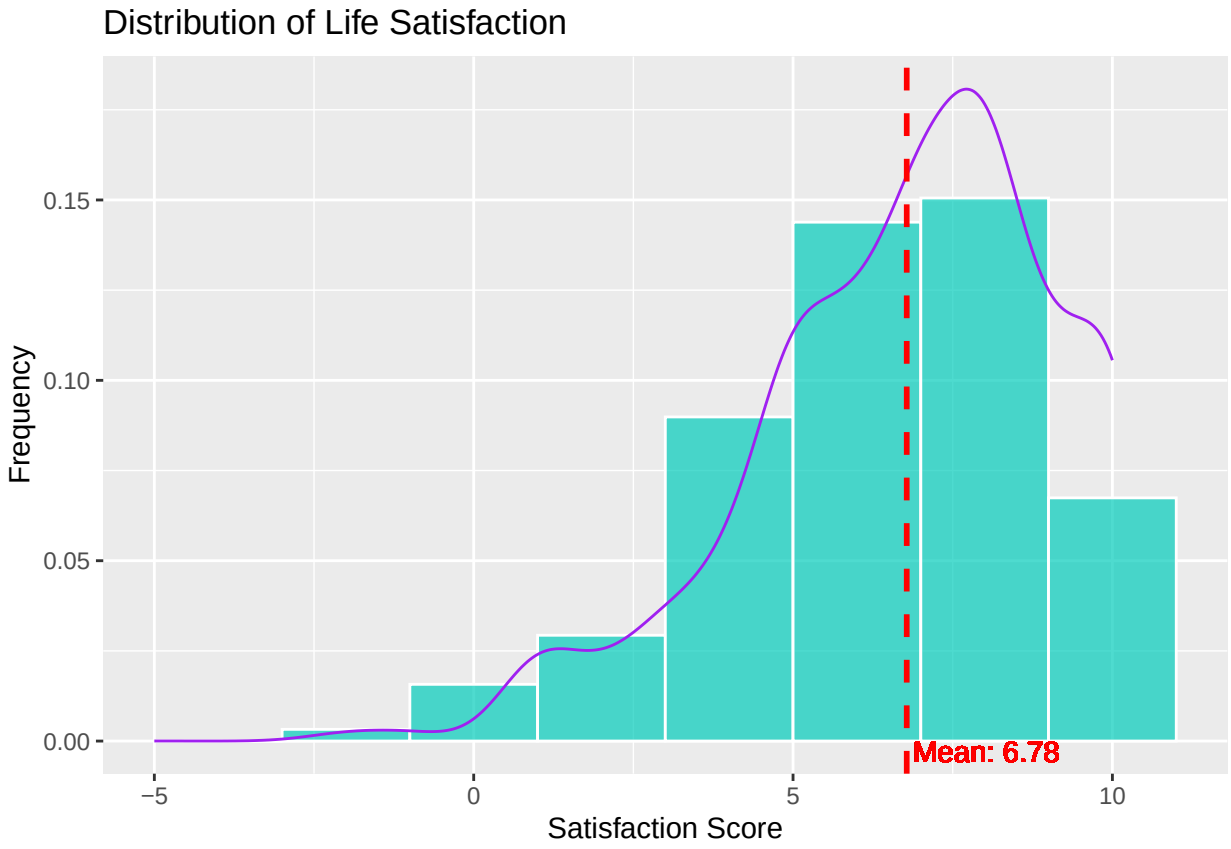
偏度为-0.7579571说明数据相对于正太分布略微向左偏斜，数据的左侧尾部比右侧尾部更重。

峰度为3.422007说明数据相对于正态分布更具有尖峰特征，即在中心处更为尖锐

生活满意度直方图

```
library(ggplot2)

p <- ggplot(WVS, aes(x=as.integer(V23))) +
  geom_histogram(aes(y = after_stat(density)),
        binwidth = 2, fill = "#00ccbb",
        color = "white",
        alpha = 0.7) +
  geom_density(color = "purple",
        linetype = "solid",
        adjust=3) +
  geom_vline(aes(xintercept = mean(V23)),
        color = "red",
        linetype = "dashed",
        linewidth = 1) +
  labs(title = "Distribution of Life Satisfaction",
       x = "Satisfaction Score",
       y = "Frequency") +
  geom_text(aes(label = paste("Mean:", round(mean(V23), 2))),
       x = mean(WVS$V23) + 0.1, y = 0,
       color = "red", hjust = 0, vjust = 1)
print(p)
```




```
# 可以看出直方图呈现左偏态，左侧有较长的尾巴，而右侧的尾巴相对较短
# 低数值的生活满意度出现频率较高，而高数值的生活满意度出现频率较低
# 从而使得平均值小于中位数
# 这里呈现一种单峰分布，说明大多数人的生活满意度集中在6.78的区间附近
```

统计每个国家人民的平均生活满意度分数

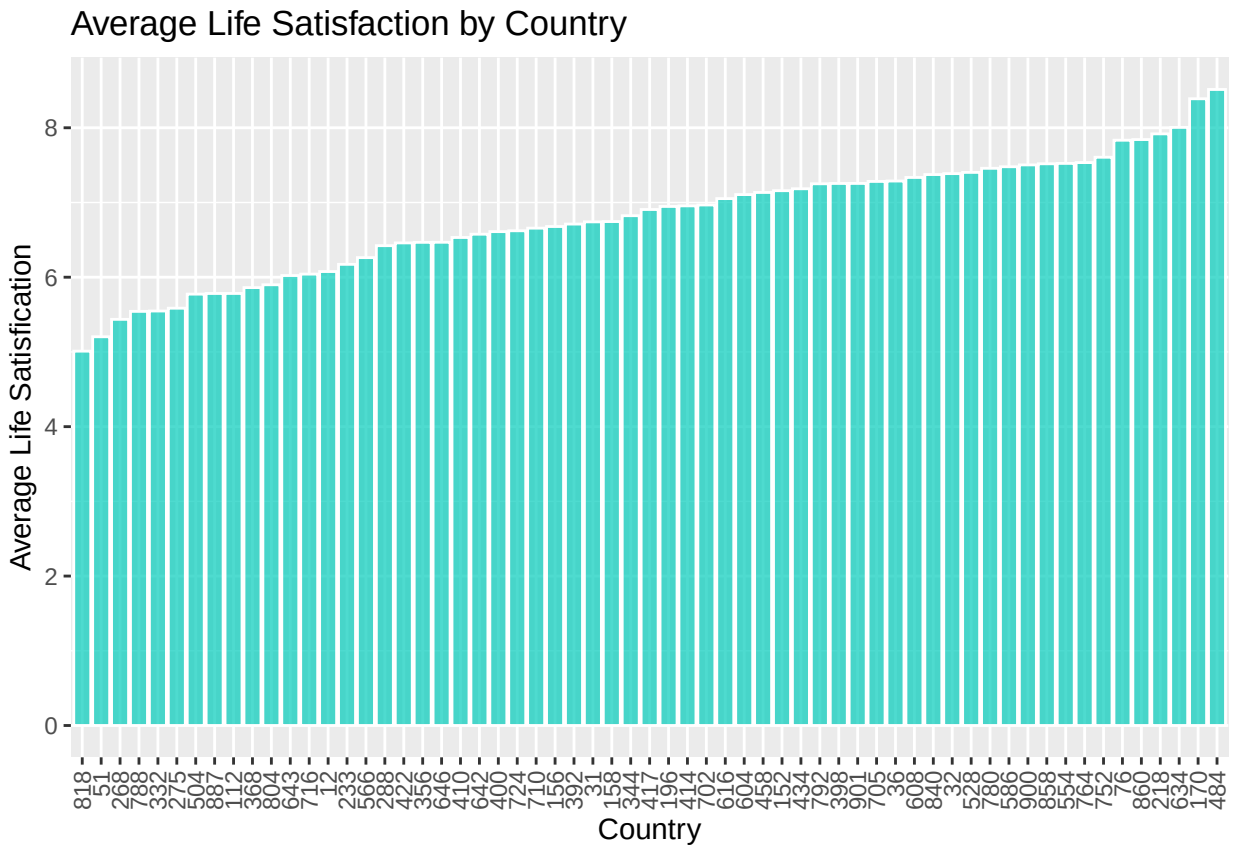
```
country <- WVS$V2A
satisfication <- WVS$V23

avg_satisfication <- aggregate(satisfication ~ country, WVS, mean)

avg_satisfication <- avg_satisfication[order(avg_satisfication$satisfication), ]

p <- ggplot(avg_satisfication,
        aes(
            x = reorder(country, satisfication),
```

```
            y = satisfication)) +
  geom_bar(stat = "identity", fill = "#00ccbb", color = "white", alpha = 0.7) +
  labs(title = "Average Life Satisfaction by Country",
       x = "Country",
       y = "Average Life Satisfication") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = .5))
print(p)
```



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