

QSS_Chapter3_Exercise

my

2021/12/15

目录

1 改变对待同性恋婚姻的看法（再探）	2
1.1 研究 1 中，计算控制组中受访者感受评分在第一波和第二波之间的相关性	2
1.2 使用研究 2 重复上一题	2
1.3 找出异常值; 创建散点图，可视化研究 2 控制组中第一波回答和后续回答之间的关系	3
1.4 比较 ccap2012 和 gayreshaped 两个数据集	5
1.5 通过分位数图比较两个样本分布	5
2 中国和墨西哥的政治效力	6
2.1 用条形图表示中国和墨西哥的受访者回答分布，计算每个国家受访者回答平均分	6
2.2 检查中国和墨西哥的分数差异是否由年龄分布差异引起	8
2.3 对比自我评估的问题分数和情景问题的分数（计算两个国家自我评估比 Moses 分数低的比 例）	10
2.4 创建一个变量记录受访者自我评估相对于情景评估的排名	11
2.5 老年人是否比年轻人更严重?	12
3 联合国大会投票表决	14
3.1 柏林墙倒塌的前后 10 年（即 1980 年和 2000 年）的各国理想分数分布	14
3.2 随着时间的推移，与美国投票一致的国家的数量变化	16
3.3 绘制美国和俄罗斯的理想点随时间变化的图象	17
3.4 曾属于苏联的国家和不属于苏联的国家有什么不同	18
3.5 观察曾属于苏联的国家和其他所有国家的理想点在过去几年中的变化	19
3.6 国家是否仍然分为两个意识形态派别?	20

1 改变对待同性恋婚姻的看法（再探）

```
setwd("D:/QSS/Chapter3_Measurement/Exercise")
gayreshaped <- read.csv("gayreshaped.csv")
ccap2012 <- read.csv("ccap2012.csv")
# View(gayreshaped)
# View(ccap2012)
table(gayreshaped$treatment)

##
##                               No Contact
##                               6441
##      Recycling Script by Gay Canvasser
##                               1046
##      Recycling Script by Straight Canvasser
##                               1039
##      Same-Sex Marriage Script by Gay Canvasser
##                               2389
## Same-Sex Marriage Script by Straight Canvasser
##                               1033
```

1.1 研究 1 中，计算控制组中受访者感受评分在第一波和第二波之间的相关性

```
## 将研究 1 中控制组的数据进行分集
s1nc <- subset(gayreshaped, subset = (study == "1" & treatment == "No Contact"))
## 计算第一波和第二波的感受评分的相关性，并将 use 参数设置为 "complete.obs" 以处理缺失值
## (listwise 的删除方式)
cor(s1nc$therm1, s1nc$therm2, use = "complete.obs")

## [1] 0.9975817
```

1.2 使用研究 2 重复上一题

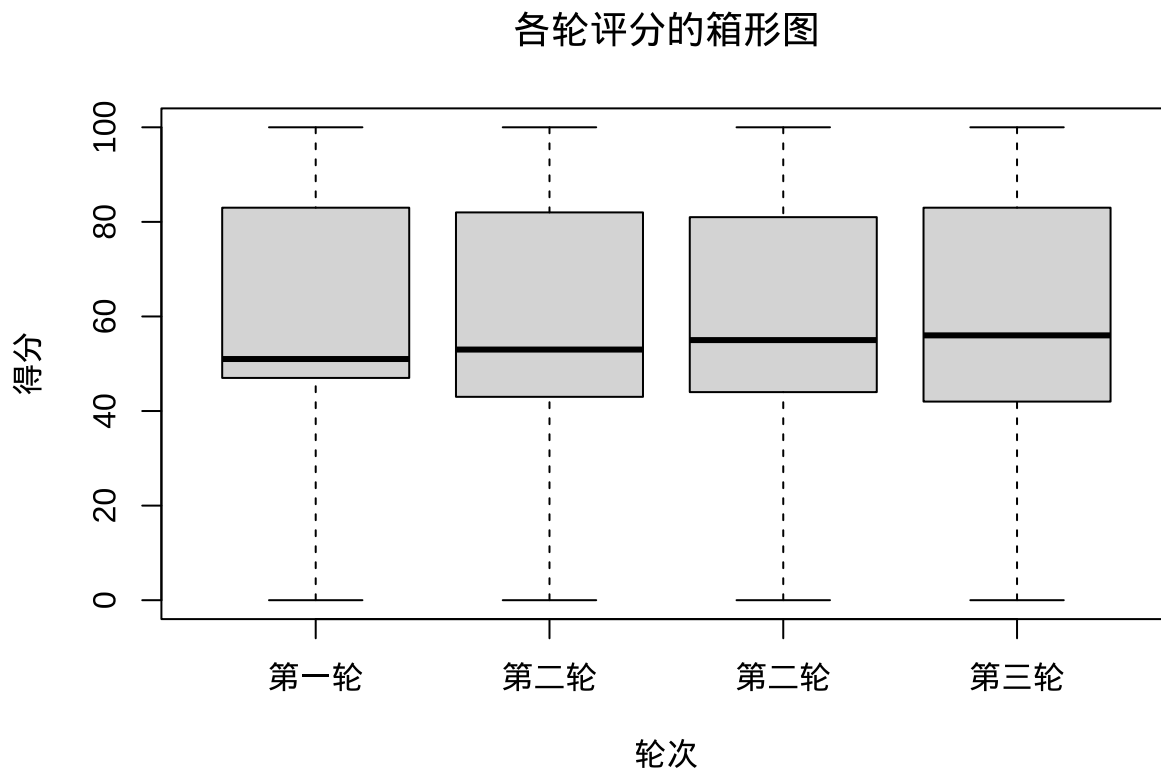
```
## 将研究 2 中控制组的数据进行分集
s2nc <- subset(gayreshaped, subset = (study == "2" & treatment == "No Contact"))
## 比较所有调查中感受评分的相关性，并将 use 参数设置为 "pair.complete.obs" 以处理缺失值
## cor() 函数可以接收一个数据框，取出需要计算相关性的变量形成的数据集
```

```
s2nc_alltherm <- subset(s2nc, select = c("therm1", "therm2", "therm3", "therm4"))  
## (pairwise 的删除方式)  
cor(s2nc_alltherm, use = "pairwise.complete.obs")
```

```
##           therm1    therm2    therm3    therm4  
## therm1 1.0000000 0.9734449 0.9594085 0.9709017  
## therm2 0.9734449 1.0000000 0.9308287 0.9436621  
## therm3 0.9594085 0.9308287 1.0000000 0.9343249  
## therm4 0.9709017 0.9436621 0.9343249 1.0000000
```

1.3 找出异常值; 创建散点图, 可视化研究 2 控制组中第一波回答和后续回答之间的关系

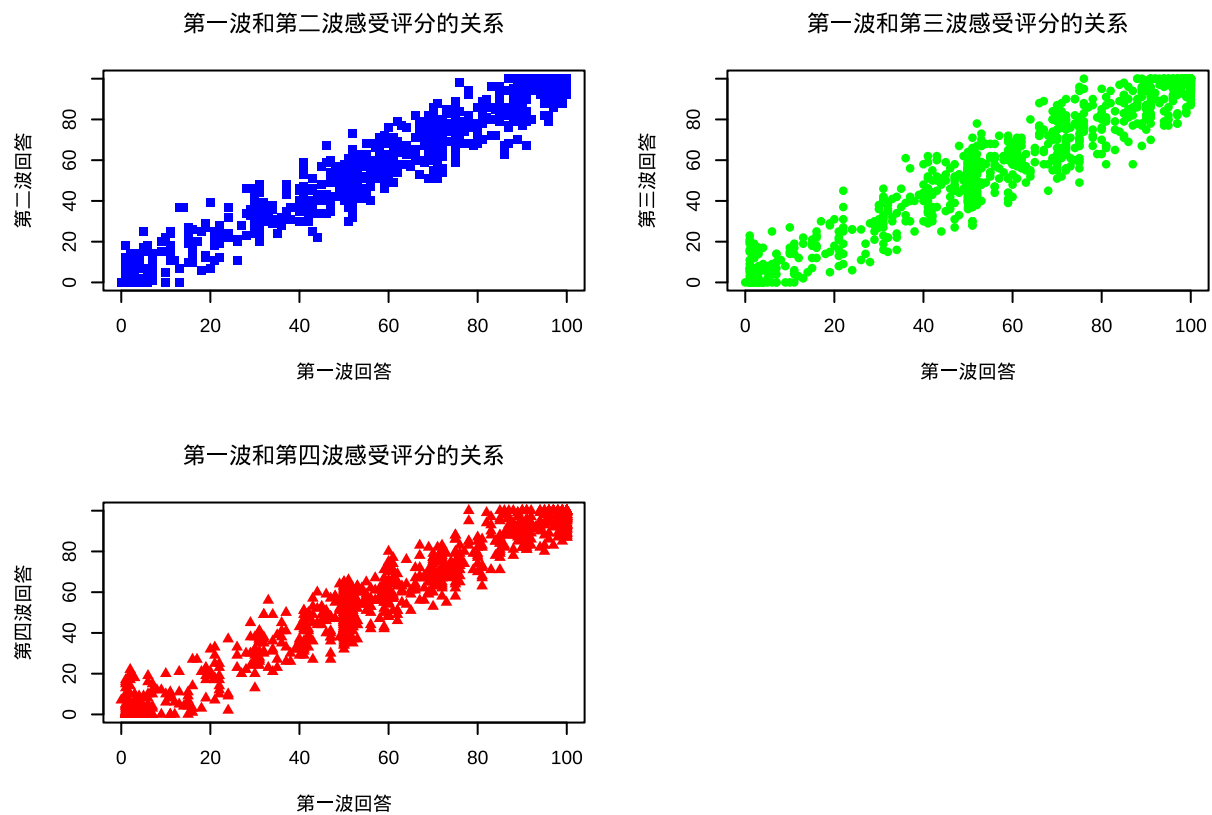
```
## 绘制一个箱型图以找出异常值  
boxplot(s2nc_alltherm, data = s2nc_alltherm,  
        names = c(" 第一轮", " 第二轮", " 第二轮", " 第三轮"),  
        xlab = " 轮次", ylab = " 得分",  
        ylim = c(0,100), main = " 各轮评分的箱形图")
```



```

par(mfrow = c(2, 2), cex = 0.6)
## 绘制第一波受访者和第二波的关系散点图
## 横坐标为第一波回答, 纵坐标为第二波的回答
plot(s2nc$therm1, s2nc$therm2, pch = 15, col = "blue",
     xlab = " 第一波回答", ylab = " 第二波回答",
     main = " 第一波和第二波感受评分的关系")
## 绘制第一波受访者和第三波的关系散点图
## 横坐标为第一波回答, 纵坐标为第三波的回答
plot(s2nc$therm1, s2nc$therm3, pch = 16, col = "green",
     xlab = " 第一波回答", ylab = " 第三波回答",
     main = " 第一波和第三波感受评分的关系")
## 绘制第一波受访者和第四波的关系散点图
## 横坐标为第一波回答, 纵坐标为第四波的回答
plot(s2nc$therm1, s2nc$therm4, pch = 17, col = "red",
     xlab = " 第一波回答", ylab = " 第四波回答",
     main = " 第一波和第四波感受评分的关系")

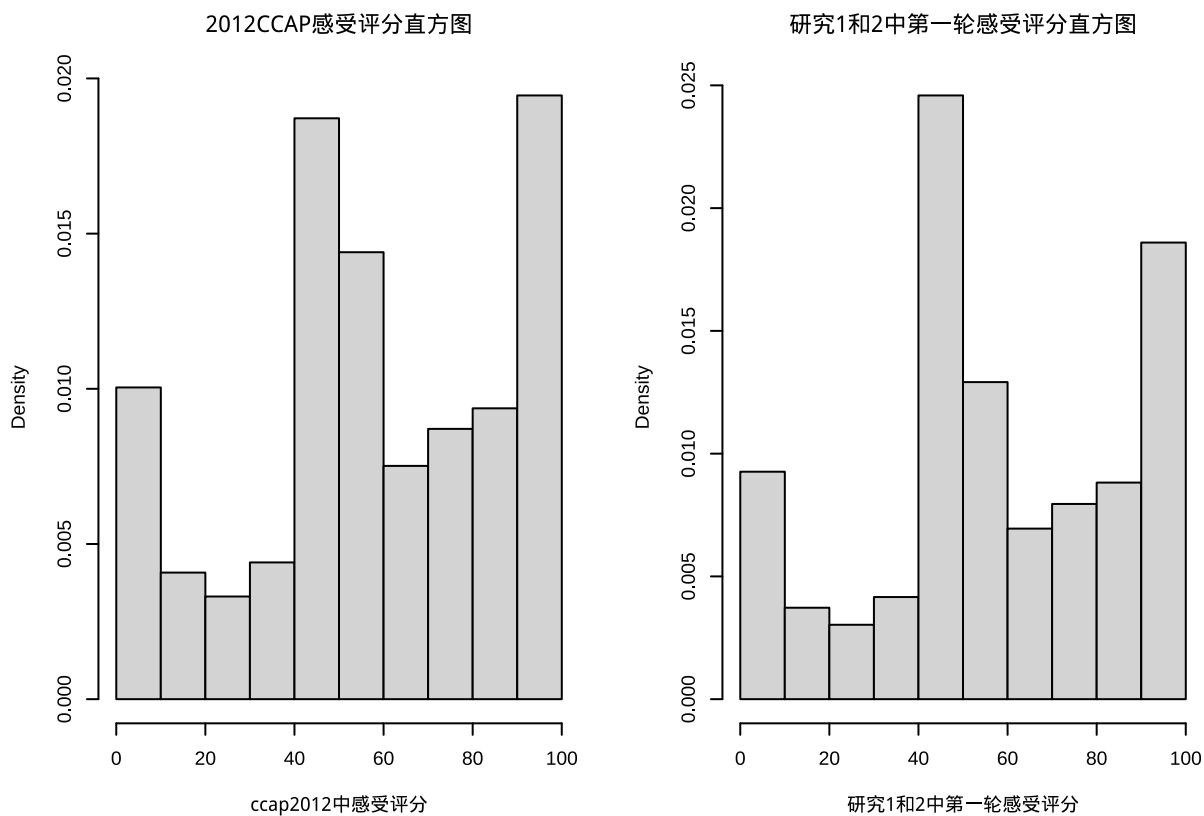
```



第一波回答和后续回答的感受评分之间成正相关关系

1.4 比较 ccap2012 和 gayreshaped 两个数据集

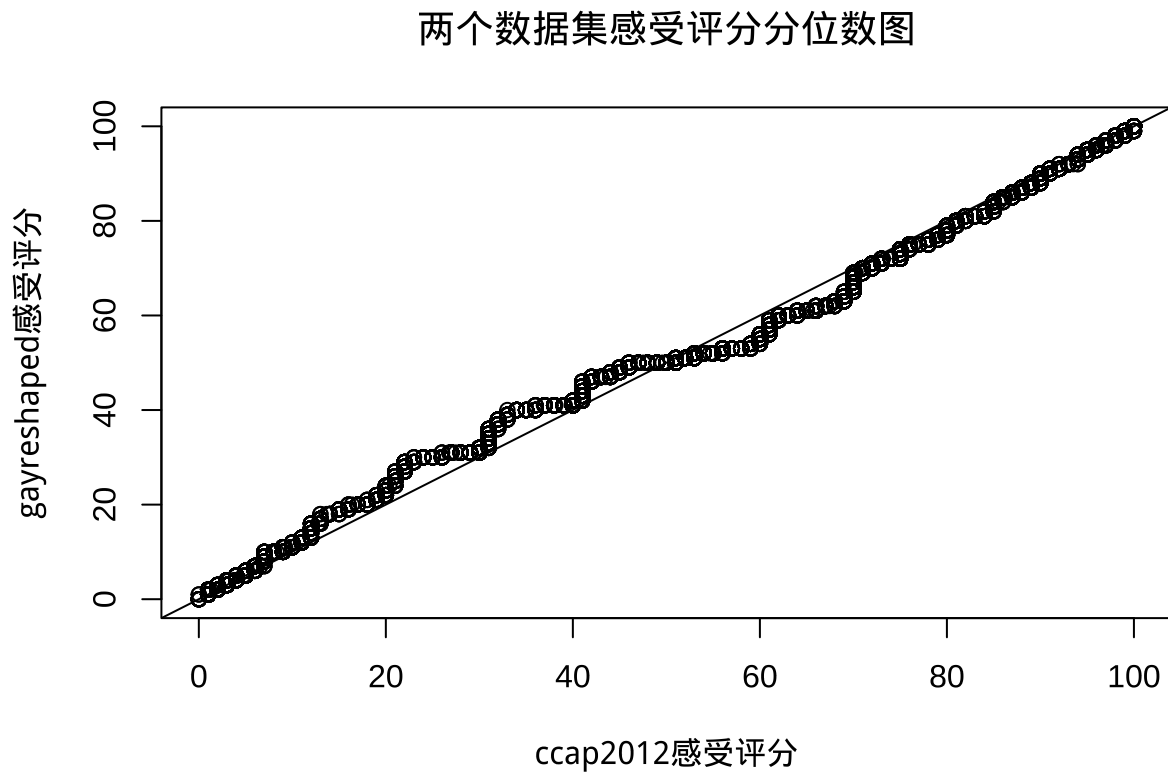
```
par(mfrow = c(1, 2), cex = 0.6)
# 创建 2012CCAP 感受评分的直方图
hist(ccap2012$gaytherm, freq = FALSE,
     breaks = seq(from = 0, to = 100, by = 10),
     xlab = "ccap2012 中感受评分",
     main = "2012CCAP 感受评分直方图")
hist(gayreshaped$therm1, freq = FALSE,
     breaks = seq(from = 0, to = 100, by = 10),
     xlab = "研究 1 和 2 中第一轮感受评分",
     main = "研究 1 和 2 中第一轮感受评分直方图")
```



1.5 通过分位数图比较两个样本分布

```
## 依旧以第一轮感受评分为例
qqplot(ccap2012$gaytherm, gayreshaped$therm1, xlab = "ccap2012 感受评分",
```

```
ylab = "gayreshaped 感受评分", xlim = c(0,100), ylim = c(0,100),
main = " 两个数据集感受评分分位数图")
abline(0,1)
```



点形成的线与 45 度线基本重合，说明两个样本分布几乎相同

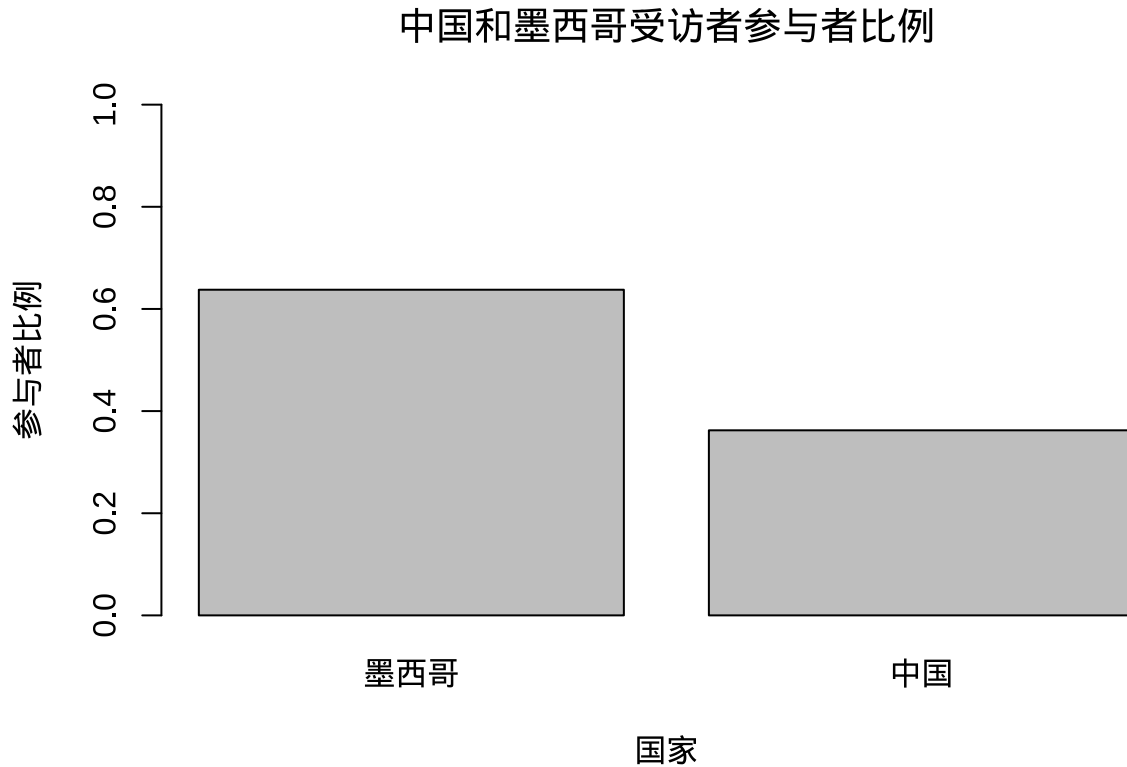
2 中国和墨西哥的政治效力

```
vignettes <- read.csv("vignettes.csv")
# View(vignettes)
```

2.1 用条形图表示中国和墨西哥的受访者回答分布，计算每个国家受访者回答平均分

```
# 计算每个国家受访者参与者比例
prop.respondents <- prop.table(table(vignettes$china))
prop.respondents
```

```
##  
##          0          1  
## 0.637644 0.362356  
  
# 条形图表示中国和墨西哥的受访者回答分布图  
barplot(prop.respondents,  
        names.arg = c(" 墨西哥", " 中国"), # 指定每个小节标签  
        main = " 中国和墨西哥受访者参与者比例",  
        xlab = " 国家",  
        ylab = " 参与者比例",  
        ylim = c(0, 1)  
        )
```

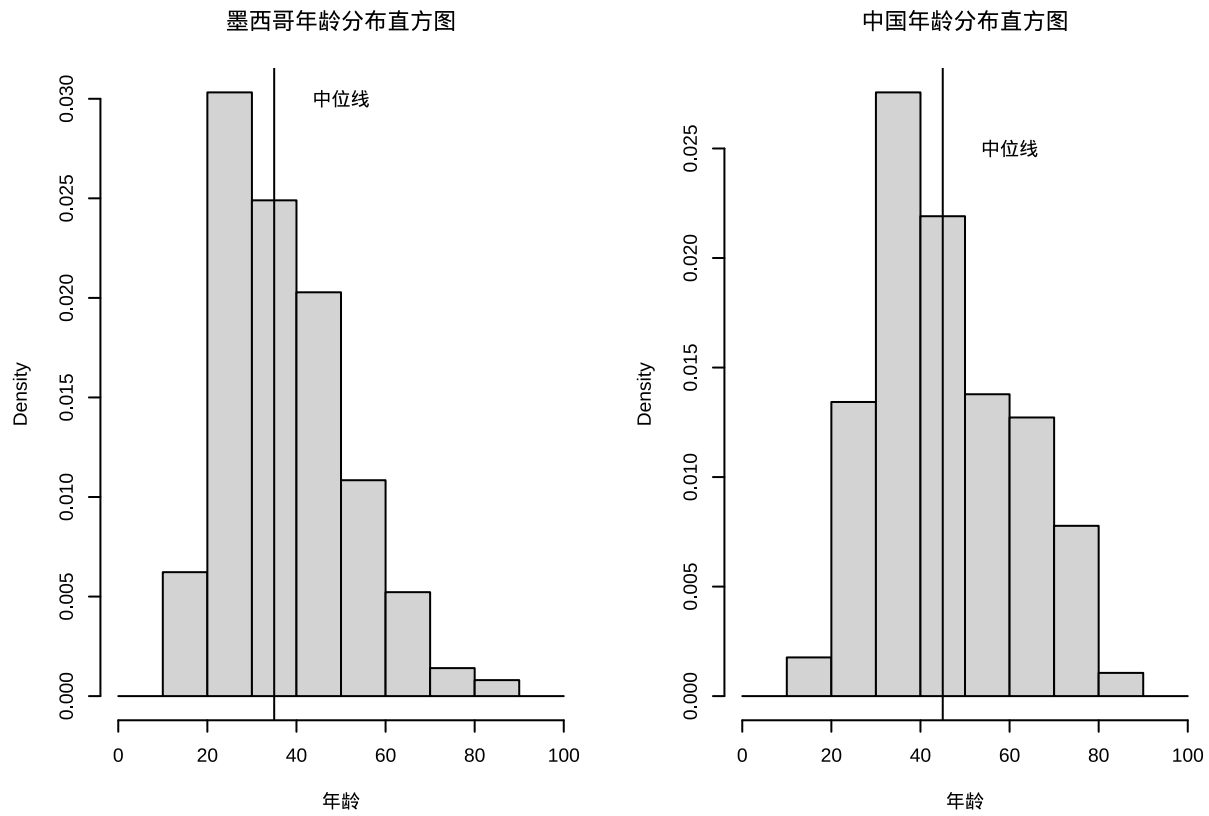


```
## 计算每个国家受访者回答平均分, china 中的 1 表示中国, 0 表示墨西哥  
tapply(vignettes$self, vignettes$china, mean)
```

```
##          0          1  
## 1.825301 2.621908
```

2.2 检查中国和墨西哥的分数差异是否由年龄分布差异引起

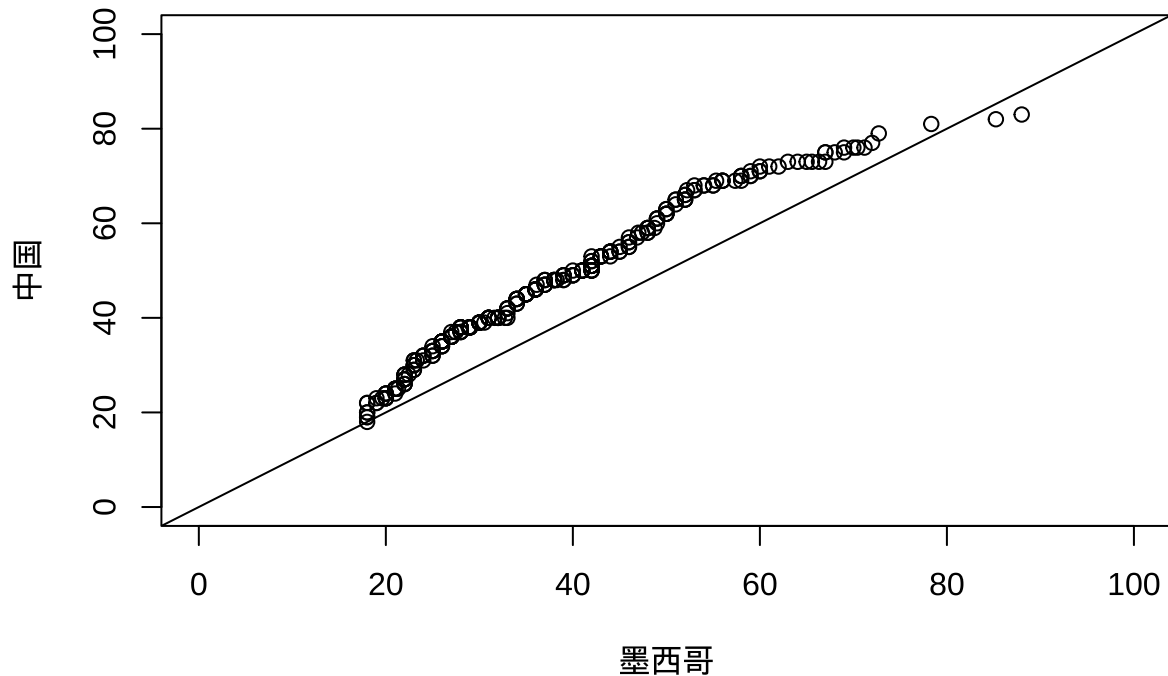
```
## 创建墨西哥的年龄变量的直方图，并添加中位年龄的垂直线
par(mfrow = c(1, 2), cex = 0.6)
chinaset <- subset(vignettes, subset = (vignettes$china == 1))
mexicoset <- subset(vignettes, subset = (vignettes$china == 0))
hist(mexicoset$age, freq = FALSE,
     breaks = seq(from = 0, to = 100, by = 10),
     xlab = " 年龄",
     main = " 墨西哥年龄分布直方图")
text(x = 50, y = 0.03, " 中位线") # 文本标签“中位线”出现在 (3, 0.5) 的位置
abline(v = median(mexicoset$age)) # 在中位数处绘制一条垂直线
## 创建中国的年龄变量的直方图，并添加中位年龄的垂直线
hist(chinaset$age, freq = FALSE,
     breaks = seq(from = 0, to = 100, by = 10),
     xlab = " 年龄",
     main = " 中国年龄分布直方图")
text(x = 60, y = 0.025, " 中位线") # 文本标签“中位线”出现在 (3, 0.5) 的位置
abline(v = median(chinaset$age)) # 在中位数处绘制一条垂直线
```

使用分位数图比较两个国家年龄分布情况

```
qqplot(mexicoset$age, chinaset$age, xlab = " 墨西哥",
       ylab = " 中国", xlim = c(0, 100), ylim = c(0, 100),
       main = " 墨西哥和中国年龄分布 Q-Q 图")
abline(0, 1)
```

墨西哥和中国年龄分布Q-Q图



大部分分位数点出现在 45 度线以上：说明中国答复者的年龄普遍比墨西哥答复者的年龄大 - 对于年轻人群，中国受访者的年龄分布更分散，对于年老人群，中国受访者年龄分布集中

2.3 对比自我评估的问题分数和情景问题的分数（计算两个国家自我评估比 Moses 分数低的比例）

```
# 中国自我评估比 Moses 分数低的比例
china <- subset(chinaset, subset = (self < moses))
prop1 <- nrow(china)/nrow(chinaset)
prop1
```

```
## [1] 0.5618375
```

```
# 墨西哥自我评估比 Moses 分数低的比例
mexico <- subset(mexicoset, subset = (self < moses))
prop2 <- nrow(mexico)/nrow(mexicoset)
prop2
```

```
## [1] 0.248996
```

- 中国政治效力低
- 但之前第一问的分析中，中国民众看起来有更高的政治效力

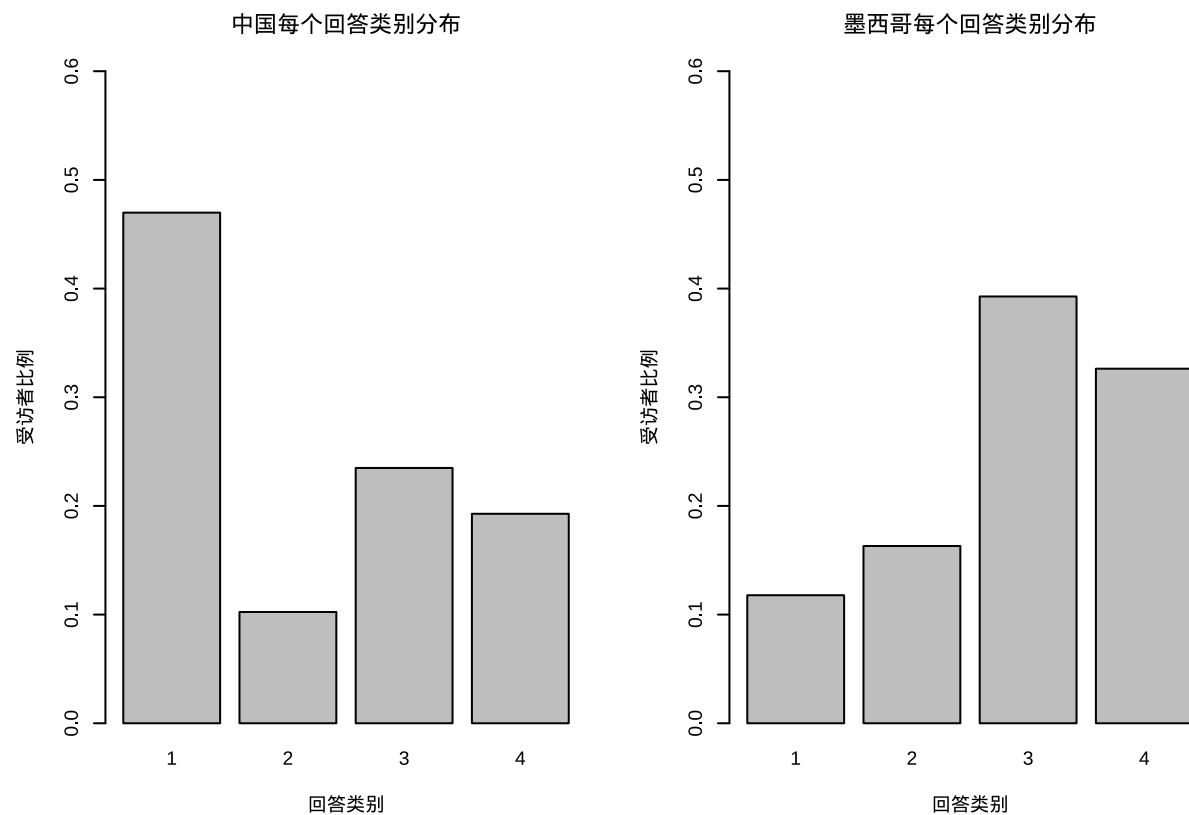
2.4 创建一个变量记录受访者自我评估相对于情景评估的排名

```
# 将符合预期排序的数据分出一个子集
expecteddata <- subset(vignettes, subset = ((alison >= jane) & (jane >= moses)))
# 新建一个变量 rank
expecteddata$rank <- NA
expecteddata$rank[expecteddata$self < expecteddata$moses] <- 1
expecteddata$rank[expecteddata$self >= expecteddata$moses
                  & expecteddata$self < expecteddata$jane] <- 2
expecteddata$rank[expecteddata$self >= expecteddata$jane
                  & expecteddata$self < expecteddata$alison] <- 3
expecteddata$rank[expecteddata$self >= expecteddata$alison] <- 4

par(mfrow = c(1, 2), cex = 0.6)

# 条形图表示中国每个回答类别受访者比例
china.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 1]))
barplot(china.ptable,
        names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
        main = " 中国每个回答类别分布",
        xlab = " 回答类别",
        ylab = " 受访者比例",
        ylim = c(0, 0.6)
        )

# 条形图表示墨西哥每个回答类别受访者比例
mexico.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 0]))
barplot(mexico.ptable,
        names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
        main = " 墨西哥每个回答类别分布",
        xlab = " 回答类别",
        ylab = " 受访者比例",
        ylim = c(0, 0.6)
        )
```



```
# 计算中国和墨西哥这个新变量的均值
```

```
tapply(expecteddata$rank, expecteddata$china, mean)
```

```
##          0          1
```

```
## 2.927492 2.150602
```

- 中国等级 1 的比例较高，而墨西哥等级 3 和 4 的比例较高
- 中国相对于情景而言的自我评估分数比墨西哥低
- 墨西哥的民众看似具有更高的政治效力

2.5 老年人是否比年轻人更严重？

```
par(mfrow = c(2, 2), cex = 0.6)
```

```
# 条形图表示中国 40 岁以下每个回答类别受访者比例
```

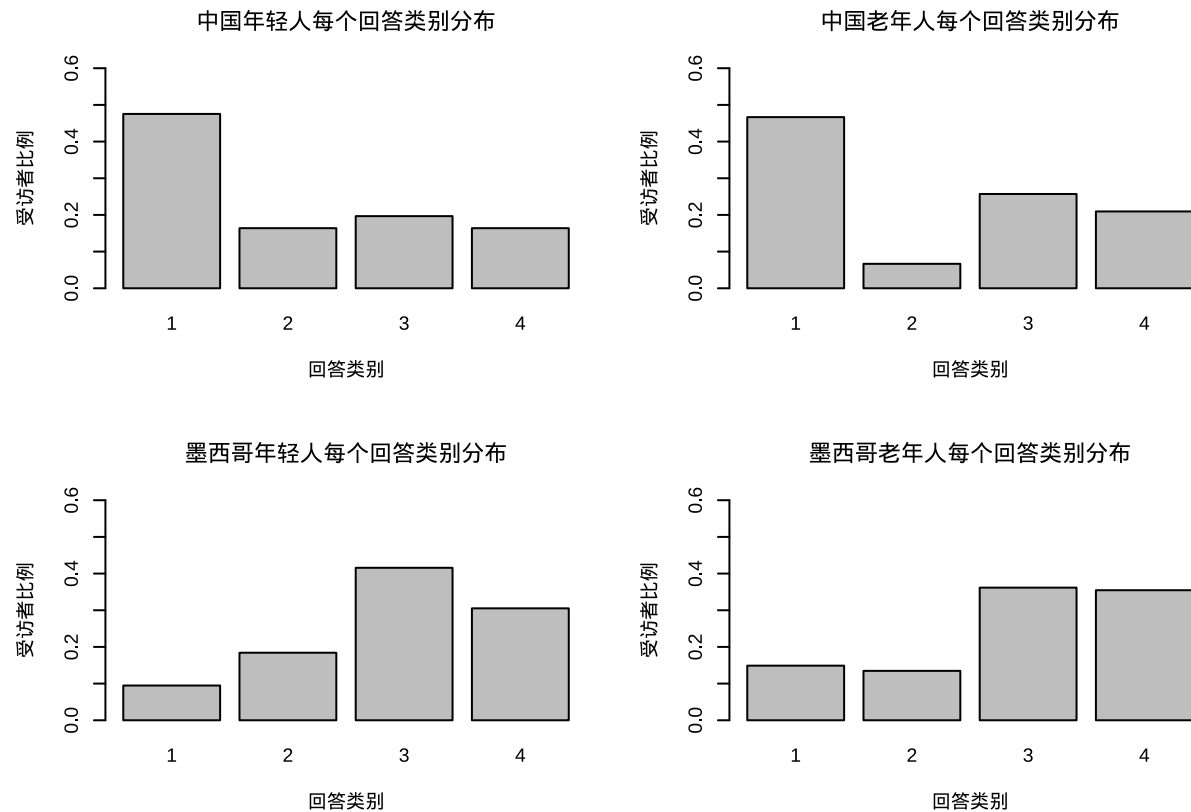
```
china.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 1  
                                     & expecteddata$age < 40]))
```

```
barplot(china.ptable,
```

```

names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
main = " 中国年轻人每个回答类别分布",
xlab = " 回答类别",
ylab = " 受访者比例",
ylim = c(0, 0.6)
)
# 条形图表示中国 40 岁以下每个回答类别受访者比例
china.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 1
                                & expecteddata$age >= 40]))
barplot(china.ptable,
        names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
        main = " 中国老年人每个回答类别分布",
        xlab = " 回答类别",
        ylab = " 受访者比例",
        ylim = c(0, 0.6)
)
# 条形图表示墨西哥 40 岁以下每个回答类别受访者比例
mexico.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 0
                                & expecteddata$age < 40]))
barplot(mexico.ptable,
        names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
        main = " 墨西哥年轻人每个回答类别分布",
        xlab = " 回答类别",
        ylab = " 受访者比例",
        ylim = c(0, 0.6)
)
# 条形图表示墨西哥 40 岁以上每个回答类别受访者比例
mexico.ptable <- prop.table(table(expecteddata$rank[expecteddata$china == 0
                                & expecteddata$age >= 40]))
barplot(mexico.ptable,
        names.arg = c("1", "2", "3", "4"), # 指定每个小节标签
        main = " 墨西哥老年人每个回答类别分布",
        xlab = " 回答类别",
        ylab = " 受访者比例",
        ylim = c(0, 0.6)
)

```



在中国，老年人在高自我评估等级的比例大于年轻人，老年人组政治效力更高 - 在墨西哥，老年人在低自我评估等级的比例大于年轻人，年轻人组政治效力更高

3 联合国大会投票表决

```
unvoting <- read.csv("unvoting.csv")
# View(unvoting)
```

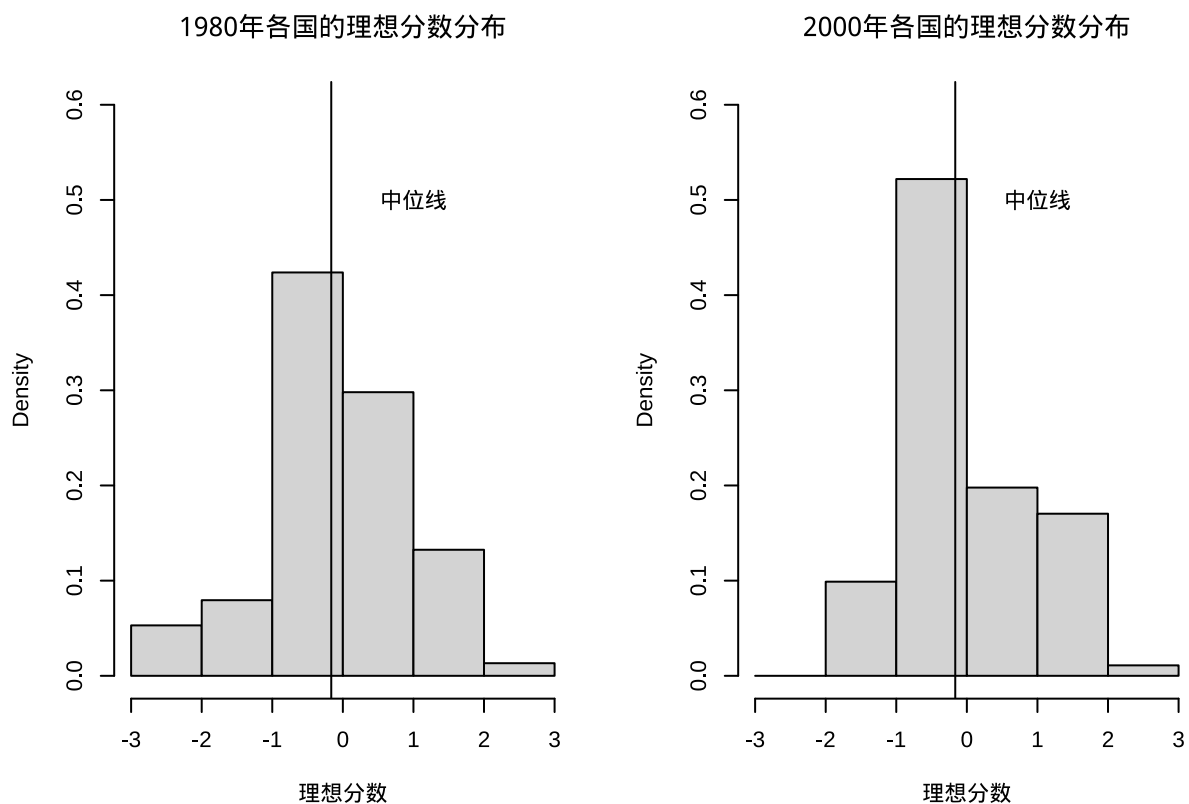
3.1 柏林墙倒塌的前后 10 年（即 1980 年和 2000 年）的各国理想分数分布

```
par(mfrow = c(1, 2), cex = 0.7)
# 1980 年各国的理想分数分布
hist(unvoting$idealscore[unvoting$Year == 1980], freq = FALSE,
     breaks = seq(from = -3, to = 3, by = 1),
     xlab = "理想分数", ylim = c(0, 0.6),
     main = "1980 年各国的理想分数分布")
```

```

text(x = 1, y = 0.5, " 中位线") # 文本标签“中位线”出现在 (1, 0.5) 的位置
abline(v = median(unvoting$idealpoint)) # 在中位数处绘制一条垂直线
# 2000 年各国的理想分数
hist(unvoting$idealpoint[unvoting$Year == 2000], freq = FALSE,
     breaks = seq(from = -3, to = 3, by = 1),
     xlab = " 理想分数", ylim = c(0, 0.6),
     main = "2000 年各国的理想分数分布")
text(x = 1, y = 0.5, " 中位线") # 文本标签“中位线”出现在 (1, 0.5) 的位置
abline(v = median(unvoting$idealpoint)) # 在中位数处绘制一条垂直线

```



```
quantile(unvoting$idealpoint[unvoting$Year == 1980])
```

```
##           0%          25%          50%          75%         100%
## -2.35749800 -0.34559800 -0.09465057  0.34365995  2.31760700
```

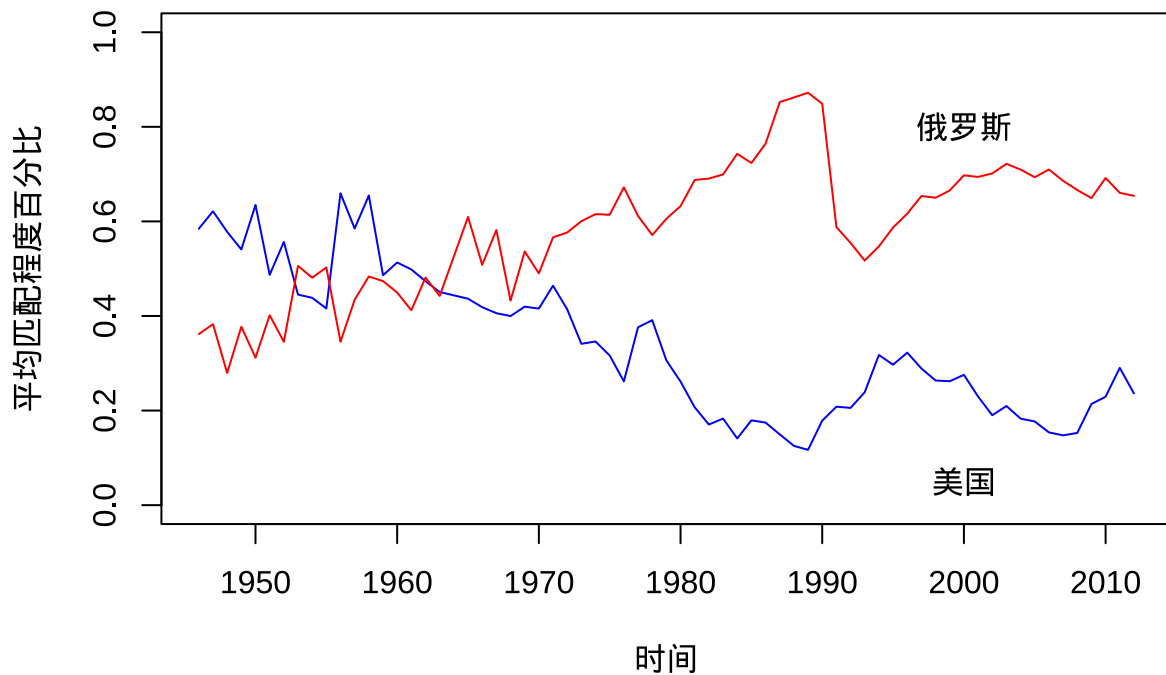
```
quantile(unvoting$idealpoint[unvoting$Year == 2000])
```

```
##           0%          25%          50%          75%         100%
## -1.7018510 -0.7021018 -0.3514302  0.6200154  2.6105410
```

- 1980 年比 2000 年拥有更多极端保守的国家
- 随着时间的变化, 越来越多国家意识形态偏向中立

3.2 随着时间的推移, 与美国投票一致的国家数量变化

```
## 计算每年各个国家与美国的匹配程度的平均百分比
PctAgreeUS.mean <- tapply(unvoting$PctAgreeUS, unvoting$Year, mean, na.rm = TRUE)
## 计算每年各个国家与俄罗斯的匹配程度的平均百分比
PctAgreeRUSSIA.mean <- tapply(unvoting$PctAgreeRUSSIA, unvoting$Year, mean, na.rm = TRUE)
## 创建一个折线图, 观察各个国家与美国和俄罗斯的平均匹配程度百分比如何随时间变化
plot(names(PctAgreeUS.mean), PctAgreeUS.mean, col = "blue",
     type = "l", xlim = c(1946, 2012), ylim = c(0, 1),
     xlab = " 时间", ylab = " 平均匹配程度百分比") # 与美国
lines(names(PctAgreeRUSSIA.mean), PctAgreeRUSSIA.mean, col = "red") # 与俄罗斯
text(2000, 0.8, " 俄罗斯")
text(2000, 0.05, " 美国")
```



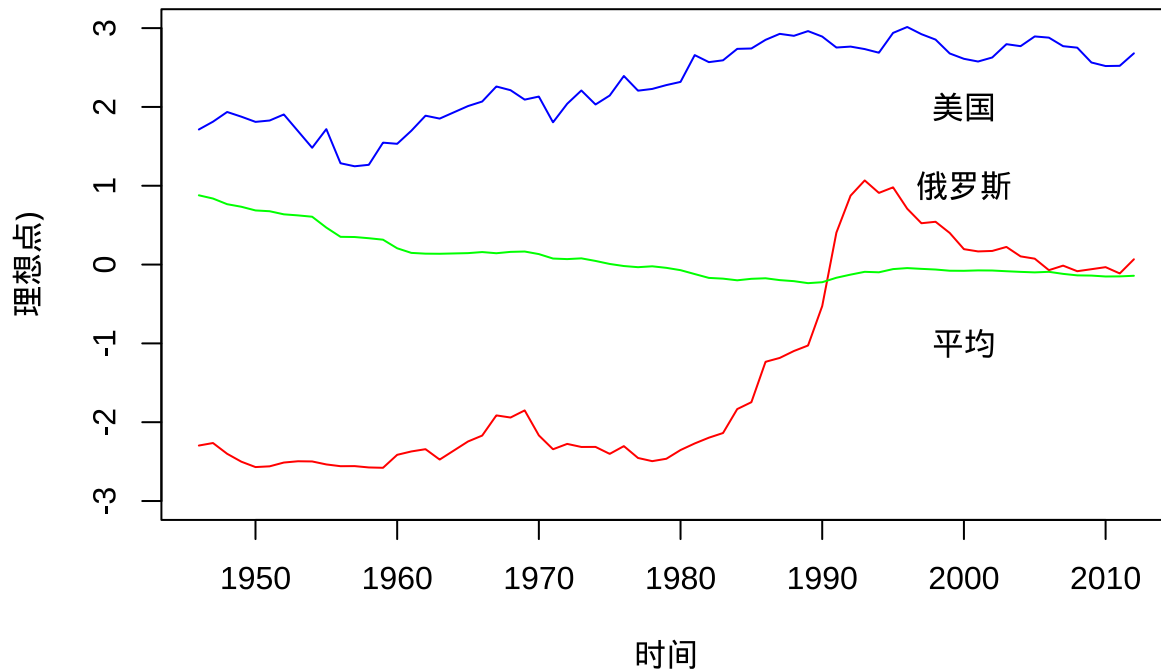

```
## 与俄罗斯相比，美国随着时间的增加变得越来越孤立
```

```
## 找出一些一贯支持美国的国家和一贯支持俄罗斯的国家
```

3.3 绘制美国和俄罗斯的理想点随时间变化的图象

```
## 创建一个折线图，观察美国、俄罗斯和其他国家平均的理想点随时间变化
```

```
plot(names(PctAgreeUS.mean), unvoting$idealpoint[unvoting$CountryAbb == "USA"],  
      col = "blue", type = "l", xlim = c(1946, 2012), ylim = c(-3, 3),  
      xlab = " 时间", ylab = " 理想点") # 美国  
lines(names(PctAgreeUS.mean), unvoting$idealpoint[unvoting$CountryAbb == "RUS"],  
       col = "red") # 俄罗斯  
# 计算每年所有国家平均的理想点数值  
idealpoint.mean <- tapply(unvoting$idealpoint, unvoting$Year, mean, na.rm = TRUE)  
lines(names(PctAgreeUS.mean), idealpoint.mean, col = "green") # 所有国家平均  
text(2000, 1, " 俄罗斯")  
text(2000, 2, " 美国")  
text(2000, -1, " 平均")
```



美国的意识形态立场有微小变化，变得越来越自由主义 - 俄罗斯的意识形态立场有很大变化，先是大幅度变得更自由主义，之后又逐渐保守 - 所有国家的平均意识形态立场整体上变得越来越保守

3.4 曾属于苏联的国家和不属于苏联的国家有什么不同

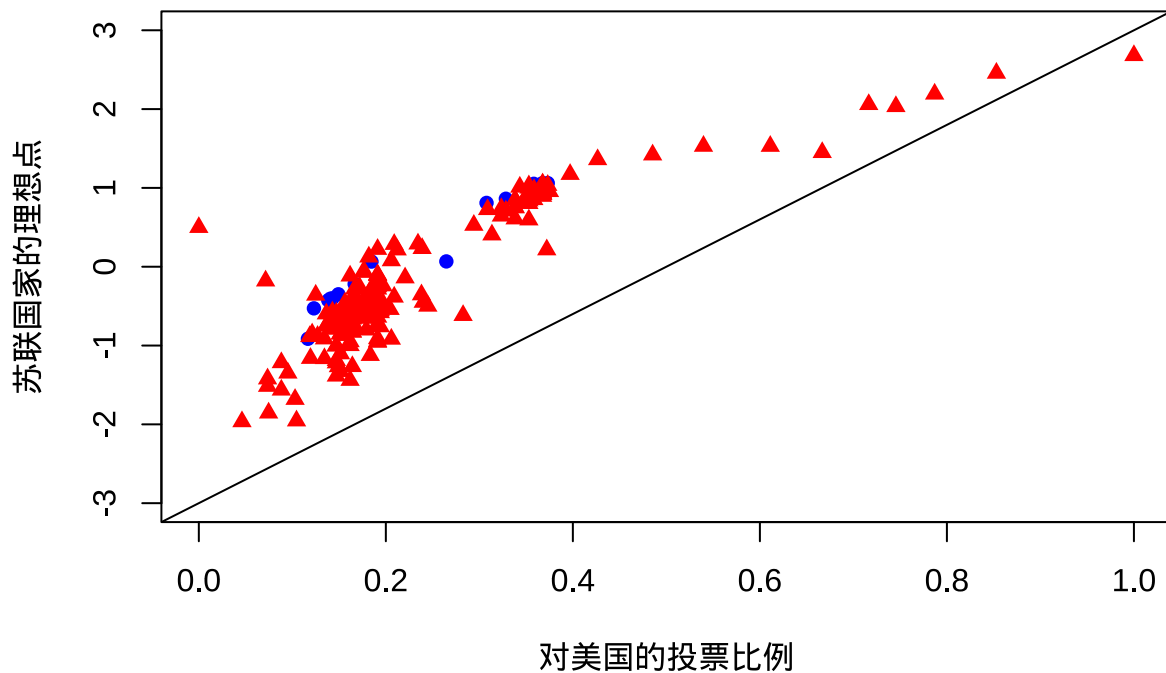
```
## 记录一个包含苏联国家的向量
RUSSIA <- c("Estonia", "Latvia", "Lithuania", "Belarus", "Moldova",
            "Ukraine", "Armenia", "Azerbaijan", "Georgia", "Kazakhstan",
            "Kyrgyzstan", "Tajikistan", "Turkmenistan", "Uzbekistan", "Russia")
## 新建一个判断所有国家哪些曾属于苏联的逻辑向量
unvoting$is.RUSSIA <- unvoting$CountryName %in% RUSSIA
## 将曾属于苏联国家的理想点与对美国的投票比例进行对比 (2012 年)
## 首先将 2012 年曾属于苏联国家的数据分出来
unvotingRUSSIA2012 <- subset(unvoting, subset = (Year == 2012 & is.RUSSIA == TRUE))
unvotingOTHERS2012 <- subset(unvoting, subset = (Year == 2012 & is.RUSSIA == FALSE))
## 使用 Q-Q 图比较理想点和投票比例分布
qqplot(unvotingRUSSIA2012$PctAgreeUS, unvotingRUSSIA2012$idealpoint,
       xlab = " 对美国的投票比例", ylab = " 苏联国家的理想点", xlim = c(0, 1),
```

```

ylim = c(-3, 3), main = " 曾属于苏联国家的理想点与对美国投票比例 Q-Q 图",
pch = 16, col = "blue")
abline(-3, 6)
points(unvotingOTHERS2012$PctAgreeUS, unvotingOTHERS2012$idealpoint,
       pch = 17, col = "red")

```

曾属于苏联国家的理想点与对美国投票比例Q-Q图



对于曾属于苏联的国家而言，对美国支持率较低的国家的理想点分布比较分散 - 对于不属于苏联的国家而言，对美国支持率较低的国家的理想点分布比较分散，对美国支持率较高的国家的理想点分布比较集中 - 说明支持美国的国家一般意识形态比较类似

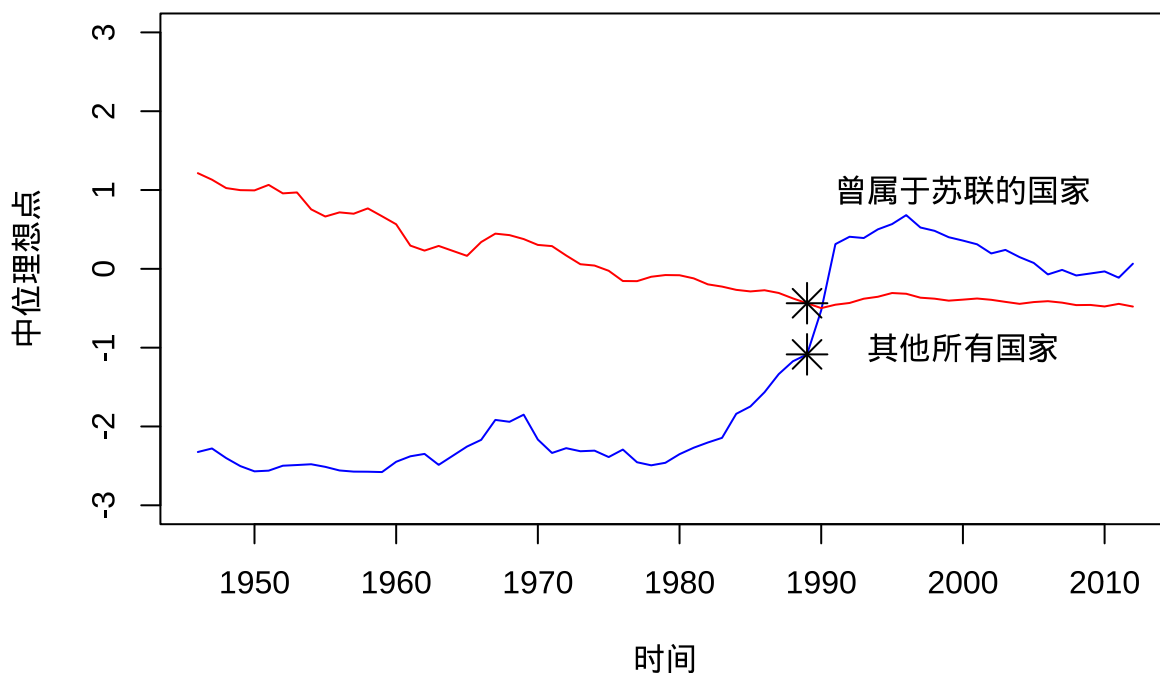
3.5 观察曾属于苏联的国家和其他所有国家的理想点在过去几年中的变化

```

## 首先将曾属于苏联国家的数据分出来
unvotingRUSSIA <- subset(unvoting, subset = (is.RUSSIA == TRUE))
unvotingOTHERS <- subset(unvoting, subset = (is.RUSSIA == FALSE))
## 计算曾属于苏联的国家的每年的中位理想点
RUSSIA.median <- tapply(unvotingRUSSIA$idealpoint, unvotingRUSSIA$Year, median)
OTHERS.median <- tapply(unvotingOTHERS$idealpoint, unvotingOTHERS$Year, median)

```

```
## 创建一个折线图，观察各个国家与美国和俄罗斯的平均匹配程度百分比如何随时间变化
plot(names(RUSSIA.median), RUSSIA.median, col = "blue",
     type = "l", xlim = c(1946, 2012), ylim = c(-3, 3),
     xlab = " 时间", ylab = " 中位理想点") # 曾属于苏联国家
lines(names(OTHERS.median), OTHERS.median, col = "red") # 其他所有国家
## 添加 1989 年的两个点
points(1989, -1.08524800, pch = 8, cex = 2)
points(1989, -0.43633980, pch = 8, cex = 2)
text(2000, 1, " 曾属于苏联的国家")
text(2000, -1, " 其他所有国家")
```



曾属于苏联的国家的理想点中位数这些年来逐渐上升，变得越来越自由，直到 1995 年左右开始下降，又开始变得保守 - 其他所有国家的理想点中位数这些年来逐渐下降，变得越来越保守

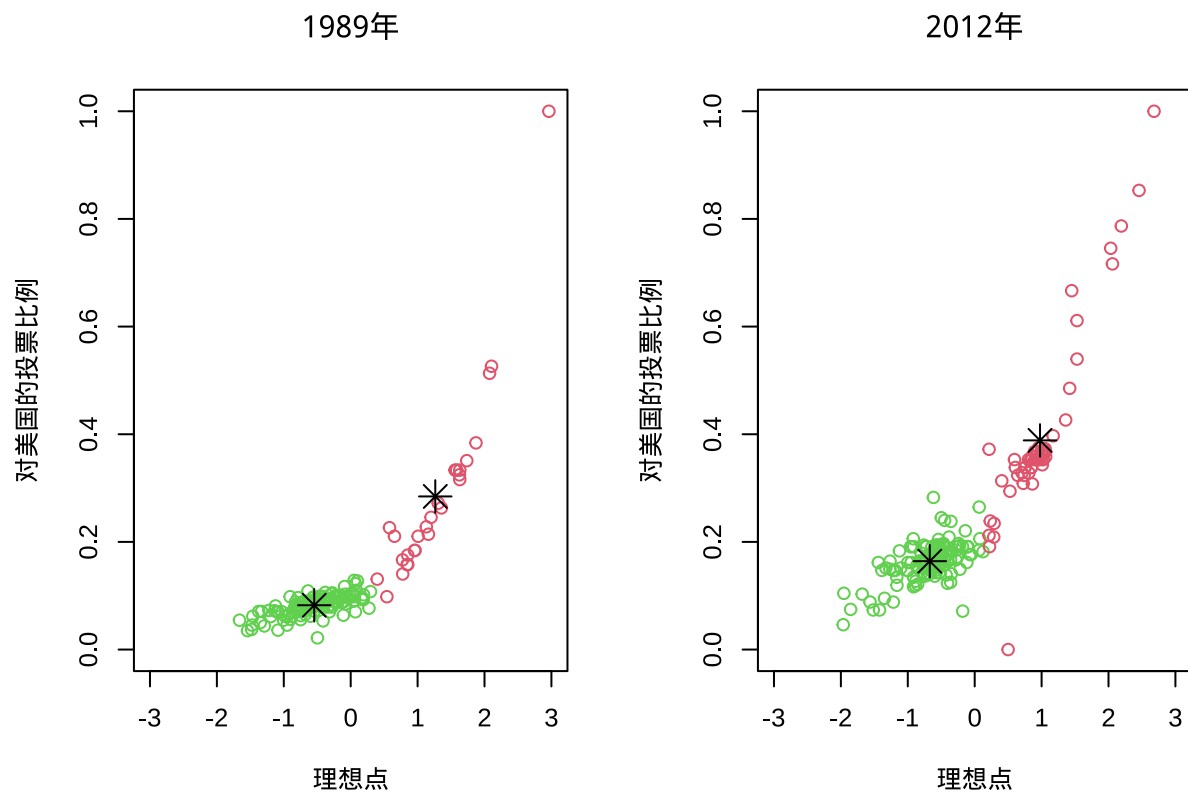
3.6 国家是否仍然分为两个意识形态派别？

```
## 将 idealpoint 和 PctAgreeUS 变量组合
k1989 <- cbind(unvoting$idealpoint[unvoting$Year == 1989],
```

```
        unvoting$PctAgreeUS[unvoting$Year == 1989])
k2012 <- cbind(unvoting$Idealpoint[unvoting$Year == 2012],
              unvoting$PctAgreeUS[unvoting$Year == 2012])
## 实行 kmeans 算法进行聚类
k1989.out <- kmeans(k1989, centers = 2, nstart = 2)
k2012.out <- kmeans(k2012, centers = 2, nstart = 2)

## 可视化
par(mfrow = c(1, 2), cex = 0.8)
## 绘制 1989 年两个聚类的散点图
plot(k1989, col = k1989.out$cluster + 1, xlab = "理想点", ylab = "对美国的投票比例",
      xlim = c(-3, 3), ylim = c(0, 1), main = "1989 年")
## 绘制质心
points(k1989.out$centers, pch = 8, cex = 2)

## 绘制 2012 年两个聚类的散点图
plot(k2012, col = k2012.out$cluster + 1, xlab = "理想点", ylab = "对美国的投票比例",
      xlim = c(-3, 3), ylim = c(0, 1), main = "2012 年")
## 绘制质心
points(k2012.out$centers, pch = 8, cex = 2)
```



从所有国际的情况上看，大致还是分成两个意识形态，但是随着时间的推移，确实变得更多元化