

# Gini Coefficients of OECD Countries

20173253 Minyoung-Choi

## Data 작업

- OECD 국가들의 Gini계수 읽어들이기.

```
library(knitr)
library(magrittr)
library(printr)
```

```
## Registered S3 method overwritten by 'printr':
##   method          from
##   knit_print.data.frame rmarkdown
```

```
Gini_b_tax <- read.table(file = "../data/Gini_before_tax.txt",
                        header = FALSE,
                        sep = "Wt")
Gini_a_tax <- read.table(file = "../data/Gini_after_tax.txt",
                        header = FALSE,
                        sep = "Wt")

str(Gini_b_tax)
```

```
## 'data.frame':   34 obs. of  8 variables:
## $ V1: chr  "Australia" "Austria" "Belgium" "Canada" ...
## $ V2: num  NA NA NA 0.385 NA NA NA NA 0.343 NA ...
## $ V3: num  NA NA 0.449 0.395 NA NA 0.373 NA 0.387 0.38 ...
## $ V4: num  NA NA NA 0.403 NA NA 0.396 NA NA 0.37 ...
## $ V5: num  0.467 NA 0.472 0.43 0.441 0.442 0.417 NA 0.479 0.473 ...
## $ V6: num  0.476 NA 0.464 0.44 NA 0.472 0.415 NA 0.478 0.49 ...
## $ V7: num  0.465 0.433 0.494 0.436 0.414 0.474 0.417 0.504 0.483 0.485 ...
## $ V8: num  0.468 0.472 0.469 0.441 0.426 0.444 0.416 0.458 0.465 0.483 ...
```

```
str(Gini_a_tax)
```

```
## 'data.frame':   34 obs. of  8 variables:
## $ V1: chr  "Australia" "Austria" "Belgium" "Canada" ...
## $ V2: num  NA NA NA 0.304 NA NA NA NA 0.235 NA ...
## $ V3: num  NA 0.236 0.274 0.293 NA NA 0.221 NA 0.209 0.3 ...
## $ V4: num  NA NA NA 0.287 NA 0.232 0.226 NA NA 0.29 ...
## $ V5: num  0.309 0.238 0.287 0.289 0.427 0.257 0.215 NA 0.218 0.277 ...
## $ V6: num  0.317 0.252 0.289 0.318 NA 0.26 0.226 NA 0.247 0.287 ...
## $ V7: num  0.315 0.265 0.271 0.317 0.403 0.268 0.232 0.349 0.254 0.288 ...
## $ V8: num  0.336 0.261 0.259 0.324 0.394 0.256 0.248 0.315 0.259 0.293 ...
```

```
lapply(list(Gini_b_tax, Gini_a_tax), str)
```

```
## 'data.frame':  34 obs. of  8 variables:
## $ V1: chr  "Australia" "Austria" "Belgium" "Canada" ...
## $ V2: num  NA NA NA 0.385 NA NA NA NA 0.343 NA ...
## $ V3: num  NA NA 0.449 0.395 NA NA 0.373 NA 0.387 0.38 ...
## $ V4: num  NA NA NA 0.403 NA NA 0.396 NA NA 0.37 ...
## $ V5: num  0.467 NA 0.472 0.43 0.441 0.442 0.417 NA 0.479 0.473 ...
## $ V6: num  0.476 NA 0.464 0.44 NA 0.472 0.415 NA 0.478 0.49 ...
## $ V7: num  0.465 0.433 0.494 0.436 0.414 0.474 0.417 0.504 0.483 0.485 ...
## $ V8: num  0.468 0.472 0.469 0.441 0.426 0.444 0.416 0.458 0.465 0.483 ...
## 'data.frame':  34 obs. of  8 variables:
## $ V1: chr  "Australia" "Austria" "Belgium" "Canada" ...
## $ V2: num  NA NA NA 0.304 NA NA NA NA 0.235 NA ...
## $ V3: num  NA 0.236 0.274 0.293 NA NA 0.221 NA 0.209 0.3 ...
## $ V4: num  NA NA NA 0.287 NA 0.232 0.226 NA NA 0.29 ...
## $ V5: num  0.309 0.238 0.287 0.289 0.427 0.257 0.215 NA 0.218 0.277 ...
## $ V6: num  0.317 0.252 0.289 0.318 NA 0.26 0.226 NA 0.247 0.287 ...
## $ V7: num  0.315 0.265 0.271 0.317 0.403 0.268 0.232 0.349 0.254 0.288 ...
## $ V8: num  0.336 0.261 0.259 0.324 0.394 0.256 0.248 0.315 0.259 0.293 ...
```

```
## [[1]]
## NULL
##
## [[2]]
## NULL
```

- 2000년 후반 자료만 모아서 새로운 data frame 구성

```
(Gini_b_a<-data.frame(Country = Gini_b_tax$V1,
                      Before = Gini_b_tax$V8,
                      After = Gini_a_tax$V8))
```

Country	Before	After
Australia	0.468	0.336
Austria	0.472	0.261
Belgium	0.469	0.259
Canada	0.441	0.324
Chile	0.426	0.394
Czech_Republic	0.444	0.256
Denmark	0.416	0.248
Estonia	0.458	0.315
Finland	0.465	0.259
France	0.483	0.293
Germany	0.504	0.295
Greece	0.436	0.307
Hungary	0.466	0.272

Country	Before	After
Iceland	0.382	0.301
Ireland	NA	0.293
Israel	0.498	0.371
Italy	0.534	0.337
Japan	0.462	0.329
Luxembourg	0.482	0.288
Mexico	0.494	0.476
Netherlands	0.426	0.294
New_Zealand	0.455	0.330
Norway	0.410	0.250
Poland	0.470	0.305
Portugal	0.521	0.353
Slovak_Republic	0.416	0.257
Slovenia	0.423	0.236
South_Korea	0.344	0.315
Spain	0.461	0.317
Sweden	0.426	0.259
Switzerland	0.409	0.303
Turkey	0.470	0.409
United_Kingdom	0.456	0.345
United_States	0.486	0.378

- 세전과 세후의 Gini 계수 차이를 개선도(Improvement)라고 명명.

```
Gini_b_a$Improvement <- Gini_b_a[, 2] - Gini_b_a[, 3]
Gini_b_a
```

Country	Before	After	Improvement
Australia	0.468	0.336	0.132
Austria	0.472	0.261	0.211
Belgium	0.469	0.259	0.210
Canada	0.441	0.324	0.117
Chile	0.426	0.394	0.032
Czech_Republic	0.444	0.256	0.188

Country	Before	After	Improvement
Denmark	0.416	0.248	0.168
Estonia	0.458	0.315	0.143
Finland	0.465	0.259	0.206
France	0.483	0.293	0.190
Germany	0.504	0.295	0.209
Greece	0.436	0.307	0.129
Hungary	0.466	0.272	0.194
Iceland	0.382	0.301	0.081
Ireland	NA	0.293	NA
Israel	0.498	0.371	0.127
Italy	0.534	0.337	0.197
Japan	0.462	0.329	0.133
Luxembourg	0.482	0.288	0.194
Mexico	0.494	0.476	0.018
Netherlands	0.426	0.294	0.132
New_Zeland	0.455	0.330	0.125
Norway	0.410	0.250	0.160
Poland	0.470	0.305	0.165
Portugal	0.521	0.353	0.168
Slovak_Republic	0.416	0.257	0.159
Slovenia	0.423	0.236	0.187
South_Korea	0.344	0.315	0.029
Spain	0.461	0.317	0.144
Sweden	0.426	0.259	0.167
Switzerland	0.409	0.303	0.106
Turkey	0.470	0.409	0.061
United_Kingdom	0.456	0.345	0.111
United_States	0.486	0.378	0.108

- 개선도가 낮은 순서로 나열. 아일랜드는 세전 자료가 없기 때문에 맨 뒤로 위치.

```
Gini_b_a[order(Gini_b_a$Improvement), ]
```

Country	Before	After	Improvement
---------	--------	-------	-------------

	Country	Before	After	Improvement
20	Mexico	0.494	0.476	0.018
28	South_Korea	0.344	0.315	0.029
5	Chile	0.426	0.394	0.032
32	Turkey	0.470	0.409	0.061
14	Iceland	0.382	0.301	0.081
31	Switzerland	0.409	0.303	0.106
34	United_States	0.486	0.378	0.108
33	United_Kingdom	0.456	0.345	0.111
4	Canada	0.441	0.324	0.117
22	New_Zealand	0.455	0.330	0.125
16	Israel	0.498	0.371	0.127
12	Greece	0.436	0.307	0.129
1	Australia	0.468	0.336	0.132
21	Netherlands	0.426	0.294	0.132
18	Japan	0.462	0.329	0.133
8	Estonia	0.458	0.315	0.143
29	Spain	0.461	0.317	0.144
26	Slovak_Republic	0.416	0.257	0.159
23	Norway	0.410	0.250	0.160
24	Poland	0.470	0.305	0.165
30	Sweden	0.426	0.259	0.167
7	Denmark	0.416	0.248	0.168
25	Portugal	0.521	0.353	0.168
27	Slovenia	0.423	0.236	0.187
6	Czech_Republic	0.444	0.256	0.188
10	France	0.483	0.293	0.190
13	Hungary	0.466	0.272	0.194
19	Luxembourg	0.482	0.288	0.194
17	Italy	0.534	0.337	0.197
9	Finland	0.465	0.259	0.206
11	Germany	0.504	0.295	0.209
3	Belgium	0.469	0.259	0.210

	Country	Before	After	Improvement
2	Austria	0.472	0.261	0.211
15	Ireland	NA	0.293	NA

```
Gini_b_a[order(Gini_b_a$Improvement, decreasing = TRUE), ]
```

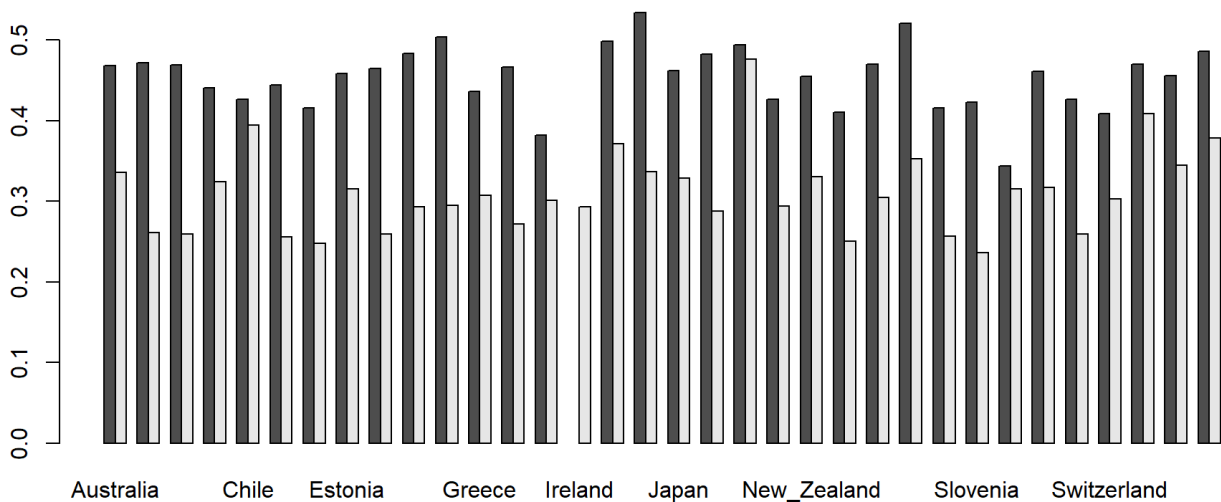
	Country	Before	After	Improvement
2	Austria	0.472	0.261	0.211
3	Belgium	0.469	0.259	0.210
11	Germany	0.504	0.295	0.209
9	Finland	0.465	0.259	0.206
17	Italy	0.534	0.337	0.197
13	Hungary	0.466	0.272	0.194
19	Luxembourg	0.482	0.288	0.194
10	France	0.483	0.293	0.190
6	Czech_Republic	0.444	0.256	0.188
27	Slovenia	0.423	0.236	0.187
25	Portugal	0.521	0.353	0.168
7	Denmark	0.416	0.248	0.168
30	Sweden	0.426	0.259	0.167
24	Poland	0.470	0.305	0.165
23	Norway	0.410	0.250	0.160
26	Slovak_Republic	0.416	0.257	0.159
29	Spain	0.461	0.317	0.144
8	Estonia	0.458	0.315	0.143
18	Japan	0.462	0.329	0.133
1	Australia	0.468	0.336	0.132
21	Netherlands	0.426	0.294	0.132
12	Greece	0.436	0.307	0.129
16	Israel	0.498	0.371	0.127
22	New_Zealand	0.455	0.330	0.125
4	Canada	0.441	0.324	0.117
33	United_Kingdom	0.456	0.345	0.111
34	United_States	0.486	0.378	0.108

	Country	Before	After	Improvement
31	Switzerland	0.409	0.303	0.106
14	Iceland	0.382	0.301	0.081
32	Turkey	0.470	0.409	0.061
5	Chile	0.426	0.394	0.032
28	South_Korea	0.344	0.315	0.029
20	Mexico	0.494	0.476	0.018
15	Ireland	NA	0.293	NA

# Graphic representation

- 세전 세후 Gini 계수를 시각적으로 비교하려면 `barplot()` 이 적합함. `barplot(height, ...)` 에서 `height` 가 매트릭스일 때는 막대는 열의 각 요소를 크기대로 쌓아놓은 형태가 되므로, `t()` 를 이용하여 `transpose`시킨 후 `barplot()` 을 적용. 또한 `transpose`를 시켜도 여전히 `data frame` 이기 때문에 매트릭스로 강제 변환함. 세전, 세후 비교를 위해 쌓아 놓기 보다는 옆에 늘어세우는 게 나으므로 `beside = TRUE` 를 적용하고 각 막대의 이름으로 나라이름을 사용.

```
barplot(as.matrix(t(Gini_b_a[, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country)
```



- 개선도 순서(내림차순)를 `o_improvement` 로 저장하여 지속적으로 활용.

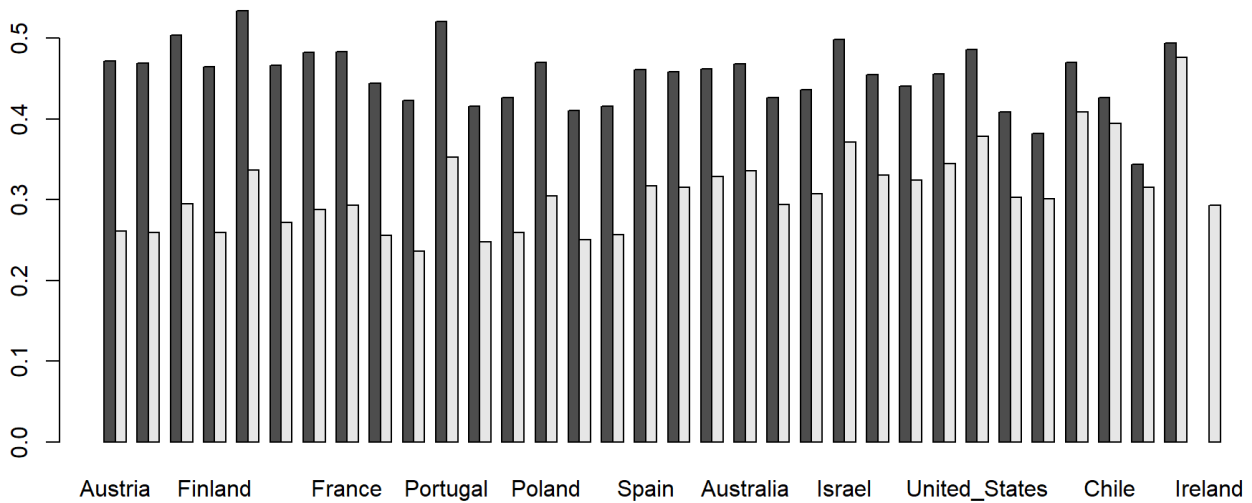
```
o_improvement <- order(Gini_b_a$Improvement, decreasing = TRUE)
Gini_b_a$Country[o_improvement]
```

```
## [1] "Austria"      "Belgium"      "Germany"      "Finland"
## [5] "Italy"        "Hungary"      "Luxembourg"   "France"
## [9] "Czech_Republic" "Slovenia"    "Portugal"    "Denmark"
## [13] "Sweden"      "Poland"      "Norway"      "Slovak_Republic"
## [17] "Spain"       "Estonia"     "Japan"       "Australia"
## [21] "Netherlands" "Greece"      "Israel"      "New_Zealand"
## [25] "Canada"      "United_Kingdom" "United_States" "Switzerland"
## [29] "Iceland"     "Turkey"      "Chile"       "South_Korea"
## [33] "Mexico"      "Ireland"
```

- 개선도 순서대로 막대를 늘어세우면,

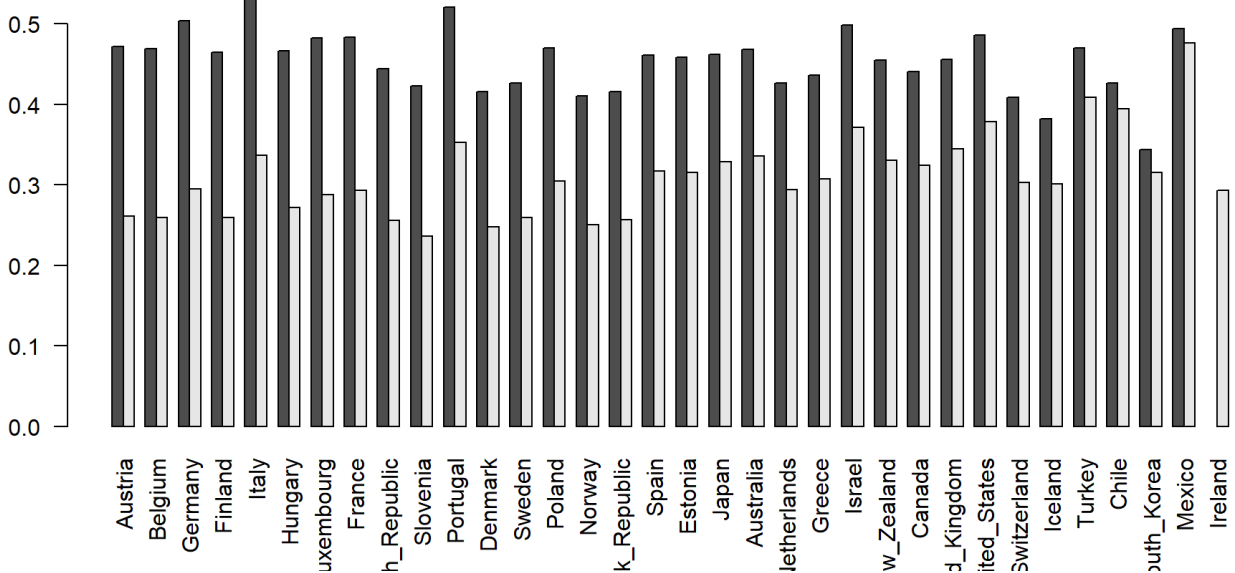
```
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country[o_improvement])
```





- las = 2 를 이용하여 막대 이름을 눕힘.

```
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        las = 2)
```

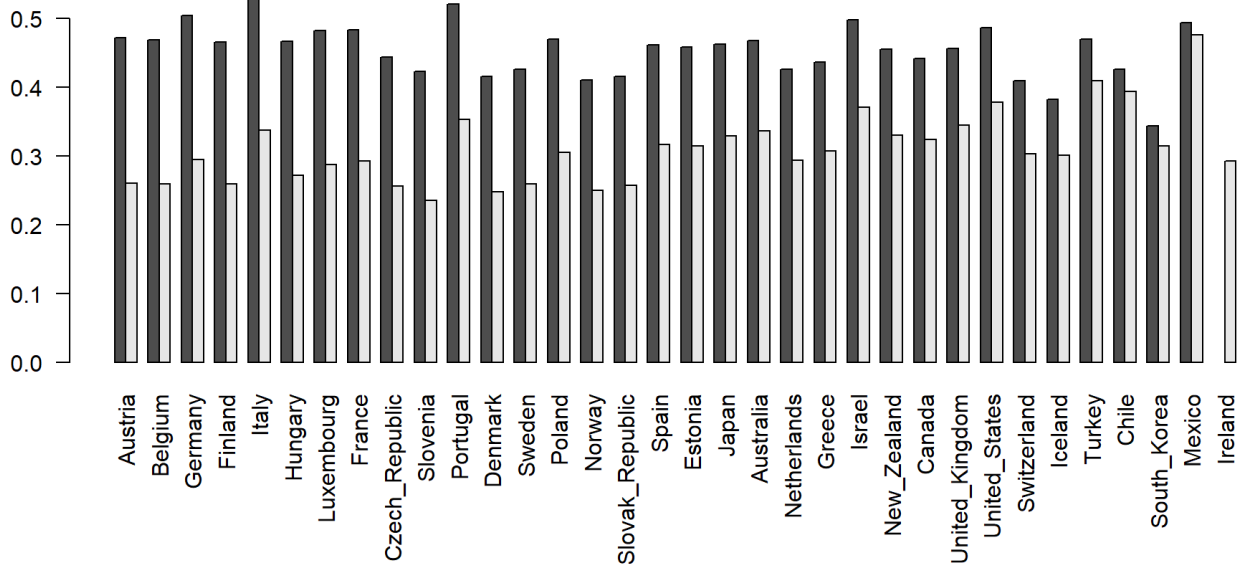


- 나라 이름이 가리지 않도록 par("mai") 를 조정

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.5, 0.8, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        las = 2)
```



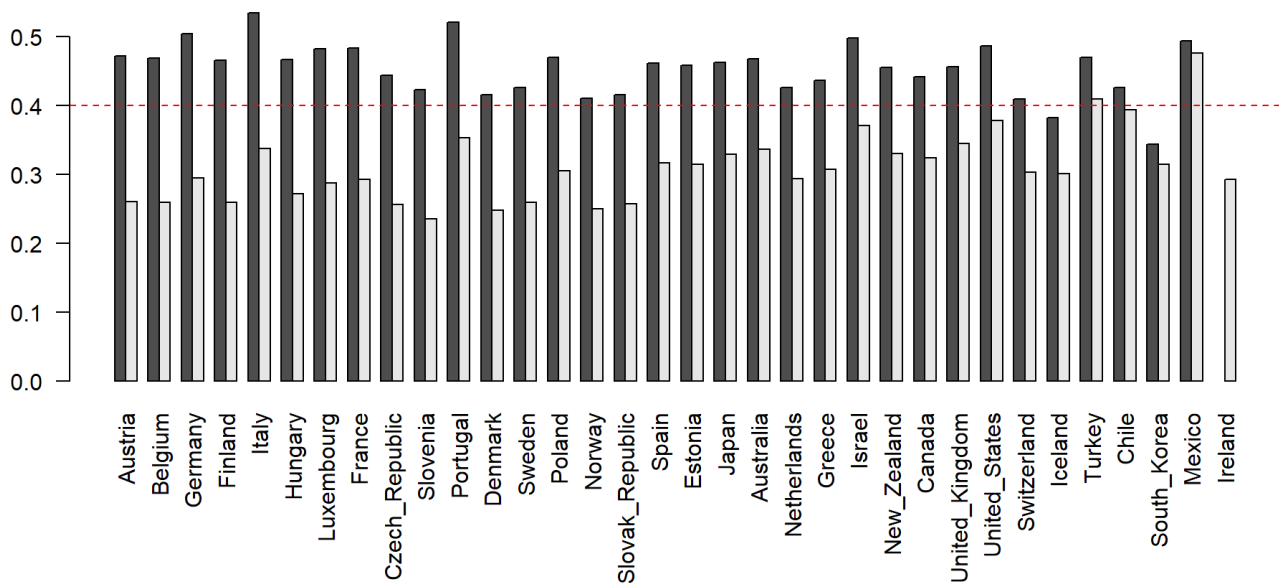
```
par(old_par)
```

- 불평등이 심하다고 판단하는 Gini 계수 0.4를 경계로 나눠 보면,

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.5, 0.8, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        las = 2)
abline(h = 0.4, lty = 2, col = "red")
```



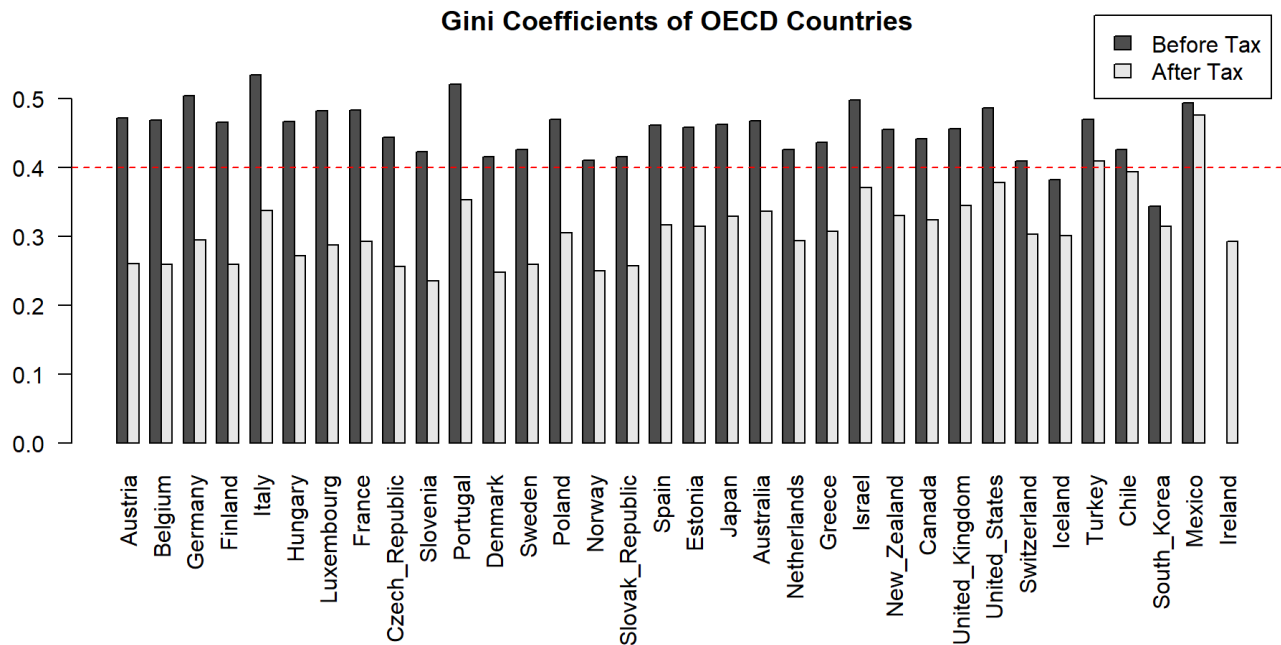
```
par(old_par)
```

- 범례와 메인 타이틀 추가. 좌표에 유의

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

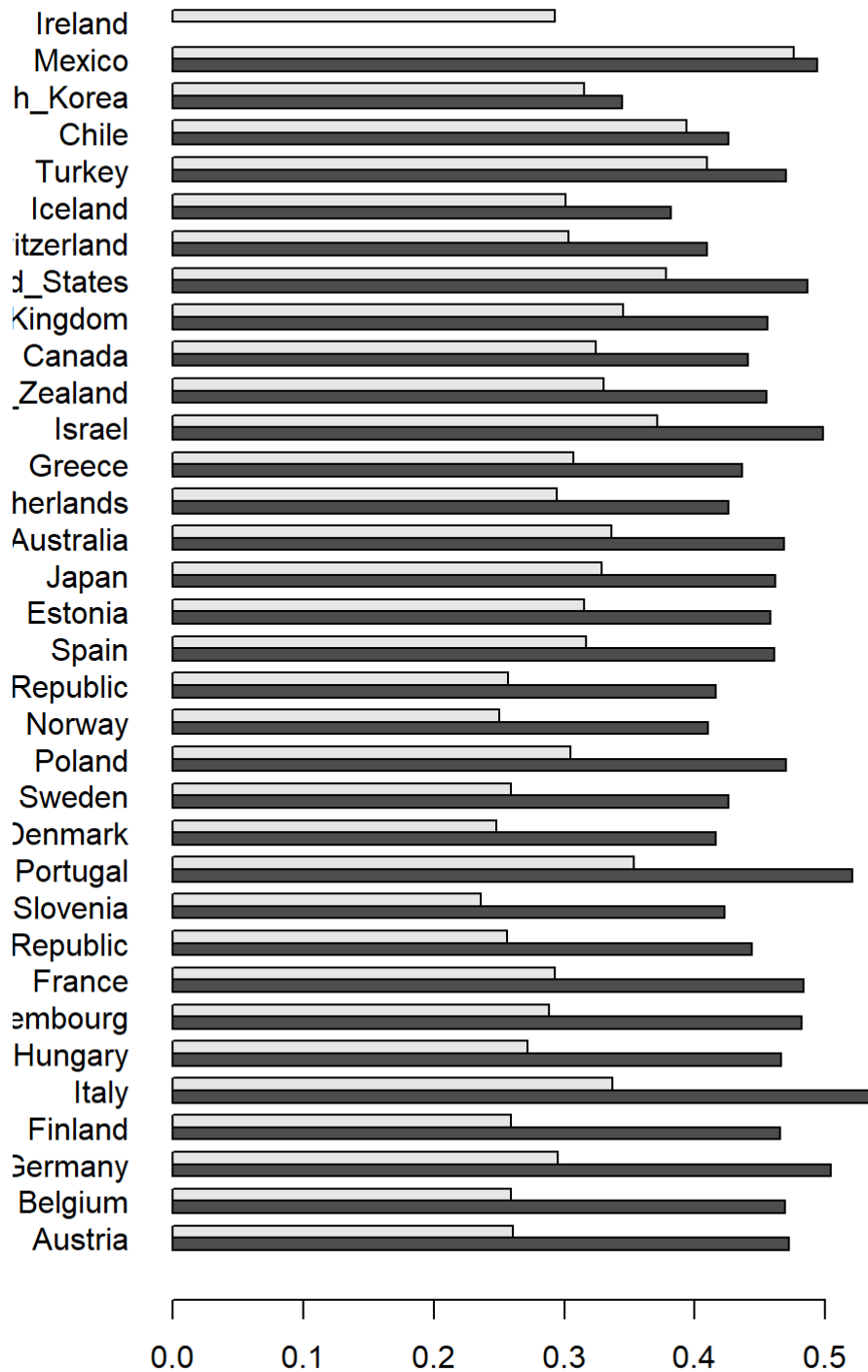
```
par("mai" = c(1.5, 0.8, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        legend.text = c("Before Tax", "After Tax"),
        args.legend = list(x = 105, y = 0.62),
        las = 2)
abline(h = 0.4, lty = 2, col = "red")
title(main = "Gini Coefficients of OECD Countries")
```



```
par(old_par)
```

- 이번에는 막대를 눕히는 방법을 생각해 보자. 옆으로 눕히면서 `las = 1` 로 설정하면,

```
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        horiz = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        las = 1)
```

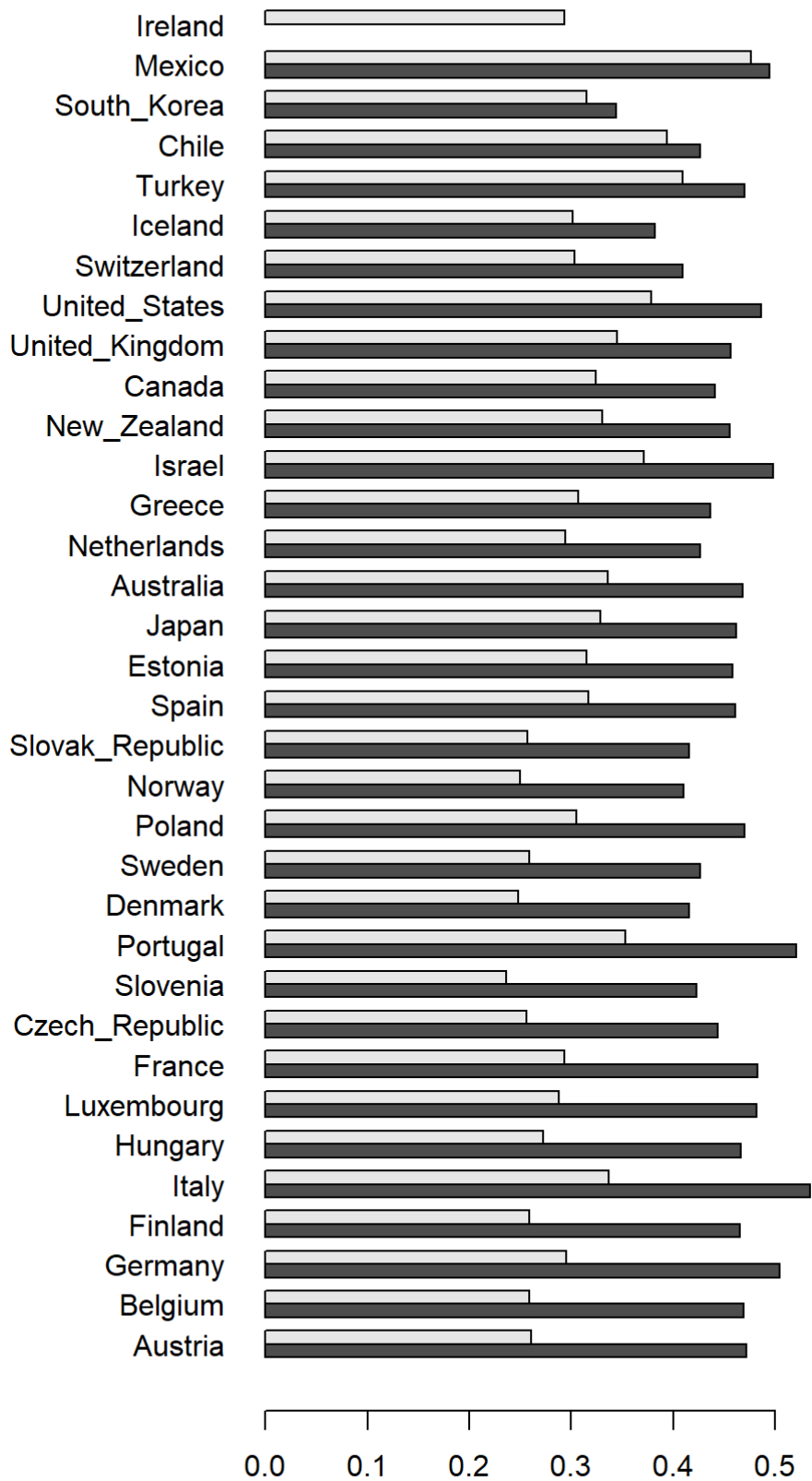


- 역시 나라 이름이 가리지 않도록 `par("mai")` 를 조정.

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.0, 1.5, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])),
        beside = TRUE,
        horiz = TRUE,
        names.arg = Gini_b_a$Country[o_improvement],
        las = 1)
```



```
par(old_par)
```

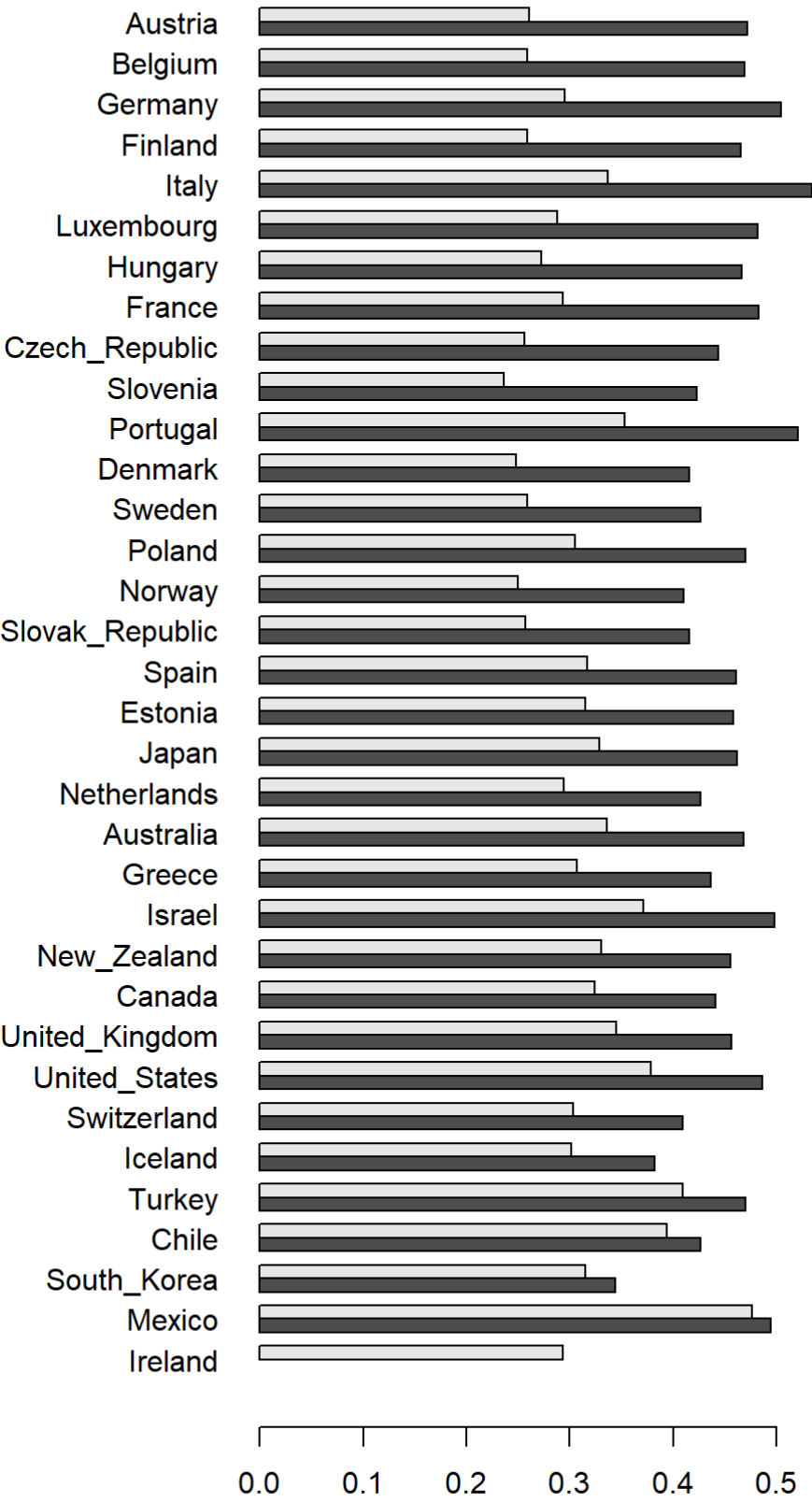
- 개선도가 낮은 순서대로 밑에서 올라가도록 다시 그리면,

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.0, 1.5, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[order(Gini_b_a$Improvement,
                                   na.last = FALSE), 2:3])),
        beside = TRUE,
        horiz = TRUE,
        names.arg = Gini_b_a$Country[order(Gini_b_a$Improvement,
                                             na.last = FALSE)],
        las = 1)
```



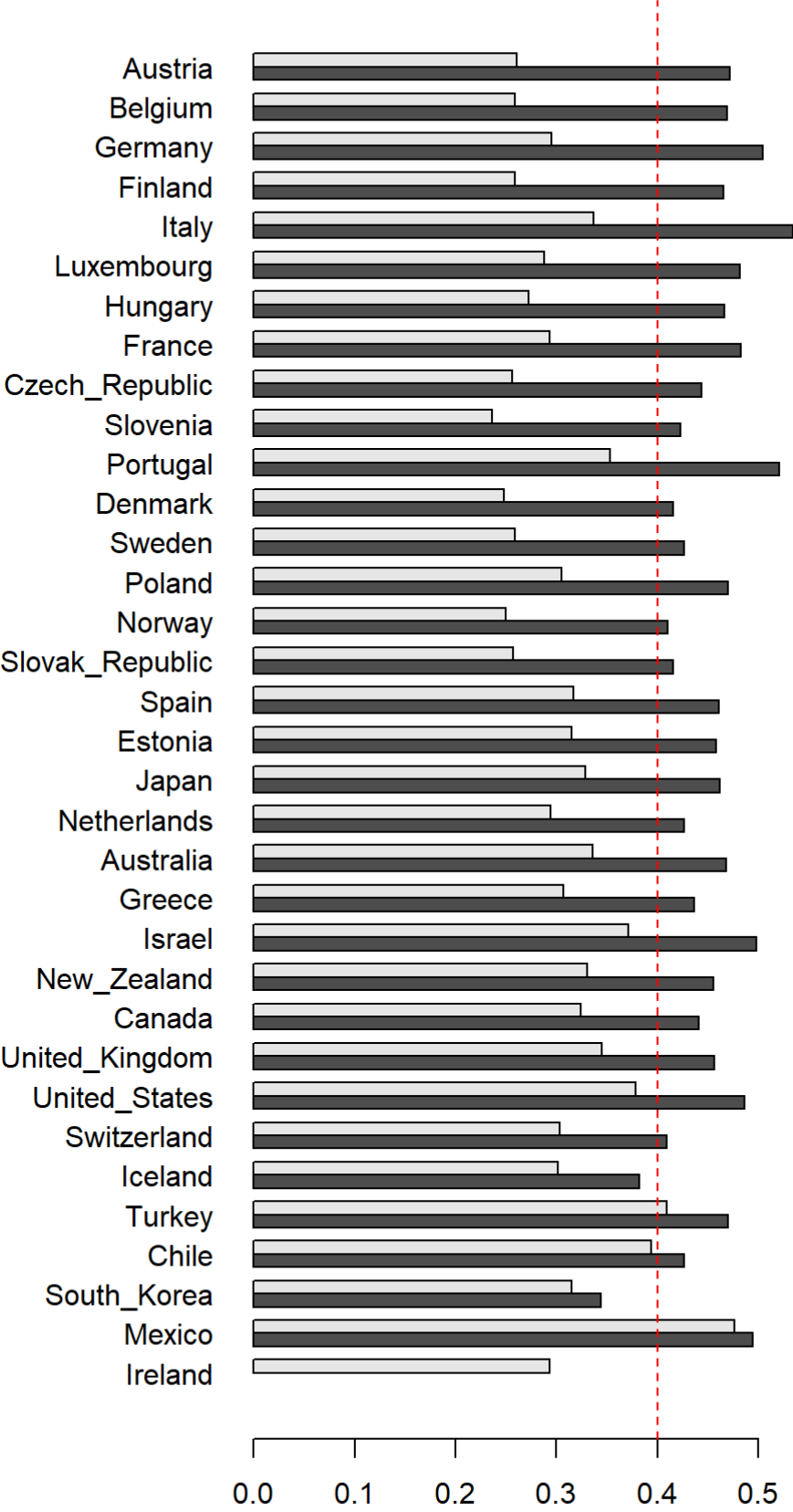


par(old\_par)

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.0, 1.5, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[order(Gini_b_a$Improvement,
                                   na.last = FALSE), 2:3])),
        beside = TRUE,
        horiz = TRUE,
        names.arg = Gini_b_a$Country[order(Gini_b_a$Improvement,
                                             na.last = FALSE)],
        las = 1)
abline(v = 0.4, lty = 2, col = "red")
```



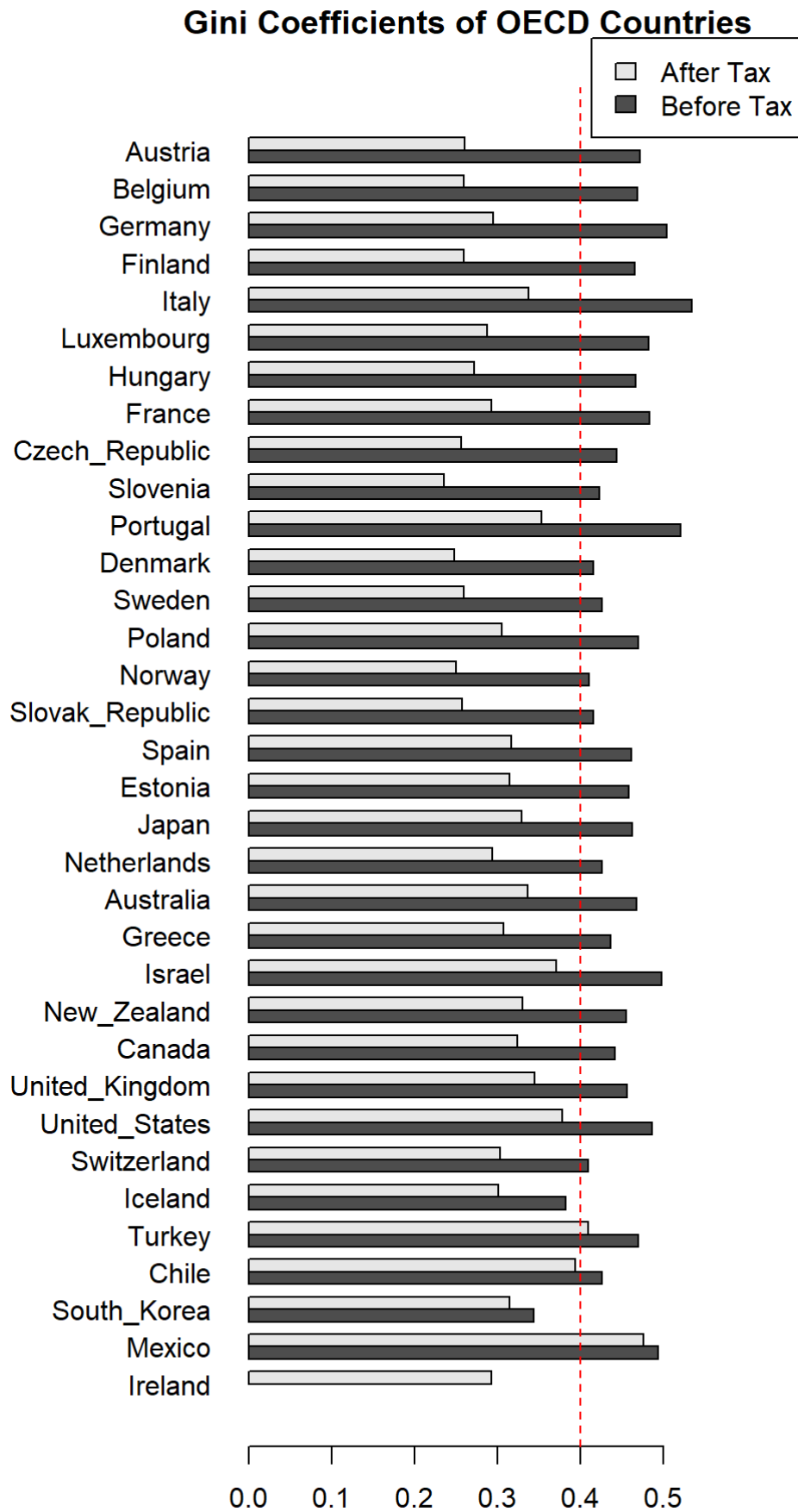
par(old\_par)

- 범례 및 메인 타이틀 추가.

```
old_par <- par(no.readonly = TRUE)
par("mai")
```

```
## [1] 1.02 0.82 0.82 0.42
```

```
par("mai" = c(1.0, 1.5, 0.8, 0.8))
barplot(as.matrix(t(Gini_b_a[order(Gini_b_a$Improvement,
                                   na.last = FALSE), 2:3])),
        beside = TRUE,
        horiz = TRUE,
        names.arg = Gini_b_a$Country[order(Gini_b_a$Improvement,
                                             na.last = FALSE)],
        legend.text = c("Before Tax", "After Tax"),
        args.legend = list(x = 0.67, y = 110),
        las = 1)
abline(v = 0.4, lty = 2, col = "red")
title(main = "Gini Coefficients of OECD Countries")
```



```
par(old_par)
```

# ggplot

## Data reshaping

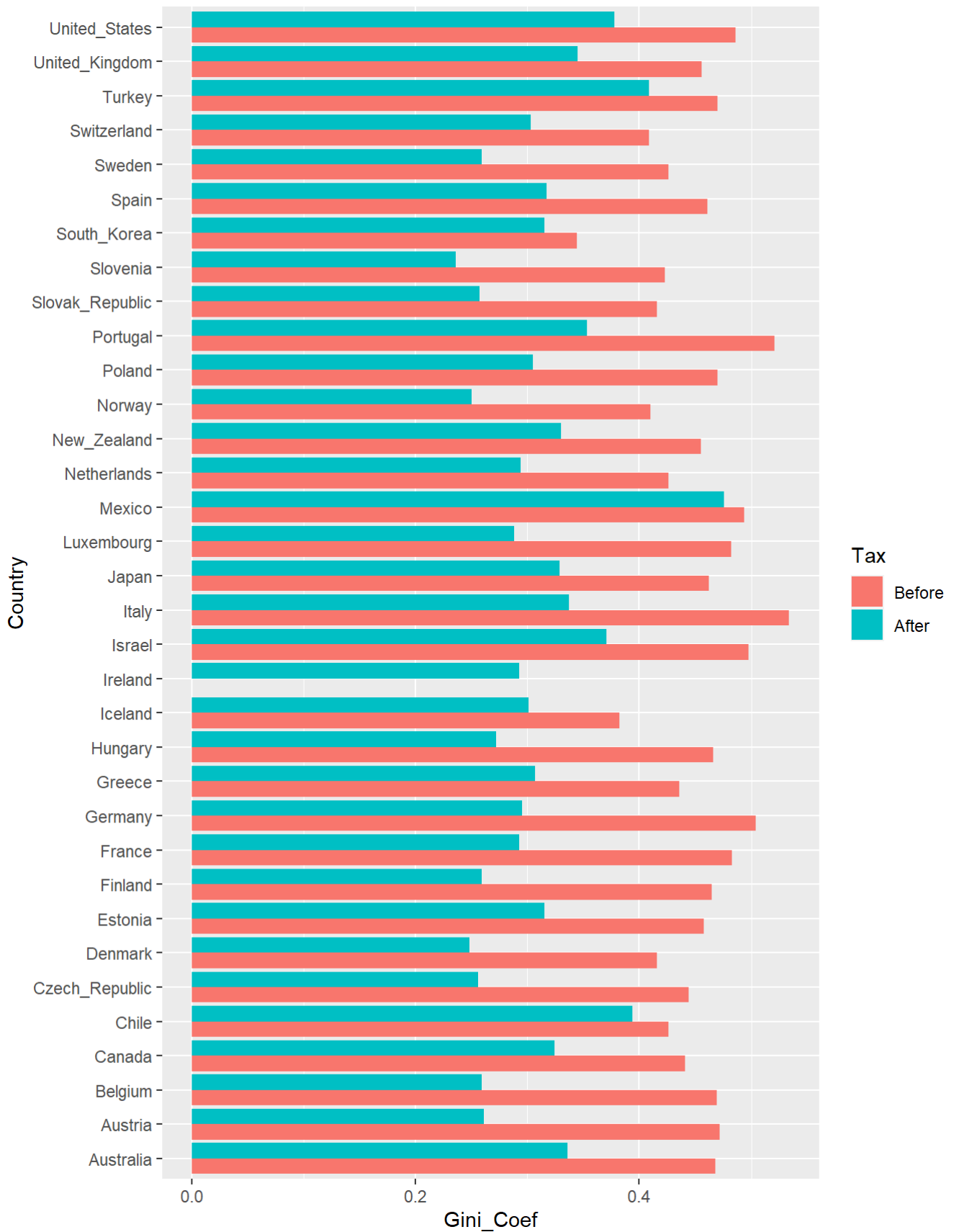
- reshape2 package 를 검색 목록에 등록

```
# library(reshape2)
# (Gini_b_a_melt <- melt(Gini_b_a,
#                         id.vars = "Country",
#                         measure.vars = c("Before", "After"),
#                         variable.name = "Tax",
#                         value.name = "Gini_Coef"))
# str(Gini_b_a_melt)
Gini_Coef <- c(as.matrix(Gini_b_a[, 2:3]))
N <- length(Gini_Coef)
Country <- rep(Gini_b_a[, 1], length.out = N)
Tax <- gl(2, length(Gini_b_a[, 1]), N,
          labels = c("Before", "After"))
Gini_b_a_tbl <- data.frame(Country, Gini_Coef, Tax)
```

- ggplot2 등록 후 geom\_bar()

```
library(ggplot2)
ggplot(data = Gini_b_a_tbl,
       mapping = aes(x = Country,
                     y = Gini_Coef,
                     fill = Tax)) +
  geom_bar(stat = "identity",
          position = "dodge") +
  coord_flip()
```

```
## Warning: Removed 1 rows containing missing values (`geom_bar()`).
```



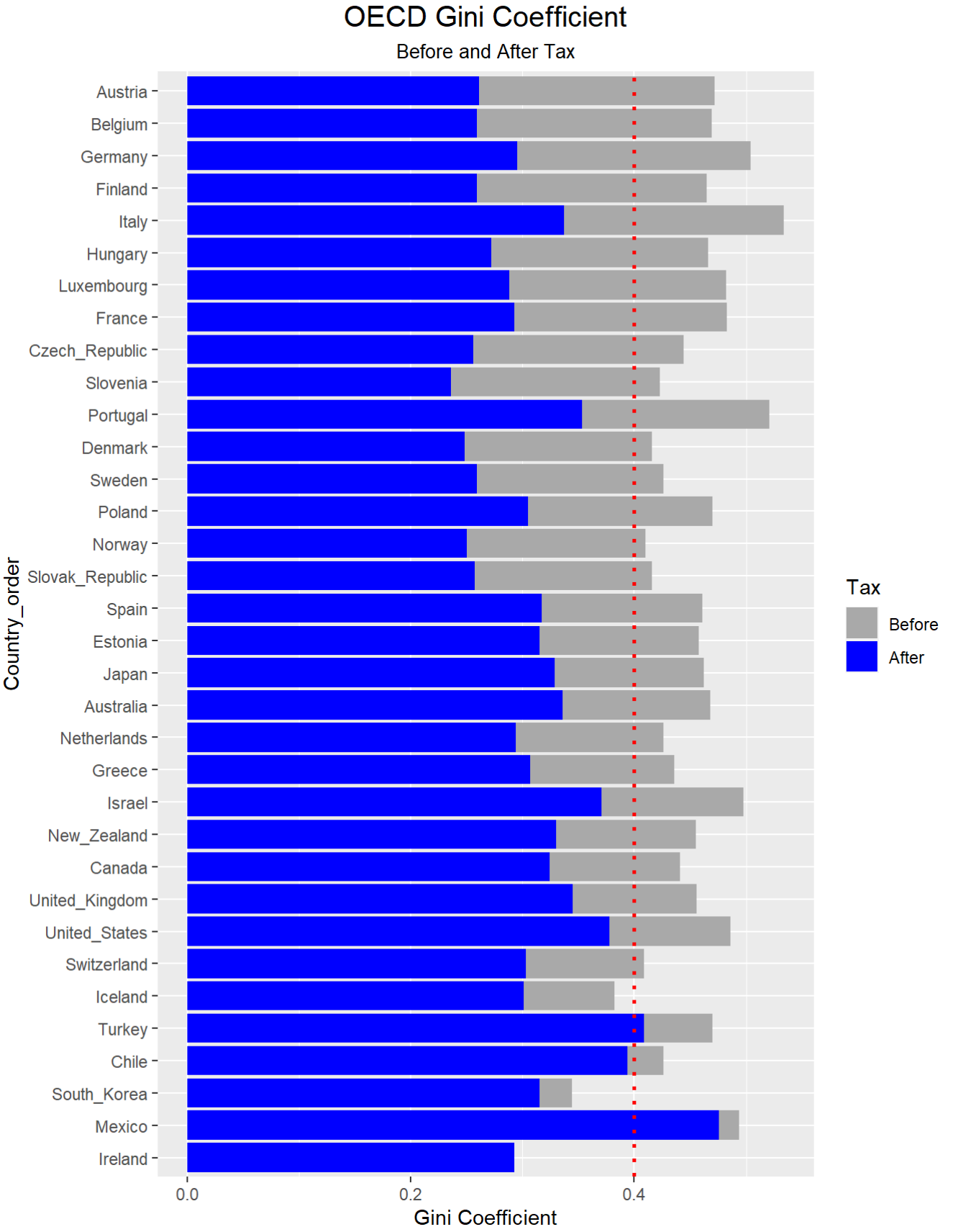
```
Gini_b_a$Country_order <- factor(Gini_b_a$Country,
                                levels = Gini_b_a$Country[rev(o_improvement)])
# Gini_b_a_order_melt <- melt(Gini_b_a,
#                             id.vars = "Country_order",
#                             measure.vars = c("Before", "After"),
#                             variable.name = "Tax",
#                             value.name = "Gini_Coef")
# str(Gini_b_a_order_melt)
Country_order <- rep(Gini_b_a[, "Country_order"], length.out = N)
Tax <- gl(2, length(Gini_b_a[, "Country_order"]), N,
        labels = c("Before", "After"))
Gini_b_a_order_tbl <- data.frame(Country_order, Gini_Coef, Tax)
```

- Gini\_b\_a\_order\_tbl 의 Country\_order 가 개선도 올림차순으로 정리되어 있는 factor 이기 때문에 그대로 활용.

```
ggplot(data = Gini_b_a_order_tbl,
       mapping = aes(x = Country_order,
                     y = Gini_Coef,
                     fill = Tax)) +
  geom_bar(stat = "identity",
          position = "identity",
          na.rm = TRUE) +
  geom_hline(yintercept = 0.4,
            color = "red",
            linetype = 3,
            size = 1) +
  scale_fill_manual(values = c("darkgrey", "blue")) +
  # scale_fill_brewer(type = "qual", palette = "Set1", direction = -1) +
  labs(title = "OECD Gini Coefficient",
       subtitle = "Before and After Tax",
       y = "Gini Coefficient") +
  theme(plot.title = element_text(size = 15, hjust = 0.5),
        plot.subtitle = element_text(size = 10, hjust = 0.5)) +
  coord_flip()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ■ Please use `linewidth` instead.
```





```
save.image(file = "Gini_OECD.RData")
```