Gini_OECD

coop711

2018-04-15

Data 작업

- OECD 국가들의 Gini계수 읽어들이기. 세전과 세후로 구분. 자료구조로 인하여 `sep="\t"` 을 사용한 것에 유의

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

```
## 'data.frame':    34 obs. of  8 variables:
##  $ V1: Factor w/ 34 levels "Australia","Austria",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ 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: Factor w/ 34 levels "Australia","Austria",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ 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 ...
```

- 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
## 1        Australia  0.468 0.336
## 2          Austria  0.472 0.261
## 3          Belgium  0.469 0.259
## 4           Canada  0.441 0.324
## 5            Chile  0.426 0.394
## 6   Czech_Republic  0.444 0.256
## 7          Denmark  0.416 0.248
## 8          Estonia  0.458 0.315
## 9          Finland  0.465 0.259
## 10          France  0.483 0.293
## 11         Germany  0.504 0.295
## 12          Greece  0.436 0.307
## 13         Hungary  0.466 0.272
## 14         Iceland  0.382 0.301
## 15         Ireland     NA 0.293
## 16          Israel  0.498 0.371
## 17           Italy  0.534 0.337
## 18           Japan  0.462 0.329
## 19      Luxembourg  0.482 0.288
## 20          Mexico  0.494 0.476
## 21     Netherlands  0.426 0.294
## 22     New_Zealand  0.455 0.330
## 23          Norway  0.410 0.250
## 24          Poland  0.470 0.305
## 25        Portugal  0.521 0.353
## 26 Slovak_Republic  0.416 0.257
## 27        Slovenia  0.423 0.236
## 28     South_Korea  0.344 0.315
## 29           Spain  0.461 0.317
## 30          Sweden  0.426 0.259
## 31     Switzerland  0.409 0.303
## 32          Turkey  0.470 0.409
## 33  United_Kingdom  0.456 0.345
## 34   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
## 1        Australia  0.468 0.336       0.132
## 2          Austria  0.472 0.261       0.211
## 3          Belgium  0.469 0.259       0.210
## 4           Canada  0.441 0.324       0.117
## 5            Chile  0.426 0.394       0.032
## 6   Czech_Republic  0.444 0.256       0.188
## 7          Denmark  0.416 0.248       0.168
## 8          Estonia  0.458 0.315       0.143
## 9          Finland  0.465 0.259       0.206
## 10          France  0.483 0.293       0.190
## 11         Germany  0.504 0.295       0.209
## 12          Greece  0.436 0.307       0.129
## 13         Hungary  0.466 0.272       0.194
## 14         Iceland  0.382 0.301       0.081
## 15         Ireland     NA 0.293          NA
## 16          Israel  0.498 0.371       0.127
## 17           Italy  0.534 0.337       0.197
## 18           Japan  0.462 0.329       0.133
## 19      Luxembourg  0.482 0.288       0.194
## 20          Mexico  0.494 0.476       0.018
## 21     Netherlands  0.426 0.294       0.132
## 22     New_Zealand  0.455 0.330       0.125
## 23          Norway  0.410 0.250       0.160
## 24          Poland  0.470 0.305       0.165
## 25        Portugal  0.521 0.353       0.168
## 26 Slovak_Republic  0.416 0.257       0.159
## 27        Slovenia  0.423 0.236       0.187
## 28     South_Korea  0.344 0.315       0.029
## 29           Spain  0.461 0.317       0.144
## 30          Sweden  0.426 0.259       0.167
## 31     Switzerland  0.409 0.303       0.106
## 32          Turkey  0.470 0.409       0.061
## 33  United_Kingdom  0.456 0.345       0.111
## 34   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
## 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
## 2          Austria  0.472 0.261       0.211
## 15         Ireland     NA 0.293          NA
```

- 개선도가 높은 순서로 나라명을 나열하려면, `decreasing = TRUE` 추가.

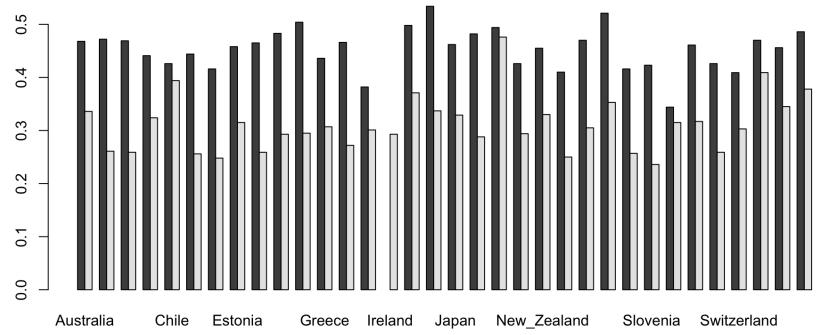
```
Gini_b_a[order(Gini_b_a$Improvement, decreasing = TRUE), ]
```

```
##            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
## 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 transpose시킨 후 `barplot()` 을 적용. 또한 transpose를 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)
```



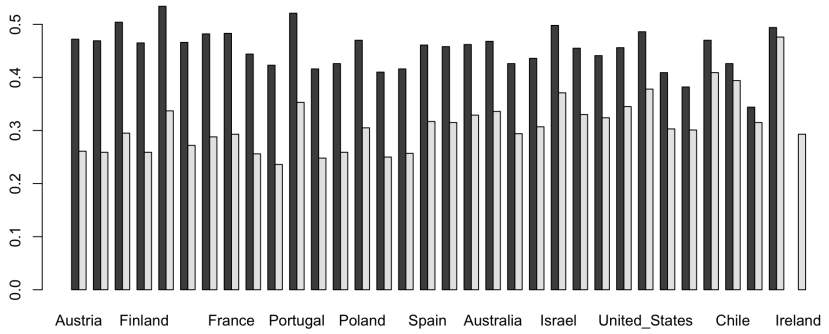
- 개선도 순서(내림차순)를 `o_improvement` 로 저장하여 지속적으로 활용.

```
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
## 34 Levels: Australia Austria Belgium Canada Chile ... United_States
```

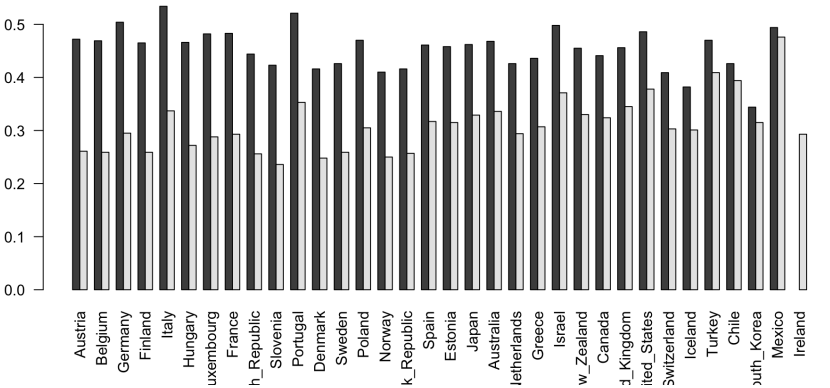
- 개선도 순서대로 막대를 늘어세우면,

```
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)
```

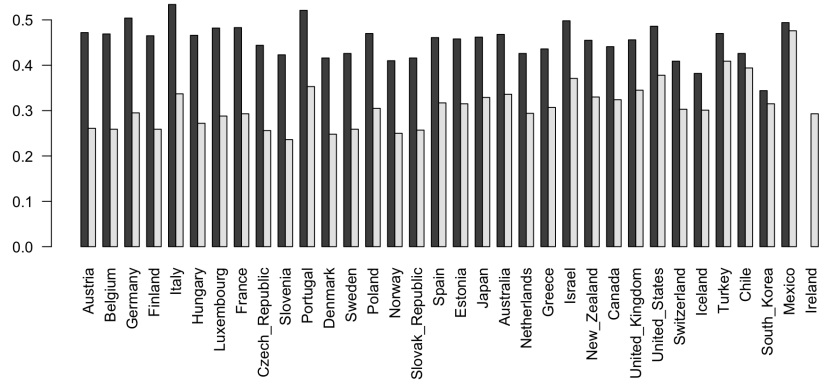


- 나라 이름이 가리지 않도록 `par("mai")` 를 조정

```
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$C
ountry[o_improvement], las=2)
```



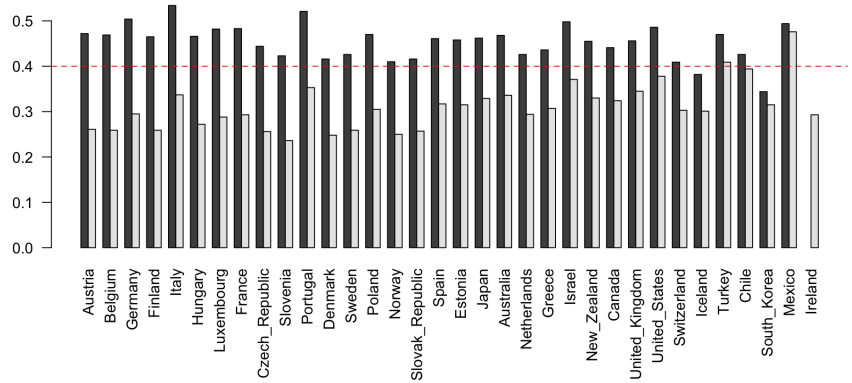

```
par(old_par)
```

- 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$C
ountry[o_improvement], las=2)
abline(h=0.4, lty=2, col="red")
```



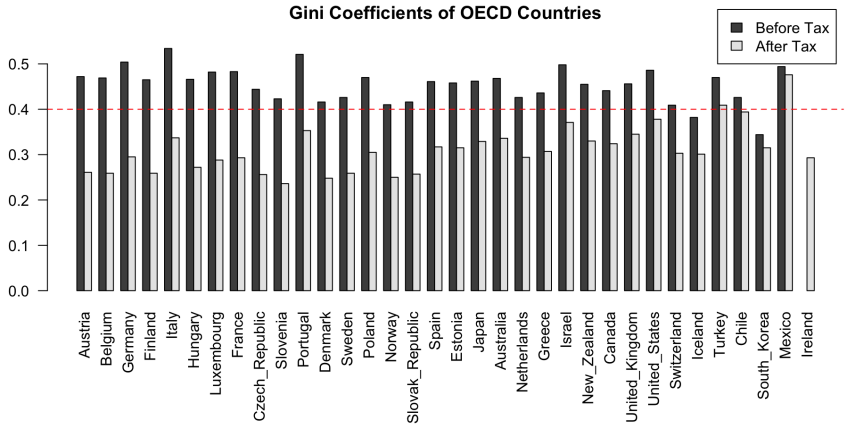

```
par(old_par)
```

- 범례와 메인 타이틀 추가. 좌표에 유의

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

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

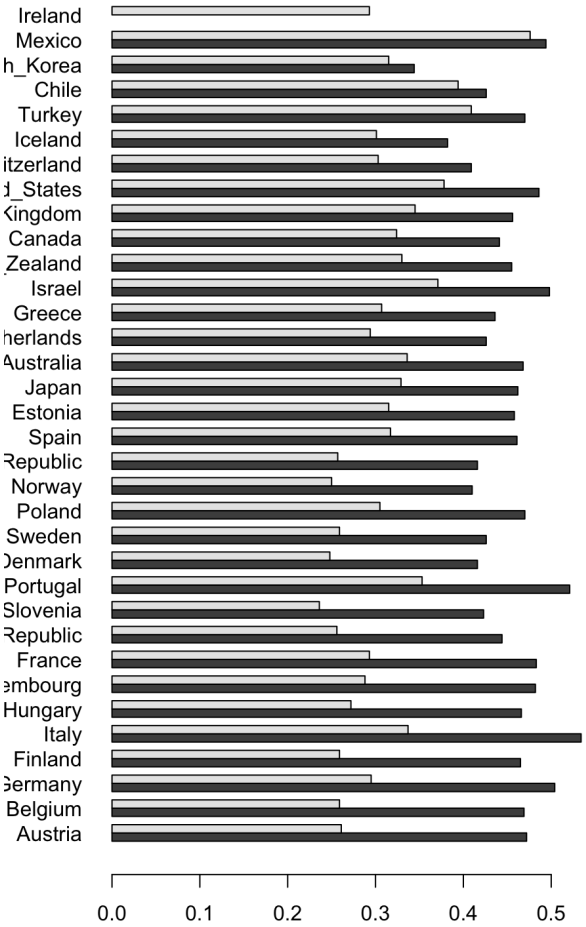
```
par("mai"= c(1.5, 0.8, 0.8, 0.4))
barplot(as.matrix(t(Gini_b_a[o_improvement, 2:3])), beside=TRUE, names.arg=Gini_b_a$C
ountry[o_improvement], legend.text=c("Before Tax", "After Tax"), args.legend=list(x=1
05, y=0.62), las=2)
abline(h=0.4, lty=2, col="red")
title(main="Gini Coefficients of OECD Countries")
```




```
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.ar
g=Gini_b_a$Country[o_improvement], las=1)
```

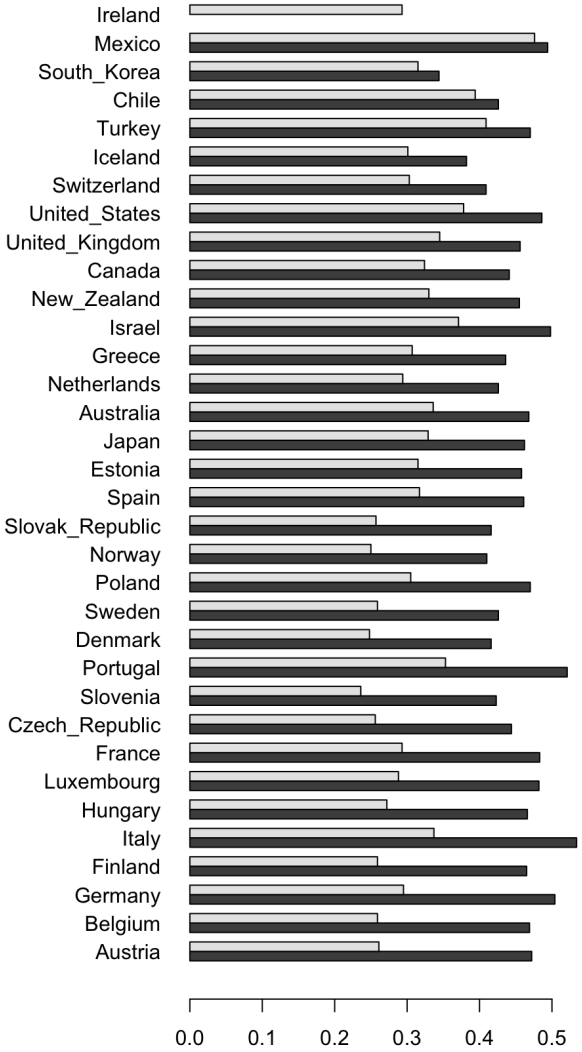


- 역시 나라 이름이 가리지 않도록 `par("mai")` 를 조정.

```
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)
```



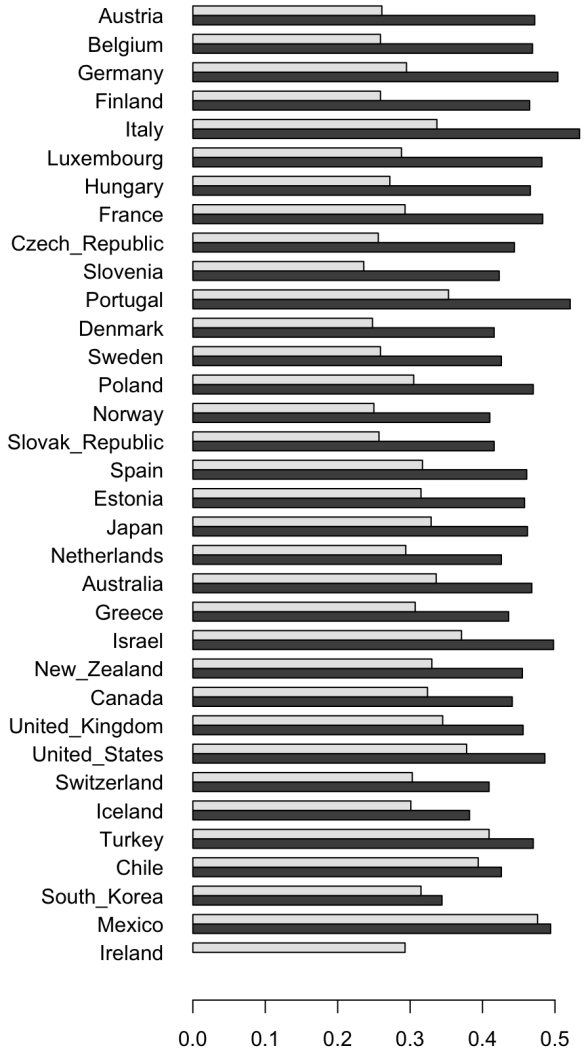

```
par(old_par)
```

- 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)
```

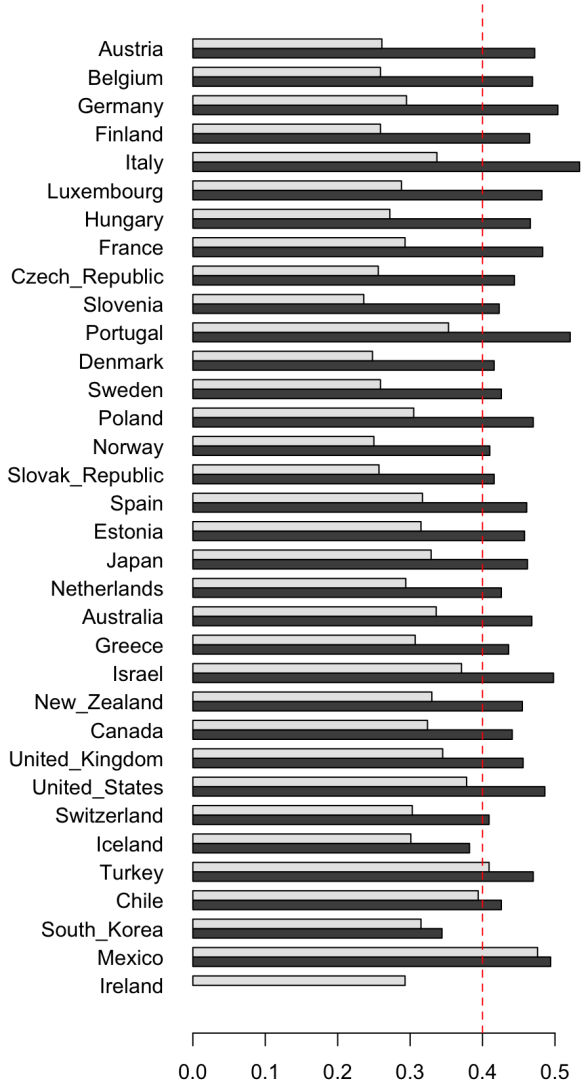
- 이 때, Ireland가 맨 위에 올라오는 게 보기 좋지 않으므로, `na.last=FALSE` 를 추가한 것임.

    - 세전 Gini 계수 0.4를 경계로 나눠보면

```
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")
```



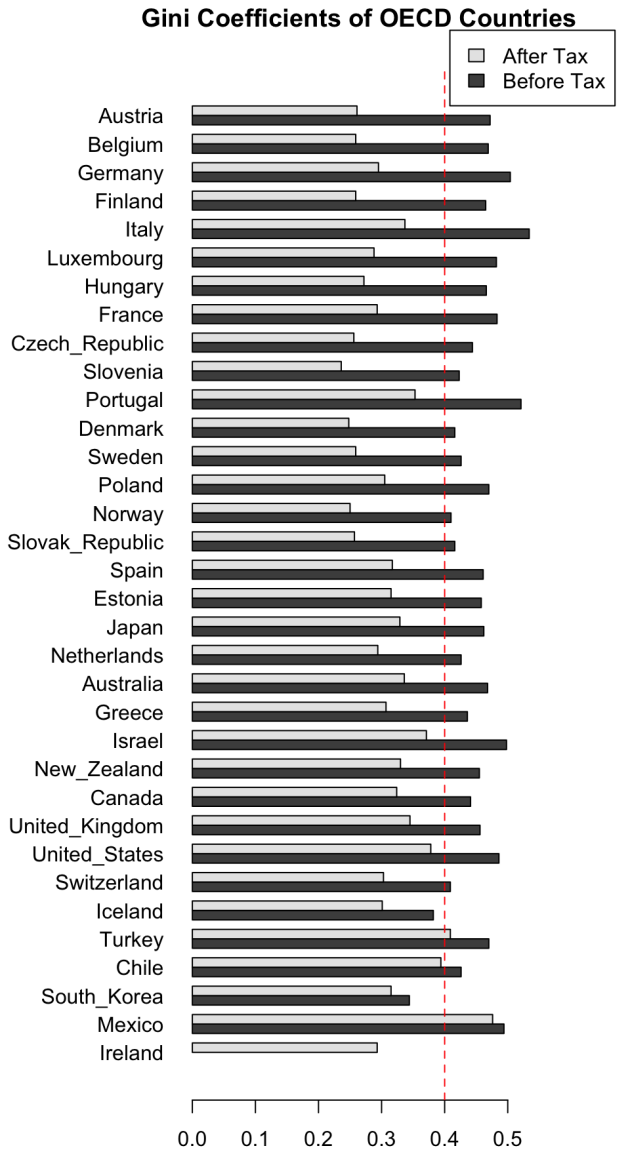

```
par(old_par)
```

- 범례 및 메인 타이틀 추가. 시행착오를 거쳐 구한 좌표에 유의할 것.

```
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"))
```

```
##            Country    Tax Gini_Coef
## 1        Australia Before     0.468
## 2          Austria Before     0.472
## 3          Belgium Before     0.469
## 4           Canada Before     0.441
## 5            Chile Before     0.426
## 6   Czech_Republic Before     0.444
## 7          Denmark Before     0.416
## 8          Estonia Before     0.458
## 9          Finland Before     0.465
## 10          France Before     0.483
## 11         Germany Before     0.504
## 12          Greece Before     0.436
## 13         Hungary Before     0.466
## 14         Iceland Before     0.382
## 15         Ireland Before        NA
## 16          Israel Before     0.498
## 17           Italy Before     0.534
## 18           Japan Before     0.462
## 19      Luxembourg Before     0.482
## 20          Mexico Before     0.494
## 21     Netherlands Before     0.426
## 22     New_Zealand Before     0.455
## 23          Norway Before     0.410
## 24          Poland Before     0.470
## 25        Portugal Before     0.521
## 26 Slovak_Republic Before     0.416
## 27        Slovenia Before     0.423
## 28     South_Korea Before     0.344
## 29           Spain Before     0.461
## 30          Sweden Before     0.426
## 31     Switzerland Before     0.409
## 32          Turkey Before     0.470
## 33  United_Kingdom Before     0.456
## 34   United_States Before     0.486
## 35       Australia  After     0.336
## 36         Austria  After     0.261
## 37         Belgium  After     0.259
## 38          Canada  After     0.324
## 39           Chile  After     0.394
## 40  Czech_Republic  After     0.256
## 41         Denmark  After     0.248
## 42         Estonia  After     0.315
## 43         Finland  After     0.259
## 44          France  After     0.293
## 45         Germany  After     0.295
## 46          Greece  After     0.307
## 47         Hungary  After     0.272
## 48         Iceland  After     0.301
## 49         Ireland  After     0.293
## 50          Israel  After     0.371
## 51           Italy  After     0.337
## 52           Japan  After     0.329
## 53      Luxembourg  After     0.288
## 54          Mexico  After     0.476
## 55     Netherlands  After     0.294
## 56     New_Zealand  After     0.330
## 57          Norway  After     0.250
## 58          Poland  After     0.305
## 59        Portugal  After     0.353
## 60 Slovak_Republic  After     0.257
## 61        Slovenia  After     0.236
## 62     South_Korea  After     0.315
## 63           Spain  After     0.317
## 64          Sweden  After     0.259
## 65     Switzerland  After     0.303
## 66          Turkey  After     0.409
## 67  United_Kingdom  After     0.345
## 68   United_States  After     0.378
```

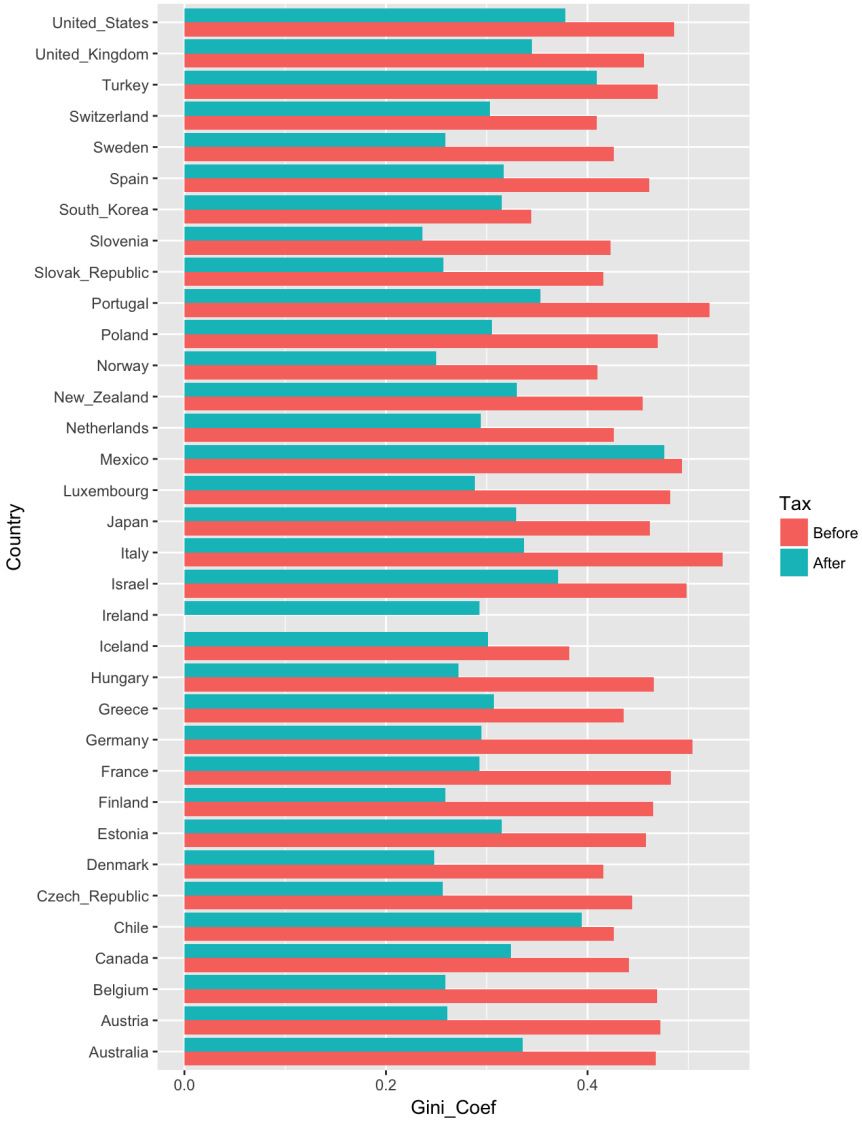
```
str(Gini_b_a_melt)
```

```
## 'data.frame':    68 obs. of  3 variables:
##  $ Country  : Factor w/ 34 levels "Australia","Austria",..: 1 2 3 4 5 6 7 8 9 10
...
##  $ Tax      : Factor w/ 2 levels "Before","After": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Gini_Coef: num  0.468 0.472 0.469 0.441 0.426 0.444 0.416 0.458 0.465 0.483 ...
```

- ggplot2 등록 후 geom_bar()

```
library(ggplot2)
ggplot(data = Gini_b_a_melt,
       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).
```



- 개선도 순서대로 늘어세우려면 그 순서를 `level` 로 갖는 `factor` 로 만들어야함. `o_improvement` 가 내림차순으로 정리되어 있는 순서이기 때문에 `rev(o_improvement)` 는 올림차순으로 정리되어 있는 순서임. 따라서,

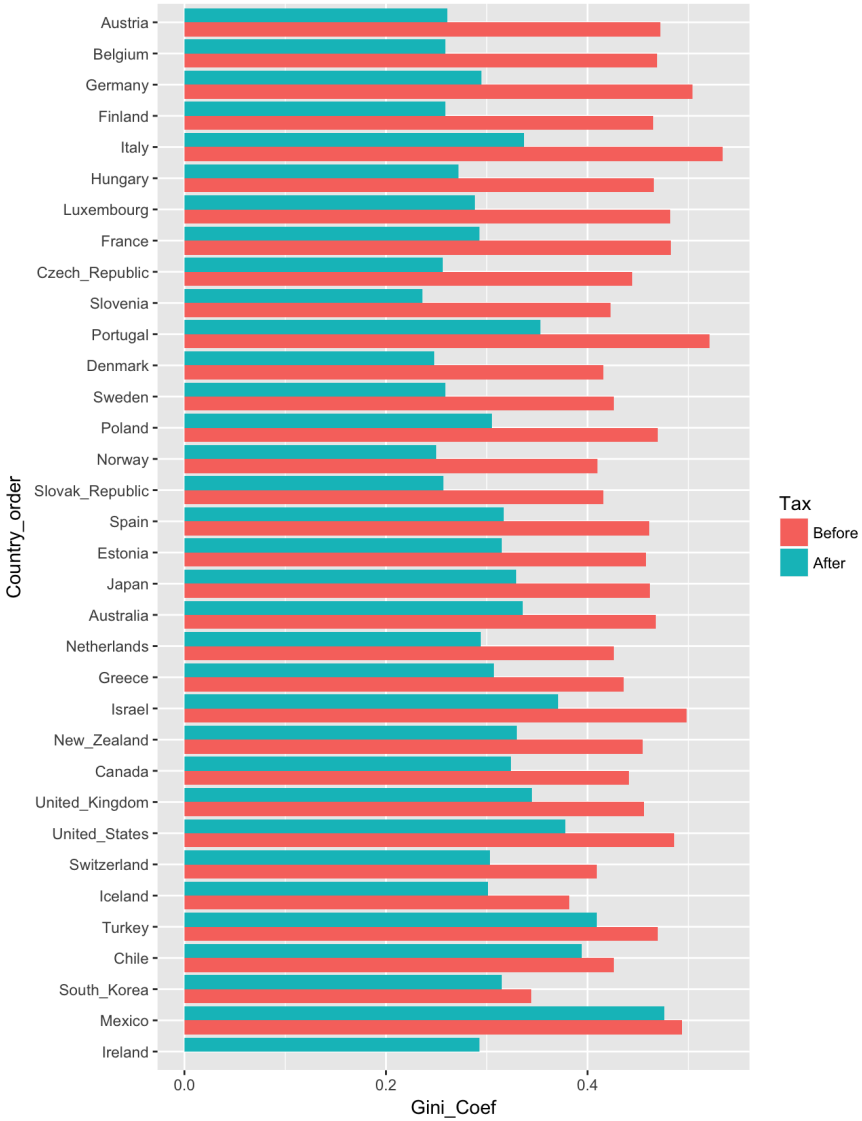
```
Gini_b_a$Country_order <- factor(Gini_b_a$Country,
                                 levels = Gini_b_a$Country[rev(o_improvement)])
Gini_b_a_order_melt <- melt(Gini_b_a,
                            id.vars = "Country_order",
                            measure.vars = c("Before", "After"),
                            variable.name = "Tax",
                            value.name = "Gini_Coef")
str(Gini_b_a_order_melt)
```

```
## 'data.frame':    68 obs. of  3 variables:
##  $ Country_order: Factor w/ 34 levels "Ireland","Mexico",..: 15 34 33 10 4 26 23 1
7 31 27 ...
##  $ Tax          : Factor w/ 2 levels "Before","After": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Gini_Coef    : num  0.468 0.472 0.469 0.441 0.426 0.444 0.416 0.458 0.465 0.483
...
```

- `Gini_b_a_order_melt` 의 `Country_order` 가 개선도 올림차순으로 정리되어 있는 `factor` 이기 때문에 그대로 활용하면 됨.

```
ggplot(data = Gini_b_a_order_melt,
       mapping = aes(x = Country_order, y = Gini_Coef, fill = Tax)) +
  geom_bar(stat = "identity", position = "dodge") +
coord_flip()
```

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



- 한글 제목 등의 세부 작업은 차후에

뒷 마무리

```
save(file = "Gini_OECD1801.RData", list = ls())
# savehistory("Gini_OECD102.Rhistory")
```