

# Income Inequality vs Index of Health and Social Problems : Piping

coop711

2018-05-09

## Data Reading

Equality Trust에서 기부금을 받고 제공하는 두 종류의 자료 중 23개 국가의 각종 지표를 비교한 자료에 World Bank에서 발표하는 GDP자료 ([https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(PPP\)\\_per\\_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita))를 추가한 자료를 data 단계에서 읽어들이고 필요한 부분만 정리한 RData 파일을 읽어들이면,

```
library(magrittr)
library(knitr)
library(RColorBrewer)
load("Inequality_Index_HS.RData", verbose = TRUE)
```

```
## Loading objects:
##   data_full
##   v_names
##   country_na
##   v_names_index
##   data_21
```

```
# data_full <- read.csv("../data/international-inequality_GDP_csv", stringsAsFactors
= FALSE)
# data_full <- read.csv("../data/international-inequality_GDP_csv", stringsAsFactors
= TRUE)
data_21 %>% str
```

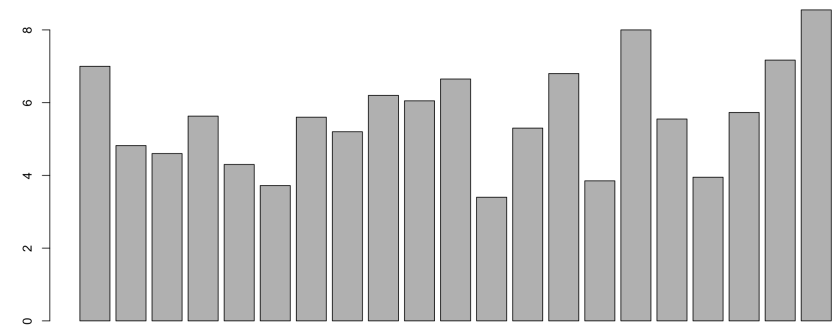
```
## 'data.frame':   21 obs. of  4 variables:
##  $ Country      : chr  "Australia" "Austria" "Belgium" "Canada" ...
##  $ Income_inequality: num  7 4.82 4.6 5.63 4.3 3.72 5.6 5.2 6.2 6.05 ...
##  $ Index_HS      : num  0.07 0.01 -0.23 -0.07 -0.19 -0.43 0.05 -0.06 0.38 0.25
##  ...
##  $ GDP_WB       : int  45926 47682 43435 45066 45537 40676 39328 46401 26851 4
9393 ...
```

## Plots

### Barplots for Income Inequalities

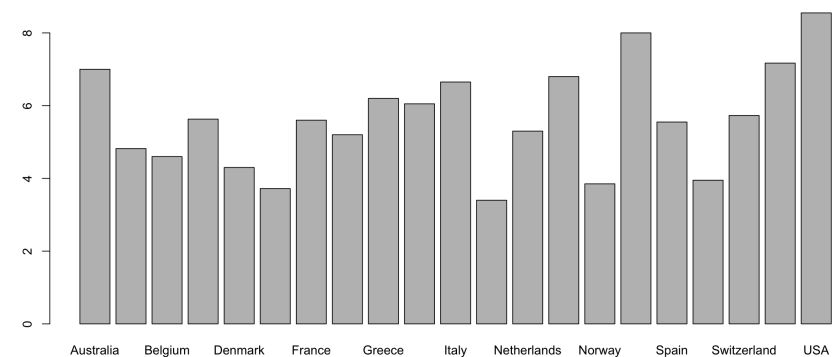
기본형

```
# par(mai = c(2.0, 0.8, 0.8, 0.4) + 0.2)
data_21 %>%
  .$Income_inequality %>%
  barplot
```



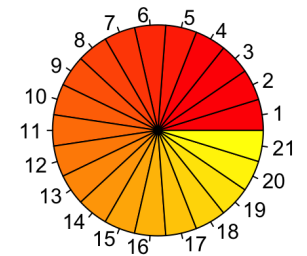
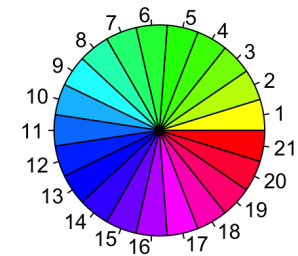
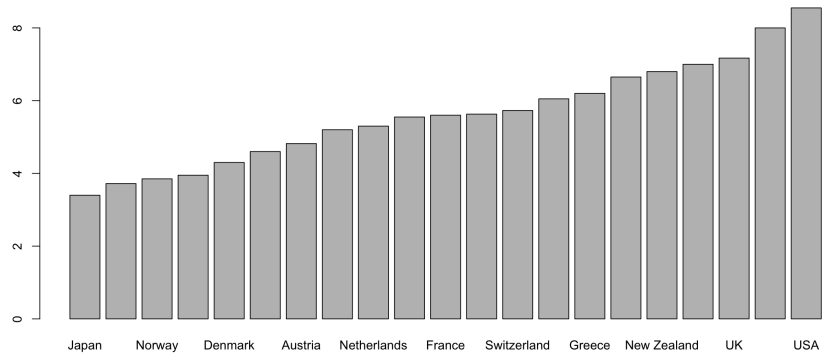
나라이름 붙이기

```
data_21 %>%
  .$Income_inequality %>%
  barplot(names.arg = data_21$Country)
```



## 5분위배율 순으로 정리하기

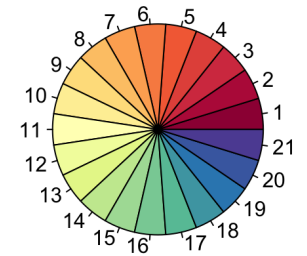
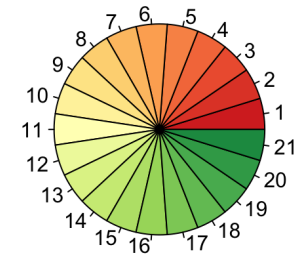
```
o_fifth <- data_21 %>% .$Income_inequality %>% order
b <- data_21 %>%
  .$Income_inequality %>%
  `[^(o_fifth) %>%
  barplot(names.arg = data_21$Country %>% `[^(o_fifth))
```



## Colors

```
N <- data_21 %>% nrow
cols_brewer <- colorRampPalette(brewer.pal(9, "RdYlGn"))(N)
cols_brewer_2 <- colorRampPalette(brewer.pal(11, "Spectral"))(N)
par(mfrow = c(1, 2))
pie(rep(1, N), col = rainbow(N, start = 1/6, end = 1))
pie(rep(1, N), col = rainbow(N, start = 0, end = 1/6))
```

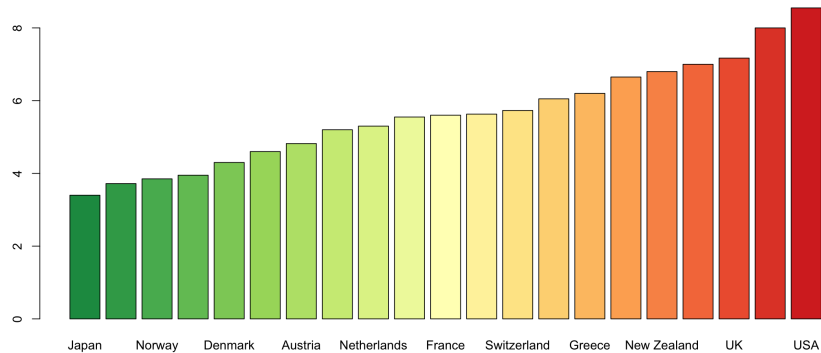
```
pie(rep(1, N), col = cols_brewer)
pie(rep(1, N), col = cols_brewer_2)
```



```
par(mfrow = c(1, 1))
```

## 색깔로 구분하기

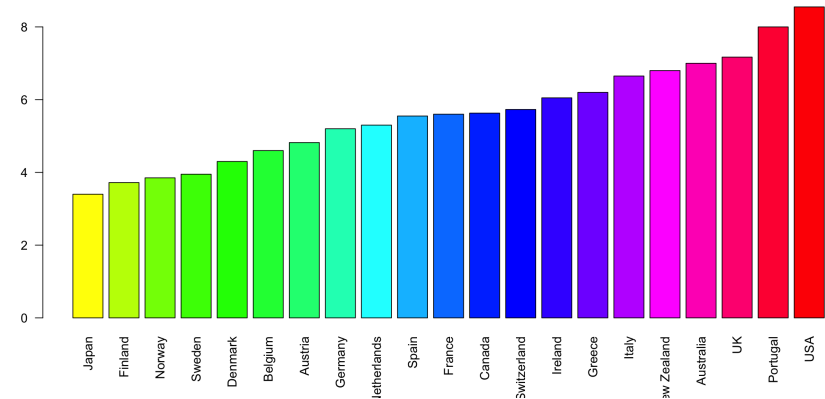
```
data_21 %>%
  .$Income_inequality %>%
  `[`(o_fifth) %>%
  barplot(names.arg = data_21$Country %>% `[`(o_fifth),
          col = cols_brewer %>% rev)
```



```
# col = rainbow(N, start = 1/6, end = 1))
```

## 나라 이름 눕히기

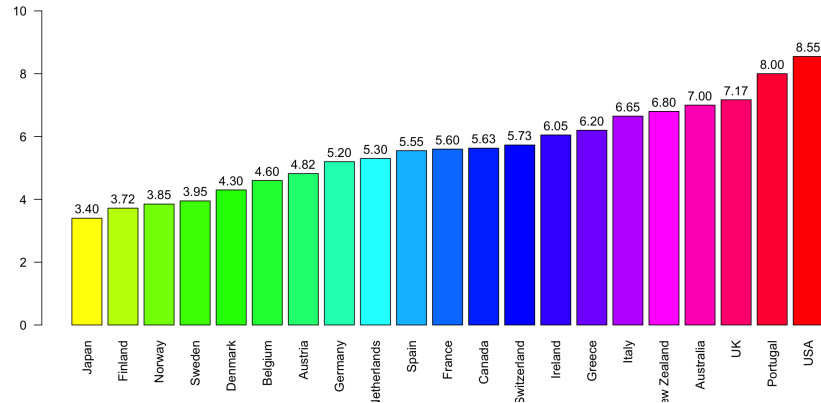
```
data_21 %>%
  .$Income_inequality %>%
  `[`(o_fifth) %>%
  barplot(names.arg = data_21$Country %>% `[`(o_fifth),
          col = rainbow(N, start = 1/6, end = 1),
          las = 2)
```



## 추가정보 텍스트와 그림 제목 붙이기

```
data_21 %>%
  .$Income_inequality %>%
  `[`(o_fifth) %>%
  barplot(names.arg = data_21$Country %>% `[`(o_fifth),
          col = rainbow(N, start = 1/6, end = 1),
          ylim = c(0, 10),
          las = 2)
text(x = b,
     y = data_21 %>%
       .$Income_inequality %>%
       `[`(o_fifth) %>%
       `+`(0.3),
     labels = data_21 %>%
       .$Income_inequality %>%
       `[`(o_fifth) %>%
       format(digits = 3))
title(main = "Fifth Ratios of Selected Countries", cex.main = 2)
```

Fifth Ratios of Selected Countries



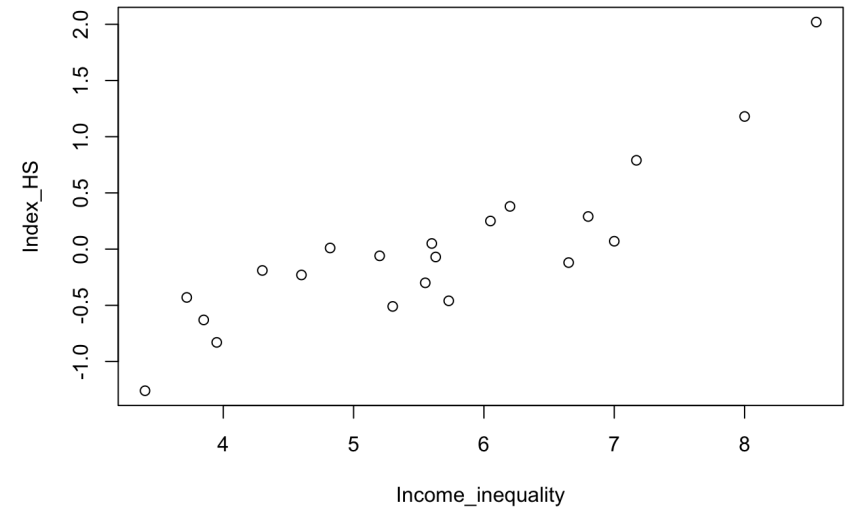
## Scatter Diagram

우선 소득불평등과 건강 및 사회문제 지표의 관계를 대략적으로 살펴보면,

```
Index_inequality_df <- data_21 %>% `(`(c("Income_inequality", "Index_HS"))
Index_inequality_df %>% str
```

```
## 'data.frame':    21 obs. of  2 variables:
## $ Income_inequality: num  7 4.82 4.6 5.63 4.3 3.72 5.6 5.2 6.2 6.05 ...
## $ Index_HS         : num  0.07 0.01 -0.23 -0.07 -0.19 -0.43 0.05 -0.06 0.38 0.25
## ...
```

```
Index_inequality_df %>% plot
```



```
Index_inequality_df %>% cor
```

```
##           Income_inequality  Index_HS
## Income_inequality      1.0000000 0.8735785
## Index_HS               0.8735785 1.0000000
```

```
cor_1 <- data_21 %>% `(`("Income_inequality") %>% cor(data_21 %>% `(`("Index_HS"))
cor_1
```

```
##           Index_HS
## Income_inequality 0.8735785
```

매우 높은 양의 상관관계( $r = 0.87$ )가 관찰됨을 알 수 있다. 자주 사용하는 `data_21[c("Income_inequality", "Index_HS")]`를 간단한 R 오브젝트로 assign하여 반복 사용하고 있다. `cor()` 에도 data frame을 사용하면 어떻게 되는지 다음 결과와 비교해 보자.

```
Index_inequality_df %>% cor
```

```
##           Income_inequality  Index_HS
## Income_inequality      1.0000000 0.8735785
## Index_HS               0.8735785 1.0000000
```

각 점이 어느 나라를 나타내는지 표시하기 위하여 `text()`를 활용하자. 동그라미 대신 까만 점으로 표시하고, 나라 이름을 올려보자. 우선 `[`와 `[[`의 용법에 대하여 정리하여 보자

```
(Country <- data_21 %>% `[.(, , "Country")])
```

```
## [1] "Australia" "Austria" "Belgium" "Canada" "Denmark"  
## [6] "Finland" "France" "Germany" "Greece" "Ireland"  
## [11] "Italy" "Japan" "Netherlands" "New Zealand" "Norway"  
## [16] "Portugal" "Spain" "Sweden" "Switzerland" "UK"  
## [21] "USA"
```

```
(Country_2 <- data_21 %>% `[("Country")])
```

```
## Country  
## 1 Australia  
## 2 Austria  
## 3 Belgium  
## 4 Canada  
## 5 Denmark  
## 6 Finland  
## 7 France  
## 8 Germany  
## 9 Greece  
## 10 Ireland  
## 12 Italy  
## 13 Japan  
## 14 Netherlands  
## 15 New Zealand  
## 16 Norway  
## 17 Portugal  
## 19 Spain  
## 20 Sweden  
## 21 Switzerland  
## 22 UK  
## 23 USA
```

```
(Country_3 <- data_21 %>% `[("Country") %>% .$Country)
```

```
## [1] "Australia" "Austria" "Belgium" "Canada" "Denmark"  
## [6] "Finland" "France" "Germany" "Greece" "Ireland"  
## [11] "Italy" "Japan" "Netherlands" "New Zealand" "Norway"  
## [16] "Portugal" "Spain" "Sweden" "Switzerland" "UK"  
## [21] "USA"
```

```
(Country_4 <- data_21 %>% `[(`("Country"))
```

```
## [1] "Australia" "Austria" "Belgium" "Canada" "Denmark"  
## [6] "Finland" "France" "Germany" "Greece" "Ireland"  
## [11] "Italy" "Japan" "Netherlands" "New Zealand" "Norway"  
## [16] "Portugal" "Spain" "Sweden" "Switzerland" "UK"  
## [21] "USA"
```

```
Country %>% str
```

```
## chr [1:21] "Australia" "Austria" "Belgium" "Canada" "Denmark" ...
```

```
Country_2 %>% str
```

```
## 'data.frame': 21 obs. of 1 variable:  
## $ Country: chr "Australia" "Austria" "Belgium" "Canada" ...
```

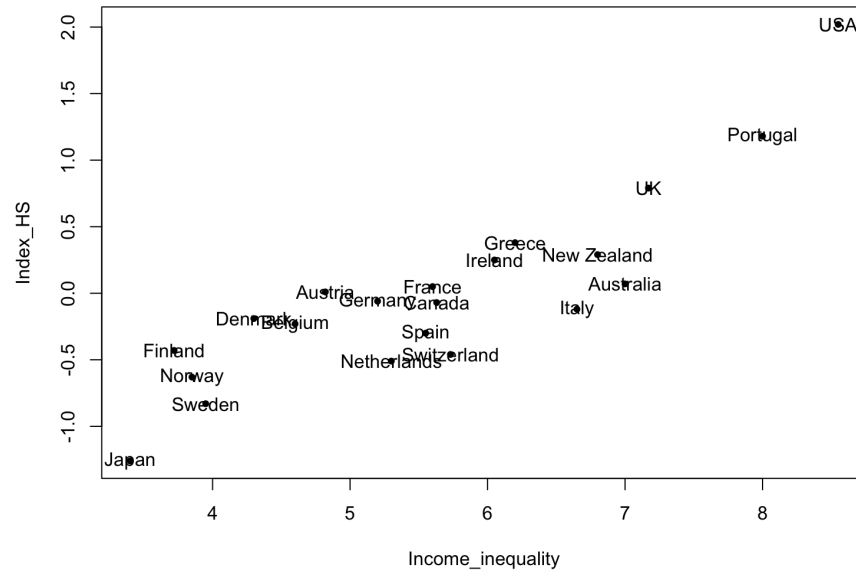
```
Country_3 %>% str
```

```
## chr [1:21] "Australia" "Austria" "Belgium" "Canada" "Denmark" ...
```

```
Country_4 %>% str
```

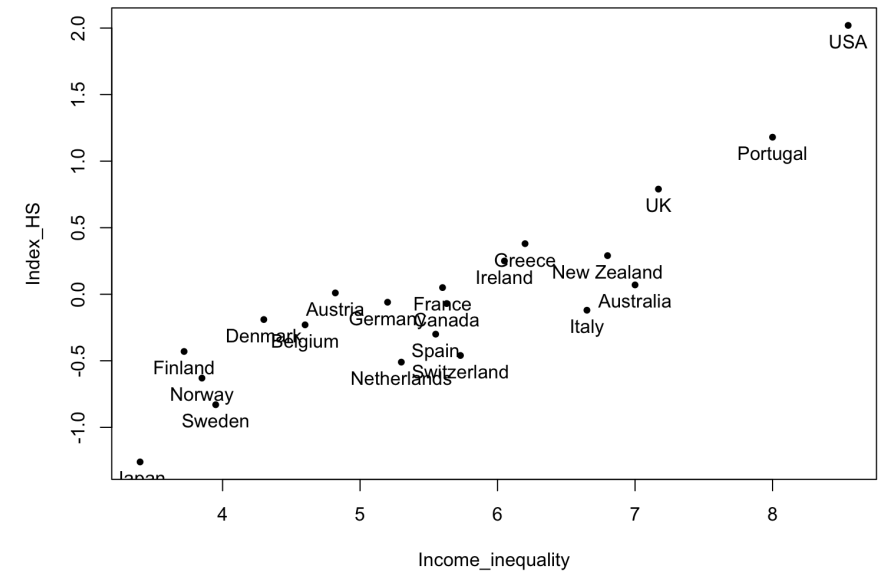
```
## chr [1:21] "Australia" "Austria" "Belgium" "Canada" "Denmark" ...
```

```
Index_inequality_df %>% plot(pch = 20)
Index_inequality_df %>% text(labels = Country)
```

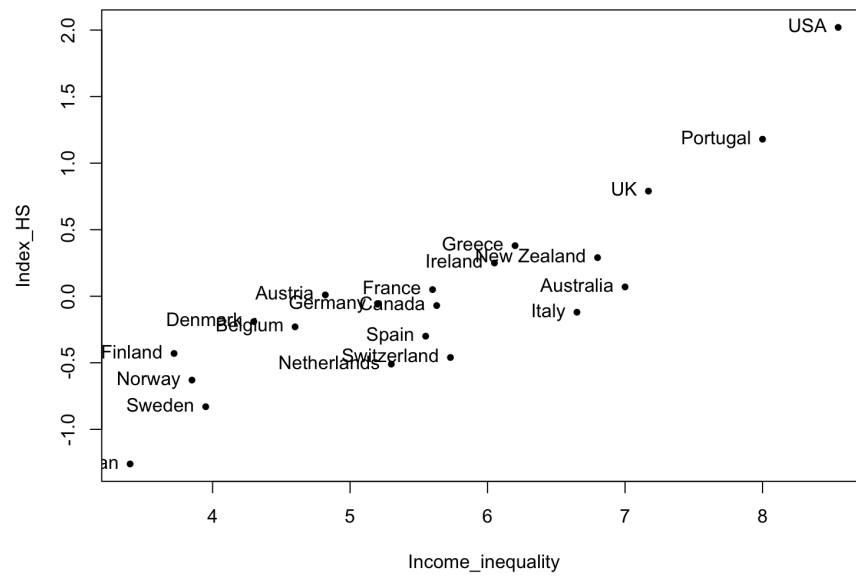


text label의 위치 기본값은 바로 점 위임을 알 수 있다. 위치 선정에 가능한 값들을 넣어보자.

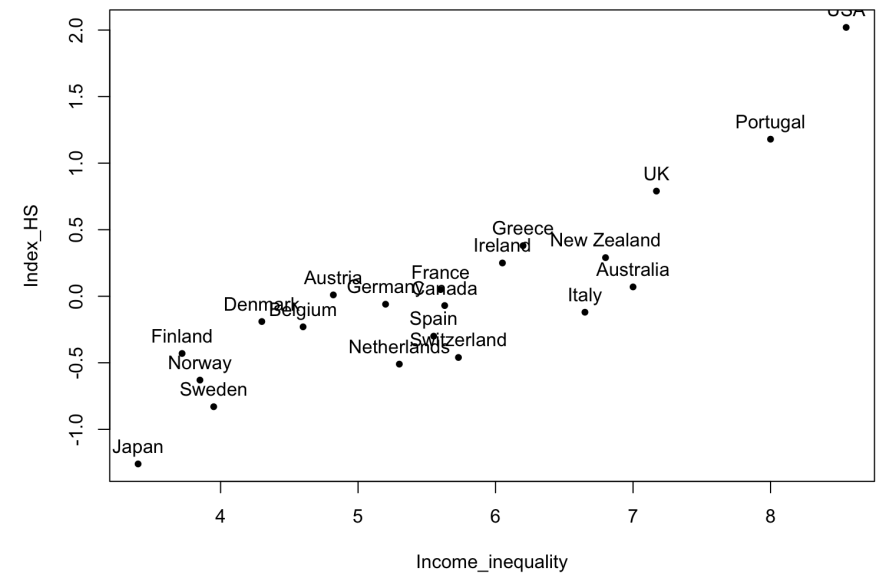
```
Index_inequality_df %>% plot(pch = 20)
Index_inequality_df %>% text(labels = Country, pos = 1)
```



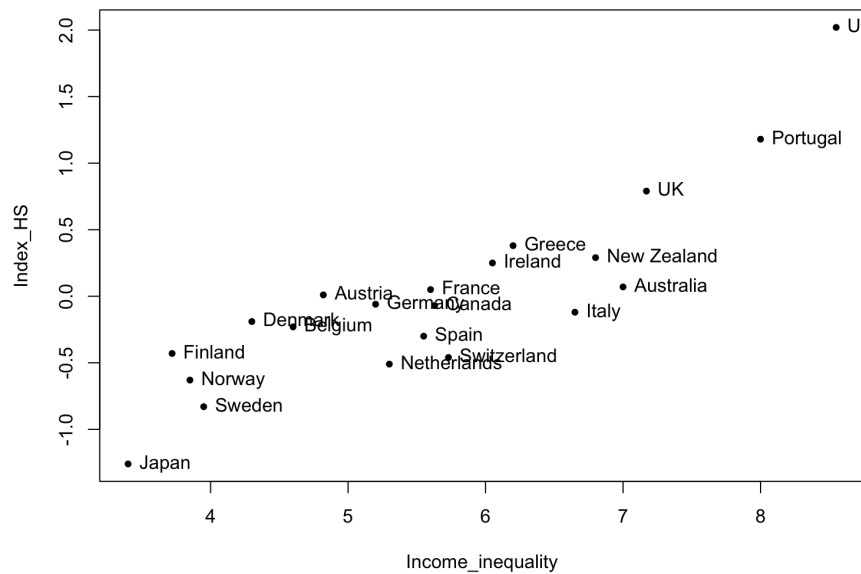
```
Index_inequality_df %>% plot(pch = 20)
Index_inequality_df %>% text(labels = Country, pos = 2)
```



```
Index_inequality_df %>% plot(pch = 20)
Index_inequality_df %>% text(labels = Country, pos = 3)
```

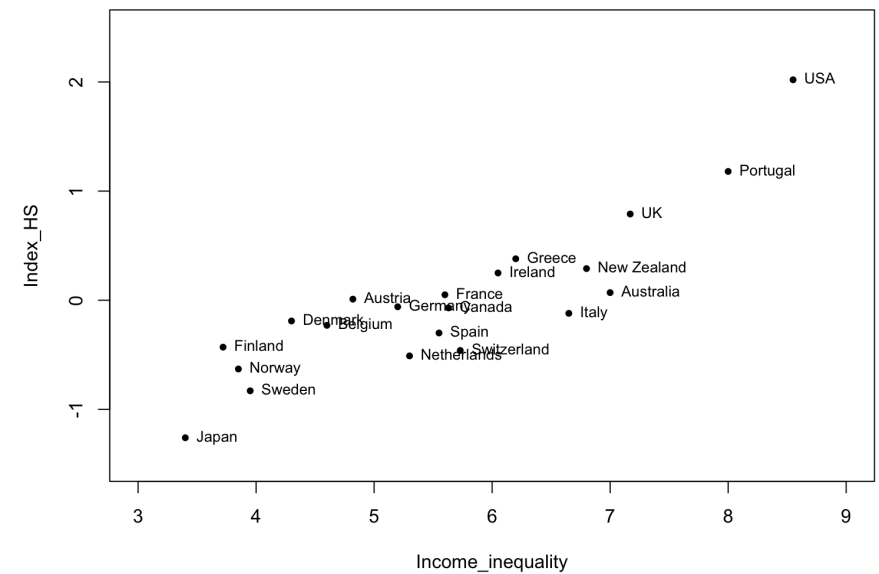


```
Index_inequality_df %>% plot(pch = 20)
Index_inequality_df %>% text(labels = Country, pos = 4)
```



우선 x-축과 y-축의 범위를 `xlim = c(3, 9)`, `ylim = c(-1.5, 2.5)` 로 하여 미국과 일본의 라벨이 도표 밖으로 나가 지 않게 하자. `pos = 4` 로 하고 `cex = 0.8` 로 하여 글자 크기를 줄여보면,

```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5))
Index_inequality_df %>% text(labels = Country, pos = 4, cex = 0.8)
```



오스트리아, 덴마크, 독일, 네덜란드의 라벨만 점 왼쪽에 위치시켜 보자. 각 인덱스를 찾아보면,

```
Country %in% c("Austria", "Denmark", "Germany", "Netherlands") %>% which
```

```
## [1] 2 5 8 13
```

```
text_left <- Country %in% c("Austria", "Denmark", "Germany", "Netherlands") %>% which
text_left
```

```
## [1] 2 5 8 13
```

```
text_right <- 1:nrow(data_21) %>% setdiff(text_left)
text_right
```

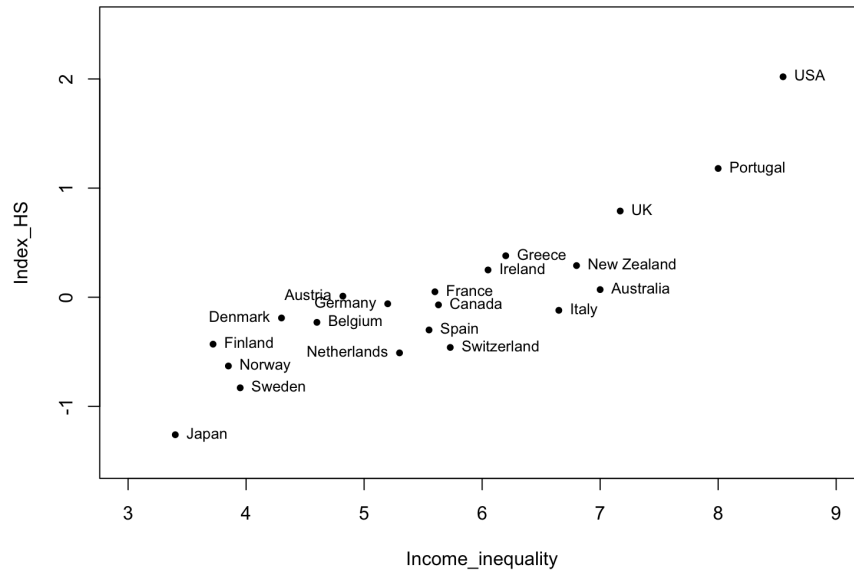
```
## [1] 1 3 4 6 7 9 10 11 12 14 15 16 17 18 19 20 21
```

```
pos_text <- 1:nrow(data_21) %in% text_left %>% ifelse(2, 4)
pos_text
```

```
## [1] 4 2 4 4 2 4 4 2 4 4 4 4 2 4 4 4 4 4 4 4 4
```



```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5))
Index_inequality_df %>% text(labels = Country, pos = pos_text, cex = 0.8)
```



독일의 라벨을 위로 붙이면 보기가 나아질 것으로 생각되므로,

```
Country %in% "Germany" %>% which
```

```
## [1] 8
```

```
text_up <- Country %in% "Germany" %>% which
text_up
```

```
## [1] 8
```

```
text_left <- 1:nrow(data_21) %>% setdiff(c(text_right, text_up))
text_left
```

```
## [1] 2 5 13
```

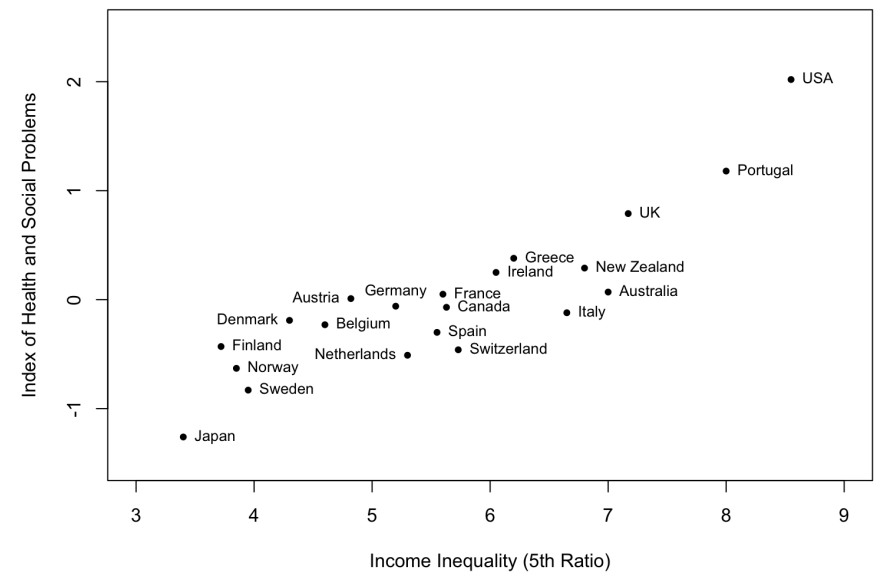
```
pos_text <- 1:nrow(data_21) %in% text_up %>%
  ifelse(3, 1:nrow(data_21) %in% text_left %>% ifelse(2, 4))
pos_text
```

```
## [1] 4 2 4 4 2 4 4 3 4 4 4 4 2 4 4 4 4 4 4 4
```

이제 조정된 text 외에 x-축과 y-축에 적절한 라벨과 메인 타이틀을 넣어보자.

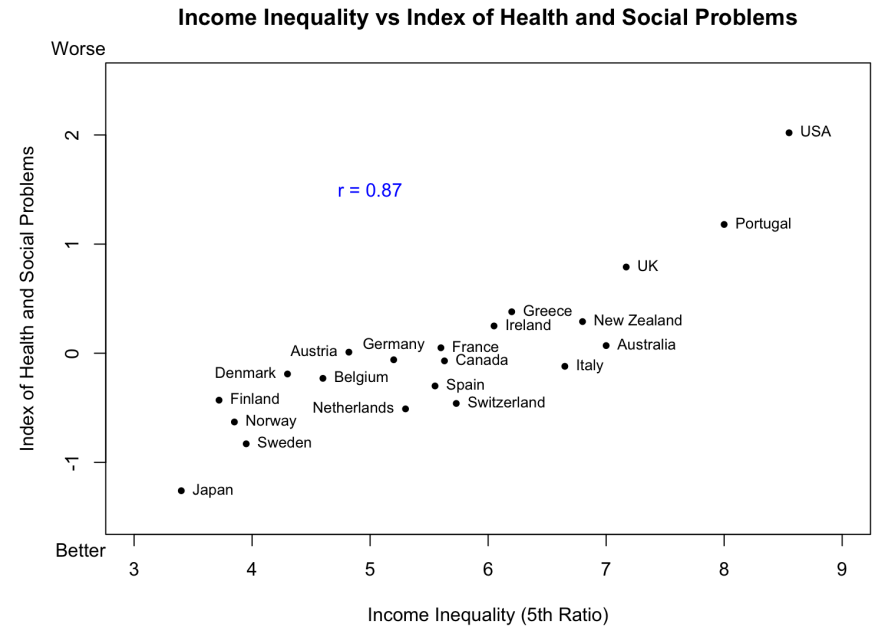
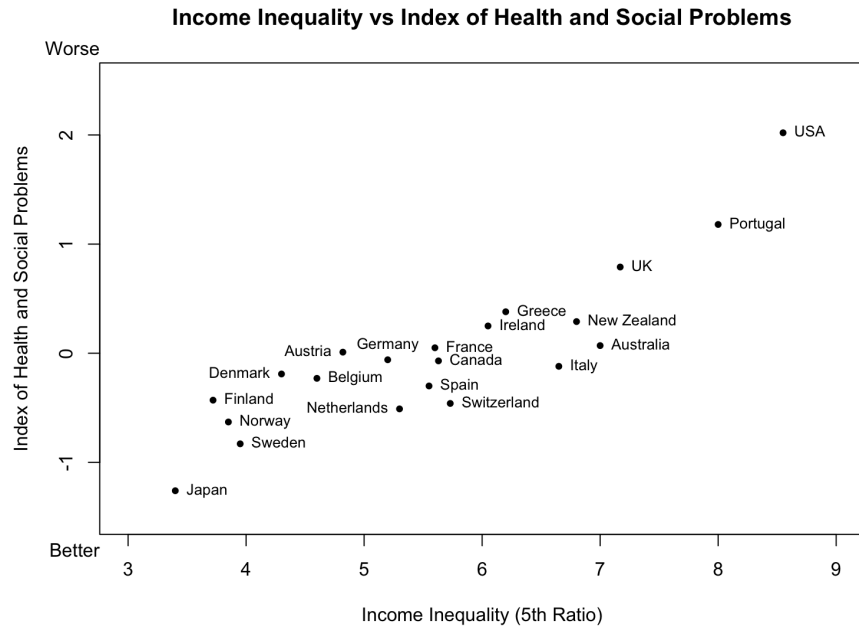
```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5), ann = FALSE)
Index_inequality_df %>% text(labels = Country, pos = pos_text, cex = 0.8)
main_title <- "Income Inequality vs Index of Health and Social Problems"
x_lab <- "Income Inequality (5th Ratio)"
y_lab <- "Index of Health and Social Problems"
title(main = main_title, xlab = x_lab, ylab = y_lab)
```

### Income Inequality vs Index of Health and Social Problems



건강 및 사회문제 지표의 경우 어느 방향이 좋은지 알 수 없으므로 친절하게 도표의 주변에(margin)에 알려주려면,

```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5), ann = FALSE)
Index_inequality_df %>% text(labels = Country, pos = pos_text, cex = 0.8)
main_title <- "Income Inequality vs Index of Health and Social Problems"
x_lab <- "Income Inequality (5th Ratio)"
y_lab <- "Index of Health and Social Problems"
title(main = main_title, xlab = x_lab, ylab = y_lab)
mtext(c("Better", "Worse"), side = 2, at = c(-1.8, 2.8), las = 1)
```



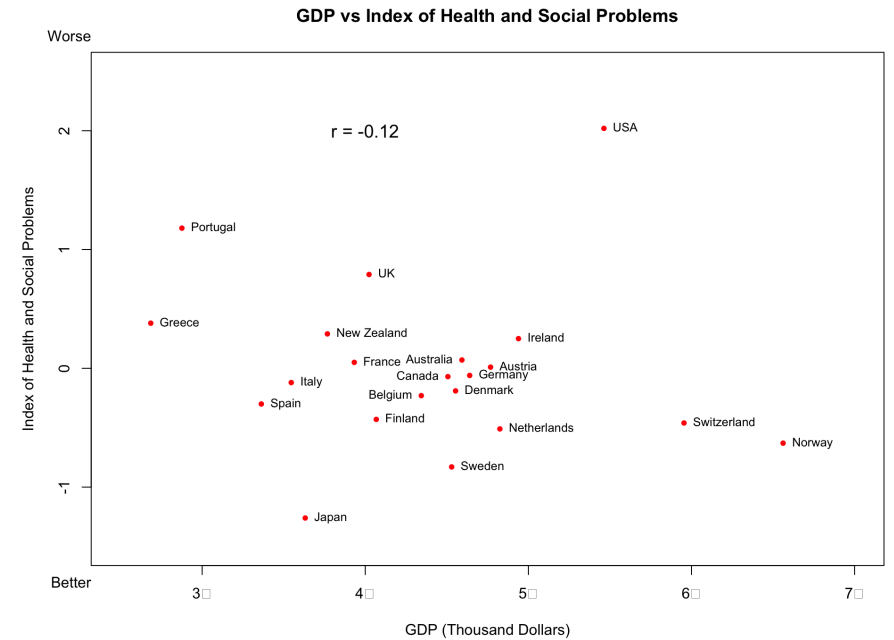
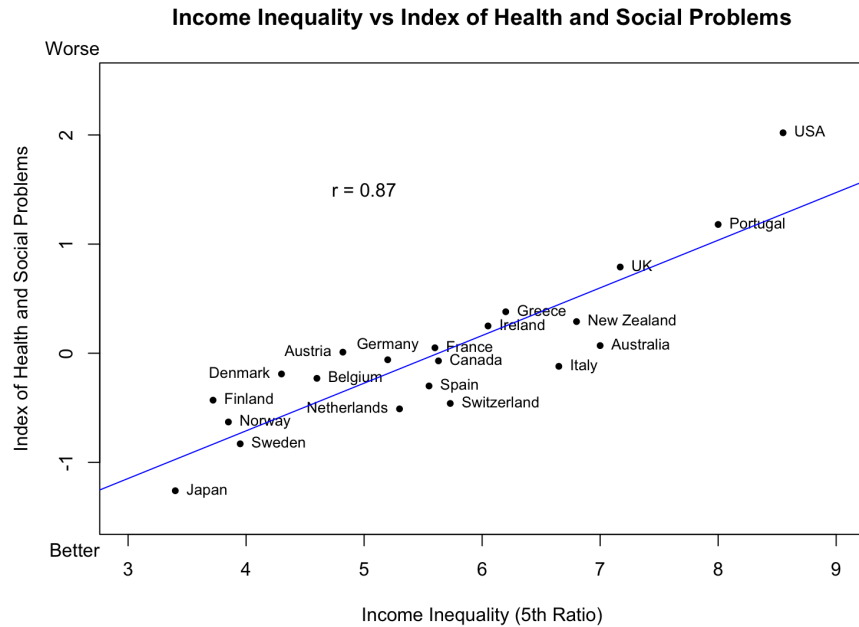
상관계수를 텍스트로 그림 안에 넣어주고 여기까지 작업한 내용을 별도의 파일로 저장해 놓으려면,

```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5), ann = FALSE)
Index_inequality_df %>% text(labels = Country, pos = pos_text, cex = 0.8)
main_title <- "Income Inequality vs Index of Health and Social Problems"
x_lab <- "Income Inequality (5th Ratio)"
y_lab <- "Index of Health and Social Problems"
title(main = main_title, xlab = x_lab, ylab = y_lab)
mtext(c("Better", "Worse"), side = 2, at = c(-1.8, 2.8), las = 1)
text(x = 5, y = 1.5, labels = paste("r =", cor_1 %>% round(digits = 2)), col = "blue")
)
```

```
# dev.copy(png, file = "../pics/inequality_health_social_en_72dpi.png", width = 640, height = 480)
# dev.off()
```

선형회귀선을 추가하여 대체적인 추세를 보려면 `lm()` 을 이용하되, `x`, `y` 의 순서를 제대로 바꿔야 함에 유의.

```
Index_inequality_df %>% plot(pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5), ann = FALSE)
Index_inequality_df %>% text(labels = Country, pos = pos_text, cex = 0.8)
main_title <- "Income Inequality vs Index of Health and Social Problems"
x_lab <- "Income Inequality (5th Ratio)"
y_lab <- "Index of Health and Social Problems"
title(main = main_title, xlab = x_lab, ylab = y_lab)
mtext(c("Better", "Worse"), side = 2, at = c(-1.8, 2.8), las = 1)
text(x = 5, y = 1.5, labels = paste("r =", cor_1 %>% round(digits = 2)))
# lm_ineq <- lm(Index_inequality_df[2:1])
lm_ineq <- lm(Index_HS ~ Income_inequality, data = Index_inequality_df)
lm_ineq$coef %>% abline(col = "blue")
```



GDP와 건강 및 사회문제 지수

```
Index_GDP_df <- data_21[c("GDP_WB", "Index_HS")]
text_left_2 <- which(Country %in% c("Canada", "Belgium", "Australia"))
text_right_2 <- setdiff(1:nrow(data_21), c(text_left_2))
pos_text_2 <- ifelse(1:nrow(data_21) %in% text_left_2, 2, 4)
plot(Index_GDP_df, pch = 20, col = "red", xlim = c(25000, 70000), ylim = c(-1.5, 2.5),
     xaxt = "n", ann = FALSE)
axis(side = 1, at = seq(30000, 70000, by = 10000), labels = paste(3:7, "만", sep = ""))
text(Index_GDP_df, labels = Country, pos = pos_text_2, cex = 0.8)
cor_2 <- cor(Index_GDP_df["GDP_WB"], Index_GDP_df["Index_HS"])
text(x = 40000, y = 2, labels = paste("r =", round(cor_2, digits = 2)), cex = 1.2)
main_title_2 <- "GDP vs Index of Health and Social Problems"
x_lab_2 <- "GDP (Thousand Dollars)"
y_lab_2 <- "Index of Health and Social Problems"
title(main = main_title_2, xlab = x_lab_2, ylab = y_lab_2)
mtext(c("Better", "Worse"), side = 2, at = c(-1.8, 2.8), las = 1)
```

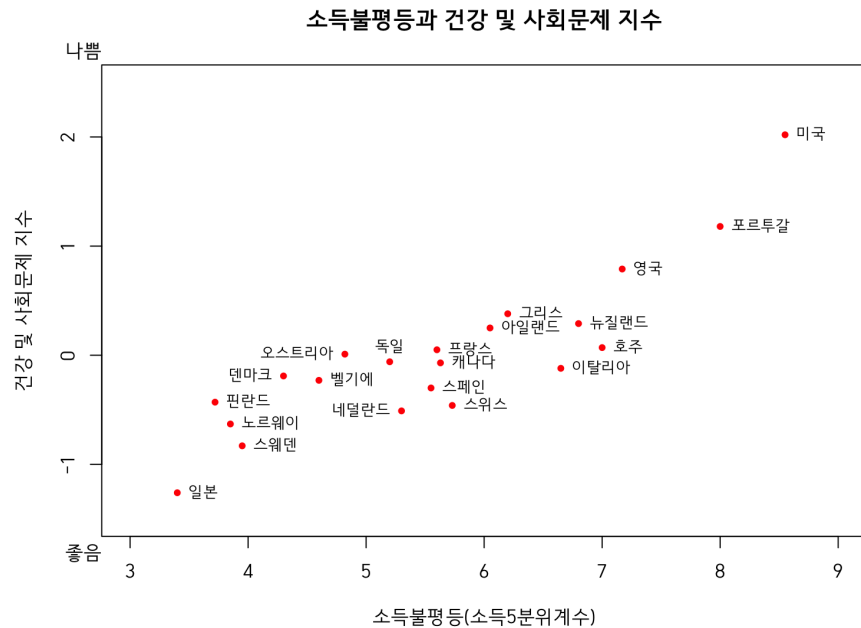
```
# dev.copy(png, file = "../pics/GDP_health_social_en_72dpi.png", width = 640, height = 480)
# dev.off()
```

## 한글화

국가명을 한글로 만들어 country\_kr 로 저장하자.

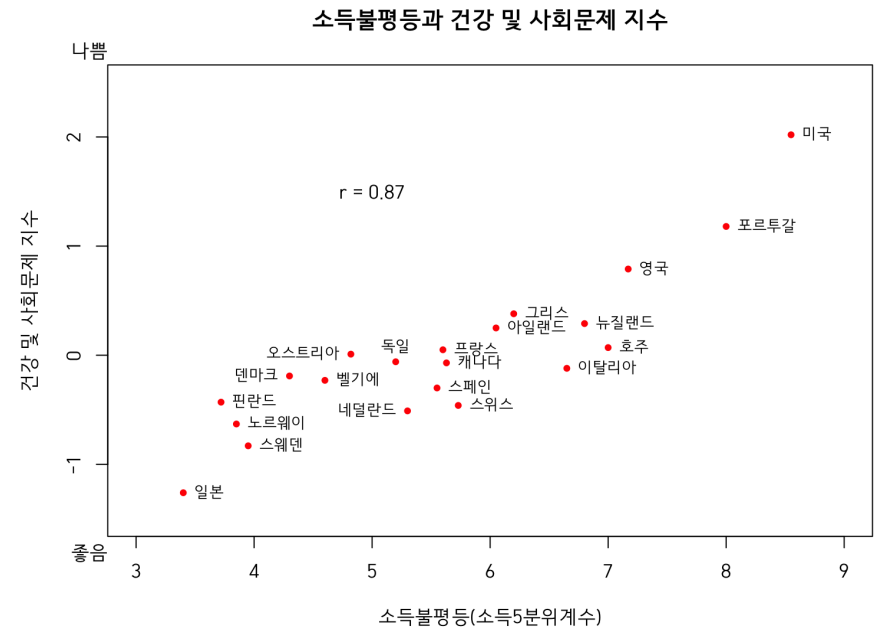
```
Country_kr <- c("호주", "오스트리아", "벨기에", "캐나다", "덴마크",
  "핀란드", "프랑스", "독일", "그리스", "아일랜드", "이탈리아",
  "일본", "네덜란드", "뉴질랜드", "노르웨이", "포르투갈",
  "스페인", "스웨덴", "스위스", "영국", "미국")
```

```
# library(extrafont)
par(family = "HCR Dotum LVT")
plot(Index_inequality_df, pch = 20, col = "red", xlim = c(3, 9), ylim = c(-1.5, 2.5),
     ann = FALSE)
text(Index_inequality_df[text_right, ], labels = Country_kr[text_right], pos = 4, cex = 0.8)
text(Index_inequality_df[text_left, ], labels = Country_kr[text_left], pos = 2, cex = 0.8)
text(Index_inequality_df[text_up, ], labels = Country_kr[text_up], pos = 3, cex = 0.8)
)
main_title_kr <- "소득불평등과 건강 및 사회문제 지수"
x_lab_kr <- "소득불평등(소득5분위계수)"
y_lab_kr <- "건강 및 사회문제 지수"
title(main = main_title_kr, xlab = x_lab_kr, ylab = y_lab_kr)
mtext(c("좋은", "나쁨"), side = 2, at = c(-1.8, 2.8), las = 1)
```



상관계수  $r = 0.87$  를 도표 안에 표시하고 별도의 파일로 출력하려면,

```
par(family = "HCR Dotum LVT")
plot(Index_inequality_df, pch = 20, col = "red", xlim = c(3, 9), ylim = c(-1.5, 2.5),
     ann = FALSE)
text(Index_inequality_df, labels = Country_kr, pos = pos_text, cex = 0.8)
main_title_kr <- "소득불평등과 건강 및 사회문제 지수"
x_lab_kr <- "소득불평등(소득5분위계수)"
y_lab_kr <- "건강 및 사회문제 지수"
title(main = main_title_kr, xlab = x_lab_kr, ylab = y_lab_kr)
mtext(c("좋은", "나쁨"), side = 2, at = c(-1.8, 2.8), las = 1)
text(x = 5, y = 1.5, labels = paste("r =", round(corr(Index_inequality_df[,1], Index_in
equality_df[,2]), digits = 2)))
```



```
# dev.copy(png, file = "../pics/inequality_health_social_72dpi.png", width = 640, hei
ght = 480)
# dev.off()
```

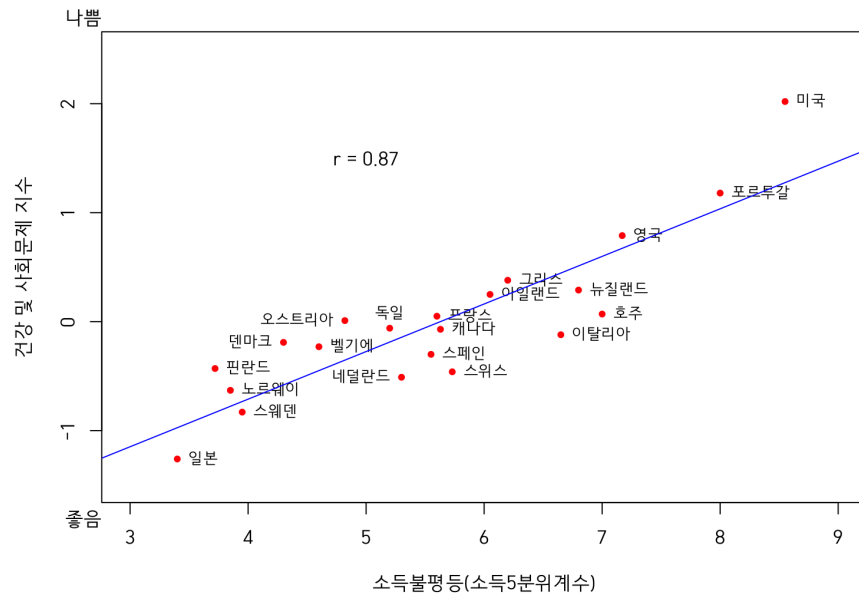
선형회귀선을 이번에는 `lsfit` 을 이용하여 삽입

```

par(family = "HCR Dotum LVT")
plot(Index_inequality_df, pch = 20, col = "red", xlim = c(3, 9), ylim = c(-1.5, 2.5),
     ann = FALSE)
text(Index_inequality_df, labels = Country_kr, pos = pos_text, cex = 0.8)
main_title_kr <- "소득불평등과 건강 및 사회문제 지수"
x_lab_kr <- "소득불평등(소득5분위계수)"
y_lab_kr <- "건강 및 사회문제 지수"
title(main = main_title_kr, xlab = x_lab_kr, ylab = y_lab_kr)
mtext(c("좋은", "나쁨"), side = 2, at = c(-1.8, 2.8), las = 1)
text(x = 5, y = 1.5, labels = paste("r =", round(cor(Index_inequality_df[1], Index_in
equality_df[2]), digits = 2)))
lsfit_ineq <- lsfit(x = Index_inequality_df[, 1], y = Index_inequality_df[, 2])
abline(lsfit_ineq$coefficients, col = "blue")

```

소득불평등과 건강 및 사회문제 지수



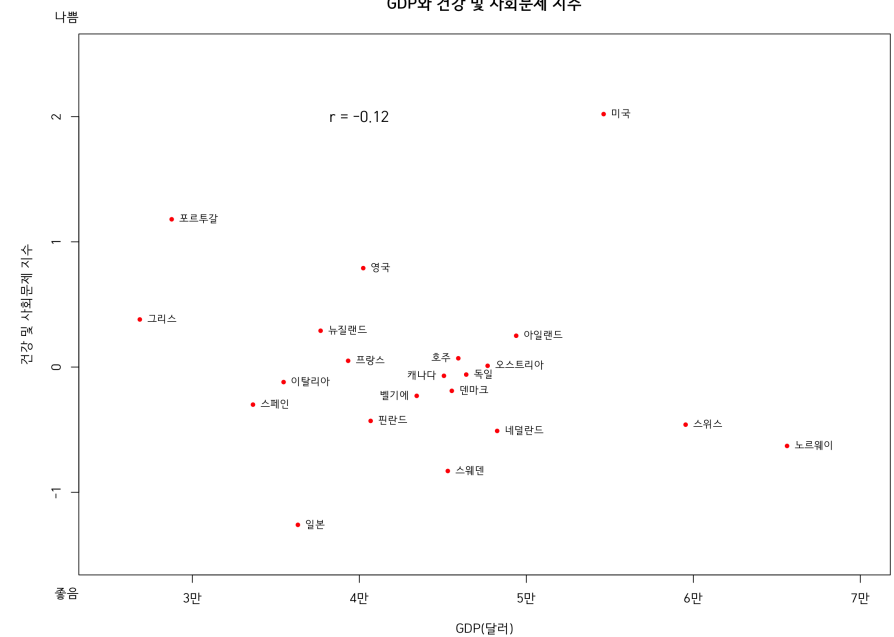
GDP와의 관계

```

par(family = "HCR Dotum LVT")
Index_GDP_df <- data_21[c("GDP_WB", "Index_HS")]
text_left_2 <- which(Country %in% c("Canada", "Belgium", "Australia"))
text_right_2 <- setdiff(1:nrow(data_21), c(text_left_2))
pos_text_2 <- ifelse(1:nrow(data_21) %in% text_left_2, 2, 4)
plot(Index_GDP_df, pch = 20, col = "red", xlim = c(25000, 70000), ylim = c(-1.5, 2.5),
     xaxt = "n", ann = FALSE)
axis(side = 1, at = seq(30000, 70000, by = 10000), labels = paste(3:7, "만", sep = ""))
text(Index_GDP_df, labels = Country_kr, pos = pos_text_2, cex = 0.8)
text(x = 40000, y = 2, labels = paste("r =", round(cor(Index_GDP_df[1], Index_GDP_df[
2]), digits = 2)), cex = 1.2)
main_title_2_kr <- "GDP와 건강 및 사회문제 지수"
x_lab_2_kr <- "GDP(달러)"
y_lab_2_kr <- "건강 및 사회문제 지수"
title(main = main_title_2_kr, xlab = x_lab_2_kr, ylab = y_lab_2_kr)
mtext(c("좋은", "나쁨"), side = 2, at = c(-1.8, 2.8), las = 1)

```

GDP와 건강 및 사회문제 지수



```

# dev.copy(png, file = "../pics/GDP_health_social_72dpi.png", width = 640, height = 480)
# dev.off()

```