

Income Inequality vs Index of Health and Social Problems

coop711

2016-05-10

Data Preparation

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 추가하여 읽어들이면,

```
library(knitr)
load("Inequality_Index_HS.rda")
# data.full <- read.csv("../data/international-inequality_GDP.csv", stringsAsFactors = FALSE)
str(data.full)
```

```
## 'data.frame': 23 obs. of 30 variables:
## $ Country : chr "Australia" "Austria" "Belgium"
## $ Income.inequality : num 7 4.82 4.6 5.63 4.3 3.72 5.6 5.2
## $ Trust : num 39.9 33.9 30.7 38.8 66.5 58 22.2
## $ Life.expectancy : num 79.2 78.5 78.8 79.3 76.6 78 79 7
## $ Infant.mortality : num 4.9 4.8 5 5.3 5.3 3.7 4.4 4.4 5
## $ Obesity : num 18.4 14.5 13.5 12.8 15 ...
## $ Mental.illness : num 23 NA 12 19.9 NA NA 18.4 9.1 NA NA
## $ Maths.and.literacy.scores : num 524 498 518 530 503 ...
## $ Teenage.births : num 18.4 14 9.9 20.2 8.1 9.2 9.3 13.1
## $ Homicides : num 16.9 11.6 13 17.3 12.7 28.2 21.5
## $ Imprisonment..log. : num 4.61 4.52 4.28 4.77 4.17 4.11 4.5
## $ Social.mobility : num NA NA NA 0.14 0.14 0.15 NA 0.17 NA
## $ Index.of.health...social_problems : num 0.07 0.01 -0.23 -0.07 -0.19 -0.43
## $ Child.overweight : num NA 11.9 10.4 19.5 10.3 13.3 11.2
## $ Drugs.index : num 1.71 -0.02 -0.18 0.61 -0.09 -0.88
## $ Calorie.intake : int 3142 3753 3632 3167 3405 3197 357
## $ Public.health.expenditure : num 67.9 69.3 71.7 70.8 82.4 75.6 76
## $ Child.wellbeing : num -0.21 -0.07 0.05 0.04 0.21 0.34 -
## $ Maths.education.science.score : num 525 496 515 526 494 ...
## $ Child.conflict : num NA 0.31 0.33 0.24 -0.14 -1.25 0.5
## $ Foreign.aid : num 0.25 0.52 0.53 0.34 0.81 0.47 0.4
## $ Recycling : num 7.4 NA NA NA NA NA 6 3.4 NA NA
## $ Peace.index : num 1.66 1.48 1.49 1.48 1.38 1.45 1.7
## $ Maternity.leave : int 0 16 15 17 18 18 16 14 17 18 ...
## $ Advertising : num 1.24 0.97 0.82 0.77 0.75 0.9 0.71
## $ Police : int 304 305 357 186 192 160 NA 303 NA
## $ Social.expenditure : num 17.8 27.5 26.5 17.2 27.6 25.8 29
## $ Women.s_status : num 0.46 -0.81 0.61 0.56 0.83 1.08 -
```

```
## $ Lone.parents      : int  21 15 12 17 22 19 12 21 3 14 ...
## $ GDP_WB            : int  45926 47682 43435 45066 45537 406
76 39328 46401 26851 49393 ...
```

이 자료 중 소득불평등을 나타내는 지표는 5분위계수로서 두번째 컬럼에 `Income.inequality` 라는 이름으로 나와 있고, 건강과 사회문제 지표는 13번째 컬럼에 `Index.of.health...social_problems` 라는 이름으로 주어져 있다. 나라들은 `Country` 라는 변수명으로 첫번째 컬럼에 나와 있다. 그리고, 건강과 사회문제 지표에 결측치들이 있기 때문에 먼저 이 나라들을 제외하고 분석작업을 수행하여야 한다. `which()` 를 이용하여 해당 인덱스를 찾고, 나라명을 추출한다.

```
(country.na <- which(is.na(data.full$Index.of.health...social_problems)))
```

```
## [1] 11 18
```

```
data.full$Country[country.na]
```

```
## [1] "Israel"      "Singapore"
```

결측치가 있는 나라를 빼고, 필요한 변수만 챙겨서 새로운 `data frame` 을 구성하기 위하여 건강과 사회문제 지표의 위치를 찾아보자.

```
names(data.full)
```

```
## [1] "Country"
## [2] "Income.inequality"
## [3] "Trust"
## [4] "Life.expectancy"
## [5] "Infant.mortality"
## [6] "Obesity"
## [7] "Mental.illness"
## [8] "Maths.and.literacy.scores"
## [9] "Teenage.births"
## [10] "Homicides"
## [11] "Imprisonment..log."
## [12] "Social.mobility"
## [13] "Index.of.health...social_problems"
## [14] "Child.overweight"
## [15] "Drugs.index"
## [16] "Calorie.intake"
## [17] "Public.health.expenditure"
## [18] "Child.wellbeing"
## [19] "Maths.education.science.score"
## [20] "Child.conflict"
## [21] "Foreign.aid"
## [22] "Recycling"
## [23] "Peace.index"
## [24] "Maternity.leave"
## [25] "Advertising"
## [26] "Police"
## [27] "Social.expenditure"
## [28] "Women.s_status"
## [29] "Lone.parents"
## [30] "GDP_WB"
```

```
which(names(data.full) == "Index.of.health...social_problems")
```

```
## [1] 13
```

새로운 `data frame` 을 `data.21` 으로 저장하자. 시각적 가독성을 높이기 위하여 자릿수를 조정한다.

```
options(digits = 2)
v.names <- c("Country", "Income.inequality", "Index.of.health...social_problem
s", "GDP_WB")
data.21 <- data.full[-c(11, 18), v.names]
names(data.21)[3] <- "Index.HS"
kable(data.21)
```

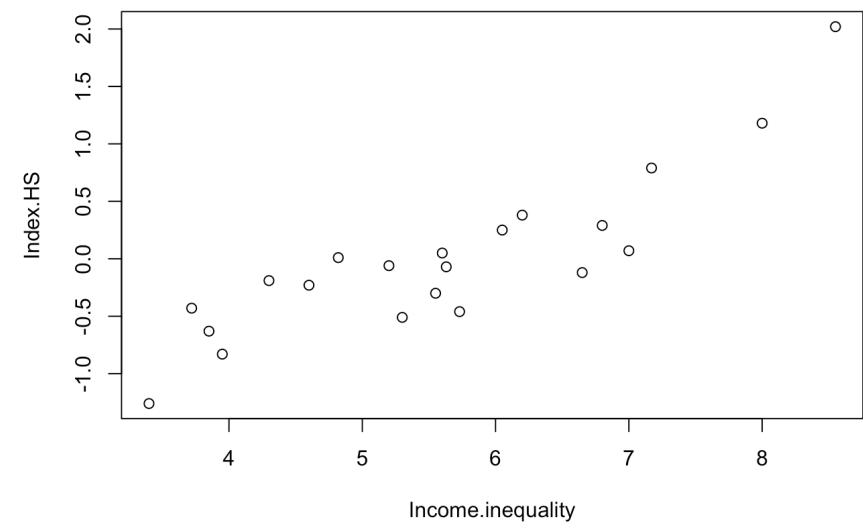
	Country	Income.inequality	Index.HS	GDP_WB
1	Australia	7.0	0.07	45926
2	Austria	4.8	0.01	47682
3	Belgium	4.6	-0.23	43435

4	Canada	5.6	-0.07	45066
5	Denmark	4.3	-0.19	45537
6	Finland	3.7	-0.43	40676
7	France	5.6	0.05	39328
8	Germany	5.2	-0.06	46401
9	Greece	6.2	0.38	26851
10	Ireland	6.0	0.25	49393
12	Italy	6.7	-0.12	35463
13	Japan	3.4	-1.26	36319
14	Netherlands	5.3	-0.51	48253
15	New Zealand	6.8	0.29	37679
16	Norway	3.9	-0.63	65615
17	Portugal	8.0	1.18	28760
19	Spain	5.5	-0.30	33629
20	Sweden	4.0	-0.83	45297
21	Switzerland	5.7	-0.46	59540
22	UK	7.2	0.79	40233
23	USA	8.6	2.02	54630

Plots

우선 소득불평등과 건강 및 사회문제 지표의 관계를 대략적으로 살펴보면,

```
plot(data.21[c("Income.inequality", "Index.HS")])
```



매우 높은 양의 상관관계($r = 0.87$)가 관찰됨을 알 수 있다.

각 점이 어느 나라를 나타내는지 표시하기 위하여 `text()`를 활용하자. 동그라미 대신 까만 점으로 표시하고, 나라 이름을 올려보자. 자주 사용하는 `data.21[c("Income.inequality", "Index.HS")]`를 간단한 R 오브젝트로 assign하여 반복 사용하자.

```
Index_inequality.df <- data.21[c("Income.inequality", "Index.HS")]
str(Index_inequality.df)
```

```
## '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 ...
```

```
(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"
```

```
str(Country)
```

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

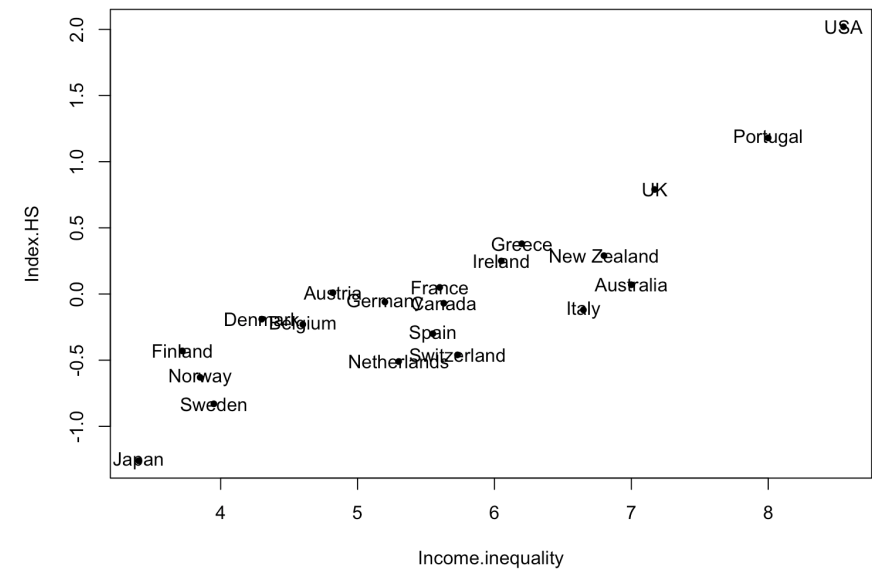
```
str(Country.2)
```

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

```
str(Country.3)
```

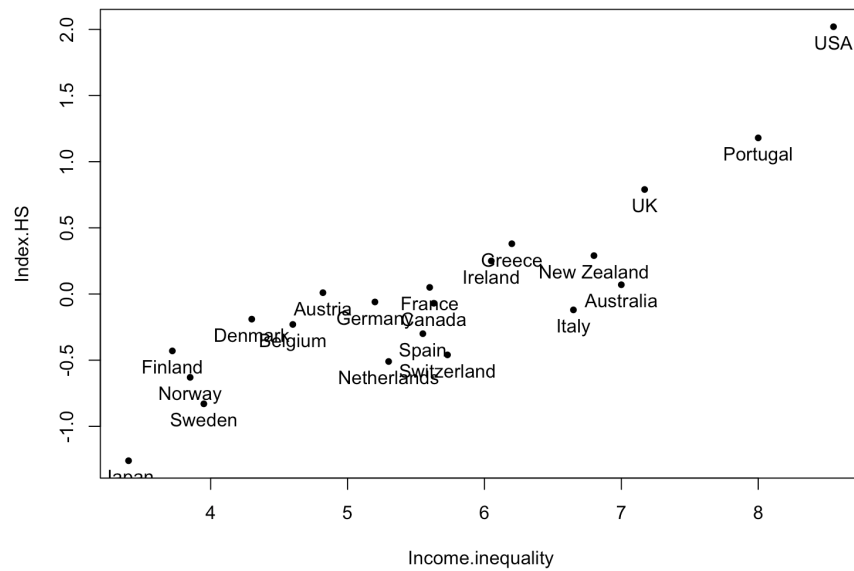
```
## chr [1:21] "Australia" "Austria" "Belgium" "Canada" ...
```

```
plot(Index_inequality.df, pch = 20)
text(Index_inequality.df, labels = Country)
```

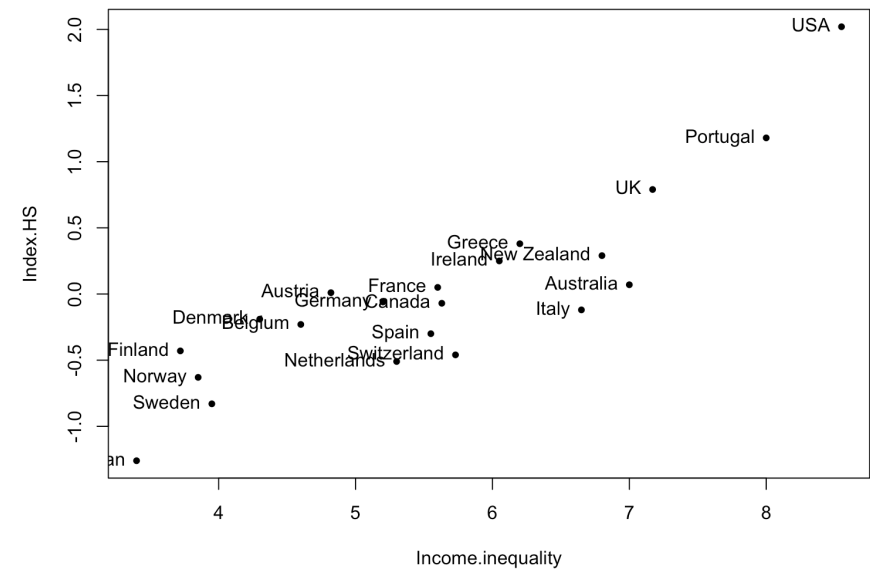


text label의 위치 기본값은 바로 점 위임을 알 수 있다. 위치 선정에 가능한 값들을 넣어보자.

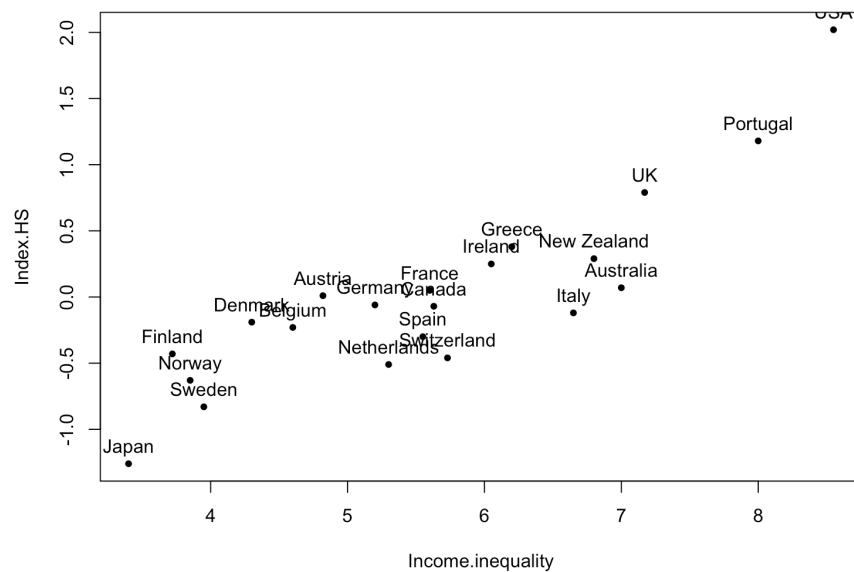
```
plot(Index_inequality.df, pch = 20)
text(Index_inequality.df, labels = Country, pos = 1)
```



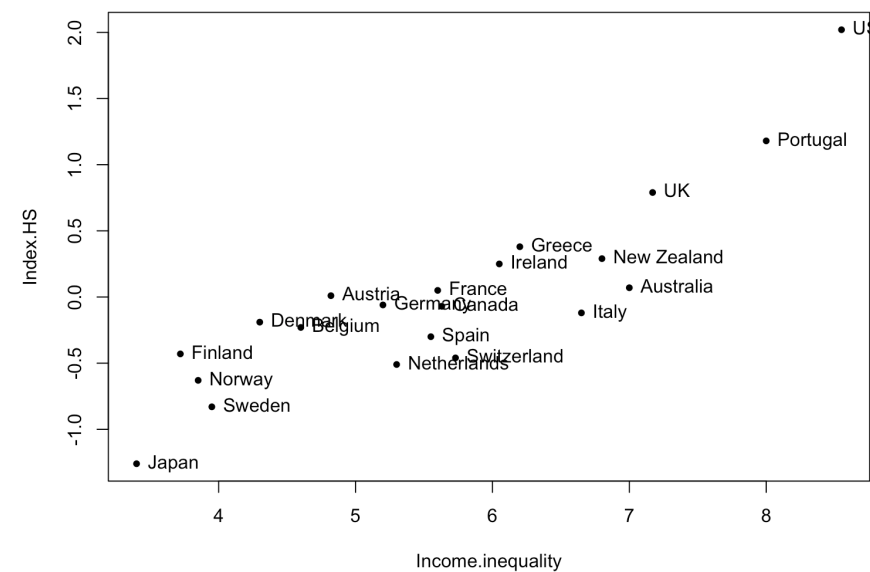
```
plot(Index_inequality.df, pch = 20)
text(Index_inequality.df, labels = Country, pos = 2)
```



```
plot(Index_inequality.df, pch = 20)
text(Index_inequality.df, labels = Country, pos = 3)
```

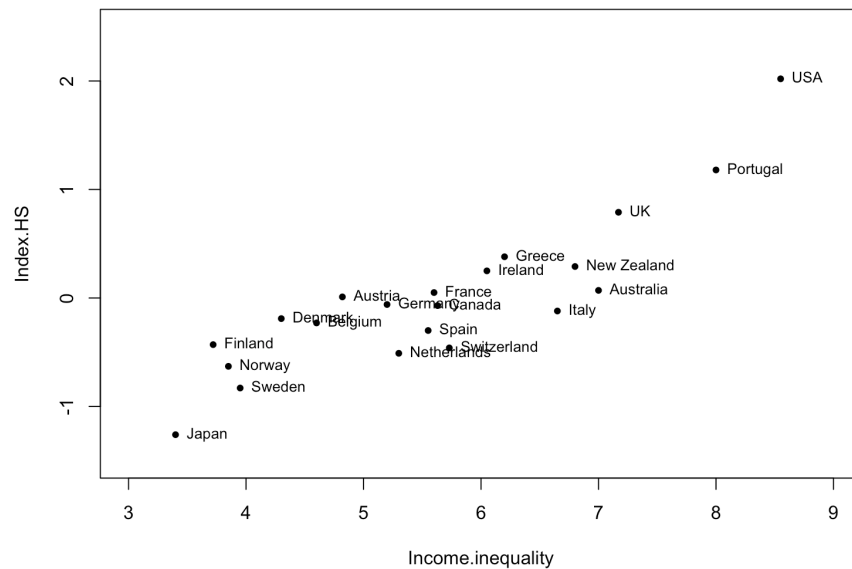


```
plot(Index_inequality.df, pch = 20)
text(Index_inequality.df, labels = Country, pos = 4)
```



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

```
plot(Index_inequality.df, pch = 20, xlim = c(3, 9), ylim = c(-1.5, 2.5))
text(Index_inequality.df, labels = Country, pos = 4, cex = 0.8)
```



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

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

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

```
text.left <- which(Country %in% c("Austria", "Denmark", "Germany", "Netherlands"))
text.left
```

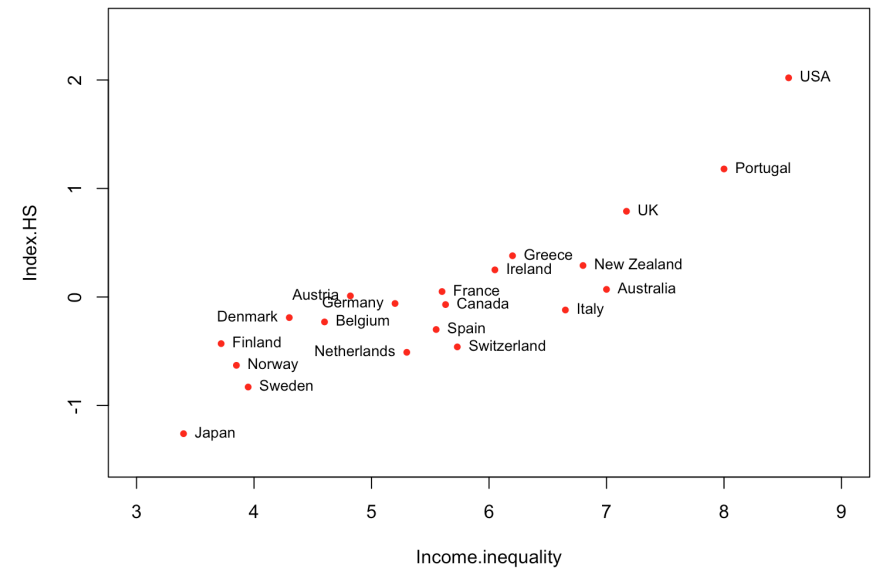
```
## [1] 2 5 8 13
```

```
text.right <- setdiff(1:nrow(data.21), text.left)
text.right
```

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

```
pos.text <- ifelse(1:nrow(data.21) %in% text.left, 2, 4)
```

```
plot(Index_inequality.df, pch = 20, col = "red", xlim = c(3, 9), ylim = c(-1.5, 2.5))
text(Index_inequality.df, labels = Country, pos = pos.text, cex = 0.8)
```



독일의 라벨을 위로 붙이면 보기가 나아질 것으로 생각되므로,

```
which(Country %in% "Germany")
```

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

```
text.up <- which(Country %in% "Germany")
text.up
```

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

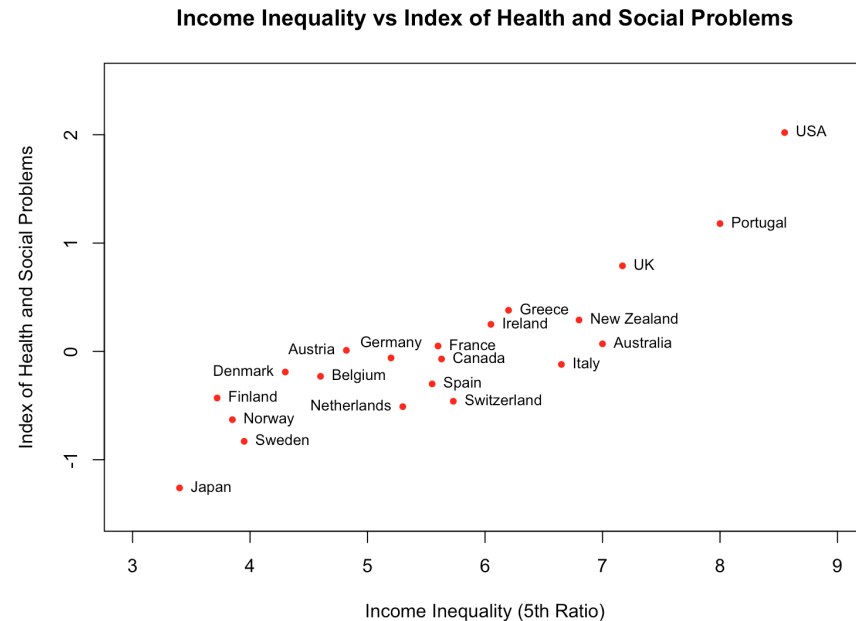
```
text.left <- setdiff(1:nrow(data.21), c(text.right, text.up))
text.left
```

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

```
pos.text <- ifelse(1:nrow(data.21) %in% text.up, 3, ifelse(1:nrow(data.21) %in%
text.left, 2, 4))
```

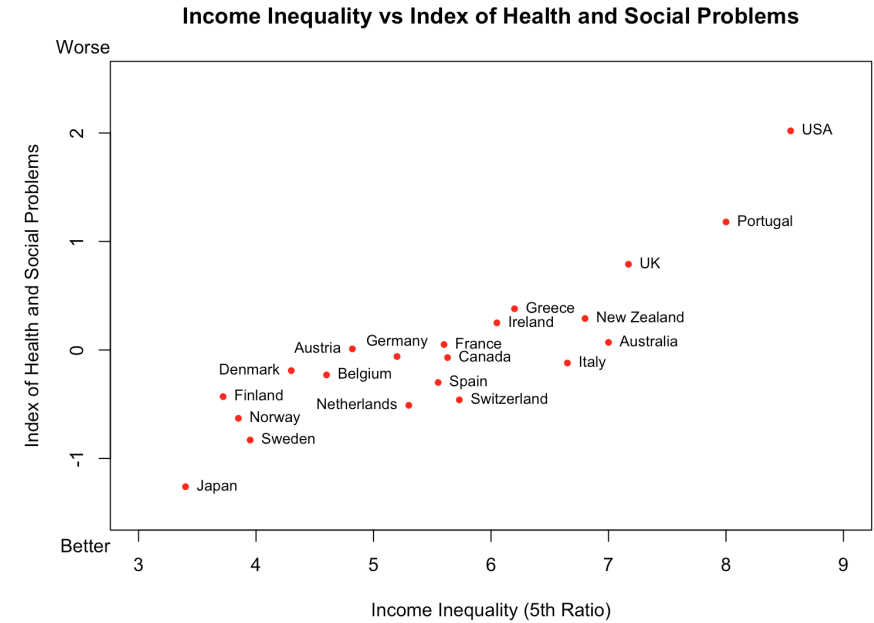
이제 조정된 text 외에 x-축과 y-축에 적절한 라벨과 메인 타이틀을 넣어보자.

```
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, 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)
```



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

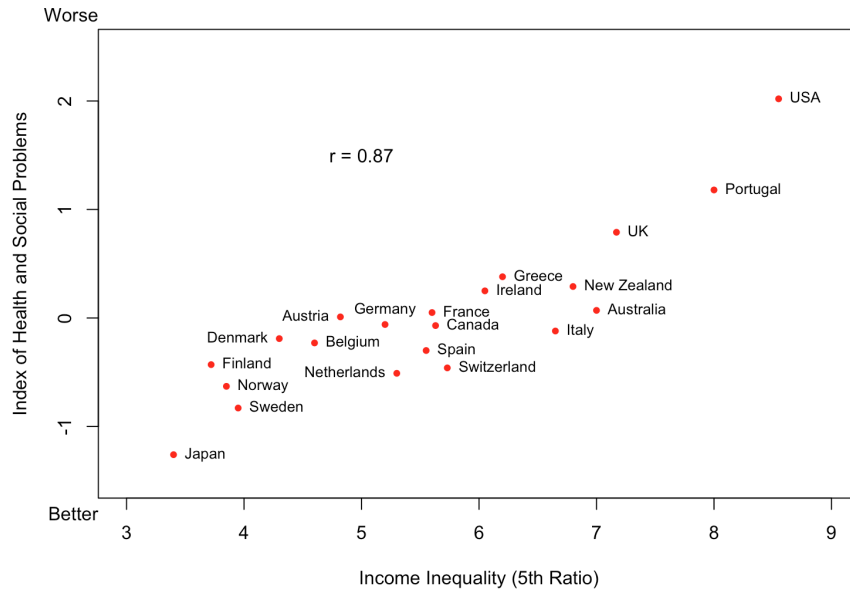
```
plot(Index_inequality.df, pch = 20, col = "red", xlim = c(3, 9), ylim = c(-1.5,
2.5), ann = FALSE)
text(Index_inequality.df, labels = Country, pos = pos.text, cex = 0.8)
main.title <- "Income Inequality vs Index of Health and Social Problems"
x.lab <- "Income Inequality (5th Ratio)"
y.lab <- "Index 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,

```
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, 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 =", round(cor(Index_inequality.df[1], In
dex_inequality.df[2]), digits = 2)))
```


Income Inequality vs Index of Health and Social Problems



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

```
## quartz_off_screen
## 3
```

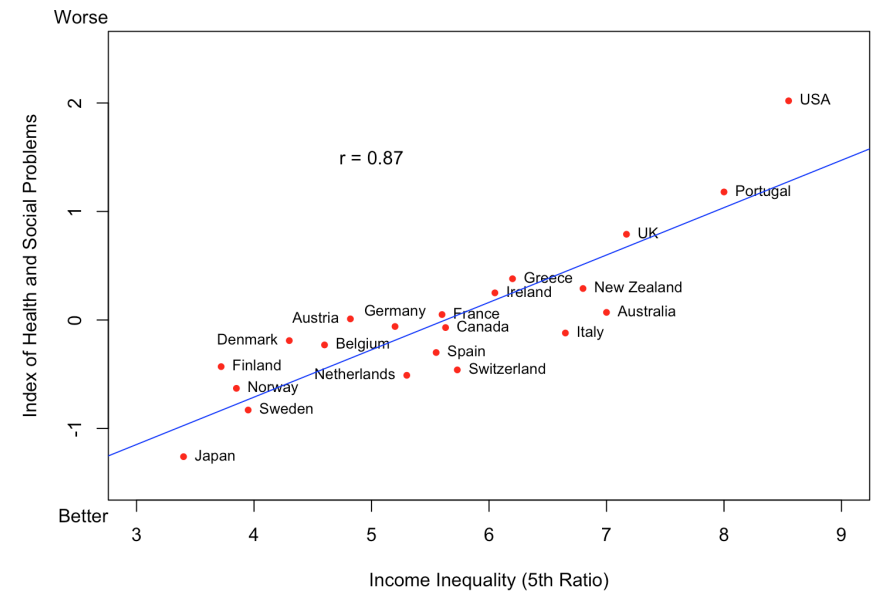
```
dev.off()
```

```
## quartz_off_screen
## 2
```

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

```
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, 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 =", round(cor(Index_inequality.df[1], Index_inequality.df[2]), digits = 2)))
lm.ineq <- lm(Index.HS ~ Income.inequality, data = Index_inequality.df)
# lm.ineq <- lm(Index_inequality.df[2:1])
abline(lm.ineq$coef, col = "blue")
```

Income Inequality vs Index of Health and Social Problems

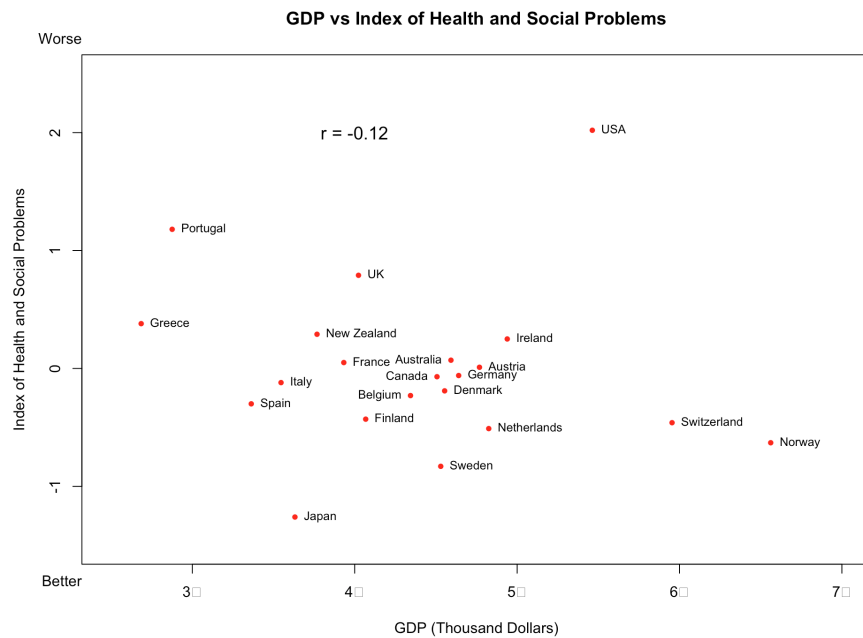


GDP와 건강 및 사회문제 지수

```

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)
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 <- "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)

```

```

## quartz_off_screen
##
3

```

```

dev.off()

```

```

## quartz_off_screen
##
2

```

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")

```

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

```

par(family = "Gulim")
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_inequality.df[2]), digits = 2)))
dev.copy(png, file = "../pics/inequality_health_social_72dpi.png", width = 640, height = 480)
dev.off()

```

선형회귀선을 이번에는 lsfit 을 이용하여 삽입

```

par(family = "Batang")
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_inequality.df[2]), digits = 2)))
lsfit.ineq <- lsfit(x = Index_inequality.df[, 1], y = Index_inequality.df[, 2])
abline(lsfit.ineq$coefficients, col = "blue")

```

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

미국의 경우

```
# data.usa <- read.xlsx("../data/USA-inequality.xls", 1)
str(data.usa)
```

```
## 'data.frame':   50 obs. of  20 variables:
## $ State      : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
## $ State.Abbrev : chr  "AL" "AK" "AZ" "AR" ...
## $ Income.Inequality : num  0.475 0.402 0.45 0.458 0.475 ...
## $ Trust      : num  23 NA 47 29 43 46 49 NA 37 38 ...
## $ Life.expectancy : num  74.6 76.7 77.5 75.1 78.3 ...
## $ Infant.mortality : num  9.1 5.5 6.4 8.3 5.5 ...
## $ Obesity     : num  32 30 28.5 31 31 21.5 26.5 27 27.5 30.5 ...
## $ Mental.health : num  3.3 2.8 2.2 3.2 3.3 ...
## $ Maths.and.literacy.scores : num  258 268 263 262 259 ...
## $ Teenage.births : num  62.9 42.4 69.1 68.5 48.5 ...
## $ Homicides     : num  78.9 85.6 80.4 56.1 60.5 ...
## $ Imprisonment  : num  509 413 507 415 478 357 372 429 447 502 ...
## $ Index.of.health...social.problems: num  1.385 0.137 0.212 0.948 0.327 ...
## $ Overweight.children : num  35 31 30 33 30 22 27 35 32 32 ...
## $ Child.wellbeing : num  8.5 4.4 4.9 9.3 -3.4 ...
## $ Women.s.status : num  -0.932 0.74 -0.147 -1.318 0.969 ...
## $ Juvenile.homicides : num  12 8 7 6 10 4 4 0 NA 8 ...
## $ High.school.drop.outs : num  24.7 11.7 19 24.7 23.2 ...
## $ Child.mental.illness : num  11.5 8.2 8.7 11.8 7.5 ...
## $ Pugnacity      : num  41.8 NA 36.3 38.4 37.7 ...
```

당장 필요한 변수들만 모아서 data frame으로 재구성한다. 변수명 설정에 유의한다.

```
data.usa.1 <- data.frame(Gini = data.usa$Income.Inequality, HS.index = data.usa$Index.of.health...social.problems)
str(data.usa.1)
```

```
## 'data.frame':   50 obs. of  2 variables:
## $ Gini      : num  0.475 0.402 0.45 0.458 0.475 ...
## $ HS.index: num  1.385 0.137 0.212 0.948 0.327 ...
```

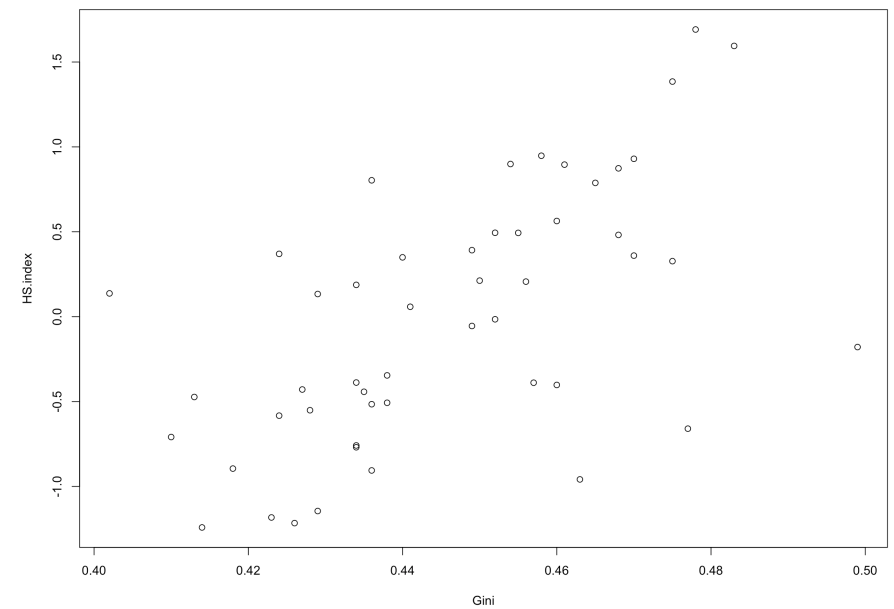
```
Abb <- data.usa$State.Abbrev
options(digits = 3)
kable(data.frame(State = Abb, data.usa.1))
```

State	Gini	HS.index
AL	0.475	1.385
AK	0.402	0.137
AZ	0.450	0.212
AR	0.458	0.948
CA	0.475	0.327
CO	0.438	-0.507
CT	0.477	-0.660
DE	0.429	0.133
FL	0.470	0.360
GA	0.461	0.896
HI	0.434	-0.388
ID	0.427	-0.429
IL	0.456	0.206
IN	0.424	0.370
IA	0.418	-0.895
KS	0.435	-0.442
KY	0.468	0.874
LA	0.483	1.595
ME	0.434	-0.769
MD	0.434	0.187
MA	0.463	-0.959

MI	0.440	0.349
MN	0.426	-1.216
MS	0.478	1.692
MO	0.449	0.392
MT	0.436	-0.906
NE	0.424	-0.583
NV	0.436	0.803
NH	0.414	-1.242
NJ	0.460	-0.402
NM	0.460	0.564
NY	0.499	-0.179
NC	0.452	0.494
ND	0.429	-1.145
OH	0.441	0.058
OK	0.455	0.494
OR	0.438	-0.346
PA	0.452	-0.015
RI	0.457	-0.389
SC	0.454	0.899
SD	0.434	-0.759
TN	0.465	0.788
TX	0.470	0.930
UT	0.410	-0.709
VT	0.423	-1.183
VA	0.449	-0.055
WA	0.436	-0.516
WV	0.468	0.482
WI	0.413	-0.473
WY	0.428	-0.551

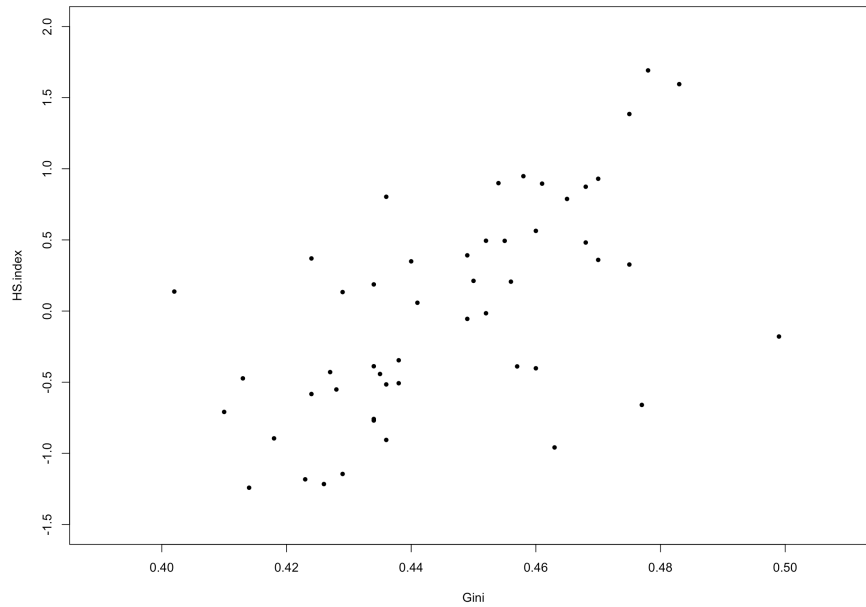
간단한 산점도를 그리고, 추가 작업을 생각한다.

```
plot(data.usa.1)
```



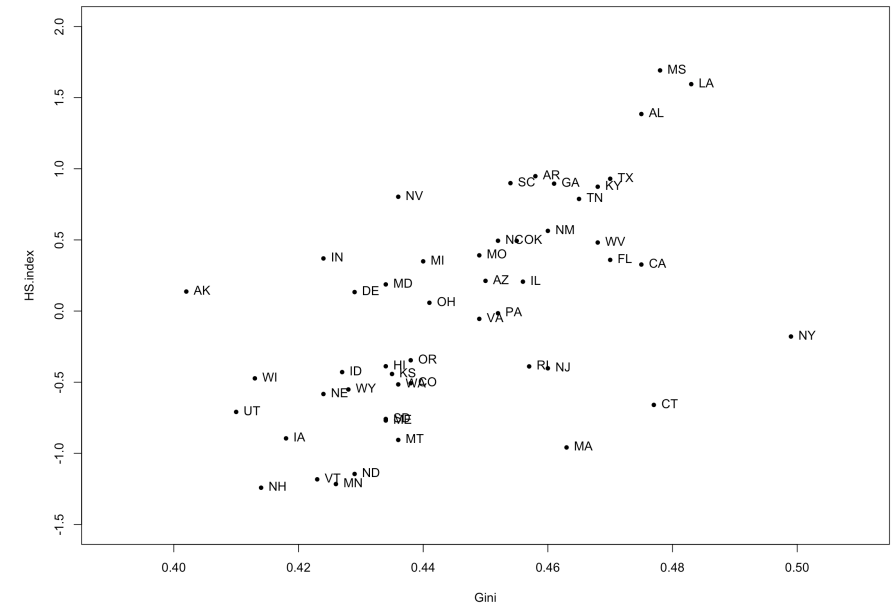
x-축과 y-축의 범위를 설정하고, pch = 20 으로 다시 그린다.

```
plot(data.usa.1, pch = 20, xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
```



각 주의 약칭을 새겨넣는다.

```
plot(data.usa.1, pch = 20, xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
text(data.usa.1, labels = Abb, pos = 4)
```



겉쳐보이는 주의 약칭들로부터 인덱스를 추출한다.

```
which(Abb %in% c("VT", "ME", "NE", "WA", "VA", "HI", "RI", "SC", "AR", "NC", "GA", "KY"))
```

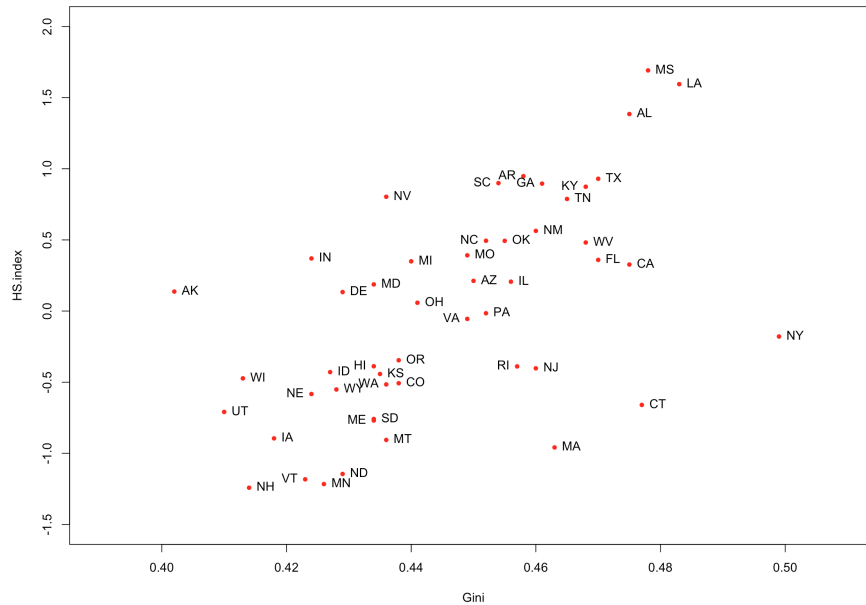
```
## [1] 4 10 11 17 19 27 33 39 40 45 46 47
```

점 왼쪽에 약칭을 넣을 주들의 인덱스를 저장한다. 나머지 인덱스는 오른쪽에 넣을 것으로 따로 저장한다.

```
text.left.us <- which(Abb %in% c("VT", "ME", "NE", "WA", "VA", "HI", "RI", "SC", "AR", "NC", "GA", "KY"))
text.right.us <- setdiff(1:nrow(data.usa.1), text.left.us)
pos.text.us <- ifelse(1:nrow(data.usa.1) %in% text.left.us, 2, 4)
```

왼쪽, 오른쪽 위치를 조정한 주 약칭을 다시 넣는다.

```
plot(data.usa.1, pch = 20, col = "red", xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
text(data.usa.1, labels = Abb, pos = pos.text.us)
```



점 아래에 약칭을 넣을 주들의 인덱스를 찾는다. 왼쪽 인덱스, 오른쪽 인덱스에서 조정한다.

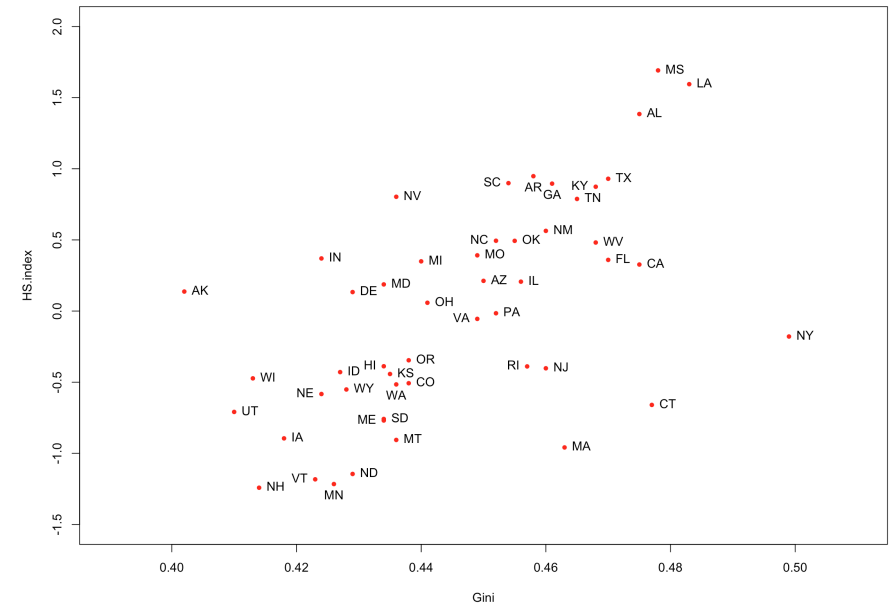
```
text.down.us <- which(Abb %in% c("WA", "AR", "GA", "MN"))
which(text.left.us %in% text.down.us)
```

```
## [1] 1 2 12
```

```
text.left.us <- setdiff(text.left.us, text.down.us)
text.right.us <- setdiff(text.right.us, text.down.us)
pos.text.us <- ifelse(1:nrow(data.usa.1) %in% text.down.us, 1, ifelse(1:nrow(data.usa.1) %in% text.left.us, 2, 4))
```

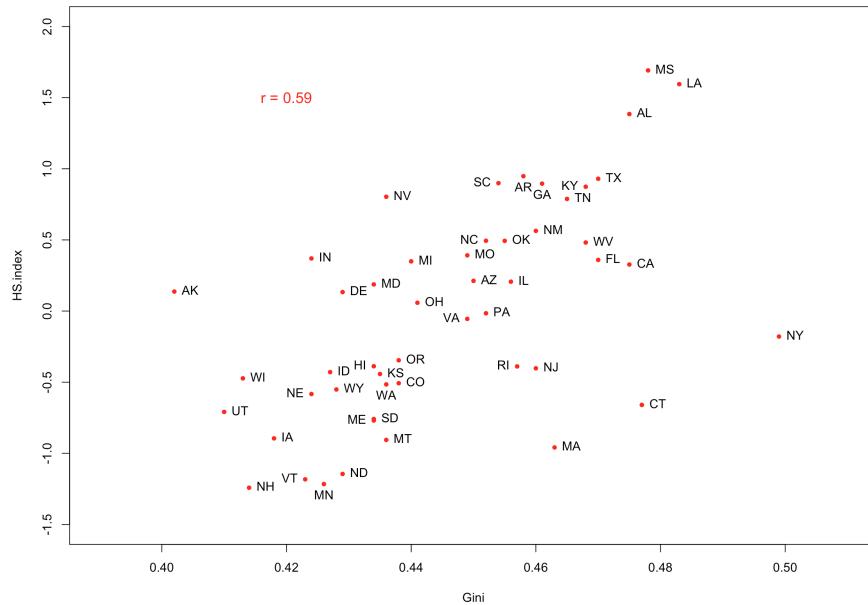
약칭 위치를 아래로 조정한 산점도를 다시 그린다.

```
plot(data.usa.1, pch = 20, col = "red", xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
text(data.usa.1, labels = Abb, pos = pos.text.us)
```



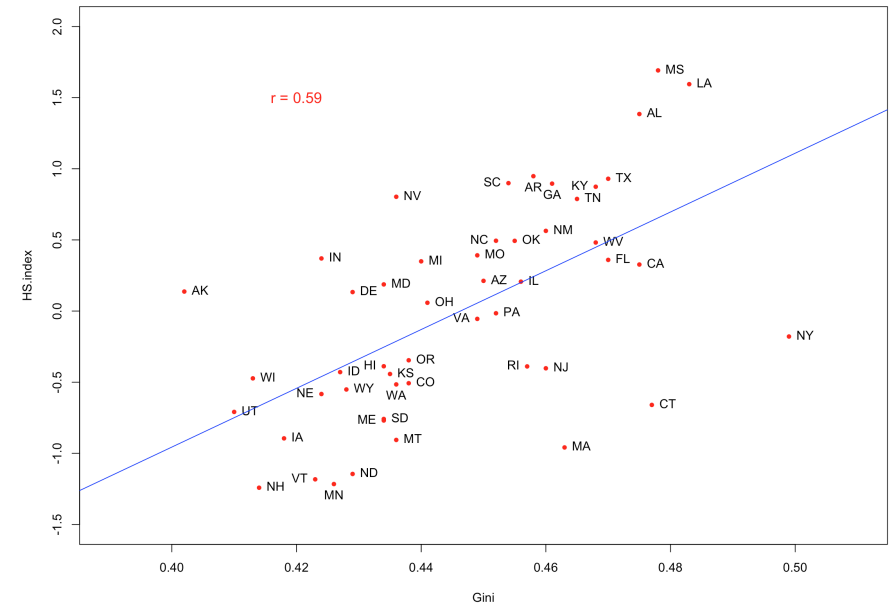
상관계수를 추가한다.

```
plot(data.usa.1, pch = 20, col = "red", xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
text(data.usa.1, labels = Abb, pos = pos.text.us)
text(x = 0.42, y = 1.5, labels = paste("r =", round(cor(data.usa.1$HS.index, data.usa.1$Gini), digits = 2)), col = "red", cex = 1.2)
```



단순회귀선을 추가한다.

```
plot(data.usa.1, pch = 20, col = "red", xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0))
text(data.usa.1, labels = Abb, pos = pos.text.us)
text(x = 0.42, y = 1.5, labels = paste("r =", round(cor(data.usa.1$HS.index, data.usa.1$Gini), digits = 2)), col = "red", cex = 1.2)
lm.ineq.us <- lm(data.usa.1[2:1])
abline(lm.ineq.us$coef, col = "blue")
```



```
# abline(lm(HS.index ~ Gini, data = data.usa.1)$coef)
```

주제목을 추가하고, xlab, ylab 을 수정한다. 수직축의 의미를 명확히 한다.

```
plot(data.usa.1, pch = 20, col = "red", xlim = c(0.39, 0.51), ylim = c(-1.5, 2.0), ann = FALSE)
text(data.usa.1, labels = Abb, pos = pos.text.us)
text(x = 0.42, y = 1.5, labels = paste("r =", round(cor(data.usa.1[1], data.usa.1[2]), digits = 2)), col = "red", cex = 1.2)
abline(lm(HS.index ~ Gini, data = data.usa.1)$coef, col = "blue")
mtext(c("Better", "Worse"), side = 2, at = c(-1.8, 2.3), las = 1)
main.title.us <- "Income Inequality vs Health and Social Index (USA)"
x.lab.us <- "Gini Coefficients"
y.lab.us <- "Index of Health and Social Problems"
title(main = main.title.us, xlab = x.lab.us, ylab = y.lab.us)
```

