

Income Inequality vs Index of Health and Social Problems

coop711

2015년 5월 3일

Data Preparation

xlsx package은 Excel 자료를 다루는 데 매우 유용한데, read.xlsx(filename, n) 의 구조로 되어 있으며, 여기서 n 은 엑셀 시트의 번호이다.

```
install.packages("xlsx", repos="http://cran.rstudio.com")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/_h/tg1th9bd4h98rjjb5vy9gn3m0000gn/T//Rtmp0LpwZ6/downloaded_packages
```

```
library(xlsx)
```

Equality Trust에서 기부금을 받고 제공하는 두 종류의 자료 중 23개 국가의 각종 지표를 비교한 것을 읽어들이면,

```
data.full<-read.xlsx("international-inequality.xls", 1)  
str(data.full)
```

```
## 'data.frame':    23 obs. of  29 variables:
## $ Country          : chr  "Australia" "Austria" "Belgium"
"Canada" ...
## $ Income.inequality : num  7 4.82 4.6 5.62 4.3 ...
## $ Trust             : num  39.9 33.9 30.7 38.8 66.5 ...
## $ Life.expectancy   : num  79.2 78.5 78.8 79.3 76.6 ...
## $ Infant.mortality  : num  4.9 4.8 5 5.3 5.3 3.7 4.4 4.4 5
5.9 ...
## $ Obesity           : num  18.4 14.5 13.5 12.8 15 ...
## $ Mental.illness    : num  23 NA 12 19.9 NA ...
## $ Maths.and.literacy.scores : num  524 498 518 530 503 ...
## $ Teenage.births    : num  18.4 14 9.9 20.2 8.1 ...
## $ Homicides         : num  16.9 11.6 13 17.3 12.7 ...
## $ Imprisonment..log. : num  4.61 4.52 4.28 4.77 4.17 ...
## $ Social.mobility    : num  NA NA NA 0.143 0.143 ...
## $ Index.of.health...social_problems: num  0.0652 0.01 -0.2333 -0.0707 -0.18
83 ...
## $ Child.overweight  : num  NA 11.9 10.4 19.5 10.3 ...
## $ Drugs.index       : num  1.7074 -0.0182 -0.1778 0.6068
-0.0948 ...
## $ Calorie.intake    : num  3142 3753 3632 3167 3405 ...
## $ Public.health.expenditure : num  67.9 69.3 71.7 70.8 82.4 75.6 76
74.9 56 76 ...
## $ Child.wellbeing   : num  -0.2147 -0.0679 0.047 0.0395 0.21
4 ...
## $ Maths.education.science.score : num  525 496 515 526 494 ...
## $ Child.conflict    : num  NA 0.309 0.329 0.237 -0.137 ...
## $ Foreign.aid       : num  0.25 0.52 0.53 0.34 0.81 ...
## $ Recycling         : num  7.4 NA NA NA NA ...
## $ Peace.index       : num  1.66 1.48 1.49 1.48 1.38 ...
## $ Maternity.leave   : num  0 16 15 17 18 18 16 14 17 18 ...
## $ Advertising       : num  1.24 0.97 0.82 0.77 0.75 ...
## $ Police            : num  304 305 357 186 192 160 NA 303 NA
NA ...
## $ Social.expenditure : num  17.8 27.5 26.5 17.2 27.6 25.8 29
27.3 19.9 15.8 ...
## $ Women.s_status    : num  0.457 -0.808 0.606 0.563 0.827
...
## $ Lone.parents      : num  21 15 12 17 22 19 12 21 3 14 ...
```

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(is.na(data.full$Index.of.health...social_problems))
```

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

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"
```

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

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

새로운 data frame 을 data.1 으로 저장하자. 시각적 가독성을 높이기 위하여 자릿수를 조정한다.

```
options(digits=2)
data.1<-data.full[-c(11, 18), c(1, 2, 13)]
data.1
```

##	Country	Income.inequality	Index.of.health...social_problems
## 1	Australia	7.0	0.065
## 2	Austria	4.8	0.010
## 3	Belgium	4.6	-0.233
## 4	Canada	5.6	-0.071
## 5	Denmark	4.3	-0.188
## 6	Finland	3.7	-0.433
## 7	France	5.6	0.051
## 8	Germany	5.2	-0.065
## 9	Greece	6.2	0.380
## 10	Ireland	6.1	0.254
## 12	Italy	6.7	-0.116
## 13	Japan	3.4	-1.256
## 14	Netherlands	5.3	-0.510
## 15	New Zealand	6.8	0.291
## 16	Norway	3.9	-0.633
## 17	Portugal	8.0	1.180
## 19	Spain	5.6	-0.304
## 20	Sweden	4.0	-0.834
## 21	Switzerland	5.7	-0.455
## 22	UK	7.2	0.789
## 23	USA	8.5	2.021

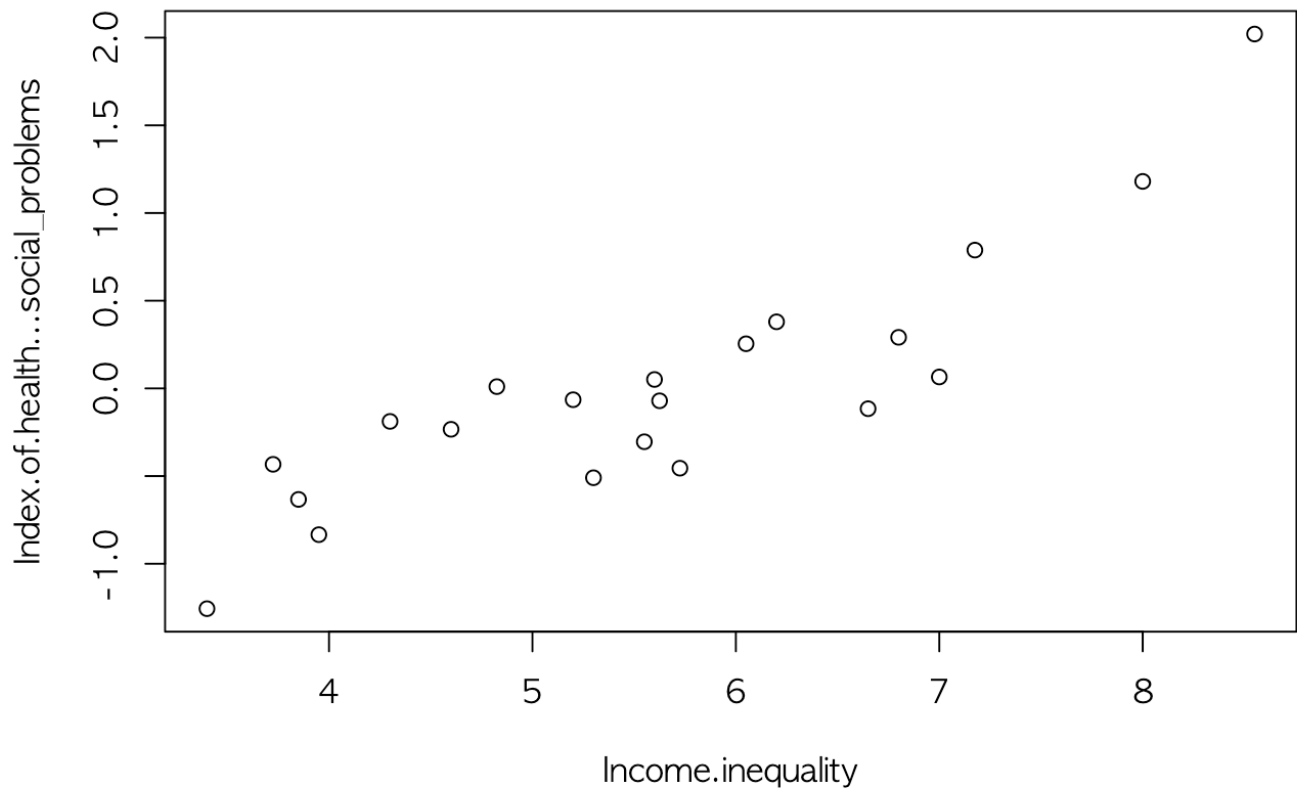
Plots

plot 과 기타 작업의 편의를 위하여 data.1 을 검색 목록에 올린다.

```
attach(data.1)
```

우선 소득불평등과 건강 및 사회문제 지표의 관계를 대략적으로 살펴보면,

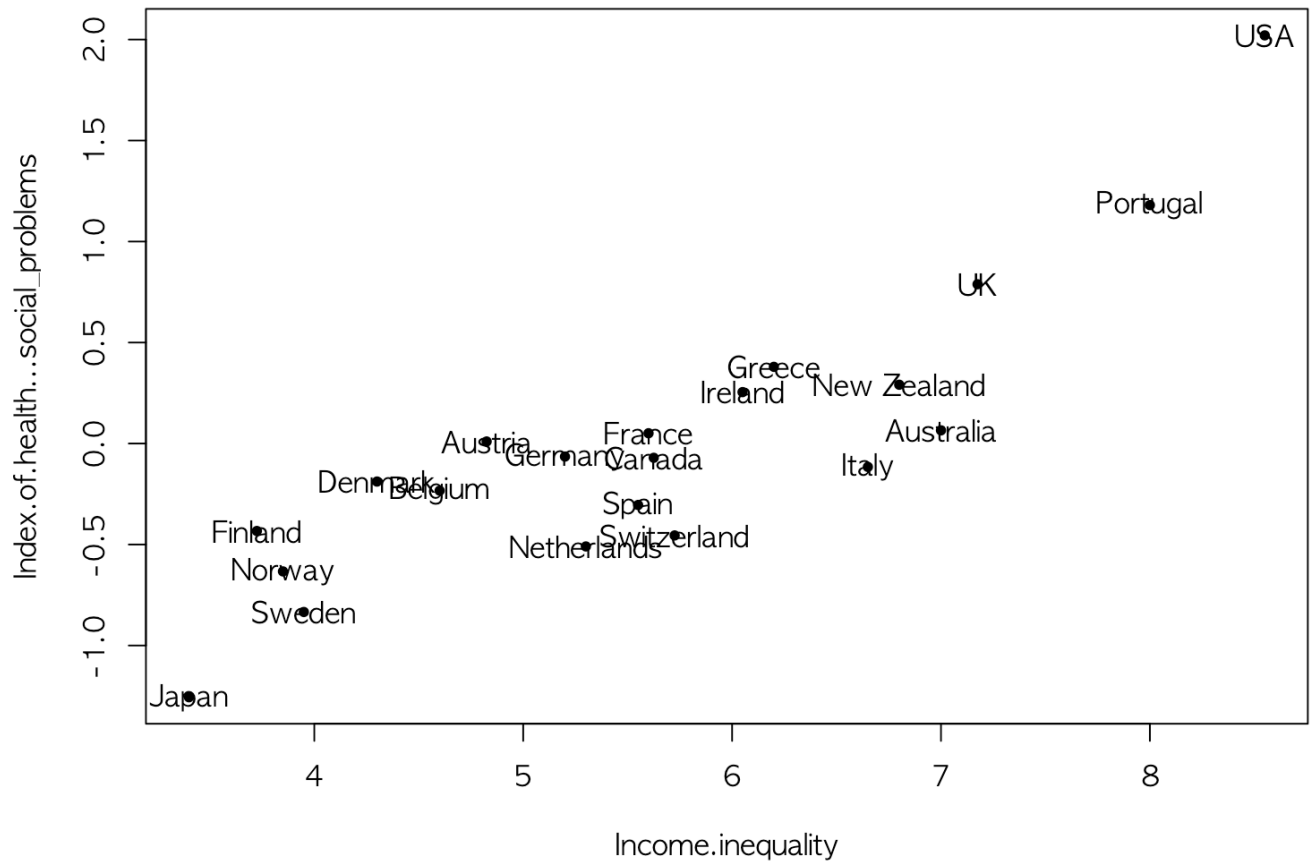
```
plot(Index.of.health...social_problems~Income.inequality, data=data.1)
```



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.

```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20)
text(Income.inequality, Index.of.health...social_problems, labels=Country)
```

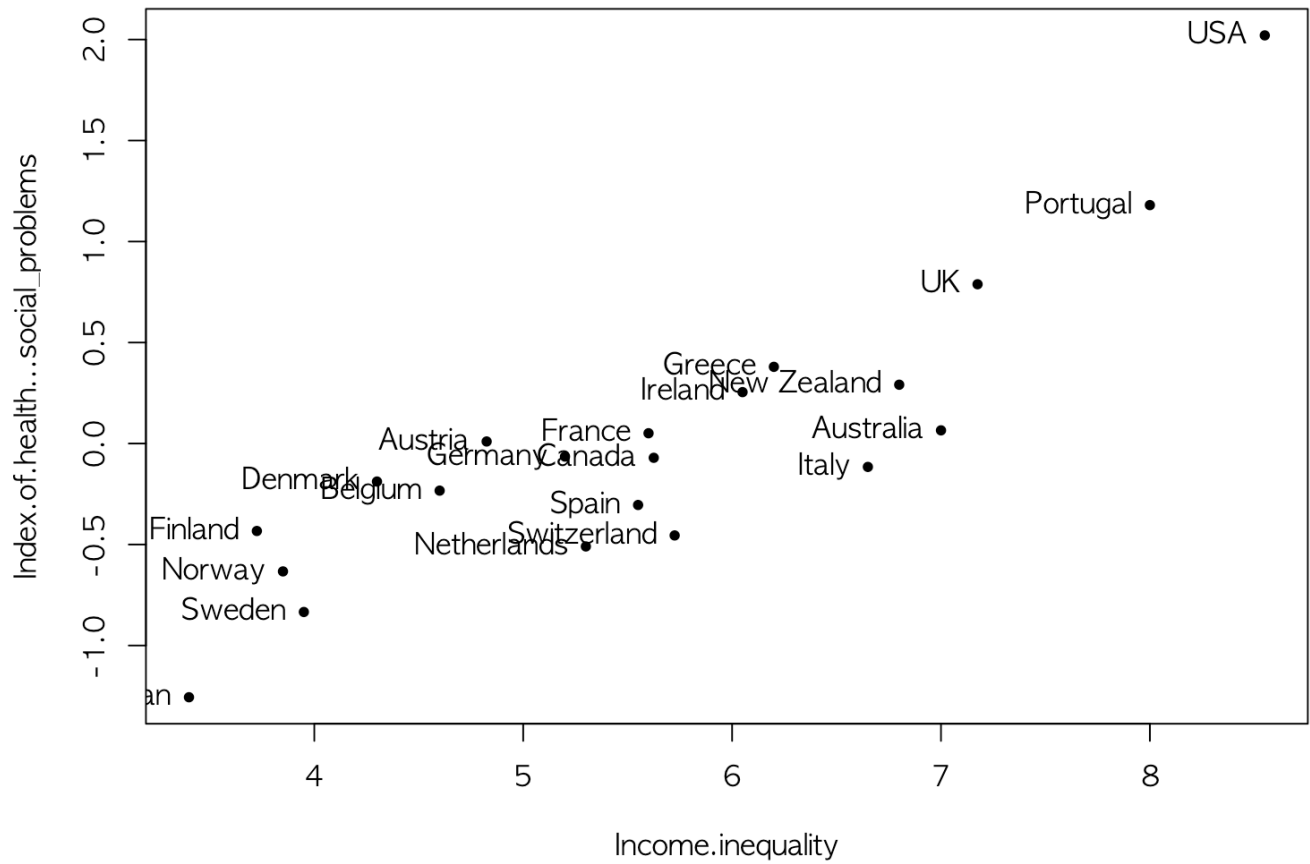


text label의 위치 기본값은 바로 점 위임을 알 수 있다. 위치 선정에 가능한 값들을 넣어보자.

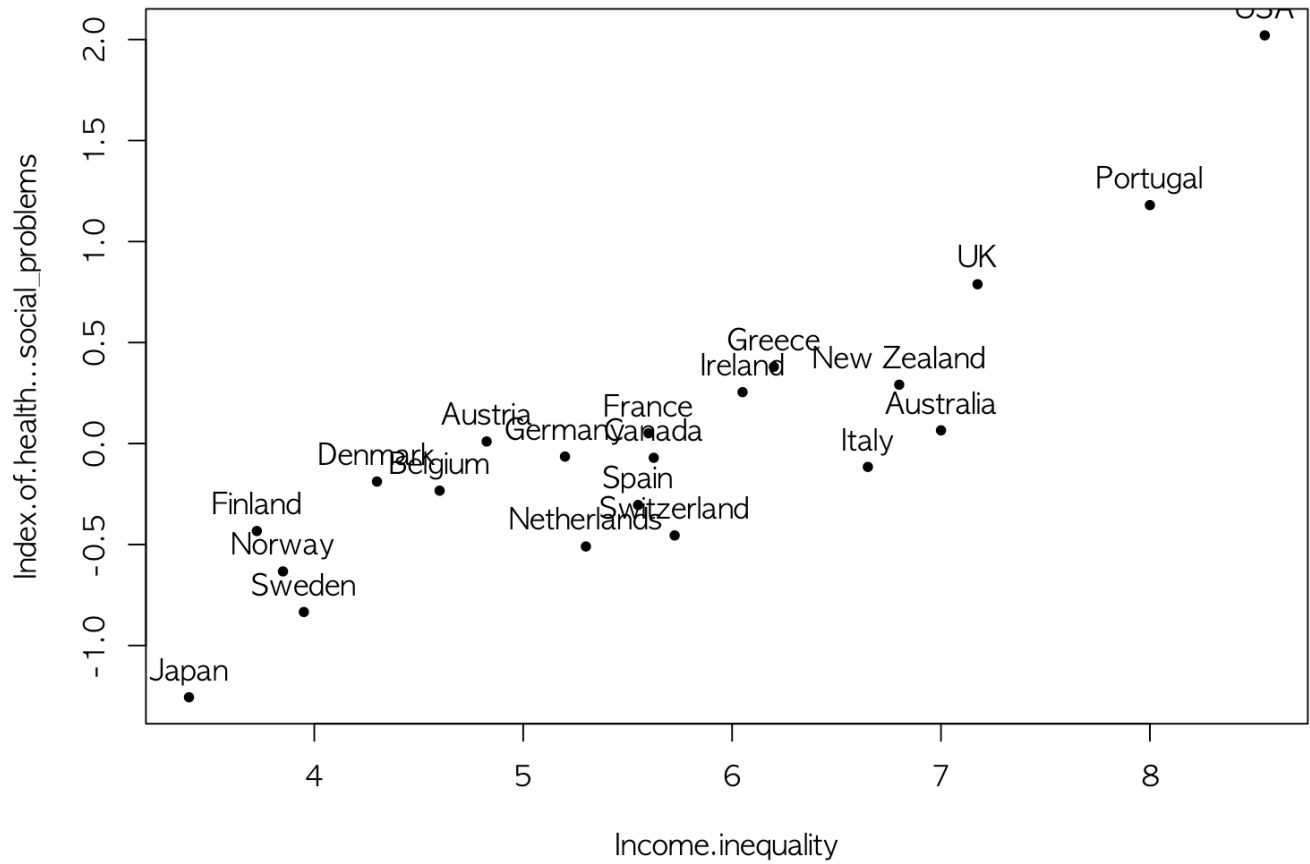
```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20)
text(Income.inequality, Index.of.health...social_problems, labels=Country, pos=1)
```



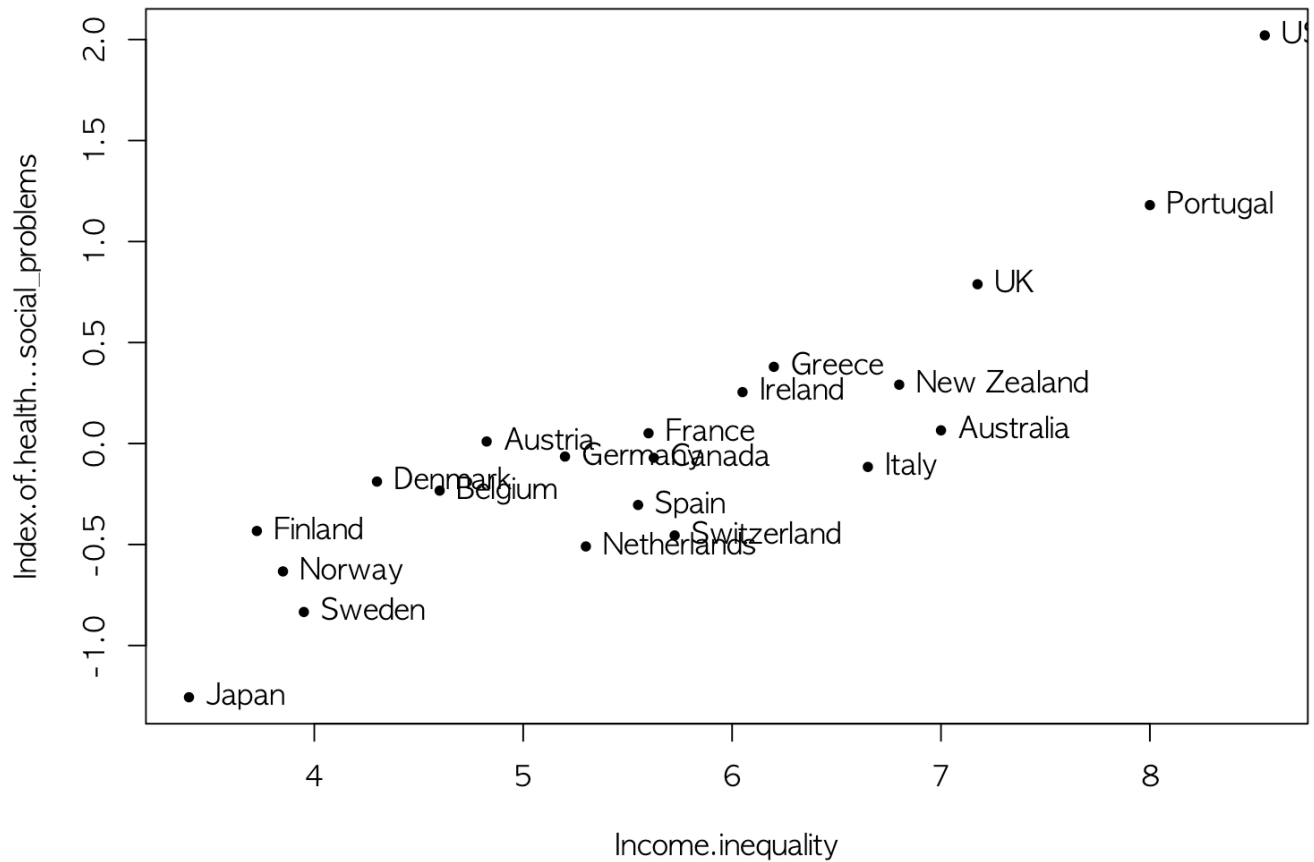
```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20)
text(Income.inequality, Index.of.health...social_problems, labels=Country, pos=2)
```



```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20)
text(Income.inequality, Index.of.health...social_problems, labels=Country, pos=3)
```

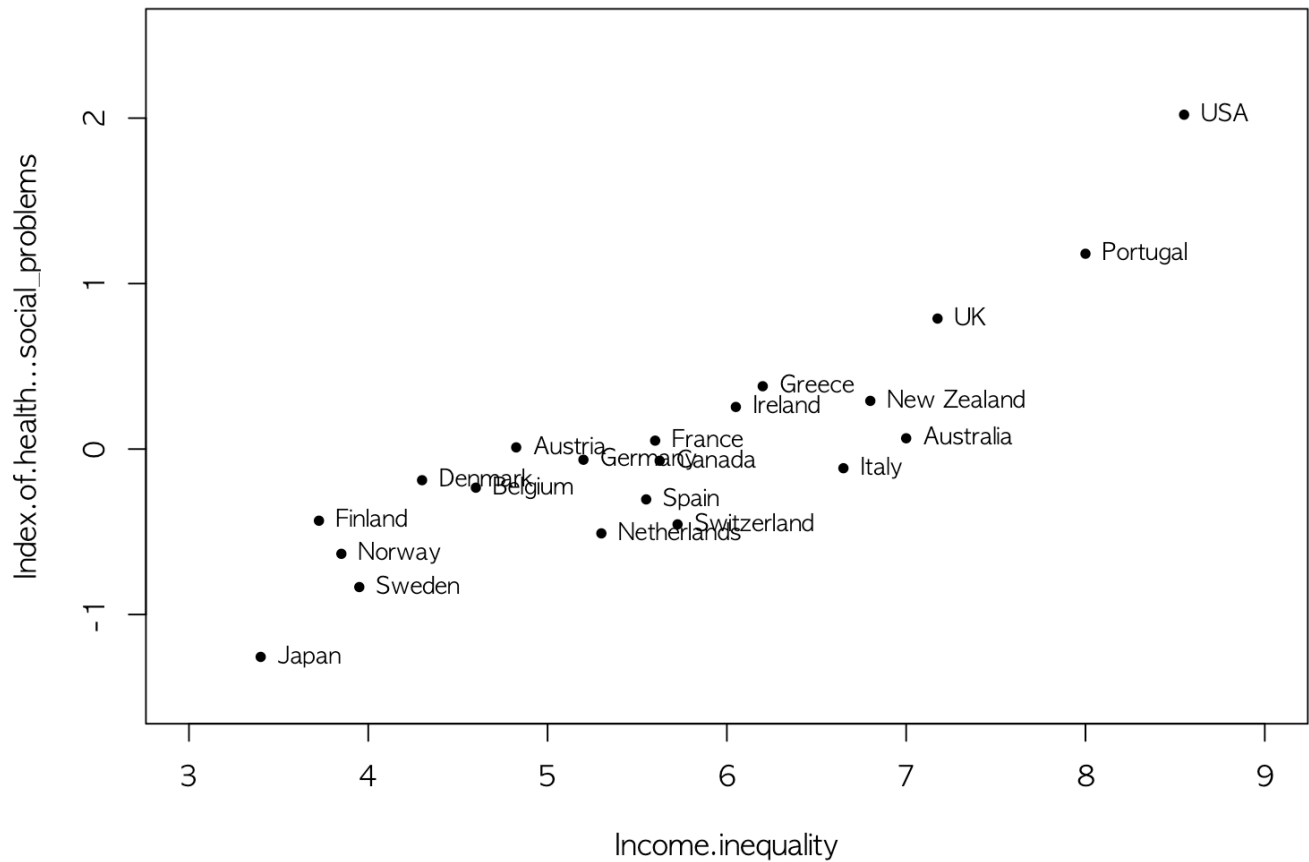



```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20)
text(Income.inequality, Index.of.health...social_problems, labels=Country, pos=4)
```



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.of.health...social_problems~Income.inequality, data=data.1, pch=20,
      xlim=c(3, 9), ylim=c(-1.5, 2.5))
text(Income.inequality, Index.of.health...social_problems, labels=Country, pos=4, cex=0.8)
```



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

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

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

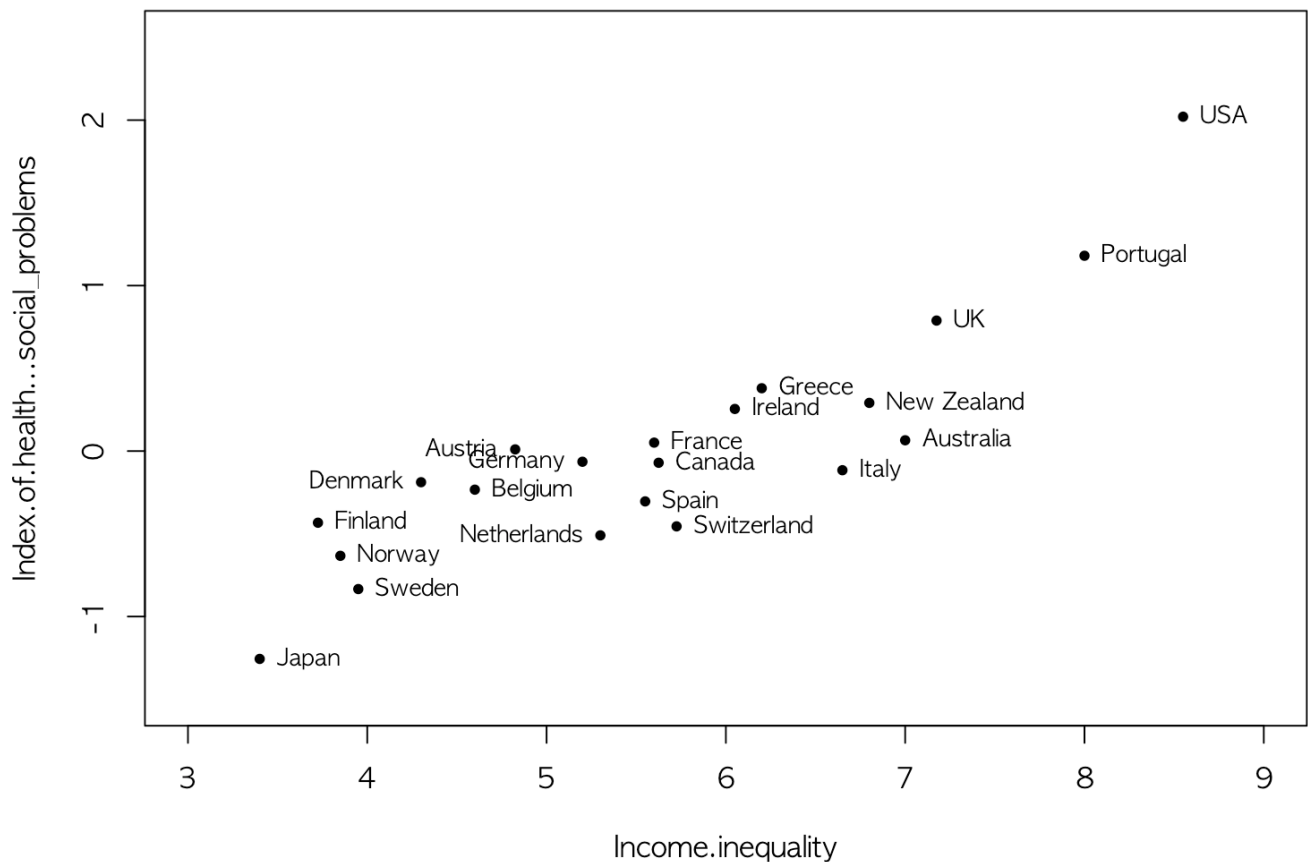
```
text.left<-which(Country %in% c("Austria", "Denmark", "Germany", "Netherlands"))
text.left
```

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

```
text.right<-(1:21)[-text.left]
text.right
```

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

```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20,
xlim=c(3, 9), ylim=c(-1.5, 2.5))
text(Income.inequality[text.right], Index.of.health...social_problems[text.righ
t], labels=Country[text.right], pos=4, cex=0.8)
text(Income.inequality[text.left], Index.of.health...social_problems[text.lef
t], labels=Country[text.left], pos=2, cex=0.8)
```



독일의 라벨을 위로 붙이면 보기가 나아질 것으로 생각되므로,

```
which(Country == "Germany")
```

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

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

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

```
which(text.left==text.up)
```

```
## [1] 3
```

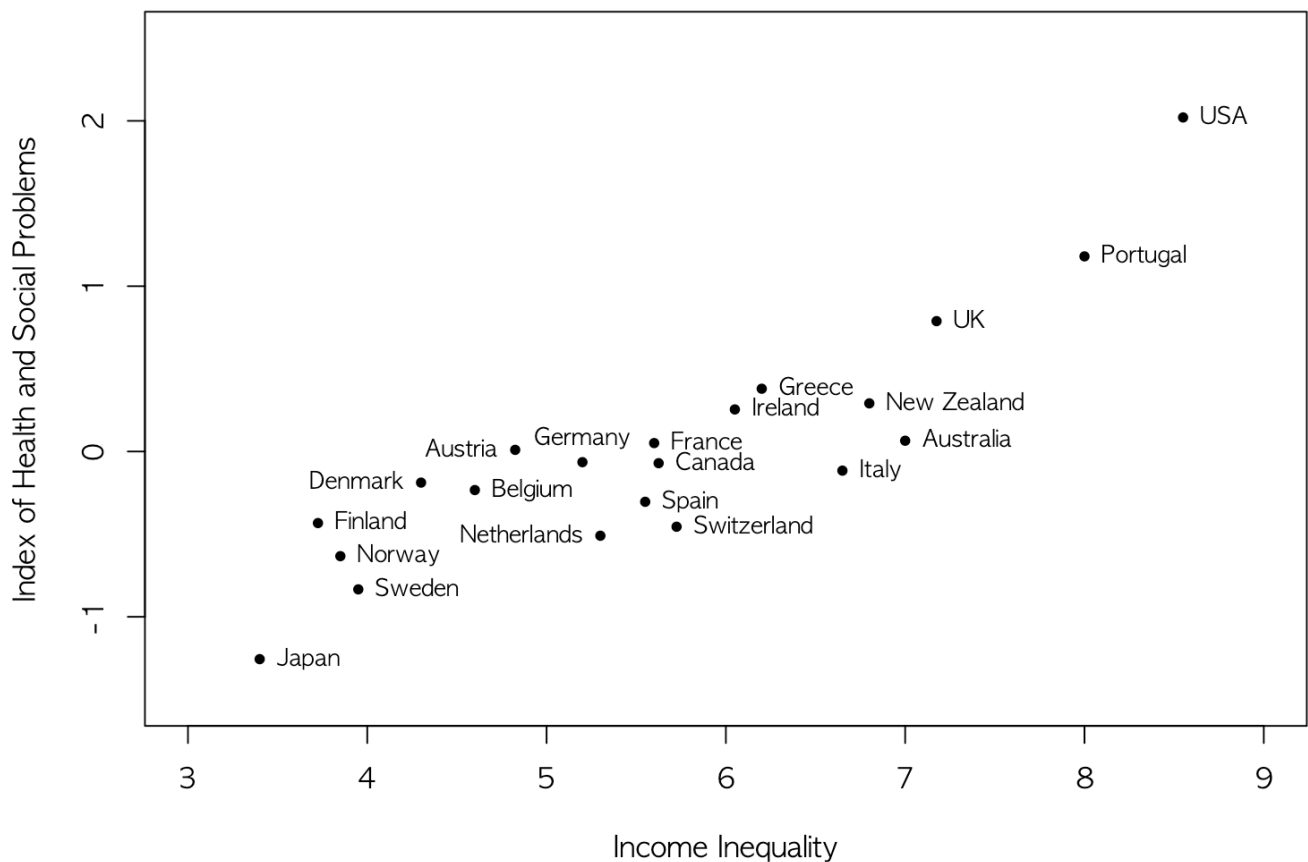
```
text.left<-text.left[!(text.left == text.up)]
text.left
```

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

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

```
plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20,
xlim=c(3, 9), ylim=c(-1.5, 2.5), ann=FALSE)
text(Income.inequality[text.right], Index.of.health...social_problems[text.righ
t], labels=Country[text.right], pos=4, cex=0.8)
text(Income.inequality[text.left], Index.of.health...social_problems[text.lef
t], labels=Country[text.left], pos=2, cex=0.8)
text(Income.inequality[text.up], Index.of.health...social_problems[text.up], la
bels=Country[text.up], pos=3, cex=0.8)
title(main="Income Inequality vs Index of Health and Social Problems", xlab="In
come Inequality", ylab="Index of Health and Social Problems")
```

Income Inequality vs Index of Health and Social Problems

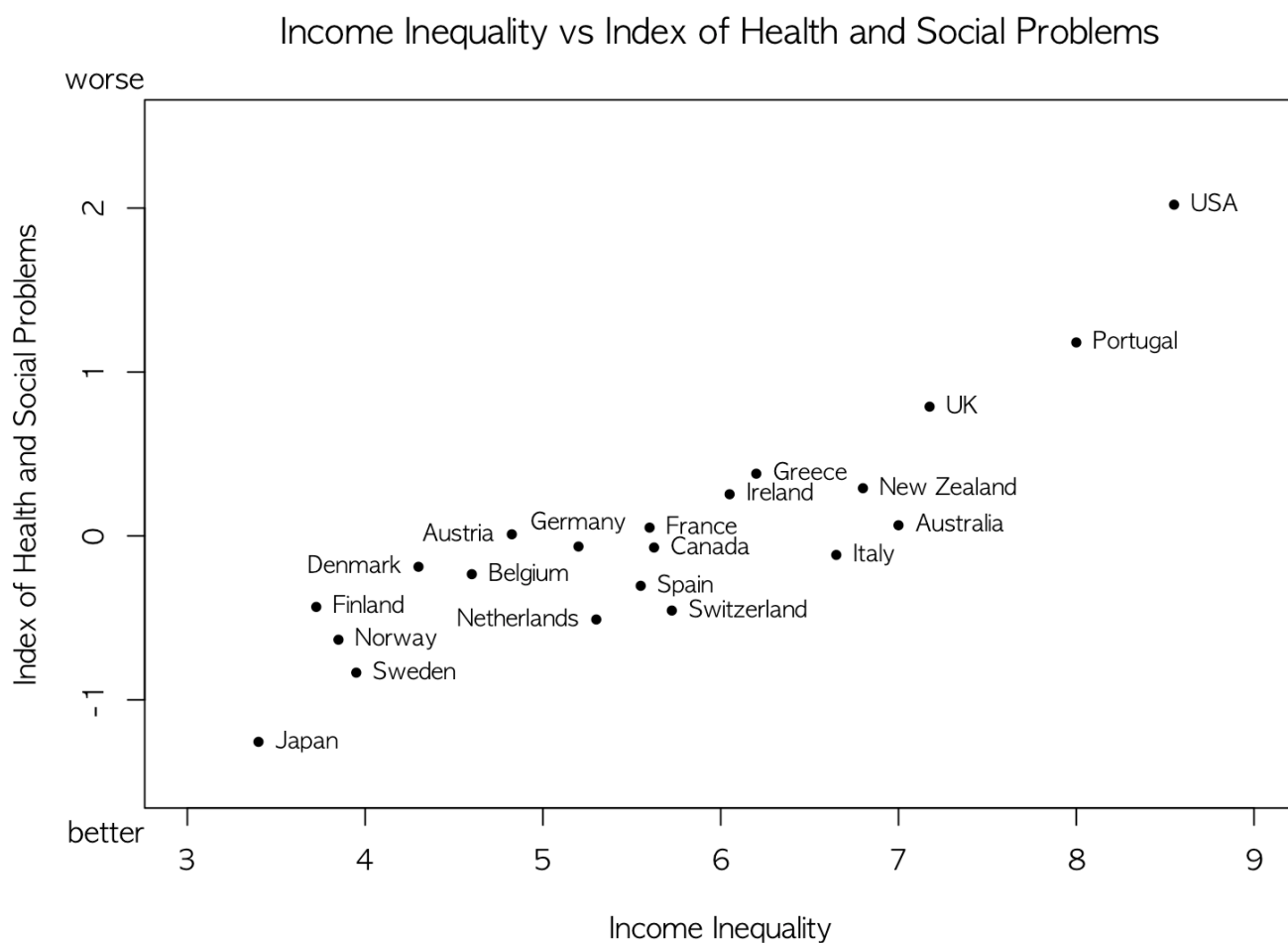


건강 및 사회문제 지표의 경우 어느 방향이 좋은지 알 수 없으므로 친절하게 도표의 주변에(margin)에 알려주려면,

```

plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20,
xlim=c(3, 9), ylim=c(-1.5, 2.5), ann=FALSE)
text(Income.inequality[text.right], Index.of.health...social_problems[text.righ
t], labels=Country[text.right], pos=4, cex=0.8)
text(Income.inequality[text.left], Index.of.health...social_problems[text.lef
t], labels=Country[text.left], pos=2, cex=0.8)
text(Income.inequality[text.up], Index.of.health...social_problems[text.up], la
bels=Country[text.up], pos=3, cex=0.8)
title(main="Income Inequality vs Index of Health and Social Problems", xlab="In
come Inequality", ylab="Index of Health and Social Problems")
mtext("worse", side=2, at=2.8, las=1)
mtext("better", side=2, at=-1.8, las=1)

```



한글화 작업

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")

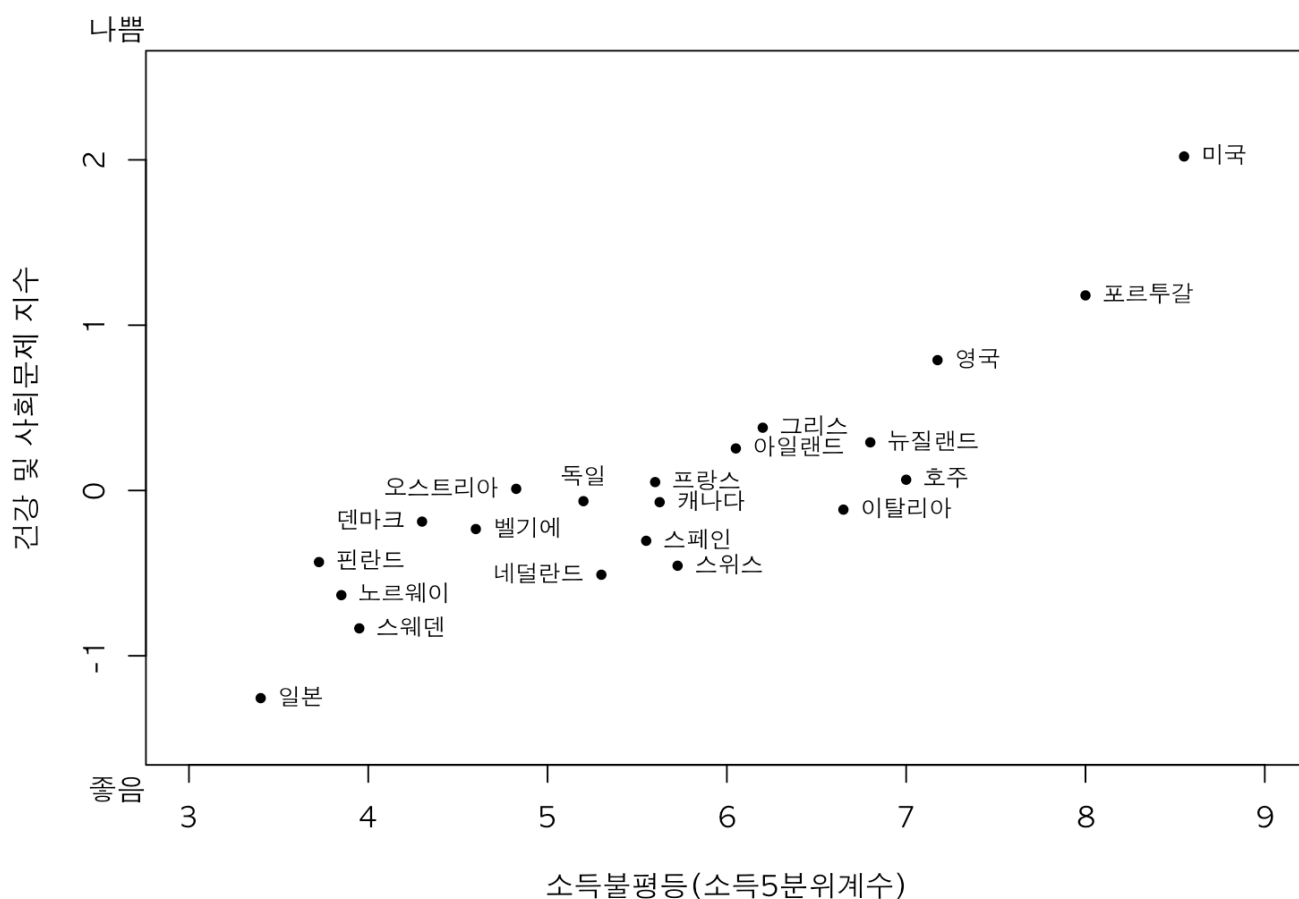
```

```

plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20,
xlim=c(3, 9), ylim=c(-1.5, 2.5), ann=FALSE)
text(Income.inequality[text.right], Index.of.health...social_problems[text.righ
t], labels=Country.kr[text.right], pos=4, cex=0.8)
text(Income.inequality[text.left], Index.of.health...social_problems[text.lef
t], labels=Country.kr[text.left], pos=2, cex=0.8)
text(Income.inequality[text.up], Index.of.health...social_problems[text.up], la
bels=Country.kr[text.up], pos=3, cex=0.8)
title(main="소득불평등과 건강 및 사회문제 지수", xlab="소득불평등(소득5분위계수)", ylab="건강
및 사회문제 지수")
mtext("나쁨", side=2, at=2.8, las=1)
mtext("좋음", side=2, at=-1.8, las=1)

```

소득불평등과 건강 및 사회문제 지수



끝으로 두 변수 간의 상관계수 $r = 0.87$ 를 도표 안에 표시하려면, 좌표를 잡아서

```

plot(Index.of.health...social_problems~Income.inequality, data=data.1, pch=20,
xlim=c(3, 9), ylim=c(-1.5, 2.5), ann=FALSE)
text(Income.inequality[text.right], Index.of.health...social_problems[text.righ
t], labels=Country.kr[text.right], pos=4, cex=0.8)
text(Income.inequality[text.left], Index.of.health...social_problems[text.lef
t], labels=Country.kr[text.left], pos=2, cex=0.8)
text(Income.inequality[text.up], Index.of.health...social_problems[text.up], la
bels=Country.kr[text.up], pos=3, cex=0.8)
text(x=5, y=1.5, labels=paste("r =", round(cor(data.1[,2], data.1[,3]), digit
s=2)))
title(main="소득불평등과 건강 및 사회문제 지수", xlab="소득불평등(소득5분위계수)", ylab="건강
및 사회문제 지수")
mtext("나쁨", side=2, at=2.8, las=1)
mtext("좋음", side=2, at=-1.8, las=1)

```

소득불평등과 건강 및 사회문제 지수

