Wilkinson_USA

coop711

2015년 5월 5일

미국의 경우

`xlsx` 패키지를 이용하여 자료를 읽어들인다.

```
library(xlsx)
```

```
## Loading required package: rJava
## Loading required package: xlsxjars
```

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

```
## 'data.frame':    50 obs. of  20 variables:
##  $ State                           : chr  "Alabama" "Alaska" "Arizona" "Ark
ansas" ...
##  $ 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 2
7.5 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 4
47 502 ...
##  $ 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 ...
```

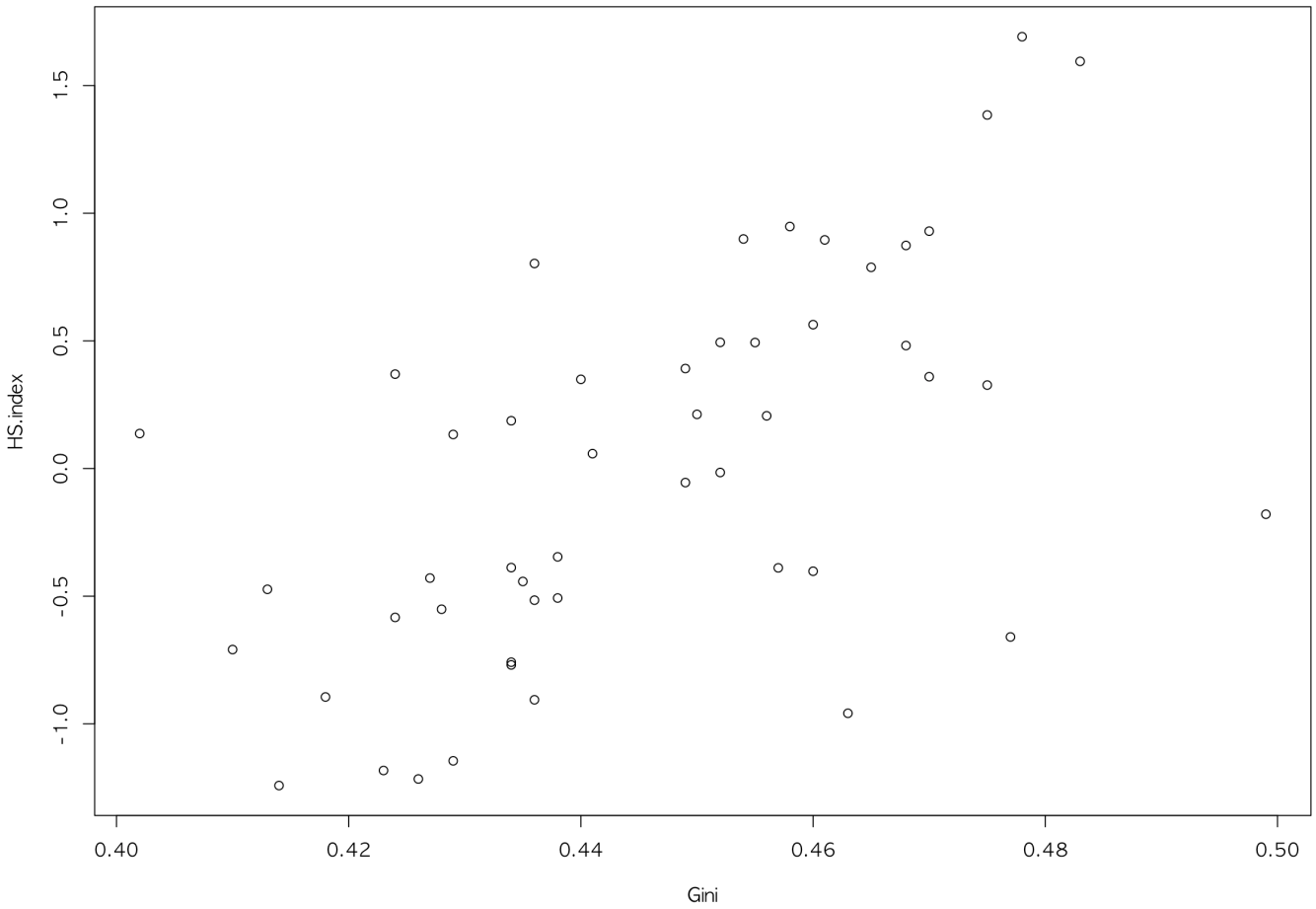
당장 필요한 변수들만 모아서 data frame으로 재구성한다. 변수명 설정에 유의한다.

```
data.usa.1<-data.frame(Abb=data.usa$State.Abbrev, Gini=data.usa$Income.Inequali
ty, HS.index=data.usa$Index.of.health...social.problems)
options(digits=3)
data.usa.1
```

```
##    Abb  Gini HS.index
## 1   AL 0.475   1.3849
## 2   AK 0.402   0.1371
## 3   AZ 0.450   0.2122
## 4   AR 0.458   0.9480
## 5   CA 0.475   0.3270
## 6   CO 0.438  -0.5072
## 7   CT 0.477  -0.6597
## 8   DE 0.429   0.1334
## 9   FL 0.470   0.3596
## 10  GA 0.461   0.8956
## 11  HI 0.434  -0.3880
## 12  ID 0.427  -0.4291
## 13  IL 0.456   0.2065
## 14  IN 0.424   0.3698
## 15  IA 0.418  -0.8948
## 16  KS 0.435  -0.4423
## 17  KY 0.468   0.8738
## 18  LA 0.483   1.5948
## 19  ME 0.434  -0.7692
## 20  MD 0.434   0.1873
## 21  MA 0.463  -0.9586
## 22  MI 0.440   0.3494
## 23  MN 0.426  -1.2160
## 24  MS 0.478   1.6915
## 25  MO 0.449   0.3917
## 26  MT 0.436  -0.9058
## 27  NE 0.424  -0.5831
## 28  NV 0.436   0.8032
## 29  NH 0.414  -1.2417
## 30  NJ 0.460  -0.4022
## 31  NM 0.460   0.5636
## 32  NY 0.499  -0.1790
## 33  NC 0.452   0.4942
## 34  ND 0.429  -1.1450
## 35  OH 0.441   0.0583
## 36  OK 0.455   0.4935
## 37  OR 0.438  -0.3459
## 38  PA 0.452  -0.0155
## 39  RI 0.457  -0.3891
## 40  SC 0.454   0.8992
## 41  SD 0.434  -0.7585
## 42  TN 0.465   0.7881
## 43  TX 0.470   0.9299
## 44  UT 0.410  -0.7090
## 45  VT 0.423  -1.1828
## 46  VA 0.449  -0.0550
## 47  WA 0.436  -0.5156
## 48  WV 0.468   0.4817
## 49  WI 0.413  -0.4731
## 50  WY 0.428  -0.5512
```

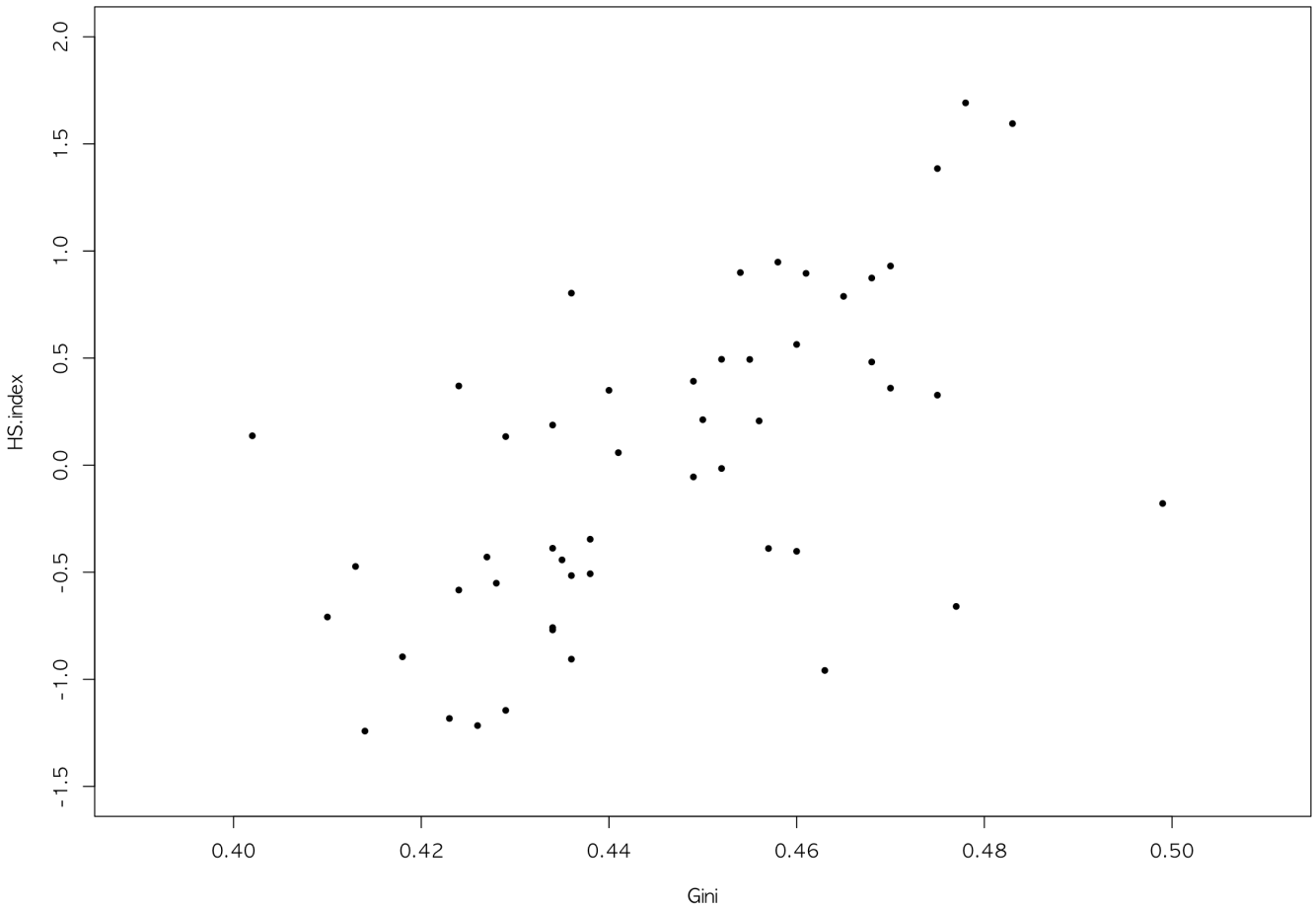
간단한 산점도를 그리고, 추가 작업을 생각한다.

```
plot(HS.index~Gini, data=data.usa.1)
```



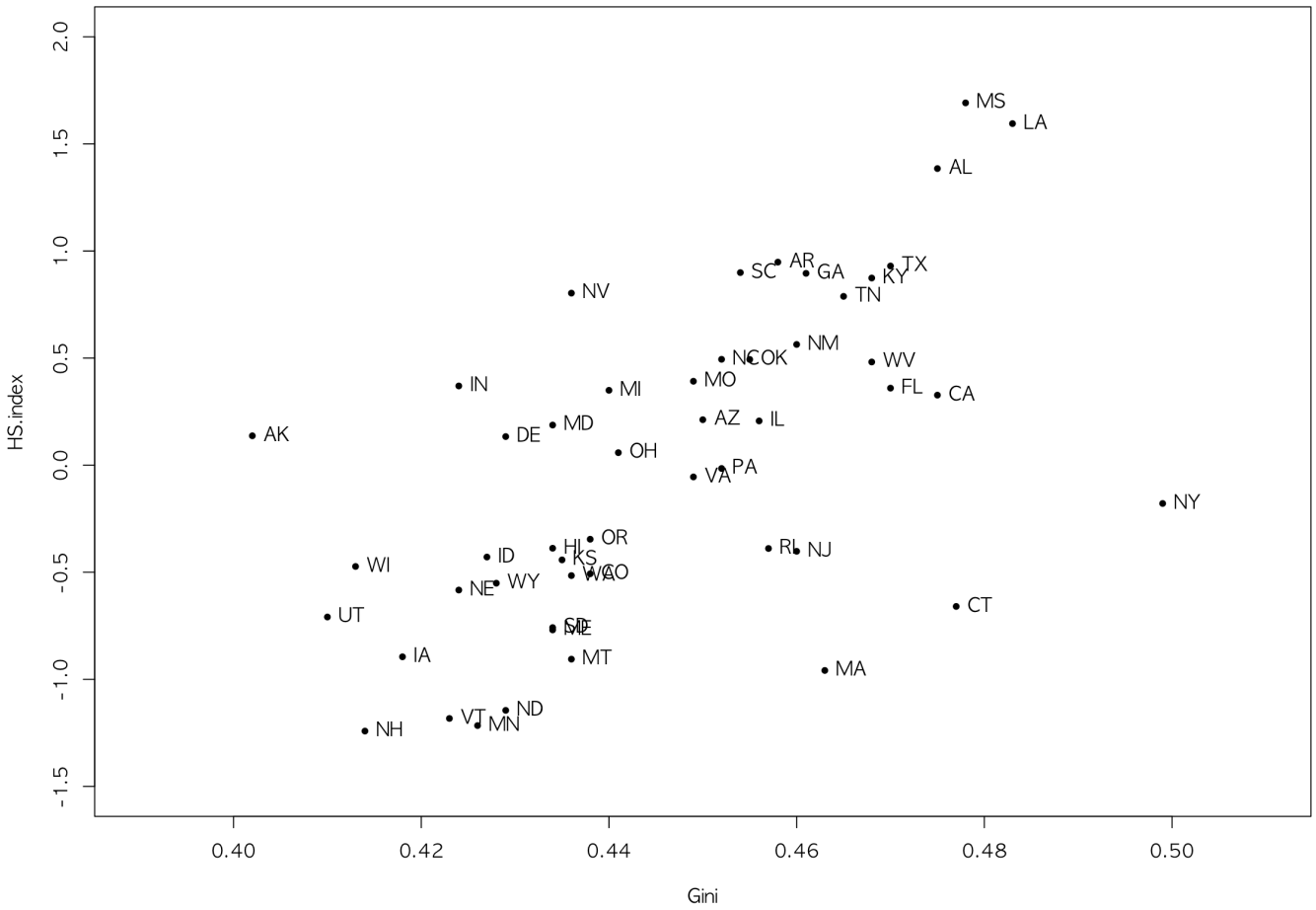
x-축과 y-축의 범위를 설정하고, `pch=20` 으로 다시 그린다.

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



각 주의 약칭을 새겨넣는다.

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



겹쳐보이는 주의 약칭들로부터 인덱스를 추출한다.

```
which(data.usa.1$Abb %in% c("VT", "ME", "NE", "WA", "VA", "HI", "RI", "SC", "A
R", "NC", "GA", "KY"))
```

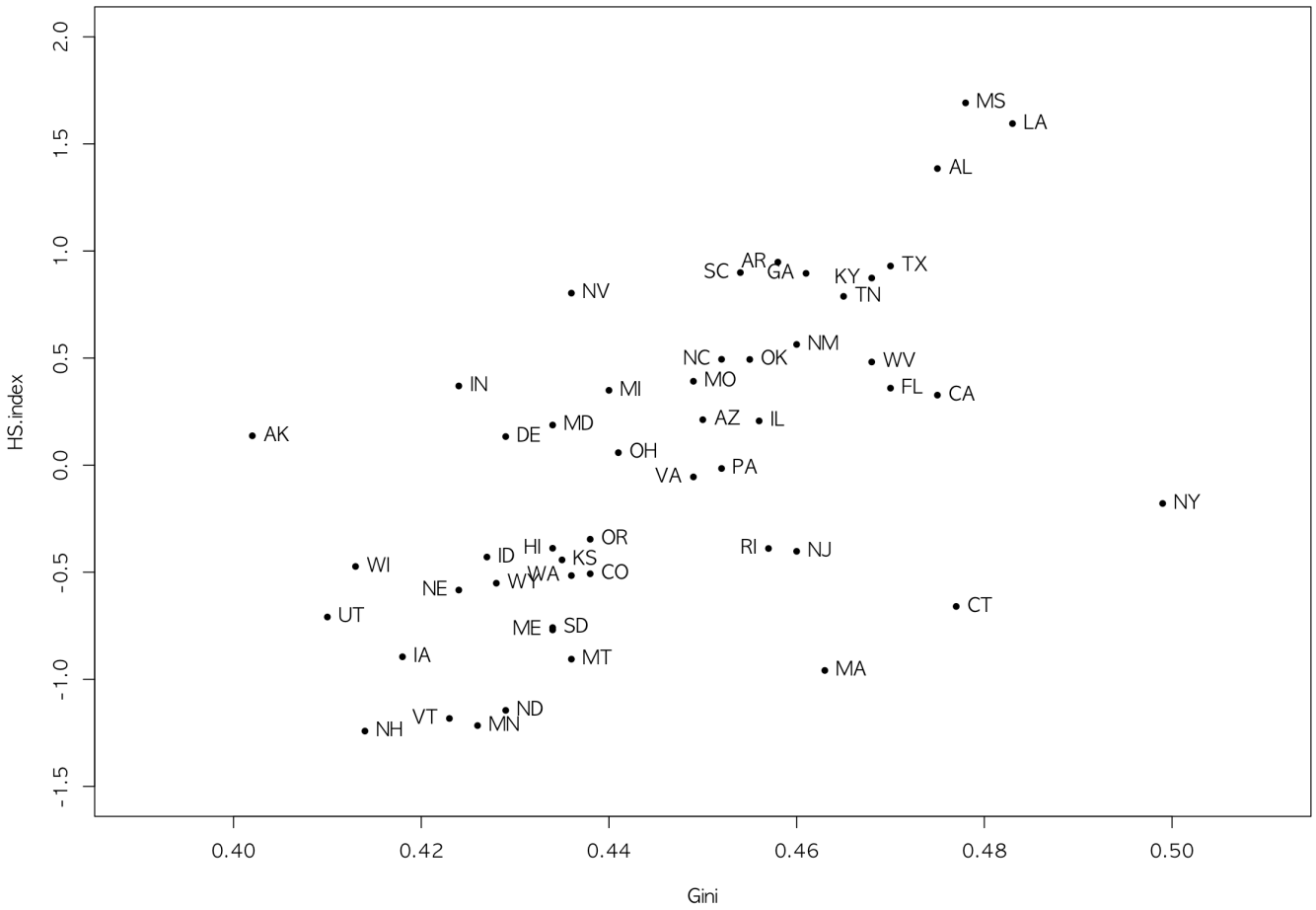
```
##  [1]  4 10 11 17 19 27 33 39 40 45 46 47
```

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

```
text.left<-which(data.usa.1$Abb %in% c("VT", "ME", "NE", "WA", "VA", "HI", "R
I", "SC", "AR", "NC", "GA", "KY"))
text.right<-(1:50)[-text.left]
```

왼쪽, 오른쪽 위치를 조정한 주 약칭을 다시 넣는다.

```
plot(HS.index~Gini, data=data.usa.1, pch=20, xlim=c(0.39, 0.51), ylim=c(-1.5,
2.0))
text(data.usa.1$Gini[text.right], data.usa.1$HS.index[text.right], labels=dat
a.usa.1$Abb[text.right], pos=4)
text(data.usa.1$Gini[text.left], data.usa.1$HS.index[text.left], labels=data.us
a.1$Abb[text.left], pos=2)
```



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

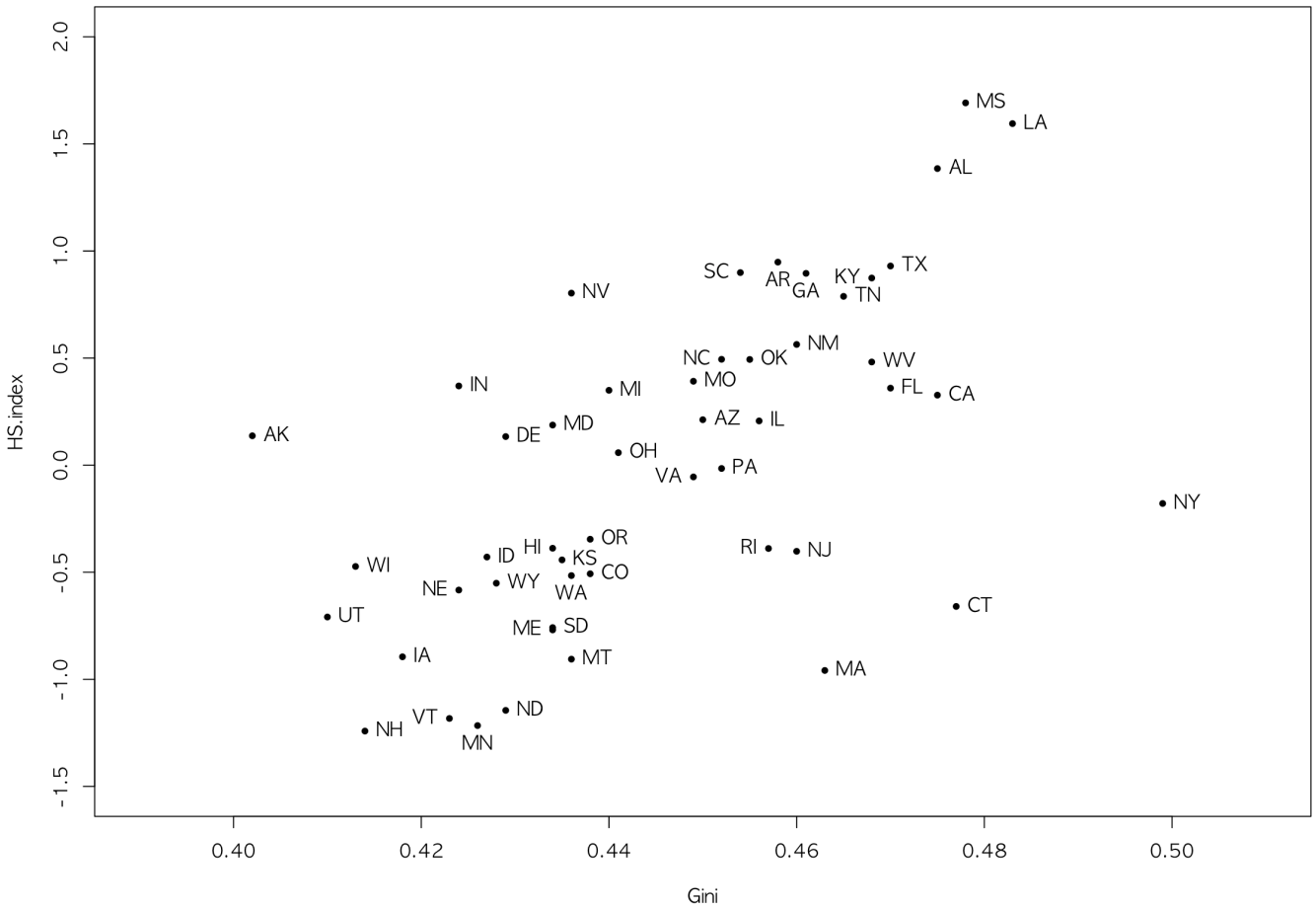
```
text.down<-which(data.usa.1$Abb %in% c("WA", "AR", "GA", "MN"))
which(text.left %in% text.down)
```

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

```
text.left<-text.left[!(text.left %in% text.down)]
text.right<-text.right[!(text.right %in% text.down)]
```

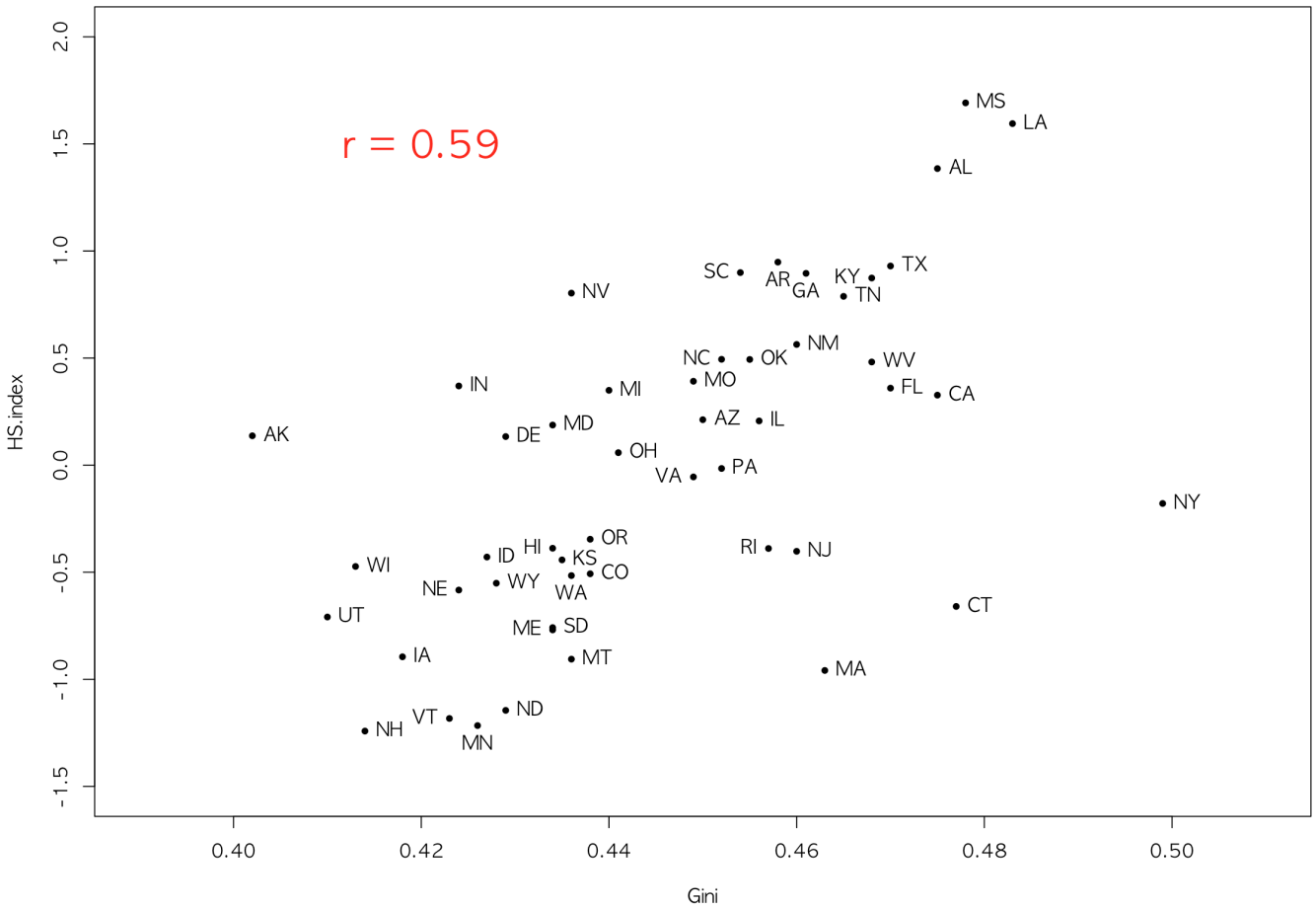
약칭 위치를 아래로 조정한 산점도를 다시 그린다.

```
plot(HS.index~Gini, data=data.usa.1, pch=20, xlim=c(0.39, 0.51), ylim=c(-1.5,
2.0))
text(data.usa.1$Gini[text.right], data.usa.1$HS.index[text.right], labels=dat
a.usa.1$Abb[text.right], pos=4)
text(data.usa.1$Gini[text.left], data.usa.1$HS.index[text.left], labels=data.us
a.1$Abb[text.left], pos=2)
text(data.usa.1$Gini[text.down], data.usa.1$HS.index[text.down], labels=data.us
a.1$Abb[text.down], pos=1)
```



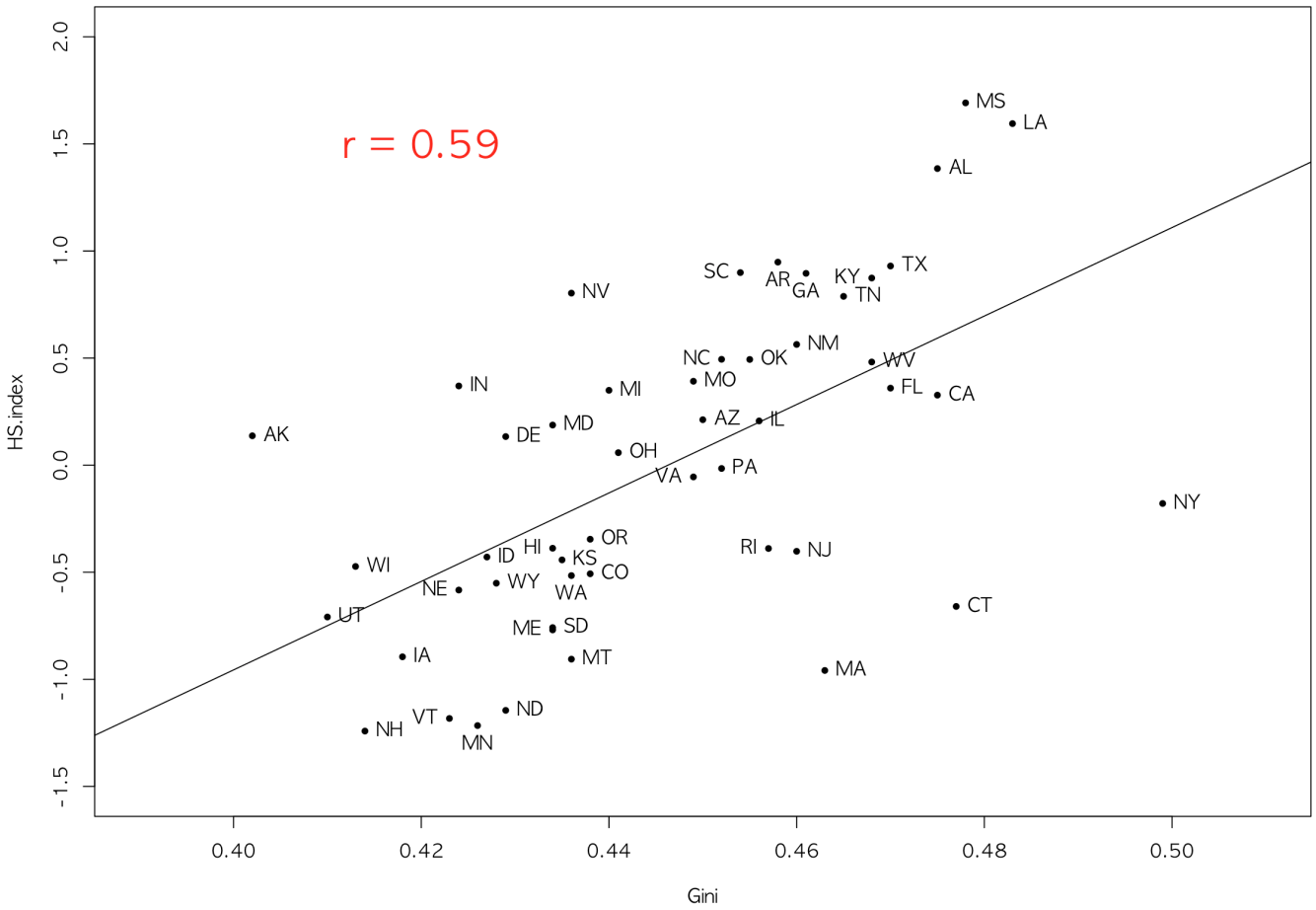
상관계수를 추가한다.

```
plot(HS.index~Gini, data=data.usa.1, pch=20, xlim=c(0.39, 0.51), ylim=c(-1.5,
2.0))
text(data.usa.1$Gini[text.right], data.usa.1$HS.index[text.right], labels=dat
a.usa.1$Abb[text.right], pos=4)
text(data.usa.1$Gini[text.left], data.usa.1$HS.index[text.left], labels=data.us
a.1$Abb[text.left], pos=2)
text(data.usa.1$Gini[text.down], data.usa.1$HS.index[text.down], labels=data.us
a.1$Abb[text.down], pos=1)
text(x=0.42, y=1.5, labels=paste("r =", round(cor(data.usa.1$HS.index, data.us
a.1$Gini), digits=2)), col="red", cex=2)
```



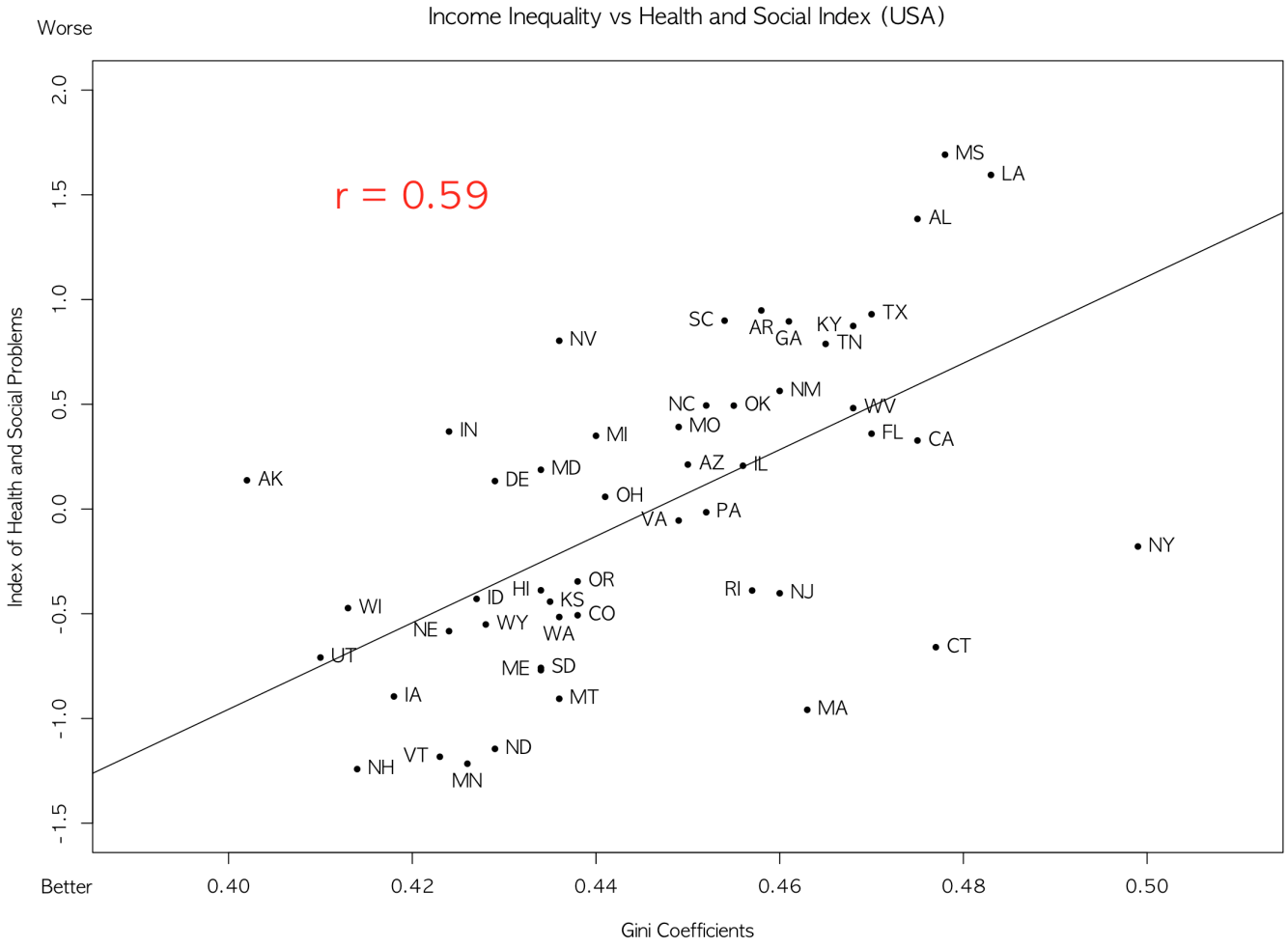
단순회귀선을 추가한다.

```
plot(HS.index~Gini, data=data.usa.1, pch=20, xlim=c(0.39, 0.51), ylim=c(-1.5,
2.0))
text(data.usa.1$Gini[text.right], data.usa.1$HS.index[text.right], labels=dat
a.usa.1$Abb[text.right], pos=4)
text(data.usa.1$Gini[text.left], data.usa.1$HS.index[text.left], labels=data.us
a.1$Abb[text.left], pos=2)
text(data.usa.1$Gini[text.down], data.usa.1$HS.index[text.down], labels=data.us
a.1$Abb[text.down], pos=1)
text(x=0.42, y=1.5, labels=paste("r =", round(cor(data.usa.1$HS.index, data.us
a.1$Gini), digits=2)), col="red", cex=2)
abline(lm(HS.index~Gini, data=data.usa.1)$coefficient)
```



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

```
plot(HS.index~Gini, data=data.usa.1, pch=20, xlim=c(0.39, 0.51), ylim=c(-1.5,
2.0), ann=FALSE)
text(data.usa.1$Gini[text.right], data.usa.1$HS.index[text.right], labels=dat
a.usa.1$Abb[text.right], pos=4)
text(data.usa.1$Gini[text.left], data.usa.1$HS.index[text.left], labels=data.us
a.1$Abb[text.left], pos=2)
text(data.usa.1$Gini[text.down], data.usa.1$HS.index[text.down], labels=data.us
a.1$Abb[text.down], pos=1)
text(x=0.42, y=1.5, labels=paste("r =", round(cor(data.usa.1$HS.index, data.us
a.1$Gini), digits=2)), col="red", cex=2)
abline(lm(HS.index~Gini, data=data.usa.1)$coefficient)
mtext("Worse", side=2, at=2.3, las=1)
mtext("Better", side=2, at=-1.8, las=1)
title(main="Income Inequality vs Health and Social Index (USA)", xlab="Gini Coe
fficients", ylab="Index of Health and Social Problems")
```



뒷 마무리

```
save(file="Income_inequality_vs_health_social_index_USA.rda", list=ls())
```

```
savehistory("Income_inequality_vs_health_social_index_USA.Rhistory")
```