

# Income Inequality vs Index of Health and Social Problems : US Data

coop711

2016-10-07

## Data Preparation

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

```
library(knitr)
# install.packages("xlsx", repos = "https://cran.rstudio.com")
library(xlsx)
```

xlsx 패키지를 이용하여 자료를 읽어들인다.

```
data.usa <- read.xlsx("../data/USA-inequality.xls", 1, stringsAsFactors = FALSE)
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)
```

```

State 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 ...

```

```

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

```

State	State.Abb	Gini	HS.index
Alabama	AL	0.475	1.385
Alaska	AK	0.402	0.137
Arizona	AZ	0.450	0.212
Arkansas	AR	0.458	0.948
California	CA	0.475	0.327
Colorado	CO	0.438	-0.507
Connecticut	CT	0.477	-0.660
Delaware	DE	0.429	0.133
Florida	FL	0.470	0.360
Georgia	GA	0.461	0.896
Hawaii	HI	0.434	-0.388
Idaho	ID	0.427	-0.429
Illinois	IL	0.456	0.206
Indiana	IN	0.424	0.370
Iowa	IA	0.418	-0.895
Kansas	KS	0.435	-0.442
Kentucky	KY	0.468	0.874
Louisiana	LA	0.483	1.595
Maine	ME	0.434	-0.769
Maryland	MD	0.434	0.187
Massachusetts	MA	0.463	-0.959
Michigan	MI	0.440	0.349
Minnesota	MN	0.426	-1.216
Mississippi	MS	0.478	1.692
Missouri	MO	0.449	0.392
Montana	MT	0.436	-0.906

State	State.Abb	Gini	HS.index
Nebraska	NE	0.424	-0.583
Nevada	NV	0.436	0.803
New Hampshire	NH	0.414	-1.242
New Jersey	NJ	0.460	-0.402
New Mexico	NM	0.460	0.564
New York	NY	0.499	-0.179
North Carolina	NC	0.452	0.494
North Dakota	ND	0.429	-1.145
Ohio	OH	0.441	0.058
Oklahoma	OK	0.455	0.494
Oregon	OR	0.438	-0.346
Pennsylvania	PA	0.452	-0.015
Rhode Island	RI	0.457	-0.389
South Carolina	SC	0.454	0.899
South Dakota	SD	0.434	-0.759
Tennessee	TN	0.465	0.788
Texas	TX	0.470	0.930
Utah	UT	0.410	-0.709
Vermont	VT	0.423	-1.183
Virginia	VA	0.449	-0.055
Washington	WA	0.436	-0.516
West Virginia	WV	0.468	0.482
Wisconsin	WI	0.413	-0.473
Wyoming	WY	0.428	-0.551

Save

```
save.image(file = "Inequality_Index_HS_US.RData")
```