Income Inequality vs Index of Health and Social Problems : Data

coop711

2018-04-29

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)를 추가하여 읽어들이면,

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

```
## 'data.frame':    23 obs. of  30 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 ...
##  $ Trust                            : num  39.9 33.9 30.7 38.8 66.5 58 22.2 34.8 23.7 35.2 ...
##  $ Life.expectancy                  : num  79.2 78.5 78.8 79.3 76.6 78 79 78.3 78.3 77 ...
##  $ 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 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 11.8 18.7 ...
##  $ Homicides                        : num  16.9 11.6 13 17.3 12.7 28.2 21.5 13.7 13.9 8.6 ...
##  $ Imprisonment..log.               : num  4.61 4.52 4.28 4.77 4.17 4.11 4.5 4.51 3.33 4.17 ...
##  $ Social.mobility                  : num  NA NA NA 0.14 0.14 0.15 NA 0.17 NA NA ...
##  $ Index.of.health...social_problems: num  0.07 0.01 -0.23 -0.07 -0.19 -0.43 0.05 -0.06 0.38 0.25 ...
##  $ Child.overweight                 : num  NA 11.9 10.4 19.5 10.3 13.3 11.2 11.3 16 12.1 ...
##  $ Drugs.index                      : num  1.71 -0.02 -0.18 0.61 -0.09 -0.88 -0.35 -0.3 -0.99 -0.03 ...
##  $ Calorie.intake                   : int  3142 3753 3632 3167 3405 3197 3576 3395 3687 3656 ...
##  $ Public.health.expenditure        : num  67.9 69.3 71.7 70.8 82.4 75.6 76 74.9 56 76 ...
##  $ Child.wellbeing                  : num  -0.21 -0.07 0.05 0.04 0.21 0.34 -0.17 -0.01 -0.04 -0.04 ...
##  $ Maths.education.science.score    : num  525 496 515 526 494 ...
##  $ Child.conflict                   : num  NA 0.31 0.33 0.24 -0.14 -1.25 0.59 -0.7 0.4 -0.06 ...
##  $ Foreign.aid                      : num  0.25 0.52 0.53 0.34 0.81 0.47 0.47 0.35 0.24 0.41 ...
##  $ 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 1.73 1.52 1.79 1.4 ...
##  $ 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 0.99 1.04 1 ...
##  $ Police                           : int  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  0.46 -0.81 0.61 0.56 0.83 1.08 -0.17 -0.21 -0.85 -0.21 ...
##  $ Lone.parents                     : int  21 15 12 17 22 19 12 21 3 14 ...
##  $ GDP_WB                           : int  45926 47682 43435 45066 45537 40676 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 나라들

을 제외하고 분석작업을 수행하여야 한다. `which()` 를 이용하여 해당 인덱스를 찾고, 나라명을 추출한다.

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

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

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
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_problems", "GDP_WB")
names(data.full) %in% v.names
```

```
##  [1]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
```

```
(v.names.index <- which(names(data.full) %in% v.names))
```

```
## [1]  1  2 13 30
```

```
data.21 <- data.full[-c(11, 18), v.names]
# data.21 <- data.full[-c(11, 18), v.names.index]
str(data.21)
```

```
## '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.of.health...social_problems: 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 49393 ...
```

```
names(data.21)[3] <- "Index.HS"
str(data.21)
```

```
## '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 49393 ...
```

```
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

Save

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