

Income Inequality vs Index of Health and Social Problems : Data with Piping

coop711

2018-05-09

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(magrittr)
library(knitr)
data_full <- "../data/international-inequality_GDP.csv" %>%
  read.csv(stringsAsFactors = FALSE)
data_full %>% names
```

```
## [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"
```

이 자료 중 소득불평등을 나타내는 지표는 5분위배율로서 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.

```
country_na <- data_full %>%
  `[` (., "Index.of.health...social_problems") %>%
  is.na %>%
  which
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.

```
data_full %>%
  names %>%
  `==` ("Index.of.health...social_problems") %>%
  which
```

```
## [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")
v_names_index <- data_full %>%
  names %>%
  `%in%` (v_names) %>%
  which
data_21 <- data_full %>%
  `[` (., -c(11, 18), v_names_index)
data_21 %>% names
```

```
## [1] "Country" "Income.inequality"
## [3] "Index.of.health...social_problems" "GDP_WB"
```

```
names(data_21)[2:3] <- c("Income_inequality", "Index_HS")
data_21 %>% names %>% `[` (2:3)
```

```
## [1] "Income_inequality" "Index_HS"
```

```
data_21 %>% str
```

```
## '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 4
9393 ...
```

```
data_21 %>% kable
```

	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_v2.RData")
```