Poverty and Health Equity

Lujain Alsaleh

Course: PBHL-B481/581

## Primary Aim:

Poverty and health equity is a major health topic. It is important for policy makers to make decisions based on data. My app will provide an interface to health and poverty data and analysis. Using this data might help them to find out what policy will work.

### Analysis work flow:

The data will be taken from the World Data Bank (WDB). Using the R app WDI to download the data as the user request it.

WDB has lots of data, but the app will present the user with a good selection of that data. This will save the user from having explore the thousands of series on the website.

The series can be categorized as follows (The user will choose from options in these categories):

Outcome measure: Outcome measurements related to poverty or economic health such as GINI index.

Intervention: Measurements of government intervention or things can be modified by government policies such as Income share held by fourth 20%. For example, the democrats claim that they can improve this number.

Confounders: Variables that cannot be modified by government policies such as age distribution, or ethnic variety. These will be used to adjust measures of the association between interventions and outcomes.

### Data

The data will be downloaded from World Data Bank in real time.

This website allows you to select data by series year and country.

## Proof of Concept:

The dataset below is one example from Poverty and Equity. This include GINI index and Income share held by fourth 20%. Other series can be selected as needed.

p <- read\_csv("data-raw/76a73c18-13c7-4003-aa16-d497274f8db4\_Data.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_character()  
## )

## See spec(...) for full column specifications.

head(p)

## # A tibble: 6 x 50  
## `Series Name` `Series Code` `Country Name` `Country Code` `1974 [YR1974]`  
## <chr> <chr> <chr> <chr> <chr>   
## 1 GINI index (~ SI.POV.GINI United States USA 35.3   
## 2 GINI index (~ SI.POV.GINI China CHN ..   
## 3 GINI index (~ SI.POV.GINI Sweden SWE ..   
## 4 Income share~ SI.DST.04TH.~ United States USA 23.2   
## 5 Income share~ SI.DST.04TH.~ China CHN ..   
## 6 Income share~ SI.DST.04TH.~ Sweden SWE ..   
## # ... with 45 more variables: `1975 [YR1975]` <chr>, `1976 [YR1976]` <chr>,  
## # `1977 [YR1977]` <chr>, `1978 [YR1978]` <chr>, `1979 [YR1979]` <chr>, `1980  
## # [YR1980]` <chr>, `1981 [YR1981]` <chr>, `1982 [YR1982]` <chr>, `1983  
## # [YR1983]` <chr>, `1984 [YR1984]` <chr>, `1985 [YR1985]` <chr>, `1986  
## # [YR1986]` <chr>, `1987 [YR1987]` <chr>, `1988 [YR1988]` <chr>, `1989  
## # [YR1989]` <chr>, `1990 [YR1990]` <chr>, `1991 [YR1991]` <chr>, `1992  
## # [YR1992]` <chr>, `1993 [YR1993]` <chr>, `1994 [YR1994]` <chr>, `1995  
## # [YR1995]` <chr>, `1996 [YR1996]` <chr>, `1997 [YR1997]` <chr>, `1998  
## # [YR1998]` <chr>, `1999 [YR1999]` <chr>, `2000 [YR2000]` <chr>, `2001  
## # [YR2001]` <chr>, `2002 [YR2002]` <chr>, `2003 [YR2003]` <chr>, `2004  
## # [YR2004]` <chr>, `2005 [YR2005]` <chr>, `2006 [YR2006]` <chr>, `2007  
## # [YR2007]` <chr>, `2008 [YR2008]` <chr>, `2009 [YR2009]` <chr>, `2010  
## # [YR2010]` <chr>, `2011 [YR2011]` <chr>, `2012 [YR2012]` <chr>, `2013  
## # [YR2013]` <chr>, `2014 [YR2014]` <chr>, `2015 [YR2015]` <chr>, `2016  
## # [YR2016]` <chr>, `2017 [YR2017]` <chr>, `2018 [YR2018]` <chr>, `2019  
## # [YR2019]` <chr>

There is also an r app called WDI that allows downloading data. The search function below find series based on key words “Poverty” is an example.

r <- WDI::WDIsearch("poverty")  
r %>% head

## indicator name   
## [1,] "1.0.HCount.1.90usd" "Poverty Headcount ($1.90 a day)"   
## [2,] "1.0.HCount.2.5usd" "Poverty Headcount ($2.50 a day)"   
## [3,] "1.0.HCount.Ofcl" "Official Moderate Poverty Rate-National"  
## [4,] "1.0.HCount.Poor4uds" "Poverty Headcount ($4 a day)"   
## [5,] "1.0.PGap.1.90usd" "Poverty Gap ($1.90 a day)"   
## [6,] "1.0.PGap.2.5usd" "Poverty Gap ($2.50 a day)"

Once the series has beed located the WDI app can be used to download the data as shown below:

#d <- WDI::WDI(indicator= c("SI.POV.GINI" , "SI.DST.04TH.20") )  
d %>%   
 na.omit() %>%   
head()

## iso2c country year SI.POV.GINI SI.DST.04TH.20  
## 421 AE United Arab Emirates 2014 32.5 25.5  
## 586 AL Albania 1996 27.0 23.3  
## 592 AL Albania 2002 31.7 22.2  
## 595 AL Albania 2005 30.6 22.5  
## 598 AL Albania 2008 30.0 22.2  
## 602 AL Albania 2012 29.0 22.8

Here the user will be able to filter the data by country and create a long version of the data set. Both wide and long version are useful for different types of analysis. Here we demonstrate how to change from wide to long format.

dd <- d %>%   
 filter(country %in% c("United States", "China" , "Sweden")) %>%   
 na.omit() %>%   
 pivot\_longer(col = c('SI.POV.GINI','SI.DST.04TH.20'),  
 names\_to = "series",  
 values\_to = "measure")   
dd %>% head

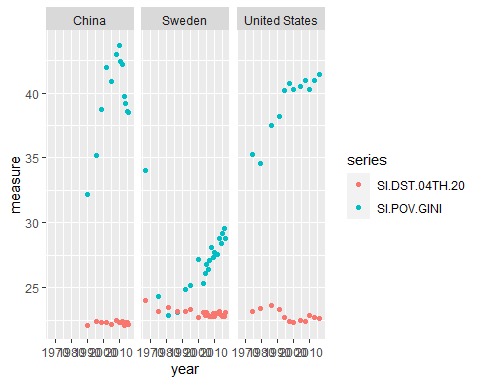
## # A tibble: 6 x 5  
## iso2c country year series measure  
## <chr> <chr> <int> <chr> <dbl>  
## 1 CN China 1990 SI.POV.GINI 32.2  
## 2 CN China 1990 SI.DST.04TH.20 22.1  
## 3 CN China 1996 SI.POV.GINI 35.2  
## 4 CN China 1996 SI.DST.04TH.20 22.4  
## 5 CN China 1999 SI.POV.GINI 38.7  
## 6 CN China 1999 SI.DST.04TH.20 22.3

Save data tables to the package if needed.

#usethis::use\_data(dd,overwrite = TRUE)  
#usethis::use\_data(d,overwrite = TRUE)

Users will be able to select graphs to be used to visualize associations.

dd %>%   
 ggplot(aes(x=year , y=measure , color=series))+  
 geom\_point()+  
 facet\_wrap(~country)

 In addition to plots, users can select tables to be produce summaries using gtsummary or flextable.

dd %>%   
 filter(series == "SI.DST.04TH.20") %>%   
 select(country,measure) %>%   
 tbl\_summary(by='country',statistic = list(measure~"{mean} ({sd})"))

## Table printed with {flextable}, not {gt}. Learn why at  
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

| Characteristic | China, N = 131 | Sweden, N = 221 | United States, N = 121 |
| --- | --- | --- | --- |
| measure | 22.28 (0.11) | 23.06 (0.29) | 22.83 (0.44) |
| 1Statistics presented: Mean (SD) | | | |

In addition, statistical models will also be available.

## Challanges:

1. Finding good data series from WDB website.
2. Identifying which series are outcomes, predictors, intervention and Confounders.
3. Select some good modeling technique.
4. Learn Rshiny.
5. Learn how to deploy Rshiny.

## To do:

1. Find other data series. (10/15/2020)
2. Develop a set of a data exploration summaries and graphs for the user to choose from. (10/22/2020)
3. Develop a selection of statistical models. (10/29/2020)
4. Program interface between Rshiny and WDB. (11/05/2020)
5. Develop a shiny user interface.(11/10/2020)
6. Develop the shiny server code. (11/18/2020)
7. Deploy the shiny app. (11/19/2020)