Exposure and Fixed-Effects modelling

# Introduction

This aim of this research is to build a series of models which include the following response variables:

1. Male age-standardised mortality rates (all-cause)
2. Female age-standardised mortality rates (all-cause)
3. Relative difference in all-cause rates (1 / 2)
4. Absolute difference in all-cause rates (1 - 2)
5. Male age-standardised mortality rates (NCDs)
6. Female age-standardised mortality rates (NCDs)
7. Relative difference in NCD rates (5 / 6)
8. Absolute differnece in NCD rates (5 - 6)

This will be done for each country (level 3 in location hierarchy).

And a series of exposures from the World Bank will be used:

1. GDP per cap by purchasing power parity
2. Male literacy rates
3. Female literacy rates
4. etc etc

# Loading GBD data

* Source is [here](http://s3.healthdata.org/gbd-api-2016-production/3f1a784fe33bf6d9447c7b90e8828c3f_files/IHME-GBD_2016_DATA-3f1a784f-1.zip)

require(tidyverse)

## Loading required package: tidyverse

## -- Attaching packages --------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

source("scripts/download\_completed\_request.R")

Will read\_csv\_directly work?

asr\_dta <- read\_csv\_directly(  
 url\_loc = "http://s3.healthdata.org/gbd-api-2016-production/3f1a784fe33bf6d9447c7b90e8828c3f\_files/IHME-GBD\_2016\_DATA-3f1a784f-1.zip",  
 csv\_name = "IHME-GBD\_2016\_DATA-3f1a784f-1.csv")

*Note:* Though read\_csv\_directly() works as a function, the links are not permanent, so downloading them is preferable. The aim should be to reproduce the above and check with the online webpage to confirm.

download\_completed\_request("http://s3.healthdata.org/gbd-api-2016-production/67c84b9e938574bd56e66ab1c58653db\_files",  
"IHME-GBD\_2016\_DATA-67c84b9e-",  
"data/age\_standardised\_allcause\_ncd/"  
 )

asr\_dta <- read\_csv("data/age\_standardised\_allcause\_ncd/1.csv")

## Parsed with column specification:  
## cols(  
## measure\_id = col\_integer(),  
## measure\_name = col\_character(),  
## location\_id = col\_integer(),  
## location\_name = col\_character(),  
## sex\_id = col\_integer(),  
## sex\_name = col\_character(),  
## age\_id = col\_integer(),  
## age\_name = col\_character(),  
## cause\_id = col\_integer(),  
## cause\_name = col\_character(),  
## rei\_id = col\_integer(),  
## rei\_name = col\_character(),  
## metric\_id = col\_integer(),  
## metric\_name = col\_character(),  
## year = col\_integer(),  
## val = col\_double(),  
## upper = col\_double(),  
## lower = col\_double()  
## )

Now to link this to the hierarchy

gbd\_locations <- readxl::read\_excel("raw\_data/IHME\_GBD\_2016\_CODEBOOK/IHME\_GBD\_2016\_GBD\_LOCATION\_HIERARCHY\_Y2018M05D23.XLSX")  
  
gbd\_country\_ids <- gbd\_locations %>%   
 filter(location\_id %in% 1:218) %>%   
 filter(level == 3) %>%   
 pull(location\_id)  
  
asr\_dta %>%   
 filter(location\_id %in% gbd\_country\_ids) %>%   
 filter(measure\_name == "Deaths") %>%   
 select(location = location\_name, year, sex = sex\_name, cause = cause\_name, std\_mort\_rate = val) -> mort\_rate\_data  
  
asr\_dta %>%   
 filter(location\_id %in% gbd\_country\_ids) %>%   
 filter(measure\_name == "DALYs (Disability-Adjusted Life Years)") %>%   
 select(location = location\_name, year, sex = sex\_name, cause = cause\_name, std\_mort\_rate = val) -> daly\_rate\_data

And now to join this to the world bank measures

wb\_data <- read\_rds("data/world\_bank/extracted\_indicators.RData")  
nrow(wb\_data)

## [1] 58414

gdp\_percap\_data <- wbstats::wb(indicator= "NY.GDP.PCAP.PP.CD") %>% as\_tibble()   
nrow(gdp\_percap\_data)

## [1] 6405

wb\_data <- bind\_rows(wb\_data, gdp\_percap\_data)  
rm(gdp\_percap\_data)  
nrow(wb\_data)

## [1] 64819

join to mort rate

mort\_rate\_data\_joined <- mort\_rate\_data %>%  
 mutate(country = location) %>%   
 inner\_join(  
 wb\_data %>%   
 mutate(year = as.integer(date)) %>%   
 select(year, country, indicator\_id = indicatorID, indicator, indicator\_value = value)   
 )

## Joining, by = c("year", "country")

daly\_rate\_data\_joined <- daly\_rate\_data %>%  
 mutate(country = location) %>%   
 inner\_join(  
 wb\_data %>%   
 mutate(year = as.integer(date)) %>%   
 select(year, country, indicator\_id = indicatorID, indicator, indicator\_value = value)   
 )

## Joining, by = c("year", "country")

Models by gender and indicator

pull\_indicator\_coeff <- function(x){x %>% filter(term == "indicator\_value") %>% pull(estimate) }  
pull\_indicator\_se <- function(x){x %>% filter(term == "indicator\_value") %>% pull(std.error) }  
  
simple\_models <- mort\_rate\_data\_joined %>%   
 filter(cause == "All causes") %>%   
 select(country, year, sex, std\_mort\_rate, indicator\_id, indicator, indicator\_value) %>%   
 filter(!is.na(indicator\_value)) %>%   
 group\_by(indicator\_id) %>%   
 nest() %>%   
 mutate(male\_model = map(  
 data,  
 function(x) {  
 x %>%  
 filter(sex == "Male") %>%  
 lm(std\_mort\_rate ~ year + indicator\_value, data = .)   
 }  
 )  
 ) %>%  
 mutate(female\_model = map(  
 data,   
 function(x) {  
 x %>%   
 filter(sex == "Female") %>%   
 lm(std\_mort\_rate ~ year + indicator\_value, data = .)   
 }  
 )  
 ) %>%   
 mutate(  
 male\_model\_tidy = map(male\_model, broom::tidy),  
 female\_model\_tidy = map(female\_model, broom::tidy)   
 ) %>%   
 mutate(  
 male\_indicator\_coeff = map\_dbl(male\_model\_tidy, pull\_indicator\_coeff),  
 female\_indicator\_coeff = map\_dbl(female\_model\_tidy, pull\_indicator\_coeff),  
 male\_indicator\_se = map\_dbl(male\_model\_tidy, pull\_indicator\_se),  
 female\_indicator\_se = map\_dbl(female\_model\_tidy, pull\_indicator\_se)  
 )

Note: for some of these indicators the log should probably be used

simple\_models %>%   
 select(indicator\_id, male\_indicator\_coeff:female\_indicator\_se) %>%   
 left\_join(wb\_data %>% select(indicator\_id = indicatorID, indicator) %>% distinct) %>%  
 mutate(  
 male\_t = male\_indicator\_coeff / male\_indicator\_se,   
 female\_t = female\_indicator\_coeff / female\_indicator\_se  
 ) %>%   
 select(indicator\_id, indicator, male\_t, female\_t) %>%   
 arrange(desc(abs(male\_t))) %>%  
 mutate(  
 male\_rank = 1 + length(male\_t) - rank(abs(male\_t)),   
 female\_rank = 1 + length(female\_t) - rank(abs(female\_t))) %>%   
 filter(abs(male\_t) > 2 | abs(female\_t) > 2) %>%  
 mutate(male\_t = round(male\_t, 2), female\_t = round(female\_t, 2)) %>%   
 mutate(diff = male\_rank - female\_rank) %>%   
 knitr::kable(caption = "Signal strength of indicators on male and female age-standardised all-cause mortality trends")

## Joining, by = "indicator\_id"

Signal strength of indicators on male and female age-standardised all-cause mortality trends

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | male\_t | female\_t | male\_rank | female\_rank | diff |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | -18.36 | -22.30 | 1 | 2 | -1 |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | -17.89 | -25.56 | 2 | 1 | 1 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | -17.53 | -18.66 | 3 | 3 | 0 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | -8.88 | -9.68 | 4 | 8 | -4 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | -7.20 | -12.69 | 5 | 4 | 1 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | -7.12 | -8.00 | 6 | 9 | -3 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | -6.68 | -9.83 | 7 | 7 | 0 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | -5.81 | -11.40 | 8 | 5 | 3 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | -5.04 | -10.60 | 9 | 6 | 3 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | -4.61 | -7.60 | 10 | 10 | 0 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | 4.06 | 6.14 | 11 | 11 | 0 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | -3.52 | -5.40 | 12 | 13 | -1 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | 3.05 | 4.66 | 13 | 17 | -4 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | -3.00 | -5.28 | 14 | 14 | 0 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -2.93 | -1.82 | 15 | 24 | -9 |
| UNDP.HDI.XD | Human development index (HDI) | -2.79 | -5.27 | 16 | 15 | 1 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | 2.75 | 5.79 | 17 | 12 | 5 |
| 5.51.01.07.gender | Gender equality in education | -2.27 | -3.40 | 18 | 18 | 0 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | 2.10 | 5.25 | 19 | 16 | 3 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | -1.69 | -2.71 | 20 | 20 | 0 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | 1.61 | 2.90 | 21 | 19 | 2 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | -1.43 | -2.71 | 23 | 21 | 2 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 0.32 | -2.20 | 31 | 22 | 9 |

The above shows the ‘signal strength’ of the different indicators once year is controlled for. For both genders GDP per capital has the strongest signal, followed by gender parity in primary school enrollment. For males, age-standardised mortality rates are then most strongly predicted by the female urban population (% of total), followed by male urban population as % of total. For females, the third strongest signal is gender parity in primary and secondary school enrollment, followed by gender parity in secondary school (only) enrollment rates.

# As above, but CVD only

produce\_table <- function(DTA, this\_cause, this\_caption){  
  
 DTA %>%   
 filter(cause == this\_cause) %>%   
 select(country, year, sex, std\_mort\_rate, indicator\_id, indicator, indicator\_value) %>%  
 filter(!is.na(indicator\_value)) %>%  
 group\_by(indicator\_id) %>%  
 nest() %>%  
 mutate(male\_model = map(  
 data,  
 function(x) {  
 x %>%  
 filter(sex == "Male") %>%  
 lm(std\_mort\_rate ~ year + indicator\_value, data = .)  
 }  
 )  
 ) %>%  
 mutate(female\_model = map(  
 data,  
 function(x) {  
 x %>%  
 filter(sex == "Female") %>%  
 lm(std\_mort\_rate ~ year + indicator\_value, data = .)  
 }  
 )  
 ) %>%  
 mutate(  
 male\_model\_tidy = map(male\_model, broom::tidy),  
 female\_model\_tidy = map(female\_model, broom::tidy)  
 ) %>%  
 mutate(  
 male\_indicator\_coeff = map\_dbl(male\_model\_tidy, pull\_indicator\_coeff),  
 female\_indicator\_coeff = map\_dbl(female\_model\_tidy, pull\_indicator\_coeff),  
 male\_indicator\_se = map\_dbl(male\_model\_tidy, pull\_indicator\_se),  
 female\_indicator\_se = map\_dbl(female\_model\_tidy, pull\_indicator\_se)  
 ) %>%  
 select(indicator\_id, male\_indicator\_coeff:female\_indicator\_se) %>%  
 left\_join(wb\_data %>% select(indicator\_id = indicatorID, indicator) %>% distinct) %>%  
 mutate(  
 male\_t = male\_indicator\_coeff / male\_indicator\_se,  
 female\_t = female\_indicator\_coeff / female\_indicator\_se  
 ) %>%  
 select(indicator\_id, indicator, male\_t, female\_t) %>%  
 arrange(desc(abs(male\_t))) %>%  
 mutate(  
 male\_rank = 1 + length(male\_t) - rank(abs(male\_t)),  
 female\_rank = 1 + length(female\_t) - rank(abs(female\_t))) %>%  
 filter(abs(male\_t) > 2 | abs(female\_t) > 2) %>%  
 mutate(male\_t = round(male\_t, 2), female\_t = round(female\_t, 2)) %>%  
 mutate(diff = male\_rank - female\_rank) %>%  
 knitr::kable(caption = this\_caption)  
  
   
}

mort\_rate\_data\_joined %>%   
 produce\_table(  
 this\_cause = "Non-communicable diseases",   
 this\_caption = "Signal strength of indicators on male and female age-standardised NCD mortality trends"  
 )

## Joining, by = "indicator\_id"

Signal strength of indicators on male and female age-standardised NCD mortality trends

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | male\_t | female\_t | male\_rank | female\_rank | diff |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | -8.71 | -16.26 | 1 | 2 | -1 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | -6.06 | -10.32 | 2 | 4 | -2 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 4.99 | -3.18 | 3 | 14 | -11 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | -4.89 | -8.41 | 4 | 5 | -1 |
| UNDP.HDI.XD | Human development index (HDI) | 4.71 | -1.99 | 5 | 18 | -13 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 4.42 | -4.95 | 6 | 8 | -2 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | -4.37 | -14.85 | 7 | 3 | 4 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 4.11 | -5.22 | 8 | 7 | 1 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 3.30 | -6.50 | 9 | 6 | 3 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -3.28 | -0.33 | 10 | 28 | -18 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -3.09 | 0.34 | 11 | 27 | -16 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 2.97 | -1.07 | 12 | 23 | -11 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 2.86 | -1.00 | 13 | 24 | -11 |
| SG.COK.CROP.ZS | Main cooking fuel: agricultural crop (% of households) | 2.74 | 1.20 | 14 | 22 | -8 |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | -2.72 | -16.98 | 15 | 1 | 14 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 2.54 | 0.09 | 16 | 33 | -17 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 2.45 | -4.00 | 17 | 10 | 7 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -2.12 | 1.70 | 18 | 20 | -2 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | 2.07 | -0.83 | 19 | 25 | -6 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | 2.07 | -0.20 | 20 | 31 | -11 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 1.99 | -2.48 | 21 | 15 | 6 |
| 5.51.01.07.gender | Gender equality in education | -1.46 | -4.14 | 22 | 9 | 13 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 1.33 | -2.43 | 24 | 16 | 8 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -1.21 | 3.89 | 25 | 11 | 14 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -1.20 | 3.62 | 26 | 12 | 14 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -0.76 | 2.22 | 28 | 17 | 11 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | -0.21 | -3.20 | 33 | 13 | 20 |

Gender differences in the signal strength of these indicators appear much larger for NCD mortality than all cause mortality. Note also that the direction of effects goes in the opposite direction for some indicators, though only for the Mo Ibrahim index are the effects in opposite directions and the t values > 2.

The rank importance of access to anti-retroviral drugs is increased, to 2nd and 3rd place after GDP per capita.

# Signal strength, all cause, DALY

Now to look at the equivalent for DALYs

daly\_rate\_data\_joined %>%   
 produce\_table(  
 this\_cause = "All causes",   
 this\_caption = "Signal strength of indicators on male and female age-standardised all-cause DALY rate trends")

## Joining, by = "indicator\_id"

Signal strength of indicators on male and female age-standardised all-cause DALY rate trends

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | male\_t | female\_t | male\_rank | female\_rank | diff |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | -21.71 | -24.90 | 1 | 1 | 0 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | -21.12 | -21.90 | 2 | 2 | 0 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | -17.93 | -17.15 | 3 | 3 | 0 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | -11.70 | -14.84 | 4 | 4 | 0 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | -10.17 | -13.38 | 5 | 5 | 0 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | -9.29 | -10.66 | 6 | 7 | -1 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | -9.11 | -12.14 | 7 | 6 | 1 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | -8.84 | -8.70 | 8 | 8 | 0 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | -7.17 | -7.20 | 9 | 10 | -1 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | -6.30 | -7.83 | 10 | 9 | 1 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | 5.39 | 6.28 | 11 | 11 | 0 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | -5.09 | -6.07 | 12 | 12 | 0 |
| UNDP.HDI.XD | Human development index (HDI) | -4.73 | -5.84 | 13 | 15 | -2 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | -4.65 | -5.91 | 14 | 14 | 0 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | 4.58 | 5.31 | 15 | 16 | -1 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | 4.37 | 5.92 | 16 | 13 | 3 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | 3.56 | 5.23 | 17 | 17 | 0 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | 2.89 | 3.56 | 18 | 18 | 0 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -2.85 | -2.07 | 19 | 23 | -4 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | -2.43 | -3.05 | 20 | 19 | 1 |
| 5.51.01.07.gender | Gender equality in education | -2.10 | -2.55 | 21 | 21 | 0 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | -1.95 | -2.56 | 22 | 20 | 2 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | -1.92 | -2.31 | 23 | 22 | 1 |

For DALY rate trends, the signal strength of school enrollment gender parity is stronger for females than that of GDP per capita.

Now, finally, to look at DALY rates related to NCDs

daly\_rate\_data\_joined %>%   
 produce\_table(  
 this\_cause = "Non-communicable diseases",   
 this\_caption = "Signal strength of indicators on male and female age-standardised NCD DALY rate trends")

## Joining, by = "indicator\_id"

Signal strength of indicators on male and female age-standardised NCD DALY rate trends

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | male\_t | female\_t | male\_rank | female\_rank | diff |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | -10.28 | -15.74 | 1 | 3 | -2 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | -7.19 | -10.19 | 2 | 4 | -2 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | -6.53 | -15.81 | 3 | 2 | 1 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | -5.84 | -8.17 | 4 | 5 | -1 |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | -5.04 | -18.61 | 5 | 1 | 4 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 4.07 | -3.62 | 6 | 13 | -7 |
| UNDP.HDI.XD | Human development index (HDI) | 3.69 | -2.65 | 7 | 17 | -10 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 3.01 | -5.88 | 8 | 8 | 0 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -2.80 | 0.16 | 9 | 32 | -23 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 2.63 | -6.24 | 10 | 7 | 3 |
| SG.COK.CROP.ZS | Main cooking fuel: agricultural crop (% of households) | 2.60 | 1.05 | 11 | 26 | -15 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 2.53 | -1.95 | 12 | 21 | -9 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -2.31 | 1.05 | 13 | 27 | -14 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 2.13 | -1.84 | 14 | 22 | -8 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 2.08 | -0.52 | 15 | 29 | -14 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 1.67 | -7.79 | 17 | 6 | 11 |
| 5.51.01.07.gender | Gender equality in education | -1.54 | -3.94 | 19 | 12 | 7 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -1.50 | 2.01 | 20 | 19 | 1 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 1.40 | -4.64 | 22 | 9 | 13 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 1.18 | -3.07 | 23 | 15 | 8 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -0.67 | 4.53 | 25 | 10 | 15 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -0.63 | 4.30 | 27 | 11 | 16 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 0.56 | -2.99 | 28 | 16 | 12 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -0.52 | 2.58 | 29 | 18 | 11 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | -0.49 | -3.45 | 30 | 14 | 16 |

Let’s now try to regress this on abs or rel difference

diffs\_mort\_joined <- mort\_rate\_data\_joined %>%   
 select(location, year, sex, cause, std\_mort\_rate) %>%   
 distinct() %>%   
 spread(sex, std\_mort\_rate) %>%   
 mutate(rel = Male / Female, abs = Male - Female, logrel = log(rel)) %>%   
 select(location, year, cause, rel, abs, logrel) %>%   
 mutate(country = location) %>%   
 inner\_join(  
 wb\_data %>%   
 mutate(year = as.integer(date)) %>%   
 select(year, country, indicator\_id = indicatorID, indicator, indicator\_value = value)   
 )

## Joining, by = c("year", "country")

diffs\_daly\_joined <- daly\_rate\_data\_joined %>%   
 select(location, year, sex, cause, std\_mort\_rate) %>%   
 distinct() %>%   
 spread(sex, std\_mort\_rate) %>%   
 mutate(rel = Male / Female, abs = Male - Female, logrel = log(rel)) %>%   
 select(location, year, cause, rel, abs, logrel) %>%   
 mutate(country = location) %>%   
 inner\_join(  
 wb\_data %>%   
 mutate(year = as.integer(date)) %>%   
 select(year, country, indicator\_id = indicatorID, indicator, indicator\_value = value)   
 )

## Joining, by = c("year", "country")

Now produce similar tables as above

produce\_diff\_table <- function(DTA, this\_cause, this\_caption){  
  
 DTA %>%   
 gather(rel:logrel, key = "outcome", value = "outcome\_value") %>%   
 filter(cause == this\_cause) %>%   
 select(country, year, outcome, outcome\_value, indicator\_id, indicator, indicator\_value) %>%  
 filter(!is.na(indicator\_value)) %>%  
 group\_by(indicator\_id) %>%  
 nest() %>%  
 mutate(rel\_model = map(  
 data,  
 function(x) {  
 x %>%  
 filter(outcome == "rel") %>%  
 lm(outcome\_value ~ year + indicator\_value, data = .)  
 }  
 )  
 ) %>%  
 mutate(abs\_model = map(  
 data,  
 function(x) {  
 x %>%  
 filter(outcome == "abs") %>%  
 lm(outcome\_value ~ year + indicator\_value, data = .)  
 }  
 )  
 ) %>%  
 mutate(  
 rel\_model\_tidy = map(rel\_model, broom::tidy),  
 abs\_model\_tidy = map(abs\_model, broom::tidy)  
 ) %>%  
 mutate(  
 rel\_indicator\_coeff = map\_dbl(rel\_model\_tidy, pull\_indicator\_coeff),  
 abs\_indicator\_coeff = map\_dbl(abs\_model\_tidy, pull\_indicator\_coeff),  
 rel\_indicator\_se = map\_dbl(rel\_model\_tidy, pull\_indicator\_se),  
 abs\_indicator\_se = map\_dbl(abs\_model\_tidy, pull\_indicator\_se)  
 ) %>%  
 select(indicator\_id, rel\_indicator\_coeff:abs\_indicator\_se) %>%  
 left\_join(wb\_data %>% select(indicator\_id = indicatorID, indicator) %>% distinct) %>%  
 mutate(  
 rel\_t = rel\_indicator\_coeff / rel\_indicator\_se,  
 abs\_t = abs\_indicator\_coeff / abs\_indicator\_se  
 ) %>%  
 select(indicator\_id, indicator, rel\_t, abs\_t) %>%  
 arrange(desc(abs(rel\_t))) %>%  
 mutate(  
 rel\_rank = 1 + length(rel\_t) - rank(abs(rel\_t)),  
 abs\_rank = 1 + length(abs\_t) - rank(abs(abs\_t))) %>%  
 filter(abs(rel\_t) > 2 | abs(abs\_t) > 2) %>%  
 mutate(rel\_t = round(rel\_t, 2), abs\_t = round(abs\_t, 2)) %>%  
 knitr::kable(caption = this\_caption)  
  
   
}

diffs\_mort\_joined %>%   
 produce\_diff\_table(  
 this\_cause = "All causes",   
 this\_caption = "Signal strength of indicators: gender inequalities in health; mort due to all causes")

## Joining, by = "indicator\_id"

Signal strength of indicators: gender inequalities in health; mort due to all causes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | rel\_t | abs\_t | rel\_rank | abs\_rank |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | 21.16 | 8.54 | 1 | 4 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | 13.77 | 4.36 | 2 | 11 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 12.79 | 10.99 | 3 | 3 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | 12.70 | 0.47 | 4 | 32 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 12.67 | 11.75 | 5 | 2 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 11.89 | 12.08 | 6 | 1 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 10.16 | 7.67 | 7 | 5 |
| UNDP.HDI.XD | Human development index (HDI) | 9.85 | 4.78 | 8 | 10 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 7.81 | 3.16 | 9 | 16 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | 7.53 | 0.48 | 10 | 31 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -7.29 | -6.12 | 11 | 7 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 6.99 | 5.25 | 12 | 9 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -6.72 | -5.26 | 13 | 8 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | 6.59 | 0.93 | 14 | 30 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 6.45 | 7.42 | 15 | 6 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 5.91 | 3.90 | 16 | 12 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -5.43 | -2.62 | 17 | 23 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 5.10 | 3.30 | 18 | 15 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -4.84 | -2.92 | 19 | 21 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | 4.59 | 3.10 | 20 | 17 |
| SL.UEM.1524.FM.NE.ZS | Ratio of female to male youth unemployment rate (%) (national estimate) | -4.58 | -2.97 | 21 | 20 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -4.56 | -3.05 | 22 | 18 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | 4.19 | 3.38 | 23 | 14 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 3.44 | 3.63 | 24 | 13 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -3.34 | -2.78 | 25 | 22 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -3.04 | -3.02 | 26 | 19 |
| SG.DMK.DPCH.FN.ZS | Women participating in making daily purchase decisions (% of women age 15-49) | 2.54 | 1.34 | 27 | 29 |
| 5.51.01.07.gender | Gender equality in education | 2.47 | 2.16 | 28 | 25 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | 2.40 | 1.58 | 29 | 26 |
| SG.COK.DUNG.ZS | Main cooking fuel: dung (% of households) | -1.10 | -2.18 | 31 | 24 |

diffs\_mort\_joined %>%   
 produce\_diff\_table(  
 this\_cause = "Non-communicable diseases",   
 this\_caption = "Signal strength of indicators: gender inequalities in health; mort due to NCDs")

## Joining, by = "indicator\_id"

Signal strength of indicators: gender inequalities in health; mort due to NCDs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | rel\_t | abs\_t | rel\_rank | abs\_rank |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | 21.28 | 13.45 | 1 | 3 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | 14.37 | 8.82 | 2 | 6 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | 13.28 | 4.74 | 3 | 16 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 13.13 | 13.77 | 4 | 2 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 13.12 | 13.38 | 5 | 4 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 12.65 | 13.80 | 6 | 1 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 10.69 | 10.91 | 7 | 5 |
| UNDP.HDI.XD | Human development index (HDI) | 9.31 | 8.55 | 8 | 7 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 8.92 | 6.36 | 9 | 12 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | 7.78 | 2.22 | 10 | 27 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 7.46 | 7.58 | 11 | 8 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -7.41 | -6.99 | 12 | 10 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -6.96 | -6.56 | 13 | 11 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -6.72 | -5.16 | 14 | 14 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | 6.63 | 2.02 | 15 | 29 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 6.13 | 7.50 | 16 | 9 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 5.79 | 5.75 | 17 | 13 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 5.25 | 4.78 | 18 | 15 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -4.89 | -4.62 | 19 | 17 |
| SL.UEM.1524.FM.NE.ZS | Ratio of female to male youth unemployment rate (%) (national estimate) | -4.81 | -3.31 | 20 | 20 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -4.62 | -3.87 | 21 | 18 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | 4.32 | 3.79 | 22 | 19 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | 3.47 | 3.07 | 23 | 24 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 3.33 | 3.18 | 24 | 22 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -3.14 | -2.19 | 25 | 28 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -3.08 | -3.28 | 26 | 21 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | 3.01 | 3.16 | 27 | 23 |
| 5.51.01.07.gender | Gender equality in education | 2.79 | 2.54 | 28 | 26 |
| SG.DMK.DPCH.FN.ZS | Women participating in making daily purchase decisions (% of women age 15-49) | 2.25 | 1.68 | 29 | 31 |
| SG.COK.CROP.ZS | Main cooking fuel: agricultural crop (% of households) | 2.04 | 2.73 | 30 | 25 |

diffs\_daly\_joined %>%   
 produce\_diff\_table(  
 this\_cause = "All causes",   
 this\_caption = "Signal strength of indicators: gender inequalities in health; DALYs due to all causes")

## Joining, by = "indicator\_id"

Signal strength of indicators: gender inequalities in health; DALYs due to all causes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | rel\_t | abs\_t | rel\_rank | abs\_rank |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | 19.67 | 1.71 | 1 | 20 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 12.50 | 6.04 | 2 | 4 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | 12.35 | -1.93 | 3 | 17 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 12.29 | 6.70 | 4 | 2 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 11.32 | 6.87 | 5 | 1 |
| UNDP.HDI.XD | Human development index (HDI) | 10.55 | 1.51 | 6 | 22 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | 10.07 | -5.69 | 7 | 5 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 9.07 | 4.37 | 8 | 6 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 8.12 | 6.37 | 9 | 3 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 7.70 | 0.68 | 10 | 30 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -7.20 | -3.62 | 11 | 7 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 6.92 | 1.56 | 12 | 21 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 6.55 | 2.45 | 13 | 13 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -6.48 | -2.78 | 14 | 9 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | 5.89 | -2.36 | 15 | 14 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 5.38 | 1.09 | 16 | 26 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -5.32 | -1.37 | 17 | 24 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -5.02 | -0.33 | 18 | 32 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | 4.89 | -1.48 | 19 | 23 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -4.73 | -1.86 | 20 | 19 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | 4.68 | 1.89 | 21 | 18 |
| SL.UEM.1524.FM.NE.ZS | Ratio of female to male youth unemployment rate (%) (national estimate) | -4.21 | -2.32 | 22 | 15 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | 3.49 | 2.04 | 23 | 16 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -3.48 | -3.34 | 24 | 8 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 3.35 | 2.73 | 25 | 10 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -2.86 | -2.59 | 26 | 12 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | 2.78 | 0.86 | 27 | 29 |
| SG.DMK.DPCH.FN.ZS | Women participating in making daily purchase decisions (% of women age 15-49) | 2.44 | 0.54 | 28 | 31 |
| 5.51.01.07.gender | Gender equality in education | 2.20 | 1.22 | 29 | 25 |
| SG.COK.DUNG.ZS | Main cooking fuel: dung (% of households) | -1.01 | -2.70 | 32 | 11 |

diffs\_daly\_joined %>%   
 produce\_diff\_table(  
 this\_cause = "Non-communicable diseases",   
 this\_caption = "Signal strength of indicators: gender inequalities in health; DALYs due to NCDs")

## Joining, by = "indicator\_id"

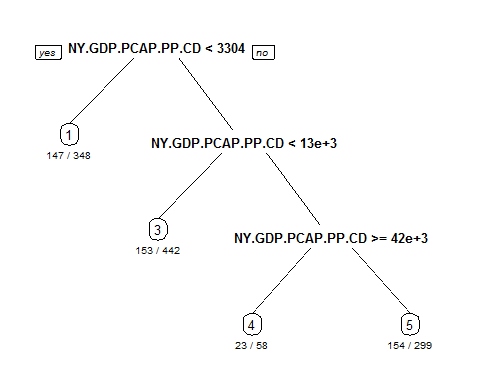
Signal strength of indicators: gender inequalities in health; DALYs due to NCDs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| indicator\_id | indicator | rel\_t | abs\_t | rel\_rank | abs\_rank |
| SP.URB.TOTL.FE.ZS | Urban population, female (% of total) | 18.43 | 10.25 | 1 | 4 |
| SE.ENR.PRIM.FM.ZS | School enrollment, primary (gross), gender parity index (GPI) | 12.59 | 11.68 | 2 | 3 |
| SE.ENR.PRSC.FM.ZS | School enrollment, primary and secondary (gross), gender parity index (GPI) | 12.47 | 11.86 | 3 | 2 |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) | 11.97 | 11.98 | 4 | 1 |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) | 11.83 | 5.99 | 5 | 9 |
| MO.INDEX.XQ | Overall Mo Ibrahim index | 10.09 | 9.61 | 6 | 5 |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) | 9.62 | 1.44 | 7 | 29 |
| UNDP.HDI.XD | Human development index (HDI) | 8.47 | 7.54 | 8 | 6 |
| SG.GEN.TECH.ZS | Female professional and technical workers (% of total) | 7.93 | 7.08 | 9 | 7 |
| SE.NED.HIAT.FE.ZS | Educational attainment, no schooling, population 25+ years, female (%) | -7.28 | -5.43 | 10 | 10 |
| SG.H2O.PRMS.HH.ZS | Households with water on the premises (%) | 6.96 | 5.06 | 11 | 13 |
| SE.NED.HIAT.MA.ZS | Educational attainment, no schooling, population 25+ years, male (%) | -6.82 | -5.13 | 12 | 12 |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) | 6.80 | 6.03 | 13 | 8 |
| SG.COK.LPGN.ZS | Main cooking fuel: LPG/natural gas/biogas (% of households) | 5.50 | 5.22 | 14 | 11 |
| SH.HIV.ARTC.MA.ZS | Access to anti-retroviral drugs, male (%) | 5.48 | 0.38 | 15 | 32 |
| SG.H2O.TM30.HH.ZS | Households with water 30 minutes or longer away round trip (%) | -5.39 | -4.25 | 16 | 15 |
| SE.ADT.1524.LT.FM.ZS | Literacy rate, youth (ages 15-24), gender parity index (GPI) | 4.97 | 3.91 | 17 | 16 |
| SG.COK.WOOD.ZS | Main cooking fuel: wood (% of households) | -4.74 | -4.32 | 18 | 14 |
| SL.UEM.1524.FM.NE.ZS | Ratio of female to male youth unemployment rate (%) (national estimate) | -4.50 | -3.01 | 19 | 21 |
| SH.HIV.ARTC.FE.ZS | Access to anti-retroviral drugs, female (%) | 4.49 | 0.32 | 20 | 33 |
| SG.DMK.HLTH.HB.ZS | Decision maker about a woman’s own health care: mainly husband (% of women age 15-49) | -4.34 | -3.73 | 21 | 17 |
| SG.DMK.HLTH.FN.ZS | Women participating in own health care decisions (% of women age 15-49) | 4.17 | 3.59 | 22 | 18 |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) | -3.67 | -2.67 | 23 | 24 |
| SG.COK.ELEC.ZS | Main cooking fuel: electricity (% of households) | 3.63 | 3.25 | 24 | 19 |
| SG.DMK.FOOD.FN.ZS | Women participating in decision of what food to cook daily (% of women age 15-49) | 3.06 | 2.60 | 25 | 25 |
| SG.COK.CHCO.ZS | Main cooking fuel: charcoal (% of households) | -2.99 | -3.05 | 26 | 20 |
| SG.COK.HOUS.ZS | Location of cooking: inside the house (% of households) | 2.88 | 3.01 | 27 | 22 |
| 5.51.01.07.gender | Gender equality in education | 2.55 | 2.12 | 28 | 26 |
| SG.COK.CROP.ZS | Main cooking fuel: agricultural crop (% of households) | 2.41 | 2.99 | 29 | 23 |
| SG.DMK.DPCH.FN.ZS | Women participating in making daily purchase decisions (% of women age 15-49) | 2.01 | 1.53 | 30 | 28 |

Now a CART to group countries into quintile of gender inequality in outcomes

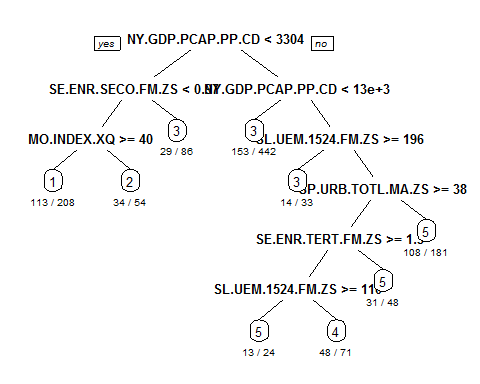
tmp\_dta <- diffs\_mort\_joined %>%   
 filter(cause == "All causes") %>%   
 mutate(quint\_rel = cut(rel, breaks = quantile(rel, seq(0, 1, by = 0.2)), labels = F) %>% as.factor() ) %>%   
 select(location, year, cause, quint\_rel, indicator\_id, indicator\_value) %>%   
 spread(indicator\_id, indicator\_value)   
  
mod\_simple <- tmp\_dta %>%   
 rpart::rpart(quint\_rel ~ year + NY.GDP.PCAP.PP.CD, data = ., method = "class")  
  
# plot(model\_01)  
# text(model\_01, cex = 0.8, use.n = T, xpd = T)  
  
rpart.plot::prp(mod\_simple, faclen = 0, cex = 0.8, varlen = 0, extra = 2, under = T)

## Warning: Cannot retrieve the model data (will ignore roundint=TRUE).  
## To make this warning go away:  
## Call prp with roundint=FALSE,  
## or rebuild the rpart model with model=TRUE.



mod\_full <- tmp\_dta %>%   
 select(-location) %>%   
 rpart::rpart(quint\_rel ~ ., data = ., method = "class")  
  
rpart.plot::prp(mod\_full, faclen = 0, cex = 0.8, varlen = 0, extra = 2, under = T)

## Warning: Cannot retrieve the model data (will ignore roundint=TRUE).  
## To make this warning go away:  
## Call prp with roundint=FALSE,  
## or rebuild the rpart model with model=TRUE.



|  |  |
| --- | --- |
| indicator\_id | indicator |
| NY.GDP.PCAP.PP.CD | GDP per capita, PPP (current international $) |
| SE.ENR.SECO.FM.ZS | School enrollment, secondary (gross), gender parity index (GPI) |
| MO.INDEX.XQ | Overall Mo Ibrahim index |
| SL.UEM.1524.FM.ZS | Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate) |
| SP.URB.TOTL.MA.ZS | Urban population, male (% of total) |
| SE.ENR.TERT.FM.ZS | School enrollment, tertiary (gross), gender parity index (GPI) |

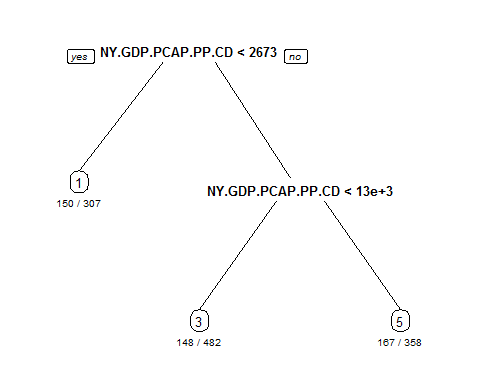
So, the above model seems to be suggesting:

* If GDP per cap < $3k, gender parity in 2ndry school enrolment is low, and Mo Ibrahim index high -> lowest/negative male gender disadvantage
* If GDP per cap < $3k, gender parity in 2ndry school enrolment is now, and Mo Ibrahim index low -> 2nd quintile
* If GDP per cap < $3k, gender parity in 2ndry school enrolment is high -> 3rd quintile
* If GDP per cap < $13k -> 3rd quintile
* If GDP per cap > $13k, high female:male youth unemployment rate -> 3rd quintile
* If GDP per cap > $13k, < 38% of males in urban areas (?) -> 5th quintile
* If GDP per cap > $13k, >= 38% of males in urban areas (?) but low gender parity index in tertiary school enrolment -> 5th quintile
* If GDP per cap > $13k, >= 38% of males in urban areas, high gender parity in tertiary school enrolment, but female:male youth unemployment >= 110 -> 5th quintile
* Otherwise 4th quintile

Now the same for NCDs

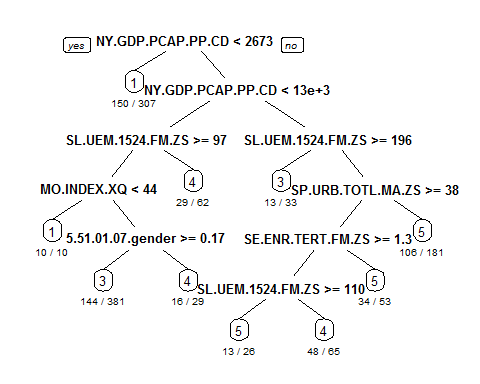
tmp\_dta <- diffs\_mort\_joined %>%   
 filter(cause == "Non-communicable diseases") %>%   
 mutate(quint\_rel = cut(rel, breaks = quantile(rel, seq(0, 1, by = 0.2)), labels = F) %>% as.factor() ) %>%   
 select(location, year, cause, quint\_rel, indicator\_id, indicator\_value) %>%   
 spread(indicator\_id, indicator\_value)   
  
mod\_simple <- tmp\_dta %>%   
 rpart::rpart(quint\_rel ~ year + NY.GDP.PCAP.PP.CD, data = ., method = "class")  
  
# plot(model\_01)  
# text(model\_01, cex = 0.8, use.n = T, xpd = T)  
  
rpart.plot::prp(mod\_simple, faclen = 0, cex = 0.8, varlen = 0, extra = 2, under = T)

## Warning: Cannot retrieve the model data (will ignore roundint=TRUE).  
## To make this warning go away:  
## Call prp with roundint=FALSE,  
## or rebuild the rpart model with model=TRUE.



mod\_full <- tmp\_dta %>%   
 select(-location) %>%   
 rpart::rpart(quint\_rel ~ ., data = ., method = "class")  
  
rpart.plot::prp(mod\_full, faclen = 0, cex = 0.8, varlen = 0, extra = 2, under = T)

## Warning: Cannot retrieve the model data (will ignore roundint=TRUE).  
## To make this warning go away:  
## Call prp with roundint=FALSE,  
## or rebuild the rpart model with model=TRUE.



But what are these thresholds?

diffs\_mort\_joined %>%   
 filter(cause == "All causes") %>%   
 pull(rel) %>%   
 quantile(seq(0, 1, by = 0.2)) -> tmp1  
  
diffs\_mort\_joined %>%   
 filter(cause == "Non-communicable diseases") %>%   
 pull(rel) %>%   
 quantile(seq(0, 1, by = 0.2)) -> tmp2  
  
tab <- rbind(tmp1, tmp2); rm(tmp1, tmp2)  
row.names(tab) <- c("All causes", "NCDs")  
  
tab

## 0% 20% 40% 60% 80% 100%  
## All causes 0.7704889 1.0797595 1.235633 1.473530 1.729629 2.642346  
## NCDs 0.4499314 0.9707666 1.206926 1.428568 1.692264 2.655742