R Notebook

# Further notes

* From Vittal

Really good to see this progress.

OK, here’s a suggestion for the structure of the paper:

* *Figure 1*: Relative and absolute gaps in (a) life expectancy and (b) HALYs, stratified by SDI groups
* *Figure 2*: (a) Relative and (b) absolute gaps in all cause mortality, stratified by SDI and age groups (i.e. what you’ve already done)
* *Figure 3*: (a) Relative and (b) absolute gaps in DALYs, stratified by SDI and age groups (i.e. what you’ve already done)
* *Figure 4*: Decomposition of (a) absolute gaps and (b) relative gaps in life expectancy by risk factors (i.e. behavioural and other), stratified by SDI groups
* *Figure 5*: Decomposition of (a) absolute gaps and (b) relative gaps in HALYs by risk factors (i.e. behavioural and other), stratified by SDI groups
* *Table 1*: Fixed-effects regression models of the association between social indicators and (a) life expectancy and (b) HALYs

Exposures would be:

* [GDP per capita](https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD)
* [UN HDR Gender Inequality Index (GII)](http://hdr.undp.org/en/content/gender-inequality-index-gii):
* World Bank Health expenditure per capita (also get info on health expenditure as % of GDP)
* [World Income Inequality Database](https://www.wider.unu.edu/database/world-income-inequality-database-wiid34) Unfortunately, this would require a little processing to create a time series that only includes one observation per country-year. There is advice on how to do this. See what you think - if v time-consuming, we could drop.

It may be worthwhile looking at additional exposures. I’ve tried to focus on data which cover a large number of countries and across a long time series (so we do not lose too many country-years in the analysis). I’d suggest doing a fixed-effects analysis where we cluster by country.

*Appendix*: Repeat Figs 2 and 3 decomposed by risk factors. Repeat Figs 1, 4 and 5 stratified by continent. Repeat Figs 1, 4 and 5 for each individual country within a large table for selected years i.e. 1990 and most recent available year.

I’d probably avoid decomposing by disease because the GBD papers that are coming out this year are beginning to do that a bit (I think based on my suggestion). What do you think? Please do disagree if you have other thoughts.

Does the above make sense? It’s quite possible it doesn’t - if so, let’s meet to discuss. I could do Mon 21st 1530 at SPHSU or Fri 25th (not 14-1500) venue TBC.

It probably would be sensible to have most/all of Figs 1-5 drafted and then we can circulate to the broader group. I wouldn’t bother doing the Appendix and Table just now, but prepping the exposure data in the meantime may be worthwhile.

Cheers.

Vittal

New permalink: <http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2016-permalink/49f19f6b2e1fafca509aeeea1adda399>

[This file](./download_ro2_csv_files_directly.R) contains the code for downloading the zip files directly by url.

To start let’s load and bind the files together

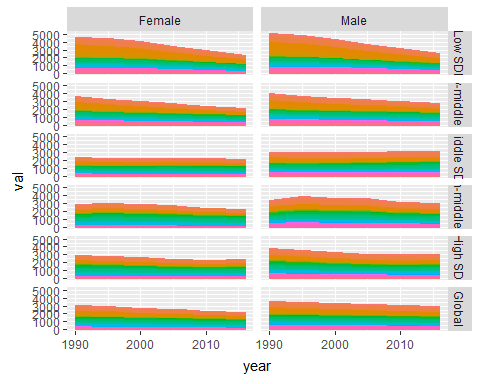
## # A tibble: 1,008,504 x 18  
## measure\_id measure\_name location\_id location\_name sex\_id sex\_name  
## <int> <chr> <int> <chr> <int> <chr>   
## 1 1 Deaths 1 Global 2 Female   
## 2 1 Deaths 1 Global 1 Male   
## 3 1 Deaths 1 Global 2 Female   
## 4 1 Deaths 1 Global 1 Male   
## 5 1 Deaths 1 Global 2 Female   
## 6 1 Deaths 1 Global 1 Male   
## 7 1 Deaths 1 Global 2 Female   
## 8 1 Deaths 1 Global 1 Male   
## 9 1 Deaths 1 Global 2 Female   
## 10 1 Deaths 1 Global 1 Male   
## # ... with 1,008,494 more rows, and 12 more variables: age\_id <int>,  
## # age\_name <chr>, cause\_id <int>, cause\_name <chr>, rei\_id <int>,  
## # rei\_name <chr>, metric\_id <int>, metric\_name <chr>, year <int>,  
## # val <dbl>, upper <dbl>, lower <dbl>

Let’s do some sense-checking exploratory analyses

* All cause mortality rate by rei

dta %>%   
 select(sex = sex\_name, age = age\_name, risk\_factor = rei\_name, cause = cause\_name, sdi = location\_name, measure = measure\_name, metric = metric\_name, year = year, val) %>%   
 select(cause, age, year, sex, sdi, measure, metric, risk\_factor, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age == "All Ages") %>%   
 filter(metric == "Rate") %>%   
 filter(measure == "Deaths") %>%   
 select(-cause, -age, -metric, -measure) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 ggplot(aes(colour = risk\_factor, fill = risk\_factor, x = year, y = val)) +   
 geom\_area(position = 'stack') +   
 facet\_grid(sdi ~ sex) +   
 guides(fill = FALSE, colour = FALSE)

## Warning: package 'bindrcpp' was built under R version 3.4.4



ggsave("figures/risk\_factor\_facets.png", width = 60, height = 60, units = "cm", dpi = 150)

I’ve redone the query to include all countries as well as separate SDI groups.

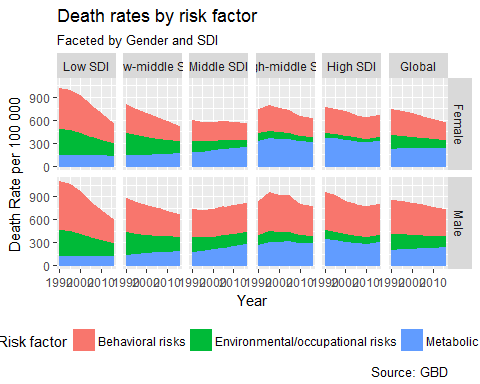
I now want to learn more about the main risk disease categories and hierarchies. Do I have this info in a lookup already?

Yes. Thse are in REI hierarchy

rei\_codebook <- readxl::read\_excel("raw\_data/IHME\_GBD\_2016\_CODEBOOK/IHME\_GBD\_2016\_REI\_HIERARCHY\_Y2018M04D26.XLSX")

# Let's keep all level 1  
  
rei\_codebook %>%   
 right\_join(dta) %>%   
 filter(level == 1) %>%   
 select(sex = sex\_name, age = age\_name, risk\_factor = rei\_name, cause = cause\_name, sdi = location\_name, measure = measure\_name, metric = metric\_name, year = year, val) %>%   
 select(cause, age, year, sex, sdi, measure, metric, risk\_factor, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age == "All Ages") %>%   
 filter(metric == "Rate") %>%   
 filter(measure == "Deaths") %>%   
 select(-cause, -age, -metric, -measure) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 ggplot(aes(colour = risk\_factor, fill = risk\_factor, x = year, y = val)) +   
 geom\_area(position = 'stack') +   
 facet\_grid(sex ~ sdi) +   
 labs(x = "Year", y = "Death Rate per 100 000", title = "Death rates by risk factor",  
 subtitle = "Faceted by Gender and SDI", fill = "Risk factor", colour = "Risk factor",  
 caption = "Source: GBD") +  
 theme(legend.position = "bottom")

## Joining, by = c("rei\_id", "rei\_name")



Now let’s take the coding scheme/approach used previously to show trends by SDI over time. We start by porting over some convenience functions I made:

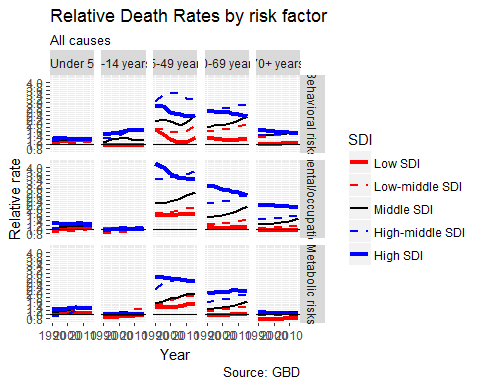
order\_age\_sdi <- function(x, age\_var, location\_var){  
   
 # enquo finds the value associated with the var name and makes sure that   
 # when !! is used, the value is used as the name  
   
 # quo\_name used on LHS before special asignment operator :=   
 # for naming columns   
   
 age\_var <- enquo(age\_var)  
 location\_var <- enquo(location\_var)  
   
 x %>%   
 mutate(!!quo\_name(age\_var) := factor(  
 !!age\_var, ordered = T,   
 levels = c("Under 5", "5-14 years", "15-49 years", "50-69 years", "70+ years")  
 )  
 ) %>%   
 mutate(!!quo\_name(location\_var) := factor(  
 !!location\_var, ordered = T,  
 levels = c("Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")  
 )  
 ) -> out  
 out  
  
}  
  
  
# Function for adding consistent aesthetics to comparisons between SDI groups  
  
sdi\_scaling <- function(){  
 list(  
 scale\_colour\_manual("SDI", values = c("red", "red", "black", "blue", "blue")),   
 scale\_linetype\_manual("SDI", values = c("solid", "dashed", "solid", "dashed", "solid")),   
 scale\_size\_manual("SDI", values = c(1.5, 1, 1, 1, 1.5))   
 )  
}

# to start. let's   
# 1) make age\_name ordered   
# 2) filter on cause\_name == "all causes")  
# 3) calc relative and absolute   
rei\_codebook %>%   
 right\_join(dta) %>%   
 filter(level == 1) %>%   
 select(sex = sex\_name, age = age\_name, risk\_factor = rei\_name, cause = cause\_name, sdi = location\_name, measure = measure\_name, metric = metric\_name, year = year, val) %>%   
 select(cause, age, year, sex, sdi, measure, metric, risk\_factor, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age != "All Ages") %>%   
 filter(metric == "Rate") %>%   
 filter(measure == "Deaths") %>%   
 filter(sdi != "Global") %>%   
 select(-cause, -metric, -measure) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(rel = Male / Female, abs = Male - Female) %>%   
 select(-Female, -Male) -> sdi\_death\_rates

## Joining, by = c("rei\_id", "rei\_name")

Here’s the graph of relative risk, all causes

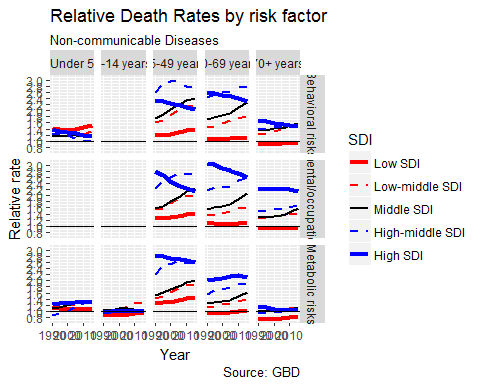
sdi\_death\_rates %>%   
 ggplot(aes(x = year, y = rel, colour = sdi, linetype = sdi, size = sdi)) +   
 facet\_grid(risk\_factor ~ age) +   
 geom\_line() +  
 labs(title = "Relative Death Rates by risk factor", subtitle = "All causes", caption = "Source: GBD",  
 y = "Relative rate", x = "Year") +  
 sdi\_scaling() +   
 scale\_y\_continuous(breaks = seq(0.8, 4, by = 0.2)) +  
 geom\_hline(yintercept = 1.0)



And here’s the same for NCDs only

rei\_codebook %>%   
 right\_join(dta) %>%   
 filter(level == 1) %>%   
 select(sex = sex\_name, age = age\_name, risk\_factor = rei\_name, cause = cause\_name, sdi = location\_name, measure = measure\_name, metric = metric\_name, year = year, val) %>%   
 select(cause, age, year, sex, sdi, measure, metric, risk\_factor, val) %>%   
 filter(cause == "Non-communicable diseases") %>%   
 filter(age != "All Ages") %>%   
 filter(metric == "Rate") %>%   
 filter(measure == "Deaths") %>%   
 filter(sdi != "Global") %>%   
 select(-cause, -metric, -measure) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(rel = Male / Female, abs = Male - Female) %>%   
 select(-Female, -Male) %>%   
 ggplot(aes(x = year, y = rel, colour = sdi, linetype = sdi, size = sdi)) +   
 facet\_grid(risk\_factor ~ age) +   
 geom\_line() +  
 labs(title = "Relative Death Rates by risk factor", subtitle = "Non-communicable Diseases", caption = "Source: GBD",  
 y = "Relative rate", x = "Year") +  
 sdi\_scaling() +   
 scale\_y\_continuous(breaks = seq(0.8, 4, by = 0.2)) +  
 geom\_hline(yintercept = 1.0)

## Joining, by = c("rei\_id", "rei\_name")



*To do: Work out why values missing for age 5-14 years for Behavioural and Mental/Occupational. And Under 5 for Mental/Occupational - this is likely due to these death/risk factor combinations being very rare and so estimates unreliable.*

*Suggestion: Start comparisons after age 14 years*

# Reminder of suggested main figures:

* *Figure 1*: Relative and absolute gaps in (a) life expectancy and (b) HALYs, stratified by SDI groups
* *Figure 4*: Decomposition of (a) absolute gaps and (b) relative gaps in life expectancy by risk factors (i.e. behavioural and other), stratified by SDI groups
* *Figure 5*: Decomposition of (a) absolute gaps and (b) relative gaps in HALYs by risk factors (i.e. behavioural and other), stratified by SDI groups

Let’s start with figure 1, including all necessary downloading etc

## Figure 1

To start need to download e0 by SDI, gender, year, and death status (all, NCD)

This will require another query. Having wrapped up the data grabber into a function this should now be a bit easier.

First e0.

url <- "http://s3.healthdata.org/gbd-api-2016-production/0960785dcee27c60760f6ac02d2025d7\_files/IHME-GBD\_2016\_DATA-0960785d-1.zip"  
url\_before <- "http://s3.healthdata.org/gbd-api-2016-production/0960785dcee27c60760f6ac02d2025d7\_files"  
url\_after <- "IHME-GBD\_2016\_DATA-0960785d-"  
source("download\_completed\_request.R")  
download\_completed\_request(  
 url\_before\_filenm = url\_before,   
 url\_filenm\_root = url\_after,   
 outdir = "raw\_data/e0",   
 flush = FALSE  
 )

Now to download from local dir

dta\_e0 <- read\_csv("raw\_data/e0/1.csv")

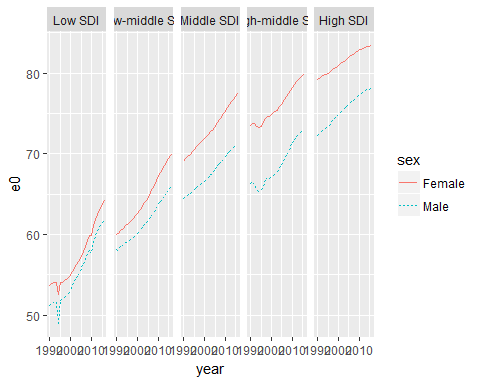
## Parsed with column specification:  
## cols(  
## measure\_id = col\_integer(),  
## measure\_name = col\_character(),  
## location\_id = col\_integer(),  
## location\_name = col\_character(),  
## sex\_id = col\_integer(),  
## sex\_name = col\_character(),  
## age\_id = col\_integer(),  
## age\_name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

dta\_e0

## # A tibble: 324 x 14  
## measure\_id measure\_name location\_id location\_name sex\_id sex\_name  
## <int> <chr> <int> <chr> <int> <chr>   
## 1 26 Life expectancy 1 Global 1 Male   
## 2 26 Life expectancy 1 Global 2 Female   
## 3 26 Life expectancy 1 Global 1 Male   
## 4 26 Life expectancy 1 Global 2 Female   
## 5 26 Life expectancy 44634 High-middle SDI 1 Male   
## 6 26 Life expectancy 44634 High-middle SDI 2 Female   
## 7 26 Life expectancy 44635 High SDI 1 Male   
## 8 26 Life expectancy 44635 High SDI 2 Female   
## 9 26 Life expectancy 44636 Low-middle SDI 1 Male   
## 10 26 Life expectancy 44636 Low-middle SDI 2 Female   
## # ... with 314 more rows, and 8 more variables: age\_id <int>,  
## # age\_name <chr>, metric\_id <int>, metric\_name <chr>, year <int>,  
## # val <dbl>, upper <dbl>, lower <dbl>

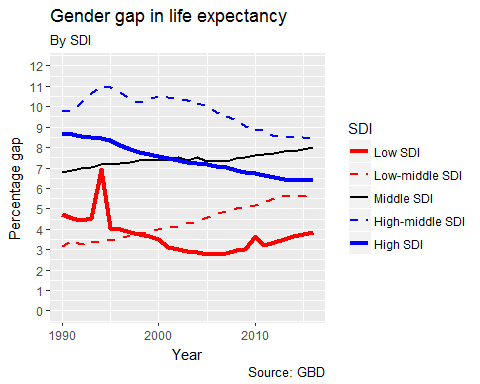
To visualise, values

dta\_e0 %>%   
 filter(location\_name != "Global") %>% filter(age\_name == "All Ages") %>%   
 filter(measure\_name == "Life expectancy") %>%   
 select(sdi = location\_name, year, sex = sex\_name, e0 = val) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>% ggplot(aes(x = year, y = e0, colour = sex, linetype = sex)) +   
 geom\_line() +   
 facet\_wrap(~sdi, nrow = 1)



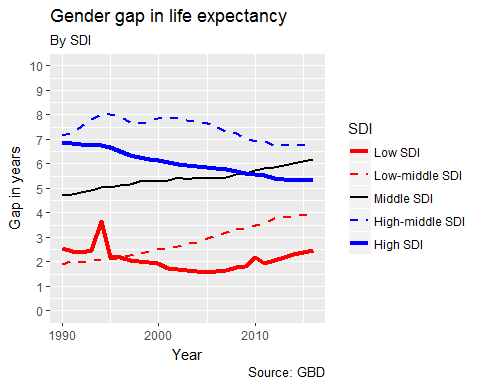
Now relative differences

dta\_e0 %>%   
 filter(location\_name != "Global") %>% filter(age\_name == "All Ages") %>%   
 filter(measure\_name == "Life expectancy") %>%   
 select(sdi = location\_name, year, sex = sex\_name, e0 = val) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 spread(sex, e0) %>%   
 mutate(rel = Male / Female) %>%   
 mutate(perc\_gap = (1 - rel) \* 100) %>%   
 ggplot(aes(x = year, y = perc\_gap, colour = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +  
 sdi\_scaling() +  
 scale\_y\_continuous(breaks = seq(0, 12, by = 1), limits = c(0, 12)) +   
 labs(title = "Gender gap in life expectancy",   
 subtitle = "By SDI",   
 x = "Year", y = "Percentage gap",  
 caption = "Source: GBD")



Now absolute differences

dta\_e0 %>%   
 filter(location\_name != "Global") %>% filter(age\_name == "All Ages") %>%   
 filter(measure\_name == "Life expectancy") %>%   
 select(sdi = location\_name, year, sex = sex\_name, e0 = val) %>%   
 mutate(sdi = forcats::fct\_relevel(sdi, "Low SDI", "Low-middle SDI", "Middle SDI", "High-middle SDI", "High SDI")) %>%   
 spread(sex, e0) %>%   
 mutate(abs = Female - Male) %>%   
 ggplot(aes(x = year, y = abs, colour = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +  
 sdi\_scaling() +  
 scale\_y\_continuous(breaks = seq(0, 10, by = 1), limits = c(0, 10)) +   
 labs(title = "Gender gap in life expectancy",   
 subtitle = "By SDI",   
 x = "Year", y = "Gap in years",  
 caption = "Source: GBD")



Now to so the same with HALE

* Tried to query this - “Your query returned no data”

Will try DALYs instead First download the data

url <- "http://s3.healthdata.org/gbd-api-2016-production/1ad84f3d429a369aeb4b401311b852d0\_files/IHME-GBD\_2016\_DATA-1ad84f3d-1.zip"  
url\_before <- "http://s3.healthdata.org/gbd-api-2016-production/1ad84f3d429a369aeb4b401311b852d0\_files"  
url\_after <- "IHME-GBD\_2016\_DATA-1ad84f3d-"  
source("download\_completed\_request.R")  
download\_completed\_request(  
 url\_before\_filenm = url\_before,   
 url\_filenm\_root = url\_after,   
 outdir = "raw\_data/dalys",   
 flush = FALSE  
 )

DALYs by gender, age group, SDI - for all cause and NCD only

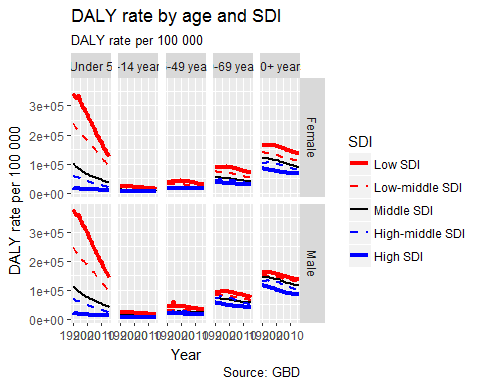
First load the data

dta\_daly <- read\_csv("raw\_data/dalys/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(),  
## metric\_id = col\_integer(),  
## metric\_name = col\_character(),  
## year = col\_integer(),  
## val = col\_double(),  
## upper = col\_double(),  
## lower = col\_double()  
## )

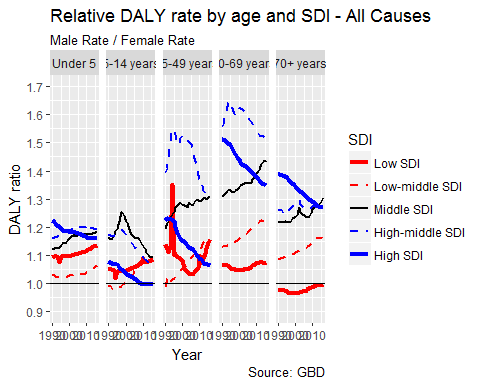
First all cause

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 ggplot(aes(x = year, y = val, colour = sdi, group = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_grid(sex ~ age) +   
 sdi\_scaling() +  
 labs(title = "DALY rate by age and SDI",  
 subtitle = "DALY rate per 100 000",  
 x = "Year", y = "DALY rate per 100 000",  
 caption = "Source: GBD")



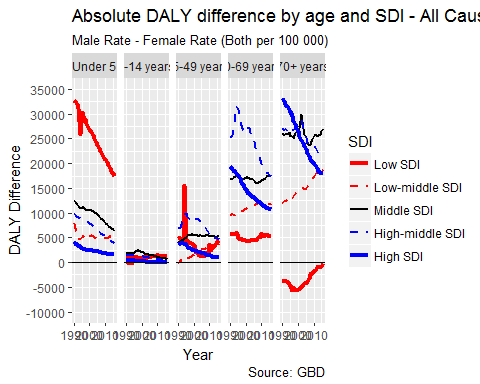
All cause - relative difference by gender

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(rel = Male / Female) %>%   
 ggplot(aes(x = year, y = rel, group = sdi, col = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_wrap(~age, nrow = 1) +  
 sdi\_scaling() +  
 labs(title = "Relative DALY rate by age and SDI - All Causes",  
 subtitle = "Male Rate / Female Rate",  
 x = "Year", y = "DALY ratio",  
 caption = "Source: GBD") +   
 scale\_y\_continuous(breaks = seq(0.9, 1.7, by = .1), limits = c(0.9, 1.7)) +  
 geom\_hline(yintercept = 1)



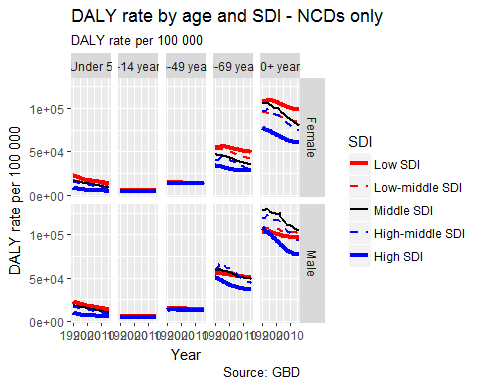
Now absolute - all cause

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "All causes") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(abs = Male - Female) %>%   
 ggplot(aes(x = year, y = abs, group = sdi, col = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_wrap(~age, nrow = 1) +  
 sdi\_scaling() +  
 labs(title = "Absolute DALY difference by age and SDI - All Causes",  
 subtitle = "Male Rate - Female Rate (Both per 100 000)",  
 x = "Year", y = "DALY Difference",  
 caption = "Source: GBD") +   
 scale\_y\_continuous(breaks = seq(-10000, 35000, by = 5000), limits = c(-10000, 35000)) +  
 geom\_hline(yintercept = 0)



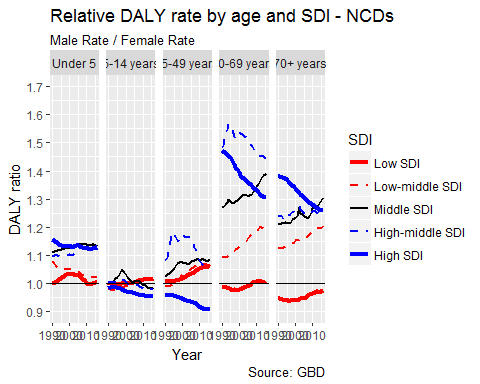
## Now NCD only

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "Non-communicable diseases") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 ggplot(aes(x = year, y = val, colour = sdi, group = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_grid(sex ~ age) +   
 sdi\_scaling() +  
 labs(title = "DALY rate by age and SDI - NCDs only",  
 subtitle = "DALY rate per 100 000",  
 x = "Year", y = "DALY rate per 100 000",  
 caption = "Source: GBD")



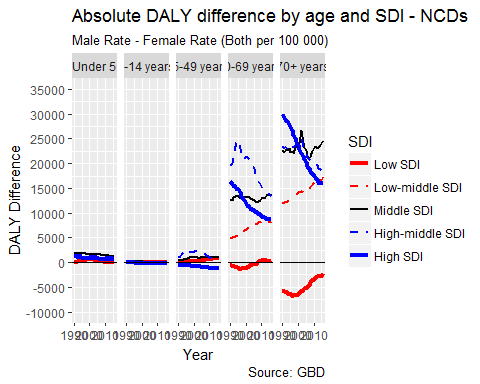
All cause - relative difference by gender

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "Non-communicable diseases") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(rel = Male / Female) %>%   
 ggplot(aes(x = year, y = rel, group = sdi, col = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_wrap(~age, nrow = 1) +  
 sdi\_scaling() +  
 labs(title = "Relative DALY rate by age and SDI - NCDs",  
 subtitle = "Male Rate / Female Rate",  
 x = "Year", y = "DALY ratio",  
 caption = "Source: GBD") +   
 scale\_y\_continuous(breaks = seq(0.9, 1.7, by = .1), limits = c(0.9, 1.7)) +  
 geom\_hline(yintercept = 1)



Now absolute - all cause

dta\_daly %>%   
 select(-measure\_id, -measure\_name) %>%   
 rename(sdi = location\_name) %>%   
 select(-location\_id, -sex\_id, -age\_id) %>%   
 select(sex = sex\_name, sdi, age = age\_name, year, cause = cause\_name, val) %>%   
 filter(cause == "Non-communicable diseases") %>%   
 filter(age != "All Ages") %>%   
 order\_age\_sdi(age\_var = age, location\_var = sdi) %>%   
 spread(sex, val) %>%   
 mutate(abs = Male - Female) %>%   
 ggplot(aes(x = year, y = abs, group = sdi, col = sdi, linetype = sdi, size = sdi)) +   
 geom\_line() +   
 facet\_wrap(~age, nrow = 1) +  
 sdi\_scaling() +  
 labs(title = "Absolute DALY difference by age and SDI - NCDs",  
 subtitle = "Male Rate - Female Rate (Both per 100 000)",  
 x = "Year", y = "DALY Difference",  
 caption = "Source: GBD") +   
 scale\_y\_continuous(breaks = seq(-10000, 35000, by = 5000), limits = c(-10000, 35000)) +  
 geom\_hline(yintercept = 0)



# Decomposition figures

Want the following

* Age standardised (then age-stratified)
* All cause mortality rates
* DALY rates
* By gender
* By level 2 risk factors
* By SDI group

# Returning to fig 4 and 5

* *Figure 4*: Decomposition of (a) absolute gaps and (b) relative gaps in life expectancy by risk factors (i.e. behavioural and other), stratified by SDI groups
* *Figure 5*: Decomposition of (a) absolute gaps and (b) relative gaps in HALYs by risk factors (i.e. behavioural and other), stratified by SDI groups

Re Fig 4: I think decomposition of e0 into risk factors is perhaps more of a complex actuarial task requiring more age-disaggregated data than it may first appear. *Suggestion: Speak to Frank Popham for further advice* - However - might want to look at YLLs, and differences in YLLs by risk factor.

Re Fig 5: So far more data appears to be available by DALYs than HALE.

*Key Suggestion*: **Run GBD Results Tool with preferred queries set. Save permalink and send to me (with description). I’ll be able to download the files directly from this link. (n.b. data split into multiple files are fine too. The script I’ve written will just iterate through these files.)**