Risk Factors

# Intro

This doc will explore mortality and DALYs by risk factors

First to load the pre requisite packages and script for downloading data.

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

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

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

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

Next I will download the data from the request. (Only has to be done once so now set not to evaluate (run)).

url\_body <- "http://s3.healthdata.org/gbd-api-2016-production/6790a3f2d5948a86929939c83525f81e\_files"  
url\_head <- "IHME-GBD\_2016\_DATA-6790a3f2-"  
outdir <- "raw\_data/risk"  
  
download\_completed\_request(url\_body, url\_head, outdir, flush = T)

The files can now be loaded and joined locally as follows

dta <- read\_csv("raw\_data/risk/1.csv") %>% bind\_rows(read\_csv("raw\_data/risk/2.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()  
## )  
## 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()  
## )

dta

## # A tibble: 501,900 x 18  
## measure\_id measure\_name location\_id  
## <int> <chr> <int>  
## 1 2 DALYs (Disability-Adjusted Life Years) 1  
## 2 2 DALYs (Disability-Adjusted Life Years) 1  
## 3 2 DALYs (Disability-Adjusted Life Years) 1  
## 4 2 DALYs (Disability-Adjusted Life Years) 1  
## 5 2 DALYs (Disability-Adjusted Life Years) 1  
## 6 2 DALYs (Disability-Adjusted Life Years) 1  
## 7 2 DALYs (Disability-Adjusted Life Years) 1  
## 8 2 DALYs (Disability-Adjusted Life Years) 1  
## 9 2 DALYs (Disability-Adjusted Life Years) 1  
## 10 2 DALYs (Disability-Adjusted Life Years) 1  
## # ... with 501,890 more rows, and 15 more variables: location\_name <chr>,  
## # sex\_id <int>, sex\_name <chr>, 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>

glimpse(dta)

## Observations: 501,900  
## Variables: 18  
## $ measure\_id <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...  
## $ measure\_name <chr> "DALYs (Disability-Adjusted Life Years)", "DALYs...  
## $ location\_id <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...  
## $ location\_name <chr> "Global", "Global", "Global", "Global", "Global"...  
## $ sex\_id <int> 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, ...  
## $ sex\_name <chr> "Male", "Female", "Male", "Female", "Male", "Fem...  
## $ age\_id <int> 1, 1, 1, 1, 1, 1, 22, 22, 22, 22, 22, 22, 23, 23...  
## $ age\_name <chr> "Under 5", "Under 5", "Under 5", "Under 5", "Und...  
## $ cause\_id <int> 294, 294, 294, 294, 294, 294, 294, 294, 294, 294...  
## $ cause\_name <chr> "All causes", "All causes", "All causes", "All c...  
## $ rei\_id <int> 161, 161, 161, 161, 161, 161, 161, 161, 161, 161...  
## $ rei\_name <chr> "Occupational exposure to silica", "Occupational...  
## $ metric\_id <int> 1, 1, 2, 2, 3, 3, 1, 1, 2, 2, 3, 3, 1, 1, 2, 2, ...  
## $ metric\_name <chr> "Number", "Number", "Percent", "Percent", "Rate"...  
## $ year <int> 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, ...  
## $ val <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.0000...  
## $ upper <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.0000...  
## $ lower <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.0000...

The \_id suffix columns can be used to join to the lookup tables. In this case for risk (by rei\_id)

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

## # A tibble: 144 x 5  
## rei\_id rei\_name parent\_id level  
## <dbl> <chr> <dbl> <dbl>  
## 1 169 All risk factors 169 0  
## 2 202 Environmental/occupational risks 169 1  
## 3 203 Behavioral risks 169 1  
## 4 104 Metabolic risks 169 1  
## 5 171 Etiologies 171 0  
## 6 191 Impairments 191 0  
## 7 82 Unsafe water, sanitation, and handwashing 202 2  
## 8 83 Unsafe water source 82 3  
## 9 84 Unsafe sanitation 82 3  
## 10 238 No access to handwashing facility 82 3  
## # ... with 134 more rows, and 1 more variables: sort\_order <dbl>

glimpse(lookup)

## Observations: 144  
## Variables: 5  
## $ rei\_id <dbl> 169, 202, 203, 104, 171, 191, 82, 83, 84, 238, 85, ...  
## $ rei\_name <chr> "All risk factors", "Environmental/occupational ris...  
## $ parent\_id <dbl> 169, 169, 169, 169, 171, 191, 202, 82, 82, 82, 202,...  
## $ level <dbl> 0, 1, 1, 1, 0, 0, 2, 3, 3, 3, 2, 3, 3, 3, 2, 3, 3, ...  
## $ sort\_order <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ...

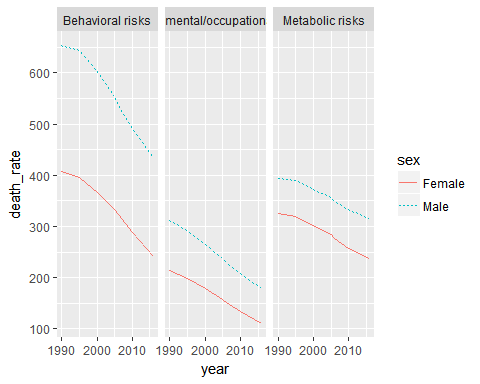
In this example I’ll join then filter only level 1 categories in the rei hierarchy

dta %>%   
 left\_join(lookup) %>%   
 filter(level == 1) -> dta\_lvl1

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

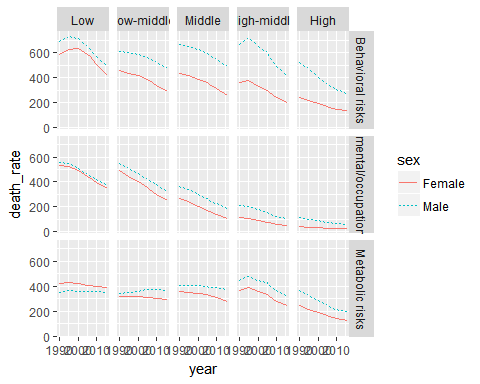
Now to start exploring: age-standardised, global, deaths, by lvl 1 risk factor and gender

dta\_lvl1 %>%  
 filter(measure\_name == "Deaths") %>%   
 filter(location\_name == "Global") %>%   
 filter(age\_name == "Age-standardized") %>%   
 filter(metric\_name == "Rate") %>%   
 filter(cause\_name == "All causes") %>%   
 select(year, sex = sex\_name, risk\_factor = rei\_name, death\_rate = val) %>%   
 ggplot(aes(x = year, y = death\_rate, linetype = sex, colour = sex)) +   
 facet\_wrap(~ risk\_factor) +  
 geom\_line()



Now to do the same by SDI

dta\_lvl1 %>%  
 filter(measure\_name == "Deaths") %>%   
 filter(location\_name != "Global") %>%   
 filter(age\_name == "Age-standardized") %>%   
 filter(metric\_name == "Rate") %>%   
 filter(cause\_name == "All causes") %>%   
 select(year, sex = sex\_name, sdi = location\_name, risk\_factor = rei\_name, death\_rate = val) %>%   
 mutate(sdi = stringr::str\_replace(sdi, " SDI", "")) %>%   
 mutate(sdi = factor(sdi, levels = c("Low", "Low-middle", "Middle", "High-middle", "High"), ordered = T)) %>%   
 ggplot(aes(x = year, y = death\_rate, linetype = sex, colour = sex)) +   
 facet\_grid(risk\_factor ~ sdi) +  
 geom\_line()



And now let’s see what this implies for relative inequalities

dta\_lvl1 %>%  
 filter(measure\_name == "Deaths") %>%   
 filter(location\_name != "Global") %>%   
 filter(age\_name == "Age-standardized") %>%   
 filter(metric\_name == "Rate") %>%   
 filter(cause\_name == "All causes") %>%   
 select(year, sex = sex\_name, sdi = location\_name, risk\_factor = rei\_name, death\_rate = val) %>%   
 mutate(sdi = stringr::str\_replace(sdi, " SDI", "")) %>%   
 mutate(sdi = factor(sdi, levels = c("Low", "Low-middle", "Middle", "High-middle", "High"), ordered = T)) %>% spread(sex, death\_rate) %>% mutate(ratio = Male / Female) %>%   
 ggplot(aes(x = year, y = ratio, colour = risk\_factor)) +   
 geom\_line() +   
 facet\_grid(. ~ sdi) +   
 geom\_hline(yintercept = 1)

