# European Doctoral School of Demography (EDSD)
# Decomposition Techniques - Final Assignment

Ainhoa-Elena Leger
Gonzalo Daniel Garcia
Liliana Patricia Calderon Bernal
Marilyn-Anne Tremblay
Özge Elif Özer

04/6/2021

```
#Loading the packages
library(ggplot2)
library(DemoDecomp)
library(HMDHFDplus)
library(demography)
library(tidyverse)
library(eurostat)

# HFD credentials
user <- ""
pass <- ""

# HMD credentials
HMD_user=""
HMD_pass=""

# Directory of the data
setwd("C:/Users/Ainhoa/Desktop/Github Repositories/EDSD202021/Decomposition Techniques")
```

## Challenge 1

**Proof Kitagawa decomposition (1995) without interactions**

Define the difference between the crude death rates as $\Delta$.

$$\Delta\text{CDR} = \sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)} - \sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)}$$

I divide each of the terms into two equal parts and add and subtract some additional terms, thereby keeping the difference ($\Delta$) constant.

$$\Delta\text{CDR} = \frac{\sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)}}{2} + \frac{\sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)}}{2} - \frac{\sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)}}{2} - \frac{\sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)}}{2}$$
$$+ \frac{\sum_x M_x(t_1)\frac{N_x(t_2)}{N(t_2)}}{2} - \frac{\sum_x M_x(t_1)\frac{N_x(t_2)}{N(t_2)}}{2} + \frac{\sum_x M_x(t_2)\frac{N_x(t_1)}{N(t_1)}}{2} - \frac{\sum_x M_x(t_2)\frac{N_x(t_1)}{N(t_1)}}{2}$$

I now combine the eight terms in $\Delta$ into four:

$$\Delta\text{CDR} = \sum_x \frac{N_x(t_2)}{N(t_2)}\left(\frac{M_x(t_2) + M_x(t_1)}{2}\right) - \sum_x \frac{N_x(t_1)}{N(t_1)}\left(\frac{M_x(t_2) + M_x(t_1)}{2}\right)$$
$$+ \sum_x M_x(t_2)\left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2}\right) - \sum_x M_x(t_1)\left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2}\right).$$

Finally, we combine the terms into two:

$$\Delta\text{CDR} = \sum_x \left(\frac{M_x(t_2) + M_x(t_1)}{2}\right)\left(\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right) + \sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2}\right)(M_x(t_2) - M_x(t_1)).$$

The first terms is the difference in age composition weighted by the average age-specific mortality, while the second term is the difference in rate schedules weighted by the average age composition. Therefore, $\Delta$ is equal to the sum of the contribution of age compositional differences and the contribution of rate schedule differences.

# Challenge 2

**With data on fertility (e.g. HFD) select 3 countries and analyze the change in their crude fertility rate (CFR) in a recent period (10 years) and decompose these changes following Kitagawa's decomposition and describe your results. Then for the most recent period select the two countries (among the 3) with the highest and lowest CFR and decompose their difference and describe your results.**

For doing this challenge we decided to take Spain, Korea and Czechia as the three countries to compare. The data was retrieved from the Human Fertility Database (HFD), (Jasilioniene et al. 2015) using Tim Riffe's package `HMDHFDplus` (Riffe 2015). The selected measure is the General Fertility Rate, as defined in page 93 of Preston's book (Preston, Heuveline, and Guillot 2000):

$$GFR[0,T] = \frac{Births[0,T]}{Person-years-lived[0,T]}$$

We used this measure instead of the Crude Birth Rate to avoid a bias of different age-structure of populations.

**A glimpse on data and some literature**

After some data wrangling (code provided), we have the following time series of the GFR for the three countries:

GFR for selected countries
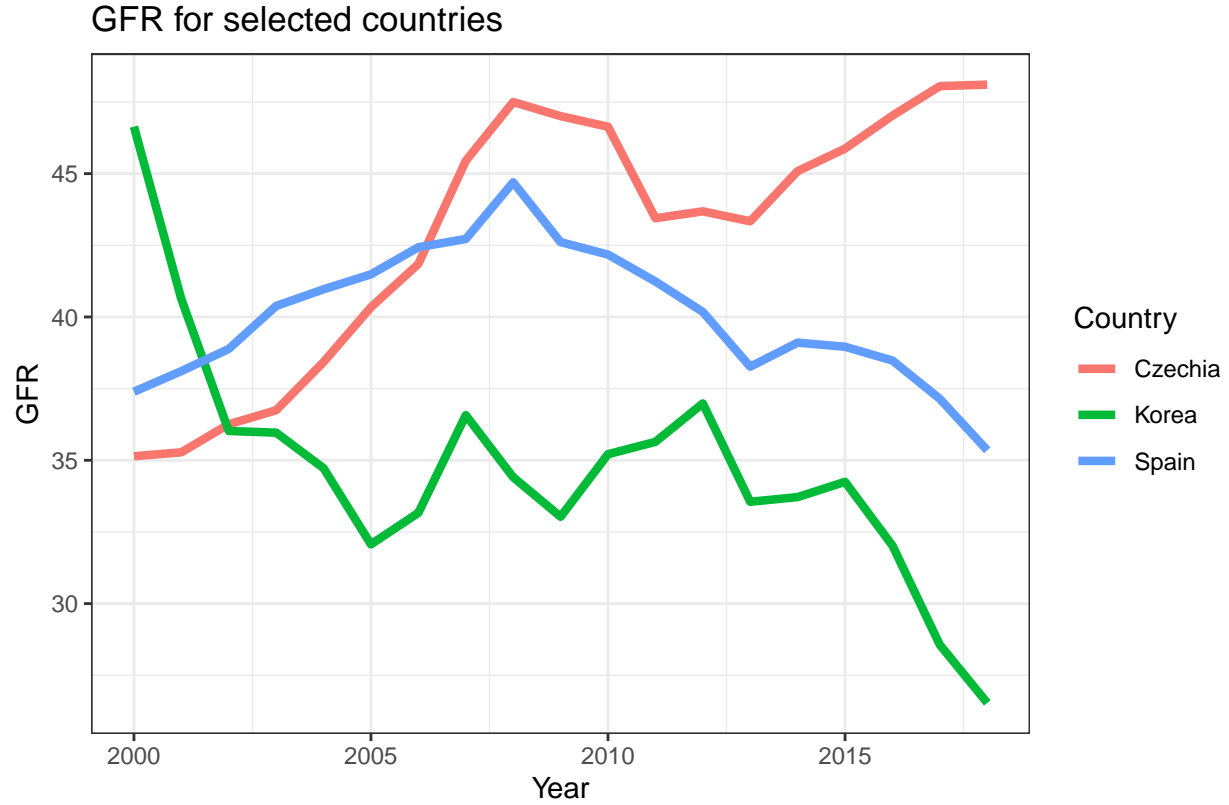
Source: HFD

As we can see, Czechia presents an increasing GFR for the period, contrasting to Korea's GFR which is decreasing and the lowest one for the selected countries. As for Spain, the GFR was increasing until the 2008/9 Financial Crisis, and from that point onward decreases.

Literature suggests that Korea has a decreasing fertility due to socioeconomic factors, especially pertaining to the domestic division of labor and the availability of State help to childcare (Kim 2017; Seo 2019). The variability of the time series could be related to tempo effects (Yoo and Sobotka 2018).

When analyzing Spain, most of literature relates fertility to socioeconomic factors (Barbieri et al. 2015; Sobotka, Skirbekk, and Philipov 2011). The GFR curve, with a peak around the Financial Crisis, shows this very clearly.

As for the case of Czechia, the behavior could be related to the country experiencing a Second Demographic Transition after the end of the USSR (Billingsley 2010). We can also observe that the Financial Crisis impacted the GFR of this country, but the increasing trend remains after the economic recovery (Matysiak, Sobotka, and Vignoli 2021).

**Time to Decompose**

We will be using the Kitagawa decomposition method (Kitagawa 1955) components to unravel changes in composition of the population under analysis from the changes in the rate.

Basically, Kitagawa tells us that a change of any rate `R = A * B` can be decompose in this way (Tønnessen 2019):

$$\Delta R = (\Delta A * \overline{B}) + (\Delta B * \overline{A})$$

The following plot shows the decomposition of the GFR for each country for each of the years observed:

Kitagawa decomposition: Czechia, Korea & Spain, 2000–2018
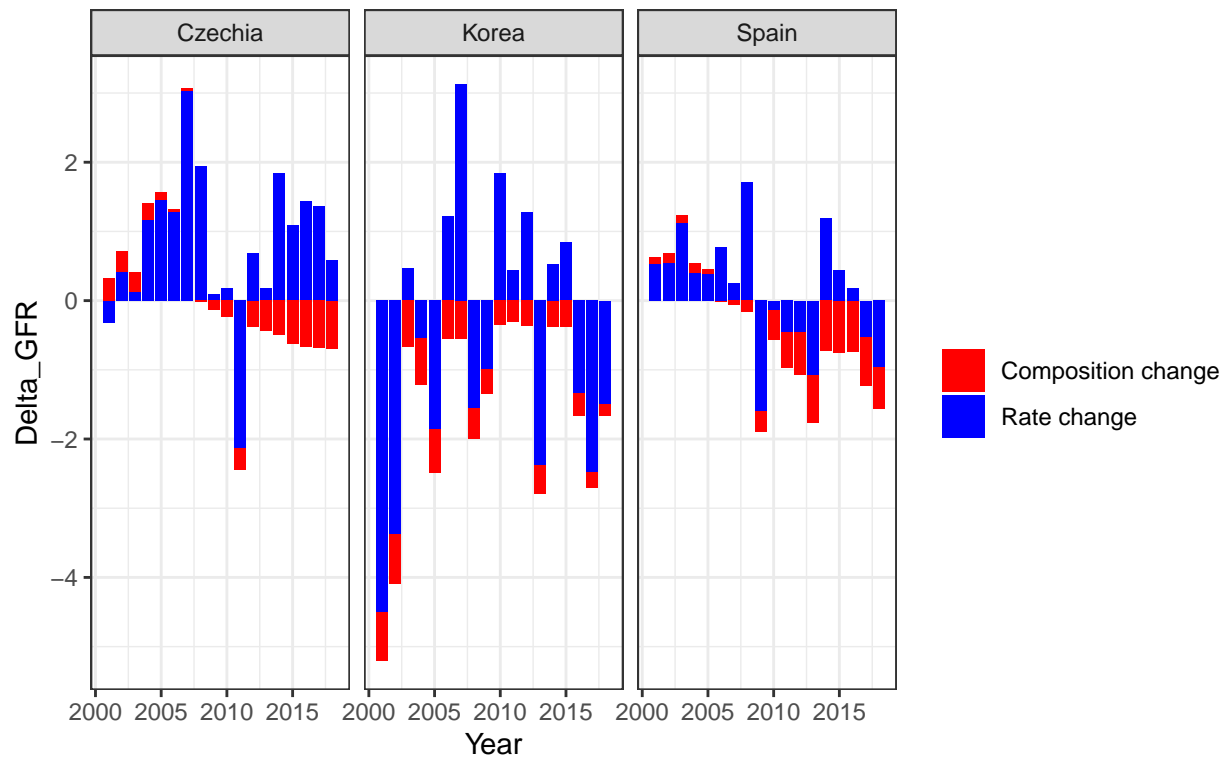
Source: HFD

We can observe that the source of change in the GFR for all countries is mostly due to changes in rate and not in population composition. But it is interesting that for both Czechia and Spain, the amount of change due to changes in composition has been increasing during this 2000-2018 period. This is related to the presence of aging populations, a common problem in European countries.

As for Korea, most part of the change in the GFR is related to changes in rate and not in population composition.

**Comparing Countries**

To finish this challenge, we can also decompose the difference in the GFR of Czechia and Korea. Since Czechia will be used as the second population and this country has experienced increasing GFR during the period, the decomposition should give positive numbers for most of the observed periods.

## Kitagawa decomposition: Czechia vs Korea, 2000–2018
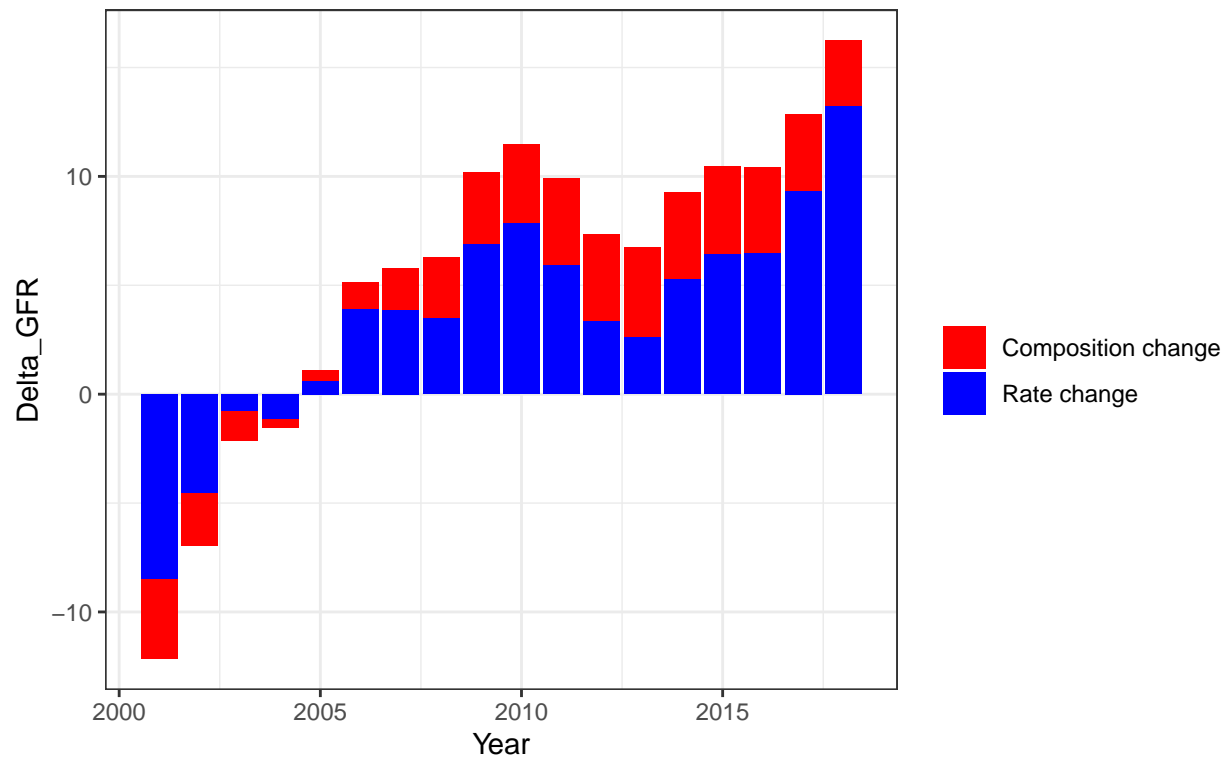


Source: HFD

The plot shows that the difference in GFR between the two countries is related more to changes in rates, rather than in population composition. But in those years where the change in rates is small, the composition factor plays a major role in explaining this differences.

## Challenge 3

**Get (any source) age-specific prevalences and apply the Sullivan method in R. Follow the practical guide if you want to and be aware of the method's limitations (e.g. same mortality schedule).**

```
options(scipen = 10)
```

Check the names of the countries and the the items available in the HMDweb.

```
# getHMDcountries()
# getHMDitemavail(CNTRY="LTU",username=HMD_user, password=HMD_pass)
# getHMDitemavail(CNTRY="NLD",username=HMD_user, password=HMD_pass)
```

We will extract data from Lithuania (LTU) and the Netherlands (NLD), the two European countries with the highest (9.6) and the lowest (3.1) gender gap in life expectancy in 2019 according to Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Life_expectancy_gender_gap_2021-01.jpg

We first extract data on Deaths and Exposures by 5 year age group from the HMD for Lithuania and the Netherlands.

```
## Get the Deaths 5x1 (by 5-year age group and year)
LTU.Dx <- readHMDweb(CNTRY="LTU",item="Deaths_5x1",
```

```
                            username=HMD_user, password=HMD_pass, fixup=F)
NLD.Dx <- readHMDweb(CNTRY="NLD",item="Deaths_5x1",
                            username=HMD_user, password=HMD_pass, fixup=F)


## Get the Exposures to risk 5x1 (by 5-year age group and year)
LTU.Exp <- readHMDweb(CNTRY="LTU",item="Exposures_5x1",
                            username=HMD_user, password=HMD_pass, fixup=F)
NLD.Exp <- readHMDweb(CNTRY="NLD",item="Exposures_5x1",
                            username=HMD_user, password=HMD_pass, fixup=F)


# We create a data frame of ages allowing to gather the age-groups over 85.
Ages <- data.frame(Age = c("0", "1-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
                            "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79",
                            "80-84", "85-89", "90-94", "95-99", "100-104", "105-109", "110+"),
                    age = c(0, 1, seq(5, 85, 5), rep(85,5)))
```

We compute the Age-Specific mortality rates by 5-year age groups from the HMD for each country and sex separately.

```
#### Lithuania mx for both sexes separately with 85+ as open-ended age interval ####

lithuania.Dx2 <- LTU.Dx %>%
  filter(Year == 2019) %>%
  select(Age, Female, Male) %>%
  pivot_longer(cols=c(Female, Male), names_to = "Sex", values_to = "Deaths")

lithuania.Exp2 <- LTU.Exp %>%
  filter(Year == 2019) %>%
  select(Age, Female, Male) %>%
  pivot_longer(cols=c(Female, Male), names_to = "Sex", values_to = "Exposures")

lithuania.mx <- lithuania.Dx2 %>%
  left_join(lithuania.Exp2, by = c("Age", "Sex")) %>%
  left_join(Ages) %>%
  group_by(Sex, age) %>%
  mutate(mx = sum(Deaths)/sum(Exposures)) %>%
  ungroup() %>%
  # We filter data after 90+ since the rate for mx85+ has already been computed
  filter(Age != "90-94", Age !="95-99", Age != "100-104", Age != "105-109", Age != "110+") %>%
  select(Sex, age, mx)

#### Netherlands mx for both sexes separately with 85+ as open-ended age interval ####

netherlands.Dx2 <- NLD.Dx %>%
  filter(Year == 2019) %>%
  select(Age, Female, Male) %>%
  pivot_longer(cols=c(Female, Male), names_to = "Sex", values_to = "Deaths")

netherlands.Exp2 <- NLD.Exp %>%
  filter(Year == 2019) %>%
  select(Age, Female, Male) %>%
  pivot_longer(cols=c(Female, Male), names_to = "Sex", values_to = "Exposures")
```

```
netherlands.mx <- netherlands.Dx2 %>%
  left_join(netherlands.Exp2, by = c("Age", "Sex")) %>%
  left_join(Ages) %>%
  group_by(Sex, age) %>%
  mutate(mx = sum(Deaths)/sum(Exposures)) %>%
  ungroup() %>%
  # We filter data after 90+ since the rate for mx85+ has already been computed
  filter(Age != "90-94", Age !="95-99", Age != "100-104", Age != "105-109", Age != "110+") %>%
  select(Sex, age, mx)


# Deleting intermediate tables
rm(lithuania.Dx2, lithuania.Exp2, netherlands.Dx2, netherlands.Exp2)
```

To apply the Sullivan method, we will use data on self-reported chronic morbidity (i.e. the presence of long-term (chronic) symptoms, health conditions or diseases) https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Chronic_morbidity.
This dimension of health is captured by one of the three questions of the Minimum European Health Module (MEHM): "Do you have any long-standing illness or health problem?". THe MEHM is currently implemented in the EU Statistics on Income and Living Conditions (EU-SILC) and the European Health Interview Survey (EHIS) https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Minimum_European_Health_Module_(MEHM) The EU-SILC target population consists of all individuals Individuals aged 16 years old and over living in private households. Hence, people living in collective households and in institutions are generally excluded from the survey.

Data on chronic morbidity from the EU-SILC is expressed as percentages within the population combining various breakdowns: sex, age-group, labour status, educational attainment level, country of birth, country of citizenship, degree of urbanisation and income quintile (group). https://ec.europa.eu/eurostat/web/health/methodology
Since we are only interested un the total population by sex and age-group, we will use the first data-set with labour status.

We extract the dataset of people having a long-standing illness or health problem, by sex, age and labour status (hlth_silc_04) and get a glimpse on it.

```
hlth_silc <- get_eurostat("hlth_silc_04", time_format = "num", type ="code")


# Get a glimpse on the data
# glimpse(hlth_silc)

# Check the values for wstatus and age
# unique(hlth_silc$wstatus)
# unique(hlth_silc$age)
```

We filter the data to keep only information for Lithuania and the Netherlands in 2009 for each sex separately, as well as the total population (wstatus == POP) and the age values corresponding to a distribution by 10 years age group, since it is the most regular age-grouping across the whole dataset.

```
hlth_silc2 <- hlth_silc %>%
  filter(time == 2019 & geo %in% c("LT","NL"), wstatus == "POP", sex %in% c("F","M"),
         age %in% c("Y16-24", "Y25-34", "Y35-44", "Y45-54","Y55-64", "Y65-74", "Y75-84", "Y_GE85")) %>%
  select(c(age, sex, geo, values)) %>%
  mutate(Sex = if_else(sex == "F", "Female", "Male"),
         age_gr = case_when(age == "Y16-24" ~ 15,
                            age == "Y25-34" ~ 25,
                            age == "Y35-44" ~ 35,
```

```
                                                age == "Y45-54" ~ 45,
                                                age == "Y55-64" ~ 55,
                                                age == "Y65-74" ~ 65,
                                                age == "Y75-84" ~ 75,
                                                TRUE ~ 85),
                    pix = values/100) %>%
  select(-c(age, sex, values))

lithuania.px <- hlth_silc2 %>%
  filter(geo == "LT") %>%
  select(-geo)

netherlands.px <- hlth_silc2 %>%
  filter(geo == "NL") %>%
  select(-geo)
```

Since the death rates are available by 5-year age group, but chronic morbidity prevalence (pix) by 10-year age group, we need to add a correspondence 10-year age group value to the death rates, allowing us to join both tables. We made the assumption that the prevalence is the same for the two 5-year age groups included in each 10-year each group. We also split each dataset by sex.

```
# Lithuania data frame with mx and pix for 15-85
lithuania <- lithuania.mx %>%
  mutate(age_gr = case_when(age %in% c(0, 1, 5, 10) ~ 0,
                            age %in% c(15,20) ~ 15,
                            age %in% c(25,30) ~ 25,
                            age %in% c(35,40) ~ 35,
                            age %in% c(45,50) ~ 45,
                            age %in% c(55,60) ~ 55,
                            age %in% c(65,70) ~ 65,
                            age %in% c(75,80) ~ 75,
                            age %in% c(75,80) ~ 75,
                            TRUE ~ 85)) %>%
  left_join(lithuania.px, by= c("Sex", "age_gr")) %>%
  filter(age_gr != 0) %>%
  select(-age_gr)

# We split the dataset by sex
lithuania.f <- lithuania %>%
  filter(Sex=="Female") %>% select(-Sex) %>% rename(nmx=mx)
lithuania.m <- lithuania %>%
  filter(Sex=="Male") %>% select(-Sex) %>% rename(nmx=mx)
# Netherlands data frame with mx and pix for 15-85
netherlands <- netherlands.mx %>%
  mutate(age_gr = case_when(age %in% c(0, 1, 5, 10) ~ 0,
                            age %in% c(15,20) ~ 15,
                            age %in% c(25,30) ~ 25,
                            age %in% c(35,40) ~ 35,
                            age %in% c(45,50) ~ 45,
                            age %in% c(55,60) ~ 55,
                            age %in% c(65,70) ~ 65,
                            age %in% c(75,80) ~ 75,
                            age %in% c(75,80) ~ 75,
                            TRUE ~ 85)) %>%
```

```
  left_join(netherlands.px, by= c("Sex", "age_gr")) %>%
  filter(age_gr != 0) %>%
  select(-age_gr)

# We split the dataset by sex
netherlands.f <- netherlands %>%
  filter(Sex=="Female") %>% select(-Sex) %>% rename(nmx=mx)
netherlands.m <- netherlands %>%
  filter(Sex=="Male") %>% select(-Sex) %>% rename(nmx=mx)
# Delete original data
rm(LTU.Dx, LTU.Exp, NLD.Dx, NLD.Exp, hlth_silc, hlth_silc2)

# Delete partial tables
rm(Ages, lithuania.mx, lithuania.px, lithuania, netherlands.mx, netherlands.px, netherlands)
```

To apply the Sullivan Method, we will consider the non-presence of chronic morbidity as an indicator of health status. We create a function allowing to compute general Life Expectancy (LE) and Morbidity Free Life Expectancy (MFLE) above age 15.

```
Sullivan <- function(age, nmx, pix){

  # Assign n and nax values
  n      <- c(diff(age), 999)
  nax    <- 0.5 * n

  # Computation of the life table e0
  nqx          <- (n * nmx)/(1 + (n - nax) * nmx)
  nqx          <- c(nqx[-(length(nqx))], 1)
  nqx[nqx > 1] <- 1
  npx          <- 1 - nqx
  lx           <- cumprod(c(1, npx))
  ndx          <- -diff(lx)
  lxpn         <- lx[-1]
  nLxpn        <- n * lxpn + ndx * nax
  nLx          <- c(nLxpn[-length(nLxpn)],
                    lxpn[length(lxpn)-1]/nmx[length(nmx)])
  Tx           <- rev(cumsum(rev(nLx)))
  lx           <- lx[1:length(age)]
  ex           <- Tx/lx
  e15           <- ex[1]

  # Sullivan method (Morbidity-Free Life Expectancy)
  nLx.MF        <- nLx*(1-pix)
  Tx.MF         <- rev(cumsum(rev(nLx.MF)))
  ex.MF         <- (Tx.MF)/lx
  e15.MF         <- ex.MF[1]

  return(data.frame(Age = age,
                    LE = ex,
                    MFLE = ex.MF))
}
```

We apply the Sullivan function, to each dataset to compute the LE and the MFLE. We merge all datasets in a unique one to plot.

```
# Lithuania
MFLE.ltu.f <- Sullivan(age=lithuania.f$age, nmx=lithuania.f$nmx,
                       pix=lithuania.f$pix) %>%
  mutate(Sex = "Females")
MFLE.ltu.m <- Sullivan(age=lithuania.m$age, nmx=lithuania.m$nmx,
                       pix=lithuania.f$pix)  %>%
  mutate(Sex = "Males")

MFLE.ltu <- bind_rows(MFLE.ltu.f, MFLE.ltu.m) %>%
  pivot_longer(cols = c("LE", "MFLE"), names_to = "Measure", values_to = "Years") %>%
  mutate(Country = "Lithuania")

## Netherlands
MFLE.nld.f <- Sullivan(age=netherlands.f$age, nmx=netherlands.f$nmx,
                       pix=netherlands.f$pix) %>%
  mutate(Sex = "Females")
MFLE.nld.m <- Sullivan(age=netherlands.m$age, nmx=netherlands.m$nmx,
                       pix=netherlands.f$pix)  %>%
  mutate(Sex = "Males")

MFLE.nld <- bind_rows(MFLE.nld.f, MFLE.nld.m) %>%
  pivot_longer(cols = c("LE", "MFLE"), names_to = "Measure", values_to = "Years") %>%
  mutate(Country = "Netherlands")

MFLE <- bind_rows(MFLE.ltu, MFLE.nld)

# Remove intermediate tables
rm("lithuania.f", "lithuania.m", "MFLE.ltu", "MFLE.ltu.f", "MFLE.ltu.m",
   "MFLE.nld", "MFLE.nld.f", "MFLE.nld.m", "netherlands.f", "netherlands.m")
```
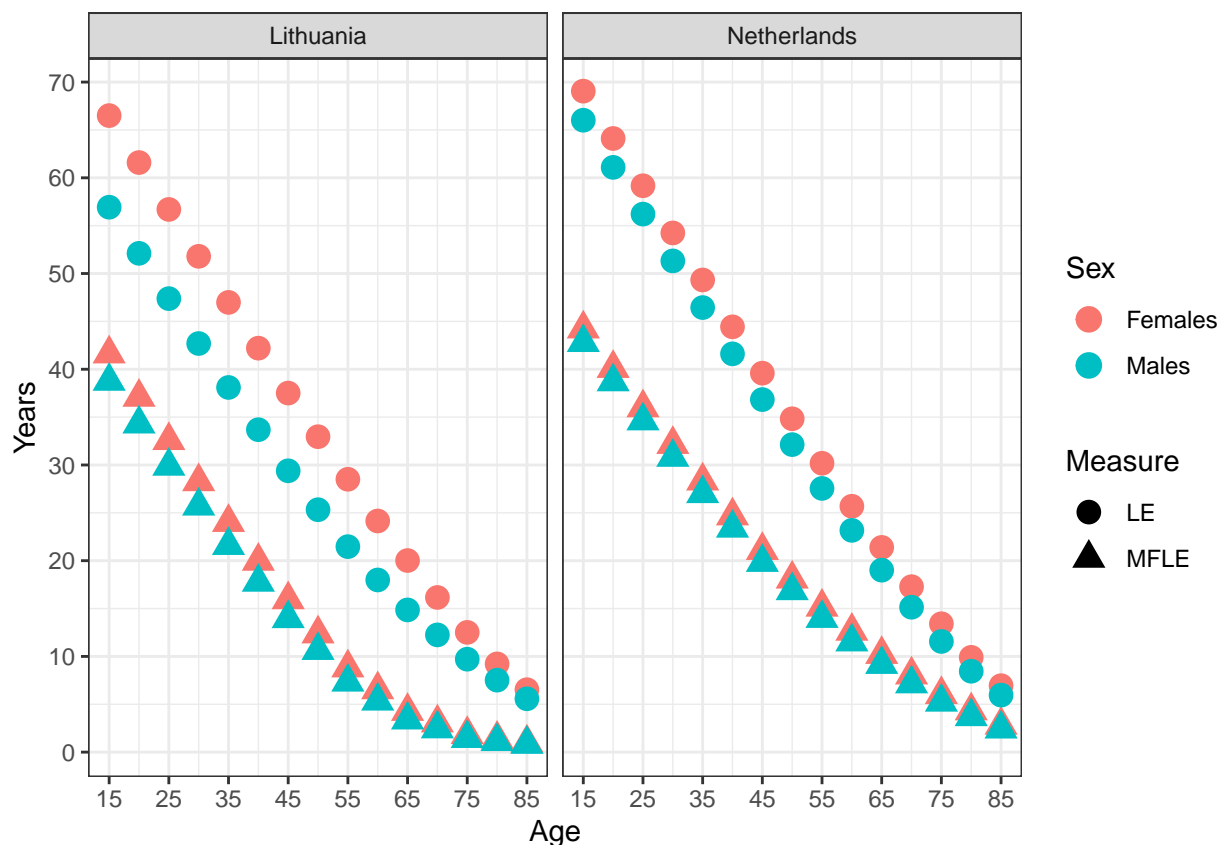
We can now plot all together.

**Remaining Life Expectancy (LE) and Morbidity Free Life Expectancy (MFLE) by sex in Lithuania and the Netherlands in 2019.**

```
ggplot(MFLE, aes(x=Age, y = Years, color = Sex, shape = Measure)) +
  theme_bw() +
  facet_grid(. ~ Country) +
  geom_point(size = 4) +
  scale_x_continuous(breaks = seq(15,85, by=10)) +
  scale_y_continuous(breaks = seq(0,70, by=10))
```

## Challenge 4

**Use the linear integral model to decompose the change in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death for 3 countries you might be interested in (over time or between them). Interpret the results of life expectancy alongside standard deviation. Make it interesting. You can use data from HCoD, HMD, WHO, GBD.**

In this exercise, we aim to see the changes in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death in Latvia, Russia and Poland from the Soviet Union dissolution to becoming independent states. Since the impact of macrolevel societal changes can take time to be observed in demographic behavior, we decided to look at the change from 1990 to 2010.

We will use the data from the Human Mortality Database (HMD) and the Cause of Death Database (CDD). As requested, we benefited from the Horiuchi and colleagues' (2008) linear integral decomposition model to decompose the changes in standard deviation and life expectancy. First we will look at it only by age and then with cause of mortality as well.

We first needed to arrange the dataset to make it ready for decomposition. The code for this is available on the rmd.

```
########## DATA PREPARATION: LATVIA ##########

## Data All Cause ####

# Obtain Deaths and Exposures
latvia.death <- readHMDweb(CNTRY="LVA",item="Deaths_5x1",
                           username=HMD_user,password=HMD_pass,fixup=T)
```

```r
latvia.expo <- readHMDweb(CNTRY="LVA",item="Exposures_5x1",
                          username=HMD_user,password=HMD_pass,fixup=T)

latvia.death <- latvia.death %>%
  select("Year", "Age", "Male") %>%
  rename(Death = Male)
latvia.expo <- latvia.expo %>%
  select("Year", "Age", "Male") %>%
  rename(Exposure = Male)

latvia <- left_join(latvia.death, latvia.expo, by = c("Year", "Age"))

## Data Wrangling All Cause ####

# Death rates in 1990
latvia.90 <- subset(latvia, Year ==1990)
latvia.90 <- latvia.90 %>%
  mutate(Death =
           ifelse(latvia.90$Age == 85,
                  sum(latvia.90[latvia.90$Age %in% c(85:110),]$Death),
                  latvia.90$Death),
         Exposure =
           ifelse(latvia.90$Age == 85,
                  sum(latvia.90[latvia.90$Age %in% c(85:110),]$Exposure),
                  latvia.90$Exposure),
         mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)

# Death rates in 2010
latvia.10 <- subset(latvia, Year ==2010)
latvia.10 <- latvia.10 %>%
  mutate(Death =
           ifelse(latvia.10$Age == 85,
                  sum(latvia.10[latvia.10$Age %in% c(85:110),]$Death),
                  latvia.10$Death),
         Exposure =
           ifelse(latvia.10$Age == 85,
                  sum(latvia.10[latvia.10$Age %in% c(85:110),]$Exposure),
                  latvia.10$Exposure),
         mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)


## Data by Cause of death ####
latvia.cod <- read.csv("LVA_m_short_idr.csv")
latvia.cod.90 <- latvia.cod %>%
  filter(sex==1, year==1990,) %>%
  select( - c(sex, list, agf))

latvia.cod.10 <- latvia.cod %>%
  filter(sex==1, year==2010,) %>%
```

```r
    select( - c(sex, list, agf))


## Data Wrangling Cause of Death Latvia ####

# Death rates by cause in 1990
latvia.c.90 <- as.data.frame(t(latvia.cod.90))
latvia.c.90 <- as.data.frame(lapply(latvia.c.90, as.numeric))
colnames(latvia.c.90) <- latvia.c.90[1,]
latvia.c.90 <- latvia.c.90[2:20,]
latvia.c.90$Age <- c(0,1,seq(5,85,5))
latvia.c.90$Year <- 1990
latvia.c.90[,c(2:17)] <- latvia.c.90[,c(2:17)]/1000000
latvia.c.90<- latvia.c.90[,c(2:19)]
row.names(latvia.c.90) <- NULL

# Death rates by cause in 2010
latvia.c.10 <- as.data.frame(t(latvia.cod.10))
latvia.c.10 <- as.data.frame(lapply(latvia.c.10, as.numeric))
colnames(latvia.c.10) <- latvia.c.10[1,]
latvia.c.10 <- latvia.c.10[2:20,]
latvia.c.10$Age <- c(0,1,seq(5,85,5))
latvia.c.10$Year <- 1990
latvia.c.10[,c(2:17)] <- latvia.c.10[,c(2:17)]/1000000
latvia.c.10<- latvia.c.10[,c(2:19)]
row.names(latvia.c.10) <- NULL

########## DATA PREPARATION: RUSSIA ##########

## Data All Cause ####

# Obtain Deaths and Exposures
russia.death <- readHMDweb(CNTRY="RUS",item="Deaths_5x1",
                     username=HMD_user,password=HMD_pass, fixup=T)
russia.expo <- readHMDweb(CNTRY="RUS",item="Exposures_5x1",
                        username=HMD_user,password=HMD_pass,fixup=T)

russia.death <- russia.death %>%
  select("Year", "Age", "Male") %>%
  rename(Death = Male)
russia.expo <- russia.expo %>%
  select("Year", "Age", "Male") %>%
  rename(Exposure = Male)

russia <- left_join(russia.death, russia.expo, by = c("Year", "Age"))

## Data Wrangling All Cause ####

# Death rates in 1990
russia.90 <- subset(russia, Year ==1990)
russia.90 <- russia.90 %>%
  mutate(Death =
           ifelse(russia.90$Age == 85,
                  sum(russia.90[russia.90$Age %in% c(85:110),]$Death),
```

```
                          russia.90$Death),
                 Exposure =
                   ifelse(russia.90$Age == 85,
                             sum(russia.90[russia.90$Age %in% c(85:110),]$Exposure),
                             russia.90$Exposure),
                 mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)

# Death rates in 2010
russia.10 <- subset(russia, Year ==2010)
russia.10  <- russia.10  %>%
  mutate(Death =
             ifelse(russia.10 $Age == 85,
                       sum(russia.10 [russia.10 $Age %in% c(85:110),]$Death),
                       russia.10$Death),
             Exposure =
               ifelse(russia.10 $Age == 85,
                         sum(russia.10 [russia.10 $Age %in% c(85:110),]$Exposure),
                         russia.10$Exposure),
             mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)


## Data by Cause of death ####
russia.cod <- read.csv("RUS_m_short_idr.csv")
russia.cod.90 <- russia.cod %>%
  filter(sex==1, year==1990,) %>%
  select( - c(sex, list, agf))

russia.cod.10 <- russia.cod %>%
  filter(sex==1, year==2010,) %>%
  select( - c(sex, list, agf))


## Data Wrangling Cause of Death Russia ####

# Death rates by cause in 1990
russia.c.90 <- as.data.frame(t(russia.cod.90))
russia.c.90 <- as.data.frame(lapply(russia.c.90, as.numeric))
colnames(russia.c.90) <- russia.c.90[1,]
russia.c.90 <- russia.c.90[2:20,]
russia.c.90$Age <- c(0,1,seq(5,85,5))
russia.c.90$Year <- 1990
russia.c.90[,c(2:17)] <- russia.c.90[,c(2:17)]/1000000
russia.c.90<- russia.c.90[,c(2:19)]
row.names(russia.c.90) <- NULL

# Death rates by cause in 2010
russia.c.10 <- as.data.frame(t(russia.cod.10 ))
russia.c.10 <- as.data.frame(lapply(russia.c.10, as.numeric))
colnames(russia.c.10) <- russia.c.10[1,]
```

```r
russia.c.10 <- russia.c.10[2:20,]
russia.c.10$Age <- c(0,1,seq(5,85,5))
russia.c.10$Year <- 1990
russia.c.10[,c(2:17)] <- russia.c.10[,c(2:17)]/1000000
russia.c.10<- russia.c.10[,c(2:19)]
row.names(russia.c.10) <- NULL
```

```r
########## DATA PREPARATION: POLAND ##########

## Data All Cause ####

# Obtain Deaths and Exposures
poland.death <- readHMDweb(CNTRY="POL",item="Deaths_5x1",
                           username=HMD_user,password=HMD_pass,fixup=T)
poland.expo <-  readHMDweb(CNTRY="POL",item="Exposures_5x1",
                           username=HMD_user,password=HMD_pass,fixup=T)


poland.death <- poland.death %>%
  select("Year", "Age", "Male") %>%
  rename(Death = Male)
poland.expo <- poland.expo %>%
  select("Year", "Age", "Male") %>%
  rename(Exposure = Male)

poland <- left_join(poland.death, poland.expo, by = c("Year", "Age"))

## Data Wrangling All Cause ####

# Death rates in 1990
poland.90 <- subset(poland, Year ==1990)
poland.90  <- poland.90  %>%
  mutate(Death =
           ifelse(poland.90$Age == 85,
                  sum(poland.90[poland.90 $Age %in% c(85:110),]$Death),
                  poland.90$Death),
         Exposure =
           ifelse(poland.90 $Age == 85,
                  sum(poland.90[poland.90 $Age %in% c(85:110),]$Exposure),
                  poland.90$Exposure),
         mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)

# Death rates in 2010
poland.10 <- subset(poland, Year ==2010)
poland.10  <- poland.10  %>%
  mutate(Death =
           ifelse(poland.10$Age == 85,
                  sum(poland.10[poland.10 $Age %in% c(85:110),]$Death),
                  poland.10$Death),
         Exposure =
           ifelse(poland.10 $Age == 85,
                  sum(poland.10[poland.10 $Age %in% c(85:110),]$Exposure),
                  poland.10$Exposure),
```

```r
        mx = Death /Exposure) %>%
  filter(Age < 90) %>%
  select(Year, Age, mx)

## Data by Cause of death ####
poland.cod <- read.csv("POL_m_short_idr.csv")
poland.cod.90 <- poland.cod %>%
  filter(sex==1, year==1990,) %>%
  select( - c(sex, list, agf))

poland.cod.10 <- poland.cod %>%
  filter(sex==1, year==2010,) %>%
  select( - c(sex, list, agf))

## Data Wrangling Cause of Death Poland ####

# Death rates by cause in 1990
poland.c.90 <- as.data.frame(t(poland.cod.90))
poland.c.90<- as.data.frame(lapply(poland.c.90, as.numeric))
colnames(poland.c.90) <- poland.c.90[1,]
poland.c.90 <- poland.c.90[2:20,]
poland.c.90$Age <- c(0,1,seq(5,85,5))
poland.c.90$Year <- 1990
poland.c.90[,c(2:17)] <- poland.c.90[,c(2:17)]/1000000
poland.c.90<- poland.c.90[,c(2:19)]
row.names(poland.c.90) <- NULL

# Death rates by cause in 2010
poland.c.10 <- as.data.frame(t(poland.cod.10 ))
poland.c.10 <- as.data.frame(lapply(poland.c.10 , as.numeric))
colnames(poland.c.10 ) <- poland.c.10[1,]
poland.c.10  <- poland.c.10 [2:20,]
poland.c.10 $Age <- c(0,1,seq(5,85,5))
poland.c.10 $Year <- 1990
poland.c.10 [,c(2:17)] <- poland.c.10 [,c(2:17)]/(1000000)
poland.c.10 <- poland.c.10 [,c(2:19)]
row.names(poland.c.10) <- NULL
```

```r
#load some functions and some info for graphs
source('Functions_4.R')

# Function for the standard deviation of the age at death distribution and life expectancy

# e0.frommx ####
e0.frommx <- function(nmx =  mx, sex=1, age = c(0, 1, seq(5, 85, 5)), nax = NULL){
  n    <- c(diff(age), 999)

  if (is.null(nax)) {
    nax <- 0.5 * n
    if (n[2] == 4) {
      if (sex == 1) {
        if (nmx[1] >= 0.107) {
          nax[1] <- 0.33
          nax[2] <- 1.352
```

```r
    }
    else {
      nax[1] <- 0.045 + 2.684 * nmx[1]
      nax[2] <- 1.651 - 2.816 * nmx[1]
    }
  }
  if (sex == 2) {
    if (nmx[1] >= 0.107) {
      nax[1] <- 0.35
      nax[2] <- 1.361
    }
    else {
      nax[1] <- 0.053 + 2.8 * nmx[1]
      nax[2] <- 1.522 - 1.518 * nmx[1]
    }
  }
}
}
nqx          <- (n * nmx)/(1 + (n - nax) * nmx)
nqx          <- c(nqx[-(length(nqx))], 1)
nqx[nqx > 1] <- 1

npx <- 1 - nqx
lx <- cumprod(c(1, npx))
ndx <- -diff(lx)
lxpn <- lx[-1]
nLxpn <- n * lxpn + ndx * nax
nLx <- c(nLxpn[-length(nLxpn)], lxpn[length(lxpn)-1]/nmx[length(nmx)])
Tx <- rev(cumsum(rev(nLx)))
lx <- lx[1:length(age)]
ex <- Tx/lx
e0 <- ex[1]

return(e0)
}

# Decomp ####

Decomp <-function (func, rates1, rates2, N, ...) {
  y1 <- func(rates1, ...)
  y2 <- func(rates2, ...)
  d <- rates2 - rates1
  n <- length(rates1)
  delta <- d/N
  x <- rates1 + d * matrix(rep(0.5:(N - 0.5)/N, length(rates1)),
                          byrow = TRUE, ncol = N)
  cc <- matrix(0, nrow = n, ncol = N)
  for (j in 1:N) {
    for (i in 1:n) {
      z <- rep(0, n)
      z[i] <- delta[i]/2
      cc[i, j] <- func((x[, j] + z), ...) - func((x[, j] -
                                                  z), ...)
```

```r
      }
    }
    return(rowSums(cc))
}

# sd.frommx ####

sd.frommx <- function(nmx =  mx, sex=1, age = c(0, 1, seq(5, 85, 5)), nax = NULL){
  n    <- c(diff(age), 999)

  if (is.null(nax)) {
    nax <- 0.5 * n
    if (n[2] == 4) {
      if (sex == 1) {
        if (nmx[1] >= 0.107) {
          nax[1] <- 0.33
          nax[2] <- 1.352
        }
        else {
          nax[1] <- 0.045 + 2.684 * nmx[1]
          nax[2] <- 1.651 - 2.816 * nmx[1]
        }
      }
      if (sex == 2) {
        if (nmx[1] >= 0.107) {
          nax[1] <- 0.35
          nax[2] <- 1.361
        }
        else {
          nax[1] <- 0.053 + 2.8 * nmx[1]
          nax[2] <- 1.522 - 1.518 * nmx[1]
        }
      }
    }
  }
  nqx          <- (n * nmx)/(1 + (n - nax) * nmx)
  nqx          <- c(nqx[-(length(nqx))], 1)
  nqx[nqx > 1] <- 1

  npx <- 1 - nqx
  lx <- cumprod(c(1, npx))
  ndx <- -diff(lx)
  lxpn <- lx[-1]
  nLxpn <- n * lxpn + ndx * nax
  nLx <- c(nLxpn[-length(nLxpn)], lxpn[length(lxpn)-1]/nmx[length(nmx)])
  Tx <- rev(cumsum(rev(nLx)))
  lx <- lx[1:length(age)]
  ex <- Tx/lx
  nax[length(nax)] <- ex[length(ex)]
  vx <- sum(ndx*(age+nax-ex[1L])^2)
  sd <- sqrt(vx)
  return(sd)
}
```

```r
# sdfrommxc ####

sdfrommxc <- function(mxcvec,sex=1){
  dim(mxcvec) <- c(19,length(mxcvec)/19)
  mx          <- rowSums(mxcvec)
  sd.frommx(mx,sex)
}

# Saving the years as vectors ####

#Russia
russia.mx1 <- russia.90$mx
russia.mx2 <- russia.10$mx
#Poland
poland.mx1 <- poland.90$mx
poland.mx2 <- poland.10$mx
#Latvia
latvia.mx1 <- latvia.90$mx
latvia.mx2 <- latvia.10$mx

########## DECOMPOSITION: LIFE EXPECTANCY ##########

## Decomposition ####

# Latvia from 1990 to 2010
latvia.results <- horiuchi(func = e0.frommx, pars1 = latvia.mx1,
                           pars2 = latvia.mx2,N = 100)
# Russia from 1990 to 2010
russia.results <- horiuchi(func = e0.frommx, pars1 = russia.mx1,
                           pars2 = russia.mx2,N = 100)
# Poland from 1990 to 2010
poland.results <- horiuchi(func = e0.frommx, pars1 = poland.mx1,
                           pars2 = poland.mx2,N = 100)

## Check ####

# Original
russia.original <- e0.frommx(russia.mx2) - e0.frommx(russia.mx1)
latvia.original <- e0.frommx(latvia.mx2) - e0.frommx(latvia.mx1)
poland.original <- e0.frommx(poland.mx2) - e0.frommx(poland.mx1)

# From the decomposition
russia.with.decomp <- sum(russia.results)
latvia.with.decomp <- sum(latvia.results)
poland.with.decomp <- sum(poland.results)

# Comparison
c(latvia.original, latvia.with.decomp,
  russia.original, russia.with.decomp,
  poland.original, poland.with.decomp)
```

```
## [1]  3.7151230  3.7151224 -0.7273611 -0.7273611  5.9420555  5.9420509
```

```
# Errors
c(latvia.with.decomp - latvia.original,
  russia.with.decomp - russia.original,
  poland.with.decomp - poland.original)
```

```
## [1] -0.00000057451298  0.00000002265514 -0.00000455006681
```
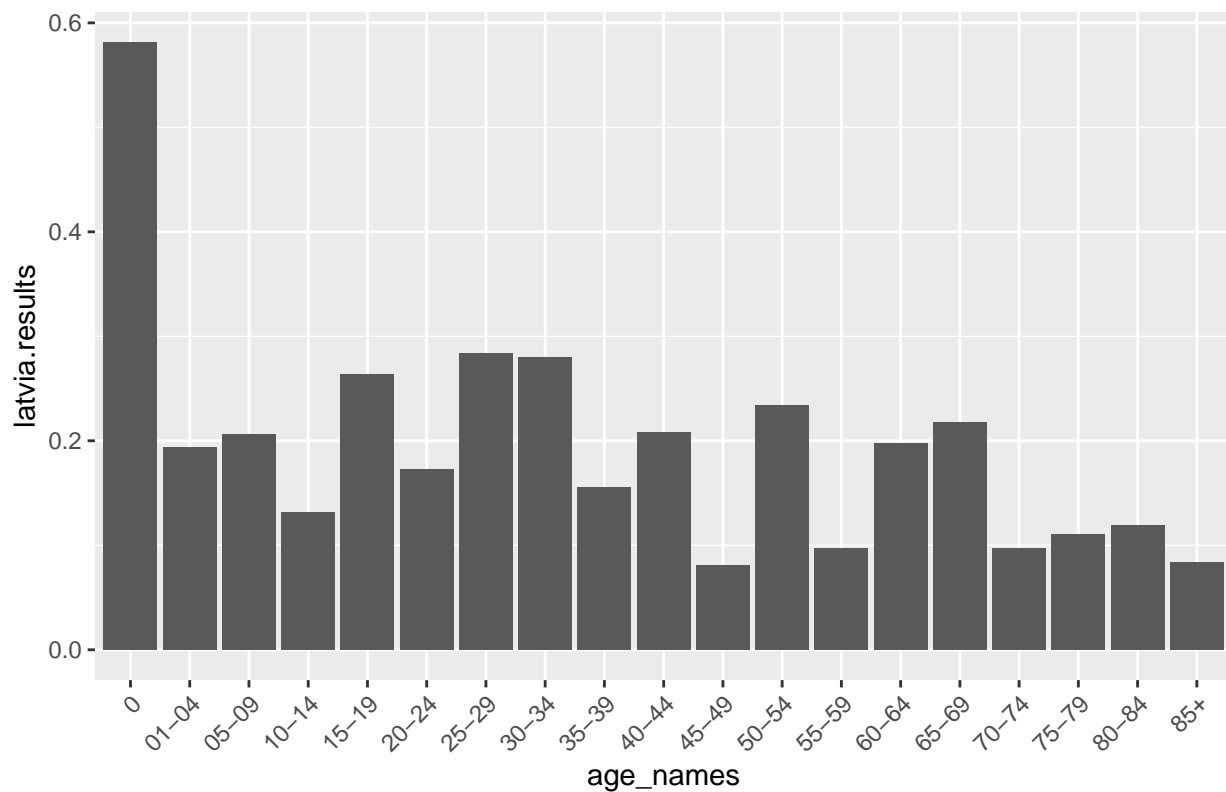
The graphs

```
## Labels for age-groups ####
age_names<-as.factor(c("0"="0","01"="01-04","05"="05-09","10"="10-14",
                       "15"="15-19","20"="20-24","25"="25-29",
                       "30"="30-34","35"="35-39","40"="40-44",
                       "45"="45-49","50"="50-54","55"="55-59",
                       "60"="60-64","65"="65-69","70"="70-74",
                       "75"="75-79","80"="80-84","85"="85+"))

## Graphs ####

# Latvia
ggplot()+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010' ))+
  geom_bar(aes(x = age_names, y= latvia.results), stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Russia
ggplot()+
```

```
ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010' ))+
geom_bar(aes(x = age_names, y= russia.results), stat = "identity", position = "stack") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Change in $e_0$ 1990–2010
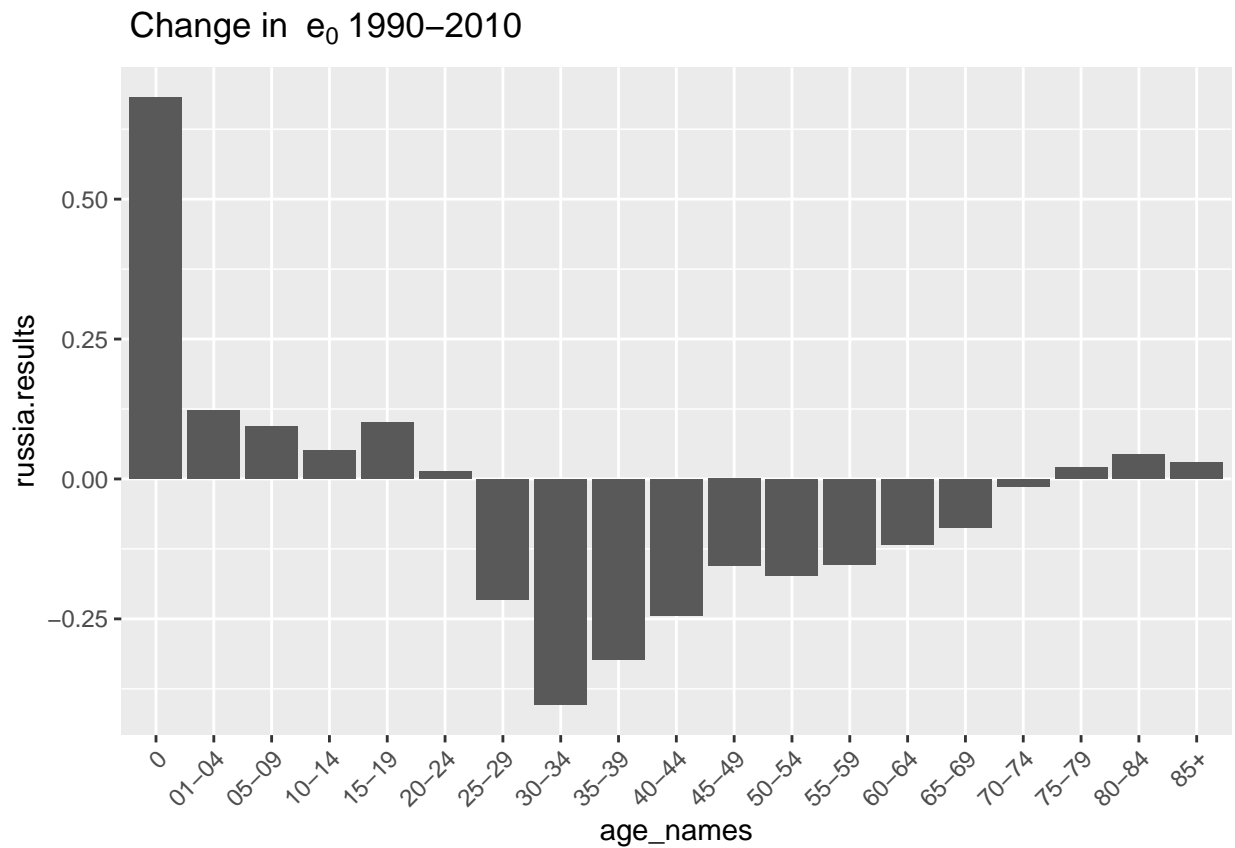


```
# Poland
ggplot()+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010' ))+
  geom_bar(aes(x = age_names, y= poland.results), stat = "identity", position = "stack")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

## Change in $e_0$ 1990–2010
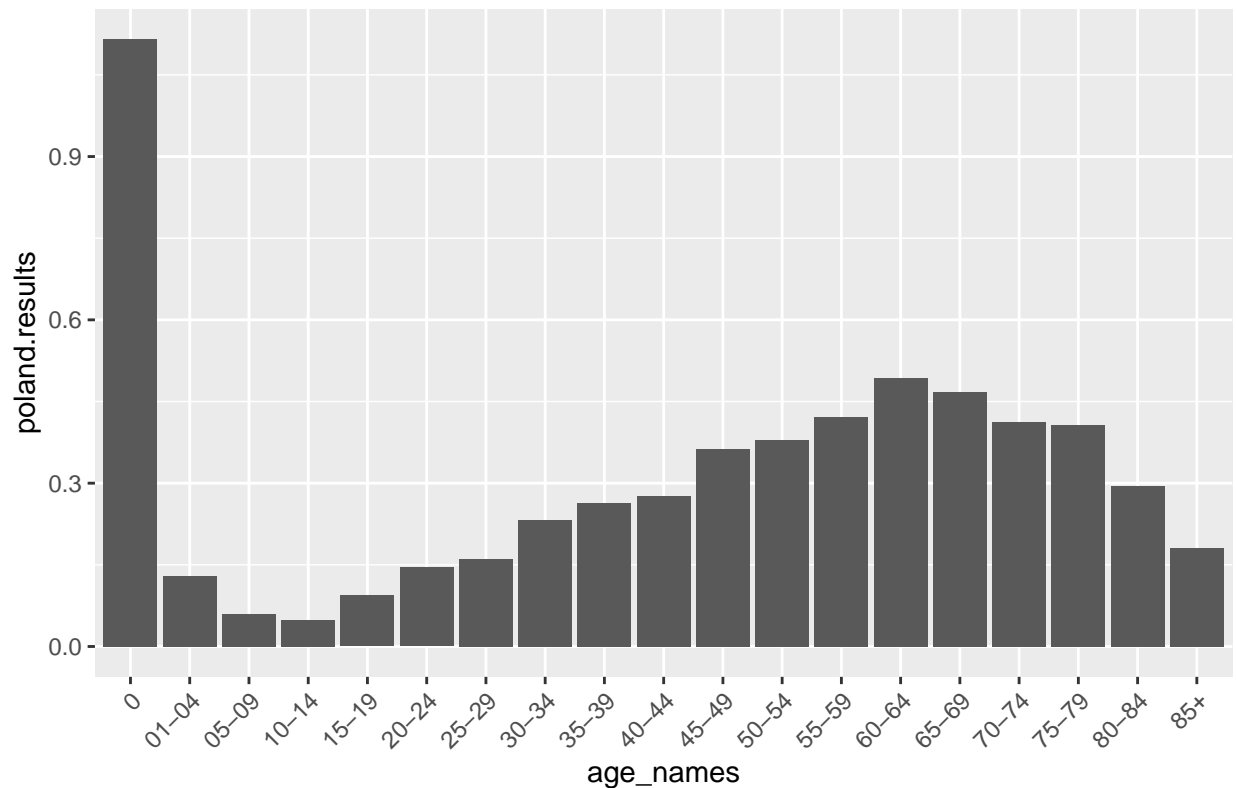


Unlike the graph for Russia, Latvia and Poland shows that there has been overall a positive change in the life expactancy.In all of the countries, the age group that has experienced the highest gain in life expectancy is the 0-5 ages. This is expected considering the increase in availability of communicable disease treatment for the infants and children. Russia shows a substantial decrease in life expectancy for the middle age groups. It should be noted that there are big differences between Russian males and females and we are only observing males here. The increase in alcohol consumtopn and smoking can be a reason for that and additionally the economic and political instability was much prolonged in the case of Russia after the dissolution of Soviet Union, and these can also be considered as causes of changes in the life expectancy.

We now extend our findings to cause of death decomposition to see which cause of death plays major role in the life expectancy changes in Latvia, Russia and Poland from 1990 t0 2010.

```
### Extending to cause specific results

## Load some functions and some info for graphs ####
source('Functions_5.R')

e0frommxc <- function(mxcvec,sex=1){
  dim(mxcvec) <- c(19,length(mxcvec)/19)
  mx <- rowSums(mxcvec)
  e0.frommx(mx,sex)
}

## Age-cause specific mortality rates in each period ####

# Latvia
COD1.latvia <- as.matrix(latvia.c.90[,1:16])
```

```r
COD2.latvia <- as.matrix(latvia.c.10[,1:16])

# Russia
COD1.russia <- as.matrix(russia.c.90[,1:16])
COD2.russia <- as.matrix(russia.c.10[,1:16])

# Poland
COD1.poland <- as.matrix(poland.c.90[,1:16])
COD2.poland  <- as.matrix(poland.c.10[,1:16])
```

```r
########## DECOMPOSITION: LIFE EXPECTANCY ##########

## Decomposition ####
results.c.latvia <- horiuchi(func = e0frommxc, pars1 = c(COD1.latvia),
                             pars2 = c(COD2.latvia), N = 100)
results.c.russia <- horiuchi(func = e0frommxc, pars1 = c(COD1.russia),
                             pars2 = c(COD2.russia), N = 100)
results.c.poland <- horiuchi(func = e0frommxc, pars1 = c(COD1.poland),
                             pars2 = c(COD2.poland), N = 100)


## Data wrangling ####

# Latvia
# Go back to a matrix
dim(results.c.latvia) <- dim(COD1.latvia)
# original
original.c.latvia <- e0frommxc(COD2.latvia) - e0frommxc(COD1.latvia)
# with decomp
with.decomp.c.latvia <- sum(results.c.latvia)

# Russia
# Go back to a matrix
dim(results.c.russia) <- dim(COD1.russia)
# original
original.c.russia <- e0frommxc(COD2.russia) - e0frommxc(COD1.russia)
# with decomp
with.decomp.c.russia <- sum(results.c.russia)

# Poland
# Go back to a matrix
dim(results.c.poland) <- dim(COD1.poland)
# original
original.c.poland <- e0frommxc(COD2.poland) - e0frommxc(COD1.poland)
# with decomp
with.decomp.c.poland <- sum(results.c.poland)

## Check ####

# Original
russia.original <- e0.frommx(russia.mx2) - e0.frommx(russia.mx1)
latvia.original <- e0.frommx(latvia.mx2) - e0.frommx(latvia.mx1)
poland.original <- e0.frommx(poland.mx2) - e0.frommx(poland.mx1)

# From the decomposition
```

```r
russia.with.decomp <- sum(russia.results)
latvia.with.decomp <- sum(latvia.results)
poland.with.decomp <- sum(poland.results)

# Comparison
c(original.c.latvia, with.decomp.c.latvia,
  original.c.russia, with.decomp.c.russia,
  original.c.poland, with.decomp.c.poland)
```

```
## [1]  6.118467  6.118466  3.198322  3.198322 11.577461 11.577441
```

```r
# Errors
c(with.decomp.c.latvia - original.c.latvia,
  with.decomp.c.russia - original.c.russia,
  with.decomp.c.poland - original.c.poland)
```

```
## [1] -0.00000081415119 -0.00000006621032 -0.00001934737402
```

The graphs for the life expectancies decomposed by age and cause of death are below.

```r
## Labels ####
cause_names<-c("1"="Infectious diseases",
               "2"="Neoplasm",
               "3"="Diseases of blood & blood-forming organs",
               "4"="Endocrine, nutritional & metabolic diseases",
               "5"="Mental & behavioral disorders",
               "6"= "Diseases of nervous system & sense organs",
               "7"="Heart diseases",
               "8"="Cerebrovascular disease",
               "9"="Other circulatory disorders",
               "10"="Acute respiratory diseases",
               "11" = "Other respiratory diseases",
               "12" = "Diseases of digestive system",
               "13" = "Diseases of skin & musculoskeletal system",
               "14" = "Genitourinary diseases & complications of pregnancy+childbirth",
               "15" = "Perinatal conditions & malformations",
               "16" = "External causes")


## Data wrangling ####

# Latvia
results.c.latvia <- data.frame(results.c.latvia)
colnames(results.c.latvia) <- cause_names
results.c.latvia$Age <- c(0,1,seq(5,85,5))
rownames(results.c.latvia) <- age_names
results.c.latvia <- gather(data = results.c.latvia,key = Cause,value = Contribution,-Age)

# Russia
results.c.russia <- data.frame(results.c.russia)
colnames(results.c.russia) <- cause_names
results.c.russia$Age <- c(0,1,seq(5,85,5))
rownames(results.c.russia) <- age_names
results.c.russia <- gather(data = results.c.russia,key = Cause,value = Contribution,-Age)

# Poland
```

```r
results.c.poland <- data.frame(results.c.poland)
colnames(results.c.poland) <- cause_names
results.c.poland$Age <- c(0,1,seq(5,85,5))
rownames(results.c.poland) <- age_names
results.c.poland <- gather(data = results.c.poland,key = Cause,value = Contribution,-Age)


## Graphs ####

# Latvia
ggplot(data=results.c.latvia,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010 in Latvia' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```
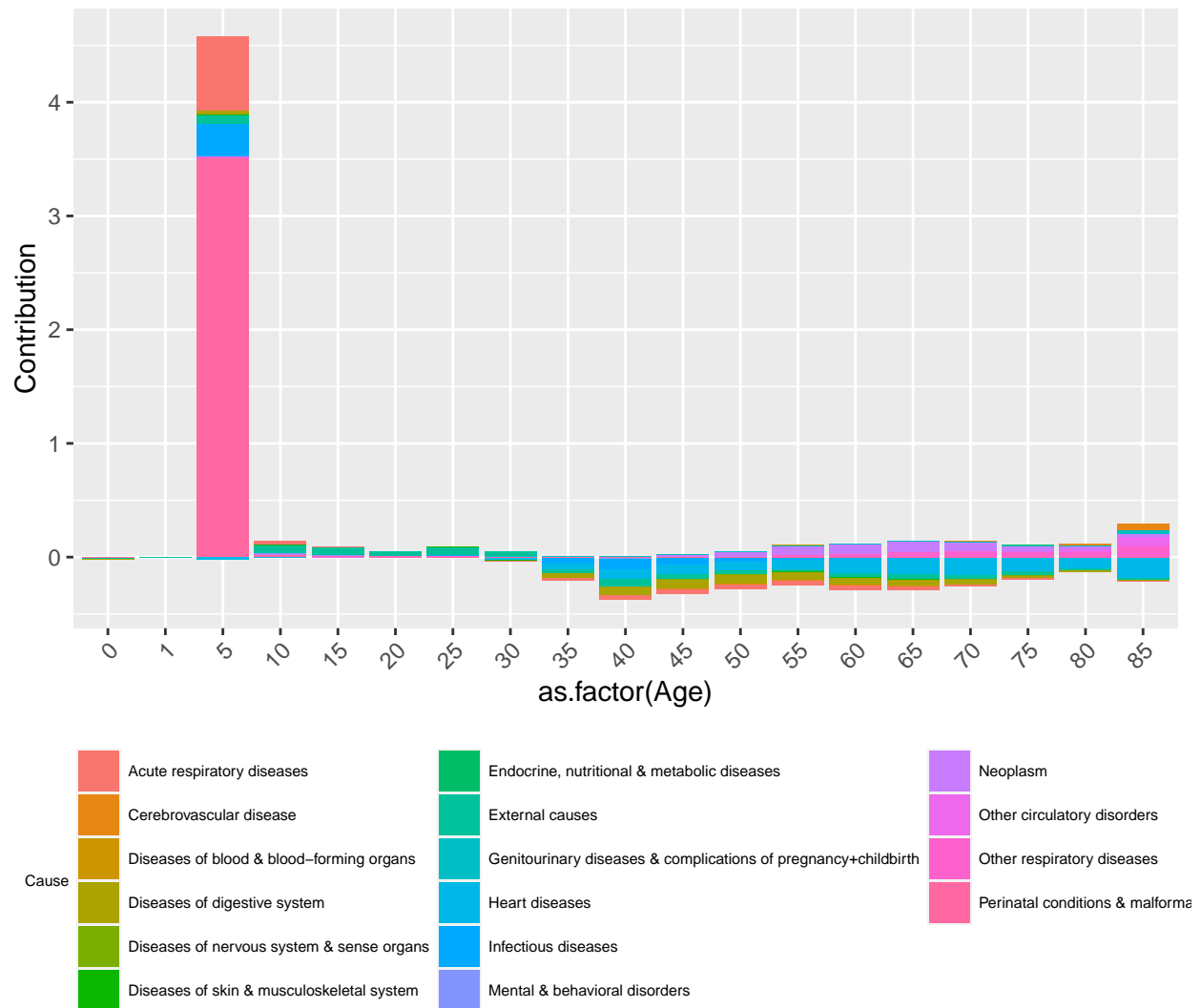
# Change in $e_0$ 1990–2010 in Latvia
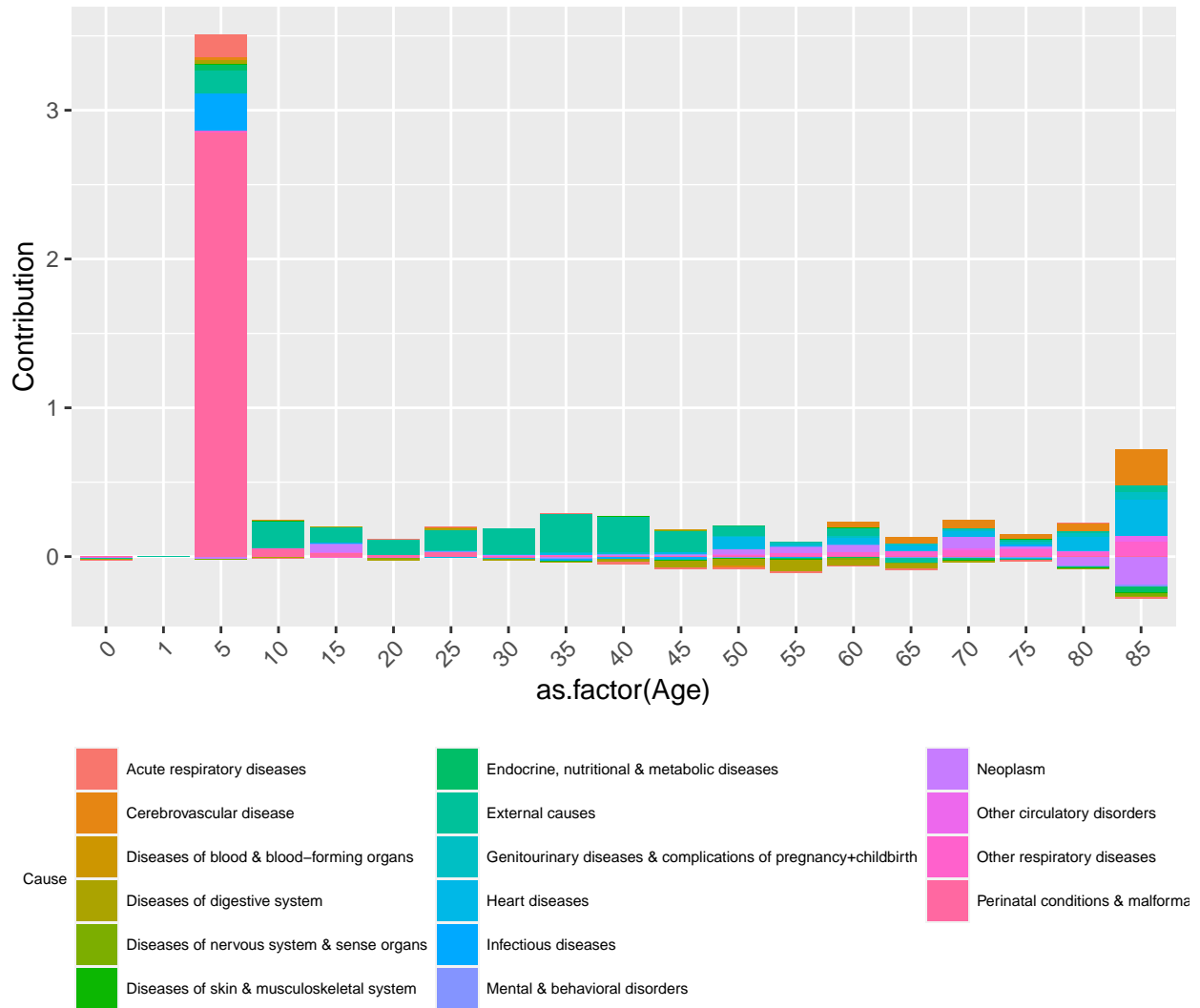


```
# Russia
ggplot(data=results.c.russia,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010 in Russia' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```

# Change in $e_0$ 1990–2010 in Russia



```
# Poland
ggplot(data=results.c.poland,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010 in Poland' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```

# Change in $e_0$ 1990–2010 in Poland
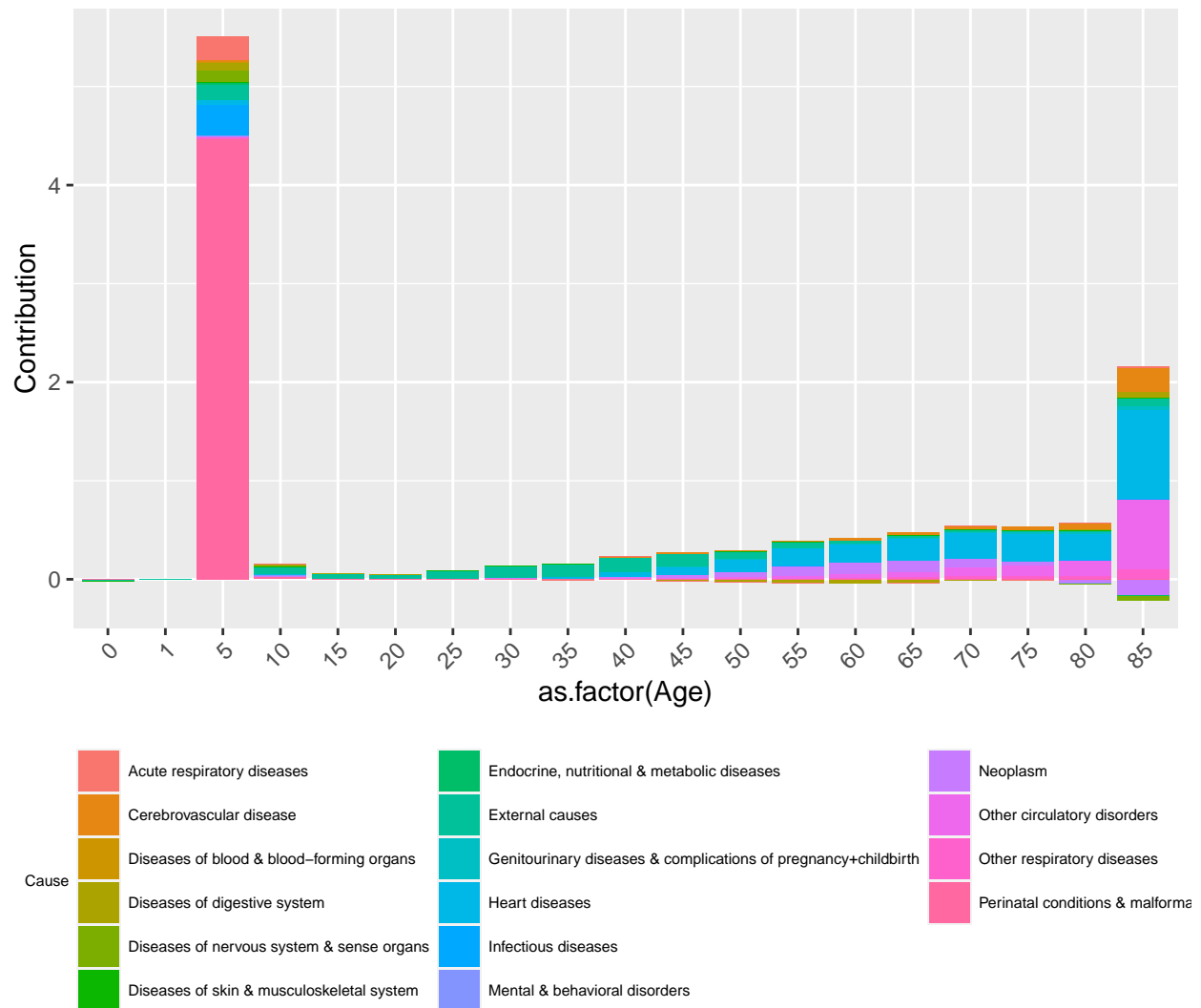


These set of graphs show the change in life expectancy decomposed by age and cause of death. For the all the countries, we can see that as suspected from the first graphs, the increase in the life expectancy at earlier ages is mostly from perinatal conditions and malformations. For the Russian case, the middle age groups had a decrease in life expectancy and the cause of death seemed to show that this is due to a rise in heart diseases. This can be due to change in alcohol and smoking habits as in Soviet Union, Gorbachov had campaigned against alcohol consumption and tobacco usage. For Poland, we observe that below age 55, there has been a decrease in death from external causes and for later ages there is a decrease for heart diseases so these are the main causes of deaths that contributed to the increase life expectancies in these age categories. For Latvia, between age 10 and 55, the main contributors to the change are again in external causes. These can be attributed to the stabilization of political, social and economical conditions for Latvia and Poland.

Now we can look at lifespan variation by calculating the standard deviation.

```
########## DECOMPOSITION: LIFESPAN VARIATION ##########

## Decomposition ####
results.sd.russia <- horiuchi(func = sdfrommxc, pars1 = c(COD1.russia),
                               pars2 = c(COD2.russia), N = 100)
```

```r
results.sd.latvia <- horiuchi(func = sdfrommxc, pars1 = c(COD1.latvia),
                              pars2 = c(COD2.latvia), N = 100)
results.sd.poland<- horiuchi(func = sdfrommxc, pars1 = c(COD1.poland),
                             pars2 = c(COD2.poland), N = 100)


## Data wrangling ####

# Latvia
# Go back to a matrix
dim(results.sd.latvia) <- dim(COD1.latvia)
# original
original.sd.latvia <- sdfrommxc(COD2.latvia) - sdfrommxc(COD1.latvia)
# with the ones obtained with the decomposition
with.decomp.sd.latvia <- sum(results.sd.latvia)


# Russia
# Go back to a matrix
dim(results.sd.russia) <- dim(COD1.russia)
# original
original.sd.russia <- sdfrommxc(COD2.russia) - sdfrommxc(COD1.russia)
# with the ones obtained with the decomposition
with.decomp.sd.russia <- sum(results.sd.russia)

# Poland
# Go back to a matrix
dim(results.sd.poland) <- dim(COD1.poland)
# original
original.sd.poland <- sdfrommxc(COD2.poland) - sdfrommxc(COD1.poland)
# with the ones obtained with the decomposition
with.decomp.sd.poland <- sum(results.sd.poland)


## Check ####

# Comparison
c(original.sd.latvia, with.decomp.sd.latvia,
  original.sd.russia, with.decomp.sd.russia,
  original.sd.poland, with.decomp.sd.poland)
```

```
## [1] -3.647308 -3.647304 -4.108706 -4.108705 -3.818840 -3.818833
```

```r
# Errors
c(with.decomp.sd.latvia - original.sd.latvia,
  with.decomp.sd.russia - original.sd.russia,
  with.decomp.sd.poland- original.sd.poland)
```

```
## [1] 0.000003980185 0.000001251437 0.000007661509
```

```r
## Data wrangling ####

# Latvia
results.sd.latvia <- data.frame(results.sd.latvia)
colnames(results.sd.latvia) <- cause_names
results.sd.latvia$Age <- c(0,1,seq(5,85,5))
rownames(results.sd.latvia) <- age_names
```

```r
results.sd.latvia <- gather(data = results.sd.latvia,key = Cause,value = Contribution,-Age)

# Russia
results.sd.russia <- data.frame(results.sd.russia)
colnames(results.sd.russia) <- cause_names
results.sd.russia$Age <- c(0,1,seq(5,85,5))
rownames(results.sd.russia) <- age_names
results.sd.russia <- gather(data = results.sd.russia,key = Cause,value = Contribution,-Age)

# Poland
results.sd.poland <- data.frame(results.sd.poland)
colnames(results.sd.poland) <- cause_names
results.sd.poland$Age <- c(0,1,seq(5,85,5))
rownames(results.sd.poland) <- age_names
results.sd.poland <- gather(data = results.sd.poland,key = Cause,value = Contribution,-Age)


## Graphs ####

# Latvia
ggplot(data=results.sd.latvia,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ sd[0] ~'1990-2010 in Latvia' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```
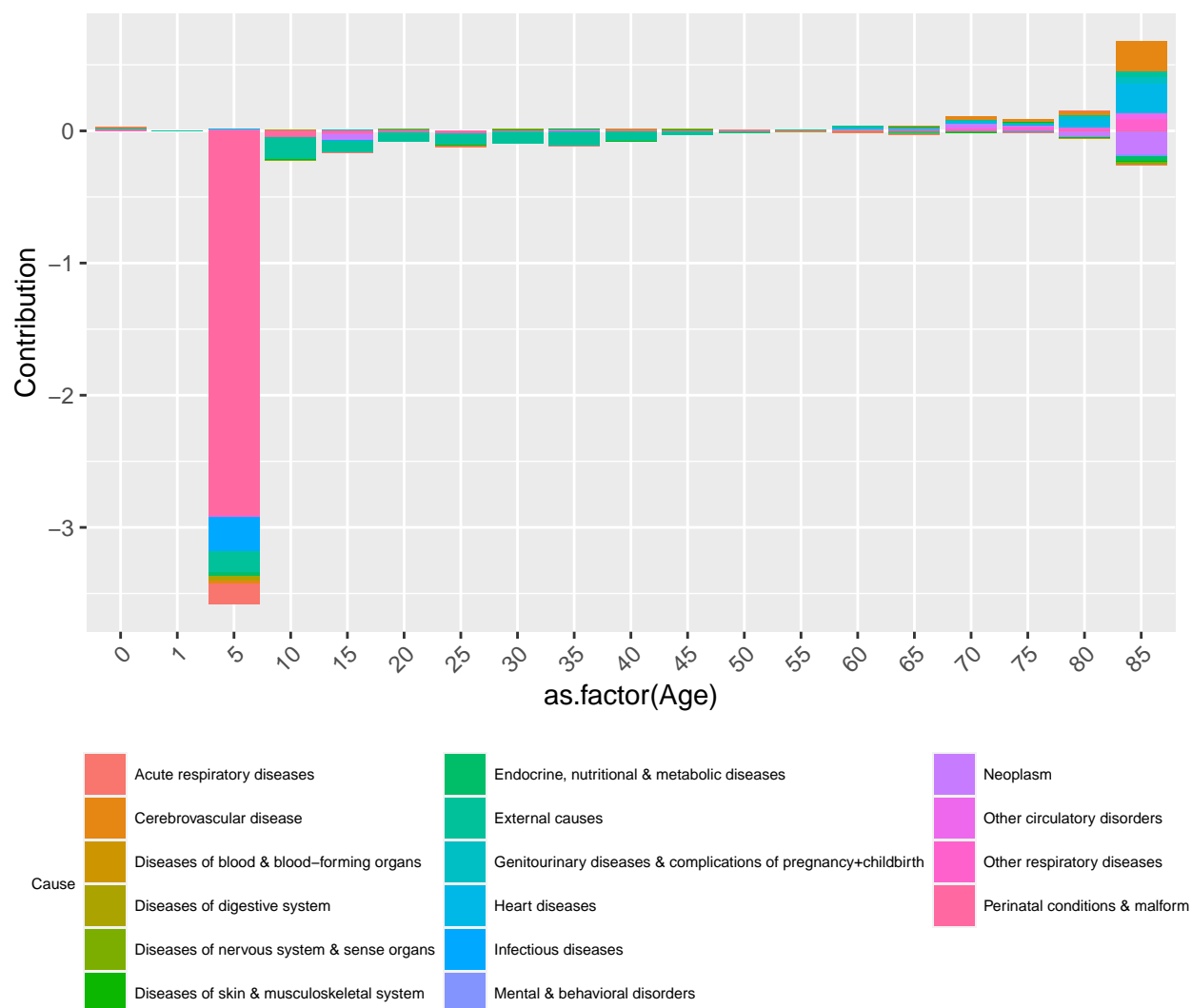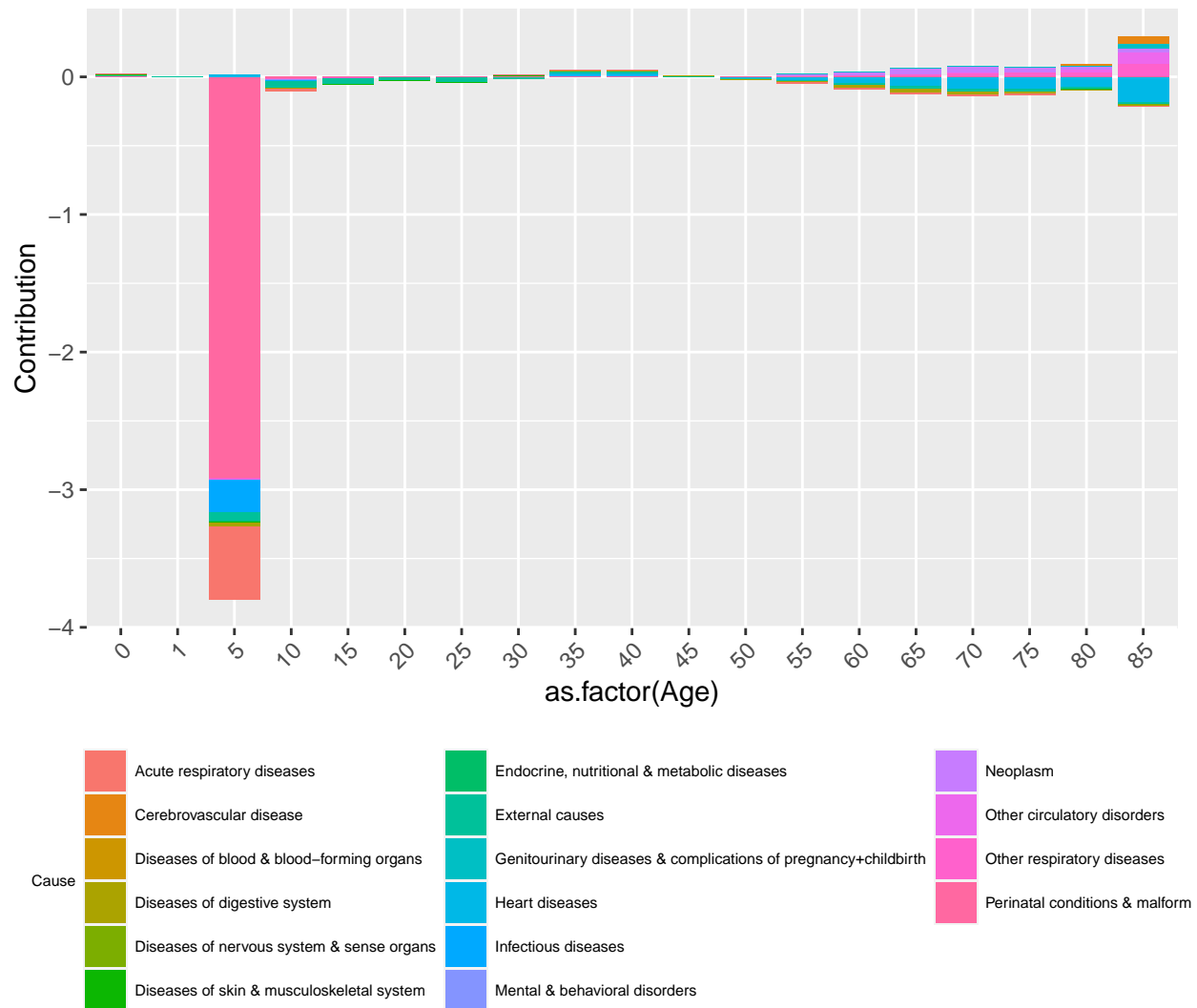
# Change in $sd_0$ 1990–2010 in Latvia



```
# Russia
ggplot(data=results.sd.russia,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ sd[0] ~'1990-2010 in Russia' ))+
  geom_bar(stat = "identity", position = "stack")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```

# Change in $sd_0$ 1990–2010 in Russia



```
# Poland
ggplot(data=results.sd.poland,
       aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ sd[0] ~'1990-2010' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.direction = "horizontal", legend.position = "bottom",
        legend.title = element_text(size = 6),
        legend.text = element_text(size = 6)) +
  guides(fill=guide_legend(ncol=3))
```

## Change in $sd_0$ 1990–2010
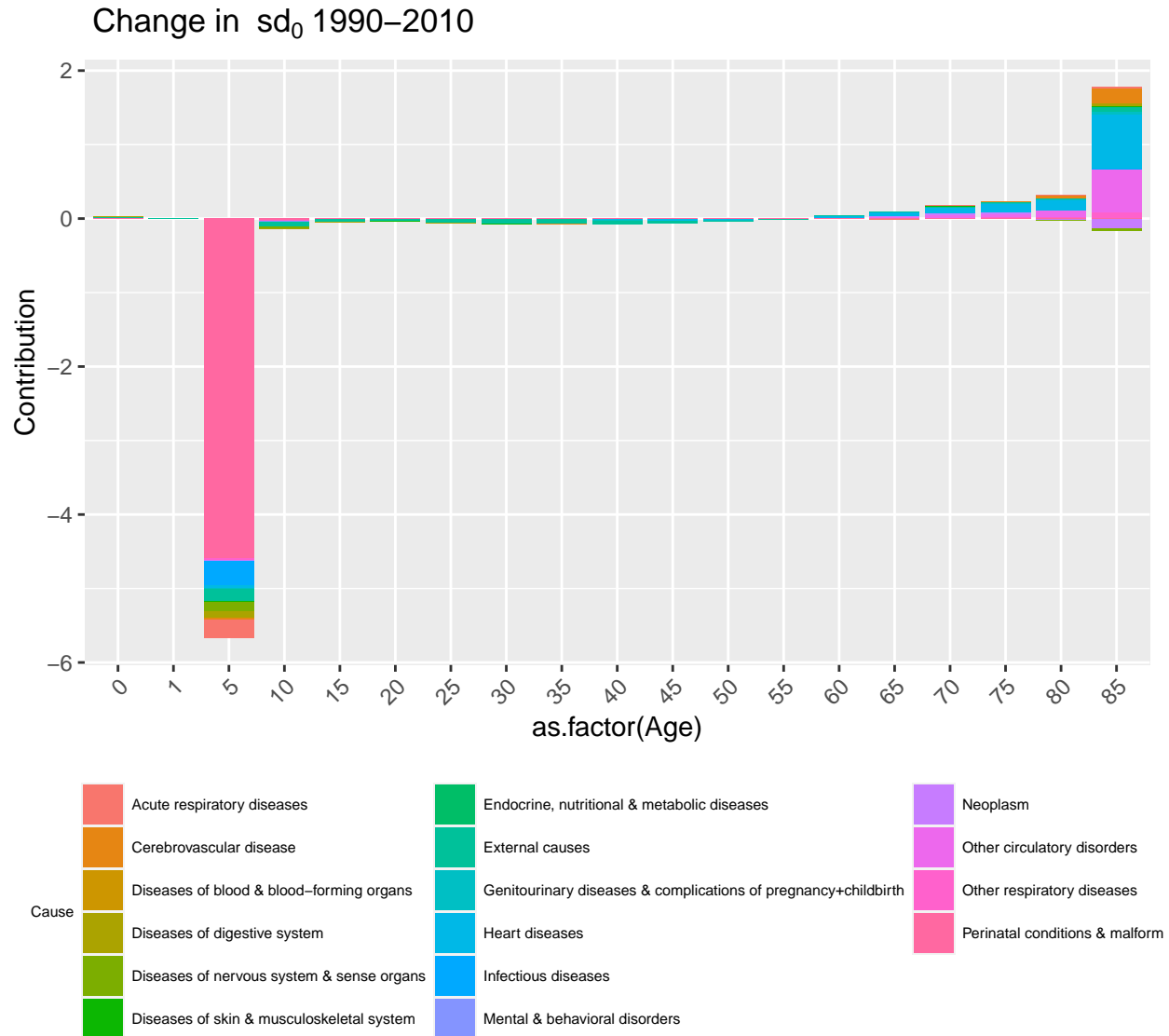


These last plots are for the changes in the standard deviation of the age-at-death distribution decomposed by age and cause of death for Russia, Poland and Latvia. We can observed that for Russian males, the main contributors for the decrease in people ages less than 5 are perinatal conditions and acute respiratory diseases. For the people aged more than 60, this seems to be heart diseases. For Latvia, the main contributor for the ages between 10 and 50 were external causes while for people above 50, the decrease is due to decrease in neoplasm and increase is due to heart diseases and cerebrovascular diseases. For Poland, there does not seem to be one main contributor for the changes in ages between 10 and 70. However, above 70 the main causes of deaths are increase in heart diseases, circulatory disorders and cerebrovascular disease. Overall, we can say that the increase observed in the selected post communist countries are mostly in the 0-5 ages. This can be due to multiple reasons such as development in medicine during these times and increase in availability of early child care. However, it is possible to say that there does not seem to be a homogenous pattern of change in life expectancy by ages and causes of death in these three countries.

## References

Barbieri, Paolo, Rossella Bozzon, Stefani Scherer, Raffaele Grotti, and Michele Lugo. 2015. "The Rise of a Latin Model? Family and Fertility Consequences of Employment Instability in Italy and Spain." *European*

*Societies* 17 (4): 423–46.

Billingsley, Sunnee. 2010. "The Post-Communist Fertility Puzzle." *Population Research and Policy Review* 29 (2): 193–231.

Jasilioniene, Aiva, Dmitri A Jdanov, Tomáš Sobotka, Evgueny M Andreev, Kryštof Zeman, Vladimir M Shkolnikov, Joshua R Goldstein, Dimiter Philipov, and German Rodriguez. 2015. "Methods Protocol for the Human Fertility Database." *Rostock: Max Planck Institute for Demographic Research.*

Kim, Erin Hye-Won. 2017. "Division of Domestic Labour and Lowest-Low Fertility in South Korea." *Demographic Research* 37: 743–68.

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50 (272): 1168–94.

Matysiak, Anna, Tomáš Sobotka, and Daniele Vignoli. 2021. "The Great Recession and Fertility in Europe: A Sub-National Analysis." *European Journal of Population* 37 (1): 29–64.

Preston, Samuel, Patrick Heuveline, and Michael Guillot. 2000. "Demography: Measuring and Modeling Population Processes. 2000." *Malden, MA: Blackwell Publishers.*

Riffe, Tim. 2015. "Reading Human Fertility Database and Human Mortality Database Data into R." *Rostock: Max Planck Institute for Demographic Research (MPIDR Technical Report TR-2015-004).*

Seo, Seung Hyun. 2019. "Low Fertility Trend in the Republic of Korea and the Problems of Its Family and Demographic Policy Implementation." *Population and Economics* 3: 29.

Sobotka, Tomáš, Vegard Skirbekk, and Dimiter Philipov. 2011. "Economic Recession and Fertility in the Developed World." *Population and Development Review* 37 (2): 267–306.

Tønnessen, Marianne. 2019. "Declined Total Fertility Rate Among Immigrants and the Role of Newly Arrived Women in Norway." *European Journal of Population*, 1–27.

Yoo, Sam Hyun, and Tomáš Sobotka. 2018. "Ultra-Low Fertility in South Korea: The Role of the Tempo Effect." *Demographic Research* 38: 549–76.