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

08/6/2021

Challenge 1

Proof Kitagawa decomposition (1995) without interactions

Define the difference between the crude death rates as Δ.

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

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

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

I now combine the eight terms in Δ into four:

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

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, Δ 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.

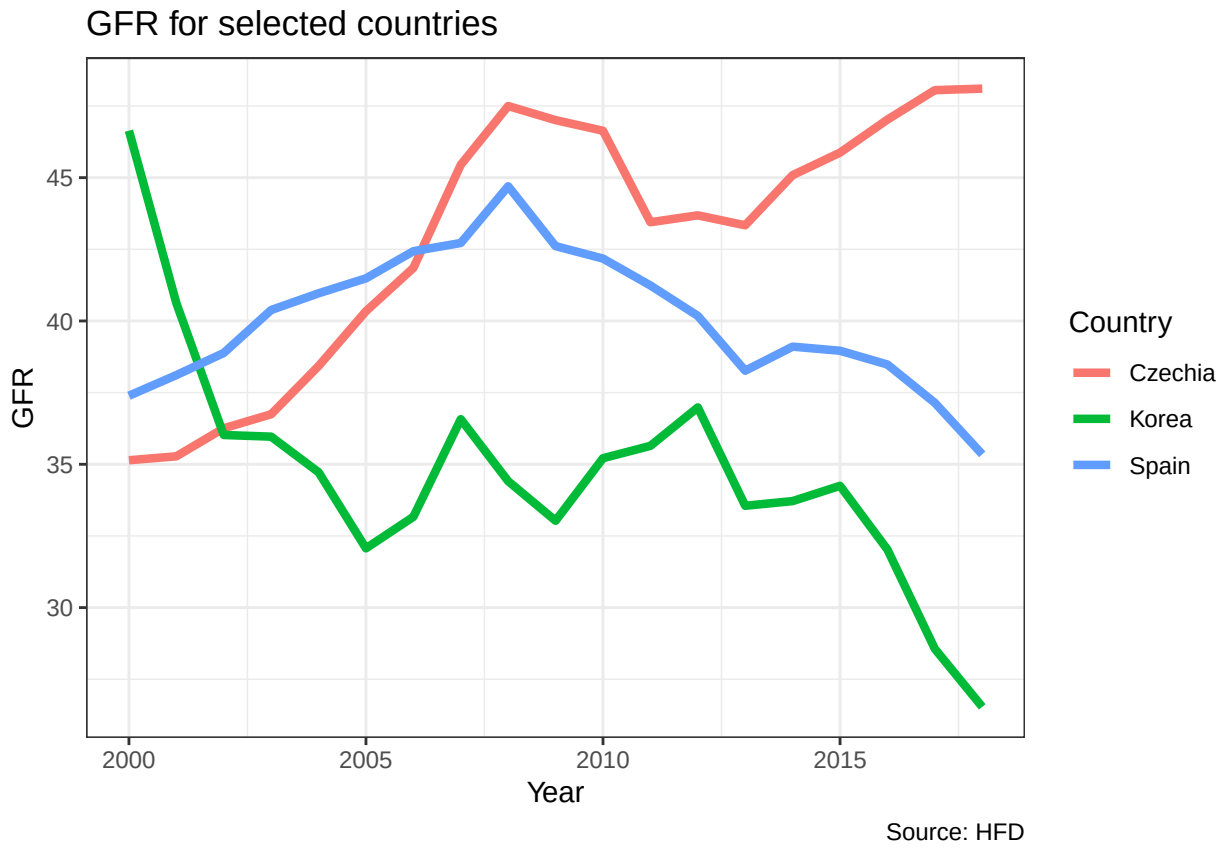
For doing this challenge we decided to take Spain, Korea and Czechia as the three countries to compare. The data was retrieved from the Human Fertility Database (HFD), (Jasilioniene et al. 2015) using Tim Riffe's package `HMDHFDplus` (Riffe 2015). The selected measure is the General Fertility Rate, as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



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).

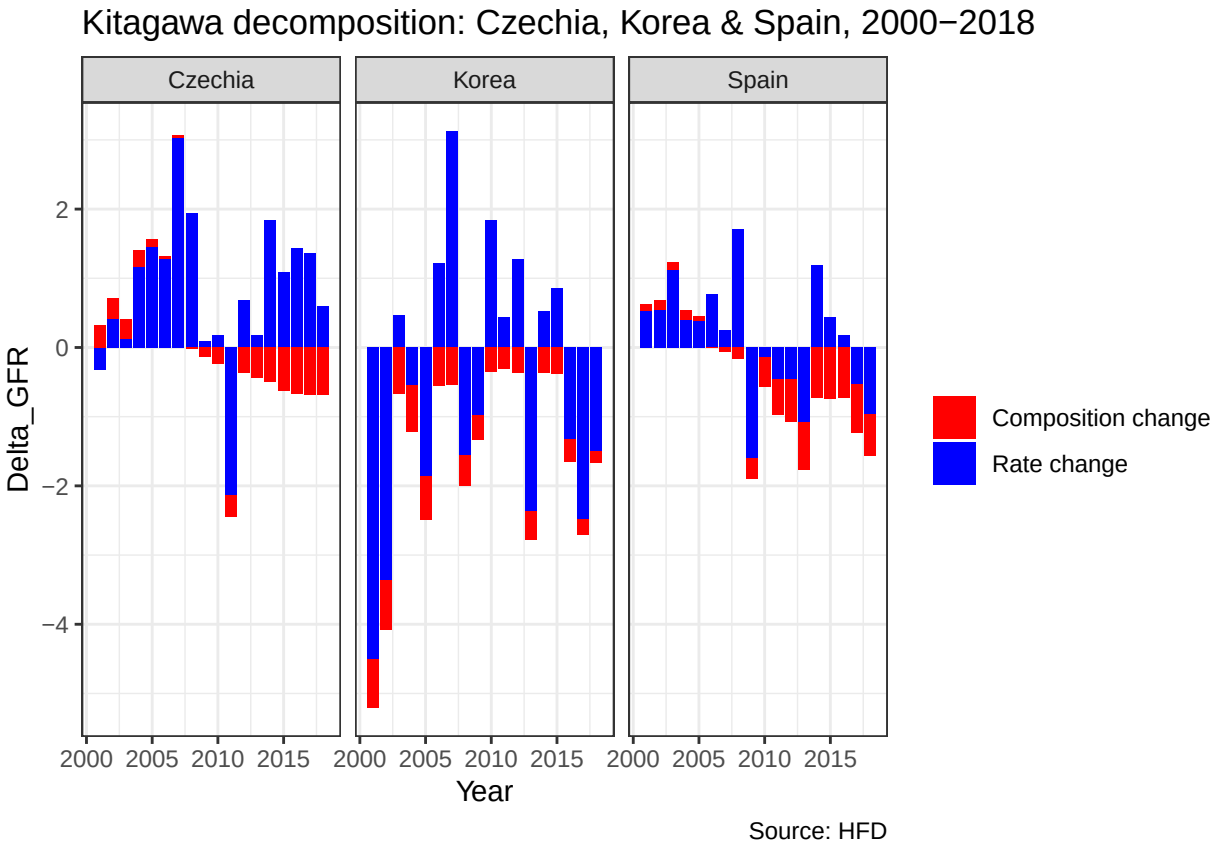
Time to Decompose

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:

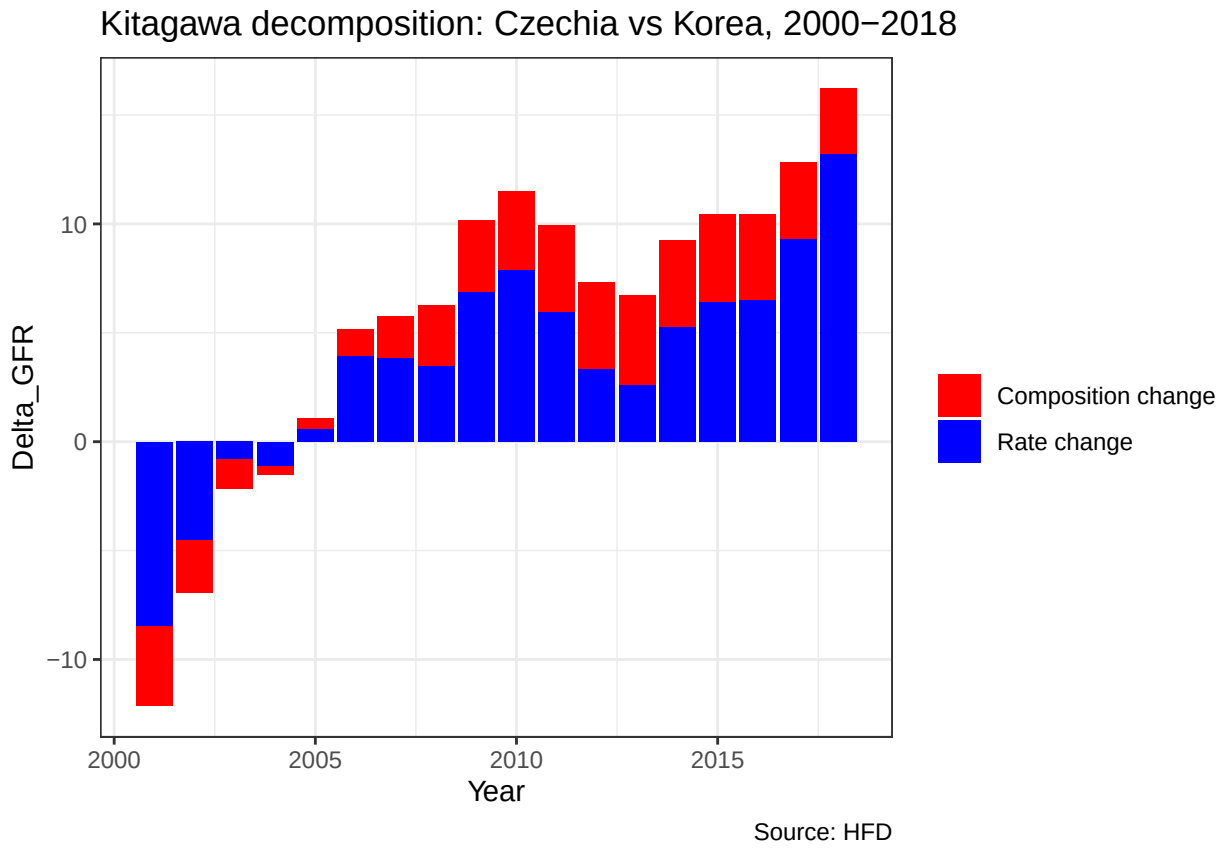


We can observe that the source of change in the GFR for all countries is mostly due to changes in rate and not in population composition. But it is interesting that for both Czechia and Spain, the amount of change due to changes in composition has been increasing during this 2000-2018 period. This is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



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).

In this exercise we analyse data from Lithuania and the Netherlands, the two European countries with the highest (9.6) and the lowest (3.1) gender gap in life expectancy in 2019 according to **Eurostat**. After checking the items availability for these two countries in the **Human Mortality Database**, we get data

from Lithuania (LTU) and the Netherlands (NLD) using the package `HMDHFDplus` (Riffe 2015). We will follow the Practical Guide for Health Expectancy Calculation by the Sullivan Method published by the European Health Expectancy Monitoring Unite (EHEMU) team (Jagger et al. 2006).

First, we extract data on Deaths and Exposures by 5 year age group in 2019 from the HMD for Lithuania and the Netherlands. Then, we compute the Age-Specific mortality rates by 5-year age groups using HMD data for each country and sex separately (cf. the provided code for all the data wrangling procedures).

To apply the Sullivan method, we use data on **self-reported chronic morbidity**, i.e. the presence of long-term (chronic) symptoms, health conditions or diseases. 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 implemented in the European Health Interview Survey (EHIS) and the EU Statistics on Income and Living Conditions (EU-SILC). We will use data from the former.

The EU-SILC target population consists of all 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. According to the **EU-SILC methodology**, data on chronic morbidity 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 quantile (group). As we are only interested in the total population by sex and age-group, we use the first data-set with labour status **(hlth_silc_04)**.

Using the `eurostat package`, we extract the data of people having a long-standing illness or health problem, by sex, age and labour status and get a glimpse on it.

We filter the data to keep only information for Lithuania and the Netherlands in 2019 for each sex separately, as well as of the total population `(wstatus == POP)` and eight age categories corresponding to a 10 years age group distribution, going from 15-24 years old to 85 years and over. This is the most regular age-grouping classification across the whole dataset.

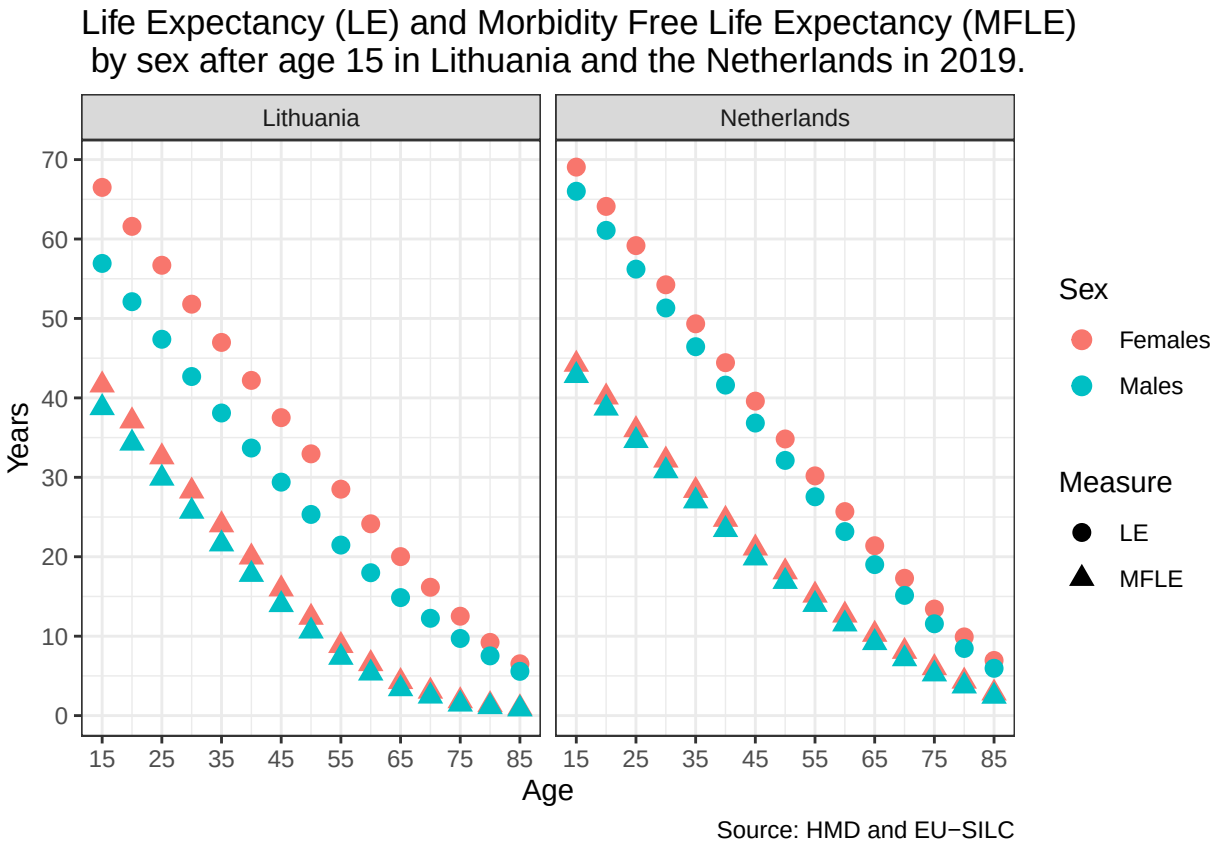
The death rates computed from the HMD are available by 5-year age group, but the chronic morbidity prevalences (pix) obtained from the EU-SILC are only available by 10-year age group. Hence, we need to add a correspondence 10-year age group value to the death rates to join both datasets and continue the calculation. We made the assumption that both 5-year age groups included in each 10-year group (e.g., 25-29 and 30-34 years old) have the same prevalence of chronic morbidity. We split each dataset by sex.

To apply the Sullivan Method, we will consider the non-presence of chronic morbidity as an indicator of good health status. We create the following function allowing to compute both the general Life Expectancy (LE) and the 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)))
```

```r
  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 function to each country and sex specific dataset to compute the LE and the MFLE. Finally we merge all the data in a single data frame, and plot the results.



As it is shown in the previous graph, Life Expectancy (LE) and Morbidity Free Life expectancy (MFLE) after age 15 were clearly higher in the Netherlands than in Lithuania in 2019, when comparing each combination of sex and type of measure. Regarding the (total) LE after age 15, the gender-gap was greater in Lithuania than in the Netherlands, as it could be expected given the above-mentioned differences in the gender gap in LE at birth (of more than 6 years between both countries). The MFLE was also overall higher in the Netherlands than in Lithuania for both sexes. Nevertheless, when considering the prevalence of chronic morbidities, the gap between sexes seemed to diminish, specially in Lithuania, and almost disappeared at old-ages in both countries. Therefore, when examining not only the average length of life, but the number of years that an individual would expect to live without suffering from chronic diseases the gender gap is nuanced in the countries with the highest and lowest differences between sexes.

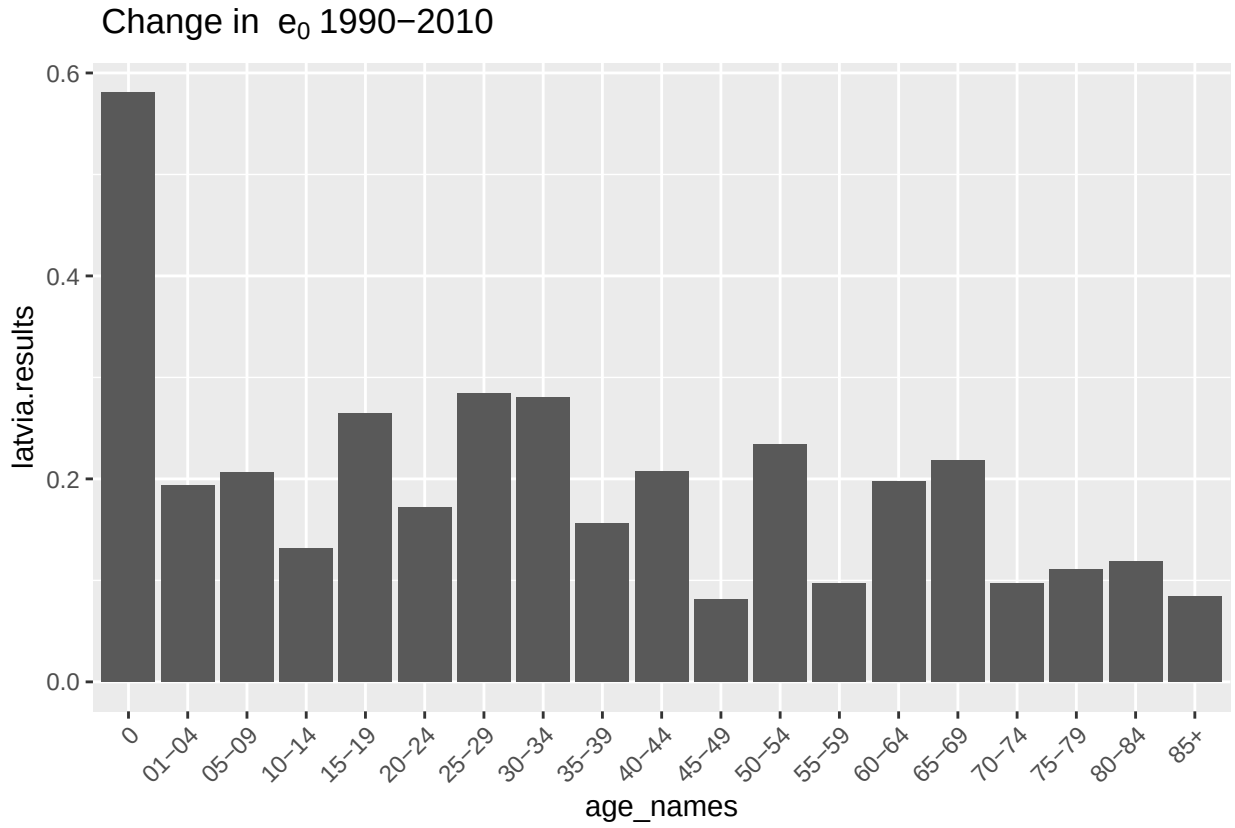
Challenge 4

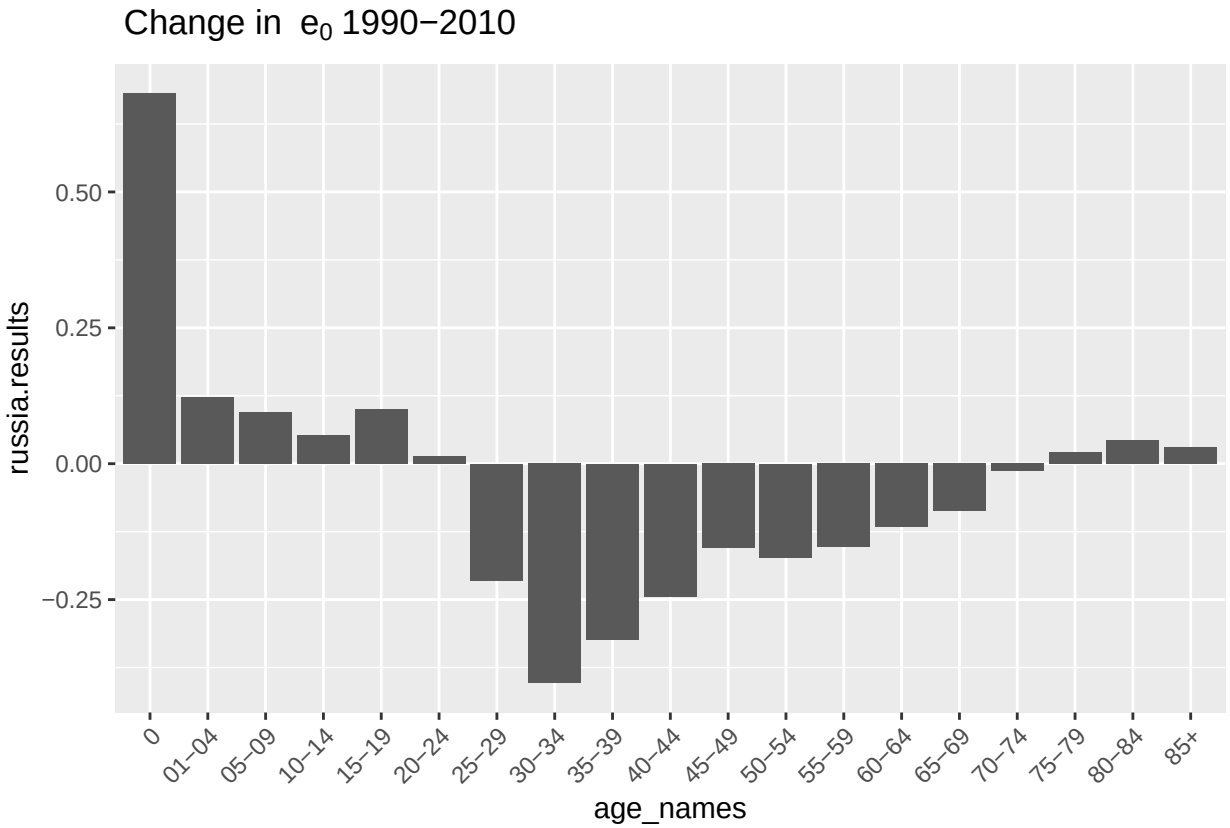
Use the linear integral model to decompose the change in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death for 3 countries you might be interested in (over time or between them). Interpret the results of life expectancy alongside standard deviation.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

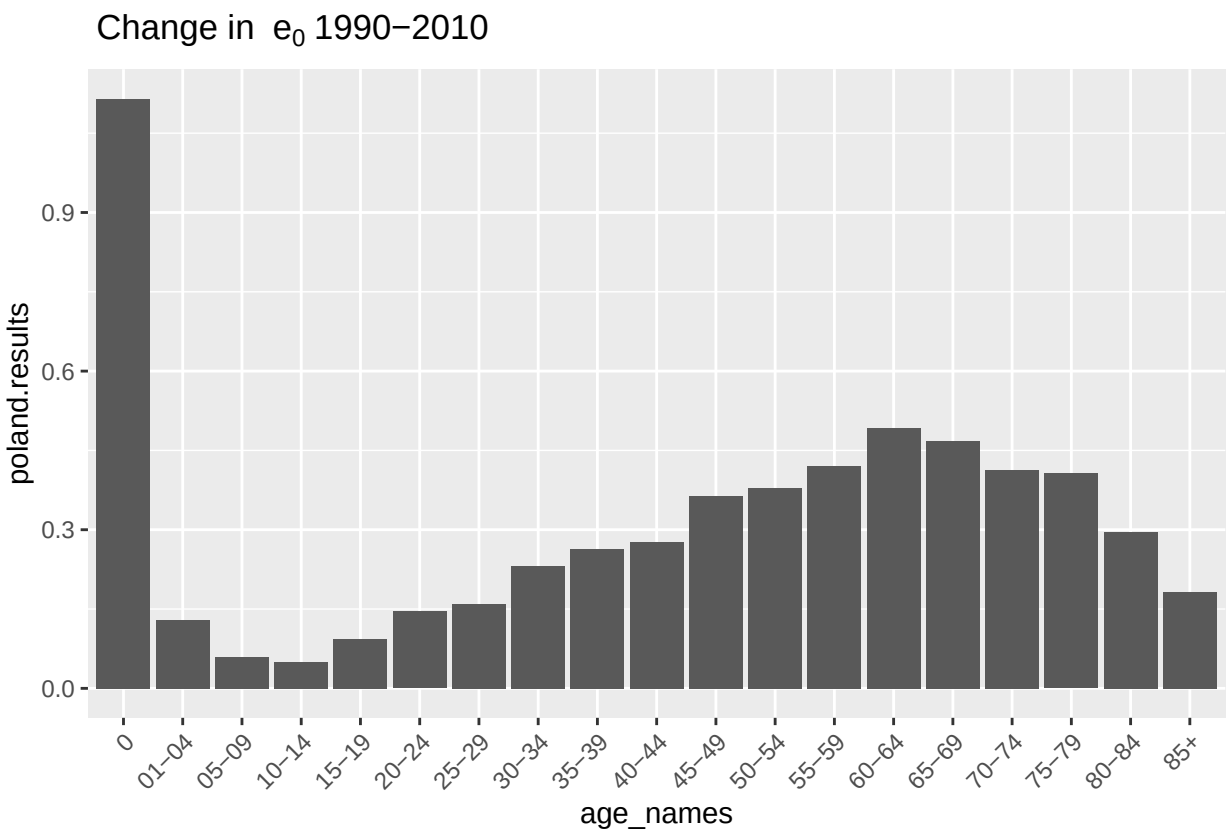
In this exercise, we aim to see the changes in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death in Latvia, Russia and Poland from the Soviet Union dissolution to becoming independent states. Since the impact of macrolevel societal changes can take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



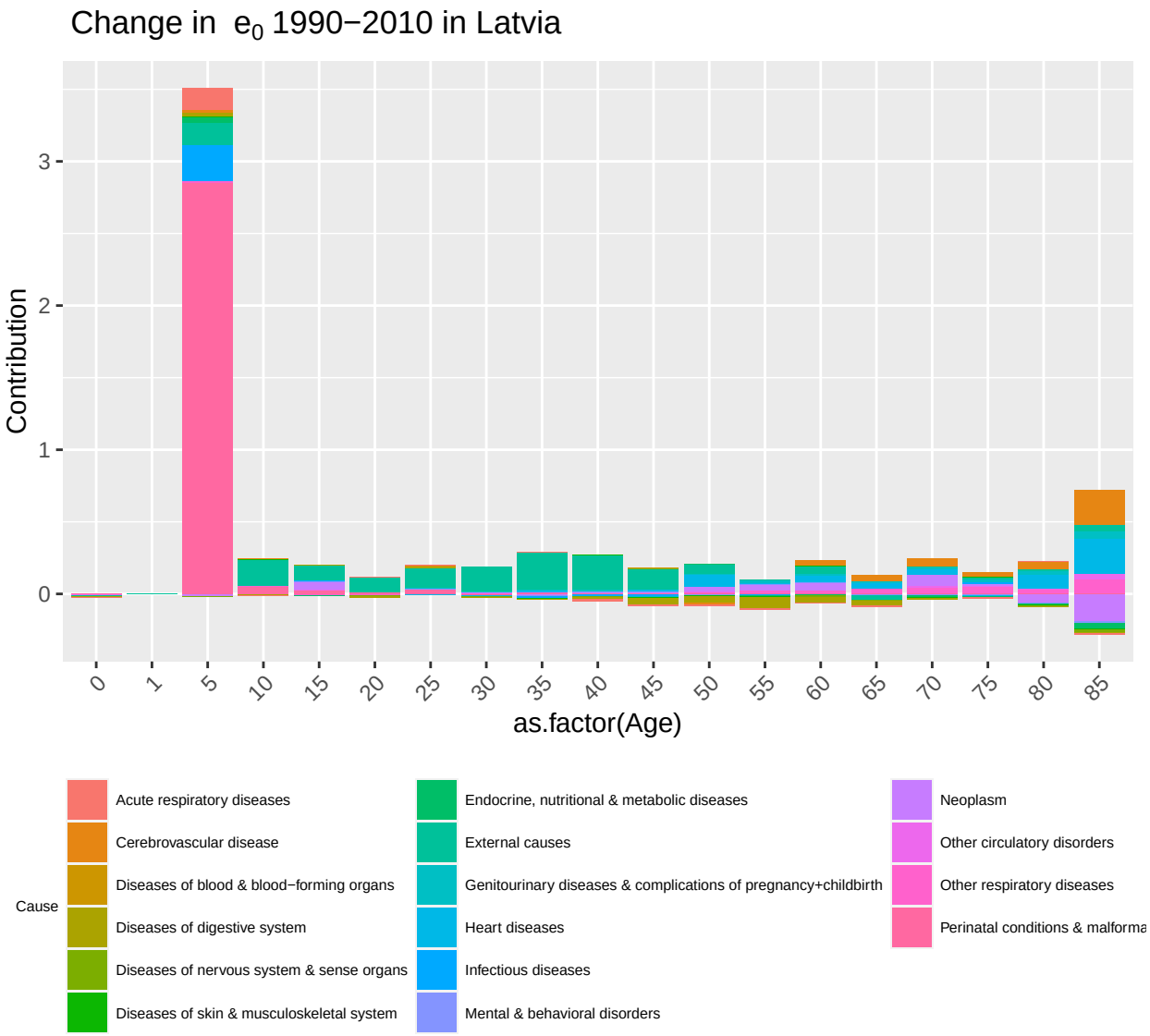

Change in e_0 1990−2010
russia.results
0.50
0.25
0.00
−0.25
0
01–04
05–09
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85+
age_names

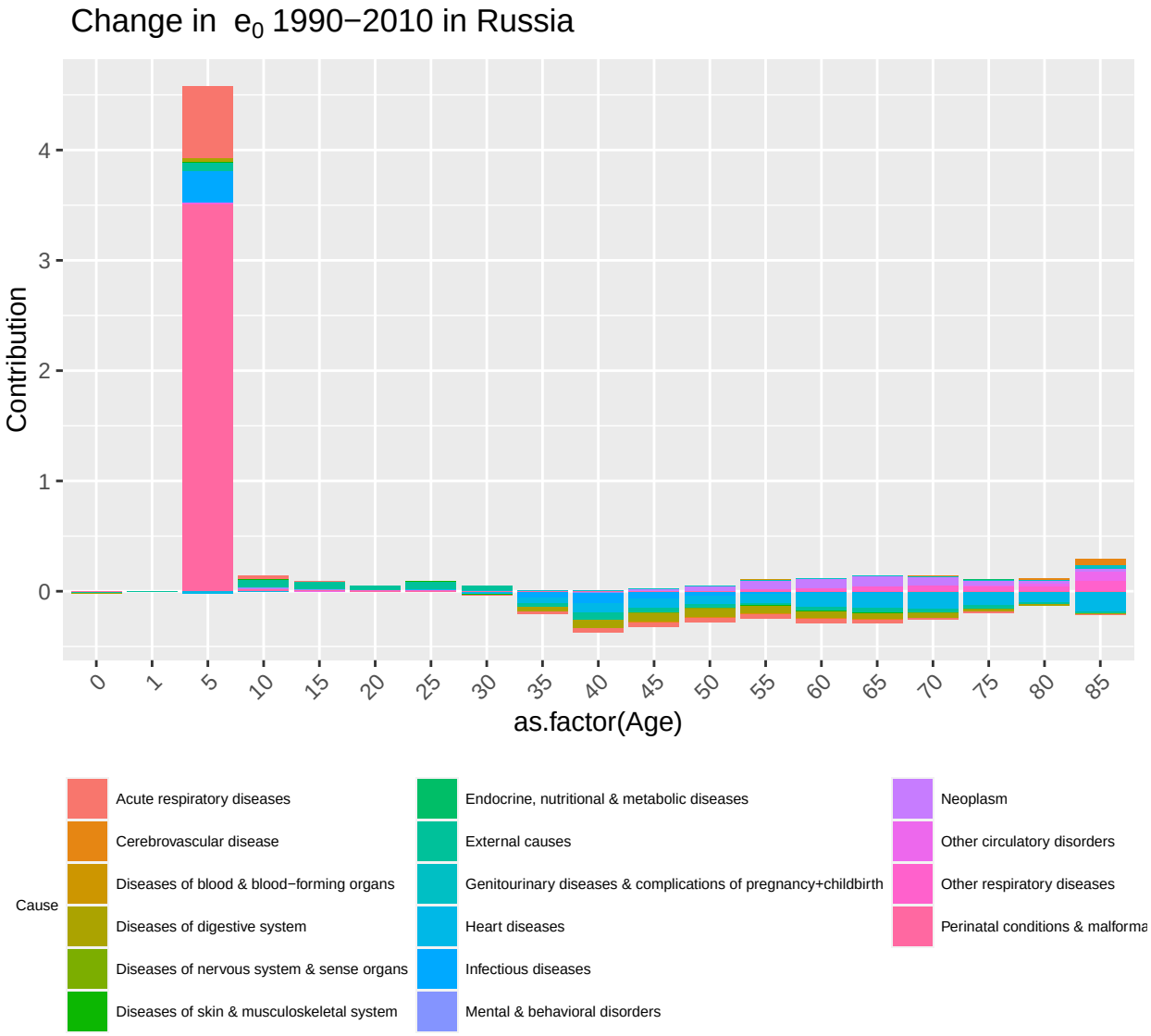

Change in e_0 1990−2010
poland.results
0.9
0.6
0.3
0.0
0
01–04
05–09
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85+
age_names

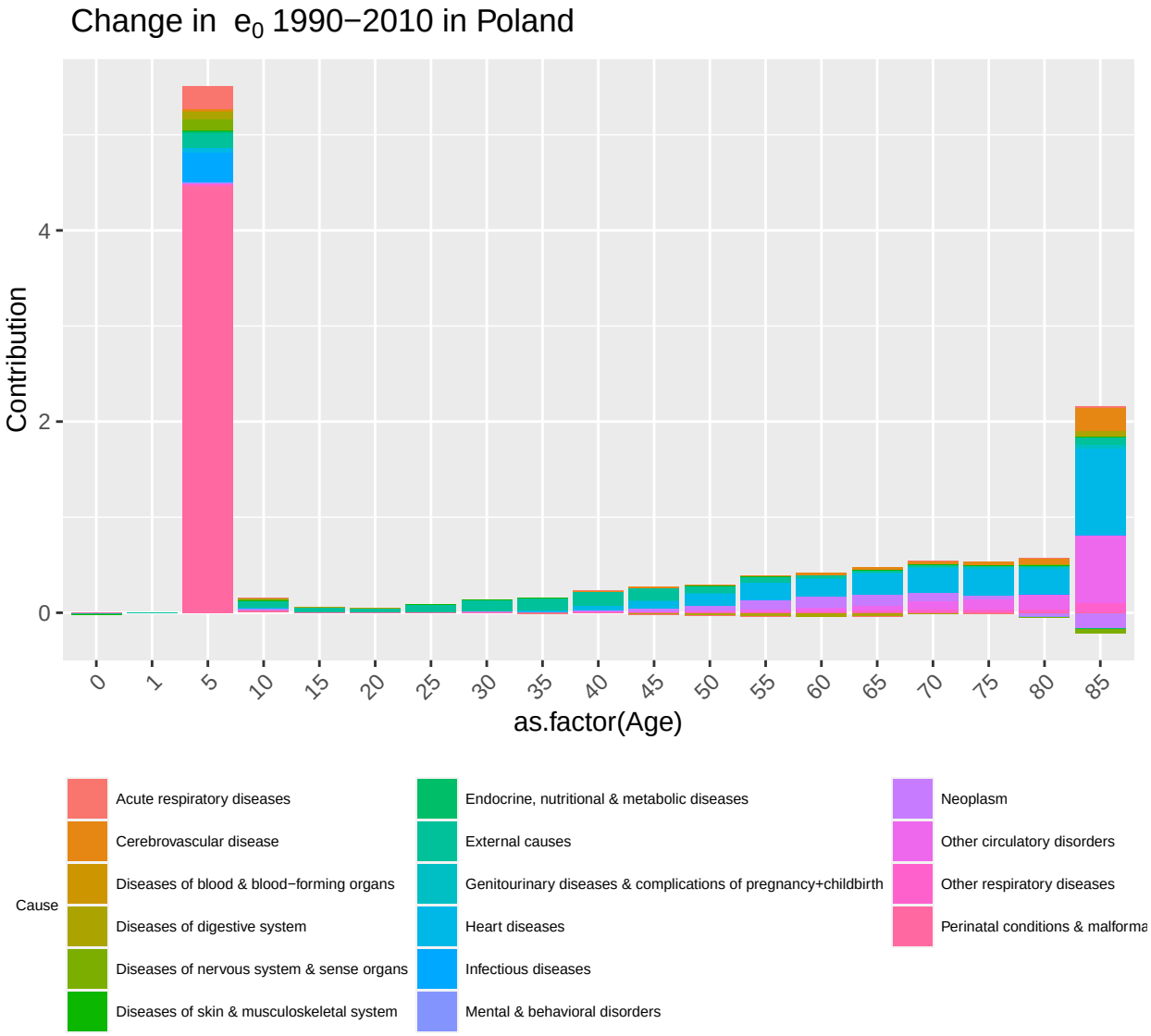
Unlike the graph for Russia, Latvia and Poland shows that there has been overall a positive change in the life expactancy.In all of the countries, the age group that has experienced the highest gain in life expectancy is the 0-5 ages. This is expected considering the increase in availability of communicable 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.

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

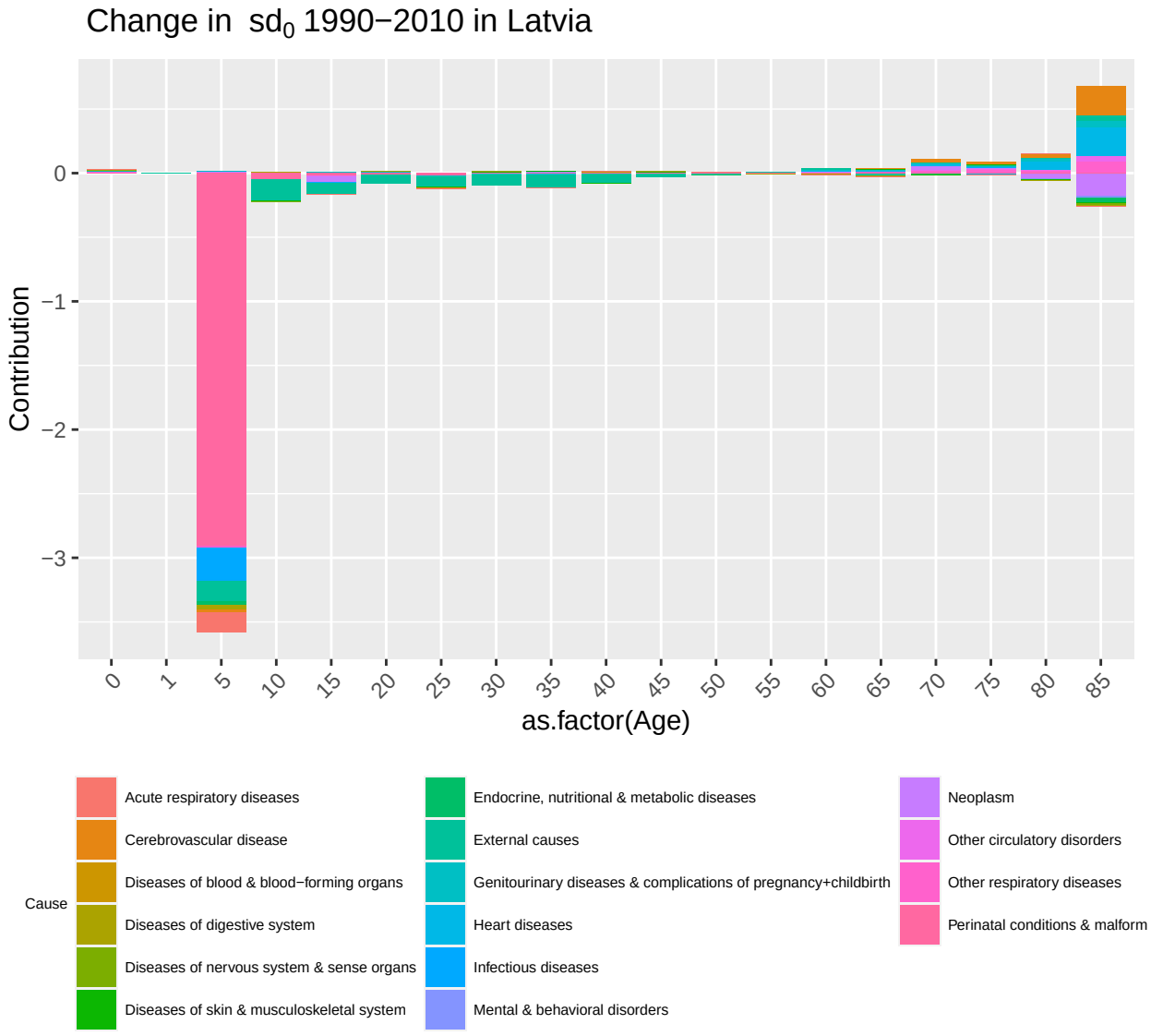


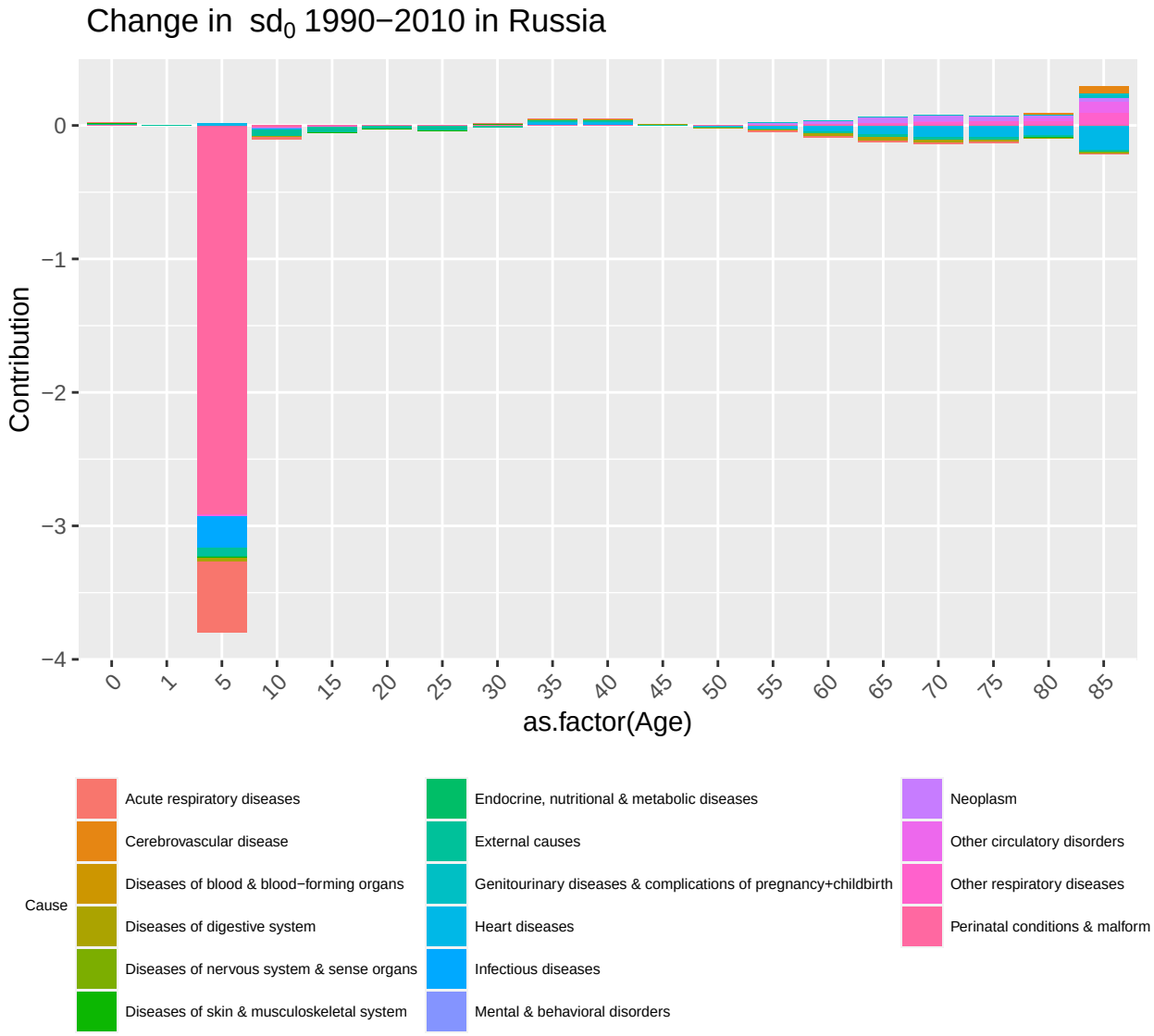

Change in e_0 1990−2010 in Russia
Contribution
4
3
2
1
0
0
1
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
as.factor(Age)
Cause
Acute respiratory diseases
Cerebrovascular disease
Diseases of blood & blood−forming organs
Diseases of digestive system
Diseases of nervous system & sense organs
Diseases of skin & musculoskeletal system
Endocrine, nutritional & metabolic diseases
External causes
Genitourinary diseases & complications of pregnancy+childbirth
Heart diseases
Infectious diseases
Mental & behavioral disorders
Neoplasm
Other circulatory disorders
Other respiratory diseases
Perinatal conditions & malforma

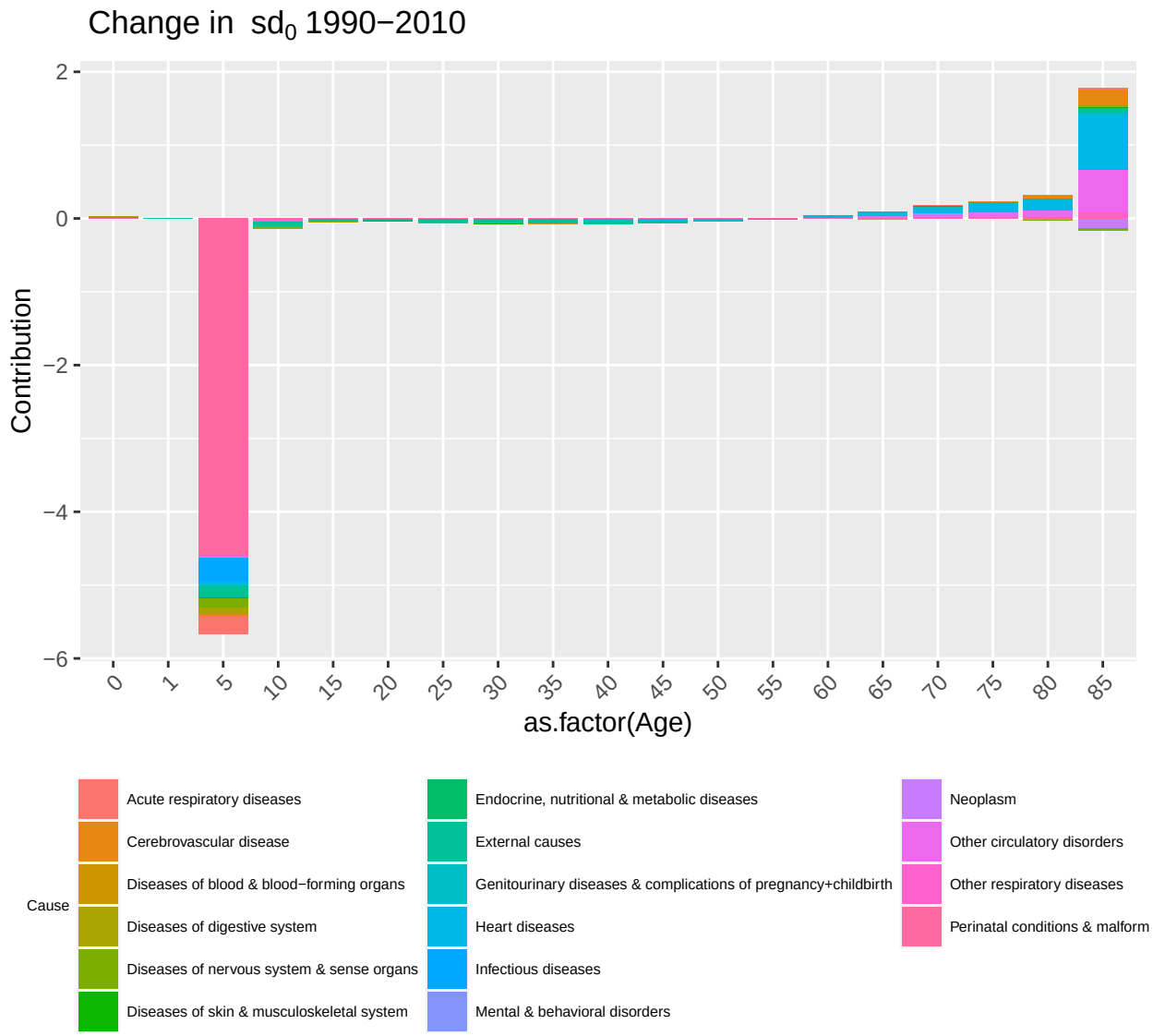


These set of graphs show the change in life expectancy decomposed by age and cause of death. For the all the countries, we can see that as suspected from the first graphs, the increase in the life expectancy at earlier ages is mostly from perinatal conditions and malformations. For the Russian case, 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.


Change in sd0 1990−2010 in Latvia
Contribution
0
−1
−2
−3
0
1
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
as.factor(Age)
Cause
Acute respiratory diseases
Cerebrovascular disease
Diseases of blood & blood−forming organs
Diseases of digestive system
Diseases of nervous system & sense organs
Diseases of skin & musculoskeletal system
Endocrine, nutritional & metabolic diseases
External causes
Genitourinary diseases & complications of pregnancy+childbirth
Heart diseases
Infectious diseases
Mental & behavioral disorders
Neoplasm
Other circulatory disorders
Other respiratory diseases
Perinatal conditions & malform


Change in sd0 1990−2010 in Russia
Contribution
0
−1
−2
−3
−4
0
1
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
as.factor(Age)
Cause
Acute respiratory diseases
Cerebrovascular disease
Diseases of blood & blood−forming organs
Diseases of digestive system
Diseases of nervous system & sense organs
Diseases of skin & musculoskeletal system
Endocrine, nutritional & metabolic diseases
External causes
Genitourinary diseases & complications of pregnancy+childbirth
Heart diseases
Infectious diseases
Mental & behavioral disorders
Neoplasm
Other circulatory disorders
Other respiratory diseases
Perinatal conditions & malform



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.

Jagger, Carol, Biana Cox, Sophie Le Roy, EHEMU Group, and others. 2006. "Health Expectancy Calculation by the Sullivan Method." *EHEMU Technical Report*.

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.