

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

01/6/2021

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

# HFD credentials
username=""
password=""

# 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 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.

```
# Check countries
getHFDcountries()

## [1] "AUT"      "BLR"      "BGR"      "CAN"      "CHL"      "HRV"      "CZE"
## [8] "DNK"      "EST"      "FIN"      "FRATNP"   "DEUTNP"   "DEUTW"    "DEUTE"
## [15] "HUN"      "ISL"      "ITA"      "JPN"      "LTU"      "NLD"      "NOR"
## [22] "POL"      "PRT"      "KOR"      "RUS"      "SVK"      "SVN"      "ESP"
## [29] "SWE"      "CHE"      "TWN"      "UKR"      "GBR_NP"   "GBRTENW"  "GBR_SCO"
## [36] "GBR_NIR" "USA"

### Downloading data for 3 countries: Spain, Bulgaria, Korea ###

# Spain
spain_birth <- readHFDweb(CNTRY = "ESP", item = "birthsTR",
                        username = username, password = password)
spain_exposure <- readHFDweb(CNTRY = "ESP", item = "exposTR",
                           username = username, password = password)

# Czechia
cze_birth <- readHFDweb(CNTRY = "CZE", item = "birthsTR",
                      username = username, password = password)
cze_exposure <- readHFDweb(CNTRY = "CZE", item = "exposTR",
                         username = username, password = password)

# Korea
kor_birth <- readHFDweb(CNTRY = "KOR", item = "birthsTR",
```

```

        username = username, password = password)
kor_exposure <- readHFDweb(CNTRY = "KOR", item = "exposTR",
        username = username, password = password)

## Wrangling time ####

# Spain
spain_birth2 <- spain_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()
spain_exposure2 <- spain_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

# Czechia
cze_birth2 <- cze_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()
cze_exposure2 <- cze_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

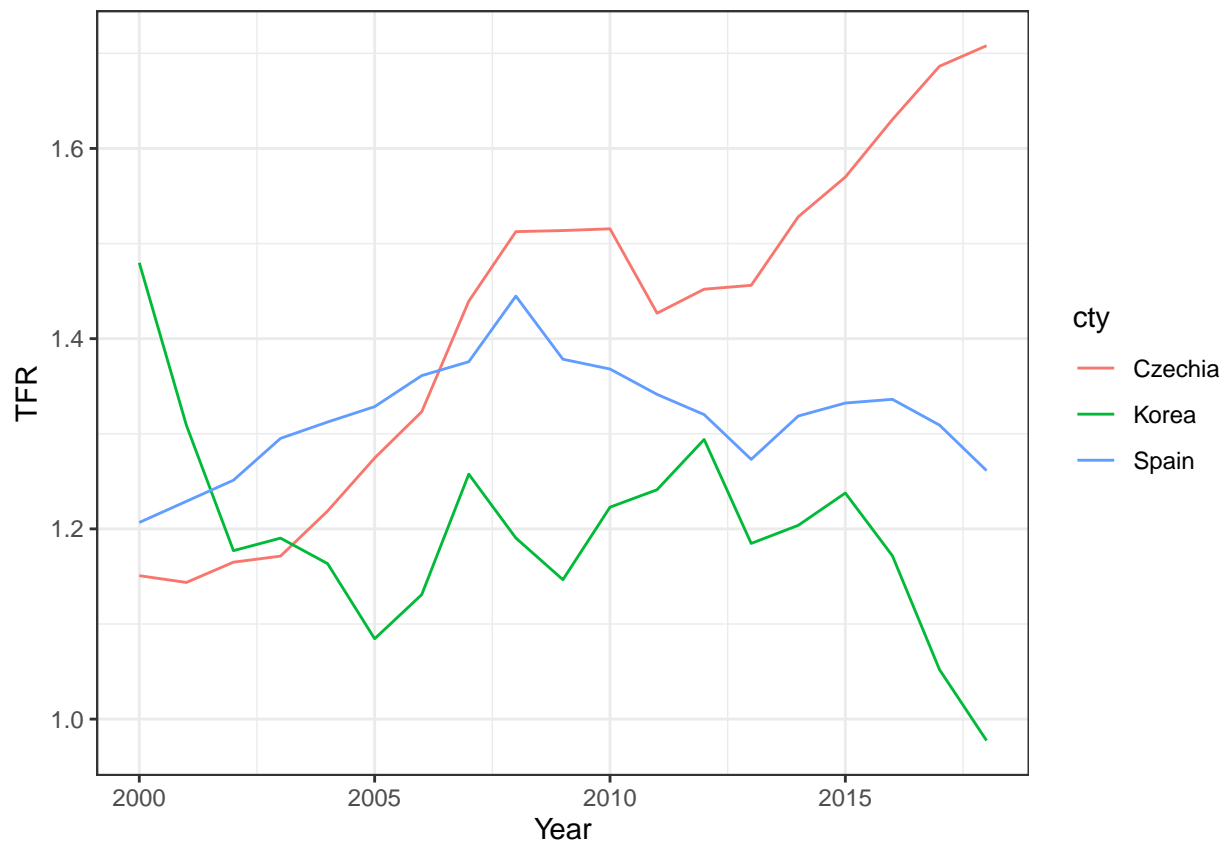
# Korea
kor_birth2 <- kor_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()
kor_exposure2 <- kor_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

# All together now
spain <- spain_exposure2 %>% left_join(spain_birth2, by = c("Year", "Age"))
czechia <- cze_exposure2 %>% left_join(cze_birth2, by = c("Year", "Age"))
korea <- kor_exposure2 %>% left_join(kor_birth2, by = c("Year", "Age"))

# All together in 1 data.frame
spain$cty <- "Spain"
czechia$cty <- "Czechia"
korea$cty <- "Korea"
challenge2 <- rbind(spain, czechia, korea)

```

```
# Some plotting
challenge2 %>%
  filter(Age >= 15, Age <= 49) %>%
  mutate(TFR_age = Births/Exposure) %>%
  group_by(cty, Year) %>%
  summarise(TFR = sum(TFR_age, na.rm = T)) %>%
  filter(Year >= 2000) %>%
  ggplot(aes(x=Year, y = TFR, color = cty)) + geom_line() +
  theme_bw()
```



```
### Function for the Kitagawa decomposition ###
```

```
# Input: exposures and deaths for A and B
# Output: ASCDRs, CDRs, CDR(A)-CDR(B), compositional diff, ASR diff
```

```
stand.decom <- function(P.A, B.A, P.B, B.B){

  # CRUDE RATES
  A.crude <- sum(B.A) / sum(P.A)
  B.crude <- sum(B.B) / sum(P.B)
  Diff.crude <- A.crude - B.crude

  # age distribution of country A and B
  C.A <- P.A / sum(P.A)
  C.B <- P.B / sum(P.B)

  # age-specific fertility rate in country A and B
  F.A <- B.A / P.A
```

```

F.B <- B.B / P.B
# average age distribution
C.ave <- (C.A + C.B)/2

#### DECOMPOSITION OF DIFFERENCES BETWEEN RATES
comp.diff <- sum((C.A - C.B) * ((F.A + F.B)/2))
ASR.diff <- sum((F.A - F.B) * C.ave)

# preparing the outcomes
outcome <- c( Diff.crude = Diff.crude,
              Diff.comp = comp.diff,
              Diff.rates = ASR.diff)

# giving the outcomes
return(outcome)
}

### Loop to apply the decomposition for consecutive years ###

decomp <- matrix(NA,11,3)

for(i in c(2007:2017)){

  # births and exposures
  spain_birth2a <- spain_birth2 %>% filter(Year==i) %>% pull(Births)
  spain_birth2b <- spain_birth2 %>% filter(Year==i+1) %>% pull(Births)

  spain_exposure2a <- spain_exposure2 %>% filter(Year==i) %>% pull(Exposure)
  spain_exposure2b <- spain_exposure2 %>% filter(Year==(i+1)) %>% pull(Exposure)

  # standardization + decomposition
  out <- stand.decom(P.A = spain_exposure2a, B.A = spain_birth2a,
                    P.B = spain_exposure2b, B.B = spain_birth2b)

  for(j in 1:3){ decomp[i-2006,j] <- out[[j]] }

}

decomp <- as.data.frame(decomp)
colnames(decomp) <- c("Diff.crude", "Diff.comp", "Diff.rates")
rownames(decomp) <- c("2007-2008", "2008-2009", "2009-2010", "2010-2011",
                    "2011-2012", "2012-2013", "2013-2014", "2014-2015",
                    "2015-2016", "2016-2017", "2017-2018")

decomp

##           Diff.crude    Diff.comp    Diff.rates
## 2007-2008 -0.0016370674 0.0001391127 -0.0017761801
## 2008-2009 0.0019547642 0.0002993576 0.0016554066
## 2009-2010 0.0005636163 0.0004150469 0.0001485694
## 2010-2011 0.0009715521 0.0004915310 0.0004800211
## 2011-2012 0.0010494831 0.0005735937 0.0004758894
## 2012-2013 0.0017861523 0.0006621398 0.0011240126
## 2013-2014 -0.0005179004 0.0007274552 -0.0012453557
## 2014-2015 0.0002855512 0.0007375776 -0.0004520265

```

```
## 2015-2016 0.0005510885 0.0007440104 -0.0001929219
## 2016-2017 0.0012558197 0.0007013645 0.0005544551
## 2017-2018 0.0016002491 0.0005962147 0.0010040344
```

Challenge 3

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=username,password=password,fixup=T)
latvia.expo <- readHMDweb(CNTRY="LVA",item="Exposures_5x1",
                          username=username,password=password,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,) %>%
  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=username,password=password, fixup=T)
russia.expo <- readHMDweb(CNTRY="RUS",item="Exposures_5x1",
                          username=username,password=password,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,]
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

##### DATA PREPARATION: POLAND #####

## Data All Cause ###

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

#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
        }
        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), ...)
    }
  }
  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 * nm[x][1]
          nax[2] <- 1.651 - 2.816 * nm[x][1]
        }
      }
    }
  }
}

```

```

    if (sex == 2) {
      if (nm[x][1] >= 0.107) {
        nax[1] <- 0.35
        nax[2] <- 1.361
      }
      else {
        nax[1] <- 0.053 + 2.8 * nm[x][1]
        nax[2] <- 1.522 - 1.518 * nm[x][1]
      }
    }
  }
}
nqx <- (n * nm[x]) / (1 + (n - nax) * nm[x])
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]/nm[x][length(nm[x])])
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)
}

# sdfrommx ####

sdfrommx <- 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] -5.745130e-07 2.265514e-08 -4.550067e-06

```

The graphs

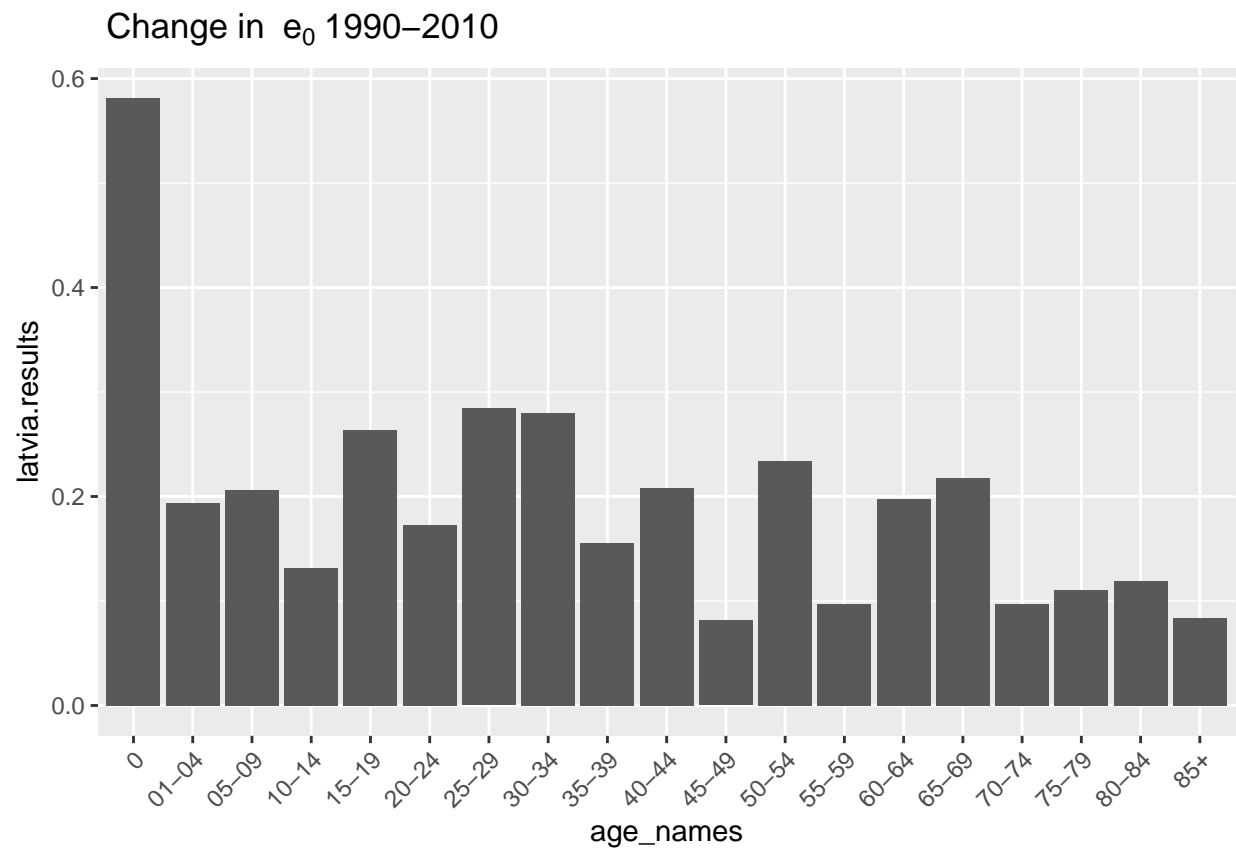
```

## 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 ####

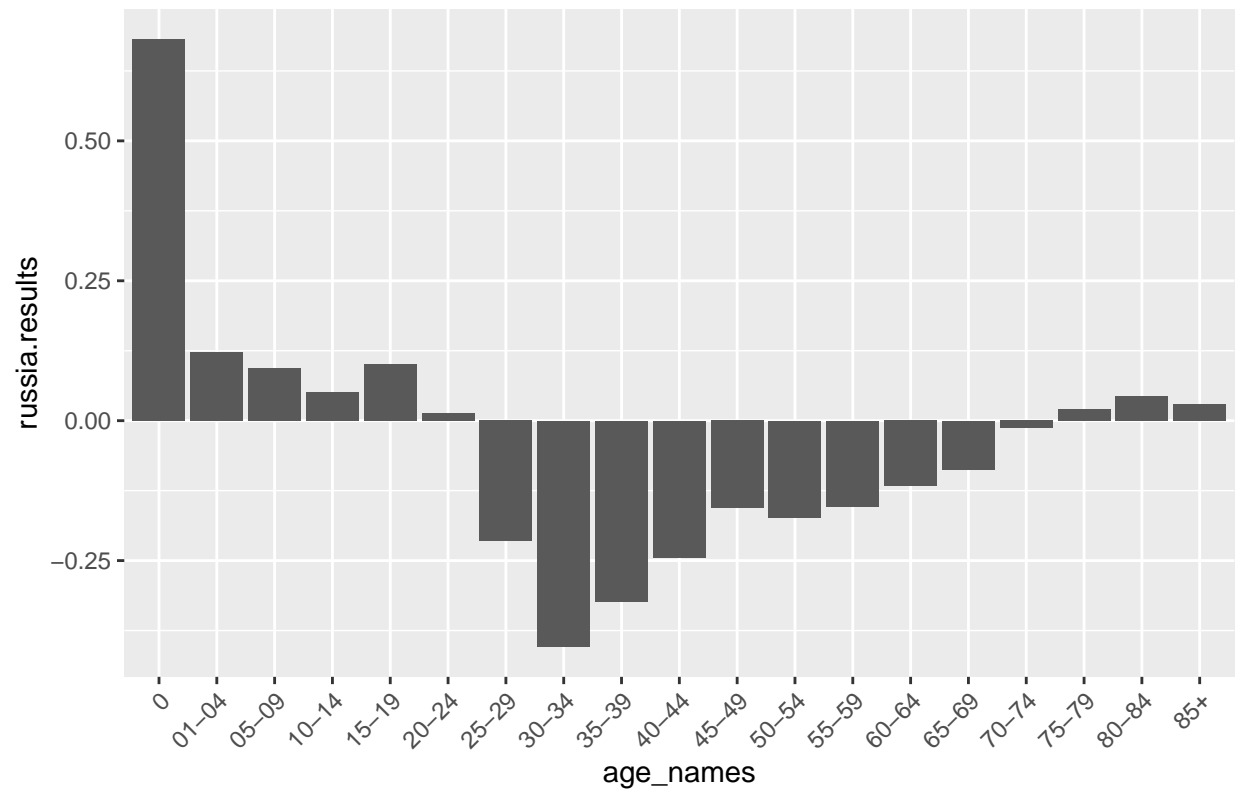
# Latvia
ggplot()+
  ggtitle(bquote(~'Change in ' ~ e[0] ~ '1990-2010' ))+
  geom_bar(aes(x = age_names, y = latvia.results), stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



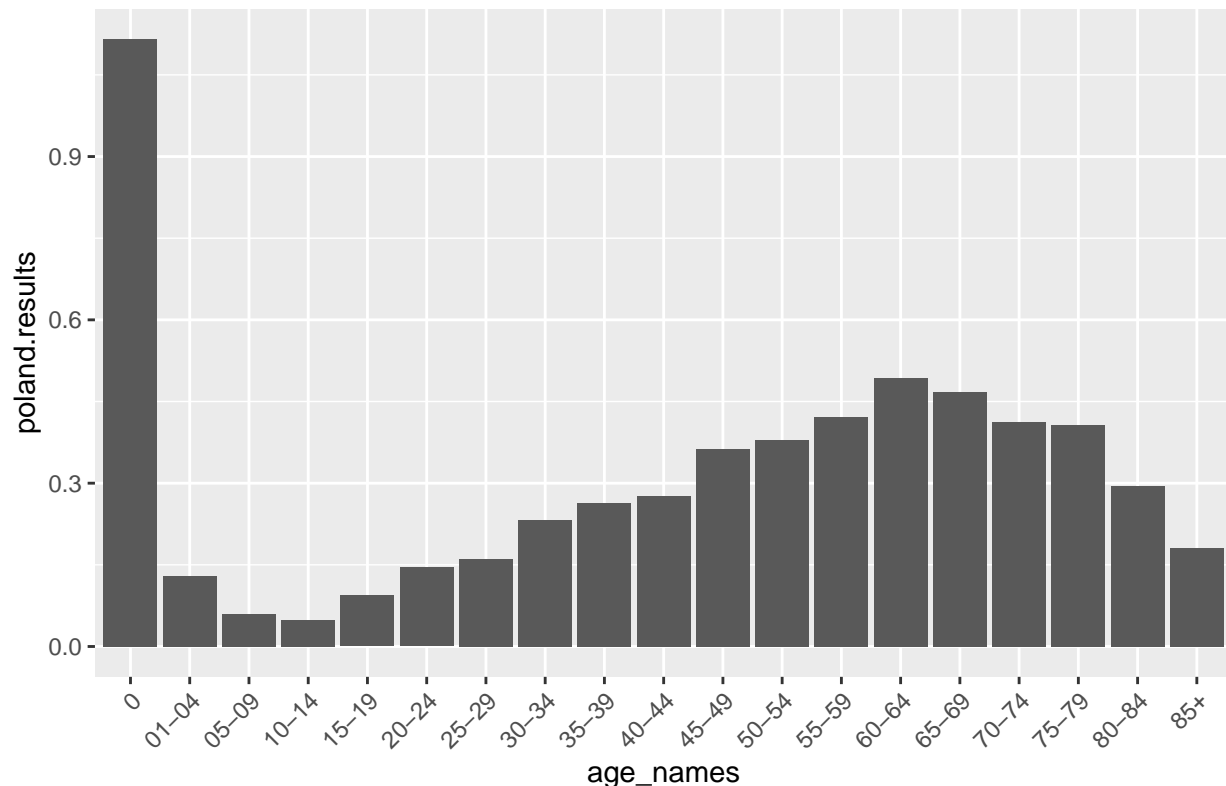
```
# Russia
ggplot()+
  ggtitle(bquote(~'Change in ' ~ e[0] ~'1990-2010' ))+
  geom_bar(aes(x = age_names, y= russia.results), stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

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 expectancy. 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 consumption 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 to 2010.

```
### Extending to cause specific results

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

e0frommx <- 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])
```

```

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])

##### 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

```

```

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

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

## [1] -8.141512e-07 -6.621032e-08 -1.934737e-05

```

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

```

## 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

```

```

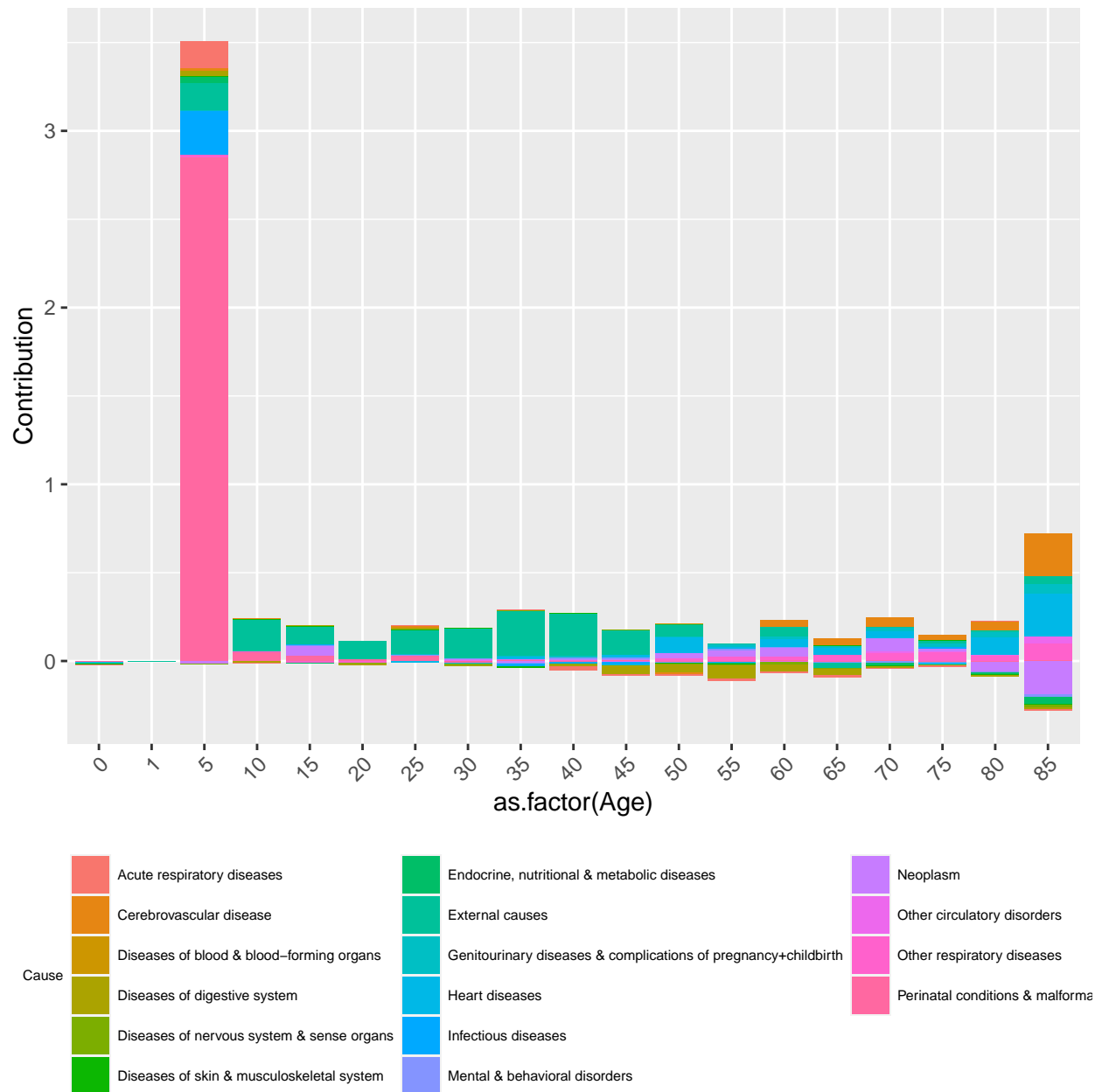
results.c.poland <- data.frame(results.c.poland)
colnames(results.c.poland) <- cause_names
results.c.poland$Age <- c(0,1,seq(5,85,5))
rownames(results.c.poland) <- age_names
results.c.poland <- gather(data = results.c.poland,key = Cause,value = Contribution,-Age)

## 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))

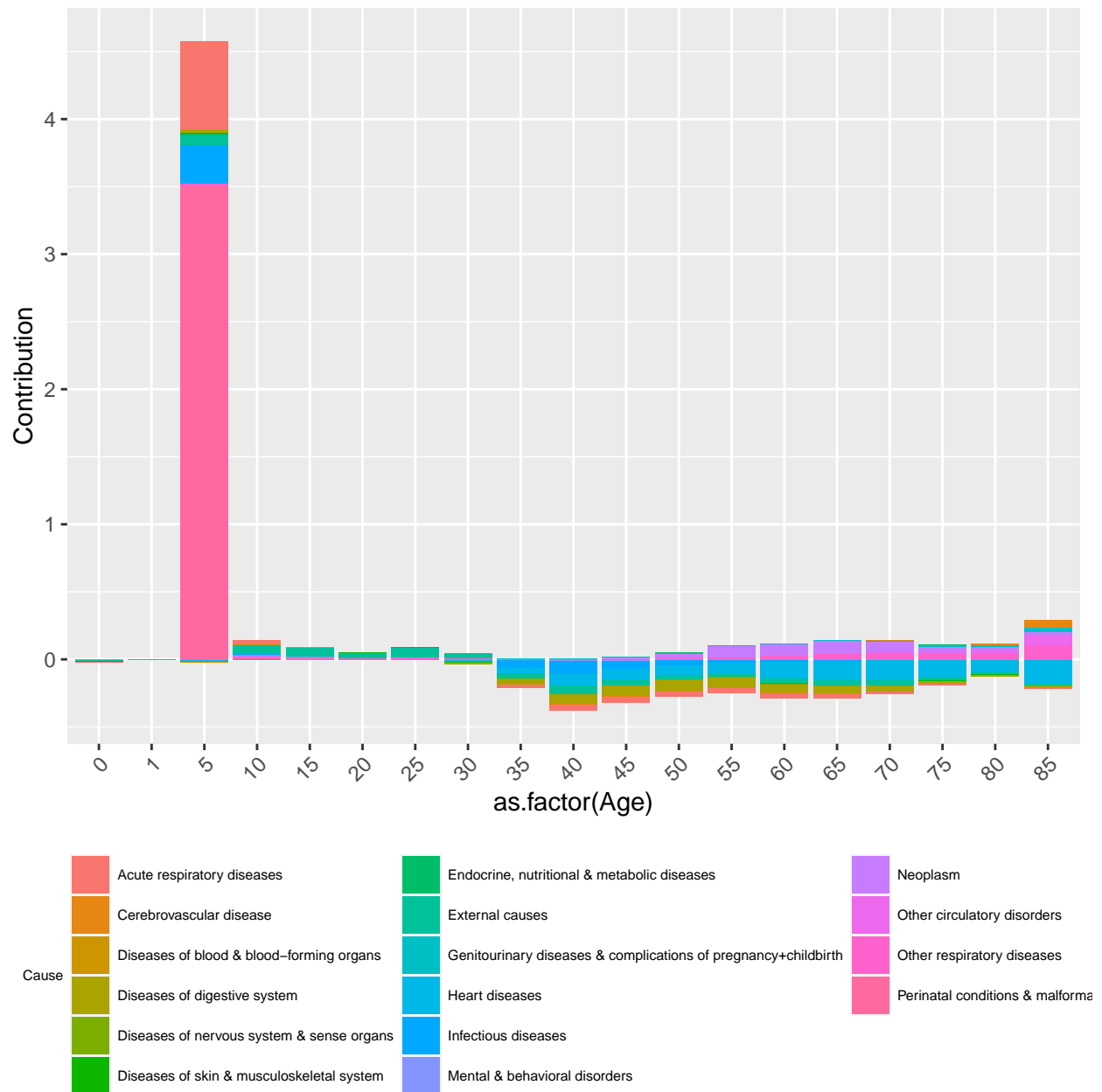
```

Change in e_0 1990–2010 in Latvia



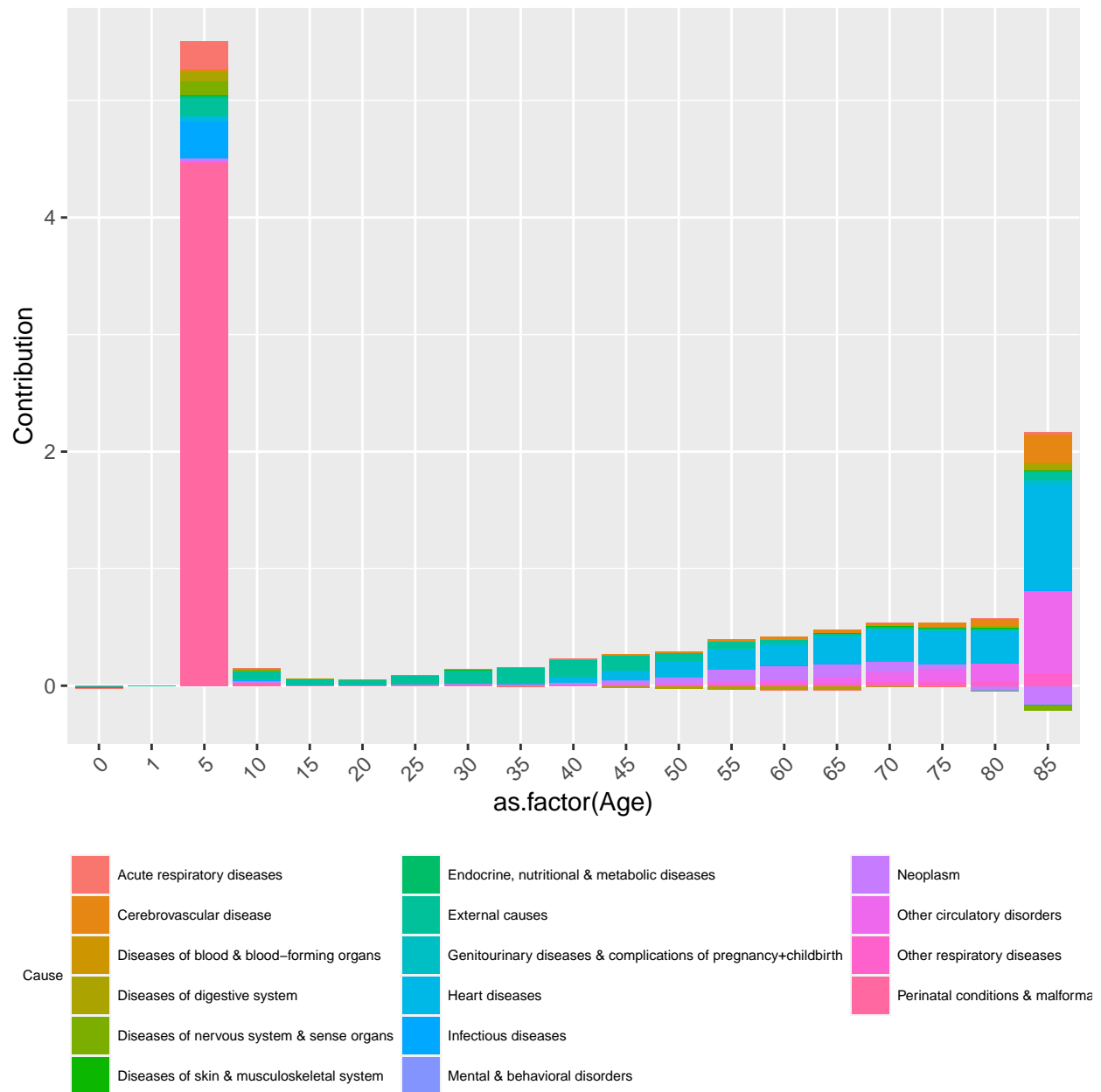
```
# Russia
ggplot(data=results.c.russia,
  aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ e[0] ~'1990-2010 in Russia' ))+
  geom_bar(stat = "identity", position = "stack") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
    legend.direction = "horizontal",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

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

## [1] 3.980185e-06 1.251437e-06 7.661509e-06

## 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
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)

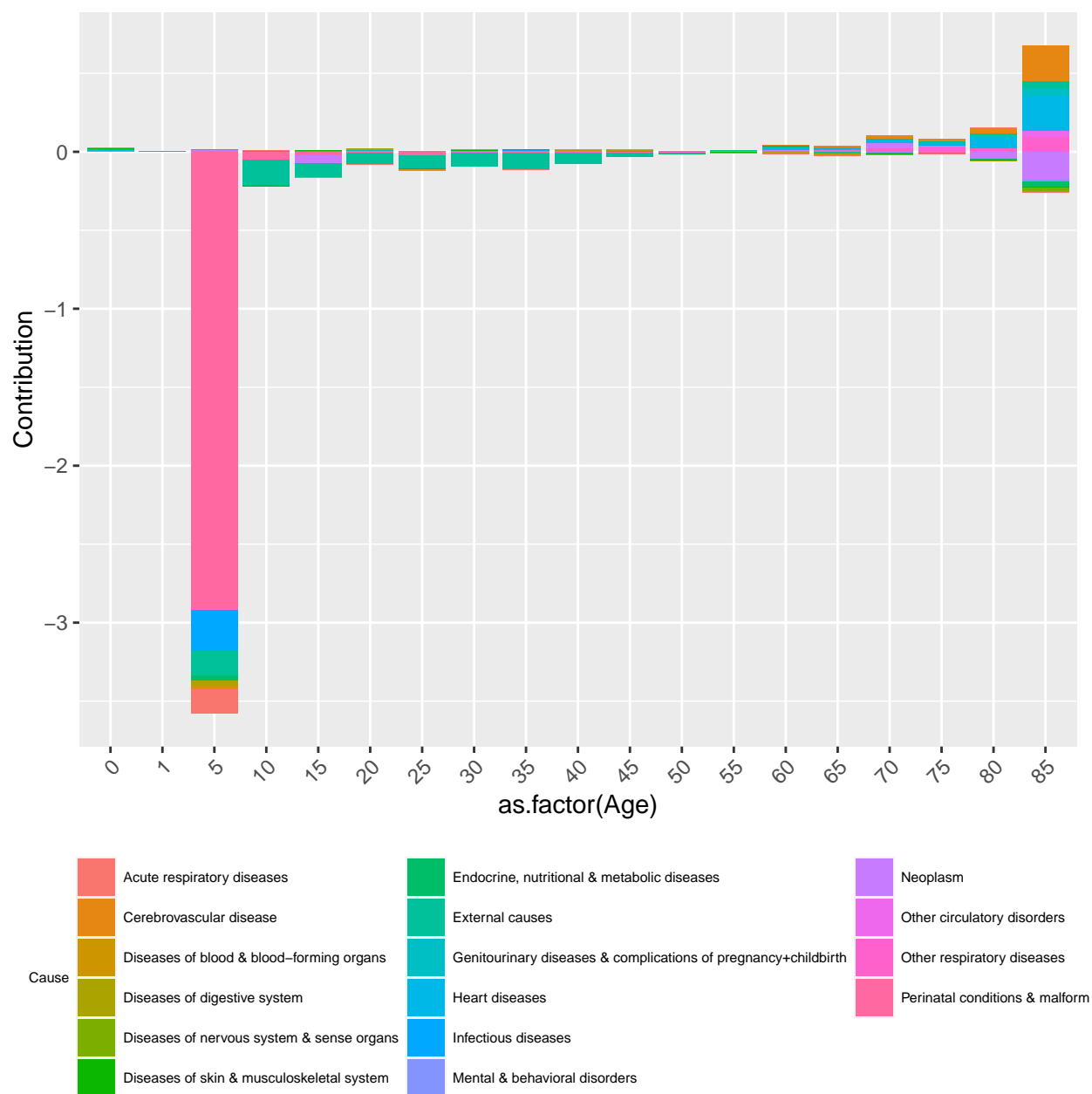
# Poland
results.sd.poland <- data.frame(results.sd.poland)
colnames(results.sd.poland) <- cause_names
results.sd.poland$Age <- c(0,1,seq(5,85,5))
rownames(results.sd.poland) <- age_names
results.sd.poland <- gather(data = results.sd.poland,key = Cause,value = Contribution,-Age)

## 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))

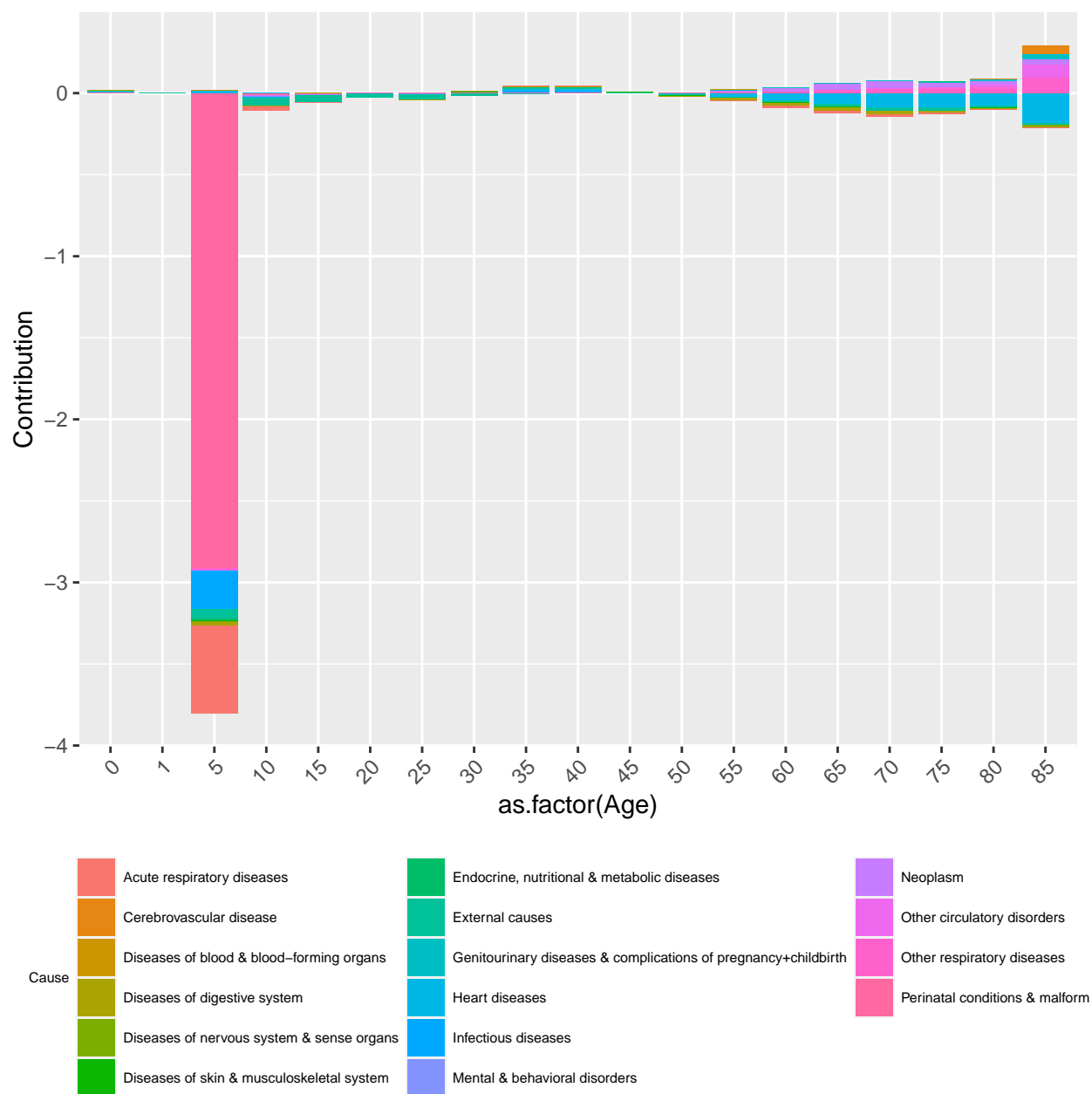
```

Change in sd_0 1990–2010 in Latvia

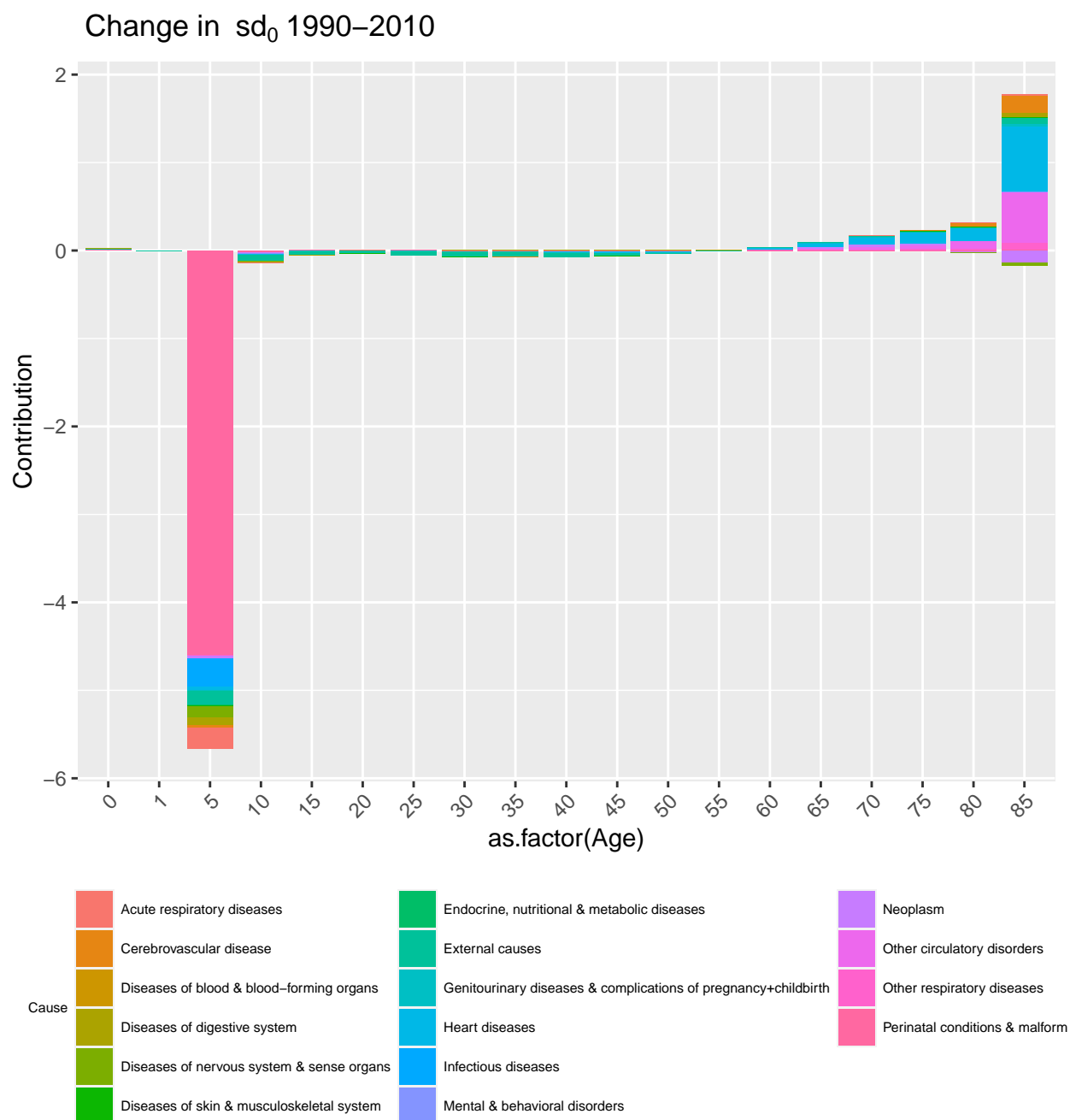


```
# Russia
ggplot(data=results.sd.russia,
  aes(x=as.factor(Age), y=Contribution, fill=Cause))+
  ggtitle(bquote(~'Change in '~ sd[0] ~'1990-2010 in Russia' ))+
  geom_bar(stat = "identity", position = "stack")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
    legend.direction = "horizontal",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



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 observe 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.