Day 2, exercise 4: Age- and cause-specific decomposition of a difference between life expectancies

2023-05-18

The aim of this exercise is to illustrate how to decompose by age and cause the difference between two life expectancies, using the Arriaga method. While there are many methods that would lead to similar results, the Arriaga method is the most common for this specific calculation. Specifically, we will use data from Aburto et al. (2022), analysing the change in male life expectancy at birth in the USA between 2019 and 2020. The dataset contains the relative lifetables.

First load the data and the necessary packages.

```
load('Arriaga_COVID_USA.RData')

library(tidyverse)
library(data.table)
```

First, let's look at the change in male life expectancy.

```
data %>% filter(year==2020,age==0,sex==0) %>% pull(ex) %>% unique() -
  data %>% filter(year==2019,age==0,sex==0) %>% pull(ex) %>% unique()
```

```
## [1] -1.525293
```

Where does this decline come from? Let's investigate with the Arriaga decomposition.

Let's start by decomposing by age only. We need to extract the information needed by the Arriaga formulas from the lifetables.

```
l1 <- data %>%
  filter(year==2019,sex==0, cause=="cvd") %>%
  pull(lx)
l2 <- data %>%
  filter(year==2020,sex==0, cause=="cvd") %>%
  pull(lx)
d1 <- data %>%
  filter(year==2019,sex==0, cause=="cvd") %>%
  pull(dx)
d2 <- data %>%
  filter(year==2020,sex==0, cause=="cvd") %>%
  pull(dx)
L1 <- data %>%
  filter(year==2019,sex==0, cause=="cvd") %>%
  pull(Lx)
L2 <- data %>%
  filter(year==2020,sex==0, cause=="cvd") %>%
  pull(Lx)
T1 <- data %>%
  filter(year==2019,sex==0, cause=="cvd") %>%
  pull(Tx)
```

```r
T2 <- data %>%
  filter(year==2020,sex==0, cause=="cvd") %>%
  pull(Tx)
```

Now we can calculate each component. We will use the implementation suggested in Preston, Heuveline and Guillot (2001), which is the one you saw in class.

```r
LAG <- length(l1)

DE <- (l1/l1[1])*((L2/l2)-(L1/l1))
# Indirect and interaction effects
IE <- (T2[-1]/l1[1])*((l1[-LAG]/l2[-LAG])-(l1[-1]/l2[-1]))
# one extra value for the indirect component
# since there is only direct component in the last age group
IE <- c(IE,0)

## add both to get the overall age-decomposition
ALL_age <- DE+IE
```

Let's check our results by comparing them with the actual difference between the two life expectancies.

```r
# check
# difference in life expectancies
data %>% filter(year==2020,sex==0,age==0) %>% pull(ex) %>% unique() -
  data %>% filter(year==2019,sex==0,age==0) %>% pull(ex) %>% unique()
```

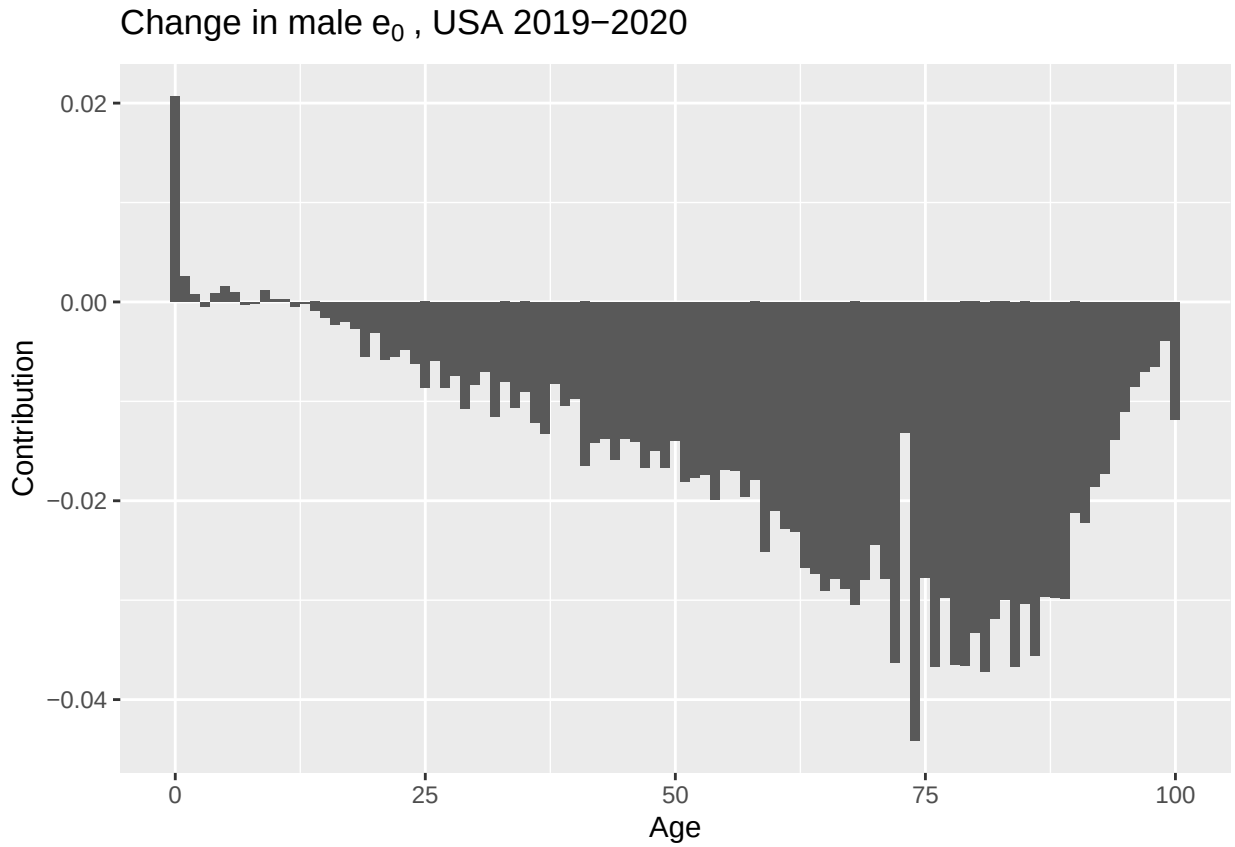
```
## [1] -1.525293
```

```r
# sum of age-specific contributions
sum(ALL_age)
```

```
## [1] -1.525293
```

Now let's plot the effects to see which ages contributed the most.

```r
ggplot() +
  ggtitle(bquote("Change in male"~"e"[0]~", USA 2019-2020")) +
  geom_bar(aes(x = unique(data$age), y = (ALL_age)), stat="identity") +
  scale_x_continuous("Age") +
  scale_y_continuous("Contribution")
```



Which ages contributed the most to the changes in life expectancy in the USA? Is this surprising?

Now let's try to add the cause-decomposition. We need to extract some additional elements.

```
mx1 <- data %>% filter(year==2019,sex==0) %>% pull(mx)
mx2 <- data %>% filter(year==2020,sex==0) %>% pull(mx)
rxi1 <- data %>% filter(year==2019,sex==0) %>% pull(prop)
rxi2 <- data %>% filter(year==2020,sex==0) %>% pull(prop)
causes <- data %>% filter(year==2020,sex==0) %>% pull(cause)
```

With these elements we can calculate the age- and cause-specific contributions, again following the formula in Preston, Heuveline and Guillot (2001).

```
ALL_cause <- ((rxi2*mx2-rxi1*mx1)/(mx2-mx1)) %>%
  cbind(age=0:100,
        cause = causes) %>%
  as.data.table() %>%
  mutate(cause = factor(cause,
                        labels=levels(causes))) %>%
  rename(cause_multiplier = ".") %>%
  inner_join(ALL_age %>%
               cbind(age=0:100) %>%
               as.data.table() %>%
               rename(C = "."), by="age") %>%
  mutate(C_cause = cause_multiplier*C)
```

Do the results match?

```r
# difference in life expectancies
data %>% filter(year==2020,sex==0,age==0) %>% pull(ex) %>% unique() -
  data %>% filter(year==2019,sex==0,age==0) %>% pull(ex) %>% unique()
```

```
## [1] -1.525293
```

```r
# sum of age-specific contributions
sum(ALL_age)
```

```
## [1] -1.525293
```

```r
sum(ALL_cause$C)/8
```

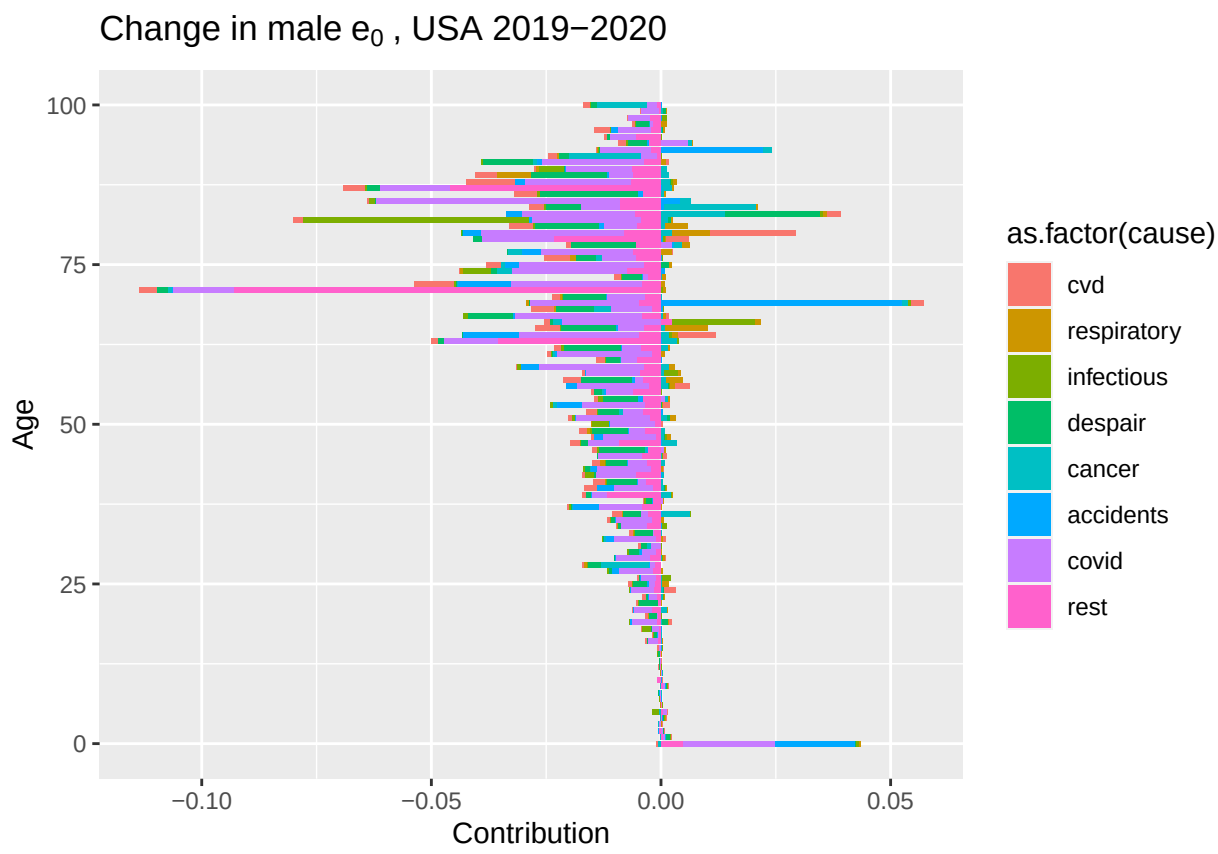
```
## [1] -1.525293
```

```r
# sum of age- and cause-specific contributions
sum(ALL_cause$C_cause)
```

```
## [1] -1.521742
```

Now we can plot the results

```r
ALL_cause %>%
ggplot() +
  ggtitle(bquote("Change in male"~"e"[0]~", USA 2019-2020")) +
  geom_bar(aes(x = age, y = (C_cause), fill=as.factor(cause)), stat="identity", position = "stack") +
  scale_x_continuous("Age") +
  scale_y_continuous("Contribution") +
  coord_flip()
```



What are the ages and causes that contributed the most to the decline in male life expectancy at birth in the uSA between 2019 and 2020?

As was pointed out in class, this code has (at least) a mistake. What is it?

```
mx1 <- data %>% filter(year==2019,sex==0) %>% pull(mx)
mx2 <- data %>% filter(year==2020,sex==0) %>% pull(mx)
rxi1 <- data %>% filter(year==2019,sex==0) %>% pull(prop)
rxi2 <- data %>% filter(year==2020,sex==0) %>% pull(prop)
causes <- data %>% filter(year==2020,sex==0) %>% pull(cause)
```

In the previous snippet of code, I extract the cause-specific proportions of deaths and the cause labels separately. This was supposed to be a nice shortcut, because the causes are ordered, so they should stay in the same order. And they do.

```
causes
```

```
##   [1] cvd         respiratory infectious  despair     cancer      accidents
##   [7] covid       rest        cvd         respiratory infectious  despair
##  [13] cancer      accidents   covid       rest        cvd         respiratory
##  [19] infectious  despair     cancer      accidents   covid       rest
##  [25] cvd         respiratory infectious  despair     cancer      accidents
##  [31] covid       rest        cvd         respiratory infectious  despair
##  [37] cancer      accidents   covid       rest        cvd         respiratory
##  [43] infectious  despair     cancer      accidents   covid       rest
##  [49] cvd         respiratory infectious  despair     cancer      accidents
##  [55] covid       rest        cvd         respiratory infectious  despair
##  [61] cancer      accidents   covid       rest        cvd         respiratory
##  [67] infectious  despair     cancer      accidents   covid       rest
##  [73] cvd         respiratory infectious  despair     cancer      accidents
##  [79] covid       rest        cvd         respiratory infectious  despair
##  [85] cancer      accidents   covid       rest        cvd         respiratory
##  [91] infectious  despair     cancer      accidents   covid       rest
##  [97] cvd         respiratory infectious  despair     cancer      accidents
## [103] covid       rest        cvd         respiratory infectious  despair
## [109] cancer      accidents   covid       rest        cvd         respiratory
## [115] infectious  despair     cancer      accidents   covid       rest
## [121] cvd         respiratory infectious  despair     cancer      accidents
## [127] covid       rest        cvd         respiratory infectious  despair
## [133] cancer      accidents   covid       rest        cvd         respiratory
## [139] infectious  despair     cancer      accidents   covid       rest
## [145] cvd         respiratory infectious  despair     cancer      accidents
## [151] covid       rest        cvd         respiratory infectious  despair
## [157] cancer      accidents   covid       rest        cvd         respiratory
## [163] infectious  despair     cancer      accidents   covid       rest
## [169] cvd         respiratory infectious  despair     cancer      accidents
## [175] covid       rest        cvd         respiratory infectious  despair
## [181] cancer      accidents   covid       rest        cvd         respiratory
## [187] infectious  despair     cancer      accidents   covid       rest
## [193] cvd         respiratory infectious  despair     cancer      accidents
## [199] covid       rest        cvd         respiratory infectious  despair
## [205] cancer      accidents   covid       rest        cvd         respiratory
## [211] infectious  despair     cancer      accidents   covid       rest
## [217] cvd         respiratory infectious  despair     cancer      accidents
## [223] covid       rest        cvd         respiratory infectious  despair
## [229] cancer      accidents   covid       rest        cvd         respiratory
```

```
## [235] infectious  despair     cancer      accidents   covid       rest
## [241] cvd         respiratory infectious  despair     cancer      accidents
## [247] covid       rest        cvd         respiratory infectious  despair
## [253] cancer      accidents   covid       rest        cvd         respiratory
## [259] infectious  despair     cancer      accidents   covid       rest
## [265] cvd         respiratory infectious  despair     cancer      accidents
## [271] covid       rest        cvd         respiratory infectious  despair
## [277] cancer      accidents   covid       rest        cvd         respiratory
## [283] infectious  despair     cancer      accidents   covid       rest
## [289] cvd         respiratory infectious  despair     cancer      accidents
## [295] covid       rest        cvd         respiratory infectious  despair
## [301] cancer      accidents   covid       rest        cvd         respiratory
## [307] infectious  despair     cancer      accidents   covid       rest
## [313] cvd         respiratory infectious  despair     cancer      accidents
## [319] covid       rest        cvd         respiratory infectious  despair
## [325] cancer      accidents   covid       rest        cvd         respiratory
## [331] infectious  despair     cancer      accidents   covid       rest
## [337] cvd         respiratory infectious  despair     cancer      accidents
## [343] covid       rest        cvd         respiratory infectious  despair
## [349] cancer      accidents   covid       rest        cvd         respiratory
## [355] infectious  despair     cancer      accidents   covid       rest
## [361] cvd         respiratory infectious  despair     cancer      accidents
## [367] covid       rest        cvd         respiratory infectious  despair
## [373] cancer      accidents   covid       rest        cvd         respiratory
## [379] infectious  despair     cancer      accidents   covid       rest
## [385] cvd         respiratory infectious  despair     cancer      accidents
## [391] covid       rest        cvd         respiratory infectious  despair
## [397] cancer      accidents   covid       rest        cvd         respiratory
## [403] infectious  despair     cancer      accidents   covid       rest
## [409] cvd         respiratory infectious  despair     cancer      accidents
## [415] covid       rest        cvd         respiratory infectious  despair
## [421] cancer      accidents   covid       rest        cvd         respiratory
## [427] infectious  despair     cancer      accidents   covid       rest
## [433] cvd         respiratory infectious  despair     cancer      accidents
## [439] covid       rest        cvd         respiratory infectious  despair
## [445] cancer      accidents   covid       rest        cvd         respiratory
## [451] infectious  despair     cancer      accidents   covid       rest
## [457] cvd         respiratory infectious  despair     cancer      accidents
## [463] covid       rest        cvd         respiratory infectious  despair
## [469] cancer      accidents   covid       rest        cvd         respiratory
## [475] infectious  despair     cancer      accidents   covid       rest
## [481] cvd         respiratory infectious  despair     cancer      accidents
## [487] covid       rest        cvd         respiratory infectious  despair
## [493] cancer      accidents   covid       rest        cvd         respiratory
## [499] infectious  despair     cancer      accidents   covid       rest
## [505] cvd         respiratory infectious  despair     cancer      accidents
## [511] covid       rest        cvd         respiratory infectious  despair
## [517] cancer      accidents   covid       rest        cvd         respiratory
## [523] infectious  despair     cancer      accidents   covid       rest
## [529] cvd         respiratory infectious  despair     cancer      accidents
## [535] covid       rest        cvd         respiratory infectious  despair
## [541] cancer      accidents   covid       rest        cvd         respiratory
## [547] infectious  despair     cancer      accidents   covid       rest
## [553] cvd         respiratory infectious  despair     cancer      accidents
```

```
## [559] covid       rest        cvd         respiratory infectious  despair
## [565] cancer      accidents   covid       rest        cvd         respiratory
## [571] infectious  despair     cancer      accidents   covid       rest
## [577] cvd         respiratory infectious  despair     cancer      accidents
## [583] covid       rest        cvd         respiratory infectious  despair
## [589] cancer      accidents   covid       rest        cvd         respiratory
## [595] infectious  despair     cancer      accidents   covid       rest
## [601] cvd         respiratory infectious  despair     cancer      accidents
## [607] covid       rest        cvd         respiratory infectious  despair
## [613] cancer      accidents   covid       rest        cvd         respiratory
## [619] infectious  despair     cancer      accidents   covid       rest
## [625] cvd         respiratory infectious  despair     cancer      accidents
## [631] covid       rest        cvd         respiratory infectious  despair
## [637] cancer      accidents   covid       rest        cvd         respiratory
## [643] infectious  despair     cancer      accidents   covid       rest
## [649] cvd         respiratory infectious  despair     cancer      accidents
## [655] covid       rest        cvd         respiratory infectious  despair
## [661] cancer      accidents   covid       rest        cvd         respiratory
## [667] infectious  despair     cancer      accidents   covid       rest
## [673] cvd         respiratory infectious  despair     cancer      accidents
## [679] covid       rest        cvd         respiratory infectious  despair
## [685] cancer      accidents   covid       rest        cvd         respiratory
## [691] infectious  despair     cancer      accidents   covid       rest
## [697] cvd         respiratory infectious  despair     cancer      accidents
## [703] covid       rest        cvd         respiratory infectious  despair
## [709] cancer      accidents   covid       rest        cvd         respiratory
## [715] infectious  despair     cancer      accidents   covid       rest
## [721] cvd         respiratory infectious  despair     cancer      accidents
## [727] covid       rest        cvd         respiratory infectious  despair
## [733] cancer      accidents   covid       rest        cvd         respiratory
## [739] infectious  despair     cancer      accidents   covid       rest
## [745] cvd         respiratory infectious  despair     cancer      accidents
## [751] covid       rest        cvd         respiratory infectious  despair
## [757] cancer      accidents   covid       rest        cvd         respiratory
## [763] infectious  despair     cancer      accidents   covid       rest
## [769] cvd         respiratory infectious  despair     cancer      accidents
## [775] covid       rest        cvd         respiratory infectious  despair
## [781] cancer      accidents   covid       rest        cvd         respiratory
## [787] infectious  despair     cancer      accidents   covid       rest
## [793] cvd         respiratory infectious  despair     cancer      accidents
## [799] covid       rest        cvd         respiratory infectious  despair
## [805] cancer      accidents   covid       rest
## Levels: cvd respiratory infectious despair cancer accidents covid rest
```

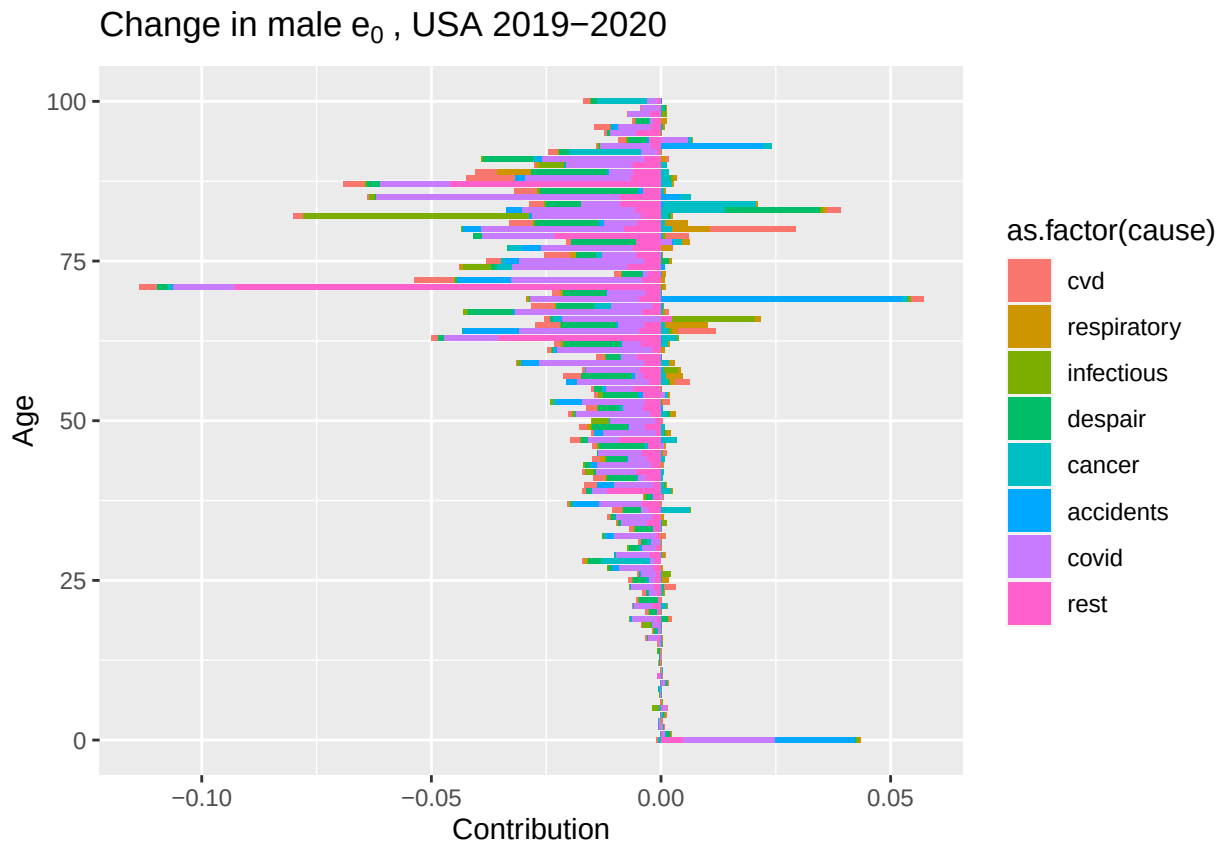
However, the original datasets repeats each age eight times in a row. Instead, when I put the cause labels and the cause-specific multiplier together, I added the age as a vector of 0 to 100, to be repeated eight times. Now, the causes do not match their contributions anymore.

```
ALL_cause <- ((rxi2*mx2-rxi1*mx1)/(mx2-mx1)) %>%
  cbind(age=0:100, # here is the age vector
        cause = causes) %>%
  as.data.table() %>%
  mutate(cause = factor(cause,
                        labels=levels(causes))) %>%
  rename(cause_multiplier = ".") %>%
```

```
  inner_join(ALL_age %>%
               cbind(age=0:100) %>%
               as.data.table() %>%
               rename(C = "."), by="age") %>%
  mutate(C_cause = cause_multiplier*C)
```

We can also see that in how disorganised the graph looks. Causes have very different contributions from one age to the next (this was also pointed out in class), while they usually have smoother patterns.

```
ALL_cause %>%
ggplot() +
  ggtitle(bquote("Change in male"~"e"[0]~", USA 2019-2020")) +
  geom_bar(aes(x = age, y = (C_cause), fill=as.factor(cause)), stat="identity", position = "stack") +
  scale_x_continuous("Age") +
  scale_y_continuous("Contribution") +
  coord_flip()
```



Let's correct this.

Instead of pulling out each element, let's reshape the data, while keeping the ages, causes and proportions always together.

```
ALL_cause_correct <- data %>%
  filter(year==2019, sex==0) %>%
  inner_join(data %>%
               filter(year==2020, sex==0), by=c("age","cause"), suffix=c("_2019","_2020")) %>% # By hav
  mutate(cause_multiplier = (prop_2020*mx_2020-prop_2019*mx_2019)/(mx_2020-mx_2019)) %>%
  select(age, cause, cause_multiplier) %>%
```

```
  inner_join(ALL_age %>%
               cbind(age=0:100) %>%
               as.data.table() %>%
               rename(C = "."), by="age") %>%
  mutate(C_cause = cause_multiplier*C)
```

Let's check the results. The sums of contributions look much better now.

```
sum(ALL_cause_correct$C_cause)
```

```
## [1] -1.525293
```

```
sum(ALL_cause_correct$C)/8
```

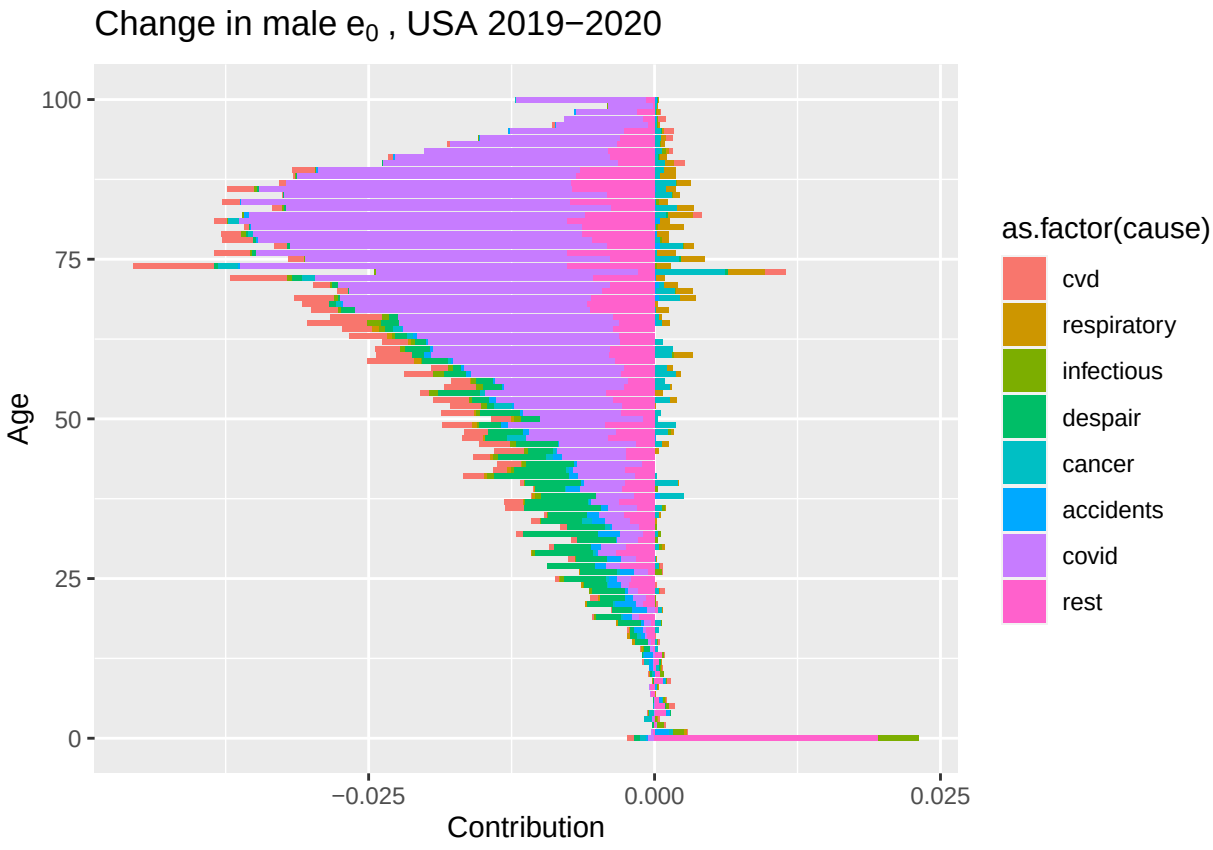
```
## [1] -1.525293
```

```
sum(ALL_age)
```

```
## [1] -1.525293
```

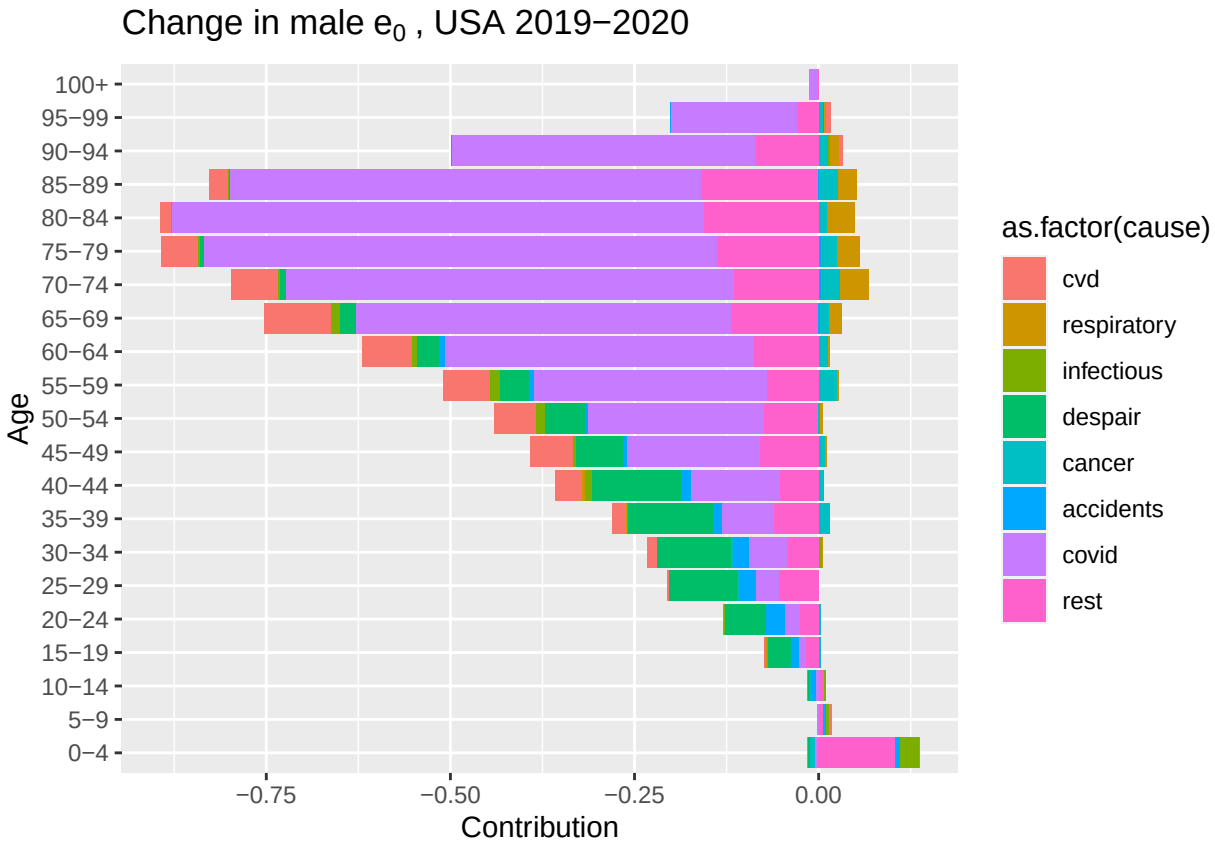
As does the graph.

```
ALL_cause_correct %>%
  ggplot() +
  ggtitle(bquote("Change in male"~"e"[0]~", USA 2019-2020")) +
  geom_bar(aes(x = age, y = (C_cause), fill=as.factor(cause)), stat="identity", position = "stack") +
  scale_x_continuous("Age") +
  scale_y_continuous("Contribution") +
  coord_flip()
```



Even though the graph looks smoother now, we can group ages in 5-year groups.

```
ALL_cause_correct %>%
  mutate(age_group = case_when(age %in% 0:4 ~ "0-4",
                               age %in% 5:9 ~ "5-9",
                               age %in% 10:14 ~ "10-14",
                               age %in% 15:19 ~ "15-19",
                               age %in% 20:24 ~ "20-24",
                               age %in% 25:29 ~ "25-29",
                               age %in% 30:34 ~ "30-34",
                               age %in% 35:39 ~ "35-39",
                               age %in% 40:44 ~ "40-44",
                               age %in% 45:49 ~ "45-49",
                               age %in% 50:54 ~ "50-54",
                               age %in% 55:59 ~ "55-59",
                               age %in% 60:64 ~ "60-64",
                               age %in% 65:69 ~ "65-69",
                               age %in% 70:74 ~ "70-74",
                               age %in% 75:79 ~ "75-79",
                               age %in% 80:84 ~ "80-84",
                               age %in% 85:89 ~ "85-89",
                               age %in% 90:94 ~ "90-94",
                               age %in% 95:99 ~ "95-99",
                               age %in% 100 ~ "100+"),
         age_group = factor(age_group, levels = c("0-4",
                                                  "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+"))) %>%
  group_by(age_group, cause) %>%
  mutate(C_cause_group = sum(C_cause)) %>%
  ggplot() +
  ggtitle(bquote("Change in male"~"e"[0]~", USA 2019-2020")) +
  geom_bar(aes(x = age_group, y = (C_cause_group), fill=as.factor(cause)), stat="identity", position =
  scale_x_discrete("Age") +
  scale_y_continuous("Contribution") +
  coord_flip()
```


Change in male e_0 , USA 2019−2020
100+
95−99
90−94
85−89
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
5−9
0−4
Age
−0.75
−0.50
−0.25
0.00
Contribution
as.factor(cause)
cvd
respiratory
infectious
despair
cancer
accidents
covid
rest