ECON 125 Problem Set 2 Solution

Question 1

Calculate female, male, and total age-specific mortality rates. Create a table that reports the average age-specific mortality rate by sex. On average, do females or males have higher age-specific mortality rates?

Answer: To calculate age-specific mortality rates per 1000, we divide deaths by population and multiply by 1000. To estimate average age-specific mortality rates, we use `summarise()`.

```r
# calculate mortality rates per 1000
df <-
  df |>
  mutate(female_mortrate = 1000*female_deaths/female_pop,
         male_mortrate = 1000*male_deaths/male_pop,
         total_mortrate = 1000*total_deaths/total_pop)

# summarize the means of the mortality rates
df |> summarise(female_mean = mean(female_mortrate),
                male_mean = mean(male_mortrate))
```

```
## # A tibble: 1 x 2
##   female_mean male_mean
##         <dbl>     <dbl>
## 1        84.1      96.8
```

We see that males on average have higher age-specific mortality rates

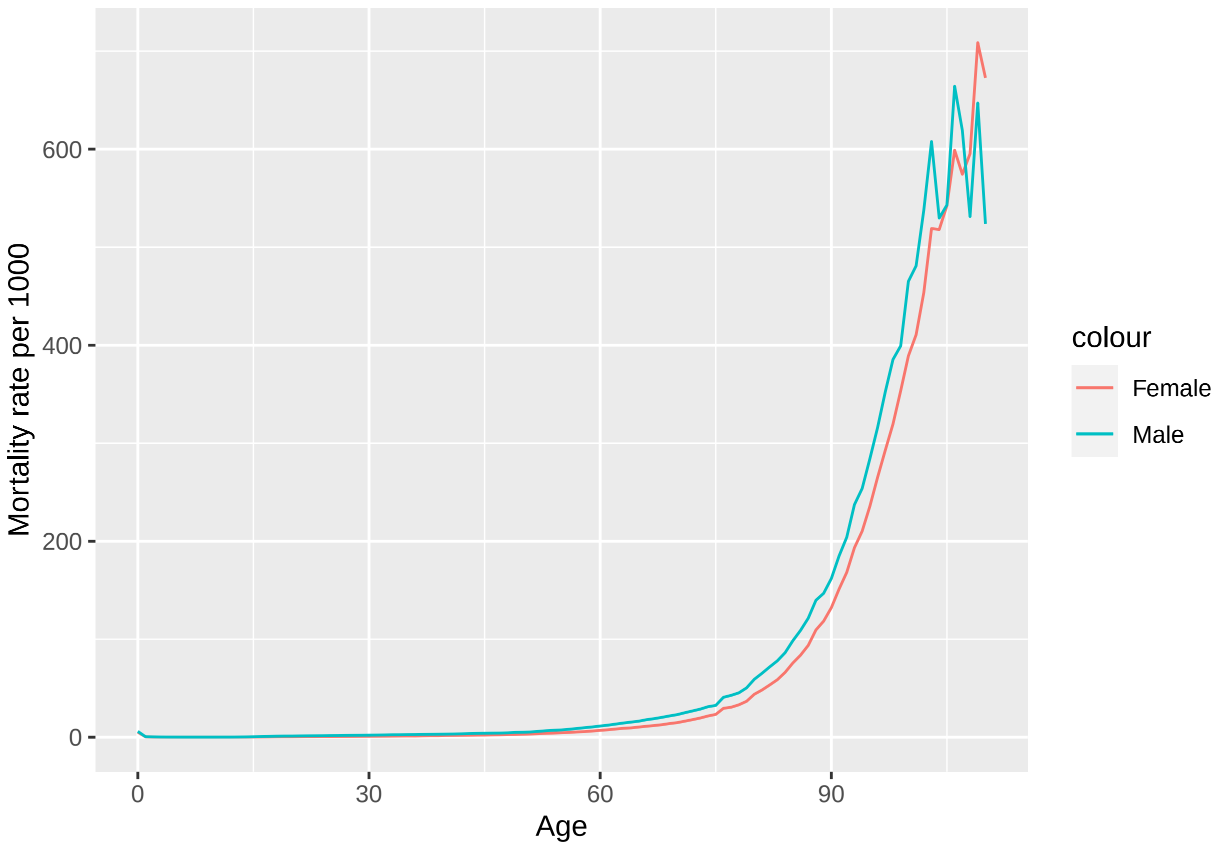
Question 2

Using only 2023 data, create two graphs. In the first, plot the age-specific mortality rate against age separately for females and males. In the second, plot the natural logarithm of the age-specific mortality rate against age separately for females and males. Are male mortality rates higher than female at every age? Is the male-female difference in mortality rates between larger among 20-30 year-olds or 80-90 year-olds? Is the male-female ratio of mortality rates larger among 20-30 year-olds or 90-100 year-olds?

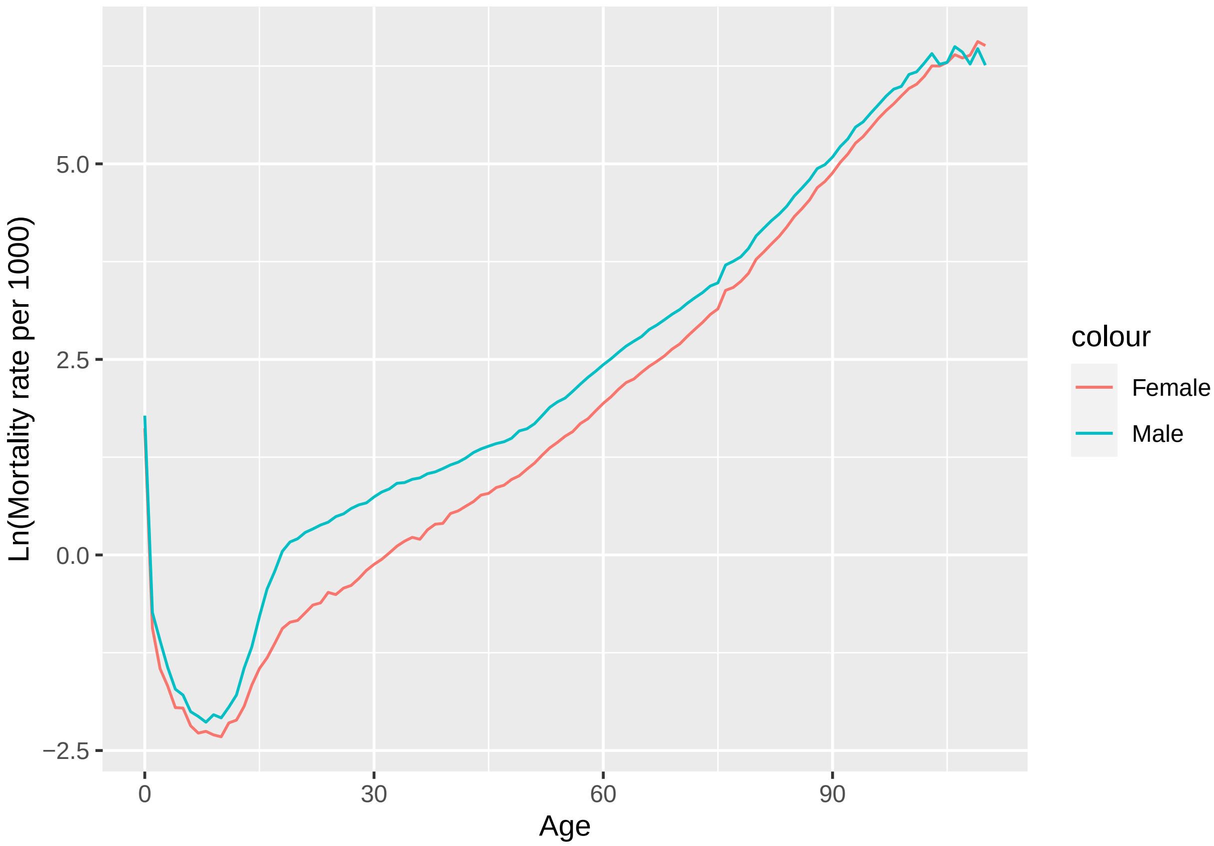
Answer: To plot 2023 rates by age, we split off a 2023-specific data frame. We then use `ggplot()` to draw two graphs.

```r
# create a 2023-specific data frame
df23 <- df |> filter(year==2023)

# plot female vs male mortality rates by age in 2023
ggplot(df23, aes(x = age)) +
  geom_line(aes(y = female_mortrate, color = "Female")) +
  geom_line(aes(y = male_mortrate, color = "Male")) +
  scale_y_continuous("Mortality rate per 1000") +
  scale_x_continuous("Age")
```




```r
# plot female vs male log mortality rates by age in 2023
ggplot(df23, aes(x = age)) +
  geom_line(aes(y = log(female_mortrate), color = "Female")) +
  geom_line(aes(y = log(male_mortrate), color = "Male")) +
  scale_y_continuous("Ln(Mortality rate per 1000)") +
  scale_x_continuous("Age")
```



Male mortality is higher than female mortality at every age but the very oldest, over age 100. The first graph shows that the male-female mortality *difference* is larger in the 80s than in the 20s. The second graph shows that the male-female mortality *ratio* is larger in the 20s than in the 80s, since $\ln(\text{male}) - \ln(\text{female}) = \ln(\frac{\text{male}}{\text{female}})$.

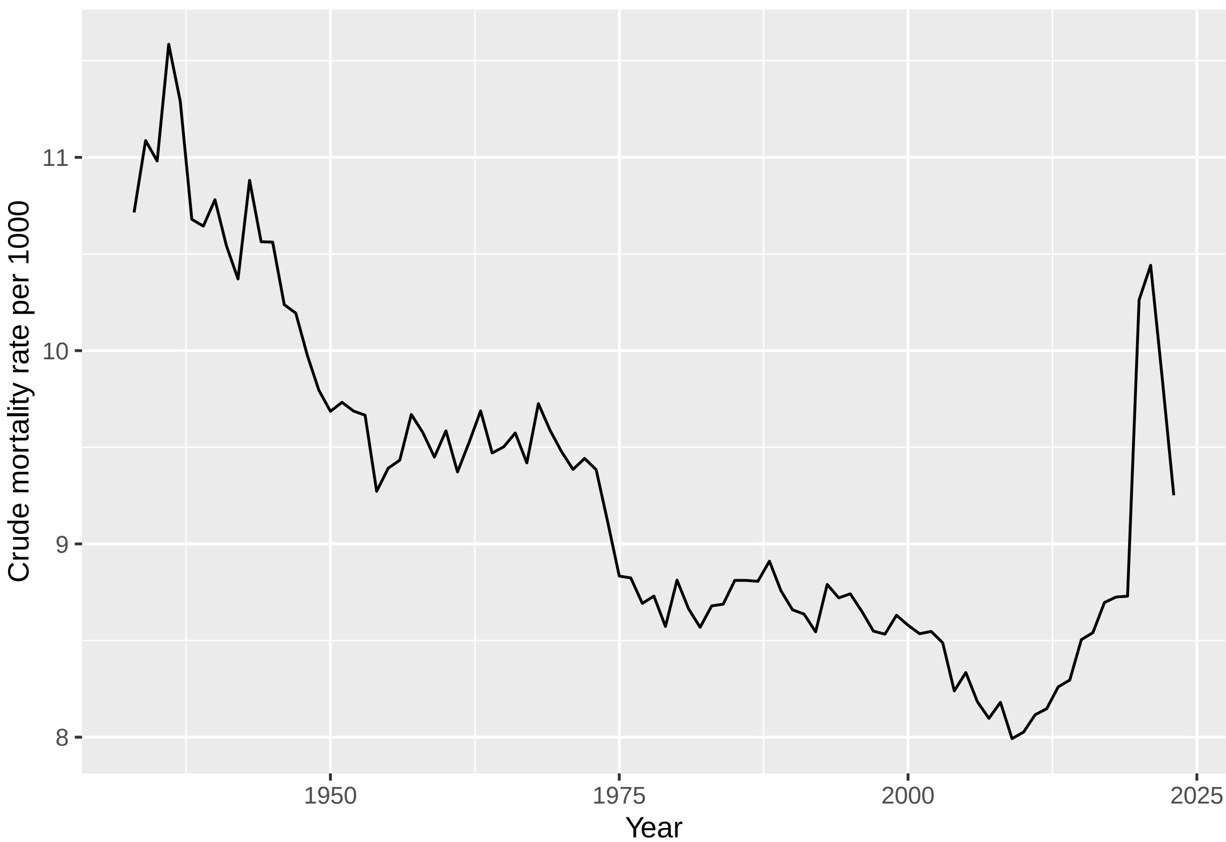
Question 3

Calculate the crude mortality rate in each year for the total population. Plot the crude mortality rate over time. Do you think the graph accurately reflects changes in the population's mortality burden? Why or why not?

Answer: To calculate the crude mortality rate, we can either (i) take the weighted sum of age-specific mortality rates, where the weights are age shares, or (ii) divide the sum of all deaths by the sum of all population (multiplied by 1000 if we want a rate per 1000). We'll need to save a new data frame at the year level, instead of the original age-year level. We can then use `ggplot()` to graph the results.

```r
# save crude mortality rate in a data frame with one row per year
annual_df <-
  df |>
  group_by(year) |>
  summarize(cmr = 1000*sum(total_deaths)/sum(total_pop))

# plot cmr
ggplot(annual_df, aes(x = year, y = cmr)) +
  geom_line() +
  scale_y_continuous("Crude mortality rate per 1000") +
  scale_x_continuous("Year")
```



This plot probably does not represent changes in the mortality burden very well because the crude mortality rate does not account for changes in age structure. If the population got older between 1933 and 2023, then the decline in the crude mortality rate understates the decline in age-specific mortality.

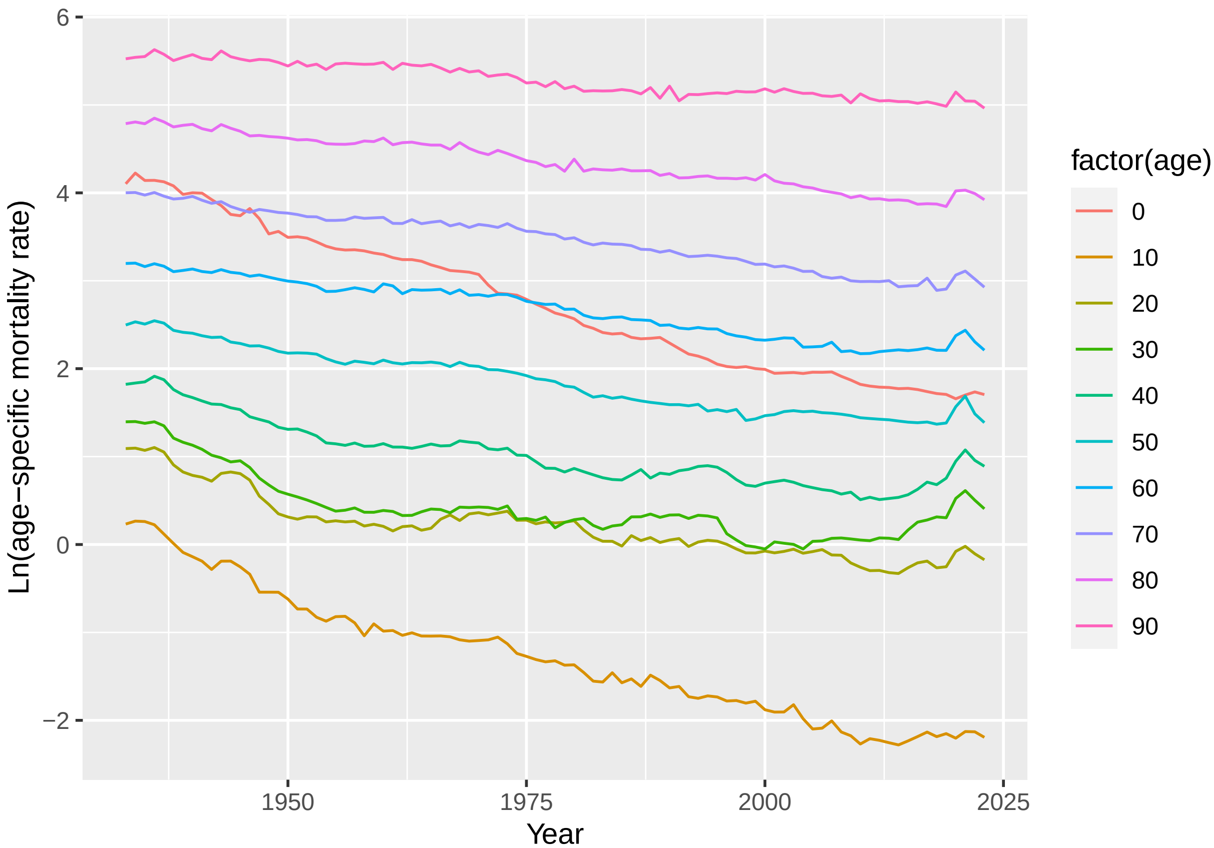
Question 4

Split off a new data frame that only keeps every tenth age (0, 10, $\cdots$ 90). Use it to draw a graph in which the x-axis is year, the y-axis is the natural logarithm of the age-specific mortality rate, and each age gets its own line plot with a different color. Which ages saw the largest proportional declines in mortality over the period? Which ages show evidence of a mortality increase during the COVID-19 pandemic?

Answer: To split off a data frame containing age multiples of 10, we use `filter()`. Inside this function, we could use a long "or" statement (`age==0|age==10|`$\cdots$) as we have done so far in class. Alternatively, we could use the modulo (remainder) operator `%%`: `age%%10` returns the remainder after dividing `age` by 10.

```
# Create new data frame with only age multiples of 10 from 0 to 90
df10 <- df |> filter(age%%10==0 & age<=90)

# Draw the plot
ggplot(df10, aes(x=year, y=log(total_mortrate), color=factor(age))) +
  geom_line() +
  scale_x_continuous("Year") +
  scale_y_continuous("Ln(age-specific mortality rate)")
```



The largest proportional declines were at younger ages like 0 and 10. The increase in mortality during the COVID-19 pandemic starting in 2020 is most visible in adulthood, past age 20.

Question 5

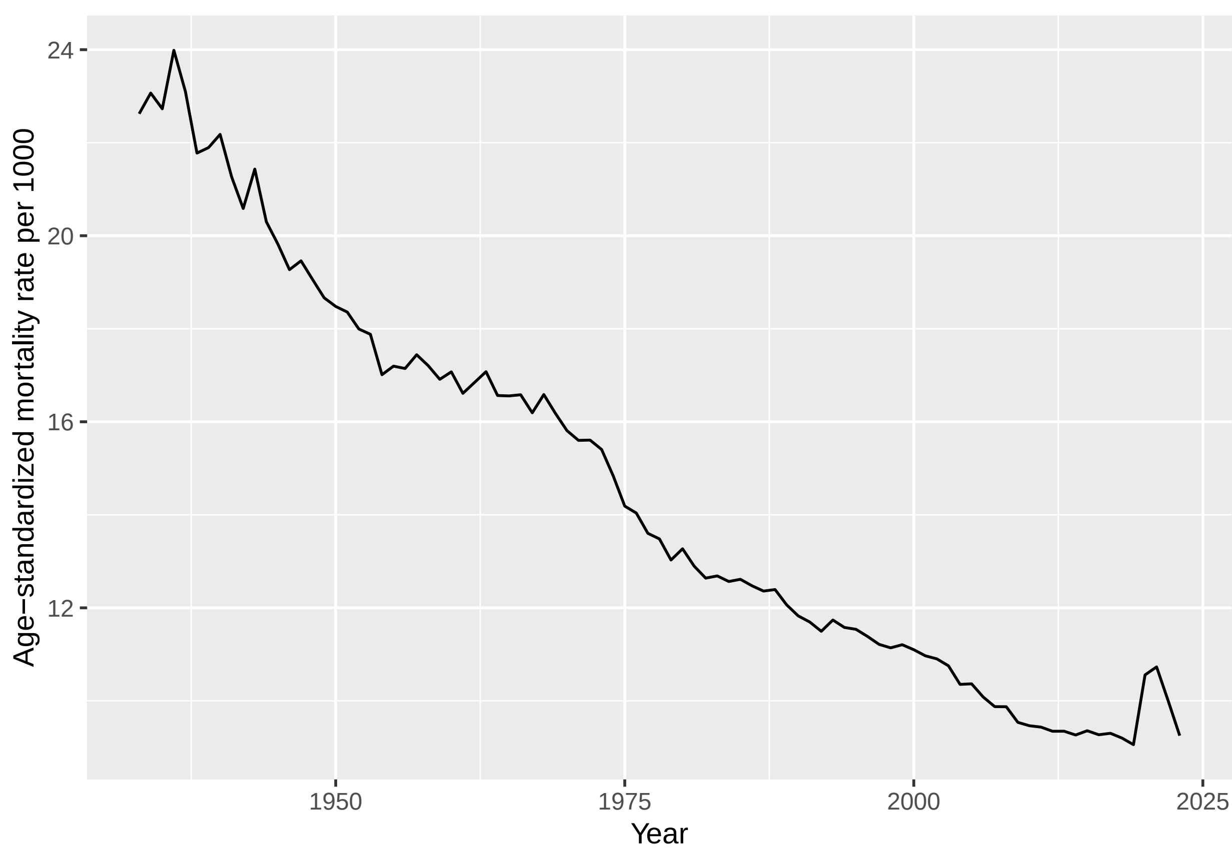
Calculate the age-standardized mortality rate using the population age structure in 2023. Plot the age-standardized mortality rate over time. How does the graph for the age-standardized mortality rate compare with the graph for the crude mortality rate in Question 3? Explain.

Answer: We apply the 2023 age structure to other years.

```
# replace the 2023 data frame with just ages and shares
df23 <-
  df23 |>
  mutate(share23 = total_pop / sum(total_pop)) |>
  select(age, share23)

# merge in 2023 shares, compute asmr, save new annual data frame
annual_df <-
  df |>
  left_join(df23, by = "age") |>
  group_by(year) |>
  summarise(asmr = sum(share23 * total_mortrate))

# plot asmr
ggplot(annual_df, aes(x = year, y = asmr)) +
  geom_line() +
  scale_y_continuous("Age-standardized mortality rate per 1000") +
  scale_x_continuous("Year")
```



The age-standardized mortality rate shows much steeper mortality decline between 1933 and present. The crude mortality rate declined from 11 to just under 10. The age-standardized mortality rate declined from over 20 to just under 10. That is consistent with the population getting older.

Question 6

Compute the period life table functions (q, l, d, L, T, e) in each year for the total population. Assume that deaths happen halfway through the year on average, and that anyone who survives to age 110 dies within the next year. Plot life expectancy at birth over time. How much did life expectancy at birth increase between 1933 and 2019 (just before the COVID-19 pandemic)? How much did life expectancy at birth decrease during the pandemic?

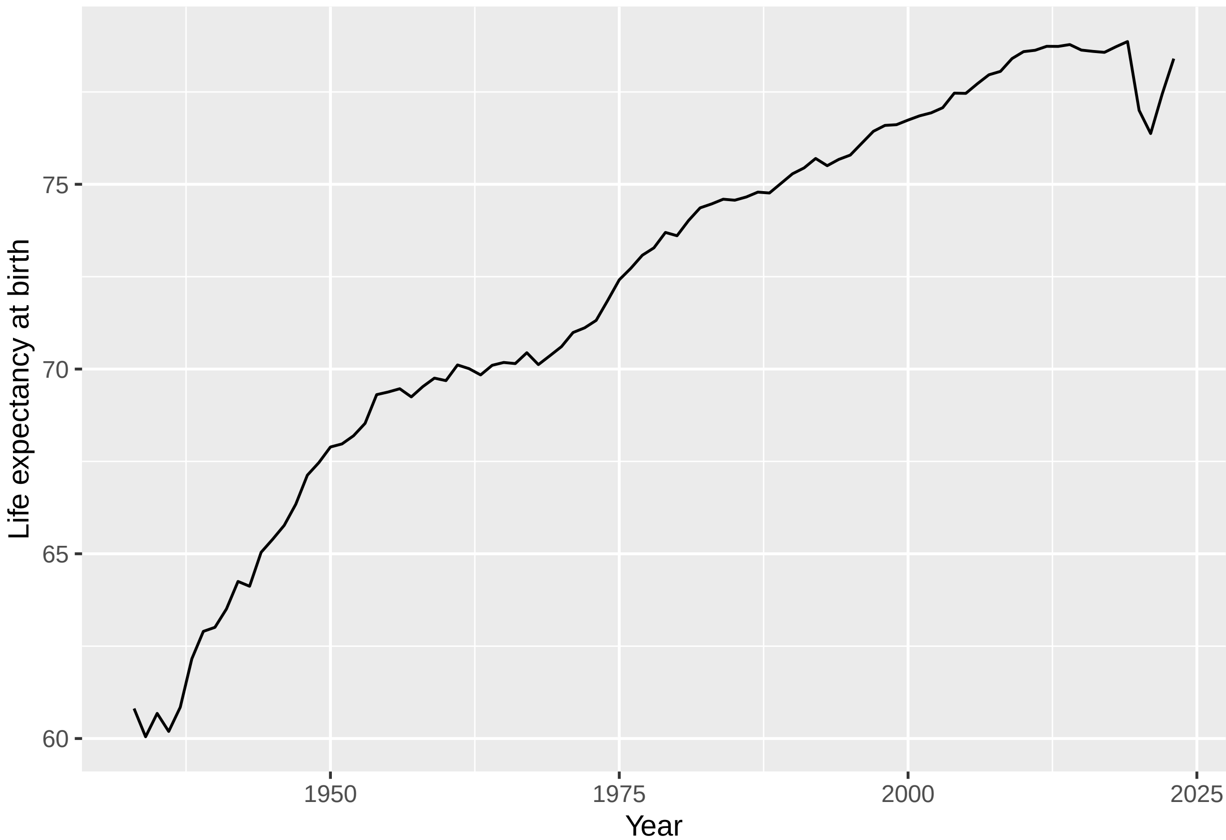
Answer: We need to follow the same procedures here as in the methods lecture, except we will group by year instead of country.

```r
# compute q as m/(1+.5m) for ages 0-109, and as 1 for age 110
df <-
  df |>
  mutate(q = if_else(age<110,
                     (total_mortrate/1000) / (1 + 0.5 * (total_mortrate/1000)),
                     1))

# compute the other period life table functions for every year
df <-
  df |>
  group_by(year) |>
  arrange(age) |>
  mutate(q = (total_mortrate/1000) / (1 + 0.5 * (total_mortrate/1000)),
         l = 100000 * cumprod(lag(1-q, default=1)),
         d = l*q,
         L = 0.5*d + (l-d),
         T = sum(L) - cumsum(L) + L,
         e = T/l)

# write over annual data frame with life expectancy at birth
annual_df <- df |> filter(age==0) |> select(year, age, e)

# graph life expectancy at birth over time
ggplot(annual_df, aes(x=year, y=e)) +
  geom_line() +
  scale_y_continuous("Life expectancy at birth") +
  scale_x_continuous("Year")
```



Life expectancy at birth rose from about 60 years in the 1930s to about 79 years in the 2010s. It declined by 2-3 years during the COVID-19 pandemic.

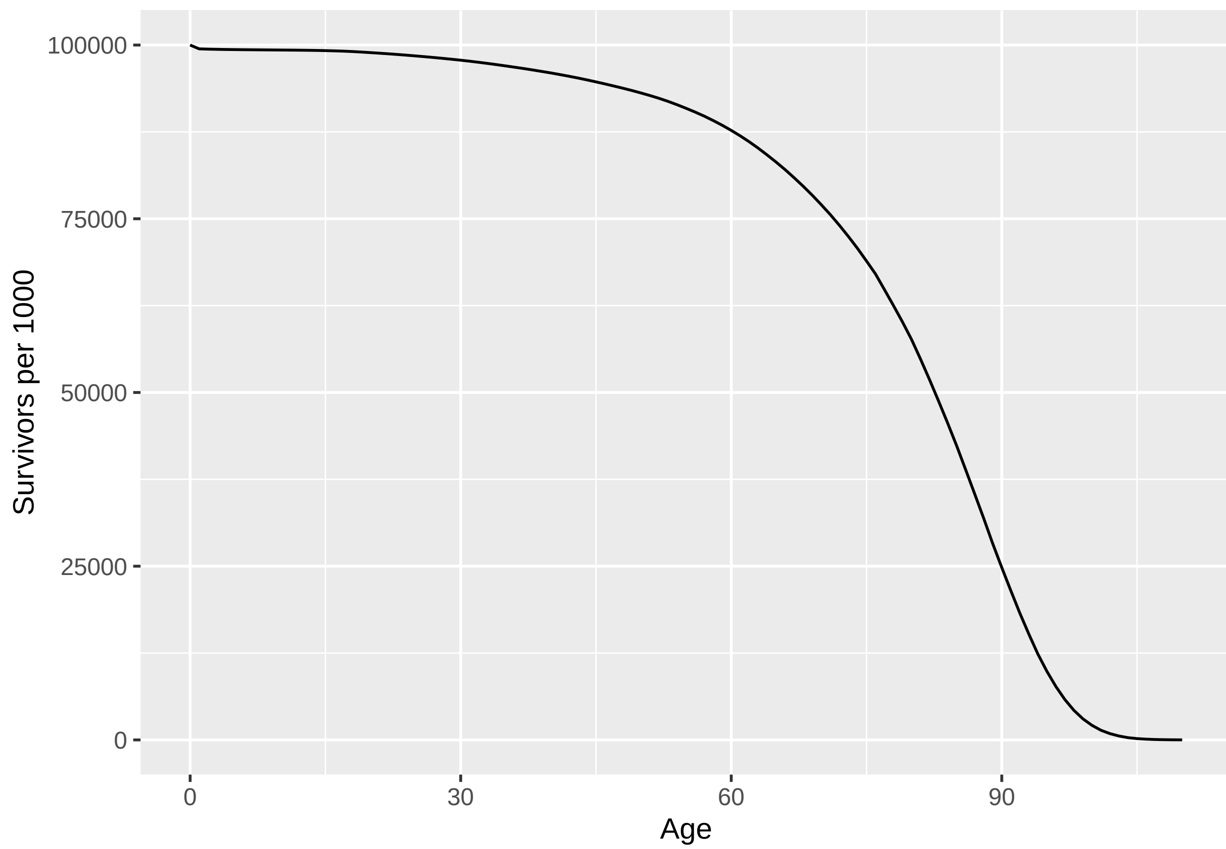
Question 7

Draw the survivorship curve (with age on the x-axis and l on the y-axis) for the total population in 2023. If a group of people were born and then experienced 2023 age-specific mortality rates over their lives, at roughly what age would half of them have died?

Answer: We already computed the survivorship function for every year in Question 6, so we just need to split off a 2023 data frame and draw a graph.

```r
# rewrite the 2023-specific data frame
df23 <- df |> filter(year==2023)

# plot survivorship curve in 2023
ggplot(df23, aes(x = age, y=l)) +
  geom_line() +
  scale_y_continuous("Survivors per 1000") +
  scale_x_continuous("Age")
```



The curve falls below 500,000 in the 80s, implying that half of the hypothetical group would have died by their 80s.

Question 8

Create a table of q_0 in every tenth year (1933, 1943, $\cdots$ 2023). How did the probability of dying in infancy change between 1933 and 2023?

Answer: Here again we can either (i) use a long "or" statement or (ii) rely on the modulo function `%%` as before.

```
df |>
  filter((year-3)%%10==0 & age==0) |>
  select(year, q)
```

```
## # A tibble: 10 x 2
## # Groups:   year [10]
##     year       q
##    <dbl>   <dbl>
##  1  1933 0.0588
##  2  1943 0.0462
##  3  1953 0.0308
##  4  1963 0.0252
##  5  1973 0.0171
##  6  1983 0.0111
##  7  1993 0.00849
##  8  2003 0.00705
##  9  2013 0.00595
## 10  2023 0.00549
```

In 1933, a newborn had a probability of 5.88 percent of dying within the next year. By 2023, that probability had fallen to 0.55 percent.

Question 9

Change infant mortality in all years to the 1933 level. Recompute l, d, L, T, and e. Graph the life expectancy at birth over time. How much lower would life expectancy at birth have been in 2023 if infant mortality were at the 1933 level?

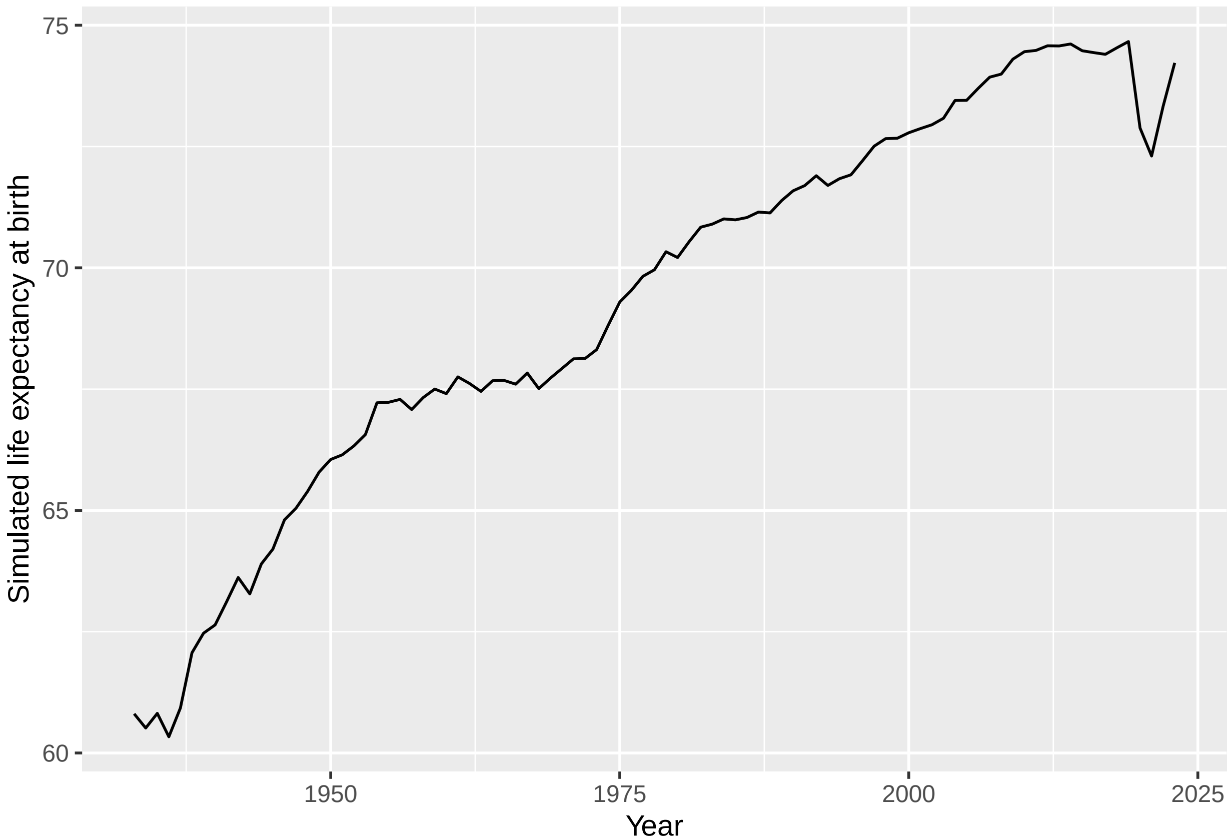
Answer: We start with an `if_else()` statement that reassigns `q` to be 0.0588 in every year. We then recompute the other life table functions and graph the results as before.

```
# change q0 to 0.0588 in all years
df <- df |> mutate(q = if_else(age==0, 0.0588, q))

# compute life table functions for every year
df <-
  df |>
  group_by(year) |>
  arrange(age) |>
  mutate(l = 100000 * cumprod(lag(1-q, default=1)),
         d = l*q,
         L = 0.5*d + (l-d),
         T = sum(L) - cumsum(L) + L,
         e = T/l)

# write over annual data frame with life expectancy at birth
annual_df <- df |> filter(age==0) |> select(year, age, e)

# graph life expectancy at birth over time
ggplot(annual_df, aes(x=year, y=e)) +
  geom_line() +
  scale_y_continuous("Simulated life expectancy at birth") +
  scale_x_continuous("Year")
```



We can see that life expectancy at birth in 2023 would have been about 4 years lower with the infant mortality rate at the 1933 level.

Question 10

You calculated *period* life expectancy above. How is it different from *cohort* life expectancy? Calculate cohort life expectancy at age 65 (e_{65}) for as many cohorts as possible. Plot e_{65} against year of birth. Describe your results.

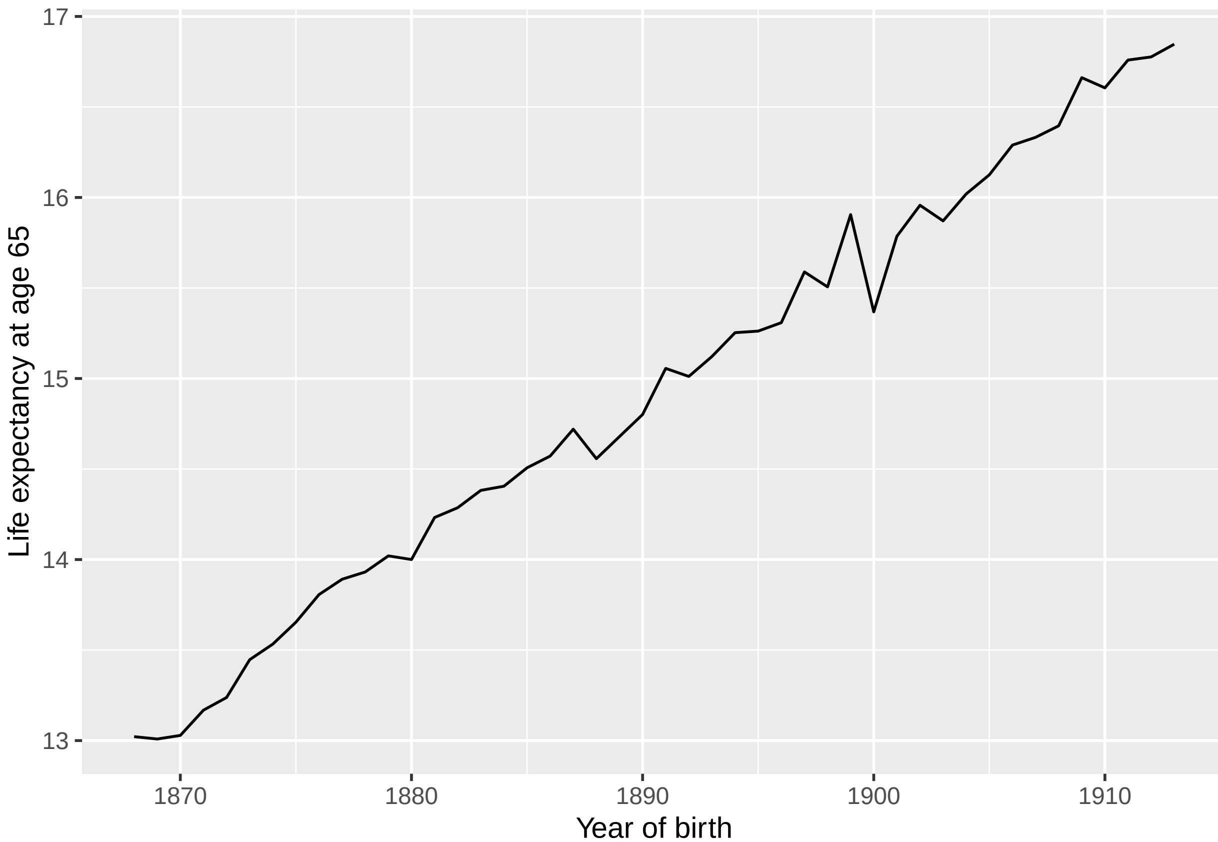
Answer: Period life expectancy e_x^o tells us how many more years a person aged x would live if they experienced *current* age-specific mortality rates for the rest of their lives. Cohort life expectancy e_x tells many more years a cohort lived past age x on average. We can approximate cohort life expectancy by moving diagonally through the dataset: for example, q_{65} in 1980, q_{66} in 1981, and so on. We just need to group by year of birth instead of by year.

```
# create new variable year of birth
df <- df |> mutate(birthyear = year-age)

# compute life table functions for every year of birth -
# this is different from before because we are grouping
# by year of birth instead of year
df <-
  df |>
  group_by(birthyear) |>
  arrange(age) |>
  mutate(l = 100000 * cumprod(lag(1-q, default=1)),
         d = l*q,
         L = 0.5*d + (l-d),
         T = sum(L) - cumsum(L) + L,
         e = T/l)

# keep age 65 and keep cohorts born at least 110 years before 2023
df65 <- df |> filter(age==65 & birthyear<=1913)

# draw graph
ggplot(df65, aes(x=birthyear, y=e)) +
  geom_line() +
  scale_y_continuous("Life expectancy at age 65") +
  scale_x_continuous("Year of birth")
```



Life expectancy at age 65 rose from 13 years among people born in 1870 to just over 16.5 years among people born in 1910.