HMD Analysis

"PSTAT 296A"

2025-11-01

```r
# import packages
library(gnm)
library(forecast)
library(StMoMo)
library(demography)
library(ggplot2)
library(tidyverse)
library(reshape2) # For data transformation
```

```r
# load data
US_data <- hmd.mx(country="USA", username="amandahuang@ucsb.edu", password="Ilovepstat296!")
US_data99 <- extract.years(US_data, 1999:2020)
```

Exploratory Analysis

Age Profile Over Time

```r
# extract age and year vector
Ages <- US_data$age
Years <- US_data$year

# extract mortality rate matrices
Rates_Male <- US_data$rate$male
Rates_Female <- US_data$rate$female
Rates_Total <- US_data$rate$total

# --- Plot 1: Total Mortality
plot(US_data, series="total", ylim = c(-10, 0), xlim = c(0,110))
legend("bottomright", legend=unique(Years),
  col=rainbow(length(Years)*1.25), ncol=8, pch=19,
  title="Year", cex=0.44)
```
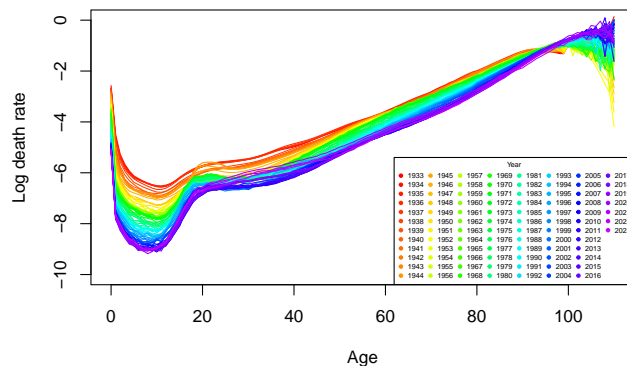
USA: total death rates (1933–2023)
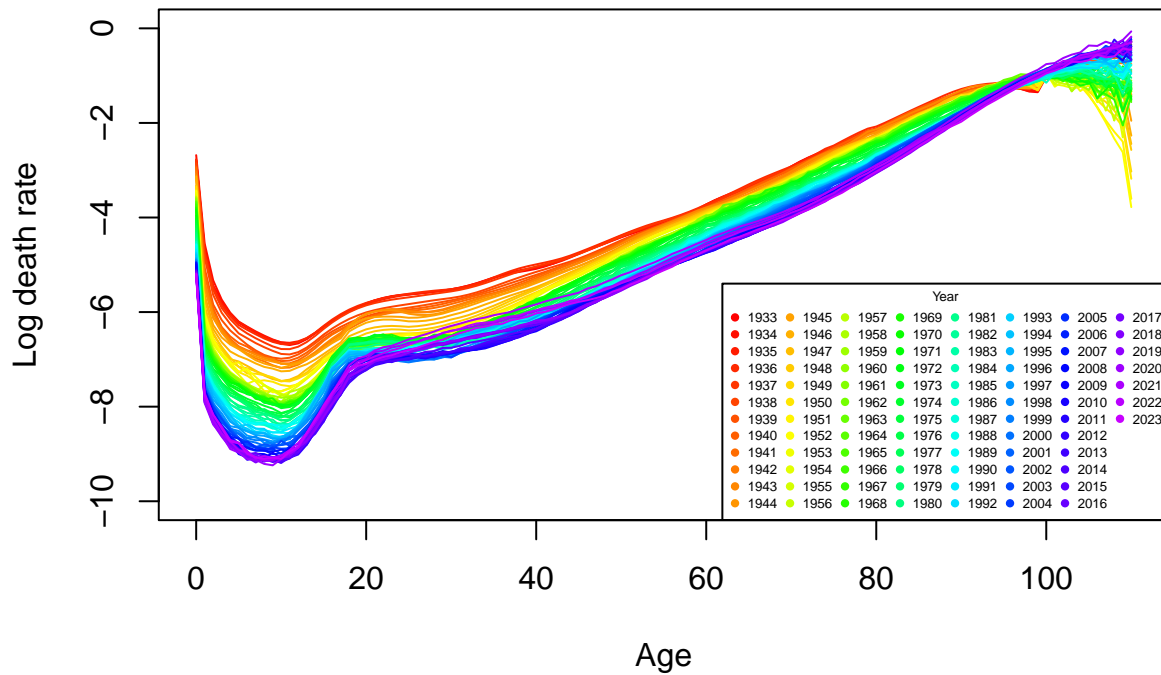

```
# --- Plot 2: Male Mortality
plot(US_data, series="male",  ylim = c(-10, 0))
legend("bottomright", legend=unique(Years),
  col=rainbow(length(Years)*1.25), ncol=8, pch=19,
  title="Year", cex=0.44)

# --- Plot 3: Female Mortality
plot(US_data, series = "female", ylim = c(-10, 0), xlim = c(0,120))
legend("bottomright", legend=unique(Years),
  col=rainbow(length(Years)*1.25), ncol=8, pch=19,
  title="Year", cex=0.44)
```

USA: male death rates (1933–2023)

USA: female death rates (1933–2023)
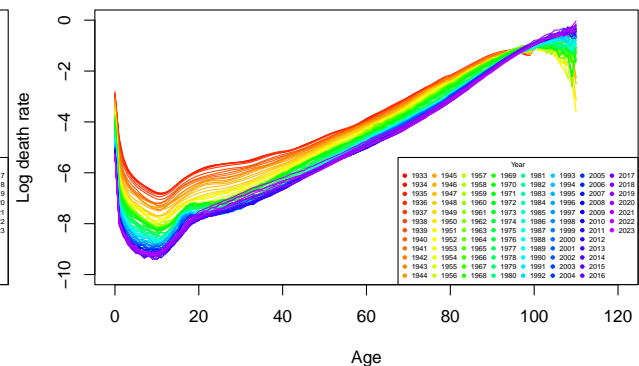
These plots display the age_specific log-mortality rate for every years from 1933 to 2023. Each line represents a single year, with the colors often used to denote the year (e.g., the warm colors for early years, cool colors for recent years.)

Mortality Curve Shape: All log-mortality curves show a J-shape: high mortality in infancy, started

declining to a lowest point around age 10 in young adulthood, followed by a continuous, near-exponential rise with age. In early years, some fluctuations at the older ages, reflecting data variability, probably due to lack of sample. **Mortality Improvement:** The mortality curve vertically shift downward over time, which represent mortality improvement (Period effect). **Gender Comparision:** The Female lines tend to be lower than the Male lines.

```r
# Define color
col_all <- rainbow(length(Years)*1.25)  # default colors
recent_years <- 2016:2023                # years to make black
col_all[Years %in% recent_years] <- "black"

# --- Total Mortality, 2016-2023
plot(US_data, series="total", ylim = c(-10, 0), xlim = c(0,120))
lines(US_data, series="total", ylim = c(-10, 0), year = 2016:2023, col = "black")
legend("bottomright", legend=unique(Years), col=col_all, ncol=8, pch=19,
  title="Year", cex=0.45)
```


USA: total death rates (1933–2023)
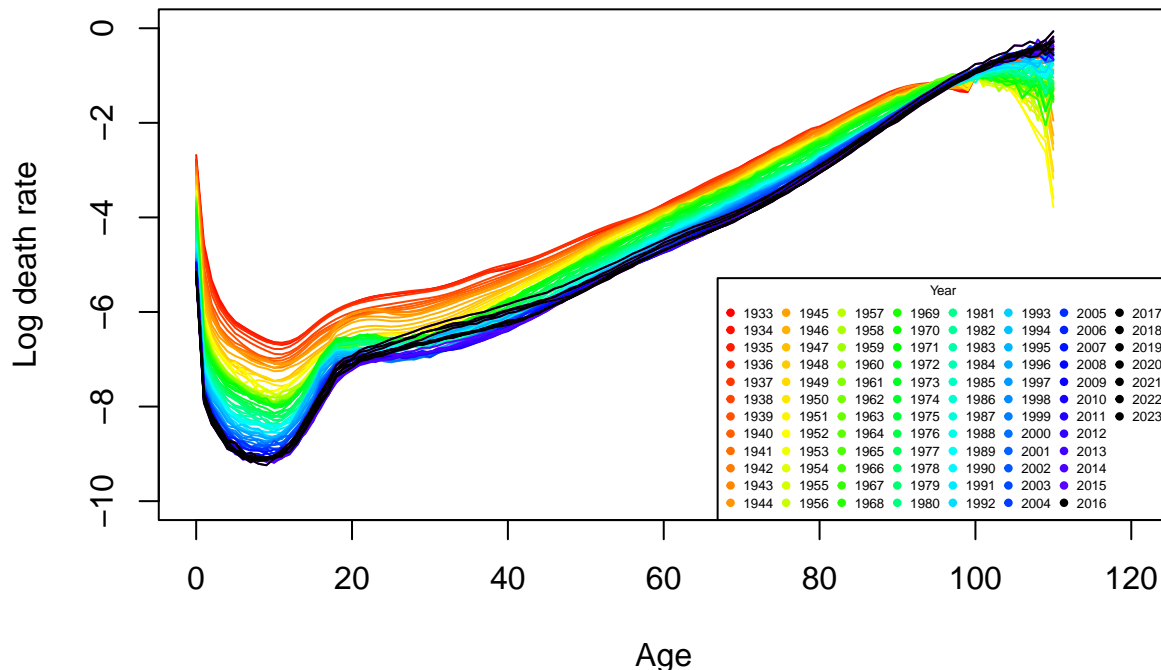
From years 2016-2023, log death rate fro age 20-40 increased compared to previous years.

Mean Mortality Profit

```r
# Cal mean log-death rate across all year for each age
# strat sex
Alpha_Male <- mean(US_data, series = names(US_data$rate)[2], transform = TRUE,
                na.rm = TRUE)$y
Alpha_Female <- mean(US_data, series = names(US_data$rate)[1], transform = TRUE,
                na.rm = TRUE)$y
Alpha_Total <- mean(US_data, series = names(US_data$rate)[3], transform = TRUE,
```

```
                            na.rm = TRUE)$y

alpha_df <- data.frame(Ages, Male = Alpha_Male, Female = Alpha_Female,
                       Total = Alpha_Total)
# alpha_df

# modify dataset for plotting
alpha_long <- alpha_df %>%
  pivot_longer(
    cols = c(Male, Female, Total),        # columns to pivot
    names_to = "Gender",                  # new variable name
    values_to = "Mean_LogRate"                # value column name
  )
# alpha_long

# plot mean log-death rate
ggplot(alpha_long, aes(x = Ages, y = Mean_LogRate, color = Gender)) +
  geom_line(linewidth = 1.2) +
  labs(
    title = "Figure 4. Mean Log-Mortality Rate by Age and Gender (a_x Profile)",
    x = "Age (x)",
    y = "Mean Log(Mortality Rate)"
  ) +
  theme_minimal()
```
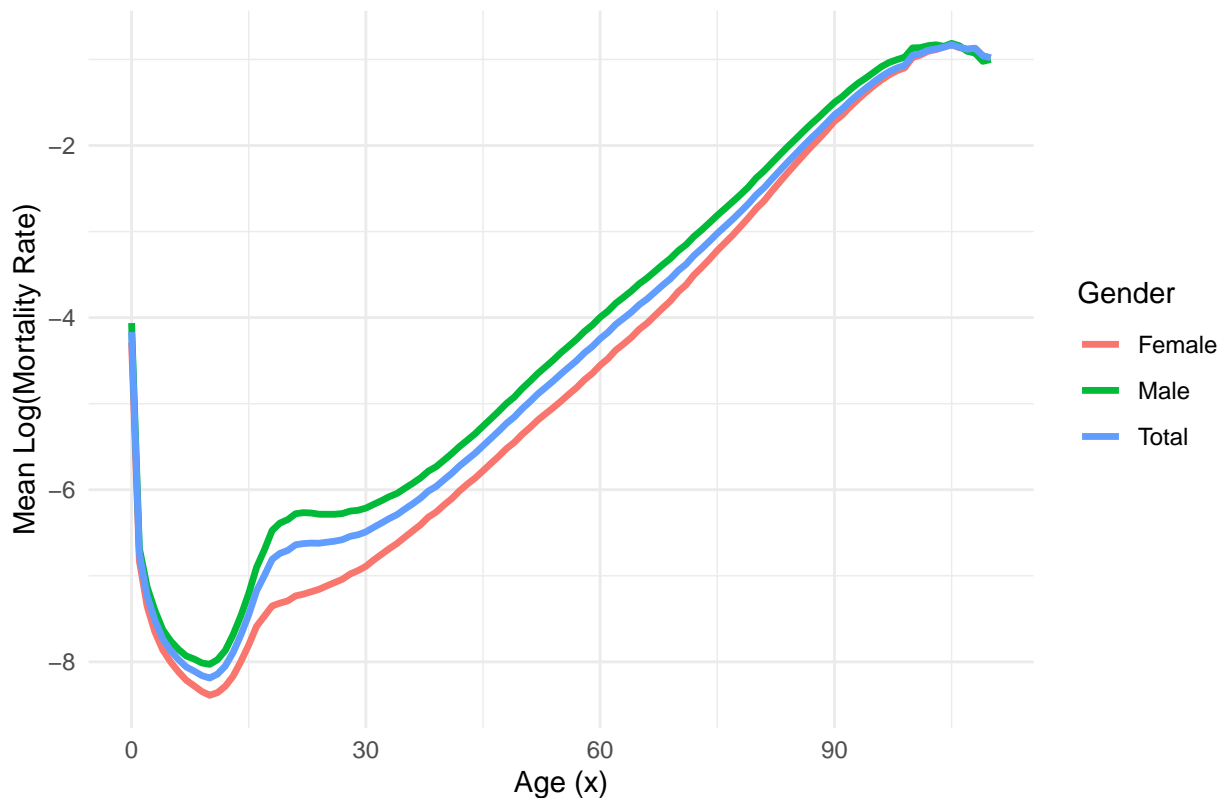
Figure 4. Mean Log–Mortality Rate by Age and Gender (a_x Profile)

Across nearly all ages, males exhibit higher mortality rate than females, with the gap most noticeable in young adulthood, reflecting behavioral and environmental factors such as greater risk exposure. The

4

difference narrows in later life when biological aging dominates mortality risk.

Change in Mortality Over Time

```
# change in mortality over years, age specific
plot(US_data, series = "total", plot.type = "time", xlab = "Year")

# Age group
plot(US_data, series = "total", plot.type = "time", xlab = "Year", age = 0:25,
     main = "USA: total death rates (1933-2023) Age: 0-24")
plot(US_data, series = "total", plot.type = "time", xlab = "Year", age = 26:40,
     main = "USA: total death rates (1933-2023) Age: 26-40")
plot(US_data, series = "total", plot.type = "time", xlab = "Year", age = 41:75,
     main = "USA: total death rates (1933-2023) Age: 41-75")
plot(US_data, series = "total", plot.type = "time", xlab = "Year", age = 76:100,
     main = "USA: total death rates (1933-2023) Age: 76-100")
plot(US_data, series = "total", plot.type = "time", xlab = "Year", age = 101:110,
     main = "USA: total death rates (1933-2023) Age: 101-110")
```

USA: total death rates (1933–2023) Age: 76–100 **USA: total death rates (1933–2023) Age: 101–110**
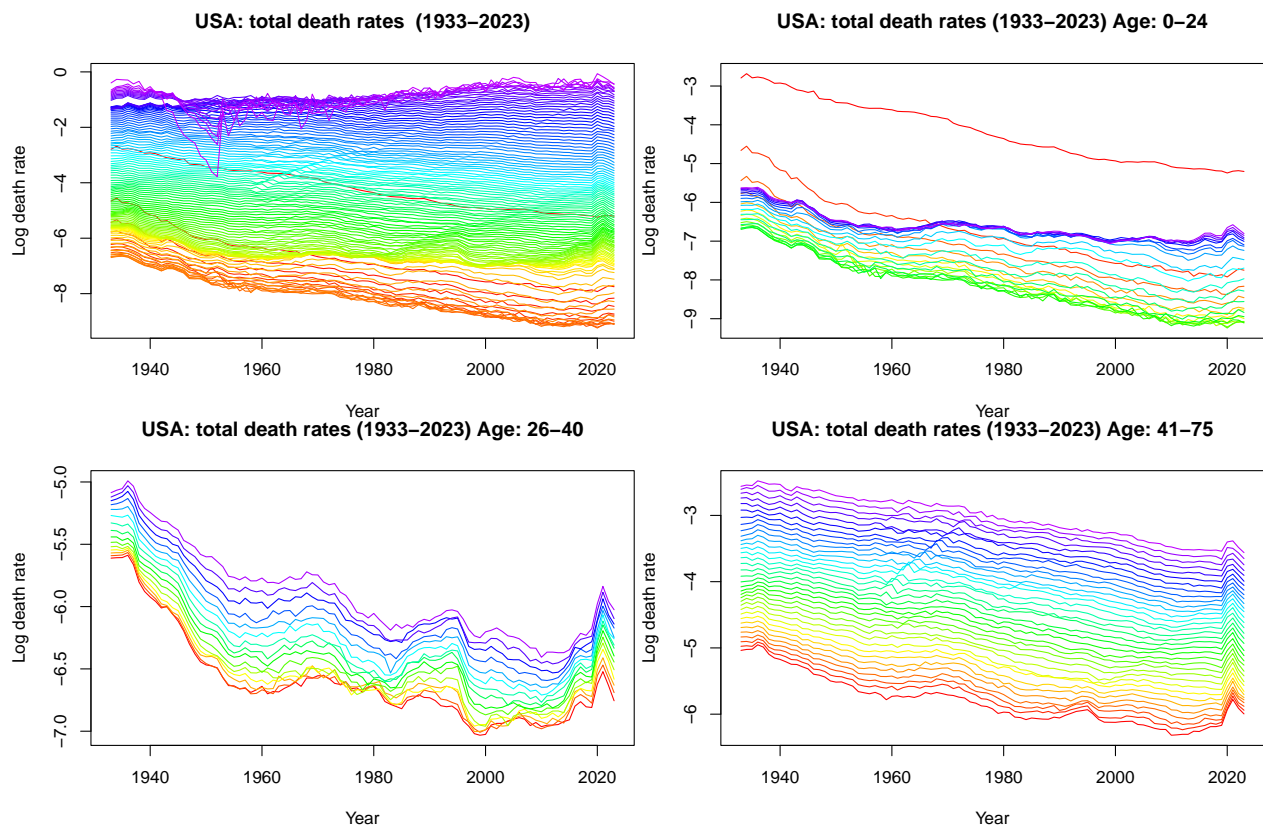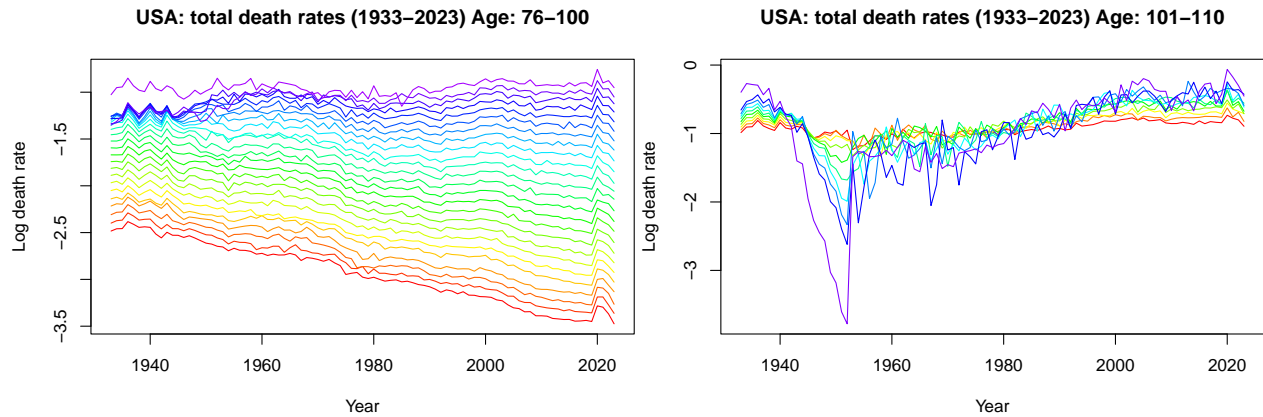
This graph displays the evolution of age-specific mortality over time, where each line represent an age group and shows how its log death rate has changed across year. And the warm color (red, orange) is younger ages and cool color (blue/ purple) represent older ages.

The overall pattern reveals a steady downward shift in death rate across nearly all ages, indicating continuous mortality improvement over years. This downward movement could potentially reflect the period effect. At year 2020 and after, there is a noticeable increase in death mortality across all ages.

From age 0-24, mortality rate have declined substantially across the entire period, particularly from 1930s-1960. From age 26-40, mortality rate has more fluctuation. Mortality rate show a notable decline from 1930s to 1970s, reaching a minimum around 1980. However, the curves rise again after 1990, particularly after 2010. This may correspond to behavioral or socioeconomic factors, like substance abuse.

From age 41-100, mortality rate shows a steady downward trend over time. From ages 101-110, the curves display considerable volatility probably due to limited data availability and sample size variation. The lack of a clear trend indicates that statistical uncertainty dominates at these extreme ages.

```r
# difference of log-mortality between gender
diff_data <- Rates_Male - Rates_Female

# plot delta(log-mort)
matplot(Years, t(diff_data), type = "l",
        col = adjustcolor(rainbow(nrow(diff_data)), alpha.f = 0.4),
        lwd = 0.8,
        ylab = "Male - Female (log mortality difference)",
        xlab = "Year",
        main = "Gender Difference in Log Mortality Rates (1933-2023)")
abline(h = 0, col = "black", lty = 2)  # reference line
```

6

Gender Difference in Log Mortality Rates (1933–2023)
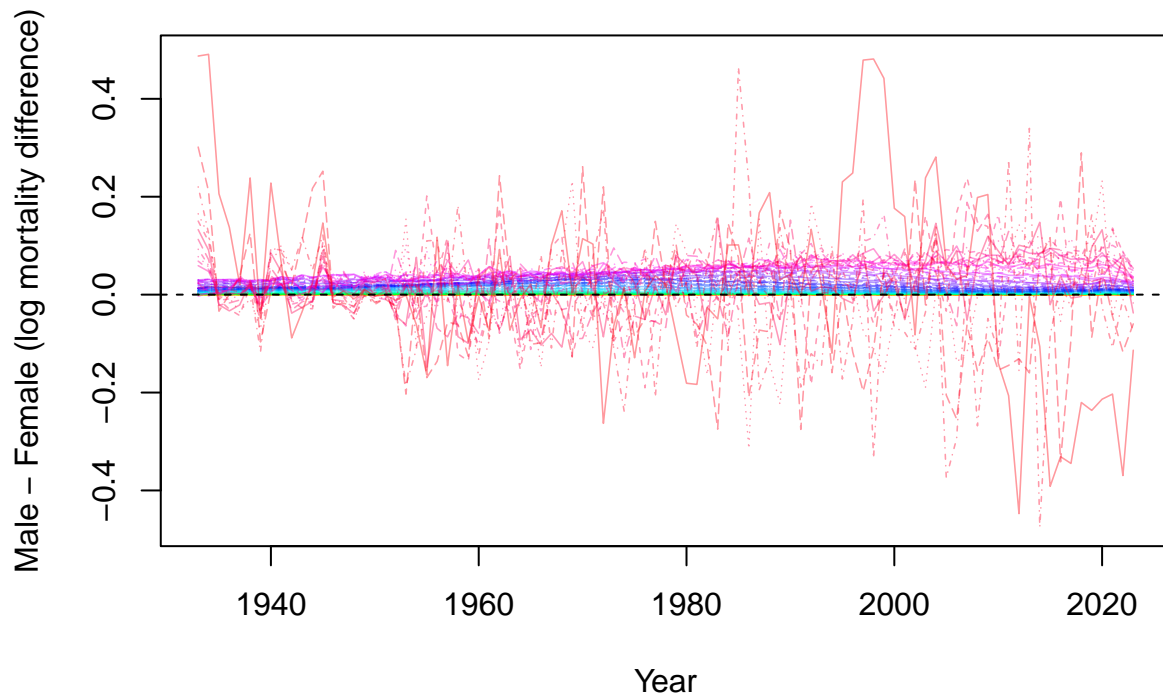
This graph shows the difference between male and female log mortality rate across all ages over time, where each line represents a age. The curves above zero indicate higher male mortality while below zero indicate higher female mortality.

Most of the curves locate above zero, meaning male mortality rates have consistently been higher than female mortality rate across almost all age groups and time periods. But during the early age, such as infant time and child time, the mortality curves fluctuate around zero, indicating no gender advantage on mortality rate during that time.

After 1970s, the gender gap appears wider and reach its maximum from 1980 to 2000. Previously, we discover that total mortality for age 26-40 increase from 1980 to 2000. Combined with the increase of gender gap from 1980 to 2000, we suspect that for age group 26-40, from 1980 to 2000, the males mortality increase or decrease at a lower rate than females while females mortality rate decrease or at a higher rate.

```r
# Define gradient colors for each gender
male_colors   <- colorRampPalette(c("darkblue", "skyblue"))(length(26:40))
female_colors <- colorRampPalette(c("darkred", "pink"))(length(26:40))

plot(US_data, series = "male", plot.type = "time", age = 26:40,
    col = male_colors, xlab = "Year", ylim = c(-8,-4.9),
     main = "USA: male vs female death rates age: 26-40")
lines(US_data, series = "female", plot.type = "time", age = 26:40,
       col = female_colors)
legend("topright",
        legend = c("Male (blue gradient)", "Female (red gradient)"),
        col = c("blue", "red"), lwd = 2, bty = "n")
```

USA: male vs female death rates age: 26–40
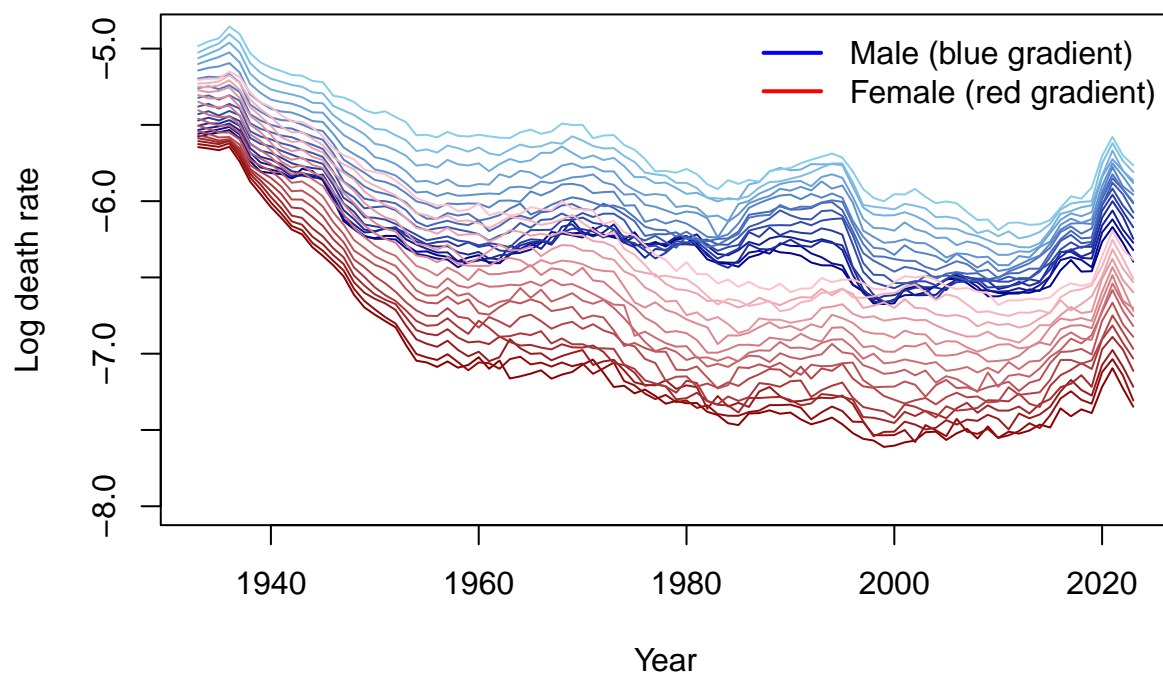
The blue gradient represents males, while the red gradient represents females. Between 1980 and 2000, male death rates stabilize and increase slightly while females rate continue to improve.