Life tables, Lee-Carter Modeling

2024-09-02

```
stopifnot(
  require(patchwork),
  require(httr),
  require(glue),
#  require(ineq),
  require(here),
  require(skimr),
  require(magrittr),
  require(plotly),
  require(tidyverse)
)

old_theme <- theme_set(theme_minimal())
```

- **M1 MIDS MA7BY020**
- **Université Paris Cité**
- Année 2024-2025
- Course Homepage

- Moodle

> **Objectives**

Data sources

Life data tables are downloaded from https://www.mortality.org.

See also https://www.lifetable.de.

If you install and load package https://cran.r-project.org/web/packages/demography/index.html, you will also find life data tables.

We investigate life tables describing countries from Western Europe (France, Great Britain –actually England and Wales–, Italy, the Netherlands, Spain, and Sweden) and the United States.

We load the one-year lifetables for female, male and whole population for the different countries.

```
life_table |>
  dplyr::mutate(Country = forcats::as_factor(Country)) |>
  dplyr::mutate(Country = forcats::fct_relevel(Country, "Spain", "Italy", "France", "England &
  dplyr::mutate(Gender = forcats::as_factor(Gender)) -> life_table

life_table |>
  dplyr::mutate(Area =forcats::fct_collapse(Country,
                                            SE = c("Spain", "Italy", "France"),
                                            NE = c("England & Wales", "Netherlands", "Sweden")
                                            USA="USA")) -> life_table
```

Check on http://www.mortality.org the meaning of the different columns:

Document Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle contains detailed information on the construction of Life Tables for France.

Two kinds of Life Tables can be distinguished: *Table du moment* which contain for each calendar year, the mortality risks at different ages for that very year; and *Tables de génération* which contain for a given birthyear, the mortality risks at which an individual born during that year has been exposed.

The life tables investigated in this homework are *Table du moment*. According to the document by Vallin and Meslé, building the life tables required ,decisions and doctoring.

See (among other things)

- p. 19 Abrupt changes in mortality quotients at some ages for a given calendar year
- Estimating mortality quotients at great age.

Have a look at Lexis diagram.

Definitions can be obtained from www.lifeexpectancy.org. We translate it into mathematical (rather than demographic) language. Recall that the quantities define a probability distribution over $\mathbb{N}$. This probability distribution is a *construction* that reflects the health situation in a population at a given time. This probability distribution does not describe the sequence of sanitary situations experienced by a *cohort* (people born during a specific year).

> One works with a period, or current, life table (*table du moment*). This summarizes the mortality experience of persons across all ages in a short period, typically one year or three years. More precisely, the death probabilities $q(x)$ for every age x are computed for that short period, often using census information gathered at regular intervals. These $q(x)$'s are then applied to a hypothetical cohort of 100000 people over their life span to produce a life table.

```
life_table |>
  filter(Country=='France', Year== 2010, Gender=='Female', Age < 10 | Age > 80)
```

```
# A tibble: 39 x 13
    Year   Age      mx      qx    ax     lx    dx    Lx      Tx    ex Country
   <int> <int>   <dbl>   <dbl> <dbl>  <int> <int> <int>   <int> <dbl> <fct>
 1  2010     0 0.00325 0.00324  0.14 100000   324 99722 8465207  84.6 France
```

```
 2  2010     1 0.00032 0.00032  0.5   99676    32 99660 8365484  83.9 France
 3  2010     2 0.00015 0.00015  0.5   99645    15 99637 8265824  83.0 France
 4  2010     3 0.00011 0.00011  0.5   99630    11 99624 8166187  82.0 France
 5  2010     4 0.00008 0.00008  0.5   99619     8 99615 8066563  81.0 France
 6  2010     5 0.00005 0.00005  0.5   99611     5 99608 7966948  80.0 France
 7  2010     6 0.00008 0.00008  0.5   99606     8 99602 7867339  79.0 France
 8  2010     7 0.00008 0.00008  0.5   99598     8 99594 7767737  78.0 France
 9  2010     8 0.00008 0.00008  0.5   99590     8 99586 7668143  77   France
10  2010     9 0.00007 0.00007  0.5   99582     7 99578 7568557  76   France
# i 29 more rows
# i 2 more variables: Gender <fct>, Area <fct>
```

In the sequel, we denote by F_t the *cumulative distribution function* for year t. We agree on $\overline{F}_t = 1 - F_t$ and $F_t(-1) = 0$.

```
life_table |>
  filter( Year>=1948) |>
  group_by(Country, Year, Gender) |>
  summarise(m1 =max(abs(lx -dx -lead(lx)), na.rm = T),
            m2 =max(abs(lx * qx -dx), na.rm=T),
            m3 =max(abs(Lx -lx * (1 + qx * (ax-1))), na.rm=T),
            m4 =max(abs(1-exp(-mx)-qx), na.rm=T)) |>
  select(Year, Country, Gender, m1, m2, m3, m4) |>
  ungroup() |>
  group_by(Country, Gender) |>
  slice_max(order_by = desc(m4), n = 1)
```

```
# A tibble: 21 x 7
# Groups:   Country, Gender [21]
    Year Country         Gender    m1    m2    m3      m4
   <int> <fct>           <fct>  <int> <dbl> <dbl>   <dbl>
 1  1948 Spain           Both       1 0.874 2.20  0.00838
 2  1948 Spain           Female     1 0.789 1.56  0.00816
 3  1952 Spain           Male       1 0.802 5.5   0.0119
 4  2004 Italy           Both       1 0.836 0.968 0.0150
 5  2004 Italy           Female     1 0.875 1.03  0.0149
 6  1984 Italy           Male       1 0.774 5.56  0.0146
 7  2007 France          Both       1 0.887 0.976 0.0152
 8  2007 France          Female     1 0.890 0.980 0.0151
 9  1979 France          Male       1 0.764 4.97  0.0161
10  1992 England & Wales Both       1 0.898 2.42  0.0135
# i 11 more rows
```

qx (age-specific) risk of death at age x, or mortality quotient at given age x for given year t: $q_{t,x} = \frac{\overline{F}_t(x) - \overline{F}_t(x+1)}{\overline{F}_t(x)}$.

For each year, each age, $q_{t,x}$ is determined by data. We also have

$$\overline{F}_t(x+1) = \overline{F}_t(x) \times (1 - q_{t,x+1}) .$$

`mx` *central death rate* at age x during year t. This is connected with $q_{t,x}$ by

$$m_{t,x} = -\log(1 - q_{t,x})\,,$$

or equivalently $q_{t,x} = 1 - \exp(-m_{t,x})$.

`lx` the so-called *survival function*: the scaled proportion of persons alive at age x. These values are computed recursively from the $q_{t,x}$ values using the formula

$$l_t(x+1) = l_t(x) \times (1 - q_{t,x})\,,$$

with $l_{t,0}$, the "radix" of the table, arbitrarily set to 100000. Function $l_{t,\cdot}$ and $\overline{F}_t$ are connected by

$$l_{t,x+1} = l_{t,0} \times \overline{F}_t(x)\,.$$

Note that in Probability theory, $\overline{F}$ is also called the survival or tail function.

`dx` $d_{t,x} = q_{t,x} \times l_{t,x}$

`Tx` Total number of person-years lived by the cohort from age x to $x+1$. This is the sum of the years lived by the $l_{t,x+1}$ persons who survive the interval, and the $d_{t,x}$ persons who die during the interval. The former contribute exactly 1 year each, while the latter contribute, on average, approximately half a year, so that $L_{t,x} = l_{t,x+1} + 0.5 \times d_{t,x}$. This approximation assumes that deaths occur, on average, half way in the age interval x to x+1. Such is satisfactory except at age 0 and the oldest age, where other approximations are often used; *We will stick to a simplified vision* $L_{t,x} = l_{t,x+1}$

`ex:` Residual Life Expectancy at age x and year t

Loading `life_table` onto an in memory database

We load `life_table` into an in memory database, unleashing the full power of SQL. This is helpful if we have to use window functions.

```
<SQL>
SELECT `dbplyr_mosaxLTcTe`.*
FROM `dbplyr_mosaxLTcTe`
WHERE (`Gender` = 'Female') AND (`Country` = 'USA') AND ('Year' = 1948.0)
```

Object `lt` can be queried like any other data frame.

```
con <- DBI::dbConnect(RSQLite::SQLite(), ":memory:")
src <- dbplyr::src_dbi(con, auto_disconnect = TRUE)

dplyr::copy_to(src, lt)
```

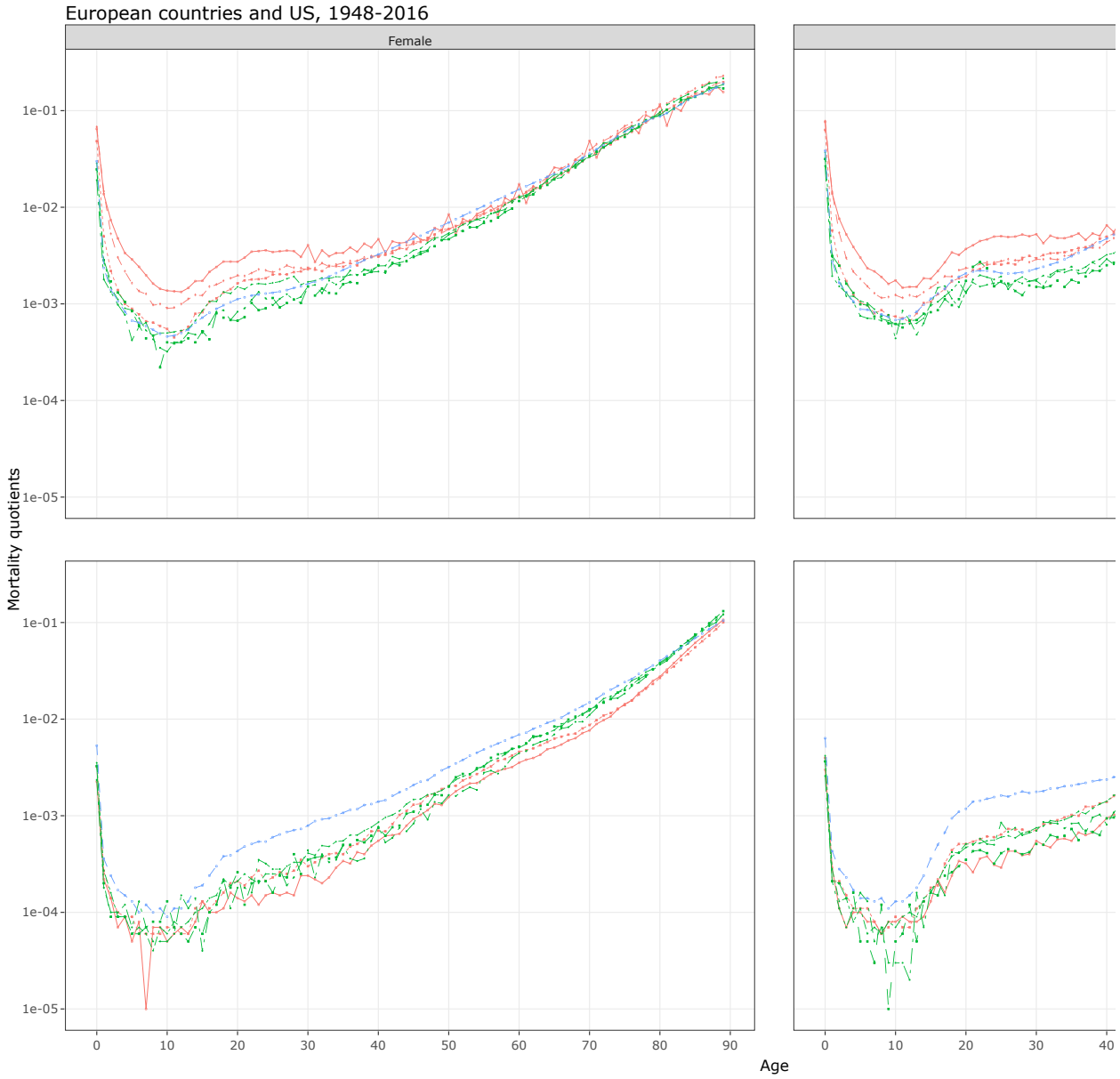
Computing residual life expectancies at all ages can also be completed using SQL queries.

Western countries in 1948

Several pictures share a common canvas: we plot central death rates against ages using a logarithmic scale on the y axis. Countries are identified by aesthetics (shape, color, linetypes). Abiding to the DRY principle, we define a prototype `ggplot` (alternatively `plotly`) object. The prototype will be fed with different datasets and decorated and arranged for the different figures.

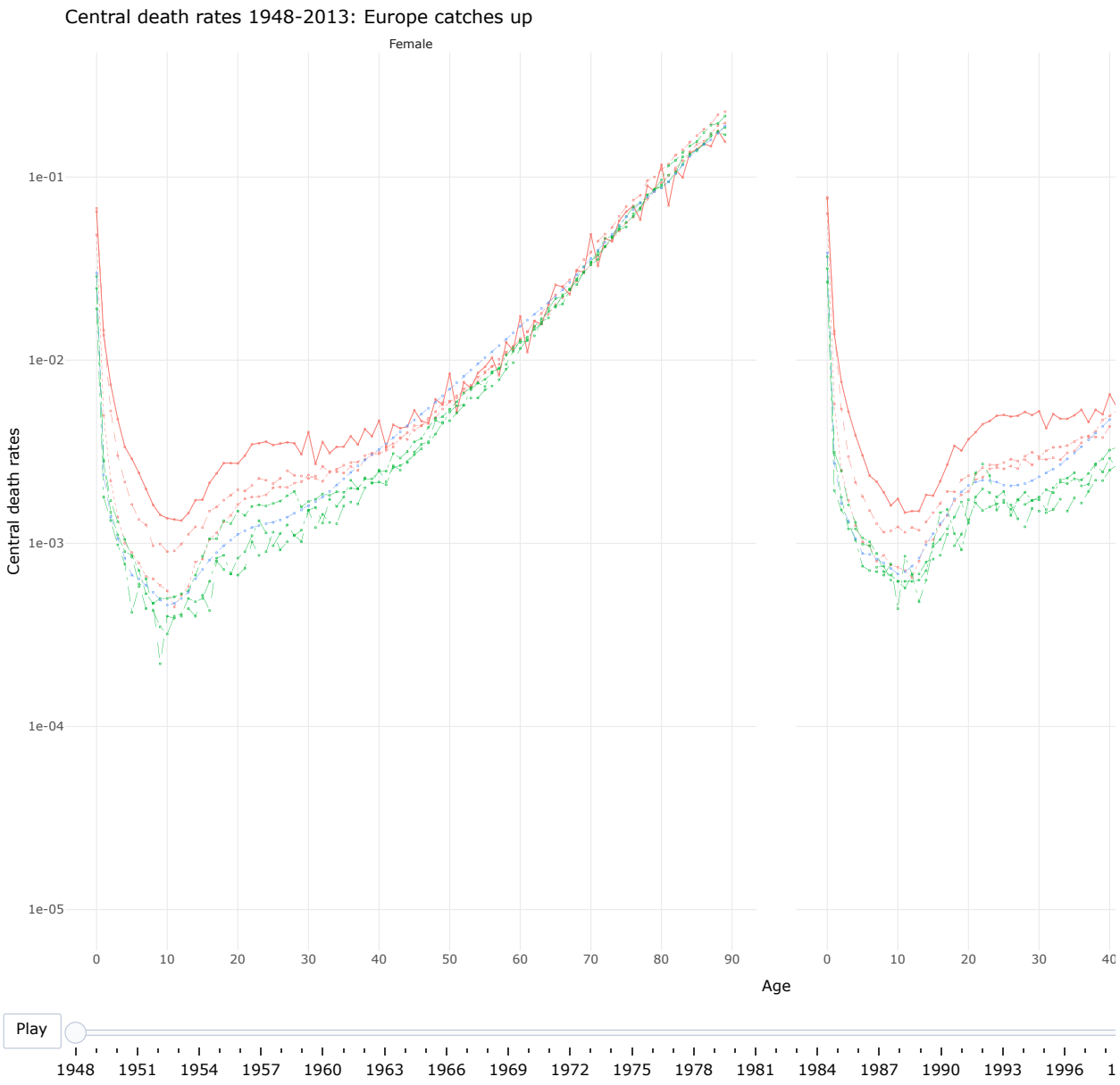
```r
dummy_data <- dplyr::filter(life_table, FALSE)

proto_plot <- ggplot(dummy_data,
                     aes(x=Age,
                         y=qx,
                         col=Area,
                         linetype=Country,
                         shape=Country)) +
              scale_y_log10() +
              scale_x_continuous(breaks = c(seq(0, 100, 10), 109)) +
              ylab("Mortality quotients") +
              labs(linetype="Country") +
              theme_bw()
```

- Plot `qx` of all Countries at all ages for years 1948 and 2013.




```
proto_plt2 <-
  ggplot() +
  aes(x=Age, y=qx, colour=Area, frame=Year, linetype=Country) +
  geom_point(size=.1) +
  geom_line(size=.1) +
  scale_y_log10() +
  labs(linetype=c("Country")) +
  scale_x_continuous(breaks = c(seq(0, 100, 10), 109)) +
  xlab("Age") +
  ylab("Central death rates") +
  facet_grid(cols=vars(Gender))
```

```
with(params,
  (proto_plt2 %+%
    (life_table |>
      filter(between(Year, year_p, year_e),
             Gender != 'Both',
             Age < 90))  +
    ggtitle("Central death rates 1948-2013: Europe catches up"))) |>
  plotly::ggplotly()
```



The animated plot allows to spot more details. It is useful to use color so as to distinguish threee areas: USA; Northern Europe (NE) comprising England and Wales, the Netherlands, and Sweden; Southern Europe (SE) comprising Spain, Italy, and France. In 1948, NE and the USA exhibit comparable central death reates at all ages for the two genders, the USA looking like a more dangerous place for young adults. Spain lags behind, Italy and Frane showing up at intermediate positions.

By year 1962, SE has almost caught up the USA. Italy and Spain still have higher infant mortality while central death rates in the USA and France are almost identical at all ages for both genders. Central death rates attain a minimum around 10-12 for both genders. In Spain the minium central death rate has been divided by almost ten between 1948 and 1962.

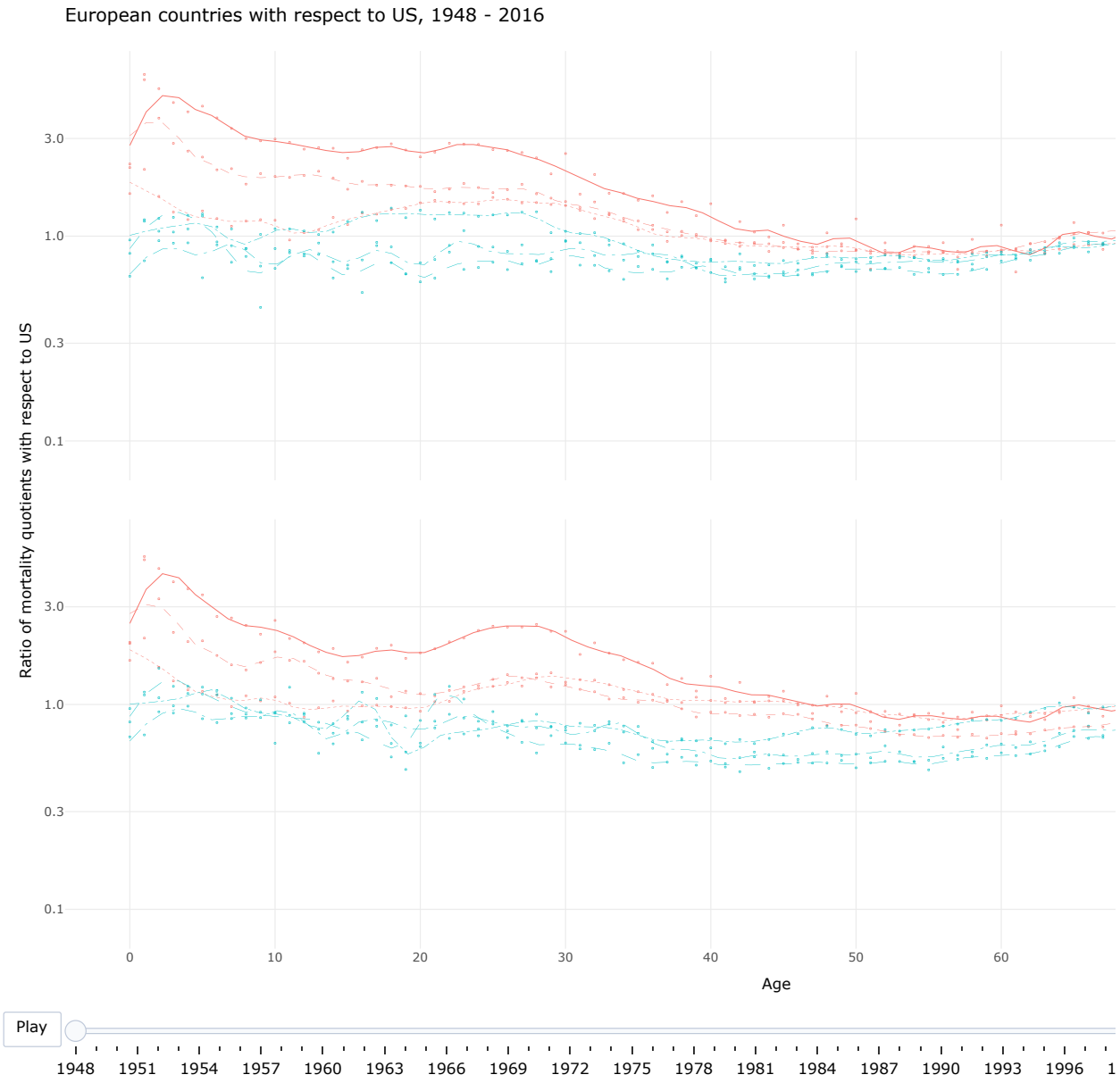
If we dig further we observe that the shape of the male central death rates curve changes over time. In 1962, in the USA and France, central death rates exhibit a sharp increase between years 12 and 18, then remain almost constant between 20 and 30 and afterwards increase again. This pattern shows up in other countries but in a less spectacular way.

Twenty years afterwards, during years 1980-1985, death rates at age 0 have decreased at around 1% in all countries while it was 7% in Spain in 1948. The male central death curve exhibits a plateau between ages 20 and 30. Central death rates at this age look higher in France and the USA.

By year 2000, France is back amongst European countries (at least with respect to central death rates). Young adult mortality rates are higher in the USA than in Europe. This phenomenon became more pregnant during the last decade.

Plot ratios between central death rates (`qx`) in European countries and central death rates in the USA in 1948.

```
with(params,
(eur_us_table  |>
  ggplot(aes(x=Age,
             y=Ratio,
             col=Area,
             frame=Year,
             linetype=Country)) +
  scale_y_log10() +
  scale_x_continuous(breaks = c(seq(0, 100, 10), 109)) +
  geom_point(size=.1) +
  geom_smooth(method="loess", se=FALSE, span=.1, size=.1) +
  ylab("Ratio of mortality quotients with respect to US") +
  labs(linetype="Country", color="Area") +
  # scale_colour_brewer(direction=-1) +
  ggtitle(label = stringr::str_c("European countries with respect to US,", year_p,'-', year_e,
  facet_grid(rows = vars(Gender)))) |>
  ggplotly()
```

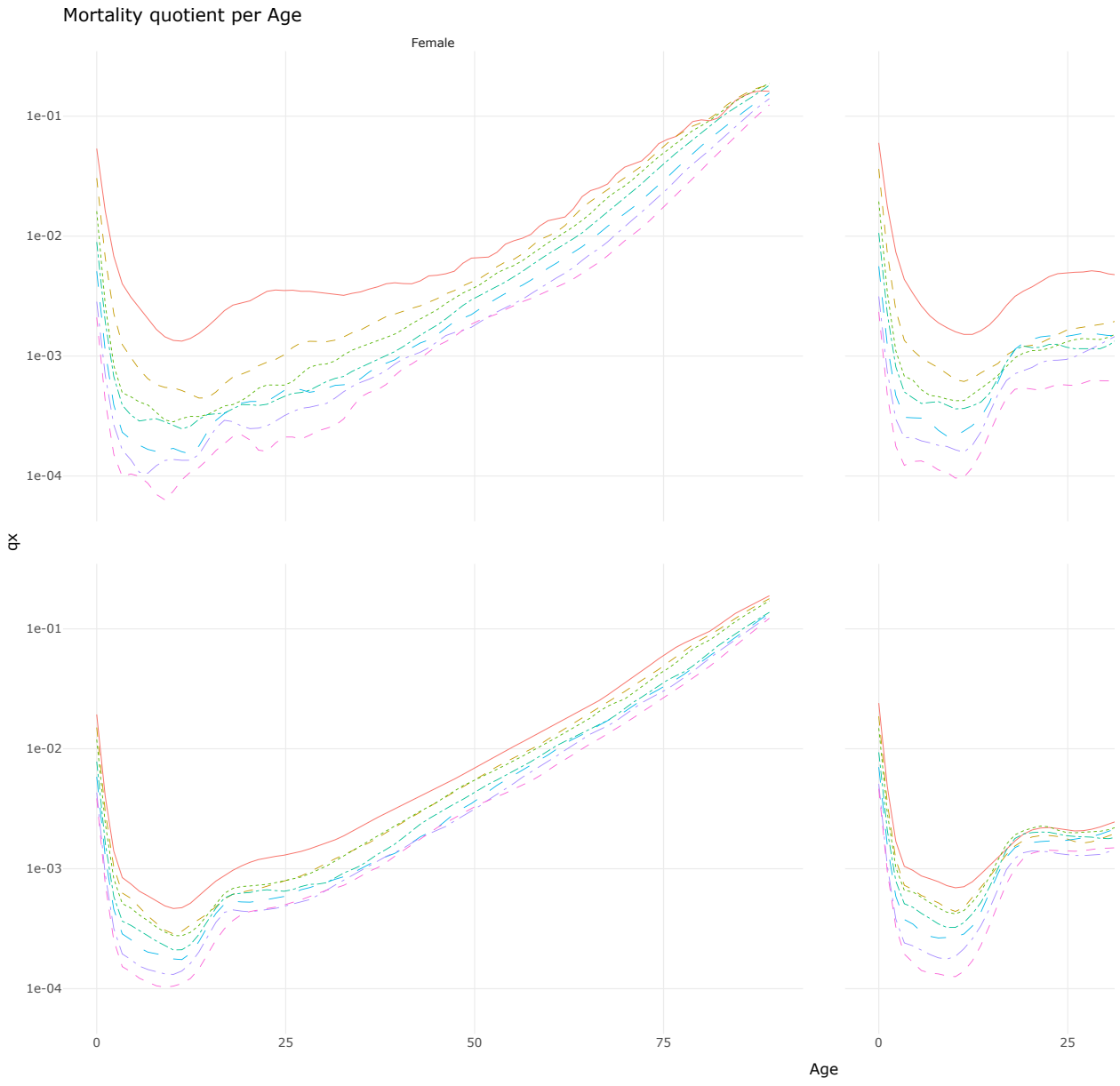


☒ Comment. This animation reveals less than the preceding one since we just have ratios with respect to the USA. But the patterns followed by European societies emerge in a more transparent way. The divide between northern and southern Europe at the onset of the period is even more visible. The ratios are important across the continent: there is a factor of 10 between spanish and swedish infant mortality rates. But the ratios at ages 50 and above tend to be similar. By the early 60s, the gap between southern and northern Europe has shrinked. By now, the ratios between central death rates tend to be within a factor of 2 across all ages, and even less at ages 50 and above.

Death rates evolution since WW II

☒ Plot mortality quotients (column `mx`) for both genders as a function of `Age` for years `1946, 1956, ...` up to `2016` . Use aesthetics to distinguish years. You will need to categorize the `Year` column (`forcats::` may be helpful).

1. Facet by `Gender` and `Country`
2. Pay attention to axes labels, to legends. Assess logarithmic scales.



☒ Write a function `ratio_mortality_rates` with signature `function(df, reference_year=1946, target_years=seq(1946, 2016, 10))` that takes as input:

- a dataframe with the same schema as `life_table`,
- a reference year `ref_year` and

- a sequence of years `target_years`

and that returns a dataframe with schema:

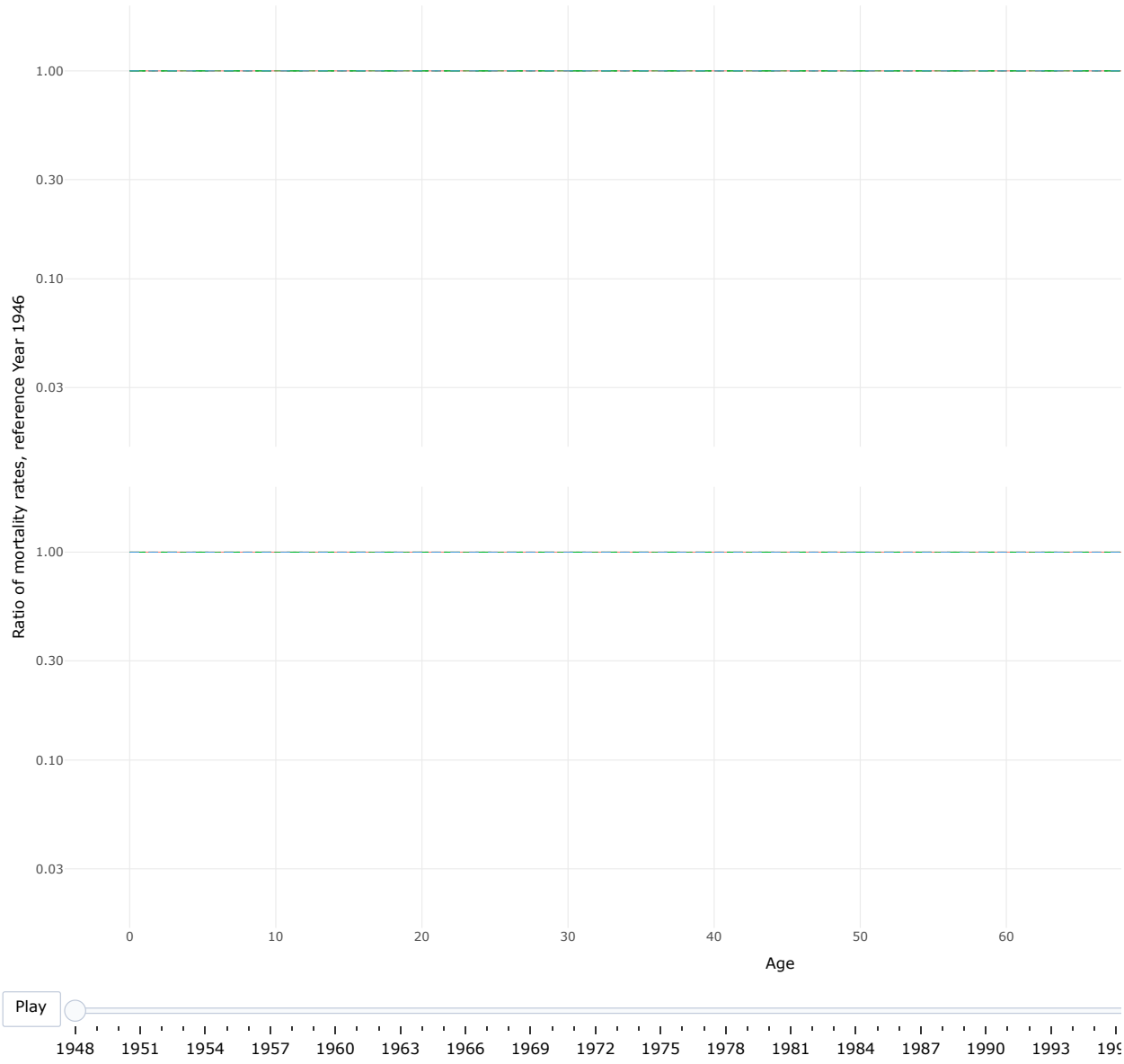
Column Name	Column Type
Year	integer
Age	integer
mx	double
mx.ref_year	double
Country	factor
Gender	factor

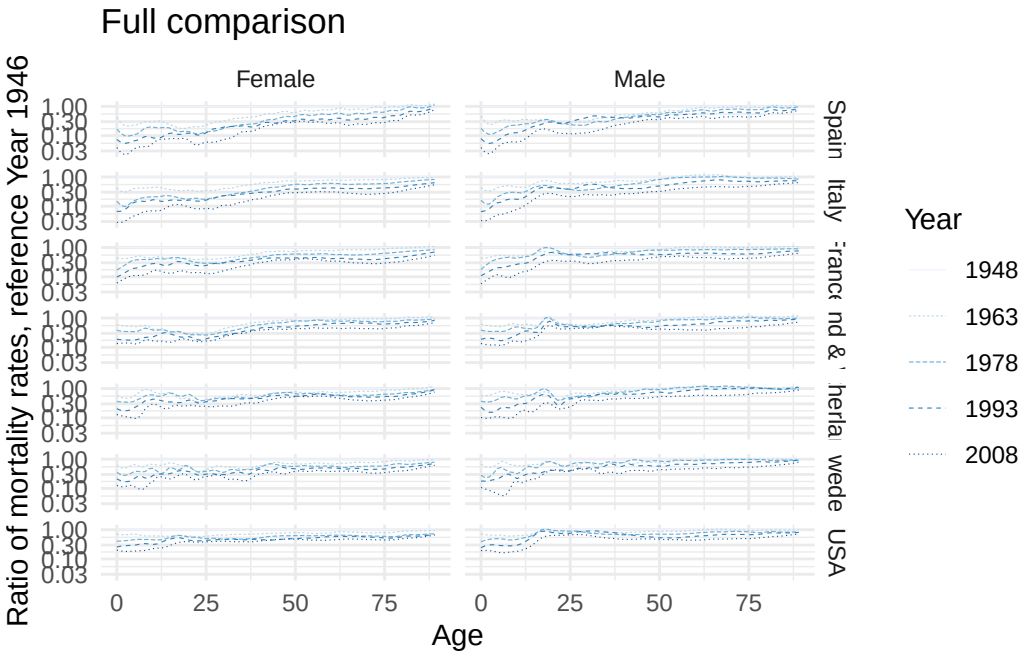
where `(Country, Year, Age, Gender)` serves as a *primary key*, `mx` denotes the central death rate at `Age` for `Year` and `Gender` in `Country` whereas `mx_ref_year` denotes central death rate at `Age` for argument `reference_year` in `Country` for `Gender`.

```
ratio_mortality_rates <- function(df,
                                  reference_year=1946,
                                  target_years=seq(1946, 2016, 10)){
  dplyr::filter(df, Year %in% target_years, Age <90) |>
  dplyr::select("Age", "Area", "Gender", "Country", "qx", "Year") |>
  dplyr::inner_join(y=df[df$Year==reference_year,
                         c("Age", "Gender", "Country", "qx")],
                    by=c("Age", "Gender", "Country"))
}
```

☒ Draw plots displaying the ratio $m_{x,t}/m_{x,1946}$ for ages $x \in 1, \dots, 90$ and year t for $t \in 1946, \dots, 2016$ where $m_{x,t}$ is the central death rate at age x during year t.

1. Handle both genders and countries `Spain`, `Italy`, `France`, `England & Wales`, `USA`, `Sweden`, `Netherlands`.
2. One properly facetted plot is enough.


1.00
0.30
0.10
0.03
1.00
0.30
0.10
0.03
Ratio of mortality rates, reference Year 1946
0
10
20
30
40
50
60
Age
Play
1948
1951
1954
1957
1960
1963
1966
1969
1972
1975
1978
1981
1984
1987
1990
1993

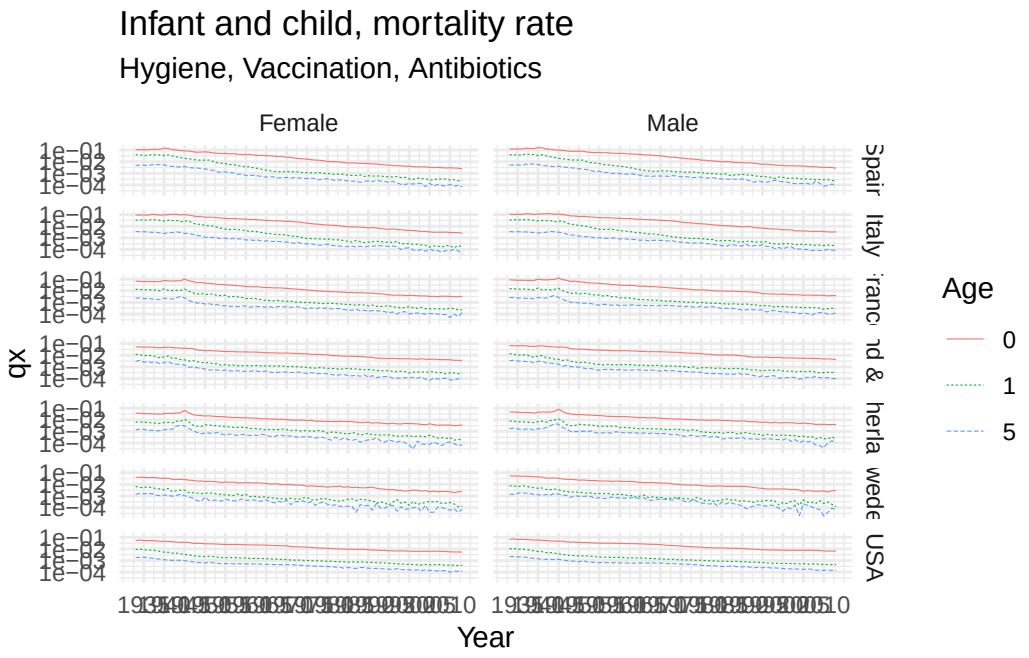


☒ Comment. During the last seventy years, death rates decreased at all ages in all seven countries. This progress has not been uniform across ages, genders and countries. Across most countries, infant mortality dramatically improved during the first post-war decade while death rates at age 50 and above remained stable until the mid seventies.

Trends

We noticed that central death rates did not evolve in the same way across all ages: first, the decay has been much more significant at low ages; second, the decay of central death rates at old ages (above 60) mostly took place during the last four decades. It is worth digging separately at what happened for different parts of life.

☒ Plot mortality quotients at ages $0, 1, 5$ as a function of time. Facet by Gender and Country

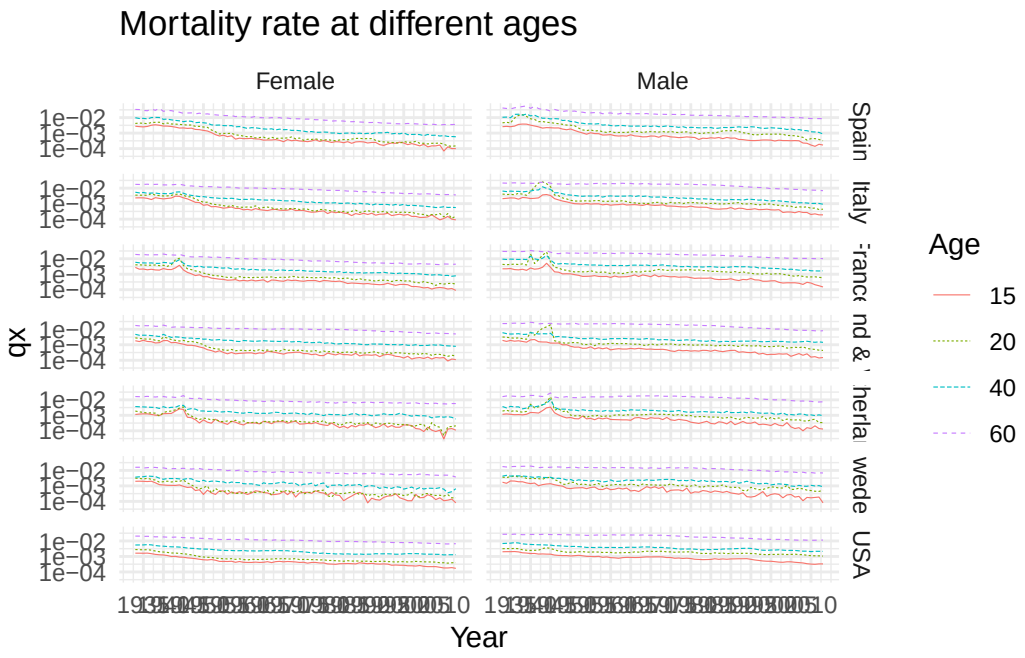


☒ Comment. All European countries achieved the same infant mortality rates after year 2000. The USA now lag behind.

During years 1940-1945, in the Netherlands and France, gains obtained before 1940 were reversed. Year 1945 was particularly difficult in the Netherlands.

☒ Plot mortality quotients at ages $15, 20, 40, 60$ as a function of time. Facet by `Gender` and `Country`

```
ages <- c(15, 20, 40, 60)

p_children %+%
  filter(life_table,
         Age %in% ages,
         Gender != "Both",
         Year %in% 1933:2013) +
  ggtitle("Mortality rate at different ages")
```



☒ Comment. While death rates at ages 15 and 20 among women are close across all societies, death rates are higher at age 20 than at age 15 among men. In France, at age 20, death rates declined from 1945 until 1960, and then increased back to their initial level until 1980. Male death rates at age 60 started to decline around 1980. Female death rates at age 60 declined steadily throughout the 7 decades. Years 1940-1945 exhibit disruptions with different shapes and intensities in Italy, France, England & Wales, and the Netherlands.

Rearrangement

☒ From dataframe `life_table`, compute another dataframe called `life_table_pivot` with primary key `Country`, `Gender` and `Year`, with a column for each `Age` from `0` up to `110`. For each age column, the entry should be the central death rate at the age defined by column, for `Country`, `Gender` and `Year` identifying the row.

You may use functions `pivot_wider`, `pivot_longer` from `tidyr::` package.

The resulting schema should look like:

Column Name	Type
Country	factor
Gender	factor
Year	integer
0	double
1	double
2	double
3	double
⋮	⋮

☒ Using `life_table_pivot` compute life expectancy at birth for each Country, Gender and Year

PCA and SVD over log-mortality tables

☒ facet screeplots for gender and countries
☒ comment the screeplots
☐ comment the correlation circles
☐ comment the biplots

In the next chunks we compute the PCA (with standardization and centering) for all Countries and Genders in the database. The `dplyr` pipeline prepares grouped tibble corresponding to the different countries and genders. The collection is fed to `group_map` which attempts to compute pca for each group. This output is a list of (key, value) pairs.

Henceforth, we exclude data from Sweden since for several years after 2000, mortality risks at some ages between 5 and 10 are zero. I do not know how reliable this information is. It might just be the outcome of random fluctuations: a modern Swedish generation typically comprises 110000 children. Data obtained from larger (and possibly less healthy) European countries suggest that mortality risks between ages 5 and 10 are less than 5×10^{-5}. This suggests that the yearly number of girls or boys dying ate age x between 5 and 10 in Sweeden is dominated by a Poisson distribution with parameter 3. The probability of outcome 0 under Poisson distribution with parameter 3 is approximately 0.05.

```r
wg_lt <- wide_grouped_life_table(life_table)

group_pca <- function(df, scale_or_not=TRUE){
  df |>
    tibble::column_to_rownames("Year") |>
    dudi.pca(scannf = FALSE, nf=5, center = TRUE, scale = scale_or_not)
}

pcas_by_country_gender <- dplyr::group_map(wg_lt,
                  .f = ~ list(key=.y , value=try(group_pca(.x, TRUE))))
```

```
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
```

```
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
```

```r
pcas_by_country_gender_ns <- dplyr::group_map(wg_lt,
                 .f = ~ list(key=.y , value=try(group_pca(.x, FALSE))))
```

```
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
Error in dudi.pca(tibble::column_to_rownames(df, "Year"), scannf = FALSE,  :
  could not find function "dudi.pca"
```

Screeplots

The screeplots for standardized-centered PCAs for countries and genders shows

- the inertia projected on the first two principal components accounts for no less than 90% of total inertia.
- for each country the inertia projected on the first two principal components is larger for Females than for Males

- The approximation by rank two matrices is worst for the Dutch matrices

The stacked screeplots obtained from centred but non-standardized arrays tell a different (but obviously related) story. Recall that the total inertia of the six different arrays are different. But for all six countries and two genders, most of the inertia is captured by the first two axes.

Screeplots from PCA without standardization tell us in advance how well Lee-Carter approximation will perform for a given population. The larger the share of inertia carried by the first component, the better the Lee-Carter approximation. From the figure above, we can infer that Lee-Carter approximation will be better for Female populations and than for Male populations.

Correlation circles

For a given PCA, the correlation circle tells us how close the different variables are from the plane spanned by the first two components and where the projection of the variables lay on this plan. This can be the basis of some variables clustering procedures.

The correlation circle delivered by `ade4` is not readable: - all variables are aligned with the first component as could have been predicted from the screeplot - labels overlapping hides a possible structure

Plotting the correlation circle is just making another scatter plot. Indeed, attribute `co` of the object returned by `dudi.pca` is a data frame which rows correspond to the variables of the original data frame (so to ages).

For UK, France and Italy, as could have been guessed from screeplots, all centered and standardized variables are highly correlated with the first component. This is even more striking for women than for men. For the Netherlands, Spain and the US, especially for men, variables associated with old ages are substantially correlated with the second component.

A common pattern emerges across countries and genders. Recall that PCA (and SVD) are uniquely defined up to signs of eigenvectors (singular vectors). The following description is taken from the correlation circle for women in France.

- Age groups `60-79` and `80-99` are packed in the same region. The two age groups are almost aligned along parallel curves and ranked according to age along each curve. The two curves look oriented into opposite directions
- Age group `40-59` is also packed in a small region and almost ordered along the second component
- Age group `20-39` forms a SE-NW oriented cloud, points are ordered according to ages
- Age group `0-19` spans the whole plot. Ages 0-7 are ordered around the second component. Ages `8-15` tend to cluster in the same region. Age `18` almost look like an outlier. Ages `16-17` are at an intermediate position between the outlier `18` and the `10-15` group.

Looking at the variables on the plane generated by the second and third principal components is also interesting.

- Ages below 20 lie around a SW-NE axis
- Ages between 20 and 40 lie on the south-west
- Ages between 40 and 60 are clustered and almost aligned on the sout-east
- Ages above 60 are almost aligned along a line

- Ages above 80 are almost aligned along another line on the north east quarter

If we plot the rows (years) on the plane spanned by the first two principal axes, patterns across genders and countries again look very similar. Axis 1 roughly corresponds to a time axis. If we look at points corresponding to the first 15 years, plots for European countries are very much alike. For Female plots, the four points corresponding to years 1948-1951 follow the same paths in all six European countries; for the US the path is different. For England and Wales, France and Italy, the projections on the planed spanned by the first two principal axes look strikingly similar.

Performing PCA without centering and normalizzation is disappointing. The first principal component catches almost all the inertia.

Assessing quality of reconstruction through truncated SVD

Canonical Correlation Analysis

- ☐ Build a function that takes as input
  - a dataframe like `life_table_pivot`,
  - a couple of countries, say `Spain` and `Sweeden`,
  - a vector of `Year`, say `1948:1998`
  - a `Gender` say `Female` returns a matrix called Z with rows corresponding to `Year` and columns corresponding to couples `(Country, Age)`.

R4Data Science Tidy