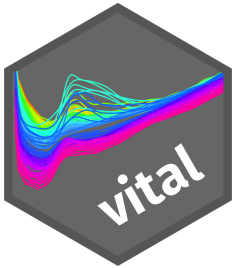


Tidy data analysis for demography using R

Rob J Hyndman

26 June 2024

MONASH University

Outline

Outline

tibble objects

Australian Deaths 1901–2020

```
tibble::as_tibble(aus) |>
  arrange(desc(State), Year, Age, Sex)
```

```
# A tibble: 145,440 x 7
    Year   Age Sex    State Mortality Exposure Deaths
   <int> <int> <chr>  <chr>     <dbl>    <dbl>  <dbl>
 1  1901     0 female WA       0.129      2511    325
 2  1901     0 male   WA       0.158      2634    416
 3  1901     1 female WA       0.0275     2219     61
 4  1901     1 male   WA       0.0391     2175     85
 5  1901     2 female WA       0.00688    2180     15
 6  1901     2 male   WA       0.0131     2208     29
 7  1901     3 female WA       0.00584    1884     11
 8  1901     3 male   WA       0.00503    1988     10
 9  1901     4 female WA       0.00290    1722      5
10  1901     4 male   WA       0.00287    1743      5
```

tsibble objects

Australian Deaths 1901–2020

```
tsibble::as_tsibble(aus) |>
  arrange(desc(State), Year, Age, Sex)
```

```
# A tsibble: 145,440 x 7 [1Y]
# Key:       Age, Sex, State [1,212]
    Year   Age Sex    State Mortality Exposure Deaths
   <int> <int> <chr>  <chr>     <dbl>    <dbl>  <dbl>
 1  1901     0 female WA       0.129      2511    325
 2  1901     0 male   WA       0.158      2634    416
 3  1901     1 female WA       0.0275     2219     61
 4  1901     1 male   WA       0.0391     2175     85
 5  1901     2 female WA       0.00688    2180     15
 6  1901     2 male   WA       0.0131     2208     29
 7  1901     3 female WA       0.00584    1884     11
 8  1901     3 male   WA       0.00503    1988     10
 9  1901     4 female WA       0.00290    1722      5
```

tsibble objects

Australian Deaths 1901–2020

```
tsibble::as_tsibble(aus) |>
  arrange(desc(State), Year, Age, Sex)
```

```
# A tsibble: 145,440 x 7 [1Y]
# Key:       Age, Sex, State [1,212]
    Year   Age Sex    State Mortality Exposure Deaths
   <int> <int> <chr>  <chr>     <dbl>    <dbl>  <dbl>
 1  1901     0 female WA       0.129      2511    325
 2  1901     0 male   WA       0.158      2634    416
 3  1901     1 female WA       0.0275     2219     61
 4  1901     1 male   WA       0.0391     2175     85
 5  1901     2 female WA       0.00688    2180     15
 6  1901     2 male   WA       0.0131     2208     29
 7  1901     3 female WA       0.00584    1884     11
 8  1901     3 male   WA       0.00503    1988     10
 9  1901     4 female WA       0.00290    1722      5
```

Variables

Index:

- Year

Keys:

- Age
- Sex
- State

tsibble objects

Australian Deaths 1901–2020

```
tsibble::as_tsibble(aus) |>
  arrange(desc(State), Year, Age, Sex)
```

```
# A tsibble: 145,440 x 7 [1Y]
# Key:       Age, Sex, State [1,212]
    Year   Age Sex    State Mortality Exposure Deaths
   <int> <int> <chr>  <chr>     <dbl>    <dbl>  <dbl>
 1  1901     0 female WA       0.129      2511    325
 2  1901     0 male   WA       0.158      2634    416
 3  1901     1 female WA       0.0275     2219     61
 4  1901     1 male   WA       0.0391     2175     85
 5  1901     2 female WA       0.00688    2180     15
 6  1901     2 male   WA       0.0131     2208     29
 7  1901     3 female WA       0.00584    1884     11
 8  1901     3 male   WA       0.00503    1988     10
 9  1901     4 female WA       0.00290    1722      5
```

Variables

Index:

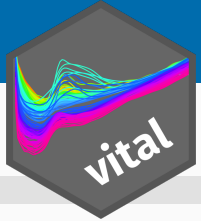
- Year

Keys:

- Age
- Sex
- State

Every row must have a unique combination of Index and Keys

vital objects



Australian Deaths 1901–2020

```
aus
```

```
# A vital: 145,440 x 7 [1Y]
# Key:     Age, Sex, State [1,212]
    Year   Age Sex    State Mortality Exposure Deaths
   <int> <int> <chr>  <chr>     <dbl>    <dbl>  <dbl>
 1  1901     0 female WA       0.129       2511    325
 2  1901     0 male   WA       0.158       2634    416
 3  1901     1 female WA       0.0275      2219     61
 4  1901     1 male   WA       0.0391      2175     85
 5  1901     2 female WA       0.00688     2180     15
 6  1901     2 male   WA       0.0131      2208     29
 7  1901     3 female WA       0.00584     1884     11
 8  1901     3 male   WA       0.00503     1988     10
 9  1901     4 female WA       0.00290     1722      5
10  1901     4 male   WA       0.00287     1743      5
```

Variables

Index:

- Year

Keys:

- Age
- Sex
- State

Variables denoting age, sex, deaths, births and population can also be specified.

Every row must have a unique combination of Index and Keys

Outline

Human Mortality Database

mortality.org

HMD

HOME PROJECT PEOPLE METHODS DATA RESEARCH LINKS

Registration Login

Human Mortality Database

Reliability and Accuracy Matter

The Human Mortality Database (HMD) is the world's leading scientific data resource on mortality in developed countries. The HMD provides detailed high-quality harmonized mortality and population estimates to researchers, students, journalists, policy analysts, and others interested in human longevity. The HMD follows open data principles.

- Short-Term Mortality Fluctuations
- Cause-of-Death Data Series
- Subnational Mortality Databases
- Citing HMD

Data by country or area

Australia	Denmark	Ireland	Norway	Switzerland
Austria	Estonia	Israel	Poland	Taiwan
Belarus	Finland	Italy	Portugal	U.K.
Belgium	France	Japan	Republic of Korea	U.S.A.

Human Fertility Database

humanfertility.org

HFD HOME PROJECT PEOPLE METHODS DATA RESEARCH LINKS Registration Login

Human Fertility Database

The Human Fertility Database (HFD) is the leading scientific data resource on fertility in the developed countries. This open access database provides detailed and high-quality historical and recent data on period and cohort fertility by age of mother and birth order. The HFD is entirely based on official vital statistics and places a great emphasis on rigorous data checking and documentation. The HFD adopts uniform methodology to warrant data comparability across time and between countries. The database follows open data principles.

- Short-Term Fertility Fluctuations
- Human Fertility Collection
- Citing HFD
- What's new

For users who seek fast access to the most commonly used summary indicators of period and cohort fertility, we provide excel tables comprising the following indicators for all the HFD countries:

HFD summary indicators					
Total fertility rate	Tempo-adjusted TFR	Mean age at birth	Mean age at first birth	Completed cohort fertility	Cohort parity

HMD imports

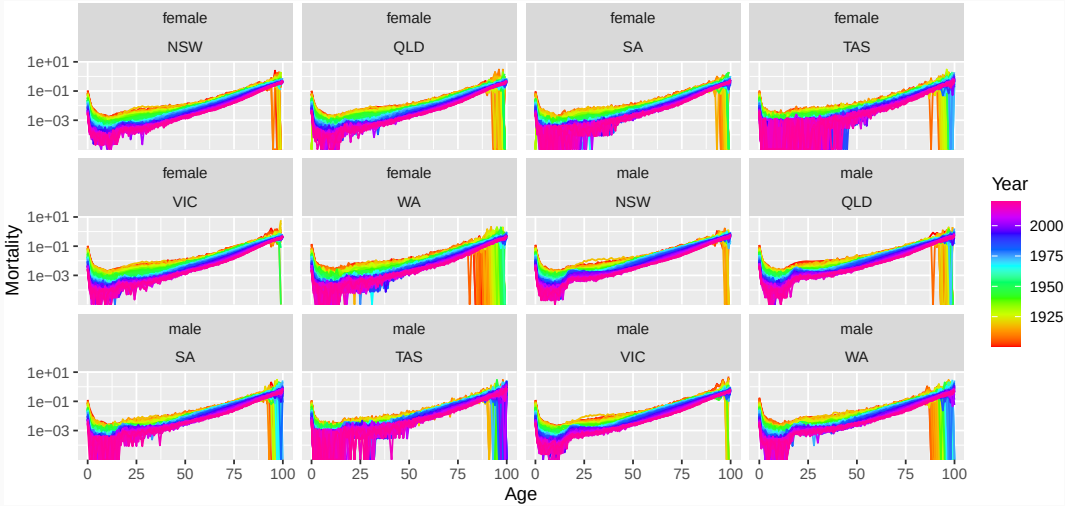
```
norway <- read_hmd(
  country = "NOR",
  username = "Nora.Weigh@mymail.com",
  password = "FF!5xeEFa6"
)
norway_births <- read_hmd(
  country = "NOR",
  username = "Nora.Weigh@mymail.com",
  password = "FF!5xeEFa6",
  variables = "Births"
)
```

- Uses `HMDHFDplus` package to handle the downloads.
- Default variables: `Deaths`, `Exposures`, `Population`, `Mx`
- Only 1x1 data supported.
- `read_hmd_files()` and `read_hfd_files()` allow reading of downloaded files.

Outline

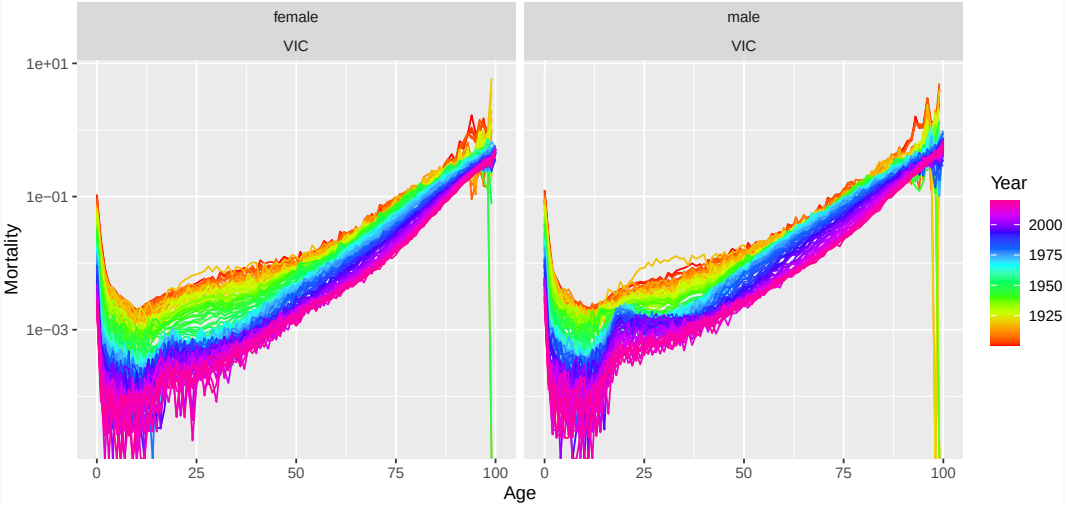
Rainbow plots

```
aus |> autoplot(Mortality) + scale_y_log10()
```



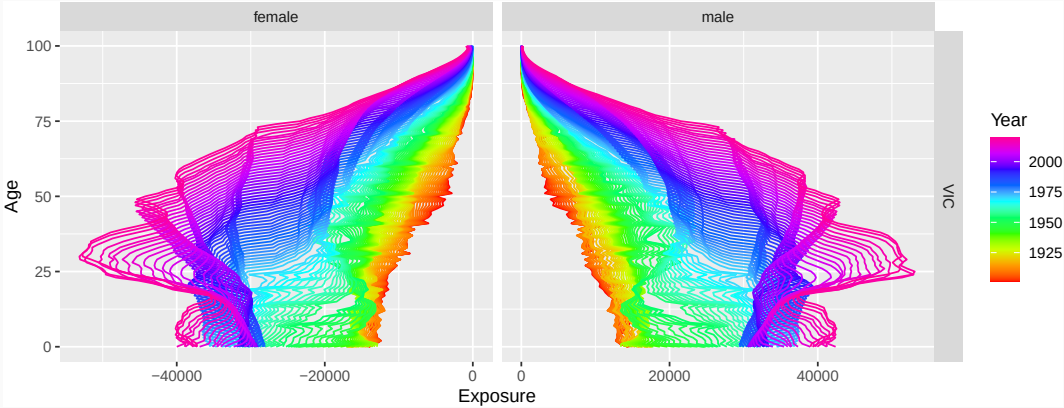
Rainbow plots

```
aus |> filter(State == "VIC") |> autoplot(Mortality) + scale_y_log10()
```



Rainbow plots

```
aus |> filter(State == "VIC") |>
  mutate(Exposure = if_else(Sex == "female", -Exposure, Exposure)) |>
  autoplot(Exposure) +
  facet_grid(State ~ Sex, scales = "free_x") + coord_flip()
```



Outline

Life tables

```
life_table(aus)
```

```
# A vital: 145,440 x 14 [1Y]
# Key:     Age, Sex, State [1,212]
    Year   Age Sex   State      mx       qx    lx       dx    Lx    Tx    ex    rx
   <int> <int> <chr> <chr>   <dbl>    <dbl> <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>
 1  1901     0 fema~ NSW   0.107   0.100   1      1.00e-1 0.935  56.2  56.2 0.935
 2  1901     1 fema~ NSW   0.0247  0.0244  0.900 2.20e-2 0.889  55.3  61.5 0.951
 3  1901     2 fema~ NSW   0.00686 0.00683 0.878 6.00e-3 0.875  54.4  62.0 0.984
 4  1901     3 fema~ NSW   0.00441 0.00441 0.872 3.84e-3 0.870  53.5  61.4 0.994
 5  1901     4 fema~ NSW   0.00374 0.00374 0.868 3.24e-3 0.867  52.7  60.7 0.996
 6  1901     5 fema~ NSW   0.00274 0.00274 0.865 2.37e-3 0.864  51.8  59.9 0.997
 7  1901     6 fema~ NSW   0.00252 0.00251 0.863 2.17e-3 0.861  50.9  59.1 0.997
 8  1901     7 fema~ NSW   0.00216 0.00216 0.860 1.86e-3 0.859  50.1  58.2 0.998
 9  1901     8 fema~ NSW   0.00169 0.00169 0.859 1.45e-3 0.858  49.2  57.3 0.998
10  1901     9 fema~ NSW   0.00109 0.00109 0.857 9.36e-4 0.857  48.4  56.4 0.999
# i 145,430 more rows
```

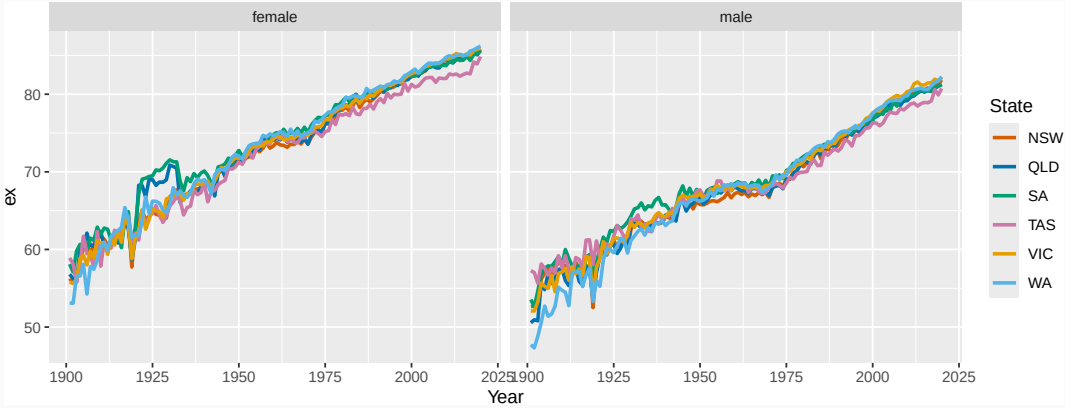
Life expectancy

```
life_expectancy(aus)
```

```
# A vital: 1,440 x 8 [1Y]
# Key:     Age, Sex, State [12]
    Year   Age Sex    State    ex    rx    nx    ax
   <int> <int> <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1  1901     0 female NSW    56.2 0.935     1 0.352
 2  1901     0 female QLD    56.8 0.937     1 0.338
 3  1901     0 female SA     58.1 0.939     1 0.324
 4  1901     0 female TAS    58.9 0.946     1 0.275
 5  1901     0 female VIC    55.8 0.937     1 0.334
 6  1901     0 female WA     53.1 0.922     1 0.35
 7  1901     0 male   NSW    52.6 0.925     1 0.33
 8  1901     0 male   QLD    50.6 0.924     1 0.33
 9  1901     0 male   SA     53.5 0.922     1 0.33
10  1901     0 male   TAS    57.3 0.930     1 0.33
# i 1,430 more rows
```

Life expectancy

```
life_expectancy(aus) |>
  ggplot(aes(x = Year, y = ex, colour = State)) +
  geom_line(linewidth = 1) +
  facet_grid(. ~ Sex)
```



Outline

Benchmark models

```
fit <- aus |>
  model(
    naive = FNAIVE(Mortality),
    mean = FMEAN(Mortality)
  )
fit
```

```
# A mable: 12 x 4
# Key:     Sex, State [12]
   Sex    State    naive    mean
   <chr>  <chr>  <model> <model>
 1 female NSW    <FNAIVE> <FMEAN>
 2 female QLD    <FNAIVE> <FMEAN>
 3 female SA     <FNAIVE> <FMEAN>
 4 female TAS    <FNAIVE> <FMEAN>
 5 female VIC    <FNAIVE> <FMEAN>
 6 female WA     <FNAIVE> <FMEAN>
```

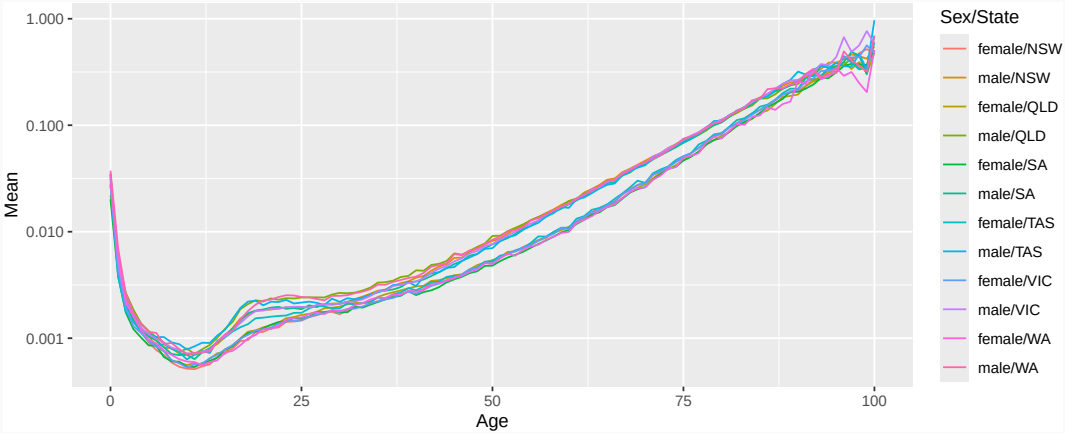
Benchmark models

```
fit |>
  filter(Sex == "female", State == "NSW") |>
  select(mean) |>
  report()
```

```
Series: Mortality
Model: FMEAN

# A tibble: 101 x 3
     Age     mean    sigma
   <int>    <dbl>    <dbl>
 1     0 0.0279   0.0256
 2     1 0.00528  0.00648
 3     2 0.00223  0.00245
 4     3 0.00145  0.00151
 5     4 0.00115  0.00116
 6     5 0.000915 0.000913
```

Benchmark models

```
fit |>
  select(mean) |>
  autoplot() + scale_y_log10()
```



Lee-Carter models

```
fit <- aus |>
  model(
    naive = FNAIVE(Mortality),
    mean = FMEAN(Mortality),
    lc = LC(log(Mortality))
  )
fit
```

```
# A mable: 12 x 5
# Key:     Sex, State [12]
   Sex    State    naive    mean      lc
   <chr>  <chr>  <model> <model> <model>
 1 female NSW    <FNAIVE> <FMEAN>    <LC>
 2 female QLD    <FNAIVE> <FMEAN>    <LC>
 3 female SA     <FNAIVE> <FMEAN>    <LC>
 4 female TAS    <FNAIVE> <FMEAN>    <LC>
 5 female VIC    <FNAIVE> <FMEAN>    <LC>
```

Lee-Carter models

```
fit |>
  filter(Sex == "female", State == "NSW") |>
  select(lc) |>
  report()
```

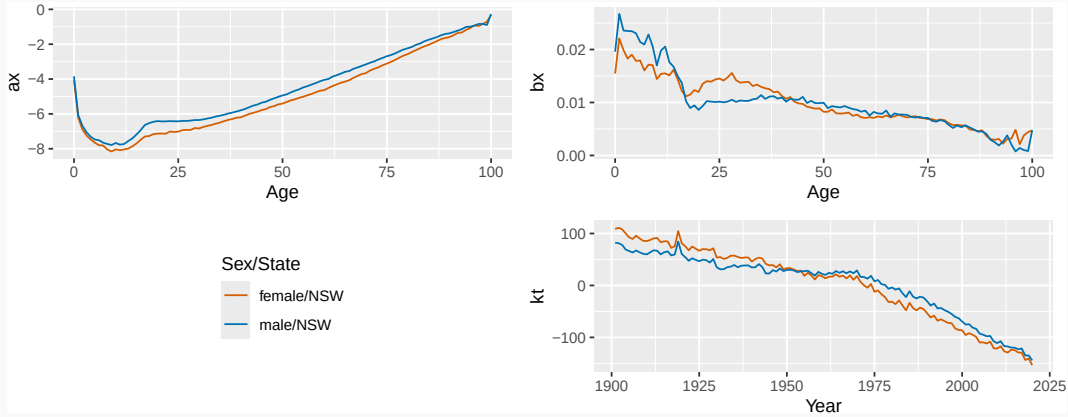
```
Series: Mortality
Model: LC
Transformation: log(Mortality)

Options:
  Adjust method: dt
  Jump choice: fit

Age functions
# A tibble: 101 x 3
    Age    ax     bx
  <int> <dbl>  <dbl>
1     0 -4.07 0.0155
2     1 -6.20 0.0221
```

Lee-Carter models

```
Age functions
# A tibble: 101 × 3
    Age    ax      bx
  <int> <dbl>   <dbl>
1     0 -4.07 0.0155
2     1 -6.20 0.0221
3     2 -6.89 0.0199
# i 98 more rows

Time coefficients
# A tsibble: 120 x 2 [1Y]
   Year    kt
  <int> <dbl>
1  1901 109.
2  1902 111.
3  1903 108.
# i 117 more rows

Time series model: RW w/ drift

Variance explained: 86.61%
```

Lee-Carter models

```
fit |>
  filter(State == "NSW") |>
  select(lc) |>
  autoplot()
```



Lee-Carter models

```
fit |> select(lc) |> age_components()
```

```
# A tibble: 1,212 x 5
   Sex    State   Age     ax      bx
   <chr>  <chr> <int> <dbl>   <dbl>
 1 female NSW       0 -4.07 0.0155
 2 female NSW       1 -6.20 0.0221
 3 female NSW       2 -6.89 0.0199
 4 female NSW       3 -7.24 0.0183
 5 female NSW       4 -7.47 0.0190
 6 female NSW       5 -7.65 0.0178
 7 female NSW       6 -7.80 0.0179
 8 female NSW       7 -7.81 0.0160
 9 female NSW       8 -8.05 0.0171
10 female NSW       9 -8.15 0.0170
# i 1,202 more rows
```

Lee-Carter models

```
fit |> select(lc) |> time_components()
```

```
# A tsibble: 1,440 x 4 [1Y]
# Key:       Sex, State [12]
   Sex    State  Year    kt
   <chr>  <chr> <int> <dbl>
 1 female NSW    1901 109.
 2 female NSW    1902 111.
 3 female NSW    1903 108.
 4 female NSW    1904 100.
 5 female NSW    1905  92.7
 6 female NSW    1906  89.5
 7 female NSW    1907  95.7
 8 female NSW    1908  90.5
 9 female NSW    1909  85.9
10 female NSW    1910  85.4
# i 1,430 more rows
```

Functional data models

Coherent FDM (generalizing Lee-Li)

Outline

Future plans

- Stochastic population forecasting (as per Booth-Hyndman 2008)
- StMoMo
- other tools in the demography package
- demography package to be maintained but not developed