Estimating Net Migration

Guy J. Abel

Net migration

- At the most basic level demographers are typically interested in the net balance of migration as a component of population change.
- Might not have an interest in the complexities involved in the different scales of migration to and from each region.
- Net migration tends to be used as it is readily available.
    - Data for in- and out-migration require specialized migration question in surveys or censuses.
    - Net migration does not require any questions on migration.
- Most censuses measure population changes accurately enough in order to develop a good estimate of net migration.
- Net migration has many weaknesses for the study of migration patterns, migration trends and population projections, see for example Rogers (1990)
    - Net migrants do not exist

Net migration estimation

- Three groups of methods to derive net migration
- First two are residual methods
  1. Vital statistics based on population change and natural increase data
  2. Survival methods based on population change data
  3. Place of birth methods based on changes in migrant stock data.
- United Nations Department of Economic and Social Affairs Population Division (1983) provides a nice discussion on the relative merits of each method

Vital statsitics

- The most elementary method to estimate net migration is using the demographic accounting equation

$$M = P^{t+n} - P^{t} - B + D$$

- Simple to calculate.
- Careful data preparation is required.
- Commonly applied to estimate net migration by sub-groups of populations where (e.g. sex)
- Less commonly applied to estimate net migration by age

$$M(x) = P^{t+n}(x+n) - P^{t}(x) + D(x)$$

  where parenthesis represent age groups x of size n
- Not easy to accurately estimate age-specific death counts that align to period between censuses

Vital statsitics

- The *migest* package has a `net_vs()` function to help obtain net migration estimates using vital statistics.
- Demonstrate using the `alabama_1970` data set in *migest*
  - Births are given in the under 10 age groups for `pop_1960`

```
> library(tidyverse)
> library(migest)
> alabama_1970
# A tibble: 68 x 6
   age_1970 sex    race  pop_1960 pop_1970 us_census_sr
   <fct>    <chr>  <chr>    <dbl>    <dbl>        <dbl>
 1 0-4      female white   104556   100224        0.965
 2 5-9      female white   119478   115269        0.956
 3 10-14    female white   120463   121922        0.997
 4 15-19    female white   114627   115128        1.01
 5 20-24    female white   113551   107480        0.998
 6 25-29    female white    93665    87706        0.989
 7 30-34    female white    76348    77285        0.996
 8 35-39    female white    74278    75115        0.994
 9 40-44    female white    79572    78924        0.989
10 45-49    female white    80719    78284        0.968
# ... with 58 more rows
# i Use `print(n = ...)` to see more rows
```

Vital statsitics

- Obtain race and sex population totals
- Need to remove those not born in the original population pop_1960.

```
> d <- alabama_1970 %>%
+   group_by(race, sex) %>%
+   summarise(births = sum(pop_1960[1:2]),
+             pop_1960 = sum(pop_1960) - births,
+             pop_1970 = sum(pop_1970)) %>%
+   ungroup()
`summarise()` has grouped output by 'race'. You can override using the
`.groups` argument.
> d
# A tibble: 4 x 5
  race      sex    births pop_1960 pop_1970
  <chr>     <chr>   <dbl>    <dbl>    <dbl>
1 non-white female 126886   515483   483882
2 non-white male   131767   467648   426452
3 white     female 224034  1159548  1298342
4 white     male   236481  1124061  1235489
```

Vital statsitics

- Given the vital statistics `net_vs()` estimate net migration and returns three additional columns
    - `pop_change` for the population difference
    - `natural_inc` for the difference in births and deaths
    - `net` for the net migration based on the two previous columns
- The `net_vs()` function assumes `births_col = "births"` and `deaths_col = "deaths"`.
    - Can alter from default if not the case

```
> d %>%
+   mutate(deaths = c(51449, 58845, 86880, 123220)) %>%
+   net_vs(pop0_col = "pop_1960", pop1_col = "pop_1970")
# A tibble: 4 x 9
  race      sex    births pop_1960 pop_1970 deaths pop_change natural_~1     net
  <chr>     <chr>   <dbl>    <dbl>    <dbl>  <dbl>      <dbl>      <dbl>   <dbl>
1 non-white female 126886   515483   483882  51449     -31601      75437 -107038
2 non-white male   131767   467648   426452  58845     -41196      72922 -114118
3 white     female 224034  1159548  1298342  86880     138794     137154    1640
4 white     male   236481  1124061  1235489 123220     111428     113261   -1833
# ... with abbreviated variable name 1: natural_inc
```

Difficulties

- Strictly speaking should refer to net migration estimates as a mixture of net migration and net balance of errors from the other data sources
- Assumes international migration is nil or negligible.
- Bogue, Hinze, and White (1982) list six difficulties with the vital statistics methods, most of which are due to the estimate is a residual from the combination of other data sources

1. Requires a stable administrative geography, where regions or countries do not change or at least enumerate population, births and deaths for the same units throughout the interval.
2. Adjustments will be required if there has been a big change in the method to collect census, for example switching from *de jure* to *de facto* for defining place of residence
3. Adjustments will be required if the birth and death periods do not align with the census dates. Typically vital statistics are annual measures starting from 1st January where as census dates are not usually on 1st January.

Difficulties

4. Births need to be tabulated or adjusted to mothers place of residence and deaths need to be tabulated or adjusted to place of residence or deceased. If place of occurrence is used for either then additional potential for error is created
5. Births and deaths need to be corrected for under-registration if it is known to exist.
6. Adjustments might be required to include/exclude population groups such as military or students - depending on how each are counted in the censuses and vital statistics registrations.

Survival methods

- Survival ratios can be used to compute mortality over the period, to then determine net migration as a residual.
- Survival ratios are an estimate of what proportion of a hypothetically closed population would be present at the end of the period.
    - Survival measures the force of mortality, rather than an overall population change
- Methods can be applied to total population or age-specific populations
- Preferred for age-specific net migration estimates as does not require age-specific death counts.
- Three related approaches using:
    - Forward survival ratios
    - Reverse survival ratios
    - Average survival ratios

Forward survival ratios

- Difference between the surviving expected population and observed population at the end of the period is an estimate of net migration during the interval

$$M'(x) = P^{t+n}(x+n) - s(x)P^{t}(x)$$

where:

- $M'(x)$ is the net migration between t and $t+n$ for age group x
- $P^{t+n}(x+n)$ is the observed population at the end of the period $(t+n)$ for age group x
- $s(x)$ is survival rate between t and $t+1$ for age group x
- $P^{t}(x)$ is the observed population at the start of the period (t) for age group x

Reverse survival ratios

- An alternative method is based on the reverse of the previous method
- Estimate the number of persons that would have been x years of age at the earlier census from the number who are enumerated as $x + n$ years old in the second census by applying *reverse survival ratios*

$$M''(x) = \frac{1}{s(x)} P^{t+n}(x+n) - P^{t}(x)$$

Average survival ratios

- The average survival ratios averages the net migration estimates form the forward and reverse survival ratios

$$\bar{M}(x) = \frac{1}{2}(M'(x) + M''(x))$$

- Siegel and Hamilton (1952) found the average survival ratio method provides the most exact approximation under normal circumstances
- Summary of assumptions for deaths:
  - Forward method: all deaths of migrants are not counted as migrants, equivalent to assuming that they all died at the place of origin.
  - Reverse method: the opposite is assumed. All migrants that die are counted as migrants, as are as those that would have moved had they survived the interval.
  - Average method: only those that died after moving are counted as migrants (approximately).

Survival methods in R

- The *migest* package contains the `net_sr()` function to calculate all three survival ratio estimates of net migration.
- Demonstrate using the `bombay_1951` data
    - Survival ratios come from a UN model life table

```
> bombay_1951
# A tibble: 13 x 5
   age_1941 age_1951 pop_1941 pop_1951    sr
   <fct>    <fct>       <dbl>    <dbl> <dbl>
 1 0-4      10-14       77135   132870 0.909
 2 5-9      15-19       85434   170227 0.957
 3 10-14    20-24       79185   263971 0.947
 4 15-19    25-29       82603   253964 0.931
 5 20-24    30-34      126247   195373 0.922
 6 25-29    35-39      155344   151259 0.916
 7 30-34    40-44      138843   118383 0.905
 8 35-39    45-49      109356    76421 0.885
 9 40-44    50-54       81626    65897 0.855
10 45-49    55-59       47062    32265 0.812
11 50-54    60-64       36908    22248 0.754
12 55-59    65-69       15134     9655 0.673
13 60+      70+         25094    10100 0.387
```

Survival methods in R

```
> net_sr(bombay_1951, pop0_col = "pop_1941", pop1_col = "pop_1951")
# A tibble: 13 x 10
   age_1~1 age_1~2 pop_1~3 pop_1~4    sr net_f~5 net_r~6 net_a~7 pop1_~8 pop0_~9
   <fct>   <fct>     <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 0-4     10-14     77135  132870 0.909  62777.  69085.  65931.  70093. 146220.
 2 5-9     15-19     85434  170227 0.957  88441.  92386.  90413.  81786. 177820.
 3 10-14   20-24     79185  263971 0.947 188975. 199530. 194252.  74996. 278715.
 4 15-19   25-29     82603  253964 0.931 177077. 190242. 183659.  76887. 272845.
 5 20-24   30-34    126247  195373 0.922  78935.  85585.  82260. 116438. 211832.
 6 25-29   35-39    155344  151259 0.916   8948.   9768.   9358. 142311. 165112.
 7 30-34   40-44    138843  118383 0.905  -7228.  -7990.  -7609. 125611. 130853.
 8 35-39   45-49    109356   76421 0.885 -20359. -23005. -21682.  96780.  86351.
 9 40-44   50-54     81626   65897 0.855  -3877.  -4535.  -4206.  69774.  77091.
10 45-49   55-59     47062   32265 0.812  -5959.  -7337.  -6648.  38224.  39725.
11 50-54   60-64     36908   22248 0.754  -5562.  -7382.  -6472.  27810.  29526.
12 55-59   65-69     15134    9655 0.673   -524.   -779.   -652.  10179.  14355.
13 60+     70+       25094   10100 0.387    399.   1031.    715.   9701.  26125.
# ... with abbreviated variable names 1: age_1941, 2: age_1951, 3: pop_1941,
#   4: pop_1951, 5: net_forward, 6: net_reverse, 7: net_average,
#   8: pop1_forward, 9: pop0_reverse
```

Survival methods in R

- Second example using `manila_1970` where survivor ratios come from census life tables for all of the Philippines
- Births and survival rates of children are unknown

```
> manila_1970
# A tibble: 16 x 4
   age_1970 pop_1960 pop_1970 phl_census_sr
   <fct>       <dbl>    <dbl>         <dbl>
 1 0-4            NA    85870        NA
 2 5-9            NA    83054        NA
 3 10-14       80275    79489         1.12
 4 15-19       70875   101410         0.992
 5 20-24       63250    90410         0.973
 6 25-29       85618    56055         0.889
 7 30-34       75793    44648         0.841
 8 35-39       60037    36963         0.957
 9 40-44       34813    28873         0.951
10 45-49       31927    23678         0.904
11 50-54       24297    19063         0.930
12 55-59       20207    14484         0.797
13 60-64       13714    10205         0.877
14 65-69        9366     6405         0.835
15 70-74        7921     3746         0.712
16 75+         11114     4779         0.562
```

Survival methods in R

- Estimate age-specific net migration for all ages, except children

```
> net_sr(manila_1970, pop0_col = "pop_1960", pop1_col = "pop_1970",
+        survival_ratio_col = "phl_census_sr")
# A tibble: 16 x 9
   age_1970 pop_1960 pop_1970 phl_cens~1 net_f~2 net_r~3 net_a~4 pop1_~5 pop0_~6
   <fct>       <dbl>    <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 0-4            NA    85870     NA           0       0       0     NA      NA
 2 5-9            NA    83054     NA           0       0       0     NA      NA
 3 10-14       80275    79489      1.12   -10196.  -9126.  -9661.  89685.  71149.
 4 15-19       70875   101410      0.992   31134.  31400.  31267.  70276. 102275.
 5 20-24       63250    90410      0.973   28877.  29683.  29280.  61533.  92933.
 6 25-29       85618    56055      0.889  -20082. -22582. -21332.  76137.  63036.
 7 30-34       75793    44648      0.841  -19117. -22723. -20920.  63765.  53070.
 8 35-39       60037    36963      0.957  -20497. -21416. -20957.  57460.  38621.
 9 40-44       34813    28873      0.951   -4244.  -4462.  -4353.  33117.  30351.
10 45-49       31927    23678      0.904   -5189.  -5739.  -5464.  28867.  26188.
11 50-54       24297    19063      0.930   -3521.  -3788.  -3655.  22584.  20509.
12 55-59       20207    14484      0.797   -1613.  -2025.  -1819.  16097.  18182.
13 60-64       13714    10205      0.877   -1822.  -2078.  -1950.  12027.  11636.
14 65-69        9366     6405      0.835   -1417.  -1697.  -1557.   7822.   7669.
15 70-74        7921     3746      0.712   -1890.  -2657.  -2274.   5636.   5264.
16 75+         11114     4779      0.562   -1472.  -2617.  -2045.   6251.   8497.
# ... with abbreviated variable names 1: phl_census_sr, 2: net_forward,
#   3: net_reverse, 4: net_average, 5: pop1_forward, 6: pop0_reverse
```

Survival methods in R

- Estimate children net migration setting `net_children = TRUE`.
- Uses method of Shryock and Siegel (1976, p381)
  - Age 0-4: $1/4 \times$ ratio of 0-4 population to 15-44 female population) $\times$ net migration for females aged 15-44
  - Age 5-9: $3/4 \times$ ratio of 5-9 population to 20-49 female population) $\times$ net migration for females aged 20-49.
- Can alter weights in `maternal_exposure` argument
  - default is `c(0.25, 0.75)`

```
> net_sr(manila_1970, pop0_col = "pop_1960", pop1_col = "pop_1970",
+         survival_ratio_col = "phl_census_sr", net_children = TRUE)
# A tibble: 16 x 9
   age_1970 pop_1960 pop_1970 phl_cens~1 net_f~2 net_r~3 net_a~4 pop1_~5 pop0_~6
   <fct>       <dbl>    <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 0-4            NA    85870     NA        -235.   -605.   -420.     NA      NA
 2 5-9            NA    83054     NA       -8935. -10486.  -9710.     NA      NA
 3 10-14       80275    79489      1.12   -10196.  -9126.  -9661.  89685.  71149.
 4 15-19       70875   101410      0.992   31134.  31400.  31267.  70276. 102275.
 5 20-24       63250    90410      0.973   28877.  29683.  29280.  61533.  92933.
 6 25-29       85618    56055      0.889  -20082. -22582. -21332.  76137.  63036.
 7 30-34       75793    44648      0.841  -19117. -22723. -20920.  63765.  53070.
 8 35-39       60037    36963      0.957  -20497. -21416. -20957.  57460.  38621.
 9 40-44       34813    28873      0.951   -4244.  -4462.  -4353.  33117.  30351.
```

Survival ratios

- The success of the above methods depend on the survival ratios.
- Ratios can be obtained from
  - Life table survival ratios (LTSR) as in `bombay_1951` example
  - Census survival ratios (CSR) as in `manila_1970` example
- Life table survival ratios are derived from the L_x columns of the life table; the ratio of persons in stationary population at age group x that are alive in comparisons to a previous age group $x - n$.

$$s_n(x) = \frac{L(x+n)}{L(x)}$$

- Can also be derived from mortality rates, if known.

Life table survival ratios

- For an accurate net migration estimate, $s_n(x)$ should
  - Measure the average mortality conditions of the period
  - Reasonably applicable to the area and population for which migration estimates are required.
- Age data to derive life tables may be inaccurate. Errors will impact the net migration estimates.

Census survival ratios

- Where appropriate life tables are not available or not appropriate, survival ratios can be computed from census age distributions
- A census survival ratio (CSR) is the ratio of the population aged $x + n$ at a given census to the population aged x at the census n years earlier.
- Computed for a nation as a whole assuming a "closed" population.
    - Adjust data for international migration before calculating CSR if international migration is a influential part of population change for a given area or group.

Age-specfic birthplace data

- Example to derive the birthplace-age-specific survival ratios from the 1950 and 1960 census data, given in usa_1960

```
> usa_1960
# A tibble: 288 x 7
   birthplace  race      sex    age_1950 age_1960 pop_1950 pop_1960
   <fct>       <fct>     <fct>  <fct>    <fct>       <dbl>    <dbl>
 1 New England white     male   0-4      10-14      465097   467291
 2 New England white     female 0-4      10-14      445100   450248
 3 New England non-white male   0-4      10-14        8419     8927
 4 New England non-white female 0-4      10-14        8205     8896
 5 New England white     male   5-9      15-19      378265   368524
 6 New England white     female 5-9      15-19      361845   359141
 7 New England non-white male   5-9      15-19        5421     5475
 8 New England non-white female 5-9      15-19        5501     5977
 9 New England white     male   10-19    20-29      606335   567349
10 New England white     female 10-19    20-29      591111   582993
# ... with 278 more rows
# i Use `print(n = ...)` to see more rows
```

Age-specfic birthplace data

- Focus on white males for example later on

```
> s <- usa_1960 %>%
+   filter(sex == "male",
+          race == "white") %>%
+   mutate(sr = pop_1960/pop_1950) %>%
+   select(-contains("pop"))
> s
# A tibble: 72 x 6
   birthplace       race  sex   age_1950 age_1960    sr
   <fct>            <fct> <fct> <fct>    <fct>    <dbl>
 1 New England      white male  0-4      10-14    1.00
 2 New England      white male  5-9      15-19    0.974
 3 New England      white male  10-19    20-29    0.936
 4 New England      white male  20-29    30-39    1.00
 5 New England      white male  30-39    40-49    0.996
 6 New England      white male  40-49    50-59    0.946
 7 New England      white male  50-59    60-69    0.825
 8 New England      white male  60+      70+      0.488
 9 Middle Atlantic white male  0-4      10-14    1.01
10 Middle Atlantic white male  5-9      15-19    0.975
# ... with 62 more rows
# i Use `print(n = ...)` to see more rows
```

Census survival ratios

- The CSR method tends to correct for systematic errors in the age data.
  - For example, get $s_n(x)$ in adolescent years greater than one as larger under-registration in 0-4 compared to 5-9 or 10-14 age groups.
- Systematic errors in the censuses might lead to survivor ratios to incorporate net census errors, that might lead to better estimate of net migration compared to LTSR.
- CSR tend to be less smooth than LTSR,
  - Perhaps more realistic age-patterns of net migration.

Limitations of census survival ratios

A number of weaknesses for CSR

- Assumes a closed population, so data must be adjusted for international migration before calculating CSR.
  - Good data on international migration data not always available
- Mortality may vary greatly in each region, so using a CSR based on national level data not always appropriate.
  - Build in regional correction factors
- Census enumeration may vary greatly in each region.
  - Build in regional correction factors
- A single census can not provide CSR for children.
  - Use birth statistics to approximate new born population for CSR calculation
  - If birth statistics are not reliable, use an approximation method using the ratio of women to children and female estimated net migration

Birthplace

- If data on lifetime migration at the start and end of the period are available, net migration can be estimated for each migrant group.
- Different procedure can be applied, depending on the availability of data
  1. Lifetime migration totals without age characteristics
  2. Lifetime migration data with age characteristics
- Both rely on a survival approach
  - Survival ratios are calculated by birthplace (and possibly other factors)
- If you view birthplace as just another dimension (such as sex) then the method is near identical to the survival ratio methods.
  - Can use the `net_sr()` function in *migest* once data is in correct format

Birthplace totals

- To demonstrate arranging birthplace totals with no age dimension and the application of `net_sv()` we use the `indian_sub` data in the *migest* package.

```
> indian_sub
# A tibble: 164 x 7
   zone              state                  sex    year in_migra~1 out_m~2 net_m~3
   <chr>             <chr>                  <chr> <int>      <dbl>   <dbl>   <dbl>
 1 United Provinces  United Provinces       male   1901     259836  878864 -619028
 2 East Zone         East Zone              male   1901     883052  529216  353836
 3 East Zone         Bihar-Orissa & Bengal  male   1901     466126  498082  -31956
 4 East Zone         Assam                  male   1901     416926   31134  385792
 5 Burma             Burma                  male   1901     352924    4489  348435
 6 South Zone        South Zone             male   1901     347416  509163 -161747
 7 South Zone        Madras                 male   1901     115290  450068 -334778
 8 South Zone        Travancore-Cochin      male   1901      42927    8515   34412
 9 South Zone        Mysore                 male   1901     189199   50580  138619
10 Bombay            Bombay                 male   1901     311720  248149   63571
# ... with 154 more rows, and abbreviated variable names 1: in_migrants,
#   2: out_migrants, 3: net_migrants
# i Use `print(n = ...)` to see more rows
```

Birthplace totals

- Separate columns for populations depending on birthplace
  - In state of birth or out of the state of birth.
- Rearrange data using `pivot_longer()` and `pivot_wider()` in the *tidyr* package
  - Location in its own column
  - Populations in each year in own columns
  - Work with male populations between 1921 and 1931 for those born in four selected states
  - Drop `net_migrants`, `sex` and `zone` columns

```
> d <- indian_sub %>%
+   filter(between(year, 1921, 1931),
+                  sex == "male",
+                  state %in% c("Assam", "Madras", "Mysore", "Bombay")) %>%
+   select(-net_migrants, -zone, -sex) %>%
+   pivot_longer(cols = contains("migrants"), names_to =  "location",
+                values_to = "pop") %>%
+   rename(birthplace = state)
```

Birthplace totals

```
> d
# A tibble: 16 x 4
   birthplace  year location        pop
   <chr>      <int> <chr>         <dbl>
 1 Assam       1921 in_migrants  671195
 2 Assam       1921 out_migrants  44136
 3 Madras      1921 in_migrants   97105
 4 Madras      1921 out_migrants 580136
 5 Mysore      1921 in_migrants  187000
 6 Mysore      1921 out_migrants  45349
 7 Bombay      1921 in_migrants  474553
 8 Bombay      1921 out_migrants 197593
 9 Assam       1931 in_migrants  754821
10 Assam       1931 out_migrants  41785
11 Madras      1931 in_migrants  119621
12 Madras      1931 out_migrants 723755
13 Mysore      1931 in_migrants  204260
14 Mysore      1931 out_migrants  54410
15 Bombay      1931 in_migrants  480557
16 Bombay      1931 out_migrants 202197
```

Birthplace totals

```
> d <- d %>%
+   mutate(location = case_when(
+     location == "in_migrants" ~ "in-state",
+     location == "out_migrants" ~ "out-of-state"
+   )) %>%
+   pivot_wider(names_from = year, values_from = pop, names_prefix = "pop_")
> d
# A tibble: 8 x 4
  birthplace location     pop_1921 pop_1931
  <chr>      <chr>           <dbl>    <dbl>
1 Assam      in-state       671195   754821
2 Assam      out-of-state    44136    41785
3 Madras     in-state        97105   119621
4 Madras     out-of-state   580136   723755
5 Mysore     in-state       187000   204260
6 Mysore     out-of-state    45349    54410
7 Bombay     in-state       474553   480557
8 Bombay     out-of-state   197593   202197
```

Birthplace totals

- Can now apply survival ratios to estimate net migration over a period
- Use a censuses survival ratio of 0.81 for both in migrants and out migrants

```
> d <- d %>%
+   mutate(sr = 0.81) %>%
+   net_sr(pop0_col = "pop_1921", pop1_col = "pop_1931")
> d
# A tibble: 8 x 10
  birthp~1 locat~2 pop_1~3 pop_1~4    sr net_f~5 net_r~6 net_a~7 pop1_~8 pop0_~9
  <chr>    <chr>     <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1 Assam    in-sta~  671195  754821  0.81 211153. 260683. 235918. 543668. 931878.
2 Assam    out-of~   44136   41785  0.81   6035.   7450.   6743.  35750.  51586.
3 Madras   in-sta~   97105  119621  0.81  40966.  50575.  45771.  78655. 147680.
4 Madras   out-of~  580136  723755  0.81 253845. 313389. 283617. 469910. 893525.
5 Mysore   in-sta~  187000  204260  0.81  52790   65173.  58981. 151470  252173.
6 Mysore   out-of~   45349   54410  0.81  17677.  21824.  19751.  36733.  67173.
7 Bombay   in-sta~  474553  480557  0.81  96169. 118727. 107448. 384388. 593280.
8 Bombay   out-of~  197593  202197  0.81  42147.  52033.  47090. 160050. 249626.
# ... with abbreviated variable names 1: birthplace, 2: location, 3: pop_1921,
#   4: pop_1931, 5: net_forward, 6: net_reverse, 7: net_average,
#   8: pop1_forward, 9: pop0_reverse
```

Birthplace totals

- To derive the net migration flow estimate for each of the states we need to make one more step
  - Subtract the net migration for the out-of-state migrants from the net migration for the in-state migrants

```
> d %>%
+   group_by(birthplace) %>%
+   summarise(net = net_forward[location == "in-state"] -
+                 net_forward[location == "out-of-state"])
# A tibble: 4 x 2
  birthplace      net
  <chr>         <dbl>
1 Assam       205118.
2 Bombay       54022.
3 Madras     -212879.
4 Mysore       35113.
```

Age-specfic birthplace data

- To demonstrate arranging age-specific birthplace data and the application of `net_sv()` we use the `new_england_1960` data in the *migest* package.
    - New England population totals by birthplace for white males.

```
> new_england_1960
# A tibble: 72 x 4
   birthplace         age_1960 pop_1950 pop_1960
   <fct>              <fct>       <dbl>    <dbl>
 1 New England        10-14      442577   417069
 2 Middle Atlantic    10-14        7651    17077
 3 East North Central 10-14        1831     4376
 4 West North Central 10-14         719     1313
 5 South Atlantic     10-14        3451     5578
 6 East South Central 10-14         679      960
 7 West South Central 10-14         830     1413
 8 Mountain States    10-14         533      819
 9 Pacific            10-14        1730     2687
10 New England        15-19      354131   314048
# ... with 62 more rows
# i Use `print(n = ...)` to see more rows
```

Age-specfic birthplace data

- Apply the age-sex-race-birthplace specific census survivorship rate based on the US census (see previous CSR slide)

```
> d <- new_england_1960 %>%
+   left_join(s)
Joining, by = c("birthplace", "age_1960")
> d
# A tibble: 72 x 8
   birthplace         age_1960 pop_1950 pop_1960 race  sex   age_1950     sr
   <fct>              <fct>       <dbl>    <dbl> <fct> <fct> <fct>     <dbl>
 1 New England        10-14      442577   417069 white male  0-4       1.00
 2 Middle Atlantic    10-14        7651    17077 white male  0-4       1.01
 3 East North Central 10-14        1831     4376 white male  0-4       1.01
 4 West North Central 10-14         719     1313 white male  0-4       1.00
 5 South Atlantic     10-14        3451     5578 white male  0-4       1.01
 6 East South Central 10-14         679      960 white male  0-4       1.01
 7 West South Central 10-14         830     1413 white male  0-4       1.02
 8 Mountain States    10-14         533      819 white male  0-4       1.02
 9 Pacific            10-14        1730     2687 white male  0-4       1.01
10 New England        15-19      354131   314048 white male  5-9       0.974
# ... with 62 more rows
# i Use `print(n = ...)` to see more rows
```

Age-specfic birthplace data

- Use the national age-sex-race-birthplace CSR to estimate net migration by birthplace and age in New England for white males

```
> d %>%
+   net_sr(pop0_col = "pop_1950", pop1_col = "pop_1960") %>%
+   relocate(contains("net"))
# A tibble: 72 x 13
   net_for~1 net_r~2 net_a~3 birth~4 age_1~5 pop_1~6 pop_1~7 race  sex   age_1~8
       <dbl>   <dbl>   <dbl> <fct>   <fct>     <dbl>   <dbl> <fct> <fct> <fct>
 1   -27596. -27466. -27531. New En~ 10-14    442577  417069 white male  0-4
 2     9333.   9222.   9278. Middle~ 10-14      7651   17077 white male  0-4
 3     2531.   2511.   2521. East N~ 10-14      1831    4376 white male  0-4
 4      594.    593.    593. West N~ 10-14       719    1313 white male  0-4
 5     2086.   2062.   2074. South ~ 10-14      3451    5578 white male  0-4
 6      271.    267.    269. East S~ 10-14       679     960 white male  0-4
 7      567.    556.    562. West S~ 10-14       830    1413 white male  0-4
 8      276.    270.    273. Mounta~ 10-14       533     819 white male  0-4
 9      932.    918.    925. Pacific 10-14      1730    2687 white male  0-4
10   -30963. -31782. -31373. New En~ 15-19    354131  314048 white male  5-9
# ... with 62 more rows, 3 more variables: sr <dbl>, pop1_forward <dbl>,
#   pop0_reverse <dbl>, and abbreviated variable names 1: net_forward,
#   2: net_reverse, 3: net_average, 4: birthplace, 5: age_1960, 6: pop_1950,
#   7: pop_1960, 8: age_1950
# i Use `print(n = ...)` to see more rows, and `colnames()` to see all variable name
```

Exercise (ex4.R)

```r
# 0.  a) Load the KOSTAT2022.Rproj file.
#     Run the getwd() below. It should print the directory where the
#     KOSTAT2022.Rproj file is located.
getwd()
#     b) Load the packages used in this exercise
library(tidyverse)
library(migest)
##
##
##
# 1. Run the code below to read in the population age structure data for Quebec
#    and a range of survival ratios
q <- read_csv("./data/quebec_1956.csv")
q
# 2. Estimate the age specific net migration counts based on the national census
#    survival ratio (column national_csr)
d1 <- #####(.data = q,
             p##### = "pop1951",
             pop1_col = #####,
             survival_ratio_col = #####)
d1
# 3. Find the total net migration estimates for the net_average method for the
#    estimates in the previous question
#####(d1$net_average)
```

References

Bogue, Donald J, Kenneth Hinze, and Michael White. 1982. *Techniques of Estimating Net Migration*. Chicago, USA: Community; Family Study Center. University of Chicago.

Rogers, Andrei. 1990. "Requiem for the Net Migrant." *Geographical Analysis* 22 (4): 283–300. http://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1990.tb00212.x/abstract.

Shryock, Henry S., and Jacob S. Siegel. 1976. *The Methods and Materials of Demography.* Edited by Edward G. Stockwell. Condensed. San Diego, California: Academic Press.

Siegel, Jacob S., and C. Horace Hamilton. 1952. "Some Considerations in the Use of the Residual Method of Estimating Net Migration." *Journal of the American Statistical Association* 47 (259): 475–500. https://doi.org/10.1080/01621459.1952.10501186.

United Nations Department of Economic and Social Affairs Population Division. 1983. *Methods of measuring internal migration.* New York, New York, USA: United Nations Publication. https://www.un.org/en/development/desa/population/publications/manual/migration/measuring-migration.asp.