Summary Migration Indices

Guy J. Abel

Background

- Compared with fertility and mortality, little attention given to the way that internal or domestic migration varies between nations.
- Comparisons of migration over time or between spatial units is complicated by many factors including:
  - Different or changing definitions of migration
  - Different or changing collection systems for migration
  - Different sizes of regions
  - Different or changing number of regions
  - Different ways of measuring distances
  - Different and changing underlying population sizes and structures
- Bell et al. (2002) brought together and proposed a number of summary measures to enable better comparisons
  - A number of these incorporated into the IMAGE software for comparisons of internal migration between many countries; Bell, Charles-Edwards, Ueffing, et al. (2015), Bell, Charles-Edwards, Kupiszewska, et al. (2015), Stillwell et al. (2016), Rees et al. (2016), Bernard, Bell, and Charles-Edwards (2014)

Background

- Bell et al. (2002) identified four main groups of migration indices:
  - Intensity of migration
  - Distance of migration
  - Migration connectivity
  - Effect of migration on the redistribution of populations

Migration intensity

- Migration intensity measures attempt to capture the overall level, or incidence, of mobility.
- Provide a single measure for comparison of migration intensities over time or space
- Some indices based on age-specific migration data
    - Will discuss later in the age-schedule section

Crude migration probability

- Crude migration intensity is a simple measure of the overall propensity to migrate
  - Similar to crude birth or death rate
- If using migration transition data, the crude migration probability (CMP) is

$$\mathtt{CMP} = 100M/P$$

where M is the total number of migrants in a given time period and P is the population at risk

Migration intensity

- Courgeau (1973) discussed how the intensity of migration is directly related to the number of regions n in the country

$$\texttt{CMP} = k \log(n^2)$$

- No intrinsic meaning to a single Courgeau's k , but can be used to compare migration intensity that cannot be seen from the raw data because of differences in their zonal systems.
- Higher the value of k, greater the intensity of migration

Migration intensity

- The `index_intensity()` function in the *migest* package calculates both intensity measures, given a migration and population data
- The *migest* package also contains a data set on Korean internal migration and populations of first level administrative districts
  - Data originally downloaded from `https://kosis.kr/eng`

```
> library(tidyverse)
> library(migest)
> korea_reg
# A tibble: 2,601 x 4
   orig  dest          year    flow
   <fct> <fct>        <int>   <int>
 1 Seoul Seoul         2012 1069300
 2 Seoul Busan         2012   21437
 3 Seoul Daegu         2012   13838
 4 Seoul Incheon       2012   32216
 5 Seoul Gwangju       2012   11811
 6 Seoul Daejeon       2012   14570
 7 Seoul Ulsan         2012    6799
 8 Seoul Sejong        2012    1015
 9 Seoul Gyeonggi-do   2012  254175
10 Seoul Gangwon-do    2012   21324
# ... with 2,591 more rows
```

Migration intensity

- The korea_pop contains the resident population in each region and year

```
> korea_pop
# A tibble: 153 x 3
   region  year population
   <chr>  <int>      <dbl>
 1 Seoul   2012   10195318
 2 Seoul   2013   10143645
 3 Seoul   2014   10103233
 4 Seoul   2015   10022181
 5 Seoul   2016    9930616
 6 Seoul   2017    9857426
 7 Seoul   2018    9765623
 8 Seoul   2019    9729107
 9 Seoul   2020    9668465
10 Busan   2012    3538484
# ... with 143 more rows
```

Migration intensity

- Calculate migration and population totals in 2020

```
> m <- korea_reg %>%
+   filter(year == 2020,
+          orig != dest) %>%
+   pull(flow) %>%
+   sum()
> m
[1] 2534114
> p <- korea_pop %>%
+   filter(year == 2020) %>%
+   pull(population) %>%
+   sum()
> p
[1] 51829023
>
> index_intensity(mig_total = m, pop_total = p,
+                 n = n_distinct(korea_pop$region))
# A tibble: 2 x 2
  measure     value
  <chr>       <dbl>
1 cmp         4.89
2 courgeau_k  0.863
```

Migration intensity

```
> mm <- korea_reg %>%
+   group_by(year) %>%
+   filter(orig != dest) %>%
+   summarise(m = sum(flow))
> mm
# A tibble: 9 x 2
   year       m
  <int>   <int>
1  2012 2512740
2  2013 2423429
3  2014 2507796
4  2015 2551424
5  2016 2453342
6  2017 2410930
7  2018 2429184
8  2019 2384948
9  2020 2534114
```

Migration intensity

```
> pp <- korea_pop %>%
+   group_by(year) %>%
+   summarise(p = sum(population))
> pp
# A tibble: 9 x 2
   year        p
  <int>    <dbl>
1  2012 50948272
2  2013 51141463
3  2014 51327916
4  2015 51529338
5  2016 51696216
6  2017 51778544
7  2018 51826059
8  2019 51849861
9  2020 51829023
```

Migration intensity

- Passing the vectors of migration and population totals can lead to confusing output

```
> d <- mm %>%
+   left_join(pp)
> d
# A tibble: 9 x 3
   year       m        p
  <int>   <int>    <dbl>
1  2012 2512740 50948272
2  2013 2423429 51141463
3  2014 2507796 51327916
4  2015 2551424 51529338
5  2016 2453342 51696216
6  2017 2410930 51778544
7  2018 2429184 51826059
8  2019 2384948 51849861
9  2020 2534114 51829023
```

Migration intensity

```
> index_intensity(mig_total = d$m, pop_total = d$p,
+                 n = n_distinct(korea_pop$region))
# A tibble: 18 x 2
   measure    value
   <chr>      <dbl>
 1 cmp        4.93
 2 courgeau_k 0.870
 3 cmp        4.74
 4 courgeau_k 0.836
 5 cmp        4.89
 6 courgeau_k 0.862
 7 cmp        4.95
 8 courgeau_k 0.874
 9 cmp        4.75
10 courgeau_k 0.838
11 cmp        4.66
12 courgeau_k 0.822
13 cmp        4.69
14 courgeau_k 0.827
15 cmp        4.60
16 courgeau_k 0.812
17 cmp        4.89
18 courgeau_k 0.863
```

Migration intensity

- Set `long = FALSE` to put each indicator in their own column

```
> index_intensity(mig_total = d$m, pop_total = d$p,
+                 n = n_distinct(korea_pop$region), long = FALSE)
# A tibble: 9 x 2
    cmp courgeau_k
  <dbl>      <dbl>
1  4.93      0.870
2  4.74      0.836
3  4.89      0.862
4  4.95      0.874
5  4.75      0.838
6  4.66      0.822
7  4.69      0.827
8  4.60      0.812
9  4.89      0.863
```

Migration intensity

- Use the `map2()` function in *purrr* to apply the function to subsets of the data
  - Provides results alongside the `year`
- Code below produces a nested list `i` containing the intensity measures for each year

```
> d <- mm %>%
+   left_join(pp) %>%
+   mutate(i = map2(.x = m, .y = p,
+                   .f = ~index_intensity(mig_total = .x,
+                                         pop_total = .y,
+                                         n = n_distinct(korea_pop$region),
+                                         long = FALSE)))
> d
# A tibble: 9 x 4
   year       m        p i
  <int>   <int>    <dbl> <list>
1  2012 2512740 50948272 <tibble [1 x 2]>
2  2013 2423429 51141463 <tibble [1 x 2]>
3  2014 2507796 51327916 <tibble [1 x 2]>
4  2015 2551424 51529338 <tibble [1 x 2]>
5  2016 2453342 51696216 <tibble [1 x 2]>
6  2017 2410930 51778544 <tibble [1 x 2]>
7  2018 2429184 51826059 <tibble [1 x 2]>
8  2019 2384948 51849861 <tibble [1 x 2]>
```

Migration intensity

- The `unnest()` function in the *tidyr* package binds each component of column `i` on top of each other
  - Easier to see changes over time

```
> unnest(d, cols = i)
# A tibble: 9 x 5
   year       m        p   cmp courgeau_k
  <int>   <int>    <dbl> <dbl>      <dbl>
1  2012 2512740 50948272  4.93      0.870
2  2013 2423429 51141463  4.74      0.836
3  2014 2507796 51327916  4.89      0.862
4  2015 2551424 51529338  4.95      0.874
5  2016 2453342 51696216  4.75      0.838
6  2017 2410930 51778544  4.66      0.822
7  2018 2429184 51826059  4.69      0.827
8  2019 2384948 51849861  4.60      0.812
9  2020 2534114 51829023  4.89      0.863
```

Migration distance

- As migration is a spatial activity, based on movements between two locations, comparisons should take account of the way that intensities of movement vary across space.
- There are a number of measures that summarize the effects of distance across a migration system
- The distance measure between each region is not straightforward
  - Ideally measure the typical distance that migrants travel.
  - Straight line distance between population-weighted centroids of each region provide a good approximation
- The costs faced by a migrant may not be represented well by the inter-centroid distance.
  - Locations in different regions might be very close to a border, so centroids will exaggerate the distance
  - Areas can take many shapes and sizes
  - Doughnut shaped regions might have centroids not in region
  - Indented coastlines might make regions look closer than they might be (culturally, travel cost)
  - Distance measures for within region moves cause another set of problems

Average migration distance

- Summary of the average migration distance can be calculated by taking a weighted average of the migration counts, where the corresponding distances are the weights
- Bell et al. (2002) note a median average as clearly preferable to a mean average as the distribution of distances is negatively skewed, reflecting the strong distance decay effect which consistently occurs
- Comparison with the mean average distance provides a guide to the skewness

Distance decay

- A more complete method to account for the skewness in migration distances is to fit a model to predict migration counts using the distances between each regions and extract the distance parameter
- Different models could be potentially used, but tend to be based on a `log(distance)` terms with categorical control variables for the origin and destination
- The distance decay parameter (β_3) in a Poisson log-linear model;

$$\log(m_{ij}) = \beta_0 + \beta_{1i}\texttt{origin} + \beta_{2j}\texttt{destination} + \beta_3 \log(\texttt{distance})$$

- The distance decay parameter of interest (β_3 in the equation above) is almost always negative, indicating an increase in migration leads to fewer predicted migrations.
    - The set of β_{1i} and β_{2j} represent some form of push and pull factors for each region (i and j)

Migration distance

- The `korea_dist` matrix provides estimates of the 2020 population weighted distances in kilometers between the 17 first level administrative districts in Korea
  - Data based on WorldPop estimates of Edwards et al. (2021)

```
> korea_dist
                    dest
orig                 Busan Chungcheongbuk-do Chungcheongnam-do Daegu Daejeon
  Busan                  0               216               255    88     201
  Chungcheongbuk-do    216                 0                58   130      45
  Chungcheongnam-do    255                58                 0   174      54
  Daegu                 88               130               174     0     122
  Daejeon              201                45                54   122       0
  Gangwon-do           286               122               159   203     166
  Gwangju              202               184               170   175     139
  Gyeonggi-do          312                96                83   225     125
  Gyeongsangbuk-do     107               125               176    32     128
  Gyeongsangnam-do      43               190               222    72     168
  Incheon              328               111                85   242     134
  Jeju                 306               381               368   333     337
  Jeollabuk-do         201               111                96   143      66
  Jeollanam-do         190               205               197   177     162
  Sejong               222                37                34   141      22
  Seoul                324               108                95   236     138
```

Migration distance

- The `index_distance()` function in the *migest* package provides three summary distance measures given a set of migration and distance measures between each origin and destination.
- The origin-destination migration flows can be given as a matrix to `m` or as a data frame, where the column names are assumed to be `orig`, `dest` and `flow`.
  - Can change using `orig_col`, `dest_col` and `flow_col` arguments
- The distance values can also be given as a matrix to `d` or as a data frame, where the column names are assumed to be `orig`, `dest` and `dist`.
- Origin and destination names in `m` and `dist` must match
- Removes all within migration moves from calculations

```
> # single year
> index_distance(m = filter(korea_reg, year == 2020),
+                d = korea_dist)
# A tibble: 3 x 2
  measure   value
  <chr>     <dbl>
1 mean    105.
2 median   68.7
3 decay    -0.852
```

Migration distance

```
> korea_reg %>%
+   nest(m = c(orig, dest, flow)) %>%
+   mutate(d = list(korea_dist)) %>%
+   mutate(i = map2(.x = m, .y = d,
+                   .f = ~index_distance(m = .x, d = .y, long = FALSE))) %>%
+   unnest(i)
# A tibble: 9 x 6
   year m                    d                   mean median  decay
  <int> <list>               <list>             <dbl>  <dbl>  <dbl>
1  2012 <tibble [289 x 3]> <dbl [17 x 17]>  108.    76   -0.784
2  2013 <tibble [289 x 3]> <dbl [17 x 17]>  107.    76   -0.795
3  2014 <tibble [289 x 3]> <dbl [17 x 17]>  108.    76   -0.816
4  2015 <tibble [289 x 3]> <dbl [17 x 17]>  108.    76   -0.828
5  2016 <tibble [289 x 3]> <dbl [17 x 17]>  107.    76   -0.820
6  2017 <tibble [289 x 3]> <dbl [17 x 17]>  108.    75.8 -0.839
7  2018 <tibble [289 x 3]> <dbl [17 x 17]>  107.    74.8 -0.839
8  2019 <tibble [289 x 3]> <dbl [17 x 17]>  107.    75.8 -0.836
9  2020 <tibble [289 x 3]> <dbl [17 x 17]>  105.    68.7 -0.852
```

Calculating distances

- There are a number of functions to a calculate distance matrices in R
- Require a set of longitude and latitudes
- If population weighted centriods are not available from national statistics offices, a number of research centers provide estimates
  - POPGRID Data Collaborative `https://www.popgrid.org/`
- Example for Ghana using WorldPop 2020 population weighted centroids
  - CSV from `https://www.worldpop.org/doi/10.5258/SOTON/WP00703`

```
>
> g <- read_csv("../data/PWD_2020_sub_national_100m.csv") %>%
+   filter(ISO == "GHA")
> g
# A tibble: 10 x 25
    year ISO   ISO_No Country_N Adm_N GID_1 HASC  PWC_Lat PWC_Lon     Pop Density
   <dbl> <chr>  <dbl> <chr>     <chr> <chr> <chr>   <dbl>   <dbl>   <dbl>   <dbl>
 1  2020 GHA      288 Ghana     Asha~ GHA.~ GH.AH    6.69  -1.60  6.43e6   258.
 2  2020 GHA      288 Ghana     Bron~ GHA.~ GH.BA    7.50  -2.04  2.73e6    70.6
 3  2020 GHA      288 Ghana     Cent~ GHA.~ GH.CP    5.47  -0.986 2.87e6   296.
 4  2020 GHA      288 Ghana     East~ GHA.~ GH.EP    6.21  -0.517 3.12e6   188.
 5  2020 GHA      288 Ghana     Grea~ GHA.~ GH.AA    5.64  -0.159 5.18e6  1414.
 6  2020 GHA      288 Ghana     Nort~ GHA.~ GH.NP    9.48  -0.569 3.25e6    47
 7  2020 GHA      288 Ghana     Uppe~ GHA.~ GH.UE   10.9   -0.680 1.11e6   129.
 8  2020 GHA      288 Ghana     Uppe~ GHA.~ GH.UW   10.4   -2.44  8.02e5    42.2
 9  2020 GHA      288 Ghana     Volta GHA.~ GH.TV    6.90   0.504 2.66e6   144.
```

Calculating distances

```
> g <- g %>%
+   filter(ISO == "GHA") %>%
+   select(Adm_N, PWC_Lon, PWC_Lat)
> g
# A tibble: 10 x 3
   Adm_N         PWC_Lon PWC_Lat
   <chr>           <dbl>   <dbl>
 1 Ashanti        -1.60     6.69
 2 Brong Ahafo    -2.04     7.50
 3 Central        -0.986    5.47
 4 Eastern        -0.517    6.21
 5 Greater Accra  -0.159    5.64
 6 Northern       -0.569    9.48
 7 Upper East     -0.680   10.9
 8 Upper West     -2.44    10.4
 9 Volta           0.504    6.90
10 Western        -2.16     5.45
```

Calculating distances

- The `distm()` function in the *geosphere* package provides great circle distance estimates in meters between centroids

```
> library(geosphere)
> ghana_dist <- g %>%
+   select(PWC_Lon, PWC_Lat) %>%
+   distm()
>
> ghana_dist
          [,1]     [,2]      [,3]      [,4]      [,5]     [,6]     [,7]
 [1,]      0.0 102992.7 151088.53 130977.06 197276.08 329699.1 471774.1
 [2,] 102992.7      0.0 254060.16 221397.05 293464.70 272448.3 399289.5
 [3,] 151088.5 254060.2      0.00  97220.64  93573.16 446740.6 596619.3
 [4,] 130977.1 221397.1  97220.64      0.00  74652.44 362233.8 513803.4
 [5,] 197276.1 293464.7  93573.16  74652.44      0.00 427847.5 579597.5
 [6,] 329699.1 272448.3 446740.58 362233.78 427847.51      0.0 151789.3
 [7,] 471774.1 399289.5 596619.34 513803.44 579597.52 151789.3      0.0
 [8,] 422969.2 325054.7 570615.97 511378.47 585433.81 229629.3 198393.6
 [9,] 233944.5 289137.3 229081.18 136579.93 158124.96 308917.8 455693.6
[10,] 149904.5 227534.2 130136.28 200358.81 222698.68 479309.9 619145.3
          [,8]     [,9]    [,10]
 [1,] 422969.2 233944.5 149904.5
 [2,] 325054.7 289137.3 227534.2
 [3,] 570616.0 229081.2 130136.3
 [4,] 511378.5 136579.9 200358.8
```

Calculating distances

- Add the origin and destination names to the `matrix` row and columns using `dimnames()`
- Allows to combine with migration data
- Divide by 1000 to get to kilometers

```
> dimnames(ghana_dist) <- list(orig = g$Adm_N, dest = g$Adm_N)
> round(ghana_dist/1000)
               dest
orig            Ashanti Brong Ahafo Central Eastern Greater Accra Northern
  Ashanti             0         103     151     131           197      330
  Brong Ahafo       103           0     254     221           293      272
  Central           151         254       0      97            94      447
  Eastern           131         221      97       0            75      362
  Greater Accra     197         293      94      75             0      428
  Northern          330         272     447     362           428        0
  Upper East        472         399     597     514           580      152
  Upper West        423         325     571     511           585      230
  Volta             234         289     229     137           158      309
  Western           150         228     130     200           223      479
               dest
orig            Upper East Upper West Volta Western
  Ashanti              472        423   234     150
  Brong Ahafo          399        325   289     228
  Central              597        571   229     130
```

Migration connectivity

- The size of migration flows in different migration corridors vary due to many reasons other than population sizes and distance.
  - Also reflect the strength of many other factors linking regions such as the strength of historical, cultural, social and economic ties, between regions.
- Bell et al. (2002) note a fragmentation in the literature on measures of connections in a migration system and the use of a range of terms including *spatial connectivity*, *spatial concentration*, *spatial inequality* and *spatial focusing*
- There are many indices on migration connectivity. See the *migration.indices* package for example.
- The `index_connectivity()` function provides 12 different measures, that can broadly be placed into 5 groups.
  - Requires a matrix or data frame of migration flows
  - When providing a data frame, function is assuming flows are in a column with name `flow`; change with `flow_col` if not.

Migration connectivity

```
> korea_reg %>%
+   filter(year == 2020) %>%
+   index_connectivity()
# A tibble: 11 x 2
   name                        value
   <chr>                       <dbl>
 1 connectivity               1
 2 inequality_equal           0.541
 3 inequality_sim             0.281
 4 gini_total                 0.709
 5 gini_orig_standardized     0.0493
 6 gini_dest_standardized     0.0517
 7 mwg_orig                   0.0370
 8 mwg_dest                   0.0389
 9 mwg_mean                   0.0379
10 cv                        17.9
11 acv                        3.43
```

Connectivity and Inequality

- The `connectivity` measure evaluates the proportion of the flows (excluding within region flows) that are non-zeros
  - More useful when many regions where populations are smaller
- Bell et al. (2002) inequality measures are based on a distributions of flows compared to distributions of expected flows
  - `inequality_equal` measures the distance of the observed flows to an expected distribution where all flows are equal
  - `inequality_sim` measures the distance of the observed flows to an expected distribution from a spatial interaction model equivalent to a Poisson regression mode for an independence fit

$$\widehat{\log(m_{ij})} = \beta_0 + \beta_{1i}\texttt{origin} + \beta_{2j}\texttt{destination}$$

- In both cases, a value of 0 shows the observed flows match the expected values (some form of equality) and 1 shows the maximum distance between the observed flows and the expected flows, i.e. maximum inequality.

Gini measures

- The Gini measures provide a value for the spatial focusing in a set of migration flow - i.e. how much of the migration is focused on a particular set of migration corridors
- Compares each flow with every other flow in the migration matrix.
- A `gini_total` value of zero indicates all flows are of equal size (no spatial focusing) to 1, only one single flow (maximum focusing).
- The `gini_orig_standardized` values provide a similar measure but compare every outflow from each origin with every other outflow from that origin.
  - Measures the extent to which the destination choices of out-migrants are spatially focused.
  - The `gini_dest_standardized` does the same but for the spatial focusing of origins of in-migrants.
  - The standardized values ensure a range of 0 and 1 - zero is no focusing

Migration totals Gini and Coefficient of Variation

- The migration weighted Gini indexes provide a measure of the focusing for the in-migration and out-migration totals (`mwg_orig` and `mwg_dest`)
- The `mwg_mean` is a simple average of `mwg_orig` and `mwg_dest` to provide a system wide measure of focusing for all migration totals.
  - As with the `gini_` measures from `index_connectivity()` values vary between zero (no focusing) and 1 (all migration goes through a single origin or destination)
- Rogers and Raymer (1998) proposed a coefficient of variation, provided by `cv` which compares the mean of the flows to the standard deviation of the flows.
  - Is not limited to 0 and 1
- The `acv` provides a similar measures of variation but based on the aggregate of coefficient of variations of in- and out-migration totals (based on the means and standard deviations of in- and out-migration totals)
  - Again, not limited to 0 and 1, but useful for comparisons across time or countries, where rising `cv` or `acv` would indicate greater inequality in migration flows or flow totals

Migration impact

- The impact of migration measures the extent to which migration acts to transform the pattern of human settlement
- Migration is already or becoming predominant mechanism leading to the redistribution of population in many regions of the world
- Descriptive studies tend to focus on regional net migration patterns
- Additional measures exist that summarize the overall effect of migration in redistributing a population across the entire system of regions

Migration effectiveness index

- The migration effectiveness index (MEI) of Shryock and Siegel (1976) compares the sum of net migration as a proportion of migration turnover, measuring the amount asymmetry or equilibrium in the migration network

$$\texttt{MEI} = 100\frac{\sum_i |\texttt{net}|}{\sum_i \texttt{turnover}_i} = 100\frac{\sum_i |m_{+i} - m_{i+}|}{\sum_i m_{+i} + m_{i+}}$$

- MEI range between 0 and 100.
- High values indicate migration is an efficient mechanism of population redistribution, based on large net totals for the given turnover.
- Values closer to zero are generated from more balanced migration systems with less population redistribution

Aggregate net migration rate

- The aggregate net migration rate (ANMR) of Bell et al. (2002) attempts to measure the overall effect of migration on the population settlement patterns by replacing the denominator of *MEI* with the each regions population P_i
- Index measures the net shift of population between regions per 100 residents in the country
  - No upper limit

$$\texttt{ANMR} = 100\frac{1}{2}\frac{\sum_i |\texttt{net}_i|}{\sum_i P_i} = 100\frac{1}{2}\frac{\sum_i |m_{+i} - m_{i+}|}{\sum_i P_i}$$

- Product of the CMI and MEI

Preference and velocity

- The manual by United Nations Department of Economic and Social Affairs Population Division (1983) provides two other impact measures, that seem to have fallen out of favor
- The preference index is based on an expected model of migration intestines based on population shares and the overall level of migration $M\frac{p_i}{P}\frac{p_j}{P}$, where M is the total migration flow and P is the total population based on the sum of populations in each region (p_i)
- Index compares the observed flows to an expected model where migration rates in all populations are the same
  - No upper limit

$$\texttt{preference} = \sum_{ij} \frac{m_{ij}}{M\frac{p_i}{P}\frac{p_j}{P}}$$

- The velocity index is based on a migration velocity measure $\frac{m_{ij}}{p_i p_j}$, multiplied by the total population and summed
- Index compares observed flows to an expected models where flows sizes are determined by population sizes alone
  - No upper limit

$$\texttt{velocity} = \sum_{ij} \frac{m_{ij}}{p_i p_j} P$$

Migration impact

- The `index_impact()` function in *migest* calculates all four measures given a set of migration flows and population sizes in each region
  - The `p` parameters assumes column names `region` and `pop` for region and population. Change from defaults using `reg_col` and `pop_col`

```
> index_impact(
+   m = subset(korea_reg, year == 2020),
+   p = subset(korea_pop, year == 2020),
+   pop_col = "population"
+ )
# A tibble: 4 x 2
  measure        value
  <chr>          <dbl>
1 effectivness   7.67
2 anmr           0.375
3 preference   375.
4 velocity      18.3
```

Migration impact

- Multiple years require nesting the migration and population data bases besides each other

```
> d <- korea_reg %>%
+   nest(m = c(orig, dest, flow)) %>%
+   left_join(korea_pop) %>%
+   nest(p = c(region, population))
Joining, by = "year"
> d
# A tibble: 9 x 3
   year m                       p
  <int> <list>                  <list>
1  2012 <tibble [289 x 3]> <tibble [17 x 2]>
2  2013 <tibble [289 x 3]> <tibble [17 x 2]>
3  2014 <tibble [289 x 3]> <tibble [17 x 2]>
4  2015 <tibble [289 x 3]> <tibble [17 x 2]>
5  2016 <tibble [289 x 3]> <tibble [17 x 2]>
6  2017 <tibble [289 x 3]> <tibble [17 x 2]>
7  2018 <tibble [289 x 3]> <tibble [17 x 2]>
8  2019 <tibble [289 x 3]> <tibble [17 x 2]>
9  2020 <tibble [289 x 3]> <tibble [17 x 2]>
```

Migration impact

- Apply the `index_impact()` function to each row and unnest

```
> d %>%
+   mutate(i = map2(.x = m, .y = p,
+                   .f = ~index_impact(m = .x, p = .y,
+                                      pop_col = "population",
+                                      long = FALSE))) %>%
+   unnest(i)
# A tibble: 9 x 7
   year m                      p                 effectivness  anmr preference velocity
  <int> <list>                 <list>                   <dbl> <dbl>      <dbl>    <dbl>
1  2012 <tibble [289 x 3~ <tibble [17 x ~         6.07 0.300       409.     20.2
2  2013 <tibble [289 x 3~ <tibble [17 x ~         5.72 0.271       371.     17.6
3  2014 <tibble [289 x 3~ <tibble [17 x ~         5.36 0.262       434.     21.2
4  2015 <tibble [289 x 3~ <tibble [17 x ~         7.73 0.383       459.     22.7
5  2016 <tibble [289 x 3~ <tibble [17 x ~         8.47 0.402       401.     19.0
6  2017 <tibble [289 x 3~ <tibble [17 x ~         7.99 0.372       417.     19.4
7  2018 <tibble [289 x 3~ <tibble [17 x ~         9.29 0.435       398.     18.7
8  2019 <tibble [289 x 3~ <tibble [17 x ~         6.94 0.319       391.     18.0
9  2020 <tibble [289 x 3~ <tibble [17 x ~         7.67 0.375       375.     18.3
```

Exercise (ex3.R)

```
# 0.  a) Load the KOSTAT2021.Rproj file.
#     Run the getwd() below. It should print the directory where the
#     KOSTAT2021.Rproj file is located.
getwd()
#     b) Load the packages used in this exercise
library(tidyverse)
library(migest)
library(geosphere)
##
##
##
# 1. Run the code below to read in the bilateral data in brazil_census_tidy.csv
#    from the 1991, 2000 and 2010 Brazilian censuses
br <- read_csv("./data/brazil_census_tidy.csv")
br
# 2. Run the code below to read in the WorldPop population data for Brazil in
#    2000 and check that the orig and dest names in the br migration data match
#    the region names in br_pop
br_pop <- read_csv("./data/PWD_2000_sub_national_100m.csv",
                   locale = readr::locale(encoding = "latin1")) %>%
  filter(ISO == "BRA") %>%
  select(Adm_N, contains("PWC"), Pop)
br_pop
# check names match
unique(br$orig) %in% br_pop$Adm_N
```

References I

Bell, Martin, Marcus Blake, Paul Boyle, O. Duke-Williams, Philip H. Rees, John Stillwell, and Graeme John Hugo. 2002. "Cross-national comparison of internal migration: issues and measures." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165 (3): 435–64. https://doi.org/10.1111/1467-985X.00247.

Bell, Martin, Elin Charles-Edwards, Dorota Kupiszewska, Marek Kupiszewski, John Stillwell, and Yu Zhu. 2015. "Internal Migration Data Around the World: Assessing Contemporary Practice." *Population, Space and Place* 21 (1): 1–17. https://doi.org/10.1002/psp.1848.

Bell, Martin, Elin Charles-Edwards, Philipp Ueffing, John Stillwell, Marek Kupiszewski, and Dorota Kupiszewska. 2015. "Internal Migration and Development: Comparing Migration Intensities Around the World." *Population and Development Review* 41 (1): 33–58. https://doi.org/10.1111/j.1728-4457.2015.00025.x.

Bernard, Aude, Martin Bell, and Elin Charles-Edwards. 2014. "Improved measures for the cross-national comparison of age profiles of internal migration." *Population Studies* 68 (2): 179–95. https://doi.org/10.1080/00324728.2014.890243.

Courgeau, Daniel. 1973. "Migrations et découpages du territoire." *Population (French Edition)* 28 (3): 511. https://doi.org/10.2307/1530704.

Edwards, Robin, Maksym Bondarenko, Alessandro Sorichetta, and Andrew J. Tatem. 2021. "Unconstrained subnational Population Weighted Density in 2000, 2005, 2010, 2015 and 2020 (100m resolution)." Southampton, UK: WorldPop, University of Southampton, UK. https://doi.org/10.5258/SOTON/WP00703.

References II

Rees, Philip H., Martin Bell, Marek Kupiszewski, Dorota Kupiszewska, Philipp Ueffing, Aude Bernard, Elin Charles-Edwards, and John Stillwell. 2016. "The Impact of Internal Migration on Population Redistribution: an International Comparison." *Population, Space and Place*, no. April. https://doi.org/10.1002/psp.2036.

Rogers, Andrei, and James Raymer. 1998. "The Spatial Focus of US Interstate Migration Flows." *International Journal of Population Geography* 4 (1): 63–80. https://doi.org/10.1002/(SICI)1099-1220(199803)4%3A1%3C63%3A%3AAID-IJPG87%3E3.0.CO%3B2-U.

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.

Stillwell, John, Martin Bell, Philipp Ueffing, K. Daras, Elin Charles-Edwards, Marek Kupiszewski, and Dorota Kupiszewska. 2016. "Internal migration around the world: comparing distance travelled and its frictional effect." *Environment and Planning A* 48 (8): 1657–75. https://doi.org/10.1177/0308518X16643963.

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.