

## **County Migration and Income - notes and links**

### **Introduction**

These maps allow users to answer questions about income and migration patterns at the county level in the United States. For example, these maps and the associated graphs let users track how many people move in and out of a particular county. Users can also analyze the incomes of immigrants and emigrants relative to the non-migrant population of the county as well as the average family size of each group.

The maps are based on the Statistics of Income (SOI) data, originally made available by a joint project of the Census Bureau and the Internal Revenue Service (IRS). The IRS has produced the migration data files on its own starting with the 2011/2012 match. The maps are a data visualization that allows for quick browsing through the data and analysis of income and migration patterns at the county level.

The SOI data contain information about every county over time. To make the visualization of panel data easier, two dimensions of the data are simultaneously displayed. As the user focuses on a year-county pair, the map of the entire country shows the distribution of the variable of interest across all counties in the given year. At the same time, the graph shows how the situation in the selected county evolves over time.

### **Sources of data**

The primary source of data is the SOI Tax Stats. The tax-based migration data had historically been a joint project undertaken by IRS and Census Bureau but it has been entirely under IRS since 2011. The data are publicly available and can be found on the IRS website (<https://www.irs.gov/uac/SOI-Tax-Stats-Migration-Data>). The data contain information about migration flows between US counties, beginning in year 1990, and are available in form of separate files for each period. These files include three variables: number of tax returns, total number of exemptions on these tax returns, and total income on these tax returns. For each county, the numbers are broken down into groups that include non-migrants as well as migrants to and from other counties. Due to privacy reasons, a migration flow between two counties is reported only if the number of households migrating is 10 or greater. Otherwise the migration between the two counties is not reported but is still included into the aggregate county migration statistic.

Migration data between years X and X+1 are based on the tax returns filed by individuals in years X and X+1. That is, the actual tax years for which incomes were reported are the X-1 and X years. Tax returns from these two years are matched by the taxpayer's identity. People who are matched and do not change the county of residence are treated as non-migrants. People who are matched and do change the county of residence are accounted for as inflow (immigration) for the destination county and outflow (emigration) for the origin county. Tax returns which are not matched (for example because the person filed tax return in only one of the two years) are not reported. The reported income and the number of exemptions are taken from the tax year X (that is returns filed in the year X+1).

Beginning with the data for the year pair 2011/2012, the methodology used for matching tax returns changed. Prior to 2011, households were matched from year to year based on the primary filer, but beginning in 2011, the SOI instituted matching that included secondary and dependent filers. This increased matches by about 5%. Also with the 2011/2012 match, the period over which tax returns were collected was extended from the initial nine months of the year to the entire year. This resulted in inclusion of people who file their tax returns later in the year, and who were previously excluded from the

data. This group includes many families with high incomes who were previously not reported in the statistics. The change in the data collection methodology can be thus easily seen by analyzing time series for counties like New York County, NY and San Francisco County, CA. More information about the new methodology and how it affects the time series data can be found on the IRS website (<https://www.irs.gov/pub/irs-soi/soi-a-inmig-id1509.pdf>).

With the year pair 2014/2015 an apparent change in the way migration is defined is reflected in the data. The immigration and emigration ratios for 2014-2015 were significantly lower than in past years. The median county immigration ratio drops from 5.5 to 4.1 and that the median county emigration ratio drops from 5.7 to 4.3 between 2013 and 2014.

The SOI documentation “2014-2015, Migration Data Users Guide” is available on the IRS website at:

<https://www.irs.gov/pub/irs-soi/1415inpublicmigdoc.pdf>

The documentation for the 2014/2015 file notes that a change has been made that affects the migration rates. The following note has been added in section C.1, page 2.

"Due to continuing efforts to combat identity theft, the method in which the IRS processes returns may undergo changes. These processing changes may have an impact on the migration data and should be taken into account when comparing the data across years."

Communication with the IRS SOI confirms that the change in the immigration/emigration rates are due to the new procedures and are not reflective of a processing error.

Subsequent year-to-year matches have produced emigration and immigration ratios similar to the trend prior to the 2014/2015 match.

The income reported by the IRS is reported in nominal dollars. Here the income data is rebased to real 2020 dollars using PCE price index available on the St. Louis Fed website. We use average annual values. (<https://research.stlouisfed.org/fred2/series/PCEPI>)

### Data manipulation

Unfortunately, the data for years 1990/1991 and 1991/1992 do not have the full set of variables. Moreover, the data for years 1992/1993, 1993/1994, and 1994/1995 have slightly different format and are incompatible with the subsequent data. Therefore, the maps and graphs use only data from years 1995/1996 and thereafter. We label data obtained from the  $X/X+1$  year pair as data in year  $X$ .

For each county-year pair  $(c, y)$  we extract nine variables:  $r_{cyn}, e_{cyn}, i_{cyn}, r_{cye}, e_{cye}, i_{cye}, r_{cye}, e_{cye},$  and  $i_{cye}$ , which are respectively: (a) non-migrant number of returns, (b) non-migrant number of exemptions, (c) non-migrant total income, (d) immigrant number of returns, (e) immigrant total exemptions, (f) immigrant total income, (g) emigrant number of returns, (h) emigrant total exemptions, and (i) emigrant total income for county  $c$  in year  $y$ . Denote the inverse of the price index by  $\delta_y$  where  $y$  is the year, and  $\delta_{2020} = 1$ . Then, for each county-year pair  $(c, y)$  we calculate the following variables:

- |                                 |                                      |
|---------------------------------|--------------------------------------|
| 1) $e_{cyn}$                    | – number of non-migrants.            |
| 2) $\delta_y i_{cyn} / e_{cyn}$ | – average non-migrant income.        |
| 3) $e_{cyn} / r_{cyn}$          | – non-migrant exemptions per return. |

- |  |                                    |
|--|------------------------------------|
| 4) $100 \times e_{c_{yi}} / (e_{c_{ye}} + e_{c_{yn}})$ | – immigration ratio.               |
| 5) $\delta_y i_{c_{yi}} / e_{c_{yi}}$                  | – average immigrant income.        |
| 6) $e_{c_{yi}} / r_{c_{yi}}$                           | – immigrant exemptions per return. |
| 7) $100 \times e_{c_{ye}} / (e_{c_{ye}} + e_{c_{yn}})$ | – emigration ratio.                |
| 8) $\delta_y i_{c_{ye}} / e_{c_{ye}}$                  | – average emigrant income.         |
| 9) $e_{c_{ye}} / r_{c_{ye}}$                           | – emigrant exemptions per return.  |

USA median non-migrant and migrant income series are depicted with each county's average non-migrant, immigrant, and emigrant income series.

### Data presentation

The color-coding of the county is based on the percentile the county was assigned in the national distribution of the given variable for the given year. Counties with missing data are in gray. The underlying map changes by selecting a different year or series option. Users can counties by clicking on the county of interest. The graph on the right-hand side of the panel reflects the information for the selected county.

### Contact us

Contact us at [perc@tamu.edu](mailto:perc@tamu.edu) for more information.