IRS County-to-County Migration Data, 1990-2010 *

BACKGROUND: The Internal Revenue Service's (IRS) county-to-county migration data are an incredible resource for understanding migration in the United States. Produced annually since 1990 in conjunction with the US Census Bureau, the IRS migration data represent 95 to 98 percent of the tax-filing universe and their dependents, making the IRS migration data one of the largest sources of migration data. However, any analysis using the IRS migration data must process at least seven legacy formats of these public data across more than 2000 data files – a serious burden for migration scholars.

OBJECTIVE: To produce a single, flat data file containing complete county-to-county IRS migration flow data and to make the computer code to process the migration data freely available.

METHODS: This paper uses R to process more than 2,000 IRS migration files into a single, flat data file for use in migration research.

CONTRIBUTION: To encourage and facilitate the use of this data, we provide a single, standardized, flat data file containing county-to-county 1-year migration flows for the period 1990-2010 and provide the full R script to download, process, and flatten the IRS migration data.

Introduction

Migration flow data (ie, the number of migrants from location i to location j) are typically difficult to obtain information despite their importance (Willekens et al., 2016; Rogers, Little and Raymer, 2010). Migration scholars typically focus on cross-border, international migration flow data and recent country-to-country migration data are vital for understanding

*The data and code that supports the creation of this data are available in the Supplementary Materials and online at https://osf.io/wgcf3/?view_only=c5ba62fb4821421ea0621bfd0d723e61.

migration processes (Abel and Sander, 2014; Abel, 2017, 2013). However, there is growing demonstrated importance surrounding subnational migration flows (Sorichetta et al., 2016; Curtis, Fussell and DeWaard, 2015).

In the United States, subnational migration flow data is available from three primary sources depending on the time period: the Decennial Census, the American Community Survey, and the Internal Revenue Service's (IRS) county-to-county migration data (described in detail in the corresponding section below). The IRS migration data are a pioneering use of administrative records to estimate demographic processes and are available on an annual basis since 1990. Because of the annual availability, relatively long time series, large universe due to the administrative records, and long history of use, the IRS data are an attractive data source for conducting migration research in the United States (e.g. (Curtis, Fussell and DeWaard, 2015; Molloy, Smith and Wozniak, 2011; Shumway and Otterstrom, 2001; Frey, 2009)). Unfortunately, these data exist in seven legacy formats, split between 2,000+ data files making analysis with this data rather burdensome and has likely hindered the widespread adoption of this valuable resource for US migration scholarship.

To encourage and facilitate the use of this tremendous migration resource, we make two contributions: (1) we publish a single, flat, standardized data file containing all county-to-county 1-year migration flows for the period 1990-2010, and (2) we publish the open-source R code used to process the IRS data into the single, flat, standardized data file for reproducibility, transparency, and educational purposes. Scholars who wish to use these data should still familiarize themselves with the strengths and weaknesses, idiosyncrasies, and design of these data (see (Gross, 2005; Engels and Healy, 1981; Franklin and Plane, 2006; Pierce, 2015) for discussions on the IRS data) and with the procedures outlined in this document and in the corresponding R code[1].

[1]The R code used to produce these data is available in the **Supplementary Materials** and can also be found in an online repository located at https://osf.io/wgcf3/?view_only=c5ba62fb4821421ea0621bfd0d723e61

We have attempted to introduce as little post-processing as possible to process the data into a common format. US Counties are fairly stable geographic units but some changes in county boundaries, names, and FIPS codes due occasionally occur[2]. To try and keep as close to the original data fidelity as possible, we did not recode any geographic changes and present the IRS migration data as-is. For instance, Broomfield County, Colorado (FIPS 08014) was created out of parts of Adams, Boulder, Jefferson, and Weld counties in 2001 and thus has data only after 2002. Users should be aware of any changes in county boundaries, county names, or FIPS changes that could substantially alter any analysis of this data[3].

We organize the following document as follows. First, we describe the IRS county-to-county migration data to provide an overview of the data for scholars who might be unfamiliar with the IRS migration data. Second, we provide usage notes providing important information that may assist other researchers who want to use our data. Third, we describe our single, flat, standardized file and document important nuances in the raw IRS migration data. Finally, we describe parts of the R code used to download the IRS migration data and process it into a common format.

The IRS migration data are an incredible tool for understanding migration. By providing these data in a readily available format and the subsequent open-source computer code used to process these data, we hope to facilitate their use in descriptive, exploratory, and analytical analyses of migration in the United States using administrative data. This data is particularly useful for understanding migration as a spatial entity and for investigating the evolution of migration systems over time.

[2]The Federal Information Processing Standard Publication (FIPS) is a 5-digit code used to uniquely identify US counties and county equivalents.

[3]More detailed information about county boundary, name, or FIPS changes can be found at the following locations https://www.census.gov/geo/reference/county-changes.html http://www.nber.org/asg/ASG_release/County_City/FIPS/FIPS_Changes.pdf https://www.cdc.gov/nchs/data/nvss/bridged_race/County_Geography_Changes.pdf https://www.ddorn.net/data/FIPS_County_Code_Changes.pdf

IRS Migration Data

The IRS began using tax data to estimate migration in the 1970s and 1980s (Engels and Healy, 1981; Franklin and Plane, 2006) and began releasing migration data in 1990. The IRS uses individual federal tax returns, matches these individual returns between two tax years (for instance tax year 2000 and tax year 2001), and identifies both migrants and non-migrants. Beginning with tax year 1991 (migration year 1990), the IRS produces these data in conjunction with the US Census Bureau using the IRS Individual Master File which contains every Form 1040, 1040A, and 1040EZ (Gross, 2005). Migration is identified when a current years' tax form contains an address that is different from the matched preceding years' return. A non-migrant is identified when there is no change in address between two years. For the 2002 tax year, the IRS migration data contained approximately 130 million returns (Gross, 2005).

The annual series of county-to-county migration data cover 95 to 98 percent of the tax-filing universe (or approximately 87% of US households (Molloy, Smith and Wozniak, 2011)) and their dependents making these data the largest migration data source for count flows between counties in the United States. The IRS derives migration information from tax-filings making those who do not file taxes most likely to be underrepresented in the migration data (Gross, 2005; DeWaard, Curtis and Fussell, 2016), namely undocumented populations, the poor, the elderly, and college students (Gross, 2005). However, the overwhelming majority of householders file US tax returns in the United States (Molloy, Smith and Wozniak, 2011).

The IRS reports a number of important variables in their data. They identify both the origin and destination counties; the number of tax returns or filers associated with those moves (roughly analogous to the number of households and listed as the `returns` field in the raw data) who moved from county i to county j and the number of tax exemptions associated with those moves (roughly analogous to the number of individuals and listed as the `exemptions` field in the raw data). They also report the number of non-migrants,

reported as the number of tax returns and exemptions associated with migrants from county i to county i. We treat the `exemptions` field as the total number of migrants.

Between 1990 and 2010, the IRS processed the county-to-county migration data using the same procedures. However, in 2011 the IRS introduced a new method for processing the migration data and introduced "enhancements" to improve the overall quality of the data (Pierce, 2015). The IRS introduced three major changes. First, they began basing migration on a full year of data as opposed to a partial year of data. To meet Census Bureau deadlines, the IRS processed all income tax returns filed before the end of September and did not process the returns filed between the end of September and the end of the calendar year. Beginning with migration year 2011, the IRS included the approximately 4% of returns that are filed between the end of September and December 31, allowing the IRS to produce a full calendar years' worth of migration. Second, the IRS improved the year-to-year matching, increasing the number of matched returns by 5 percent. Prior to 2011, the IRS used only the primary filer's taxpayer identification number (TIN), potentially excluding individuals who may be listed as a dependent in year 1 but file on their own in year 2 or in cases where a secondary filer in year 1 (such as a spouse) files as a primary filer in year 2. After 2011, the IRS broadened their matching process to include primary, secondary, and dependent TINs to improve the matching process by 5 percent. Third, the IRS began tabulating gross migration at the US State level by size of adjusted gross income (AGI) and the age of the primary taxpayer.

These changes to the processing of returns create a break in the historic time series. For this reason, we limit the data we process to the period 1990-2010, the last year before the new processing rules. If a scholar wishes to process any IRS migration data after 2010, the R code that we provide can be easily adapted to do so.

Comparisons to other US migration data

As stated in the preceding section, the three main sources of migration data in the US are the Decennial Census long form, the American Community Survey, and the IRS county-to-county migration data.

Up to and including Census 2000, on the long form of the Decennial Census the Census Bureau asked "Where did you live five years ago?" providing 5-year migration data once every decade. With the discontinuation of the long-form with Census 2010, the Census Bureau began collecting migration information on the American Community Survey (ACS) with the question "Where did you live one year ago?" providing 1-year migration data with each ACS release.

The Decennial long-form was a robust sample, gsurveying approximately one in every six or 16.7% of US households. The ACS is a smaller survey with a sample size of approximately 2 million US households per year. Due to the smaller sample size, the Census Bureau pools responses 5-year averages for county-to-county migration data. Thus, ACS migration data represents 1-year migration data over a 5-year period. The Census Bureau processes the ACS migration data and releases county-to-county migration data sets on an annual basis reflecting the 5-year average (2010-2014, 2011-2015, etc.).

The ACS migration products and the IRS migration data both have strengths and weaknesses. Table 1 compares the ACS migration products with the IRS migration data in some key areas. The ACS universe is more complete than the IRS migration universe, however the ACS migration data contains approximately 2% of the observations in the IRS migration data. The IRS releases the migration data annually allowing annual comparisons while the Census Bureau suggests only non-overlapping 5-year products should be compared to eachother (ie 2005-2009 and 2010-2014) (Brown, 2009).

Figure 1 demonstrates detectable changes in migration flows for four sample counties. These four sample counties are just some of the easily detectable impacts of major US events such as Hurricane Katrina in 2005 (Curtis, Fussell and DeWaard, 2015) or the Great

Table 1: Comparison between American Community Survey and IRS county-to-county migration data.

Issue	**ACS Migration Products**	**IRS Migration Data**
Sample Size	Approximately 2 million households per year	116 million+ households
Data universe	Sample is all US households	Universe is tax filing households
Coverage period	2005-2016	1990-2016
Time period reported	5-year average	Annual
Demographic Characteristics	Each five-year product reports different sociodemographic characteristics (e.g. 2011-2015 contains age/sex/race/hispanic origin, 2010-2016 contains relationship, household type, and tenure)	No demographic characteristics

Recession. These migration changes are largely be undetectable in the ACS migration data or our ability to detect such changes is hampered by the 5-year release.

While the IRS migration data allows for analysis of annual changes, the IRS migration data contains no sociodemographic information. The ACS and Decennial Census migration data, on the other hand, contain county-to-county migration information crossed by sociodemographic information for some releases.

Usage Notes

The dataset generated here provides detailed county-to-county 1-year migration data based on administrative records. Users of these data should be aware that although the data have been prepared in a transparent manner with documentation of their creation and post-processing, and with open-source computer code, little was done to post-process the data to correct any possible inconsistencies or errors. These data should be used only with full awareness of the inherent limitations of the IRS migration data and with the knowledge of

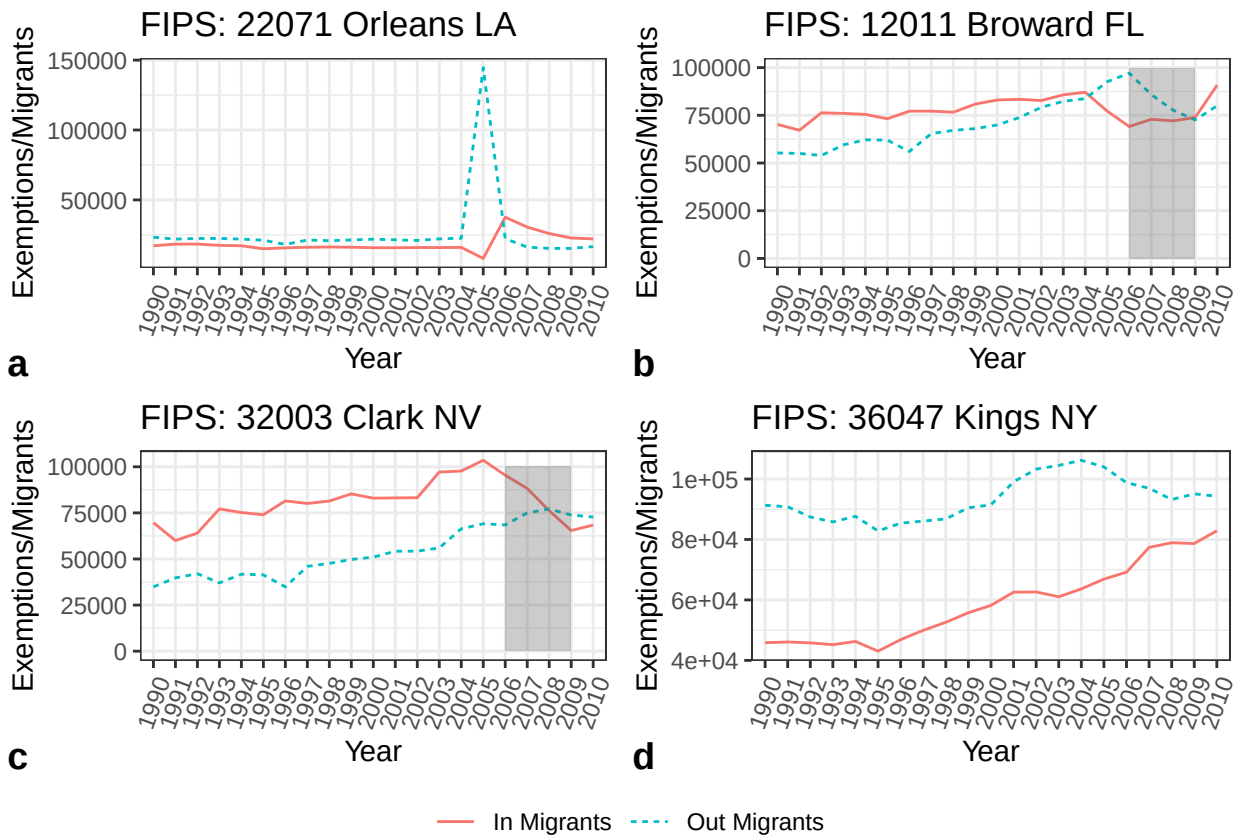


Figure 1: **Sample migration streams from the IRS migration data.** The annual release of the IRS migration data allows for detection of changes. The effect of Hurricane Katrina on New Orleans LA (a) is clearly visible; the moderate effect of the US housing bubble burst and Great Recession is detectable in Broward FL (b) and a much greater effect in Clark NV (c); and even migration streams nearly uneffected by major US changes is also detectable (d). These are just a few examples of what is possible with the IRS migration data.

the procedures outlined in this document and in the corresponding R code. Caveat emptor – users beware.

Users should be aware of several limitations of the IRS data. Namely, that any origin-destination pair with fewer than 10 tax filers is censored or suppressed by the IRS for privacy reasons. We have collected these censored flows into a unique FIPS code (FIPS 99999) by subtracting all uncensored flows from the total number of migrants. Any origin-destination pair with fewer than 10 tax filers over the entire period is thus excluded from the final datafile since no data would be recorded in the IRS datafile due to censoring.

Users should also be mindful of possible geographic changes to county boundaries that could affect the data.

The county migration data we present come from the `exemptions` field of the IRS migration data. The original IRS migration data contains two consistent fields across all years of data: a `returns` field and an `exemptions` field. Returns are the number of tax returns filed while exemptions are a proxy for the members of the household. We use the number of exemptions to better mimic the number of individuals migrating rather than the number of households.

Table 2 demonstrates the general structure of our flat migration data file.

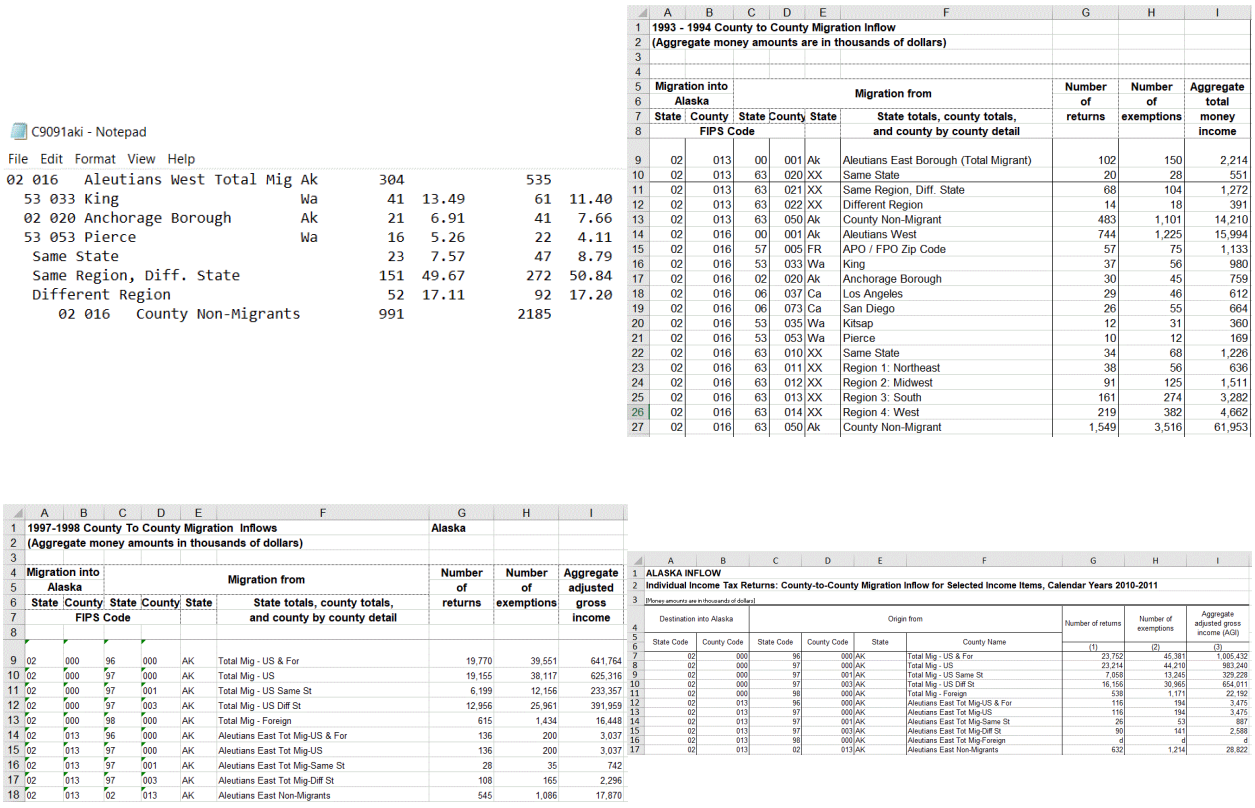
Table 2: Extract from the final migration data file. Origins and Destinations are the five-digit FIPS codes with 99999 representing all destinations with flows fewer than 10 filers. The counts represent the number of `exemptions` in the IRS data. Non-migrants are identified as having the same FIPS in the Origin and Destination fields.

Origin	Destination	1990	1991	1992	...	2010
01001	01001	26703	27278	28677	...	40643

Origin	Destination	1990	1991	1992	...	2010
01001	01003	0	0	27	...	39
01001	01013	0	0	0	...	22
01001	01021	101	94	112	...	149
...	...	...	...	...	...	...
01001	99999	1324	1020	1200	...	1758

Data Processing

The IRS migration data for the period 1990-2010 are available in seven legacy formats. Table 3 summarizes some of the similarities and differences in these formats. For every year, the IRS publishes approximately 104 data files. (52 state entities by in/out-migration. These are the 50 US states, DC, and a total US migration. Some years contain .csv and .dat summary files. The underlying file organization, file format, naming schema, and coding can differ between these legacy formats. Migration years 1990 and 1991 are available as fixed-width text files, while 1992-2010 are available as excel files. For years 1990-2003, the IRS separated in/out migration into separate folders while 2004-2010 are published in a single folder. Each legacy format utilizes a different file naming scheme as well, making pattern matching of file names (called grepping) difficult. Importantly, the IRS treats non-migrants and total migrants differently in the seven legacy formats. For 1990 and 1991, the IRS simply has a field that reads "County Non-Migrants" for non-migrants; for 1992-1994, the IRS introduced a State code 63 but two different County codes (010 for 1992 and 1994 and 050 for 1993) creating a 5-digit FIPS code of 63010 or 63050. After 1995, the IRS smartly set the origin FIPS equal to the destination FIPS for non-migrants. Lastly, Total Migrants are treated differently too. For 1990 and 1991, the destination field simply reads "Total Migration." For 1992-1994, the IRS introduced a State Code 00 and county code 001 for total migrants. After 1995, the IRS used State Code 96 and county Code 000 for a combined



C9091aki - Notepad

File Edit Format View Help

```
02 016   Aleutians West Total Mig Ak        304              535
  53 033 King                      Wa        41   13.49       61   11.40
  02 020 Anchorage Borough         Ak        21    6.91       41    7.66
  53 053 Pierce                    Wa        16    5.26       22    4.11
   Same State                              23    7.57       47    8.79
   Same Region, Diff. State               151   49.67      272   50.84
   Different Region                        52   17.11       92   17.20
      02 016   County Non-Migrants         991              2185
```

1993 - 1994 County to County Migration Inflow
(Aggregate money amounts are in thousands of dollars)

Migration into Alaska		Migration from			Number of returns	Number of exemptions	Aggregate total money income
State	County	State	County	State	State totals, county totals, and county by county detail			
FIPS Code		FIPS Code					
02	013	00	001	Ak	Aleutians East Borough (Total Migrant)	102	150	2,214
02	013	63	020	XX	Same State	20	28	551
02	013	63	021	XX	Same Region, Diff. State	68	104	1,272
02	013	63	022	XX	Different Region	14	18	391
02	013	63	050	Ak	County Non-Migrant	483	1,101	14,210
02	016	00	001	Ak	Aleutians West	744	1,225	15,994
02	016	57	005	FR	APO / FPO Zip Code	57	75	1,133
02	016	53	033	Wa	King	37	56	980
02	016	02	020	Ak	Anchorage Borough	30	45	759
02	016	06	037	Ca	Los Angeles	29	46	612
02	016	06	073	Ca	San Diego	26	55	664
02	016	53	035	Wa	Kitsap	12	31	360
02	016	53	053	Wa	Pierce	10	12	169
02	016	63	010	XX	Same State	34	68	1,226
02	016	63	011	XX	Region 1: Northeast	38	56	636
02	016	63	012	XX	Region 2: Midwest	91	125	1,511
02	016	63	013	XX	Region 3: South	161	274	3,282
02	016	63	014	XX	Region 4: West	219	382	4,662
02	016	63	050	Ak	County Non-Migrant	1,549	3,516	61,953

1997-1998 County To County Migration Inflows — Alaska
(Aggregate money amounts in thousands of dollars)

Migration into Alaska		Migration from			Number of returns	Number of exemptions	Aggregate adjusted gross income
State	County	State	County	State	State totals, county totals, and county by county detail			
FIPS Code		FIPS Code					
02	000	96	000	AK	Total Mig - US & For	19,770	39,551	641,764
02	000	97	000	AK	Total Mig - US	19,155	38,117	625,316
02	000	97	001	AK	Total Mig - US Same St	6,199	12,156	233,357
02	000	97	003	AK	Total Mig - US Diff St	12,956	25,961	391,959
02	000	98	000	AK	Total Mig - Foreign	615	1,434	16,448
02	013	96	000	AK	Aleutians East Tot Mig-US & For	136	200	3,037
02	013	97	000	AK	Aleutians East Tot Mig-US	136	200	3,037
02	013	97	001	AK	Aleutians East Tot Mig-Same St	28	35	742
02	013	97	003	AK	Aleutians East Tot Mig-Diff St	108	165	2,296
02	013	02	013	AK	Aleutians East Non-Migrants	545	1,086	17,870

ALASKA INFLOW
Individual Income Tax Returns: County-to-County Migration Inflow for Selected Income Items, Calendar Years 2010-2011
[Money amounts are in thousands of dollars]

Destination into Alaska		Origin from				Number of returns	Number of exemptions	Aggregate adjusted gross income (AGI)
State Code	County Code	State Code	County Code	State	County Name	(1)	(2)	(3)
02	000	96	000	AK	Total Mig - US & For	23,752	45,381	1,005,432
02	000	97	000	AK	Total Mig - US	23,214	44,210	983,240
02	000	97	001	AK	Total Mig - US Same St	7,058	13,245	329,228
02	000	97	003	AK	Total Mig - US Diff St	16,156	30,965	654,011
02	000	98	000	AK	Total Mig - Foreign	538	1,171	22,192
02	013	96	000	AK	Aleutians East Tot Mig-US & For	116	194	3,475
02	013	97	000	AK	Aleutians East Tot Mig-US	116	194	3,475
02	013	97	001	AK	Aleutians East Tot Mig-Same St	26	53	887
02	013	97	003	AK	Aleutians East Tot Mig-Diff St	90	141	2,588
02	013	98	000	AK	Aleutians East Tot Mig-Foreign	d	d	d
02	013	02	013	AK	Aleutians East Non-Migrants	632	1,214	28,822

Figure 2: **Sample extracts from the raw IRS migration data.** Here are four sample raw data extracts for 1990, 1993, 1997, and 2010. Notice all four have different file formats, structures, and coding schemes.

5-digit FIPS code of 96000. Figure 2 shows some sample extracts of the raw IRS migration data for 1990, 1993, 1997, and 2010.

The differences described above and in Table 3 are only some of the differences that are of interest to the data we produce here. Total Migrants, ie FIPS 96000 for migration data after 1995, is also broken down into Total Mig - US (FIPS 97000), Total Mig - US Same State (FIPS 97001), Total Mig - US Diff St (FIPS 97003), and Total Mig - Foreign (FIPS 98000). The IRS did not code these migration flows in this manner for all years, and in some cases (such as Total Mig - Foreign) migration flows are not reported. For simplicity and data continuity purposes, we simply create a new origin/destination (FIPS 99999) that contains all unspecified migration flows. We do this by subtracting the number of enumerated migrants

Table 3: Select differences in the file formats, file organizations, naming, and treatment of various migration statistics.

Years	Data Format	File Organization	Sample File naming	Coding of non-migrants	Coding of Total Migrants
1990-1991	txt	Separate in/out migration	C9091alo.txt	Destination field reads 'County Non-Migrants'	Destination field reads 'Total Migration'
1992, 1994	xls		C9293Alo.xls	State code = 63, County code = 010	State code = 00, County code = 001
1993			co934alo.xls	State code = 63, County code = 050	
1995-2003			co956alor.xls	Origin FIPS = Destination FIPS	State code = 96, County = 000
2004-2006		Single folder	co0405ALo.xls		
2007-2008			co0708oAl.xls		
2009-2010			co0910oAL.xls		

(the migration flows with greater than 10 migrants) from the total number of migrants. This way, the sum of all enumerated migrants in our dataset will equal the total number of migrants in the IRS dataset. And the sum of all migrants and non-migrants for any origin in a given year should roughly approximate the county population estimate for the previous year.

The aggregation to FIPS 99999 is the only mathematical post-processing of the IRS data.

R Code

The R code used to produce these data is available in the **Supplementary Materials** and in an online repository[4]. The code makes use of multi-core processing to speed up computation time. There are three main sections in the code: A setup section; a data download section; and a data processing section. The final flat file, `county_migration_data.txt`, contains the # of exemptions and can be either downloaded at github or produced by running the R code.

Setup

The script `000-libraries.R` simply sets up the R workspace to facilitate the data processing. The appropriate R packages are downloaded and installed if the user does not already have these packages installed. The parallel computing environment is also set up as

[4] https://osf.io/wgcf3/?view_only=c5ba62fb4821421ea0621bfd0d723e61

`DetectCores() - 1` to ensure the computer has appropriate resources for other tasks. The script requires a single reference tab separated (tsv) file in this section and we load it into the local environment. `ref_state.tsv` contains FIPS code information for US states. we simply add an additional FIPS state code for 'unknown' and assign it FIPS state 99.

Data Download

The script `001-download_data.R` will download and unzip the migration data from the IRS' websites into a folder standardized format into subdirectory `MigData/`. The IRS data is in two primary formats: 1990-2003 and 2004-onward. The IRS includes eight files in their zip archives that contain no data (these are in years 1998, 1999, 2000, and 2001). We delete these files after downloading and unzipping them. If they are not deleted, they will cause the subsequent `for loops` to fail in the next section. These files do not contain any migration information, their names suggest they represent aggregation of migration flows (for example 'co990usi.xls' suggests county (co) years 1999-2000 (990) for US (us) in-migration (i)), and we are unsure exactly why the IRS included these files or their purpose.

Data Processing

The third and final section contains several `foreach` parallel processing loops to process the seven legacy formats into a common data format. These files are then row-bound using `rbindlist` and transformed into a 'short' data frame. **Table 2** demonstrates the general file layout. We process the in- and out-migration files separately and keep only unique dyadic in the final flat file.

References

Abel, Guy J. 2013. "Estimating global migration flow tables using place of birth data." *Demographic Research* 28:505–546.

Abel, Guy J. 2017. "Estimates of global bilateral migration flows by gender between 1960 and 2015." *International Migration Review* .

Abel, Guy J and Nikola Sander. 2014. "Quantifying global international migration flows." *Science* 343(6178):1520–1522.

Brown, Warren A. 2009. *A compass for understanding and using American Community Survey data: What researchers need to know*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Curtis, Katherine J, Elizabeth Fussell and Jack DeWaard. 2015. "Recovery migration after Hurricanes Katrina and Rita: Spatial concentration and intensification in the migration system." *Demography* 52(4):1269–1293.

DeWaard, Jack, Katherine J Curtis and Elizabeth Fussell. 2016. "Population recovery in New Orleans after Hurricane Katrina: exploring the potential role of stage migration in migration systems." *Population and environment* 37(4):449–463.

Engels, Richard A and Mary K Healy. 1981. "Measuring interstate migration flows: an origin—destination network based on internal revenue service records." *Environment and Planning A* 13(11):1345–1360.

Franklin, Rachel S and David A Plane. 2006. "Pandora's box: The potential and peril of migration data from the American Community Survey." *International Regional Science Review* 29(3):231–246.

Frey, William. 2009. "The great American migration slowdown." *Brookings Institution, Washington, DC* .

Gross, Emily. 2005. Internal revenue service area-to-area migration data: Strengths, limitations, and current trends. In *Proceedings of the Section on Government Statistics*. p. 2005.

Molloy, Raven, Christopher L Smith and Abigail Wozniak. 2011. "Internal migration in the United States." *Journal of Economic perspectives* 25(3):173–96.

Pierce, K. 2015. "SOI migration data. A new approach: Methodological improvements for SOIC's United States population migration data, calendar years 2011–2012." *Statistics of Income, Internal Revenue Service* .

Rogers, Andrei, Jani Little and James Raymer. 2010. *The indirect estimation of migration: Methods for dealing with irregular, inadequate, and missing data*. Vol. 26 Springer Science & Business Media.

Shumway, J Matthew and Samuel M Otterstrom. 2001. "Spatial patterns of migration and income change in the Mountain West: the dominance of service-based, amenity-rich counties." *The Professional Geographer* 53(4):492–502.

Sorichetta, Alessandro, Tom J Bird, Nick W Ruktanonchai, Elisabeth zu Erbach-Schoenberg, Carla Pezzulo, Natalia Tejedor, Ian C Waldock, Jason D Sadler, Andres J Garcia, Luigi Sedda et al. 2016. "Mapping internal connectivity through human migration in malaria endemic countries." *Scientific data* 3:160066.

Willekens, Frans, Douglas Massey, James Raymer and Cris Beauchemin. 2016. "International migration under the microscope." *Science* 352(6288):897–899.