Los Angeles Out Migration Analysis Notes

This goal of this project was to gain insight on various demographic and economic characteristics of former Los Angeles residents. The primary data source for this information was 1-year US census data from IPUMS (<https://usa.ipums.org/usa/index.shtml>).

After discussion on which geographic variables to use, the decision was made to filter by metropolitan area, since smaller geographic units did not yield enough data or were not available for all years covered in this analysis (2006-2017). City or county would have been more ideal geographic units than metro area had these limitations not existed.

The MET2013 variable indicates the current metro area of residence while MIGMET1 & MIGPUMA1 indicate the previous metro of residence. After comparing the geographic variables, it was discovered that MET2013 essentially categorizes Los Angeles metro residents as anyone residing in either Los Angeles or Orange counties. Therefore, this same definition was used in filtering for previous metro area of residence via MIGMET1 & MIGPUMA1 (i.e. anyone previously residing in LA or Orange counties).

\*The complete data filtering can be found in the following notebook: <https://github.com/CityOfLosAngeles/Data-Angels/blob/master/Fall%202018%20Projects/LA_Outmigration/Code/data_cleaning_summary.ipynb>

\*\*The cleaned LA out migrators data (*ipums\_clean.csv*) is stored in the Data folder

**Geographic Variables:**

1. **MET2013** (<https://usa.ipums.org/usa-action/variables/MET2013#description_section>):

\*Indicates the *metropolitan area* an individual *currently* resides in.

- The data were filtered to exclude missing metro data (indicated by the value: 0) and records of *current* LA metro residents (indicated by the value: 31080).

1. **MIGMET1** (<https://usa.ipums.org/usa-action/variables/MIGMET1#description_section>)

\*Indicates the *county* an individual resided in during the previous year; this variable was only available for 2006-2011 data due to a different census methodology being implemented in 2012.

- The data were filtered to include records of individuals who previously resided in Los Angeles or Orange counties (indicated by the values: 4480 & 4482, respectively).

1. **MIGPUMA1** (<https://usa.ipums.org/usa-action/variables/MIGPUMA1#description_section>):

\*Indicates the *county* an individual resided in during the previous year; this variable was used for data from 2012 onwards, after a new census methodology was implemented.

- The data were filtered to include records of individuals who previously resided in Los Angeles or Orange counties (indicated by the values: 3700 & 5900, respectively). This variable is required to use in conjunction with MIGPLAC1, since the values in this variable are not unique to each state.

1. **MIGPLAC1** (<https://usa.ipums.org/usa-action/variables/MIGPLAC1#description_section>):

\*Indicates the *state* an individual resided in during the previous year.

- The data were filtered to only include records of individuals who resided in California during the previous year (indicated by the value: 06).

**Other Variables:**

1. **PERWT** (<https://usa.ipums.org/usa-action/variables/PERWT#description_section>):

\*Indicates the number of individuals a record represents.

- This variable is required to produce accurate estimates of the number of people represented in the data (refer to the jupyter notebooks in the visuals folder for a demonstration on how to use it, such as cell #6 in this one: <https://github.com/CityOfLosAngeles/Data-Angels/blob/master/Fall%202018%20Projects/LA_Outmigration/Visuals/Education.ipynb>)

**Data Quality:**

* Less than 20% of resident and out migrator records were missing individual income (INCTOT) values.

\*About 29% of resident records were missing family income (FTOTINC) values.

* About 17% of resident records and 29% of out migrator records were missing rent & mortgage values (i.e. 83% and 71% of resident and out migrator records, respectively, contained either rent or mortgage values).

\*About 35% for records of out migrators with no family income were missing values.

* The total number of LA metro residents represented in the data (through applying PERWT) is consistent with commonly reported figures (around 13 million).

**Miscellaneous Notes:**

* PERWT was not applied in charts highlighting economic variables (namely: income, rent, & mortgage); the reason for this is that the values returned tended to be higher than commonly reported LA city or county values.

\*One hypothesis is that since LA *metro* includes Orange county (which has a higher median income than LA county), this would cause the LA metro economic variables to return higher values than just LA county.

* Some variables contained relatively few records with certain values, which resulted in skewed data (e.g. few records contained a RACE value of “Mixed,” which returned an abnormally high rent-to-income percentage; consequently, “Mixed” was excluded from the rent-to-income percentage chart)
* Detailed geographic data is available in the PUMA variable for migration destinations (though it is not available for the origin of migration).
* RACE values were manipulated to combine several similar values under one description (e.g. “Chinese” & “Japanese” were merged under the value: “Asian”). Also, since Hispanic origin was documented under the HISPAN variable (and not in RACE), it was merged with RACE so all records reflect a single race.

\*The details of this merging are outlined in the data cleaning notebook (in the Data folder)