PUMS Personal Project: Los Angeles & Migration

ARIEGE BESSON

May 6, 2025

Project Description

An exploration of migration into & away from Los Angeles, using PUMS variables from the 2023 ACS. Where are people coming from (and moving to), and are there patterns related to income? These are really just some of the first questions that come to mind. No real purpose here except to explore PUMS data relating to moving and migration.

- Citation: Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C.W. Williams. IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS, 2025. https://doi.org/10.18128/D010.V16.0

Some Questions

Coming To Los Angeles

- What is the median income of people coming to LA?
- From what countries are the most people coming from? And from what states and counties?
- Looking at the wealthiest movers by percentile: where are they from?
- What is the median income of Venezuelans (or any particular national origin) who came to LA in x year? How does it vary across years?
- How does median income for international Latino migrants vary by the number of years they have been in the US? How does it vary across national origin?

Leaving Los Angeles

- Similar to the questions above above: what is the median income of people leaving LA? What are the top destination counties and states for people leaving? Countries? and do these patterns differ for the wealthiest people leaving LA?
- On that theme: what is the average education of people leaving Los Angeles and how does it compare to the education levels of Californians overall?

Part 1: Setup

- load packages

- run code to clear the environment

```
rm(list = ls())
```

- set your working directory

  -or not, because in .Rmd better to use here() as knitr controls the working directory.

- call in data: select variables from the 2023 ACS

```
ddi <- read_ipums_ddi(here("usa_00005.xml"))
data <- read_ipums_micro(ddi)
```

```
## Use of data from IPUMS USA is subject to conditions including that users should cite the data approp
```

Part 2: Take A Look

To start: exploratory things, see/note type of each variable

```
View(data)
names(data)
str(data)
```

Part 3: Make 2 datasets, one for people coming to LA and one for people leaving LA

Coming

To find people moving to Los Angeles, we will filter for respondents in LA (use the county FIPS code (06037)) whose value for the variable MIGRATE1 is 2, 3 or 4.

```
coming <- data %>%
  filter(STATEFIP == 6,
          COUNTYFIP == 037,
          MIGRATE1 %in% c(2, 3, 4))
```

Leaving

To find people who left Los Angeles, we will filter for respondents for whom MIGPLAC1 is 06 (California) and MIGCOUNTY1 is 037 (Los Angeles).

```
leaving <- data %>%
  filter(MIGPLAC1 == 6,
         MIGCOUNTY1 == 037)
```

Part 4: Find the median incomes of people coming and leaving LA

- I'll save these as objects/use pull() to be able to center graphs around medians later
- notes on INCTOT: for now, I will restrict this variable to positive numbers, filter out topcodes, and use only personal income (there's also a variable for total family income).

Coming

```r
medc <- coming %>%
  filter(
    INCTOT > 0,
    INCTOT != 9999999) %>%
  summarise(
    medianincome = round(wtd.quantile(INCTOT, weights =PERWT, probs = 0.5, na.rm = TRUE), 0)
  ) %>%
  pull() %>%
  print()
```

```
##   50%
## 38500
```

Leaving

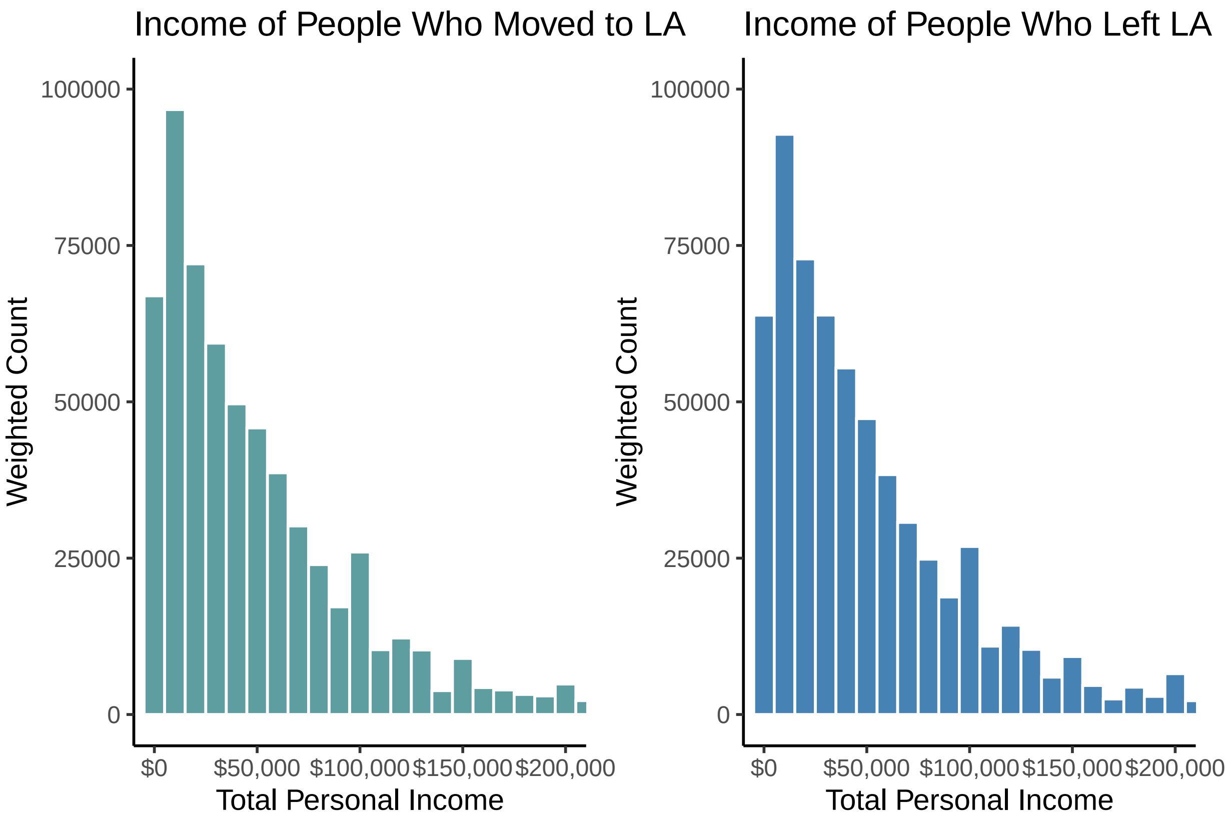
```r
medl <- leaving %>%
  filter(
    INCTOT > 0,
    INCTOT != 9999999) %>%
  summarise(
    medianincome = round(wtd.quantile(INCTOT, weights =PERWT, probs = 0.5, na.rm = TRUE), 0)
  ) %>%
  pull() %>%
  print()
```

```
##   50%
## 40000
```

So, the median income of people coming and going differs by less than $2,000. To get a better sense of the distribution of incomes of people coming and going, I'll make histograms.

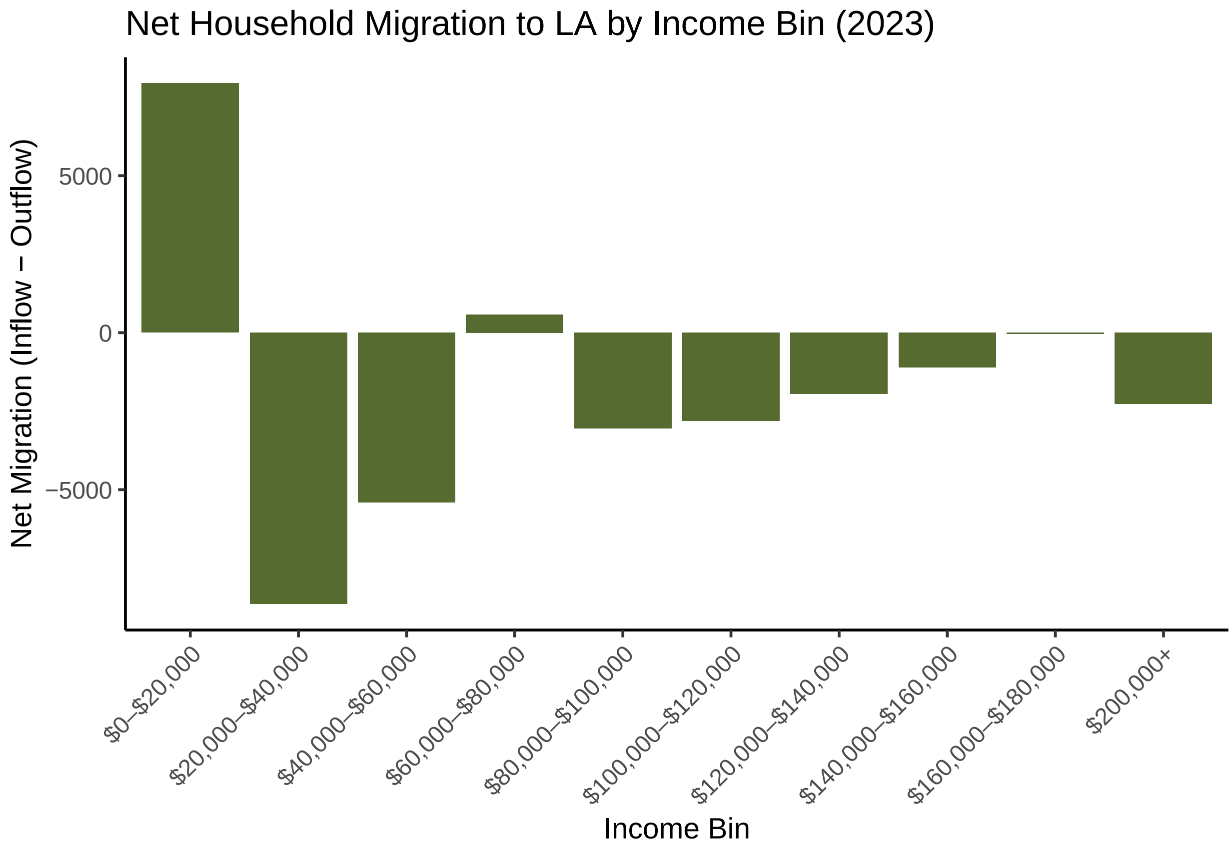
Part 5: Make and compare histograms of median income for people coming and leaving LA

- Note: using the person-level variable weight here, PERWT, will result in replicate-weighted counts for the histogram, which I want

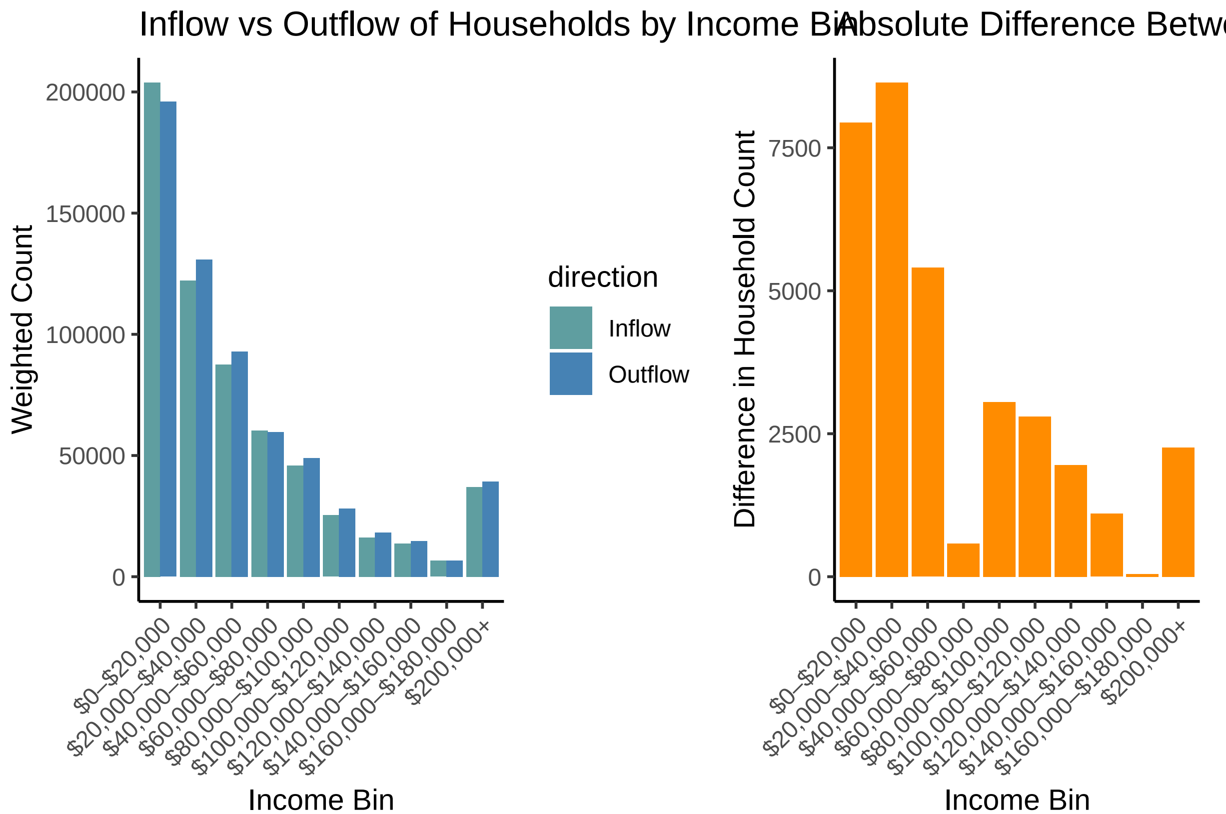


Because the patterns between people coming and people leaving are similar, across income bins, let's try another way to see the difference.

First I'll look at the net difference in aggregate households, acrors income bins. This will show us the net migration effect in Los Angeles in 2023 per income bin. Or in other words, how much each income group grow or shrink due to migration?



Second, I'll try a line segment chart that is plotted to show the total inflow and outflow households by income bin. Here, the focus is on movement intensity and symmetry. In other words, how many households came and went for each income group, regardless of net change? Ok so line segment doesn't show much because difference is small. Instead, here are two plots to compare inflow and outflow by income bins.



Part 6: Find the top countries, states and counties that people are coming to LA from

- Of course…and this caveat goes for anything calculated using ACS data, there are limitations to census data. Census workers don't get to everyone and immigrant populations especially have had some credible fear of participating in the census since at least 2016.
- Here I'll cut to just the top 20 locations in each category.

Countries

```
#filter
comingcountry <- coming %>%
  filter(MIGPLAC1 >= 100) %>%
  count(MIGPLAC1, sort = TRUE) %>%
  slice_head(n = 20)

#bring in key
countrykey <- read.csv(file.path("MIGPLAC1COUNTRYLOOKUP.csv"))

#join:
comingcountry <- left_join(comingcountry, countrykey, by = c("MIGPLAC1" = "Value"))

#add in a percent total column
comingcountry <- comingcountry %>%
  mutate(percent_total = round((n/sum(n)), 4))
```

Top 20 Countries of Origin

```
comingcountry
```

```
## # A tibble: 20 x 4
##    MIGPLAC1                                                 n Label   percent_total
##    <int+lbl>                                            <int> <chr>           <dbl>
##  1 200 [Mexico]                                           106 Mexico         0.170
##  2 500 [China]                                            103 China          0.166
##  3 521 [India]                                             35 India          0.0563
##  4 551 [Other Western Asia]                                31 Other ~        0.0498
##  5 599 [Asia, nec]                                         31 Asia, ~        0.0498
##  6 214 [Guatemala]                                         29 Guatem~        0.0466
##  7 413 [United Kingdom (excluding England: 2005ACS)]       27 United~        0.0434
##  8 503 [Taiwan]                                            27 Taiwan         0.0434
##  9 515 [Philippines]                                       27 Philip~        0.0434
## 10 150 [Canada]                                            26 Canada         0.0418
## 11 219 [Central America, nec]                              22 Centra~        0.0354
## 12 325 [Colombia]                                          22 Colomb~        0.0354
## 13 465 [USSR]                                              22 USSR           0.0354
## 14 699 [Africa, nec]                                       21 Africa~        0.0338
## 15 213 [El Salvador]                                       20 El Sal~        0.0322
## 16 501 [Japan]                                             16 Japan          0.0257
## 17 519 [Other South East Asia]                             16 Other ~        0.0257
## 18 453 [Germany]                                           14 Germany        0.0225
## 19 502 [Korea]                                             14 Korea          0.0225
## 20 350 [Peru]                                              13 Peru           0.0209
```

States & Counties

```
#filter
comingcounty <- coming %>%
  filter(MIGPLAC1 <= 100) %>%
  mutate(MIGFIPS = str_pad(MIGPLAC1, width = 2, pad = "0") %>%
         paste0(str_pad(MIGCOUNTY1, width = 3, pad = "0"))) %>%
  count(MIGFIPS, sort = TRUE) %>%
  slice_head(n = 20)

#bring in keys
cacountykey <- read.delim("californiacountykey.txt", sep = "|",
                          header = TRUE, stringsAsFactors = FALSE)
othercounty <- read.csv(file.path("xtrafipskey.csv"))

#format fips
cacountykey <- cacountykey %>%
  mutate(FIPS = str_pad(STATEFP, width = 2, pad = "0") %>%
         paste0(str_pad(COUNTYFP, width = 3, pad = "0")))
othercounty <- othercounty %>%
  mutate(FIPS = str_pad(FIPS, width = 5, pad = "0"))

#bind rows cacounty and othercounty
countykey <- bind_rows(cacountykey, othercounty)
```

```r
#join
comingcounty <- comingcounty %>%
  left_join(
    countykey %>% select(FIPS, STATE, COUNTYNAME),
    by = c("MIGFIPS" = "FIPS")
  )

#add in a percent total column
comingcounty <- comingcounty %>%
  mutate(percent_total = round((n/sum(n)), 4))
```

Top 20 Counties of Origin

```r
comingcounty
```

```
## # A tibble: 20 x 5
##    MIGFIPS      n STATE COUNTYNAME                                  percent_total
##    <chr>    <int> <chr> <chr>                                               <dbl>
##  1 06037     5864 CA    Los Angeles County                                 0.835
##  2 06059      256 CA    Orange County                                      0.0364
##  3 06071      145 CA    San Bernardino County                              0.0206
##  4 06073       95 CA    San Diego County                                   0.0135
##  5 06065       74 CA    Riverside County                                   0.0105
##  6 06085       67 CA    Santa Clara County                                 0.0095
##  7 06111       58 CA    Ventura County                                     0.0083
##  8 06001       53 CA    Alameda County                                     0.0075
##  9 06029       44 CA    Kern County                                        0.0063
## 10 04013       42 AZ    Maricopa County                                    0.006
## 11 06000       38 CA    000 = Not in universe, or county not ident~        0.0054
## 12 36061       38 NY    New York County                                    0.0054
## 13 17031       37 IL    Cook County                                        0.0053
## 14 06041       35 CA    Marin County                                       0.005
## 15 06013       34 CA    Contra Costa County                                0.0048
## 16 06075       34 CA    San Francisco County                               0.0048
## 17 36047       29 NY    Kings County                                       0.0041
## 18 08000       28 CO    000 = Not in universe, or county not ident~        0.004
## 19 32003       27 NV    Clark County                                       0.0038
## 20 53033       27 WA    King County                                        0.0038
```

Part 7: 2023 median income for California immigrant Latinos by number of years in the US

I'll skip ahead to an interesting question (for me). How does median income for international Latino migrants in 2023 vary by the number of years that they have been here? Does median income increase the longer someone has been in the states? And how about national origin– how does median income for international Latino migratns in 2023 vary by their national origin?

- This is likely better answered by looking at a larger geography than Los Angeles... we'll start with the state of California.

What variables to use?

- First: how to identify international migrants?
  - The variable YRSUSA1 (discussed below) automatically filters for international migrants: YRSUSA1 "reports how long a person who was born in a foreign country or U.S. outlying area had been living in the United States" so we are all good
- Second: how to identify Latinos?
  - Racial and ethnic identity variables include RACE, ANCESTR1 (ancestry) and HISPAN (hispanic origin). Hispanic origin takes into account Spanish ancestry but oh well. HISPAND can be used for more detailed national origin. I'll check first that the same population has responses for HISPAN and HISPAND:

```
data %>% filter(!(HISPAN %in% c(0, 9))) %>% sum()
data %>% filter(!(HISPAND %in% c(000, 900))) %>% sum()
```

- Third: how to find the number of years since immigration?
  - Use YRSUSA1, cut out values of 00 (N/A or less than one year). Other option: YRIMM
- Fourth: the variable to measure income
  - Use INCTOT, and as previously restrict to positive numbers, filter out 9999999

Creating the table

```
incbyyrs <- data %>%
  filter(STATEFIP == 6,
         YRSUSA1 >= 1,
         !(HISPAND %in% c(000, 900)),
         INCTOT > 0,
         INCTOT != 9999999
  ) %>%
  group_by(YRSUSA1) %>%
  summarise(median_income = wtd.quantile(INCTOT, weights = PERWT, probs = 0.5, na.rm = TRUE),
            mean_income = weighted.mean(INCTOT, w = PERWT, na.rm = TRUE),
            n_ppl = n(),
            n_ppl_weighted = sum(PERWT, na.rm = TRUE)
            )
```

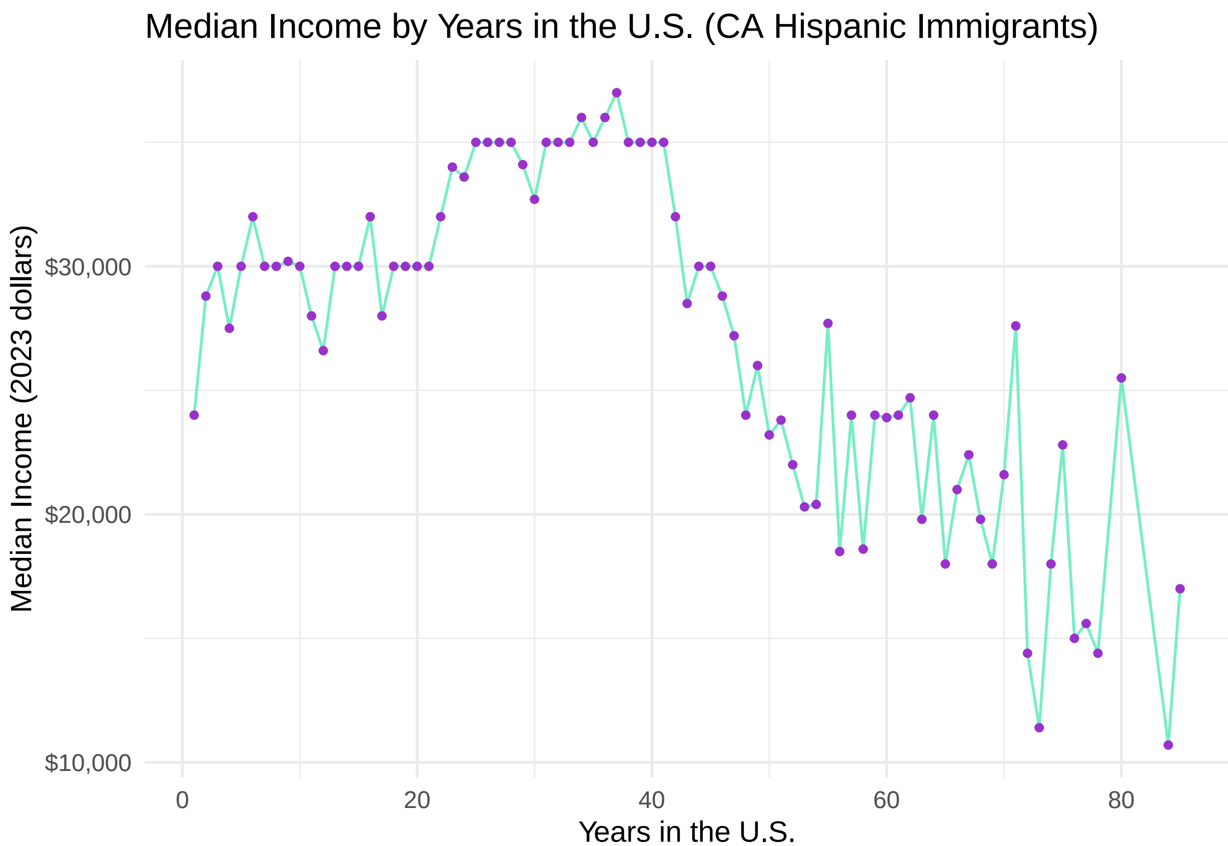
Table Preview

```
head(incbyyrs, 10)
```

```
## # A tibble: 10 x 5
##    YRSUSA1   median_income mean_income n_ppl n_ppl_weighted
##    <int+lbl>         <dbl>       <dbl> <int>          <dbl>
##  1  1                24000      30718.   487          62649
##  2  2                28800      35065.   429          57929
##  3  3                30000      35136.   344          43199
##  4  4                27500      32124.   382          47375
##  5  5                30000      39914.   383          48971
##  6  6                32000      41996.   361          43853
##  7  7                30000      37767.   397          48219
##  8  8                30000      39122.   439          50543
##  9  9                30200      40930.   309          36263
## 10 10                30000      37181.   362          44072
```

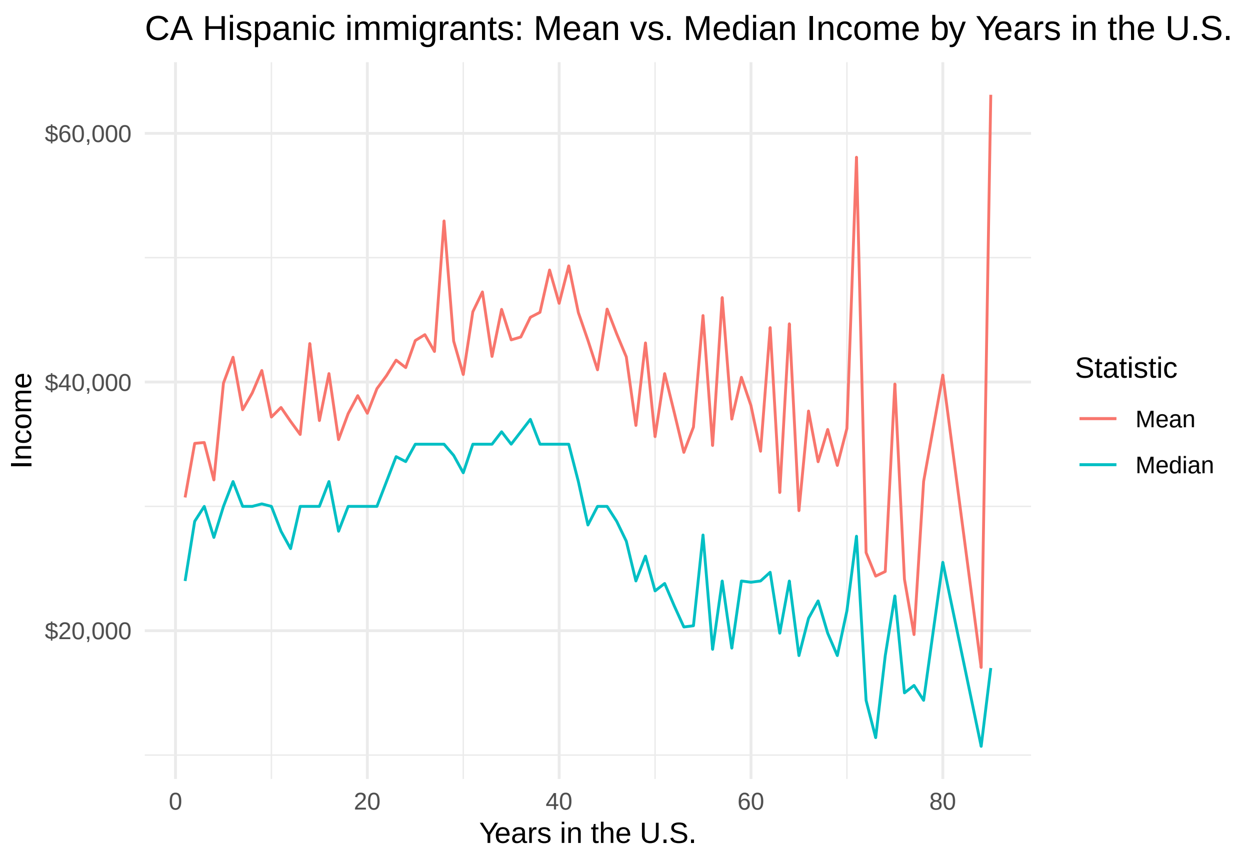
Graphing the results

- I'll try a line plot to start



Looking at this graph, we can see that the median income for CA Latino immigrants does tend to increase with the number of years since coming the US, at least until about 40 years presence in the US, where the median drops off. Generally speaking, Latino immigrants who have been in the US for 20-40 years have the highest median incomes. The drop in later years could also be due to age; older adults may not be working and may have less personal income.

- And here is a visual comparison of the median and mean values.



As expected, the mean varies more, is higher than median (pulled by high outliers) but follows the same shape as the median across the Y axis of years in the US.

Part 8: 2023 median income for California immigrant Latinos by national origin

Creating a table

```
#make the table
incbynat <- data %>%
  filter(STATEFIP == 6,
         YRSUSA1 >= 1,
         !(HISPAND %in% c(000, 900)),
         INCTOT > 0,
         INCTOT != 9999999
  ) %>%
  group_by(HISPAND) %>%
  summarise(median_income = wtd.quantile(INCTOT, weights = PERWT, probs = 0.5, na.rm = TRUE),
            mean_income = weighted.mean(INCTOT, w = PERWT, na.rm = TRUE),
            n_ppl = n(),
            n_ppl_weighted = sum(PERWT, na.rm = TRUE)
  ) %>%
  arrange(desc(median_income))

#add key to see countries of origin named
hispand_key <- read.csv(file.path("pums_hispand_key.csv"))
incbynat <- incbynat %>% left_join(hispand_key, by = c("HISPAND" = "Value"))
```

Here's the table

```r
head(incbynat, 20)
```

```
## # A tibble: 20 x 6
##    HISPAND                      median_income mean_income n_ppl n_ppl_weighted Label
##    <int+lbl>                            <dbl>       <dbl> <int>          <dbl> <chr>
##  1 425 [Paraguayan]                    113000      96500.    11           1064 Para~
##  2 431 [South American, n.~            103000     117924.     7            528 Sout~
##  3 420 [Argentinean]                    60000      89588.   247          23716 Arge~
##  4 421 [Bolivian]                       50000      67572.    95           9745 Boli~
##  5 428 [Venezuelan]                     45000      81130.   158          16745 Vene~
##  6 450 [Spaniard]                       45000      78748.   233          20557 Span~
##  7 427 [Uruguayan]                      44800      87694.    31           3551 Urug~
##  8 422 [Chilean]                        42000      71337.   131          13142 Chil~
##  9 415 [Panamanian]                     37000      55878.    78           8347 Pana~
## 10 300 [Cuban]                          36500      61300.   285          24960 Cuban
## 11 424 [Ecuadorian]                     35000      54050.   222          22621 Ecua~
## 12 426 [Peruvian]                       35000      56190.   531          54938 Peru~
## 13 414 [Nicaraguan]                     33500      45712.   482          56432 Nica~
## 14 423 [Colombian]                      32000      50251.   563          59961 Colo~
## 15 460 [Dominican]                      32000      58423.    70           8076 Domi~
## 16 200 [Puerto Rican]                   31200      50984.   352          35719 Puer~
## 17 498 [Other, n.s.]                    30800      48155.   816          72011 Othe~
## 18 100 [Mexican]                        30000      39875. 28469        3156824 Mexi~
## 19 412 [Guatemalan]                     30000      37444.  1984         238291 Guat~
## 20 413 [Honduran]                       30000      33946.   449          54041 Hond~
```

In reference to the table, Paraguayan immigrants are making significantly more money on average than other Latino immigrant groups. However, the sample size is small: My sample includes 11 Paraguayan immigrant respondents in California.

and that's all for now!