Californian Internal Emigration

Cooper Komatsu

Movement Data and Analysis (GISC 27104)

Dr. Crystal Bae

Autumn 2023

Introduction

Problem Statement: In this paper, I aim to address the nature of domestic out-migration from California. While California remains a desirable place to live, and international immigrants continue to arrive, its population has decreased in both 2021 and 2022. Empirically, many Californians, specifically in urban areas, have moved elsewhere because they have been priced out of the state due to its high housing prices and overall cost of living[1] [2]. On the other side of the trajectory, non-Californians (e.g. Arizonans, North Carolinians) have noticed Californians moving into their neighborhoods, buying up property, and driving rent and house prices up[3] [4]. These trends, whether real or imagined, seem to impact people's impressions of Californians, even stoking resentment for them. Many of the places where Californians are moving *en masse* (Phoenix, Dallas) do not have rent control in place, which could make their impact especially harmful to current residents[5]. They may, in turn, be priced out, causing further domestic migration. I seek to uncover the underlying patterns through GIS and statistical analysis.

---

1 https://investors.redfin.com/news-events/press-releases/detail/841/redfin-reports-homebuyers-are-looking-to-relocate-to

2 https://www.latimes.com/california/story/2023-11-06/we-fit-in-a-lot-better-here-californians-who-flee-to-texas-find-good-company-camaraderie

3 https://www.marketplace.org/2022/12/05/she-left-california-for-arizona-but-found-high-rent-and-lower-wages/

4 https://time.com/6170497/phoenix-fastest-growing-home-prices/

5 https://www.wcnc.com/article/money/markets/real-estate/charlotteans-feeling-high-rent-increases-nc-lawmaker-renters-relief/275-567588d2-10f0-46d7-938c-bdbc4936b0b9

Research Question: To what extent are the arrivals of Californian migrants to other states associated with house price increases? Which places do urban Californians from different counties tend to move to when they do move?

Hypotheses: Overall, I believe cities in the Southwest and Southeast will be regions with the largest immigration from California. Counties like Denton, Maricopa, and Mecklenburg will be exemplary of this trend. While Californian immigration will almost certainly not be the sole cause (if a *cause* at all), I hypothesize that it will be *correlated* with larger house price increases, even when compared to immigration in general.

Literature Review

Dietz, J., Li, B., & Castañeda, E. (2023, July 8). *Keeping in motion or staying put: Internal migration in the United States and China*. MDPI. https://www.mdpi.com/2075-4698/13/7/162

As the title suggests, this paper is a comparative study on the push and pull factors affecting migration choices in the US and in China. China has seen many of its rural citizens, especially young people, flock to cities, while the US has been relatively stable, but is characterized more by the opposite. They discuss a similar concept to the "gravity" that Gunderson and Sanderson pick up on in their "sticky" and "magnetic" notions. However, this analysis is done on a state scale, which doesn't seem nearly granular enough to me to produce meaningful data on this demographic movement that I want to analyze. Internal migration in China provides a fascinating alternative model to broaden my view of what the phenomenon looks like.

Ellis, M. (2011, March 14). *Reinventing US Internal Migration Studies in the Age of International Migration*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/psp.665

Ellis focuses on the relationship between international and internal migration. Much of what people hear on the news relates to international migration, but Ellis urges focus on internal migration as another potent source of information, with as much, if not more, related factors and demographic trends that could fascinate urbanity.

Gunderson, R. J., & Sorenson, D. J. (2010). *An examination of domestic migration from California counties*. AgEcon Search. https://ageconsearch.umn.edu/record/132439/

This is perhaps the most directly relevant of the three papers to my research question. Gunderson and Sorenson remark on the trend of lack of internal migration to California since 1990. Focused on the period 2006-2007, the authors seek to explain why Californians are trending toward moving out of the state, and find out where they are going. They note both a significant movement of Californian city dwellers out to the suburbs of that city, and to other cities. A shrinking job market, high cost of living, and lack of preferred amenities were factors in this. They also discovered differences between Northern and Southern California (in that Southern California was bleeding more people at the time), something I'd like to look into further. They have a map of outmigration to all US counties that could serve as inspiration to me (although the maps are simply choropleths, with little explicit allusion to movement). Lastly, they develop a notion of "gravity" toward migrants that they analyze statistically at the county scale.

Rees, Philip, and Nik Lomax. "Ravenstein revisited: The analysis of migration, then and now." *Comparative Population Studies*, vol. 44, 2020, https://doi.org/10.12765/cpos-2020-10.

In this paper, Rees and Lomax review the works of Ernst Ravenstein, a 19th-century geographer and demographer whose work is foundational to the study of migration patterns. He set out some basic, broad laws of human migration and spatial movement and discussed the reasons why people might want to move. The authors argue set forth

a definition of human migration (which is based around housing and households), then survey the ways in which migration data is used today, with much greater access to knowledge of people's movement. This paper is most applicable to my work in that it provides a broad framework for thinking about migration, especially internal migration. The authors argue that "internal migration measurement is particularly sensitive to the number, size and shapes of regions," which tells me I will have to be very intentional in choosing which counties I analyze, and noting the limitations of the geographies I work with (357).

Data, Study Area, and Methods

Data: The primary dataset I used in this analysis was county-to-county internal migration data in the four-year period from 2016 to 2020. For each county in a representative year aggregated over the period 2016-2020, it displays the number of people who didn't move out, moved out to another home in the county, moved out to a different county in the state, and moved out to a different state; likewise for people who moved in. Most importantly, it displays average yearly county-to-county flows for any two US counties. For smaller counties with smaller flows, the margins of error are quite high, which sometimes poses problems in analysis. These could amount to large outliers, especially when examining the ratio between one variable and another. Because of these concerns, for the most part, I excluded flows for which the margin of error was greater than the value of the flow itself—in all, 81% of flows (197889/243918). I also used the Federal Housing Finance Agency's house price index, which uses mortgage data to aggregate county-level values that estimate how expensive single-family housing has been since 1975, relative to that 1975 baseline. While single-family housing is by no means a perfect analog for housing in general, papers support[6] that house prices and rent tend to covary significantly. Thus, I decided to use it as an imperfect proxy for the increase in the cost of housing for residents of a county. One of my peers suggested that I use housing burden instead of house price to do this analysis (to do this, I might divide house price by median income), but I reasoned that, if immigration to counties by

[6] https://www.federalreserve.gov/econres/feds/files/2020044pap.pdf

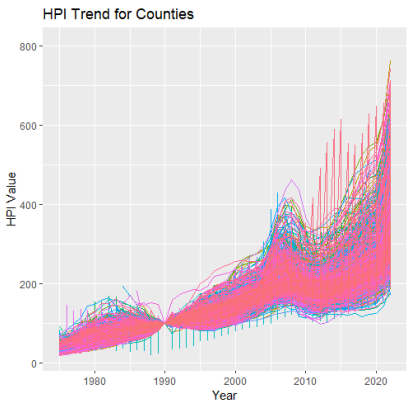
people with higher median income was indeed a factor in house price, then increases in median income would reflect the same trend and cancel itself out. Lastly, I used shapefiles from the US Census for counties and county-equivalents.

Study Area: For this study, I concentrated on the United States from 2016-2020. Perhaps this trend had been going on for years, but it seems to be especially salient nowadays: between California's first year-over-year population decrease in its history (2021), first decrease in electoral votes (2020), and the spate of recent coverage about Californians moving to other Sun Belt. So, I decided that the most recent frame of reference (2016-2020 for the Census county-to-county migration data) would be the most important for my analysis. When defining "urban California," I chose the 10 counties of Los Angeles, San Diego, Orange, Riverside, San Bernardino, Santa Clara, Alameda, Sacramento, Contra Costa, and San Francisco. The first nine are the nine most populous counties in California, all with populations over 1 million; San Francisco County is a small but densely populated and heavily urbanized county, so I thought it would be worth it to include. The next largest counties were the Central Valley counties of Fresno and Kern (home to Bakersfield), which, while significant cities, do not have the reputation for high home prices that the other counties do, so I decided to draw the line there.

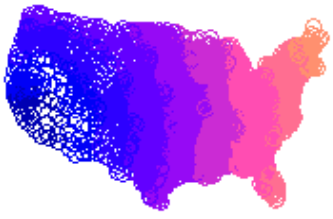
Methods: I primarily worked in R to wrangle and analyze the data for this project. The Census County-to-County Flows data comes in Excel files, but I imported them into R to

work with them. I primarily used packages *sf*, *dplyr*, *stats*, and *ggplot2*. After reading in and concatenating the sheets from the Excel file, I calculated the percent of total interstate inflow that came from each individual county-to-county flow. Then, I loaded in the datasets for House Price Index (HPI) and counties for later joining with the migration data. I calculated the 2016-2020 percent increase (change) in HPI for the former, and reformatted the counties to be usable for spatial joins with the rest of the data.

To check that the HPI data loaded in well, I plotted it to make sure it looked right. Sure enough, HPI (which I set with a baseline at 1990) trended upward over time. In 2008, HPI fell due to the Great Recession, but it has since rebounded. It seems that HPI doubles every 20 years or so based on the visible trends.



Then, I used a variety of dplyr tools to reformat the County-to-County flows into total migration from all 10 counties, subset them for each of the 10 counties, and join each of the datasets with the HPI data for later regressions. Based on results that I will mention later, I found it necessary to calculate the distance from each of the 10 urban California counties to the corresponding county for each immigration flow, in part because of my results and in part because of Ravenstein's first law of migration, which states that people tend to migrate shorter than longer distances.. I used the centroids of each county in the *counties* dataset, and then used *st_distance* to



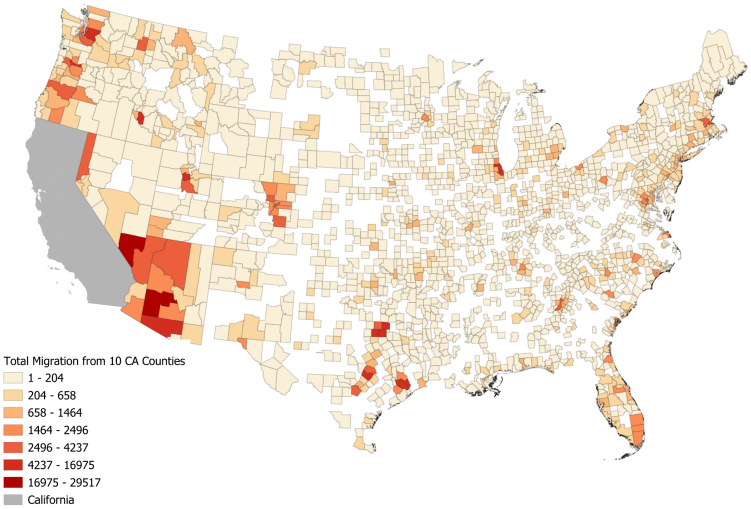
calculate the distance between each pairwise county flow. To confirm the distance calculation worked, I plotted the distance of all counties to the centroid of Los Angeles County, and it seemed to be correct.

Once I had the flows, I began performing the regressions using the *lm* function, deciding on the most instructive regressions to perform in a recursive process. Results of these regressions are below.

Finally, to analyze the clusters of emigration, I

Results

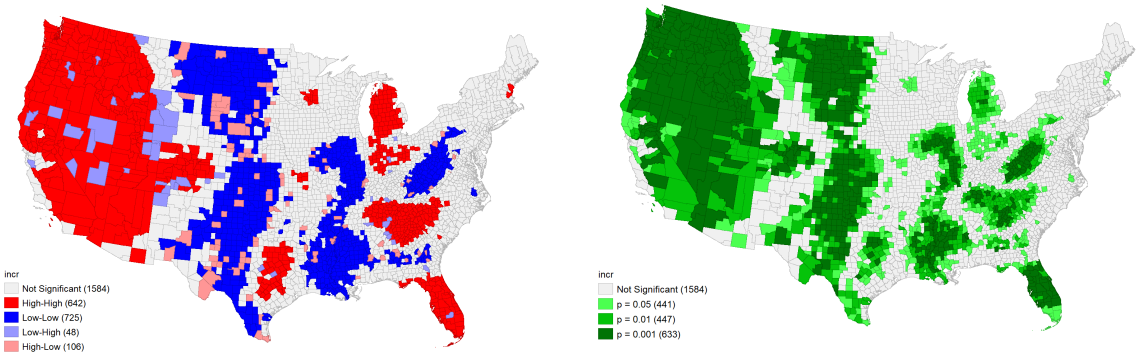
The first thing I did was to plot where the emigrations from the 10 counties in question were going to. This was available overall on the US Census website already, but I combined the 10 counties of interest to get the following map:



From the map, it seems that Californians tend to migrate more to populous counties with large cities, although that is perhaps a given, since more people live there overall. The Phoenix, Dallas, Las Vegas, Seattle, Atlanta, and Miami metropolitan areas seem to have particularly large flows. And other counties in the West that are not particularly large (Washoe, NV; Mohave, AZ; Lane, OR; Ada, ID) seem to have large flows into them. But, I felt I had to dig deeper to fully understand these patterns.

For looking at the HPI correlation, I got a suggestion from a peer to use spatial cluster analysis on this data, but unfortunately the flow data was spotty enough, and the differences in sizes and shapes were large enough, that cluster analysis (especially when based on contiguity-based weights systems) did not work very well. However, I *was* able

to use spatial cluster analysis just for the HPI increase/decrease data. Maps are shown below using Univariate Local Moran's I with queen-2 contiguity. Broadly, we are able to see that the West (aside from the California coast), the Dallas-to-San Antonio corridor, the Florida Panhandle, the Piedmont region (Atlanta, Nashville, Research Triangle), and Michigan's Lower Peninsula had large, significant high-HPI-increase clusters at the $p<0.01$ level. On the other hand, the Great Plains, the Mississippi River Valley south of Minnesota had significant low-HPI-increase clusters at the $p<0.01$ level. The first cluster, in the West, is concerning with regard to the fact that these are the closest regions of the US to California, and thus the likeliest for Californians to move to. Overall, a Moran's I value of 0.512 suggests moderately high spatial clustering.



One concern was that of regional effects—maybe Western states are independently likely to have high housing price increases, and that these were also the counties urban Californians were likely to move to, since they were closest. One method I had for controlling for regional effects was to do multivariate regression in R.

My initial regression analysis centered around running regressions between variables of county-to-county flow and house price index increase. I started by

examining each of the 10 California counties of study individually. I wanted to isolate the effects of just *those counties* and their emigrants. So, rather than just the raw magnitudes of the flows, I calculated the percent of interstate migrants to a given county who came from the applicable California county. For example, if Cook County had a yearly flow of 10,000 out-of-state migrants, and 500 of them came from Los Angeles County, the measure would be 5%. I will call this measure *California county percent* from now on. I used just interstate migrations, as intra-state or intra-county migrations would overwhelm incomers from any specific county, and because the phenomenon I want to examine involves "outsiders" coming in anyway. In the same spirit, I excluded flows from the urban California counties to other California counties[7].

The results of this regression between California county percent and 2016 Housing Price Index increase showed some significance, p-values are quite low (<0.00001), while R-squared values are still fairly low (>0.05), but not negligible. Full results are displayed below.

County	**Regression: incr* ~ ctypct****	**Regression: incr ~ dist*****	**Regression: incr ~ ctypct + dist**	**Regression: ctypct ~ dist**
Los Angeles	$R^2 = 0.0662$ $p < 10^{-5}$ ***	$R^2 = 0.2578$ $p < 10^{-15}$ ***	$R^2 = 0.2622$ $p < 10^{-15}$ ***	$R^2 = 0.1441$ $p < 10^{-10}$ ***
San Diego	$R^2 = 0.0321$ p < 0.01 **	$R^2 = 0.2559$ $p < 10^{-15}$ ***	$R^2 = 0.2960$ $p < 10^{-15}$ ***	$R^2 = 0.0018$ p = 0.51
Orange	$R^2 = 0.0987$ p < 0.0005 ***	$R^2 = 0.3107$ $p < 10^{-12}$ ***	$R^2 = 0.3191$ $p < 10^{-11}$ ***	$R^2 = 0.1728$ $p < 10^{-6}$ ***
Riverside	$R^2 = 0.0001$ p = 0.95	$R^2 = 0.2613$ $p < 10^{-6}$ ***	$R^2 = 0.2932$ $p < 10^{-6}$ ***	$R^2 = 0.1023$ p < 0.005 **

[7] With California counties included, p-values for the correlations between urban county and House Price Index are very high (>0.9) and R-squared values are very low (<0.001), suggesting little correlation. Fellow California counties are outliers as destinations that may obscure patterns in the rest of the data.

San Bernardino	$R^2 = 0.0312$ $p = 0.11$	$R^2 = 0.2335$ $p < 10^{-5}$ ***	$R^2 = 0.2338$ $p < 10^{-4}$ ***	$R^2 = 0.1160$ $p < 0.005$ **
Alameda	$R^2 = 0.0037$ $p = 0.55$	$R^2 = 0.3207$ $p < 10^{-9}$ ***	$R^2 = 0.3663$ $p < 10^{-9}$ ***	$R^2 = 0.0677$ $p < 0.005$ **
Santa Clara	$R^2 = 0.0012$ $p = 0.72$	$R^2 = 0.3657$ $p < 10^{-11}$ ***	$R^2 = 0.3880$ $p < 10^{-10}$ ***	$R^2 = 0.0856$ $p < 0.005$ **
Sacramento	$R^2 = 0.0002$ $p = 0.91$	$R^2 = 0.3938$ $p < 10^{-6}$ ***	$R^2 = 0.3951$ $p < 10^{-5}$ ***	$R^2 = 0.0006$ $p = 0.86$
Contra Costa	$R^2 = 0.0212$ $p = 0.32$	$R^2 = 0.4067$ $p < 10^{-5}$ ***	$R^2 = 0.4071$ $p < 10^{-5}$ ***	$R^2 = 0.0403$ $p = 0.17$
San Francisco	$R^2 = 0.0017$ $p = 0.73$	$R^2 = 0.3093$ $p < 10^{-6}$ ***	$R^2 = 0.3271$ $p < 10^{-5}$ ***	$R^2 = 0.0194$ $p = 0.25$

*incr: HPI percentage increase between 2016 and 2020

**ctypct: Percent of total interstate immigrants to a county from given California county

***dist: Distance between a county to given California county

RED: LA Region

BLUE: Bay Area

From these results, we see that distance from a county to the given California county (*dist*) is significantly correlated with its House Price Increase (*incr*). This confirms the impact of the Western high-HPI=increase cluster in our model. Whichever California county one might choose, it is nearest to other counties in the West, where HPI increased significantly from 2016-2020. When adding *ctypct* to *dist* as an explanatory factor, the R^2 and p values are largely unchanged. However, there were some interesting patterns in the *incr ~ ctypct* regression. Three counties (Los Angeles, San Diego, Orange) had statistically significant correlations ($R<0.01$) there. *ctypct* explained ~7% of the variance in *incr* in LA County, ~3% in San Diego County, and ~10% in Orange

County. These three counties, all in Southern California, (and not neighboring San Bernardino and Riverside) might have some different pattern occurring than the rest.

To further test the impact of this West high-HPI-increase cluster, I subset the flows out of Los Angeles, Orange, and San Diego so that only counties outside the West cluster were in the dataset (a radius about as far as Denver from Southern California). With this subsetting, I ran the regressions again. The results were that, outside of the West, there is virtually no correlation between *California County Percent* and *HPI Increase*: all three regressions gave a p-value greater than 0.8.

So, it seems that urban Californians, in particular, do not cause (or correlate with) house price increase more than migrants from any other place in the US. However, interstate immigration in general does seem to correlate with house price increase. Running a correlation between the two yields an R-squared of 0.059 and a very small p-value $< 10^{-15}$. This means that the correlation is quite significant, but that interstate immigration explains only 6% of the variation in house price increase. It may be that interstate migrants tend to move to places that are already experiencing house price increase (“booming” areas, perhaps), or that they cause house price increase when they move there.

Discussion and Conclusions

In all, the data seem to suggest that urban Californian interstate emigrants exert no more influence on increase in housing price than any other migrants. This is with the potential small exception of Los Angeles, Orange, and San Diego county emigrants driving up HPI in the West, which I did not account for in my analysis. This may make sense, given that the flows from Southern California to Clark (Las Vegas) and Maricopa (Phoenix) counties are particularly large. Home prices there are far lower than in California, especially in Las Vegas, and it is often millennials looking to buy homes who are moving there, potentially driving up house prices as they make housing scarcer[8]. It is also interesting to see that Riverside and San Bernardino counties, despite also being in Southern California, did not have the same effects. This is perhaps because they are already landing spots from people from the central city seeking more remote life with cheaper housing.

But the reason why Californians largely do not seem to be particular drivers of house price increase may be because they are not wealthy people who are moving there, with money to spend and gentrification to do. Instead, many of these emigrants are likely (at least by my mechanism) people who can no longer afford California and have been pushed out. They are more likely in search of lower rent or house prices rather than amenities in their new states of residence. While it may be that Californians are moving *en masse* to other states, and that house prices are increasing in those places, it is likely not that the Californians moving there are particularly prone to causing that trend.

[8] https://www.storagecafe.com/blog/top-states-people-are-moving-to/

The overall significant (albeit small) positive correlation between emigration overall warrants discussion, though. New interstate immigrants to a county demand housing, and when demand increases faster than supply can keep up, prices increase. If this explains the correlation, halting the house price increases can come in the form of building more housing (especially affordable housing, given that immigrants may be in search of more affordable housing) or capping or stability rent to keep the region affordable for those who have already been living there. Otherwise, they may continue the trajectory, moving to places with *lower* housing prices and increasing them there. Granted, interstate domestic immigration is at most a small piece of the puzzle of housing price.

The described pattern contradicts Ravenstein's second law of migration. Ravenstein suggested that the "natural" way for migration to flow is from remote areas to dense areas in a pattern of "absorption". The gap of people in the remote area allows for even more remote people to move in. However, this process is what Ravenstein terms "dispersion," the exact inverse of absorption. With the Bay Area and Los Angeles being two of the most urban places in the US, it would be difficult to describe a way for absorption *from* them to occur. Nevertheless, this does seem to be a process of dispersion, partially caused in quantity/magnitude, if not quality, by Californians.

References

Barbosa, Hugo, et al. “Human mobility: Models and applications.” *Physics Reports*, vol. 734, 2018, pp. 1–74, https://doi.org/10.1016/j.physrep.2018.01.001.

Dietz, Joshua, et al. “Keeping in Motion or Staying Put: Internal Migration in the United States and China.” *MDPI*, Multidisciplinary Digital Publishing Institute, 8 July 2023, www.mdpi.com/2075-4698/13/7/162.

“Droves of Californians Are Moving to Texas. Here’s the Life They Are Finding.” *Los Angeles Times*, Los Angeles Times, 7 Nov. 2023, www.latimes.com/california/story/2023-11-06/we-fit-in-a-lot-better-here-californians-who-flee-to-texas-find-good-company-camaraderie.

Ellis, Mark. “Reinventing US internal migration studies in the age of international migration.” *Population, Space and Place*, vol. 18, no. 2, 2011, pp. 196–208, https://doi.org/10.1002/psp.666.

Gatea, Maria. “Where Are People Moving to, and from, in the US? [Study].” *StorageCafe*, 12 Oct. 2023, www.storagecafe.com/blog/top-states-people-are-moving-to/.

Gunderson, Ronald, and David Sorenson. “An Examination of Domestic Migration from California Counties.” *AgEcon Search*, University of Minnesota, 2010, ageconsearch.umn.edu/record/132439/s.

Hollenhorst, Maria. “She Left California for Arizona, but Found High Rent and Lower Wages.” *Marketplace*, 5 Dec. 2022, www.marketplace.org/2022/12/05/she-left-california-for-arizona-but-found-high-rent-and-lower-wages/.

Luscombe, Belinda. “Why Phoenix Has the Fastest Growing Home Prices in the U.S.” *Time*, Time, 3 May 2022, time.com/6170497/phoenix-fastest-growing-home-prices/.

Molloy, Raven, et al. “Housing Supply and affordability: Evidence from rents, housing consumption and household location.” *Finance and Economics Discussion Series*, vol. 2020, no. 044, 2020, https://doi.org/10.17016/feds.2020.044.

Pierre, Jesse. “NC Lawmaker Taking Step to Provide Renters Some Relief | Wcnc.Com.” *WCNC Charlotte*, 20 Mar. 2023, www.wcnc.com/article/money/markets/real-estate/charlotteans-feeling-high-rent-increases-nc-lawmaker-renters-relief/275-567588d2-10f0-46d7-938c-bdbc4936b0b9.

"Redfin Reports Homebuyers Are Looking to Relocate to Affordable Areas–Especially in Florida–amid High Rates, Prices." *Redfin Corporation*, 29 Nov. 2022, investors.redfin.com/news-events/press-releases/detail/841/redfin-reports-homebuyers-are-looking-to-relocate-to.

Rees, Philip, and Nik Lomax. "Ravenstein revisited: The analysis of migration, then and now." *Comparative Population Studies*, vol. 44, 2020, https://doi.org/10.12765/cpos-2020-10.