The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data

Monica Alexander*[1], Kivan Polimis[2], and Emilio Zagheni[3]

[1]University of Toronto
[2]MAANA
[3]Max Planck Institute for Demographic Research

Abstract

Natural disasters such as hurricanes can cause substantial population out-migration. However, the magnitude of population movements is difficult to estimate using only traditional sources of migration data. We utilize data obtained from Facebook's advertising platform to estimate out-migration from Puerto Rico in the months after Hurricane Maria. We find evidence to indicate a 17.0% increase in the number of Puerto Rican migrants present in the US over the period October 2017 to January 2018. States with the biggest increases were Florida, New York and Pennsylvania, and there were disproportionately larger increases in the 15-30 age groups and for men compared to women. Additionally, we find evidence of subsequent return migration to Puerto Rico over the period January 2018 to March 2018. These results illustrate the power of complementing social media and traditional data to monitor demographic indicators over time, particularly after a shock, such as a natural disaster, to understand large changes in population characteristics.

*monica.alexander@utoronto.ca

1 Introduction

Understanding migration patterns, including emigration destinations and the characteristics of those who moved and those who are left behind, has important implications for designing policy to best help those in need. It is particularly relevant to be able to track changes in population after disasters of large-scale environmental events. A recent notable example is Hurricane Maria, which hit Puerto Rico in September 2017. While the initial official death toll was reported by the government as 64, this was recently revised up to 2,975 (Santos-Burgoa et al., 2018), after independent researchers observed that a higher death count was more likely than the initial total (Kishore et al., 2018; Santos-Lozada and Howard, 2018). Additionally, lack of basic amenities, like power and clean water, still remains a problem in many parts of this U.S. territory almost one year after the hurricane (Kishore et al., 2018).

Besides the impact on mortality, natural disasters can also cause substantial out-migration from the location affected. For example, approximately 1.5 million New Orleans' residents evacuated around the time of Hurricane Katrina in 2005, and two years later, more than a third of the pre-hurricane residents of New Orleans had not returned to the city (Fussell et al., 2010). Initial estimates for Puerto Rico suggest that out-migration was roughly 6% of the total population (Echenique and Melgar, 2018). A decrease in the size of the total population of this magnitude has important implications for both the origin and the receiving locations. In addition to that, if not properly accounted for, large flows of migrants skew estimates of mortality rates, as they affect the denominator for mortality rates, i.e. population counts across different groups.

Although data on migration are key for demographic estimation and projections — and to understand societal implications of events like natural disasters — they often do not exist, and those that do exist are typically not produced in a timely

manner. For example, in the United States, there is no publicly available dataset on monthly migration movements between states and outlying territories. The American Community Survey (ACS) provides details on birthplace and place of residence one year before the interview. This is a good resource to estimate migration rates by state and other demographic characteristics. However, ACS estimates are only produced annually, with no information available about changes in migration stocks from month to month. In addition, ACS data are published with a delay: as of October 2018, the 2017 estimates had not been released. As such, other sources of data are often used to track short-term migration changes. For example, previous work has used mobile phone data, as well as change of address and school enrollment changes, or flight passenger data, to track movements from Puerto Rico to the US after Hurricane Maria (Echenique and Melgar, 2018; Meléndez and Hinojosa, 2017; Rayer, 2018). Even when some approximate estimates are produced, the basic demographic characteristics of migrants are hard to ascertain using these types of data sources.

To overcome issues related to timeliness and lack of available demographic information, and to estimate changes in out-migration after Hurricane Maria, we propose the use of social media data. Recently, there has been an increased interest in applications that use social media and Web data to estimate and track demographic indicators over time. Facebook, for example, can be thought of as a large 'digital census' that is regularly updated. An issue with using such data is that they are not representative of underlying populations, and can lead to biased results. However, previous research has illustrated how social media data can be adjusted to account for some of the major biases in order to produce estimates of demographic indicators (Yildiz et al., 2017; Zagheni et al., 2014, 2017). For our purposes, Facebook data are particularly well-suited to studying Puerto Rican migration due to the large proportion of the population that uses Facebook on a regular basis (Rosado, 2017). Importantly, although Facebook data have a number of limitations, in the context of estimating population movement

from Puerto Rico after Hurricane Maria, Facebook is a key data source that we believe should be taken seriously.

We show how changes in the population of Facebook users can be used to estimate out-migration after Hurricane Maria. We collected aggregated data from Facebook through its Marketing Application Programming Interface (API) roughly every three months for a period of 12 months. We estimate initial out-migration as the change in Puerto Rican migrant stocks over the period October 2017 to January 2018. The data available allow us to describe migration destinations in the US, as well as changes in the age and sex profiles of the out-migrants. We combine Facebook data with more traditional demographic data and use difference-in-differences as an estimation procedure to adjust for known biases in the Facebook population compared to the broader population.

2 Data

Facebook for Business has developed a targeted advertising platform, called Ads Manager that uses a graphical user interface to allow advertisers to target specific audiences. The dimensions that can be targeted include information directly reported by Facebook users, such as age or sex, and information indirectly inferred from the use of Facebook platform or affiliated websites, such as location and behavioral interests. Before launching an advertisement, an advertiser can select a variety of characteristics (e.g., Australians living in California, who are female, and aged 30-35) and get an estimate of the 'potential reach' (monthly active users) to this subgroup. These estimates can be obtained, in a programmatic way, for a variety of different expatriate ('expat') groups by age, sex, and education.

It is the estimates of potential reach by expat group, age and sex that we use to track sizes of migration stocks over time. These estimates can be obtained before the launch

of an advertisement, and as such are obtained free of charge. We use the Ads Manager backend application, Facebook's Marketing API, to extract estimates of potential reach over time programmatically with the Python module pySocialWatcher (Araujo et al., 2017). With pySocialWatcher, we collected data across 11 age groups (10 UN age groups from 15-19 to 60-65; an 11th group for the entire available Facebook population of 13-65 was also used), three gender groups (female, male, and total population) and multiple education categories. Data was collected using Amazon Web Services (AWS) EC2 Instance servers.

As part of a broader project on using social media in demographic research, we started data collection in January 2017. This data collection is twofold for each wave, state-level estimates of all Facebook users (by age, sex, and gender) as well as state-level estimates of 90 expat groups. Since then, we have collected a new wave of data every 2-3 months, specifically:

- Wave 1: January 2017
- Wave 2: April 2017
- Wave 3: June 2017
- Wave 4: October 2017
- Wave 5: January 2018
- Wave 6: March 2018

Wave 4 overlaps with the period when Hurricane Maria impacted Puerto Rico. We use the change in the Facebook potential reach from Wave 4 to Wave 5 as the basis of our estimates of out-migration. We also use the change from Wave 5 to Wave 6 to assess subsequent migration of Puerto Ricans in the next three months.

2.1 Data Challenges

An important challenge with the Facebook data is that the population of Facebook users is not representative of the broader population. In particular, the age distribution of Facebook users is on average younger than the broader population. We compare the age distribution of Puerto Rican expats in the first wave of the Facebook data with the age distribution of Puerto Rican migrants in the 2016 5-year ACS. Figure 1 shows the two age distributions for the nine states with the largest populations of Puerto Rican migrants. In all states, the ACS has an older age distribution, while the Facebook distributions are concentrated in the age groups less than 30 years. Appropriate adjustments are thus needed to extract meaningful information from non-representative samples.

[Figure 1 about here.]

A second major source of bias is related to lack of detailed information about the algorithms that Facebook uses internally in order to estimate migrants (called 'expats' or 'ex-pats' in the Marketing API). Estimates of expat users in the Facebook data appear to vary over time (wave), independently of underlying changes in migrant stocks, because of rounding and possibly because of adjustments of the algorithms internally used. For example, there was a marked decrease of approximately 10% in the size of all expat populations from wave 3 to wave 4, followed by an approximate 10% increase going from wave 4 to wave 5 (see Supplementary materials). These changes are independent of migrant origin and do not appear to be linked to any broad-scale migration trends. It seems more likely that these large changes are related to Facebook's method of estimating 'population at reach'. If we only looked at changes in the Puerto Rican population over waves, without accounting for potential platform-wide changes, we would over-estimate the increase in migrants after Hurricane Maria.

3 Methods

In order to address the data challenges described above, we used two main methods. Firstly, we use Difference-in-Differences (DiD) to account for unrelated changes in migrant stocks over time, and to more accurately estimate the change in migrant stocks after Hurricane Maria. We compare the change in Puerto Rican migrants to the change in migrants from all other origins except Mexico, who would not have been affected by the hurricane. We remove Mexico from the control group to avoid large fluctuations in the Mexican migrant population due to reasons other than Hurricane Maria. We calculated DiD based on the difference in the proportion of the total population who are Puerto Rican compared to the proportion of the total population who are migrants from all other origins, excluding Mexico. We used DiD to estimate changes in the Puerto Rican migrant population for the whole US, by state, age and sex. Standard errors were calculated based on the Binomial approximation to the Normal distribution.

Note that DiD inherently assumes that, had Hurricane Maria not occurred, the relationship between the two migrant groups would have remained the same. One way to assess the validity of this assumption is to check trends in the migrant groups before the hurricane. As shown in the Supplementary materials, the trends are generally parallel before October, thus supporting the validity of our assumption.

Secondly, to account for the biased age distributions observed, we adjust the Facebook data using the age distributions observed in the ACS data, which are expected to be a closer representation of the underlying 'true' distribution. Note that as we are considering DiD of the Facebook data over time, the bias in the Facebook distributions will be differenced out. The ACS correction only comes into play when we convert percentage changes into the implied number of migrants. Lastly, to avoid issues with rounding in the Facebook data, for the state-by-state analysis, we only consider states that had a total Puerto Rican expat population in the Facebook data of at least 18,000

people. This corresponds to at least 1,000 people in each of the 18 age by sex groups (2 sexes and 9 five-year age groups from 15-60).

4 Results

Over the period October 2017 to January 2018, at the national level we estimate a 17.0% increase in the number of Puerto Rican migrants present in the US. This is equivalent to an increase of approximately 180,200 people, corresponding to approximately 5.4% of the total population of Puerto Rico (US Census Bureau, 2018).

On a state-by-state basis, the states with the largest existing populations of Puerto Rican migrants also had the largest increases after Hurricane Maria (Table 1, Figure 2). Florida had the largest percentage increase in Puerto Rican migration, increasing 21.6% over the period October 2017 to January 2018. This corresponded to an increase of around 62,000 people. Pennsylvania, New York, Connecticut and Massachusetts also had substantial increases of 8,000-15,000 migrants.

[Table 1 about here.]

[Figure 2 about here.]

4.1 Changes by age and sex

Figure 3 shows the proportional change in each five-year age group aged 15-60 in the nine states with the largest Puerto Rican migrant populations. In general, there is a noticeable increase in the 15-30 age range. This suggests those people moving away from Puerto Rico after the hurricane were those in the working age groups.

[Figure 3 about here.]

We define the sex ratio of migrants as the number of men divided by the number of women. All of the nine states have sex ratios greater than one, meaning that there were more men than women. In the period after the hurricane, the sex ratios remained fairly stable, suggesting that the number of women migrating compared to men did not change much. However, as shown in Table 2, there was an increase in the sex ratios for states such as Florida and Texas, suggesting that the distribution of migrants shifted to include more men. The opposite was true for Pennsylvania, which saw a 5.3% decrease in the sex ratio, suggesting an increase in women.

[Table 2 about here.]

4.2 Evidence of return migration since January 2018

In addition to assessing changes in Puerto Rican population in the period October 2017 to January, we also looked at the change in migrant stocks from January 2018 to March 2018 to assess the extent to which migration out of Puerto Rico that occurred in the three months following the Hurricane was counteracted with return migration in subsequent months. The results suggest that at the national level there was a 1.8% decrease in the Puerto Rican population on the mainland US, corresponding to around 18,900 people. In terms of specific states, there was a 7.1% decrease in the Puerto Rican population in Florida, corresponding to a decrease in population of around 20,400 people (Table 3). Other states such as Massachusetts and Connecticut saw decreases of 3 to 4%, corresponding to decreases of around 2,000-4,000 people.

[Table 3 about here.]

5 Discussion

Data obtained from Facebook's advertising platform highlight the substantial out-migration from Puerto Rico in the three months following Hurricane Maria. The data suggest a 17% increase in Puerto Rican migrant population in mainland US in the three months following the Hurricane, corresponding to around 180,200 people. People who moved were disproportionately in the younger working age groups and men. Additionally, we saw evidence of return migration to Puerto Rico, as suggested by a 1.8% decrease in Puerto Ricans in the continental USA since January 2018 (Table 3).

This case study illustrates the value in collecting large-scale social media data for improving the timeliness of monitoring migration after a natural disaster. Estimates obtained from Facebook data are complementary to official statistics and other methods of dynamic population mapping (Echenique and Melgar, 2018; Meléndez and Hinojosa, 2017; Deville et al., 2014). For example, the estimated outflows for the total population and population by state are in line with previous studies that utilized other sources of data, including cell phone records (Echenique and Melgar, 2018), flight passenger data (Rayer, 2018), port authority data (Santos, 2018), and forecasts from the American Community Survey (Meléndez and Hinojosa, 2017). In particular, cellphone data have been used for mapping population density and changes after crises, with the potential to produce extremely granular estimates of internal migration. However, those data are limited to internal movements. Data obtained from Facebook offer a richness of information that allows trends across international borders to be considered, as well as mapping population movements by key demographics such as age and sex. Facebook data and other non-traditional sources are particularly useful for complementing survey data. Social media data will never replace traditional surveys. However, we believe that there are some specific cases, like in the context of natural disasters, where digital

breadcrumbs can add relevant information to existing imperfect sources and prove particularly useful, at least in order to generate provisional estimates. There are several limitations of our approach, which are common to all studies that use data from the Facebook Advertising Platform or similar sources (e.g. (Zagheni et al., 2017; Garcia et al., 2018; Rampazzo et al., 2018). Firstly, the users of Facebook are in general not representative of the broader population for which we are interested in obtaining estimates. For example, previous research has shown that Facebook data tends to have a lower proportion of migrants, and a younger age distribution, than the general population (Zagheni et al., 2017). To overcome issues of representativeness, we implemented a difference-in-differences approach, allowing us to assess changes in the Puerto Rican population relative to non-Puerto Rican migrants. Additionally, the magnitude of changes was estimated based on known population sizes in the American Community Survey.

Secondly, we only had collected data in three-month intervals, and so did not have observations for the months between October and January. We are potentially missing more short-term moves, and underestimating the initial out-migration around the time of the Hurricane. The results presented in this paper offer a further motivation for more frequent data collection in the future, as well as the potential benefit of new and increased collaborations with social media companies such as the 'Social Science One' initiative (Social Science One, 2018).

Lastly, the Facebook data sample displays a level of variability across collection waves that appears to be unrelated to migration events. These changes make it more difficult to study substantive changes in migrant populations over time. By using a difference-in-differences approach, we were able to filter out systematic changes in the Facebook data, due for example, to changes in the Advertising platform or how the demographics are estimated. However, given the nature of the data and the lack of documentation from Facebook, we may be underestimating our uncertainty. In the

future, closer collaborations with social media companies would be important in order to map in a systematic way all potential sources of errors and uncertainty.

Notwithstanding, these results illustrate the power of using social media data to monitor demographic indicators over time, particularly after a shock, such as a natural disaster, to understand large changes in population characteristics. Social media data are especially powerful when coupled with other, complementary, data sources, such as data from traditional surveys and administrative sources (which are representative but often not timely or with sample sizes that are too small for state or local estimates) and mobile phone movements (which are highly granular but do not have demographic information). We found evidence that the majority of Puerto Ricans who left after Hurricane Maria traveled to areas where there was already a large existing population of Puerto Rican migrants, confirming the existence of strong social networks of Puerto Ricans across continental United States. In addition, the observation that the increase in Puerto Rican migrants was disproportionately male and in younger age groups may suggest that the more vulnerable were left behind. This work highlights how rapidly population movements can occur in the months after a natural disaster, and the value in collecting 'digital censuses' to track these short-term changes in populations.

References

Araujo, M., Y. Mejova, I. Weber, and F. Benevenuto (2017). Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. WebSci '17, New York, NY, USA. ACM.

Deville, P., C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences 111*(45), 15888–15893.

Echenique, M. and L. Melgar (2018, May). Where Puerto Rico's Residents Migrated Since Maria - CityLab.

Fussell, E., N. Sastry, and M. VanLandingham (2010). Race, socioeconomic status, and return migration to New Orleans after Hurricane Katrina. *Population and environment 31*(1-3), 20–42.

Garcia, D., Y. M. Kassa, A. Cuevas, M. Cebrian, E. Moro, I. Rahwan, and R. Cuevas (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences*, 201717781.

Kishore, N., D. Marqués, A. Mahmud, M. V. Kiang, I. Rodriguez, A. Fuller, P. Ebner, C. Sorensen, F. Racy, J. Lemery, and others (2018). Mortality in Puerto Rico after Hurricane Maria. *New England journal of medicine*.

Meléndez, E. and J. Hinojosa (2017). Estimates of post-hurricane María exodus from Puerto Rico. *Center for Puerto Rican Studies, Research Brief, October. https://centropr. hunter. cuny. edu/research/data-center/research-briefs/estimates-post-hurricane-maria-exodus-puerto-rico*.

Rampazzo, F., E. Zagheni, I. Weber, M. R. Testa, and F. Billari (2018). Mater certa

est, pater numquam: What can Facebook Advertising Data Tell Us about Male Fertility Rates? *arXiv preprint arXiv:1804.04632*.

Rayer, S. (2018, October). Estimating the migration of puerto ricans to florida using flight passenger data. URL: https://www.bebr.ufl.edu/sites/default/files/Research%20Reports/puerto_rican_migration.pdf.

Rosado, L. (2017, October). Facebook continues to dominate social-network use in Puerto Rico. URL: http://gaitherinternational.com/facebook-continues-to-dominate-social-network-use-in-puerto-rico/.

Santos, A. (2018). Estimates of excess passenger traffic in Puerto Rico following Hurricane María.

Santos-Burgoa, C., A. Goldman, E. Andrade, N. Barrett, U. Colon-Ramos, M. Edberg, A. Garcia-Meza, L. Goldman, A. Roess, J. Sandberg, and others (2018). Acertainment of the Estimated Excess Mortality from Hurricane Maria in Puerto Rico.

Santos-Lozada, A. R. and J. T. Howard (2018). Use of Death Counts From Vital Statistics to Calculate Excess Deaths in Puerto Rico Following Hurricane Maria. *JAMA*.

Social Science One (2018). Social science one. URL: https://socialscience.one/.

US Census Bureau (2018, October). Us Census Bureau quickfacts: Puerto rico. URL: https://www.census.gov/quickfacts/pr.

Yildiz, D., J. Munson, A. Vitali, R. Tinati, and J. A. Holland (2017). Using Twitter data for demographic research. *Demographic Research 37*, 1477–1514.

Zagheni, E., V. R. K. Garimella, I. Weber, and others (2014). Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*, pp. 439–444. ACM.

Zagheni, E., I. Weber, and K. Gummadi (2017). Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review 43*(4), 721–734.

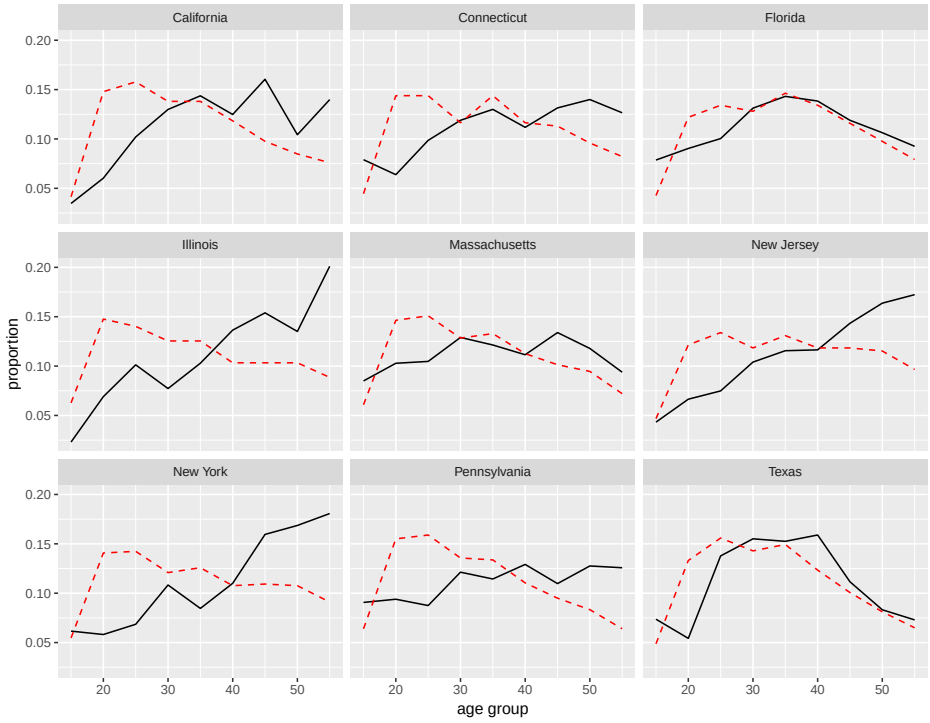

Figure 1: Age distribution of Puerto Rican migrants in Facebook data in January 2017 (red dashed line) and 2016 American Community Survey data (black solid line).

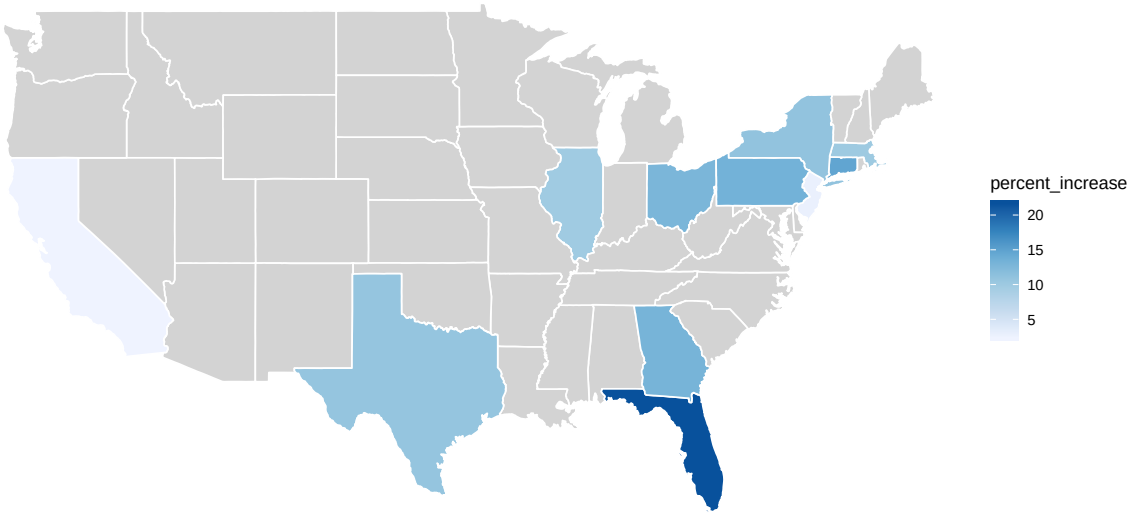

Figure 2: Estimated increase in Puerto Rican migrant stocks from October 2017 to January 2018. Note that only the states with a Puerto Rican migrant population of at least 18,000 are shown.

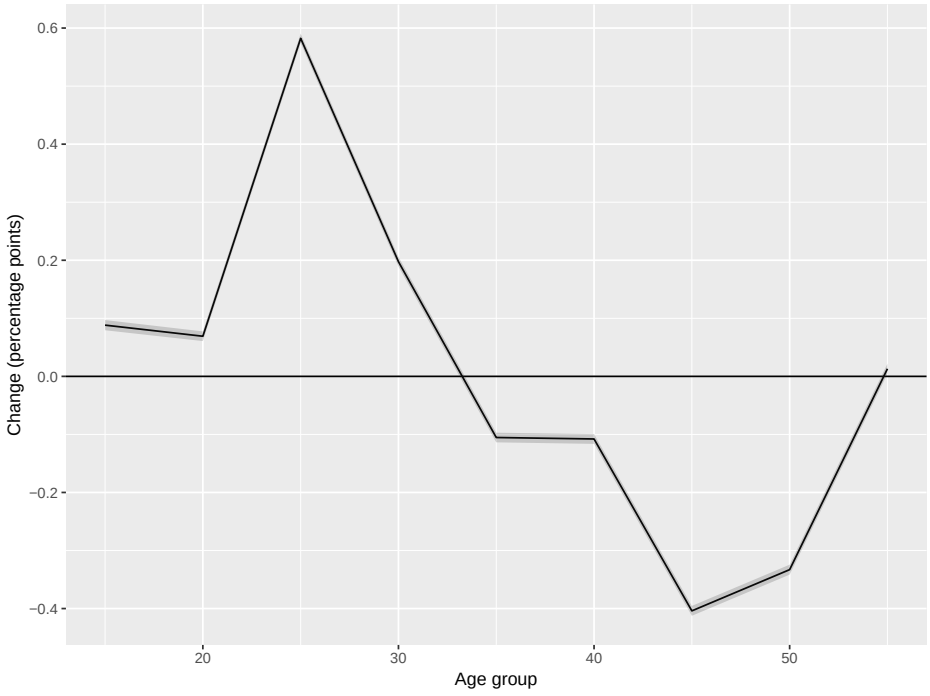

Figure 3: Estimated change in Puerto Rican migrant age distribution from October 2017 to January 2018.

Table 1: Estimated increase in Puerto Rican migrant stocks from October 2017 to January 2018. The 95% confidence intervals are shown in parentheses.

State (95% CI)	% Increase (95% CI)	Population Increase
Florida	21.6 (20.9, 22.3)	61992 (60010, 63973)
New York	11 (10.3, 11.7)	14662 (13758, 15567)
Pennsylvania	13.4 (12.7, 14.1)	13151 (12427, 13875)
Connecticut	14.7 (12.9, 16.5)	9541 (8365, 10716)
Massachusetts	10.1 (8.82, 11.4)	8824 (7708, 9940)
Texas	10.8 (10.4, 11.2)	5394 (5179, 5609)
Ohio	12.8 (12.2, 13.4)	3002 (2865, 3139)
Illinois	9.9 (9.15, 10.6)	2684 (2482, 2887)
Georgia	13.1 (12.4, 13.8)	2600 (2464, 2736)
New Jersey	2.9 (1.56, 4.24)	2282 (1228, 3335)
California	2.4 (1.86, 2.94)	576 (446, 706)

Table 2: Estimated change in Puerto Rican migrant sex ratios (male/female) from October 2017 to January 2018. The 95% confidence intervals are shown in parentheses.

State	Sex ratio (October 2017)	% Change (95% CI)
Florida	1.16	0.08 (0.076, 0.083)
Texas	1.01	0.065 (0.059, 0.071)
Georgia	1.14	0.018 (0.006, 0.031)
Connecticut	1.35	0.014 (0.004, 0.024)
New Jersey	1.19	0.014 (0.005, 0.023)
California	0.962	0.011 (0, 0.021)
Massachusetts	1.28	0.008 (0, 0.016)
New York	1.22	0.006 (0, 0.011)
Ohio	1.13	0.005 (-0.007, 0.016)
Illinois	1.13	-0.002 (-0.013, 0.008)
Pennsylvania	1.19	-0.053 (-0.059, -0.047)

Table 3: Return migration: Estimated change in Puerto Rican migrant stocks from January 2018 to March 2018. The 95% confidence intervals are shown in parentheses.

State	% Change (95% CI)	Population Change
Florida	-7.1 (-7.77, -6.43)	-20377 (-22289, -18465)
Massachusetts	-4.5 (-5.52, -3.48)	-3931 (-4826, -3037)
Connecticut	-3.6 (-5.04, -2.16)	-2336 (-3273, -1400)
Texas	-3.5 (-3.91, -3.09)	-1748 (-1952, -1544)
Pennsylvania	-1.4 (-2.03, -0.773)	-1374 (-1989, -759)
Ohio	-1.7 (-2.14, -1.26)	-399 (-503, -295)
New York	0.4 (-0.241, 1.04)	533 (-322, 1388)
Illinois	2.8 (2.15, 3.45)	759 (583, 935)
Georgia	6.4 (5.86, 6.94)	1270 (1163, 1377)
New Jersey	2.3 (1.15, 3.45)	1810 (908, 2711)
California	8.5 (7.97, 9.03)	2039 (1911, 2167)