

Mapping New York City’s Undocumented Residents

Alex Chohlas-Wood, Yilong Zha, Gang Zhao, Jane S Adams

March 27, 2014

In his inaugural State of the City address, Mayor DeBlasio announced that the city of New York would begin issuing municipal ID cards to all residents, regardless of immigration status. DeBlasio framed the promise by stating, “New York City is your home too, and we will not force *any* of our residents to live their lives in the shadows.” [6]

In identifying a municipal ID program as a central part of his agenda, we believe that Mayor DeBlasio would benefit from a data-driven snapshot of life for those without documentation. Such a snapshot would quantify the magnitude of the need (i.e., how many municipal IDs the city should be prepared to offer), the current capacity of various services that require documentation, and the resources required to scale those services to meet the increased demand presented by a more fully documented population. A municipal ID will have little to no impact on the lives of the currently undocumented if the services that they will gain access to with identification lack the resources to serve them. In short, offering municipal IDs alone is not enough; without the cooperation of adequate resources, such a program would have little impact. It is necessary, then, to quantify the resources required to make a municipal ID program effective.

Furthermore, offering municipal IDs and having the requisite resources to support increased demand does not guarantee that the current undocumented population will 1) apply for municipal IDs, and 2) utilize the new services at their disposal. We identify two key steps that we think will increase the likelihood that current undocumented residents will apply for and use municipal IDs, i.e., that would render the city’s investment in such a program worthwhile. The first step requires developing an outreach program that mitigates existing reservations regarding how the government and its authorities respond to citizenship status. Different groups may have different experiences of the government and its authorities, and may have different resources available to them. To approach the undocumented population as a homogenous entity would be to fundamentally misunderstand the problem. We are not at all suggesting that anyone plans to approach the undocumented population as a homogenous entity; what we are suggesting is that we lack the information to do much more than that. Very simply, our project aims to provide that information.

The second step in an effective municipal ID program is ensuring, as far as possible, that the ID is powerful enough as a tool and means of access to supplant existing tools and means of access used by the undocumented population to meet their needs. This second

step requires a comprehensive understanding of: the most important services that undocumented residents need access to; what, if anything, is preventing them from accessing them; and whether or not a municipal ID would provide that access and, if so, how. Part of this second step is, of course, marketing. Marketing has the greatest potential impact when it is driven by awareness of the client and their primary concerns and needs.

The challenge with obtaining a snapshot of the undocumented situation, one that captures the information needed to make informed steps in launching an effective municipal ID program, is precisely that these individuals are undocumented. This report provides an estimate of the size of the undocumented population residing in New York City by country of origin. Our estimate was obtained using the residual method employed for previous estimates of the unauthorized population.[1] We take the unauthorized foreign-born population to be the remainder, or “residual”, after the legally resident foreign-born population is subtracted from the total foreign-born population. Data to estimate the legally resident population were obtained primarily from the Department of Homeland Security (DHS), while estimates of the total foreign-born population are taken from the American Community Survey (ACS) conducted by the U.S. Census Bureau.

1 Project Objectives

To effectively target a municipal ID outreach program, it is important to have a sense of the number, location, and language (which we assume based on country of origin) of undocumented residents. The primary objective of this project is to estimate the size of, and map, the undocumented population of New York City by country of origin. The challenge boils down to estimating what portion of the foreign-born population of New York City is illegally present, assuming for our purposes that legal residents will have, or have access to, some form of documentation.¹

This data will enable an outreach program to identify the areas where they should focus their efforts, and will allow them to track their progress. To be specific, if we estimate that there are 1,166 undocumented residents living in Census Tract 6, an outreach group can use that number to determine: how many outreach members they need; how many residents they need to register each day in order to stay on schedule, if they have a schedule; and their progress, namely when they have registered a sufficient portion of the undocumented residents in that census tract and can move on to the next. If we estimate further that, of the 1,166 undocumented residents living in Census Tract 6, 1,090 are from China, an outreach group can deploy more of their members with Chinese language skills to that area. Our estimates can also be used to identify local, and culture-specific, community centers

¹Jorge Montalvo at the State Office of New Americans informed us that this was not necessarily the case; indeed, according to 311 Immigration Hotline call data for 2013, most calls were from lawful permanent residents, this group placing more than twice as many calls as the next group which comprised persons with no documentation. The top two immigration concerns were eligibility and documentation, and most people calling about immigrant visa applications were calling for information about relative petitions.

and institutions that outreach groups can coordinate with to reach the undocumented populations in those areas. For example, if we estimate that Hondurans make up the majority of the undocumented population in Census Tract 225, and we estimate that there are neither undocumented Indians nor undocumented El Salvadorans there, an outreach group should focus on coordinating with the local Honduran community through its centers and institutions, and can reasonably set Indian and El Salvadoran community centers and institutions aside.

After obtaining an estimate, our secondary objective was to identify those features that are predictive of the presence of undocumented residents in an area. We hoped to find a “signal” that would not only augment our existing estimate of the size and location of the undocumented population, but also hopefully shed light on the current situation of undocumented residents living in New York City. For example, is there a correlation between illegal building conversion complaints and undocumented residents, and does that indicate that building hazards are a burden borne disproportionately by undocumented residents? However, in the available time and with the available data, we were unable to identify features that were predictive of the presence of undocumented residents.

2 Definitions

Before deriving our estimate, we would like to clarify the major terms that we will use throughout.

Foreign-born is defined by the Census Bureau as any individual that was not a United States citizen at birth, i.e., individuals that were neither 1) born in the United States, nor 2) born abroad to United States citizen parents. The foreign-born population includes naturalized citizens, lawful permanent residents (LPRs), migrants, as well as individuals that are in the United States illegally, such as temporary workers and students.

Legal foreign-born or *authorized* refers to legal permanent residents, naturalized citizens, asylees, refugees, and non-immigrants, and we assume that these residents have, or at least have access to, some form of documentation.

Unauthorized refers to residents that are illegally present in the United States.

Undocumented refers to any person that does not have some form of documentation, whether they are authorized or unauthorized, foreign-born or domestic. We would like to emphasize that not all undocumented persons are unauthorized. However, as authorized residents have theoretical access to forms of documentation other than a municipal ID, we take them to be secondary clients of a municipal ID program.

To make our estimate a truer estimate of the undocumented population, we could incorporate an estimate of the portion of the authorized population, foreign-born and domestic, that does not have some form of documentation. As a data problem, obtaining such an estimate is the same as the problem addressed here. It requires establishing a set of assumptions about the circumstances under which an authorized resident wouldn't have documentation, and then mapping the datasets that could validate those assumptions (if such datasets exist) against one another to, hopefully, allow an estimate to emerge where one did not exist before. Specifically, we looked into estimating the homeless, underbanked, and low-income populations of New York City. We were able to derive a preliminary estimate of the low-income population, but did not make satisfactory progress in estimating the homeless and underbanked populations to present estimates of these populations here.

3 Data

We used the following datasets for our population estimates:

- U.S. Bureau of the Census, “2012 American Community Survey 1-Year Estimates” http://factfinder2.census.gov/faces/help/jsf/pages/metadata.xhtml?lang=en&type=dataset&id=dataset.en.ACS_12_1YR
- U.S. Bureau of the Census, “2008–2012 American Community Survey 5-Year Estimates” http://factfinder2.census.gov/faces/help/jsf/pages/metadata.xhtml?lang=en&type=dataset&id=dataset.en.ACS_12_5YR²
- U.S. Department of Homeland Security, “2012 Yearbook of Immigration Statistics” <http://www.dhs.gov/yearbook-immigration-statistics>
- U.S. Department of Homeland Security, “Profiles on Legal Permanent Residents” <http://www.dhs.gov/profiles-legal-permanent-residents>
- U.S. Department of Homeland Security, “Profiles on Naturalized Citizens” <http://www.dhs.gov/profiles-naturalized-citizens>
- U.S. Department of Labor, “Foreign Labor Certification” <http://www.foreignlaborcert.doleta.gov/performance/cfm>
- Institute of International Education, “Open Doors Fact Sheet: New York” <http://www.iie.org/~media/Files/Corporate/Open-Doors/Fact-Sheets-2012/State/New-York-Open-Doors-2012.ashx>
- TRAC Immigration, “U.S. Deportation Outcomes by Charge: Completed Cases in Immigration Courts,” Syracuse University http://trac.syr.edu/phptools/immigration/court_backlog/deport_outcome_charge.php

²The ACS data are based on a sample of the population, and thus are subject to sampling variability.

We used the following datasets for analysis and predictive learning:

- NYC 311, “311 Service Requests from 2010 to Present” <https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9?>
- U.S. Internal Revenue Service, “SOI Tax Stats–Individual Income Tax Statistics” <http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-%28SOI%29>
- NYPD Crash Data Band-Aid <http://nypd.openscrape.com/#/>
- U.S. Bureau of the Census, “2008–2012 American Community Survey 5-Year Estimates” http://factfinder2.census.gov/faces/help/jsf/pages/metadata.xhtml?lang=en&type=dataset&id=dataset.en.ACS_12_5YR
- Environmental Control Board, “Notice of Violations” <https://data.cityofnewyork.us/City-Government/ECB-Notice-of-Violations/y6h5-jvss>

4 Methodology

4.1 Estimate

While the U.S. Census Bureau does not collect immigration data, the DHS Office of Immigration Statistics publishes statistics and reports based on administrative data, such as the number of persons obtaining legal permanent resident status, refugees and asylees, naturalizations, non-immigrant admissions, and enforcement actions. The DHS also publishes estimates of the unauthorized population at the national and state level. They obtain their estimates via the “residual” method whereby the unauthorized population is taken to be the residual, or, what is left over, after subtracting the legal foreign-born population from the total foreign-born population. This method has been employed in previous estimates of unauthorized residents at the national level.[1] Using administrative data from the DHS, we estimate that the total foreign-born population of New York State on January 1, 2013 was 3,545,375. Via the residual method, we estimate that the total unauthorized population of New York State at that time was 582,774. What follows is a detailed derivation of the estimates that can be found in Table 1 below.

We begin with the total foreign born population for New York State provided by the ACS 2012 and 5-year estimates. This data was obtained from the ACS’s FactFinder, an online data portal maintained by the U.S. Census Bureau for accessing ACS estimates from the full sample for a particular year. From the total foreign-born population present in 2012, we subtracted those that entered the United States prior to 1980 on the assumption that most were eligible for legal permanent resident (LPR) status.³ We shifted the reference

³Please refer to section 249 of the INA, the registry provision.

	National	New York
Total foreign-born population	39,784,305	4,268,020
Foreign-born population (entered before 1980)		948,033
Foreign-born population (entered 1981–2012)		3,319,987
Shifted reference date		56,440
Undercount of non-immigrants		16,939
Undercount of legally resident immigrants		93,732
Undercount of unauthorized immigrants		58,227
Estimated total foreign-born population		3,545,375
LPR flow (1981–1990)	7,338,062	787,219
LPR flow (1991–2004)		1759876
LPR flow (2005–2012)		1,176,704
Estimated total LPR population		3,723,169
Refugee flow (2.3 years)		106,856
Asylee flow (4.4 years)		136,551
Mortal		243,703
Emigrate		712,363
Non-immigrant population		169,386
Estimated total legally present population		2,962,601
Estimated total unauthorized population		582,774

Table 1: Population estimates (shifted from 2012 reference data)

date for the 2012 ACS to January 1, 2013 by multiplying the population of 2012 entrants by 1.72. This number comes from the average of three ratios: the ratio of the estimated population in the 2010 ACS that entered the United States during 2009 versus the estimated population in 2009 that entered in 2009, and the comparable ratios for 2009 entrants in 2008 versus 2009, and for 2008 entrants in 2007 versus 2008.⁴

To account for people that were not captured by the census, we apply “undercount” rates to different portions of the population based on rates used by the DHS in its own estimates. We apply an undercount rate of 10% to the non-immigrant population, 2.5% to the legal foreign-born population, and 10% to the unauthorized population. It is important to note that these estimated rates are sensitive to the assumptions the DHS makes about the non-immigrant, legal foreign-born, and unauthorized populations; their assumptions, in this respect, are not benign. Via these steps, we arrive at an estimate of the total foreign-born population of New York State on January 1, 2013: 3,545,375.

We then estimate what portion of the above estimate are legal residents, again assuming for our purposes that legal residents will have some form of documentation. To do this, we take the flow of LPRs, refugees, and asylees, mortality rates, and emigration rates for

⁴We adopted this method from the DHS. Please see: Hoefer, Rytina, & Baker, 2012.

the legally present population. “Flow” refers to the year of entry or year of adjustment for LPRs, refugees, and asylees gaining LPR status.⁵ LPRs include 1) all new arrivals to the United States with immigrant visas issued by the State Department, and 2) qualified persons adjusting to LPR status. We use 2010 estimates from the DHS of the average time that refugees and asylees live in the United States before they adjust to LPR status to estimate the flow: 2.3 years for refugees, and 4.4 years for asylees. We take the national flow for three periods (1981–1990, 1991–2004, and 2005–2012) and multiply each by the ratio of the total foreign-born population of New York State to the total foreign-born population of the United States. We take the summation of these three periods and apply a 6.5% mortality rate and a 19% emigration rate. The mortality rate was taken from the 1999–2011 life tables[2], while the emigration rate was taken from estimates of emigration of the foreign-born population based on 1980 and 1990 Census data[4], after accounting for mortality. We subtract the portion of the population that died and the proportion of the population that emigrated from the legally present flow to arrive at an estimate of the legal resident population: 2,793,215.

To this estimate we then added the non-immigrant population, which includes temporary workers and students. We obtained data for temporary workers from fiscal year 2012 data from the Department of Labor, and for students from 2012 data from the Institute of International Education. Importantly, these data are not actual population counts, but come from admissions data. Adding the non-immigrant population to our previous estimate, we estimate the total legally present population: 2,962,601. Finally, we subtract this estimate from the total estimated foreign-born population of New York State to estimate the unauthorized population on January 1, 2013: 582,774.

To estimate the total unauthorized population of New York City, by census tract, we first apply the breakdown by country of origin of successful deportations for New York State obtained from fiscal year 2014 deportation data available from Syracuse University which can be found in Table 2. We apply this proportion to our estimate of the unauthorized population of New York State. This gives us an estimate of the total number of unauthorized residents by country of origin for New York State.

We assume that the country breakdown of successful deportations is representative of the breakdown of the total unauthorized population. Explicitly, since Chinese persons make up 18.5% of successful deportations in New York State, we assume that Chinese persons make up 18.5% of our estimated total unauthorized population. We do this for the top 10 countries represented in successful deportations: Chinese, Mexican, Guatemala, El Salvadoran, Ecuadorian, Honduran, Dominican, Indian, Nepalese, Colombian, and Other. Due to time and data constraints, we were unable to control for potential deportation biases.

While the Census Bureau does not gather citizenship data, they do gather data on coun-

⁵Immigration analysis have expressed some concern about capturing valid and reliable data on year of entry from the Census survey. Additionally, there is some concern about converting the DHS administrative data on legally resident immigrants to year of entry dates.

United States		New York		New York City	
Mexico	0.489	China	0.185	China	0.214
Guatemala	0.117	Mexico	0.146	Mexico	0.103
El Salvador	0.093	Guatemala	0.102	El Salvador	0.102
Honduras	0.089	El Salvador	0.092	Guatemala	0.1
China	0.021	Ecuador	0.075	Ecuador	0.081
Dominican Republic	0.016	Honduras	0.075	Honduras	0.064
Ecuador	0.012	Dominican Republic	0.028	Dominican Republic	0.029
Cuba	0.012	India	0.019	India	0.023
Jamaica	0.01	Nepal	0.018	Nepal	0.016
India	0.009	Colombia	0.015	Colombia	0.016
Others	0.132	Others	0.245	Others	0.246

Table 2: Proportion of successful deportations, top 10 countries of origin (fiscal year 2014)

try of origin. Using the spatial distribution of the total foreign-born population weighted by country of origin provided in the ACS, we determine what proportion of the total foreign-born population for each country of origin is present in each census tract. We then apply that proportion (again, for each country of origin in each census tract) to our estimate of the total number of unauthorized residents to determine the number of unauthorized residents by country of origin present in each census tract. This map can be found in Figure 1 below. Finally, we sum over the census tracts in New York City to estimate the total unauthorized population: 388,352.

4.2 Correlations

Before moving onto this step, we showed our map to various city and state agencies, as well as non-profit organizations and lawyers that work in the immigration space, to validate our results. Their reaction was resoundingly positive, and they confirmed that the map was in line with their intuitions. Some of these groups focus on particular nationalities (for example, Make the Road New York is a primarily Hispanic-facing organization), and they were excited to see a more comprehensive picture of the unauthorized immigrant population, one that went beyond what they were perhaps explicitly aware of in their own experience. After obtaining an estimate of the unauthorized population and validating our results within the community, we looked for spatial correlations between our estimate and various different datasets including 311 call data, tax returns, NYPD summonses, and building violation data.

We use monthly reports for 311 complaints about building code enforcements and other building development activities received by the Department of Buildings.⁶ We mapped the

⁶A full list of the complaint categories can be found [here](#).

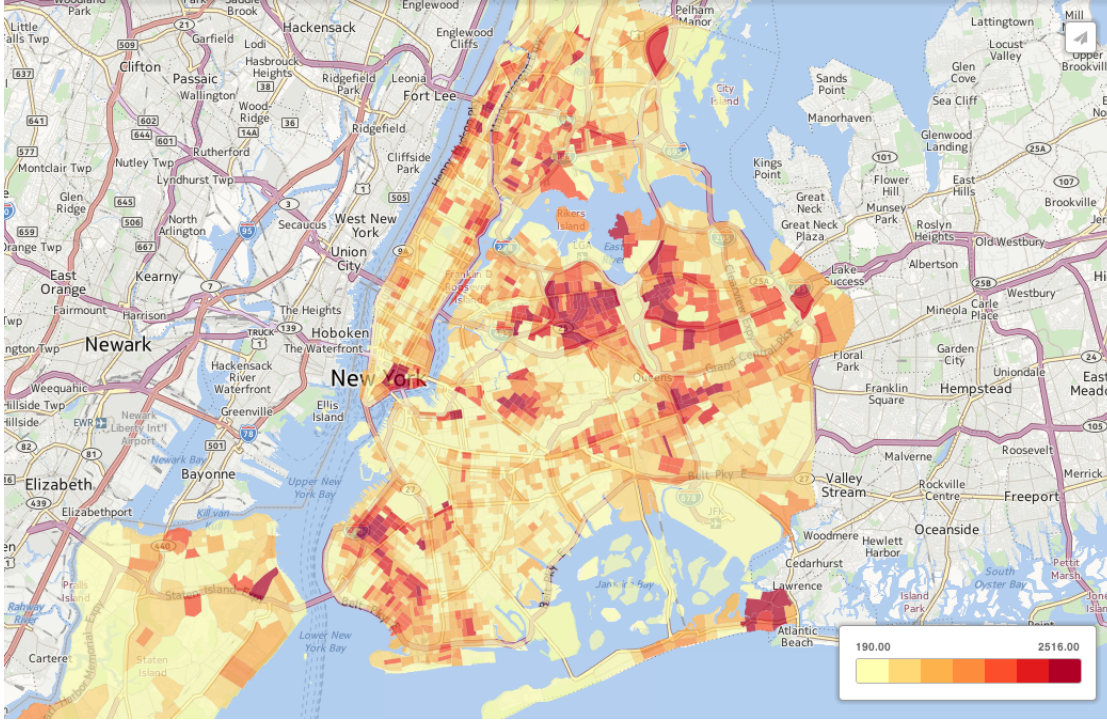


Figure 1: Density of unauthorized population by census tract

number of incidents per zip code, normalized to population. We measured the correlation between the number of incidents, the total population, and the estimated undocumented population to get a sense of which complaint types were most strongly correlated with the existence of undocumented residents.

We examined illegal conversions complaints specifically, on the intuition that incidents of illegal conversion might occur more frequently in areas with higher concentrations of undocumented residents. This intuition was not motivated by the correlation between building use complaints and the estimated undocumented population, and not borne out by the correlation between illegal conversion complaints and the estimated undocumented population. Figure 2 below shows the spatial or “geo” correlation of the total population and the estimated unauthorized population with different complaint types, arranged along the x-axis by rank. As you can see, there does not appear to be a steady signal of the unauthorized population in the data.

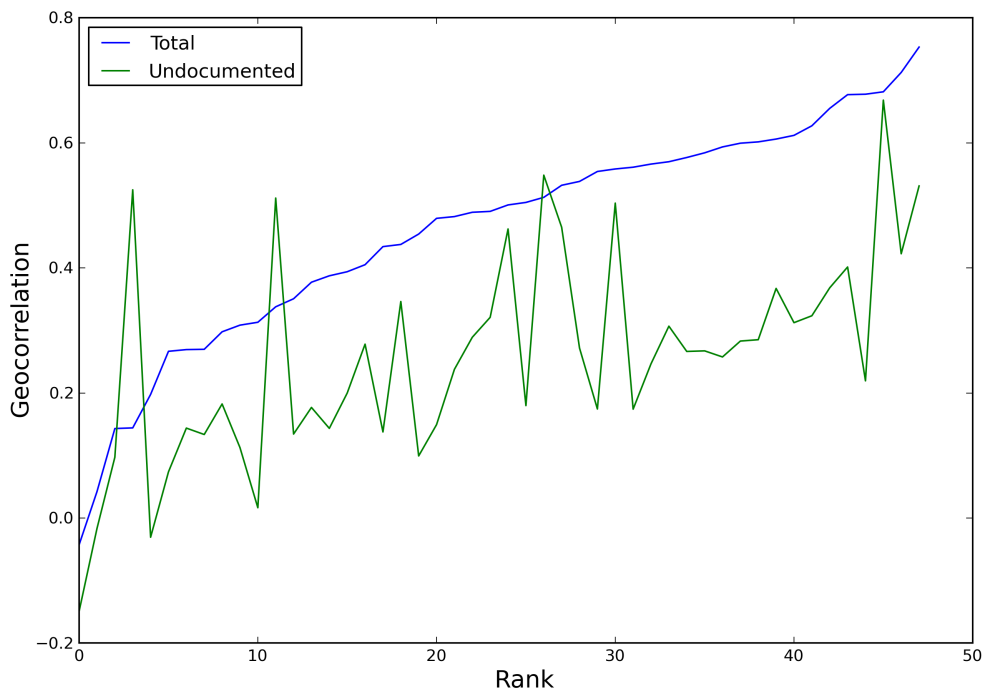


Figure 2: Geocorrelation with 311 Complaints by Type Rank

We mapped the number of NYPD summonses for unlicensed operators and unregistered vehicles for each precinct, normalized to the total number of summonses per precinct. Our intuition here was that neither a driver’s license nor vehicle registration can be obtained without identification. While this data looks promising, it is at very low granularity.

4.3 Predictive modeling

After getting a sense of the initial correlations, we tried to identify features that were predictive of high concentrations of undocumented residents. We used the residual estimation explained above as targets for our models. This is an unusual approach. But given confirmation by immigration activists that our map was largely correct, we wanted to see if there were other tracts in the city that had many of the same features, but weren’t picked up by the residual method. In order not to contaminate our predictions with the training data, we left out the borough of Queens during the training phase. Given the high diversity and concentration of immigrants in Queens, we figured that this was the most likely place for

hidden areas to emerge. The other four boroughs were used for training purposes.

We chose a Random Forest regressor to pick out important features in 311 call data. Random forests are good at handling large numbers of features and picking out which features are most predictive. Our hypothesis was that there might be a general class of complaints that was predictive of higher concentrations of undocumented residents. Unfortunately, there was almost no signal; the model relied heavily on total population from the ACS, but on little else from the 311 call data.

5 Social Impact

There are a bevy of services, both public and private, that require documentation. Access to libraries and neighborhood recreation centers both require some form of identification. Many doctors in hospitals will not authorize treatment in the absence of patient identification. Undocumented residents cannot apply for a legal driver’s license, excluding them from drivers’ education classes to learn the “rules of the road,” even as they drive and endanger others. And residents without identification may be hesitant to contact the police or other authorities, allowing crime and vice—like severe building violations—to continue without appropriate enforcement. Each of these services can be taken as evidence of the necessity of a municipal ID for New York City. However, it is important to understand the potential impact of a municipal ID on life as an unauthorized resident. Would a municipal ID address language, legal, and employment needs, which one could argue are the primary needs of undocumented residents?

One key potential of our map is in identifying staging locations, high foot traffic areas, and established community and civic organizations in areas of interest. One of the biggest hurdles that will have to be overcome is gaining the trust of these communities; airdropping into them is probably not the most effective way of engaging the undocumented population and having them volunteer very personal information about themselves. Utilizing already established networks (both private and public) is the surest way of gaining individuals’ trust and running the most effective program to distribute municipal IDs.

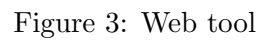
While our primary target is a forthcoming municipal ID outreach group, our insights are not limited to such a group. Over the course of the project, we met with several city and state agencies, as well as non-profit organizations, that operate in different parts of the immigration space. Each had a unique perspective on the situation of undocumented residents, and a unique opinion about how the insights of our project could have the greatest and most positive impact.

For instance, Joe Salvo at the New York City Department of City Planning told us that what the city really needs to roll out a municipal ID Program is numbers; they want to target the areas with the most people. The Office of New Americans has found that one of the primary needs of immigrants in New York City is learning English. To address this need, the ONA informed us that data on the languages spoken by the immigrant

population would have the greatest impact. Deportation lawyers told us that having data on the country of origin and number of dependents per immigrant adult would help them strengthen their cases for keeping their clients in the country. Estimating the number of dependents per immigrant adult, and having data on their documented status, which could be different, would be a major step toward improving our estimate.

We have designed our maps and made a web tool that addresses the needs of these different groups, and will make our data, estimates, and maps available to them. Because of the high sensitivity of the information, we will not be making these widely available. We have used only data that is publicly available, and thus it is possible for anyone to recreate what we have done here. There is reasonable concern, for example, that the Immigration & Customs Enforcement Office, upon seeing our map, might target their deportation efforts. As this is counter to our project goals, and the goals of Mayor DeBlasio's municipal ID program, our public-facing webpage contains only general information about the project, and our contact information so individuals can reach out to us about acquiring the data and a more detailed description of our methodology and derivations. The public-facing web tool can be found [here](#). In closing, Figure 3 below is a screenshot of the web tool we have made available to city agencies, non-profit organizations, and researchers.

Alex Chohlas-Wood, Frank Zha, Gang Zhao, Jane Stewart Adams
NYU CUSP



References

- [1] Hoefler, M., N. Rytina, & B. C. Baker. (2012). “Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2011.” Office of Immigration Statistics, Policy Directorate, U.S. Department of Homeland Security. http://www.dhs.gov/xlibrary/assets/statistics/publications/ois_ill_pe_2010.pdf
- [2] Arias, E., L. R. Curtin, R. Wei, & R. N. Anderson. (2008). “U.S. Decennial Life Tables for 1999-2001, United States Life Tables.” *National Vital Statistics Report* 57(1), National Center for Health Statistics, Centers for Disease Control. http://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_01.pdf
- [3] Warren, R. & J. S. Passel. (1987). “A Count of the Uncountable: Estimates of Undocumented Aliens Counted in the 1980 United States Census/” *Demography* 24: 375–393.
- [4] Ahmed, B. & J. G. Robinson. (1994). “Estimates of Emigration of the Foreign-Born Population: 1980-1990.” Technical Working Paper No. 9, U.S. Bureau of the Census. <http://www.census.gov/population/www/documentation/twps0009/twps0009.html>
- [5] U.S. Bureau of the Census. “Foreign-Born Population Frequently Asked Questions.” Accessed 25 March 2014. <https://www.census.gov/population/foreign/about/faq.html>
- [6] “Text of Bill DeBlasio’s First State of the City Address.” *The New York Times*. Last updated 10 February 2014. http://www.nytimes.com/2014/02/11/nyregion/text-of-bill-de-blasios-first-state-of-the-city-address.html?_r=0