Florida
Data Science
for Social Good

Florida Philanthropic Network

Developing a Methodology to Estimate Census Undercounts in Florida Counties

November 2023

Table of Contents

Executive Summary

Every decade, the United States Census Bureau attempts to count or estimate the number and characteristics of the U.S. population. The U.S. government uses these estimates to determine the level of federal funding that is allocated to each state. If the population within a state is underestimated, then funding resources to address critical social needs within the state are limited. The Florida Philanthropic Network (FPN) is working to understand the potential Census undercount in Florida and to minimize the undercount in the 2030 census. Using estimates by Florida Tax Watch, **demographic data provided by the U.S. Census, the Opportunity Atlas, Florida voter registration and participation, data from the Florida Nonprofit Alliance, as well as methodology developed by the Texas Census Institute**, the Florida Data Science for Social Good (FL-DSSG) estimated the levels of undercount for each Florida county. Dimensions used to estimate the undercount include Personal, Geographic, and Census Campaigning. **Rural counties tended to experience a greater percent undercount than larger counties**. The exception was Miami-Dade County, which received a high percentage undercount and a high total population undercount based on its population size. **Counties with the largest population showed the highest undercount population values, including Miami-Dade, Broward, Palm Beach, Orange, Hillsboro, and Duval**. Patterns related to the "digital divide" in rural counties and social integration in larger counties were explored. FL-DSSG has developed an interactive dashboard and data estimates for FPN to help with state-wide planning to mitigate the potential undercount in the 2030 census.

Tableau Dashboard of Florida Census Undercount Estimates: https://tabsoft.co/3u7xYKI

DSSG

The ability to affect change and do good in one's community increasingly depends on having the right information at the right time to make the right decisions about the most important things. Directors of community programs and funding agencies want evidence of impact and demonstrated efficiency in programs that serve our communities. Often, the information available to meet these needs is not well organized, poorly understood, and not packaged in a way that helps those working in the community do their best.

The Florida Data Science for Social Good (FL-DSSG) Program is a summer internship program that matches data science expertise with real-world problems. The FL-DSSG program works with community organizations that are trying to affect change in their communities and who have data management, analysis, and data visualization projects that have the potential to shift understanding around a community issue, influence planning, revise practices, or see efforts in supporting community

initiatives more focused or renewed. The FL-DSSG aims to promote data-informed approaches among nonprofits to solve wicked social problems while creating educational programs for aspiring data scientists.

"We are social trustees of knowledge with a unique capacity to do social good."

Wicked Problem

The Florida Philanthropic Network (FPN) approached the Florida Data Science for Social Good (FL-DSSG) to address the challenge of determining the probable undercount in counties in Florida during the 2020 Census. Census undercount refers to the population that was not accurately counted in the Census, where the population reported by the Census for a demographic characteristic was less than the actual population in that area. Having an undercount within a state or county has significant implications for federal funding. Florida Tax Watch estimates that Florida's population was undercounted by 3.48% in the 2020 Census[1], potentially resulting in the loss of up to $11 billion[2] in federal funds by the end of the decade. To prevent a similar undercount in the 2030 Census, FPN and its partners aim to understand the following questions:

1. Who was missed in the 2020 Census?
2. Where were they located, and what was their demographic profile?
3. What was the method of response for those who were counted?

The current methods employed by the U.S. Census Bureau primarily focus on estimating undercounts at the state level, rather than at the more granular county level. Therefore, developing a methodology to accurately determine the undercounts at the county level becomes a crucial task. This information is vital for identifying areas that require additional community partnerships, conducting targeted focus groups to refine messaging, and helping counties, municipalities, and other groups understand the significance of a complete count in future census endeavors.

Partner Perspectives

The Florida Philanthropic Network (FPN) played a crucial role in establishing a collaborative effort aimed at increasing awareness and participation in the 2020 Census. FPN's staff members have been actively involved in Census work at the state-wide level for many years. They continue to engage in fundraising, awareness-raising activities, and leading bimonthly meetings related to the Census. FPN has provided over $130,000 in grants to support community organizations' ongoing involvement in this significant civic engagement work.

During FPN's annual summit in February, they organized a session emphasizing the importance of philanthropy's sustained engagement in the Census. Civic engagement and the Census are key public

[1] U.S Census Bureau (2022), https://www.census.gov/library/stories/2022/05/2020-census-undercount-overcount-rates-by-state.html
[2] Florida Tax Watch (2022), https://floridataxwatch.org/Research/Full-Library/looking-back-at-census-2020-4

policy priorities for FPN, closely tied to one of the primary pillars of their current strategic plan. FPN is fully committed to making this work a long-term organizational priority.

To enhance their efforts, FPN reached out to the Texas Census Institute. The Institute had conducted similar research looking into the undercount in Texas. FPN gained insights from their experiences and subsequently began collecting additional data on various factors related to the Census. With the assistance of the FL-DSSG program, FPN aims to understand and communicate with visuals the potential undercounts in each County in Florida. These visual representations of undercounts will be instrumental in creating awareness programs, helping FPN further its objectives.

Data Sources

Initially, we received data from FPN, which included data sources from the U.S. Census Bureau. We started with these datasets:

- B98021-Housing unit response and non-response unit,
- DP05-ACS demographic and housing estimates
- Florida self-response data 2020-2010.

In addition to the data received, FL-DSSG reviewed the Texas Census Institute research methodology[3]. From the dimensions that were relevant in Texas, we started to gather data that would be suitable for Florida. We accessed additional public data from the U.S. Census Bureau. We included data from Demographic and Housing Estimates (DP1) and the American Community Survey, as well as non-response data and race/ethnicity populations (to estimate racial population integration across block group boundaries). As income and mobility were of concern in Florida, we accessed public data regarding Asset-Limited, Income Constrained, Employed (ALICE) reports[4]. Working with FPN, we understood the importance of non-profit organizations, civic engagement, and philanthropy. We explored philanthropic giving and the number of non-profit organizations using data from the Florida Non-profit Alliance[5]. To address civic engagement, we accessed voter registration and voter turnout data from the Division of Elections in the Florida Department of State[6]. We accessed economic connectedness data from the Opportunity Atlas[7] to address social mobility and engagement issues.

Using these data sources, we were able to estimate the undercount for each Florida county. We used a similar methodology as the Texas Census Institute, focusing on three main dimensions: Personal, Geography, and Census Campaigning. Below are the data sources for the same.

[3] Castellanos-Sosa, F.A. (2022). Undercounting and Overcounting: Population in Texas Counties A Determinants-Side Approach and its Application to Texas. http://texascensus.org
[4] United for ALICE. (2023). https://www.unitedforalice.org/essentials-index
[5] Florida Nonprofit Alliance. (2023). Florida's Nonprofit Sector. https://flnonprofits.org/page/Research
[6] Florida Division of Elections. (2023). https://dos.fl.gov/elections/data-statistics/elections-data/
[7] Opportunity Atlas. (2023). https://www.opportunityatlas.org/

Personal Dimension:

Civic Engagement and Social Capital were the main focus of this dimension, using data sources: Florida Nonprofit Alliance (Philanthropic Giving, Number of Nonprofits per Population), ALICE Report (Percent in ALICE category), Voter Registration, Voter Participation, the Opportunity Atlas (Economic Connectedness), and Demographic Estimates (DP1) to calculate Racial Integration.

Geographic Dimension:

Being Difficult to Reach and Difficult to Register were the main concerns of this dimension. The U.S. Census Bureau was our main source for this dimension: DP1- Profile of General Population and Housing Characteristics, Local Update of Census Addresses (LUCA), and Unable to Geocode (estimates new housing).

Census Campaigning Dimension:

Non-Response Rate and the Digital Divide were the main focus areas for this dimension, using data: U.S. Census Bureau B98021- Census Response Rates (Non-response Rates), and B28002-Internet Access (digital divide).

Data Cleaning

Since undercounting was an issue in the 2010 census, we gathered data for 2 decades, that is, 2010 and 2020. After going through various data sources, we have made master files for 2010 and 2020, and a metadata file that includes information and data sources for each parameter. Most of the factors of interest to calculate the dimensions were missing in the 2010 census, so we limited our Masterfile to 2020 data, allowing for the complete implementation of our methodology. For data cleaning and merging, we mostly used Excel and Python.

Personal Dimension:

Personal Dimensions were often reported in percentages (e.g., Voting Registration Percent, Voting Participation Percent). We did estimate the number of non-profit organizations and philanthropic giving per population within the County. Racial Integration was calculated using a procedure developed by Iceland (2004)[8] for multi-group entropy.

Geography Dimension:

1. For population density, data was available in the required format. Hence, we just merged it directly to our CSV file.
2. For population share, we calculated it using the below formula:

[8] Iceland, J. (2004). The Multigroup Entropy Index (Also Known as Theil's H or the Information Theory Index). https://www.census.gov/topics/housing/housing-patterns/about/multi-group-entropy-index.html

Population share = Total population (county) / Total population (State)

3. For addresses unable to locate factor, each county had a file associated with it, so there were 67 different files.

Census Campaigning dimension:

The data for Census campaigning were reported in percentages of households. We merged these measures directly into our CSV file.

After selecting all the measures in each dimension, we bounded the cleaned data, controlling for outliers. Outliers were determined by visual inspection of the univariate distributions of each measure.

Methodology

We started familiarizing ourselves with the Texas Census Institute research paper on the undercount topic to know how much it related to Florida. From this research, we adapted the 3 dimensions: personal, geography, and census campaigning. For the personal dimension, we chose Non-profits, Economic connectedness, and voter registrations as they represent social capital and civic engagement, which will be useful in determining undercount. The factors that make a location easy to reach and accurate in The Master Address File (MAF) make up the geographic dimension. Using population density and the percentage of the state's population that each county represents, we can roughly estimate the Easiness-to-reach dimension. The Master Address File's accuracy in accurately identifying each housing unit is another feature of the geography dimension. We calculate this factor using the percentage of housing units that the U.S. Census Local Update of Census Addresses (LUCA) could not geocode. The third dimension incorporates marketing strategies and internet/technological accessibility. We selected nonresponse rates and internet accessibility as the contributing factors to this dimension. After conducting distribution and correlation analyses using the data we obtained for Florida, we finalized these features.

We calculated the Z-score of each measure with the bounded data. If X is a measure of a dimension, we calculated Z-score by the following formula:

$$X = \frac{\pm(X - mean(column\ X))}{standard\ deviation\ (column\ X)}$$

Here, $\pm$ means the undercounting impact of each measure. For example, 'Voter's turnout percentage' negatively impacts undercounting; the more voters turn out, the less undercount will occur. Similarly, 'Percent_ALICE' has a positive impact on undercounting. The larger the share of the asset-limited and income-constrained population, the higher undercounting value is likely. We worked on the Z-

score column to measure the undercount value for each measure that results in each dimension. The calculation is shown below:

Personal Dimension: $\frac{1}{2}$ (Civic Engagement undercount value + Social Capital undercount value)

[Civic Engagement undercount = ½ (Voter Turnout Z-score + Nonprofit Z-Score),
Social Capital undercount = 1/3 (Economic Connectedness Z-score + Percent ALICE Z-score + Racial Integration Z-score)]

Geographic Dimension: $\frac{1}{2}$ (Difficult to Register undercount + Difficult to Reach undercount)

[Difficult to Register undercount = Unable to Geocode Z-score,
Difficult to Reach undercount = ½ (Population Share Percentage Z-score + Renters Percentage Z-score)]

Census Campaigning dimension: $\frac{1}{2}$ (Non-Response Rate undercount + Digital Divide undercount)

[Non-Response Rate undercount = ½ (No One Home Percentage Z-score + Maximum Contact Attempts Reached percent Z-score),
Digital Divide undercount = Without Internet Percentage Z-score]

After calculating each dimension, we merged these three dimensions to calculate our final undercount. The final undercount was estimated using the below formula:

Undercount Index by County = 1/3 [Personal Dimension + Geographic Dimension + Census Campaigning Dimension]

After getting the undercounting value by county, we estimated the population undercount for each county.

Undercounted Population Estimate by County:

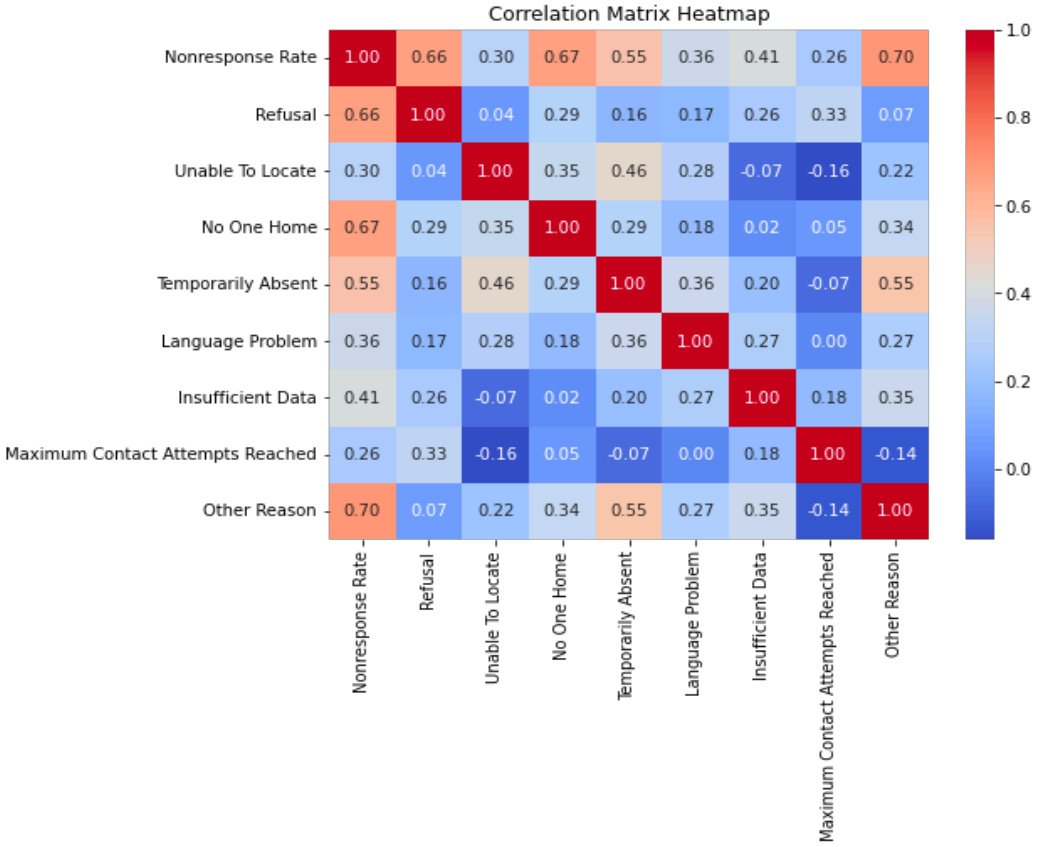
$$CLI = 0.0348 * Undercount_{index}$$

$$Undercount_{Population\_county} = 0.0348 + CLI$$

We used the CLI value by multiplying the undercount index with the Florida undercount estimate of 3.48% or 0.0348. This value gives us the proper estimated population by each county by another formula where we added the CLI value with 0.0348 to get the proper population estimate.

Analysis and Findings

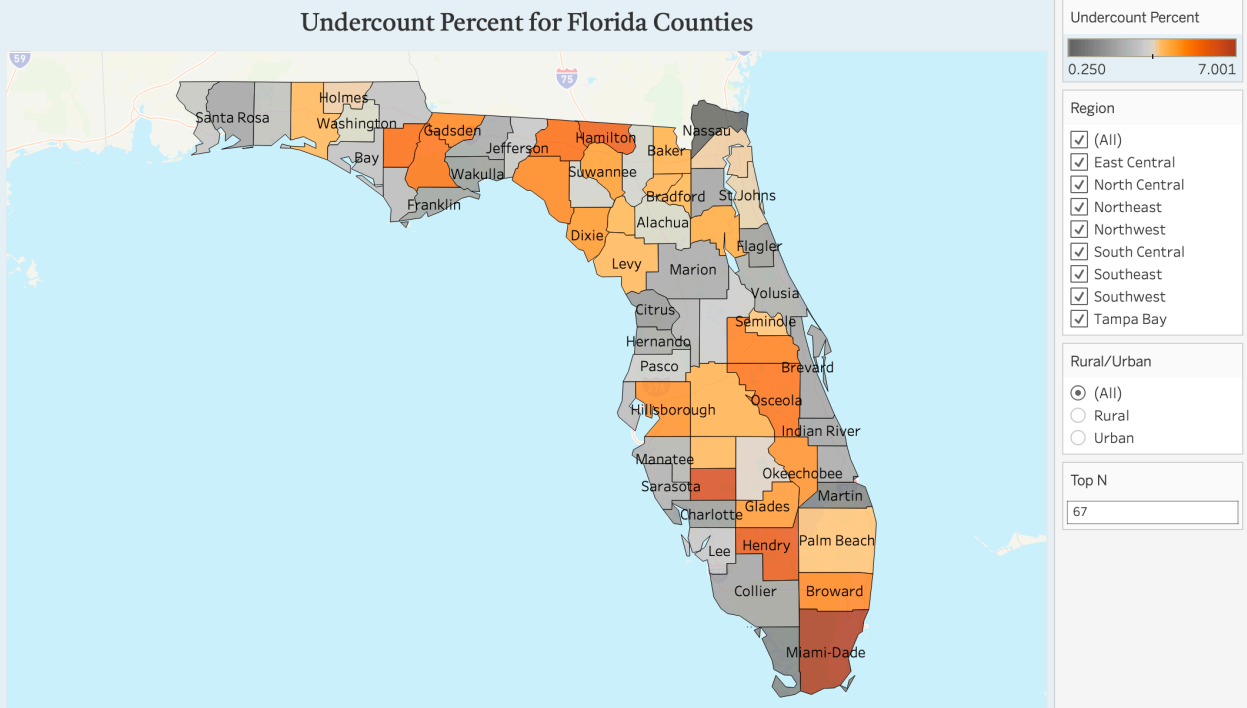
Our first analysis step was to find the right variables for the Florida undercounting. Hence, we calculated correlations between the chosen factors from all our data sources. Finally, we selected the resulting factors by looking for measures that were associated but not overlapping. For example, for the Census Campaigning Dimension, the Nonresponse Rate positively correlates with the Refusal rate, No-one Home rate, Temporarily Absent, and Maximum Contact Attempts Reached. We also performed univariate analysis and could see that the Temporarily Absent Parameter had outliers. Hence, we finalized our dimensions with the other three factors.



Once all the variables were selected and we estimated the undercount percent and count for each county, we moved to Tableau to analyze undercounting patterns in each county. We made multiple visualizations representing undercount estimate, undercount percent, factor-wise maps, dimension-wise undercount rate, and correlation scatterplots for factors.

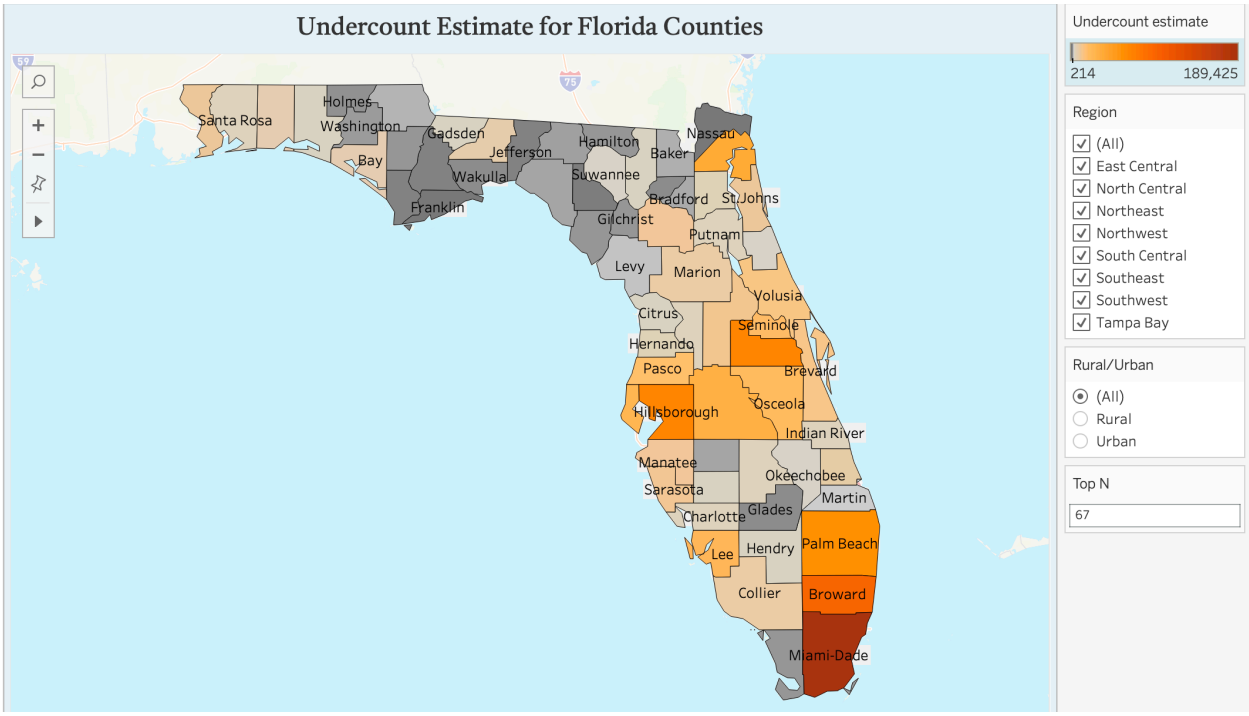
Undercount Percent:

The below figure shows the undercount percent visualization. The key findings from the undercount percent are Miami Dade has the highest undercount percent, and Nassau has the lowest percent value. North Central, Southeast, and Tampa Bay regions have the highest percent undercount.



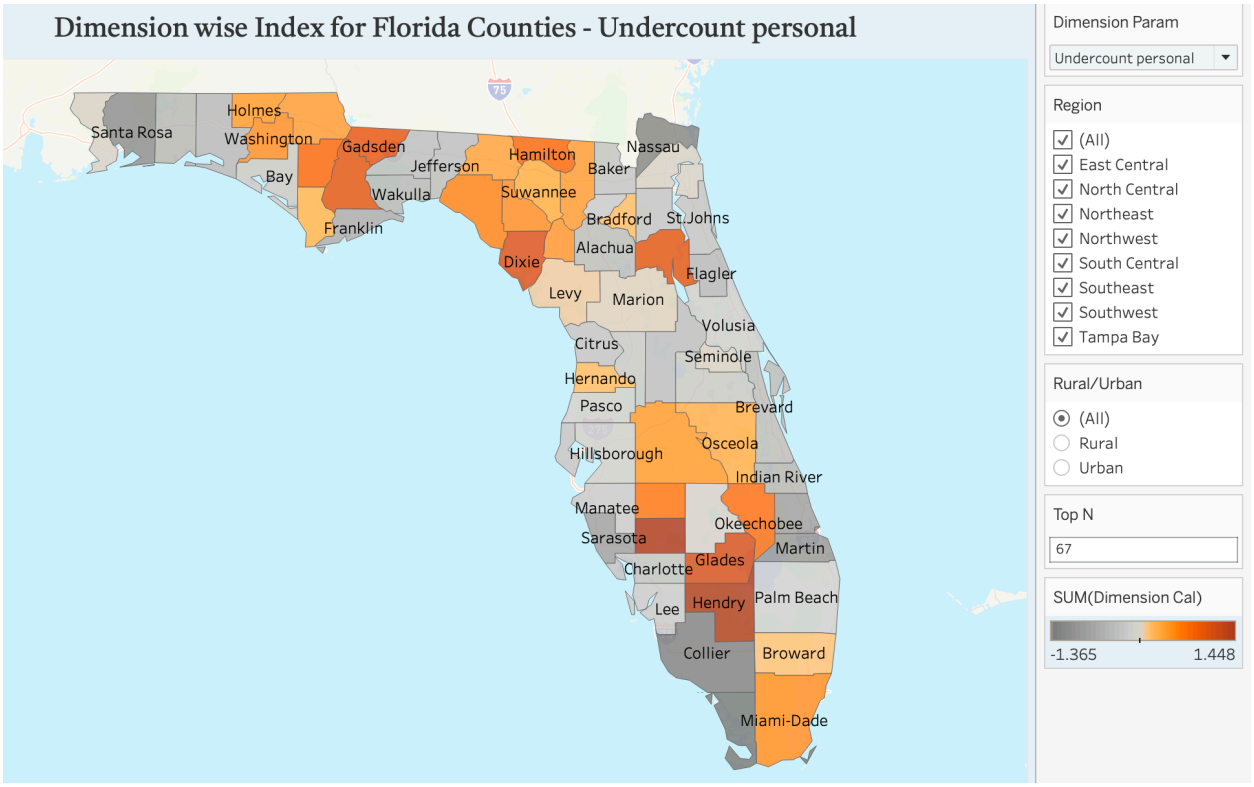
Undercount Population Estimates:

The below figure shows the undercount population estimates visualization. The key findings from this visual are Miami Dade and Broward counties have the highest population estimates for undercounting, and Nassau has the lowest estimates. Most Southeast, South Central, and Tampa Bay regions show the highest undercount population estimates.



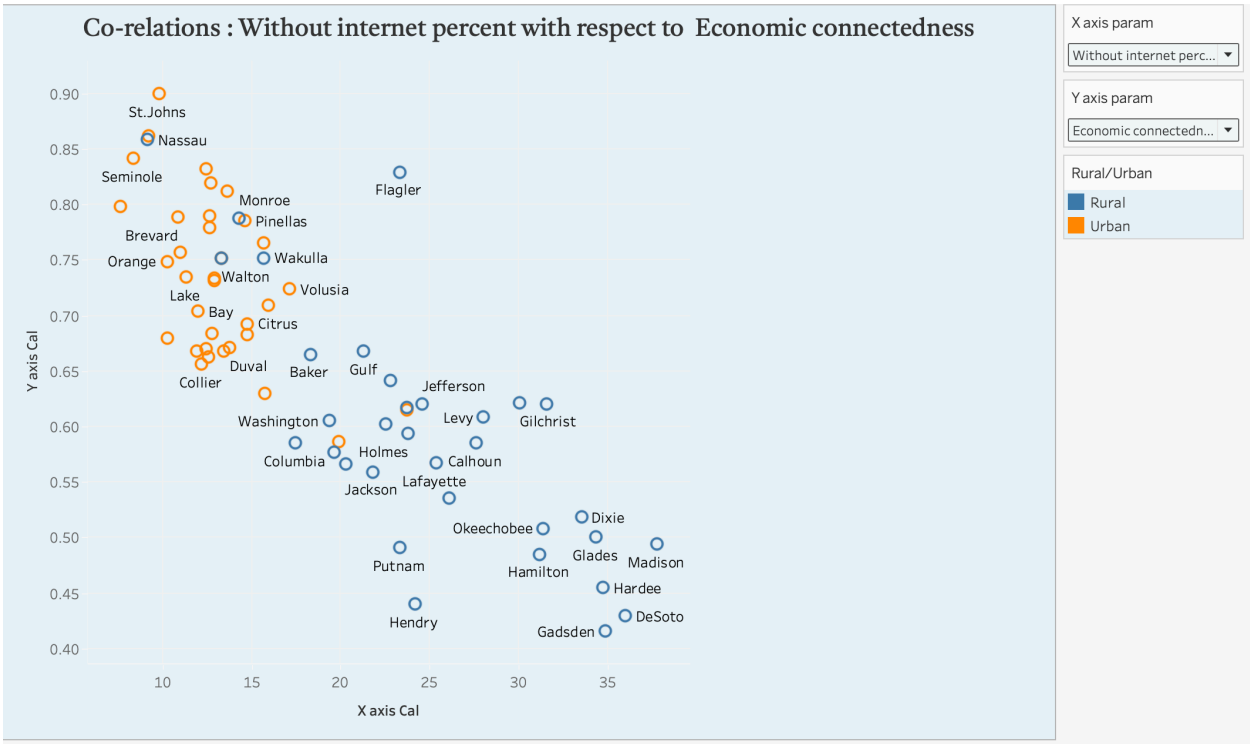
Dimension-wise Undercount Rates:

For each of the three dimensions of the undercount estimates, we created visual displays of each dimension to highlight different aspects of undercounting that FPN could address. The map below is an example of the Personal Dimension. In this dashboard, we have created a parameter called "Dimension Param" in which the user can choose the focus dimensions for the visual analysis.



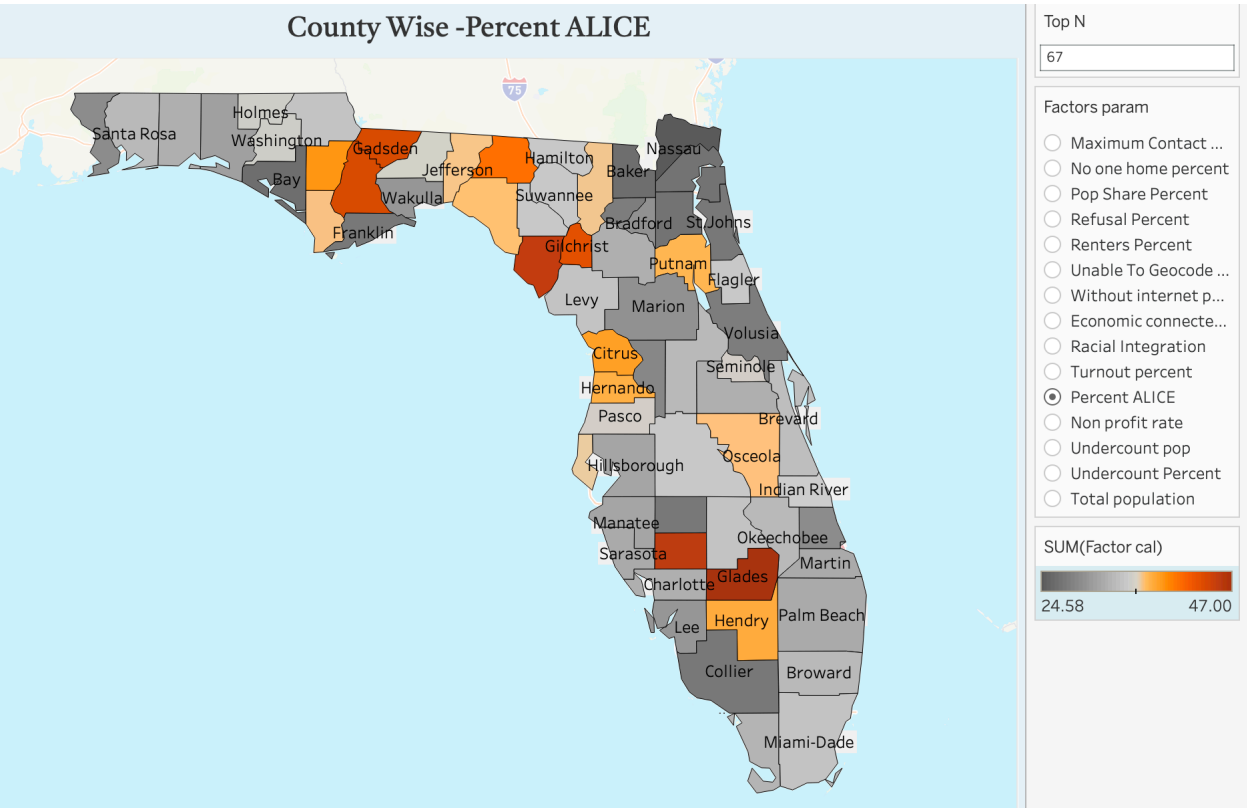
Correlations:

We also created a correlations dashboard for users to explore interrelationships among the factors that comprised each of the three primary dimensions. For example, in the next visualization, the scatterplot includes the percentage of households without internet and the level of economic connectedness within the same County. The graph shows a negative correlation, indicating that as the number of individuals without internet access in a County increases, the economic connectedness of individuals within that county decreases.



Factor-wise Analysis:

In addition, to allow further exploration of factors that contribute to undercounting, we created a visualization that allows for mapping the factors by county.



Conclusions and Recommendations

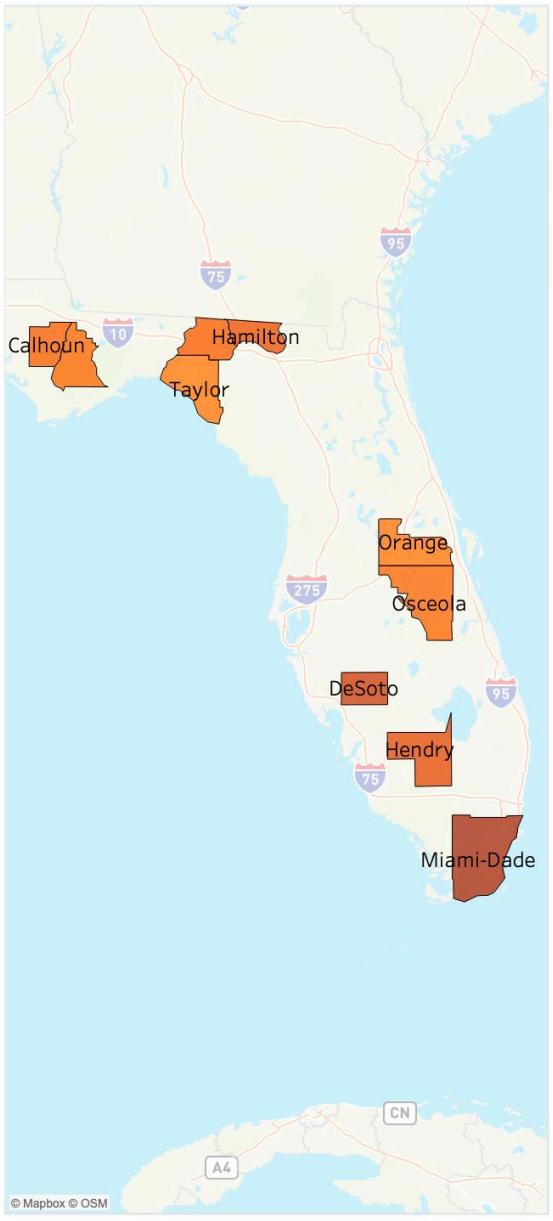
From the resulting map (below), the top 10 Counties with the highest undercount percent are Calhoun, Hamilton, Taylor, Madison, Liberty, Orange, Osceola, DeSoto, Hendry, and Miami-Dade. Counties with the highest undercount population estimates are Duval, Orange, Osceola, Polk, Hillsborough, Lee, Palm Beach, Broward, and Miami-Dade.

Both maps should be prioritized, and necessary measures should be taken to address the undercounting issues. Also, using the dimension-wise analysis, we saw that:

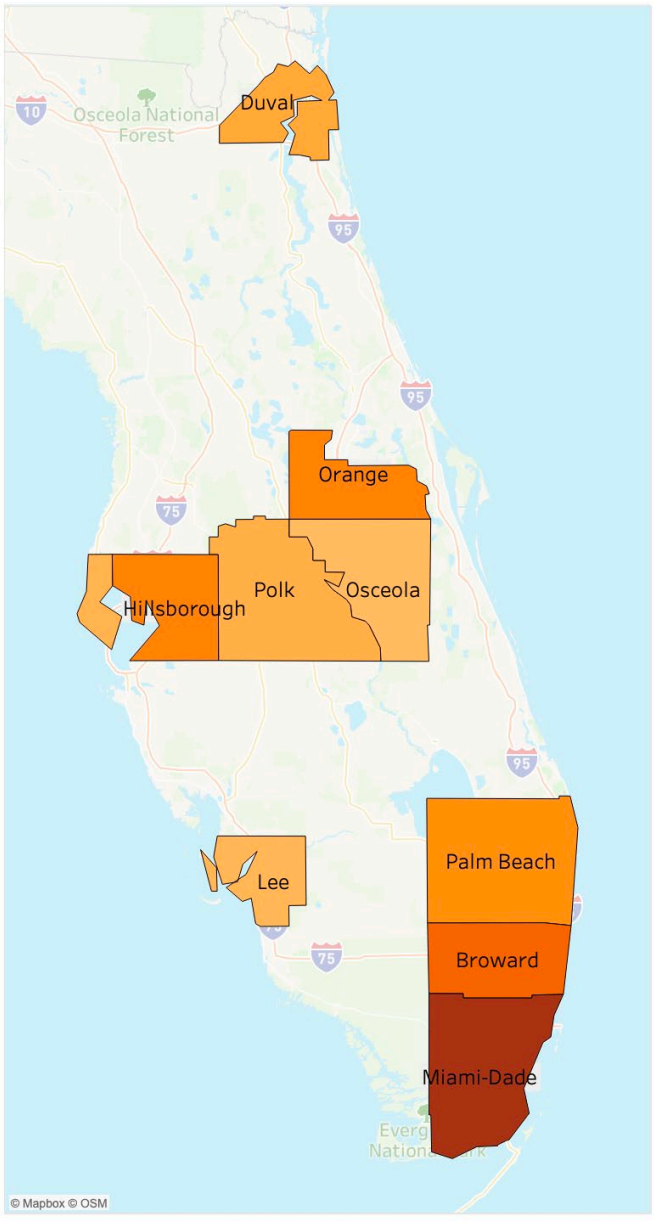
1. For the Personal Dimension: Most Northwest counties had higher undercount rates. We recommend the focus in these counties should be on media campaigns building trust and social cohesion to support Census participation.
2. For the Geography Dimension: Most urban counties had higher undercount rates. For urban and high-growth counties, FPN can work with County and municipal agencies to support registration of addresses in a timely fashion.

3. For the Census Campaigning Dimension: Most rural counties had higher undercount rates. For rural counties, FPN can work with non-profit and for-profit partners to address the digital divide, increasing internet access and economic mobility.

Undercount Percent for Florida Counties



Undercount Estimate for Florida Counties



Next Steps

With these analyses and data visualization tools, the Florida Philanthropic Network will have the information needed to make critical decisions in the coming years leading up to the 2030 Census on how best to address the Census undercount in different regions and Counties. The availability of

visualizations will help FPN develop strategies in each of the highly affected Counties and communicate with their partners throughout Florida. FPN is taking the steps necessary to form a large and sophisticated network of partners to address the Census undercount in Florida, with hopes that undercounting will be mitigated or eliminated in 2030. Addressing the underlying causes of the undercount will help FPN maximize resources in the future.

Get More Information

For additional information about the FL-DSSG program, its process, methodologies, tools, and techniques, and how to participate in the program, contact FL-DSSG program directors. Visit the FL-DSSG website for contact information.

FL-DSSG interns work on wicked problems that require technical know-how and relevant domain knowledge. To help interns complete the projects, we obtain mentors from industry professionals and research faculty members. All mentors offer their assistance voluntarily. We deeply appreciate the assistance provided by mentors.

FPN DSSG Project Resources

Project Presentation YouTube Video: https://youtu.be/qRgmHScrtBo

2023 Big Reveal Presentation Slides: https://bit.ly/23FLDSSGBigRevealSlides

Tableau Dashboard of Florida Census Undercount Estimates: https://tabsoft.co/3u7xYKI

2023 FL-DSSG Interns

Partha Protim Datta, Master of Science – Data Science, University of North Florida
Patrick Harrell, Bachelor of Science – Psychology, University of North Florida
Vinaya Rajaram Nayak, Master of Science – Business Analytics and Information Systems, University of South Florida
Aishwarya Pawar, Bachelor of Science – Data Science, University of North Florida
Sri Ram Sripada, Master of Science – Business Analytics and Information Systems, University of South Florida
Yuanyuan Yang, Doctor of Philosophy – Social Work, Washington University in St. Louis
Amanda Yelverton - Bachelor of Science – Psychology, University of North Florida

2023 Sponsors

NLP Logix, Jacksonville Jaguars, Community First Cares Foundation, Cathedral Arts Project, Florida Philanthropic Network, Miller Electric Company, MarketOnce, HC Brands, PGA Tour, ADVOS Legal, iVenture Solutions, and Leanovation Labs

FL-DSSG Program Directors

Dr. Dan Richard, Associate Professor, Psychology Department
Dr. Karthikeyan Umapathy, Associate Professor, School of Computing

Florida Data Science for Social Good (FL-DSSG)
University of North Florida
November 2023
https://dssg.unf.edu/
https://www.facebook.com/FLDSSG/