Projecting the 2020 Census: A Study of Hard to Count Communities

**Abstract**

This paper utilized Census Bureau data to examine the … The results of this project found no correlation between 2010 Census response rates and Texas residents broken down by race and ethnicity.

**Background**

The purpose of this project is to use US Census data and trends to analyze if there is any similarity in demographic information (e.g., race and ethnicity.) for areas of Texas (e.g., Census tracts) that are labelled as “Hard to Count Communities” in. This project examined publicly available data from the Census 2020 Hard to Count map by the City University of New York. This data was originally provided by U.S. Census Bureau from the previous decennial Census in 2010 and from American Community Survey (ACS) population estimates from 2013-2017. Specific data includes mail-in response rates from the 2010 Census and demographic population information and internet access from the 2013-2017 ACS.

With the approach of the 2020 Census and the concern about a potential citizenship question being added to the Census, there is a large amount of concern over obtaining an accurate and clear Census count. This research question is important as Census data has far reaching implications. For example, federal funding for states and local governments is allocated based on Census data (citation). Census data is also used in planning for population growth and development, like investments in infrastructure, transportation, and education. In Texas specifically, Census data is incorporated into determining state legislative and Congressional districts and representation, a practice that has been questioned in relation to ethics (citation).

Populations that have historically been undercounted in the US Census tend to include communities of color, immigrant households, households with limited English proficiency, households of lower socioeconomic status, children, and rural communities (citation). This undercounting then goes on the disproportionally impact these groups in a myriad of ways. As a result of [Include theoretical framework in paper – transformational theory; critical race theory; advocacy coalition framework?]

**Research Question**

The research question for this project is what relationship do demographic factors have to Census 2010 response rates in Hard to Count Communities? For the purposes of this study, only the demographic variables of race and ethnicity for the total population will be measured. Race and ethnicity were studied as the independent variable while mail-in response rate to the 2010 Census was the dependent variable.

The hypothesis for this question was that response rates would have a positive correlation with white communities and a negative correlation with populations of color based on literature about Hard to Count Communities and who usually is at risk of being undercounted in the Census.

**Methodology**

In its initial form, this project intended to merge and compare the previously mentioned Texas HTC data with another dataset that contained information about Texas general elections. To acquire the data needed for this project, a search for this data on the *Texas Secretary of State*’s website occurred, however, the available data was not disaggregated according to demographic information.

The *United States Elections Project* (McDonald) site was then searched for a data source that contained voter turnout information, however, the Voting Age Population data for “State Turnout Rates” was not disaggregated by race and ethnicity, or by Census tract.

The IPUMS website was also visited to search for panel data that may have contained demographic voting data, however, this data was also not sufficient for the confines of the original research question.

After consulting with the course instructor and teaching assistant, it was decided to restructure the project focusing on the data source from CUNY’s Center for Urban Research. Instead, the project would look at different categorical variables (e.g., population demographic information) and examine how they were correlated with mail-in response rates for the 2010 Census.

**Data Management Plan**

With the use of only one dataset, this simplified the original data management plan significantly. Many of the previously planned for steps were no longer applicable to the process since there would no be a merging of data sets.

**Data Acquisition**

**Workflow and Replication**

* Data sources and complied data sets
  + IPUMS panel data – search demographic voting data
  + Census response rates by demographics (and county?)
* Data dictionary explaining merged data files
* Github and readme file
* Visualizations?
  + Census 2020 Hard to Count Map – <https://www.censushardtocountmaps2020.us/>
  + Census Engagement Navigator – <https://www.census.gov/library/visualizations/interactive/engagement.html>
* Run regression in Stata
* Codebook
* Talk about how project has changed in workflow

**Validation**

* Challenge of validating Census data

**Limitations**

A major limitation of this project is that presence of bias in Census data collection. Census data is subject to nonresponse bias from populations that are undercounted. communities that are undercounted don’t receive benefits; non-representative results

Selection bias is also an issue in Census data. In communities with low response rates, the sample may not be representative of the overall population. Furthermore, the next decennial Census is at risk of experiencing more bias and potential undercounting with the possible inclusion of a citizenship question. This question may lower the response rate for communities that have larger immigrant populations or non-US citizens.

Another limitation is the challenge of validation Census data. Since the data is so large, the raw data is largely inaccessible, which the Census only publishing what it wants.

Census Bureau’s website transition. This transition may explain why may links to data on the Census site were broken or “could not be found”

**For Future Studies**

The original purpose of this project was to investigate the relationship between response rates and demographic characteristics for communities labeled as “Hard to Count” in Texas and data on voter turnout (in both midterm and presidential elections). This paper would have discussed the findings upon comparison of this data, speaking on the ethical implications of the results, and examining any limitations of this process, as well as potential solutions to any challenges that may arise or concerning trends found. Any correlations would have been examined to see if there was a connection between reported voter disenfranchisement and suppression in Texas. However, voting data in Texas (to the desired extent) was unattainable within the parameters of this project.

In the future, available data would

**Discussion**

The utilization of Census data is important to examine

* Policy implications
  + Complete Count Committees to increase response rates
  + Redistricting and gerrymandering
  + Funding

**Conclusion**

References

Center for Urban Research, CUNY Graduate Center. (2017, September). Census 2020 Hard to

Count Map. Retrieved from <https://www.censushardtocountmaps2020.us/>

Center for Urban Research, CUNY Graduate Center. (2017, September). *Tract Data – TX Excel*

*file* [Data file]. Retrieved from <https://www.censushardtocountmaps2020.us/>

Center for Urban Research, CUNY Graduate Center. (2019, March 3) Census 2020 Hard to

Count Communities in Texas. Retrieved from

https://www.censushardtocountmaps2020.us/img/mappdfs/Texas.pdf

McDonald, Michael P. (2019). "Voter Turnout: State Turnout Rates." *United States Elections*

*Project*. Retrieved from: <http://www.electproject.org/home/voter-turnout/voter-turnout->

data/

US Census Bureau, [*American Community Survey, 2013-2017*](https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t), tables B03002, B02009-B02012.

**Appendix**

A github repository containing all files, folders, a README file, a Google Colab document of the Python code, lab notebook of STATA commands, and graphs can be found at <https://github.com/mirandabad/AEM_final_project>.