Final Project Paper Draft

**Intro/Abstract**

**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, ethnicity, income, etc.) for areas of Texas (e.g., counties, zip codes, or Census tracts) that are labelled as “Hard to Count Communities” and communities in Texas that have experienced voter suppression and disenfranchisement. This project [will examine] publicly available data from the US Census Bureau from the previous decennial Census in 2010 and from American Community Survey estimates that have been administered since the 2010 Census. This data [will be compared] data on Texas general election. This data would ideally come from the Texas Secretary of State, however, this depends on the availability of datasets. My paper will largely be a discussion of findings comparing this data, speak on the ethical implications of the results, and examine any limitations of this process as well as potential solutions to any challenges that may arise or concerning trends found.

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. 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.

**Research Question**

Question (still refining): What impact does voter suppression have on response to government survey (i.e., decennial Census)?

Hypothesis (still refining): Demographic information for Hard to Count Communities (i.e., race, ethnicity, socioeconomic status) will be similar to demographic information for communities that are disenfranchised from voting?

I would hypothesize that there is a positive correlation between communities identified as harder to count and areas/populations that have had a harder time voting. Populations that have historically been undercounted in the US Census tend to include communities of color, immigrant households, households with limited English proficiency, and households of lower socioeconomic status. Research has shown that these same populations are more likely to be subject to stricter Voter ID laws in Texas and are less likely to vote.

**Methodology**

* Download data sets from US Census Bureau for Census data and compare to Texas voter data (from Pew Research Center?)
* Merge necessary data sets to compare info

**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>

**Validation**

**Critiques/limitations**

* Bias in Census data collection
  + Selection bias

**Discussion**

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

**Conclusion**