

University of Mississippi

eGrove

---

Population Briefs

Center for Population Studies

---

11-25-2020

## The Potential Impact of the Census Bureau's Differential Privacy System: Understanding Characteristics of the Mississippi Population

Jamiko Deleveaux

*The University of Mississippi*, [jvdeleve@olemiss.edu](mailto:jvdeleve@olemiss.edu)

Elizabeth Young Sweeney

*The University of Mississippi*, [eyoung@olemiss.edu](mailto:eyoung@olemiss.edu)

Follow this and additional works at: [https://egrove.olemiss.edu/population\\_brief](https://egrove.olemiss.edu/population_brief)

---

### Recommended Citation

Deleveaux, Jamiko and Sweeney, Elizabeth Young, "The Potential Impact of the Census Bureau's Differential Privacy System: Understanding Characteristics of the Mississippi Population" (2020). *Population Briefs*. 5.

[https://egrove.olemiss.edu/population\\_brief/5](https://egrove.olemiss.edu/population_brief/5)

This Article is brought to you for free and open access by the Center for Population Studies at eGrove. It has been accepted for inclusion in Population Briefs by an authorized administrator of eGrove. For more information, please contact [egrove@olemiss.edu](mailto:egrove@olemiss.edu).

# The Potential Impact of the Census Bureau's Differential Privacy System: Understanding Characteristics of the Mississippi Population

Jamiko Deleveaux and Elizabeth Sweeney  
*Population Brief* November 24, 2020



In 2018 the United States Census Bureau released a public notice in the Federal Register regarding a change in their differential privacy system. Following the release, the Census Bureau sought consultation with the data user community through presentations and workshops with the State Data Centers Network (SDC), the Federal-State Cooperative for Population Estimates (FSCPE), the Committee on National Statistics (CNSTAT), and the Population Association of America (PAA), among others. The change in the differential privacy system will have an impact on the way researchers, organizations, and the general public can analyze future Census Bureau publicly accessible data. To examine the potential effects of differential privacy on Mississippi population totals, this brief explains differential privacy and compares the Census Bureau's 2010 Decennial Census and 2010 Demonstration Data Product.

## ***What is Differential Privacy?***

The United States Census Bureau is in the process of retooling the way the organization protects privacy. Why is this change important to everyday users and stakeholders who rely on Census Bureau information to develop policy, improve infrastructure, and learn about their communities? First, we must realize that the Census Bureau under Title 13 is legally required to provide safeguards that protect individuals' ability to remain anonymous within the data. The Census Bureau has achieved this mission through a statistical disclosure system which aims "to protect the confidentiality of its respondents and their data" in all publicly released statistical datasets [1]. Second, the Census Bureau is an organization always aiming to improve between each Decennial Census; an example of this is the transition from the long form to the short form between 2000-2010.

Why has the Census Bureau deemed it necessary to make the change from the statistical disclosure system to the differential privacy system? The major reason lies with the Census Bureau legal mandate to protect the anonymity of individuals. For the past 50 years, the Census Bureau has relied on data suppression, swapping, and partial synthetic data for group quarters as a way to ensure that individuals remained anonymous within the data [2]. Research indicated that the statistical disclosure limitation systems used during the 1970-2010 Censuses were severely flawed in protecting individual identifiable information [3]. In response to the identified flaws and to modernize the system to meet new threats in protecting data such as database reconstruction, in 2020 the Census Bureau along with partners plan to implement a private disclosure system using a differential privacy method [4]. According to Census Bureau Associate Director John Abowd, the construction of the differential privacy system was designed to address two problems: provide maximum accuracy while accounting for privacy-loss budget, and equip policymakers with information to make the appropriate accuracy/privacy-loss choice [5].

The key measure in differential privacy revolves around the privacy-loss budget. The privacy-loss budget determines “how much of an individual’s privacy an analysis may utilize [6].” For example, data users conducting one analysis privacy-loss budget can be fully applied to one dataset. However, most data users typically conduct several analyses on a dataset, thus privacy-loss budget is calculated using total utilization [7]. Traditionally, researchers aim to have a smaller privacy-loss value to indicate better protection of individuals’ privacy.

A differential privacy system is built on the use of mathematical formulas. Incorporating the use of mathematical formulas within the differential privacy system introduces random variation or “noise” into the data [8]. The Census Bureau is in the process of finalizing its disclosure limitation algorithm, TopDown Algorithm (TDA), for implementation with the 2020 Census. TDA uses two steps (Imprecise Histogram and Optimize) to introduce random variation to account for accuracy and privacy-loss budget. In the first step of TDA, “variation from a carefully chosen distribution to the stratified counts of individuals” [9]. The by-product of the addition creates a set of counts with illogical inconsistencies creating an imprecise histogram [10]. The second step “finds optimized counts for each most-detailed cell in the histogram” [11]. To achieve optimization, the algorithm aims to make the count as close as possible to the count in the imprecise histogram while maintaining three assumptions: be non-negative, be consistent among cells and higher levels of hierarchy, and satisfy the invariants and inequalities. The two steps are applied to all six geographic levels (national, state, county, census tracts, block groups, and blocks). We can think of this as a ladder process where each lower level total helps to calculate the sum of the level above it. This process is achieved through the use of a privacy-loss budget (Epsilon) which determines how much variation to add during the initial step [12].

### ***Effects of Differential Privacy on Mississippi***

In this section we will compare the 2010 Census to the September 2020 release of the 2010 Demonstration Data (a public data release by the Census Bureau for users to see how 2010 Census raw data would look when differential privacy is applied). Mississippi cities and towns saw a net loss of nearly 555 residents in the 2010 Demonstration Data when compared to the 2010 Census (see Figure 1). Three cities experienced more than a 15 percent population gain. While this might appear to be positive for the cities of DeKalb, Sallis, and Silver Creek, four cities experienced more than a 20 percent loss, and 11 cities and places experienced a population loss of more than 10 percent. The varying differences in population loss and gain is very concerning and will have long term ramifications for local cities and municipalities. The concern is magnified when one considers that many cities and municipalities facing budget constraints will be impacted by population loss and gains, undoubtedly affecting the distribution of state and federal funding that rely on Census Bureau population figures for federally mandated population formulas.

The work of academics, non-profits, and professionals who use Census publicly released data to inform public policy decisionmakers and the public will become more challenging with the application of differential privacy. One area that has a potential for impact is educational funding between rural and urban school districts. In 2017-2018, a total of 827 schools in Mississippi qualified for Title I funding [13]. Qualifying for the Department of Education Title I funding relies on the use of basic demographic information and poverty characteristics of the community collected by the Census Bureau. These funds are then administered to local school districts to support students at risk of not meeting state learning requirements. The way school districts measure trends, plan for the future, and project potential federal funding may potentially be hampered by the introduction of differential privacy. Figure 2 indicates the potential for some school districts to qualify for increased funding due to population

gains, whereas some communities in need of government allocations will see a decrease in their funding due to a decline in population.

Mississippi has a high percentage of residents living in rural counties and cities. Comparing the 2010 Census to the 2010 Demonstration Data Product, major urban counties of Mississippi (DeSoto, Harrison, and Hinds) are similar. Figure 3 shows that rural communities are often the beneficiaries of population gains with the use of differential privacy. While most counties (45 out of 82 counties) saw modest or similar percentage change between the two datasets, Sharkey County (-0.12) and Quitman County (-0.11) were 2 of 37 counties to see the largest negative percentage change.

What does the change to differential privacy mean for Census Bureau data users in Mississippi? The 2020 Census has direct implications for Mississippi communities with regard to federal funding as well as information on population characteristics and distribution. Figures 4 and 5 compare the population age and sex distribution of Clarksdale, MS and Jackson, MS using the 2010 Decennial Census and 2010 Demonstration Data Product. In the case of Clarksdale (Figure 4), a rural community, the 2010 Demonstration Data shows an increased male and female population. Observing Jackson, MS (Figure 5), an urban community, both the 2010 Decennial Census and Demonstration Data Product are similar in terms of age and sex distribution.

The United States Census Bureau has made significant strides between their two iterations of the Demonstration Data Product (May 2020 and September 2020). However, one issue still exists for data users. Accuracy for small areas with smaller populations has improved but continues to be challenging for conducting analysis when compared to larger geographies. Additionally, in some cases, data users who have relied on certain types of Census data may no longer be able to access these data, because the Census Bureau will not be able to adequately protect the data or because the injection of noise makes it unusable. It is important for the Mississippi Census data user community to become aware of the advantages and disadvantages differential privacy might present when the Census Bureau releases 2020 Census data applying differential privacy in March 2021.

**Figure 1. Mississippi Towns and Places Differential Privacy Comparison between the 2010 Decennial Census and 2010 Demonstration Data Product Population Counts.**

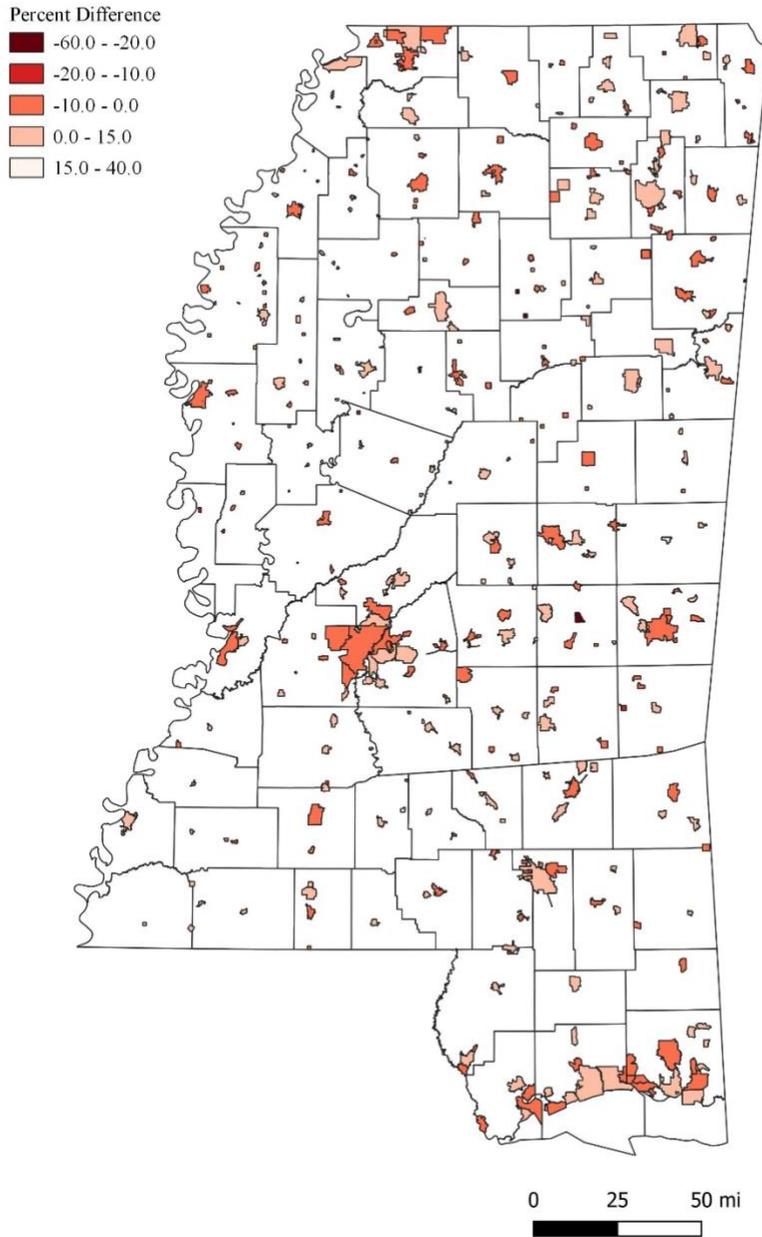


Figure Using Authors' Calculations Created by the State Data Center at the University of Mississippi.  
Source: United States Census Bureau 2010 Decennial Census and 2010 Demonstration Data Product.

**Figure 2. Mississippi Unified School Districts Differential Privacy Comparison between the 2010 Decennial Census and 2010 Demonstration Data Product Population Counts.**

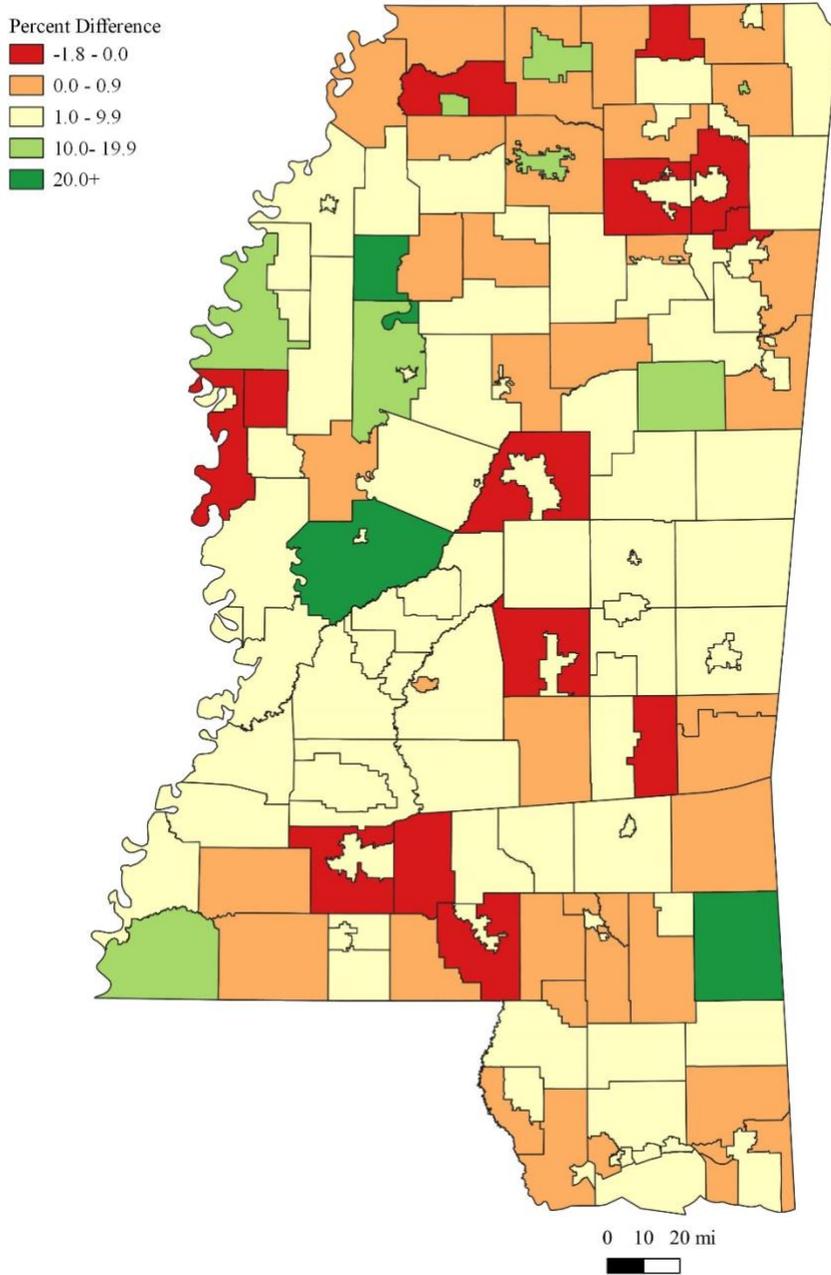


Figure Using Authors' Calculations Created by the State Data Center at the University of Mississippi.  
Source: United States Census Bureau 2010 Decennial Census and 2010 Demonstration Data Product.

**Figure 3. Mississippi County Differential Privacy Comparison between the 2010 Decennial Census and 2010 Demonstration Data Product Population Counts.**

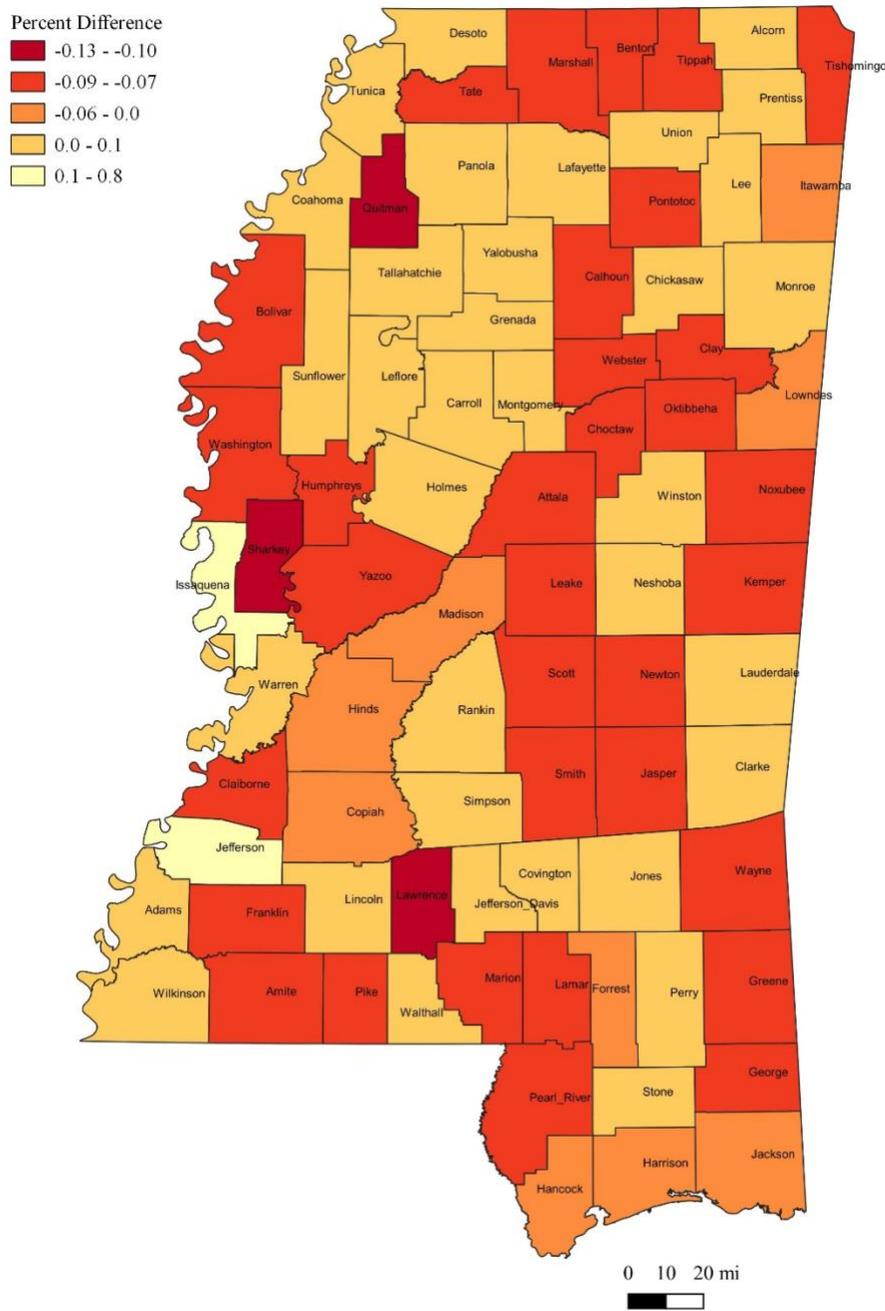
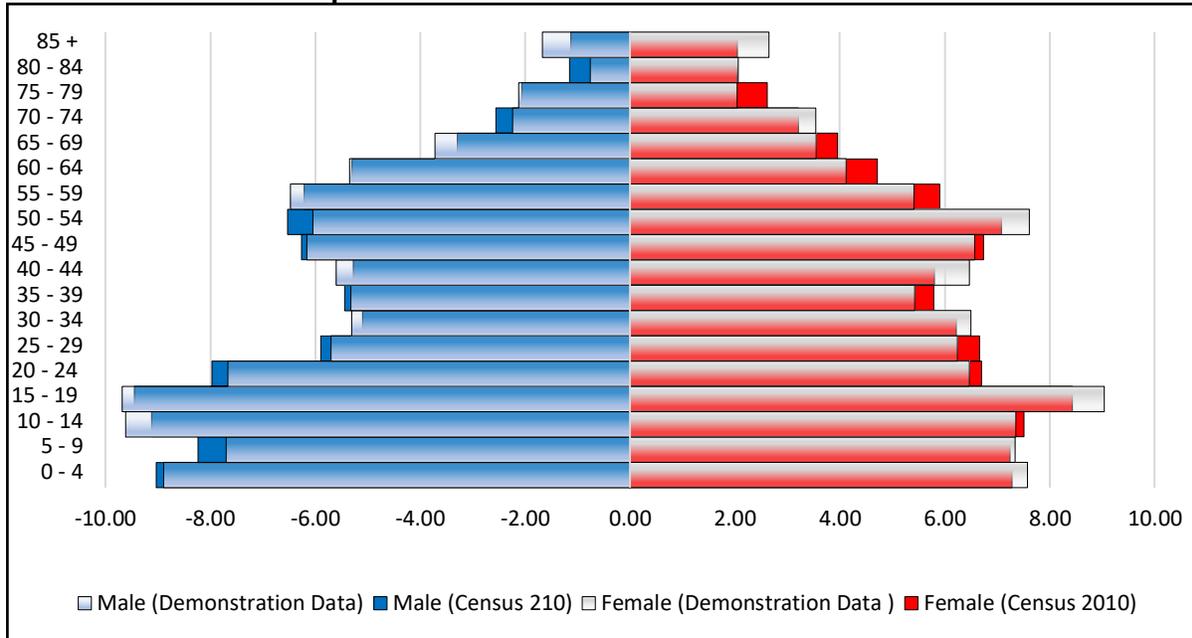


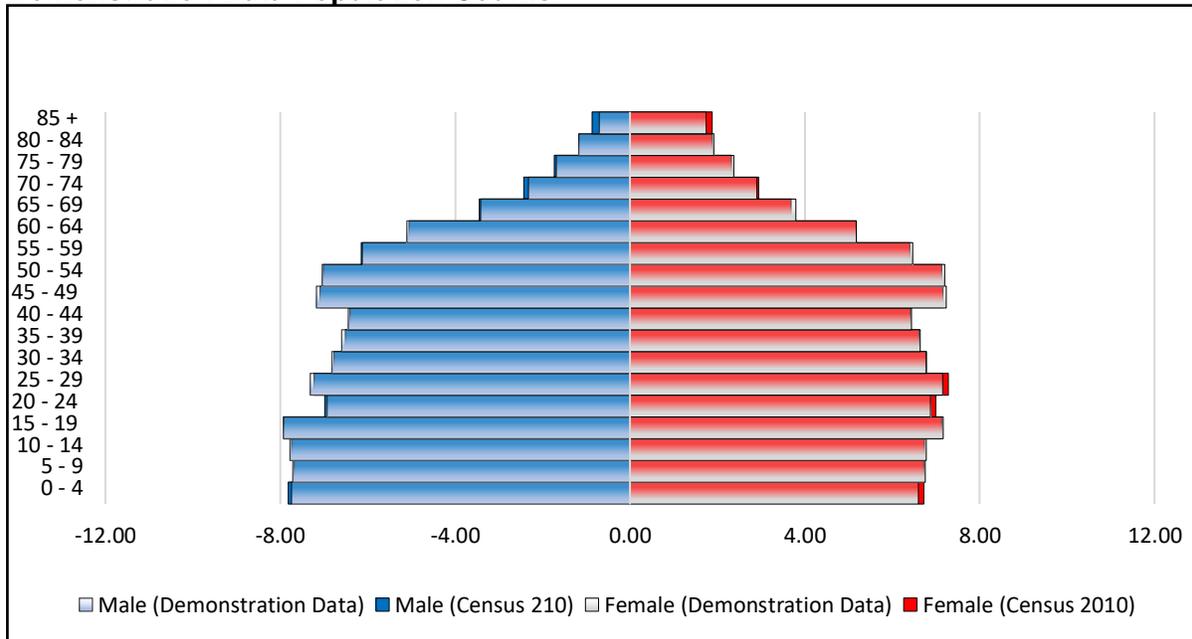
Figure Using Authors' Calculations Created by the State Data Center at the University of Mississippi.  
 Source: United States Census Bureau 2010 Decennial Census and 2010 Demonstration Data Product.

**Figure 4. Population Pyramid for Clarksdale, MS using 2010 Decennial Census and Demonstration Data Population Counts.**



Source: United States Census Bureau 2010 Decennial Census and 2010 Demonstration Data Product.

**Figure 5. Population Pyramid for Jackson, MS using 2010 Decennial Census and Demonstration Data Population Counts.**



Source: United States Census Bureau 2010 Decennial Census and 2010 Demonstration Data Product.

## References

- [1] McKenna, Laura. 2018. "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing." Washington, DC: United States Census Bureau Research and Methodology Directorate.
- [2] McKenna, Laura. 2018. "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing." Washington, DC: United States Census Bureau Research and Methodology Directorate.
- [3] Dinur, Irit and Kobbi Nissim. 2003. "Revealing Information while Preserving Privacy." Pp. 202-210 in *PODS '03: Proceedings of the Twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. San Diego, CA: Association for Computing Machinery.
- [4] Abowd, John. 2018. "The U.S. Census Bureau Adopts Differential Privacy." Paper presented at the Annual Meeting of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August, London, United Kingdom.
- [5] Abowd, John. 2018. "The U.S. Census Bureau Adopts Differential Privacy." Paper presented at the Annual Meeting of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August, London, United Kingdom.
- [6] Nissim, Kobbi, Thomas Steinke, Alexandra Wood, Mark Bun, Marco Gaboardi, David R. O'Brien, and Salil Vadhan. 2017. *Differential Privacy: A Primer for a Non-technical Audience*. Privacy Tools for Sharing Research Data project at Harvard University.
- [7] Elamurugaiyan, Aaruran. 2018. "A Brief Introduction to Differential Privacy." Retrieved August 13, 2020 (<https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b>).
- [8] Pettie, Samantha and Abraham Flexman. 2020. "Differential Privacy in the 2020 US Census: What Will it Do? Quantifying the Accuracy/Privacy Tradeoff." *Gates Open Research*, 3:1722.
- [9] Dwork, Cynthia and Aaron Roth. 2014. "The Algorithmic Foundations of Differential Privacy." *Foundations and Trends in Theoretical Computer Science* 9(3): 211-407.
- [10] U.S. Census Bureau. 2020. *2010 Demonstration Data Product*. Washington, DC: Author.
- [11] Pettie, Samantha and Abraham Flexman. 2020. "Differential Privacy in the 2020 US Census: What Will it Do? Quantifying the Accuracy/Privacy Tradeoff." *Gates Open Research*, 3:1722.
- [12] U.S. Census Bureau. 2020. *2010 Demonstration Data Product*. Washington, DC: Author.
- [13] Mississippi Department of Education. 2018. *Title I Schools 2017-2018*. Jackson, Mississippi: Author.