Greenbelt Reparations Commission Data Dashboard Project

Authors: Parker Leipzig, Brandon Fung, Isabella MacDonald, Ilyas Nur, and Larissa Musaga

Abstract

This project tasked us with analyzing demographic data for the Greenbelt Reparations Committee to gain deeper insights into the persistent racial inequalities experienced by African Americans and Native Americans in Greenbelt. Throughout the project we remained in contact with our main liaison for the project, Bob Rand, mainly via email: bob.rand@verizon.net. By analyzing the harm inflicted on these communities through comprehensive reports and data findings, our project aimed to assist the committee in addressing current issues effectively. We put a lot of care and thought into where we got our data and made sure it was the most accurate data we could get our hands on because of the sensitive nature of this project and how this could be used by the commission to inform their future projects. We looked at multiple publicly available datasets including but not limited to data from the US Census, GIS data, data from the Greenbelt healthcare industry, and data from the Maryland Statistical Analysis Center. The data in our dashboard is from the US Census data, specifically the American Community Survey Data (ACS) from 2021. This source was chosen because it is heavily peer-reviewed and while it does have a margin of error, at least we can show that in our dashboard and be transparent about that rather than use a data source that may have issues that are hidden. One thing we dealt with during the data collection process was needing to format it in a way that is easier for Tableau to digest and utilize for visualizations in our data dashboard. Some of our group members cleaned the data, organized it, and normalized it so that the end result of the data takes into account the different population sizes of the main demographics we were looking at (African-American, Native, and white). Then we experimented in Tableau to see the best way to visualize this for our client. We had them look at our first iteration of the dashboard/charts and then used their advice and feedback to create our final, edited version of the data dashboard that is live on Tableau public and can be edited once data is updated.

Methods

One of the main technical resources we needed to complete this project was quality data about the demographics of Greenbelt, which includes but is not limited to race/ethnicity, gender, and socio-economic status. We wanted to investigate if there are any possible disparities experienced within these demographics concerning education, disabilities, and household income statistics for residents of the city of Greenbelt. We investigated different sources such as health data from Prince George's County, the Maryland Open Data Portal, the U.S. Census, and data from the Maryland Statistical Center. While the Census data may not be fully complete, we found it to be one of the most holistic data sources to use. Our client recommended using the 10-year Census data along with the American Community Survey (ACS) data to get more granular information and to also evaluate county data, which could potentially help us in comparing Greenbelt's demographic data to the rest of Prince George's County. We knew that communicating with key individuals who are dedicated to helping those underserved in Greenbelt was incredibly important in learning what would be important to know for this project.

Our team was in constant communication with Bob Rand, our point-of-contact, to ensure we had all the resources and information we needed to help discover any inequities that are present in the city of Greenbelt.

Data Sources

The data was sourced from the American Community Survey (ACS), which is an annual survey conducted by the United States Census. The ACS collects data yearly and offers different estimates across years. We chose to use two different 5-year estimates, 2012-2016 and 2017-2021 so that we could compare the data between these years.

We specifically looked at four different ACS tables covering different topics regarding the population in Greenbelt: housing type, household income, education, and disability status. Each of these tables can be filtered for each race/ethnicity that is recorded by the ACS. Below are the names of each table, with links to each per race/ethnicity. We did not include data that was for two or more races.

2017-2021 ACS 5 Year Estimates

	African American	Native American/Alaskan Native	White	All Race/Ethnicity
Housing Type	B25003B	B25003C	B25003H	x
Household Income	B19001B	B19001C	B19001H	x
Education	C15002B	C15002C	C15002H	x
Disability Status	B18101B	B18101C	B18101H	x
Population	x	x	x	DP05

2012-2016 ACS 5 Year Estimates

	African American	Native American/Alaskan Native	White	All Race/Ethnicity
Housing Type	B25003B	B25003C	B25003H	x
Household Income	B19001B	B19001C	B19001H	x
Education	C15002B	C15002C	C15002H	x
Disability Status	B18101B	B18101C	B18101H	x
Population	x	x	x	DP05

Data Cleaning

The ACS data was not in a usable format upon downloading each table as a CSV file. We used Google Sheets to clean the data. We eliminated whitespace in each cell. We also added columns for the groupings. For example, for disability data we added a disability column and an age column. Below is what the data looked like before we made any of our changes.

Label (Grouping)	Greenbelt city, Maryl	Greenbelt city, Mary
Total:	17	±27
Under 18 years:	0	±22
With a disability	0	±22
No disability	0	±22
18 to 64 years:	17	±27
With a disability	0	±22
No disability	17	±27
65 years and over:	0	±22
With a disability	0	±22
No disability	0	±22

In Tableau, we decided to merge all the data according to its category. The four categories we decided to analyze were education, household income, tenure, and disability. This data covered Native American, African American, and Caucasian communities in Greenbelt. After merging, we had four spreadsheets for each category, where each spreadsheet covered all races in the year 2021. We also normalized the data so that all estimates were proportional to
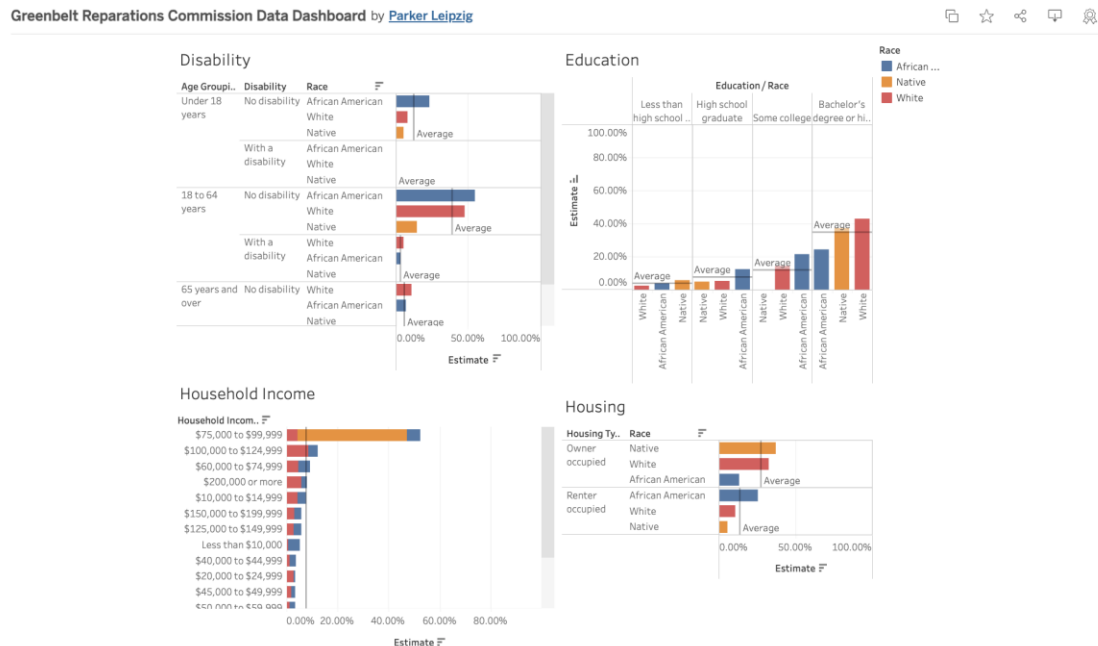
each race's population in Greenbelt to avoid any bias. Put simply, we took each estimate and divided it by its corresponding race population in Greenbelt in 2021. After this was complete, we were ready to analyze the data and create the dashboard in Tableau. Here is an example of what the disability spreadsheet looks like after cleaning:

disability	age_grouping	race	estimate	estimate(num)	margin_of_error	margin_of_error(num)	year
With a disability	Under 18 years	African American	0.004691689	56	0.004607909	55	2021
No disability	Under 18 years	African American	0.232573727	2776	0.054540885	651	2021
With a disability	18 to 64 years	African American	0.034014745	406	0.013153485	157	2021
No disability	18 to 64 years	African American	0.548592493	6548	0.053703083	641	2021
With a disability	65 years and over	African American	0.044989946	537	0.036947051	441	2021
No disability	65 years and over	African American	0.069872654	834	0.021866622	261	2021
With a disability	Under 18 years	Native	0	0	0.096899225	25	2021
No disability	Under 18 years	Native	0.054263566	14	0.093023256	24	2021
With a disability	18 to 64 years	Native	0	0	0.096899225	25	2021
No disability	18 to 64 years	Native	0.143410853	37	0.244186047	63	2021
With a disability	65 years and over	Native	0.372093023	96	0.445736434	115	2021
No disability	65 years and over	Native	0	0	0.096899225	25	2021
With a disability	Under 18 years	White	0	0	0.003495526	25	2021
No disability	Under 18 years	White	0.08291387	593	0.031180089	223	2021
With a disability	18 to 64 years	White	0.055508949	397	0.025307606	181	2021
No disability	18 to 64 years	White	0.476370246	3407	0.071588367	512	2021
With a disability	65 years and over	White	0.046420582	332	0.01761745	126	2021
No disability	65 years and over	White	0.110738255	792	0.025307606	181	2021

Tableau Dashboard

We wanted to create a dashboard that was both meaningful and easy-to-understand. The dashboard should show the disparities between all the different races in Greenbelt. In all the graphs in the dashboard, we assigned a color to each race to easily show their estimates. The axis measuring the estimate is represented as a percentage from 0 to 100 of that race's total population. Since there were concerns regarding comparing ACS data from two different years, we decided to use only data from 2021 and represent these disparities as stacked/unstacked bar charts. The ACS data was also useful because it is a five-year estimate it does encompass the years 2016-2021. After consulting with our client about the first iteration of our charts on the data, we found that the commission would prefer simpler charts that they can easily read. We also wanted to make sure our charts could be understandable by anyone, not just our group and the commission who may have been looking at this kind of data for a long time. We ended up choosing to do solely bar/column charts for our data dashboard to meet these requirements. This type of visualization made it very simple to compare the different races' estimates for each category. We posted our data dashboard on Tableau public so that anyone can see the dashboard and our clients can access the dashboard as long as they may wish. The dashboard allows users to scroll over the different bars of data and receive about six or seven different kinds of more

specific data points rather than just the view of the bars from a first glance. In each of the four charts, the pop-up shows you the race, the actual number of the estimate and the percentage of the estimate, and the margin of error number and percentage. Here is a screenshot of a glance at the data dashboard we constructed.

Data Analysis

The visualizations we've crafted, accessible here, offer valuable insights into the demographic landscape of Greenbelt. These visualizations employ normalized and scaled data, focusing on Education, Disability, Housing conditions, and Household Income within our targeted Greenbelt community. Our aim is to aid the Greenbelt Reparation Commission in gaining insights to address historical racial and ethnic inequities in the Greenbelt community.

Our educational visualization, based on 2021 data, reveals that the Native American and White populations in Greenbelt surpass the average in attaining Bachelor's degrees. Conversely, the African American population falls below the average in this aspect. This educational gap may impact their access to well-paying jobs, subsequently affecting household income and home ownership.

The disability visualization, based on the five-year estimate of the ACS data, showed that the Native American population had the highest number of people with disabilities of the Native American, African-American, and White populations for the 65 and older group. For the other two age groups, under 18 years of age and between 18 and 64 years of age, it was a little surprising to learn from the data that the African-American population had the highest number of the population in Greenbelt without disabilities for the 18 to 64 age group, with the White population not falling far behind.

Regarding housing conditions, as anticipated, the Native American and White populations in Greenbelt exceed the average in homeownership rates and exhibit lower rates of renting. Conversely, the African American population shows lower rates of homeownership and relatively higher rates of renting.

Finally, for the household income data, we learned that $75,000 to $99,999 was the most common household income for all of Greenbelt, but more specifically for the Native American population. The White population had the highest average household income of any group, falling into the $100,000 to $124,999 category.

For all of this data analysis, it is essential to acknowledge that the figures for Native Americans have a notably high margin of error due to the relatively small population available for comparison. Nevertheless, we've taken additional steps to normalize and scale the data to the best of our ability.

Conclusions

As previously mentioned in our data analysis, we discovered some differences between different races (White, Native American, and African American) for different socioeconomic factors that may lead to inequities in Greenbelt. We hope that this dashboard created by our team of UMD students will be a valuable tool for the commission to investigate racial demographic data in the city of Greenbelt. In this holistic report, we detailed our methods, how we cleaned our data, and the research we conducted before creating this Tableau dashboard. To further assist the commission after delivering our project, we will also be attaching a Tableau user guide on how to read our dashboard and how to navigate it as this may be some users' first time interacting with this software. We are thankful and honored for the opportunity to assist the Greenbelt Reparations Commission in its mission of combating racial inequities. Finally, we would like to extend our gratitude to Bob Rand, our point-of-contact, for joining us in numerous meetings and answering all of our questions to help us better understand our project requirements and scope.

Future Steps/Recommendations

There is much more that can be examined when it comes to examining the harm of racism in Greenbelt. One future project could involve creating an interactive map that can be used to examine some of the same topics as above in a geographic manner. The ACS data we examined does not include any geographic data. However, the Census Bureau does offer this data with geographic components. So a future project could involve geospatial analysis as the entire focus, as there is a lot that can be examined in that aspect.

Additionally, the ACS does offer more granular datasets. One such table, for example, aggregates poverty status by disability status and by employment status– table B23024. Future projects with more time could look into these more specific datasets.

Another idea for a future project would involve examining the intersectionalities of this data in more depth. For example, we looked at disability data by age group, so in this case future work could include looking at disability by sex as well. This could apply to all of the types of data we looked at– housing, education, etc.

After our presentation, we found that our client was impressed by our final result and we believe our dashboard will be very useful to them. We hope that they can use this to help inform whatever reparations decisions they make in the future, but we do think more can certainly be done and added to this dashboard to help flesh out the intricacies of small ways inequities have impacted racial communities in Greenbelt.