# ***Final Project Write Up***

**Midterm Proposal Outline Update**

**The Original Proposal:**

**Name:** Ainsley McLaughlin

**Project Title:** A Look Homeschooling in the United States, 2016 vs 2019

**I. Introduction / Background**

I chose to contrast a subset of the data collected **during 2016 and 2019** by the **Parent and Family Involvement in Education (PFI) survey which is collected by the government. I have chosen to use the subset of this data concerning homeschooling in the United States. In a paper by Harvard Law Professor Elizabeth Bartholet published in 2020, she voiced concerns about homeschooling in the US, and engaged academic, political, and public interest in issues surrounding homeschooling. Due to her analysis, she proposed a near ban on homeschooling in the US and received heavy criticism or support from various, different groups. Due to the fact that very little research is done on homeschooling, Bartholet relied heavily on the national survey mentioned above which is the primary source of information collected on homeschooling in the US.**

**To my knowledge, no research or data has been collected on homeschooling where the study participants have been previously homeschooled adults. Most, if not all, information collected is from the perspective of parents of homeschooled children.**

**Homeschooling in the US has always been legal. However, each state regulates this practice differently, and homeschooling can look different for each participating family and child. This makes it very difficult to collect data concerning the practice.**

**II. Audience**

The audience for this project could be any person interested in learning more about homeschooling in the US. Some examples could be previously or currently homeschooled individuals, homeschooling parents, persons interested in becoming homeschooling parents, educators, lawmakers, among many others.

**III. Dataset(s)**

*source URL: https://nces.ed.gov/nhes/dataproducts.asp#2019dp*

In this data, there are 16447 rows and 827 columns since this is a national survey. However, I will only be using a select number of those columns which concern homeschooling.

The data is not easily useable, and I will need to subset the data and add meaningful headers although the excel files have abbreviation headers. I will need to do a lot of work to subset and compile the data I need from that larger data set.

**IV. Proposed Visualizations**

I would like to look at where homeschoolers are primarily located in the USA, using a map. I will allow the user to only select certain regions, and include the different years, 2019 vs 2016.

A map with blue dots

Description automatically generated

I will also include a Bar chart of the motivations or reasons for homeschooling and allow the user to choose the regions on the sidebar. I will have one of each for 2016 vs 2019 and put them in a workbook.

A screenshot of a graph

Description automatically generated

I will also include a pie chart of the type of curriculum used by the homeschooling family and a pie chart of the mode of instruction used by homeschooling families, both for 2016 and 2019. I will allow the user to filter by region for all four of these charts.

A colorful pie chart with white text

Description automatically generated

*Source:* [How to filter on the pie chart (tableau.com)](https://community.tableau.com/s/question/0D54T00000C5zAySAJ/how-to-filter-on-the-pie-chart)

**V. Plan**

I plan to use Tableau to create the visualizations and put them in a story. However, I will likely use R and Excel to work with the raw data that will be used in Tableau. I will publish the visualizations through Tableau itself so they will be public. I will also put this in a portfolio as I apply to grad school in data science and data analytics.

*Plan A:*

My ideal plan would be to collect some of my own data from adults who were homeschooled and compare it with the year 2019 data on the same metrics and using the same visualizations as above.

Plan B:

My fallback plan is to use the 2016 vs the 2019 data using the visualizations and metrics listed above.

*Plan C:*

My second fallback plan is to use this data set: [Gun Violence Data (kaggle.com)](https://www.kaggle.com/datasets/jameslko/gun-violence-data/data), and explore gun violence in the US using similar visualizations as described above.

Helpful link for myself: [Homeschooling in the United States: Results from the 2012 and 2016 Parent and Family Involvement Survey Web Tables](https://nces.ed.gov/pubs2020/2020001.pdf)

https://nces.ed.gov/nhes/data/2019/ecpp/cbook\_ecpp\_pu.pdf

**Instructor Feedback on the Midterm Proposal:**

You've chosen an interesting topic with a useful dataset!

The three initial visualizations you've described will give you a good starting point. I'm sure once you begin to explore the data, you'll have ideas for additional visualizations and/or ways to supplement your existing visualizations (e.g. filtering for different homeschooling reasons on the map).

Since viewers typically have difficulty judging the area of pie chart regions, you'll want to either add labels for the quantities (missing in the example chart, but provided in some examples completed in class) or you may consider using different visualization options.

I strongly suggest that any backup plan takes advantage of work completed towards the more ambitious plan(s). So, I would discard plan C. Your fallback plan B definitely sounds doable.

I look forward to seeing your work on this project!

**Peer Feedback on the Midterm Proposal:**

By Zac Simpson:

The specific data that is going to be shown should be specified.

You should subset and compile the data.

You might consider renaming all of the columns.

You might have a hard time getting the info so maybe try to use Plan B.

**Changes Made and the Reasons:**

The first and most obvious change is that I decided to use Plan B since I did not have adequate time to collect my own data responses as well as prepare it for use.

Secondly, I was only able to use fields that were common to the 2016 and 2019 files which did not include data that would be used for a map. This also did not allow for me to allow the user to filter by region for any of the data. I used primarily bar charts since these are familiar to users. I also included a few pie charts and one other chart.

Below I outline the visualizations I created and the interactivity of them.

*Additional Visualizations I decided to create based on the data that I have:*

* Homeschooled vs not for 2016 vs 2019: A bar chart on a logarithmic scale. There is no interactivity.
* Race: 2016 vs 2019: A bar chart one may filter on homeschooled vs not.
* Total Income of the household for 2016 and 2019: A highlight table one may filter on homeschooled vs not.
* Grade level of surveyed child 2016 vs 2019: A bar graph one may filter on homeschooled vs not.
* Parent highest level of education 2016 vs 2019: A pie chart one may filter on homeschooled vs not.
* Parent’s expectation for child’s future educational attainment 2016 vs 2019: A horizontal bar chart one may filter on homeschooled vs not.
* Parent’s primary motivations for homeschooling 2016 vs 2019: A bar chart one may filter on total household income.
* Homeschool teaching style (following formal curriculum or not) 2016 vs 2019; A bar chart one may filter on primary reason for homeschooling.
* Zip code classification by community type for those who answered homeschooled 2016 vs 2019: A bar chart one may filter on primary reason for homeschooling.
* How much homeschooling 2016 vs 2019; A pie chart one may filter on primary reason for homeschooling
* Primary provider of the homeschool education 2016 vs 2019: A bar chart one may filter on total household income.

**Codes used:** HOMESCHLX, RACEETH2, TTLHHINC, ALLGRADEX, PARGRADEX, SEFUTUREX, ZIPLOCL, HMSCHARR, HSMOSTX, HSSTYL, HSWHOX

**Summary / Visualization Access**

I chose to contrast a subset of the data collected **during 2016 and 2019** by the **Parent and Family Involvement in Education (PFI) survey which is collected by the National Center for Education Statistics. I have chosen to use the subset of this data concerning homeschooling in the United States. In a paper by Harvard Law Professor Elizabeth Bartholet published in 2020, she voiced concerns about homeschooling in the US, and engaged academic, political, and public interest in issues surrounding homeschooling. Due to her analysis, she proposed a near ban on homeschooling in the US and received heavy criticism or support from various, different groups. Due to the fact that very little research is done on homeschooling, Bartholet relied heavily on the national survey mentioned above which is the primary source of information collected on homeschooling in the US and is collected from the perspective of the parents. Additionally, each state regulates this practice differently, and homeschooling can look different for each participating family and child.**

**The goal of the story is to first allow users to explore the differences in homeschooled children versus those who are not homeschooled. The story allows the user to see that homeschoolers make up a small percentage of children. Additionally, the participants who were homeschooled were slightly more racially diverse than the general population. In 2016 especially, those who were homeschooled seemed concentrated in the total household income range of $0-10,000 or $75,000-$100,000. In 2019, they seemed more concentrated in the $75,000-$150,000 range. In the general population, the concentration of people got higher as the income range got larger. Furthermore, the grade level distribution of the general population slightly increased as the grade level increased. When examining homeschoolers alone, this was not the case. In 2016, the most common grade levels were Kindergarten, nineth and twelfth grade. In 2019, one can observe that the grade distribution is fairly even with the most common grade levels being first, fourth, eighth, and twelfth. When examining educational attainment, it appeared that parents of homeschooled children expected their children to obtain a bachelor’s or graduate degree primarily.**

**When examining homeschoolers alone, it appears that the greatest reasons families decide to homeschool their children is for academic instruction, school environment, and unspecified other reasons. Thus, in the future, those creating the survey should attempt to explore other reasons families decide to homeschool their children to get a more accurate representation of reasons. Especially as this data was filtered by low-income level, those three reasons were highlighted. In opposition to the perception that homeschoolers are educated in an unstructured way, exploring the teaching style of homeschooling parents highlighted the fact that homeschoolers are primarily educated using a mostly formal or strictly formal curriculum. This was especially true when the data was filtered by homeschooling parents whose reason for homeschooling their child was academic instruction or school environment which were found to be two of the top reasons. When examining the community types of homeschoolers, it was found that in 2016 they primarily lived in large cities or large suburbs and by 2019 the most common areas were large suburbs or fringe rural areas, followed closely by large cities. It seemed to be a trend that homeschooling was becoming more popular in rural areas as time went on. Interestingly, of those who were homeschooled, in 2016 about seventy-five percent of children were only partially homeschooled and participated in other academic institutions such as public or private schools. By 2019, only about nine percent answered they were partially homeschooled and most instead claimed to be completely homeschooled if homeschooled at all. These percentages were even more drastic when filtered by the top reasons for homeschooling: academic instruction and school environment. Finally, it was found that the most common facilitators of homeschool education are mothers, although the use of virtual schools or curriculums grew from 2016 to 2019.**

The Tableau story may be accessed through this link [A Look at Homeschooling in the U.S...2016 vs 2019 | Tableau Public](https://public.tableau.com/app/profile/ainsley.mclaughlin/viz/ALookatHomeschoolingintheU_S___2016vs2019/Proj?publish=yes). The first tab is an introduction that includes a link to the data source, and the seventh tab transitions the user to part two which only concerns data on surveyed children who were homeschooled for at least some classes. On every other tab, each visualization has the option to hover over the graphic to highlight different elements. Additionally, clicking on any element in the visualizations for 2016 will highlight the corresponding element in the visualization for 2019. Finally, almost every tab in the story has a filter on the right panel which users can select various options to explore the data.

**Challenges Encountered and Addressed**

The first challenge I faced was the size of the data set. Since I did not need most of the data from the survey, I subset the data using R. Then, I found that the column titles did not always match between the 2016 and 2019 data. Thus, in Excel and with help, I was able to find the column headings I was interested in which were common to the 2016 and 2019 data. Next, I found that Tableau did not recognize the categorical data I was using since those categories were input as numbers. For example, being homeschooled was input as 1, not being homeschooled was input as 2 and valid skip was input as -1. Thus, for every column, I changed the inputs to string categorical variables. Additionally, since some of the inputs were analogous but different even in the columns which were common between the 2016 and 2019 data, I had to standardize the inputs for both to make them common between the sets. For example, in 2016, the inputs for race were very simplistic and in 2019 they were more diverse. Additionally, for some of the columns just the numeric inputs were different, but the options were the same between the 2016 and 2019 data sets. For example, under grade level, in the 2016 data the input number for first grade was 1 and in the 2019 data it was 4.

**Design Decisions**

I chose primarily bar charts since users are very familiar with this type of visualization and it is easy to compare different variables in these charts. I also used some pie charts with labels and a highlight table. I chose pie charts as these are also familiar to users and it is easy to imagine percentages with them. I chose a highlight table for the data on household income since this data has so many categories and the color scheme made it easy to see which categories were more common than others. Since all the data I had was categorical, my options for visualizations were fairly limited. I tried using bubble charts but perceiving the sizes of the bubbles was difficult. I also tried using a tree diagram, but the labels did not fit well, and it did not seem to highlight the important points well. I also tried to use a stacked bar chart, but the chart became one bar and thus was not interesting or interpretable. Finally, I really wanted to use a side-by-side bar chart for data from 2016 versus 2019 but Tableau does not allow for one chart to use data from different data sets. Thus, this was not a possibility.

I chose various color schemes for different visualizations. For the grade level charts I grouped different age groups by color scheme to help distinguish patterns. Additionally, I chose different color schemes for different types of zip code classifications so that users could easily see the four main zip code classifications which were broken into three smaller classifications each. In tab 1.5 which concerned educational attainment, since the inputs were not exactly the same between the parents’ educational attainment and their expectation for their children, I made two of the inputs in the latter the same color in order to make the inputs analogous in color. Furthermore, I tried to choose contrasting colors for adjacent bars and sections of visualizations in order to easily distinguish them. Finally, I made the colors for the inputs the same between the 2016 and 2019 visualizations to make comparisons easy.

For interactivity, I chose to allow the user to filter the visualizations by various other categories in order to highlight different points. For example, when exploring if the child was completely or partially homeschooled, I allowed the user to filter by the reason they chose to homeschool the child. Thus, one might be able to see that if the family chose to homeschool the child due to concerns about academic instruction or school environment, the chart would highlight that these families more often completely homeschooled the child. In the other visualizations, I chose the filters based on similar logic given my experience and predictions on what the data should show. However, some of the results were surprising and contradicted what I expected. For additional interaction, I wanted the user to be able to highlight and element in one of the charts from either year, and it would highlight the corresponding element in the chart from the other year. However, it was only possible to make selecting an element in one of charts highlight the corresponding element in the other. Thus, I chose the visualizations from 2016 as the chart which when clicked in, would highlight the same element in the 2019 visualization.

**Discussion of Future Work**

The goal of this project is to give a visual context for the discussion surrounding the data on homeschoolers in the U.S. which was brought into the limelight by **Elizabeth Bartholet in 2020. However, there is very little data on homeschooling in the U.S. and virtually none from adults who were homeschooled. Most of the data comes from parents who homeschool or homeschooled their children. Thus, next semester, I plan to extend this project with a personal research project on adults who were previously homeschooled. There is especially a lack of data from formerly homeschooled adults on their perception or experience with homeschooling. I would like to collect this data and contrast it with data from parents of homeschooled children.**

**Additionally, homeschooling grew and changed tremendously during the Covid-19 pandemic. Thus, I would like to contrast the data collected during 2016 and 2019 with more recent data in order to see trends in this growth and these predicted changes. Thus, I would like to include some line charts that measured changes over different years in order to see trends.**

**One of the biggest concerns that Professor Bartholet had in her work was that of the unsupervised aspect of homeschooling which allows it to be used by abusers. She used the example of Tera Westover, whose autobiography “Educated” was the First Year Reading Experience during my first year at the University of South Carolina. This project could be extended to include data and graphics which represent the correlation or lack thereof between homeschooling and cases of abuse or neglect.**

**Another very valid concern Professor Bartholet had concerning homeschooling in the U.S. was the lack of national regulation or standards. Each state creates its own regulations or lack thereof and thus, there are varying amounts of oversight and accountability. I would like to use data from every state on how they regulate homeschooling and put these methods in a visualization which could be filtered by state. This filter could also apply to cases of abuse or neglect of homeschooled children in the state to see if there is any correlation between the two.**