Wellesley College

How the Quality of Abortion Clinics Suggested by Google Changes when Searching
from High vs Low Income US Counties

Sofia Kobayashi
CS 234 Data, Analytics, and Visualization
Professor Eni Mustafaraj
21 December 2022

Abstract

With the fall of Roe v Wade and the subsequent banning of abortion in many American states, the subject of abortion access is more contentious and important than ever. With Google being one of the first tools that people turn to when trying to find abortion care[1], it's important to know if there are differences in the clinics it suggests when searched for by different groups of people and what those differences are. This work looks at abortion clinics that Google Places suggests from a search for 'abortion near me' in 83% of US counties in which abortion is still legal, and attempts to determine what the differences in star rating, number of reviews, and distance to closest one are and why those differences might exist.

Motivation

Since access to abortion is currently under very real threat in America and because of how important Google is to finding abortion care, it's important to know how your location and community may impact the abortion clinics that Google Places suggest when searching for abortion near you. This paper aims to provide important context to both people seeking an abortion and in what ways our current system affects different people's access to abortion care.

Research Questions

The main question that guided this paper was 'How does a person's location and community affect the Google Places abortion clinics that google suggests?'. I hypothesized that when searching from lower income counties, the clinics that Google Places suggested would be worse, defined in this paper as having lower star ratings and being further away.

Methods

There were two main sources of data for this project: US Census Population data and the data scraped from Google when searching for abortion care. The first data set was created around all the counties in America: their geographical location, population by gender, percent of the population that identify as one race and white, median household income, and political alignment and I mostly used the pandas library to clean and format the data. Once the variables had been identified, the internet was searched for appropriate data sets, which were found on the US Census Bureau's website or on Wikipedia who originally got it from the Census Bureau. These data sets were all for 2021.

---

[1] https://www.washingtonpost.com/technology/2022/08/25/google-maps-abortions/

The data sets were then read into Jupyter Notebook and, with the use of the pandas Python library, the relevant variables were picked out of them, then each data set was cleaned and standardized so that they could then be merged together into one big data set. After this merge, all counties in the states in which abortion was illegal were dropped, since I assumed there would not be enough Google Places data from those states. Then each county was assigned a political alignment, either republican or democrat, based on how the state had voted in the 2020 presidential election. Once this was completed, there were 1880 counties with the variable shown below (Figure 1A).

```
Data columns (total 12 columns):
 #    Column                          Non-Null Count
---   ------                          --------------
 0    state                           1880 non-null
 1    county                          1880 non-null
 2    total_population                1880 non-null
 3    total_female                    1880 non-null
 4    total_male                      1880 non-null
 5    total_one_race_white            1880 non-null
 6    percent_female                  1880 non-null
 7    percent_white                   1880 non-null
 8    latitude                        1880 non-null
 9    longitude                       1880 non-null
 10   num_households                  1880 non-null
 11   dollar_household_median_income  1880 non-null
```

Figure 1A. Display of all variables by US county from the US Census Bureau, 2021

The Google Places search results were captured using the Python library Selenium. Our script changed the location from which the search came for each county in our dataframe, searched on Google for 'abortion clinics near me', then saved that search page result to an HTML file. This process took about 2 hours with six computers running simultaneously, each search took about 20 seconds. Once all 1880 of the HTML files were saved, a separate script was run that found the Google Places section on the search page and passed to the adding-script the average star rating and average number of reviews of all abortion clinics in that section, the number of clinics Google Places suggested, and the distance in miles from the county to the closest clinic. The adding-script then added this

information to each county in the dataframe (seen below, Figure 1B), then dropped any rows that in the star ratings, number of reviews, or distance to nearest clinic could not be found so that t-test could be carried out. Once this was complete 1556 counties remained (about 83% of the original 1880 in our population).

```
 #    Column                          Non-Null Count
---   ------                          --------------
 0    state                           1880 non-null
 1    county                          1880 non-null
 2    total_population                1880 non-null
 3    total_female                    1880 non-null
 4    total_male                      1880 non-null
 5    total_one_race_white            1880 non-null
 6    percent_female                  1880 non-null
 7    percent_white                   1880 non-null
 8    latitude                        1880 non-null
 9    longitude                       1880 non-null
10    num_households                  1880 non-null
11    dollar_household_median_income  1880 non-null
12    numGooglePlaces                 1804 non-null
13    avgStarRating                   1556 non-null
14    closestClinic_mi                1556 non-null
15    avgNumReviews                   1556 non-null
16    poli_align                      1880 non-null
```

Figure 1B. Display of the Google Places search-related variables (seen in the red box) that were added to the county dataframe

Methods and Results

Three two-sample t-tests were conducted to determine whether or not a county's income level, median household income in this case, would affect the star rating, number of reviews, or distance to nearest clinic that Google Places would suggest. The two samples in this case were high vs low income counties, determined by whether or not a county was below the median household income average. In all three t-tests, it was found that county income level did significantly affect all three, even with all the t-test's alphas being 0.001, with the the distance to closest clinic being the most obviously impacted.

The difference between star ratings and number of reviews in high vs low income counties was that they both were higher on average in low income counties. This was very surprising as I had hypothesized the opposite because I had assumed clinics in high income counties would have more resources and thus be rated higher and with more people leaving good reviews, but seeing these results, I can think of two reasons this may be. The more social and speculative reason: people from lower income counties are likely to have a harder time getting an abortion and thus be less negative if they receive the care they need. While the more stats-based reason is that because lower income counties had higher numbers of reviews, which had a correlation of 0.59 to average star rating, thus, the higher star rating may be because of higher numbers of reviews rather than a difference in service or resources.

However, the results of the distance to closest clinic suggested by Google Places was the opposite, in that on average someone from a low income county would have to drive almost 50% longer to get to the closest clinic—90.21 miles compared to 61.61 miles for someone in a high income county. These two contradictory results make it difficult to determine if a person's location affects the quality of abortion clinics suggested by Google Places, but it certainly shows that it is more difficult to access abortion care if a person lives in a lower income county.

Visualizations

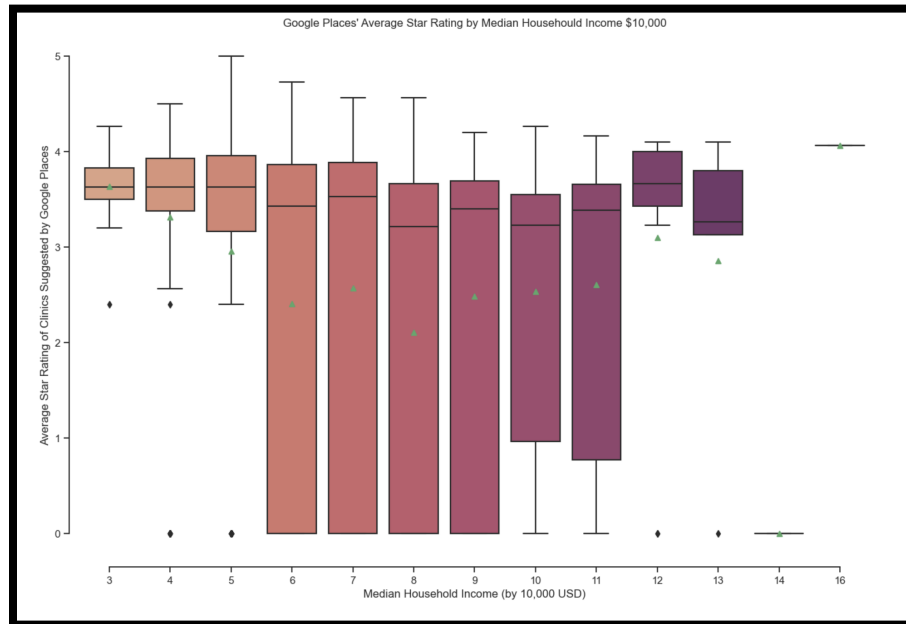Our group made [a website showing each member's project and our group work](#).

Figure 2A. Boxplot showing showing how the average star rating of the Google Places abortion clinic suggestions change when searching from US counties with different levels of median household income (by the 10,000 USD)
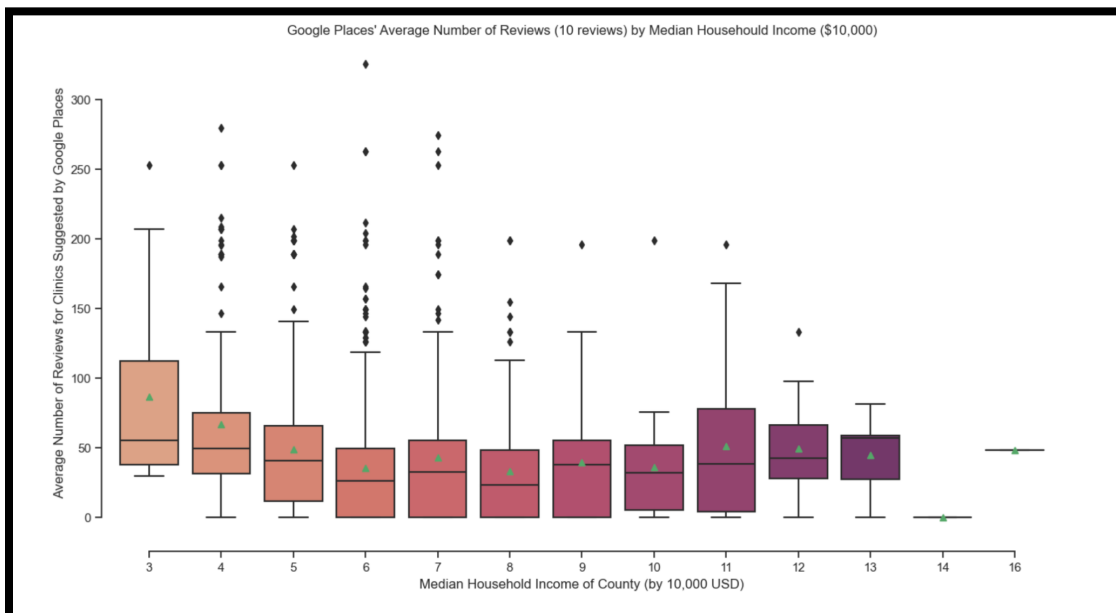

Figure 2B. Boxplot showing showing how the average number of reviews of the Google Places abortion clinic suggestions change when searching from US counties with different levels of median household income (by the 10,000 USD)
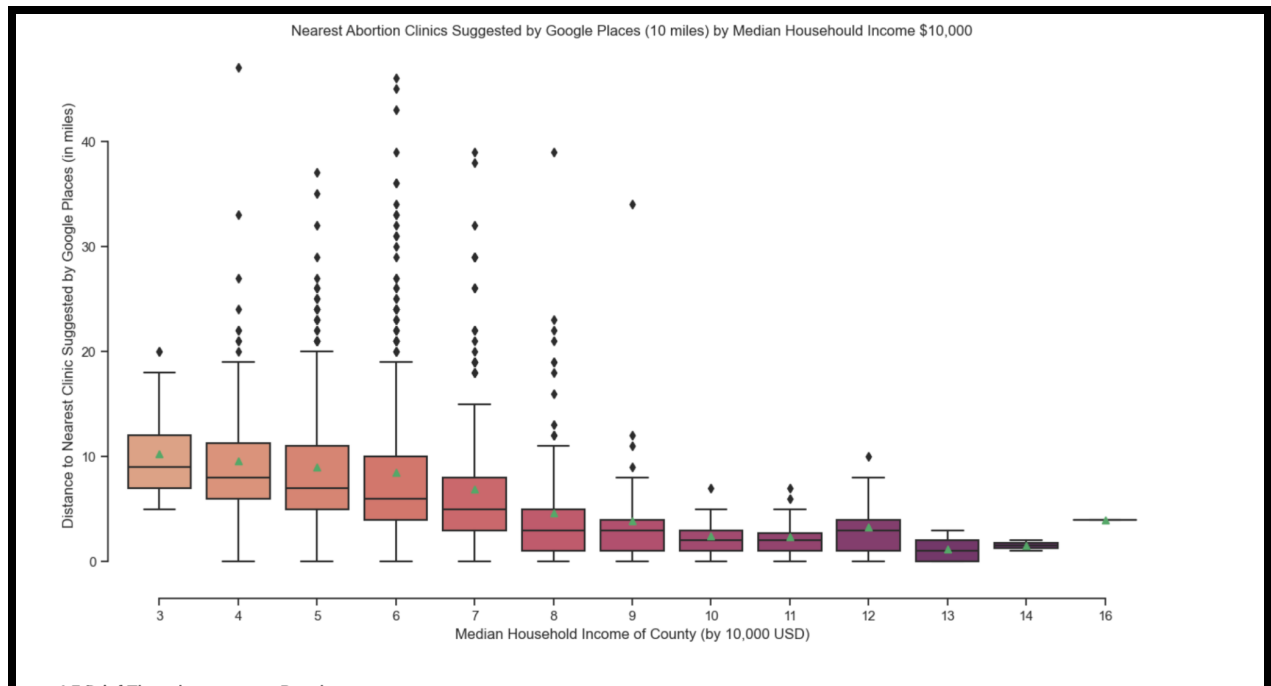
Figure 2C. Boxplot showing showing how the distance (in miles) of the closest clinic from the Google Places abortion clinic suggestions change when searching from US counties with different levels of median household income (by the 10,000 USD)

Conclusions and Future Work

I was very surprised by the star ratings and number of reviews being higher in low income counties than high ones and the magnitude of the difference in distance to the closest suggested clinic between high and low income counties. I hypothesize there would be a significant change when searching between the two samples, but I didn't think it would be by 50%.

For future work, I would like to continue this analysis by the race and gender breakdown of each county's population, as well as by other population and political metrics such as size of family, male vs female provider for family, or political alignment of region officials. Mapping all the abortion clinics in the US, their location, ratings, and reviews would also add a critical component to this project that I wish I could have explored further—if the way Google was suggesting abortion clinics was the 'best' it could be and whether or not it was just a vessel for societal disparaties in abortion access or if Google was actively affecting abortion access one way or another. For example, if Google was always suggesting the closest abortion clinic to a person's location, and if not, why? It would also be an

interesting project to look into how a state banning or heavily restricting abortion affects a person's usage of tech-based abortion resources, such as Google, sexual and reproductive health websites, and period trackers.