Racial Demographics' Impact on Online Reviews in Philadelphia

James M. Moore

Data Analytics, Dickinson College

Introduction

This research was conducted on the effect that the racial makeup of a neighborhood has on the online reviews of businesses in that particular municipality. The focus of this particular project is on Philadelphia, one of America's most segregated cities, and how the Yelp reviews vary from neighborhood to neighborhood within the city. We have collected data from every business that has received at least five Yelp reviews and has been open within the last two years.

The primary goal of this study is to determine if racial makeup, in particular white/non-whiteness has a measurable effect on the reviews given to businesses in that area. It is important that sites that claim to be unbiased authorities be able to withstand scrutiny and that preexisting biases do not affect the communities they serve.

Methods

Data was collected through Yelp in March-April 2023, on as many businesses as possible in Philadelphia and surrounding incorporated neighborhoods. Neighborhood data was collected from the OpenPhillyData project, as well as the city's official website. The demographic data was obtained via census records.

To determine meaningful data, the information was run first through GIS software to determine overt graphical trends within the data and have a visual representation that could serve as an easy comparison, as well as give a baseline of what a reader should expect of the data moving forward

Secondly the data was parsed through a variety of regression methods and models to determine both the significance of any given trend, as well as provide the best possible explanations for any given variations or patterns found within the research itself.

Maps

Maps of Philadelphia: Left, a map of the average star reviews of each neighborhood. Right, a map of the percentage white each neighborhood is, with a red dot for each business
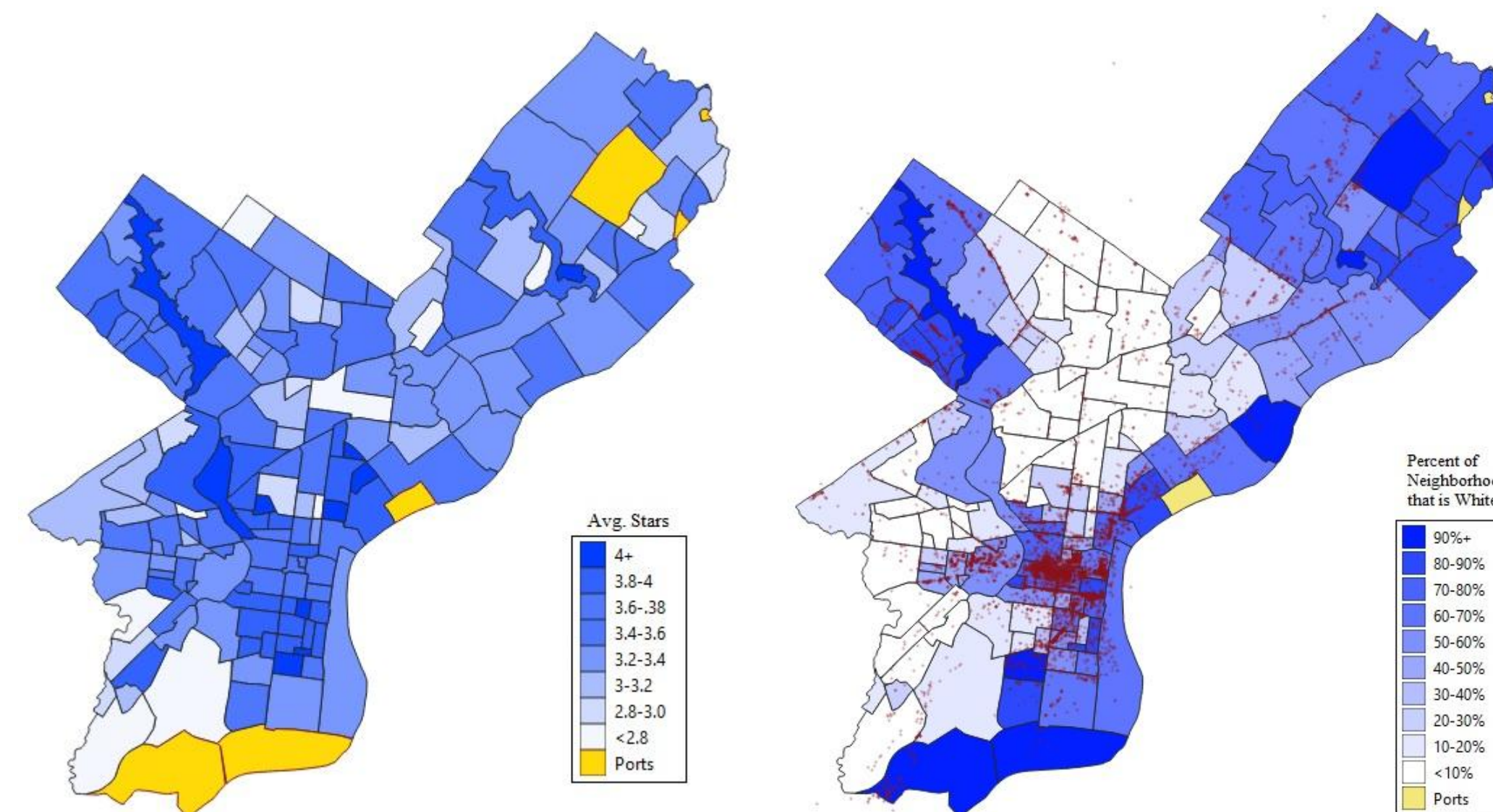
The maps were made with QGIS

Upon initial look there are some initial trends visible. The primarily white areas in the northeast adn northwest appear to be much bluer, with higher reviews. In fact the long jagged neighborhood in the Northeast, Wissahickon Park, is 100% white and one of highest average ranked neighborhoods in the city

Graphs
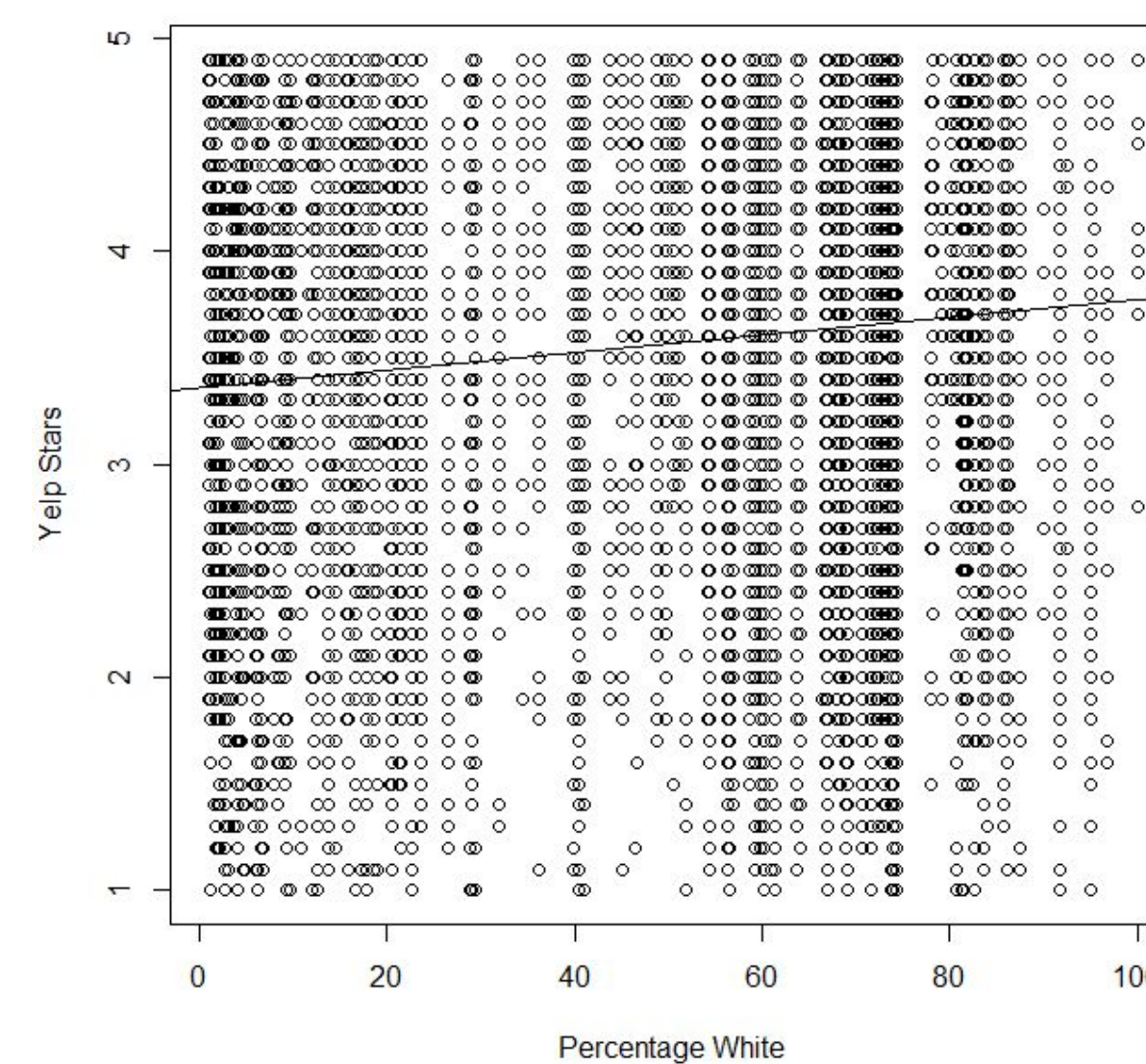
The graph on the left shows all of the business star rating mapped against the percentage of white residents in their neighborhood as well as a line of best fit.

What is most interesting about this graph at first glance is just how apparent the segregation in the city is. There are incredibly visible bands of data, particularly around the ~2% mark and the ~70% mark, further demonstrating the divide within the city itself.

Regression Analysis

While there are various models were fitted to this data, the most striking was just the basic linear model as it tells the starkest tale. There is a direct linear correlation between the whiteness of a neighborhood and the

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.360069   0.021393  157.07   <2e-16 ***
data$X.White 0.004159  0.000344   12.09   <2e-16 ***
```

star ratings that businesses in that neighborhood receive. While the slope estimate is rather small, but it does show an almost half star difference in average reviews between the whitest neighborhoods and those that are not. Furthermore with such a small p-value, this implies a strong correlation between the two values which would suggest an immensely problematic trend going forward.

Discussion

The relationship found in this research points to a very concerning trend that businesses in whiter neighborhoods are receiving higher average reviews. This further perpetuates a cycle of inequality as sites like Yelp are often the first resource people looking for services will use to find those desired services. Without some sort of compensation for this bias, a vicious cycle will occur and businesses built in less white neighborhoods will suffer from lack of demand.

Going forward the goal would be filtering out even more noise within the data to pinpoint the exact trends, as well as creating a weighted compensation algorithm to ensure that each business is assessed on a fair and equitable scale.

References

1. "Data Summary." Data Summary - Pennsylvania Spatial Data Access, https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=7047.
2. "Race and Ethnicity in Philadelphia, Pennsylvania (City)." The Demographic Statistical Atlas of the United States - Statistical Atlas, https://statisticalatlas.com/place/Pennsylvania/Philadelphia/Race-and-Ethnicity.
3. "Search Businesses in Philadelphia, PA." Yelp, https://www.yelp.com/search?find_loc=Philadelphia%2C%2BPA.

mooreji@dickinson.edu