Analysis of OpenStreetsPGH Data for BikePGH

Delivered: March 24, 2019

Lead:
Sagnik Rana
sagnikr@andrew.cmu.edu

Team:
Arnaud Avondet, Jordan Bender, Michell Li, Preeya Mody, Maitri Vasa
{aavondet, benderj, mli5, mpreeya, mvasa} @andrew.cmu.edu

1. Introduction

Main Objective: Help BikePGH understand who attends OpenStreetsPGH based on race/ethnicity, economic bracket, and neighborhoods to help them better understand and service City of Pittsburgh residents with their programs.

Our given dataset consists of approximately 190 survey responses gathered from online surveys completed after three OpenStreetsPGH events in the summer of 2018. The survey consisted of mixture of 23 questions: single-select, select-all-that-apply, and open-ended.

According to BikePGH's website, 45,000 people attended the Downtown/Strip District/Lawrenceville event alone. The number of survey responses (N = 190) over the duration of three separate events is limited compared to the total attendees. Therefore, we advise caution when reading this report due to those limitations and suggest the findings as guideposts for future opportunity explorations.

2. Analysis

2.1 Understanding the Attendees: Demographics

In order to better understand the survey respondents profile, we did some basic exploratory analysis. Most of this could be done using Microsoft Excel or basic R, but we provided graphs below for quick reference:
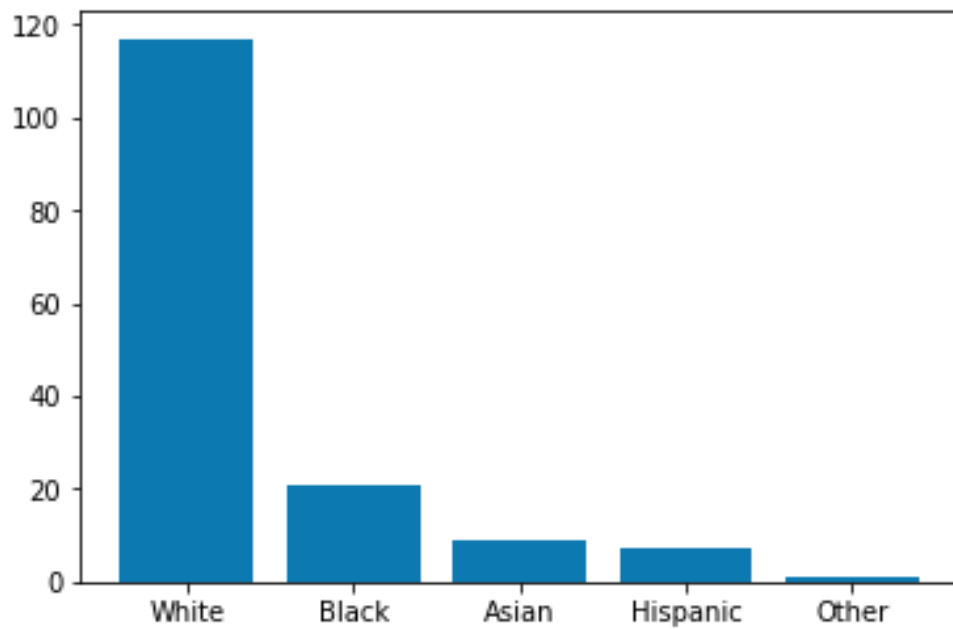
Figure 1: Race/Ethnicity of Attendees (Count) *Races of the survey respondents correspond to city demographics.*
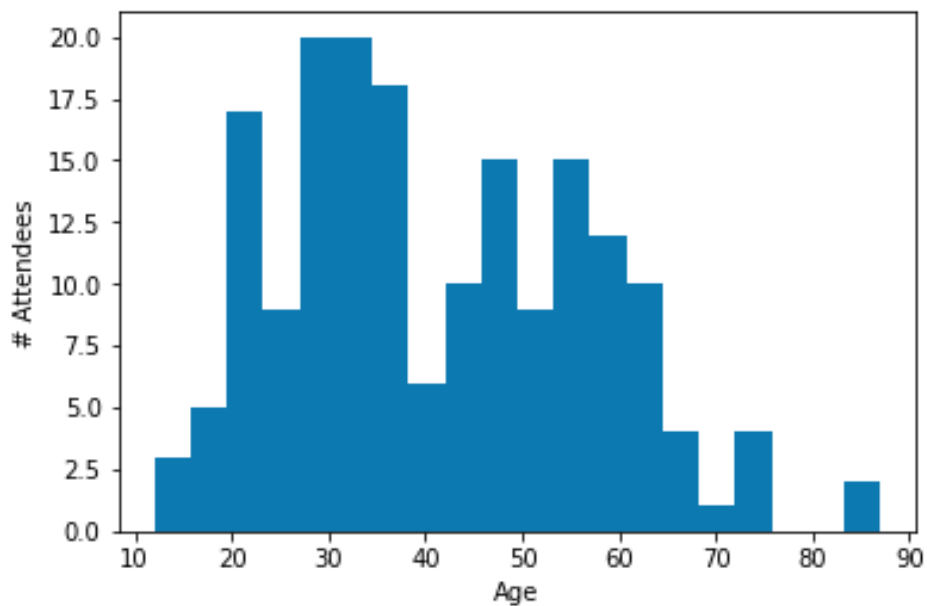

Figure 2: Age Distribution of Attendees (Count) *Notice there are two distinct sets: Under 40 and Over 40.*

Figure 3: Income
Distribution of
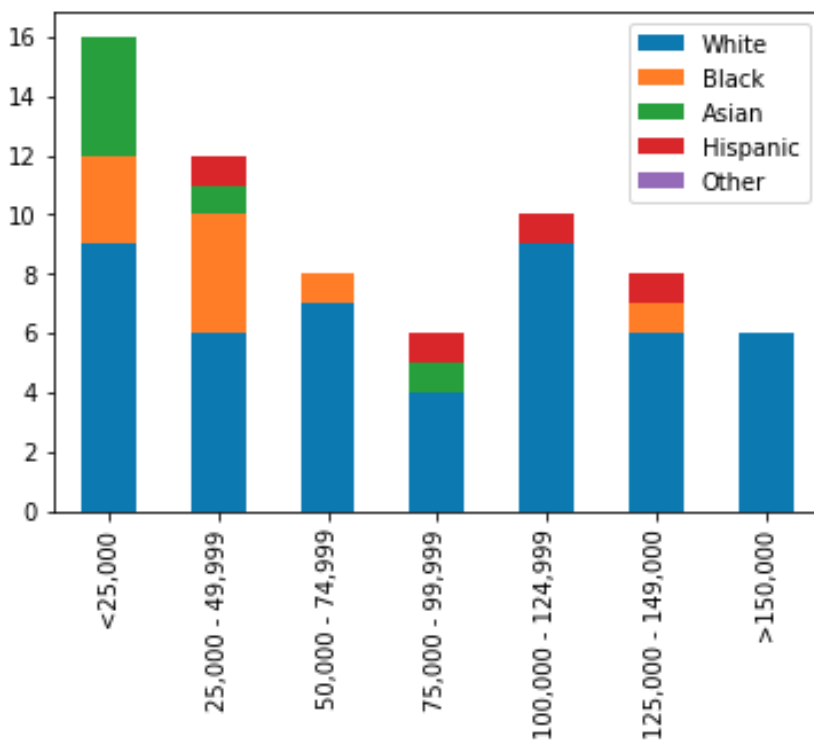Attendees (Count)
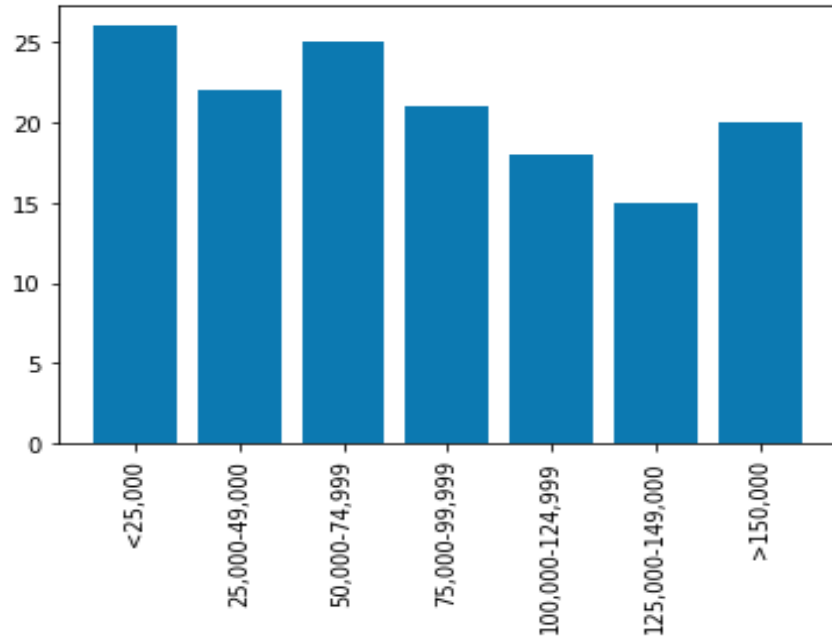*Notice that it is
relatively uniform.*

Figure 4: Income by Race
for Attendees under 40
years of age. (Count) *By
looking at the two wider
age groups and their
respective incomes, we
can further segment
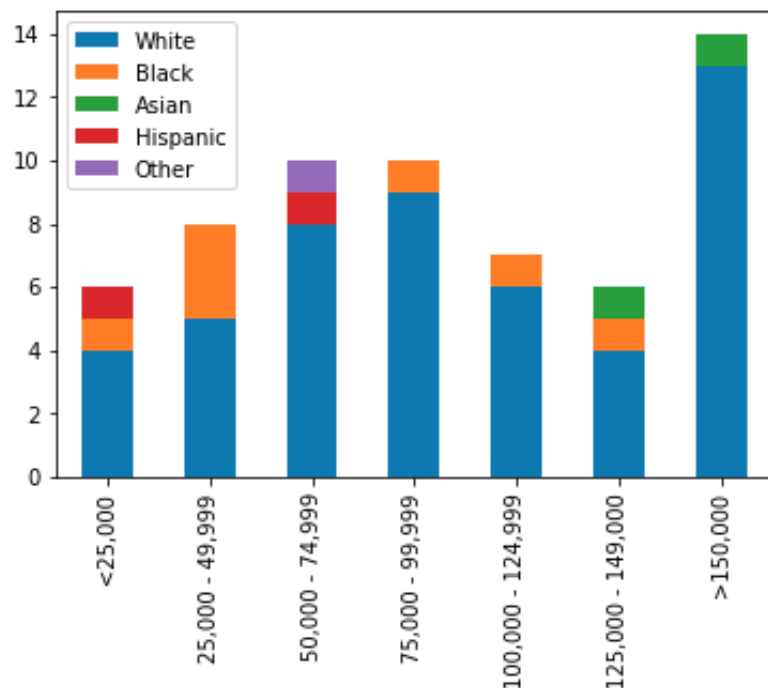participants and their
possible interests.*

Figure 5: Income by
Race for Attendees over
40 years of age.

2.2 Understanding the Attendees: Interests & Motivations

We focused much of our analysis on the activities that brought people to the event. In further sections we break interest down by race and by zip codes.
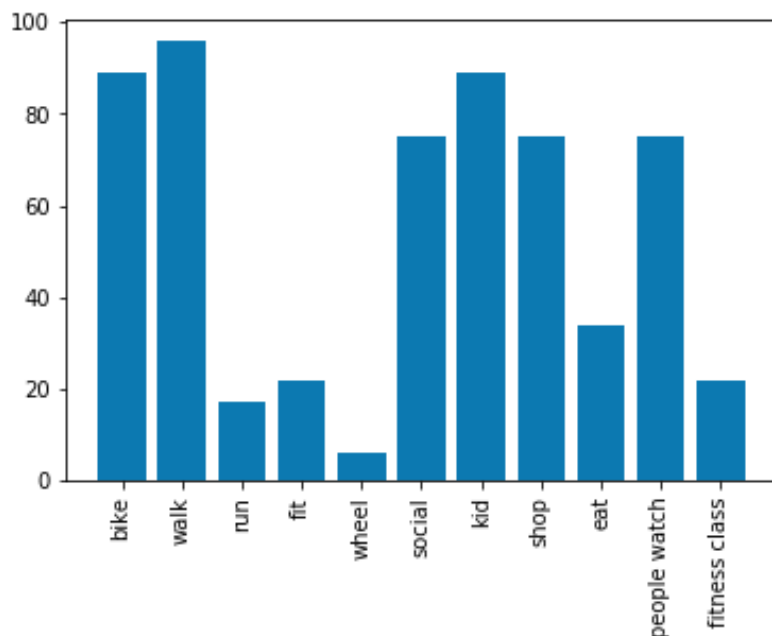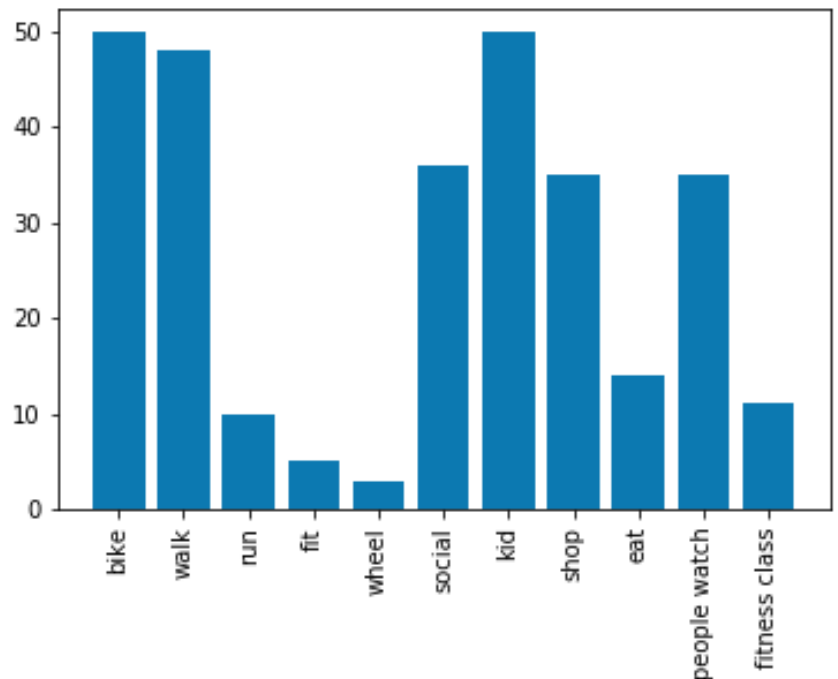

Figure 7: What activities do they like the most? (Count) *It appears that the social interaction is a huge component for many attendees.*

Figure 8: What are the reasons that people who've attended OpenStreetsPGH more than once like about the event? (Count) *Notice that attendees interests remain consistent as compared to Figure. 7.*

- Most of the participants were joining Open Streets event for the first time
- People who come to the event for second time, mostly continues to come at least 4-5 times.

Figure 9: How often do participants attend OpenStreetsPGH? *Based on the surveys, if an attendee goes to a second event, they are likely to continue going. Therefore, capturing interest during their first event is important.*

2.3 Activities Based on Race

One of our main interests was to determine how different racial groups interact with OpenStreetsPGH. BikePGH notes 'Equity, Diversity, and Inclusion' as a core value and we believe leveraging data analyses may determine focus areas to promote wider participation.
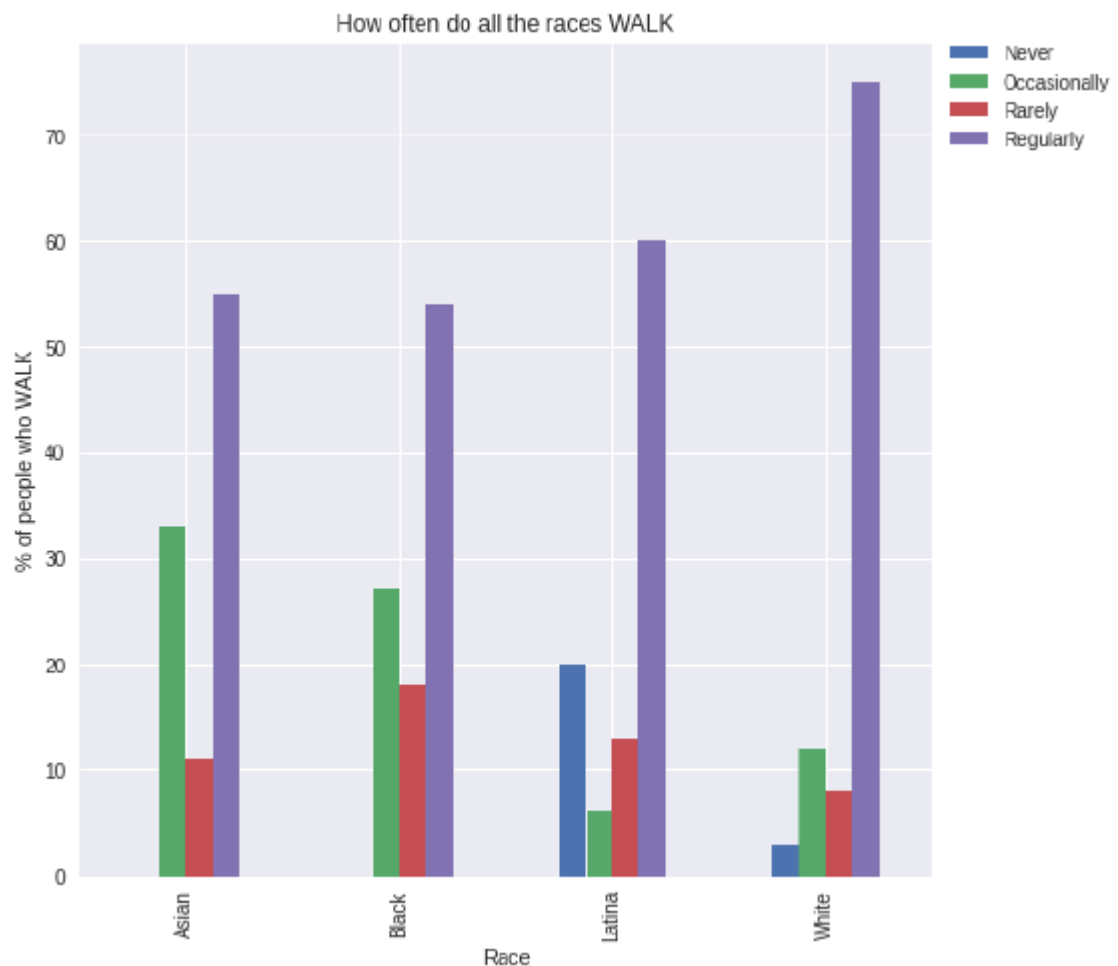

Figure 10: How often do people walk? Percentage based on race. *While the surveys are self-selecting (i.e. these people chose to come to an event focused on exercise and interaction), it is reassuring there are not many "Never" walkers. We see that Asian and Black groups have a higher percentage of "Occasional" walkers and may be a good starting point to encourage more walking.*
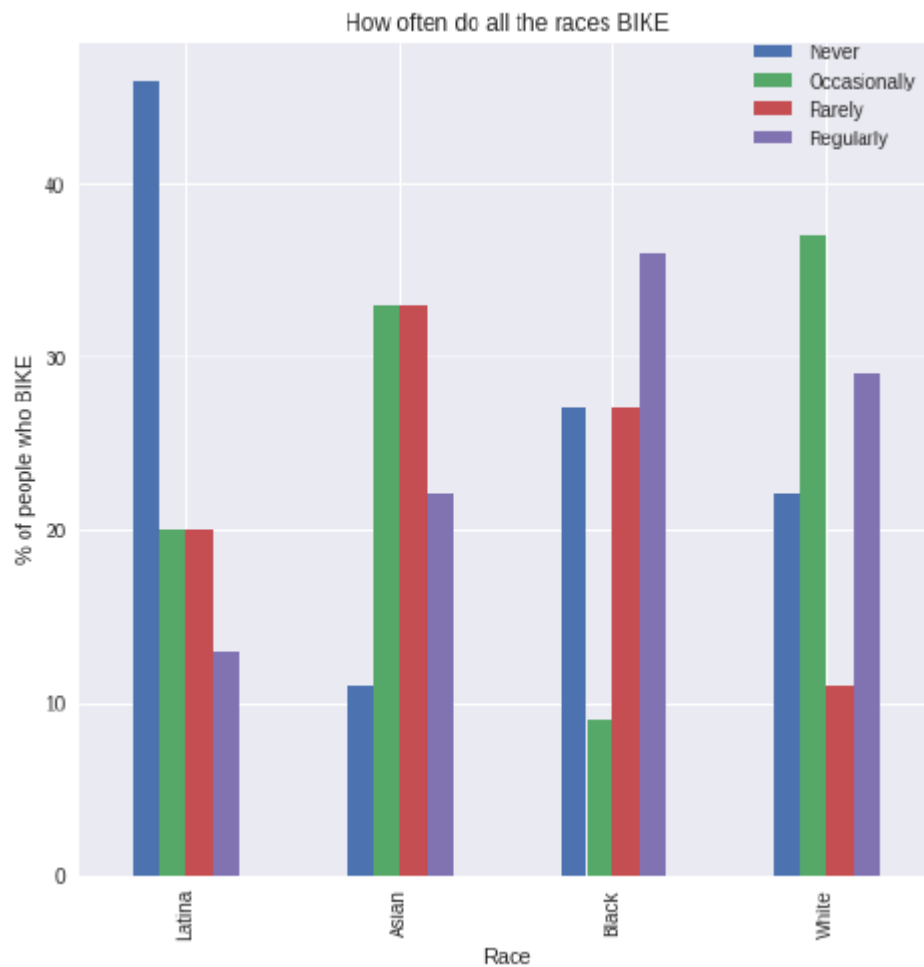
Figure 11: How often do people bike? Percentage based on race. *The prime opportunities to increase bike usage based on this data is among White and Asian groups, as these groups have the highest "Occasional" biking rates. However, the high percentage of 'Never' and 'Rarely' biking among Latinx and Black groups may have hidden factors. (Also note we had a limited sample size, especially among Latinx survey respondents.) These groups may be biking less due to factors discovered in our other analyses: lower average income and safe biking infrastructure in their neighborhoods (Zip Codes). Therefore, these groups may have huge potential to increase biking participation, but the underlying factors will be harder to address.*
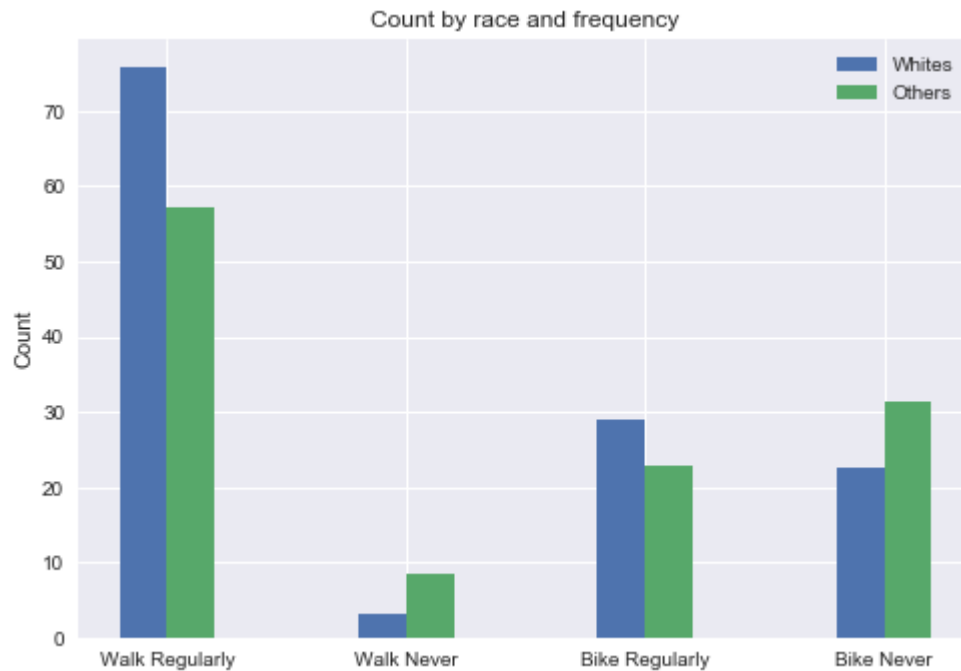
Figure 12: How often do people walk? Percentage based on Whites and Combination of all other races. (Count) *Because our data is so heavily skewed towards White groups, we broke down walking/biking frequency by White or Others for easier comparison. However, even taking into account this distinction, we still see that White groups walk and bike more regularly than Other groups. This suggests that more data is required to determine lower reasons that Other groups do not participate as readily and to increase outreach to those peoples.*
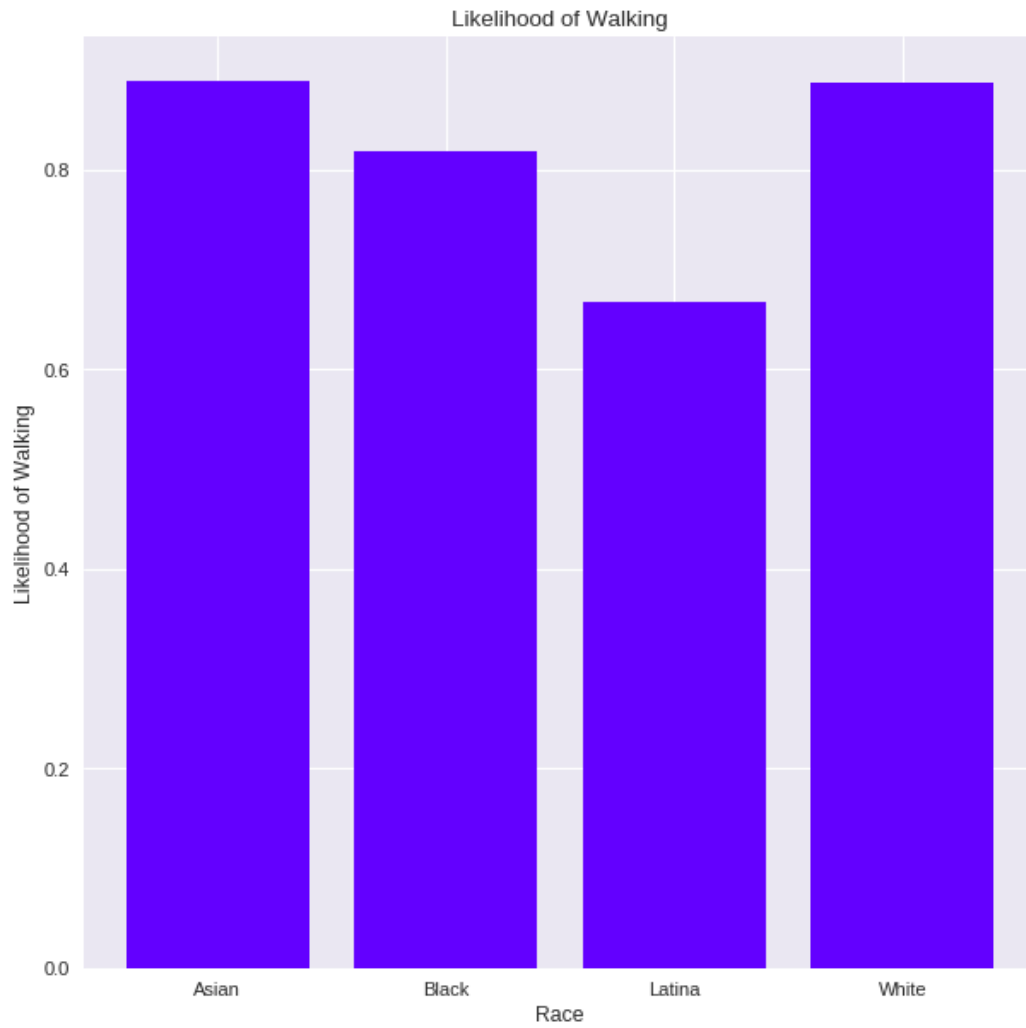
Figure 13: Likelihood of Respondents Walking by Race. *Comparing the frequency of walking in Figure 10 to this graph shows that even groups that only "occasionally" walk (e.g. Asian and Black groups), nearly all groups have a high level of likelihood of walking. The lowest group is Latinx, but this could be biased due to low number of respondents of that race.*
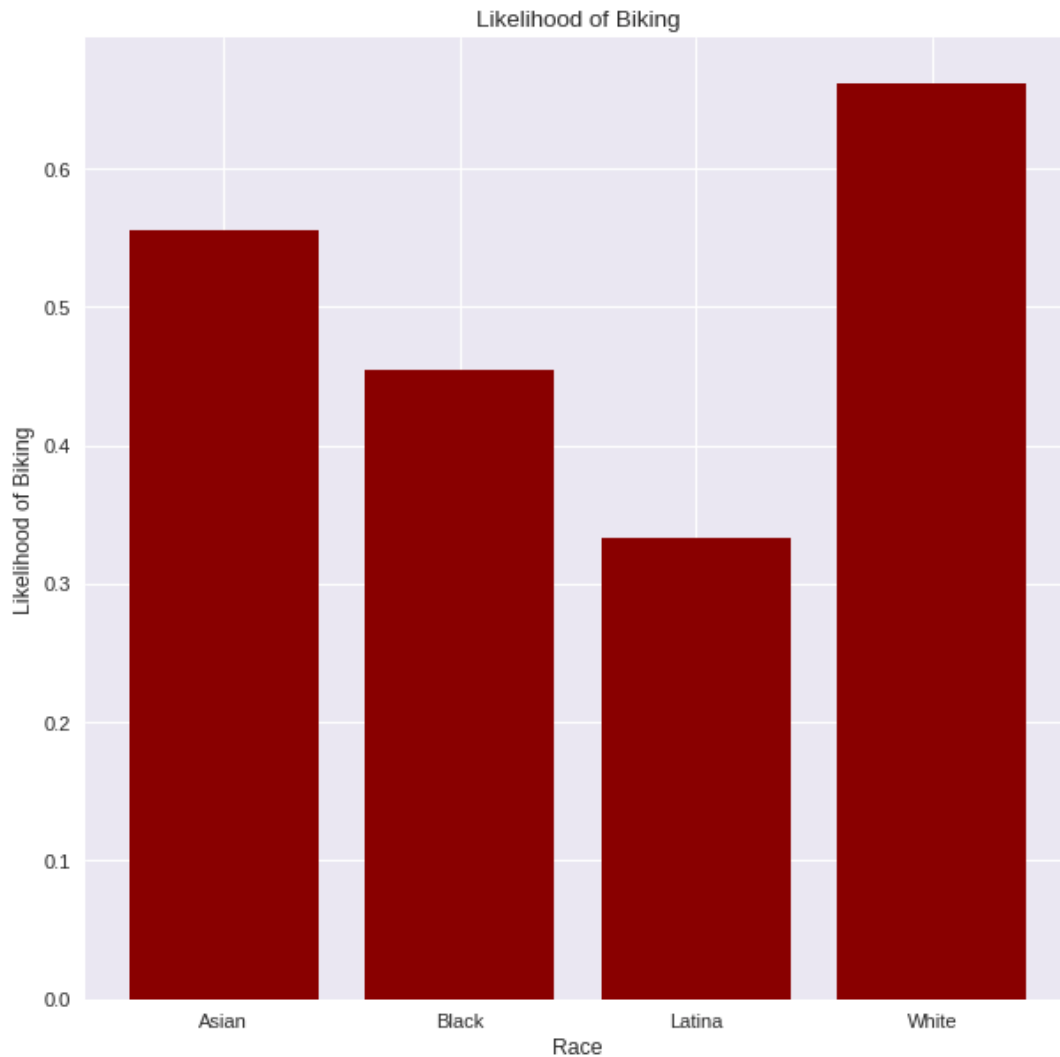
Figure 14: Likelihood of Respondents Biking by Race: *Comparing the frequency of walking in Figure 11 to this graph shows the variability of respondents likelihood to bike. Comparing this graph to Figure 13 also shows that people are less likely to bike than walk. Latinx is particularly low, but this could be due to the limited survey respondents of that race.*

2.4 Activities Based on Zip Code

As OpenStreetsPGH moves locations throughout the summer, we thought it would be valuable to see if specific events could increase participation in certain neighborhoods.
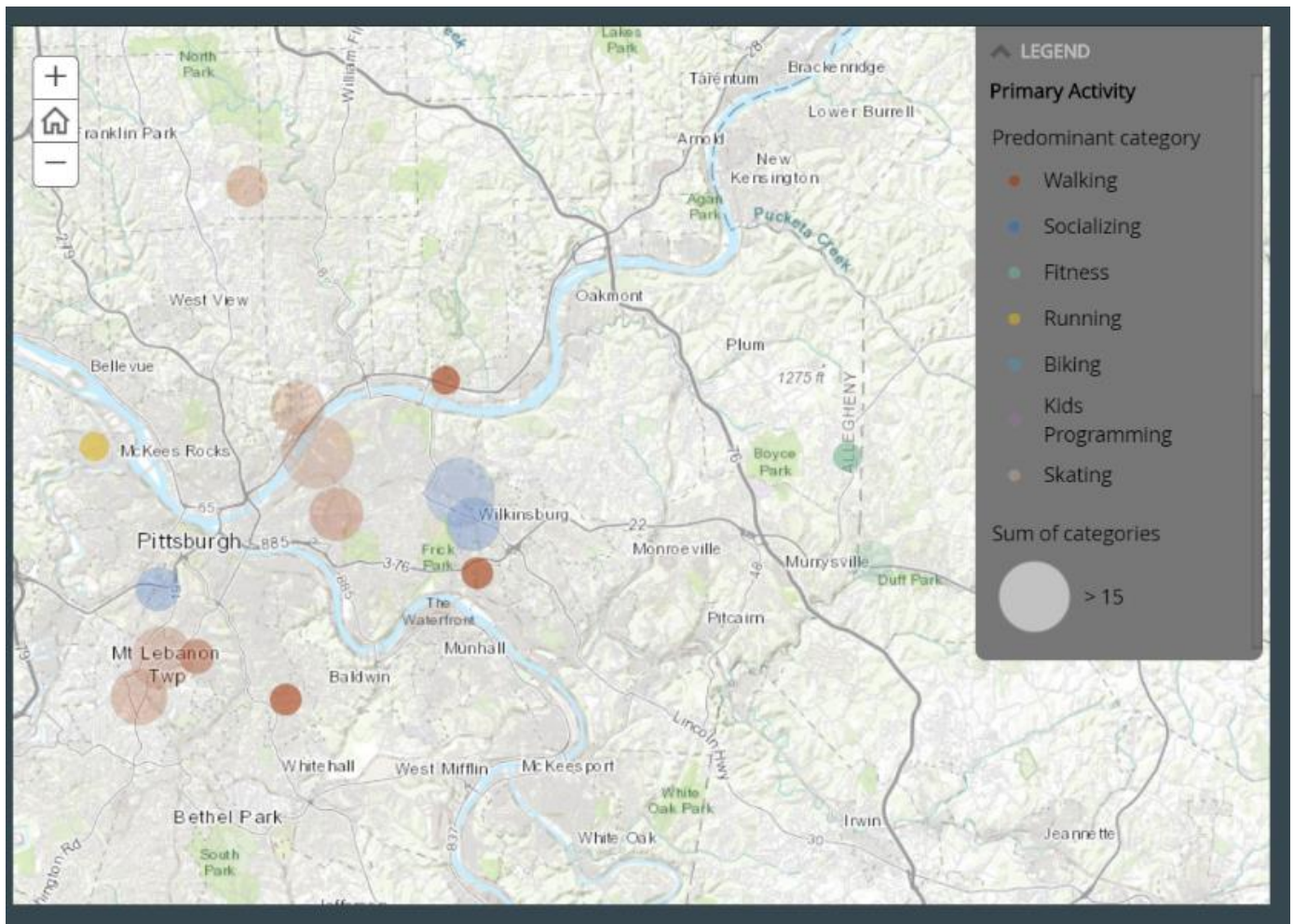

Figure 15: Primary Activity by Zip Code (Mapped using ArcGIS). *Based on this map, socializing seems to be a large draw in far east neighborhoods such as Wilkinsburg and Mt. Washington, whereas North Side, Strip District, and Downtown residents prefer walking.*
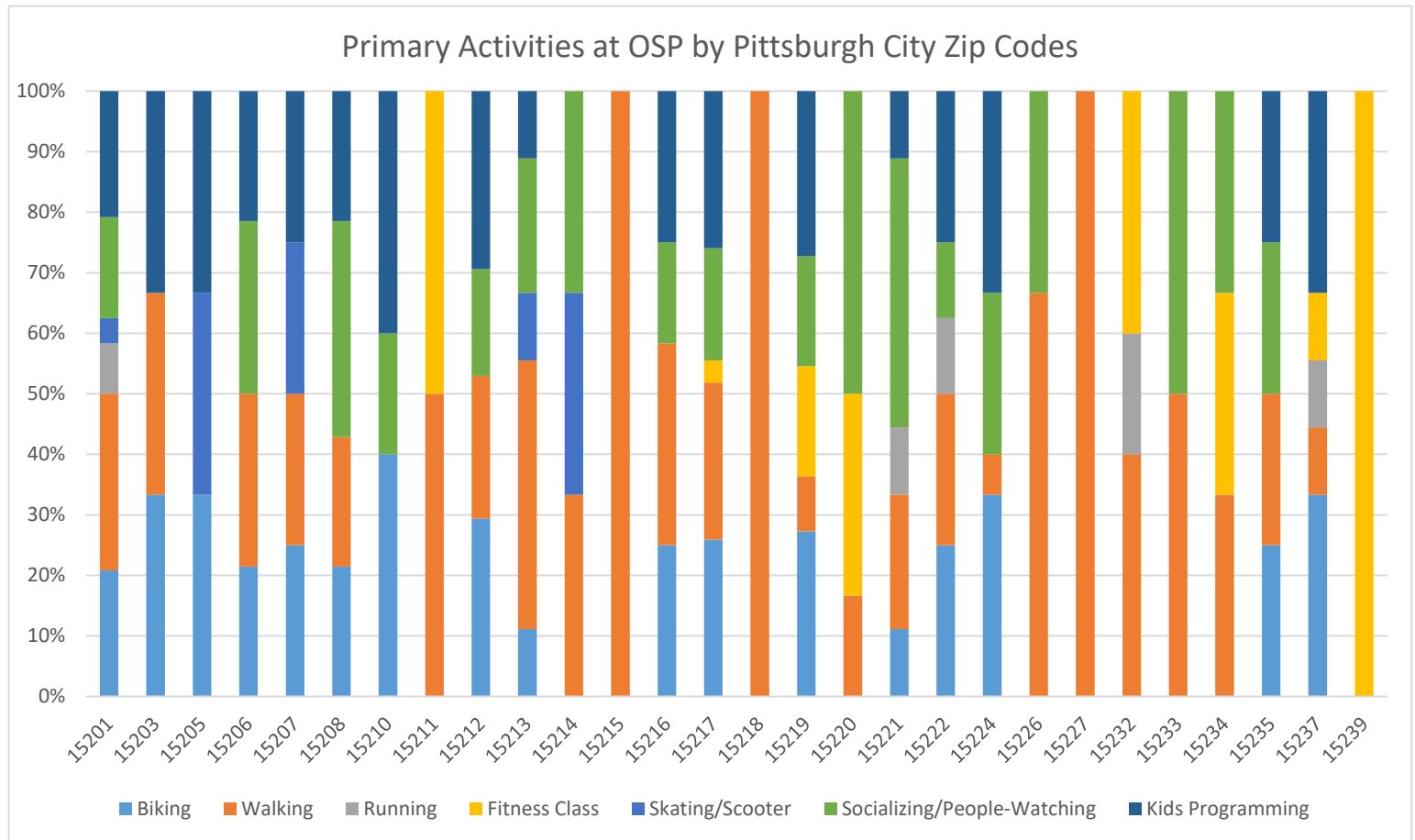
Figure 16: Primary activity by Pittsburgh city limits Zip Codes (Percentage). *Due to limited survey respondents per Zip Code, this graphic should be viewed as a general overview of primary activities. Some Zip Codes had < 5 respondents, so this is an example of how increasing survey responses could increase reliability of this study.*

3. Suggestions for Future Surveys

Below are some suggestions to improve the integrity and analyzability of future iterations of the survey.

1. The survey should be segmented by the specific OSP event. This comes in as an important point because in order to have a better understanding of who comes to which events and why, it is important to know which event people attended. This will
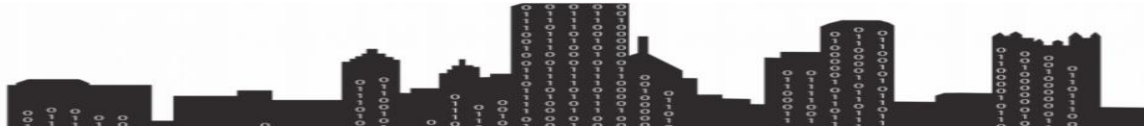
add whole new levels of analysis based on location and time of year that was noticeably lacking in our analysis. We propose this to be the first question on the survey.

2. Race or Ethnicity could be a Select One option. While giving participants the option write-in this answer allows fuller self-expression, the manual work required to clean answers for categorization takes time and is prone to error. We suggest the following bins:
    a. American Indian or Alaska Native
    b. Asian
    c. Black or African American
    d. White
    e. Hispanic or Latinx
    f. Two or More Races

3. Consider the length of the survey, especially for groups "on the go" such as parents or younger attendees. While we can't comment on the time this survey took, a shorter selection of questions could result in more people agreeing to complete the survey. SurveyMonkey has a great article about ideal survey length: https://www.surveymonkey.com/curiosity/survey_completion_times/

4. Lastly, more responses would be ideal for better analyses. Regardless, a push for more respondents over the course of the events will allow more in-depth and accurate information that could aid BikePGH with their mission.

Appendices

Appendix A: Map Analyses

These maps show different attributes based on the survey respondents Zip Code. They were created in ArcGIS. Please see links at the end of this section to further explore the maps.
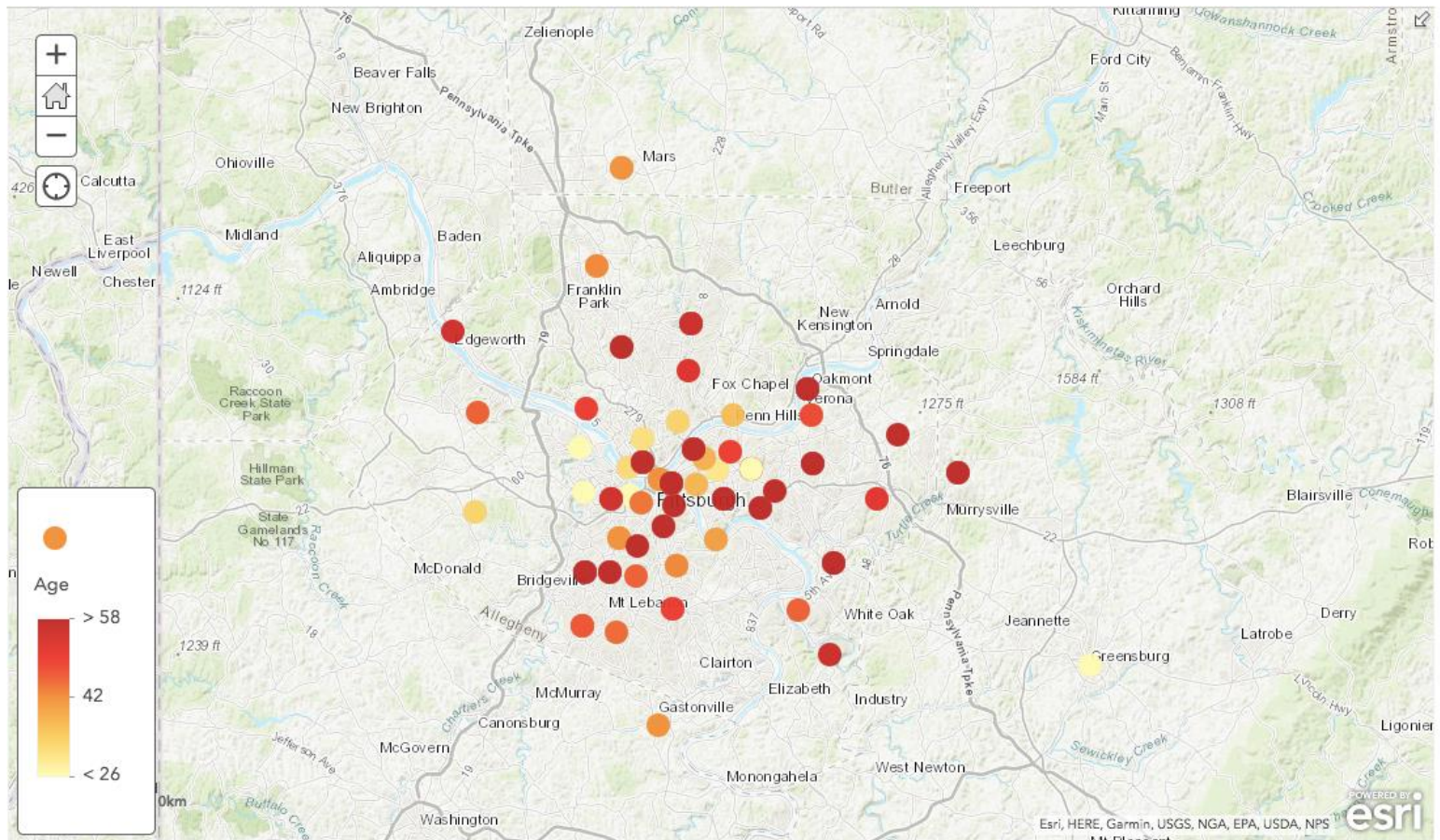

Figure A-1: Age by Zip Code. *The Pittsburgh area skews older than the national average. This is reflected in our analysis, as OSP attendees were mostly above 40 in age. Younger people were more prominent in the East End and North Side neighborhoods.*
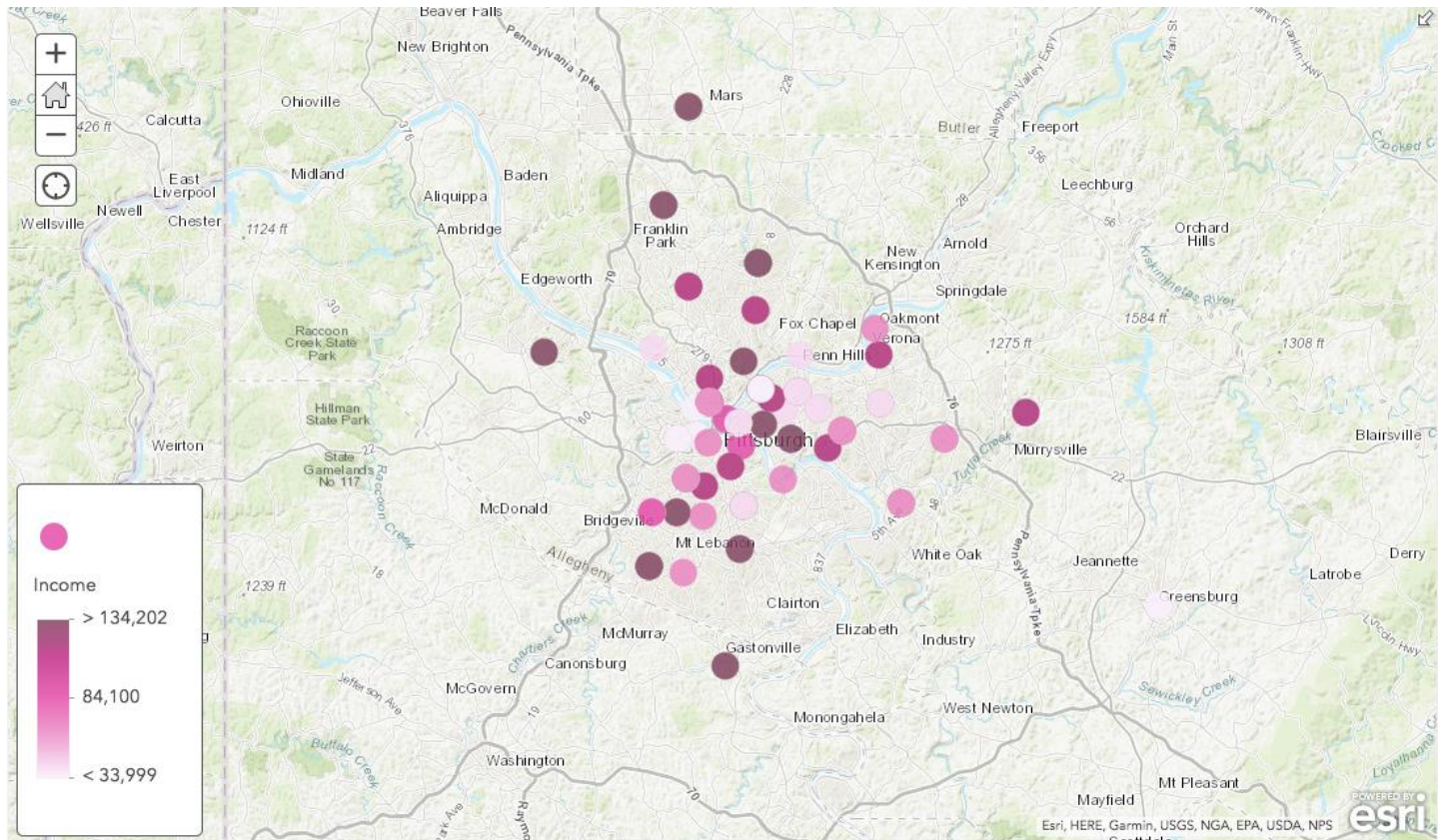
Figure A-2: Income by Zip Code. *Higher incomes were reported in some East End neighborhoods and outer suburbs. As we discovered in our earlier analyses, this event seems to have a fairly equal distribution of attendees by incomes.*
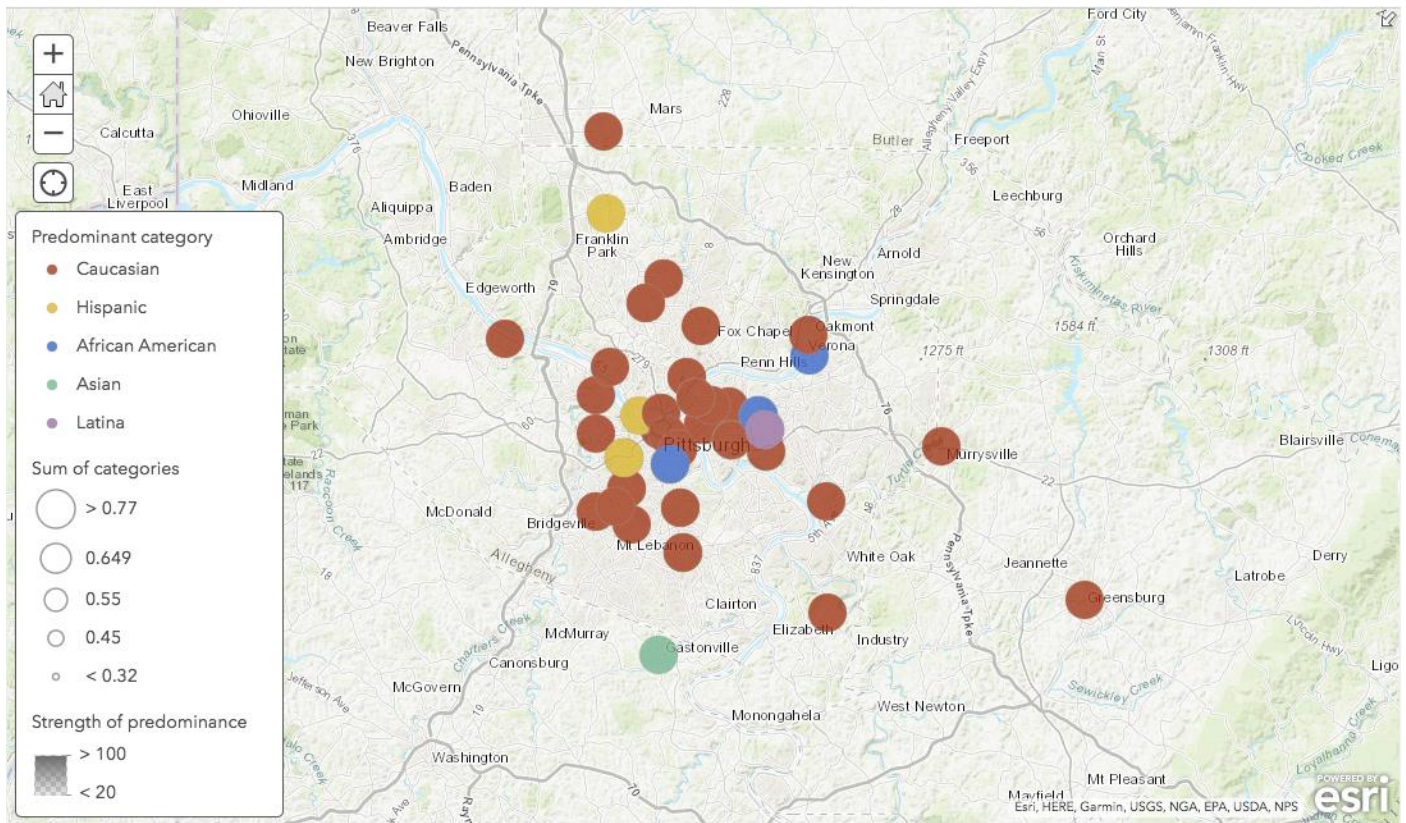
Figure A-3: Race by Zip Code. *This maps aligns with assumptions that Pittsburgh has neighborhoods that are predominately white.*

Appendix A: Map Links

Race ArcGIS:
https://www.arcgis.com/home/item.html?id=2145750f4abd46908a89aacdc9bfb03a

Age ArcGIS:

https://www.arcgis.com/home/item.html?id=fe6084e90b9e444cb62dc451d9f4fcdf

Income ArcGIS;

https://www.arcgis.com/home/item.html?id=efde1cbe303d4c6a8b7fa7588ea21fbb

Primary Activity ArcGIS:

https://www.arcgis.com/home/item.html?id=d62dc4fd966c408ab872c0f415c2be78

Appendix B: Links to Code

Github Links to various analyses of this project.

https://github.com/ml42322/BikePGH/blob/master/BikePGH.ipynb

https://github.com/sagnikrana/BikePGH---SUDS/blob/master/SUDS_Maitri.ipynb