Ali Carmichael

ADS4SJ 410

06/09/25

Eugene and Springfield Demographics That CAHOOTS is Serving

Background

The research questions I would like to answer in my final project are about the demographics that CAHOOTS works with. The questions I chose are: What populations are CAHOOTS serving, and how do these populations differ between Eugene and Springfield? I chose these questions because I think it would be interesting to look closer at the demographics of the calls at CAHOOTS and discover patterns in the populations they are serving. This could be helpful in identifying communities that might be affected by the loss of CAHOOTS which I am interested in discovering.

The data set that we are provided with does not contain any demographic data for the calls so by using another data set in combination with the CAHOOTS data set I can determine the demographics of the different call types over time and in different locations. This is important in understanding if certain populations in Eugene are more vulnerable to certain call types and provides CAHOOTS with information on areas that may need more attention in the future of the organization. It might also help determine if there are temporal patterns in certain populations that might coincide with larger events during that period. Overall, I think these questions can help in seeing if certain groups of people are more impacted by high or low-priority calls and which could help us understand what groups will be most impacted by the loss of CAHOOTS

Data

The data I will use is the CAD call data that provides the call type, length of the call, date of call, priority, reason for the call, and zip code. The other data set that I will use is a compilation of census data from the Census Reporter for the different zip codes found in the call data set. This data set will include things like median income, age, race/ethnicity, and education. This data will be several different data sets for each zip code found in the CAHOOTS dataset.

Call Data

- Strengths: Detailed call log of the calls that CAHOOTS has responded to
- Limitations: Does not include demographic data and is not sorted by CAHOOTS calls only.

Demographic data

- Strengths: No NA values in the data set and already set up in a format where values are counts
- Limitations: Data is formatted weirdly so it is not easy to load into a notebook as is. Some races are also excluded and little detail is given about people who are more than one race.

- **Age Data:** https://censusreporter.org/data/table/?table=B01001&geo_ids=86000US97401,05000US41039,31000US21660,04000US41,01000US,860|31000US21660&primary_geo_id=86000US97401#valueType|estimate
  - Columns: Zip Code for Eugene-Springfield, OR Metro Area
  - Rows: Age intervals (ex: 25 to 29 years)
  - Values: Counts of people in that age interval and zip code
- **Income Data:** https://censusreporter.org/data/table/?table=B19001&geo_ids=86000US97401,05000US41039,31000US21660,04000US41,01000US,860|31000US21660&primary_geo_id=86000US97401#valueType|estimate
  - Columns: Zip Code for Eugene-Springfield, OR Metro Area
  - Rows: Household income intervals (ex: $60,000 to $74,999)
  - Values Counts of people in that age interval and zip code
- **Education Data:** https://censusreporter.org/data/table/?table=B15002&primary_geo_id=86000US97401&geo_ids=86000US97401,05000US41039,31000US21660,04000US41,01000US
  - Columns: Zip Code for Eugene-Springfield, OR Metro Area
  - Rows: Education Level (ex: Bachelor's degree)
  - Values Counts of people with that education level and zip code

- **Race/Ethnicity Data:** https://censusreporter.org/data/table/?table=B03002&primary_geo_id=86000US97402&geo_ids=86000US97402,05000US41039,31000US21660,04000US41,01000US
  - Columns: Zip Code for Eugene-Springfield, OR Metro Area
  - Rows: Race or Ethnicity (ex: Asian alone)
  - Values Counts of people of that race or ethnicity and zip code

Data Cleaning:

- Load in call data using PANDAS read_csv() function
- Create a function called unit_call_sign to determine the responding unit to the call
- Using unit_call_sign filter the call. Data to just calls responded to by CAHOOTS
- Drop columns that are unnecessary for analysis and filter out any NA values in any of the rows
- Convert zip codes to integer values
- Load in the four demographic data sets using PANDAS read_csv() function

- Rename columns for better formatting
- Convert demographic data to a long format using the PANDAS melt() function, changing variable names to zip code and value names to count_demographic
- Convert long demographic data to a wide format using the sum function
- Drop the total column from the wide tables
- Combine the demographic data into one data set using PANDAS merge() function on zip code using an outer merge
- Create a call volume table by counting the number of calls for each zip code
- Merge the call volume table and demographic combined data frame into one data frame using PANDAS merge() function.
- Create a function called eugene_or_springfield that assigns each entry a city in a new column called "city".
- Save the final clean data frame as a csv to load into the analysis notebook.

Data Cleaning Notebook:

https://github.com/alicarmichael/ADS4SJ_Final_Project/tree/main

Packages: PANDAS and re

Methods

Data Analysis:

- Load in clean full data frame as a csv using PANDAS read_csv() function
- Create a total population column for each zip code by summing an exclusive demographic variable
- Normalize the call volume across varying demographic populations by creating a calls_per_1000 columns
- Calculate the correlations between the demographic variables and the normalized call volume and plot them using a bar plot
- Using the correlations, regroup the age, income, and education intervals for analysis so each interval contains a relatively even population size
- Drop all the columns after regrouping that will not be used
- Create call volume quartile areas using the PANDAS qcut() function and save them as a column in the data frame
- Create a new data frame that groups by the quartiles saving them as rows and each of the demographic intervals as columns, leaving the call volume counts as values
- Subset the data frame by demographic columns and then melt them into separate data frames for plotting
- Create subplots of the demographics by call area quartile
- Plot the call area quartiles as a heatmap
- Create scatterplots for each demographic and normalized call volume coloring by city
- Create new correlations between demographics and call volume normalized

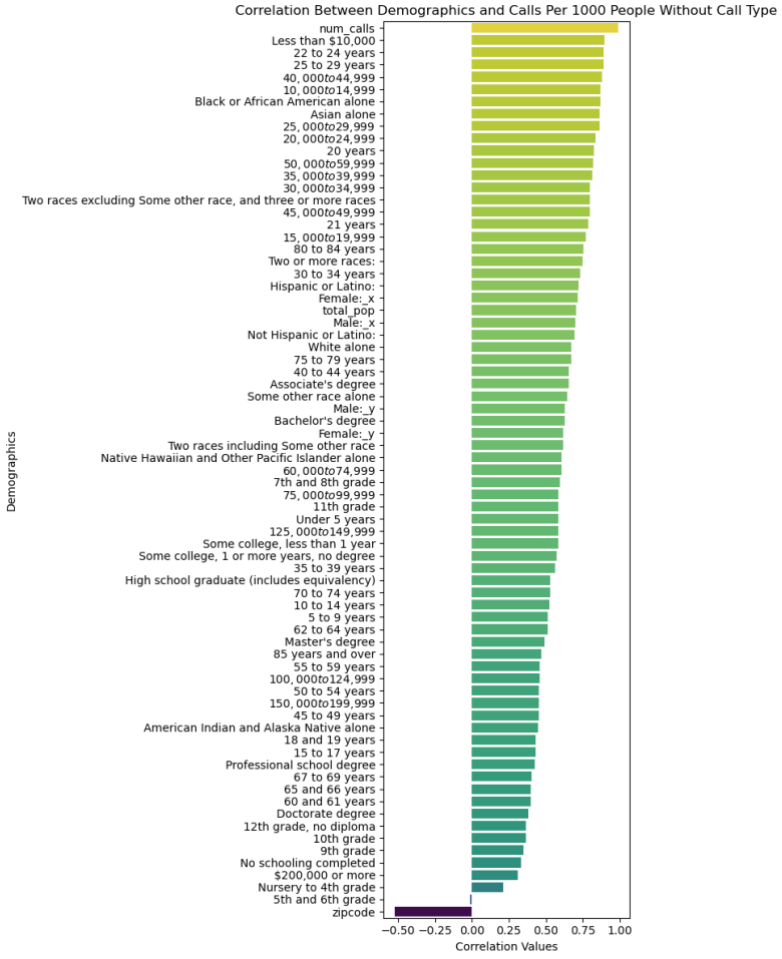
- Drop columns only keeping city as rows and demographic values as columns
- Create a chi_squared_test() function to calculate the observed – expected difference and plot it as a heatmap for each demographic group and city

Data Analysis Notebook:
https://github.com/alicarmichael/ADS4SJ_Final_Project/tree/main

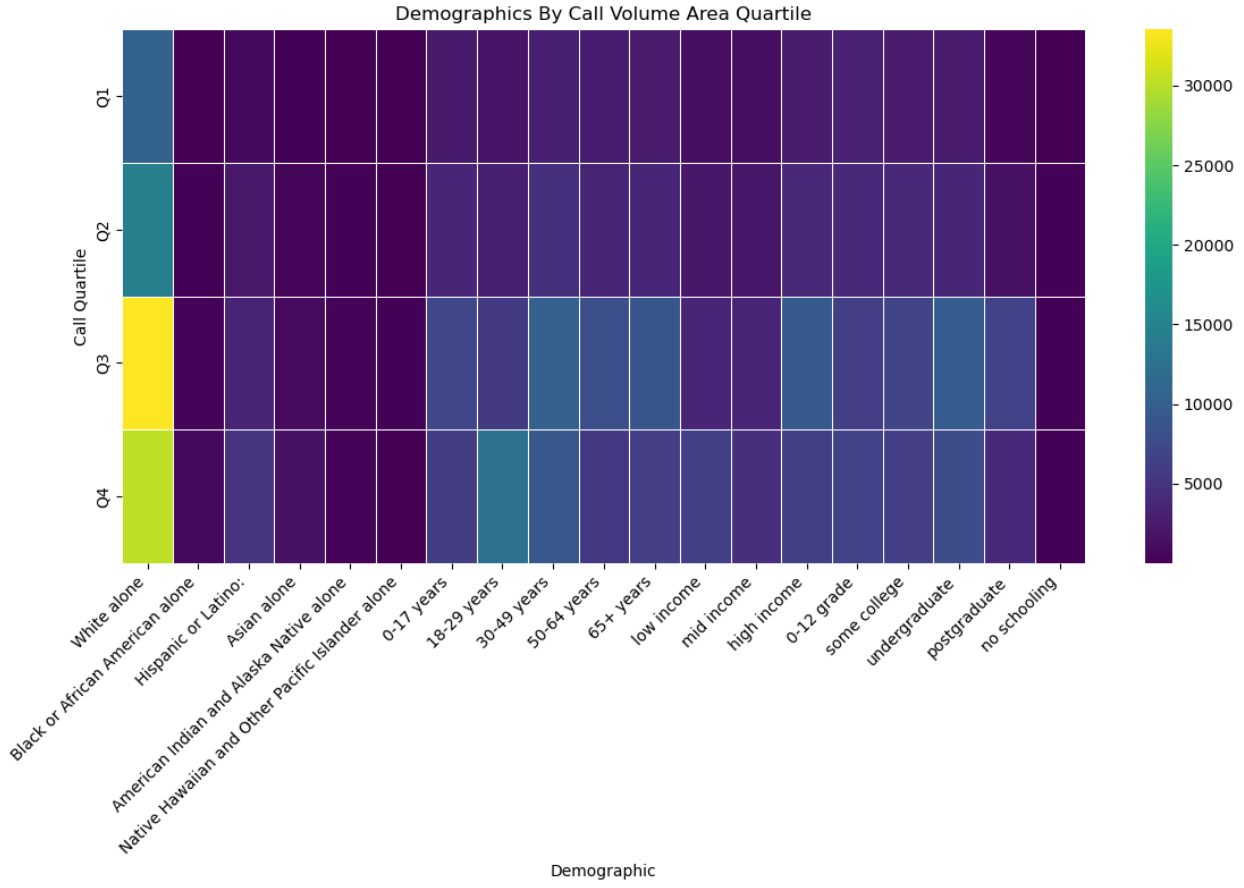
Packages: PANDAS, Seaborn, Matplotlib, Scipy.stats

Results

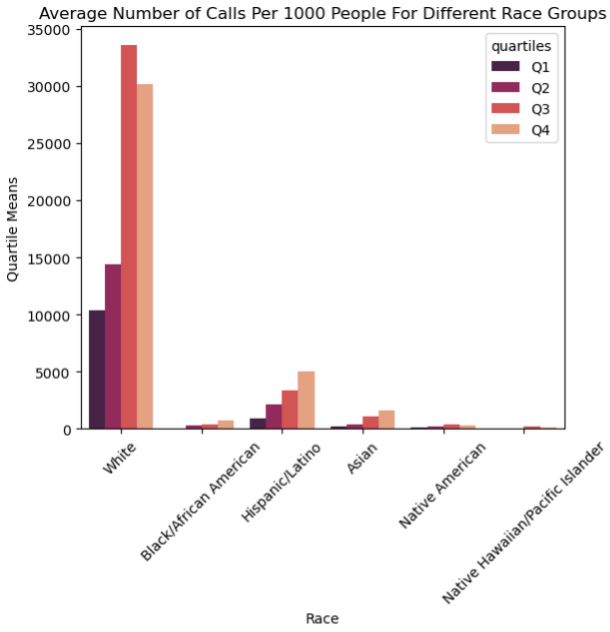
My first plot that I made was a bar plot of the correlation values for each of the demographic variables and the calls per 1000 people. This was before I regrouped the demographic variables into more even intervals. The demographic variables with the highest correlation values were area with income less than $10,000 at a correlation value of around 0.96, ages 22 to 24 at around 0.84, ages 25 to 29 at around 0.83, and income $40,000 to $44,999 at around 0.81. Black or African American had the highest correlation value of any of the race demographics at about 0.80. The lowest correlation values were income over $200,000, which had a correlation value of around 0.28, and elementary schooling levels, which were close to zero.

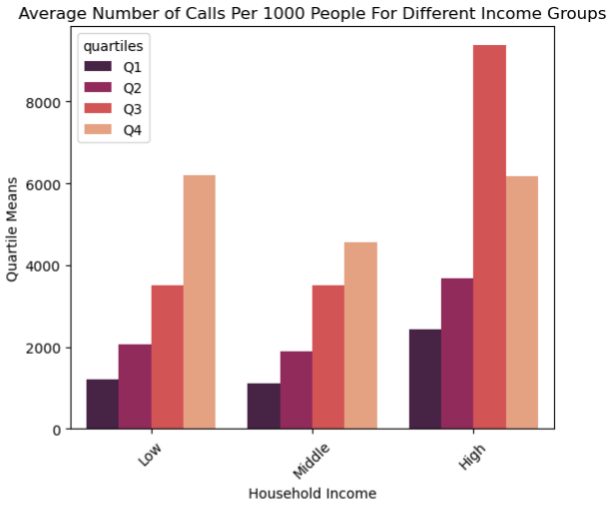


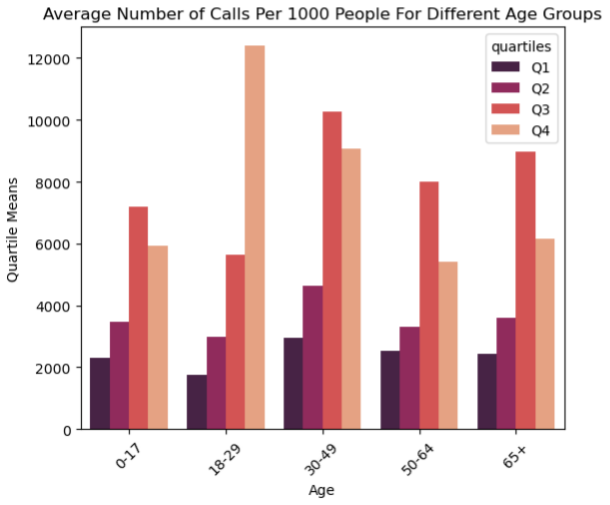
This plot is a heatmap of the demographics split up into the new demographic groups. This heatmap shows the call area quartiles, with Q4 being areas with the highest number of calls and Q1 being areas with the lowest number of calls. The squares contain the number of people that live in these areas, who are of the different demographic variables. In high call areas, there is a large portion of white people and individuals aged 18-29. In lower call areas, there are more young people and fewer individuals with more schooling. The high call areas also feature these demographics, likely due to the high call areas being more populated areas of Eugene and Springfield, so they contain more diverse demographics. A call hotspot area was downtown Eugene, which includes the University and all the college students who live nearby; these areas tend to be more dense and diverse.

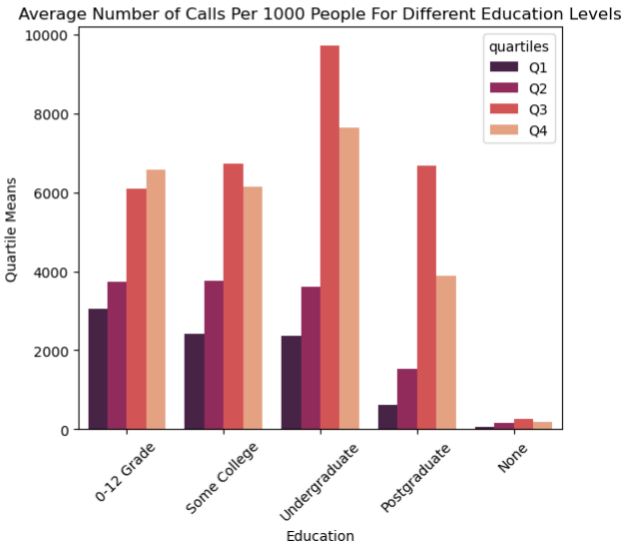


My next plot is similar to the plot above, but shown as a bar plot. Each plot is split up into the demographic categories, Race, Income, Age, and Education. The x-axis is the demographic intervals for each category, and the y-axis is the average number of people for each call quartile. The bars are colored by quartile, with Q4 again being the area with the highest number of calls and Q1 being the area with the lowest. Again, there is a large population of white people who live in areas with moderately high to high call volume, as well as a large population of people aged 18-29. Most other age demographics are found in areas that have moderately high call volumes. Areas with lots of people who are high-income and people who have an undergraduate degree are also found in moderately high call volume areas. There are no unexpected results like a certain demographic having a high population in a low call area.

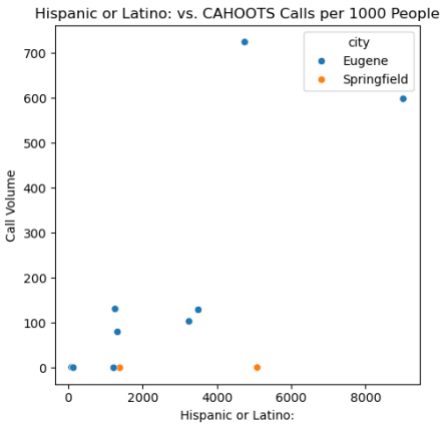


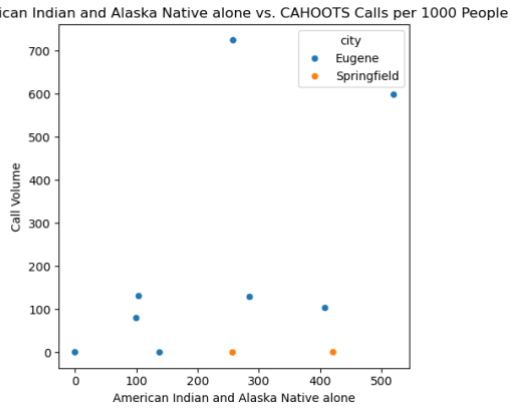


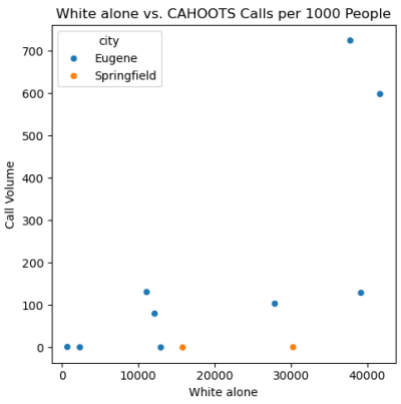


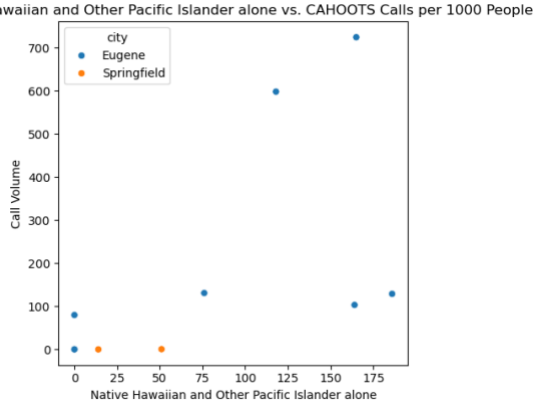


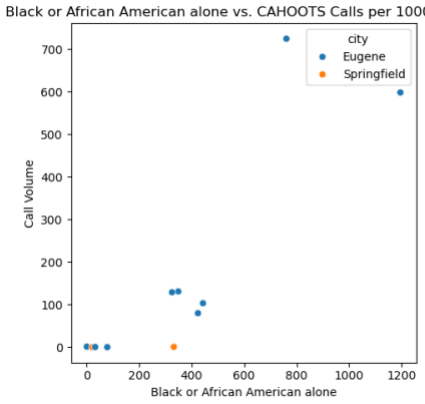
These are the scatterplots I made to show call volume by city for each demographic. The number of calls for demographics in Eugene tends to fluctuate depending on the demographic variable, while Springfield stays much lower. An important thing of note, there are more zip codes for Eugene than there are for Springfield, which leads to unbalanced analysis. Eugene has some zip codes that receive a lot of calls in areas where there are more People of Color, like Hispanic, Asian, and Black, as well as areas with lower income and older aged individuals.

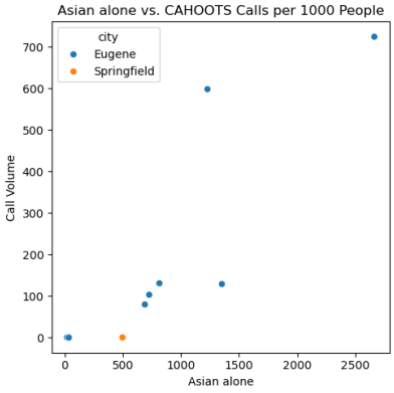


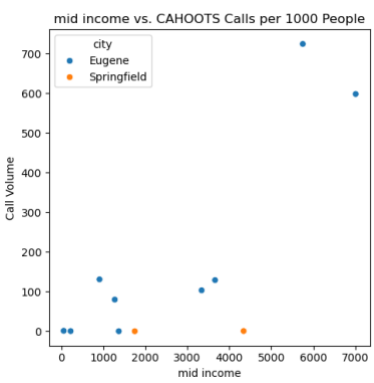


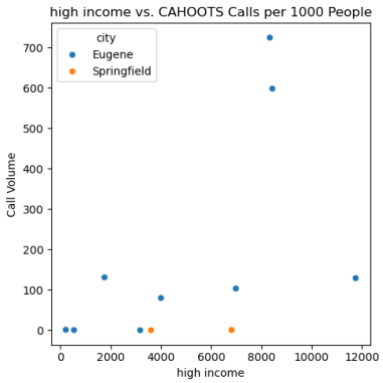


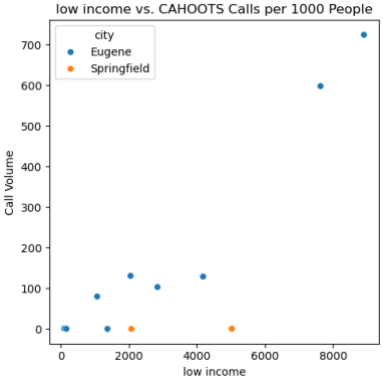


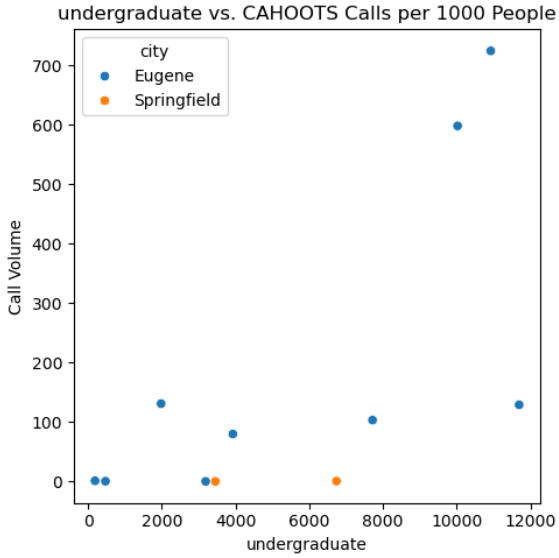


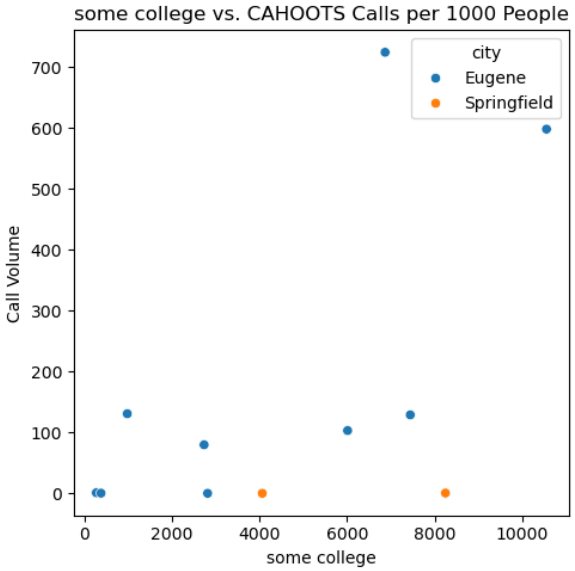


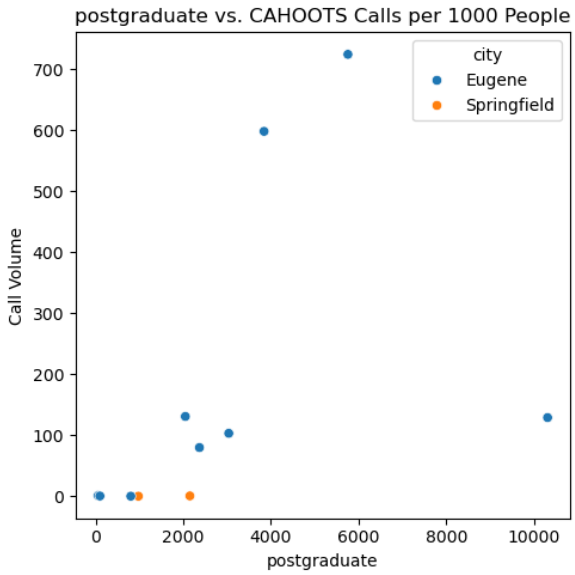

mid income vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
mid income
0
100
200
300
400
500
600
700
0
1000
2000
3000
4000
5000
6000
7000

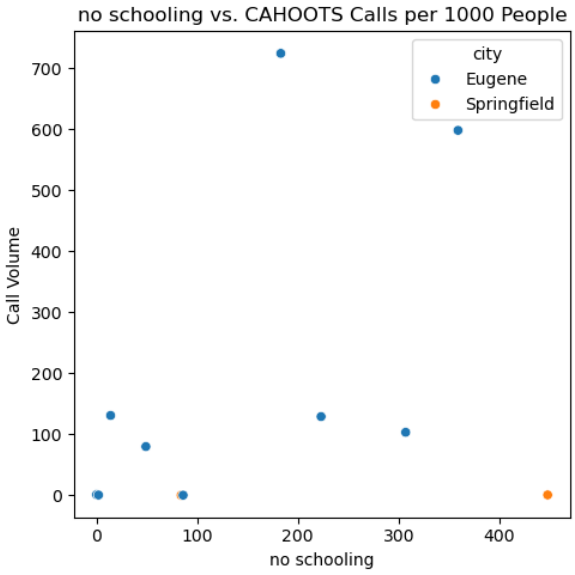

high income vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
high income
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000
10000
12000

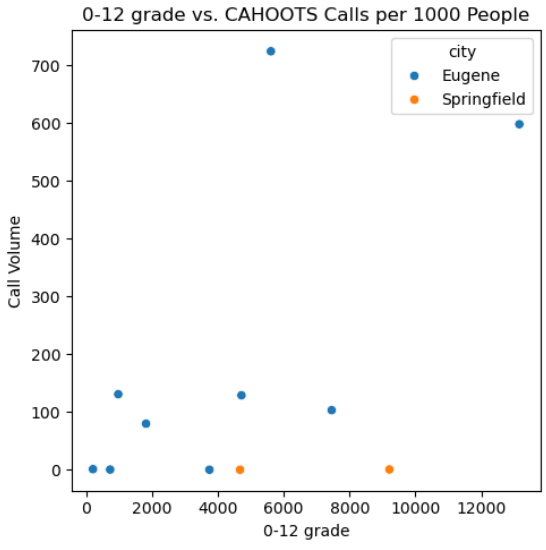

low income vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
low income
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000

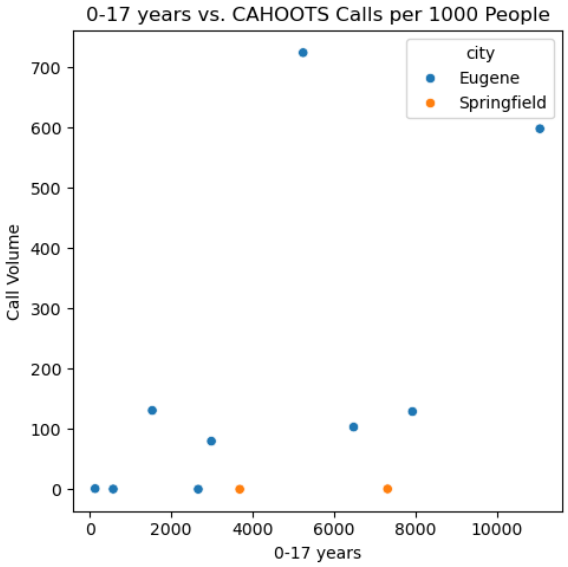

undergraduate vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
undergraduate

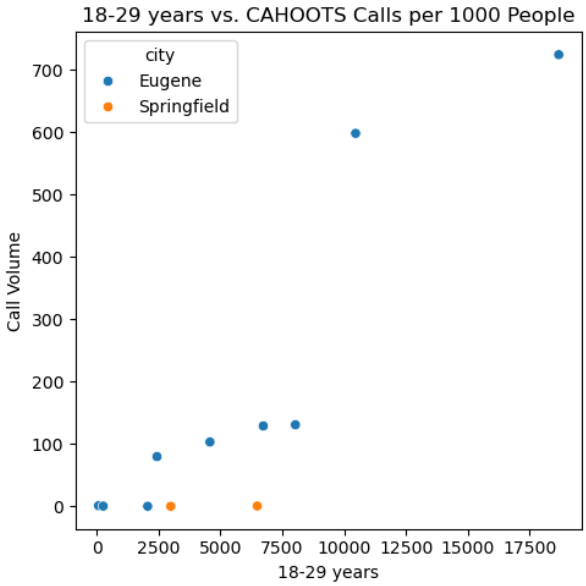

some college vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
some college

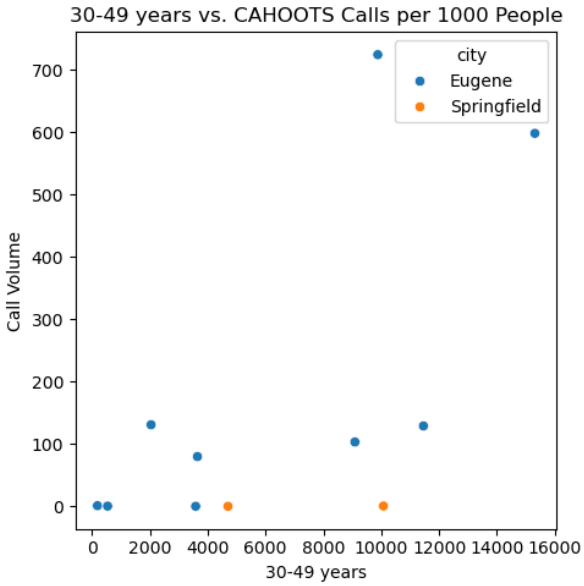

postgraduate vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
postgraduate

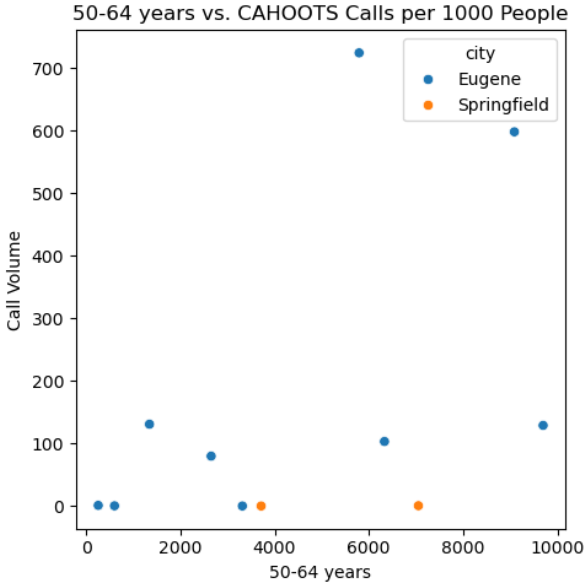

no schooling vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
no schooling

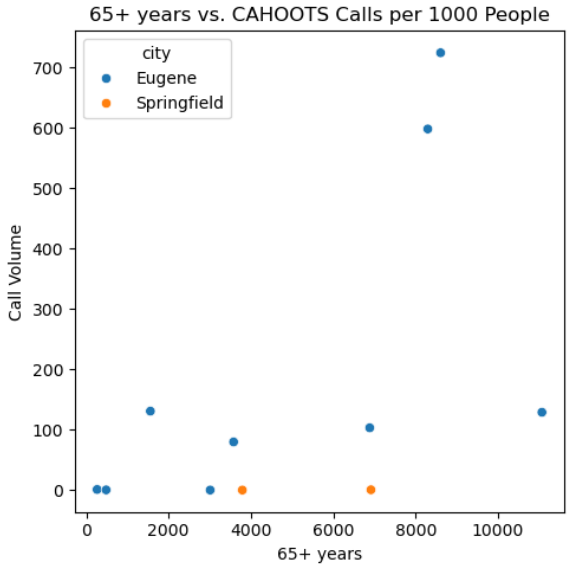

0-12 grade vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
0-12 grade


0-17 years vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
0-17 years
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000
10000

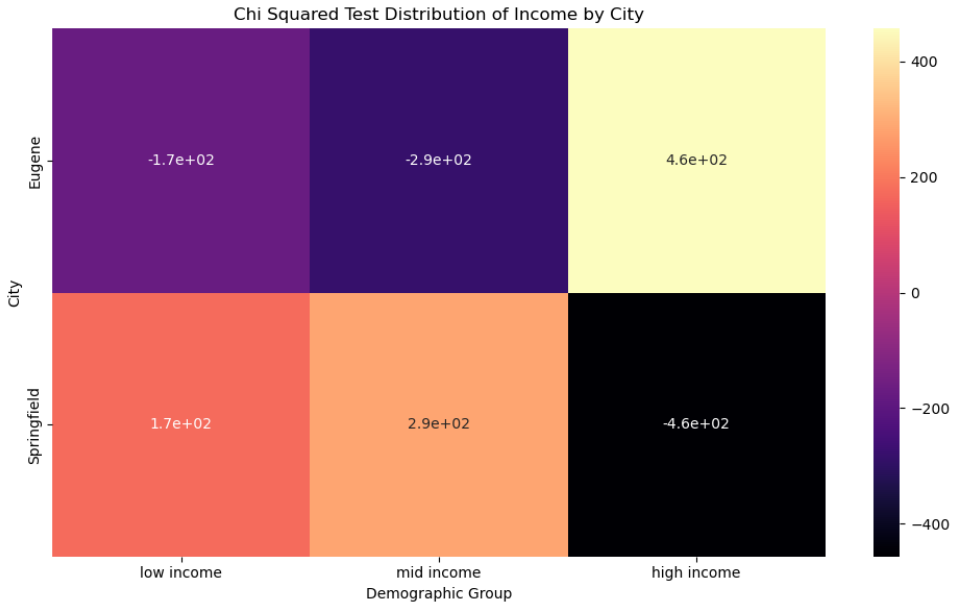

18-29 years vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
18-29 years
0
100
200
300
400
500
600
700
0
2500
5000
7500
10000
12500
15000
17500

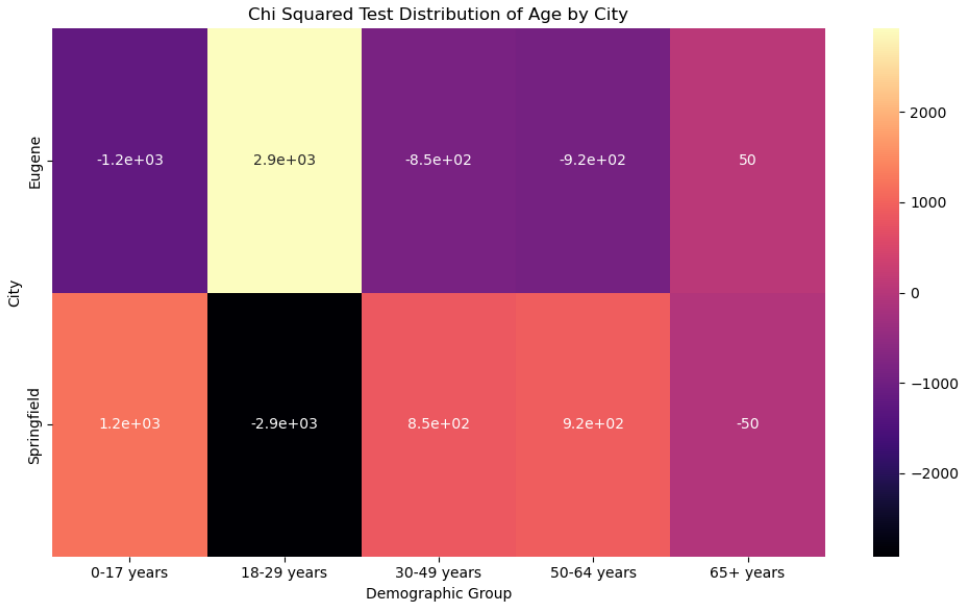

30-49 years vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
30-49 years
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000
10000
12000
14000
16000

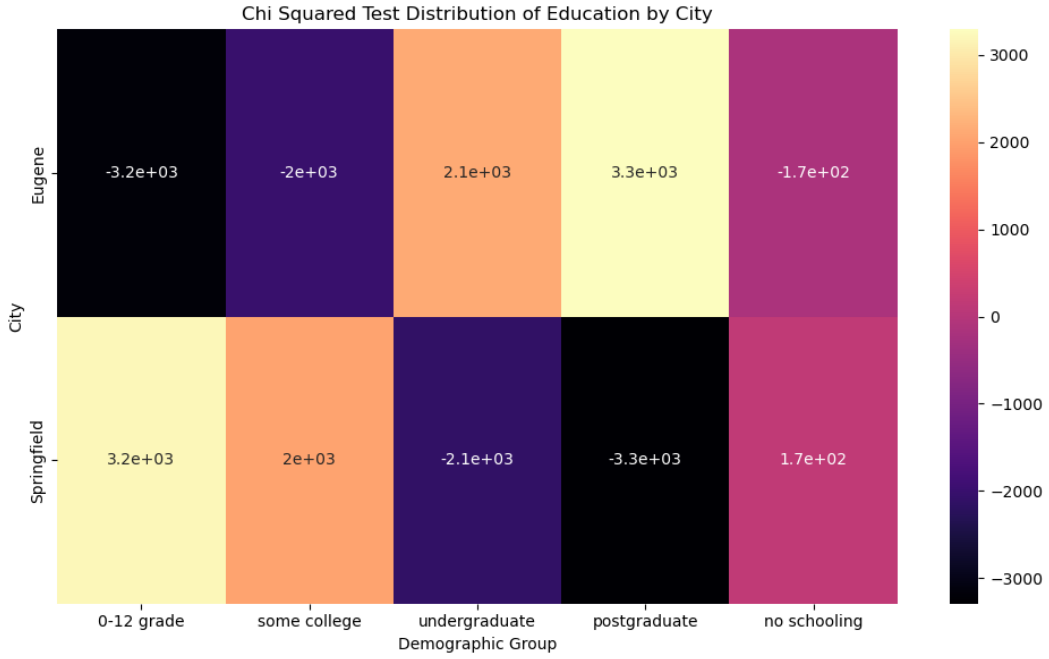

50-64 years vs. CAHOOTS Calls per 1000 People
city
Eugene
Springfield
Call Volume
50-64 years
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000
10000



For my statistical analysis, I did a chi-squared test to see if there was a difference in the Eugene and Springfield demographics. For all four demographic groups, the p-values were significant, and all were very close to zero. Looking at this further, I realized there was a much larger spread in the sample sizes than I originally thought so while the tests gives support to my hypothesis that Eugene and Springfield will have a difference in their demographics, these results may be not significant and are just due to the lack of data for Springfield, further analysis is needed. These heatmaps show the difference in the observed and expected values and show what demographics may be underrepresented and overrepresented in each city. High-income individuals and those aged 18-29 are both underrepresented in Springfield, while Eugene's Hispanic and mid-income populations are underrepresented.


Chi Squared Test Distribution of Income by City
Eugene
Springfield
City
-1.7e+02
-2.9e+02
4.6e+02
1.7e+02
2.9e+02
-4.6e+02
low income
mid income
high income
Demographic Group
400
200
0
−200
−400


Chi Squared Test Distribution of Age by City
Eugene
Springfield
City
-1.2e+03
2.9e+03
-8.5e+02
-9.2e+02
50
1.2e+03
-2.9e+03
8.5e+02
9.2e+02
-50
0-17 years
18-29 years
30-49 years
50-64 years
65+ years
Demographic Group
2000
1000
0
−1000
−2000



Discussion

In conclusion, I found that CAHOOTS serves a wide variety of communities in both Eugene and Springfield. Areas where CAHOOTS receives the most calls are often highly populated areas like downtown Eugene, where there is a larger population of younger people. There is a much larger population of White people than any other race in both Eugene and Springfield, but there are still some Eugene zip codes that receive a decent amount of CAHOOTS calls that have more Hispanic, Asian, and Black people who live there.

CAHOOTS also serves people of all different educational backgrounds, and there is no particular area they visit more often that has more people living there with less schooling. They often get calls from areas where there is a large number of people with some college or a bachelor's degree. Calls also come from areas with different levels of income, and there was no particular area that received a high call volume that was of a certain income interval.

From my analysis, I found that demographic populations do differ between Eugene and Springfield, but further investigation is needed to determine if this difference is significant or just due to a lack of data for Springfield. Specifically, Springfield has a smaller population of younger individuals (18-29) and a higher proportion of people with higher incomes. Eugene has a smaller population of Hispanic individuals and a higher proportion of middle-income households. Both Eugene and Springfield have an overwhelmingly high population of White people in comparison to the populations of other races.

If I were to continue this project, it would be nice to get demographic data from CAD that was directly linked to the calls. It was difficult to make direct conclusions on any trends or significance I found because the demographic data I used was separate from the CAD call data and was only linked by zip code, so I could only infer correlations and not direct causations. I also think that it would be interesting to do a similar analysis to what I did with the call area quartiles and the number of calls for demographics with call type to see if there was an association between call types and demographics. It might also be interesting to take demographic data for Eugene and Springfield rather than separating each zip code into their respective cities to do more analysis on the difference between Eugene and Springfield demographics that CAHOOTS works with. Overall, I believe this project was helpful in understanding what communities might be impacted by the loss of CAHOOTS services in Eugene, as well as future impacts with the loss of funding for CAHOOTS.