Does light density have an impact on youth experiencing homelessness in the SF area.

2) My motivation for this question stems from my personal experiences as a former foster youth who had their own hurdles of fighting housing insecurity. I am a data analyst for the SFYEAH research team currently geolocating areas of reported violence against youth experiencing homelessness in SF. I'm interested in this topic as I want to dig deeper into the factors that are leading to violence against youth experiencing homelessess in SF. I hope that this research will assist policy makers and programs like Larkin serving homeless youth in SF.

3) I plan to use the regression discontinuity model to analyze my data because I want to see the correlation between light and reported areas that youth labeled "unsafe".

4) $Y = \beta 0 + \beta 1 \cdot (light) + \beta 2 \cdot (pop) + \beta 3(med\_income) + \beta 4(pop * light * med\_income) + [...] + u$

b)

- Light = light radiance recorded by satellite images of the night sky in SF.
- Med_income = Medium Income Per household in SF neighborhoods
- Population density = population density per district based on square mile.

c) What the ideal dataset would look like:
The ideal dataset that I sought after would include the population density of each district of SF with the entailed light radiance of said district with the proportion of violence recorded against homeless youth. Also the time period of each varies, the light distribution I was able to collect is from 2016. The crime data I collected for homeless youth was collected only from the summer of 2018. If the homeless youth data was collected over a longer duration of time we would have a more accurate distribution of data to depict the true nature of the homeless youth population's experiences. The current youth data is pooled cross sectional data. Another factor that I would like to include is police data in comparison to the homeless youth data. Police data required too much cleaning for me to include in a timely manner for this project unfortunately. Lastly, additional factor(s) such as the exact time these youth experienced what they reported in addition to socio-economic information about the SF districts would also be ideal to include.

5) A quick description of the anticipated sign, magnitude and significance of the coefficients associated with my main explanatory variables.
Through my analysis I found that light does not play a significant role in crime/violence against homeless youth. Initially I was under the assumption that lower light levels would contribute to higher levels of reported violence. What my model showed was that homeless youth actually reported feeling safe and unsafe in areas where there were higher levels of light radiance recorded in the SF area. Through extended research I found that this was reinforced by published articles that light and crime had no significant correlation. I expect the light coefficient to be neither positive nor negative, but it's very possible that it's slightly positive. I expect medium income per household to be significant with a positive coefficient as in my model you can see that 'unsafe' experiences were recorded in lower medium income neighborhoods.

Population density seemed to be significantly correlated with light radiance. I expect this to have a positive sign. Where people congregate light follows, and where light is people congregate. Population density was more significantly impactful in crime than purely the degree of light radiance. Despite this there are still a myriad of socio-economic confounding variables that lead

to the youth experiencing homelessness that I was unable to identify within the time span of this project.

What's interesting to note is that a regression done on solely the data of the youth experiences one finds that safety and basic needs are not surprisingly very correlated. Sadly, when you regress unsafe and basic needs you also find a significant correlation that has a higher p-value.

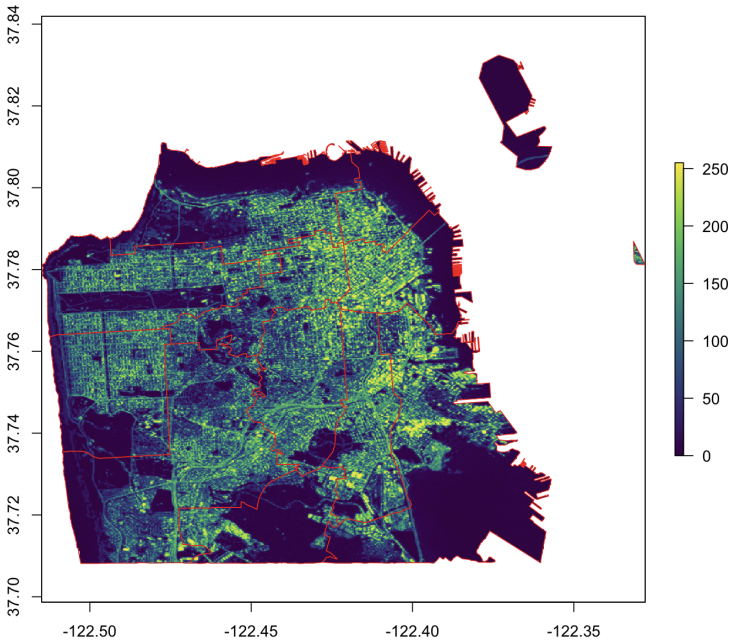
- *lm(formula = pl_safe ~ re_basic, data = crime_df):*
  Coefficients:
  Estimate Std. Error t value Pr(>|t|)
  *(Intercept) 1.000e+00 1.293e-16 7.732e+15 <2e-16 ****
  re_basic -2.225e-16 1.957e-16 -1.137e+00 0.257
- *lm(formula = pl_unsafe ~ re_basic, data = crime_df):*
  Coefficients:
  Estimate Std. Error t value Pr(>|t|)
  *(Intercept) 1.000e+00 2.069e-16 4.833e+15 <2e-16 ****
  re_basic 2.190e-16 6.080e-16 3.600e-01 0.72

6) *The main assumptions for my model to properly identify the effects you are trying to measure:*
I assumed that light radiance of night lights in SF would be correlated with the homeless youth's, 'safe' and 'unsafe', experiences. I also assumed that population density and medium income per household in SF neighborhoods would be correlated as well but not as much.

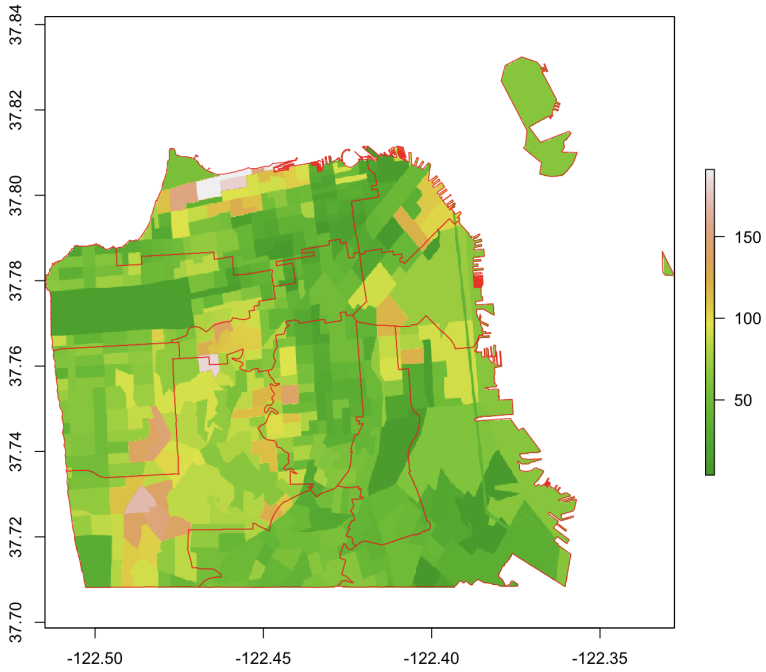
7) Discussion of the factors that could still hinder identification, e.g. if I used country-fixed effects, these would not account for unobservables that vary over time for a country.

The factors that inhibit my analysis from making any assumptions about the experiences reported by youth experiencing homelessness stems from biased collection of data and confounding variables. In terms of bias, my data is skewed because those that collected the data about homeless youth is not over a long duration of time. Other factors, like affluence of neighborhoods in SF also play an important role in the role that light and population density come into play for a youth's recorded experiences. One can see from the graph below that 'unsafe' experiences are recorded more frequently in low medium income household neighborhoods. Surprisingly, the youth predominantly felt safer in low income neighborhoods as well which can be explained partially by people wanting to keep their property values high so they may call the police or something along those lines to have them removed from their property proximity. Thereof making the youth feel unwelcomed, and unsafe. There are confounding factors like age, time of day, and the myriad of others influencers that play into the homeless population experiences that are not included in this research study. If I had the time I would have accessed police data but I was unable to clean the dataset in time for usage. All of the analysis should also be in the same time period but is not.

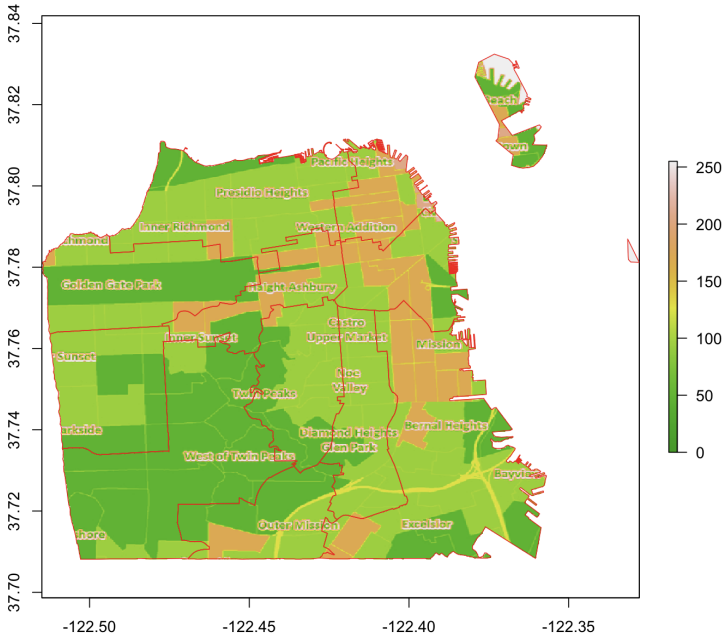
A. Satellite Image of Night Lights:

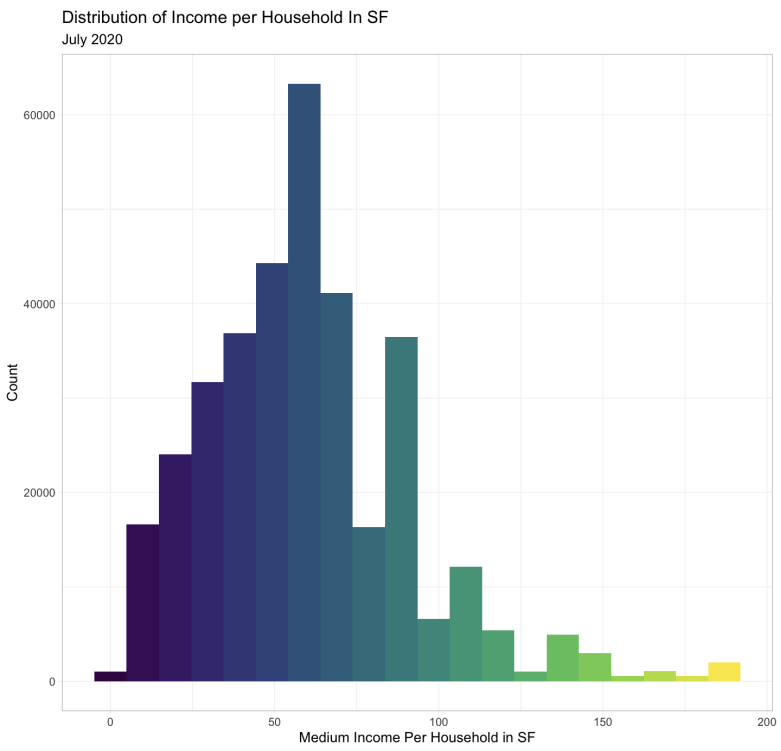


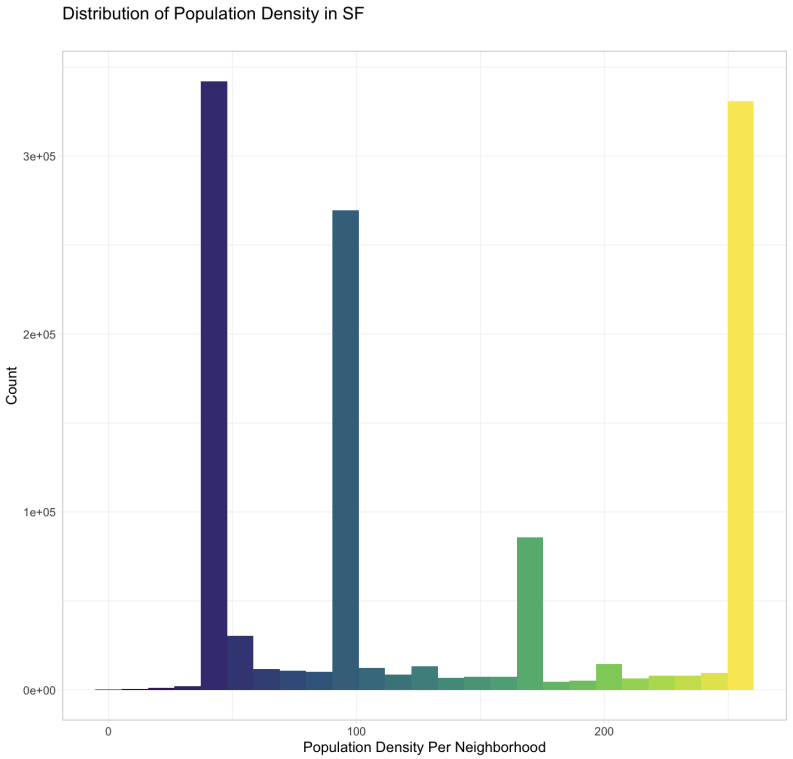
Medium Income Per Household

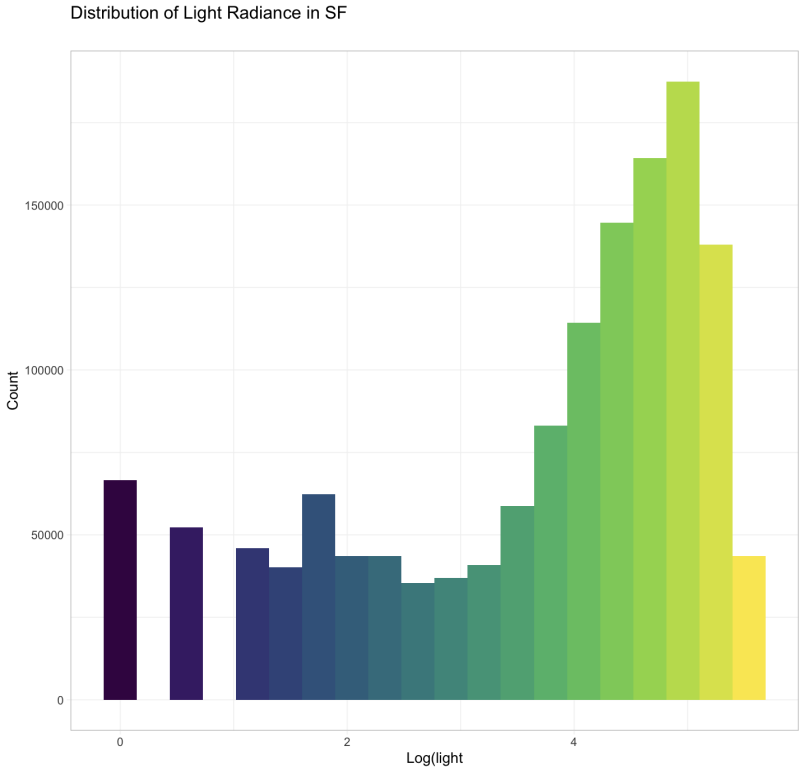


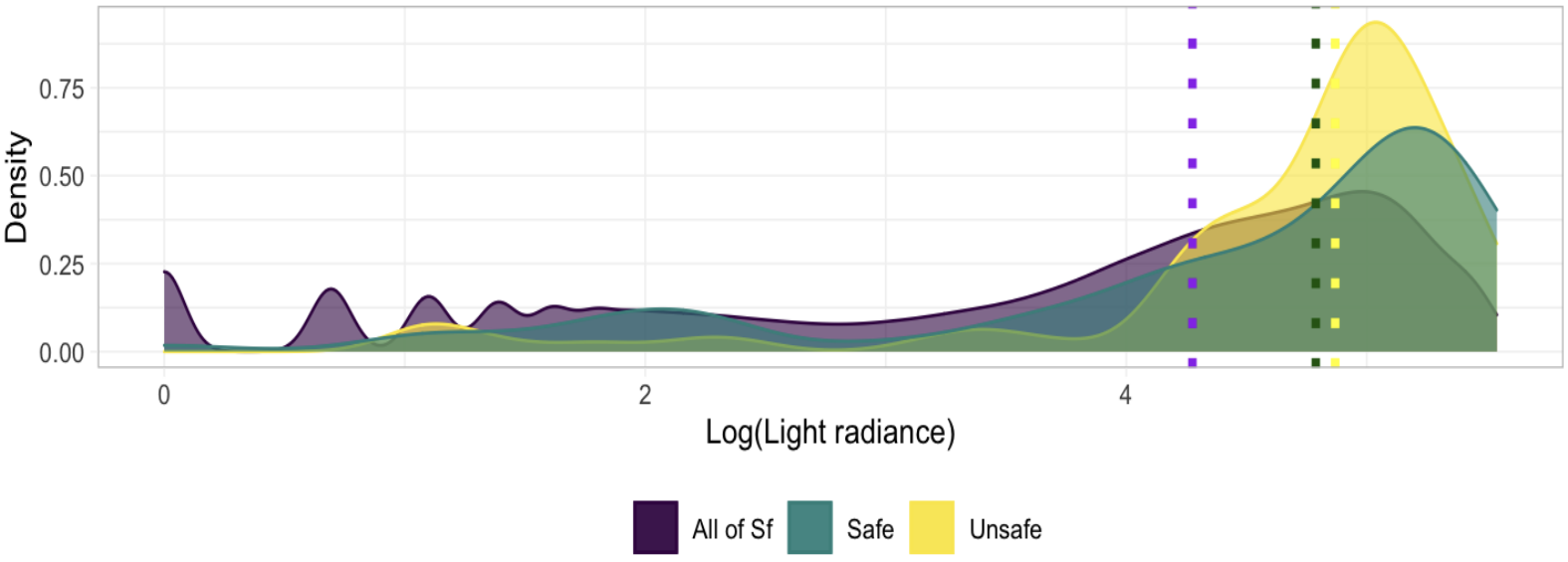
C. Population Density

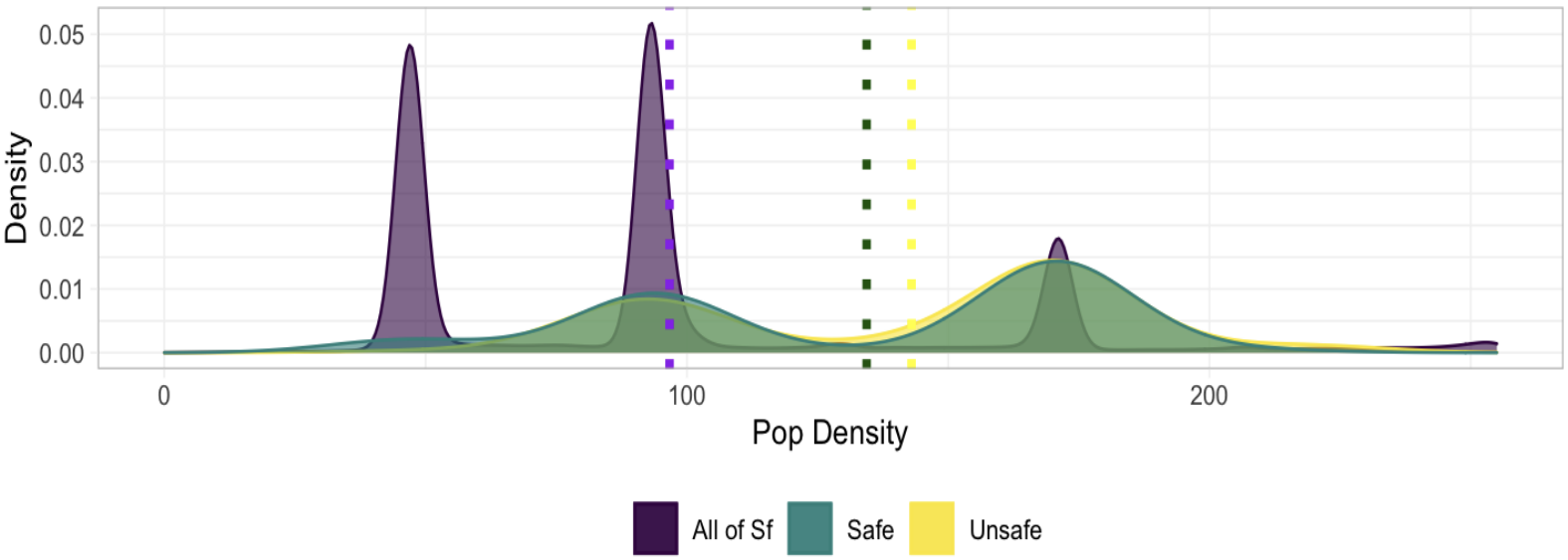


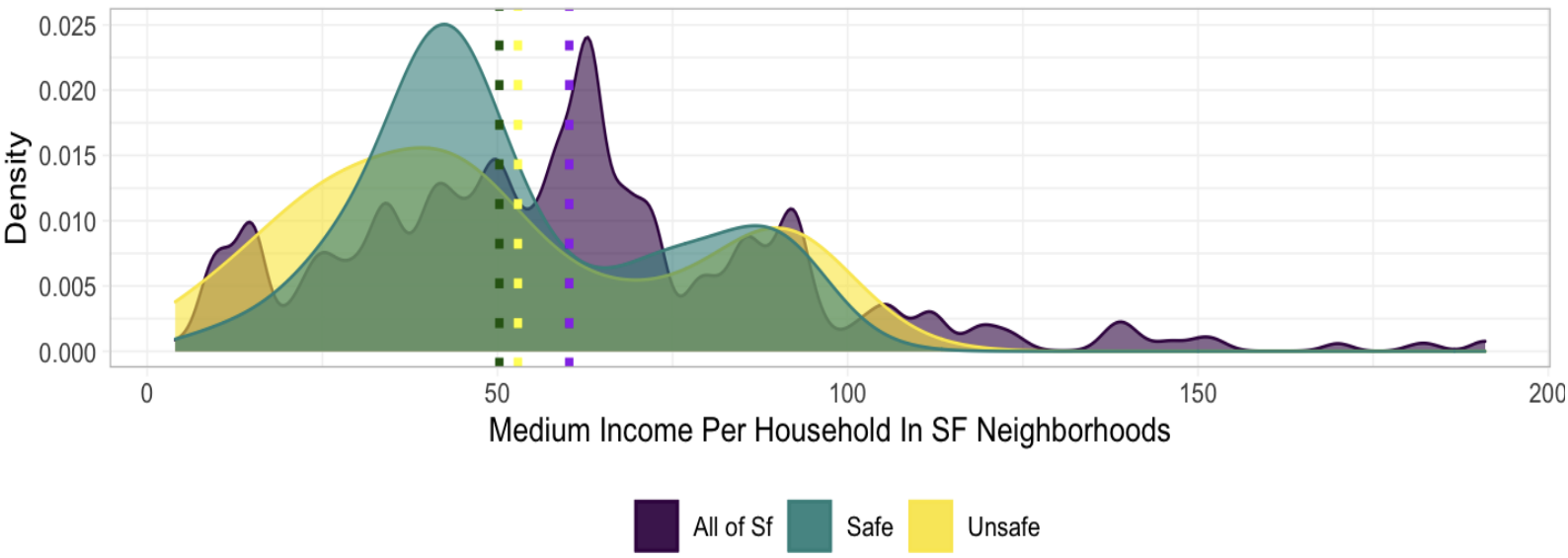

Distribution of Income per Household In SF
July 2020
Count
60000
40000
20000
0
0
50
100
150
200
Medium Income Per Household in SF


Distribution of Population Density in SF
Count
3e+05
2e+05
1e+05
0e+00
0
100
200
Population Density Per Neighborhood


Distribution of Light Radiance in SF
Count
150000
100000
50000
0
0
2
4
Log(light


Density
0.75
0.50
0.25
0.00
0
2
4
Log(Light radiance)
All of Sf
Safe
Unsafe


Density
0.05
0.04
0.03
0.02
0.01
0.00
0
100
200
Pop Density
All of Sf
Safe
Unsafe


Density
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Medium Income Per Household In SF Neighborhoods
All of Sf
Safe
Unsafe