EDA Report

Team SNAP

I. Introduction

"The right to adequate housing is a human right recognized in international human rights law as part of the right to an adequate standard of living. One of the first references to it is in article 25 (1) of the Universal Declaration of Human Rights." (UN)

Homelessness not only affects individuals in society, forcing those who face homelessness to be subjected to constant economic pressure but also has negative repercussions for a community. Homelessness can cripple a person's chance of getting back into society and hampers the overall economic growth of a community. Thus, especially with the current situation of a global pandemic which foreshadows another economic recession, we want to investigate the housing situation in 2008 when a severe economic crisis hit globally. The population of our study is all renters in California. Our study could be of interest to human rights activists, policymakers, or nonprofit organizations dealing with the ongoing housing crisis.

The great recession of 2008 undoubtedly has had a lasting impact on the economy. The housing crisis devastated the mortgage market in the US. By the end of 2009, over 25 percent of all mortgages in the US were underwater (Bowles et al., 2017: 471). Millions of people lost vast amounts of wealth, and an equal number of them were pushed into poverty.

Our group decided to pick the data from 2008, as the recession was at its height during this time; the financial institutions in the US were failing. In fact, on September 15, 2008, Lehman Brothers, one of the biggest financial institutions in the US, went bankrupt (Bowles et al., 2017: 473). Consequently, the unemployment rate went up to double digits, and people were struggling to afford housing, daily essentials, and so on.

As California is one of the states that have a wide pool of diversity with alarming housing-related issues, we will look into different districts of California with respect to racial

percentage and eviction rate. With our knowledge in systematic biases, we predict that there will be some differences in eviction rates with regard to different categories of the white population proportion, specifically the higher the white population, the lower the eviction rate as white people normally come from a more well-off background and thus should have fewer problems with rent and housing.

Currently, amid the Covid-19 pandemic, we can see a lot of parallels between the economy then and now with millions of people also facing hard economic times. Since we are starting to find similarities between then and now, we would like to understand how the poverty rate affected eviction rates in 2008. We want to study the effects of the poverty rate on the eviction rate by counties, in order to glean information about possible housing issues today. In particular, we are interested in investigating whether an increase in poverty results in a noticeable increase in evictions.

Inspired by the recent movements in unraveling and understanding how systematic biases work, and since eviction is one of the pressing issues recently with the rise in population density, we believe that there might be some relationship between races and eviction issues. Hence, one of the aspects that we want to cover in our research and analysis is whether there is a difference in the eviction rate between counties that are more ethnically diverse and the counties that are more ethnically uniform, and thus, how that difference is influenced by different levels of the white population proportion.

Furthermore, to get a better understanding of the impact of diversity on the housing situation in the US. It was imperative for us to analyze the housing situation for renters in the US. Our particular interest was to see if issues, such as segregation and racial preference, still had a foothold in the American society. California, a state with high diversity, was the perfect data set to test our hypothesis.

Table 1.1. Description of univariate variables

Variable name	Type	Description
Rent burden level	Categorical, response	The proportion of rent burden rate, binned into 3 distinct categories: *low* (≤ 18.5%)*, medium* (18.5 - 35.8%)*, high* (≥ 35.8%)
Eviction rate	Quantitative, explanatory	The ratio of the total number of renter-occupied households in a county in the state of California that received an eviction judgment in which renters were ordered to leave, over the period of one year.
White Population Proportion	Categorical, explanatory	The proportion of white people to people of color, binned into 4 distinct categories: *low* (<25%)*, low medium* (25 - 50%)*, high medium* (50 - 75%)*, and *high* (>75%).
Poverty rate	Quantitative, explanatory	The percent of the population with income in the past 12 months below the poverty level.

There are a few potential problems associated with these variables, mainly emanating from the fact that these variables involve some arbitrary choices. Firstly, we decided to base the eviction rate on counties, although this could also have been done on the basis of a different demographic division, such as a district. Secondly, we chose to study California as we hypothesized that it is a state with high diversity, although this is not necessarily representative of the U.S. as a whole. That prevents us from drawing any general conclusions about the eviction rate on a national level. Thirdly, "white population proportion" is binned into four separate categories of equal size, which keeps the investigation practical and simple for our scope, but it may not be the best way to measure diversity. Finally, the "rent burden level" has

been categorized using the quartile levels for convenience, thus, as an example, a rent burden of 15 percent, which falls under low "rent burden", may not be actually considered as low rent burden figures in reality.

II. Analysis of White Population Proportion

By Figure 2.1 and the data in Table 2.2, we can easily observe that the white population proportion levels are uniformly distributed among the counties, meaning there is roughly the same number of counties that have low, medium, and high populations of white people, with around 22 counties for each category. Furthermore, it is important to note that we have more than 5 observations for each category, which is necessary for further analysis.

On the other hand, from Table 2.2, we can also see that while the counts of counties are roughly the same, there are still more counties with a higher white population proportion. Specifically, there are 26.73 percent of counties that have a white population proportion from 50 percent to 75 percent and 29.92 percent that have a white population proportion bigger than 75 percent. This is expected because while California is diverse, which is shown in the fact that the levels of proportion look quite uniform, the U.S is still predominantly white.
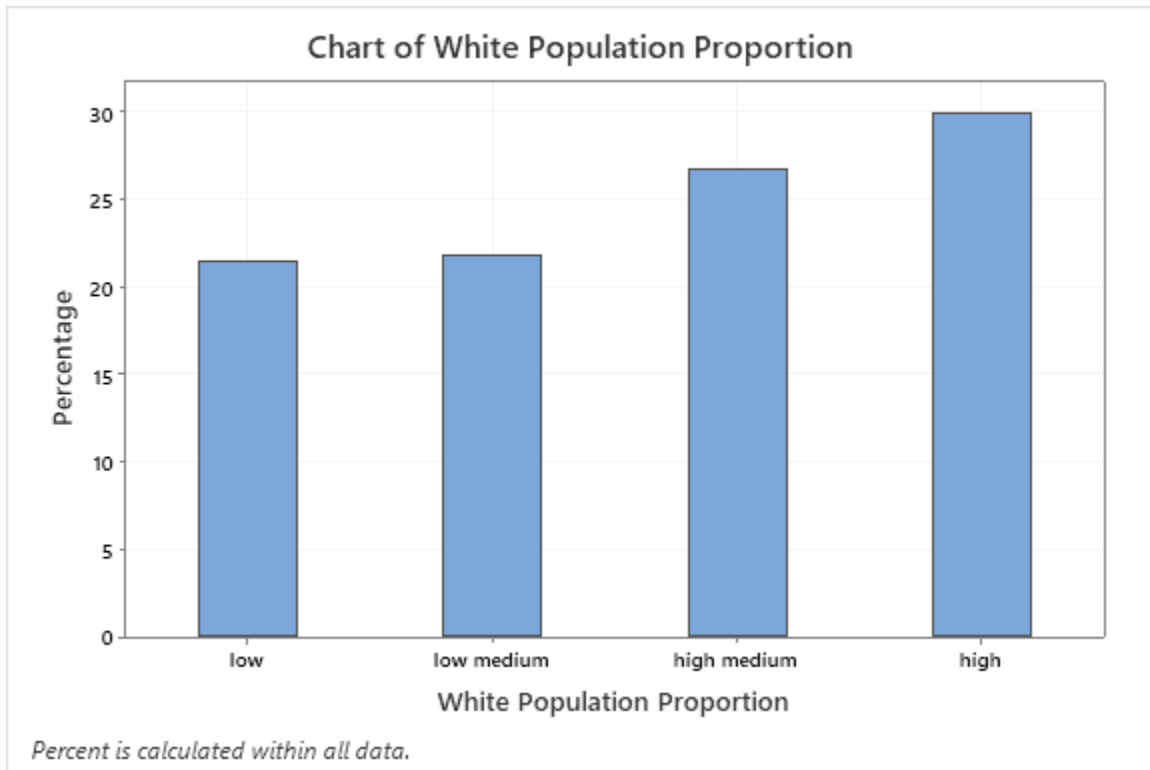

Figure 2.1: Proportion of white population of each level

Table 2.2: Count and percentage of each levels of white population

White Population Proportion	Count	Percent
Low	189	21.5
Low Medium	192	21.84
High Medium	235	26.73
High	263	29.92
Total (N) =	879	

III. Analysis of Poverty Rate

In 2008, the median poverty rate in California of the dataset was 7.89; 50 percent of the counties in California had a poverty rate of less than 7.89 percent, while 50 percent had a poverty rate higher than the median. Additionally, the average poverty rate (mean) in California, 10.69 percent, was relatively lower than the average, 13.20 percent, in the US, back then. Thus, it seems California fared better than most states, overall. However, the distribution of the poverty rate is heavily skewed to the right with a lot of outliers (about 4 percent of all poverty rates are outliers), so it appears a significant number of counties were struggling with poverty. For instance, Raisin city had the highest poverty rate in California, with a poverty rate of 79.59 percent. On the other hand, 56 counties, including Alpine Village, seemed to have no poverty issues, with a poverty rate of 0.00 percent. Similarly, the bottom 25 percent of the states had a poverty rate ranging from 0.00-3.74 percent, while the upper 25 percent had a poverty-rate ranging from 14.6 to 79.6 percent. Hence, the variation in the data is pretty high, and this might affect our analysis of the relationship between the poverty rate and the eviction rate in unpredictable ways. So, we carry out the analysis with caution.
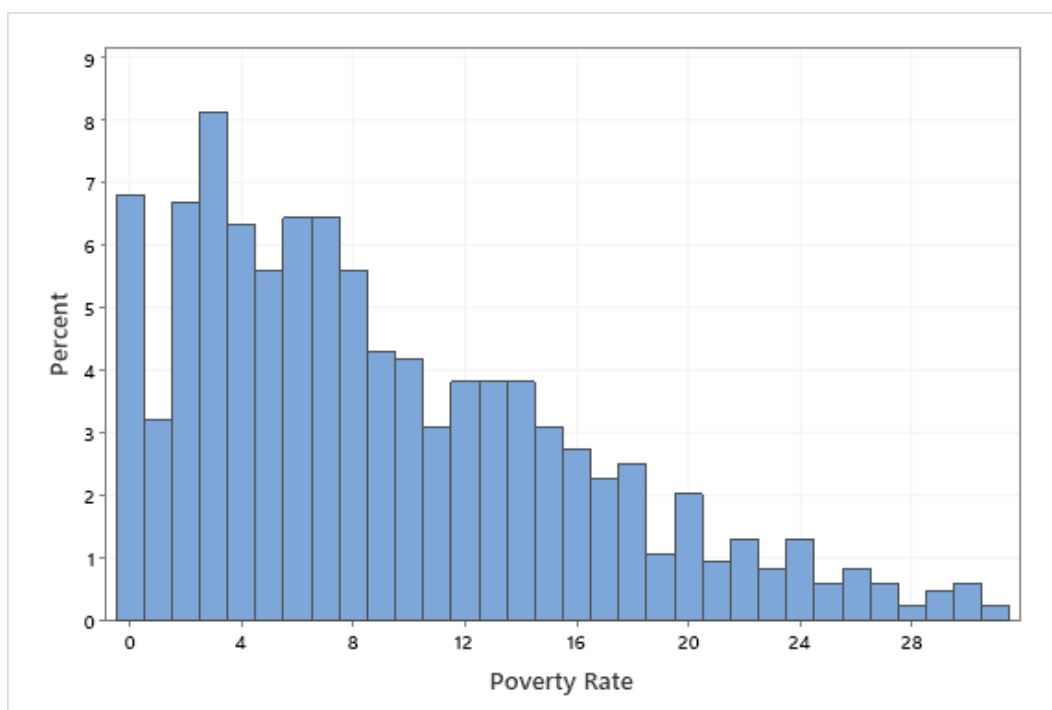

Figure 3.1: Histogram of Poverty Rate in California by county

Table 3.2: 5-number summary of Poverty Rate.

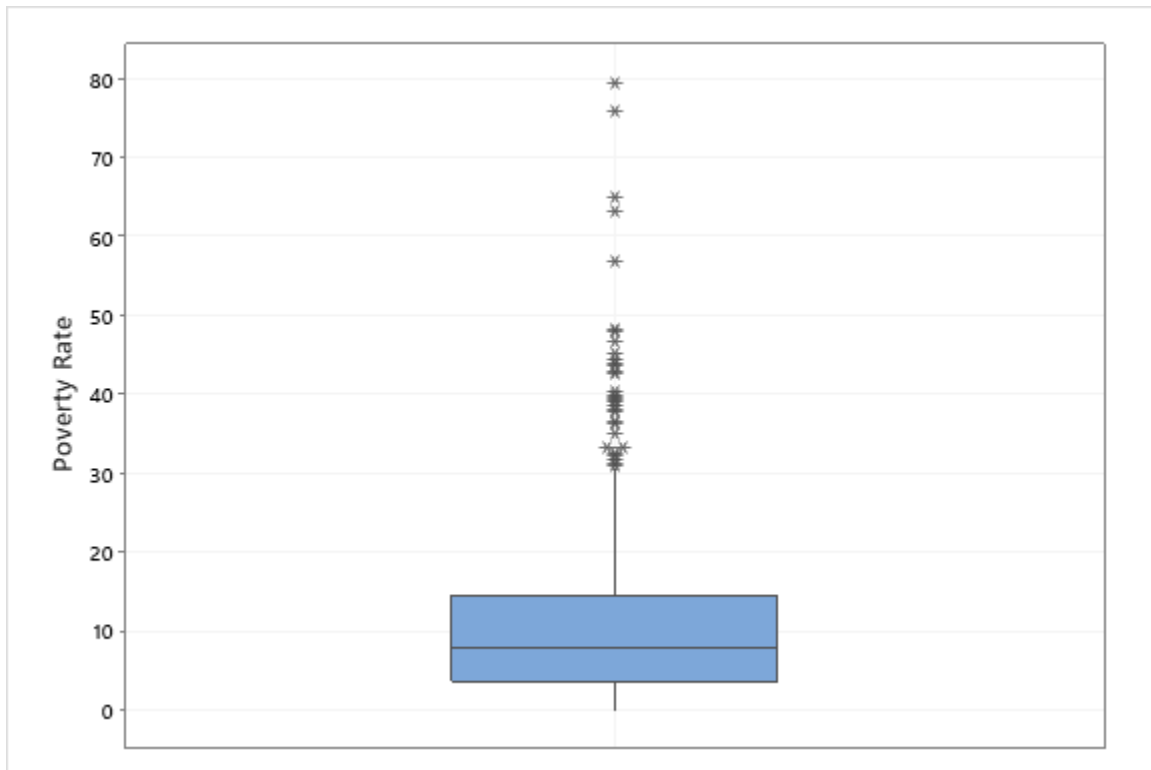| Minimum	First Quartile	Median	Third Quartile	Maximum
0.00	3.74	7.89	14.6	79.6

Figure 3.3: Boxplot of Poverty Rate in California counties.

IV. Analysis of Eviction Rate

The histogram of the Eviction Rate in Figure 4.1 shows us that the data is highly right-skewed, so the median of 1.72 is arguably a better indicator of central tendency than the mean of 2.0687 in this case, as the value of the mean becomes "skewed" higher than expected. This may be a consequence of the fact that there are a few counties with a disproportionately high relative eviction rate, for example, one county has an eviction rate of 12.5. Hence, the median value of 1.72 tells us that the number of evictions in counties in the state of California tends to be just below 2 percent measured against the total number of households. Because of the skewed data set, it is not appropriate to use the standard deviation, so instead, we use the range to determine what this data means more broadly. We can see that 75 percent of the counties are within a range of 0.82 - 2.87, which tells us that the percentage of evictions in the large majority of California counties measure more or less around 1 -3 percent compared to the total number of households.

The percent of outliers in the data set is 4.1 percent, which is a fairly small number of counties relative to the state as a whole and should therefore not have a very significant effect on the results calculated above. However, since these are all outliers on the higher end it does indicate that a small number of counties have a much higher eviction rate than the majority of the counties. This might potentially affect our analysis of the relationship between the poverty rate and the eviction rate since our data is skewed. However, for both Poverty Rate and Eviction Rate variables, we have about 4 percent of the data as outliers so for both variables the percentage of outliers seems to match up. In any case, we proceed with this analysis with caution.
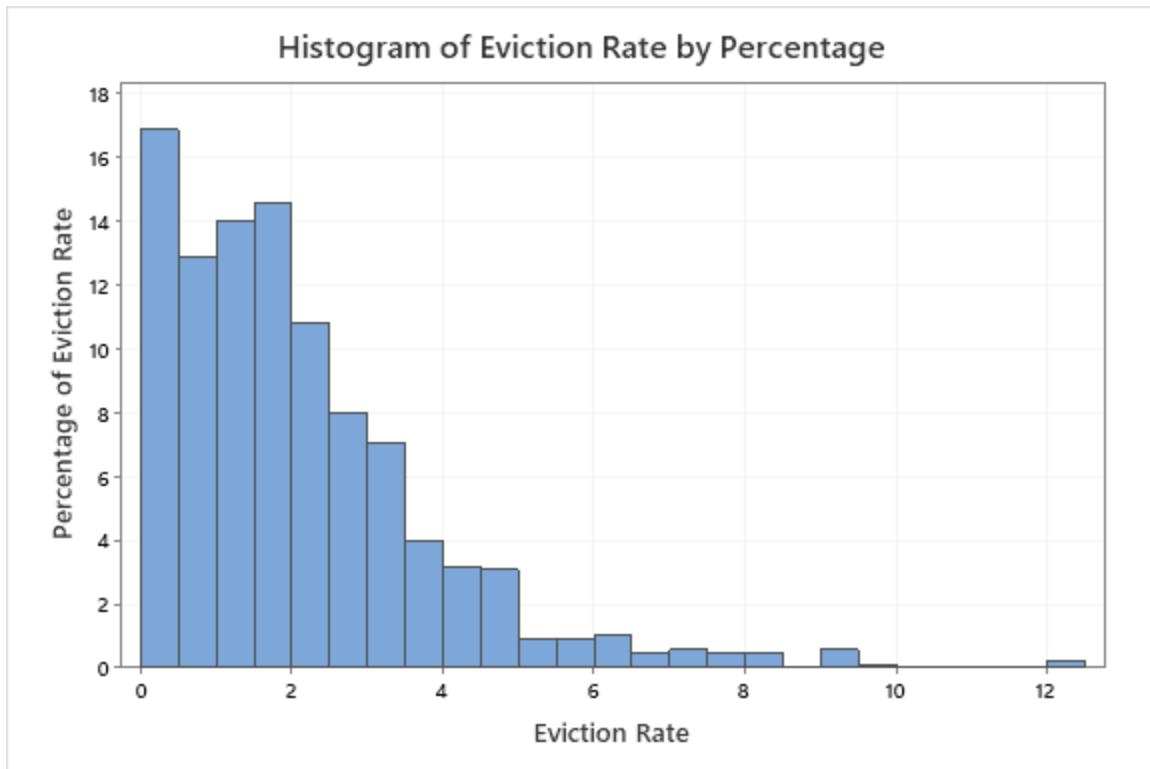
Figure 4.1: Histogram of Eviction Rate from different counties in California.

Table 4.2: Mean, STDEV and 5-number summary of Eviction Rate.

Variable	Mean	STDEV	Q1	Median	Q3	Min	Max
eviction-rate	2.0687	1.7742	0.82	1.72	2.87	0	12.5

Table 4.3: Frequency and Percentage of Outliers [1] and Non-outliers of Eviction Rate.

Outlier	Count	Percent
No	843	95.9
Yes	36	4.10
Total (N) =	879	

---

[1] Outliers represent the extreme values within our statistics. More on outliers and how to calculate them could be found here: https://www.statisticshowto.com/find-outliers/

V. Analysis of Rent Burden Level

In Figure 5.1 we see that the rent burden levels are distributed mostly normally, this is what we should expect since we binned our data based on the first and third quartiles. So, we should see about the same percent of high and low rent burden levels. Although this was an arbitrary choice and might not be the best way to divide up rent burdens, we thought this was reasonable since it was both easy in terms of data generation and what we would expect in the real world. We hypothesise rent burden to follow a normal distribution in the population since we think it's reasonable to see most people fall into the medium rent burden level and about an equal percentage of people in the high and low categories of rent burden. Furthermore, according to Table 5.2, there are more than 4 counts for each level of rent burden, which is important for further analysis.
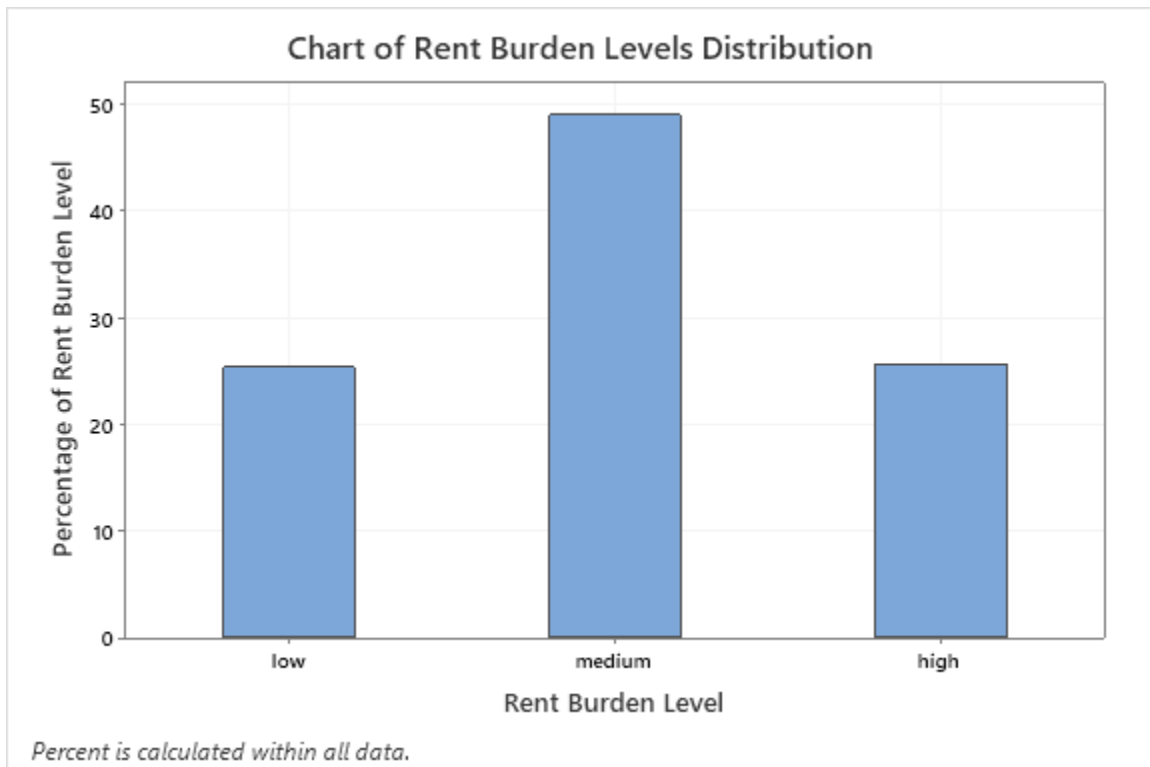

Figure 5.1: Distribution of rent burden levels

Table 5.2: Counts and percentages of the rent burden levels

Rent Burden Level	Count	Percent
low	223	25.37
medium	431	49.03
high	225	25.60
Total (N) =	879	

VI. Bivariate Analysis of the relationship between White Population Proportion and Eviction Rate

From Figure 6.1 and Table 6.2, we can see that there are some differences among the mean eviction rates according to different levels of white population proportion. Specifically, the eviction rate in the counties that have denser white populations have the lowest mean (approximately 1.45 percent of households being evicted) out of the four categories, whereas the counties that have white population proportion ranging from 25-50 percent have the highest eviction rate mean, with roughly 2.58 percent of households being evicted. This is closely followed by the counties that have a low white population proportion, with a mean of 2.487 percent of evicted households. For counties that have a high medium white population proportion, the mean eviction rate is closer to that of counties with a high white population proportion (around 2 percent eviction rate). Thus, we can see some distinctive gaps here between the lower and the higher white population counties.

At the same time, the standard deviations of eviction rates for each category are roughly similar, with an exception of the low white population proportion group of counties. While other groups' eviction rate standard deviation ranges from 1.35 (for predominantly white counties) to 1.75 (for lower medium white population proportion counties), the group for counties with low white population proportion has an eviction rate with a standard deviation of 2.171. Hence, while this category of low white population counties have a lower eviction rate mean than the group with the low medium white population, its standard deviation might reveal that there is a variation in which we have to be cautious of the extreme cases, in particular, there might be some counties with low white population and high eviction rates.
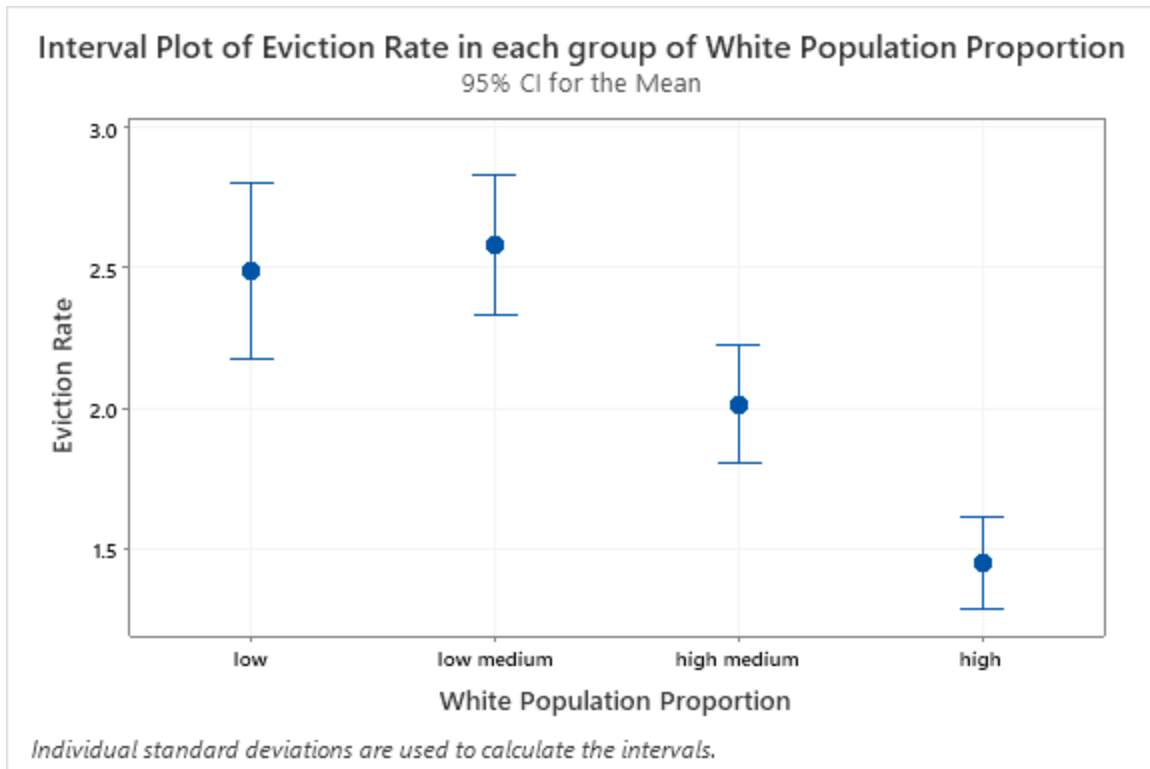
Figure 6.1: Interval Plot of Eviction Rate with regard to White Population Proportion.

Table 6.2: Means and standard deviations of Eviction Rate for each level of White Population Proportion.

Variable	Group (of White Population Proportion)	Mean	STDEV
Eviction Rate	Low	2.487	2.171
	Low Medium	2.581	1.753
	High Medium	2.011	1.630
	High	1.4458	1.3579

Indeed, according to Figure 6.3, when we will take into account the outliers, we will see some cases of extreme eviction rate. In this case, there are no low outliers but only high outliers,

which represent a distinctively high eviction rate. As a result, we observe that while every group has some number of outliers, the group with low white population proportion yields the most extreme outliers, with counties that have eviction rates of more than 10 percent. On the contrary, the group with high white population proportion has the least extreme outliers, among which the most extreme case is a county with eviction rate of around 7 percent.
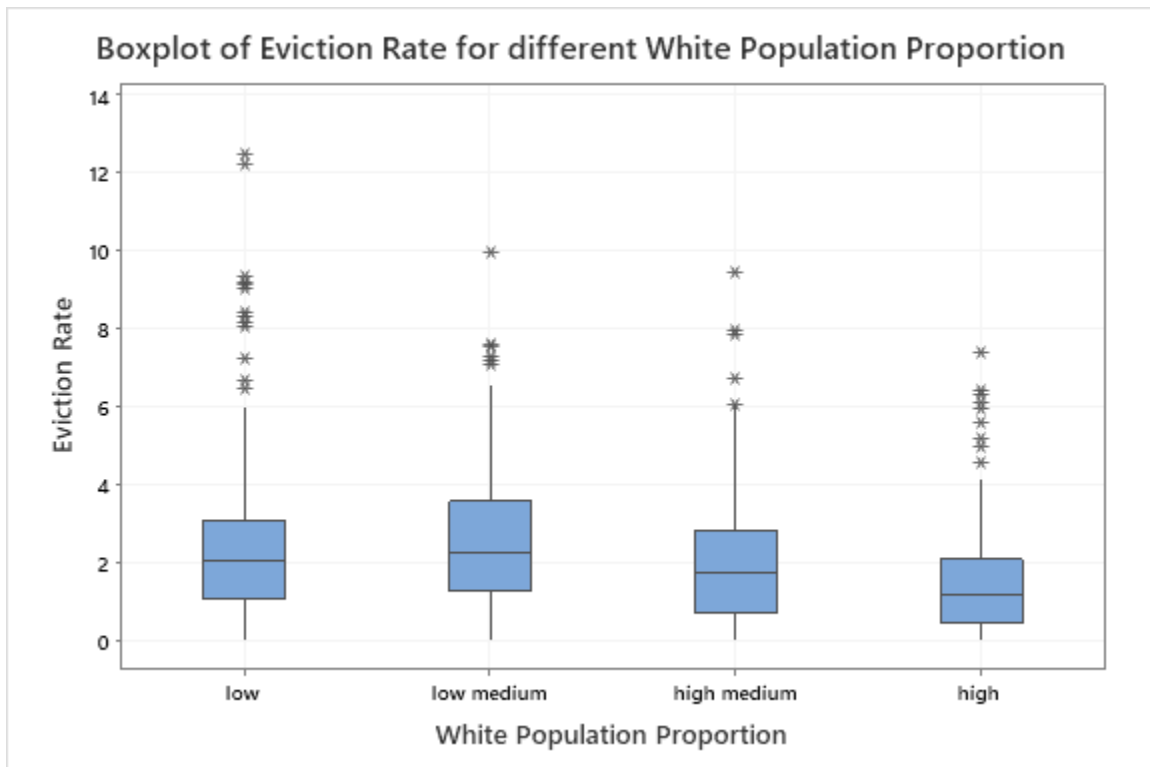

Figure 6.3: Boxplot of Eviction Rate grouped by White Population Proportion Levels.

VII. Bivariate Analysis of the relationship between Poverty Rate and Eviction Rate

Based on Figure 7.1, we do not see any clear linear relationship between the Poverty Rate and Eviction Rate. The correlation coefficient between poverty rate and eviction rate is only 0.103. So, this confirms that there is an extremely weak and positive association between eviction rate and poverty. Furthermore, the coefficient of determinant (R^2) is only 1.05 percent. Hence, the graph isn't a good predictor of the Eviction Rate as poverty level goes up.
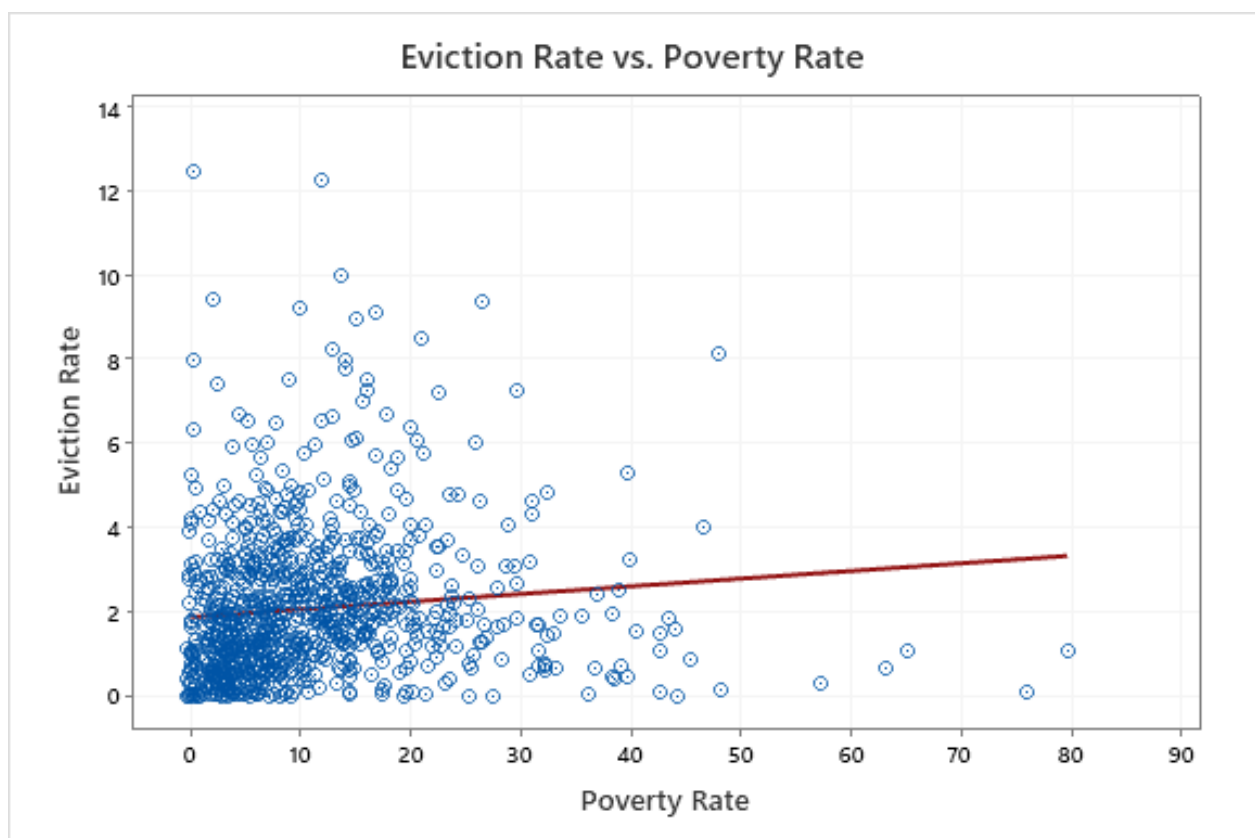

Figure 7.1: Scatterplot of Eviction Rate with Poverty Rate with least-square fit linear regression.

Moreover, according to Figure 7.2, the residual plot for Eviction Rate appears to be clustered, without sufficient randomness to validate the use of the resulting regression line.

Moving on, the outliers potentially introduce inaccuracy to the scatter plot. We hypothesize outliers in the Poverty Rate data were caused by other underlying conditions, thereby introducing lurking variables. For instance, in 2008, a wildfire might have destroyed a county engendering widespread Poverty Rate, but resulting in almost no evictions. Similarly, a county might have had strict rent laws, so even if the poverty rate was low, the eviction rate might have been high. Thus, to get a better picture of the relationship between these variables, we decided to remove the outliers.
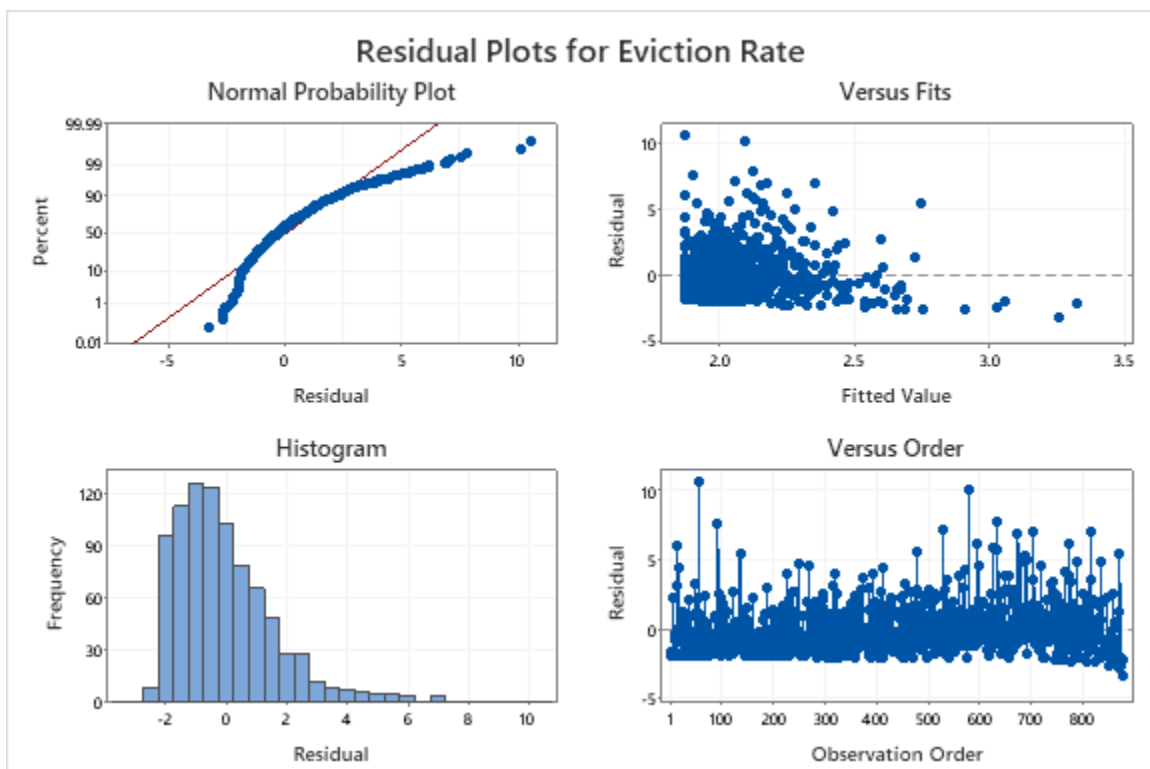

Figure 7.2: Residual Plots for Eviction Rate.

After removing the outliers, our correlation coefficient improves slightly. The correlation coefficient becomes 0.229. So there is still an extremely weak, positive correlation between poverty rate and eviction rate; Moreover, based on the coefficient of determination, R^2 = 5.2%, we can say that this is not a good model to predict the eviction rate with a change in Poverty rate. However, the residual plot, after the removal of outliers, has improved showing us that this model is still better than the model that includes the outliers.

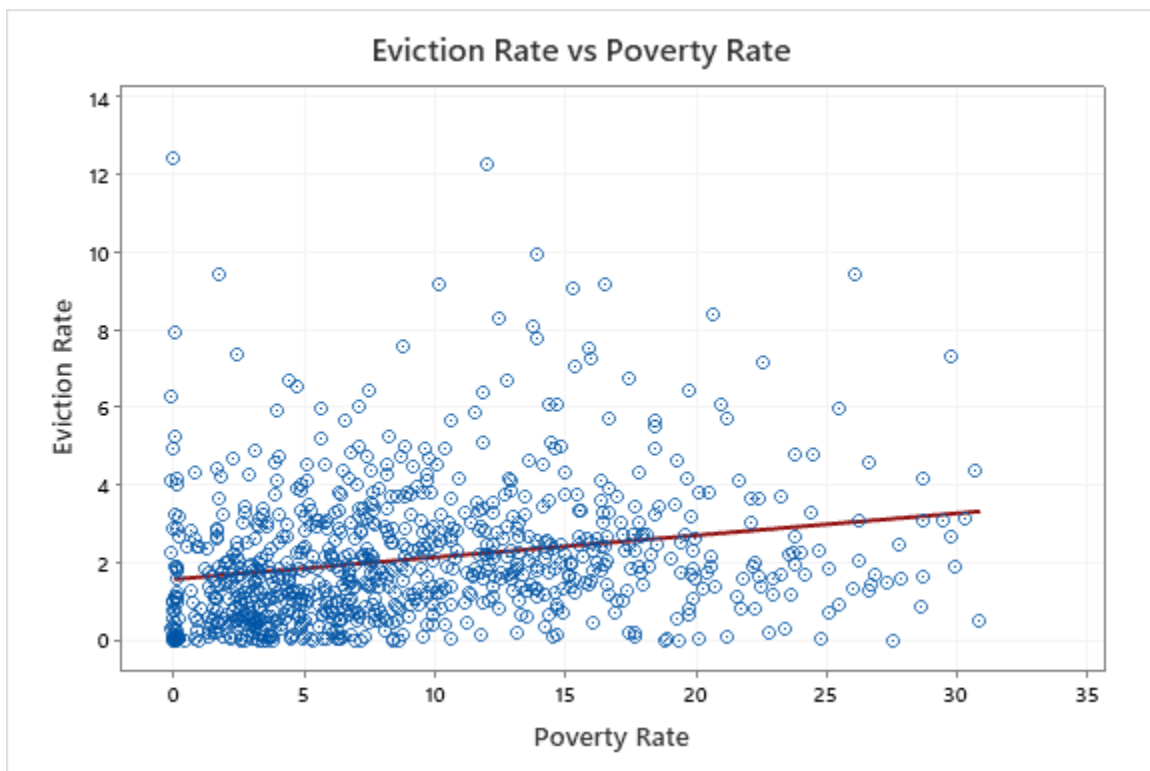After removing the outliers we get the following scatter plot:

Figure 7.3: Scatter plot of Eviction Rate and Poverty Rate with the least-square linear regression line without outliers.

Looking at Figure 7.4, The residual plot tells us that the data is spread mostly evenly around 0 with standard deviation of about 2 with minimal display of heteroscedasticity. At the end, even though this seems to be a better model, it still isn't a good model for predicting eviction rate.
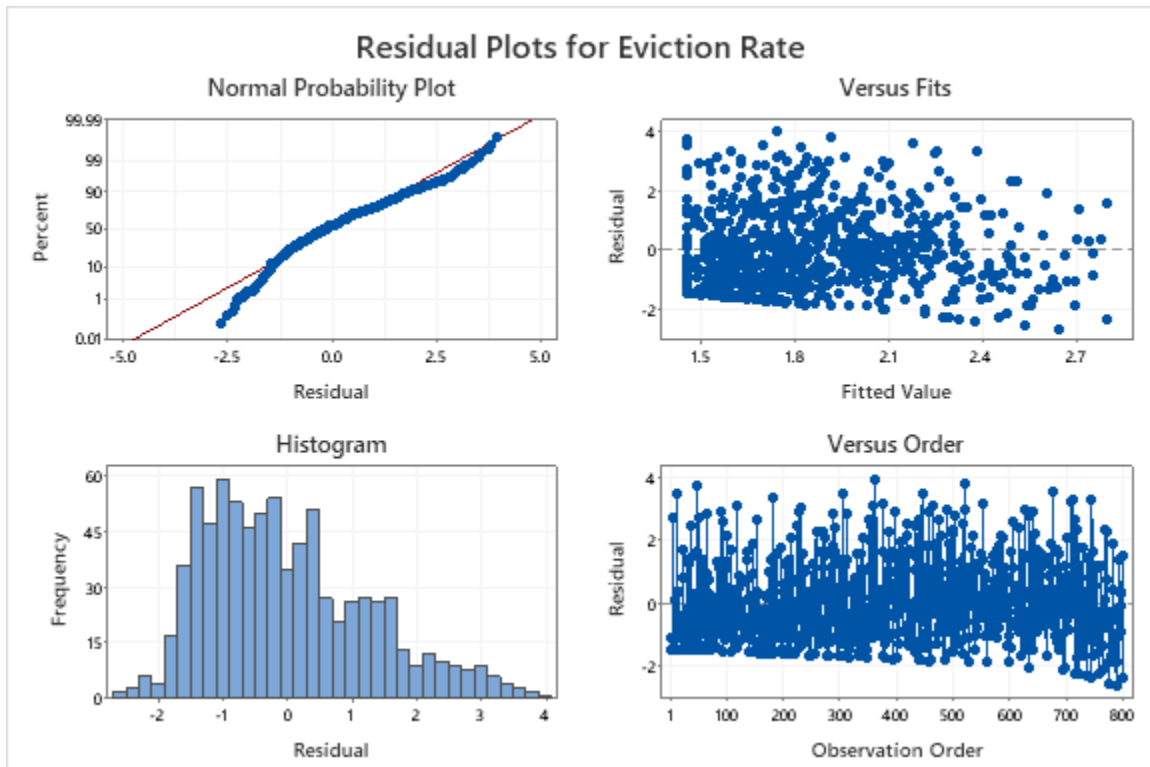

Figure 7.4: Residual Plots for Eviction Rate with LSLR without outliers.

VIII. Bivariate Analysis of Rent Burden Level and White Population Proportion

Normally we would expect all the distributions for each of the white population proportions to look the same. However, in Figure 8.1 we see that the distributions for the rent burden level are similar in low, low medium, and high medium levels of white population proportion, while the distribution for high white population level is very different from the other three. In fact, we see that in the other three levels there is a clear peak in the medium rent burden level while the rent burden levels are mostly the same in the final level. Additionally, by inspecting Table 8.2 and Figure 8.1, we see that as the white population proportion increases (from low to high) the percent of low rent burden relative to the other two categories goes up while high rent burden level goes down. So, we can guess that in counties that have a high white population proportion, there is a bigger portion of people in the low and medium rent burden level which could mean that there is a bigger disparity between rent burden levels in counties with high white population proportion.
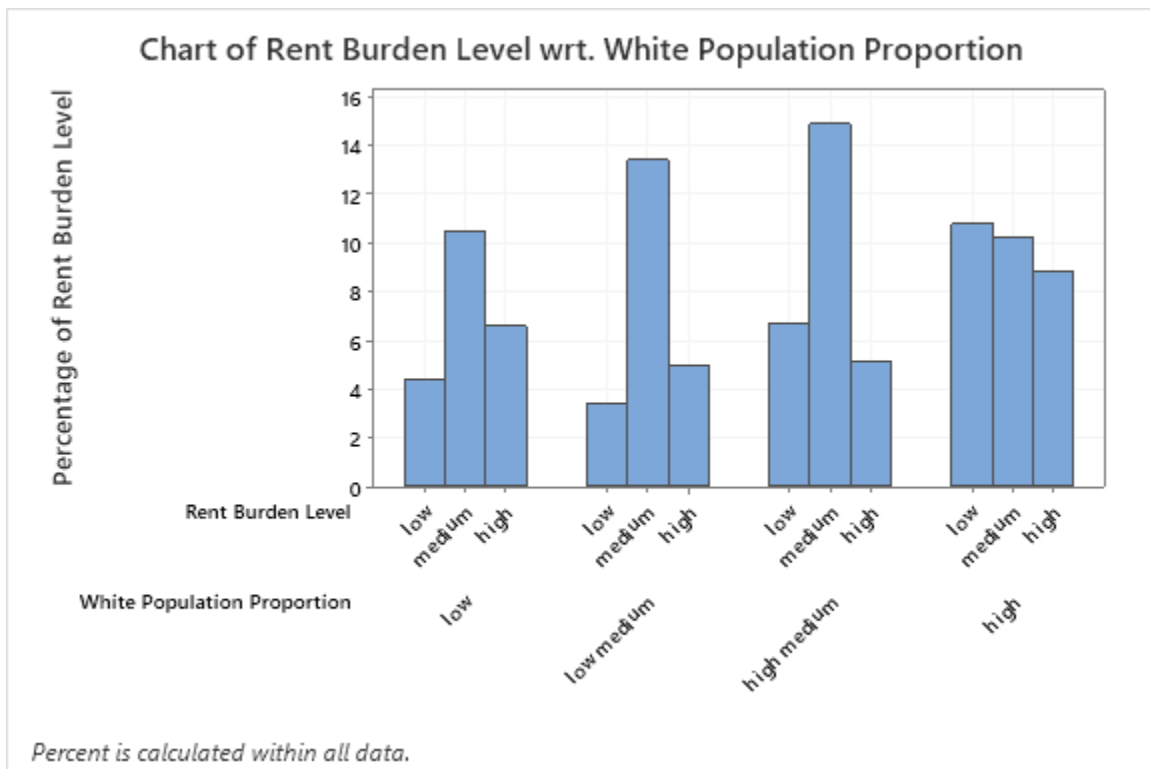

Figure 8.1: Clustered Chart of Rent Burden Level with respect to White Population Proportion

Table 8.2: Conditional distribution of the response variable for each level of the categorical variable

	low	Low medium	High medium	high	all
low	4.44	3.41	6.71	10.81	25.37
medium	10.47	13.42	14.90	10.24	49.03
high	6.60	5.01	5.12	8.87	25.60
all	21.50	21.84	26.73	29.92	100.00

IX. Conclusion

Our first univariate analysis, on the proportion of white people in California counties, showed that the proportion of white was fairly equally distributed across the four bins. That is, roughly a fourth of the counties was populated by 0-25 percent white, another fourth by 25-50 percent, a third 50-75 percent and 75-100 percent in the last fourth, as shown in Figure 2.1 (although this was the group with the highest number of counties; 29.92 percent). There were no outliers in this analysis, which strengthens these findings.

The second univariate analysis we conducted, poverty rate in California counties was less uniformly distributed and more spread. In this case, 4 percent of the total number of counties turned out to be high outliers, that is, counties with high numbers of people with income below the poverty level. This is also reflected in the distribution in Figure 3.1, which turned out to be highly right-skewed. Whereas the median poverty rate was roughly 8 percent, the county with the highest poverty rate was almost 80 percent (Table 3.3).

Similarly, the eviction rate in California counties was heavily right-skewed, with a median of 1.72 percent of households being evicted in 2008. We saw that 75 percent of the counties had an eviction rate from 0.82 to 2.87. There also existed 4.1 percent of outliers in the dataset, all of which belonged to the higher outlier groups. Thus, even though the median eviction rate remained under 2 percent, there were still districts with really high eviction rates, specifically the highest eviction rate was 12.5 percent.

With our analysis of each variable and their relationship, we can observe some features about the eviction rate with regard to the white population proportion groups. In general, the higher white population district groups obtain lower eviction rates, whilst the lower white population district groups bear higher eviction rates.

On the other hand, we cannot accurately predict eviction rates from the poverty rate using the data retrieved from counties in California. We believe that there are lurking variables that may influence the relationship between eviction rate and poverty rate. Even after removing the outliers, that we deemed as influential points, to re-examine the relationship between these

two variables, we were not able to see a significant improvement. The correlation went up by a mere 0.126.

Lastly, we saw that in counties with high white population proportions, there was a higher percentage of low rent burden levels than high rent burden level. We also saw that the rent burden tends to be lower for counties with high white population proportion. If we assume that there should be no difference between the distributions of rent burden levels for each of the white population proportion levels then we would have expected to see a peak in the medium rent burden levels. However, that is not what we saw; and so, we hypothesise that in high white population proportion counties, we might expect to see a bigger gap between the different economic classes which is the cause of not seeing a peak in the middle rent burden level and instead seeing about an equal percentage for the three levels.