1. In areas of low population density there seems to be a much higher rate of theft. In some areas of low population density, such as the southwest, there are low rates of theft, however in other areas of low population density such as the east there are high rates of theft.
2. In some areas there appears to be a positive relationship between median income (right) and car theft rates (left), but this relationship does not hold across the entire study region. For example, the southeastern-most census tract of San Francisco has relatively low median income but high car theft rates.

A picture containing graphical user interface

Description automatically generated

1. As shown by the following histogram, car theft rates are very skewed and therefore abnormally distributed.

Chart, histogram

Description automatically generated

1. The following histogram shows that car theft rates, after natural log transformation, is much more normally distributed. It is still skewed, however.

Chart, histogram

Description automatically generated

1. There appears to be a negative linear relationship between SD\_POPDENS and LOG\_RATE. As SD\_MEDINC increases, LOG\_RATE stays roughly constant (slope near 0), indicating little or no systematic relationship.Graphical user interface, chart, scatter chart

   Description automatically generated

A picture containing text, receipt

Description automatically generated

1. The R2 value of the OLS model is 0.075, meaning that this model does not fully explain the variance of the dependent variable.
2. The p-value for population density is 0.00 which is less than the significance level of 0.05, indicating that it is a significant predictor. P-value for median income is 0.16, which is greater than 0.05 and indicates that it is not a significant predictor of car theft rates. A change of 1 unit in population density corresponds to a -0.25 change in car theft rates as told by the regression coefficient of -0.26. This model has a Akaike Information Criterion (AIC) of 468.93 and a Log Likelihood of -2312.47. The large AIC and the small Log Likelihood both suggest an ill-fitting model.
3. The Jarque-Bera test returns a probability of 0.00, less than 0.05, indicating that the error term in non-normally distributed.
4. The Breusch-Pagan test has a probability of 0.005 and the Koenker-Basset test has a probability of 0.43. Since the p-value of the Breusch-Pagan test is less than 0.05, there is indication of heteroscedasticity.
5. Positive residuals indicate datapoints with values higher than the model’s line, and negative residuals indicate datapoints with values falling below the line. There appear to be more negative residuals in the west and more positive residuals in the east. Interestingly, the map of OLS residuals appears very similar to the map of car theft rate.
6. The Moran’s I-value of the residuals is 0.2505. With 999 permutations, the pseudo p-value is 0.001, which is less than the significance level 0.05 and therefore statistically significant, meaning that we can reject the null hypothesis that residuals are randomly distributed across the study region. The Moran’s I value indicates there is little autocorrelation and therefore little use for a spatial regression model.

Chart

Description automatically generated

A picture containing text, receipt

Description automatically generated



A close up of a newspaper

Description automatically generated



|  |  |  |  |
| --- | --- | --- | --- |
|  | **OLS** | **Spatial Lag** | **Spatial Error** |
| **R2** | 0.075 | 0.14 | 0.19 |
| **AIC** | 468.93 | 459.96 | 450.12 |
| **Log Likelihood** | -231.47 | -225.98 | -222.06 |
| **Diagnostic for Spatial Dependence** | n/a | 10.97 | 18.81 |

The Spatial Error Model has the highest R2, but it also has the lowest AIC, smallest Log Likelihood, and highest diagnostic for spatial dependence. This model therefore fits the data best.