Unaffordable Housing:

An Analysis of Rental Prices in Denver

By: Jessica Murphy

April 30th, 2019

Introduction

Affordable housing describes the level of housing access that residents of a specific area possess. It is usually defined in terms of monthly income compared to monthly rent or mortgage. In Denver, affordable housing is an evolving and ongoing issue as the population continues to grow rapidly. The demand for housing far outweighs the supply, significantly increasing rental prices.

This study aims to determine which of the following factors are related to unaffordable (or increased) rental prices in 75 local Denver neighborhoods: age, race (white or Hispanic), poverty, vacant housing, renter occupied housing, home value, education (no college), family households, income, size (area), police stations, schools, health clinics, crimes, and foreclosures. The DIA neighborhood was removed from this study since the airport would likely skew the results. The neighborhoods of Sun Valley and Kennedy were also removed because they did not contain any home value information.

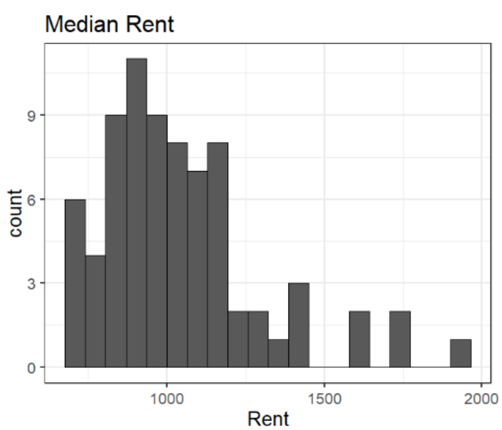
All of the data used in this analysis was obtained from observational studies. Most of the data are from the American Community Survey collected by the City and County of Denver. The data contain 5-year averages from 2010–2014 aggregated at the census tract level and summarized into neighborhoods. The crime data was also collected by the City and County of Denver for the Denver Police Department. The data includes criminal offenses for the previous five calendar years and the current year to date based on the National Incident Based Reporting System (NIBRS). The number of crimes per year, excluding traffic incidents, were then totaled and averaged together for each neighborhood. This produced 5-year averages from 2014-2018.

The neighborhood characteristics and foreclosure data were already prepped and available from the Auraria Library Data to Policy Resources website for the purpose of this project. The neighborhood characteristics data was obtained in the last couple of years and the foreclosure data is from 2003–2016. However, an average of the foreclosure data per neighborhood from 2010–2014 was used to match the survey data.

Methods/Results

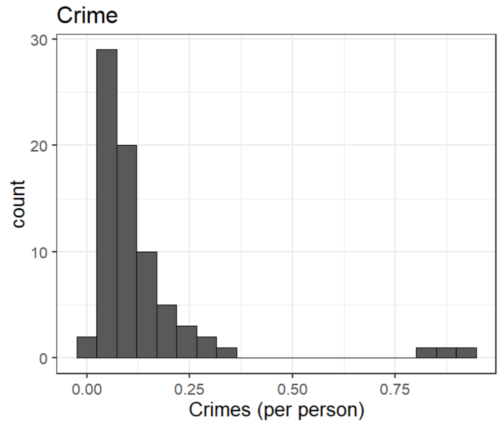
Data Summaries

Below are some initial exploratory graphical and numerical summaries of the data.



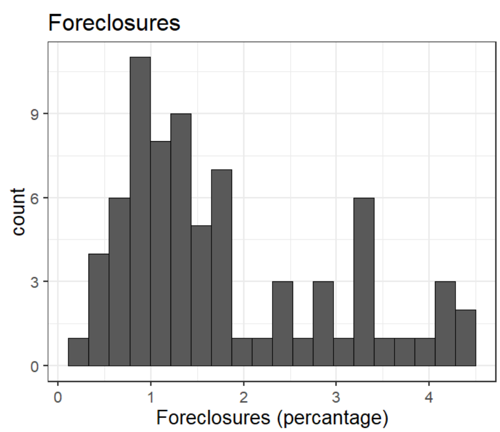
Min:	681.00
1st Qu:	871.50
Median:	976.00
Mean:	1,038.20
3rd Qu:	1,143.00
Max:	1,905.00

The histogram above shows that the variable *rent* is positively skewed because the long tail is in the positive direction. The largest outlier corresponds to the Wellshire neighborhood.



Min:	0.0239
1st Qu:	0.0584
Median:	0.0841
Mean:	0.1348
3rd Qu:	0.1309
Max:	0.9460

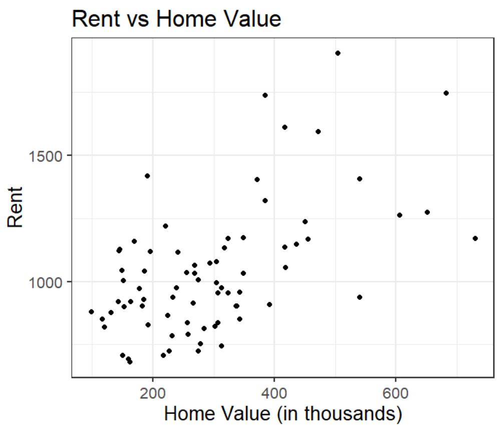
The histogram above shows that the variable *crimes* is very positively skewed because there are several extreme outliers in the positive direction. These outliers correspond to the neighborhoods of Auraria, CBD, and Civic Center.

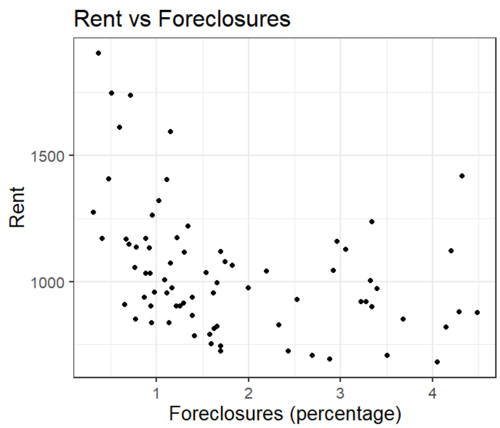


Min:	0.3089
1st Qu:	0.9392
Median:	1.3874
Mean:	1.8006
3rd Qu:	2.6068
Max:	4.4814

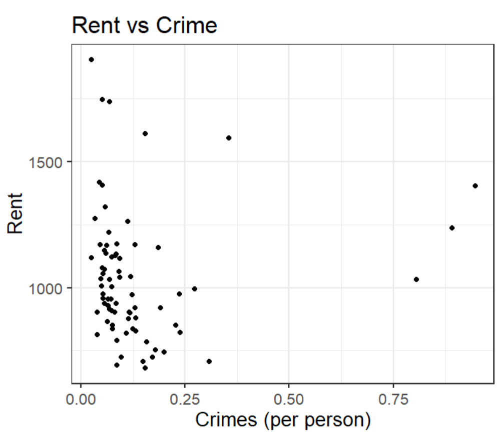
The histogram above shows that the variable *foreclosures* is positively skewed because the long tail is in the positive direction.

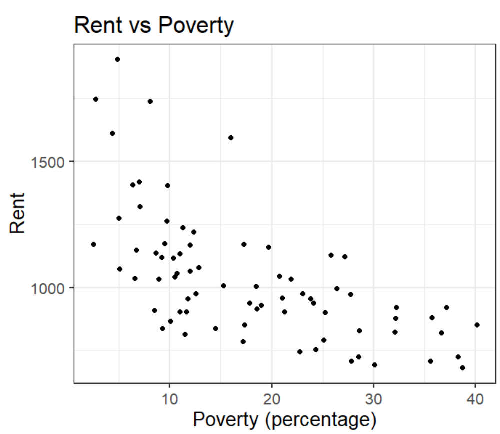
Below are some bivariate plots of the data.



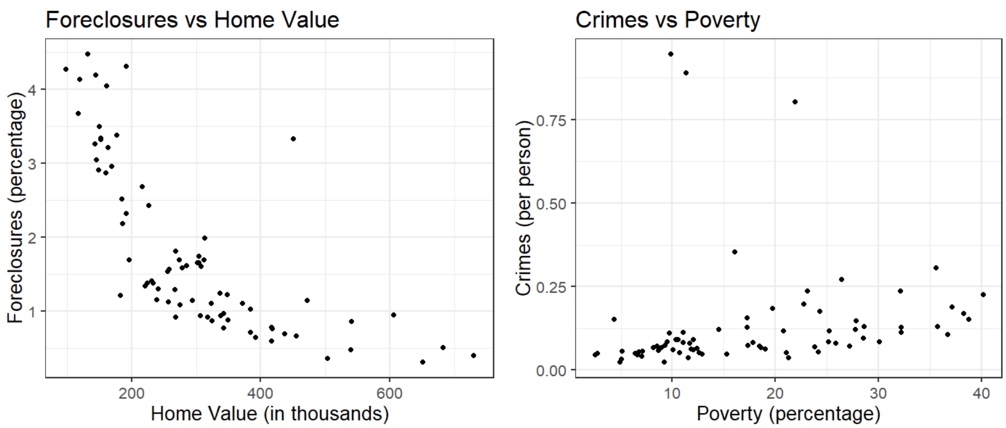


The first plot above shows a slight positive linear relationship between *rent* and *value*, suggesting rent increases when homes are more expensive. The second plot shows a slight negative nonlinear relationship between *rent* and *foreclosures*, suggesting rent decreases when neighborhoods have more foreclosures.





The first plot above shows a potential negative relationship between *rent* and *crimes*, suggesting rent decreases when there are more crimes. The plot also shows a few large outliers that could be significantly affecting the relationship. The second plot shows a negative nonlinear relationship between *rent* and *poverty*, suggesting rent decreases with an increase in poverty.



The first plot above a negative nonlinear relationship between *foreclosures* and *value*, suggesting foreclosures decrease when homes are more expensive. The second plot shows a potential positive relationship between *crimes* and *poverty*, suggesting crimes increase with an increase in poverty.

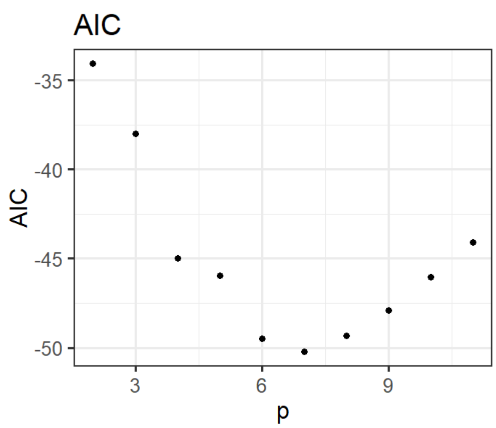
Collinearity

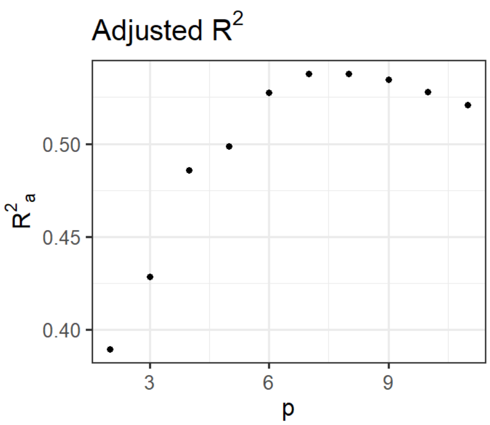
First, a linear model was constructed with *rent* as the response and the previously mentioned factors as predictors. The 16 predictors were then evaluated for collinearity. The variance inflation factors (VIFs) were checked and the predictor associated with the largest VIF was removed from the model. The VIFs were calculated for the new model and the predictor with the largest VIF was again removed from the model. This process was repeated until all of the VIFs were less than 5. The *no_college* variable (VIF = 51.42) was removed first, followed by the *white* (VIF = 18.29), *income* (VIF = 14.06), and *family* (VIF = 6.74) variables.

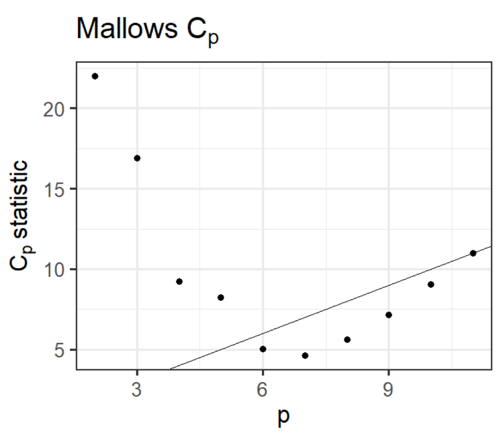
The condition indices were then checked to see if a collinearity problem still existed. Two condition indices were greater than 30, so the variable with the largest variance decomposition proportion associated with the largest condition index was removed. The condition indices were calculated for the new model and the variable with the largest variance decomposition proportion was again removed from the model. This process was repeated until all of the condition indices were less than 30. The *police* variable was removed first followed by the *clinics* variable. Six predictors were removed in total from these processes, leaving ten left in the updated model.

Variable Selection

The variable selection process was then performed by using the Akaike information criterion (AIC), adjusted R^2, and Mallows C_p as seen below.







The plots on the previous page show that both the AIC and Mallows C_p are minimized when p = 7 and the adjusted R^2 is maximized when p = 8.

Cross validation was then used to decide between these two models. Both leave one out and 10-fold cross validation showed that the p = 7 model had the smaller root mean square error (RMSE). This model corresponds to the variables *renter*, *value*, *poverty*, *foreclosures*, *crimes*, and *size* based on the logical matrix below.

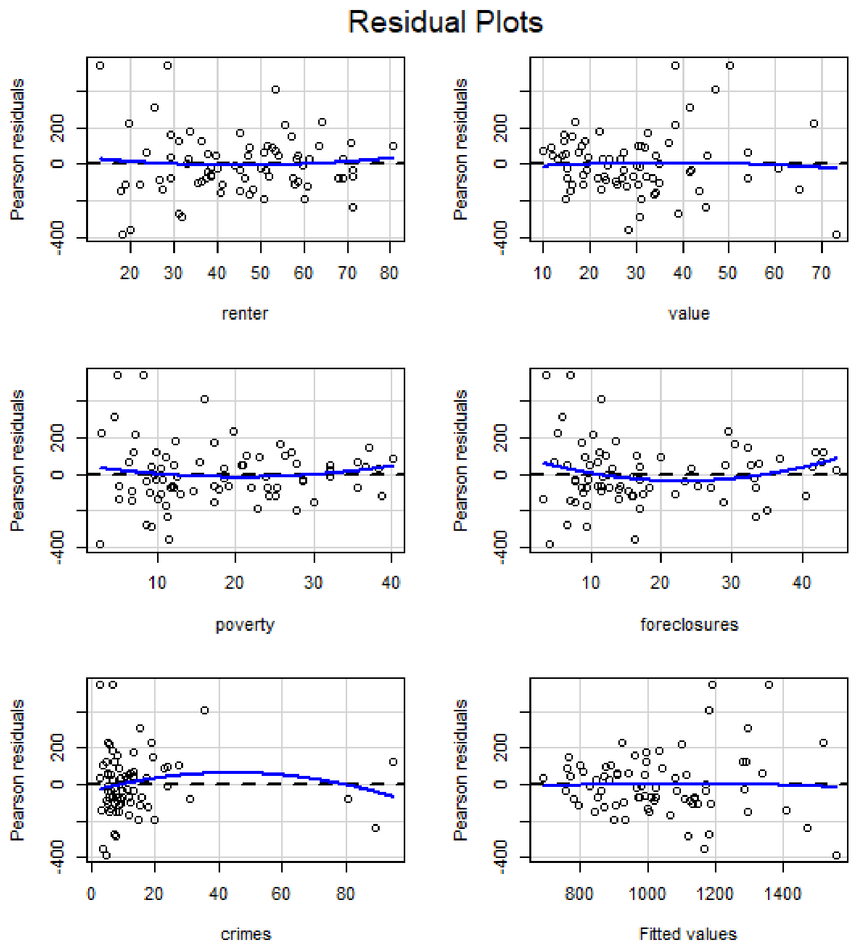
Decision Matrix

	(Intercept)	hispanic	age	renter	vacant	value	poverty	foreclosures	schools	crimes	size
1	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
2	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
3	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
4	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
5	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
6	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
7	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
8	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
9	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
10	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

After updating the model by removing the variables *vacant*, *schools*, *hispanic*, and *age*, the new model summary showed that all of the remaining variables except for *size* were significant. Since *size* was not of much interest, it was removed as well.

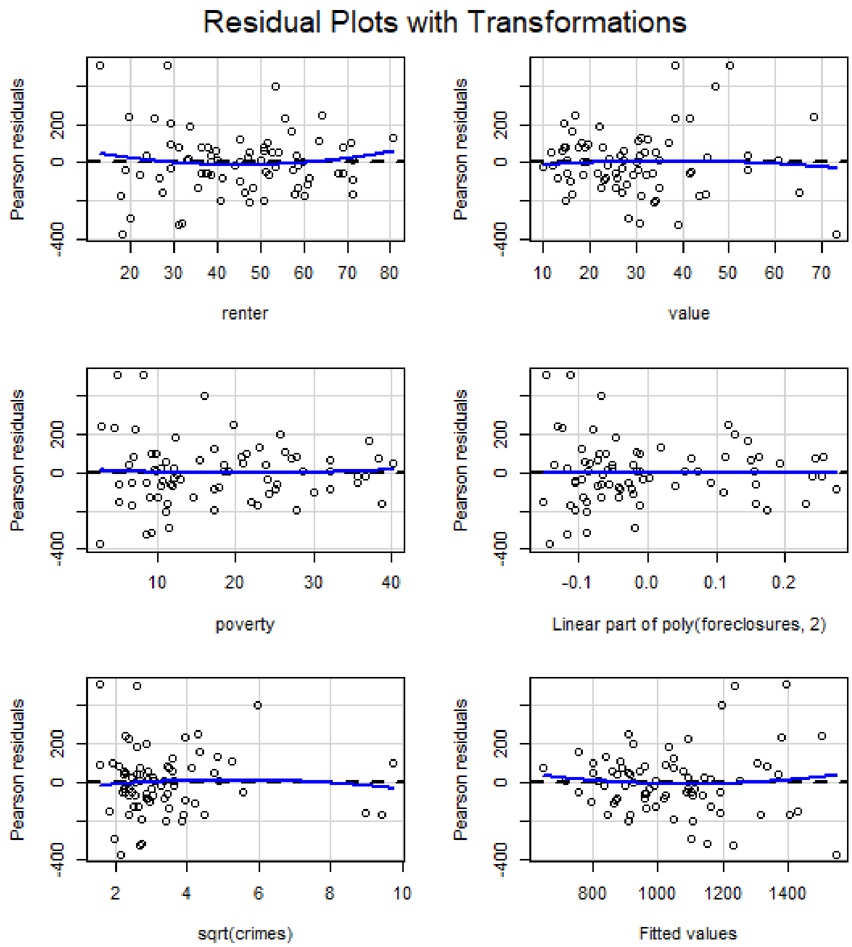
Model Structure

The structure and assumptions of the updated model were then checked using residual plots, added variable plots, marginal model plots, and component plus residual plots. The residual plots can be seen on the following page.

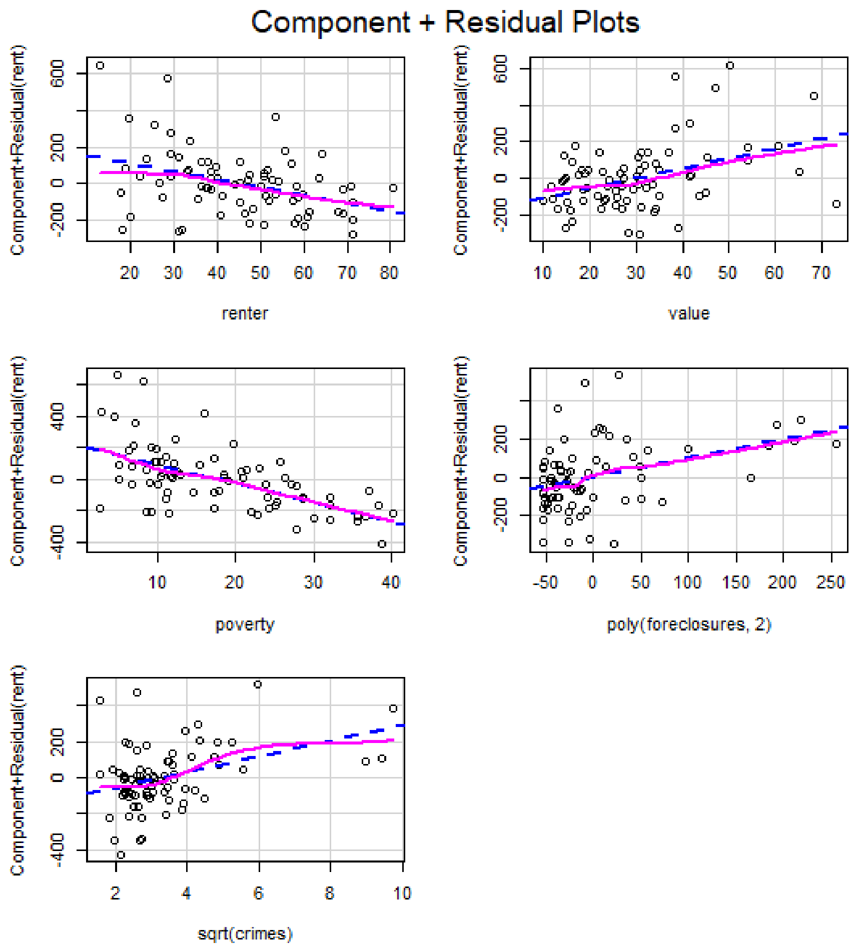


The above plots show that there is slight nonlinearity in the *foreclosures* residual plot and potential nonlinearity in the *crimes* residual plot, suggesting these variables may need to be transformed. The new plots after using a square root transformation for *crimes* and a second order polynomial transformation for *foreclosures* can be seen on the following page.

Residual Plots with Transformations

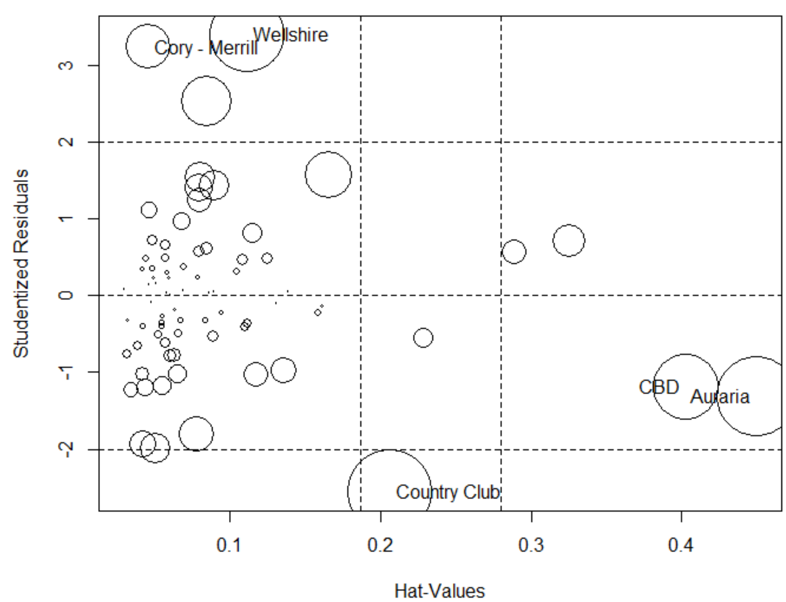


These new plots do not show any systematic patterns in the residuals for any of the variables, suggesting the transformations were effective. They also show there is no pattern in the residuals versus fitted values plot, meaning a linear model is appropriate for the data. The component plus residual plots on the following page were then used to verify the structure of the transformed model.



The above plots do not show clear nonlinear relationships for any of the variables. The fitted and smoothed lines are also similar for each variable and follow the pattern of the data. However, there do appear to be a few leverage points for the *crimes* variable, possibly affecting the model structure. These specific influential points are identified in the following influence plot.

Influence Plot



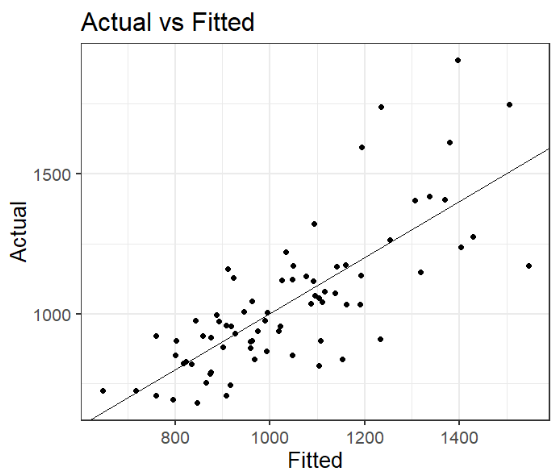
The plot above shows that the neighborhoods with larger circles are potentially more influential: Auraria, CBD, Country Club, Wellshire, and Cory-Merrill. These observations were removed from the model one-at-a-time to see if the signs of the variables' coefficients changed. None of them did, so the points were kept in the model. However, when all five of the points were removed from the model, the R^2 value increased by 8%, suggesting a better fit.

The final model after applying the transformations can be seen below.

Final Model

	Coefficient	Std. Error	t value	P-value
(Intercept)	1,145.54	102.52	11.17	4.81E-17
renter	-4.44	1.74	-2.55	0.01298
value	5.4	2.59	2.08	0.04087
poverty	-12.09	3.26	-3.71	0.00042
foreclosures	437.09	317.35	1.38	0.17294
foreclosures^2	416.64	210.68	1.98	0.05203
sqrt(crimes)	44.2	15.78	2.8	0.00662

The table on the previous page shows that the *renter*, *value*, *poverty*, and *crimes* variables are significant when compared to an alpha level of 0.05. The second order *foreclosures* term is right at the significance threshold and the first order *foreclosures* term is not significant. However, this term cannot be removed because of model hierarchy.



The above plot of actual versus fitted values shows that 59% of the variation in the response (*rent*) is explained by this model.

Hypothesis Testing

A permutation test was then performed to see if the crime variable was needed in the final model. The hypotheses for the test were set up as follows.

$$H_0:\ \beta_{crimes} = 0 \qquad \Omega = \left\{ \begin{array}{c} renter,\ value,\ poverty, \\ foreclosures,\ foreclosures^2,\ \sqrt{crimes} \end{array} \right\}$$

$$H_A:\ \beta_{crimes} \neq 0$$

The observed test statistics was 2.8 and the resulting p-value was 0.009. Therefore, there is convincing evidence that the model regressing *rent* on *renter*, *value*, *poverty*, *foreclosures*, and *foreclosures*^{2} would be improved by adding the square root of *crimes* regressor.

Predictions

The rent for a typical neighborhood, as exemplified by the median of each predictor variable, was estimated to be $979.18. The 95% confidence interval for this estimate was found to be (927.38, 1030.98), meaning we are 95% confident that the average rental price is between $979.18 and $1,030.98 for neighborhoods with 45% renter occupied housing, a home value of $274,075, 16% poverty, 1.39% foreclosures, and 0.084 crimes per person. The 95% prediction interval for this estimate was found to be (635.37, 1,322.99), meaning we are 95% confident that a specific neighborhood with 45% renter occupied housing, a home value of $274,075, 16% poverty, 1.39% foreclosures, and 0.084 crimes per person will have a rental price between $635.37 and $1,322.99.

Conclusions

The final model shows that home value, foreclosures, and crime are related to increased rental prices while renter occupied housing and poverty are related to decreased prices. So out of the 16 initial variables, home value, foreclosures, and crime were the only ones shown in this model to be related to unaffordable housing. However, these results cannot be extended to a larger population outside the city of Denver because the data did not come from a random sample. They also did not come from a randomized experiment, so causal conclusions cannot be made either. Though, these results are helpful to make recommendations to improve affordable housing within Denver.

One policy suggestion would be to encourage renting versus buying, which would reduce the effects of home value and foreclosures on rental prices. There would be less demand to buy, decreasing home value, and fewer people signing mortgages they possibly cannot afford. This would also increase the percentage of renter occupied housing, which the model showed reduced rental prices. This proposal could be achieved by reducing property tax and mortgage interest deductions, which would decrease the financial benefits of buying. A federal example is the Tax Cuts and Job Reform Act.

Another suggestion to reduce the effects of crime on rental prices would be to implement justice reinvestment initiatives. These initiatives cut spending on prisons and then reinvest that money into programs that improve safety and lower crime. One such program is an accountability court, which aims to address the cause of the criminal activity and not just punish the behavior.

The initiatives would also subject offenders to a risk and needs assessment in order to reduce recidivism rates. A federal example is the Sentencing Reform and Corrections Act.

It should be noted that the American Community Survey and foreclosure data these proposals are based on was from 5-10 years ago and could have significantly changed since then. The crime and neighborhood characteristic data were more recent, but comparing datasets from different time periods could lead to inaccurate results. Ideally, this analysis could be performed again with more recent data from the same time period to provide better policy recommendations. It would also be informative to use mortgage payments as the response variable instead of rent with the same predictor variables and compare the two models.

Another topic worth noting is the counterintuitive relationship between rent and crimes described by the model. It is usually assumed that neighborhoods with more crime are less expensive because they are not as favorable. This assumption can be seen in the bivariate plot between *rent* and *crimes*, but the final model suggests otherwise. Maybe the *crimes* outliers should have been removed at the very beginning of the analysis instead of assessing their effects at the end. Or maybe there is another underlying reason behind the unusual relationship.

Based on the previous examples, this project has shown the difficulty of working with real data. Some of these problems could potentially be mitigated by utilizing a robust linear model. This would more accurately account for the extreme values present in the data. Either way, these topics should be further investigated in future research.

References

Auraria Library. (n.d.). Data to Policy Project Resources. Retrieved from https://library.auraria.edu/d2pproject/resources

Avallone, M. (2018, March 20). Rent Versus Buy: How Tax Reform Adds To The Decision. Retrieved from https://www.forbes.com/sites/markavallone/2018/03/17/home-ownership-versus-renting-how-tax-reform-adds-to-the-decision/#40d9d93e7d82

Denver Office of Economic Development. (n.d.). About Affordable Housing. Retrieved from https://www.denvergov.org/content/denvergov/en/denver-office-of-economic-development/housing-neighborhoods/about-affordable-housing.html

Pye, J., & Rizer, A. (2017, October 31). Federal 'tough on crime' policies have failed - here's a new path forward. Retrieved from https://www.washingtonexaminer.com/federal-tough-on-crime-policies-have-failed-mdash-heres-a-new-path-forward