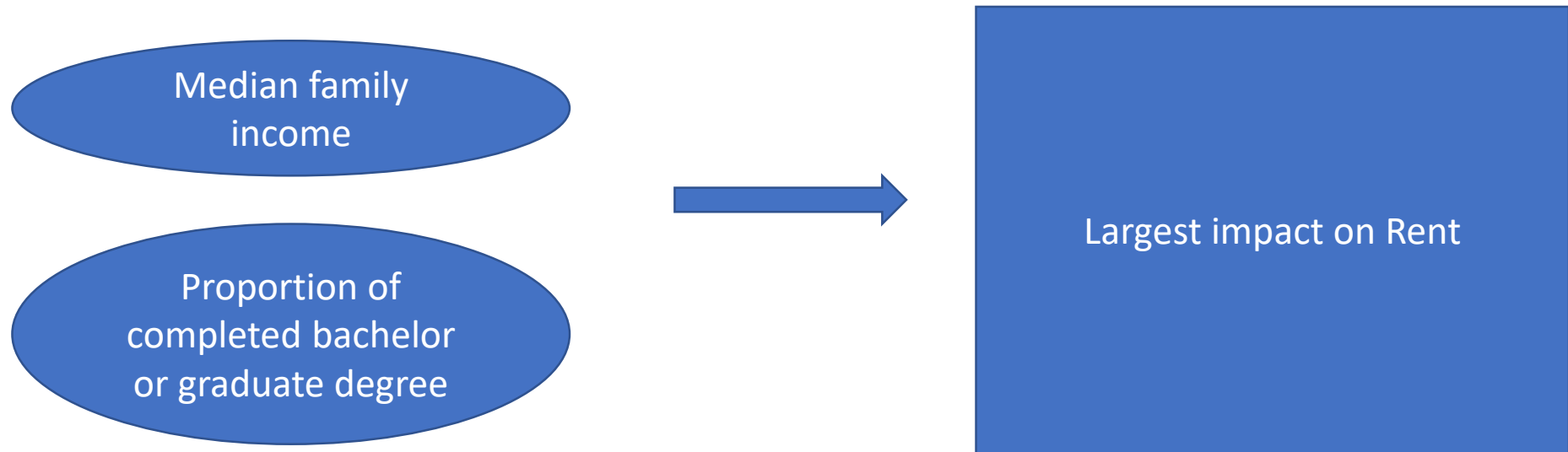


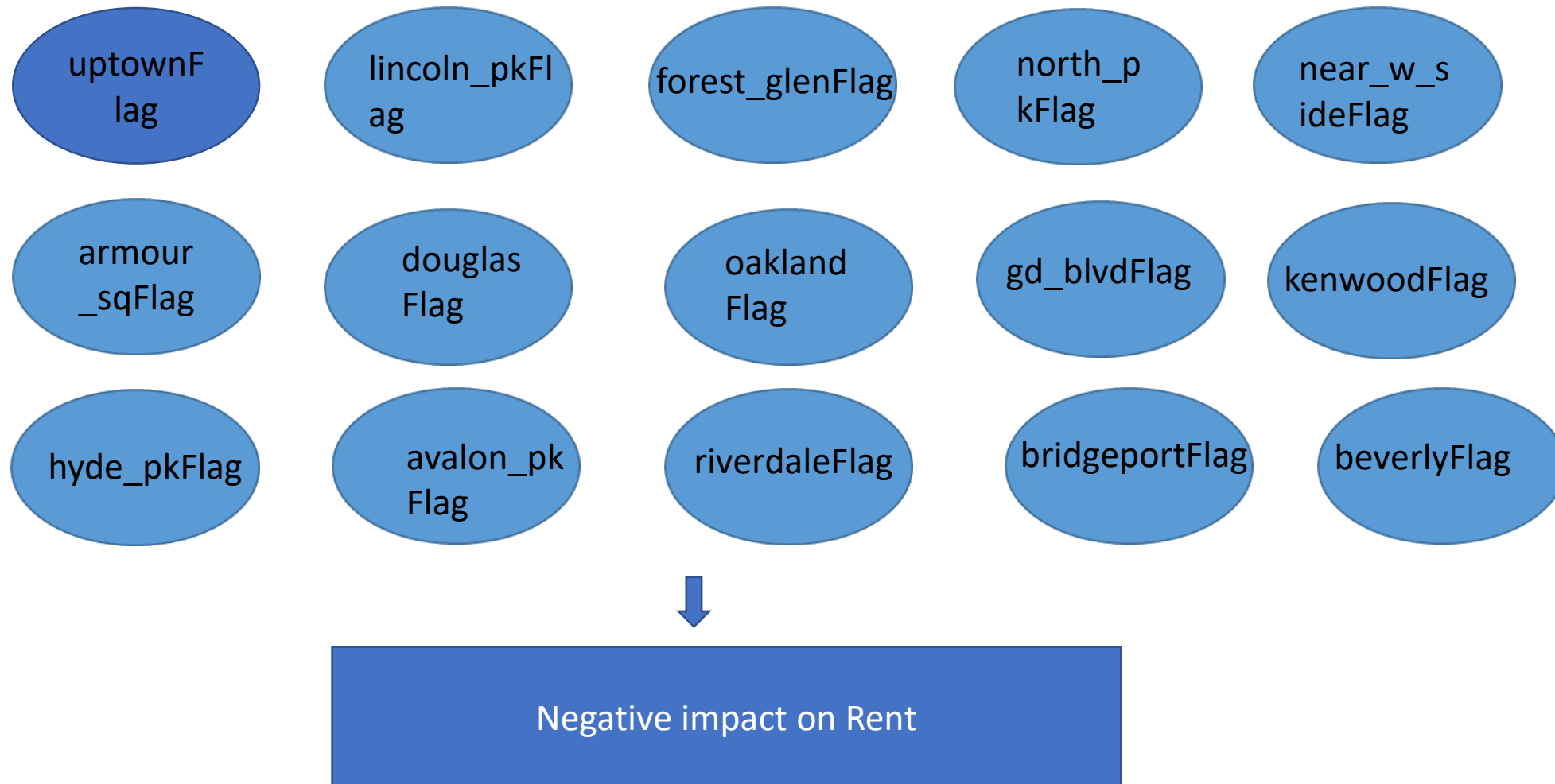
Chicago Rent Based On Socio-Economic Factors

Chicago Rent Data Summary

- Analyze which factors can be used to predict the median rent value for renter occupied houses in Chicago.



Dummy areas included in final model

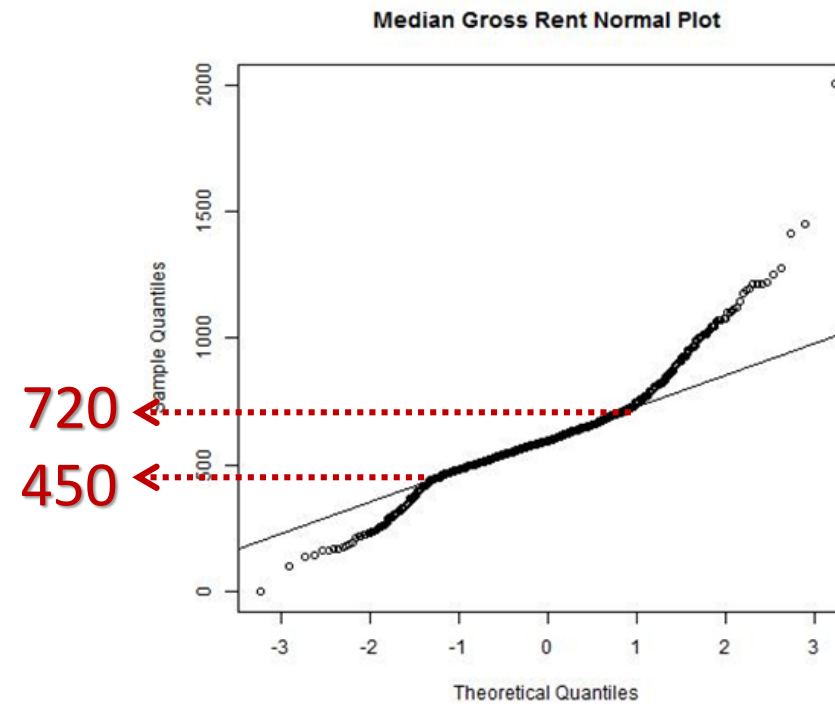
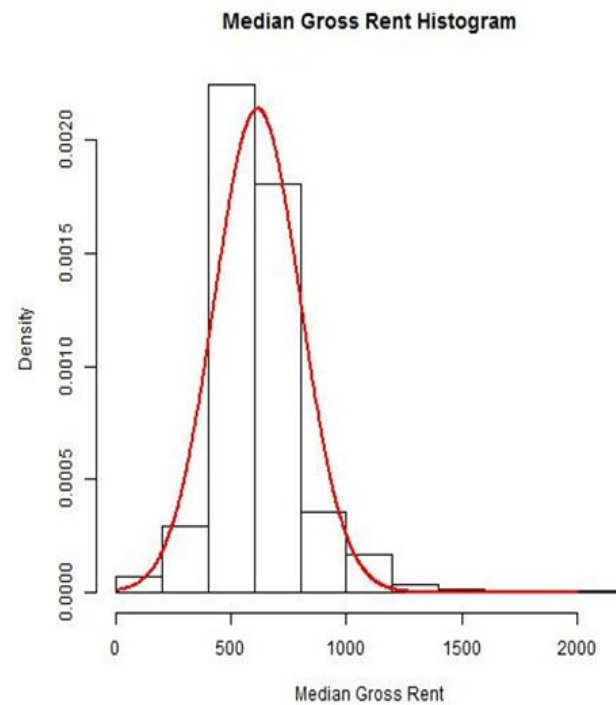


Exploratory Data Analysis

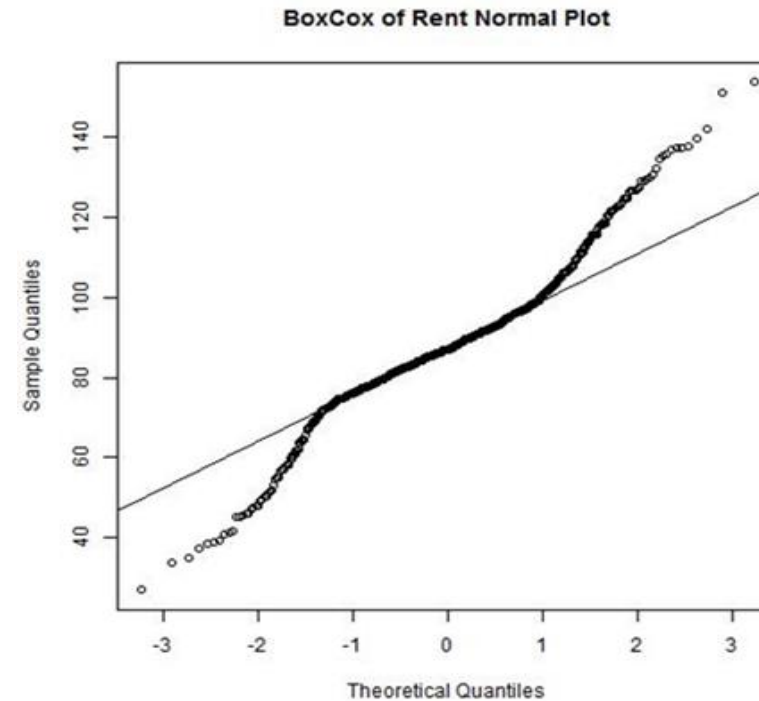
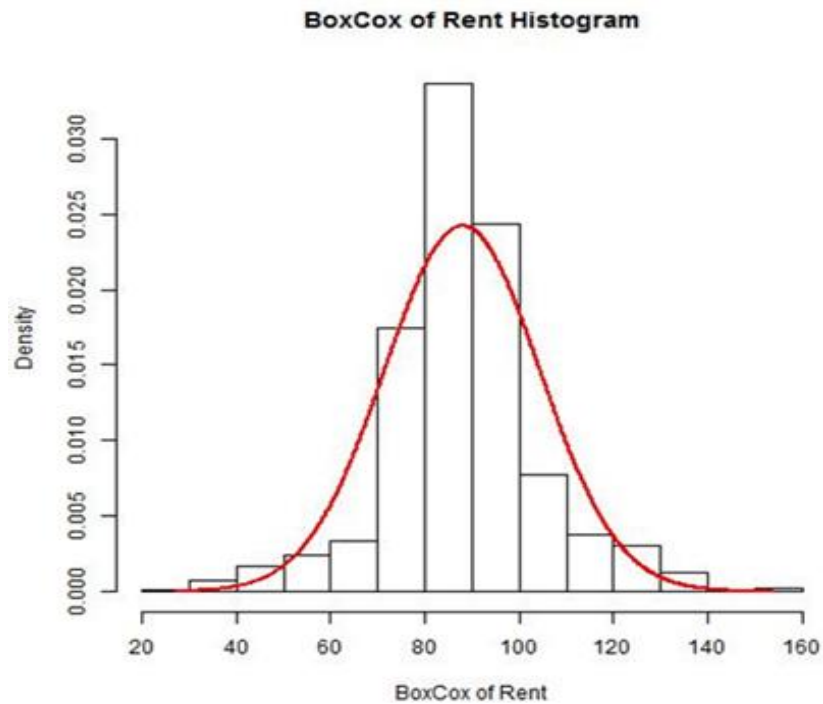
Data Treatment

- Created 76 dummy variables for canam77 (base o'hare)
- Removed rows which have data with dashes(55 rows)
- Removed a point with a value of 0 for Rent
- Converted foreclosure data from percentages to proportions
 - Divided data by 100

Rent Distribution Analysis



Box Cox of Rent Distribution Analysis



Rent Versus Independent Variable Analysis

Box cox of Rent vs x variables → Correlation Matrix

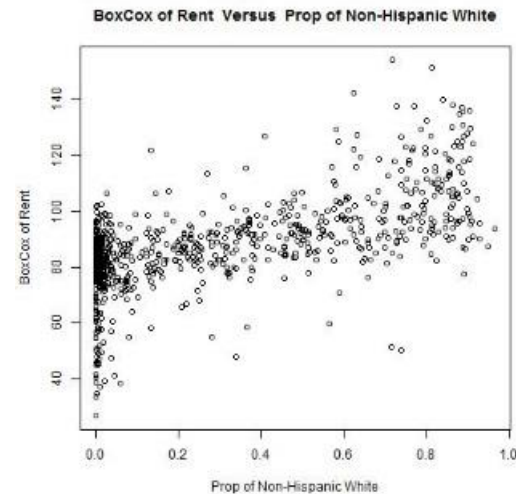
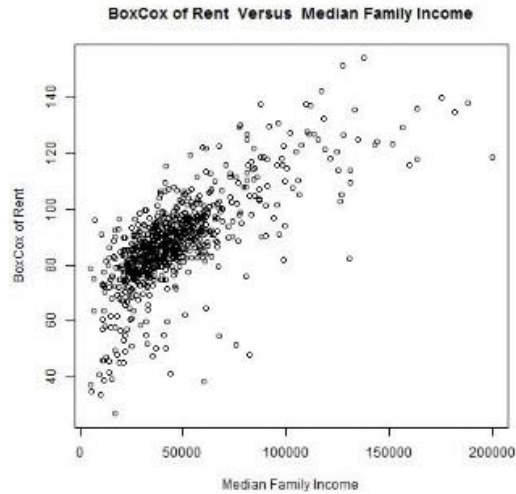
Box cox of Rent vs x variables (non-dummy) → Scatter plots

Box cox of Rent vs x variables (dummy) → Box plots

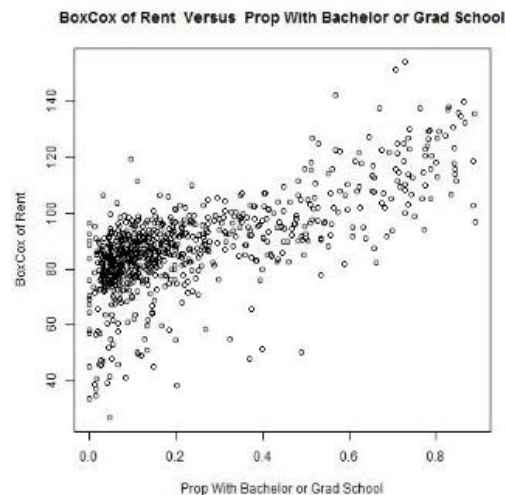
Rent Correlation Matrix

| | bcRent | | bcRent | | bcRent | | bcRent |
|----------------|--------------|-------------------|--------------|-----------------|--------------|---------------------|--------------|
| bcRent | 1 | near_n_sideFlag | 0.11475718 | loopFlag | 0.1561445 | garfield_ridgeFlag | 0.020689923 |
| mdfamy0 | 0.6612591 | edison_pkFlag | 0.03328108 | near_s_sideFlag | -0.09528046 | archer_htsFlag | 0.02333124 |
| educ160pro | 0.6332159 | norwood_pkFlag | 0.06808377 | armour_sqFlag | -0.05597019 | brighton_pkFlag | -0.036880389 |
| whitenhp | 0.5875034 | jefferson_pkFlag | 0.04370951 | douglasFlag | -0.09593001 | mckinley_pkFlag | -0.024609961 |
| nonWhitenhp | -0.587503447 | forest_glenFlag | 0.04405974 | oaklandFlag | -0.17695757 | bridgeportFlag | -0.041963733 |
| nonAanhp | 0.421659469 | north_pkFlag | -0.002199407 | fuller_pkFlag | -0.0741633 | new_cityFlag | -0.093419116 |
| renter | -0.3245187 | albany_pkFlag | 0.003709372 | gd_blvdFlag | -0.15106625 | w_elsdonFlag | 0.006050738 |
| educ8 | -0.3069567 | portage_pkFlag | 0.018534158 | kenwoodFlag | -0.01282956 | gage_pkFlag | -0.02363323 |
| asianp | 0.1391911 | irving_pkFlag | 0.023875923 | wash_pkFlag | -0.09915982 | clearingFlag | 0.01402115 |
| Noheat | -0.16642686 | dunningFlag | 0.061645354 | hyde_pkFlag | 0.06760334 | w_lawnFlag | 0.01337655 |
| totcrime | -0.01949634 | montclareFlag | 0.019718005 | woodlawnFlag | -0.09574908 | chgo_lawnFlag | -0.01569983 |
| murder | -0.12239347 | belcraginFlag | 0.021980487 | s_shoreFlag | -0.0161149 | w_englewoodFlag | -0.01085611 |
| hisp | -0.05278583 | hermosaFlag | -0.005221939 | chathamFlag | -0.01219246 | englewoodFlag | -0.08063842 |
| aanhp | -0.4216595 | avondaleFlag | 0.003975664 | avalon_pkFlag | -0.03242099 | grtr_grnd_crsngFlag | -0.07178369 |
| property | 0.07602817 | logan_sqFlag | 0.080025257 | s_chgoFlag | -0.06810314 | ashburnFlag | 0.04824505 |
| forecl | -0.10395209 | humboldt_pkFlag | -0.01471096 | burnsideFlag | -0.00329918 | ashburn_greshFlag | -0.02729073 |
| robbery | -0.24447906 | w_townFlag | 0.09641833 | calumet_htsFlag | 0.01187254 | beverlyFlag | 0.0480926 |
| personal | -0.35228226 | austinFlag | 3.06128E-05 | roselandFlag | 0.02069463 | wash_htsFlag | 0.014685914 |
| rentAffNdx | 0.000465493 | w_garfield_pkFlag | -0.03425865 | pullmanFlag | -0.03347029 | mt_greenwoodFlag | 0.017869933 |
| uptownFlag | -0.01186518 | e_garfield_pkFlag | -0.0843912 | s_deeringFlag | -0.01519793 | morgan_pkFlag | 0.019094673 |
| lincoln_sqFlag | 0.04014348 | near_w_sideFlag | -0.07727448 | e_sideFlag | -0.03701793 | rogers_pkFlag | 0.01122531 |
| n_centerFlag | 0.10451594 | n_lawndaleFlag | -0.1150703 | w_pullmanFlag | -0.011902478 | w_ridgeFlag | 0.05464501 |
| lakeviewFlag | 0.26512215 | s_lawndaleFlag | -0.1123547 | riverdaleFlag | -0.097419883 | edgewaterFlag | 0.02611834 |
| lincoln_pkFlag | 0.26513777 | lower_w_sideFlag | -0.08306181 | hegewischFlag | -0.015878366 | lowRentFlag | -0.643384867 |
| | | | | | | highRentFlag | -0.664152921 |

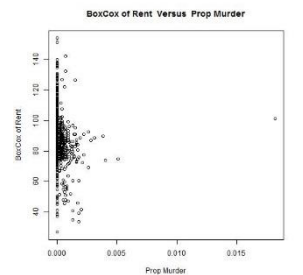
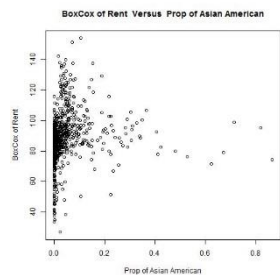
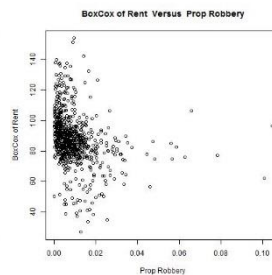
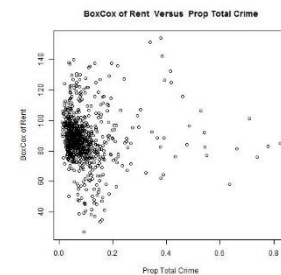
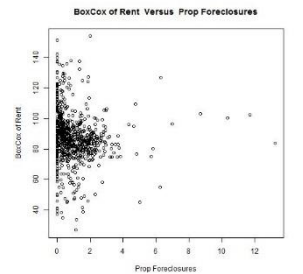
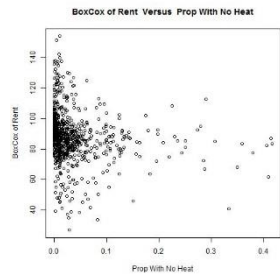
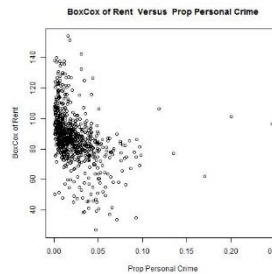
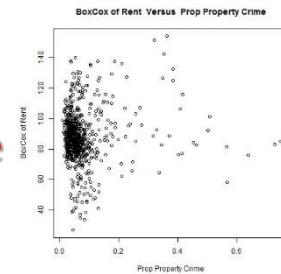
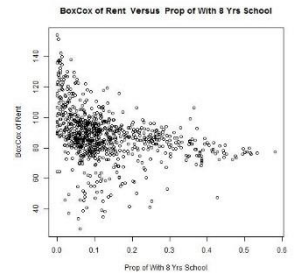
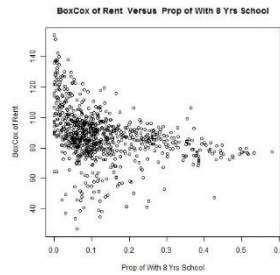
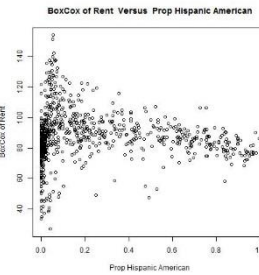
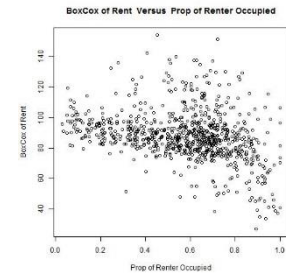
Rent Versus Non-Dummy Variable Scatterplots



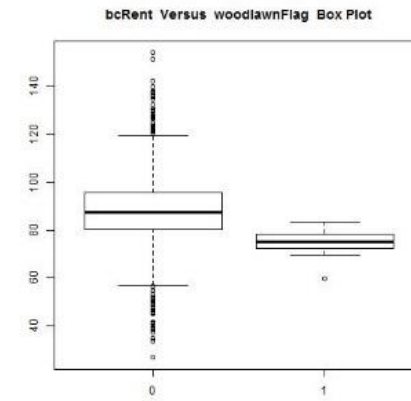
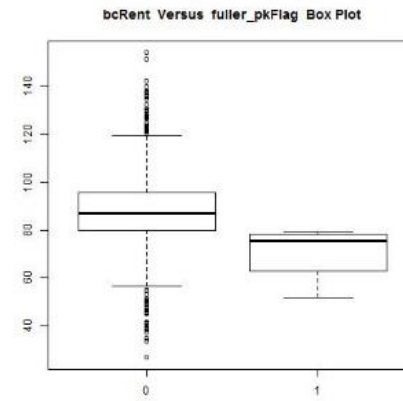
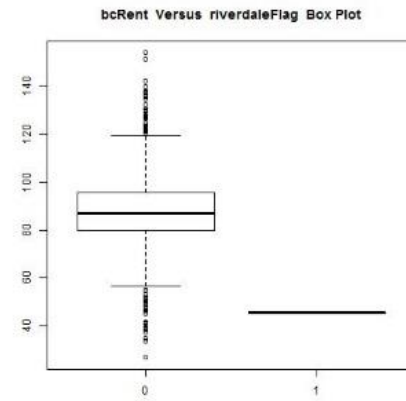
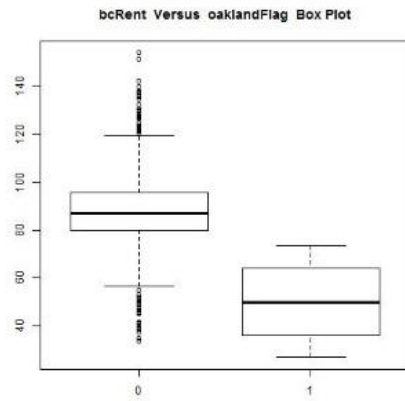
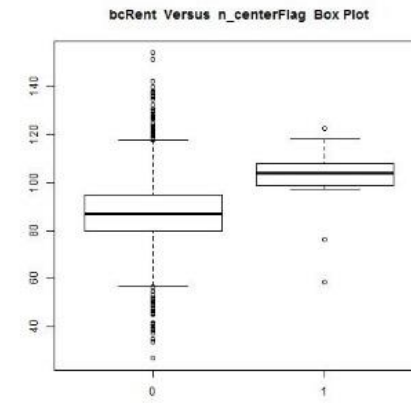
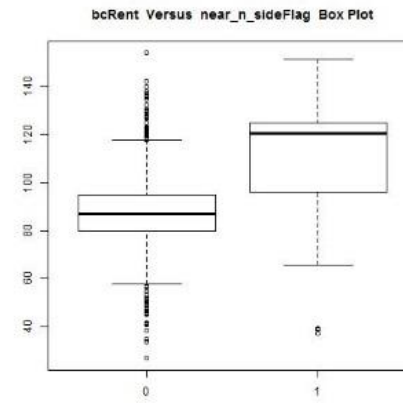
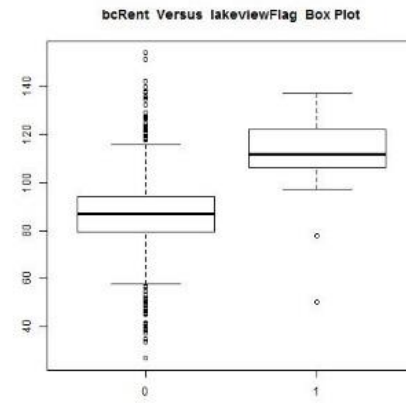
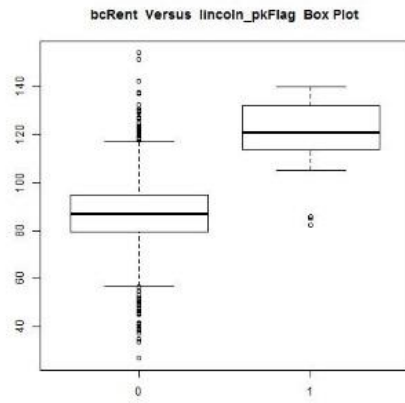
Median Family Income Proportion of Non-Hispanic White



Proportion with Bachelor or Grad school



Rent Versus Dummy Variable Boxplots



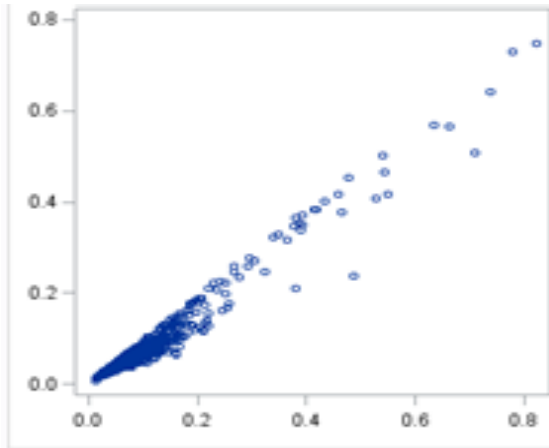
Independent Variable Multicollinearity

| | mdfamy0 | asianp | whitenhp | aanhp | hisp | renter | educ8 | educ160pro | Noheat | totcrime | murder | robbery | personal | property | forecl |
|------------|---------|---------|----------|---------|---------|---------|---------|------------|---------|----------|---------|---------|----------|----------|---------|
| mdfamy0 | 1 | 0.1194 | 0.7235 | -0.4488 | -0.1538 | -0.3667 | -0.3766 | 0.7884 | -0.2297 | -0.0316 | -0.1776 | -0.2725 | -0.4000 | 0.0753 | -0.1725 |
| asianp | 0.1194 | 1 | 0.2310 | -0.3419 | -0.0413 | 0.0638 | -0.0230 | 0.3004 | 0.0226 | -0.0787 | -0.0781 | -0.1610 | -0.2341 | -0.0259 | -0.1666 |
| whitenhp | 0.7235 | 0.2310 | 1 | -0.7325 | -0.0642 | -0.3181 | -0.3073 | 0.7002 | -0.1894 | -0.1512 | -0.1950 | -0.4168 | -0.5181 | -0.0306 | -0.3579 |
| aanhp | -0.4488 | -0.3419 | -0.7325 | 1 | -0.6035 | 0.2075 | -0.3239 | -0.4166 | 0.2384 | 0.2139 | 0.2085 | 0.4738 | 0.5845 | 0.0850 | 0.4280 |
| hisp | -0.1538 | -0.0413 | -0.0642 | -0.6035 | 1 | 0.0150 | 0.8395 | -0.2344 | -0.1620 | -0.1345 | -0.0762 | -0.2093 | -0.2428 | -0.0884 | -0.2043 |
| renter | -0.3667 | 0.0638 | -0.3181 | 0.2075 | 0.0150 | 1 | 0.1369 | -0.0137 | 0.4308 | 0.1511 | 0.0642 | 0.2662 | 0.3211 | 0.0857 | 0.0862 |
| educ8 | -0.3766 | -0.0230 | -0.3073 | -0.3239 | 0.8395 | 0.1369 | 1 | -0.4751 | -0.1124 | -0.0893 | -0.0259 | -0.0437 | -0.0525 | -0.0892 | -0.1049 |
| educ160pro | 0.7884 | 0.3004 | 0.7002 | -0.4166 | -0.2344 | -0.0137 | -0.4751 | 1 | -0.0116 | 0.0251 | -0.1415 | -0.2303 | -0.3468 | 0.1264 | -0.1970 |
| Noheat | -0.2297 | 0.0226 | -0.1894 | 0.2384 | -0.1620 | 0.4308 | -0.1124 | -0.0116 | 1 | -0.0082 | 0.0191 | 0.0757 | 0.1095 | -0.0402 | 0.1591 |
| totcrime | -0.0316 | -0.0787 | -0.1512 | 0.2139 | -0.1345 | 0.1511 | -0.0893 | 0.0251 | -0.0082 | 1 | 0.2490 | 0.6805 | 0.6606 | 0.9776 | 0.1280 |
| murder | -0.1776 | -0.0781 | -0.1950 | 0.2085 | -0.0762 | 0.0642 | -0.0259 | -0.1415 | 0.0191 | 0.2490 | 1 | 0.1911 | 0.4497 | 0.1635 | 0.0672 |
| robbery | -0.2725 | -0.1610 | -0.4168 | 0.4738 | -0.2093 | 0.2662 | -0.0437 | -0.2303 | 0.0757 | 0.6805 | 0.1911 | 1 | 0.8859 | 0.5430 | 0.2209 |
| personal | -0.4000 | -0.2341 | -0.5181 | 0.5845 | -0.2428 | 0.3211 | -0.0525 | -0.3468 | 0.1095 | 0.6606 | 0.4497 | 0.8859 | 1 | 0.4879 | 0.3066 |
| property | 0.0753 | -0.0259 | -0.0306 | 0.0850 | -0.0884 | 0.0857 | -0.0892 | 0.1264 | -0.0402 | 0.9776 | 0.1635 | 0.5430 | 0.4879 | 1 | 0.0629 |
| forecl | -0.1725 | -0.1666 | -0.3579 | 0.4280 | -0.2043 | 0.0862 | -0.1049 | -0.1970 | 0.1591 | 0.1280 | 0.0672 | 0.2209 | 0.3066 | 0.0629 | 1 |

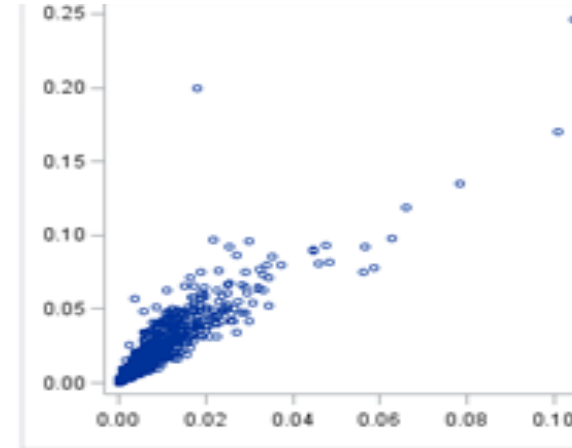
property & totcrime → 0.9776
 personal & robbery → 0.8859
 educ8 & hisp → 0.8395

educ160pro & mdfamy0 → 0.7884
 aanhp & whitenhp → 0.7325
 whitenhp & mdfamy0 → 0.7235
 educ160pro & whitenhp → 0.7002

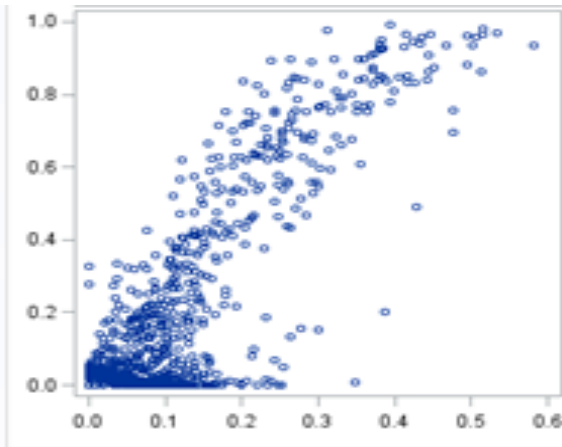
Independent Variable Multicollinearity Plots



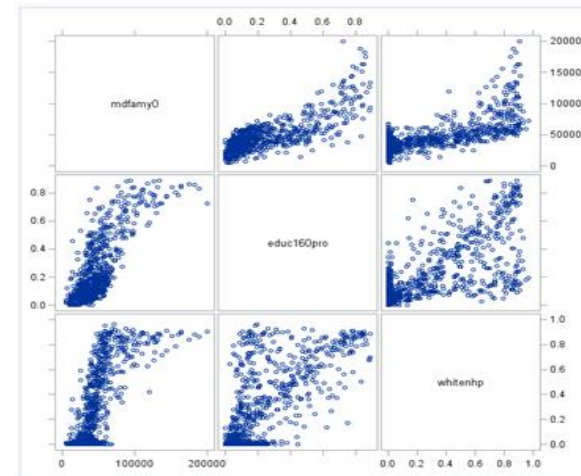
property & totcrime
→ 0.9776



personal & robbery
→ 0.8859



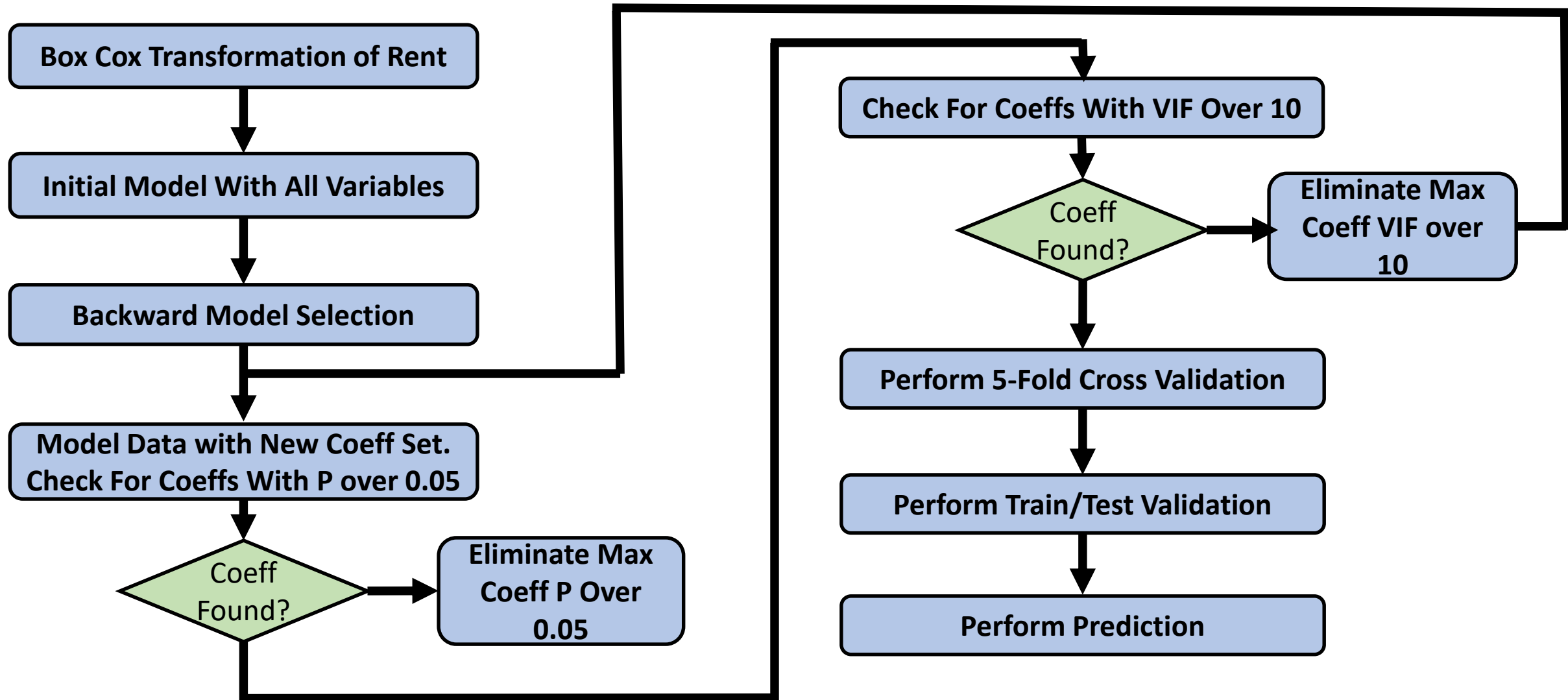
educ8 & hisp
→ 0.8395



educ160pro
& mdfamy0
& whitenhp

Model Analysis

Final Model Creation



Model Interpretation

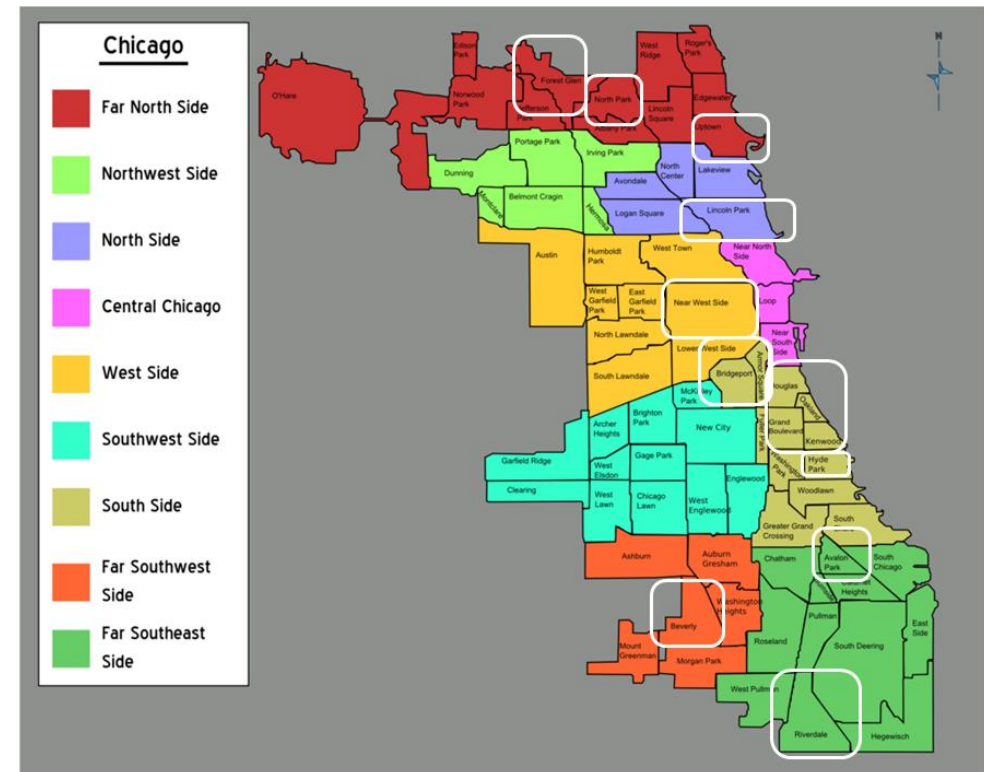
- Socio-economic variables with positive effects on rent
 - Median family income
 - Proportion of people over 25 that completed bachelor or graduate degrees
- Socio-economic variables with negative effects on rent
 - Proportion of people over 25 that completed 0 to 8 years of school
 - Proportion of African American non-Hispanic residents
 - Proportion of white non-Hispanic residents
 - Proportion of total renter occupied housing
 - Proportion of foreclosures

| | Coefficient Estimate |
|-------------|----------------------|
| (Intercept) | 88.000000 |
| mdfamy0 | 0.000199 |
| whitenhp | -9.030000 |
| aanhp | -10.100000 |
| renter | -10.200000 |
| educ8 | -24.800000 |
| educ160pro | 33.200000 |
| forecl | 115.000000 |

Model Interpretation Continued

- Neighborhoods with negative effects on rent
 - Uptown, Lincoln Park, Forest Glen, North Park, Near W Side, Armour Square, Douglas, Oakland, Grand Boulevard, Kenwood, Hyde Park, Avalon Park, Riverdale, Bridgeport, Beverly

| | Coefficient Estimate |
|-----------------|----------------------|
| lincoln_pkFlag | -8.330000 |
| forest_glenFlag | -9.920000 |
| north_pkFlag | -10.900000 |
| near_w_sideFlag | -9.330000 |
| armour_sqFlag | -11.400000 |
| douglasFlag | -12.400000 |
| oaklandFlag | -27.200000 |
| gd_blvdFlag | -7.160000 |
| kenwoodFlag | -7.390000 |
| hyde_pkFlag | -11.400000 |
| avalon_pkFlag | -14.100000 |
| riverdaleFlag | -26.700000 |
| bridgeportFlag | -7.910000 |



Model Coefficient Analysis

- Most significant variables (in descending order)
 - (+) Proportion of people over 25 that completed Bachelor or graduate degree
 - (+) Median family income
 - (-) Proportion of African American non-Hispanic residents

| Standard Coefficients | | | | | | | | |
|-----------------------|--------------|-----------------|-------------|----------------|------------|-------------|-----------------|---------------|
| educ160pro | mdfamy0 | aanhp | whitenhp | educ8 | renter | oaklandFlag | near_w_sideFlag | hyde_pkFlag |
| 0.4503 | 0.3257 | -0.2671 | -0.1736 | -0.1659 | -0.1372 | -0.1156 | -0.1136 | -0.0905 |
| forecl | douglasFlag | lincoln_pkFlag | beverlyFlag | bridgeportFlag | uptownFlag | gd_blvdFlag | riverdaleFlag | armour_sqFlag |
| 0.0846 | -0.0831 | -0.0784 | -0.078 | -0.0626 | -0.0606 | -0.0605 | -0.0569 | -0.054 |
| avalon_pkFlag | north_pkFlag | forest_glenFlag | kenwoodFlag | | | | | |
| -0.0522 | -0.052 | -0.0422 | -0.0415 | | | | | |

- All variable VIF values are under 10

| VIF Values | | | | | | | | |
|---------------|--------------|-----------------|-------------|----------------|------------|-------------|-----------------|---------------|
| educ160pro | mdfamy0 | aanhp | whitenhp | educ8 | renter | oaklandFlag | near_w_sideFlag | hyde_pkFlag |
| 6.31 | 6.03 | 9.34 | 8.78 | 5.28 | 2.1 | 1.03 | 1.07 | 1.18 |
| forecl | douglasFlag | lincoln_pkFlag | beverlyFlag | bridgeportFlag | uptownFlag | gd_blvdFlag | riverdaleFlag | armour_sqFlag |
| 1.29 | 1.04 | 1.64 | 1.06 | 1.05 | 1.05 | 1.1 | 1.01 | 1.02 |
| avalon_pkFlag | north_pkFlag | forest_glenFlag | kenwoodFlag | | | | | |
| 1.03 | 1.02 | 1.03 | 1.04 | | | | | |

Model Residual Analysis

- Standardized residual normal plot shows non-normal behavior
 - Plot has thick tails at high and low predicted rents
 - Plot behavior matches behavior of rent variable
- Addition of other variables did not fix non-normal behavior
 - Added non-African American and non-white dummy variables
 - Added rent affordability index
 - Added high rent and low rent dummy variables
 - Improved normal plot
 - Created bias in predicted scatterplot

Model Validation

- Performed 5-fold cross validation on model
 - Used final model with all data
- Performed validation with a 75% training set and a 25% testing set
 - Split data into sets randomly
 - Created model using variables given in final model and the training set
 - Predicted data in the testing set with training set model

| All Data Set Cross Validation | |
|-------------------------------|-------|
| Model R-Sq | 0.672 |
| CV Model RMSE | 9.44 |
| CV Model MAE | 6.62 |
| CV Model MAPE | 8.46 |
| CV Test RMSE | 9.92 |
| CV Test MAE | 6.91 |
| CV Test MAPE | 8.85 |
| CV Test R-Sq | 0.638 |
| CV RMSE Diff | -0.48 |
| CV MAE Diff | -0.29 |
| CV MAPE Diff | -0.39 |
| Model R-Sq - CV R-Sq | 0.034 |

| Training and Testing Data Set Validation | |
|--|-------|
| Model R-Sq | 0.67 |
| Val Train RMSE | 9.46 |
| Val Train MAE | 6.62 |
| Val Train MAPE | 8.45 |
| Val Test RMSE | 9.6 |
| Val Test MAE | 6.84 |
| Val Test MAPE | 8.91 |
| Val Test R-Sq | 0.664 |
| Val RMSE Diff | -0.14 |
| Val MAE Diff | -0.22 |
| Val MAPE Diff | -0.46 |
| Train R-Sq - Test R-Sq | 0.006 |

Model Predictions

- Performed 4 predictions using final model
 - Used 4 eliminated data points
 - Replaced missing data with column means
 - First 2 were only missing 2 variables
 - Last 2 were missing 3 variables

| All Data Model | Predicted Value | Pred Int LL | Pred Int UL | Conf Int LL | Conf Int UL | Actual Value | Actual - Pred | Diff % of Actual |
|----------------|-----------------|-------------|-------------|-------------|-------------|--------------|---------------|------------------|
| | \$762.00 | \$553.00 | \$994.00 | \$738.00 | \$786.00 | \$728.00 | -\$34.00 | -4.67 |
| | \$662.00 | \$465.00 | \$884.00 | \$640.00 | \$685.00 | \$545.00 | -\$117.00 | -21.47 |
| | \$746.00 | \$539.00 | \$978.00 | \$723.00 | \$771.00 | \$925.00 | \$179.00 | 19.35 |
| | \$689.00 | \$489.00 | \$915.00 | \$667.00 | \$713.00 | \$1,033.00 | \$344.00 | 33.30 |

Model Inconsistencies

- Proportion of white non-Hispanic
 - Correlation is positive – Model coefficient is negative
- Proportion of foreclosure
 - Correlation is negative – Model coefficient is positive
- Lincoln Park and North Park neighborhoods
 - Boxplot shows Lincoln Park and North Park increase rent - Model coefficients are negative

| Correlation of Rent and Box Cox of Rent to Independent Variables (Only > 0.1 Shown) | | | | | |
|---|--------------|-----------------|-----------------|--------------|-----------------|
| | Rent | Box Cox of Rent | | Rent | Box Cox of Rent |
| mdfamy0 | 0.703704412 | 0.6612591 | property | 0.116441188 | 0.07602817 |
| asianp | 0.132488176 | 0.1391911 | forecl | -0.117521105 | -0.10395209 |
| whitenhp | 0.603970093 | 0.5875034 | n_centerFlag | 0.10811778 | 0.10451594 |
| aanhp | -0.409192145 | -0.4216595 | lakeviewFlag | 0.29245562 | 0.26512215 |
| reenter | -0.282186323 | -0.3245187 | lincoln_pkFlag | 0.30330934 | 0.26513777 |
| educ8 | -0.3340027 | -0.3069567 | near_n_sideFlag | 0.17151754 | 0.11475718 |
| educ160pro | 0.67587045 | 0.6332159 | n_lawndaleFlag | -0.11283053 | -0.1150703 |
| Noheat | -0.169565777 | -0.16642686 | s_lawndaleFlag | -0.11511727 | -0.1123547 |
| murder | -0.117009175 | -0.12239347 | loopFlag | 0.19101863 | 0.1561445 |
| robbery | -0.219150652 | -0.24447906 | oaklandFlag | -0.13316763 | -0.17695757 |
| personal | -0.318707323 | -0.35228226 | gd_blvdFlag | -0.13566259 | -0.15106625 |

Model Recommendation

- Divide data set into three groups
 - Low rent group (rents lower than \$450)
 - Middle rent group (rents between \$450 and \$720)
 - High rent group (rents higher than \$720)
- Create interaction terms for highly correlated independent variables

| Independent Variable Multicollinearity Summary (Only > -.6 Shown) | | | | | |
|---|------------|-------------|--|------------|--------------|
| Income Multicollinearity Correlations | | | White Proportion Multicollinearity Correlations | | |
| mdfamy0 | whitenhp | 0.723461691 | whitenhp | mdfamy0 | 0.723461691 |
| mdfamy0 | educ160pro | 0.788419205 | whitenhp | aanhp | -0.732528761 |
| | | | whitenhp | educ160pro | 0.700232204 |
| Total Crime Multicollinearity Correlations | | | Hispanic Proportion Multicollinearity Correlations | | |
| totcrime | robbery | 0.680450686 | hisp | educ8 | 0.83948022 |
| totcrime | personal | 0.660623077 | hisp | aanhp | -0.603465283 |
| totcrime | property | 0.9776299 | | | |
| Robbery Multicollinearity Correlations | | | African American Proportion Multicollinearity Correlations | | |
| robbery | personal | 0.8858553 | aanhp | hisp | -0.603465283 |
| robbery | totcrime | 0.680450686 | aanhp | whitenhp | -0.732528761 |

Conclusion

- Model is not sufficient for use in predictions
 - Model has non-normal residuals
 - Model results are inconsistent with rent and independent correlation, scatterplots, and boxplots
 - Prop white non-Hispanic, foreclosures, Lincoln Park, North Park
- Model should be done again using recommendations
 - Data set division based on rent
 - Interaction term additions