Crime2018

Fionnuala McPeake

December 2, 2018

Abstract

This report outlines the attempt made to predict house prices based on crime rates in the surrounding area. In order to capture the influence of different types of crimes, offenses were categorized into one of several categories. Multiple models were made with the purpose of capturing this relationship, though due to the limitations of the data, none sufficiently explained it.

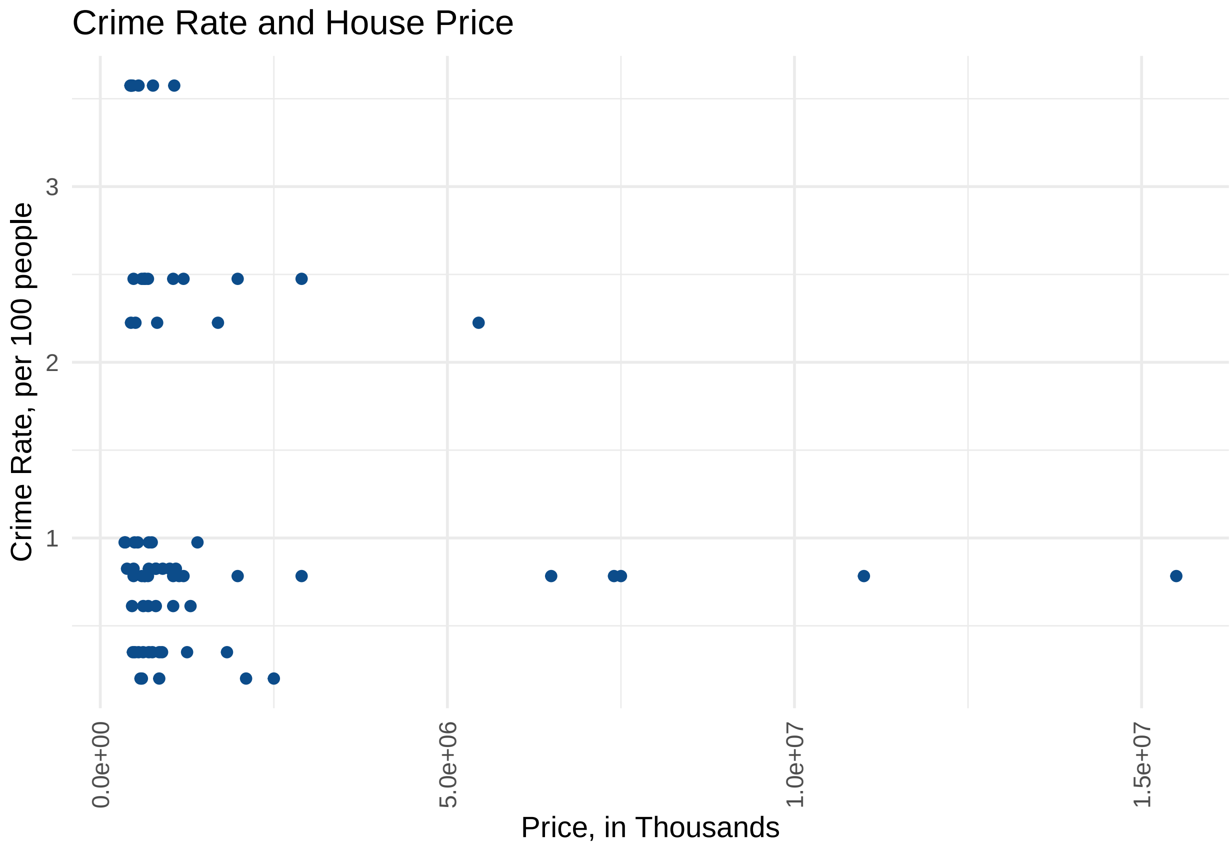
Introduction

The purpose of this work is to predict house price based primarily on crime. While intuition may have a person believe that safer areas may be more expensive, this may not be the case, especially when different types of offenses are considered. For this reason, the type of crime committed will be taken into account. The results of this work may help neighborhoods predict the type of crime they are most vulnerable to, such as burglary or vandalism, and set up precautions accordingly. Furthermore, results contrary to intuition can help disprove stereotypes and improve the reputation of areas that are more affordable. It should be acknowledged that many factors contribute to the price of a house, and area in general, and that crime is only one factor. While studies have been performed to explore the relationship between crime and house prices, even the authors recognize the fluid nature of crime movement and the multiple elements that influence house prices. However, despite this, there has been a connection shown between house prices and crimes, particularly robbery and aggravated assault. One study found that when homicide and robberies decreased in Rio de Janeiro, property values increased 5-10% [1]. The following report will chronicle the effort made to explore the relationship between house price and crime committed in the surrounding area.

Method

The primary data source for this analysis is the Boston crime incidents report [2], in which the Boston Police Department report details of incidents to which they respond. This data set included date, hour, offense code group, offense description, and latitude and longitude for the years 2015-2018. In order to make the data more manageable, all of the unique offense descriptions were observed and classified into more board description of the incident, such as domestic, violent, and non-violent. To find the location of the incident in more interpretable terms, the latitude and longitude for each occurrence was run through the Federal Communications Commission API [3] to determine the census block to which it belongs. In order to create a range of economic situations, the median house values for different neighborhoods and zip codes in Boston was researched on Zillow to select neighborhoods to include in the model. House prices of properties currently on the market were then researched for each area. After selecting appropriate groups, the census block number for the desired locations was determined by examining both Google maps and a map of census blocks [4] for Boston. The crime rate for each area was calculated based on the fact that census blocks, optimally, 4,000 people per block [5]. It should be noted that Boston is a very expensive, and while there is a wide range of house prices, it is not as variable as would be found in other cities.

In order to explore how different crime rates influence property values, models were run controlling for the total crime rate, the rates of all individual crime types, and only including rates for crimes that studies have indicated are the most influential on house prices. In order to add more information regarding the areas, the percentage of the population that is white [6] [7] [8] [9] [10] [11] [12] [13] was added as a group level predictor. This leaves city as the only random effect. To capture the differences between areas, a mixed effects model was used. Exploratory data analysis shows that there are different rates of crime associated with different property prices, weakly indicating that crime rates and property values are negatively correlated.



Graph of crime rate versus house price. Graph indicates that for the data gathered, higher priced areas generally have lower crime rates.

However, this trend was not captured in the models. In fact, despite the different inputs, all of the models had nearly identical outputs, with little variation between areas being captured. The same phenomenon occurred when the models were fit linearly. Though the results were marginally better, they still did not indicate anything of substance.

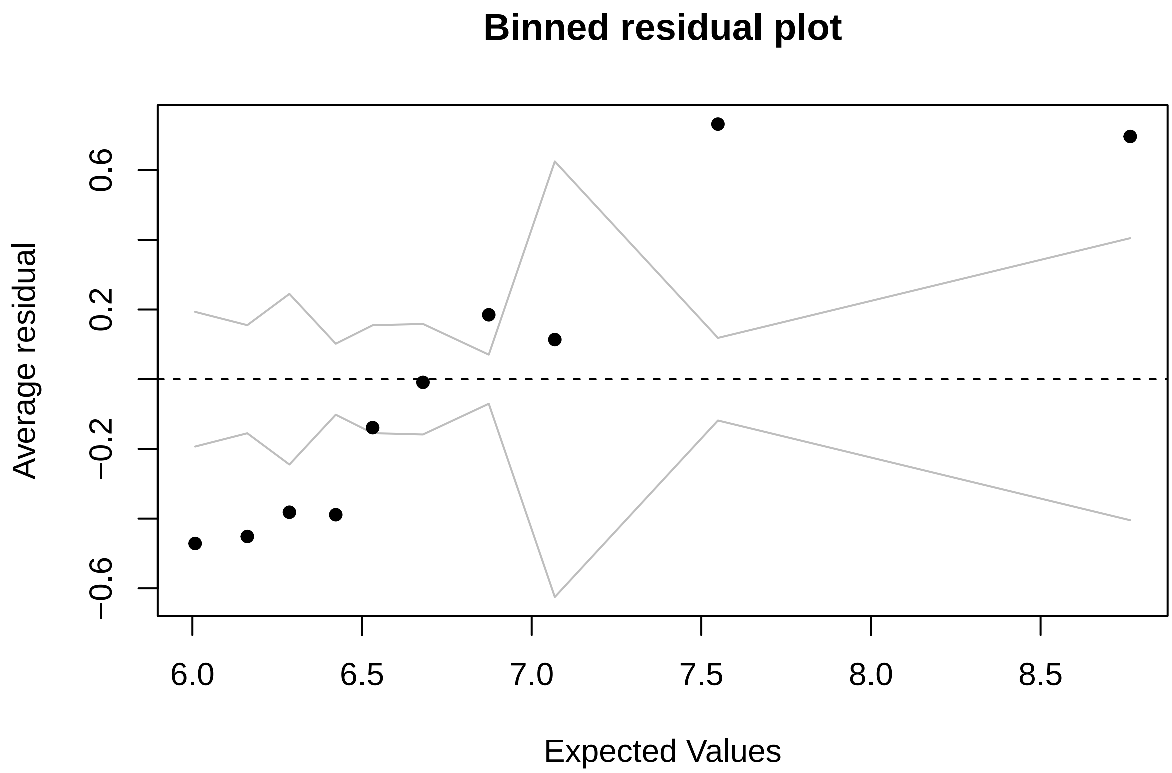
Result

Because the fits of the models were the same within their respective class, it is difficult to determine which one is the most appropriate. Based on the goal of this work, neither complete pooling nor no-pooling modeling is optimal. The information provided by the analysis would ideally be used to identify areas with low house prices due to crime, and as complete pooling ignores variation between areas, the subject of this work would be pooled away. In contrast, no-pooling modeling would overestimate the variation between areas, skewing the results. Thus, multilevel modeling is the most appropriate. To make use of all of the data available and illustrate differences between areas, a varying intercept model with group level predictors should produce the most accurate results. Because of the variation within cities/neighborhoods that contribute to both the overall crime level and the type of crime that is likely to be committed, the model should include crime rate by type, rather than overall crime rate. The results of this model are shown below:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## log_price ~ Percent_White + Vio_100 + Nvio_100 + Fin_100 + Dom_100 +
##     Oth_100 + Drug_100 + MP_100 + Car_100 + (1 | City)
```

```
##    Data: Long2018
##
## REML criterion at convergence: 102
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.93842 -0.65608 -0.05798  0.51185  2.69800
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  City     (Intercept) 0.4734   0.6880
##  Residual             0.3474   0.5894
## Number of obs: 72, groups:  City, 10
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)     19.842     15.822   1.254
## Percent_White  -22.449     28.047  -0.800
## Vio_100          2.628      2.846   0.923
## Nvio_100         6.451      5.837   1.105
## Fin_100         78.920     79.562   0.992
## Dom_100        -49.438     68.339  -0.723
## Oth_100        -62.447     53.895  -1.159
## Drug_100         9.856     12.861   0.766
## MP_100         108.806    137.060   0.794
## Car_100         -5.851      3.623  -1.615
##
## Correlation of Fixed Effects:
##             (Intr) Prcn_W Vi_100 Nv_100 Fn_100 Dm_100 Ot_100 Dr_100 MP_100
## Percent_Wht -0.999
## Vio_100      0.938 -0.940
## Nvio_100     0.963 -0.964  0.927
## Fin_100      0.985 -0.988  0.930  0.938
## Dom_100     -0.992  0.994 -0.941 -0.970 -0.973
## Oth_100     -0.983  0.983 -0.934 -0.961 -0.988  0.965
## Drug_100     0.991 -0.994  0.943  0.959  0.985 -0.992 -0.974
## MP_100       0.989 -0.992  0.932  0.978  0.976 -0.995 -0.976  0.990
## Car_100     -0.443  0.431 -0.509 -0.485 -0.475  0.421  0.487 -0.455 -0.427
```

The intercept shows the expected value, on the logarithmic scale, of a house in an area with no crime and a 0% Caucasian population. This hypothetical house is incredibly expensive, and, regardless of the accuracy of this model, is difficult to believe would ever be a reality. The coefficients of the model indicate that an increase in the Caucasian population, domestic violence, crimes under the “other” category, and car related incidents seem to decrease the house price. Although this is mostly contrary to studies, burglaries are listed under “other”, which is consistent with research. The random effects of this model are incredibly small, indicating little variation between neighborhoods defined in the model.



Residual binned plot of selected model

The residual plots of this model are the same as all others generated for lmer fit models. It is clear that the model does not fit the data well, and that there is an underlying pattern, possibly indicating that the model type is not suitable.

Discussion

As the results of the selected model are dubious, it is difficult to draw meaningful conclusions. If the model is correct, it would indicate that home owners would have a financial incentive to keep their area free of crime, and their property would be a great investment if they lived in an area that became safer over time. This is especially true of domestic crime, which may motivate some people to take a greater interest in the personal lives of their friends and neighbors, as domestic crime often happens behind closed doors. The results of this model also indicate that neighborhoods would be very concerned about their cars and parking violations, as well as miscellaneous crimes listed in the "other" category, including firearm violations, assembly or gathering violations, and burglaries, which is one of the only believable aspects of this model.

There are several limitations to this model. Although the original crime data included hundreds of thousands of crime points for the entire time span, and tens of thousands for 2018, the prices of current houses for sale was the limiting factor in the size of the final data used for the model. This resulted in groups being very small, making it hard to represent the story fully. Another large limitation of this work is the granular level of data that is available. Property prices in some areas vary a substantial amount. However, it is difficult to focus in on the crimes that happened around a single house for which you know the value, and repeat this for a sufficient number of houses in each area, which would be the best way of capturing the relationship between crime and real estate prices. Because of this limitation, the crime rate for each house within an area is the same. Additionally, crime type was subjectively categorized, influencing the basis of the model. In regards to the nature of crime itself, many crimes go unreported, which can have a huge influence on what researches find.

The inclusion of several predictors would greatly improve the accuracy of the model. Points of interest include the hour in which the crime took place, as some crimes are more likely to occur when people are typically not at home; the proportion of employed and/or educated population in the area; and the age distribution of the area, as some crimes strongly follow age profiles. For example, teenagers are more likely to commit acts of vandalism than their older counterparts [14]. To factor in gentrification and determine where crime moves because of it, a time series could be created, with the inclusion of the predictors mentioned above.

Acknowledgement

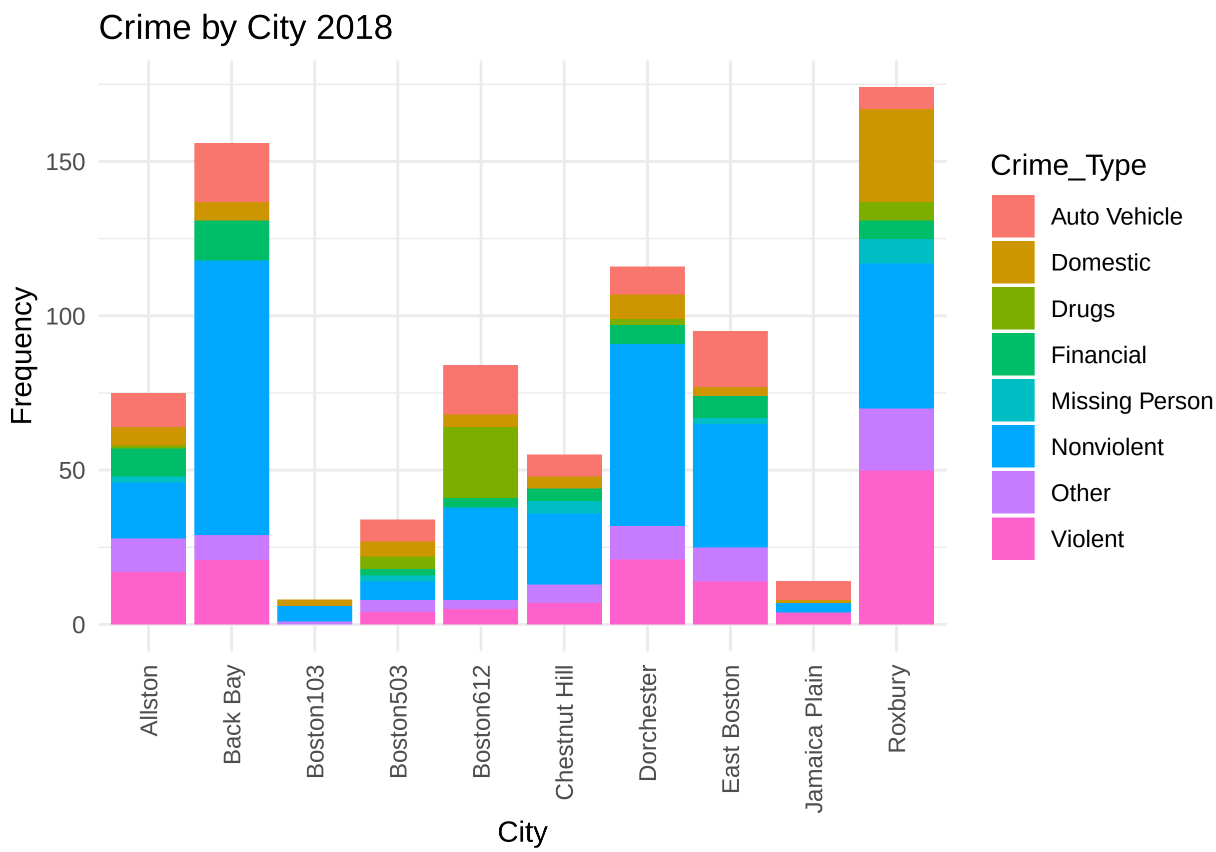
I would like to thank Masanao Yajima for being a patient educator, Haviland Wright for helping me with my code, Andrew Gelman and Jennifer Hill for writing an understandable book on a complicated topic, and Conor and Jackie McPeake for discussing ideas with me.

Reference

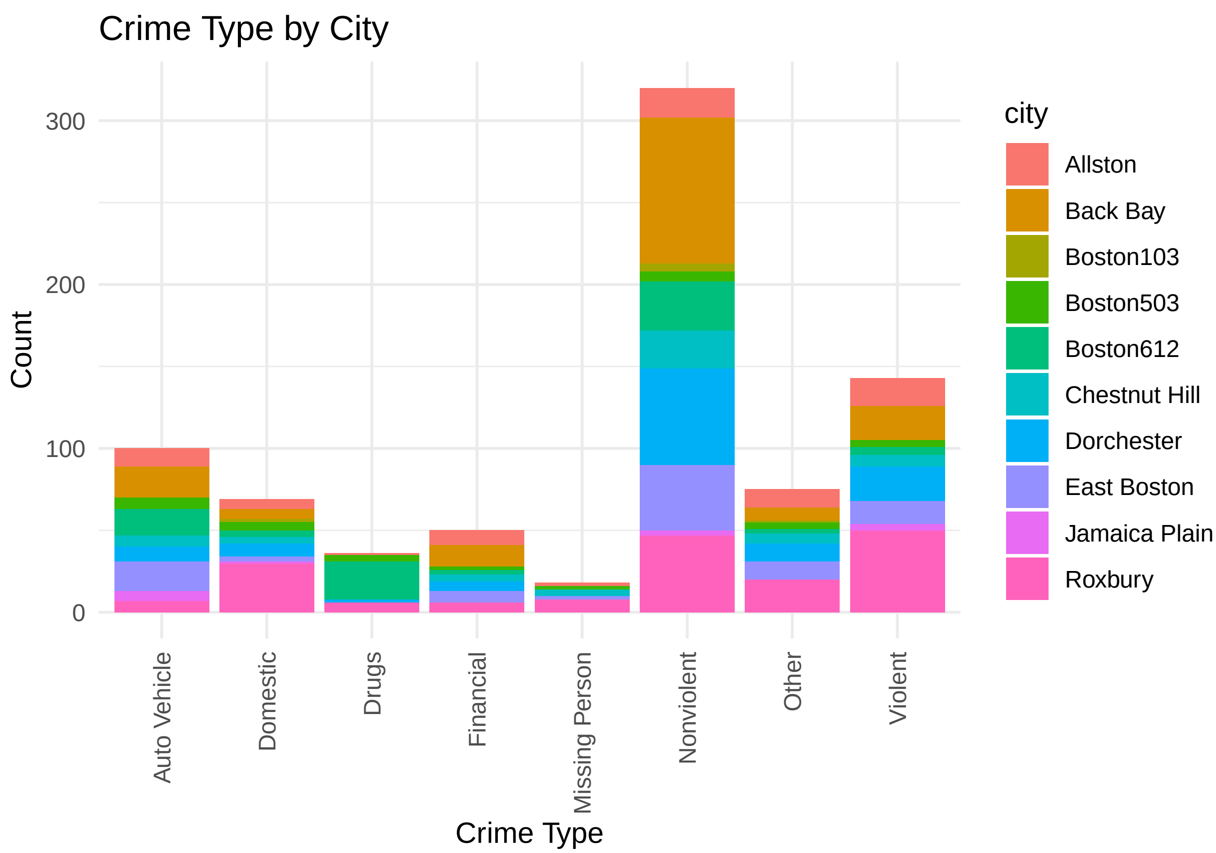
[1] Maximino, Martin. "The Impact of Crime on Property Values: Research Roundup." Journalist's Resource, 16 Feb. 2017, journalistsresource.org/studies/economics/real-estate/the-impact-of-crime-on-property-values-research-roundup. [2] "Crime Incident Reports (August 2015 - To Date) (Source: New System)." Analyze Boston, data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system. [3] "Federal Communications Commission." FCC Area API, geo.fcc.gov/api/census/. [4] "2010 Census - Census Tract Reference Map - Suffolk County, MA." Census Bureau, 2010. 2.census.gov/geo/maps/dc10map/tract/st25_ma/c25025_suffolk/DC10CT_C25025_001.pdf [5] Branch, Geographic Products. "2010 Geographic Terms and Concepts - Census Tract." Census Bureau QuickFacts, United States Census Bureau, 1 Sept. 2012, www.census.gov/geo/reference/gtc/gtc_ct.html. [6] "Allston." Wikipedia, Wikimedia Foundation, 23 Nov. 2018, en.wikipedia.org/wiki/Allston. [7] "Back Bay, Boston." Wikipedia, Wikimedia Foundation, 5 Dec. 2018, en.wikipedia.org/wiki/Back_Bay,_Boston. [8] "Chestnut Hill Demographics and Statistics." Niche, Niche, www.niche.com/places-to-live/n/chestnut-hill-newton-ma/residents/. [9] "Fenway-Kenmore." Wikipedia, Wikimedia Foundation, 5 Nov. 2018, en.wikipedia.org/wiki/Fenway%E2%80%93Kenmore. [10] "Roxbury, Boston." Wikipedia, Wikimedia Foundation, 3 Dec. 2018, en.wikipedia.org/wiki/Roxbury,_Boston. [11] "Dorchester, Boston." Wikipedia, Wikimedia Foundation, 8 Dec. 2018, en.wikipedia.org/wiki/Dorchester,_Boston. [12] "East Boston." Wikipedia, Wikimedia Foundation, 1 Dec. 2018, en.wikipedia.org/wiki/East_Boston. [13] "Jamaica Plain." Wikipedia, Wikimedia Foundation, 23 Nov. 2018, en.wikipedia.org/wiki/Jamaica_Plain. [14] "Who Are the Most Likely Offenders of Graffiti." Western Australia Police, 19 Jan. 2018, www.goodbyegraffiti.wa.gov.au/Schools/Facts-for-Students/Who-are-the-most-likely-offenders-of-graffiti.

Appendix

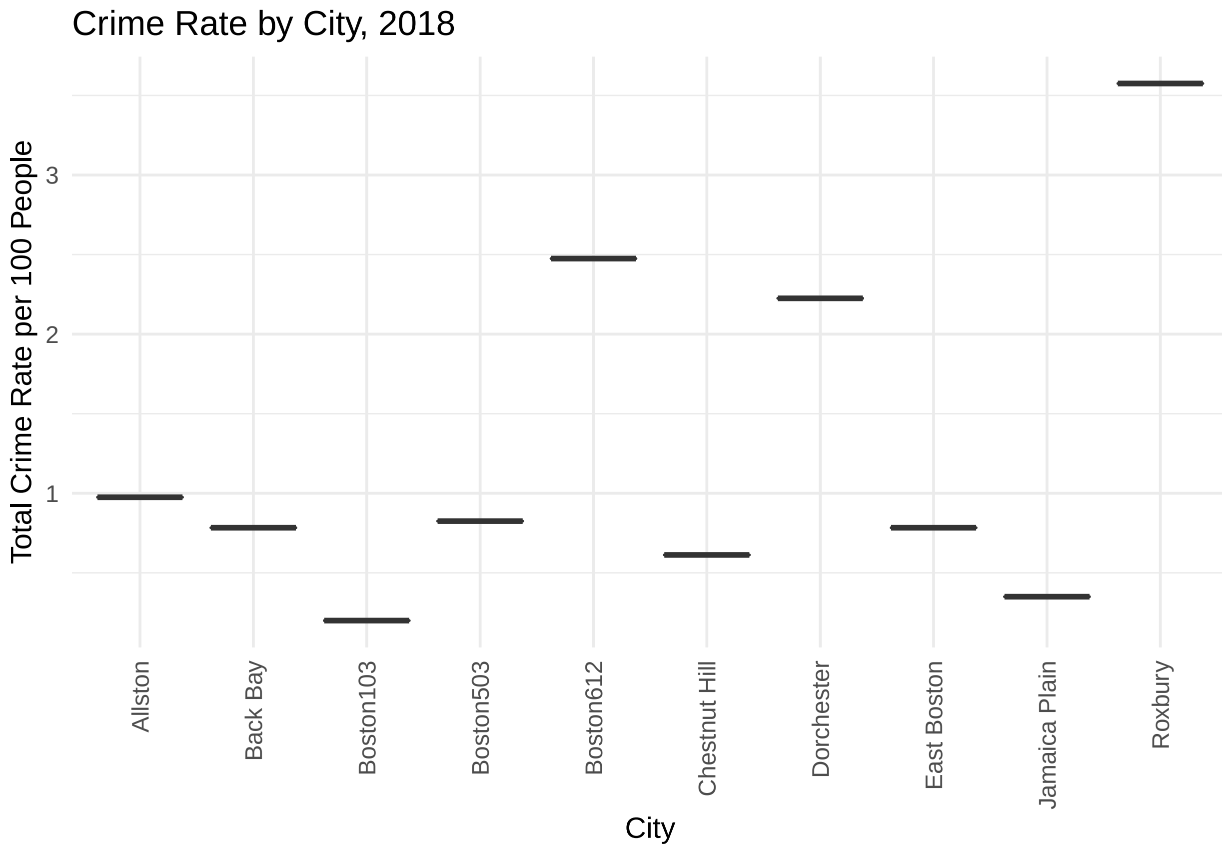
Plots



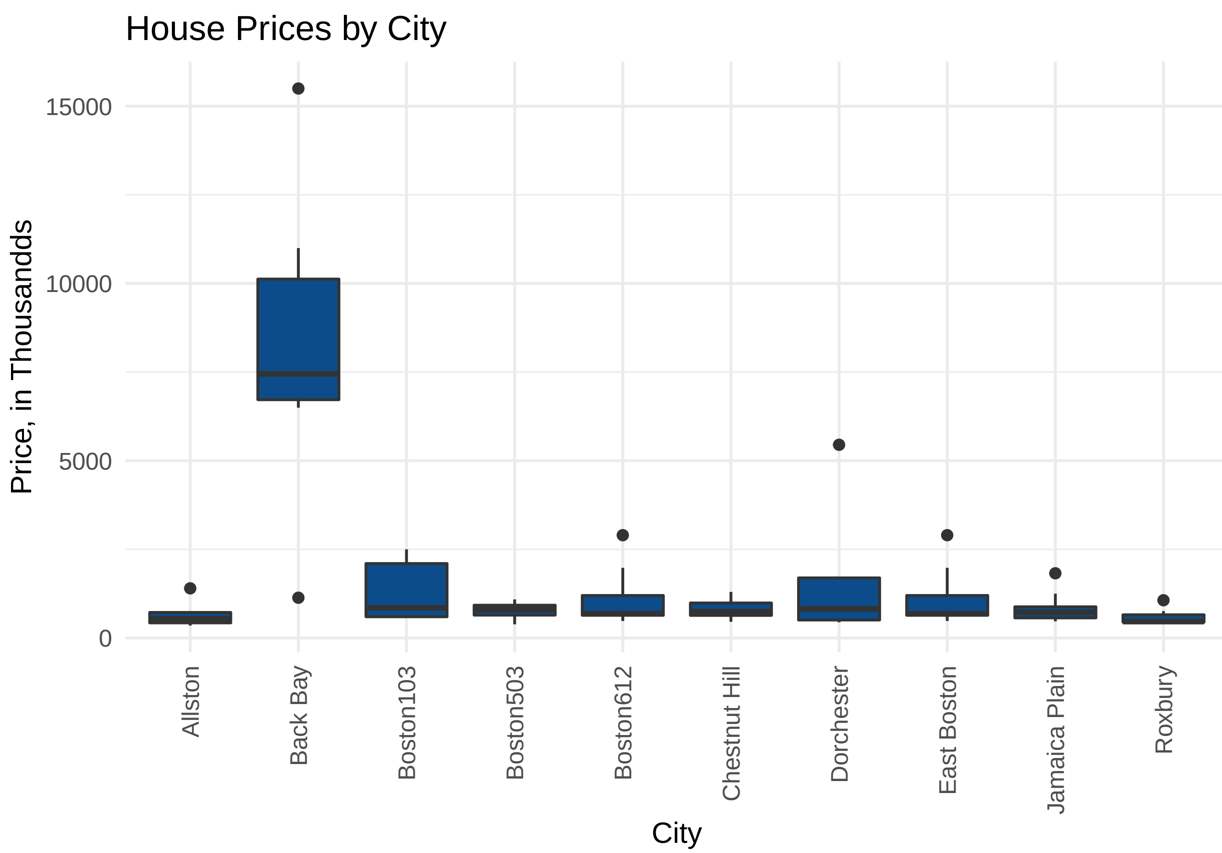
The number of crimes for each city included in the data. Filled in for crime type.



Count of crime type, filled in for each city included in the data.



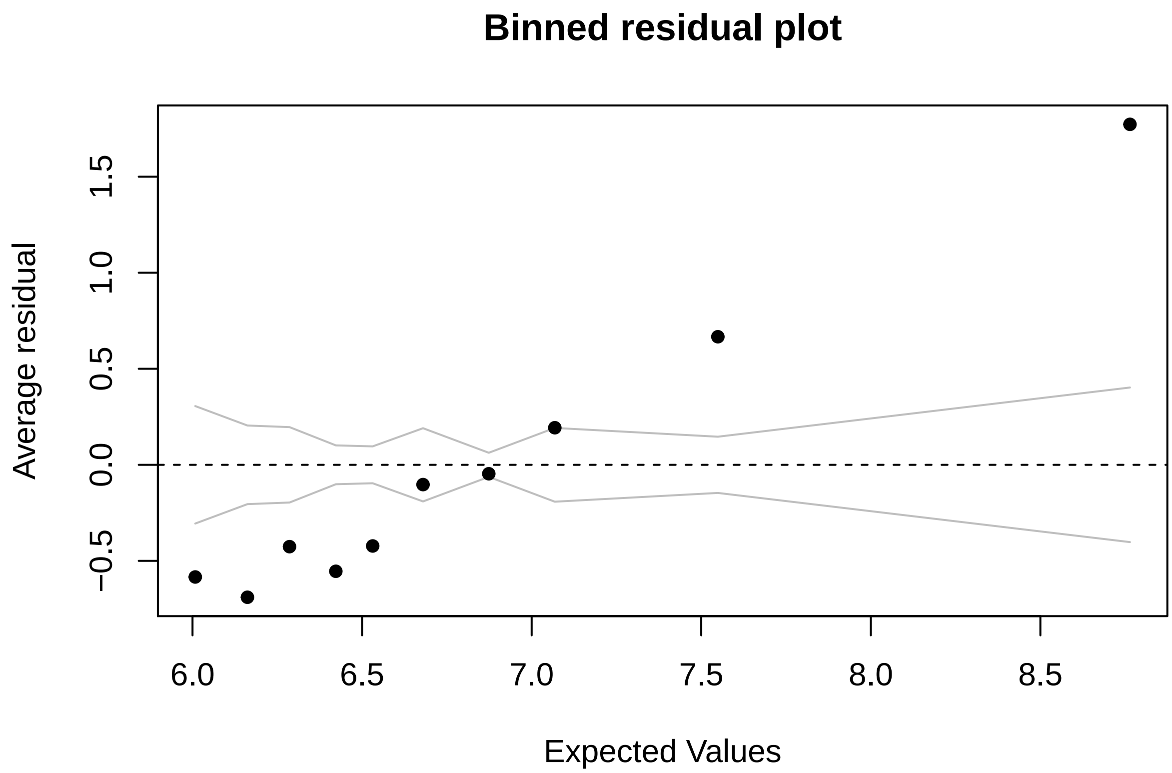
Overall crime rate for each city. As the crime rate is flat for each area, there is only one data point for each group.

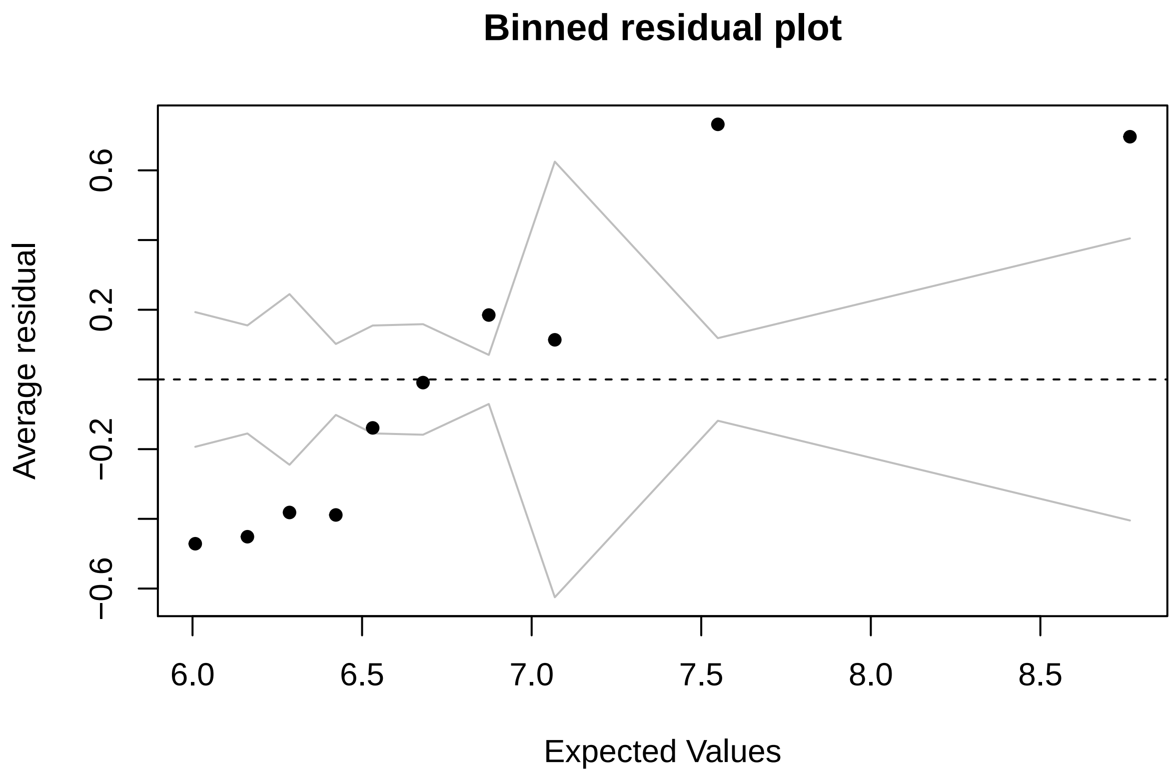


House prices for each city included in the data.

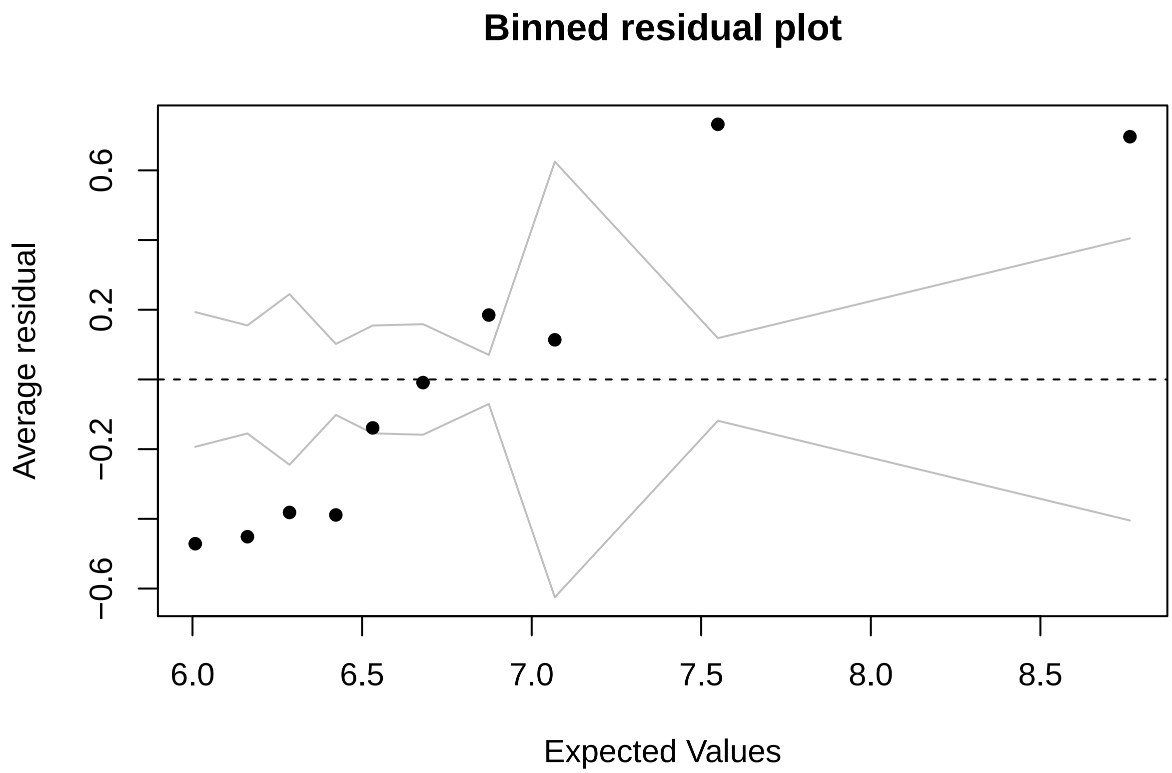
Complete pooling models

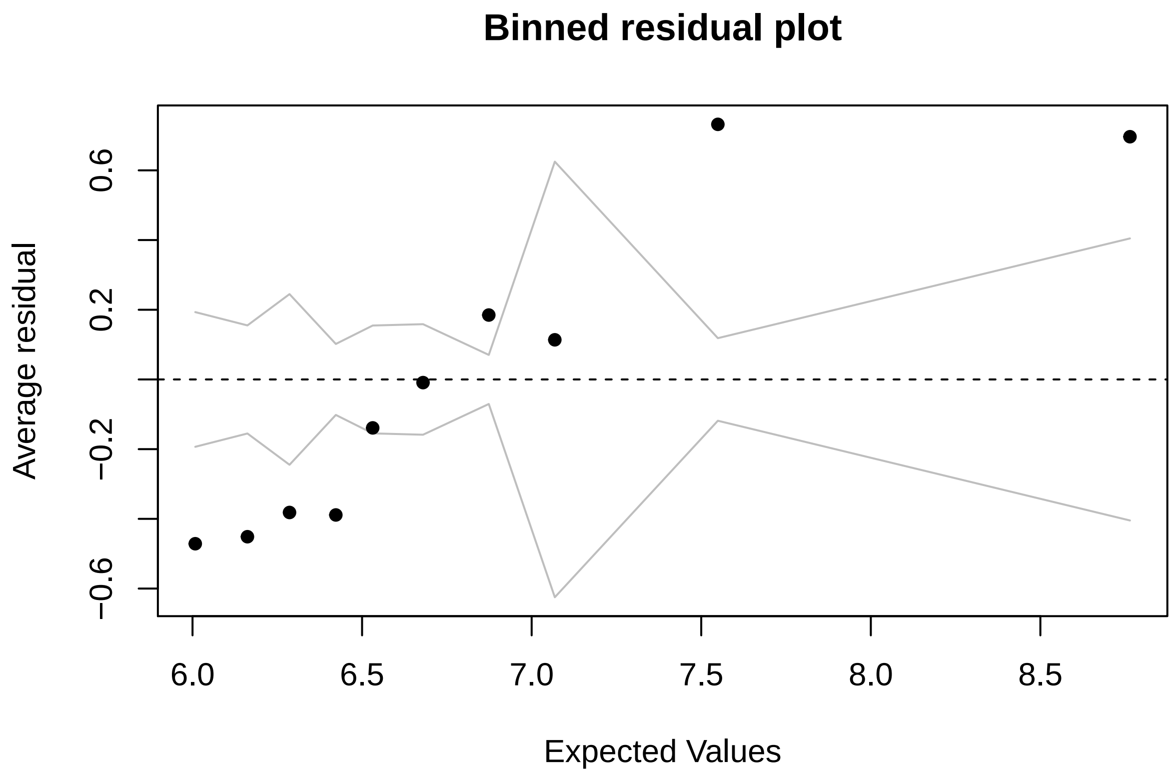
```
##
## Call:
## lm(formula = log_price ~ Crime_100 + Percent_White, data = Long2018)
##
## Coefficients:
##   (Intercept)      Crime_100  Percent_White
##       6.14697        0.01274        1.18216
```

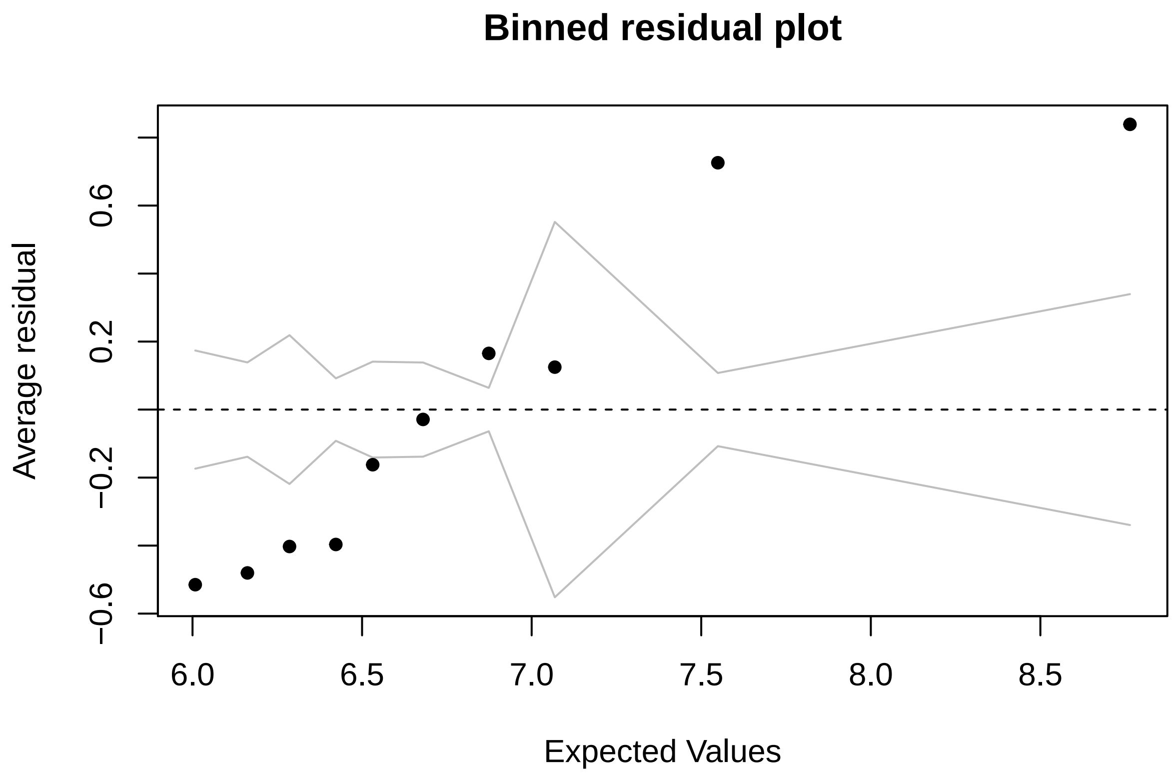

Binned residual plot
Average residual
1.5
1.0
0.5
0.0
−0.5
6.0
6.5
7.0
7.5
8.0
8.5
Expected Values




```
##
## Call:
## lm(formula = log_price ~ Crime_100 + Percent_White + factor(City) -
##     1, data = Long2018)
##
## Coefficients:
##                Crime_100              Percent_White
##                    1.508                     11.607
##       factor(City)Allston       factor(City)Back Bay
##                   -2.382                     -1.364
##     factor(City)Boston103      factor(City)Boston503
##                   -1.211                     -1.905
##     factor(City)Boston612  factor(City)Chestnut Hill
##                   -6.078                     -3.443
##    factor(City)Dorchester    factor(City)East Boston
##                   -0.312                     -1.588
## factor(City)Jamaica Plain       factor(City)Roxbury
##                       NA                         NA
```

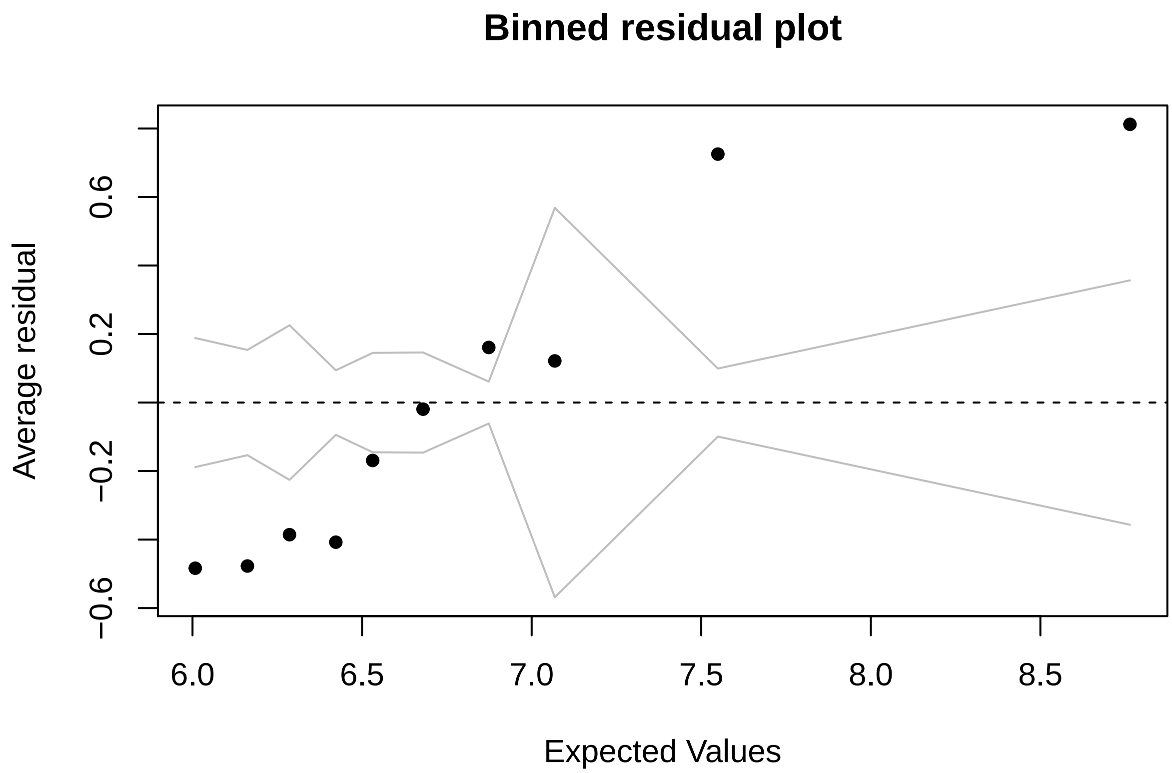

Binned residual plot
Average residual
0.6
0.2
−0.2
−0.6
6.0
6.5
7.0
7.5
8.0
8.5
Expected Values


Binned residual plot
Average residual
0.6
0.2
−0.2
−0.6
6.0
6.5
7.0
7.5
8.0
8.5
Expected Values




```
##   (Intercept)     Crime_100 Percent_White
##     6.1265728     0.0266426     1.2638893
```

```
## $City
##               (Intercept)
## Allston       -0.514517929
## Back Bay       1.463275819
## Boston103      0.001969433
## Boston503     -0.315763796
## Boston612     -0.319459629
## Chestnut Hill -0.436994042
## Dorchester     0.343908418
## East Boston   -0.082875982
## Jamaica Plain -0.146493617
## Roxbury        0.006951326
```




```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## log_price ~ Percent_White + Vio_100 + Nvio_100 + Fin_100 + Dom_100 +
##     Oth_100 + Drug_100 + MP_100 + Car_100 + (1 | City)
##    Data: Long2018
##
## REML criterion at convergence: 102
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.93842 -0.65608 -0.05798  0.51185  2.69800
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  City     (Intercept) 0.4734   0.6880
##  Residual             0.3474   0.5894
## Number of obs: 72, groups:  City, 10
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)     19.842     15.822   1.254
## Percent_White  -22.449     28.047  -0.800
## Vio_100          2.628      2.846   0.923
## Nvio_100         6.451      5.837   1.105
## Fin_100         78.920     79.562   0.992
## Dom_100        -49.438     68.339  -0.723
```

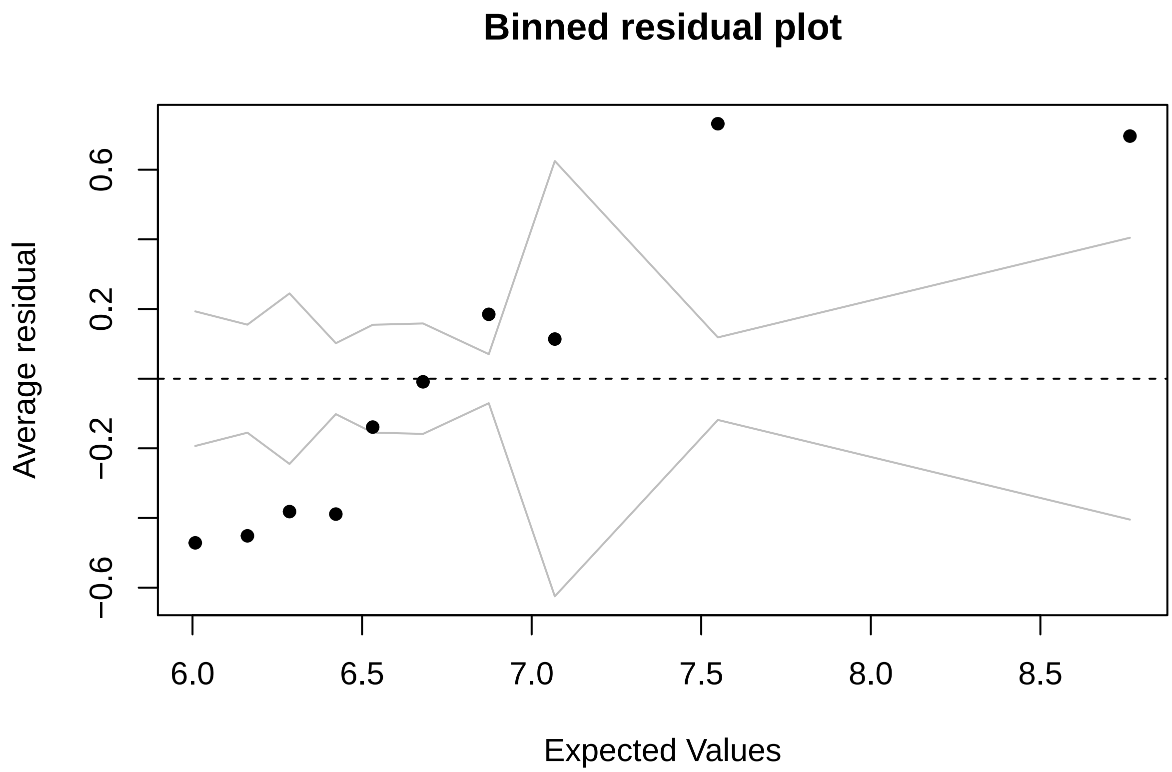
```
## Oth_100        -62.447     53.895  -1.159
## Drug_100         9.856     12.861   0.766
## MP_100         108.806    137.060   0.794
## Car_100         -5.851      3.623  -1.615
##
## Correlation of Fixed Effects:
##             (Intr) Prcn_W Vi_100 Nv_100 Fn_100 Dm_100 Ot_100 Dr_100 MP_100
## Percent_Wht -0.999
## Vio_100      0.938 -0.940
## Nvio_100     0.963 -0.964  0.927
## Fin_100      0.985 -0.988  0.930  0.938
## Dom_100     -0.992  0.994 -0.941 -0.970 -0.973
## Oth_100     -0.983  0.983 -0.934 -0.961 -0.988  0.965
## Drug_100     0.991 -0.994  0.943  0.959  0.985 -0.992 -0.974
## MP_100       0.989 -0.992  0.932  0.978  0.976 -0.995 -0.976  0.990
## Car_100     -0.443  0.431 -0.509 -0.485 -0.475  0.421  0.487 -0.455 -0.427
```

```
## $City
##               (Intercept) Percent_White  Vio_100 Nvio_100 Fin_100
## Allston          19.84217      -22.4489 2.627906 6.451065 78.9197
## Back Bay         19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston103        19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston503        19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston612        19.84217      -22.4489 2.627906 6.451065 78.9197
## Chestnut Hill    19.84217      -22.4489 2.627906 6.451065 78.9197
## Dorchester       19.84217      -22.4489 2.627906 6.451065 78.9197
## East Boston      19.84217      -22.4489 2.627906 6.451065 78.9197
## Jamaica Plain    19.84217      -22.4489 2.627906 6.451065 78.9197
## Roxbury          19.84217      -22.4489 2.627906 6.451065 78.9197
##                 Dom_100   Oth_100 Drug_100   MP_100   Car_100
## Allston       -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Back Bay      -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston103     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston503     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston612     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Chestnut Hill -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Dorchester    -49.43812 -62.44736 9.855846 108.8059 -5.851264
## East Boston   -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Jamaica Plain -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Roxbury       -49.43812 -62.44736 9.855846 108.8059 -5.851264
##
## attr(,"class")
## [1] "coef.mer"
```

```
##   (Intercept) Percent_White       Vio_100      Nvio_100       Fin_100
##     19.842166    -22.448896      2.627906      6.451065     78.919702
##       Dom_100       Oth_100      Drug_100        MP_100       Car_100
##    -49.438116    -62.447356      9.855846    108.805905     -5.851264
```

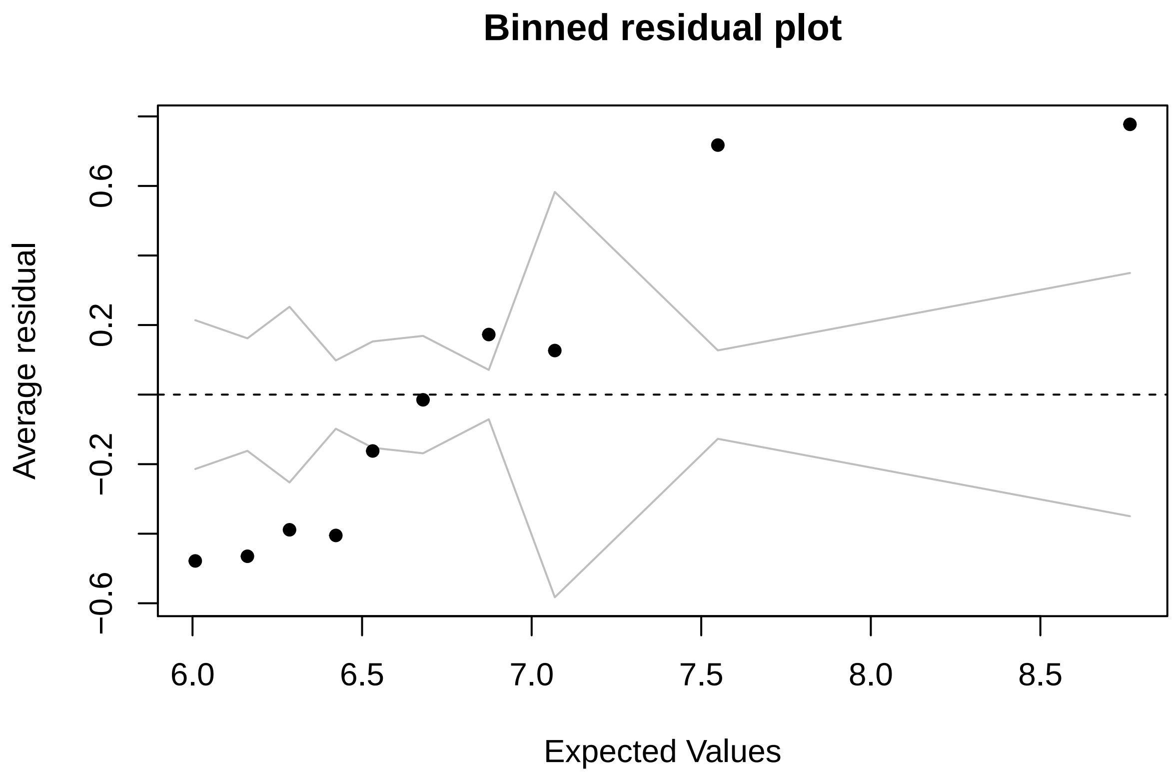
```
## $City
##                 (Intercept)
## Allston        7.730172e-12
## Back Bay      -8.525247e-12
## Boston103     -3.070450e-12
## Boston503     -4.964861e-13
```

```
## Boston612      2.080445e-12
## Chestnut Hill  9.050973e-12
## Dorchester     4.310054e-12
## East Boston   -1.636002e-11
## Jamaica Plain  8.218530e-12
## Roxbury       -3.115192e-12
```

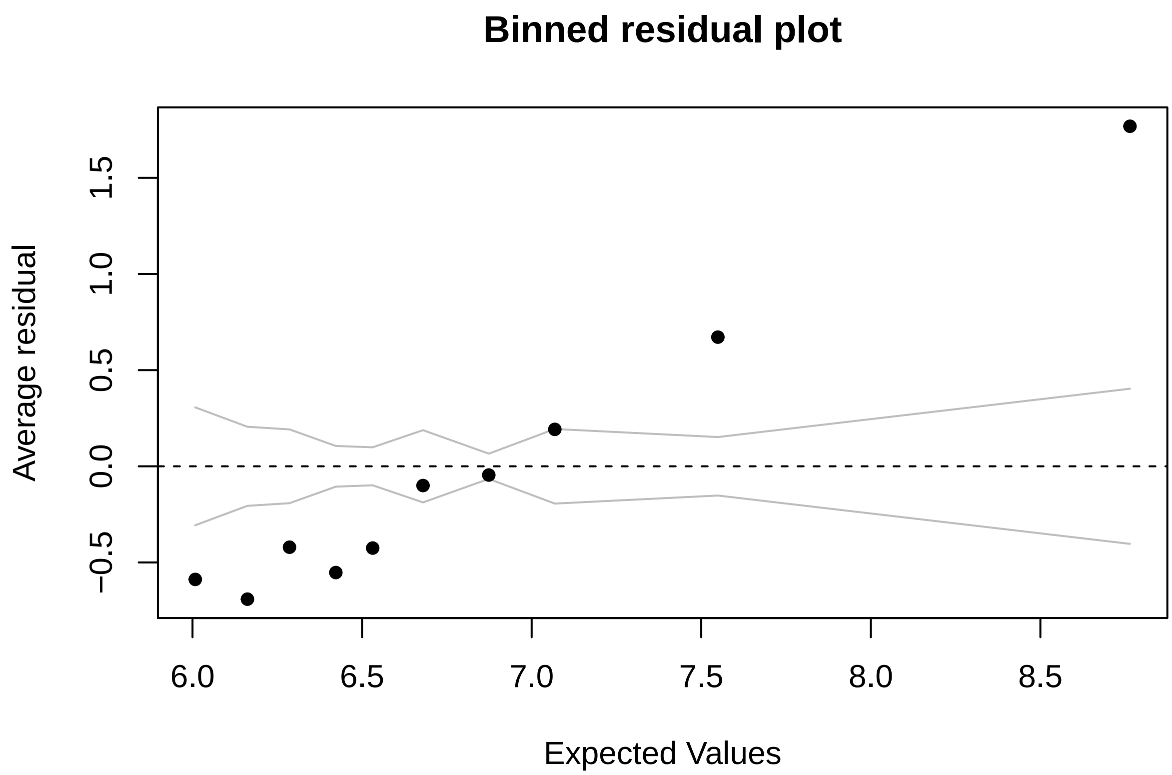



```
##   (Intercept) Percent_White       Vio_100        MP_100      Nvio_100
##     5.9166899     1.2769123    -0.0243061    -3.5015519     0.6976470
```

```
## $City
##                (Intercept)
## Allston       -0.367135902
## Back Bay       1.204337092
## Boston103      0.105805280
## Boston503     -0.022192686
## Boston612     -0.555102559
## Chestnut Hill -0.265967343
## Dorchester    -0.310690450
## East Boston   -0.035555854
## Jamaica Plain  0.004712455
## Roxbury        0.241789966
```


Binned residual plot
Average residual
Expected Values
0.6
0.2
−0.2
−0.6
6.0
6.5
7.0
7.5
8.0
8.5

Try another type of transformation




```
## $City
##                (Intercept) Percent_White  Vio_100 Nvio_100 Fin_100
## Allston           19.84217      -22.4489 2.627906 6.451065 78.9197
## Back Bay          19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston103         19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston503         19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston612         19.84217      -22.4489 2.627906 6.451065 78.9197
## Chestnut Hill     19.84217      -22.4489 2.627906 6.451065 78.9197
## Dorchester        19.84217      -22.4489 2.627906 6.451065 78.9197
## East Boston       19.84217      -22.4489 2.627906 6.451065 78.9197
## Jamaica Plain     19.84217      -22.4489 2.627906 6.451065 78.9197
## Roxbury           19.84217      -22.4489 2.627906 6.451065 78.9197
##                  Dom_100   Oth_100 Drug_100   MP_100   Car_100
## Allston        -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Back Bay       -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston103      -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston503      -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston612      -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Chestnut Hill -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Dorchester     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## East Boston    -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Jamaica Plain -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Roxbury        -49.43812 -62.44736 9.855846 108.8059 -5.851264
##
```

```
## attr(,"class")
## [1] "coef.mer"
```

```
## $City
##               (Intercept) Percent_White  Vio_100 Nvio_100 Fin_100
## Allston          19.84217      -22.4489 2.627906 6.451065 78.9197
## Back Bay         19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston103        19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston503        19.84217      -22.4489 2.627906 6.451065 78.9197
## Boston612        19.84217      -22.4489 2.627906 6.451065 78.9197
## Chestnut Hill    19.84217      -22.4489 2.627906 6.451065 78.9197
## Dorchester       19.84217      -22.4489 2.627906 6.451065 78.9197
## East Boston      19.84217      -22.4489 2.627906 6.451065 78.9197
## Jamaica Plain    19.84217      -22.4489 2.627906 6.451065 78.9197
## Roxbury          19.84217      -22.4489 2.627906 6.451065 78.9197
##                 Dom_100   Oth_100 Drug_100   MP_100   Car_100
## Allston       -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Back Bay      -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston103     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston503     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Boston612     -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Chestnut Hill -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Dorchester    -49.43812 -62.44736 9.855846 108.8059 -5.851264
## East Boston   -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Jamaica Plain -49.43812 -62.44736 9.855846 108.8059 -5.851264
## Roxbury       -49.43812 -62.44736 9.855846 108.8059 -5.851264
##
## attr(,"class")
## [1] "coef.mer"
```