A Statistical Study of Single Family Home Sales in Atlanta

Table of Contents

1. Introduction

This study focuses on understanding single family residential home sales in certain areas of metro Atlanta, specifically Cobb County and North Fulton County. The key goals of the study is to build a model that will predict the appropriate list price for the sale of a single family home in these areas and, further, to estimate what the sales price likely will be. We apply simple and multiple regression analysis to help us to understand the contributions of the variables we have collected in our sample data set.

2. Description of Data

An Atlanta Realtor with whom we consulted indicated that in the First Multiple Listing Service (FMLS) as of March 9, 2013, the number of Cobb county homes for sale or shown as sold was 4,108 and the number of N. Fulton county homes for sale or shown as sold was 1,474. From these county populations, two samples comprising information on 100 homes in Cobb county and 100 homes in North Fulton county were provided by the Realtor based on current listing. After consulting the realtor, we selected the following predictors (adjusted as noted below from the raw FMLS data) with which to model single family home list and sales prices:

- *Size* = floor size (thousands of square feet)
- *Lot* = lot size category (see below)
- *Bath* = number of bathrooms (between 1 and 6, with half-baths counting as 0.2)
- *Bed* = number of bedrooms (between 2 and 6)
- *Basement* = basement status (0 = crawlspace or half, 1 = unfinished full, 2 = finished full)
- *Age* = age of home since originally built (number of decades since 1980)
- *Garage* = garage size (0, 1, 2, or 3 cars)
- *Brick* = brick siding construction
- *Frame* = wood, vinyl or aluminum siding construction
- *Stucco* = Stucco or concrete siding construction
- *SchoolQuality* = indicator of quality of school (see below)

The Realtor informed us that lot size is best viewed in terms of categories because lot size does not have a linear impact on price and using actual lot size would not produce as realistic an effect on price. For example, increasing a lot size by 1,000 would have a greater effect on price where the lot size increases from 4,000 to 5,000 square feet than from 19,000 to 20,000 square feet. The categories used for lot size are as follows:

Lot Size (acres)	<1/3	1/3 to 1/2	1/2 to 3/4	3/4 to 1	1 to 2	2 to 3	3 to 5
Category	1	2	3	4	5	6	7

Moreover, the number of baths may not be a whole number due to the presence of half-baths. The Realtor informed us that a half-bath typically is not valued only marginally by buyers and not literally as one half of a whole bath. For this reason we assign the value of 0.2 to a half-bath rather than 0.5. The Realtor also believes, based on her experience, that the homes in the Cobb County and North Fulton areas were built largely from the

1950s through 2012 with a mean around 1980. Like lot size, the age of the home is best reflected in categories, and a simple and convenient categorization is to define age in terms of decades from the mean. Hence, the predictor *Age* is determined by the subtracting 1980 from the year a home was originally built and dividing by 10.

The Realtor further informed us that in her experience the quality of the schools in the district to which the home belongs can have a significant influence on price. We account for quality of the schools associated with each home as categories of the average rankings assigned to the elementary and high schools based on data from the Georgia Public Education Report Card as provided by the Georgia Department of Education. These categories are as follows for the variable *SchoolQuality*: 1 = tier 1 (schools of excellence), 2 = tier 2 (high quality), 3 = tier 3 (average quality), and 4 = 4th tier (below average quality).

3. Overview of Predictors

Using Minitab, we determined the summary statistics for each predictor and performed a simple regression analysis to explore the linear relationship between each predictor and ListPrice, as summarized below in Tables 1 and 2. Based on this analysis, only some of the predictors appear to have a linear relationship with ListPrice and, thus, a significant influence on its variation. In Cobb county, the predictors Garage and Stucco appear insignificant with respect to ListPrice (having Pvalues of the t-test for $\hat{\beta}_1$ greater than 0.05). In N. Fulton county, the predictors Basement, Frame and Stucco appear insignificant.

Table 1. Summary of Analysis of Predictors for the Cobb County Sample

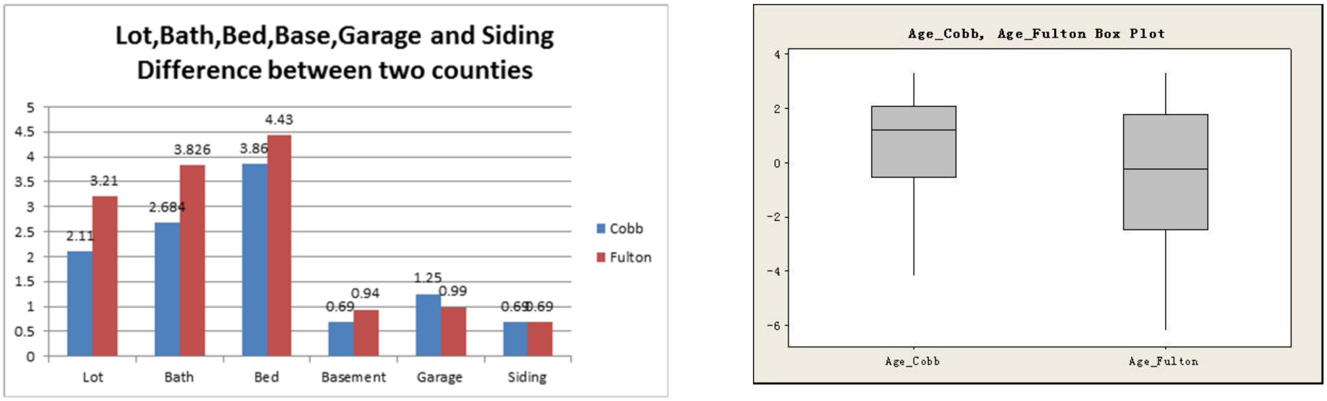
	Size	Lot	Age	Bed	Bath	Basement	Garage	Brick	Frame	Stucco	SchQlty
Mean	2708	2.11	0.729	3.86	2.684	0.69	1.25	0.57	0.30	0.13	3.23
Std Err	1277	1.392	1.836	0.9849	1.090	0.8002	0.9468	0.4976	0.4606	0.3380	0.7921
Min	768	0	-4.2	2	1	0	0	0	0	0	1
Max	6620	6	3.3	6	6	2	2	1	1	1	4
$\hat{\beta}_0$	-51074	176379	235266	-58268	-46471	201935	273592	210156	305041	62598	636128
$\hat{\beta}_1$	115.734	42137	37068	119432	116155	91816	-9043	91459	-142510	55269	-115740
T_0	15.17	3.29	4.14	10.80	9.18	4.25	-0.48	2.64	-3.97	1.14	-7.07
T pvalue	0.000	0.001	0.000	0.000	0.000	0.000	0.632	0.010	0.000	0.257	0.000
R^2	0.632	0.099	0.149	0.543	0.462	0.156	0.002	0.051	0.138	0.013	0.337
R^2 (adj)	0.629	0.090	0.140	0.539	0.457	0.147	0.000	0.041	0.129	0.003	0.338
SS_E	1.26E+12	3.08E+12	3.04E+12	2.38E+12	1.84E+12	2.90E+12	3.41E+12	3.25E+12	3.69E+12	3.39E+12	3.03E+12
MS_E	1.28E+10	3.15E+10	3.11E+10	2.43E+10	1.88E+10	2.96E+10	3.48E+10	3.32E+10	3.12E+10	3.45E+10	3.09E+10
F_0	168.57	10.80	17.10	42.92	116.67	18.08	0.45	5.27	15.72	1.30	49.91
F pvalue	0.000	0.001	0.000	0.000	0.000	0.000	0.632	0.010	0.000	0.257	0.000

Table 2. Summary of Analysis of Predictors for the N. Fulton County Sample

Predictor	Size	Lot	Age	Bed	Bath	Basement	Garage	Brick	Frame	Stucco	SchQlty
Mean	4083	3.21	-0.523	4.43	3.826	0.94	0.99	0.71	0.09	0.20	1.84
Std Err	2373	1.659	2.472	1.183	1.620	0.8969	0.9692	0.4560	0.2876	0.4020	0.8958
Min	1176	0	-6.2	2	2.0	0	0	0	0	0	1
Max	12705	7	3.3	7	10.2	2	2	1	1	1	4
$\hat{\beta}_0$	60920	346277	922323	-301343	-125712	783577	1080573	924212	844431	54599	1574277
$\hat{\beta}_1$	205.24	168162	93620	267731	266454	101877	-199229	-448612	174599	256728	-265750
T_0	11.63	5.07	4.05	6.15	7.75	1.52	-3.39	-2.31	-1.52	1.18	--7.22
T pvalue	0.000	0.000	0.000	0.000	0.000	0.132	0.001	0.023	0.132	0.243	0.000
R²	0.58	0.208	0.143	0.278	0.484	0.023	0.105	0.052	0.023	0.010	0.347
R² (adj)	0.576	0.200	0.134	0.271	0.479	0.013	0.096	0.042	0.013	0.004	0.340
SS_E	5.15E+13	8.11E+13	1.03E+14	8.56E+13	6.41E+13	1.033E+14	9.59E+13	1.02E+14	1.01E+14	1.03E+14	8.53E+13
MS_E	5.26E+11	8.28E+11	1.05E+12	8.74E+11	6.54E+11	1.05E+12	9.78E+11	1.05E+12	1.03E+12	1.05E+12	8.70E+11
F_0	135.22	25.74	16.37	37.78	91.94	2.30	11.52	5.33	2.31	1.38	52.07
F pvalue	0.000	0.000	0.000	0.000	0.000	0.132	0.001	0.023	0.132	0.243	0.000

4. The Initial Multiple Regression Model for List Price

As an initial matter, we first determined whether the county made a significant difference in *ListPrice*. Using the combined sample data, we ran a simple regression analysis in Minitab for both the Cobb variable (indicating that a home is located in Cobb county) and the N. Fulton variable (indicating that a home is located in N. Fulton county). For the t-test of the $\hat{\beta}_1$ coefficient we obtained P-values of 0.000 for the coefficients associated with each county, indicating that the identity of the county is significant. This is consistent with the observation that the means of the predictors differ by county, as shown in Figure 1. Based on this, we conducted further testing separately on the sample data for each county rather than as an aggregate sample.



(a) **(b)**

Figure 1. Comparison of Predictors by County.

We ran an initial multiple regression analysis in Minitab using all predictors showing t-test P-values for $\hat{\beta}_j$ less than 0.05 for the two counties as shown below. These initial results suggested that only some of the predictors have a significant influence on *ListPrice* and these vary by county. For Cobb county, the predictors Size, Lot and SchoolQuality are shown to be significant. For N. Fulton county, the predictors Size and SchoolQuality are significant.

Cobb County:

```
ListPrice = 184122 + 98.4 Size + 21188 Lot - 1518 Age - 25167 Bed + 31293 Bath
            - 22513 Basement - 27683 Brick - 46145 Frame - 53755 SchoolQuality

Predictor         Coef  SE Coef      T      P
Constant        184122    66403   2.77  0.007
Size             98.43    13.30   7.40  0.000
Lot              21188     5869   3.61  0.001
Age              -1518     5636  -0.27  0.788
Bed             -25167    14714  -1.71  0.091
Bath             31293    16302   1.92  0.058
Basement        -22513    12576  -1.79  0.077
Brick           -27683    24691  -1.12  0.265
Frame           -46145    27559  -1.67  0.098
SchoolQuality   -53755     9670  -5.56  0.000

S = 76006.6   R-Sq = 83.1%   R-Sq(adj) = 81.5%
```

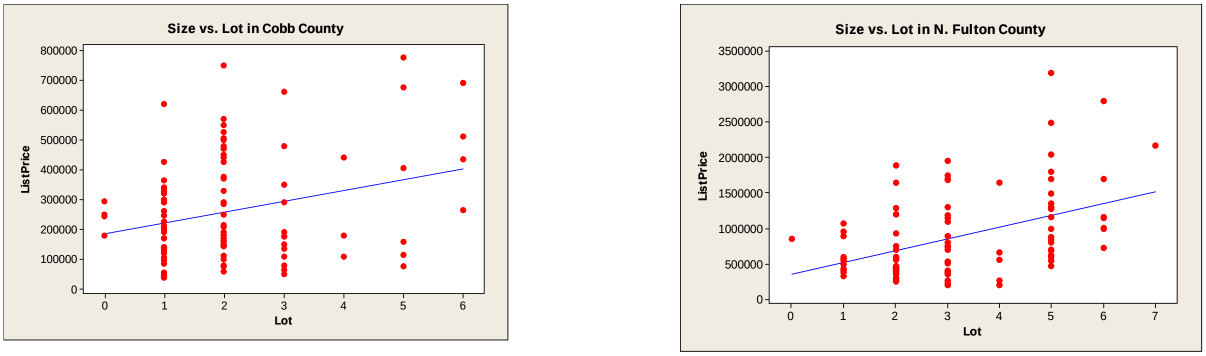
N. Fulton County:

```
ListPrice = 623994 + 144 Size + 22786 Lot + 6519 Age - 19392 Bed + 30149 Bath
            - 83039 Garage + 19550 Brick - 35408 Frame - 180790 SchoolQuality

Predictor          Coef  SE Coef      T      P
Constant         623994   210277   2.97  0.004
Size             144.23    30.91   4.67  0.000
Lot               22786    27231   0.84  0.405
Age                6519    19144   0.34  0.734
Bed              -19392    47674  -0.41  0.685
Bath              30149    50511   0.60  0.552
Garage           -83039    40097  -2.07  0.051
Brick             19550    91708   0.21  0.832
Frame            -35408   150590  -0.24  0.815
SchoolQuality   -180790    42285  -4.28  0.000

S = 341786   R-Sq = 70.0%   R-Sq(adj) = 67.0%
```

The regression equation for Cobb county differs from N. Fulton county by showing that Lot is significant. As suggested in Figure 2, Size is more correlated to Lot in N. Fulton county than in Cobb county, which helps to explain the difference. Further, Size is sufficiently correlated to Lot, Bed and Bath, as shown in Figures 3 and 4, which helps to explain why Size is significant to the regression equation for ListPrice while Bed and Bath (and Lot for N. Fulton county only) are not despite being linearly related to ListPrice as indicated earlier by the simple regression analysis. Likewise, as shown in Figures 5 and 6, Size appears sufficiently correlated to Age, Basement and Garage to explain why they are not significant for the multiple regression model. Essentially, the square footage size of homes has increased over time, particularly since 1980, and as homes have gotten larger the number of bedrooms and baths have increased in step. Likewise, over time, homes have tended to increase the capacity of the garage to hold two cars and to have full basements. In the decades prior to 1980, for example, the typical N. Fulton county home size was less than 2,500 square feet with two and three bedrooms being common, but in the period following 1980, the typical home grew well past 2,500 square feet with very few two bedroom homes built with the number of bedrooms increasing to the range of 4 to 6.



(a) **(b)**

Figure 2. Scatterplots of Size vs. Lot in Cobb (a) and N. Fulton (b) counties.

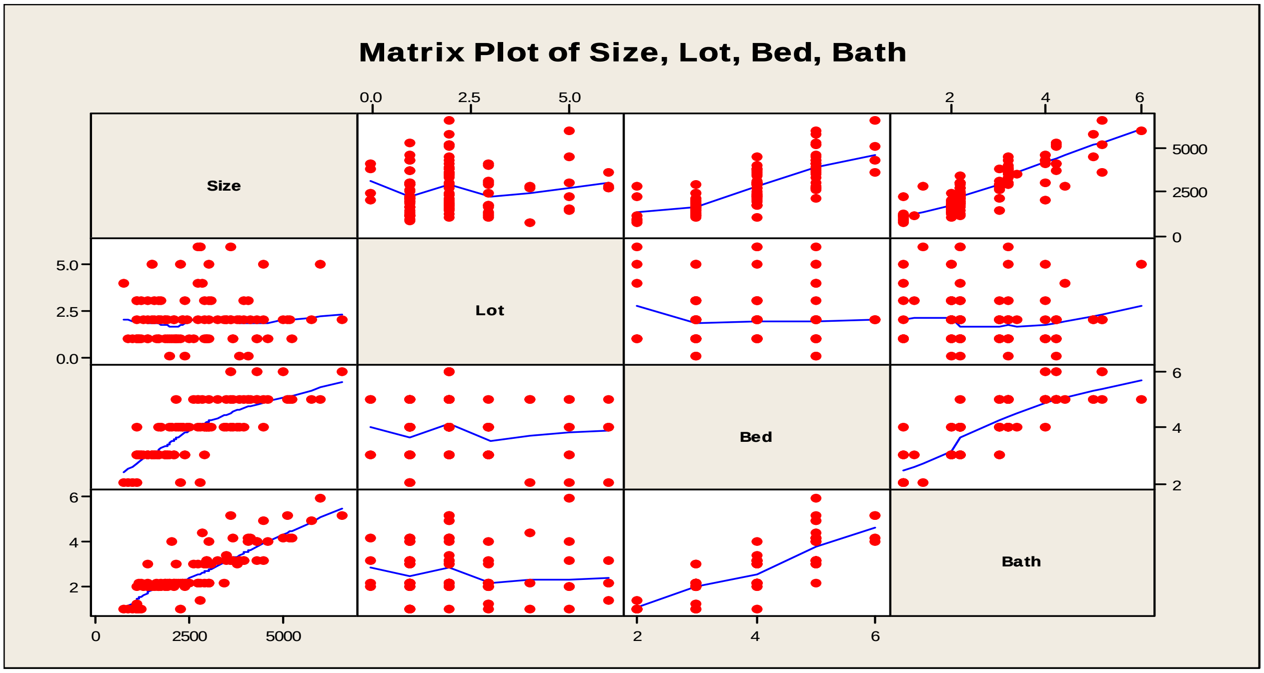


Figure 3. Scatterplots of Size vs. Lot, Bed and Bath in Cobb County.

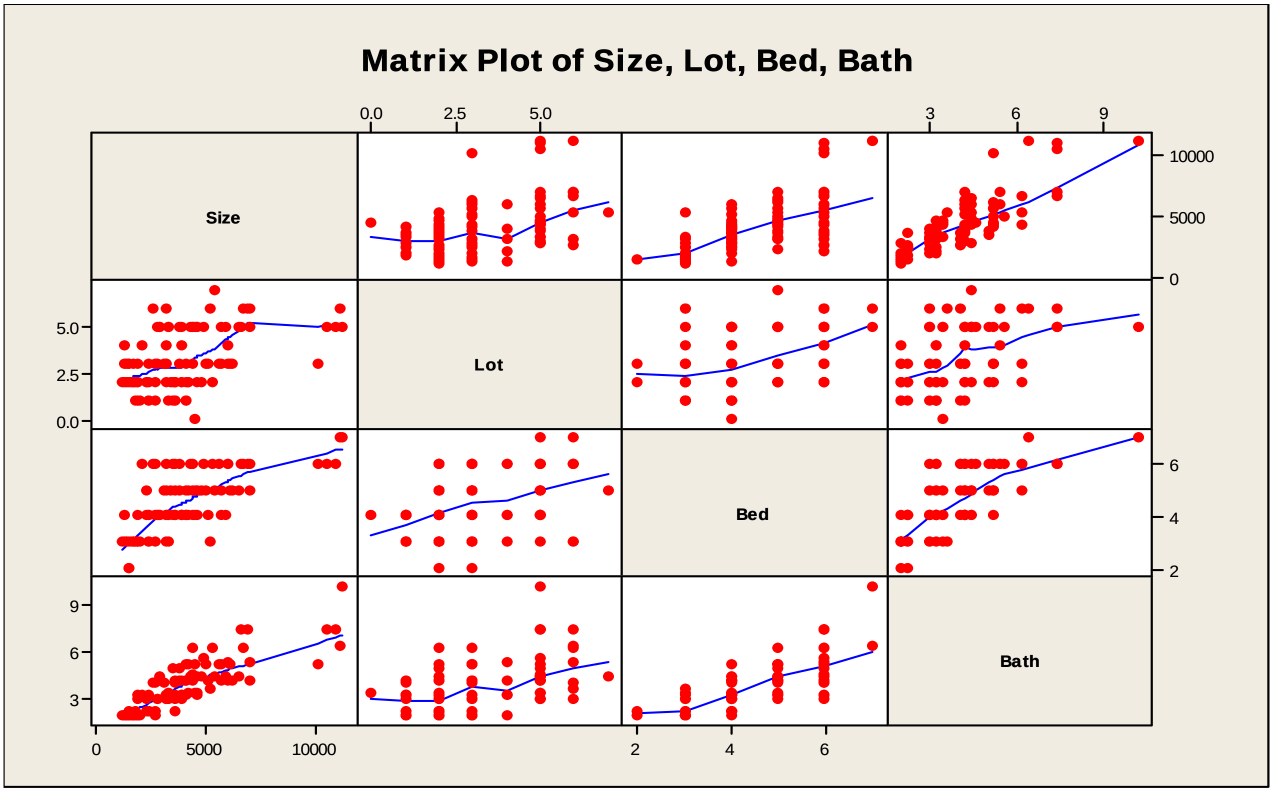


Figure 4. Scatterplots of Size vs. Size vs. Lot, Bed and Bath in N. Fulton County.

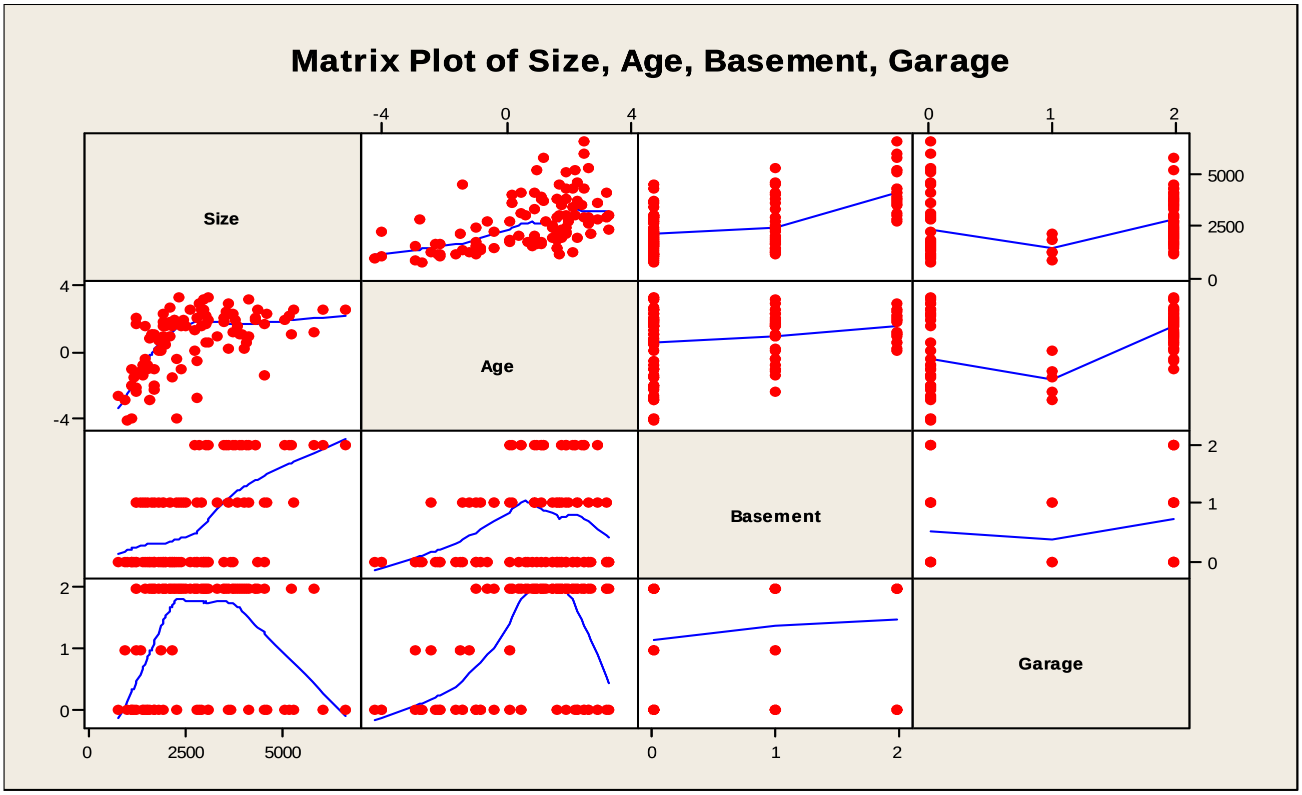


Figure 5. Scatterplots of Size vs. Age, Basement and Garage in Cobb County.

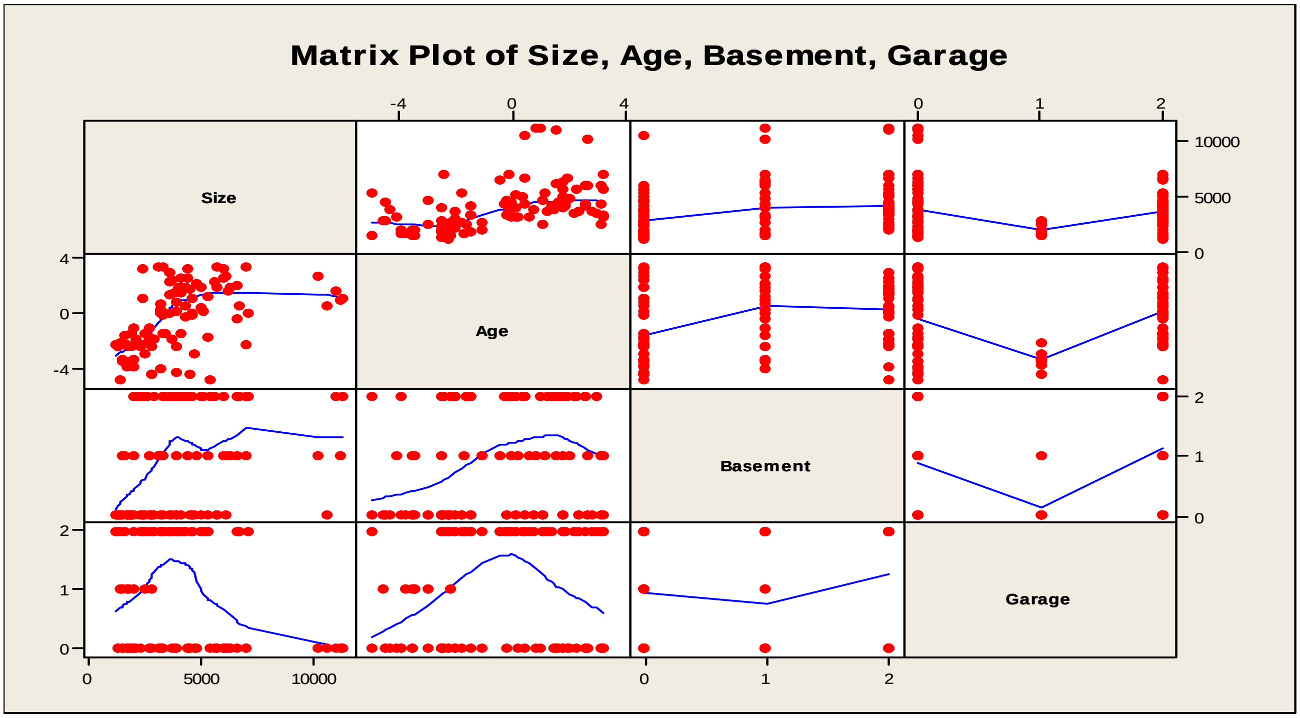


Figure 6. Scatterplots of Size vs. Age, Basement and Garage in N. Fulton County.

5. The Refined Multiple Regression Model for List Price

Removing the insignificant predictors from the initial multiple regression models provides us with the refined models for each county, as shown in the following Minitab output.

Cobb County:

```
ListPrice = 117318 + 100 Size + 22110 Lot - 53474 SchoolQuality

Predictor          Coef  SE Coef      T      P
Constant         117318    45947   2.55  0.012
Size            100.103    6.602  15.16  0.000
Lot               22110     5760   3.84  0.000
SchoolQuality    -53474     9747  -5.49  0.000

S = 77694.6   R-Sq = 81.2%   R-Sq(adj) = 80.6%

Analysis of Variance

Source          DF           SS           MS       F      P
Regression       3  2.50530E+12  8.35099E+11  138.34  0.000
Residual Error  96  5.79499E+11   6036443842
Total           99  3.08480E+12
```

N. Fulton County:

```
ListPrice = 609925 + 169 Size - 212740 SchoolQuality

Predictor          Coef  SE Coef      T      P
```

```
Constant        609925   122612   4.97  0.000
Size            168.92    16.86  10.02  0.000
SchoolQuality  -212740    38847  -5.48  0.000

S = 340726   R-Sq = 67.9%   R-Sq(adj) = 67.2%

Analysis of Variance

Source          DF           SS           MS       F      P
Regression       2  2.38234E+13  1.19117E+13  102.60  0.000
Residual Error  97  1.12611E+13  1.16094E+11
Total           99  3.50845E+13
```

6. Adequacy of the Refined Regression Model for List Price

The R^2 (adj) value for the Cobb county regression equation indicates that 80.6% of variability in the value of ListPrice can be explained by the predictors Size, Lot and SchoolQuality. This appears to be a surprisingly good result given the simplicity of the model and the subjectivity involved in the judgment of the seller and listing agent in pricing a home for sale. The R^2 (adj) value for the N. Fulton county regression equation indicates that 67.2% of variability in the value of ListPrice can be explained by the predictors Size and SchoolQuality. While not as high as the R^2 (adj) value for Cobb county, it is a respectable result nonetheless, accounting for a solid majority of the variability of ListPrice. What is interesting is the weightier influence that Size has on ListPrice in both counties. As shown in Table 2, the standardized regression coefficients indicate that one standard deviation increase in Size influences ListPrice more than a corresponding one standard deviation increase in any of the other predictors.

Table 2. Standardized Regression Coefficients

Results for: Cobb Sample
Standardized Regression Coefficients for ListPrice

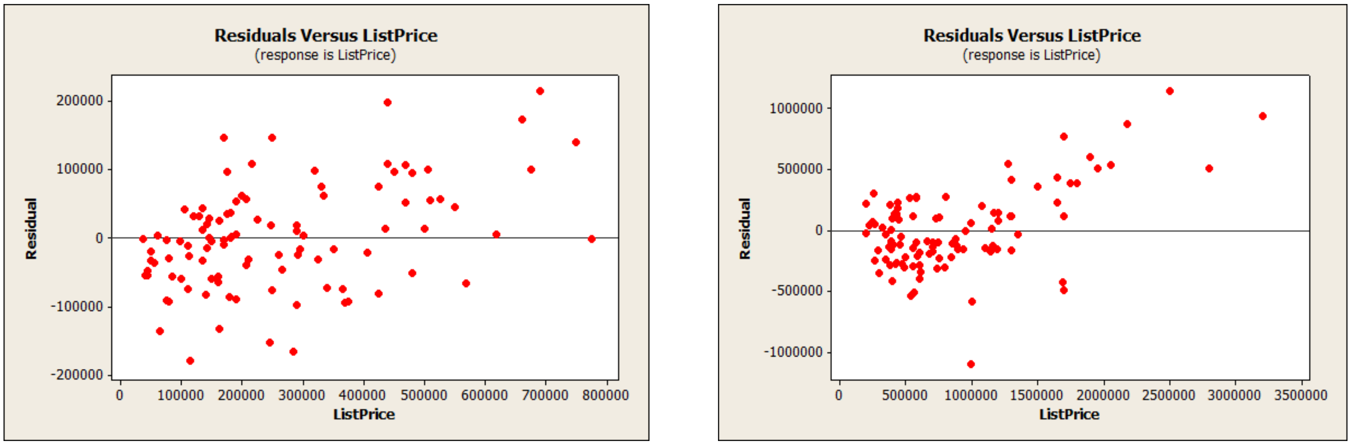
```
Row  Predictors       StdCoef
  1  Size            0.724386
  2  Lot             0.174335
  3  SchoolQuality  -0.268389
```

Results for: Fulton Sample
Standardized Regression Coefficients for ListPrice

```
Row  Predictors       StdCoef
  1  Size            0.626722
  2  SchoolQuality  -0.342619
```

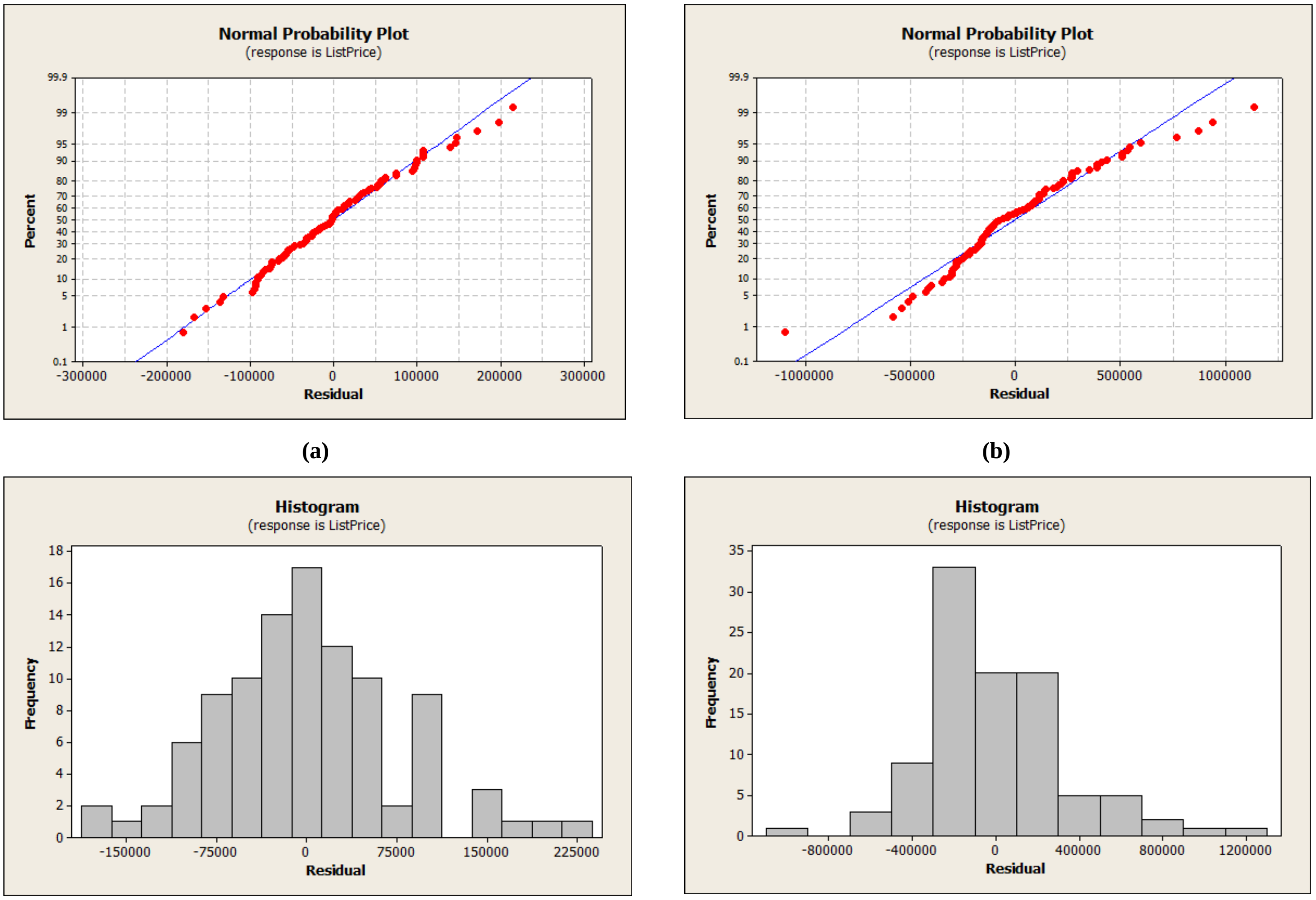
Based on ANOVA, the f-test statistic $F_0 = 49.33$ (Cobb county) > $F_{0.05,99,300} \cong 0.0000$ and $F_0 = 23.37$ (N. Fulton county) $F_{0.05,99,200} \cong 0.0000$ so the null hypothesis is rejected in both instances; i.e. there is not sufficient support for the claim that $\hat{\beta}_0 = \hat{\beta}_1 = \hat{\beta}_2 \dots = \hat{\beta}_j = 0$, indicating that at least one coefficient is non-zero. Thus, based on ANOVA, the regression equations are useful for estimating ListPrice because there is a significant linear relationship between ListPrice and at least one predictor.

Further, as shown in Figures 7 and 8, residual analysis for the multiple regression models associated with each county indicates that the following assumptions are validated: (1) the residuals are independent, (2) the residuals have a constant variance, and (3) the residuals are normally distributed with a mean of zero.



(a) (b)

Figure 7. Plot of Residuals ($\hat{e}_j$) vs. Predicted ListPrice ($\hat{y}_j$) for Cobb and N. Fulton Counties.



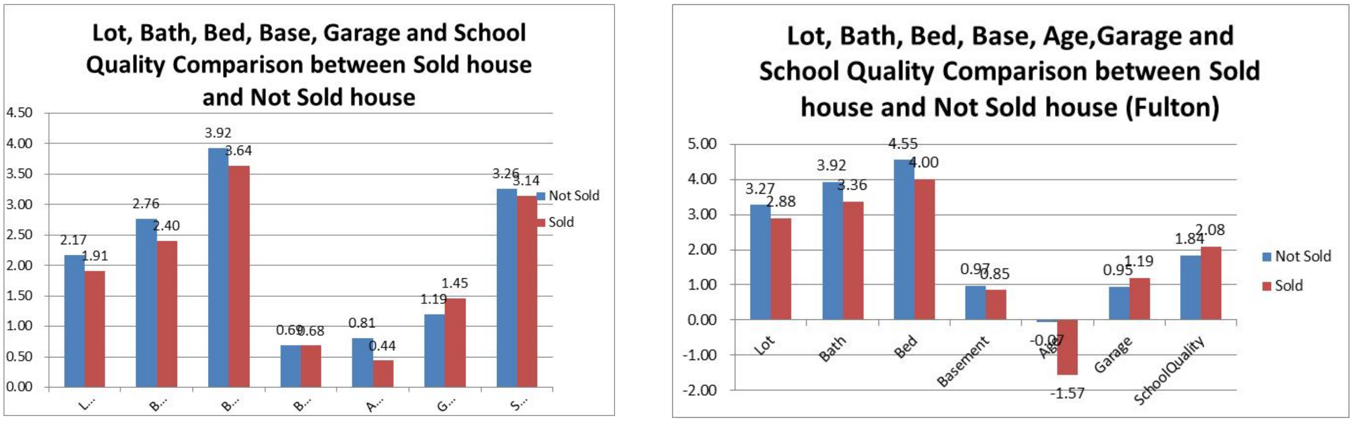
(a) (b)

(c) (d)

Figure 8. Normal Probability Plot and Histograms of Residuals for Cobb and N. Fulton Counties.

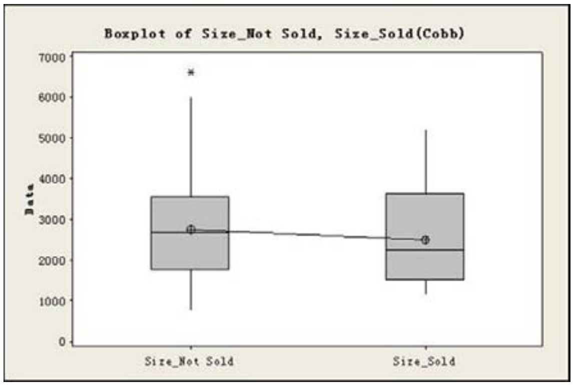
7. Initial Exploration of Sale Price: Comparison of Homes Sold to Homes Not Sold

As shown in Figure 9, we find compared to the homes that are not sold, the sold homes are smaller in size, lot, bath, bed, and basement measures. There are 22 out of 100 houses are sold in Cobb area with average sale price $232,695 and 26 out of 100 houses are sold in Fulton area with average sale price $627,546. The differences, however, are insubstantial enough to lead us to expect that the homes in the sample data providing a non-zero sale price (meaning the home is sold), as a subset of the sample data, is an adequate representation of the population in like manner to the overall sample data. For example, the boxplots and hypothesis tests for the significance of the differences in the means in *Size* are shown in Figure 10.



(a) **(b)**

Figure 9. Size comparison between houses that are sold and not sold (Cobb and Fulton).



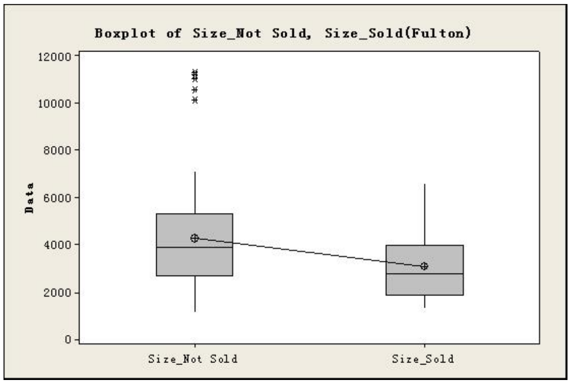
(a)

```
Hypothesis test:

Two-sample T for Size_Not Sold vs Size_Sold

                N  Mean  StDev  SE Mean
Size_Not Sold  78  2744   1280      145
Size_Sold      22  2503   1183      252


Difference = mu (Size_Not Sold) - mu (Size_Sold)
```



(b)

```
Hypothesis test:

Two-sample T for Size_Not Sold vs Size_Sold

                N  Mean  StDev  SE Mean
Size_Not Sold  74  4298   2349      273
Size_Sold      26  3106   1455      285


Difference = mu (Size_Not Sold) - mu (Size_Sold)
```

```
Estimate for difference:  241
95% CI for difference:  (-349, 831)
T-Test of difference = 0 (vs not =): T-Value =
0.83  P-Value = 0.412  DF = 36
```

```
Estimate for difference:  1192
95% CI for difference:  (405, 1980)
T-Test of difference = 0 (vs not =): T-Value =
3.02  P-Value = 0.004  DF = 71
```

(c) **(d)**

Figure 10. Box plots and T-tests for *Size* in Cobb and Fulton Counties.

8. The Regression Analysis of Sale Price

After developing models to predict the list price, we next examine the subset of data for which a sale price is available (houses that have already been sold). The first goal was to examine the relationship between list price and sale price in order to answer the question: "What is the expected sale price for a home with a given list price?"

Since home buyers often negotiate a lower price for a home—and rarely offer to pay more than the list price—it is reasonable to hypothesize that the mean sale price will be less than the mean list price. Due to the dependency between the samples (each house has a pair of prices: a list price and a sale price), a paired-t test was the appropriate hypothesis test. For all of the sold houses, a paired-t test was conducted with the alternative hypothesis stated above (mean sale price < mean list price). The T-value was -4.03 and the P-value was 0.000. The 95% upper bound for the mean difference was -12799 dollars. Therefore, the null hypothesis was rejected, as expected. The sample size of N = 48 (for both counties combined) should be large enough to ensure the validity of the results.

After confirming that the mean sale price was indeed less than the mean list price, a linear regression model was created to try to quantify the relationship between the list price and the sale price. The goal was to develop a model to allow potential home buyers to determine a sale price to negotiate based on the list price of the house. Initially, the county was also included as a categorical predictor; however, it was not a statistically significant predictor. The final model, which included only list price as a predictor, was, `SalePrice = 14003 + 0.923 ListPrice` with the residuals shown in Figure 11.

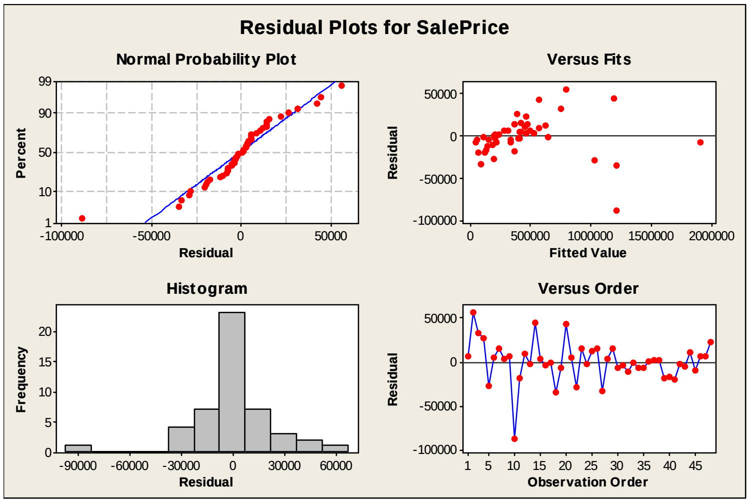


Figure 11: Residuals for Sale Price vs. List Price Regression

Although the adjusted coefficient of determination, $\mathrm{R}^2_{\mathrm{adj}} = 99.6\%$, is very high, the residuals appear to be clustered and do not pass a normality test. There appear to be just a few outliers that are responsible for causing this violation of the normality assumption for the residuals. Eliminating the four unusual observations identified by Minitab results in the slightly modified regression equation `SalePrice = 10319 + 0.933 ListPrice` which has $R^2_{adj} = 99.6\%$, and the residuals shown in Figure 12.

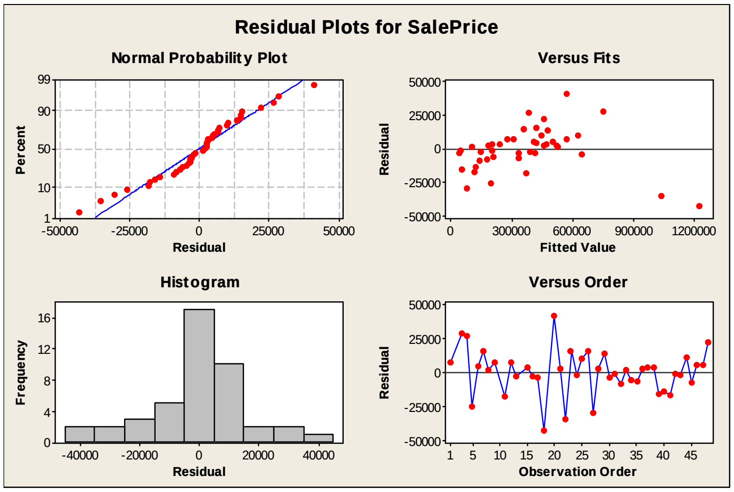


Figure 12: Residuals for Revised Sale Price vs. List Price Regression

The residuals now pass a normality test (with a P-value of 0.102); however, based on the distribution in the upper right subplot, it is possible that the assumption of constant variance could be violated. Nevertheless, this model should provide a useful tool for home buyers to get a rough idea of the expected sale price for a given list price.

Next, regression models were developed to predict sale price based on the predictors in the data. The goal was to determine whether list price or sale price could be predicted more accurately based on the information in the dataset. In order to determine which predictors might be important to include in a multiple linear regression model, simple linear regressions were generated for each predictor vs. Sale price. The R^2 values and P-values (for the hypothesis test with H0: β_1=0) are shown in Table 3.

Table 3: R2 and P-values for simple regressions of each predictor vs. Sale Price

Both Counties										
Predictor	**Size**	**Lot**	**Bath**	**Bed**	**Basement**	**Age**	**Garage**	**Siding**	**School Quality**	**County**
R-sq(adj), %	66.30	24.70	70.00	42.90	3.40	0.00	0.00	8.81	39.30	28.65
P-value	0.000	0.000	0.000	0.000	0.111	0.926	0.601	0.047	0.000	0.000
Fulton Only										
Predictor	**Size**	**Lot**	**Bath**	**Bed**	**Basement**	**Age**	**Garage**	**Siding**	**School Quality**	
R-sq(adj), %	76.60	31.10	64.70	46.40	0.00	7.30	0.00	1.30	21.40	

P-value	0.000	0.002	0.000	0.000	0.613	0.097	0.400	0.330	0.010
					Cobb Only				
Predictor	Size	Lot	Bath	Bed	Basement	Age	Garage	Siding	School Quality
R-sq(adj), %	81.60	0.00	75.80	61.80	37.00	7.60	20.30	16.34	38.30
P-value	0.000	0.917	0.000	0.000	0.002	0.114	0.020	0.071	0.001

All predictors with a P-value less than 0.05 (reject the null hypothesis) were included in the multiple regression model. However, as shown in Figure 13, there is a clear correlation between size, number of bathrooms, and number of bedrooms. Therefore, of these, only size was included in the models.

The procedure for creating the multiple regression model was to initially include all predictors with a P-value greater than 0.05 from the simple regression table (except for Bed and Bath, as mentioned). Then, the predictor with the highest P-value was removed in successive iterations until all remaining predictors in the model had a P-value less than 0.05. County, Size and Lot were found to be the significant predictors. The regression equation was,

```
  Cobb   SalePrice  =  -309275 + 182.707 Size + 44343 Lot
Fulton  SalePrice  =  -67791.1 + 182.707 Size + 44343 Lot
```

which has $R^2_{adj} = 82.82\%$, and the residuals shown in Figure 14. The P-value is 0.000 for the Analysis of Variance test of multiple coefficients.

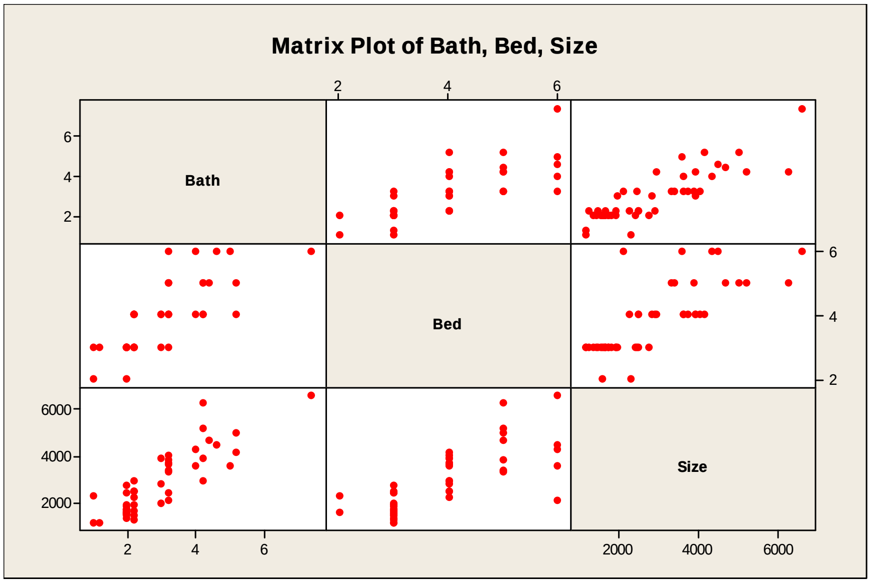


Figure 13: Correlation between Size, # of Bathrooms, and # of Bedrooms

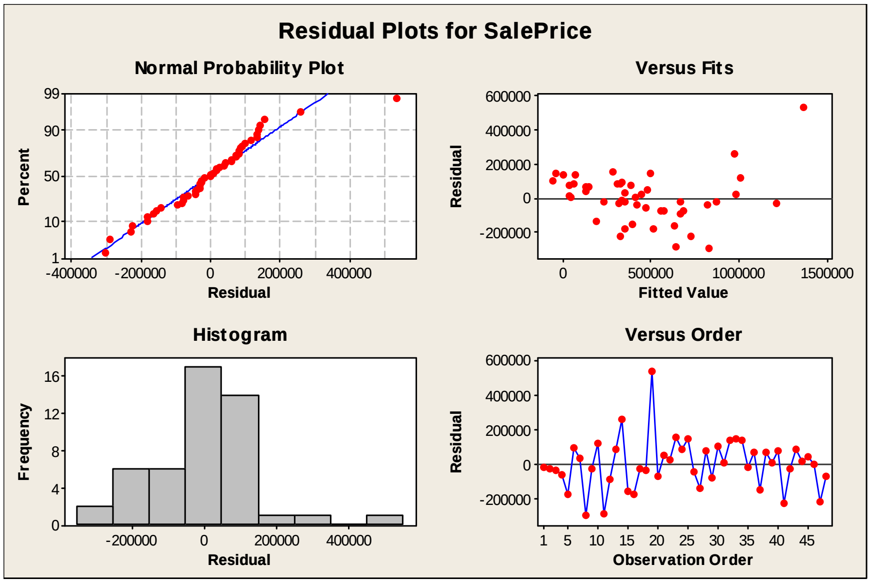


Figure 14: Residuals for model to predict Sale Price

The residuals appear to be random with constant variance, and pass a Kolmogorov-Smirnov normality test with a P-Value > 0.150.

Since we have confirmed that County is an important predictor, the next step was to generate separate models for each county. The procedure was the same as for the regression model for both counties. For Fulton County, Size and Lot were once again the important predictors. The regression equation was, `SalePrice = - 186568 + 206 Size + 60460 Lot` which has $R_{adj}^2 = 80.4\%$, and the residuals shown in Figure 15. The P-value is 0.000 for the Analysis of Variance test of multiple coefficients. The residuals appear to be random with constant variance, and pass a Kolmogorov-Smirnov normality test with a P-Value > 0.150.

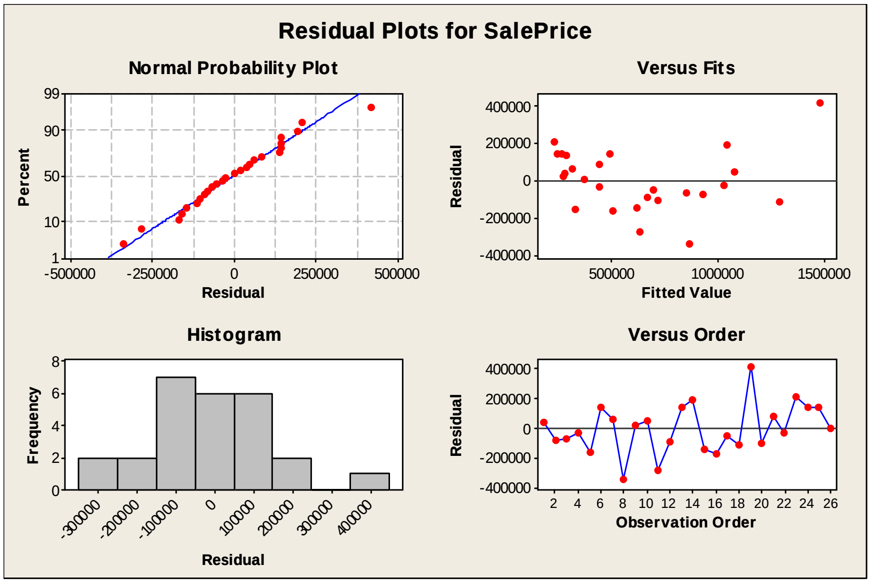


Figure 15: Residuals for Fulton County Model to predict Sale Price

For Cobb County, only Size was an important predictors. The regression equation was, `SalePrice = - 81737 + 126 Size` which has $R^2_{adj} = 81.6\%$, and the residuals shown in Figure 16. The P-value is 0.000 for the Analysis of Variance test.

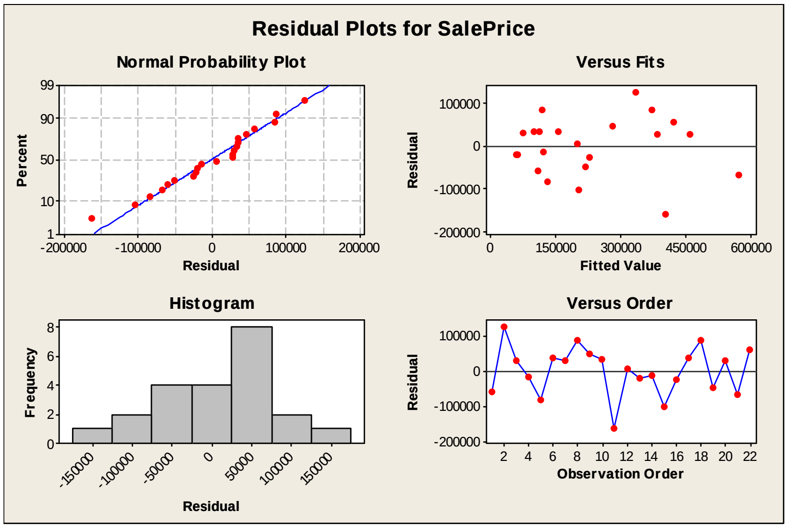


Figure 16: Residuals for Cobb County Model to predict Sale Price

The residuals appear to be random with constant variance, and pass a Kolmogorov-Smirnov normality test with a P-Value > 0.150. Since all three of the regression models for sale price have a higher R^2 value than the models to predict list price, we conclude that we can more accurately predict sale price than list price based on the predictors in this data set. This supports the idea that sale price may be a more accurate reflection of the true market value than list price (but is not sufficient evidence to confirm this hypothesis).

9. Conclusion

Based on our analysis, we determined that the regression models differ significantly by county (i.e. in this sense, location matters). The size of a home is by far the most important predictor of both list price and sale price. To a lesser extent, the quality of the schools also show a significant influence on list price, but not on sale price. One interpretation is that some of the perceived value by the seller of being located in a good school district is illusory and is negotiated away to arrive at a lower sale price. This is consistent with the observation that the predictors are better at estimating sale price that list price (as indicated by higher R^2_{adj} values).

10. Future Research

Were we to follow up our analysis with further research, we would recommend the following avenues: (1) explore different data representations (e.g. do not categorize Age or Lot); (2) use of other predictors, esp. to better define quality differences in homes (e.g. indicators of renovation) and neighborhoods (e.g. distance to stores, private schools, parks); and (3) explore other locations: more granular focus, e.g. by neighborhood or zip code rather than county, and other counties and cities.