Journal of Statistics Education

ISSN: (Print) 1069-1898 (Online) Journal homepage: https://www.tandfonline.com/loi/ujse20

Modeling Home Prices Using Realtor Data

Iain Pardoe

To cite this article: Iain Pardoe (2008) Modeling Home Prices Using Realtor Data, Journal of Statistics Education, 16:2, , DOI: 10.1080/10691898.2008.11889569

To link to this article: https://doi.org/10.1080/10691898.2008.11889569

Journal of Statistics Education

Modeling Home Prices Using Realtor Data

Iain Pardoe
Lundquist College of Business, University of Oregon

Journal of Statistics Education Volume 16, Number 2 (2008), www.amstat.org/publications/jse/v16n2/pardoe.html

---

Key Words: Graphics; Indicator variables; Interaction; Linear regression; Model building; Quadratic; Transformations.

Abstract

It can be challenging when teaching regression concepts to find interesting real-life datasets that allow analyses that put all the concepts together in one large example. For example, concepts like interaction and predictor transformations are often illustrated through small-scale, unrealistic examples with just one or two predictor variables that make it difficult for students to appreciate how these concepts might be applied in more realistic multi-variable problems. This article addresses this challenge by describing a complete multiple linear regression analysis of home price data that covers many of the usual regression topics, including interaction and predictor transformations. The analysis also contains useful practical advice on model building—another topic that can be hard to illustrate realistically—and novel statistical graphics for interpreting regression model results. The analysis was motivated by the sale of a home by the author. The statistical ideas discussed range from those suitable for a second college statistics course to those typically found in more advanced linear regression courses.

1. Introduction

This article describes a complete multiple linear regression analysis of home price data for a city in Oregon, USA in 2005. At the time the data were collected, I was preparing to place my home on the market and it was important to come up with a reasonable asking price. Whereas realtors use experience and local knowledge to subjectively value a home based on its characteristics (size, amenities, location, etc.) and the prices of similar homes nearby, regression analysis provides an alternative approach that more objectively models local home prices using these same data. Better still, realtor experience can help guide the modeling process to fine-tune a final predictive model.

The article discusses statistical ideas ranging from those suitable for the regression component of a second college statistics course to those typically found in more advanced linear regression courses. The analysis includes many elements covered in typical regression components of second statistics courses such as indicator variables for coding qualitative information, model building, hypothesis testing, diagnostics, and model interpretation. The analysis also provides a compelling application of more challenging topics including predictor interactions, predictor transformations, and understanding model results through the use of graphics. I have used this material in my own second statistics course, which is taken by business undergraduates at the University of Oregon. The example generates much discussion with students able to strongly relate to questions about home values either directly or through their parents' homes. Students can engage with this application for a variety of reasons. For example, they could predict the sale price of a home similar to the one in which they grew up, or explore the relative values of different home characteristics such as, "Is an additional bathroom valued more than an additional bedroom?"

This article is based on a case study in Pardoe (2006), which also contains further details on the more routine aspects of a regression analysis. Here I complement that case study by providing additional motivation for the analysis and further background on actual use of this dataset in the classroom. The article is organized as follows: Section 2 describes the dataset; Section 3 shows the model building process to develop suitable multiple linear regression models; Section 4 discusses the final model found in Section 3 along with its potential use; Section 5 describes how to construct graphs to better understand the results of the final model; and Section 6 concludes with ideas for extending the analysis in class or in student assignments.

2. Data Description

The data file containing information on 76 single-family homes in Eugene, Oregon during 2005 was provided by Victoria Whitman, a Eugene realtor. We will model single-family home sale prices (*Price*, in thousands of dollars), which range from $155,000 to $450,000, using these predictor variables:

- *Size* = floor size (thousands of square feet)
- *Lot* = lot size category (from 1 to 11—explained below)
- *Bath* = number of bathrooms (with half-bathrooms counting as 0.1—explained below)
- *Bed* = number of bedrooms (between 2 and 6)
- *Age* = age (standardized: (year built - 1970)=10—explained below)
- *Garage* = garage size (0, 1, 2, or 3 cars)
- *Active* = indicator for "active listing" (reference: pending or sold)
- *Edison* = indicator for Edison Elementary (reference: Edgewood Elementary)
- *Harris* = indicator for Harris Elementary (reference: Edgewood Elementary)
- *Adams* = indicator for Adams Elementary (reference: Edgewood Elementary)
- *Crest* = indicator for Crest Elementary (reference: Edgewood Elementary)
- *Parker* = indicator for Parker Elementary (reference: Edgewood Elementary)

It seems reasonable to expect that homes built on properties with a large amount of land area command higher sale prices than homes with less land, all else being equal. However, an increase in land area of (say) 2000 square feet from 4000 to 6000 should probably make a larger difference (to sale price) than going from 24,000 to 26,000. Thus, realtors have constructed lot size "categories," which in their experience correspond to approximately equal-sized increases in sale price. The categories (variable Lot) used in this dataset are:

Lot size	0-3k	3-5k	5-7k	7-10k	10-15k	15-20k	20k-1ac	1-3ac	3-5ac	5-10ac	10-20ac
Category	1	2	3	4	5	6	7	8	9	10	11

Lot sizes ending in "k" represent thousands of square feet, while "ac" stands for acres—there are 43,560 square feet in an acre. This will prove to be important when we come to use Lot in a multiple linear regression model in Section 3. In a multiple linear regression model, predictors necessarily have "linear" impacts on the response variable (*Price*), such that a unit change in *Lot* is associated with a fixed change in *Price*, whether going from categories 2 to 3 or 7 to 8. By contrast, using actual lot size in square feet in a model would produce less realistic results in which an increase in land area from 4000 to 6000 square feet would be no different (in terms of sale price) than going from 24,000 to 26,000.

To reflect the belief that half-bathrooms (i.e., those without a shower or bath-tub) are not valued by home-buyers nearly as highly as full bathrooms, the variable Bath records half-bathrooms with the value 0.1.

This particular housing market has a mix of homes that were built from 1905 to 2005, with a mean around 1970. Since from the realtor's experience both very old homes and very new homes tend to command a price premium relative to "middle age" homes in this market, a quadratic effect might be expected for an age variable in a multiple linear regression model to predict price. To facilitate this we use a rescaled *Age* variable by subtracting 1970 from "year built" and dividing by 10. The resulting variable has a mean close to zero and a standard deviation just over 2, and represents the number of decades away from 1970.

This dataset includes three types of sales listings: homes that have recently sold, "pending sales" for which a sale price had been agreed but paperwork still needed to be completed, and homes that are "active listings" offered for sale but which have not yet sold. Since at the time these data were collected the final sale price of a home could sometimes be considerably less than the price for which it was initially offered, we define an indicator variable, Active, to model the average difference between actively listed homes (*Active*=1) and pending or sold homes (*Active*=0).

Different neighborhoods in this housing market have potentially different levels of housing demand. The strongest predictor of demand that is available with this dataset relates to the nearest elementary school (out of six) for each home. Thus, we define five indicator variables to serve as a proxy for the geographic neighborhood of each home. The most common elementary school in the dataset, Edgewood Elementary, is the "reference level" and the indicator variables *Edison* to *Parker* represent the difference in price between those schools and Edgewood. Figure 3 in the Appendix contains a map that shows the location of the six elementary schools used in the analysis. The numbers of homes in neighborhoods near to *Crest* (6 homes) and *Adams* (3 homes) are relatively small, which may limit our ability to say much about systematic differences in home prices for these two neighborhoods. We return to this question at the end of Section 3.

3. Regression Model Building

The first step of any data analysis should be to explore the data through drawing appropriate graphs and calculating various summary statistics—Section 6.1 of Pardoe (2006) contains examples for this dataset. The dataset is sufficiently rich and varied that in my experience students engage with the data quite readily and appear to enjoy constructing scatterplots and boxplots for the easily understood variables. After the students get a feel for the data we next apply multiple linear regression modeling to see whether the various home characteristics allow us to model sale prices with any degree of accuracy. All of the topics considered in this section (and also Section 4) — namely R^2, adjusted R^2, regression standard error, residual analysis, indicator variables, variable transformations, interactions, regression assumptions, variable selection, and nested model F-tests — would typically be covered during the regression component of a second college statistics course.

We first try a model with each of the predictors "as is" (no transformations or interactions):

$E(Price) = b_0 + b_1 Size + b_2 Lot + b_3 Bath + b_4 Bed + b_5 Age + b_6 Garage + b_7 Active + b_8 Edison + b_9 Harris + b_{10} Adams + b_{11} Crest + b_{12} Parker$.

However, the residuals from model 1 fail to satisfy the zero mean (linearity) assumption in a plot of the residuals versus *Age*, displaying a relatively pronounced curved pattern. The left-hand plot in Figure 1 displays this plot, together with the *loess fitted line* that provides a graphical representation of the average value of the residuals as we move across the plot (i.e., as *Age* increases). Pardoe (2006, p. 107) and Cook and Weisberg (1999, p. 44) provide more details on the use of loess fitted lines for assessing patterns in scatterplots—this might be a more suitable topic for a more advanced regression course.

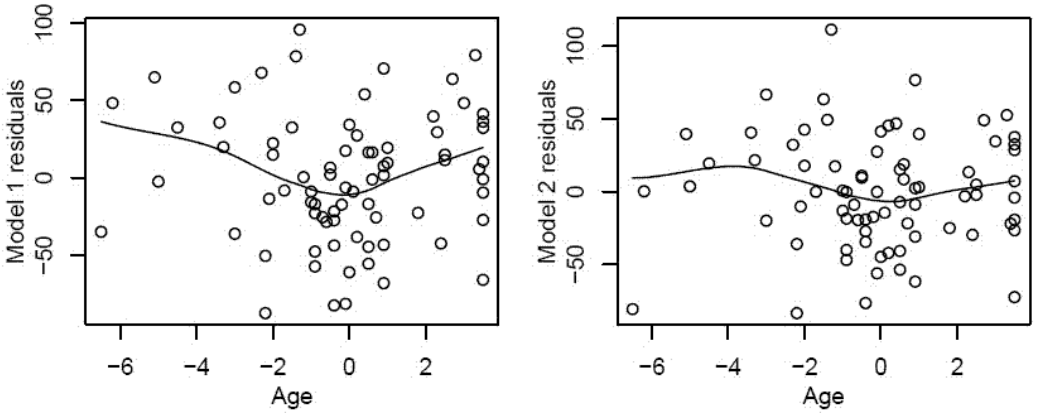

Figure 1: Residual plots for the first model (left) and second model (right), both with Age on the horizontal axis and loess fitted lines superimposed.

To attempt to correct this failing, we will add an Age^2 transformation to the model, which as discussed above was also suggested from the realtor's experience. The finding that the residual plot with *Age* has a curved pattern does not necessarily mean that an Age^2 transformation will correct this problem, but it is certainly worth trying.

In addition, both *Bath* and *Bed* have relatively large individual t-test p-values in model 1, which appears to contradict the notion that home prices should increase with the number of bedrooms and bathrooms. This provides an ideal opportunity to ask students why this might be happening. For example, do they think that adding extra bathrooms to homes with just two or three bedrooms might just be considered a waste of space and so have a negative impact on price. Conversely, is there a clearer benefit for homes with four or five bedrooms to have more than one bathroom, so that adding bathrooms for these homes probably has a positive impact on price. If they agree with these statements, how can the model be extended to allow it to capture these price impacts? In my experience, interaction in regression is a challenging concept, but this kind of plausible and realistic example can greatly ease understanding. The instructor can guide the students in seeing that to model such a relationship we need to add a $Bath \times Bed = BathBed$ interaction term to the model. I show in Section 5 how this interaction can then be displayed graphically too.

Therefore, we next try the following model:

$E(Price) = b_0 + b_1 Size + b_2 Lot + b_3 Bath + b_4 Bed + b_{34} BedBath + b_5 Age + b_{52} Age^2 + b_6 Garage + b_7 Active + b_8 Edison + b_9 Harris + b_{10} Adams + b_{11} Crest + b_{12} Parker$.

Model 2 results in an increased value of R^2 (coefficient of determination), a reduced value of the regression standard error, and residuals that appear to satisfy the four regression model assumptions of zero mean (linearity), constant variance, normality, and independence reasonably well (the residual plot with *Age* on the horizontal axis is displayed as the right-hand plot in Figure 1).

However, the model includes some terms with large individual t-test p-values, suggesting that perhaps it is more complicated than it needs to be and includes some redundant terms. In particular, the last three elementary school indicators (*Adams*, *Crest*, and *Parker*) have p-values of 0.310, 0.683, and 0.389. We can conduct a nested model F-test (also known as an "analysis of variance" test or "extra sum of squares" test) to see whether we can safely remove these three indicators from the model without significantly worsening its fit. The resulting p-value of 0.659 is more than any sensible significance level, so we cannot reject the null hypothesis that the last three school indicator regression parameters are all zero. In addition, removing these three indicators increases the value of adjusted R^2 and reduces the regression standard error. Removing these three indicators from the model means that the school reference level now comprises Edgewood, Adams, Crest, and Parker (so that there are no systematic differences between these four schools with respect to home prices). This also provides the instructor with an opportunity to explore a challenging concept—the use of indicator variables to model qualitative data—within a realistic context.

Thus, a final model for these data is

$$E(Price) = b_0 + b_1 Size + b_2 Lot + b_3 Bath + b_4 Bed + b_{34} BedBath + b_5 Age + b_{52} Age^2 + b_6 Garage + b_7 Active + b_8 Edison + b_9 Harris,$$

with estimated regression equation:

$$\widehat{Price} = 332.48 + 56.72 Size + 9.92 Lot - 98.16 Bath - 78.91 Bed + 30.39 BathBed + 3.30 Age + 1.64 Age^2 + 13.12 Garage + 27.42 Active + 67.06 Edison + 47.27 Harris.$$

Model 3 results in residuals that appear to satisfy the regression model assumptions reasonably well (residual plots not shown). Also, each of the individual t-test p-values is below the usual 0.05 threshold (including *Bath*, *Bed*, and the *BathBed* interaction), except *Age* (which is included to retain hierarchy since Age^2 is included in the model) and *Garage* (which is nonetheless retained since its p-value is low enough to suggest a potentially important effect).

4. Model Interpretation and Use

A potential use for the final model might be to narrow the range of possible values for the asking price of a home about to be put on the market. For example, consider a home with the following features: 1879 square feet, lot size category 4, two and a half bathrooms, three bedrooms, built in 1975, two-car garage, and near Parker Elementary School (this was my home at the time). A 95% prediction interval ignoring the model comes to ($164,800, $406,800); this is based on the formula: sample mean ± t-percentile × sample standard deviation × $\sqrt{(1 + 1/n)}$. By contrast, a 95% prediction interval using the model results comes to ($197,100, $369,000), which is about 70% the width of the interval ignoring the model. A realtor could advise the vendors to price their home somewhere within this range depending on other factors not included in the model (e.g., toward the upper end of this range if the home is on a nice street, the property is in good condition, and landscaping has been done to the yard). As is often the case, the regression analysis results are more effective when applied in the context of expert opinion and experience.

These results also illustrate that "prediction is hard;" whereas the final model might be considered quite successful in terms of its ability to usefully explain the variation in sale price (R^2 = 58.8%), more than 40% of the variation in price remains unexplained by the model. Further, the above 95% prediction interval of ($197,100, $369,000) is perhaps disappointingly wide. The dataset predictors can only go so far in helping to explain and predict home prices in this particular housing market. Students might like to consider ways to explain more variation and to tighten up this interval, for example, by collecting more data observations or thinking of new predictor variables, such as other factors related to the geographical neighborhood, condition of the property, landscaping, and features such as fireplaces and updated kitchens.

A further use for the model might be to utilize the specific findings relating to the effects of each of the predictors on the price. For example, since $\hat{b}_1$= 56.72, we expect sale price to increase by $5672 for each 100 square foot increase in floor size, all else held constant. Interpretations of $\hat{b}_2$ for lot size and $\hat{b}_6$ for garage size are similarly straightforward, but the interpretations for numbers of bedrooms/bathrooms and age are more complicated. Section 5 suggests graphical ideas for increasing understanding of regression parameter estimates when interactions and transformations are involved, as they are here.

The preceding discussion contains most of the standard topics that would typically be covered during the regression component of a second college statistics course. Such courses sometimes also cover more "advanced" topics, such as the role that individual data observations can play in a multiple linear regression model (e.g., outliers or high leverages); these topics would certainly be covered in a more advanced course dealing only with regression. To illustrate, calculation of studentized residuals, leverages, and Cook's distances can help to identify overly influential observations. Finding such observations can suggest the need to investigate possible data errors, to add additional predictors to the model, to respecify the model in some other way, or to consider removing the influential observations from the dataset. However, in this case none of the final model studentized residuals are outside the ± 3 range, and so none of the observations would probably be considered outliers. Home 76 (with a large floor size) has the highest leverage, although home 54 (the oldest home) is not far behind. These two homes also have the two highest Cook's distances, although neither is above a 0.5 threshold (see Cook and Weisberg 1999, p. 358), and neither dramatically changes the regression results if excluded.

5. Predictor Effect Plots

Interpretation of the parameter estimates for *Bath*, *Bed*, and *Age* are complicated somewhat by their interactions and transformations. In such circumstances it can be helpful to use statistical graphics to help interpret the model results. Section 5.4 in Pardoe (2006, pp. 188–194) introduces "predictor effect plots," line plots that show graphically how a regression response variable (home price in this case) is associated with changes to the predictor variables. For the variables *Size*, *Lot*, and *Garage*, these line plots simply represent the corresponding parameter estimates as straightforward slopes. However, in the case of *Bath*, *Bed*, and *Age* the plots provide additional insights due to the presence of complicating interaction effects and transformations.

Note that estimated regression parameters cannot usually be interpreted *causally*. We can really only use the regression models described in this article to quantify relationships and to identify whether a change in one variable is associated with a change in another variable, not to establish whether changing one variable "causes" another to change. The term "predictor effect" in this section indicates how the model expects *Price* to change as each predictor changes (and all other predictors are held constant), but without suggesting at all that this is some kind of causal effect.

The basic ideas behind predictor effect plots are sufficiently straightforward that they could be comfortably covered in the regression component of a second college statistics course. First, use the estimated regression equation to consider how *Price* changes as *Size* changes when we hold the remaining predictors constant (say, at sample mean values for the quantitative predictors and zero for the indicator variables):

$$\textit{Size} \text{ effect on } \textit{Price} = 135.1 + 56.72\textit{Size}.$$

This shows how *Price* changes as *Size* changes for homes with average values for *Lot*,..., *Garage* that are in the Edgewood, Adams, Crest, or Parker neighborhoods. The corresponding predictor effect plot has this *Size* effect drawn as a single straight line on the vertical axis with *Size* on the horizontal axis. It is straightforward to construct similar predictor effect plots for *Lot* and *Garage*.

However, the "*BathBed* effect on *Price*" involves an interaction and so is more complicated:

$$\textit{BathBed} \text{ effect on } \textit{Price} = 504.2 - 98.16\textit{Bath} - 78.91\textit{Bed} + 30.39\textit{BathBed}.$$

The left-hand plot in Figure 2 shows a line plot with this *BathBed* effect on *Price* on the vertical axis, *Bath* on the horizontal axis, and lines marked by the value of *Bed*. Now students can easily visualize the interaction concept previously considered verbally in Section 3. In homes with just two or three bedrooms, additional bathrooms are associated with lower prices (holding all else constant), particularly two-bedroom homes. Conversely, in homes with four or five bedrooms, additional bathrooms are associated with higher prices (all else constant), particularly five-bedroom homes. The scale on the plot shows the approximate magnitude of average prices for different numbers of bathrooms and bedrooms (for homes with average values for *Size*, *Lot*, *Age*, and *Garage* that are in the Edgewood, Adams, Crest, or Parker neighborhoods). Homes with other predictor values tend to have price differences of a similar magnitude for similar changes in the numbers of bathrooms and bedrooms. Similar interpretations can be made for the right-hand plot in Figure 2, which shows a line plot with the *BathBed* effect on *Price* on the vertical axis and *Bed* on the horizontal axis, and lines marked by the value of *Bath*.

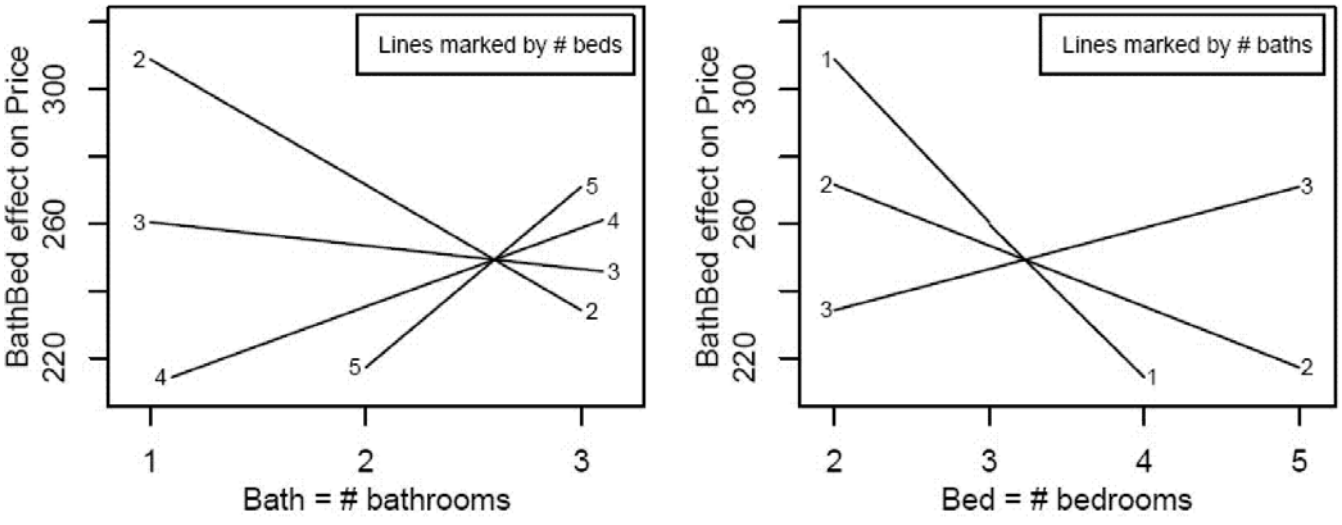

Figure 2: Predictor effect plots for Bath and Bed in the home prices example. In the left plot, the BathBed effect on Price of 504.2 − 98.16Bath − 78.91Bed + 30.39BathBed is on the vertical axis while Bath is on the horizontal axis and the lines are marked by the value of Bed. In the right plot, the BathBed effect on Price is on the vertical axis while Bed is on the horizontal axis and the lines are marked by the value of Bath.

The "*Age* effect on *Price*" is complicated by the transformation:

$$\textit{Age} \text{ effect on } \textit{Price} = 246.9 + 3.30\textit{Age} + 1.64\textit{Age}^2.$$

However, the corresponding predictor effect plot (not shown) is simply a quadratic line that shows average prices decreasing from approximately \$295k to \$245k from the early 1900s to 1960, and then increasing again up to approximately \$280k in 2005 (for homes with average values for *Size*, *Lot*, *Bath*, *Bed*, and *Garage* that are in the Edgewood, Adams, Crest, or Parker neighborhoods).

6. Discussion

This article has described a complete multiple linear regression analysis of a compelling real-life dataset on home prices. The analysis covers many of the usual topics in a regression course, but there are additional possibilities for investigating this dataset further. The following suggestions provide ideas for students to continue working with this dataset. Some, such as the first suggestion below would be reasonable for the regression component of a second college statistics course, while others, such as the final suggestion, might be better suited for a more advanced course dealing only with regression.

1. It is possible that the final model could be improved by considering interactions between the quantitative predictors and the indicator variables, for example, ActiveSize. Investigate whether there are any such interactions that significantly improve the model.
   This suggestion can lead some students to become a little overzealous in their model building, trying every interaction and transformation they can think of to try to improve the fit. This presents an opportunity to discuss the concept of overfitting in regression, and to advise that model building guided by careful thought, as illustrated by using the realtor's experience, is generally preferable to aimless trial and error methods.

2. Investigate whether an alternative measure of lot size might be more appropriate than the categories used in the dataset. For example, define a new predictor variable that is the natural logarithm of the mid-point of the lot size range (in thousands of square feet) represented by each category (i.e., ln(1.5) = 0.41 for category 1, ln(4) = 1.39 for category 2, and so on). Reanalyze the data with this new predictor in place of *Lot*. Do model results change drastically when you do this?
   This can be a useful exercise for students that struggle with mathematical concepts such as logarithmic transformations. Here the predictor effect plots from Section 5 come into their own, since students should find that there is little qualitative difference between the Lot predictor effect plots, whether Lot is defined using the dataset categories or using logarithms.

3. Investigate whether counting half-bathrooms as 0.1 is reasonable. For example, change values of *Bath* ending in .1 in the dataset to end in .5 instead, and reanalyze the data. Do model results change drastically when you do this?
   This gives students an opportunity to think about the choices that are made when data are coded and how model

conclusions might potentially depend critically on those choices. In more advanced courses, this could lead to discussions about robustness and sensitivity tests. In this particular example, students should find that the model results are fairly robust to the choice of .1 versus .5 for half-bathrooms.

4. Investigate whether there appear to be any systematic differences between pending sale prices and actual sales prices (all else equal). The analysis just described assumes no difference since the only indicator variable for "status" is *Active*, which is 1 for active listings and 0 for both pending sales and sold homes. Add an indicator variable that is 1 for pending sales and 0 for both active listings and sold homes, and reanalyze the data. Do model results change drastically when you do this?
   The end of Section 3 noted an opportunity to explore a concept that I have found students generally struggle with—the use of indicator variables for qualitative predictor variables. This suggestion too provides an opportunity to practice this concept, particularly since I have found that while students can often grasp the use of a single indicator variable to model differences between two categories, they can have a harder time grasping the use of two indicator variables to model differences among three categories.

5. The values for *Price* are slightly skewed in a positive direction, suggesting perhaps that transforming *Price* to ln(*Price*) might result in an improved multiple linear regression model. Reanalyze the data, but use ln(*Price*) as the response variable instead of *Price*. Interpret the results, remembering that regression parameter estimates such as $\hat{b}_1$ will need to be transformed to $\exp(\hat{b}_1)-1$, where they now represent the expected proportional change in *Price* from increasing *Size* by one unit (all else constant). Justification for this transformation comes from the following. Partition the model predictors into *Size* and **X** (a vector representing all remaining predictors), and note that the expected change in *Price* after increasing *Size* by one unit (all else constant) is

$$\exp(\hat{b}_1(Size+1) + \hat{\mathbf{b}}\mathbf{X}) - \exp(\hat{b}_1 Size + \hat{\mathbf{b}}\mathbf{X}) = (\exp(\hat{b}_1)-1)\exp(\hat{b}_1 Size + \hat{\mathbf{b}}\mathbf{X}),$$

   where $\hat{\mathbf{b}}$ is a vector representing the estimated regression parameters for the predictors in **X**. Thus, the expected *proportional* change in *Price* after increasing *Size* by one unit (all else constant) is simply $\exp(\hat{b}_1)-1$.
   In my experience, the symbolic algebra involved here remains a bit of a mystery for many students, but they often have an easier time with the empirical results of this model when presented with illustrative examples, for example calculating the predicted prices for two homes that differ only in their size (by 1000 square feet) and noting that this is the same as the proportional price difference implied by $\exp(\hat{b}_1)-1$.

6. Obtain similar data for a housing market near you (e.g., home listings are commonly available on the internet), and perform a regression analysis to explain and predict home prices in that market. Compare and contrast your results with the results presented here.
   Teaching in a quarter system, I haven't had an opportunity to assign this challenge in the course in which I teach this material. However, in elective courses that I teach I have assigned projects in which students collect their own data to analyze, and these projects are often among the most successful aspects of the course.

Appendix

Figure 3 shows the location of the six elementary schools used in the analysis.

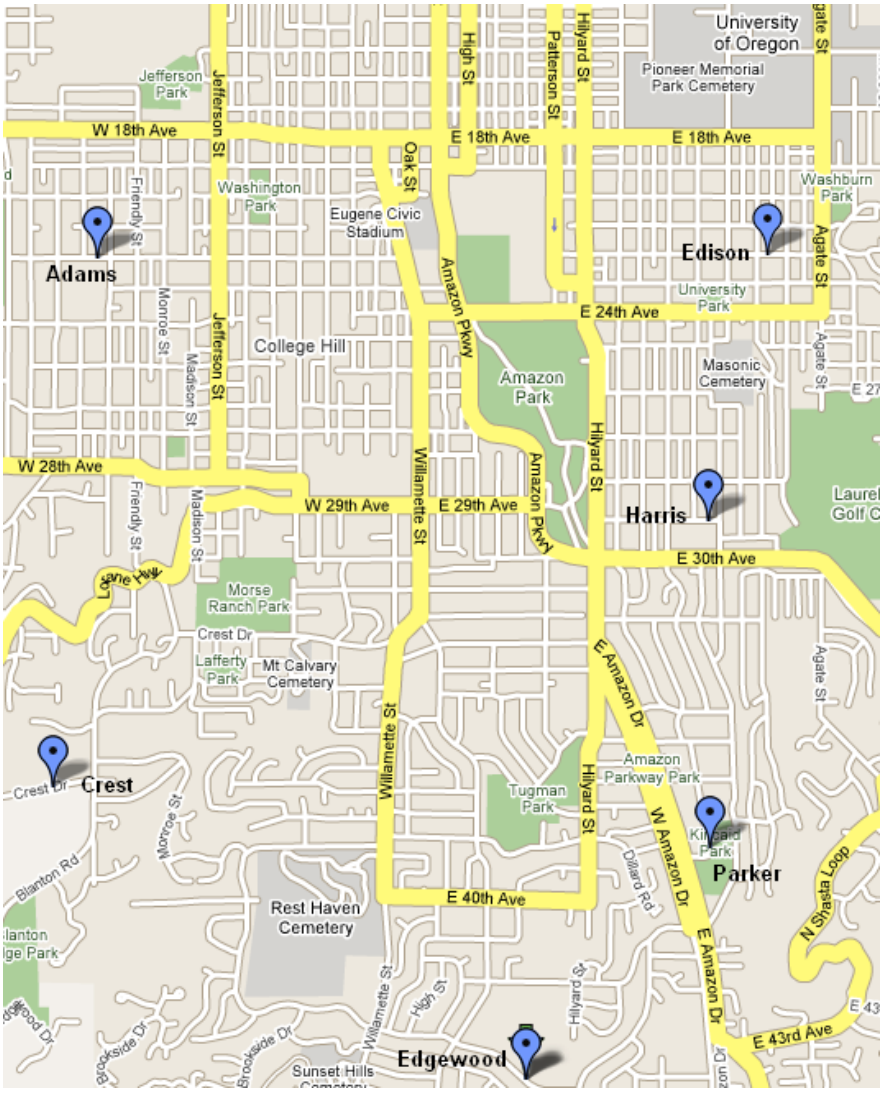

Figure 3: Map of the local area in Eugene, Oregon covered by the analysis, including the locations of the six elementary schools.

Acknowledgments

I thank Victoria Whitman for providing the data and two anonymous reviewers and editors Roger Johnson and Dex Whittinghill for their many helpful suggestions.

References

Cook, R. D. and S. Weisberg (1999). Applied Regression Including Computing and Graphics. Hoboken, NJ: Wiley.

Pardoe, I. (2006). Applied Regression Modeling: A Business Approach. Hoboken, NJ: Wiley.

Iain Pardoe
Lundquist College of Business
University of Oregon
Eugene, Oregon
U.S.A.
ipardoe@lcbmail.uoregon.edu