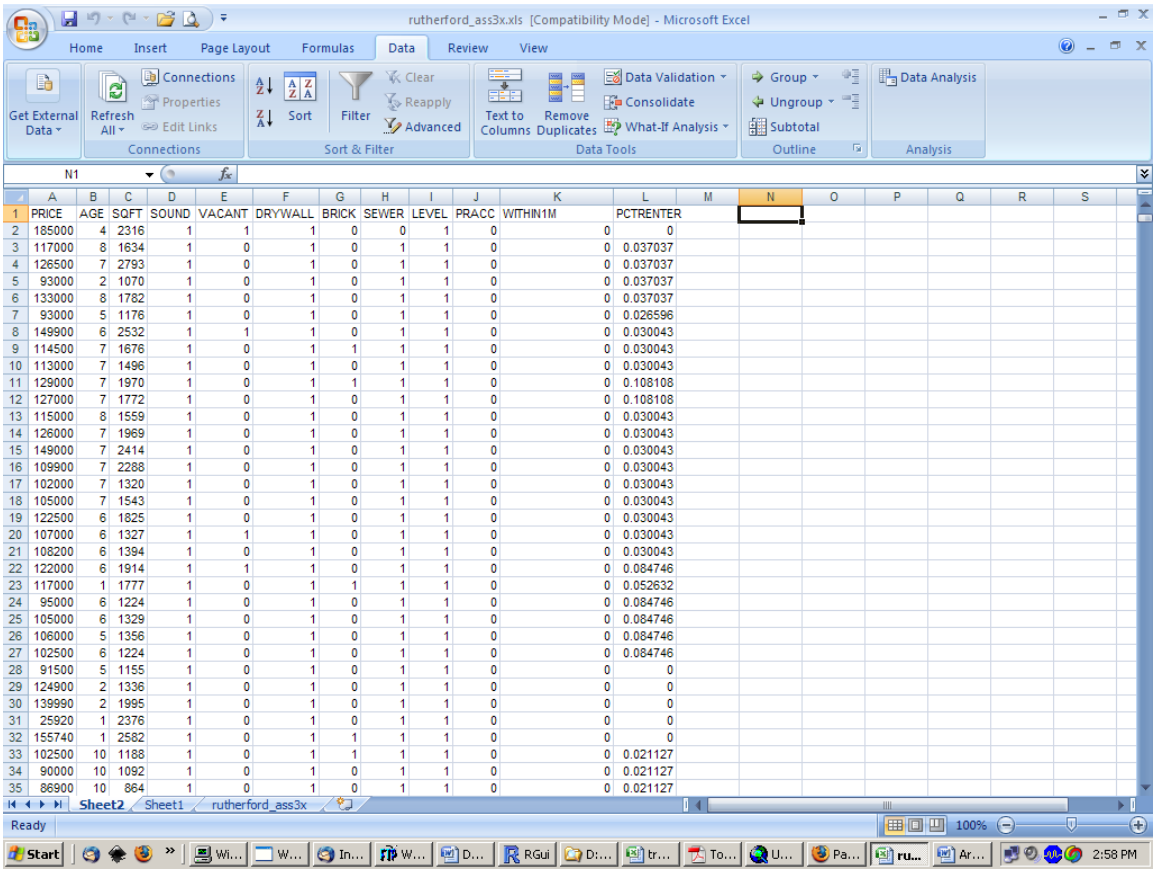
This week we will use the geocoded housing shapefile to estimate a hedonic house price model. We will estimate the model with Excel, and then use ArcGIS to view model results.

Open Excel. Open the dbf file that makes up part of the shapefile you just created, and immediately save it as an *xls* file in your *My Documents/ArcGIS* folder. Add several worksheets to the workbook (click the rightmost tab at the bottom of your worksheet).

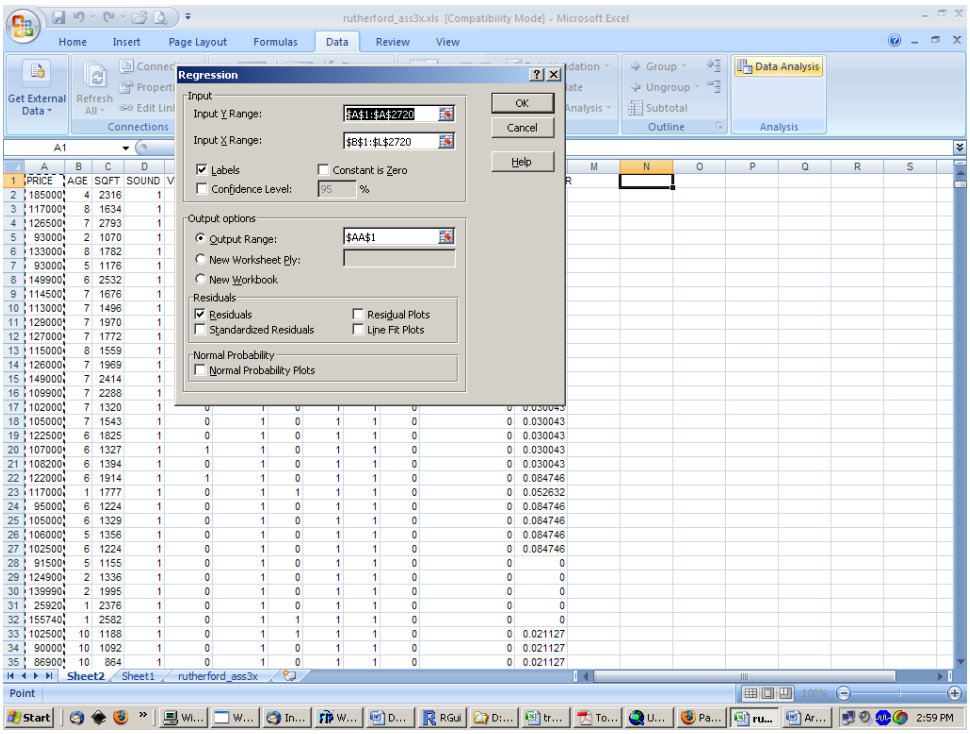


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	PRICE	AGE	SQFT	SOUND	VACANT	DRYWALL	BRICK	SEWER	LEVEL	PRACC	WITHIN1M	PCTRENTER							
2	185000	4	2316	1	1	1	0	0	1	0	0	0							
3	117000	8	1634	1	0	1	0	1	1	0	0	0.037037							
4	126500	7	2793	1	0	1	0	1	1	0	0	0.037037							
5	93000	2	1070	1	0	1	0	1	1	0	0	0.037037							
6	133000	8	1782	1	0	1	0	1	1	0	0	0.037037							
7	93000	5	1176	1	0	1	0	1	1	0	0	0.026596							
8	149900	6	2532	1	1	1	0	1	1	0	0	0.030043							
9	114500	7	1676	1	0	1	1	1	1	0	0	0.030043							
10	113000	7	1496	1	0	1	0	1	1	0	0	0.030043							
11	129000	7	1970	1	0	1	1	1	1	0	0	0.108108							
12	127000	7	1772	1	0	1	0	1	1	0	0	0.108108							
13	115000	8	1559	1	0	1	0	1	1	0	0	0.030043							
14	126000	7	1969	1	0	1	0	1	1	0	0	0.030043							
15	149000	7	2414	1	0	1	0	1	1	0	0	0.030043							
16	109900	7	2288	1	0	1	0	1	1	0	0	0.030043							
17	102000	7	1320	1	0	1	0	1	1	0	0	0.030043							
18	105000	7	1543	1	0	1	0	1	1	0	0	0.030043							
19	122500	6	1825	1	0	1	0	1	1	0	0	0.030043							
20	107000	6	1327	1	1	1	0	1	1	0	0	0.030043							
21	108200	6	1394	1	0	1	0	1	1	0	0	0.030043							
22	122000	6	1914	1	1	1	0	1	1	0	0	0.084746							
23	117000	1	1777	1	0	1	1	1	1	0	0	0.052632							
24	95000	6	1224	1	0	1	0	1	1	0	0	0.084746							
25	105000	6	1329	1	0	1	0	1	1	0	0	0.084746							
26	106000	5	1356	1	0	1	0	1	1	0	0	0.084746							
27	102500	6	1224	1	0	1	0	1	1	0	0	0.084746							
28	91500	5	1155	1	0	1	0	1	1	0	0	0							
29	124900	2	1336	1	0	1	0	1	1	0	0	0							
30	139990	2	1995	1	0	1	0	1	1	0	0	0							
31	25920	1	2376	1	0	1	0	1	1	0	0	0							
32	155740	1	2582	1	0	1	1	1	1	0	0	0							
33	102500	10	1188	1	0	1	1	1	1	0	0	0.021127							
34	90000	10	1092	1	0	1	0	1	1	0	0	0.021127							
35	86900	10	864	1	0	1	0	1	1	0	0	0.021127							

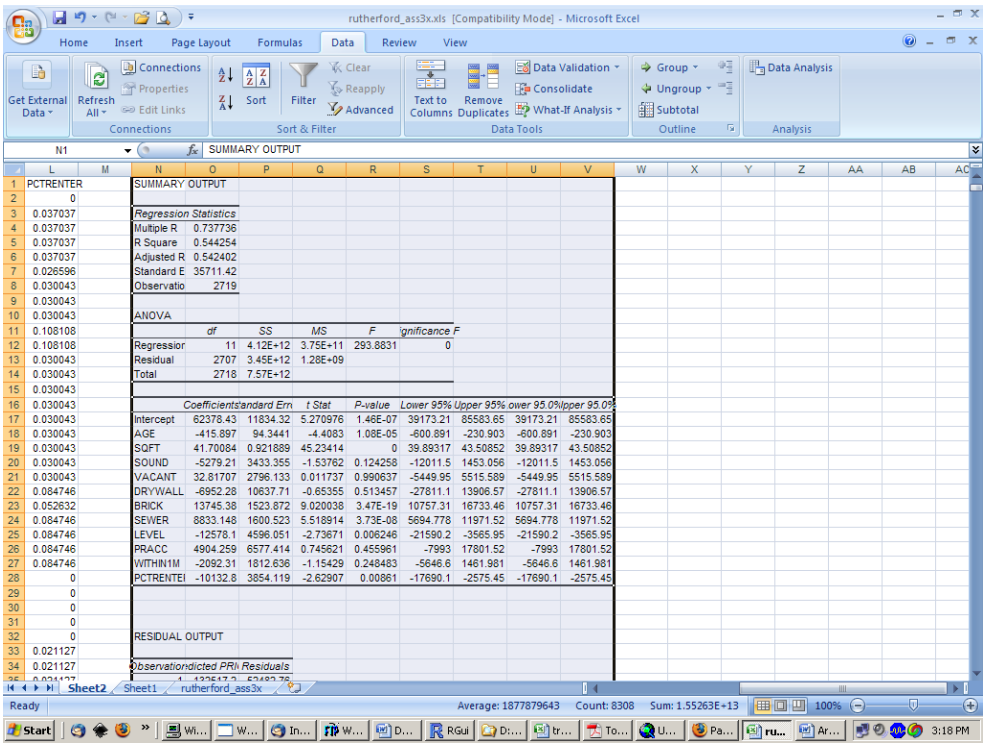
Begin to set up a worksheet for linear regression. Copy the column for 'PRICE' and paste it in the far left column of a blank worksheet. Next to 'PRICE' paste all of the variables that you wish to use in your regression explaining PRICE; you may paste these one at a time or severally—the important point is that all of the variables lie side-by-side in a block. Not all of the variables in the dBase file would be suitable explanatory variables (definitely don't use any character variables, and don't use ACRES or IMPVAL). You will wish to transform some of the Census variables by normalization (i.e., divide the values by population, households, or housing units, as appropriate). Be sure to convert the cells from formulas to values, and then be sure to convert any cells that say '#DIV/0!' to zero.

Once you have a worksheet set up, go to that sheet and click the "File" tab at far left. Click the button for "Options", and you will be presented with a new menu box. On the left is a list of options—click on "Add-Ins". Yet another menu box will appear—highlight "Analysis ToolPak" and click "OK". Now click the "Data" tab toward the top middle of the Excel view. On the far right you should now have a "Data Analysis" icon. Click this, then highlight "Regression" in the resulting list, and click "OK".

A ‘Regression’ window opens (see following screen shot). With the cursor blinking in the ‘Input Y Range’ box, highlight cell A1, where the word ‘PRICE’ stands. Hold down the ‘Shift’ key and the ‘Control’ key and hit the down-arrow key one time: the entire range of price figures should now be represented in the ‘Input Y Range’ box. Now click in the ‘Input X Range’ box, so that the cursor is blinking there. Hold down the ‘Control’ key and hit the ‘Home’ key to return to the top left corner of the worksheet. Then click cell B1, the cell containing the name of your first explanatory variable. Hold down the ‘Shift’ key and the ‘Control’ key and hit the ‘End’ key one time: the entire range of explanatory variable figures should now be represented in the ‘Input X Range’ box. Next put a check mark next to ‘Labels.’ Click the white circle next to ‘Output Range’ and then click a cell in row 1, in a blank column perhaps two columns beyond your last explanatory variable. Finally, click ‘Residuals.’ Click OK to run the regression.



The regression results will be similar to the following screen shot:



There are four items worthy of special attention:

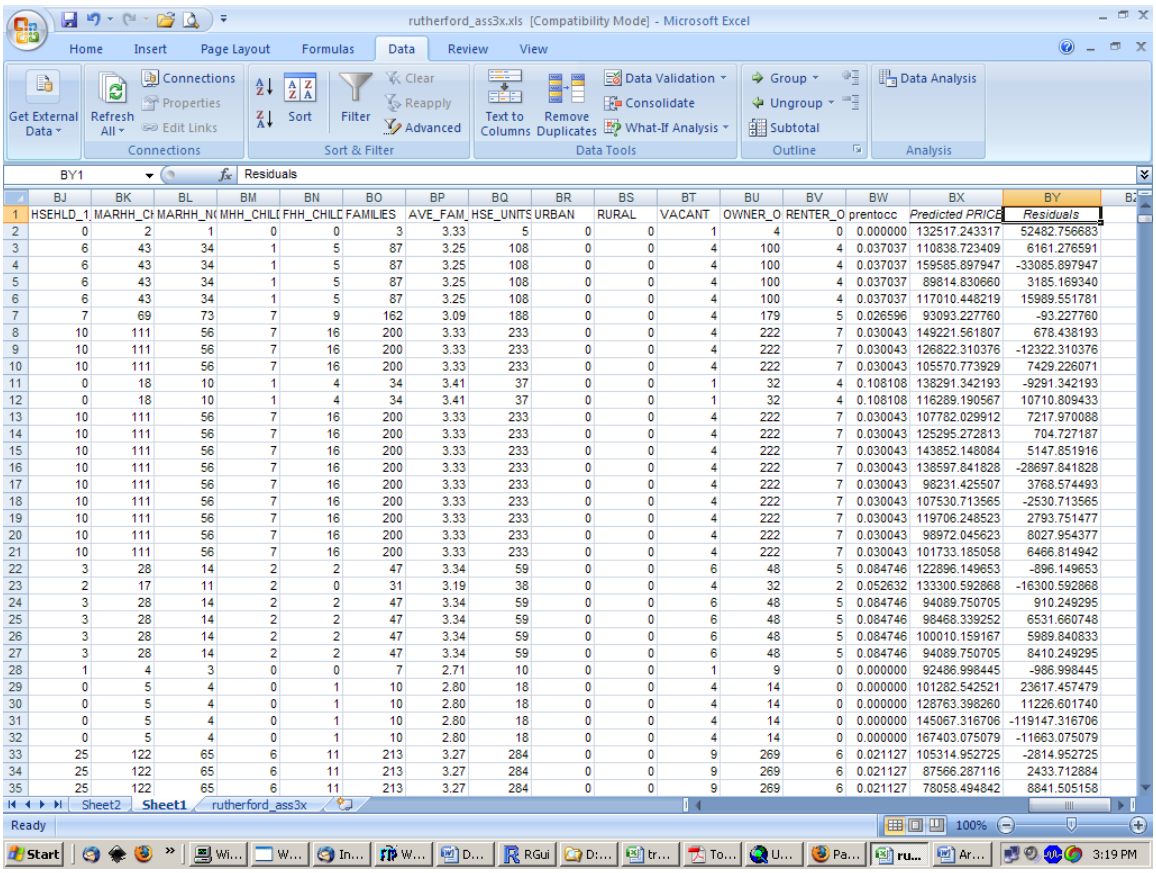
1. The ‘R Square’: this figure gives the proportion of the variation in PRICE explained by the model. If it equaled one (it never should) it would mean that the model explained 100% of the variation in price; if it equaled zero (it never should) it would mean that the model explained none of the variation in price.

2. The ‘Coefficients’: one is really interested in the *sign* of the coefficients. If positive, then an increase in the explanatory variable leads to an increase in price; if negative, then an increase in the explanatory variable leads to a decrease in price.

3. The ‘P-value’: this gives the probability that the true value of the coefficient is zero (i.e., that the explanatory variable has no effect on price). In general, if the P-value is less than 0.1, we consider that the true value of the coefficient is not zero.

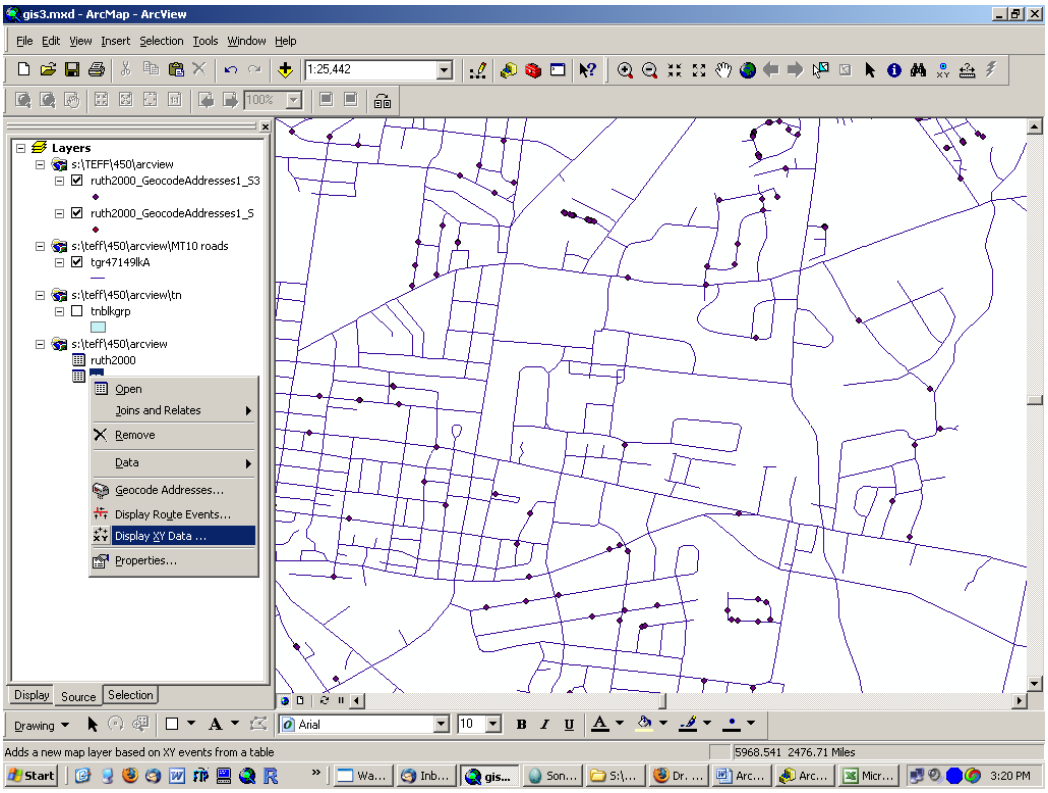
4. The column of ‘Residuals.’ The residual is the difference between PRICE and the predicted PRICE. If the residual is positive, then PRICE is higher than one would have expected, given the values of the explanatory variables; if the residual is negative, then price is lower than one would have expected, given the values of the explanatory variables.

Copy the entire range of predicted PRICE and residuals, and then paste this range in your original worksheet (the one with all of the data brought in from the dBase file), in the two blank columns adjacent to your data (see screen shot below). Go to cell A1 (shortcut: simultaneously press ‘Control’ and ‘Home’), then press ‘Shift’, ‘Control’, and hit ‘End’ one time. The entire block of data should now be highlighted; copy this block of data to a blank worksheet. Save your Excel file (make sure you can find this file).



	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY
1	HSEHLD_1	MARHH_CH	MARHH_NO	MHH_CHILD	FHH_CHILD	FAMILIES	AVE_FAM_	HSE_UNITS	URBAN	RURAL	VACANT	OWNER_O	RENTER_O	prentocc	*Predicted PRICE*	*Residuals*
2	0	2	1	0	0	3	3.33	5	0	0	1	4	0	0.000000	132517.243317	52482.756683
3	6	43	34	1	5	87	3.25	108	0	0	4	100	4	0.037037	110838.723409	6161.276591
4	6	43	34	1	5	87	3.25	108	0	0	4	100	4	0.037037	159585.897947	-33085.897947
5	6	43	34	1	5	87	3.25	108	0	0	4	100	4	0.037037	89814.830660	3185.169340
6	6	43	34	1	5	87	3.25	108	0	0	4	100	4	0.037037	117010.448219	15989.551781
7	7	69	73	7	9	162	3.09	188	0	0	4	179	5	0.026596	93093.227760	-93.227760
8	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	149221.561807	678.438193
9	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	126822.310376	-12322.310376
10	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	105570.773929	7429.226071
11	0	18	10	1	4	34	3.41	37	0	0	1	32	4	0.108108	138291.342193	-9291.342193
12	0	18	10	1	4	34	3.41	37	0	0	1	32	4	0.108108	116289.190567	10710.809433
13	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	107782.029912	7217.970088
14	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	125295.272813	704.727187
15	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	143852.148084	5147.851916
16	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	138597.841828	-28697.841828
17	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	98231.425507	3768.574493
18	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	107530.713565	-2530.713565
19	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	119706.248523	2793.751477
20	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	98972.045623	8027.954377
21	10	111	56	7	16	200	3.33	233	0	0	4	222	7	0.030043	101733.185058	6466.814942
22	3	28	14	2	2	47	3.34	59	0	0	6	48	5	0.084746	122896.149653	-896.149653
23	2	17	11	2	0	31	3.19	38	0	0	4	32	2	0.052632	133300.592868	-16300.592868
24	3	28	14	2	2	47	3.34	59	0	0	6	48	5	0.084746	94089.750705	910.249295
25	3	28	14	2	2	47	3.34	59	0	0	6	48	5	0.084746	98468.339252	6531.660748
26	3	28	14	2	2	47	3.34	59	0	0	6	48	5	0.084746	100010.159167	5989.840833
27	3	28	14	2	2	47	3.34	59	0	0	6	48	5	0.084746	94089.750705	8410.249295
28	1	4	3	0	0	7	2.71	10	0	0	1	9	0	0.000000	92486.998445	-986.998445
29	0	5	4	0	1	10	2.80	18	0	0	4	14	0	0.000000	101282.542521	23617.457479
30	0	5	4	0	1	10	2.80	18	0	0	4	14	0	0.000000	128763.398260	11226.601740
31	0	5	4	0	1	10	2.80	18	0	0	4	14	0	0.000000	145067.316706	-119147.316706
32	0	5	4	0	1	10	2.80	18	0	0	4	14	0	0.000000	167403.075079	-11663.075079
33	25	122	65	6	11	213	3.27	284	0	0	9	269	6	0.021127	105314.952725	-2814.952725
34	25	122	65	6	11	213	3.27	284	0	0	9	269	6	0.021127	87566.287116	2433.712884
35	25	122	65	6	11	213	3.27	284	0	0	9	269	6	0.021127	78058.494842	8841.505158

Now we will read the file in ArcMap. Double click on the menu bar icon that looks like a plus sign on a yellow triangle to bring in your Excel file. Then right click on the filename in the table of contents and click *Display x,y Data* in the drop-down menu. Select *point_X* and *point_Y* as the x and y coordinates. You will see that you created an "event theme" in the table of contents; save this as a new permanent shapefile.



Maps and regression output due next class period.

Normalize your predicted Price with actual Price. Identify properties that are particularly good or bad deals. Try to think of new explanatory variables that might explain why these observations are high or low. Make a map to show to the class next class, along with your regression output. Be prepared to explain what your coefficients mean and why you used the explanatory variables you did. Identify a specific property that you think would have been a good purchase; try to include a photo of the property (Google StreetViews, for example) and find some more information (Zillow, etc).

Next week we have Spring Break. When we return, on March 13, we will present our findings from this house price work, followed by book report presentations.