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).

C	rutherford_ass3x.xls [Compatibility Mode] - Microsoft Excel														σx					
~	🤊 н	ome	In	sert	Page Lay	out Fo	rmulas	Data	Rev	riew Vie	w								🥝 –	<b>□</b> X
				Conn	ections			🖊 🕅 Cle	ar		-	Data Valio	lation -	🔶 Group 👻 📲		Data Analysis				
			<u>e</u> 7	Prope	rties	Z + Z Ă	l II	Reapply				🗧 🖓 Consolidate			👙 Ungroup 🛪 🖷					
Ge	t External	al Refresh				Z↓ Sort	Filte	er Vand	anced	Text to	Remove	fill Culture L								
	Data *	All + See Eure Enrics							anceu	Columns	Duplicates	Subtotal								
			Cor	nection	s		Sort &	Filter		Л	Data Io	Outli	ne 🤘	Ana	lysis					
	N1			- (	f <sub>x</sub>													×		
	Α	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Q	R	S	
1	PRICE	AGE	SQFT	SOUND	VACANT	DRYWALL	BRICK	SEWER LE	VEL PF	RACC WITHIN	11M	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	1/82	1	0	1	0	1	1	0	0	0.037037								
-	93000	5	11/6	1	0	1	0	1	1	0	0	0.026596								_
0	149900	7	1676	1	1		1	1	1	0	0	0.030043								_
10	113000	7	1496	1	0		0	1	1	0	0	0.030043								
11	129000	7	1970	1	0		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								
20	102500	6	1224	1	0		0	1	1	0	0	0.004740								
28	91500	5	1155	1	0		0	1	1	0	0	0.004740								
29	124900	2	1336	1	0		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								-
М	4 F FI	Shee	t2 / 9	Sheet1	rutherfo	ord_ass3x	<u>(</u>										_			
Re	ady																100%	Θ	-0	-+
<b>a</b>	Start	<b>(3)</b>	ک 🌲	»	💐 Wi	🗖 w	🌀 In	<b>11</b> W	1	) 🔣 🧖 RG	iui 🚺 🙆 D:	🕙 tr	🔁 То	Q U	🥮 Pa	🕙 ru	💇 Ar	B 0 0	<b>)</b> 🧑	58 PM

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.

Image: State of the state o																				
Home Insert Page Layout Formulas Data Review View														- ×						
				Conne	d:(				٦٢	·····		a	idation *	Grour	- 0E	Data	Analysis			
	Regression													- Croup		a journ	Analy 312			
Ge	Cat External Pafrarb												late	🖓 Ungro	oup * "=					
Data * All * © Edit Lini Inout V Paper												ж	Analysis 👻	Subto	tal					
Connections											Ca	ancel		Outlin	ne 🕞	Ana	lysis			
A1 ■ Input <u>X</u> Range: \$8\$1:\$L\$2720 II														<u></u>		π				×
	A			6						_		who I				-	-			~
4	A	B	C	D		Labels		Const	ant is <u>Z</u> ero			<u></u>	М	N	0	Р	Q	R	S	-
1	PRICE	AGE	SUFT	SUUND		Confidence	Level:	95	%				ĸ							-
2	117000	4	1634	1				,												
4	126500	7	2793	1	Outp	ut options														
5	93000	2	1070	1		Output Rapo	10'	\$A	4\$1	1										
6	133000	8	1782	1		Zachar vauð														
7	93000	5	1176	1	- 0	Vew Worksh	eet <u>P</u> ly:													
8	149900	6	2532	1	0	New <u>W</u> orkbo	iok.													
9	114500	7	1676	1	Resi	iduals														
10	113000	7	1496	1	E F	Residuals		Г	Residual	Plots										
11	129000	7	1970	1			d Residua	s T	Line Fit P	lots										
12	127000	7	1772	1																
13	115000	8	1559	1	Norr	mal Probabili	ty													
14	140000	7	2414	1	-    🗆 [	<u>N</u> ormal Prob	ability Plo	:5												
16	109900	7	2288	1																
17	102000	7	1320	1						U		0.000040								
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								
20	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	0.004740								
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	U	0	0.021127								
14		Snee	LZ / 5	neet1 /	rumern	ord_ass3x												0		
Po	int																E 100%	e	1	(+)
<b>a</b> y	Start	<b>(3)</b>	ک چ	»	🗐 Wi	🔲 W	🎯 In	519 W	🕑 D	💦 RGui	😂 D:	🕙 tr	🔁 То	Q U	🥹 Pa	🕙 ru	🖭 Ar	100	🎾 🧑 2	:59 PM

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

😭 🖾 🤊 - 🕐 😂 🔔 = rutherford_ass3xxis (Compatibility Mode) - Microsoft Excel														- 🖷 X				
C	Home Insert Page Layout Formulas Data Review View													0 -	⇒ x			
Ge	External Data *	Refresh All *	Connections <sup>™</sup> Properties <sup>∞</sup> Edit Links		Sort Filter		Clear Reapply Advanced	Text to Remove Columns Duplicate		Data Validation *		<ul> <li>⇒ Group * </li> <li>◆ Ungroup * </li> <li>■ Subtotal</li> </ul>		-1	Data Analysis			
		0	nnections	£ SUMM		t & Filter			Dat	a lools		J O	utline	19				
_	1		▼ (G	Jz Somm		0	D	6	т		M	W.	V	V	7	A A	AR	
1	PCTRENT	R	SUMMARY	OUTPUT	-	u	R	3		U	v	**	~	1	2	MA	AD	AL
2	(	0																
3	0.03703	7	Regression	n Statistics														
4	0.03703	7	Multiple R	0.737736														
5	0.03703	7	R Square	0.544254														
6	0.03703	7	Adjusted R	0.542402														
7	0.026596	6	Standard E	35711.42														
8	0.030043	3	Observatio	2719														
9	0.03004	3																
10	0.030043	3	ANOVA															
11	0.10810	3		df	SS	MS	F	gnificance	F									
12	0.10810	3	Regression	11	4.12E+12	3.75E+11	293.8831	0										
13	0.03004	3	Residual	2707	3.45E+12	1.28E+09												
14	0.03004	3	Total	2718	7.57E+12													
15	0.03004	3																
16	0.03004	3		Coefficients	andard Ern	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%							
17	0.03004	3	Intercept	62378.43	11834.32	5.2/09/6	1.46E-07	39173.21	85583.65	39173.21	85583.65							
18	0.03004	5	AGE	-415.897	94.3441	-4.4083	1.08E-05	-600.891	-230.903	-600.891	-230.903							
20	0.03004	) )	SOUND	41.70004	0.921009	45.23414	0.424259	12011 5	43.30032	12011 5	43.50652							
20	0.03004	2	VACANT	22 94707	2706 122	-1.55702	0.124230	-12011.5	EE1E E00	-12011.5	FE1E 590							
22	0.08474	8	DRYWALL	-6952.28	10637 71	-0.65355	0.513457	-27811.1	13906 57	-27811.1	13906 57							
23	0.05263	2	BRICK	13745.38	1523 872	9.020038	3 47E-19	10757 31	16733.46	10757 31	16733.46							
24	0.08474	3	SEWER	8833,148	1600.523	5.518914	3.73E-08	5694,778	11971.52	5694,778	11971.52							
25	0.08474	3	LEVEL	-12578.1	4596.051	-2.73671	0.006246	-21590.2	-3565.95	-21590.2	-3565.95							
26	0.084746	3	PRACC	4904.259	6577.414	0.745621	0.455961	-7993	17801.52	-7993	17801.52							
27	0.084746	6	WITHIN1M	-2092.31	1812.636	-1.15429	0.248483	-5646.6	1461.981	-5646.6	1461.981							
28	(	)	PCTRENTEI	-10132.8	3854.119	-2.62907	0.00861	-17690.1	-2575.45	-17690.1	-2575.45							
29	(	0									1							
30	(	)																
31		)																
32	(	0	RESIDUAL	OUTPUT														
33	0.02112	7																
34	0.021127 Observation-dicted PRI/ Residuals											-						
25		heet2 /	Sheet1 / n	utherford a	ss3x	2/					4							
Re	adv								Average: 1	877879643	Count: 830	08 Sun	1.552638	+13	<b>D D</b> 1009	. —		<b>(</b>
		A 4	s » I 🔳			- 1-20				. 1.00.	1 -	1	1 40 0					2.40.014
2	Start	ු 🌸 🍯	) » 📃	wi 📃 '	w   🧐 I	n   🎵 V	v   🖭 c	)   🖳 R	:Gui 🛛 🙆 D	: 🛛 😢 tr	🛛 💏 To	.   🧿 U.	🧐 P	a    🖾 r	u 🕺 Ar	. 🛃 🤨	: 🔊 🐽 🔍 ا	3:18 PM

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).

rutherford_ass3x.xls [Compatibility Mode] - Microsoft Excel																				
_	~ +	lome	Inse	rt P	age Layou	it F	ormula	s Dat	a Rev	iew Vi	ew								- 10	a x
Get	Externa	Connections Properties			tions A es z	2 ZA K Clear			Clear Reapply	Data Validation ▼						oup + 4				
[	Data 🕆	All	- Ge	Edit Lin	ks Ä	1 2011		<b>1</b>	Advanced	Columns Duplicates 🔛 What-If Analysis 🔻						btotal				
			Conn	ections			Sort	& Filter		Data Tools						utline	Gi .	Analysis		
BV4 - G			£. Des	iduale														×		
_	01				Jx 100	DM DN						DO DT				517	5.00	51/	514	V
4	BJ	1 MAD		BL MADUU I	BW BW			BU	BP AVE EAM	BQ HCE UNITE	BR	BUDAI	BI		BU ANNER O	BV DENTED O	BW	BX Deadlated DRICE	BY	B/
2	HOEHLU	0	2	MARIN_	1	0		AMILES	2 33	Hac_UNITS	ORDAN	RURAL	0 VACA	1	WINER_0	RENTER_0	0.000000	132517 243317	52482 756683	-
3		6	43	3	4	1	5	87	3.25	108	0		0	4	100	4	0.037037	110838 723409	6161 276591	
4		6	43	3	4	1	5	87	3.25	108	0		0	4	100	4	0.037037	159585 897947	-33085 897947	
5		6	43	3	4	1	5	87	3.25	108	0		0	4	100	4	0.037037	89814.830660	3185,169340	
6		6	43	3	4	1	5	87	3.25	108	0		0	4	100	4	0.037037	117010.448219	15989.551781	
7		7	69	7	3	7	9	162	3.09	188	0	)	0	4	179	5	0.026596	93093.227760	-93.227760	
8		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	149221.561807	678.438193	
9		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	126822.310376	-12322.310376	
10		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	105570.773929	7429.226071	
11		0	18	1	0	1	4	34	3.41	37	0	)	0	1	32	4	0.108108	138291.342193	-9291.342193	
12		0	18	1	0	1	4	34	3.41	37	0	)	0	1	32	4	0.108108	116289.190567	10710.809433	
13		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	107782.029912	7217.970088	
14		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	125295.272813	704.727187	
15		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	143852.148084	5147.851916	
16		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	138597.841828	-28697.841828	
17		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	98231.425507	3768.574493	
18		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	107530.713565	-2530.713565	
19		10	111	5	6	7	16	200	3.33	233	0	)	0	4	222	7	0.030043	119706.248523	2793.751477	
20		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	98972.045623	8027.954377	
21		10	111	5	6	7	16	200	3.33	233	0		0	4	222	7	0.030043	101733.185058	6466.814942	
22		3	28	1	4	2	2	47	3.34	59	0		0	6	48	5	0.084746	122896.149653	-896.149653	
23		2	17	1	1	2	0	31	3.19	38	0		0	4	32	2	0.052632	133300.592868	-16300.592868	
24		3	28	1	4	2	2	47	3.34	59	0		0	6	48	5	0.084746	94089.750705	910.249295	
25		3	28	1	4	2	2	47	3.34	59	0		0	6	48	5	0.084746	98468.339252	6531.660748	
20		2	20		4	2	2	47	3.34	59			0	0	40	5	0.004740	04090 750705	5909.040033	
21		3	20		9	2	2	4/	3.34	59	0		0	4	40	0	0.004740	94069.750705	0410.249295	
20		0			4	0	1	10	2.71	10	0		0	4	14	0	0.000000	101292 542521	22617 457470	
30		0	5		4	0	1	10	2.00	18	0		0	4	14	0	0.000000	128763 398260	11226 601740	
31		0	5		4	0	1	10	2.00	18			0	4	14	0	0.000000	145067 316706	-119147 316706	
32		0	5		4	0	1	10	2.00	18	0		0	4	14	0	0.000000	167403 075079	-11663.075079	
33		25	122	6	5	6	11	213	3.27	284	0		0	9	269	6	0.021127	105314.952725	-2814.952725	
34		25	122	6	5	6	11	213	3.27	284	0		0	9	269	6	0.021127	87566.287116	2433.712884	
35		25	122	6	5	6	11	213	3.27	284	0	)	0	9	269	6	0.021127	78058.494842	8841.505158	_
14 4	I F FI	Sheet2	Sh	eet1 /	rutherfor	d_ass3x	10	1					1	1						
Rea	adv		-			-											m	<b>II</b> III 100% (		<b>(+)</b>
<u>a</u>	Start	<b>(3)</b>	e 🍪	» 🔳	. wi	] w [	🄄 In	<b>ED</b> É V	v   🗐 c	)   限 R	Gui 🛛 🙆 🕻	o 🗐	tr   🕇	🖣 To	.   🥥 u.	🛛 🕲 Pa	🕲 n	J 🔯 Ar		:19 PM

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.