By Hayden Lamprecht

Stata

About Stata

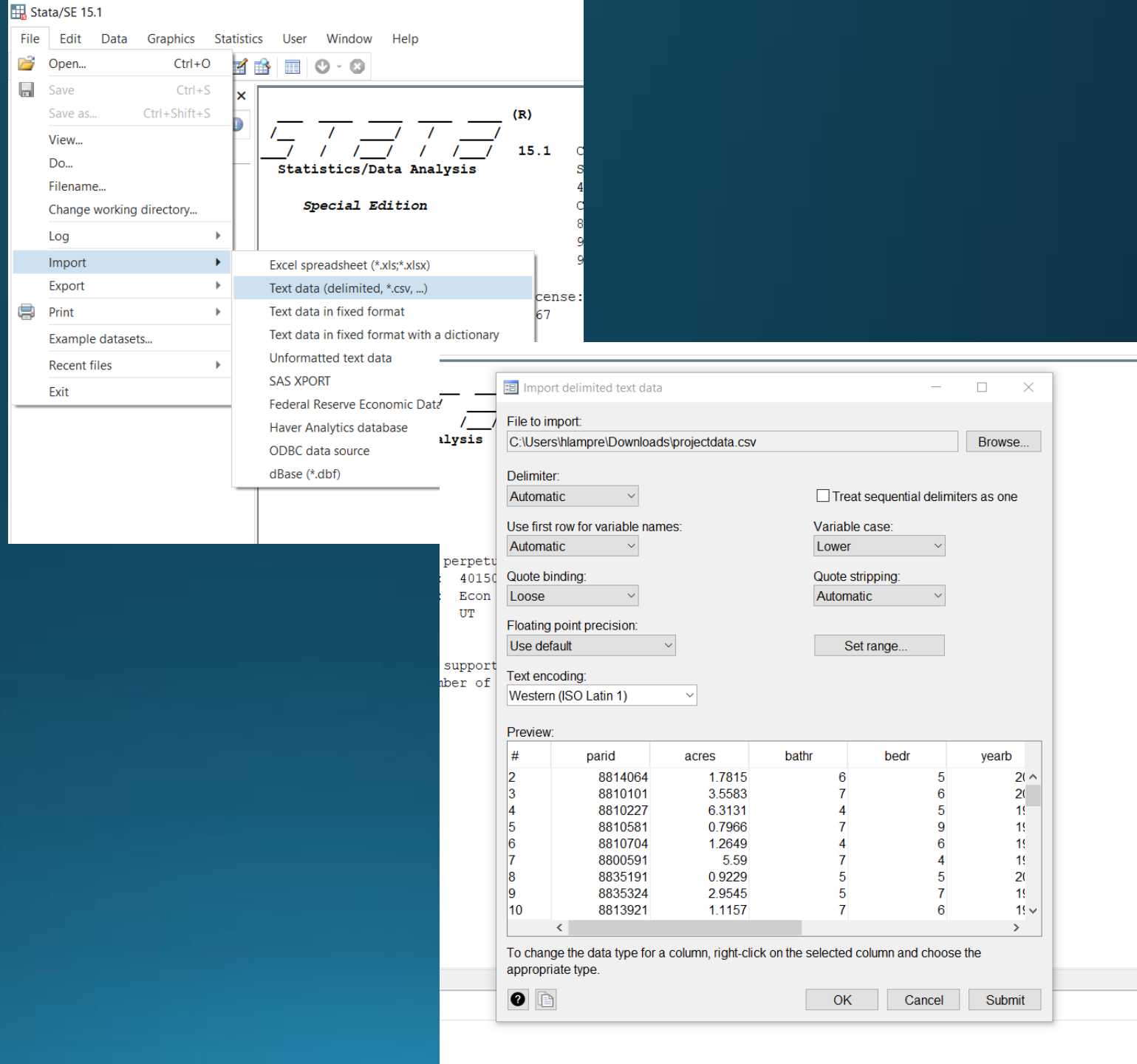
- Stata was founded in 1985 in Texas by William Gould
- Very user friendly
- Great to manipulate data (importing, cleaning, etc.)
- Regression analysis is easily accessible and understandable
- Some data visualization tools

Using Stata

- Stata is not free to use
- Must purchase a license from their website
- Licenses can be purchased for single use or company use
- As cheap as $160 for a single student for 1 year up to $720 a year for 8 cores, which is the highest performance
- Stata is available at UT in UH 4150

Loading Data

- Use CSV version of excel document
- File > Import > Text data
- Import from files on computer



Creating a Dummy Variable

- "Generate" creates a new variable in Stata
- Using parentheses around the condition creates a dummy variable

```
. generate apartment = (acres == -9)
```

Replacing

- "Replace" can be used as seen

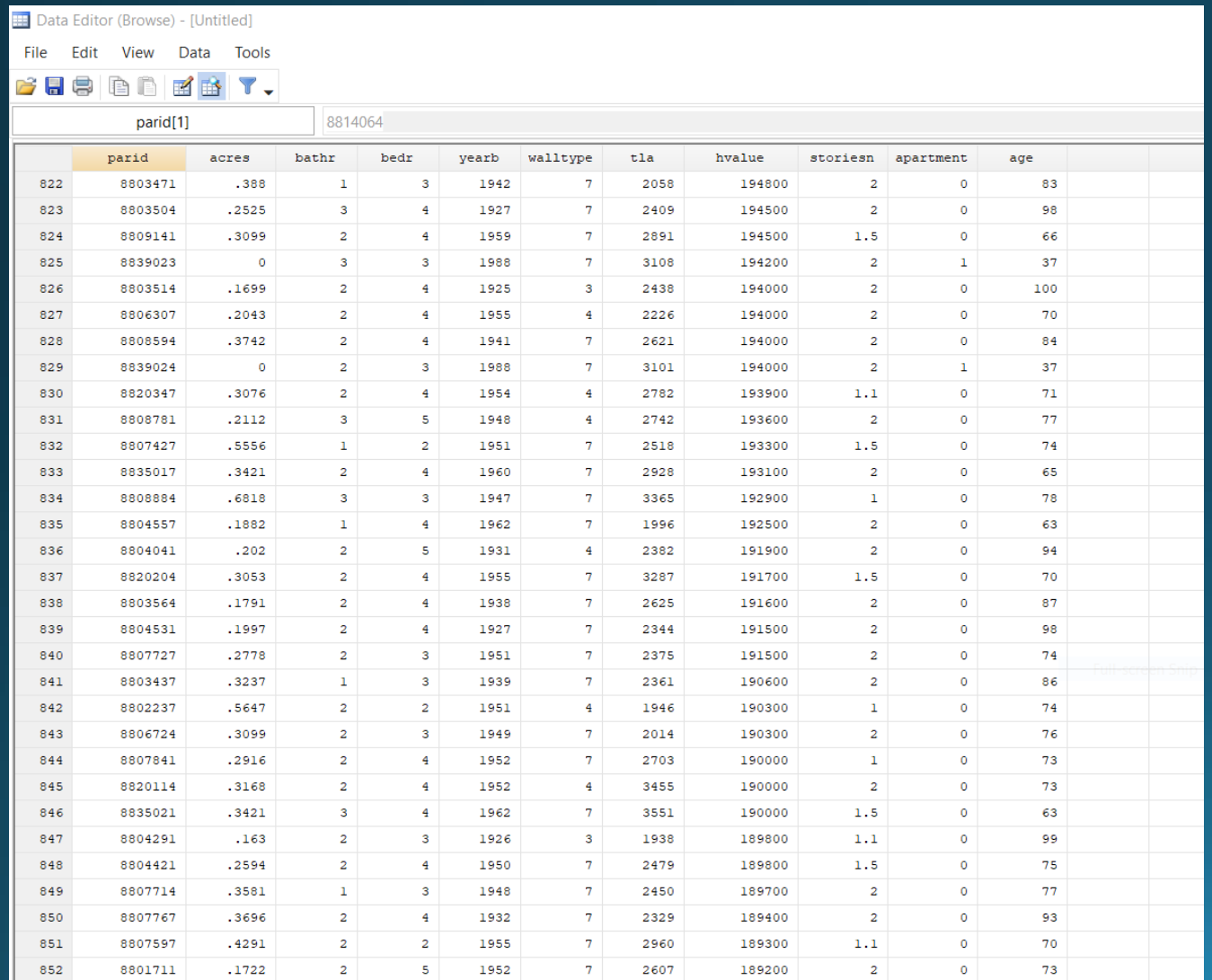
```
. replace acres=0 if acres == -9
(106 real changes made)
```

Creating New Variable

- "Generate" can be used again
- However, no parenthesis are needed

```
. generate age = 2025 - yearb
```

Final Data


Data Editor (Browse) - [Untitled]

File Edit View Data Tools

parid[1] 8814064

	parid	acres	bathr	bedr	yearb	walltype	tla	hvalue	storiesn	apartment	age
822	8803471	.388	1	3	1942	7	2058	194800	2	0	83
823	8803504	.2525	3	4	1927	7	2409	194500	2	0	98
824	8809141	.3099	2	4	1959	7	2891	194500	1.5	0	66
825	8839023	0	3	3	1988	7	3108	194200	2	1	37
826	8803514	.1699	2	4	1925	3	2438	194000	2	0	100
827	8806307	.2043	2	4	1955	4	2226	194000	2	0	70
828	8808594	.3742	2	4	1941	7	2621	194000	2	0	84
829	8839024	0	2	3	1988	7	3101	194000	2	1	37
830	8820347	.3076	2	4	1954	4	2782	193900	1.1	0	71
831	8808781	.2112	3	5	1948	4	2742	193600	2	0	77
832	8807427	.5556	1	2	1951	7	2518	193300	1.5	0	74
833	8835017	.3421	2	4	1960	7	2928	193100	2	0	65
834	8808884	.6818	3	3	1947	7	3365	192900	1	0	78
835	8804557	.1882	1	4	1962	7	1996	192500	2	0	63
836	8804041	.202	2	5	1931	4	2382	191900	2	0	94
837	8820204	.3053	2	4	1955	7	3287	191700	1.5	0	70
838	8803564	.1791	2	4	1938	7	2625	191600	2	0	87
839	8804531	.1997	2	4	1927	7	2344	191500	2	0	98
840	8807727	.2778	2	3	1951	7	2375	191500	2	0	74
841	8803437	.3237	1	3	1939	7	2361	190600	2	0	86
842	8802237	.5647	2	2	1951	4	1946	190300	1	0	74
843	8806724	.3099	2	3	1949	7	2014	190300	2	0	76
844	8807841	.2916	2	4	1952	7	2703	190000	1	0	73
845	8820114	.3168	2	4	1952	4	3455	190000	2	0	73
846	8835021	.3421	3	4	1962	7	3551	190000	1.5	0	63
847	8804291	.163	2	3	1926	3	1938	189800	1.1	0	99
848	8804421	.2594	2	4	1950	7	2479	189800	1.5	0	75
849	8807714	.3581	1	3	1948	7	2450	189700	2	0	77
850	8807767	.3696	2	4	1932	7	2329	189400	2	0	93
851	8807597	.4291	2	2	1955	7	2960	189300	1.1	0	70
852	8801711	.1722	2	5	1952	7	2607	189200	2	0	73

Summary Statistics

- Generally, "sum" can be used to summarize all variables, however it does not include median
- To include median, "tabstat, statistics()" can be used

```
. tabstat acres bathr bedr yearb walltype tla hvalue storiesn apartment age, statistics(mean median sd min max)
```

stats	acres	bathr	bedr	yearb	walltype	tla	hvalue	storiesn	apartm~t	age
mean	.4304181	2.183799	3.626276	1933.692	5.507148	2976.641	255352.5	1.592648	.0721579	91.30837
p50	.3237	2	4	1951	7	2763	213700	2	0	74
sd	.4725215	1.49303	1.610978	198.5468	2.257771	1290.961	147276.5	1.161631	.2588373	198.5468
min	0	-9	-9	-9	-9	-9	500	-9	0	12
max	6.3131	7	9	2013	7	9966	1437700	3	1	2034

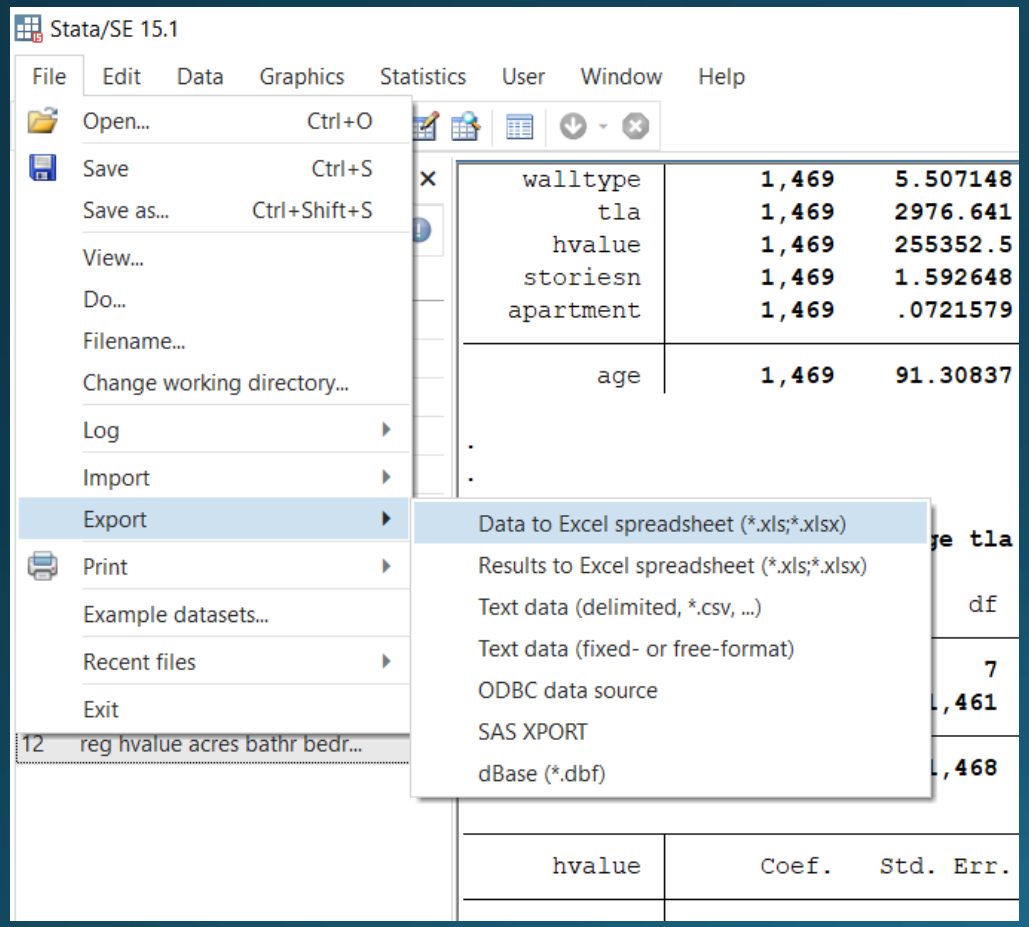
Regression

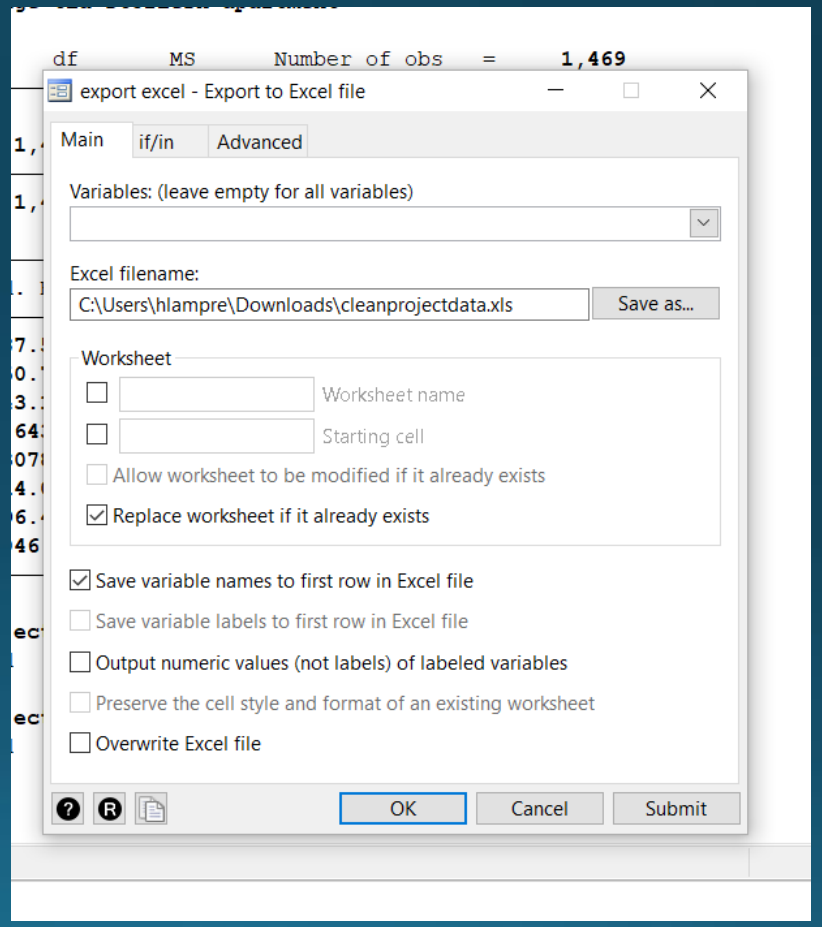
```
. reg hvalue acres bathr bedr age tla storiesn apartment
```

Source	SS	df	MS		
Model	2.6199e+13	7	3.7427e+12	Number of obs	= 1,469
Residual	5.6428e+12	1,461	3.8623e+09	F(7, 1461)	= 969.03
				Prob > F	= 0.0000
				R-squared	= 0.8228
				Adj R-squared	= 0.8219
Total	3.1841e+13	1,468	2.1690e+10	Root MSE	= 62147

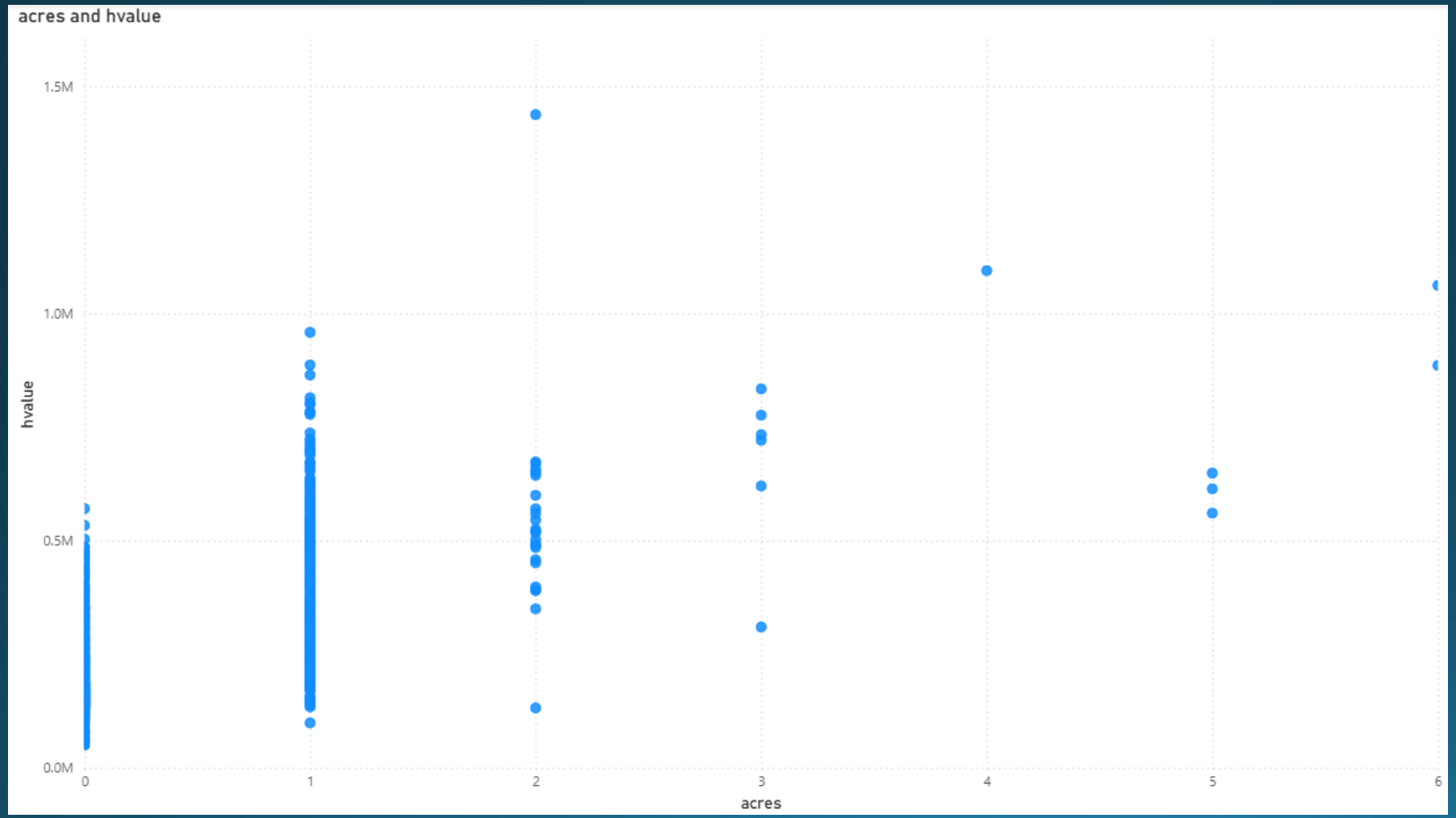
hvalue	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
acres	69937.21	4387.538	15.94	0.000	61330.66	78543.76
bathr	9080.64	2750.765	3.30	0.001	3684.77	14476.51
bedr	-13774.1	2543.142	-5.42	0.000	-18762.7	-8785.498
age	80.47655	24.64334	3.27	0.001	32.13645	128.8166
tla	84.87467	2.307853	36.78	0.000	80.34761	89.40173
storiesn	16046.85	4114.043	3.90	0.000	7976.792	24116.91
apartment	-41082.46	7696.453	-5.34	0.000	-56179.74	-25985.19
_cons	-27213.5	10046.58	-2.71	0.007	-46920.77	-7506.231

Exporting


Stata/SE 15.1
File Edit Data Graphics Statistics User Window Help
Open... Ctrl+O
Save Ctrl+S
Save as... Ctrl+Shift+S
View...
Do...
Filename...
Change working directory...
Log
Import
Export
Print
Example datasets...
Recent files
Exit
12 reg hvalue acres bathr bedr...
Data to Excel spreadsheet (*.xls;*.xlsx)
Results to Excel spreadsheet (*.xls;*.xlsx)
Text data (delimited, *.csv, ...)
Text data (fixed- or free-format)
ODBC data source
SAS XPORT
dBase (*.dbf)
walltype 1,469 5.507148
tla 1,469 2976.641
hvalue 1,469 255352.5
storiesn 1,469 1.592648
apartment 1,469 .0721579
age 1,469 91.30837
hvalue Coef. Std. Err.


df MS Number of obs = 1,469
export excel - Export to Excel file
Main if/in Advanced
Variables: (leave empty for all variables)
Excel filename:
C:\Users\hlampre\Downloads\cleanprojectdata.xls
Save as...
Worksheet
Worksheet name
Starting cell
Allow worksheet to be modified if it already exists
Replace worksheet if it already exists
Save variable names to first row in Excel file
Save variable labels to first row in Excel file
Output numeric values (not labels) of labeled variables
Preserve the cell style and format of an existing worksheet
Overwrite Excel file
OK
Cancel
Submit

Power BI



Conclusion

Pros	Cons
• User friendly • Great for beginners • Many capabilities for data analysis	• Cost • Not much flexibility for advanced programming • Graphing ability is subpar to R or Python

Economics/Data Analytics at UT

- Find an aspect of Econ/Data Analytics that you love and get good at it.