

Absolutely no late midterms will be accepted. You may hand it in early if you like.

Midterms are to be completed independently (no corroborating with other students), and to be emailed as a PDF. Ensure all questions are answered clearly. **You must include the following statement at the beginning of your midterm submission to acknowledge that you clearly understand the rules and consequences of the take home midterm:**

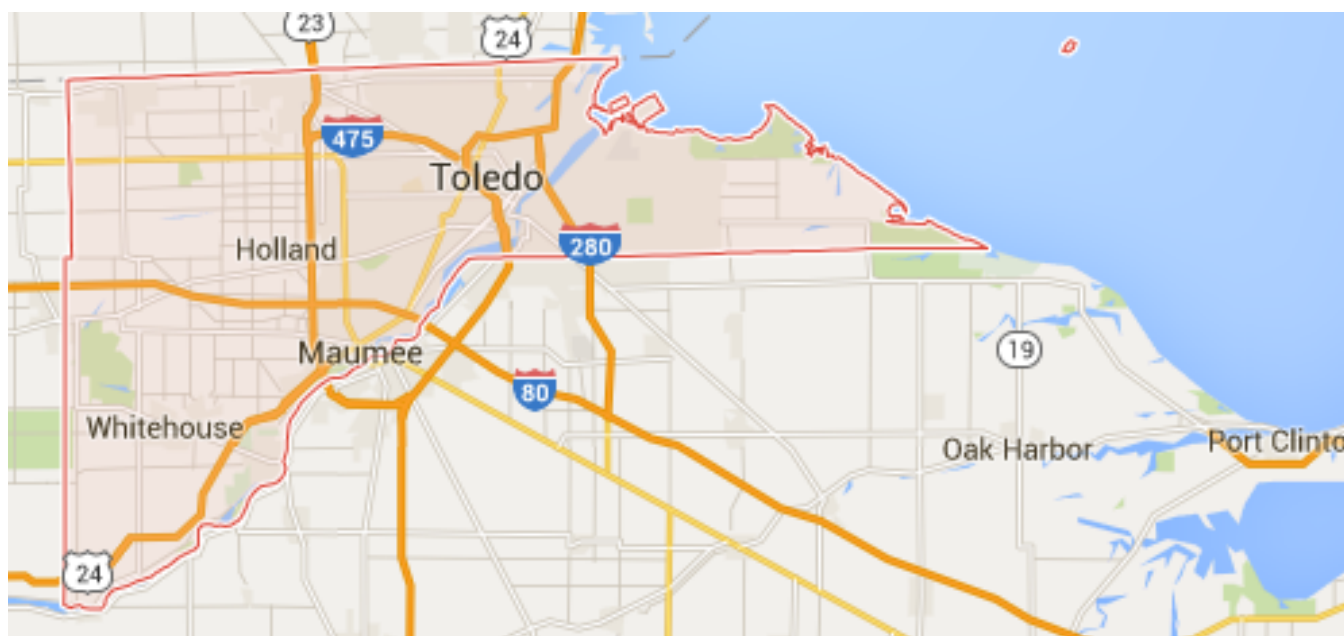
“I, [name], understand that this midterm is to be completed independently. I have asked no questions to any other human, with the exception of the TA or the professor, regarding this midterm. I understand that any suspicion of plagiarism or academic dishonesty will result in a zero grade for this midterm.
This statement and my name act as an understanding and compliance of the terms.
Signed, [name].”

You will be using data on single family homes sold in Lucas County, Ohio, between 1993-1998 as reported from the county auditor (see `house.dta` on Latte). The file is in Stata data format, so use,
> `library(foreign)` > `read.dta` command, or > `library(haven)` > `read.dta` command.

N.B. property assessment values are used to determine the amount of property tax the homeowner pays. See <http://co.lucas.oh.us/358/Real-Estate-Appraisal-and-Assessment> for more details. The important thing about the assessed value is that it is a completely deterministic process: the government takes publicly known features about the house (ex. square footage, lot size, number of windows, zip code, etc.)—all of which are in the dataset—plugs these into a formula, and the output is the assessed value (which is what property taxes are based on). There is no error term, things like “outdated kitchen” have no effect on the assessed value, but certainly do on the market price.

One degree of latitude is approximately 69 miles; one degree of longitude (at about 40 degrees latitude) is approximately 53 miles.

Map of Lucas County, OH:



1. Data Inspection and Statistical Inference:

- (a) How many variables and observations does the dataset contain? Which variables are dummy variables? Which variables are categorical variables (more than 2 categories)?
- (b) Show **price**, the dependent variable, as a histogram. Describe the distribution.
- (c) Determine the relationship between **sqft** and **price**, and **garage_type** and **price**.
- (d) Show **saledate** as a histogram (because it is a time variable, declare how you want your bins divided, `hist(saledate, "break")`, where **break** is days, weeks, months, or years). Is there any useful information here?

2. Model Selection and Output

- (a) Construct your model by theory and statistical inference. What are your determinants of sale price? Are there determinants that are missing from the dataset (you may need to you can create new variables). Present your regressors in a table and explain why you include them (include what your *expected* effect of each is on your price).
- (b) Are there any variables missing from the dataset? Which ones may cause omitted variable bias?
- (c) Run your regressions and diagnoses to determine if there are any OLS violations. Is heteroskedasticity present? Demonstrate how you came to your conclusion. If heteroskedasticity is present, use robust standard errors (`> library(sandwich) > coeftest(regressionname, vcov=vcovHC(regressionname, type = 'HC1'))`). Are there any other clear patterns?
- (d) What is your choice model(s). Defend your model selection. Present the results in a table. What corrections did you make to your initial theorized model, and why did you select it/them?
- (e) Interpret all your the coefficients (mainly dummy variables, and where you use natural log transformation, or polynomials).
- (f) Provide an explanation for any surprising or counterintuitive coefficients.

3. Hypothesis Testing (at the $\alpha_{max} = 0.05$ significance level) for your selected model.

- (a) What variables are statistically significant?
- (b) Interpret the R^2 , and the F-statistic.
- (c) Test the null hypothesis that only the physical characteristics of the house matter (*i.e.* jointly test if all variables you include other than the physical characteristics like square footage, number of bathrooms, etc. are equal to zero). Interpret the result.