Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**BA 253: ICE 8 – Zillow.com**

**Part 1**

1. Collect and list 10 data points from Zillow.com for single family homes in Durango. For each property, list the zestimate, number of bedrooms, number of bathrooms and square footage.
2. Graph **and print** just the square feet (as the x-variable) versus the zestimate (as the y-variable). Add the trend line and note the r2 value. Interpret the slope and the y-intercept. About how much is my house worth (at 1441 square feet)?
3. Re-do part b) but now forcing the y-intercept to 0. What does a square foot of house cost in Durango? What does this say about how much my house is worth?
4. Now, run multiple linear regression with the zestimate as the y-variable and all three other columns (beds, baths, sqft) as the x-variables. **Print this output.**
   1. Interpret both the r value and the r2 value.
   2. The “Significance F” is really the overall p-value. What does this tell us?
   3. List the multiple linear regression equation. Estimate what is my house worth – 3 beds, 1.5 baths, 1441 sqft.
   4. What are the individual p-values for the y-intercept and x-variables?
   5. According to the p-values and the 20% Huggins Rule, which variables are significant and should be included in the model and which should be eliminated? Do either beds or baths make the cut?
5. With your data, which makes more sense: the simpler linear model from part b) or the more complicated multiple regression model from part d)? Does the multiple regression yield enough additional accuracy to justify the additional complexity?

**Part 2**

1. As a class, figure out a way to put all of your data together into one big spreadsheet, eliminating any repeats. Good luck.
2. Repeat parts b) – e) above.