GEOG 400 Project 2 **Name:**

First step: Simple linear regressions. Again.

Check next to your hypothesized direction of association between each Xi and Y (direct or inverse):

Y and X1	Direct	Inverse
Y and X2	Direct	Inverse
Y and X3	Direct	Inverse
Y and X4	Direct	Inverse
Y and X5	Direct	Inverse
Y and X6	Direct	Inverse

Alpha you plan to use to test the null hypotheses, "there is no association between Xi & Y"

Alpha =

Justification for the alpha you picked:

Which associations turned out significant at your chosen alpha level?
And which of the **significant** associations had the same direction you predicted (direct or inverse)?

	Calculated P-Value	Is P-Value < alpha?	Do significant associations match your prediction?
Y and X1	Prob-value	Significant?	Predicted direction?

Y and X2	Prob-value	Significant?	Predicted direction?
Y and X3	Prob-value	Significant?	Predicted direction?
Y and X4	Prob-value	Significant?	Predicted direction?
Y and X5	Prob-value	Significant?	Predicted direction?
Y and X6	Prob-value	Significant?	Predicted direction?

Interpretation/speculations:

Kitchen sink multiple regression model

The two variables in your **best** simple linear regression model above:

Y		Xi	
Best model's R	*vs.*	Multiple regression R	
Best model's R2	*vs.*	Multiple regression R2adj	

By comparing movement in R2/R2adj, did you produce a much better explanation of Y with 6 Xis?

At four decimal places of accuracy, do you see a noticeable improvement in the significance of the model?

Whether it was worth your bother or not, please write down your model, showing co-efficients at **4** decimal places:

a	B1	B2	B3	B4	B5	B6

Y = _____ + _____X1 + _____X2 + _____X3 + _____ X4 + _____X5 + _____X6

Looking at the t-scores and prob-values for each variable in the kitchen sink model and comparing them with their corresponding values in the simple linear regressions you did earlier, which (if any) have significant p-values once they're allowed to interact?

X1	X2	X3	X4	X5	X6

Why is it that several of the variables are significant considered alone (simple linear regressions) but drop out of significance when they are put together in a common model?

Prune the model of all Xi variables that have P-values larger than your pre-selected alpha standard. Name them:

Refining the multiple regression model through backwards elimination:

Rerun the multiple regression, but ONLY with the Xi variables that have SMALLER P-values.
While R2 can be expected to decline, the key diagnostics are the changes in significance and in the F statistic that defines it.
Did F:

Increase	Decrease	Stay roughly the same

Write down your new model:

Y =

Now, re-examine the new t-scores and P-values. The new regression will have altered them from either the simple linear regression or the kitchen sink everything-in-it regression. Did the interactions among variables in the new model cause any of your Xi variables to generate P-values **higher** than your alpha standard (meaning they're disposable)?

Yes	No	Which?

Dump any Xi variable with effects that are no longer significant in comparison with your alpha standard. Now, rerun the regression.

Write down your newest model:

Y =

What happened to the new(est) R2adj?

Increased	Decreased	Stayed roughly the same

Dealing with an outlier

After doing your six scatterplots of each Xi on Y, identify the outlier record:

Redo the kitchen-sink multiple regression (all Xi variables at once) but with the outlier removed.

Write down this newest kitchen-sink masterpiece:

a	B1	B2	B3	B4	B5	B6
Y = _____ + _____	X1 + _____	X2 + _____	X3 + _____	X4 + _____	X5 + _____	X6

As before, refine through backwards elimination. Throw out all Xi variables with P-values above your alpha standard.

Write down the newest model.

Y =

Which of the two outlier-free models (kitchen-sink or the backwards pruned one) has the best F score/significance value?

Interpretation of the performance of your models, both with and without the outlier. Use another sheet, if necessary.