STAT460 – Homework 2 Due: Feb. 4 at the start of class.

1. These data record the level of atmospheric ozone concentration from eight daily meteorological measurements made in the Los Angeles basin in 1976. We have the 330 complete cases¹. We want to find climate/weather factors that impact ozone readings. Ozone is a hazardous byproduct of burning fossil fuels and can harm lung function. The data set for this problem is:

Variable Name	Definition
ozone	Log Maximum Ozone
vh	Vandenberg 500 mb Height
wind	Wind Speed (mph)
humidity	Humidity (%)
temp	Sandburg AFB Temperature
ibh	Inversion Base Height
dpg	Daggot Pressure Gradient
ibt	Inversion Base Temperature
vis	Visibility (miles)
doy	Day of the Year

- (a) Report the full linear regression of ozone on the other variables. Comment.
- (b) Report the selected variables using the following model selection techniques (use: either BIC, AIC, or Mallow's Cp)
 - i. All subsets (plot this using regsubsets, plot)
 - ii. Forward stepwise...
 - A. ... using the step approach
 - B. ... using the regsubsets approach
 - iii. Backwards stepwise (choose any method)
 - iv. Both stepwise (choose any method)
- (c) Compare the outcome of these methods with the significant variables found in the full linear regression in part (a)
- (d) Potentially, other transformations of covariates might be important. What happens if you attempt to do all subsets with the original covariates and their square? That is, for all covariates, put both

X and X^2

as possible terms.

¹Note that this dataset violates some assumptions of linear regression. Do you know which one(s)? For this assignment, ignore this fact.