The chapter on linear models (“Lining Up Our Models”) introduces linear predictive modeling using the tool known as **multiple regression**. The term “multiple regression” has an odd history, dating back to an early scientific observation of a phenomenon called “regression to the mean.” These days, multiple regression is just an interesting name for using linear modeling to assess the connection between one or more predictor variables and an outcome variable.

In this exercise, you will predict Ozone air levels from three predictors.

Please make sure you have included an attribution statement (see syllabus if you have questions).

1. We will be using the *airquality* data set available in R. Copy it into a dataframe called *air* and use the appropriate functions to summarize the data.
2. In the analysis that follows, *Ozone* will be considered as the outcome variable, and *Solar.R*, *Wind*, and *Temp* as the predictors. Add a comment to briefly explain the outcome and predictor variables in the dataframe using *?airquality*.
3. Inspect the outcome and predictor variables – are there any missing values? Show the code you used to check for that.
4. Use the *na\_interpolation()* function from the *imputeTS* package from HW 6 to fill in the missing values in each of the 4 columns. Make sure there are no more missing values using the commands from Step C.
5. Create 3 bivariate scatterplots (X-Y) plots for each of the predictors with the outcome. *Hint:* In each case, put *Ozone* on the Y-axis, and a predictor on the X-axis. Add a comment to each, describing the plot and explaining whether there appears to be a linear relationship between the outcome variable and the respective predictor.
6. Next, create a simple regression model predicting *Ozone* based on *Wind*. Refer to page 202 in the text for syntax and explanations of the *lm( )* command. In a comment, report the coefficient (aka slope or beta weight) of *Wind* in the regression output and, if it is statistically significant, interpret it with respect to *Ozone.* Report the adjusted R-squared of the model and try to explain what it means.
7. Create a multiple regression model predicting *Ozone* based on *Solar.R*, *Wind*, and *Temp*. **Make sure to include all three predictors in one model** – NOT three different models each with one predictor.
8. Report the adjusted R-Squared in a comment – how does it compare to the adjusted R-squared from Step F? Is this better or worse? Which of the predictors are statistically significant in the model? In a comment, report the coefficient of each predictor that is statistically significant. Do not report the coefficients for predictors that are not significant.
9. Create a one-row data frame like this:   
   *predDF <- data.frame(Solar.R=290, Wind=13, Temp=61)* and use it with the *predict( )* function to predict the expected value of *Ozone*.
10. Create an additional multiple regression model, with *Temp* as the outcome variable, and the other 3 variables as the predictors. Review the quality of the model by commenting on its adjusted R-Squared.