The Data

Bikeshare systems are popular in major cities around the world and are increasingly viewed as an important mechanism to reduce auto traffic, improve air quality, reduce use of fossil fuels, and improve the health of the population. It is particularly important to be able to predict the use level of the bikesharing system in order to plan maintenance, distribution of the bikes, and make decisions on whether and where to add more bike stations.

This data is from Washington, DC's Capital bikeshare system from the years 2011-2012.   Riders can rent bikes from one location and return to a different location in the system. There are two types of users: CASUAL users, who rent the bike for a one-time fee, and REGISTERED users, who pay a yearly membership fee in exchange for unlimited bike rental.

The data in the accompanying file **bikeshare.csv** (posted on Canvas) is publicly available at the UC Irvine Machine Learning Repository:

http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset

The file contains daily information on the number of trips taken. The variables in the data set are:

1. MONTH
2. HOLIDAY: whether the day is a US holiday or not
3. WEEKDAY: if day is neither weekend nor holiday WEEKDAY is YES, otherwise WEEKDAY is NO.
4. WEATHERSIT: The values are (1) Clear/Few clouds (2) Misty (3) Light snow or light rain (4) Heavy rain, snow, or thunderstorms
5. TEMP Normalized temperature in Celsius. The values are derived as $\frac{t-t_{min}}{t_{max}-t_{min}}$, where $t_{min} = -8$ and $t_{max} = +39$
6. ATEMP: Normalized "feels like" temperature in Celsius. The values are derived as $\frac{t-t_{min}}{t_{max}-t_{min}}$, where $t_{min} = -16$ and $t_{max} = +50$
7. HUMIDITY: Normalized humidity on a scale of 0 to 1.
8. WINDSPEED: Normalized wind speed on a scale of 0 to 1.
9. CASUAL: count of bikes rented by casual bikeshare users.
10. REGISTERED: count of registered users.
11. COUNT: count of total rental bikes including both casual and registered.

There are two goals. The first is to predict the variable COUNT as a function of the other variables. The second is to determine the effect of bad weather on the number of bikes rented.