36-315: Statistical Graphics and Visualization

Lab 12

Date: April 9, 2002   **Due: start of class April 15, 2002**

In this lab you will continue the final project by looking more closely at the relationships you found in lab 11. Specifically, you should look for *outliers* and *interactions*.

To explain outliers with graphs, bubble plots are very effective, as shown in class for the outliers in `PCTFEMAL`. Just add another variable to be the z variable and see if the outliers pop out. Another approach is to highlight the same outlier on many different scatterplots, using a different color or plotting symbol (see the `points` command).

To see interactions, the best tool is the contour plot, since the relevance of different variables can be immediately seen in the shape of the contours. This technique was discussed in class on Feb 20 and is summarized below:

- If the contours are straight parallel lines, then the z variable is a quasi-additive function of x and y, i.e. the x and y variables do not interact. This is independent of whether x and y are correlated.

- If transforming x and y makes the contours into straight parallel lines, then x and y also do not interact.

- If the contours are all horizontal, then x is irrelevant for predicting z when y is present. Vertical means y is irrelevant.

- If none of the above are true, then x and y are both relevant and their relative importance to predicting z changes as the contours change slope.

We saw that car prices followed the last case: horsepower and weight both affect price in general, but weight doesn't matter when horsepower is high. In using these rules, remember that contours should not be trusted where there is no data. (The latest version of `three.s` will only show contours within the range of the data.) If you have questions about contour plots, please ask the TA during the scheduled lab time.

Grading will be the same as for lab 11.