# 36-350: Data Mining

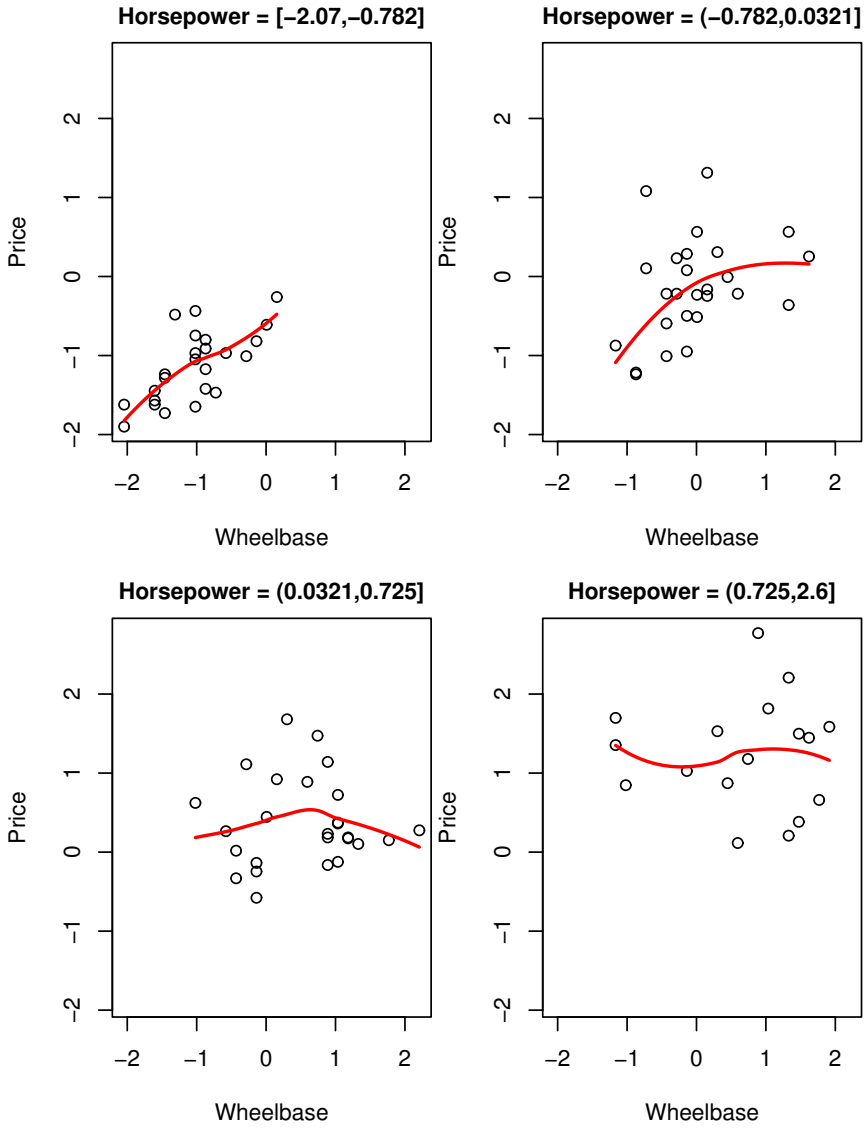
**Handout 13**
**October 8, 2003**

---

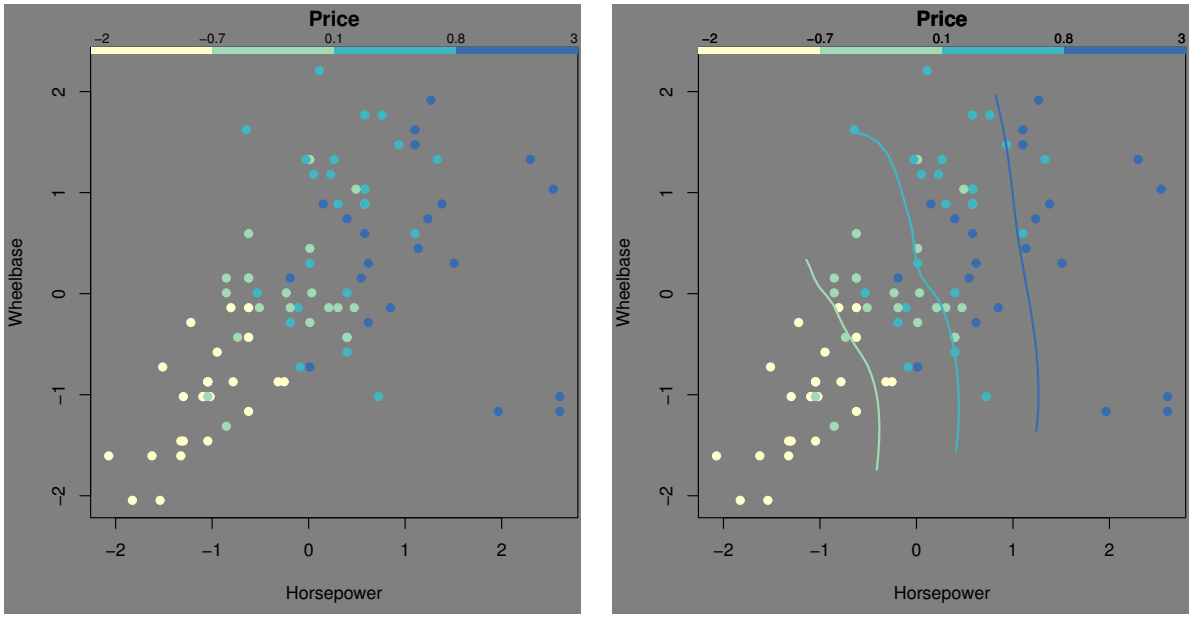
Visualizing interactions with contour plots

If $y = f(x_1) + g(x_2)$ for some functions $f$ and $g$, then the relationship is **additive**. For example, linear relationships are additive, as well as $y = x_1^2 + x_2^2$. Otherwise, there is an **interaction term**, such as $x_1 x_2$ or $\max(x_1, x_2)$. Slice plots reveal interaction terms.

For car price, there is an interaction term of the form $(\texttt{Horsepower} < 0) * \texttt{Wheelbase}$. It would be hard to guess this without the slice plot.

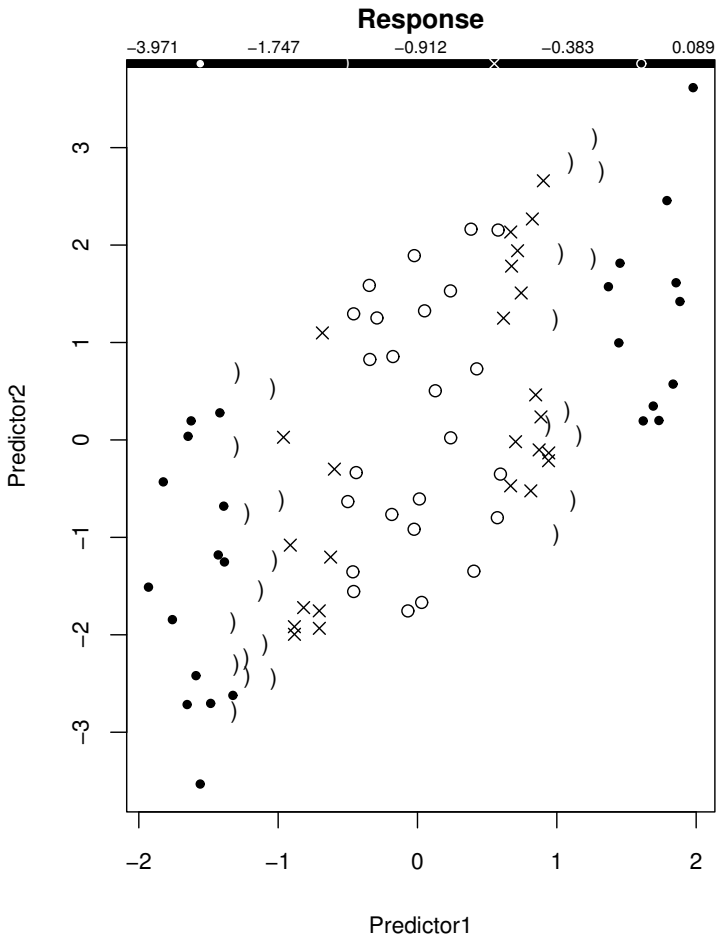


**Color plot**—Predictor 1 versus predictor 2, with points colored according to the response. You are looking at the slice plots from "above" rather than from the "sides." Each row of points is a slice.

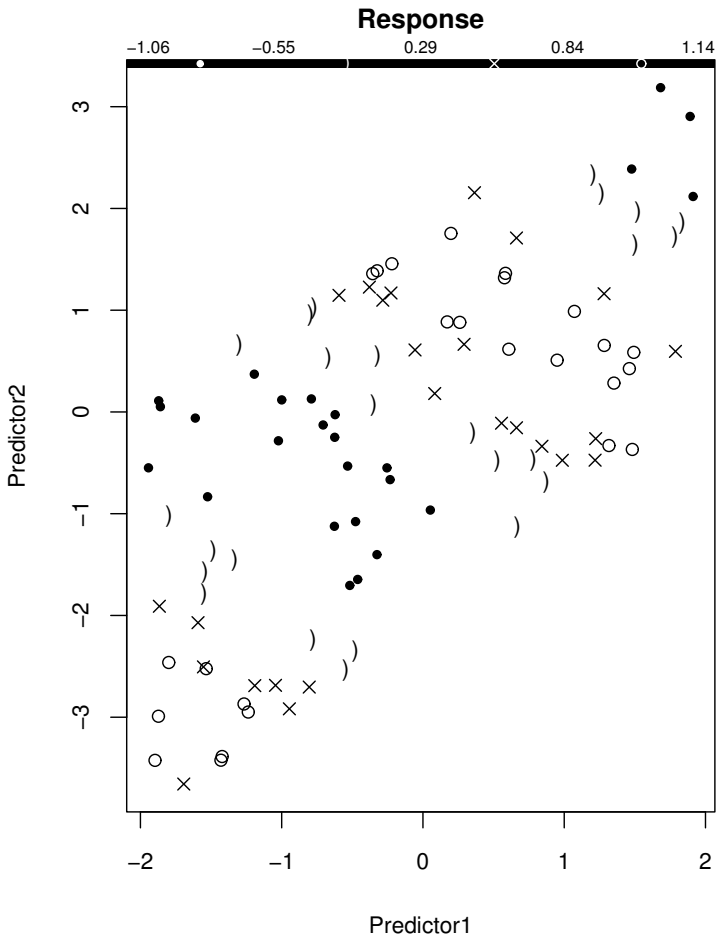



As you change a predictor, look at how the *average* response varies, in other words, the average color in the region. Contour lines (on the right) delineate regions of different average color, allowing you to see the trends more clearly. When you cross a contour line, the average response is changing, so the predictor is important. As you move along a contour line, the average response is not changing, so the predictor is not important.

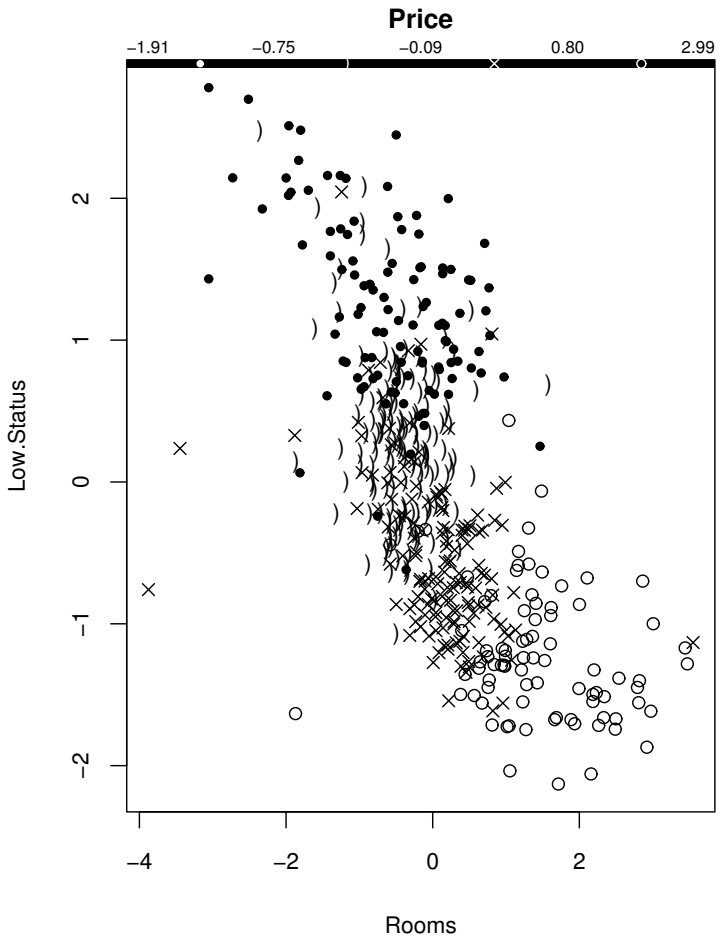
If the response is linear in the predictors, then the contour lines will be straight with the same slope. If the response is nonlinear, or there is an interaction term, then the contours will bend. (Compare slices to see which case you are in.)



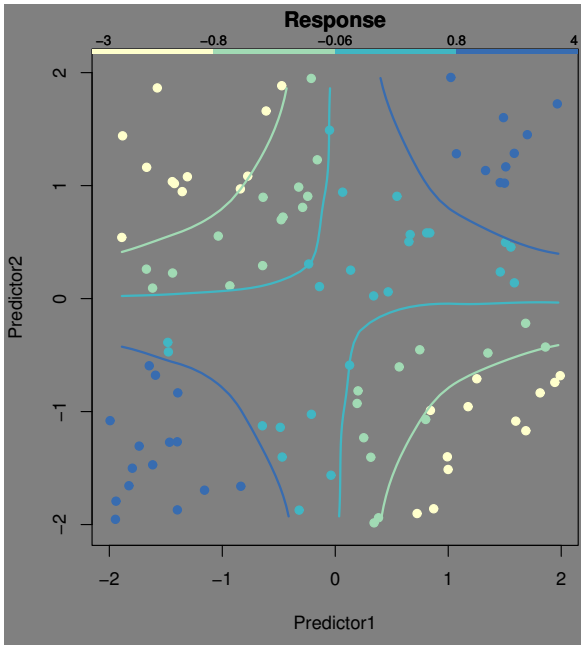

1. (a) Draw smooth boundaries that approximately separate neighboring response groups.
   (b) Is Predictor1 sufficient to separate the groups?
   (c) Is Predictor2 sufficient to separate the groups?
   (d) Is an interaction term needed?
   (e) Write down a possible equation for the response in terms of the predictors.



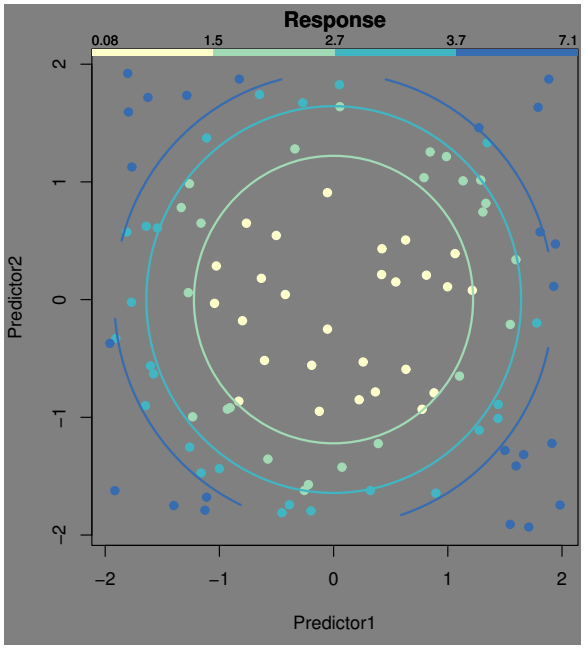

2. (a) Draw smooth boundaries that approximately separate neighboring response groups.
   (b) Is Predictor1 sufficient to separate the groups?
   (c) Is Predictor2 sufficient to separate the groups?
   (d) Is an interaction term needed?
   (e) Write down a possible equation for the response in terms of the predictors.




3. (a) Draw smooth boundaries that approximately separate neighboring response groups.
   (b) Is an interaction term needed?
   (c) When is Predictor1 important? When is Predictor2 important?




Interaction term ($y = x_1 x_2$)




No interaction term ($y = x_1^2 + x_2^2$)

A **contour plot**, like an elevation map, depicts a surface by drawing lines where the value is equal—isocontour lines.

These plots are made by computing a **prediction surface** and then tracing paths of equal predicted response. The surface is estimated by local regression, just like trend lines, which can make the contours noisy. In particular, they tend to "flare" at the edges.

(Return to handout12)