36-315: Statistical Graphics and Visualization

Lab 14

Date: April 22, 2003 **Due: end of lab**

By now, you are familiar with the pairwise relationships between your predictors and the response. In this lab, you will try to find other predictors that may be relevant, rank them by strength, and spot interactions between them.

In this lab, you are given the commands to type in, but with question marks (`??`) denoting things you must provide, based on the dataset you are using.

1. Download the files for this lab from the course web page.
2. Open a Word document to record your work (plots and commands).

Start R

3. `Start -> Programs -> Class software -> R 1.5.1`
4. Set the working directory to `My Documents`:

   `File -> Change dir...`

5. Load the special functions for this lab:

   ```
   source("lab14.r")
   ```

   The file `variables.r` contains a suggested set of variables and transformations, in a formula called `goodvars`. It has been changed a bit from the previous lab.

   ```
   source("variables.r")
   ```

Load your data

6. From the class web page, download the census data files for your state. Unzip them into `My Documents`. Then they can be loaded via

   ```
   a = read.csv("tracta.csv")
   b = read.csv("tractb.csv")
   frame = clean.census.data(a,b)
   ```

Projections

7. A discriminative projection can help find new variables relevant to the response. The first step is to identify all tracts whose response value is above (say) the 80% quantile:

```
y = frame[,??]
b = quantile(y,0.8)
i = which(y > b)
```

Now separate these tracts from the rest:

```
x = model.frame(goodvars,frame)
x.not.y = not(x,??)
w = separate(x.not.y,i,"high response")
```

One variable has to be removed first for this plot to make sense. Which one? Make a list of the important variables thus found.

8. A regression projection tries to separate all levels of the response, and is useful for highlighting interactions. The first step is to make a data frame containing the response and transformed versions of the important predictors:

   ```
   x = model.frame(? ~ ? + ? + ...,frame)
   ```

   Next perform the scaling, projection, and plotting steps described at the end of the lab. The strongest predictors are the ones most perpendicular to the color boundaries. *Which are the strongest predictors?*

9. If the color boundaries change direction at all in this plot, then there are interactions. The orientation of the boundaries with respect to an axis tells you how the importance of a variable is changing. *Are there any interactions between predictors? Which variables are changing strength?*

Contour plot

10. Pick two predictors that appear to have the strongest interaction. Fit a smooth prediction surface and make a contour plot. The commands should be something like:

    ```
    fit = smooth(? ~ ? + ?, frame, span=??)
    color.plot(fit)
    ```

    Which variable is stronger? What is the interaction, and is it surprising?

11. Show us your graphs.

Regression projection Let `x` be a data frame with a designated response variable. A regression projection can be made via the following steps:

```
sx = scale(x)
w = projection(sx,k=2)
px = project(sx,w)
color.plot(px)
plot.axes(w)
```

The projection weights are placed in `w`. The projected dataset itself is in `px`. (The `separate` function performs the same steps, but with a special response variable designating the groups to separate.)