36-315: Statistical Graphics and Visualization

Lab 7
Date: February 26, 2002 **Due: start of class March 4, 2002**

---

1 Introduction

The purpose of this lab is to learn principles of graphical design. The task is to redesign your plots from lab 4, and practice judging the quality of a graph. To judge a graph we will focus on the three tasks of visual perception: detection, assembly, and estimation.

2 Detection

Detection is seeing the data. To enable detection in a scatterplot, the symbol representing each point should be prominent enough to be seen compared to other parts of the graph like connecting lines, labels, or a grid. On the other hand, points should not overplot and obscure each other. You want to minimize clutter.

The plotting symbol used by S-PLUS is controlled by the parameter `pch`, used in lab 6. Symbols are specified by numerical or character code. To see the numerical codes available, type this:

```
n <- 28
plot(1:n,rep(1,n),type="n")
for(p in 1:n) {
  points(p,1,pch=p,cex=2)
  text(p,0.9,p)
}
```

In S-PLUS 2000, the codes go up to 28. In S-PLUS 6 for Unix, the codes go up to 19. In R, the codes go up to 23. You can also specify a character to use as a symbol, like `pch="."` or `pch="+"`. Remember that an open circle is usually the best choice, if there is only one group of data on the plot.

The size of the plotting symbol is controlled by the parameter `cex` (**c**haracter **ex**pansion). A value of 2 means twice as big, 0.5 means half as big. For example:

```
plot(1:100,sqrt(1:100),cex=0.5)
```

Unfortunately this will also scale the axis labels. To scale only the points, first draw the axes using `type="n"`, then draw the points on top, like so:

```
x <- 1:100
y <- sqrt(1:100)
plot(x,y,type="n")
points(x,y,pch=1,cex=0.5)
```

3 Assembly

Assembly is perceiving groups and patterns in the data. In a scatterplot, assembly is aided by drawing a smooth curve through the cloud. When different things are superposed on the same plot, assembly is aided by choosing plotting symbols, line styles, and colors that are easy to discriminate.

4 Estimation

Estimation is perceiving magnitudes and differences. Estimation is influenced by several things. One is the type of encoding: position, length, area, etc. In a scatterplot, the encoding is positional. There is also the choice of scale: zooming in enhances estimation. Besides the plotting limits, scale is also controlled by aspect ratio, which can enhance estimation by maximizing the change in curve direction. The third influence is transformation. Transforming for symmetry and transforming for linearity both aid the perception of trends.

5 Making your own plots

For each pair of variables below, make a scatterplot which best shows their relationship. Choose the x and y variables carefully so that a lowess curve is reasonable.

1. MEDHHINC vs. POPPSQMI
2. MEDYRBLT vs. MEDHHINC
3. PCTVACNT vs. MEDRENT
4. MEDHHINC vs. PCTELEM

Your plots should improve on the ones you made in lab 4, by incorporating comments from the grader as well as principles of design discussed in class and repeated above. Essentially you are being given the chance to redo that lab. You do not need to identify outliers on these plots. You should include the plot and your code for the plot. You will be graded on how well your plot meets the criteria for good detection, assembly, and estimation. Do not try to use `set.aspect` for the aspect ratio; set it by hand.

> **Question 1:** For each plot, briefly describe how your plotting choices facilitate the tasks of detection, assembly, and estimation.

> **Question 2:** Interpret your plots. What are they saying about your state?