

36-315: Statistical Graphics and Visualization

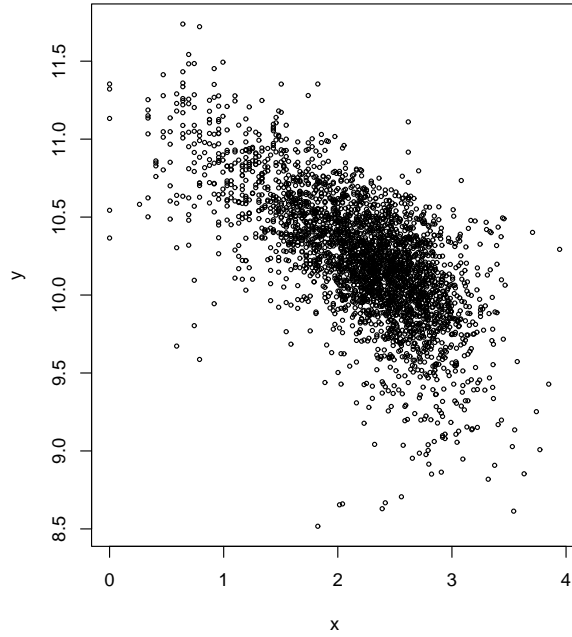
Homework 5

Date: February 11, 2002

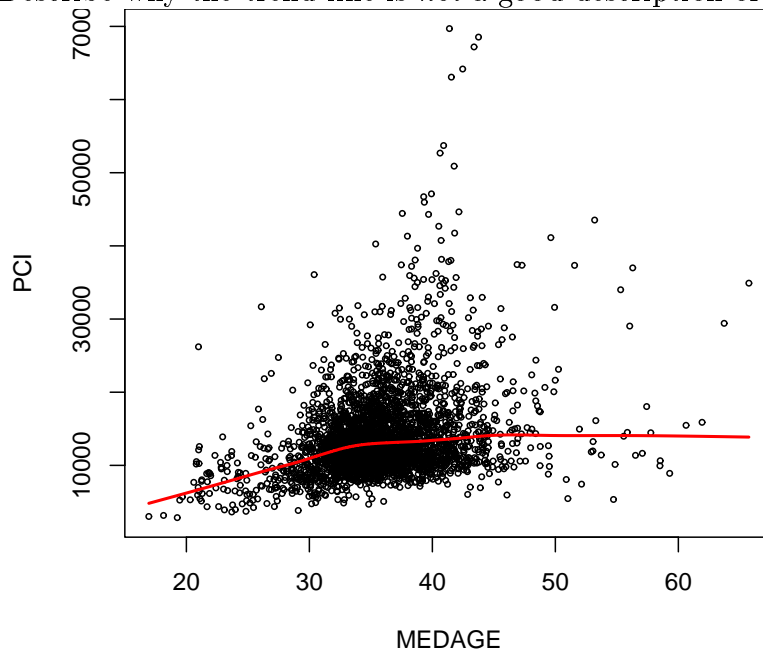
Due: start of class February 17, 2002

1. Download `lab5.csv` and source `lab5.r` just like in lab. In this problem, you will make some scatterplots, which will be graded according to the criteria given in class, especially aspect ratio and symbol size. Note that when you copy a figure into Word or print it out, the aspect ratio and symbol size may change. (In most cases, though, you shouldn't have to alter the default plot.)
 - (a) Make a scatterplot, without transformation, to predict per-capita income from population density `POPPSQMI`. What trend is visible on this plot?
 - (b) Transform the variables to remove skew and add a trend line. Adjust the smoothness of the trend line so that it follows the data correctly on the left side. Describe the trend. What areas of Pennsylvania would naturally correspond to low, medium, and high population density?
 - (c) What population density has the greatest disparity of income? What population density has the least disparity of income?
 - (d) Transform the variables so that differences of income on the plot correspond to percentage differences instead of absolute differences. What population density has the greatest disparity of income on a percentage basis?
 - (e) Describe the sort of aspect ratio one would use in order to obscure the differences in income with respect to population density. Describe the sort of aspect ratio one would use in order to emphasize the differences in income with respect to population density.
 - (f) Now flip the axes of the plot and draw a new trend line to predict population density from income. Which is more predictable: income from density or density from income?

2. On the scatterplot below, draw a trend line which predicts y from x , using the rule given in class.



3. Below is a scatterplot of income versus age in Pennsylvania census tracts, with a trend line on top. Describe why the trend line is *not* a good description of the relationship.



4. The scatterplot depicting “The Pace of City Life” (Handout 8) has a number of interesting features to admire. List as many features as you can find which make this an effective and readable plot. (Hint: in addition to the checklist given in class, notice the grid lines, how the symbols touch each other and the trend line, and the placement of labels. Compare to, for example, the scatterplot of Florida counties in Handout 8.)
5. In a paper on particle physics, a scatterplot was given depicting three groups of particles. The plot is reprinted on the next page. On the following page is one possible revision of the plot, which significantly reduces the clutter and improves comprehension. In order to understand how this was done, make a list of all the specific changes you can find in the revision which would have the effect of reducing the clutter and enhancing the data in the plot. (Note that not all of the changes actually do this.)
6. The mathematical properties of coastlines are nicely illustrated in a scatterplot in B. Mandelbrot’s book, “The Fractal Geometry of Nature”. In this plot, reprinted on the last page, the author is trying to show that data corresponding to each coastline and land-frontier is linear on a log-log scale.
 - (a) List some desirable and effective features of this plot. (For example, notice how the labels are drawn).
 - (b) List some undesirable features of this plot.
 - (c) How does the aspect ratio used in this plot compare to the rule given in class? What crucial change could arise from using a different aspect ratio?