36-315: Statistical Graphics and Visualization

Lab 5

Date: February 11, 2003   **Due: end of lab**

In the previous labs, you examined income in cities and percentage of women in cities. Now you will put them together. Is the reduced income due to the wages that women earn, or some other factor?

Interspersed throughout this lab are some useful thought questions. You will be asked about them at check-off.

1. Download the files for this lab from the course web page.

2. Open a Word document to record your work.

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("lab5.r")`

Load the data

6. `frame = read.csv("lab5.csv")`

   `frame` contains three census variables, measured for each census tract in Pennsylvania:

PCI	Per-capita income
PCTFEMHE	Percent of households which are female-headed
PCTUNEMP	Percent unemployed 16+

PCI vs. PCTFEMHE

7. Make a scatterplot to predict income from female-headed households. *Are there any bivariate outliers on this plot?*

8. Make another scatterplot with a trend line. *Is the trend line straight or curved?*

9. Using the mouse, resize the plot window so that it is very wide and short. *Does this alter your impression of the trend line?*

10. Now resize the plot window so that it is very thin and tall. *Does this alter your impression of the trend line?* What you've done in the last two steps is change the **aspect ratio** (height divided by width) of the plot. This can also be done using the `asp` option.

11. Make the window normal size again (using `Windows->Tile`). Make another plot with trend line, but with span 0.1 (less smoothness). *Does this trend line tell you anything new?*

12. Zoom in on the new feature by making the x axis range from 0 to 10. Once again, resize the plot window to wide and short, then thin and tall. *How is your perception of this feature altered by the aspect ratio?*

PCTUNEMP vs. PCTFEMHE

13. So you established the relationship between income and female-headed households. Is this due to wages? Make a scatterplot of unemployment versus female-headed households. *Is one predictable from the other?*

14. The data in this plot is highly skewed on both dimensions. Use a log transformation to de-skew them. Because the variables are sometimes zero, add one first. Thus the transformation is `log(PCTUNEMP+1)` and `log(PCTFEMHE+1)`. *What new features show up in this plot? What values of* `PCTUNEMP` *and* `PCTFEMHE` *do they correspond to?*

15. Show us your graphs.