36-315: Statistical Graphics and Visualization

Lab 2
Date: January 21, 2003 **Due: end of lab**

In this lab, you will try out some modern alternatives to the histogram. 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:

   `http://www.stat.cmu.edu/~minka/courses/36-315/lab/lab2/`

   Save in the `My Documents` folder.

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 by typing in the Command Window:

   ```
   source("lab2.r")
   ```

   and press enter. If this doesn't work, the file may have been renamed to `lab2.r.txt` when you downloaded it.

6. Arrange the various sub-windows by selecting the menu option `Windows -> Tile`.

Load the first dataset

7. Read the first csv file into an R data frame, by typing into the command window:

   ```
   frame = read.csv("lab2.csv")
   ```

8. Extract the column `POPPSQMI` from `frame`:

   ```
   x = frame[,"POPPSQMI"]
   ```

   The data is now a vector of numbers called `x`. It is the population density data from lab1, but on a log scale.

Make plots for the first dataset

9. Make a histogram of `x` using the default number of bins.

10. Using the handout on the `density.curve` function, plot a density curve for `x`. You do not need to create a new window because the plot should appear in the same window, below the first one. Notice the difference from the histogram.

11. Now make a histogram using 50 bins, and below it a density curve with adjustment 0.2 (making the kernels narrower by 0.2). Hence the density curve parallels a histogram in some ways. Save this plot by pasting into your Word document.

Jitterplots

12. Make a jitterplot of `x` with default settings. In this plot, the x values are given by the data, but the y values are random. Can you tell that the x values are normally distributed? Do any points look like outliers?

13. Run jitterplot again, and you will get a second version below the first. Notice it is somewhat different. Save the plots.

14. Make a third jitterplot, but with scale 0.5. Then make a fourth jitterplot, with scale 1 but symbol size 2. Do either of these changes help your perception of the bimodal distribution? Do they change your perception of the outliers? Save the plots.

Histodot plots

15. Make a histodot plot via

    ```
    histodot(x)
    ```

    This plot is a cross between a histogram and a jitterplot. The x values are given by the data and the y values are chosen randomly to space out the symbols, but only enough to prevent overlap. The result is vertical stacks of dots. Notice how the bimodality and the outliers are represented in this display.

16. Make a second histodot plot and see how the result varies. Save the plots.

17. To get a better idea of how this plot works, you may want to try it on only the first 500 tracts:

    ```
    histodot(x[1:500])
    ```

18. You should now have saved four windows of two graphs each. Show us the graphs.