36-315: Statistical Graphics and Visualization

Lab 4
Date: February 4, 2003 **Due: end of lab**

Besides income and rent, cities also have a different family structure, as you will discover in this lab. 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("lab4.r")
   ```

Load the data

6. The data is in a special R format, so you use `load`:

   ```
   load("lab4.rda")
   ```

   This defines two variables: `frame` (a matrix) and `fam.vs.density` (a table). `frame` contains five census variables, counting different types of families across Pennsylvania:

MCFAMS	Married couple
MCWCHILD	... with related children
FEMHEAD	Female householder with relatives but no husband present
FEMHEADC	... with related children
NONFHHS	Householder has no relatives present

   ("Male householder with relatives but no wife present" is also counted by the Census but not included in this dataset.)

7. Extract the family type variables from `frame`:

   ```
   x = frame[,1:5]
   ```

State totals

8. The first set of plots will involve the totals for the state. In `x` you have values for each census tract. Add them up via

   ```
   s = colSums(x)
   ```

   Type `s` to see that it is now a vector of named values.

9. Make a pie chart of `s`, with radius small enough that the labels are visible. To test how readable this plot is, can you tell which is bigger: `MCWCHILD` or `NONFHHS`?

10. Now make a bar chart and dot chart of `s`. Is it easier to answer the question?

11. Make a dot chart of `log(s)`. Differences on this plot correspond to percentage differences in the totals. *Is the percentage difference between* `FEMHEAD` *and* `FEMHEADC` *the same as between* `MCFAMS` *and* `MCWCHILD`*?*

City vs. country

12. `fam.vs.density` is a matrix similar to `s` but with columns for low, medium, and high population density. Each column has a different total number of families, as you can verify with `colSums(fam.vs.density)`. To normalize them, compute the probability of each family type within each column:

    ```
    p = row.probs(fam.vs.density)
    ```

    The columns now each have sum 1.

13. Using one call to `pie`, make pie charts of the three columns of `p`. To test the readability of the plot: *Is the proportion of families with children (*`MCWCHILD` + `FEMHEADC`*) increasing or decreasing with population density?*

14. Now make a stacked bar chart. Is it easier to answer the question?

15. The proportions can also be encoded with shading. Make an image plot of `p` via

    ```
    image(p)
    ```

    In this plot, darker means larger. *Which family types increase with population density, and which decrease?*

16. The rows and columns of `p` can be switched with the function `t(p)`. Make two bar charts with juxtaposed bars: one of `p` and another of `t(p)`. *Which one makes the stronger point?*

17. Make a line chart of `p` and compare to juxtaposed bars. (A line chart is simply a dot chart with lines connecting data in the same row.)

18. Show us your graphs and answer this question: How would you personally rank the different encodings (angle, length, position, shading) used by these plots?