

- 1 - Ingestion dose factor
- 2 - Milk transfer factor/individual cow
- 3 - Interception parameter
- 4 - Integrated deposition
- 5 - Date cows taken off pasture
- 6 - Inhalation dose factor
- 7 - Breathing rate
- 8 - Ingestion rate of fresh fruit

Gaps in pie charts account for the fraction of uncertainty not described by linear regression.

Figure 2: Relative importance of parameters for Iodine-131 dose to the Thyroid of a child raised on milk from a pasture-fed cow, 1945; adapted from Figure 5.5, Farris *et al* (1994).

Other parameters which may be important, depending on the RI, are the rate at which a cow will transfer ingested radionuclides into her milk, the rate at which a human will absorb inhaled radioactive particles and the rate at which airborne particles will settle to the ground.

Conclusions

The modeling of radionuclide transport, exposure and dose in a wide number of scenarios over a span of 50 years across a large geographical area was an ambitious effort. The careful analyses of uncertainty and sensitivity made it possible to produce useful and meaningful results. Models are currently being modified to provide dose estimates and uncertainty assessments for specific individuals whose characteristics differ from those of the RI's defined for the HEDR project.

References

Farris, W.T., Napier, B.A., Ikenberry, T.A., Simpson, J.C. and Shieler, D.B. (1994), "Atmospheric Pathway Dosimetry Report, 1944-1992." PNWD-2228 HEDR, Draft, Battelle Pacific Northwest Laboratory, Richland Washington.

Shieler, D.B. and Napier, B.A. (1992), "HEDR Modeling Approach." PNWD-1983 HEDR, Battelle Pacific Northwest Laboratory, Richland Washington.

Simpson, J.C. and Ramsdell, J.V. Jr. (1993), "Uncertainty and Sensitivity Analyses Plan." PNWD-2124 HEDR, Draft, Battelle Pacific Northwest Laboratory, Richland Washington.

Jeremy C. York
Battelle Pacific Northwest Laboratories
jc.york@pnl.gov



TOPICS IN SCIENTIFIC VISUALIZATION

Using Gray in Plots

by Dan Carr

In the last newsletter I discussed color. A point that I forgot to make was that colors can get hard to discriminate when the colored regions become very small. The newsletter editors reminded me of this fact by reducing the area of my choropleth residual plot by a factor of ten. Perhaps if I keep this article short the key plot in the article will be full sized. (*We too were surprised at and regret the quality of the color plots in the last issue, due to their size. As we gain experience with color plots, from YOUR submissions, we hope to improve the results. Eds.*)

This article continues on the topic of using gray in plots. Figure 1 shows a summary of the EPA's Toxic Release Chemical Inventory (TRI) for 1987 and provides the discussion example. Figure 1 is a plot that reexpresses most of a visually intimidating two-page table in Courteau 1990. In Figure 1, the basic patterns can be seen at a glance. It doesn't take long to find the two states with the greatest totals. A second glance indicates that the dominant releases for the two states go underground rather than being emitted to the air, stored on land, transferred to other sites, etc. Several aspects of the Figure 1 design can be called out: background grids, panel sizing to maintain comparability and preserve resolution, separate scaling of the margin total panel and its placement, sorting of states and panels by margin totals, staggered tic labels, and general labeling. Carr (1994) discusses these design aspects in an extended treatment of row-labeled plots. In this article attention focuses on design features related to gray and on plot interpretation.

Grids are back! Grids used to be common when people plotted on gridded graph paper. Historically Tukey (1977), Tufte (1983) and others have noted that heavy grids interfered with seeing data and hence interfered

with geometric pattern perception. Soon grids disappeared entirely as some used tracing paper to improve hand drawings and many began generating plots using computers. This was taking the guidance too far. Recently Cleveland (1993a, 1993b) has demonstrated that background grids improve perceptual accuracy of extraction and help in making comparisons. The basic question now is how best to produce gentle background grids.

Grids are back! Grids used to be common when people plotted on graph paper.

The grid lines in Figure 1 are white lines on a light gray background (assuming decent color reproduction). The Washington Post commonly uses this approach. Cleveland's plots use gray grid lines on a white background. I have a slight preference for white lines on a gray background since the gray background reduces the contrast and gives the plot more of a value-added appearance. With light gray and white both methods provide unobtrusive reference grids.

In terms of production, laser printers that provide half-toning will produce gray regions and gray lines. Unfortunately low resolution half-tone gray appears sloppy and half-tone gray does not reproduce well given commonly accessible copying technology. What one really wants is gray ink. Some elegant gray ink examples can be found in Tufte (1983, 1990) and Grant (1993). While the current emphasis on black and white plotting will diminish over time as did the emphasis on black and white television, in the short term monochrome laser printing will have do for routine graphics. As for a modest number of copies, the best bet is to generate new originals from the electronic version.

Returning to the topic of grid lines, the horizontal grid lines in the plot help the eye match the state names with the bars in the different panels. The spacing after every fifth state creates smaller perceptual units (perceptual grouping is important, see Kosslyn 1994) that help in the matching process. With grid lines and grouping, the error rate for matching names to bars should be low. The vertical grid lines provide a basis for fairly accurate comparison of bar lengths. Note that in the right six panels the grid line spacing has the same units and the same physical size. Thus bar lengths in the panels are directly comparable and the grid lines make comparisons more accurate. The panels have differing widths to cover the differing ranges of the data. The margin total panel on the left has its own separate scale. The design separates this panel from the other plots (slightly)

to call attention to the fact. Also the design uses black bars in the panel to further emphasize that the scale and the vertical grid lines are different.

The symbols chosen for the display are bars. Bars, like most area filling symbols are visually pronounced. Patterns among the bars are easily seen despite separating panel lines. Unfortunately regularly spaced black bars on a white background can introduce the moire effects as described by Tufte (1983). A bit of this may be observed in the "Total" panel. The gray bars in the remaining panels diminishes the contrast with the background. This makes it easier to really look at the bars.

A weakness in the display scale is the inability to distinguish the small values from zero. Shifting the location of zero to the right of the panel edge would allow positive values to show as a thin line but this costs in terms of plot resolution and makes plot look busy. Producing a dot plot on a log scale would provide more resolution for the smaller values but is harder for many people to interpret. The bar plot as it stands focuses attention on the large values and they are the basic message.

Truth in labeling for plots seems reasonable in an age of information consumption.

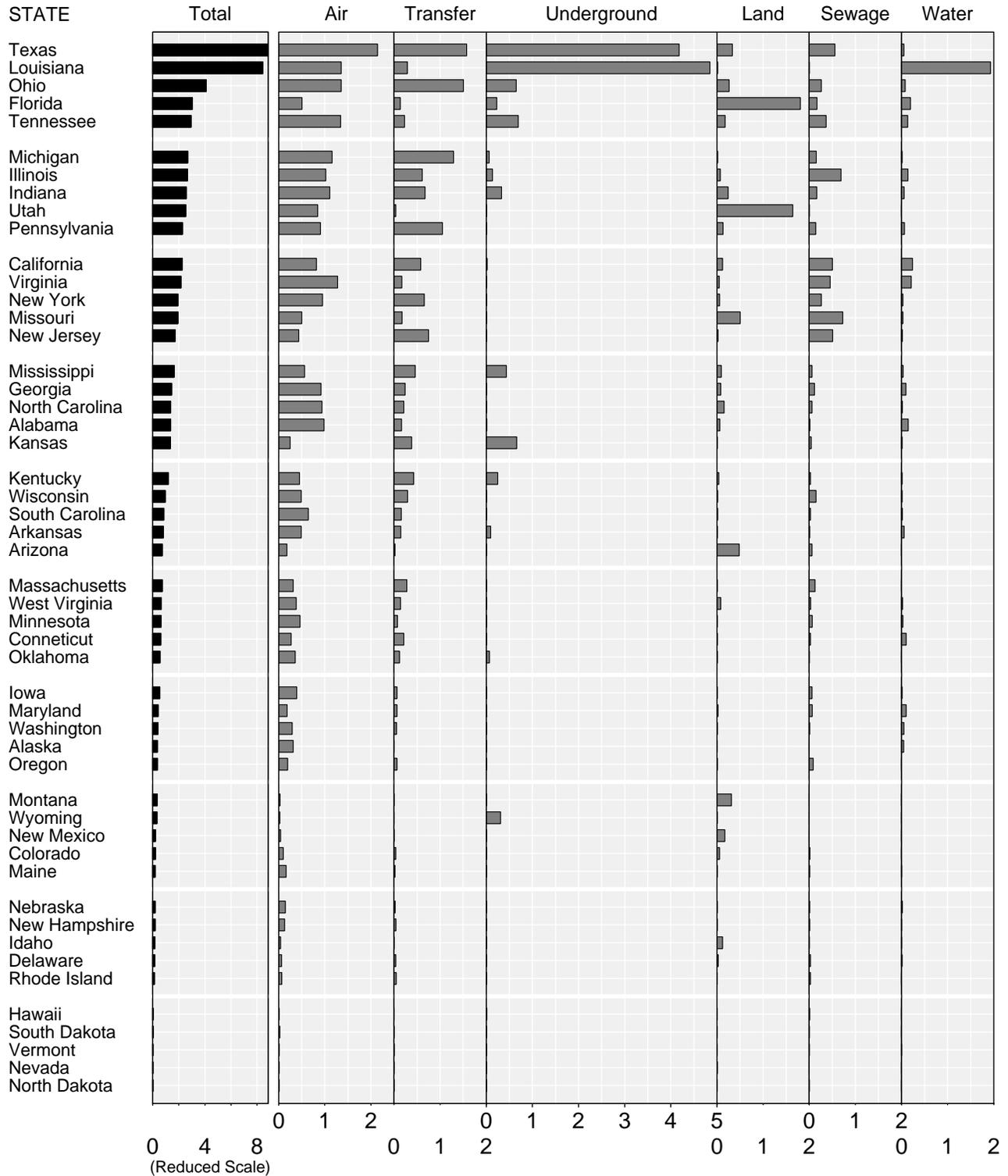
Unfortunately the interpretation of this graphic is limited by the data. The total of seven billion pounds of toxic materials being released or transferred may seem more than enough, but actually represents only a fraction of the unknown total. The regulations behind the Toxic Release Chemical Inventory only requires reports from not-small companies with certain SEC codes. In 1987, one-third of the companies who were obligated to report did not do so. The company reports are self-assessment estimates. While larger companies have environmental engineers to fill out the reports, in some cases the task falls to secretaries. Further, summarizing in units of pounds is convenient but hides the differing toxicities of the chemicals released. Figure 1 represents a summary of the information collected. The connection of the summary to the state of the nation is tenuous except as providing guesstimated lower bounds.

Metadata should be attached to plots like this to provide an appropriate context for interpretation. A significant problem is that while government documents often do an excellent job of providing the context in writing, the plots are often at risk of being extracted and used to influence public opinion and public policy. Truth in labeling for plots seems reasonable in an age of information consumption, and that is a good topic for another article.

TRI Releases And Transfers For 1987

Totals By State and Distribution Class

Grand Total = 7 Billion Pounds



Units: 100 Million Pounds

As for producing similar plots, the Splus functions, script files and data for all the row-labeled plots described in Carr (1994) are available by anonymous ftp. The computer is `galaxy.gmu.edu` and the directory is `/pub/submissions/rowplot`.

Acknowledgments

Research related to this article was supported by EPA under cooperative agreement No. CR8280820-01-0. This article has not been subjected to the review of the EPA and thus does not necessarily reflect the views of the agency and no official endorsement should be inferred.

References

- Carr, D. B. (1994), "Converting Tables to Row-Labeled Plots." Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax, VA.
- Cleveland, W. S. (1993a), "A Model for Studying Display Methods of Statistical Graphics," *Journal of Computational and Graphical Statistics*, 2 (4), 323-343.
- Cleveland, W.S. (1993b), *Visualizing Data*, Summit NJ: Hobart Press.
- Courteau, J. B. (Editor) (1988), *Toxics in the Community, 1988 National and Local Perspectives*, EPA 560/4-90-017, Washington, D.C: U.S. Government Printing Office.
- Grant, J. P. (1993), *The State of the World's Children 1993*, New York, NY: Oxford University Press.
- Kosslyn, S. M. (1994), *Elements of Graphic Design*, New York, NY: W. H. Freeman and Company.
- Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading MA: Addison-Wesley.
- Tufte, E. R. (1983), *The Visual Display of Quantitative Data*, Cheshire, CT: Graphics Press.
- Tufte, E. R. (1990), *Envisioning Information*, Cheshire, CT: Graphics Press.

Daniel B. Carr
George Mason University
`dcarr@galaxy.gmu.edu`



UNIX COMPUTING

Regular Expressions

by Phil Spector

As you spend more time in a UNIX environment, one concept which you will see again and again is regular expressions. A variety of UNIX tools and commands use regular expressions, unfortunately not always consistently. Nevertheless, regular expressions are a very powerful tool for searching and processing text, and even in their simplest form can make many seemingly complex tasks very easy. In this article, I'll define just what a regular expression is and show some simple ways to use them.

Regular expressions are a very powerful tool for searching and processing text.

A regular expression is a way of describing a pattern of characters to a program so that it can either display or modify the occurrences of that pattern. The UNIX utility `grep` (which is an acronym for "global regular expression print") is the most basic UNIX tool which supports regular expressions; given a regular expression and one or more file names as arguments (in that order), `grep` will display to standard output each line of the file which contains the regular expression. (`grep` can also be used as a filter, by using it on the right hand side of a pipe and omitting the file name.) The simplest form of a regular expression is just a literal string; for example, if we wish to display each line of a file which contains the string `dog`, we can use the command "`grep dog filename`". A number of flags increase the utility of `grep`; they are displayed in the table below.

Flag	Function
<code>-i</code>	Ignore case
<code>-n</code>	Include line numbers
<code>-w</code>	Search for words
<code>-l</code>	List filenames only
<code>-v</code>	List non-matching lines

Table 1: Options for `grep`.

Of course, it often becomes necessary to search for something more elusive than a literal string. To describe more general patterns, special characters are used. For example, in a regular expression, the period (`.`) represents exactly one occurrence of any character other than a newline. Thus, the regular expression `d.g` will be matched not only by `dog`, but by `prodigy` and `bandage` since each of them contain a `d` and a `g` separated by exactly one letter. The command "`grep d.g`