Exploratory Data Analysis

Nicola Davide D'Avanzo

04/08/2015

Principles for Analytic Graphics

- Show comparisons
- Show causality, mechanism, explanation
- Show multivariate data
- Integrate multiple modes of evidence
- Describe and document the evidence
- Content is king

Exploratory Graphs

- Exploratory plots are "quick and dirty"
- Let you summarize the data (usually graphically) and highlight any broad features
- Explore basic questions and hypotheses (and perhaps rule them out)
- Suggest modeling strategies for the "next step"

Plotting Systems in R

- Base: "artist's palette" model
- Lattice: Entire plot specified by one function; conditioning
- ggplot2: Mixes elements of Base and Lattice

Base Plotting System

- Plots in the base plotting system are created by calling successive R functions to "build up" a plot
- Plotting occurs in two stages: 1) creation of a plot; 2) annotation of a plot (adding lines, points, text, legends)
- The base plotting system is very flexible and offers a high degree of control over plotting

Graphics Devices in R

- Plots must be created on a graphics device
- The default graphics device is almost always the screen device, which is most useful for exploratory data analysis
- File devices are useful for creating plots that can be included in other documents or sent to other people
- For file devices, there are vector and bitmap formats: 1) Vector formats are good for line drawings and plots with solid colors using a modest number of points 2) Bitmap formats are good for plots with a large number of points, natural scenes or web-based plots

Lattice Plotting System

Lattice plotting system is implemented using the following package:

- lattice: cointains code for producing Trellis graphics

- grid: implements a different graphic system independent of the “base” graphics system

All plotting/annotation is done at once with a single function call.

Lattice functions

- *xyplot*: this is the main function for creating scatterplots
- *bwplot*: boxplots
- *histograms*
- *stripplot*: boxplot with actual points
- *dotplots*: plots dots on “violin strings”
- *splom*: scatterplot matrix
- *levelplot* e *counterplot*: for plotting “image” data

Lattice panel function

- controls what happens inside each panel of the plot
- both default and custom
- receives the x/y coordinates of the data points in their panelù

ggplot2

qplot() function:

- analog to *plot()* but with many built-in features
- syntax somewhere in between base/lattice
- produces very nice graphics, essentially pubblication ready