Deloitte.

Introduction to R

CAS RPM Seminar
March 19, 2012

Steve Berman, FCAS, MAAA
Jim Guszcza, FCAS, MAAA

R – Graphics

Charts
Scatterplots
Histograms
Boxplots
Other visualizations

R Graphics

- R offers a large versatility in graphics. It's easy to create a simple plot, but graphs can be customized in ways that are hard to do in Excel or most applications

2

plot

- `plot` is the simplest graphics command. It's a generic plotting function. Creates a graphics frame and allows drawing of data series. Can create quick plots or customize the output
  - Chart type (`type`)
  - Titles (`main`, `sub`)
  - Axes (`axis`)
  - Legends (`legend`)
  - Colors (`col`)
  - Sizes (`lwd`, `cex`)
  - Data labels (`text`, `mtext`)
- `lines` draws additional lines on an existing graph
- `par` controls graphical parameters

3

Bar charts

- Use `barplot` function, similar to `plot`
- Grouped bars
- Stacked bars
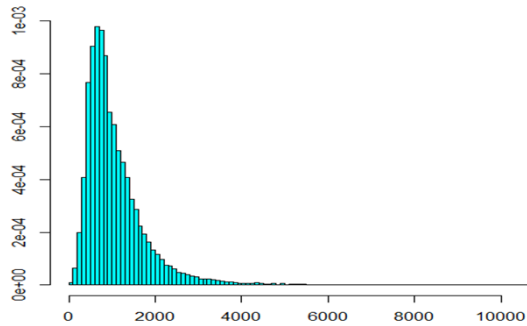- Vertical or horizontal

4

Exercise

- Create a bar chart showing losses by accident month
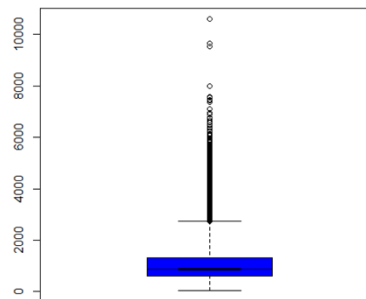- Extra credit: split by coverage, 2007 only

5

Histograms

- There are several functions for creating simple histograms
    - `hist`
    - `truehist`
- Can adjust size of bins, color, a few other options

6

boxplot

- Use to get a graphical representation of distribution
- Show center of distributions, outliers
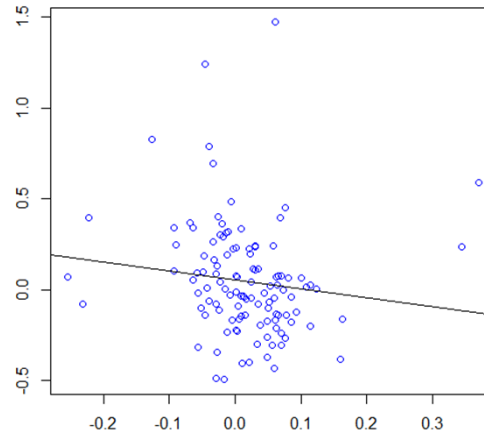- Grouping by subcategories

- Tip: 2-D version of a boxplot exists, called a bagplot (in `aplpack` library)

7

Scatterplots

- Two dimensional plot (where x axis points not ordered)
- Use plot function (type='p') or other functions

8

Exercise

- Modify the price / quantity scatterplot for the following:
  - Pricing changes in January – mark all January and February dots in a different color
  - Change the dots so that the size of the dots is proportional to sales (under $1M, $1M-2M, $2M+) (size of dots using `cex`)
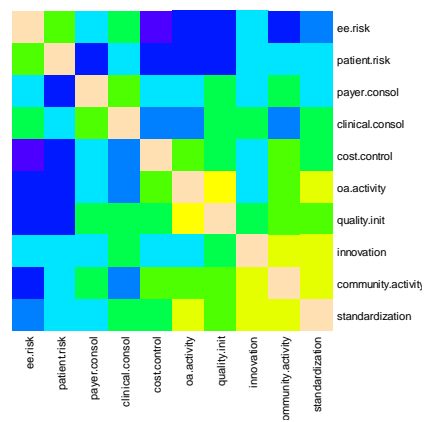
9

Q-Q plots

- To test that data came from specific distribution
  - One axis is empirical distribution
  - One axis is theoretical distribution
  - This example uses `plot` and `abline`

QQ plot
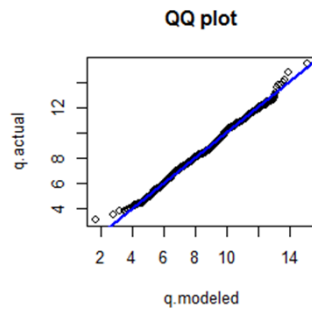
10

Correlations

- Get more information than from `cor()` function
  - `pairs`
  - `heatmap`

11

Other visualizations

- Several packages exist which add to visualizations
  - lattice
  - ggplot
  - maps
  - VMRLGen
  - Many more…

12