BIO5312 Biostatistics
R Session 02: Graph Plots in R

Dr. Junchao Xia
Center of Biophysics and Computational Biology

Fall 2016
Graphic Methods

Graphic methods of displaying data give a quick overall impression of data. The following are some graphic methods.

Bar graphs:
- used to display grouped data;
- difficult to construct;
- identity of the sample points within the respective groups is lost

Stem-and-Leaf plots:
- easy to compute the median and other quantiles
- each data point is converted into stem and leaf, e.g., 438 (stem: 43; leaf: 8)

Box plots:
- uses the relationships among the median, upper quartile, and lower quartile to describe the skewness or symmetry of a distribution

Scatter plots:
- use horizontal and vertical axes to plot data points.
- show how much one variable is affected by another (correlation).
R plotting commands can be used to produce a variety of graphical displays and to create entirely new kinds of display. Plotting commands are divided into three basic groups:

- **High-level plotting functions** create a new plot on the graphics device, possibly with axes, labels, titles and so on.

- **Low-level plotting functions** add more information to an existing plot, such as extra points, lines and labels.

- **Interactive graphics functions** allow you interactively add information to, or extract information from, an existing plot, using a pointing device such as a mouse.

In addition, R maintains a list of graphical parameters which can be manipulated to customize plots.
Bar Plots

We will continue to use the LEAD.DAT as an example.

# set path
>setwd("C:/Users/Junchao/Desktop/Biostatistics_5312_2016/lab_02")
# load the data
>lead=read.table("LEAD.DAT.txt", header=T)
# get the maxfwt column
>fwt=lead$maxfwt
# simple bar plot, showing each data point as one column
>barplot(fwt)
# simple histogram plot
>hist(fwt)
# add x and y labels and title
>hist(fwt, xlab="maximum number of finger-wrist taps", ylab="Frequency", main="Histogram")
Bar Plots: Continued

Final output from the hist() function from the previous slide
Stem-and-Leaf Plots

# use stem simple function to generate a plot
>stem(fwt)

# output

The decimal point is 1 digit(s) to the right of the |
1 | 334
1 |
2 | 3
2 | 6
3 | 4
3 | 5788
4 | 000122244
4 | 55666668888899999
5 | 000000011122222233444
5 | 5566667777778899999
6 | 011222344
6 | 55558
7 | 0224
7 | 69
8 | 34
8 |
9 |
9 | 9999999999999999999999999
A horizontal bar represents the median

The top and bottom edges show the upper and lower quartiles

An outlying value is a value \( x \) such that either
\[ x > \text{upper quartile} + 1.5 \times (\text{upper quartile} - \text{lower quartile}) \]
\[ x < \text{lower quartile} - 1.5 \times (\text{upper quartile} - \text{lower quartile}) \]

An extreme outlying value is a value \( x \) such that either
\[ x > \text{upper quartile} + 3.0 \times (\text{upper quartile} - \text{lower quartile}) \]
\[ x < \text{lower quartile} - 3.0 \times (\text{upper quartile} - \text{lower quartile}) \]

A vertical bar connects the upper quartile to the largest nonoutlying value in the sample
A vertical bar connects the lower quartile to the smallest nonoutlying value in the sample
Box Plots: Continued

We will continue to use the LEAD.DAT as an example.

```r
# simple box plot using default setting
> boxplot(fwt)
# add outline
> boxplot(fwt, outline=T)
# use a different range
> boxplot(fwt, range=1.1, outline=T)
# add y label
> boxplot(fwt, range=1.1, outline=T, ylab="maximum number of finger-wrist taps")
# generate all box plots from the lead data
> boxplot(lead)
```
Box Plots: Continued

Final output from the boxplot() function from the previous slide
# generate a scatter plot between the full scale IQ and max number of FWT

```r
> plot(lead$iqf, lead$maxfwt, xlab="Full Scale IQ", ylab="MAXFWT")
> plot(lead$iqf, lead$iqp, xlab="Full Scale IQ", ylab="Performance IQ")
```
More High-Level Plotting commands

➢ Displaying multivariate data
  o pairs(x) : produces a pairwise scatterplot matrix of the variables defined by the columns of x.
  o coplot(a ~ b | c) : produces a number of scatterplots of a against b for given values of c.

➢ Display graphics
  o qqnorm(x), qqline(x), qqplot(x, y) : distribution-comparison plots.
  o hist(x), hist(x, nclass=n), hist(x, breaks=b, ...) : produces a histogram of the numeric vector x.
  o image(x, y, z, ...), contour(x, y, z, ...), persp(x, y, z, ...): plots of three variables.
Arguments to High-level Plotting functions

- `xlab=`, `ylab=`, `main=`: add x, y labels and title to a plot.
- `add=TRUE`: Forces the function to act as a low-level graphics function, superimposing the plot on the current plot (some functions only).
- `axes=FALSE`: Suppresses generation of axes—useful for adding your own custom axes with the `axis()` function. The default, `axes=TRUE`, means include axes.
- `log="x", log="y", log="xy"`: Causes the x, y or both axes to be logarithmic. This will work for many, but not all, types of plot.
- `type=`: controls the type of plot produced, as follows:
  - `type="p"` Plot individual points (the default)
  - `type="l"` Plot lines
  - `type="b"` Plot points connected by lines (both)
  - `type="....."` Plot many others
- `col=`: controls the color of plot produced as `color="red", "green", "blue", "dark orange"` and many others.
Some Lower-Level Plotting Commands

Add extra information (such as points, lines or text) to the current plot. Some of the more useful low-level plotting functions are:

- **points(x, y), lines(x, y):** Adds points or connected lines to the current plot.
- **text(x, y, labels, ...):** Add text to a plot at points given by x, y.
- **abline(a, b), abline(h=y), abline(v=x), abline(lm.obj):** Adds a line of slope b and intercept a to the current plot.
- **polygon(x, y, ...):** Draws a polygon defined by the ordered vertices in (x, y).
- **legend(x, y, legend, ...):** Adds a legend to the current plot at the specified position.
- **title(main, sub):** Adds a title main to the top of the current plot in a large font and (optionally) a sub-title sub at the bottom in a smaller font.
- **axis(side, ...):** Adds an axis to the current plot on the side given by the first argument (1 to 4, counting clockwise from the bottom.)
The End