1-06 Intro to Data Visualization

CSI 500

Course material derived from:
An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics
Version 3.4.3 (2017-11-30)
https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

R for Visualization

- R has a very rich set of data visualization features
  - multiple plot types built-in
  - add-on packages provide even more features
- Best learned by example
  - we'll do some basic plots and graphs

Stem-and-Leaf

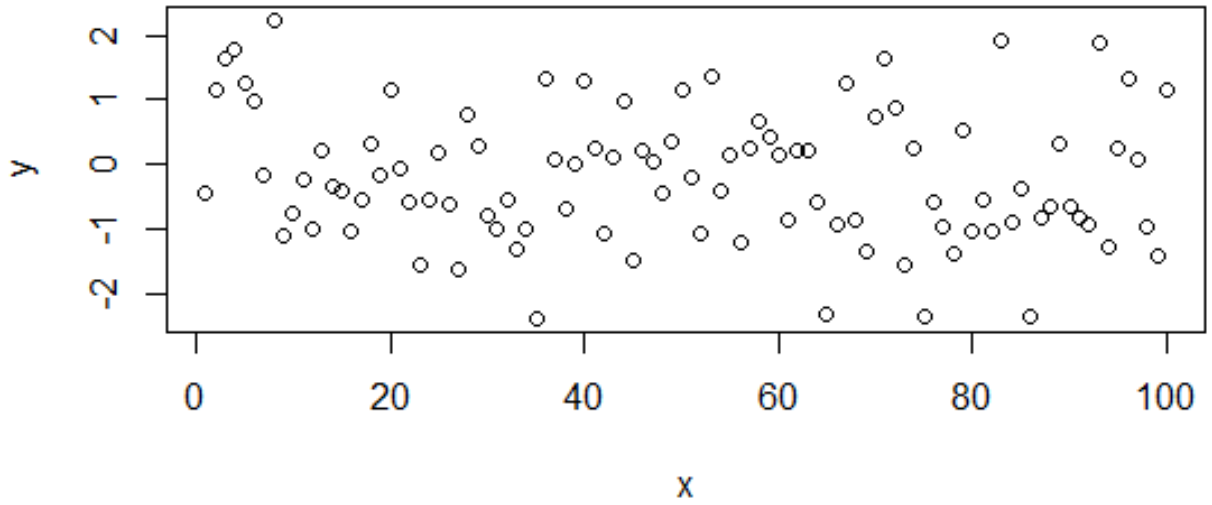
- stem()
- provides simple visualization of a set of numeric data
- Use rnorm(100,0,1) to create 100 random normal draws from a distribution with mean=0, sd=1
- Results show range of data from -2.3 up to 2.8
- Most data clustered around mean, as expected

```
# stem and leaf example
> y = rnorm(100,0,1)
> stem(y)

  The decimal point is at the |

  -2 | 33220
  -1 | 987744332111110000
  -0 | 99998888886666432222111000
   0 | 011122233333444455667777778999
   1 | 002233445677777
   2 | 8
```

Scatterplot

- plot(x, y [,options])
- plots (x,y) paired data on X-Y graph
- here we've specified a set of 100 x values, indexed from 0 to 99
- we've supplied 100 randomly generated y values

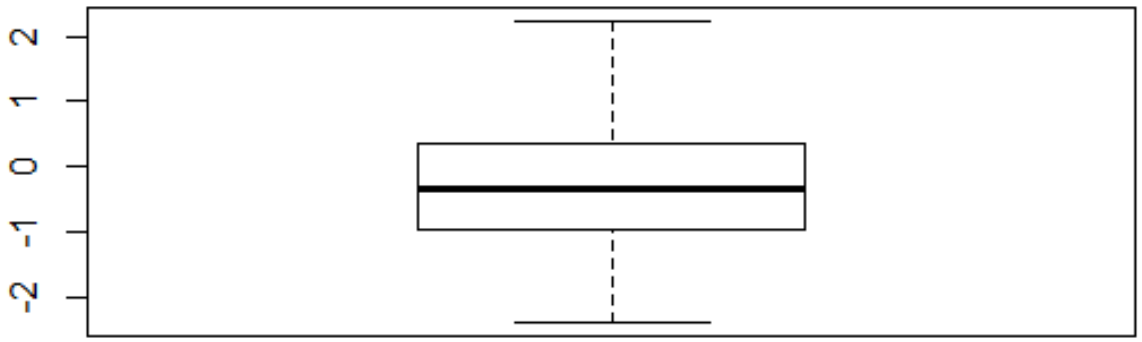
```
# scatterplot example
> x = seq(0,99)
> y = rnorm(100,0,1)
> plot(x,y)
```



Boxplot

- boxplot(y [,y2, y3…])
- plots one or more data sets on a boxplot
- we've supplied 100 randomly generated y values

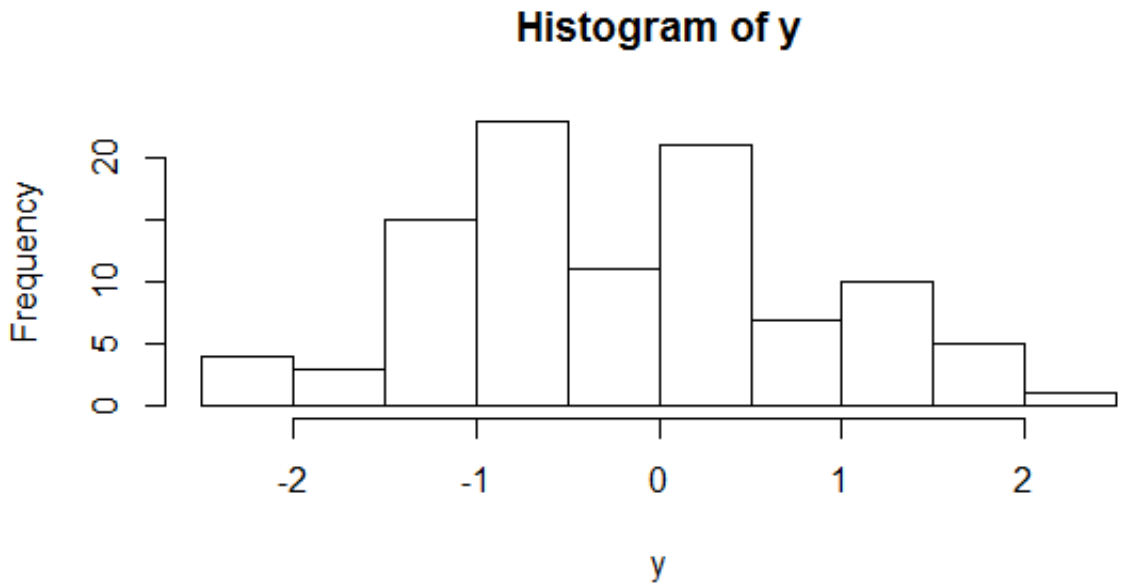
```
# boxplot example
> y = rnorm(100,0,1)
> boxplot(y)
```



Histogram

- hist(y)
- plots data set on a histogram
- we've supplied 100 randomly generated y values

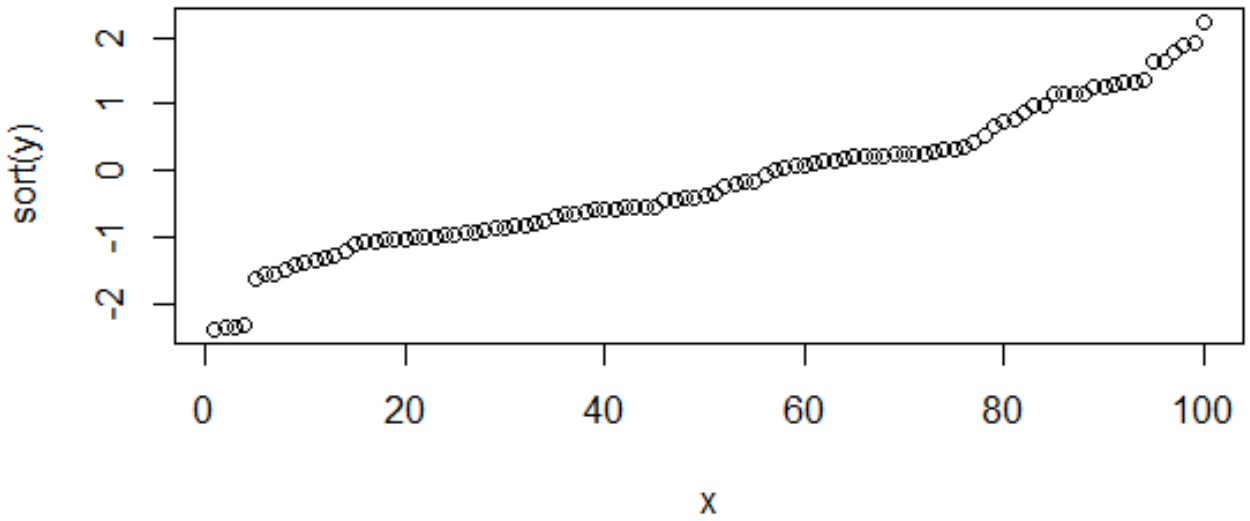
```
# histogram example
> y = rnorm(100,0,1)
> hist(y)
```



Plot

- plot(x, y [,options])
- plots (x,y) pairs on a 2D plot
- we've supplied 100 randomly generated y values
- to make it interesting, we've sorted the y values

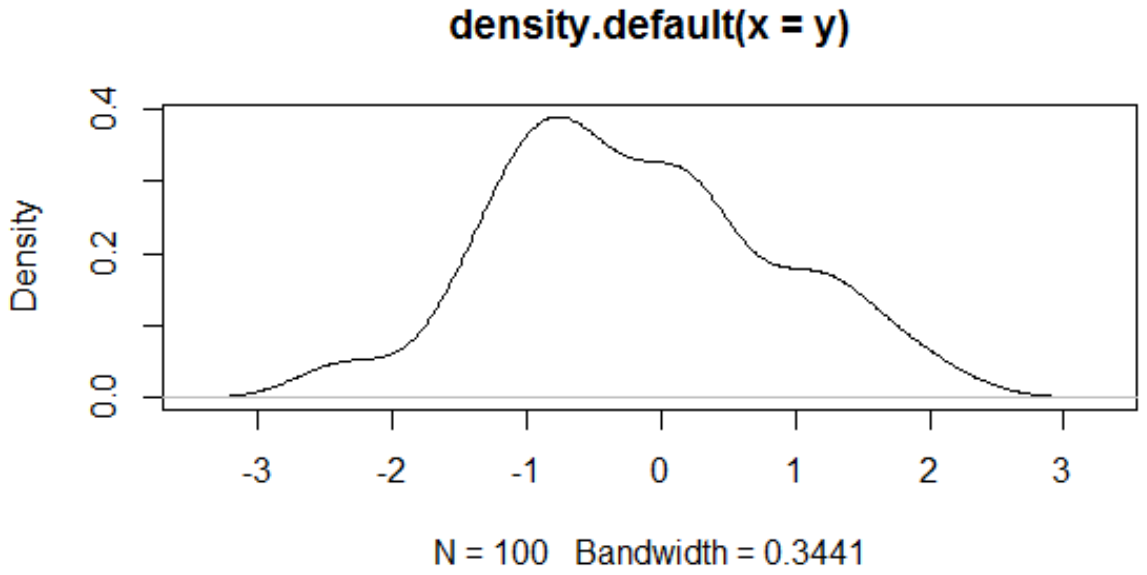
```
# plot example
> x = seq(0,100)
> y = rnorm(100,0,1)
> plot(x, sort(y))
```



Density

```
# density example
> x = seq(0,100)
> y = rnorm(100,0,1)
> plot( density(y))
```

- density(y [,options])
- computes kernel density on a 2D plot
- can be used to show "shape" of distribution
- we've supplied 100 randomly generated y values



Options and Embellishments

- Most R graphics routines use the same set of embellishment options - here are some commonly used features:

  - main="Title of your graph"
  - xlab="title of your X axis"
  - ylab="title of your y axis"
  - xlim=c(low,high) # define X limits if needed
  - ylim=c(low,high) # define Y limits if needed
  - lty=n # define line type, solid, dashed, dash-dot, etc
  - lwd=n # define line width, default is 1.0
  - col="color" # define color
  - pch=n # define point character (dot, triangle, etc.)

Example Embellished Scatter plot

- plot(x, y,

main="example scatterplot",

xlab="X (ID Num)",

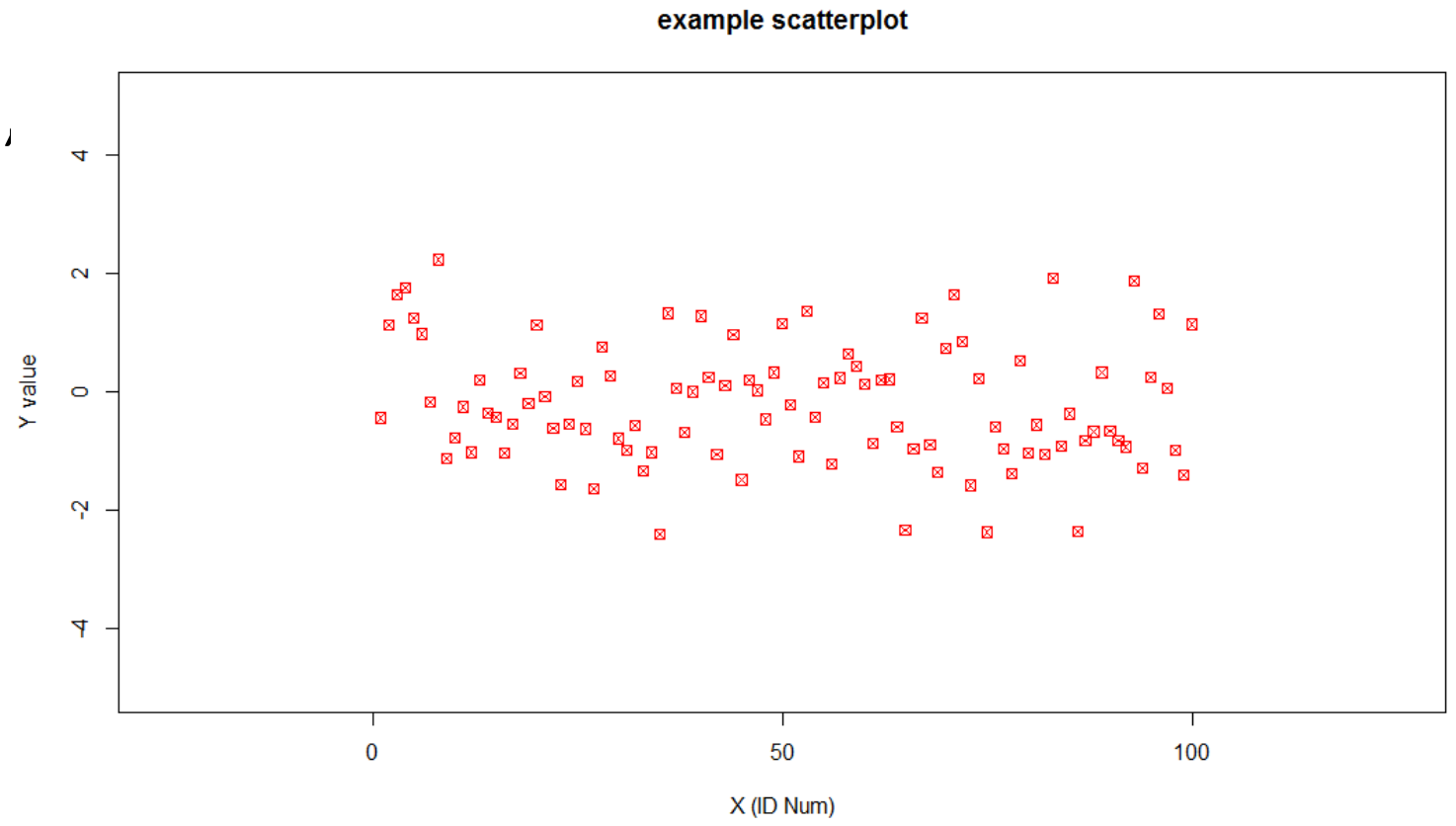
ylab="Y value",

ylim=c(-5,5),

xlim=c(-25,125),

col="red",

pch=7)



Summary

- R has lots of built-in graphical capabilities
  - Stem-Leaf plot
  - Scatterplot
  - Boxplot
  - Histograms
  - X-Y graphs
  - Kernel density
- R has lots of embellishments to enhance your graphs
  - titles, ranges, colors, legends, etc