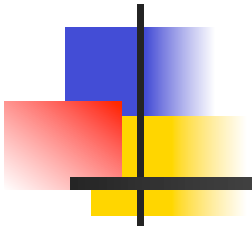
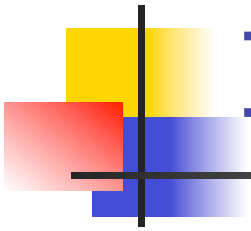


Discussion 01
Introduction to R

STAT 371
Lili Lan, lan@stat.wisc.edu
Office: Medical Science Center B248-H
Office Hours: Friday 11:Am-1:00PM

Installing, Starting & Exiting R



- Download R source files and install
  - http://www.r-project.org -> CRAN -> CRAN Mirrors, select one mirror -> Choose windows -> base -> R-2.7.2-win32.exe (See course website for R installation for Mac)
- R has base packages that can be used after installation of R
- R also has contributed packages which are not available by default. You must install them by yourself after installation of R

- Staring R: Double click the icon of R
- Exiting R:
  - type q()
  - or Click on the X in the top-right corner

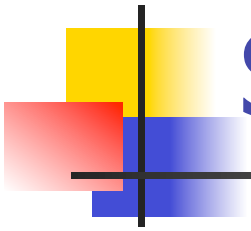
- http://www.biostat.wisc.edu/~kbroman/Rintro/Rwin.html

Working with R

- Use text editor
  - Programming in text editor, copy and paste into R
  - Easy to change and correct programs, easy to save the coding program for future use
- Use R as a calculator: +, -, *, /, ^, log, log10… (See e.g.)
- Most of the time, we work with R objects
  - Need to give an object a value before use it, the value could be numeric or character; character values must be enclosed by quotation marks; numeric values must be numerals
  - object<-value E.g.: name1<-'Jason', name2<-'Kevin'

    height1<-186, height2<-178

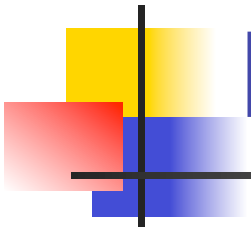
    To use the objects defined, e.g. diff=height1-height2
  - The mode() function helps you identify if an object is numeric or character. E.g. mode(name1), mode(height2)
  - In this class, vector is the mostly used object



Some R functions

- Creating a vector: c()
  - names<-c('Sarah', 'Mark', 'Tom')
  - heights<-c(168, 180, 176)
- Creating numeric sequence: seq()
  - X<-seq(from, to, interval)
  - E.g. x<-seq(1, 10, 0.5)
- A simple but very useful function :
  - It represents consecutive integer numbers
  - E.g. y<-1:20, z<-x[1:5]

Reading Data Files



- Set working directory
  - A working directory is usually a folder where you store your data files and where you want to store your analysis results
  - setwd('path of the folder')
  - setwd('C:\\Documents and Settings\\lily\\Desktop\\engine')
- Read .txt file using read.table
  - enginedata<-read.table('diesel.txt', header=T)
    - Header=T indicates that the file contains the names of the variables as its first line and we want to read the names in.

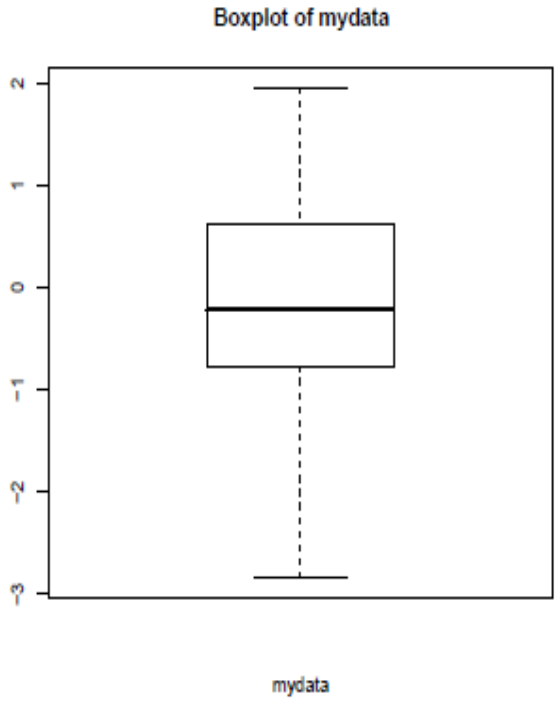
Some Statistical Graphs

- Boxplot
  - Generate the data:
    mydata<-rnorm(100,0,1)
  - Plot the data:
    boxplot(mydata,
    main='Boxplot of mydata',
    xlab='mydata')
  - Mininum, Maximum

    Lower (25%) quantile: Q1

    Median (50%): Q2

    Higher(75%) quantile: Q3



- Histogram
  - Plot the data: hist(mydata, main='Histogram of mydata',xlab='mydata')
- Area of rectangle is proportional to the number of data points in interval