Getting started with R

Goals:

Make yourself familiar with R-**basics**: be able to install the package at your home computer, open the package, find help, load data, download/install additional R-packages, save/load workspace, *etc.*

Learn how to use **simple graphs**, work with graph elements (axis labels; title; grid; scale; plot style, symbols, and colors, *etc.*); export graphs in different formats.

Calculate **simple statistics** for your data (mean, standard deviation, median, quantiles)

Assignments:

1. **Install** R on your personal computer **or** make sure you have access to a campus computer with R.
2. Run the **sample session** (see p. 5). The goal is to type what you see and observe what happens. If you are a first-time R-user, this is the best way to learn. Make sure you understand what the commands do, use help if there are questions.

Reports: No reports are required for this Lab.

Remarks:

- We will need some libraries *NOT* included in a standard R package. This includes library **stats**. Install it on your computer as described below on p. 4.

Installing R on your Windows computer

The R package can be installed on any platform, the instructions are given at the official R web site: http://www.r-project.org/ Here we discuss the Windows installation.

Step 1: Go to http://www.r-project.org/

Step 2: Click on **CRAN** in **Download** section on the left

Step 3: Choose a download mirror (I work with Berkeley)

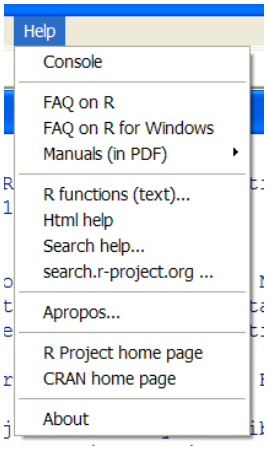
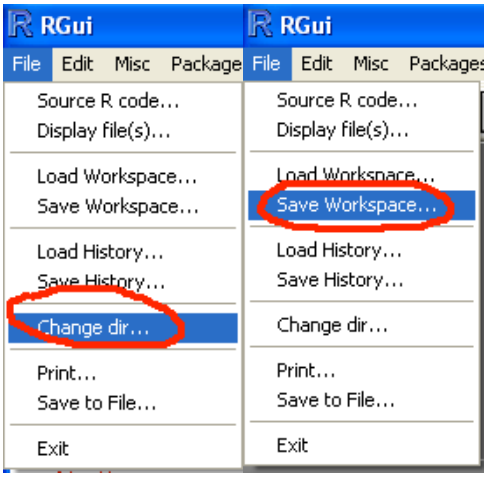
Step 4: Go to Windows section

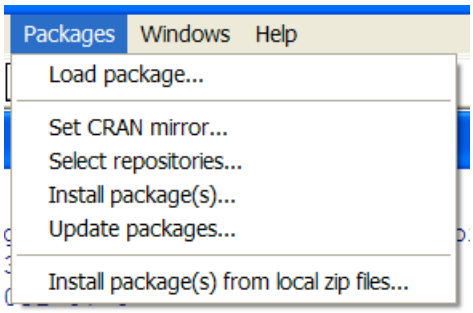
Step 5: Go to base section

Step 6: Download the setup program R-2.13.1-win.exe and save it in your local directory (note that it is about 39Mb)

Step 8: Run the setup program R-2.13.1-win.exe from your local directory

First steps with R

After successful installation, you should see the following (or similar) shortcut on your desktop. Alternatively, you can start the package from START menu (this is the case in Math Center).	
The first place to go is **Help**:	
To setup your working directory, save your workspace, *etc.* you go to **File:**	

R is an environment within which many classical and modern statistical techniques have been implemented. Some of these are built into the base R environment, but many are supplied as ***packages*** *(libraries)*. Sometimes, you will need to download additional packages; this can be done via **Packages → Install Packages**	

How to work in R

The work in R is done either from *command line*, or by running R*-codes*. Always place parentheses ( ) after your commands, even if they have no argument. Say, to ask for help you should run `> help()` not just `> help`. The command without parentheses will give you the source code.

R*-code* is an ASCII file with extension .R that lists R-commands to be executed. It is very advisable to do your work using R-codes.

To create a new R-code, go to **File → New Script**
To run an existing R-code, go to **File → Source R-code…**

Essential R commands:

Session management:

`help()`	`ls()`	`getwd()`
`setwd()`	`library()`	`data()`
`save()`	`load()`	`read.table()`
`class()`	`names()`	`rm()`

Vectors:

`c()`	`seq()`	`rep()`

Data summaries:

`mean()`	`sd()`	`median()`
`quantile()`	`summary()`	

Graphs:

`par()`	`plot()`	`points()`
`lines()`		

Sample session for Lab 1

Part 1: Basics

```
> getwd()
[1] "C:/Program Files/R/R-2.8.1"
```
Shows you the current working directory

```
> setwd("C:/Stats653/Homework")
```
Sets up the working directory, note that you can do it also from **File** menu, which might be much more convenient

```
> data()
```
Opens a window listing all the data sets available with currently installed packages

```
> help("airmiles")
```
Opens a window with help on the data set "airmiles"

```
> data("airmiles")
```
Loads the data set "airmiles" into your current working space

```
> ls()
[1] "airmiles"
```
List variables in the working space

```
> airmiles
Time Series:
Start = 1937
End = 1960
Frequency = 1
 [1]   412   480   683  1052  1385  1418  1634  2178  3362  5948  6109
5981  6753  8003 10566
[16] 12528 14760 16769 19819 22362 25340 25343 29269 30514
```
Shows the content of the data set "airmles"

```
> plot(airmiles)
```
Plots the time series "airmiles" (later we will see how to make this graph pretty)

```
> library()
```
Opens a window with all additional packages currently available

```
> help(package="lattice")
```
Opens a window with help on the package "lattice"

```
> library("lattice")
```
Loads package "lattice" (note that this can be done without commands, just by clicking as shown above)

```
> help(read.table)
```
Opens a window with help on the command "read.table"

```
> C<-read.table("coyote.txt",header=TRUE)
```

Loads table with coyote lengths from the file “coyote.txt” and creates a data frame C from it

```
> class(C)
[1] "data.frame"
```

Class of object C

```
> names(C)
[1] "length" "gender"
```

Names of variables within data frame C

```
> C
   length gender
1    93.0 female
2    97.0 female
3    92.0 female
...
81   90.5   male
82   80.0   male
83   80.0   male
```

Displays the data frame C

```
> ls()
[1] "airmiles" "C"
```

Lists variables in the current workspace

```
> save(file="tmp.tmp")
```

Saves current workspace in the file “tmp.tmp”

Part 2: Simple graphs

```
> par(bg='yellow')
```

Prepare figure background

```
> plot(airmiles,col=4,type='o',pch=19,
xlab='Time, years',ylab='Miles',
main=' Passenger Miles on Commercial US Airlines, 1937-1960')
```

Plot ‘airmiles’ time series (check what each option does to the plot)

```
> grid(col=1,lty=2)
```

Add grid

```
> abline(v=seq(1937,1960),lty=2,col=1)
```

Add additional (yearly) grid lines

Part 3: Simple numerical summaries

```
> a<-c(1,3,5,7,9)
> a<-seq(1,10,by=2)
```

Different ways to create vector a = [1,3,5,7,9]

```
> mean(C$length)
[1] 90.6988

> summary(C$length)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  71.00   86.45   91.00   90.70   95.25  105.00

> quantile(C$length,c(.1,.5,.76))
   10%   50%   76%
   83.56 91.00 95.66
```

Quantiles of input sample