Basics of R

Dennis E. Slice

Abridged History of R

- R is a dialect of the "S" language
- S was initiated in 1976 as an internal statistical analysis
- Environment—originally implemented as Fortran libraries.
- Early versions of the language did not contain functions for statistical modeling
- http://www.biostat.jhsph.edu/~rpeng/biostat776/lecture1.pdf

Alternatives to R used in psychology

- Not free: $$
  - SPSS, SAS, S, S+, Excel, Stata + many more
- Free
  - R (public domain version of R) (*google* R and PSPP)
  - PSPP (a subset of SPSS)
- Luckily, R is extremely extensive with thousands of extension packages contributed by the community
- More information:
  - http://www.coderanch.com/t/535109/ol/alternatives-open-source
  - there is no lack of choice

Why R?

- R excels at
  - data analysis, statistics, model building and predictive analytics.
  - if there is a way to analyze and understand data, someone has created a package for it.
  - broad selection of graphing types to visualize data.
  - R runs on every platform imaginable (including an iPhone)
- But ...
  - higher learning curve, but worth the effort!
  - knowledge of R increases marketability when seeking jobs and flexibility when doing research
  - knowledge of R will simplify learning alternative languages

Using R for Psychological Research

- http://www.personality-project.org/r/
  - Useful links
  - Installation of R
  - Introduction to R
- Based on a more complete introduction to R:
  - http://cran.r-project.org/doc/manuals/R-intro.html
  - use as a reference for class

Installation
(lab: Wed. Jan. 4)

- Access the following web site:
  http://cran.r-project.org/bin/
- Choose the link that corresponds to your computer (macosx, windows, linux)
- Follow the instructions

MacOSX

http://cran.r-project.org/bin/macosx/

R-2.15.2.pkg (latest version)
MD5-
hash: 8935aaa6c90e522e7b1da487c50e0d3c
(ca. 64MB)

R 2.15.2 binary for Mac OS X 10.5 (Leopard) and higher, signed package. Contains R 2.15.2 framework, R.app GUI 1.53 in 32-bit and 64-bit for Intel Macs. The above file is an Installer package which can be installed by double-clicking. Depending on your browser, you may need to press the control key and click on this link to download the file.

From the finder:
Find R-2.15.2.pkg (or appropriate file)
Double click
Follow instructions

Windows

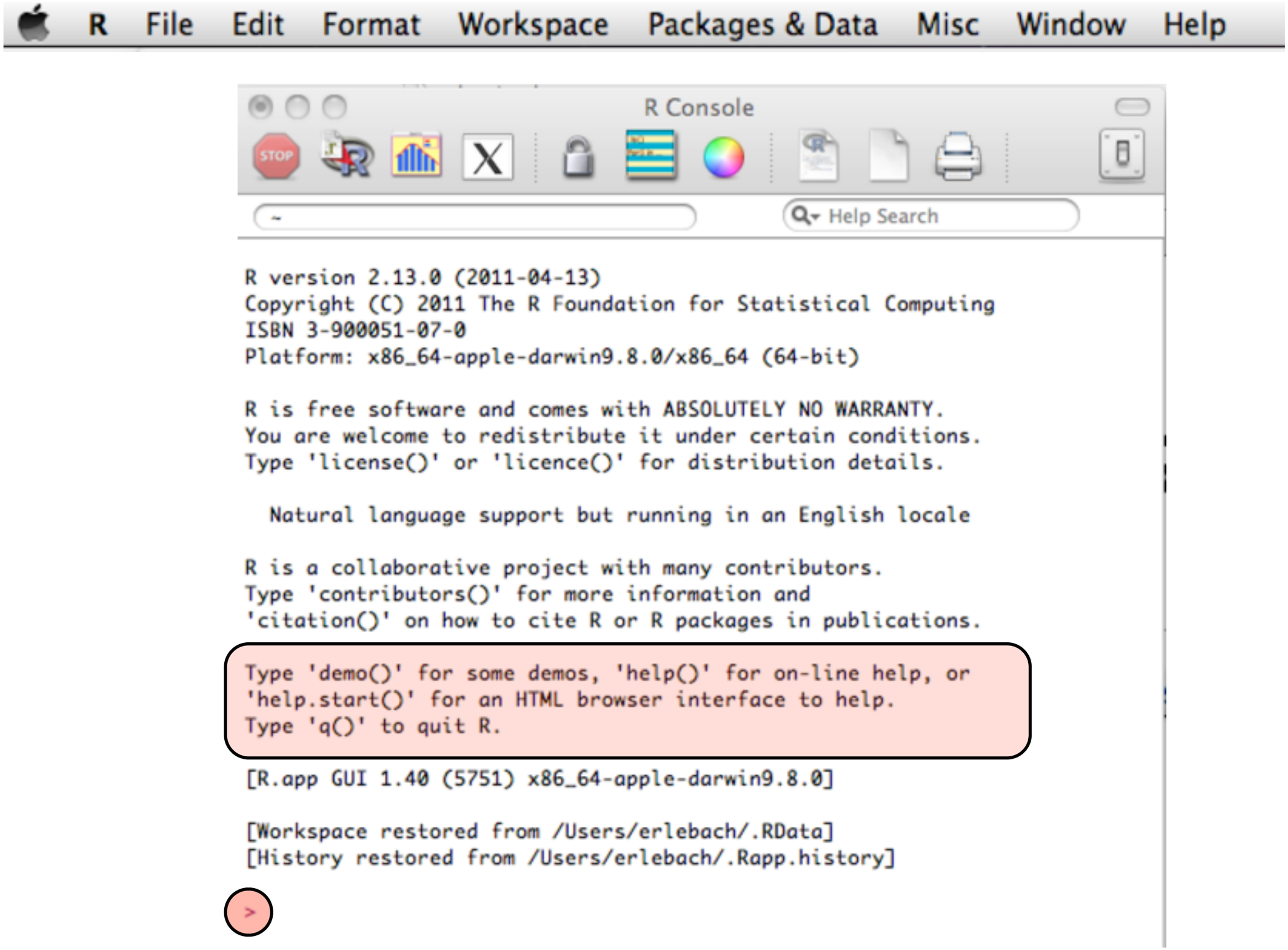
http://cran.r-project.org/bin/windows/base/

Download R 2.15.2 for Windows (47 megabytes, 32/64 bit)
The following file is downloaded:
http://cran.r-project.org/bin/windows/base/R-2.14.1-win.exe

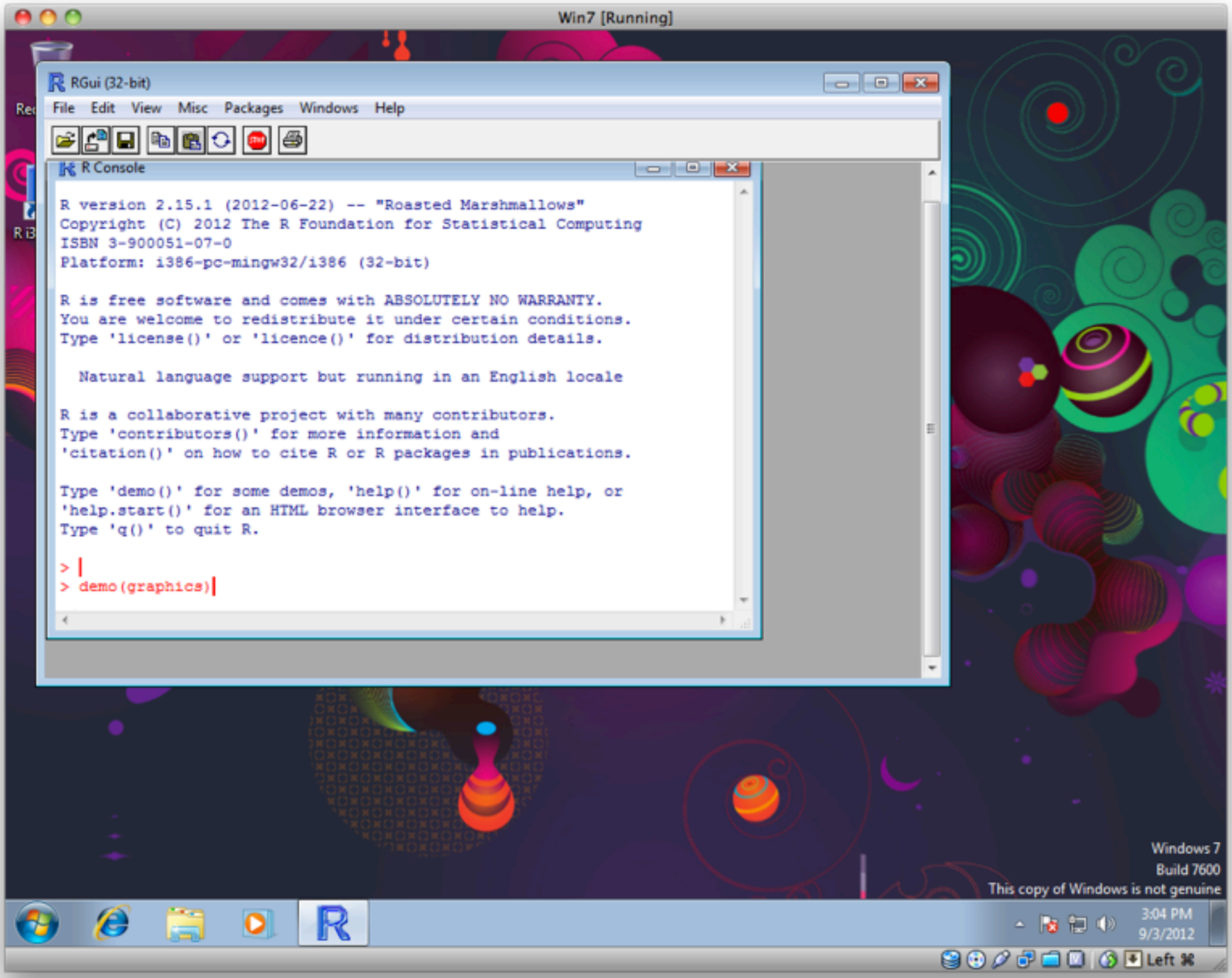
Install like any other program on Windows

Ask TA for help if there are any problems.

R Framework on MacOSX



R Framework on Windows



RStudio

The preferred Software

http://www.rstudio.com/ide/download/

Download RStudio v0.97

If you run R on your desktop:

Download RStudio Desktop

http://www.rstudio.com/ide/download/desktop

RStudio Download

http://www.rstudio.com/ide/download/desktop

RStudio v0.97.248 — Release Notes

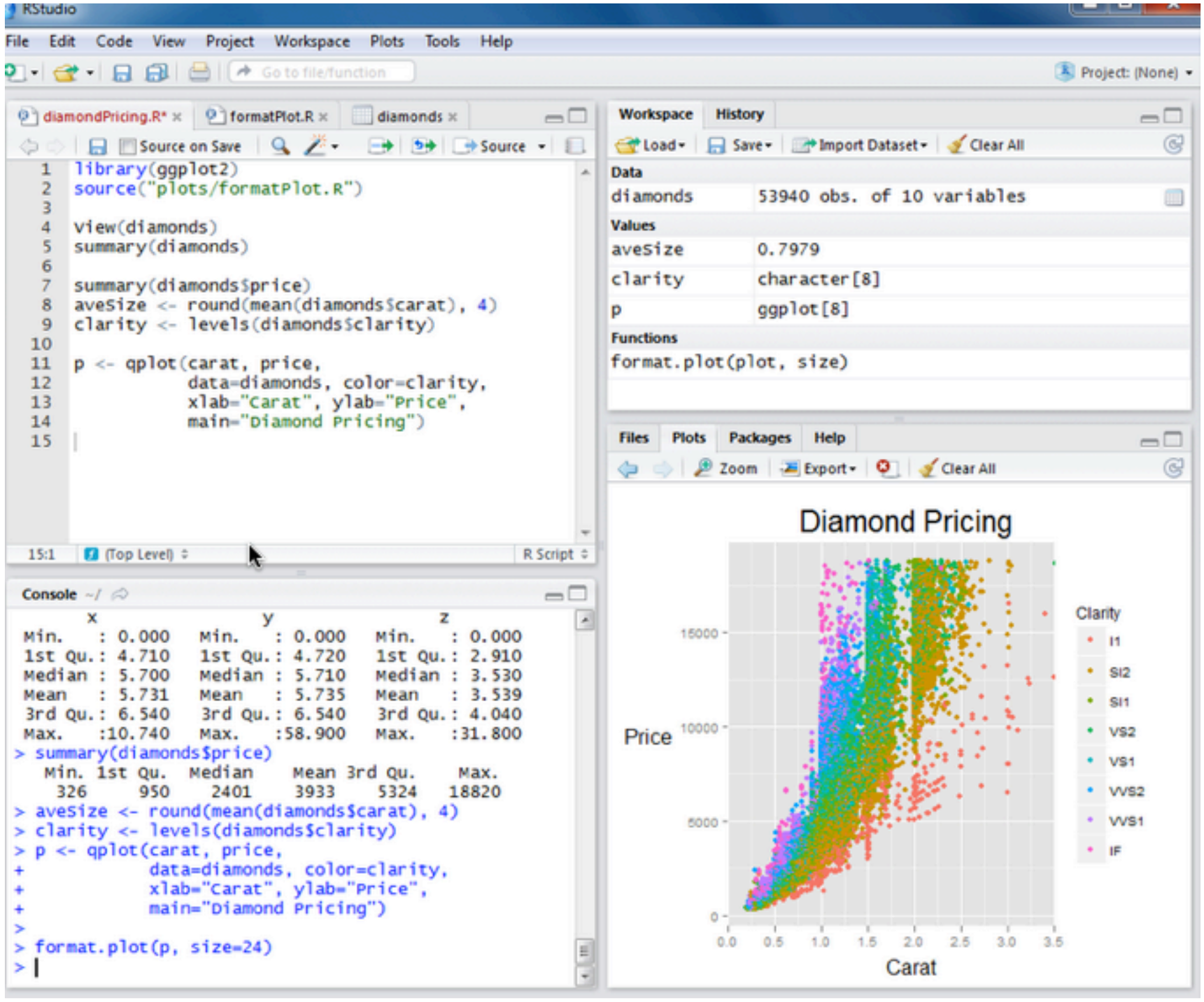
RStudio requires R 2.11.1 (or higher). If you don't already have R, you can download it here.

Recommended For Your System	Size	Date	MD5
RStudio 0.97.248 - Mac OS X 10.6+ (64-bit)	29.9 MB	2012-12-21	ba7519dc

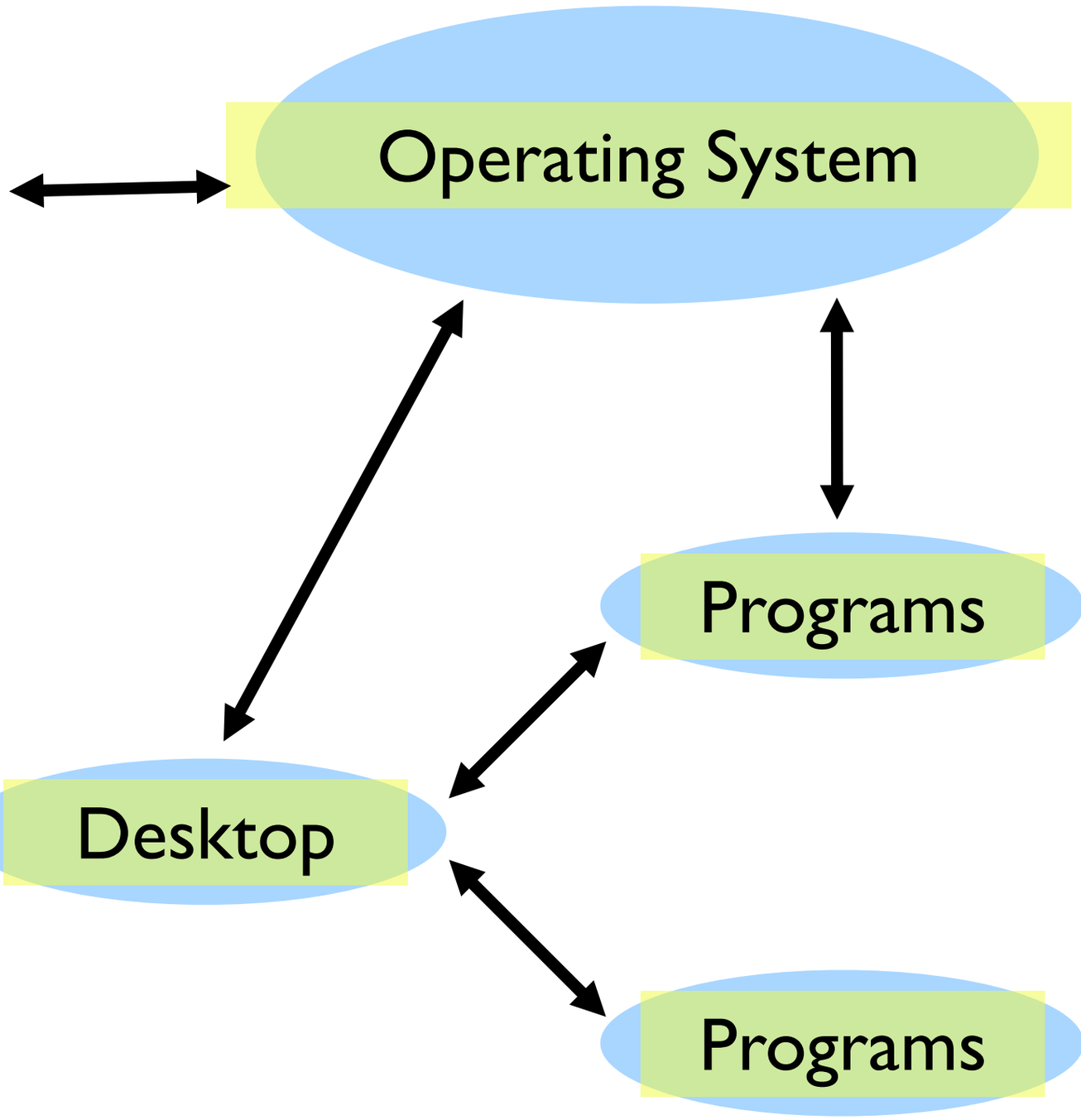
NOTE: RStudio v0.97 requires a 64-bit Mac running Snow Leopard (version 10.6) or higher

All Platforms	Size	Date	MD5
RStudio 0.97.248 - Windows XP/Vista/7	32.5 MB	2012-12-21	701285a0
RStudio 0.97.248 - Mac OS X 10.6+ (64-bit)	29.9 MB	2012-12-21	ba7519dc
RStudio 0.97.248 - Debian 6+/Ubuntu 10.04+ (32-bit)	27.5 MB	2012-12-21	38c1bcd0
RStudio 0.97.248 - Debian 6+/Ubuntu 10.04+ (64-bit)	27.8 MB	2012-12-21	a9bb9db0
RStudio 0.97.248 - Fedora 13+/openSUSE 11.4+ (32-bit)	27.6 MB	2012-12-21	0edb6ef2
RStudio 0.97.248 - Fedora 13+/openSUSE 11.4+ (64-bit)	27.7 MB	2012-12-21	2bd2bcb9

RStudio


RStudio
File Edit Code View Project Workspace Plots Tools Help
Project: (None)
diamondPricing.R* formatPlot.R diamonds
Source on Save
Source
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12 data=diamonds, color=clarity,
13 xlab="Carat", ylab="Price",
14 main="Diamond Pricing")
15
15:1 (Top Level) R Script
Console ~/
x y z
Min. : 0.000 Min. : 0.000 Min. : 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median : 5.700 Median : 5.710 Median : 3.530
Mean : 5.731 Mean : 5.735 Mean : 3.539
3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040
Max. :10.740 Max. :58.900 Max. :31.800
> summary(diamonds$price)
Min. 1st Qu. Median Mean 3rd Qu. Max.
326 950 2401 3933 5324 18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+ data=diamonds, color=clarity,
+ xlab="Carat", ylab="Price",
+ main="Diamond Pricing")
>
> format.plot(p, size=24)
>
Workspace History
Load Save Import Dataset Clear All
Data
diamonds 53940 obs. of 10 variables
Values
aveSize 0.7979
clarity character[8]
p ggplot[8]
Functions
format.plot(plot, size)
Files Plots Packages Help
Zoom Export Clear All
Diamond Pricing
Price
Carat
Clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
0
5000
10000
15000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

How Computers Work



Different Operating Systems

- Windows
- Linux
- Lion (on Apple)
- iOS (on iPad/iPhone)
- Android (created by Google)

Menus, windows, applications, utilities *all create the user experience on a computer*

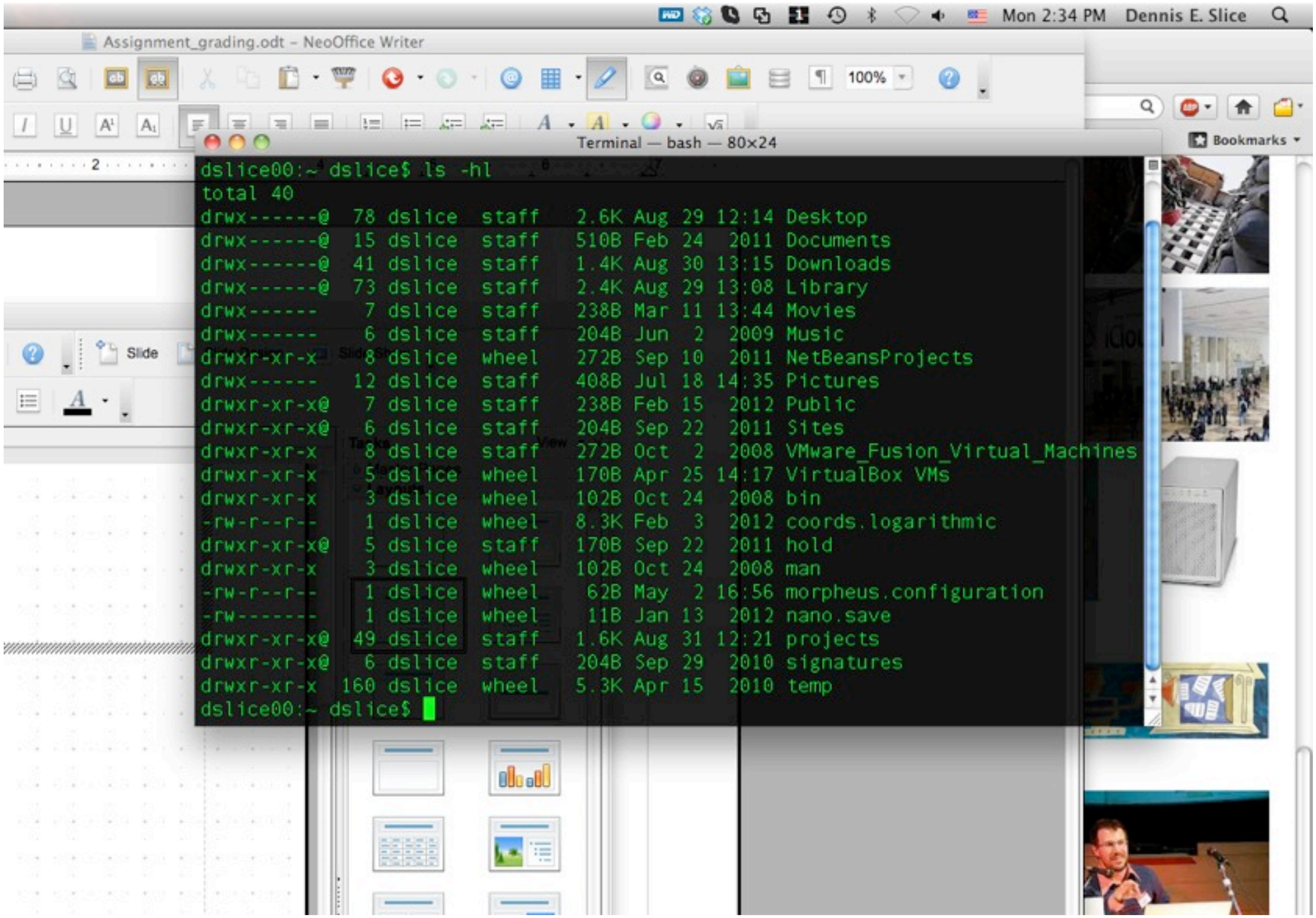
From Wikepedia

An **operating system** (**OS**) is a collection of software that manages computer hardware resources and provides common services for computer programs. The operating system is a vital component of the system software in a computer system. Application programs usually require an operating system to function.

In practice

An operating system is a collection of software, interfaces, commands through which a user interacts with a computer.

OS X Terminal



Terminals, Consoles

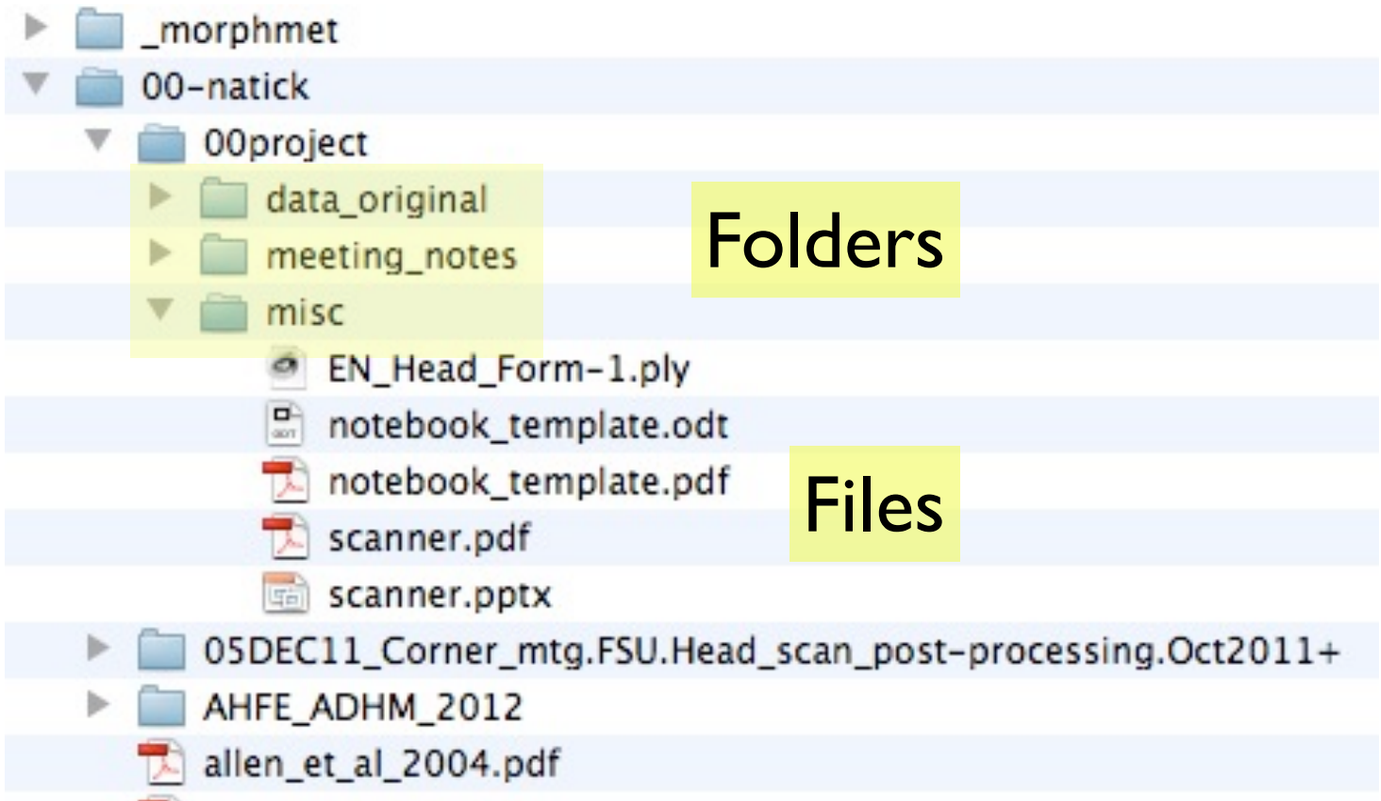
- The user types commands
  - as opposed to pressing buttons
- The computer responds

```
Gordons-MacBook-Pro:top_level erlebach$ who
erlebach console  Jan  3 19:26
erlebach ttys000  Jan  3 19:26
erlebach ttys001  Jan  3 19:28
erlebach ttys002  Jan  4 12:59
Gordons-MacBook-Pro:top_level erlebach$
```

Terminals vs GUI

- In a GUI (Graphical User Interface)
  - the user presses buttons, selects menus
  - fills text boxes
- In a terminal, there are no buttons, menus, etc.
  - the user types in the commands the computer should execute

File Systems



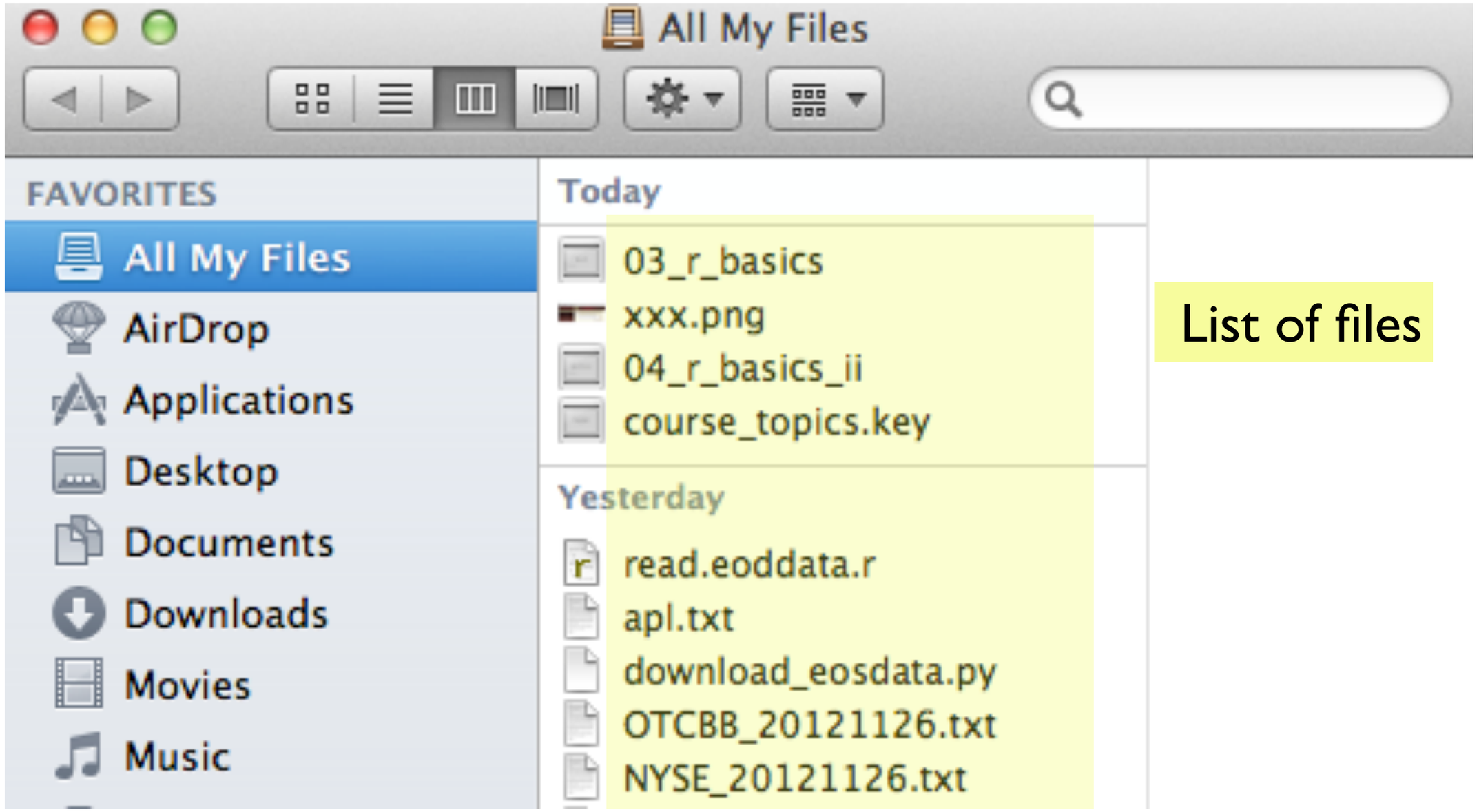
Files

- A file contains data (or information):
  - song.wav (music file)
  - homework.txt (text file)
  - suicide.doc (Microsoft Word file)
- This data is created by various applications (Excel, Word, Photoshop, converters, etc.)

Folder or Directorie

The smallest doll is a “file”.
The other dolls are “folders”

Finder (MacosX)



List of files

Folders/Directories

- In my Movie directory, I have close to 15,000 files!
- Obviously these files should be classified
  - papers in a file cabinet are stored in folders
  - files on a computer are stored in directories (also called folders)
  - A folder contains *files and folders*

blender_python_13
1_0005.MOV

Chicago_Jan_2010

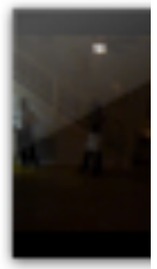
class_sachin_1_100
m152_ja...011.mov

class_sach
m152_ja..

curtis_shoko_demo
_nov_2008.mov

David_Thuy_2-26-
2011.MOV

dubravko_karen_de
mo_nov_2008.mov



dubravko_
emo_20

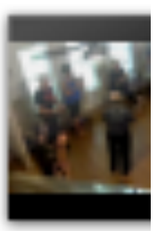
Folder

dwhelper

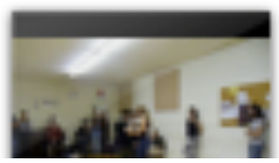
File

El_laguito.mov

erban



erban_t
Gordon...



Contents of *dwhelper* folder

Folder

File

alberto_dassieu

alejandro_hermida

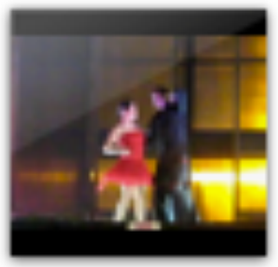
analie_centurion_walk.mov

anamaria_schapira

ANDREA MISSE IN ITHACA,...997.mp4

andrea_misse

andres_laza



Aniversario Evolucio...ngo.mp4

Paths

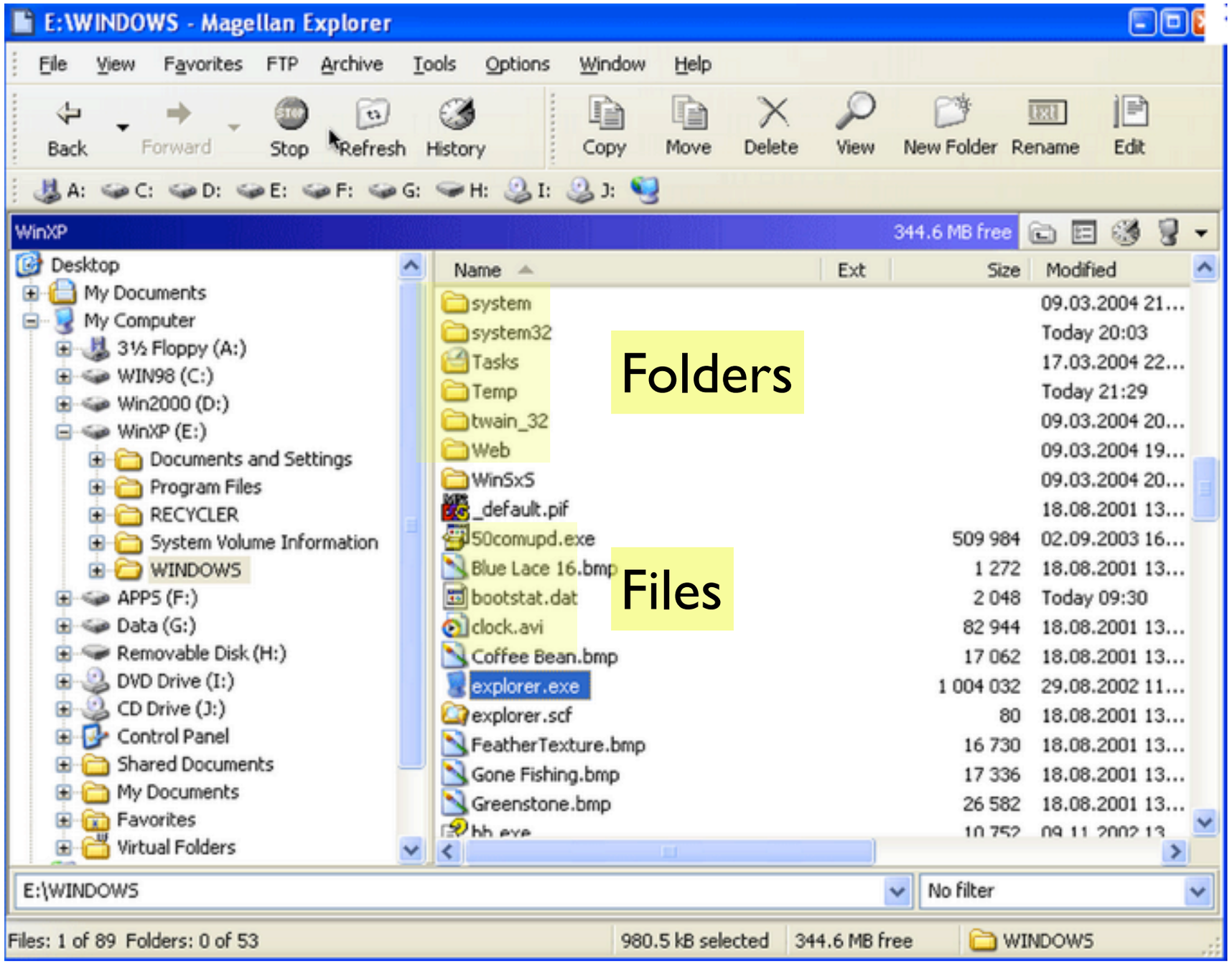
- The computer must have a way to represent the location of a file among the thousands of files and folders on a computer
- Path:

  /Users/erlebach/Movies/dwhelper/val_mimi_pau_2009.mov

  ✓ Users, erlebach, Movies, dwhelper: **folders**

  ✓ val_mimi_pau_2009.mov : **file**

Windows Explorer



Type some commands

```
> dir()
 [1] "3d_id_trim_nomiss.MDT"    "3d_id_trim_nomiss.RData"
 [3] "3d_id_trim_nomiss.RData~" "3d_id_trim_nomiss.mdt"
```

“dir” is a function.
function(arguments)

Try “dir()”,“dir”.

Typed in by user
Returned by the computer

Type some commands

```
> 3+4
[1] 7
> 7 / 3
[1] 2.333333
> 45*3/5 + 20
[1] 47
```

Typed in by user
Returned by the computer

We are not concerned with details

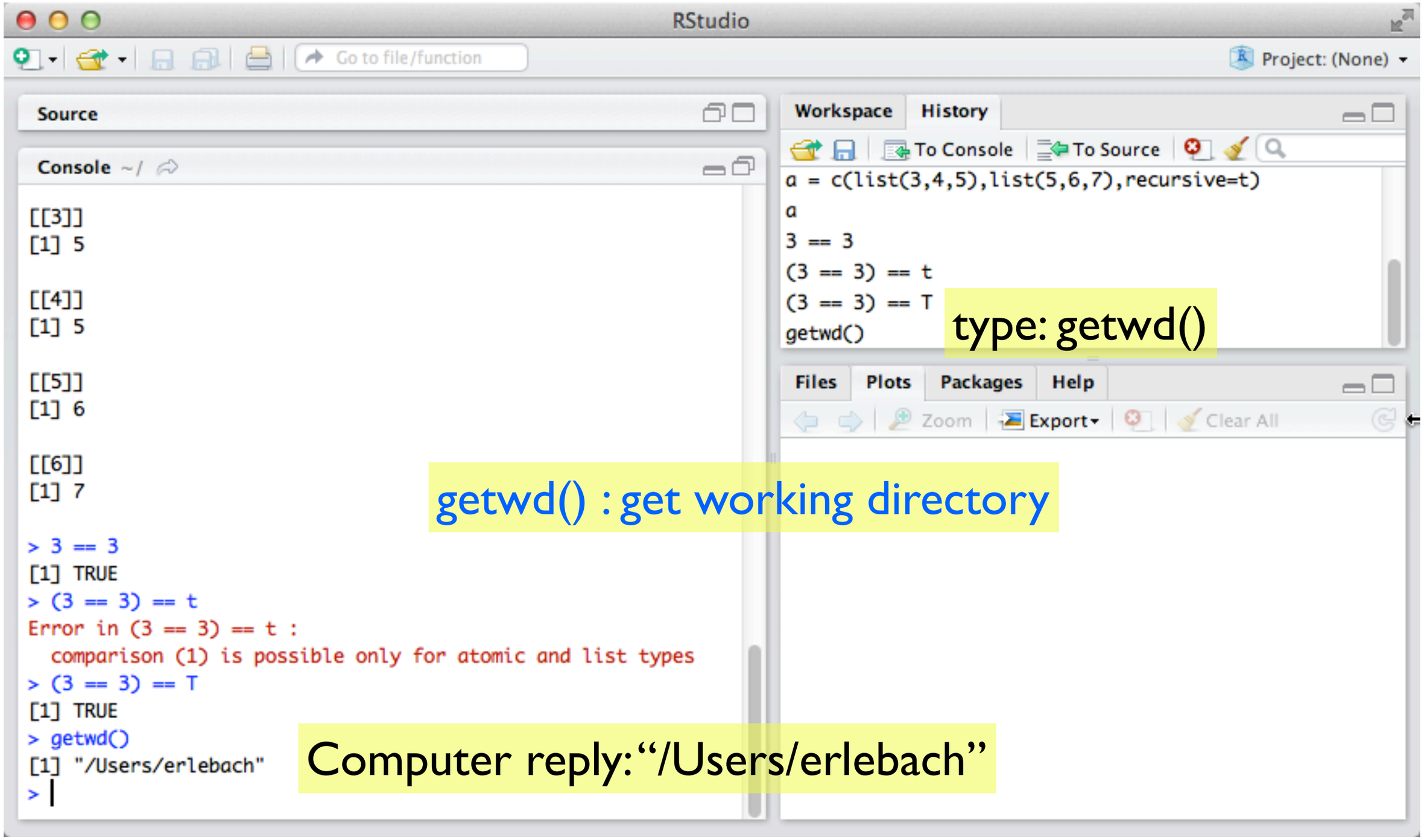
- Rather than try and understand everything, we concentrate on solving problems

First Step

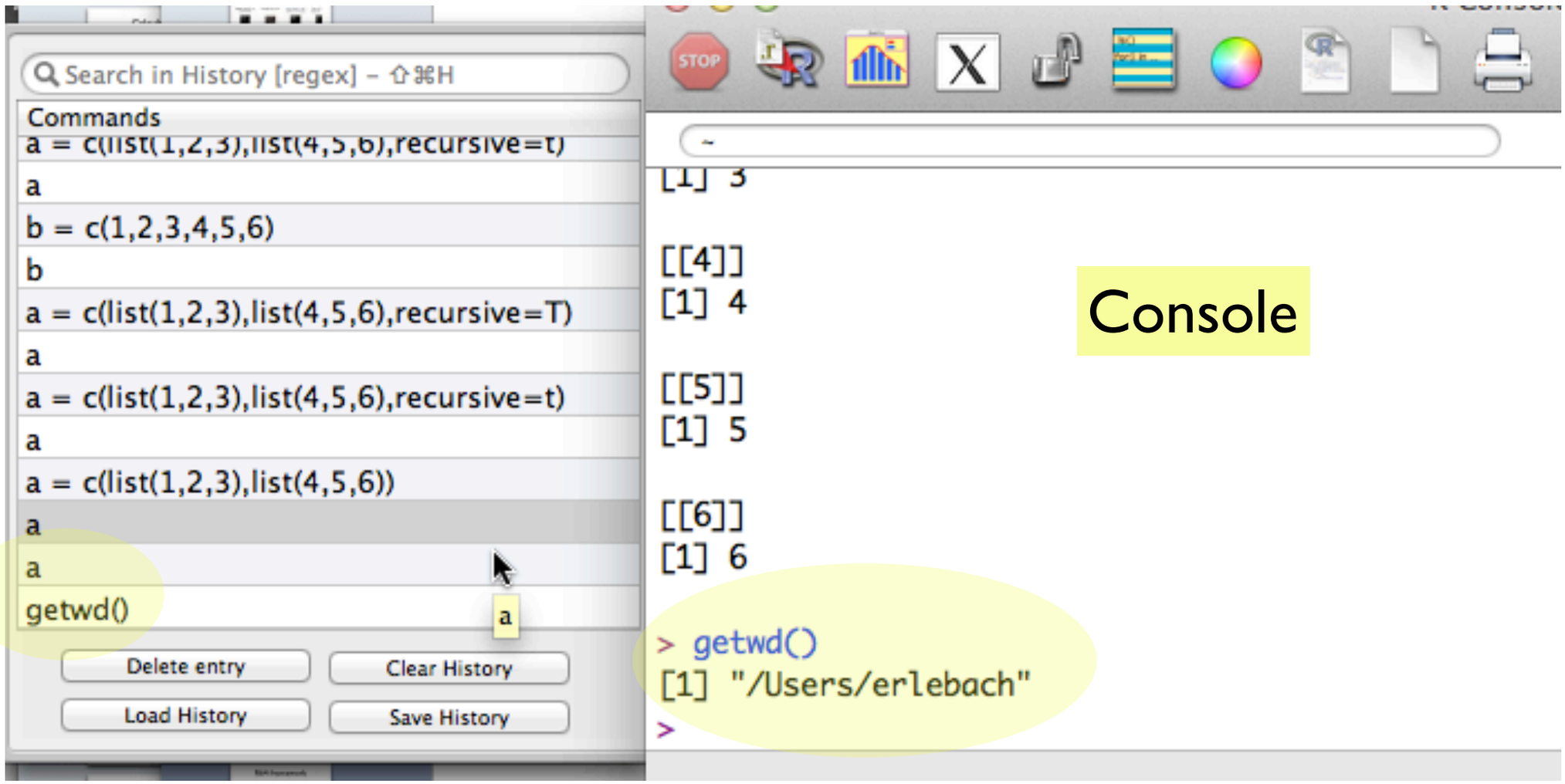
- Exploration of files, folders (or directories) using the R Framework
- Current working directory
  - location within the file system where the R Framework places us by default

```
> getwd()
[1] "/Users/erlebach/Movies"
```

RStudio


RStudio
Go to file/function
Project: (None)
Source
Console ~/
[[3]]
[1] 5
[[4]]
[1] 5
[[5]]
[1] 6
[[6]]
[1] 7
> 3 == 3
[1] TRUE
> (3 == 3) == t
Error in (3 == 3) == t :
comparison (1) is possible only for atomic and list types
> (3 == 3) == T
[1] TRUE
> getwd()
[1] "/Users/erlebach"
>
Workspace
History
To Console
To Source
a = c(list(3,4,5),list(5,6,7),recursive=t)
a
3 == 3
(3 == 3) == t
(3 == 3) == T
getwd()
type: getwd()
Files
Plots
Packages
Help
Zoom
Export
Clear All
getwd() : get working directory
Computer reply: "/Users/erlebach"

R64 framework



getwd(), setwd()

- Use getwd() and setwd() to move around the file system (using the R Framework)

```
> getwd()
[1] "/Users/erlebach"
> setwd("Movies/dwhelper/")
> getwd()
[1] "/Users/erlebach/Movies/
dwhelper"
> setwd("..")
> getwd()
[1] "/Users/erlebach/Movies"
```

Files in Current Working Directory

> **dir()**

[1] "alberto_dassieu"

[2] "alejandro_hermida"

[3] "analie_centurion_walk.mov"

....

[88] "tmp_to_classify"

[89] "val_mimi_pau_2009.mov"

[90] "valdi_guevara"

>

> **getwd()**

[1] "/Users/erlebach/Movies/dwhelper"

There are 90 files in the current directory

dir()

- List the files in the current folder

```
> getwd()
[1] "/Users/erlebach/Movies"
> dir()
 [1] "012-uv-mapping.mov"
 [2] “127_0052.MOV”
 [3] "127_0099_nathan_procrustes_reduced.MOV"
 ...
```

dir() returns a vector of **strings**

Each directory is a **string**

```
> ?dir
starting httpd help server ... done
>
```

R Help

< > Print Help Search

list.files {base} R Documentation

List the Files in a Directory/Folder

Description

These functions produce a character vector of the names of files or directories in the named directory.

Usage

```
list.files(path = ".", pattern = NULL, all.files = FALSE,
           full.names = FALSE, recursive = FALSE,
           ignore.case = FALSE, include.dirs = FALSE)

       dir(path = ".", pattern = NULL, all.files = FALSE,
           full.names = FALSE, recursive = FALSE,
           ignore.case = FALSE, include.dirs = FALSE)

list.dirs(path = ".", full.names = TRUE, recursive = TRUE
```

Arguments

path	a character vector of full path names; the default

Help

RStudio

Source

Console ~/Movies/dwhelper/

```
[65] "orlando_paiva"
[66] "oscar_casas"
[67] "osvaldo_y_coca"
[68] "osvaldo_zotto"
[69] "pablo_garcia"
[70] "pablo_pugliese"
[71] "pablo_rodriguez"
[72] "Peninsula and Jinsuk Muchacha dancing to El olivo by Juan
D'arienzo.mp4"
[73] "Peninsula y Jinsuk Muchacha dancing to Tu Corazón by Donato
Racciatti.mp4"
[74] "pibe_avellaneda"
[75] "puppy_castello"
[76] "raul_bravo"
[77] "ricardo_tango_1.mp4"
[78] "ricardo_tango_2.mp4"
[79] "roberto_herrera"
[80] "roberto_leiva_y_maricel_gomez_sunderland.mp4"
[81] "sebastian_achaval"
[82] "sebastian_arce"
[83] "sebastian_roxana_turning_technique.mp4"
[84] "silvina_valz"
[85] "Sofia et Carlitos dansent sur un tango de D' Arienzo.mp4"
[86] "Tango Argentino Fernando Galera y Vilma Vega en Club Sunderland
Mar 2011.flv"
[87] "tete_rusconi"
[88] "tmp_to_classify"
[89] "val_mimi_pau_2009.mov"
[90] "valdi_guevara"
> str(dir())
 chr [1:90] "alberto_dassieu" "alejandro_hermida" ...
> ?dir
>
```

Use "?" for help

?dir

Workspace | History

To Console | To Source

```
dir()[[76]]
dir()[76]
str(dir()[76])
str(dir()[[76]])
list(1,2,3)
str(list(1,2,3))
str(list(c(1,2,3)))
dir()
str(dir())
?dir
```

Files | Plots | Packages | Help

R: List the Files in a Directory/Folder | Find in Topic

list.files {base} R Documentation

List the Files in a Directory/Folder

Description

These functions produce a character vector of the names of files or directories in the named directory.

Usage

```
list.files(path = ".", pattern = NULL, all.files =
           full.names = FALSE, recursive = FALSE,
           ignore.case = FALSE, include.dirs = FAL

       dir(path = ".", pattern = NULL, all.files =
           full.names = FALSE, recursive = FALSE,
           ignore.case = FALSE, include.dirs = FAL

list.dirs(path = ".", full.names = TRUE, recursive
```

Read about other commands

- ?getwd
- ?setwd

Get or Set Working Directory

Description

`getwd` returns an absolute filepath representing the current working directory of the **R** process; `setwd(dir)` is used to set the working directory to `dir`.

Usage

```
getwd()
setwd(dir)
```

Habits to develop
(extremely important)

- When confronted with a command you are confused about:
  - use "?" to get help
  - use Google to get help, examples, etc.
  - ask the TAs, ask the instructor

Loose Ends

- true mydata.txt
- True MyData.txt
- TRUE MYDATA.TXT

- The same? Depends...

Case Sensitivity

- OS X, Windows Desktops – case insensitive
  ls is equivalent to Ls or LS
- Windows command line – case insensitive
- OS X command line – case sensitive or insensitive depending on installation

Case Sensitivity in R

- R – case sensitive (X≠x; true≠TRUE)
- R – TRUE FALSE (all caps, reserved words)

```
> grade = c(80,70,90)

> grade

[1] 80 70 90

> Grade

Error: object 'Grade' not found
```

Clearly, grade and Grade are *not equivalent!!*

R Basic Types

- Scalars (computing sense) – single entities
  - Integer 3
  - Real 3.14
  - Logical 0 (false) 1 (true)
  - Character/String “male”
  - ...others

Simple Math

```
> 3 + 4
[1] 7
> 7.5/2.0
[1] 3.75
>TRUE & FALSE
[1] FALSE
> TRUE + FALSE
[1] 1
>”garnet” + “gold”
```

R Data Structures

- **Vectors** – simple collection of like entities
- Matrices
- Arrays
- **Data Frames**
- Lists
- **Factors**

Most important structures taught in this course

Next steps

- Become comfortable with **integers, floating point numbers and vectors**