Note 2. Getting Started with R

Introduction to Statistical Programming

Chanmin Kim

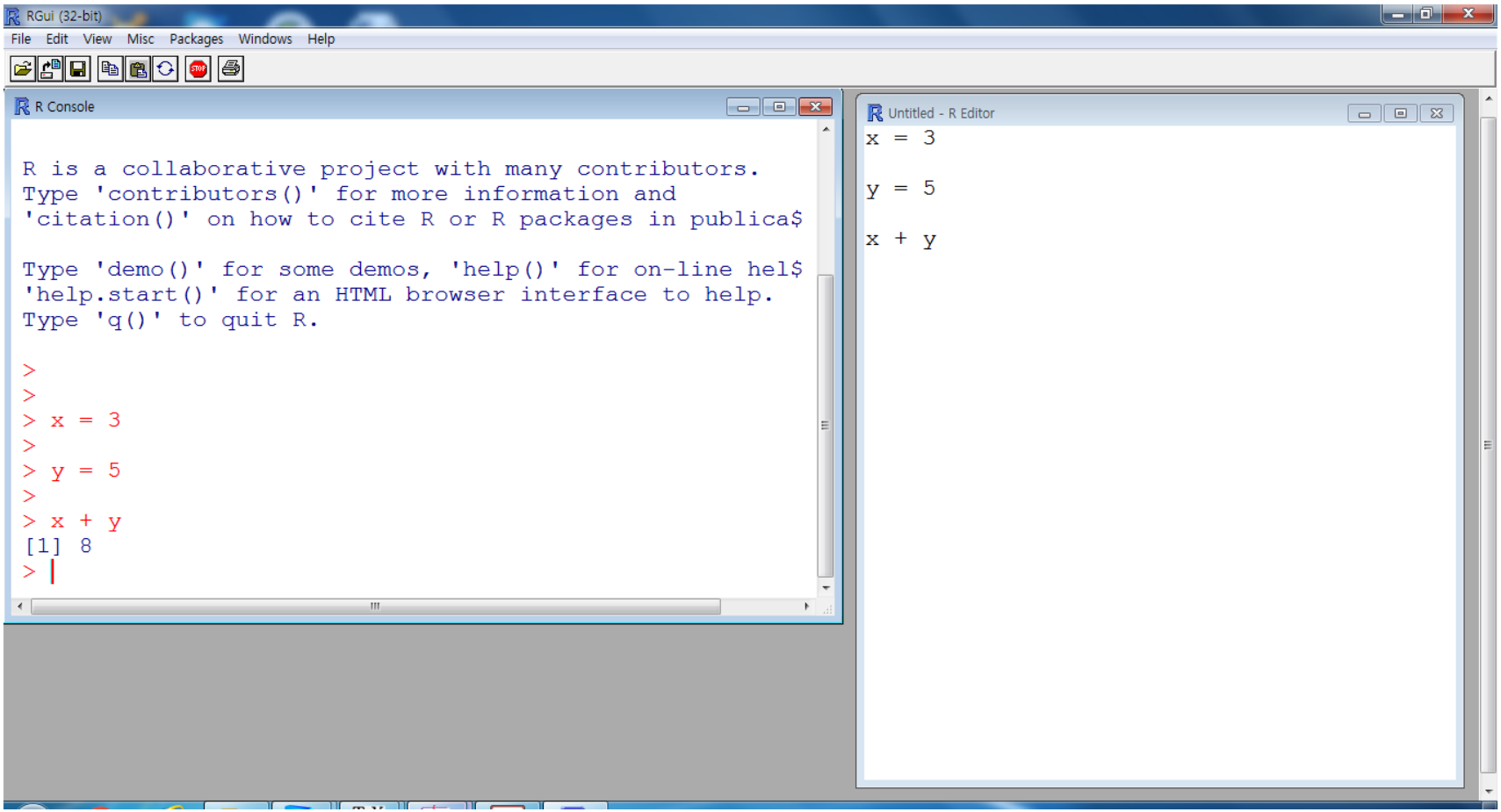
Department of Statistics
Sungkyunkwan University

2022 Spring

Installation

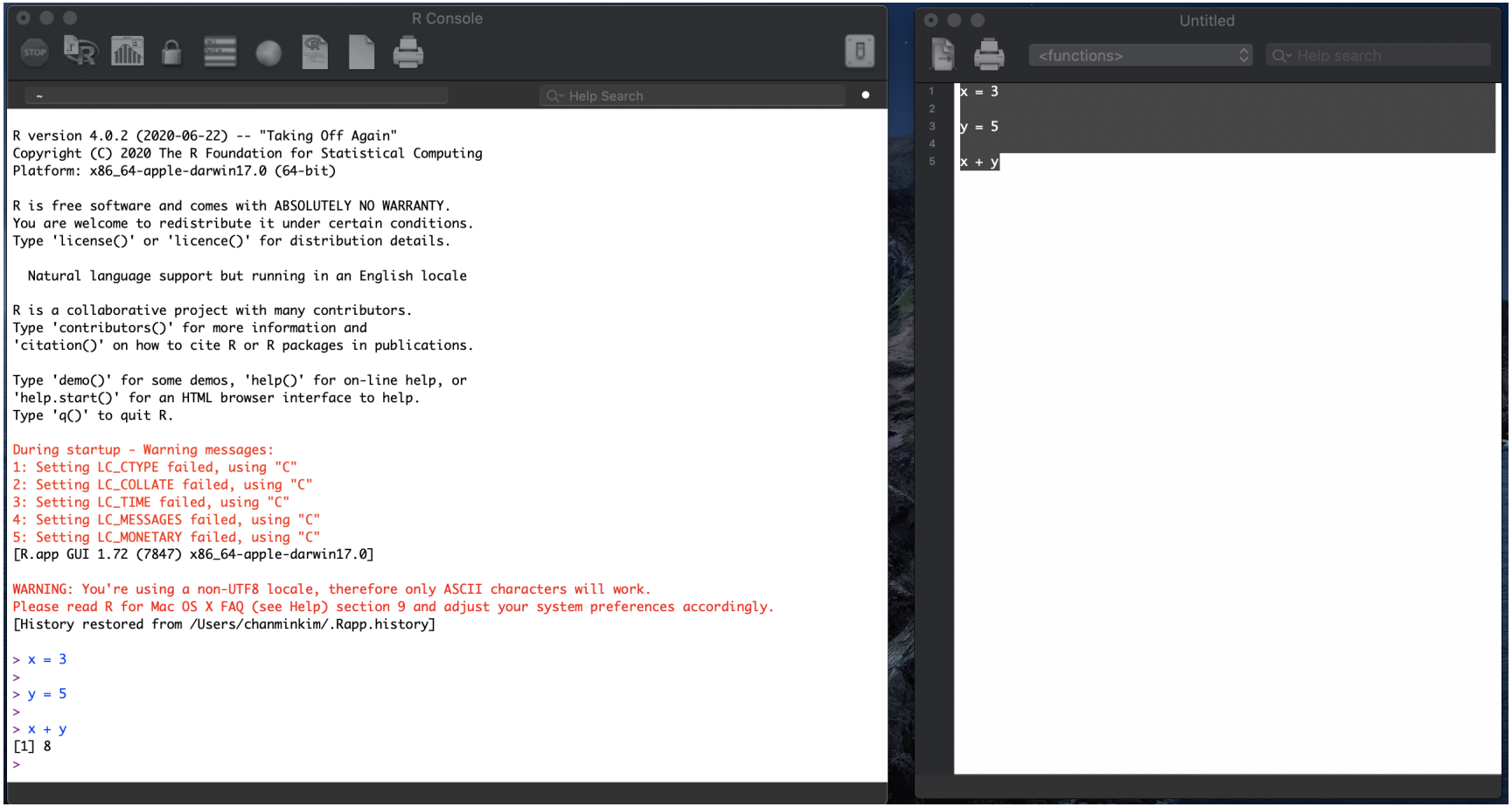
- Download from CRAN website (https://cran.r-project.org/)
- Operating system of your computer (Windows (32bit, 64bit), Mac OS X, Linux)
- Installing R on Windows 10 (YouTube)
- Free GUI (Graphic User Interface) for R
  - R studio (http://www.rstudio.org/)
  - StatET (http:/www.walware.de/goto/statet/)
  - ESS (http://ess.r-project.org/)

R System (Windows)



- R console: Window that R code is running.
- R editor: Window for editing R code (create and save .R files).

R System (Mac)

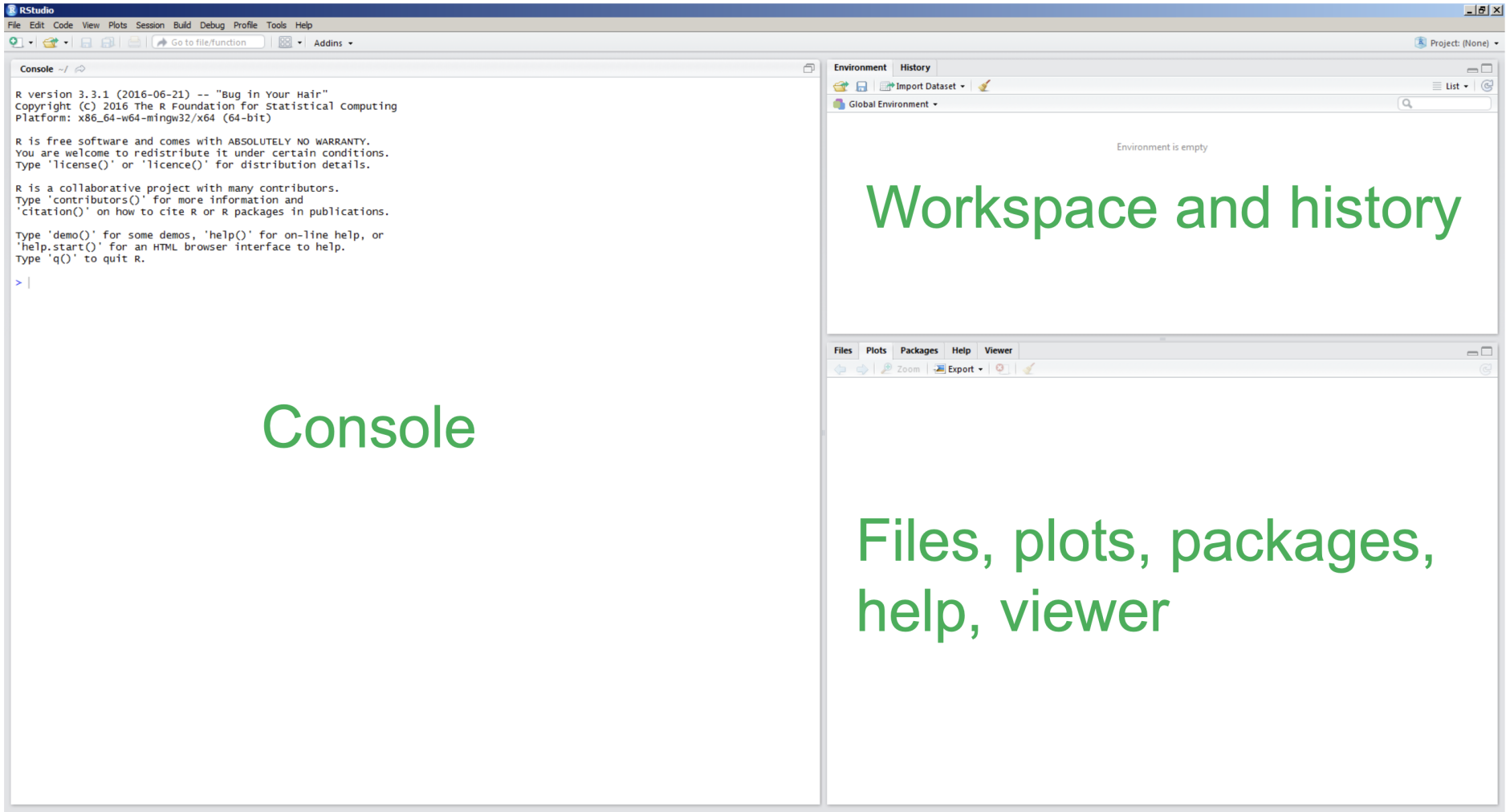


- R console: Window that R code is running.
- R editor: Window for editing R code (create and save .R files).

RStudio

- RStudio is an Integrated Development Environment (IDE) for R
- RStudio integrates the R environment
  - with an advanced text editor
  - R's help system
  - version control tool
  - into a single application
- Before you install RStudio, you need to install R
- It is possible to have multiple versions of R installed on your computer
- RStudio will use the latest version by default, but can be configured to use different version
- The desktop version of RStudio can be downloaded from https://www.rstudio.com/products/RStudio/
- Updates (check for updates by clicking on **Help** | **Check for updates**)

RStudio


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console ~/
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment History
Import Dataset
List
Global Environment
Environment is empty
Workspace and history
Files Plots Packages Help Viewer
Zoom
Export
Console
Files, plots, packages, help, viewer

Working Directory

- Working directory: A directory of a hierarchical file system associated with your work process.
- Once you define the working directory, you can access files in the working directory without specifying its path.
- `getwd()`: It returns current working directory.
- `setwd('dir')`: It sets up your working directory.
- E.g.,

```
> getwd()
[1] "D:/Class"
> setwd('C:/Users/CKIM/')
> getwd()
[1] "C:/Users/CKIM"
```

- Menu: `[File]` ⇒ `[Change dir...]`

R environment

- Interactive mode:
  - Commend $\Rightarrow$ result.
  - E.g.,

```
> x = 1
> y = 2
> z = x + y
> z
[1] 3
```

  - .R file: A file with R commends.
  - `source('a.R')`: Run all commends in `a.R` file.

- Batch mode:
  - To Run .R file without manually launching R.
  - At operating system shell commend (such as Linux system)

```
$ R CMD BATCH a.R
```

Formats of R Objects

- Five basic formats of R objects:
  1. Character
  2. Numeric (real number)
  3. Integer
  4. Complex
  5. Logical (True / False)
- Integer $\subset$ Numeric.
- If you explicitly want an integer, specify the `L` suffix (e.g., `x = 12L`).
- `Inf`: ∞ (numeric).
- `NaN`: Undefined value ('Not a Number'; e.g., `0/0`).

Formats of R Objects

- E.g.,

```
> x = 'Kim'
> mode(x)
[1] "character"
> x = 0.5
> mode(x)
[1] "numeric"
> x = 12L
> mode(x)
[1] "numeric"
> x = 2+3i
> mode(x)
[1] "complex"
> x = F
> mode(x)
[1] "logical"
```

Attributes of R objects

- Attributes: Metadata to describe R objects.
- Types of attributes:
  - Names, dimension names.
  - Dimensions (e.g., matrix, array, etc.).
  - Class (e.g., numeric, character, logical, etc.).
  - Other user-defined attributes/metadata.
- `attributes()`: returns available attributes of R objects.
  No available attributes $\Rightarrow$ `NULL`.

Attributes of R objects

- E.g.,

```
> x = c(1,2,3)
> y = c(4,5,6)
> z = rbind(x,y)
>
> attributes(z)
$dim
[1] 2 3

$dimnames
$dimnames[[1]]
[1] "x" "y"

$dimnames[[2]]
NULL
```

R Data Structure

1. Vector ($\supset$ scalar):
   - All elements should be the same format (e.g., numeric, character, logical, etc.).
2. Matrix: Row & column (All elements = same format).
3. Array: 2 or more dimensional data structure (All elements = same format).
4. List:
   - Different types of elements (e.g., numeric/character, vector/matrix/array).
   - Elements are accessed using two-part names indicated with the dollar sign \$.

R Data Structure

5. Data frame:
   - Two dimensions (row: observation; column: variable)
   - Both numeric & character columns.

6. Class:
   - An instance of R list type.
   - S3 / S4 class.
   - Class: Definition of a R object.
   - Method: A function that performs specific calculations on objects of a specific class.
   - Generic function: A function with a collection of methods.

Example: R Data Structure

```
> # Vector -------------------------------------
> x = c(10,18,12,15,9,7)
> x
[1] 10 18 12 15  9  7
> length(x)
[1] 6
> y = c('kim','92 95','A')
> y
[1] "kim"   "92 95" "A"
> is.vector(y)
[1] TRUE
>
> # Matrix -------------------------------------
> z = matrix(x,nrow=2,ncol=3)
> z
     [,1] [,2] [,3]
[1,]   10   12    9
[2,]   18   15    7
> is.matrix(z)
[1] TRUE
```

Example: R Data Structure

```
> # Array ---------------------------------------
> a = 1:12
> z = array(a,dim=c(2,3,2))
> is.array(z)
[1] TRUE
>
> # List ----------------------------------------
> b = list(x=x,y=y,z=z)
> is.list(b)
[1] TRUE
> b$x
[1] 10 18 12 15  9  7
```

Example: R Data Structure

```
> # Data frame ------------------------------------
> x = 1:3
> y = c('a','b','c')
> z = data.frame(x,y)
> z
  x y
1 1 a
2 2 b
3 3 c
> is.data.frame(z)
[1] TRUE
>
> # Class -----------------------------------------
> x = rnorm(100)
> y = hist(x)
> y
$breaks
[1] -3 -2 -1  0  1  2  3  4
```

Example: R Data Structure

```
$counts
[1]  4 11 36 33 13  2  1

$density
[1] 0.04 0.11 0.36 0.33 0.13 0.02 0.01

$mids
[1] -2.5 -1.5 -0.5  0.5  1.5  2.5  3.5

$xname
[1] "x"

$equidist
[1] TRUE

attr(,"class")
[1] "histogram"
> plot(y)
> plot(x)
```

Workspace

- Turn R off $\Rightarrow$ All R objects are gone.
- Workspace: Save all objects to .RData file.
- Save workspace:
  - `save.imgage('file name')` (e.g., `save.image('xx.RData')`)
  - Menu: [File] $\rightarrow$ [Save workspace...].
- Load workspace:
  - `load('file name')` (e.g., `load('xxx.RData')`)
  - Menu: [File] $\rightarrow$ [Load workspace...].

Help & Example

- R help documents are available within R and on the web.
- To get online help, `help(`*`commend`*`)` or `?`*`commend`*
  (e.g., `help(mean)` or `?mean`; `help('>')`; `?'for'`).
- `example()` function runs examples in the help document.
  (e.g., `example(mean)`).
- Google-style search: `help.search('`*`key word`*`')` or `??'`*`key word`*`'`
  (e.g., `help.search('normal distribution')` or `??'normal distribution'`).