

Lecture Notes - Intro to R

Curry W. Hilton

Installing R and R Studio

- R - <http://cran.r-project.org>
- RStudio - <http://www.rstudio.com/products/rstudio/download/>

RStudio Integrated Development Environment (IDE)

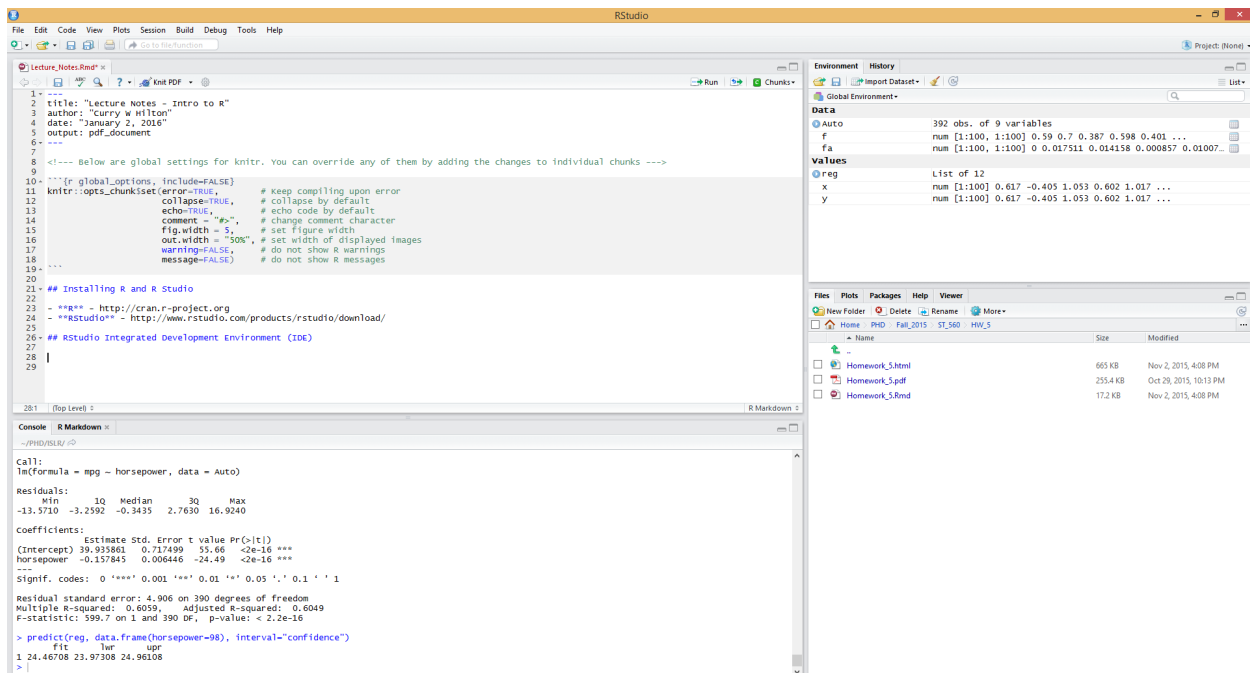


Figure 1:

Four Panes in RStudio

- Console: Where live R Code is "ran"
- Source: Where R scripts or R markdown files are produced (ability to save)
- Environment (Workspace): Where R objects live
- Utility: Where directories, plots, packages, and help live

Customization

- Tools -> Global Options ...
- Description of all Global Options: <http://support.rstudio.com/hc/en-us/articles/200549016-Customizing-RStudio>
- Under General
 - Uncheck "Restore .RData into workspace at startup"

R Project

- File -> New Project

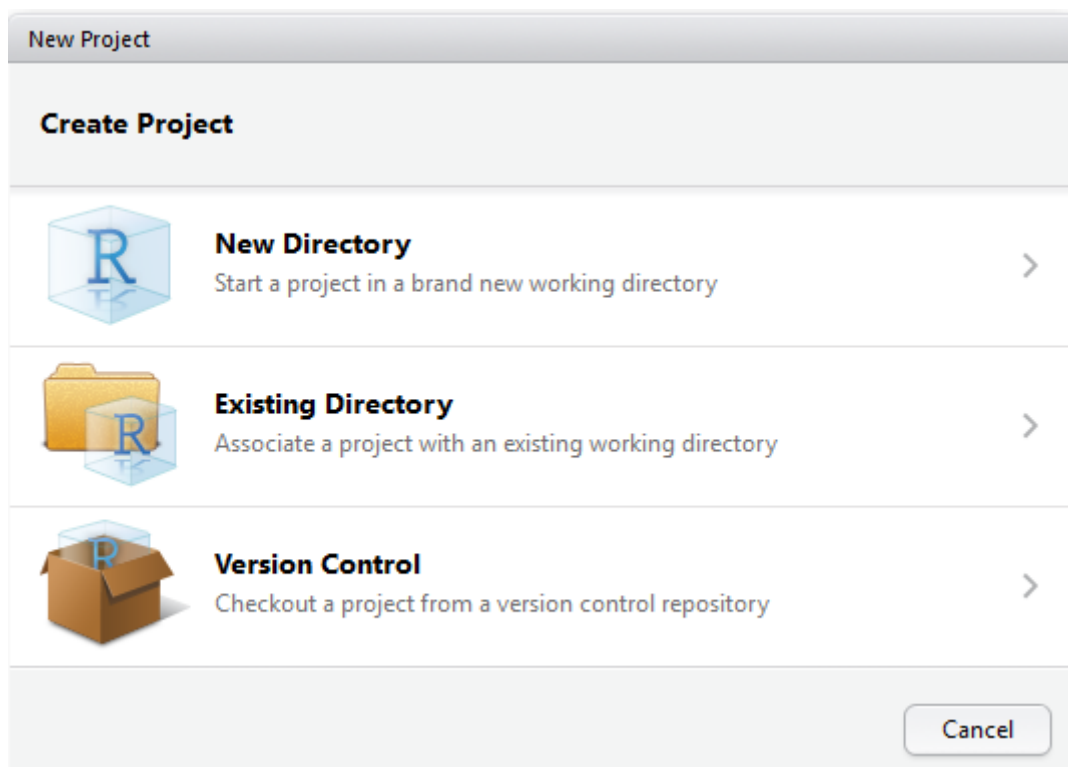


Figure 2:

- Select “Existing Directory” and find the folder system we setup for class!

Let’s Play in the Console!!! Whoop Whoop!!!

Execute the following matemathical operations:

- $5 + 9$
- $9 - 6$
- $5 / 3$
- $4 * 16$
- $16 \% 3$ (Modulo - Remainder)
- $\log(2)$
- $10 ^ 2$

Object Assignments “<-”

- Assign x a value of 5: $x <- 5$
- Print x to the console: x or $\text{print}(x)$
- Assign y an expression of $x + 7$: $y <- x + 7$
- Print y to the console: y or $\text{print}(y)$
- Assign the character string, “I love statistics!” to object z : $z <- \text{“I love statistics!”}$

- Print z to the console: z or `print(z)`
- Assign k to the logical outcome $4 > 3$: `k <- 4 > 3`
- Print k to the console: k or `print(k)`
- Assign d and c simultaneously to the value $\log(8)$: `d <- c <- log(8)`
- Print d and c to the console...are they identical in assignment value?
- Assign the following string of data, beers

```
beers <- c(1, 3, 4, 8, 2, 3)
```

- Print the string, beers to the console
- Type the function “`ls()`” in the console...what do you observe?

Indexing

- In order to index or subset a string/vector of data...
 - `x[i]`
 - `x[i:j]`

```
beers[1]           # bring back first order placement in vector (1)
```

```
## [1] 1
```

```
beers[4]           # bring back fourth order placement in vector (8)
```

```
## [1] 8
```

```
beers[4:6]         # bring back fourth to sixth order placements in vector (8, 2, 3)
```

```
## [1] 8 2 3
```

R Package Installation

- Tools -> Install Packages ...
 - Or ... in the console: `install.packages(pkgnames)`
 - Note: You will only need to install R packages **one time**
- Load Packages to the workspace: `library(pkgname)`
- Update Packages (when necessary): Tools -> Check for Package Updates ...

Required Package Installations

- openintro

Importing Data

Internal R Datasets and OpenIntro Datasets

- datasets

```
library(help = datasets)
```

- openintro

```
library(help = openintro)
```

From the Web

- Visit: <http://www.data.gov/>
 - Find and select the Agriculture link under Browse Topics
 - Choose the “data” tab
 - Choose the “Farmers Market Directory and Geographic Data” dataset
 - Copy the link address in the “Reference” box
- Tools -> Import Dataset -> From Web Url ...
 - Paste link address and press OK
- Examine dataset: Call View() function

From Directory

- Determine working directory
 - Session -> Set Working Directory -> To Project Directory
- From BlackBoard -> Module 2 -> Resources -> “New-Data-1.T15.txt” (save link as and place it in your folder system for class)
- Tools -> Import Dataset -> From Text File -> Browse and find “New-Data-1.T15.txt” in your directory folder and press **open**

Import from CSV

- Follow R script provided

```
library(readr)           # load readr package
getwd()                  # determine working directory
beers <- read_csv("dir/beers.csv") # read csv file and name it beers
View(beers)              # view beers data structure
```

Import from Excel

- Follow R script provided

```
library(readxl)          # load readxl package
getwd()                  # determine working directory
sat <- read_excel("dir/sat.xlsx") # read excel file and name it sat
View(sat)                # view sat data structure
```

Examining Data Structures - Data Frames

- View the dataset: View(x) *Note that the “V” is uppercase
- Names of the variables in the dataset: names(x)
- Structure and variable assignments of a dataset: str(x)
- Dimensions of dataset: dim(x) -> output (row, column)
- View first few observations or last few observations: head(x) or tail(x)
- Column selection: x\$y

```
library(openintro)       # open library for openintro
View(MLB)                 # view the MLB dataset
```

- Determine the names of the variables, structure, and dimensions for the MLB dataset
- Select the ‘player’ column in the MLB dataset and assign it to the name ‘player’

Help!!!!

- `help("x")`
- `?x`

```
help("openintro")
```

R Markdown

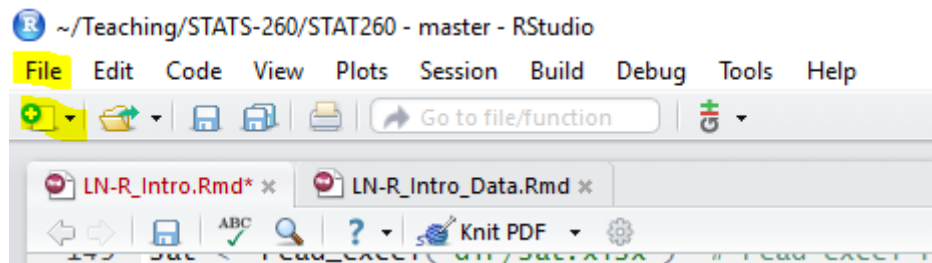


Figure 3:

File -> New File -> R Markdown or Click the “highlighted” plus symbol below the ribbon

Inserting Chunks!

To insert a Chunk go to “Chunks” and “Insert Chunk”

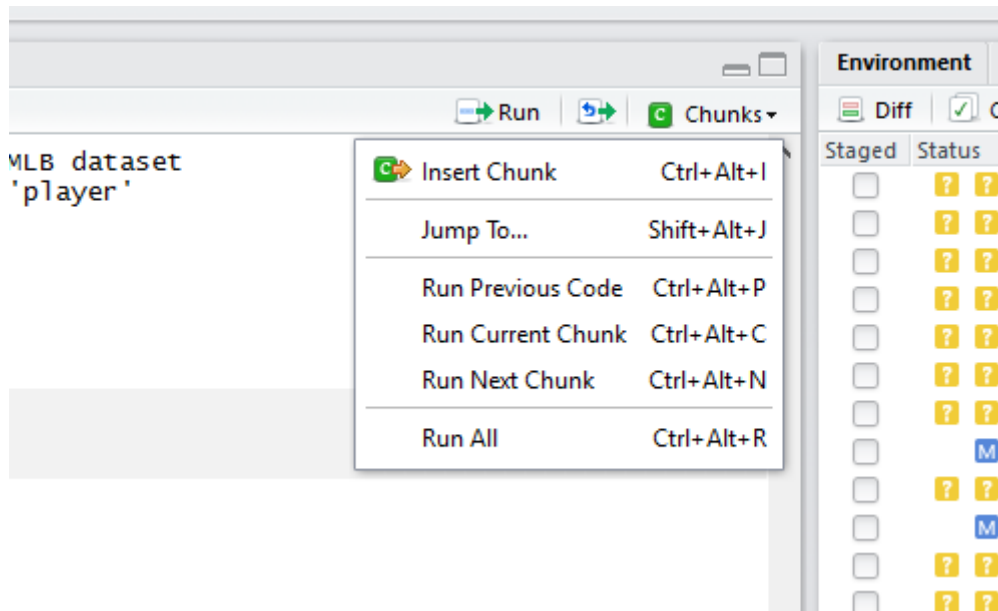


Figure 4:

Within an R Markdown file you can add “Chunks” that contain code that are executed when “knitted”
For example:

```
17*7*49      # Will execute the product of such numbers when knitted
```

```
## [1] 5831
```

We will use R Markdown for submitting all assignments. This ensures aesthetically pleasing, well-organized, and most importantly reproducible output. We will do multiple examples in class before turning you loose on your own! Also, please refer to the R Markdown reference sheet in the “Course Supplements” folder on the course homepage.