R Tutorials

Akihito Kamata, Ph.D.
Southern Methodist University
akamata@smu.edu

March 7, 2020
University of Southern Mississippi

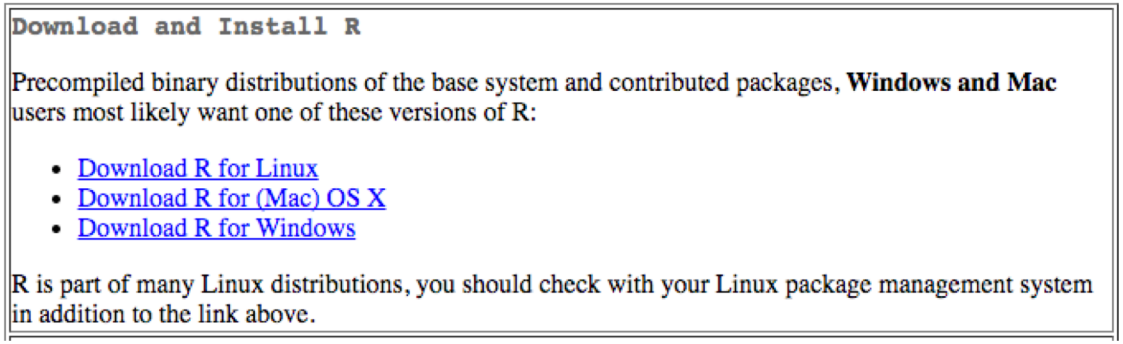
Materials

- All materials including this slide set are available at:
  - https://github.com/kamataak/WorkshopUSM2020
- Download all materials.
  - Click on “Clone or download”.
    - If you know how to use Git, clone it!
  - Otherwise, click on “Download ZIP”.
    - The entire folder will be downloaded to your download folder as a compressed ZIP file.
    - Move the ZIP file where you desire.
    - Unzip the folder.

Part I
Workflow on RStudio

Installing R

- Download R from R-Project website.
  - www.r-project.org
  - Follow links to go to the download page.
    - Go to “CRAN”
    - Then, choose one of the links, such as one under “0-Cloud”
  - R is available for Windows, Mac OS, and Linux platforms.



Download and Install R

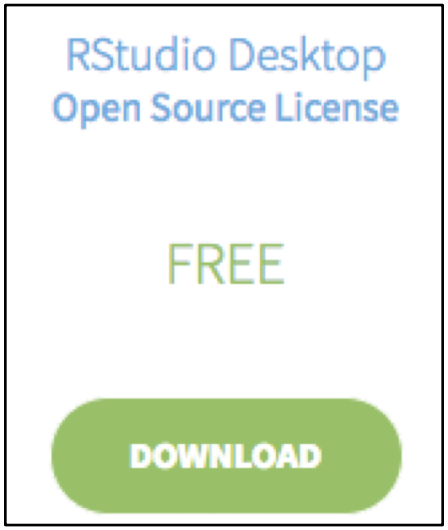
Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

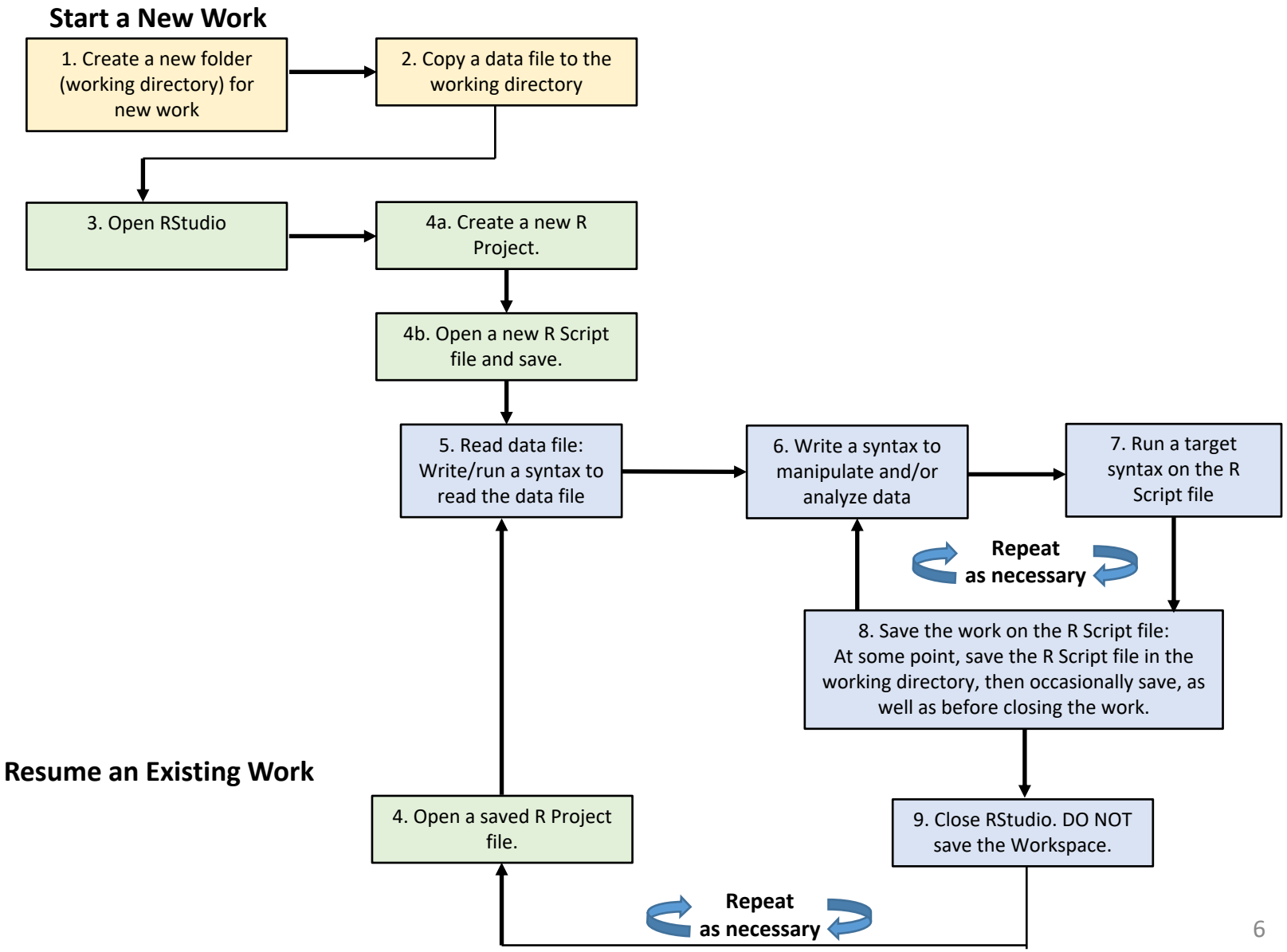
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Installing R Studio

- This is will make R much more user friendly.
- Download from R Studio website.
  - www.rstudio.com
  - Follow links to go to the download page.
  - Download the free version.



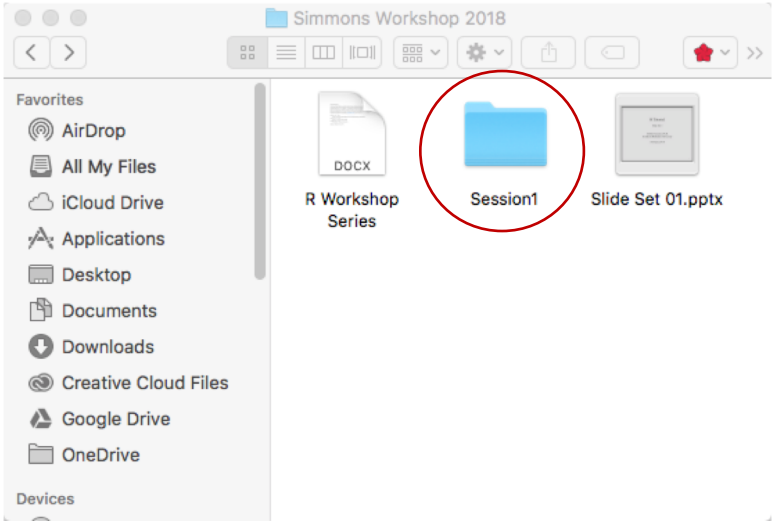
Recommended Work Management on RStudio



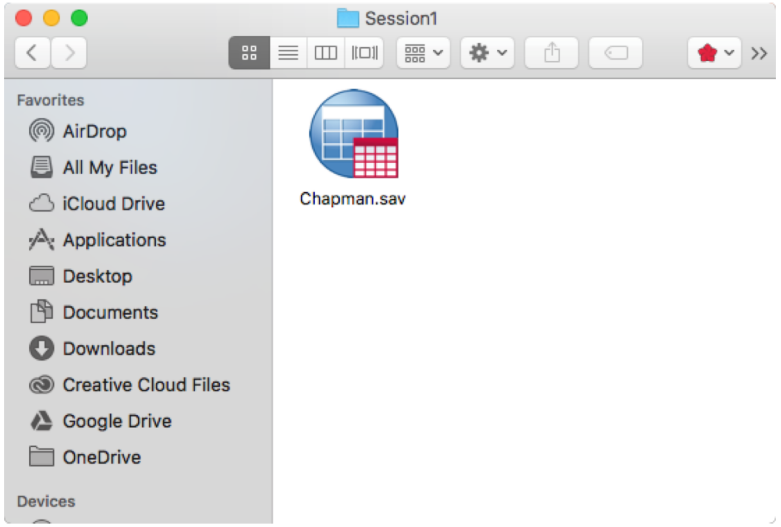
Start a New Work:
Steps 1 & 2: Preparation of Working Directory

- Create a new folder. This is a working directory.
- Copy the data file to analyze to the working directory.
  - For this demonstration, download an SPSS data from: http://akikamata.s3.amazonaws.com/Web_Files/Chapman.sav

Create a new folder (working directory).

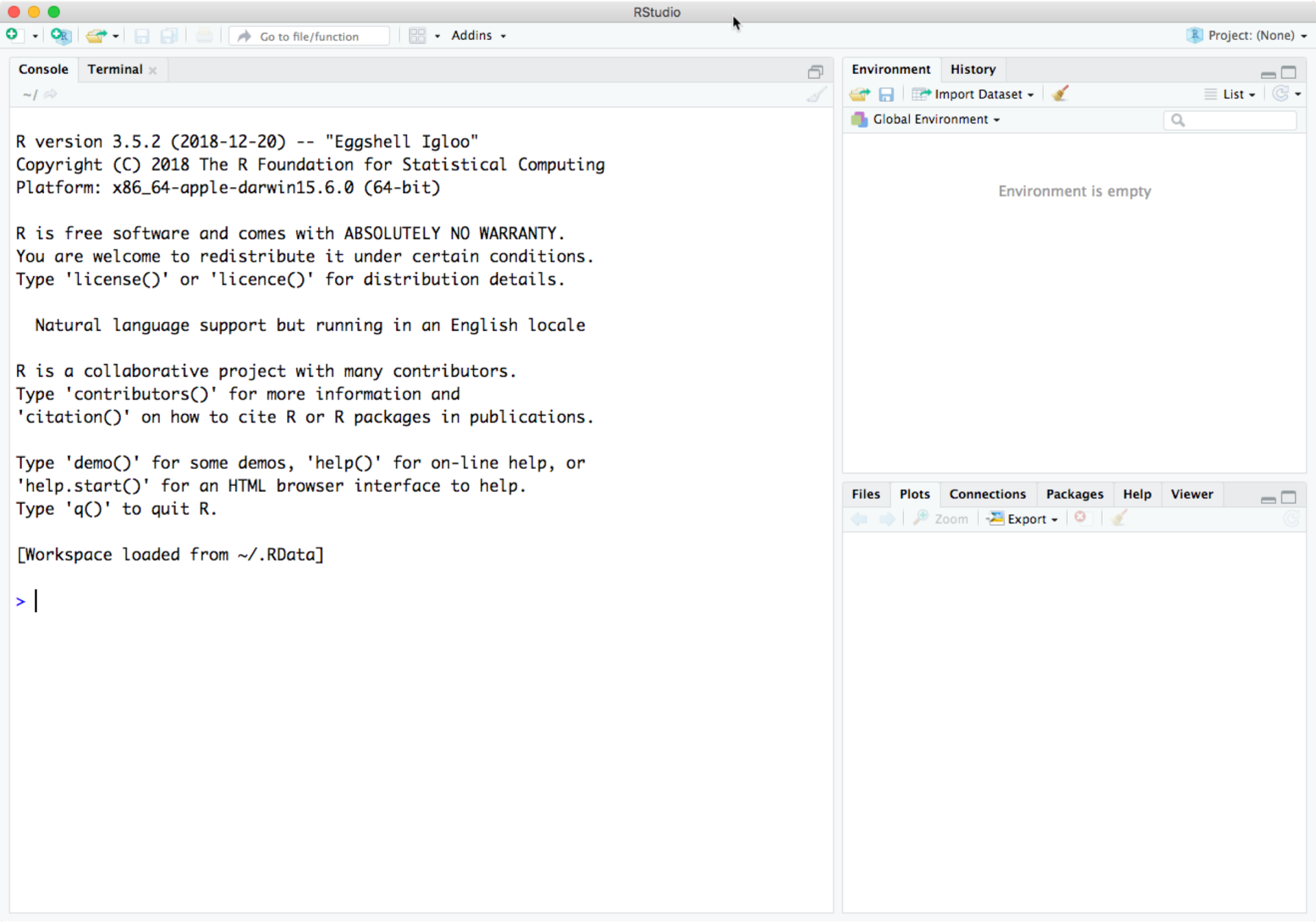


Download the data file to the working directory.



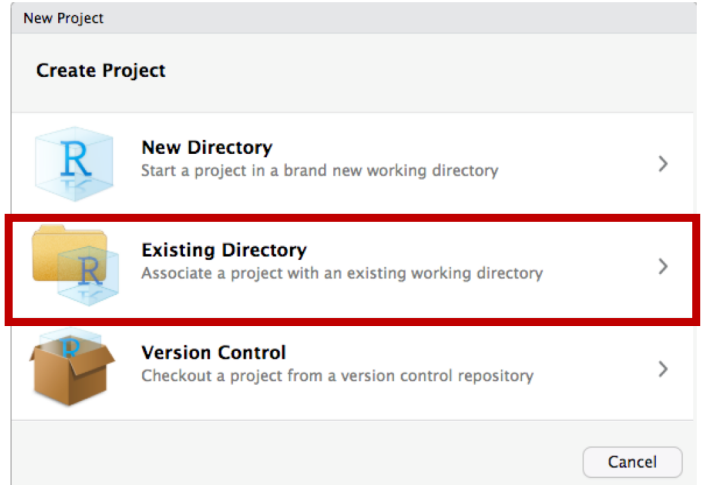
Step 3: Open R Studio

- Open RStudio, just like opening other programs on your computer.
- It will automatically link to R.

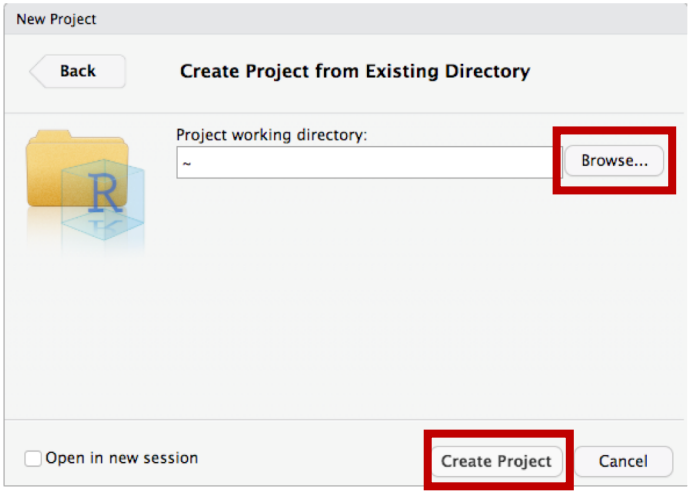

RStudio
Go to file/function
Addins
Project: (None)
Console
Terminal
~/
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
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.
[Workspace loaded from ~/.RData]
>
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Connections
Packages
Help
Viewer
Zoom
Export

Step 4a: Create a New R Project

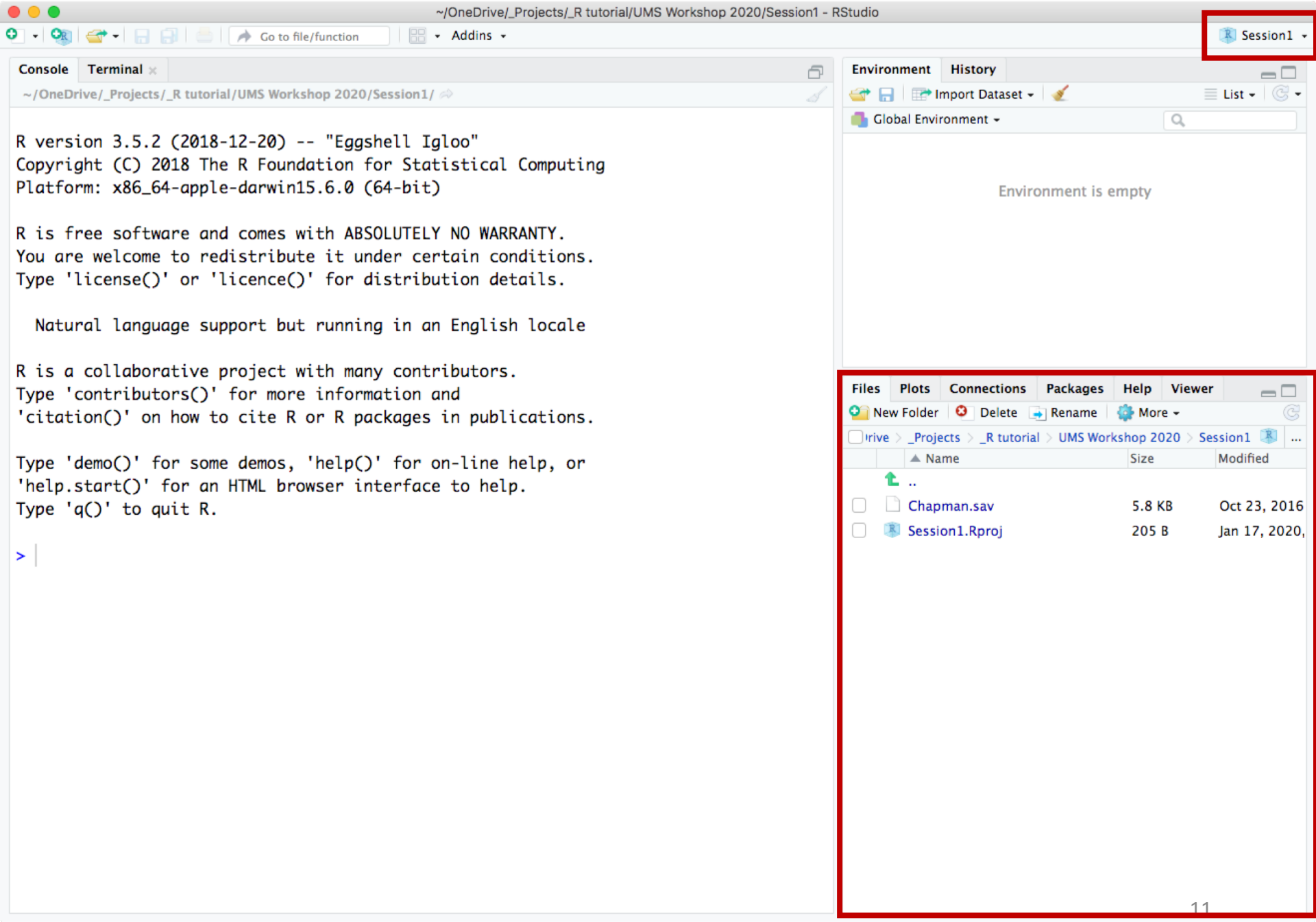
- Click on “Create a project” button  .



- Click on “Existing Directory”.

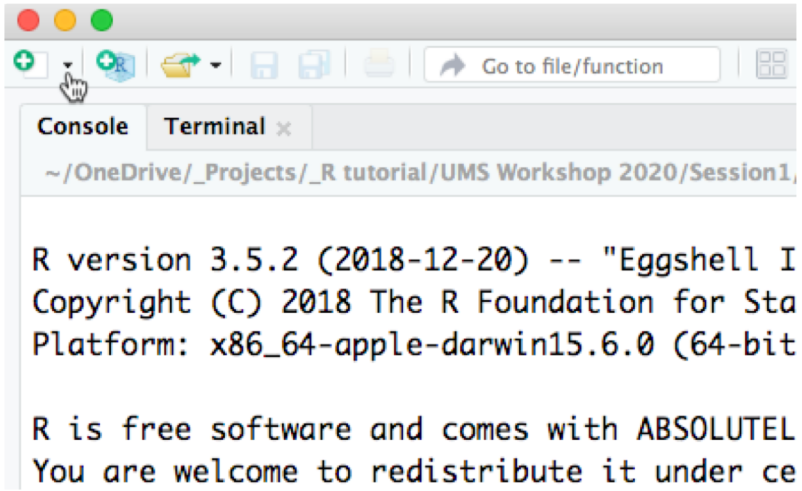


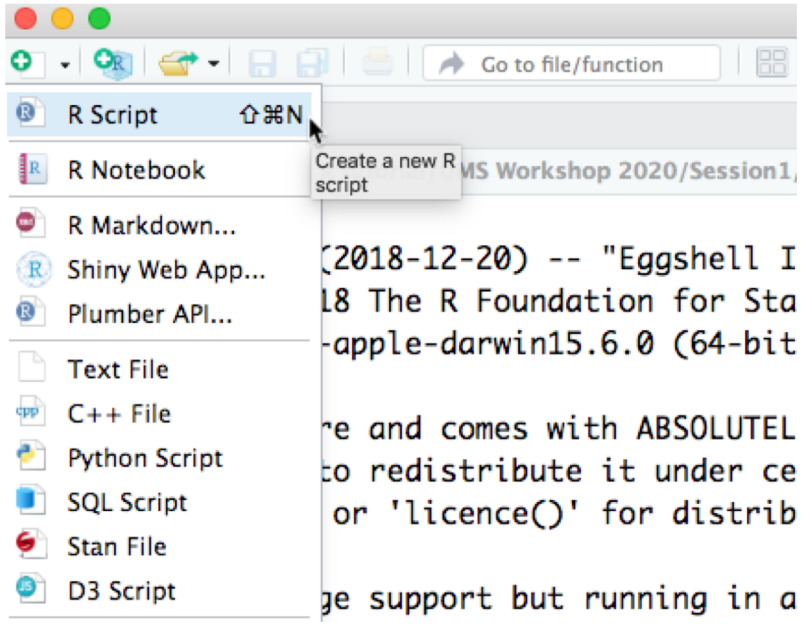
- Click on “Browse…”, and select the target working directory folder, where the data file is.
- Then, click on “Create Project”.

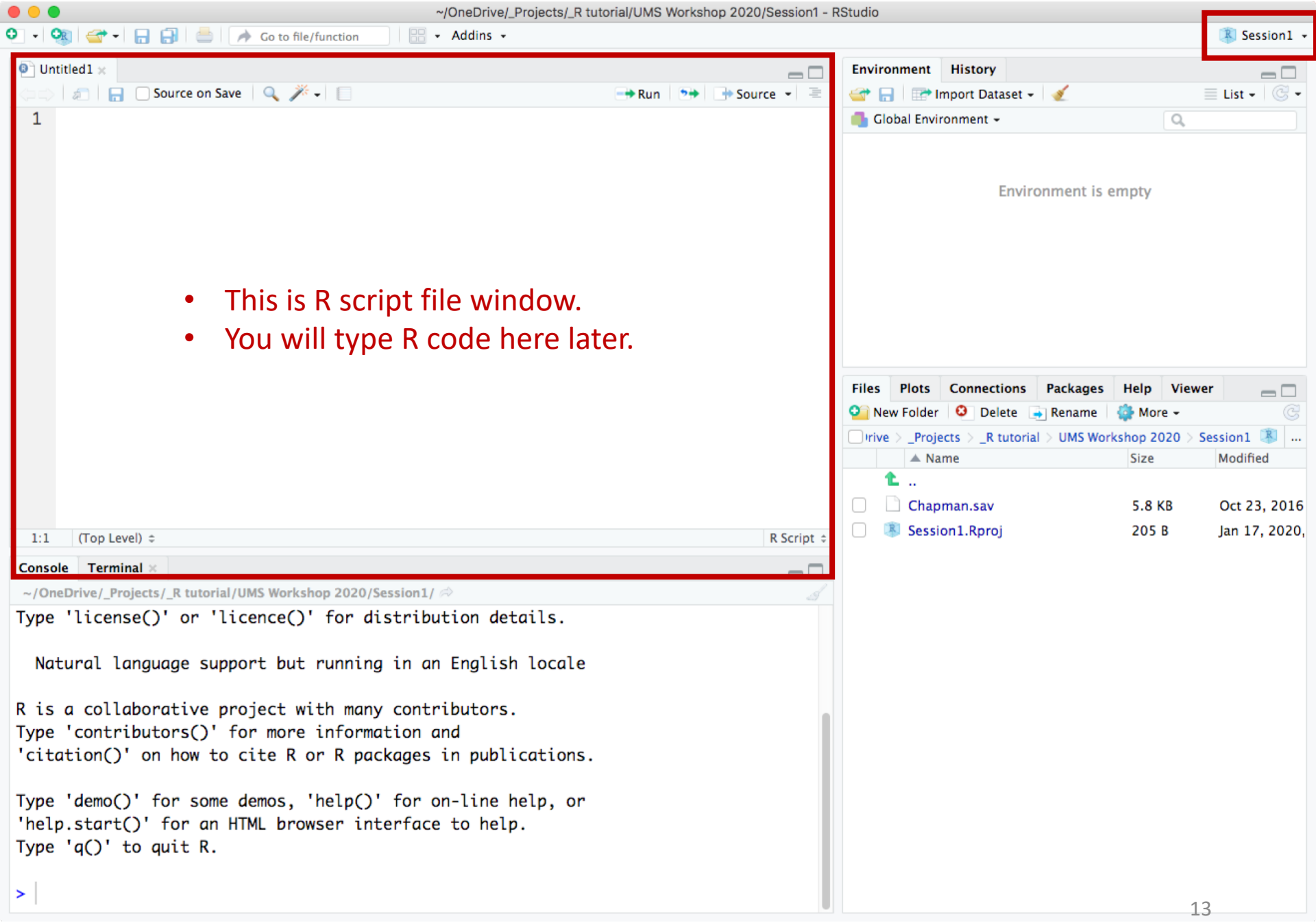

~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1 - RStudio
Go to file/function
Addins
Session1
Console
Terminal
~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1/
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.
Natural language support but running in an English locale
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
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Delete
Rename
More
rive
_Projects
_R tutorial
UMS Workshop 2020
Session1
Name
Size
Modified
..
Chapman.sav
5.8 KB
Oct 23, 2016
Session1.Rproj
205 B
Jan 17, 2020,

Steps 4b: Open a New R Script File and Save

- Open a new script file.
  - Choose an “R Script” option from the pull-down menu under  button.

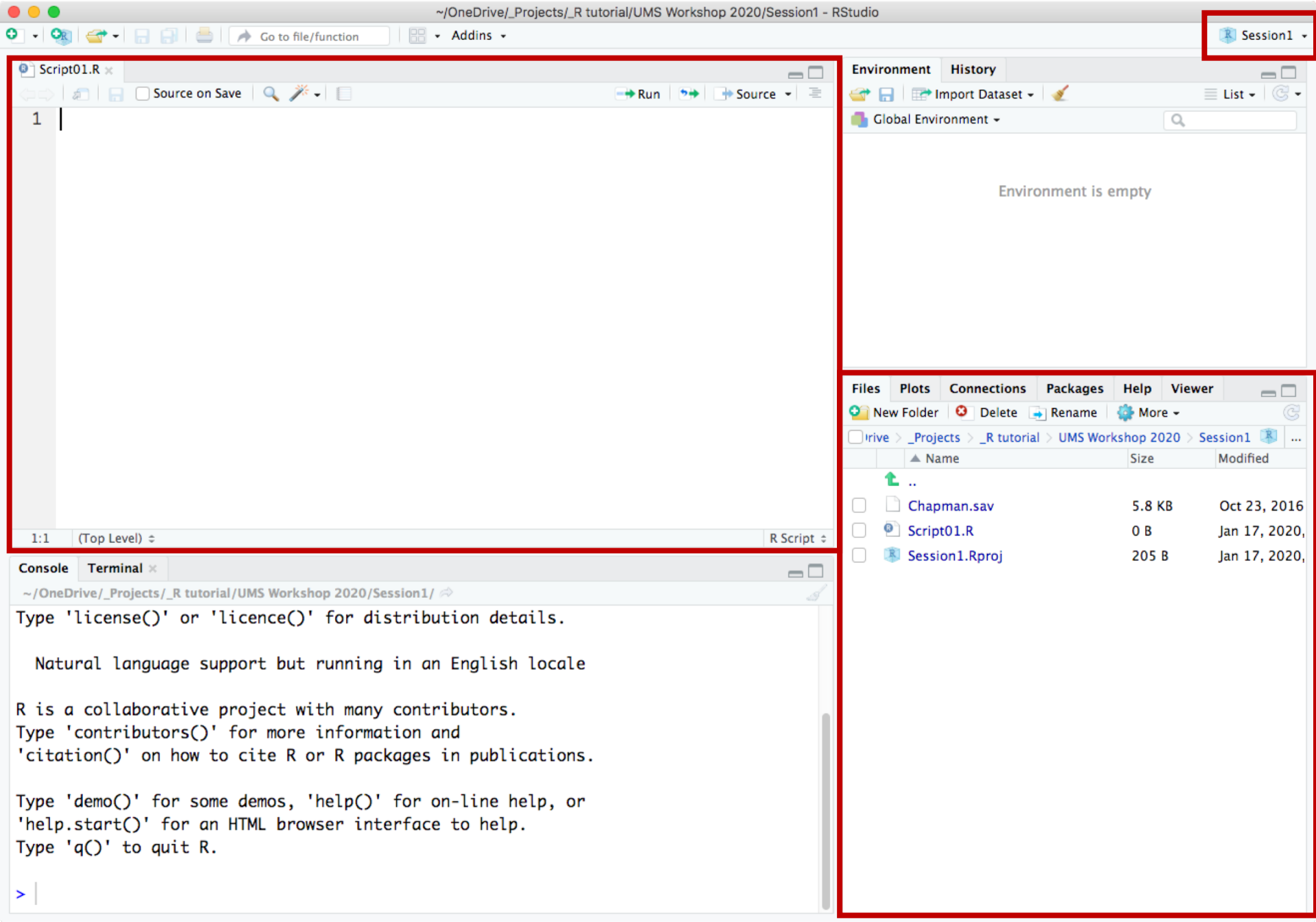


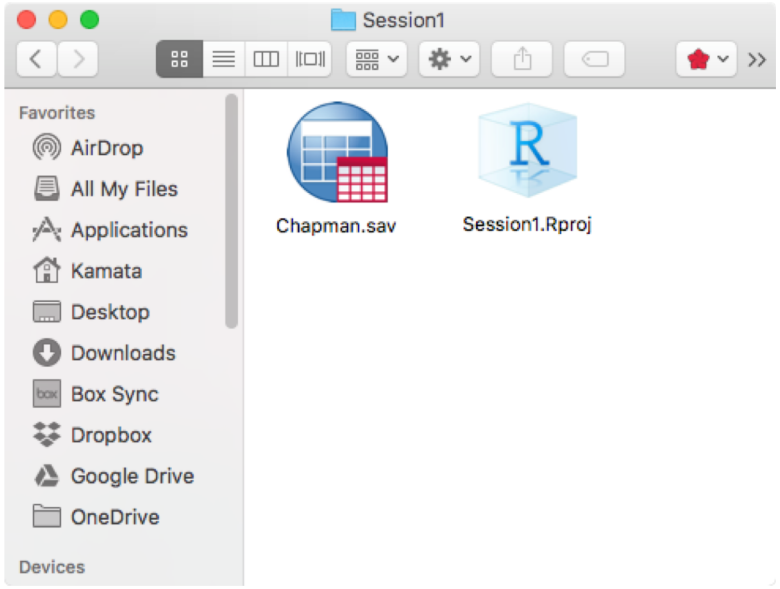


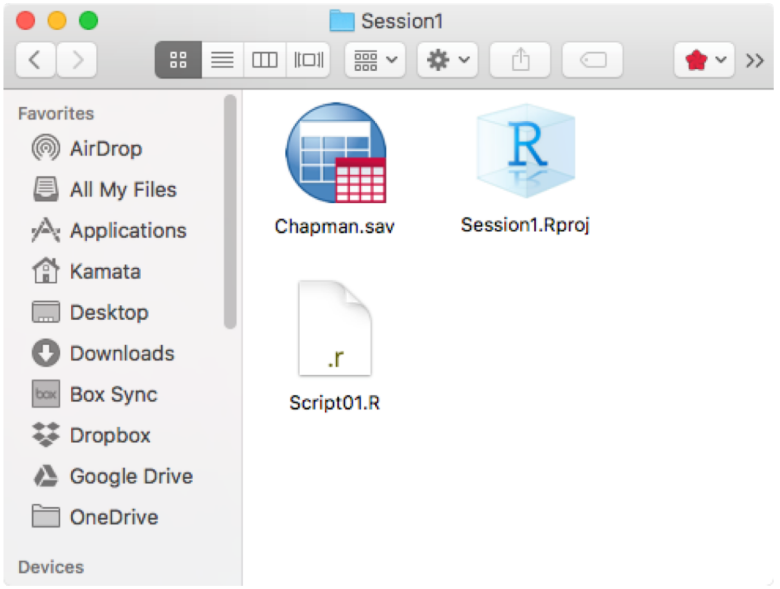

~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1 - RStudio
Go to file/function
Addins
Session1
Untitled1
Source on Save
Run
Source
1
• This is R script file window.
• You will type R code here later.
1:1
(Top Level)
R Script
Console
Terminal
~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1/
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
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
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Delete
Rename
More
_Projects
_R tutorial
UMS Workshop 2020
Session1
Name
Size
Modified
Chapman.sav
5.8 KB
Oct 23, 2016
Session1.Rproj
205 B
Jan 17, 2020,

Steps 4b: Open a New R Script File and Save

- Save the script file.
  - Click  icon.
  - The file will be saved in the target working directory by default.


~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1 - RStudio
Go to file/function
Addins
Session1
Script01.R
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Delete
Rename
More
_Projects
_R tutorial
UMS Workshop 2020
Session1
Name
Size
Modified
Chapman.sav
5.8 KB
Oct 23, 2016
Script01.R
0 B
Jan 17, 2020,
Session1.Rproj
205 B
Jan 17, 2020,
Console
Terminal
~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1/
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
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.
>


Session1
Favorites
AirDrop
All My Files
Applications
Kamata
Desktop
Downloads
Box Sync
Dropbox
Google Drive
OneDrive
Devices
Chapman.sav
Session1.Rproj


Session1
Favorites
AirDrop
All My Files
Applications
Kamata
Desktop
Downloads
Box Sync
Dropbox
Google Drive
OneDrive
Devices
Chapman.sav
Session1.Rproj
.r
Script01.R

Step 5: Read Data into R (1)

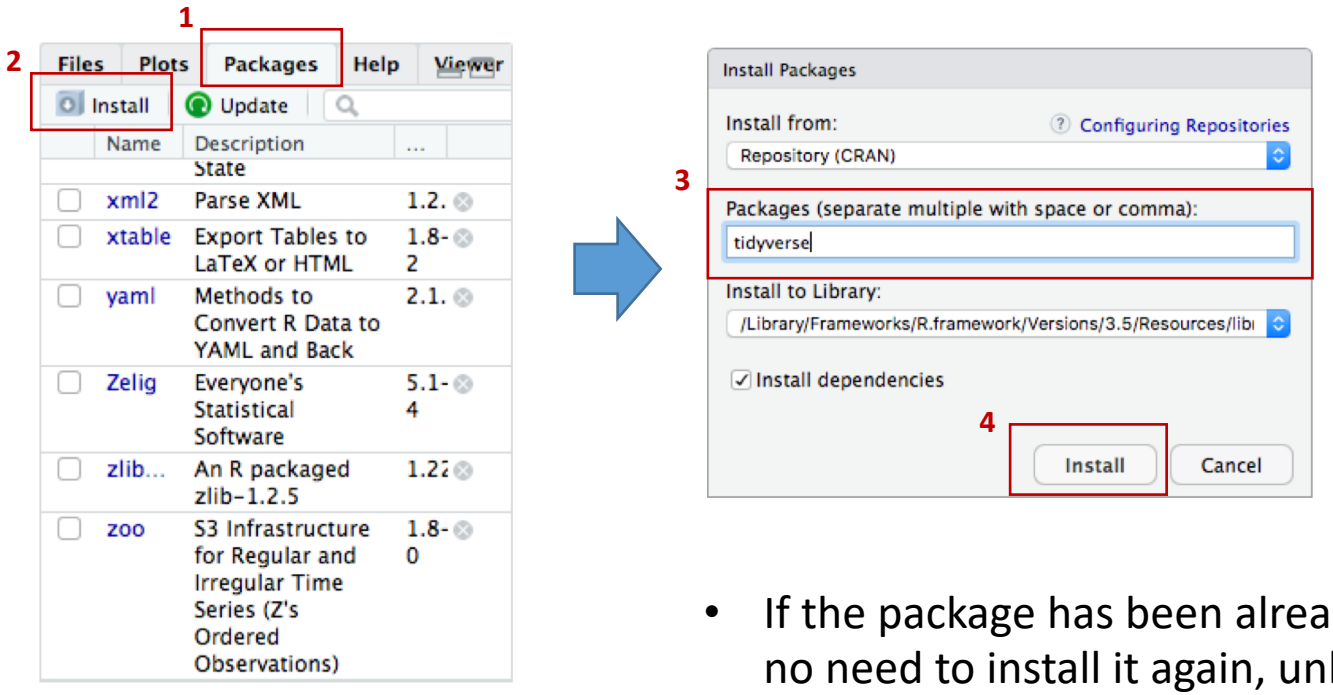
- R can read many different data file formats, including
  - SPSS, CSV, SAS, STATA, Excel, etc.
- To read SPSS data, use `read_spss` function in `haven` package.
  - Other proprietary data formats, including SAS and STATA data can be read by other functions in `haven` package.
- To read CSV data, use `read_csv` function in `readr` package.
  - Also, `read.csv` function in the base package can be used.
- To reading Excel file, use use `read_excel` function in `readxl` package, for example.

Aside: Packages for R

- R comes with only limited set of base packages.
- To do something realistic, we almost always need to use other packages.
  - For example, `haven` package is not part of the base packages, so we need to install and load it.
  - Let's install a package called `tidyverse`, which contains `haven` package and other useful packages.

Aside: Install a Package

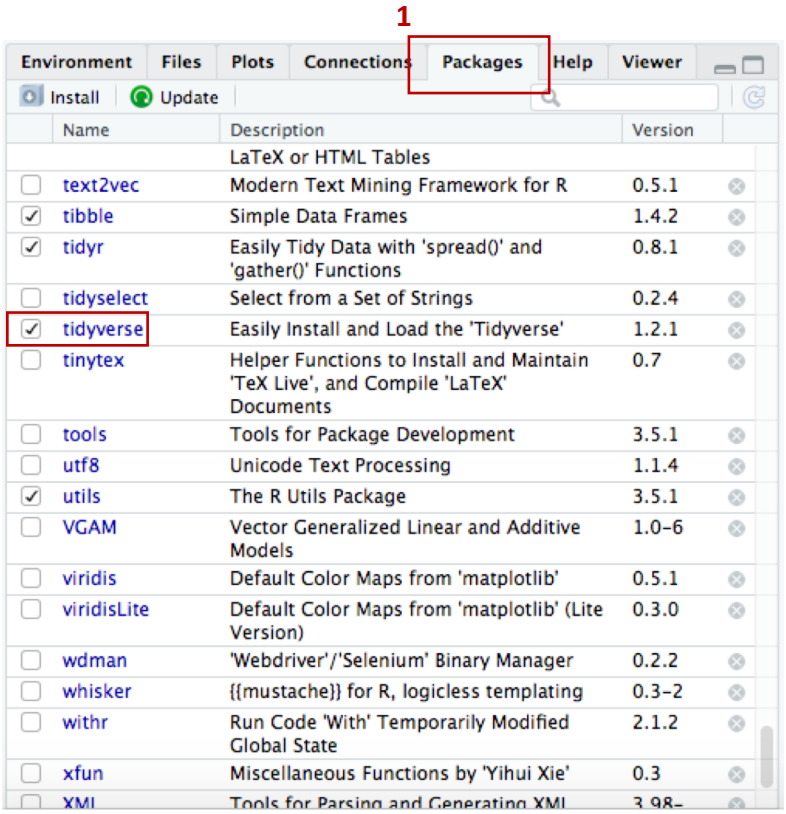
- Let's install the `tidyverse` package:
  - Run a command `install.packages("tidyverse")`
  - Alternatively, install it by point-and-click operations. (Recommended)



- If the package has been already installed, no need to install it again, unless it needs to be updated to a newer version.

Aside: Load a Package

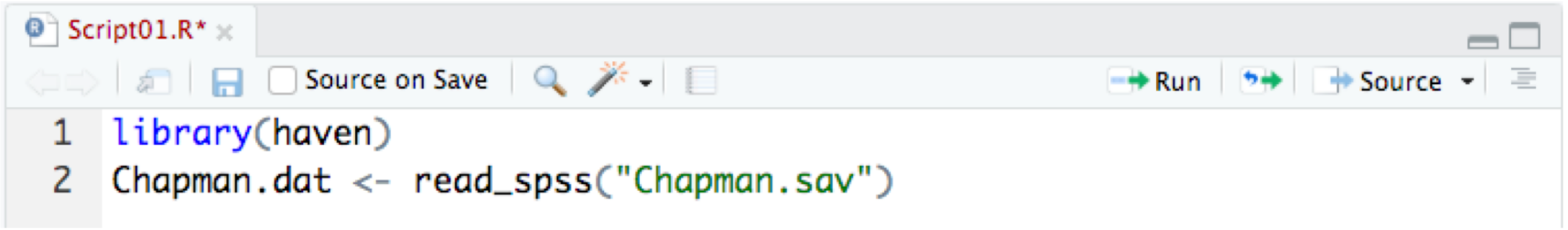
- For example, to load `tidyverse` package:
  - Run a command `library(tidyverse)` (Recommended)
  - Alternatively, we can load it by point-and-click operations.



- A package must be loaded every time you start a session, regardless starting new work or resuming existing work.

Step 5: Read SPSS Data File (2)

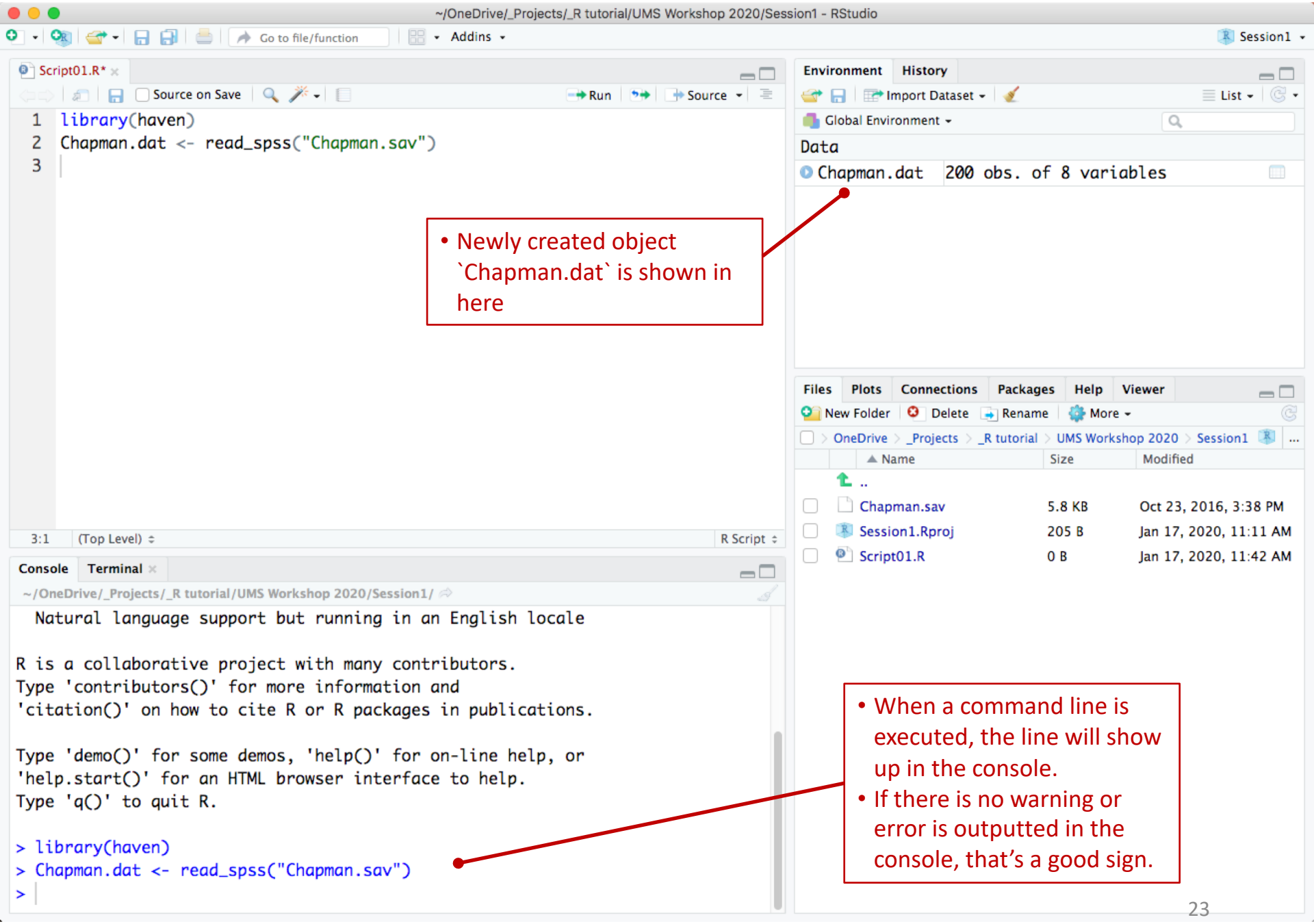
- Use `read_spss` function in `haven` package.



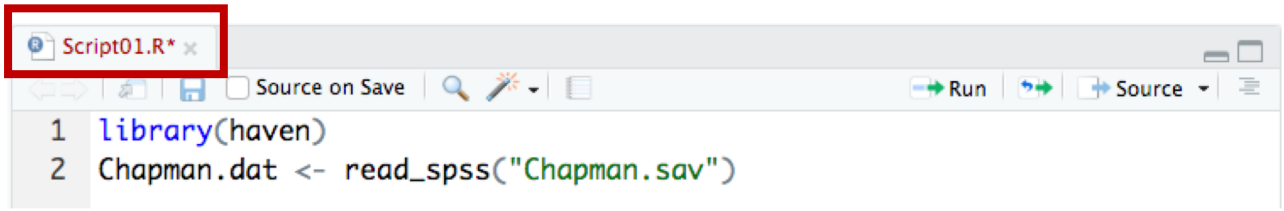
- Running line 2 will create a data object on R, called `Chapman.dat`

Aside: How to Run R Syntax Lines

- To run a syntax line by line,
  1. Place the cursor on the target line.
  2. Click the Run button.
     - Alternatively, use keyboard shortcut **Ctrl+Enter** (Win) or **Command+Enter** (Mac).
- To run multiple lines sequentially,
  1. Highlight the target lines.
  2. Click the Run button, or type **Ctrl+Enter** (Win) or **Command+Enter** (Mac).


~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1 - RStudio
Go to file/function
Addins
Session1
Script01.R*
Source on Save
Run
Source
1 library(haven)
2 Chapman.dat <- read_spss("Chapman.sav")
3
3:1
(Top Level)
R Script
Environment
History
Import Dataset
List
Global Environment
Data
Chapman.dat
200 obs. of 8 variables
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Delete
Rename
More
OneDrive
_Projects
_R tutorial
UMS Workshop 2020
Session1
Name
Size
Modified
..
Chapman.sav
5.8 KB
Oct 23, 2016, 3:38 PM
Session1.Rproj
205 B
Jan 17, 2020, 11:11 AM
Script01.R
0 B
Jan 17, 2020, 11:42 AM
Console
Terminal
~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1/
Natural language support but running in an English locale
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.
> library(haven)
> Chapman.dat <- read_spss("Chapman.sav")
>
• Newly created object `Chapman.dat` is shown in here
• When a command line is executed, the line will show up in the console.
• If there is no warning or error is outputted in the console, that's a good sign.

Aside: Saving R Script File



- When the script file name is red with an asterisk, it means that the changes on the script file is unsaved.
- It is highly recommended to save changes periodically.
  - Click the button to save.
  - Alternatively, use keyboard shortcut **Ctrl+S** (Win) or ⌘**+S** (Mac).

Step 5: Read SPSS Data File (3)

- To view the imported data in a spread sheet format, click right on the data object name, or  icon for the target data object.

Script01.R | Chapman.dat

Filter

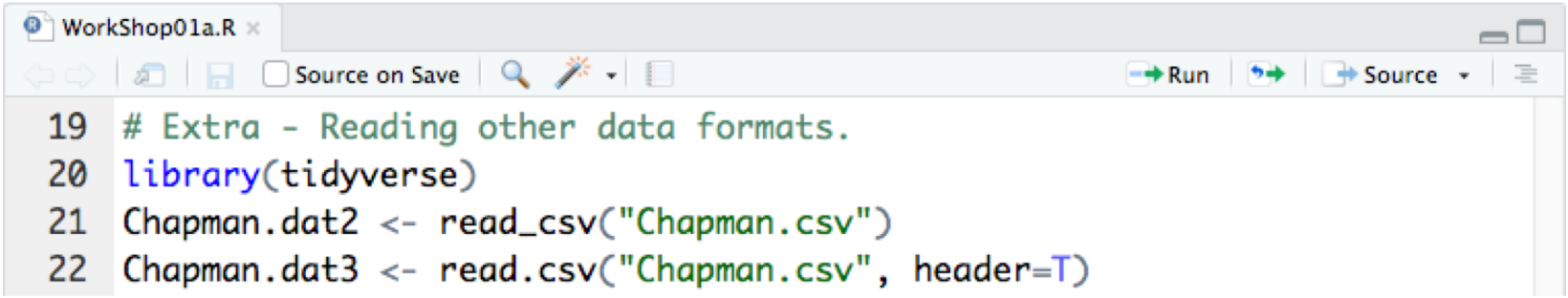
	ID CASE #	AGE	SYSBP SYSTOLIC BLOOD PRESSURE	DIABP DIASTOLIC BLOOD PRESSURE	CHOLES CHOLESTEROL LEVEL (MG/DL)	HT HEIGHT
1	1	44	124	80	254	NA
2	2	35	110	70	240	73
3	3	41	114	80	279	68
4	4	31	100	80	284	68
5	5	61	190	110	315	68
6	6	61	130	88	250	70
7	7	44	130	94	298	68
8	8	58	110	74	384	67
9	9	52	120	80	310	66
10	10	52	120	80	337	67

Aside: Variable Names, etc.

- Variable name has to start with a Roman letter.
  - No numeric, like 4 and 0.
  - No special character, like % and $.
- R is case sensitive.
  - For example, a variable name `StuID` is not the same as a variable name `stuid`.
  - Similarly, object names, function names, and package names are all case sensitive.
    - For example, the function `read_spss` will not run by typing `read_SPSS`.

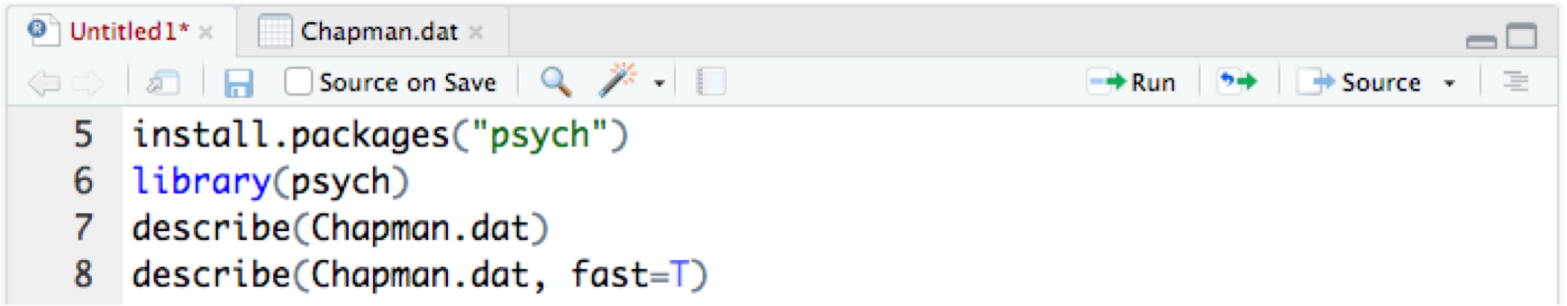
Aside: Read CSV data

- CSV data format is one of most convenient ways to manage data on R.
  - The first row of a CSV data file can be a list of variable names.
- Use `read_csv` function in the `readr` package, which will be loaded if the `tidyverse` package is loaded, or `read.csv` function in the base package just like this:



Steps 6 & 7: Let's Do Something with the Data (1)

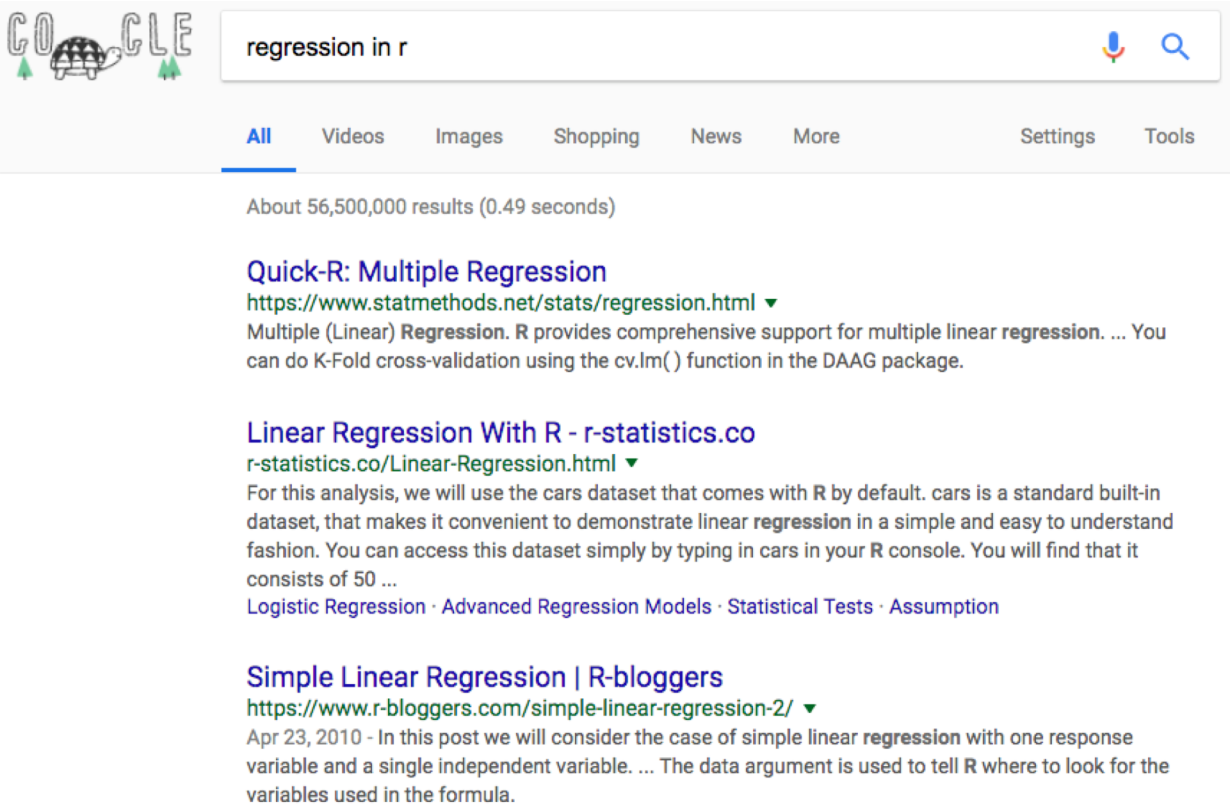
- Descriptive statistics for all variables:
  - Use `describe` function in `psych` package.
- Type the following lines and run each line.



- Again, if the package has been already installed, no need to install it again.
- However, a package must be loaded every time you start a session, regardless starting new work or resuming existing work.

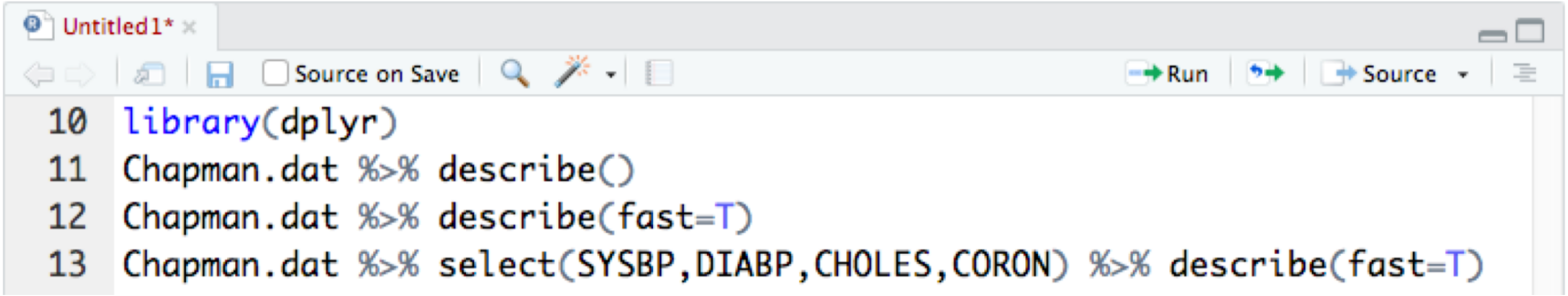
Aside: How to find a function & package

- For example, if I am interested in how to do a regression analysis, I would search by “regression in r” on Google.



Steps 6 & 7: Let's Do Something with the Data (2)

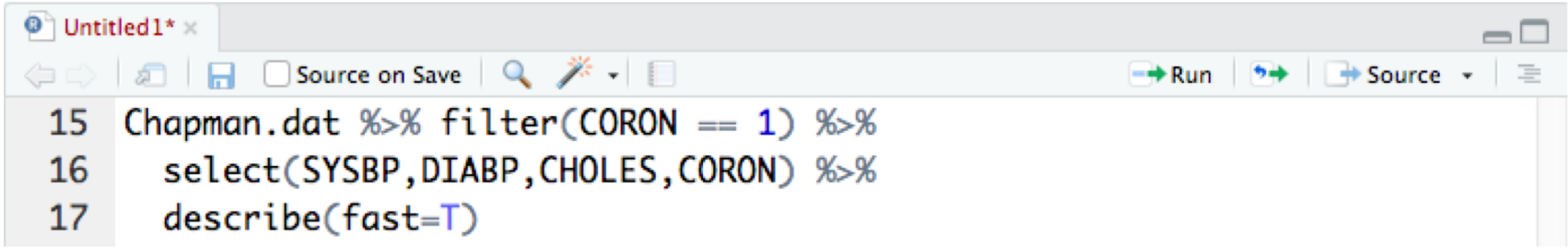
- Basic data manipulation - select only a subset of variables:
  - Use `select` function in `dplyr` package.
- Type the following lines and run each line.



- `%>%` is called a pipe operator, which is part of `dplyr` package.
- In short, the left-hand side of `%>%` will be used as the first entry of the function on the right-hand side of `%>%`.
- It is extremely useful for executing multiple operations sequentially, such as in data manipulations.

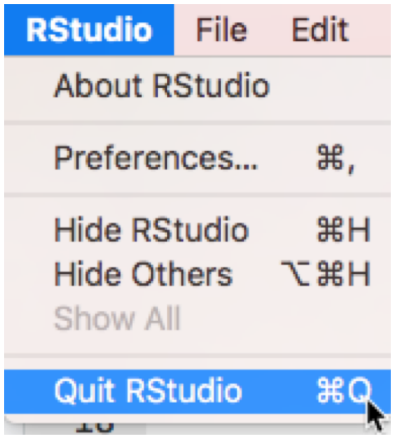
Steps 6 & 7: Let's Do Something with the Data (3)

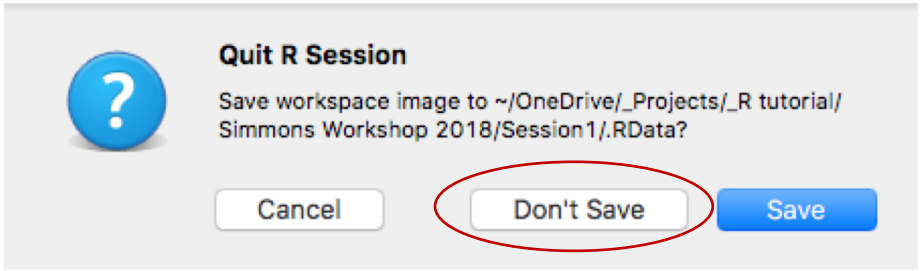
- Select only a subset of cases and variables:
  - Use `filter` and `select` functions in `dplyr` package sequentially.
- Type the following lines and run.



Step 9: Closing R Studio Session

- Be sure to save your syntax for the last time.
- It is highly recommended NOT to save the “workspace”, because you can easily replicate your work by running the saved syntax.

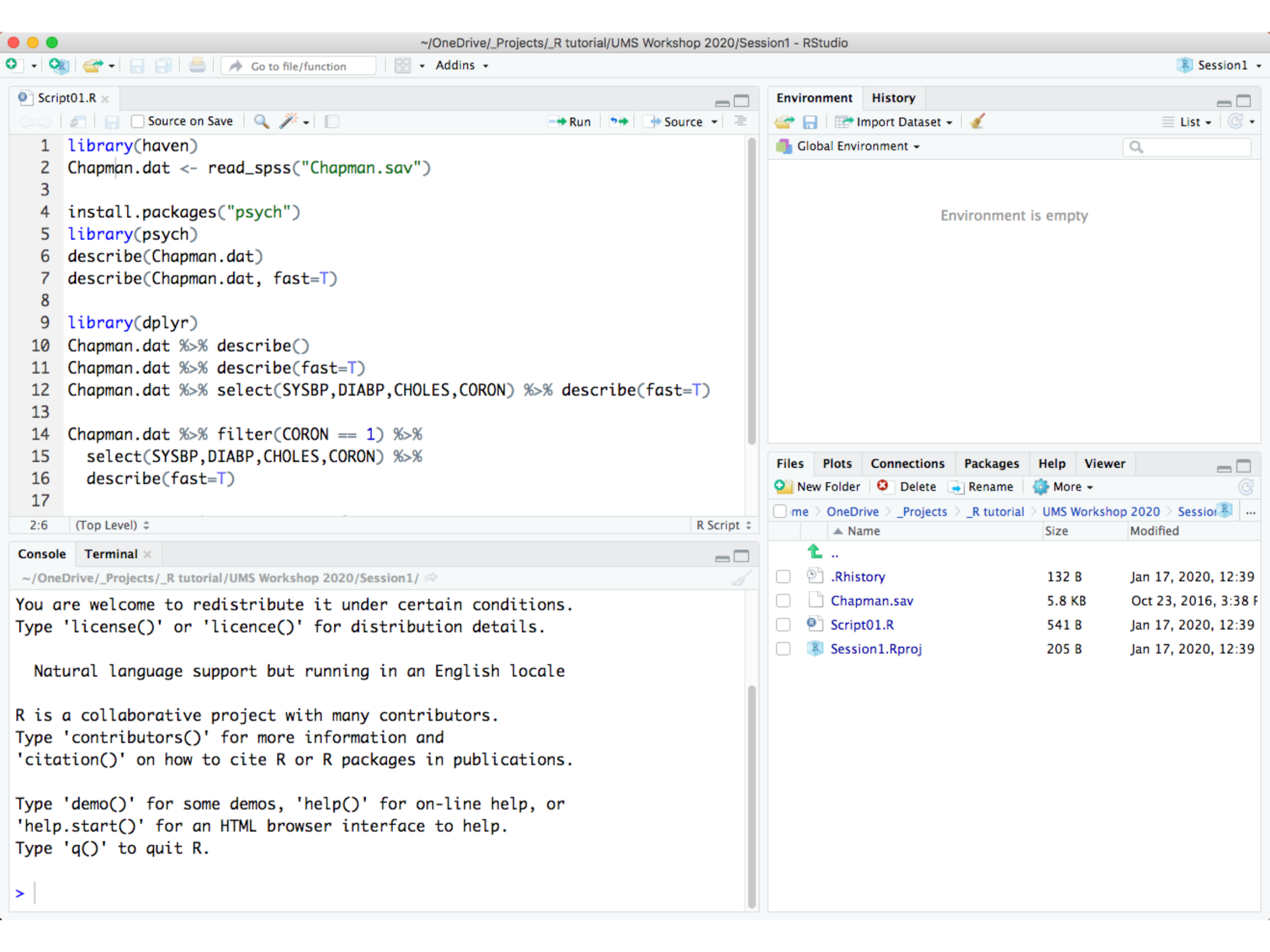




Resuming Existing Work

- Step 4: Open RStudio by double-clicking a saved R Project file (.Rproj file).

  - The target R script file will be automatically open, if it was open when you closed R Studio last time.
  - Otherwise, go to “Files” tab of Rstudion, and click the target R script file name to open.
    - Alternatively, within RStudio, click  button, and locate the target R script file and open.


~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1 - RStudio
Go to file/function
Addins
Session1
Script01.R
Source on Save
Run
Source
1 library(haven)
2 Chapman.dat <- read_spss("Chapman.sav")
3
4 install.packages("psych")
5 library(psych)
6 describe(Chapman.dat)
7 describe(Chapman.dat, fast=T)
8
9 library(dplyr)
10 Chapman.dat %>% describe()
11 Chapman.dat %>% describe(fast=T)
12 Chapman.dat %>% select(SYSBP,DIABP,CHOLES,CORON) %>% describe(fast=T)
13
14 Chapman.dat %>% filter(CORON == 1) %>%
15 select(SYSBP,DIABP,CHOLES,CORON) %>%
16 describe(fast=T)
17
2:6
(Top Level)
R Script
Console
Terminal
~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Session1/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
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
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Delete
Rename
More
me
OneDrive
_Projects
_R tutorial
UMS Workshop 2020
Name
Size
Modified
..
.Rhistory 132 B Jan 17, 2020, 12:39
Chapman.sav 5.8 KB Oct 23, 2016, 3:38 F
Script01.R 541 B Jan 17, 2020, 12:39
Session1.Rproj 205 B Jan 17, 2020, 12:39

Resuming Existing Work

- Step 5: Read data file
  - Run the syntax to read the data file.

- Steps 6-9: The rest is the same as before!

Let's Do It

- Replicate what have been demonstrated.
  1. Create a new folder `FirstRprojedt`, and copy the data file `Chapman.sav`.
  2. Open RStudio, and create a new project for the newly created folder.
  3. Install `tidyverse` package, if you have not.
  4. Open a new script file, and write lines to (1) load `heaven` package, and (2) read the `Chapman.sav` data as `Chapman.dat` .
  5. Save the R script file by naming it `FirstRscript`.
  6. Close RStudio.
  7. Re-open RStudio by double-clicking `FirstRproject.Rprj`.

Part II
Basic Statistics on R

Basic Statistics

- We will learn:
  - Basic statistics
  - Basic graphics by `ggplot`
  - Basic data manipulations
- Let’s look at:
  - `01_BasicStats.R`

Part III
R Markdown

R Markdown

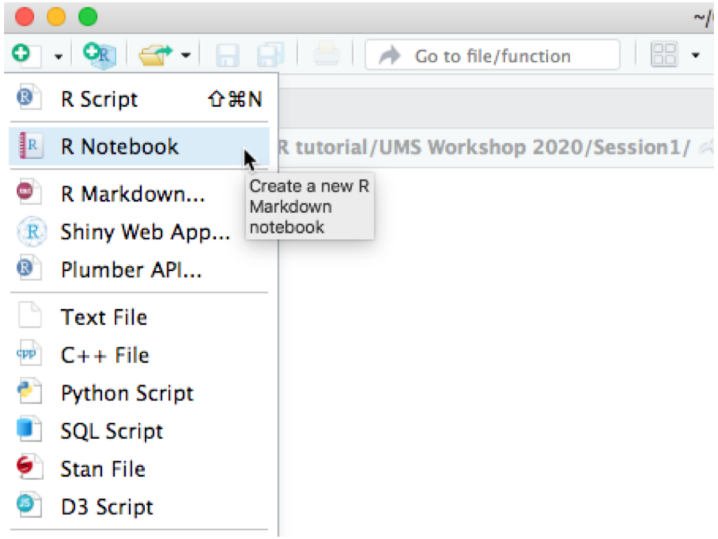
- R Markdown is a way to produce a document that can integrate what you do on R and text you want to type in.
  - R code
  - R output (e.g., table, graph)
  - Your comments about analysis
- It is an effective way to run, view, save, and share your analysis on R.
  - R Markdown can produce many kinds of document (e.g., presentation slides and APA style manuscript) in different file formats (HTML, PDF, and MSWord).
- The most basic and interactive form of R Markdown is ”R Notebook”, which produce an HTML file.

R Notebook

- Besides producing an external file, R Notebook provides a convenient platform to work on R codes.
  - R code and output will be presented on the same window.
  - Output remains on the R Notebook document unless codes are modified for the code chunk.
    - No need to re-run to review outputs from previous run or session.

Start R Notebook File

- First, `rmarkdown` package needs to be installed.
- Open an R Notebook file by choosing `R Notebook` command from `File` pull down menu  .



- This will open a template R Notebook file that contains minimal but sufficient example to produce an HTML file.
- Once this file is saved, an HTML file will be produced in the workspace.

~/OneDrive/_Projects/_R tutorial/UMS Workshop 2020/Se

Go to file/function | Addins

Notebook01.Rmd

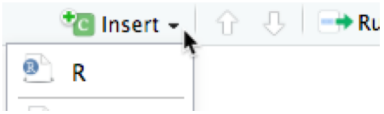
Preview | Insert | Run

```
1  ---
2  title: "R Notebook"
3  output: html_notebook
4  ---
5
6  This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you
   execute code within the notebook, the results appear beneath the code.
7
8  Try executing this chunk by clicking the *Run* button within the chunk or
   by placing your cursor inside it and pressing *Cmd+Shift+Enter*.
9
10 ```{r}
11 plot(cars)
12 ```
13
14 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or
   by pressing *Cmd+Option+I*.
15
16 When you save the notebook, an HTML file containing the code and output
   will be saved alongside it (click the *Preview* button or press
   *Cmd+Shift+K* to preview the HTML file).
17
18 The preview shows you a rendered HTML copy of the contents of the editor.
   Consequently, unlike *Knit*, *Preview* does not run any R code chunks.
   Instead, the output of the chunk when it was last run in the editor is
   displayed.
19
20
```

- This is a "YAML header", where some information about the document is provided, by being tagged by: --- at the beginning and at the end.

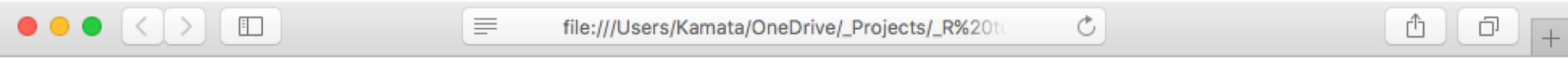
- Some texts are added here. These will appear in the HTML file to be produced.
- Some formatting appear here:
  - Hyperlink
  - Italicized text
- To change paragraph, provide an empty line in between two paragraphs.

- This is a "code chunk", where R code is placed, by being tagged by: ```{r} at the beginning and ``` at the end.
- To insert the code chunk tags:
  - Use "Insert R" pull-down menu.



  - Use keyboard shortcut:
    - ⌥ + ⌘ + I (Mac).
    - cmd + option + I (Windows).
  - Manually type.

- Some more texts are added between code chunks.



R Notebook

Code

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

Hide

```
plot(cars)
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

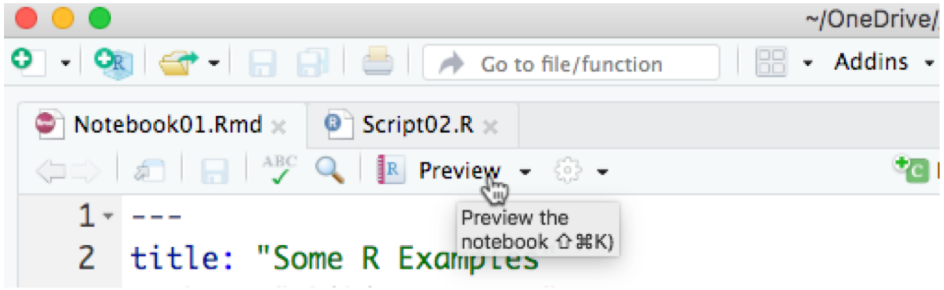
When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

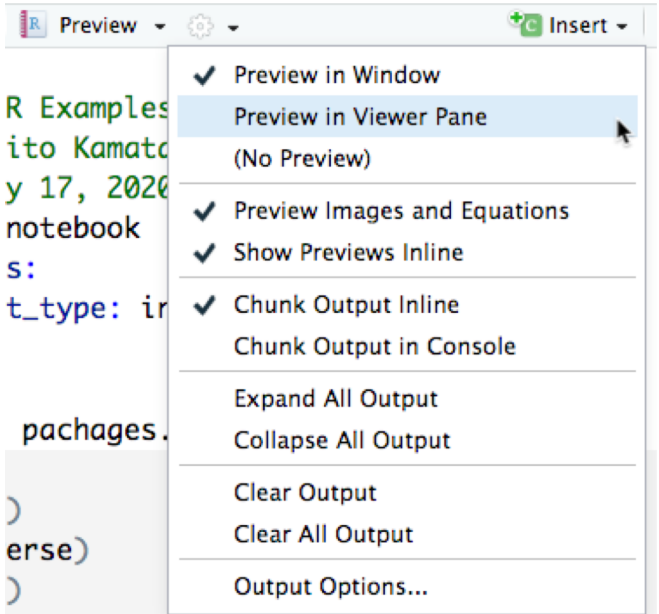
- Note that R code is in a gray box. This is a default to R Markdown.
- Texts outside of the code chunk is printed.
- However, the result of `plot(cars)` is not printed.
  - We need to actually run the code chunk.

Aside: Preview R Notebook File

- Another way to view the Notebook HTML file is by "Preview".

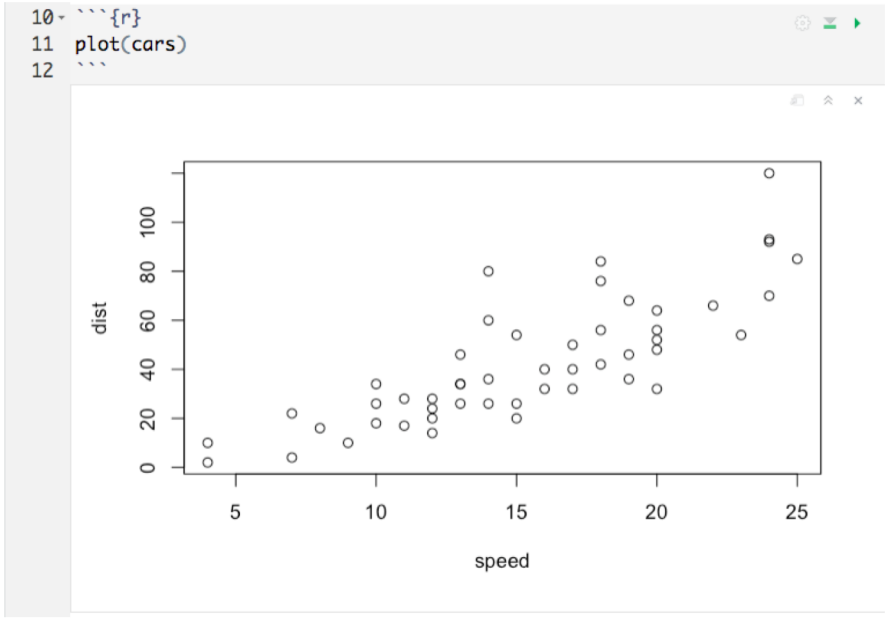


- By default, a separate preview window will open.
- Alternatively, a preview can be open within the main RStudio window.



Run Code Chunk in R Notebook File

- One way to run a code chunk is to click an arrow button  for the target code chunk.
- This will create an output window right below the code chunk.



- Once the R Notebook file is saved, the HTML file will be updated.

R Notebook

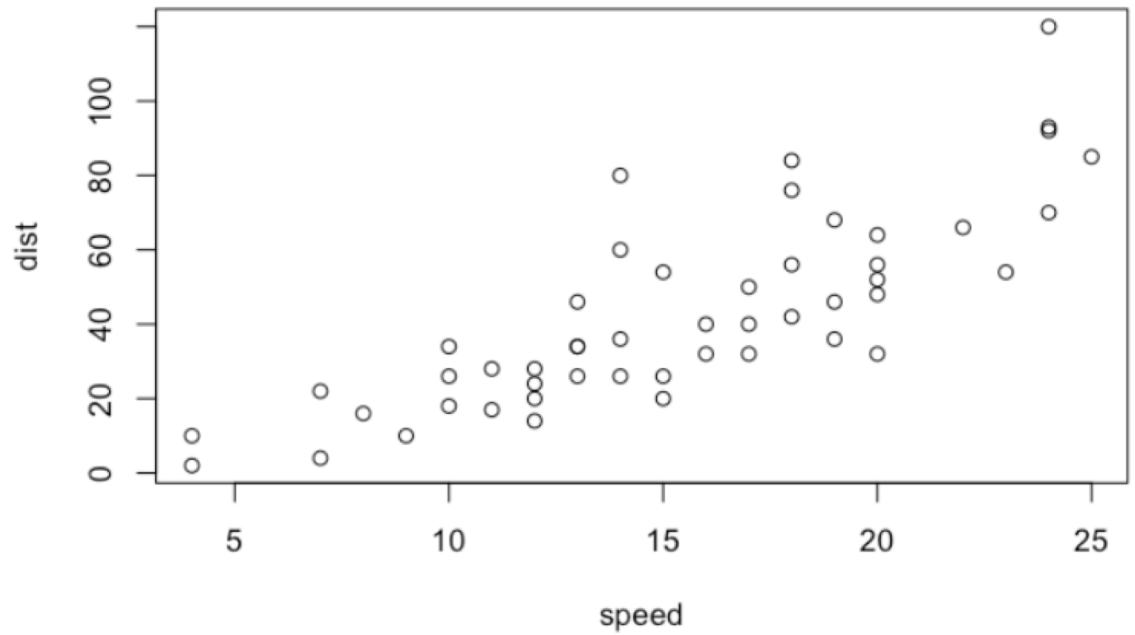
Code

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

Hide

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

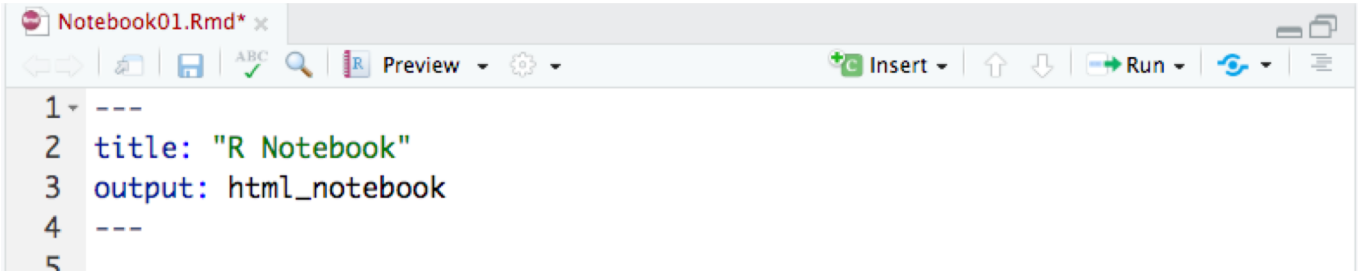
When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

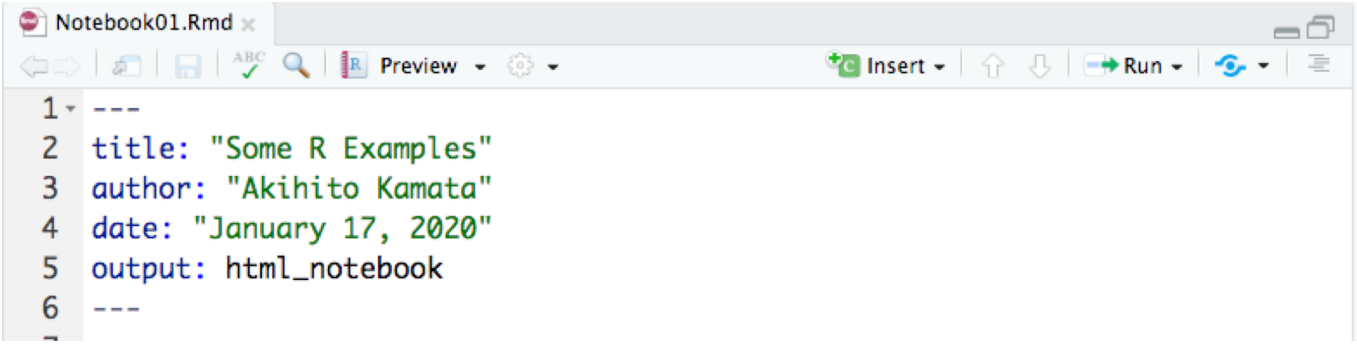
Edit R Notebook File

- We can use the template Notebook file to get what we need.

1. Delete all the contents, except the YAML header.

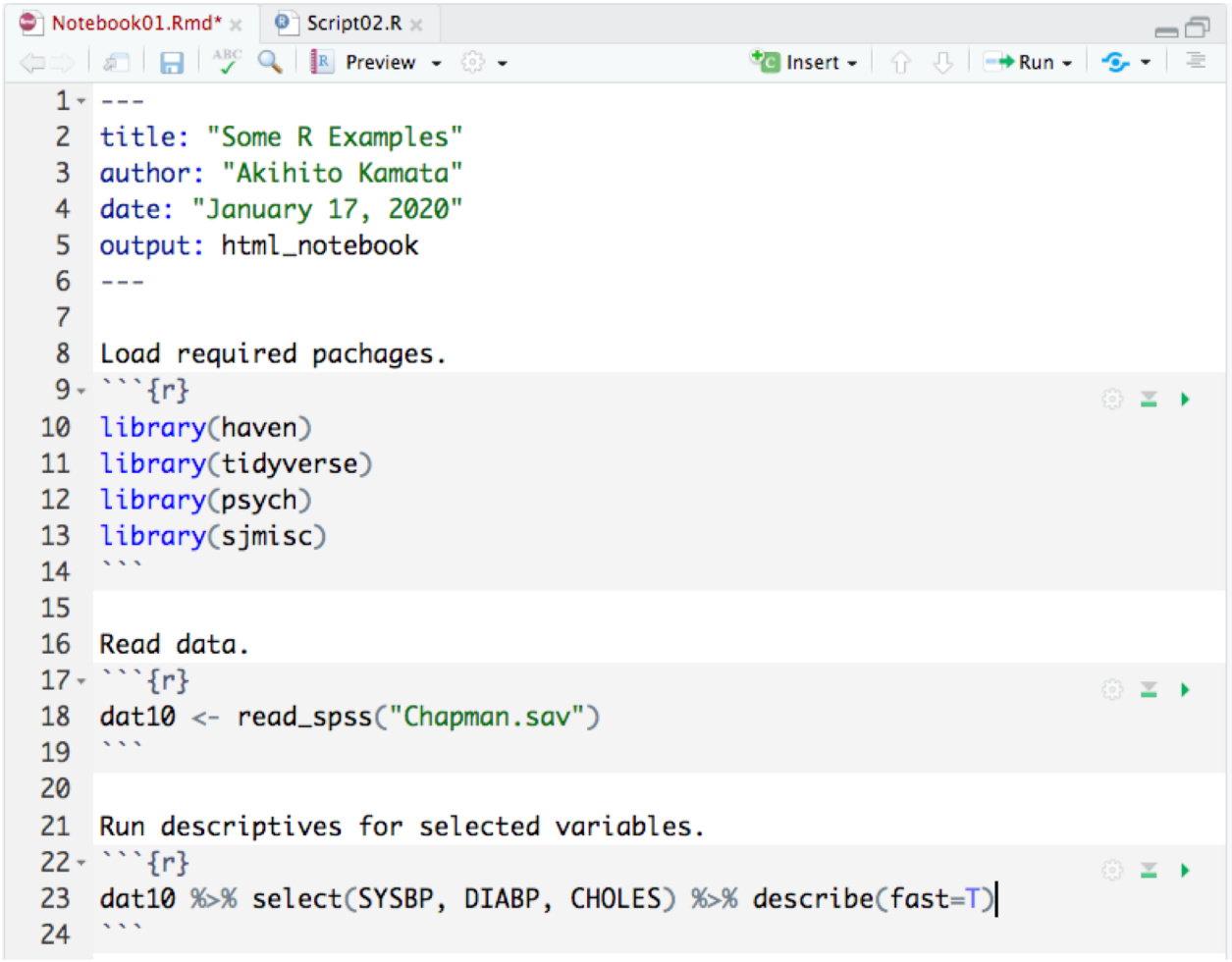


2. Change the title. Perhaps add your name as an author and date, too.



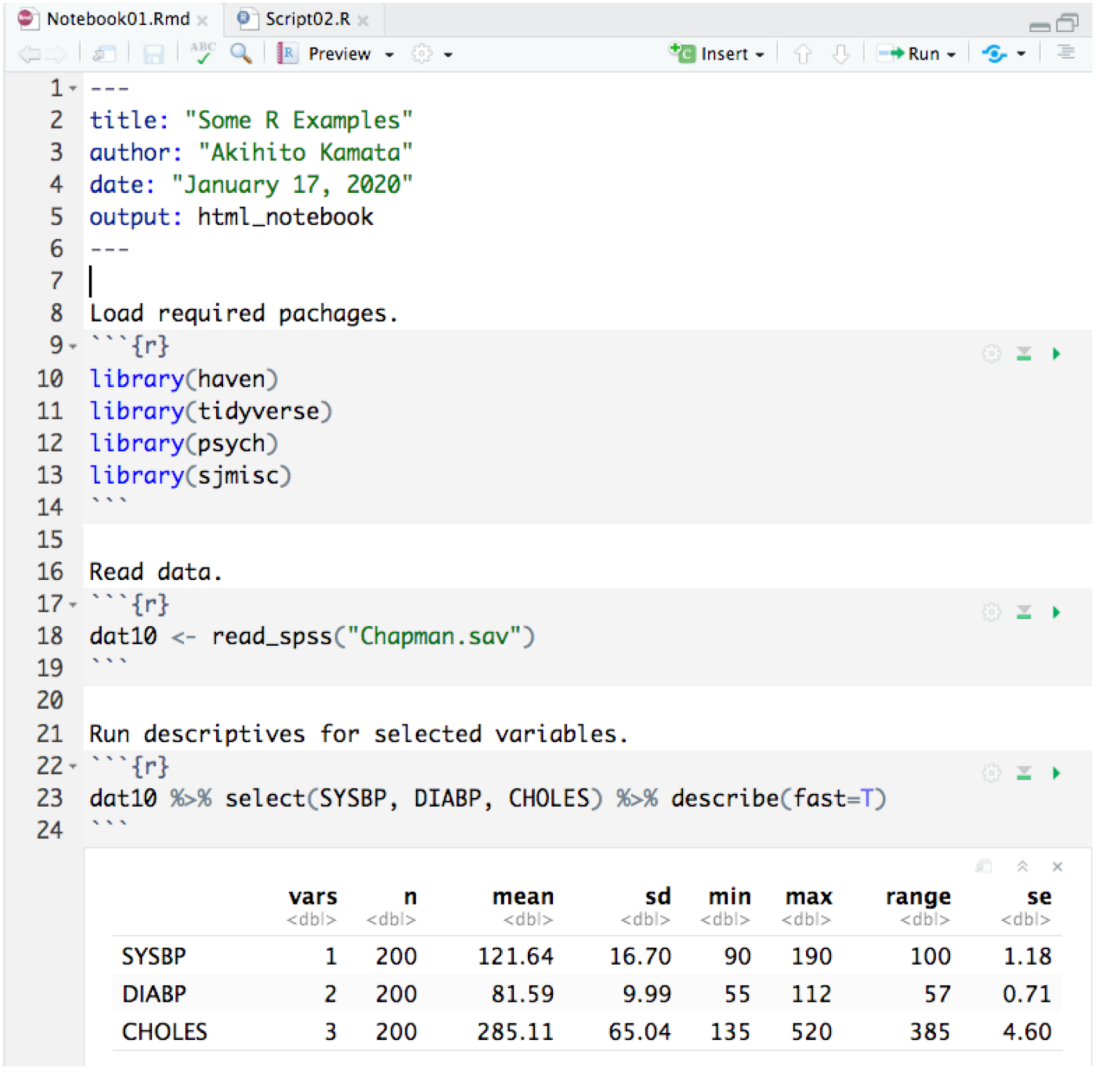
Edit R Notebook File

3. Add some texts and code chunks.



Edit R Notebook File

4. Run code chunks, and save the Markdown file.



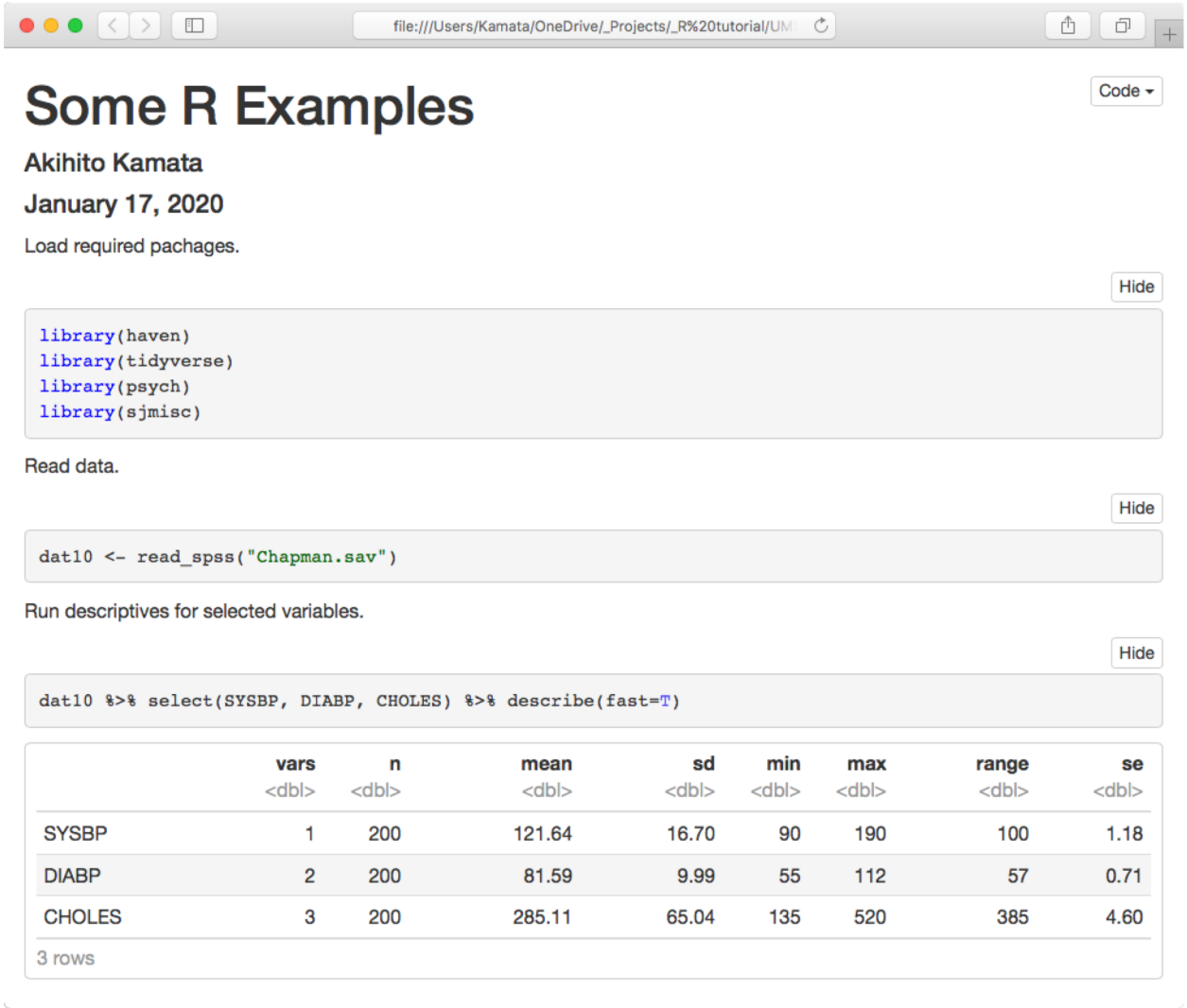
	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	min <dbl>	max <dbl>	range <dbl>	se <dbl>
SYSBP	1	200	121.64	16.70	90	190	100	1.18
DIABP	2	200	81.59	9.99	55	112	57	0.71
CHOLES	3	200	285.11	65.04	135	520	385	4.60

Actually, viewing only R Notebook file with results will enhance our productivity.

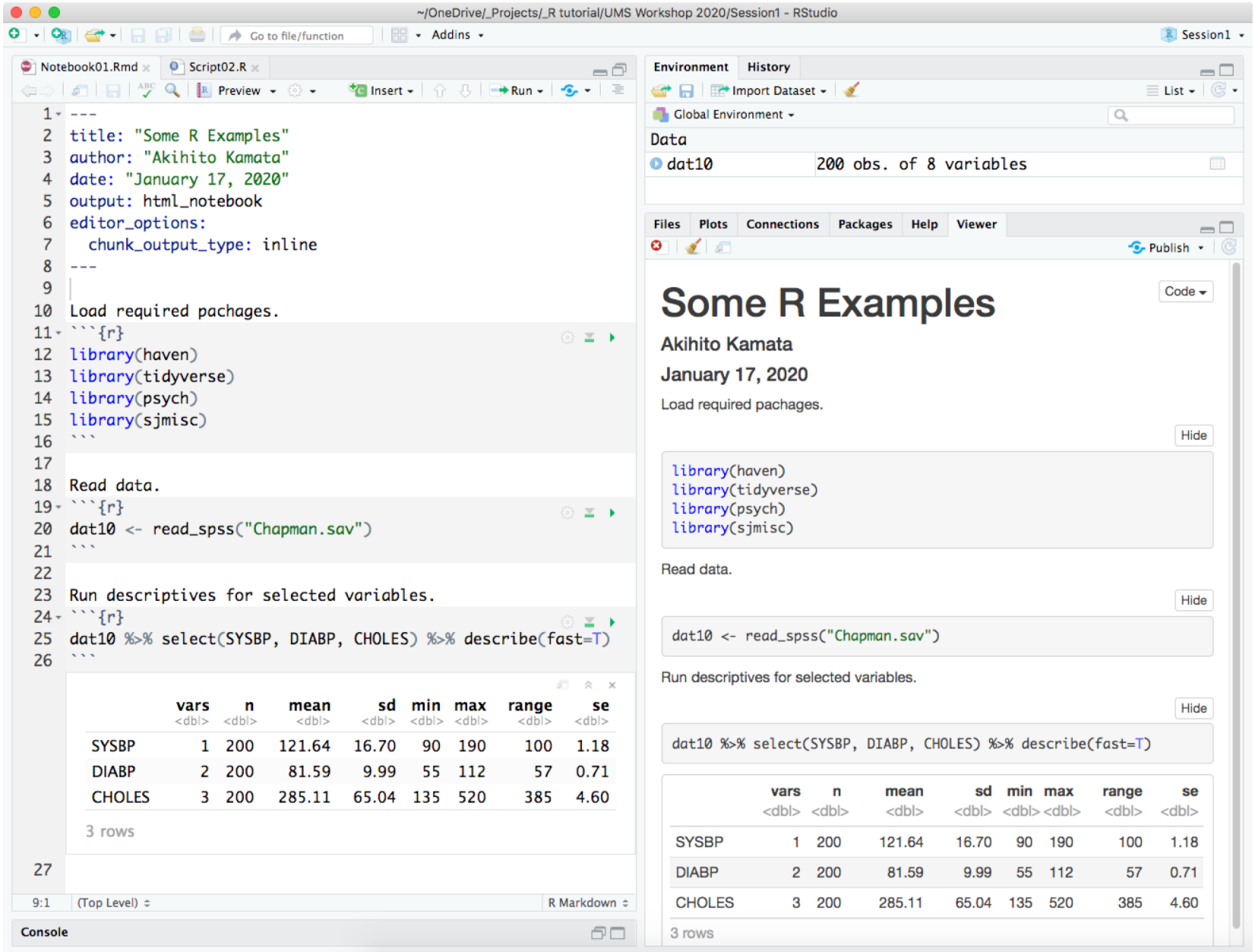
There is an option to show output only to console. However, not so much recommended.

Edit R Notebook File

5. Open or refresh the HTML file.



Preview in Viewer Pane



Aside: Write Equations on R Notebook File

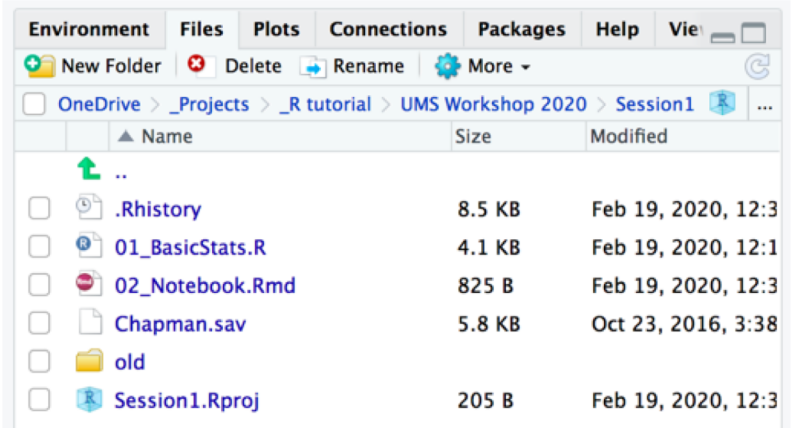
- We can insert symbols and equations on R Markdown files.
  - Use LaTeX way of writing an equation by tagging a dollar sign ($) at the beginning and the end of the equation.

    (e.g.)
    - $a=150$ will become $a = 150$.
    - $\alpha=150$ will become $\alpha = 150$.
    - $\bar{X}=150$ will become $\bar{X} = 150$.
  - There are a number of online equation editor to generate LaTeX equation code.

    (e.g.)
    - https://latex.codecogs.com/eqneditor/editor.php
    - http://www.sciweavers.org/free-online-latex-equation-editor

Let's Do It

- Some examples are provided in `02_Notebook.Rmd` file.
- Open the file from the "Files" tab within RStudio.



- Try to edit & add some analyses and comments on the file.
- Click on "Preview" to see the changes.

Part III
Structural Equation Modeling on R

SEM on R

- We will use `lavaan` package.
  - Path analysis model for mediation analysis.
  - CFA and Measurement invariance tests.
- Let’s look at:
  - `03_SEM.Rmd`

Part IV
Some Iterations on R

Iterations on R

- Let’s look at:
    - 04_Iterations.Rmd