SMU
SINGAPORE MANAGEMENT UNIVERSITY

Lab 1: Introduction to R

ISSS 616 Applied Statistical Analysis with R (ASAR)

SMU School of Computing and Information Systems

Masters of IT in Business AY 2022-2023 – Term 1

Agenda

1. Background on R and RStudio
2. Setting up R
3. Setting up RStudio
4. Package Installation
5. Basic commands
6. Self-learning resources

1. BACKGROUND ON R

1. Background on R and RStudio

- R is a language and environment for statistical computing and graphics.

- Modern implementation of the S language. There are some important differences, but much code written for S runs unaltered under R.
  - S was developed by John Chambers and colleagues at Bell Labs in 1976. The aim of the language is to “turn ideas into software, quickly and faithfully”.

- R was created in New Zealand by Ross Ihaka and Robert Gentlemen. In 2000, R version 1.0.0 was released.

- R Core Group and R foundation controls and maintains the source code for R. The R Environment is primarily written in C, Fortran and R.

1. Background on R and R Studio

- R is widely used for data analysis and is growing in popularity.

- Some reasons for its popularity are:
  - Runs on almost any standard computing platform/ OS
  - Actively developed (frequent annual releases and bug fix releases)
  - Lean software, functionality divided into modular packages
  - Sophisticated graphic capabilities compared to other statistics software
  - Useful for interactive work, and also a powerful programming language for developing new tools

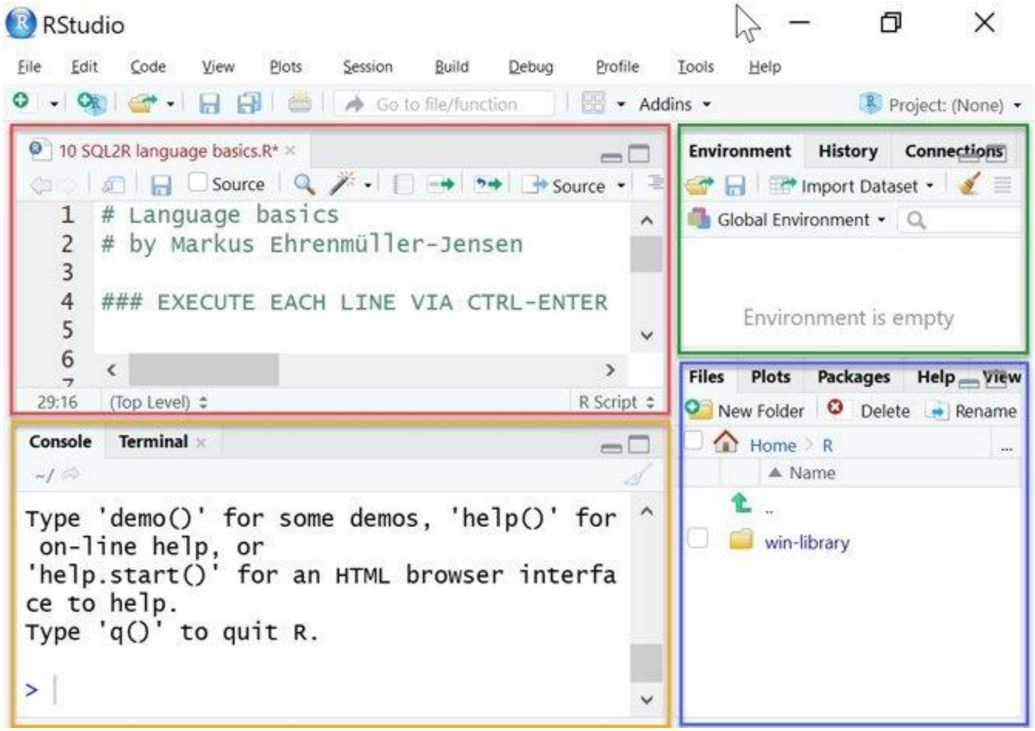
1. Background on R and R Studio

- RStudio is an integrated development environment (IDE) for R. It runs the R program in the background, and provides an interface with lots of nice features that makes coding in R easy and fun. It is also actively developed, with frequent updates.

- There are several customizable panels which are frequently used:

Source: where R scripts are edited (hidden if no scripts opened)

Console: where commands are typed and output is returned



```
# Language basics
# by Markus Ehrenmüller-Jensen

### EXECUTE EACH LINE VIA CTRL-ENTER
```

```
Type 'demo()' for some demos, 'help()' for
 on-line help, or
'help.start()' for an HTML browser interfa
ce to help.
Type 'q()' to quit R.

>
```

- Environment: list objects that exist in the current environment
- History: log commands issued

- Plots: where graphs are displayed
- Help: display R documentation

1. Background on R and R Studio

- Scripts help keep track of the complex analysis process. A script is a file that contains a sequence of commands to be executed.

- There are many advantages of coding scripts rather than from the console:
    - Reduce typing
    - Reduce errors
    - Describe code via “comments”
    - Use of white spaces to improve readability

- It is recommended to script all commands that are part of the analysis.

- R scripts are saved with the extension .R.

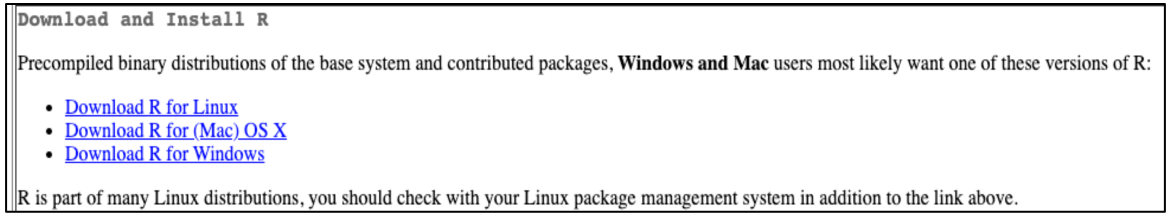
2. Setting up R

2. Installing R CRAN

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R.

1. Direct to : https://cran.r-project.org/
2. Select “Download R for your OS
3. Select “base” for windows
4. Select Download R 4.2.1 for Windows or R-4.2.1.pkg for Mac
5. Right click on downloaded file to open context menu and run .exe or .pkg file with Admin rights.
6. Install R in default setting

Download Page



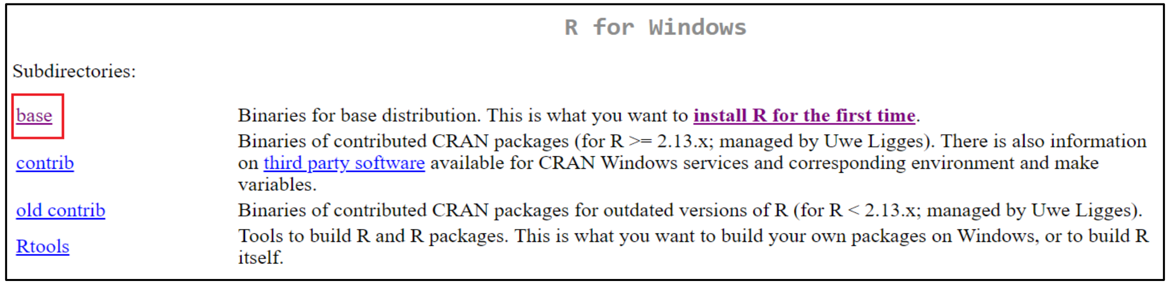
Download and Install R

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.

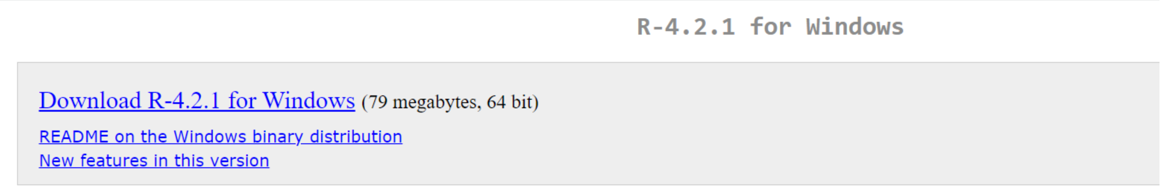
For Windows



R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.



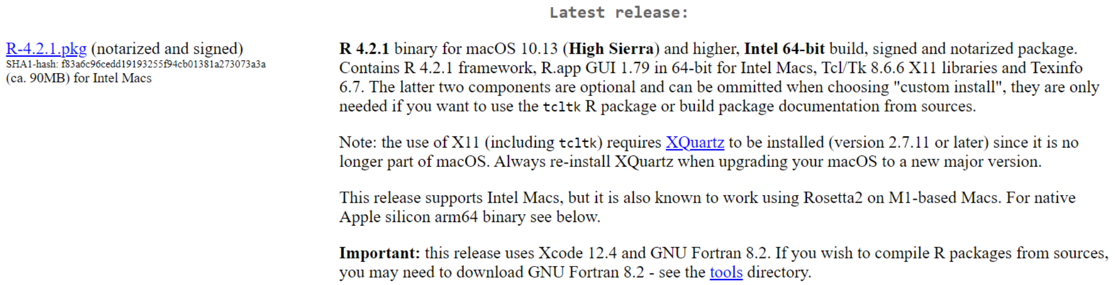
R-4.2.1 for Windows

Download R-4.2.1 for Windows (79 megabytes, 64 bit)

README on the Windows binary distribution

New features in this version

For MacOS



Latest release:

R-4.2.1.pkg (notarized and signed)
SHA1-hash: f83a6c96cedd19193255f94cb01381a273073a3a
(ca. 90MB) for Intel Macs

R 4.2.1 binary for macOS 10.13 (**High Sierra**) and higher, **Intel 64-bit** build, signed and notarized package. Contains R 4.2.1 framework, R.app GUI 1.79 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 6.7. The latter two components are optional and can be ommitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires XQuartz to be installed (version 2.7.11 or later) since it is no longer part of macOS. Always re-install XQuartz when upgrading your macOS to a new major version.

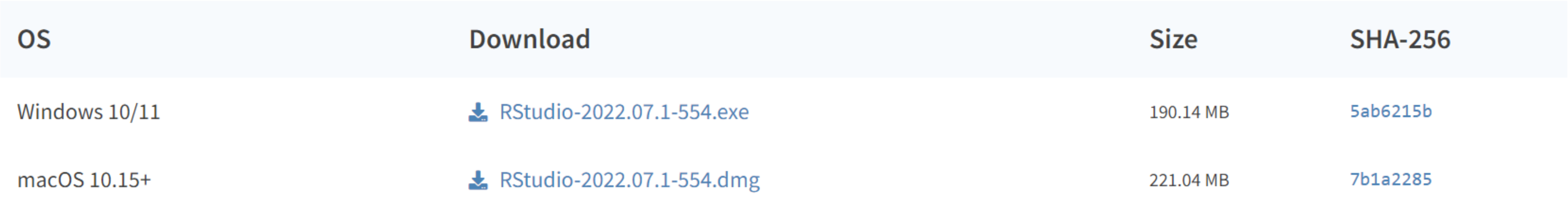
This release supports Intel Macs, but it is also known to work using Rosetta2 on M1-based Macs. For native Apple silicon arm64 binary see below.

Important: this release uses Xcode 12.4 and GNU Fortran 8.2. If you wish to compile R packages from sources, you may need to download GNU Fortran 8.2 - see the tools directory.

3. Setting up RStudio

2. Installing R Studio

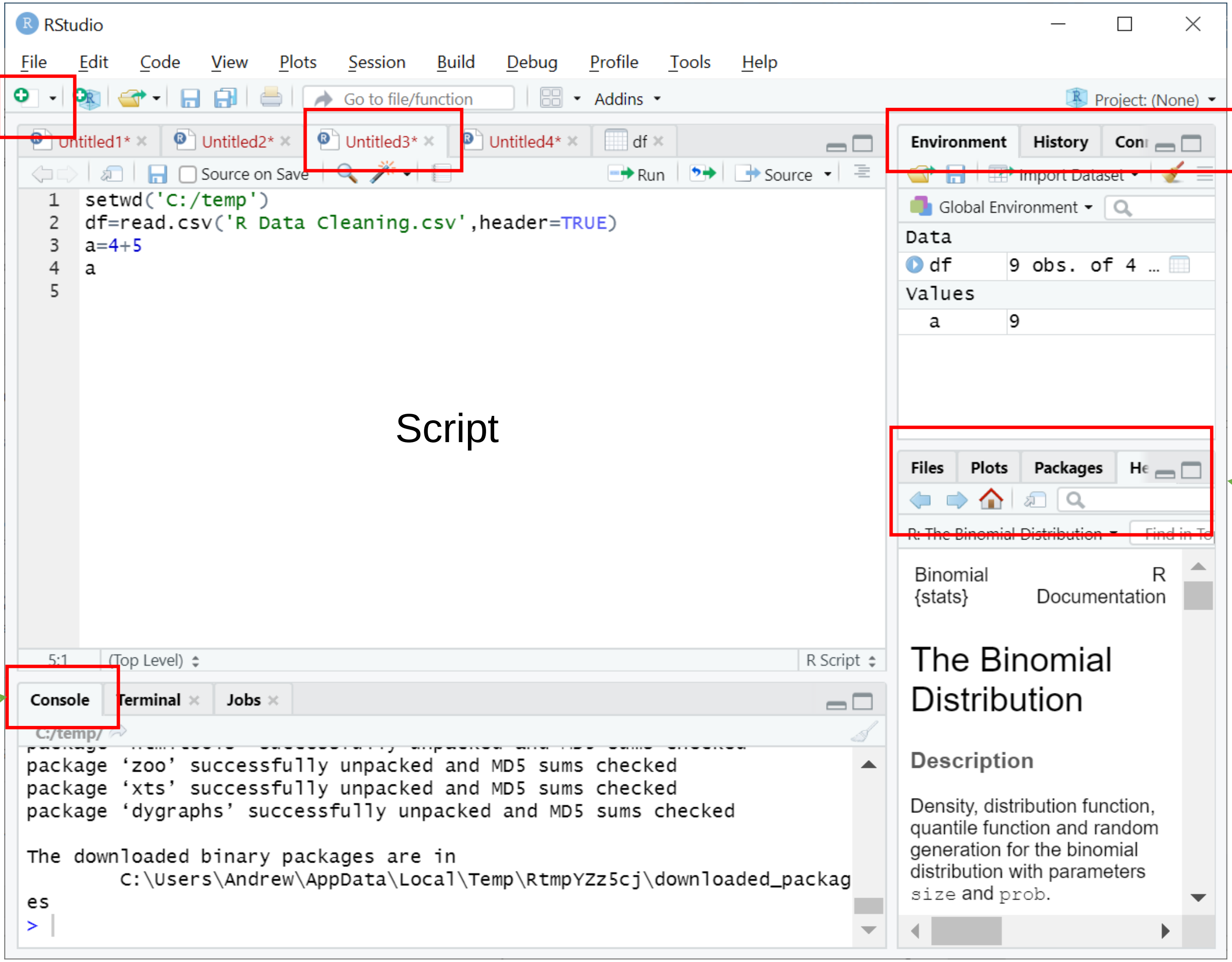
1. Direct to https://rstudio.com/products/rstudio/download/#download
2. Select file according to your OS
3. Right click on downloaded file to open context menu and run .exe or .dmg file with Admin rights.
4. Install with default setting
5. Open R Studio



OS	Download	Size	SHA-256
Windows 10/11	RStudio-2022.07.1-554.exe	190.14 MB	5ab6215b
macOS 10.15+	RStudio-2022.07.1-554.dmg	221.04 MB	7b1a2285

3. RStudio Layout and Interface

New Script



```
setwd('C:/temp')
df=read.csv('R Data Cleaning.csv',header=TRUE)
a=4+5
a
```

```
package ‘zoo’ successfully unpacked and MD5 sums checked
package ‘xts’ successfully unpacked and MD5 sums checked
package ‘dygraphs’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Andrew\AppData\Local\Temp\RtmpYZz5cj\downloaded_packag
es
>
```

Data

df 9 obs. of 4 …

Values

a 9

Binomial {stats} R Documentation

The Binomial Distribution

Description

Density, distribution function, quantile function and random generation for the binomial distribution with parameters `size` and `prob`.

Inspect environment variables and history, shows all values and arrays stored

View files, plots, Update Packages Help: packages documentation

R Console:
Runs code. Can use this for code that you do not want in your script, such as installing packages

4. Package Installation

4. Package Installation

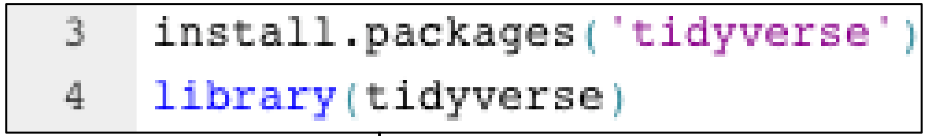
- Packages extend R’s functionality
- Some packages are pre-loaded at startup, e.g. stats, graphics. Included packages can be found in the bottom right panel, under the 'Packages' tab. Others must be loaded in every new R session using library (<package_name>)
- Most packages are not included with R, but can be found on CRAN (Comprehensive R Archive Network).
- Packages from CRAN can be installed using install.packages (“<package_name>”).
- To check for available packages updates, use old.packages ( ).
- To update all packages, use update.packages (ask = FALSE).
- Remember that a package only needs to be installed once but has to be loaded for each new R session (if you need to use it).

4. Package Installation

Suggested packages to install:

- dplyr: essential for data wrangling, provides a consistent set of verbs to manipulate data
- ggplot2: maps variables into aesthetics, declaratively creating graphics
- manipulate: interactive plotting functions for use within RStudio.
- rmarkdown: turns analyses into high quality documents, reports, presentations and dashboards.
- knitr: general-purpose literate programming engine, with lightweight API's designed to give users full control of the output without heavy coding work.

Basic Packages (How to install)

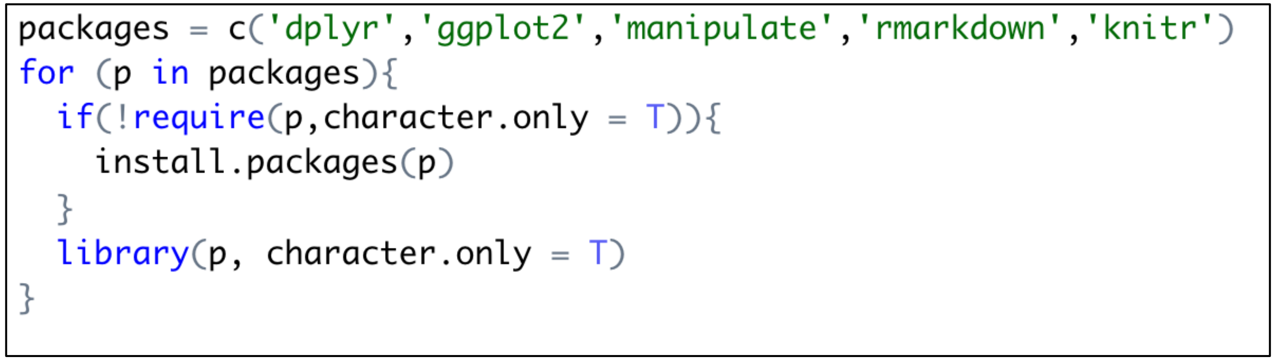


```
install.packages('tidyverse')
library(tidyverse)
```

This line helps to install packages from CRAN, therefore requiring internet access. Previously installed packages will not need internet access. Remember inverted commas.

After installing, remember to load using library(). Does not need inverted commas or internet access.

Using the code right can help you install and load multiple packages at once.



```
packages = c('dplyr','ggplot2','manipulate','rmarkdown','knitr')
for (p in packages){
  if(!require(p,character.only = T)){
    install.packages(p)
  }
  library(p, character.only = T)
}
```

4. Package Installation

Further assistance:

- If in doubt, please refer to the R documentation. It provides useful information like:
  - Description of what the function does
  - Arguments/options available
  - Examples and/or references

- Alternatively, you can google for precedent examples or check forums like **Stack Overflow**.

- Online documentation resources:
- https://www.rdocumentation.org/
- https://community.rstudio.com/
- https://shiny.rstudio.com/

5. Basic Commands

- ?function : to look up function explanation and syntax (e.g. ?summary)

Mac:

- Command + Enter : Run 1 line
- Command + Shift+Return: Run entire script
- Option + -(minus) : <- (Assignment operator in R)
- Command + Shift + C : Multiple lines comments
- Command + Z: Undo

Windows:

- Ctrl/Alt+ Enter : Run 1 line
- Ctrl + Shift + Enter : Run entire script
- Alt + -(minus) : <- (Assignment operator in R)
- Ctrl + Shift + C : Multiple lines comments
- Ctrl + Z: Undo

More on: RStudio Cheatsheets - RStudio
Keyboard Shortcuts in the RStudio IDE – RStudio Support

6. Basic Resources (Textbooks and Cheatsheets)

1. **R for Data Science** (https://r4ds.had.co.nz/)

- Written by Hadley Wickham, developer of Tidyverse (ggplot2, plyr, dplyr, and reshape2)
- Good for beginners

2. **Solutions** for questions in the book above for practice (https://lokhc.wordpress.com/r-for-data-science-solutions/)

3. **Other books** (http://ucanalytics.com/blogs/learn-r-12-books-and-online-resources/)

4. **Cheatsheets Compiled** (https://rstudio.com/wp-content/uploads/2019/01/Cheatsheets_2019.pdf)

- Recommend to start with tidyr, ggplot2, dplyr

References

Books:

1. An introduction to R by CRAN Project

2. Introduction to R by Monash Bioinformatics Platform

Last words…

All the best for your R Journey! Feel free to contact us if you have any question:

Ranice: ranice.tan.2021@mitb.smu.edu.sg
Raveena: raveenac.2021@mitb.smu.edu.sg
Tasa: tasapornv.2021@mitb.smu.edu.sg