Rough schedule for week #1 and week #2

(from 20200921 to 20200928)

**Base-line:** R for Data Science, ggplot2 by Hadley Wickham, Rattle, and Ledolter’s chapter.

Unit ONE:

1. Install [R](https://www.r-project.org/) , RStudio, and package {Rattle}. Visit [R Home](https://www.r-project.org/) to get free stuff.

2. Browse R packages with html format from R console manu

Play with R a little for fun using the R code as follows

## R code for checking installed packages currently

**getwd**()

packinfo <- **installed.packages**(fields = c("Package", "Version"))

**str**(packinfo)

packinfo[,**c**("Package", "Version")]

**ls**()

**save**(packinfo, file="check.rda")

#unlink("check.rda")

**rm**(list=ls(all=TRUE))

**ls**()

#################

**load**("check.rda")

**ls**()

**View**(packinfo)

**#################**

installed.packages()

View(installed.packages())

a <- installed.packages();str(a)

a[,c(1,4)]

\* Online R code execution

<https://rdrr.io/snippets/>

<https://www.tutorialspoint.com/execute_r_online.php> (能跑中文)

<https://onecompiler.com/r> (能跑中文)

3. Get to know installed “data” from packages (for your practices)

data {utils}

data()

data(iris)

4. Overview .rda, .Rdata, and .rds. (lectured in cluss)

5. Getting technical help: [Stack Overflow](http://stackoverflow.com/) , [RPubs](https://rpubs.com/) & [R-Bloggers](https://www.r-bloggers.com/) (Git and GitHub)

6. ggplot2, Rattle etc. 圖館均有電子書 (亦上傳至[本課程之教學平台](https://elearning.nkust.edu.tw/mooc/index.php))

7. Self-learning: For example, to learn “[Changing a Numeric Variable to Categorical Variable in R (R Tutorial 5.4)](https://www.youtube.com/watch?v=EWs1Ordh8nI)”, you can get the idea from YouTube yourself.

8. Get to know R code roughly by executing R code.

**[For fun: see below]**

###############################################################

**str(iris)**

**plot(iris$Sepal.Length~iris$Sepal.Width,pch=21, cex=3, bg=c("red","green3","blue")[unclass(iris$Species)])**

**plot(iris[,4],iris[,3], pch=21, cex=3, bg=c("red","green3","blue")[unclass(iris$Species)])**

**plot(iris$Petal.Length~iris$Petal.Width,pch=21, cex=3, bg=c("red","green3","blue")[unclass(iris$Species)])**

**rm(list=ls())**

**par(mfrow=c(1,2))**

**plot(iris[,4],iris[,3])**

**text(iris[,4],iris[,3],labels=iris[,3], col="red")**

**plot(iris[,4],iris[,3])**

**text(iris[,4],iris[,3],labels=iris$Species, col=c("red","green3","blue"))**

**table(iris[,5])**

**table(iris[,5], iris[,3])**

9. Meet “BIG” data first hand (details in「大」數據投影片範例.ppt)

Data source: [臺北捷運每日分時各站OD流量統計資料](http://163.29.157.32:8080/dataset/98d67c29-464a-4003-9f78-b1cbb89bff59)

Maybe you need a professional editor named [EmEditor](https://zh-tw.emeditor.com/download/) (20.1.1) to help you opening the “big” file.

a. unzip 「臺北捷運每日分時各站OD流量統計資料\_202007.zip」

b. fix 「臺北捷運每日分時各站OD流量統計資料\_202007.csv」with Windows notebook (記事本) 🡺 time consuming

c. or fix 「臺北捷運每日分時各站OD流量統計資料\_202007.csv」with EmEditor 🡺 time saving

d. use read.csv(), rio::import(), save(.rda), load(.rda) to practice import-export data file. (will be demo in class)

e. sum 「df07$人次」and see what we get.

10. Getting started with rattle(), focusing on Chapter 2 (install packages AUTOMATICALLY)

11 如何裝好 R rattle: (if necessary)

install.packages("<https://cran.r-project.org/bin/windows/contrib/3.3/RGtk2_2.20.31.zip>", repos=NULL)  
install.packages('rattle',repos="<http://cran.csie.ntu.edu.tw/>")  
library(rattle)  
library(RGtk2)  
rattle()  
library(stringr)

12. Surfing “data.gov.tw” for your final project

13. Data Manipulation

Electronic resource: [R for Data Science](https://r4ds.had.co.nz/) (Web reading), [hadley/r4ds](https://github.com/hadley/r4ds) (GitHub)

And read the following

1. Introduction to dplyr.pdf (資料源如下R碼)

* dplyr in html help, find *vignettes and other documentation*.
* R for Data Science, Chapter 5 Data transformation
* transform.Rmd (in [r4ds-master at GitHub](https://github.com/hadley/r4ds))
* Data Manipulation的資料源{nycflights13} #partial from: lecture01.R

install.packages(c("RCurl","imager"))

data(package='dplyr')

data(package='nycflights13')

library(RCurl)

myurl <- "http://r4ds.had.co.nz/diagrams/relational-nycflights.png"

download.file(myurl,'test.png', mode = 'wb')

library(imager)

im <- load.image('test.png')

plot(im)

另：<https://github.com/hadley/nycflights13/blob/master/R/flights.R>

<https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236>

b. dplyr and pipes the basics.pdf

c. When to fly to get there on time.pdf

d. dplyr-tutorial.pdf

14. Get to know “r4ds” and learn how to drive R markdown text format (.rdm). To fully utilize .rdm effectively, we need to execute .rdm using RStudio.

* Try to execute “corrplot-intro.rdm” uploaded to get the idea (amazing!).
* Install “[R for Data Science](https://github.com/hadley/r4ds)”: devtools::install\_github("hadley/r4ds")

15. Let’s see what else left? Q and A!

**I will append the related text anytime whenever needed! Thus, check the Web regularly.**

**練習讀取「大」檔案:以800M的資料檔為例**

rm(list=ls(all=TRUE))

getwd()

memory.limit()

memory.size()

ptm <- proc.time()#real and CPU time (in seconds) the currently running R process has already taken

# 臺北捷運每日分時各站OD流量統計資料\_202007.csv需為ANSI格式，不要用"具有BOM的UTF-8"格式

df07 <- read.csv("臺北捷運每日分時各站OD流量統計資料\_202007.csv", sep="")

proc.time() - ptm

str(df07)

退出R，再從新開R

rm(list=ls(all=TRUE))

memory.limit()

memory.size()

library(rio)

#install\_formats()

#install.packages('rio')

ptm <- proc.time()

df07 <- import("臺北捷運每日分時各站OD流量統計資料\_202007.csv")

proc.time() - ptm

str(df07)

head(df07)

memory.size()

ls()

**所以大檔案(.csv)就用盡量用import{rio}來開**

rm(list=ls(all=TRUE))

memory.limit()

memory.size()

library(rio)

#install\_formats()

#install.packages('rio')

df07 <- import("臺北捷運每日分時各站OD流量統計資料\_202007.csv")

str(df07)

head(df07)

save(df07, file="臺北捷運每日分時各站OD流量統計資料\_202007.Rda")

# re-load data file (.Rda) again.

rm(list=ls(all=TRUE))

ls()

load("臺北捷運每日分時各站OD流量統計資料\_202007.Rda")

memory.size()

ls()

str(df07)

head(df07)

sum(df07$人次)

Data Manipulation的資料源{nycflights13} #from: lecture01.R

install.packages(c("RCurl","imager"))

data(package='dplyr')

data(package='nycflights13')

library(RCurl)

myurl <- "http://r4ds.had.co.nz/diagrams/relational-nycflights.png"

download.file(myurl,'test.png', mode = 'wb')

library(imager)

im <- load.image('test.png')

plot(im)