\*\*video1.13, 1.14與老師上課內容重複，因此記錄在本份文件的上課筆記中

>x=11/y<-7是一樣的 define value. Equal

>xx <- “marin” define object with character

>print(x) print x

>rm(y) remove y

>x+y/x-y….abs正負相反 sqrt平方根 calculate

Keyboard-arrow up-move to the end

To label paragraph in rows of orders, use --The code below is for…

> x <- c(1, 3, 5, 7. 9) 定義集合

> x [3] 得5

> x [1:3] 得 135

> x[-c(1, 5)] 得 3, 5, 7

> matrix (c(1, 2, 3, 4, 5, 6, 7, 8, 9), nrow=3, byrow=TRUE)

得 1 2 3

4 5 6

7 8 9

> mat<-matrix

> mat[ ,2] 得 4 5 6

> seq(from=1, to=7 by=1/3)

> rep(1, times=10)/ rep(“marin”, times=10)/ rep(c(“m”, “f”), times=5)

# if two vectors are of the same length, we may add/subtract/mult/div corresponding elements

Video 5

.csv/.txt

\*to import

> help(read.csv)

> ?read.csv

>

> data1 <- read.csv(file.choose(), header=T)

>

> data1

> data2 <- read.table(file.choose(), header=T, sep=”,”)

Video 6

\*Directly import excel (xlsx)

= Max Rows

= Skip (skip rows counted from top)

= N/A (type-in your mark sign in excel)

= copy code to import data in script

Video 7

\*to export

**write.table**

>write.table(data to be exported from R, file=”new file name.csv”, sep=”,”)

>write.table(data to be exported from R, file=”new file name.csv”, row.names=F, sep=”,”)

write.csv: do not need sep

Video 8

>help(read.table)

>?read.table

> Data1 <- read.table(file=”file location”, header=TRUE, sep=”mark for data separation”)

or

> Data2 <- read.table(file.choose(), …)

Or

Import Dataset-set “header”, “sep”, “decimal”…

> rm(Data1)

> dim(Data2)

=show the number of rows and columns of the data

> head(Data2)

> tail(Data2)

> Data2[5:9, ]

=show row 5 to row 9 of Data2

…

Video 9

Attach, Detach, and Summary with variables

> names(Data2)

=show column titles of Data2

> mean (Data2$Age)

= calculate the mean of data under column “Age”

= $ extract

> attach(Data2)

>

> mean(Data$Age)

>

> Age

=show data

> detach(Data2)

>

> Age

=error

> level(Smoke)

> summary(Data2)

=show statistic summary of Data2, including Min, 1st Q, Median, Mean…

> x <- c(0,1,1,1,0,0,0,0,0)

>

> class(x)

[1] “numeric”

> summary(x)

>

> x <- as.factor(x)

>

> class(x)

[1] “factor”

>

> summary(x)

0 1

7 3

Video 10

Subsetting Data with []

> dim(Data2)

>

> length(Age)

=numbers of data under column “Age”

> mean (Age[Gender==”female”])

= “=”: to assign values to objects

= “==” to represent the meaning of equality mathematically

> FemData <- Data2[Gender==”female”,]

>

> dim(FemData)

=for check

>

> FemOver15 <- Data2[Gender==”female” & Age.15, ]

1015

[[1]]=1st item

(條件)成立時，執行{指令}

NA=not available

> yy<- matrix(NA, nrow=6, ncol=2)

> yy<- matrix(5:16, nrow=6, ncol=2)

wd=working directory

paste( , )=to connect strings before and after “,”

paste0( , )=to connect strings with no space

getwd()

save(list=c("dd","ee"), file=paste0(getwd(),"/output-2021-10-15.RData"))

load(paste0(getwd(),"/output-2021-10-15.RData"))

Video 11

Logic and other commands

> Age[1:5]

Show first 5 items

> temp <- Age>15

> temp[1:5]

FALSE TRUE TRUE FALSE FALSE (depends on item values)

> temp2 <- as.numeric(Age>15)

0 1 1 0 0

> FemSmoke <- Gender==”female” & Smoke==”yes”

>

> rm(list=ls())

=to clean all object in Workspace list

Video 12

Setting up Working directories

=to save works

> getwd()

=to Show the location of where the source comes from

> setwd(“location”)

=to define the location of the saved file

=go back to note 1015

>

> getwd()

>  
> load(“FimeName.Rdata”)

= to load workspace of previous file into R

>

> save.image(“FileName.Rdata”)

=Menu-Session-Save Workspace As

>

> q()

=to quit R

1022

library(stringr)

kk <- "abcdefgheijk"

pp <- str\_split(kk, "e")

pp <- str\_split(kk, “e|j”)

=to split strings

=str\_split(StringName, “the item that invokes split”)

=”|”=or

kk <- "John Wang"

yy <- str\_split(kk, " ")

zz <- paste0("Mary ", yy[[1]][2])

=paste0(“FirstItem”, “SecondItem”)

=”0”=to connect two item with NO space

library(pdftools)

pdfText <- pdf\_text(paste0(getwd(),"/05160218039.004.pdf"))

text1 <- pdfText[1]

pos1 <- str\_locate(text1, "發文字號")

>pos1

start end

[1,] 44 47

=”發” is the 44th character in text1, and “號” is the 47th character in text1.

=every row occupies 2 item position.

wantedString <- str\_sub(text1, (pos1[1,2]+2), (pos3[1,1]-3))

file.rename("05160218039.004.pdf", paste0(wantedString,".pdf"))

True

=to rename file name

1029

Sys.setlocale(locale="English\_United States")

Sys.getlocale()

>library(readxl)

>read\_xlsx(“FileName.xlsx”)

>any(aa, na.rm=T)

=to check if there is at least one “aa” in all items,

=if yes, show “True”

>all(aa, na.rm=T)

=to check if there is “aa” in all items,

=if yes, show “True”

for(j in 1763:1772){

print(getSubstring(j))

}