Video 1：

Difference between R and R studio

R studio: development environment(not necessary，但很方便)

R: coding language

X <- 1:5

Y <- 6:10

ls()會出現”x”,”y” 代表R現在有紀錄著X跟Y

> z<- 11:15

> x <- 1:5

> y <- 6:10

> sum(x,y,z)

[1] 120

Video 2：

Install(已經裝好了)

Video 3：

x = 11

print(x)

y<-7

print(y)

y <- 9

print(y)

ls()

rm(y)

ls()

y=9

print(y)

> z<- 11:15

> x <- 1:5

> y <- 6:10

> sum(x,y,z)

[1] 120

> (1+15)15/2

Error: unexpected numeric constant in "(1+15)15"

> (1+15)\*15/2

[1] 120

>

> x = 11

> print(x)

[1] 11

> x = 11

> print(x)

[1] 11

> y<-7

> print(y)

[1] 7

> y <- 9

> print(y)

[1] 9

> y <- 9

> print(y)

[1] 9

> print("你好")

[1] "你好"

> print(100)

[1] 100

> ls()

[1] "bone" "cc"

[3] "dd" "ee"

[5] "elementOne" "ff"

[7] "henry" "henry2"

[9] "i" "j"

[11] "k" "lengNew"

[13] "lengSequ" "newObject"

[15] "Newsequ" "ray"

[17] "raymond" "sequ"

[19] "x" "y"

[21] "yy" "z"

> rm(y)

> ls()

[1] "bone" "cc"

[3] "dd" "ee"

[5] "elementOne" "ff"

[7] "henry" "henry2"

[9] "i" "j"

[11] "k" "lengNew"

[13] "lengSequ" "newObject"

[15] "Newsequ" "ray"

[17] "raymond" "sequ"

[19] "x" "yy"

[21] "z"

> y=9

> print(y)

[1] 9

> x.1=14

> x.1

[1] 14

> "1x"=122

> 1x

Error: unexpected symbol in "1x"

> 1x=122

Error: unexpected symbol in "1x"

> xx="mario"

> xx

[1] "mario"

> x

[1] 11

> y

[1] 9

> x+y

[1] 20

> z=x+y

> z

[1] 20

> z^2

[1] 400

> z^2/(y+1)

[1] 40

> sqrt(y)

[1] 3

> sqrt(y)

[1] 3

> #the code below is for...

Video 4：

> x=11

> x

[1] 11

> x1=c(1,3,5,7,9)

> x1

[1] 1 3 5 7 9

> gender=c('male','female')

> gender

[1] "male" "female"

> 2:7

[1] 2 3 4 5 6 7

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

[1] 1 2 3 4 5 6 7

> seq(from=1, to=7,by=2)

[1] 1 3 5 7

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

[1] 1 4 7

> seq(from=1, to=7,by=4)

[1] 1 5

> seq(from=1, to=7,by=0.5)

[1] 1.0 1.5 2.0 2.5 3.0 3.5

[7] 4.0 4.5 5.0 5.5 6.0 6.5

[13] 7.0

> rep(1, times=10)

[1] 1 1 1 1 1 1 1 1 1 1

> rep('cool', times=5)

[1] "cool" "cool" "cool"

[4] "cool" "cool"

> rep(1:3, times=5)

[1] 1 2 3 1 2 3 1 2 3 1 2 3 1

[14] 2 3

> rep(seq(from=1,to=5,by=2), times=5)

[1] 1 3 5 1 3 5 1 3 5 1 3 5 1

[14] 3 5

> rep(c('k','e','v','i','n'), times=2)

[1] "k" "e" "v" "i" "n" "k"

[7] "e" "v" "i" "n"

> x=1:5

> x

[1] 1 2 3 4 5

> y=seq(from=1,to=9,by=2)

> y

[1] 1 3 5 7 9

> x+y

[1] 2 5 8 11 14

> x-10

[1] -9 -8 -7 -6 -5

> x+1

[1] 2 3 4 5 6

> #相同長度的向量可以加減乘除等等的

> y[3]

[1] 5

> y[-3]

[1] 1 3 7 9

> y

[1] 1 3 5 7 9

> y[1,5]

Error in y[1, 5] : incorrect number of dimensions

> y[1:3]

[1] 1 3 5

> y[c(1,5)]

[1] 1 9

> y[-c(1,5)]

[1] 3 5 7

y[y<6]

[1] 1 3 5

matrix(c(1:9),nrow=3,byrow=TRUE)

[,1] [,2] [,3]

[1,] 1 2 3

[2,] 4 5 6

[3,] 7 8 9

> matrix(c(1:9),nrow=5,byrow=TRUE)

[,1] [,2]

[1,] 1 2

[2,] 3 4

[3,] 5 6

[4,] 7 8

[5,] 9 1

Warning message:

In matrix(c(1:9), nrow = 5, byrow = TRUE) :

data length [9] is not a sub-multiple or multiple of the number of rows [5]

> matrix(c(1:10),nrow=5,byrow=TRUE)

[,1] [,2]

[1,] 1 2

[2,] 3 4

[3,] 5 6

[4,] 7 8

[5,] 9 10

> matrix(c(1:10),nrow=2,byrow=TRUE)

[,1] [,2] [,3] [,4] [,5]

[1,] 1 2 3 4 5

[2,] 6 7 8 9 10

> matrix(c(1:9),nrow=3,byrow=FALSE)

[,1] [,2] [,3]

[1,] 1 4 7

[2,] 2 5 8

[3,] 3 6 9

> made=matrix(c(1:9),nrow=3,byrow=TRUE)

> made[1,2]

[1] 2

> made

[,1] [,2] [,3]

[1,] 1 2 3

[2,] 4 5 6

[3,] 7 8 9

> made[c(1,3),2]

[1] 2 8

> made[2,]

[1] 4 5 6

> made[,2]

[1] 2 5 8

> made\*10

[,1] [,2] [,3]

[1,] 10 20 30

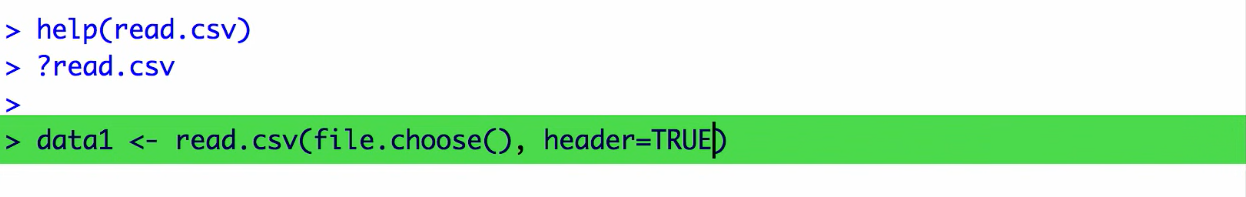
[2,] 40 50 60

[3,] 70 80 90

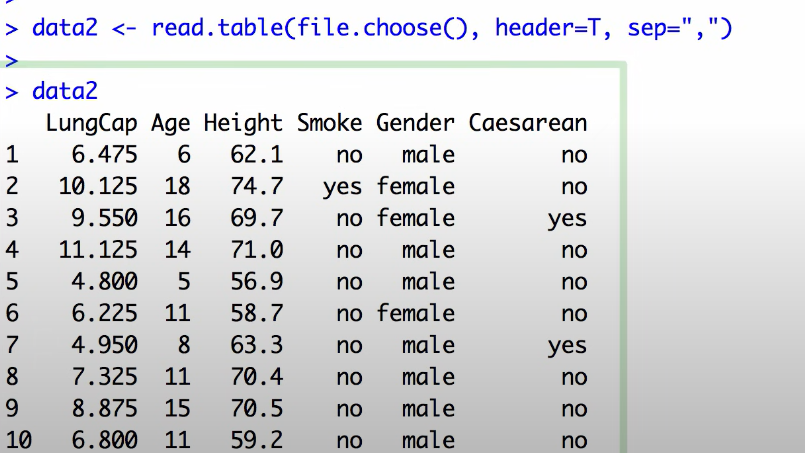
Video 5：(import from excel)

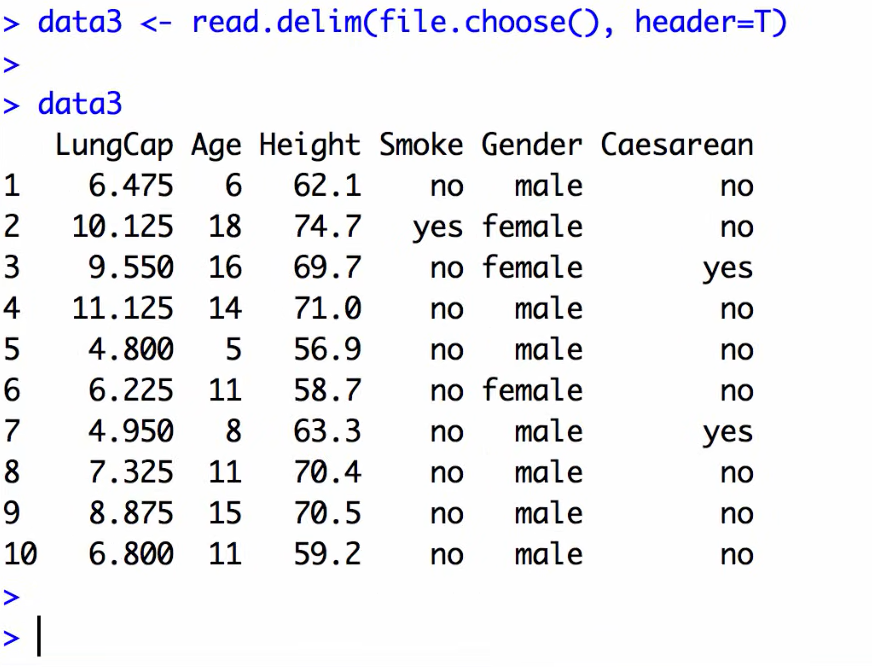
先把excel檔另存新檔成csv

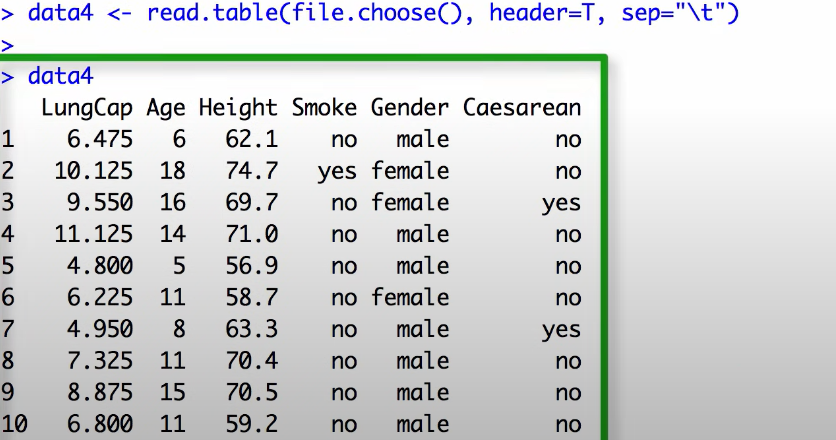
Rstudio打help(read.csv)



Header true代表第一列是標題







tmp <- data.frame(Student\_ID=c(1,2,3,4,5),

name=c("Helen", "Lun", "Leon", "Kevin", "Tommy"),

score=c(80,36, 88.9, 97.5, 60))

tmp # data frame的型態

circle\_calculator <- function(r, is\_area = TRUE) {

# R 語言有內建圓周率 pi

area <- pi \* r\*\*2

perimeter <- 2 \* pi \* r

if (is\_area) {

return(area)

} else {

return(perimeter)

}

}

circle\_calculator(3) # 預設回傳圓形面積

circle\_calculator(3, is\_area = FALSE) # 回傳圓形周長

bmi\_calculator <- function(height, weight) {

height <- height/100

bmi <- weight / height\*\*2

if (bmi < 18.5) {

bmi\_label <- "過輕"

} else if (bmi > 30) {

bmi\_label <- "肥胖"

} else if (bmi >= 18.5 & bmi < 25) {

bmi\_label <- "正常"

} else {

bmi\_label <- "過重"

}

bmi\_list <- list(

bmi = bmi,

bmiLabel = bmi\_label

)

return(bmi\_list)

}

shaq <- bmi\_calculator(216, 147)

shaq$bmi

shaq$bmiLabel

## > # 雜亂無章的資料

## > messy\_data <- data.frame(c(1, 2, 3, 4, NA), c(1, 2, 3, NA, 5), c(1, 2, NA, 4, 5))

## > names(messy\_data) <- c("var\_1", "var\_2", "var\_3")

## > messy\_data

## var\_1 var\_2 var\_3

## 1 1 1 1

## 2 2 2 2

## 3 3 3 NA

## 4 4 NA 4

## 5 NA 5 5

# 自訂函數 clean\_data

clean\_data <- function(df, impute\_value){

n\_rows <- nrow(df)

na\_sum <- rep(NA, times = n\_rows)

for (i in 1:n\_rows) {

na\_sum[i] <- sum(is.na(df[i, ])) # 計算每個觀測值有幾個 NA

df[i, ][is.na(df[i, ])] <- impute\_value # 把 NA 用某個數值取代

}

complete\_cases <- df[as.logical(!na\_sum), ] # 把沒有出現 NA 的觀測值保留下來

imputed\_data <- df

df\_list <- list(

complete\_cases = complete\_cases,

imputed\_data = imputed\_data

)

return(df\_list)

}

# 自訂函數 clean\_data

clean\_data <- function(df, impute\_value){

n\_rows <- nrow(df)

na\_sum <- rep(NA, times = n\_rows)

for (i in 1:n\_rows) {

na\_sum[i] <- sum(is.na(df[i, ])) # 計算每個觀測值有幾個 NA

df[i, ][is.na(df[i, ])] <- impute\_value # 把 NA 用某個數值取代

}

complete\_cases <- df[as.logical(!na\_sum), ] # 把沒有出現 NA 的觀測值保留下來

imputed\_data <- df

df\_list <- list(

complete\_cases = complete\_cases,

imputed\_data = imputed\_data

)

return(df\_list)

}

# 雜亂無章的資料

messy\_data <- data.frame(c(1, 2, 3, 4, NA), c(1, 2, 3, NA, 5), c(1, 2, NA, 4, 5))

names(messy\_data) <- c("var\_1", "var\_2", "var\_3")

messy\_data

cleaned\_data <- clean\_data(messy\_data, impute\_value = 999)

cleaned\_data$complete\_cases # 保留完整觀測值的資料

cleaned\_data$imputed\_data # 取代 NA 為 999 的資料

## > # 自訂函數 clean\_data

## > clean\_data <- function(df, impute\_value){

## + n\_rows <- nrow(df)

## + na\_sum <- rep(NA, times = n\_rows)

## + for (i in 1:n\_rows) {

## + na\_sum[i] <- sum(is.na(df[i, ])) # 計算每個觀測值有幾個 NA

## + df[i, ][is.na(df[i, ])] <- impute\_value # 把 NA 用某個數值取代

## + }

## + complete\_cases <- df[as.logical(!na\_sum), ] # 把沒有出現 NA 的觀測值保留下來

## + imputed\_data <- df

## + df\_list <- list(

## + complete\_cases = complete\_cases,

## + imputed\_data = imputed\_data

## + )

## + return(df\_list)

## + }

## >

## > # 雜亂無章的資料

## > messy\_data <- data.frame(c(1, 2, 3, 4, NA), c(1, 2, 3, NA, 5), c(1, 2, NA, 4, 5))

## > names(messy\_data) <- c("var\_1", "var\_2", "var\_3")

## > messy\_data

## var\_1 var\_2 var\_3

## 1 1 1 1

## 2 2 2 2

## 3 3 3 NA

## 4 4 NA 4

## 5 NA 5 5

## > cleaned\_data <- clean\_data(messy\_data, impute\_value = 999)

## > cleaned\_data$complete\_cases # 保留完整觀測值的資料

## var\_1 var\_2 var\_3

## 1 1 1 1

## 2 2 2 2

## > cleaned\_data$imputed\_data # 取代 NA 為 999 的資料

## var\_1 var\_2 var\_3

## 1 1 1 1

## 2 2 2 2

## 3 3 3 999

## 4 4 999 4

## 5 999 5 5

物件的作用範圍

開始自行寫作函數之後我們必須要意識到物件的作用範圍（scope），當程式中沒有函數區塊（function block，這指的是自訂函數大括號所間隔出來的縮排區域），所有的物件作用範圍都被視為全域變數（global variables），在物件成功宣告後的任何一段程式碼中我們都能去使用它。

當程式中開始出現函數區塊後，物件作用範圍就被分為區域變數（local variables）與全域變數；在函數區塊內的程式碼我們能夠使用兩種被成功宣告的物件，但是在函數區塊以外的程式碼就僅能使用全域變數。用文字敘述還是相當抽象，我們應該動手自訂函數來釐清觀念。在以下程式範例中，bmi 這個數值向量的作用範圍是區域變數，不能夠在函數區塊以外使用。

## > get\_bmi <- function(height, weight) {

## + height <- height/100 # local numeric

## + bmi <- weight / height\*\*2 # local numeric

## + return(bmi)

## + }

## > shaq\_height <- 216 # globel numeric

## > shaq\_weight <- 147 # global numeric

## > shaq\_bmi <- get\_bmi(shaq\_height, shaq\_weight) # global numeric

## > shaq\_bmi # global numeric

## [1] 31.5072

## > bmi # local numeric

## Error: object 'bmi' not found

## > height <- 216 # global numeric

## > weight <- 147 # global numeric

## > get\_bmi <- function() {

## + height <- height/100 # local numeric

## + bmi <- weight / height\*\*2 # local numeric

## + return(bmi)

## + }

## > shaq\_bmi <- get\_bmi() # global numeric

## > shaq\_bmi # global numeric

## [1] 31.5072

## > height # global numeric

## [1] 216

print("COOL")

newObject <- 5

newObject <- 1

newObject <- 12

class(newObject)

newObject <- 'Hello'

class(newObject)

print(newObject)

newObject <- c(5,12,36)

newObject <- (-4):(-2)

newObject <- c('Hello',"World")

newObject <- c(34, "Hey")

class(newObject)

elementOne <- newObject[1]

class(elementOne)

elementOne <- as.numeric(elementOne)

class(elementOne)

lengNew <- length(newObject)

class(lengNew)

sequ <- -20000:200000

lengSequ<-length(sequ)

matrixOne <- matrix(1:100, nrow=10, ncol=10, byrow=FALSE)

matrixOne

?matrix # help(matrix) is an alternative

matrixSub <- matrixOne[7:8,3:5]

matrixSub

matrixMinus <- matrixOne[-(7:8),]

matrixMinus

matrixMinus <- matrixOne[-(7:8),4:8]

matrixMinus

length(matrixMinus)

class(matrixMinus)

matrixMinus[2,3] <- "try"

matrixMinus

class(matrixMinus)

class(matrixMinus[1,2])

matrixOne <- matrix(1:100, nrow=10, ncol=10, byrow=FALSE)

matrixOne

?matrix # help(matrix) is an alternative

matrixSub <- matrixOne[7:8,3:5]

matrixSub

matrixMinus <- matrixOne[-(7:8),]

matrixMinus

matrixMinus <- matrixOne[-(7:8),4:8]

matrixMinus

length(matrixMinus)

class(matrixMinus)

matrixMinus[2,3] <- "try"

matrixMinus

class(matrixMinus)

class(matrixMinus[1,2])

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

ben <- seq(1,5,1)

ben1 <-rep(ben,2)

aa <- data.frame(nickname=c("John","Mary","Leo"), weight=60:62, Height=c(160,170,180))

aa

class(aa)

rm(list=c("aa","matrixMinus"))

rm(list=c("pricesM1","pricesPercentM1"))

bb <- matrix(1:12, nrow=6)

class(bb)

bb <- as.data.frame(bb)

class(bb)

colnames(bb) <- c("col1","col2")

bb <- as.matrix(bb)

cc <- c(123,45,678)

class(cc)

cc <- as.character(cc)

class(cc)

cc

cc[1]+cc[3]

dd <- list("qwe",590,"zzz")

dd

ee <- list(c("qwe","asd"),590,"zzz")

ee

ff <- list(dd,c(1,2,3))

ff <-list(dd,ee,c(1,2,3))

ff[[2]][3]

ff[[2]][3] <- 50

for (i in -3:7) {

print(i^2)

}

for (i in -3:7) {

for (j in 6:9) {

print(i^2+sqrt(j))

}

}

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

getwd()

save(list=c("aa","bb"), file=paste0(getwd(),"/output-2020-10-16.RData"))

load(paste0(getwd(),"/output-2020-10-16.RData"))

dd <- list("qwe",590,"zzz")

dd

ee <- list(c("qwe","asd"),590,"zzz")

ee

ff <- list(dd,c(1,2,3))

ff

ff[[2]][3]

ff[[2]][3] <- 50

for (i in -3:7) {

print(i^2)

}

for (i in -3:7) {

for (j in 6:9) {

print(i^2+sqrt(j))

}

}

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

getwd()

save(list=c("aa","bb"), file=paste0(getwd(),"/output-2020-10-16.RData"))

load(paste0(getwd(),"/output-2020-10-16.RData"))

Video5~10：

aaa <- c(1,8,10)

bbb <- "try"

ccc <- data.frame(col1=c(12,34),col2=c(56,78))

ddd <- list(aaa,bbb,ccc)

eee <- list(ddd,aaa)

#看每個的class

library(stringr)

kk <- "abcdefgheijk"

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

kk <- "abcdefgheijk"

pp <- str\_split(kk, c("d|h"))

kk <- "abcdefgheijk"

pp <- str\_split(kk, c("[dh]"))

#注意pp的class

#注意pp[[1]][2]等等

kk <- "abcdefgheijk"

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

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

kk <- "John Wang"

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

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

kk <- c("John Wang", "Tom Cruise", "Kelly Chen")

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

library(pdftools)

pdfText=pdf\_text(paste0(getwd(),"/2327\_GS\_20211021.pdf"))

Amber=pdfText[1]

position1=str\_locate(Amber,'James')

position2=str\_locate(Amber,'is')

position3=str\_locate(Amber,'Attractive')

#接手機圖片！！！

allFiles=list.files(getwd(),full.names=F)

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

Mat1 = matrix(1:100, nrow = 10, ncol = 10, byrow = T)

Mat2 = Mat1^2

Mat3 = matrix(NA, nrow = 10, ncol = 10, byrow = T)

for (i in 1:10) {

for (j in 1:10){

if (i < j) {

Mat3[i,j] = Mat1[i,j] + Mat2[i,j]

} else if (i == j) {

Mat3[i,j] <- i

} else {

Mat3[i,j] <- j^2+i^3

}

}

}

if () {

}

i <- 0

while (i<=6) {

print(i^2)

i <- i+1

}

squarePlusCubic <- function(one, two) {

added <- one^2 + two^3

return(added)

}

squarePlusCubic(2,3)

print(squarePlusCubic(2,3)\*(-1))

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

library(stringr)

aa <- str\_locate("Where are you?", "er")

aa

class(aa)

aa[1,1]

aa[1,2]

bb <- str\_locate(hotelNumberOfRoomsData[1,2], "瞼D簫n糧]竅I")

bb

cc <- str\_sub(hotelNumberOfRoomsData[1,2], 5, 7)

hotelNumberOfRoomsData[,2] <- as.character(hotelNumberOfRoomsData[,2])

hotelNumberOfRoomsData[1,2] <- cc

dd <- str\_locate(NA, "100")

dd

if (is.na(dd[1,1])) {

print("abc")

} else {

print("def")

}

ff <- is.na(hotelNumberOfRoomsData[,2])

gg <- which(ff==TRUE)

Sys.setlocale(locale="Chinese (Traditional)\_Taiwan")

Video11~15：

Sys.getlocale()

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

Sys.setlocale(locale='Chinese (Traditional)\_Taiwan')

aa=dim(resultA)

class(aa)

aa[2]

remove(aa)

read\_xlsx('R影片指令選擇.xlsx')

RVSU=read\_xlsx('R影片指令選擇.xlsx')

aaa=RVSU[,2]=='毛子晴'

class(aaa)

any(aaa,na.rm=T)

#上面這行是檢查有沒有true

library(stringr)

bbb=str\_locate(resultA[1,1],'院')

bbb[1,1]

ccc=str\_locate(resultA[1763,1],'新北院')

startPos=bbb[1,1]

subString=str\_sub(resultA[1,1],1,startPos)

getSubStr=function(i){

bb=str\_locate(resultA[i,1],'院')

startPos=bb[1,1]

cc=str\_sub(resultA[i,1],1,startPos)

return(cc)

}

getSubStr(1:100)

getSubStr(1763)

PGS=print(getSubStr(1:10))

PGS=print(getSubStr(1763:1772))

for(j in 1763:1772){print(getSubStr(j))}