# Basic R

## Basic operation

เราสามารถใช้ R ในการคำนวณต่างๆ ได้ เช่น บวก ลบ คูณ หาร ยกกำลัง เป็นต้น

3+2

## [1] 5

3-2

## [1] 1

3\*2

## [1] 6

3/2

## [1] 1.5

3^2

## [1] 9

log(3)

## [1] 1.098612

sqrt(3)

## [1] 1.732051

3==3 # ตรวจสอบว่าข้อมูลเหมือนกันหรือไม่

## [1] TRUE

## Variable

### Variable assignment

R สามารถเก็บข้อมูลต่างๆ ไว้ในตัวแปรได้ เพื่อที่สามารถนำมาใช้ในภายหลัง โดยการเก็บตัวแปรนั้นจะใช้เครื่องหมาย <-

x <- 2  
x

## [1] 2

y <- 3  
y

## [1] 3

x+y # เราสามารถนำตัวแปรมาทำ operation ได้ตามปกติ

## [1] 5

x\*y

## [1] 6

x <- 5 # การลงข้อมูลในตัวแปรเดิมจะเป็นการลบตัวแปรเก่า  
x

## [1] 5

hellothisisRNAseqworkshop <- (x+y)^(x-y) # สามารถตั้งชื่ออะไรก็ได้ตราบใดที่ไม่เว้นวรรค  
hellothisisRNAseqworkshop

## [1] 64

### Type of variable

R นั้นสามารถรองรับตัวแปรต่างๆ ได้หลากหลาย ซึ่งเป็นได้ทั้ง ตัวเลข หรือตัวอักษร หรือแม้กระทั่งเก็บหลายข้อมูลภายในตัวแปรเดียวได้

x <- "Hello world" # ตัวอักษร  
x

## [1] "Hello world"

y <- c(1,2,3,4) # เก็บหลายตัวข้อมูลในตัวแปรเดียว  
y

## [1] 1 2 3 4

z <- list(c(1,2,3), 4, c("hello world", "I love R")) # เก็บข้อมูลในรูปแบบ list  
z

## [[1]]  
## [1] 1 2 3  
##   
## [[2]]  
## [1] 4  
##   
## [[3]]  
## [1] "hello world" "I love R"

class(x) # เราสามารถเช็คชนิดของตัวแปรได้โดยใช้ function class()

## [1] "character"

ลักษณะตัวแปรต่างๆ ใน R มีดังนี้

| ชนิด | ตัวอย่าง | คำอธิบาย |
| --- | --- | --- |
| numeric | 1, 2.3, 5 | จำนวนจริง รวมทศนิยม |
| integer | 1, 2, 3 | จำนวนเต็ม เป็น subset ของ numeric |
| complex | 1i | จำนวนเชิงซ้อน |
| character | “สวัสดี”, “Hello world” | ตัวอักษร ต้องอยู่ในเครื่องหมาย ” ” |
| factor | “a”, “b”, “c” | คล้าย character แต่มีจำนวนตัวแปรจำกัด |
| logical | TRUE, FALSE | ตามหลักตรรกศาสตร์ |
| vector | c(1,2,3) | หลายข้อมูลใน 1 ตัวแปร โดยต้องเป็นตัวแปรชนิดเดียวกัน |
| list | list(1, c(1,3,4), “Hello”) | หลายข้อมูลใน 1 ตัวแปร โดยไม่จำเป้นต้องเป็นตัวแปรชนิดเดียวกัน |
| dataframe | data.frame(x=3, y=2) | ตาราง |

## Matrix and Dataframe

เนื่องจาก R นั้นเป็นโปรแกรมที่ส่วนมากใช้ในการวิเคราะห์ทางสถิติ ซึ่งเกี่ยวข้อมูลส่วนใหญ่จะถูกเก็บในรูปของตาราง R จึงมีตัวแปรที่เก็บข้อมูลในรูปของตารางโดยเฉพาะ เรียกว่า matrix และ dataframe ซึ่งเราจะใช้เป็นหลักในการวิเคราะห์ข้อมูลใน R

mat <- matrix(c(1,2,3,4), nrow=2)  
mat

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 4

class(mat)

## [1] "matrix" "array"

df <- data.frame(x=c(3,4),y=c(2,5),z=c(4,7))  
df

## x y z  
## 1 3 2 4  
## 2 4 5 7

class(df)

## [1] "data.frame"

โดยตารางนั้นจะประกอบด้วยสองส่วนหลักๆ คล้าย excel spreadsheet ได้แก่

* Column (คอลัมน์): คือ ข้อมูลในแนวตั้ง ซึ่งแถวบนสุดจะเป็นชื่อ column นั้นๆ
* Row (แถว): คือ ข้อมูลในแนวนอน

โดย matrix นั้น สามารถเก็บ variable ในรูปแบบเดียวกันได้เท่านั้น แต่ dataframe สามารถเก็บข้อมูลต่างชนิดร่วมกันได้ โดยมีข้อแม้ว่า column เดียวกัน จะต้องเป็นข้อมูลชุดเดียวกัน

## Subset

เราสามารถดึงข้อมูลแค่บางส่วนออกมาจาก vector, list, matrix หรือ dataframe ได้ เรียกว่าการ subset

x <- c("a","b","c","d")  
x[3] # subset โดยระบุตำแหน่ง

## [1] "c"

x[1:3] # subset หลายตำแหน่ง

## [1] "a" "b" "c"

x[c(1,3)] # subset หลากหลายตำแหน่งแบบจำเพาะ

## [1] "a" "c"

y <- list(c(1,2,3), c("a","b","c"))  
y[1] # subset list ตามตำแหน่ง (จะได้ list ย่อยออกมา)

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

y[[1]] # ดึงข้อมูลที่อยู่ใน list ออกมา

## [1] 1 2 3

ในส่วนของ matrix และ dataframe นั้น เราสามารถ subset ตามตำแหน่งได้ โดยการระบุ row และ column ตามลำดับ

mat

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 4

mat[1,2] # 1st row, 2nd column

## [1] 3

df

## x y z  
## 1 3 2 4  
## 2 4 5 7

df[1,3] # 1st row, 3rd column

## [1] 4

ในส่วนของ dataframe นั้น เราสามารถ subset ได้โดยใช้ชื่อของ column อีกด้วย

df["x"] # subset เป็น column ย่อย

## x  
## 1 3  
## 2 4

df[["x"]] # subset ข้อมูลที่อยู่ใน column นั้น

## [1] 3 4

df[[2, "x"]] # ระบุแถวด้วย

## [1] 4

df$x # เหมือนกัน df[["x"]]

## [1] 3 4