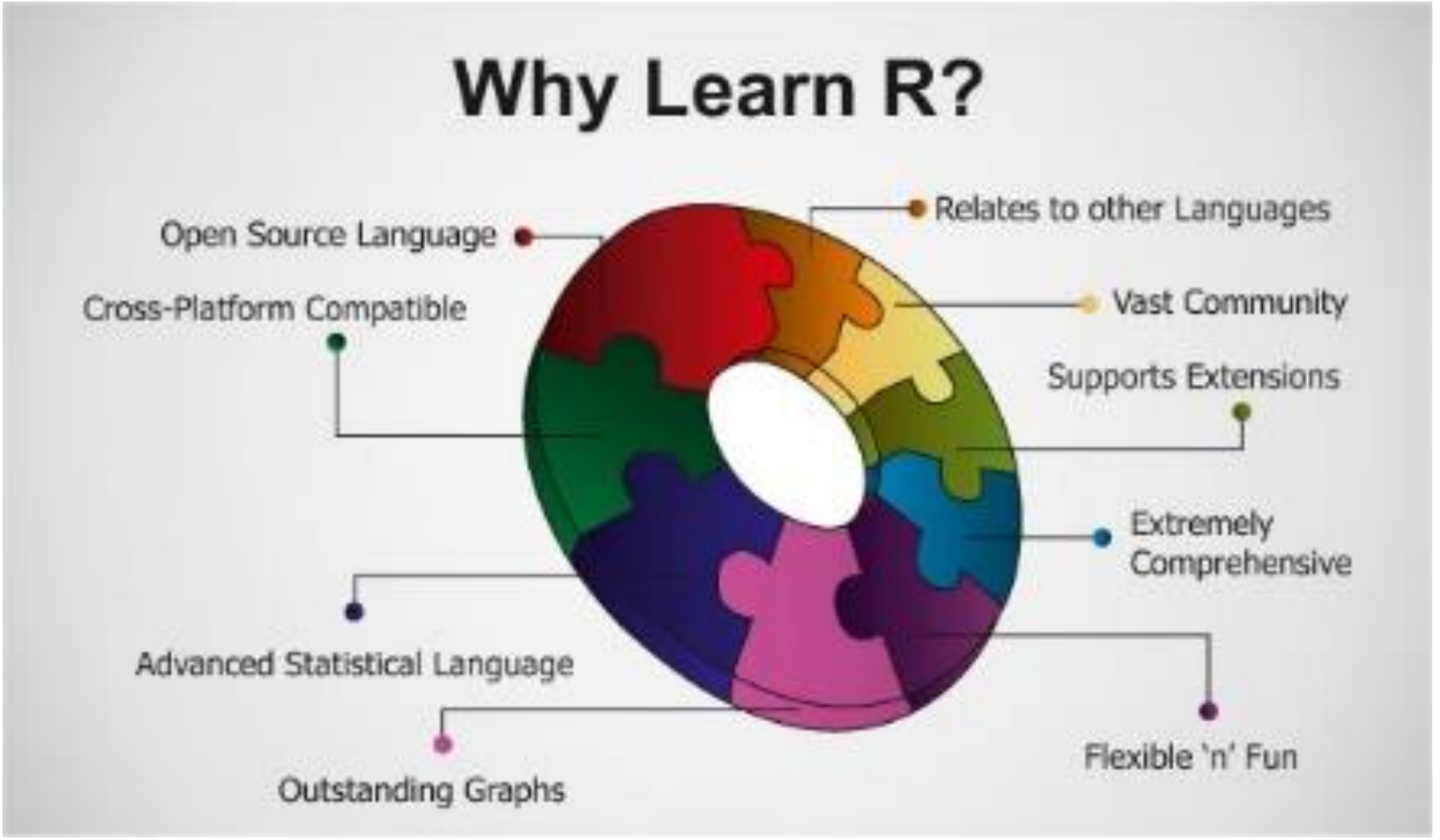
Information Visualization

Lab 1 – R Introduction

R as a Programming Language

- Open source language for statistical computing and graphics.
- Data summary and exploration, graphical presentation and data modelling .

Why R?


Why Learn R?
Open Source Language
Cross-Platform Compatible
Advanced Statistical Language
Outstanding Graphs
Relates to other Languages
Vast Community
Supports Extensions
Extremely Comprehensive
Flexible 'n' Fun

Get R to Work

- Download R
  - https://cran.r-project.org/bin/windows/base/
- Download RStudio Desktop IDE – Free edition
  - https://www.rstudio.com/products/rstudio/download/

R Quick Start

- Variable assignment
- Assignment could be done by “=“ or “ <- “ operators, but most R developers prefer “<-”

```
# '=' objects will be defined within the scope of their functions
x = 1
x
# '<-' objects will be defined within the scope of user workspace
y <- "Welcome R!"
y
```

Build-in Functions (sample)

```r
ls() # list objects in current workspace
rm(x) # remove object from current workspace
sqrt(16)
help(sqrt) # return information about the specified function
?sqrt # do the same as help()

# sequences
# sequence of numbers 1 to 10
x <- 1:10
seq(1, 10)
# seq(start, end, increment rate)
seq(1, 10, by = 2)
# sequence of 6 numbers start from 8 and end at 20
seq(8, 20, length = 6)

#others
mean(x)
var(x)
# gets structure type about an object
str(x)
# provide a collection of summary information about the object
summary(x)

# repeat
rep(0,100) # repeats number 0 for 100 times
rep(1:3, 6) # repeats the sequence '1 2 3' 6 times
```

Exercise 1.1

1. Define

```
x <- c(4,2,6)
y <- c(1,0,-1)
```

Decide what the result will be of the following:

(a) length(x)
(b) sum(x)
(c) sum(x^2)
(d) x+y
(e) x*y
(f) x-2
(g) x^2

Use **R** to check your answers.

Exercise 1.1

2. Decide what the following sequences are and use **R** to check your answers:

(a) seq(4,10,by=2)
(b) seq(3,30,length=10)
(c) seq(6,-4,by=-2)

Exercise 1.1

3. Determine what the result will be of the following **R** expressions
(a) rep(2,4)
(b) rep(c(1,2),4)
(c) rep(1:4,rep(3,4))

4. Use the rep function to define simply the following vectors in R.
(a) 6,6,6,6,6,6
(b) 5,8,5,8,5,8,5,8
(c) 5,5,5,5,8,8,8,8

R Data Types

```r
#1# Vectors
# sequence of data elements of the same basic type.
a <- c(1,2,3)
a
b <- c("x","y","z")
b
c <- c("a", 3) # both are characters
c
str(c)
str(a)

# Vector Arithmetics
x <- c(2,4,6)
y <- c(1,3,5)

x + y; x*5; y^2 # ';' needed to write multiple commands in one line

# vector index starts at 1 not 0
x[1]
x[-1] # removes the first element that's how a negative index work
x[5] # NA : Not defined cause it's out of range
```

R Data Types

```r
#2# Matrix
# A two dimentinal array of data elements of the same basic type
A <- matrix(c(1,2,3,4,5,6), nrow = 2, ncol = 3, byrow = F)
B <- matrix(c(1,2,3,4), nrow = 2, byrow = T)
A[2,3]
t(A) # Transpose
solve(B) #inverse for square matrics
```

```r
#3# Data Frames
# Store different object types organized in a set or rows and column
# much like table data structure

df <- data.frame(c(1,2,3), c("x","y","z"), c(1.3,0.5,2.2))
df
str(df)
names(df) <- c("no.","name","degree")
df
```

Exercise 1.2

1. If x<- c(5,9,2,3,4,6,7,0,8,12,2,9) decide what each of the following is:

(a) x[2]
(b) x[2:4]
(c) x[c(2,3,6)]
(d) x[c(1:5,10:12)]
(e) x[-(10:12)]

Exercise 1.2

2. The data y<-c(33,44,29,16,25,45,33,19,54,22,21,49,11,24,56) contain sales of milk in liters for 5 days in three different shops (the first 3 values are for shops 1,2 and 3 on Monday, etc.) Produce a statistical summary of the sales for each day of the week and also for each shop.

Exercise 1.2

1. Create in **R** the matrices

$$x = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix}$$

and

$$y = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Calculate the following and check your answers in R:

(a) 2*x
(b) x*x
(c) x%*%x
(d) x%*%y