CBIO-628

Unit 3

Data Analytics with R: R basics, Data Types, Data Structure: Vectors and Factors, Vector operations, Arrays & Matrices, Lists, and Dataframes. Conditions and loops, Handling Files, Data visualization, Regression Analysis and Correlation analysis.

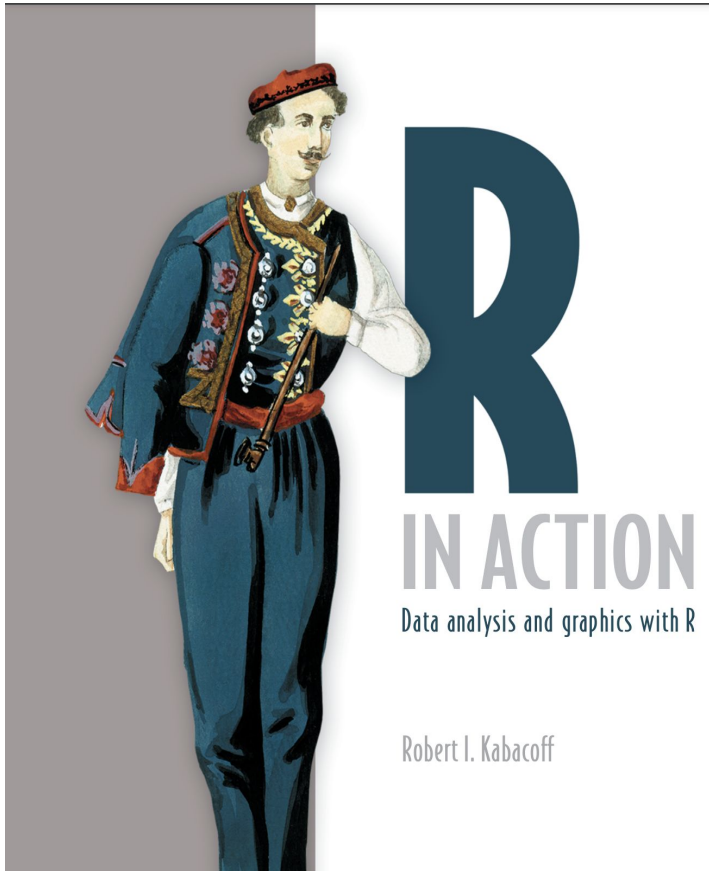
R?

- **R is a programming and statistical language for analytical purposes.**
- **R is also one of the most popular tools used for data visualization**
- **R is a simple and easy to learn, read & write**
- **Free and Open source**

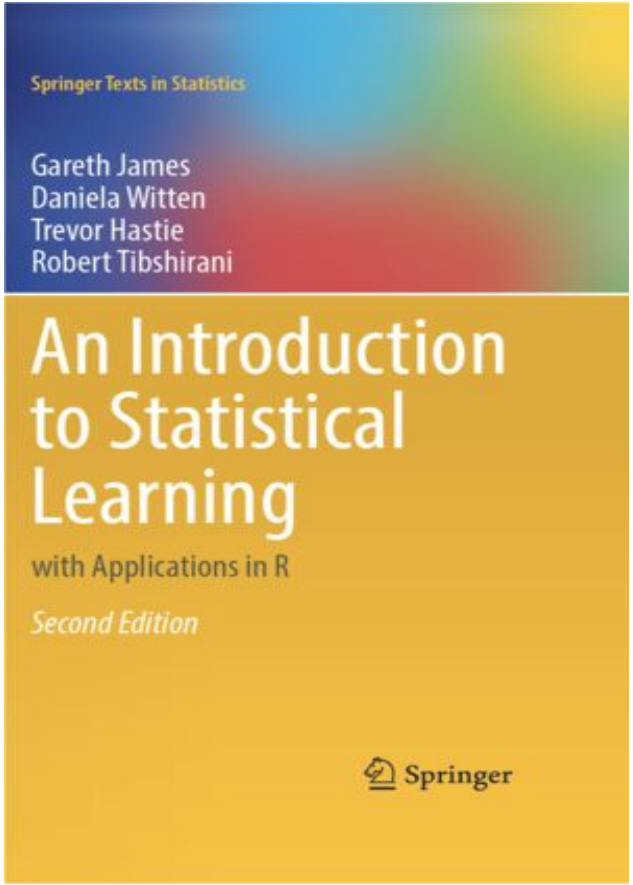
Installation steps

- Download R: *https://cran.r-project.org/bin/windows/base/*
- Download R studio: *https://www.rstudio.com/*
- Online: *https://rstudio.cloud/*

Reference



http://www.cs.uni.edu/~jacobson/4772/week11/R_in_Action.pdf



https://hastie.su.domains/ISLR2/ISLRv2_website.pdf

Variables in R

- Variables are nothing but reserved memory locations to store values. This means that when you create a variable you reserve some space in memory.

Data operators in R

- **Arithmetic Operators**

  \+ - * / ^ %% %/%

- **Assignment Operators**

  = <- <<- ->

- **Relational Operators**

  > < == != >= =<

- **Logical Operators**

  & | !

- **Special Operators**

  : %in%

Data Types

1. **Vectors**
   - *vectors is a sequence of data elements of the same basic type*
2. **Lists**
   - *Contain elements of different types like-numbers,strings, vectors and another list inside it*
3. **Arrays**
   - *Arrays hold multidimensional data.*
4. **Matices**
   - *special case of two-dimensional arrays*
5. **Factors**
   - *Factor is a data structure which are used to categorize the data and store it as levels*
6. **Data Frames**
   - *a table or a **two-dimensional array-like** structure in which each column contains values of one variable and each r**ow** contains one set of values from each **column***

N.B. *We do not need to declare a variables before using them*

Data types

Vector vs List vs Array vs Matrices

- Vector is a special cases for one dimensional arrays
- A list is a variety of objects under one name
- Matrix is a special cases for two dimensional arrays
- Array can also have any dimension level

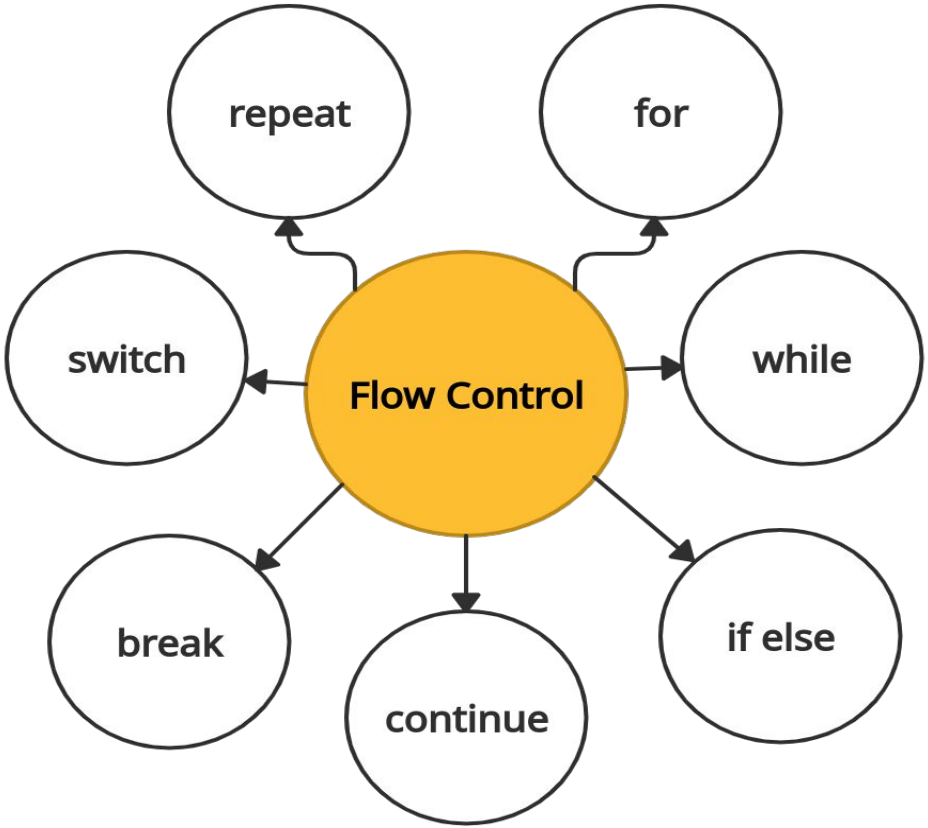
Factors

- Factor is a data structure which are used to categorize the data and store it as levels
- Can store both integers and strings
- They are useful in data analysis or modeling

Data frame

- A data frame is a table or a ***two-dimensional array***-like structure in which ***each column contains values of one variable*** and e***ach row contains one set of values from each column.***

- Characteristics of a data frame.
  - The column names should be non-empty
  - The row names should be unique
  - The data stored in a data frame can be of numeric, factor or character type
  - Each column should contain same number of data items


repeat
for
switch
Flow Control
while
break
continue
if else

Loops Repeat Actions

Repeat things until the loop condition is true

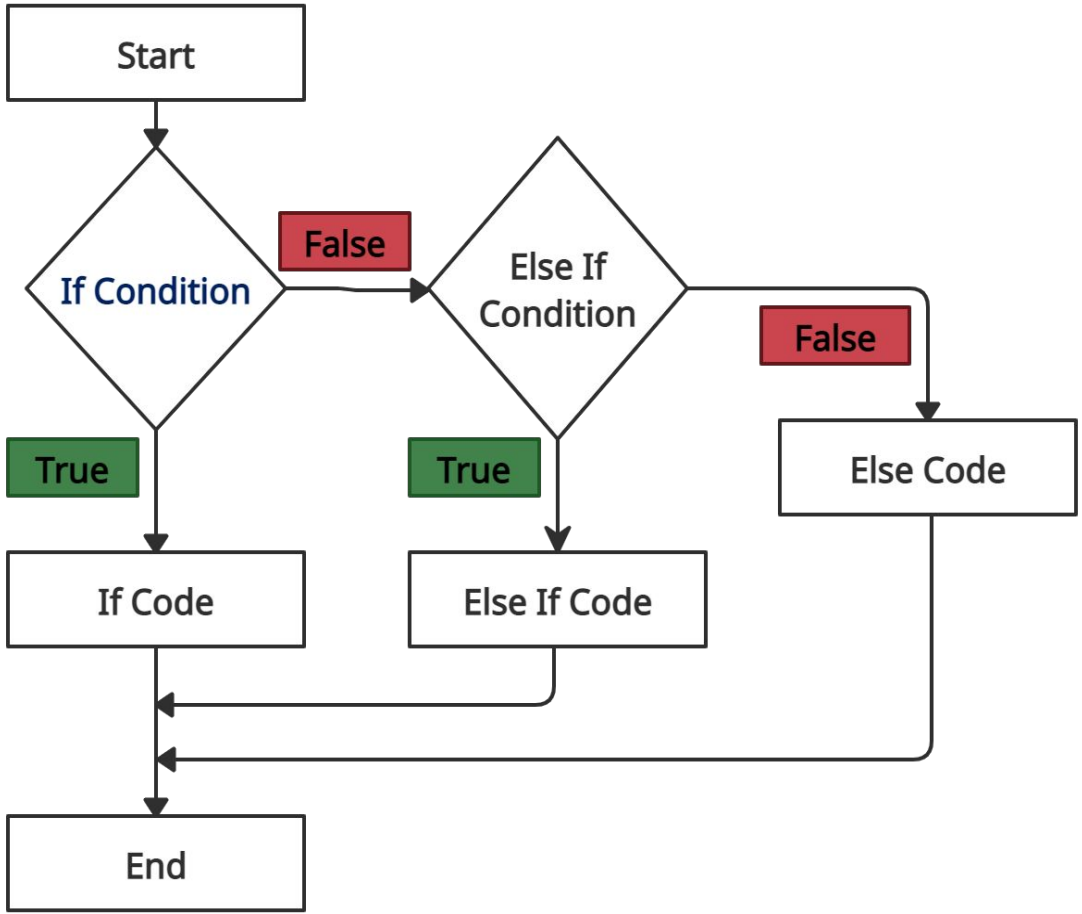
Repeat the things until the loop condition is true

Repeat things till the given number of timers

2. if...else statement

Syntax:

```
if (condition 1):
    statements 1 ...
    .
    .
    .
else
    statements n ...
```



3. Switch statement

Syntax:

```
switch (expression,
value1:   Statement1
value2:   Statement2
.
.
, default Statement
)
```

4. repeat statement

Syntax:

```
repeat {
   commands
   if(condition) {
      break
   }
}
```

5. while statement

Syntax:

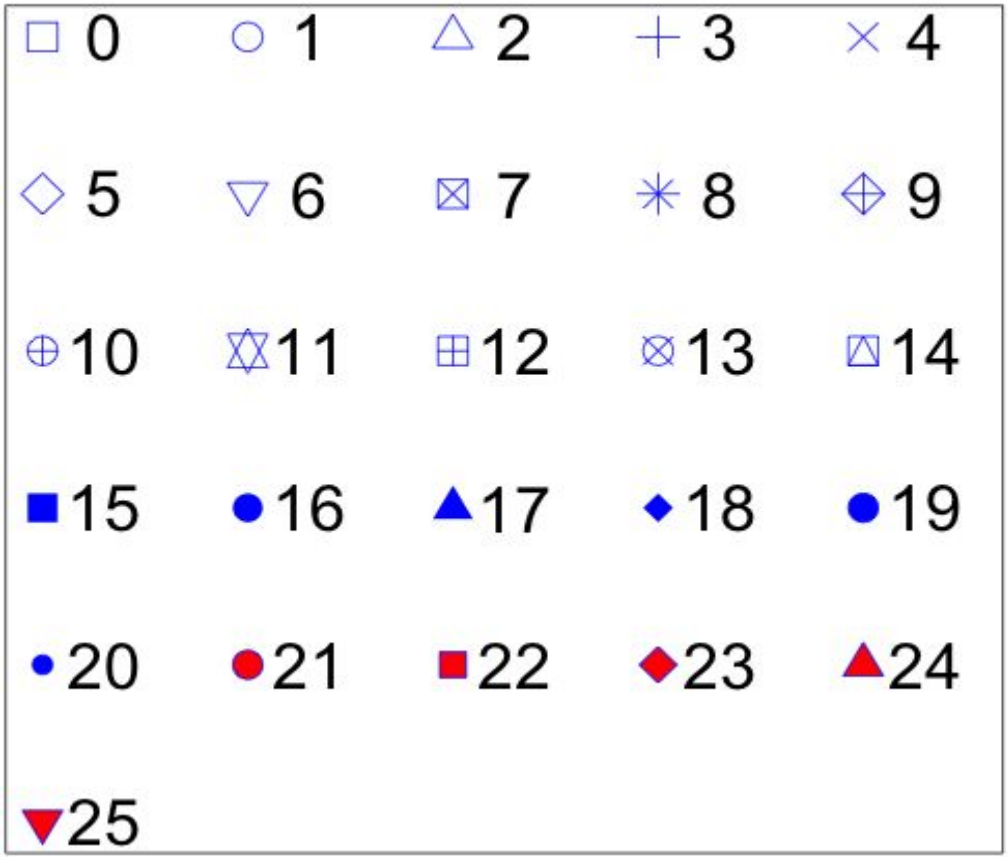
```
while (condition is True)
    {
            statements...
    }
```

6. for statement

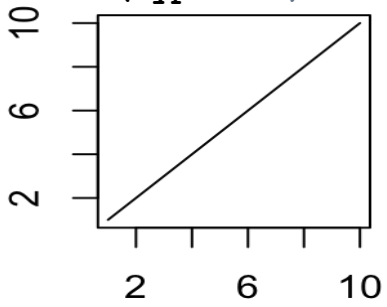
Syntax:

```
for(value in vector)
{
        statements...
}
```

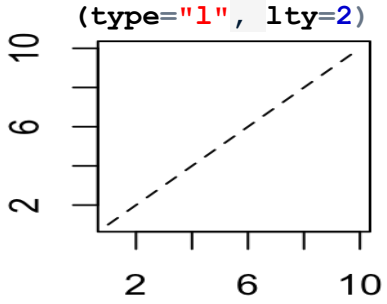
pch arguments



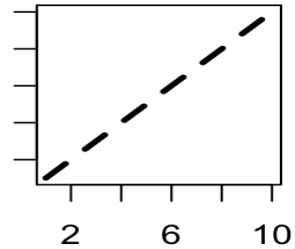
(type="l")



(type="l", lty=2)



type="l", lty=2, lwd=3)



- **lty** can be used to specify the **line type**.
- To change **line width**, the argument **lwd** can be used.

6.'twodash'

5.'longdash'

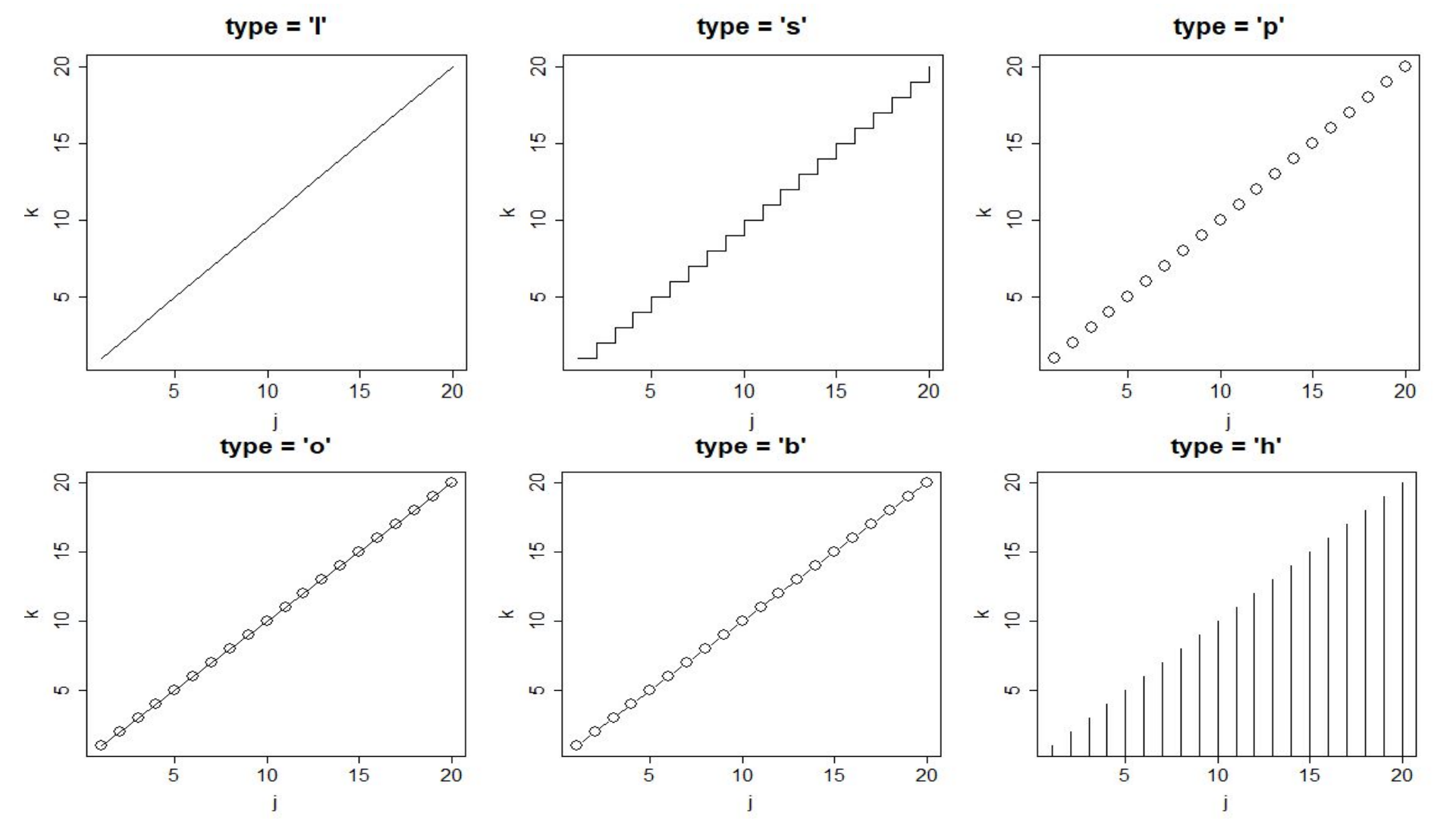
4.'dotdash'

3.'dotted'

2.'dashed'

1.'solid'

0.'blank'


type = 'l'
type = 's'
type = 'p'
type = 'o'
type = 'b'
type = 'h'
k
j
5
10
15
20