DATA SCIENCE FOR ENGINEERS

Week 1

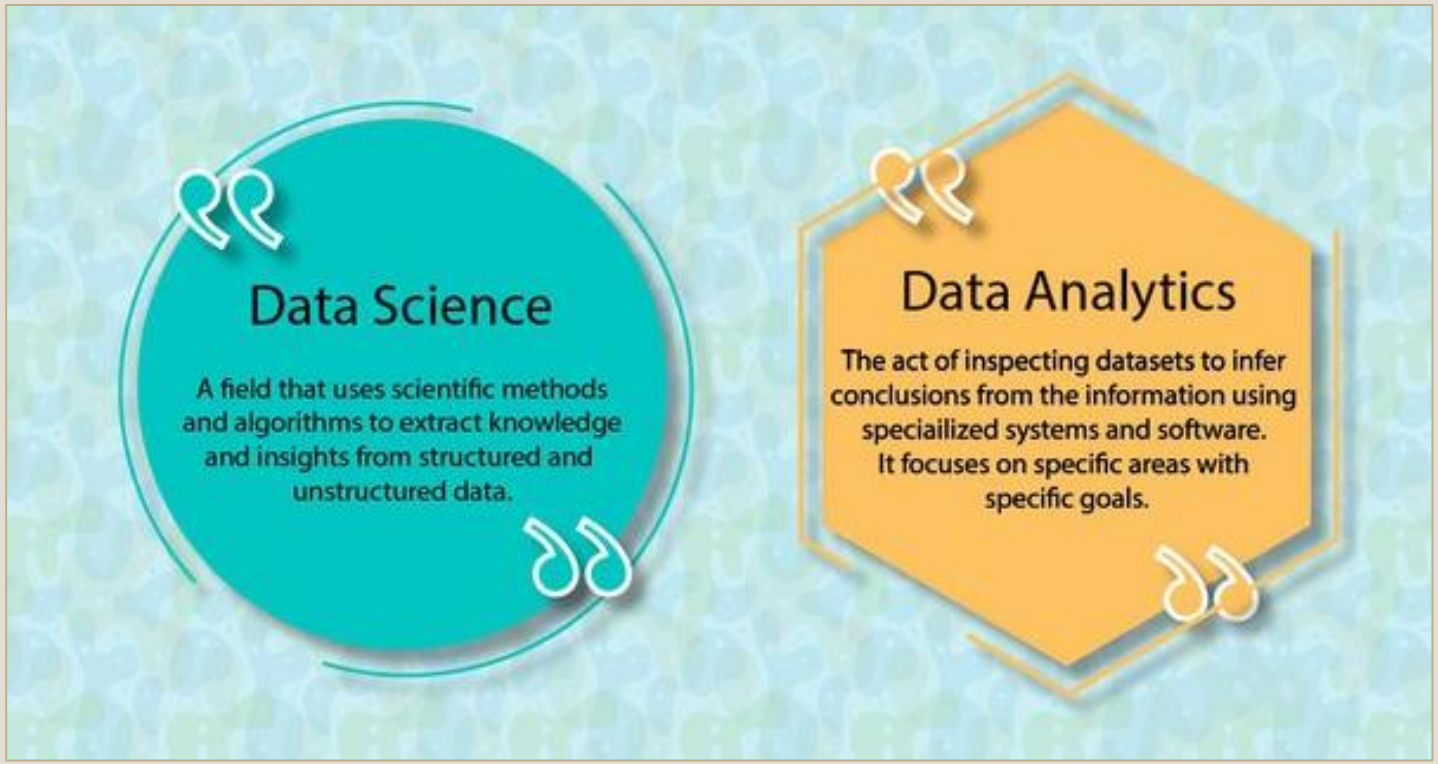
Session Co-ordinator: Abhijit Bhakte

R Code Link

https://drive.google.com/drive/folders/1bjs2_9hHMVp4iNOLO3bbMzgCej3nsl8H?usp=sharing

Data Science & Data Analytics

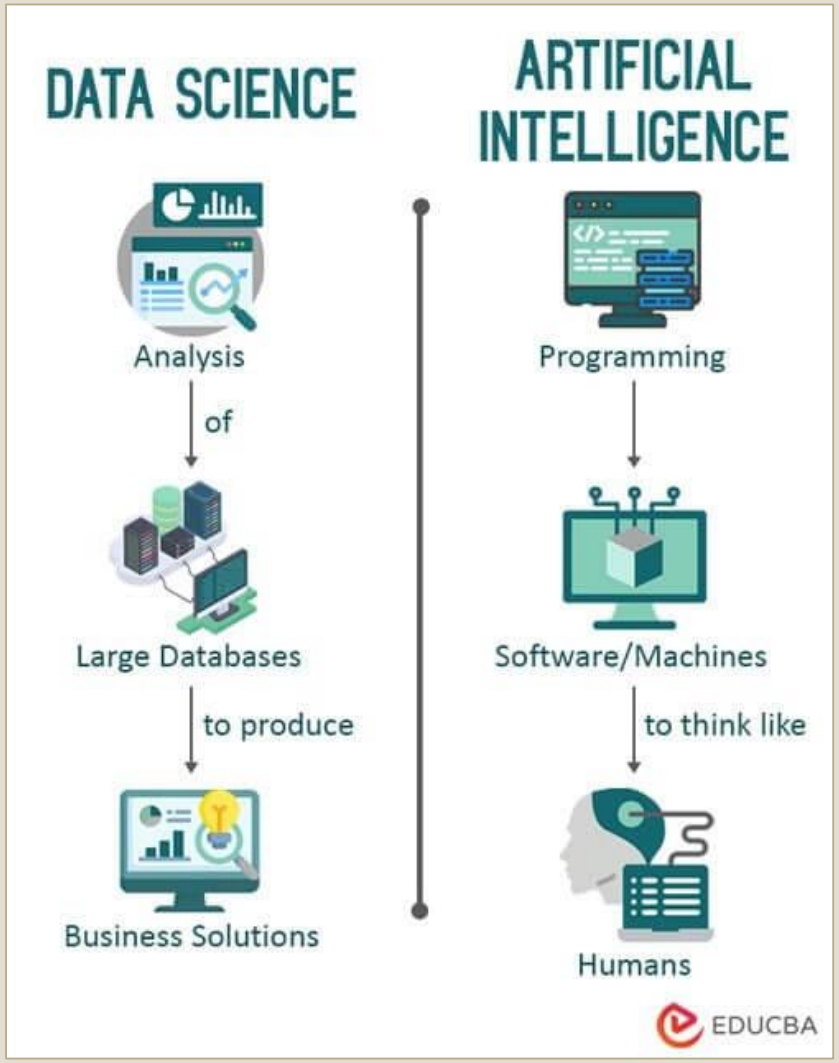
- **Data Science** often involves **using data to build models** that can predict future outcomes,
- **Data Analytics** tends to focus more on **analyzing past data to inform decisions** in the present



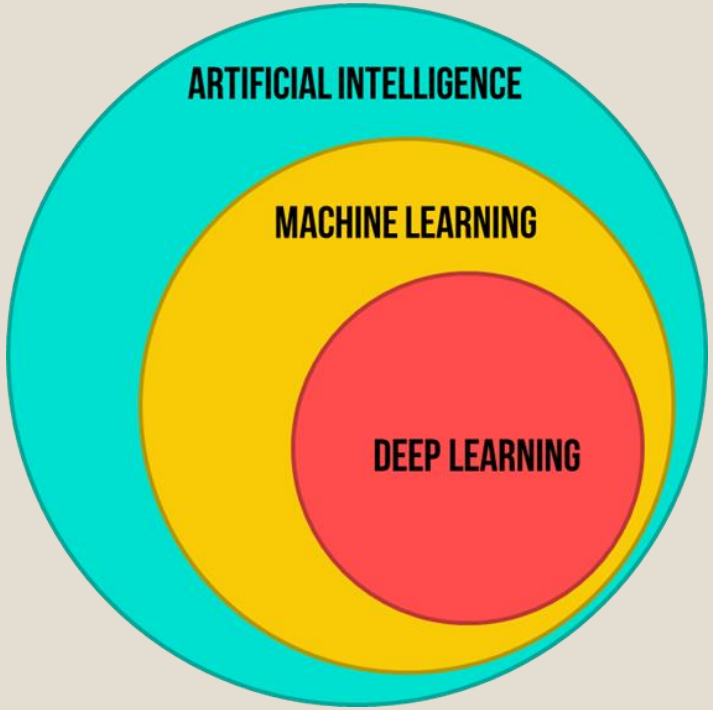
Data Science & Artificial Intelligence

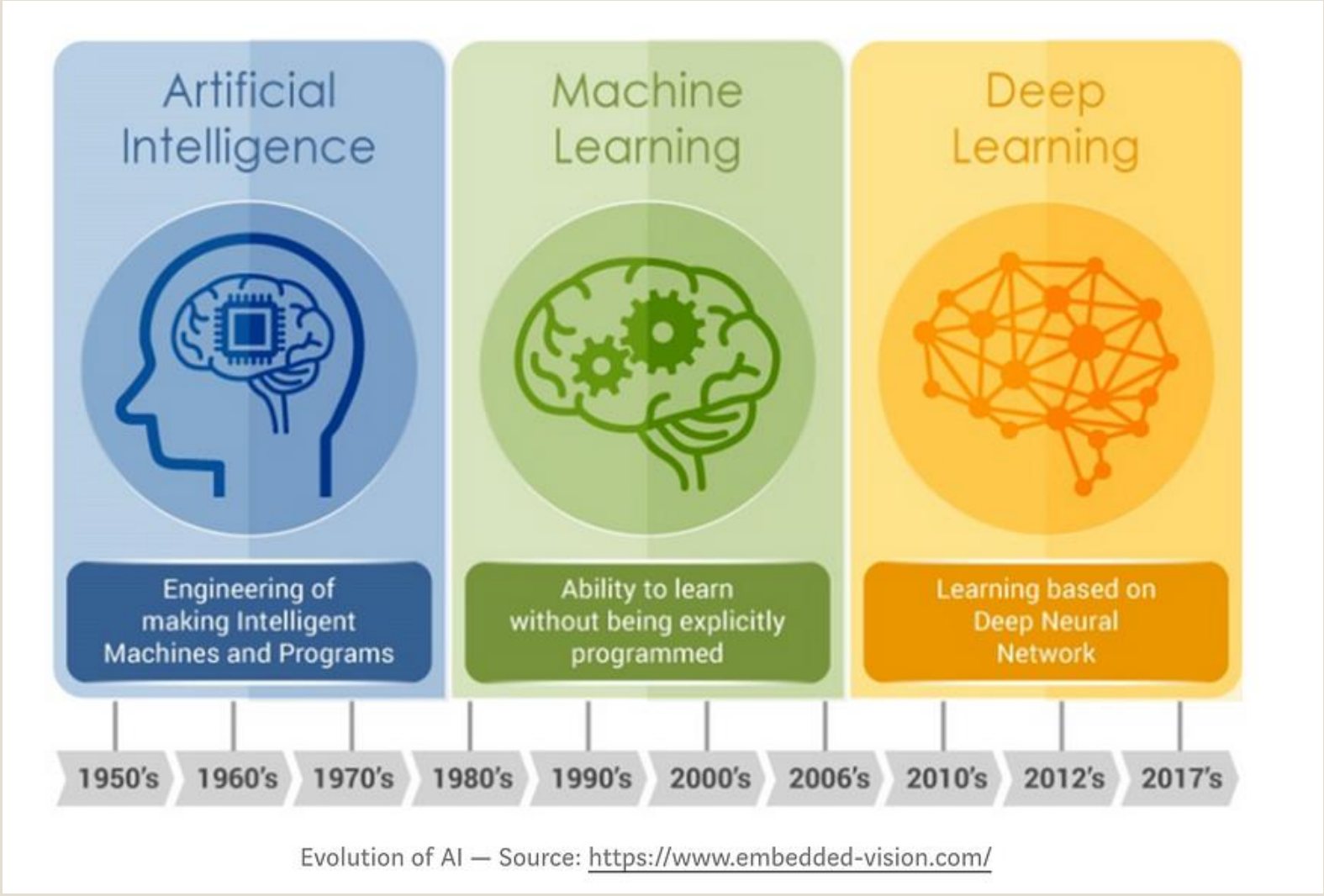
- **Data Science** is about **extracting valuable insights** from data and using them to **make informed decisions**

- **Artificial Intelligence** is concerned with **creating intelligent systems** that can perform tasks requiring **human-like intelligence**



AI, ML and DL





Evolution of AI — Source: https://www.embedded-vision.com/

Highlights of Week 1

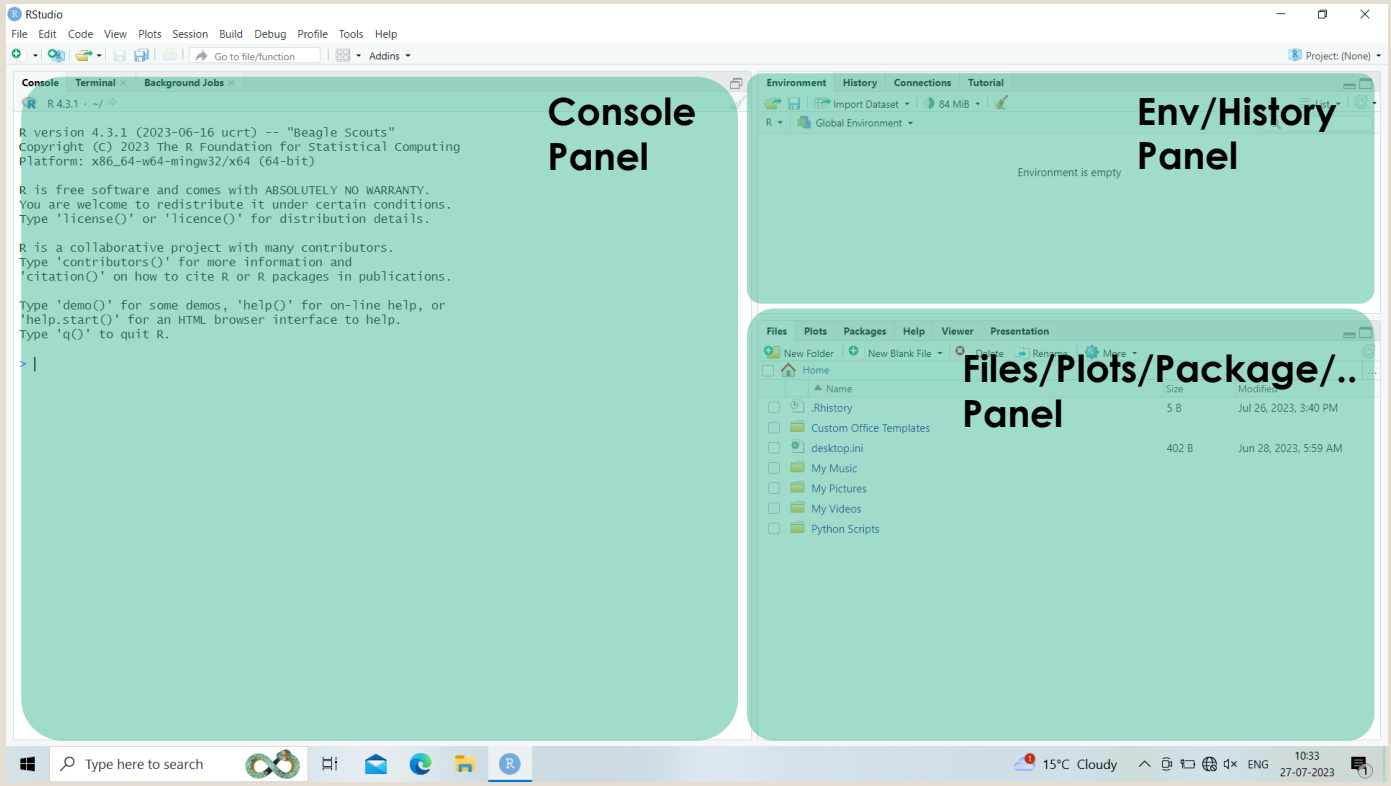
Topic	Details
Introduction to R language	Interface introduction/ How to create file/ save file /run code/ comment code/ clear workspace
Variables and Datatypes in R	Variable naming/ Types of Data types/ Task in Datatypes/ Object in R (vector, list, dataframes)
Dataframes object in R	Introduction to Dataframes/ How to Create/ Edit element /Add row / Delete row / Joining two dataframes
Different operations in R	Arithmetic / Logical / Matrix operations
Functions in R	Introduction to function/ loading function/ calling function/ MIMO function/ Looping over object
Control Structures	If-else-if statement/ for loop/ while loop/ break in loop
Data Visualization	Scatter plot/ Line plot/ Bar plot

Introduction to R language

- R is a powerful and versatile programming language **used for statistical computing, data analysis, and graphical visualization.**
- **Created by Ross Ihaka and Robert Gentleman** at the University of Auckland, New Zealand in the early 1990s.
- It is an **open-source language,** meaning it is freely available and has a large, active community of developers.
- **R provides a wide range of packages and libraries**, making it an excellent choice for data scientists, statisticians, and researchers.

Introduction to R studio

- R Studio is an **integrated development environment** (IDE) for the R programming language.
- It provides **a user-friendly interface** and powerful data analysis, visualization, and statistical modeling tools.



Introduction to R studio

- Interface Introduction
- Setting workspace (setwd('folder path'))
- Creating R file (ctrl+shift+N)
- Writing code
- Saving R file (ctrl+S)
- Running code (Run(ctrl+enter)/Source(ctrl+shift+S)/Source with eco(ctrl+shift+enter))
- Commenting inside code
- Clearing, saving, loading of workspace

Variable naming & Data type

- Variable name always starts with alphabet without any special character
  - Ex: var1,var_1,var.1 (currect) / 1var,var@ (Incorrect)
- Data Type: Indicates the type of variable

Logical	TRUE/FALSE
Integer	Set of all integer (1,6)
Complex	Set of complex numbers (6+4i)
Character	'a', 'john', "6"

- Functions in Datatype

Function	**Syntax**	**Example (a='cat')**
To find type of datatype	typeof(var)	typeof(a)→character
To verify the datatype	is.datatype(var)	is.integer(a)→ False
To convert into other datatype	as.datatype(var)	as.integer(a)→ NA(error)

R studio

Questions

Q) Which of the following is a valid variable name in R?

A) my_var-1

B) 2nd_variable

C) total score

D) My_Var_1

Q) In R, the function 'as.integer()' is used for:

A) Converting a variable to a integer datatype.

B) Declaring a new custom integer datatype.

C) Concatenating two integers.

D)All

Object in R

- **Vectors**: Ordered collection of same datatype

```
X=c(1,2,3,4,5)
Print(X)
```

OUTPUT: 1 2 3 4 5

- **List**: Ordered collection of objects

```
day=c('sat','sun')
temp=c(45,35)
weather=list(day,temp)
```

OUTPUT: [[1]] 'sat' 'sun'
[[2]] 45 35

- **dataframe**: Collection of tabular data

```
day=c('sat','sun')
temp=c(45,35)
weather=data.frame(day,temp)
```

OUTPUT:

	day	temp
1	sat	45
2	sun	35

R studio

Questions

```r
my_list <- c("apple", "banana", "orange", "grape", "kiwi")
index_result <- my_list[c(3, 1, 4)]
```

Q) What is output of index_result?

A) "orange", "apple", "grape"

B) "orange", "banana", "kiwi"

C) "banana", "orange", "grape"

D) "kiwi", "grape", "orange“

Q) How to get Bob's maths mark?

A) student_data[1, "Math"]

B) student_data[2, "Math"]

C) student_data["Bob", "Math"]

D) student_data["Math", 2]

```r
# Sample dataframe
student_data <- data.frame(
  Name = c("Alice", "Bob", "Charlie", "David", "Eva"),
  Math = c(85, 92, 78, 68, 95),
  Science = c(90, 88, 76, 85, 93),
  History = c(80, 75, 85, 90, 82)
)
```

Operators in R

- **Arithmetic:** used to perform mathematical operations

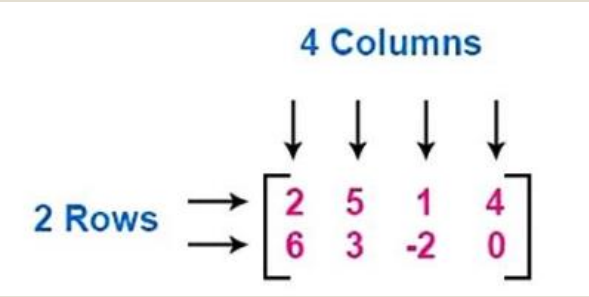
Operator	Description
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponent
%%	Modulus (Remainder from division). Give the remainder of the first vector with the second
%/%	Integer Division. The result of division of first vector with second (quotient)

- **Logical:** used to perform logical operations

Operator	Description
<	Less than
<=	Less than or equal to
>	Greater than
>=	Greater than or equal to
==	Exactly equal to
!=	Not equal to
!x	Not x
x	y
x & y	x AND y
isTRUE(x)	Test if X is TRUE

Matrix

- It is a rectangular arrangement of rows and columns
- Rows are horizontal and columns are vertical



Syntax

```
A=matrix(vec,nrow,ncol,byrow)
```

Example

```
A=matrix(c(1,2,3,4,5,6,7,8,9),nrow=3,ncol=3,byrow=True)
```

OUTPUT:
```
1 2 3
4 5 6
7 8 9
```

If-else statement

- This helps us to execute the statement if condition is TRUE

Syntax

```
If (condition){
statement
} else {
another statement
}
```

Example: Odd/Even Number

```
x=53
if (x%%2==0){
  print('The number is even')
}else{
  print('The number is odd')
}
```

OUTPUT: The number is odd

R studio

Questions

Q) What will be the value of the variable "result“ ?

A) "A"

B) "B"

C) 10

D) 15

```r
x <- 15
if (x > 10) {
  result <- "A"
} else {
  result <- "B"
}
```

Q) In a nested if-else statement, the _____ will be executed if none of the above conditions are true.

A) Code block A

B) Code block B

C) Code block C

D) None of the above

```r
if (condition1) {
    // Code block A
} else if (condition2) {
    // Code block B
} else {
    // Code block C
}
```

for loop

- For loop useful to iterate over the elements of a list, dataframe, vector, matrix, or any other object
- The syntax comprises of three elements: loop variable, sequence and statement

Syntax

```
for (iter in sequnce){
statement
}
```

Example: Square of numbers

```
for (i in 1:5){
  print(i**2)
}
```

OUTPUT: 1 4 9 16 25

R studio

Questions

Q) What is the output of following code

A) A

B) A B

C) A B C

D) C

```
v <- LETTERS[3]
for ( i in v) {
  print(v)
}
```

while loop

- while loop is used when you want to execute the statement until the condition is FALSE

Syntax

```
while (condition){
statement
}
```

Example: Square of numbers

```
x=0
while (x<=5){
  print(x**2)
  x=x+1
}
```

OUTPUT: 1 4 9 16 25

R studio

Questions

Q) which statement terminates the loop statement and transfers execution to the statement immediately following the loop?

A) goto

B) switch

C) break

D)label

Q) What is the ouput of the following code?

A) Odd Odd Odd Odd Odd

B) Even Even Even Even Even

C) Odd Even Odd Even Odd

D) Odd Even Even Odd Even

```
x <- 1
while (x <= 5) {
    if (x %% 2 == 0) {
        cat("Even ")
    } else {
        cat("Odd ")
    }
    x <- x + 1
}
```

Function in R

- A function is a set of statements organized together to perform a specific task
- source button help to load the function

Syntax

```
funct_name=function(argument){
Statements}
```

Example: cube of number

```
cube_func=function(num=6){
  cube=num**3
  return(cube)
}
```

OUTPUT: 216

R studio

Questions

Q) Which of the following options represents the correct syntax for defining a function in the R programming language?

```r
function myFunction(parameter1, parameter2) {
  # Function body
  # ...
  return(result)
}
```

(A)

```r
def myFunction(parameter1, parameter2)
  # Function body
  # ...
  return(result)
```

(B)

```r
myFunction <- function(parameter1, parameter2) {
  # Function body
  # ...
  return(result)
```

(C)

```r
function myFunction(parameter1, parameter2)
  # Function body
  # ...
  return(result)
```

(D)

Questions

```r
calculate_sum_of_squares <- function(vector) {
  sum_sq <- sum(vector^2)
  return(sum_sq)
}
```

Q) If the input vector is c(1, 2, 3, 4), what would be the output?

A) 30

B) 25

C) 55

D) 10

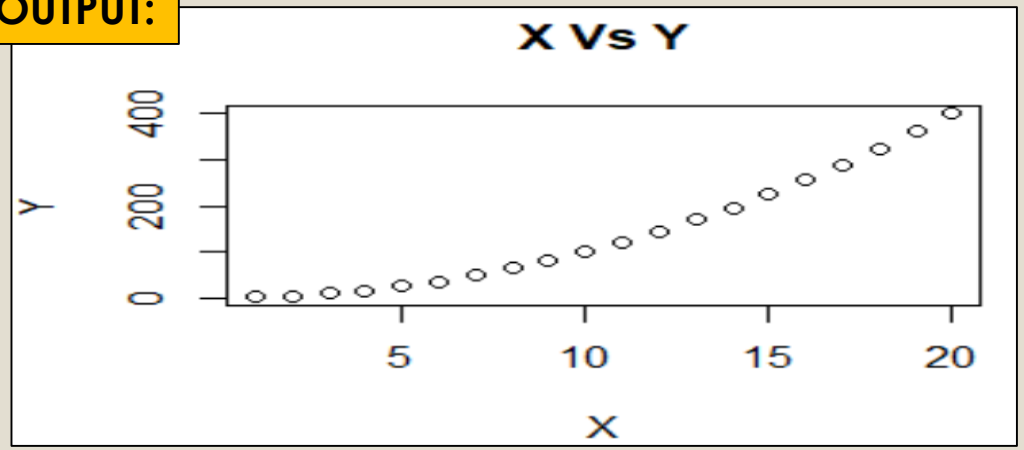
Data Visualization

❑ Scatter plot

Plot(x,y,…)

```
x=1:20
y=x**2
plot(x,y,main='X Vs Y',xlab='X',ylab='Y', pch=1)
```

OUTPUT:

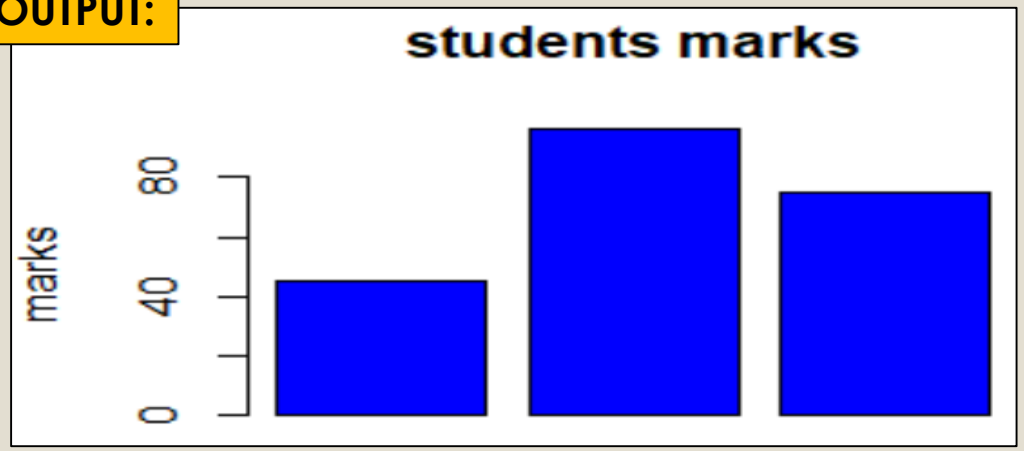


❑ Bar plot

barplo(height,…)

```
barplot(mark,ylab='marks',main='students
marks',col='blue')
```

OUTPUT:



Thank you