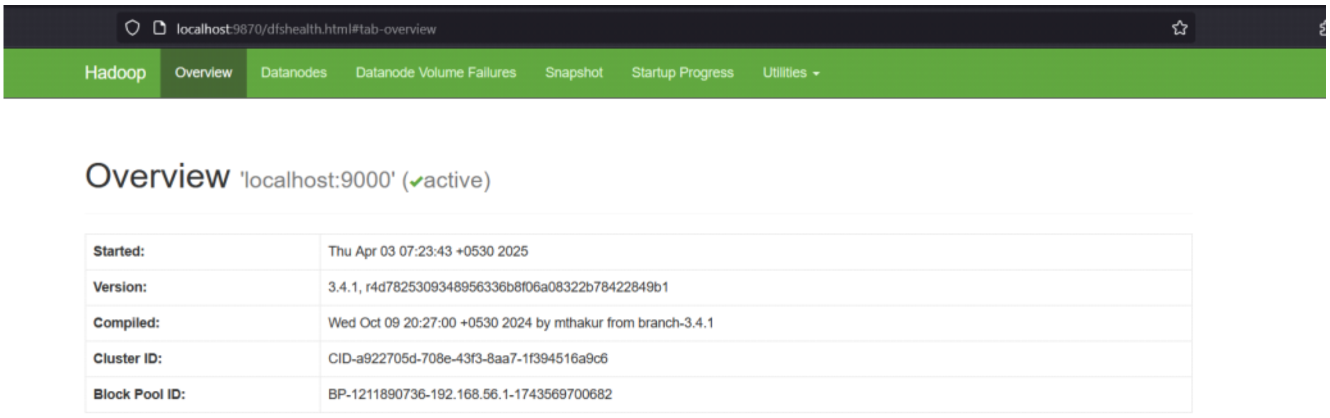
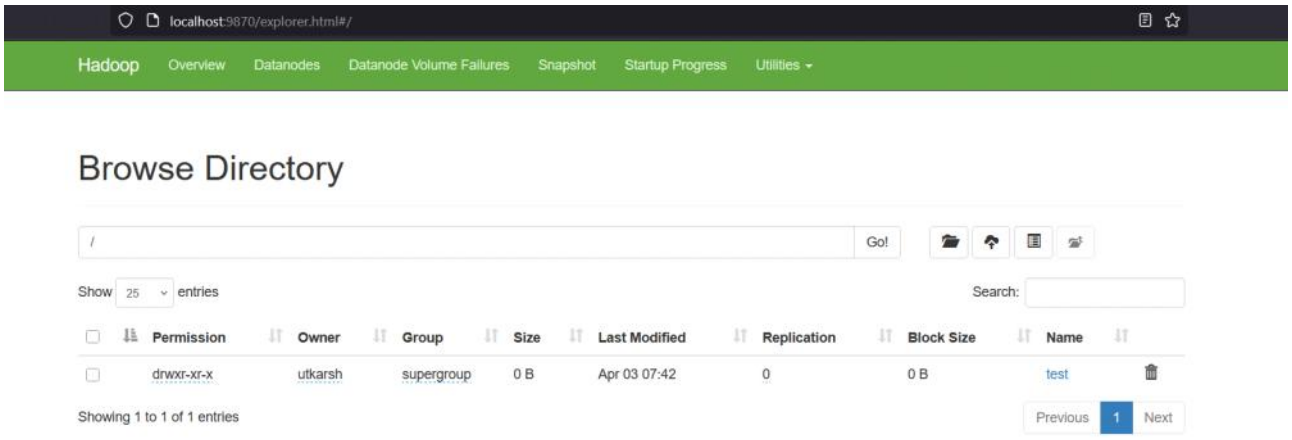
Experiment No 1



Overview 'localhost:9000' (✓active)

Started:	Thu Apr 03 07:23:43 +0530 2025
Version:	3.4.1, r4d7825309348956336b8f06a08322b78422849b1
Compiled:	Wed Oct 09 20:27:00 +0530 2024 by mthakur from branch-3.4.1
Cluster ID:	CID-a922705d-708e-43f3-8aa7-1f394516a9c6
Block Pool ID:	BP-1211890736-192.168.56.1-1743569700682



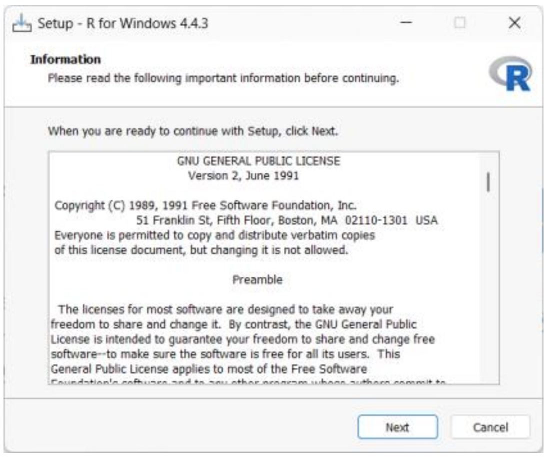
Browse Directory

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
	drwxr-xr-x	utkarsh	supergroup	0 B	Apr 03 07:42	0	0 B	test

Showing 1 to 1 of 1 entries

Experiment No 4

Install R:



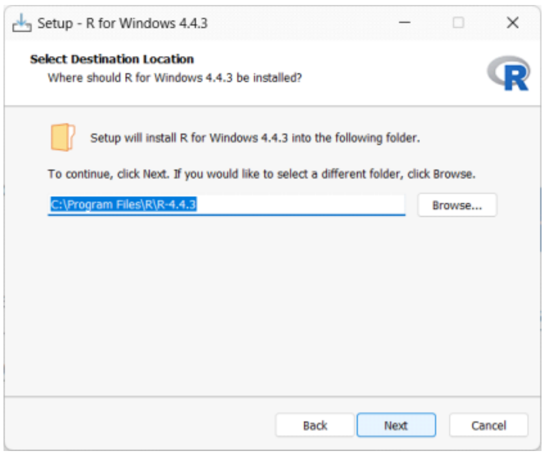
Setup - R for Windows 4.4.3

Information

Please read the following important information before continuing.

When you are ready to continue with Setup, click Next.

```
                    GNU GENERAL PUBLIC LICENSE
                       Version 2, June 1991

 Copyright (C) 1989, 1991 Free Software Foundation, Inc.
              51 Franklin St, Fifth Floor, Boston, MA  02110-1301  USA
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

                            Preamble

  The licenses for most software are designed to take away your
freedom to share and change it.  By contrast, the GNU General Public
License is intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users.  This
General Public License applies to most of the Free Software
```



Setup - R for Windows 4.4.3

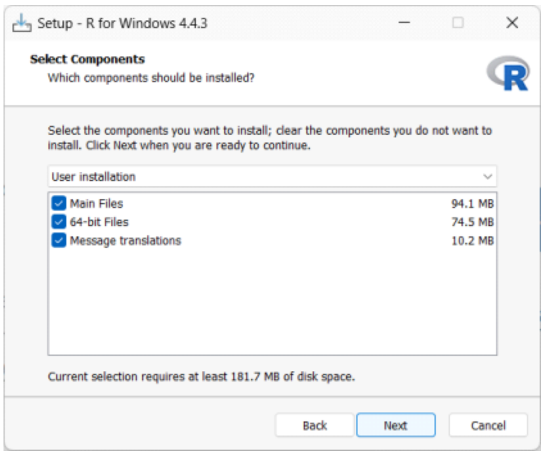
Select Destination Location

Where should R for Windows 4.4.3 be installed?

Setup will install R for Windows 4.4.3 into the following folder.

To continue, click Next. If you would like to select a different folder, click Browse.

C:\Program Files\R\R-4.4.3



Setup - R for Windows 4.4.3

Select Components

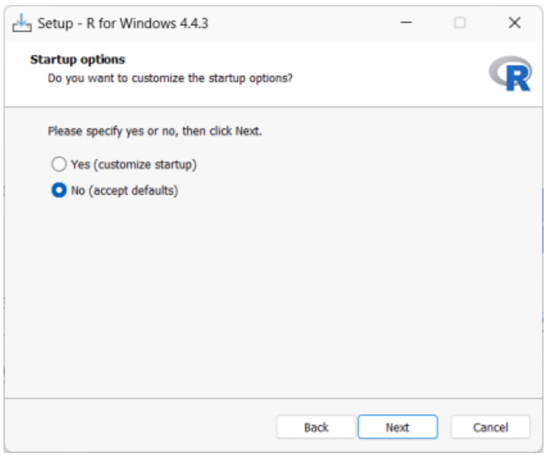
Which components should be installed?

Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

User installation

✓	Main Files	94.1 MB
✓	64-bit Files	74.5 MB
✓	Message translations	10.2 MB

Current selection requires at least 181.7 MB of disk space.



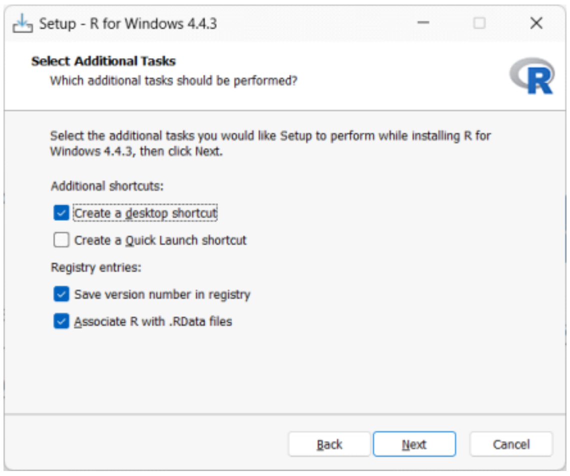
Setup - R for Windows 4.4.3

Startup options

Do you want to customize the startup options?

Please specify yes or no, then click Next.

Yes (customize startup) [✓ No (accept defaults)]



Setup - R for Windows 4.4.3

Select Additional Tasks

Which additional tasks should be performed?

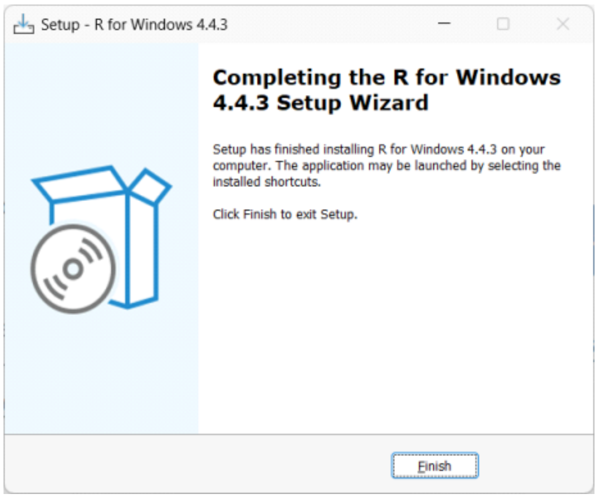
Select the additional tasks you would like Setup to perform while installing R for Windows 4.4.3, then click Next.

Additional shortcuts:

- [✓] Create a desktop shortcut
- [ ] Create a Quick Launch shortcut

Registry entries:

- [✓] Save version number in registry
- [✓] Associate R with .RData files



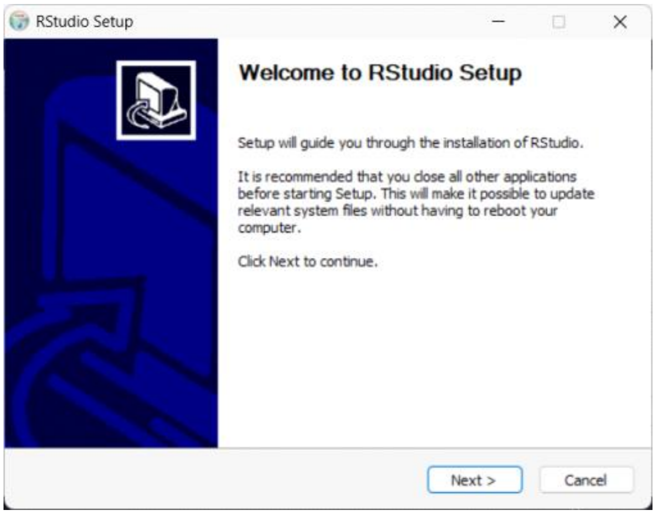
Setup - R for Windows 4.4.3

Completing the R for Windows 4.4.3 Setup Wizard

Setup has finished installing R for Windows 4.4.3 on your computer. The application may be launched by selecting the installed shortcuts.

Click Finish to exit Setup.

Install R Studio :

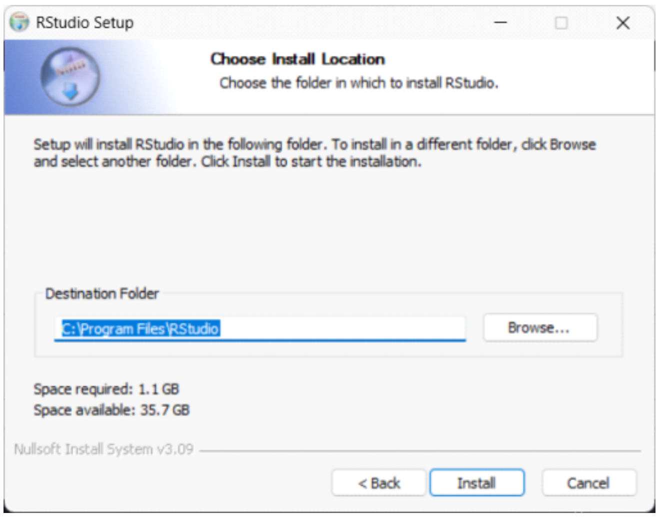

RStudio Setup

Welcome to RStudio Setup

Setup will guide you through the installation of RStudio.

It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.

Click Next to continue.


RStudio Setup

Choose Install Location

Choose the folder in which to install RStudio.

Setup will install RStudio in the following folder. To install in a different folder, click Browse and select another folder. Click Install to start the installation.

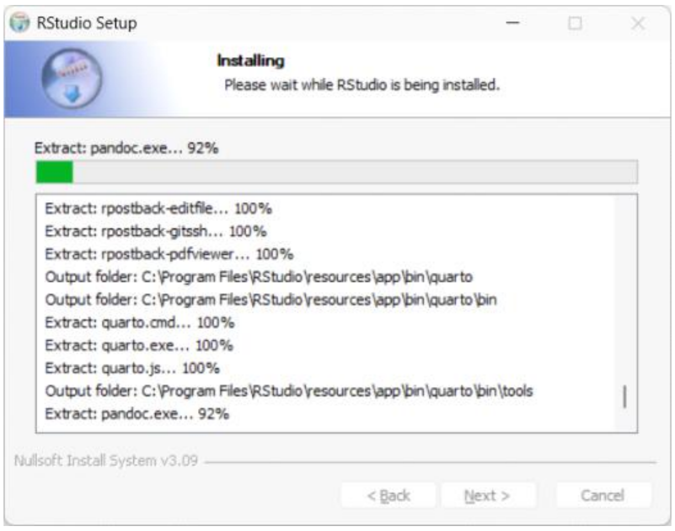
Destination Folder

C:\Program Files\RStudio

Space required: 1.1 GB

Space available: 35.7 GB

Nullsoft Install System v3.09


RStudio Setup

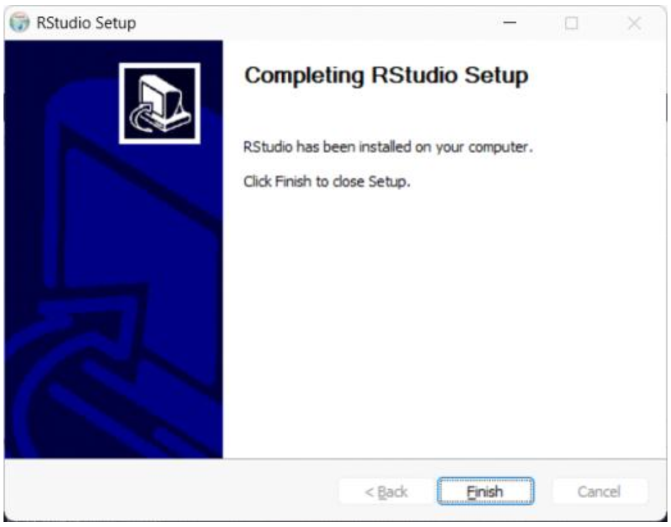
Installing

Please wait while RStudio is being installed.

```
Extract: pandoc.exe... 92%
```

```
Extract: rpostback-editfile... 100%
Extract: rpostback-gitssh... 100%
Extract: rpostback-pdfviewer... 100%
Output folder: C:\Program Files\RStudio\resources\app\bin\quarto
Output folder: C:\Program Files\RStudio\resources\app\bin\quarto\bin
Extract: quarto.cmd... 100%
Extract: quarto.exe... 100%
Extract: quarto.js... 100%
Output folder: C:\Program Files\RStudio\resources\app\bin\quarto\bin\tools
Extract: pandoc.exe... 92%
```

Nullsoft Install System v3.09


RStudio Setup

Completing RStudio Setup

RStudio has been installed on your computer.

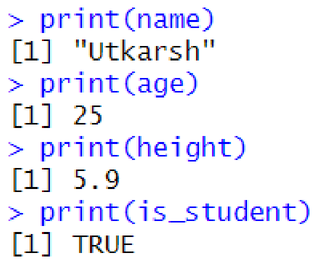
Click Finish to close Setup.

Experiment No 5

1. Declaring Variables in R

```
name <- "Utkarsh"
age <- 25
height <- 5.9
is_student <- TRUE

print(name)
print(age)
print(height)
print(is_student)
```

Output :

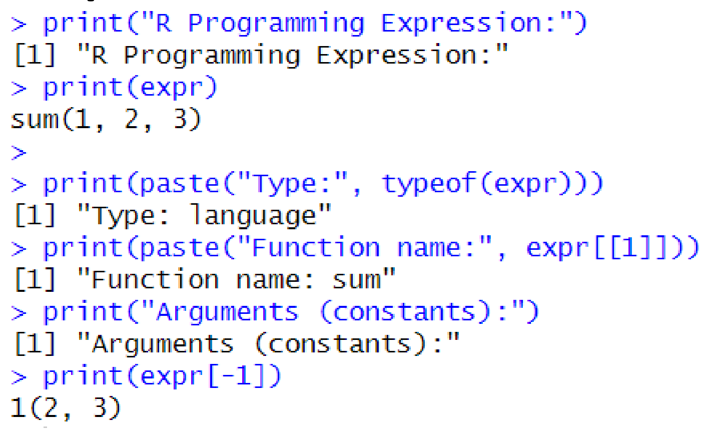


```
> print(name)
[1] "Utkarsh"
> print(age)
[1] 25
> print(height)
[1] 5.9
> print(is_student)
[1] TRUE
```

2. Expressions in R

```
r_exprs <- c("John", "Mary", "the cat")
print("Linguistic R-expressions:")
print(r_exprs)

expr <- quote(sum(1, 2, 3))

print("R Programming Expression:")
print(expr)

# Components
print(paste("Type:", typeof(expr)))       # call
print(paste("Function name:", expr[[1]]))  # sum
print("Arguments (constants):")
print(expr[-1])
```

Output:

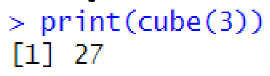


```
> print("R Programming Expression:")
[1] "R Programming Expression:"
> print(expr)
sum(1, 2, 3)
>
> print(paste("Type:", typeof(expr)))
[1] "Type: language"
> print(paste("Function name:", expr[[1]]))
[1] "Function name: sum"
> print("Arguments (constants):")
[1] "Arguments (constants):"
> print(expr[-1])
1(2, 3)
```

3.1. User-Defined Function in R

```
# User-defined function to calculate the cube of a number
cube <- function(x) {
  return(x^3)
}

print(cube(3))
```

Output :

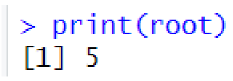


```
> print(cube(3))
[1] 27
```

3.2. Built-in Function in R

```
# Built-in function to calculate square root
num <- 25
root <- sqrt(num)

print(root)
```

Output :

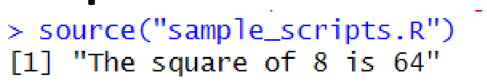


```
> print(root)
[1] 5
```

4. Scripts in R

```
# This is a simple R script
square <- function(x) {
  return(x * x)
}

number <- 8

result <- square(number)

print(paste("The square of", number, "is", result))
```

Output :



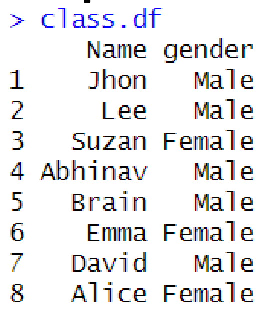
```
> source("sample_scripts.R")
[1] "The square of 8 is 64"
```

Experiment No 6

1. Build the data frame using vectors:

```
# R program to illustrate
# data frame from vector

Name <- c("Jhon", "Lee", "Suzan", "Abhinav","Brain", "Emma", "David", "Alice")
gender <- c("Male", "Male", "Female", "Male", "Male", "Female", "Male", "Female")

class.df<- data.frame( Name, gender)
class.df
```

Output :

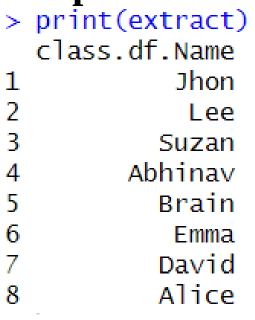


```
> class.df
     Name gender
1    Jhon   Male
2     Lee   Male
3   Suzan Female
4 Abhinav   Male
5   Brain   Male
6    Emma Female
7   David   Male
8   Alice Female
```

2. Extract Data from Data Frame

```
# R program to illustrate
# data frame from vector

Name <- c("Jhon", "Lee", "Suzan", "Abhinav",
     "Brain", "Emma", "David", "Alice")
Gender <- c("Male", "Male", "Female", "Male",
     "Male", "Female", "Male", "Female")

 extract<- data.frame(class.df$Name) print(extract)
```

Output :

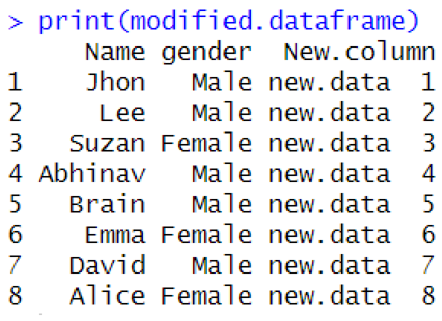


```
> print(extract)
  class.df.Name
1          Jhon
2           Lee
3         Suzan
4       Abhinav
5         Brain
6          Emma
7         David
8         Alice
```

3. Adding column:

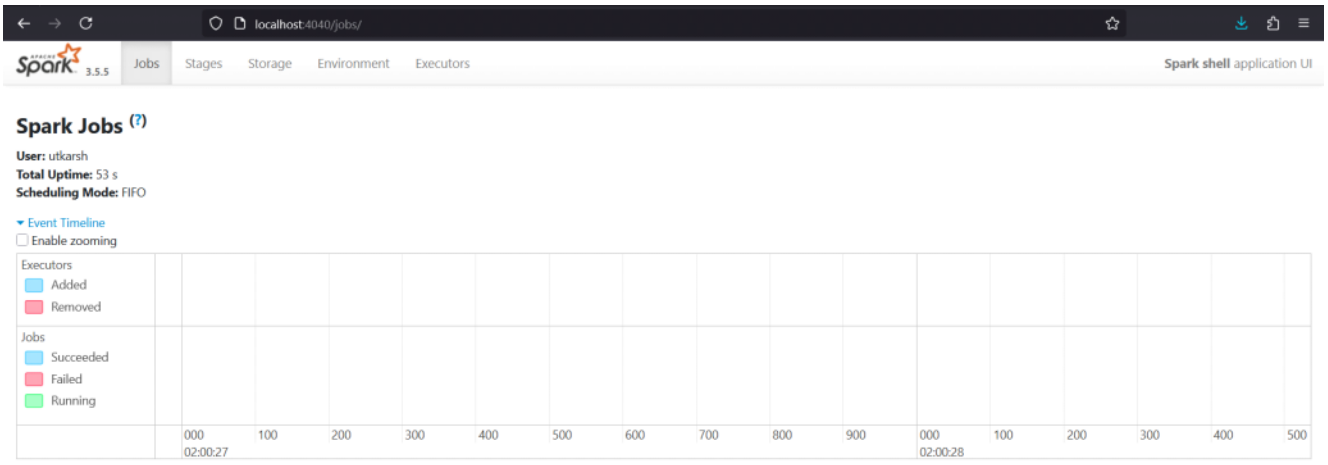
```
class.df$New.column<- sprintf("new.data % d", 1:8)
modified.dataframe <- class.df
print(modified.dataframe)
```

Output :



```
> print(modified.dataframe)
     Name gender  New.column
1    Jhon   Male new.data  1
2     Lee   Male new.data  2
3   Suzan Female new.data  3
4 Abhinav   Male new.data  4
5   Brain   Male new.data  5
6    Emma Female new.data  6
7   David   Male new.data  7
8   Alice Female new.data  8
```

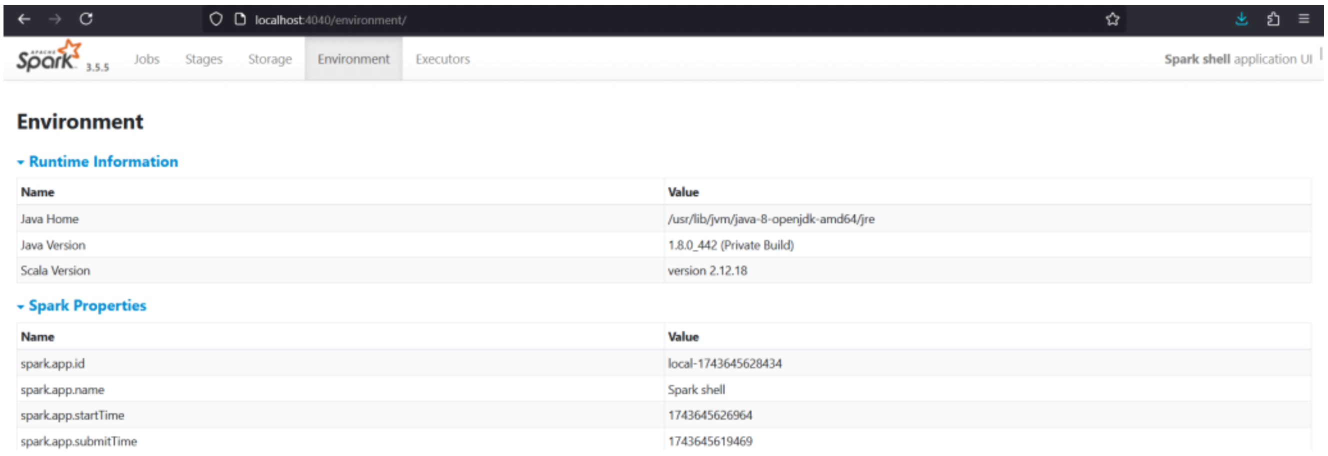
Experiment No 8



Spark Jobs (?)

User: utkarsh
Total Uptime: 53 s
Scheduling Mode: FIFO

Event Timeline
Enable zooming



Environment

Runtime Information

Name	Value
Java Home	/usr/lib/jvm/java-8-openjdk-amd64/jre
Java Version	1.8.0_442 (Private Build)
Scala Version	version 2.12.18

Spark Properties

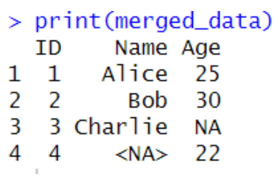
Name	Value
spark.app.id	local-1743645628434
spark.app.name	Spark shell
spark.app.startTime	1743645626964
spark.app.submitTime	1743645619469

Experiment No 10

1. Merging Datasets

```
df1 <- data.frame(ID = c(1, 2, 3), Name = c("Alice", "Bob", "Charlie"))

df2 <- data.frame(ID = c(1, 2, 4), Age = c(25, 30, 22))


# Merging datasets by the common column "ID"

merged_data <- merge(df1, df2, by = "ID", all = TRUE)

print(merged_data)
```

Output:

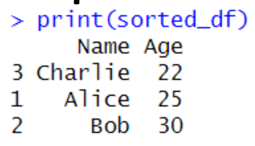


```
> print(merged_data)
  ID    Name Age
1  1   Alice  25
2  2     Bob  30
3  3 Charlie  NA
4  4    <NA>  22
```

2. Sorting Data

```
# Sorting a data frame by a column (e.g., "Age")
df <- data.frame(Name = c("Alice", "Bob", "Charlie"), Age = c(25, 30, 22))
sorted_df <- df[order(df$Age), ]
print(sorted_df)
```

Output :

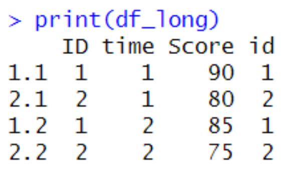


```
> print(sorted_df)
     Name Age
3 Charlie  22
1   Alice  25
2     Bob  30
```

3. Shaping Data

```
# Wide format example

df_wide <- data.frame(ID = c(1, 2), Score1 = c(90, 80), Score2 = c(85, 75))

# Reshaping from wide to long

df_long <- reshape(df_wide, varying = c("Score1", "Score2"), direction = "long", v.names = "Score")

print(df_long)
```

Output :

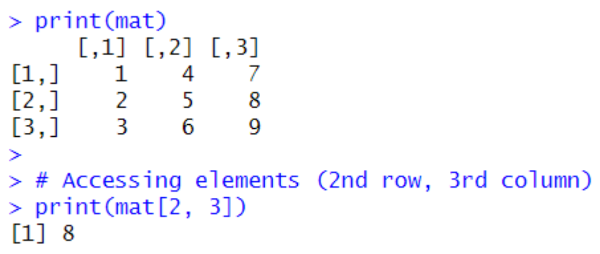


```
> print(df_long)
    ID time Score id
1.1  1    1    90  1
2.1  2    1    80  2
1.2  1    2    85  1
2.2  2    2    75  2
```

4. Managing Data with Matrices

```
# Creating a matrix
mat <- matrix(1:9, nrow = 3, ncol = 3)
print(mat)

# Accessing elements (2nd row, 3rd column)
print(mat[2, 3])
```

Output :

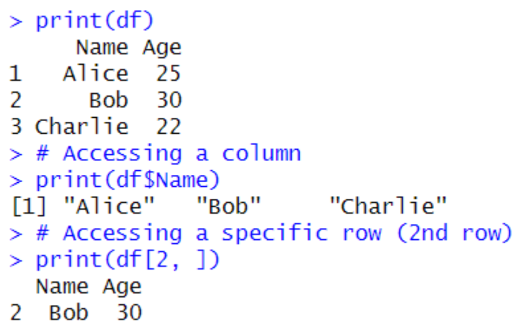


```
> print(mat)
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
>
> # Accessing elements (2nd row, 3rd column)
> print(mat[2, 3])
[1] 8
```

5. Managing Data with Data Frames

```
# Creating a data frame

df <- data.frame(Name = c("Alice", "Bob", "Charlie"), Age = c(25, 30, 22))

print(df)

# Accessing a column

print(df$Name)

# Accessing a specific row (2nd row)

print(df[2, ])
```

Output :



```
> print(df)
     Name Age
1   Alice  25
2     Bob  30
3 Charlie  22
> # Accessing a column
> print(df$Name)
[1] "Alice"    "Bob"      "Charlie"
> # Accessing a specific row (2nd row)
> print(df[2, ])
  Name Age
2  Bob  30
```