



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

DIGITAL ASSIGNMENT-#01

CSE3046-Programming for Data Science

DONE BY: D VASANTH KUMAR

REG NO: 19BDS0083



ACTIVITY-1 Installing R and R Studio in
Computer

INSTALLING R STUDIO FROM BROWSER:

Course Presurvey x Download the RStudio IDE - RSt... x


← → ↻ ⌨️ rstudio.com/products/rstudio/download/#download ☆ ⚙️ 👤

Apps Three-Phase Source... Delta and Wye 3-P... Adobe Premiere Pr... (8) Vector Calculus F... How To Hide a Viru... Preset and customi... Interstellar Main Th... Other bookmarks Reading list


users with tools to access,

RStudio Desktop 1.4.1717 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:

 **DOWNLOAD RSTUDIO FOR WINDOWS**
1.4.1717 | 156.18MB

Requires Windows 10 (64-bit)



All Installers

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

RStudio-1.4.1717.exe

System

Control Panel Home

Device Manager Remote settings System protection Advanced system settings

View basic information about your computer

Windows edition

Windows 10 Home Single Language

© Microsoft Corporation. All rights reserved.

System

Processor: Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz 2.30 GHz

Installed memory (RAM): 8.00 GB (7.81 GB usable)

System type: 64-bit Operating System, x64-based processor

Pen and Touch: No Pen or Touch Input is available for this Display

Computer name, domain, and workgroup settings

Computer name: VASANTHSWEETY

Full computer name: VASANTHSWEETY

Computer description:

Workgroup: WORKGROUP

Windows activation

Windows is activated [Read the Microsoft Software License Terms](#)

Product ID: 00327-35843-43581-AADEM [Change product key](#)

See also Security and Maintenance

Completing RStudio Setup

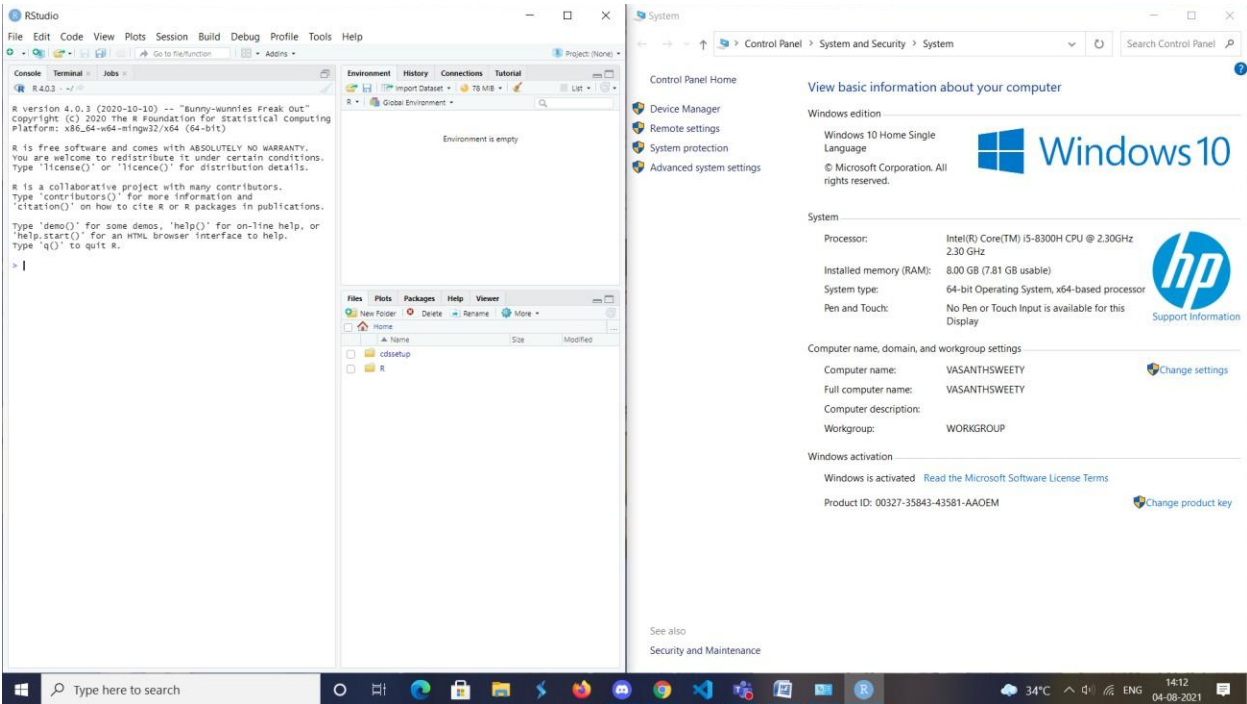
RStudio has been installed on your computer.

Click Finish to close Setup.

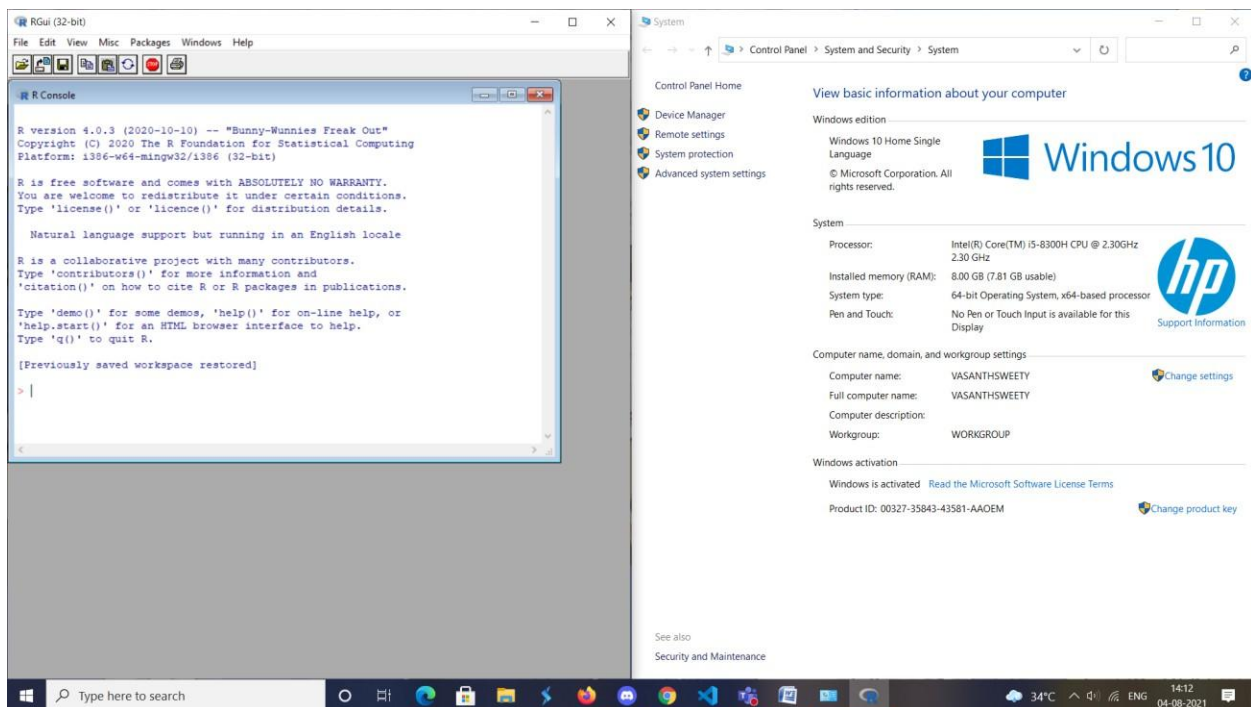
< Back Finish Cancel

Type here to search

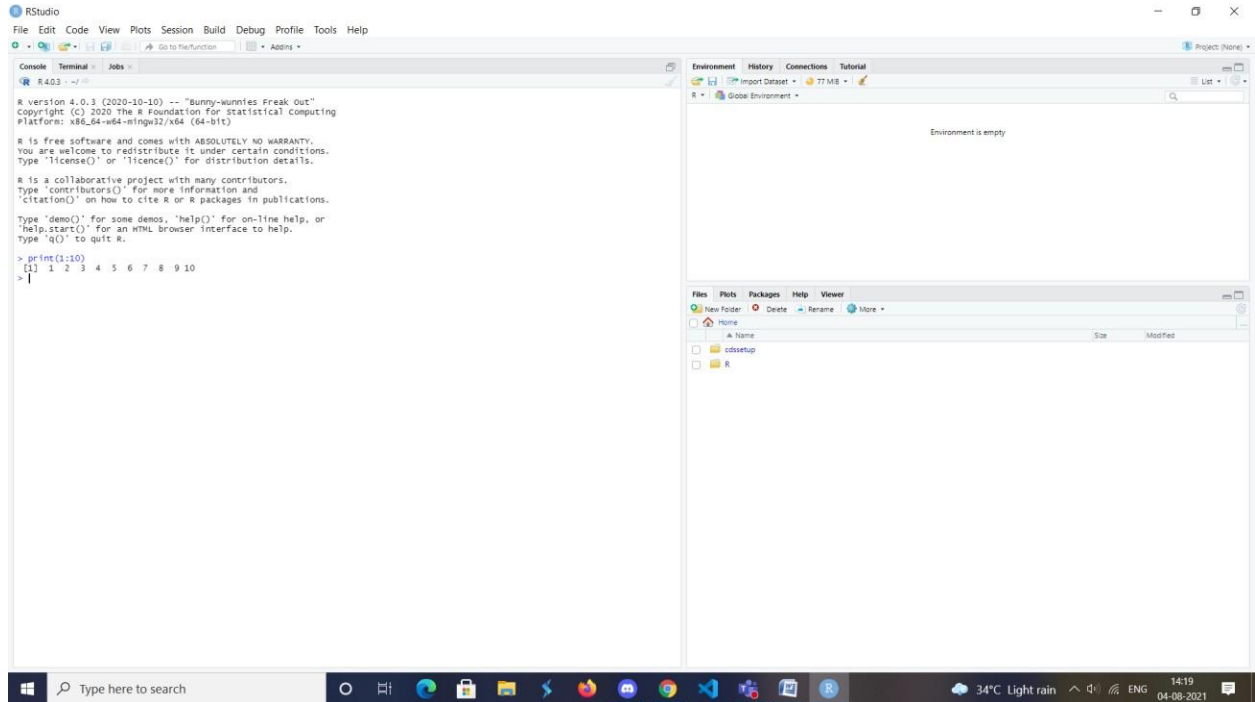
34°C ENG 14:11 04-08-2021



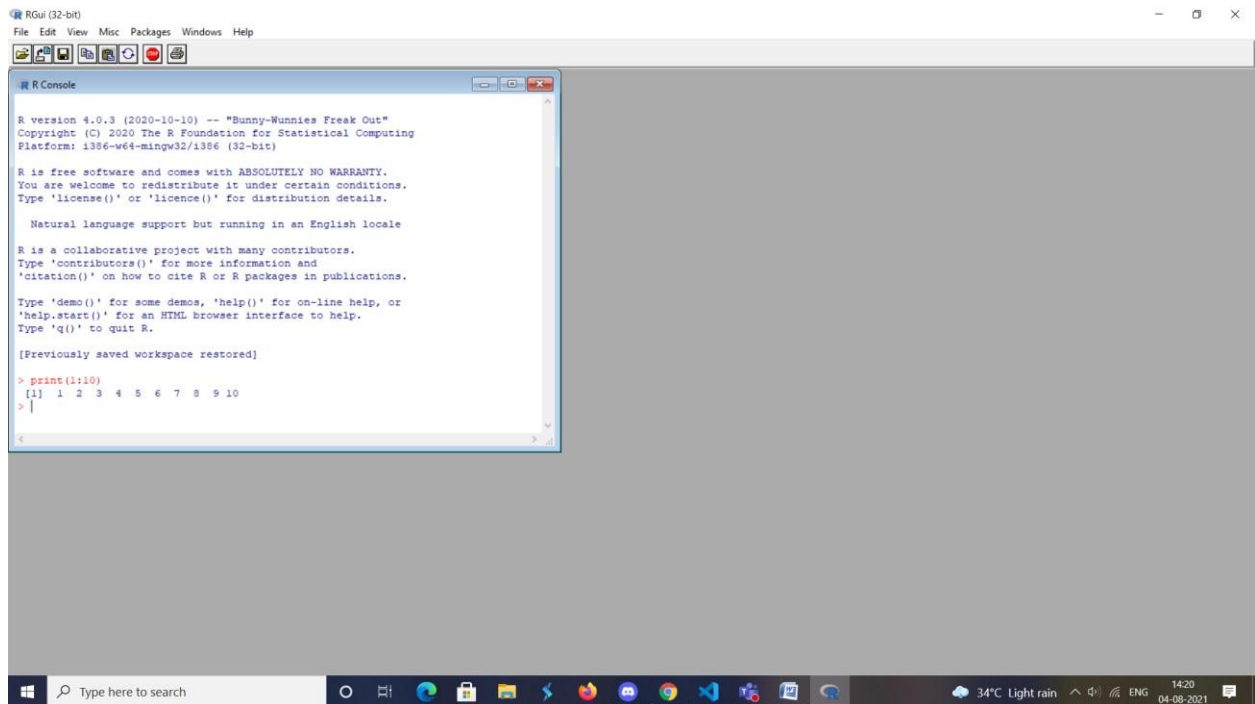
INSTALLING R IN COMPUTER:



WORKING ON R STUDIO:



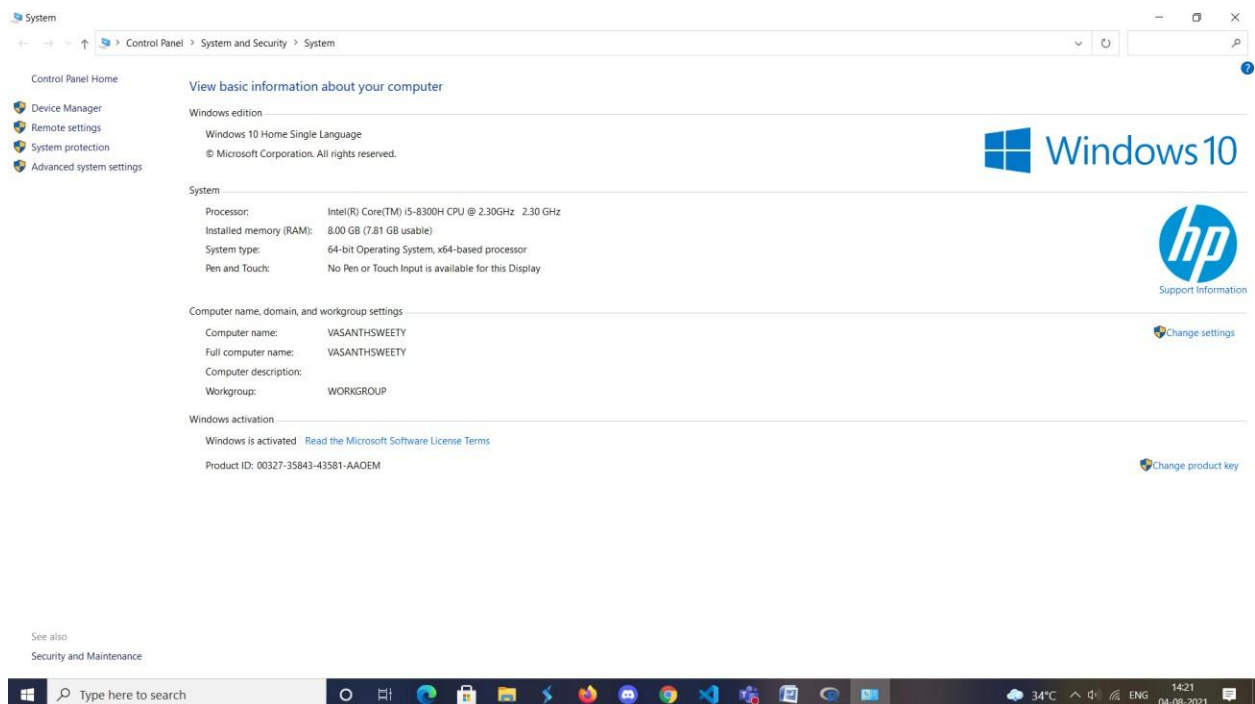
WORKING ON R:



DESKTOP IMAGE:



SYS INFO:

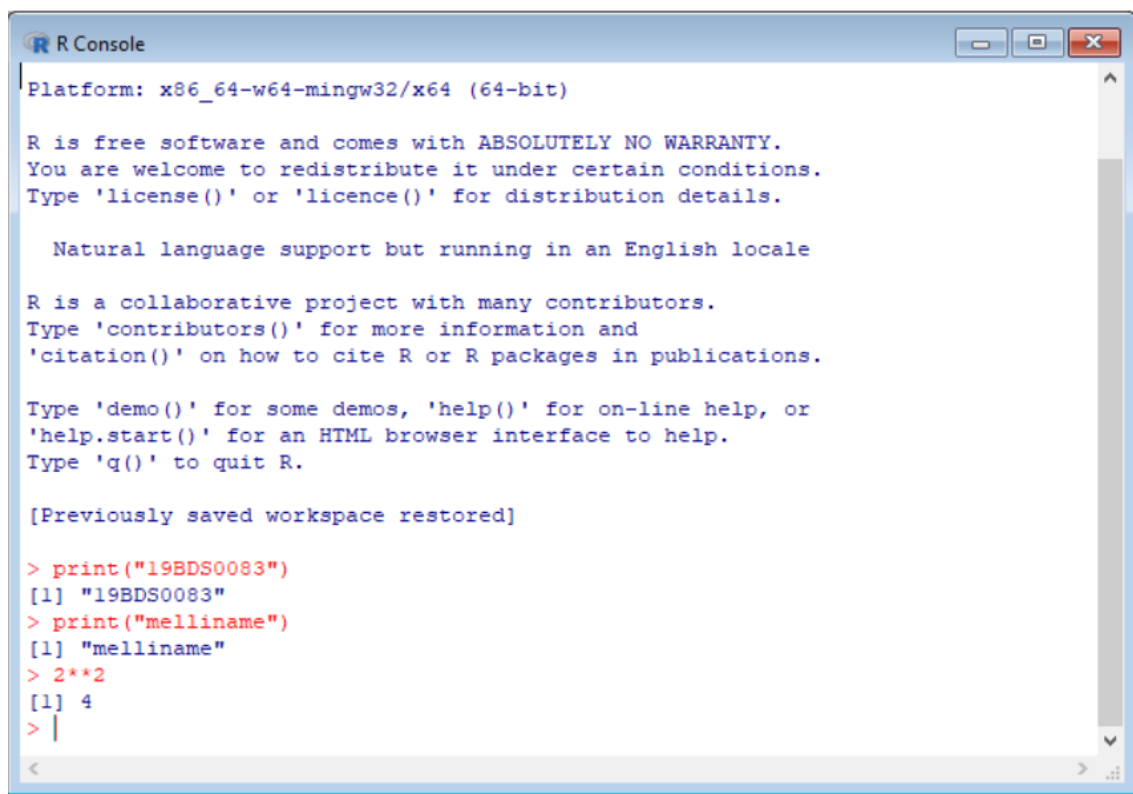


BASICS OF R(ACTIVITY-2A)

What is R?

The R programming language is an open source scripting language for predictive analytics and data visualization.

R CONSOLE (LINE BY LINE EXECUTION):



```
R Console
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

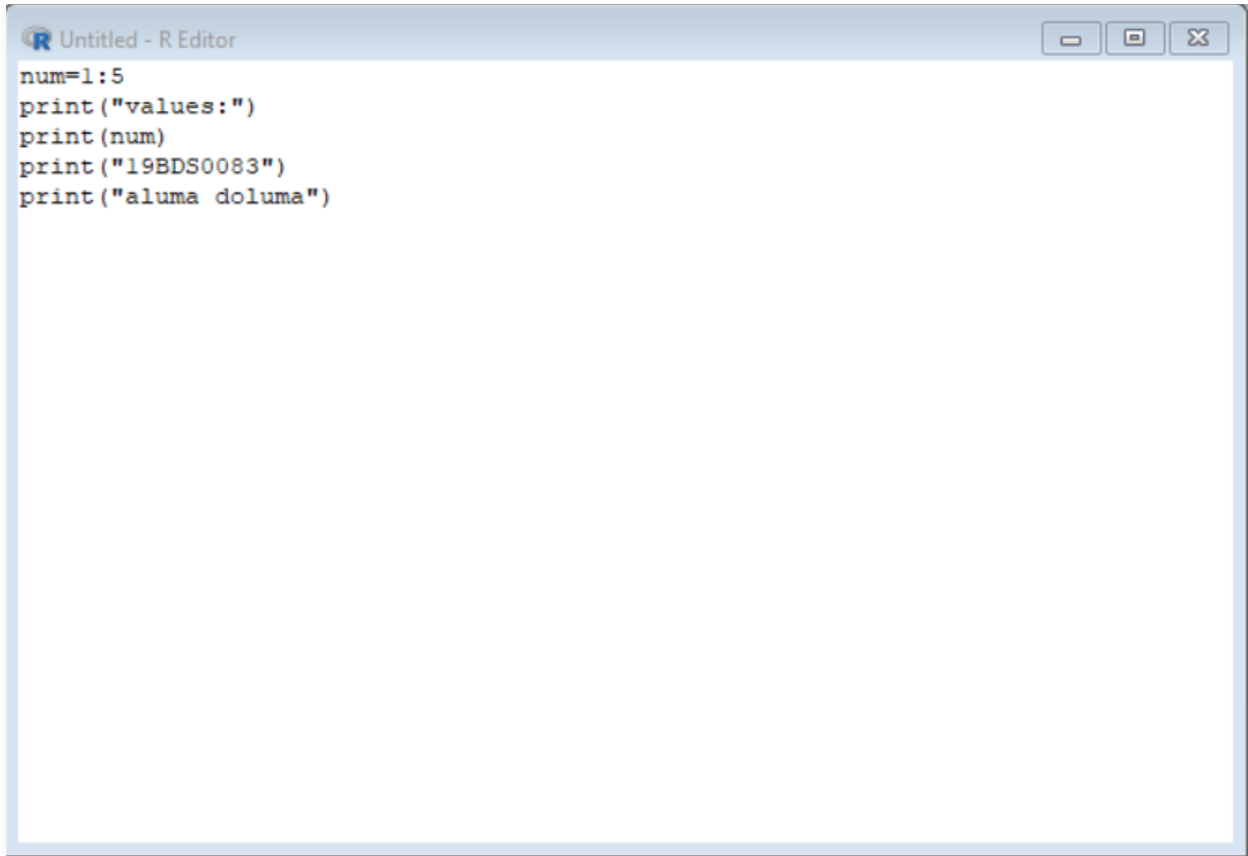
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

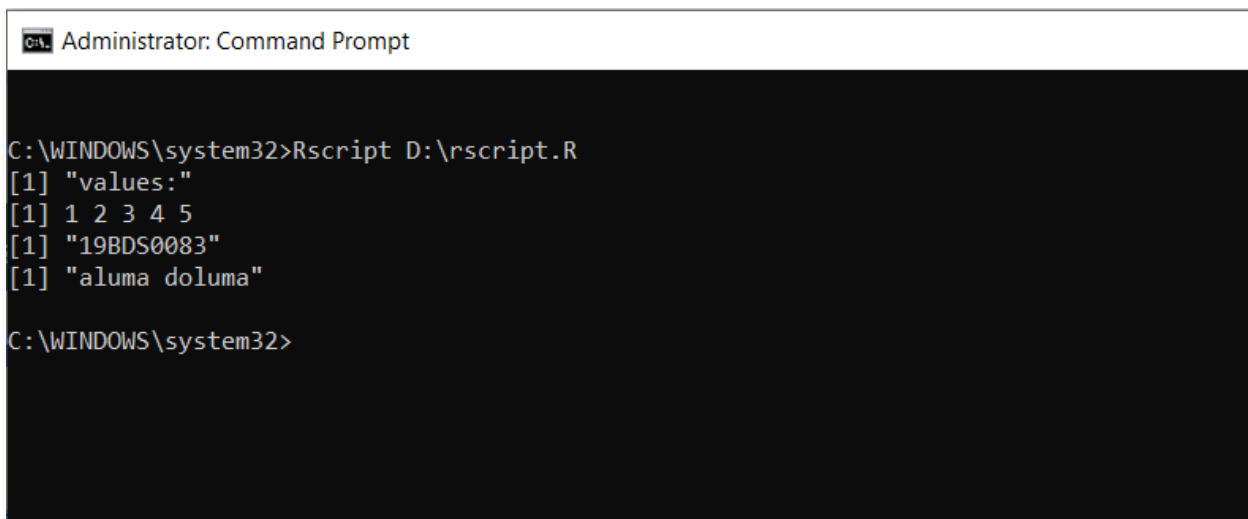
> print("19BDS0083")
[1] "19BDS0083"
> print("melliname")
[1] "melliname"
> 2**2
[1] 4
> |
```

R SCRIPT (PROGRAM WRITTEN AS SCRIPT):



```
num=1:5
print("values:")
print(num)
print("19BDS0083")
print("aluma doluma")
```

RUNNING THE SCRIPT USING CMD(TERMINAL):

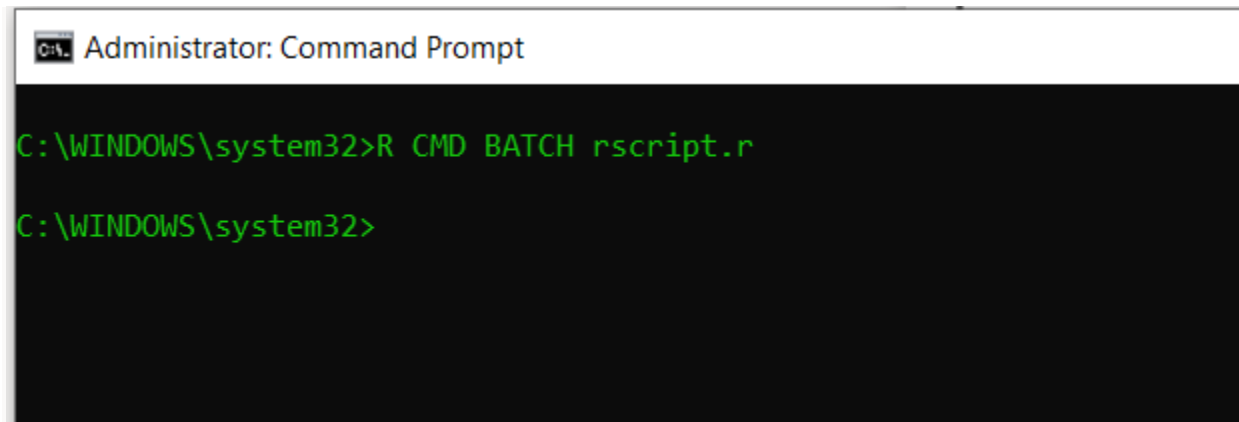


```
C:\WINDOWS\system32>Rscript D:\rscript.R
[1] "values:"
[1] 1 2 3 4 5
[1] "19BDS0083"
[1] "aluma doluma"
C:\WINDOWS\system32>
```

To run the script program we use “Rscript D:\rscript.R” command.

Log File:

To save the output as a file(.rout format)we use the command “R CMD BATCH filename .R”



```
Administrator: Command Prompt

C:\WINDOWS\system32>R CMD BATCH rscript.r

C:\WINDOWS\system32>
```

THE FILE THAT GOT CREATED:

rscript.r - Notepad
File Edit Format View Help

*R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.*

Natural Language support but running in an English locale

*R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.*

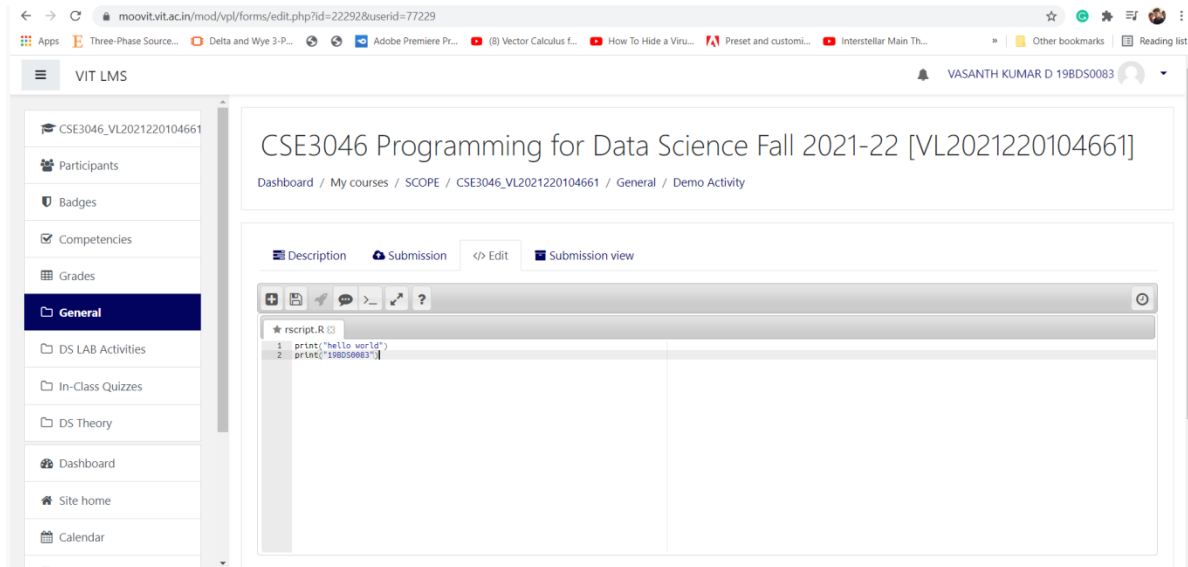
*Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.*

```
> num=1:5
> print("values:")
[1] "values:"
> print(num)
[1] 1 2 3 4 5
> print("19BDS0083")
[1] "19BDS0083"
> print("aluma doluma")
[1] "aluma doluma"
>
> proc.time()
   user  system elapsed 
  0.09   0.04   0.12
```


BASICS OF R(ACTIVITY-2B)

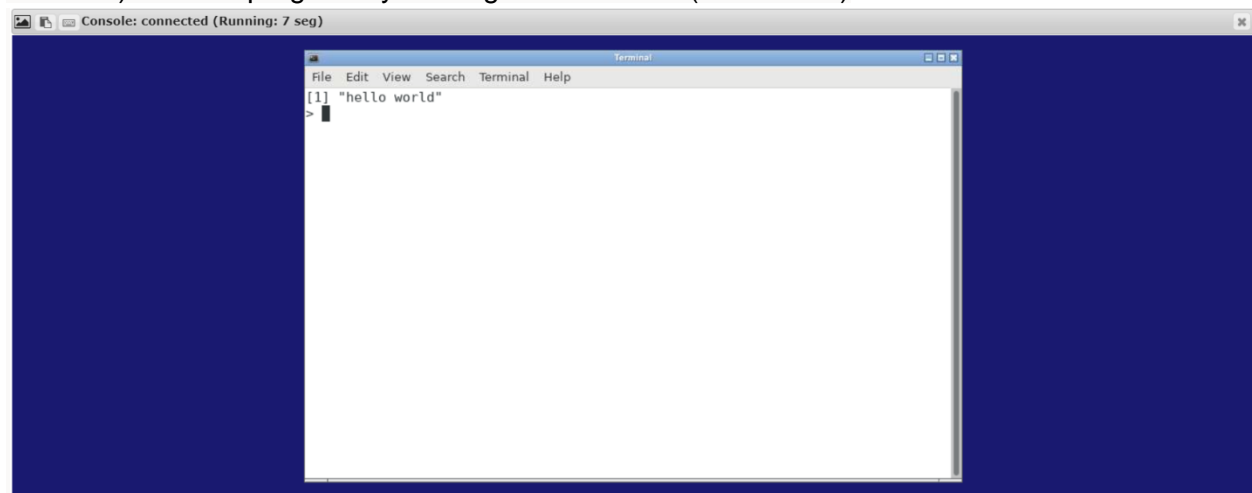
DEMO ACTIVITY (MOODLE):

SAMPLE CODE:



Procedure to run:

- 1) Create a new file with '.R' extension .
- 2) Type the code that is to be run.
- 3) Leave a blank line (by pressing enter) after writing all the code.
- 4) Save the program(ctrl+s)
- 5) Run the program by clicking the run button(rocket icon)



ACTIVITY-3

Obtaining Data with Web Scrapping using R

19BDS0083_anime_review.R:(Script file)

```
library(rvest)
```

```
library(dplyr)
```

```
require(rvest)
```

```
#install.packages("writexl")
```

```
library("writexl")
```

```
#link="https://myanimelist.net/reviews.php?t=anime"
```

```
anime_review=data.frame()
```

```
for(page_result in seq(from =1,to=15,by=1)){
```

```
  link=paste0("https://myanimelist.net/reviews.php?t=anime&p=",page_result)
```

```
  page=read_html(link)
```

```
  anime_name=page %>% html_nodes(".hoverinfo_trigger") %>% html_text()
```

```
  user_id=page %>% html_nodes("td > a") %>% html_text()
```

```
  overall_rating=page %>% html_nodes(".mb8 .spaceit+ div") %>% html_text()
```

```
  review_date=page %>% html_nodes(".mb8 div:nth-child(1)") %>% html_text()
```

```
anime_review=rbind(anime_review,data.frame(anime_name,user_id,overall_rating,review_date))

}

View(anime_review)

write_xlsx(anime_review,"V:/vit/sem5/rvest/19BDS0083_anime_review_AUG18_28.xlsx")

write.csv(anime_review,"V:/vit/sem5/rvest/19BDS0083_anime_review_AUG18_28.csv")
```

DATA SET DESCRIPTION:

The Data set of MyAnimeList consists of four columns. Each column represent different data acquired from same webpage with different information in it.

The column **anime_name** consists of the title of the anime on which reviews are being published in the website. Users mention the title of the anime and share their opinion on the anime that is mentioned. Providing reviews for such titled anime enhances popularity of the title or some bad reviews notices the quality of the anime and makes user know about it.

The column **user_id** is important data set among all other columns. In this columns the user_id of the user using this website is mentioned along with the title and review of an anime. So this information helps in identifying few people and approach them to acquire different opinions regarding the reviewed anime. This enhances interaction among the webusers of MyAnimeList.

The column **overall_rating** tells about the quality and strong base for an anime to become popular. this rating is calculated out of 10 which is the best score for depict that the anime is good at all conditions. Users provide multiple rating depending on their personal opinion and this rating is averaged and provides

overall rating for an anime. Users prefer seeing rating of an anime to watch it which is also important data.

The final column **review_date** represents the date when the review of particular anime is published. This helps us to find number of reviews is being published per day.

Link for Rscript:

<https://drive.google.com/file/d/14vOyXdB4I3DM1iXw47E0evJ6wjsQHdgN/view?usp=sharing>

CSV:

<https://drive.google.com/file/d/19-wAu5XsTbipxVdBtE3Sg2idUi0s9Ans/view?usp=sharing>

XLSX:

https://docs.google.com/spreadsheets/d/15LfZyQSJkvDDqGz85Z8MQULDLv_B6oms/edit?usp=sharing&ouid=103452182630991993013&rtpof=true&sd=true

SAMPLE DATA SET:

SAMPLE CODE:

```
1 library(rvest)
2 library(dplyr)
3
4 require(rvest)
5 #install.packages("writexl")
6 library("writexl")
7
8 #link="https://myanimelist.net/reviews.php?t=anime"
9 anime_review=data.frame()
10 for(page_result in seq(from=-1,to=15,by=1)){
11   link=paste0("https://myanimelist.net/reviews.php?t=anime&p=",page_result)
12   page=read_html(link)
13
14   anime_name=page %>% html_nodes(".hoverinfo_trigger") %>% html_text()
15   user_id=page %>% html_nodes("td > a") %>% html_text()
16   overall_rating=page %>% html_nodes(".mb8 .spaceit+ div") %>% html_text()
17   review_date=page %>% html_nodes(".mb8 div:nth-child(1)") %>% html_text()
18
19   anime_review=rbind(anime_review,data.frame(anime_name,user_id,overall_rating,review_date))
20 }
21
22 view(anime_review)
23
24 write_xlsx(anime_review,"v:/vit/sem5/rvest/19BD50083_anime_review_AUG18_28.xlsx")
25 write_csv(anime_review,"v:/vit/sem5/rvest/19BD50083_anime_review_AUG18_28.csv")
26 |
```

DATA SET:

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

anime_name	user_id	overall_rating	review_date
1 Shigatsu wa Kimi no Uso	Nibir_Kujo	Overall Rating: 8	Aug 25, 2021
2 Boku wa Tomodachi ga Sukunai Next	_hsiW	Overall Rating: 10	Aug 25, 2021
3 Tonari no Kaibutsu-kun	titogonzalez	Overall Rating: 6	Aug 25, 2021
4 Meikyuu Black Company	Circle-kun	Overall Rating: 8	Aug 25, 2021
5 Hai to Gensou no Grimgar	NewStarLabs	Overall Rating: 10	Aug 25, 2021
6 Boku no Hero Academia	AniManga_Boy07	Overall Rating: 7	Aug 25, 2021
7 Kono Sekai no Katasumi ni	Samimito	Overall Rating: 8	Aug 25, 2021
8 School Days	jossreio	Overall Rating: 10	Aug 25, 2021
9 Tsuki ga Michibiku Isekai Douchuu	olander	Overall Rating: 7	Aug 25, 2021
10 Re:Zero kara Hajimeru Isekai Seikatsu 2nd Season Part 2	oknem	Overall Rating: 7	Aug 25, 2021
11 Re:Zero kara Hajimeru Isekai Seikatsu 2nd Season	oknem	Overall Rating: 9	Aug 25, 2021
12 Re:Zero kara Hajimeru Isekai Seikatsu	oknem	Overall Rating: 8	Aug 25, 2021
13 86	oknem	Overall Rating: 9	Aug 25, 2021
14 Black Lagoon	FakeAccount1	Overall Rating: 6	Aug 25, 2021
15 Boruto: Naruto Next Generations	BosniaMan	Overall Rating: 4	Aug 25, 2021
16 Shuumatsu no Walküre	Toshito123	Overall Rating: 10	Aug 25, 2021
17 Ergo Proxy	DeltaWDunn	Overall Rating: 7	Aug 25, 2021
18 Watashi ga Motete Dousunda	giraffenanime	Overall Rating: 6	Aug 25, 2021
19 Yubisaki kara Honki no Netsujou 2: Koibito wa Shouboushi	SeaLia	Overall Rating: 1	Aug 25, 2021
20 Initial D First Stage	Patrxaa	Overall Rating: 8	Aug 25, 2021
21 Karigurashi no Arrietty	Anterbo	Overall Rating: 9	Aug 25, 2021
22 Hige wo Soru. Soshite Joshikousei wo Hirou.	RenFlexx	Overall Rating: 10	Aug 25, 2021
23 Shiguang Daliren	Zeooooo	Overall Rating: 9	Aug 25, 2021
24 Vinland Saga	Ali_Hayki	Overall Rating: 9	Aug 25, 2021
25 Tokyo Revengers	muvinco	Overall Rating: 6	Aug 25, 2021
26 Madonna (Movie)	Remjy	Overall Rating: 10	Aug 25, 2021
27

Showing 1 to 27 of 1,100 entries, 4 total columns