

MVLU COLLEGE

AIM:Creating datasets from raw data (text files, CSV files, Excel sheets) and importing data into SAS/SPSS/R.

Importing data set:

The screenshot shows the RStudio interface. The top-left pane (Console) displays the output of installing the 'Rcpp' package for Excel Import. The output indicates that the package was successfully installed. The top-right pane (Environment) shows that the environment is empty. The bottom-left pane (Files) shows the file explorer with various folders and files.

The bottom-right pane (Terminal) shows the 'Import Excel Data' dialog box. The 'File/URL' field is set to 'C:\Users\mvlu\OneDrive\Desktop\ankita tiwari\Data Analysis\student data.xls'. The 'Data Preview' table shows the following data:

ID (double)	Age (double)	Study Hours (double)	Attendance (%) (double)	Score (%) (double)
1	17	2.5	85	76
2	16	3.0	90	82
3	17	1.0	70	54
4	18	4.0	95	88
5	16	2.0	80	72
6	17	3.5	92	85
7	18	1.5	75	60
8	17	2.8	88	78
9	16	0.5	65	40
10	18	4.5	98	93
11	17	3.2	87	81
12	16	2.7	82	74
13	18	1.2	68	55

The 'Import Options' section shows the following settings:

- Name: student_data.xls
- Sheet: Default
- Range: A1:D10
- Max Rows: (empty)
- Skips: 0
- NA: (empty)
- ☒ First Row as Names
- ☒ Open Data Viewer

The 'Code Preview' section shows the following R code:

```
library(readxl)
student_data_xls <- read_excel("C:/Users/mvlu/OneDrive/Desktop/ankita tiwari/Data Analysis/student data.xls")
view(student_data_xls)
```

The 'Import' button is highlighted.

ANKITA TIWARI

SYCS

S122

Data Analysis with SAS / SPSS / R PRACTICAL NO.2

MVLU COLLEGE

The screenshot displays the RStudio interface. The main window shows a data frame named 'student_data_xls' with 20 observations and 5 variables. The variables are ID, Age, Study Hours, Attendance (%), and Score (%). The console shows the R version 4.1.2 (2021-11-01) and the 'readxl' package loaded. The file explorer on the right shows the 'student_data_xls' file in the 'Global Environment'.

ID	Age	Study Hours	Attendance (%)	Score (%)
1	17	2.5	85	76
2	16	3.0	90	82
3	17	1.0	70	54
4	18	4.0	95	88
5	16	2.0	80	72
6	17	3.5	92	85
7	18	1.5	75	60
8	17	2.8	88	78
9	16	0.5	65	40
10	18	4.5	98	93
11	17	3.2	87	81
12	16	2.7	82	74
13	18	1.2	68	55
14	17	3.8	94	89
15	16	2.3	79	69
16	18	4.2	97	91
17	17	1.8	72	58
18	16	3.6	90	84
19	18	2.1	77	66
20	17	3.9	96	90

```
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
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.

> library(readxl)
Warning message:
package 'readxl' was built under R version 4.1.3

> student_data_xls <- read_excel("C:/Users/mvlu/OneDrive/desktop/ankita tiwari/Data Analysis/student dat
a.xls.xlsx")
> view(student_data_xls)
> |
```

ANKITA TIWARI

SYCS

S122

Data Analysis with SAS / SPSS / R PRACTICAL NO.2

MVLU COLLEGE

CSV:

The screenshot displays the RStudio interface during the import of a CSV file. The console shows the following code and output:

```
R - R 4.1.2 - ~/R
> library(readxl)
Warning message:
package 'readxl' was built under R version 4.1.3
> student_data_xls <- read_excel("C:/Users/mvlu/OneDrive/Desktop/ankita tiwari/Data Analysis/student d
ata.xls.xlsx")
> view(student_data_xls)
> view(student_data_xls)
> view(student_data_xls)
> |
```

The Environment pane shows the following data frames:

- student_data_xls: 20 obs. of 5 variables
- Student.Mental.health: 101 obs. of 11 variables

The Data pane shows the following data frame:

Timestamp	Choose your gender	Age	What is your course	Your current year of Study	What is your CGPA	Marital status
08-07-2020 12:02	Female	18	Engineering	year 1	3.00 - 3.49	No
08-07-2020 12:04	Male	21	Islamic education	year 2	3.00 - 3.49	No
08-07-2020 12:05	Female	19	BIT	Year 1	3.00 - 3.49	No
08-07-2020 12:06	Female	22	Laws	Year 3	3.00 - 3.49	Yes
08-07-2020 12:13	Male	23	Mathematics	year 4	3.00 - 3.49	No
08-07-2020 12:31	Male	19	Engineering	Year 2	3.50 - 4.00	No
08-07-2020 12:32	Female	23	Pendidikan Islam	year 2	3.50 - 4.00	Yes
08-07-2020 12:33	Female	18	BCS	year 1	3.50 - 4.00	No
08-07-2020 12:35	Female	19	Human Resources	Year 2	2.50 - 2.99	No

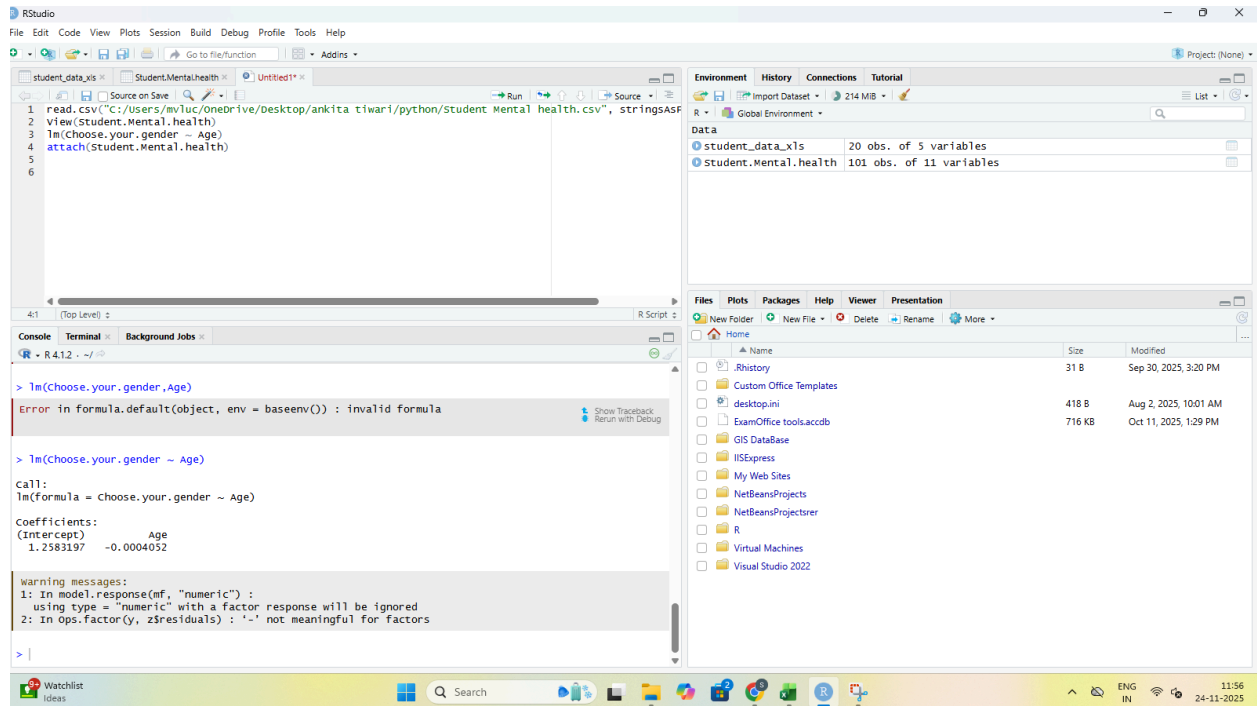
ANKITA TIWARI

SYCS

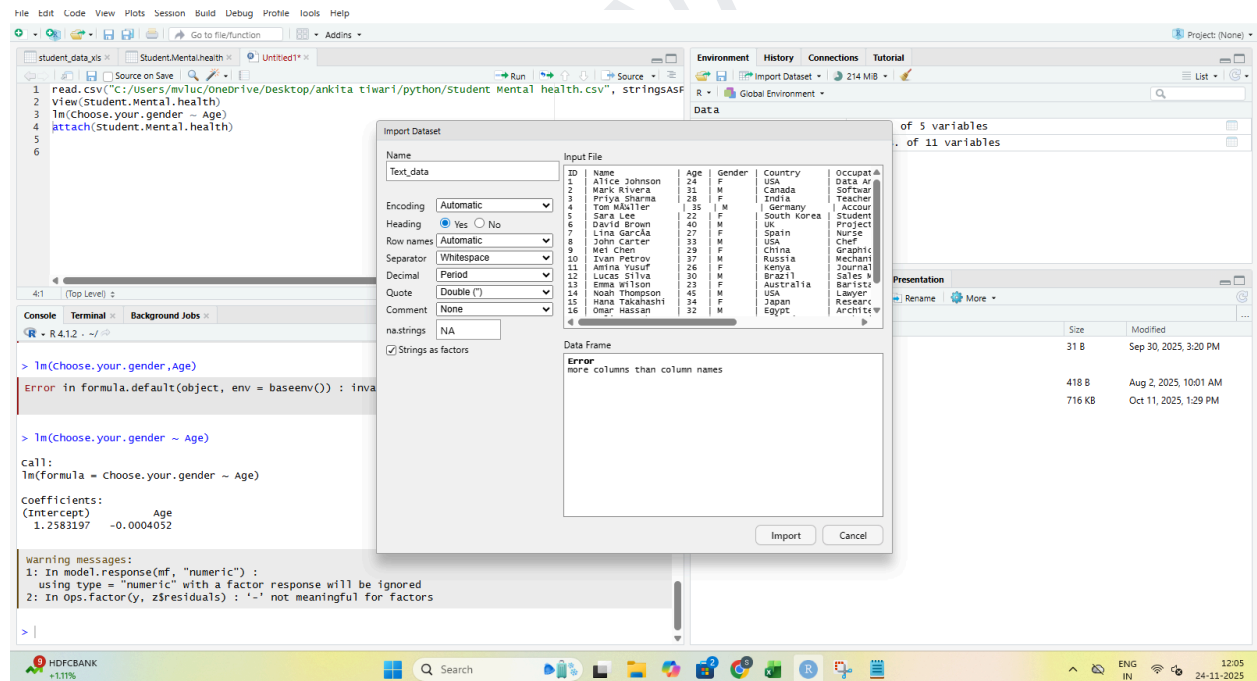
S122

Data Analysis with SAS / SPSS / R PRACTICAL NO.2

MVLU COLLEGE



Text:



ANKITA TIWARI

SYCS

S122

Data Analysis with SAS / SPSS / R PRACTICAL NO.2

MVLU COLLEGE

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for loading and analyzing data. The code includes reading a CSV file, creating a linear model, and attaching a text file.
- Environment:** Lists loaded objects: `student_data.xls` (20 obs. of 5 variables), `Student.Mental.health` (101 obs. of 11 variables), and `Text_data` (20 obs. of 8 variables).
- Console:** Shows the execution of the code, including the linear model call and coefficients.
- Files:** Displays the file explorer with various system and project files.

```
1 read.csv("C:/Users/mvlu/OneDrive/Desktop/ankita tiwari/python/Student Mental health.csv", stringsAsFactors = FALSE)
2 view(Student.Mental.health)
3 lm(Choose.your.gender ~ Age)
4 attach(Student.Mental.health)
5
6 read.delim("C:/Users/mvlu/OneDrive/Desktop/ankita tiwari/data Analysis/Text_data.txt", stringsAsFactors = FALSE)
7 view(Text_data)
8 lm(Gender ~ Age)
9 attach(Text_data)
10
```

Console Output:

```
15 15 Hana Takahashi 34 F Japan Researcher 72000 Studies AI ethics
16 16 Omar Hassan 32 M Egypt Architect 68000 Designs eco-buildings
17 17 Julia Novak 27 F Poland Marketing Spec. 41000 Social media expert
18 18 Miguel Torres 39 M Mexico Mechanic 30000 Restores classic cars
19 19 Zoe Martin 21 F France Student 0 Learning programming
20 20 Ethan Wright 36 M New Zealand Pilot 95000 Loves photography

> view(Text_data)
> lm(Gender ~ Age)

Call:
lm(formula = Gender ~ Age)

Coefficients:
(Intercept)      Age
-0.38352      0.06086

Warning messages:
1: In model.response(mf, "numeric") :
  using type = "numeric" with a factor response will be ignored
2: In Ops.factor(y, z$residuals) : '-' not meaningful for factors

> attach(Text_data)

The following objects are masked from Text_data (pos = 3):
```

ANKITA TIWARI

SYCS

S122

Data Analysis with SAS / SPSS /R PRACTICAL NO.2