

MVLU COLLEGE

Aim:

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

EXCEL

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.

[workspace loaded from ~/.RData]

> |

Import Excel Data

File/URL:
C:/Users/mvlu/OneDrive/Desktop/ANSHRAH/DATA ANALYSIS WITH SASS SPSS R/CHARACTERS_DATASET.xlsx

Data Preview:

NAME (character)	INITIAL_AGE (double)	AGE (double)	GENDER (character)	GENDER_NUMBER (double)	ROLE (character)	ALIGNMENT (character)	COMBAT_STYLE (character)	STATUS (character)	ARCHETYPE (logical)	ARC_TYPE (logical)
NEVEAH ELIXIR	9	16	Female	1	Protagonist	Good	Ranged	Alive	N/A	N/A
DAMIAN ESTEBETH	9	16	Male	0	Protagonist	Good	Close	Alive	N/A	N/A
ELIAS EVANDER	9	16	Male	0	Protagonist	Good	Stealth	Alive	N/A	N/A
CALE WERENUS	9	16	Male	0	Supporting	Neutral	Close	Alive	N/A	N/A
ADEN WERENUS	28	35	Male	0	Supporting	Neutral	Close	Alive	N/A	N/A
IDIRIA WERENUS	30	37	Female	1	Supporting	Neutral	None	Alive	N/A	N/A
VICTORIA CASABEL	10	17	Female	1	Supporting	Good	Stealth	Alive	N/A	N/A
JULIAN CASABEL	18	25	Male	0	Supporting	Good	Ranged	Missing	N/A	N/A
JULIA CASABEL	39	46	Female	1	Supporting	Good	Close	Dead	N/A	N/A
VECTOR CASABEL	41	48	Male	0	Antagonist	Evil	Stealth	Alive	N/A	N/A
AXEL	9	16	Male	0	Supporting	Good	Stealth	Alive	N/A	N/A
RAIDENCEVEL EVEVORY	9	16	Female	1	Supporting	Good	Stealth	Alive	N/A	N/A
RENOX EVEVORY	14	21	Male	1	Supporting	Good	Ranged	Alive	N/A	N/A

Import Options:

Name: CHARACTERS_DATASET Max Rows: ☒ First Row as Names
Sheet: Default Skip: ☒ Open Data Viewer
Range: A1:D10 NA:

Code Preview:

```
library(readxl)  
CHARACTERS_DATASET <- read_excel("C:/Users/mvlu/OneDrive/Desktop/ANSHRAH/DATA ANALYSIS WITH  
SASS SPSS R/CHARACTERS_DATASET.xlsx")  
View(CHARACTERS_DATASET)
```

Import Cancel

Visual Studio 2022

ANSHRAH SHAIKH

S111

SYCS

DATA ANALYSIS IN SASS / SPSS / R

MVLU COLLEGE

The screenshot displays the RStudio environment with the following components:

- Environment:** Shows the loaded dataset `CHARACTERS_DATASET` with 33 observations and 11 variables.
- Files:** Lists various files in the project, including `RData`, `Rhistory`, `class printString.java`, `Computer Network prat1.docx`, `Custom Office Templates`, `Deep GIS Prac.docx`, `desktop.ini`, `GIS DataBase`, `GIS Prac.docx`, `ISExpress`, `My Web Sites`, `NetBeansProjects`, `Office Tools Practical_1.accd`, `R`, `Virtual Machines`, and `Visual Studio 2022`.
- Console:** Displays the R version (4.1.2), copyright information, and the execution of the following code:

```
8  
9 read_excel("C:/Users/mvluc/OneDrive/Desktop/ANSHRAH/DATA ANALYSIS WITH SASS SPSS R/CHARACTERS_DATASET.XLSX")  
10 view(CHARACTERS_DATASET)  
11 lm(AGE~GENDER)  
12 attach(CHARACTERS_DATASET)
```
- Terminal:** Shows the output of the linear model:

```
> attach(CHARACTERS_DATASET)  
> lm(AGE~GENDER)  
  
call:  
lm(formula = AGE ~ GENDER)  
  
Coefficients:  
(Intercept)  GENDERMale  
33.786      -6.154
```

ANSHRAH SHAIKH

S111

SYCS

DATA ANALYSIS IN SASS / SPSS / R

MVLU COLLEGE

CSV:

The screenshot displays the RStudio interface. The 'Import Dataset' dialog box is open, showing the 'student_spending' dataset. The 'Input File' field contains a long file path. The 'Encoding' is set to 'Automatic', 'Heading' is 'Yes', 'Row names' is 'Automatic', 'Separator' is 'Comma', 'Decimal' is 'Period', 'Quote' is 'Double (")', 'Comment' is 'None', and 'na.strings' is 'NA'. The 'Strings as factors' checkbox is checked. The 'Data Frame' preview shows the first 15 rows of the dataset, with columns: X, age, gender, year_in_school, major, and a numeric variable (likely monthly_income).

X	age	gender	year_in_school	major	
0	19	Non-binary	Freshman	Psychology	
1	24	Female	Junior	Economics	1
2	24	Non-binary	Junior	Economics	
3	23	Female	Senior	Computer Science	
4	20	Female	Senior	Computer Science	
5	25	Non-binary	Sophomore	Computer Science	
6	23	Female	Freshman	Engineering	1
7	23	Female	Junior	Economics	
8	22	Non-binary	Senior	Computer Science	1
9	18	Female	Junior	Computer Science	1
10	23	Male	Junior	Biology	
11	25	Male	Freshman	Engineering	1
12	21	Male	Sophomore	Economics	
13	19	Female	Junior	Biology	1
14	24	Non-binary	Freshman	Economics	1
15	22	Non-binary	Sophomore	Computer Science	1

The RStudio console shows the following code and output:

```
call: lm(formula = AGE ~ GENDER)
```

Coefficients:

ANSHRAH SHAIKH

S111

SYCS

DATA ANALYSIS IN SASS / SPSS / R

MVLU COLLEGE

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R script:

```
1 read.csv("C:/Users/mvlu/OneDrive/Desktop/ANSHRAH/ADVANCE PYTHON FOR DATA SCIENCE/student_spending.csv", stringsAsFactors=TRUE)
2 view(student_spending)
3 lm(age~gender)
4 attach(student_spending)
```
- Console:** Shows the execution of the script:

```
> attach(student_spending)
> lm(age~gender)

Call:
lm(formula = age ~ gender)

Coefficients:
(Intercept)      genderMale  genderNon-binary 
      21.59752           0.07102           0.16260
```
- Environment:** Lists loaded objects: `CHARACTERS_DA` (33 obs. of 11 variables) and `student_spending` (1000 obs. of 18 variables).
- Files:** Shows the file explorer with various project files.

Text:

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R script:

```
1 read.csv("C:/Users/mvlu/OneDrive/Desktop/ANSHRAH/ADVANCE PYTHON FOR DATA SCIENCE/text_data.csv", stringsAsFactors=TRUE)
2 view(text_data)
3 lm(AGE~GENDER)
```
- Console:** Shows the execution of the script:

```
> attach(text_data)
> lm(AGE~GENDER)

Call:
lm(formula = AGE ~ GENDER)

Coefficients:
```
- Environment:** Lists loaded objects: `CHARACTERS_DA` (33 obs. of 11 variables), `student_spending` (1000 obs. of 18 variables), and `text_data` (20 obs. of 10 variables).
- Files:** Shows the file explorer with various project files.

ANSHRAH SHAIKH

S111

SYCS

DATA ANALYSIS IN SASS / SPSS / R

MVLU COLLEGE

The screenshot displays the RStudio environment with the following components:

- Script Editor:** Contains R code for loading and analyzing data. The code includes reading CSV and Excel files, attaching datasets, and running a linear model.
- Environment Pane:** Shows the loaded datasets:
 - `CHARACTERS_DATA`: 33 obs. of 11 variables
 - `student_spending`: 1000 obs. of 18 variables
 - `text_data`: 20 obs. of 10 variables
- Console:** Displays the output of the executed commands, including the linear model results.

```
1 read.csv("C:/Users/mvluc/OneDrive/Desktop/ANSHRAH/ADVANCE PYTHON FOR DATA SCIENCE/student_spending.csv", stringsAsFactors=TRUE)
2 view(student_spending)
3 lm(age~gender)
4 attach(student_spending)
5
6
7
8
9 read_excel("C:/Users/mvluc/OneDrive/Desktop/ANSHRAH/DATA ANALYSIS WITH SASS SPSS R/CHARACTERS_DATASET.xlsx")
10 View(CHARACTERS_DATASET)
11 lm(AGE~GENDER)
12 attach(CHARACTERS_DATASET)
13
14
15
16 read.csv("C:/Users/mvluc/OneDrive/Desktop/ANSHRAH/DATA ANALYSIS WITH SASS SPSS R/text_data.txt", stringsAsFactors=TRUE)
17 View(text_data)
18 lm(Age~Gender)
19 attach(text_data)
```

Console Output:

```
R - R 4.1.2 ~- /
> view(text_data)
> attach(text_data)
> lm(Age~Gender)

Call:
lm(formula = Age ~ Gender)

Coefficients:
(Intercept)      GenderM 
        27.2           4.2
```

ANSHRAH SHAIKH

S111

SYCS

DATA ANALYSIS IN SASS / SPSS / R