# Module 2: Reading data files in R

created by

R Team, FOSSEE, IIT Bombay

and

contributed by

Mr. Digvijay Singh, Project Research Associate, R Team, FOSSEE, IIT Bombay Mrs. Usha Viswanathan, Sr. Project Manager, FOSSEE, IIT Bombay

under the guidance of

Prof. Radhendushka Srivastava,

Mathematics Department, IIT Bombay

24 October 2022

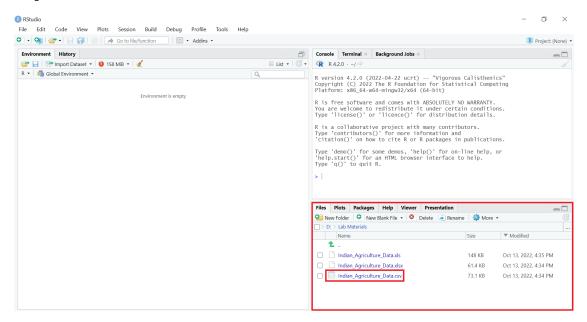
### Steps to read data files in R

This module explains the methods for importing data from a file present in the working directory to the R software. Before importing data, set the folder containing the data file as the working directory by following the instructions from Module 1.

Data from a .csv, .xlsx or .xls file can be read in R. Let us read data from a .csv file in R.

## **Using RStudio**

**Step 1.** Left-click on the name of the data file in the **Files** tab of the **Plots and Files** window.



*Figure 1: Left-click on the name of the data file to be read.* 

**Step 2.** Select the option **Import Dataset...** from the dropdown menu.

▼ Q Go to file/function Addins ▼					Project: (No		
nvironment History	Ð	Console Terminal ×	Background Jobs ×		_		
🚰 📊 📑 Import Dataset 🔹 🚯 158 MiB 👻 🔏	≡ List •   C •	🗣 R 4.2.0 · ~/ 🕫					
<ul> <li>Global Environment </li> <li>Environment is empty</li> </ul>	Q	<pre>R version 4.2.0 (2022-04-22 ucrt) "'vigorous Calisthenics" Copyright (C) 2022 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. 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.stat()' for an HML browser interface to help. Type 'q()' to quit R.</pre>					
		Files Plots Packag	ges Help Viewer Presentat	ion	_		
		💁 New Folder 🛛 🕯	vew Blank File 🔹 🤒 Delete 🕞	Rename   🎲 More			
		> D: > Lab Materials					
				Size	Modified		
		Name					
		<b>t</b>					
		1	-	148 KB	Oct 13, 2022, 4:35 PM		
		1	ulture_Data.xlsx	148 KB 61.4 KB 73.1 KB	Oct 13, 2022, 4:35 PM Oct 13, 2022, 4:34 PM Oct 13, 2022, 4:34 PM		

Figure 2: Select the Import Dataset... option.

A window will appear on the screen. It will display the data to be imported along with default data types for each column.

History					Con	ole Terminal ×	Background Jo	obs ×						
Import Text Data														
File/URL:														
D:/Lab Mate	rials/Indian_Agricu	ilture_Data.csv							Up	odate				
Data Preview:														
State (character)	Zone (character) ~	GSDP (double) ~	Area (double)	Production (double) ~	Prod./Area (double) ~	NFS (character) <sup>¬</sup>	PFS (character)	KFS (character)	District (character) *	DASM (di				
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	AMRITSAR					
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BARNALA					
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BATHINDA					
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FARIDKOT					
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FATEHGARH SAHIB	•				
										•				
Previewing first	50 entries.													
Import Options							Code Preview:							
Name: Indi	an Agriculture Da	First Re	w as Names [	Delimiter: Comma	▼ Escape: N	one 🔻	library(re		ture_Data <- read_csv("D:/Lab an_Agriculture_Data.csv")					
		Trim Sp		Quotes: Default	Comment: D		Materials/	Indian_Agricu n_Agriculture						
Skip:	0			ocale: Configure		efault ▼	view(india	n_Agriculture,	_Data)					
		C open i	Jata viewei L	ocale.	INA. D	ciduit •								

#### Figure 3: Import Text Data window.

**Step 3.** To change the default data type of a column, left-click anywhere in the box containing the column header. A drop down list will appear. Select the desired data type from the drop down list.

		Go to file/fur		<ul> <li>Addins</li> </ul>							🚯 Pro		
	History					Con	iole Terminal >	Background Jo	bs ×				
Glo	Import Text Data										-		
Giu	File/URL:												
	D:/Lab Materia	ls/Indian_Agricult	ure_Data.csv							Upd	date		
	Data Preview:												
	State (character)	Zone (character)	GSDP (double) ~	Area (double)	Production (double) ~	Prod./Area (double)	NFS (character) *	PFS (character)	KFS (character) *	District (character) *	DASM (d)		
	PUNJAB	Northern	Guess	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	AMRITSAR			
	PUNJAB	Northern	Character	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BARNALA			
	PUNJAB	Northern	Double	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BATHINDA			
	PUNJAB	Northern	Integer	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FARIDKOT			
	PUNJAB	Northern	Numeric	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FATEHGARH SAHIB	•		
	Previewing first 5	0 entries.	Logical										
	Import Options:		Date Time					Code Preview:			0		
	Name: Indiar	n_Agriculture_Dat	DateTime		Delimiter: Comma		one 🔻	<pre>library(readr) Indian_Agriculture_Data &lt;- read_csv("D:/Lab Materials/Indian_Agriculture_Data.csv")</pre>					
	Skip:	0	Factor		Quotes: Default	▼ Comment: D		View(India	n_Agriculture_	Data)	1:35		
			Include	ata Viewer	ocale: Configure	9 NA: D	efault 🔻				134		
			Skip										
	<li>2 Reading rasts</li>	ingular data using r	Only							Import Ca	incel		

Figure 4: Change default data type of a column.

In the above figure the data type of **GSDP** column is changed from *double* to *integer*.

Import Text Data         File/URI:         D/Lab Materials/Indian_Agriculture_Data.csv         Data Preview:         State       Zone         (character) *       (character) *         (character) *       (character)	History					🗇 Con	sole Terminal ×	Background Jo	bs ×				
D/Lab Materials/Indian_Agriculture_Data csv       Update         Data Preview:       State (character) * (Gouble) * (Gouble) * (Gouble) * (Gouble) * (Character) * (C	Import Text Data												
D/Lab Materials/Indian_Agriculture_Data csv       Update         Data Preview:       State (character) * (Gouble) * (Gouble) * (Gouble) * (Gouble) * (Character) * (C	File/URI :												
Data Preview:         State (review:         Open (review:         State (review:         PURJAB         Northern       51251093         917.11       734212         18.74376       LOW         PUNJAB       Northern         51251093       391.71         734212       18.74376         PUNJAB       Northern         51251093       391.71         734212       18.74376         LOW       LOW         MEDIUM       BARNALA         PUNJAB       Northern         51251093       391.71         734212       18.74376         LOW       LOW         MEDIUM       BARNALA         PUNJAB       Northern         51251093       391.71         734212       18.74376         LOW       LOW         Medium first So entries.       LOW         Import Options:       Code Pre	-	als/Indian Agricul	ture Data.csv							Upi	date		
State (character)     Zone (character)     SSDP (integer)     Area (double)     Prod/Area (double)     NFS (character)     PFS (character)     KFS (character)     District (character)     DaSM (character)       PURJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     AMRITSAR     Image: Character)     MEDIUM     BARNALA       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     BARNALA       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     BARNALA       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     BATHINDA       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     FARIDKOT       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     FARIDKOT       PUNJAB     Northem     51251093     391.71     7342.12     18.74376     LOW     LOW     MEDIUM     FARIDKOT       PUNJAB     Northem     51251093     391.71     7342.12     18.74376			-										
PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         BARNALA           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         BATHINDA           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         BATHINDA           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         BATHINDA         PUNJAB           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FATEHGARH SAHIB         P           Punviewing first 50 entries         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FATEHGARH SAHIB         P           Proviewing first 50 entries         Trimos Termines         Escape:         Norter	State												
PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         BATHINDA           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FARDKOT         V           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FARDKOT         V           PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FARDKOT         V           Previewing first 50 entries.         Common State Stat	PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	AMRITSAR			
PUNJAB         Northern         51251093         391.71         7342.12         18.7437c         LOW         LOW         MEDIUM         FARIDKOT           PUNJAB         Northern         51251093         391.71         7342.12         18.7437c         LOW         LOW         MEDIUM         FARIDKOT           Punuab         Northern         51251093         391.71         7342.12         18.7437c         LOW         LOW         MEDIUM         FARIDKOT           Previewing first 50 entries.         Import Options:         Code Preview.         Import Options:         Import Optins:         Import	PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BARNALA			
PUNJAB         Northern         51251093         391.71         7342.12         18.74376         LOW         LOW         MEDIUM         FATEHGARH SAHIB           Import Options:         Impor	PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BATHINDA			
Previewing first 50 entries.       Import Options:       Name: Indian_Agriculture_Data       Skip:       0       Ø Trim Spaces       Quotes:       Default       V       Material S/Indian_Agriculture_Data       Skip:       0         Ø Trim Spaces       Quotes:       Default         V	PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FARIDKOT			
Previewing first 50 entries.       Import Options:       Code Preview:       Import Options:       Name:       Indian_Agriculture_Data       First Row as Names       Delimiter:       Comman       Skip:       0       First Row as Names       Quotes:       Default       Comment:       Default	PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FATEHGARH SAHIB	•		
Import Options: Mame: Indian_Agriculture_Data @First Row as Names Delimiter: Comma Escape: None V Skip: 0 @Tim Spaces Quotes: Default V Comment: Default V (c) Ltypes = cols(CSDP = col_integer()))											•		
Name:     Indian_Agriculture_Data     Indian_Agric	Previewing first	50 entries.											
Name:     Indian_Agriculture_Data     Image: First Row as Names     Delimiter:     Comma     Escape:     None     Indian_Agriculture_Data < read_csv/"0:/Lab	Import Options:							Code Preview:			Ċ		
Skip: U	Name: India	in_Agriculture_Dat						Indian_Agr Materials/	Agriculture_Data <- read_csv("D:/Lab ials/Indian_Agriculture_Data.csv",				
	Skip:	0						col_ty View(India	pes = cols(GSI n_Agriculture	DP = col_integer())) _Data)	1		

*Figure 5: GSDP column with integer data type.* 

**Step 4.** Left-click on the **Import** button present at the bottom-right corner of the window.

History					Cons	ole Terminal >	Background Jo	obs ×					
Import Text Da	ta												
File/URL:													
D:/Lab Ma	terials/Indian_Agricu	lture_Data.csv							Uţ	odate			
Data Preview													
State (character)	✓ Zone (character) ▼	GSDP (integer) ¬	Area (double) ~	Production (double) *	Prod./Area (double)	NFS (character) *	PFS (character) ¬	KFS (character)	District (character)	DASM (di			
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	AMRITSAR				
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BARNALA				
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	BATHINDA				
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FARIDKOT				
PUNJAB	Northern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIUM	FATEHGARH SAHIB	-			
										•			
Previewing t	rst 50 entries.												
Import Optic	ins:						Code Preview:						
	lame: Indian Agriculture Data												
		Trim Sp		Quotes: Default	▼ Comment: D		Materials/	rials/Indian_Agriculture_Data.csv", col_types = cols(GSDP = col_integer()))					
Skip:	0					efault V		pes = cols(GS n_Agriculture		)			
		⊡ Open L	Jata viewer j	ocale: Configure	3 NA: D	elault 🔻							

*Figure 6: Select* **Import** *button to import the data.* 

The data after importing will automatically get displayed on the **Source** window.

Indian_Agriculture_Data	×														
🗈 📄 🚛 🕴 🖓 Filter										Terminal $\times$	Backgroun	d Jobs ×			-
							Q,			2.0 · ~/ ∞ rv(readr)					
* State ÷ Z	one G	isdp =	Area 🏺	Production <sup>©</sup>	Prod./Area	NFS	PFS	KFS <sup>÷</sup> I	Warning	message:					
1 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU 🔺					version 4.2.1 csv("D:/Lab Ma	terials/In	dian_Agriculture_D
2 PUNJAB	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU	csv",		-		integer()))		
B PUNJAB	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU			iculture_D		integer()))		
4 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU	>						
5 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
6 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
7 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
8 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
9 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
O PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU							
1 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU	Files Pl	ots Packag	jes Help	Viewer	Presentation		-
2 PUNJAB N	lorthern	51251093	391.71	7342.12	18.74376	LOW	LOW	MEDIU 💌			lew Blank File	• 🙂 C	Delete 🜛 Rename	🔅 More 🝷	
								•	0	Lab Materials					
owing 1 to 12 of 525 entrie	es, 15 total co	olumns							£	Name				Size	Modified
vironment History									_		ulture_Data.xls			148 KB	Oct 13, 2022, 4:35 PM
* 📊 📑 Import Datas	ot v 🚯 1	188 MiR 👻	1				= 1	.ist •   @ •			ilture_Data.xis			61.4 KB	Oct 13, 2022, 4:33 PM
<ul> <li>Global Environme</li> </ul>							0				ilture_Data.xis ilture_Data.csv			73.1 KB	Oct 13, 2022, 4:34 PM Oct 13, 2022, 4:34 PM
ta							~			indian_Agrici	liture_Data.cs	v		73.1 KB	Oct 13, 2022, 4:34 PM
		e	6.45.11	riables											

Figure 7: Data imported and viewed in R.

Follow the same steps to read data from an **.xlsx** or **.xls** file.

# Alternatively, using R console to read data files

To read and view data from **.csv** file execute the following commands:

```
# Loading "readr" library to use its "read_csv()" function
library(readr)
# Reading "Indian_Agriculture_Data.csv" file's data using the "read_csv()"
function and storing it in the "Indian_Agriculture_Data" object
Indian_Agriculture_Data <- read_csv("D:/Lab
Materials/Indian_Agriculture_Data.csv")
# Viewing the contents of "Indian_Agriculture_Data" object
View(Indian_Agriculture_Data)
```

The above commands shall read and display the data present in the **"Indian\_Agriculture\_Data.csv"** file.

To read and view data from **.xlsx** or **.xls** file execute the following commands:

```
library(readx1)
Indian_Agriculture_Data <- read_excel("D:/Lab
Materials/Indian_Agriculture_Data.xls")
View(Indian_Agriculture_Data)</pre>
```

The above commands shall read and display the data present in the **"Indian\_Agriculture\_Data.xls"** file. Same commands can be used to read data from **"Indian\_Agriculture\_Data.xlsx"** file by changing the input data file name.

## **Spoken Tutorials**

For more details, refer to the Introduction to Data Frames in R Spoken Tutorial video.