



**ACADGILD**

# SESSION 2: INTRODUCTION to working with R

---

Assignment 3



## Table of Contents

1.Introduction .....	3
2.Objective.....	3
3.Prerequisites .....	3
4.Associated Data Files .....	3
5.Problem Statement .....	3
6.Expected Output.....	3
7.Approximate Time to Complete Task.....	3

## 1. Introduction

This assignment will help you understand the concepts learnt in the session.

## 2. Objective

This assignment will test your skills on the basics of R.

## 3. Prerequisites

Not applicable.

## 4. Associated Data Files

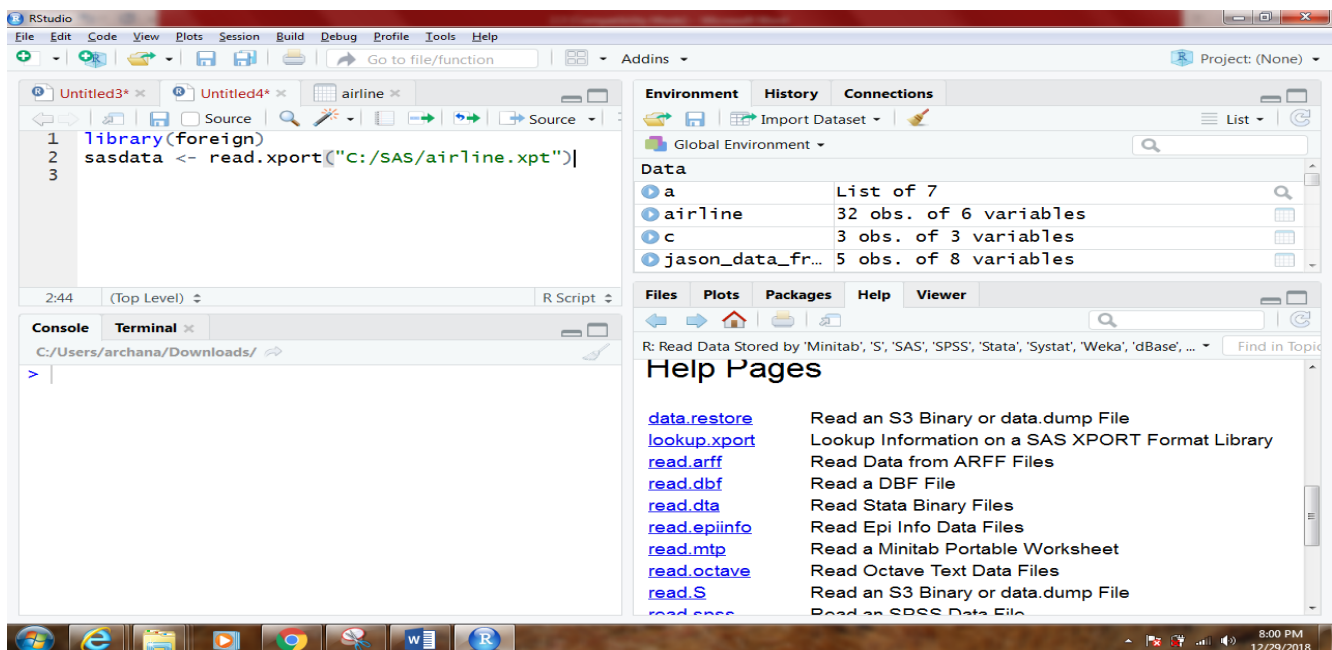
Not applicable.

## 5. Problem Statement

### 1. How to Import SAS XPORT Files into R With The foreign package

Ans-

To import SAS data files in R, first install and load the package `foreign`. The functions `read.xport()` of the package `foreign` import SAS Transport data file



## 2. How To Import SAS Files into R With The haven Package?

- Install it,
- `install.packages("haven")`,
- Load it
- `library(haven)`,
- Then pick the appropriate read function:
  - SAS: `read_sas()`

Output is shown in below screen shot: sample sas file is airline

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the following R code:

```
1 library(haven)
2 read_sas("airline.sas7bdat")
3 write_sas(airline, "C:/SAS/airline.sas7bat")
4
```
- Console:** Shows the execution of the code and the resulting tibble:

```
> library(haven)
> read_sas("airline.sas7bdat")
# A tibble: 32 x 6
  YEAR     Y     W     R     L     K
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1948  1.21 0.243 0.145  1.41 0.612
2 1949  1.35 0.260 0.218  1.38 0.559
3 1950  1.57 0.278 0.316  1.39 0.573
4 1951  1.95 0.297 0.394  1.55 0.564
5 1952  2.27 0.310 0.356  1.80 0.574
6 1953  2.73 0.322 0.359  1.93 0.711
7 1954  3.03 0.335 0.403  1.96 0.776
8 1955  3.56 0.350 0.396  2.12 0.827
9 1956  3.98 0.361 0.382  2.43 0.800
```
- Environment:** Shows the loaded objects:
  - `a`: List of 7
  - `airline`: 32 obs. of 6 variables
  - `c`: 3 obs. of 3 variables
  - `jason_da...`: 5 obs. of 8 variables
- Files:** Shows the installed packages in the User Library, including `assertt...`, `BH`, `bindr`, `bindrcpp`, `bitops`, and `boot`.

## 1. How to read Weka Attribute-Relation File Format (ARFF) files in R?

Ans:

This is done by installing package "sos"

```
install.packages("sos")
```

```
read.csv("data.arff", header=FALSE, comment.char = "@")
```

Below is screen shot. File name is train.arff

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 library(sos)
2 file.arff <- read.csv("train.arff", header=FALSE, comment.char = "@")
3 file.arff
4 library("sos")
```

The console shows the execution of the code, resulting in a data frame with 6 columns (V1-V6) and 3 rows of data:

```
> library(sos)
> file.arff <- read.csv("train.arff", header=FALSE, comment.char = "@")
> file.arff
```

	V1	V2	V3	V4	V5	V6
1	{246 28	256 99	864 85	1208 133	2274 47	2338 27
2	{100 21	246 50	555 181	655 88	656 31	890 153
3	{129 88	246 102	1056 264	1267 79	1632 68	1911 125

The Environment pane on the right shows the loaded data objects: 'a' (List of 7), 'airline' (32 obs. of 6 v...), 'c' (3 obs. of 3 va...), and 'file.a...' (444 obs. of 15...).

## 2. How to read a heavy csv/tsv file using readr package?

```
# Installing
install.packages("readr")
# Loading
library("readr")
```

```
read_delim(file, delim, col_names = TRUE)
# Read comma (",") separated values
read_csv(file, col_names = TRUE)
# Read semicolon (";") separated values
# (this is common in European countries)
read_csv2(file, col_names = TRUE)
# Read tab separated values
read_tsv(file, col_names = TRUE)
```

Read a tsv file

```
my_data <- read_tsv(file.choose())
# Read a csv file
my_data <- read_csv(file.choose())
```

The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 library(readr)
2 my_data <- read_csv("C:/Users/archana/Desktop/archana/GITS/mtcars.csv")
3 my_data
4 head(mtcars)
5
6
```

The console on the bottom left shows the output of the code:

```
> library(readr)
> my_data <- read_csv("C:/Users/archana/Desktop/archana/GITS/mtcars.csv")
> my_data
```

	model	mpg	cyl	disp	hp	drat	wt
1	Mazda RX4	21.0	6	160.0	110	3.90	2.620
2	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875
3	Datsun 710	22.8	4	108.0	93	3.85	2.320
4	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215
5	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440
6	Valiant	18.1	6	225.0	105	2.76	3.460
7	Duster 360	14.3	8	360.0	245	3.21	3.570
8	Merc 240D	24.4	4	146.7	62	3.69	3.190
9	Merc 230	22.8	4	140.8	95	3.92	3.150
10	Merc 280	19.2	6	167.6	123	3.92	3.440

The Environment pane on the right shows the 'Global Environment' with a variable 'mtcars' of type 'Data Frame' containing 32 observations and 12 variables. The Files pane on the bottom right shows the 'R: Read a delimited file (including csv & tsv) into a tibble' function, with a tooltip explaining the 'file' argument.

Data Analytics

## **6. Expected Format**

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

## **7. Approximate Time to Complete Task**

20 mins.



