



ACADGILD

SESSION 2: Introduction to working with R

Assignment 3

Submitted by: Munmun Ghosal
Login Id: munmun55@gmail.com
(M):+91-8007178659

Table of Contents

1. Problem Statement	3
2. Solution	3

1. Problem Statement

1. How to Import SAS XPORT Files into R With The foreign package
2. How to Import SAS Files into R With The haven Package?
3. How to read Weka Attribute-Relation File Format (ARFF) files in R?
4. How to read a heavy csv/tsv file using readr package?

2. Solution

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

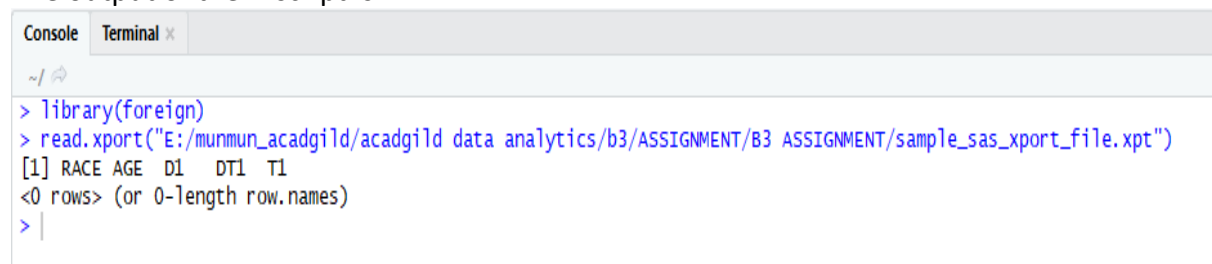
The foreign package is installed using the command: `install.packages("foreign")`

`read.xport (file)` command reads a file as a SAS XPORT format library and returns a list of `data.frames` ;where,file is character variable with the name of the file to read. The file must be in SAS XPORT format.

The R-script for importing SAS XPORT Files: `sample_sas_xport_file.xpt` into R is as follows:

```
library(foreign)
read.xport("E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3
ASSIGNMENT/sample_sas_xport_file.xpt")
```

The output of the R-script is:



```
Console Terminal x
~/
> library(foreign)
> read.xport("E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3 ASSIGNMENT/sample_sas_xport_file.xpt")
[1] RACE AGE D1 DT1 T1
<0 rows> (or 0-length row.names)
>
```

2. Import SAS Files into R With The haven Package

The haven package is installed using the command: `install.packages("haven")`

The haven package is used to Import and Export 'SPSS', 'Stata' and 'SAS' Files

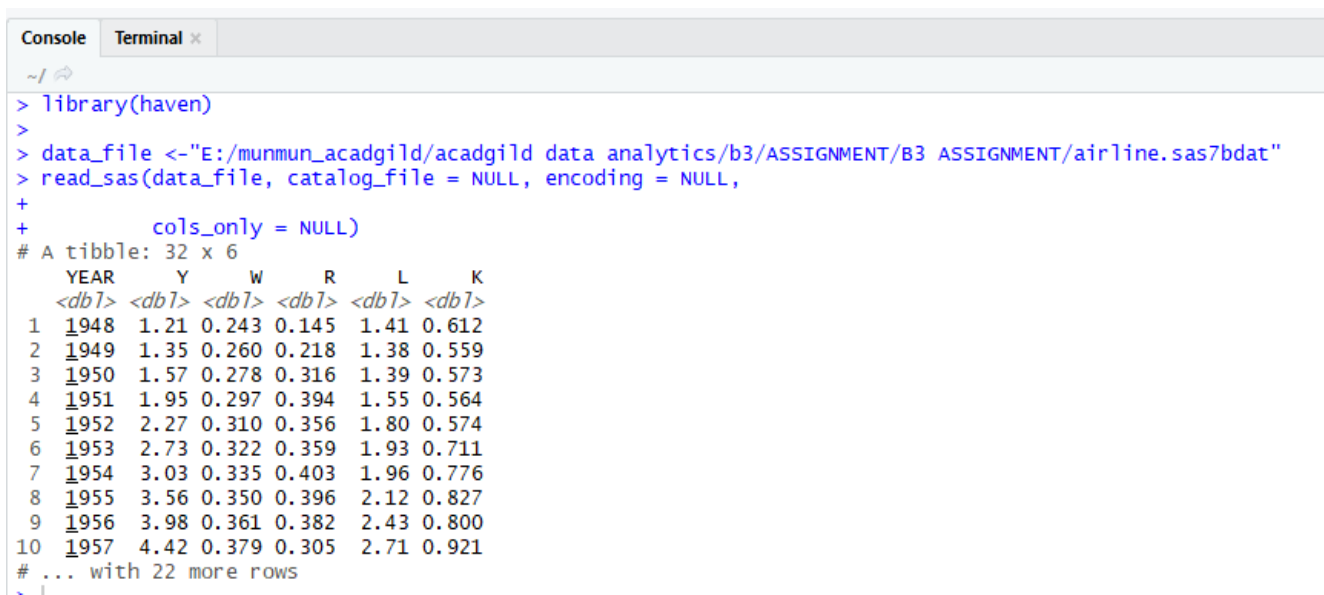
`read_sas (file)` function supports both `sas7bdat` files and the accompanying `sas7bdat` files that SAS uses to record value labels. `read_sas (file)` command reads a file `airline.sas7bdat` file.

The R-script for importing SAS Files into R With The haven Package is as follows:

```
library(haven)
data_file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3
ASSIGNMENT/airline.sas7bdat"
read_sas(data_file, catalog_file = NULL, encoding = NULL,

        cols_only = NULL)
```

The output of the R-script is:



```
Console Terminal x
~/
> library(haven)
>
> data_file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3 ASSIGNMENT/airline.sas7bdat"
> read_sas(data_file, catalog_file = NULL, encoding = NULL,
+
+         cols_only = NULL)
# 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
10 1957 4.42 0.379 0.305 2.71 0.921
# ... with 22 more rows
```

3. Read Weka Attribute-Relation File Format (ARFF) files in R

read.arff(file) reads data from Weka Attribute-Relation File Format (ARFF) files: sample_weka.arff file.

The script for reading Attribute-Relation File Format (ARFF) files in R is as follows:

```
library(foreign)
data_file<-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3
ASSIGNMENT/sample_weka.arff"
read.arff(data_file)
```

The output of the R-script is:

```
Console Terminal x
~/
> library(foreign)
> data_file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3 ASSIGNMENT/sample_weka.arff"
> read.arff(data_file)
  sepallength sepalwidth petallength petalwidth      class
1          5.1         3.5         1.4         0.2 Iris-setosa
2          4.9         3.0         1.4         0.2 Iris-setosa
3          4.7         3.2         1.3         0.2 Iris-setosa
4          4.6         3.1         1.5         0.2 Iris-setosa
5          5.0         3.6         1.4         0.2 Iris-setosa
6          5.4         3.9         1.7         0.4 Iris-setosa
7          4.6         3.4         1.4         0.3 Iris-setosa
8          5.0         3.4         1.5         0.2 Iris-setosa
9          4.4         2.9         1.4         0.2 Iris-setosa
10         4.9         3.1         1.5         0.1 Iris-setosa
> |
```

4. Read a heavy csv/tsv file using readr package

The 'readr' package provides a fast way to read rectangular data (like 'csv', 'tsv', and 'fwf'). The `read_csv()` and `read_tsv()` are special cases of the general `read_delim()`. They're useful for reading the most common types of flat file data, comma separated values and tab separated values, respectively.

a. R-script to read a heavy csv file using readr package:

```
library(readr)
file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3
ASSIGNMENT/cs2m.csv"
read_csv(file, col_names = TRUE, col_types = NULL,

  locale = default_locale(), na = c("", "NA"), quoted_na = TRUE,

  quote = "\"", comment = "#", trim_ws = TRUE, skip = 0, n_max = Inf,

  progress = show_progress())
```

The output of the R-script is:

```
Source
Console Terminal x
~/
> library(readr)
> file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3 ASSIGNMENT/cs2m.csv"
> read_csv(file, col_names = TRUE, col_types = NULL,
+           locale = default_locale(), na = c("", "NA"), quoted_na = TRUE,
+           quote = "\"", comment = "", trim_ws = TRUE, skip = 0, n_max = Inf,
+           progress = show_progress())
Parsed with column specification:
cols(
  BP = col_double(),
  chlstr1 = col_double(),
  Age = col_double(),
  Prgnt = col_double(),
  AnxtyLH = col_double(),
  DrugR = col_double()
)
# A tibble: 30 x 6
   BP chlstr1 Age Prgnt AnxtyLH DrugR
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  100    150    20     0     0     0
2  120    160    16     0     0     0
3  110    150    18     0     0     0
4  100    175    25     0     0     0
5   95    250    36     0     0     0
6  110    200    56     0     1     0
7  120    180    59     0     1     0
8  150    175    45     0     1     0
9  160    185    40     0     1     0
10 125    195    20     1     0     0
# ... with 20 more rows
> |
```

b. R-script to read a heavy tsv file using readr package

```
library(readr)
file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3
ASSIGNMENT/sample_tsv_file.tsv"
read_tsv(file, col_names = TRUE, col_types = NULL,

        locale = default_locale(), na = c("", "NA"), quoted_na = TRUE,

        quote = "\"", comment = "", trim_ws = TRUE, skip = 0, n_max = Inf,

        , progress = show_progress())
```

The output of the R-script is:

```
Console Terminal x
~/
> library(readr)
> file <-"E:/munmun_acadgild/acadgild data analytics/b3/ASSIGNMENT/B3 ASSIGNMENT/sample_tsv_file.tsv"
> read_tsv(file, col_names = TRUE, col_types = NULL,
+         locale = default_locale(), na = c("", "NA"), quoted_na = TRUE,
+         quote = "\"", comment = "", trim_ws = TRUE, skip = 0, n_max = Inf,
+         , progress = show_progress())
Parsed with column specification:
cols(
  name = col_character(),
  species = col_character(),
  foods = col_character()
)
# A tibble: 3 x 3
  name    species foods
<chr>   <chr>   <chr>
1 Meowsy  cat     tuna|catnip|ham|zucchini
2 Barky   dog     bones|carrots|tuna
3 Purrpaws cat     mice|nn|cookies
> |
```