Subject: Re: Reading data files into R Studio
Posted by Bridgette-DHS on Thu, 02 Feb 2017 11:31:19 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Specialist, Trevor Croft:

The data file you are looking at is a fixed format file, in which columns of the records determining which variable is which. Each record should have the same length, but in some files records with trailing blanks are truncated. You can find the layout of these records by looking at any of the .DCT (for Stata), .SAS (for SAS), or .SPS (for SPSS) files. These are all text files that describe the layout of the data, and you can use this information to construct code to read the data into R.

However, the easiest way to get data into R is actually to start with either the Stata or SPSS datasets. I generally prefer the Stata dataset, but they both work. You can use the read.dta() function, as follows using the Stata dataset:

dta <- read.dta("PKBR21FL.dta", convert.factors = FALSE)

read.dta() is in the package "foreign", so you will need
install.packages("foreign")
library(foreign)

I prefer not to convert variables to factors automatically so I use convert.factors = FALSE, but you may prefer to have it set to TRUE and automatically convert. If you don't automatically convert variables to factors, then you can use code such as
dta$sex <-factor(recode(dta$b4,"1='1 Male';2='2 Female';9='9 Missing';else=NA"))
or even
dta$sex = factor(dta$b4)