

Importing data into R and Writing Data in R

For small datasets, your best bet is to import the data with the `read.csv()` function.

For larger datasets (such as the ANES surveys), you will probably start off with a dataset in one of many file formats for commercial software, such as SPSS, Stata, SAS, or many others. These formats can be imported into R.

The function to do so is based on the R package ‘foreign’. First thing to do is load it into R. The easiest way to do so is to add the name `foreign` to the package commands at the top of each R lab script file. Then it will load along with `lattice`. Be sure to call it with `library(foreign)`.

The foreign package is intended to allow you to load datasets in “foreign” file formats into R. The use of the functions in the package is the same as `read.csv()`. You declare the name of an intended R dataframe, and then use the appropriate function to read the dataset into R.

For example, if a dataset called ‘polldata’ was stored in Stata format, I would import it into R with the following:

```
> poll<-read.dta("polldata.dta")
```

In my R workspace, the object “poll” now contains the dataset “polldata.dta”. Since I set my working directory to the storage location of the dataset, all that was needed was the name of the Stata dataset.

This method is recommended for downloading any of the ANES datasets into R. Importing it from Stata preserves the formatting of the factor, ordinal or nominal variables.

Each of the functions in foreign load a different dataset. Type `help(package="foreign")` to see the various dataset options. From there, you will see that each of the different `read` functions. For a file in SPSS format, you would use `read.spss()`. Each of these functions have different options, depending on the foreign dataset format. Type `?read.spss` to see to the various options for converting the dataset to R format.

Tip for those in the know: some data are stored in strange ways in SPSS format. Your best bet is to download a dataset into Stata format for importing into R.

Importing parts of the World Values Survey

You already have the dataset required for the project; if you choose to download other datasets, you are responsible for knowing which country dataset you want, and why.

Importing specific country data files for the WVS requires some intermediate skill in data manipulation. You need to download the *ENTIRE* WVS dataset, in STATA format, for example. Then, you have to subset the dataset to include just the dataset you would like.

Let’s say I’m interested in downloading the Albanian survey, from the 1999-2004 survey wave. I would download the “WVS 2000 official file compressed in ZIP format” dataset in .dta format. After downloaded and unzipped, it is called “wvs2000_v7.dta”.

After importing the dataset into R, let's say you wanted to keep the country survey of Albania. I import it into R:

```
wvs<-read.dta("wvs2000_v7.dta", warn.missing.labels=FALSE).
```

Go make a snack. It will take a few minutes. You are importing a massive survey dataset with tens of thousands of observations.

Once imported, you should save the workspace for future reference.

Two variables describe key characteristics of the dataset: `v1` describes the study wave. `v2` describes the country of the survey.

Then you'll want to use a function in R called `subset`. There are two arguments to the `subset` function that are relevant here, which is the name of the dataframe and the observation condition to select (or subset) on. We will try to create a dataset in which only the Albanian survey is included. We will create a new object in our workspace (I suggest starting a new one).

```
albania<-subset(wvs, v2=="albania")
```

In the `subset` command, "wvs" refers to the dataset, while "v2=="albania" refers to the logical condition TRUE if the observation in the wvs dataset belongs to the Albanian survey.

Now in our workspace, we have a new object called "albania", which is a dataset just for the Albanian WVS survey.

Repeat this process as needed for additional datasets.