

Making the most of census microdata

Practical 1

The purpose of this first tutorial is to familiarise yourself with R Studio, the console and the main functions. For this, you will be using a subset of variables from the safeguarded version of the census microdata 2011. This dataset will be used in all the tutorials of this workshop and is being made available on the folder **census_workshop**.

First of all, you need to set a working space on your PC. Copy the folder **census_workshop** into the work folder of your PC (C:Work).

Note: If you are doing this tutorial in your own PC, choose you desired location.

Now, let's start with R

Task 1: Set the working directory to your folder. **Note** Remember you can do this with code or using the point and click option.

```
#Solution:  
setwd("C:/Work")
```

Note: If you are doing this tutorial in your own PC, choose the folder where the data has been saved.

Alternative solution Go to:

Session... Set working directory...Choose directory... select the folder

Task 2: Installing Packages

a. Install the package dplyr and haven

```
#Solution:  
install.packages("dplyr")  
install.packages("haven")
```

- b. Load the packages `dplyr` and `haven`. Remember that the package needs to be loaded before being used.

```
#Solution:  
library(dplyr)  
library(haven)
```

Task 3: Import the microdata dataset into R

Hint: use `haven`

This is the general function:

```
read_dta(name_of_the_dataset.dta)
```

In this case, we used `read_dta`, as `dta` is the suffix for Stata files. Other formats can also be read (imported) into R, such as SPSS and SAS file. More details about that can be found on the `haven` reference manual.

Make sure to store this into the environment by assigning the dataset a name. Short but informative names are the best.

```
#Solution:  
censusd <- read_dta("census_workshop.dta")
```

You can also let R help you with this. Go to the **Files** tab (bottom right panel of R Studio) and click on the dataset name; this will give you the option of **view file** and **import dataset**. Select **import dataset**, R will import the data for you. Make sure to read the code that this operation leaves on the console, as this can help you to learn how to do it yourself!

Note: "census_workshop.dta" was the name of the data used in the workshop, so you will need to change the name of the dataset to the one that you have.

Task 4: Exploring the data

R Studio allows us to see the dataset in the same console in a similar way as other statistical software packages.

To see the data, you can write the function `View(censusd)`. This function uses the name that you give to your dataset in R. You can also click on the name of your imported dataset in the environment window, and R will open a new tab with the data.

```
#Solution:  
View(censusd)
```