# R Notebook: Warm up exercise with data frames

This is a R Notebook that helps you to start working with data frames.

A data frame is a table or a two-dimensional array-like structure in which each column contains values of one variable and each row contains one set of values from each column.

## Fist step: Establish your working directory

Set your working directory with setwd ("P:/folder/folder") or go to Session menu-> Set Working Directory ->Choose Directory

```
# Add your code
```

## Second step: Load the data

Load the file people-example.csv that needs to be saved in the working directory. You will use the function read.csv. Save the information in a variable called datos. This variable will be a data frame.

```
# Add your code
```

The structure of the data frame can be seen by using str() function.

```
#str(data frame name)
```

## Third step: Inspect the dataframe

Let's use some operations to check on the dataframe

Use the function head to see the first elements of the data frame

```
# Add your code
```

Use the function tail to see the last elements of the data frame

```
# Add your code
```

If you want to filter a specific column, for example Country column

```
# Add your code
```

If you want to filter using more than one column, for example First.Name, Country and age columns

```
# Add your code
```

Calculate the max age value and its mean

```
# Add your code
```

Use summary function to summary data frame statistics

```
# Add your code
```

# Operations with the dataframe itself

## Rename columns

Use rename function to change the names of First.Name and Last.Name columns. Use First and Last instead.

```
# Load plyr package in order to use rename function
library(plyr)

# Add your code
```

##Concatenate String in a new column Add a new column in the data frame, called FullName, which will include the name and last name.

```
# Add your code
```

## Removing columns

```
# Add your code
```

## Editing columns and programming

Making changes in the columns is usual in Data Science. Let's make a function to substract 50000 dollars in taxes to the ones who earns more than 300000 dollars. There are two different ways of doing this: 1) using loops or 2) using apply family functions. Please, try both of them!

Option 1: Using loops

Option 2: Using apply family functions (see Quick R Tutorial)

```
changeSavings<-function(saving){
  #Add your code
}
```

And now we are going to test our function before using it with the data frame

```
a=changeSavings(5000000)
a
```

```
## NULL
```

```
b=changeSavings(500)
b
```

```
## NULL
```

Once we have tested the function works correctly, you will apply it to the data frame using sapply function.

```
 #Add your code
```

#Sorting values Sort the values of the data frame by age column.

```
#Add your code
```

## Done!!