Data Import

Quantide Srl

2016-05-02

First of all, set your working directory in the *data* folder, using `setwd()` function, like in this example

```
setwd("C:/Users/Veronica/Documents/rbase/data)
```

We will work inside this folder.

Text Files

Exercise 1

a. Import text file named *"tuscany.txt"* and save it in an R object named `tuscany_df`.
Open the text file before importing it to control if the first row contains column names and to control the field and the decimal separator characters. Remember to not import the character columns as factors.

```
tuscany_df <- read.table("tuscany.txt", header = TRUE, sep = "|",
                         dec=".", stringsAsFactors = FALSE)
```

b. Visualize the first rows of `tuscany_df`

```
head(tuscany_df)
```

```
##   id sex year_of_birth marital_status  income house_number
## 1  1   M          1969        married 16101.1       5144.0
## 2  2   M          1962         single 17220.0       6158.0
## 3  3   M          1965       divorcee 28801.9      10078.0
## 4  4   F          1968         single 25964.0      11133.7
## 5  5   M          1975        married 16522.5       5078.0
## 6  6   M          1977        married 18124.0       5115.0
##               city_name province provincial_acronym
## 1            Riparbella     Pisa                 PI
## 2              Capolona   Arezzo                 AR
## 3             Pomarance     Pisa                 PI
## 4               Cascina     Pisa                 PI
## 5              Quarrata  Pistoia                 PT
## 6 Castiglion Fiorentino   Arezzo                 AR
```

Exercise 2

Import 7 rows of the text file named *"solar.txt"* skipping the first two rows. Save it in the object `solar_df`. Open the text file before importing it to control if the first row contains column names and to control the field and the decimal separator characters. Remember to not import the character columns as factors.

```
solar_df <- read.table("solar.txt", header = FALSE, sep = ",",
                       dec=".", stringsAsFactors = FALSE,
                       nrows = 7, skip = 2)
solar_df
```

```
##    V1    V2    V3    V4
## 1 mar 23877 24671 22455
## 2 apr 24377 23677 23670
## 3 mag 24581 25476 24999
## 4 giu 22154 21998 22451
## 5 lug 20924 21645 23871
## 6 ago 23183 22576 23556
## 7 set 27446 27695 28664
```

Exercise 3

Considering the following data frame, named `df`:

```
df <- data.frame(col1=1:4, col2=4:1, col3=c("one", "two", "three", "four"),
                 stringsAsFactors = FALSE)
```

Save it in a .txt file named *"exercise-3.txt"* in *data* folder.

```
write.table(df, file="exercise-3.txt")
```

Excel Files

Exercise 1

a. Import xlsx file *"flowers.xlsx"* using `XLConnect` function `loadWorkbook()` and save it in a R workbook object named `flowers`.
Remember to load `XLConnect` package, supposing it is already installed.

```
require(XLConnect)
```

```
flowers <- loadWorkbook("flowers.xlsx")
```

b. Read *iris* sheet with `readWorksheet()` function and save it in `flower_df` object. Then, visualize its first rows.

```
flowers_df <- readWorksheet(flowers, sheet = 'iris')
head(flowers_df)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Exercise 2

a. Create a new file xlsx, named *"exercise-2.xlsx"*, and save it in the R worksheet object, named `ex_2`. Use: `loadWorkbook()` and `saveWorkbook()` functions of `XLConnect`.

```
require(XLConnect)
ex_2 <- loadWorkbook(filename = "exercise-2.xlsx", create = TRUE)
saveWorkbook(ex_2)
```

b. Create a sheet in the R workbook object, named `df`, using `createSheet()` function. Remember to save the changes also in .xlsx file (use `saveWorkbook()` function).

```
createSheet(object = ex_2, name = 'df')
saveWorkbook(ex_2)
```

c. Considering the following data frame, named `numbers_df`:

```
numbers_df <- data.frame(a= 1:4, b=c("one", "two", "three", "four"), stringsAsFactors = FALSE)
numbers_df
```

```
##   a     b
## 1 1   one
## 2 2   two
## 3 3 three
## 4 4  four
```

Add it to `df` sheet of `ex_2` R workbook object, starting from row 3 and from column 2. Use the function `writeWorksheet()`. Remember to save the changes also in .xlsx file (use `saveWorkbook()` function).

```
writeWorksheet(object = ex_2, data = numbers_df, sheet = "df", startRow = 3, startCol = 3)
saveWorkbook(ex_2)
```

Databases

Exercise 1

a. Connect to *"plant.sqlite"* SQLite database, using `dbConnect()` function of `RSQLite` package. Save the connection in an R object, named `con`.
Remember to load `RSQLite` package, supposing it is already installed.

```
require(RSQLite)
```

```
con <- dbConnect(RSQLite::SQLite(), "plant.sqlite")
```

b. See the list of available tables in *"plant.sqlite"* db, using `dbListTables()` function.

```
dbListTables(con)
```

```
## [1] "PlantGrowth"
```

c. See list of fields in *"PlantGrowth"* table of *"plant.sqlite"* db, using `dbListFields()` function.

```
dbListFields(con, name = "PlantGrowth")
```

```
## [1] "weight" "group"
```

d. Send query to *"PlantGrowth"* table of *"plant.sqlite"* which select the records with `weight` greater than 5.5.

```
dbGetQuery(con, "SELECT * FROM PlantGrowth WHERE weight >= 5.5")
```

```
##   weight group
## 1   5.58  ctrl
## 2   6.11  ctrl
## 3   5.87  trt1
## 4   6.03  trt1
## 5   6.31  trt2
## 6   5.54  trt2
## 7   5.50  trt2
## 8   6.15  trt2
## 9   5.80  trt2
```

e. Disconnect from the database, using `dbDisconnect()` function.

```
dbDisconnect(con)
```

```
## [1] TRUE
```

R Data Files

Exercise 1

Given the following data frame, named `df_rdata`:

```
df_rdata <- data.frame(a=1:20, b=20:1)
```

Save it in *.Rda* format in the file *"df_rdata.Rda"*, using `save()` function.

```
save(df_rdata, file = "df_rdata.Rda")
```

```
## [1] TRUE
```

Exercise 2

Load *"drug.Rda"* file into the environment, using `load()` function.

```
load("drug.Rda")
```