`qtraining` repository

Date of documentation update: 20161220
Documentation updated by: Veronica

Introduction

`qtraining` repository contains Quantide training courses material.

The courses included are:

- **R for Beginners** (two days course)
- **R for Beginners** (one day course)
- **Statistical Models with R**
- **Base R programming**
- **Data Manipulation with R**
- **Data Visualization with R** in english and in italian (under development)
- **Data Mining with R**
- **Big Data** (under development)

`qtraining` Structure

`qtraining` repository contains both courses development material and courses manuals.
The structure is organized in the following folders:

- *00-qdata*: containing courses data organized as an R package, named `qdata`
- *010-rbase-two-days*: containing "R for Beginners" (two days) course
- *011-rbase-one-days*: containing "R for Beginners" (one day) course
- *020-models*: containing "Statistical Models with R" course
- *030-rprogramming-base*: containing "Base R programming" course
- *050-dplyr-datamanage*: containing "Data Manipulation with R" course
- *060-ggplot*: containing "Data Visualization with R" course
- *061-ggplot-ita*: containing "Data Visualization with R" course in italian (under development)
- *080-data-mining*: containing "Data Mining with R" course
- *100-bigdata-spark*: contanis "Big Data" course
- *courses-index*: containing courses index
- *include*: containing files for output structure building

All previous folders includes courses material, apart from *00-qdata*, *courses-index* and *include* folders.

An overview on courses material folders

Each course material folder name must be identified by a **number** and by the **course name**, in the following format: "number-course name".
The **number** to be assigned to the course name must consist of three digits and the **course name** must describe the course name in short. If the **course name** consists in more than one word, the words must be separated by `-`.

The courses material folders have the same base structure:

1. *input* folder, which contains source files of course material

2. *output* folder, which contains built course material files
3. *exercises* folder, which contains source files of exercises
4. *docs* folder, which contains documents useful for course development
5. *Makefile*, which is a text file containg code written in Linux kernel. It is a simple way to organize code compilation. It contains the instructions for building the content of `output` folder
6. RStudio project file, named `course name.Rproj`

In some courses material folder is included also *data* folder, which contains data that can't be included in `qdata` package.

input folder

It includes:

- *Markdown and R Markdown scripts with course content*. Course content is organized in chapters. Each chapter is identified by a script. It is recommended to add a number to the script names for ordering them (e.g. 01-first.Rmd).
- *images* folder, which contains figures to be included in the course content
- *TOC* file, which contains course index.

TOC file

TOC files includes the structure of the course manual.

The md or Rmd files must be ordered from the first to the last and must be associated with a title. In particular, you have to write firstly the title, then "|" and finally the md and Rmd name. You can also divide the rmd into sections, writing only the title of the section before the belonging Rmds.

Here an example:

First Section First Rmd | 1-first.Rmd Second Rmd | 2-second.Rmd

Second Section Third Rmd | 1-third.Rmd

output folder

It includes:

- *html* folder, which contains htmls files with the full content of the course and a zip folder containing all html files (the folder name is "course name.zip")
- *purl* folder, which contains R script with R code extracted from html (one R script for each html file), and a zip folder containing all R scripts (the folder name is "R.zip")
- *pdf* folder, which contains pdf file/s of courses exercises (this folder is not always present)

exercises folder

It includes:

- Markdown and R Markdown scripts with course exercises_.
- *images* folder, which contains figures to be included in exercises content

This folder is not populated in all material courses folders. It is populated in: *010-rbase*, *020-models*, *050-dplyr-datamanage*.
In *010-rbase* and *050-dplyr-datamanage*, exercises are organized in chapters. Each chapter is identified by a script. The pdf file/s of courses exercises are built by the *Makefile* and included in *output/pdf* folder.

In *020-models* script exercises and built pdf and html files are included in *exercises* folder. They are not built by the *Makefile*, but manually clicking "Knit PDF" from RStudio toolbar, because the exercises are not yet completed.

Guide for building course material

1. Open RStudio
2. Double-click on the project file
3. Click on "Build All" from RStudio "Build" tab or click "Ctrl+Shift+B" on the keyboard

The course building follows the instructions provided by *Makefile*. For more details see the *Makefile* of the course of interest.

00-qdata folder

This folder contains courses data organized as an R package, named `qdata`.
It is structured in the following way:

- *data* folder, which contains data in .RData format included in `qdata` package
- *R* folder, which contains two R script: `qdata.R` and `qdata-data.R` with package and data documentation (written with roxigen2)
- *pkgs* folder, which contains `qdata` package versions realized
- *doc* folder, which contains an R script `raw-data.R`. This script includes some commands of operations done on data before including them in package.
- *rowdata* folder, which contains original data files
- *DESCRIPTION* file, which contains library desciption
- *NAMESPACE* file, which contains informations about imported and exported functions (automatically created by library 'build')
- R project file, named `qdata.Rproj`

On Decembre 20 2016, the version of qdata is 0.27.

Guide for building `qdata`

Build `qdata`

`qdata` is an R package so its building works as well as any R package building.

1. Open RStudio
2. Double-click on the project file
3. Click "Build & Reload" on RStudio "Build" tab or click "Ctrl+Shift+B" on the keyboard

Note: When one ore more data file/s are added or when data documentation is modified, the package version MUST be updated. To update the package version modify *Version:* tag in *DESCRIPTION* file going forward of one digit (e.g. 027 -> 0.28)

Create source package

When you edit a new version of qdata, you MUST build a Source package of the new version:

Follow these steps:

1. Move the folders: *pkgs* and *rowdata* from *00-qdata* folder into another location (cut and paste), in order to not include them in the source package
2. Open RStudio
3. Open "More" window on RStudio "Build" Tab
4. Click on "Build Source package"
5. Reinsert the folders: *pkgs* and *rowdata* into *00-qdata* folder
6. Move the file *qdata_x.xx.tar.gz* from "~dev/qtraining/00-qdata" into "~dev/qtraining/00-qdata/pkgs"

Guide for installing and loading `qdata`

1. Open RStudio
2. Install the package typing the following lines on the R console:

```
install.packages("~/dev/qtraining/00-qdata/pkgs/qdata_0.27.tar.gz", repo = NULL )
```

3. Loading `qdata` package on the workspace, typing:

```
require(qdata)
```

4. Loading data, included in `qdata` package, on the workspace. For example, if we want to load "bank" data we have to type:

```
data("bank")
```

Otherwise, you can install `qdata` from Rstudio "Packages" tab:

1. Open RStudio
2. Click on "Install"
3. Set "Install from" filed to "Package Archive (.tar.gz)"
4. Choose "qdata" for "Packages" field

qdata locations

The materials included into *00-qdata* folder is included also in another github repository: https://github.com/quantide/qdata.

qdata is a public Quantide repository contining the data necessary for public courses available on Quantide website. So also this version has to be updated. When you update the version on *qtraining* repository you have to update also that in *qdata* repository.

Advice to update version: in *qdata* repository, add the added .Rdata files into *data* folder, and replace the scripts included in *R* folder with that modified of *qtraining* version.

Future Developments: develop *qdata* package into a single location.

include folder

This folder contains files with the instructions on output files structure building for each course. It means that the *Makefile* of each course refers to these files for building the structure of output files.

It is structured in these folders:

- *html*
- *libs*
- *r*
- *tex*

courses-index folder

The content of this folder is used for building an index of courses.
It includes:

- *course-index.html*: html file of courses index
- *images* folder: which includes figures used in *course-index.html* file

Modify courses index

1. Open *course-index.html* with a text editor
2. Modify the file. The code is written in HTML
3. Save the file

Other Details

Generate a PDF BOOK starting from rmds (bookdown)

To buid a book in PDF containing the rmds:

1. See PDF BOOK section in Makefile of 060-ggplot
2. Use *bookdown* package

Using *bookdown* is the better option.
It organizes in a better way the materials, avoiding problems with LateX (spaces, positioning, ...)

To use *bookdown*:

1. set working directory inside *input* folder
2. De-comment *bookdown::pdf_book:* part and comment html_document part
3. Modify *manual-with-written-cover.tex*, lines 174 and 176, respectitively with:

```
\begin{flushleft}\includegraphics[scale=.175]{./images/quantide.png}\end{flushleft}
\begin{flushright}\includegraphics[scale=.25]{./images/R-training.png}\end{flushright}
```

4. De-comment the first level title in each rmd (# Title)
5. Run

```
bookdown::render_book("index.md", "bookdown::pdf_book", new_session = T)
```

A *_book* folder and other files will be automathically created. *_main.pdf* is the PFD book and it is included in *_book* folder.

To use PDF BOOK section in Makefile of 060-ggplot

1. Copy *images* folder outside *input*
2. Click on *Build All* in *Build* tab