

MilanoR

4th meeting

October 24, 2013

R and Microsoft Office



Nicola Sturaro Sommacal

nicola.sturaro@quantide.com



Outline

- **Why Microsoft Office?**
- **Read a Microsoft Excel spreadsheet**
- **Produce a Microsoft Word report**
- **Access a Microsoft Access database**

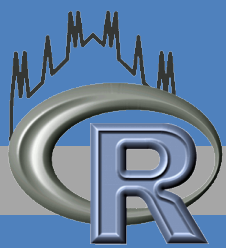


Why Microsoft Office?



Microsoft Office is a *standard de facto* in many business environments.

Many people daily write a report with Word, update some Excel spreadsheet or add new customers to the Access database.



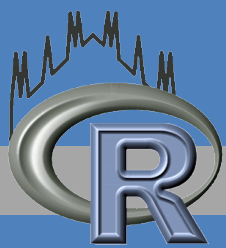
Why Microsoft Office?

In many cases, better alternatives are available.

Reports can be produced directly from R with Sweave or Markdown and easily convert into HTML or PDF, ready to be distributed. Static data should be kept in text files, while database can be easily managed with some SQL database.

Sometimes, due to «orders from above» (of some dumb boss), we have to use Microsoft Office.

Integrate them with R will save time (and money).



Read a Microsoft Excel spreadsheet



Converting Excel files to text files is the better solution.

Text files are universals, they can be opened with any computer at any time.

Sometimes, a spreadsheet is a more convenient way to modify data and re-save each time the file as a text file may be a boring and time consuming task.



Read a Microsoft Excel spreadsheet

A state-of-art solution is not available, yet.

Many solutions have been implemented to read Excel files from R: each one has advantages and disadvantages.

Get an overview of some solutions, allows the choice of the best solution case-by-case.



Read a Microsoft Excel spreadsheet

```
> require(gdata)  
> df= read.xls ("myfile.xlsx")
```

Advantages:

- cross platform

Disadvantages:

- Perl required



Read a Microsoft Excel spreadsheet

```
> require(XLConnect)  
> wb= loadWorkbook("file.xlsx")  
> df= readWorksheet(wb)
```

Advantages:

- Java-based (cross platform)
- returns satisfactory results

Disadvantages:

- very slow for large data sets



Read a Microsoft Excel spreadsheet

```
> require(xlsx)  
> df= read.xlsx("file.xlsx")  
> df= read.xlsx2("file.xlsx")
```

Advantages:

- Java-based (cross platform)
- probably returns best results

Disadvantages:

- slow (read.xlsx)
- requires some more options (read.xlsx)



Produce a Microsoft Word report



With the increasing of reproducible research and automatic reporting, many people asks how to produce a report with Microsoft Word from R.

Also in this case, there is not a state-of-art solution, mainly because it is something (relatively) new.



Produce a Microsoft Word report

A R-Markdown file can be produced and converted to Word using Pandoc.

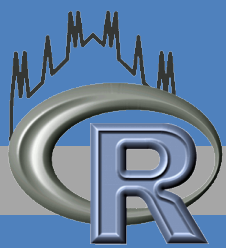
```
$ pandoc myreport.rmd -o myreport.docx
```

Advantages:

- requires Markdown: easy and fast to learn, several outputs
- good results

Disadvantages:

- requires some experience with Word styles
- integration is not perfect



Produce a Microsoft Word report

> `require(R2DOCX)`

Package hosts on GitHub, is not available from CRAN.

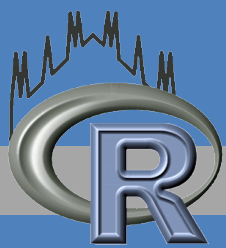
<https://github.com/davidgohel/R2DOCX>

Advantages:

- good integration with Word tables and charts

Disadvantages:

- code may be difficult to read
- code is not “interchangeable”



Access a Microsoft Access database

Connect to an Access database is not as critical as read an Excel files.



ODBC works fine through to RODBC package.

ODBC drivers installation may require some work. ODBC drivers are platform specific, but they are available for Windows, Mac and Linux.



Access a Microsoft Access database

```
> ch= odbcConnect ("odbcName")  
> df= sqlFetch (ch)
```

Advantages:

- excellent results
- very fast also with large data sets

Disadvantages:

- doesn't work with xls files



R and Microsoft Office

4th meeting MilanoR - October 24, 2013

Thank you!



www.quantide.com

Nicola Sturaro Sommacal

nicola.sturaro@quantide.com

370 718 32 71