

R for the Excel User

Julie Lowndes & Allison Horst

2019-09-26

Contents

1	Welcome	7
2	Overview	9
2.1	What to expect	9
2.2	What you'll learn	10
2.3	Learning with data that are not your own	10
2.4	Emphasizing collaboration	10
2.5	By the end of the course...	10
2.6	Prerequisites	10
2.7	Credit	11
3	readxl	13
3.1	Summary (a few sentences)	13
3.2	Objectives (more detailed, bulletpoints?)	13
3.3	Resources	14
3.4	Lessons teaching for each objective..... (objectives, examples) . .	14
3.5	Fun facts (quirky things) - making a note of these wherever possible for interest (little "Did you know?" sections)	15
3.6	Interludes (deep thoughts/openscapes)	15
3.7	Our Turn Your Turn 1	15
3.8	Our Turn Your Turn 2	15
3.9	Efficiency Tips	15
4	RMarkdown	17
4.1	Summary (a few sentences)	17
4.2	Objectives (more detailed, bulletpoints?)	17
4.3	Resources	17
4.4	Lessons teaching for each objective..... (objectives, examples) . .	17
4.5	Fun facts (quirky things) - making a note of these wherever possible for interest (little "Did you know?" sections)	17
4.6	Interludes (deep thoughts/openscapes)	17
4.7	Our Turn Your Turn 1	17
4.8	Our Turn Your Turn 2	17
4.9	Efficiency Tips	17

5	Dplyr and Pivot Tables	19
5.1	Summary (a few sentences)	19
5.2	Objectives (more detailed, bulletpoints?)	19
5.3	Resources	19
5.4	Lessons teaching for each objective..... (objectives, examples)	19
5.5	Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)	19
5.6	Interludes (deep thoughts/openscapes)	19
5.7	Our Turn Your Turn 1	19
5.8	Our Turn Your Turn 2	19
5.9	Efficiency Tips	19
6	Dplyr and vlookups	21
6.1	Summary (a few sentences)	21
6.2	Objectives (more detailed, bulletpoints?)	21
6.3	Resources	21
6.4	Lessons teaching for each objective..... (objectives, examples)	21
6.5	Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)	21
6.6	Interludes (deep thoughts/openscapes)	21
6.7	Our Turn Your Turn 1	21
6.8	Our Turn Your Turn 2	21
6.9	Efficiency Tips	21
7	Tidying	23
7.1	Better practices [needs a better name]	23
7.2	Summary (a few sentences)	23
7.3	Objectives (more detailed, bulletpoints?)	23
7.4	Resources	24
7.5	Lessons teaching for each objective..... (objectives, examples)	24
7.6	Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)	25
7.7	Interludes (deep thoughts/openscapes)	25
7.8	Our Turn Your Turn 2	25
7.9	Efficiency Tips	25
8	Formatting and Sharing	27
8.1	Summary (a few sentences)	27
8.2	Objectives (more detailed, bulletpoints?)	27
8.3	Resources	27
8.4	Lessons teaching for each objective..... (objectives, examples)	27
8.5	Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)	27
8.6	Interludes (deep thoughts/openscapes)	27
8.7	Our Turn Your Turn 1	27
8.8	Our Turn Your Turn 2	27

<i>CONTENTS</i>	5
8.9 Efficiency Tips	27
9 Synthesis	29
9.1 Summary (a few sentences)	29
9.2 Objectives (more detailed, bulletpoints?)	29
9.3 Resources	29
9.4 Lessons teaching for each objective..... (objectives, examples) . .	29
9.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)	29
9.6 Interludes (deep thoughts/openscapes)	29
9.7 Our Turn Your Turn 1	29
9.8 Our Turn Your Turn 2	29
9.9 Efficiency Tips	29

Chapter 1

Welcome

Excel is a widely used and powerful tool for working with data. As automation, reproducibility, collaboration, and frequent reporting become increasingly expected in data analysis, a good option for Excel users is to extend their workflows with R. Integrating R into data analysis with Excel can bridge the technical gap between collaborators using either software. R enables use of existing tools built for specific tasks and overcomes some limitations that arise when working with large datasets or repeated analyses. This course is for Excel users who want to add or integrate R and RStudio into their existing data analysis toolkit. Participants will get hands-on experience working with data across R, Excel, and Google Sheets, focusing on: data import and export, basic wrangling, visualization, and reporting with RMarkdown. Throughout, we will emphasize conventions and best practices for working reproducibly and collaboratively with data, including naming conventions, documentation, organization, all while “keeping the raw data raw”. Whether you are working in Excel and want to get started in R, already working in R and want tools for working more seamlessly with collaborators who use Excel, or whether you are new to data analysis entirely, this is the course for you!

If you answer yes to these questions, this course is for you!

- Are you an Excel user who wants to expand your data analysis toolset with R?
- Do you want to bridge analyses between Excel and R, whether working independently or to more easily collaborate with others who use Excel or R?
- Are you new to data analysis, and looking for a good place to get started?

Chapter 2

Overview

Welcome.

This workshop you will learn hands-on how to begin to interoperate between Excel and R.

We will learn and reinforce X main things all at the same time: coding with best practices (R/RStudio), Y, Z. This training will teach these all together to reinforce skills and best practices, and get you comfortable with a workflow that you can use in your own projects.

2.1 What to expect

This is going to be a fun workshop.

The plan is to expose you to a lot of great tools that you can have confidence using in your research. You'll be working hands-on and doing the same things on your own computer as we do live on up on the screen. We're going to go through a lot in these two days and it's less important that you remember it all. More importantly, you'll have experience with it and confidence that you can do it. The main thing to take away is that there *are* good ways to approach your analyses; we will teach you to expect that so you can find what you need and use it! A theme throughout is that tools exist and are being developed by real, and extraordinarily nice, people to meet you where you are and help you do what you need to do. If you expect and appreciate that, you will be more efficient in doing your awesome science.

You are all welcome here, please be respectful of one another. You are encouraged to help each other.

Everyone in this workshop is coming from a different place with different experiences and expectations. But everyone will learn something new here, because there is so much innovation in the data science world. Instructors and helpers learn something new every time, from each other and from your questions. If you are already familiar with some of this material, focus on how we teach, and how you might teach it to others. Use these workshop materials not only as a reference in the future but also for talking points so you can communicate the importance of these tools to your communities. A big part of this training is not only for you to learn these skills, but for you to also teach others and increase the value and practice of open data science in science as a whole.

2.2 What you'll learn

2.3 Learning with data that are not your own

2.4 Emphasizing collaboration

Collaborating efficiently has historically been really hard to do. It's only been the last 20 years or so that we've moved beyond mailing things with the postal service. Being able to email and get feedback on files through track changes was a huge step forward, but it comes with a lot of bookkeeping and reproducibility issues (did I send that report based on `analysis_final_final.xls` or `analysis_final_usethisone.xls`?). But now, open tools make it much easier to collaborate.

Working with collaborators in mind is critical for reproducibility. And, your most important collaborator is Future You. This training will introduce best practices using open tools, so that collaboration will become second nature to you!

2.5 By the end of the course...

By the end of the course,

2.6 Prerequisites

Before the training, please make sure you have done the following:

1. Download and install **up-to-date versions** of:
 - R: <https://cloud.r-project.org>

- RStudio: <http://www.rstudio.com/download>
 - Git: <https://git-scm.com/downloads> *Note: open the download and follow normal install procedures on your computer but you won't see any software installed when you're done*
2. Create a GitHub account: <https://github.com> *Note! Shorter names that kind of identify you are better, and use your work email!*
 3. Get comfortable: if you're not in a physical workshop, be set up with two screens if possible. You will be following along in RStudio on your own computer while also watching a virtual training or following this tutorial on your own.

2.7 Credit

This material builds from a lot of fantastic materials developed by others in the open data science community. In particular, it pulls from the following resources, which are highly recommended for further learning and as resources later on. Specific lessons will also cite more resources.

- R for Data Science by Hadley Wickham and Garrett Grolemund
- STAT 545 by Jenny Bryan
- Happy Git with R by Jenny Bryan
- Software Carpentry by the Carpentries

Chapter 3

readxl

Note for Allison: the `{#readxl}` above lets us reference it from another chapter...not sure how much we'll use it but it's possible (and doesn't hurt to have). Cool, thanks Julie!

3.1 Summary (a few sentences)

The **readxl** package makes it easy to import tabular data from Excel spreadsheets (.xls or .xlsx files) and includes several options for cleaning data during import. **readxl** has no external dependencies and functions on any operating system, making it an OS- and user-friendly package that simplifies getting your data from Excel into R.

3.2 Objectives (more detailed, bulletpoints?)

- Use `readr::read_csv()` to read in a comma separated value (CSV) file
- Use `readxl::read_excel()` to read in an Excel worksheet
- Use `readxl::read_excel()` to read in parts of a worksheet (by cell range)
- Specify column names when importing Excel data
- Replace a specific string/value in a spreadsheet with with 'NA'
- Skip *n* rows when importing an Excel worksheet
- Read in multiple Excel worksheets
- Something about readxl workflows (e.g. `%>% write_csv()` to keep a copy of the raw data in a project)??

3.3 Resources

- <https://readxl.tidyverse.org/>

3.4 Lessons teaching for each objective..... (objectives, examples)

3.4.1 a. Use `readr::read_csv()` to read in data from a CSV file

There are many types of files containing data that you might want to work with in R. A common one is a comma separated value (CSV) file, which contains values with each column entry separated by a comma delimiter. CSVs can be opened, viewed, and worked with in Excel just like an .xlsx or .xls file - but let's learn how to get data directly from a CSV into R where we can work with it more reproducibly.

First, create a new .R script and add a useful header to the top of it. For example, at a minimum:

```
# -----  
# A descriptive title  
# Summary of what this script is for  
# Your name  
# Contact information  
# -----  
  
# Other things you might include: required packages or datasets, relevant links (e.g.
```

3.5. *FUN FACTS (QUIRKY THINGS) - MAKING A NOTE OF THESE WHEREVER POSSIBLE FOR INTEREST*

3.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)

3.6 Interludes (deep thoughts/openscapes)

3.7 Our Turn Your Turn 1

3.8 Our Turn Your Turn 2

3.9 Efficiency Tips

Chapter 4

RMarkdown

- 4.1 Summary (a few sentences)
- 4.2 Objectives (more detailed, bulletpoints?)
- 4.3 Resources
- 4.4 Lessons teaching for each objective..... (objectives, examples)
- 4.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)
- 4.6 Interludes (deep thoughts/openscapes)
- 4.7 Our Turn Your Turn 1
- 4.8 Our Turn Your Turn 2
- 4.9 Efficiency Tips

Chapter 5

Dplyr and Pivot Tables

- 5.1 Summary (a few sentences)
- 5.2 Objectives (more detailed, bulletpoints?)
- 5.3 Resources
- 5.4 Lessons teaching for each objective..... (objectives, examples)
- 5.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)
- 5.6 Interludes (deep thoughts/openscapes)
- 5.7 Our Turn Your Turn 1
- 5.8 Our Turn Your Turn 2
- 5.9 Efficiency Tips

Chapter 6

Dplyr and vlookups

- 6.1 Summary (a few sentences)
- 6.2 Objectives (more detailed, bulletpoints?)
- 6.3 Resources
- 6.4 Lessons teaching for each objective..... (objectives, examples)
- 6.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)
- 6.6 Interludes (deep thoughts/openscapes)
- 6.7 Our Turn Your Turn 1
- 6.8 Our Turn Your Turn 2
- 6.9 Efficiency Tips

Chapter 7

Tidying

7.1 Better practices [needs a better name]

How to be a nimble useR Modern useRs are nimble internet useRs something clever about cleaning I am the worst at naming things

7.2 Summary (a few sentences)

R ecosystem evolves and improves due to contributed work by the community, and this is a good thing. Being a nimble useR means being able to navigate/keep tabs on this ecosystem and find what you need. It also means working reproducibly, so you can re-run and update things more easily. Here we will teach you how to expect things and help yourself. Pay attention to urls.

7.3 Objectives (more detailed, bulletpoints?)

- expect there is a better way, how and where to look (20 mins)
 - CRAN
 - Twitter #rstats
 - rOpenSci
 - RStudio
 - Example: how to Google.
- hands-on with janitor (30+ mins)
 - discovery and quality assurance
 - installing from GitHub
 - big payoff for little effort

- hands-on with another excel-useful example: `skimr`?
- reproducibility (20 mins)
 - it's important, scripted

7.4 Resources

- Wilson et al. 2014 “Good enough practices”

7.5 Lessons teaching for each objective..... (objectives, examples)

7.5.1 Expect there's a better way chat

- give time for them to google?

7.5.2 Janitor

janitor & other things that will make your life easier with limited effort Janitor: up till now the column names have been fine. Until now.

7.5.2.1 Our turn your turn

Walk through and example and leave our code up, and have you do it but clean another dataset. Work with a neighbor.

7.5.3 Example: How to Google

Pay attention to URLs, build github/rmarkdown savviness (ex: `raw.githubusercontent.com`)

- I read this blog: <https://blog.revolutionanalytics.com/2018/08/how-to-use-r-with-excel.html>
- I've never heard of click on `openxlsx`, what is it
- Takes me here <https://www.rdocumentation.org/packages/openxlsx/versions/4.1.0.1>, but I want more info. How recently was it worked on? Does it interface with tidyverse? Click on “news”
- Takes me here. <https://raw.githubusercontent.com/awalker89/openxlsx/master/NEWS> . Not useful. But from this URL, - I see the username so I can edit this url to be <https://github.com/awalker89/openxlsx/>

7.6. *FUN FACTS (QUIRKY THINGS) - MAKING A NOTE OF THESE WHEREVER POSSIBLE FOR INTEREST*

- 1st thing: most recent commit was a year ago. Can poke around more, are there issues open, are they taken care of? Etc. I will probably not pursue using this right now. But good to have learned about it.

7.6 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)

7.7 Interludes (deep thoughts/openscapes)

7.8 Our Turn Your Turn 2

7.9 Efficiency Tips

- browser efficiency tips
 - Rmd/github anchors for urls
 - press command to open a new tab

Reproducibility is important (this might be new to some people) Example: run everything start to finish and then closing it all and trying to do again In excel Vs R If your computer shuts off are you nervous to close it? Recreate it “What they didn’t forget to teach you about R” WTDF. uncool

Chapter 8

Formatting and Sharing

- 8.1 Summary (a few sentences)
- 8.2 Objectives (more detailed, bulletpoints?)
- 8.3 Resources
- 8.4 Lessons teaching for each objective..... (objectives, examples)
- 8.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)
- 8.6 Interludes (deep thoughts/openscapes)
- 8.7 Our Turn Your Turn 1
- 8.8 Our Turn Your Turn 2
- 8.9 Efficiency Tips

Chapter 9

Synthesis

- 9.1 Summary (a few sentences)
- 9.2 Objectives (more detailed, bulletpoints?)
- 9.3 Resources
- 9.4 Lessons teaching for each objective..... (objectives, examples)
- 9.5 Fun facts (quirky things) - making a note of these wherever possible for interest (little “Did you know?” sections)
- 9.6 Interludes (deep thoughts/openscapes)
- 9.7 Our Turn Your Turn 1
- 9.8 Our Turn Your Turn 2
- 9.9 Efficiency Tips