readxl

Jennifer Bryan
RStudio

@JennyBryan
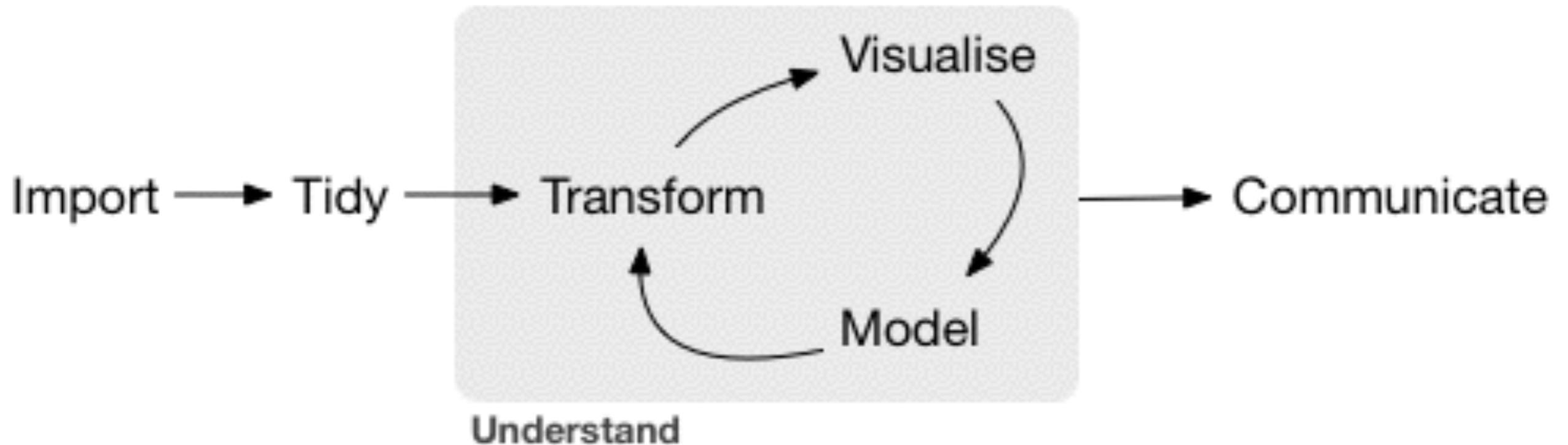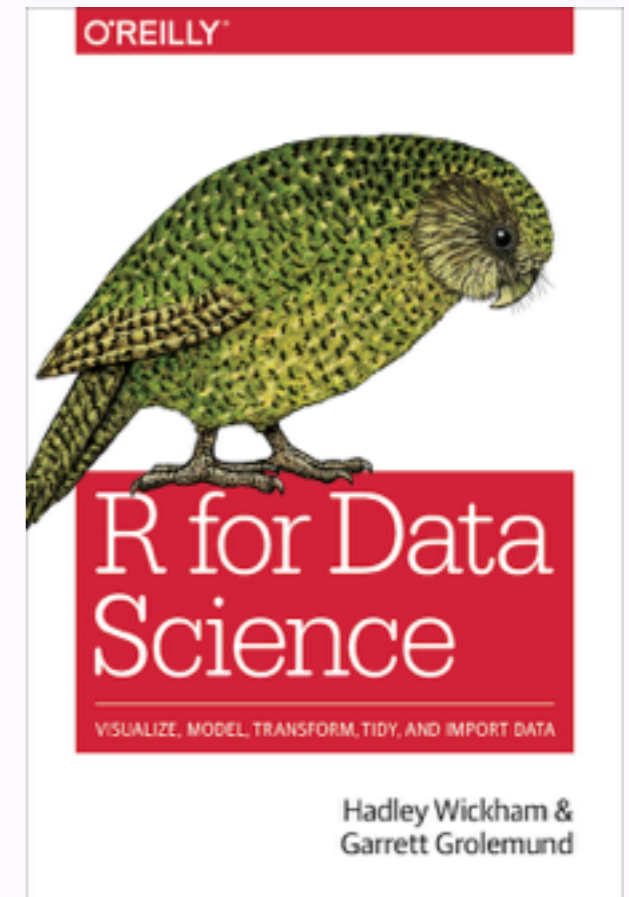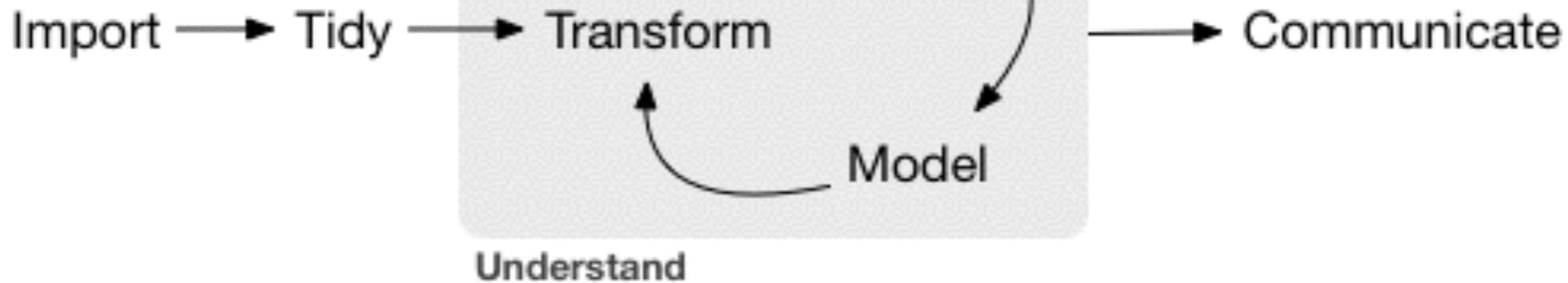@jennybc

readxl

www.rstudio.com

http://readxl.tidyverse.org

http://tidyverse.org

http://r4ds.had.co.nz

O'REILLY

R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Grolemund

Import → Tidy → Transform → Visualise → Model → Communicate

Understand

Program

2015 April

🎉 First readxl CRAN release!

v0.1.0, written by Hadley Wickham

2016 March

🤕 v0.1.1 patch release

2017 April

v1.0.0, 🙋 Jenny Bryan is new maintainer

🐞 lots of bug fixes, 🍾 some new features

```r
## install readxl + other packages
install.packages("tidyverse")

## install readxl alone
install.packages("readxl")

## regardless, load readxl like so:
library(readxl)
```
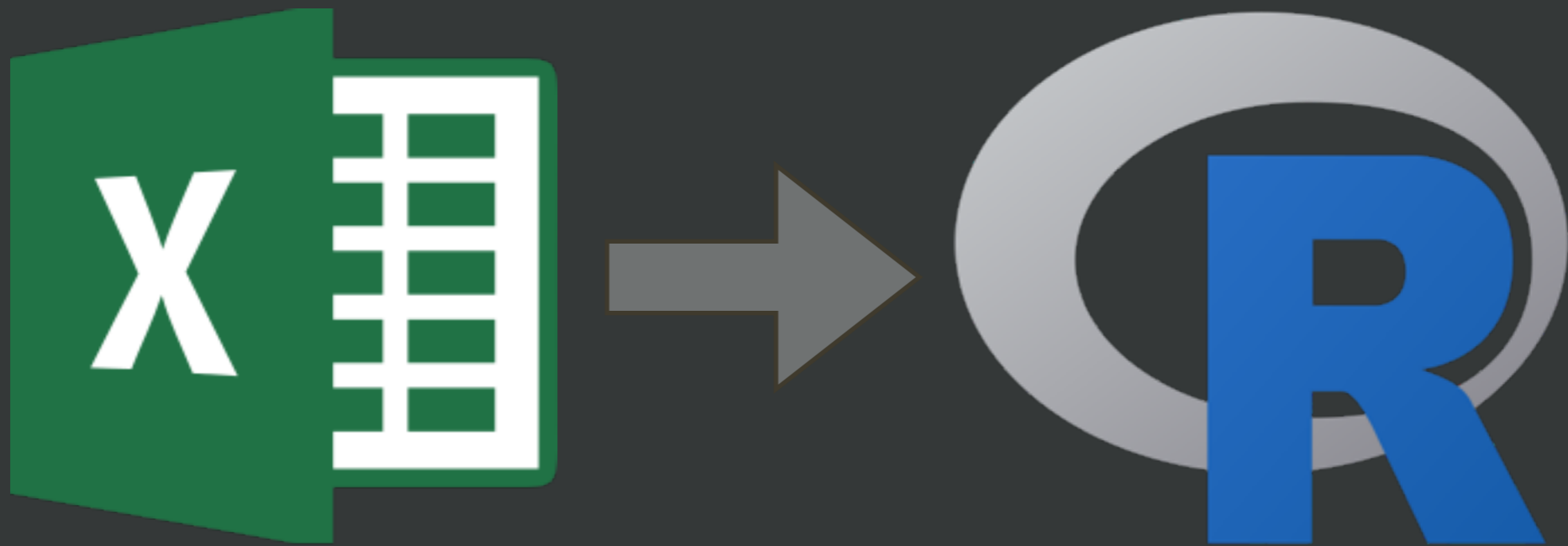
readxl, openxlsx, XLConnect, xlsx, gdata, RODBC, excel.link, and more!

🤔

.xls vs .xlsx

self-contained vs. needs Java/Perl

data (frame) vs cells & formatting

.xls vs .xlsx

self-contained vs. needs Java/Perl

data (frame) vs cells & formatting

	A	B	C	D	E	F
1	Lots of people					
2	simply cannot resist writing					some notes
3	at	the	top		of	their spreadsheet
4	or	merging				cells
5	Name	Profession	Age	Has kids	Date of birth	Date of death
6	David Bowie	musician	69	TRUE	1947-01-08	2016-01-10
7	Carrie Fisher	actor	60	TRUE	1956-10-21	2016-12-27
8	Chuck Berry	musician	90	TRUE	1926-10-18	2017-03-18
9	Bill Paxton	actor	61	TRUE	1955-05-17	2017-02-25
10	Prince	musician	57	TRUE	1958-06-07	2016-04-21
11	Alan Rickman	actor	69	FALSE	1946-02-21	2016-01-14
12	Florence Henderson	actor	82	TRUE	1934-02-14	2016-11-24
13	Harper Lee	author	89	FALSE	1926-04-28	2016-02-19
14	Zsa Zsa Gábor	actor	99	TRUE	1917-02-06	2016-12-18
15	George Michael	musician	53	FALSE	1963-06-25	2016-12-25
16	Some					
17		also like to write stuff				
18			at the	bottom,		
19						too!

1. readxl is all about the data rectangle

~/rrr/readxl - master - RStudio

Untitled1 — cell-and-column-types.Rmd

Environment History Build Git

Go to file/function — Addins

readxl

Import Excel Data

master

File/Url:

~/rrr/readxl/inst/extdata/deaths.xlsx

Update

Data Preview:

Lots of people (character)	X__1 (character)	X__2 (character)	X__3 (character)	X__4 (character)	X__5 (character)
simply cannot resist writing	NA	NA	NA	NA	some notes
at	the	top	NA	of	their spreadsheets
or	merging	NA	NA	NA	cells
Name	Profession	Age	Has kids	Date of birth	Date of death
David Bowie	musician	69	TRUE	17175	42379
Carrie Fisher	actor	60	TRUE	20749	42731
Chuck Berry	musician	90	TRUE	9788	42812
Bill Paxton	actor	61	TRUE	20226	42791
Prince	musician	57	TRUE	21343	42481

Previewing first 50 entries.

Import Options:

Name: deaths
Sheet: Default
Range: A1:D10

Max Rows:
Skip: 0
NA: Default

☑ First Row as Names
☑ Open Data Viewer

Code Preview:

```
library(readxl)
deaths <- read_excel("inst/extdata/deaths
.xlsx")
View(deaths)
```

? Reading Excel files using readxl

Import Cancel

Console

~/rrr/readx

R is a c
Type 'co
'citatio

Type 'de
'help.st
Type 'q(

>

Modified
Mar 29,
Mar 29,
May 8, 2
Nov 5, 2
Apr 17,
Apr 17,
Apr 17,
Apr 17,
Apr 17,
Apr 17,
May 9, 2

2. RStudio IDE has a nice Excel helper

```
> read_excel(readxl_example("deaths.xlsx"), range = cell_rows(5:15))
# A tibble: 10 × 6
                 Name Profession   Age `Has kids` `Date of birth`
                <chr>      <chr> <dbl>      <lgl>          <dttm>
 1       David Bowie   musician    69       TRUE      1947-01-08
 2     Carrie Fisher      actor    60       TRUE      1956-10-21
 3       Chuck Berry   musician    90       TRUE      1926-10-18
 4       Bill Paxton      actor    61       TRUE      1955-05-17
 5            Prince   musician    57       TRUE      1958-06-07
 6      Alan Rickman      actor    69      FALSE      1946-02-21
 7 Florence Henderson      actor    82       TRUE      1934-02-14
 8        Harper Lee     author    89      FALSE      1926-04-28
 9     Zsa Zsa Gábor      actor    99       TRUE      1917-02-06
10    George Michael   musician    53      FALSE      1963-06-25
# ... with 1 more variables: `Date of death` <dttm>
```
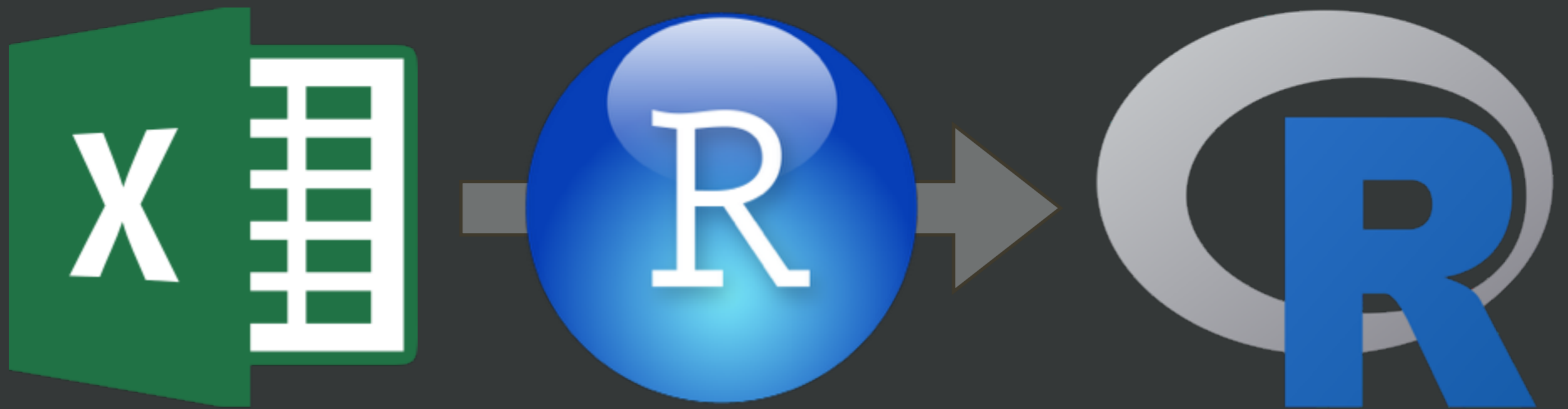
3. readxl's column typing is groovy

```r
path <- readxl_example("deaths.xlsx")
deaths <- path %>%
  excel_sheets() %>%
  set_names() %>%
  map_df(
    ~ read_excel(path = path, sheet = .x, range = "A5:F15"),
    .id = "sheet"
  )
```

4. workflows for multiple sheets and files

let's work together in the IDE …

1. readxl is all about the data rectangle

2. RStudio IDE has a nice Excel helper

3. readxl's column typing is groovy

4. workflows for multiple sheets and files

http://readxl.tidyverse.org