

# rbtl - Research Beyond the Lab

## Reference management with Zotero

Lars Schöbitz

2022-03-31

# Today

1. Homework Assignment 2
2. Reference Management - Zotero
3. Open Source - Licenses
4. Reproducible Research
5. Homework Assignment 3

# Homework Assignment 2

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator) %>%  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

# Homework Assignment 2

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator) %>%  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

- Objects that store dataframes: `dat_in` and `dat_in_sum_day`

# Homework Assignment 2

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator) %>%  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

- Objects that store dataframes: `dat_in` and `dat_in_sum_day`
- Functions: `filter()`, `mutate()`, `as_date()`, `group_by()`, `summarise()`, etc.

# Homework Assignment 2

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator) %>%  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

- Objects that store dataframes: `dat_in` and `dat_in_sum_day`
- Functions: `filter()`, `mutate()`, `as_date()`, `group_by()`, `summarise()`, etc.
- Assignment operator: `<-`

# Homework Assignment 2

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator) %>%  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

- Objects that store dataframes: `dat_in` and `dat_in_sum_day`
- Functions: `filter()`, `mutate()`, `as_date()`, `group_by()`, `summarise()`, etc.
- Assignment operator: `<-`
- Pipe operators: `%>%`

# Homework Assignment 2 - Imported raw data

```
dat_link <- "https://raw.githubusercontent.com/Global-Health-Engineering/manuscri
dat_in <- read_csv(dat_link)
dat_in
```

```
# A tibble: 203,806 × 6
  date_time      id location indicator value unit
<dtm>         <chr> <chr>    <chr>    <dbl> <chr>
1 2019-10-08 13:59:01 hos1 guardian pm2.5    19.4 uq_m3
2 2019-10-08 13:59:01 hos1 guardian pm10     27 uq_m3
3 2019-10-08 14:04:41 hos1 guardian pm2.5    44.9 uq_m3
4 2019-10-08 14:04:41 hos1 guardian pm10    56.7 uq_m3
5 2019-10-08 14:10:21 hos1 guardian pm2.5   202. uq_m3
6 2019-10-08 14:10:21 hos1 guardian pm10   240. uq_m3
# ... with 203,800 more rows
```



# Summarised derived data

```
dat_in_sum_day <- dat_in %>%  
  filter(value <= 1000) %>%  
  mutate(date = as_date(date_time)) %>%  
  group_by(date, location, indicator)  
  summarise(min = min(value),  
            median = median(value),  
            mean = mean(value),  
            sd = sd(value),  
            max = max(value))
```

```
# A tibble: 890 × 8  
# Groups:   date, location [445]  
  date          location indicator    min  
  <date>         <chr>      <chr>    <dbl>  
1 2019-10-01 Lhouse    pm10      22.8  
2 2019-10-01 Lhouse    pm2.5     12.2  
3 2019-10-02 Lhouse    pm10      24.9  
4 2019-10-02 Lhouse    pm2.5     12.8  
5 2019-10-02 Lions      pm10       7.5  
6 2019-10-02 Lions      pm2.5      4.6  
# ... with 884 more rows
```

# Reference Management

# Why?

- You will read a lot
- You want to stay organized
- You don't want to waste your time on formatting

# Which tool?

- Mendeley
- EndNote
- Zotero
- many, many more

# Which tool?

- Mendeley
- EndNote
- **Zotero**
- many, many more

# Why Zotero?



## Rest easy.

Zotero is [open source](#) and developed by an independent, nonprofit organization that has no financial interest in your private information. With Zotero, you always stay in control of your own data.

# Zotero is Open Source - Why is that good?

- Free Software
- Transparent about access to your own data
- The source code that Zotero is developed in is public
- Commitment to support open software and open standards
- Zotero developers helped create the open [Citation Style Language \(CSL\)](#)

# Open Source - Licenses

- Open Source isn't just code on the internet
- Use permissive licenses to allow others to reuse, remix and build upon (also for commercial purposes)
- Recommended licenses
  - **Text, slides, images:** Creative Commons (CC0, CC-BY, CC-BY-SA)
  - **Software:** MIT License, Hippocratic License, Unlicense for software
- <https://tldrlegal.com/> - plain english explanations of licences in bullet form.
- <https://kbroman.org/steps2rr/pages/licenses.html> - Read Karl Broman

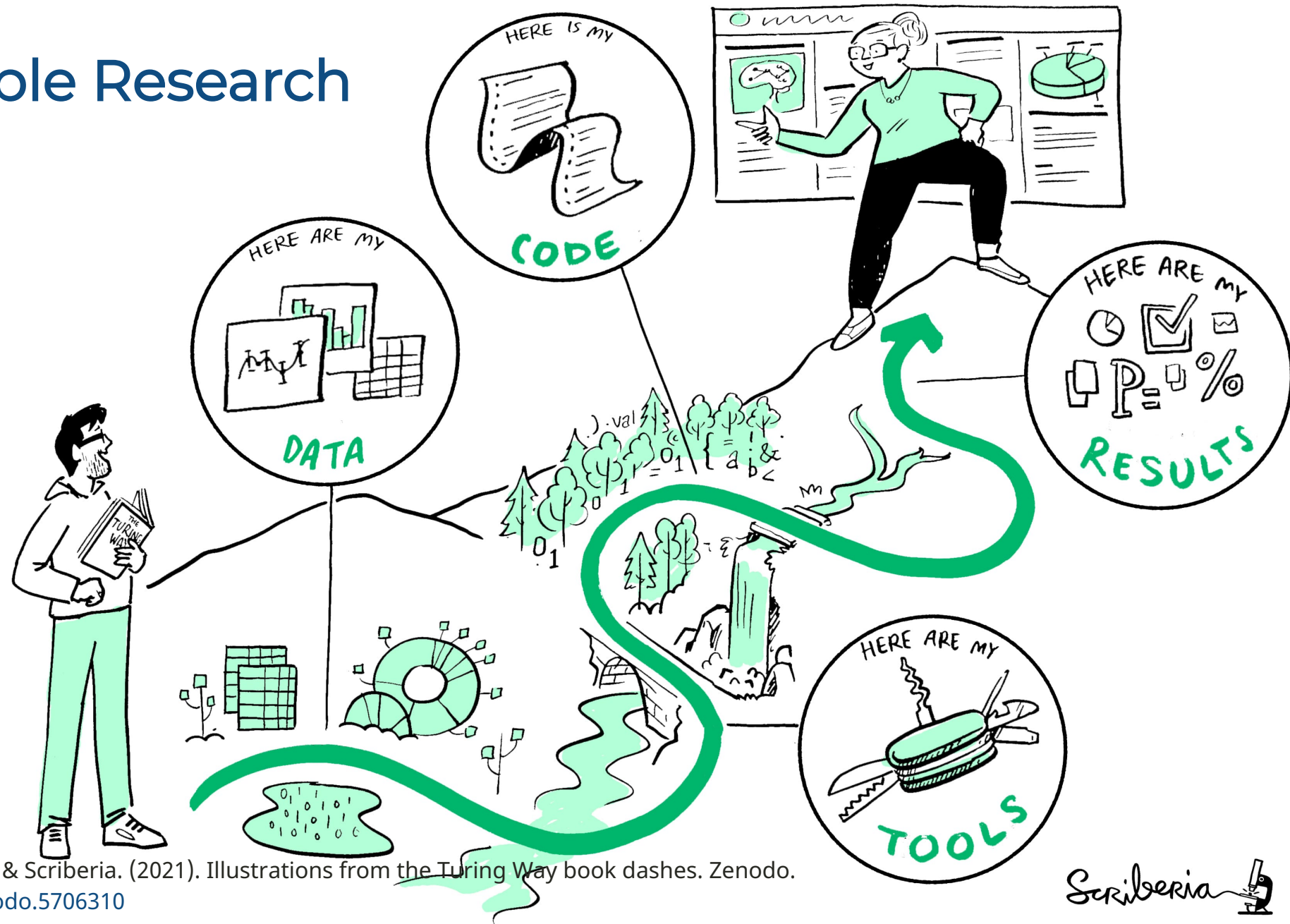


Open Source != Open Access

Open Source != Open Data

Open Source (Code) + Open Data =  
Reproducible Research

# Reproducible Research



The Turing Way Community, & Scriberia. (2021). Illustrations from the Turing Way book dashes. Zenodo.  
<https://doi.org/10.5281/zenodo.5706310>

Scriberia

# Homework Assignment

# Homework Assignment 3 - Learning Objectives

These learning objectives are related to the assignment for this week.

- Learners are able to import references to a Zotero group library
- Learners can use an exported library from Zotero in Better BibTex Format to generate an automated reference list in an R Markdown file
- Learners can edit a file in the Citation File Format (.cff) to add their name to the author list

# Homework Assignment 3 - Due Date

- Complete Assignment 2 before you complete Assignment 3
- Assignment 3, due on 15th March
- Readings on Reproducible Research

Thanks! 

Slides created via the R packages:

**xaringan**

[gadenbuie/xaringanthemer](#)

The chakra comes from [remark.js](#), **knitr**, and R Markdown.

Access slides as [PDF on GitHub](#)

All material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).