

Introduction to R Markdown

CECL Workshop

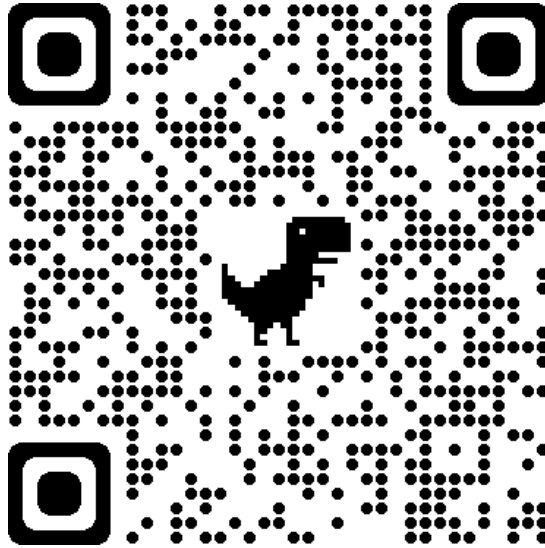
Nathan Vandeweerd

March 29th, 2021

Workshop Materials

All workshop materials (including these slides) are available on GitHub:

https://github.com/nvandeweerd/intro_rmarkdown



Replication and Reproducibility

"The replication of research results is a linchpin of the scientific process." (Nature, 2006)

"To facilitate such replication, researchers would typically be expected to **report the methodology used in some detail**. What can at first appear to be a minor feature hidden away in the "limitations" section of a paper might -- after replicating the study a number of times and in a number of ways -- turn out to be a key factor impacting the outcomes obtained." (Porte, 2012 emphasis added)



Recommendations for replication studies in Applied Linguistics

(Marsden, Morgan-Short, Thompson, and Abugaber, 2018)

- Increase the number of replication studies and the rate at which they are performed and published.
- Make systematic inquiry into the causes of low rates of published replication studies...
- Use more self-labeling with the term replication wherever appropriate.
- Apply a principled, standard nomenclature...
- Reviewers of replications should also read the initial study that is being replicated.
- Provide warrants for replication studies...
- **Increase open availability of materials, including proficiency measures, for L2 research.**
- **Make more research fully transparent and open for replication by making data available.**
- **Encourage more journals to give more and stronger incentives to their authors for systematically making materials and data openly available.**
- Ensure that replication studies are conducted by researchers independently of the initial study's authors...
- Increase multisite collaborative replication efforts.
- Encourage journal editorial boards to consider accepting Registered Reports...
- Encourage publishers to lift word limits or provide online capacity.
- When the initial study is cited, also cite replication studies.
- Increase funding to promote replication.

What is R Markdown?

- Markdown (Gruber, 2004): mark-up language to convert plain text to HTML
- The `knitr` package: a method for embedding R chunks into markdown documents
- The `pandoc` package: enhances Markdown (adds support for tables, footnotes, math expressions etc.)
- The `rmarkdown` package: combines all of the above into a single document format which can be *knit* into various output formats (pdf, HTML, Word, powerpoint etc.)

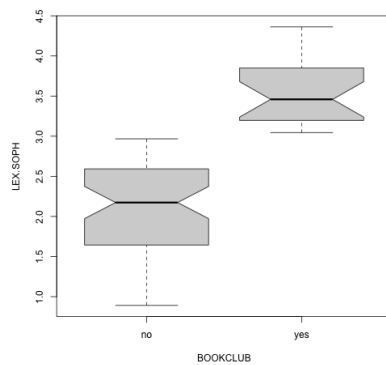


Transparent Analysis

```
data <- read.csv("../data/stay-abroad.csv")
data <- data[data$ABROAD == "yes",]
print(head(data))
```

```
##      ID ABROAD BOOKCLUB  PLACEMENT LEX.SOPH ECON.VOCAB VOCAB.TEST
## 2  102    yes      no    erasmus  2.487593   273.6371   85.98986
## 3  103    yes      no    business  0.970495   215.1088   80.58794
## 4  104    yes      no    erasmus  1.717259   324.6748   67.55379
## 6  106    yes      no lang_assist  1.377997   303.7077   50.08712
## 7  107    yes      no    business  1.443990   238.0487   59.67121
## 9  109    yes      no lang_assist  1.066004   229.0751   74.10126
```

```
boxplot(LEX.SOPH ~ BOOKCLUB, data = data, notch = TRUE)
```



```
t.test(data$LEX.SOPH ~ data$BOOKCLUB)
```

```
##
##      Welch Two Sample t-test
##
## data:  data$LEX.SOPH by data$BOOKCLUB
## t = -12.578, df = 56.78, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.733163 -1.257069
## sample estimates:
##  mean in group no mean in group yes
##           2.075463           3.570579
```

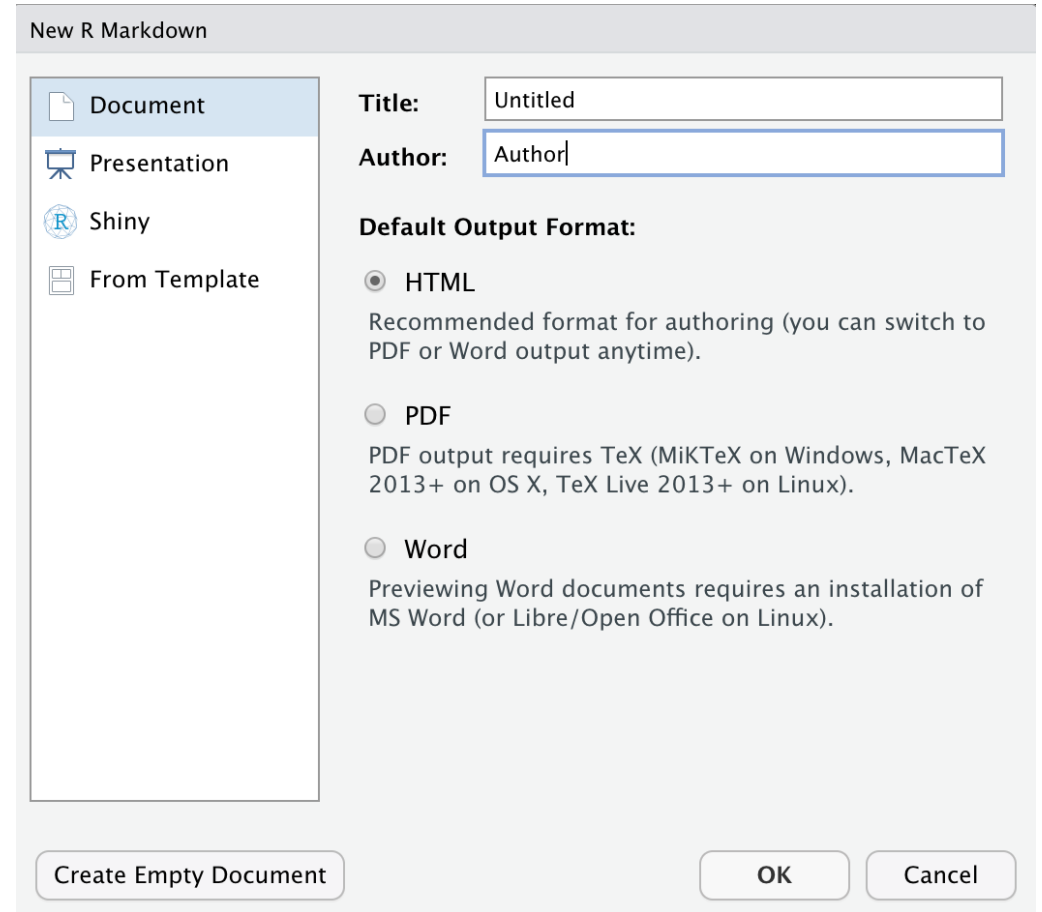
Getting started

```
# Install most recent version
install.packages("rmarkdown")

# Latex engine for PDFs
install.packages("tinytex")
```

In R Studio:

File >> New File >> R Markdown...



New R Markdown

☒ Document
☐ Presentation
☐ Shiny
☐ From Template

Title:

Author:

Default Output Format:

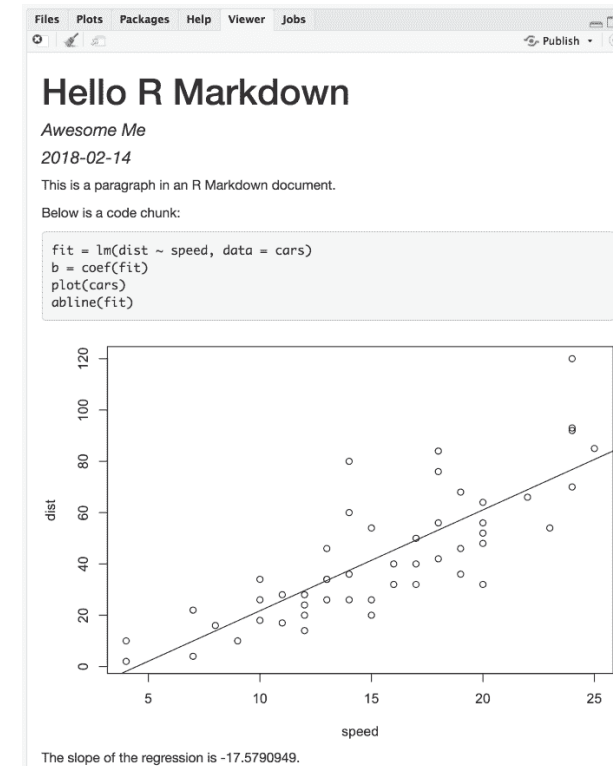
☒ **HTML**
Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ **PDF**
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ **Word**
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

Anatomy of an R Markdown file

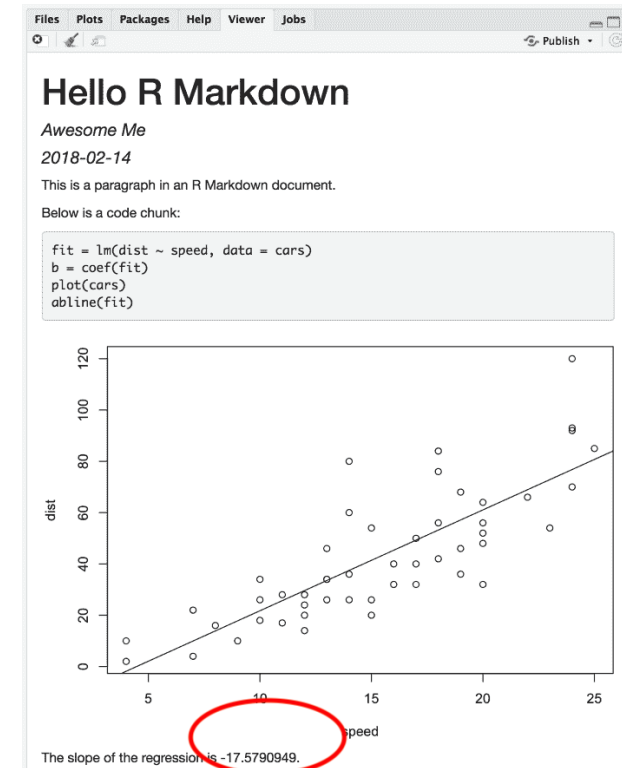
```
1 ---
2 title: "Hello R Markdown"
3 author: "Awesome Me"
4 date: "2018-02-14"
5 output: html_document
6 ---
7
8 This is a paragraph in an R Markdown document.
9
10 Below is a code chunk:
11
12 ```{r}
13 fit = lm(dist ~ speed, data = cars)
14 b = coef(fit)
15 plot(cars)
16 abline(fit)
17 ```
18
19 The slope of the regression is `r b[1]`.
20 |
```



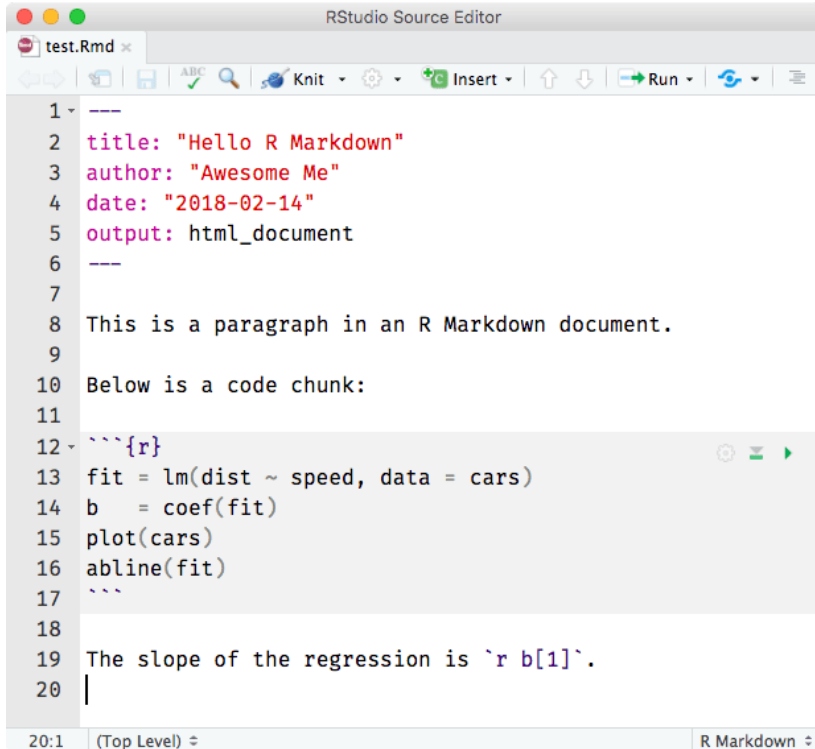
Did you spot the error?

Accountability

```
RStudio Source Editor
test.Rmd x
1 ---
2 title: "Hello R Markdown"
3 author: "Awesome Me"
4 date: "2018-02-14"
5 output: html_document
6 ---
7
8 This is a paragraph in an R Markdown document.
9
10 Below is a code chunk:
11
12 ```{r}
13 fit = lm(dist ~ speed, data = cars)
14 b = coef(fit)
15 plot(cars)
16 abline(fit)
17 ```
18
19 The slope of the regression is `r b[1]`.
20
```



Anatomy of an R Markdown file



```
1 ---
2 title: "Hello R Markdown"
3 author: "Awesome Me"
4 date: "2018-02-14"
5 output: html_document
6 ---
7
8 This is a paragraph in an R Markdown document.
9
10 Below is a code chunk:
11
12 ```{r}
13 fit = lm(dist ~ speed, data = cars)
14 b = coef(fit)
15 plot(cars)
16 abline(fit)
17 ```
18
19 The slope of the regression is `r b[1]`.
20 |
```

YAML Header

- "YAML Ain't Markup Language"
- Generated automatically when you use wizard
- Contains:
 - file metadata
 - output format
 - instructions for **knitting** the document

Markdown formatted text

Code chunks

- Three backticks followed by the language (R) in curly brackets

Inline code

- A single backtick followed by the language (without curly brackets)

YAML Header

- file metadata:
 - title, subtitle, author, date, abstract
- output file:
 - "html_document"
 - "pdf_document" (requires Tex)
 - "word_document"
- instructions for knitting:
 - styling information (page numbers, table of contents, font size)
 - path to a template document (handy for working with Microsoft Word)
 - path to a citation file (and citation format file)

Markdown formatted text

```
# Header Level 1
## Header Level 2
### Header Level 3
Text surrounded by a single asterisks is *italic*
Text surrounded by two asterisks **bold**
- Single dash renders as bullet points
  - item
  - item
- Numbered lists
  1. thing
  2. thing

hyperlinks: [R Markdown Cheatsheet](https://www.rstudio.org/links/r
>block quotes
```

Note: You must put two spaces at the end of a line if you want a line break.

Header Level 1

Header Level 2

Header Level 3

Text surrounded by a single asterisks is *italic*

Text surrounded by two asterisks is **bold**

- Single dash renders as bullet points
 - item
 - item
 - Numbered lists
 1. thing
 2. thing
- hyperlinks: R Markdown Cheatsheet

block quotes

Code chunks

- Need to specify the language (r)
- Options can be specified after the language
 - **echo**: should the code be printed? (default = TRUE)
 - **eval**: should the code be run? (default = TRUE)

```
```{r, echo = TRUE, eval = TRUE}  
mean(data$LEX.SOPH)
```
```

What will this produce in the output document?

```
mean(data$LEX.SOPH)
```

```
## [1] 2.497163
```

What about this?

```
```{r, echo = TRUE, eval = FALSE}  
mean(data$LEX.SOPH)
```
```

```
mean(data$LEX.SOPH)
```

Code chunks

- Need to specify the language (r)
- Options can be specified after the language
 - **echo**: should the code be printed? (default = TRUE)
 - **eval**: should the code be run? (default = TRUE)

What about this?

```
```{r, echo = FALSE, eval = TRUE, comment = '(:):)'}  
mean(data$LEX.SOPH)
```
```

```
(:):) [1] 2.497163
```



Inserting graphics

```
```{r}  
knitr::include_graphics('cecl.png')
```
```



Inserting graphics

```
```{r, out.width='10%', fig.cap = 'Figure 1. CECL logo', fig.align='center'}  
knitr::include_graphics('cecl.png')
```
```



Figure 1. CECL logo

Options:

- `out.width`: changes the size of the image
- `fig.cap`: adds a caption below the image
- `fig.align`: alignment of the image

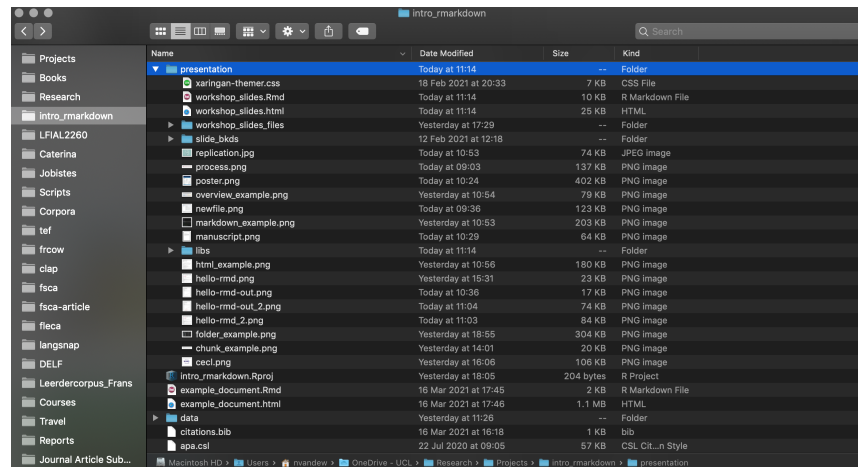


A note about file management

- By default the "working directory" of an R Markdown file is the folder where it is located
- File paths (e.g. for images, data) must be specified *relative* to the working directory
- Easiest to include all necessary files in the same folder as the R Markdown folder
- Also useful to work in "projects" (the [here](#) package is very useful for this)

Note: An R Markdown document also uses its own "environment" so packages need to be loaded using `library()` in a code chunk **or** explicitly called using double colons

```
knitr::include_graphics("folder_example.png")
```



Tables

How can we turn the raw data frame into a nicely formatted table?

```
text.lengths <- read.csv("../data/lsnp.info.csv")  
print(text.lengths)
```

| ## | TASK | min | median | mean | max |
|------|----------------|-----|--------|---------|------|
| ## 1 | lgbt_essay | 110 | 213.0 | 212.72 | 306 |
| ## 2 | drugs_essay | 143 | 209.5 | 208.00 | 279 |
| ## 3 | food_essay | 155 | 212.0 | 209.91 | 298 |
| ## 4 | cat_story | 130 | 335.0 | 349.72 | 842 |
| ## 5 | sisters_story | 211 | 402.5 | 410.27 | 916 |
| ## 6 | brothers_story | 123 | 265.0 | 297.17 | 778 |
| ## 7 | interview | 282 | 1021.0 | 1231.45 | 4158 |

Tables

The `kable()` function from the **knitr** package can be used to convert raw data frames to markdown format

```
```{r}
text.lengths <- read.csv('../data/lsnp.info.csv')
knitr::kable(text.lengths, booktabs = TRUE, caption = 'Text lengths by task')
```
```

Table: Text lengths by task

| TASK | min | median | mean | max |
|----------------|-----|--------|---------|------|
| lgbt_essay | 110 | 213.0 | 212.72 | 306 |
| drugs_essay | 143 | 209.5 | 208.00 | 279 |
| food_essay | 155 | 212.0 | 209.91 | 298 |
| cat_story | 130 | 335.0 | 349.72 | 842 |
| sisters_story | 211 | 402.5 | 410.27 | 916 |
| brothers_story | 123 | 265.0 | 297.17 | 778 |
| interview | 282 | 1021.0 | 1231.45 | 4158 |

Tables

For more flexibility (and for non HTML documents), try the `flextable` package

```
```{r}
library(flextable)
library(magrittr)
text.lengths <- read.csv('../data/lsnp.info.csv')
flextable(text.lengths) %>%
 bold(part = 'header') %>%
 bg(i = ~ median > 300, bg = 'wheat') %>%
 set_caption('Table 1. Text lengths by task')
```
```

Table 1. Text lengths by task

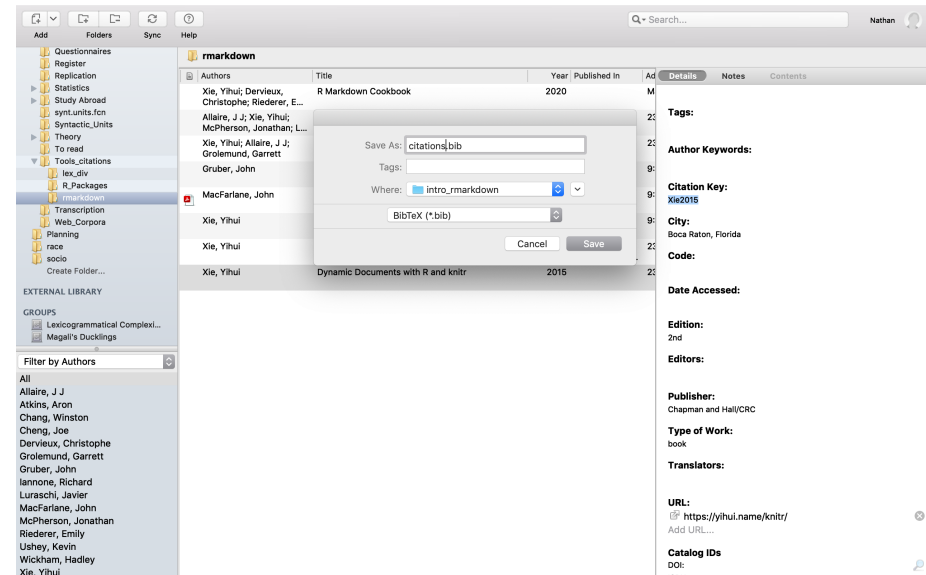
| TASK | min | median | mean | max |
|----------------|-----|---------|---------|-------|
| lgbt_essay | 110 | 213.0 | 212.7 | 306 |
| drugs_essay | 143 | 209.5 | 208.0 | 279 |
| food_essay | 155 | 212.0 | 209.9 | 298 |
| cat_story | 130 | 335.0 | 349.7 | 842 |
| sisters_story | 211 | 402.5 | 410.3 | 916 |
| brothers_story | 123 | 265.0 | 297.2 | 778 |
| interview | 282 | 1,021.0 | 1,231.5 | 4,158 |

Citations

In order to include citations, you first need to add two lines to the YAML header:

- a path to a **.bib** file containing citations in Bibtex format (can be exported from Mendeley or Zotero)
- a path to a **.csl** file containing citation formatting style guidelines (can be downloaded from the [Zotero Style Repository](#))
- easiest if you store both of these in the same folder as the R Markdown file

```
---  
title: "Introduction to R Markdown"  
author: "Nathan Vandeweerd"  
date: "`r Sys.Date()`"  
output: html_document  
bibliography: citations.bib  
csl: apa.csl  
---
```



Citations

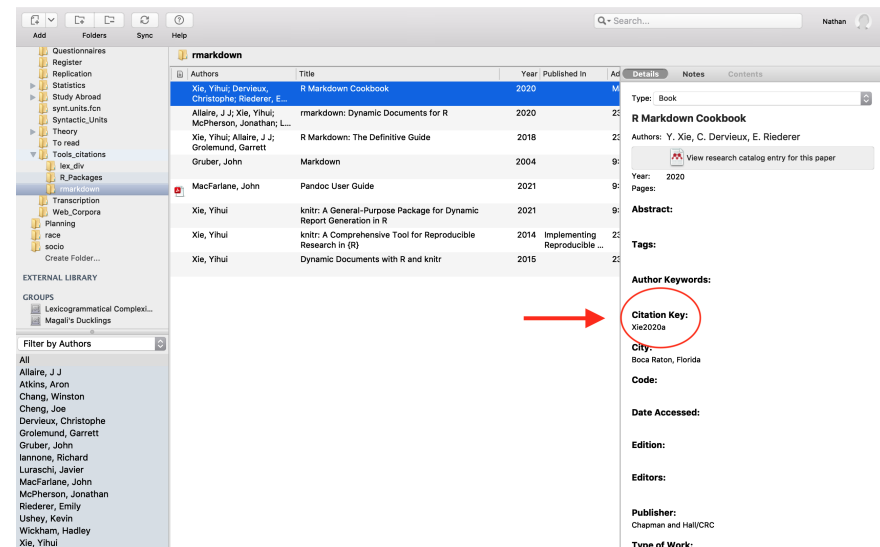
Once you have linked to a **.bib** file and a **.csi** file, you can include in-text citations using the format below:

- Multiple citations can be separated with a semi-colon
- A minus sign (-) means that only the year will be included

[@CitationKey]; @CitationKey2] --> (Author, 2005; Author, 2006)
[-@CitationKey1] --> (2005)

The Citation Key is **generated automatically** by reference tools such as Mendeley and Zotero (though you may need to change the settings to make it visible).

A bibliography will automatically be included at the end of the document which contains all of the cited references. To turn this off, simply add the following line to the YAML header:
suppress-bibliography: true



Want to learn more?

- R Markdown: The Definitive Guide
- R Markdown Cookbook

Other Useful packages

- `knitr` (included with `rmarkdown` package): working with code chunks, inserting graphics, rendering basic tables
- `officer` and `officetools`: for working with Microsoft Office
- `xaringan`: for HTML presentations (like this one)
- `flextable`: more customizable tables
- `bookdown`: for long documents (i.e. dissertations)





Q & Eh?



Practice

Download the following files from GitHub (remember to save them all in the same folder):

- `example_document.Rmd`
- `apa.csl`
- `citations.bib`
- `cecl.png`
- `lsnp.info.csv`

To do:

1. make section headers:
 - abstract (first level)
 - topics covered (second level)
2. make bullet points for the topics covered
3. add a code chunk that calculates 1+1 (show both the code and the result)
4. add inline code that calculates 1+1
5. add the CECL Logo (in a reasonable size)
6. add a table showing the lengths of the langsnap texts by task (`lsnp.info.csv`)
7. add a barplot (`barplot()`) of the same data
8. add a hyperlink to the R Markdown webpage
9. make the second paragraph a block quote
10. add the citations for the R Markdown package

References

Gruber, J. (2004). *Markdown*. URL: <https://daringfireball.net/projects/markdown/>.

Marsden, E., K. Morgan-Short, et al. (2018). "Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field". In: *Language Learning* 68.2, pp. 321-391.

Nature (2006). "Let's Replicate". In: *Nature* 442.7101, p. 330.

Porte, G. (2012). *Replication research in applied linguistics*. Cambridge: Cambridge University Press.

