

# Introduction to R Markdown

Kristen Sauby, ksauby@gmail.com

26 March 2019

Find this presentation (made with R Markdown) and more here:

<https://github.com/ksauby/R-Markdown-Introduction>

# Overview



## .Rmd files

An R Markdown (.Rmd) file is a record of your research. It contains the code that a scientist needs to reproduce your work along with the narration that a reader needs to understand your work.



## Reproducible Research

At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

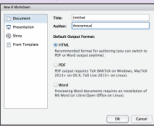


## Dynamic Documents

You can choose to export the finished report as a html, pdf, MS Word, ODT, RTF, or markdown document; or as a html or pdf based slide show.

## Workflow

- 1 Open a new .Rmd file** at File ► New File ► R Markdown. Use the wizard that opens to pre-populate the file with a template



- 2 Write document** by editing template

- 3 Knit document to create report** Use knit button or `render()` to knit

- 4 Preview Output** in IDE window

- 5 Publish** (optional) to web or server

Synch publish button to accounts at

- [rpubs.com](https://rpubs.com),
- [shinyapps.io](https://shinyapps.io)

RStudio Connect  
Reload document  
Find in document  
File path to output document

- 6 Examine build log** in R Markdown console

- 7 Use output file** that is saved alongside .Rmd

### .Rmd structure

#### YAML Header

Optional section of render (e.g. pandoc) options written as key:value pairs (YAML).

- At start of file
- Between lines of ---

#### Text

Narration formatted with markdown, mixed with:

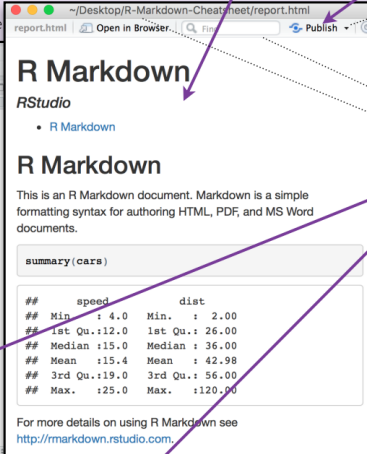
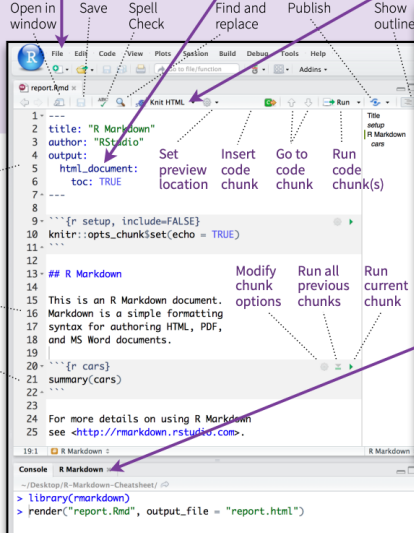
#### Code chunks

Chunks of embedded code. Each chunk:

- Begins with `{r}`
- ends with `}`

R Markdown will run the code and append the results to the doc.

It will use the location of the .Rmd file as the **working directory**



### render()

Use `rmarkdown::render()` to render/knit at cmd line. Important args:

**input** - file to render

**output\_format**

**output\_options** - List of render options (as in YAML)

**output\_file**

**output\_dir**

**params** - list of params to use

**envir** - environment to evaluate code chunks in

**encoding** - of input file


# Output to PDF

```
1  ---
2  title: A Sampling Strategy Designed to Maximize the Efficiency of Data
3  Collection of Secondary Information About Items of Interest
4  author: Kristen E. Sauby and Mary C. Christman
5  output:
6    pdf_document:
7      includes:
8        in_header: header.tex
9        number_sections: true
10 bibliography: experiment.bib
11 csl: methods-in-ecology-and-evolution.csl
12 ---
```

# Output to Word

```
1 ---
2 title: A Sampling Strategy Designed to Maximize the Efficiency of Data
3       Collection of Secondary Information About Items of Interest
4 author: Kristen E. Sauby and Mary C. Christman
5 output:
6   word_document:
7     reference_docx: "Manuscript_style.docx"
8 bibliography: experiment.bib
9 csl: methods-in-ecology-and-evolution.csl
10 ---
```

1 The Relative Importance of Herbivory and Abiotic Conditions to Demographic  
2 Rates of two Species of *Opuntia* Cacti in Florida [a1](#)

3 **Kristen E. Sauby, John Kilmer, Mary C. Christman, Robert D.**  
4 **Holt, and Travis D. Marsico** 

## 5 Introduction

## 6 Methods

## 7 Study System

8 Statistical Analysis

## 9 Relative Growth Rate

## 10 Fruiting Probability and Abundance

# Output to multiple formats

```
1 ---
2 title: A Sampling Strategy Designed to Maximize the Efficiency of Data
3       Collection of Secondary Information About Items of Interest
4 author: Kristen E. Sauby and Mary C. Christman
5 output:
6   word_document:
7     reference_docx: "Manuscript_style.docx"
8   pdf_document:
9     includes:
10      in_header: header.tex
11      number_sections: true
12 bibliography: experiment.bib
13 csl: methods-in-ecology-and-evolution.csl
14 ---
15
```



## Citations and Bibliographies

Create citations with .bib, .bibtex, .copac, .enl, .json, .medline, .mods, .ris, .wos, and .xml files

- 1 **Set bibliography file** and CSL 1.0 Style file (optional) in the YAML header

```
---  
bibliography: refs.bib  
csl: style.csl  
---
```

- 2 **Use citation keys in text**

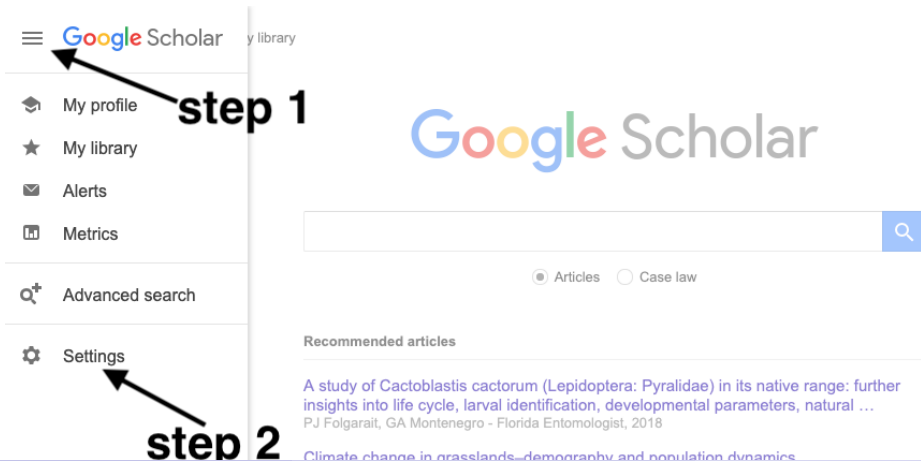
```
Smith cited [@smith04].  
Smith cited without author [-@smith04].  
@smith04 cited in line.
```

- 3 **Render.** Bibliography will be added to end of document

```
Smith cited (Joe Smith 2004).  
Smith cited without author (2004).  
Joe Smith (2004) cited in line.
```

# Set up Google Scholar, part 1

- Set up Google Scholar so that it shows your BibTeX formatting for each citation
- This assumes that you have a Google account



# Set up Google Scholar, part 2

Google Scholar

Settings

Search results

Languages

Library links

Account

Button

Collections

☒ Search articles (☒ include patents).

☐ Search case law.

Results per page

10

Google's default (10 results) provides the fastest results.

Where results open

☐ Open each selected result in a new browser window

Bibliography manager

☐ Don't show any citation import links.

☒ Show links to import citations into BibTeX

Save

Cancel

step 3

To retain settings, you must turn on cookies

# Example - Addition Citations to your .bib

- Now we can look up articles on Google Scholar and *copy and paste* the BibTeX citation to my “bibliography.bib” file
- Rarely do I have to write the BibTeX citation from scratch

The screenshot shows the Google Scholar search results for the query "p hacking". The search bar at the top contains "p hacking" and a magnifying glass icon. Below the search bar, the results are listed. The first result is "Life after p-hacking" by JP Simmons, LD Nelson, and U Simonsohn, published in 2013. The abstract mentions "Electronic copy available at: http://ssrn.com/abstract=2205186". The second result is "The extent and consequences of p-hacking in science" by ML Head, L Holman, R Lanfear, AT Kahn, et al., published in 2015. The abstract mentions "A focus on novel, confirmatory, and statistically significant results leads to substantial bias in the scientific literature." The third result is "Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015)." by U Simonsohn, JP Simmons, and LD Nelson, published in 2015. The abstract mentions "When studies examine true effects, they generate right-skewed p-curves, distributions of statistically significant results with more low (.01 s) than high (.04 s) p values." The fourth result is "P-curve and p-hacking in observational research" by SB Bruns, JPA Ioannidis, et al., published in 2016. The abstract mentions "The P-curve, the distribution of statistically significant p-values of published studies, has been used to make inferences on the proportion of true effects and on the presence of p-hacking." The "Import into BibTeX" link for the first result is highlighted with a red box.

Google Scholar p hacking

Articles About 165,000 results (0.04 sec)

Any time Since 2019 Since 2018 Since 2015 Custom range...

Sort by relevance Sort by date

☒ include patents ☒ include citations

☒ Create alert

**Life after p-hacking** [PDF] acrwebsite.org  
JP Simmons, LD Nelson, U Simonsohn - 2013 - papers.ssrn.com  
Page 1. Electronic copy available at: http://ssrn.com/abstract=2205186 Life After **P-Hacking** Joe Simmons (Wharton) Leif Nelson (Berkeley) Uri Simonsohn (Wharton) Presented at SPSP 2013 Page 2. Electronic copy available at: http://ssrn.com/abstract=2205186 We have decided not to ...  
☆ Cited by 135 Related articles All 3 versions Import into BibTeX

**The extent and consequences of p-hacking in science** [HTML] plos.org  
ML Head, L Holman, R Lanfear, AT Kahn... - PLoS ..., 2015 - journals.plos.org  
A focus on novel, confirmatory, and statistically significant results leads to substantial bias in the scientific literature. One type of bias, known as "**p-hacking**," occurs when researchers collect or select data or statistical analyses until nonsignificant results become significant ...  
☆ Cited by 335 Related articles All 25 versions Import into BibTeX

**Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015).** [PDF] upenn.edu  
U Simonsohn, JP Simmons, LD Nelson - 2015 - psycnet.apa.org  
When studies examine true effects, they generate right-skewed **p**-curves, distributions of statistically significant results with more low (.01 s) than high (.04 s) **p** values. What else can cause a right-skewed **p**-curve? First, we consider the possibility that researchers report only ...  
☆ Cited by 112 Related articles All 11 versions Import into BibTeX

**P-curve and p-hacking in observational research** [HTML] plos.org  
SB Bruns, JPA Ioannidis - PLoS one, 2016 - journals.plos.org  
The **p**-curve, the distribution of statistically significant **p**-values of published studies, has been used to make inferences on the proportion of true effects and on the presence of **p**-

# Why Use R Markdown for your Citations/Bibliography?

- formats citations according to format of your choosing
- compiles bibliography for you
- when you re-compile your R Markdown document, the bibliography will be recreated as well
  - ensures that bibliography is up-to-date
  - no extras, no missing references

# Citation Styles

← → ↻ 🔒 GitHub, Inc. [US] | https://github.com/citation-style-language/styles ☆ ABP ⓘ 🌈 📌 👤 ⚙️ 🏠 🟢 🍷 📁 📧 | 🖱️ ⋮

🐙 Search or jump to... / Pull requests Issues Marketplace Explore 🔔 + 👤

 [citation-style-language / styles](#)

 Watch ▼

78

★ Star

1.605

 Fork

2.200

&lt;&gt; Code

Issues 30

Pull requests 27

Projects 0

 Wiki

| Insights

Official repository for Citation Style Language (CSL) citation styles. <https://citationstyles.org/>

citation-style-language

bibliography

citations

[citation-styles](#)

9,255 commits

8 branches

0 releases

 674 contributors

Branch: master ▼

New pull request

Create new file

Upload files

Find File

Clone or download ▾

 adam3smith Merge pull request #3984 from anfourny/patch-1 ...

Latest commit 5e6ff65 7 hours ago

⚠ Sorry, we had to truncate this directory to 1,000 files. 941 entries were omitted from the list.

📁 .github

Update stale.yml

2 months ago

 document

Morgan will request #2070 from library/inmate. C4

42 days on.

# Tables and Figures

## Table suggestions

Several functions format R data into tables

Table with kable

eruptions	waiting
3.600	79
1.800	54
3.333	74
2.283	62

eruptionswaiting

1	3.60	79.00
2	1.80	54.00
3	3.33	74.00
4	2.28	62.00

Table with xtable

Table with stargazer

eruptionswaiting		
1	3.600	79
2	1.800	54
3	3.333	74
4	2.283	62

```
data <- faithful[1:4, ]
```

```
`` `{r results = 'asis'}  
knitr::kable(data, caption = "Table with kable")  
`` `
```

```
`` `{r results = "asis"}  
print(xtable::xtable(data, caption = "Table with xtable"),  
      type = "html", html.table.attributes = "border=0")  
`` `
```

```
`` `{r results = "asis"}  
stargazer::stargazer(data, type = "html",  
                      title = "Table with stargazer")  
`` `
```

Learn more in  
the **stargazer**,  
**xtable**, and  
**knitr** packages.

# Use Fancy LaTeX Code to Include Figure

```
1 <!-- %%%%%%%%%%% cluster sampling example %%%%%%%%%%% -->
2
3 \begin{figure}[!h]
4   \centering
5
6   \noindent
7   \includegraphics[width=\textwidth,height=\textheight,keepaspectratio]{/Users/KSauby/Documents/Projects/ACSampling_project/code/Manuscript/figures/RACS_figure.png}
8
9   \caption[Diagram of cluster sampling.]{Diagram of cluster sampling,
10    including A) secondary sampling around primary unit  $i$  for up to steps  $f=4$ ,
11    and B) the configurations of all clusters for which  $m$  is known where  $f_{\text{max}} = 2$ .}
12
13   \label{fig:ACS_diagram_cap}
14 \end{figure}
15 \clearpage
```

tell R Markdown where the image is

caption

- R Markdown will give the figure a number  
- reference this label in the text and the corresponding number will be included



# Numbering

- R Markdown numbers the tables and figures for you so you do not have to!

```
02_Methods.Rmd — ACSampling_project (git: master)
x postslides.tex Manuscript.Rmd header.tex Figures.Rmd 03_Results.Rmd 01_Introduction.Rmd 02_Methods.Rmd
49 The remaining units not satisfying  $C$  but within the cluster are edge units,
. the total number of which is  $a$ .
50 Thus, the total number of units in a cluster is  $m + a$ .
51 Each primary unit  $i$  that does not satisfy  $C$  is its own network and is
. assigned  $m=1$  and  $a=0$ 
52 (e.g., Figure \ref{fig:ACS_diagram_cap}a). |
53
54
```

151 Each primary unit  $i$  satisfying  $C$  is thus part of a neighborhood of units called a *cluster*, which includes the  
152 primary unit  $i$  and its associated secondary units. Within a cluster, the collection of units satisfying  $C$  is  
153 called a *network*,  $\psi$ , and the cardinality (the size, in terms of the number of units) of the network is denoted  
154  $m$ . The remaining units not satisfying  $C$  but within the cluster are *edge* units, the total number of which is  $a$ .  
155 Thus, the total number of units in a cluster is  $m + a$ . Each primary unit  $i$  that does *not* satisfy  $C$  is its own  
156 network and is assigned  $m = 1$  and  $a = 0$  (e.g., Figure 1a).

```

69 The HT estimator has variance  $\bar{y}_{HT}$  that is estimated by
70  $\widehat{\text{var}}(\bar{y}_{HT}) = N^{-2} \sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} y_{i\cdot} y_{j\cdot} (\pi_{ij} - \pi_i \pi_j) / (\pi_i \pi_j)$ 
    .
    .
    .
71 The number of distinct networks included in the primary sample is denoted by
    .
    .
    .
72 The total of the  $y$ -values in network  $\psi_i$  is  $y_{i\cdot} = \sum_{i \in \psi_i} y_i$ .
    .
    .
73

```

173 The HT estimator has variance  $\bar{y}_{HT}$  that is estimated by  $\widehat{\text{var}}(\bar{y}_{HT}) = N^{-2} \sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} y_{i\cdot} y_{j\cdot} (\pi_{ij} -$   
 174  $\pi_i \pi_j) / (\pi_i \pi_j)$  (Thompson 1990). The number of distinct networks included in the primary sample is  
 175 denoted by  $\kappa$ . The total of the  $y$ -values in network  $\psi_i$  is  $y_{i\cdot} = \sum_{i \in \psi_i} y_i$ .

# A Tool to Build Math Formulas

I just Google “online latex equation editor” to find website

https://www.codecogs.com/latex/eqnedit...

Clear Colors... Functions... Examples History


$\sum_{i}^n x \tfrac{1}{2}$


$\begin{bmatrix} \dots \end{bmatrix}$   $\begin{bmatrix} \dots \end{bmatrix}$   $\begin{bmatrix} \dots \end{bmatrix}$


gif Latin Modern (10pt) Normal 110 Transparent Inline Compressed


# Bookdown Dissertation!


<https://github.com/ksauby/thesisdownufl>


 **ksauby / thesisdownufl**  
forked from [ismayc/thesisdown](#)


 Unwatch ▾ 1


 Star 2


 Fork 122


 Code

 Pull requests 0

 Projects 0

 Wiki

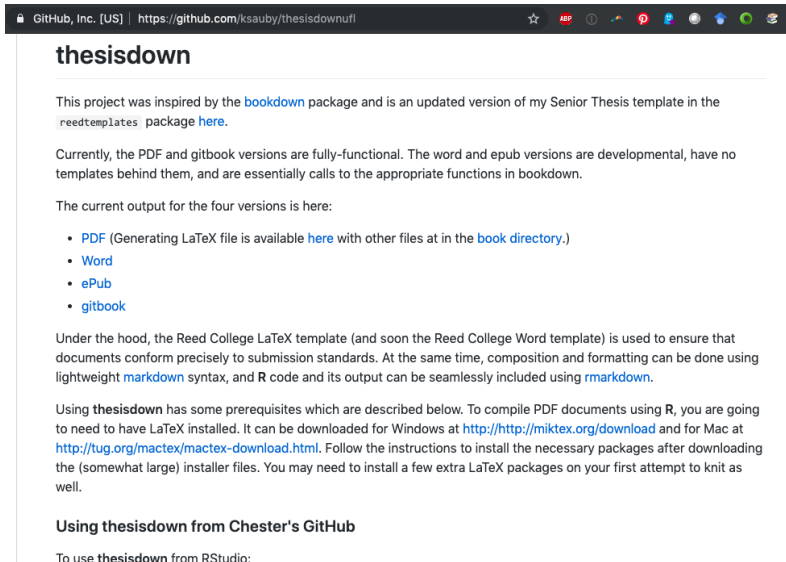
 Insights

 Settings

An updated R Markdown thesis template using the bookdown package; edited to work with the University of Florida dissertation template.

Edit

# Bookdown Dissertation!



The screenshot shows the GitHub repository page for 'thesisdown' by ksauby. The browser address bar shows the URL 'https://github.com/ksauby/thesisdown'. The repository name 'thesisdown' is prominently displayed. The description states that the project is inspired by the 'bookdown' package and is an updated version of a Senior Thesis template. It mentions that the PDF and gitbook versions are fully-functional, while the word and epub versions are developmental. A list of available outputs (PDF, Word, ePub, gitbook) is provided. The page also includes instructions on how to use the template, mentioning the Reed College LaTeX template and the need for LaTeX installation. The footer of the page indicates that the user is viewing the repository from RStudio.

thesisdown

This project was inspired by the [bookdown](#) package and is an updated version of my Senior Thesis template in the [reedtemplates](#) package [here](#).

Currently, the PDF and gitbook versions are fully-functional. The word and epub versions are developmental, have no templates behind them, and are essentially calls to the appropriate functions in bookdown.

The current output for the four versions is here:

- [PDF](#) (Generating LaTeX file is available [here](#) with other files at in the [book directory](#).)
- [Word](#)
- [ePub](#)
- [gitbook](#)

Under the hood, the Reed College LaTeX template (and soon the Reed College Word template) is used to ensure that documents conform precisely to submission standards. At the same time, composition and formatting can be done using lightweight [markdown](#) syntax, and R code and its output can be seamlessly included using [rmarkdown](#).

Using **thesisdown** has some prerequisites which are described below. To compile PDF documents using R, you are going to need to have LaTeX installed. It can be downloaded for Windows at <http://http://miktex.org/download> and for Mac at <http://tug.org/mactex/mactex-download.html>. Follow the instructions to install the necessary packages after downloading the (somewhat large) installer files. You may need to install a few extra LaTeX packages on your first attempt to knit as well.

### Using thesisdown from Chester's GitHub

To use thesisdown from RStudio:

# R Markdown Vs. Microsoft Word

## Pros of R Markdown

- incorporate code directly into your document
- you can hide text that you are ready to delete!

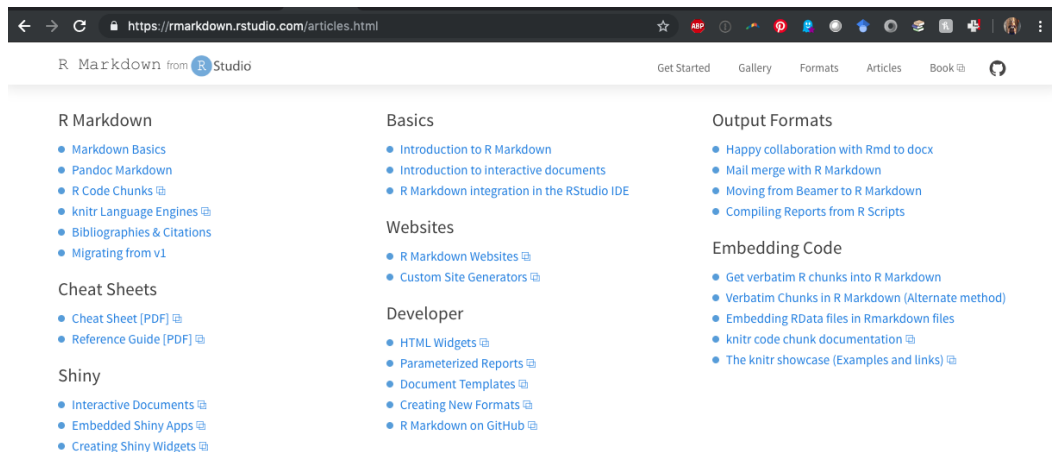
## Cons of R Markdown

- track changes - not as easy to implement changes when you have to do it in R Markdown

# Resources

<https://rmarkdown.rstudio.com/index.html>

<https://bookdown.org/yihui/rmarkdown/>

A screenshot of the R Markdown website. The browser address bar shows the URL 'https://rmarkdown.rstudio.com/articles.html'. The website header includes the title 'R Markdown from R Studio' and navigation links: 'Get Started', 'Gallery', 'Formats', 'Articles', 'Book', and a GitHub icon. The main content area is divided into three columns. The left column contains links for 'R Markdown' (including Basics, Pandoc, Code Chunks, Engines, Bibliographies, and Migration), 'Cheat Sheets' (PDF links), and 'Shiny' (Interactive Documents, Apps, and Widgets). The middle column contains links for 'Basics' (Introduction to R Markdown, Interactive documents, and RStudio IDE integration), 'Websites' (R Markdown Websites and Custom Site Generators), and 'Developer' (HTML Widgets, Parameterized Reports, Document Templates, Creating New Formats, and R Markdown on GitHub). The right column contains links for 'Output Formats' (Happy collaboration with Rmd to docx, Mail merge, Moving from Beamer, and Compiling Reports) and 'Embedding Code' (Get verbatim R chunks, Verbatim Chunks (Alternate method), Embedding RData files, knitr code chunk documentation, and The knitr showcase).