

Reproducible Reporting

Robert C. Cline, Sr., M.A., CPL

2022-06-12

Contents

1	Introduction	5
1.1	Usage	5
1.2	About the Author	6
1.3	Preview book	7
2	Implications for Reproducible Reporting	9
2.1	The Hypothetical Title Report	9
3	Cross-references	13
3.1	Chapters and sub-chapters	13
3.2	Captioned figures and tables	13
4	Parts	17
5	Footnotes and citations	19
5.1	Footnotes	19
5.2	Citations	19
6	Blocks	21
6.1	Equations	21
6.2	Theorems and proofs	21
6.3	Callout blocks	21
7	Sharing your book	23
7.1	Publishing	23
7.2	404 pages	23
7.3	Metadata for sharing	23

Chapter 1

Introduction

The scientific community has invested heavily in what is known as reproducible research to make their work reproducible for themselves and for other scientists wishing to replicate their studies to verify findings or explore the subject further. Reproducible research documents the research, making it both reproducible and verifiable. The technology available to create reproducible research is not limited to the scientific community. Any individual or organization can utilize this technology to make one's own reporting reproducible. The advantages of reproducible reporting are many, including organizing resources, and is used to automate the presentation of data analysis as data is compiled into a report for publication or review. Reproducibility turns the process into an iterative process, so as the data changes, the reports themselves simply update automatically. Stated another way, much of our reporting and document preparation stays the same while only the data source changes. An example would be the preparation of oil and gas leases. Instead of a lease and offer letter being prepared for each Lessor, all packages can be prepared at once from one data source. The names, tracts, net minerals and bonus payments change for each lessor and for each project, only the data input changes. The process of producing the lease package is one print operation. Only the data changes. Very little time is used in preparing the lease package documents. This is where reproducible reporting has a great advantage over traditional reporting methods both in savings of time and money. The time savings is huge.

1.1 Usage

While replication is a fundamental tenet of science, it is grossly under utilized in business; especially the title business. The process of reporting in the title industry is expensive and time consuming. The process of interpreting title ownership requires "hands-on" review of title documentation and a knowledge of land law and contract law. However, the process of reporting the analysis

can be shortened by utilizing the methodology of reproducible research. Reproducible reporting is a way of utilizing current technology to both create iterative types of reports and as a way to organize the title ownership data into easily managed data which can be maintained in a centralized database and updated as necessary.

1.2 About the Author

I, Robert C. Cline, Sr., hold a Masters Degree in Clinical Psychology from University of Houston-Clear Lake and I am an AAPL Certified Petroleum Landman. My family operated a sovereign title plant in Wharton Texas which was established by my grandfather in 1890. I started working in that plant at age 11, a barefooted boy copying documents in the court house, developing and printing the images for the abstractor to compile into Abstracts and Title Certificates. I worked in that plant throughout my highschool years. After college, I attended a two year program at the University of Texas San Antonio Title School, earning a certificate in the title insurance industry. I was a licensed closing officer in the title plant and was a certified Title Examiner for several title insurance underwriters including Chicago Title Insurance Agency and Minisota Title Insurance Company, a title insurance underwriter, a position that, by industry standards, required a law degree and five years of practice of land title law. I was a member of the Texas Land Title Association and was a Title Plant Examiner for that organization, examining the records of title plants for adequacy and accuracy of title plants which had applied for membership in the association.

In 1976, I joined the brokerage firm of I. Jon Brook, Jr. as a petroleum landman. In 1980, went to work for Clayton Williams, Jr. in the Austin Chalk as a landman. Robert has had many clients over the years working as an Independent landman, managing title research for clients, managing due diligence projects, leasing projects and seismic projects. Robert earned the master's degree while working full time in the oil and gas industry as a field landman. In 2010 to 2014, I managed the field operations for Fairways Exploration & Production, LLC acquiring 10,000 leases covering 400,000 acres. At that same time, I managed a 100 square mile seismic acquisition project for that same company.

Today, I am still employed as a landman. During the downturn in the industry in 2014 to 2020, I used the time to study advanced statistics, learning methods to create reproducible research. I write R code, a statistical programming language and some python. I am proficient in Excel VBA, Power Query, Power Pivot and I create MS Access applications utilizing knowledge of VBA. I use GIS in both R programming language and with QGIS, both of which are open source applications.

1.3 Preview book

This proposal is written in the format of a book with the Bookdown package and is made available online in the form of an HTML document, PDF format, and as eBook publication. I intend to maintain a live preview which will update as I edit and change the document on an as-needed basis.

```
bookdown::serve_book()
```

```
citation("plotrix")
```


Chapter 2

Implications for Reproducible Reporting

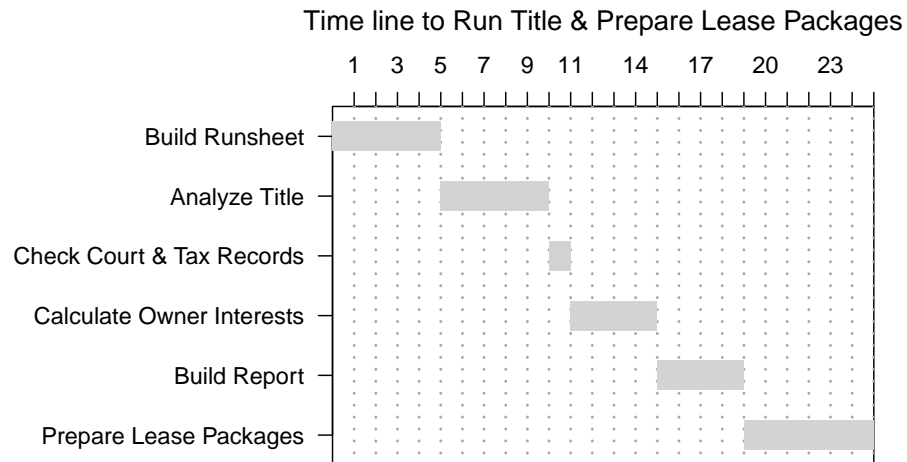
In the world of scientific research, the purpose of data analysis is to make the data analysis as open and transparent as possible to be able to exactly reproduce any of the results that were produced by it (?).

The data analysis pipeline is analogous to analyzing land title ownership. Title analysis begins with acquiring title information from public records, documenting the title chain, reading the title chain and interpreting the chain of ownership and reducing the ownership into numerical quantities. Mineral titles often have many owners with multiple classes of ownership, from royalty interests, production interests, mineral interests, term interests, life estate and remainder interests, severed or interests in both surface and minerals. These interests are commonly referred to the “bundle of sticks,” meaning there are many classes of ownerships in the mineral and surface estates. A title report analyses these interests in the form of fractional ownership. Periods of ownership is important as well. In Nebraska, for example, an owner who has not documented a claim or had a transfer of interest in his severed mineral estate, is subject to claims of abandonment of the mineral interest by the surface owner. The individual owners interest must be compiled into an ownership report together with the contact information for each owner. Claims, such as Lis Pendens, delinquent taxes, pending County and District Court suits must be documented when reporting the status of the title for each owner.

2.1 The Hypothetical Title Report

Ownership reports are commonly lengthy and contain a lot of data depending on the complexity of the title. Areas where there is mineral production, the title can expected to be more complex and more time consuming to complete.

Creating title reports, like scientific research, is a multistep process. The Gantt chart below outlines a hypothetical number of days that it could take for each step in the process of compiling ownership reports.



The Gantt chart lists six steps for the preparation of an ownership report and preparation of lease packages. I have had some mineral titles that had as few as one owner and as many as 350 owners. One never knows what title issues or how complex a title can be until one actually examines the title. The number of days chosen in the example below is simply a hypothetical case for illustration. What can remain constant is, the first four steps are relative in to steps 5 and six. That is, the number of days it takes to build the runsheet, analyze the title, calculate the owners' interests, is relative to the number of days it takes to build the report and prepare lease packages.

If for example one needs 15 days to get through calculating the owners' interests, one will need nearly that many days to build the report and prepare the lease packages. This is where reproducible reporting becomes an advantage. Using the techniques of reproducible reporting, the time required for building the reports and preparing the lease packages is reduced from days to usually *seconds*.

This advantage is multiple many fold if changes to the report must be made. Instead of re-writing a report, a simple change in the data for the report is made and the report is re-run in seconds.

2.1.1 Additional benefits

Mark Andrews (?) emphasizes that reproducible data analysis is often motivated as a means of doing more high quality and robust data analysis as a way of quality control essential to measure, identify errors, increase regour and verify results and conclusions. Mark Andrews offers as an example, a \$9bn loss of investment bank JPMorgan in 2012 (?). Keeping data in Excel also has created vulnerabilities. In 2020, the Public Health England lost Covid data as a result of using Excel to collect data. One expert commented “even a high-school computing student would know that better alternatives [to Excel] exist. ... you wouldn’t use XLS, nobody would do that” (?).

2.1.2 Summary of Savings with Reproducible Reporting

Each title examination is different from any other, no two titles are the same. No title examiner ever knows what awaits. Predicting the length of time to complete the title examination and compile the reporting is not feasible. But, as a rule of thumb the process of reporting is relative to the complexity of title. The more complex the title, the longer the process of preparing reports and lease packages. With reproducible reporting, the time process is reduced from the time it takes to examine the title to almost zero. Thus, by employing reproducible reporting, the savings in terms of time spent, is roughly one-half. Reproducible reporting saves about 50% of the time it normally takes to produce the same result.

2.1.3 Laying the groundwork for project management

Creating the infrastructure for reproducible reporting requires an understanding of the software package R and parameterized reporting. The title persons do not need to process the output. With training, they can of course. But people preparing the title reports will typically do so using templates prepared in CSV format. CSV files, or if preferred XLSX files can be batched processed to create the output into MS Word, PDF or HTML files.

Chapter 3

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

3.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
 - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world = # Hello world {#hello-world}`.
 - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 3.
 - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

3.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 3.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Don't miss Table 3.1.

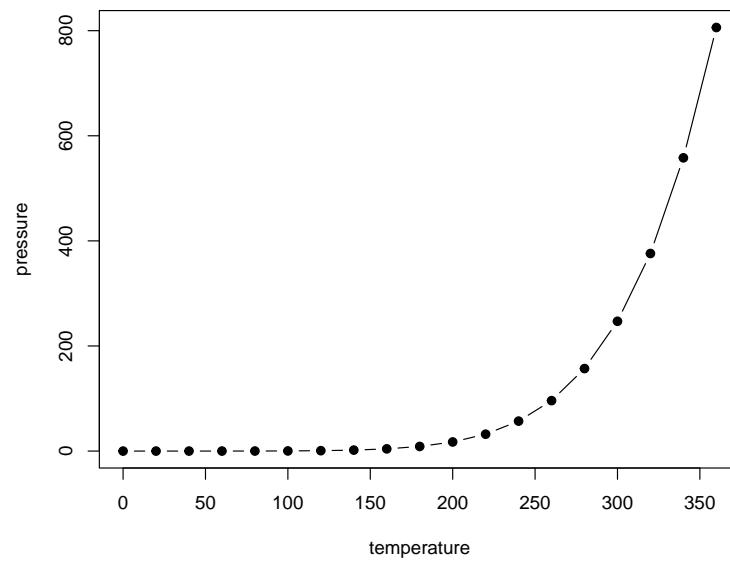


Figure 3.1: Here is a nice figure!

```
knitr::kable(  
  head(pressure, 10), caption = 'Here is a nice table!',  
  booktabs = TRUE  
)
```

Table 3.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Chapter 4

Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 5

Footnotes and citations

5.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

5.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (?) (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (?) (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `csl: chicago-fullnote-bibliography.csl` to the `index.Rmd` YAML, and include the `.csl` file. To download a new style, we recommend: <https://www.zotero.org/styles/>

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 6

Blocks

6.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (6.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (6.1).

6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 6.1.

Theorem 6.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

6.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use **markdown** inside a block.

```
head(beaver1, n = 5)
#>   day time  temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 7

Sharing your book

7.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

7.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

7.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```


Bibliography

Excel: Why using Microsoft's tool caused Covid-19 results to be lost.

Is Excel the most dangerous piece of software in the world?

Andrews, M. Chapter 7: Reproducible Data Analysis. page 37.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. <https://github.com/rstudio/bookdown>.