

RMarkdown automation.

Extending research into application

Rachael Remington; Anthony Davidson

2022-05-17

Contents

1	index	5
2	RStudio project file	7
2.1	git ignore file	7
3	BIOL8700__setup__2022	9
3.1	Tools	9
3.2	Data importing	9
3.3	Results	9
4	Experimental plan template	11
4.1	Research Question(s)?	12
4.2	Experimental Aims?	12
5	RMarkdown reports	13
6	Sampling design	17
6.1	Terry's lecture in RMarkdown	17
7	Automation	19
7.1	RMarkdown reports	19
7.2	Manual references	21
7.3	Packages	21
7.4	Download project and files	22
7.5	Usage	22
7.6	Render book	22
7.7	Preview book	22
7.8	Output files	23

Chapter 1

index

ss.

Last week we began to set up our projects using RMarkdown and R code so that we are ready for Terry Newmans sample/survey design session.

Now this is postponed till next week we can finish this off.

Chapter 2

RStudio project file

- <https://r4ds.had.co.nz/workflow-projects.html>

2.1 git ignore file

```
# History files
.Rhistory
.Rapp.history

# Session Data files
.RData

# User-specific files
.Ruserdata

# Example code in package build process
*-Ex.R

# Output files from R CMD build
/*.tar.gz

# Output files from R CMD check
/*.Rcheck/

# RStudio files
.Rproj.user/

# produced vignettes
vignettes/*.html
vignettes/*.pdf
```

```
# OAuth2 token, see https://github.com/hadley/httr/releases/tag/v0.3
.httr-oauth

# knitr and R markdown default cache directories
*_cache/
/cache/

# Temporary files created by R markdown
*.utf8.md
*.knit.md

# R Environment Variables
.Renviron

#notes
/notes2022/
/_main/
/libs/
/experimental_plan/
/Presentation/
/data/
```


Chapter 3

BIOL8700__setup__2022

The goal of BIOL8700__setup__2022 is to setup and generate documents using RStudio, RMarkdown and github

3.1 Tools

This is a *sample* book written in **Markdown** and includes embedded R code. This combination of programming languages is included in the RMarkdown package. We use this and a variation of other RMarkdown packages to make the most of reproducible reporting in R.

3.2 Data importing

There are scripts and code included within this repository to import data from csv and excel files from local and online locations.

3.3 Results

You can also embed plots, code and data in an RMarkdown document. The core results of this repository are as follows:

In that case, don't forget to commit and push the resulting figure files, so they display on GitHub.

3.3.1 Quick overview of last week

1. Run the R chunk below and check that you have all the packages running from last week.

```

library(rmarkdown)
library(rtticles)
library(bookdown)
library(bookdownplus)
library(tidyverse)
#> -- Attaching packages ----- tidyverse 1.3.1 --
#> v ggplot2 3.3.5      v purrr  0.3.4
#> v tibble  3.1.6      v dplyr  1.0.8
#> v tidyr   1.2.0      v stringr 1.4.0
#> v readr   2.1.2      v forcats 0.5.1
#> -- Conflicts ----- tidyverse_conflicts() --
#> x dplyr::filter() masks stats::filter()
#> x dplyr::lag()    masks stats::lag()
library(readxl)

```

2. Can you import the data set you found last week in class?

```

#read.csv
#read.xl
#example from

```

- What options did we have?
- Here I will take the following dataset and extend this to include some code and plots for the data

3. Follow along with the next chapter....

4.

```

library(tidyverse)
library(jtools)

```

Chapter 4

Experimental plan template

Now that you have been developing a research question over the last few months, the next step is to design a set of experiments that will specifically test your research aims.

We are often focused on the “cookbook” aspect of experiments – the protocols and steps required to conduct each experiment. However, it’s critical to spend time designing your overall experimental approach and the finer details to ensure that your research will produce robust data that can be clearly analysed without bias. When we test specific questions, we want to avoid statistical issues such as “noise” and “confounding” factors.

Terry Neeman will be delivering a workshop to help you strengthen your experimental plan – both in terms of your proposed design, and to help you more clearly and accurately explain the rationale and set-up of your experiments. This workshop will help you apply the principles taught in BIOL8291 to your own experiment. To prepare, you will create an outline of your experimental plan, focusing on the statistical framework of your design.

Below is an outline of questions for you to answer/justify for each part of your experimental plan. You will also need to draw two figures (digital drawings preferred) for each aim that show:

- 1) a simple overview of the experimental plan related to the research aim,
- 2) a detailed “snap-shot” of the experimental set up (i.e., how will the plates, plants, etc. be arranged? Will there be a row-column design? Blocking? Randomization? What treatments will be applied and how many replicates will be tested?)

You can access the template for the experimental plan on Wattle or a copy can be found within this repository. **DOWNLOAD NOW**. The template is laid out below as follows:

4.1 Research Question(s)?

#input question here

4.2 Experimental Aims?

For each experiment explain the overall experimental approach (1-3 sentences + overview diagram) list:

- 1) the response variables/outcome measure(s) for the experiment
- 2) the experimental factor(s) of interest
- 3) the experimental conditions (groups for comparisons), and the number of replicates/sample size for each condition
- 4) the experimental control(s)
- 5) are there any potentially confounding factors (“nuisance factors”)? Briefly explain how they will be tracked/or mitigated

Briefly explain the design of the experiment and provide a diagram that shows a “snap shot” of the experimental set-up (e.g., how all the plants under different experimental conditions will be arranged, all the plates in the lab, a flow chart of computational steps, etc.). Make sure to consider and include relevant design aspects like blocking, randomisation, as well as to clearly indicate treatments, replicate numbers, and controls.

TIP: Start thinking about how you will analyse your data:
“what statistical tests would you use?”

4.2.0.1 Importing data

4.2.0.2 Visualising data

4.2.0.3 Reporting

Chapter 5

RMarkdown reports

We should all know what these are and how to render a report in RMarkdown. Next we will produce a RMarkdown document for the question we have been working on ready to add data and other sampling design information.

```
#general packages used
library(tidyverse)
```

What does this tell us about how RProjects and other funky things work?

3. Data import

```
dat <- read.csv("data/Analysis_ardMods.csv")
glimpse(dat)
#> Rows: 380
#> Columns: 35
#> $ Well
#> $ Well.Position
#> $ Omit
#> $ Sample.Name
#> $ Delta.Ct.SE
#> $ Target.Name
#> $ Task
#> $ Reporter
#> $ Quencher
#> $ RQ
#> $ RQ.Min
#> $ RQ.Max
#> $ CT
#> $ EQ.Ct.Mean
#> $ EQ.Ct.SE
#> $ Quantity
```

```
<chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "1~
<chr> "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8",~
<chr> "false", "false", "false", "false", "false", "f~
<chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "1~
<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
<chr> "Gapdh", "Gapdh", "Gapdh", "Gapdh", "Gapdh", "G~
<chr> "UNKNOWN", "UNKNOWN", "UNKNOWN", "UNKNOWN", "UN~
<chr> "SYBR", "SYBR", "SYBR", "SYBR", "SYBR", "SYBR",~
<chr> "None", "None", "None", "None", "None", "None",~
<dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
<dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
<dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
<chr> "12.909", "14.253", "13.532", "13.377", "14.249~
<dbl> 12.909, 14.253, 13.532, 13.377, 14.249, 15.053,~
<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```

#> $ Delta.Ct.Mean      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Delta.Delta.Ct     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Automatic.Ct.Threshold <chr> "true", "true", "true", "true", "true", "true", ~
#> $ Ct.Threshold       <dbl> 0.062, 0.062, 0.062, 0.062, 0.062, 0.062, 0.062, ~
#> $ Automatic.Baseline <chr> "true", "true", "true", "true", "true", "true", ~
#> $ Baseline.Start     <int> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, ~
#> $ Baseline.End       <int> 9, 11, 10, 10, 11, 13, 10, 10, 11, 10, 9, 9, 10, ~
#> $ Efficiency         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Tm1                <dbl> 84.395, 84.395, 84.395, 84.395, 84.395, 84.527, ~
#> $ Comments          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Tm2               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Amp.Score         <dbl> 1.346, 1.345, 1.352, 1.349, 1.339, 1.356, 1.347, ~
#> $ Tm3               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> $ Cq.Conf           <dbl> 0.920, 0.872, 0.930, 0.933, 0.974, 0.938, 0.924, ~
#> $ MTP               <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ~
#> $ EXPFAIL           <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ~
#> $ NOISE             <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ~
#> $ NOAMP             <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ~
#> $ THOLDFAIL         <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ~

```

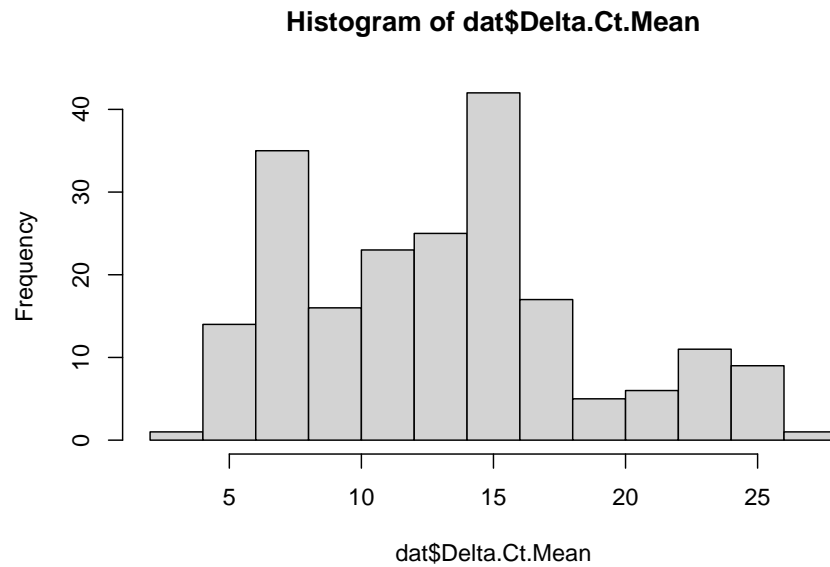
4. Data visualisation

```

# table(dat$Well.Position)
mean(dat$Delta.Ct.Mean, na.rm = TRUE)
#> [1] 13.08628
sd(dat$Delta.Ct.Mean, na.rm = TRUE)
#> [1] 5.507596

hist(dat$Delta.Ct.Mean)

```



5. Tidyverse approach

6. ggplot

Tasks ready for next week:

1. Outcome and predictor variables
2. Other studies with same sampling design
3. Other reference material.
4. Read a cool sampling design/issue paper

Watch this short (ish video) as a summary of what you should understand so far.

Chapter 6

Sampling design

```
#general packages used  
library(tidyverse)  
library(knitr)
```

We should all know what these are and how to render a report in RMarkdown. Next we will produce a RMarkdown document for the question we have been working on ready to add data and other sampling design information.

1. Outcome and predictor variables
2. Other study examples and code
3. Other studies with same sampling design
4. Other reference material.
5. Read a cool sampling design/issue paper

6.1 Terry's lecture in RMarkdown

```
knitr::include_graphics("study_design/Experimental-plan-workshop Terry Neeman 17 May 2002.pdf")
```

Experimental Plan in Biology Research

A Statistical Perspective

Terry Neeman

Australian National University

May 2022

```
#general packages used  
library(tidyverse)
```

Chapter 7

Automation

What is special about using `README.Rmd` instead of just `README.md`?

You can include R chunks like so:

You'll still need to render `README.Rmd` regularly, to keep `README.md` up-to-date. `devtools::build_readme()` is handy for this. You could also use GitHub Actions to re-render `README.Rmd` every time you push. An example workflow can be found here: <https://github.com/r-lib/actions/tree/v1/examples>.

7.1 RMarkdown reports

We should all know what these are and how to render a report in RMarkdown. Next we will produce a RMarkdown document for the question we have been working on ready to add data and other sampling design information.

1. Outcome and predictor variables
2. Other study examples and code
3. Other studies with same sampling design
4. Other reference material.
5. Read a cool sampling design/issue paper

Watch this short (ish video) as a summary of what you should understand so far.

What does this tell us about how RProjects and other funky things work?

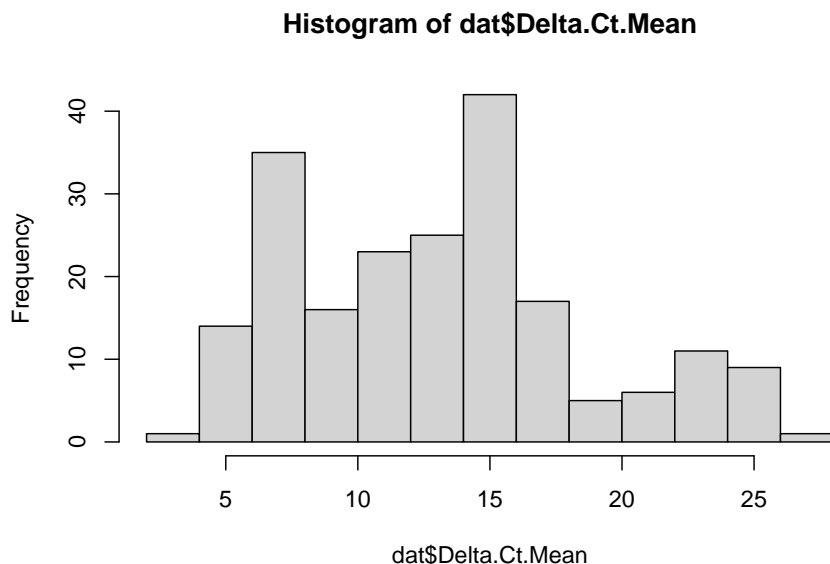
3. Data import

```
dat <- read.csv("data/Analysis_ardMods.csv")
```

4. Data visualisation

```
# table(dat$Well.Position)
mean(dat$Delta.Ct.Mean, na.rm = TRUE)
#> [1] 13.08628
sd(dat$Delta.Ct.Mean, na.rm = TRUE)
#> [1] 5.507596

hist(dat$Delta.Ct.Mean)
```



5. Tidyverse approach

```
glimpse(dat)
#> Rows: 380
#> Columns: 35
#> $ Well
#> $ Well.Position
#> $ Omit
#> $ Sample.Name
#> $ Delta.Ct.SE
#> $ Target.Name
#> $ Task
#> $ Reporter
#> $ Quencher
#> $ RQ
#> $ RQ.Min
#> $ RQ.Max
```

<chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "1~
 <chr> "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", ~
 <chr> "false", "false", "false", "false", "false", "f~
 <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "1~
 <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
 <chr> "Gapdh", "Gapdh", "Gapdh", "Gapdh", "Gapdh", "G~
 <chr> "UNKNOWN", "UNKNOWN", "UNKNOWN", "UNKNOWN", "UN~
 <chr> "SYBR", "SYBR", "SYBR", "SYBR", "SYBR", "SYBR", ~
 <chr> "None", "None", "None", "None", "None", "None", ~
 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~

```

#> $ CT <chr> "12.909", "14.253", "13.532", "13.377", "14.249~
#> $ EQ.Ct.Mean <dbl> 12.909, 14.253, 13.532, 13.377, 14.249, 15.053,~
#> $ EQ.Ct.SE <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Quantity <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Delta.Ct.Mean <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Delta.Delta.Ct <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Automatic.Ct.Threshold <chr> "true", "true", "true", "true", "true", "true",~
#> $ Ct.Threshold <dbl> 0.062, 0.062, 0.062, 0.062, 0.062, 0.062, 0.062~
#> $ Automatic.Baseline <chr> "true", "true", "true", "true", "true", "true",~
#> $ Baseline.Start <int> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
#> $ Baseline.End <int> 9, 11, 10, 10, 11, 13, 10, 10, 11, 10, 9, 9, 10~
#> $ Efficiency <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Tm1 <dbl> 84.395, 84.395, 84.395, 84.395, 84.395, 84.527,~
#> $ Comments <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Tm2 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Amp.Score <dbl> 1.346, 1.345, 1.352, 1.349, 1.339, 1.356, 1.347~
#> $ Tm3 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
#> $ Cq.Conf <dbl> 0.920, 0.872, 0.930, 0.933, 0.974, 0.938, 0.924~
#> $ MTP <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N~
#> $ EXPFAIL <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N~
#> $ NOISE <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N~
#> $ NOAMP <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N~
#> $ THOLDFAIL <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N~

```

6. Nicer plots using ggplot

7.1.0.1 Next steps

One aspect that can be challenging when working with RMarkdown documents for manuscripts is references.

The references for a bookdown or rmarkdown file can be included using the following information in the yml header of the index file. The references for a bookdown or rmarkdown file can be included using the following information in the yml header of the index file.

7.2 Manual references

7.3 Packages

The goal of BIOL877_setup_2022 repository here is to provide a starting point for using RMarkdown and RStudio to undertake research.

7.4 Download project and files

We checked and loaded an RMarkdown

This is a *sample* book written in **Markdown** and extended from the original RMD file. You can use anything that Pandoc’s Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

7.5 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: **# A good chapter**, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: **## A short section** or **### An even shorter section**.

The **index.Rmd** file is required, and is also your first book chapter. It will be the homepage when you render the book.

7.6 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and
2. Click on **Build Book**, then select your output format, or select “All formats” if you’d like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a **bookdown::pdf_book**, you’ll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.

7.7 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in “Preview book”, or from the R console:

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

7.8 Output files

One of the benefits of working in RMarkdown is that the ‘gitbook’ extension is that it can easily be hosted on github. To do this with the least hurdles is to publish the output files as a github pages site from the `.docs/`.