Introduction to R markdown

Behaviour and evolution meeting

Tom Keaney

17 September 2019

Contents


```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, tidy = TRUE)
```

1 Intro

What is R markdown?

- A way to turn your statistical analysis into a journal quality report.
- An alternative to r-scripts.
- A tool to be used in the movement towards transparent and reproducible science.

Why do I use R markdown?

- Annotating code is much easier.
- To create documents to share analyses with supervisors / collaborator.
- To create supplementary material documents.
- To report the code used to create figures and run analysis.
- To share raw data.

What else could I use it for?

- You can create presentations, websites and even write your thesis in Rmarkdown!
- Combine different coding languages within the same document.

2 Knitting

Once we have our analysis up and running, we can produce a html, word doc or pdf from this .Rmd file (Rmarkdown file), using the Knit button next to Find/replace.

A good first step is to specify how you want to refine the document by editing the **YAML header**.

Aesthetics to consider:

- Code folding: If echo is set to true (default) then your code will display alongside the output. Code folding hides the code unless it is requested. Very handy for long code chunks. Specify 'hide' to start with hidden code and 'show' to start with visible code.
- toc: table of contents. You can specify the table of contents to 'float' as you scroll through the document. When writing the document use # to denote heading to be used in the toc. 'Depth' can also be used to specify how many levels of headings you would like to be included in the toc e.g. #, ##, ### etc.
- Theme: this controls how your document will look. Check this blogpost out for a rundown on available themes. I've used the sandstone theme for this document.
- Code download: this allows people to download your markdown file from the html document. Great for data and analysis transparency.
- always_allow_html: yes. This makes it possible to render PDFs or word docs when there is html-exclusive material in the document. I wouldn't recommend.

3 Using markdown:

R markdown can be broken down into two categories: writing and code chunks.

3.1 Writing

The first thing to note is you don't need hashtags to write outside of 'code chunks'.

Writing outside of code chunks works similarly to word processing programs.

3.1.1 Syntax tips to create effects:

- **Bold** is coded by enclosing the word with **
- *Italics* are coded by enclosing the word with _
- Superscript^2 is coded by enclosing with ^
- Subscript$_2$ is coded by enclosing with ~
- Equations: $\Delta\overline{W} = V_{A,W}$, enclose with $
- Dotpoints are coded by starting a sentence with -, * or +
- Hyperlink: https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf enclose with <>
- Link: use [link] then follow with (website).
- Images: !, caption in square brackets, figure file name in standard brackets.

Figure 1: Figure 1. An R meme

3.1.2 Referencing

The easiest way to reference in markdown is to export your references from your citation manager as bibtex files. Save your .bib file in your working directory then cite references as follows:

In the YAML header include bibliography: "name of bibliography file". The bibliography will be automatically generated at the end of the document.

Include the reference like this [@. . . ref ID here]. If you wish to cite without the authors name, use [-@. . . ref ID here].

E.g. this document was created in R-studio (RStudio Team 2018). If you want to reference multiple sources simply lsit them within the square brackets e.g. (Wernick et al. 2019; Svensson and Connallon 2019).

3.2 Coding and sample analysis

This is pretty much the same as normal R-studio scripts, but there are a few important differences.

Code is written within chunks, which can be included using the green insert button. Alternatively, chunks can be manually added by enclosing lines with "‘. The first set of dashes should be followed by {}. Note that you can combine different coding programs within the same document.

Within the curly brackets, you can name the chunk and provide some additional information using the following:

- Echo = T/F: do you want the script to display with the output?
- Eval = T/F: do you want R to run this chunk?
- Include = T/F: do you want the chunk to run, but not be included in the html document?

- Cache = T/F: If true the chunk will be saved when first knitted and reloaded when knitting in the future - useful for computationally expensive outputs.
- Warning = T/F: would you like warning messages in the output?
- Message = T/F: would you like messages in the output?
- Fig.width and Fig.height: how large would you like the figure?

Now lets load in some handy packages.

Note that we are also installing `tinytex`, which is required to knit documents to PDF files. It's a simplified version of LaTex designed specifically for R users that aren't familiar with the LaTex syntax. If you already have LaTex, I recommend doing further research before installing, as compatibility issues appear possible.

```
# Write code as normal in here - hashtags required for annotation

library(tidyverse)  # a range of packages for data manipulation and plotting
library(pander)  # for nice tables
library(kableExtra)  # for scrollable tables
library(ggbeeswarm)  # an extension to ggplot2 that allows geom_quasirandom

# Install tinytex - you only need to do this once

# install.packages('tinytex') tinytex::install_tinytex() # install TinyTeX
```

Next lets load in some data of mine:

```
data <- read.csv("example_data.csv") %>% select(Individual, Sex, Wing.size..mm.) %>%
    filter(!is.na(Wing.size..mm.)) %>% rename(Wing_length = Wing.size..mm.)
```

The dataset is made up of three columns - an individual ID, the sex of the individual and the wing length of the individual. Wing length was measured for 405 fruit flies.

3.2.1 Tables

Lets create a table with the `pander` package, which makes really nice tables in html documents.

We also can create tabs using {.tabset} after the header. Here I use this to illustrate how to create bold rows, which can be useful when you want to showoff a significant result.

3.2.1.1 Table

```
# I want to calculate confidence intervals from the raw data so I make this
# easier by creating a function

SE <- function(x) sd(x)/sqrt(length(x))

# Create a summary to display in the table. Basically I use the dplyr
# package (part of the tidyverse) to split the data by sex with the group_by
# function, then tell R to find the mean wing length for each sex and
# associated CIs. Finally, I create a new column called n, which gives us
# the number of individuals within each group.
```

```
summary_data <- data %>% group_by(Sex) %>% dplyr::summarise(Mean_wing_length = mean(Wing_length),
    Lower = (Mean_wing_length - SE(Wing_length) * 1.96), Upper = (Mean_wing_length +
        SE(Wing_length) * 1.96), n = n()) %>% mutate(Sex = replace(as.character(Sex),
    Sex == "F", "Female"), Sex = replace(as.character(Sex), Sex == "M", "Male")) %>%
    rename(Wing_length = Mean_wing_length)

# Now lets make the table

pander(summary_data, split.cell = 40, split.table = Inf, col.names = c("Sex",
    "Mean wing length (mm)", "Lower CI", "Upper CI", "n"), round = 3)
```

Sex	Mean wing length (mm)	Lower CI	Upper CI	n
Female	1.071	1.058	1.083	204
Male	0.983	0.972	0.994	201

`split.cell` tells pander where to include a line break in the column names.

`split.table` tells pander where to split the table into multiple tables, Inf tells R to never split the table.

`round` specifies the number of decimal places you want presented.

More info on pander tables here.

3.2.1.2 Bolded table

```
pander(summary_data, split.cell = 40, split.table = Inf, col.names = c("Sex",
    "Mean wing length (mm)", "Lower CI", "Upper CI", "n"), emphasize.strong.rows = (2),
    round = 3)
```

Sex	Mean wing length (mm)	Lower CI	Upper CI	n
Female	1.071	1.058	1.083	204
Male	**0.983**	**0.972**	**0.994**	**201**

`split.cell` tells pander where to include a line break.

`split.table` tells pander where to split the table into multiple tables, Inf tells R to never split the table.

`round` specifies the number of decimal places you want presented.

`emphasize.strong.rows` tells pander to display certain rows in bold

More info on pander tables here.

3.2.2 Plots

Now lets explore our data visually with the package `ggplot2`

```
data <- data %>% mutate(Sex = replace(as.character(Sex), Sex == "F", "Female"),
    Sex = replace(as.character(Sex), Sex == "M", "Male"))
```

```r
# There's a bit going on below.

# First I plot the wing length of each fly with geom_quasirandom

# Then I use the mean and CIs we calculated for the table and plot them with
# geom_point and geom_errorbar

# The rest of the code is plot aesthetics


data %>% ggplot(aes(x = Sex, y = Wing_length, fill = Sex, colour = Sex)) + geom_quasirandom(data = data
    width = 0.3, size = 2, alpha = 0.5) + geom_point(data = summary_data, aes(x = Sex,
    y = Wing_length), size = 3, colour = "Black") + geom_errorbar(data = summary_data,
    aes(x = Sex, ymax = Upper, ymin = Lower), colour = "black", width = 0, size = 1.2) +
    scale_colour_manual(values = c(Female = "#fe9929", Male = "#41b6c4")) +
    labs(x = "Sex", y = "Wing length (mm)") + theme_minimal() + theme(legend.position = "none") +
    theme(panel.grid.major.x = element_blank())
```

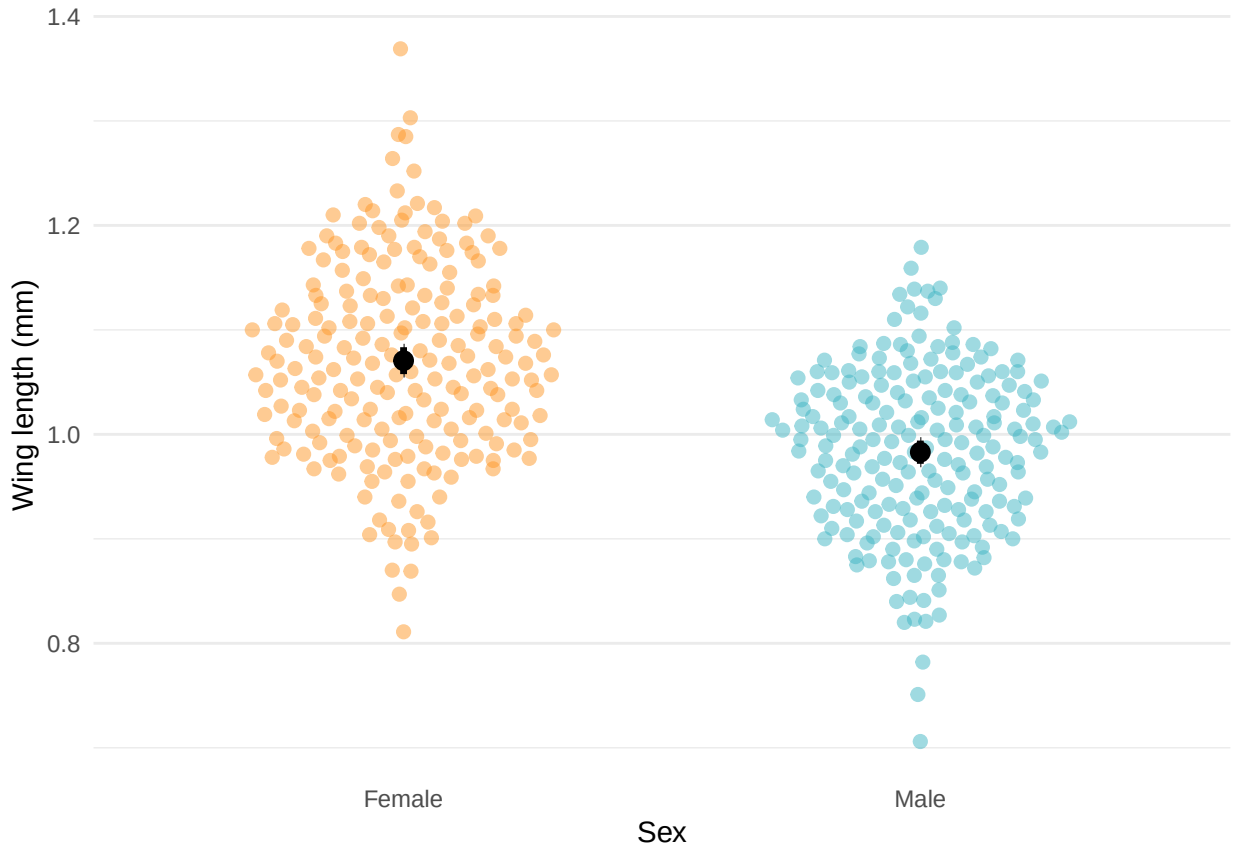

Figure 2. Female flies are larger than male flies.

For plot colours I use http://colorbrewer2.org.

3.2.3 Raw data and session info

One of the major benefits of R markdown is the ease with which you can make your data readily available. I use the `kableExtra` package to include the raw data we have used. This package gives us the ability

to include scrollable tables, which are very useful for large datasets. **However, the scrollbox is only available in html format.**

```
kable(data, format = "html") %>% kable_styling(bootstrap_options = c("hover",
    "condensed"), fixed_thead = T) %>% scroll_box(height = "400px")
```

Individual

Sex

Wing_length

601

Female

1.102

602

Male

1.009

603

Female

0.988

606

Male

0.940

609

Female

0.981

610

Male

0.820

611

Female

1.202

617

Female

0.969

620

Male

0.936

621

Female

0.811

624

Male

0.840

625

Female

1.264

626

Male

0.957

629

Female

1.179

630

Male

1.102

632

Male

1.179

636

Male

1.014

639

Female

1.194

642

Male

1.030

643

Female

1.083

644

Male

0.926

645

Female

1.287

648

Male

1.139

649

Female

1.163

650

Male

0.951

651

Female

1.020

653

Female

1.142

655

Female

1.176

656

Male

0.982

657

Female

1.108

658

Male

1.033

663

Female

1.039

664

Male

1.012

666

Male

1.050

668

Male

1.009

676

Male

0.964

678

Male

1.030

682

Male

1.050

683

Female

1.084

684

Male

0.975

685

Female

1.155

686

Male

0.931

688

Male

1.060

689

Female

1.038

690

Male

1.032

692

Male

1.035

693

Female

1.106

694

Male

1.054

696

Male

1.074

697

Female

1.167

699

Female

1.111

700

Male

1.072

705

Female

1.113

707

Female

1.210

708

Male

1.137

709

Female

1.057

710

Male

1.060

711

Female

1.074

712

Male

1.082

715

Female

1.220

716

Male

1.094

717

Female

1.105

719

Female

1.285

722

Male

1.071

723

Female

1.178

724

Male

1.087

725

Female

1.052

727

Female

1.233

730

Male

1.060

731

Female

1.179

732

Male

1.038

735

Female

1.106

738

Male

1.134

739

Female

1.214

742

Male

1.005

743

Female

1.057

744

Male

0.928

745

Female

1.068

747

Female

1.133

749

Female

1.140

752

Male

0.926

753

Female

1.369

757

Female

1.202

758

Male

1.016

761

Female

1.303

767

Female

1.174

769

Female

1.086

770

Male

0.969

773

Female

1.166

783

Female

1.190

785

Female

1.134

788

Male

1.116

793

Female

1.183

795

Female

0.940

803

Female

1.108

805

Female

1.024

807

Female

1.015

808

Male

0.955

810

Male

0.956

811

Female

1.119

815

Female

1.143

817

Female

1.172

818

Male

1.025

819

Female

1.183

820

Male

1.006

824

Male

1.068

827

Female

1.094

830

Male

1.023

831

Female

1.123

832

Male

1.055

833

Female

1.013

835

Female

1.137

838

Male

0.977

839

Female

1.022

840

Male

1.017

843

Female

1.133

844

Male

0.938

848

Male

0.875

855

Female

1.175

858

Male

1.051

860

Male

0.949

862

Male

1.033

885

Female

1.142

891

Female

1.052

892

Male

1.004

898

Male

1.086

904

Male

0.913

905

Female

1.170

907

Female

1.252

912

Male

1.056

915

Female

1.023

918

Male

1.110

919

Female

0.979

924

Male

1.088

927

Female

0.955

928

Male

1.051

932

Male

0.964

933

Female

1.126

934

Male

1.005

938

Male

1.122

939

Female

0.991

941

Female

1.204

944

Male

0.983

945

Female

1.133

947

Female

1.217

961

Female

1.019

962

Male

0.965

964

Male

1.086

965

Female

0.986

973

Female

1.133

975

Female

1.075

976

Male

0.992

978

Male

1.041

982

Male

0.883

987

Female

1.110

988

Male

0.999

989

Female

1.190

1010

Male

0.939

1024

Male

0.879

1031

Female

1.001

1040

Male

1.059

1074

Male

1.130

1076

Male

1.055

1079

Female

0.870

1093

Female

0.982

1094

Male

0.897

1095

Female

1.042

1098

Male

0.890

1110

Male

0.882

1117

Female

1.125

1121

Female

0.955

1125

Female

1.062

1132

Male

0.903

1135

Female

1.121

1137

Female

0.978

1138

Male

0.880

1139

Female

0.975

1140

Male

0.933

1141

Female

1.044

1142

Male

1.002

1145

Female

0.895

1150

Male

0.913

1154

Male

1.159

1157

Female

1.042

1163

Female

1.014

1171

Female

1.084

1174

Male

0.995

1175

Female

1.130

1177

Female

1.053

1181

Female

0.909

1184

Male

1.007

1185

Female

0.992

1187

Female

0.975

1195

Female

1.076

1196

Male

0.878

1198

Male

0.922

1202

Male

0.918

1203

Female

1.016

1209

Female

0.904

1212

Male

0.900

1216

Male

0.896

1218

Male

0.931

1219

Female

1.073

1221

Female

1.014

1224

Male

1.004

1226

Male

1.037

1227

Female

1.060

1229

Female

1.042

1232

Male

1.007

1233

Female

1.045

1244

Male

1.080

1246

Male

1.021

1252

Male

1.073

1258

Male

1.031

1274

Male

0.878

1287

Female

1.045

1292

Male

1.047

1297

Female

1.024

1325

Female

0.967

1337

Female

1.038

1356

Male

1.077

1361

Female

0.963

1364

Male

0.841

1369

Female

1.005

1382

Male

1.061

1386

Male

0.957

1404

Male

0.751

1426

Male

0.963

1429

Female

1.071

1435

Female

1.016

1438

Male

0.862

1440

Male

1.011

1441

Female

0.901

1443

Female

0.918

1444

Male

0.890

1446

Male

1.036

1454

Male

1.038

1458

Male

0.998

1464

Male

0.907

1466

Male

0.944

1476

Male

1.007

1477

Female

1.113

1481

Female

0.979

1490

Male

1.140

1491

Female

1.045

1493

Female

1.143

1495

Female

0.976

1496

Male

1.047

1501

Female

1.053

1510

Male

1.024

1516

Male

1.084

1520

Male

1.042

1521

Female

0.869

1523

Female

1.209

1535

Female

1.178

1537

Female

1.221

1542

Male

1.040

1550

Male

0.971

1556

Male

0.978

1560

Male

1.011

1574

Male

0.823

1579

Female

0.964

1580

Male

0.902

1609

Female

0.847

1612

Male

0.905

1615

Female

1.187

1617

Female

0.926

1620

Male

1.010

1621

Female

0.979

1622

Male

0.988

1623

Female

1.078

1626

Male

0.989

1630

Male

0.976

1631

Female

1.068

1642

Male

0.912

1648

Male

0.939

1651

Female

1.190

1653

Female

1.027

1655

Female

1.054

1656

Male

0.983

1657

Female

1.089

1670

Male

0.945

1672

Male

0.984

1673

Female

1.013

1677

Female

1.034

1682

Male

0.969

1684

Male

0.936

1687

Female

0.995

1690

Male

0.973

1691

Female

1.023

1697

Female

0.962

1698

Male

0.880

1709

Female

0.967

1714

Male

0.928

1718

Male

1.067

1728

Male

0.782

1730

Male

1.017

1741

Female

1.097

1751

Female

1.090

1770

Male

0.926

1772

Male

0.851

1774

Male

0.821

1779

Female

0.977

1781

Female

1.056

1785

Female

1.062

1787

Female

1.198

1804

Male

1.060

1810

Male

1.084

1811

Female

1.177

1820

Male

0.898

1823

Female

1.018

1824

Male

0.932

1828

Male

0.872

1829

Female

1.096

1832

Male

0.865

1835

Female

0.959

1855

Female

1.100

1868

Male

0.892

1875

Female

0.976

1884

Male

0.918

1902

Male

0.865

1906

Male

0.929

1918

Male

0.706

1928

Male

1.078

1930

Male

1.008

1933

Female

1.094

1938

Male

0.844

1939

Female

1.068

1940

Male

0.900

1941

Female

1.040

1955

Female

0.985

1961

Female

0.989

1962

Male

0.919

1964

Male

0.999

1966

Male

0.910

1970

Male

0.999

1973

Female

1.024

1983

Female

0.996

1987

Female

0.994

1990

Male

1.012

1993

Female

1.033

1994

Male

0.906

1995

Female

0.908

1998

Male

0.987

2003

Female

1.042

2004

Male

0.963

2006

Male

1.059

2009

Female

1.149

2012

Male

1.030

2014

Male

0.947

2015

Female

1.157

2019

Female

0.998

2021

Female

1.092

2022

Male

1.060

2023

Female

1.005

2024

Male

1.021

2025

Female

1.074

2027

Female

0.967

2028

Male

1.042

2031

Female

1.106

2033

Female

1.063

2034

Male

0.876

2035

Female

0.985

2036

Male

0.904

2037

Female

1.053

2039

Female

1.205

2041

Female

1.106

2051

Female

1.090

2052

Male

0.995

2054

Male

0.995

2056

Male

0.973

2060

Male

0.965

2063

Female

1.165

2067

Female

1.070

2069

Female

0.994

2070

Male

0.902

2071

Female

1.103

2072

Male

0.981

2073

Female

0.916

2074

Male

0.995

2075

Female

1.102

2077

Female

1.076

2078

Male

0.944

2079

Female

1.003

2081

Female

1.212

2089

Female

0.940

2093

Female

0.897

2095

Female

1.011

2096

Male

0.970

2103

Female

1.085

2104

Male

0.917

2109

Female

0.936

2135

Female

0.999

2136

Male

0.827

2155

Female

1.080

2156

Male

1.017

2158

Male

1.059

2160

Male

0.952

2170

Male

0.993

2171

Female

1.057

2178

Male

0.988

2187

Female

1.100

2189

Female

1.124

2196

Male

1.071

2197

Female

1.114

For more on `kableExtra` click here.

Finally we can also include our R session info to make it easier for others to reproduce our analysis

```
sessionInfo() %>% pander()
```

R version 3.5.3 (2019-03-11)

Platform: x86_64-w64-mingw32/x64 (64-bit)

locale: *LC_COLLATE=English_United Kingdom.1252*, *LC_CTYPE=English_United Kingdom.1252*, *LC_MONETARY=English_United Kingdom.1252*, *LC_NUMERIC=C* and *LC_TIME=English_United Kingdom.1252*

attached base packages: *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

other attached packages: *ggbeeswarm(v.0.6.0)*, *kableExtra(v.1.1.0)*, *pander(v.0.6.3)*, *forcats(v.0.4.0)*, *stringr(v.1.4.0)*, *dplyr(v.0.8.1)*, *purrr(v.0.3.2)*, *readr(v.1.3.1)*, *tidyr(v.0.8.3)*, *tibble(v.2.1.2)*, *ggplot2(v.3.1.1)* and *tidyverse(v.1.2.1)*

loaded via a namespace (and not attached): *beeswarm(v.0.2.3)*, *tidyselect(v.0.2.5)*, *xfun(v.0.7)*, *haven(v.2.1.0)*, *lattice(v.0.20-38)*, *colorspace(v.1.4-1)*, *generics(v.0.0.2)*, *htmltools(v.0.3.6)*, *viridisLite(v.0.3.0)*, *yaml(v.2.2.0)*, *rlang(v.0.3.4)*, *pillar(v.1.4.1)*, *glue(v.1.3.1)*, *withr(v.2.1.2)*, *modelr(v.0.1.4)*, *readxl(v.1.3.1)*, *plyr(v.1.8.4)*, *munsell(v.0.5.0)*, *gtable(v.0.3.0)*, *cellranger(v.1.1.0)*, *rvest(v.0.3.4)*, *evaluate(v.0.14)*, *labeling(v.0.3)*, *knitr(v.1.23)*, *vipor(v.0.4.5)*, *highr(v.0.8)*, *broom(v.0.5.2)*, *Rcpp(v.1.0.1)*, *scales(v.1.0.0)*, *backports(v.1.1.4)*, *formatR(v.1.6)*, *webshot(v.0.5.1)*, *jsonlite(v.1.6)*, *hms(v.0.4.2)*, *digest(v.0.6.19)*, *stringi(v.1.4.3)*, *grid(v.3.5.3)*, *cli(v.1.1.0)*, *tools(v.3.5.3)*, *magrittr(v.1.5)*, *lazyeval(v.0.2.2)*, *crayon(v.1.3.4)*, *pkgconfig(v.2.0.2)*, *xml2(v.1.2.0)*, *lubridate(v.1.7.4)*, *assertthat(v.0.2.1)*, *rmarkdown(v.1.13)*, *httr(v.1.4.0)*, *rstudioapi(v.0.10)*, *R6(v.2.4.0)*, *nlme(v.3.1-137)* and *compiler(v.3.5.3)*

4 Github

Github is an online data storage and web-hosting platform that is free to join. The learning curve is pretty steep but there are very helpful guides that will ease you through it. It is here where you can archive data and host your supplementary material online document.

I recommend this guide: https://cfss.uchicago.edu/setup/github/

Especially the setting up Git and Github, and using Git with R studio sections.

5 Further reading

I've only scratched the surface, for more info I recommend:

1. R Markdown: The definitive guide
2. The R markdown cheat sheet
3. Word documents knit with very basic formatting, to make them look nicer you'll need to create a 'style reference document.' More info here.
4. The package `workflowr` integrates R markdown with github - an online hosting platform where you can store data and put your html supplementary material online - to create a reproducible and intuitive workflow for your data analysis. Once you're familiar with markdown the leap to `workflowr` is small. For more details there is an extremely helpful walk-through.

References

RStudio Team. 2018. *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, Inc. http://www.rstudio.com/.

Svensson, Erik I., and Tim Connallon. 2019. "How Frequency-Dependent Selection Affects Population Fitness, Maladaptation and Evolutionary Rescue." Journal Article. *Evolutionary Applications* 12 (7): 1243–58. https://doi.org/10.1111/eva.12714.

Wernick, Riana I., Stephen F. Christy, Dana K. Howe, Jennifer A. Sullins, Joseph F. Ramirez, Maura Sare, McKenna J. Penley, Levi T. Morran, Dee R. Denver, and Suzanne Estes. 2019. "Sex and Mitonuclear Adaptation in Experimental Caenorhabditis Elegans Populations." Journal Article. *Genetics* 211 (3): 1045. https://doi.org/10.1534/genetics.119.301935.