Reproducibility in wildlife ecology

The draft manuscript extended from aaarchmiller/reproducibility_in_wildlife_ecology

Contents

The key aspects I am interested in here is the consistency of model diagrams in invasive species modelling in NZ and globally… for example…..

1 Summary

Processing data from Google form Description: Takes raw csv file from the Google form and updates it for analysis.

1.1 Coder name(s): Althea ArchMiller

1.1.1 Preamble

```
# Load libraries
library(ezknitr)
library(knitr)
library(devtools)
```

```
## Loading required package: usethis
```

Clear environment and set seed

```
remove(list=ls())
set.seed(2647)
#Load Data
raw_data <- read.csv(file = "./data/google_form_data_raw.csv", stringsAsFactors = F, header = T)
str(raw_data)
```

```
## 'data.frame':    101 obs. of  14 variables:
##  $ Timestamp                                                                    : chr  "2018
##  $ Study.ID..e.g...wsb002.or.jwm040.                                            : chr  "wsb
##  $ Reviewer.s.Initials..yours.                                                  : chr  "AAA
##  $ Was.this.paper.suitable.for.review..                                         : chr  "Tru
```

```
##  $ Was.this.paper.reviewed..                                                                  : chr  "" "
##  $ Data.readily.available.                                                                    : chr  "Fals
##  $ Was.the.data.pre.processed.or.shared.in.original.raw.format.                               : chr  "Pre
##  $ Code.available                                                                             : chr  "Tru
##  $ Analysis.was.entirely.code.based                                                           : chr  "Tru
##  $ Open.source.file.formats.used.                                                             : chr  "Tru
##  $ Can.code.be.run.as.is..                                                                    : int  5 4 
##  $ Were.the.study.s.figures.reproduced..                                                      : int  5 5 
##  $ Were.the.numbers.in.the.study.s.tables.and.or.results.text.quantitatively.reproduced..: int  5 4 
##  $ Were.the.study.s.conclusions.reproduced..                                                  : int  5 4 
```

```
## 'data.frame': 101 obs. of 14 variables: ## $ Timestamp : chr "2018/07/25 1:16:00 PM CDT" "2018/07/25
```

Fill in missing data The question "Was this study reviewed?" was added after a few survey responses were filled in, so I will update those records with "TRUE"

View responses to this column:

```
table(raw_data$Was.this.paper.reviewed..)
```

```
##
##
##                                                                    6
## False, author declined to participate (e.g., private data, no time, etc)
##                                                                   10
##                                              False, author never responded
##                                                                   41
##                                  False, for other reasons not stated above
##                                                                    6
##                                               False, unsuitable for review
##                                                                    6
##                                                                 True
##                                                                   32
```

```
## ## ## 6 ## False, author declined to participate (e.g., private data, no time, etc) ## 10 ## False, 
```

Double check that it worked

```
table(raw_data$Was.this.paper.reviewed..)
```

```
##
##
##                                                                    6
## False, author declined to participate (e.g., private data, no time, etc)
##                                                                   10
##                                              False, author never responded
##                                                                   41
##                                  False, for other reasons not stated above
##                                                                    6
##                                               False, unsuitable for review
##                                                                    6
##                                                                 True
##                                                                   32
```

```
## ## False, author declined to participate (e.g., private data, no time, etc)
## 10 ## False, author never responded ## 41 ## False, for other reasons not stated above ## 6 ## False

raw_data$Why.not.reviewed[raw_data$Was.this.paper.reviewed..=="True"] <- "NA: Study was reviewed"
```

```
# If unsuitable, add this information in:
raw_data$Why.not.reviewed[raw_data$Was.this.paper.reviewed..=="False, unsuitable for review"] <- "Unsui
# If author never responded, add this information in:
raw_data$Why.not.reviewed[raw_data$Was.this.paper.reviewed..=="False, author never responded"] <- "No c
# In two cases, a study was initially thought to be not reviewed but it was. One example of this was jw
new_data <- raw_data[!(raw_data$Study.ID..e.g...wsb002.or.jwm040.. == "jwm027" & raw_data$Was.this.pape
```

In the second case, we had to send in a written proposal to a regulatory agency to get the data. We thought that we would not get the proposal accepted, so we thought that we would be unable to review this project. But, in the end, the proposal was accepted and we reviewed this study. I will need to remove the data point that incorrectly files wsb038 as "False, for other reasons not stated above"

```
new_data <- new_data[!(new_data$Study.ID..e.g...wsb002.or.jwm040.. == "wsb038" & new_data$Was.this.pape

#Track down reasons for "False, author declined to participate…" What studies was this true for? new_da
```

This is an extension of the manuscript I found on github a while back and now have more bookdown skills to add some more notes.

Future update checks can be done by checking the github reprository: `aaarchmiller/reproducibility_in_wildlife_ecolog`

1.2 Overview

Why is this important….

- SSM modelling issues (secrets thing…)
- errors in life history structures are common (paper now)
- reproducibility is very important (SCIENCE CRISIS)
- predator free nz 2050
  - community groups are given concepts not details…. CEO etc relationship?

A literature review….

- Hard to start
- But as a computational workflow it is valueable to build
- very small collection of key points….
  - figures of concepts
- To find out key words…
  - `diagrammR`

Survey…

- extend from paper here…
- already have draft suvery
- email and ask??

1.3 Method

Lit review

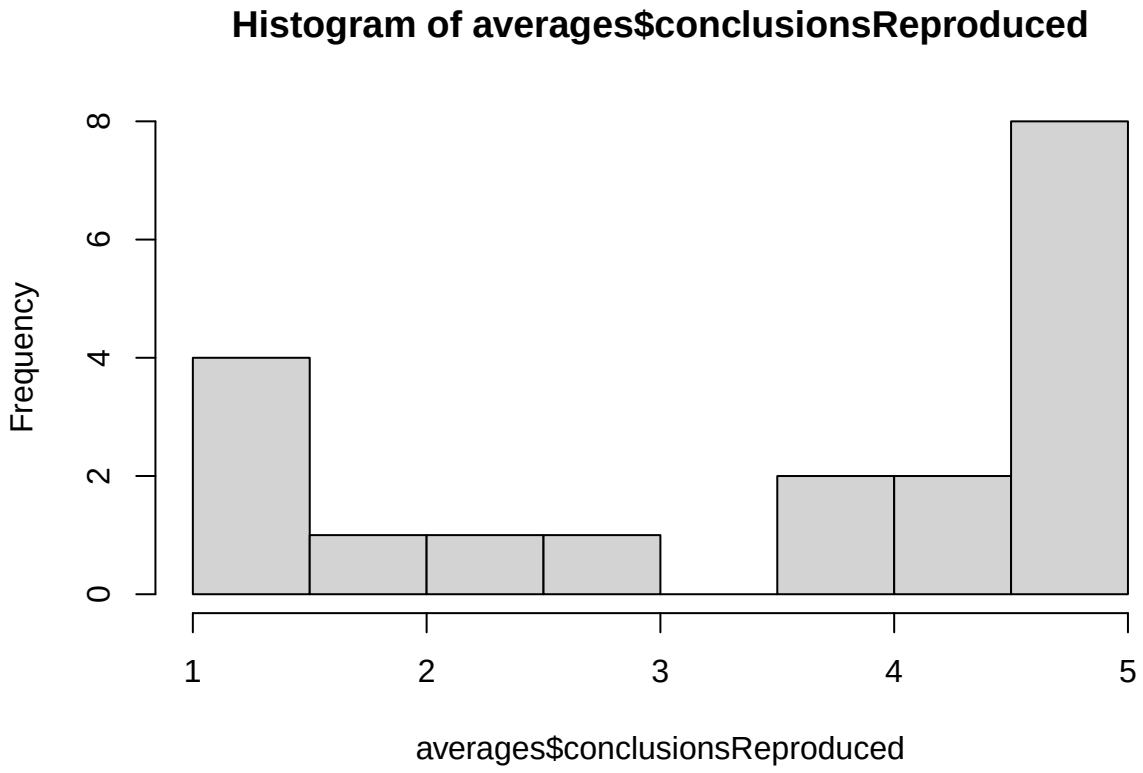
- collect all literature
  - Using this collection of keywords

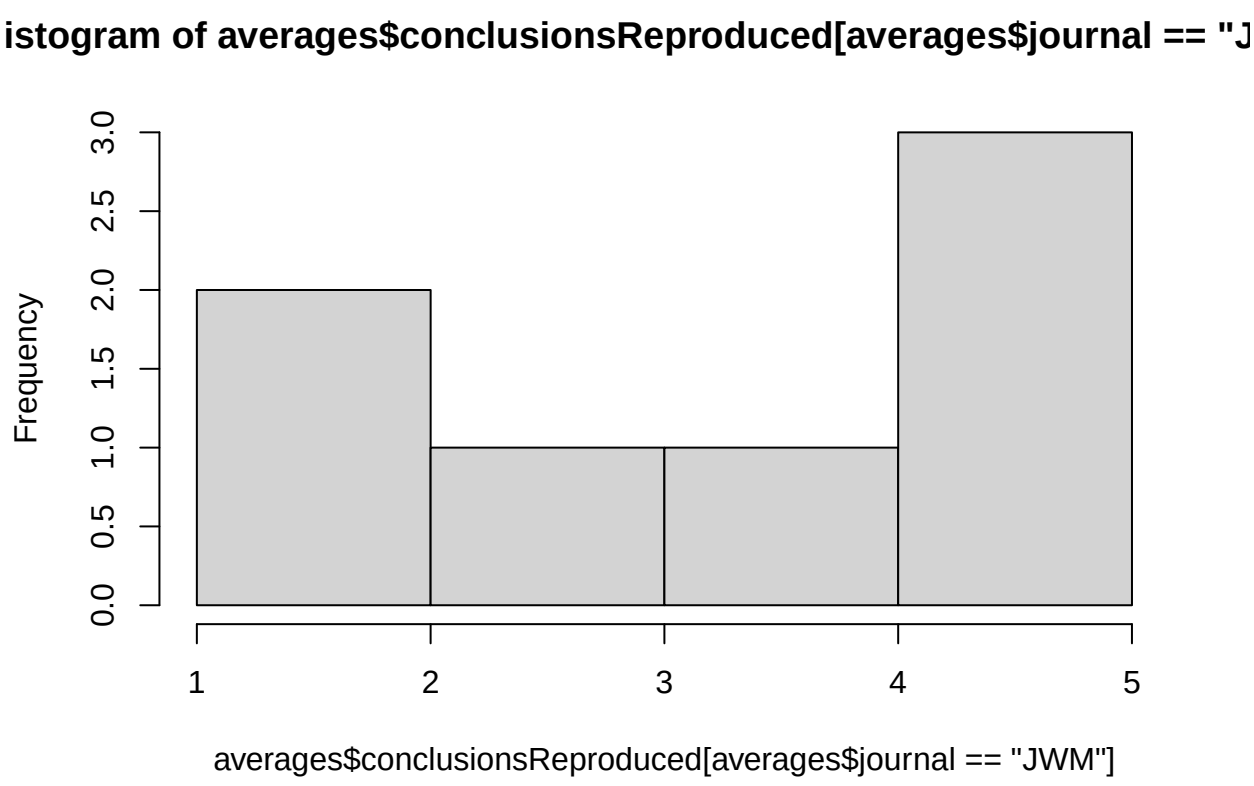
1.4 Study overview

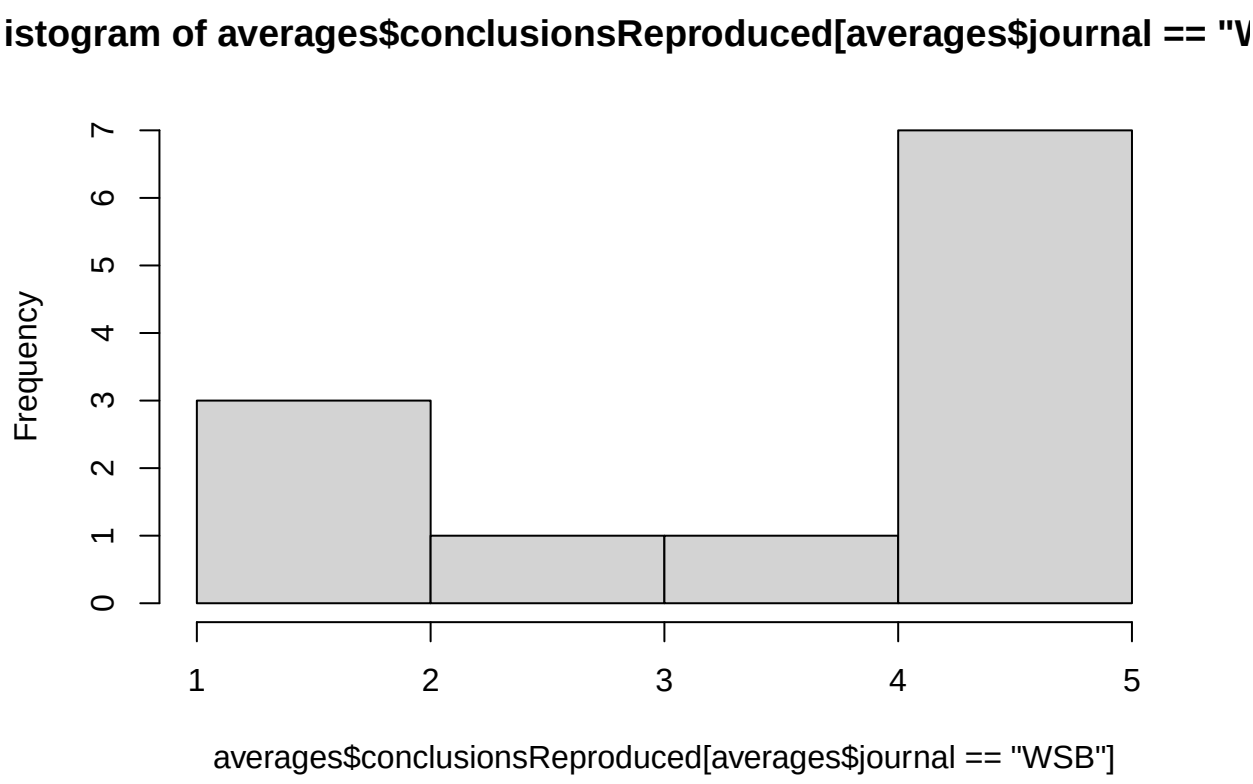
```
source("programs/02_study_overview_figure.R")
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
```




istogram of averages$conclusionsReproduced[averages$journal == "J
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
averages$conclusionsReproduced[averages$journal == "JWM"]



1.5 Scores

```
source("programs/03_scores_by_journal_figure.R")
```

1.6 Criteria

```
source("programs/04_reproducibility_criteria_figure.R")
```

```
##       Score        SD    upper     lower              Category
## 1  3.200000 1.8234583 5.023458 1.3765417     Figures reproduced
## 2  3.166667 1.5275252 4.694192 1.6391414     Figures reproduced
## 3  3.100000 1.2449900 4.344990 1.8550100     Numbers reproduced
## 4  3.153846 1.5053324 4.659179 1.6485138     Numbers reproduced
## 5  3.500000 1.6583124 5.158312 1.8416876 Conclusions reproduced
## 6  3.576923 1.6689087 5.245832 1.9080143 Conclusions reproduced
## 7  3.800000 1.3509256 5.150926 2.4490744     Figures reproduced
## 8  2.916667 1.6213537 4.538020 1.2953129     Figures reproduced
## 9  3.750000 0.8803408 4.630341 2.8696592     Numbers reproduced
## 10 2.833333 1.5423320 4.375665 1.2910014     Numbers reproduced
## 11 4.333333 1.4023789 5.735712 2.9309544 Conclusions reproduced
## 12 3.166667 1.6283474 4.795014 1.5383193 Conclusions reproduced
## 13 3.227273 1.7372915 4.964564 1.4899812     Figures reproduced
## 14 3.083333 1.3197222 4.403056 1.7636111     Figures reproduced
## 15 3.136364 1.6138604 4.750224 1.5225032     Numbers reproduced
## 16 3.142857 1.1073349 4.250192 2.0355223     Numbers reproduced
```

```
## 17 3.272727 1.8078113 5.080539 1.4649159 Conclusions reproduced
## 18 4.000000 1.2583057 5.258306 2.7416943 Conclusions reproduced
## 19 3.666667 1.6007811 5.267448 2.0658856     Figures reproduced
## 20 2.500000 2.1794495 4.679449 0.3205505     Figures reproduced
## 21 3.500000 1.5411035 5.041104 1.9588965     Numbers reproduced
## 22 2.666667 2.0816660 4.748333 0.5850007     Numbers reproduced
## 23 3.722222 1.6791201 5.401342 2.0431021 Conclusions reproduced
## 24 2.500000 2.1794495 4.679449 0.3205505 Conclusions reproduced
## 25 3.444444 1.5500896 4.994534 1.8943548     Figures reproduced
## 26 2.875000 1.6201852 4.495185 1.2548148     Figures reproduced
## 27 3.333333 1.4142136 4.747547 1.9191198     Numbers reproduced
## 28 2.944444 1.4457793 4.390224 1.4986651     Numbers reproduced
## 29 3.444444 1.5092309 4.953675 1.9352136 Conclusions reproduced
## 30 3.666667 1.8027756 5.469442 1.8638910 Conclusions reproduced
##                Question Response  n
## 1  Q5: Data available?      Yes  5
## 2  Q5: Data available?       No 12
## 3  Q5: Data available?      Yes  5
## 4  Q5: Data available?       No 13
## 5  Q5: Data available?      Yes  5
## 6  Q5: Data available?       No 13
## 7        Q6: Raw data?      Yes  5
## 8        Q6: Raw data?       No 12
## 9        Q6: Raw data?      Yes  6
## 10       Q6: Raw data?       No 12
## 11       Q6: Raw data?      Yes  6
## 12       Q6: Raw data?       No 12
## 13     Q8: Code based?      Yes 11
## 14     Q8: Code based?       No  6
## 15     Q8: Code based?      Yes 11
## 16     Q8: Code based?       No  7
## 17     Q8: Code based?      Yes 11
## 18     Q8: Code based?       No  7
## 19 Q7: Code available?      Yes  9
## 20 Q7: Code available?       No  3
## 21 Q7: Code available?      Yes  9
## 22 Q7: Code available?       No  3
## 23 Q7: Code available?      Yes  9
## 24 Q7: Code available?       No  3
## 25    Q9: Open source?      Yes  9
## 26    Q9: Open source?       No  8
## 27    Q9: Open source?      Yes  9
## 28    Q9: Open source?       No  9
## 29    Q9: Open source?      Yes  9
## 30    Q9: Open source?       No  9
```

1.7 Overall scripts

```
# source("programs/")
```

Gerenal visualz and patterns NZ? global?

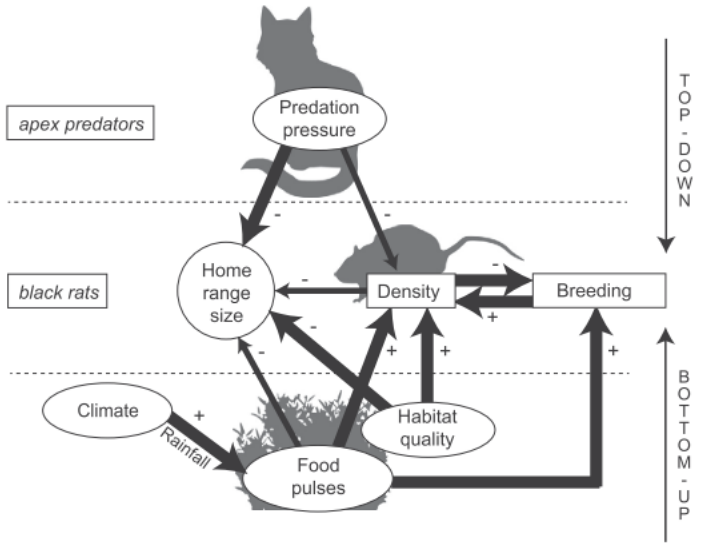


Fig. 6. Schematic representation of the main drivers of population (density, breeding) and behavioural traits (home range size) of black rats on Europa and Juan de Nova. Derived from Russell et al. (2011a) and Ringler et al. (2014). Width of arrows indicates strength of regulation.

Figure 1: image-20191202103427895

1.8 Statistical analysis

```
a <- list.files("programs/")
a
```

```
## [1] "01_data_processing.R"
## [2] "02_study_overview_figure.R"
## [3] "03_scores_by_journal_figure.R"
## [4] "04_reproducibility_criteria_figure.R"
## [5] "R_scripts.zip"
```

```
#added here package to help with file locations
source("programs/01_data_processing.R")
```

```
## here() starts at C:/Code/reproducibility_in_wildlife_ecology
```

```
## 'data.frame':    101 obs. of  14 variables:
##  $ Timestamp                                                                    : chr  "2018
##  $ Study.ID..e.g...wsb002.or.jwm040.                                            : chr  "wsb0
##  $ Reviewer.s.Initials..yours.                                                  : chr  "AAA
##  $ Was.this.paper.suitable.for.review..                                         : chr  "Tru
##  $ Was.this.paper.reviewed..                                                    : chr  "" "
##  $ Data.readily.available.                                                      : chr  "Fals
##  $ Was.the.data.pre.processed.or.shared.in.original.raw.format.                 : chr  "Pre
##  $ Code.available                                                               : chr  "Tru
##  $ Analysis.was.entirely.code.based                                             : chr  "Tru
##  $ Open.source.file.formats.used.                                               : chr  "Tru
##  $ Can.code.be.run.as.is..                                                      : int  5 4 1
##  $ Were.the.study.s.figures.reproduced..                                        : int  5 5 3
##  $ Were.the.numbers.in.the.study.s.tables.and.or.results.text.quantitatively.reproduced..: int  5 4 1
```

```
##  $ Were.the.study.s.conclusions.reproduced..                                    : int  5 4 4
## [1] "wsb028 error: data readily available"
## [1] "jwm019 error: open source formats"
## [1] "wsb013 error: code available"
## [1] "wsb013 error: open source formats"
## [1] "wsb035 error: data readily available"
## [1] "wsb035 error: was analysis code based"
## [1] "wsb004 error: was data pre processed"
## [1] "jwm010 error: was data pre processed"
## [1] "jwm010 error: was analysis code based"
## [1] "jwm010 error: Can.code.be.run.as.is.."
## [1] "wsb018 error: data readily available"
## [1] "wsb018 error: open source formats"
## [1] "wsb006 error: data readily available"
## [1] "wsb006 error: was data pre processed"
## [1] "wsb006 error: was analysis code based"
## [1] "wsb006 error: Can.code.be.run.as.is.."
## [1] "wsb006 error: open source formats"
## [1] "wsb023 error: was analysis code based"
## [1] "wsb010 error: was data pre processed"
## [1] "wsb010 error: Can.code.be.run.as.is.."
## [1] "wsb010 error: open source formats"
## [1] "wsb014 error: data readily available"
## [1] "wsb014 error: code available"
## [1] "wsb014 error: open source formats"
## [1] "jwm022 error: data readily available"
## [1] "jwm022 error: was data pre processed"
## [1] "jwm022 error: code available"
## [1] "jwm022 error: Can.code.be.run.as.is.."
## [1] "wsb008 error: data readily available"
## [1] "wsb008 error: code available"
## [1] "wsb008 error: Can.code.be.run.as.is.."
## [1] "jwm024 error: was analysis code based"
## [1] "jwm024 error: Can.code.be.run.as.is.."
## [1] "jwm028 error: was data pre processed"
## [1] "jwm028 error: was analysis code based"
## [1] "jwm028 error: Can.code.be.run.as.is.."
## [1] "jwm004 error: open source formats"
## [1] "wsb038 error: Can.code.be.run.as.is.."
## [1] "jwm027 error: open source formats"
```

Survey?

1.9 Results

1.10 Discussion

Need to just find and extract images from PDF….

Time and increase in diagram structure…

1.10.0.1 package to make them reproducible iswell….. here….