

oxforddown:
An Oxford University Thesis
Template for R Markdown



Author Name
Your College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2018

For Yihui Xie

Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (<http://pandoc.org>) which converts Markdown to a large number of output formats, and to Yihui Xie for creating `knitr` which introduced R Markdown as a way of embedding code in Markdown documents, and `bookdown` which added tools for technical and longer-form writing.

Special thanks to [Chester Ismay](#), who created the `thesisdown` package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John McManigle, whose adaption of Sam Evans' adaptation of Keith Gillow's original maths template for writing an Oxford University DPhil thesis in LaTeX provided the template that I in turn adapted for R Markdown.

Finally, profuse thanks to JJ Allaire, the founder and CEO of [RStudio](#), and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.

Ulrik Lyngs
Linacre College, Oxford
2 December 2018

Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown` and the `OxThesis` L^AT_EX template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to write a thesis in R Markdown, and largely follows the structure from [this R Markdown workshop](#).

Congratulations for taking a step further into the lands of open, reproducible science by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

Contents

List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
1 Einleitung	1
2 Accounting Fraud und Machine Learning	3
3 Datensatz	6
4 Methoden	8
5 Customisations and extensions	15
5.1 Front matter	16
5.2 Shorten running header (PDF)	17
5.3 Unnumbered chapters	17
5.4 Beginning chapters with quotes (PDF)	17
5.5 Highlighting corrections (HTML & PDF)	18
5.6 Apply custom font color and highlighting to text (HTML & PDF) .	19
5.7 Including another paper in your thesis - embed a PDF document . .	20
5.8 Including another paper in your thesis - R Markdown child document	24
5.9 Customizing referencing	28
5.10 Customizing the page headers and footers (PDF)	31
5.11 Diving in to the OxThesis LaTeX template (PDF)	32
5.12 Customising to a different university	32
6 Troubleshooting	34
6.1 Error: Failed to build the bibliography via biber	34

Contents

Conclusion	36
More info	36
Appendices	
A The First Appendix	38
B The Second Appendix, for Fun	39
Works Cited	40

List of Figures

List of Tables

List of Abbreviations

- 1-D, 2-D** . . . One- or two-dimensional, referring **in this thesis** to spatial dimensions in an image.
- Otter** One of the finest of water mammals.
- Hedgehog** . . . Quite a nice prickly friend.

1

Einleitung

Vergangene und gegenwärtige Bilanzskandale wie ENRON, Worldcom oder Wirecard führen immer wieder zu Diskussionen über die Vertrauenswürdigkeit des Kapitalmarkts sowie über die Vertrauenswürdigkeit der Abschlussprüfung. Bilanzskandale führen weit-reichende Konsequenzen mit sich welche eine genauere Analyse des Warums und wie anstoßen (vgl. Boecker/Zwirner 2012, S. 1). Diese Arbeit hat das Ziel, den Einsatz neuronaler Netze gegenüber der gängigeren Methode der logistischen Regression (vgl. Bao et al., 2020) zu vergleichen und dabei herauszufinden, ob erstere signifikant besser darin sind, so viele Betrugsfälle wie möglich zu identifizieren, ohne die triviale Annahme zu treffen, dass jeder Fall ein Betrugsfall ist. Bei der Modellevaluation wird angenommen, dass ein nicht-identifizierter Betrugsfall doppelt so schwer wiegt, wie ein prognostizierter Betrugsfall, der sich als Nicht-Betrug herausstellt. Die Metrik zur Messung der Ergebnisse ist der F_β -Score (vgl. Tharwat 2020, S. 174). Dieser bildet das harmonische Mittel aus Präzision und Sensitivität. Die Präzision sagt aus, wie viele vorhergesagte Betrugsfälle tatsächlich Betrugsfälle sind, wobei die Sensitivität dabei auf die Frage antwortet, wie viele der tatsächlichen Betrugsfälle als solche identifiziert wurden. Weil die Sensitivität im Kontext der Fraud Detection bedeutungs-voller erscheint, wird sie in dieser Arbeit

1. *Einleitung*

doppelt so hoch gewichtet wie die Präzision. Die Formel dieser Metrik lautet dabei:

$$F_{\beta=2} = (1 + \beta^2) * \frac{P * R}{\beta * P * R}$$

2

Accounting Fraud und Machine Learning

Das Handelsgesetzbuch (HGB) sieht gemäß § 317 Absatz (Abs.) 1 Satz (S.) 3 vor, dass der Abschlussprüfer im Rahmen der Abschlussprüfung, Unrichtigkeiten und Verstöße welche der Ordnungsmäßigkeit des Abschlusses entgegen stehen, erkennt und entsprechend deklariert. Was unter den Begrifflichkeiten Unrichtigkeit und Verstöße zu verstehen ist, wird durch den Gesetzgeber an dieser Stelle nicht weiter konkretisiert (vgl. Zwernemann et al. 2015, S. 22; § 317 Abs.1 HGB).

Dem entgegen versucht das Institut der Wirtschaftsprüfer (IDW) mit dem veröffentlichten Prüfungsstandard 210 (IDW PS 210) Rechnung zu tragen. Diesem zu entnehmen ist, dass ein Fehlerhafter Abschluss entweder auf Fraud (Verstoß) oder Error (Unrichtigkeit) zurückzuführen ist. Unter dem Begriff Unrichtigkeit wird eine unabsichtliche Angabe im Abschluss verstanden. Konkret bedeutet dies begangene Rechenfehler, eine unbewusst falsche Anwendung von Rechnungslegungsgrundsätzen sowie die falsche Einschätzung von Sachverhalten (vgl. Hlavica et al. 2016, S. 209f.). Der Begriff Verstoß dagegen umfasst eine beabsichtigte Handlung mit dem Ziel rechtswidrige Vorteile zu realisieren. Diese Handlungen konkretisiert der IDW PS 210 als Vermögensschädigungen, Täuschungen und Gesetzesverstöße, welche eine Auswirkung auf die Rechnungslegung zur Folge haben (vgl. Zwernemann et al. 2015, S. 8). Um die Gründe für eine betrügerische Handlung nachvollziehen zu können,

2. Accounting Fraud und Machine Learning

entwickelte Donald Cressey in den 1940er Jahren das sogenannte „Fraud-Triangle“. Dieses Dreieck wird ferner dem IDW PS 210 zugrunde gelegt (vgl. Boecker/Zwirner 2012, S. 2f.). Demnach tritt ein Verstoß dann auf, wenn drei Gegebenheiten als erfüllt angesehen werden können. So muss der Täter eine Gelegenheit zu der Tat haben und einen Anreiz (Motivation) für die Tat verspüren. Als letztes muss der Täter die Tat als moralisch akzeptabel rechtfertigen vor sich selbst rechtfertigen (vgl. Schuchter/Levi 2016, S. 3f.). Aber nicht nur durch psychologische Ansätze versucht die Wissenschaft Verstöße einzuordnen und zu identifizieren, sondern auch durch eine Vielzahl an Machine Learning Ansätzen, welche Verstöße mittels Algorithmen identifizieren sollen. Vor dem Hintergrund der potenziellen Gefahren des Bilanzbetrugs werden Letztere zunehmend für die Vorhersage und Aufdeckung von diskretionärer Bilanzpolitik eingesetzt. Die Benchmark in diesem Bereich ist das Dechow et al. Modell, welches auf Grundlage von Accounting and Auditing Enforcement Releases (AAERs) der U.S. Securities and Exchange Commission (SEC) mit Hilfe einer logistischen Regression die Wahrscheinlichkeiten von (bewusst) fehlerhaften Darstellungen schätzt und klassifiziert (Dechow et al. 2011). Hierbei gelten die AAERs als ProxyVariable für die Manipulation der Bilanz. Durch das Voraussetzen der Untersuchungshandlungen seitens der SEC ergibt sich der Vorteil, dass der Typ I Fehler – das Modell sagt fälschlicherweise ein misstatement voraus – deutlich geringer ausfällt. Durch einige wenige Transformationen der logistischen Funktion kann der Einfluss einer jeden unabhängigen Variable durch den entsprechenden Regressionskoeffizienten hinsichtlich der Effektgröße verglichen werden, weswegen die Ergebnisse gut interpretierbar sind. Aus dem Dechow et al. Modell folgt eine korrekte Klassifizierung von misstatements und nonmisstatements von ungefähr 63% (Dechow et al. 2011, S.59). Die Sensitivität, d.h. in wie vielen Fällen das Modell ein misstatement richtig vorhergesagt hat, liegt bei etwas mehr als 68% (339 von 494). Der Typ II Fehler (das Modell klassifiziert ein misstatement als nonmisstatement), der im Rahmen des accounting frauds schwerwiegender ist als der Typ I Fehler (vgl. Lin et al. 2015, S. 468), liegt bei etwas mehr als 31% (155 von 494). Ein Vergleich der Performance der

2. Accounting Fraud und Machine Learning

logistischen Regression mit den neuronalen Netzen, einer weiteren Methode zur Vorhersage von Bilanzbetrug, findet sich in dem Paper von Lin et al. (2015). Aus diesem geht hervor, dass die neuronalen Netze hinsichtlich der Aufdeckung von accounting fraud bessere Ergebnisse liefern als die logistische Regression. Die artificial neural networks (ANNs) erreichen bei dem Testdatensatz eine Genauigkeit von fast 93%. Die Sensitivität von fast 83% liegt zudem deutlich höher als bei der logistischen Regression, bei der 72% der misstatements richtig vorhergesagt wurden (vgl. Lin et al. 2015, S. 465f.). Die Interpretierbarkeit und verhältnismäßig einfache Anwendbarkeit haben die logistische Regression zu einem beliebten Instrument gemacht, die Ergebnisse hinsichtlich der Vorhersage von Bilanzbetrug werden allerdings von anderen Modellen übertroffen (vgl. Dutta et al. 2017, S. 375). Im Falle der neuronalen Netze ergibt sich wiederum der Nachteil der geringeren Transparenz hinsichtlich der Arbeitsweise des Algorithmus (vgl. Bao et al. 2020, S. 228). Schlussendlich ergibt sich ein Trade-Off zwischen der Interpretierbarkeit und Vorhersagekraft. Nachfolgend liegt der Fokus dieser Ausarbeitung auf der reinen Performance respektive Vorhersagekraft des Modells. Ein Modell, welches misstatements richtig vorhersagt, erscheint bei der Klassifizierung von Bilanzbetrug wichtiger als ein Modell, aus dem abgelesen werden kann, welche Variablen den misstatement wie stark beeinflussen. An dieser Stelle sollte erwähnt werden, dass auch ein Modell, welches in 99% der Fälle die richtige Vorhersage trifft, hinsichtlich der Aufdeckung von accounting fraud nicht geeignet sein muss. Durch die Problematik der signifikanten sample imbalance – Betrugsfälle sind stark unterrepräsentiert – ist es möglich, dass misstatements durch das Modell nicht erkannt respektive falsch klassifiziert werden, also Typ II Fehler auftreten können.

3

Datensatz

Ein besonderer Fokus liegt aus diesem Grund auf der Minimierung des Typ II Fehlers respektive der Maximierung der Sensitivität des Modells. Eine höchstmögliche Genauigkeit ist gut, aufgrund der sehr geringen Anzahl an bilanziellen Verfehlungen verglichen zu der Stichprobengröße sollte hier allerdings kein Schwerpunkt liegen. Zum Training und Test des Machine Learning Modells, wird ein Datensatz aus der Veröffentlichung von Bao et al. (2020) verwendet. Die Autoren haben diesen im Internet auf der Seite „GitHub“ zur Verfügung gestellt (vgl. Bao et al. 2020, GitHub Repository). Dieser besteht aus allen öffentlich gelisteten US-amerikanischen Firmen im Zeitraum von 1991 bis 2008. Die Accounting-Betrugsfälle aus den „Accounting and Auditing Enforcement Releases“ (AAER), die von der United States Securities and Exchange Commission (SEC) im gleichen Zeitraum veröffentlicht worden sind (vgl. Bao et al. 2020, S. 207). Der Datensatz listet für jeden Eintrag 28 verschiedene finanzielle Items auf. Diese setzen sich aus den Veröffentlichungen von Cecchini et al. (2010) und Dechow et al. (2011) zusammen. Die finanziellen Items stammen aus vier verschiedenen Bereichen, der Bilanz (z.B. gesamte Forderungen), der Gewinn- und Verlustrechnung (z.B. Nettoumsatz), der Kapitalflussrechnung (z.B. Langzeitemission von Schuldtiteln) und dem Marktwert (z.B. Common Shares Outstanding). Da Accounting-Betrugsfälle eher seltener vorkommen (vgl. Dutta et

3. Datensatz

al. 2011, S. 381), weist der Datensatz eine große Verteilungsungleichheit zwischen den Betrugs- und Nicht-betrugsfällen auf. Weniger als ein Prozent aller Einträge im Datensatz sind hierbei Betrugsfälle. Weitere Betrugsfälle könnten über andere Datenbanken gesucht werden, hierbei gibt es allerdings das Problem, dass die Betrugsfälle identifiziert und einzeln herausgesucht werden müssen. Daher besteht hier ein „class imbalance“ Problem, das mithilfe des Machine Learning Algorithmus gelöst werden muss. Zudem sind nur Betrugsfälle bis zum Jahr 2008 in diesem Datensatz erhalten, da die SEC nach der Finanzkrise ihre Prioritäten geändert hat (vgl. Bao et al. 2020, S. 208). Sollte sich die Art und Weise mit der Accounting-Betrug durchgeführt wird in den Jahren danach geändert haben, so können diese Fälle unter Umständen nicht vom Algorithmus erkannt werden.

4

Methoden

Diese Arbeit vergleicht die Vorhersagequalität der in der Accounting Fraud Detection gängigen logistischen Regression nach Dechow et al. (vgl. Bao et al. 2020, S. 2) mit der von neuronalen Netzen. Die logistische Regression ist eng mit der linearen Regression verwandt und wird zur binären Schätzung einer Klassenzugehörigkeit verwandt (vgl. Géron 2019, S. 144). Dabei berechnet sie eine gewichtete Summe von Inputfaktoren und aggregiert sie zu einer Wahrscheinlichkeit zwischen 0 und 1. Liegt diese Wahrscheinlichkeit bei mindestens 0.5, so wird die Klasse „1“ vorhergesagt, welcher in dieser Arbeit der Klasse „fraud“ entspricht (vgl. Géron 2019, S. 144). Hierzu wird für jedes Attribut einer Beobachtung eine Sigmoid-Funktion verwendet, welche S-förmig vom Minimum bis zum Maximum des jeweiligen Attributs verläuft und die Verteilungen der Merkmalsausprägungen möglichst gut nach Klassenzugehörigkeit abgrenzt. Je weiter eine Merkmalsausprägung von dieser Grenze entfernt ist, desto näher ist der Funktionswert an der 1 oder der 0 (vgl. Géron 2019, S. 148). Neuronale Netze bestehen aus drei Sorten von Schichten: Input-Schichten, welche Daten einlesen, versteckte Schichten, welche die Daten verarbeiten und Output-Schichten, welche aus den verarbeiteten Daten eine Prognose ableiten (vgl. Géron 2019, S. 286). In dieser Arbeit besteht die Output-Schicht aus lediglich einem Knoten, welche die Klassen „fraud“ abbildet. Die Anzahl der Knoten der

4. Methoden

Input-Schicht entspricht der Anzahl der Variablen im Datensatz. Die Knoten einer Schicht sind jeweils mit jedem Knoten seiner nachfolgenden Schicht durch „Gewichte“ verbunden. Jeder einzelne Knoten aggregiert die Signale, die er empfängt über deren Gewichte zu einer Zahl und wendet eine Aktivierungsfunktion an (vgl. Géron 2019, S. 282). Übersteigt der Funktionswert einen gegebenen Schwellenwert, so „feuert“ das Neuron, was bedeutet, dass es ein Signal größer 0 an die Neuronen der nächsten Schicht weitergibt (vgl. Géron 2019, S. 282-283). Die Gewichte und alle Schwellenwerte werden durch Backpropagation unter Zuhilfenahme des Gradient Descent Algorithmus verbessert (vgl. Géron 2019, S.119 und S, 286). Sind alle Trainingsdaten einmal zum Training herangezogen worden, bedeutet das, dass das Netz für „eine Epoche“ trainiert wurde (vgl. Géron 2019, S. 127). In dieser Arbeit wird ein Netz über 100 Epochen hinweg trainiert.

```
library(readr)
data <- read_csv("data/data.csv")

##
## -- Column specification -----
## cols(
##   .default = col_double()
## )
## i Use 'spec()' for the full column specifications.

# str(data)

dim(data)

## [1] 31990    51
```

Wir haben 31990 Beobachtungen auf 51 Variablen. Wir müssen die fehlenden Werte identifizieren und die Spalten, welche zu viele davon enthalten löschen. Danach löschen wir die übrigen Zeilen, welche NaNs enthalten. Dies dient dem Umstand, dass NN mit NaN-Werten nicht arbeiten können. Wir verwenden für LogReg und NN den gleichen Datensatz, um die Performance der Modelle vergleichen zu können.

4. Methoden

```
for (col in colnames(data)){  
  nan_number = sum(is.na(data[col]))  
  print(paste(col, nan_number ))  
}
```

```
## [1] "fyear 0"  
## [1] "gvkey 0"  
## [1] "sich 199"  
## [1] "insbnk 0"  
## [1] "understatement 0"  
## [1] "option 0"  
## [1] "p_aaer 31846"  
## [1] "new_p_aaer 31846"  
## [1] "misstate 0"  
## [1] "act 0"  
## [1] "ap 0"  
## [1] "at 0"  
## [1] "ceq 0"  
## [1] "che 0"  
## [1] "cogs 0"  
## [1] "csho 0"  
## [1] "dlc 0"  
## [1] "dltis 0"  
## [1] "dltt 0"  
## [1] "dp 0"  
## [1] "ib 0"  
## [1] "invst 0"  
## [1] "ivao 0"  
## [1] "ivst 0"  
## [1] "lct 0"  
## [1] "lt 0"
```

4. Methoden

```
## [1] "ni 0"  
## [1] "ppegd 0"  
## [1] "pstk 0"  
## [1] "re 0"  
## [1] "rect 0"  
## [1] "sale 0"  
## [1] "sstk 0"  
## [1] "txp 0"  
## [1] "txt 0"  
## [1] "xint 0"  
## [1] "prcc_f 0"  
## [1] "dch_wc 723"  
## [1] "ch_rsst 734"  
## [1] "dch_rec 678"  
## [1] "dch_inv 681"  
## [1] "soft_assets 13"  
## [1] "ch_cs 3631"  
## [1] "ch_cm 3763"  
## [1] "ch_roa 3380"  
## [1] "issue 0"  
## [1] "bm 6"  
## [1] "dpi 1084"  
## [1] "reoa 13"  
## [1] "EBIT 13"  
## [1] "ch_fcf 836"
```

Die Variable, “p_aaer” und “new_p_aaer” fallen durch extrem viele NaN-Werte auf und werden komplett entfernt. Die übrigen Variablen bestehen zu maximal ca 10% aus NaN-Werten und werden daher als unbedenklich eingestuft. “new_p_aaer” und alle Beobachtungen, welche NaN-Werte enthalten, werden beseitigt.

4. Methoden

```
data <- data[, -match(c("new_p_aaer", "p_aaer"), names(data))]  
data <- data[complete.cases(data),]  
dim(data)
```

```
## [1] 27792    49
```

```
print(data[5,])
```

```
## # A tibble: 1 x 49
```

```
##   fyear gvkey  sich insbnk understatement option misstate  act    ap    at  
##   <dbl> <dbl> <dbl>  <dbl>          <dbl>  <dbl>    <dbl> <dbl> <dbl> <dbl>  
## 1  1990  1028  7385      0              0      0      0  17.3  3.52  27.5  
## # ... with 39 more variables: ceq <dbl>, che <dbl>, cogs <dbl>, csho <dbl>,  
## #   dlc <dbl>, dltis <dbl>, dlts <dbl>, dp <dbl>, ib <dbl>, invt <dbl>,  
## #   ivao <dbl>, ivst <dbl>, lct <dbl>, lt <dbl>, ni <dbl>, ppeg <dbl>,  
## #   pstk <dbl>, re <dbl>, rect <dbl>, sale <dbl>, sstk <dbl>, txp <dbl>,  
## #   txt <dbl>, xint <dbl>, prcc_f <dbl>, dch_wc <dbl>, ch_rsst <dbl>,  
## #   dch_rec <dbl>, dch_inv <dbl>, soft_assets <dbl>, ch_cs <dbl>, ch_cm <dbl>,  
## #   ch_roa <dbl>, issue <dbl>, bm <dbl>, dpi <dbl>, reoa <dbl>, EBIT <dbl>,  
## #   ch_fcf <dbl>
```

Damit ein NN Inputdaten verwerten kann, müssen diese zwischen 0 und 1 normiert sein.

```
normalize <- function(x, na.rm = TRUE) {  
  return((x- min(x)) /(max(x)-min(x)))  
}  
# as.data.frame(apply(df$name, normalize))  
for (col in colnames(data))  
{  
  data[col] <- normalize(data[col])  
}
```

4. Methoden

```
print(data[5,])

## # A tibble: 1 x 49
##   fyear    gvkey  sich insbnk understatement option misstate    act    ap
##   <dbl>    <dbl> <dbl>  <dbl>          <dbl>  <dbl>    <dbl>    <dbl>
## 1      0 0.000156 0.736      0          0      0      0 0.000262 0.000180
## # ... with 40 more variables: at <dbl>, ceq <dbl>, che <dbl>, cogs <dbl>,
## #   csho <dbl>, dlc <dbl>, dltis <dbl>, dlтт <dbl>, dp <dbl>, ib <dbl>,
## #   invt <dbl>, ivao <dbl>, ivst <dbl>, lct <dbl>, lt <dbl>, ni <dbl>,
## #   ppgt <dbl>, pstk <dbl>, re <dbl>, rect <dbl>, sale <dbl>, sstk <dbl>,
## #   txp <dbl>, txt <dbl>, xint <dbl>, prcc_f <dbl>, dch_wc <dbl>,
## #   ch_rsst <dbl>, dch_rec <dbl>, dch_inv <dbl>, soft_assets <dbl>,
## #   ch_cs <dbl>, ch_cm <dbl>, ch_roa <dbl>, issue <dbl>, bm <dbl>, dpi <dbl>,
## #   reoa <dbl>, EBIT <dbl>, ch_fcf <dbl>
```

Jetzt wird ein NN trainiert und eine erste Prognose abgegeben. Seed = 42

```
library(nnet)
set.seed(42)
inTrain <- runif(nrow(data)) < 0.2

model <- nnet(misstate ~ ., data=data[inTrain,],
              size=15, maxit=100, rang=0.1, decay=5e-4)

## # weights: 751
## initial value 1761.201764
## iter 10 value 33.571159
## iter 20 value 24.129969
## iter 30 value 24.030039
## iter 40 value 24.029168
## iter 50 value 24.028222
## iter 60 value 24.027147
```

4. Methoden

```
## iter 70 value 24.025834
## iter 80 value 24.024014
## iter 90 value 24.020810
## iter 100 value 24.011132
## final value 24.011132
## stopped after 100 iterations
```

```
pred <- predict(model, data[-inTrain,], type="raw")
```

Nun werden die Ergebnisse für einen Score verwendet.

```
cm <- table(pred=pred,
true=data$misstate[-inTrain])
accuracy = sum(diag(cm))/sum(sum(cm))
print(accuracy)
```

```
## [1] 3.598287e-05
```

Haha :D Verdammt, da lief irgendwas falsch. Ich denke, dass `type="raw"` bei der prediction quatsch ist und dass da stattdessen `"class"` rein müsste. Allerdings. funktioniert das dann aus Gründen nicht. morgen oder die Tage mal weiterprobieren.

There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

— Charles Darwin (Darwin 1859)

5

Customisations and extensions

Contents

5.1	Front matter	16
5.1.1	Shorten captions shown in the list of figures (PDF)	16
5.1.2	Shorten captions shown in the list of tables (PDF)	16
5.2	Shorten running header (PDF)	17
5.3	Unnumbered chapters	17
5.4	Beginning chapters with quotes (PDF)	17
5.5	Highlighting corrections (HTML & PDF)	18
5.5.1	Short, inline corrections	18
5.5.2	Blocks of added or changed material	19
5.5.3	Stopping corrections from being highlighted	19
5.6	Apply custom font color and highlighting to text (HTML & PDF)	19
5.7	Including another paper in your thesis - embed a PDF document	20
5.8	Including another paper in your thesis - R Markdown child document	24
5.8.1	An example paper in another folder	24
5.8.2	Step 1: Include paper as a child document	25
5.8.3	Step 2: Make file paths compatible	25
5.8.4	Step 3: Make sure header levels are correct	26
5.8.5	Step 4: Make sure figure widths are correct	27
5.9	Customizing referencing	28
5.9.1	Using a .csl file with pandoc instead of biblatex	28
5.9.2	Customizing biblatex and adding chapter bibliographies	28
5.10	Customizing the page headers and footers (PDF)	31
5.11	Diving in to the OxThesis LaTeX template (PDF)	32
5.12	Customising to a different university	32
5.12.1	The minimal route	32

This chapter describes a number of additional tips and tricks as well as possible customizations to the `oxforddown` thesis.

5.1 Front matter

5.1.1 Shorten captions shown in the list of figures (PDF)

You might want your list of figures (which follows the table of contents) to have shorter (or just different) figure descriptions than the actual figure captions.

Do this using the chunk option `fig.scap` ('short caption'), for example `{r captain-image, fig.cap="A very long and descriptive (and potentially boring) caption that doesn't fit in the list of figures, but helps the reader understand what the figure communicates.", fig.scap="A concise description for the list of figures"}`

5.1.2 Shorten captions shown in the list of tables (PDF)

You might want your list of tables (which follows the list of figures in your thesis front matter) to have shorter (or just different) table descriptions than the actual table captions.

If you are using `knitr::kable` to generate a table, you can do this with the argument `caption.short`, e.g.:

```
knitr::kable(mtcars,
              caption = "A very long and descriptive (and potentially
              boring) caption that doesn't fit in the list of figures,
              but helps the reader understand what the figure
              communicates.",
              caption.short = "A concise description for the list of tables")
```

5.2 Shorten running header (PDF)

You might want a chapter's running header (i.e. the header showing the title of the current chapter at the top of page) to be shorter (or just different) to the actual chapter title.

Do this by adding the latex command `\chaptermark{My shorter version}` after your chapter title.

For example, chapter ??'s running header is simply 'Cites and cross-refs', because it begins like this:

```
# Citations, cross-references, and collaboration {#cites-and-refs}
\chaptermark{Cites and cross-refs}
```

5.3 Unnumbered chapters

To make chapters unnumbered (normally only relevant to the Introduction and/or the Conclusion), follow the chapter header with `{-}`, e.g. `# Introduction {-}`.

When you do this, you must also follow the heading with these two latex commands:

```
\adjustmtc
\markboth{The Name of Your Unnumbered Chapter}{{}}
```

Otherwise the chapter's mini table of contents and the running header will show the previous chapter.

5.4 Beginning chapters with quotes (PDF)

The OxThesis LaTeX template lets you inject some wittiness into your thesis by including a block of type `savequote` at the beginning of chapters. To do this, use the syntax ````\block type='savequote'}`.¹

¹For more on custom block types, see the relevant section in *Authoring Books with R Markdown*.

5. Customisations and extensions

Add the reference for the quote with the chunk option `quote_author="my author name"`. You will also want to add the chunk option `include=knitr::is_latex_output()` so that quotes are only included in PDF output.

It's not possible to use markdown syntax inside chunk options, so if you want to e.g. italicise a book name in the reference use a `'text reference'`: Create a named piece of text with `'(ref:label-name) My text'`, then point to this in the chunk option with `quote_author='(ref:label-name)'`.

5.5 Highlighting corrections (HTML & PDF)

For when it comes time to do corrections, you may want to highlight changes made when you submit a post-viva, corrected copy to your examiners so they can quickly verify you've completed the task. You can do so like this:

5.5.1 Short, inline corrections

Highlight **short, inline corrections** by doing `[like this]{.correction}` — the text between the square brackets will then **be highlighted in blue** in the output.

Note that pandoc might get confused by citations and cross-references inside inline corrections. In particular, it might get confused by `"[what @Shea2014 said]{.correction}"` which becomes (what Shea et al. 2014, said){.correction}. In such cases, you can use LaTeX syntax directly. The correction highlighting uses the `soul` package, so you can do like this:

- If using biblatex for references, use `"\hl{what \textcite{Shea2014} said}"`
- If using natbib for references, use `"\hl{what \cite{Shea2014} said}"`

Using raw LaTeX has the drawback of corrections then not showing up in HTML output at all, but you might only care about correction highlighting in the PDF for your examiners anyway!

5. Customisations and extensions

5.5.2 Blocks of added or changed material

Highlight entire **blocks of added or changed material** by putting them in a block of type `correction`, using the syntax ````{block type='correction'}`.² Like so:

For larger chunks, like this paragraph or indeed entire figures, you can use the `correction` block type. This environment **highlights paragraph-sized and larger blocks** with the same blue colour.

Note that correction blocks cannot be included in word output.

5.5.3 Stopping corrections from being highlighted

To turn off correction highlighting, go to the YAML header of `index.Rmd`, then:

- PDF output: set `corrections: false`
- HTML output: remove or comment out - `templates/corrections.css`

5.6 Apply custom font color and highlighting to text (HTML & PDF)

The lua filter that adds the functionality to highlight corrections adds two more tricks: you can apply your own choice of colour to highlight text, or change the font color. The syntax is as follows:

Here's `[some text in pink highlighting]{highlight="pink"}`
Becomes: Here's `some text in pink highlighting`.

`[Here's some text with blue font]{color="blue"}`
Becomes: `Here's some text with blue font`

Finally — never, ever actually do this — `[here's some text with black highlighting and yellow font]{highlight="black" color="yellow"}`
Becomes: `here's some text with black highlighting and yellow font`

²In the `.tex` file for PDF output, this will put the content between `\begin{correction}` and `\end{correction}`; in gitbook output it will be put between `<div class="correction">` and `</div>`.

5. Customisations and extensions

The file `scripts__and__filters/colour__and__highlight.lua` implements this, if you want to fiddle around with it. It works with both PDF and HTML output.

5.7 Including another paper in your thesis - embed a PDF document

You may want to embed existing PDF documents into the thesis, for example if your department allows a ‘portfolio’ style thesis and you need to include an existing typeset publication as a chapter.

In gitbook output, you can simply use `knitr::include_graphics` and it should include a scrollable (and downloadable) PDF. You will probably want to set the chunk options `out.width='100%'` and `out.height='1000px'`:

```
knitr::include_graphics("figures/sample-content/pdf_embed_example/Lyngs2020_FB.pdf")
```

In LaTeX output, however, this approach can cause odd behaviour. Therefore, when you build your thesis to PDF, split the PDF into an alphanumerically sorted sequence of **single-page** PDF files (you can do this automatically with the package `pdftools`). You can then use the appropriate LaTeX command to insert them, as shown below (for brevity, in the `oxforddown` PDF sample content we’re only including two pages). *Note that the chunk option `results='asis'` must be set.* You may also want to remove margins from the PDF files, which you can do with Adobe Acrobat (paid version) and likely other software.

```
# install.packages(pdftools)
# split PDF into pages stored in
  figures/sample-content/pdf_embed_example/split/
#
  pdftools::pdf_split("figures/sample-content/pdf_embed_example/Lyngs2020_FB.pdf")
# output = "figures/sample-content/pdf_embed_example/split/"
# grab the pages
```

5. Customisations and extensions

```
pages <- list.files("figures/sample-content/pdf_embed_example/split",
  full.names = TRUE)

# set how wide you want the inserted PDFs to be:
# 1.0 is 100 per cent of the oxforddown PDF page width;
# you may want to make it a bit bigger
pdf_width <- 1.2

# for each PDF page, insert it nicely and
# end with a page break
cat(stringr::str_c("\\newpage \\begin{center}
  \\makebox[\\linewidth][c]{\\includegraphics[width=", pdf_width,
  "\\linewidth]{", pages, "}} \\end{center}"))
```

‘I Just Want to Hack Myself to Not Get Distracted’: Evaluating Design Interventions for Self-Control on Facebook

Ulrik Lyngs¹, Kai Lukoff², Petr Slovak³, William Seymour¹, Helena Webb¹,
Marina Jirotko¹, Jun Zhao¹, Max Van Kleek¹, Nigel Shadbolt¹

¹Department of Computer Science, University of Oxford, UK, {first.last}@cs.ox.ac.uk

²Human Centered Design & Engineering, University of Washington, Seattle, US, kai.l@uw.edu

³Department of Informatics, King’s College London, UK, petr.slovak@kcl.ac.uk

ABSTRACT

Beyond being the world’s largest social network, Facebook is for many also one of its greatest sources of digital distraction. For students, problematic use has been associated with negative effects on academic achievement and general wellbeing. To understand what strategies could help users regain control, we investigated how simple interventions to the Facebook UI affect behaviour and perceived control. We assigned 58 university students to one of three interventions: goal reminders, removed newsfeed, or white background (control). We logged use for 6 weeks, applied interventions in the middle weeks, and administered fortnightly surveys. Both goal reminders and removed newsfeed helped participants stay on task and avoid distraction. However, goal reminders were often annoying, and removing the newsfeed made some fear missing out on information. Our findings point to future interventions such as controls for adjusting types and amount of available information, and flexible blocking which matches individual definitions of ‘distraction’.

Author Keywords

Facebook; problematic use; self-control; distraction; ICT non-use; addiction; focus; interruptions

CCS Concepts

•Human-centered computing → Empirical studies in HCI;

INTRODUCTION

Research on ‘Problematic Facebook Use’ (PFU) has investigated correlations between Facebook use and negative effects on outcomes such as level of academic achievement [35] and subjective wellbeing [58, 57]. A cross-cutting finding is that negative outcomes are associated with difficulty at exerting self-control over use, as well as specific use patterns including viewing friends’ wide-audience broadcasts rather than receiving targeted communication from strong ties [13, 58].

Much of this work has focused on self-control over Facebook use in student populations [2, 44, 46], with media multitasking research finding that students often give in to use which provides short-term ‘guilty pleasures’ over important, but aversive academic tasks [76, 88, 60]. In the present paper, we present a mixed-methods study exploring how two interventions to Facebook — goal reminders and removing the newsfeed — affect university students’ patterns of use and perceived control over Facebook use. To triangulate self-report with objective measurement, our study combined usage logging with fortnightly surveys and post-study interviews.

We found that both interventions helped participants stay on task and use Facebook more in line with their intentions. In terms of use patterns, goal reminders led to less scrolling, fewer and shorter visits, and less time on site, whereas removing the newsfeed led to less scrolling, shorter visits, and less content ‘liked’. However, goal reminders were often experienced as annoying, and removing the newsfeed made some participants fear missing out on information. After the study, participants suggested a range of design solutions to mitigate self-control struggles on Facebook, including controls for filtering or removing the newsfeed, reminders of time spent and of use goals, and removing features that drive engagement. As an exploratory study, this work should be followed by confirmatory studies to assess whether our findings replicate, and how they may generalise beyond a student population.

RELATED WORK

Struggles with Facebook use

Whereas many uses of Facebook offer important benefits, such as social support, rapid spread of information, or facilitation of real-world interactions [78], a substantial amount of research has focused on negative aspects [58]. For example, studies have reported correlations between patterns of Facebook use and lower academic achievement [77, 86], low self-esteem, depression and anxiety [51], feelings of isolation and loneliness [2], and general psychological distress [15]. Such ‘Problematic Facebook Use’ (PFU) has been studied under various names (including ‘Facebook dependence’ [87] and ‘Facebook addiction’ [5]), but a recent review summarised a common definition as ‘problematic behaviour characterised by addictive-like symptoms and/or self-regulation difficulties related to Facebook use leading to negative consequences in personal and social life’ [58].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI ’20, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-6708-0/20/04.

<http://dx.doi.org/10.1145/3313831.3376672>

5. Customisations and extensions

CHI 2020 Paper

CHI 2020, April 25–30, 2020, Honolulu, HI, USA

REFERENCES

- [1] Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless Computing: Designing Technologies to Subtly Influence Behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 719–730. DOI: <http://dx.doi.org/10.1145/2750858.2805843>
- [2] Sami Abdo Radman Al-Dubai, Kurubaran Ganasegeran, Mustafa Ahmed Mahdi Al-Shagga, Hematram Yadav, and John T. Arokiasamy. 2013. Adverse Health Effects and Unhealthy Behaviors among Medical Students Using Facebook. <https://www.hindawi.com/journals/tswj/2013/465161/>. (2013). DOI: <http://dx.doi.org/10.1155/2013/465161>
- [3] All Party Parliamentary Group on Social Media and Young People's Mental Health and Wellbeing. 2019. *#NewFilters to Manage the Impact of Social Media on Young People's Mental Health and Wellbeing*. Technical Report. UK Parliament.
- [4] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2019. *The Welfare Effects of Social Media*. Working Paper 25514. National Bureau of Economic Research. DOI: <http://dx.doi.org/10.3386/w25514>
- [5] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Staale Pallesen. 2012. Development of a Facebook Addiction Scale. *Psychological Reports* 110, 2 (apr 2012), 501–517. DOI: <http://dx.doi.org/10.2466/02.09.18.PR0.110.2.501-517>
- [6] Yummy Apps. 2019. Todobook. (May 2019).
- [7] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37, 2 (1982), 122–147. DOI: <http://dx.doi.org/10.1037/0003-066x.37.2.122>
- [8] Fanni Bányaí, Ágnes Zsila, Orsolya Király, Aniko Maraz, Zsuzsanna Elekes, Mark D. Griffiths, Cecilie Schou Andreassen, and Zsolt Demetrovics. 09-Jan-2017. Problematic Social Media Use: Results from a Large-Scale Nationally Representative Adolescent Sample. *PLOS ONE* 12, 1 (09-Jan-2017), e0169839. DOI: <http://dx.doi.org/10.1371/journal.pone.0169839>
- [9] Elliot T Berkman, Cendri A Hutcherson, Jordan L Livingston, Lauren E Kahn, and Michael Inzlicht. 2017. Self-Control as Value-Based Choice. *Current Directions in Psychological Science* 26, 5 (2017), 422–428. DOI: <http://dx.doi.org/10.1177/0963721417704394>
- [10] Walter R. Boot, Daniel J. Simons, Cary Stothart, and Cassie Stutts. 2013. The Pervasive Problem with Placebos in Psychology. *Perspectives on Psychological Science* 8, 4 (jul 2013), 445–454. DOI: <http://dx.doi.org/10.1177/1745691613491271>
- [11] Amara Brook. 2011. Ecological Footprint Feedback: Motivating or Discouraging? *Social Influence* 6, 2 (April 2011), 113–128. DOI: <http://dx.doi.org/10.1080/15534510.2011.566801>
- [12] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual Review of Economics* 2, 1 (Sept. 2010), 671–698. DOI: <http://dx.doi.org/10.1146/annurev.economics.102308.124324>
- [13] Moira Burke and Robert E. Kraut. 2016. The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication* 21, 4 (2016), 265–281. DOI: <http://dx.doi.org/10.1111/jcc4.12162>
- [14] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social Network Activity and Social Well-Being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1909–1912. DOI: <http://dx.doi.org/10.1145/1753326.1753613>
- [15] Wenhong Chen and Kye-Hyoung Lee. 2013. Sharing, Liking, Commenting, and Distressed? The Pathway between Facebook Interaction and Psychological Distress. *Cyberpsychology, Behavior and Social Networking* 16, 10 (oct 2013), 728–734. DOI: <http://dx.doi.org/10.1089/cyber.2012.0272>
- [16] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 199:1–199:13. DOI: <http://dx.doi.org/10.1145/3290605.3300429>
- [17] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. DOI: <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- [18] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1389–1397. DOI: <http://dx.doi.org/10.1145/2851581.2892410>
- [19] Helen Creswick, Liz Dowthwaite, Ansgar Koene, Elvira Perez Vallejos, Virginia Portillo, Monica Cano, and Christopher Woodard. 2019. "... They don't really listen to people". *Journal of Information, Communication and Ethics in Society* 17, 2 (May 2019), 167–182. DOI: <http://dx.doi.org/10.1108/jices-11-2018-0090>
- [20] Angela L. Duckworth, Katherine L. Milkman, and David Laibson. 2018. Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological Science in the Public Interest* 19, 3 (Dec. 2018), 102–129. DOI: <http://dx.doi.org/10.1177/1529100618821893>

5.8 Including another paper in your thesis - R Markdown child document

Sometimes you want to include another paper you are currently writing as a chapter in your thesis. Above 5.7, we described the simplest way to do this: include the other paper as a pdf. However, in some cases you instead want to include the R Markdown source from this paper, and have it compiled within your thesis. This is a little bit more tricky, because you need to keep careful track of your file paths, but it is possible by [including the paper as a child document](#). There are four main steps:

1. Include the paper as a child document
2. Make file paths compatible with knitting the article on its own, as well as when it's include in your thesis
3. Make header levels correct
4. Make figure widths correct

5.8.1 An example paper in another folder

Take this simple example (files for this are in [this GitHub repository](#)):

```
|--paper_to_include
|   |--my_paper.Rmd
|   |--data
|   |   |--cat_salt.csv
|   |--figures
|   |   |--cat.jpg
|
|--thesis
```

As the chart suggests, you have another folder, **paper_to_include/** living in the same containing folder as your thesis folder. In the **paper_to_include** folder, the file **my_paper.Rmd** is where you write the paper. In **my_paper.Rmd**,

5. Customisations and extensions

you read in a CSV file found in the subfolder **data/cats.csv**, and also an image from the subfolder **figures/cat.jpg**.

5.8.2 Step 1: Include paper as a child document

In your thesis folder, create an Rmd file for the chapter where you want to include another paper. Add one or more code chunks that include R Markdown files from that paper as child documents:

```
# Including an external chapter

```{r child = "../paper_to_include/my_paper.Rmd"}
```
```

5.8.3 Step 2: Make file paths compatible

Use [parameters](#) to adjust the file path of images based on values you set in the YAML header of an R Markdown file. In **my_paper.Rmd**, create a parameter called **other_path** and set it to an empty string:

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
---
```

In **my_paper.Rmd**, put this at the start of the filepath when you read in data or include images:

```
library(tidyverse)
library(knitr)

cat_data <- read_csv(str_c(params$other_path, "data/cats.csv"))
include_graphics(str_c(params$other_path, "figures/cat.jpg"))
```

5. Customisations and extensions

Finally, in your thesis folder's **index.Rmd** file, also create the parameter `other_path`. But here, set it to where the **paper_to_include/** folder is relative to your thesis folder:

```
params:
  other_path: "../paper_to_include/"
```

Note on HTML output

Note that if you want to host an HTML version on your thesis online, you will need to include graphics in the content that you host online - the internet obviously won't be able to see filepaths that are just referring to stuff in another folder on your computer!

5.8.4 Step 3: Make sure header levels are correct

Unless the paper you want to include is also written as a book, your header levels are probably going to be off. That is, the level 1 headers (`# Some header`) you use for main sections in the other paper turns into chapter titles when included in your thesis.

To avoid this, first *increment all heading levels by one in **paper_to_include/my_paper.Rmd*** (`# Some header` -> `## Some header`). Then in **paper_to_include/** create a [lua filter](#) that decrements header levels by one: Create a text file, save it as **reduce_header_level.lua**, and give it the content below.

```
function Header(el)
  if (el.level <= 1) then
    error("I don't know how to decrease the level of h1")
  end
  el.level = el.level - 1
  return el
end
```

In the YAML header of **paper_to_include/my_paper.Rmd**, use this filter:

5. Customisations and extensions

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
output:
  pdf_document:
    pandoc_args: ["--lua-filter=reduce_header_level.lua"]
---
```

Now, your header levels will be correct both when you knit the paper on its own and when its included in your thesis.

5.8.5 Step 4. Make sure figure widths are correct

It might be that your figure widths when knitting your paper on its own, and when including it in your thesis, need to be different. You can again use parameters to set figure widths.

Imagine you want figure width to be 80% of the page width when knitting your paper on its own, but 100% in your thesis. In `paper_to_include/my_paper.Rmd`, first add a parameter we could call `out_width` and set it to the string “80%”:

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
  out_width: "80%"
output:
  pdf_document:
    pandoc_args: ["--lua-filter=reduce_header_level.lua"]
---
```

Then, make sure use that parameter to set the output width when you include figures in `paper_to_include/my_paper.Rmd`:

5. Customisations and extensions

```
```{r, out.width=params$out_width, fig.cap="A very funny cat"}  
include_graphics(str_c(params$other_path, "figures/cat.jpg"))
```
```

Finally, create the parameter `out_width` in your thesis' **index.Rmd** file:

```
params:  
  other_path: "../paper_to_include/"  
  out_width: "80%"
```

Now, the output width of your figure will be 80% when knitting your paper on its own, and 100% when knitting it as child document of your thesis.

5.9 Customizing referencing

5.9.1 Using a .csl file with pandoc instead of biblatex

The `oxforddown` package uses biblatex in LaTeX for referencing. It is also possible to use pandoc for referencing by providing a .csl file in the YAML header of **index.Rmd** (likely requiring commenting out the biblatex code in **templates/template.tex**). This may be helpful for those who have a .csl file describing the referencing format for a particular journal. However, note that this approach does not support chapter bibliographies (see Section 5.9.2).

```
csl: ecology.csl
```

5.9.2 Customizing biblatex and adding chapter bibliographies

This section provides one example of customizing biblatex. Much of this code was combined from searches on Stack Exchange and other sources (e.g. [here](#)).

In **templates/template.tex**, one can replace the existing biblatex calls with the following to achieve referencing that looks like this:

(Charmantier and Gienapp 2014)

5. Customisations and extensions

Charmantier, A. and P. Gienapp (2014). Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications* 7(1):15–28. doi: 10.1111/eva.12126.

```
\usepackage[backend=biber,  
  bibencoding=utf8,  
  refsection=chapter, % referencing by chapter  
  style=authoryear,  
  firstinits=true,  
  isbn=false,  
  doi=true,  
  url=false,  
  eprint=false,  
  related=false,  
  dashed=false,  
  clearlang=true,  
  maxcitenames=2,  
  mincitenames=1,  
  maxbibnames=10,  
  abbreviate=false,  
  minbibnames=3,  
  uniquelist=minyear,  
  sortcites=true,  
  date=year  
{biblatex}  
\AtEveryBibitem{%  
  \clearlist{language}%  
  \clearfield{note}  
}  
  
\DeclareFieldFormat{titlecase}{\MakeTitleCase{#1}}
```

5. Customisations and extensions

```

\newrobustcmd{\MakeTitleCase}[1]{%
  \ifthenelse{\ifcurrentfield{booktitle}\OR\ifcurrentfield{booksubtitle}%
    \OR\ifcurrentfield{maintitle}\OR\ifcurrentfield{mainsubtitle}%
    \OR\ifcurrentfield{journaltitle}\OR\ifcurrentfield{journalsubtitle}%
    \OR\ifcurrentfield{issuetitle}\OR\ifcurrentfield{issuesubtitle}%
    \OR\ifentrytype{book}\OR\ifentrytype{mvbook}\OR\ifentrytype{bookinbook}%
    \OR\ifentrytype{booklet}\OR\ifentrytype{suppbook}%
    \OR\ifentrytype{collection}\OR\ifentrytype{mvcollection}%
    \OR\ifentrytype{suppcollection}\OR\ifentrytype{manual}%
    \OR\ifentrytype{periodical}\OR\ifentrytype{suppperiodical}%
    \OR\ifentrytype{proceedings}\OR\ifentrytype{mvproceedings}%
    \OR\ifentrytype{reference}\OR\ifentrytype{mvreference}%
    \OR\ifentrytype{report}\OR\ifentrytype{thesis}}
    {#1}
    {\MakeSentenceCase{#1}}}

% \renewbibmacro{in:}{}
% suppress "in" for articles
%
\renewbibmacro{in:}{%
  \ifentrytype{article}{}{\printtext{\bibstring{in}\intitlepunct}}
  %-- no "quotes" around titles of chapters/article titles
\DeclareFieldFormat{article, inbook, incollection, inproceedings, misc, thesis, unp
{title}{#1}
%-- no punctuation after volume
\DeclareFieldFormat{article]
{volume}{{#1}}
%-- puts number/issue between brackets
\DeclareFieldFormat{article, inbook, incollection, inproceedings, misc, thesis, unp

```

5. Customisations and extensions

```
{number}}{\mkbibparens{#1}}
%-- and then for articles directly the pages w/o any "pages" or "pp."
\DeclareFieldFormat[article]
{pages}{#1}
%-- for some types replace "pages" by "p."
\DeclareFieldFormat[inproceedings, incollection, inbook]
{pages}{p. #1}
%-- format 16(4):224--225 for articles
\renewbibmacro*{volume+number+eid}{
  \printfield{volume}%
  \printfield{number}%
  \printunit{\addcolon}
}
```

If you would like chapter bibliographies, in addition insert the following code at the end of each chapter, and comment out the entire REFERENCES section at the end of template.tex.

```
\printbibliography[segment=\therefsection,heading=subbibliography]
```

5.10 Customizing the page headers and footers (PDF)

This can now be done directly in **index.Rmd**'s YAML header. If you are a LaTeX expert and need further customisation that what's currently provided, you can tweak the relevant sections of **templates/template.tex** - the relevant code is beneath the line that begins `\usepackage{fancyhdr}`.

5.11 Diving in to the OxThesis LaTeX template (PDF)

For LaTeX minded people, you can read through `templates/template.tex` to see which additional customisation options are available as well as `templates/ociamthesis.cls` which supplies the base class. For example, `template.tex` provides an option for master’s degree submissions, which changes identifying information to candidate number and includes a word count. At the time of writing, you must set this directly in `template.tex` rather than from the YAML header in `index.Rmd`.

5.12 Customising to a different university

5.12.1 The minimal route

If the front matter in the OxThesis LaTeX template is suitable to your university, customising `oxforddown` to your needs could be as simple as putting the name of your institution and the path to your university’s logo in `index.Rmd`:

```
university: University of You
university-logo: figures/your-logo-here.pdf
```

5.12.2 Replacing the entire title page with your required content

If you have a `.tex` file with some required front matter from your university that you want to replace the OxThesis template’s title page altogether, you can provide a filepath to this file in `index.Rmd`. `oxforddown`’s sample content includes an example of this — if you use the YAML below, your front matter will look like this:

```
alternative-title-page: front-and-back-matter/alt-title-page-example.tex
```

5. Customisations and extensions

| | | |
|--|---|--|
| <p>Title of your Thesis</p> <p>John Doe</p> | <p>Thesis committee</p> <p>Promotor:
Prof.dr. J. Smith
Professor of Geo-information Science and Remote Sensing
Wageningen University</p> <p>Co-promotors:
Dr. Name of co-promotor
Assistant Professor, Laboratory of Geo-information Science and Remote Sensing
Wageningen University</p> <p>Other members:
Prof.dr. Jury member 1, Wageningen University
Prof.dr. Jury member 2, Affiliation
Prof.dr. Jury member 3, Affiliation
Prof.dr. Jury member 4, Affiliation</p> <p>This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PERC)</p> | <p>Title of your thesis</p> <p>John Doe</p> <p>Thesis
submitted in fulfillment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Date of your defense
at 4 p.m. in the Aula</p> |
| <p>John Doe
Title of your thesis
77 pages
PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summary in English
ISBN XXX-YYY</p> | <p>For Yihai Xie</p> | <p>Acknowledgements</p> <p>This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our thanks to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.</p> <p>We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (http://pandoc.org) which converts Markdown to a large number of output formats, and to Yihai Xie for creating knitr which introduced R Markdown as a way of embedding code in Markdown documents, and bookdown which added tools for technical and longer-form writing.</p> <p>Special thanks to Chetani Ismay, who created the <i>thesisdown</i> package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John MacFarlane, whose adoption of Stan Evans' adaptation of Keith Gilmer's original <i>manila</i> template for writing an Oxford University DPhil thesis in L^AT_EX provided the template that I in turn adapted for R Markdown.</p> <p>Finally, please thank to JJ Allaire, the founder and CEO of RStudio, and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.</p> <p>Ulrik Lenga
Linacre College, Oxford
2 December 2018</p> |

6

Troubleshooting

This chapter describes common errors you may run into, and how to fix them.

6.1 Error: Failed to build the bibliography via biber

This can happen if you've had a failed build, perhaps in relation to RStudio shutting down abruptly.

Try doing this:

1. type `make clean-knits` in the terminal tab (or run `file.remove(list.files(pattern = "*.log|mtc|maf|aux|bbl|blg|xml"))` in the R console) to clean up files generated by LaTeX during a build
2. restart your computer

If this does not solve the problem, try using the [natbib](#) LaTeX package instead of [biblatex](#) for handling references. To do this, go to **index.Rmd** and

1. set `use-biblatex: false` and `use-natbib: true`
2. set `citation_package: natbib` under

6. *Troubleshooting*

```
output:
```

```
  bookdown::pdf_book:
```

```
    citation_package: natbib
```

*Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken.
All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.*
— Johann Wolfgang von Goethe (von Goethe 1829)

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

This paragraph, by contrast, *will* be indented as it should because it is not the first one after the 'More info' heading. All hail LaTeX. (If you're reading the HTML version, you won't see any indentation - have a look at the PDF version to understand what in the earth this section is babbling on about).

Appendices



The First Appendix

This first appendix includes an R chunk that was hidden in the document (using `echo = FALSE`) to help with readability:

In 02-rmd-basics-code.Rmd

And here's another one from the same chapter, i.e. Chapter ??:

B

The Second Appendix, for Fun

Works Cited

- Darwin, Charles (1859). *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Von Goethe, Johann Wolfgang (1829). *Wilhelm Meisters Wanderjahre oder die Entsagenden*. de. Cotta.
- Shea, Nicholas et al. (2014). “Supra-personal cognitive control and metacognition”. In: *Trends in Cognitive Sciences* 18.4, pp. 186–193. DOI: [10.1016/j.tics.2014.01.006](https://doi.org/10.1016/j.tics.2014.01.006). URL: <http://dx.doi.org/10.1016/j.tics.2014.01.006>.