



My Title

My subtitle

Petra Muster

Student Number: 00-123-456

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
MASTER OF MARKETING MANAGEMENT

Supervisor: Professor Dr. Arnold Arnoldson
Second Reader: Dr. Brenda Barber

Tilburg University
Department of Marketing
School of Economics and Management
Tilburg, The Netherlands

30. Juni 2018

Management Summary

There is a max. 1 page management summary.

Acknowledgements

I want to thank a few people.

Table of Contents

Guide	1
Chapter 1: R Markdown Basics	3
1.1 Lists	3
1.2 Line breaks	4
1.3 R chunks	4
1.4 Inline code	5
1.5 Including plots	5
1.6 Loading and exploring data	6
1.7 Additional resources	10
Chapter 2: Mathematics and Science	11
2.1 Math	11
Chapter 3: Graphics, References, and Labels	13
3.1 Figures	13
3.2 Footnotes and Endnotes	17
3.3 Bibliographies	17
3.4 Anything else?	19
Conclusion	20
Appendix A: The First Appendix	21
Appendix B: The Second Appendix, for Fun	22
References	23

Guide

Welcome to the *R Markdown* thesis template. This template is based on (and in many places copied directly from) university LaTeX templates from a variety of sources, but hopefully it will provide a nicer interface for those that have never used TeX or LaTeX before.

Using *R Markdown* will also allow you to easily keep track of your analyses in **R** chunks of code, with the resulting plots and output included as well.

The hope is this *R Markdown* template gets you in the habit of doing reproducible research, which benefits you long-term as a researcher, but also will greatly help anyone that is trying to reproduce or build onto your results down the road.

Hopefully, you won't have much of a learning period to go through and you will reap the benefits of a nicely formatted thesis.

The use of LaTeX in combination with *Markdown* is more consistent than the output of a word processor, much less prone to corruption or crashing, and the resulting file is smaller than a Word file. While you may have never had problems using Word in the past, your thesis is likely going to be about twice as large and complex as anything you've written before, taxing Word's capabilities.

After working with *Markdown* and **R** together for a few weeks, we are confident this will be your reporting style of choice going forward.

Why use it?

R Markdown creates a simple and straightforward way to interface with the beauty of LaTeX. Packages have been written in **R** to work directly with LaTeX to produce nicely formatting tables and paragraphs. In addition to creating a user friendly interface to LaTeX, *R Markdown* also allows you to read in your data, to analyze it and to visualize it using **R** functions, and also to provide the documentation and commentary on the results of your project.

Further, it allows for **R** results to be passed inline to the commentary of your results. You'll see more on this later.

Who should use it?

Anyone who needs to use data analysis, math, tables, a lot of figures, complex cross-references, or who just cares about the final appearance of their document should use *R Markdown*. Of particular use should be anyone in the sciences, but the user-friendly nature of *Markdown* and its ability to keep track of and easily include figures, automatically generate a table of contents, index, references, table of figures, etc. should make it of great benefit to nearly anyone writing a thesis project.

For additional help with bookdown

Please visit the free online bookdown reference guide.

Chapter 1

R Markdown Basics

Here is a brief introduction into using *R Markdown*. *Markdown* is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. *R Markdown* provides the flexibility of *Markdown* with the implementation of **R** input and output.

For more details on using *R Markdown* see <https://rmarkdown.rstudio.com>.

Be careful with your spacing in *Markdown* documents.

While whitespace largely is ignored, it does at times give *Markdown* signals as to how to proceed.

As a habit, try to keep everything left aligned whenever possible, especially as you type a new paragraph.

In other words, there is no need to indent basic text in the Rmd document (in fact, it might cause your text to do funny things if you do). Starting each sentence on a new line is not required, but I have found it makes things easier to version control.

1.1 Lists

It's easy to create a list. It can be unordered like

- Item 1
- Item 2

or it can be ordered like

1. Item 1
2. Item 2

Notice that I intentionally mislabeled Item 2 as number 4. *Markdown* automatically figures this out!

You can put any numbers in the list and it will create the list. Check it out below.

To create a sublist, just indent the values a bit (at least four spaces or a tab). (Here's one case where indentation is key!)

1. Item 1
2. Item 2
3. Item 3
 - Item 3a
 - Item 3b

1.2 Line breaks

Make sure to add white space between lines if you'd like to start a new paragraph. Look at what happens below in the outputted document if you don't:

Here is the first sentence.

Here is another sentence.

Here is the last sentence to end the paragraph. This should be a new paragraph.

Now for the correct way:

Here is the first sentence.

Here is another sentence.

Here is the last sentence to end the paragraph.

This should be a new paragraph.

1.3 R chunks

When you click the **Knit** button above a document will be generated that includes both content as well as the output of any embedded **R** code chunks within the document. You can embed an **R** code chunk like this (`cars` is a built-in **R** dataset):

```
summary(cars)
```

	speed		dist
Min.	: 4.0	Min.	: 2.00
1st Qu.	:12.0	1st Qu.	: 26.00
Median	:15.0	Median	: 36.00
Mean	:15.4	Mean	: 42.98
3rd Qu.	:19.0	3rd Qu.	: 56.00
Max.	:25.0	Max.	:120.00

1.4 Inline code

If you'd like to put the results of your analysis directly into your discussion, add inline code like this:

The `cos` of 2π is 1.

Another example would be the direct calculation of the standard deviation:

The standard deviation of `speed` in `cars` is 5.2876444.

One last neat feature is the use of the `ifelse` conditional statement which can be used to output text depending on the result of an **R** calculation:

The standard deviation is less than 6.

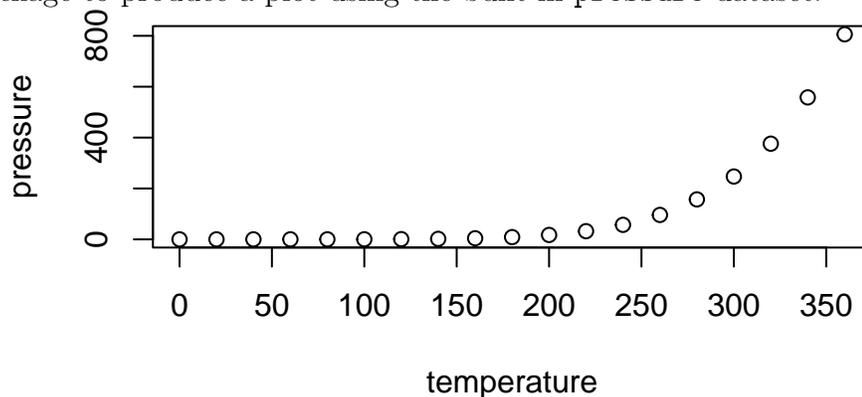
Note the use of `>` here, which signifies a quotation environment that will be indented.

As you see with `\pi` above, mathematics can be added by surrounding the mathematical text with dollar signs.

More examples of this are in Mathematics and Science if you uncomment the code in Math.

1.5 Including plots

You can also embed plots. For example, here is a way to use the base **R** graphics package to produce a plot using the built-in `pressure` dataset:



Note that the `echo=FALSE` parameter was added to the code chunk to prevent printing of the **R** code that generated the plot. There are plenty of other ways to add chunk options (like `fig.height` and `fig.width` in the chunk above).

More information is available at <https://yihui.org/knitr/options/>.

Another useful chunk option is the setting of `cache=TRUE` as you see here. If document rendering becomes time consuming due to long computations or plots that are expensive to generate you can use knitr caching to improve performance. Later in this file, you'll see a way to reference plots created in **R** or external figures.

1.6 Loading and exploring data

Included in this template is a file called `flights.csv`.

This file includes a subset of the larger dataset of information about all flights that departed from Seattle and Portland in 2014. More information about this dataset and its **R** package is available at <https://github.com/ismayc/pnwflights14>. This subset includes only Portland flights and only rows that were complete with no missing values. Merges were also done with the `airports` and `airlines` data sets in the `pnwflights14` package to get more descriptive airport and airline names.

We can load in this data set using the following commands:

```
# flights.csv is in the data directory
flights_path <- here::here("data", "flights.csv")
# string columns will be read in as strings and not factors now
flights <- read.csv(flights_path, stringsAsFactors = FALSE)
```

The data is now stored in the data frame called `flights` in **R**. To get a better feel for the variables included in this dataset we can use a variety of functions. Here we can see the dimensions (rows by columns) and also the names of the columns.

```
dim(flights)
```

```
[1] 12649    16
```

```
names(flights)
```

```
[1] "month"      "day"        "dep_time"   "dep_delay"
[5] "arr_time"   "arr_delay"  "carrier"    "tailnum"
[9] "flight"     "dest"       "air_time"   "distance"
[13] "hour"       "minute"     "carrier_name" "dest_name"
```

Another good idea is to take a look at the dataset in table form. With this dataset having more than 20,000 rows, we won't explicitly show the results of the command here. I recommend you enter the command into the Console *after* you have run the **R** chunks above to load the data into **R**.

```
View(flights)
```

While not required, it is highly recommended you use the `dplyr` package to manipulate and summarize your data set as needed.

It uses a syntax that is easy to understand using chaining operations.

Below I've created a few examples of using `dplyr` to get information about the Portland flights in 2014.

You will also see the use of the `ggplot2` package, which produces beautiful, high-quality academic visuals.

We begin by checking to ensure that needed packages are installed and then we load them into our current working environment:

```
# List of packages required for this analysis
pkg <- c("dplyr", "ggplot2", "knitr", "bookdown")
# Check if packages are not installed and assign the
# names of the packages not installed to the variable new.pkg
new.pkg <- pkg[!(pkg %in% installed.packages())]
# If there are any packages in the list that aren't installed,
# install them
if (length(new.pkg)) {
  install.packages(new.pkg, repos = "https://cran.rstudio.com")
}
# Load packages
library(tisemdown)
library(dplyr)
library(ggplot2)
library(knitr)
```

The example we show here does the following:

- Selects only the `carrier_name` and `arr_delay` from the `flights` dataset and then assigns this subset to a new variable called `flights2`.
- Using `flights2`, we determine the largest arrival delay for each of the carriers.

```
flights2 <- flights %>%
  select(carrier_name, arr_delay)
max_delays <- flights2 %>%
  group_by(carrier_name) %>%
  summarize(max_arr_delay = max(arr_delay, na.rm = TRUE))
```

A useful function in the `knitr` package for making nice tables in *R Markdown* is called `kable`.

It is much easier to use than manually entering values into a table by copying and pasting values into Excel or LaTeX.

This again goes to show how nice reproducible documents can be! (Note the use of `results="asis"`, which will produce the table instead of the code to create the table.)

The `caption.short` argument is used to include a shorter title to appear in the List of Tables.

```
kable(max_delays,
  col.names = c("Airline", "Max Arrival Delay"),
  caption = "Maximum Delays by Airline",
  caption.short = "Max Delays by Airline",
  longtable = TRUE,
  booktabs = TRUE
)
```

Table 1.1: Maximum Delays by Airline

Airline	Max Arrival Delay
Alaska Airlines Inc.	338
American Airlines Inc.	1539
Delta Air Lines Inc.	371
Frontier Airlines Inc.	166

Hawaiian Airlines Inc.	116
JetBlue Airways	256
SkyWest Airlines Inc.	321
Southwest Airlines Co.	315
United Air Lines Inc.	319
US Airways Inc.	347
Virgin America	366

The last two options make the table a little easier-to-read.

We can further look into the properties of the largest value here for American Airlines Inc.

To do so, we can isolate the row corresponding to the arrival delay of 1539 minutes for American in our original `flights` dataset.

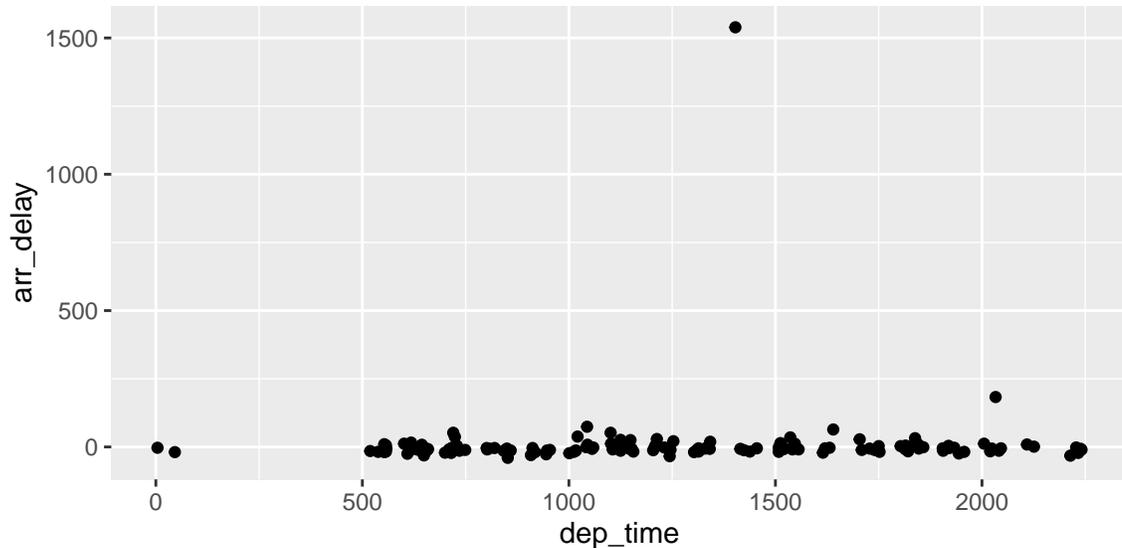
```
flights %>%
  filter(
    arr_delay == 1539,
    carrier_name == "American Airlines Inc."
  ) %>%
  select(-c(
    month, day, carrier, dest_name, hour,
    minute, carrier_name, arr_delay
  ))
```

```
dep_time dep_delay arr_time tailnum flight dest air_time distance
1      1403      1553      1934 N595AA  1568 DFW      182      1616
```

We see that the flight occurred on March 3rd and departed a little after 2 PM on its way to Dallas/Fort Worth.

Lastly, we show how we can visualize the arrival delay of all departing flights from Portland on March 3rd against time of departure.

```
flights %>%
  filter(month == 3, day == 3) %>%
  ggplot(aes(x = dep_time, y = arr_delay)) +
  geom_point()
```



1.7 Additional resources

- *Markdown Cheatsheet* - <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>
- *R Markdown*
 - Reference Guide - <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>
 - Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf>
- *RStudio IDE*
 - Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf>
 - Official website - <https://rstudio.com/products/rstudio/>
- Introduction to dplyr - <https://cran.rstudio.com/web/packages/dplyr/vignettes/dplyr.html>
- ggplot2
 - Documentation - <https://ggplot2.tidyverse.org/>
 - Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf>

Chapter 2

Mathematics and Science

2.1 Math

\TeX is the best way to typeset mathematics. Donald Knuth designed \TeX when he got frustrated at how long it was taking the typesetters to finish his book, which contained a lot of mathematics.

One nice feature of *R Markdown* is its ability to read LaTeX code directly.

If you are doing a thesis that will involve lots of math, you will want to read the following section which has been commented out. If you're not going to use math, skip over or delete this next commented section.

Some example equations so you get a sense of what to do:

$$\sum_{j=1}^n (\delta\theta_j)^2 \leq \frac{\beta_i^2}{\delta_i^2 + \rho_i^2} \left[2\rho_i^2 + \frac{\delta_i^2 \beta_i^2}{\delta_i^2 + \rho_i^2} \right] \equiv \omega_i^2$$

From Informational Dynamics, we have the following (Dave Braden):

After n such encounters the posterior density for θ is

$$\pi(\theta|X_1 < y_1, \dots, X_n < y_n) \propto \pi(\theta) \prod_{i=1}^n \int_{-\infty}^{y_i} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) dx$$

Another equation:

$$\det \begin{vmatrix} c_0 & c_1 & c_2 & \dots & c_n \\ c_1 & c_2 & c_3 & \dots & c_{n+1} \\ c_2 & c_3 & c_4 & \dots & c_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ c_n & c_{n+1} & c_{n+2} & \dots & c_{2n} \end{vmatrix} > 0$$

Lapidus and Pindar, Numerical Solution of Partial Differential Equations in Science and Engineering. Page 54

$$\int_t \left\{ \sum_{j=1}^3 T_j \left(\frac{d\phi_j}{dt} + k\phi_j \right) - kT_e \right\} w_i(t) dt = 0, \quad i = 1, 2, 3.$$

L&P Galerkin method weighting functions. Page 55

$$\sum_{j=1}^3 T_j \int_0^1 \left\{ \frac{d\phi_j}{dt} + k\phi_j \right\} \phi_i dt = \int_0^1 k T_e \phi_i dt, \quad i = 1, 2, 3$$

Another L&P (p145)

$$\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 f(\xi, \eta, \zeta) = \sum_{k=1}^n \sum_{j=1}^n \sum_{i=1}^n w_i w_j w_k f(\xi, \eta, \zeta).$$

Another L&P (p126)

$$\int_{A_e} (\cdot) dx dy = \int_{-1}^1 \int_{-1}^1 (\cdot) \det[J] d\xi d\eta.$$

Chapter 3

Graphics, References, and Labels

3.1 Figures

If your thesis has a lot of figures, *R Markdown* might behave better for you than that other word processor.

One perk is that it will automatically number the figures accordingly in each chapter. You'll also be able to create a label for each figure, add a caption, and then reference the figure in a way similar to what we saw with tables earlier.

If you label your figures, you can move the figures around and *R Markdown* will automatically adjust the numbering for you. No need for you to remember!

So that you don't have to get too far into LaTeX to do this, a couple **R** functions have been created for you to assist. You'll see their use below.

In the **R** chunk below, we will load in a picture stored as `tilburg_color.png` in the figures subdirectory.

We then give it the caption of "Tilburg logo," the label of "tilburglogo," and specify that this is a figure.

Make note of the different **R** chunk options that are given in the R Markdown file (not shown in the knitted document).

```
include_graphics(path = "figure/tilburg_color.png")
```



Figure 3.1: Tilburg logo

Here is a reference to the Tilburg logo: Figure 3.1.

Note the use of the `fig:` code here. By naming the **R** chunk that contains the figure, we can then reference that figure later as done in the first sentence here.

We can also specify the caption for the figure via the R chunk option `fig.cap`.

Below we will investigate how to save the output of an **R** plot and label it in a way similar to that done above.

Recall the `flights` dataset from Chapter 1. (Note that we've shown a different way to reference a section or chapter here.)

We will next explore a bar graph with the mean flight departure delays by airline from Portland for 2014.

```
mean_delay_by_carrier <- flights %>%  
  group_by(carrier) %>%  
  summarize(mean_dep_delay = mean(dep_delay))  
ggplot(mean_delay_by_carrier, aes(x = carrier, y = mean_dep_delay)) +  
  geom_bar(position = "identity", stat = "identity", fill = "red")
```

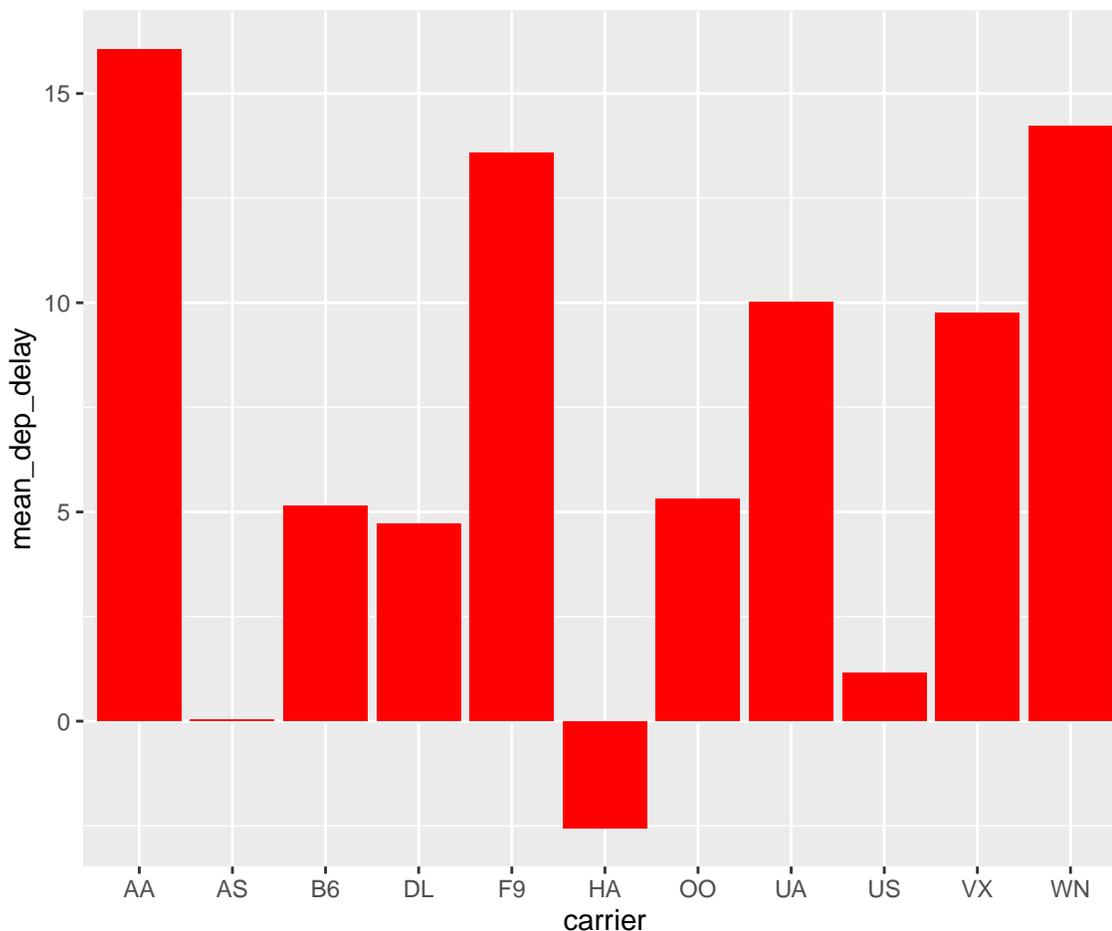


Figure 3.2: Mean Delays by Airline

Here is a reference to this image: Figure 3.2.

A table linking these carrier codes to airline names is available at <https://github.com/ismayc/pnwflights14/blob/master/data/airlines.csv>.

Next, we will explore the use of the `out.extra` chunk option, which can be used to shrink or expand an image loaded from a file by specifying `"scale= "`. Here we use the mathematical graph stored in the “subdivision.pdf” file.

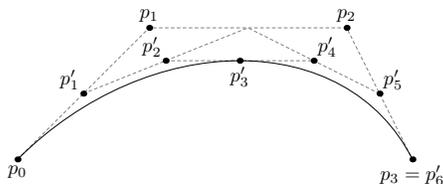


Figure 3.3: Subdiv. graph

Here is a reference to this image: Figure 3.3.
Note that `echo=FALSE` is specified so that the **R** code is hidden in the document.

More Figure Stuff

Lastly, we will explore how to rotate and enlarge figures using the `out.extra` chunk option.

(Currently this only works in the PDF version of the book.)

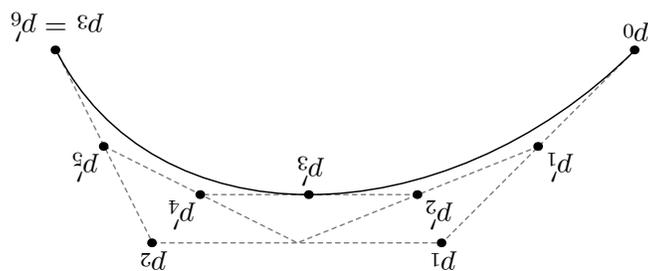


Figure 3.4: A Larger Figure, Flipped Upside Down

As another example, here is a reference: Figure 3.4.

3.2 Footnotes and Endnotes

You might want to footnote something.¹ The footnote will be in a smaller font and placed appropriately. Endnotes work in much the same way.

3.3 Bibliographies

Of course you will need to cite things, and you will probably accumulate an armful of sources. There are a variety of tools available for creating a bibliography database

¹footnote text

(stored with the .bib extension).

In addition to BibTeX suggested below, you may want to consider using the free and easy-to-use tool called Zotero.

R Markdown uses *pandoc* (<https://pandoc.org/>) to build its bibliographies. One nice caveat of this is that you won't have to do a second compile to load in references as standard LaTeX requires. To cite references in your thesis (after creating your bibliography database), place the reference name inside square brackets and precede it by the "at" symbol.

For example, here's a reference to a book about worrying: (Molina & Borkovec, 1994). This `Molina1994` entry appears in a file called `thesis.bib` in the `bib` folder. This bibliography database file was created by a program called BibTeX. You can call this file something else if you like (look at the YAML header in the main `.Rmd` file) and, by default, is to placed in the `bib` folder.

If you look at the YAML header at the top of the main `.Rmd` file you can see that we can specify the style of the bibliography by referencing the appropriate `csl` file. You can download a variety of different style files at <https://www.zotero.org/styles>.

Make sure to download the file into the `csl` folder.

Tips for Bibliographies

- Like with thesis formatting, the sooner you start compiling your bibliography for something as large as thesis, the better. Typing in source after source is mind-numbing enough; do you really want to do it for hours on end in late April? Think of it as procrastination.
- The cite key (a citation's label) needs to be unique from the other entries.
- When you have more than one author or editor, you need to separate each author's name by the word "and" e.g. `Author = {Noble, Sam and Youngberg, Jessica},.`
- Bibliographies made using BibTeX (whether manually or using a manager) accept LaTeX markup, so you can italicize and add symbols as necessary.
- To force capitalization in an article title or where all lowercase is generally used, bracket the capital letter in curly braces.

3.4 Anything else?

If you'd like to see examples of other things in this template, please contact the Lachlan Deer (email lachlan.deer@gmail.com) We love to see people using *R Markdown* for their theses, and are happy to help.

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.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readability and/or setup.

In the main Rmd file

In Chapter 3:

```
library(dplyr)
library(ggplot2)
library(knitr)
flights <- read.csv("data/flights.csv", stringsAsFactors = FALSE)
```

Appendix B

The Second Appendix, for Fun

References

- Angel, E. (2000). *Interactive computer graphics : A top-down approach with OpenGL*. Boston, MA: Addison Wesley Longman.
- Angel, E. (2001a). *Batch-file computer graphics : A bottom-up approach with Quick-Time*. Boston, MA: Wesley Addison Longman.
- Angel, E. (2001b). *Test second book by angel*. Boston, MA: Wesley Addison Longman.
- Molina, S. T., & Borkovec, T. D. (1994). The Penn State worry questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265–283). New York: Wiley.