Wasian gone Bayesian

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Josh Yamamoto

May 2023

Approved for the Division
(Mathematics & Statistics)

_____

Leonard Wainstein

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

List of Abbreviations

ABC   American Broadcasting Company
CBS   Colombia Broadcasting System
CUS   Computer User Services
NBC   National Broadcasting Company
PBS   Public Broadcasting System

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

Introduction

Welcome to the *R Markdown* thesis template. This template is based on (and in many places copied directly from) the Reed College LaTeX template, 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

Introduction

In this thesis, I will concerned with modeling a response variable from various explanatory variables in data that exhibits two distinctive features: (i) the response variable that is "zero-inflated", and (ii) the data is "clustered".

1.1 Zero-Inflated Data

As the name suggests, data are canonically classified as being zero-inflated when they contain a significant proportion of zeroes. While it's hardly ever very productive to spell out a definition for a phrase that is its own definition, I do so here to emphasize the fact that to call data zero-inflated is to only say something very broad about how that data is distributed. There is no commonly accepted cutoff for at what proportion of zeros our data deserves the label zero-inflation, and there is no restriction on the distribution of the non-zero data. While the work done in this thesis concerns zero-inflated data with no constraint on the level of "zeroness", I do require that the non-zero data is positive and continuously distributed. For example, the response variable might look like this:
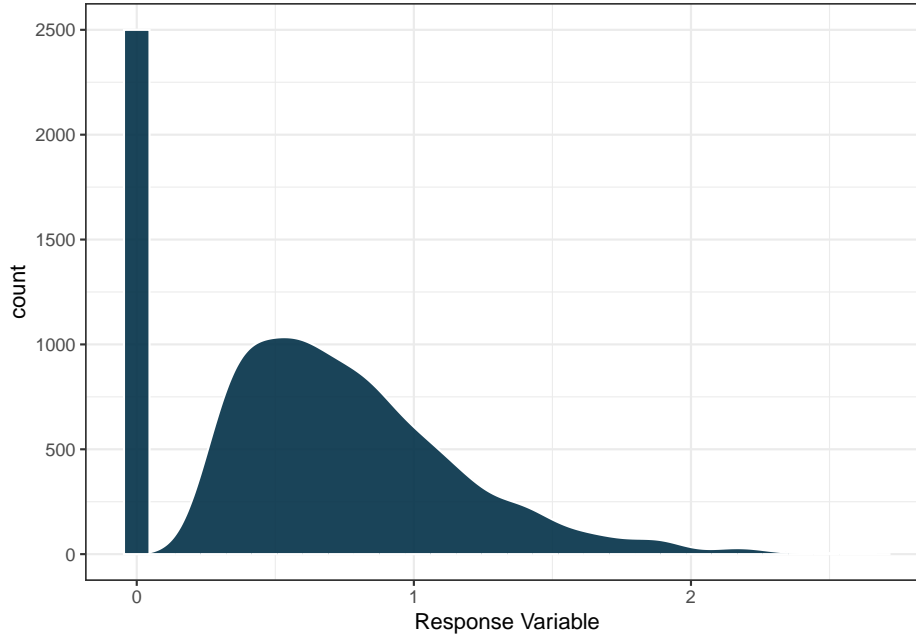
And in reality, this form of zero-inflated data is one that we see quite often in the real world. Importantly, an abundance of zeros in a measured variable might come about for a variety of different reasons. Sometimes it could be a characteristic of the data itself, for example if we collected data on the total weight of fish caught at a lake by individuals on a given day, we likely would see a lot of individuals who caught zero fish leading to a significant portion of zeros in our data. But other times it could be a characteristic of the data collection process itself, for example a measurement error or a sampling error could cause data to be zero-inflated as well.

Importantly, as I am working in a modeling setting, *when I say that data is zero-inflated I mean that the response variable is zero-inflated.* Although I will simply refer to my data as being zero-inflated and my models as being suited for zero-inflated data for the duration of this thesis, this is simply a matter of convenience and not a statement that the methods work for any situation in which data can be considered zero-inflated. Put simply, *I explore, present, and evaluate a model that is suited for a response variable which has a significant portion of zeroes, with the non-zero portion of that variable belonging to a positive continuous distribution.*

1.2 Clustered Data

Furthermore, we will be operating in a setting where the data is not only zero-inflated, but where it also exhibits a clustered structure. Again, the notion of data being clustered is a very non precise one. For this thesis we will not put a very strong

restriction on what this looks like. Our data will be clustered in the sense that there is meaningful grouping in the data structure that makes data points within the same cluster more alike on average than points across clusters.

For example, going back to the fishing in a lake example. If we looked at data on the weight of each individual fish caught, we would imagine that fish of the same species would generally be more similar in weight than two fish from different species.

1.3   Forestry Setting

One specific setting that exhibits both of these data features is data on the United States' forests. In particular, I will focus on forestry data collected by the Forestry Inventory & Analysis Program (FIA) of the U.S. Government. The FIA monitors the nation's forests by collecting data on, and providing estimates for, a wide array of forest attributes. Not only is this work vitally important, but it's essential that it be done accurately and efficiently: "The FIA is responsible for reporting on dozens, if not hundreds, of forest attributes relating to merchantable timber and other wood products, fuels and potential fire hazard, condition of wildlife habitats, risk associated with fire, insects or disease, biomass, carbon storage, forest health, and other general characteristics of forest ecosystems."[1].

These sampled locations are referred to as plot-level data and the FIA sends a crew out to physically measure a wealth of forest attributes at that location. As you might expect, not only is this method extremely time intensive, but it is also very expensive. The vastness of the nation's forests in tandem with the resources needed to collect plot-level data, make it impossible to collect census level data on forest metrics. Thus, the need for additional data sources as well as statistical models are vital to the work that the FIA does. The main secondary data source that the FIA employees is remote sensed data. The remote sensed data typically includes climate metrics (e.g. temperature and precipitation), geomorphological measures (e.g. elevation and eastness), as well as metrics like tree canopy cover which can be measured from a satellite.

While the main use of the additional remote sensed data sources are to increase the accuracy of the estimators that the FIA builds, they are also used to make rational decisions about the aforementioned plot-level data collection. Before sending a crew out to a given sampled location, the FIA will first look at the remote sensed data for that location. If that location happens to be in a place where there is clearly

---

[1]McConville, Moisen, & Frescino (2020)

no forest, for example in the middle of a parking lot, the FIA will not send a crew out and instead will mark all forest attributes for that location as being zero. As you might imagine, this happens quite a bit, and so an interesting characteristic of many forest attribute variables collected by the FIA is that they are zero-inflated. Importantly, this is an example of where the data is zero-inflated because of the data collection process.

Importantly, the FIA groups the continental U.S. into smaller domains called Eco-Subsections. These Eco-Subsections are drawn with the goal of maintaining internal ecologically homogeneity as best as possible. Thus each data point belongs to a specific Eco-Subsection and it's this grouping that gives us a clustered data structure.

If we look at the distribution of the FIA collected forest attribute "Dry Above Ground Biomass From live Trees", we can see that it is indeed quite zero-inflated.
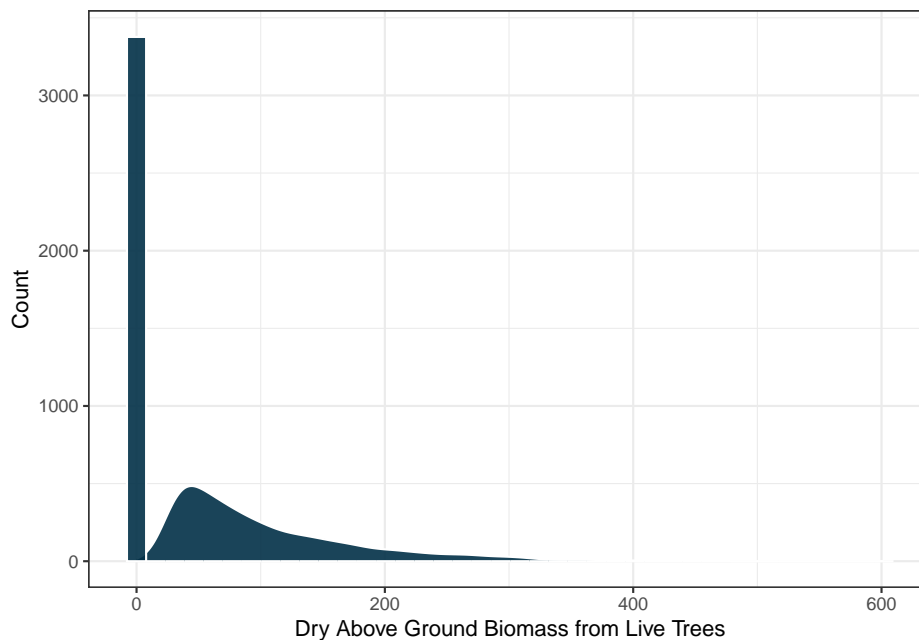
Not only is the FIA data a good real life example of when we see this zero-inflated and clustered data structure, but it's also a setting in which it's quite important that the models used to estimate these forest attributes are sufficiently accurate and efficient.

1.3.1   Immediate Modeling Struggles

To motivate using a more complex method, I'll first show what happens when I try to just fit a simple linear regression to this type of data. If we regress our response variable on a useful covariate and plot both the data and the simple linear regression line together we get the following
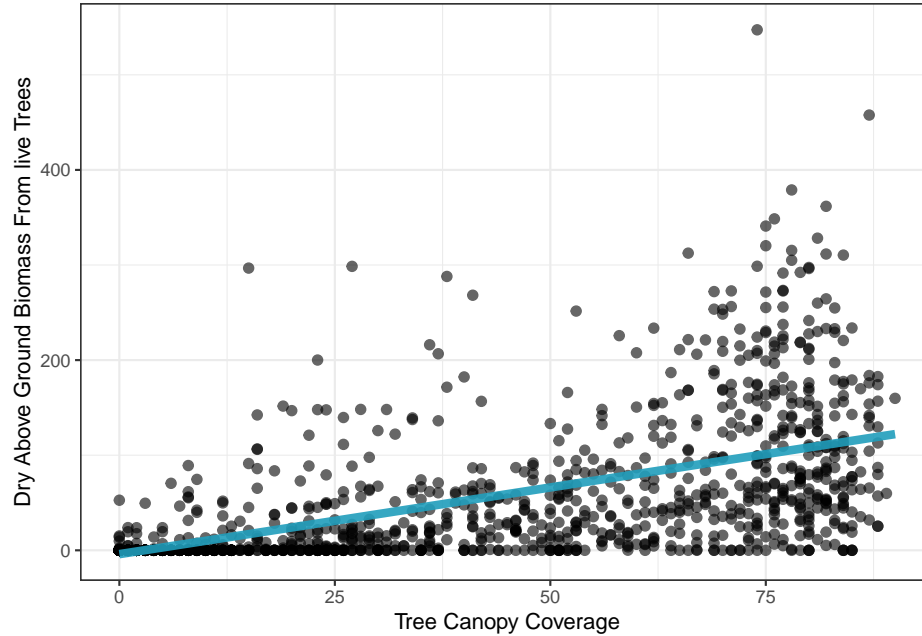
While this model isn't awful it's certainly misspecified. What I mean by this is that a simple straight line doesn't appropriately capture the dynamics of the relationship between our covariate and our response. The zero-inflation in the response variable pulls the regression line down so that it doesn't properly capture the relationship between the explanatory variable and the *non-zero* response, but more importantly it doesn't capture the structure of zeros in the response at all. We can see that the only time this model will predict a near zero response is when the covariate value is very close to zero, but this is an extreme limitation of the model since we observe zero response values across almost the entire range observed values for the covariate. What's more, a simple linear regression model does not allow us to understand how the probability our response variable being zero, changes with our covariate.

We would call this model statistically biased, as it is overly simple and thus doesn't properly capture the structure of the data. While it's perhaps feels unfair to motivate my method by piting it against the simplest of statistical models, the reality is that linear regression is a very powerful and widely used model. Moreover, in a setting such as this one where the data looks plausibly linear, the principle of parsimony might make a linear regression model a well reasoned choice. While there's certainly a need for a model that is better fit to the data, I won't go down the route of constructing an incredibly opaque and complex deep learning model to do so. Instead they model I present is interpretable and intuitive while being flexible enough to capture the structure of the zero-inflated data.

1.3.2   The New Model

While I will exhaustively describe details of, and the math behind, the exact model in a later section, I'll go through a non-technical overview of how it will function here.

The defining characteristic of the model is that it is a two-part model. Instead of trying to fit the data with a singular model, we instead fit two different models and then combine them at the end. The two models are

1. A classification model fit to the entire data set that predicts how likely it is that a certain data point has a non-zero response value.

2. A regression model fit to the non-zero portion of the dataset that predicts the continuous response variable.

To get a final prediction for a data point we take the prediction from model (1) and multiply it by the prediction from model (2).

$$\text{final prediction} = \underbrace{\Big(\text{regression model output}\Big)}_{\text{Model (2)}} \times \underbrace{\Big(\text{how likely it is that that point is non-zero}\Big)}_{\text{Model (1)}}$$

The intuition here is that if our classification model is sufficiently accurate, then points that indeed have zero-response will get sent towards zero due to the fact that the regression output will be multiplied by a number close to zero, while points that have a non-zero response will remain unchanged when multiplied by a number close to one. This method also operates under the idea that a regression model can be well fit to the non-zero portion of the response variable, and thus our regression component should be less biased than if we just fit a single regression model to the entire data set like we described in the previous section.

1.3.3   Building the Model

Now, as the title suggests I'll be building these models in a Bayesian frame. But what does that even mean and why would one want to do that? While most of the thesis will be devoted to answering the second question, I'll spend some time in the next section describing Bayesian methods, and walking through how they differ from a Frequentist approach.

Importantly, since this thesis is simply an earnest exploration of a Bayesian method, it has no intention to participate in the deep and opaque philosophical dia-

logue regarding whether Bayesian or Frequentist methods are a more "correct" way to do statistics.

That being said, the word Bayesian is so overwhelmingly ideologically tied to this statistical dichotomy that it is, by nature, very difficult to talk about a Bayesian method without talking about Frequentism as well. Because classical statistical methods are all Frequentist ones, there is often a pressure to validate a Bayesian method by standing it next to its Frequentist counterpart. While this Bayesian thesis will indeed feature an alternative Frequentist method, it does so, not to argue for one side or the other, but rather to illustrate some of the key differences in, and logic behind, Bayesian and Frequentist analyses.

1.4   Looking Ahead

In order to introduce, study, and implement these models I will structure the research in the following way:

- Chapter 2 gives a thorough functional overview of how Bayesian and Frequentist methods differ in the simple setting of inference for a mean. The goal here is primarily to provide a gentle introduction to Bayesian data analysis, so as not to drop the reader into the deep end when the main model is introduced.

- Chapter 3 gives a detailed overview of all the methods employed in the thesis. It starts with a high-level description of the Zero-Inflation model, before moving on to detailed descriptions of how each model will be built. Next, prediction for Bayesian models is illustrated both theoretically and computationally. Finally, a mathematical proof is presented to justify building each part of the Bayesian two part model separately.

- Chapter 4 sets up the simulation study that serves as the main process by which we evaluate the various models.

- Chapter 5 showcases the results of each model's performance. Beyond comparing the performance metrics of each model, I also describe the challenges associated with making comparisons between Bayesian and Frequentist models in a complex setting like this one.

- Chapter 6 gives an overview of the R Package written to accompany the methods explored in this thesis. A vignette is provided the applies the R Package to the Forestry data setting.

1.5   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.

2.2 Chemistry 101: Symbols

Chemical formulas will look best if they are not italicized. Get around math mode's automatic italicizing in LaTeX by using the argument `$\mathrm{formula here}$`, with your formula inside the curly brackets. (Notice the use of the backticks here which enclose text that acts as code.)

So, $\mathrm{Fe_2^{2+}Cr_2O_4}$ is written `$\mathrm{Fe_2^{2+}Cr_2O_4}$`.

Exponent or Superscript: O^-

Subscript: CH_4

To stack numbers or letters as in Fe_2^{2+}, the subscript is defined first, and then the superscript is defined.
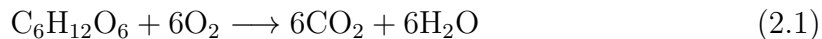
Bullet: $CuCl \bullet 7H_2O$

Delta: Δ

Reaction Arrows: $\longrightarrow$ or $\xrightarrow{solution}$

Resonance Arrows: $\leftrightarrow$

Reversible Reaction Arrows: $\rightleftharpoons$

2.2.1   Typesetting reactions

You may wish to put your reaction in an equation environment, which means that LaTeX will place the reaction where it fits and will number the equations for you.

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O \tag{2.1}$$

We can reference this combustion of glucose reaction via Equation (2.1).

2.2.2   Other examples of reactions

$NH_4Cl_{(s)} \rightleftharpoons NH_{3(g)} + HCl_{(g)}$

$MeCH_2Br + Mg \xrightarrow[below]{above} MeCH_2 \bullet Mg \bullet Br$
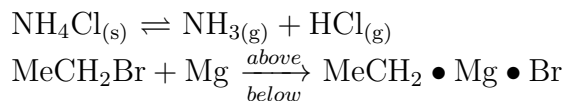
2.3   Physics

Many of the symbols you will need can be found on the math page `https://web.reed.edu/cis/help/latex/math.html` and the Comprehensive LaTeX Symbol Guide (`https://mirror.utexas.edu/ctan/info/symbols/comprehensive/symbols-letter.pdf`).

2.4   Biology

You will probably find the resources at `https://www.lecb.ncifcrf.gov/~toms/latex.html` helpful, particularly the links to bsts for various journals. You may also be interested in TeXShade for nucleotide typesetting (`https://homepages.uni-tuebingen.de/beitz/txe.html`). Be sure to read the proceeding chapter on graphics and tables.

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 `reed.jpg` in our main directory. We then give it the caption of "Reed logo", the label of "reedlogo", 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/reed.jpg")
```


Figure 3.1: Reed logo

Here is a reference to the Reed 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 **??**. (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")
```
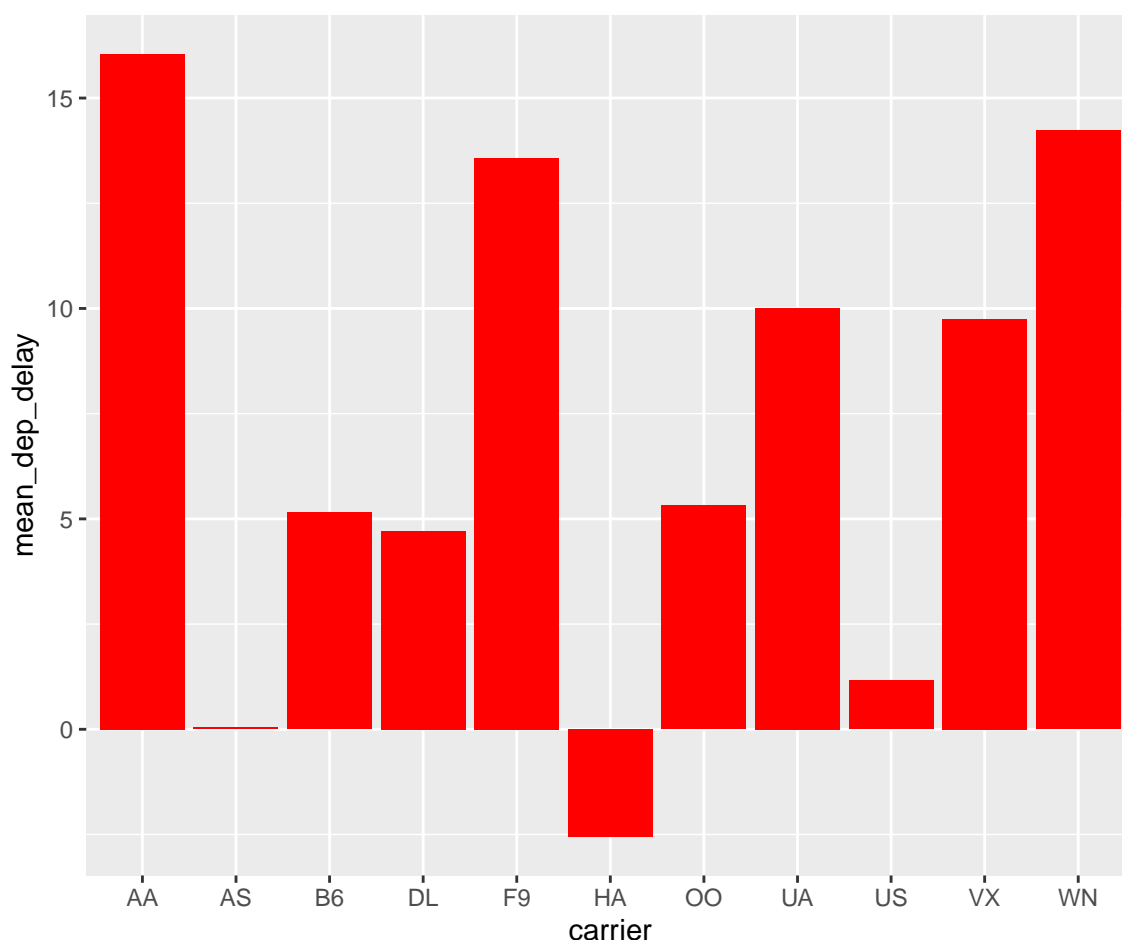

Figure 3.2: Mean Delays by Airline

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. More information can be found about both on the CUS site or feel free to reach out to `data@reed.edu`.

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

---

[1] footnote text

database (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. The Reed librarians have created Zotero documentation at `https://libguides.reed.edu/citation/zotero`. In addition, a tutorial is available from Middlebury College at `https://sites.middlebury.edu/zoteromiddlebury/`.

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.

For more information about BibTeX and bibliographies, see our CUS site (`https://web.reed.edu/cis/help/latex/index.html`)[2]. There are three pages on this topic: *bibtex* (which talks about using BibTeX, at `https://web.reed.edu/cis/help/latex/bibtex.html`), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at `https://web.reed.edu/cis/help/latex/bibtexstyles.html`) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at `https://web.reed.edu/cis/help/latex/bibman.html`). The last page will not be useful unless you have only a few sources.

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.

---

[2]Reed College (2007)

- 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.
- You can add a Reed Thesis citation[3] option. The best way to do this is to use the phdthesis type of citation, and use the optional "type" field to enter "Reed thesis" or "Undergraduate thesis."

3.4   Anything else?

If you'd like to see examples of other things in this template, please contact the Data @ Reed team (email `data@reed.edu`) with your suggestions. We love to see people using *R Markdown* for their theses, and are happy to help.

---

[3]Noble (2002)

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 readibility and/or setup.

In the main Rmd file

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
if (!require(remotes)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("remotes", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste('You need to run install.packages("remotes")',
            "first in the Console.')
    )
  }
}
if (!require(thesisdown)) {
  if (params$`Install needed packages for {thesisdown}`) {
    remotes::install_github("ismayc/thesisdown")
  } else {
    stop(
      paste(
        "You need to run",
```

```r
      'remotes::install_github("ismayc/thesisdown")',
      "first in the Console."
    )
   )
  }
}
library(thesisdown)
# Set how wide the R output will go
options(width = 70)
```

In Chapter 3:

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis and also two functions
# used for labeling and referencing
if (!require(remotes)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("remotes", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("remotes")',
        "first in the Console."
      )
    )
  }
}
if (!require(dplyr)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("dplyr", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("dplyr")',
        "first in the Console."
```

```r
      )
    )
  }
}
if (!require(ggplot2)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("ggplot2", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("ggplot2")',
        "first in the Console."
      )
    )
  }
}
if (!require(bookdown)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("bookdown", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("bookdown")',
        "first in the Console."
      )
    )
  }
}
if (!require(thesisdown)) {
  if (params$`Install needed packages for {thesisdown}`) {
    remotes::install_github("ismayc/thesisdown")
  } else {
    stop(
      paste(
        "You need to run",
        'remotes::install_github("ismayc/thesisdown")',
```

```
        "first in the Console."
      )
    )
  }
}
library(thesisdown)
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.

McConville, K. S., Moisen, G. G., & Frescino, T. S. (2020). A tutorial on model-assisted estimation with application to forest inventory. *Forests*, *11*(2), 244.

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.

Noble, S. G. (2002). *Turning images into simple line-art* (Undergraduate thesis). Reed College.

Reed College. (2007). LaTeX your document. Retrieved from `https://web.reed.edu/cis/help/LaTeX/index.html`