

My thesis title

A dissertation presented

by

My Name

to

Department of my department

in partial fulfillment of the requirements

for the degree of

PhD

in the subject of

My field

Harvard University

Cambridge, Massachusetts

May 2020

© 2020 My Name

All rights reserved.

Dissertation Advisor:
My advisor

Author:
My Name

My thesis title

ABSTRACT

“Here is my abstract”

Contents

Title Page	i
Copyright	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Acknowledgments	viii
 Introduction	 1
 1 R Markdown Basics	 3
1.1 Lists	3
1.2 Line breaks	4
1.3 R chunks	5
1.4 Inline code	5
1.5 Including plots	6
1.6 Loading and exploring data	7
1.7 Additional resources	11
 2 Mathematics and Science	 12
2.1 Math	12
2.2 Chemistry 101: Symbols	12
2.3 Physics	13
2.4 Biology	14

3	Tables, Graphics, References, and Labels	15
3.1	Tables	15
3.2	Figures	22
3.3	Footnotes and Endnotes	25
3.4	Cross-referencing chapters and sections	26
3.5	Bibliographies	27
	Conclusion	29
3.6	More Info	29
A	The First Appendix	30
A.1	In the main Rmd file	30
A.2	In Chapter 3	30
B	The Second Appendix, for Fun	31
	References	32

List of Tables

1.1	Max Delays by Airline	9
3.1	Sample of Airline Data	16
3.2	Sample of Airline Data	19

List of Figures

3.1	Harvard logo	23
3.2	Mean Delays by Airline	24
3.3	Subdiv. graph	25
3.4	A Larger Figure, Flipped Upside Down	25

Acknowledgments

“My acknowledgments”

“My dedication”

Introduction

Welcome to the *R Markdown* thesis template. This template is based on the Harvard Graduate School of Arts and Sciences thesis formatting guidelines, from which a LaTeX class was written, but hopefully it will provide a nicer interface for those that have never used TeX or LaTeX before. The structure of this package and most of this document comes directly from Ben Marwick's huskydown package.

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 at least 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.

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 <http://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).

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:

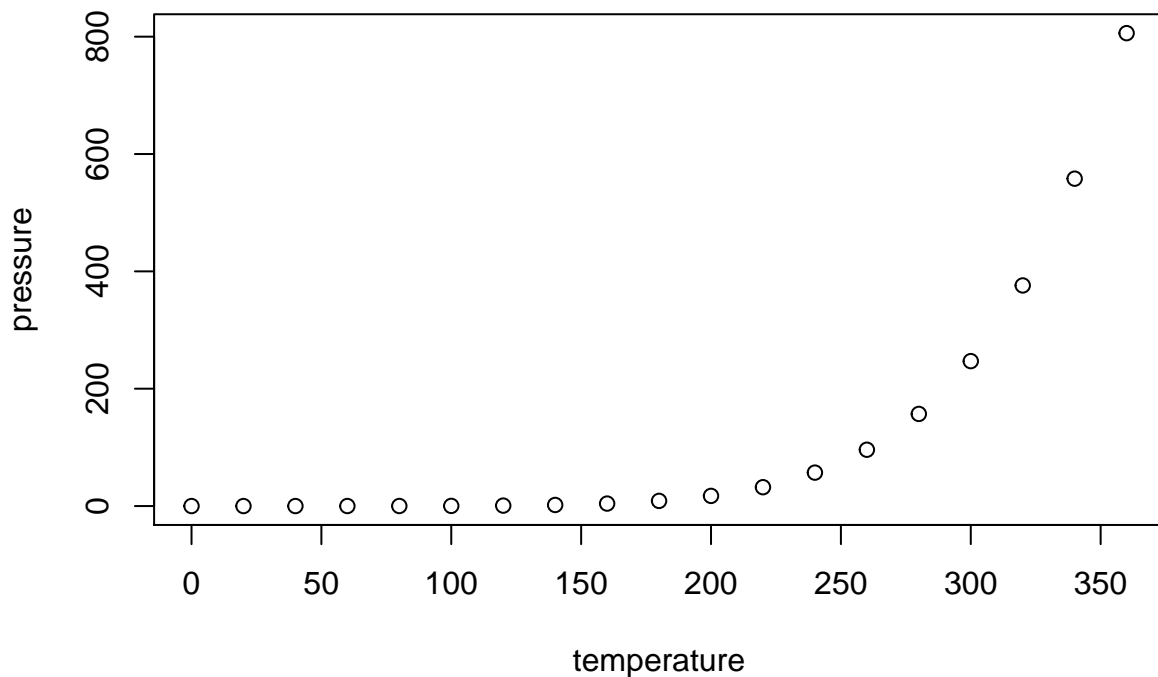
The standard deviation is less than 6.

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. More information is available at <http://yihui.name/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 <http://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 command:

```
flights <- read.csv("data/flights.csv")
```

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] 52808    16
```

```
names(flights)
```

```
[1] "month"      "day"        "dep_time"   "dep_delay"  "arr_time"
[6] "arr_delay"  "carrier"    "tailnum"    "flight"     "dest"
[11] "air_time"   "distance"   "hour"       "minute"     "carrier_name"
[16] "dest_name"
```

Another good idea is to take a look at the dataset in table form. With this dataset having more than 50,000 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", "kableExtra")
# 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 = "http://cran.rstudio.com")
library(dplyr)
library(ggplot2)
library(knitr)
library(kableExtra)
```

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.	651
Frontier Airlines Inc.	575
Hawaiian Airlines Inc.	407

JetBlue Airways	273
SkyWest Airlines Inc.	421
Southwest Airlines Co.	694
United Air Lines Inc.	472
US Airways Inc.	347
Virgin America	366

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

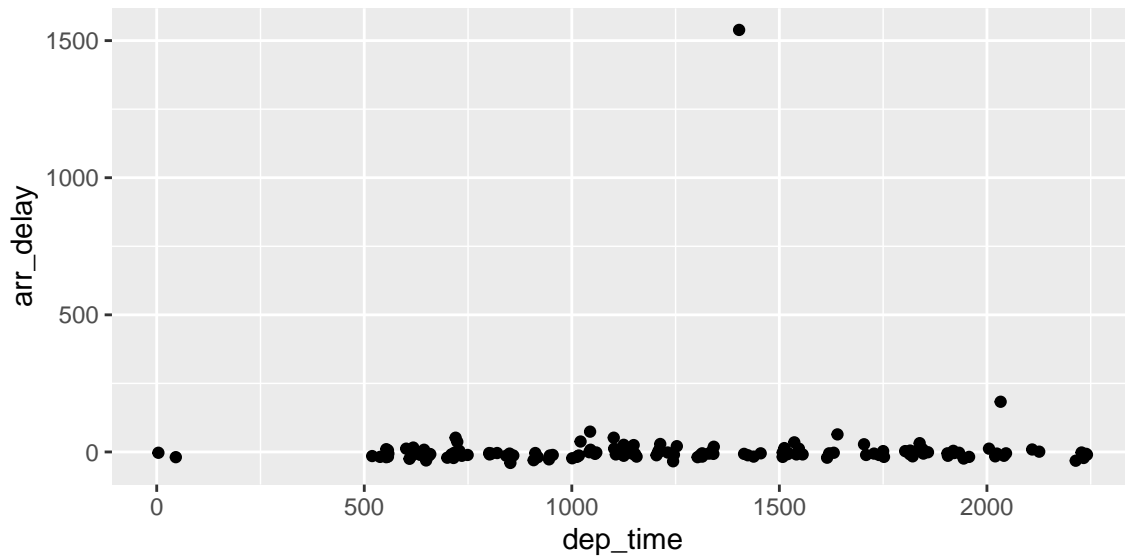
We can further look into the properties of the largest value here for American Airlines Inc. To do so, we can isolate the row corresponding to the arrival delay of 1539 minutes for American in our original flights dataset.

```
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

- *Bookdown* Book: <https://bookdown.org/yihui/bookdown/>
- *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>
- Introduction to *dplyr*: <https://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html>
- *ggplot2* Documentation: <http://docs.ggplot2.org/current/>

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.

L&P Galerkin method weighting functions. Page 145

$$\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 f(\xi, \eta, \zeta) = \sum_{k=1}^n \sum_{j=1}^n \sum_{i=1}^n w_i w_j w_k f(\xi, \eta, \zeta). \quad (2.1)$$

Another L&P (p126), which we'll give an equation number to.

$$\int_{A_e} (\cdot) dx dy = \int_{-1}^1 \int_{-1}^1 (\cdot) \det[J] d\xi d\eta. \quad (2.2)$$

We can then reference this as Equation (2.2).

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, $\text{Fe}_2^{2+}\text{Cr}_2\text{O}_4$ 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: $\text{CuCl} \bullet 7\text{H}_2\text{O}$

Delta: Δ

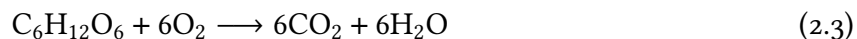
Reaction Arrows: \longrightarrow or $\xrightarrow{\text{solution}}$

Resonance Arrows: \leftrightarrow

Reversible Reaction Arrows: \rightleftharpoons

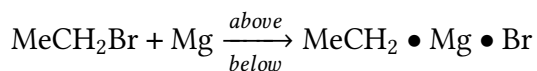
2.2.1 TYPESETTING REACTIONS

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.



We can reference this combustion of glucose reaction via Equation (2.3).

2.2.2 OTHER EXAMPLES OF REACTIONS



2.3 PHYSICS

Many of the symbols you will need can be found on the math page <http://web.reed.edu/cis/help/latex/math.html> and the Comprehensive LaTeX Symbol Guide (<http://mirror.utexas.>

[edu/ctan/info/symbols/comprehensive/symbols-letter.pdf](http://www.ctan.org/info/symbols/comprehensive/symbols-letter.pdf)).

2.4 BIOLOGY

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

Tables, Graphics, References, and Labels

3.1 TABLES

By far the easiest way to present tables in your thesis is to store the contents of the table in a CSV or Excel file, then read that file in to your R Markdown document as a data frame. Then you can style the table with the `kable` function, or functions in the `kableExtra` package.

If you are presenting tables of descriptive characteristics or of regression output, there are some R packages you may want to look into, include `tableone` and `broom`. Other helpful packages for LaTeX and html output include `xtable` and `gt`.

If you go back to the section on loading and exploring data in Chapter 1 and look at the `kable` table, we can create a table which we can reference here too: Table 1.1.

3.1.1 LONG AND WIDE TABLES

This time, when we make a table, we'll make one that takes up several pages to demonstrate how long tables work. Whenever a table continues onto the next page, the caption should be followed by "continued," which we can achieve with the `kableExtra` package.

```
#options(knitr.table.format = "latex")
kable(flights[1:60, 1:6],
      caption = "Sample of Airline Data",
```



```

longtable = TRUE,
booktabs = TRUE) %>%
kable_styling(latex_options = c("repeat_header"))

```

Table 3.1: *Sample of Airline Data*

month	day	dep_time	dep_delay	arr_time	arr_delay
1	1	1	96	235	70
1	1	8	13	548	-4
1	1	28	-2	800	-23
1	1	526	-4	1148	15
1	1	541	1	911	4
1	1	549	24	907	12
1	1	559	-1	916	-9
1	1	602	-3	1204	7
1	1	606	6	746	3
1	1	618	-2	1135	-30
1	1	622	2	1412	-19
1	1	630	-10	834	-14
1	1	637	-8	858	-7
1	1	637	-3	854	-6
1	1	638	-2	1003	-5
1	1	638	10	1408	-4
1	1	639	4	1219	-11
1	1	648	-12	910	-11
1	1	654	-6	1047	-33
1	1	655	-5	1210	-7

Table 3.1: *Sample of Airline Data (continued)*

month	day	dep_time	dep_delay	arr_time	arr_delay
1	1	656	-4	1242	-28
1	1	656	-4	853	8
1	1	700	0	844	-1
1	1	704	-1	910	-9
1	1	705	0	931	9
1	1	708	-2	1453	-20
1	1	710	-5	906	-21
1	1	715	-5	1055	-47
1	1	722	-3	933	3
1	1	754	-6	1011	-14
1	1	801	-4	932	-18
1	1	809	-6	1326	-5
1	1	819	14	1343	-7
1	1	827	2	1414	-16
1	1	840	-5	1226	16
1	1	843	-2	1159	-14
1	1	846	-4	1114	-16
1	1	919	-6	1719	-16
1	1	925	-2	1106	-17
1	1	944	-6	1727	-8
1	1	946	-4	1220	-38
1	1	949	-1	1108	-12
1	1	951	-4	1716	-24

Table 3.1: *Sample of Airline Data (continued)*

month	day	dep_time	dep_delay	arr_time	arr_delay
1	1	1005	-5	1608	2
1	1	1011	-4	1535	-29
1	1	1014	-1	1357	-23
1	1	1021	-9	1347	-10
1	1	1023	-7	1407	-38
1	1	1027	-2	1242	-16
1	1	1036	81	1408	63
1	1	1043	-8	1226	-13
1	1	1045	40	1303	40
1	1	1052	2	1603	-2
1	1	1054	94	1807	84
1	1	1055	-5	1312	-1
1	1	1059	44	1325	32
1	1	1102	2	1425	-5
1	1	1113	-2	1427	-8
1	1	1117	37	1436	36
1	1	1123	-2	1442	-12

We might also have tables that are very wide, and for which we need to switch to landscape. The code here allows for landscape mode in both pdf and word (via officedown).

Table 3.2: *Sample of Airline Data*

month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	dest	air_time	distance
1	1	1	96	235	70	AS	N508AS	145	ANC	194	1542
1	1	8	13	548	-4	UA	N37422	1609	IAH	201	1825
1	1	28	-2	800	-23	US	N547UW	466	CLT	251	2282
1	1	526	-4	1148	15	UA	N813UA	229	IAH	217	1825
1	1	541	1	911	4	UA	N36476	1569	DEN	130	991
1	1	549	24	907	12	US	N548UW	649	PHX	122	1009
1	1	559	-1	916	-9	F9	N210FR	796	DEN	125	991
1	1	602	-3	1204	7	UA	N77295	1573	ORD	203	1739
1	1	606	6	746	3	AS	N796AS	406	SJC	87	569
1	1	618	-2	1135	-30	AA	N4XMAA	1650	DFW	184	1616
1	1	622	2	1412	-19	UA	N36472	1021	EWR	270	2434
1	1	630	-10	834	-14	OO	N219AG	3456	BUR	105	817
1	1	637	-8	858	-7	AS	N408AS	560	LAX	108	834
1	1	637	-3	854	-6	AS	N588AS	572	SAN	120	933
1	1	638	-2	1003	-5	AS	N508AS	646	PHX	125	1009

1	1	638	10	1408	-4	UA	N526UA	250	IAD	253	2327
1	1	639	4	1219	-11	WN	N8308K	401	MDW	207	1751
1	1	648	-12	910	-11	AS	N611AS	582	SNA	122	859
1	1	654	-6	1047	-33	AS	N586AS	833	HNL	327	2603
1	1	655	-5	1210	-7	DL	N963DN	2090	MSP	168	1426

We make sure to label the code chunks so that we can reference the long table as 3.1 and wide table as 3.2.

3.2 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 `harvard.png` in our main directory. We then give it the caption of "Harvard logo," the label of "harvardlogo," 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/harvard.png")
```

Here is a reference to the Harvard 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`.

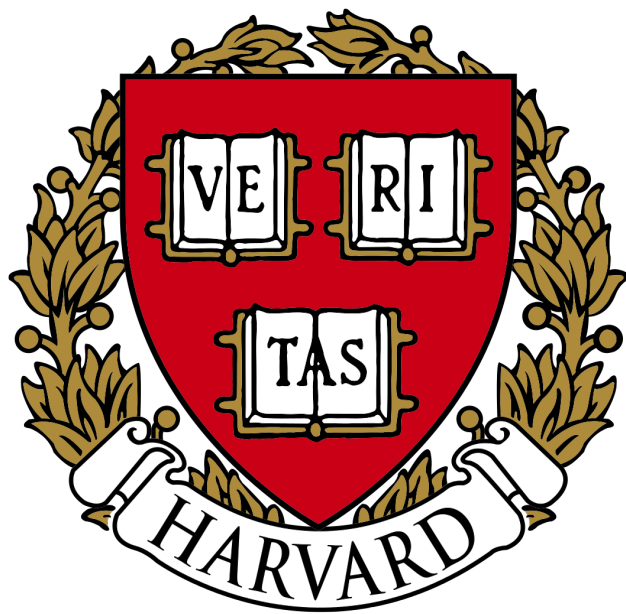


Figure 3.1: *Harvard logo*

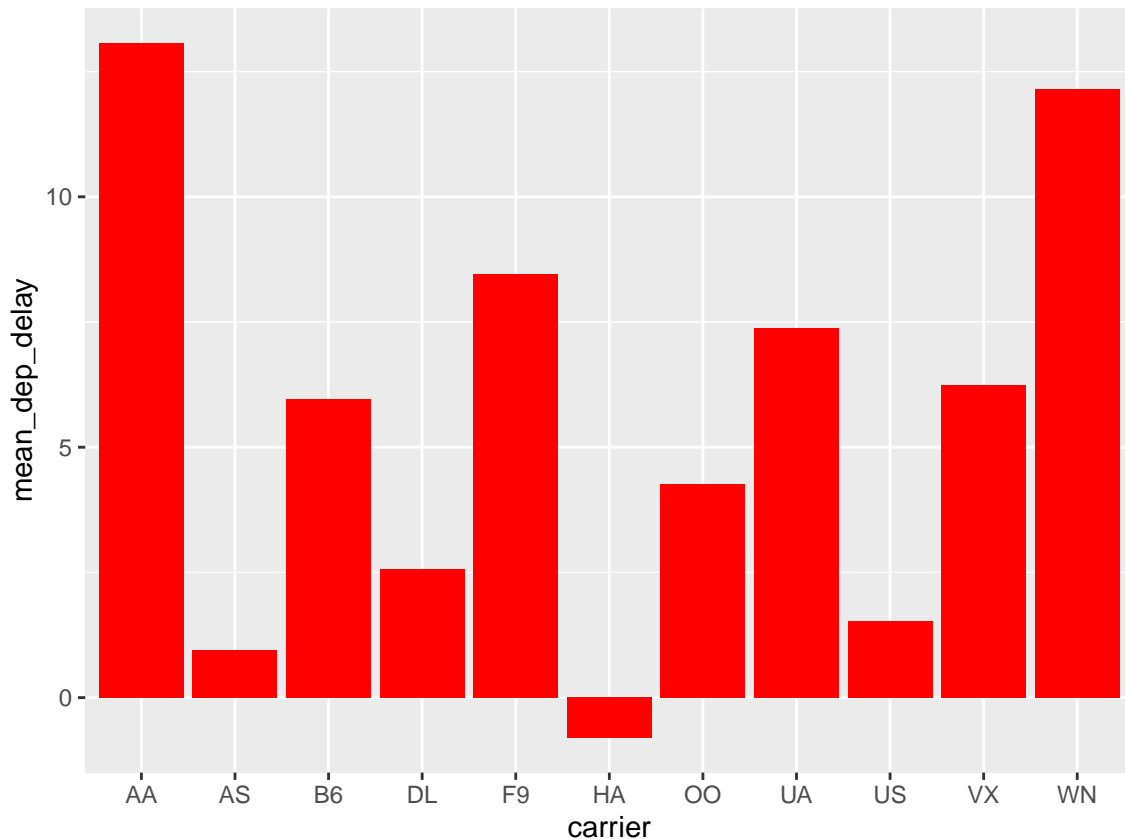


Figure 3.2: *Mean Delays by Airline*

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. Note also the use of the `scale` parameter which is discussed on the next page.

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

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>.

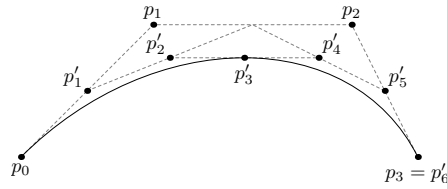


Figure 3.3: *Subdiv. graph*

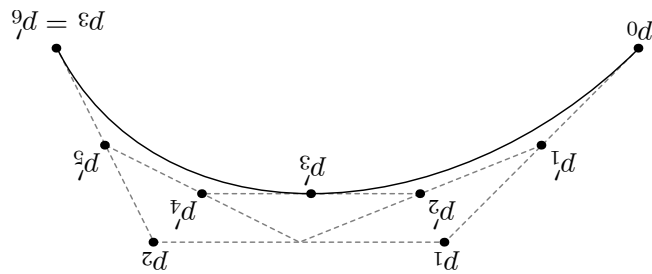


Figure 3.4: *A Larger Figure, Flipped Upside Down*

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

3.3 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.

¹footnote text

3.4 CROSS-REFERENCING CHAPTERS AND SECTIONS

The bookdown documentation is an excellent source for learning how to cross-reference in a bookdown project such as a crimsondown document. Here we only cover the most common uses for a typical thesis. If you want something more complex or fancy, please refer to the bookdown documentation and seek help from the developers of that package.

By default, all of your chapter and section headers will get an auto-generated ID label. For example, e.g., `# Chapter 1` will have an auto-generated ID `chapter-1`. Note that the ID label is all lower case, and has no spaces. If you have any kind of punctuation in your header, such as a colon (:), it will not appear in the ID label. Then in your text you can reference chapter one in your Rmd file like this: ‘as discussed in Chapter `\@ref(chapter-1)`’, which will print as ‘as discussed in Chapter 1’

We strongly recommend that you to manually assign ID labels to your chapter header to make it easy to cross-reference. For example, at the top of the Rmd file for this chapter, you can see:

```
# Tables, Graphics, References, and Labels {#ref-labels}
```

The `{#ref-labels}` part of this header is the ID label. It doesn’t show in the output, but is there for us to use for easy cross-referencing, because it can be short, and we don’t need to change it elsewhere our document when we update the chapter header. We can use this custom ID label in our Rmd document like this: ‘as discussed in Chapter `\@ref(ref-labels)`’, which will print as ‘as discussed in Chapter 3.’ If you need to show custom text instead of the chapter number, you use this syntax in your Rmd document: `see [my chapter about labels](#ref-labels)` for more details which will appear as ‘see my chapter about labels for more details’

To cross-reference a specific section in the same chapter, we recommend adding a custom ID label to the section header, and using that to cross-reference. For example, earlier in this chapter we have a section on tables and in the Rmd file we see `## Tables {#tables}`. We can cross-reference that in the text like this ‘as discussed in the section on `[tables](#tables)`’ which

will appear as ‘as discussed in the above section on tables’

To cross-reference a section in a different chapter we can use the ID label from that section directly. For example, we can write in our Rmd document as discussed in the section on [R code chunks](#r-chunks) in Chapter \@ref(rmd-basics) which will appear as ‘as discussed in the section on R code chunks in Chapter 1.’

If you prefer to cross-reference by the section number, we can use custom ID labels in our Rmd document. For example, to refer to a section in our first chapter, we can write in the Rmd document: as discussed in section \@ref(r-chunks) in Chapter \@ref(rmd-basics). This will appear with section and chapter numbers like so: as ‘as discussed in section 1.3 in Chapter 1.’

3.5 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 (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. Some Zotero documentation is at <http://libguides.reed.edu/citation/zotero>. In addition, a tutorial is available from Middlebury College at <http://sites.middlebury.edu/zoteromiddlebury/>.

R Markdown uses *pandoc* (<http://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 placed in the bib folder.

For more information about BibTeX and bibliographies, see (<http://web.reed.edu/cis/help/latex/index.html>)². There are three pages on this topic: *bibtex* (which talks about using BibTeX, at <http://web.reed.edu/cis/help/latex/bibtex.html>), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at <http://web.reed.edu/cis/help/latex/bibtexstyles.html>) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at <http://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.
- 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.

²Reed College (2007)

Conclusion

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

3.6 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.

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.

A.1 IN THE MAIN RMD FILE

A.2 IN CHAPTER 3

B

The Second Appendix, for Fun

Any files between the first appendix and the references will also be appendices.

References

- 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.
- Reed College. (2007). LaTeX your document. Retrieved from <http://web.reed.edu/cis/help/LaTeX/index.html>