

IMPERIAL

IMPERIAL COLLEGE LONDON

Department of Dinosaurs

Imperial College PhD Thesis

Ross Geller

November 2024

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Dinosaurs of Imperial College London and
the Diploma of Imperial College London.

Declaration of Originality

I, Ross Geller, declare that the work in this thesis is my own. The work of others has been appropriately referenced. A full list of references is given in the bibliography.

Copyright

The copyright of this thesis rests with the author and its contents are made available under a Creative Commons Attribution Non-Commercial Share-Alike 4.0 International (CC BY-NC-SA 4.0) License. You may copy and redistribute the material in any medium or format. You may also remix, transform or build upon the material. In doing so, you must give appropriate credit to the author, provide a link to the license and indicate if any changes were made. If you remix, transform or build upon this material, you must redistribute your contributions under the same license. You may not use the material for commercial purposes.

Please seek permission from the copyright holder for uses of this work that are not included in the license mentioned above.

Abstract

Some abstract that manages to somehow summarise everything the entire thesis did.

Good luck.

Acknowledgements

Firstly, the loved ones. For example your family is a great starting point.

Secondly, your supervisor.

Finally, your colleagues and friends.

Dedication

For my loved one(s).

Contents

1	R Markdown Basics	10
1.1	Lists	10
1.2	Line breaks	11
1.3	R chunks	11
1.4	Inline code	11
1.5	Including plots	12
1.6	Loading and exploring dinosaur diet data	12
1.7	Additional resources	17
2	Mathematics and Science	19
2.1	Math	19
2.2	Chemistry 101: Symbols	19
2.2.1	Typesetting reactions	20
2.2.2	Other examples of reactions	20
2.3	Physics	20
2.4	Biology	20
3	Graphics, References, and Labels	21
3.1	Figures	21
3.2	Footnotes and Endnotes	23
3.3	Bibliographies	24
3.4	Anything else?	24
	Conclusion	25
A	The First Appendix	26
B	The Second Appendix, for Fun	29
	References	30

List of Tables

1.1	Average consumption	14
1.2	Diet summary	14
1.3	Food preferences	16

List of Figures

1.1	Feeding Patterns Throughout the Day by Diet Type	15
1.2	Distribution of Meal Sizes by Diet Type	17
1.3	Daily Food Consumption Trends by Diet Type	18
3.1	Imperial logo	21
3.2	Food Eaten by Each Dinosaur	22
3.3	Subdiv. graph	23
3.4	A Larger Figure, Flipped Upside Down	23

Some abstract that manages to somehow summarise everything the entire thesis did.

Good luck.

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).

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):

```
knitr::kable(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 dinosaur diet data

This analysis explores feeding records of various dinosaur species, including their diet types, food preferences, and amount of food consumed across multiple days.

```
# Read the CSV data
dino_data <- read.csv('data/dino_data.csv')

# Clean column names by removing spaces and making them R-friendly
names(dino_data) <- c("Day", "Time", "Species", "Diet", "Food_Type", "Kg_Food")
```

Let's examine the basic structure and summary of our dataset:

```
dim(dino_data)
```

```
[1] 49 6
```

```
knitr::kable(summary(dino_data[1:4]))
```

	Day	Time	Species	Diet
	Min. : 1.000	Length:49	Length:49	Length:49
	1st Qu.: 3.000	Class :character	Class :character	Class :character
	Median : 5.000	Mode :character	Mode :character	Mode :character
	Mean : 5.408	NA	NA	NA
	3rd Qu.: 8.000	NA	NA	NA
	Max. :10.000	NA	NA	NA

Loading required packages for analysis:

```
pkg <- c("dplyr", "ggplot2", "knitr")
new.pkg <- pkg[!(pkg %in% installed.packages())]
if (length(new.pkg)) {
  install.packages(new.pkg, repos = "https://cran.rstudio.com")
}
library(dplyr)
library(ggplot2)
library(knitr)
```

Warning: package 'knitr' was built under R version 4.3.3

Let's analyze feeding patterns by species:

```
# Calculate average daily consumption by species
avg_consumption <- dino_data %>%
  group_by(Species) %>%
  summarize(
    avg_daily_kg = mean(Kg_Food),
    total_kg = sum(Kg_Food),
    diet_type = first(Diet),
    n_meals = n()
  ) %>%
  arrange(desc(avg_daily_kg))
# Simple table output
kable(avg_consumption,
      caption = "Average consumption",
      col.names = c("Species", "Avg Kg per Meal", "Total Kg Eaten",
                    "Diet Type", "Number of Meals"))
```

Analyzing diet types:

```
diet_summary <- dino_data %>%
  group_by(Diet) %>%
  summarize(
    avg_meal_size = mean(Kg_Food),
    max_meal = max(Kg_Food),
    total_consumption = sum(Kg_Food),
    n_meals = n()
  )

kable(diet_summary,
      caption = "Diet summary",
      col.names = c("Diet Type", "Avg Meal Size (kg)",
                    "Max Meal Size (kg)", "Total Food (kg)", "Number of Meals"))
```

Table 1.1: Average consumption

Species	Avg Kg per Meal	Total Kg Eaten	Diet Type	Number of Meals
Brachiosaurus	125.00000	375.0	Herbivore	3
Diplodocus	82.50000	165.0	Herbivore	2
Tyrannosaurus	75.00000	225.0	Carnivore	3
Carnotaurus	62.50000	125.0	Carnivore	2
Triceratops	54.33333	163.0	Herbivore	3
Edmontosaurus	52.50000	105.0	Herbivore	2
Allosaurus	50.00000	150.0	Carnivore	3
Iguanodon	47.50000	95.0	Herbivore	2
Spinosaurus	42.33333	127.0	Carnivore	3
Corythosaurus	41.00000	82.0	Herbivore	2
Stegosaurus	34.33333	103.0	Herbivore	3
Parasaurolophus	33.50000	67.0	Herbivore	2
Therizinosaurus	29.50000	59.0	Herbivore	2
Ankylosaurus	27.66667	83.0	Herbivore	3
Maiasaura	26.00000	52.0	Herbivore	2
Pachycephalosaurus	12.50000	25.0	Herbivore	2
Velociraptor	12.00000	36.0	Carnivore	3
Gallimimus	5.50000	11.0	Omnivore	2
Oviraptor	4.00000	12.0	Omnivore	3
Compsognathus	1.25000	2.5	Carnivore	2

Table 1.2: Diet summary

Diet Type	Avg Meal Size (kg)	Max Meal Size (kg)	Total Food (kg)	Number of Meals
Carnivore	41.59375	80	665.5	16
Herbivore	49.07143	130	1374.0	28
Omnivore	4.60000	6	23.0	5

Visualizing feeding patterns:

```
ggplot(dino_data, aes(x = Time, y = Kg_Food, color = Diet)) +
  geom_point(size = 3, alpha = 0.7) +
  facet_wrap(~Diet) +
  theme_minimal() +
  labs(title = "Feeding Patterns by Diet Type",
       x = "Time of Day",
       y = "Food Consumed (kg)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Distribution of meal sizes:

```
ggplot(dino_data, aes(x = Diet, y = Kg_Food, fill = Diet)) +
  geom_boxplot() +
  geom_jitter(width = 0.2, alpha = 0.3) +
```

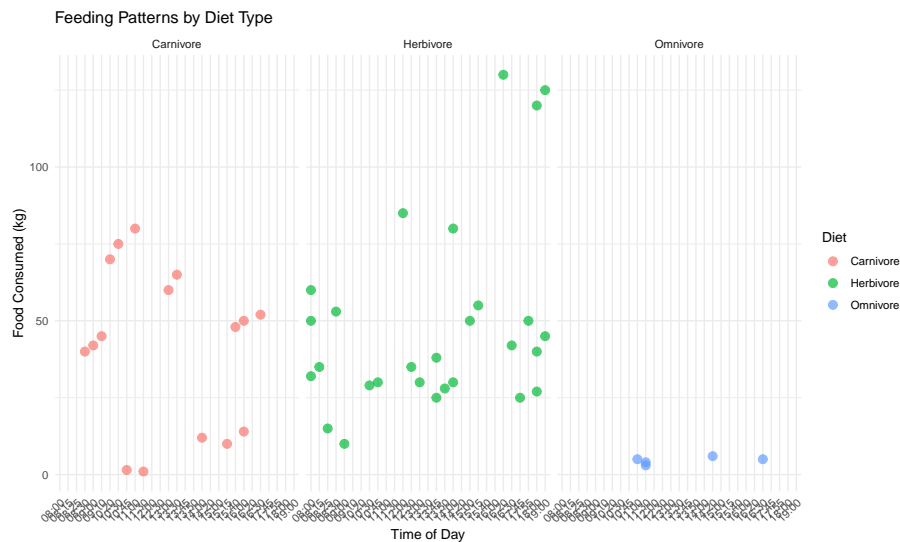


Figure 1.1: Feeding Patterns Throughout the Day by Diet Type

```
theme_minimal() +
labs(title = "Meal Sizes by Diet Type",
     x = "Diet Category",
     y = "Food Consumed (kg)") +
theme(legend.position = "none")
```

Food type preferences:

```
food_preferences <- dino_data %>%
  group_by(Diet, Food_Type) %>%
  summarize(
    times_eaten = n(),
    avg_portion = mean(Kg_Food)
  ) %>%
  arrange(Diet, desc(times_eaten))
```

'summarise()' has grouped output by 'Diet'. You can override using the '.groups' argument.

```
kable(food_preferences,
      caption = "Food preferences",
      col.names = c("Diet Type", "Food Type",
                    "Times Eaten", "Average Portion (kg)"))
```

Daily consumption trends:

```
daily_total <- dino_data %>%
  group_by(Day, Diet) %>%
  summarize(total_food = sum(Kg_Food))
```

Table 1.3: Food preferences

Diet Type	Food Type	Times Eaten	Average Portion (kg)
Carnivore	Fish	3	42.33333
Carnivore	Small Dinosaur	3	24.66667
Carnivore	Small Herbivore	2	30.00000
Carnivore	Edmontosaurus	1	70.00000
Carnivore	Herbivore	1	52.00000
Carnivore	Iguanodon	1	80.00000
Carnivore	Insects	1	1.00000
Carnivore	Large Herbivore	1	65.00000
Carnivore	Medium Herbivore	1	60.00000
Carnivore	Reptiles	1	1.50000
Carnivore	Triceratops	1	75.00000
Herbivore	Ferns	6	39.66667
Herbivore	Conifer Leaves	3	108.33333
Herbivore	Moss	3	39.33333
Herbivore	Aquatic Plants	2	45.00000
Herbivore	Aquatic Vegetation	2	48.50000
Herbivore	Conifer Foliage	1	85.00000
Herbivore	Cycads	1	60.00000
Herbivore	Fruits	1	15.00000
Herbivore	High Trees	1	130.00000
Herbivore	Low Plants	1	30.00000
Herbivore	Low Vegetation	1	29.00000
Herbivore	Low-growing Plants	1	25.00000
Herbivore	Low-lying Plants	1	25.00000
Herbivore	Plants	1	30.00000
Herbivore	Seeds	1	10.00000
Herbivore	Soft Plants	1	32.00000
Herbivore	Soft Vegetation	1	35.00000
Omnivore	Plants	2	5.50000
Omnivore	Small Animals	2	4.50000
Omnivore	Eggs	1	3.00000

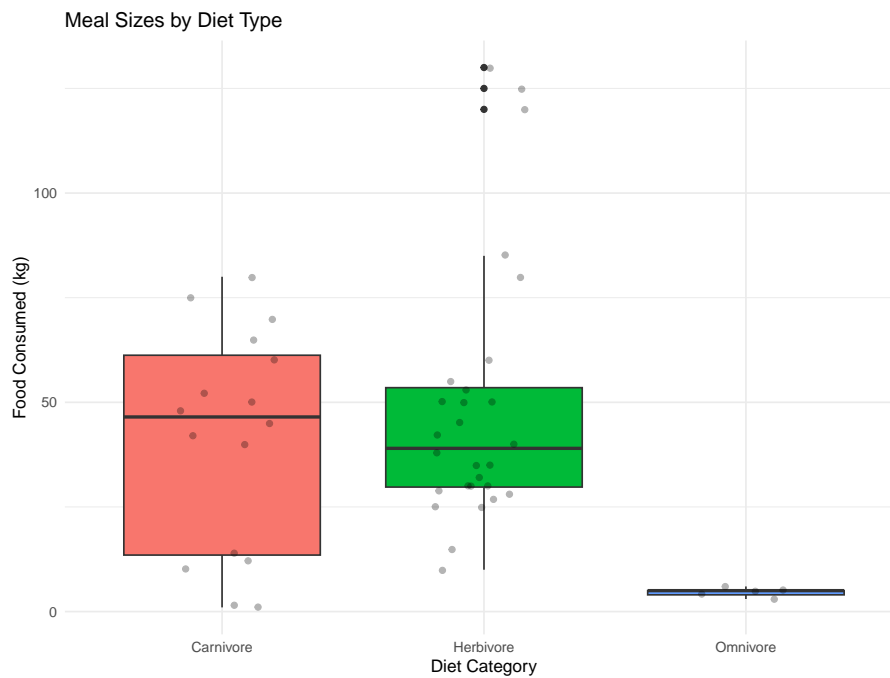


Figure 1.2: Distribution of Meal Sizes by Diet Type

'summarise()' has grouped output by 'Day'. You can override using the '.groups' argument.

```
ggplot(daily_total, aes(x = Day, y = total_food, color = Diet, group = Diet)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Daily Consumption by Diet Type",
       x = "Day",
       y = "Total Food Consumed (kg)")
```

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/>

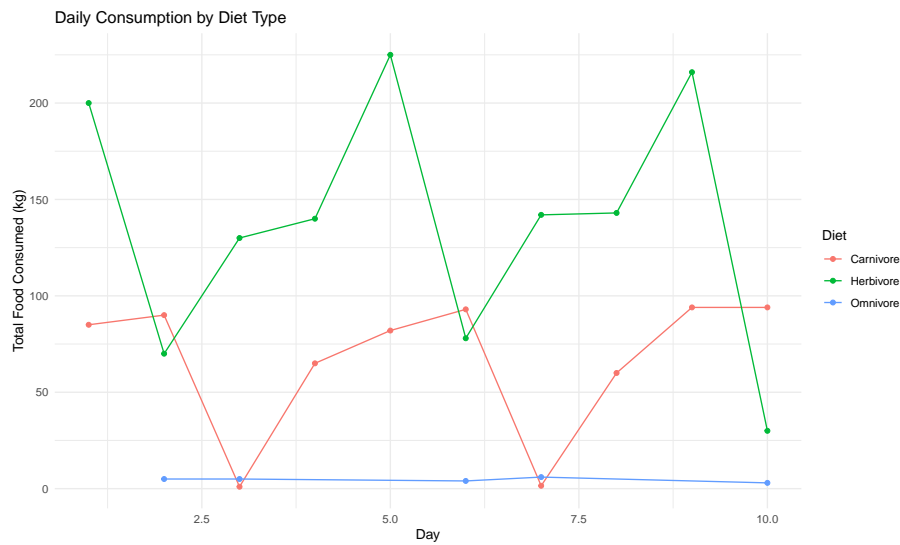


Figure 1.3: Daily Food Consumption Trends by Diet Type

- 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-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, $\text{Fe}_2^{2+}\text{Cr}_2\text{O}_4$ 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

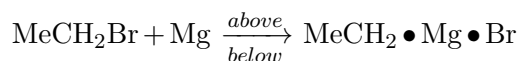
2.2.1 Typesetting reactions

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 (2.1).

2.2.2 Other examples of reactions



2.3 Physics

Many of the symbols you will need can be found on this 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 bst's 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 `logo.pdf` in our main directory. We then give it the caption of "Imperial logo", the label of "iclogo", 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/logo.pdf")
```

Here is a reference to the Imperial 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`.

IMPERIAL

Figure 3.1: Imperial logo

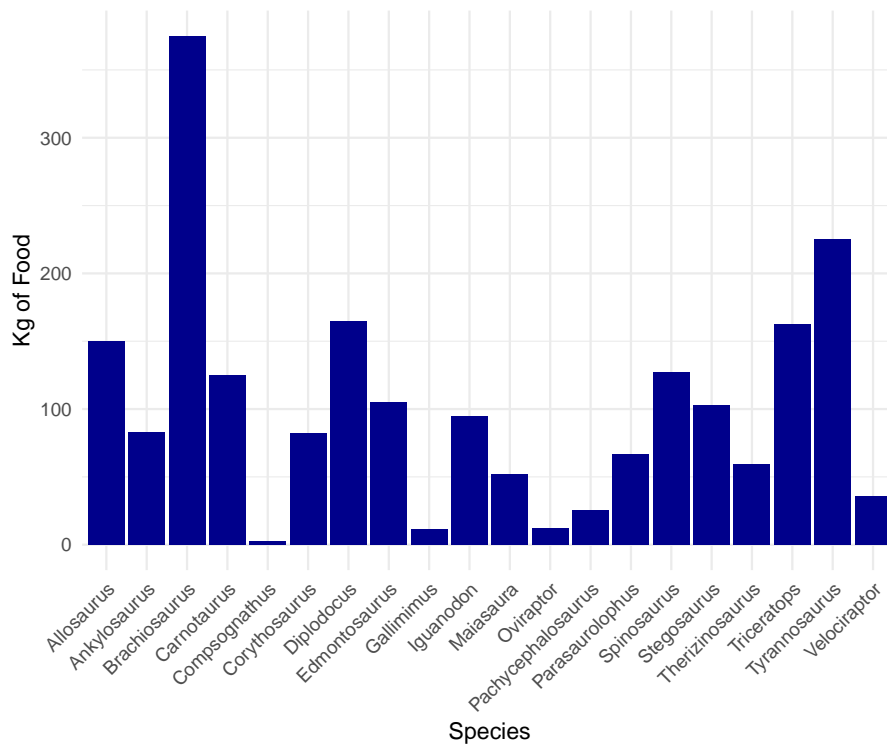


Figure 3.2: Food Eaten by Each Dinosaur

Below we will create a basic bar plot showing how much food each dinosaur ate.

```
ggplot(dinodata, aes(x = Dinosaur.Species, y = Kg.Food.Eaten)) +
  geom_bar(stat = "identity", fill = "darkblue") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(x = "Species",
       y = "Kg of Food")
```

Here is a reference to this image: Figure 3.2.

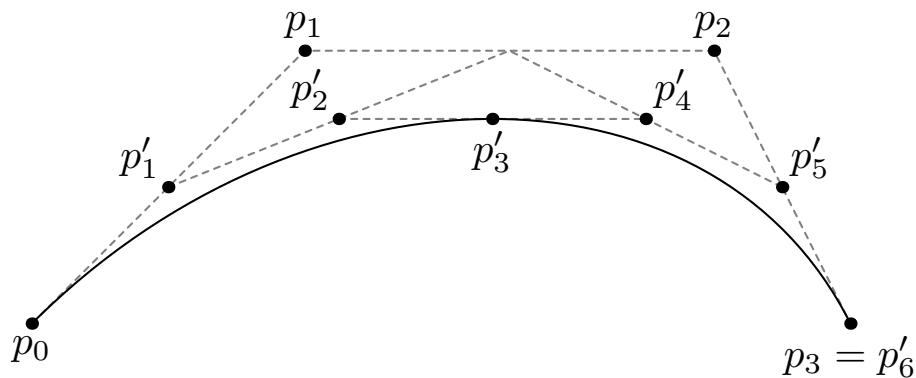


Figure 3.3: Subdiv. graph

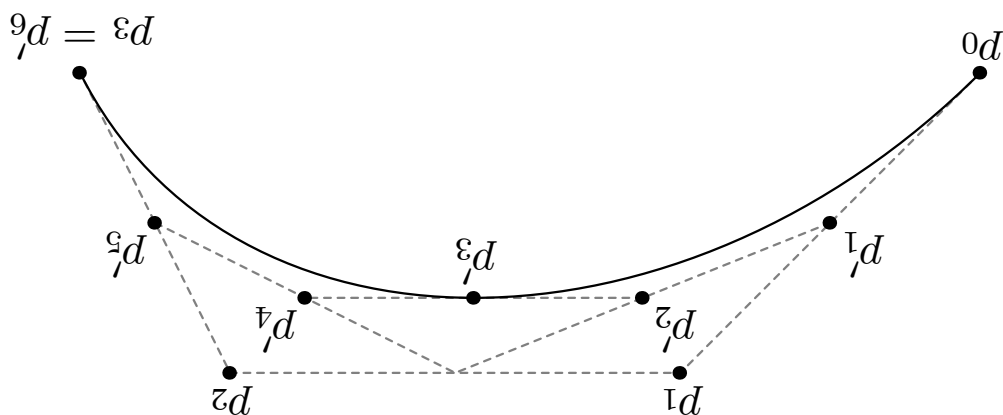


Figure 3.4: A Larger Figure, Flipped Upside Down

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.

3.2 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.

¹footnote text

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 (stored with the .bib extension).

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 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.

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 Jay at ECRI (email j.des@imperial.ac.uk) with your suggestions. 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:

```
# 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."
      )
    )
  }
}
```

```

    )
  )
}
}
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)
dinodata <- read.csv("data/dino_data.csv")
```

Appendix B

The Second Appendix, for Fun

References

- Angel, E. (2001a) *Batch-file computer graphics : A bottom-up approach with QuickTime*. Boston, MA, Wesley Addison Longman.
- Angel, E. (2000) *Interactive computer graphics : A top-down approach with OpenGL*. Boston, MA, Addison Wesley 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*. New York, Wiley. pp. 265–283.