Characterizing Heme Pockets

Universitat Autònoma de Barcelona

Patrick Finnerty

Biosciences

Universtat Autonoma de Barcelona

A thesis submitted for the degree of

Master of Science

Fall 2021

For those who gave me inspiration but now give never more

Acknowledgements

In case anyone reads this in the future, some context may be appreciated: I attended and completed this Master's during the COVID-19 global pandemic from September 2020 to September 2021.

Thanks professors

Thanks lab

Thanks UAB

Thanks Spain, and Catalonia, allowing me in and then also having public health measures unlike Donny's America

Thanks classmates

Thanks fam, friends

Thanks to the media and the creators of media that facilitated the survival of my sanity through the pandemic.

Finally, I'd like to quote a well-known artist from California. He was referencing his own work, but I wholly identify with his appreciation for the subject of his esteem:

"Last but not least, I wanna thank me. I wanna thank me for believing in me. I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for, for never quitting. I wanna thank me for always being a giver, and trying to give more than I receive. I wanna thank me for trying to do more right than wrong. I wanna thank me for just being me at all times." – Calvin Cordozar Broadus Jr.

Abstract

Metalloproteins compose approximately 40 percent (look up how to do percents in latex) of all known proteins, and use some metallic group to accomplish their chemistry. One such metallic group is heme. Heme is a member of the porphyrin family, which are able to catalyze a broad range of reactions. Heme in particular catalyzes many different reactions and is present in many proteins. However, the underlying structural requirements to host heme in a protein are not well studied.

In this study, all heme or heme-c containing proteins as of xx were downloaded and processed in order to determine underlying structural characteristics these proteins may have in common. Parameters that were examined include: xx. Overall, we found: xx. These results may have implications for protein engineering; or if I fucked up this illustrates the difficulty of the field and demonstrate the wide range of acceptable environments of heme; it may therefore be more appropriate to take a more hands-on approach until perhaps other computational methods evolve to better examine structure-function relationships.

See? Not so bad of a worst-case scenario. Just, an unusual sentiment to see in modern science.

Contents

Contents

List of Figures

List of Tables

List of Abbreviations

1-D, 2-D . . . One- or two-dimensional, referring **in this thesis** to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

Lay Summary

I investigated how heme, a molecule involved in many biological processes, binds to proteins. I did this by...

Introduction

Proteins may catalyze reactions, and many require ligands to enable their chemistry. A significant portion of all proteins, approximately 40%, require a metallic group as a ligand in order to function correctly - these proteins are known as metalloproteins.

One of these metallic groups is heme. Heme is a member of the porphyrin family, a group of molecules capable of catalyzing a broad range of reactions. Heme can catalyze many different reactions and is present in many proteins. However, the underlying structural requirements to host heme in a protein are not well understood. [MAY ADD CITATIONS]

There have only been a handful of studies dedicated to understanding the structure-chemical relationship between heme and the proteins that use heme for their chemistry (these proteins are known as hemoproteins).

In the most significant previous work, approximately ~125 hemoproteins were studied(Li et al. [1]). Although pdbs were thoroughly examined and the datasets were culled, the sample size of this study is very small compared to the amount of hemoproteins available in the pdb a decade later (~10,000 HEM-containing proteins and xx). The dataset is also limited in that there is a somewhat homogenous group of proteins examined (?). The characteristics examined were limited to: xx.

It is hypothesized that the following characteristics all have an impact on the binding of heme and function of the hemoprotein: XXXXXXXXX.

In this study, some of these characteristics were examined. They include: XX. The remainder are thus far not feasible to calculate.

All of these characteristics have implications in the field of protein engineering or basic research into hemoproteins. Examples of the uses of these results include [SUPER BLOOD STUDY] and [OTHER PROTEIN ENGINEERING STUFF]. Not

sure how much we can reference those other papers besides doing that besides in the conclusion.

Notable results from some of the prior studies include: xx and xx. These characteristics are also examined in this dataset, while some are not due to different study approaches.

1
How to use

Contents

1.1   How `oxforddown` is structured

```
.
+-- index.Rmd

+-- _bookdown.yml

+-- 00-introduction.Rmd

|   ...

+-- 07-conclusion.Rmd

+-- front-and-back-matter
```

```
|   +-- _abstract.Rmd
|   +-- 98-appendices.Rmd
|   ...
+-- bibliography
|   +-- references.bib
|   ...
+-- figures
|   ...
+-- docs
|   +-- _main.pdf
|   ...
+-- scripts_and_filters
|   +-- knit-functions.R
|   ...
+-- templates
|   +-- template.tex
|   ...
```

1.1.1   index.Rmd: metadata and layout options

In index.Rmd, set your thesis' basic metadata (e.g., title, author name)

```
title: |
  `oxforddown`: \
  An Oxford University Thesis \
  Template for R Markdown
author: Author Name
college: Your College
```

Also set filepath(s) to your abstract, acknowledgements, abbreviations, and bibliography (one or more **.bib** files):

1. How to use

```
abstract: |
  `r paste(readLines("front-and-back-matter/_abstract.Rmd"), collapse = '\n  ')`
acknowledgements: |
  `r paste(readLines("front-and-back-matter/_acknowledgements.Rmd"), collapse = '\n
dedication: For Yihui Xie
abbreviations: |
  `r paste(readLines("front-and-back-matter/_abbreviations.Rmd"), collapse = '\n

#####################
## bibliography path ##
#####################
bibliography: [bibliography/references.bib, bibliography/additional-references.bib]
```

Finally, **index.Rmd** is also where you customise layout options. For example,
in PDF output what should the heading for the bibliography section say? How
should page numbers be positioned? Should line numbers be shown? In HTML
output, what CSS files should be used for styling?

```
### citation and bibliography style ###
bibliography-heading-in-pdf: Works Cited
...

### position of page numbers ###
ordinary-page-number-foot-or-head: foot #'foot' puts page number in footer, 'head
ordinary-page-number-position: C
...

includeline-num: false #show line numbering in PDF?
...

  bookdown::bs4_book:
```

```
css:
  - templates/bs4_style.css
  - templates/corrections.css # remove to stop highlighting corrections
```

1.1.2 other .Rmd files in root folder: thesis chapters

- each chapter of your thesis should have its own **.Rmd** file in the root directory

- when you knit **index.Rmd**, these chapters are merged together in alphabetical order, based on their filenames

1.1.3 front-and-back-matter/

- this folder holds the front and back matter of your thesis

- it has **.Rmd** files for your abstract, acknowledgements, abbreviations, and a welcome note that is included in HTML output. Note how these files start with an underscore (e.g. **_abstract.Rmd**). This means they will not automatically be merged into the thesis – they are explicitly included in **index.Rmd**

- **98-appendices** and **99-references.Rmd** are automatically merged into thesis, however - therefore their file names start with a high number, so that they will be included by the very end (merging is done alphabetically)

- **99-references.Rmd** sole purpose is to set the heading for the references section in HTML and Word output

1.1.4 _bookdown.yml: build options

- Set output directory for your thesis files (**docs/** is the default, as it makes it easy to publish HTML output on GitHub pages)

- Should R Markdown automatically merge **.Rmd** files in alphabetical order? Alternatively, specify explicitly which files should be included.

1.1.5 scripts-and-filters

- **knit-function.R** has the functions that are used when you build the entire thesis by knitting **index.Rmd**

- **create_chunk_options.R** lets you include cute quotes at the start of a chapter in PDF output

- **colour_and_highlight.lua** lets you color text or apply background color to text

1.1.6 templates

- **template.tex** is the LaTeX template used to build the entire thesis to PDF in the OxThesis layout (relies on **ociamthesis.cls**)

- **brief-template.tex** is the LaTeX template used to build a single chapter to PDF in the OxThesis layout (relies on **ociamthesis.cls**)

- **beltcrest.pdf**: the oxford logo used on the front page of the PDF output

1.2 Building your entire thesis

- Build the entire thesis by opening **index.Rmd** and clicking the 'knit' button.
- The generated thesis files are saved in the **docs/** folder
- To choose output formats, go to the top of **index.Rmd**'s YAML header and edit the line `thesis_formats <- "pdf";` to the format(s) you want (options are "pdf", "bs4", "gitbook", and "word")
- You can build to multiple formats simultaneously with, e.g., `thesis_formats <- c("pdf", "bs4", "word")`
- If you want to customise the build function, edit **scripts_and_filters/knit-functions.R**

1. How to use

PDF output

```
knit: (function(input, ...) {
    thesis_formats <- "pdf";
    ...
```

When you build the entire thesis to PDF, Latex generates a whole bunch of auxillary files - these are automatically removed after the build process end by the custom knit function that is used when you knit **index.Rmd**.

To change how this removal is done, edit **scripts_and_filters/knit-functions.R**. The line `file.remove(list.files(pattern = "*\\.(log|mtc\\d*|maf|aux|bcf|lof|lot|out|t` within `if ("pdf" %in% output_format){` is the one that removes files after PDF output is generated.

BS4 book output (HTML)

```
knit: (function(input, ...) {
    thesis_formats <- "bs4";
    ...
```

- NOTE: the bs4 book output requires the `downlit` and `bslib` R packages (install them with `install.packages`)
- Note also that to deploy a BS4 book on GitHub Pages, there must be a **.nojekyll** file in the **docs/** folder, otherwise GitHub does some voodoo that causes some filepaths not to work. This file is generated automatically by `oxforddown`s knitting function.

Gitbook output (HTML)

```
knit: (function(input, ...) {
    thesis_formats <- "gitbook";
    ...
```

- Note that to deploy a gitbook on GitHub Pages, there must be a **.nojekyll** file in the **docs/** folder, otherwise GitHub does some voodoo that causes some filepaths not to work. This file is generated automatically by `oxforddown`s knitting function.

Word output

```
knit: (function(input, ...) {
    thesis_formats <- "word";
    ...
```

- Note that the Word output has no templates behind it, and many things do not work (e.g. image rotation, highlighting corrections). **I encourage pull requests that optimise the Word output, e.g. by using tools from the `officer` package.**

1.3   Building a single chapter

To knit an individual chapter without compiling the entire thesis you:

1. open the **.Rmd** file of a chapter
2. add a YAML header specifying the output format(s) (e.g. `bookdown::word_document2` for a word document you might want to upload to Google Docs for feedback from collaborators)
3. click the `knit` button (the output file is then saved in the root folder)

As shown in the sample chapters' YAML headers, to output a single chapter to PDF, use e.g.:

```
output:
  bookdown::pdf_document2:
    template: templates/brief_template.tex
    citation_package: biblatex
```

```
documentclass: book
bibliography: references.bib
```

The file **templates/brief_template.tex** formats the chapter in the OxThesis style but without including the front matter (table of contents, abstract, etc).

*Neque porro quisquam est qui dolorem ipsum quia
dolor sit amet, consectetur, adipisci velit. . .*

*There is no one who loves pain itself, who seeks after
it and wants to have it, simply because it is pain. . .*

— Cicero's *de Finibus Bonorum et Malorum.*

2

R Markdown basics

Contents

Here is a brief introduction to using *R Markdown. Markdown* is a simple formatting

syntax for authoring HTML, PDF, and MS Word documents and much, much more. *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.

2.1 Basic markdown syntax

2.1.1 Whitespace

Be careful with your spacing. 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).

2.1.2 Italics and bold

- *Italics* are done like *this* or _this_
- **Bold** is done like **this** or ___this___
- ***Bold and italics*** is done like ***this***, ____this____, or (the most transparent solution, in my opinion) **_this_**

2.1.3 Inline code

- `Inline code` is created with backticks like `this`

2.1.4 Sub and superscript

Sub$_2$ and super^2 script is created like this~2~ and this^2^

2.1.5 Strikethrough

- ~~Strikethrough~~ is done ~~like this~~

2.1.6   'Escaping' (aka "What if I need an actual asterisk?")

- To include an actual *, _ or \, add another \ in front of them: *, _, \\

2.1.7   Endash (–), emdash (—)

- – and — with -- and ---

2.1.8   Blockquotes

Do like this:

> Put a > in front of the line.

2.1.9   Headings

Section headers are created with #'s of increasing number, i.e.

- # First-level heading
- ## Second-level heading
- ### Etc.

In PDF output, a level-five heading will turn into a paragraph heading, i.e. `\paragraph{My level-five heading}`, which appears as bold text on the same line as the subsequent paragraph.

2.1.10   Lists

Unordered list by starting a line with an * or a -:

- Item 1
- Item 2

Ordered lists by starting a line with a number. Notice that you can mislabel the numbers and *Markdown* will still make the order right in the output:

1. Item 1

2. Item 2

To create a sublist, indent the values a bit (at least four spaces or a tab):

1. Item 1
2. Item 2
3. Item 3

    - Item 3a
    - Item 3b

2.1.11   Line breaks

The official *Markdown* way to create line breaks is by ending a line with more than two spaces.

Roses are red. Violets are blue.

This appears on the same line in the output, because we didn't add spaces after red.

Roses are red.
Violets are blue.

This appears with a line break because I added spaces after red.

I find this is confusing, so I recommend the alternative way: Ending a line with a backslash will also create a linebreak:

Roses are red.
Violets are blue.

To create a new paragraph, you put a blank line.

Therefore, this line starts its own paragraph.

2.1.12   Hyperlinks

- This is a hyperlink created by writing the text you want turned into a clickable link in `[square brackets followed by a](https://hyperlink-in-parentheses)`

2.1.13 Footnotes

- Are created[1] by writing either ^[my footnote text] for supplying the footnote content inline, or something like [^a-random-footnote-label] and supplying the text elsewhere in the format shown below [2]:

  [^a-random-footnote-label]: This is a random test.

2.1.14 Comments

To write comments within your text that won't actually be included in the output, use the same syntax as for writing comments in HTML. That is, <!-- this will not be included in the output -->.

2.1.15 Math

The syntax for writing math is stolen from LaTeX. To write a math expression that will be shown **inline**, enclose it in dollar signs. - This: $A = \pi*r^{2}$ Becomes: $A = \pi * r^2$

To write a math expression that will be shown in a block, enclose it in two dollar signs.

This: $$A = \pi*r^{2}$$

Becomes:

$$A = \pi * r^2$$

To create numbered equations, put them in an 'equation' environment and give them a label with the syntax (\#eq:label), like this:

```
\begin{equation}
  f\left(k\right) = \binom{n}{k} p^k\left(1-p\right)^{n-k}
  (\#eq:binom)
\end{equation}
```

---

[1]my footnote text

[2]This is a random test.

Becomes:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2.1}$$

For more (e.g. how to theorems), see e.g. the documentation on bookdown.org

2.2 Executable code chunks

The magic of R Markdown is that we can add executable code within our document to make it dynamic.

We do this either as *code chunks* (generally used for loading libraries and data, performing calculations, and adding images, plots, and tables), or *inline code* (generally used for dynamically reporting results within our text).

The syntax of a code chunk is shown in Figure 2.1.
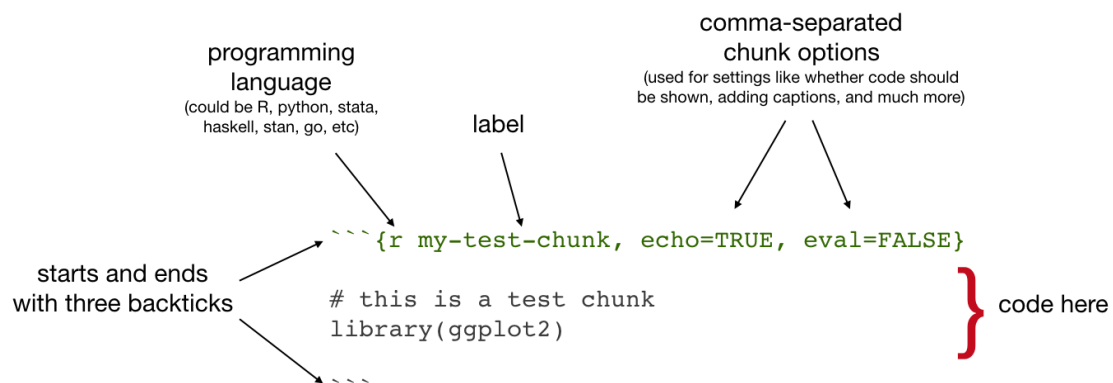
Figure 2.1: Code chunk syntax

Common chunk options include (see e.g. bookdown.org):

- `echo`: whether or not to display code in knitted output
- `eval`: whether or to to run the code in the chunk when knitting
- `include`: whether to include anything from the from a code chunk in the output document
- `fig.cap`: figure caption
- `fig.scap`: short figure caption, which will be used in the 'List of Figures' in the PDF front matter

IMPORTANT: Do *not* use underscoores in your chunk labels - if you do, you are likely to get an error in PDF output saying something like "! Package caption Error: \caption outside float".

2.2.1 Setup chunks - setup, images, plots

An R Markdown document usually begins with a chunk that is used to **load libraries**, and to **set default chunk options** with `knitr::opts_chunk$set`.

In your thesis, this will probably happen in **index.Rmd** and/or as opening chunks in each of your chapters.

```
```{r setup, include=FALSE}
# don't show code unless we explicitly set echo = TRUE
knitr::opts_chunk$set(echo = FALSE)


library(tidyverse)
```
```

2.2.2 Including images

Code chunks are also used for including images, with `include_graphics` from the `knitr` package, as in Figure 2.2

```
knitr::include_graphics("figures/sample-content/beltcrest.png")
```

Useful chunk options for figures include:

- `out.width` (use with a percentage) for setting the image size
- if you've got an image that gets waaay to big in your output, it will be constrained to the page width by setting `out.width = "100%"`

Figure rotation

You can use the chunk option `out.extra` to rotate images.

The syntax is different for LaTeX and HTML, so for ease we might start by assigning the right string to a variable that depends on the format you're outputting to:

Figure 2.2: Oxford logo

```r
if (knitr::is_latex_output()){
  rotate180 <- "angle=180"
} else {
  rotate180 <- "style='transform:rotate(180deg);'"
}
```

Then you can reference that variable as the value of `out.extra` to rotate images, as in Figure 2.3.

2.2.3 Including plots

Similarly, code chunks are used for including dynamically generated plots. You use ordinary code in R or other languages - Figure 2.4 shows a plot of the `cars` dataset of stopping distances for cars at various speeds (this dataset is built in to **R**).

```r
cars %>%
  ggplot() +
    aes(x = speed, y = dist) +
    geom_point()
```
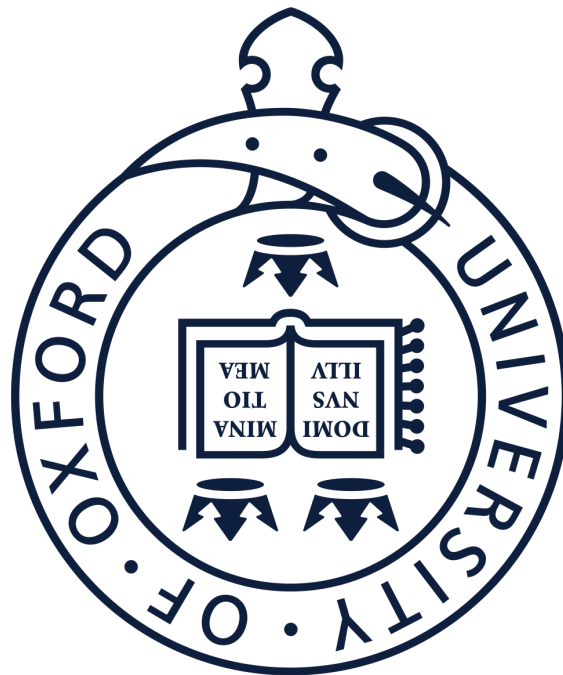
Figure 2.3: Oxford logo, rotated
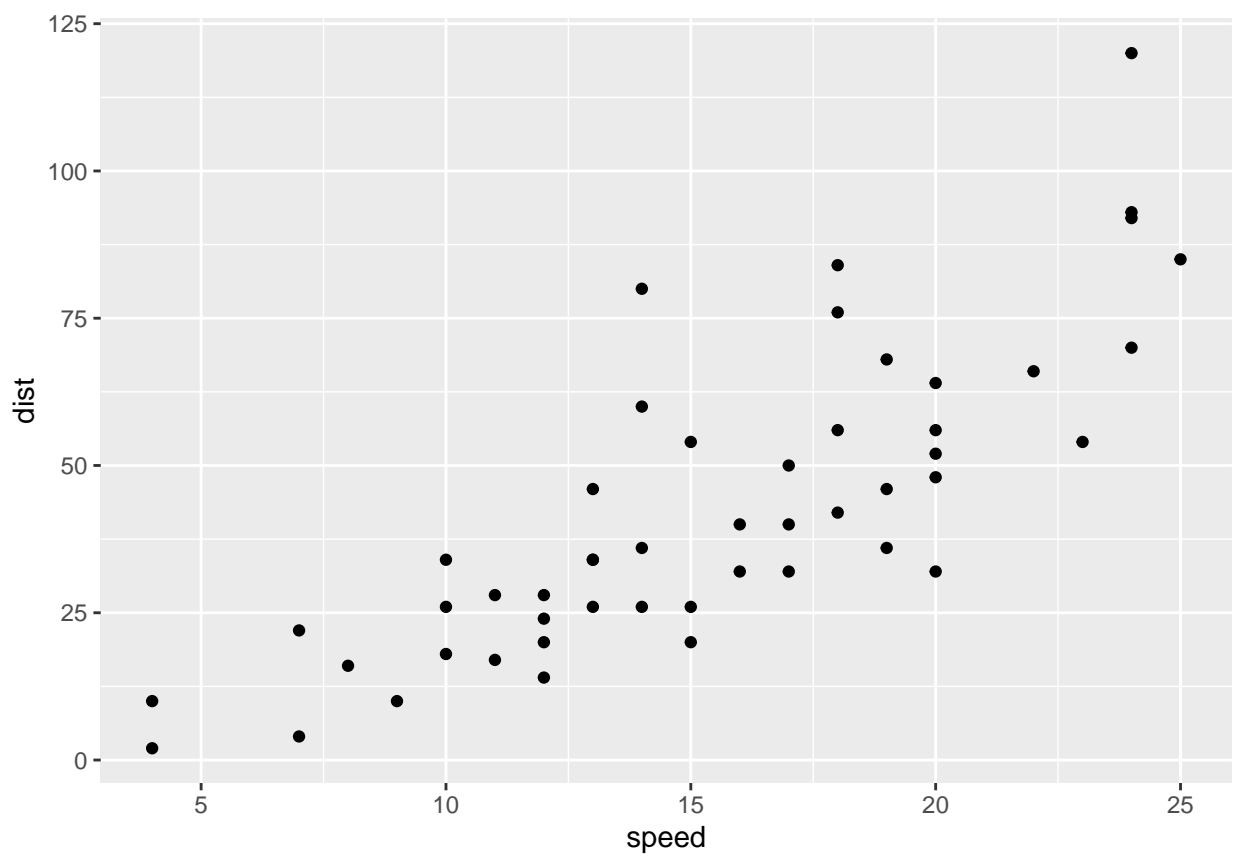
Figure 2.4: A ggplot of car stuff

Table 2.1: A knitr kable table

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10

Under the hood, plots are included in your document in the same way as images - when you build the book or knit a chapter, the plot is automatically generated from your code, saved as an image, then included into the output document.

2.2.4 Including tables

Tables are usually included with the `kable` function from the `knitr` package.

Table 2.1 shows the first rows of that cars data - read in your own data, then use this approach to automatically generate tables.

```r
cars %>%
  head() %>%
  knitr::kable(caption = "A knitr kable table")
```

- Gotcha: when using `kable`, captions are set inside the `kable` function
- The `kable` package is often used with the `kableExtra` package

2.2.5 Control positioning

One thing that may be annoying is the way *R Markdown* handles "floats" like tables and figures. In your PDF output, LaTeX will try to find the best place to put your object based on the text around it and until you're really, truly done writing you should just leave it where it lies.

In general, you should allow LaTeX to do this, but if you really *really* need a figure to be positioned where you put in the document, then you can make LaTeX attempt to do this with the chunk option `fig.pos="H"`, as in Figure 2.5:

```
knitr::include_graphics("figures/sample-content/beltcrest.png")
```


Figure 2.5: An Oxford logo that LaTeX will try to place at this position in the text

As anyone who has tried to manually play around with the placement of figures in a Word document knows, this can have lots of side effects with extra spacing on other pages, etc. Therefore, it is not generally a good idea to do this - only do it when you really need to ensure that an image follows directly under text where you refer to it (in this document, I needed to do this for Figure 4.1 in section 4.1.4). For more details, read the relevant section of the R Markdown Cookbook.

2.3   Executable inline code

'Inline code' simply means inclusion of code inside text. The syntax for doing this is `r R_CODE` For example, `r 4 + 4` will output 8 in your text.

You will usually use this in parts of your thesis where you report results - read in data or results in a code chunk, store things you want to report in a variable, then insert the value of that variable in your text. For example, we might assign the number of rows in the `cars` dataset to a variable:

```r
num_car_observations <- nrow(cars)
```

We might then write:

"In the `cars` dataset, we have `` `r num_car_observations` `` observations."

Which would output:

"In the `cars` dataset, we have 50 observations."

2.4 Executable code in other languages than R

If you want to use other languages than R, such as Python, Julia C++, or SQL, see the relevant section of the *R Markdown Cookbook*

3

Citations, cross-references, and collaboration

Contents

3.1 Citations

The usual way to include citations in an *R Markdown* document is to put references in a plain text file with the extension **.bib**, in **BibTex** format.[1] Then reference the path to this file in **index.Rmd**'s YAML header with `bibliography: example.bib`.

---

[1] The bibliography can be in other formats as well, including EndNote (**.enl**) and RIS (**.ris**), see rmarkdown.rstudio.com/authoring_bibliographies_and_citations.

Most reference managers can create a .bib file with you references automatically. However, the **by far** best reference manager to use with *R Markdown* is Zotero with the Better BibTex plug-in, because the `citr` plugin for RStudio (see below) can read references directly from your Zotero library!

Here is an example of an entry in a **.bib** file:

```
@article{Shea2014,
  author =        {Shea, Nicholas and Boldt, Annika},
  journal =       {Trends in Cognitive Sciences},
  pages =         {186--193},
  title =         {{Supra-personal cognitive control}},
  volume =        {18},
  year =          {2014},
  doi =           {10.1016/j.tics.2014.01.006},
}
```

In this entry highlighed section, 'Shea2014' is the **citation identifier**. To default way to cite an entry in your text is with this syntax: `[@citation-identifier]`.

So I might cite some things [**Shea2014**, **Lottridge2012**].

3.1.1   PDF output

In PDF output, the bibliography is handled by the OxThesis LaTeX template. If you set `bib-humanities: true` in **index.Rmd**, then in-text references will be formatted as author-year; otherwise references will be shown as numbers.

If you choose author-year formatting, a number of variations on the citation syntax are useful to know:

- Put author names outside the parenthesis

  – This: `@Shea2014 says blah.`
  – Becomes: **Shea2014** says blah.

- Include only the citation-year (in parenthesis)

- This: `Shea et al. says blah [-@Shea2014]`
- Becomes: Shea et al. says blah [**Shea2014**]

- Add text and page or chapter references to the citation

  - This: `[see @Shea2014, pp. 33-35; also @Wu2016, ch. 1]`
  - Becomes: Blah blah [**Shea2014**, **Wu2016**].

3.1.2 Gitbook output

In gitbook output, citations are by default inserted in the Chicago author-date format.

To change the format, add `csl: some-other-style.csl` in **index.Rmd**'s YAML header. You can browse through and download styles at zotero.org/styles.
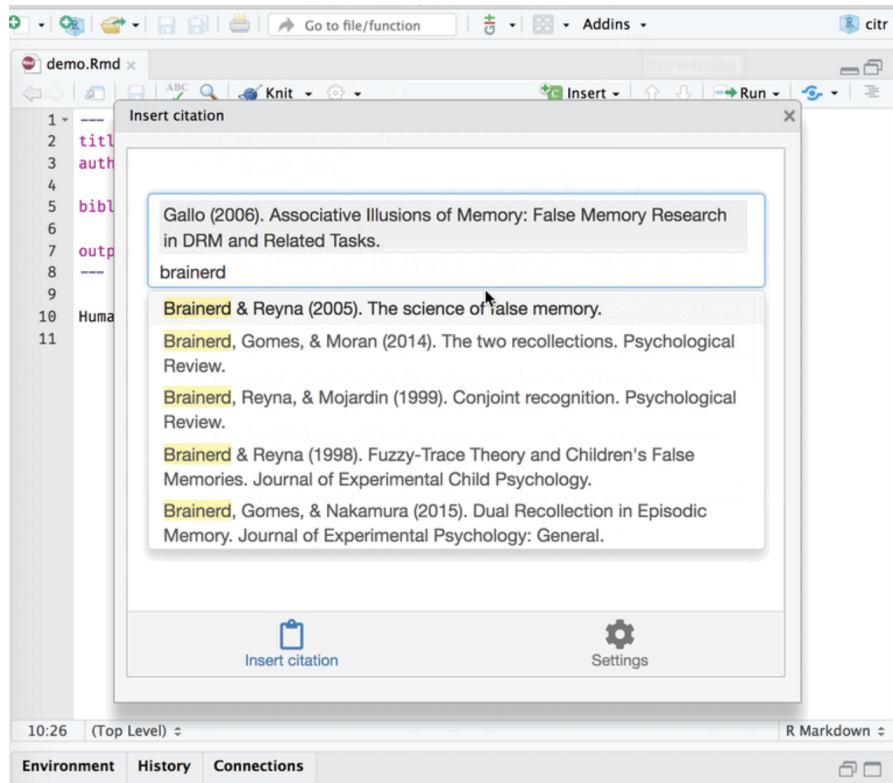
Figure 3.1: The 'citr' add-in

3.1.3 Insert references easily with the `citr` add-in

For an easy way to insert citations, try the `citr` RStudio add-in (Figure 3.1). You can install this add-in by typing `install.packages("citr")` in the R Console.

3.2 Cross-referencing

We can make cross-references to **sections** within our document, as well as to **figures** (images and plots) and **tables**.

The general cross-referencing syntax is `\@ref(label)`

3.2.1 Section references

Headers are automatically assigned a reference label, which is the text in lower caps separated by dashes. For example, `# My header` is automatically given the label `my-header`. So `# My header` can be referenced with `\@ref(my-section)`

Remember what we wrote in section 3.1?

We can also use **hyperlink syntax** and add # before the label, though this is only guaranteed to work properly in HTML output:

- So if we write `Remember what we wrote up in [the previous section](#citations)`?
- It becomes Remember what we wrote up in the previous section?

Creating custom labels

It is a very good idea to create **custom labels** for our sections. This is because the automatically assigned labels will change when we change the titles of the sections - to avoid this, we can create the labels ourselves and leave them untouched if we change the section titles.

We create custom labels by adding `{#label}` after a header, e.g. `# My section {#my-label}`. See our chapter title for an example. That was section 3.

3.2.2 Figure (image and plot) references

- To refer to figures (i.e. images and plots) use the syntax `\@ref(fig:label)`
- **GOTCHA**: Figures and tables must have captions if you wish to cross-reference them.

Let's add an image:

```
knitr::include_graphics("figures/sample-content/captain.jpeg")
```

We refer to this image with `\@ref(fig:captain)`. So Figure 3.2 is this image. And in Figure 2.4 we saw a cars plot.

3.2.3 Table references

- To refer to tables use the syntax `\@ref(tab:label)`

Let's include a table:

Figure 3.2: A marvel-lous meme

Table 3.1: Stopping cars

speed	dist
4	2
4	10
7	4
7	22
8	16

```
knitr::kable(cars[1:5,],

            caption="Stopping cars")
```

We refer to this table with `\@ref(tab:cars-table2)`. So Table 3.1 is this table.

And in Table 2.1 we saw more or less the same cars table.

3.2.4   Including page numbers

Finally, in the PDF output we might also want to include the page number of a reference, so that it's easy to find in physical printed output. LaTeX has a command for this, which looks like this: `\pageref{fig/tab:label}` (note: curly

braces, not parentheses)

When we output to PDF, we can use raw LaTeX directly in our .Rmd files. So if we wanted to include the page of the cars plot we could write:

- This: `Figure \@ref(fig:cars-plot) on page \pageref(fig:cars-plot)`
- Becomes: Figure 2.4 on page 20

Include page numbers only in PDF output

A problem here is that LaTeX commands don't display in HTML output, so in the gitbook output we'd see simply "Figure 2.4 on page".

One way to get around this is to use inline R code to insert the text, and use an `ifelse` statement to check the output format and then insert the appropriate text.

- So this: `` `r ifelse(knitr::is_latex_output(), "Figure \\@ref(fig:cars-plot) on page \\pageref{fig:cars-plot}", "")` ``
- Inserts this (check this on both PDF and gitbook): Figure 2.4 on page 20

Note that we need to escape the backslash with another backslash here to get the correct output.

3.3   Collaborative writing

Best practices for collaboration and change tracking when using R Markdown are still an open question. In the blog post **One year to dissertate** by Lucy D'Agostino, which I highly recommend, the author notes that she knits .Rmd files to a word document, then uses the `googledrive` R package to send this to Google Drive for comments / revisions from co-authors, then incorporates Google Drive suggestions *by hand* into the .Rmd source files. This is a bit clunky, and there are ongoing discussions among the *R Markdown* developers about what the best way is to handle collaborative writing (see issue #1463 on GitHub, where CriticMarkup is among the suggestions).

For now, this is an open question in the community of R Markdown users. I often knit to a format that can easily be imported to Google Docs for comments, then go over suggested revisions and manually incorporate them back in to the .Rmd source files. For articles, I sometimes upload a near-final draft to Overleaf, then collaboratively make final edits to the LaTeX file there. I suspect some great solution will be developed in the not-to-distant future, probably by the RStudio team.

3.4   Additional resources

- *R Markdown: The Definitive Guide* - `https://bookdown.org/yihui/rmarkdown/`

- *R for Data Science* - `https://r4ds.had.co.nz`

4

Tables

Contents

4.1 Making LaTeX tables play nice

Dealing with tables in LaTeX can be painful. This section explains the main tricks you need to make the pain go away.

(Note: if you are looking at the ebook version, you will not see much difference in this section, as it is only relevant for PDF output!)

4.1.1 Making your table pretty

When you use `kable` to create tables, you will almost certainly want to set the option `booktabs = TRUE`. This makes your table look a million times better:

```
library(knitr)

library(tidyverse)


head(mtcars) %>%

  kable(booktabs = TRUE)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Compare this to the default style, which looks terrible:

```
head(mtcars) %>%

  kable()
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

4.1.2 If your table is too wide

You might find that your table expands into the margins of the page, like the tables above. Fix this with the `kable_styling` function from the kableExtra package:

```
library(kableExtra)


head(mtcars) %>%

  kable(booktabs = TRUE) %>%

  kable_styling(latex_options = "scale_down")
```

This scales down the table to fit the page width.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

4.1.3   If your table is too long

If your table is too long to fit on a single page, set `longtable = TRUE` in the `kable`
function to split the table across multiple pages.

```
a_long_table <- rbind(mtcars, mtcars)
```

```
a_long_table %>%

  select(1:8) %>%

  kable(booktabs = TRUE, longtable = TRUE)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1

4. Tables

Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0

4. Tables

Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1

When you do this, you'll probably want to make the header repeat on new pages.

Do this with the `kable_styling` function from `kableExtra`:

```
a_long_table %>%

  kable(booktabs = TRUE, longtable = TRUE) %>%

  kable_styling(latex_options = "repeat_header")
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

Unfortunately, we cannot use the `scale_down` option with a `longtable`. So if a `longtable` is too wide, you can either manually adjust the font size, or show the table in landscape layout. To adjust the font size, use kableExtra's `font_size` option:

4. Tables

```
a_long_table %>%

  kable(booktabs = TRUE, longtable = TRUE) %>%

  kable_styling(font_size = 9, latex_options = "repeat_header")
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4

4. Tables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

To put the table in landscape mode, use kableExtra's `landscape` function:

```
a_long_table %>%

  kable(booktabs = TRUE, longtable = TRUE) %>%

  kable_styling(latex_options = "repeat_header") %>%

  landscape()
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

4.1.4   Max power: manually adjust the raw LaTeX output

For total flexibility, you can adjust the raw LaTeX output from `kable`/`kableExtra` that generates the table. Let us consider how we would do this for the example of adjusting the font size if our table is too wide: Latex has a bunch of standard commands that set an approximate font size, as shown below in Figure 4.1.

\tiny	Lorem ipsum
\scriptsize	Lorem ipsum
\footnotesize	Lorem ipsum
\small	Lorem ipsum

Figure 4.1: Font sizes in LaTeX

You could use these to manually adjust the font size in your longtable in two steps:

1. Wrap the longtable environment in, e.g., a `scriptsize` environment, by doing a string replacement in the output from `kable`/`kableExtra`
2. Add the attributes that make R Markdown understand that the table is a table (it seems R drops these when we do the string replacement)

```
our_adjusted_table <- a_long_table %>%
  kable(booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(latex_options = "repeat_header") %>%
  # wrap the longtable in a tiny environment
  str_replace('\\\\begin\\{longtable\\}',
              '\\\\begin\\{scriptsize\\}\n\\\\begin\\{longtable\\}') %>%
  str_replace('\\\\end\\{longtable\\}',
              '\\\\end\\{longtable\\}\n\\\\end\\{scriptsize\\}')
```

4. Tables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1

4. Tables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

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

— Charles Darwin [**Darwin1859**]

5

Customisations and extensions

Contents

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

5.1 Front matter

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

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

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

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

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

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

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

5.2   Shorten running header (PDF)

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

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

For example, chapter 3's running header is simply 'Cites and cross-refs', because it begins like this:

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

5.3   Unnumbered chapters

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

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

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

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

5.4   Beginning chapters with quotes (PDF)

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

---

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

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

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

5.5   Highlighting corrections (HTML & PDF)

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

5.5.1   Short, inline corrections

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

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

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

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

5.5.2   Blocks of added or changed material

Highlight entire **blocks of added or changed material** by putting them in a block

of type `correction`, using the syntax ```` ```{block type='correction'}.````[2] Like so:

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

Note that correction blocks cannot be included in word output.

5.5.3   Stopping corrections from being highlighted

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

- PDF output: set `corrections: false`

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

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

The lua filter that adds the functionality to highlight corrections adds two more

tricks: you can apply your own choice of colour to highlight text, or change the

font color. The syntax is as follows:

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

---

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

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

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

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

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

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

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

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


# grab the pages
```

5. Customisations and extensions

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


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


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

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

Ulrik Lyngs[1]**, Kai Lukoff**[2]**, Petr Slovak**[3]**, William Seymour**[1]**, Helena Webb**[1]**,**
Marina Jirotka[1]**, Jun Zhao**[1]**, Max Van Kleek**[1]**, Nigel Shadbolt**[1]
[1]Department of Computer Science, University of Oxford, UK, {first.last}@cs.ox.ac.uk
[2]Human Centered Design & Engineering, University of Washington, Seattle, US, kai1@uw.edu
[3]Department of Informatics, King's College London, UK, petr.slovak@kcl.ac.uk

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

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

CCS Concepts
•**Human-centered computing → Empirical studies in HCI;**

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

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

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

RELATED WORK

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

5. Customisations and extensions

54

REFERENCES

[1] Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless Computing: Designing Technologies to Subtly Influence Behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 719–730. DOI: http://dx.doi.org/10.1145/2750858.2805843

[2] Sami Abdo Radman Al-Dubai, Kurubaran Ganasegeran, Mustafa Ahmed Mahdi Al-Shagga, Hematram Yadav, and John T. Arokiasamy. 2013. Adverse Health Effects and Unhealthy Behaviors among Medical Students Using Facebook. https://www.hindawi.com/journals/tswj/2013/465161/. (2013). DOI: http://dx.doi.org/10.1155/2013/465161

[3] All Party Parliamentary Group on Social Media and Young People's Mental Health and Wellbeing. 2019. *#NewFilters to Manage the Impact of Social Media on Young People's Mental Health and Wellbeing*. Technical Report. UK Parliament.

[4] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2019. *The Welfare Effects of Social Media*. Working Paper 25514. National Bureau of Economic Research. DOI: http://dx.doi.org/10.3386/w25514

[5] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Staale Pallesen. 2012. Development of a Facebook Addiction Scale. *Psychological Reports* 110, 2 (apr 2012), 501–517. DOI: http://dx.doi.org/10.2466/02.09.18.PR0.110.2.501-517

[6] Yummy Apps. 2019. Todobook. (May 2019).

[7] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37, 2 (1982), 122–147. DOI: http://dx.doi.org/10.1037/0003-066x.37.2.122

[8] Fanni Bányai, Ágnes Zsila, Orsolya Király, Aniko Maraz, Zsuzsanna Elekes, Mark D. Griffiths, Cecilie Schou Andreassen, and Zsolt Demetrovics. 09-Jan-2017. Problematic Social Media Use: Results from a Large-Scale Nationally Representative Adolescent Sample. *PLOS ONE* 12, 1 (09-Jan-2017), e0169839. DOI: http://dx.doi.org/10.1371/journal.pone.0169839

[9] Elliot T Berkman, Cendri A Hutcherson, Jordan L Livingston, Lauren E Kahn, and Michael Inzlicht. 2017. Self-Control as Value-Based Choice. *Current Directions in Psychological Science* 26, 5 (2017), 422–428. DOI: http://dx.doi.org/10.1177/0963721417704394

[10] Walter R. Boot, Daniel J. Simons, Cary Stothart, and Cassie Stutts. 2013. The Pervasive Problem with Placebos in Psychology. *Perspectives on Psychological Science* 8, 4 (jul 2013), 445–454. DOI: http://dx.doi.org/10.1177/1745691613491271

[11] Amara Brook. 2011. Ecological Footprint Feedback: Motivating or Discouraging? *Social Influence* 6, 2 (April 2011), 113–128. DOI: http://dx.doi.org/10.1080/15534510.2011.566801

[12] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual Review of Economics* 2, 1 (Sept. 2010), 671–698. DOI: http://dx.doi.org/10.1146/annurev.economics.102308.124324

[13] Moira Burke and Robert E. Kraut. 2016. The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication* 21, 4 (2016), 265–281. DOI: http://dx.doi.org/10.1111/jcc4.12162

[14] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social Network Activity and Social Well-Being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1909–1912. DOI: http://dx.doi.org/10.1145/1753326.1753613

[15] Wenhong Chen and Kye-Hyoung Lee. 2013. Sharing, Liking, Commenting, and Distressed? The Pathway between Facebook Interaction and Psychological Distress. *Cyberpsychology, Behavior and Social Networking* 16, 10 (oct 2013), 728–734. DOI: http://dx.doi.org/10.1089/cyber.2012.0272

[16] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 199:1–199:13. DOI: http://dx.doi.org/10.1145/3290605.3300429

[17] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. DOI: http://dx.doi.org/10.1037/0033-2909.112.1.155

[18] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1389–1397. DOI: http://dx.doi.org/10.1145/2851581.2892410

[19] Helen Creswick, Liz Dowthwaite, Ansgar Koene, Elvira Perez Vallejos, Virginia Portillo, Monica Cano, and Christopher Woodard. 2019. "... They don't really listen to people". *Journal of Information, Communication and Ethics in Society* 17, 2 (May 2019), 167–182. DOI: http://dx.doi.org/10.1108/jices-11-2018-0090

[20] Angela L. Duckworth, Katherine L. Milkman, and David Laibson. 2018. Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological Science in the Public Interest* 19, 3 (Dec. 2018), 102–129. DOI: http://dx.doi.org/10.1177/1529100618821893

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

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

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

5.8.1 An example paper in another folder

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

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

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

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

5.8.2   Step 1: Include paper as a child document

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

```
# Including an external chapter


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

5.8.3   Step 2: Make file paths compatible

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

```
---

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

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

```
library(tidyverse)
library(knitr)


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

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

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

Note on HTML output

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

5.8.4 Step 3: Make sure header levels are correct

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

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

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

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

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

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

NOTE: There might be no need to use a lua filter to shift heading - it seems you could simply use `pandoc_args: ["--shift-heading-level-by=-1"]` (see https://pandoc.org/MANUAL.html#reader-options)

5.8.5 Step 4. Make sure figure widths are correct

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

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

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
  out_width: "80%"
output:
  pdf_document:
```

```
    pandoc_args: ["--lua-filter=reduce_header_level.lua"]
---
```

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

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

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

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

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

5.9   Customizing referencing

5.9.1   Using a .csl file with pandoc instead of biblatex

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

```
csl: ecology.csl
```

5.9.2   Customizing biblatex and adding chapter bibliographies

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

5. Customisations and extensions

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

(Charmantier and Gienapp 2014)

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

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

5. Customisations and extensions

```
  \clearfield{note}
}


\DeclareFieldFormat{titlecase}{\MakeTitleCase{#1}}


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


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

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

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

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

5.10 Customizing the page headers and footers (PDF)

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

5.11 Diving in to the OxThesis LaTeX template (PDF)

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

5.12 Customising to a different university

5.12.1 The minimal route

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

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

5.12.2 Replacing the entire title page with your required content

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

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

5. Customisations and extensions

**Title of your
Thesis**

John Doe

Thesis committee

Promotor:
Prof.dr J. Smith
Professor of Geo-information Science and Remote Sensing
Wageningen University

Co-promotors:
Dr. Name of co-promotor
Assistant Professor, Laboratory of Geo-information Science and Remote Sensing
Wageningen University

Other members:
Prof.dr Jury member 1, Wageningen University
Prof.dr Jury member 2, Affiliation
Prof.dr Jury member 3, Affiliation
Prof.dr Jury member 4, Affiliation

This research was conducted under the auspices of the C.T. de Wit Graduate School
of Production Ecology & Resource Conservation (PE&RC)

**Title of your
thesis**

John Doe

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Date of your defense
at 4 p.m. in the Aula.

John Doe
Title of your thesis
?? pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summary in English

ISBN XXX-YYY

For Yihui Xie

Acknowledgements

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

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

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

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

Ulrik Lyngs
Linacre College, Oxford
2 December 2018

6

Troubleshooting

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

6.1 Error: Failed to build the bibliography via biber

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

Try doing this:

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

2. restart your computer

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

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

2. set `citation_package: natbib` under

6. Troubleshooting

```
output:
  bookdown::pdf_book:
    citation_package: natbib
```

Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken.

All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.

— Johann Wolfgang von Goethe
[**von_goethe_wilhelm_1829**]

Conclusion

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

More info

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

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

7

Methods

Contents

7.1 Datasets

The primary dataset of heme-containing proteins (HEM) was composed by searching for proteins containing "HEM" in the PDB. 58 PDBs were kept after culling a much larger list. Not all PDBs were suitable for the investigation - some contain two ligands in one pocket, or two proteins per PDB (IDK HOw to say this properly rn)

A similar culling was done for the other ligands. From this work we ended up with the following dataset sizes: HEM 58, HEC . . .

Notes for each dataset.

8
Results

Trial

Here we'll attempt to use the r block of code to show a dataframe to the right. And there we are. In the future I should find if there's a more reproducible solution than just saving the global environment to a data file. Perhaps if I could just be like main.R and then call stuff. But that takes FOREVER to run, so this for now.

Residue	Freq
LEU	8
ALA	7
TYR	7
GLY	6
SER	6
ARG	5
HIS	5
PHE	5
VAL	5
GLU	4
LYS	4
ASN	3
GLN	3
ILE	3
MET	3
THR	3
ASP	2

Appendices

A

The First Appendix

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

In 02-rmd-basics-code.Rmd

```
library(tidyverse)
knitr::include_graphics("figures/sample-content/chunk-parts.png")
```

These are the HEM figures, which will likely need to be split up to reference properly.

```
load("mainData.RData")


library(dplyr)
library(data.table)

##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
## The following object is masked from 'package:purrr':
```

A. The First Appendix

```
##
##        transpose
```

```
library(tidyr)
library(ggplot2)
library(stringr)
library(knitr)
library(clipr) # Linux) sudo apt-get install xclip ... R)install.packages("clipr"
```

```
## Welcome to clipr. See ?write_clip for advisories on writing to the clipboard in
```

```
source("~/heme-binding/scripts/r/addpdbcol.R")


# this last one was tricky, see comments underneath:
library(kableExtra)
#install.packages("kableExtra")
# I had to download more stuff, in ubuntu terminal:
# sudo apt-get install libxml2-dev
# sudo apt-get install libssl-dev
# sudo apt-get install libcurl4-openssl-dev #not sure if competely necessary
# sudo apt-get install libfontconfig1-dev

# you could probably also tease out what you're missing from the error messages
# just keep trying to install kableExtra and download what you need, until it wor

#get 5A data real quick so we overwrite all DFs it uses, below:
source("~/heme-binding/scripts/r/data5A.R")
ls5A <- data5A_fn()
```

```
## [1] "1B2V.pdb.mono.pdb.processed.txt"
```

```
## [1] "1B5M.pdb.mono.pdb.processed.txt"
```

```
## [1] "1DK0.pdb.mono.pdb.processed.txt"
```

```
## [1] "1DKH.pdb.mono.pdb.processed.txt"
```

```
## [1] "1ICC.pdb.mono.pdb.processed.txt"
## [1] "1IPH.pdb.mono.pdb.processed.txt"
## [1] "1N45.pdb.mono.pdb.processed.txt"
## [1] "1P3T.pdb.mono.pdb.processed.txt"
## [1] "1QHU.pdb.mono.pdb.processed.txt"
## [1] "1QJS.pdb.mono.pdb.processed.txt"
## [1] "1SI8.pdb.mono.pdb.processed.txt"
## [1] "1SY2.pdb.mono.pdb.processed.txt"
## [1] "1U9U.pdb.mono.pdb.processed.txt"
## [1] "1VGI.pdb.mono.pdb.processed.txt"
## [1] "1ZVI.pdb.mono.pdb.processed.txt"
## [1] "2BHJ.pdb.mono.pdb.processed.txt"
## [1] "2CJ0.pdb.mono.pdb.processed.txt"
## [1] "2CN4.pdb.mono.pdb.processed.txt"
## [1] "2CPO.pdb.mono.pdb.processed.txt"
## [1] "2E2Y.pdb.mono.pdb.processed.txt"
## [1] "2FC2.pdb.mono.pdb.processed.txt"
## [1] "2IIZ.pdb.mono.pdb.processed.txt"
## [1] "2IPS.pdb.mono.pdb.processed.txt"
## [1] "2J0P.pdb.mono.pdb.processed.txt"
## [1] "2J18.pdb.mono.pdb.processed.txt"
## [1] "2O6P.pdb.mono.pdb.processed.txt"
## [1] "2Q6N.pdb.mono.pdb.processed.txt"
## [1] "2R7A.pdb.mono.pdb.processed.txt"
## [1] "2SPL.pdb.mono.pdb.processed.txt"
## [1] "2VEB.pdb.mono.pdb.processed.txt"
## [1] "3HX9.pdb.mono.pdb.processed.txt"
## [1] "3MVF.pdb.mono.pdb.processed.txt"
## [1] "3QZN.pdb.mono.pdb.processed.txt"
## [1] "3QZZ.pdb.mono.pdb.processed.txt"
```

A. The First Appendix

```
## [1] "3SIK.pdb.mono.pdb.processed.txt"
## [1] "3TGC.pdb.mono.pdb.processed.txt"
## [1] "3VP5.pdb.mono.pdb.processed.txt"
## [1] "3ZJS.pdb.mono.pdb.processed.txt"
## [1] "4B8N.pdb.mono.pdb.processed.txt"
## [1] "4CAT.pdb.mono.pdb.processed.txt"
## [1] "4CDP.pdb.mono.pdb.processed.txt"
## [1] "4I3Q.pdb.mono.pdb.processed.txt"
## [1] "4JET.pdb.mono.pdb.processed.txt"
## [1] "4MF9.pdb.mono.pdb.processed.txt"
## [1] "4MYP.pdb.mono.pdb.processed.txt"
## [1] "4NL5.pdb.mono.pdb.processed.txt"
## [1] "4UZV.pdb.mono.pdb.processed.txt"
## [1] "4XZD.pdb.mono.pdb.processed.txt"
## [1] "4Y1Q.pdb.mono.pdb.processed.txt"
## [1] "5CN5.pdb.mono.pdb.processed.txt"
## [1] "5GJ3.pdb.mono.pdb.processed.txt"
## [1] "5KZL.pdb.mono.pdb.processed.txt"
## [1] "5O1L.pdb.mono.pdb.processed.txt"
## [1] "5O1M.pdb.mono.pdb.processed.txt"
## [1] "5VEU.pdb.mono.pdb.processed.txt"
## [1] "6A2J.pdb.mono.pdb.processed.txt"
## [1] "7C74.pdb.mono.pdb.processed.txt"
## [1] "7DMR.pdb.mono.pdb.processed.txt"
## [1] "1BBH.pdb.mono.pdb.processed.txt"
## [1] "1S56.pdb.mono.pdb.processed.txt"
## [1] "1W2L.pdb.mono.pdb.processed.txt"
## [1] "2BC5.pdb.mono.pdb.processed.txt"
## [1] "2BH5.pdb.mono.pdb.processed.txt"
## [1] "3EAH.pdb.mono.pdb.processed.txt"
```

```
## [1] "3X15.pdb.mono.pdb.processed.txt"

## [1] "4B2N.pdb.mono.pdb.processed.txt"

## [1] "5KPF.pdb.mono.pdb.processed.txt"

## [1] "5LFT.pdb.mono.pdb.processed.txt"

## [1] "5T8W.pdb.mono.pdb.processed.txt"

## [1] "6VDQ.pdb.mono.pdb.processed.txt"

## [1] "6WZA.pdb.mono.pdb.processed.txt"

## [1] "6XNK.pdb.mono.pdb.processed.txt"

## [1] "1ZJ8.pdb.mono.pdb.processed.txt"

## [1] "2AKJ.pdb.mono.pdb.processed.txt"

## [1] "2AOP.pdb.mono.pdb.processed.txt"

## [1] "3BOG.pdb.mono.pdb.processed.txt"

## [1] "3VKP.pdb.mono.pdb.processed.txt"

## [1] "3VLX.pdb.mono.pdb.processed.txt"

## [1] "3VLY.pdb.mono.pdb.processed.txt"

## [1] "3VLZ.pdb.mono.pdb.processed.txt"

## [1] "5H8V.pdb.mono.pdb.processed.txt"

## [1] "1TWN.pdb.mono.pdb.processed.txt"

## [1] "1TWR.pdb.mono.pdb.processed.txt"

## [1] "2ZVU.pdb.mono.pdb.processed.txt"

## [1] "3MOO.pdb.mono.pdb.processed.txt"

## Warning: Expected 2 pieces. Additional pieces discarded in 244 rows [1, 36,
## 60, 61, 62, 63, 95, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132,
## 133, ...].


## Warning: Expected 2 pieces. Additional pieces discarded in 244 rows [1, 36,
## 60, 61, 62, 63, 95, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132,
## 133, ...].

## Warning: Expected 2 pieces. Additional pieces discarded in 104 rows [1, 2, 3, 4,
## 5, 40, 41, 42, 43, 44, 79, 123, 124, 125, 126, 127, 128, 129, 130, 131, ...].
```

A. The First Appendix

```
## Warning: Expected 2 pieces. Additional pieces discarded in 104 rows [1, 2, 3, 4,
## 5, 40, 41, 42, 43, 44, 79, 123, 124, 125, 126, 127, 128, 129, 130, 131, ...].
## Warning: Expected 2 pieces. Additional pieces discarded in 27 rows [1, 2, 3, 4,
## 5, 6, 7, 46, 90, 151, 225, 297, 370, 444, 514, 515, 516, 517, 518, 519, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 27 rows [1, 2, 3, 4,
## 5, 6, 7, 46, 90, 151, 225, 297, 370, 444, 514, 515, 516, 517, 518, 519, ...].
## Warning: Expected 2 pieces. Additional pieces discarded in 7 rows [1, 2, 3, 29,
## 30, 31, 32].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 7 rows [1, 2, 3, 29,
## 30, 31, 32].
## Warning: Expected 2 pieces. Additional pieces discarded in 8 rows [1, 31, 32,
## 33, 34, 35, 36, 37].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 8 rows [1, 31, 32,
## 33, 34, 35, 36, 37].
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
```

A. The First Appendix

```
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found
```

```
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found
## Warning: Expected 3 pieces. Additional pieces discarded in 4577 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 1290 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 796 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
```

A. The First Appendix

```
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 137 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 154 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 378037 rows [1,
## 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 73108 rows [1, 2,
## 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 68932 rows [1, 2,
## 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
```

```
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 8715 rows [1, 2,
## 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 6905 rows [1, 2,
## 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 378037 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 73108 rows [1, 2,
## 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 68932 rows [1, 2,
## 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 8715 rows [1, 2,
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
## Warning: Expected 2 pieces. Missing pieces filled with 'NA' in 6905 rows [1, 2,
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
```

```r
### actual code start lol
activeLigand = "HEM"

  # AA FReq ----
  rm(tmp5A,tmp7A,tmpBoth)
  head(HEM_aaFreqDf)
```

```
##   Residue Freq
## 1     LEU  261
## 2     PHE  224
## 3     ALA  188
## 4     ILE  161
## 5     VAL  158
## 6     TYR  156
```

A. The First Appendix

```
tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_aaFreqDf",sep='
                              #HEM_aaFreqDf)
tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_aaF
tmp7A$ang = '7'
tmp5A$ang = '5'
head(tmp7A)
```

```
##   df7A.Residue df7A.Freq ang
## 1          LEU       261   7
## 2          PHE       224   7
## 3          ALA       188   7
## 4          ILE       161   7
## 5          VAL       158   7
## 6          TYR       156   7
```

```
tmp7A
```

```
##    df7A.Residue df7A.Freq ang
## 1           LEU       261   7
## 2           PHE       224   7
## 3           ALA       188   7
## 4           ILE       161   7
## 5           VAL       158   7
## 6           TYR       156   7
## 7           ARG       146   7
## 8           HIS       142   7
## 9           THR       142   7
## 10          GLY       133   7
## 11          SER       129   7
## 12          GLU       104   7
## 13          ASP        99   7
## 14          LYS        95   7
```

```
## 15              PRO        84    7
## 16              ASN        78    7
## 17              GLN        78    7
## 18              MET        72    7
## 19              TRP        60    7
## 20              CYS        17    7
```

```r
tmp7A %>%
   dplyr::rename(
      df5A.Residue = df7A.Residue
   ) -> tmp7A
merge(tmp5A,tmp7A,by = 'df5A.Residue') -> tmpBoth
#alternative: left_join(tmp5A,tmp7A,by='df5A.Residue')
tmpBoth %>%
   dplyr::rename(
      Residue = df5A.Residue
   ) -> tmpBoth
tmpBoth
```

```
##     Residue df5A.Freq ang.x df7A.Freq ang.y
## 1       ALA        87     5       188     7
## 2       ARG        95     5       146     7
## 3       ASN        33     5        78     7
## 4       ASP        36     5        99     7
## 5       CYS        12     5        17     7
## 6       GLN        37     5        78     7
## 7       GLU        35     5       104     7
## 8       GLY        70     5       133     7
## 9       HIS        97     5       142     7
## 10      ILE        97     5       161     7
## 11      LEU       152     5       261     7
```

```
## 12      LYS        49     5        95    7

## 13      MET        46     5        72    7

## 14      PHE       153     5       224    7

## 15      PRO        31     5        84    7

## 16      SER        64     5       129    7

## 17      THR        60     5       142    7

## 18      TRP        37     5        60    7

## 19      TYR        97     5       156    7

## 20      VAL        97     5       158    7
```

```r
library(reshape) #ONLY LOAD THIS HERE. IF YOU LOAD EALIER,
```

```
##
## Attaching package: 'reshape'
## The following object is masked from 'package:data.table':
##
##     melt
## The following object is masked from 'package:dplyr':
##
##     rename
## The following objects are masked from 'package:tidyr':
##
##     expand, smiths
```

```r
# ALL CASE OF RENAME() FROM DPLYR PACKAGE IN ALL SCRIPTS GETS FUCKED UP.
# PUTTING IT HERE, ONLY BELOW THIS LIBRARY LOAD DO YOU NEED NEED TO SPECIFY DP

to_plot <- reshape::melt((data.frame(x=tmpBoth$Residue,`Distance_5A`=tmpBoth$df5
#   detach("package:reshape", unload=TRUE) #maybe avoid the problem anyway lol

print(ggplot(to_plot,aes(x= reorder(x,-value),y=value,fill=variable)) +
        geom_bar(stat="identity",position = "identity", alpha=.4) +
```

```
        labs(x = "Residue",y="Frequency", title = paste(activeLigand,": AA Freq
        scale_fill_discrete(name = "Distance Cutoff"))
```


HEM: AA Frequency
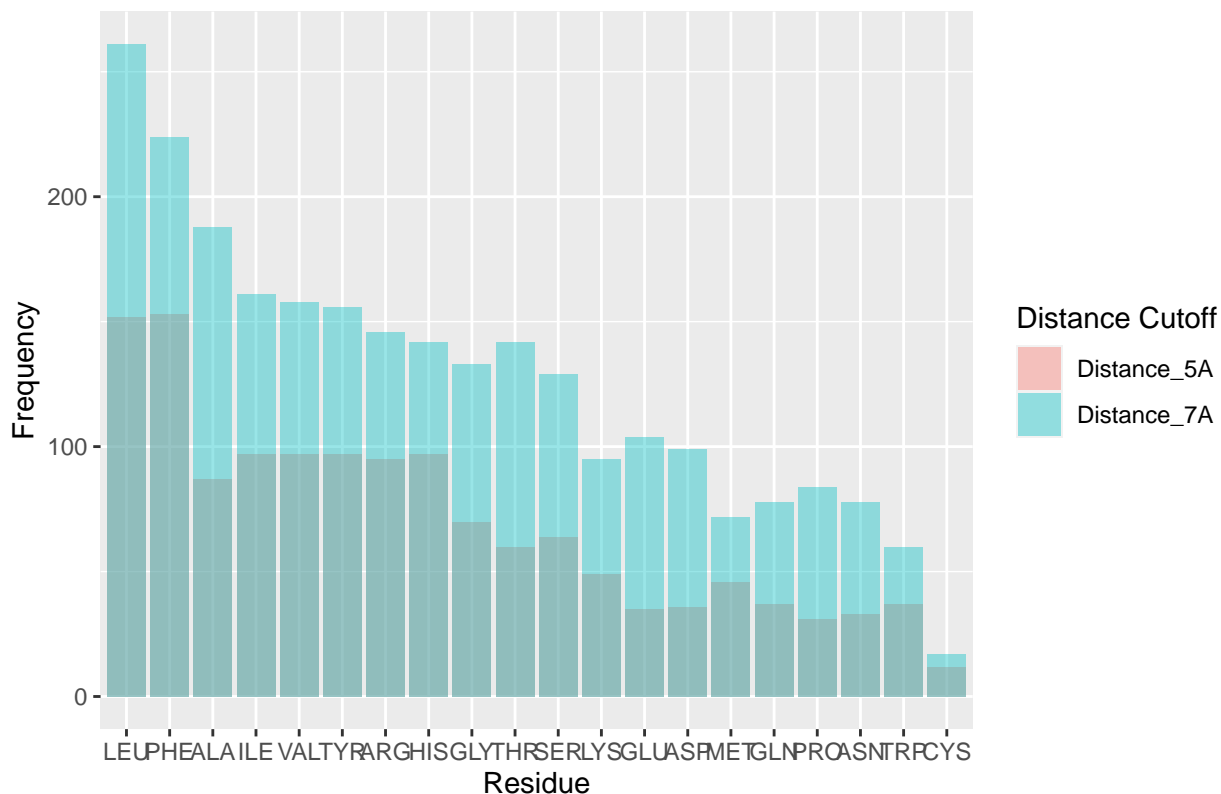
```
# ### VOLUME ####


rm(tmp5A,tmp7A,tmpBoth)

colName = 'volume_data'

tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_MERGED_DF$",col

tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_MERG

tmp7A$Distance_Cutoff = '7A'

tmp5A$Distance_Cutoff = '5A'

tmp7A %>%

    dplyr::rename(

        df5A = df7A #hopefuly does not append, rm() should take care of that

    ) -> tmp7A

tmpBoth <- rbind(tmp5A,tmp7A)
```

```
print(
    ggplot(tmpBoth, aes(x=df5A,fill=Distance_Cutoff,color=)) +
        geom_histogram(position="identity",alpha=0.4)
    + labs(x="Volume (A³)",y="Frequency",title = paste(activeLigand,": Volume (A³
)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


HEM : Volume (A³)

```
# Ligand Excluded Surface Area
```

```
rm(tmp5A,tmp7A,tmpBoth)
colName = paste(activeLigand,"_Excluded_SA",sep="")
tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_MERGED_DF$",col
tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_MERG
tmp7A$Distance_Cutoff = '7A'
tmp5A$Distance_Cutoff = '5A'
tmp7A %>%
```

```
    dplyr::rename(
        df5A = df7A #hopefuly does not append, rm() should take care of that
    ) -> tmp7A
tmpBoth <- rbind(tmp5A,tmp7A)
print(
    ggplot(tmpBoth, aes(x=df5A,fill=Distance_Cutoff,color=)) +
        geom_histogram(position="identity",alpha=0.4)
    + labs(x="Surface Area (A²)",y="Frequency",title = paste(activeLigand,": Liga
)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```
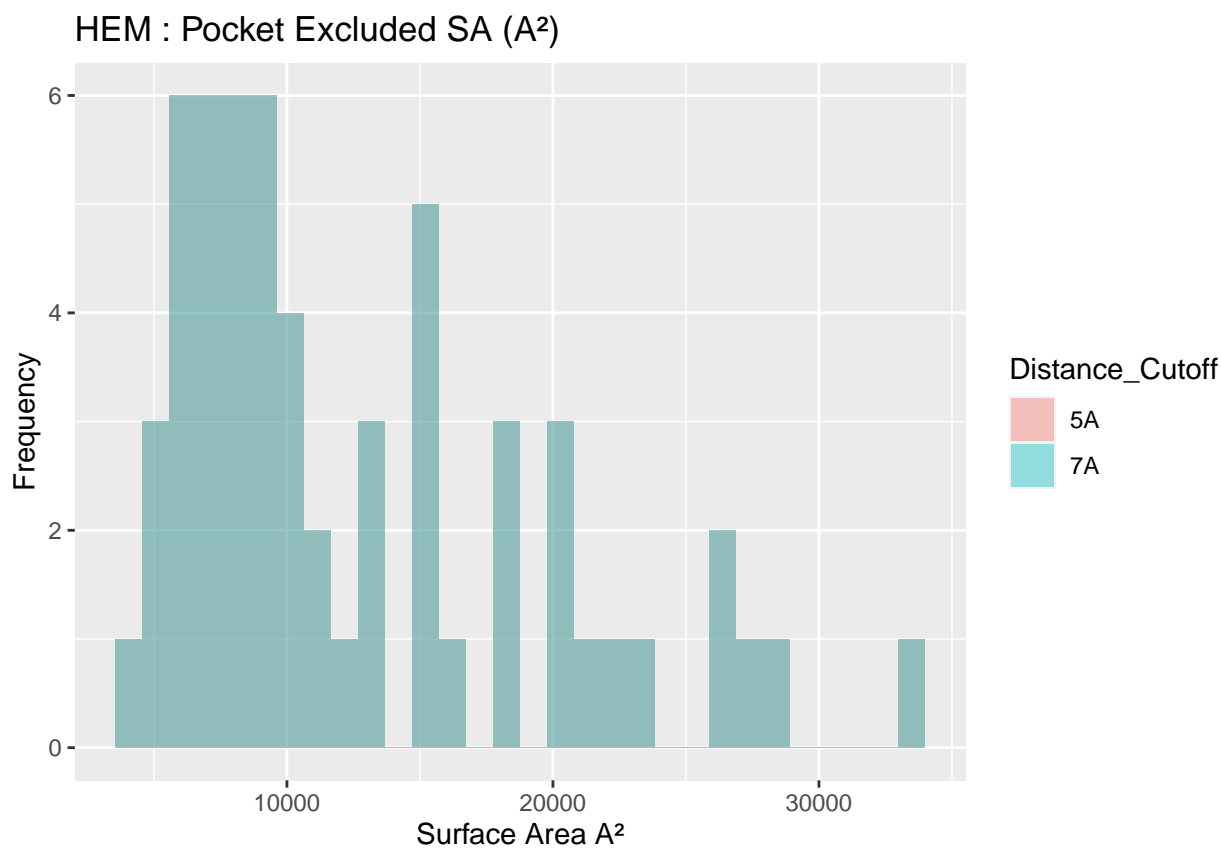


```
# Ligand Accessible Surface Area


rm(tmp5A,tmp7A,tmpBoth)
colName = paste(activeLigand,"_Accessible_SA",sep="")
tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_MERGED_DF$",col
```

A. The First Appendix

```
tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_MER
tmp7A$Distance_Cutoff = '7A'
tmp5A$Distance_Cutoff = '5A'
tmp7A %>%
   dplyr::rename(
      df5A = df7A #hopefuly does not append, rm() should take care of that
   ) -> tmp7A
tmpBoth <- rbind(tmp5A,tmp7A)
print(
   ggplot(tmpBoth, aes(x=df5A,fill=Distance_Cutoff,color=)) +
      geom_histogram(position="identity",alpha=0.4)
   + labs(x="Surface Area (A²)",y="Frequency",title = paste(activeLigand,": Liga
)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



HEM : Ligand Accessible SA (A²)
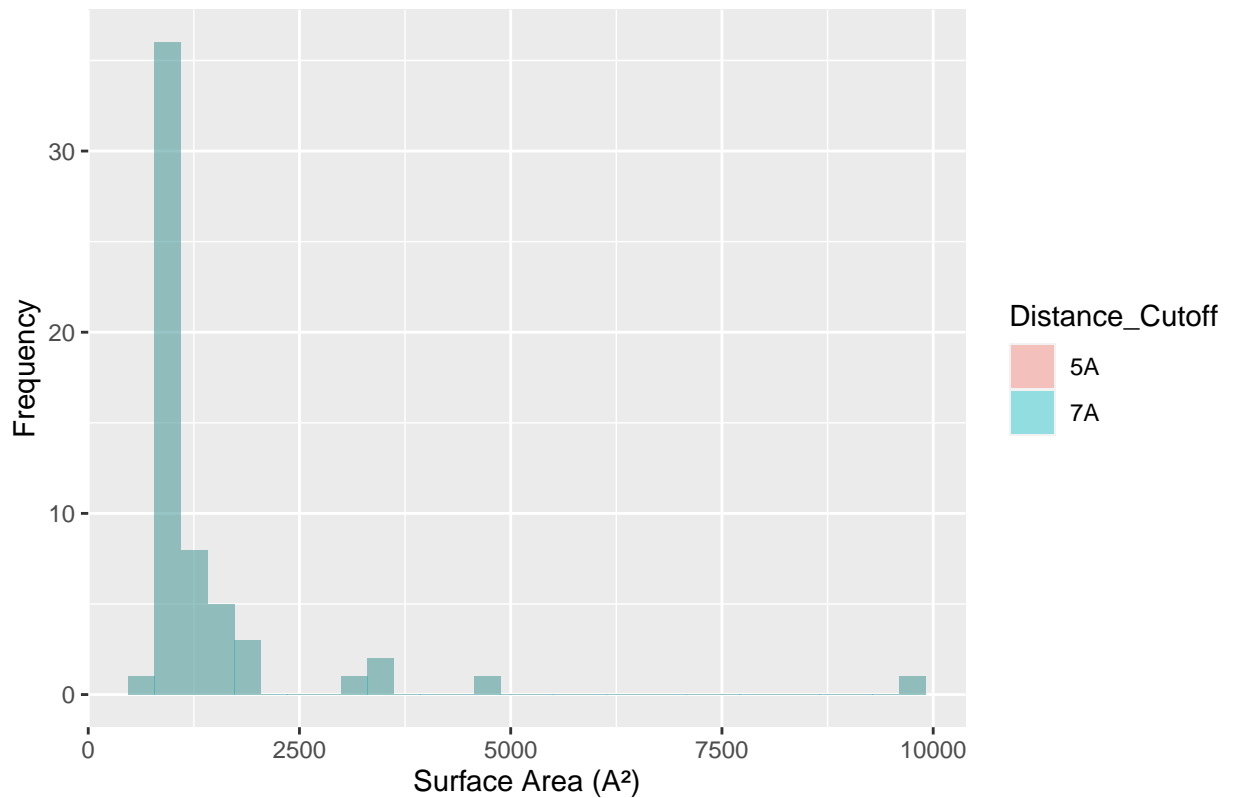
## A. The First Appendix

```
# Pocket Excluded SA

rm(tmp5A,tmp7A,tmpBoth)
colName = "Pocket_Excluded_SA"
tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_MERGED_DF$",col
tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_MERG
tmp7A$Distance_Cutoff = '7A'
tmp5A$Distance_Cutoff = '5A'
tmp7A %>%
    dplyr::rename(
        df5A = df7A #hopefuly does not append, rm() should take care of that
    ) -> tmp7A
tmpBoth <- rbind(tmp5A,tmp7A)
print(
    ggplot(tmpBoth, aes(x=df5A,fill=Distance_Cutoff,color=)) +
        geom_histogram(position="identity",alpha=0.4)
    + labs(x="Surface Area A²",y="Frequency",title = paste(activeLigand,": Pocket
)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

HEM : Pocket Excluded SA (A²)


```
# Pocket Acessible histo
rm(tmp5A,tmp7A,tmpBoth)
colName = "Pocket_Accessible_SA"
tmp7A <- data.frame(df7A = eval(parse(text=(paste(activeLigand,"_MERGED_DF$",col
tmp5A <- data.frame(df5A = eval(parse(text=(paste("ls5A$",activeLigand,"_5A_MERG
tmp7A$Distance_Cutoff = '7A'
tmp5A$Distance_Cutoff = '5A'
tmp7A %>%
    dplyr::rename(
        df5A = df7A #hopefuly does not append, rm() should take care of that
    ) -> tmp7A
tmpBoth <- rbind(tmp5A,tmp7A)
print(
    ggplot(tmpBoth, aes(x=df5A,fill=Distance_Cutoff,color=)) +
        geom_histogram(position="identity",alpha=0.4)
```

```
      + labs(title = paste(activeLigand,"Pocket Acessible SA (A²)"))
  )
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

HEM Pocket Acessible SA (A²)
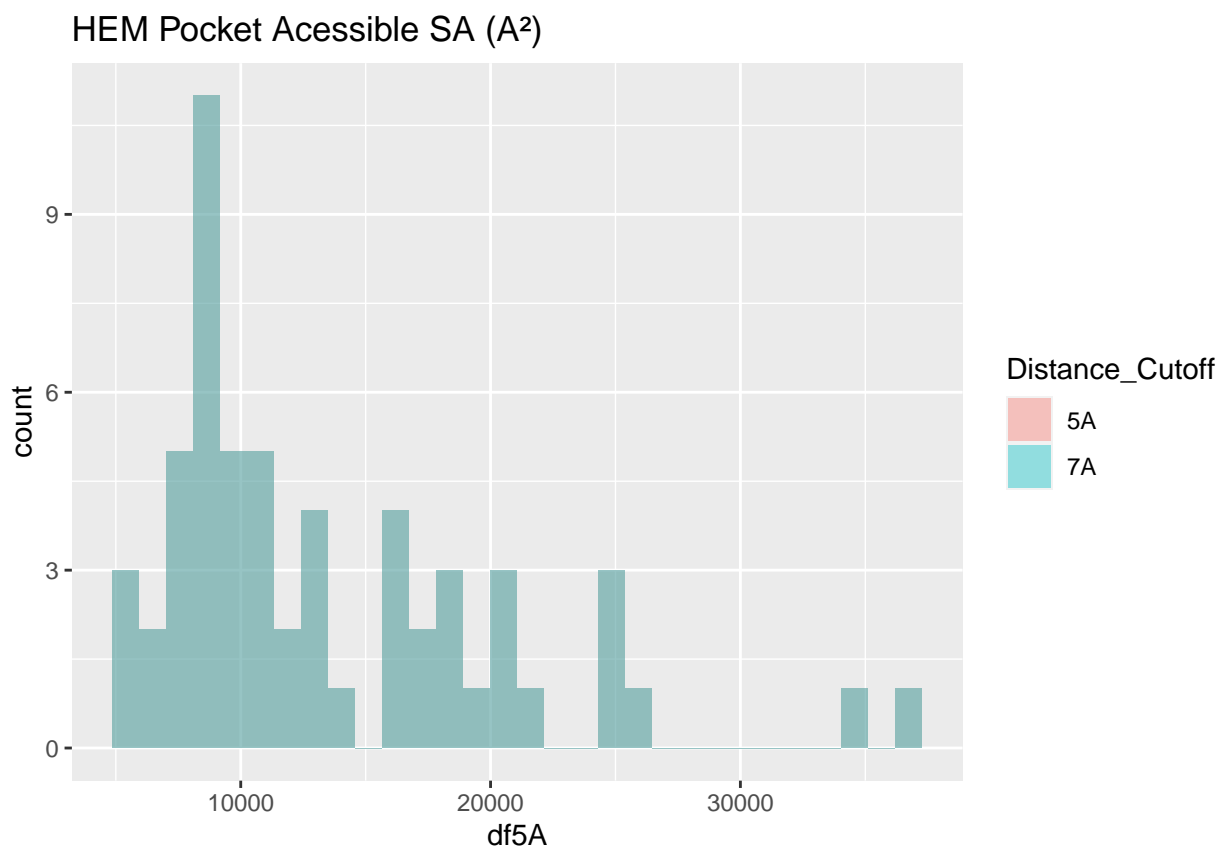

```
# plots of intersected distances with...

  # all planar angles
  # yeah keep the as.numeric(as.character),easiest way to solve the problem
  planarAllPlot <- ggplot(eval(parse(text=paste(activeLigand,"_allDistPlanarDf",se
                  aes(x=Residue_Code.x,y=as.numeric(as.character(Angle)),f
    geom_violin(trim=FALSE) +
    labs(title = paste(activeLigand,": All Planar Angles",sep=''), x="Residue",y=
  print(planarAllPlot)
```
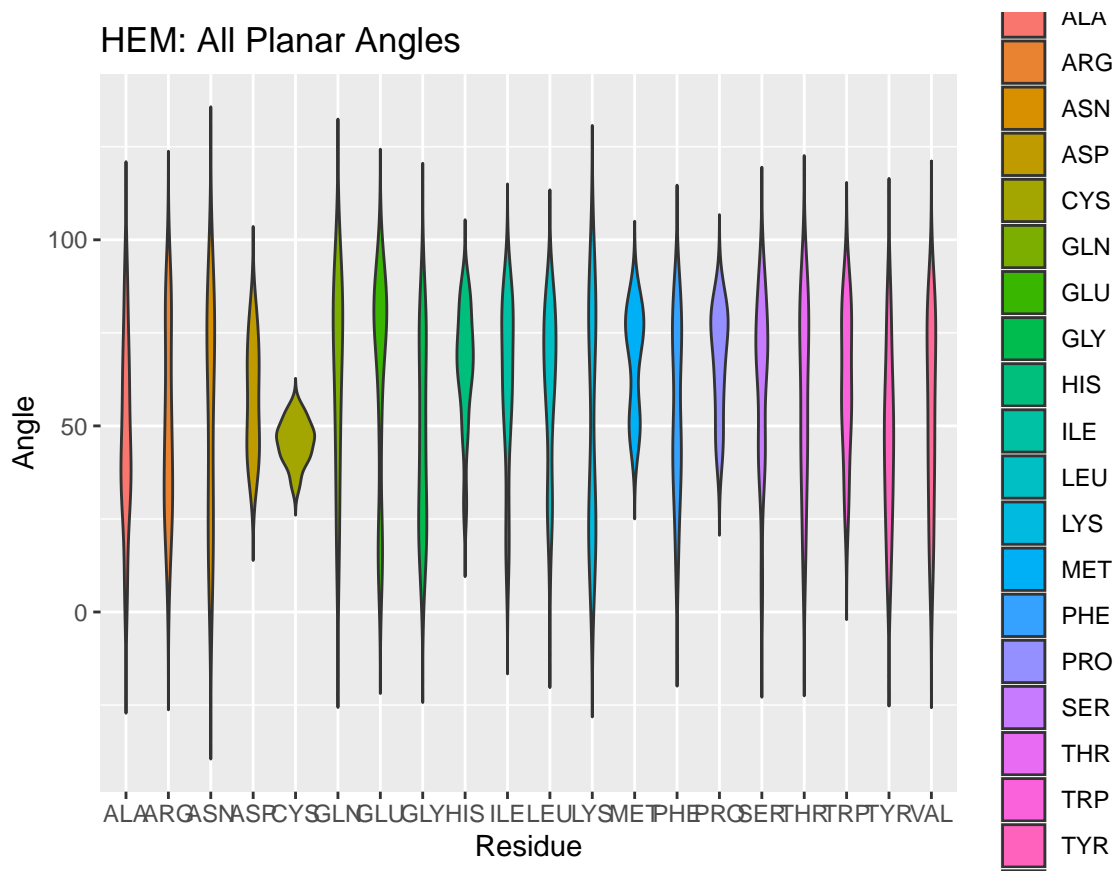
```
# closest residues' planar angles
planarMinPlot <- ggplot(eval(parse(text=paste(activeLigand,"_minDistPlanarDf",se
                        aes(x=Residue_Code.x,y=as.numeric(as.character(Angle)),f
    geom_violin(trim = FALSE) +
    labs(title = paste(activeLigand,": Planar Angles of Closest Residues",sep =
print(planarMinPlot)
```

```
## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.


## Warning: Groups with fewer than two data points have been dropped.


## Warning: Groups with fewer than two data points have been dropped.


## Warning: Groups with fewer than two data points have been dropped.
```
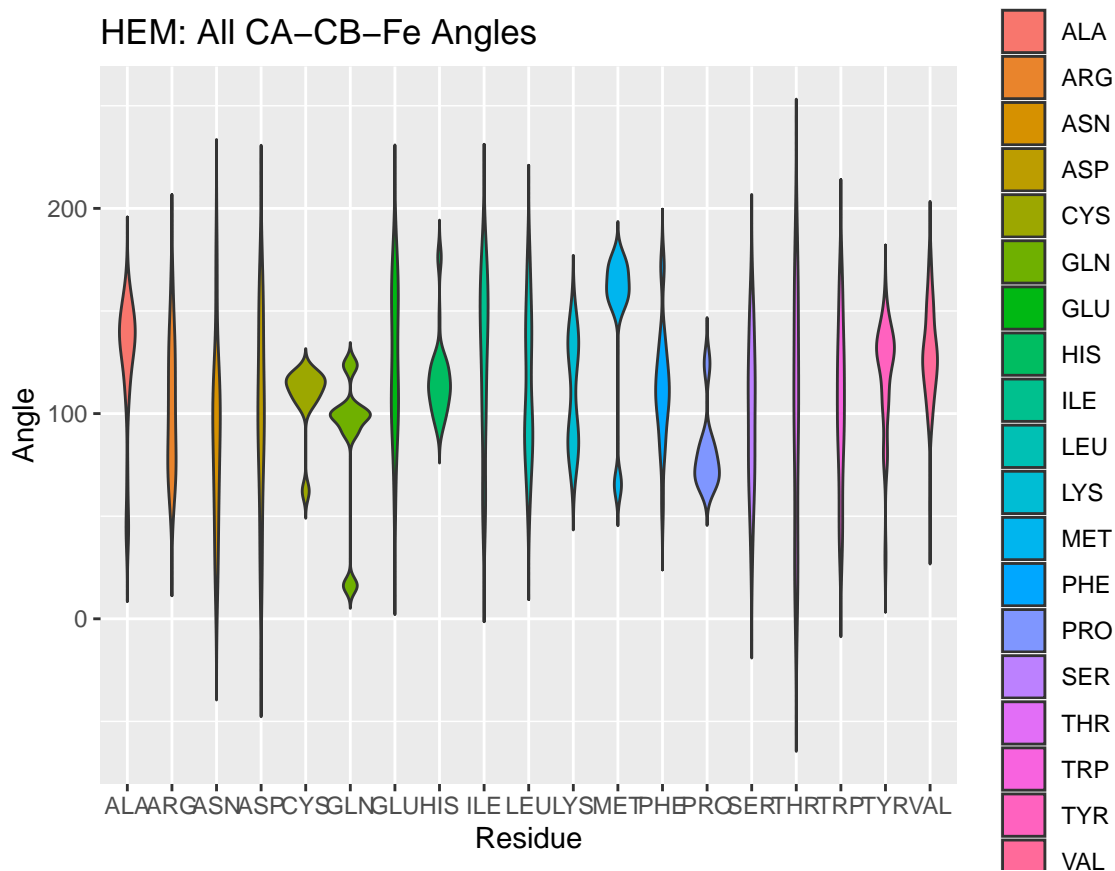
HEM: Planar Angles of Closest Residues


```
# all cabs with distance intersect
cabAllPlot <- ggplot(eval(parse(text=paste(activeLigand,"_allDistCabDf",sep=''))
                aes(x=Residue_Code.x,y=as.numeric(as.character(Angle)),fill
    geom_violin(trim=FALSE) +
    labs(title = paste(activeLigand,": All CA-CB-Fe Angles",sep=''), x="Residue",
print(cabAllPlot)
```

```r
# cabs for closest 3 residues
cabMinPlot <- ggplot(eval(parse(text=paste(activeLigand,"_minDistCabDf",sep='')
                  aes(x=Residue_Code.x,y=as.numeric(as.character(Angle)),fill
    geom_violin(trim=FALSE) +
    labs(title = paste(activeLigand,": CA-CB-Fe Angles of Closest Residues",sep=
print(cabMinPlot)
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```


HEM: CA–CB–Fe Angles of Closest Residues
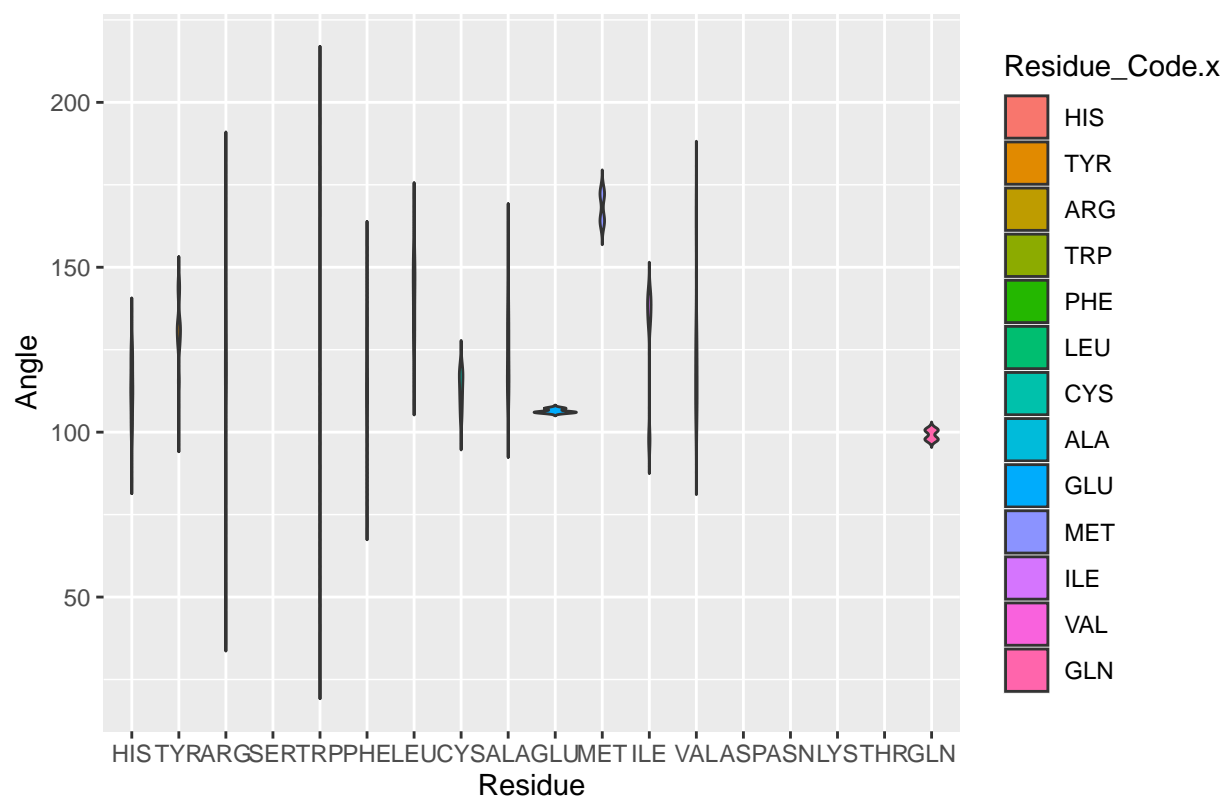
And here's another one from the same chapter, i.e. Chapter 2.2:

B

The Second Appendix, for Fun

Here we're gonna try out some DANK tables.

```r
# declarations


load("mainData.RData")


library(dplyr)

library(data.table)

library(tidyr)

library(ggplot2)

library(stringr)

library(knitr)

library(clipr) # Linux) sudo apt-get install xclip ... R)install.packages("clipr"

source("~/heme-binding/scripts/r/addpdbcol.R")


# this last one was tricky, see comments underneath:

library(kableExtra)

#install.packages("kableExtra")

# I had to download more stuff, in ubuntu terminal:
```

B. The Second Appendix, for Fun

```r
# sudo apt-get install libxml2-dev

# sudo apt-get install libssl-dev

# sudo apt-get install libcurl4-openssl-dev #not sure if competely necessary

# sudo apt-get install libfontconfig1-dev


# you could probably also tease out what you're missing from the error messages

# just keep trying to install kableExtra and download what you need, until it wor


#get 5A data real quick so we overwrite all DFs it uses, below:
source("~/heme-binding/scripts/r/data5A.R")
ls5A <- data5A_fn()
```

```
## [1] "1B2V.pdb.mono.pdb.processed.txt"
## [1] "1B5M.pdb.mono.pdb.processed.txt"
## [1] "1DK0.pdb.mono.pdb.processed.txt"
## [1] "1DKH.pdb.mono.pdb.processed.txt"
## [1] "1ICC.pdb.mono.pdb.processed.txt"
## [1] "1IPH.pdb.mono.pdb.processed.txt"
## [1] "1N45.pdb.mono.pdb.processed.txt"
## [1] "1P3T.pdb.mono.pdb.processed.txt"
## [1] "1QHU.pdb.mono.pdb.processed.txt"
## [1] "1QJS.pdb.mono.pdb.processed.txt"
## [1] "1SI8.pdb.mono.pdb.processed.txt"
## [1] "1SY2.pdb.mono.pdb.processed.txt"
## [1] "1U9U.pdb.mono.pdb.processed.txt"
## [1] "1VGI.pdb.mono.pdb.processed.txt"
## [1] "1ZVI.pdb.mono.pdb.processed.txt"
## [1] "2BHJ.pdb.mono.pdb.processed.txt"
## [1] "2CJ0.pdb.mono.pdb.processed.txt"
## [1] "2CN4.pdb.mono.pdb.processed.txt"
## [1] "2CPO.pdb.mono.pdb.processed.txt"
```

B. The Second Appendix, for Fun

```
## [1] "2E2Y.pdb.mono.pdb.processed.txt"

## [1] "2FC2.pdb.mono.pdb.processed.txt"

## [1] "2IIZ.pdb.mono.pdb.processed.txt"

## [1] "2IPS.pdb.mono.pdb.processed.txt"

## [1] "2J0P.pdb.mono.pdb.processed.txt"

## [1] "2J18.pdb.mono.pdb.processed.txt"

## [1] "2O6P.pdb.mono.pdb.processed.txt"

## [1] "2Q6N.pdb.mono.pdb.processed.txt"

## [1] "2R7A.pdb.mono.pdb.processed.txt"

## [1] "2SPL.pdb.mono.pdb.processed.txt"

## [1] "2VEB.pdb.mono.pdb.processed.txt"

## [1] "3HX9.pdb.mono.pdb.processed.txt"

## [1] "3MVF.pdb.mono.pdb.processed.txt"

## [1] "3QZN.pdb.mono.pdb.processed.txt"

## [1] "3QZZ.pdb.mono.pdb.processed.txt"

## [1] "3SIK.pdb.mono.pdb.processed.txt"

## [1] "3TGC.pdb.mono.pdb.processed.txt"

## [1] "3VP5.pdb.mono.pdb.processed.txt"

## [1] "3ZJS.pdb.mono.pdb.processed.txt"

## [1] "4B8N.pdb.mono.pdb.processed.txt"

## [1] "4CAT.pdb.mono.pdb.processed.txt"

## [1] "4CDP.pdb.mono.pdb.processed.txt"

## [1] "4I3Q.pdb.mono.pdb.processed.txt"

## [1] "4JET.pdb.mono.pdb.processed.txt"

## [1] "4MF9.pdb.mono.pdb.processed.txt"

## [1] "4MYP.pdb.mono.pdb.processed.txt"

## [1] "4NL5.pdb.mono.pdb.processed.txt"

## [1] "4UZV.pdb.mono.pdb.processed.txt"

## [1] "4XZD.pdb.mono.pdb.processed.txt"

## [1] "4Y1Q.pdb.mono.pdb.processed.txt"
```

B. The Second Appendix, for Fun

```
## [1] "5CN5.pdb.mono.pdb.processed.txt"

## [1] "5GJ3.pdb.mono.pdb.processed.txt"

## [1] "5KZL.pdb.mono.pdb.processed.txt"

## [1] "5O1L.pdb.mono.pdb.processed.txt"

## [1] "5O1M.pdb.mono.pdb.processed.txt"

## [1] "5VEU.pdb.mono.pdb.processed.txt"

## [1] "6A2J.pdb.mono.pdb.processed.txt"

## [1] "7C74.pdb.mono.pdb.processed.txt"

## [1] "7DMR.pdb.mono.pdb.processed.txt"

## [1] "1BBH.pdb.mono.pdb.processed.txt"

## [1] "1S56.pdb.mono.pdb.processed.txt"

## [1] "1W2L.pdb.mono.pdb.processed.txt"

## [1] "2BC5.pdb.mono.pdb.processed.txt"

## [1] "2BH5.pdb.mono.pdb.processed.txt"

## [1] "3EAH.pdb.mono.pdb.processed.txt"

## [1] "3X15.pdb.mono.pdb.processed.txt"

## [1] "4B2N.pdb.mono.pdb.processed.txt"

## [1] "5KPF.pdb.mono.pdb.processed.txt"

## [1] "5LFT.pdb.mono.pdb.processed.txt"

## [1] "5T8W.pdb.mono.pdb.processed.txt"

## [1] "6VDQ.pdb.mono.pdb.processed.txt"

## [1] "6WZA.pdb.mono.pdb.processed.txt"

## [1] "6XNK.pdb.mono.pdb.processed.txt"

## [1] "1ZJ8.pdb.mono.pdb.processed.txt"

## [1] "2AKJ.pdb.mono.pdb.processed.txt"

## [1] "2AOP.pdb.mono.pdb.processed.txt"

## [1] "3B0G.pdb.mono.pdb.processed.txt"

## [1] "3VKP.pdb.mono.pdb.processed.txt"

## [1] "3VLX.pdb.mono.pdb.processed.txt"

## [1] "3VLY.pdb.mono.pdb.processed.txt"
```

```
## [1] "3VLZ.pdb.mono.pdb.processed.txt"
```

```
## [1] "5H8V.pdb.mono.pdb.processed.txt"
```

```
## [1] "1TWN.pdb.mono.pdb.processed.txt"
```

```
## [1] "1TWR.pdb.mono.pdb.processed.txt"
```

```
## [1] "2ZVU.pdb.mono.pdb.processed.txt"
```

```
## [1] "3MOO.pdb.mono.pdb.processed.txt"
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 244 rows [1, 36,
```

```
## 60, 61, 62, 63, 95, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132,
```

```
## 133, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 244 rows [1, 36,
```

```
## 60, 61, 62, 63, 95, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132,
```

```
## 133, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 104 rows [1, 2, 3, 4,
```

```
## 5, 40, 41, 42, 43, 44, 79, 123, 124, 125, 126, 127, 128, 129, 130, 131, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 104 rows [1, 2, 3, 4,
```

```
## 5, 40, 41, 42, 43, 44, 79, 123, 124, 125, 126, 127, 128, 129, 130, 131, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 27 rows [1, 2, 3, 4,
```

```
## 5, 6, 7, 46, 90, 151, 225, 297, 370, 444, 514, 515, 516, 517, 518, 519, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 27 rows [1, 2, 3, 4,
```

```
## 5, 6, 7, 46, 90, 151, 225, 297, 370, 444, 514, 515, 516, 517, 518, 519, ...].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 7 rows [1, 2, 3, 29,
```

```
## 30, 31, 32].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 7 rows [1, 2, 3, 29,
```

```
## 30, 31, 32].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 8 rows [1, 31, 32,
```

```
## 33, 34, 35, 36, 37].
```

B. The Second Appendix, for Fun

```
## Warning: Expected 2 pieces. Additional pieces discarded in 8 rows [1, 31, 32,
## 33, 34, 35, 36, 37].
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found
```

```
## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'awesome_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'dist_aa_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'aa_super_df' not found

## Warning in rm(awesome_df, combined_results_df, dist_aa_df, distance_raw_df, :
## object 'ang1_df' not found

## Warning: Expected 3 pieces. Additional pieces discarded in 4577 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult

## `mutate_if()` ignored the following grouping variables:
## Column `PDB_ID`

## Warning: Expected 3 pieces. Additional pieces discarded in 1290 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
```

B. The Second Appendix, for Fun

```
## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 796 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 137 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
## 'mutate_if()' ignored the following grouping variables:
## Column 'PDB_ID'
## Warning: Expected 3 pieces. Additional pieces discarded in 154 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].


## Warning: Ignoring unknown parameters: mult


## Warning: Ignoring unknown parameters: mult
```

```
## Warning: Ignoring unknown parameters: mult

## `mutate_if()` ignored the following grouping variables:

## Column `PDB_ID`

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 378037 rows [1,

## 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 73108 rows [1, 2,

## 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 68932 rows [1, 2,

## 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 8715 rows [1, 2,

## 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 6905 rows [1, 2,

## 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 378037 rows [1,

## 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 73108 rows [1, 2,

## 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 68932 rows [1, 2,

## 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 8715 rows [1, 2,

## 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 6905 rows [1, 2,

## 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
```

```r
### actual code start lol
activeLigand = "HEM"


HEM_p1DF %>%
```

```
kable(longtable = T, booktabs = T, caption = "p1DF") %>%
 kable_styling(full_width = T, latex_options = c("striped","repeat_header"))
```

Table B.1: p1DF

PDB_ID	Molecule_Name	Source_Organism
1B2V	PROTEIN (HEME-BINDING PROTEIN A);	SERRATIA MARCESCENS;
1B5M	CYTOCHROME B5;	RATTUS NORVEGICUS;
1DK0	HEME-BINDING PROTEIN A;	SERRATIA MARCESCENS;
1DKH	HEME-BINDING PROTEIN A;	SERRATIA MARCESCENS;
1ICC	CYTOCHROME B5 OUTER MITOCHONDRIAL MEMBRANE	RATTUS NORVEGICUS;
1IPH	CATALASE HPII;	ESCHERICHIA COLI;
1N45	HEME OXYGENASE 1;	HOMO SAPIENS;
1P3T	HEME OXYGENASE 1;	NEISSERIA MENINGITIDIS;
1QHU	PROTEIN (HEMOPEXIN);	ORYCTOLAGUS CUNICULUS;
1QJS	HEMOPEXIN;	ORYCTOLAGUS CUNICULUS;
1SI8	CATALASE;	ENTEROCOCCUS FAECALIS;
1SY2	NITROPHORIN 4;	RHODNIUS PROLIXUS;
1U9U	CYTOCHROME B5;	BOS TAURUS;
1VGI	HEME OXYGENASE 1;	RATTUS NORVEGICUS;
1ZVI	NITRIC-OXIDE SYNTHASE, BRAIN;	RATTUS NORVEGICUS;
2BHJ	NITRIC OXIDE SYNTHASE;	MUS MUSCULUS;
2CJ0	CHLOROPEROXIDASE;	CALDARIOMYCES FUMAGO;
2CN4	HEMOPHORE HASA;	SERRATIA MARCESCENS;
2CPO	CHLOROPEROXIDASE;	LEPTOXYPHIUM FUMAGO;
2E2Y	MYOGLOBIN;	PHYSETER CATODON;

Table B.1: p1DF *(continued)*

PDB_ID	Molecule_Name	Source_Organism
2FC2	NITRIC OXIDE SYNTHASE;	BACILLUS SUBTILIS;
2IIZ	MELANIN BIOSYNTHESIS PROTEIN TYRA, PUTATIVE;	SHEWANELLA ONEIDENSIS;
2IPS	LACTOPEROXIDASE;	BOS TAURUS;
2J0P	HEMIN TRANSPORT PROTEIN HEMS;	YERSINIA ENTEROCOLITICA;
2J18	CHLOROPEROXIDASE;	CALDARIOMYCES FUMAGO;
2O6P	IRON-REGULATED SURFACE DETERMINANT PROTEIN C;	STAPHYLOCOCCUS AUREUS SUBSP. AUREUS;
2Q6N	CYTOCHROME P450 2B4;	ORYCTOLAGUS CUNICULUS;
2R7A	BACTERIAL HEME BINDING PROTEIN;	SHIGELLA DYSENTERIAE;
2SPL	MYOGLOBIN;	PHYSETER CATODON;
2VEB	PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;
3HX9	PROTEIN RV3592;	MYCOBACTERIUM TUBERCULOSIS;
3MVF	NITROPHORIN-4;	RHODNIUS PROLIXUS;
3QZN	IRON-REGULATED SURFACE DETERMINANT PROTEIN A;	STAPHYLOCOCCUS AUREUS SUBSP. AUREUS;
3QZZ	METHANOSARCINA ACETIVORANS PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;
3SIK	CONSERVED DOMAIN PROTEIN;	BACILLUS ANTHRACIS;
3TGC	NITROPHORIN-4;	RHODNIUS PROLIXUS;
3VP5	TRANSCRIPTIONAL REGULATOR;	LACTOCOCCUS LACTIS;
3ZJS	PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;

Table B.1: p1DF *(continued)*

PDB_ID	Molecule_Name	Source_Organism
4B8N	CYTOCHROME B5-HOST ORIGIN;	OSTREOCOCCUS TAURI VIRUS 2;
4CAT	CATALASE;	PENICILLIUM JANTHINELLUM;
4CDP	PUTATIVE HEME/HEMOGLOBIN TRANSPORT PROTEIN;	ESCHERICHIA COLI;
4I3Q	CYTOCHROME P450 3A4;	HOMO SAPIENS;
4JET	HEMOPHORE HASA;	YERSINIA PESTIS;
4MF9	HEMIN DEGRADING FACTOR;	PSEUDOMONAS AERUGINOSA;
4MYP	IRON-REGULATED SURFACE DETERMINANT PROTEIN A;	LISTERIA MONOCYTOGENES;
4NL5	HEME-DEGRADING MONOOXYGENASE HMOB;	MYCOBACTERIUM TUBERCULOSIS;
4UZV	HEMOGLOBIN;	THERMOBIFIDA FUSCA TM51;
4XZD	EXTRACELLULAR HEME ACQUISITION HEMOPHORE HASA;	YERSINIA PSEUDOTU-BERCULOSIS IP 32953;
4Y1Q	EXTRACELLULAR HEME ACQUISITION HEMOPHORE HASA;	YERSINIA PSEUDOTU-BERCULOSIS IP 32953;
5CN5	MYOGLOBIN;	EQUUS CABALLUS;
5GJ3	PERIPLASMIC BINDING PROTEIN;	ROSEIFLEXUS SP. RS-1;
5KZL	HEME OXYGENASE;	LEPTOSPIRA INTERROGANS;
5O1L	RUBBER OXYGENASE;	STREPTOMYCES SP. (STRAIN K30);
5O1M	RUBBER OXYGENASE;	STREPTOMYCES SP. (STRAIN K30);
5VEU	CYTOCHROME P450 3A5;	HOMO SAPIENS;
6A2J	HEME A SYNTHASE;	BACILLUS SUBTILIS (STRAIN 168);
7C74	LACTOPEROXIDASE;	BOS MUTUS;
7DMR	LACTOPEROXIDASE;	BOS MUTUS;

Table B.1: p1DF *(continued)*

PDB_ID	Molecule_Name	Source_Organism

```
HEM_p2DF %>%
  kable(longtable = T, booktabs = T, caption = "p2DF") %>%
    kable_styling(full_width = T, latex_options = c("striped","repeat_header"))
```

Table B.2: p2DF

PDB_ID	Volume_Data	HEM_Excluded_SA	HEM_Accessible_SA	Pocket_SA_Excluded_SA	Pocket_SA_Accessible_SA
1B2V	893.60	502.042	820.988	7276.09	8232.60
1B5M	672.79	490.050	800.780	4695.01	5512.20
1DK0	966.72	505.258	837.157	7237.94	8217.58
1DKH	1010.70	509.042	828.131	7402.34	8175.94
1ICC	1000.40	499.585	811.357	5079.72	6028.23
1IPH	1345.60	501.603	814.652	33983.80	34094.40
1N45	978.98	560.384	983.238	9944.50	10779.30
1P3T	987.05	509.939	829.611	9530.67	10410.80
1QHU	1389.20	573.686	1002.160	18503.10	18257.20
1QJS	1102.30	573.266	1000.380	18588.40	18584.10
1SI8	965.57	646.643	1184.070	23711.20	25120.40
1SY2	918.34	501.850	817.749	8960.76	9610.23
1U9U	738.55	496.132	813.773	4675.76	5632.32
1VGI	870.44	577.234	1002.530	9615.29	10248.20
1ZVI	1435.90	701.091	1129.540	19918.60	20968.20
2BHJ	1438.30	836.576	1290.530	20102.30	20762.60
2CJ0	809.62	2653.180	4835.280	12749.60	12892.20
2CN4	526.88	576.760	961.348	9617.23	11917.70
2CPO	886.17	1846.490	3329.540	13081.60	12995.60
2E2Y	994.92	811.270	1607.370	7531.94	8240.75
2FC2	1091.40	1011.190	1669.900	18383.50	18552.10
2IIZ	1015.60	731.342	1393.160	13651.70	14031.40
2IPS	1242.40	618.252	1075.560	27760.50	25814.10
2J0P	1281.80	1030.510	1873.810	15192.90	15871.10
2J18	841.67	1962.990	3556.340	12675.10	12779.00
2O6P	788.05	499.017	822.121	6234.84	7200.43
2Q6N	1030.10	644.365	1040.080	20051.10	19747.50
2R7A	1284.50	507.098	845.182	11255.10	12389.00
2SPL	1055.70	589.706	1029.660	7588.36	8105.94
2VEB	886.06	762.309	1454.750	9840.72	10401.80

Table B.2: p2DF *(continued)*

PDB_ID	Volume_Data	HEM_Excluded_SA	HEM_Accessible_SA	Pocket_SA_Excluded_SA	Pocket_SA_Accessible_SA
3HX9	1844.50	785.442	1168.200	5819.08	7189.03
3MVF	1271.40	576.502	1009.950	8559.24	9573.08
3QZN	726.52	664.858	1221.330	6133.24	7179.49
3QZZ	977.30	496.950	825.255	8523.59	9708.28
3SIK	492.15	498.621	823.565	6495.38	7739.06
3TGC	969.87	524.380	853.710	8712.77	9181.94
3VP5	1094.60	602.790	1050.820	9801.82	10810.80
3ZJS	788.74	528.419	860.137	9568.10	10130.40
4B8N	841.27	569.302	990.216	4560.39	5458.66
4CAT	1933.90	484.341	778.502	28372.40	36788.30
4CDP	1053.70	1425.050	3141.090	14733.50	15887.40
4I3Q	1220.50	510.623	845.108	21946.50	21093.70
4JET	1010.80	495.992	818.131	7887.81	8695.85
4MF9	1286.50	488.695	790.732	15669.80	16791.30
4MYP	610.72	963.019	1834.680	6285.40	7351.53
4NL5	1088.70	576.669	1003.400	5715.52	6894.72
4UZV	1184.10	526.584	844.058	7378.28	8322.74
4XZD	932.14	498.788	816.032	8028.32	8752.50
4Y1Q	952.23	494.939	806.960	7905.84	8785.04
5CN5	1070.30	663.162	1223.640	7629.45	8117.34
5GJ3	1108.20	756.603	1131.670	11394.00	12591.80
5KZL	914.22	483.760	805.567	9662.03	10431.00
5O1L	1438.70	801.519	1447.270	15538.20	16876.00
5O1M	1431.30	493.850	799.331	16096.90	15912.50
5VEU	964.76	993.578	1502.660	20900.80	20425.90
6A2J	1015.90	6183.450	9902.920	14870.30	15888.00
7C74	1155.10	497.527	820.381	26111.40	25094.20
7DMR	1083.40	1049.750	1916.950	26004.00	24563.80

Works Cited

[1]   Ting Li, Herbert L Bonkovsky, and Jun Tao Guo. "Structural analysis of heme proteins: Implications for design and prediction". In: *BMC Structural Biology* 11 (2011). DOI: 10.1186/1472-6807-11-13.