

Feasibility and Acceptance of chatbots embedded in healthcare curricula:



—
CEPEH report
—

December

2022

see cepeh.eu for more information

Acknowledgements

This work is supported by the ERASMUS+ Strategic Partnership in Higher Education “Chatbot Enhance Personalise European Healthcare Curricula (CEPEH)” (www.cepeh.eu) (2019-1-UK01- KA203-062091) project of the European Union.

CEPEH Team

Abstract

This document details the evaluation of each resource in terms of the feasibility and acceptance from the end-users. There was evidence of identifying the feasibility of such resources into formal training and studies exist on the acceptance of such resources, with promising results. However, all these studies defined the need for further research in the area until the use of chatbots in healthcare education became common. Furthermore, the creation process of CEPEH resources was significantly different and had improvements to current methods, due to the co-creation process, and use of low cost but effective technology.

Contents

List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
Introduction	1
Background	2
1 Method	4
1.1 Participants	4
1.2 Procedure	5
1.3 Design	5
1.4 Materials	5
1.5 Analysis	8
2 Results	9
2.1 Pre Usage Results	9
2.2 System Usability Scale (SUS) Scores	12
2.3 Technology Acceptance Model	12
2.4 Executable inline code	21
2.5 Executable code in other languages than R	22
3 Training Event Results	23
3.1 CEPEH Training Event C1	23
4 Overall Training Events Evalaution	24
4.1 CEPEH Training Event 2	24
4.2 Cross-referencing	29
4.3 Collaborative writing	32
4.4 Additional resources	32

Contents

5	Discussion	33
5.1	Reach, Impact, and Qualatative analysis	34
6	Text Mining, Natural Language Processing, and Sentiment Analysis	46
6.1	1 Reading in texts	46
6.2	2 Tidy text	46
6.3	3 Analysing frequencies	49
6.4	4 Comparing the vocabulary of texts	55
7	Sentiment analysis	57
	Appendix	61
	More info	61
	Appendices	
A	The First Appendix	63
B	The Second Appendix, for Fun	64
	References	65

List of Figures

2.1	Code chunk syntax	17
2.2	Oxford logo	18
2.3	Oxford logo, rotated	19
2.4	A ggplot of car stuff	20
2.5	An Oxford logo that LaTeX will try to place at this position in the text	21
4.1	A marvel-lous meme	30
5.1	Font sizes in LaTeX	44

List of Tables

2.1	A knitr kable table	19
4.1	Stopping cars	31

List of Abbreviations

CEPEH	Chatbot Enhance Personalised European Healthcare curricula
RLO	Reusable Learning Object
NLP	Natural Language Processing
NLU	Natural Language Understanding
A.I	Artificial Intelligence

Introduction

Personalised Healthcare Education is needed to meet growing demand and quality maintenance. There is a growing evidence around chatbots, namely machine conversation systems- these programs have the potential to change the way students learn and search for information.

Chatbots can quiz existing knowledge, enable higher student engagement with a learning task, or support higher-order cognitive activities. In large-scale learning scenarios with a high student-to-lecturer ratio, chatbots can help tackle the issue of individualized student support and facilitate personalised learning. However, limited examples of chatbots in European Healthcare Curricula have been utilised to combine both the continuum of cognitive processes presented in Bloom's taxonomy, with the idea that some repetitive tasks can be done with a chatbot- to provide greater access or to scale faculty time.

Thus, CEPEH strategic partnership has co-created open access chatbots utilising artificial intelligence, promoting innovative practices in digital era, by supporting current curricula and fostering open education.

CEPEH Erasmus+ strategic partnership aimed to co-design and implement new pedagogical approaches and, in particular, chatbots for European medical and nursing schools. CEPEH used participatory design to engage stakeholders (students, healthcare workforce staff, lecturers, clinicians, etc.) in order to co-design effective chatbots and release them as open access resources. Through CEPEH, effective use of digital technologies and open education were incorporated into healthcare curricula. This enabled students to increase their health and medical related skills through flexible learning.

Introduction

CEPEH expected that students adopted this new digital pedagogy and improve their skills and competences through flexible personalised learning, while the teaching staff enhanced their e-learning tool co-creation competences and make use of co-design best practices and recommendations for use. It is also expected increased cooperation between the partners. Thus, in the long term, CEPEH expects to influence the development of medical and nursing curricula with this digital innovation, foster the quality of the future healthcare workforce and further improve international competitiveness of the partners' healthcare curricula. This document details the evaluation of the resources created by the CEPEH team.

The evaluation specifically explored the feasibility and acceptance from the end-users. These end-users are learners in European healthcare higher education institutions.

There was firstly evidence for the need to identify the feasibility of chatbots and similar resources into formal education and training, with a further need to improve access to these types of learning resources. Of course, studies exist on the acceptance of chatbots, virtual patients, and many other healthcare applications, with promising results. However, through various limitations, we believed there was further research to be completed to accelerate the design, development, implementation, and evaluation processes. These have financial, stakeholder, time, and efficacy benefits. The creation process of CEPEH resources was significantly different to most in the literature, and this report highlights the approach of the CEPEH team towards enhancing personalised healthcare education can be achieved.

Background

The working practices of CEPEH are aimed at maximizing efficacy of these chatbots as learning resources, and provided a sense of shared development and ownership from all stakeholders. The process normally begins with workshops in which the project is scoped and team building occurs. The CEPEH workshops involve the widest possible team of stakeholders including tutors, students, healthcare

Introduction

workers, learning technologists, health service users and carers- depending on the materials being created.

For readers who are interested in using these high quality digital resources please access them for free at CEPEH.EU

The next section will now present the evaluation of all CEPEH chatbot resources.

1

Method

Contents

1.1	Participants	4
1.2	Procedure	5
1.3	Design	5
1.4	Materials	5
1.4.1	System Usability Scale	5
1.4.2	Computer Self-Efficacy Scale Tool	6
1.4.3	Unified Theory of Acceptance and Use of Technology . .	6
1.4.4	Technology Acceptance Model (TAM)	6
1.4.5	Qualitative Measure- Focus Group Discussions	7
1.5	Analysis	8

1.1 Participants

This dataset had 14 males and 28 females, for a total of 168. It was a repeated measure design therefore‘

There were 21 females from Greece, 3 from Cyprus, 2 from Sweden.

There were 7 males from Greece, 3 from Cyprus, 2 from Sweden.

1. Method

1.2 Procedure

1.3 Design

1.4 Materials

1.4.1 System Usability Scale

The System Usability Scale (SUS) was used [10] and is a widely used and adopted usability questionnaire. It is popular due to its unbiased and agnostic properties, a non proprietary, and quick scale of 10 questions.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The SUS was developed with a scoring system, in which the following should be performed: For each of the odd numbered questions, subtract 1 from the score. For each of the even numbered questions, subtract their value from 5. Add up these numbers to find the total score, then multiply this by 2.5. The result is a score out of 100 and can be compared against a determined average score of 68. Further, 80.3 or higher is excellent, and 51 or under suggests significant usability problems.

1. Method

1.4.2 Computer Self-Efficacy Scale Tool

The 10 question CSEST was based on the 32-item questionnaire by Murphy, Coover, and Owen (1989). Participants were provided with the facilitator stating 'Imagine you have found a new technology product that you have previously not used. You believe this product will make your life better. It doesn't matter specifically what this technology product does, only that it is intended to make your life easier and that you have never used it before. I could use the new technology...

1. If there was no one around to tell me what to do as I go
2. If I had never used a product like it before
3. If I had only the product manuals for reference
4. If I had seen someone else using it before trying it myself
5. If I could call someone for help if I got stuck
6. If someone else had helped me get started
7. If I had a lot of time to complete the job for which the product was provided
8. If I had just the built-in help facility for assistance
9. If someone showed me how to do it first
10. If I had used similar products before this one to do the same job

1.4.3 Unified Theory of Acceptance and Use of Technology

1.4.4 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) [1] was specifically developed with the primary aim of identifying the determinants involved in computer acceptance in general; secondly, to examine a variety of information technology usage behaviours; and thirdly, to provide a parsimonious theoretical explanatory model. TAM suggests that attitude would be a direct predictor of the intention to use technology, which in turn would predict the actual usage of the technology. The only modification to the nine sub-scales of the questionnaire consists of applying the items to the context of chatbots. All the items, except those measuring attitudes, utilize a

1. Method

seven-point Likert scale ranging from “strongly agree” to “strongly disagree” with a middle neutral point [2].

The nine sub-scales of the questionnaire:

Ease of use of chatbots Perceived usefulness of chatbots Intention of use.
Attitude toward usage of chatbots. Perception of personal efficacy to use a chatbot resource. Perception of external control toward chatbots. Anxiety toward chatbot use. Intrinsic motivation to use chatbot resources. Perceived costs of chatbots.

1.4.5 Qualitative Measure- Focus Group Discussions

Focus groups are a pervasive means of market research and provides credible acceptance evaluators regarding the penetration that a product or service will have on a target demographic. Focus groups are a form of qualitative research consisting of interviews or structured discussions, in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging. Questions are asked in an interactive group setting where participants are free to talk with other group members. During this process, the researcher either takes notes or records the vital points he or she is getting from the group. Researchers select members of the focus group carefully for effective and authoritative responses. Relevant stakeholders, then, can use the information collected through focus groups to receive insights on a specific product, issue, or topic focus [7].

A series of short focus group sessions identified the feasibility of CEPEH resources for formal curricular integration. These sessions, spanning no more than 1-1.5 hours and consisting of no more than 5-7 persons each explored all axes of curricular integration such as accessibility in the classroom, use case scenarios, technology requirements for curricular integration etc. These axes were formalized by the research team, in each evaluation site, to consider the curricular details of each institution.

1. Method

1.5 Analysis

2

Results

Contents

2.1	Pre Usage Results	9
2.2	System Usability Scale (SUS) Scores	12
2.3	Technology Acceptance Model	12
2.3.1	Course Learning, Recommendations, and more	13
2.3.2	Italics and bold	17
2.3.3	Hyperlinks	17
2.3.4	Footnotes	17
2.3.5	Comments	17
2.3.6	Including plots	18
2.3.7	Including tables	19
2.3.8	Control positioning	20
2.4	Executable inline code	21
2.5	Executable code in other languages than R	22

Have users in previous years shared the HELM Open RLO catalogue?

Well, we have so much data we haven't looked through it all yet.

33,571 learners told us how they found out- each answer is different.

We are through about 10% of this data and will update weekly.

2.1 Pre Usage Results

```
# A tibble: 41 x 2
```

2. Results

profession	hours
<chr>	<chr>
1 Student on a Healthcare course	1-4 hours
2 Student on a Healthcare course	Never
3 Student on a Healthcare course	Never
4 Student on a Healthcare course	Never
5 College student	1-4 hours
6 Student on a Healthcare course	1-4 hours
7 Student on a Healthcare course	Never
8 Student on a Healthcare course	Never
9 Student on a Healthcare course	Never
10 Lecturer	Never
# ... with 31 more rows	

2. Results

profession	hours
Student on a Healthcare course	1-4 hours
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	Never
College student	1-4 hours
Student on a Healthcare course	1-4 hours
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Lecturer	Never
Learning Technologist	20+ hours
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	1-4 hours
Student on a Healthcare course	1-4 hours
Student on a Healthcare course	1-4 hours
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	1-4 hours
Lecturer	1-4 hours
Student on a Healthcare course	Never
Student on a Healthcare course	1-4 hours
College student	1-4 hours
Postgraduate student	1-4 hours
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Student on a Healthcare course	Never
Mature Student	Never
Postgraduate student	1-4 hours
Doctor	5-9 hours
Doctor	10-19 hours
Lecturer	1-4 hours
College student	Never
Medical doctor	1-4 hours
Learning Technologist	Never
Student on a Healthcare course	5-9 hours
Student on a Healthcare course	Never
Student on a Healthcare course	1-4 hours

2. Results

2.2 System Usability Scale (SUS) Scores

Note= The amount of ‘agreement’ is defined as the addition of ‘Agree’ and ‘Strongly agree’ responses.

The SUS score for all data was XXX. This is within, and above the median of, 68 – which is in the range of ‘average’ usability. This is good as the resources were early demonstrations and had reduced beta alpha testing due to time constraints- future updates can improve this metric.

After reversing the scores of the negatively worded questions (odd numbered questions), participants strongly agreed the system was not complex (XX% agreements), and they did not need assistance before use (XX% agreements). All remaining questions has the most frequently observed response as ‘agree’- the lowest amount of agreement (agree and strongly agree) was XX% for question X, which was explored further in the individual Partners’ analyses.

if you don’t like boring tables, here is the same data in a graph!

2.3 Technology Acceptance Model

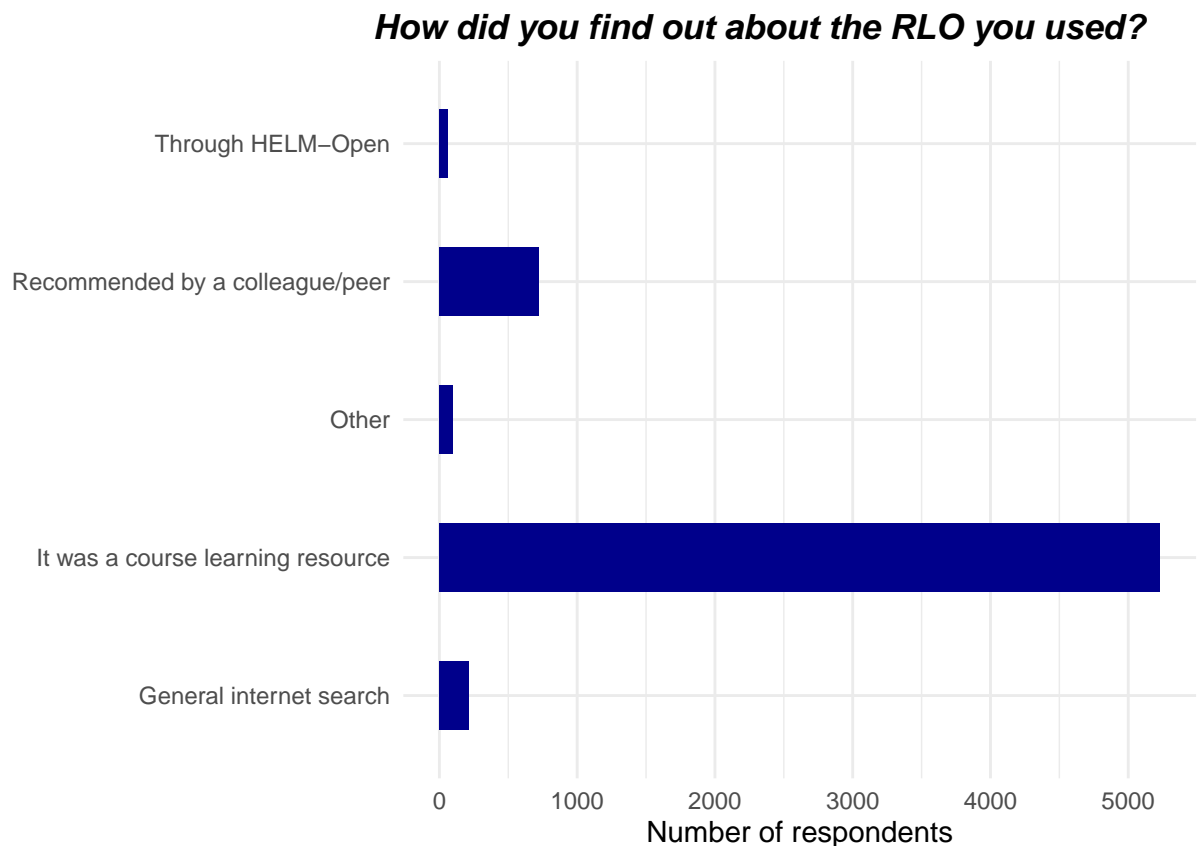
The TAM had 3 sections (Ease of Use, Perceived Usefulness, and Intention of Use). Ease of Use results showed significant increases in Users’ usage with each Chatbot. Perceived Usefulness: There were not significant findings for the Perceived usefulness. The justification for this may be due to being early versions of applications with limited functionality and functions which can be difficult for user to experience the intended further range of features and learning exercises. Intention of Use: For users’ intentions to use within their course, the result of the Mann-Whitney U test was not significant, $U =$, $z =$, $p =$. in their intentions before use (m=xx, mode=xx) compared to after (m=xx, mode=x), however there was improvement therefore the chatbots may have more benefit than expected by students.

2. Results

2.3.1 Course Learning, Recommendations, and more

The data showed that learners *strongly recommend* the RLO(s) they used to others, but how does this translate?

For the 10% of data we have, the figure below '*How did you find out about the RLO you used*' shows 700 respondents were recommended from a friend, peer, tutor, or other.



This figure also shows how more than 6000 respondents first used the RLOs as instructed by their tutors on their course.

and if you're browsing the internet for information on a healthcare topic and come across <https://www.nottingham.ac.uk/helmopen/> :- You're 1 of about 350 people finding about our resources from internet search. Hopefully that grows, but it seems social networking is the key to sharing these tools.

A random sample of other sources are: Twitter, Aim higher days, Barnardos ignite learning, and, well, '*a random Google photo*;' - our online presence seems

2. Results

to be in many places!

UP TO HERE1

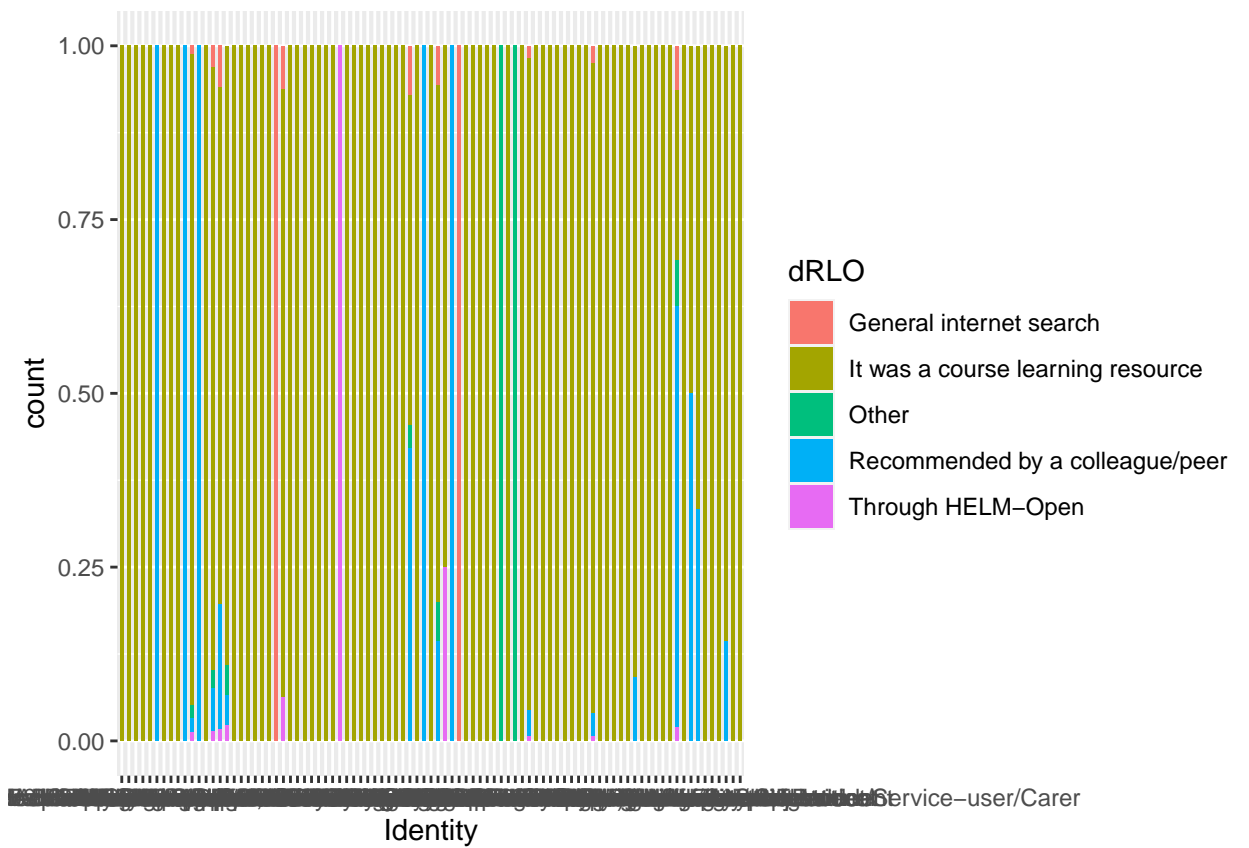
```
# A tibble: 24 x 3
```

```
# Groups:   Identity, dRLO [24]
```

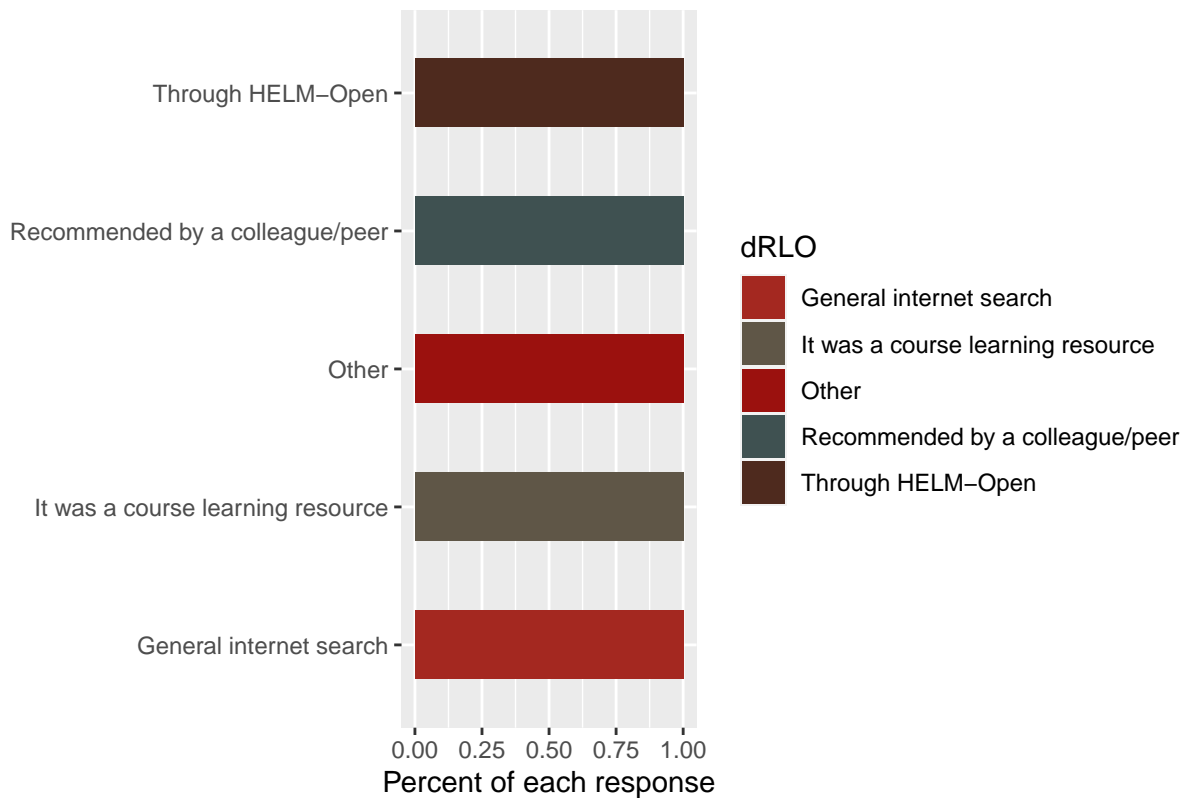
	Identity	dRLO	n
	<chr>	<chr>	<int>
1	HE student	It was a course learning resource	80
2	HE student	Other	2
3	HE student	Recommended by a colleague/peer	4
4	HE student	Through HELM-Open	1
5	Healthcare professional	General internet search	7
6	Healthcare professional	It was a course learning resource	131
7	Healthcare professional	Other	3
8	Healthcare professional	Recommended by a colleague/peer	10
9	Other	General internet search	8
10	Other	It was a course learning resource	50

```
# ... with 14 more rows
```

2. Results



How helpful has this learning object been?



2. Results

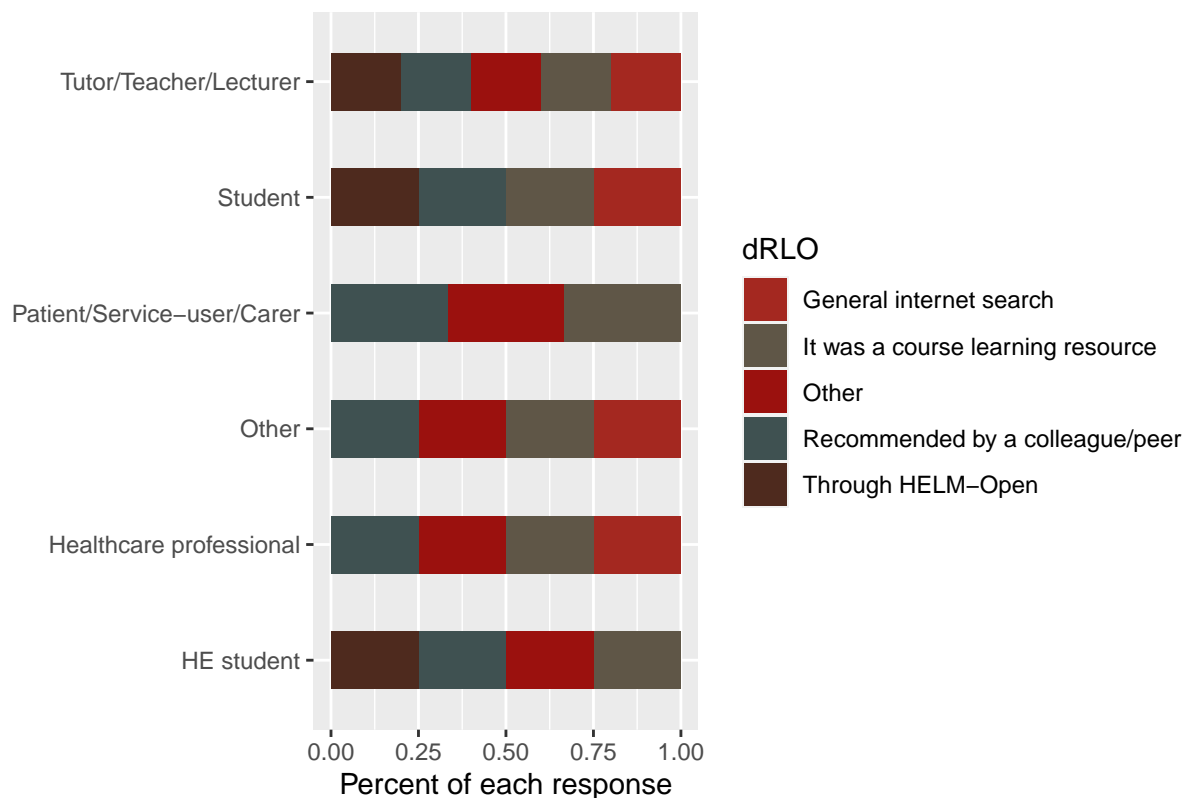
A tibble: 24 x 3

Groups: Identity, dRLO [24]

Identity	dRLO	n
<chr>	<chr>	<int>
1 HE student	It was a course learning resource	80
2 HE student	Other	2
3 HE student	Recommended by a colleague/peer	4
4 HE student	Through HELM-Open	1
5 Healthcare professional	General internet search	7
6 Healthcare professional	It was a course learning resource	131
7 Healthcare professional	Other	3
8 Healthcare professional	Recommended by a colleague/peer	10
9 Other	General internet search	8
10 Other	It was a course learning resource	50

... with 14 more rows

How did you find out about this resource?



2. Results

2.3.2 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_**`

2.3.3 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)`

2.3.4 Footnotes

- Are created¹ 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.`

2.3.5 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, .

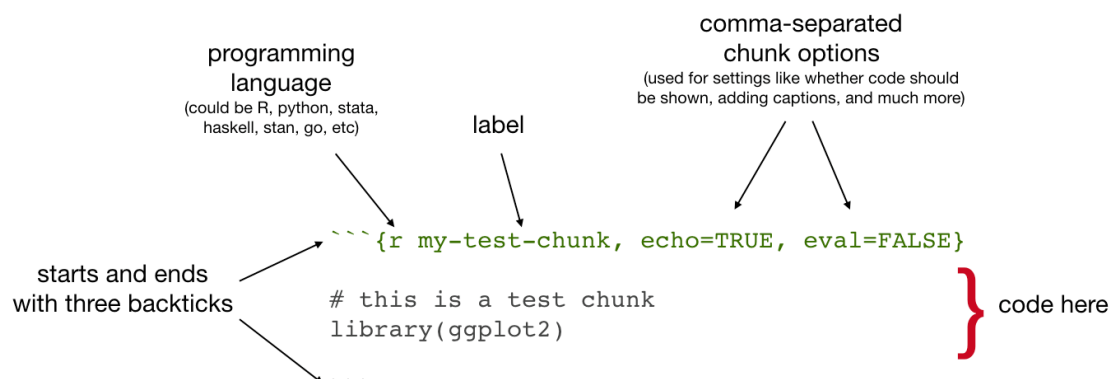


Figure 2.1: Code chunk syntax

¹my footnote text

²This is a random test.

2. Results



Figure 2.2: Oxford logo

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

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:

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

2.3.6 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

2. Results



Figure 2.3: Oxford logo, rotated

Table 2.1: A knitr kable table

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

of stopping distances for cars at various speeds (this dataset is built in to **R**).

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.

2.3.7 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.

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

2. Results

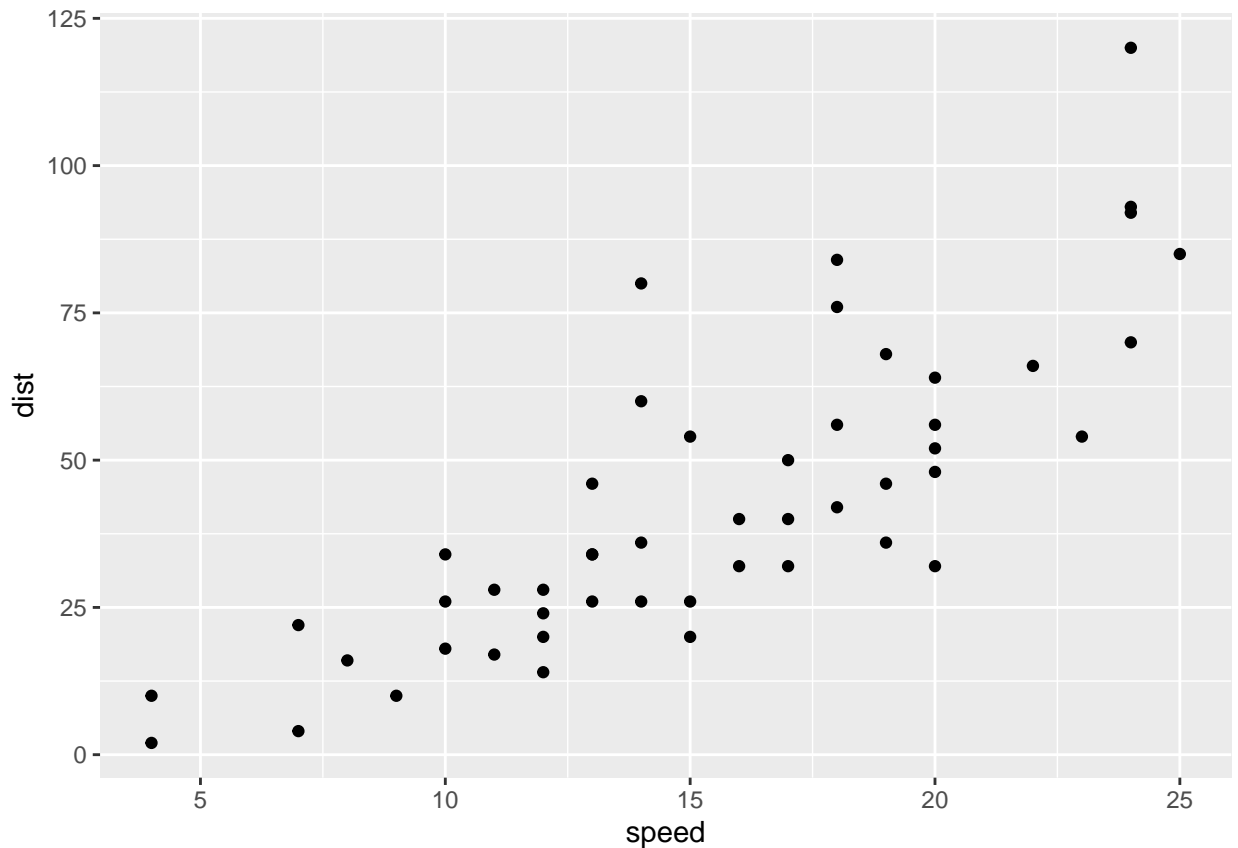


Figure 2.4: A ggplot of car stuff

2.3.8 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:



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 5.1 in section 5.1.4). For more details, read the relevant section of the [R Markdown Cookbook](#).

2.4 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:

We might then write:

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

2. Results

Which would output:

“In the `cars` dataset, we have 50 observations.”

2.5 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

Training Event Results

Contents

3.1 CEPEH Training Event C1	23
--	-----------

3.1 CEPEH Training Event C1

The CEPEH training event C1 held at the premises of University of Nottingham aiming to prepare participants for the practical elements of co-creation and implementation of chatbots as an educational resource. It combined both theoretical and hands-on training. 15 participants were from RISE, AUTH, UoN.

Project managers of partners signposted the person involved, and relevant announcements were made through social media channels to the wider public. External to the project speakers were from University of Leeds, and Computer Science Department of University of Nottingham. It included academics, medical doctors, and researchers with focus both on clinical research and digital innovations in healthcare education and IT specialist/learning technologists 11.18 years of experience (SD=7.2). A balance between male and female participants achieved.

4

Overall Training Events Evaluation

Participants were asked to highlight what they liked for each day and how each day can be improved. Findings are described below per day of the training event

Day 1

The participants comment that they liked the design method for educational resources presented using a co-creation approach, they liked the interactions with other groups, and they liked the overview of existing chatbot resources of the partners. On the areas that can be improved, more media material were requested.

Day 2 Participants enjoyed the presentation from the invited speaker from another faculty of the University of Nottingham, the CEPEH resources presented and the storyboarding process. Participants highlighted that the participation of more clinicians in the event would be an added value in regards with the storyboarding process.

Day3 Participants liked the hands-on activities of the day also enjoyed the creativity of the groups on the online chatbot development tool. As an area of improvement, participants wanted more time on hands on sections.

4.1 CEPEH Training Event 2

Pre-Training Event survey May 9th-13th 2022 Thessaloniki, Greece

4. Overall Training Events Evaluation

Twenty-six participants attended the Training Event, along with approximately 10 staff members. There were 21 undergraduate students and 5 postgraduate students, who completed the survey for a total of 26 responses. There were 86% of participants who stated they had not been to a similar event like the training event CEPEH facilitated. There were 90% of students who found the event schedule very organised, and 70% agreed most of the planned sessions were relevant to that interest with the remaining 30% not having enough experience to understand the context to determine if they are interested in the training event. There were 95% of students agreeing or strongly agreeing the training event location is great, the remaining person did not leave additional comments.

Table 1 suggested attendees had minimal intention to share their own ideas due to lack of previous experience of attending such events, or due to lack of knowledge on the area. However, most were interested in listening to other groups and hearing contextual cases in healthcare.

There were 77% of participants stated they were novices in experience with chatbots in healthcare and were attending to learn more. The remaining 23% (7 students) stated they were competent and had limited experience with chatbots in healthcare.

One day had several events regarding cybersecurity in healthcare. When asked before these events, 83% stated they were neutral or disagreed that they felt confident about their cybersecurity knowledge in healthcare. In addition, 80% stated they were neutral or disagreed that they felt they had strong cybersecurity safety in healthcare. Table 2 shows the main pre and post results suggesting a positive experience for more than 75% of attendees on all measures.

There were 90% (23) of students who heard about the event through a lecturer or a professor, the CEPEH newsletter (2), and 1 person was informed through the anatomy tutoring system at Karolinska Institute. Additionally, 60% suggested the training event to somebody else before the course started.

There were six individuals who stated neutral or disagree when asked if having issues on registration or finding the information for the event. This may have

4. Overall Training Events Evaluation

been due to being dependent on emails to receive the information, instead of a dedicated website where the information is available anytime.

As this was face-to-face, participants were asked about sufficient Covid-19 precautions in place at the facility, 94% agreed with sufficient precautions, two individuals stated no but did not give further information in the additional input box provided. In summary, most participants were undergraduate students with novice experience, happy with the training event location, felt the sessions were relevant to them, and most shared the event with their colleagues. The values of co-creation, chatbots in healthcare, and taking patient history were bestowed to students in an engaging and well-received manner. Notably, the highest ratings were for staff friendliness which is key to engagement and consistent interaction throughout the intense and long 5-day duration. The sessions were recorded there for the online recordings may be viewed with higher numbers over the subsequent weeks.

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**.

Most reference managers can create a .bib file with your 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},
```

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

4. Overall Training Events Evaluation

```
doi = {10.1016/j.tics.2014.01.006},  
}
```

In this entry highlighted 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 (Lottridge et al., 2012; Mill, 1965 [1843]; Shea et al., 2014).

4.1.1 Appearance of citations and references section (pandoc)

By default, `oxforddown` lets `Pandoc` handle how citations are inserted in your text and the references section. You can change the appearance of citations and references by specifying a CSL (Citation Style Language) file in the `cs1` metadata field of `index.Rmd`. By default, `oxforddown` by the American Psychological Association (7th Edition), which is an author-year format.

With this style, a number of variations on the citation syntax are useful to know:

- Put author names outside the parenthesis
 - This: `@Shea2014 says blah.`
 - Becomes: Shea et al. (2014) says blah.
- Include only the citation-year (in parenthesis)
 - This: `Shea et al. says blah [-@Shea2014]`
 - Becomes: Shea et al. says blah (2014)
- Add text and page or chapter references to the citation
 - This: `[see @Shea2014, pp. 33–35; also @Wu2016, ch. 1]`
 - Becomes: Blah blah (see Shea et al., 2014, pp. 33–35; also Wu, 2016, ch. 1).

4. Overall Training Events Evalaution

If you want a numerical citation style instead, try `csl: bibliography/transactions-on-comput` or just have a browse through the [Zotero Style Repository](#) and look for one you like. For convenience, you can set the line spacing and the space between the bibliographic entries in the reference section directly from the YAML header in **index.Rmd**.

If you prefer to use `biblatex` or `natbib` to handle references, see [this chapter](#).

4. Overall Training Events Evaluation

4.1.2 Insert references easily with RStudio's Visual Editor

For an easy way to insert citations, use RStudio's [Visual Editor](#). Make sure you have the latest version of RStudio – the visual editor was originally really buggy, especially in relation to references, but as per v2022.02.0, it's great!

4.2 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)`

4.2.1 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 ???

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.

4. Overall Training Events Evalaution



Figure 4.1: A marvel-lous meme

4.2.2 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:

We refer to this image with `\@ref(fig:captain)`. So Figure 4.1 is [this image](#).

And in Figure 2.4 we saw a [cars](#) plot.

4.2.3 Table references

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

Let's include a table:

We refer to this table with `\@ref(tab:cars-table2)`. So Table 4.1 is [this table](#).

And in Table 2.1 we saw more or less [the same cars table](#).

4. Overall Training Events Evalaution

Table 4.1: Stopping cars

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

4.2.4 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.

4.3 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.

4.4 Additional resources

- *R Markdown: The Definitive Guide* - <https://bookdown.org/yihui/rmarkdown/>
- *R for Data Science* - <https://r4ds.had.co.nz>

5

Discussion

Contents

5.1	Reach, Impact, and Qualatative analysis	34
5.1.1	Making your table pretty	34
5.1.2	If your table is too wide	34
5.1.3	If your table is too long	35
5.1.4	Max power: manually adjust the raw LaTeX output . .	44

Here is a (very large) table with all of the currently active RLOS.

RLO Name	Location	Start Date	SubmissionDate	Y
01NCTLR Probability and inferential Statistics	NA	09/12/2010	09/12/2010	20
01NCTLR Probability and inferential Statistics	NA	13/04/2013	13/04/2013	20
01NCTLR Probability and inferential Statistics	NA	01/05/2014	01/05/2014	20
01NCTLR Probability and inferential Statistics	NA	04/05/2013	04/05/2013	20
01NCTLR Probability and inferential Statistics	NA	15/10/2011	15/10/2011	20
01NCTLR Probability and inferential Statistics	NA	05/01/2010	05/01/2010	20

Those results can be interpreted that the learning objectives of the training event was chosen appropriately for the diverse audience including clinicians, academics, researchers, and learning technologists/IT specialist resulting to a successful training event that enable participants to take the acquired knowledge back to their organisations in order to co-design and implement. As it was expected and can be depicted from self-confidence statements that some participants being

5. Discussion

very confident before the event, not all the objectives expected to be reached by everyone, since the training was targeting both technical and non-technical participants. However, on both average and individual matched responses participants self-statements showed that they improved their knowledge and understanding in using co-creation approaches to develop digital education resources and in designing and developing chatbots as educational resources.

5.1 Reach, Impact, and Qualatative analysis

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!)

5.1.1 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:

Compare this to the default style, which looks terrible:

	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

5.1.2 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:

This scales down the table to fit the page width.

5. Discussion

	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

5.1.3 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.

	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
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

5. Discussion

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
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`:

5. Discussion

	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

5. Discussion

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
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:

	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

5. Discussion

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
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 RX4i	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 710i	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 360i	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 230i	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280i	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 128i	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 Z28i	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 X1-9i	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2i	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

5. Discussion

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
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:

	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

5. Discussion

5.1.4 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 5.1.

<code>\tiny</code>	Lorem ipsum
<code>\scriptsize</code>	Lorem ipsum
<code>\footnotesize</code>	Lorem ipsum
<code>\small</code>	Lorem ipsum

Figure 5.1: 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)

	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

5. Discussion

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
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
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

6

Text Mining, Natural Language Processing, and Sentiment Analysis

6.1 1 Reading in texts

6.1.1 1.1 txt files

Here's how you can read in one .txt file that is saved in the same location as this script (i.e. in the same folder on your computer):

If you want to read all files from a sub-folder, type the name of the folder followed by / and * to ask R to read in all files in that folder:

6.1.2 1.3 Preparing data

- convert name to ID numbers with more descriptive labels

6.2 2 Tidy text

- One word per row, facilitates analysis
- Token: “a meaningful unit of text, most often a word, that we are interested in using for further analysis”

6.2.1 2.1 the `unnest_tokens` function

- Easy to convert from full text to token per row with `unnest_tokens()` Syntax:
`unnest_tokens(df, newcol, oldcol)`
- `unnest_tokens()` automatically removes punctuation and converts to lower-case (unless you set `to_lower = FALSE`)
- by default, tokens are set to words, but you can also use `token = "characters"`, "ngrams", "sentences", "lines", "regex", "paragraphs", and even "tweets" (which will retain usernames, hashtags, and URLs)

```
## readtext object consisting of 2858 documents and 0 docvars.
## # Description: df [2,858 x 3]
##   doc_id word      text
##   <fct> <chr>     <chr>
## 1 1      p1       "\"\"\"..."
## 2 1      for      "\"\"\"..."
## 3 1      me       "\"\"\"..."
## 4 1      personally "\"\"\"..."
## 5 1      it       "\"\"\"..."
## 6 1      was      "\"\"\"..."
## # ... with 2,852 more rows
```

6.2.2 2.2 Removing non-alphanumeric characters

- `str_extract` is used to get rid of non-alphanumeric characters (because we don't want to count *word* separately from word)

6.2.3 2.3 Stop words

- Stop words: very common, "meaningless" function words like "the", "of" and "to" – not usually important in an analysis (i.e. to find out that the most common word in two books you are comparing is "the")
- `tidytext` has a built-in df called `stop_words` for English

6. Text Mining, Natural Language Processing, and Sentiment Analysis

- remove these from your dataset with `anti_join`

We can take a look:

```
## # A tibble: 1,149 x 2
##   word      lexicon
##   <chr>     <chr>
## 1 a        SMART
## 2 a's      SMART
## 3 able     SMART
## 4 about    SMART
## 5 above    SMART
## 6 according SMART
## 7 accordingly SMART
## 8 across   SMART
## 9 actually SMART
## 10 after   SMART
## # ... with 1,139 more rows

## readtext object consisting of 821 documents and 0 docvars.
## # Description: df [821 x 3]
##   doc_id word      text
##   <fct> <chr>     <chr>
## 1 1      personally "\"\"..."
## 2 1      nice      "\"\"..."
## 3 1      week       "\"\"..."
## 4 1      ive        "\"\"..."
## 5 1      feeling    "\"\"..."
## 6 1      chatbots   "\"\"..."
## # ... with 815 more rows
```

Define other stop words:

6. Text Mining, Natural Language Processing, and Sentiment Analysis

Break: Prepare your data with the steps above. 1) Unnest tokens, 2) Remove alpha-numeric characters, 3) Remove stopwords

6.3 3 Analysing frequencies

6.3.1 3.1 Find most frequent words

- Easily find frequent words using `count()`
- Data must be in tidy format (one token per line)
- `sort = TRUE` to show the most frequent words first

```
tidy_books %>% count(word, sort = TRUE)
```

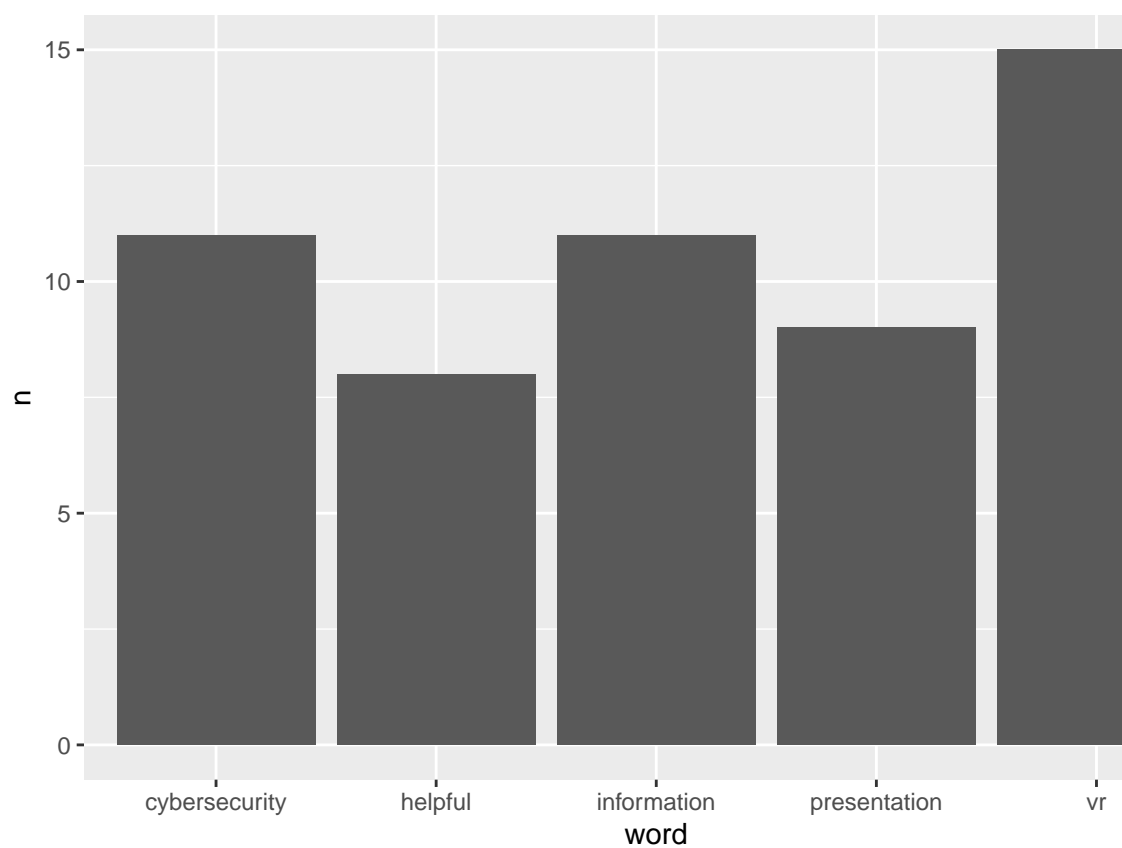
```
## # A tibble: 387 x 3
## # Groups:   doc_id [1]
##   doc_id word          n
##   <fct> <chr>        <int>
## 1 1      vr          15
## 2 1      cybersecurity 11
## 3 1      information 11
## 4 1      presentation 9
## 5 1      helpful      8
## 6 1      idea          7
## 7 1      ideas          7
## 8 1      lot           7
## 9 1      workshop      7
## 10 1     beginning      6
## # ... with 377 more rows
## # A tibble: 388 x 3
## # Groups:   doc_id [1]
##   doc_id word          n
##   <fct> <chr>        <int>
```

6. Text Mining, Natural Language Processing, and Sentiment Analysis

```
## 1 1      vr      15
## 2 1      cybersecurity 11
## 3 1      information 11
## 4 1      presentation 9
## 5 1      helpful 8
## 6 1      understand 8
## 7 1      idea 7
## 8 1      ideas 7
## 9 1      lot 7
## 10 1     workshop 7
## # ... with 378 more rows
```

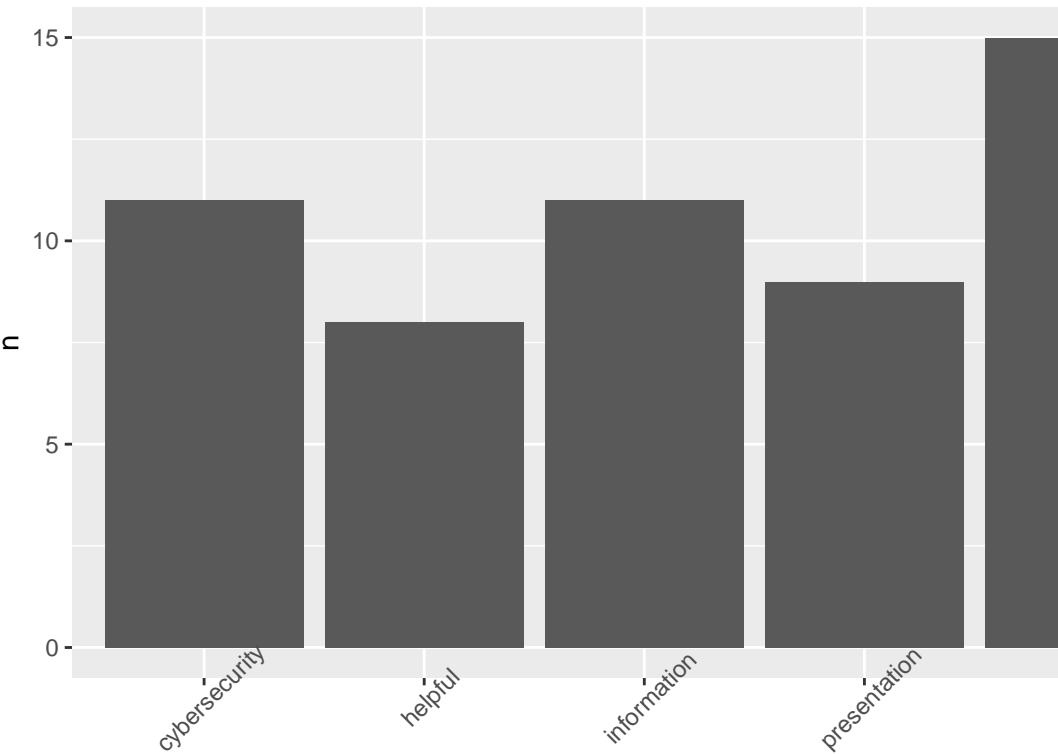
Plotting word frequencies - bar graphs

Bar graph of top words in CEPEHQ.



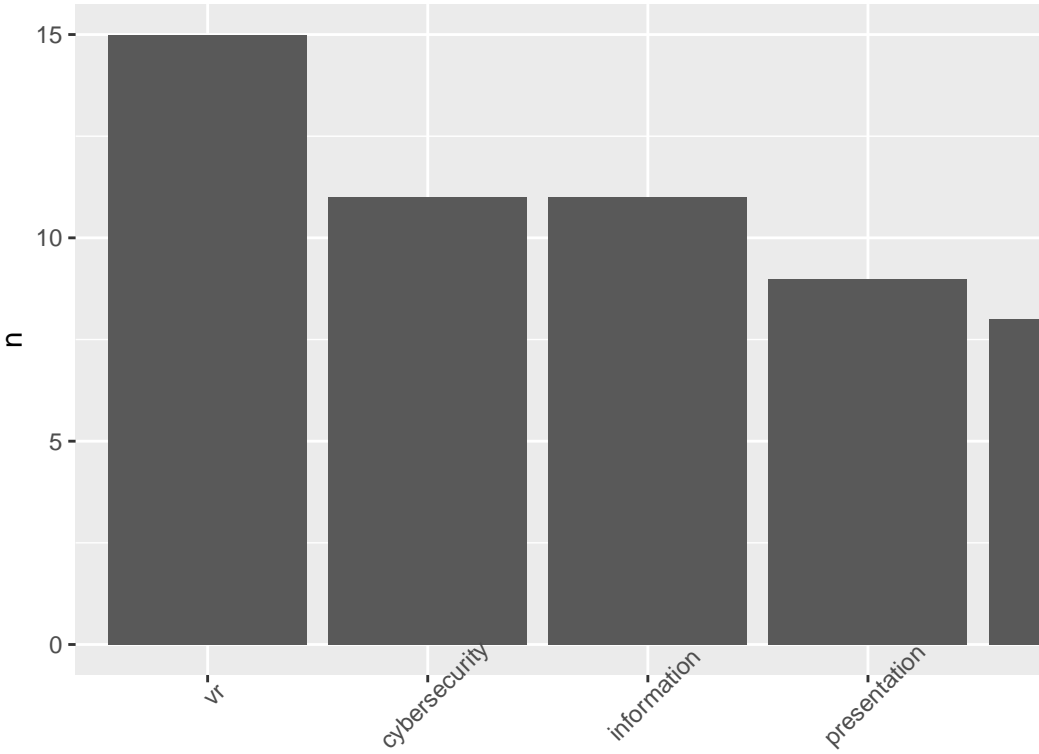
Basic graph:

6. Text Mining, Natural Language Processing, and Sentiment Analysis



Readable labels:

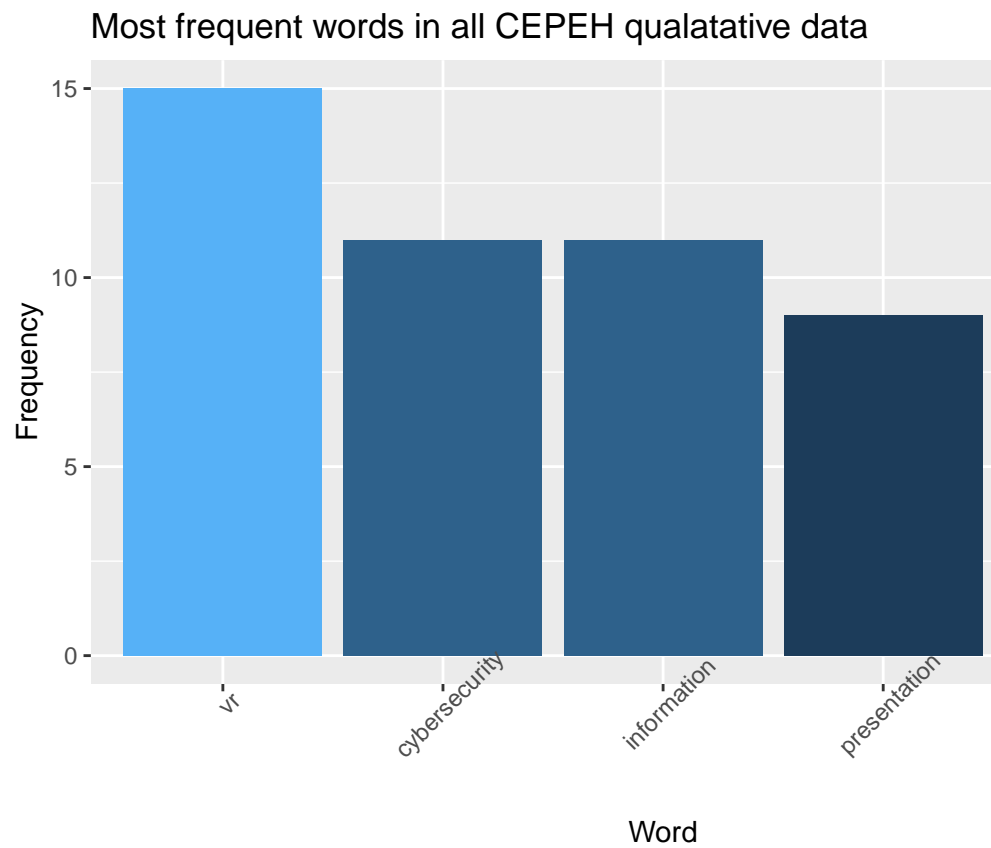
word



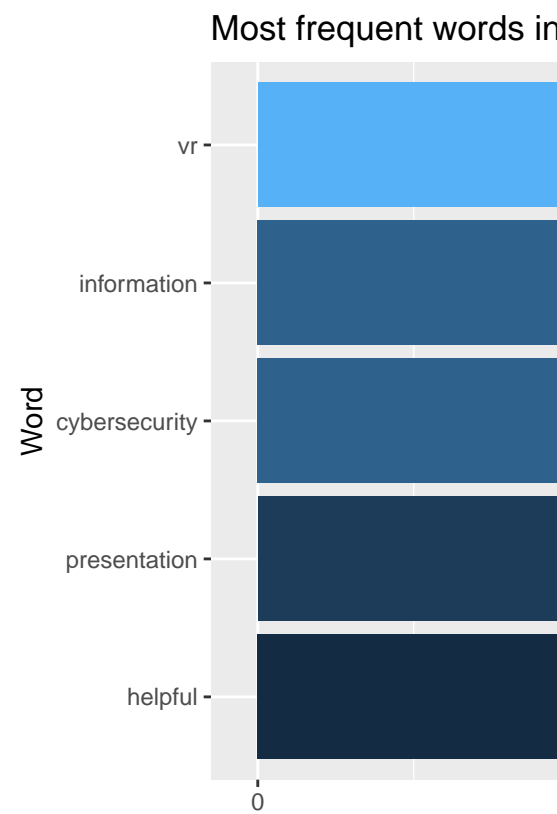
Descending order:

reorder(word, -n)

6. Text Mining, Natural Language Processing, and Sentiment Analysis



Axis names and colors:



Or: flip coordinate system to make more space for words

6.3.2 3.2 Normalised frequency

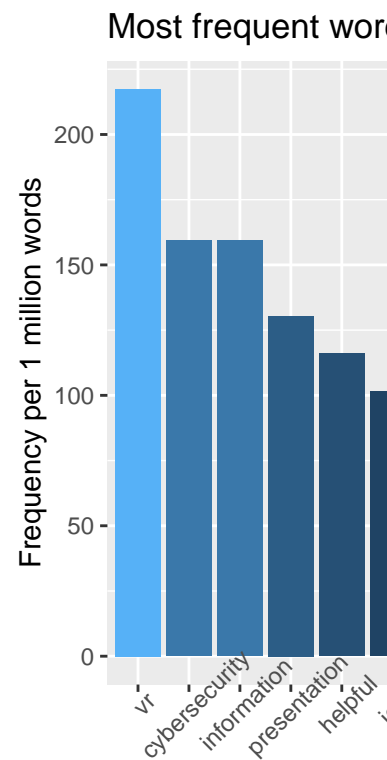
- when comparing the frequencies of words from different texts, they are commonly normalised
- convention in corpus linguistics: report the frequency per 1 million words
- for shorter texts: per 10,000 or per 100,000 words
- calculation: $\text{raw frequency} * 1,000,000 / \text{total numbers in text}$

```
## # A tibble: 1 x 2
## # Groups:   doc_id [1]
##   doc_id `sum(n)`
##   <fct>     <int>
## 1 1          696

## # A tibble: 386 x 2
##   word          pmw
##   <chr>         <dbl>
## 1 vr          217.
## 2 cybersecurity 159.
## 3 information   159.
## 4 presentation 130.
## 5 helpful      116.
## 6 idea         101.
## 7 ideas        101.
## 8 lot          101.
## 9 workshop     101.
## 10 beginning    87.0
## # ... with 376 more rows
```

6. Text Mining, Natural Language Processing, and Sentiment Analysis

Plotting normalised frequency



Now we can plot, for example, the 20 most frequent words (by pmw).

6.3.3 3.3 Word clouds

Let's visualise the most frequent words in a word cloud. Here, the size indicates the frequency, with words that occur more often being displayed in a larger font size, but this can also be used to visualise e.g. normalised frequency (pmw) or length or



anything else you pass to the `freq =` part of the command.

6.4 4 Comparing the vocabulary of texts

Next, we'll create two graphs to compare the vocabulary of our texts. First, we focus on Alice's Adventures and Anderson's CEPEHQ. The newly created `comp_2` data frame contains only the words and their frequencies in the two texts in two separate columns.

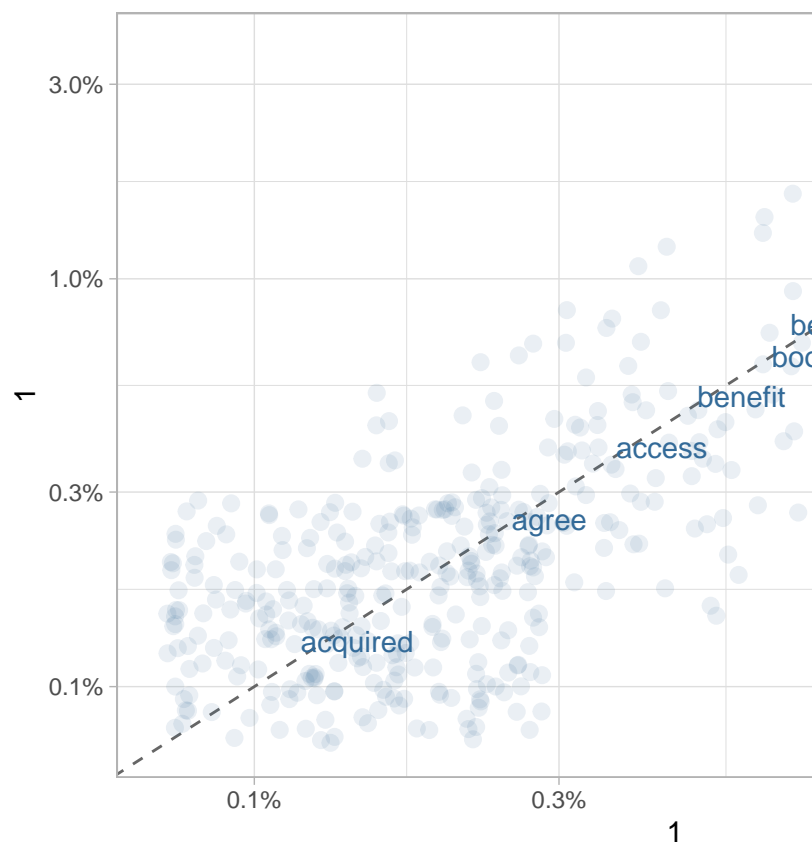
6.4.1 Comparing two texts

```
## # A tibble: 6 x 3
##   word          pmw      `1`
##   <chr>        <dbl>   <dbl>
## 1 access      43.5 0.00435
## 2 acquired    14.5 0.00145
## 3 add         14.5 0.00145
```


6. Text Mining, Natural Language Processing, and Sentiment Analysis

```
## 4 administrator 14.5 0.00145
## 5 advance       14.5 0.00145
## 6 advanced      14.5 0.00145
```

Now, we can plot the words. Their placement depends on the word frequencies. Additionally, colour coding shows how different the frequencies are - darker items are more similar in terms of their frequencies, lighter-coloured ones more frequent in one text compared to the other. We'll discuss the interpretation in more detail once



we've created the threeway comparison.

7

Sentiment analysis

```
# 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
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 Lyngs¹, Kai Lukoff², Petr Slovak³, William Seymour¹, Helena Webb¹,
Marina Jirotko¹, Jun Zhao¹, Max Van Kleek¹, Nigel Shadbolt¹

¹Department of Computer Science, University of Oxford, UK, {first.last}@cs.ox.ac.uk

²Human Centered Design & Engineering, University of Washington, Seattle, US, kai.l@uw.edu

³Department 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].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI ’20, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-6708-0/20/04.

<http://dx.doi.org/10.1145/3313831.3376672>

7. Sentiment analysis

CHI 2020 Paper

CHI 2020, April 25–30, 2020, Honolulu, HI, USA

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 Bányaí, 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>

7. Sentiment analysis

<p>Title of your Thesis</p> <p>John Doe</p>	<p>Thesis committee</p> <p>Promotor: Prof.dr. J. Smith Professor of Geo-information Science and Remote Sensing Wageningen University</p> <p>Co-promotors: Dr. Name of co-promotor Assistant Professor, Laboratory of Geo-information Science and Remote Sensing Wageningen University</p> <p>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</p> <p>This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PERC)</p>	<p>Title of your thesis</p> <p>John Doe</p> <p>Thesis submitted in fulfillment 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</p>
<p>John Doe Title of your thesis 77 pages PhD thesis, Wageningen University, Wageningen, NL (2015) With references, with summary in English ISBN XXX-YYY</p>	<p>For Yihai Xie</p>	<p>Acknowledgements</p> <p>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 thanks 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 thses in R Markdown.</p> <p>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 Yihai 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.</p> <p>Special thanks to Charles Ismay, who created the <i>thesisdown</i> package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John McManis, whose adoption of Stan Evans' adaptation of Keith Gilmer's original <i>maths</i> template for writing an Oxford University DPhil thesis in L^AT_EX provided the template that I in turn adapted for R Markdown.</p> <p>Finally, please 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.</p> <p>Ulrik Lenga Linacre College, Oxford 2 December 2018</p>

Appendix

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).

Appendices



The First Appendix

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

In 02-rmd-basics-code.Rmd

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

And here's another one from the same chapter, i.e. Chapter ??:

```
knitr::include_graphics("templates/download.png")
```


B

The Second Appendix, for Fun

References

- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. John Murray.
- Goethe, J. W. von. (1829). *Wilhelm Meisters Wanderjahre oder die Entsagen-den*. Cotta.
- Lottridge, D., Marschner, E., Wang, E., Romanovsky, M., & Nass, C. (2012). Browser design impacts multitasking. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. <https://doi.org/10.1177/1071181312561289>
- Mill, J. S. (1965 [1843]). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. Longmans.
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4), 186–193. <https://doi.org/10.1016/j.tics.2014.01.006>
- Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Knopf Publishing Group.