

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	6
1.4 Materials and Measures	7
2 Results	12
2.1 Learner Characteristics1	12
2.2 Chatbot Usability Questionnaire (CUQ)	19
2.3 System Usability Scale (SUS) Scores	20
2.4 Technology Acceptance Model	21
3 Text Mining, Natural Language Processing, and Sentiment Analysis	29
3.1 CEPEH Qualitative Feedback	29
3.2 Tokenising	29
3.3 The vocabulary of Texts	37
3.4 Sentiment analysis	38
3.5 Cross-referencing	66
3.6 Collaborative writing	68
3.7 Additional resources	69

Contents

Appendix	70
Appendices	
A The First Appendix	72
B The Second Appendix, for Fun	73
References	74

List of Figures

2.1	A marvel-lous meme	20
-----	------------------------------	----

List of Tables

3.1	Stopping cars	67
-----	-------------------------	----

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

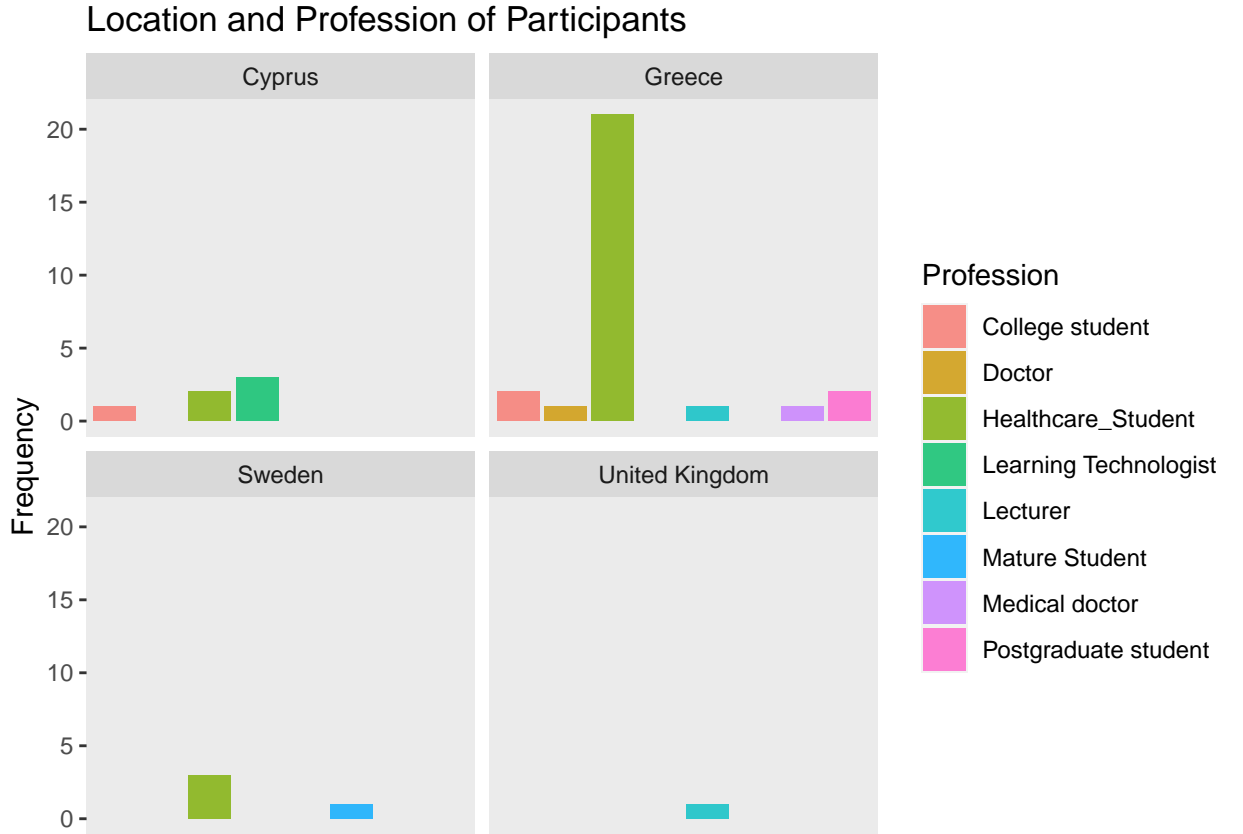
1.1	Participants	4
1.2	Procedure	5
1.3	Design	6
1.4	Materials and Measures	7
1.4.1	Chatbot Usability Questionnaire (CUQ)	7
1.4.2	UTAUT2 (Unified Theory of Acceptance and Use of Technology)	7
1.4.3	System Usability Scale	8
1.4.4	Computer Self-Efficacy Scale Tool	9
1.4.5	Technology Acceptance Model (TAM)	9
1.4.6	Qualitative Measure- Focus Group Discussions	10

1.1 Participants

This dataset had 14 males and 28 females therefore a total of 42 participants. It was a repeated measure design whereby each participant used the 4 chatbots developed by the CEPEH team. Therefore, there are 42 points of data in the condition before testing, and 126 data points after testing the chatbots- for a total of 168 row of data, 5 per participant. There were 78 questions asked in total, therefore the full dataset has over 4000 cells recorded.

1. Method

There were 22 females and 7 males from Greece. There were 3 females and 4 males from Cyprus. There were 2 females and 2 males from Sweden, and there were 2 participants from the United Kingdom.



The majority 36 participants, were student, with 3 being learning technologists, 2 were lecturers, and 1 was a doctor. Although there could be a difference in these groups, the design was within- groups therefore each participants pre-usage metrics were the comparative control data, and participant differences did not affect the evaluation.

1.2 Procedure

For each resource created by the Partners, the same experimental methodology was followed. For each resource created by partners, students performed a study within an online or face to face workshop or course. Student participants joined from Greece, Cyprus, Sweden, and the United Kingdom. A repeated measures

1. Method

design was used as the same group measures were taken before and after usage of the chatbots. They were recruited via staff members in the CEPEH group.

Participants were asked prior to the study if they agree to participate, providing them with a PIS form. Participants had the opportunity to discuss with the research team prior to the study and before consent is given. Then, participants used the chatbot resources independently and technical support was provided. Finally, post-intervention measures were recorded.

Some of the participants were invited to participate in Focus Group Discussions (FGD), and each FGD lasted between 15 to 25 minutes, with 5-10 participants. Participants were asked if they would like to be informed of the findings of the study.

1.3 Design

The data captured from the participants were their initials and numerical day of birth, used as anonymous identifier for pre-post analysis. Their institution was captured (Aristotle University of Thessaloniki, CYENS Centre of Excellent, Karolinska Institute, and The University of Nottingham), and Sex (Male/Female/Other).

Before any interaction with the learning resources, various perceptions of chatbot such as confidence and easy of use, usefulness, Influence from others, and current learning resources (videos, textbooks, Google, friends etc), were captured. Descriptive data was produced alongside repeated measures t-tests. Repeated measures t-tests were the appropriate test to use as this explores differences between groups, there were no covariates and we did not have several dependant variables. There was one Independent factor being Chatbot use having 2 levels (pre/post). There were 3 chatbots therefore there was option for ANOVA to determine where differences lie if statistical differences were found however this was not wholly appropriate for the data type and not necessary for pre-post comparison.

1.4 Materials and Measures

The measures used fit within a newly developed Chatbot Evaluation Framework which takes the best measures of 5 previous frameworks. Denecke and Warren [2] derived several quality dimensions and attributes from previous chatbot literature. They formed six perspectives from their review of articles and mobile health applications.

These six perspectives were: 1) Task-oriented, 2) Artificial intelligence, 3) System quality perspective, 4) Linguistic perspective, 5) UX Perspective, 6) Healthcare quality perspective.

To capture these perspectives, we used several validated materials that can distinguish these elements of the CEPEH chatbots.

1.4.1 Chatbot Usability Questionnaire (CUQ)

The Chatbot Usability Questionnaire (CUQ) [4] is a new questionnaire specifically designed for measuring the usability of chatbots by an interdisciplinary team from the Ulster University. CUQ can be used alongside the prevalent System Usability Scale Score (SUS) [5]. Multiple metrics are more appropriate when measuring usability of chatbots [6] therefore a combination of two scores can provide an all-inclusive overview.

1.4.2 UTAUT2 (Unified Theory of Acceptance and Use of Technology)

The underpinning theory of the UTAUT2 is that there are four key constructs to the intentions of using technology based resources: 1) performance expectancy, 2) effort expectancy, 3) social influence, and 4) enabling conditions.

The TAM and the UTAUT2 have cross over in measuring technology acceptance, however the UTAUT2 has more applied probing questions. Few studies exist that use technology acceptance theories for the intention to use products that explicitly incorporate AI. A recent extension of the UTAUT2 model added five (health,

1. Method

convenience comfort, sustainability, safety, security, and personal innovativeness) additional influencing factors to accommodate for AI [7]. This can be used for products in either health, household use, or mobility and can help to explain behavioural intention and use behaviour of chatbots.

1.4.3 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

1.4.4 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

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

1. Method

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.

1.4.6 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



Figure 1: Flow diagram of the recruitment process

2

Results

Contents

2.1	Learner Characteristics1	12
2.2	Chatbot Usability Questionnaire (CUQ)	19
2.2.1	CUQ Calculation tool	19
2.3	System Usability Scale (SUS) Scores	20
2.4	Technology Acceptance Model	21
2.4.1	Other Findings	22

2.1 Learner Characteristics1

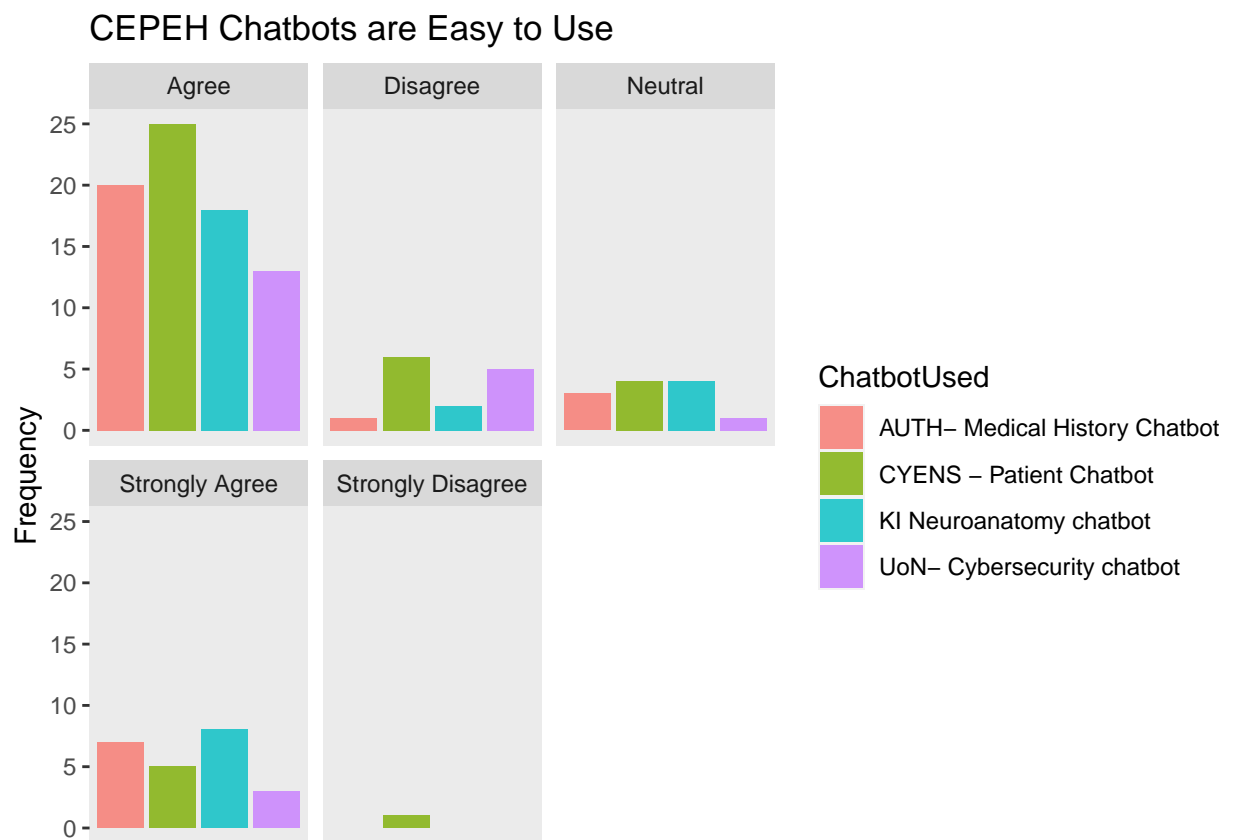
Most learners use books or online books as resources. Of course, they may use multiple sources however they were asked to note the primary source. Only 6 stated their primary sources were *Online videos/interactive materials* which includes such tools as chatbots.

This can be supported when asked the amount of time participants have used a chatbot- in any form or subject: 23 stated they had never used a chatbot, being educational or not. 2 individuals had spent what would be extensive time with usage- these were the Learning Technologist and Mature Student. Therefore, we can state that the sample used did not regularly use chatbots for their course learning,

2. Results

with 18/42 having used a chatbot at least once for between 0-4 hours of use in total.

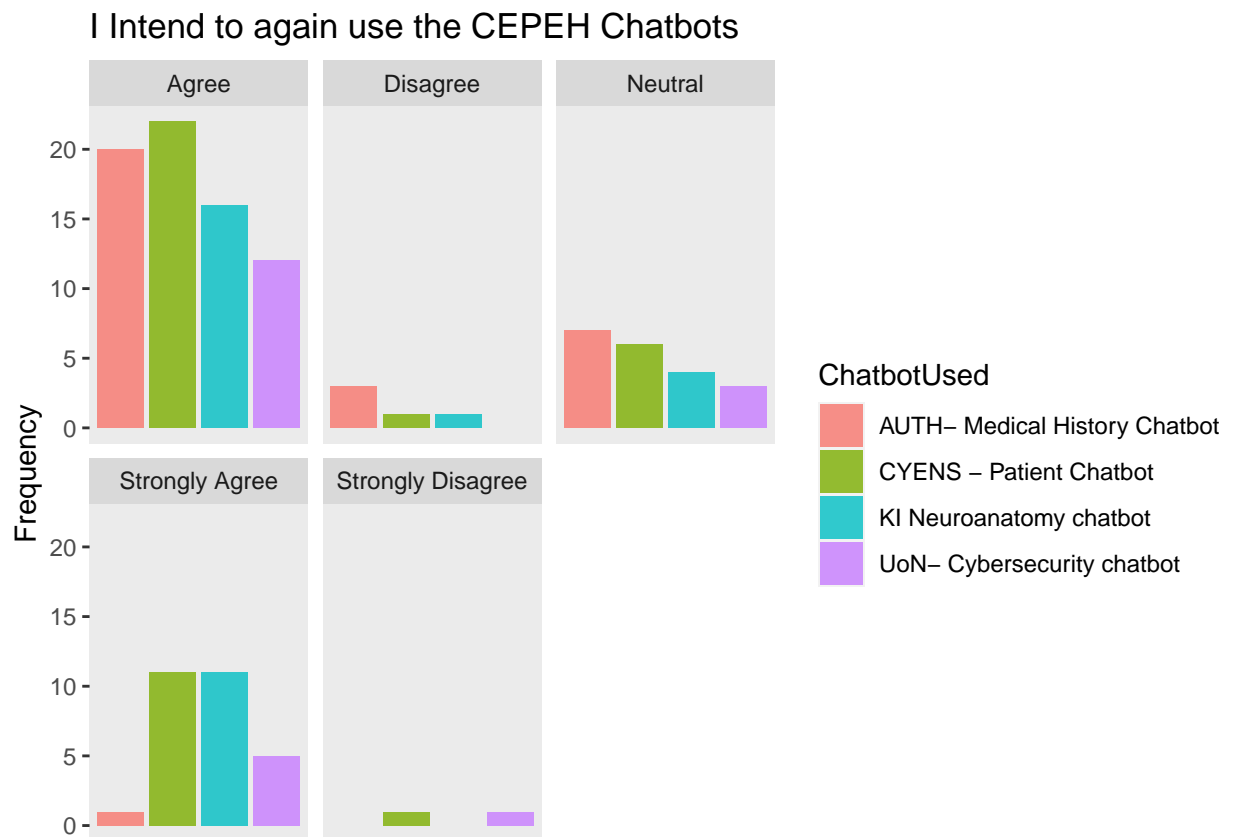
The first boxplot here shows



There was only 1 'Strongly Disagree' response. The agreement options counted

for the majority of the data.

2. Results

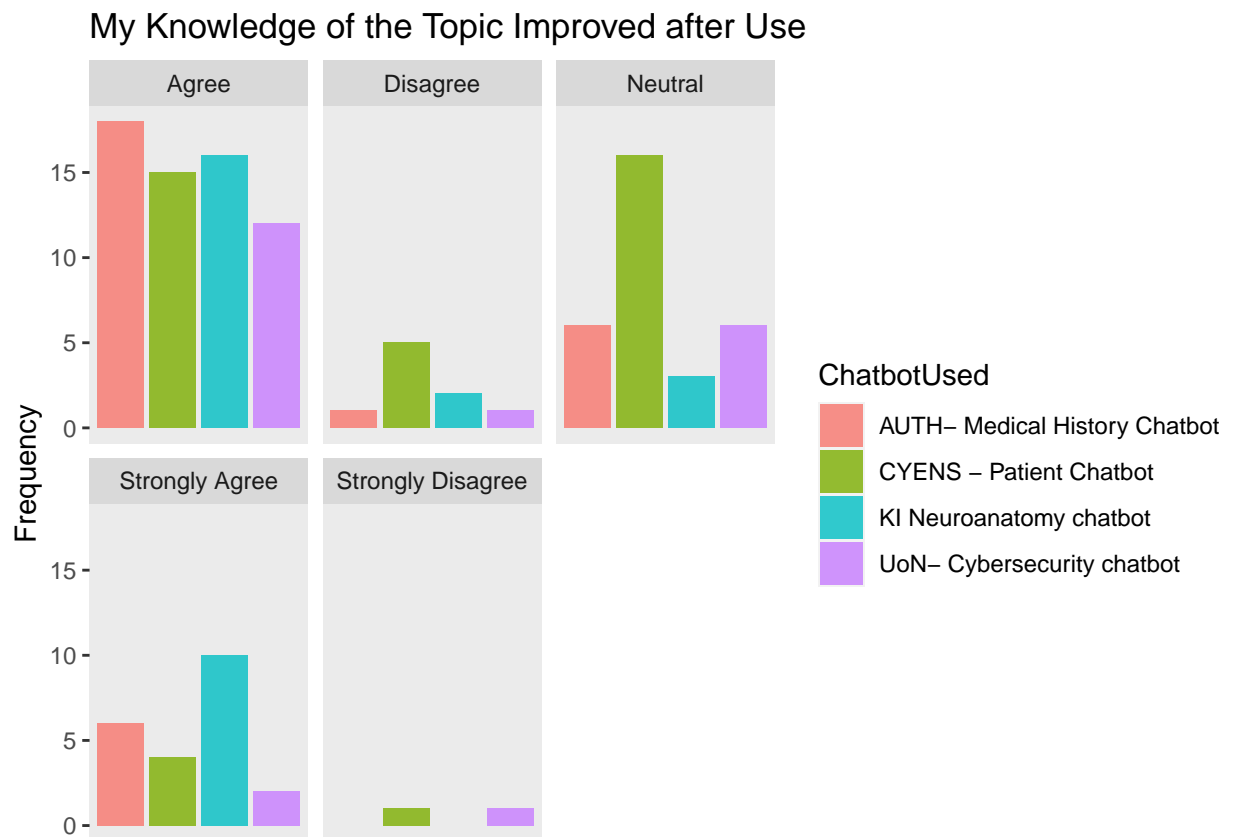


For CYENS, even though the knowledge of the topic was not perceived to improve

by some participants, this box plot shows how 34/42 stated they would reuse the

chatbot developed by CYENS.

2. Results

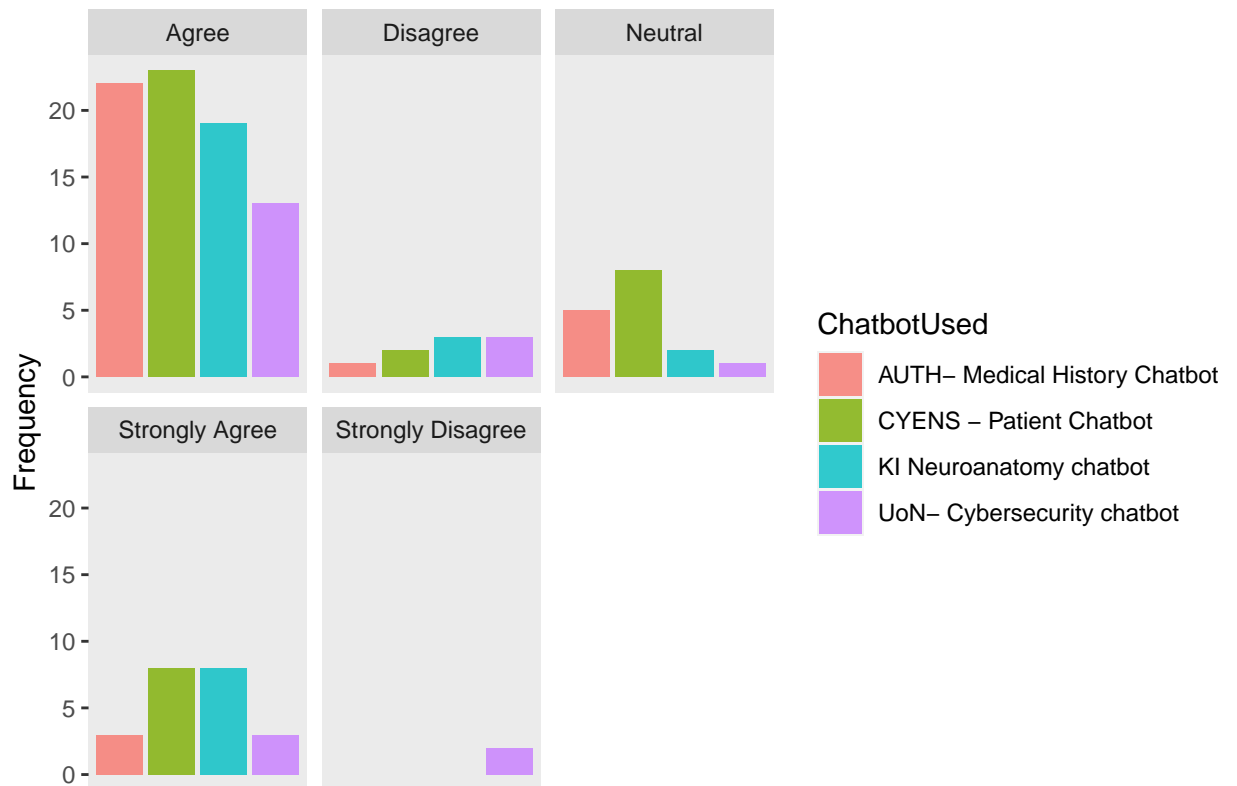


CYENS chatbot had around 10 more participants stating that they were neutral

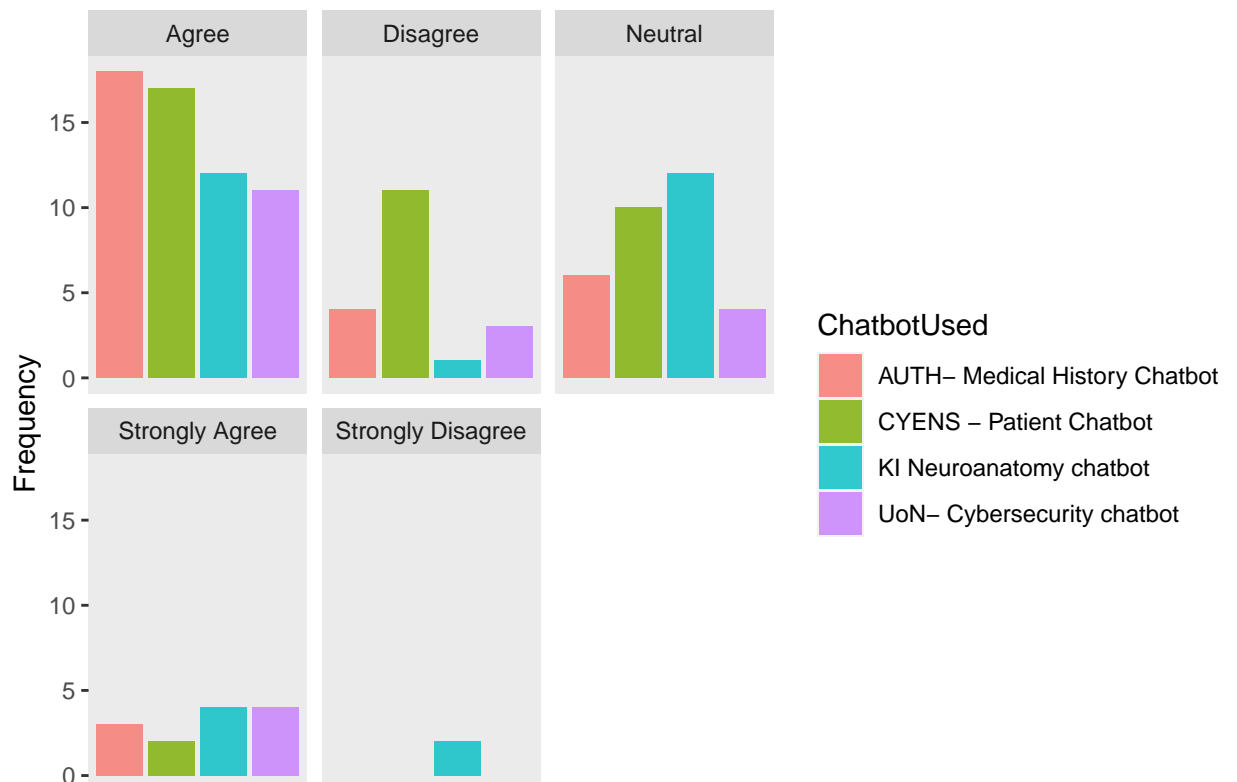
on gaining knowledge oo the topic

2. Results

I Trust CEPEH Chatbots to Provide me with my Course Information



CEPEH Chatbot Personality was Realistic and Engaging



2. Results

There was mixed results for the chatbot used being realistic and engaging. This

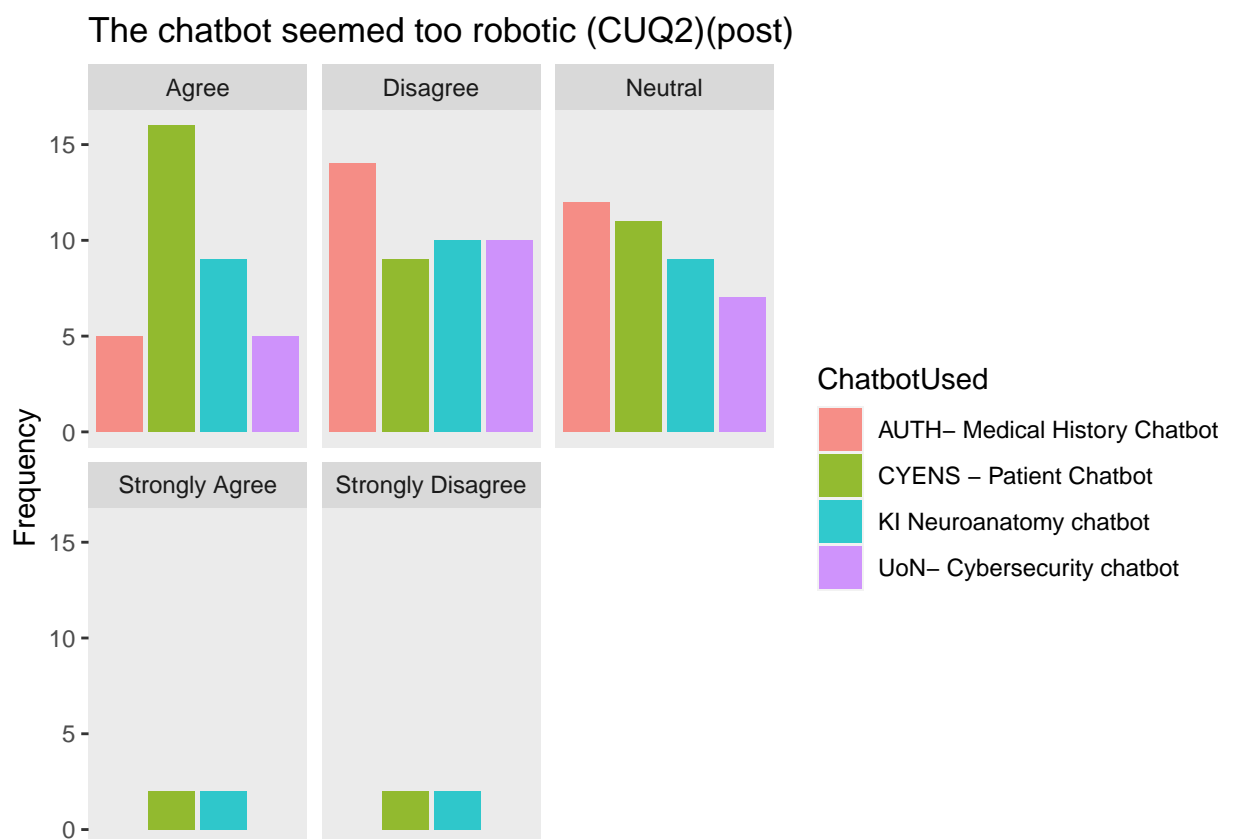
question has two descriptive terms however based on the other results we under-

stand that the chatbots' NLP logic, or ability to respond required improvement to

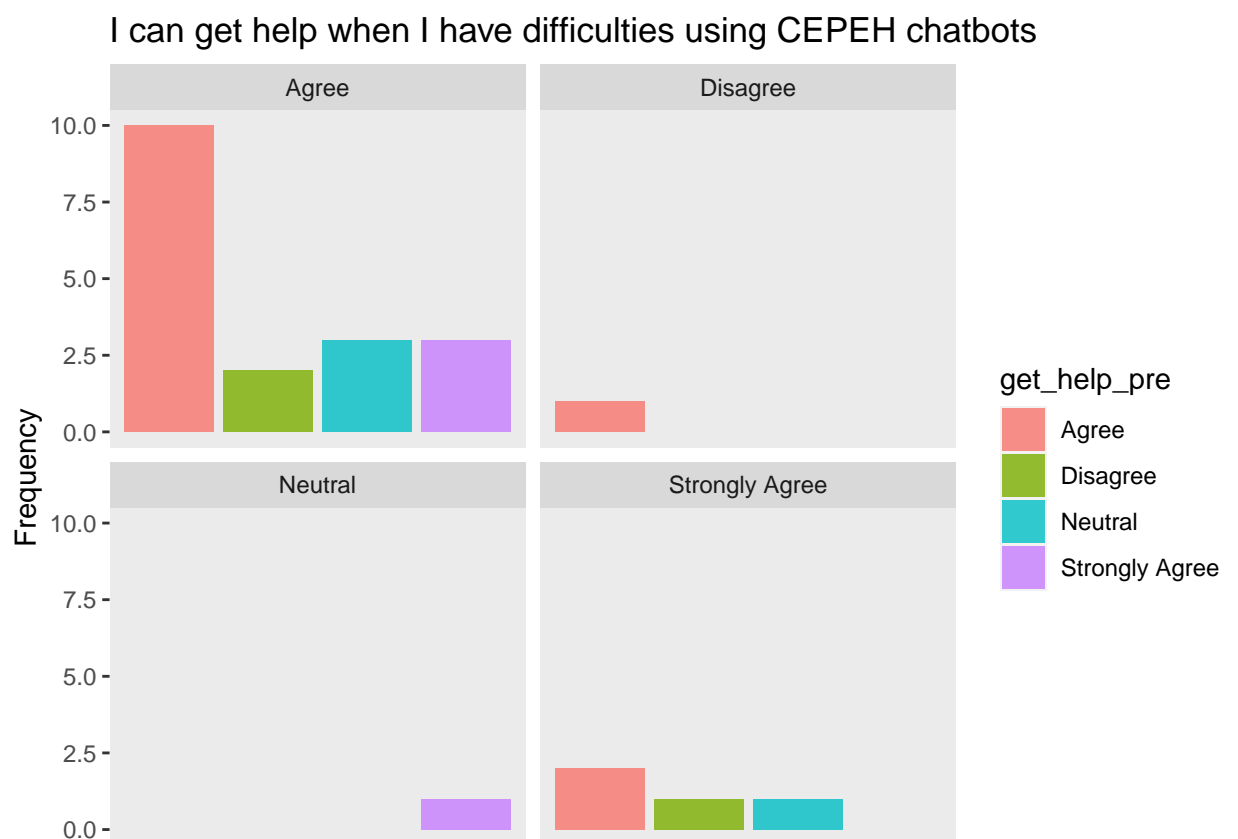
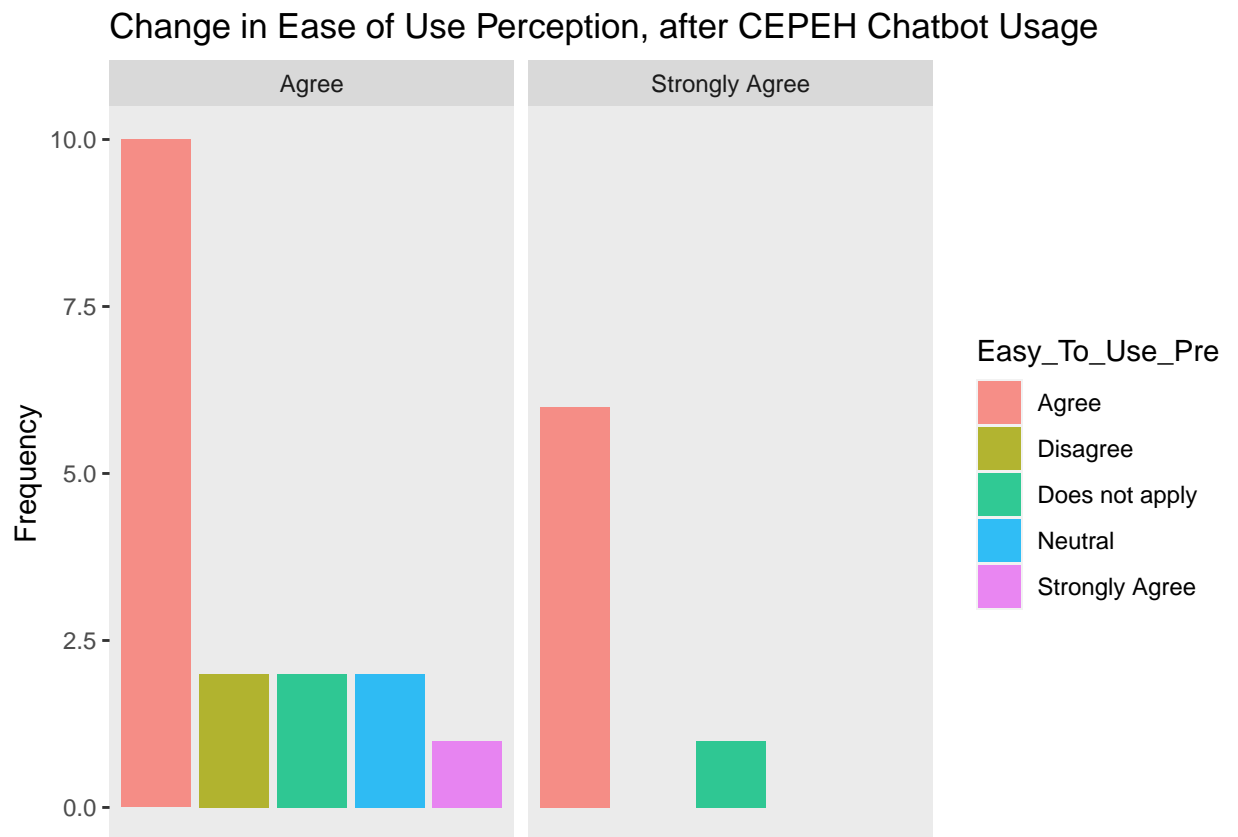
be more 'smooth' in replying. The primary limitation was found in the 'robotic'

interactions(See Figure 10). This was investigated further in the 'Text Mining'

and 'Sentiment Analysis' sections.



2. Results



2. Results

Those who disagreed or were neutral in the pre usage measure, improved their understanding that help was available with the CEPEH chatbots. After usage, 40 participants agreed they could get help if they had difficulty using the resources.

Previous_Chatbot_Usage	n
1-4 hours	15
10-19 hours	1
20+ hours	1
5-9 hours	2
Never	23

2.2 Chatbot Usability Questionnaire (CUQ)

2.2.1 CUQ Calculation tool

The CUQ was developed by researchers at Ulster University (see <https://www.ulster.ac.uk/research/topic/computer-science/artificial-intelligence/projects/cuq>) and as the calculation can be complex a dedicated calculation tool has been created.

Please download the CEPEH CUQ calculation tool which has all of the data entered, so you can see the CEPEH CUQ scoring.

[click here to download CUQ calc tool](#)

[click here to download CUQ score image](#)

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

The score for all 3 chatbots grouped was 65.2/100, This scoring system was designed to be comparable to SUS and may be freely used alongside it, or in combination with other usability metrics. There has been evidence of correlation of 76% between the CUQ and SUS therefore we expect the SUS scored to be between 48.75 and 81%. We believe the CUQ has more validity towards measuring the concepts of interest on this study.

[Read the CUQ development paper, see page 3 for correlation](#)

2. Results

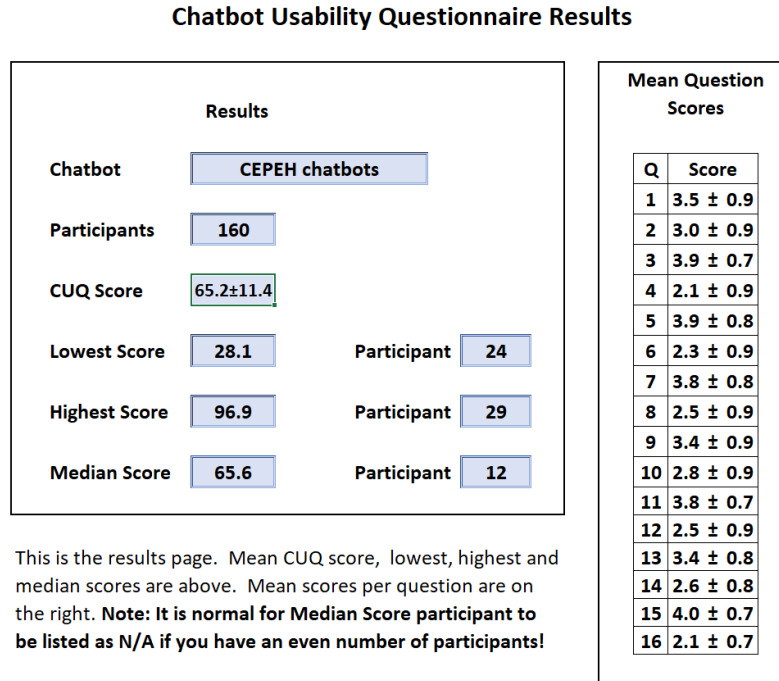


Figure 2.1: A marvel-lous meme
(#fig:cuq image)

Figure shows the CUQ scores as a box plot to highlight the range of Usability of the resources. Further exploration is required to understand which elements are causing this spread.

2.3 System Usability Scale (SUS) Scores

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

The SUS score should consist of 10 items. However, some SUS questions were improved upon by 1 or more CUQ questions, specifically to this Chatbot study. The SUS results would be overshadowed by the CUQ scores, expect 2 that did not have cross-over. The two questions were:

- I would like to use the CEPEH chatbot I tested, more frequently (SUS1)(post)
- I felt confident using the CEPEH chatbot (SUS2)(post)

2. Results

This meant the score of the SUS was not created, however the CUQ score better represented the Learners' perceptions of the CEPEH chatbot in terms of feasibility of reuse and acceptability in healthcare curricula.

Keep_Using_Chatbots	Confident	Count
Agree	Agree	44
Agree	Disagree	5
Agree	Neutral	11
Agree	Strongly Agree	6
Disagree	Agree	6
Disagree	Disagree	5
Disagree	Neutral	4
Neutral	Agree	10
Neutral	Disagree	1
Neutral	Neutral	6
Not Applicable	Not Applicable	3
Strongly Agree	Agree	10
Strongly Agree	Not Applicable	1
Strongly Agree	Strongly Agree	12
Strongly Disagree	Agree	1
Strongly Disagree	Strongly Agree	1

2.4 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

2.4.1 Other Findings

Other questions

I intend to continue using chatbots in the future (BI1)

The chatbot provided the information I needed with minimal commands

My knowledge of the topic improved after i had used the Chatbot

My confidence in understanding the topic improved after I had used the Chatbot

The chatbot provided me with the type of response i expected from asking
a tutor/lecturer

The information provided was reliable

The chatbot has a high level of trustworthiness

The duration of conversations to find my answer was too long

The videos/images provided were useful to my questions

The chatbot exceeded my expectation of how it could help me

The chatbot exceeded my expectation of how it could engage with me

I think this learning method could help me to acquire knowledge

I would use this tool again as it has some value to me

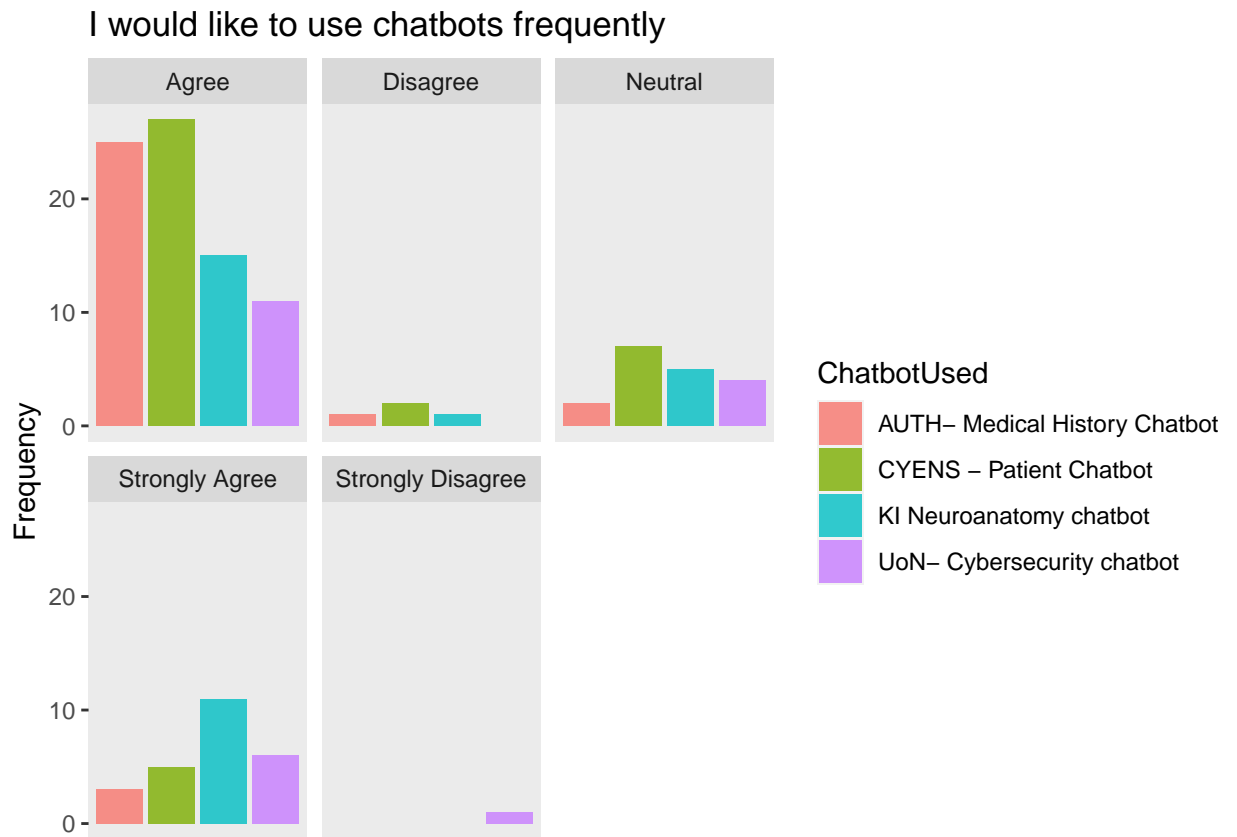
I think i will actively use this learning method

I believe I had some choice about learning during chatbot use

I would trust the chatbot to provide me with information for my course

One piece of knowledge i learned from the chatbot was..

2. Results



Repeated Measures t-test, aka paired t-test (before and after measurements)

```
## [1] "sex"
## [2] "before"
## [3] "after"
## [4] "Code"
## [5] "Location"
## [6] "Profession"
## [7] "CurrentMaterials"
## [8] "Previous_Chatbot_Usage"
## [9] "Use chatbots frequently (pre)"
## [10] "...10"
## [11] "I find chatbots can be useful in my daily life (PE1)(pre)"
## [12] "Using chatbots increases my chances of achieving things that are important"
## [13] "Using chatbots can help me accomplish things more quickly (PE3)(pre)"
## [14] "UseChatbotsFrequently"
```


2. Results

```
## [15] "Learning how to use chatbots is, or would be, for me (EE1)(pre)"
## [16] "My interactions with chatbots are clear and understandable (EE2)(pre)"
## [17] "Easy_To_Use_Pre"
## [18] "It is easy for me to become skilful at using chatbots (EE4)(pre)"
## [19] "People who are important to me think that I should use chatbots (SI1)(pre)"
## [20] "People who influence my behaviour think that I should use chatbots (SI2)(pre)"
## [21] "People whose opinions that I value prefer that I use chatbots (SI3)(pre)"
## [22] "I have the resources necessary to use chatbots (FC1)(pre)"
## [23] "knowledge_necessary_pre"
## [24] "Chatbots are compatible with other technologies I use (FC3)(pre)"
## [25] "get_help_pre"
## [26] "Using chatbots is enjoyable (HM2)(pre)"
## [27] "Using chatbots is very entertaining (HM3)(pre)"
## [28] "I intend to continue using chatbots in the future (BI1)(pre)"
## [29] "ChatbotUsed"
## [30] "Information_with_minimal_command_post"
## [31] "knowledge_improved_after_use"
## [32] "My confidence in understanding the topic improved after I had used the Chatbot"
## [33] "The chatbot provided me with the type of response i expected from asking a question"
## [34] "The information provided was reliable (post)"
## [35] "Personailty_real_engage"
## [36] "Robotic_CUQ2_post"
## [37] "The chatbot was welcoming during initial setup (CUQ3)(post)"
## [38] "The chatbot seemed very unfriendly (CUQ4)(post)"
## [39] "The chatbot explained its scope and purpose well (CUQ5)(post)"
## [40] "The chatbot gave no indication as to its purpose (CUQ6)(post)"
## [41] "The chatbot was easy to navigate (CUQ7)(post)"
## [42] "It would be easy to get confused when using the chatbot (CUQ8)(post)"
## [43] "The chatbot understood me well (CUQ9)(post)"
## [44] "The chatbot failed to recognise a lot of my inputs (CUQ10)(post)"
```

2. Results

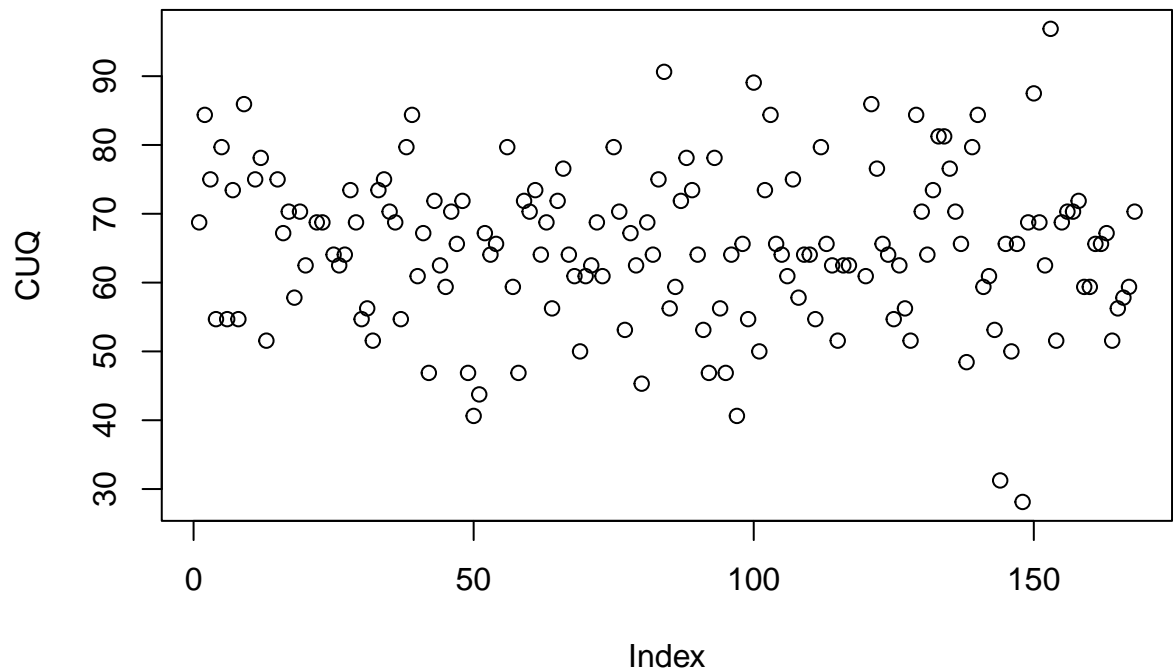
[45] "Chatbot responses were useful, appropriate and informative (CUQ11)(post)"
[46] "Chatbot responses were irrelevant (CUQ12)(post)"
[47] "The chatbot coped well with any errors or mistakes (CUQ13)(post)"
[48] "The chatbot seemed unable to handle any errors (CUQ14)(post)"
[49] "The chatbot was very easy to use (CUQ15)(post)"
[50] "The chatbot was very complex (CUQ16)(post)"
[51] "I would like to use the CEPEH chatbot i tested, more frequently (SUS1)(post)"
[52] "...52"
[53] "...53"
[54] "The chatbot has a high level of trustworthiness(post)"
[55] "The duration of conversations to find my answer was too long(post)"
[56] "I found the CEPEH chatbots useful in my daily life (PE1)(post)"
[57] "Using CEPEH chatbots increases my chances of achieving things important to me (PE2)(post)"
[58] "Using CEPEH chatbots can help me accomplish things more quickly (PE3)(post)"
[59] "Using CEPEH chatbots increases my productivity (PE4)(post)"
[60] "Learning how to use CEPEH chatbots is easy for me (EE1)(post)"
[61] "My interaction with CEPEH chatbots is clear and understandable (EE2)(post)"
[62] "CEPEH_Post_Easy_To_Use"
[63] "It is easy for me to become skilful at using CEPEH chatbots (EE4)(post)"
[64] "People who are important to me think that I should use CEPEH chatbots (SI1)(post)"
[65] "People who influence my behaviour think that I should use CEPEH chatbots (SI2)(post)"
[66] "People whose opinions that I value prefer that I use CEPEH chatbots (SI3)(post)"
[67] "I have the resources necessary to use CEPEH chatbots (FC1)(post)"
[68] "I have the knowledge necessary to use CEPEH chatbots (FC2)(post)"
[69] "CEPEH Chatbots are compatible with other technologies I use (FC3)(post)"
[70] "get_help"
[71] "Using CEPEH chatbots is enjoyable (HM2)(post)"
[72] "Intend"
[73] "The videos/images provided were useful to my questions(post)"
[74] "The chatbot exceeded my expectation of how it could help me(post)"

2. Results

```
## [75] "The chatbot exceeded my expectation of how it could engage with me(post)"
## [76] "The chatbot exceeded my expectation of how entertaining it was to use(post)"
## [77] "I think this learning method could help me to acquire knowledge(post)"
## [78] "ReuseChatbot"
## [79] "I think i will actively use this learning method(post)"
## [80] "I believe i had some choice about learning during chatbot use(post)"
## [81] "POST_Trust"
## [82] "Informed of results(post)"
## [83] "CUQ"

## # A tibble: 168 x 3
##   sex      before after
##   <chr>    <dbl> <dbl>
## 1 Male          2      2
## 2 Female        3      2
## 3 Female        3      2
## 4 Male          1      2
## 5 Male          1      2
## 6 Female        2      2
## 7 Male          1      2
## 8 Male          2      2
## 9 Female        1      2
## 10 Male         1      2
## # ... with 158 more rows
```

2. Results



```
##  
## Paired t-test  
##  
## data: CUQ and after  
## t = 64.48, df = 117, p-value < 2.2e-16  
## alternative hypothesis: true mean difference is not equal to 0  
## 9.5 percent confidence interval:  
## 62.88102 63.11474  
## sample estimates:  
## mean difference  
## 62.99788
```

This t-test compares confident using mobile chatbots before and after CEPEH chatbot usage.

output: #bookdown::html_document2: default #bookdown::word_document2:

2. Results

default bookdown::pdf_document2: template: templates/template.tex document-
class: book

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

3

Text Mining, Natural Language Processing, and Sentiment Analysis

3.1 CEPEH Qualatative Feedback

The focus group discussions provided a lot of feedback for how the participants experienced their interactions with the chatbots, and how the CEPEH team can improve them, improve the design and development processes, and improve uptake and sharing.

One method of analysing this data is with use of text mining and data manipulation, creating word clouds, sentiment analysis, and using a model which can distinguish the unique themes in text, and highlights for us what text is used to create these themes.

Therefore, we have created a model to allow efficient and intelligent analysis of this open/free focus group data.

3.2 Tokenising

Firstly, we tokenised the words from the FGDs. A Token is “a meaningful unit of text, most often a word, that we are interested in using for further analysis”. For each word we give it a property that we can call upon later.

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

The data manipulation for this included removing punctuation, converting to lower-case, and setting word type to word (and not such types as “characters”, “ngrams”, “sentences”, “lines” etc)

```
## readtext object consisting of 6 documents and 0 docvars.
## # Description: df [6 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        "\"\"..."
```

3.2.1 Stop words

The model then removed words with meaningless function. These are called stop words. Words like “the”, “of” and “to” are the most frequent words found, technically, but are of little interest to us.

```
## # 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
```

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

```
## 10 after      SMART
## # ... with 1,139 more rows
## readtext object consisting of 806 documents and 0 docvars.
## # Description: df [806 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 800 more rows
```

We also created a custom list of stop words for CEPEH. We know participants may mention other objects, and the list was as followed: found; chatbot; chatbots; presentation.

The data was ready for analysis by the model. We ordered it to find the most frequent words.

```
## # A tibble: 385 x 3
## # Groups:   doc_id [1]
##   doc_id word      n
##   <fct> <chr>    <int>
## 1 1      information 11
## 2 1      helpful    8
## 3 1      understand  8
## 4 1      idea        7
## 5 1      ideas        7
## 6 1      lot          7
## 7 1      workshop    7
## 8 1      beginning    6
```


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

```
## 9 1      dont      6

## 10 1     medical    6

## # ... with 375 more rows

## readtext object consisting of 669 documents and 0 docvars.

## # Description: df [669 x 3]

##   doc_id word      text

##   <fct> <chr>     <chr>

## 1 1      personally "\"\"..."

## 2 1      nice      "\"\"..."

## 3 1      week      "\"\"..."

## 4 1      ive       "\"\"..."

## 5 1      feeling   "\"\"..."

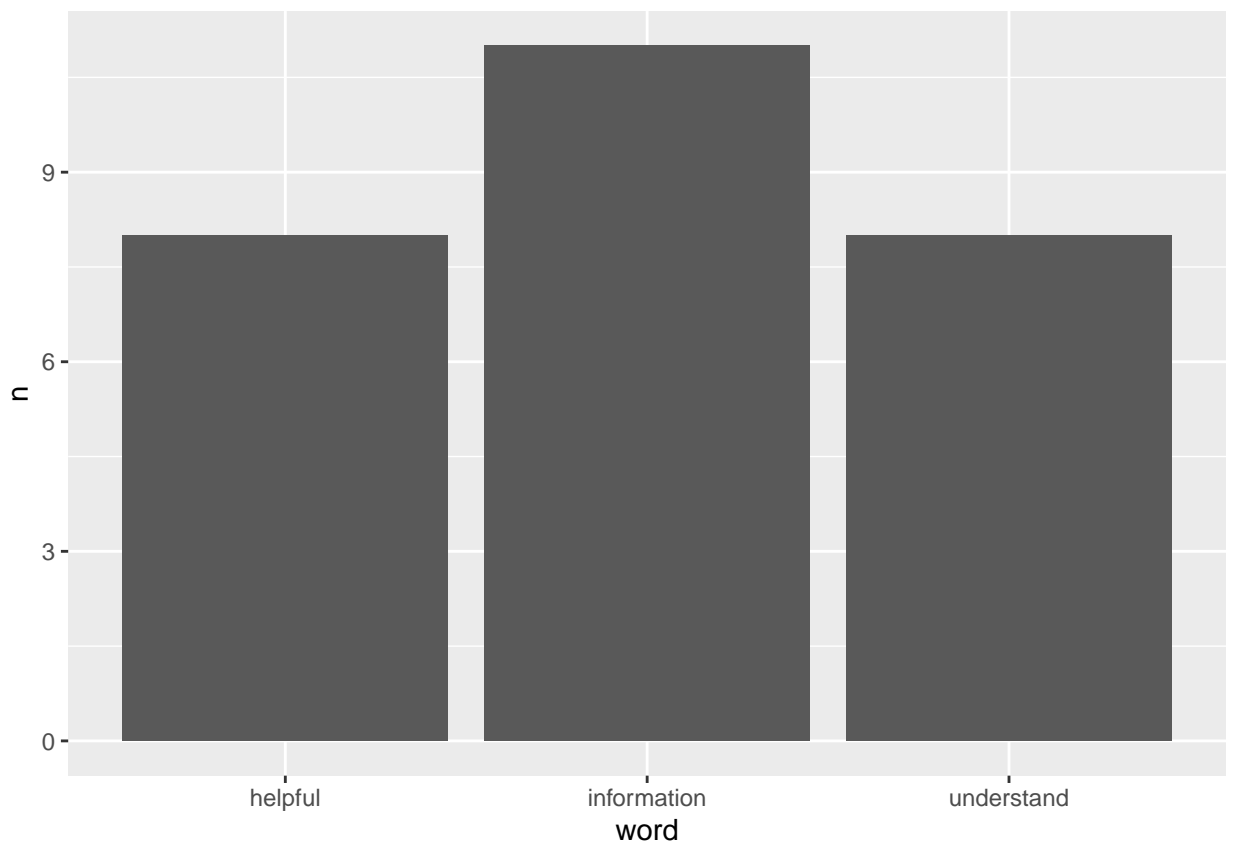
## 6 1      technical "\"\"..."

## # ... with 663 more rows
```

Plotting word frequencies - bar graphs

With this information a Bar graph of top words from the participants in the FGD can be rendered.

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

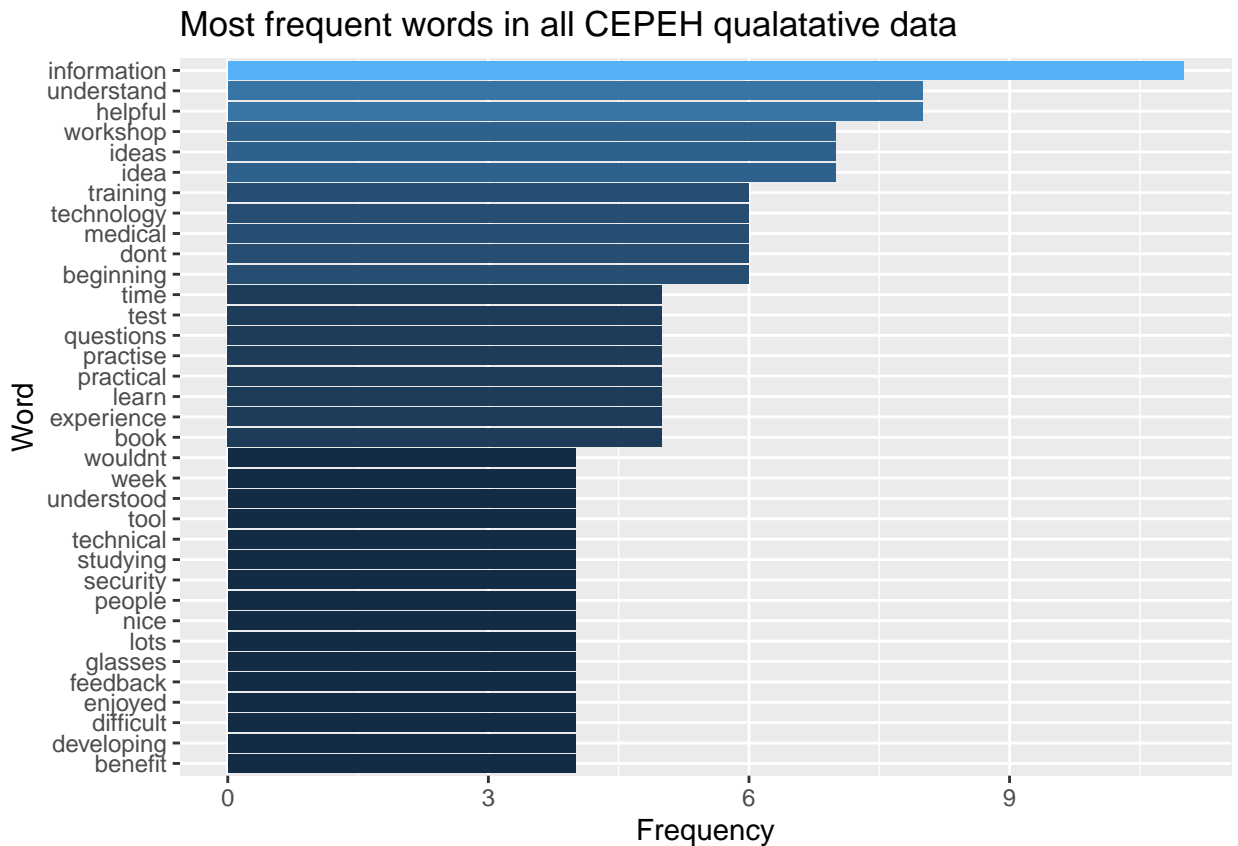


and after some modifications, a graph of the top 35 words is produced, with

better aesthetics. The most frequent words present in focus group discussions after

using the 4 chatbots, are in the Figure below.

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



Although the frequency is not high for each word, we are able to get a general picture of the sentiments, intensities, and concerns which would be immediately occurring when plotted.

3.2.2 3.2 Normalised frequency

A better way to understand this data is to normalise the frequency of occurrences in accordance with the source text. The raw text had 2827 words in total. Therefore we can mutate the ratios to suit.

```
## # A tibble: 1 x 2
## # Groups:   doc_id [1]
##   doc_id `sum(n)`
##   <fct>   <int>
## 1 1      656
## # A tibble: 383 x 2
```

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

```
##      word      pmw

##      <chr>      <dbl>

##  1 information  47.4

##  2 helpful     34.5

##  3 understand  34.5

##  4 idea        30.2

##  5 ideas       30.2

##  6 workshop    30.2

##  7 beginning   25.9

##  8 dont        25.9

##  9 medical     25.9

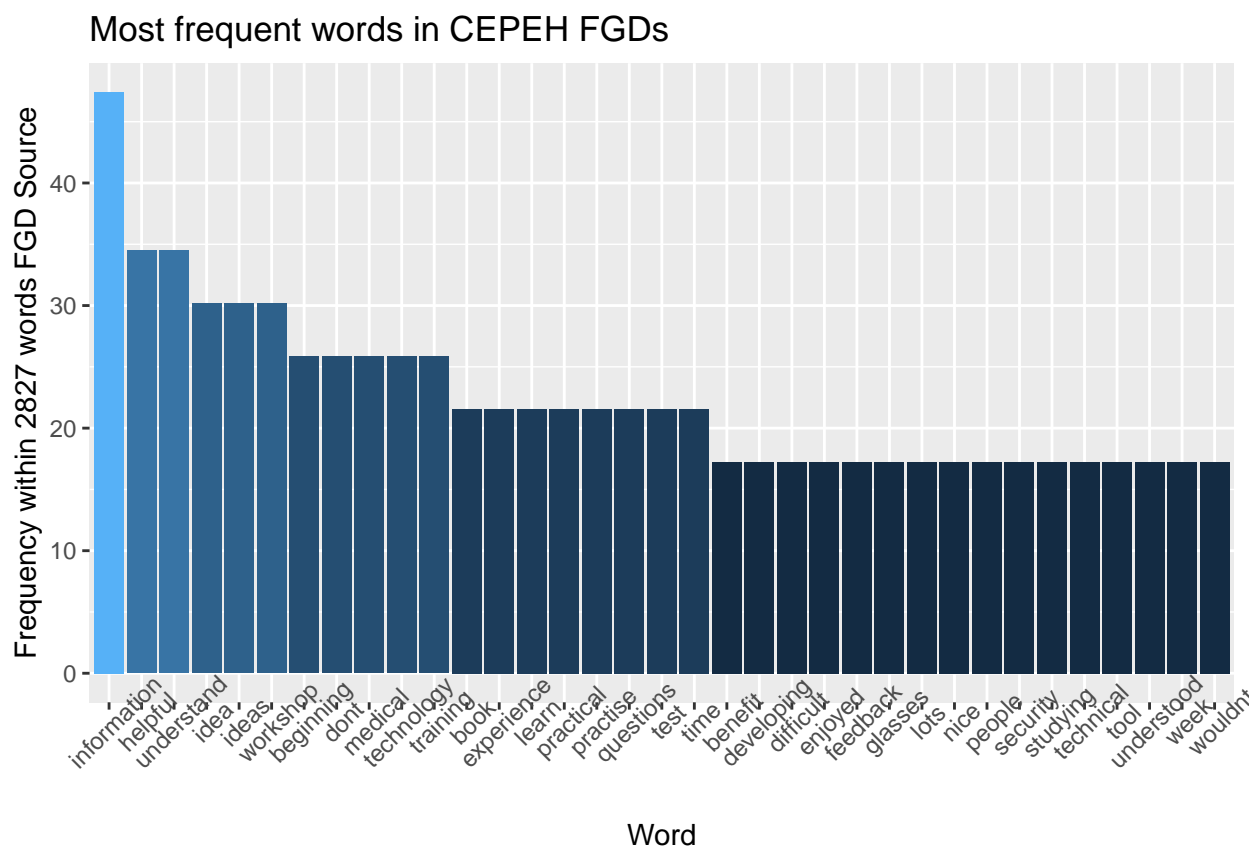
## 10 technology  25.9

## # ... with 373 more rows
```

Plotting normalised frequency

Now we can plot, for example, the 20 most frequent words when normalised by the source text.

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



In summary, this understanding of frequent words can help to understand common concurrences and extrapolate to a larger audience. If scope and impact of CEPEH chatbots increased we can understand the type of themes and trends may occur, based on such FGD analysis.

3.2.3 3.3 Word clouds

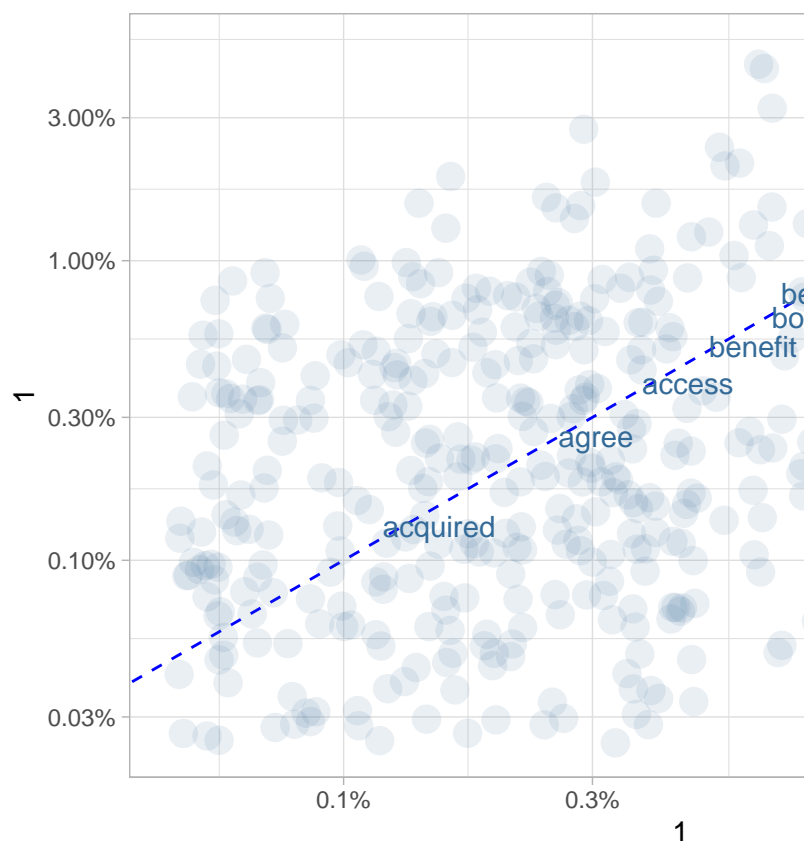
To visualise the most frequent words in another format, below is a word cloud which presents the word size to indicate the frequency- words that occur more often being displayed in a larger font size. This has a normalised data frequency in accordance to the FGD source document analysed.



We understand the context has been reduced for each word. However, in general there can be categorised positive/negative words from the word cloud: Positive words are- benefit, practical, nice, helpful, learn, ideas, and enjoyed Negative words are- difficult, test (who likes a test?), don't, and 'lot' may be negative if there is a 'lot' of information.

3.3 The vocabulary of Texts

Here is a graph that has plotted the words in places depending 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 fre-



quent in one text compared to the other.

3.4 Sentiment analysis

What is the sentiment of all participants? What is types of emotional words are being used? The preparation of these words has some use in understanding the frequencies, but their emotional valence are not compared. The table above has the word *'helpful'* which has a positive connotation, however there are 386 words, with many having several occurrences.

```
## # A tibble: 6 x 2
##   word      sentiment
##   <chr>    <chr>
## 1 2-faces  negative
## 2 abnormal negative
## 3 abolish negative
## 4 abominable negative
```

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

```
## 5 abominably negative
## 6 abominate negative

## readtext object consisting of 6 documents and 1 docvar.
## # Description: df [6 x 4]
##   doc_id word      sentiment text
##   <fct> <chr>      <chr>    <chr>
## 1 1      nice      positive  "\"\"\"..."
## 2 1      invaluable positive  "\"\"\"..."
## 3 1      facilitate positive  "\"\"\"..."
## 4 1      difficult negative  "\"\"\"..."
## 5 1      benefits  positive  "\"\"\"..."
## 6 1      easier    positive  "\"\"\"..."

## readtext object consisting of 2 documents and 0 docvars.
## # Description: df [2 x 3]
##   sentiment      n text
##   <chr>      <int> <chr>
## 1 negative      24 "\"\"\"..."
## 2 positive      62 "\"\"\"..."

## readtext object consisting of 6 documents and 1 docvar.
## # Description: df [6 x 4]
##   word      sentiment      n text
##   <chr>      <chr>      <int> <chr>
## 1 helpful    positive      8 "\"\"\"..."
## 2 benefit    positive      4 "\"\"\"..."
## 3 difficult  negative      4 "\"\"\"..."
## 4 enjoyed    positive      4 "\"\"\"..."
## 5 nice       positive      4 "\"\"\"..."
## 6 easy       positive      3 "\"\"\"..."

## readtext object consisting of 2 documents and 1 docvar.
## # Description: df [2 x 4]
```


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

```
##   doc_id sentiment      n text
## * <fct>  <chr>      <int> <chr>
## 1 1      negative     24 "\"\"\"..."
## 2 1      positive     62 "\"\"\"..."

## readtext object consisting of 2 documents and 0 docvars.
## # Description: df [2 x 3]
##   sentiment      n text
##   <chr>      <int> <chr>
## 1 negative     24 "\"\"\"..."
## 2 positive     62 "\"\"\"..."

## readtext object consisting of 6 documents and 1 docvar.
## # Description: df [6 x 4]
##   word      sentiment      n text
##   <chr>      <chr>      <int> <chr>
## 1 helpful    positive      8 "\"\"\"..."
## 2 benefit    positive      4 "\"\"\"..."
## 3 difficult  negative      4 "\"\"\"..."
## 4 enjoyed    positive      4 "\"\"\"..."
## 5 nice       positive      4 "\"\"\"..."
## 6 easy       positive      3 "\"\"\"..."
```

```
CEPEHQ_tidy_bing %>%
  summarize(max(total_score), min(total_score))
```

```
##   max(total_score) min(total_score)
## 1                38                38
```

```
knitr::kable(CEPEHQ_tidy_bing)
```

negative	positive	total_score
24	62	38

```
<!--chapter:end:03-TM.Rmd-->
```

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

— — —

output:

```
#bookdown::html_document2: default
```

```
#bookdown::word_document2: default
```

bookdown::pdf_document2:

```
template: templates/template.tex
```

```
keep_tex: true
```

```
documentclass: book
```

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

— — —

Discussion {#Discussion }

```
\minitoc <!-- this will include a mini table of contents-->
```

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

```
\begin{tabular}{l}
```

\toprule

sex & time & Code & Location & Profession & CurrentMaterials & Previous_Chatbot_U

\midrule

Female & Pre & 666 & Cyprus & Student on a Healthcare course & Books & Never & Agree

[illegible]

Male & Pre & A.S.1507 & Greece & Student on a Healthcare course & Online journals\

NA & Post & A.N. 31082003 & NA & NA & NA & NA & NA & NA & NA & NA & NA & NA &

[illegible]

\addlinespace

Female & Pre & AG 26062002 & Greece & Student on a Healthcare course & Books & Neve

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

```
\bottomrule  
\end{tabular}
```

Those results can be interpreted that the learning objectives of the training event

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

Making your table pretty

When you use ``kable`` to create tables, you will almost certainly want to set the op

This makes your table look a million times better:

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

Compare this to the default style, which looks terrible:

```
\begin{tabular}{l|r|r|r|r|r|r|r|r|r|r|r|r}
```

	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

```
\end{tabular}
```

If your table is too wide

You might find that your table expands into the margins of the page, like the table above. Fix this with the `kable_styling` function from the `[kableExtra]` (<https://haozhu23.com/kableExtra/>)

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

```
\begin{table}
\centering
\resizebox{\linewidth}{!}{
\begin{tabular}{lrrrrrrrrrrrr}
\toprule
& mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
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\\
\addlinespace
Valiant & 18.1 & 6 & 225 & 105 & 2.76 & 3.460 & 20.22 & 1 & 0 & 3 & 1\\
\bottomrule
\end{tabular}}
\end{table}
```

This scales down the table to fit the page width.

If your table is too long

If your table is too long to fit on a single page, set ``longtable = TRUE`` in the ``k``

```
\begin{longtable}{lrrrrrrrr}
\toprule
& mpg & cyl & disp & hp & drat & wt & qsec & vs\\
\midrule
```

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

```
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
```

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

```
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\addlinespace
```

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

```

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\\
\\addlinespace
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\\
\\addlinespace
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\\
\\bottomrule
\\end{longtable}

```

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``:

```

\\begin{longtable}{lrrrrrrrrrrrr}
\\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\\midrule
\\endfirsthead
\\multicolumn{12}{@{}l}{\\textit{(continued)}}\\
\\toprule

```


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

& mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\

\midrule

\endhead

\endfoot

\bottomrule

\endlastfoot

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\\

\addlinespace

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\\

\addlinespace

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\\

\addlinespace

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\\

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

\addlinespace

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\\

\addlinespace

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\\

\addlinespace

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\\

\addlinespace

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\\

\addlinespace

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\\

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

\addlinespace

```
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\\
```

\addlinespace

```
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\\
```

\addlinespace

```
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\\
```

\addlinespace

```
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\\*
```

\end{longtable}

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 s

To adjust the font size, use kableExtra's ``font_size`` option:

\begingroup\fontsize{9}{11}\selectfont

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

```

\begin{longtable}{lrrrrrrrrrrrr}
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endfirsthead
\multicolumn{12}{@{}l}{\textit{(continued)}}\\
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endhead

\endfoot
\bottomrule
\endlastfoot
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\\
\addlinespace
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\\
\addlinespace
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\\

```

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

```
Volvo 142E1 & 21.4 & 4 & 121.0 & 109 & 4.11 & 2.780 & 18.60 & 1 & 1 & 4 & 2\\*
\end{longtable}
\endgroup{}
```

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

```
\begin{landscape}
\begin{longtable}{lrrrrrrrrrrrr}
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endfirsthead
\multicolumn{12}{@{}l}{\textit{(continued)}}\\
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endhead

\endfoot
\bottomrule
\endlastfoot

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\\
\addlinespace
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\\
```

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

```
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\\
\addlinespace
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\\*
\end{longtable}
\end{landscape}
```

Max power: manually adjust the raw LaTeX output `{#max-power}`

For total flexibility, you can adjust the raw LaTeX output from ``kable`/`kableExtra``

Let us consider how we would do this for the example of adjusting the font size if

Latex has a bunch of standard commands that set an approximate font size, as shown

```
\begin{figure}[H]
```

```
{\centering \includegraphics[width=0.5\linewidth]{figures/sample-content/latex_font
```

```
}
```

```
\caption{Font sizes in LaTeX}\label{fig:latex-font-sizing}
```

```
\end{figure}
```

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
2. Add the attributes that make R Markdown understand that the table is a table (it

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

```
\begin{scriptsize}
\begin{longtable}{lrrrrrrrrrrrr}
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endfirsthead
\multicolumn{12}{@{}l}{\textit{(continued)}}\\
\toprule
    & mpg & cyl & disp & hp & drat & wt & qsec & vs & am & gear & carb\\
\midrule
\endhead

\endfoot
\bottomrule
\endlastfoot
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\\
\addlinespace
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\\
\addlinespace
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\\
```

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\
\addlinespace
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\\

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

```
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\\*
\end{longtable}
\end{scriptsize}
```

```
<!--chapter:end:04-Discussion.Rmd-->
```

output:

```
#bookdown::html_document2: default
#bookdown::word_document2: default
bookdown::pdf_document2:
  template: templates/template.tex
documentclass: book
#bibliography: [bibliography/references.bib, bibliography/additional-references.bib]
---
```

```
# Training Event Results {#cites-and-refs}
```

```
\chaptermark{Citations and cross-refs}
```

```
\minitoc <!-- this will include a mini table of contents-->
```

```
## CEPEH Training Event C1
```

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

The CEPEH training event C1 held at the premises of University of Nottingham aiming 15 participants were from RISE, AUTH, UoN.

Project managers of partners signposted the person involved, and relevant announcements

Overall Training Events Evaluation

Participants were asked to highlight what they liked for each day and how each day

Day 1

The participants comment that they liked the design method for educational resources

Day 2

Participants enjoyed the presentation from the invited speaker from another faculty

Day3

Participants liked the hands-on activities of the day also enjoyed the creativity of

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

CEPEH Training Event 2

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

Twenty-six participants attended the Training Event, along with approximately 10 students.

Table 1 suggested attendees had minimal intention to share their own ideas due to limited time.

There were 77% of participants stated they were novices in experience with chatbots.

One day had several events regarding cybersecurity in healthcare. When asked before the event, participants were asked if they had any experience with cybersecurity in healthcare.

There were 90% (23) of students who heard about the event through a lecturer or a peer.

There were six individuals who stated neutral or disagree when asked if having issues with chatbots.

As this was face-to-face, participants were asked about sufficient Covid-19 precautions. In summary, most participants were undergraduate students with novice experience, had limited experience with chatbots, and had minimal intention to share their own ideas.

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

The usual way to include citations in an **R Markdown** document is to put references in a separate file. Then reference the path to this file in ***index.Rmd***'s YAML header with ``bibliography``.

[`~bib-formats`]: The bibliography can be in other formats as well, including EndNote.

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].

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

```
```\bibtex
@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 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).

### 3.4.1 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).

If you want a numerical citation style instead, try `cs1: 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](#).

### 3.4.2 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!

## 3.5 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)`

### 3.5.1 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. Text Mining, Natural Language Processing, and Sentiment Analysis

**Table 3.1:** Stopping cars

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

#### 3.5.2 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 ?? is [this image](#).

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

#### 3.5.3 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 3.1 is [this table](#).

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

#### 3.5.4 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:

### 3. Text Mining, Natural Language Processing, and Sentiment Analysis

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

#### 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 ?? 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 ?? on page ??

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

## 3.6 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](#),

### *3. Text Mining, Natural Language Processing, and Sentiment Analysis*

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.

## **3.7 Additional resources**

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

# Appendix

# 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`**

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

B

The Second Appendix, for Fun

## References

- 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/10.1177/10](https://doi.org/10.1177/1071181312561289)  
[71181312561289](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.