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

1 Text Mining, Natural Language Processing, and Sentiment Analysis

```
library(tidyverse) #for various data manipulation tasks
library(tidytext) #for text mining specifically, main package in
↪  book
library(stringr) #for various text operations
library(gutenbergr) #to access full-text books that are in the
↪  public domain
library(scales) # for visualising percentages
library(readtext) # for reading in txt files
library(wordcloud) # for creating wordclouds
```

1.1 1 Reading in texts

1.1.1 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):

```
CEPEHQ_raw <- readtext("P1.txt")

CEPEHQ_raw $doc_id <- sub(".txt", "", CEPEHQ_raw $doc_id) # this
↪  gets rid of .txt in the play titles
```

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:

1.1.2 1.3 Preparing data

- convert name to ID numbers with more descriptive labels

```
CEPEHQ_raw$doc_id <- as.factor(CEPEHQ_raw$doc_id)
CEPEHQ_raw$doc_id <- plyr::revalue(CEPEHQ_raw$doc_id,
                                                c("P1" = "1"))
```

1.2 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"

1.2.1 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 lowercase (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)

```
CEPEHQ_tidy <- CEPEHQ_raw%>%
  unnest_tokens(word, text)


CEPEHQ_tidy
```

```
## readtext object consisting of 2888 documents and 0 docvars.
## # Description: df [2,888 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 2,882 more rows
```

1.2.2 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)

```
CEPEHQ_tidy <- CEPEHQ_tidy %>%
  mutate(word = str_extract(word, "[a-z']+"))
```

1.2.3 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
- remove these from your dataset with anti_join

We can take a look:

```
stop_words
```

```
## # 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
```

```
CEPEHQ_tidy <- CEPEHQ_tidy %>%
  anti_join(stop_words)
```

```
## Joining, by = "word"
```

```
CEPEHQ_tidy
```

```
## readtext object consisting of 816 documents and 0 docvars.
## # Description: df [816 x 3]
##   doc_id word       text
##   <fct>  <chr>      <chr>
## 1 1      personally "\"\"..."
## 2 1      nice       "\"\"..."
## 3 1      week       "\"\"..."
## 4 1      ve         "\"\"..."
## 5 1      feeling    "\"\"..."
## 6 1      chatbots   "\"\"..."
## # ... with 810 more rows
```

Define other stop words:

```
meaningless_words <- tibble(word = c("found", "chatbot", "chatbots",
↪  "presentations", "NA", "video", "videos", "didnt", "na", "nil",
↪  ""))

CEPEHQ_tidy <- CEPEHQ_tidy %>%
  anti_join(meaningless_words)
```

```
## Joining, by = "word"
```

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

1.3 3 Analysing frequencies

1.3.1 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)

```
CEPEHQ_freq <- CEPEHQ_tidy %>%
  group_by(doc_id) %>% #including this ensures that the counts are
  ↪  by book and the id column is retained
  count(word, sort=TRUE)
CEPEHQ_freq <- CEPEHQ_freq[-6,]

CEPEHQ_freq
```

```
## # A tibble: 385 x 3
## # Groups:   doc_id [1]
##    doc_id word             n
##    <fct>  <chr>        <int>
##  1 1      vr              15
##  2 1      information     11
##  3 1      presentation     9
```

```
##  4 1      security         9
##  5 1      cyber            8
##  6 1      understand       8
##  7 1      idea             7
##  8 1      ideas            7
##  9 1      lot              7
## 10 1      workshop         7
## # ... with 375 more rows
```

```
CEPEHQ_tidy %>%
  group_by(doc_id) %>%
  count(word, sort=TRUE) %>%
  filter(doc_id == "1")
```

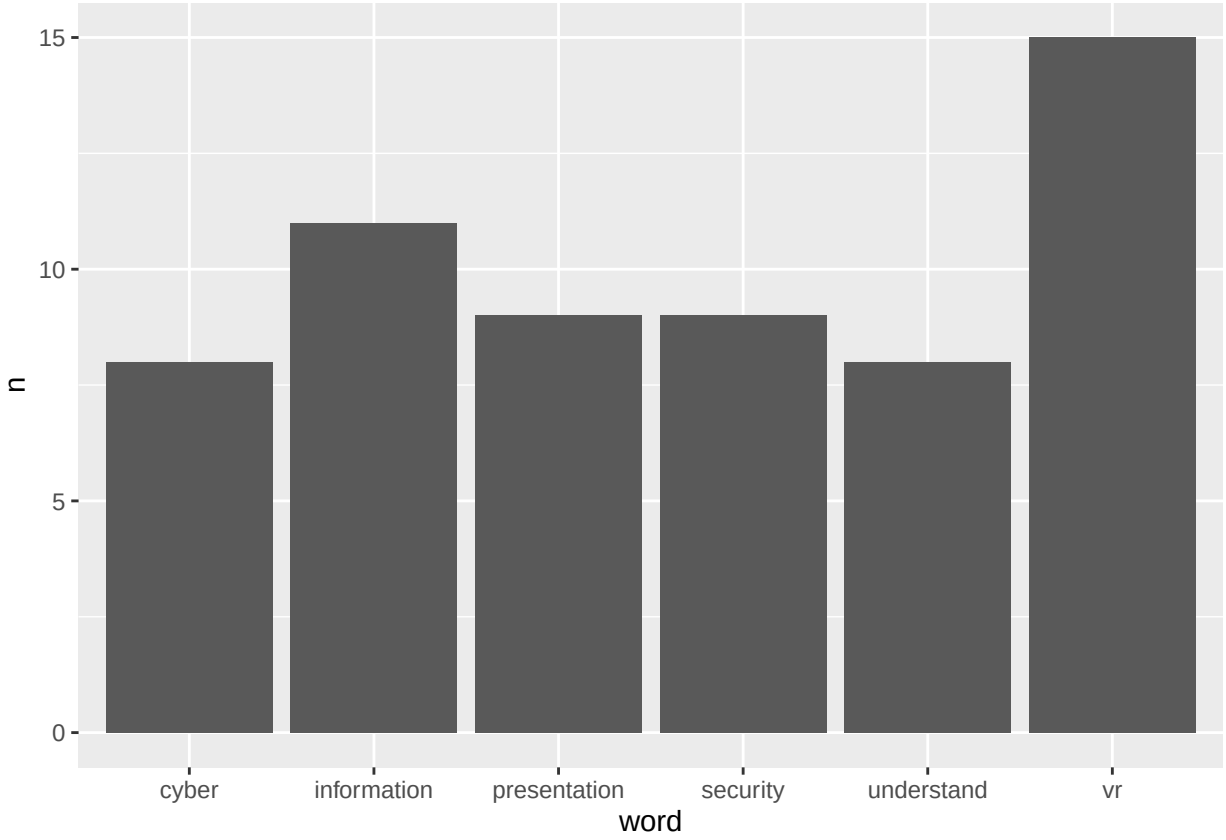
```
## # A tibble: 386 x 3
## # Groups:   doc_id [1]
##    doc_id word             n
##    <fct>  <chr>        <int>
##  1 1      vr              15
##  2 1      information     11
##  3 1      presentation     9
##  4 1      security         9
##  5 1      cyber            8
##  6 1      helpful          8
##  7 1      understand       8
##  8 1      idea             7
##  9 1      ideas            7
## 10 1      lot              7
## # ... with 376 more rows
```

Plotting word frequencies - bar graphs

Bar graph of top words in CEPEHQ.

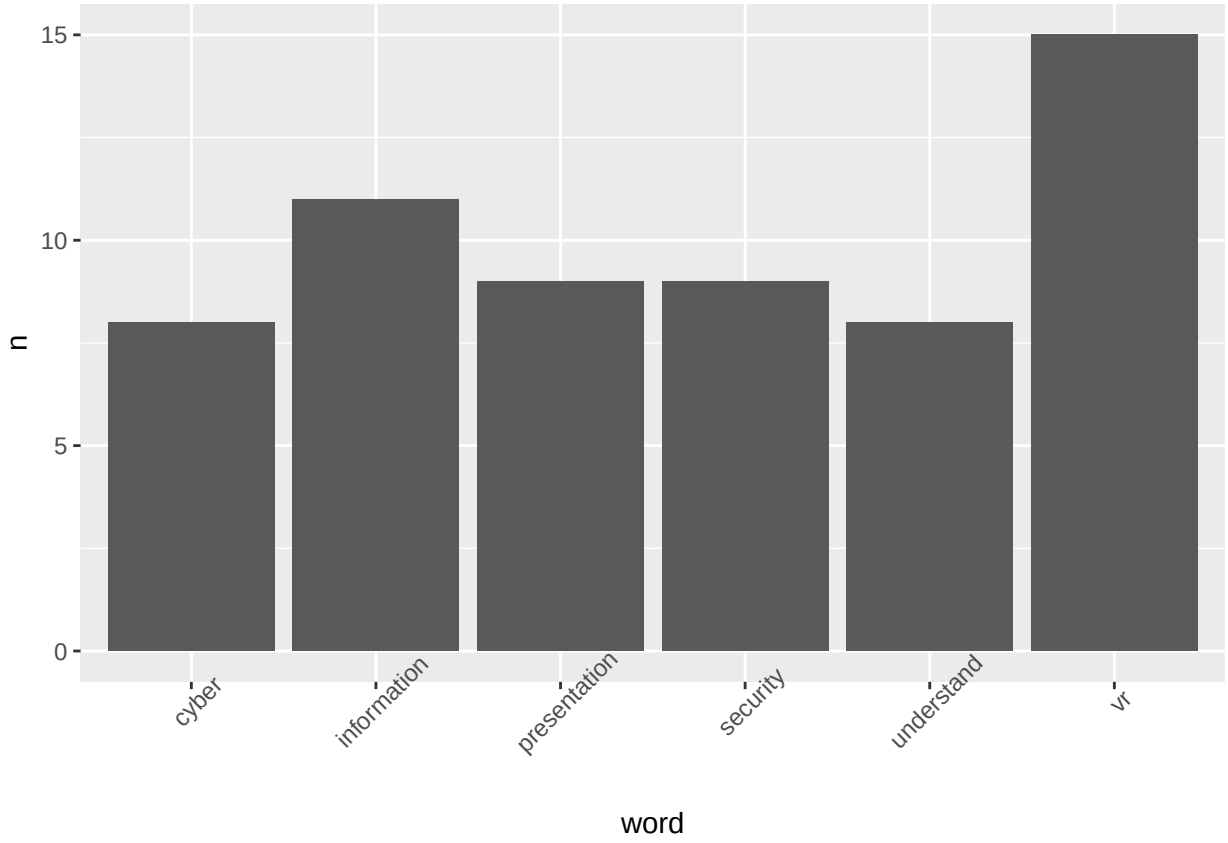
Basic graph:

```
CEPEHQ_freq %>%
  filter(n>7& doc_id == "1") %>%
  ggplot(aes(x=word, y=n)) +
  geom_col()
```



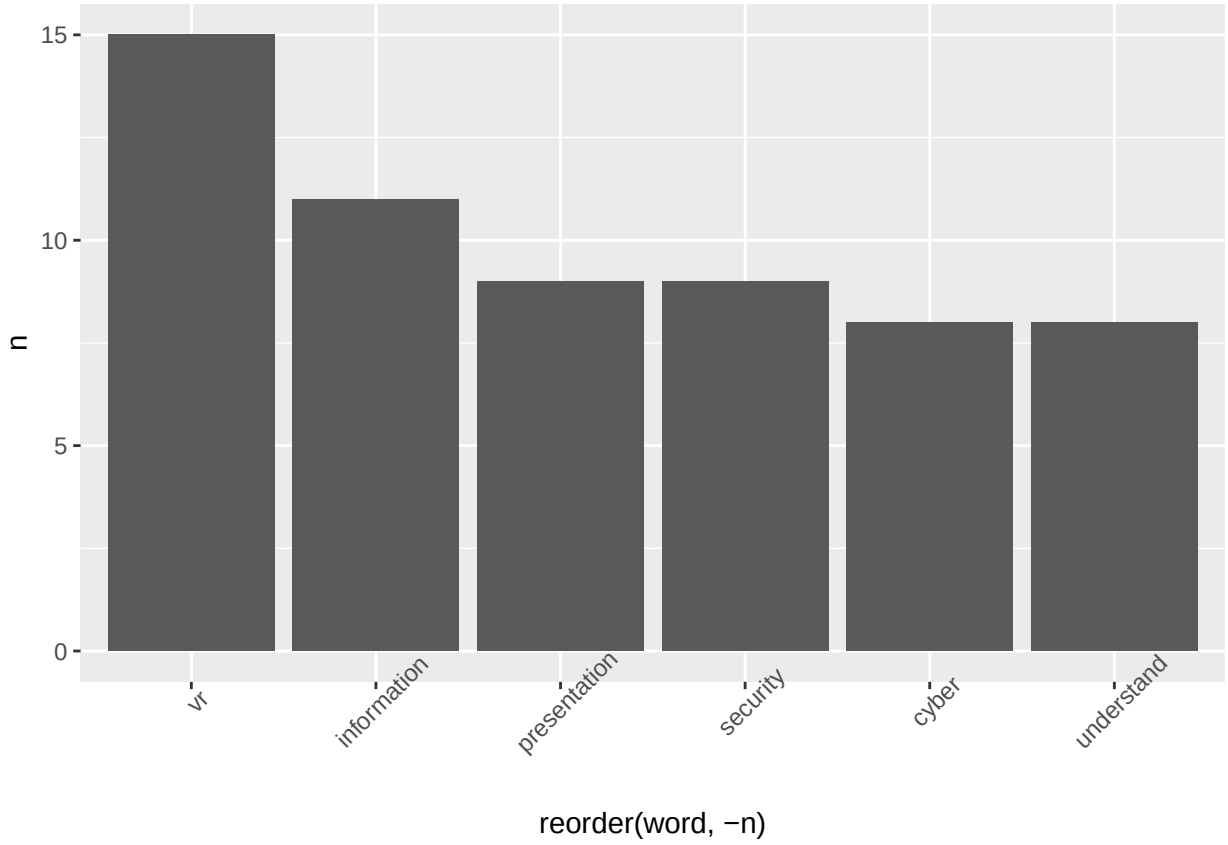
Readable labels:

```
CEPEHQ_freq %>%
  filter(n>7 & doc_id == "1") %>%
  ggplot(aes(x=word, y=n)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45))
```



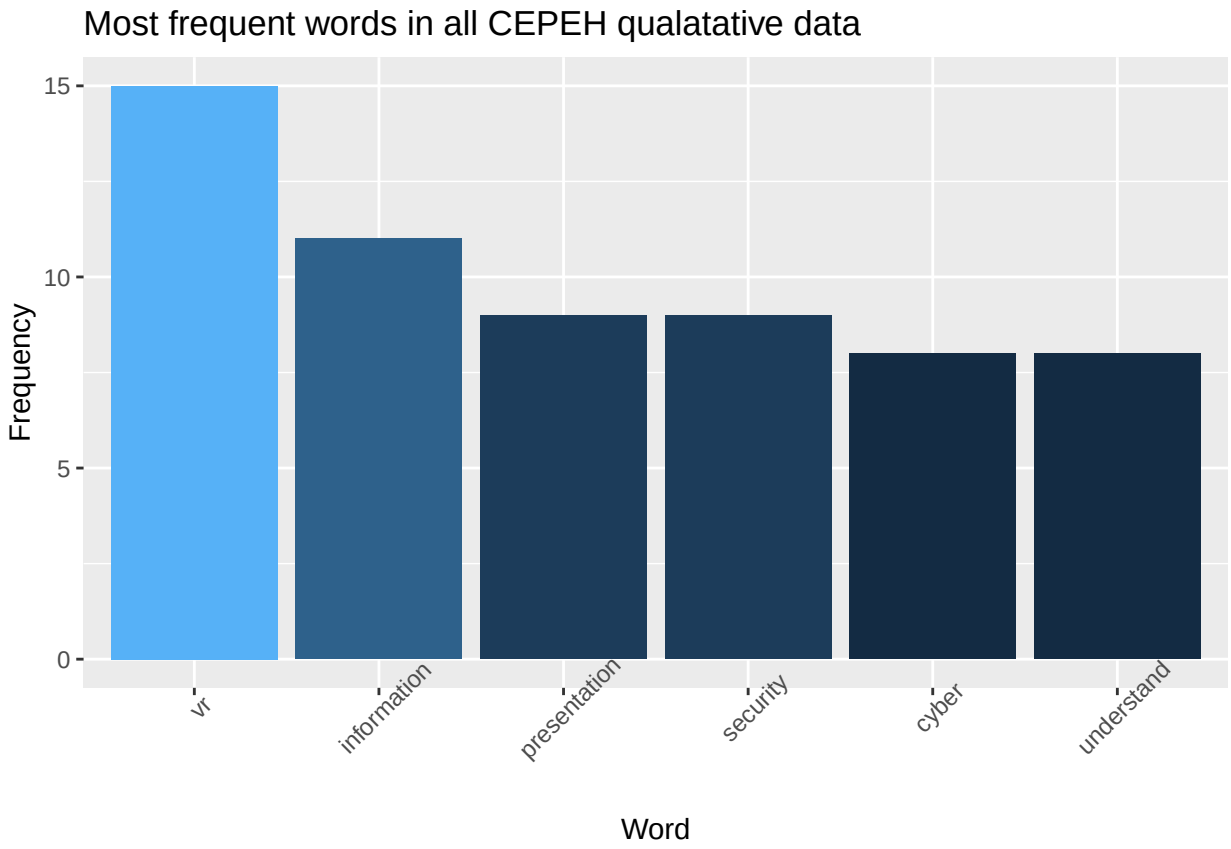
Descending order:

```
CEPEHQ_freq %>%
  filter(n>7 & doc_id == "1") %>%
  ggplot(aes(x=reorder(word, -n), y=n)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45))
```



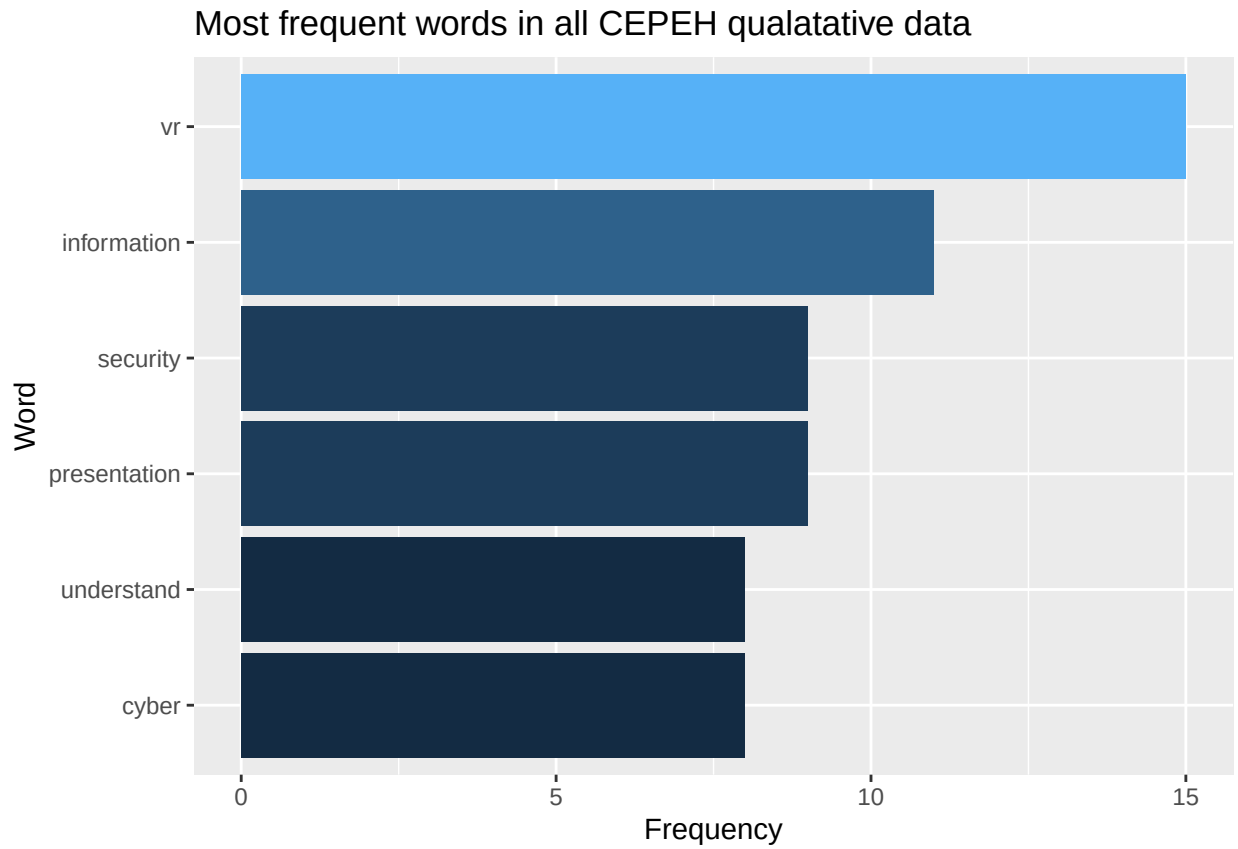
Axis names and colors:

```
CEPEHQ_freq %>%
  filter(n>7 & doc_id == "1") %>%
  ggplot(aes(x=reorder(word, -n), y=n, fill=n)) +
  geom_col(show.legend=FALSE) +
  theme(axis.text.x = element_text(angle = 45)) +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Most frequent words in all CEPEH qualatative data")
```



Or: flip coordinate system to make more space for words

```
CEPEHQ_freq %>%
  filter(n>7,doc_id == "1") %>%
  ggplot(aes(x=reorder(word, n), y=n, fill=n)) +
  geom_col(show.legend=FALSE) +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Most frequent words in all CEPEH qualatative data") +
  coord_flip()
```



1.3.2 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: raw frequency * 1,000,000 / total numbers in text

```
# see the total number of words per play (doc_id)
CEPEHQ_freq %>%
  group_by(doc_id) %>%
  mutate(sum(n)) %>%
  distinct(doc_id, sum(n))
```

```
## # A tibble: 1 x 2
## # Groups:   doc_id [1]
##   doc_id `sum(n)`
```

```
##   <fct>     <int>
## 1 1           694
```

```
CEPEHQ_freq <- CEPEHQ_freq %>%
  na.omit() %>%
  group_by(doc_id) %>%
  mutate(pmw = n*1000000/sum(n)) %>% # creates a new column called
  ↪  pmw
  ungroup() %>%
  anti_join(stop_words) # removing stopwords afterwards
```

```
## Joining, by = "word"
```

```
CEPEHQ_freq %>% select(word, pmw)
```

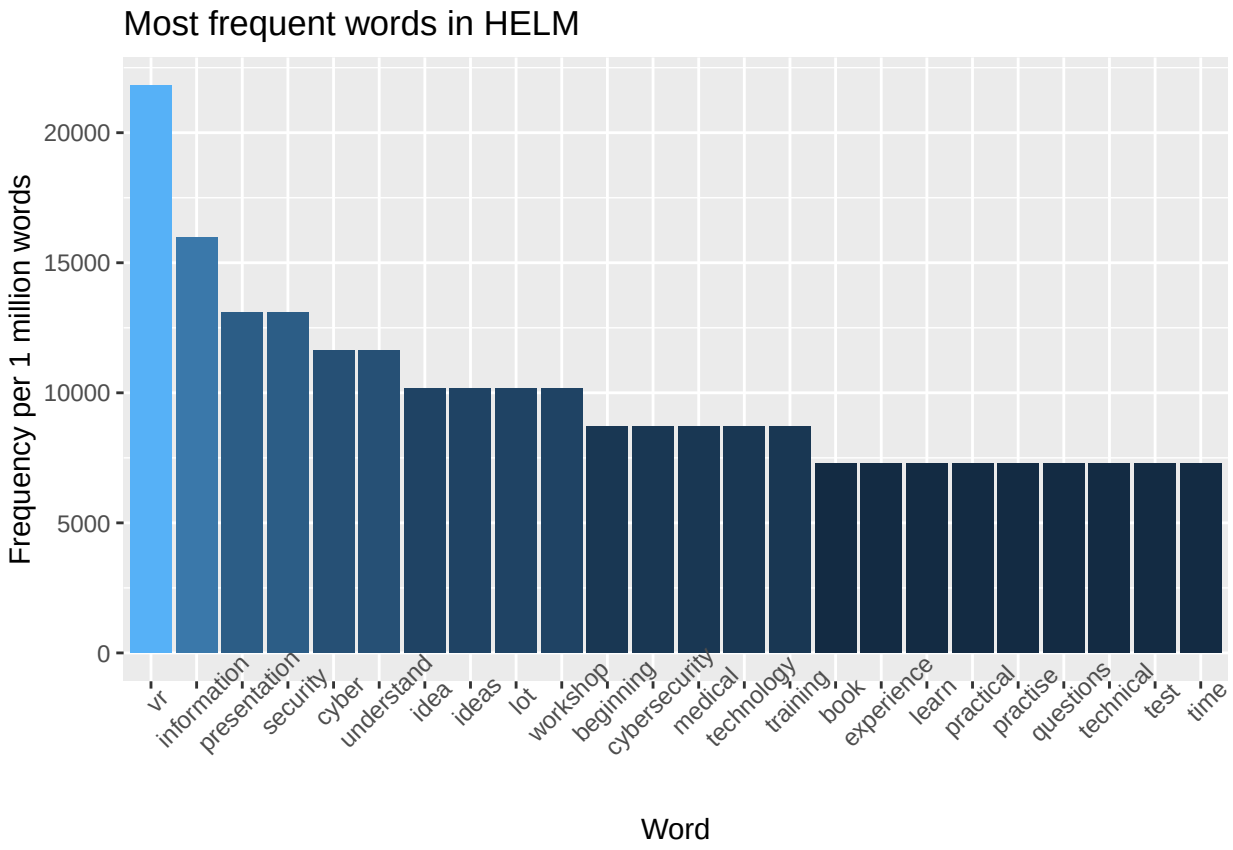
```
## # A tibble: 384 x 2
##    word            pmw
##    <chr>         <dbl>
##  1 vr          21802.
##  2 information 15988.
##  3 presentation 13081.
##  4 security    13081.
##  5 cyber       11628.
##  6 understand  11628.
##  7 idea        10174.
##  8 ideas       10174.
##  9 lot         10174.
## 10 workshop    10174.
## # ... with 374 more rows
```

Plotting normalised frequency

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

```
CEPEHQ_freq %>%
  filter(doc_id == "1") %>%
  top_n(20, pmw) %>%
```

```
ggplot(aes(x=reorder(word, -pmw), y=pmw, fill=pmw)) +
geom_col(show.legend=FALSE) +
theme(axis.text.x = element_text(angle = 45)) +
xlab("Word") +
ylab("Frequency per 1 million words") +
ggtitle("Most frequent words in HELM")
```



1.3.3 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.

```
wordcloud(words = CEPEHQ_freq$word, freq = CEPEHQ_freq$n,
          min.freq = 6, max.words=2000, random.order=FALSE,
          ↪  rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

1.4 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.

1.4.1 Comparing two texts

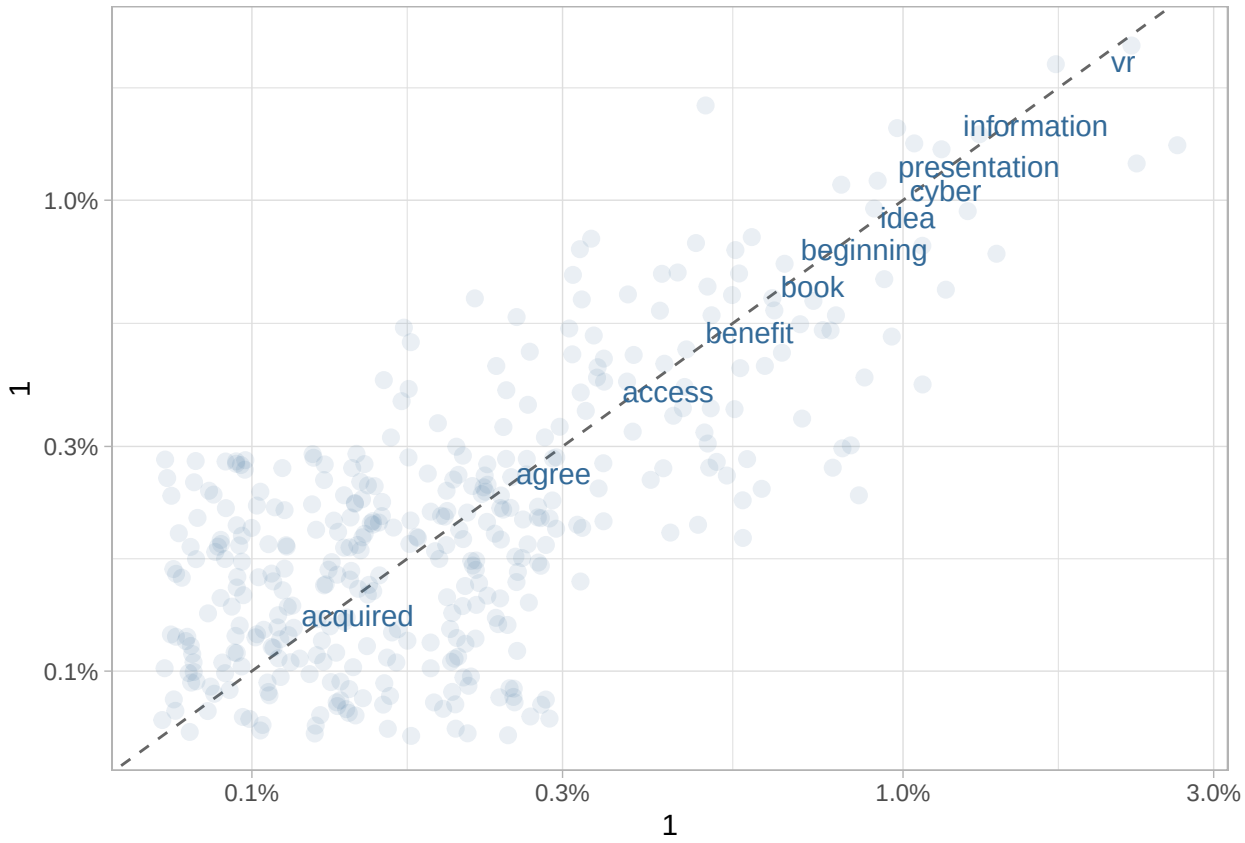
```
comp_2 <- CEPEHQ_freq %>%
  filter(doc_id == "1"|doc_id == "1") %>%
  group_by(doc_id) %>%
  mutate(proportion = n / sum(n)) %>% #creates proportion column
  ↪  (word frequency divided by overall frequency per author)
  select(-n) %>%
  spread(doc_id, proportion)
```

```
head(comp_2)
```

```
## # A tibble: 6 x 3
##   word            pmw     `1`
##   <chr>         <dbl>   <dbl>
## 1 access        4360. 0.00436
## 2 acquired      1453. 0.00145
## 3 add           1453. 0.00145
## 4 administrator 1453. 0.00145
## 5 advance       1453. 0.00145
## 6 advanced      1453. 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.

```
ggplot(comp_2,
       aes(x = `1`, y = `1`,
           color = abs(`1` - `1`))) +
  geom_abline(color = "gray40", lty = 2) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.3, height = 0.3) +
  geom_text(aes(label = word), check_overlap = TRUE, vjust = 1.5) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  theme_light() +
  theme(legend.position="none") +
  labs(y = "1", x = "1")
```




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

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

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

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

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

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

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

ACM ISBN 978-1-4503-6708-0/20/04.
http://dx.doi.org/10.1145/3313831.3376672

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].

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Title of your Thesis

John Doe

Thesis committee

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

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

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

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

Title of your thesis

John Doe

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

John Doe
Title of your thesis
77 pages.

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

ISBN XXX-YYY

For Yihui Xie

Acknowledgements

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

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

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

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

Ulrik Lyngs
Linacre College, Oxford
2 December 2018