A Case Study of the Shift to Online Education using Data Analysis Techniques

Aniruddha Das

(Student ID: 201715695)

April 29th, 2021

Simon Tam
Co-operative Education Office
Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, NL A1B 3X5

Dear Dr. Tam,

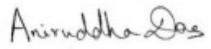
Hi,

I am currently completing my third work term under the supervision of Dr. Bipul Hawlader, Civil Engineering Department, Faculty of Engineering and Applied Science, at the Memorial University of Newfoundland. During this work term, I was primarily involved with designing new websites for the civil department professors.

The current report focuses on the shift to online education. Data analysis techniques have been used to study the collected data better to understand the common advantages and disadvantages of the system. This allows the proposal of a model to solve the resultant common drawbacks. The model is in the form of a phone companion app with several features to improve classroom interaction (a significant problem in online education). The report can also be used to identify the major requirements for the success of the same platform for different user categories.

Any questions regarding the report can be directly addressed to me at aniruddhad@mun.ca. I will answer all the questions to the best of my abilities!

Regards,



Aniruddha Das

A Case Study of the Shift to Online Education using Data Analysis Techniques

Prepared for:

Dr. Simon Tam

Faculty of Engineering and Applied Sciences

Memorial University of Newfoundland

Originating from:

Faculty of Engineering and Applied Sciences

Civil Engineering

Submitted by: Aniruddha Das (Student ID: 201715695)

Summary

The report studies the shift to online education with the help of data analysis techniques. User feedback is collected, and then data analysis is performed on the collected data. We obtain four outputs, a graph showing the most commonly used words and the frequency of their occurrence, an average sentimentality value of the data, a word cloud providing a quick glimpse of the positive and negative comments, and a graph detailing the sentimentality contribution of the relevant words, to the collected data. Finally, the most common problems are discussed, which include lack of interaction, increase in stress, lack of invigilated exams, connection issues, and noisy environment during exam time. A solution has been discussed in the form of a phone companion app, where the phone replaces the classroom interaction, while the student and professor computers emulate the class lecture. The results of this project can be improved by increasing the sample size (increasing collected data size). This project also demonstrates the usage of fundamental data analysis techniques to obtain the main problems/advantages of a system quickly.

Table of Contents

Acknowledgements iv

Statement of Scope v

1 Introduction – What is Online Technology? 1

2 Online Technology in Education 1

3 The basic working idea behind online education 2

4 Analyzing the Data 2-5

4.1 The Question for the Data 2-3

4.2 Data Storage 3

4.3 Data Analysis 3

4.4 Tidying Data 3-4

4.5 Output Results 4-5

5 Results Summary 6-11

6 Interpretation of the Obtained Results 12-14

7 Possible ways of Improving the System 14-16

7.1 Summary of Problems Discussed 14

7.2 The Solution – Digital Classroom in a Home Environment 15-16

7.2.1 What is a Digital Classroom? 15

7.2.2 Digital Classroom in a Home Environment – A Possible Working Model 15-16

8 Probable Solutions for the Other Problems 17

9 Concluding Remarks 18

10 Challenges of this Project 18

11 Future Possible Project Utilization 19

References 20

Appendix 21-25

Acknowledgments

I would like to thank Dr. Cheng Li for his approval for my project research involving professors from the ECE department.

I would like to thank Sarah Mansour for helping me distribute my email for research help to the professors in the ECE department.

I would like to thank Dr. Simon Tam, my co-op coordinator, for helping me visualize the final version of the project and for providing tips to improve my project report quality.

I would like to thank Dr. Bipul Hawlader, my work term supervisor, for his inputs regarding this report.

I would like to thank all the professors who have helped me by contributing to my research for this project.

I would like to thank Rajdeep Konar for helping me distribute my research form among students in India.

I would like to thank Daniel Grillo, Pinaak Nemade, Dani Amisi, Gavin Healy, and Kenzie Allen, among others, for helping me with my research on the side of students from Canada.

Statement of Scope

The accuracy of the results obtained by Data Analysis techniques in this report is limited by the number of users participating in the research. More responders would result in higher accuracy in obtained results.

1 Introduction - What is online technology?

The group of technologies that allow users to access information and communicate over the internet can be regarded as online technology. In other words, online technology can range from the simple web browsers that we use to access the internet from our personal computers (e.g., Google Chrome, Mozilla Firefox, Safari, Edge etc.) to the more complicated Cloud technology. Every technology we utilize to take advantage of the internet services (e.g., Email, Social Media, Food Delivery, Online drives/storages, Remoting into servers etc.) can be considered to be online technology.

2 Online technology in education

In the current report, the emphasis would be given to the educational aspect of online technology. Reasons for focusing on online education technology are –

- Online education has had a global effect with students and teachers from all over the world moving from a well established in-class education system to a comparatively new technology for delivering education
- The importance to maintain and improve the standards of education in the online education system to ensure the quality of education

3 The basic working idea behind online education

The basic idea behind online education is that students are delivered course material through internet platforms. Professors either record their lectures and upload them to an online platform to make the lessons available for students. Or, these lectures can be streamed online by the professors during regular class hours. Technologies like D2L Brightspace, Google Classroom, Blackboard Learning etc., are often utilized to create a virtual class environment. This is achieved by providing a common platform for all the students and the professor to interact in real-time, and often professors utilize these technologies to stream their lectures. These platforms are usually equipped with communication features like chats, microphones, camera (for live video streaming), and screen sharing, which effectively allows the students and professors to carry out all class activities, ranging from class discussions, tutorials to presentations and even examinations, in the most convenient way possible under the new system.

4 Analyzing the Data

4.1 The Questions for the Data

For this report, data was collected from professors and students regarding their experiences with online education. The four main points that were explored are–

- What are the advantages/disadvantages of the online teaching system?
- Ways in which the system can be improved/made easier to use.
- Additional features that could be added to the system to make it more functional.
- If there is a choice between in-class teaching and a near-perfect online teaching system, which alternative would be preferable and why?

4.2 Data Storage

The data was collected and stored in .txt format. The information was categorized into three major groups, “Professors teaching in Canada”, “Students studying in Canada”, and “Students studying in India”. Each data group has a corresponding “total.txt” file containing all the data entries of individuals in their respective groups.

4.3 Data Analysis

The primary data analysis process is done by using R language on RStudio. R scripts were written, namely “pCAN_DA.R”, “sCAN_DA.R” and “sIND_DA.R”, for analyzing their corresponding data group text files.

4.4 Tidying Data

- Inside the R script for data analysis, we first read in the “total.txt” file and convert it into a tibble.

(A tibble or tbl_df, is a modern reimagining of data frame. Tibbles are “lazy” data frames, i.e., keeps variable name unchanged, no partial matching, and

throws more complaints when a variable does not exist). This forces the problems to be confronted earlier, leading to cleaner code).**[1]**

- Each row of this tibble stores the individual lines from the text file.
- This tibble is then further modified such that the new tibble contains each individual word in each row.
- The words are then counted and sorted in descending order (words with the highest occurring frequency at the top). In case numbers (in their numeric form) are entered by the individuals, they are detected and removed from the final tibble.
- We also do a sentimentality analysis with the help of general-purpose sentiment lexicons, *bing,* and *afinn*.

(The *bing* lexicon categorizes words in a binary fashion into positive and negative categories. The *afinn* lexicon assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indication positive sentiment).**[2]**

- Finally, all the finalized data tables are exported in the ".csv" format to the working directory.

4.5 Output Results

From the tidied data table, we are outputting four results. In the following section, each result will be discussed, along with their importance to this project–

- We output a plot containing the top commonly used words in the collected data. This gives us a quick glimpse of the direction in which the

majority of the responders in that specific data category were thinking while commenting about the online education system.

- We obtain an average sentimentality score from the collected data. This is done with the help of the *afinn* lexicon. The sentimentality scores of all the words are added up and divided by the total number of words that are being analyzed.

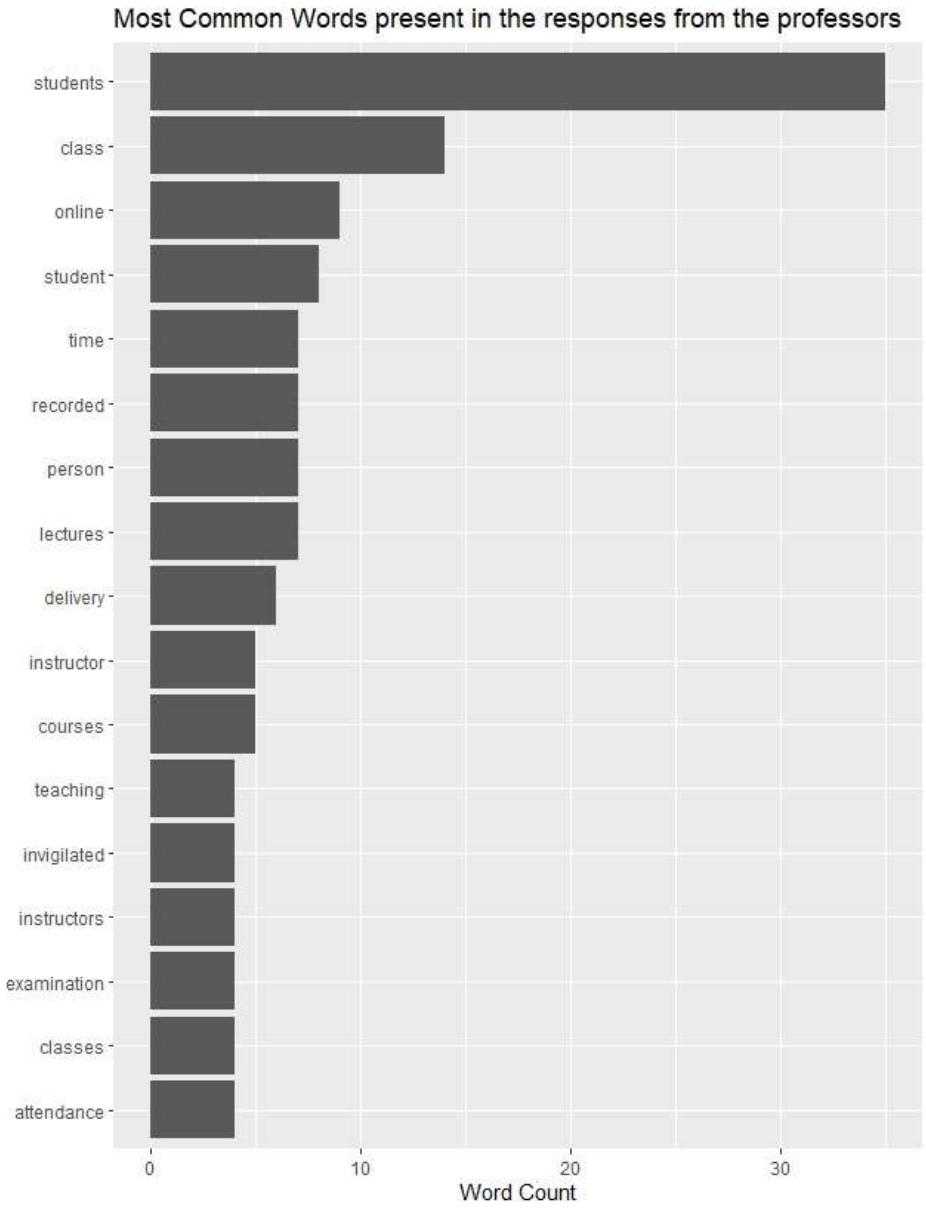
(Formula used: - Σsentimentality score/Σword.

Logic behind the formula: We are assuming that the sentimentality is uniformly spread out in the collected data, and the average sentimentality score is the representation of the tone of the collected data)

- We output a “wordcloud”, where dark grey words represent the negative words and the light grey words represent the positive words. The size of the words represents the frequency of their usage in the collected data. This gives us a quick look at the most commonly used positive and negative words in the collected data.
- We output a plot comparing the sentimentality contribution of the words to the collected data. This gives us a detailed look at the positive and negative words used in the document and how much they affect the average sentimentality value of the collected data.

5 Results Summary

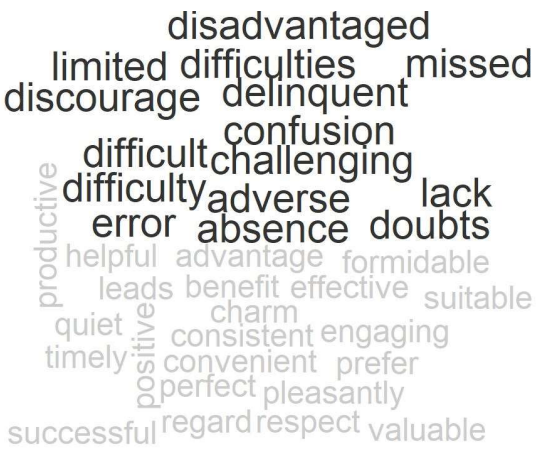
- ***Professors in Canada***
    - Plot showing the most commonly used words



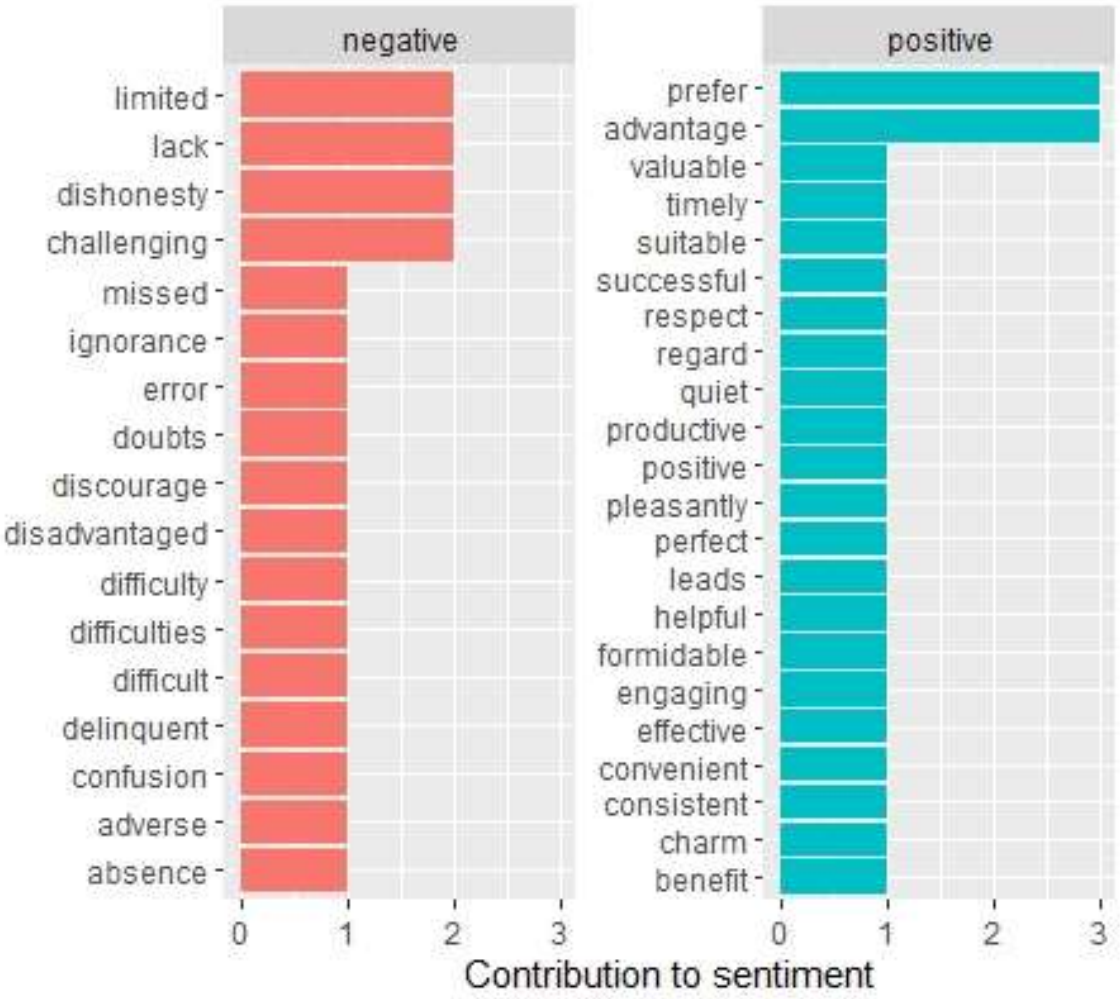
    - Average sentimentality score

      0.4444

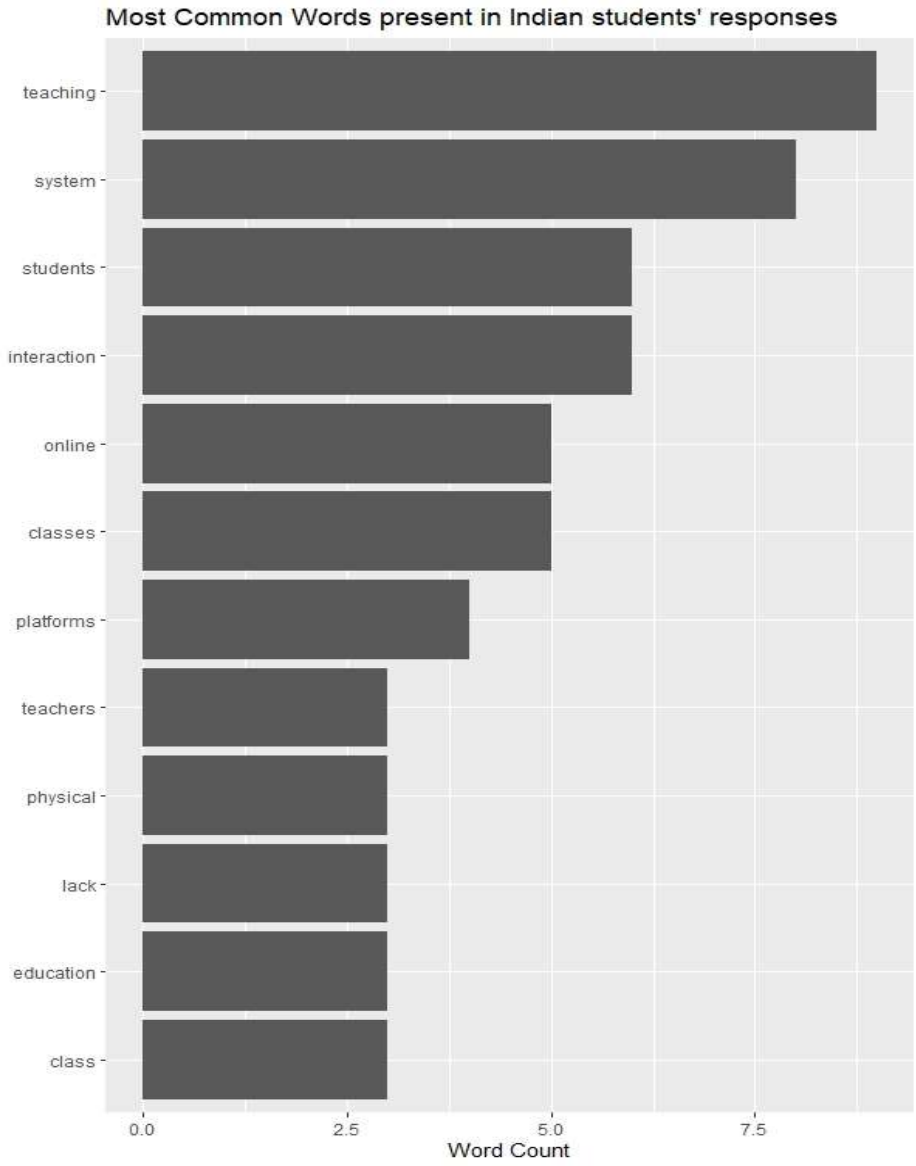
- Word cloud



- Plot demonstrating the contribution to the sentimentality of the collected text of each contributing word



- ***Students in India***
    - Plot showing the most commonly used words

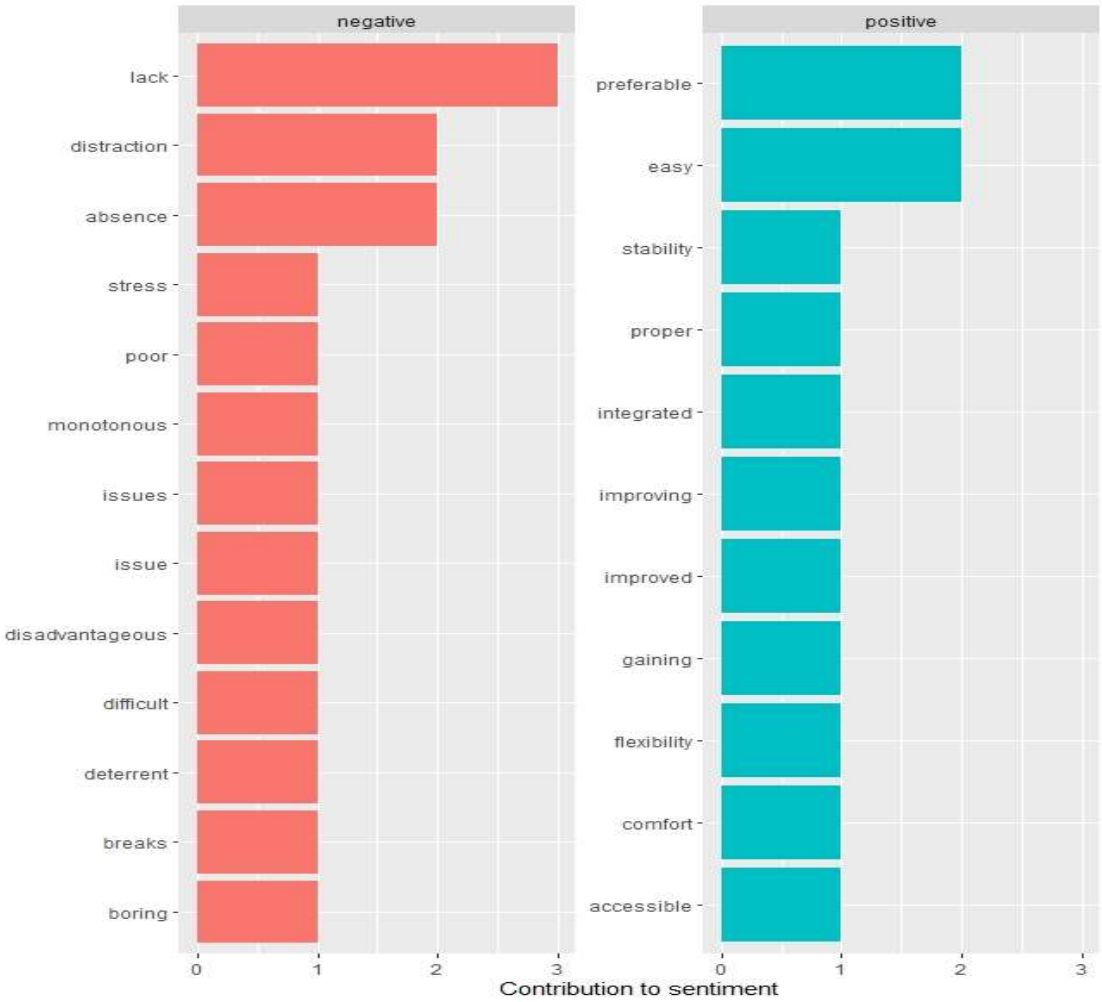


    - Average sentimentality score

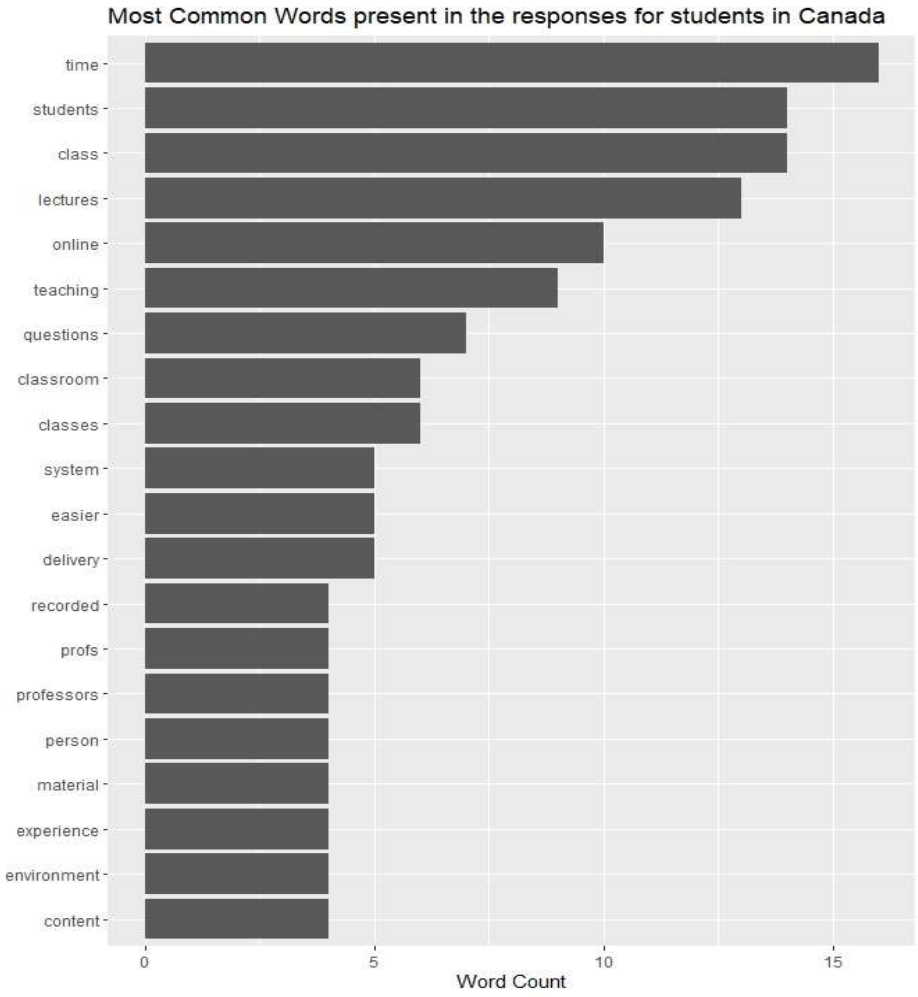
      0.2

- Word cloud

- Plot demonstrating the contribution to the sentimentality of the collected text of each contributing word



- ***Students in Canada***
    - Plot showing the most commonly used words

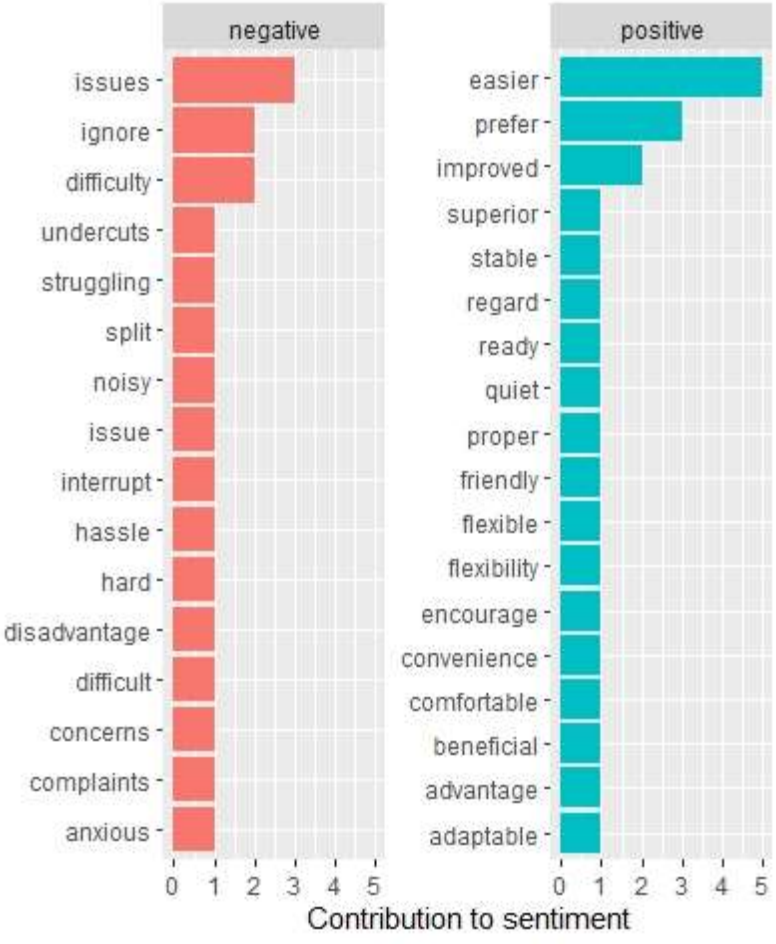


    - Average sentimentality score

      0.27273

- Word cloud

- Plot demonstrating the contribution to the sentimentality of the collected text of each contributing word



6 Interpretation of the Obtained Results

From the above results, we can interpret the following remarks –

- The most common word used by the professors is interestingly "students", with the word "student", coming at the fourth place. This indicates how much the professors think of their students every time they think about the platform which is used for education. This can also give rise to an interesting notion, for the professors, a better education platform is the platform that is better for the students.

  "Teaching" is the most common word used by the students from India in their data entry, followed by "system", "students", "interaction", and so on. From this, we can remark that for Indian students, a better education platform is one where the teaching process is the smoothest, with a focus on the students and the interaction between students and professors.

  Surprisingly, for students in Canada, "teaching" is pushed down to sixth place. The list is headed by "time", followed by "students", "class", "lectures", etc. We can say that for students in Canada, the online platform provides a good alternative, with its advantage in time management. Focus is given to "students", while achieving the best possible class environment and lecture sessions for the platform.

- The average sentimentality score of professors is the highest at 0.4444. For students from India and Canada, the scores are pretty close to each other, being 0.2 and 0.27273, respectively. This could indicate that the

professors are more accepting of the online education system. Students from India and Canada have very similar acceptability, with students from Canada having a slightly better reception to the online platform.

It could also represent the two major age groups in our data, one representing the professors and the other representing the students

- From the word cloud and “Contributions to sentimentality chart”, we can look at the positive and negative terms used in the document, which could give us an idea of the advantages and disadvantages of the system, as thought of by the responders of the three data categories.

  For professors, some advantages talked about are productivity of the system, engaging nature, successful experiences, possibility for quiet study environments, etc. Some disadvantages spoken about are the limited capabilities of the system, the difficulties faced and challenges involved in using the system properly, student absences, lack of exam invigilation, etc.

  For students from India, some advantages talked about are the flexibility and adaptability of the system, increased accessibility, and comfort, etc. Disadvantages talked about are the monotonous nature of online lectures, increased stress, and distractions, boring lectures, etc.

For students from Canada, some advantages discussed are flexibility and adaptability of the system, increased comfortable nature of the education system, etc. Problems with the system include noisy environment (lack of proper class environment for studying), interruptions between lectures, increased anxiety and difficulty of lessons and more struggling with lessons, etc.

7 Possible ways of improving the system

7.1 Summary of problems discussed –

- One of the biggest problems in the online education system is the lack of interaction between the professors and students and the absence of a class environment. This severely impacts the in-class discussions and/or activities, which could otherwise be used to make lectures more appealing for both students and professors.
- On the student side (both from India and Canada), an increase in stress/anxiety is observed.
- Lack of invigilated exams poses a significant challenge to professors.
- Connectivity issues
- Unsuitable (noisy) environment for giving an exam

7.2 The Solution – Digital Classroom in a Home Environment

7.2.1 What is a Digital Classroom?

A digital classroom can be defined as a learning environment incorporating electronic devices and software into the learning experience. It is fully immersed in technology and relies on websites and applications to enhance student learning. Digital classroom holds the potential for collaboration and/or continuing work beyond class walls. **[3]**

Even before the advent of Covid-19, and the subsequent shift to the online education system, there has been an increasing shift towards a digital classroom education system. This trend can be utilized by modeling our solution in the form of a digital classroom system but in a home environment.

7.2.2 Digital Classroom in a Home Environment – A possible working model

One of the biggest challenges of the online education system has been the reduced class interactions between students and professors and students. This problem can be addressed by utilizing our phones as the interaction medium.

The student computers and the professor's computer emulate a class lecture, and the phone interactions simulate the class interactions. Companion applications could be designed for iOS and Android to sync with the online classroom platforms being used. While most online classroom platforms do

come equipped with chat features, continuously using this chat feature can be highly disruptive to the lecture.

Instead, the phone companion app could have a chat feature, allowing students to directly send their questions to the instructor. The instructor can then address these questions at his own convenience during class hours. Students could use the application to make private student groups during lectures, increasing student interaction. The professor can also use this application to post in-class questions, or carry out in-class activities like quizzes, debates etc., seamlessly during lectures. The phone application will include camera features, which would allow students/professors to snap pictures of problems they want to share on the platform.

An increase in stress/anxiety in students can be reduced by introducing the concept of office hours on the phone application. Office hour times could be set by the professor. The professor would be online on the companion app for the duration of active online hours. Students can request bookings or drop by if the chat room is available to clear their doubts. The application would be equipped with audio and video calling features to make office hours more realistic.

A feedback system should also be included in the companion app. Feedback could be given for midterms and quizzes to further improve the examination experience during finals. Feedback can also be asked for at the end of every month to constantly improve the quality of lectures and reduce anxiety among students.

8 Probable Solutions for the other problems

- The use of earbuds and earmuffs can be popularized to help students focus in a noisy environment during exam hours.
- It is challenging to have invigilated exams in an online environment. While proctored exams are definitely the closest alternative, they are not preferred by the majority of the students. One possible way to address this situation would be to have two distinct exam paper styles–
    - Moderate to high difficulty paper including questions requiring out-of-the-box thinking and application of the logic involved in the questions. Exams could be open book with sufficient time. This would significantly decrease the ease of cheating among students while answering, even if they have lengthy answering hours.
    - Low to low-moderate difficulty paper including short questions emphasizing on testing the understandings of the basic working logic behind the question. Paper duration would be short, preventing students from having the time to look up answers.

  (For question selection, a list of questions could be prepared (e.g., twenty questions), out of which a certain number of questions (e.g., five questions) can be selected at random for each student. This would further randomize the chance of having common questions between any communicating students during exam hours).

9 Concluding Remarks

On a scale of -5 to 5, the average sentimentality score of the collected data for each data group is around 0 (midway), indicating the acceptability of the online education system by students and professors and the room for improvement.

Using Data Analysis techniques, the most common ideas in the data submissions of the professors and students have been explored in this report. Subsequently, a working model was proposed to address the most common challenges in the online education system, as indicated by the results from data analysis of the collected data.

Through this report, we have also demonstrated how fundamental Data Analysis techniques can be used to explore user feedback to identify the most common problems and advantages of a system.

10 Challenges of this project

The results of this project would be more accurate with more data to explore (i.e., increasing the sample space).

11 Future possible project utilization

The logic and the code used in this project can be utilized to study the advantages and disadvantages of the online education system in terms of country, age group, level of education, and so on.

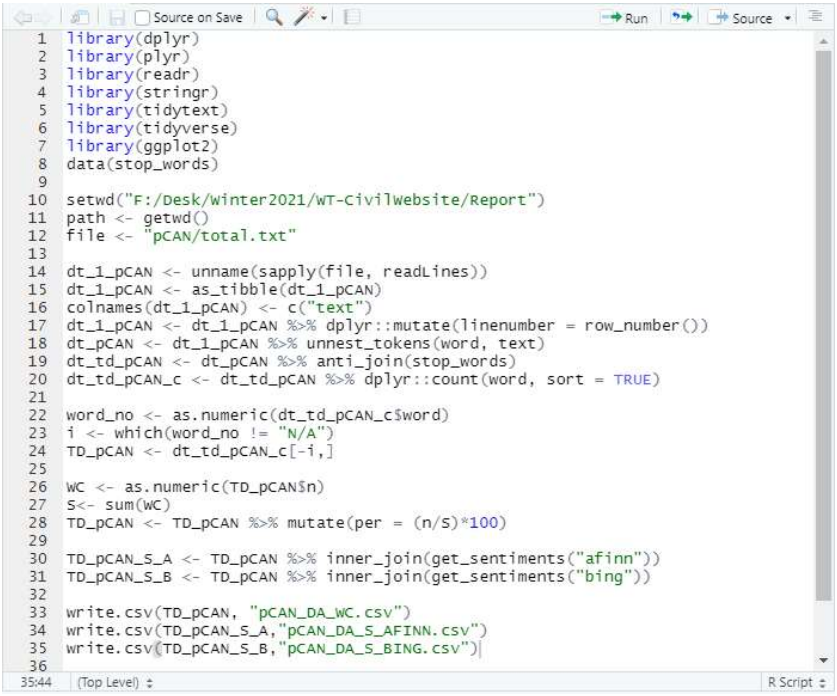
References

[1] Krill Muller, Hadley Wickham. "tibble" tibble.tidyverse.org https://tibble.tidyverse.org/#:~:text=Overview,modern%20reimagining%20of%20the%20data.&text=Tibbles%20are%20data.,a%20variable%20does%20not%20exist). (last accessed 29 April, 2021)

[2] Julia Silge & David Robinson. "Sentiment Analysis with tidy data" in *Text Mining with R: A Tidy Approach*, ch 2, sec. 2.1. [Online]. Available: https://www.tidytextmining.com/sentiment.html

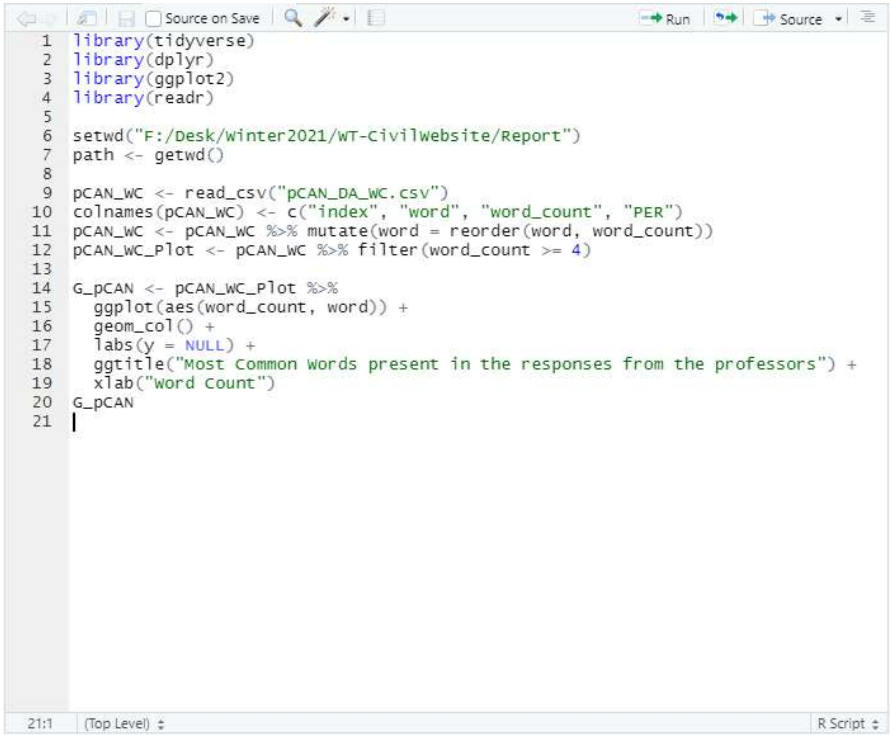
[3] Top Hat. "Digital Classroom" https://tophat.com/glossary/d/digital-classroom/#:~:text=A%20digital%20classroom%20is%20typically,software%20into%20the%20learning%20environment.&text=A%20digital%20classroom%20refers%20to,websites%20to%20enhance%20student%20learning. (last accessed 29 April, 2021)

Appendix



```
library(dplyr)
library(plyr)
library(readr)
library(stringr)
library(tidytext)
library(tidyverse)
library(ggplot2)
data(stop_words)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()
file <- "pCAN/total.txt"

dt_1_pCAN <- unname(sapply(file, readLines))
dt_1_pCAN <- as_tibble(dt_1_pCAN)
colnames(dt_1_pCAN) <- c("text")
dt_1_pCAN <- dt_1_pCAN %>% dplyr::mutate(linenumber = row_number())
dt_pCAN <- dt_1_pCAN %>% unnest_tokens(word, text)
dt_td_pCAN <- dt_pCAN %>% anti_join(stop_words)
dt_td_pCAN_c <- dt_td_pCAN %>% dplyr::count(word, sort = TRUE)

word_no <- as.numeric(dt_td_pCAN_c$word)
i <- which(word_no != "N/A")
TD_pCAN <- dt_td_pCAN_c[-i,]

WC <- as.numeric(TD_pCAN$n)
S<- sum(WC)
TD_pCAN <- TD_pCAN %>% mutate(per = (n/S)*100)

TD_pCAN_S_A <- TD_pCAN %>% inner_join(get_sentiments("afinn"))
TD_pCAN_S_B <- TD_pCAN %>% inner_join(get_sentiments("bing"))

write.csv(TD_pCAN, "pCAN_DA_WC.csv")
write.csv(TD_pCAN_S_A,"pCAN_DA_S_AFINN.csv")
write.csv(TD_pCAN_S_B,"pCAN_DA_S_BING.csv")
```

Fig1. Code for tidying data entry for professors in Canada and outputting required data tables

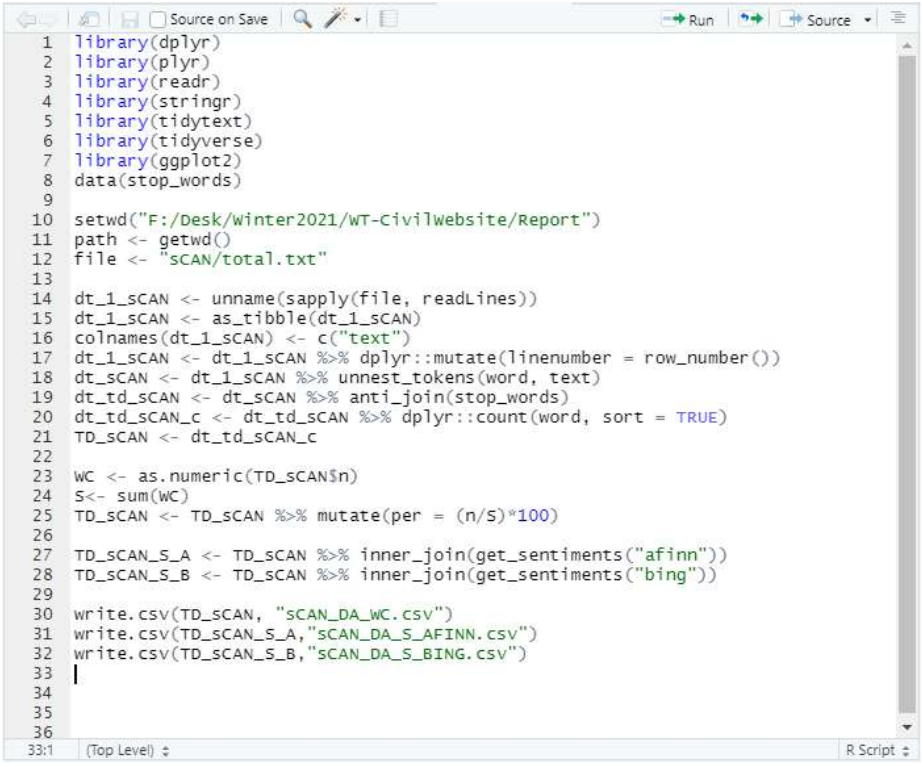


```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()

pCAN_WC <- read_csv("pCAN_DA_WC.csv")
colnames(pCAN_WC) <- c("index", "word", "word_count", "PER")
pCAN_WC <- pCAN_WC %>% mutate(word = reorder(word, word_count))
pCAN_WC_Plot <- pCAN_WC %>% filter(word_count >= 4)

G_pCAN <- pCAN_WC_Plot %>%
  ggplot(aes(word_count, word)) +
  geom_col() +
  labs(y = NULL) +
  ggtitle("Most Common Words present in the responses from the professors") +
  xlab("Word Count")
G_pCAN
```

Fig2. Code for plot for most commonly words used in collected data

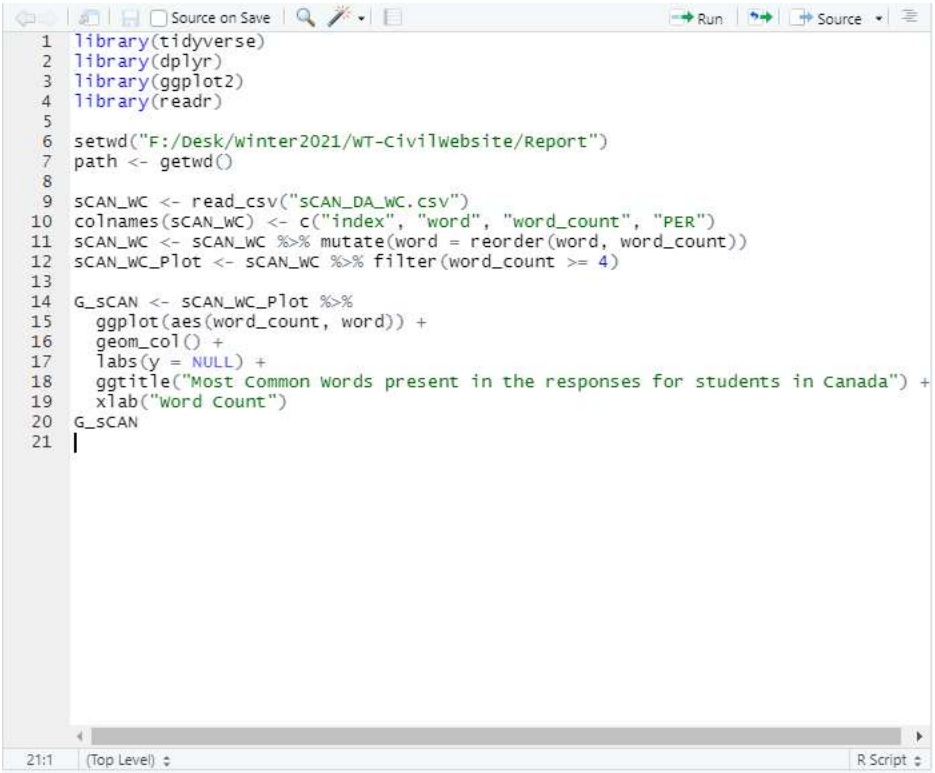
```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)
library(reshape2)
library(tm)
library(wordcloud)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()

pCAN_S_A <- read_csv("pCAN_DA_S_AFINN.csv")
pCAN_S_B <- read_csv("pCAN_DA_S_BING.csv")
colnames(pCAN_S_A) <- c("index", "word", "word_count", "PER", "value")
colnames(pCAN_S_B) <- c("index", "word", "word_count", "PER", "sentiment")

AVG_pCAN <- sum(pCAN_S_A$value*pCAN_S_A$word_count)/sum(pCAN_S_A$word_count)

I_pCAN_pnw <- pCAN_S_B %>%
  dplyr::count(word, sentiment, sort=TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors =c("gray20", "gray80"))

G_pCAN_S <- pCAN_S_B %>%
  group_by(sentiment) %>%
  ungroup() %>%
  mutate(word = reorder(word, word_count)) %>%
  ggplot(aes(word_count, word, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment,scales = "free_y") +
  labs(x = "Contribution to sentiment", y = NULL)
G_pCAN_S
```

Fig3. Code for calculating average data sentimentality score and producing the wordcloud and the graph representing the sentimentality contribution of the relevant words



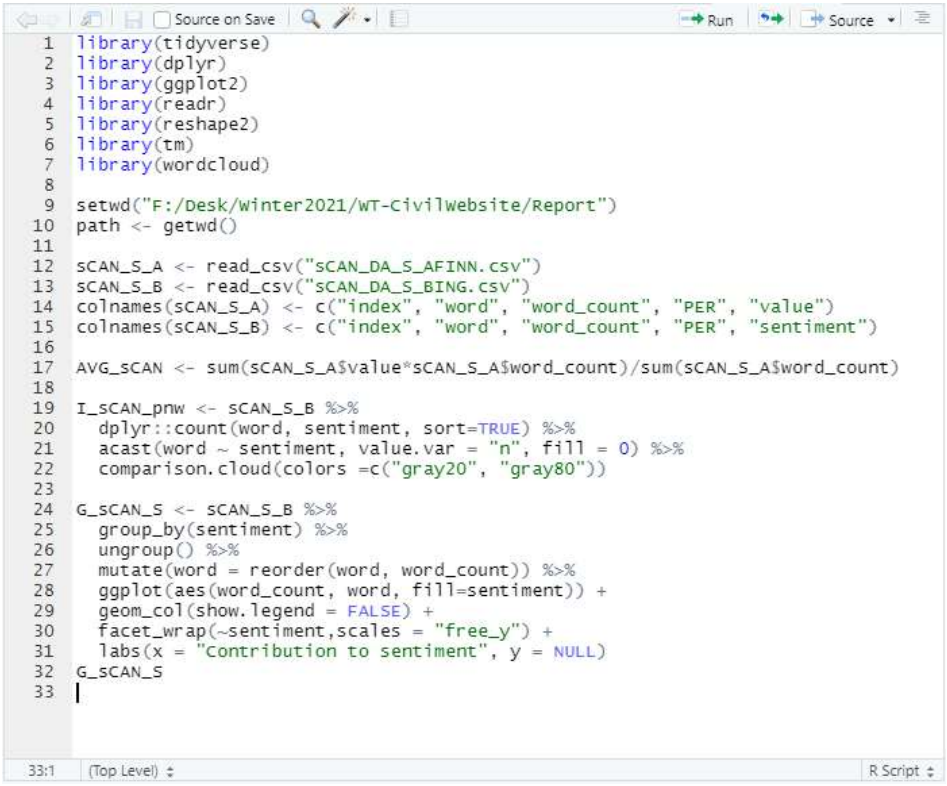
```
library(dplyr)
library(plyr)
library(readr)
library(stringr)
library(tidytext)
library(tidyverse)
library(ggplot2)
data(stop_words)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()
file <- "sCAN/total.txt"

dt_1_sCAN <- unname(sapply(file, readLines))
dt_1_sCAN <- as_tibble(dt_1_sCAN)
colnames(dt_1_sCAN) <- c("text")
dt_1_sCAN <- dt_1_sCAN %>% dplyr::mutate(linenumber = row_number())
dt_sCAN <- dt_1_sCAN %>% unnest_tokens(word, text)
dt_td_sCAN <- dt_sCAN %>% anti_join(stop_words)
dt_td_sCAN_c <- dt_td_sCAN %>% dplyr::count(word, sort = TRUE)
TD_sCAN <- dt_td_sCAN_c

WC <- as.numeric(TD_sCAN$n)
S<- sum(WC)
TD_sCAN <- TD_sCAN %>% mutate(per = (n/S)*100)

TD_sCAN_S_A <- TD_sCAN %>% inner_join(get_sentiments("afinn"))
TD_sCAN_S_B <- TD_sCAN %>% inner_join(get_sentiments("bing"))

write.csv(TD_sCAN, "sCAN_DA_WC.csv")
write.csv(TD_sCAN_S_A,"sCAN_DA_S_AFINN.csv")
write.csv(TD_sCAN_S_B,"sCAN_DA_S_BING.csv")
```

Fig4. Code for tidying data entry for students in Canada and outputting required data tables



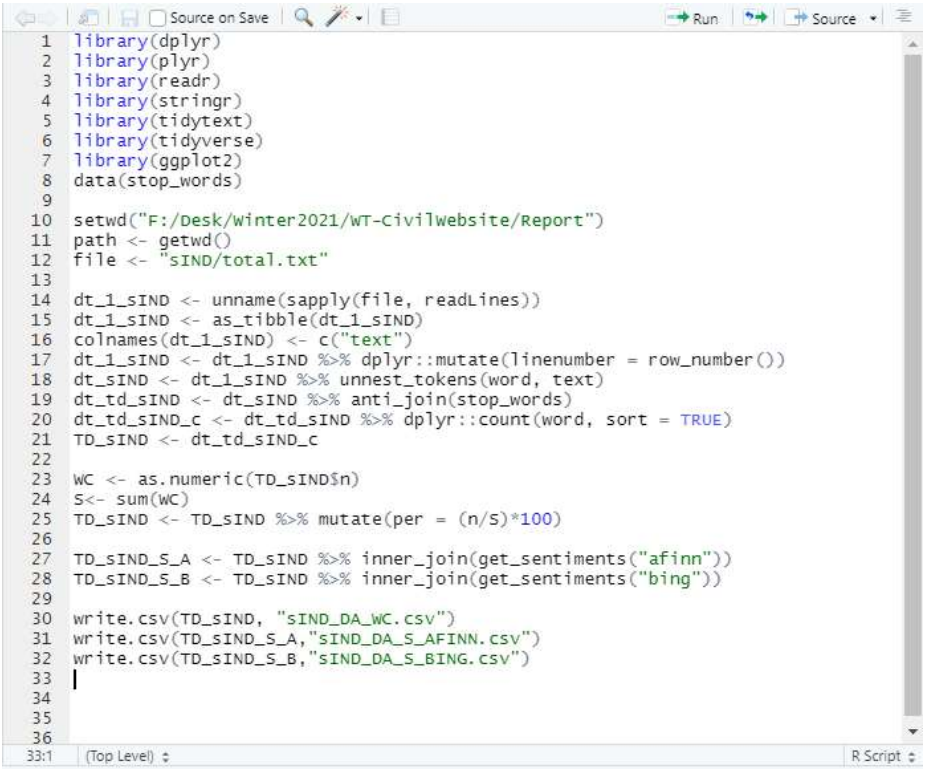
```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()

sCAN_WC <- read_csv("sCAN_DA_WC.csv")
colnames(sCAN_WC) <- c("index", "word", "word_count", "PER")
sCAN_WC <- sCAN_WC %>% mutate(word = reorder(word, word_count))
sCAN_WC_Plot <- sCAN_WC %>% filter(word_count >= 4)

G_sCAN <- sCAN_WC_Plot %>%
  ggplot(aes(word_count, word)) +
  geom_col() +
  labs(y = NULL) +
  ggtitle("Most Common Words present in the responses for students in Canada") +
  xlab("Word Count")
G_sCAN
```

Fig5. Code for plot for most commonly words used in collected data



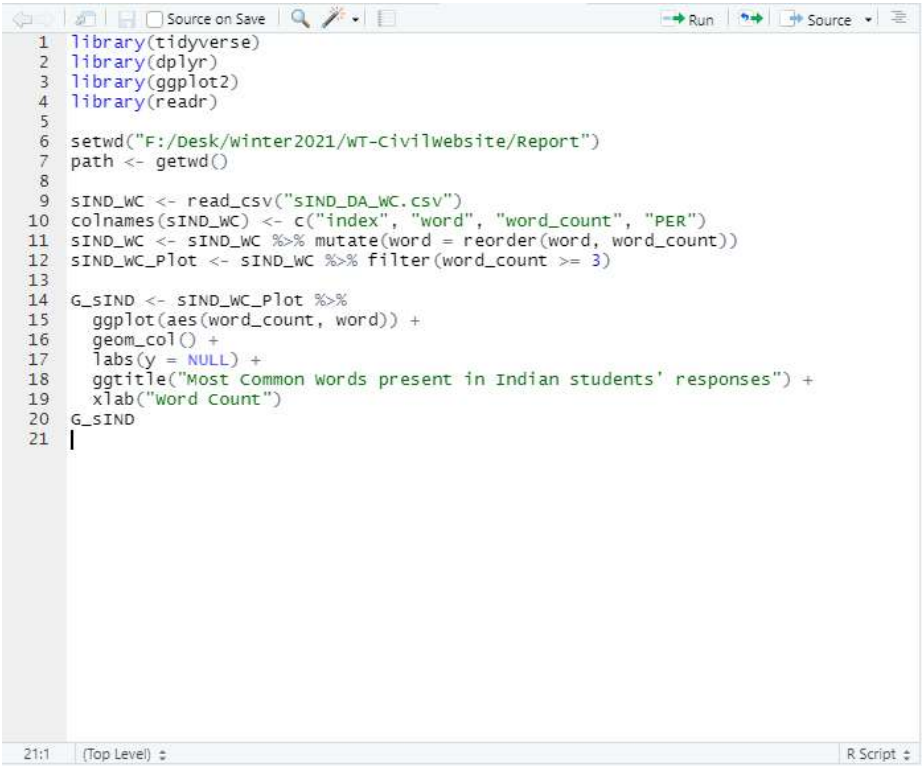
```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)
library(reshape2)
library(tm)
library(wordcloud)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()

sCAN_S_A <- read_csv("sCAN_DA_S_AFINN.csv")
sCAN_S_B <- read_csv("sCAN_DA_S_BING.csv")
colnames(sCAN_S_A) <- c("index", "word", "word_count", "PER", "value")
colnames(sCAN_S_B) <- c("index", "word", "word_count", "PER", "sentiment")

AVG_sCAN <- sum(sCAN_S_A$value*sCAN_S_A$word_count)/sum(sCAN_S_A$word_count)

I_sCAN_pnw <- sCAN_S_B %>%
  dplyr::count(word, sentiment, sort=TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors =c("gray20", "gray80"))

G_sCAN_S <- sCAN_S_B %>%
  group_by(sentiment) %>%
  ungroup() %>%
  mutate(word = reorder(word, word_count)) %>%
  ggplot(aes(word_count, word, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment,scales = "free_y") +
  labs(x = "Contribution to sentiment", y = NULL)
G_sCAN_S
```

Fig6. Code for calculating average data sentimentality score and producing the wordcloud and the graph representing the sentimentality contribution of the relevant words



```r
library(dplyr)
library(plyr)
library(readr)
library(stringr)
library(tidytext)
library(tidyverse)
library(ggplot2)
data(stop_words)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()
file <- "sIND/total.txt"

dt_1_sIND <- unname(sapply(file, readLines))
dt_1_sIND <- as_tibble(dt_1_sIND)
colnames(dt_1_sIND) <- c("text")
dt_1_sIND <- dt_1_sIND %>% dplyr::mutate(linenumber = row_number())
dt_sIND <- dt_1_sIND %>% unnest_tokens(word, text)
dt_td_sIND <- dt_sIND %>% anti_join(stop_words)
dt_td_sIND_c <- dt_td_sIND %>% dplyr::count(word, sort = TRUE)
TD_sIND <- dt_td_sIND_c

WC <- as.numeric(TD_sIND$n)
S<- sum(WC)
TD_sIND <- TD_sIND %>% mutate(per = (n/S)*100)

TD_sIND_S_A <- TD_sIND %>% inner_join(get_sentiments("afinn"))
TD_sIND_S_B <- TD_sIND %>% inner_join(get_sentiments("bing"))

write.csv(TD_sIND, "sIND_DA_WC.csv")
write.csv(TD_sIND_S_A,"sIND_DA_S_AFINN.csv")
write.csv(TD_sIND_S_B,"sIND_DA_S_BING.csv")
```

Fig7. Code for tidying data entry for students in India and outputting required data tables



```r
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)

setwd("F:/Desk/Winter2021/WT-CivilWebsite/Report")
path <- getwd()

sIND_WC <- read_csv("sIND_DA_WC.csv")
colnames(sIND_WC) <- c("index", "word", "word_count", "PER")
sIND_WC <- sIND_WC %>% mutate(word = reorder(word, word_count))
sIND_WC_Plot <- sIND_WC %>% filter(word_count >= 3)

G_sIND <- sIND_WC_Plot %>%
  ggplot(aes(word_count, word)) +
  geom_col() +
  labs(y = NULL) +
  ggtitle("Most Common Words present in Indian students' responses") +
  xlab("Word Count")
G_sIND
```

Fig8. Code for plot for most commonly words used in collected data

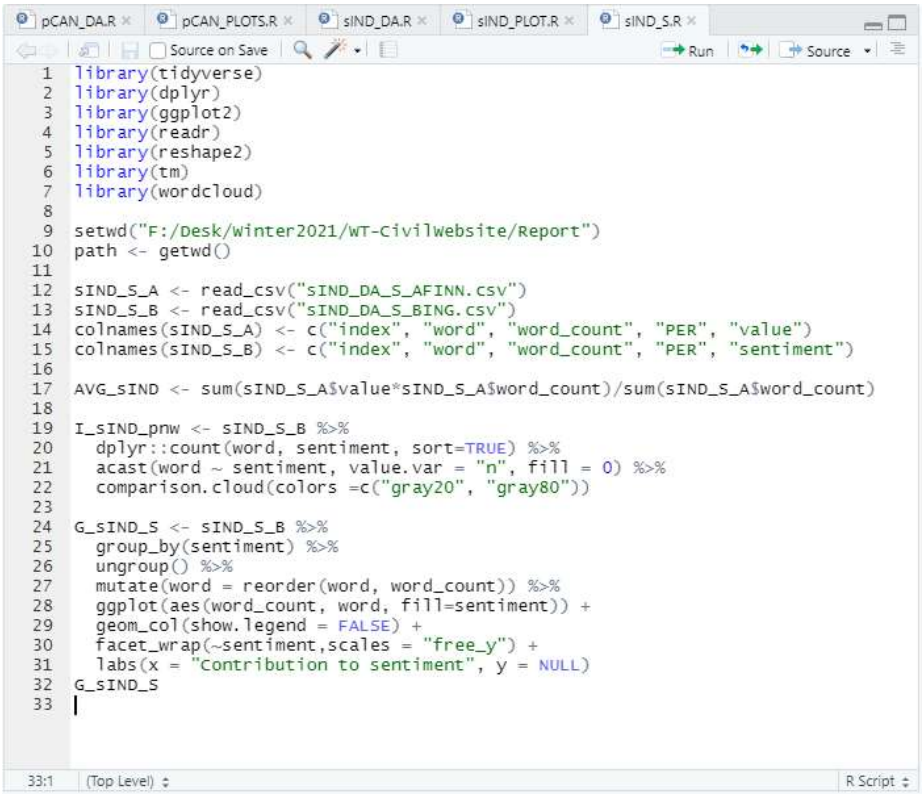


Fig6. Code for calculating average data sentimentality score and producing the wordcloud and the graph representing the sentimentality contribution of the relevant words

Fig1, Fig2 and Fig3 work with data group for professors in Canada Fig4, Fig5 and Fig6 work with data group for students in Canada and Fig7, Fig8 and Fig9 work with data group for students in India.

All R script code, raw data and data tables used in this project are available at https://github.com/AniDas2403/WT3_DataAnalysis .