RESEARCH TITLE:

Analysis of Qualitative Data from StudentSurvey.ie (the Irish Survey of Student Engagement) 2022 using Power BI

AUTHORS:

Dr. Stephen Erskine & David Harmon

ORGANISATION/INSTITUTION:

Insight Statistical Consulting

Student Survey.ie

PGR Student Survey.ie

Foreword

The StudentSurvey.ie Steering Group is pleased to publish the results of five research projects analysing the qualitative data generated by the free-text response questions in StudentSurvey.ie and PGR StudentSurvey.ie. The results contained within this report make up one part of this research series.

Five projects were funded by research bursaries offered by StudentSurvey.ie in October 2022. The aim of the bursary awards was to promote greater ownership and encourage wider use of the StudentSurvey.ie and PGR StudentSurvey.ie data. Proposals for the analysis of the qualitative data emerging from StudentSurvey.ie and PGR StudentSurvey.ie were invited from members of the research community within the participating institutions, as well as commercial data analysis companies. The projects were completed in May 2023.

Each project is an independent project undertaken by qualified and experienced researchers on behalf of StudentSurvey.ie. Each project took a unique approach. The commonalities between all five projects are that they all utilised well-grounded methodologies, offer mechanisms for replication of the analysis in future years, and are innovative and authentic.

These results are the first of their kind for StudentSurvey.ie and PGR StudentSurvey.ie and we hope they are the first of many research projects involving the qualitative results of these surveys.

What are StudentSurvey.ie and PGR StudentSurvey.ie?

StudentSurvey.ie (the Irish Survey of Student Engagement) is an annual national survey of student engagement among first year undergraduate, final year undergraduate and taught postgraduate students in higher education institutions in Ireland.

PGR StudentSurvey.ie (the Irish Survey of Student Engagement for Postgraduate Research Students) is a biennial national survey of student engagement among Masters by Research students and PhD students in higher education institutions in Ireland.

Both surveys are designed to focus on student engagement, namely the amount of time and effort that students put into meaningful and purposeful educational activities, and the extent to which institutions provide such opportunities and encourage students to engage with them. The data collected reflect students' self-reported perceptions of their experiences.

Report on the Analysis of Qualitative Data from StudentSurvey.ie (the Irish Survey of Student Engagement) 2022 using Power BI

Authors: Dr. Stephen Erskine
David Harmon

Insight Statistical Consulting
60 Merrion Square
Dublin 2, Ireland
www.insightsc.ie

Date: March 2023

Contents

1. Introduction and Background

StudentSurvey.ie (the Irish Survey of Student Engagement) is an annual national survey of student engagement among first year undergraduate, final year undergraduate, and taught postgraduate students in 21 higher education institutions in Ireland.

The survey is designed to focus on student engagement, namely the amount of time and effort that students put into meaningful and purposeful educational activities, and the extent to which institutions provide such opportunities and encourage students to engage with them. The data collected reflect students' self-reported perceptions of their experiences.

The qualitative data relevant to this report are generated by two questions:

1. What does your institution do best to engage students in learning?
2. What could your institution do to improve students' engagement in learning?

In 2022, these questions were answered by approximately 22,700 respondents. Students may answer in Irish or English, with the majority choosing to answer in English. These questions have been in use since 2016, though the number of students who answer them has increased significantly over the years since then. Significant quantitative data are collected by survey from 67 multiple choice questions, and numerous student and course characteristic variables are also recorded.

1.1 Methodological Approach

The two qualitative questions used in the survey have generated an extremely large corpus of data which as a resource, for understanding students' thoughts and experiences in higher education, has until recently been under-examined[1]. This under-examination of the material

[1] Irish Survey of Student Engagement. National Report 2020, p79.
- Erskine, S and Harmon, D. Report on the Analysis of Qualitative Data from StudentSurvey.ie (the Irish Survey of Student Engagement) 2016 – 2020.

was perfectly understandable as the volume of material produced by these two questions each year is significant, and by its nature, requires a different approach to standard qualitative analysis because of the sheer scale of the material.

For example, in our previous analysis, if one was to take the comments provided by students from 2016 to 2020 and place them into a standard A4 document, it would amount to almost 8,000 pages of A4 text. For an average reader this would take approximately 200 hours to read through once, and even longer if one wanted to code the material, identify themes, add comments, elaborate to set of generalisations, and collate all of the material found into a document to report the findings as a sequence of tasks necessary to conduct qualitative research[2].

With this amount of data, steps had to be taken to handle the corpus of information in a suitable manner as hand-coding of material becomes too unwieldly to be practical. As such, our previous work combined researcher-led coding of the material with computer-assisted content analysis methods to draw out key information on what students think about student engagement in their higher education institutions. In addition, this report provided a framework for further research to build upon as our analysis utilised the open-source statistical software *R* and its graphical interface R Studio, along with a suite of R packages. Because of this, our project is replicable, and our coding framework and code can be extended to further iterations of the Irish Survey of Student Engagement.

However, from the discussions arising from the presentation of our work during the first round of qualitative data analysis projects, it was evident that although we designed our project to be available to as many people as were interested in it at no cost, there still remained a significant knowledge gap, in that people who were not experienced in using the R programming language felt that their absence of experience with the language meant that they were excluded from utilising efficiently the methods we had developed.

[2] Robson, C. Real World Research. (Blackwell Publishing, 2002), p459.

To rectify this, we proposed a means to bridge this knowledge gap and allow researchers immediate access to the StudentSurvey.ie qualitative data by developing an interactive data visualisation platform using Power BI, which is an intuitive software product developed by Microsoft.

At the outset, we did not know if it would be possible to replicate what we had done in R in another software format. We thought that there would be a trade-off between detail and accessibility in that our work with R was more in-depth but less accessible to researchers without a background or experience in programming. In contrast, this project would not be as in-depth but will be immediately accessible to more users from varying research backgrounds. Ideally, interested researchers could use this project to explore the qualitative data collected and if the results of this merit further analysis they can utilise our earlier work to assist them with going further with their analysis.

However, for the most part, the translation between programs has worked well, and by our estimation, we have been able to convert over eighty percent of the material covered in the previous report into a usable Power BI format, and to extend this in ways that were not possible in R. As such, this report covers in detail, the steps taken to replicate our work done in R as thoroughly as was possible, and a 'how to' guide for other researchers to use the Power BI template we have built.

At the outset, we were uncertain how much actual qualitative analysis could be conducted in Power BI itself, it was only when we learned that a recent addition to Power BI allowed users to incorporate R programming, certain R packages, and run R scripts from within Power BI itself did we gain confidence that this project would not be for naught. Instead, we found that using R within Power BI is another good way of bridging the experience gap, and in addition, Power BI supplemented the R work through its addition of filters which makes extending the analysis across groups of interest extremely easy, and will be discussed further below.

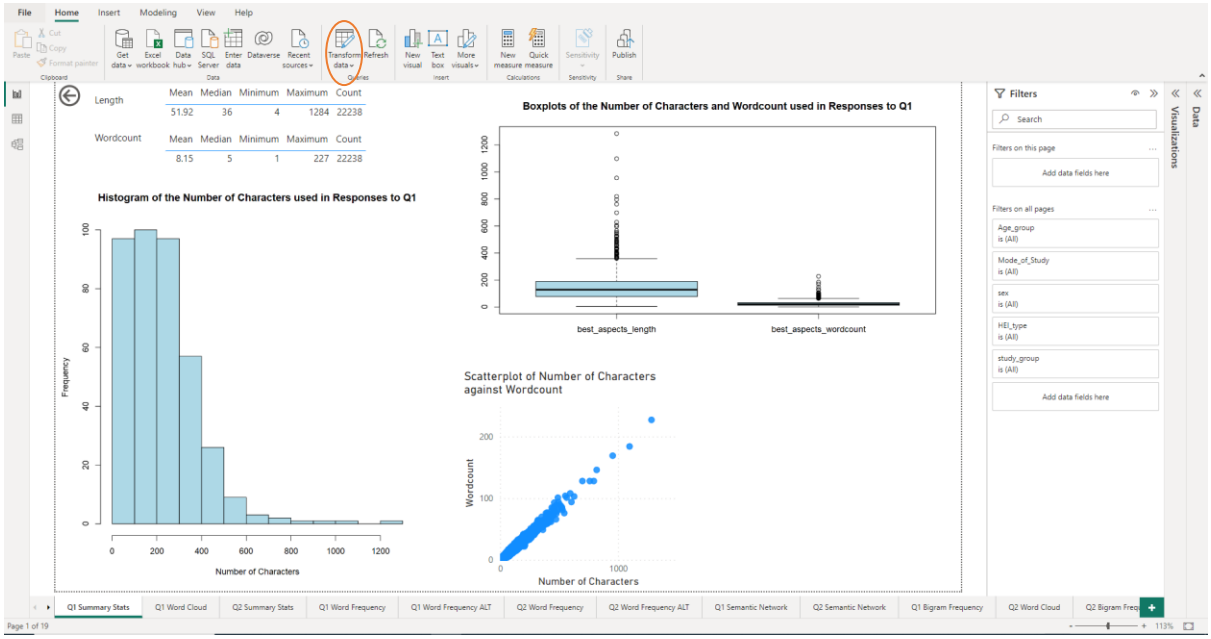
1.2 Set-up of the Power BI Report

To examine the Power BI report, one needs three things:

1. Power BI desktop installed on their computer,
2. The Power BI Report called ‘ISSE22.pbix’, and
3. The data itself, called ‘StudentSurvey.ie Data 2022.xlsx’.

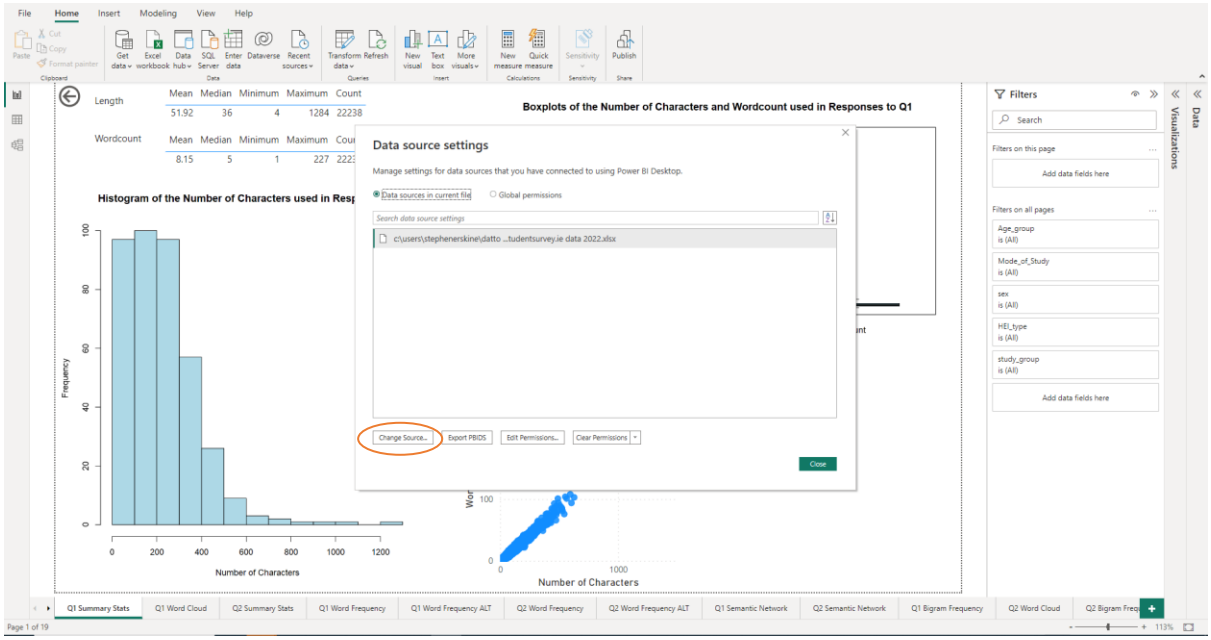
The data and the Power BI report should be saved to a known location on the computer as it will be necessary when using the Power BI Report to navigate to the data so that all the tabs present in the report can be populated from the data. The picture below shows the Power BI Report with the Transform data button encircled.

Figure 1.1: Setting up the data source within the Power BI report (1)



This needs to be pressed and a menu will appear. ‘Data source settings’ should be selected then a pop-up menu will appear.

Figure 1.2: Setting up the data source within the Power BI report (2)



The picture above shows this pop-up menu, at this point the 'change source' button on the bottom right of the menu should be selected, then the full location of the data be entered by the user. Once this has been done, the report will draw upon this location as the source of the data used by each tab of the report.

2. "What does your institution do best to engage students in learning?"

This chapter of the report covers the analysis undertaken in Power BI to provide some answers to the open-text question posed to students; "What does your institution do best to engage students in learning?" Which for the shorthand purposes will hereby be referred to as Q1. This chapter covers the analysis conducted to break this material into more readily comprehensible units first, and then discusses the subsequent analysis which reintegrates these units to gain a systematic understanding of students' responses which can then be used to illuminate areas in which HEIs are succeeding in engaging students.

As noted already, the sheer volume of data provided by students in their responses to this question is the key difficulty in evaluating the content of students' responses. As such, with this amount of information it becomes all too easy to be overwhelmed unless a strategy to break the data down into more manageable units of analysis is utilised.

This was achieved by taking the data through a series of steps which changed it from a 'raw' format to being 'cleaned' to ensure that all the data contained is treated equally by the statistical software. A copy of the original Q1 variable is kept throughout as a reference and quality check.

The first step in this case was gaining a familiarity with the data by reading over many of the comments. This was very time-intensive but invaluable in understanding how students approached answering the question at hand.

This step was conducted in tandem with a comprehensive spell check which tried to ensure that any common incorrect spellings present in the data were rectified and that any common differences were changed to a single keyword. For example, a common answer to Q1 was some variant of "the student union" however there were a myriad of ways that students provided this response such as "the su", "the SU", "the <name of their HEI> SU", "the students union", "the students' union", "the student's union" and other combinations thereof.

Because of this, all these variants were changed to “student_union” so that they would be identified as being all of one group by the statistical software. The underscore symbol was used to identify that this is a keyword which tends to be separated by a space, but which should be evaluated together in the analysis as being different from the individual words “student” and “union”. This is discussed further below as a number of items of interest in the analysis are multiword phrases which have been compounded to ensure that they are correctly identified as being distinct from their constitutive components and are not analysed separately.

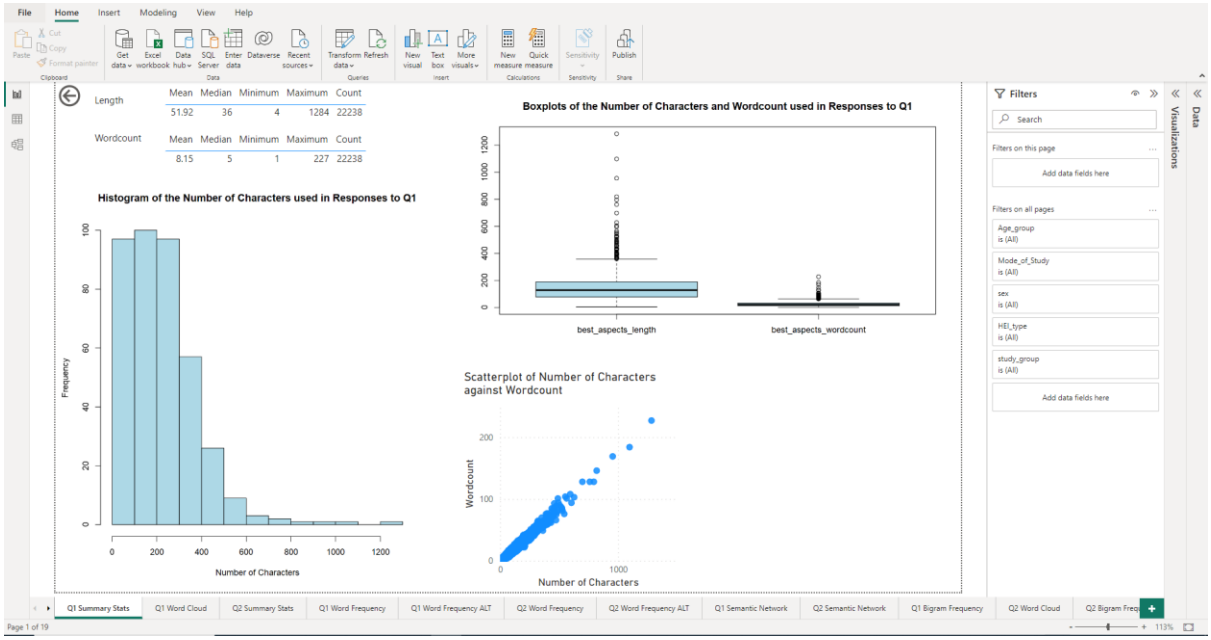
The vast majority of responses to Q1 were coherent answers which were left untouched by the cleaning process. A reasonable proportion were corrected for spelling errors as noted above. However, a small proportion of the text for Q1 was unusable for a variety of reasons. Firstly, the longer the text provided by a respondent the easier it is to gain some comprehension about the intent of the student even if their original text was corrupted in some way (with a spelling error for example). However, under a certain length it was impossible to ascertain what the respondent was presumably trying to say. In this analysis, comments of one or two characters fell into this category. Secondly, a similar problem was evident with comments like “Yes” or “No”, as the intent of these remarks could not be accurately determined. Thirdly, there were a number of random placeholder text strings which had been entered by students which did not actually contain any content, for example, “asdf” or “gygygygy”. Finally, some text contained expletives, and raises the question of whether these should be taken at face value or incidences of sarcasm or irony. As disentangling this and the other unusable text could not be adequately resolved, all of these cases were removed from the analysis by being coded as unusable and filtered out from the subsequent analysis.

Once the preliminary stages of cleaning the corpus of data had been completed, the data was ready to be entered into Power BI.

2.1 Meta-analysis and Frequency of Characters

To begin, the question asked in this portion of the survey was open-text and had no limit on the amount of material that respondents could provide. A result of this is that there is considerable variation in the amount of text students provided. Some provided terse answers, other miniature essays. Figure 2.1 provides the Power BI 'Q1 Summary Statistics' tab and in the top-left hand of the picture shows the number of characters used by students in their responses, that is the number of letters, spaces and punctuation marks provided in a response, and the number of words used in their responses. This top-line in this table shows that on average, the length of a response was close to 52 characters in length, though the median length was only 36 characters. As discussed above, the minimum length of a response was four characters, and the maximum length was 1,284 characters, and there were 22,238 responses to the question.

Figure 2.1: Q1 Summary Statistics Tab in Power BI



The bottom-line in this table shows that on average, the number of words used in a response was 8, though the median was only 5 words. The minimum length of a response was one, and the maximum was 227 words.

A large difference in mean and median values is typically an indication of a non-normal or skewed distribution, and the histogram on the bottom-left of the tab supports this.

From a meta-perspective, the number of characters that students use when answering tells us something about the variation in the amount of information students are willing to provide, but it tells us nothing about the information itself. To disaggregate this and begin to understand the information provided we need to be reasonably confident in the quality of the material to begin with.

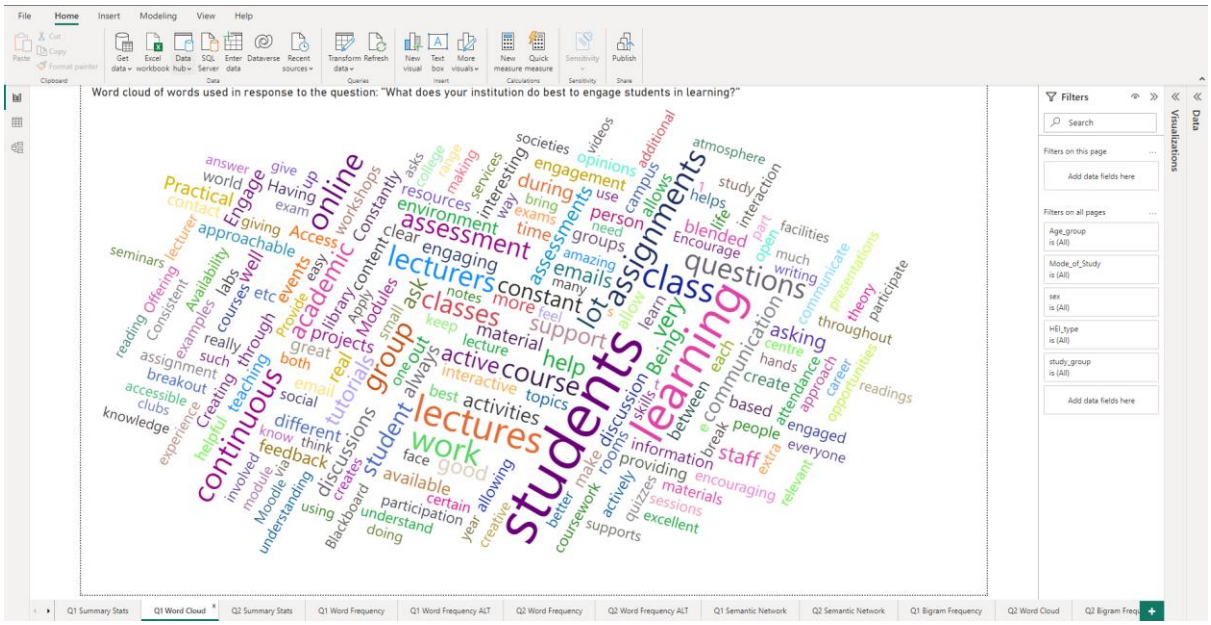
The first step of this was to remove punctuation and symbols that cluttered the corpus. The second step was to segment the corpus of data into individual tokens, usually words separated by white space. However, as noted above, through immersion in the material from the outset, we noted some words which tended to be associated with one another. To avoid these being lost to the analysis, underscores were inserted between these words, so that these multiword phrases would be compounded and seen by the software as being one unit. Examples of these phrases included phrases were: “continuous_assessment”, and “small_classes”.

The final step was to remove common ‘stop words’ that are of little intrinsic value to the analysis (such as “the”, “a”, “an”, and “in”) and whose presence would suppress words and phrases that are important to the analysis. As noted above, Power BI allows users to incorporate R programming, certain R packages, and run R scripts from within Power BI itself. The *Quanteda* R package is one supported by Power BI and has a built-in list of these common words, though this list was also supplemented further by examining the frequency of individual words and removing extraneous and unnecessary words by adding them to the stop word list. A full list of the stop words removed is provided in Appendix A.

Traditional text analysis would also stem words removing suffixes and common endings of words so that word stems could be examined together. This has not been done here as to do so would lose nuance and mean that we would be unable to distinguish between words that are very similar but have crucial differences. For example, lectures and lecturers; tutorials and tutors; practical and practicals.

At this point we are left with a corpus of data that is now ready for analysis. One initial way of examining the corpus would be to view the frequency of keywords. Figure 2.2 presents the 75 most frequently occurring words in a word cloud, with the relative size of the word being an indication of its frequency. As we can see from this graphic, “students”, “learning”, and “lecturers” are the most commonly used words in the corpus.

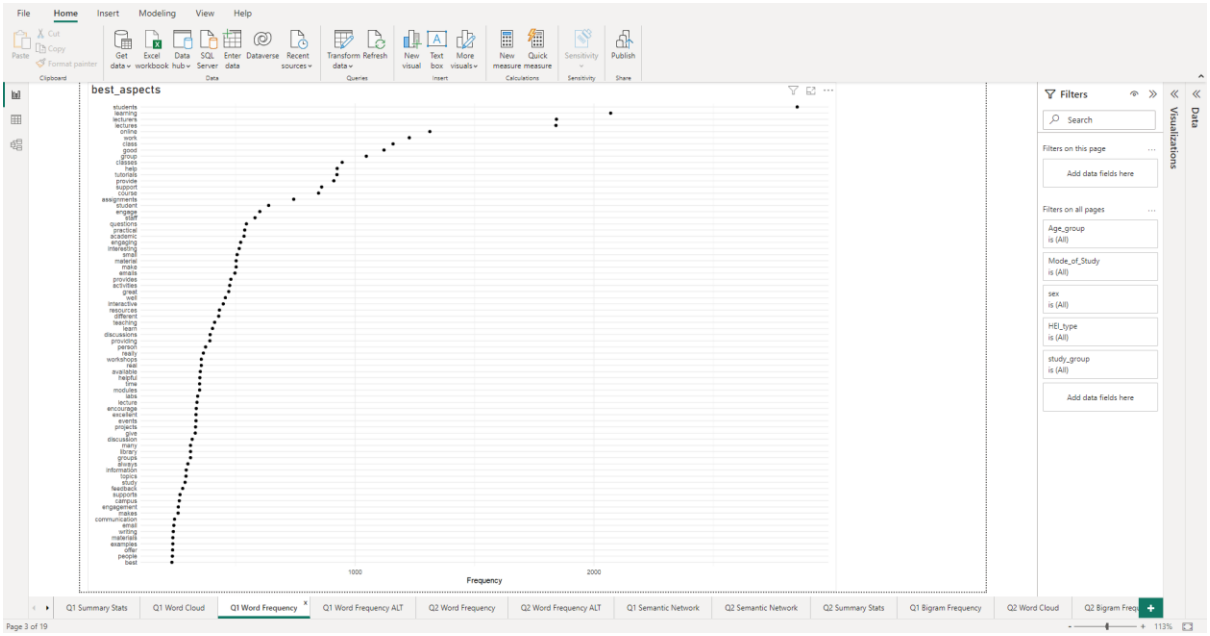
Figure 2.2: Word cloud of the 75 most frequently used words



However, by themselves word clouds are not particularly enlightening. They are a good way of viewing the corpus quickly but beyond the most frequently used words it is difficult to evaluate the relative placement of terms.

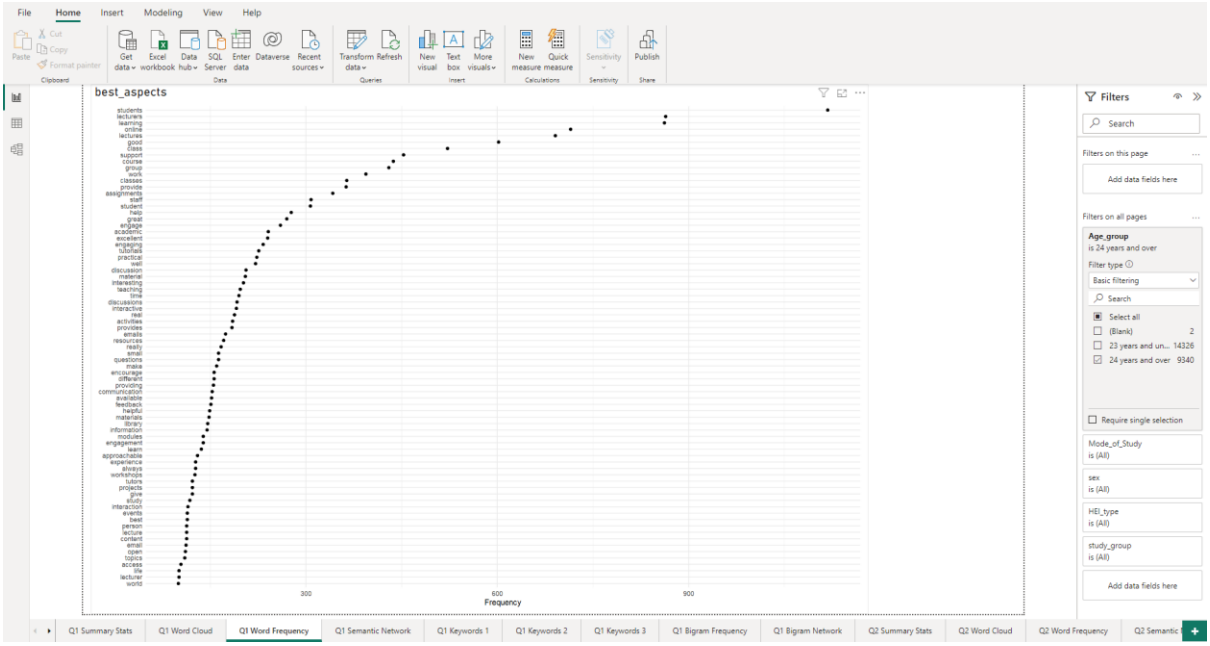
Figure 2.3 uses the same information as Figure 2.2 but plots the relative frequency of the top 75 words instead. From this we can see that “students” is the most used word with close to 3,000 instances in the corpus. After this there is a large gap between the most used and second most used with “learning” being used around 2,000 times. In third place, close to “learning” is “lecturers” with around 1,800 instances. The chart also shows that the most frequently used words are used a lot. For example, the top 9 words are each used over 1,000 times in the corpus.

Figure 2.3: Relative frequency of the top 75 most frequently used words (Q1)



It was mentioned in the previous chapter that Power Bi allows for the addition of filters to reports which makes examination of subsets and subgroups of interest extremely easy. The report at the moment has five filters: age, mode of study, sex, HEI type, and study group[3].

Figure 2.4: Relative frequency of the top 75 most frequently used words (Q1) (over 24)



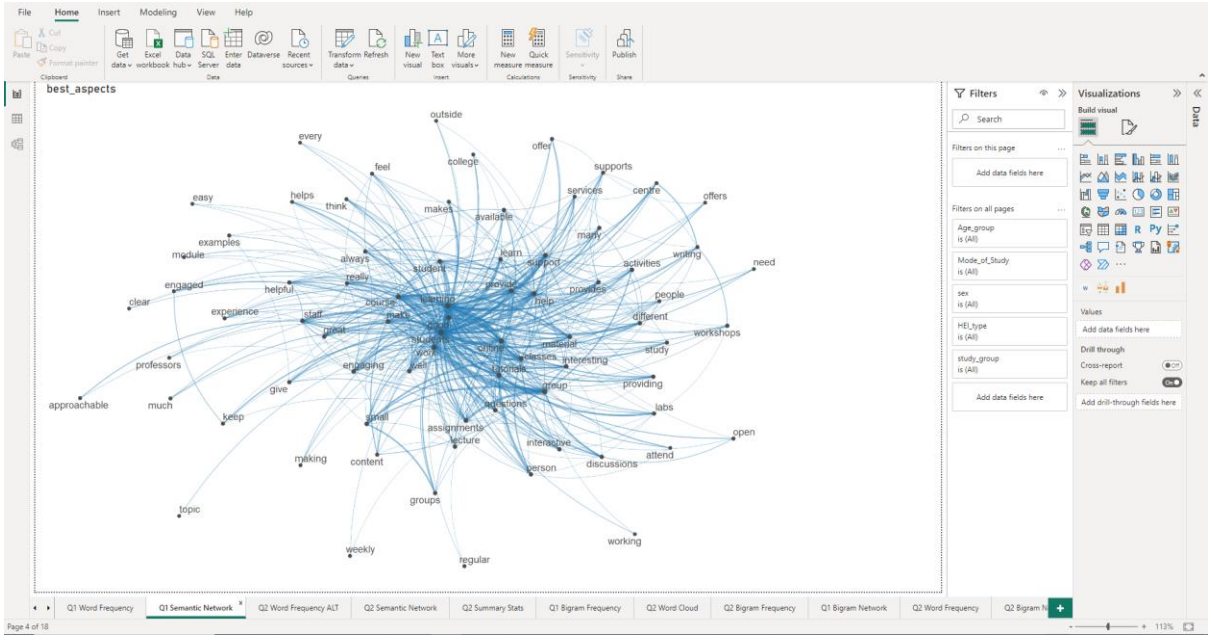
[3] Further filters can of course be added by interested researchers by dragging new ones from the data and dropping them into the filters.

The chart above shows the relative frequency of the top 75 most used words in the corpus for respondents over 24 years old. One can select and deselect across filters and the Power BI report will automatically update, thereby easily extending what in the previous report was quite static.

2.2 Semantic Networks

So far, the analysis has been of individual words, which has shown us the frequency of words used by students at an aggregate level, and some patterns with how these frequencies change when student groups are disaggregated. The next step is then to identify which words are most associated with one another. Within the statistical software this was done by creating a feature co-occurrence matrix which records the number of co-occurrences of tokens.

Figure 2.5: Semantic network of feature co-occurrence matrix



This feature co-occurrence matrix can then be visualised in a semantic network to illustrate which words are most associated with one another. The width of the bars linking words indicates the strength of the relationship between the words. Figure 2.5 presents a semantic network and as we would expect the three most frequently used words in the corpus thus

far, “students, “learning”, and “lecturers” form the central axis from which all the other words branch.

2.3 Coding Student Responses and Sentiment Analysis

The final step taken in this chapter is to extend the sentiment analysis beyond individual words and instead evaluate whole sentences. Within the survey, students were asked the question, “What does your institution do best to engage students in learning?” As we have seen already students were free to write as much, or as little, as they wanted in answering the question.

In addition, while the question asked could be seen as a direct question, students approached answering it in a myriad of ways. For example, student A could answer the question by saying:

> “I think that the lectures and tutorials provided my institution are the best for engaging students.”

Alternatively, student B could say:

> “In the institution the lecturers know my name, and this makes me feel I am more than just a number. I think this helps me engage in learning”.

Both are equally valid responses but demonstrate different approaches to answering the question. The first providing a more tangible institutional perspective, the second a more personal and experiential answer.

While conducting the preliminary review of students’ comments we noted themes and concurrent keywords associated with these themes which were repeated throughout the corpus. This provided us with an initial coding framework. This framework was then supplemented by the analysis reported so far. The coding framework provided in Table 2.1 categorises how students approach answering the question and it is purposely designed along a continuum, with one end being the institutional, and the other personal while simultaneously capturing general versus specific sentiments.

Table 2.1: Coding framework applied to the Q1 corpus

Institutional	
Theme	**Associated Keywords**
On-campus, direct learning/teaching	Tutorials; Workshops; Lectures; Seminars; Labs; Small class/es; Guest speakers; Lecturers; Academic staff; Attendance
On-campus, outside of direct learning/teaching	Academic Learning Centre; Writing centre; Peer-based Learning *(including CÉIM)*; Library; Students' Union; Clubs and Societies; Sport; Facilities
Off-campus/Remote Learning/Online Teaching Environment/ Technology	SULIS/Moodle/Blackboard/Panopto/Brightspace -> *All coded together as Online Teaching Environments*; Clicker; Email; Quizzes
Coursework and Exams	Feedback; Essays; Exams; Dissertation/Thesis; Ask questions; Continuous Assessment
Personal and Experiential	
Theme	**Associated Keywords**
Experience	Collaborative; Discussions; Work placement; Teaching placement; Work experience; Practical elements; Working together; Active participation; Interactive/Interaction; Communication; Hands-on; Help/Support/Assistance; Real life/Real world examples; Problem based learning; Case studies
Personal	Respect; Encourage/Engage; Treated like an adult/equal; "Know my/your name"; Fun; Interesting; Friendly; Approachable; Enjoyable

Within the software, we designed an algorithm that searches for these each of these keywords and iterations thereof, and records where they occur. Of the almost 23,000 comments recorded by students there are over 17,000 cases where at least one of the keywords are mentioned. The analysis presented in the remainder of this section is based on this subset of students.

As has been mentioned already, examining the frequency of themes and how often keywords are mentioned could provide us with a false impression of the sentiment associated with the either. The question asked of students is 'What does your institution do best to engage students in learning?', and we may look at the number of students who say 'lectures' in their responses as a proxy for what these students think their institutions do best. However, we must be careful here not to divorce keywords of interest from contextual information which provides nuance to the response.

For example, student C may have provided the following response:

> "I love tutorials. They give me a chance to share my reading and discuss with my classmates. They are the complete opposite to my experience with lectures. I find that being handed information in a lecture is a really dry and boring approach to education."

For this student, tutorials are evaluated positively and lectures negatively, but from a naïve frequency perspective, without any further information, as lectures are mentioned twice and tutorials only once, we may incorrectly assume that lectures are viewed more positively because they are mentioned more than tutorials.

While this is easy to distinguish from a single comment, when we are faced with a corpus of the magnitude provided by this question it becomes very appealing to use a frequency heuristic as is it time and resource efficient. We should also be cautious in handling the data not to go the other way and view everything that a student writes with scepticism because at some point we have to view the factors mentioned by students as subjectively important to them, otherwise they would not have mentioned them. As such, the analysis conducted in this section balances these perspectives through combining a frequency-based approach in tandem with sentiment analysis.

Sentiments can be broadly classified as positive, neutral or negative. They can also be represented on a numeric scale, to better express the degree to which a body of text contains positive or negative sentiment. The lexicons utilised here evaluate the words present in a students' comment and attach a score to each comment based on the positive, negative and neutral content present in the comment.

The analysis below uses the R package, Syuzhet within Power BI for generating sentiment scores, and within this package we have three dictionaries available to us called Afinn, Bing and Syuzhet[4].

The correlation coefficients between each dictionary were very high and statistically significant, and they all appear to perform in a similar manner. As such, we have presented the Syuzhet results purely because its scores are continuous thus have greater potential variation than scores with discrete positions on an axis (for example 0.0, 0.5, 1.0, 1.5, and so on).

At this point, we ran our initial algorithm which recorded the presence or absence of each of our keywords as a set of indicator variables, and passed the corpus of students' comments through a sentiment dictionary which gave each comment a sentiment score.

The next step was parsing the keywords with the sentiment analysis and calculating the general sentiment score associated with each keyword. These are presented in the tables below as a set of summary statistics, with the mean being the one most discussed as this is the best indication of the general sentiment associated with the keyword.

The scale of the sentiment scores is simple to interpret in that higher scores indicate higher levels of positive sentiment, and lower scores indicate higher levels of negative sentiment. This scale also allows for the comparison of scores across keywords, so it is evident below which keywords have significantly higher sentiment scores and lower sentiment scores than others.

[4] Jockers, M (2020). Syuzhet Package Documentation.
Available from: https://cran.rstudio.com/web/packages/syuzhet/syuzhet.pdf

While it is important to see how students are evaluating various aspects of their courses, we can also test if different student groups evaluate their courses in the same way. While not presented here, the Power BI report allows users to filter across student groups and allow us to see if, for example, first year undergraduates have a more positive sentiment score for lectures than taught undergraduates.

It is worth highlighting that this approach is not without its imperfections, but there is an inherent trade-off between human coding, time, and error. In general, from examining the scores associated with each sentence in the corpus, the sentiment dictionaries do a good job of capturing the sentiment contained within comments even if they lack the ability to parse nuance, irony and sarcasm. It is more than this methodology excels at evaluating the underlying sentiment when the wealth of material is too much for a human coder and would fall prey to biases in human coders.

The greatest strength of computer-assisted content analysis is that all material is treated equally. In contrast for a human coder there is just too much material. This amount of effort and concentration would bring with it the risk of error due to fatigue, recency bias, and so on. Whereas our software evaluates the corpus in seconds.

Figure 2.6: Summary statistics for 'on-campus, direct learning/teaching' and 'on-campus, outside of direct learning/teaching' keywords

On-campus, direct learning/teaching keywords

Academic Staff	Mean	Median	SD	Minimum	Maximum	Count
>	0.92	0.70	1.16	-0.50	6.30	168

Attendance	Mean	Median	SD	Minimum	Maximum	Count
>	0.56	0.25	0.80	-0.75	3.70	180

Guest Speakers	Mean	Median	SD	Minimum	Maximum	Count
>	0.66	0.38	0.85	-0.60	3.30	82

Labs	Mean	Median	SD	Minimum	Maximum	Count
>	0.64	0.25	0.93	-1.80	6.75	1062

Lecturer(s)	Mean	Median	SD	Minimum	Maximum	Count
>	0.64	0.40	0.87	-3.75	6.75	2139

Lectures	Mean	Median	SD	Minimum	Maximum	Count
>	0.70	0.50	0.95	-2.50	9.10	2220

Seminars	Mean	Median	SD	Minimum	Maximum	Count
>	0.65	0.33	0.87	-0.60	4.80	218

Small Classes	Mean	Median	SD	Minimum	Maximum	Count
>	0.63	0.25	1.01	-0.60	9.80	295

Tutorials	Mean	Median	SD	Minimum	Maximum	Count
>	0.68	0.40	0.91	-1.50	6.15	1517

Workshops	Mean	Median	SD	Minimum	Maximum	Count
>	0.66	0.40	0.89	-0.75	5.85	444

On-campus, outside of direct learning/teaching keywords

Academic Learning Centre	Mean	Median	SD	Minimum	Maximum	Count
>	0.69	0.40	0.78	0.00	2.25	16

Clubs and Societies	Mean	Median	SD	Minimum	Maximum	Count
>	0.61	0.25	0.93	-1.75	5.90	215

Facilities	Mean	Median	SD	Minimum	Maximum	Count
>	0.75	0.50	1.00	-0.50	6.75	156

Library	Mean	Median	SD	Minimum	Maximum	Count
>	0.79	0.50	0.99	-0.75	7.35	350

Peer Based Learning	Mean	Median	SD	Minimum	Maximum	Count
>	0.62	0.25	0.87	0.00	3.30	31

Sport	Mean	Median	SD	Minimum	Maximum	Count
>	0.65	0.75	0.70	0.00	2.10	21

Student Union	Mean	Median	SD	Minimum	Maximum	Count
>	0.71	0.60	0.78	-0.10	3.05	130

Writing Centre	Mean	Median	SD	Minimum	Maximum	Count
>	0.63	0.00	0.92	-0.85	3.55	55

To move onto the sentiment analysis itself. The frequency of keywords associated with on-campus and having direct interaction with teaching and learning, and on-campus but outside of direct/teaching and learning keywords are presented in Figure 2.6. From the left-hand side of this chart, one can see that lectures are most often mentioned in students' comments, closely followed by lectures, and then tutorials. The other keywords are mentioned much less often.

The sentiment associated with each keyword does not follow the pattern demonstrated by the frequency of keywords, illustrating the point made above that by itself, frequency is not a good indicator of what students think helps them engage in learning at their institution. Academic staff have the highest mean sentiment score in comments where the keyword is used at 0.92.

While being present in more students' comments, lectures, lecturers, and tutorials have mean sentiment scores of 0.70, 0.64, and 0.68 respectively. This indicates that comments that contain these keywords are a combination of positive and negative sentiments. The keyword attendance is typically combined with 'compulsory' in the comments and does not appear to be evaluated positively by students with a mean score of 0.56.

The right-hand side of Figure 2.6 presents the average sentiment scores for 'on-campus, outside of direct teaching/learning' keywords. The library is the most often mentioned keyword here, and has a relatively high average sentiment score of 0.79

Figure 2.7: Summary statistics for 'off-campus/remote learning/online teaching environment/technology/coursework and exams' keywords

Off-campus/remote learning/online teaching environment/technology keywords

Email	Mean	Median	SD	Minimum	Maximum	Count
>	0.71	0.40	0.97	-1.50	5.65	980

Online Learning Enviro	Mean	Median	SD	Minimum	Maximum	Count
>	0.69	0.50	0.86	-0.55	5.30	368

Quizzes	Mean	Median	SD	Minimum	Maximum	Count
>	0.70	0.50	0.88	-0.70	4.35	210

Coursework and exam Keywords

Continuous Assessment	Mean	Median	SD	Minimum	Maximum	Count
>	0.70	0.40	1.00	-1.00	7.80	391

Essays	Mean	Median	SD	Minimum	Maximum	Count
>	0.59	0.25	0.83	0.00	3.45	49

Exams	Mean	Median	SD	Minimum	Maximum	Count
>	0.67	0.30	0.93	-1.00	4.00	110

Feedback	Mean	Median	SD	Minimum	Maximum	Count
>	0.66	0.40	0.87	-0.75	4.55	309

Keywords associated with the technology that assisted them to work off-campus (online learning environments/email) or interacted with on-campus (clicker or quizzes in lectures) are presented on the left-side of Figure 2.7. From this chart, one can see email is most often mentioned in students' comments and has the highest average sentiment score within this theme of 0.71.

Keywords associated with coursework and exams are presented on the right-side of Figure 2.7. Within this theme, continuous assessment is most often mentioned in students' comments, closely followed by feedback.

Figure 2.1: Summary statistics for 'experience' and 'personal' keywords

Experience keywords

Active Participation	Mean	Median	SD	Minimum	Maximum	Count
0	0.56	0.25	0.80	-0.75	3.70	180

Communication	Mean	Median	SD	Minimum	Maximum	Count
>	0.73	0.50	1.01	-1.00	7.50	337

Discussions	Mean	Median	SD	Minimum	Maximum	Count
>	0.70	0.40	0.94	-1.85	7.75	778

Hands On	Mean	Median	SD	Minimum	Maximum	Count
>	0.63	0.40	0.91	-0.50	4.65	170

Interactive	Mean	Median	SD	Minimum	Maximum	Count
>	0.68	0.50	0.89	-1.85	5.90	851

Real World Examples	Mean	Median	SD	Minimum	Maximum	Count
>	0.70	0.60	0.79	-0.75	4.15	297

Personal Keywords

Approachable	Mean	Median	SD	Minimum	Maximum	Count
>	0.77	0.50	1.03	-1.25	8.90	556

Encouraging	Mean	Median	SD	Minimum	Maximum	Count
>	0.68	0.40	0.91	-1.85	6.75	652

Engaging	Mean	Median	SD	Minimum	Maximum	Count
>	0.68	0.40	0.93	-1.25	6.75	1051

Enjoyable	Mean	Median	SD	Minimum	Maximum	Count
>	0.76	0.60	1.07	-0.80	6.30	57

Friendly	Mean	Median	SD	Minimum	Maximum	Count
>	0.64	0.38	0.93	-0.80	6.75	206

Fun	Mean	Median	SD	Minimum	Maximum	Count
>	0.66	0.50	0.80	-1.75	4.65	175

Interesting	Mean	Median	SD	Minimum	Maximum	Count
>	0.77	0.50	1.03	-1.25	8.90	556

Keywords associated with experiential factors and personal keywords are presented in Figure 2.8. Discussions are the most often used keyword within the experience set of keywords, though communication is the most positively evaluated with an average sentiment score of 0.73. On the right-side of the chart, engaging is the most often used personal keyword, and interesting is the most positively evaluated with an average sentiment score of 0.77.

With regard to the average sentiment scores for these keywords, a case can be made to err on the side of caution here, because of the connotations present in each word. It is highly likely that the average sentiment scores for each of these keywords are higher than they should be because the words themselves have positive associations. As a result, their presence in a sentence biases the overall score towards being more positive than it would be if the keyword was not present.

This of course is not to say that these factors are not important to students, as has already been noted, the fact that these keywords have been mentioned (outside of 'engage' which they are primed to use due to its presence in the original question) can be taken as some measure of the importance of these for students, regardless of whether their institution actually meets them.

2.4 Conclusions

This chapter has broken down the analysis undertaken to address the question posed to students; "What does your institution do best to engage students in learning?" The analysis conducted has demonstrated that students answered this question in a myriad of ways, but from this almost overwhelming amount of material certain patterns emerged. From breaking down students' comments into keywords, and then examining the general sentiment present in sentences that contained these keywords, we can provide some answers to the question asked of students.

Table 2.2: Ranking of keywords with highest average sentiment scores.

Rank	Keyword	Mean Sentiment Score
1	Academic staff	0.92
2	Library	0.79
3	Facilities	0.75
4	Communication	0.73
=5	Student Union	0.71
=5	Email	0.71
=6	Lectures	0.70
=6	Continuous Assessment	0.70
=6	Discussions	0.70
=6	Real World Examples	0.70

In general, it appears that students think that the academic staff of their institution are the best for engaging students in learning. This is followed by the library services, and general facilities provided by their HEI.

Keywords associated with 'personal' experiences have been left from the table due to the fact that their mean scores may be positively biased due to the positive associations inherent to the keywords themselves. This, however, does not mean that HEIs should ignore these facets of students' experiences. Instead, if greater attention were paid to these facets by HEIs, academic staff and other points of contact between academic institutions and students, one could see a virtuous cycle which reinforces other aspects of students experiences and increases student engagement.

The next chapter moves onto addressing how students evaluated the second qualitative question they were asked; "What could your institution do to improve students' engagement in learning?"

3. “What could your institution do to improve students' engagement in learning?”

This chapter of the report covers the analysis undertaken to provide us with some answers to the second open-text question posed to students; “What could your institution do to improve students’ engagement in learning?” Which for the shorthand purposes in the rest of this chapter will be referred to as Q2.

In contrast to Q1, which was worded in a manner to evoke a precise response, such as an item, or institutional provision that students regarded as being successful in engaging students, Q2 poses a more hypothetical query, which asks about a change that could be made, and asks students to weigh up what they have received from their HEI and from this, evaluate what change could be made to improve students’ engagement in learning.

The structure of the question is similar to an equation which asks, what change needs to be made to X to see an improvement/increase in Y where Y is students’ engagement in learning. As a result of the way the question is asked, responses tend to point to an item and provide a description of degree, or provide a modifier/intensifier, for example, typical responses to Q2 included statements such as ‘’provide more continuous assessment”, “better lecturers”, “fewer exams”, “longer library opening hours” and so on.

The analysis that this chapter discusses follows a similar format to that used in Q1 but with a few significant deviations to capture this fact that standard responses from students provide **items** that students think should be altered to improve student engagement and some **measure of direction and degree** (more/less/better/fewer etc).

In addition, the sentiment analysis used in the previous chapter is of less utility here because of the prevalence of a small group of words used to indicate direction and degree. This is discussed further below. As such, there is less need to capture what students are feeling about keywords through the words used in their responses because there is:

a. Much less variety in the words used to describe what they would change about these keywords.
b. The words used to describe direction and degree specify through their usage how they would affect items, thus rendering it unnecessary to further examine the sentiment that these words contain.

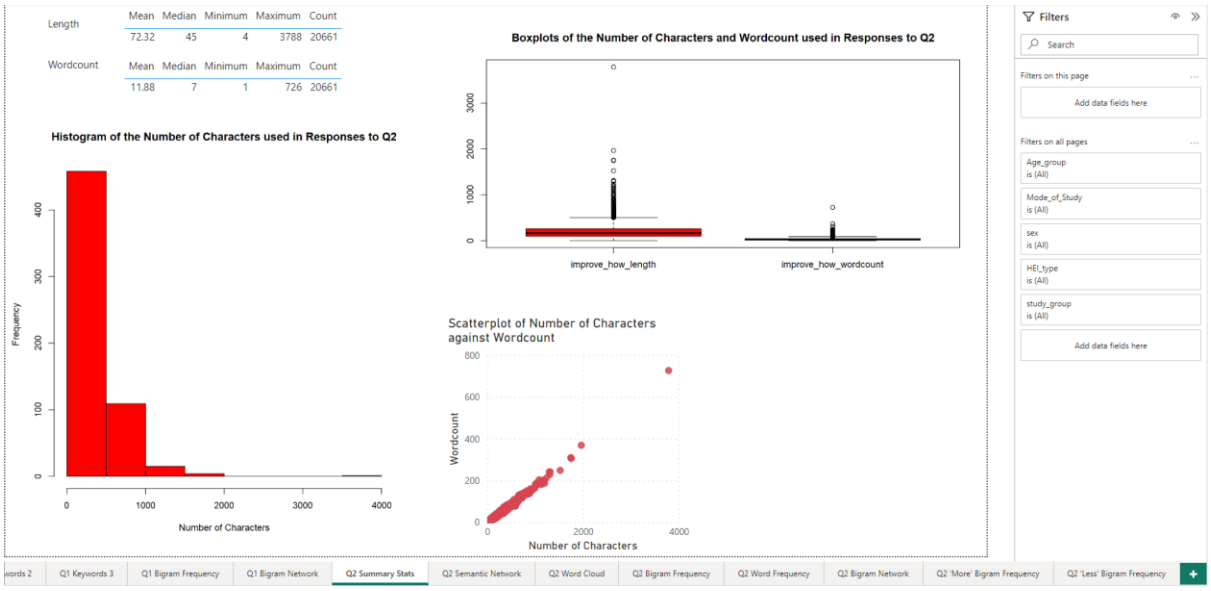
However, the first task facing the researcher is parsing the material into a comprehensible format. This is again, a daunting task due to the volume of material students have provided. The next section covers the preliminary meta-analysis of Q2 and describes how the corpus was cleaned for further analysis.

3.1 Meta-analysis and Frequency of Characters

As with Q1, responses to Q2 were open-text and free of any character limits. As such, respondents could provide as much or as little text as they wanted. This freedom has resulted in considerable variation in the amount of text students provided. Figure 3.1 provides some summary statistics of the number of characters used by students in their responses. Characters include letters, numbers, punctation marks and spaces. This table shows that on average, the length of a response just over 72 characters long. The median length was 45 characters. The minimum length was 4 characters, and the maximum was 3,788. In addition, 20,661 students provided a response to Q2.

Much like that seen in Figure 2.1 in the previous chapter, the distance between the mean and the median again indicate a non-normal distribution of characters. The skewed distribution of characters is again, not unexpected, and would only be a concern if different groups of students had different patterns of responding to the question. To ensure that this is not the case, one can use the filters on the right-hand side of the Power BI report to examine differences across groups.

Figure 3.1: Q2 Summary Statistics Tab in Power BI



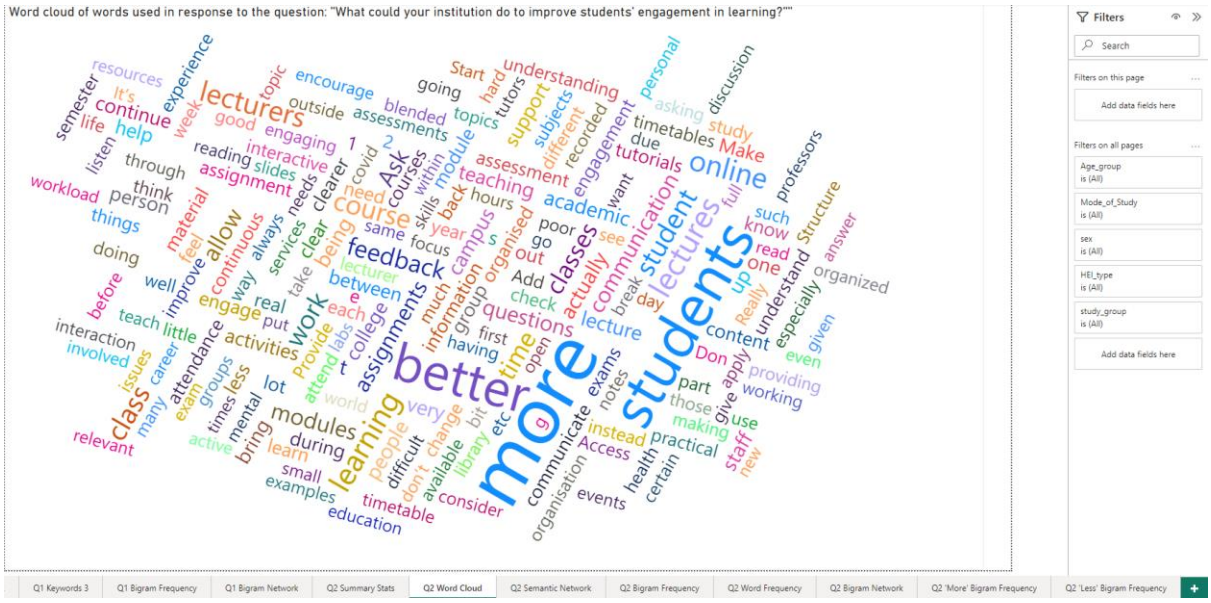
As mentioned in Chapter 2, the number of characters students use to provide a response tells us something about the willingness of students to provide information, but does not tell us anything about the information provided. To gain an understanding of the material provided, we must be confident of the quality of the material to begin with, and to ensure this, the data goes through a series of steps to remove any extraneous content which would undermine the analysis. These steps take the corpus of material from a 'raw' to 'clean' state, though throughout these steps a copy of the original data is kept as a reference and quality control check.

The first invaluable check is gaining a familiarity with the data through reading through the material provided. This is time-intensive but provides a basis from which to build an understanding of how students have approached the question asked of them. Because of the volume of material provided by students it remains impractical to read everything but by combining the immersion in the text with the spellcheck component it is still possible to get a solid grounding in recurring topics, keywords and approaches. Comments below a certain length (three characters and fewer) were removed, and the same compounding of common multiphrase words (continuous assessment, academic staff, and so on) which should be examined by the statistical software together was also conducted[5].

[5] Chapter Two contains a more thorough discussion of the steps taken to get raw data into its cleaned state.

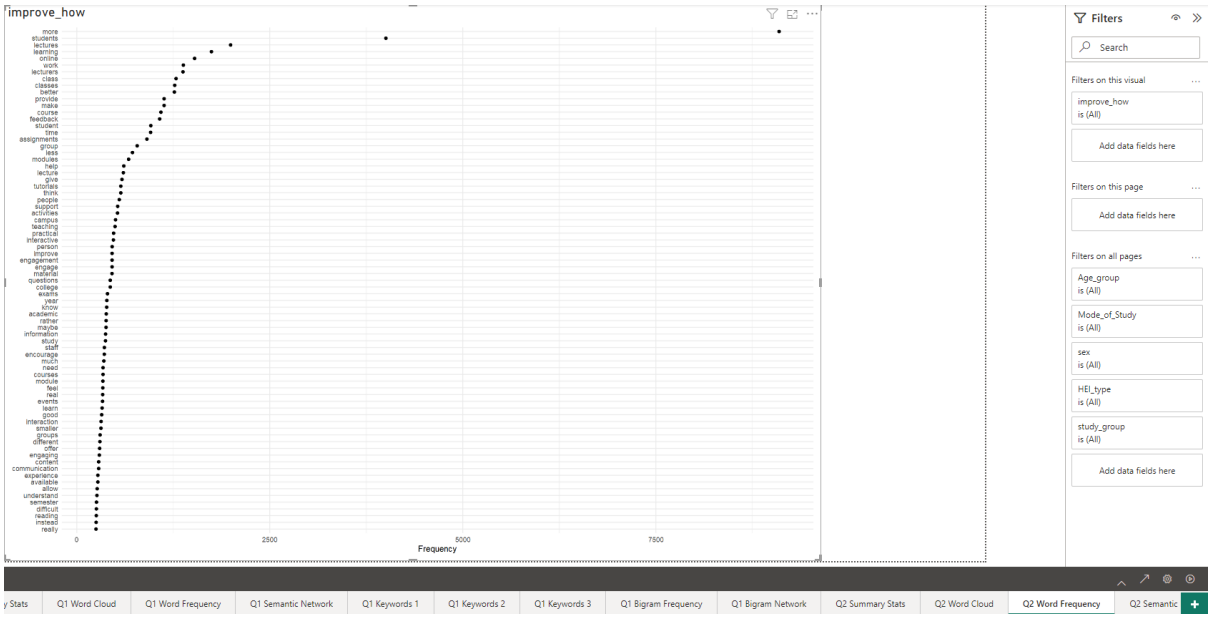
Once these steps have been completed the data is ready for entry into the software, so that punctuation and symbols can be removed, along with common stop words. However, for this portion of the project the word ‘more’ was added back into the list of allowed words rather than being removed, as this word was invaluable in the analysis below, as it is the most common way students had of articulating the changes they wished to see in their HEIs to improve student engagement.

Figure 3.2: Word cloud of the 75 most frequently used words



The remaining corpus was then ‘tokenised’ based on the whitespace between words so the analysis could begin in earnest. Figure 3.2 presents the 75 most frequently occurring keywords in the cleaned corpus in a word cloud, with the relative size of a word being an indication of its frequency. As can be seen by its relative size, “more” is the most frequently occurring word in the corpus used by students in responding to Q2.

Figure 3.3: Relative frequency of the top 75 most frequently used words (Q2)



As has been noted before, word clouds are good at providing an initial indication of the frequency of words used in the corpus but not much beyond this. As such, Figure 3.3 plots the relative frequency of words instead, and reiterates the fact that the word 'more' is by far and away the most popular word in the Q2 corpus.

3.2 Bigrams

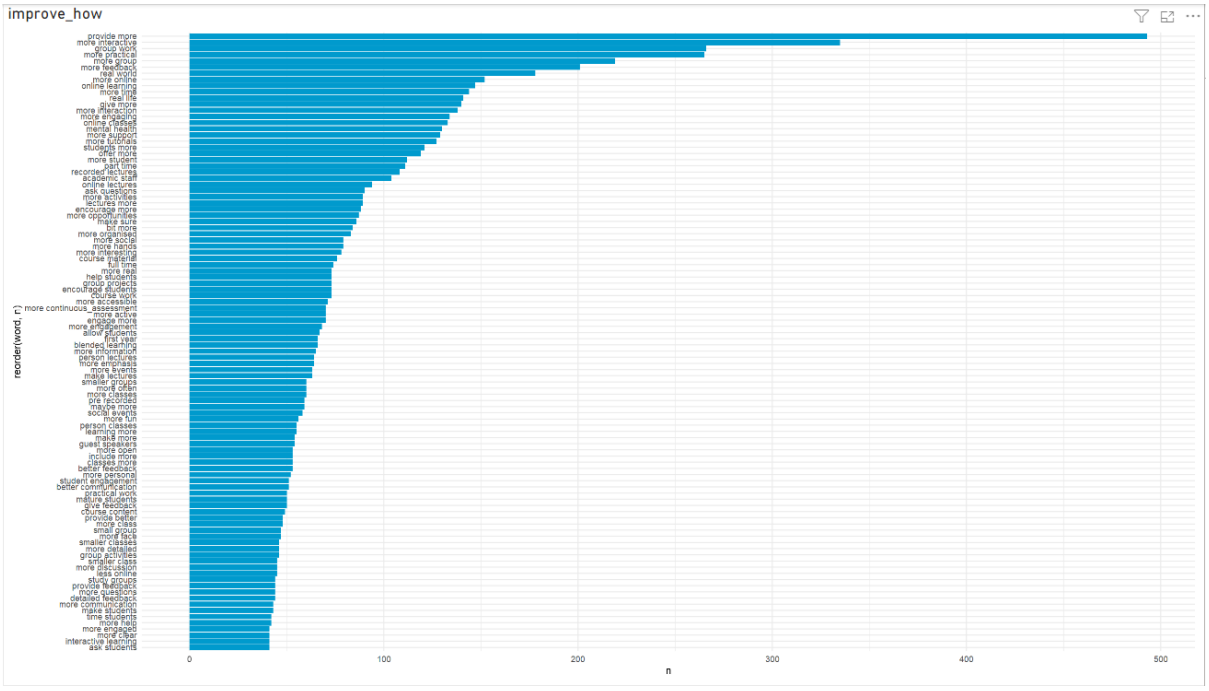
At this point in the analysis of Q1 we utilised sentiment analysis of words in order to gain an understanding of whether words used by students had positive or negative connotations. This is of less use here where we would expect answers to contain both what students think can be improved by listing an item or items, along with how it can be improved. As a result of this, the positive and negative connotations associated with a modifier is of less utility as the usage of the word encapsulates its definition, more means more, less means less, and so on. Furthermore, the sentiment contained in the comments provided to Q2 are of less use to us in *providing answers* to Q2.

Instead, because students tended to follow a very similar format in answering Q2 through mentioning some form of direction and degree along with an item, it is possible to move beyond individual words in the corpus and tokenise consecutive sequences of words within each comment provided by students, and then examine these sequences as they tend to

capture the format with which students provide answers to Q2 very well. These sequences are called n-grams and the ones with greatest use here, are bigrams which are sequences of two words. So by seeing how often word X if followed by word Y, we can then begin to build a model of the relationships between the words, and obtain a picture of the answers students are providing for Q2.

Figure 3.4 presents the most frequently occurring bigrams in the corpus. Some bigrams are completely descriptive such as 'part time', 'mature students' and 'work experience'. However, what is striking is how often modifiers are used within the corpus. Throughout this list 'more' is often the first or second word in the bigram. Other modifiers include 'smaller', 'better' 'less'. These modifiers are sometimes linked with verbs, for example, the most frequently used bigram is 'provide more', but more often the modifier is linked to an item, as can be seen in the second most frequently used bigram is 'more interactive'.

Figure 3.4: Relative frequency of most frequently occurring bigrams



As such, this simple construction of modifier plus item appears to be a simple and precise way of capturing students' answers to Q2. To demonstrate this further, Figures 3.5 to 3.15 present the frequency of bigrams with the common modifiers 'more', 'less', 'better', 'fewer',

‘improve’ and ‘reduce’ as the first word in the pair. There are a few points to note in these charts. Firstly, due to how Q2 is worded, there appears to be a natural inclination to providing a positive response, as a result, positive modifiers such as ‘more’, ‘better’, and ‘improve’ are more commonly used than the negative modifiers such as ‘less’, ‘fewer’, and ‘reduce’ as can be seen in the number of bigrams along the x-axis. Secondly, note how often the bigrams concisely summarise students’ main points in their responses. Though it is also worth noting that some bigrams remain vague because the bigrams do not end in an item, for example, ‘more practical’ in Figure 3.10 or ‘less emphasis’ in Figure 3.11. These are discussed further in the next section as bigrams are combined with Markov chains.

Figure 3.5: Relative frequency of bigrams containing ‘more’ as first word

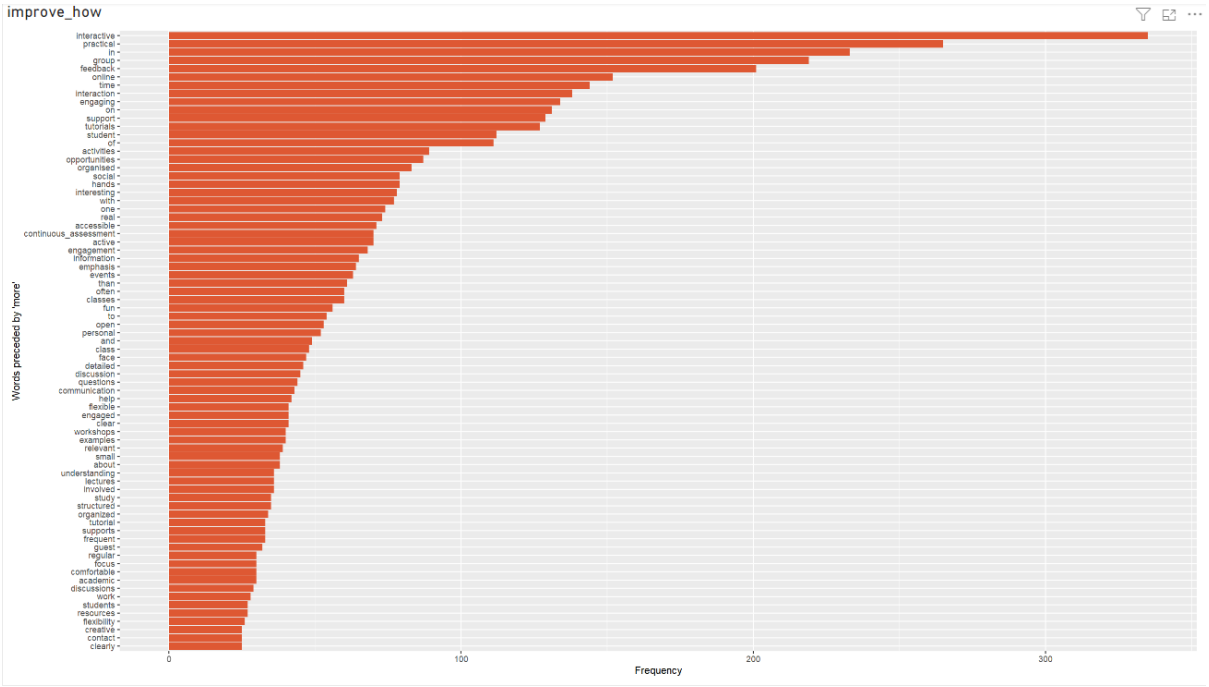


Figure 3.6: Relative frequency of bigrams containing ‘less’ as first word

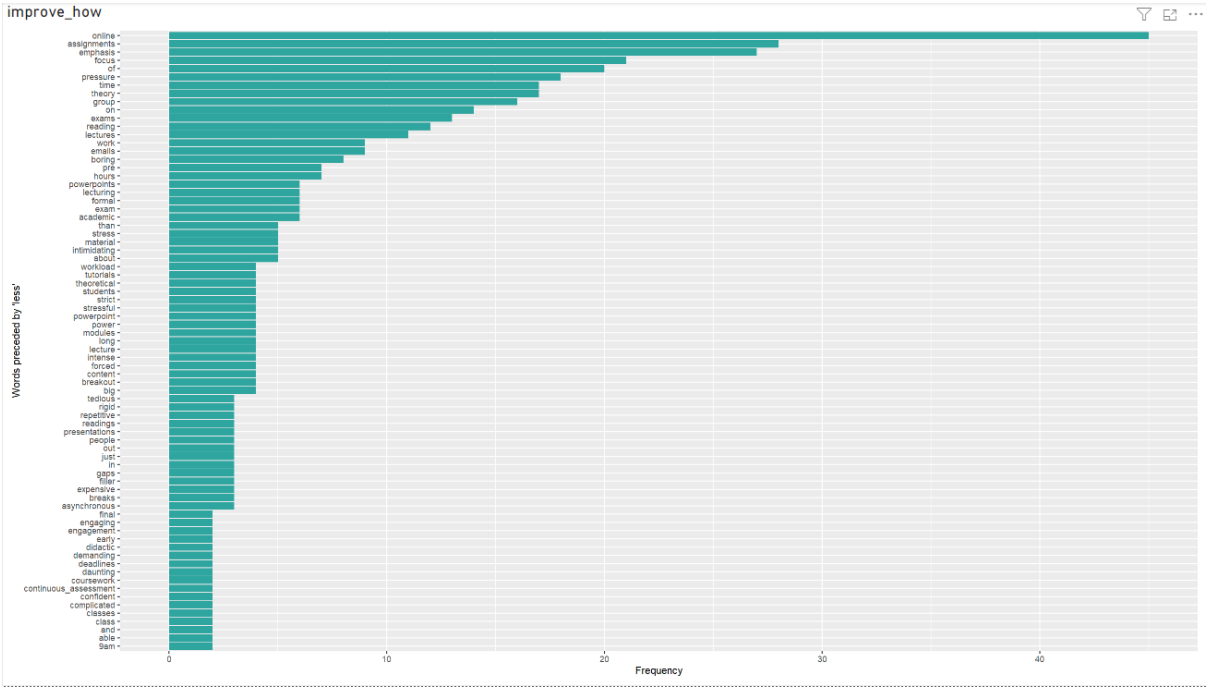


Figure 3.7: Relative frequency of bigrams containing ‘better’ as first word

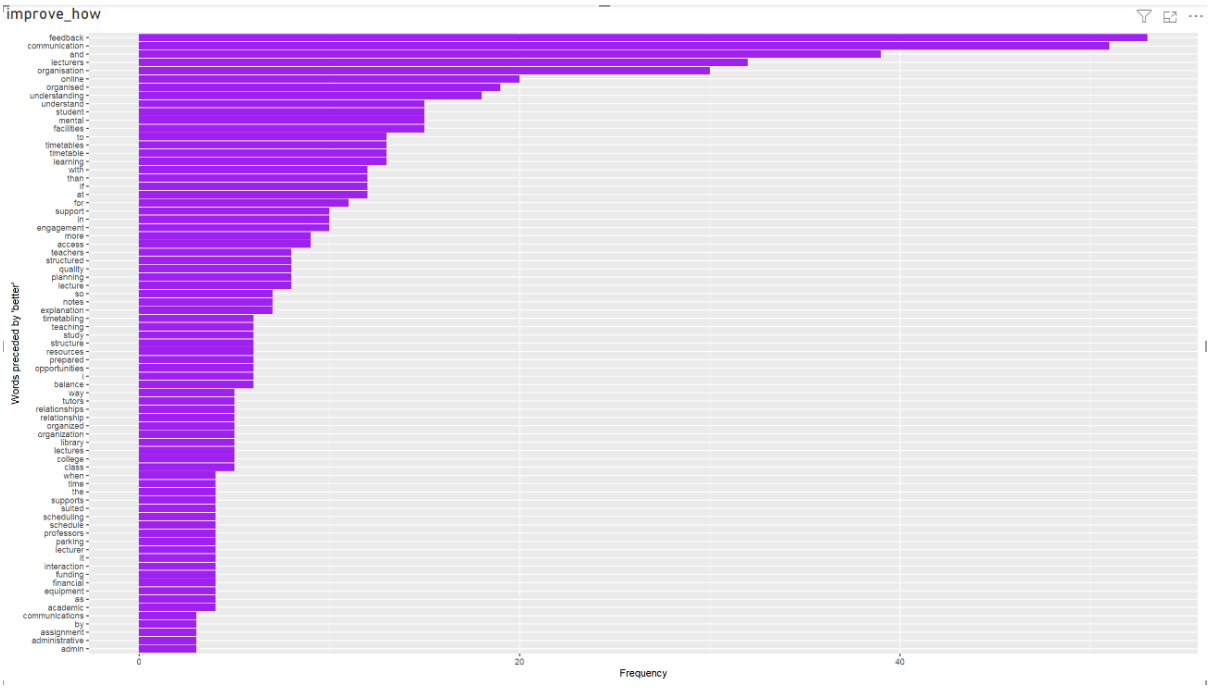


Figure 3.1: Relative frequency of bigrams containing ‘fewer’ as first word

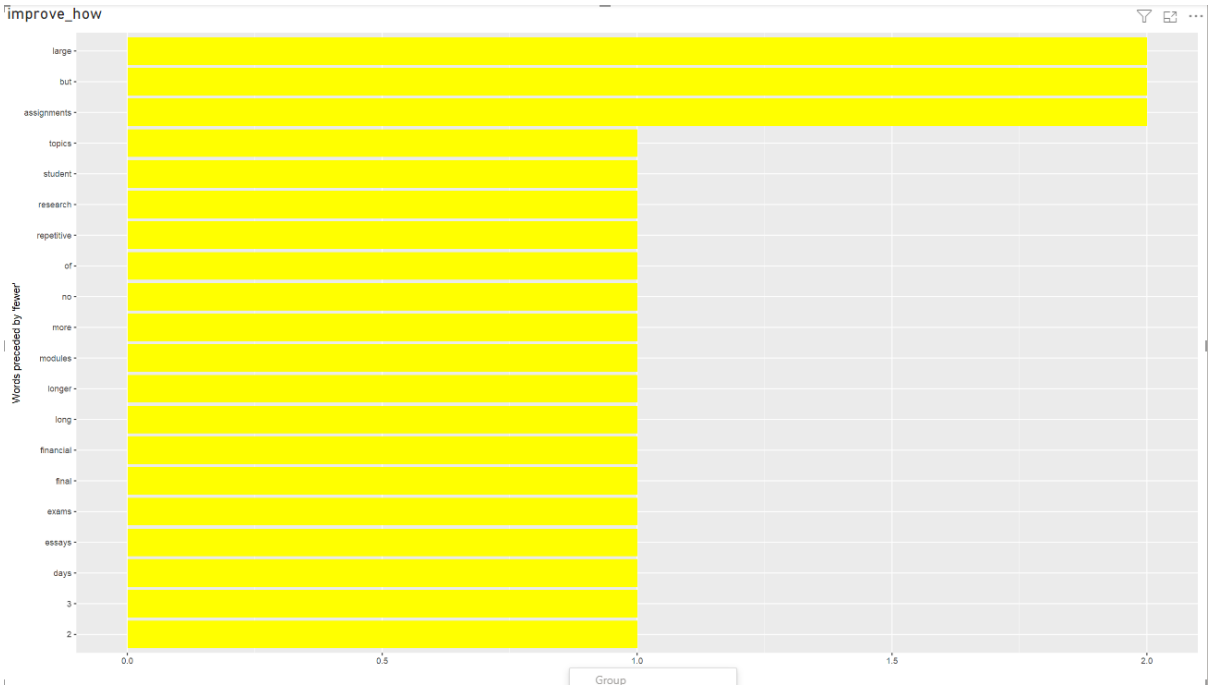


Figure 3.9: Relative frequency of bigrams containing ‘improve’ as first word

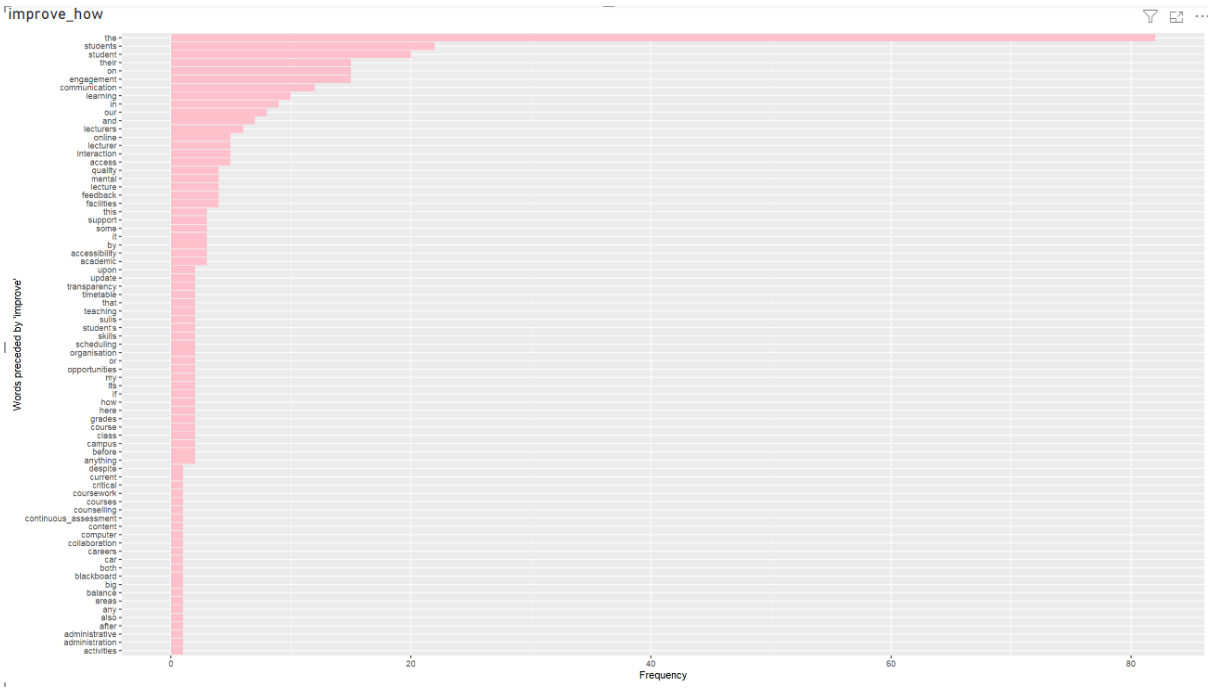
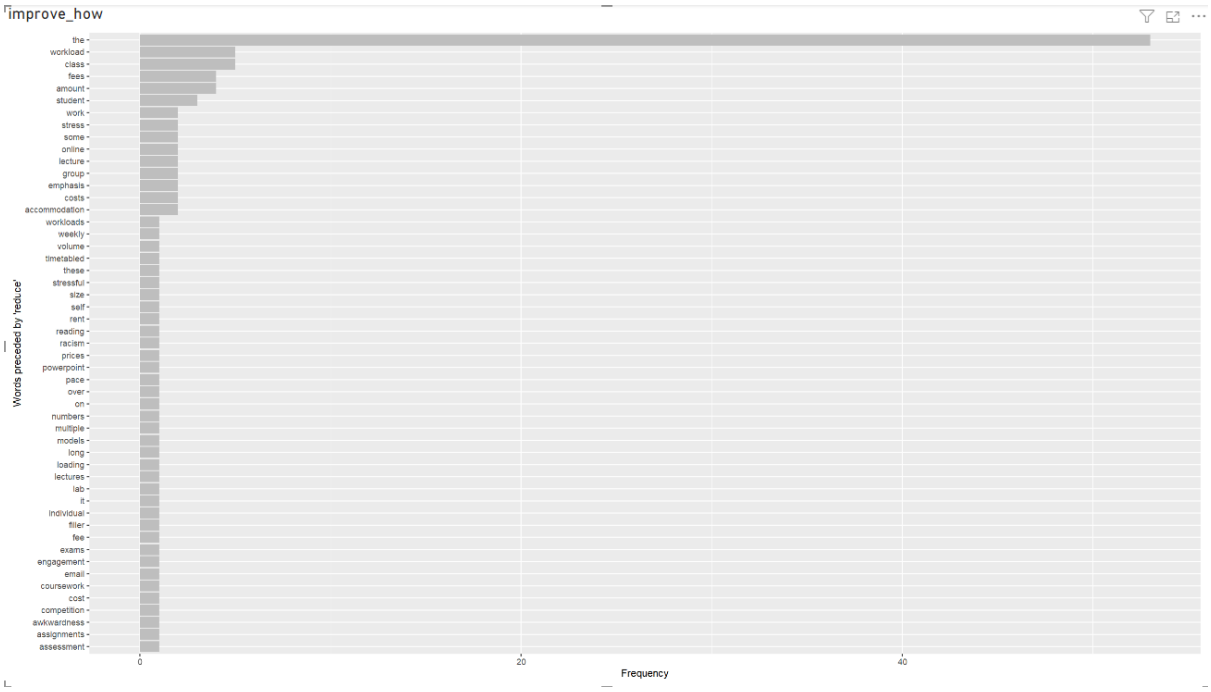


Figure 3.2: Relative frequency of bigrams containing ‘reduce’ as first word



3.3 Bigram Networks

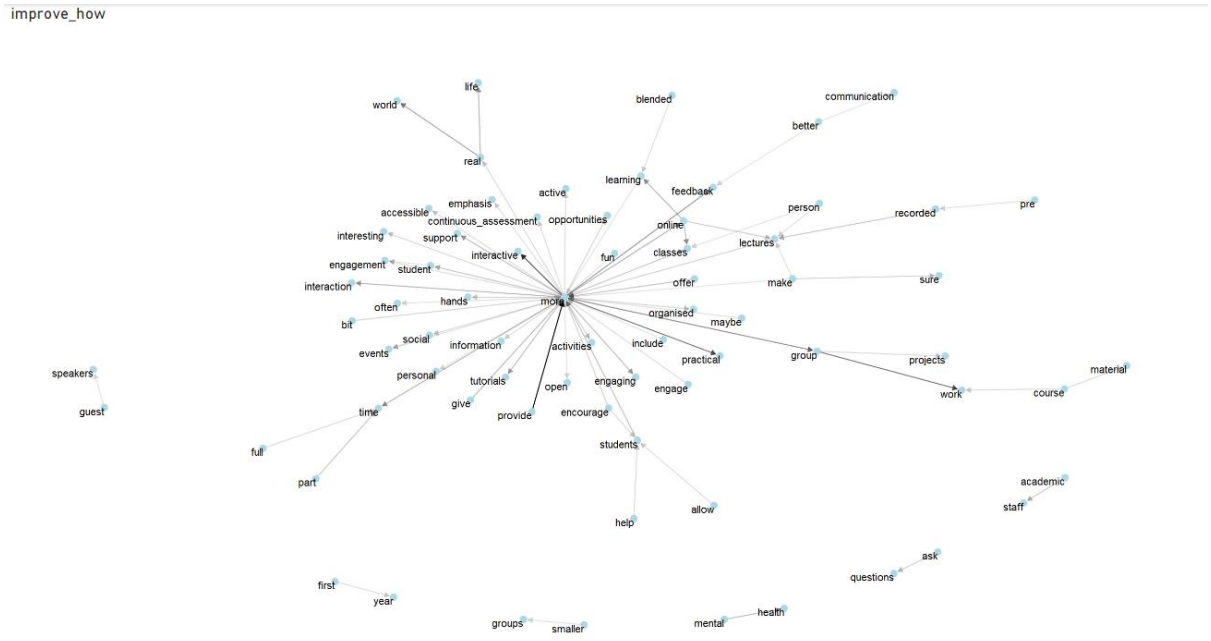
The analysis conducted so far has shown that responses to Q2 tend to follow a formula of modifier plus item and has shown how effective bigrams are in outlining the responses of students. Presenting the frequency of certain modifiers allows us to see how students have responded but requires us to be selective; what word are we interested in and do we want to use the word as word 1 or word 2 in a bigram? However, doing this is only ever going to provide us with a snapshot of responses that use the terms we are interested in. This is not in itself a poor choice of direction in analysis, but it is also not going to provide any answers about how students answered Q2 as a whole. The next step is to see how the bigrams relate to each other rather than just viewing them in isolation from one another.

The charts presented in the section show bigrams where there are at least 250 cases and visualise the relationships among words simultaneously, rather than just a selected word, and words frequently associated with each. As such, the charts are a visualisation of a Markov chain, which is a common model in text processing, where the choice of a word only

depends on its previous word. The arrows linking words show the direction of association. For example, the words ‘full’ and ‘part’ are commonly associated with ‘time’, and ‘time’ is associated with ‘students’. Examining the links between words through following the arrows builds commonly used phrases within the corpus.

Figure 3.11 presents the Markov chain bigram network for all students and as we would expect at this point with everything that has been covered so far, ‘more’ is the central node from which most words emanate, and repeat bigrams seen already in Figures 3.4 and 3.5 such as ‘more feedback’, ‘more continuous assessment’, and ‘more tutorials’. Beyond this, there are a few words on the bottom of the chart which stand apart from the hub around ‘more’, meaning they are most commonly associated with each other than words in the hub. These include ‘mental health’, ‘smaller groups’, and ‘first year’.

Figure 3.11: Markov chain bigram network (all students)



3.4 Conclusions

This chapter has been dominated by the word ‘more’. When asked what students would like to see from their institution to improve student engagement, they have typically articulated their thoughts by using the word ‘more’ as a component of their comment. It has been an inescapable facet of the responses’ students provided to this question, but at this point the charts in the Power BI report provide users with a coherent idea of what students would like to see ‘more’ of. Taken together the bigram networks present an overview of topics discussed by students in a broader context than the frequency of individual bigrams was able to provide.

4. Conclusion

The aim of this study has been to bridge the gap between a comprehensive and flexible statistical software package with a steep learning curve (R and RStudio) and a more accessible and intuitive one, though also less powerful (Power BI). Going into this study we wanted to make our prior research more accessible to interested researchers, but we were unsure if it was possible to replicate our work in Power BI. Luckily, we found a way to square this circle through integrating R code into Power BI itself. This means that interested researchers can use our report to see the analysis done in the previous chapters and also extend the work further in any direction they wish. In addition, as the R code is available to them within the report, if they wanted to, they could adjust the code further, and take initial steps using the code within an environment where they can always reset the code if it doesn't work, and move it into R if they want to take it even further.

At the outset, we thought that there would be a trade-off between detail and accessibility in that our work with R was more in-depth but less accessible to researchers without a background or experience in programming. In contrast, this project would not be as in-depth but would be immediately accessible to more users from varying research backgrounds. Ideally, interested researchers could use this project to explore the qualitative data collected and if the results of this merit further analysis they can utilise our earlier work to assist them with going further with their analysis.

However, for the most part, the integration of R and Power BI has gone beyond our own expectations, and by our estimation, we have been able to convert over eighty percent of the material covered in the previous report into a usable Power BI format, and to extend this in ways that were not possible in R. The key one being the use of filters which update the data instantaneously, which makes the report 'live' where it was previously 'static'. Table 4.1 below covers the analysis conducted for the previous report and whether we have been able to replicate it in the Power BI report.

Table 4.1: Analysis conducted for the previous report and whether it has been replicated in this Power BI report.

Q1 Analysis	
Meta-analysis and frequency of characters	☑
Word cloud	☑
Relative frequency of most-used words	☑
Relative frequency analysis (keyness)	☒
Semantic networks	☑
Latent semantic scaling	☒
Sentiment analysis	☑
Q2 Analysis	
Meta-analysis and frequency of characters	☑
Word cloud	☑
Relative frequency of most-used words	☑
Relative frequency analysis (keyness)	☒
Semantic networks	☑
Bigrams	☑
Bigram networks	☑
Weighted sentiment analysis	☒

In summary, this research project has again demonstrated the utility of using computer-assisted analytical techniques to process data which other researchers would have great difficulty in processing and evaluating due to the scale of the material involved. Furthermore, this research has provided a framework for further research to build upon. Our analysis is available in each tab of the Power BI report and is largely based upon the open-source statistical software *R* and its graphical interface R Studio, along with a suite of R packages and because of this, all of the techniques utilised within the report are replicable, and our coding framework and code can be extended for further research within the current corpus or further iterations of the Irish Survey of Student Engagement.

Appendix A: Stopwords

Standard Stopwords:

A	About	Above	After	Again	Against	All
Am	An	And	Any	Are	Aren't	As
At	Be	Because	Been	Before	Being	Below
Between	Both	But	By	Cannot	Can't	Could
Couldn't	Did	Didn't	Do	Does	Doesn't	Doing
Don't	Down	During	Each	Few	For	From
Further	Had	Hadn't	Has	Hasn't	Have	Haven't
Having	He	He'd	He'll	Her	Here	Here's
Hers	Herself	He's	Him	Himself	His	How
How's	I	I'd	If	I'll	I'm	In
Into	Is	Isn't	It	It's	Its	Itself
I've	Let's	Me	More	Most	Mustn't	My
Myself	No	Nor	Not	Of	Off	On
Once	Only	Or	Other	Ought	Our	Ours
Ourselves	Out	Over	Own	Same	Shan't	She
She'd	She'll	She's	Should	Shouldn't	So	Some
Such	Than	That	That's	The	Their	Theirs
Them	Themselves	Then	There	There's	These	They
They'd	They'll	They're	They've	This	Those	Through
To	Too	Under	Until	Up	Very	Was
Wasn't	We	We'd	We'll	We're	Were	Weren't
We've	What	What's	When	When's	Where	Where's
Which	While	Who	Whom	Who's	Why	Why's
Will	With	Won't	Would	Wouldn't	You	You'd
You'll	Your	You're	Yours	Yourself	Yourselves	You've

Custom Stopwords:

Lot	Also	One	Us	Get	Can	Lots
Like	???	Its	Just