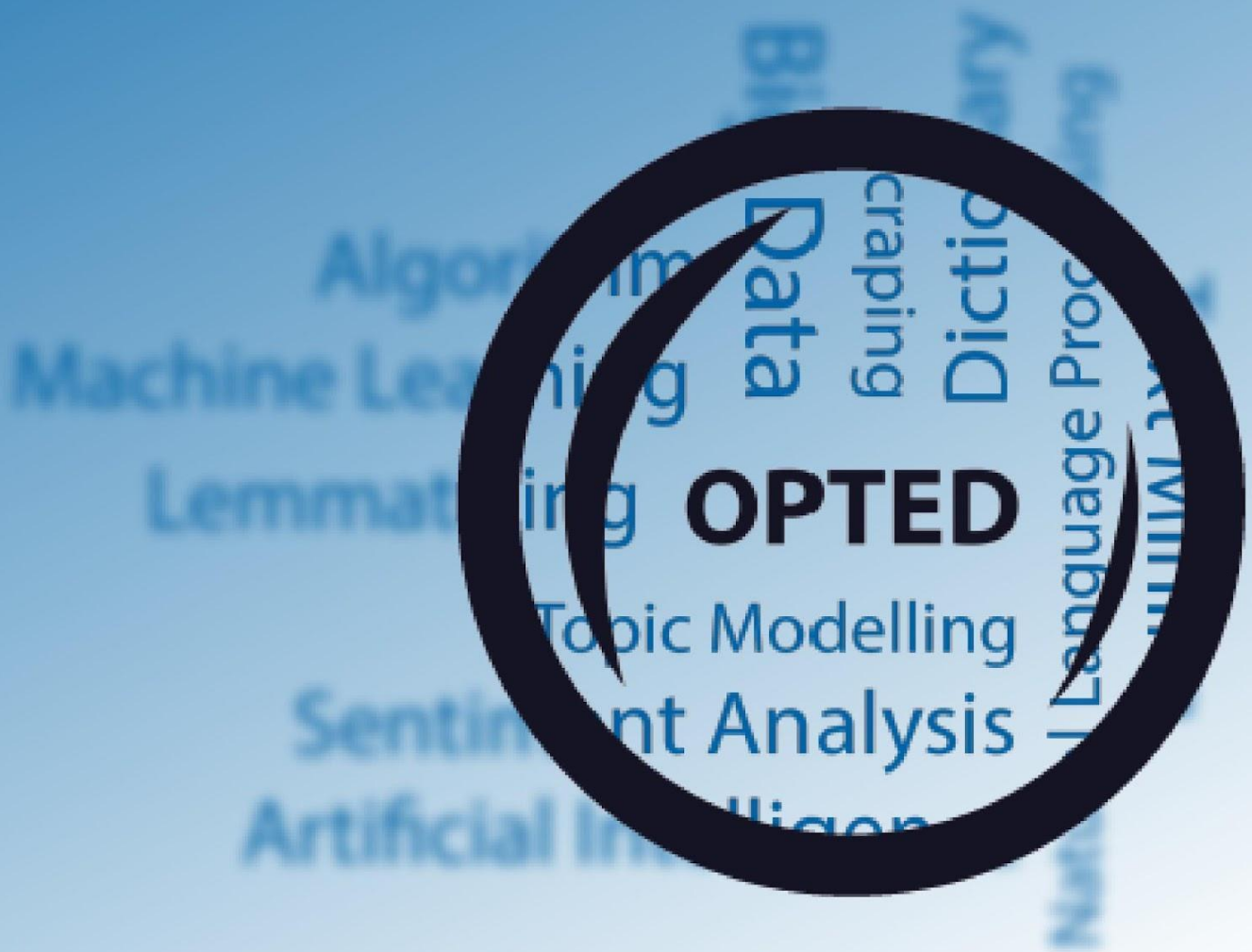
OPTED

Report on Training and Research needs and opportunities for the OPTED Platform

Susan Banducci, Michele Scotto di Vettimo, Shota Gelovani & Yannis Theocharis



Disclaimer

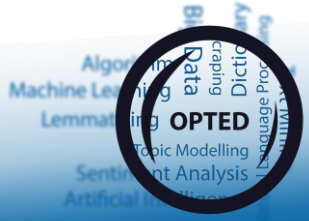
This project has received funding from the European Union's Horizon 2020 Research & Innovation Action under Grant Agreement no. 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

Dissemination level

Confidential

Type

Report



OPTED
Observatory for Political Texts in European Democracies:
A European research infrastructure

Report on Training and Research needs and opportunities for the OPTED Platform

Deliverable D9.3

Susan Banducci[1], Michele Scotto di Vettimo[1], Shota Gelovani[2], Yannis Theocharis[2]

[1] University of Exeter
[2] TU Munich

Due date: May 2022

1 Executive Summary

1.1 Purpose of the analysis of the training and research needs

The OPTED platform aims to empower interested audiences to easily find the political text data and analysis tools that help them to answer their questions. Previous reports presented our initial work and results on the framework for a curated resource classification system, mapping the principles that will help us organise, classify, and link the appropriate resources including text sources, documentation, software packages and tools and training materials that will feature on the OPTED platform.

In this report, we **analyse the training and research needs of the user community** the platform aims to reach, with a view **to devise a structure for training opportunities and provide exemplary training**. By asking evaluations about the training opportunities available to our user community, we are also better placed to assess the quality, accessibility and perceived usefulness of existing training resources and make evidence-based recommendations for changes, as well as to determine the contribution that can be brought by the OPTED infrastructure.

Our report proceeds as follows: first, we clarify the rationale behind the assessment of the training needs for the community of scholars working on text analysis, and explain the strategy devised to conduct this training needs assessment. Then, Section 3 reports the results of the survey with regard to respondents' fields of interest and techniques used, as well as concerning their engagement with – and evaluations of – available training opportunities. Subsequently, Section 4 covers the part of the questionnaire related to the preferences for future training events. In particular, we assess the attitudes towards the format of the event, the instructor and the skill-level of the needed training. Finally, in Section 5 we summarise the main results and outline the next steps necessary to devise a structure for training opportunities within the OPTED platform.

Our main results in this report are:

- 83% of the respondents are currently using text analysis techniques in their research.
- A majority of respondents currently use or plan to use computational methods. Yet, this aspect varies across research fields and on the basis of academic rank.
- The time or effort required, the need for funding, the availability of required training, as well as concerns for the measurement validity and the limited guidance offered in tools' documentation represent the main reasons why respondents decided not to use computational methods.
- Self-led online materials or online examples and posts are the first resources respondents use to acquire their skills in using text analysis methodologies. Only 18% of the respondents claim that they relied on training offered during their undergraduate, master or PhD studies.
- Around 45% of the respondents have never participated in a training event related to text analysis. Those who did participate were more likely to attend conferences or schools where some sessions where text analysis was not the exclusive focus. However, more than 82% of the respondents report that they are somewhat or very likely to participate in a training even in the next two years.
- In terms of typology of future training events, there is a slight preference for offline/in-person events over more online-based alternatives. 62% of the respondents reported that they prefer academics from their own field who are also using text analysis techniques to teach at these events. Finally, respondents report a need for introductory, intermediary, and advanced level training.

2 Analysis of training and research needs for the OPTED platform

An ever-increasing availability of digital texts as the most tangible traces of social and political phenomena has opened up new research opportunities for social scientists. Text analysis, with its many types and shapes, will be a vital tool in the endeavour to make sense of such data and their interactions for understanding of European democracies. The breadth of possibly relevant textual resources to study the functioning of and challenges to representative democracy and the need to do so over vast periods of time and across many countries, however, also pose new challenges to researchers.

One key challenge for the research field of political text analysis is posed by its fragmentation. On the

one hand, this is problematic from a resource point of view, and impairs the quality of scientific work in terms of reproducibility and validity. On the other hand, this fragmentation reflects also in the fact that expertise and training facilities appear to be centring on few countries with strong research infrastructures. Hence, researchers not only face a different set of training needs, but they also experience different challenges with regard to the access to relevant training resources. This is particularly true for the computational dimension of text analysis. The lack of comprehensive training in computational methods in the social sciences results in the creation of inequalities among scholars, and the myriad of approaches and standards prevents fruitful collaborative endeavours across social science fields (Theocharis and Jungherr 2020). Unfortunately, institutions have so far struggled in embracing this emerging multidisciplinary environment: "integrating computational training directly into social science (e.g., teaching social scientists how to code) and social science into computational disciplines (e.g., teaching computer scientists research design) has been slow. […] Computational researchers and social scientists tend to be in different units in distinct corners of the university, and there are few mechanisms to bring them together" (Lazer et al. 2020, 1060).

With this in mind, one objective of OPTED is to offer the training opportunities needed to foster the application of text analysis tools and techniques. While some organisations do offer more or less regular training in this area (e.g., GESIS), most of the time these are rather ad hoc organised workshops. Training events often remain in the initiative of individual researchers, while a structure within which services and innovation can take place would yield a more long-term perspective for the field. Therefore, the OPTED infrastructure plans to host training sessions for researchers in the various subdomains of text analysis (e.g., data scraping and data management, supervised, unsupervised methods). In addition, it will provide overviews of state-of-the art, free online learning opportunities for researchers across Europe and beyond.

With a view to devise a structure for training opportunities and to eventually provide exemplary training, one task of WP9 was to first identify such training needs in the research community, and evaluate strategies to best address them. The strategy devised to identify such needs is a consultation of the user community through a needs assessment questionnaire. In the following we clarify the characteristics of the respondents that we wanted to reach with our questionnaire and the type of training and research needs that have been assessed in the survey.

2.1 Identification and selection of the respondents from the user community

The OPTED platform aims at serving text analysis researchers and practitioners of different seniority and skills level. More specifically, we identified as the primary target group academic researchers and scholars that work in the realm of text analysis. However, a second important target group of OPTED are academic researchers with an interest in text-as-data approaches but limited knowledge or resources to invest in such methodologies. Therefore, we wanted to assess the training and research needs of a very heterogeneous research community that comprises PhD students, early- and mid-career researchers as well as senior researchers. Also, we wanted to reach researchers that are interested in text analysis, though have not used it already in their research.

The aim of reaching this heterogeneous audience informed our strategy for sampling respondents. Other OPTED WPs had previously collected contact details of scholars and practitioners working on text analysis. The specific inclusion criteria varied from one WP to another. For instance, WP6 targeted authors who published any studies using quantitative text-based research over the past five years in top journals in political science, communication, sociology, and psychology (see Baden et al. 2021), whereas WP2 collected email addresses of authors studying citizen-produced political texts. However, all these populations are contained within the primary target group of OPTED and represent a key part of the wider user community. Therefore, we asked all OPTED members to disseminate the training needs survey to all the list of contacts that they had in their possession.

However, a second important target group is represented by researchers that perhaps have not already employed text analytical techniques, but that are nevertheless interested in these approaches. To reach these individuals and make sure that we target also more junior researchers like PhD students, we decided to contact organisers of events related to text analysis held in Europe from 2018 onwards (e.g., method schools, conferences, or workshops), and to ask them to share the survey with all the event participants. Additionally, we also asked members of the steering committees of some ECPR Standing Groups to circulate the survey to the group members. Appendix 1 provides the list of events and ECPR groups contacted.

2.2 Types of needs assessed and dissemination

In order to devise appropriate training opportunities, we are interested in assessing different dimensions of training and research needs. First, we wanted to understand what types of political texts and text analysis methodologies garner more attention in the research community, either because they are used more frequently or because researchers claim to be interested in them. Secondly, we investigate more specifically the attitudes towards using computational methods for text analysis. We focus on the main challenges that prevented researchers from employing these methodologies or that were faced by those respondents who report using computational text analysis. Thirdly, we sought to determine the developments in terms of training material and database availability that would benefit our respondents the most. We aim at understanding what are the key improvements that researchers would like to see with regard to existing text analysis resources but also the most important gaps that should be filled with the development of new resources. Additionally, we ask more specifically about their experience with training resources, and their involvement in training events. This allows us to know more about the perceived usefulness of these events, their most successful and most problematic aspects, and the users' preferences in terms of format and instructors.

Although the majority of these questions had predetermined answer options, from a best practices point of view we also included qualitative questions in order to ensure that more nuanced feedback and comments were captured (Gurwitz et al. 2020). The survey was programmed on Qualtrics and has received the approval of the Research Ethics committee of the University of Exeter. The actual dissemination of the questionnaire started on February 23, 2022, when we asked OPTED members and event organisers to forward the survey to their lists of contacts. We sent out a reminder 14 days later and we also shared the survey on social media various times.

3 Results of the analysis of training and research needs

3.1 Profile of the respondents

As of May 22, 2022, 286 respondents have participated in the study. Given that the dissemination of the survey involved the collaboration of many intermediaries (e.g., event organisers) and the use of social media, we could not have a precise estimation of the response rate. However, we do know that WP2 – which collaborated with WP9 in the preparation and dissemination of the survey – sent the survey to around 2.700 addresses in their possession, and that 157 of them participated in the study. This results in a response rate of 5.8% among the respondents working on citizen-produced political texts.

Figure 3.1 (left panel) shows the seniority levels of the respondents. A majority of respondents are mid-career researchers (36%) and senior researchers (26%), but the sample also contains a high share of PhD students (20%) and early-career researchers (18%). Male respondents represent the 58% of the participants and female respondents the 40% (2% prefer not to say or do not identify as either male or female). 68% of the respondents are affiliated with a European institution,[1] 17% with an American one, 9% an Asian one, and 4% with either an African or Oceanian institution. In terms of fields of interest (Figure 3.1, right panel), 33.2% of the respondents reported to research mostly on political science, 30.7% communications, 13.7% sociology, 5.1% psychology, 3.2% economics, and 2.6% linguistics, with other fields representing less than 1% of the respondents each.[2]

[1] 34% from Western Europe, 17% from Northern Europe, 12% from Southern Europe and 5% from Eastern Europe.

[2] Note also that respondents could report more than one field of research.

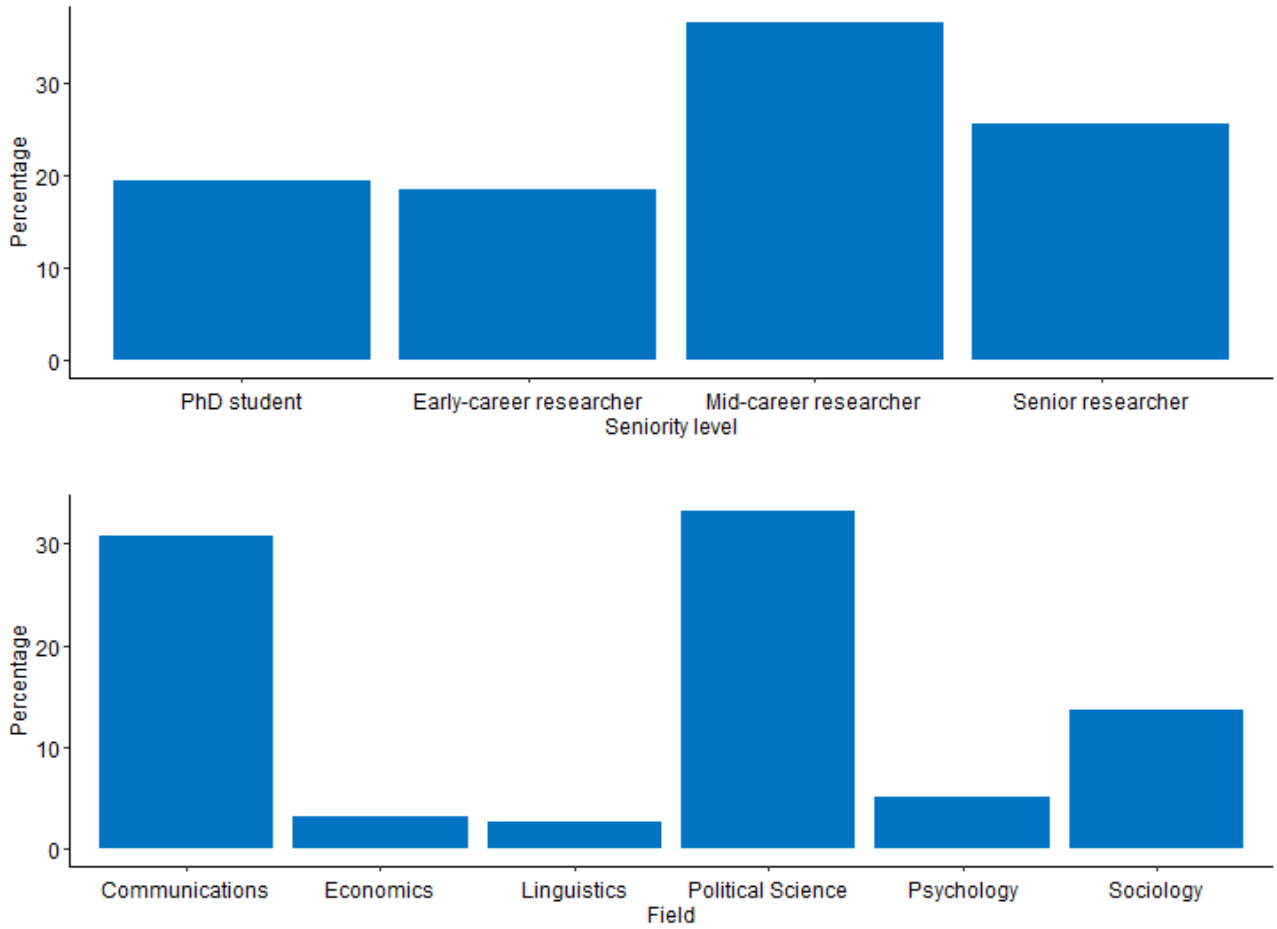


Figure 3.1 DISTRIBUTION OF RESPONDENTS BY ACADEMIC RANK AND FIELD OF RESEARCH

3.2 Experience with text analysis

We begin by investigating the current usage of text analysis techniques. Firstly, we ask about the type of text they use or wish to use in their research. In general, 83% of the respondents say that they are indeed currently performing text analysis in their works. As Figures A.1-A.8 in Appendix 3 show, more than 60% of respondents are currently using (or used in the past) texts produced by individual politicians, journalists or mass media outlets, or citizens; between 40% and 60% of the users are familiar with texts produced by legislative, bureaucratic and government organisations, political organisations, non-governmental organisations or social movements; and only around 30% of the respondents focus on non-media commercial organisations (e.g., enterprises or lobby groups). Finally, around 17% of the respondents reported using a different typology of political texts.[3]

In terms of perspective usage, more than 40% of users say that they plan also to analyse texts from non-media commercial organisations, non-governmental organisations and social movements, and for all other text typologies there are between 20% to 40% of users that say that they plan to include them in their research.

Secondly, we examine the usage of three types of text analysis methods (qualitative, quantitative manual and quantitative computational). Figure 3.2 shows that qualitative text analysis garners less interest than other quantitative techniques. Around 60% of respondents report that they rarely or regularly use qualitative text analysis, 22% of them do not use this technique, and less than 10% of the respondents say that they would like to use it in the future. On the contrary, a majority of respondents (52%) report that they regularly use

[3] When asked to indicate which other type of text they were using, these respondents indicated 34 different types of texts. Among those, most frequent entries had to do with documents from courts and judicial organisations, comments and captions from social media platforms, scientific or academic reports, biographical data and memoirs.

computational methods and 22% of them that they would like to use in the future. Around half of the users are also regularly using manual quantitative methods, with 27% of them also saying that they rarely use these techniques. Yet, only 7% of respondents say that they would like to use a manual methodology in the future.

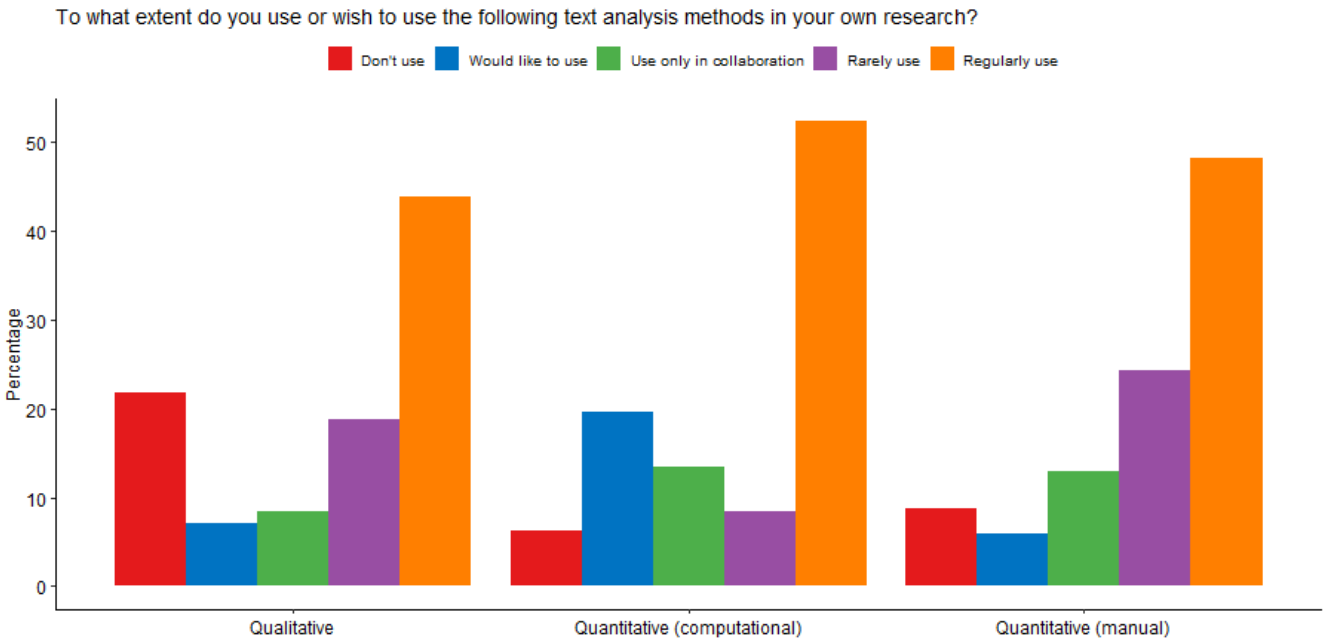


Figure 3.2 USAGE OF DIFFERENT TEXT ANALYSIS METHODS

Figures A.9-A.16 show, however, that the usage of different text analysis methodologies varies according to the field or the rank of the respondents.[4] For instance, political scientists show a clearer preference for quantitative methodologies rather than qualitative ones if compared to scholars doing communication research or sociology. The former are employing qualitative and quantitative techniques at a similar rate, whereas sociologists show a predilection for qualitative approaches. Similarly, PhD students signal a great interest in computational methods in particular, whereas senior researchers appear to use the three different text analysis approaches at a relatively similar rate. In between the two, early-career researchers also appear to behave like PhD students, whereas mid-career researchers show a more balanced pattern similar to that of senior researchers, with the exception that more than 20% of mid-career researchers indicates that they would like to use computational methods in the future, but less than 5% of them says that they would like to employ qualitative or manual approaches.

3.2.1. Reasons for *not* using computational methods

We then focus on users reporting either that they do not currently use computational text analysis or that they use it only when collaborating with others (19% of the sample), and ask what challenges were relevant (or might be relevant in the future) for their choice to not use computational methods. The results indicate (Figure A.17) that the required time or effort, the need for funding and the availability of required training were the top three "major challenges" identified. Additionally, training availability is also among the top three "minor challenges" identified in the questionnaire, along with concerns for the measurement validity and the limited guidance offered in tools' documentation.

Additionally, we allowed respondents to comment on the challenges faced. The comments received tap into the fact that training is perceived to be poorly designed as "some [of them] are too broad, others are too specific", whereas other respondents showed a more fundamental scepticism towards the capacity of computational methods in accounting for the nuances and "context" of the text analysed, signalling however

[4] Note that not all respondents filled the demographic questions at the end of the survey and, therefore, data about rank and field are not available for all participants.

that they are open to the use of computational methodologies if they are capable of accounting for these nuances.

3.2.2 Experience with computational methods for text analysis

Then, we pose similar questions about encountered or potential challenges to the respondents that are currently using computational text analysis (Figure A.19). Similarly to what happens to respondents not using computational methods, the time or effort required represent the top major challenge encountered by computational users. The other two top major challenges were the concerns for measurement validity and the need for funding. Respondents using computational methods also list the key "minor challenges" encountered: the availability of relevant training, the limited guidance offered by the tools' documentation, and the availability of a tool for a specific language. Interestingly, therefore, 4 out of 6 top major and minor challenges are shared by both actual users of computational methods and by those respondents that are not currently using such techniques. Computational users saying that the time or effort required was too challenging also reported that this was mostly due to high commitment to research work and other professional commitments, and only few of them mentioned personal or caring commitments (see Figure A.20).

Respondents also reported more specific challenges using the open-ended question. Overall, these answers related to concerns with reviewers' scepticism (for instance, one respondent mentions that in specific subfields a computational approach has to be accompanied by a more qualitative one in order for a research to be published in main journals); measurement validity and quality of the tools' documentation , that rarely help the user in understanding how to interpret and assess the quality of the output of a tool; and the availability of machine readable input data.

3.3 Skill-building, user community and participation to training events

The survey then moves more specifically to the analysis of the skill set of the respondents. In particular, we are interested in understanding how respondents enhance their skills with regard to text analysis methodologies, and how easily they can solve problems thanks to the available materials and connections with other users in the community.

We first ask how they acquired the skills necessary to utilise text analysis methodologies. As Figure 3.2 shows, the most common way is through self-led online material. Overall, 40% of the respondents report the use of either self-led online materials or online examples and posts. Around 27% say that community-led workshops, conference workshops or tutorials or third-party training courses helped them in acquiring their skills, whereas only 18% of the respondents claim that they relied on training offered during their undergraduate, master or PhD studies. Therefore, respondents seem to rely primarily on the resources available online, then on various types of training events and, only residually, on training provided during their university courses.

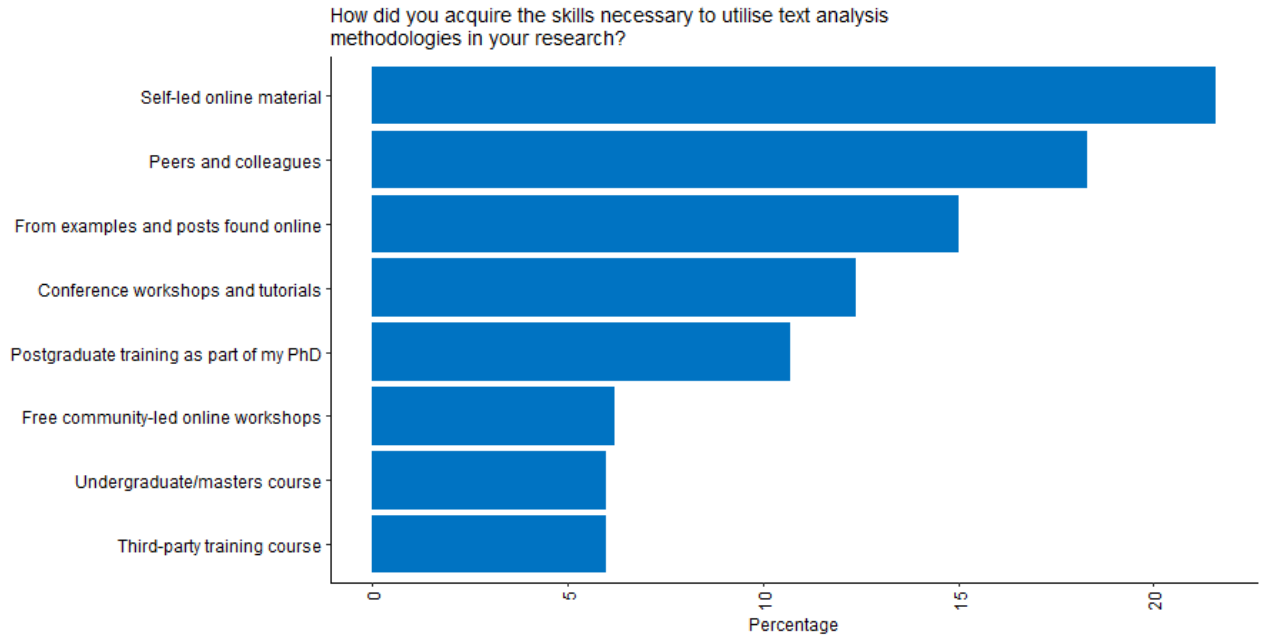


Figure 3.2 SKILLS ACQUISITION FOR TEXT ANALYSIS METHODOLOGIES

More specifically with regard to users employing computational methodologies, we also assess their agreements with statements tapping into the fora used to discuss (and solve) problems related to computational methodologies, and the perceived relevance of computational techniques for their competitiveness in the job market (see Figures A.21-A.25).

Firstly, 64% of the participants using computational methods think that it is necessary to have computational text analysis skills in order to be competitive on the academic job market, and only 19% of them claim that it is not the case. Secondly, a majority of respondents (53%) report that they can easily discuss problems related to computational text analysis with colleagues from their department. Hence, though peers and colleagues are not the primary channel for acquiring necessary text analysis skills, they prove to be useful for computational users in solving everyday problems. Nonetheless, 37% of the users express disagreement with the same statement. Thirdly, most of the participants report that they can easily solve problems by just looking at the documentation of the tools (55%) and that they rely on public platforms to further discuss problems related to computational methods (65%).

Nonetheless, 49% of the participants agree with the statement that it is sometimes hard to find a user community with which they can discuss problems related to computational methods, and only 24% of the respondents disagree with it. This indicates that, despite the importance of colleagues and the presence of online forums, for half of the computational users it is hard to find peers and practitioners with which they can discuss their everyday methodological problems.

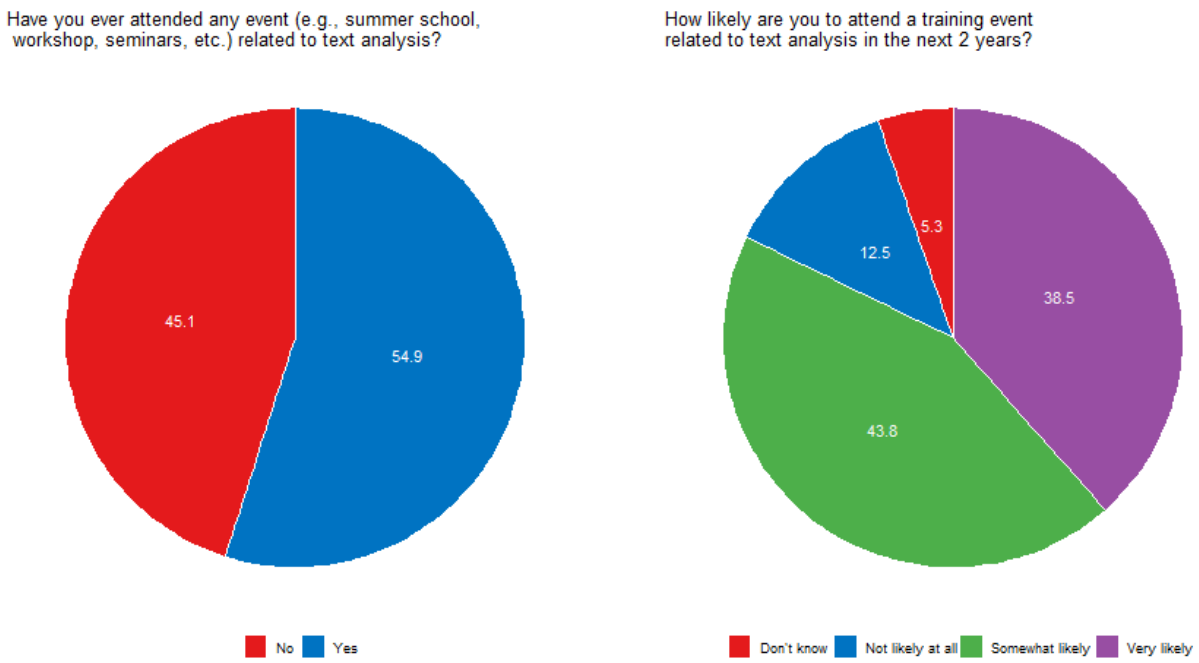


Figure 3.3 PAST AND PERSPECTIVE ATTENDANCE TO TRAINING EVENTS

Finally, we asked all our respondents their attitudes towards training events. First, we ask about attendance to past training events like summer schools, workshops, seminars, etc. It appears that around 45% of the respondents have never participated in a training event (Figure 3.3, left panel). Further inspection (Figure A.26) indicates that this relatively high percentage of respondents who never attended a training event is not explained by the more junior rank of some of the survey participants. 42% among the PhD students never participated in a training event. The share goes to 40.5% for early-career researchers, to 44% for mid-career researchers, and to 50% for senior researchers.

In terms of the type of training event attended, respondents did not focus on a very specific format (see Figure A.27). In general, they were more likely to participate in conferences or schools where some sessions were dedicated to text analysis (29.4% of the responses), rather than to similar events exclusively focused on text analysis (21.6%). Also, participation in seminars or workshops is more common than attendance of training sessions (27.4% compared to 21.6%).

The evaluations respondents gave of these events are fairly positive (Figures A.28-A.31). 83% of the respondents claim that the events were either somewhat or very helpful for learning new software, resources and techniques, and 87% say that they were helpful in improving their training in software, resources and techniques they already knew. Training events are also seen as a positive networking opportunity, with 79% of respondents saying that they were helpful in getting to know a community of scholars working with similar methodologies as theirs. Finally, "only" 65% of the users claim that such events were helpful in getting feedback on their ongoing research.

We gave respondents the possibility to comment on what were – from their perspective – the key strengths and weaknesses of the events in which they participated. In terms of strengths, users emphasised the possibility of having discussion and exchange with peers (even between sessions), sometimes actively promoted through a collaborative component; the fact that instructors were key experts in the field (sometimes even the creator of a specific software); the fact that the analysis and training was very applied, with an hands-on approach, and relevant for their own research; the possibility of getting an informed overview of the state of the field and alternative techniques and methodologies; and the fact that the training material was made available in advance so that they could benefit the most from the actual training sessions.

Conversely, with regard to the weaknesses, respondents pointed out various elements that could be linked to the issue of the appropriateness of the level of training and great diversity in terms of participants' starting points. Some respondents are dissatisfied with the fact that complex concepts were often taken for granted even though the participation was not restricted to experienced users, others, instead, complained that most events cover fairly introductory topics and that it is difficult to find some more advanced training. Respondents

also found that these events should also cover more practical issues like privileging training with open access or not to costly software, addressing problems related to data storage and memory usage. In terms of course design, users also felt that training should give more space to working on one's own data. As one respondent put it: "classes often work with data where the output is useful, since the dataset has been tested before and the instructor knows which data cleaning steps to apply. But once you use your own data, this rarely happens. It would be super helpful to instead use one's own data and discuss in the class how to improve the results". Finally, respondents also emphasise the need to cover more issues related to interdisciplinarity by bringing in examples from as many domains as possible and by combining more qualitative and quantitative approaches.

Finally, when asked how likely they are to participate in a training event related to text analysis in the next two years, 82.3% of the respondents report that they are somewhat or very likely to participate, with only 12.5% saying that they are not likely at all (Figure 3.3, right panel). Further inspection (Figure A.32) suggests that there is some variation across different respondents' ranks. Early-career researchers are the ones with the highest share of prospective participants (97.2%), followed by PhD students (87.2%), mid-career researchers (80.4%), and senior researchers (68.7%).

4 Preferences for future training

The survey, then, assesses respondents' preferences for future training. We tap into preferences about the format of the event, the type of instructor, and the level of the training. We also ask scholars directly involved in the training of other researchers what they think are the most important training needs and who is in the best position to train young researchers. The feedback received on these questions will be key in informing the design of training resources which will be made available on the OPTED platform.

Firstly, with regard to the type of training event respondents would like to attend in the next two years, there is a slight preference for offline/in-person events. The open answers to the question about weaknesses of past training events already hinted at the fact that respondents do not prefer events that are entirely run online, as one respondent reported that most benefits of participating in such events like networking opportunities, discussions and the possibility of having constructive debates about one own's research needs are somewhat reduced in an online setting. In fact, 32% of users preferred purely offline events over the other alternatives. 28% of them prefer online resources but with pre-recorded material, whereas 24% of the users say that they are likely to attend hybrid events with both online and offline sessions. However, only 16% of the respondents are likely to attend events that are only online and with synchronous sessions (see Figure 4.1).

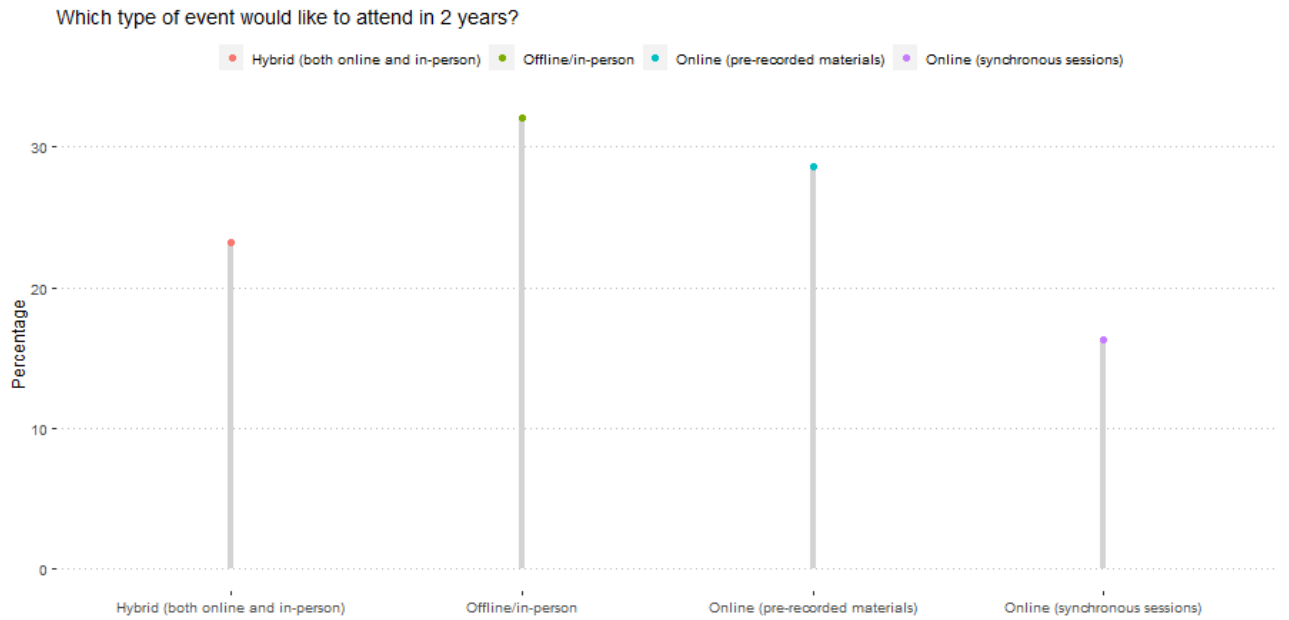


Figure 4.1 PREFERRED FORMAT FOR TRAINING EVENTS

When asked about whom they would like the most to deliver the training, a solid majority of 62% of the respondents reported that they prefer academics from their own field who are also using text analysis

techniques. 33% of the respondents would instead like academic experts in text analysis but that are not necessarily familiar with their field, and only 5% of the respondents would like non-academic experts and practitioners coming from the industry (Figure 4.2). Interestingly, we receive very similar results when we ask supervisors of researchers who need to learn text analysis who is in the best position to deliver training sessions for such researchers. 67% of the respondents reported being currently involved in supervising researchers working on text analysis. Among the supervisors, 61% of them believe academics from the same field as the researcher are in the best position to run training sessions, compared to 37% for academics expert in text analysis but not necessarily familiar with the field, and 2% of non-academic trainers (Figure A.33).

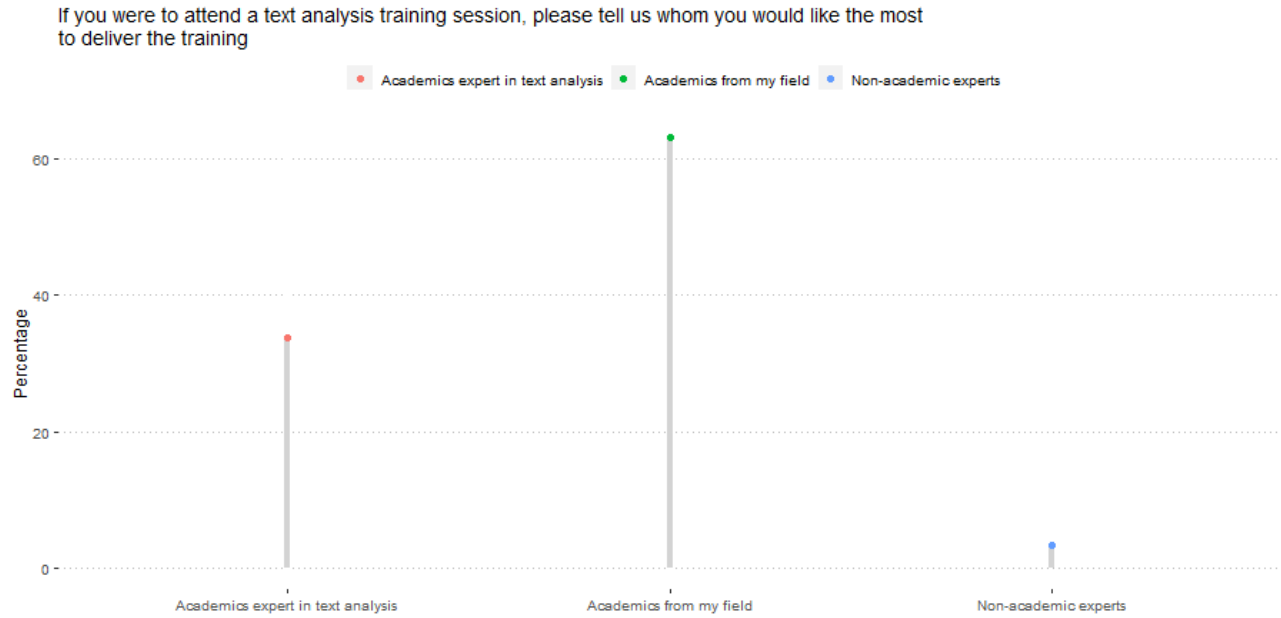


Figure 4.2 PREFERRED INSTRUCTOR FOR TRAINING EVENTS

Another question we ask only to respondents involved in the supervision of researchers working on text analysis is about the most important training needs of these researchers. An area in which 87% of the supervisors believe there are either very or extremely important training needs has to do with theory and concepts related to text analysis. Additionally, 78% of the supervisors think that data and open access tools represent one area where there are either very important or extremely important training needs. Similarly, 75% of the trainers report that programming and software skills represent an area with (very or extremely) important training needs. Finally, there is also a concern with regard to matters related to research integrity and ethics, with 76% of the respondents mentioning this as an area with either very or extremely important training needs (see Figures A.34-A.37). Additionally, when asked to comment on additional areas where training is needed, supervisors mention data management planning and research methodology. Few other answers tap into needs that are more specific to a field or type of text, like multilingualism or software such as LexisNexis for newspaper articles.

Finally, we ask our respondents whether they are interested in receiving further training and at what level would then need to receive it. On the first question, 83% of the respondents showed interest in receiving further training. Among those saying they are not interested, 7% of them claim to have sufficient skills already, 4% say that they do not plan to use text analysis in the future, and 6% of them say they are not interested for other reasons (mostly because of lack of time). When further asked about the level at which they would need this additional training, 22% of those interested say that they would need introductory level training, 36% an intermediate level training and another 42% an advanced level training (Figure 4.3).

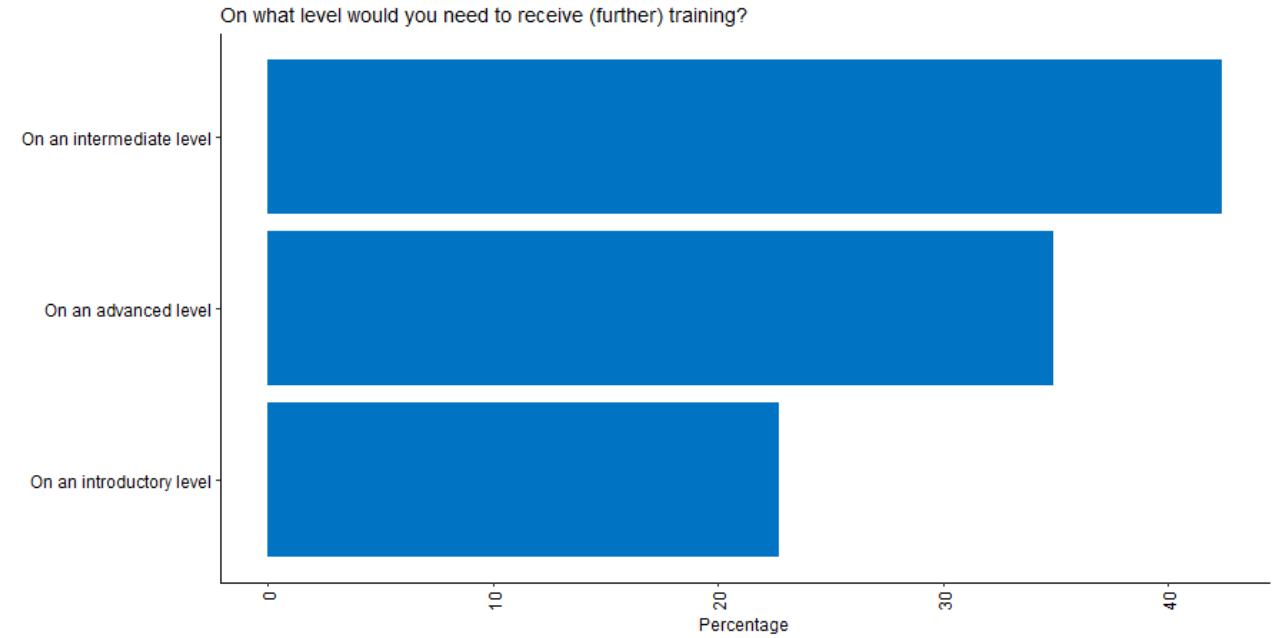


Figure 4.3 PREFERRED LEVEL FOR TRAINING EVENTS

5 Summary and next steps

This report presented the results of a training needs survey run in February-May 2022 and targeting researchers working or that are potentially interested in working on text analysis. Overall, 286 respondents participated in the survey. A majority of respondents are mid-career researchers (36%) and senior researchers (26%), but the sample also contains a high share of PhD students (20%) and early-career researchers (18%). 33.2% of the respondents reported to research mostly on political science, 30.7% communications, 13.7% sociology, 5.1% psychology, 3.2% economics, and 2.6% linguistics, with other fields representing less than 1% of the respondents each.

First, 823% of the respondents say that they are indeed currently performing text analysis in their works, and from the analysis emerge that they are interested in all the different types of political texts covered by the OPTED work packages. We show that qualitative text analysis garners less interest than other quantitative techniques, although the usage of different text analysis methodologies varies according to the research field or the rank of the respondents. When asked about what challenges were relevant (or might be relevant in the future) for their choice to not use computational methods, respondents indicated that the required time or effort, the need for funding and the availability of required training were the top three "major challenges". Additionally, training availability is also among the top three "minor challenges" identified in the questionnaire, along with concerns for the measurement validity and the limited guidance offered in tools' documentation. Similar challenges were also actually faced by users employing text analysis.

Respondents also indicate that the most common way to acquire the skills necessary to utilise text analysis methodologies is through self-led online material, and only 18% of the respondents claim that they relied on training offered during their undergraduate, master or PhD studies. Therefore, respondents seem to rely primarily on the resources available online, then on various types of training events and, only residually, on training provided during their university courses.

Around 45% of the respondents have never participated in a training event. Only 42% among the PhD students never participated in a training event. The share goes to 40.5% for early-career researchers, to 44% for mid-career researchers, and to 50% for senior researchers. In terms of the type of training event attended, respondents did not focus on a very specific format. In general, they were more likely to participate in conferences or schools where some sessions were dedicated to text analysis (29.4% of the responses), rather than to similar events exclusively focused on text analysis (21.6%). Also, participation in seminars or workshops is more common than attendance of training sessions (27.4% compared to 21.6%).

5.1 Implementing user feedback when designing training opportunities

The results presented in Section 4 can be used to design training opportunities that build on existing users' experiences in an attempt to learn from good practices and at the same time to address particular gaps highlighted by the respondents.

In summary, we find that respondents have a slight preference for offline/in-person events. Respondents do not value events that are entirely run online, as one participant reported that most benefits of participating in such events like networking opportunities, and the possibility of having constructive debates about one's own research needs are somewhat reduced in an online setting. Eventually, only 16% of the respondents said that they are likely to attend events that are exclusively online and with synchronous sessions. When asked about whom they would like the most to deliver the training, a solid majority of 62% of the respondents reported that they prefer academics from their own field who are also using text analysis techniques. Finally, 82.3% of the respondents showed interest in receiving further training. When asked about the level at which they would need this additional training, 22% of those interested say that they would need introductory-level training, 36% an intermediate-level training and another 42% an advanced-level training.

In light of these results, we wanted to highlight few general recommendations for the future development of training resources within the OPTED platform:

- Format: though respondents showed a preference for in person events, these can be inserted in a wider constellation of online pre-recorded training materials. This will ensure a wider and perhaps ad hoc fruition from the user community perspective. What is to be avoided is to limit the platform to the organisation of online synchronous events that might have a more limited impact and are not seen as valuable as other types of events by the users.
- Instructors and content: respondents unambiguously signalled that what they value more in the training received is that fact that it can have a direct and clear relevance for their research. Hence, they prefer academics working in their field delivering such training (rather than text-as-data experts with a reduced familiarity with a specific field). This means also that training materials should be designed so that they move beyond generic examples and embrace as much as possible the different types of texts covered by the OPTED platform with specific examples and resources.
- Relevance for the users' research work: respondents indicated that one key successful feature of training events is the fact that what they learn brings direct benefits to their research work. One task of the OPTED platform should be to design opportunities that are more tailored to the data and research needs of the participants, perhaps by involving prospective participants in the selection of example data to be used during the training or with the provision of a more personalised feedback to participants.
- Level of the training: the survey highlighted that the mismatch between the skill-level of the participants and the level of the training event is one of the key issues that negatively impacts the user satisfaction with the training received. At the same time, the respondents consulted indicate that there is a demand for training opportunities at an introductory, intermediate and advanced level. Hence, the OPTED platform should be as comprehensive as possible in terms of the level of the training resources hosted, but also very transparent to the users in terms of skill requirements for a resource and in terms of learning outcomes upon completion of a training opportunity. Ideally, when more resources cover similar topics at different skill-levels, the users should be guided towards the selection of the most appropriate resource for their needs.

References

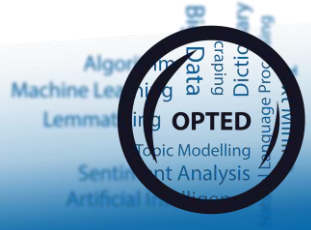
Baden, C., Pipal, C., Schoonvelde, M., and van der Velden, M. (2021). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, 16(1), 1-18. https://doi.org/10.1080/19312458.2021.2015574

Gurwitz, K. T., Singh Gaur, P., Bellis, L. J., Larcombe, L., Alloza, E., Balint, B. L., Botzki, A., Dimec, J., Dominguez del Angel, V., Fernandes, P. L., Korpelainen, E., Krause, R., Kusak, M., Le Pera, L., Leskosek, B., Lindvall, J. M., Marek, D., Martinez, P. A., Muyldermans, T., Nygard, S., Palagi, P. M., Peterson, H., Psomopoulos, F., Spiwok, V., Van Gelder, C. W. G., Via, A., Vidak, M., Wibberg, D.,

Morgan, S. L., and Rustici, G. (2020). A framework to assess the quality and impact of bioinformatics training across ELIXIR. *PLoS Computational Biology*, 16(7), e1007976. https://doi.org/10.1371/journal.pcbi.1007976

Lazer, D., Pentland, A., Watts, D., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M., Strohmaier, M., Vespignani, A. and Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060-1062. https://doi.org/10.1126/science.aaz8170

Theocharis, Y. and Jungherr, A. (2020). Computational social science and the study of political communication. *Political Communication*, 38(1-2), 1-22. https://doi.org/10.1080/10584609.2020.1833121



Appendix 1 – Text analysis events and ECPR Standing Groups contacted

1. Text analysis events

- COMPTEXT Conference (2018, 2019, 2020, 2022)
- Computational Text Analysis Workshop – Florence (2020)
- ECPR Summer School – Text analysis courses (2018, 2019, 2020, 2021)
- ECPR Winter School – Text analysis courses (2018, 2019, 2020, 2021)
- Essex Summer School – Quantitative text analysis course (2020, 2021)
- Quantitative Text Analysis Workshop – Dublin (2019)

2. ECPR Standing Groups

- Elites and Political Leadership
- European Union
- Interest Groups
- Internet and Politics
- Laws and Courts
- Parliaments
- Participation and Mobilization
- Political Communication
- Political Methodology
- Political Networks
- Political Parties
- Political Representation
- Politics and Technology
- Regulatory Governance

Appendix 2 – Survey Questionnaire

Dear Colleague,

This survey is part of the European Union funded research project OPTED: "Observatory for Political Texts in European Democracies: A European research infrastructure" (Horizon 2020 Grant 951832).

This survey is aimed towards social scientists who use text analysis methods in their research. It takes about 15 minutes to complete. You will be asked about your experience with using text analysis methods and, in particular, about your experience with existing resources and materials, and preferences for additional training resources. The purpose of the survey is to learn more about the challenges of conducting text analysis research. The outcomes of this survey will be used to create actionable insights to advance the state of text analysis in the social sciences to the benefit of the social science community.

Before you start the study, it is important that you are aware of the voluntary nature of the survey and how we protect your privacy. All the data is anonymous and will be treated in an aggregated manner. Please read information available at this link and do not hesitate to ask for clarification if you have questions (see contact details below).

By clicking on "I consent", you confirm the following: I am 18 years or older. I have read and understood the information for participants. I agree to participate in the research and to the use of data obtained from it. I reserve the right to revoke this consent without giving any reason. I reserve the right to discontinue the research at any time.

Thank you for your participation in this study!

For questions please contact: Prof. Susan Banducci at the University of Exeter (s.a.banducci@exeter.ac.uk) or Prof. Karolina Koc-Michalska at the Audencia Business School (kkocmichalska@audencia.com).

consent
If you do not consent to participate in the study, you can now end your participation by clicking on "I prefer not to participate".

- ○ I consent
- ○ I prefer not to participate

End of Block: Informed Consent

Start of Block: Uses text data or not

q1
First, we would like to learn about your use of textual data in your research, as well as the methods that you use to study text.

What sorts of text do you use, plan to use or wish to use in your research? Please consider any uses of textual data in your research, even if this may not be your primary research focus.

	Don't use and no plans to use	Don't use but would maybe use in future	Currently using or used in the past	Don't know
legislative, bureaucratic and government organizations (e.g, parliamentary speeches, bills, laws)	○	○	○	
political organizations such as political parties (e.g., manifestos, speeches, press releases)	○	○	○	
individual politicians (e.g., speeches, interviews, social media posts)	○	○	○	
journalists and/or published by mass media outlets (e.g., news articles, op-eds)	○	○	○	
non-media commercial organizations such as enterprises, lobby groups (e.g., reports, campaign materials, press releases)	○	○	○	
non-governmental organisations and social movements (e.g., forums, pamphlets, campaign materials)	○	○	○	
individual citizens (e.g., tweets, Facebook posts, blog posts, petitions)	○	○	○	
Other (please specify):	○	○	○	

End of Block: Uses text data or not

Start of Block: Methods & Texts

q2
To what extent do you use or wish to use the following text analysis methods in your own research?

	I don't use it	I would like to use it in the future	I only use it while collaborating with others	I rarely use it	I regularly use it
Qualitative text analysis (e.g., discourse analysis, conversation analysis)	○	○	○	○	○
Quantitative manual text analysis (e.g., manual content analysis)	○	○	○	○	○
Computational text analysis (e.g., automated content analysis, machine learning, text mining)	○	○	○	○	○

q2_open
If you use other methods than those mentioned above, please specify:

q3

To what extent do you use or wish to use the following kinds of software and tools in your own research on text?

	I don't use it and don't plan to use it in the future	I currenly don't use it, but would like to use in the future	I only use it while collaborating with others	I rarely use it	I regularly use it	I primarily use it
Statistical software (e.g., SPSS, Stata)	○	○	○	○	○	○
Mathematical software (e.g., Matlab, Octave)	○	○	○	○	○	○
Network analysis software (e.g., Gephi, Pajek, UCInet)	○	○	○	○	○	○
Qualitative data analysis software (e.g., MaxQDA, Atlas.ti, Nvivo)	○	○	○	○	○	○
Specialized text mining software (e.g., LIWC, SentiStrength, Wordscore)	○	○	○	○	○	○
Generalized text analysis platforms (e.g., AmCAT)	○	○	○	○	○	○
Open source coding platforms (e.g., Python, R, Julia)	○	○	○	○	○	○
Hosting and version control services (e.g., GitHub)	○	○	○	○	○	○

q4

Please indicate how much you agree or disagree with the following statements, relating to reliability testing (such as, e.g.

inter-coder reliability tests).

	Strongly disagree	Somewhat disagree	Somewhat agree	Strongly agree	Don't know
It is important to conduct a reliability test	○	○	○	○	○
Conducting reliability tests is difficult	○	○	○	○	○
It is at times unclear what reliability test should be used	○	○	○	○	○
I always report reliability in papers based on text data	○	○	○	○	○
Academic journals require reliability tests for the type of text analysis I conduct	○	○	○	○	○

End of Block: Methods & Texts

Start of Block: IF Computational

comp1
Which kinds of computational tools do you use to study text?

	I'm not familiar with	I know, but did not apply	I have applied
Text statistics (e.g., word frequencies)	○	○	○
Automated extraction (e.g., link/hashtag extraction, concordancers, keywords-in-context)	○	○	○
Dictionaries/keyword searches	○	○	○
Natural language processing tools (e.g., Part-of-Speech taggers, dependency parsers)	○	○	○
Sentiment scoring	○	○	○
Semantic network tools	○	○	○
Topic models/text clustering tools	○	○	○
Word embeddings	○	○	○
Text similarity scoring (e.g., plagiarism software)	○	○	○
Document scoring (e.g. Wordfish, Wordscore, Wordshoal)	○	○	○
Supervised machine learning	○	○	○
Machine translation	○	○	○

comp1_open
I use other computational tools (please specify):

comp3
In your experience, what challenges did you encounter (or you think you might encounter in the future) when using

computational methods for text analysis?

	not a challenge	a minor challenge	a major challenge
Time/effort required (e.g. technical requirements, experience)	○	○	○
Funding required (e.g., for training, fees, licenses)	○	○	○
Availability of required training	○	○	○
Training is available but you have no capacity to engage	○	○	○
Limited methodological guidance/documentation of tools	○	○	○
Infrastructure is not available (e.g., archives, computers with access to specialist software, etc.)	○	○	○
Availability of suitable computational tools for specific measurement purposes	○	○	○
Availability or comparability of suitable computational tools for the language(s) that I study	○	○	○
Issues concerning measurement validity/limited nuance	○	○	○
Skepticism (of myself or others) toward computational methods	○	○	○

comp3_open
Other challenges (please specify):

Display This Question: If comp3 = 1 [2] Or comp3 = 1 [3]

comp4
To what extent was the required time and effort challenging because of the following? (please tick all that apply)

- ☐ Childcare/other caring commitments
- ☐ Other personal commitments
- ☐ High commitment to research work
- ☐ High teaching load
- ☐ Other professional commitments
- ☐ Other, please specify:
- ☐ None of the above

comp6
Please indicate your agreement with the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I can easily discuss problems related to computational text analysis with colleagues from my department	○	○	○	○	○
I easily solve most problems related to resources for computational text analysis simply by consulting the documentation made available with the relevant resource	○	○	○	○	○
I rely on public platforms (e.g., Stack Overflow) to discuss problems related to computational text analysis methods and resources	○	○	○	○	○
Sometimes it is hard to find a user community with which I can discuss problems related to computational methods	○	○	○	○	○
It is necessary to have computational text analysis skills in order to be competitive on the academic job market	○	○	○	○	○

comp8
How important would the following be for your research activities?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
A single platform for text analysis tools and resources	○	○	○	○	○
Access to text analysis tools for programming	○	○	○	○	○

End of Block: IF Computational

Start of Block: IF Not Computational

nocomp1
What challenges were relevant (or might be relevant in the future) for your choice to not use computational methods for text analysis?

	not a challenge	a minor challenge	a major challenge
Time/effort required (e.g., technical requirements, experience)	○	○	○
Funding required (e.g., for training, fees, licenses)	○	○	○
Availability of required training	○	○	○
Training is available but I have no capacity to engage	○	○	○
Limited methodological guidance/documentation of tools	○	○	○
Infrastructure is not available (e.g., archives, computers with access to specialist software, etc.)	○	○	○
Availability of suitable computational tools for specific measurement purposes	○	○	○
Availability or comparability of suitable computational tools for the language(s) that I study	○	○	○
Issues concerning measurement validity/limited nuance	○	○	○
Reviewers'/editors' skepticism toward computational methods	○	○	○
Skepticism (of myself or others) toward computational methods	○	○	○

nocomp1_open
Other challenges (please specify):

Display This Question: If nocomp1 = 1 [2] Or nocomp1 = 1 [3]

nocomp2
To what extent was the required time and effort challenging because of the following? (please tick all that apply)

- ☐ Childcare/other caring commitments
- ☐ Other personal commitments
- ☐ High commitment to research work
- ☐ High teaching load
- ☐ Other professional commitments
- ☐ Other, please specify:
- ☐ None of the above

End of Block: IF Not Computational

Start of Block: Access & Publishing

q5
In your experience, which were challenges that you encountered relating to **accessing text data**?

	Not applicable	Not a challenge	A minor challenge	A major challenge
Restricted access to data by companies owning or storing the data (e.g., Twitter, Facebook, CrowdTangle, Crimson Hexagon/Brandwatch)	○	○	○	○
Restricted access to data due to content removal (e.g., online content that is flagged and removed by moderators)	○	○	○	○
Difficulty finding or identifying all relevant text data (e.g., because words that are spelled correctly by users are not collected)	○	○	○	○
Difficulty identifying or using the tools needed to access data	○	○	○	○
National ethical research rules or guidelines do not enable accessing certain text	○	○	○	○

q5_open
Other challenges (please specify):

q6
Please indicate how much you agree with the following statements, relating to data privacy, confidentiality and consent.

	Strongly disagree	Somewhat disagree	Somewhat agree	Strongly agree
I am aware of the best practices for ensuring confidentiality of those whose text data I collect/analyse	○	○	○	○
I have received sufficient training in the legal and ethical guidelines for data protection and privacy of those whose data I collect/analyse	○	○	○	○
It has not always been possible in my research to ask for consent to participate, even if gaining consent may have been best practice.	○	○	○	○

End of Block: Access & Publishing

Start of Block: Databases

q16
In order to undertake political text analysis, researchers need access to textual data (e.g., national legislation, news stories). Thinking about the future, how important are the following improvements to these types of databases for researchers?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Expanding the geographical coverage of existing datasets	○	○	○	○	○
Expanding the temporal coverage of existing datasets	○	○	○	○	○
Develop new datasets	○	○	○	○	○
Facilitating the linkage of existing datasets	○	○	○	○	○
Develop public web portals	○	○	○	○	○
Develop procedures for local installations of datasets	○	○	○	○	○
Develop API-web services	○	○	○	○	○

q16_open
Please provide more details as to what sort of database development is a priority, or suggest other types of database developments to prioritise: _____

End of Block: Databases

Start of Block: Languages

q7
In which language(s) are those texts that you study (or wish to study)? (please tick all that apply)
[Follows boxes with languages]

End of Block: Languages

Start of Block: IF Multiple Languages

lang1
To what extent were the following statements reasons for you to conduct textual analysis in multiple languages? These are languages...

	Not at all a reason	Minor reason	Major reason
that I speak well myself	○	○	○
that are spoken in a single country that I study (e.g., Switzerland, India)	○	○	○
that are particularly relevant for my research	○	○	○
in which it is possible/easy to access textual material	○	○	○
for which it is possible/easy to find qualified collaborators/ assistants	○	○	○
for which it is possible/easy to find suitable tools	○	○	○
for which the quality of machine translation suffices	○	○	○
for which it is possible/easy to conduct a comparative analysis	○	○	○

End of Block: IF Multiple Languages

Start of Block: Training

q8
How did you acquire the skills necessary to utilise text analysis methodologies in your research? (Tick all that apply):

- ☐ Self-led online material
- ☐ From examples and posts found online
- ☐ Free community-led online workshops
- ☐ Conference workshops and tutorials
- ☐ Peers and colleagues
- ☐ Undergraduate/master course
- ☐ Postgraduate training as part of my PhD
- ☐ Third-party training course
- ☐ Other (please specify): _____

q9
Have you ever attended any event (e.g., summer school, workshop, seminars, etc.) related to text analysis?

- ○ Yes
- ○ No

q10
How likely are you to attend a training event related to text analysis in the next 2 years?

- ○ Not likely at all
- ○ Somewhat likely
- ○ Very likely
- ○ Don't know

q11
What is the type of event you would most like to attend in the next two years?

- ○ Offline/in-person
- ○ Online with mostly synchronous sessions
- ○ Online with pre-recorded training materials (e.g., available to watch on demand)
- ○ Hybrid (both online and in-person)
- ○ Other (please specify): _____

q12
If you were to attend a text analysis training session, please tell us whom you would like the most to deliver the training

- ○ Academics from my field who are also using text analysis techniques
- ○ Academics expert in text analysis techniques, but who do not necessarily operate in my field
- ○ Non-academic experts and practitioners (e.g., coming from the industry)

q13
Are you currently involved in the training or supervision of social scientists? (e.g., as part of a degree program, doctoral/postdoctoral supervision, specialized training)

- ○ Yes
- ○ No

Display This Question: If q13 = 1

q13a
Do you believe that, in the context of your training activities for students and social scientists, there are important training needs in relation to text analysis in any of the following areas?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Data and open access tools	○	○	○	○	○
Programming and software skills	○	○	○	○	○
Theory and concepts	○	○	○	○	○
Research integrity, ethics	○	○	○	○	○

Display This Question: If q13 = 1

q13a_open
Other training needs (please specify): _____

q14
As a supervisor of researchers who need to learn text analysis, who do you think is suitable to lead the training sessions for these researchers? (skip if it does not apply to you)

- ○ Academics from my field who are also using text analysis techniques
- ○ Academics expert in text analysis techniques, but who do not necessarily operate in my field
- ○ Non-academic experts and practitioners (e.g., coming from the industry)

q15
Would you be interested in receiving (further) training in computational text analysis methods?

- ○ **Yes**
- ○ **No**, my skills are sufficient
- ○ **No**, I do not plan to use computational text analysis anytime soon
- ○ **No**, for other reasons (please specify): _____

Display This Question: If q15 = 1

q15a
On what level would you need to receive (further) training?

- ☐ On an introductory level
- ☐ On an intermediate level
- ☐ On an advanced level

End of Block: Training

Start of Block: IF Training events

train1
What kind of events did you attend? (Select all that apply).

- ☐ Method school exclusively focused on text analysis methods
- ☐ Method school where some courses covered text analysis methods
- ☐ Conference exclusively focused on text analysis methods
- ☐ Conference where some sessions/panels were dedicated to text analysis methods
- ☐ Seminar or workshop organised by another university
- ☐ Seminar or workshop organised at my university
- ☐ Training event organised by another university
- ☐ Training event organised by my university

train2
Overall, how would you rate the events you attended with regard to the following:

	Not helpful at all	Somewhat unhelpful	Neither helpful nor unhelpful	Somewhat helpful	Very helpful
Learning new software, resources and techniques	○	○	○	○	○
Improving your training in software, resources and techniques you already know	○	○	○	○	○
Getting to know a community of scholars working with similar methodologies	○	○	○	○	○
Receiving feedback on your ongoing research	○	○	○	○	○

train3
Thinking about the training events you participated in, what do you think was the key strength which should definitely be kept so as to ensure that they benefit users like you?

train4
Thinking about the training events you participated in, what do you think was the key issue which should be addressed so as to make them more helpful for users like you?

End of Block: IF Training events

Start of Block: OPTED proto-type

q17
The OPTED network is currently working on the proto-type of a platform for text analysis resources for the academic and the non-academic community interested in political texts (you can find more information at: opted.eu). We highly value any user input on this proto-type, as it will allow us to better meet the needs of the user community we aim to serve.

Which aspects do you think deserve to be prioritised to best satisfy the users' community needs, including your own?

	Not a priority at all	Definitely a priority	Definitely a priority **and** I would make use of it
An open repository for different types of **data sources and relevant documentation**	○	○	○
An open repository for different types of **training materials**	○	○	○
An open repository for different types of **tools, software and packages**	○	○	○
A platform where users can find **recommendations for relevant resources** for their research	○	○	○
A platform users can **contribute to with their own resources** (e.g., new databases, software or packages)	○	○	○
A platform with tools and standards for the **validation** of computational methods	○	○	○
A platform that can be used to work **collaboratively on the discovery, creation and sharing of text analysis resources** (e.g., codes or data)	○	○	○
A platform that can **host discussions** about text analysis resources	○	○	○
A platform for where users can **access training materials** (e.g., videos, slides, Shiny apps)	○	○	○
A platform that can be used to **"re-appraise" existing resources**, highlight and solve issues (e.g., bugs) and track updates	○	○	○

q18
What do you think are the key features the platform should possess in order to satisfy the user community needs?

_____

End of Block: OPTED proto-type

Start of Block: Demographics

field
What is your main area of research or study? (please tick all that apply)

- ☐ Communications
- ☐ Economics
- ☐ Political Science
- ☐ Psychology
- ☐ Sociology
- ☐ Other:

country
In what country is the university or employer with which you are affiliated?

▼ Afghanistan (1) ... Zimbabwe (1357)

rank
I am...

- ○ a PhD student
- ○ an early-career researcher (<5 years since PhD)
- ○ a mid-career researcher (5-15 years since PhD)
- ○ a senior researcher (>15 years since PhD)
- ○ Other (please specify):

gender
I identify as

- ○ a man
- ○ a woman
- ○ I do not identify as a man nor as a woman
- ○ I prefer not to say

mailing_list
Would you like to be contacted in the upcoming weeks with more information about the OPTED platform, so as to remain updated on its development and to have the possibility to try and comment on its proto-type?

- ○ Yes, I would like to be added to the OPTED mailing list
- ○ No, I do not want to be added to the OPTED mailing list

End of Block: Demographics

Appendix 3 – Answers to selected questions

q1

What sorts of text do you use, plan to use or wish to use in your research? Please consider any uses of textual data in your research, even if this may not be your primary research focus.

Figure A.1 USAGE OF LEGISLATIVE, BUREAUCRATIC AND GOVERNMENT TEXTS

Figure A.2 USAGE OF POLITICAL ORGANISATIONS' TEXTS

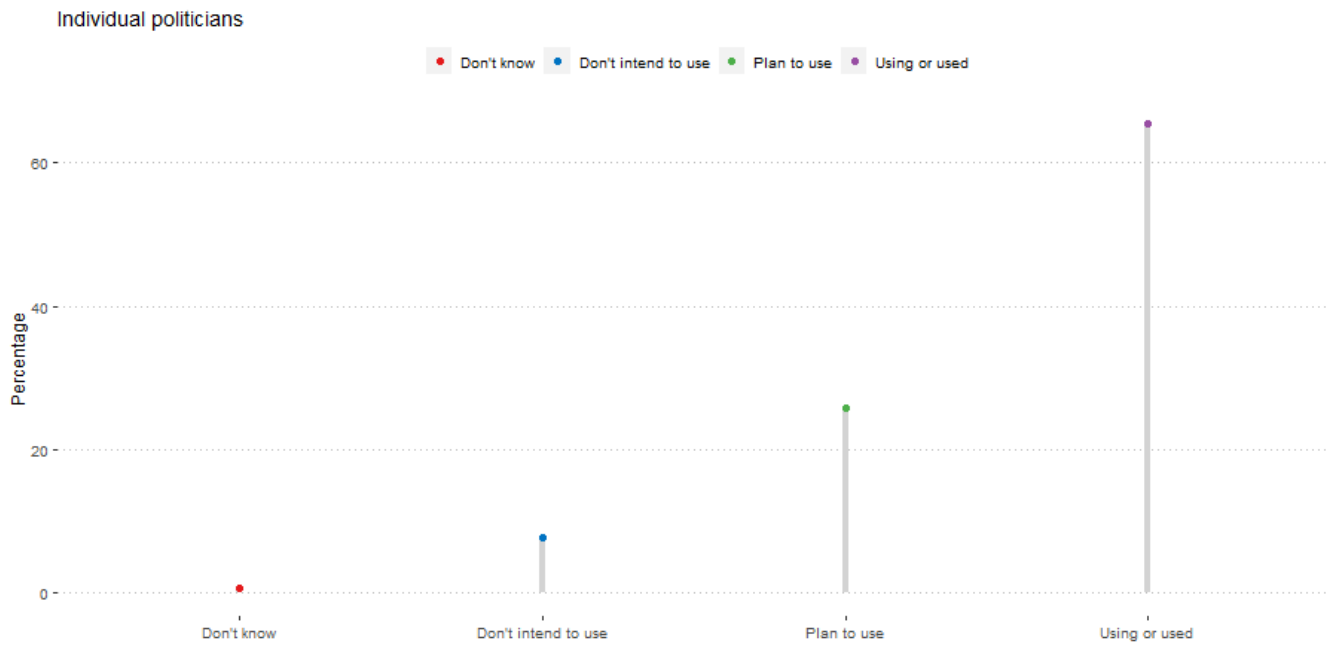


Figure A.3 USAGE OF TEXTS FROM INDIVIDUAL POLITICIANS

Figure A.4 USAGE OF JOURNALISTIC OR MASS MEDIATED POLITICAL TEXTS

Figure A.5 USAGE OF TEXTS FROM NON-MEDIA COMMERCIAL ORGANISATIONS

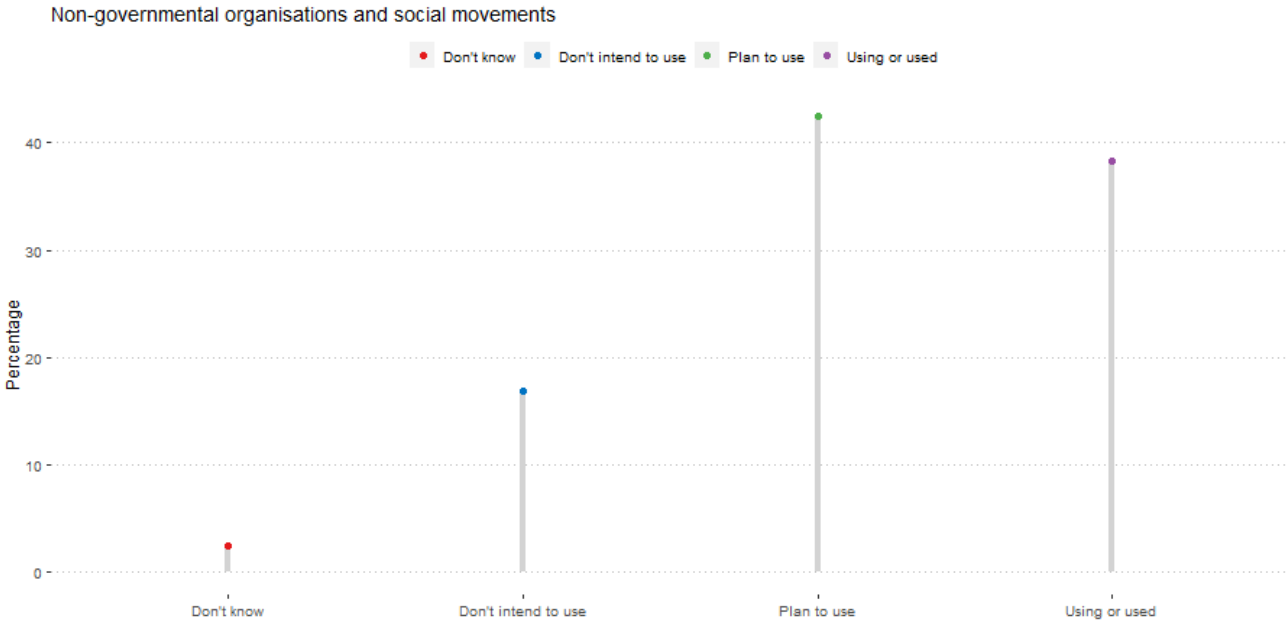


Figure A.6 USAGE OF TEXTS FROM NGOS AND SOCIAL MOVEMENTS

Figure A.7 USAGE OF TEXTS FROM INDIVIDUAL CITIZENS

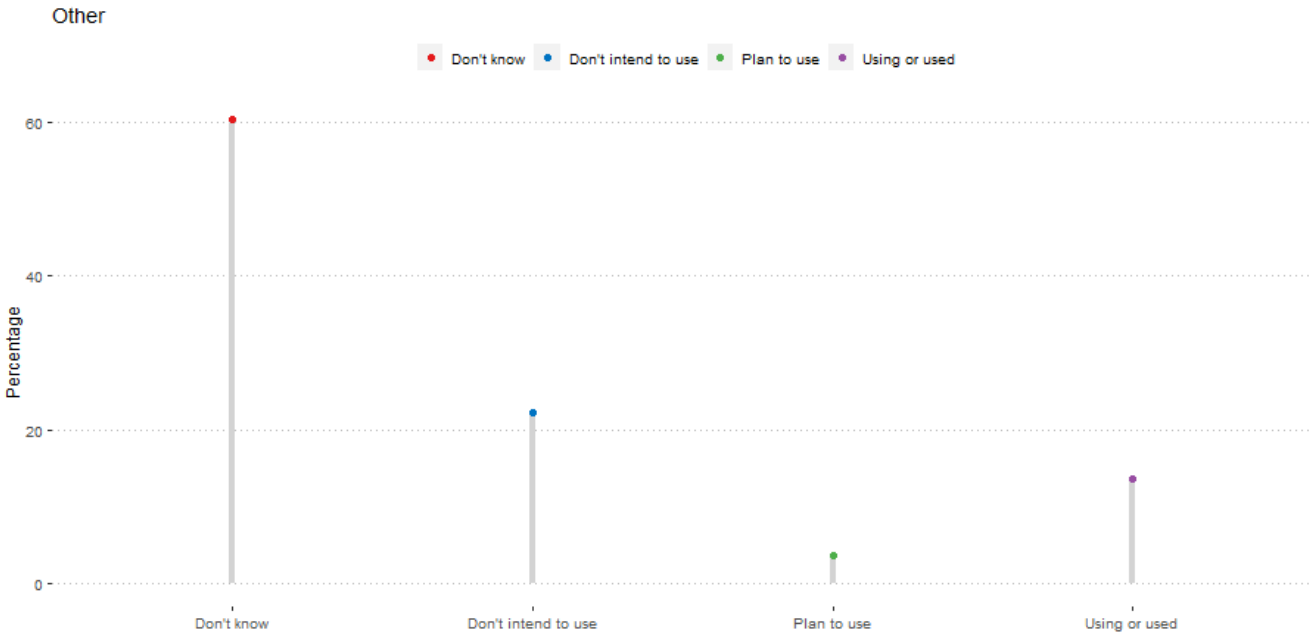


Figure A.8 USAGE OF OTHER TYPES OF TEXTS

q2

To what extent do you use or wish to use the following text analysis methods in your own research?

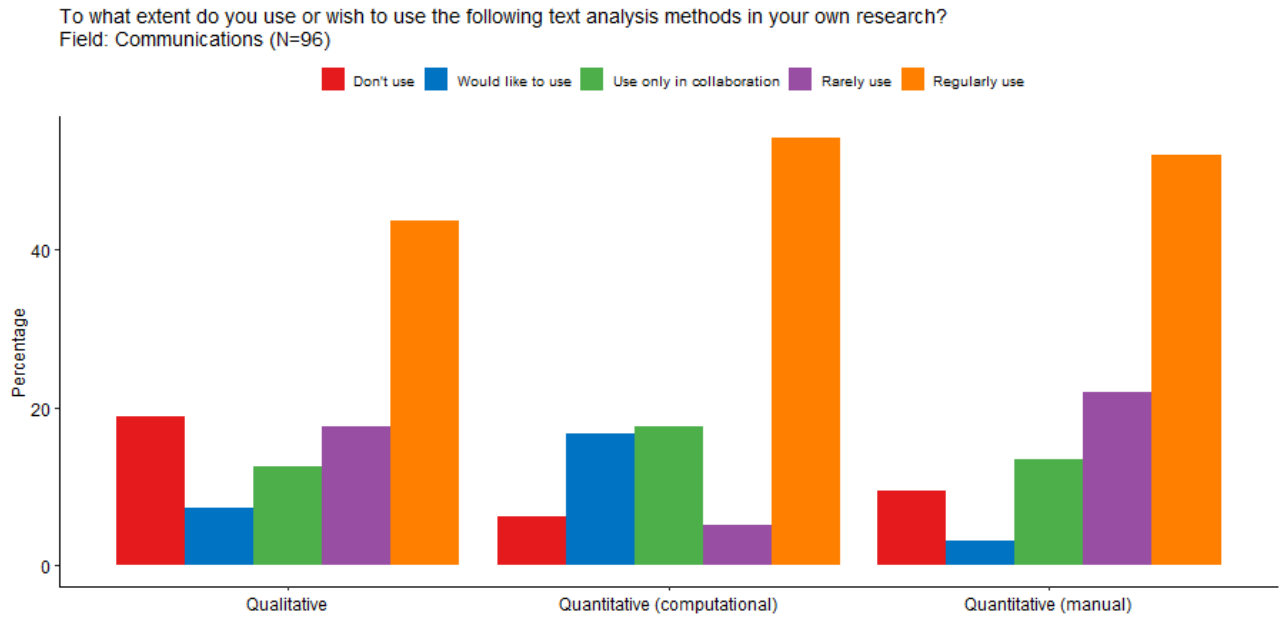


Figure A.9 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: COMMUNICATION RESEARCH

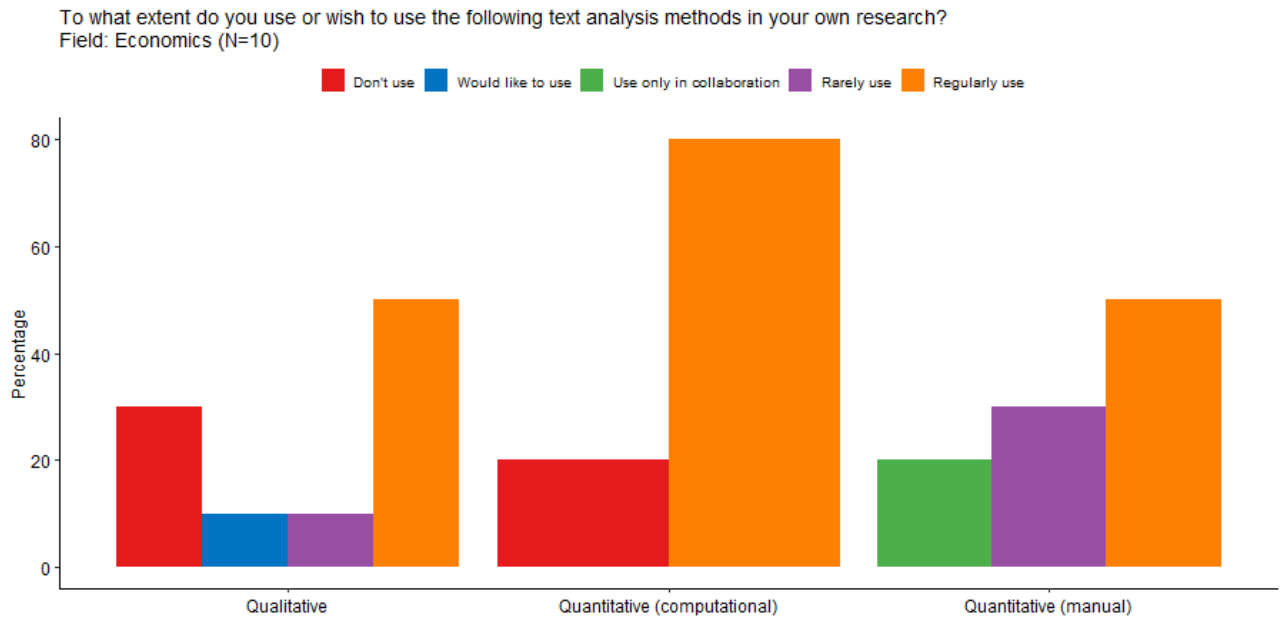


Figure A.10 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: ECONOMICS RESEARCH

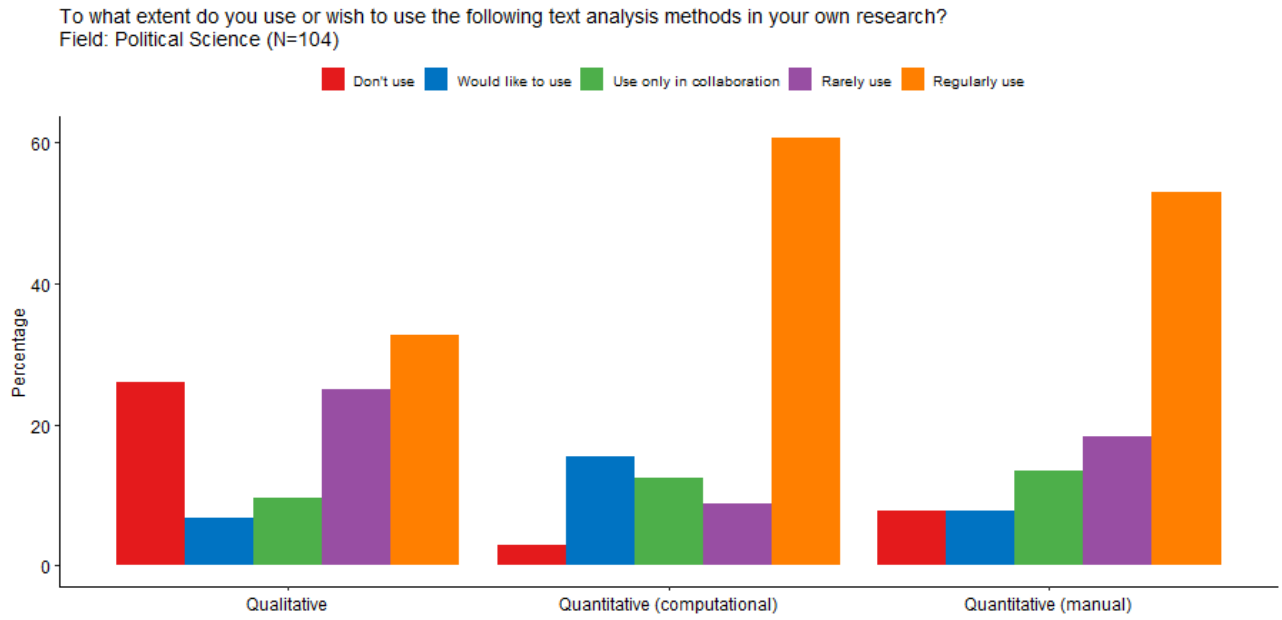
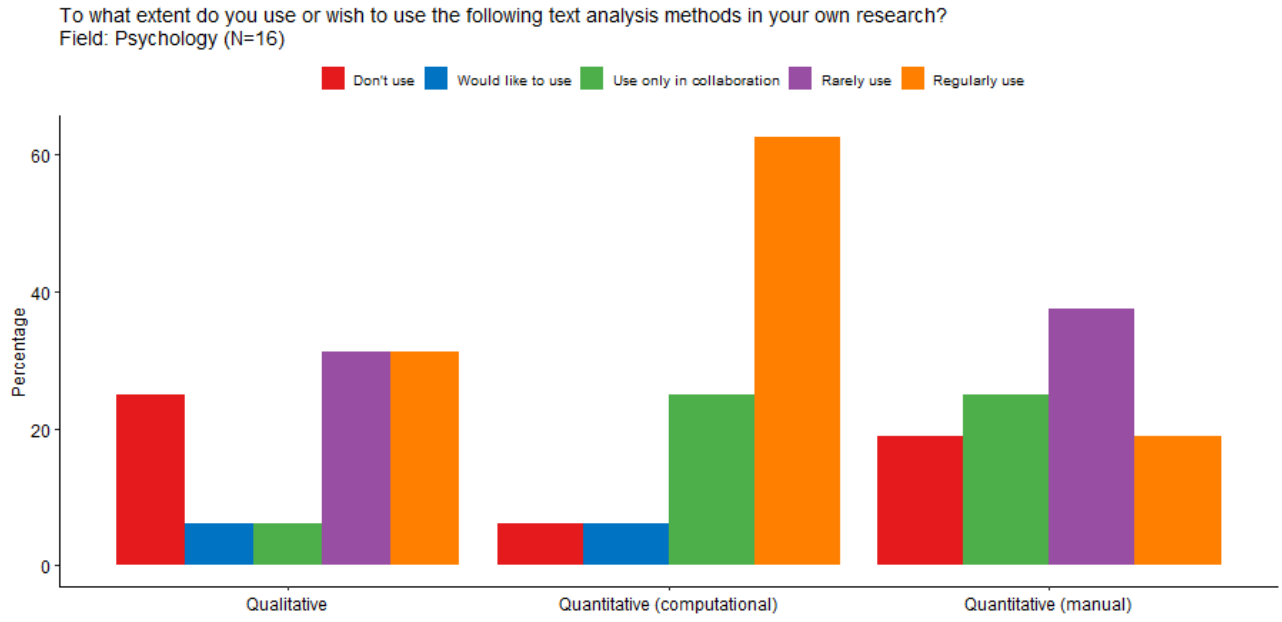


Figure A.10 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: POLITICS RESEARCH



,

Figure A.11 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: PSYCHOLOGY RESEARCH

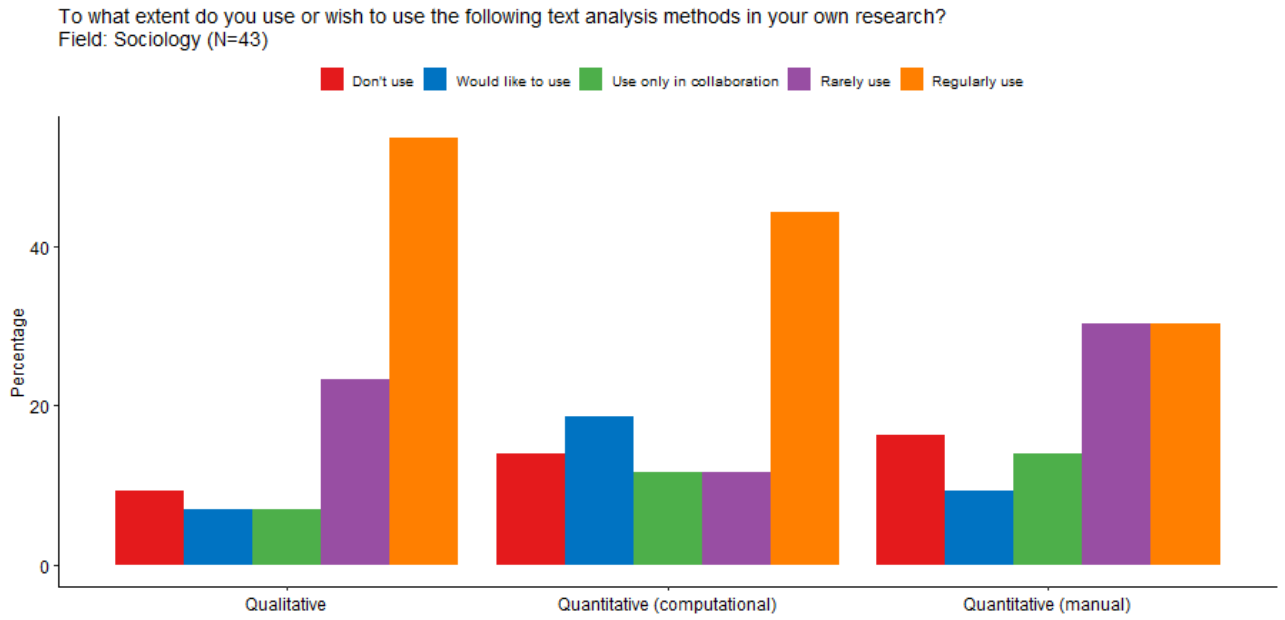


Figure A.12 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: SOCIOLOGY RESEARCH

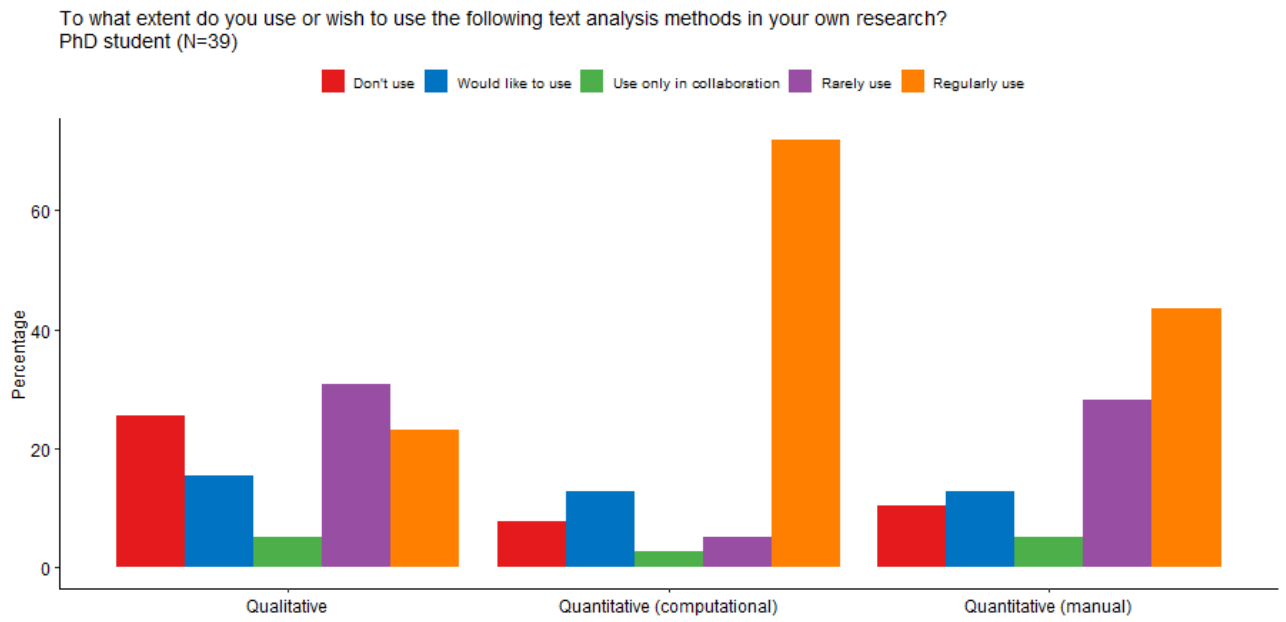


Figure A.13 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: PHD STUDENTS

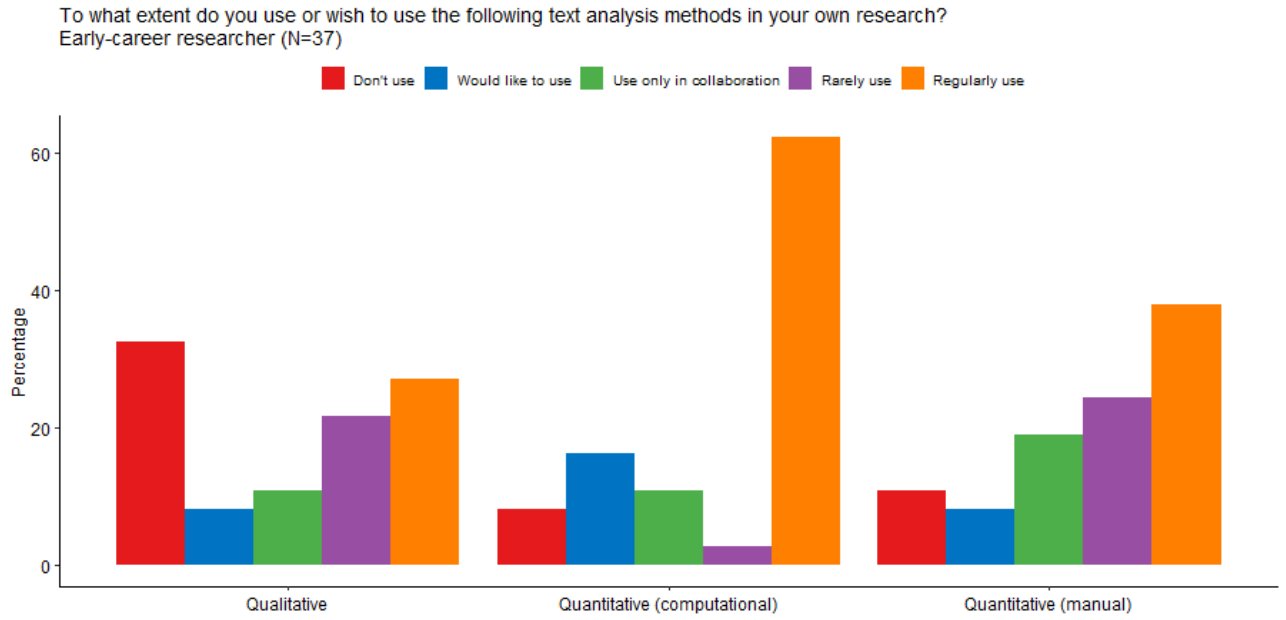


Figure A.14 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: EARLY-CAREER RESEARCHERS

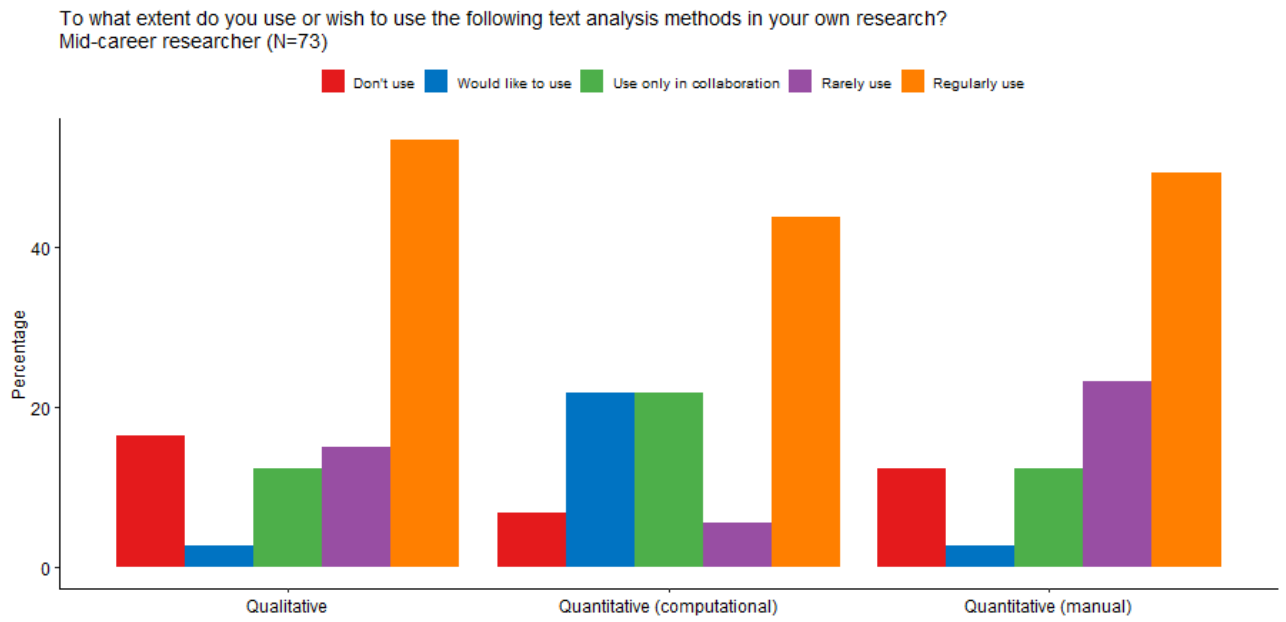


Figure A.15 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: MID-CAREER RESEARCHERS

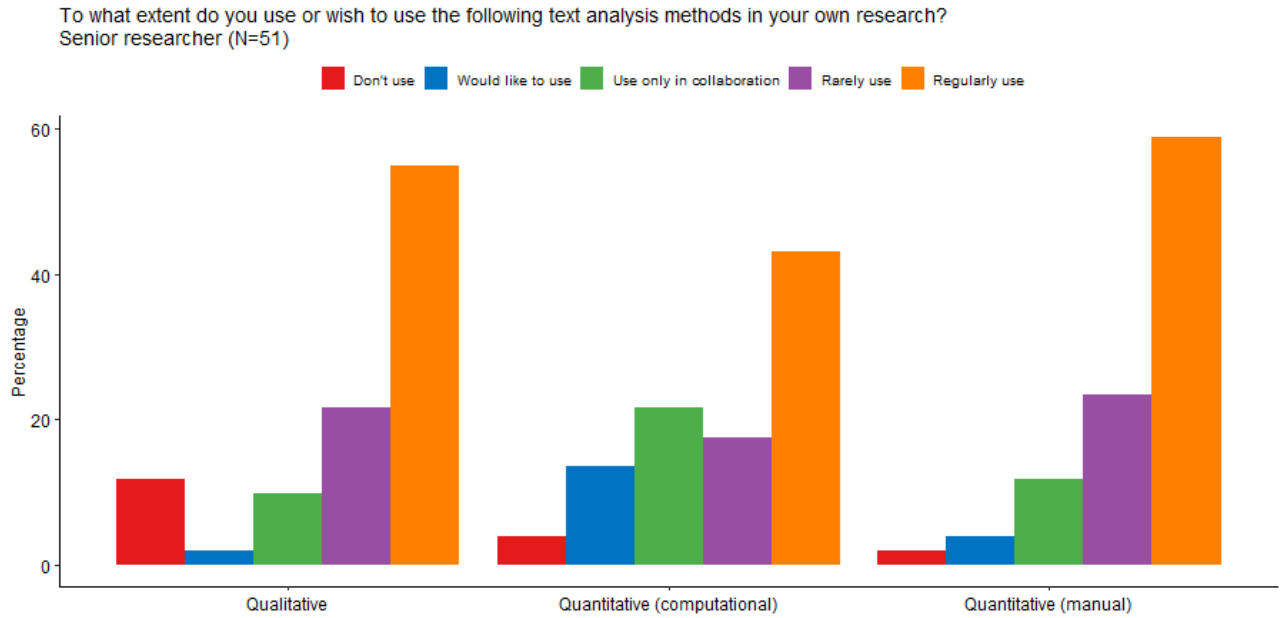


Figure A.16 USAGE OF DIFFERENT TEXT ANALYSIS METHODS: SENIOR RESEARCHERS

nocomp1
What challenges were relevant (or might be relevant in the future) for your choice to not use computational methods for text analysis?

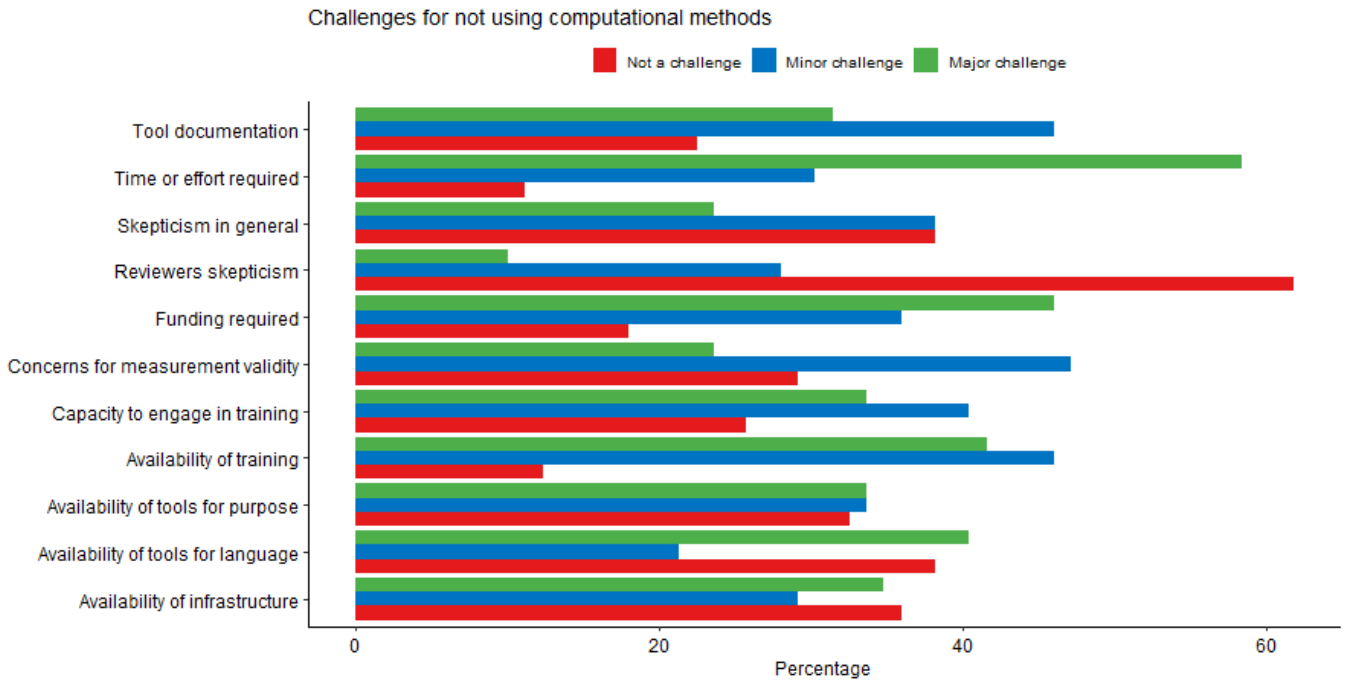


Figure A.17 CHALLENGES REVELANT FOR THE DECISION OF NOT TO USE COMPUTATIONAL METHODS

nocomp2
To what extent was the required time and effort challenging because of the following?

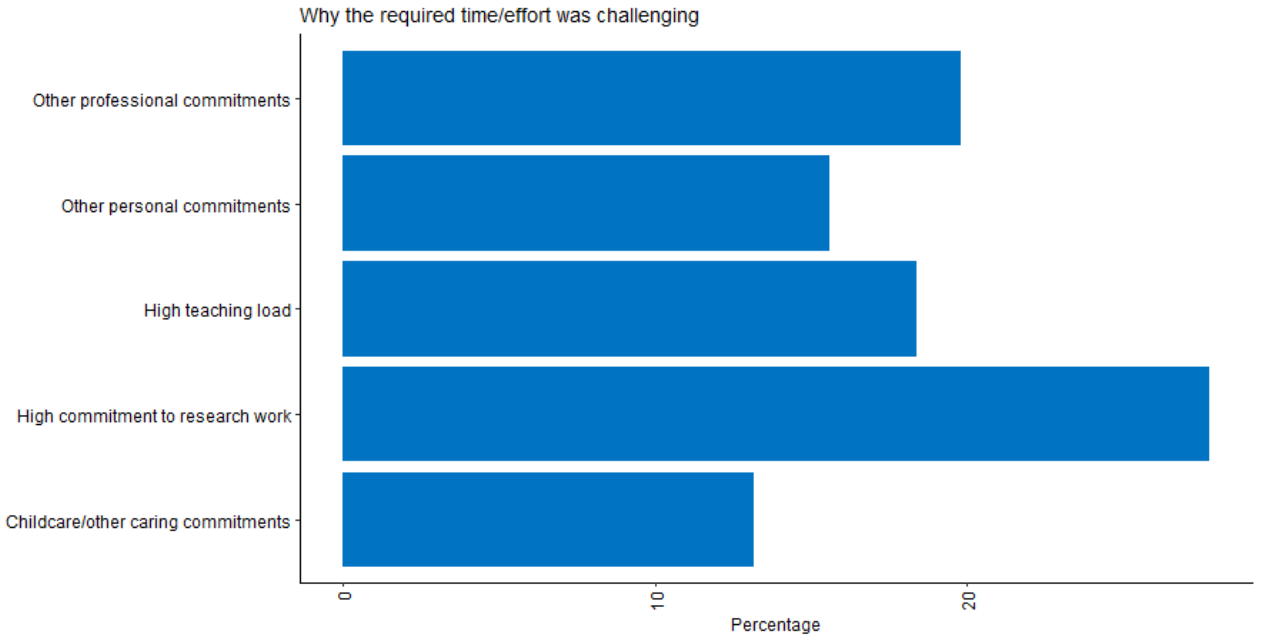


Figure A.18 WHY THE REQUIRED TIME OR EFFORT WAS CHALLENGING

comp3

In your experience, what challenges did you encounter (or you think you might encounter in the future) when using computational methods for text analysis?

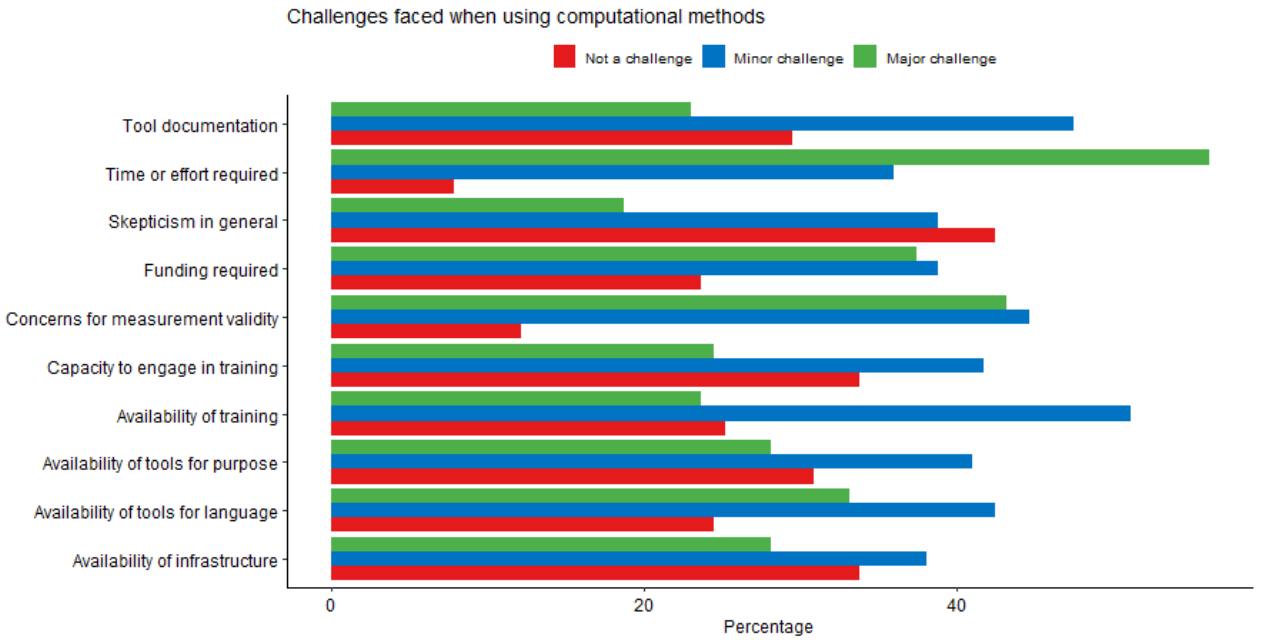


Figure A.19 CHALLENGES ENCOUNTERED WHEN USING COMPUTATIONAL METHODS

comp4

To what extent was the required time and effort challenging because of the following?

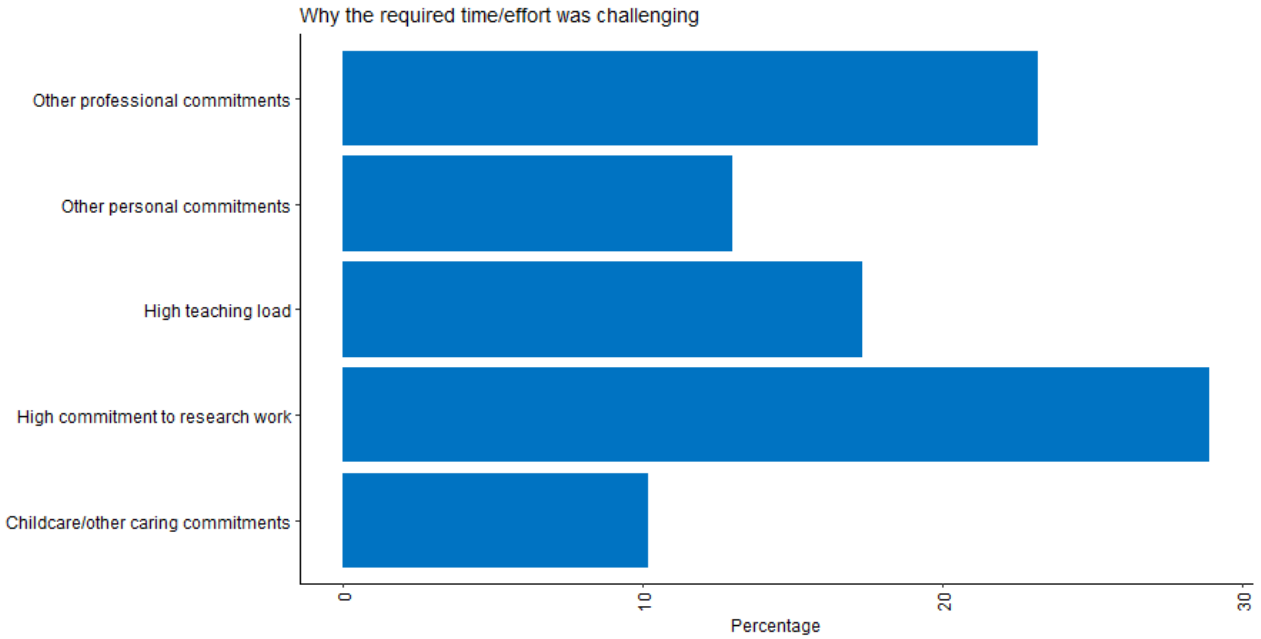


Figure A.20 WHY THE REQUIRED TIME OR EFFORT WAS CHALLENGING

comp6

Please indicate your agreement with the following statements:

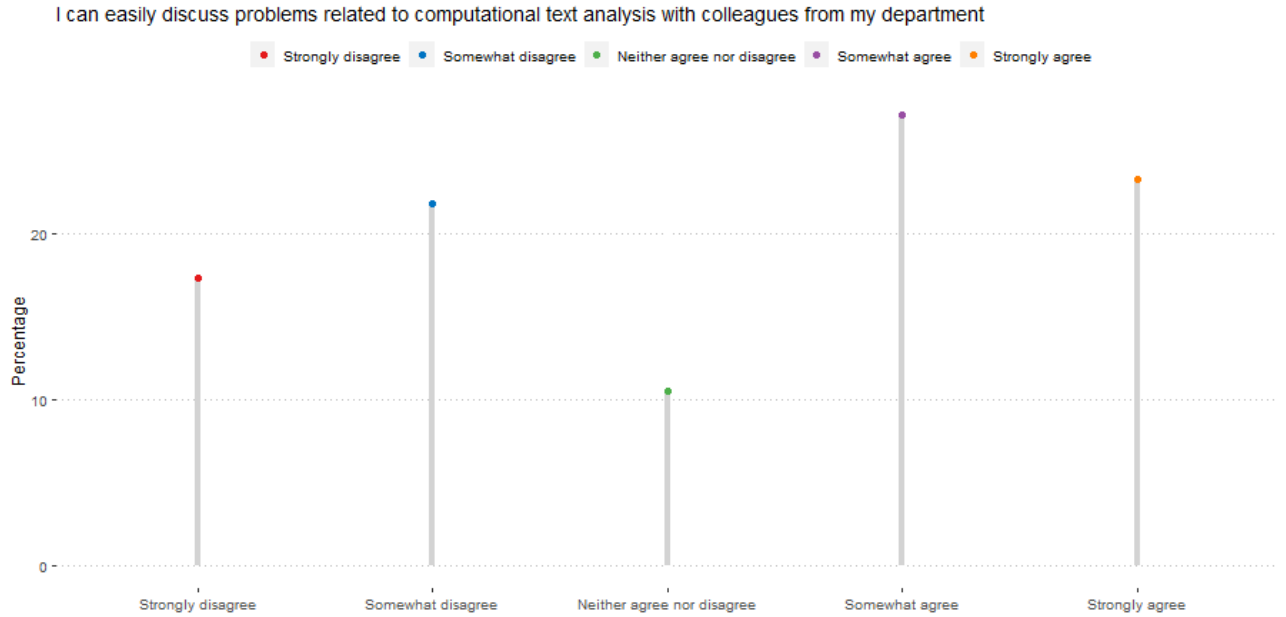


Figure A.21 STATEMENT: CAN DISCUSS PROBLEMS WITH COLLEAGUES IN DEPARTMENT

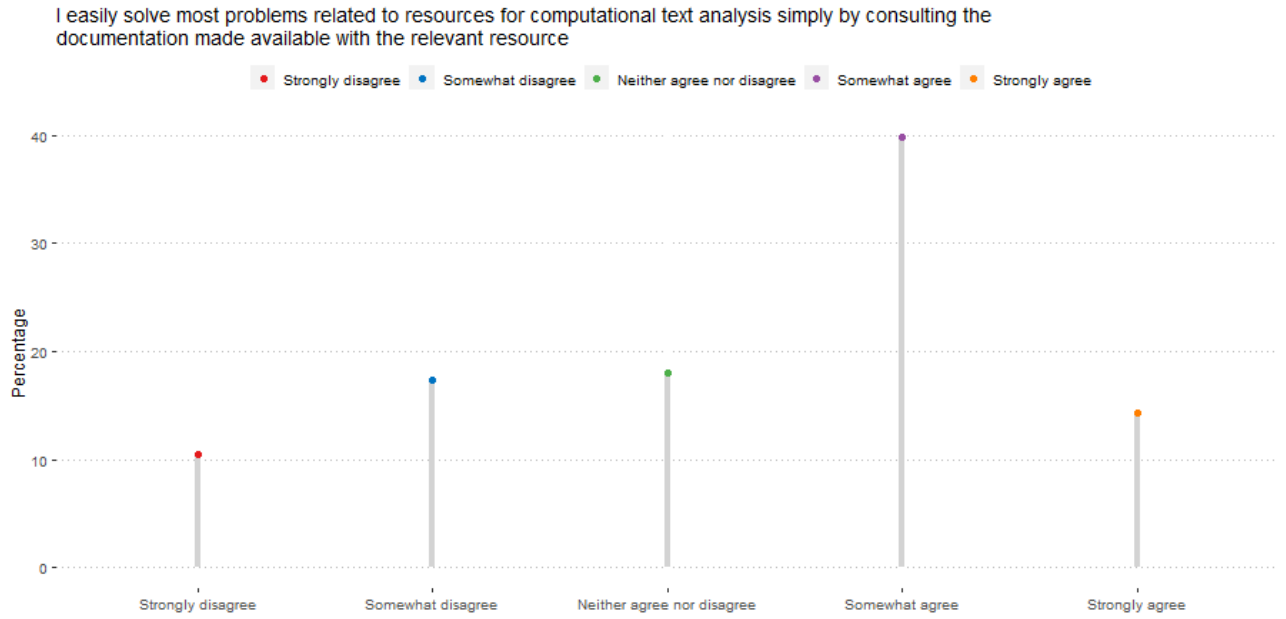


Figure A.22 STATEMENT: CAN SOLVE PROBLEMS USING DOCUMENTATION PROVIDED

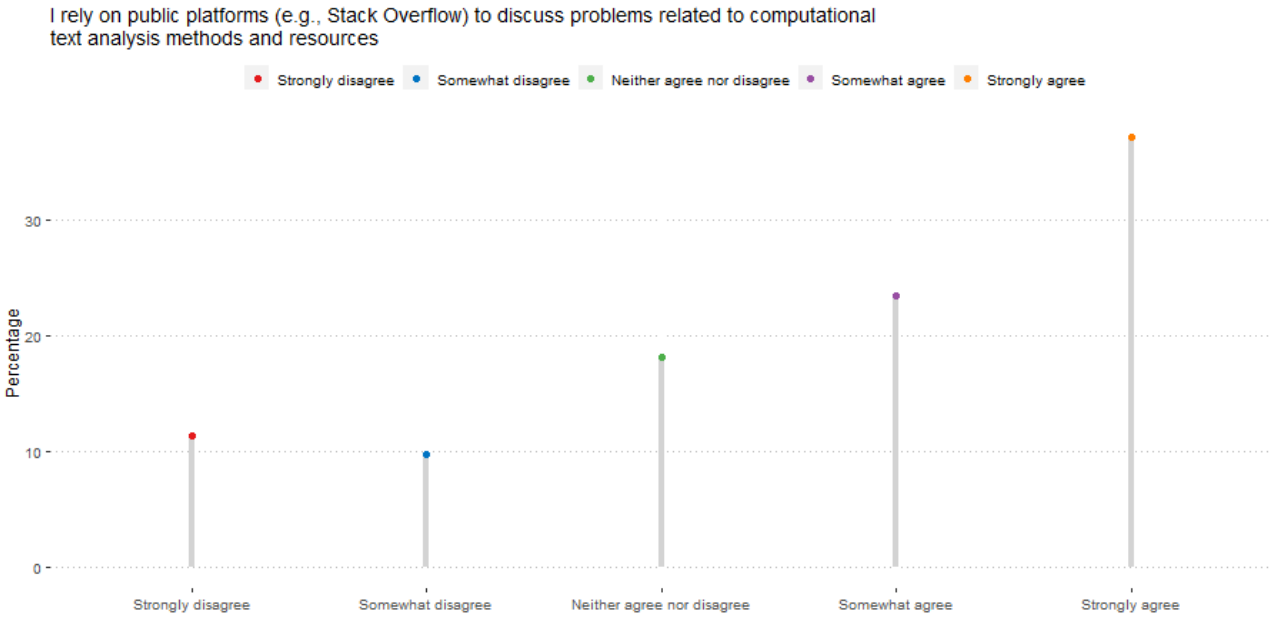


Figure A.23 STATEMENT: RELIANCE ON PUBLIC PLATFORMS TO DISCUSS PROBLEMS

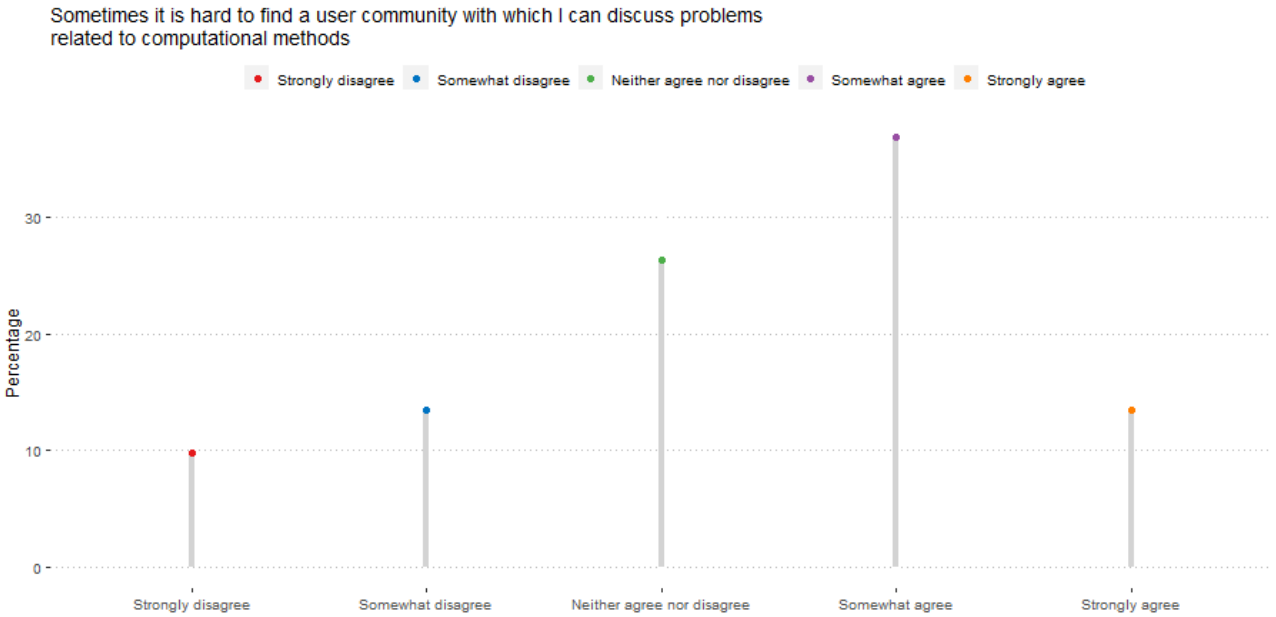


Figure A.24 STATEMENT: HARD TO FIND AN USER COMMUNITY TO DISCUSS WITH

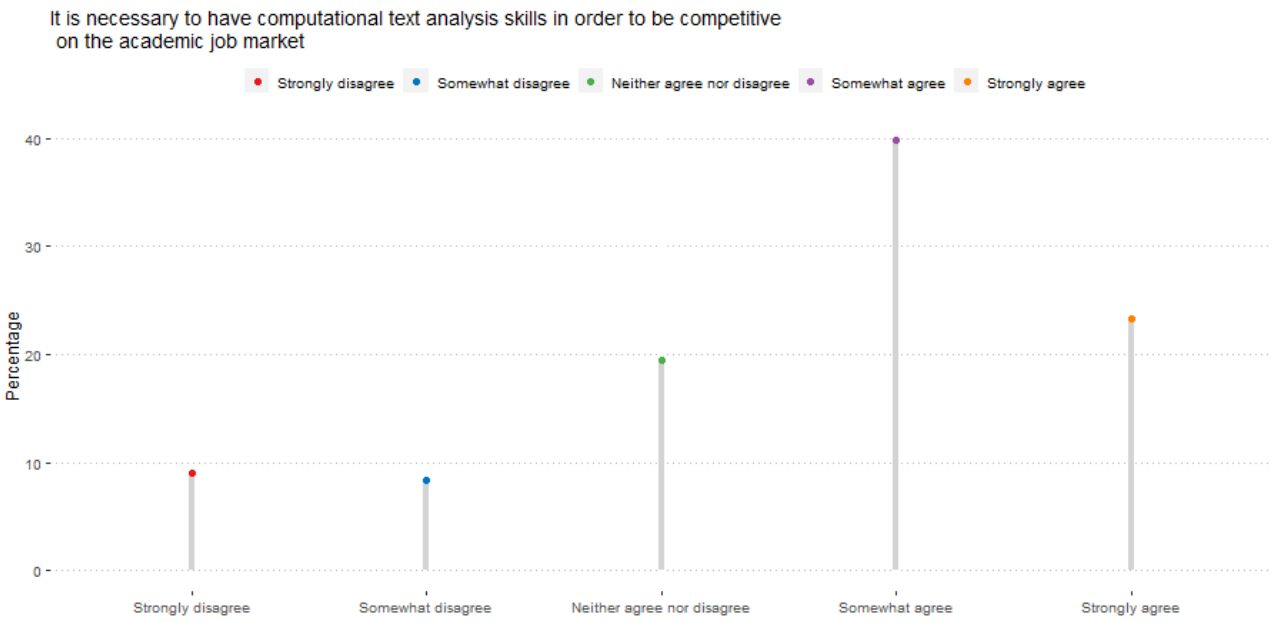


Figure A.25 STATEMENT: COMPUTATIONAL METHODS NECESSARY FOR BEING COMPETITIVE

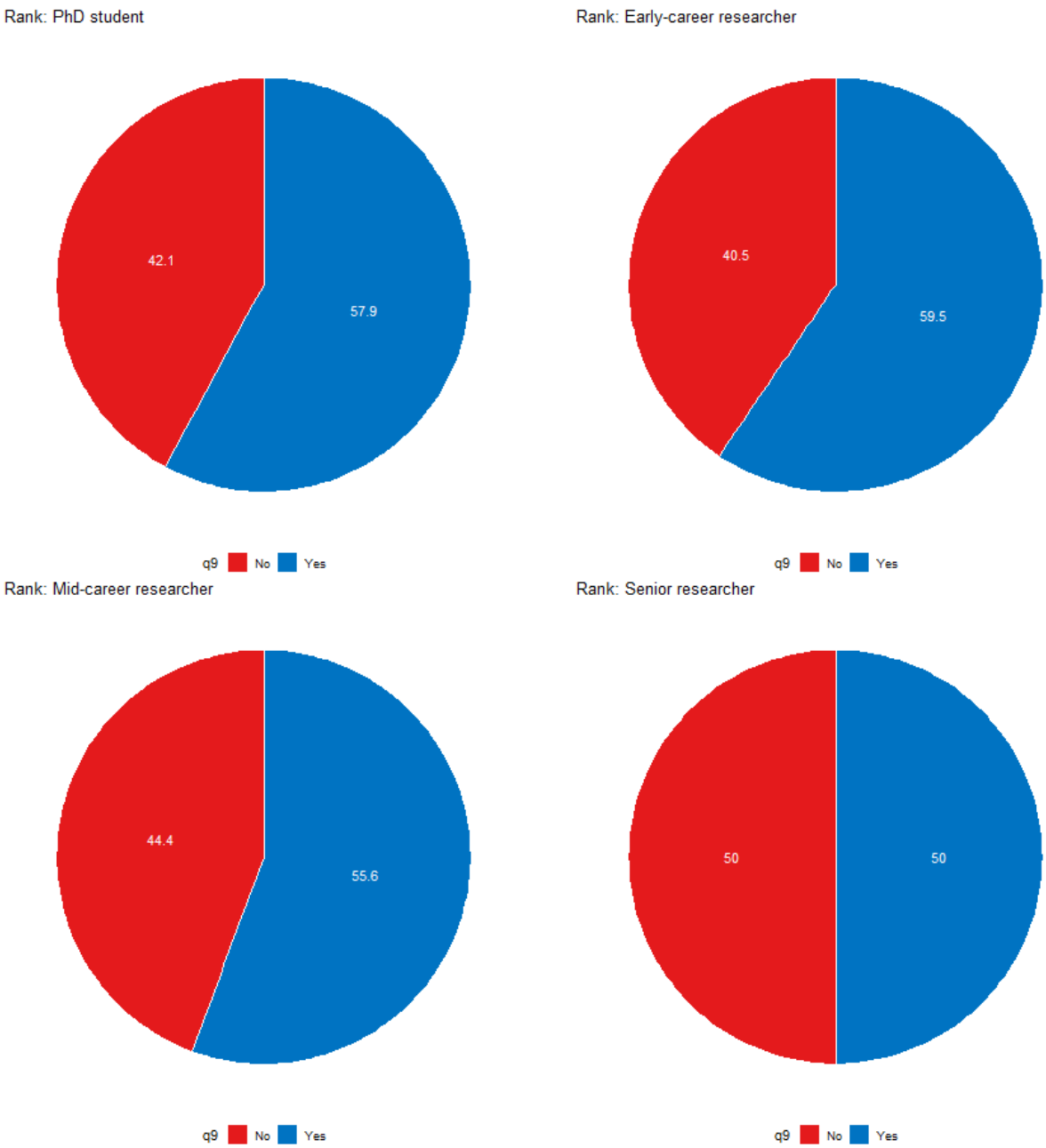


Figure A.26 ATTENDANCE TO TRAINING EVENT RELATED TO TEXT ANALYSIS

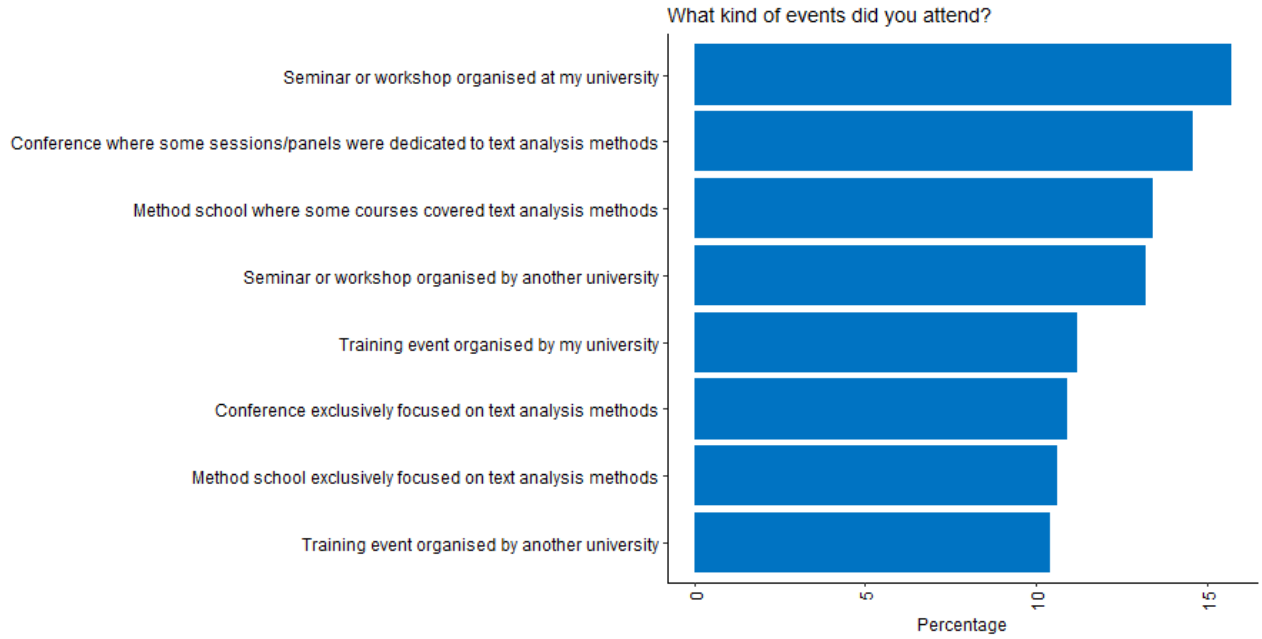


Figure A.27 TYPES OF TEXT ANALYSIS EVENTS ATTENDED

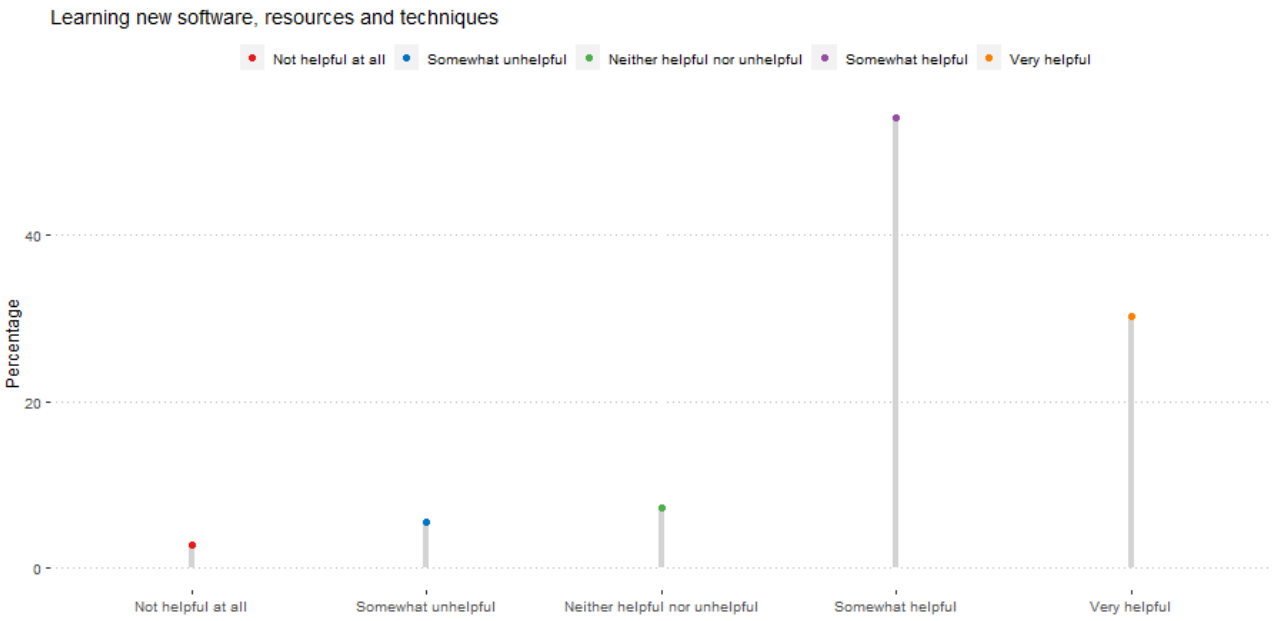


Figure A.28 STATEMENT: EVENT HELPFUL FOR LEARNING NEW RESOURCES

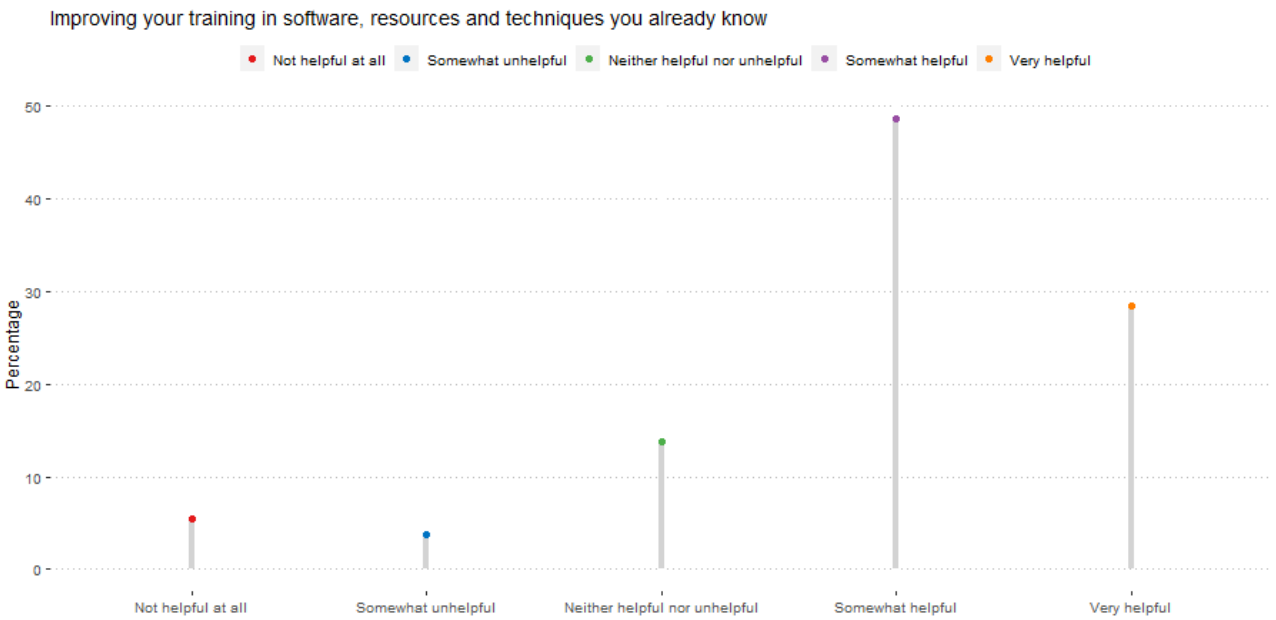


Figure A.29 STATEMENT: EVENT HELPFUL FOR IMPROVING KNOWLEDGE OF RESOURCES

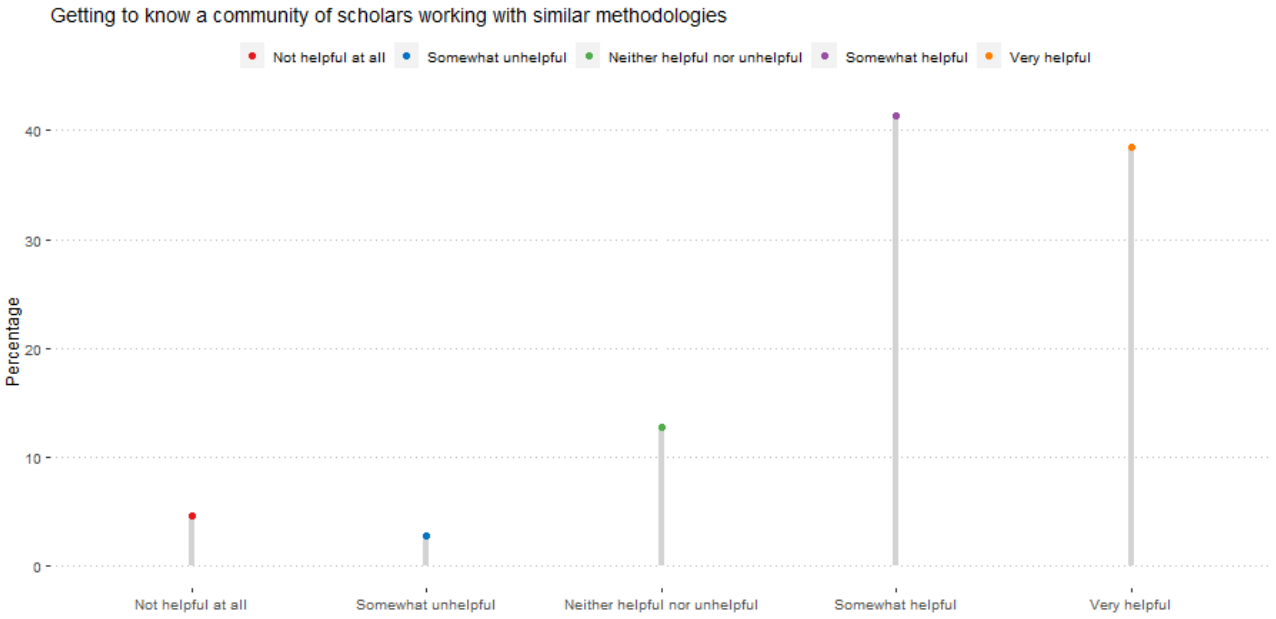


Figure A.30 STATEMENT: EVENT HELPFUL FOR KNOWING SCHOLARS WITH SAME INTERESTS

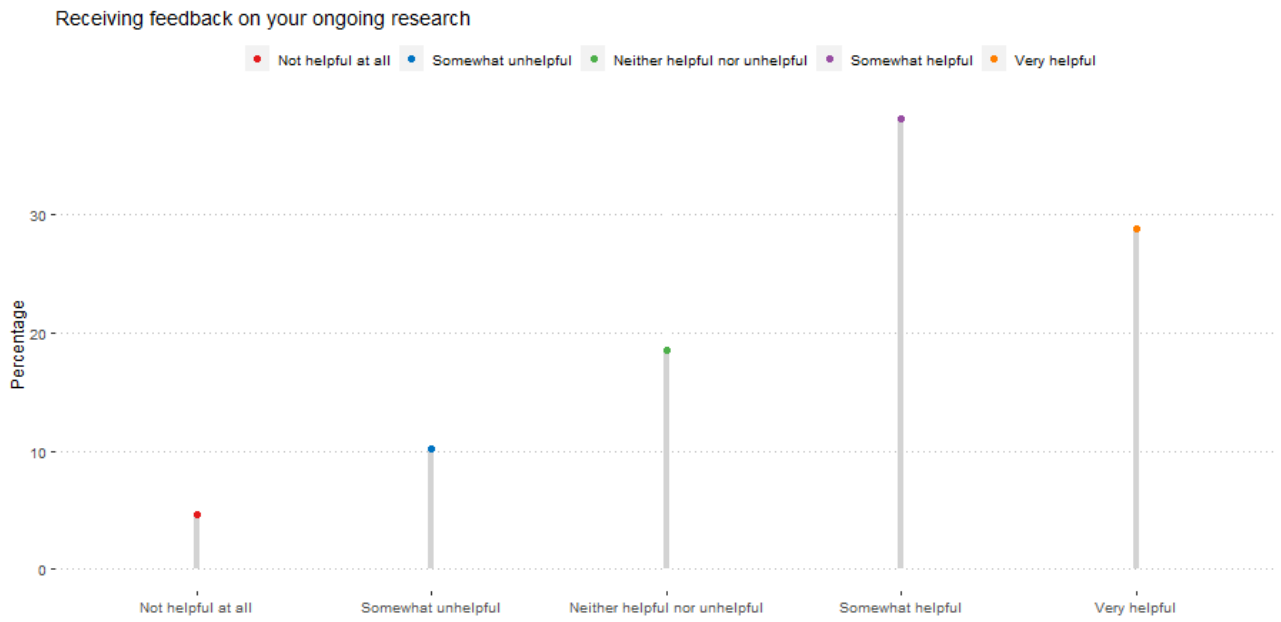


Figure A.31 STATEMENT: EVENT HELPFUL FOR GETTING FEEDBACK ON ONE'S WORK

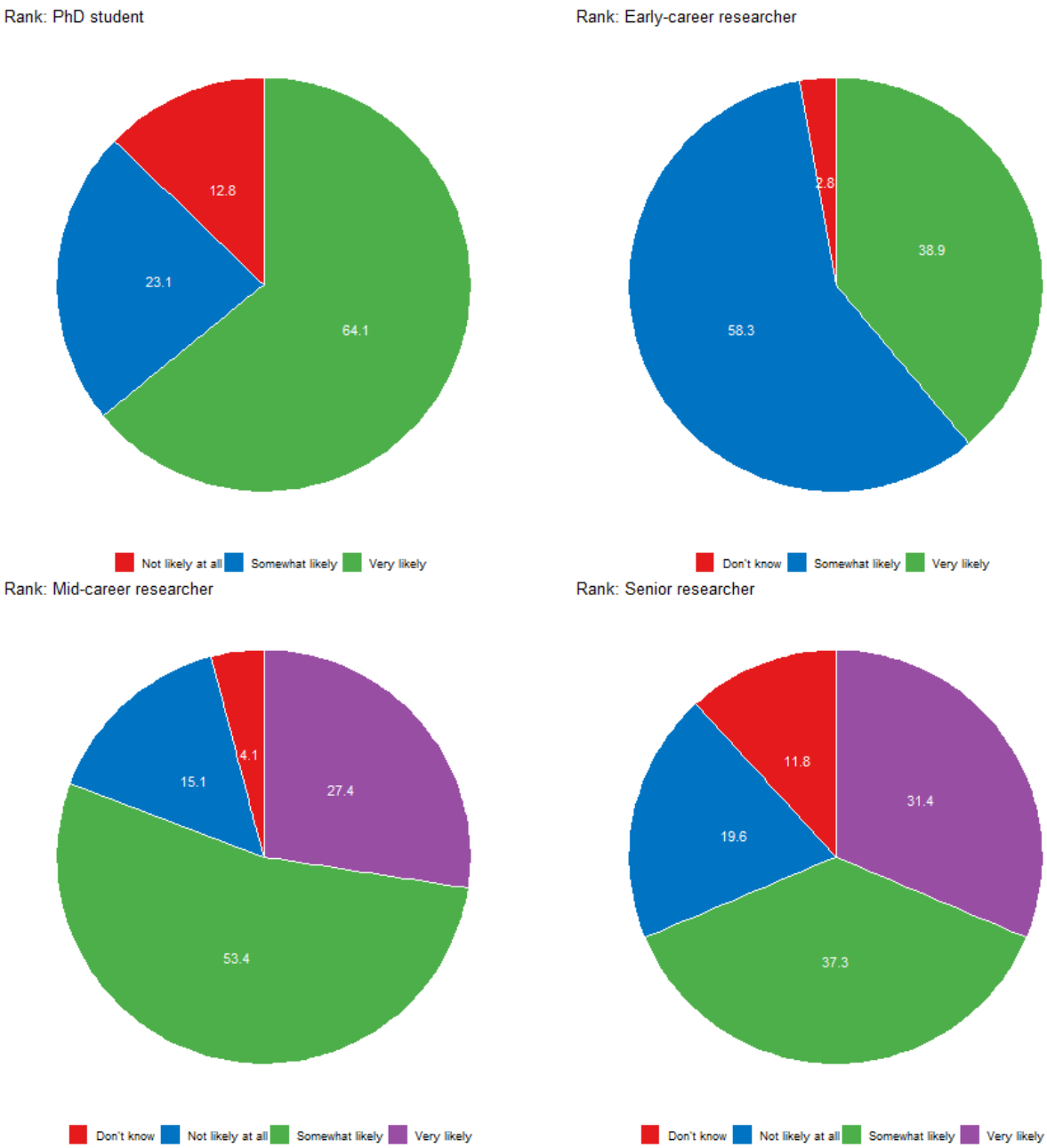


Figure A.32 LIKELIHOOD OF ATTENDANCE TO TRAINING EVENT RELATED TO TEXT ANALYSIS

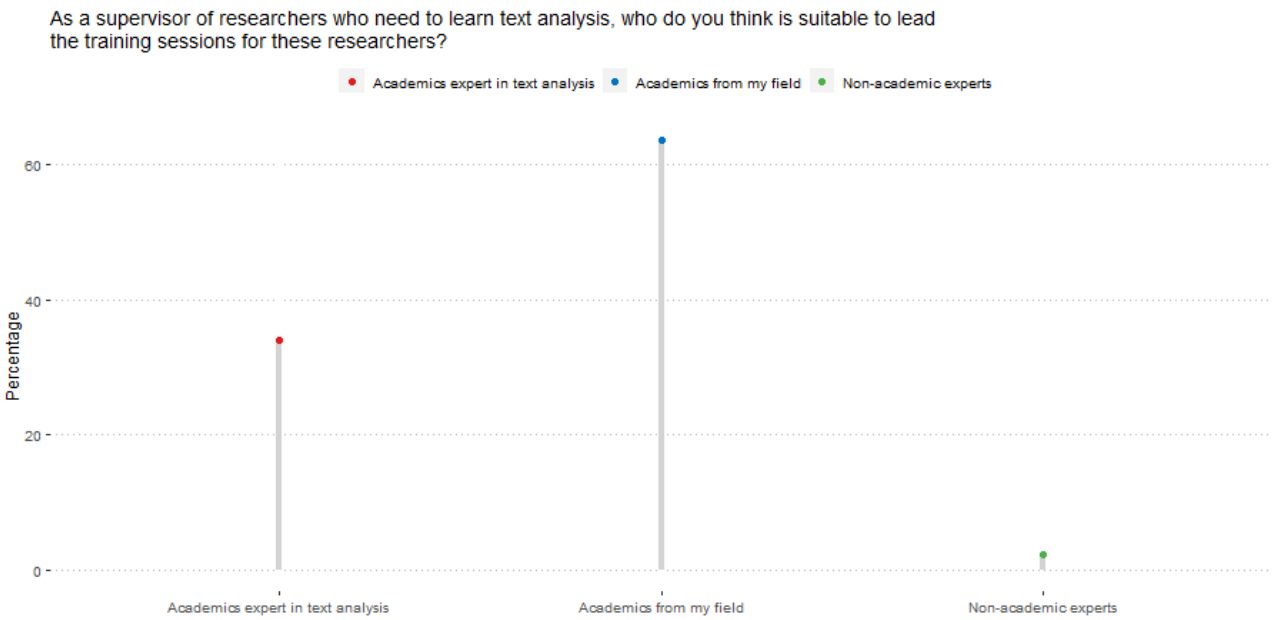


Figure A.33 MOST SUITABLE INSTRUCTOR FOR RESEARCHERS LEARNING TEXT ANALYSIS

q13a

Do you believe that, in the context of your training activities for students and social scientists, there are important training needs in relation to text analysis in any of the following areas?

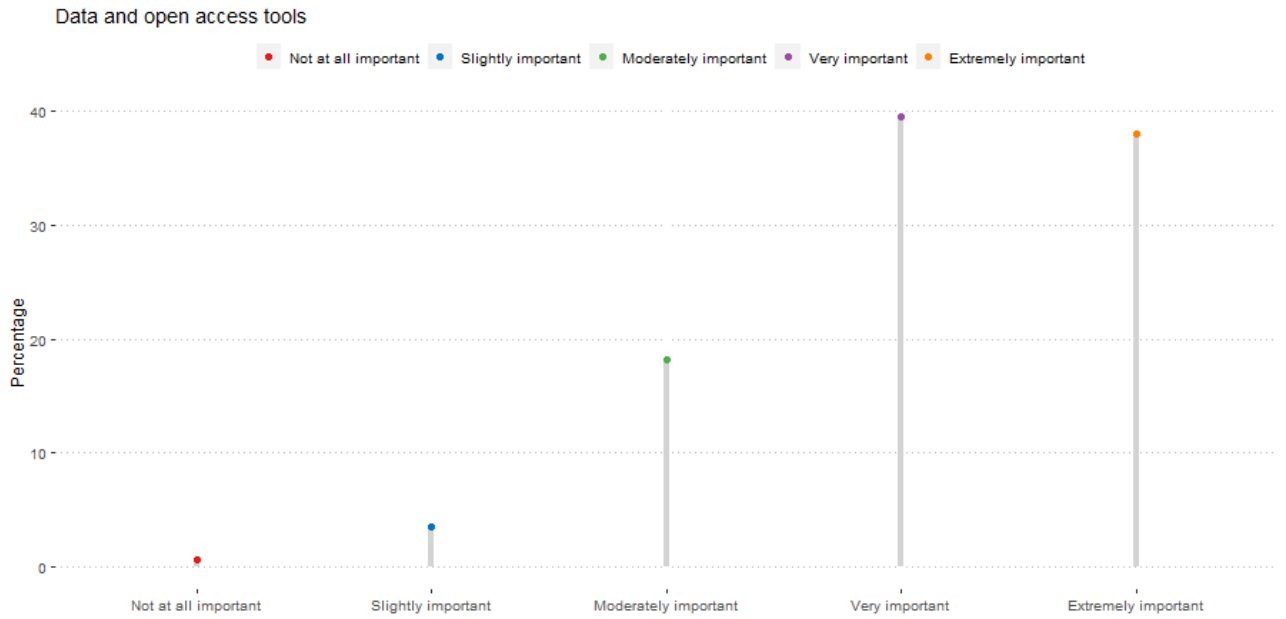


Figure A.34 IMPORTANT TRAINING NEED: DATA AND OPEN ACCESS TOOLS

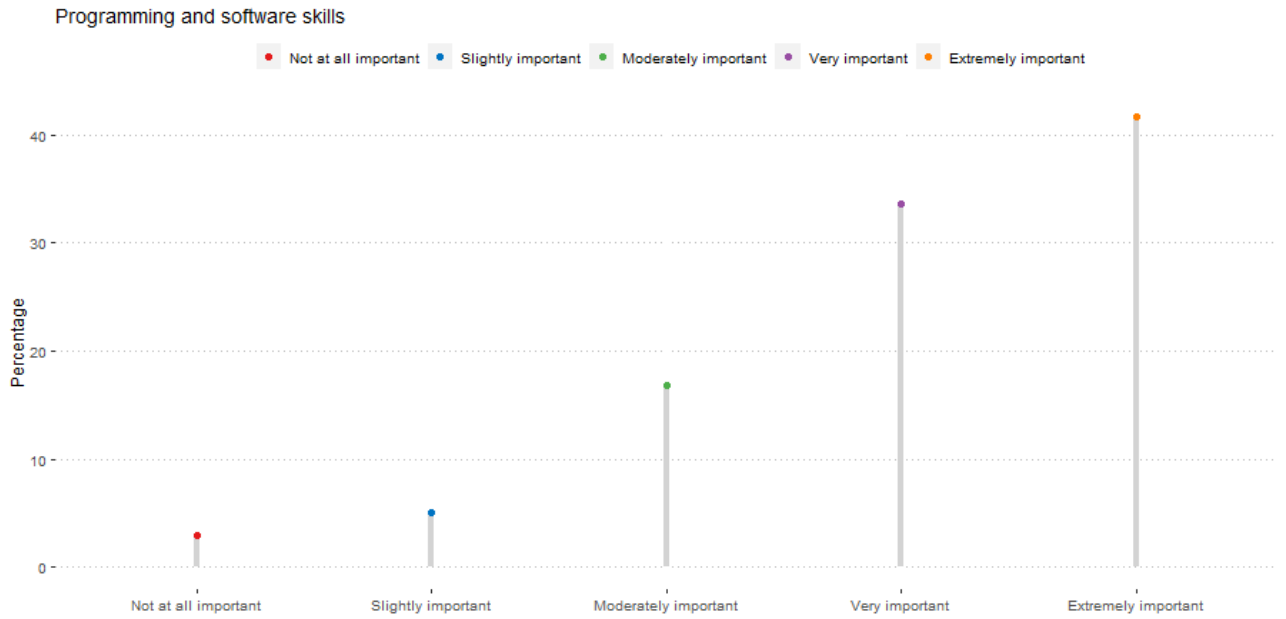


Figure A.35 IMPORTANT TRAINING NEED: PROGRAMMING AND SOFTWARE SKILLS

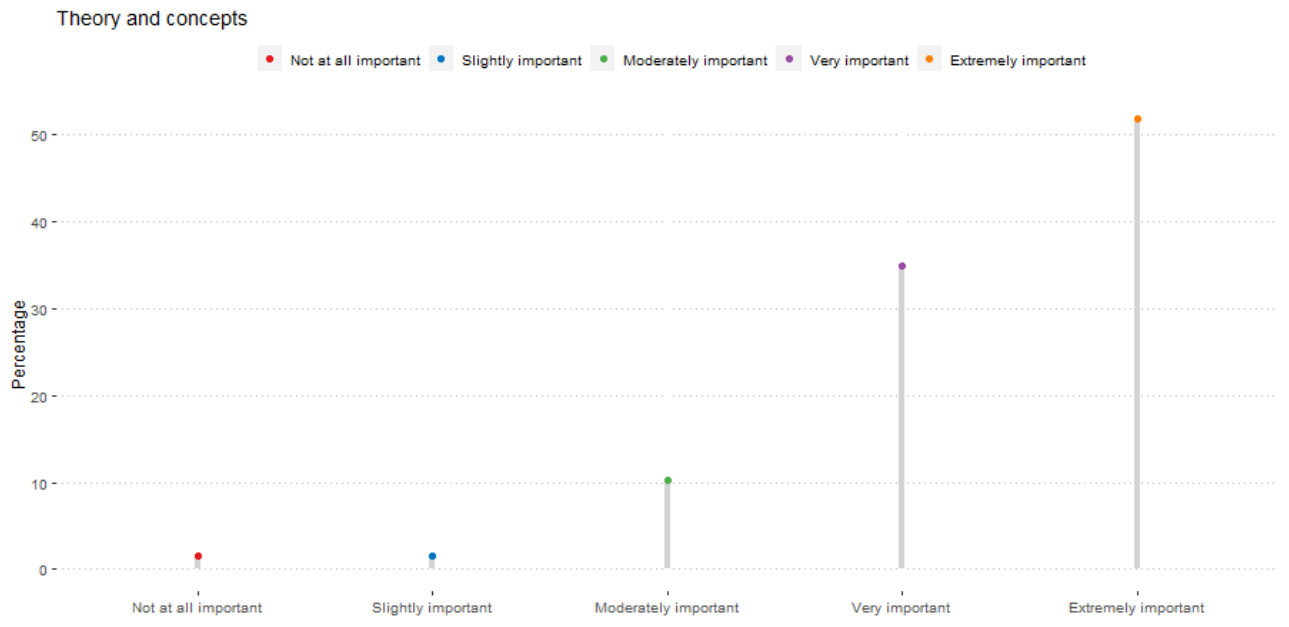


Figure A.36 IMPORTANT TRAINING NEED: THEORY AND CONCEPTS

Figure A.37 IMPORTANT TRAINING NEED: RESEARCH INTEGRITY AND ETHICS