September

2022

A CASE STUDY IN PHILANTHROPY:

Piloting the use of natural language processing to understand funding decisions

Jonathan Scaccia, PhD
Victoria C. Scott, PhD, MBA

DAWN CHORUS GROUP

EXECUTIVE SUMMARY

The grant making process is essential for philanthropic organizations to build community capacity for health improvement. However, little is known regarding what application attributes influence funding decisions. The current study reviewed retrospective grant narratives to examine associations between applicant characteristics and funding decisions. Specifically, we explore the following research questions:

What trends exist in funding decisions over time?
What textual insights can be uncovered from applications?
What application attributes are predictive of positive funding?

In collaboration with a philanthropy organization's (deidentified) project staff, we identified the central variables of interest and assembled a philanthropy organization Creative Ideas project applicant database. Then, we employed natural language processing (NLP), which uses automatic computational processing to extract meaning in text-based data, to analyze over 4,000 grantee applications.

This study yielded three major insights. First, funding outcomes were not predicted by the amount of requested funding, location, narrative length, or readability. Second, technology-based words, “research,” and the terms that aligned with the stated mission of the Creative Ideas portfolio were most likely to be funded. Lastly, NLP is a powerful synthesis method for examining grant making when used in conjunction with human expertise. NLP methods become more robust as grant text volume increases.

We believe NLP is the wave of the future for philanthropy organizations. Insights from rapid synthesis of the data can be used by philanthropy organizations to understand grant making trends and to increase organizational transparency about funding decisions. This project illuminated trends in the Creative Ideas project portfolio that has instructive value for the philanthropy organization's grant making activities and stimulated new questions ripe for prospective examination. Additionally, it demonstrated the value and constraints of NLP methods for analyzing grant application data and associations with funding decisions.

TABLE OF CONTENTS

INTRODUCTION

Philanthropy aims to have the greatest impact with the most efficient use of resources. The grant making process is a vital mechanism through which philanthropic organizations build community capacity for health improvement. In this process, funders assess grant applications based on organizational characteristics (e.g., staff capacity, mission, longevity) and fit with the call for proposals. Certain organizational conditions and capacities (e.g., positive organizational culture, high leadership engagement, presence of program champions) make it more likely that a program will be successfully implemented in a target setting (Scaccia et al., 2015). However, the grant application attributes that most influence grant making decisions are not commonly assessed retrospectively.

Existing research points to important variations in grant disbursement that have implications for health equity across communities. For example, the New Jersey Health Initiatives and the Walter Rand Institute have identified particular geographic locations that are less likely to apply for funding, which correspond to areas with poorer health outcomes (Atkins et al., 2020). This type of trend perpetuates a *grant inequities paradox,* whereby organizations with greater capacity (i.e., better resourced) are able to apply for and secure funding. This enhances the organization's position for future funding opportunities while poorly resourced organizations miss out on funding opportunities, which can ultimately result in widening health inequities across communities. These disparities are unintended but plausible consequences of grant disbursement activities. A better understanding of trends in grant-funding decisions (e.g., specific organizational attributes that are associated with successful funding) can foster organizational self-awareness and equity-minded review of applicants.

PROJECT AIM

Our Creative Ideas project involved a retrospective review of grant narratives to examine associations between grant application characteristics and grantmaking decisions. We examined whether particular grant application characteristics can distinguish organizations that do and do not successfully obtain funding from philanthropic organizations through Natural Language Processing (NLP) methods. Specifically, we examined the following three (3) research questions:

1. **What trends exist in funding decisions over time?**
2. **What textual insights can be uncovered from applications?**
3. **What application attributes are predictive of positive funding?**

NLP is commonly employed in the medical and customer service fields but has not yet widely impacted social services. Thus, a secondary aim of this project involved the testing of NLP methods as an innovative methodology for efficiently processing large amounts of grant data. We explored the value and constraints of varying NLP methods for data analysis and sensemaking. We view this project as a demonstration to illustrate how natural language processing can be used to understand trends in grantmaking.

METHODS

Sample

The philanthropy organization Creative Ideas project grant mechanism "seeks proposals that are primed to influence health equity in the future." For this project, we were provided with access to 4,199 grantee applications in the Creative Ideas portfolio dated from 2013 to 2020. These represented all 50 states, the District of Columbia, Puerto Rico, the US Virgin Islands, and the Federated States of Micronesia. Approximately 3,611 organizations were represented.

Within these applications, we examined the following variables: application ID (for tracking), application title, the application narrative, amount requested, and submitting organizations. The application narratives, averaged about 995 words, with a standard deviation of 295 words (range 47-1501 words).

Procedure

With the philanthropy organization's project staff, we identified the central variables of interest and assembled a philanthropy organization Creative Ideas project applicant database. We then cleaned, organized, and quality checked the data. The process of data analysis, interpretation, and report development were supported by subject matter expert consultants through a formal project Advisory Panel. The Advisory Panel was composed of four racially and ethnically diverse individuals with expertise in NLP methodologies, ethical considerations, and grant making (see table below).

Name	Degree	Position
Jeehye Yun	BA, Computer Science	CEO, Redshred, AI and CNN software development
Ben Kinsella	PhD, Hispanic Linguistics	Program Manager, Tribe AI
Jason Timm	PhD, Linguistics	Research Assistant Process, University of New Mexico
Nimu Sidhu	MS, Biophysics	Solutions Architect, Deloitte

Data Analyses

Broadly, the majority of analyses involved NLP methods. NLP is a subfield of artificial intelligence that uses automatic computational processing to extract meaning and nuance in text-based data (i.e., words that people use). This approach employs statistical inference to automatically learn how to process text through the analysis of a large, general corpus and the application of these rules to unfamiliar input. By treating words and clustering of words as meaningful, NLP extracts concepts and relationships from texts more efficiently than humans are capable of.

A summary of the specific analyses used for each research question is described below. All analyses were conducted in R using functions from the following packages: *tidymodels, textrecipes, quanteda, keras, tokenizers.bpe,*and *LIME.*

Research Questions 1. We examined trends in funding decisions through descriptive statistics. Specifically, we calculated frequency counts in relation to funding decisions, requested amounts of funding, major sources of applicants at the organizational and state level, and funding distribution by year.

Research Question 2. Textual insights in grant narratives were examined primarily using two NLP methods: Bag of words and latent Dirichlet allocation.

Bag of Words. The core of NLP is tokenization, which treats the words as “tokens” that are meaningful information units in and of themselves. Generally, the more frequently that a token occurs, the more relevant it is in descriptive analysis. The *Bag of Words* approach is a way of extracting features from text to describe the occurrence of words within a dataset. This approach treats each token as meaningful in and of itself and ignores any information about the structure of words in the document. A token is either a single word (unigram) or a sequence of words (e.g., bigrams, trigrams). The more frequently a token occurs, the more relevant it is to the descriptive analysis. To prepare the data for this analytic method, we completed a series of preprocessing steps, including the removal of stop words (i.e., words that add insignificant value to the overall text ; e.g., “the”, “with”, “any”) and lemmatization (i.e., converting words down to their root form).

Latent Dirichlet Allocation (LDA). Topic modeling is a method that automatically uncovers and extracts main ideas from a collection of documents, such as grant proposal narratives. We applied the Latent Dirichlet Allocation (LDA) topic modeling approach. LDA assumes that each topic is a cluster of words that co-occur together, and that documents (i.e., proposal narratives) are clusters of topics. Topics are used to represent narrative content areas. We examined perplexity scores to determine the ideal number of topics (k), ultimately selecting a parameter of k=20.

Research Question 3. First, we used inferential statistics and topic modeling to assess associations between application attributes and funding decisions. Then, to examine what application attributes are predictive of positive funding, we applied eight (8) predictive models and one neural network model. The use of multiple models enables identification of the best model. See table below for the model components.

Model
Model I: Term Frequency + glmnet.
Model 2: Term Frequency + upsampling + glmnet.
Model 3 - tf w/ 2000 tokens + upsample + glmnet
Model 4 - tf + stem + upsample + glmnet.
Model 5 - tf + bigrams + upsample + glmnet.
Model 6 - tf + uni+bigrams + upsample + glmnet
Model 7 - tf + tokenizers.bpe + upsample + glmnet
Model 8 - tfidf + scaling + uni+bigrams + upsample + glmnet
Model 9 - Neural Network- LSTM

To ensure a generalizable model, we developed two data sets: a training data set (based on a random selection of 70% of the available data set) and a test data set (based on the remaining 30%). All algorithms were applied to both the training and test set to evaluate model performance. This is a standard approach used in applied machine learning toward building a model that can be used with other similar sets of data (e.g., other grant narrative portfolios). We evaluated each model using a matrix table (see Figure 1) of potential outcomes and the four metrics: accuracy, precision, recall, FI-score.

Figure I. Outcomes Matrix Table

		Predicted Outcome	
		Turned Application Down	**Proceed with Application**
True Outcome	**Turned Application Down**	True Negative	False Positive *Type I Error*
	Proceed with Application	False Negative *Type II error*	True Positive

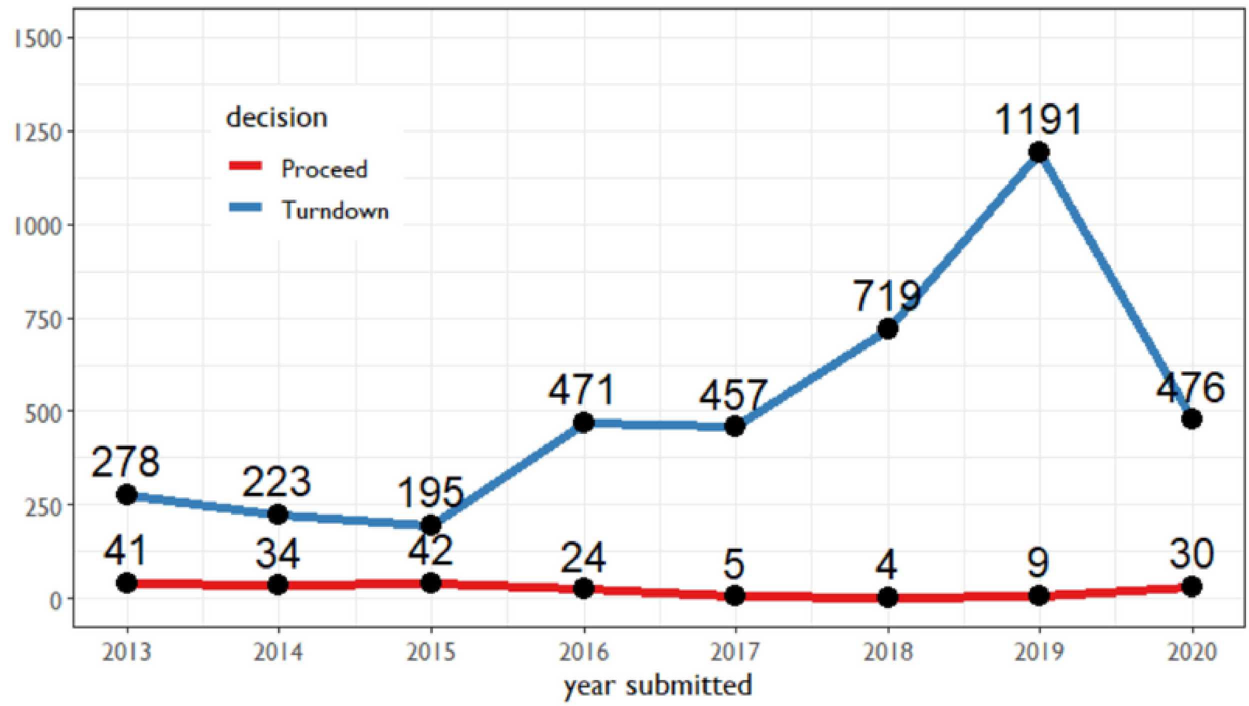
Cross-validation is a statistical method used to estimate the performance of machine learning models. It is a resampling procedure to evaluate a model on a limited dataset. Specifically, this approach estimates how the model is expected to perform in general when used to make predictions on data that were not used during the training of the model. For all models, we used 10-fold cross-validation. This means that we compute predictive statistics using ten different sample configurations, and then identify the average across all these implementations.

RESULTS

RQ 1. What trends exist in funding decisions over time?

The philanthropy organization's Creative Ideas project portfolio is a grant that has solicited submissions every year since 2013. Of 4,199 submissions, 189 were accepted for grant funding, while 4,010 were rejected between 2013 and 2020. The average acceptance rate was 5% and varied between 0.5% (2017) to 22% (2015). Figure 2 depicts the number of applications that were submitted and approved by year. This figure also previews one of the central challenges of this project– that is, successful funding is rare.

Figure 2. Number of Funded and Unfunded Creative Ideas Applications during 2013-2020



Note: Both the number of applications and the number of applications funded (see red line) for 2017 through 2019 appeared to be anomalous. We considered excluding these years and running our predictive analysis on 2013-2016 and 2020 separately. However, this would have resulted in cutting our data set roughly in half, significantly reducing an already small sample. We found that the amount that applicants requested (regardless of whether they received funding) was normally distributed on a logarithmic (increasing nonlinear) scale.

Figure 3 (next page) shows the number of states that submitted the most applications by year. Figure 4 depicts the distribution of applications by city. The greatest number of submissions generally reflects population trends (states with the most people) and proximity to the philanthropy organization (DC, New Jersey).

Figure 3. States Submitting the Most Creative Ideas Applications during 2013-2020

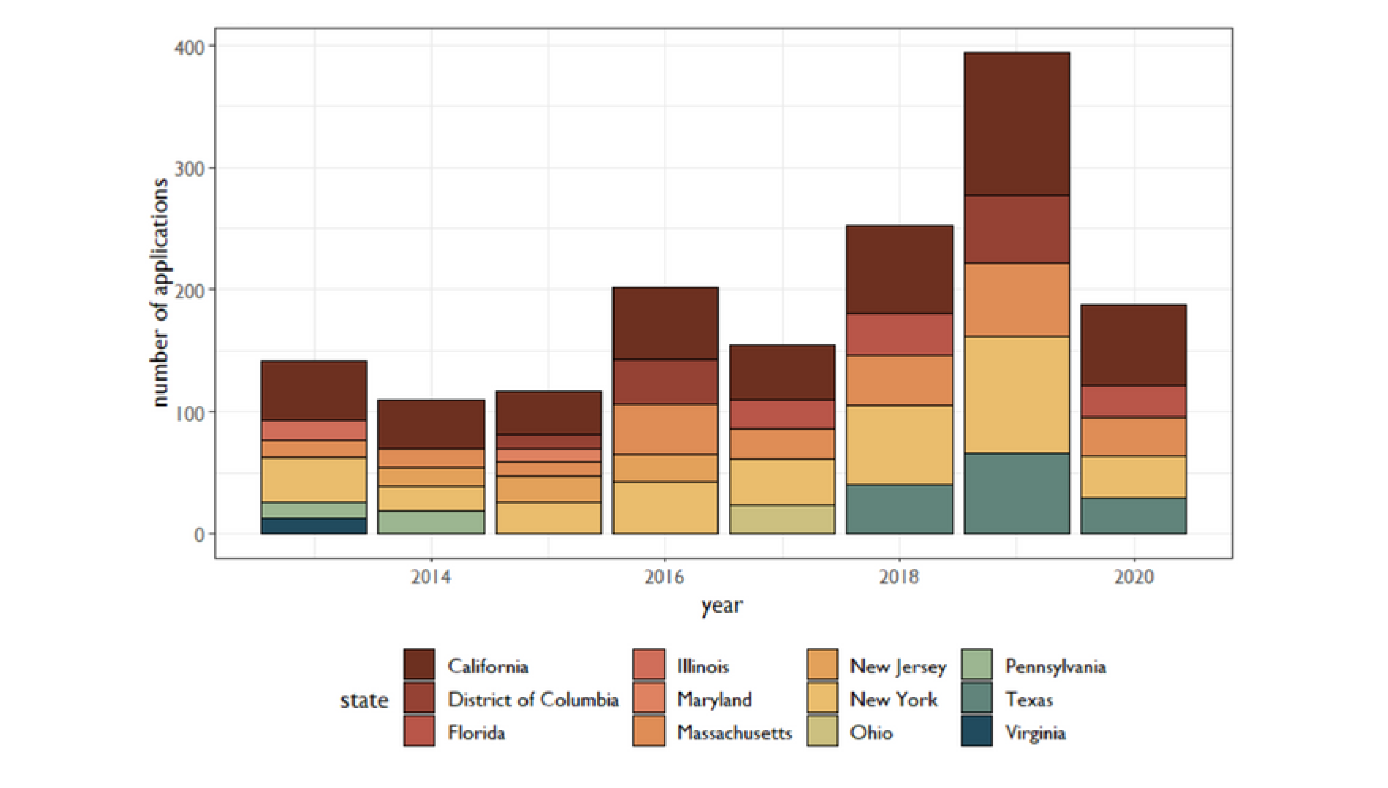
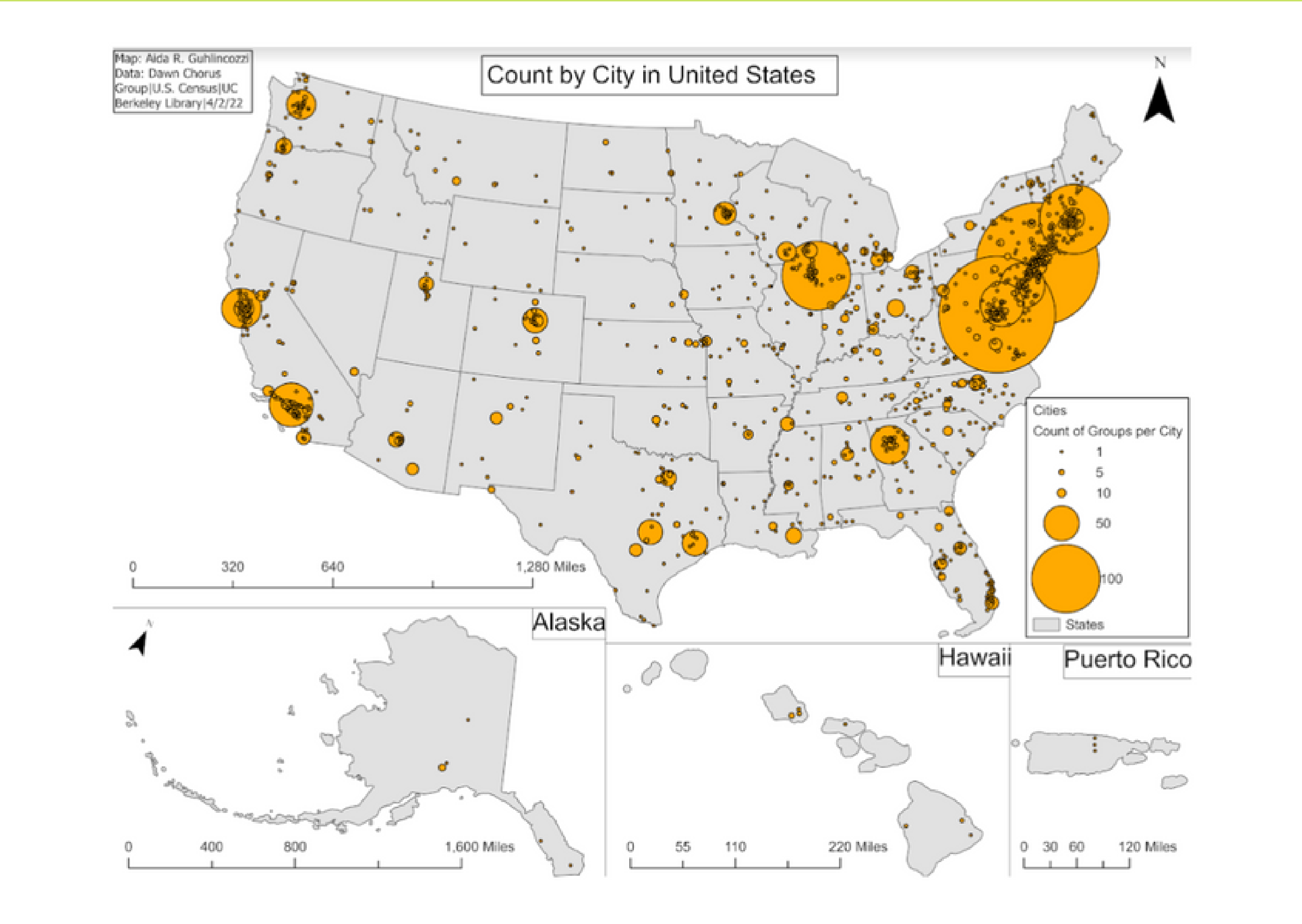
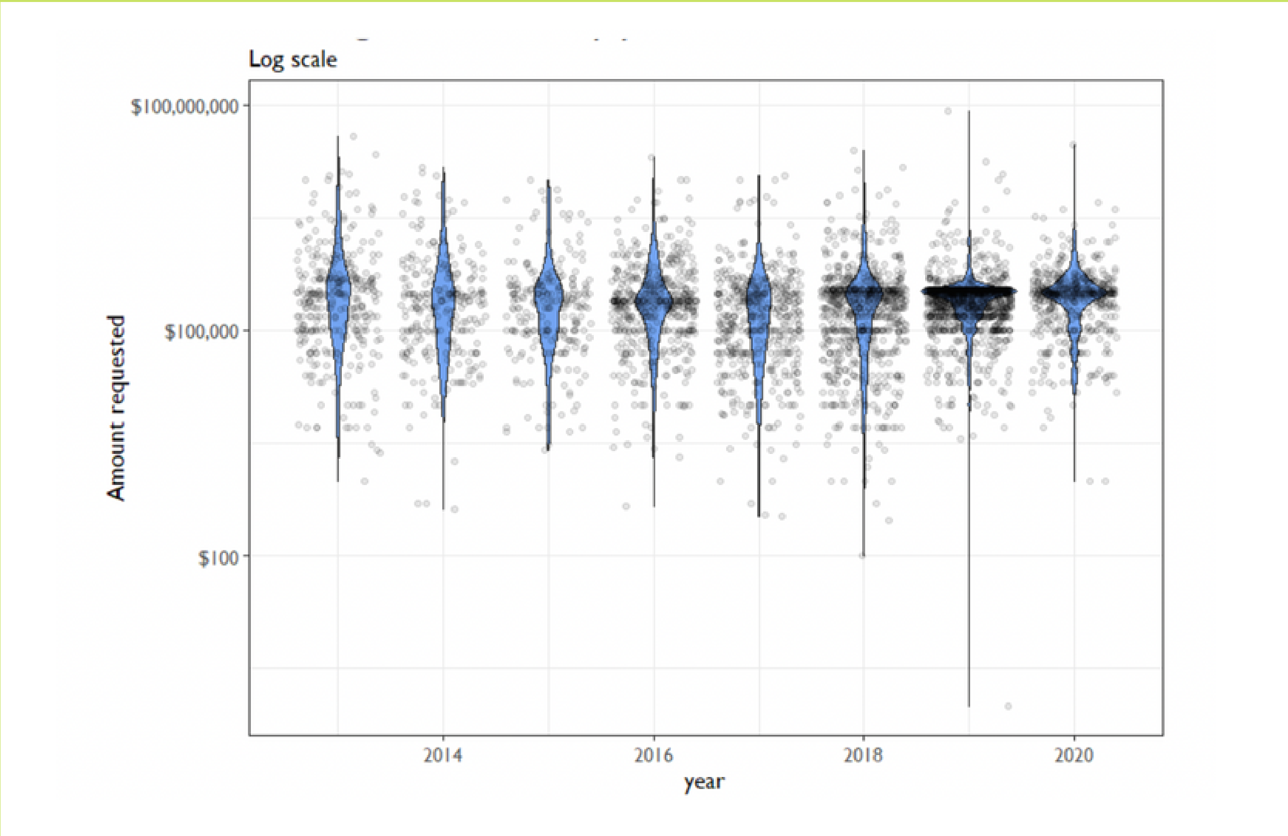


Figure 4. Distribution of Applications by City



Overall, 3,611 distinct organizations submitted applications. We found that universities are the major source of applications, with Johns Hopkins, the University of Washington, Emory, and the University of Pennsylvania all submitting ten or more applications over the past seven years. The average amount requested was $490,038 (standard deviation = $1,916.012), and ranged from $0 (five of them) to $84,693,252. Figure 5 shows a violin plot for funding by year, with each point representing an application. We used a log scale to account for the skewed distribution.

Figure 5. Funding Distribution by Year



RQ2. What textual insights can be uncovered from applications?

First, we examined the data for the most frequently occurring tokens (words or sequence of words). The term “health” appeared most frequently across narratives, followed by “community,” and “care,” which are consistent with the philanthropy organization’s general focus. Figures 6 and 7 present two depictions of the top 30 most common words reflected in the dataset. These words reflect the general nature of applications and point to broad themes for subsequent NLP analyses.

Figure 6. Frequency of Top 30 Most Commonly Reflected Terms

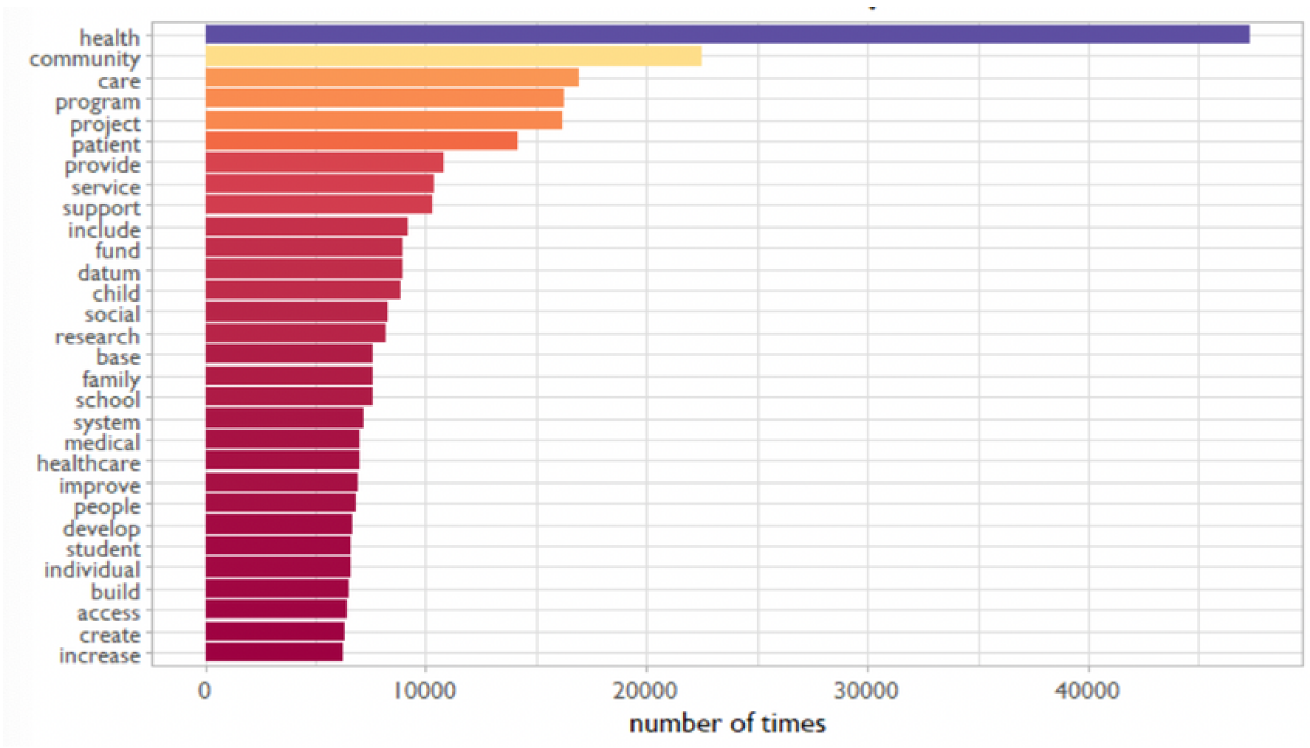


Figure 7. Term Frequency Word Cloud

Figure Note. The font size corresponds to frequency of appearance.

We examined the composition of content areas reflected in project narratives through the association of words within topics. Figure 8 depicts the top five words associated with each topic. We used five terms to avoid crowding the figure and to aid data visualization. We observed that topics reflect a range of community health issues such as health disparities (Topic 3), nutrition and healthy food programs (Topic 6), experiences of trauma, stress and violence (Topic 10), mental health and substance use (Topic 13), cancer (Topic 15) and (dis-)ability (Topic 20). Additionally, topic clusters reflected diverse settings (e.g. Topic - schools, Topic 16- healthcare, Topic 2- general public). We also observed topic clusters associated with “administrative aspects” of community health improvement, namely information technology (Topic 7) and issues of project/foundation costs/fund(ing) (Topic 11). Some topics, notably Topic 9, are not readily interpretable.

Figure 8. Key Terms for Each Topic

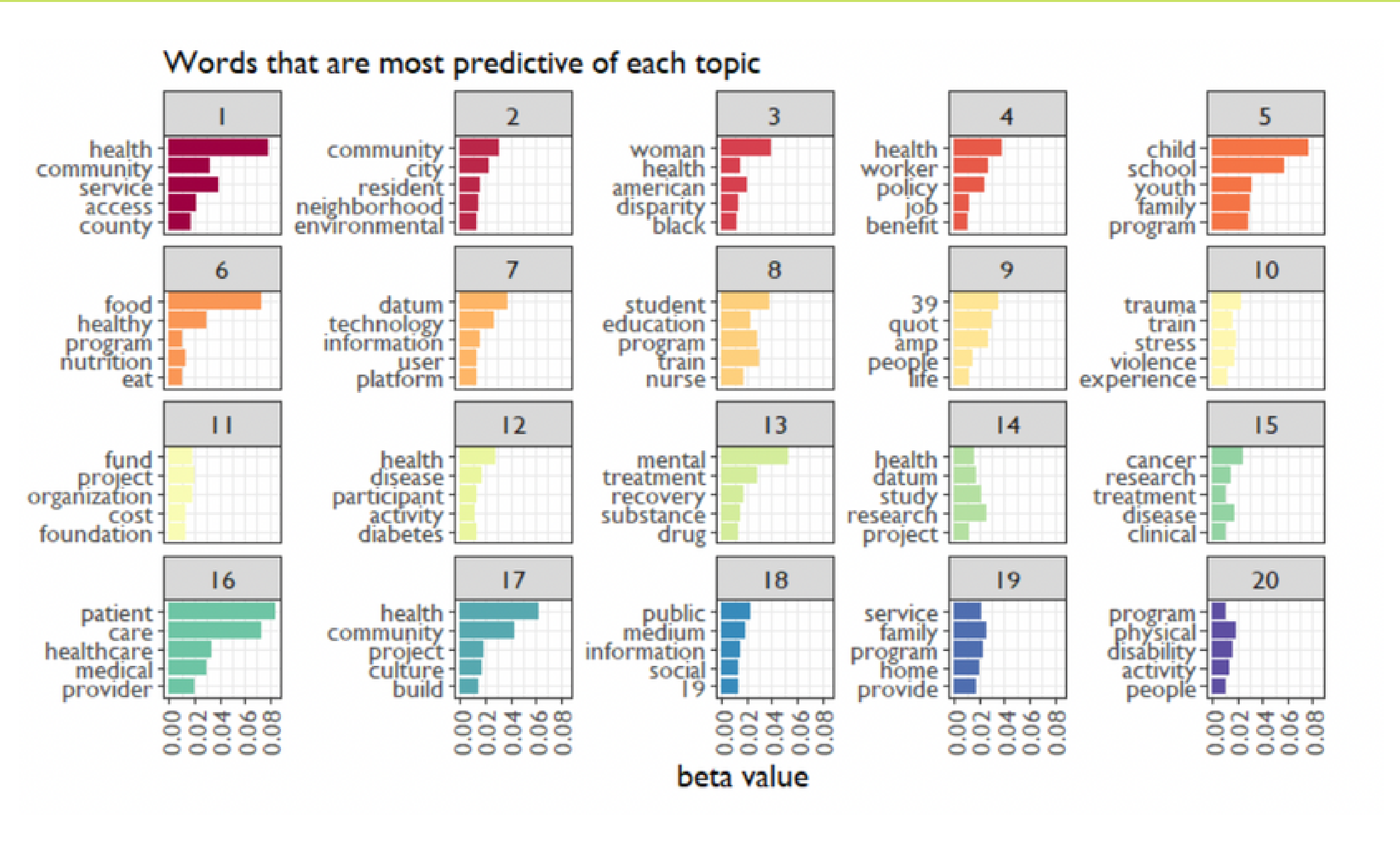
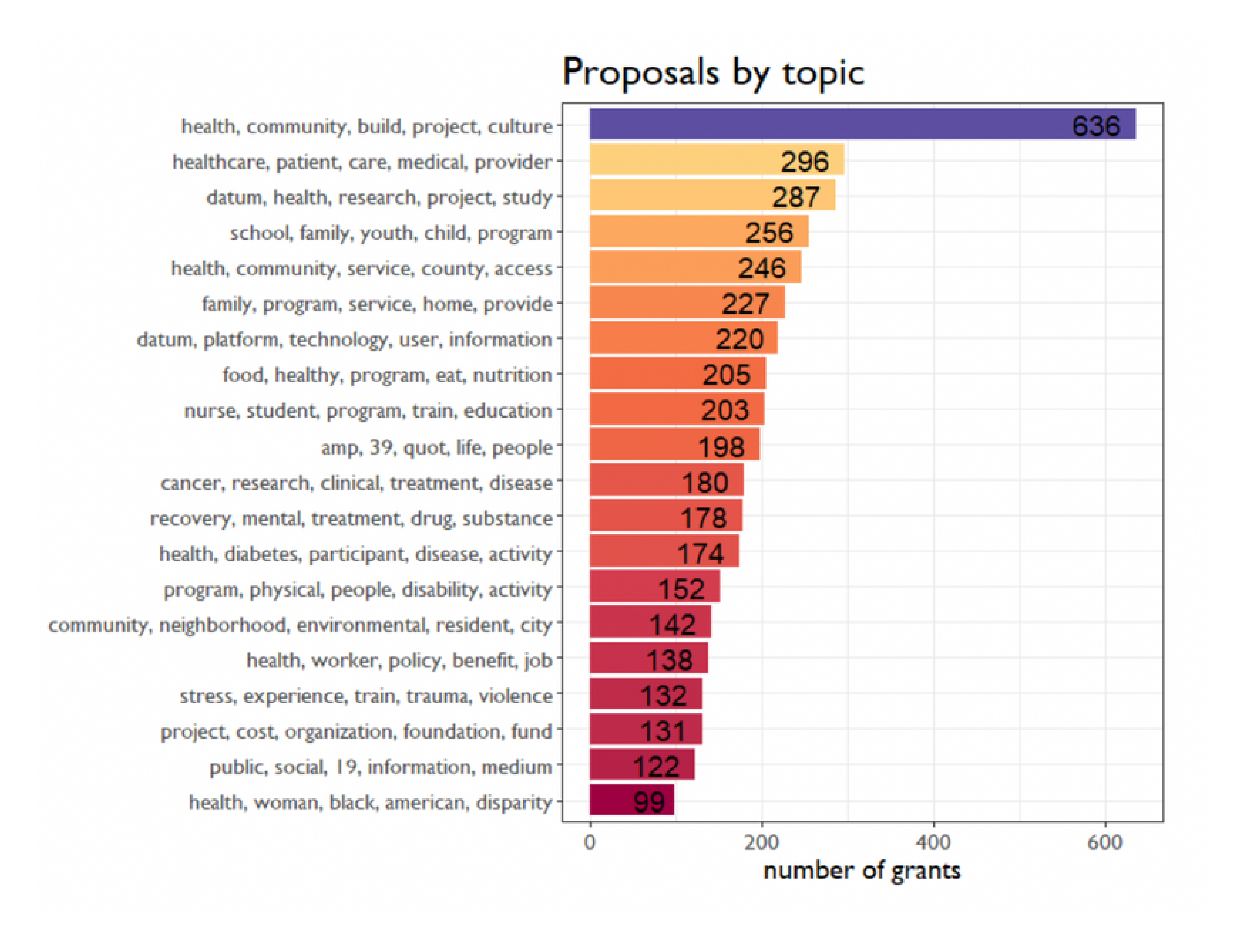


Figure 9 shows the frequency of proposals by topic. By far, the largest topic reflected applications that referenced the philanthropy organization's culture of health and its impact on the community. Surprisingly, the topic pertaining to health disparities appeared in the least number of proposals. This is not to say that other topics did not address issues of health disparities/equity. Rather, in this topic configuration, the number of primary-focused health disparities proposals showed up in 99 grant applications.

Figure 9. Frequency of Application Proposals by Topic



RQ3. What application attributes are predictive of positive funding?

Our analysis of Research Question 3 began with examining associations between application attributes and funding outcomes. We asked, i)Does the amount that grantees requested impact their decision? And, ii)Does the length of their application impact the decision? For both these questions, we found no meaningful difference in how these variables impacted the decision to proceed or not (see Figures 10 & 11 below).

Figure 10. Non-significant relationship between Amount Requested and Funding Decision

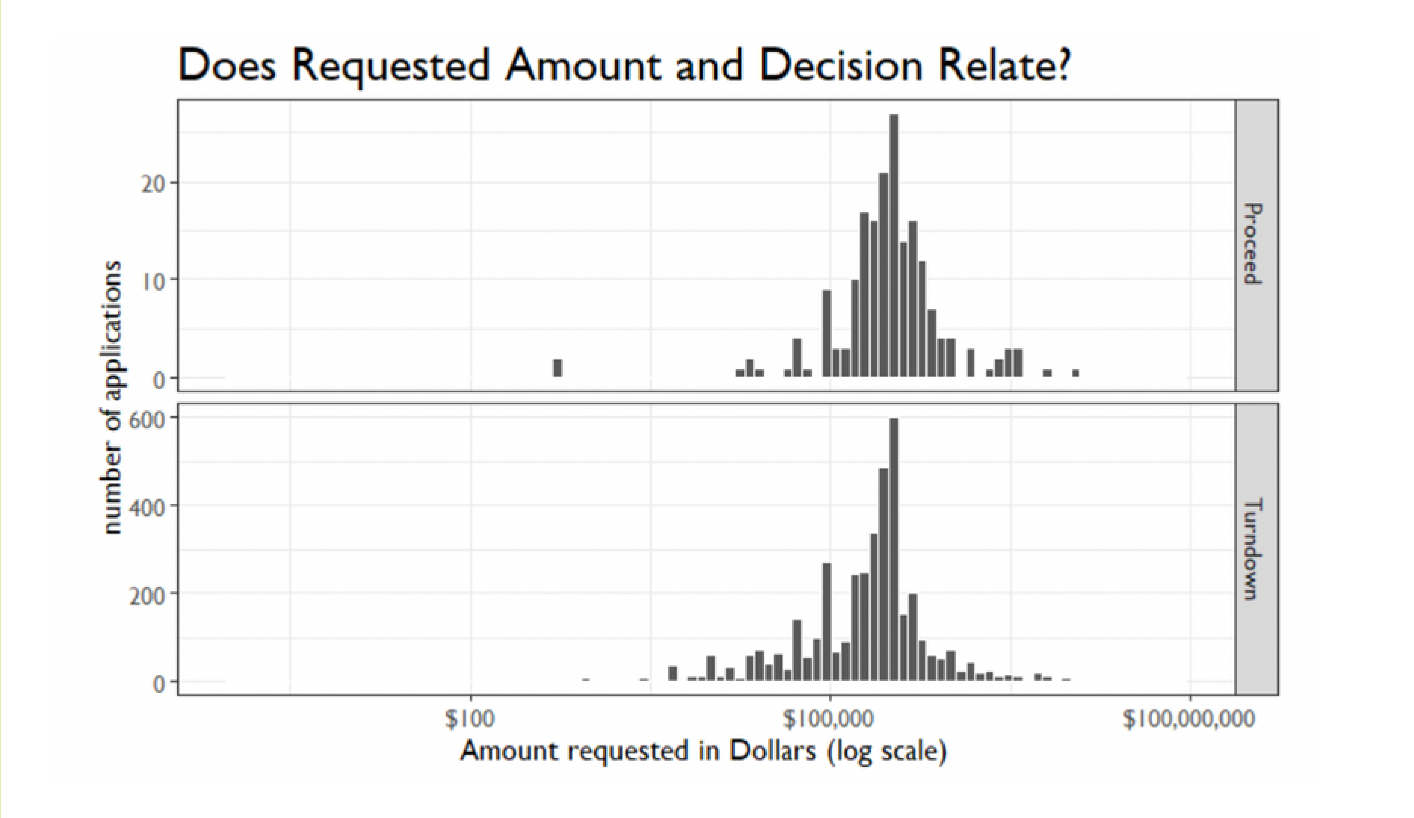
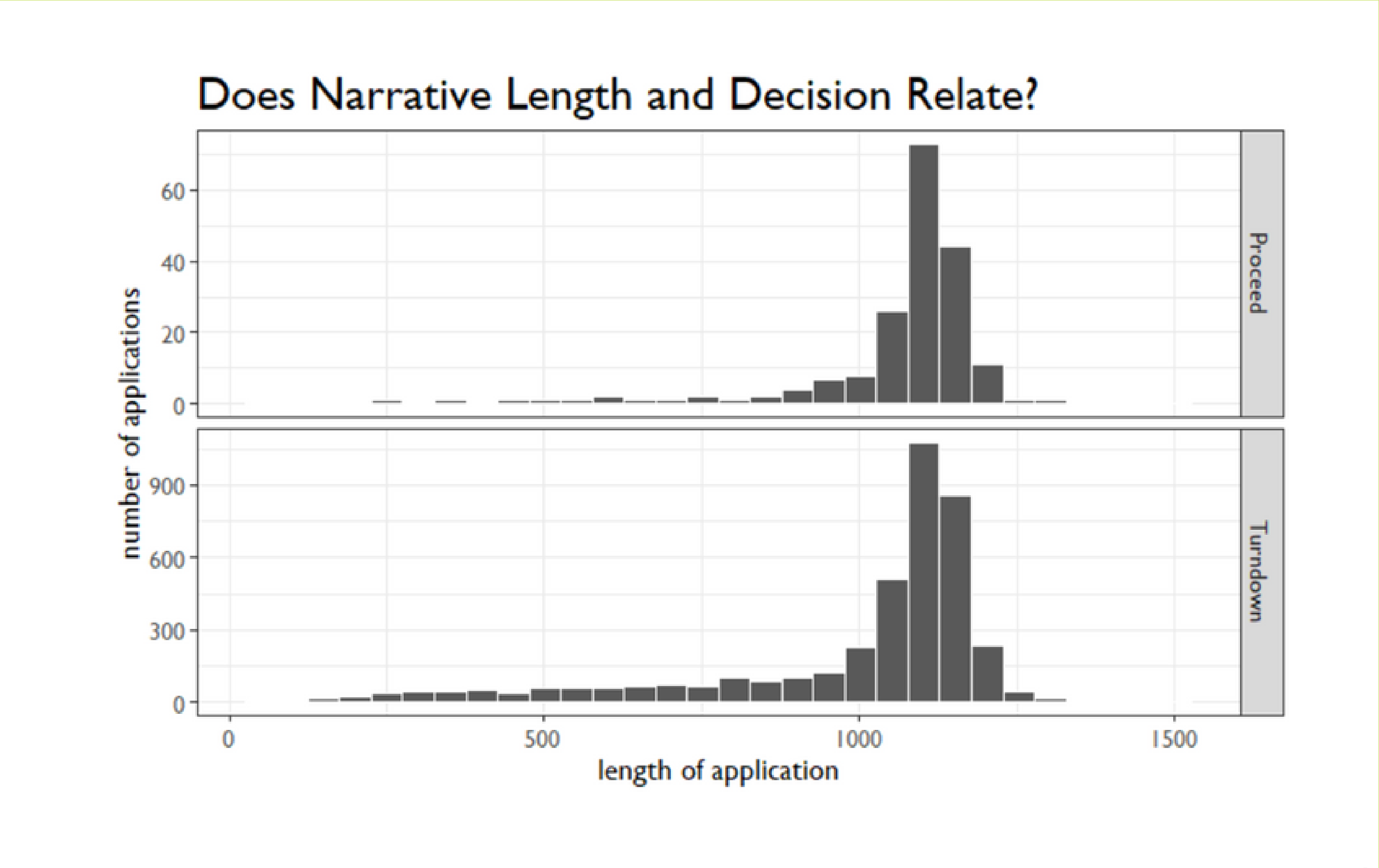
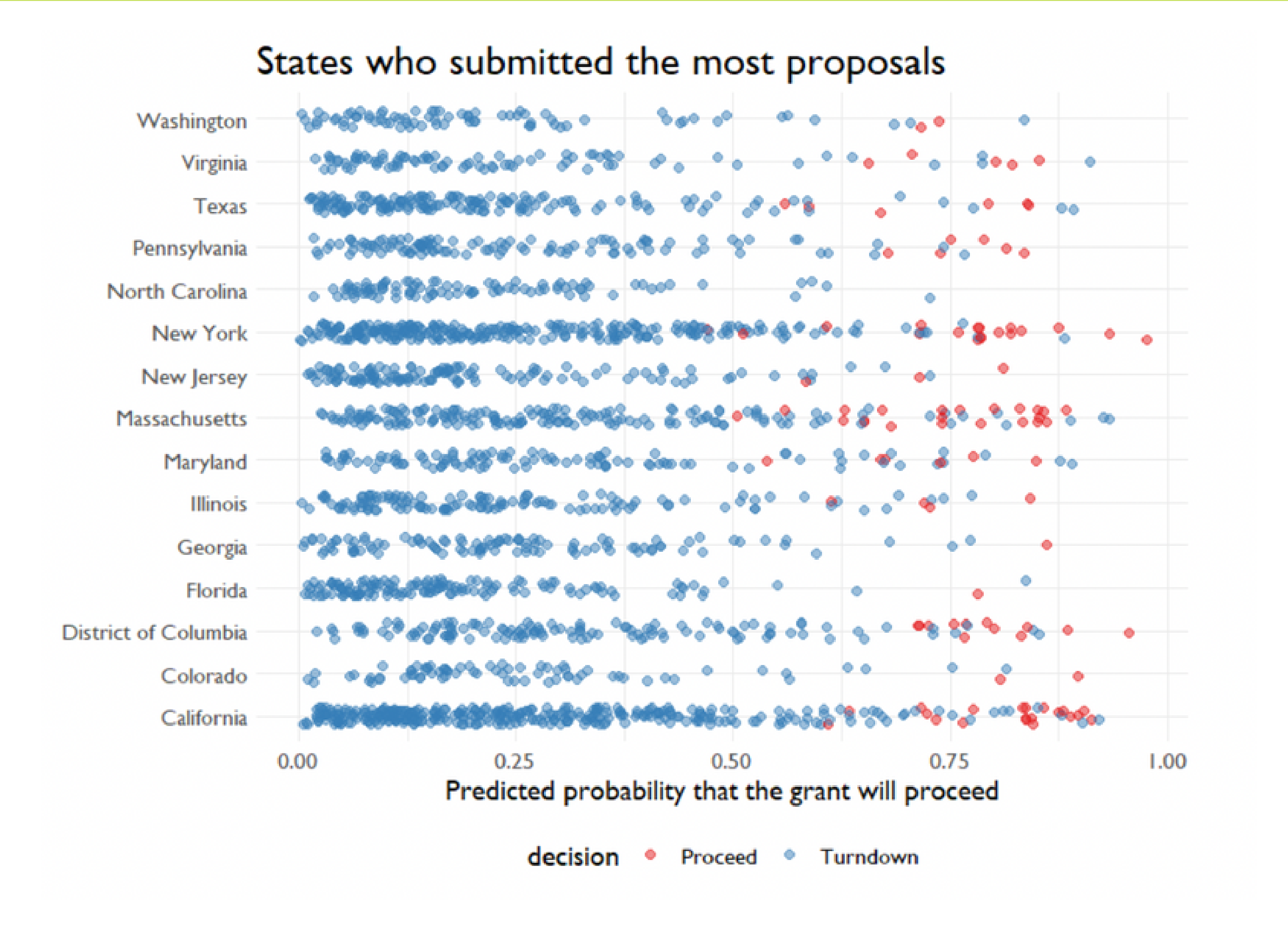


Figure 11. Non-significant relationship between Narrative Length and Funding Decision



We considered examining associations between organizational level characteristics and funding decisions. However, there were challenges that made this analysis unfeasible. For example, the number of applications per organization was too small (e.g. the top submitted organization - Johns Hopkins - had only 19 proposals) for analysis. The Creative Ideas portfolio is represented by a highly diverse set of organizations (a good thing from a funder's perspective!). We also considered examining the organization's geographic location but believed that depiction could be misleading, as an organization's headquarters may not be where projects are implemented or where the majority of employees work. Given these challenges, we deferred to the state level as the smallest unit of analysis. We examined the top 15 states who submitted the most proposals between 2013-2020. We found no meaningful differences among states, indicating that the state of the application is not associated with likelihood of funding. The values in this graph are derived from the predictions in our final machine learning model.

Figure 12. Non-significant Relationship between State of Application and Funding Decision

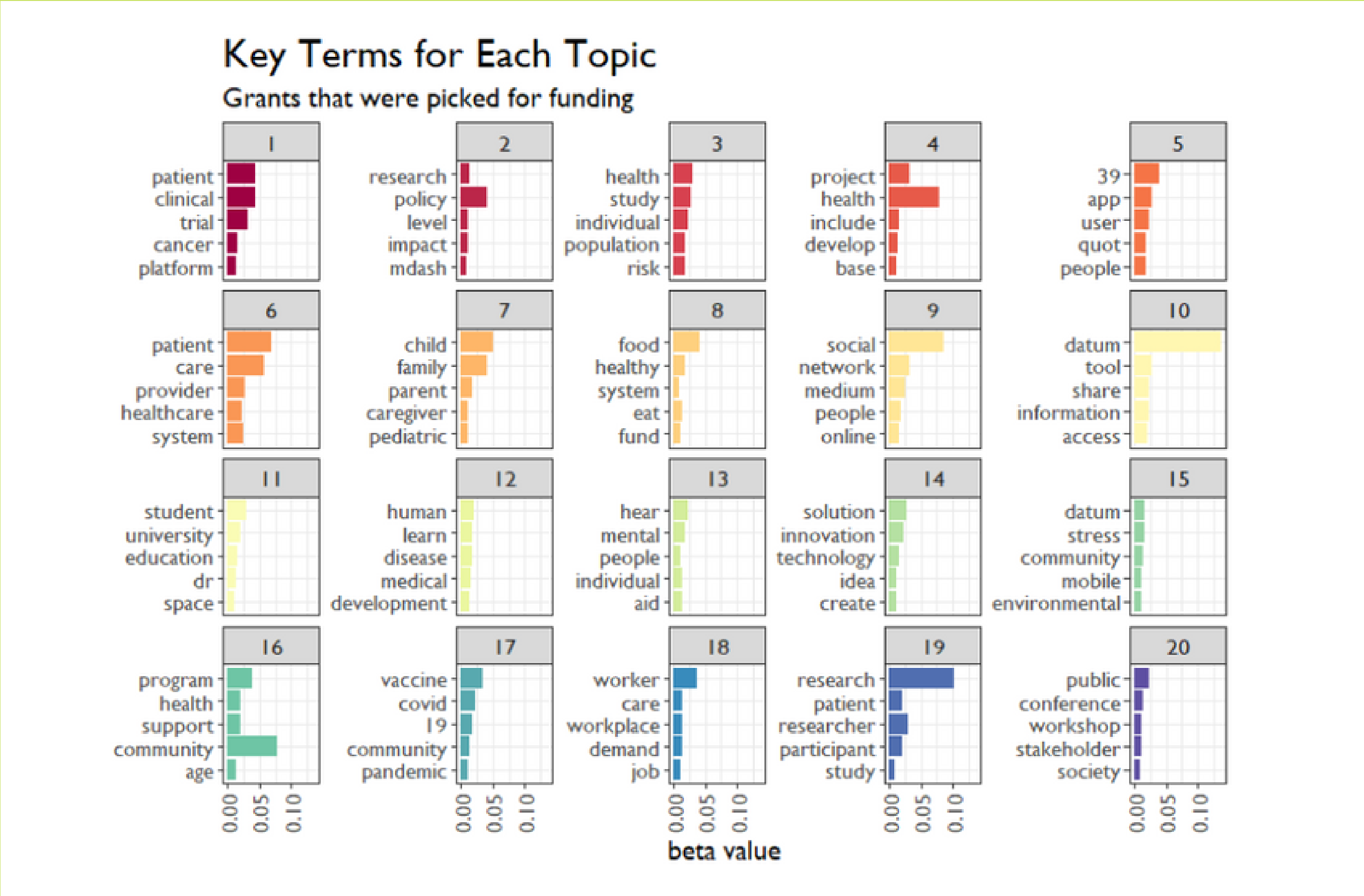


What topics appear in funded and unfunded proposals? We stratified by funding decision and replicated the topic modeling analysis from Research Question 2 to examine what topics appear in funded and unfunded proposals. Based on perplexity scores testing, we maintained the use of k=20.

Within applications that were selected to proceed, we found:

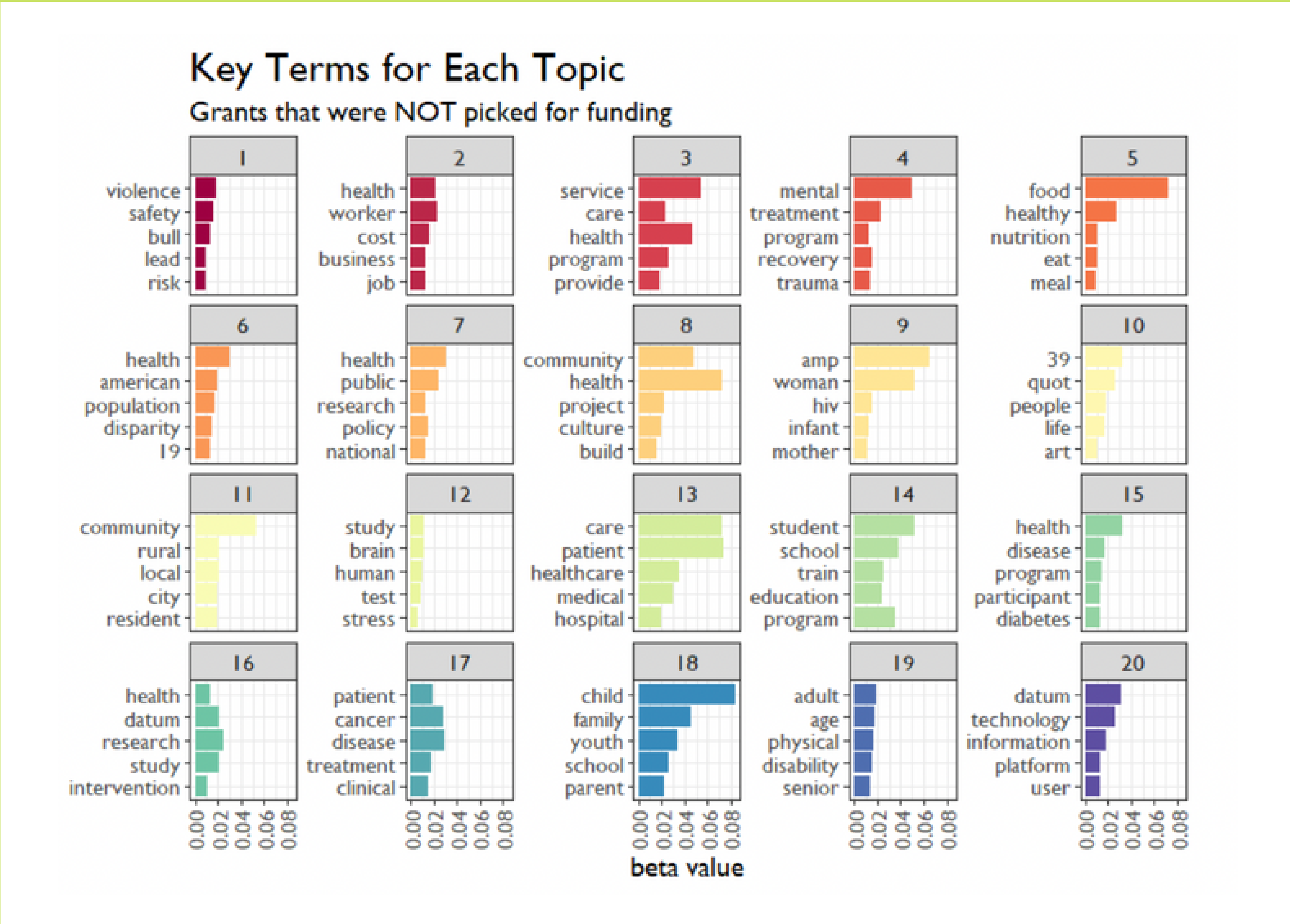
- *Health areas* pertained to cancer (Topic 1), risk (Topic 3), stress (Topic 15), age (Topic 16), Covid-19 (Topic 17), workplace/job (Topic 18).
- *Settings* explicitly reflected were healthcare (Topic 6), online (Topic 9), university (Topic 11), community (Topic 15, 16, 17), and workplace (Topic 18).
- *Populations* reflected were patients (Topic 1, 6, 19), providers (Topic 6), family/child/parent/caregiver (Topic 7), student (Topic 11), worker (Topic 19), researcher (Topic 19).
- Other notable topics reflected were research (e.g., clinical trial, study; Topic 1, 2, 3, 19), policy (Topic 2), use of technology (Topic 5, 9, 14, 15), and public forums (Topic 20). (see Figure 13)

Figure 13. Key Terms for Each Topic in Funded Applications



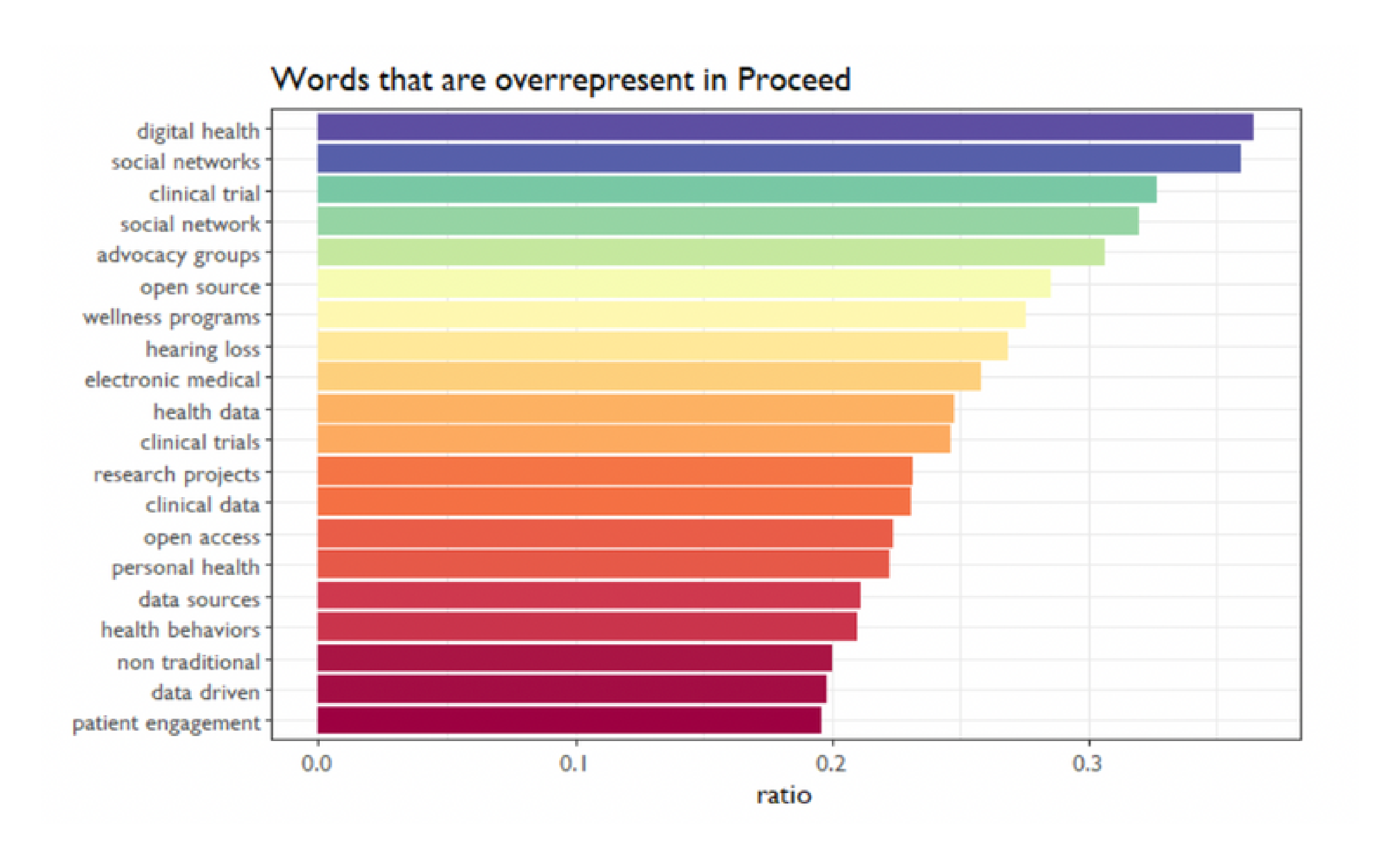
Applications that were not selected reflected some of the topics that appeared in the funded set (e.g., health areas: cancer, stress, age, risk). However, notable terms that appeared only in unfunded topics related to bullying (Topic 1), trauma/mental health recovery (Topic 4), school (Topic 14, 18), woman/mother/infant (maternal child health; Topic 9), and HIV (Topic 9). See Figure 14 for key terms and topics in unfunded applications.

Figure 14. Key Terms for Each Topic in Unfunded Applications



Next, we examined the most frequent bigrams (two-word phrases) occurring in funded proposals relative to unfunded proposals (see Figure 15). To compute this, we compared the ratio of bigram frequency in the proceed vs. turndown applications. These are not words that predicted funding per se. Rather, they are phrases that showed up more in the applications that received successful funding. Analyses revealed that technology-based terms are most represented (e.g., digital health, social network(s), open source, electronic medical, and health data), which may be especially appropriate for the future-oriented focus of the Creative Ideas portfolio.

Figure 15. Ratio of Bigram Frequency (Proceed:Turndown)



Piloting Machine Learning Models: Predictive and Neural Network Modeling

A goal of this project was to pilot the utility of NLP methods for predicting funding decisions. We tested nine machine learning models to identify the best model for predicting funding divisions. The fit of each model was assessed according to four model evaluation metrics: accuracy, precision, recall, FI; see table below for summary of fit indices. According to the set of model evaluation metrics, the machine learning strategies tested yielded two acceptable, but not high-performing models: Models 8 and 9. Given the size and imbalance of available data, we could not discern a robust relationship between each model's input and the likelihood of an application receiving funding. To some degree, the up-sampling did help alleviate challenges associated with an unbalanced dataset (i.e., significantly more unfunded than funded proposals). Upsampling is an approach commonly used with imbalanced datasets, such that synthetically generated data corresponding to the minority class is included in the data.

Evaluation Metrics for Nine Machine Learning Models

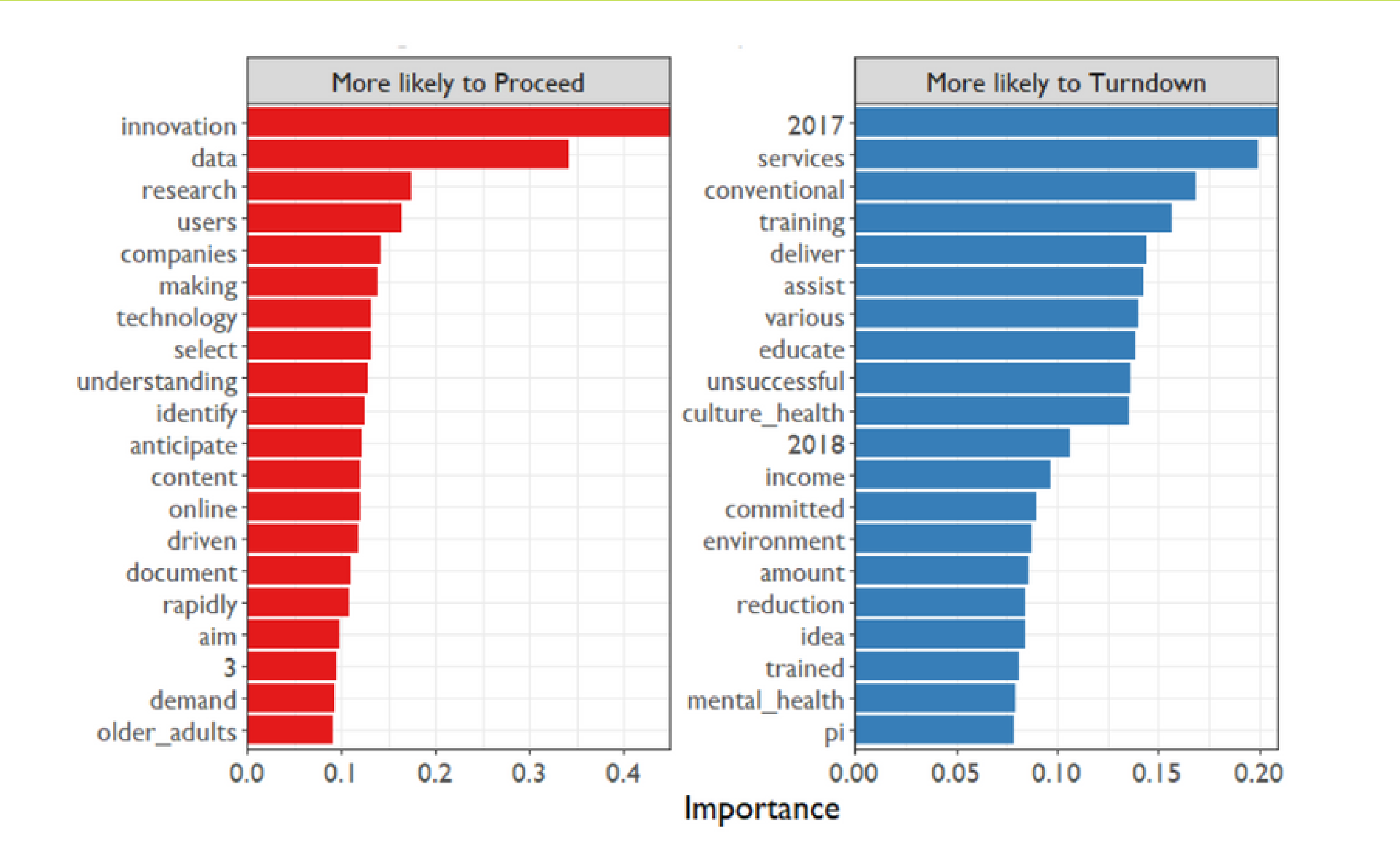
Model	Accuracy	Precision	Recall	F_1 Score
Model 0. Null model - reject all applications	0.970	0	0	0
Model 1. Tf + glmnet	0.941	0.223	0.159	0.189
Model 2. Term Frequency + upsampling + glmnet	0.896	0.156	0.312	0.208
Model 3. tf w/ 2000 tokens + upsample + glmnet	0.910	0.170	0.276	0.210
Model 4. tf + stem + upsample + glmnet	0.897	0.155	0.308	0.206
Model 5. tf + bigrams + upsample + glmnet	0.880	0.123	0.288	0.172
Model 6. tf + uni+bigrams + upsample + glmnet	0.910	0.168	0.274	0.208
Model 7. tf + tokenizers.bpe + upsample + glmnet	0.895	0.134	0.259	0.177
Model 8. tfidf + scaling + uni+bigrams + upsample + glmnet	0.903	0.148	0.260	0.189
Model 9. Neural Network - LSTM	0.920	0.241	0.310	0.272

Model Evaluation Metrics

- *Accuracy:* refers to how often the classifier is correct. Accuracy is computed by *True Positive + True Negative / Total*. However, Given dataset imbalance, accuracy alone is not a good indicator for model performance. For example, by using a null model, that is, we turn down everyone, we end up with about 97% accuracy.
- *Precision: refers to correct predictions (e.g.,* When the model predicts yes, how often is it correct?) Precision is computed by *True Positive / (True Positive + False Positive).* Precision is a good measure to determine when the cost of False Positive is high. In cases where there is a limited amount of resources to distribute, it may be necessary to make sure that decisions are "correct"
- *Recall:* When the outcome is actually yes, how often does the model predict yes? This is computed by *True Positive / (True Positive + False Negative.).* This is a good measure when the cost of non-detection is high. In the current COVID/Omicron world, not knowing a positive case may lead to further infections and spread.
- F1-score: The F1 Score balances between Precision and Recall. It is computed by 2 x (Precision * Recall)/(Precision + Recall). In most cases, we do not need to pay much attention to our True Negatives, whereas False Negative and False Positive have varying levels of tangible and intangible repercussions.

Working with the best performing predictive model (i.e., Model 8), we examine text features for predicting funding decisions. Figure 16 shows the top 20 terms that were most predictive of funding decisions. Terms associated with funded applications include innovation, data, research, and technology. Terms associated with being turned down include 2017 (the term within the text, not the date of submission) services, conventional, and training.

Figure 16. Top 20 Terms Most Predictive of Funding Decisions



Readability

Our final predictive analysis addressed the general readability of the narrative. Plausibly, more readable narratives could be easier to engage with and thus communicate their ideas better. Readability captures the ease of reading document text (narratives), which is shaped by textual features including number of words, morphological complexity, and syntactic richness of a sentence. Alternatively, less readable and more jargon-filled narratives may convey a higher level of sophistication, as is sometimes the case in academic writing. To test this hypothesis, we computed 45 different metrics of readability and whether or not they predicted proceed. At first glance, there appeared to be instances of differences (see Figure 17). However, these were driven by outliers in the data, as seen in the violin plot Figure 18..

Figure 17. Average Readability of Applications based on Funding Decisions

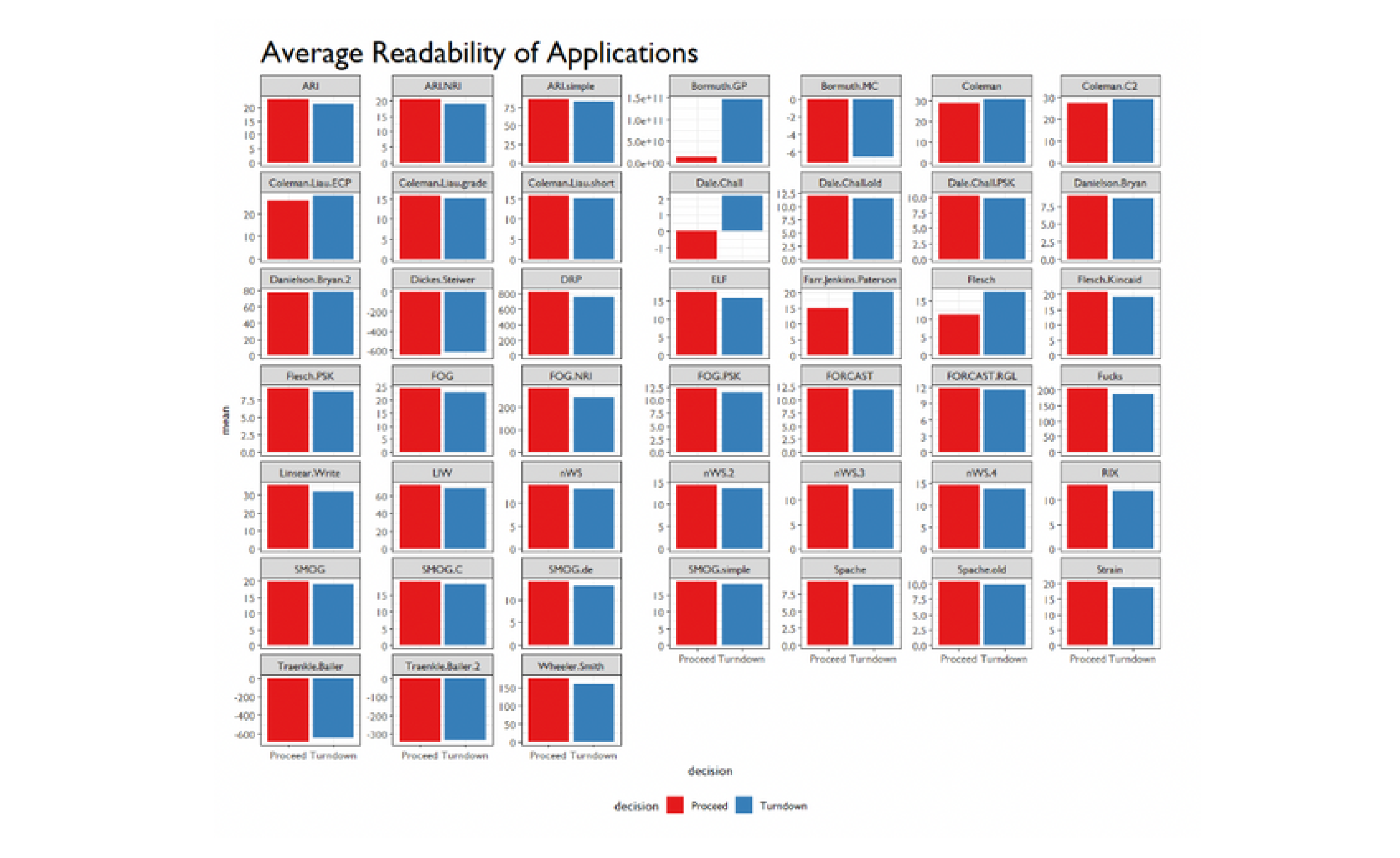
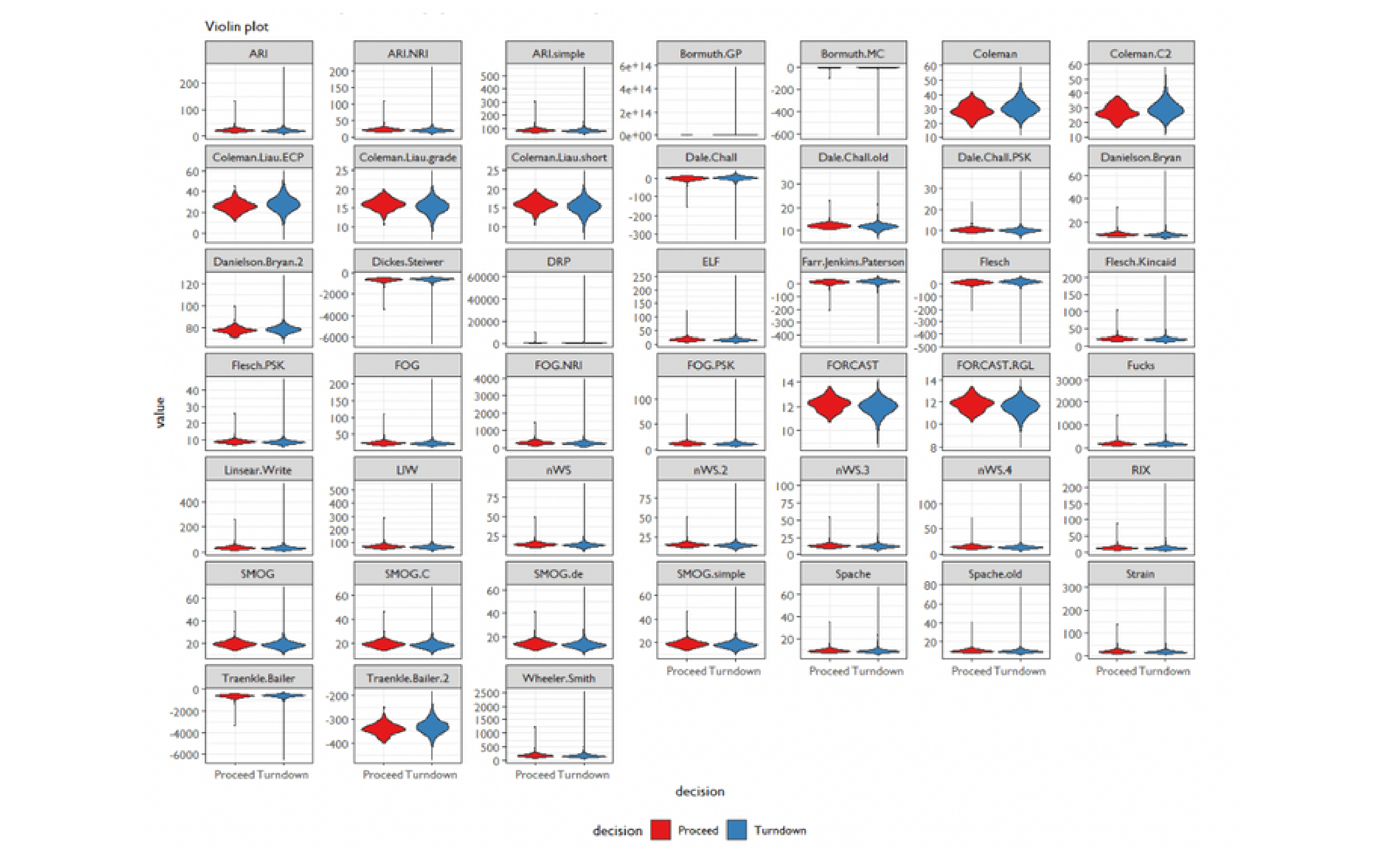


Figure 18. Readability of Applications by Decision - Violin Plot



To be certain, we ran another predictive model to see whether the decision could be a function of readability. We normalized all readability indices and similarly upsampled the outcome to ensure balanced training. Based on the four evaluation metrics (accuracy, precision, recall, FI-Score; see summary table below), the model performed poorly. Our analyses suggest that readability did not predict funding decisions in the examined dataset.

Model	Accuracy	Precision	Recall	F_1 Score
GLMNET + upsampling + normalization	0.593	0.063	0.598	0.113

DISCUSSION

This project began with a broad question: how do grant application characteristics shape funding decisions? And, what are the benefits and constraints of NLP methods for analyzing grant application data? Through this Creative Ideas project, we arrived at the following general insight: NLP is a powerful synthesis method for examining grant making. However, it requires large amounts of textual data and performs better as text volume increases. A key challenge in this particular analysis was the significant imbalance in available textual data between funded and unfunded applications, with funded applications comprising an average of 5% of the overall dataset. Despite our efforts at balancing the data, the models deployed were unable to effectively identify textual features highly predictive of successful funding. Nevertheless, the data lent itself to discerning several important trends.

First, funding outcomes in the examined dataset are not predicted by amount of requested funding, state, narrative length, or readability. An implication of this finding is that funding decisions are informed by other application attributes of greater interest to the Creative Ideas project portfolio and the philanthropy organization. Analyses of geographic data revealed that the majority of applicants are located in the northeast region of the U.S.. This is likely due to the regional influence of the philanthropy organization. Deliberate steps might be taken to solicit Creative Idea applications representing states in the southern, mid-western, and western regions. It is useful to note that the trend as it relates to narrative length and readability may exist as a byproduct of the data. Both of these variables had minimal variance such that narrative length and readability of applications were quite comparable across applications. The homogenous nature of variables can preclude the identification of statistical associations.

Second, technology-based words and “research” are most highly predictive of positive funding. Overall, the terms most associated with positive funding align with the stated mission of the Creative Ideas portfolio. One exception is the term “equity,” a concept resting at the heart of the philanthropy organization's organizational mission. In supplemental analyses (informed by conversations with Creative Ideas leadership), we found that the term equity first appears in 2020 for analyses limited to the “top 35 most commonly appearing words”. When we expanded the analytic parameters to “top 100 most commonly appearing words,” the term “equity” first appeared in 2017.

Third, NLP methods can be useful for understanding trends in grant funding when used in conjunction with human expertise. Our analyses were iteratively informed by input from Creative Ideas staff as well as the project advisory panel. The collective and continuous feedback loop contributed to appropriate interpretation of the data and emergent trends. Complementing NLP methods with human expertise enabled an efficient synthesis of trends based on over 4,000 applications across an eight year period.

Our exploratory analyses were focused on a specific portfolio of grant applications. The potential for understanding associations between application characteristics and funding decisions expands with the availability of data. This may involve analyses that include a greater number of years, more grant portfolios, or applications across philanthropic organizations. An expanded and deeper dive can be instructive for both funders and prospective applicants. For instance, insights from larger data samples could inform modifications to application solicitation and review criteria. We draw on findings from our analyses to provide a tangible example: based on the low frequency of applications reflecting the term “equity,” a funder might modify the application call and review criteria to elevate the importance of “equity.” In doing so, the funder seeks to increase disbursement toward projects with an equity focus.

Enhancing transparency about funding disbursement decisions and communication with prospective applicants can be another benefit to examining grantmaking trends using NLP methods. In the case of insights from our project, the funder may publicly share trends, such as conveying that applications with technology-based words and research historically have had a higher probability of funding and that applications focused on “training” were rarely funded.

Additionally, the funder might share that particular application characteristics (e.g., state, amount requested) are not predictive of funding decisions. In sharing this information, a funder can provide useful guidance about the specific application elements that are of central interest to the funder. This type of guidance would optimize the use of resources for the funder by promoting the submission of applications that are arguably better aligned with the intent of the funding call. It would also optimize the use of resources for prospective applicants by preemptively generating guidance that helps prospective applicants “score” the application and iteratively improve it by adding elements that are associated with funding decisions. In the worst of light, this might simply enhance grantsmanship. In the best of light, the transparency and preemptive guidance can deepen community trust in a funding organization and optimize the use of community and funder resources.

Limitations to Consider

Several limitations are important to note along with insights from this project. First, the insights from our analysis are specific to the dataset provided. As such, generalizations cannot be extended to other philanthropy organization portfolios. Second, while we were able to train the data for an operable predictive model, the size of the data sample (4,199 applications) precluded the attainment of a strong performing predictive model. A replication study with a larger data sample is recommended for increased confidence in the identified trends. Lastly, while the NLP methods illuminated patterns in the text, the patterns are not strongly associated with funding. This may be an artifact of the data sample size rather than a representation of actual phenomenon.

Future Directions

Based on insights from this project, we extend the following recommendations for the philanthropy organization\ to consider as future directions:

01 **Replicating this analysis with additional application sets.**

This could produce a number of findings that would be of significant value to any organization involved in grant making. This includes topics that are over or underrepresented in the Creative Ideas application narratives.

02 Replicating this analysis with other applications sets

Perhaps there are factors that make the Creative Ideas portfolio unique and distinct from the other portfolios that the philanthropy organization funds. Indeed, as seen in the figures below, the Creative Ideas Project portfolio represents only 0.5% of the philanthropy organization's total funding outlays.

There is good reason to believe that an analysis of the Creative Ideas portfolio would not be indicative of the Foundations as a whole. For example, the types of people who submit to the Creative Ideas portfolio are likely to be creative and opportunistic. That creativity, as expressed by content variation between applications, may not be present in other types of RFP, which are seeking specific solutions to specific problems. This may constrain the text in the narrative and prove to be a better method to illuminate the unique factors that relate to funding decisions.

03 Expanding this analysis with new questions

The following set of new questions emerged for us as a result of this project. These are potential future areas of inquiry for philanthropy organization.

- How do these textual trends ultimately compare to the foundation's portfolio overall, which constitutes 99.5% of the total investment? Would a more complete model show more reliability in funding decisions?
- What were the specific criteria used for selecting applications? Do attributes of those criteria appear in textual analysis, or is there something else in the grant determination process that comes into play (e.g. pre-existing funding, organizational size, existing relationships)? How can we explain the unexplained variance in our model?
- How do funded applications align with the mission of philanthropy organization (e.g., focus on health equity; culture of health)?
- How do organizations/sectors compare between those that submit and those that are funded? Do universities tend to apply the most and get the most funding?

Conclusion

We believe NLP is the wave of the future for philanthropy organizations. Across sectors, NLP is increasingly leveraged by organizations to understand group trends and to inform organizational decisions. This project illuminated trends in the Creative Ideas portfolio that has instructive value for philanthropy organization grant making activities and stimulated new questions ripe for prospective examination. Additionally, it demonstrated the value and constraints of NLP methods for analyzing grant application data and associations with funding decisions. While NLP methods are sophisticated and can rapidly synthesize large amounts of data, we believe it is essential that use of NLP is done in conjunction with human expertise.