# Proof of Concept for Data-Driven Philanthropy Project

This document represents the preliminary analysis of a sample of the data I expect to use to develop my Data Incubator Project. There are 3 sections of this document with between one and five subsections.

# About Us Statements

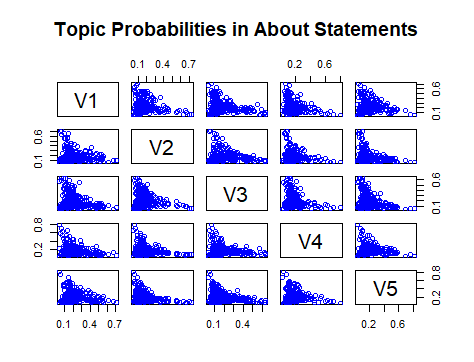
In this section, I offer an analysis of the “About Us” statements available on the web from a sample of 261 nonprofit organizations accredited by the Better Business Bureau. In each of the subsections below, I show how the difference in the number of expected topics (k) alters the arrangement of the probability of a statement including these topics and the number of documents within which these topics are represented.

K=5

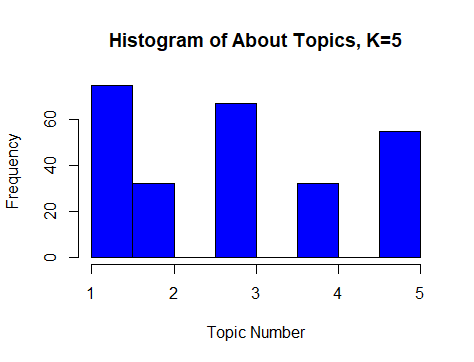
Many unsupervised learning programs use the k=5 convention for preliminary analysis. To stay within expected convention, a k=5 analysis is presented first with no supporting context information implied to choose k.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| People | National | Research | Children | World |
| Work | Public | Cancer | Organization | people |
| Communities | Center | Foundation | Hope | Help |
| Programs | Society | Organization | Education | Need |
| International | States | Support | First | every |
| Food | Rights | Families | Program | care |

The probability distribution of organizations having one of these topics in their About Us statement is represented in the probability distribution matrix below.



A histogram of the distribution of the number of organizations with each topic is presented below.



At the conventional k=5 level, a large number of organizations are obviously represented in topic 1, a set of relatively generic “helping” terms.

K=26

The National Center for Charitable Statistics offers the NTEE (National Taxonomy of Exempt Entities) codes, an A-Z alphabetical classification method for nonprofit organization activities. This code list is used by the IRS, researchers, and philanthropic monitoring organizations to categorize nonprofit organizations according to high level (A-Z) and specific (000-999) activities. Recent research suggests that the NTEE codes may not be an adequate method for categorization of smaller or newer nonprofit organizations whose mission and activities are more variable than older, established, organizations. This analysis of k=26 is presented to demonstrate whether there are 26 distinct topic within this random sample of BBB accredited nonprofits.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| Food | God | Watch | Volunteers | Water |
| Programs | Jesus | Outreach | Relief | Conservation |
| America | Christ | Throughout | Service | national |
| Community | Church | Media | Serve | future |
| Providing | Believe | Educational | Help | environmental |
| Bank | Mission | Judicial | Disaster | rivers |

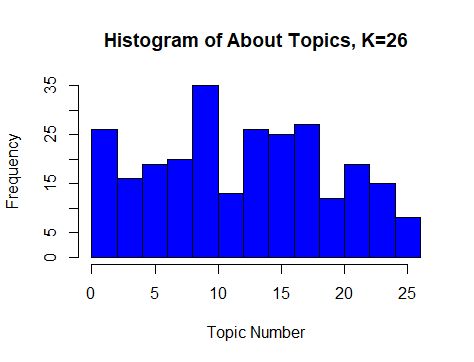
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 6** | **Topic 7** | **Topic 8** | **Topic 9** | **Topic 10** |
| Educational | Aid | International | Rights | People |
| School | New | Social | American | Work |
| Program | Farm | Development | Free | Communities |
| First | Services | Impact | Right | World |
| Schools | Family | Human | Justice | Lives |
| Students | rural | organizations | religious | Local |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 11** | **Topic 12** | **Topic 13** | **Topic 14** | **Topic 15** |
| Public | Animals | Cancer | Make | Children |
| Policy | Animal | Fund | Every | Hope |
| Advocacy | Wildlife | Research | People | Asia |
| Government | Paws | Years | Help | Home |
| Nations | Work | National | Well | Child |
| Food | Since | Help | Ocean | Care |

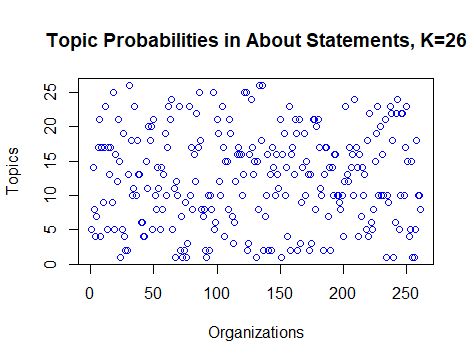
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 16** | **Topic 17** | **Topic 18** | **Topic 19** | **Topic 20** |
| Care | Research | Young | Society | Center |
| Health | Foundation | Women | National | Safety |
| Medical | Patients | Girls | Board | Safe |
| World | Disease | Support | Autism | Also |
| International | Alzheimers | Leaders | Including | Vital |
| Need | Information | Every | Directors | Year |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Topic 21** | **Topic 22** | **Topic 23** | **Topic 24** | **Topic 25** | **Topic 26** |
| Families | Foster | Organization | Organization | Also | Women |
| Association | Redwings | Foundation | Founded | Jewish | Lilly |
| Support | Housing | Service | First | Blind | Many |
| Programs | Mission | Veterans | Conference | JBI | Began |
| Children | New | Assistance | Known | Israel | Boys |
| National | Day | Support | UNFPA | Books | Provided |

The histogram of topics is presented below. In a random sample, a more uniform distribution might be expected.



The probability distribution of topics is presented below. This probability distribution appears mostly random, which is consonant with the findings of recent research that the 26 NTEE codes may not be any better than random for organization of nonprofit missions, depending on size and age of the organization.



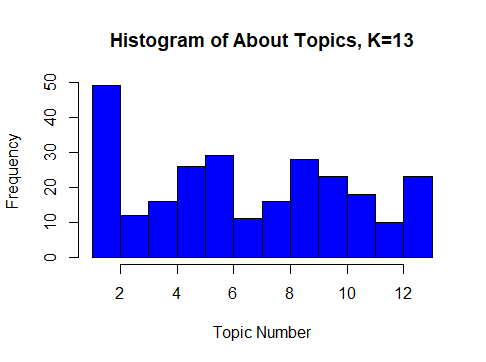
K=13

Given the findings from the NTEE code plots, an analysis of half the number (k=13) was conducted to assess whether the number of expected topics effects distribution.

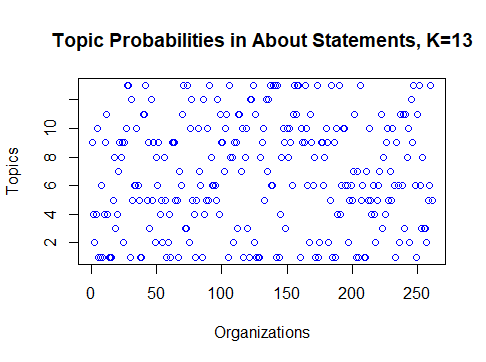
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** | **Topic 6** |
| Research | Public | Food | Every | International | Animals |
| Cancer | Rights | Water | World | Social | Animal |
| Foundation | Conservation | America | Make | Services | Work |
| Patients | American | Public | Help | Development | Sanctuary |
| Support | Issues | Action | Years | Impact | Paws |
| Leading | Environmental | Hunger | Can | Donors | Asia |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 | Topic 13 |
| People | Children | National | Health | Education | Also | Believe |
| Work | Hope | Families | Care | First | Center | God |
| Communities | Asias | Support | Medical | School | Safety | Church |
| Local | Organization | Association | Need | Program | America | Life |
| Partners | Child | Organization | Programs | Medical | Women | Jesus |
| Poverty | Home | Society | Countries | Years | Jewish | Mission |

A histogram of the number of organizations whose About Us statement includes these topics is presented below.



A distribution of the probability of the topics and organizations is presented below.



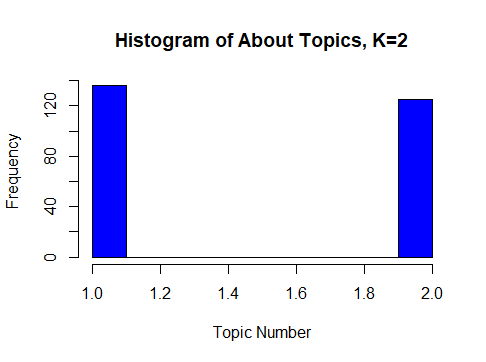
It appears that, with fewer clusters, more regular patterns appear.

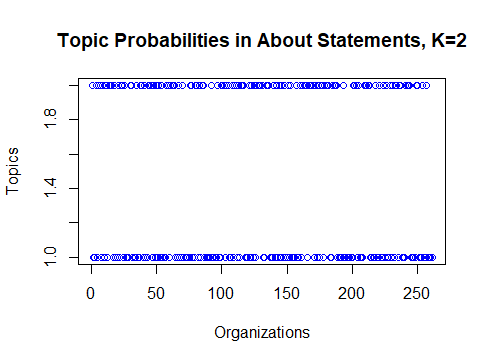
K=2

To test a minimum limit of the number of clusters, particularly given that an attempt at an elbow plot was inconclusive, a k=2 analysis was run to examine the topics at the least feasible level. At this level, the general topics that nonprofit organizations use to describe themselves become more obvious. In future analysis, these may be usefully included in a list of specific “stop words” to remove to reveal more substantive commitments.

|  |  |
| --- | --- |
| Topic 1 | Topic 2 |
| People | Children |
| Work | Organization |
| World | Support |
| Communities | Research |
| Help | National |
| International | Care |

This distribution of general terms is clear in the histogram and probabilities plots below.





Summary

About Us statements are a rich source of organization descriptions of their orientation towards the service tasks they choose to adopt as their purpose. These statements may or may not be included in the major texts used by organizations, like the IRS, or researchers to identify what organizations do and why. To evaluate this, the mission statements of organizations are analyzed below.

# Mission Statements

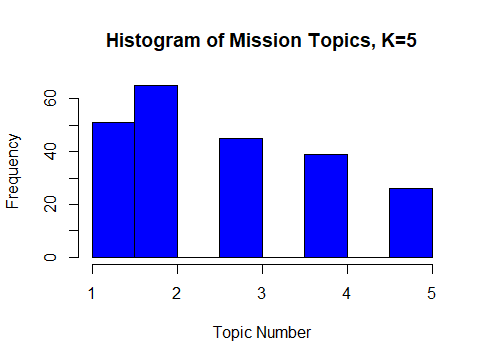
Mission statements are arguably the most salient expression of an organizations purpose. Mission statements are also expected to be included in the IRS 990 forms. In a full version of this proposed project, scraping the mission statements is expected to be a robust component of the presentation of organizations to users. 226 Mission statements were analyzed in this section.

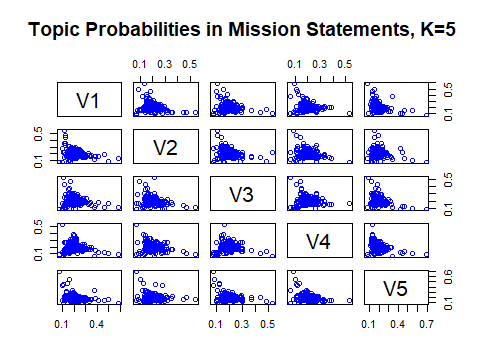
K=5

The conventional k=5 analysis is presented here below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| Mission | People | Health | Communities | Life |
| Cancer | Children | Education | Families | National |
| Research | World | International | America | Prolife |
| Support | Lives | Need | Public | Human |
| Foundation | Help | Programs | Americas | Child |
| Provide | Hope | Quality | Wildlife | Members |

The histogram presented below shows the larger number of organizations with a mission encompassing either cancer research or aid to children.





k=26

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| Support | Action | Mission | Every | Mission |
| Health | Poverty | Needs | World | Quality |
| Research | Lasting | Providing | Christ | Education |
| Programs | World | Quality | Ministries | Institute |
| Affected | Promotes | Child | People | Engage |
| Services | Foundation | Family | Lives | Parks |

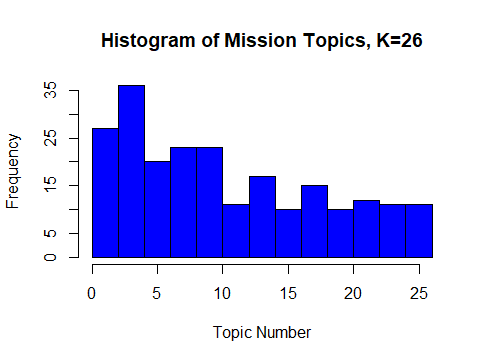
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 6** | **Topic 7** | **Topic 8** | **Topic 9** | **Topic 10** |
| People | Wildlife | Training | Young | Mission |
| Communities | America | Help | Women | Cancer |
| Network | Protect | Hope | Cancer | Cure |
| Partners | Believe | Nations | Fund | Disease |
| Abuse | Conservation | Ensure | Provides | Lung |
| Service | Legal | Families | Survivors | Foundation |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 11** | **Topic 12** | **Topic 13** | **Topic 14** | **Topic 15** |
| Life | Public | Children | Watch | Human |
| Prolife | Families | Missing | Healthy | Child |
| National | American | Suffering | Potential | Organization |
| Hunger | Worldwide | Educational | Together | Rights |
| Members | Groups | Nonprofit | Haiti | Effective |
| Alliance | Funding | Local | Future | Advocacy |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 16** | **Topic 17** | **Topic 18** | **Topic 19** | **Topic 20** |
| Mission | Organization | Work | One | Education |
| America | National | Innovative | Life | Catholic |
| Communities | Disaster | Promote | Mission | Providing |
| Social | Lives | Aid | Americas | Association |
| Dedicated | Organizations | Development | Founded | Poor |
| Around | Also | People | Rural | Communities |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Topic 21** | **Topic 22** | **Topic 23** | **Topic 24** | **Topic 25** | **Topic 26** |
| People | Creating | Provide | Programs | Lives | International |
| Build | Worldwide | Resources | Providing | Care | Need |
| Among | Dogs | Food | Public | Eye | Solar |
| Issues | Church | Individuals | Local | Treatment | People |
| Economic | Leaders | Financial | Promotes | Homes | Changing |
| Nonprofit | Individual | Friends | Opportunities | Society | World |

Although the probability distribution of the mission language encompassing one of these topics could not be plotted by R, the histogram is illustrative of the comparative generality of mission statement language when a larger number of topics is considered.

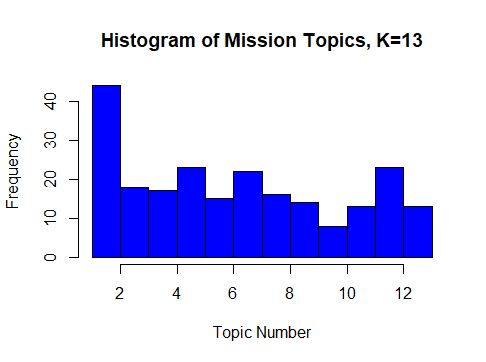


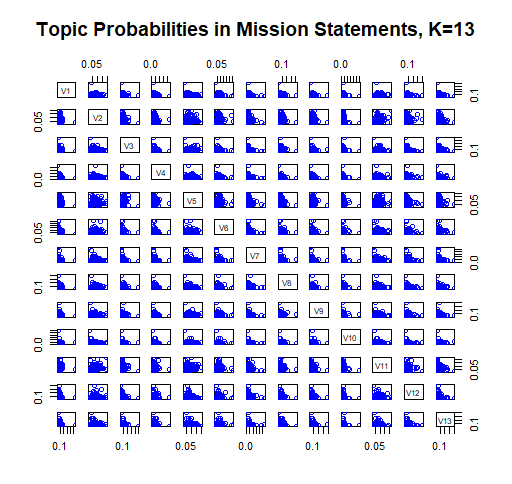
K=13

The NTEE codes are routinely used in conjunction with the mission statement of an organization to grasp their overall purpose. As with the About Us statement analysis, a k=26 analysis may prove too many expected topics. Half that number is run here.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** | **Topic 6** |
| Communities | Education | America | Need | World | Public |
| Organization | Community | Christ | International | Change | Network |
| Help | Social | United | Lives | Partners | Mission |
| Providing | Families | Every | Poverty | Food | Make |
| Local | American | Catholic | Around | Resources | Government |
| Association | Working | Ministries | Provides | Development | Watch |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Topic 7** | **Topic 8** | **Topic 9** | **Topic 10** | **Topic 11** | **Topic 12** | **Topic 13** |
| Mission | Children | Lives | National | People | Health | Wildlife |
| Cancer | Foundation | Serve | Prolife | Provide | Support | Dedicated |
| Research | Cure | Can | Work | Human | Programs | Education |
| Young | Mission | Mission | Life | Affected | Quality | Animals |
| Disease | Child | Needs | Hunger | Rights | Services | Protect |
| Women | One | Nonprofit | End | Organizations | Care | Commitment |





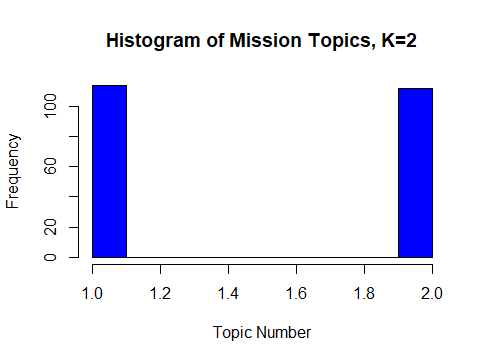
Both the histogram and probability distribution plots show considerable commonalities across topics expressed in mission statements.

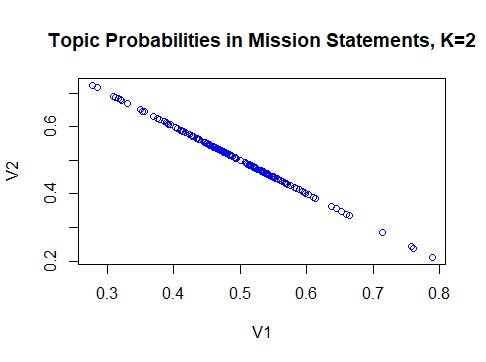
K=2

The findings in the about statements are similar to those in the mission statements, suggesting that a very small number of k would be a useful analysis to estimate unique stopwords needed for this data. A k=2 analysis is presented below.

|  |  |
| --- | --- |
| **Topic 1** | **Topic 2** |
| Mission | People |
| Life | Children |
| Cancer | Communities |
| Support | Lives |
| Families | Health |
| Research | World |

The corresponding plots of topics and topic probabilities is as expected.

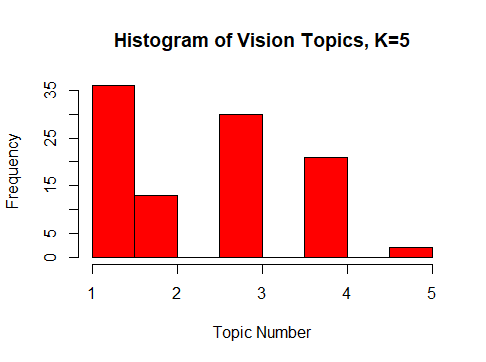


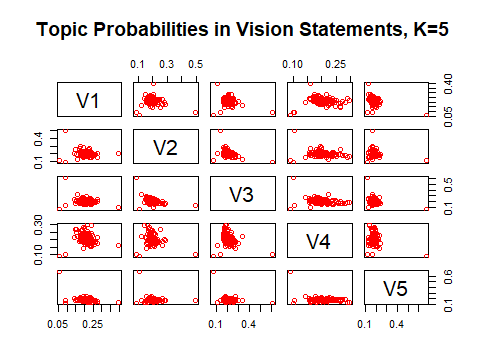


# Vision Statements

As organizations are asked to compete on branding, vision statements—short, easily recalled, tag lines—are a more important part of nonprofit statements of purpose. Although these short statements are presented here, the analysis is truncated due to insufficient quantity of data to analyze from the 102 organizations who publish an explicit vision statement. Only a k=5 analysis is presented.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 |
| World | Future | Vision | Communities | Will |
| Life | Children | Every | Provide | People |
| Free | Lives | America | Access | Jesus |
| Quality | Education | Healthy | Envisions | Need |
| Everyone | Envision | Without | Disease | Jews |
| Food | One | Families | Model | Work |





In the probability distribution for the vision statements, there appears to be more variation. However, this may be a visual artifact due to the significantly smaller number of organizations’ visions analyzed than missions or about statements.

# Lessons learned

A careful interrogation of the best unsupervised learning methods for this project is needed. More refined techniques, perhaps bi-topic modeling, is needed for the visions data due to the low number of terms overall, as well as the low number of unique terms in this data. A more careful evaluation of the differences in substantive insights from the About Us versus Mission statements is needed.