Milestone Report

Introduction

Across the country, many public school teachers use their own money to buy supplies for their classrooms, including basic essentials like notebooks and pencils. If schools and teachers cannot afford these supplies, students may miss out on learning opportunities. DonorsChoose.org was established to help teachers get the resources they need for classroom projects. Through the DonorsChoose.org website, teachers can request materials for their classrooms. Donors can browse through requests, and select a project they would like to support financially.

The goal of this capstone project is to provide guidance to teachers who are using DonorsChoose.org to fund their classroom projects. This will include suggestions on how to select a school project that has a greater chance of receiving funding. Information will be provided on how to structure essay requests, and the preferred tone of the requests to improve the chance of getting donations.

Dataset

Data for this capstone project was obtained from the DonorsChoose.org open datasets:

https://data.donorschoose.org/open-data/overview/

This includes information on donations, school projects, resources requested, school information, and written essays describing the projects.

Exploratory Data Analysis

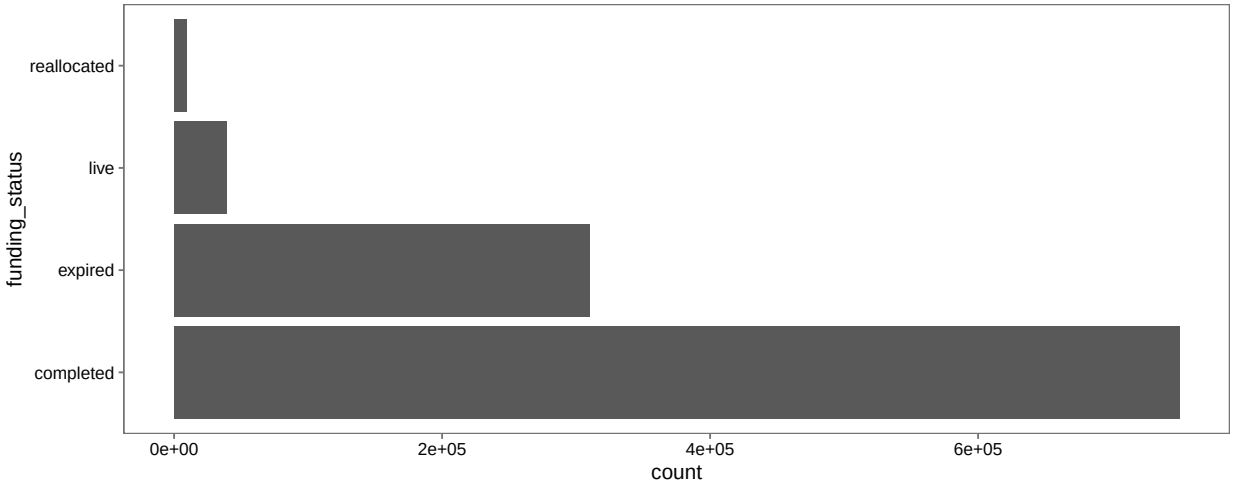
Data analysis was conducted using R. The datasets from DonorsChoose.org were cleaned, including changing data types, and reordering factor levels.

For the exploratory data analysis, the factors that may affect a project's chances of getting funding (including subject area, resources requested, school location, grade level, poverty level, etc.) were analyzed. Essay requests were examined using text sentiment analysis.

Project Data

Funding status

In the data set, there are four different options for funding status, including completed, expired, reallocated and live. Since the goal of this capstone project is to predict whether or not a project will be successfully funded (completed or expired), any live or reallocated projects will be excluded from the analysis.

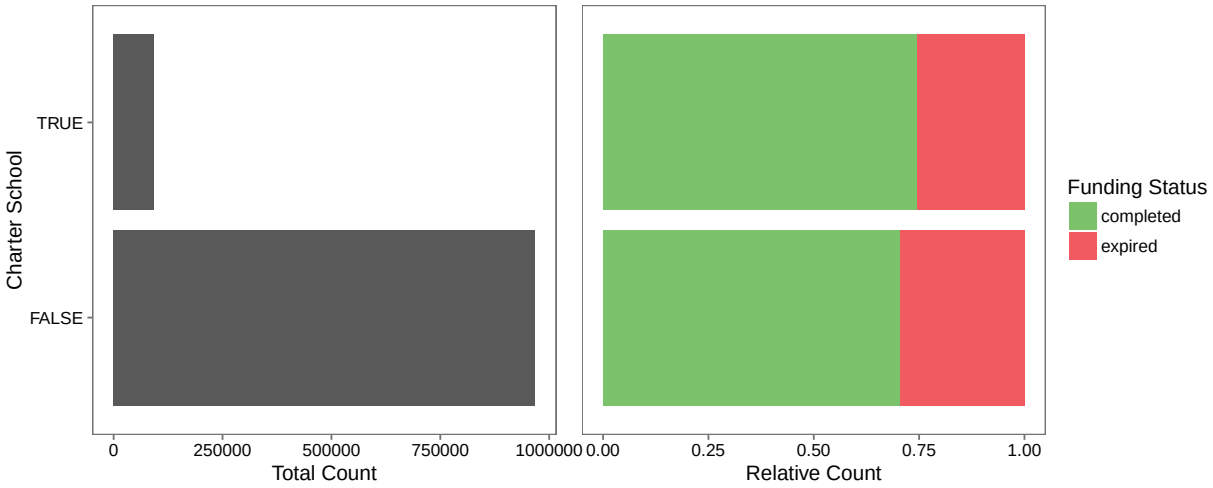


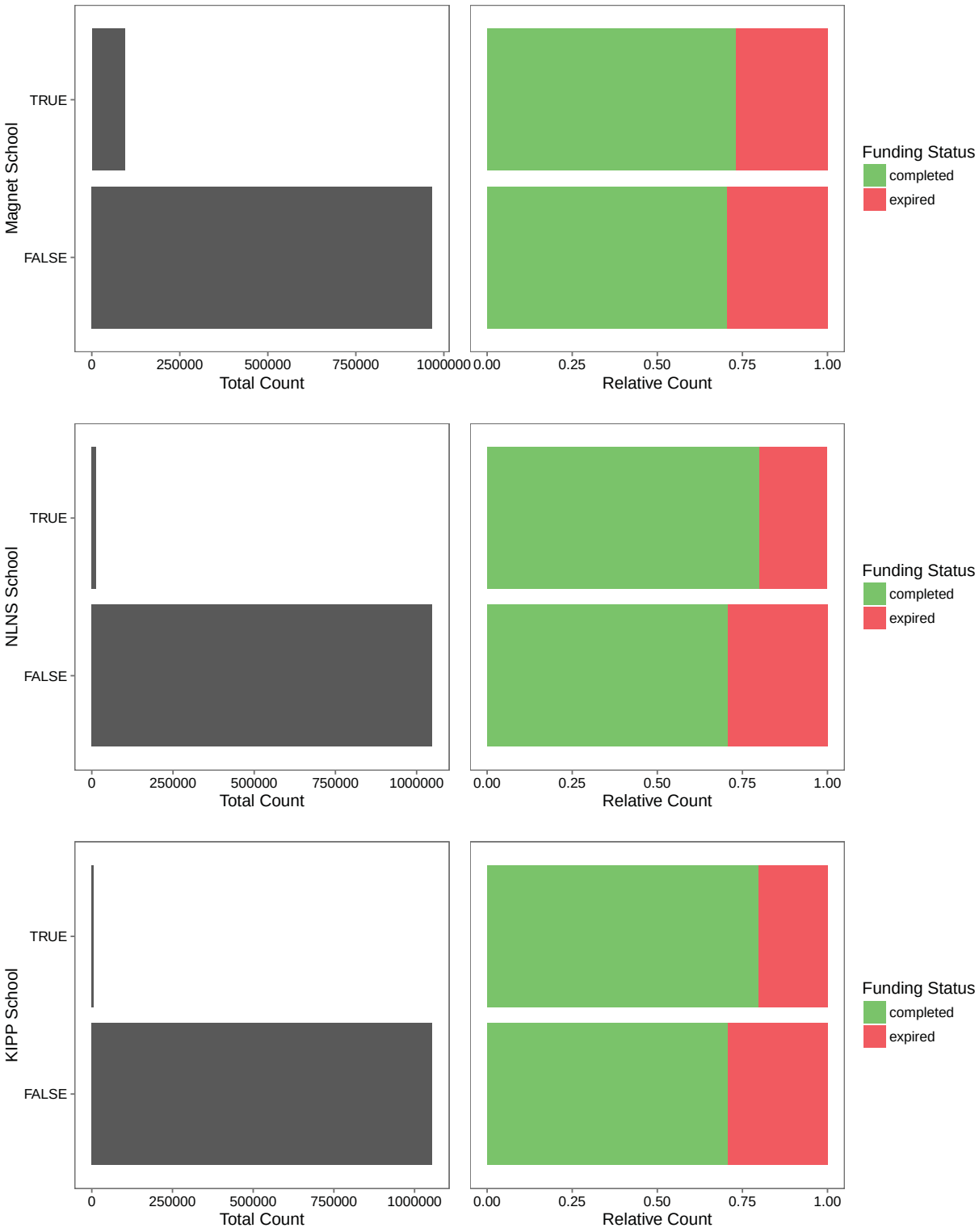
Type of School

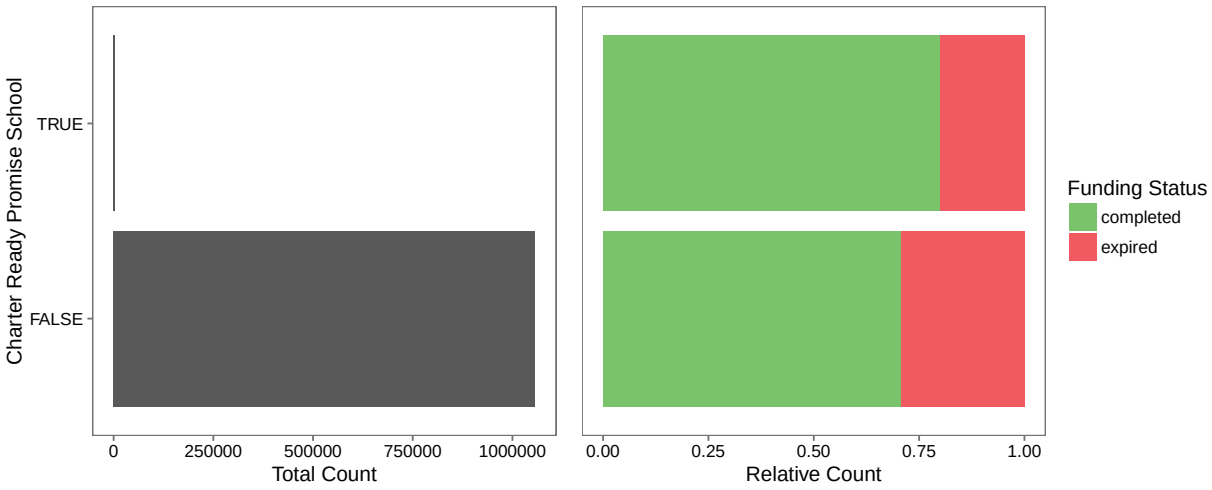
There are several properties which indicate the type of school, including:

- school_charter
- school_magnet
- school_year_round
- school_nlns (New Leaders for New Schools program)
- school_kipp
- school_charter_ready_promise

The relationship between each of these properties and funding status was examined.



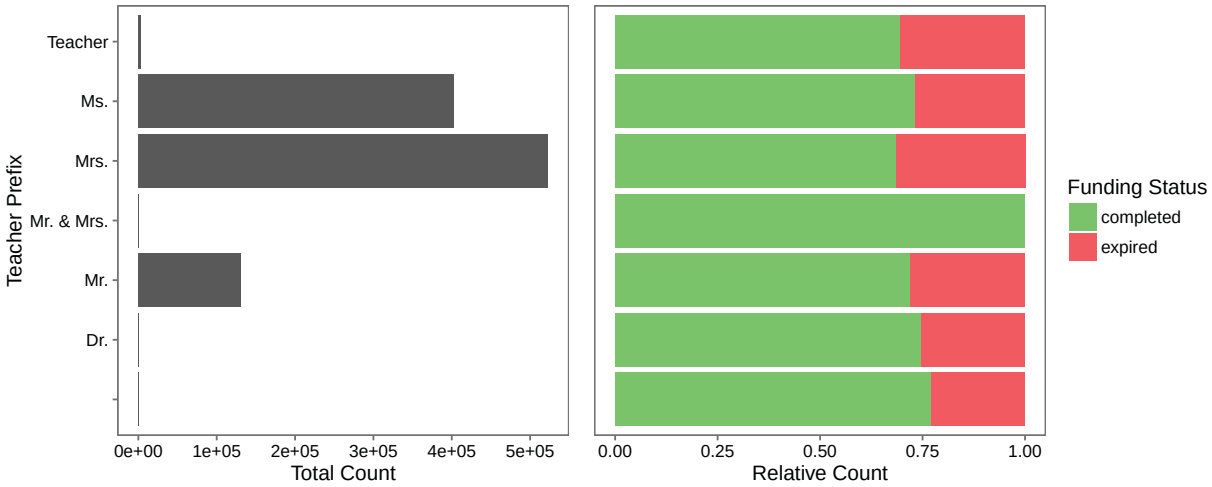

Magnet School
TRUE
FALSE
0
250000
500000
750000
1000000
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired
NLNS School
TRUE
FALSE
0
250000
500000
750000
1000000
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired
KIPP School
TRUE
FALSE
0
250000
500000
750000
1000000
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired

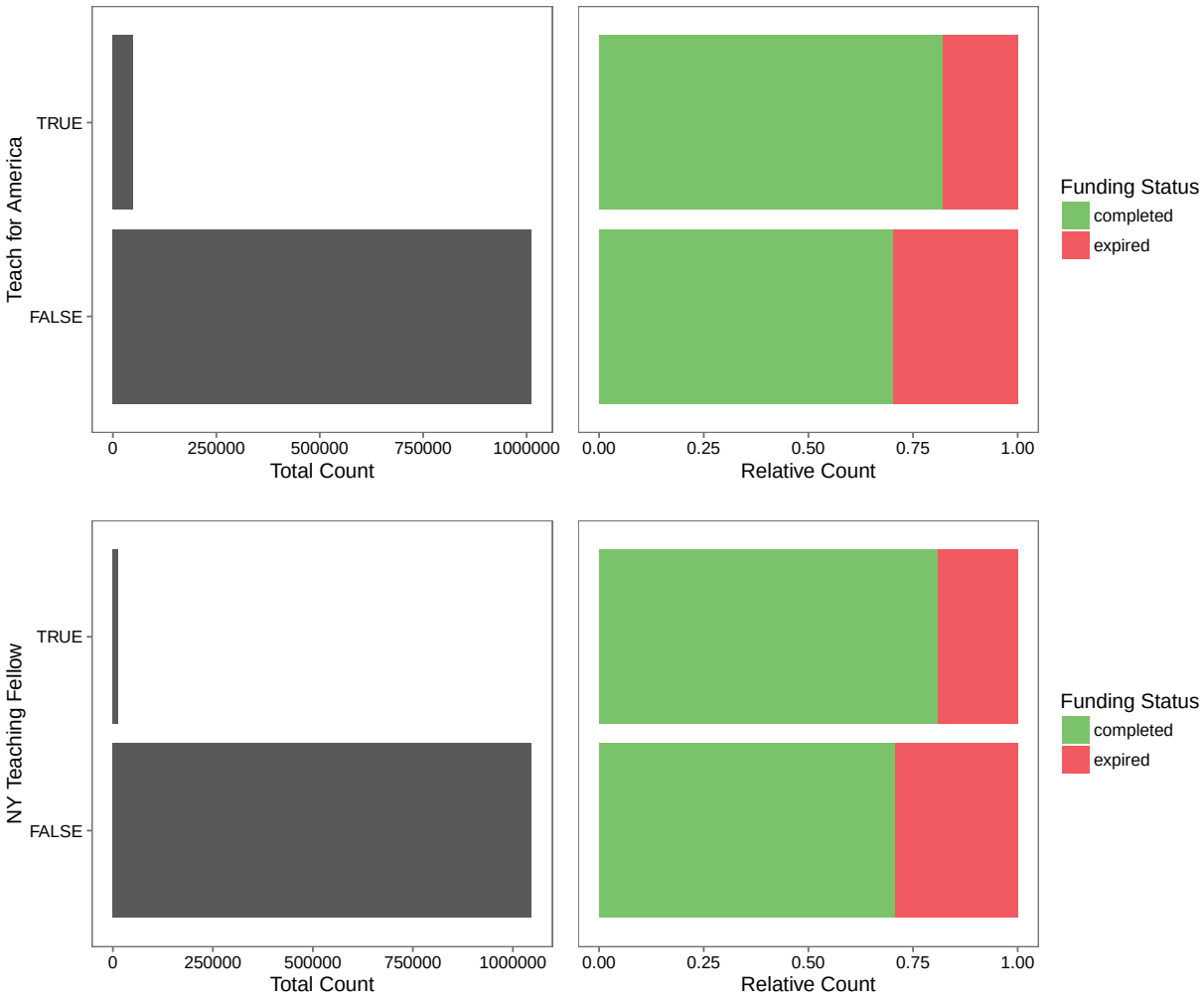


Teacher Attributes

The attributes of the teacher include:

- teacher_prefix
- teacher_teach_for_america
- teacher_ny_teaching_fellow

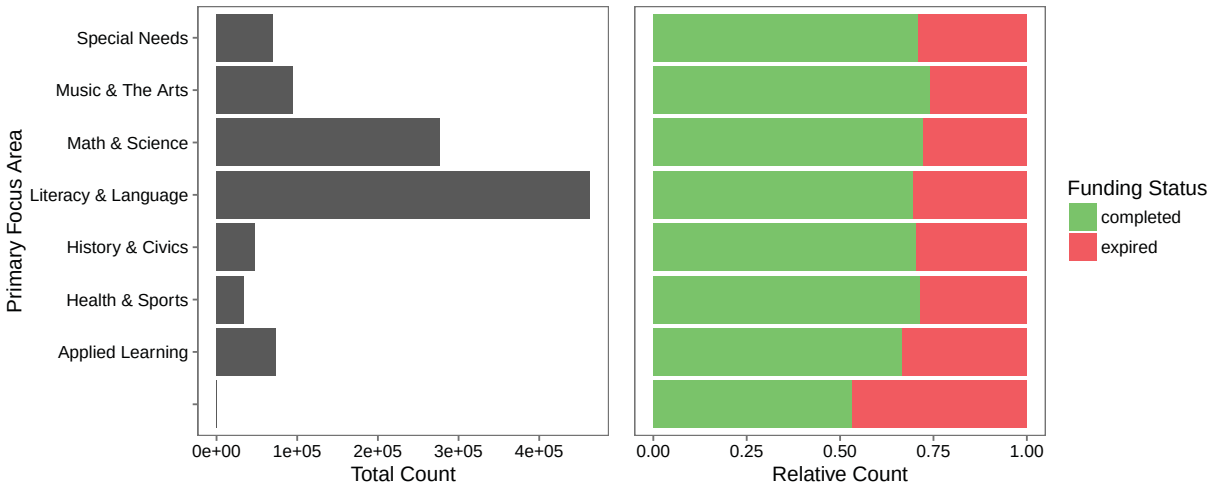


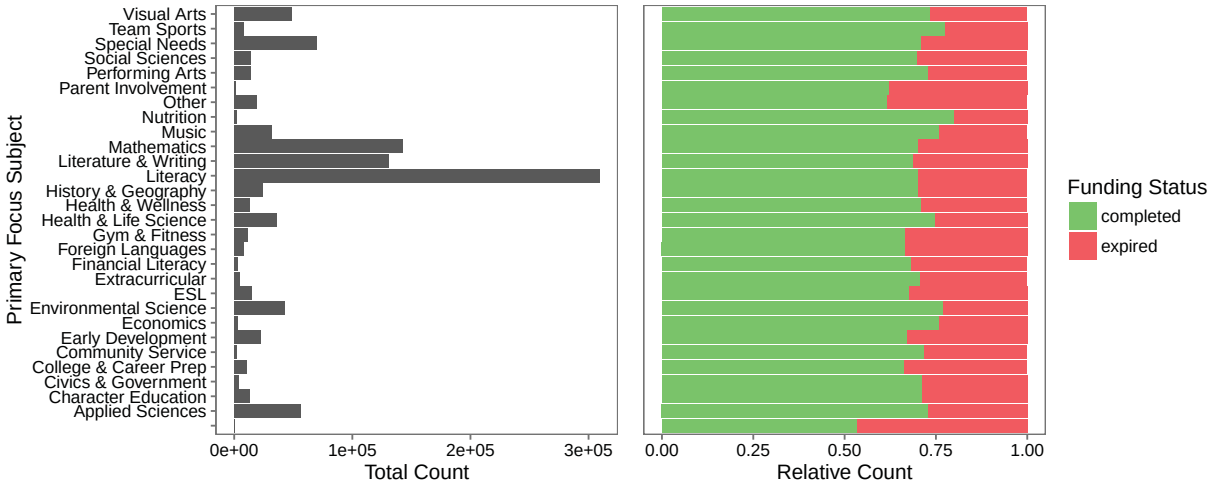


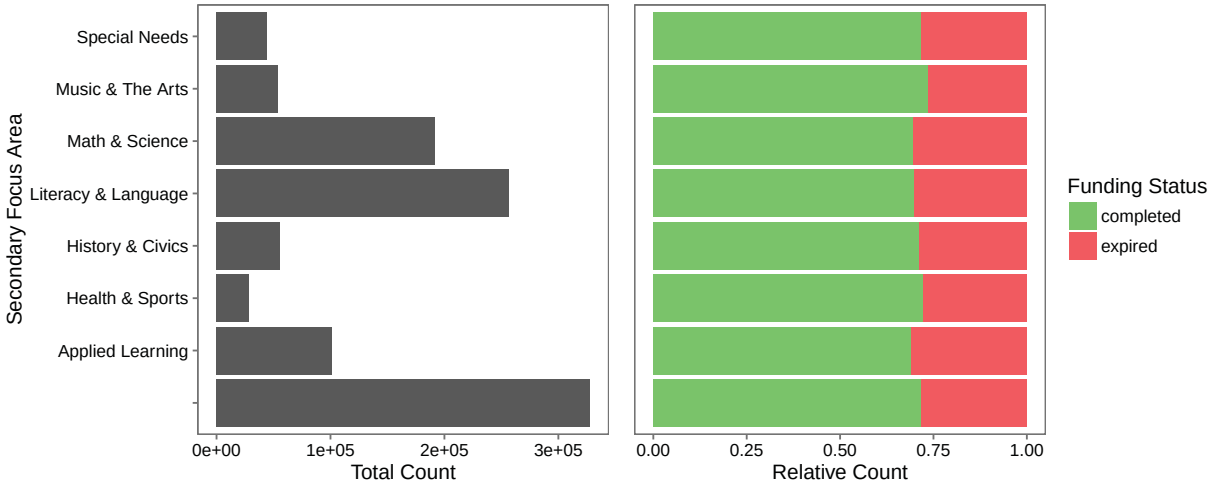
Project Categories

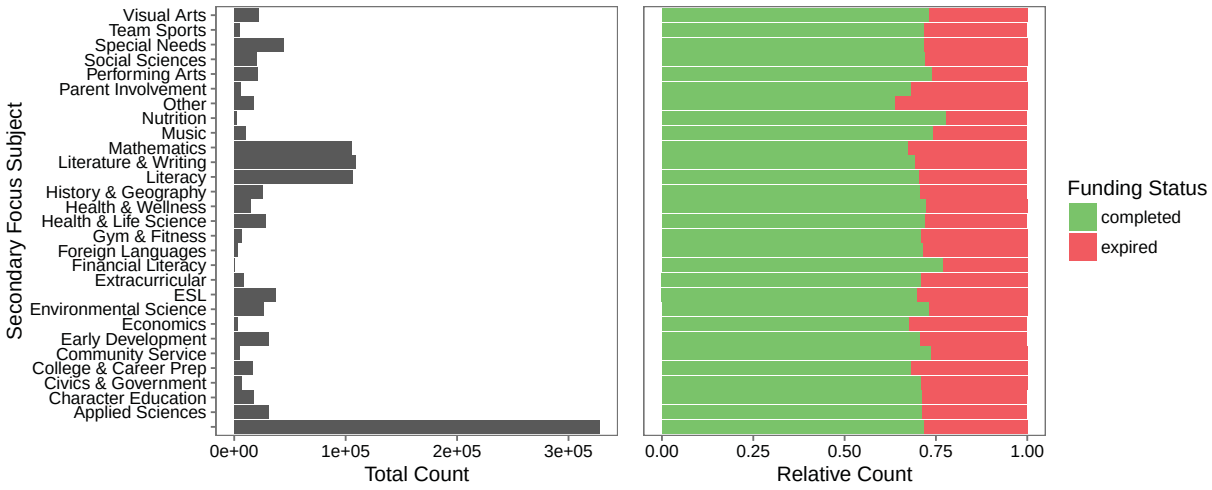
The project categories include:

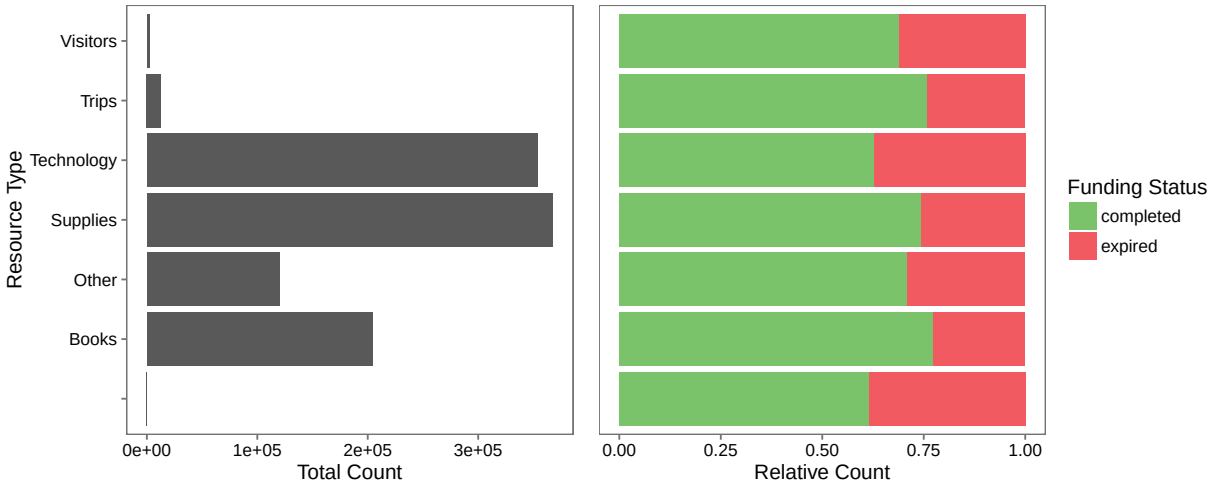
- primary_focus_subject
- primary_focus_area
- secondary_focus_subject
- secondary_focus_area
- resource_type
- poverty_level
- grade_level

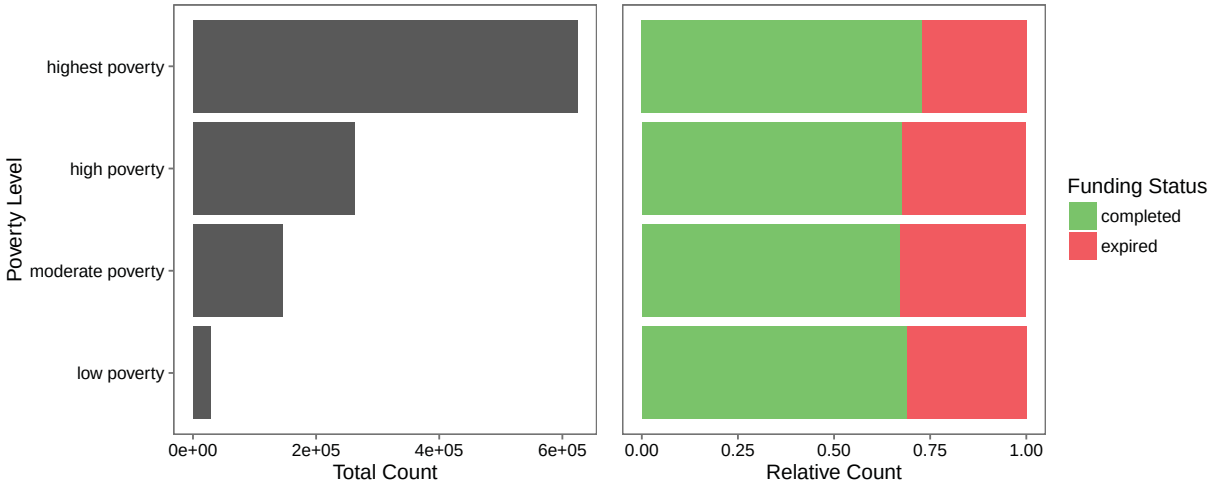

Primary Focus Area
Special Needs
Music & The Arts
Math & Science
Literacy & Language
History & Civics
Health & Sports
Applied Learning
0e+00
1e+05
2e+05
3e+05
4e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired

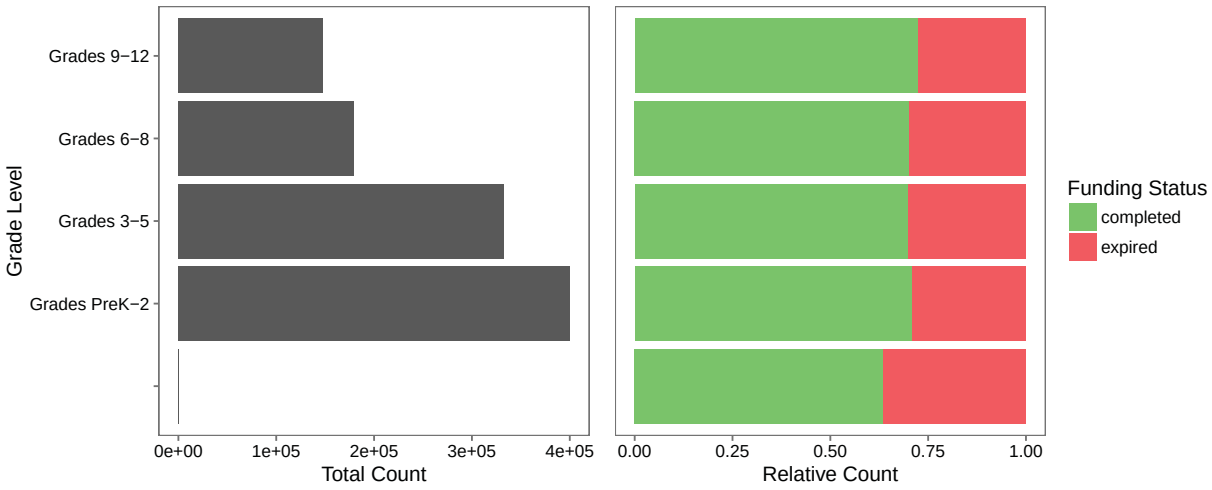

Primary Focus Subject
Visual Arts
Team Sports
Special Needs
Social Sciences
Performing Arts
Parent Involvement
Other
Nutrition
Music
Mathematics
Literature & Writing
Literacy
History & Geography
Health & Wellness
Health & Life Science
Gym & Fitness
Foreign Languages
Financial Literacy
Extracurricular
ESL
Environmental Science
Economics
Early Development
Community Service
College & Career Prep
Civics & Government
Character Education
Applied Sciences
0e+00
1e+05
2e+05
3e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired


Secondary Focus Area
Special Needs
Music & The Arts
Math & Science
Literacy & Language
History & Civics
Health & Sports
Applied Learning
0e+00
1e+05
2e+05
3e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired


Secondary Focus Subject
Visual Arts
Team Sports
Special Needs
Social Sciences
Performing Arts
Parent Involvement
Other
Nutrition
Music
Mathematics
Literature & Writing
Literacy
History & Geography
Health & Wellness
Health & Life Science
Gym & Fitness
Foreign Languages
Financial Literacy
Extracurricular
ESL
Environmental Science
Economics
Early Development
Community Service
College & Career Prep
Civics & Government
Character Education
Applied Sciences
0e+00
1e+05
2e+05
3e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired


Resource Type
Visitors
Trips
Technology
Supplies
Other
Books
0e+00
1e+05
2e+05
3e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired


Poverty Level
highest poverty
high poverty
moderate poverty
low poverty
0e+00
2e+05
4e+05
6e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired

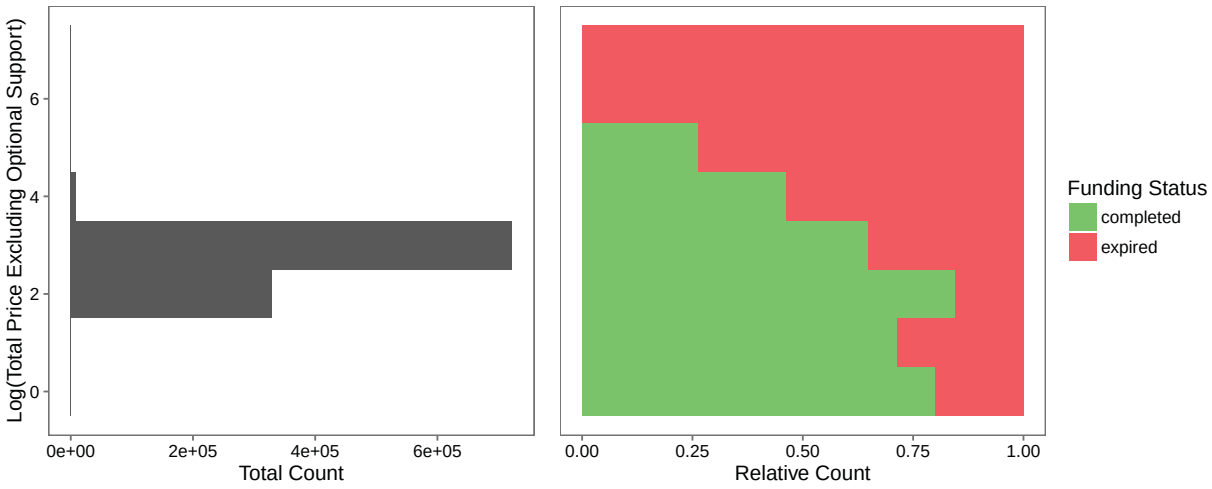


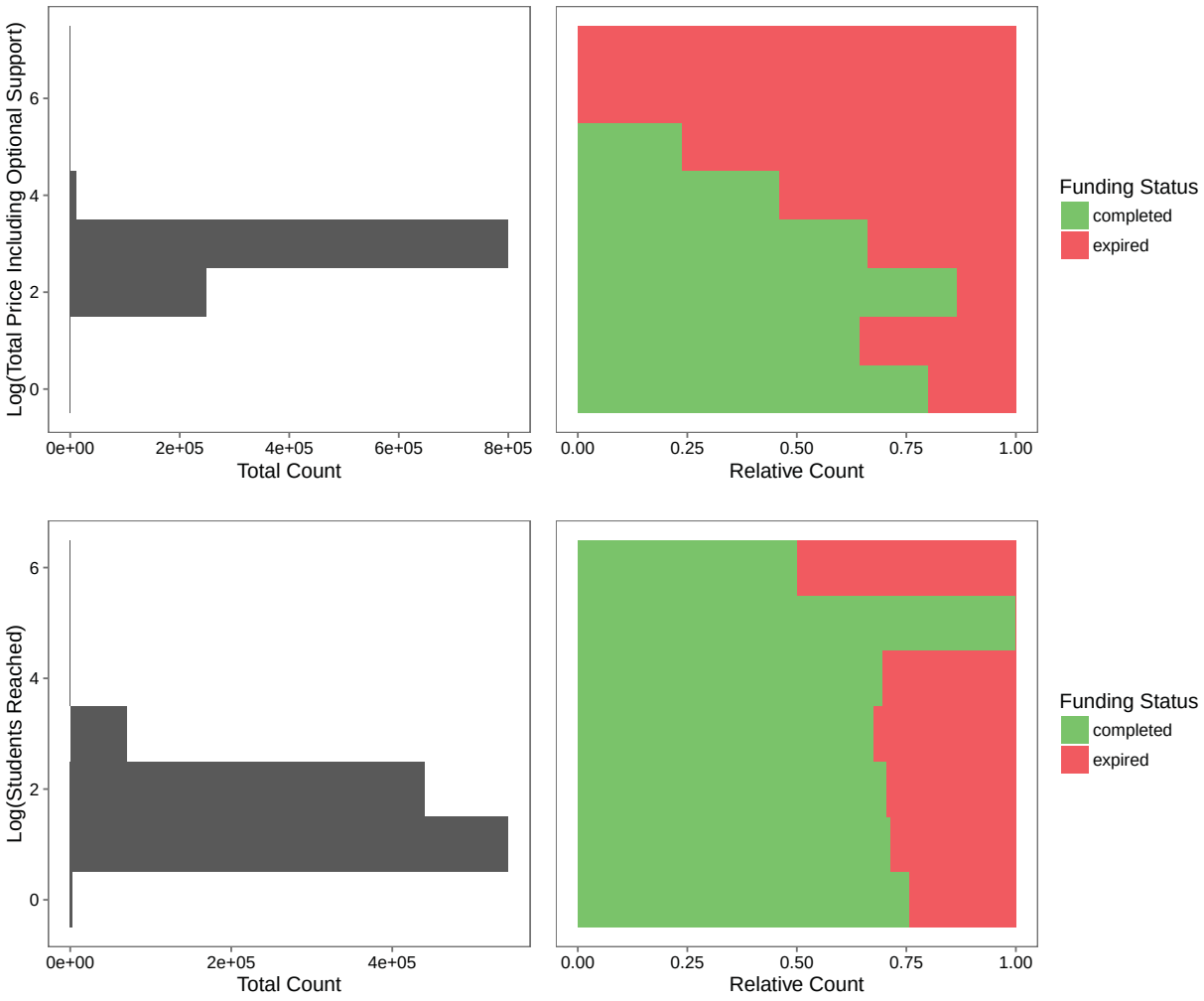
Project Pricing and Impact

These fields are available under Project Details on a project page, breaking down the cost of a project into its parts.

- vendor_shipping_charges
- sales_tax
- payment_processing_charges
- fulfillment_labor_materials
- total_price_excluding_opt_donation
- total_price_including_opt_donation
- students_reached

Some of these properties were examined in more detail:

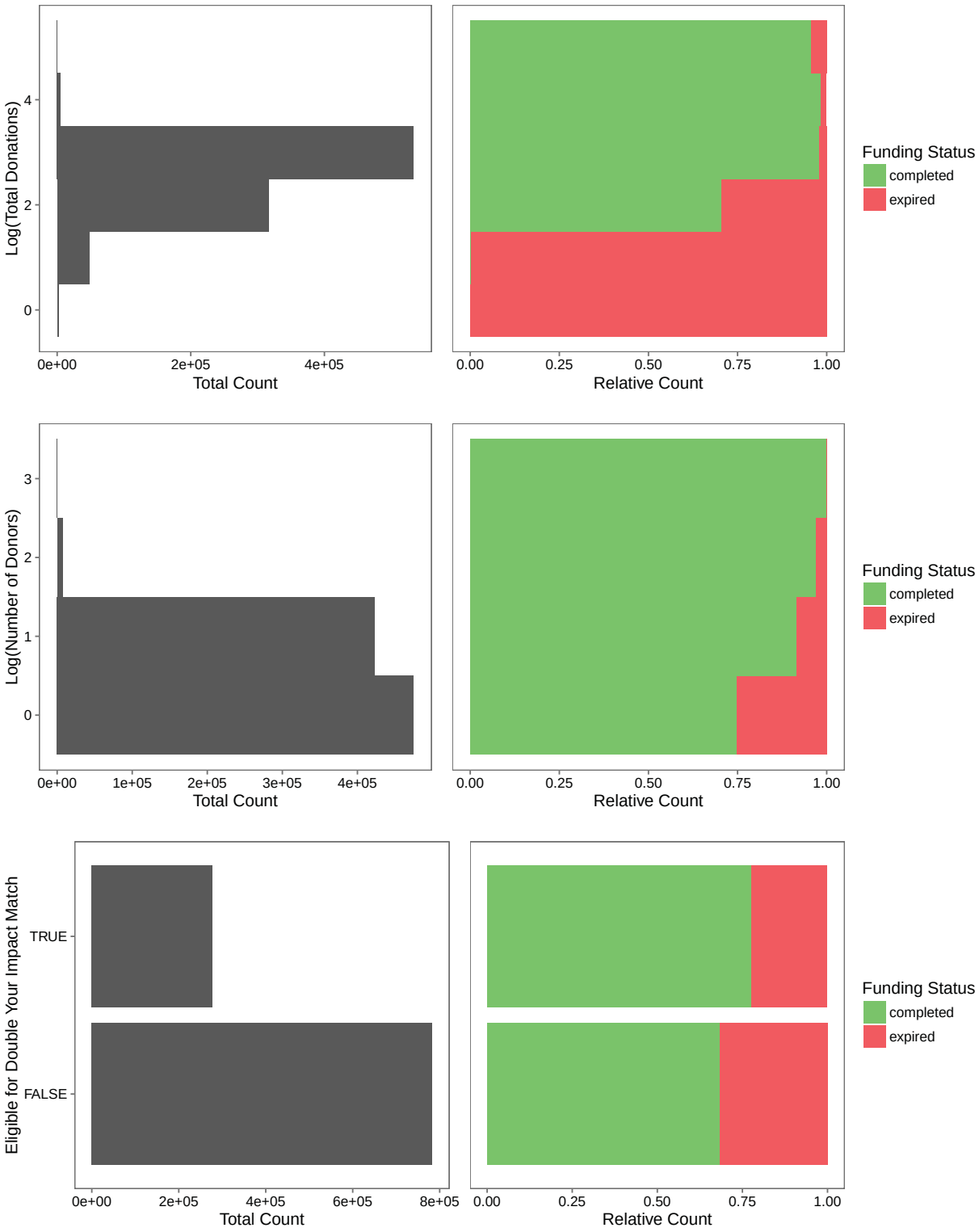


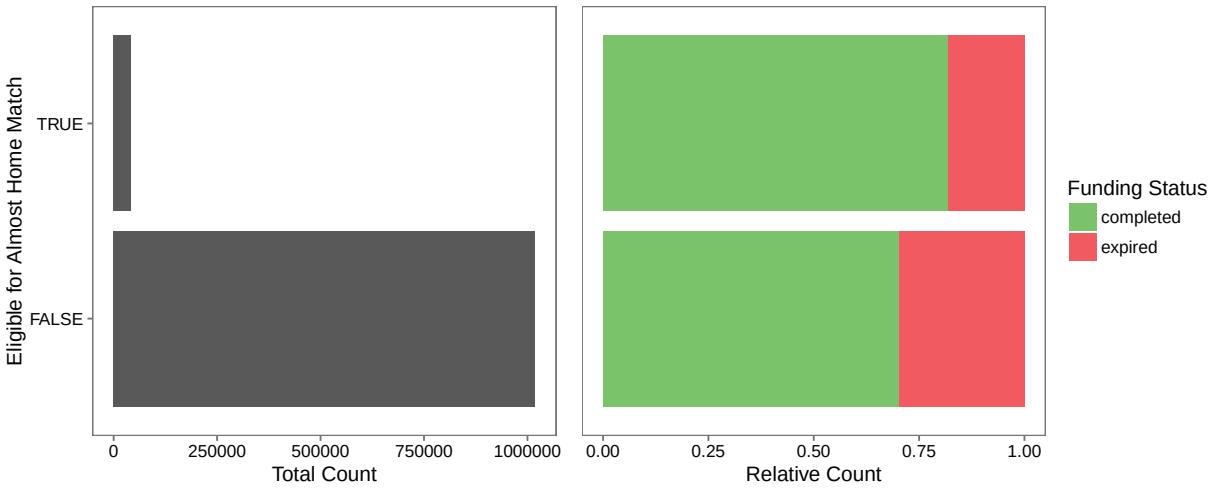


Project Donations

These fields provide information about the donations made to a particular project.

- total_donations: Total donation amount
- num_donors: Number of unique donors giving to this project
- eligible_double_your_impact_match: True if the project was ever eligible for a Double Your Impact match offer
- eligible_almost_home_match: True if the project was ever eligible for an Almost Home match offer

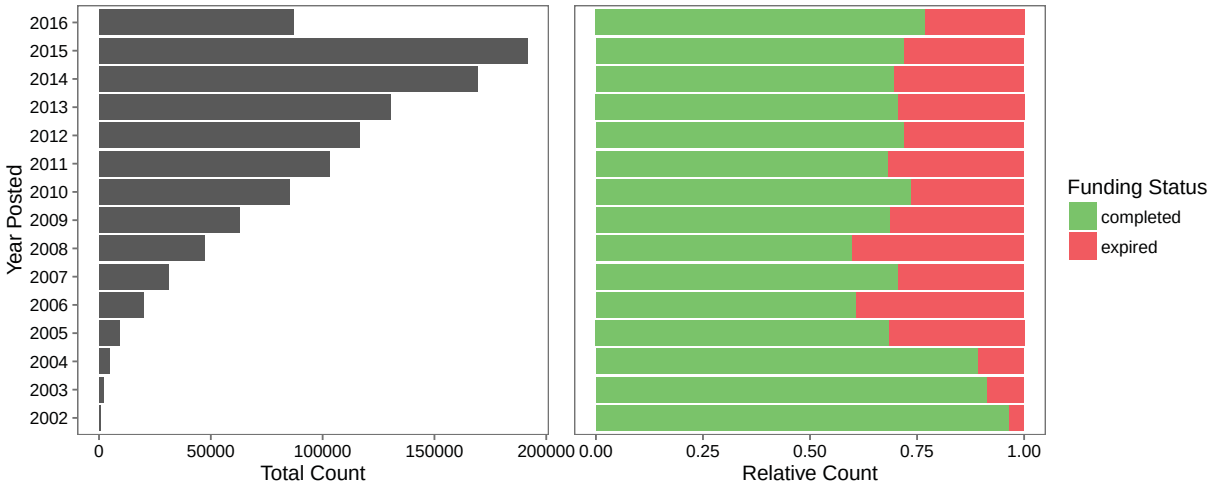

Log(Total Donations)
0
2
4
0e+00
2e+05
4e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired
Log(Number of Donors)
0
1
2
3
0e+00
1e+05
2e+05
3e+05
4e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired
Eligible for Double Your Impact Match
TRUE
FALSE
0e+00
2e+05
4e+05
6e+05
8e+05
Total Count
0.00
0.25
0.50
0.75
1.00
Relative Count
Funding Status
completed
expired



Project Status

These fields provide status information and dates for when a project reached their milestones.

- funding_status: Completed, Expired, Live, or Reallocated
- date_posted: Date a project was approved by staff to be listed on the site
- date_completed: Date a project become fully funded
- date_thank_you_packet_mailed: Date that the project went through the final stage of sending out a thank you package to donors
- date_expiration: Date the project was set to expire and be delisted from the site

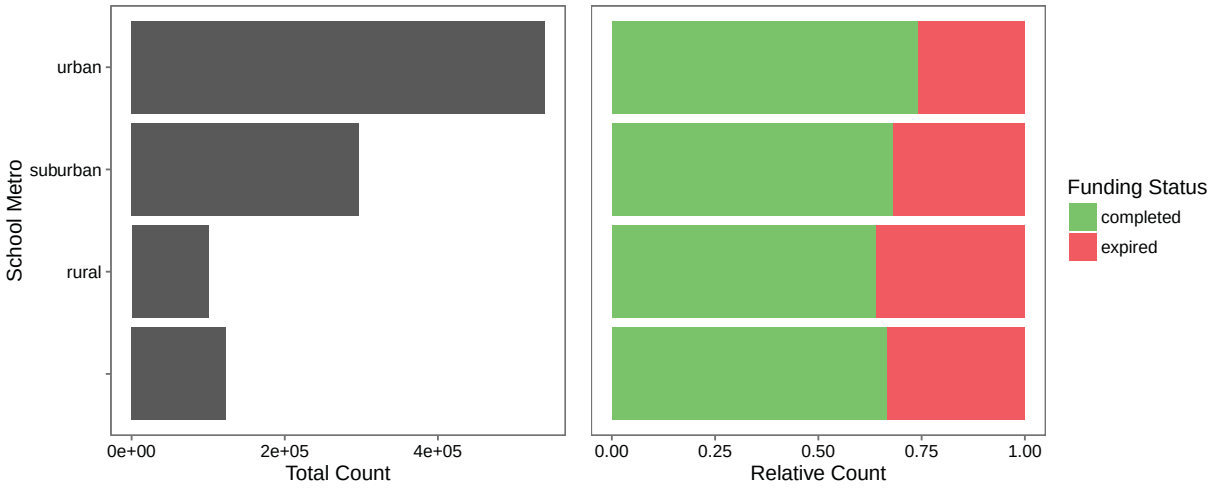


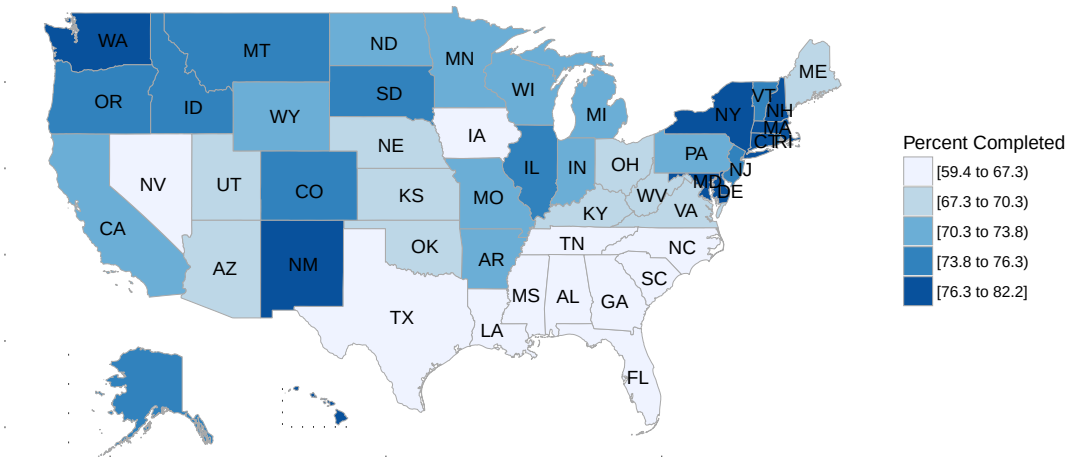
School Location

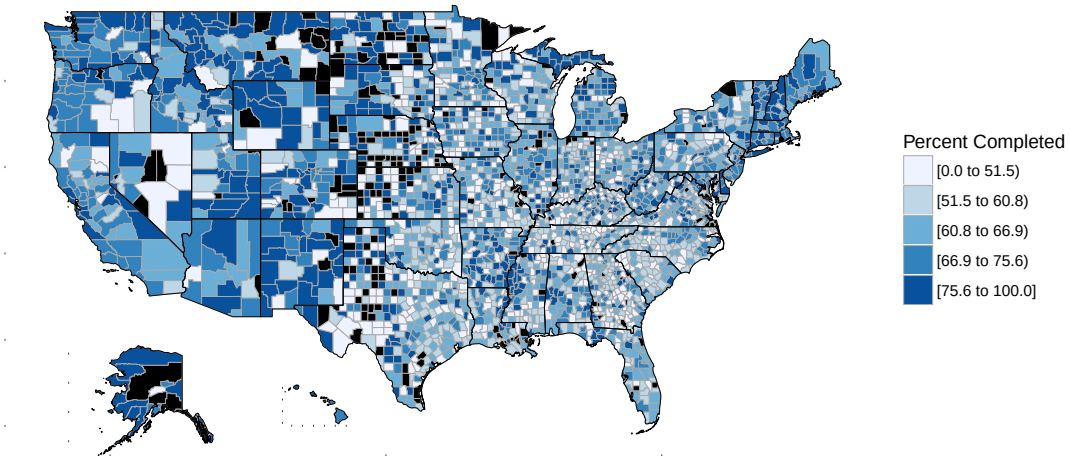
The fields describing school location include:

- school_latitude
- school_longitude
- school_city
- school_state

- school_zip: 5 Digit Zip
- school_metro: Urban, Suburban, or Rural
- school_district
- school_county





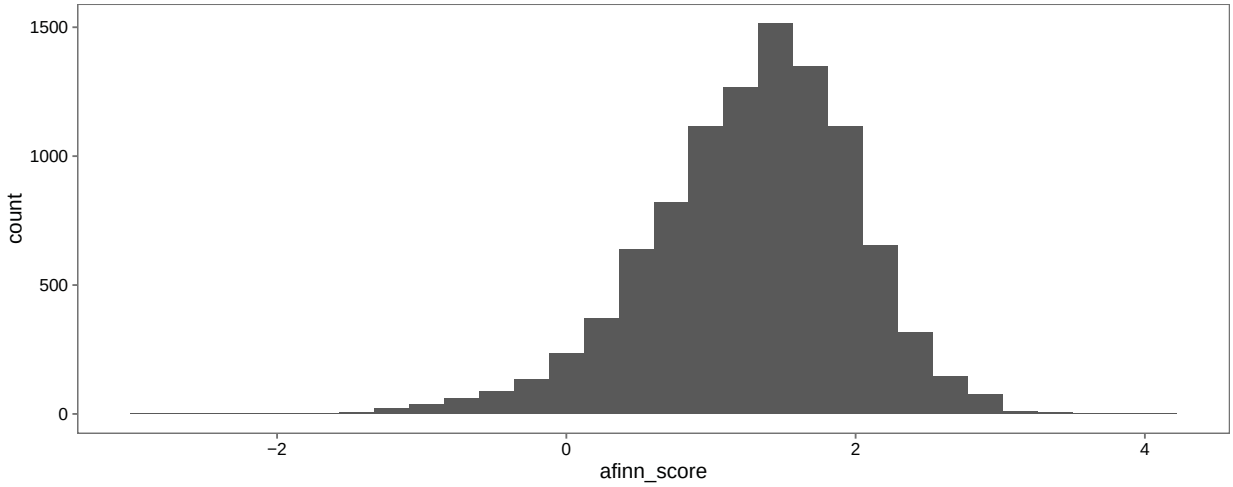


Essay Data

The project essay data includes the full text of the teacher-written requests accompanying all classroom projects. Since the dataset is quite large, the exploratory data analysis was limited to the first 10000 essays.

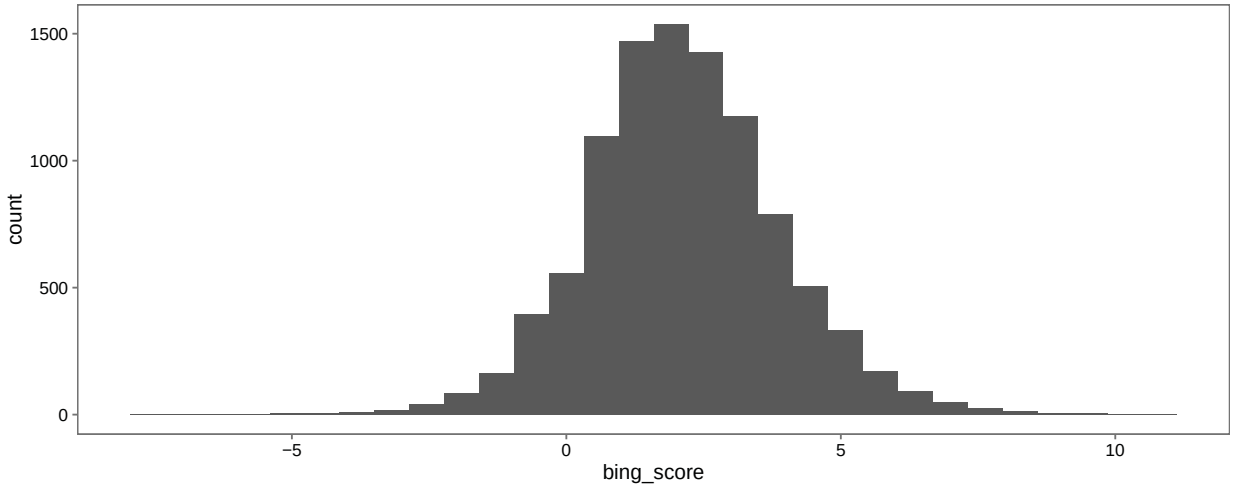
Sentiment Analysis using AFINN Lexicon

Sentiment analysis was performed on each essay using the AFINN lexicon, which provides a positivity score for each word, from -5 (most negative) to 5 (most positive). The positivity scores for each word in the essay were then averaged.



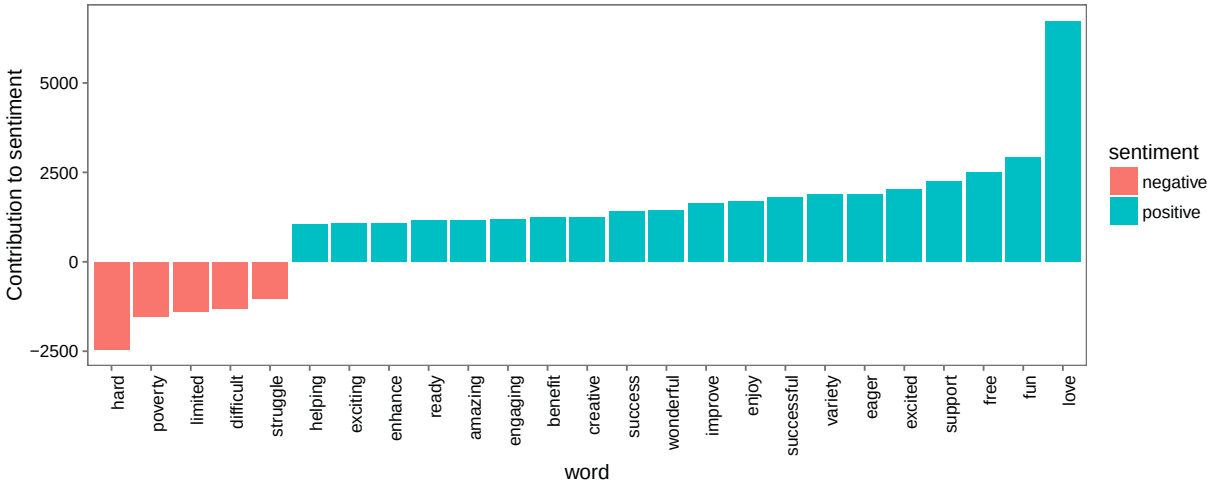
Sentiment Analysis using Bing Lexicon

Sentiment analysis was also conducted using the Bing lexicon, which rates words as positive or negative. The difference between the number of positive and negative words in each essay was then calculated.



Most common positive and negative words

The bing lexicon was used to figure out the most commonly used positive and negative terms.



Next Steps

The next steps for this project include:

- Project and essay data will be joined into a single data frame using the projectid as the common variable between the two data sets.
- Due to the extensive size of the data sets (over 1 million projects and essays) and limited computational resources (not enough RAM), the data analysis will likely need to be limited to a smaller date range.
- A model will be developed to predict the likelihood of funding based on these factors.