Final Project Design Review: Donation statistics

Xinwen Liu (xinwenl), Xinyu Lin (xinyulin), Yuxi Luo (yuxiluo), Shaobo Guan (shaobog)
https://github.com/CMU-IDS-2020/fp-vectorization
November 20, 2020

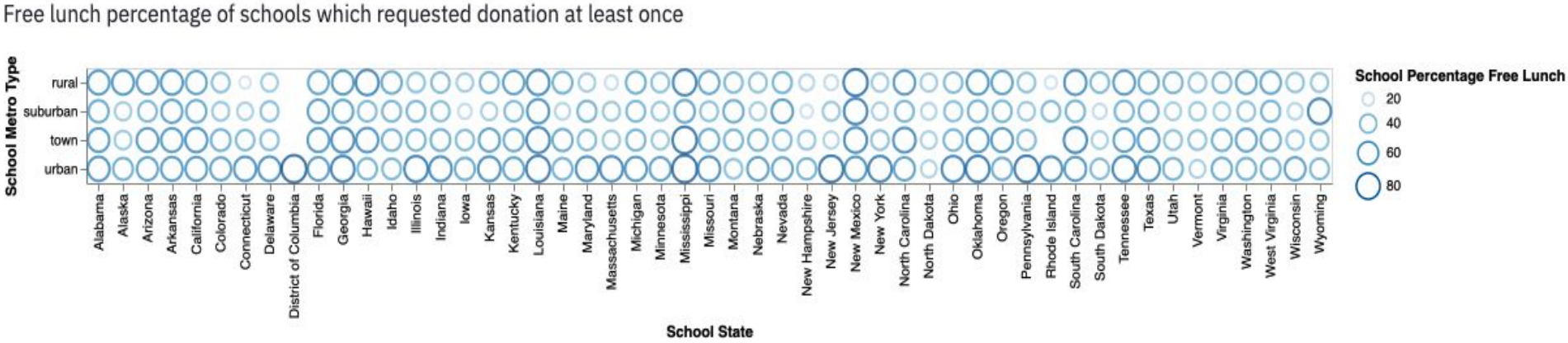
Overview

- Track:
  - Narrative
- Problem:
  - What is the current state of funding for public schools?
- Data:
  - DonorsChoose.org: a platform that hosts teachers' project proposals to seek for donations
- Interactive System:
  - Visualization on general donation statistics
    - Help the donors collect and understand current situation and make effective donations
  - Model prediction and analysis on project fully-funded status
    - Explore potential factors for projects to get fully-funded
    - Help teachers to formulate better project proposal

Visualization -- Free lunch percentage

for donors

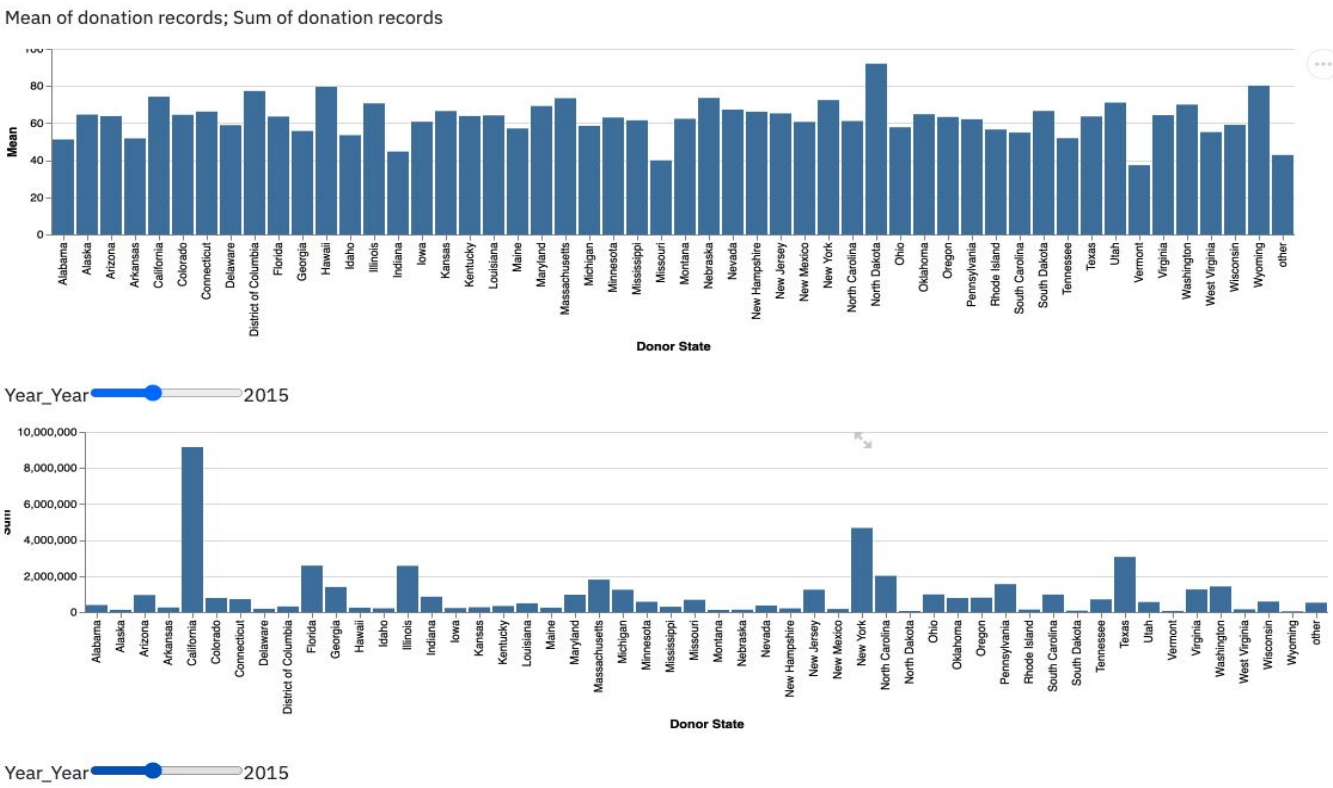
- Current visualization:
  - The percentage of free lunch shown by the size of bubbles
  - Categorized based on metro types and states



Visualization -- Donations given out

for donors & platform

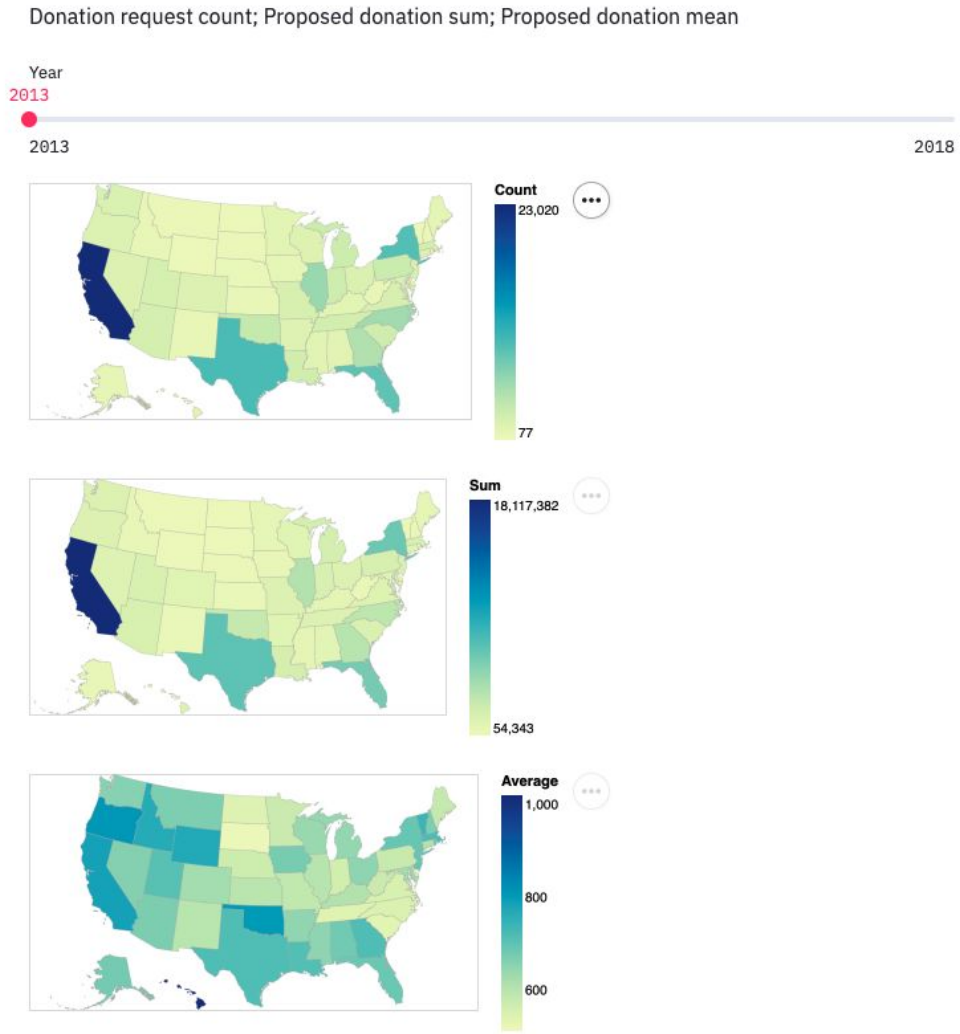
- Current visualization:
  - Donation records
  - Based on mean & sum
  - Shown by years



Visualization -- Donation requests

for donors

- Current visualization:
  - Donation request info geographically
    - Based on count, sum & average
    - Shown by years
- Future work:
  - Combine the maps into one big interactive map if possible



Visualization -- Resource needed

for donors

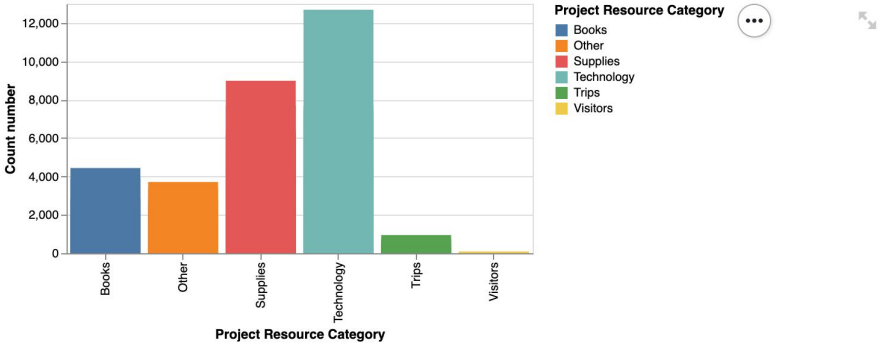
Explore the needs by

Project Resource Category

See the change of counts by year

2015

2013 2018



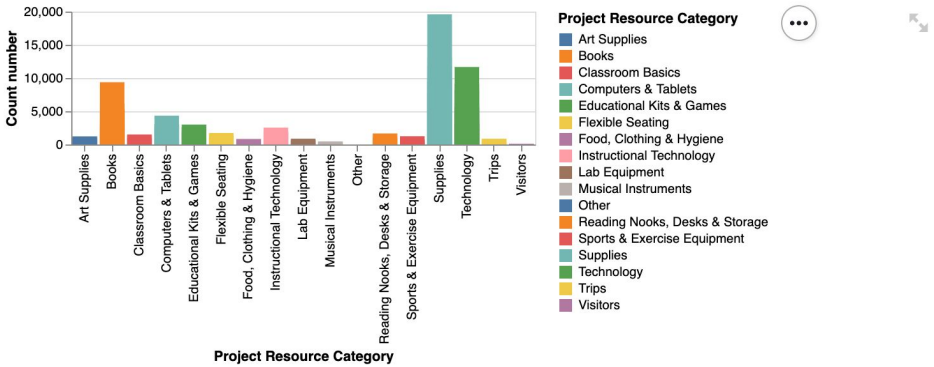
Explore the needs by

Project Resource Category

See the change of counts by year

2017

2013 2018



We could see the changes of resource needed based on year change

Visualization -- Grade level

for donors

It’s also interesting for donors to learn which grades have the most needs

Model prediction and analysis

- Purpose:
  - Understand what factors contribute to the probability of projects to be fully-funded
  - Help teachers predict whether the project proposal can be fully-funded
- Model:
  - Logistic Regression
- Current features:
  - Project description features:
    - length, number of “?!”, numerical expressions, 85 words of high frequency
  - Project cost, year, valid duration, subcategory, resource category
- Current results:
  - 64.8% accuracy on a balance dataset

Model prediction and analysis

- Sample project description:

  Imagine having tables and having to stand to **play** and complete **work activities**. I **teach** preschool special education in an urban **school** district. Most of the **students** are on **free** and reduced **lunch**. My **students** are between the ages of 3-5 years old. My **students** have a variety of disabilities including Autism, Down Syndrome, and language impairments. With the chairs, my **students** will be able to sit down to **play** with table toys like Legos and stringing beads. My **students** will also be able to sit to complete **work activities** such as cutting with scissors, tracing their names and completing letter and number crafts. Donations will **help** my **children** sit and concentrate on their **work**. My **students** will be better prepared for kindergarten. They will be able to sit and learn basic **skills** such as colors, shapes, numbers and letters.

Model prediction and analysis

- Future works:
- Improve the model accuracy
  - Refine current features
  - Incorporate more data
- Present the model to users
  - Fully-funded status prediction:
    - Allow users to indicate whether they want to donate on a given project proposal, and compare the user's choice with the predicted fully-funded status
  - Project proposal refinement:
    - Allow users to input a project proposal, show model prediction and highlight the parts that contribute positively/negatively towards the fully-funded status
    - Potentially help teachers to formulate better project proposal

Questions for the course staff

- Visualizations:
  - What dimensions of the data should be further investigated?
- Model:
  - Should we focus on presenting the model and building interactions based on the model, or improving the model performance?
  - Any recommendation for our way of presenting the model?
  - What are some potential approaches we could apply to better improve our model?

Thank you for listening!