

Project 1 Proposal

Min Su Kim, Zach Kamran, Varshant Dhar

The central idea for the project is to use a dataset available online by DonorsChoose.org, an online charity that makes it easy to help students in need through school donations. We want to visualize the success of the projects throughout the country based on the school poverty levels. We wish to understand whether or not a particular school's funding level somehow influences the outcome of the project proposals.

1. Dataset

The data was released on Kaggle, an online platform for predictive modelling and analytics with a wide variety of datasets uploaded by companies and users. The files contain a wide range of information, of which we will use:

- donations.csv - contains information about the donations to each project
- projects.csv - contains information about each project, including school location and poverty level
- resources.csv - contains information about the resources requested for each project
- outcomes.csv - contains information about the outcomes of projects

The website explains that the projects starting January 2014 may contain inaccurate information, as these may still be live. Therefore, our visualization will focus on projects prior to this cut-off date. Preprocessing will be minimal and will only require averaging based on location for poverty level data.

2. Visualization

Our ideal end visualization will have a color scale (e.g., red-to-blue) to represent the poverty level, each school represented by either a triangle or a circle on a map of the United States. The triangle corresponds to projects that failed, and the circle represents those that succeeded, such that the viewer will be able to distinguish the following:

- The general trend - does the poverty level somehow influence the success of the projects?
- Map saturation - which areas were more funded? Which areas were more likely to succeed?

3. Additional Comments

Depending on the level of saturation and overall distribution, we may choose to limit the mapped region to a single city, since the dataset is quite large. Additionally, there are other factors that we could include or switch to represent, such as excitement factor (based on the criteria

provided by Kaggle). However, any possible changes to the project would not largely detract from the outline presented above.