

Molly Braun
Experiment Write-up
March 2, 2017

Data

The data is from the 2014 Public Library Survey, the most recent available survey from the Institute of Museum and Library Services. Three CSV files are publicly available, broken down by state, administrative entity, and library outlet. I will be using the administrative entity data set since it's the most comprehensive. There are 9,305 entries and 159 variables.

The data dictionary can be found here, starting at page 24 of the pdf.

https://www.ims.gov/sites/default/files/fy2014_pls_data_file_documentation.pdf

The Institute of Museum and Library Services releases an annual report to summarize the findings of each survey. The 2014 report has not been released yet, but reports from previous years are available.

Problem

I would like to predict library circulation numbers. Since this would be a continuous variable, I think I would start with a linear regression. I would like to use characteristics of the service area and library in the model. The service area population, budget, and the size of the collection are all provided in the dataset.

The dataset also includes census tracts. I'd like to use those to look up census data and include AMI in the model as well.

The dataset also includes children's and electronic materials circulations. If I have time, I'd like to see if these numbers can be predicted as well, either from the same variables as the total circulation or others from the dataset or census data.