
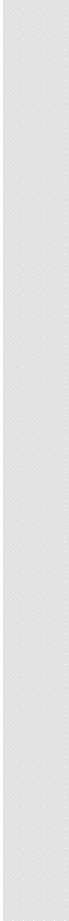


Library Circulation

Molly Braun

- 
- 
- PROBLEM: Can circulation be predicted?
 - HYPOTHESIS: Library system circulation can be predicted using library and geographical characteristics.

Data

- Institute of Museum and Library Services Public Library Survey (PLS)
 - Survey results of 9,305 library systems nationwide
- 2014 American Community Survey 5-year Estimates
 - Area Median Income (AMI) by state

Data

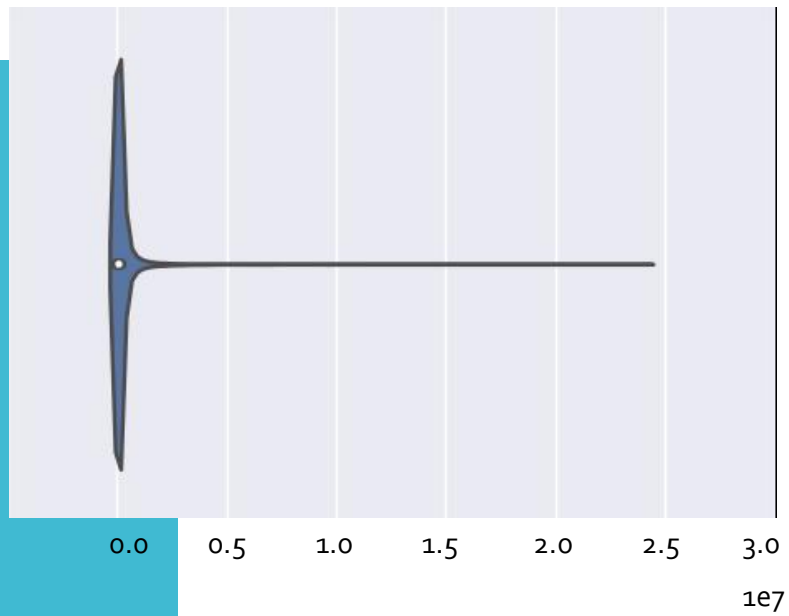
Variable Name	Description	Type of Variable
TOTCIR	Total annual circulation	Continuous
POPU_LSA	Population of the legal service area	Continuous
STINCOME	Household median income (dollars)	Continuous
BKVOL	Print materials	Continuous
EBOOK	Electronic books	Continuous
STGVT	Operating revenue from state government	Continuous
FEDGVT	Operating revenue from federal government	Continuous
LOCGVT	Operating revenue from local government	Continuous
TOTSTAFF	Total paid full time employees	Continuous
GEOCODE	Geographic code	Categorical
OBereg	Region	Categorical
LOCALE	Geographic location	Categorical

Data

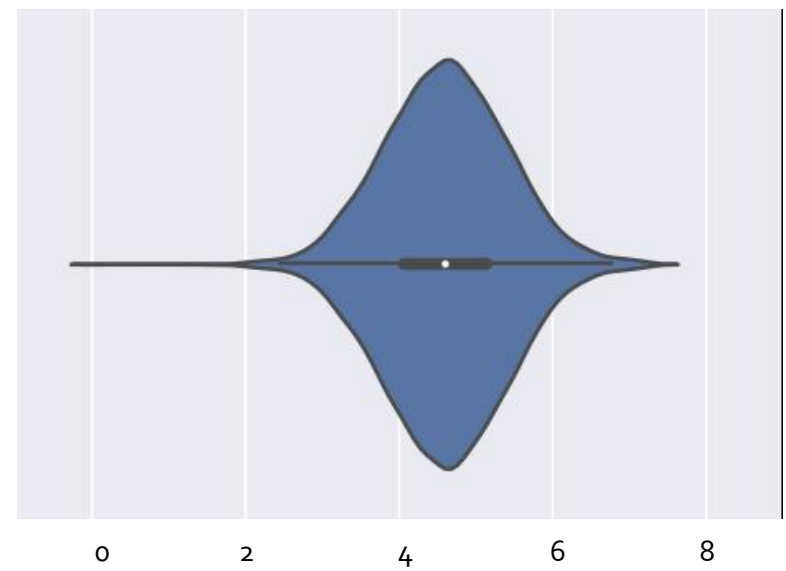
- Dropped missing income data
- Imputed median and mode for library data
- Log of continuous data because of skew
- Dummy variables for categorical data

Data

Total Circulation, original



Total Circulation, transformed



Model

- Two sets of variables predicting:
 - 1: Total circulation
 - 2: Circulation per capita
- Two types of models
 - Linear regression
 - Random forest regressor

Model 1: Total Circulation

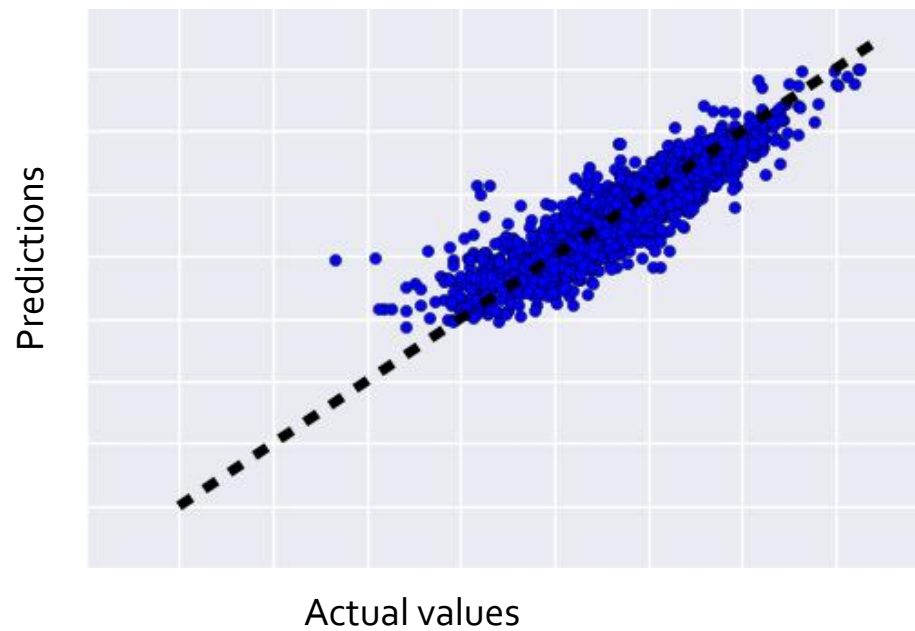
- Input Data
 - Population of service area
 - State income
 - State government revenue
 - Geographic code
 - Geographic location
 - Region

Model 1: Total Circulation

- Random Forest Regressor
 - Best estimator: Max depth = 27, Min sample split = 30
 - Mean squared error = 0.139

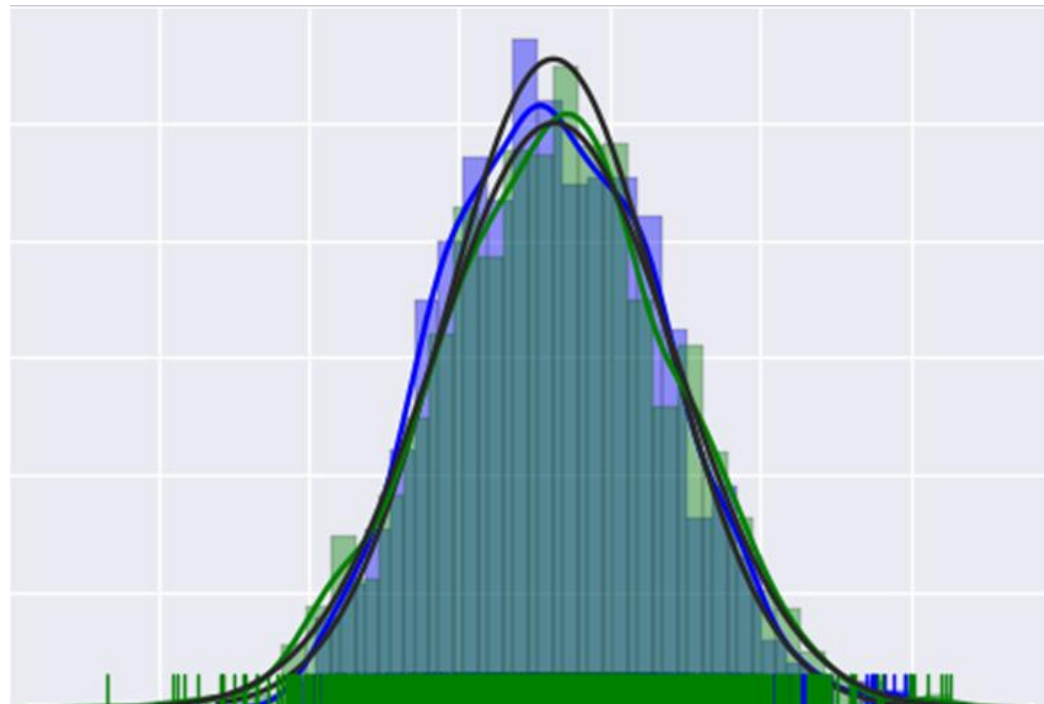
Model 1: Total Circulation

Total Circulation: Predictions vs. Actual Values



Model 1: Total Circulation

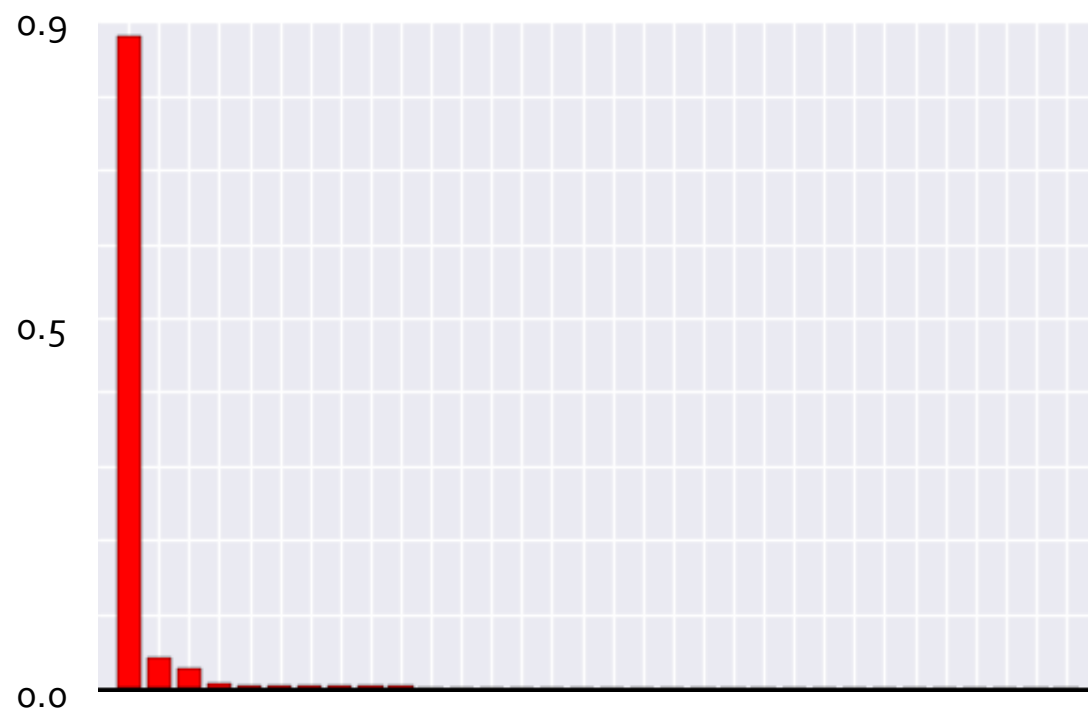
Total Circulation: Predictions and Actual Values



Green=actual values
Blue=predictions

Model 1: Total Circulation

- Feature importance
 - Most important: Population of service area



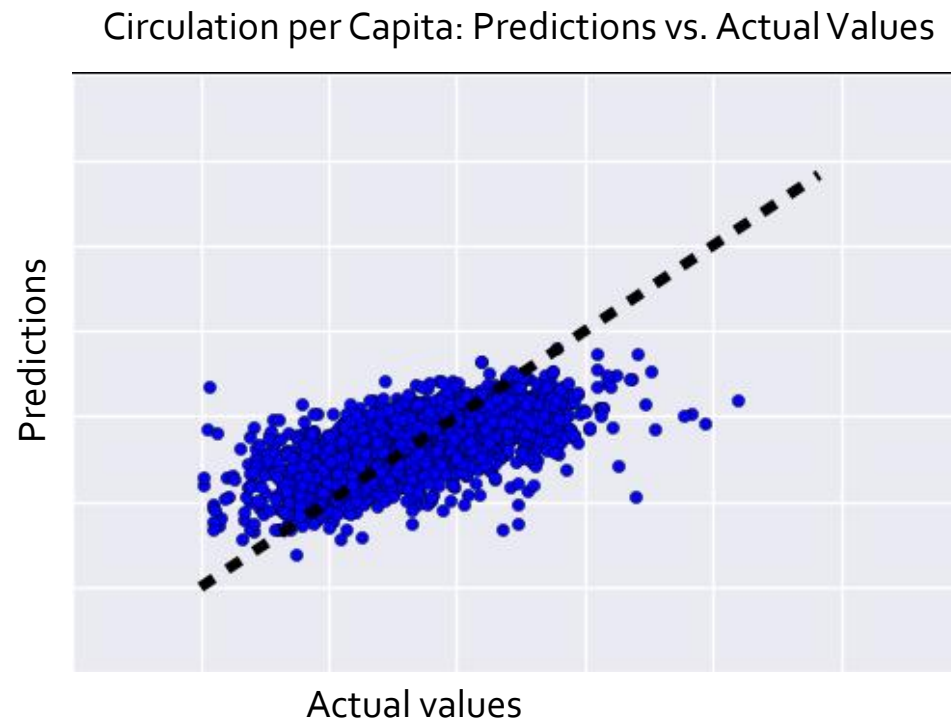
Model 2: Circulation per Capita

- Input Data
 - State income
 - Print materials
 - Electronic books
 - State government revenue
 - Federal government revenue
 - Local government revenue
 - Full time employees
 - Geographic code
 - Geographic location
 - Region

Model 2: Circulation per Capita

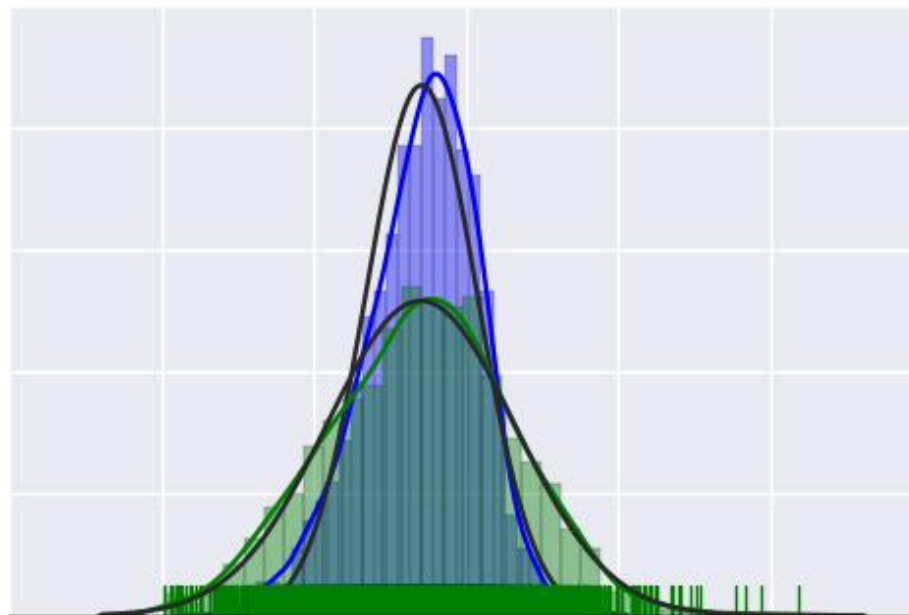
- Random Forest Regressor
 - Best estimator: Max depth = 22, Min sample split = 20
 - Mean squared error = 0.061

Model 2: Circulation per Capita



Model 2: Circulation per Capita

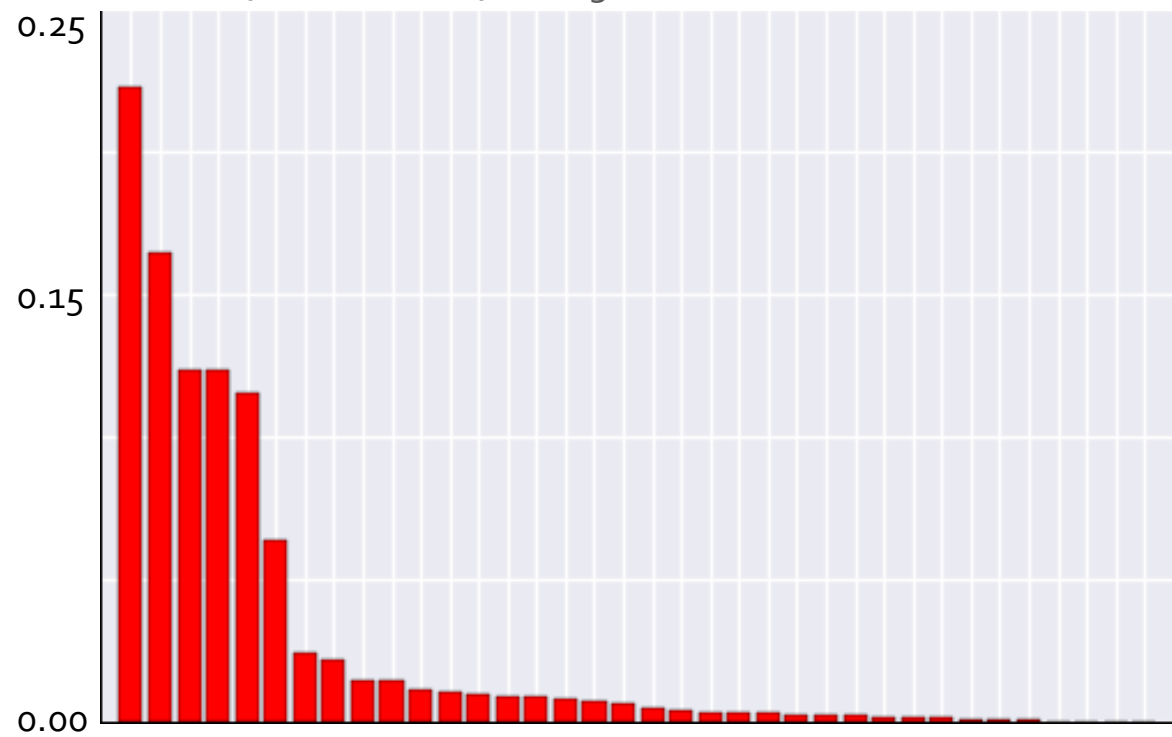
Circulation per Capita: Predictions and Actual Values



Green=actual values
Blue=predictions

Model 2: Circulation per Capita

- Feature importance
 - Most important: electronic books, state income, local government revenue, book volume, state government revenue



Results

- Random forest regressor had lower mean squared error for both sets of models
- Population of service area had a big effect on total circulation

Next Steps

- Add more demographic data
 - Age
 - AMI by census tract