Brad Pawlow, Sebastian Reid, Noah Slosberg

04/21/2023

STAT 390

Progress Report 1

Goals from Week Plan 1:

* Begin work on an EDA
* Perform univariate, bivariate, and multivariate analysis
* Perform simple data cleaning procedures to better understand the data

Tasks Completed:

* Began work on an EDA
  + We visually mapped the programs on a map of Chicago
  + We found the breakdown of programs by category type
  + We looked ar average capacity for programs and discovered some issues with the capacity variable
* We added demographic data by community area
  + This includes data for race, income, and diversity indexes
* We cleaned the data a bit by performing feature selection and creating new columns that condense the community area data down to geographic clusters like west side, north side, etc. We also saw that the category name variable would have names like “Sports + Wellness.” and “Sports + Wellness”. We removed any periods from the category names to make it unified.

Roadblocks/Concerns:

* Deciding which program categories to further our EDA analysis
* Identifying which problem to address within different neighborhoods
* What target variables would be most valuable to predict and apply ML models?

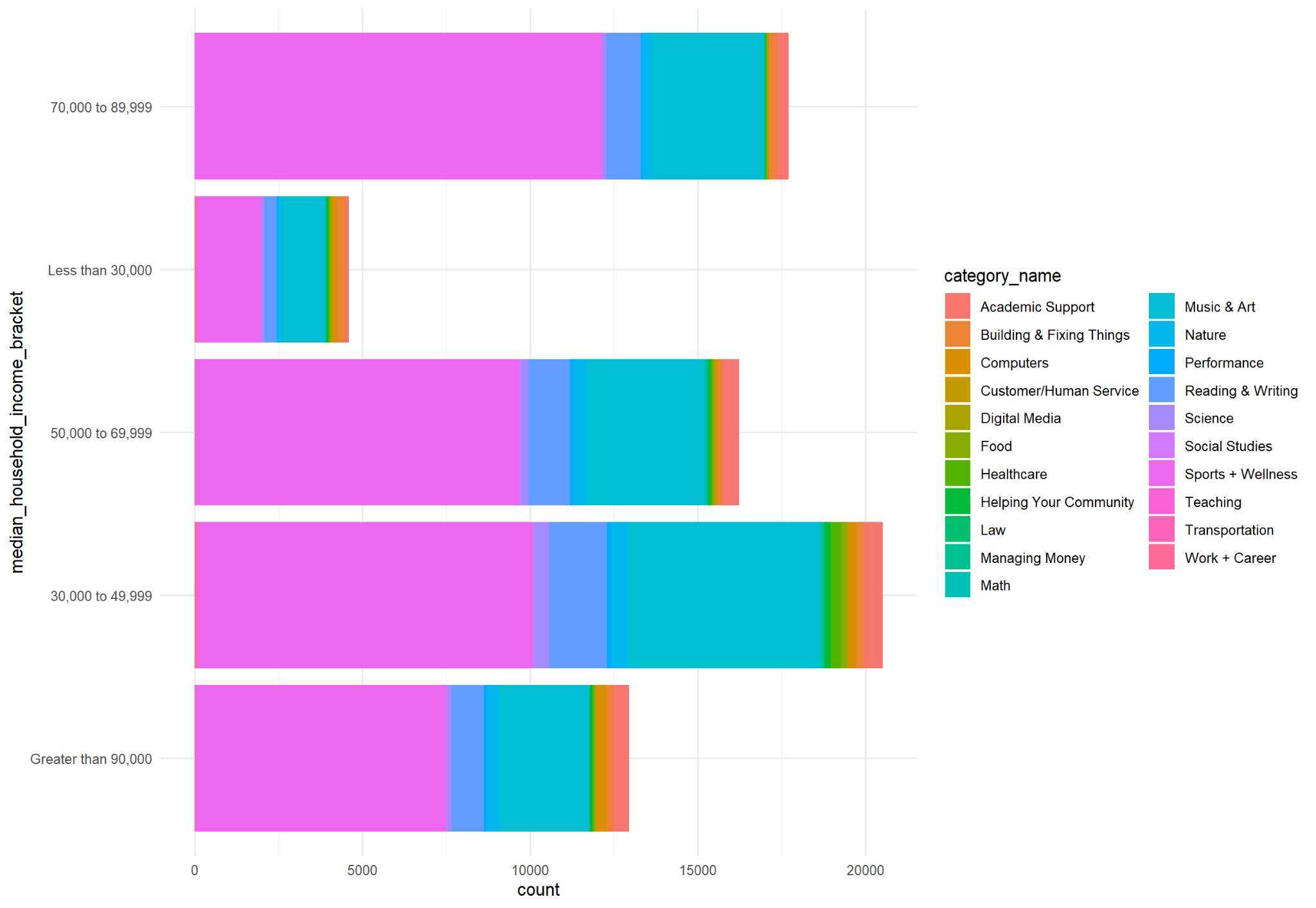
Goals for Week 2:

* Finish data cleaning to filter out the predictor variables that are significant
  + Use Lasso and Ridge regression to find which features best predict and affect the data.
  + We will also work on coming up with what we actually want to predict
* Flatten dataset to only contain unique observations
* Discard programs where the minimum age is over 25
* Reorganize dataset with more geographic data
* Maybe Perform PCA and factor analysis to condense dimensionality of dataset

*Appendix*

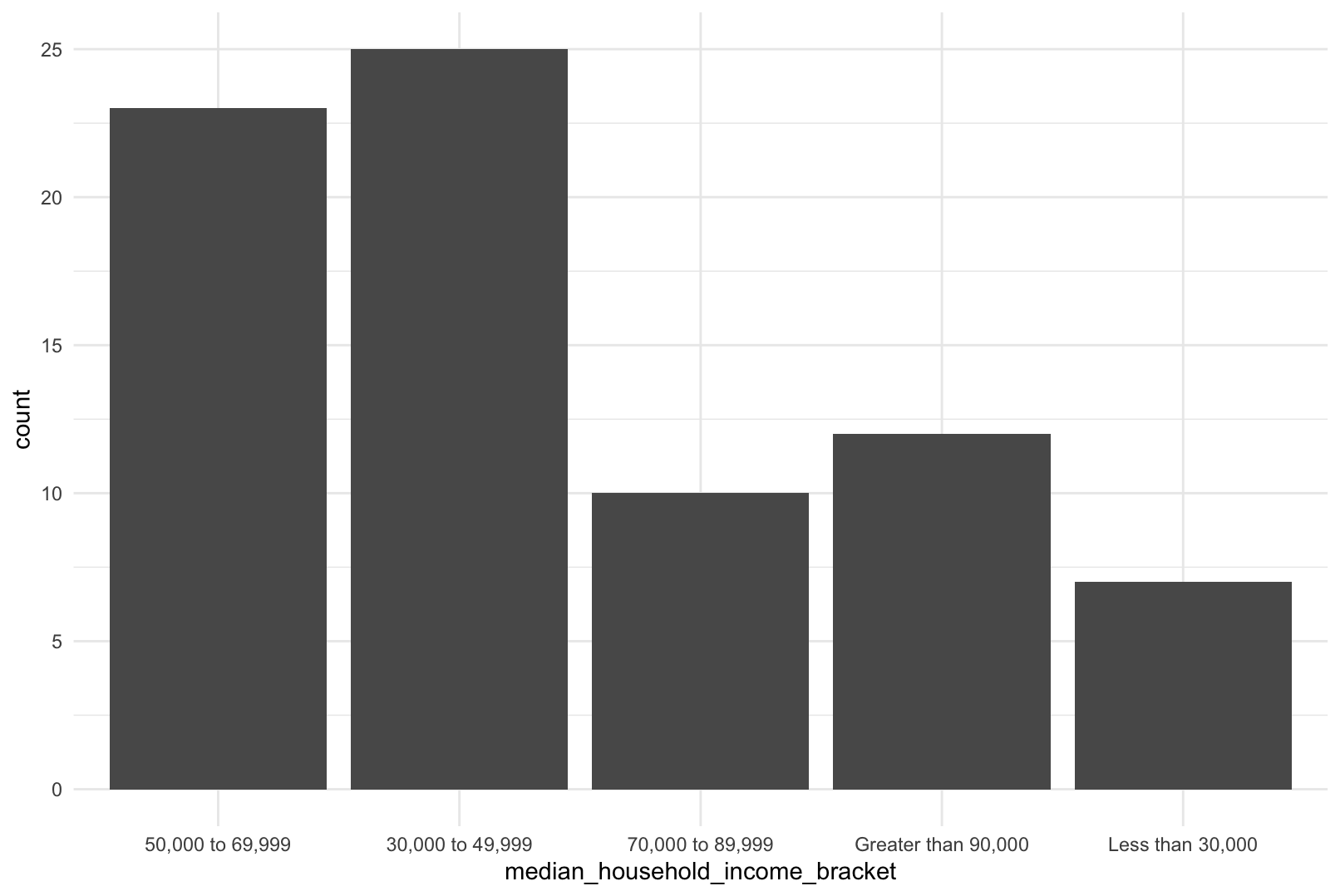
*Plot 1*

*Number of programs by income and category*

**

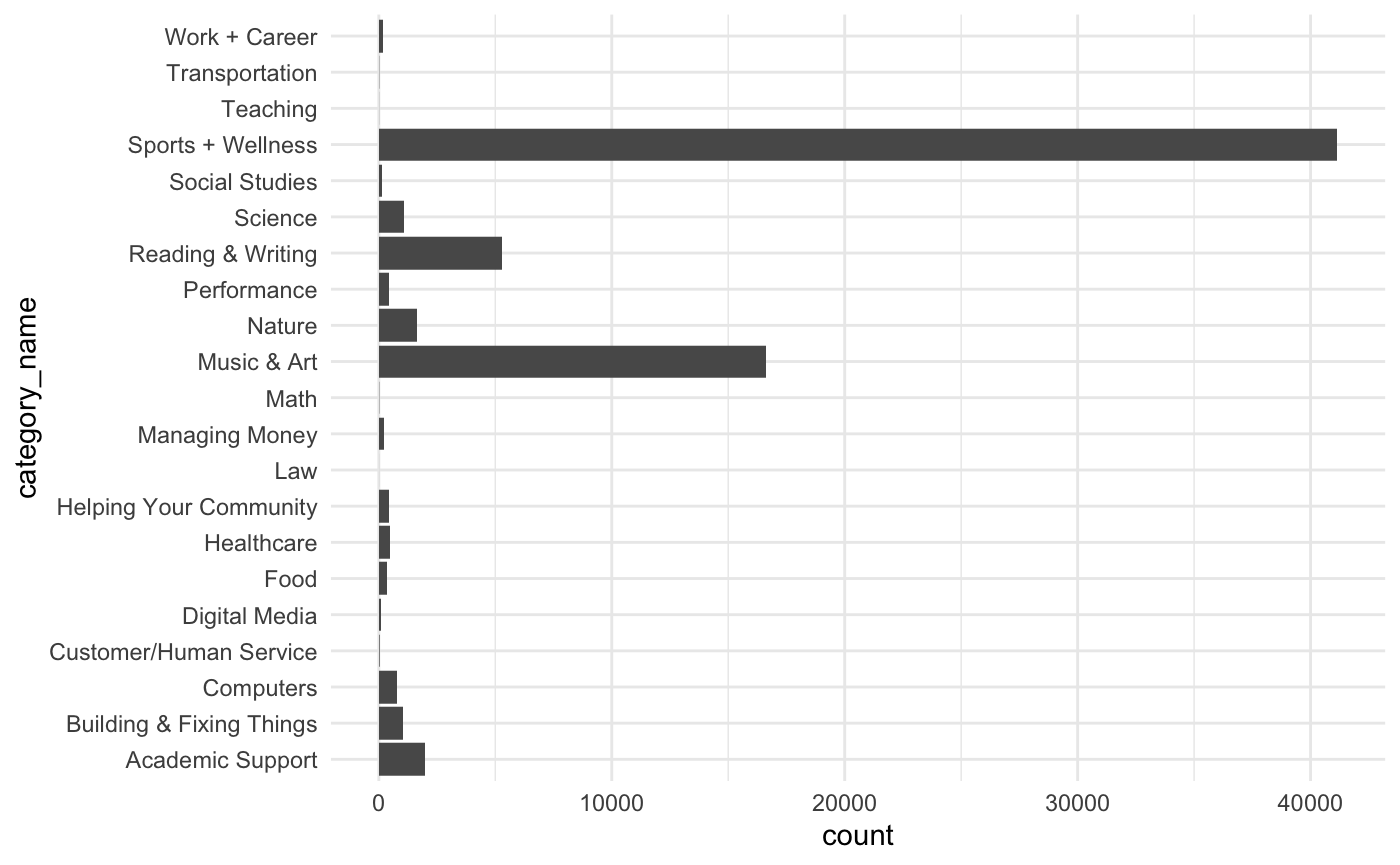
*Plot 2*

*Income Breakdown of Chicago by Community Area*

**

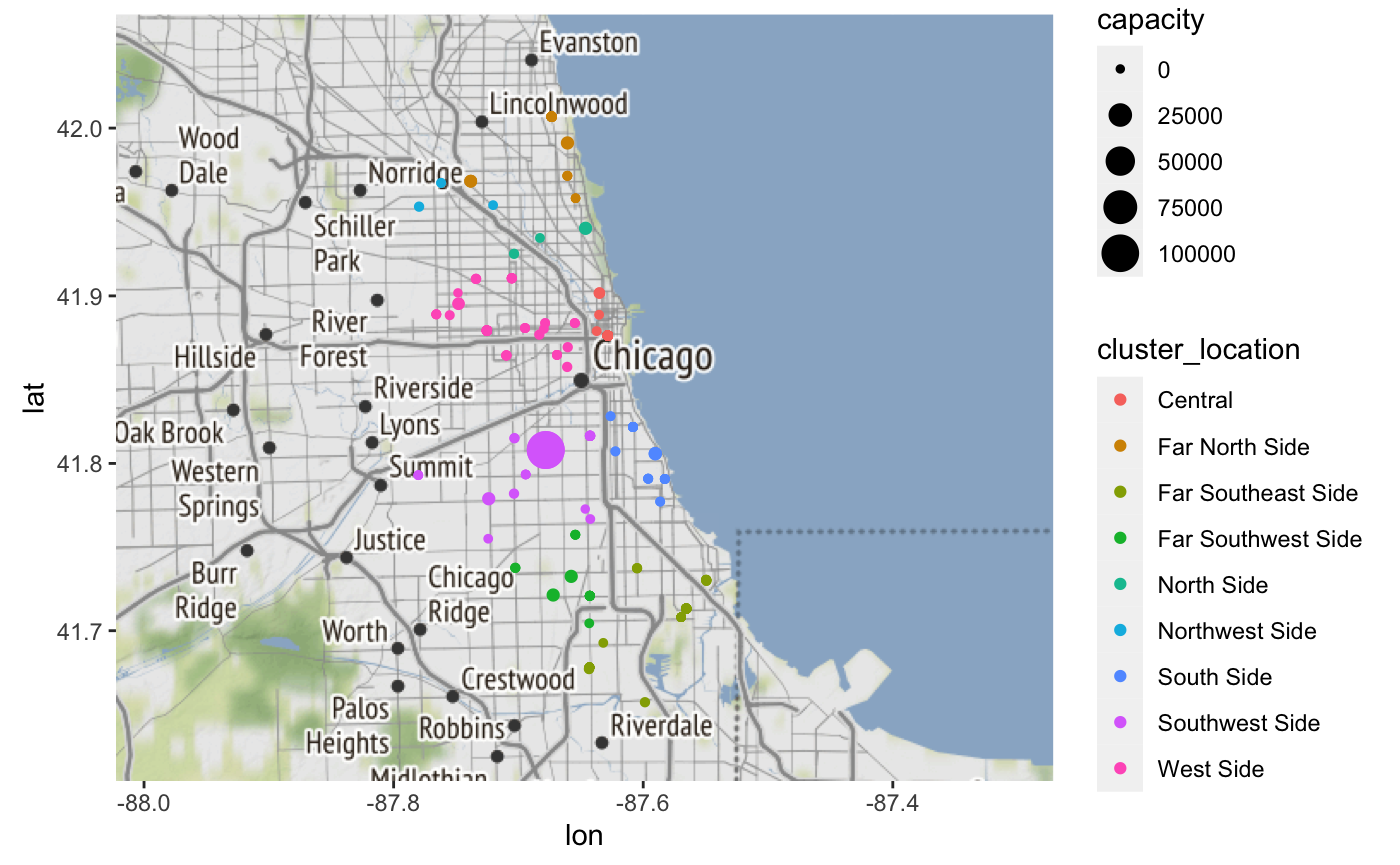
*Plot 3*

*Number of programs by category*

**

*Plot 4*

*Computer Programs Offered by Capacity and region*

**