Noah Slosberg, Sebastian Reid, Bradley Pawlow

4/14/23

Stat 390

Weekly Report 1

For the next week, we will start working on our EDA. To start, we need to begin to understand the dataset provided by the partners. First and foremost, we need to understand the completeness of the data and any issues that may arise from problems in the dataset. We will need to clean the data and fix the issues with it before we can even start to analyze the data.

We will start by performing the following analyses:

Univariate:

* Number of Geographic Areas
* Capacity of programs

Bivariate:

* Number of Programs by Geographic Areas
* Number of Programs by predominant Race of community
* Number of Programs by income data
* Number of programs by category

Multivariate:

* Number of programs by Geographic area and program category

We will then attempt to find outside data sources that will be useful to our analysis. We will likely have to bring in the following sources:

* Geographic boundaries for the community areas
* Transportation data between community areas
* Census data for income and other demographic factors