Data2020 Midterm

Emily Jaekle, Cynthia Vu, Tony Wu

Objective

The data set we are using is New York City census data. In this project we aim to create good models for median household income and income per capita in NYC and the surrounding area.

Cleaning Data

The data set given to us had multiple NA values, which we had to address. The data set had 36 columns and 2167 rows. There were a total of 1269 NA values. Many of these values were in rows where the total population was zero. When we remove these rows there were only 177 NA values and 2128 rows. We then chose to use na.omit to get rid of the rest of these 177 values. This left us with a data frame with 2095 rows. Thus we only lost 33 rows (since the other 39 had total populations of 0). This would have been the same as if we had done na.omit on the dataset at the very beginning, but now we know we did not lose as much data as we intially thought. We proceeded with this data set.

Exploratory Analysis

```
## [1] "example!"
```

Appendix (R code)

Cleaning Data R Code

Exploratory Analysis R Code

```
print("example!")
```