Are Fatal collisions Correlated with Socioeconomic Status?

In this project, we investigated whether socioeconomic status (defined by median income in the investigation) effects the number of fatal road collisions. In order to investigate this question, data from 5 major North American cities was analyzed; Toronto, Montreal, New York, Vancouver and Boston. Our hypothesis was that lower income neighbourhoods would have more fatal collisions than neighbourhoods with higher median incomes. To test this hypothesis, neighbourhood collisions and median income would need to show a strong, negative correlation. Data was obtained from respective municipal police departments (collisions) and citywide census (income), whereby data was grouped by neighbourhood. A final summary table was created for each city displaying median income, collision count, population (for scaling) and location. Using matplotlib, scatter plots with regression analysis was conducted for data visualization purposes. Collision fatalities were then mapped using the google maps API, the heatmap clearly displays hotspots for fatal collisions.

The null hypothesis is rejected (r=-.207 in combined city graph), there is no correlation between fatal collisions and median income. If time permitted, conducting further hypothesis testing would be a great next step to see if our results are significant. The data sets used were quite limited, conducting further analysis with larger datasets or comparing other road collision types to neighbourhood median income would also be another interesting next step. Many studies show that factors including walkability, traffic flow and general maintenance mean that lower income neighbourhoods do have higher traffic collisions (especially at risk: pedestrians).

Our data did provide us an answer to a business question, which is ‘where do collisions most frequently?’. These hotspots (seen on the heatmaps) have created multiple additional questions including; are pedestrians at greater risk in lower income neighbourhoods, are there certain areas with more cyclist collisions/speeding etc?

A further question that was as of the data following the first analysis was whether fatal accidents involving a pedestrian is correlated to the median neighbourhood income in Toronto. A new dataframe was made, limited to collisions if they involved a pedestrian. The same data visualization method was then done (scatter plow of collision count vs median income) to study if there was a correlation between pedestrian risk and the neighbourhood SES. Once again the null hypothesis was rejected, as the correlation (r=-0.15) between collisions involving pedestrians and neighbourhood median income was not significant.