Propose a project that uses a large, publically accessible dataset. Motivate the problem you are tackling, discuss the data source(s) you are using, and explain the analysis you are performing. Do enough exploratory data analysis to convince the project is viable and to generate two interesting non-trivial plots. Explain them and give url links to those plots.

The Urban Forest as a Factor in Social Wellbeing

This project is a look at the association between urban trees and several indicators of social wellbeing. Research has shown a positive relationship between street trees and income levels in many American cities, where wealthier neighborhoods have more trees. Another recent study found that mothers who live in neighborhoods with more trees and green space have babies with higher birth weights, even after adjusting for income, air pollution, noise levels, and walkability. These studies suggest that the design of urban spaces can impact the wellbeing of the people who occupy them, and may have implications for public health and health care costs.

For example, we know that birth weight is associated with health outcomes as the person ages. Green space may therefore contribute to potential long-term benefits for the individual, and for society in terms of health care costs. Second, it’s possible green space is related to health outcomes in other populations, such as those with chronic illness, cancer, or cardiovascular disease. Third, green space may affect other indicators of social wellbeing, such as crime rates, education levels, and income - if not directly, then perhaps by influencing the perception of different neighborhoods and their residents. A example of this type of effect is the “Broken Windows Theory” (not to be confused with the Broken Windows Fallacy), which proposes that signs of disorder and vandalism will increase the likelihood of further crimes in an area, due to it’s perception as a place where criminal activity is likely to go unnoticed. Adapting urban planning to accommodate green space has the potential to make a cost-effective contribution towards improve public health and social wellbeing.

This project is a preliminary step in determining the potential impacts of green space on health and social outcomes: it is an assessment of the relationship between income, unemployment levels, education, and street trees in Vancouver, Canada.

Three publically available data sets were obtained for this project, all from the City of Vancouver’s Open Data website:

1. Vancouver Street Trees, a data set mapping and describing every street tree in the city. <http://data.vancouver.ca/datacatalogue/streetTrees.htm>
2. Statistics Canada 2006 census data, compiled by Vancouver neighborhood. <http://data.vancouver.ca/datacatalogue/censusLocalAreaProfiles2006.htm> -Note: the most recent census was in 2011. As it did not include many social and economic indicators collected in previous years, I used the 2006 data.
3. A GIS dataset containing the geographical boundaries for Vancouver’s 22 neighborhoods. <http://data.vancouver.ca/datacatalogue/localAreaBoundary.htm>

Analysis: I performed an exploratory data analysis to understand the associations between the primary factors in the datasets across the 22 neighborhoods of Vancouver. First I derived the area (square km) of each neighborhood from the GIS dataset. Then I aggregated the trees by neighborhood, and created an index of ‘tree mass’ for each. The index is a factor of tree density (per square km) and average tree size (as a function of height and diameter). I selected a few indicators of social wellbeing from the census data, including median income, education levels, percent of population living under the poverty line, and the unemployment rate. I ran some simple correlations between tree mass and the social indicators, and then used the GIS data to map the tree index and median household income.

Outcome: The economic indicators also all showed a strong correlation with the tree mass index. This is not unexpected, as they all measure aspects of a single construct: economic success. Education was also associated with tree mass, but to a lesser degree. Not shown (but interesting) is that while median household income and the unemployment rate were both highly correlated with the % of households that were low income, they were less highly correlated with each other. It’s clear that the relationship between income and trees that has been observed in some American cities also holds true in Vancouver. What impact this has on the wellbeing of the residents of these neighborhoods, including via how they are perceived both within and from outside their neighborhood, remains an open question.

|  |  |
| --- | --- |
| Factor | Correlation with Tree Mass (r2) |
| Median Household Income | 0.53 |
| % of Households that are Low Income | -0.49 |
| % of adults aged 25 to 65 with some Post-Secondary Education | 0.20 |
| Unemployment Rate | -0.49 |

Graph 1: This graph shows the correlations between the tree mass index, and median household income and the unemployment rate: neighborhoods with higher incomes have more trees, and neighborhoods with high unemployment rates have fewer trees. <http://goo.gl/RX4JfM>

Graph 2: Map of Vancouver neighborhoods. The color gradient represents the tree mass, with darker green indicating more trees. Clicking on each neighborhood will provide the name of the neighborhood, the tree mass index and the median household income. <http://goo.gl/zjJyGr>

<https://www.google.com/fusiontables/embedviz?q=select+col3%3E%3E0+from+13O8yWdo9gFaMw86ThyxiJ2wybnMC4zXDDzZE0Zxa&viz=MAP&h=false&lat=49.25113290751412&lng=-123.124052&t=1&z=12&l=col3%3E%3E0&y=2&tmplt=2&hml=KML>

<https://github.com/sarawxyz/data-incubator/blob/master/corr_graphs.png>

<https://www.youtube.com/watch?v=XHiuBDvjyJc>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Correlation Matrix | | | | | |
|  | Median Household Income | % of Households that are Low Income | % of adults aged 25 to 65 with some Post-Secondary Education | Unemployment Rate | Tree Mass Index |
| Median Household Income | 1 | -0.82 | -0.41 | -0.52 | 0.53 |
| % of Households that are Low Income |  | 1 | 0.13 | 0.79 | -0.49 |
| % of adults aged 25 to 65 with some Post-Secondary Education |  |  | 1 | -0.26 | 0.20 |
| Unemployment Rate |  |  |  | 1 | -0.49 |
| Tree Mass Index |  |  |  |  | 1 |