High-dimensional analysis of census tracts within the City of Surrey
Natasha Mattson, Sarah Neubauer, Rashedul Hoque, and Tony Hui

CASCADIA URBAN ANALYTICS COOPERATIVE

Introduction

PROJECT: Economic Development project with the City of Surrey for the University of British Columbia’s (UBC) 2017 Data Science for Social Good (DSSG) fellowship program

PURPOSE: To create an economic profile of Surrey, describing different features which have the potential to affect the economic health of Surrey as a whole and to find out what regions within Surrey (on a census tract level) are distinctive with these different features.

DATASETS:
• Geographic data
• 2011 National Household Survey (NHS)
• Business licenses
• Commercial rental listings
• Job postings
• Property assessment data
• Business break and enters
• New building permit data

Hierachical clustering reveals distinct clusters

• Upon inspection of the resulting dendrogram, we decided to go with 5 clusters (as colored).

Automatic removal of redundant variables

• Many variables containing redundant information (left)
• Simultaneously reduce dimensionality and remove redundancy with Principle Component Analysis (PCA)
• Decided to use 4 PCs as PC5’s loadings appeared to contain redundant information (right)

Towards a new data-driven vocabulary

Right: Cluster results visualized in VisualSurrey. Each color represents one cluster. Clusters have a tendency to be grouped together in geographical space. This enables a new vocabulary to describe the different neighborhoods in the city.

Right: High-level descriptions of the clusters

Below: Example boxplots of some key variables defining each cluster of census tracts

Conclusions and Future Directions

• Built a visualization tool to visualize arbitrary =
• Clustered census tracts in high dimensional space
• Interpreted the defining characteristics for each cluster
• More data to add to VisualSurrey
• More detailed descriptions of census tract clusters

VisualSurrey allows for visualization of important metrics

• A data visualization platform written in Javascript/Python
• Hosted on Microsoft Azure
• Accessible at bit.ly/visualsurrey

VisualSurrey allows for visualization of important metrics

• A data visualization platform written in Javascript/Python
• Hosted on Microsoft Azure
• Accessible at bit.ly/visualsurrey

VisualSurrey allows for visualization of important metrics

• A data visualization platform written in Javascript/Python
• Hosted on Microsoft Azure
• Accessible at bit.ly/visualsurrey