Exploring Spatial Patterns and Relationships in GIS Datasets for Vision Zero Boston

Youshe Li

Department of Innovation and Technology

City of Boston

VISION ZERO BOSTON

- Vision Zero is an international road traffic safety initiative that aims to achieve a road traffic system with no fatalities or serious injuries. It started in Sweden and was approved by their parliament in October 1997.

- The City of Boston joined New York, Washington DC and other cities worldwide in adopting Vision Zero and issuing similar plans to prioritize transportation safety in December 2015.

- Vision Zero Boston is our commitment to focus the city's resources on proven strategies to eliminate fatal and serious traffic crashes.

- See http://www.visionzeroboston.org for more details.

VISION ZERO BOSTON – SPATIAL PATTERNS AND RELATIONSHIPS

- Objectives:

  To explore the question of what data should dictate the engineering focus of Boston Transportation Department and to understand how well 311 Engineering Requests align with other data such as Traffic Crashes, Safety Concerns and Waze Traffic Jams.

- Methodologies:

  Two spatial statistical tools (ArcGIS and R):

  1, Kernel Density Estimation (KDE) used to create the cluster maps.

  2, Ripley's K/L Function (both Univariate and Bivariate) used to explore the spatial relationships between different GIS datasets.

- Website: http://app01.cityofboston.gov/proxy/vzanalysis

VISION ZERO DATASET

- Traffic Crashes
  Totaled 9,229 cases from 7/1/2014 - 7/31/2016.

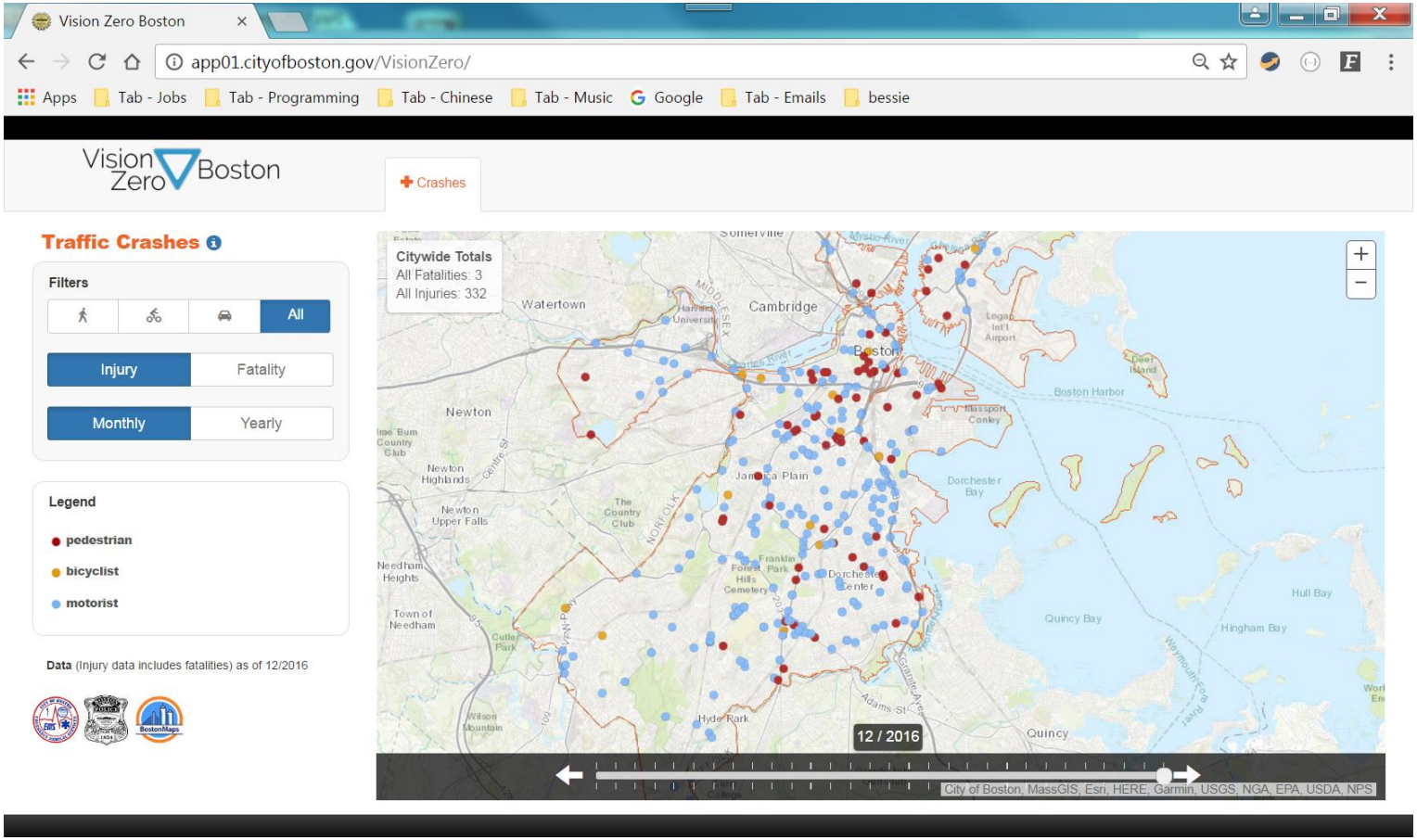
- Safety Concerns
  Totaled 8,188 cases from 1/15/2016 - 10/12/2016.

- 311 Engineer Requests
  Totaled 4,677 cases from 1/1/2014 - 6/26/2016.

- Waze
  Totaled 42,538 cases on one day.

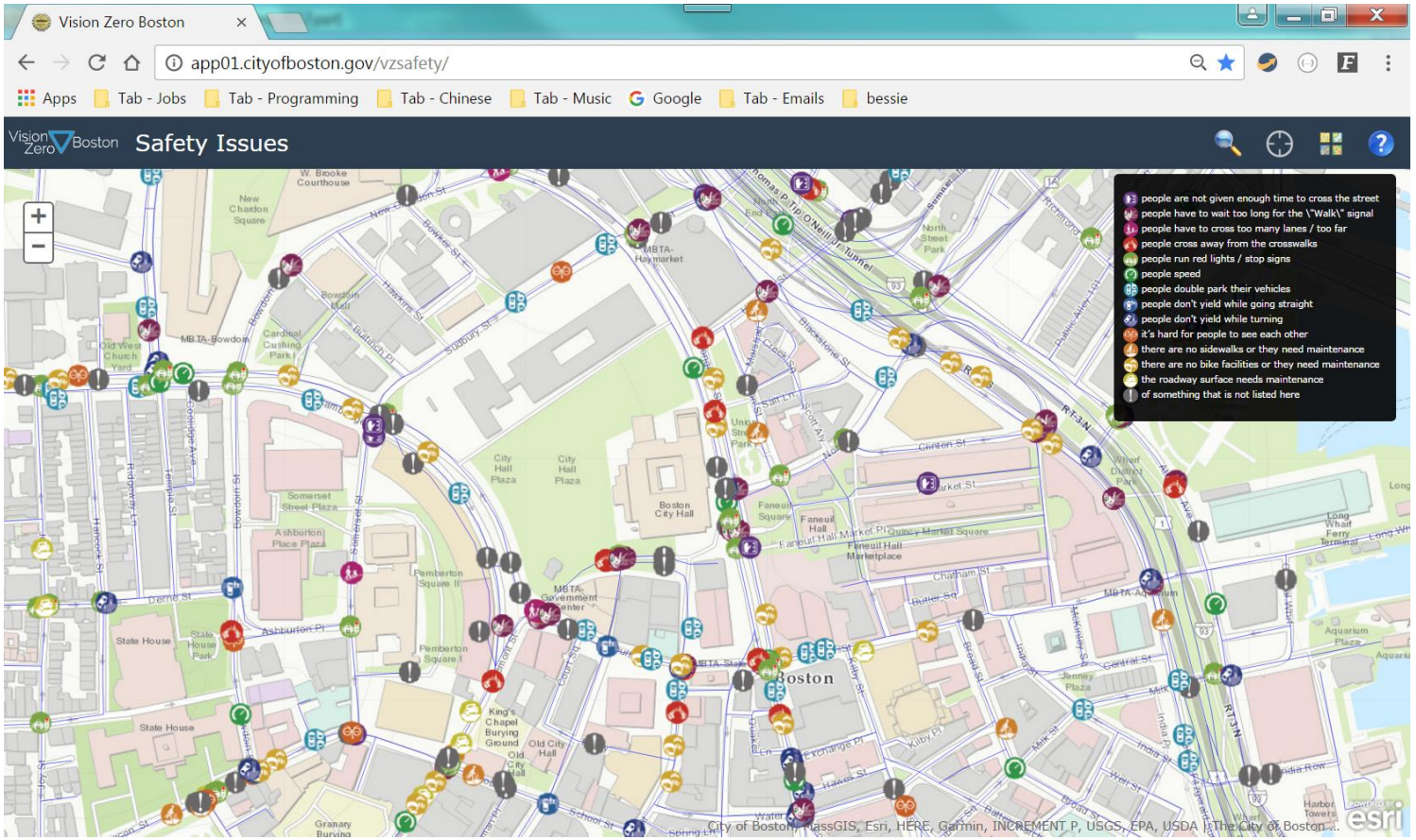
VISION ZERO BOSTON CRASH MAP

- http://app01.cityofboston.gov/VisionZero



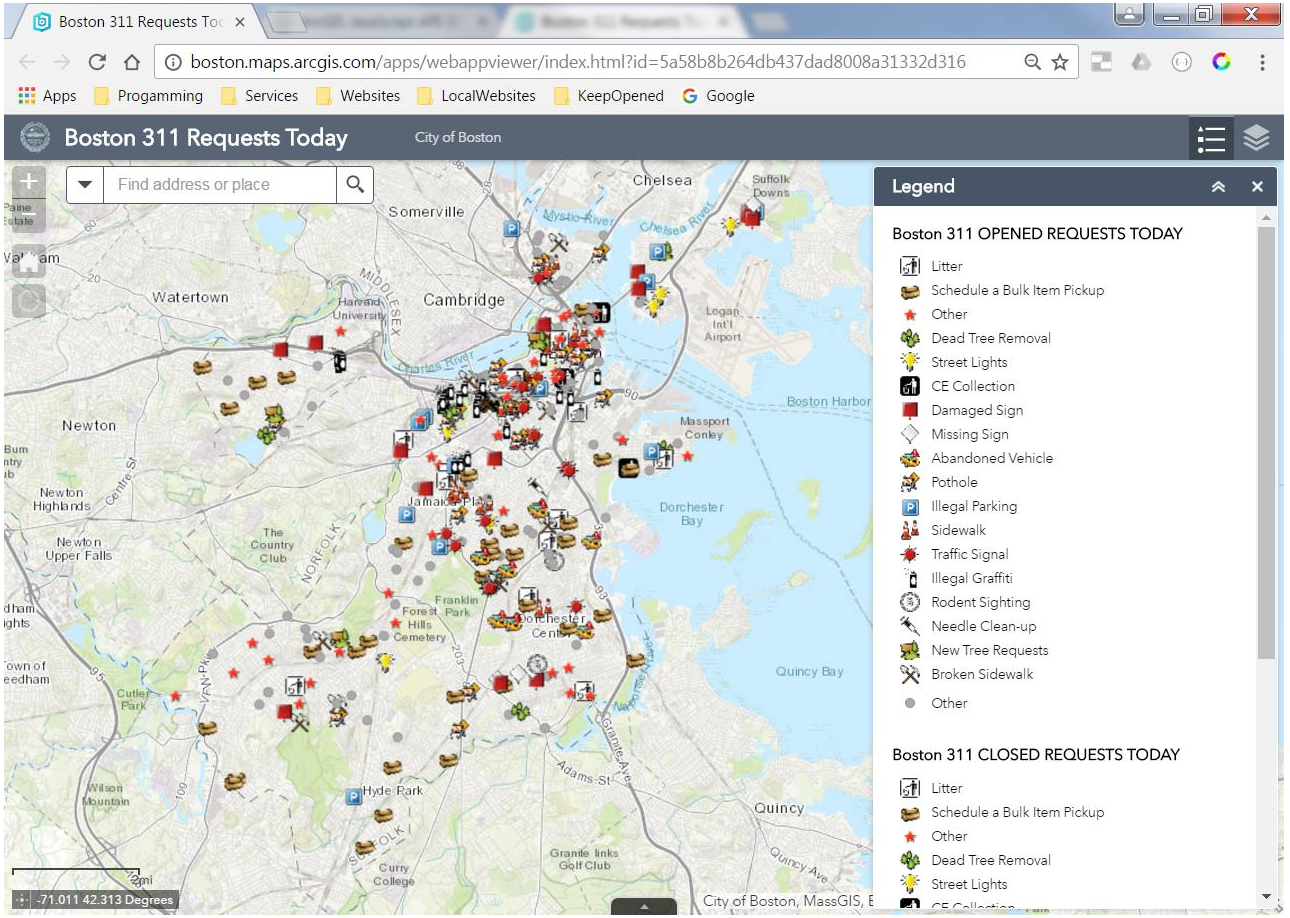
VISION ZERO BOSTON
SAFETY CONCERNS MAP

- http://app01.cityofboston.gov/vzsafety



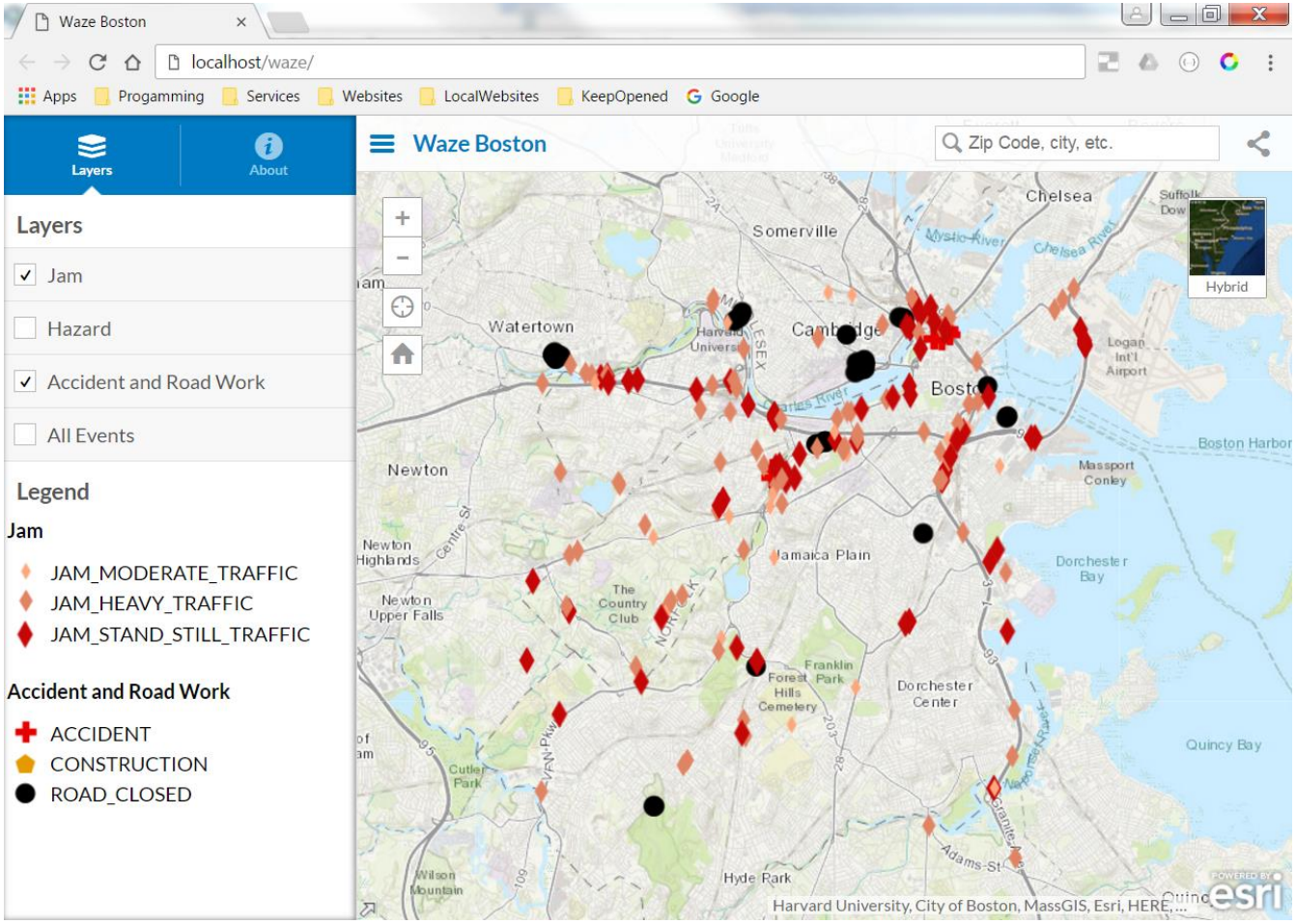
BOSTON 311 REQUESTS

- http://boston.maps.arcgis.com/apps/webappviewer/index.html?id=5a58b8b264db437dad8008a31332d316



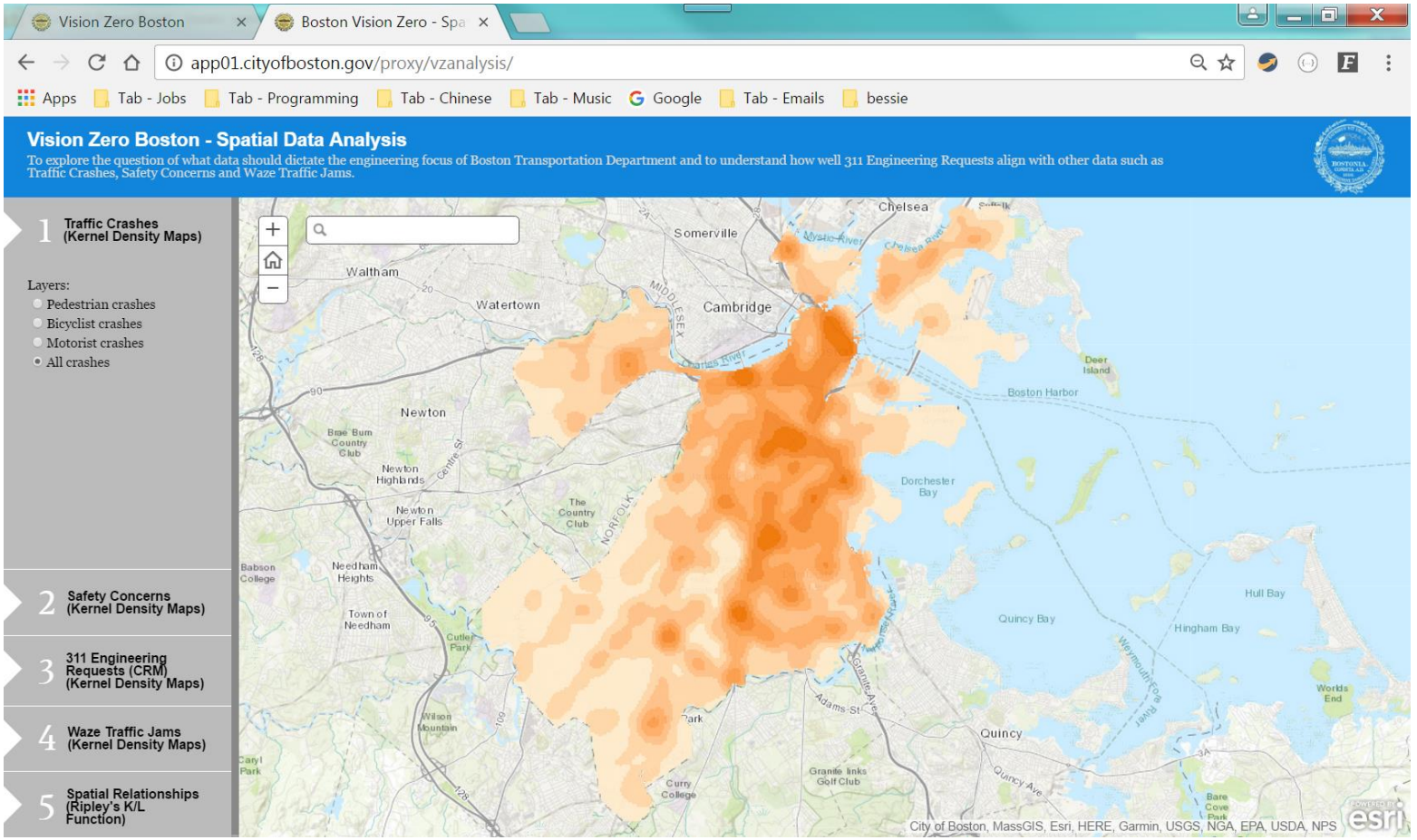
WAZE BOSTON

- http://localhost/waze (Internal website)



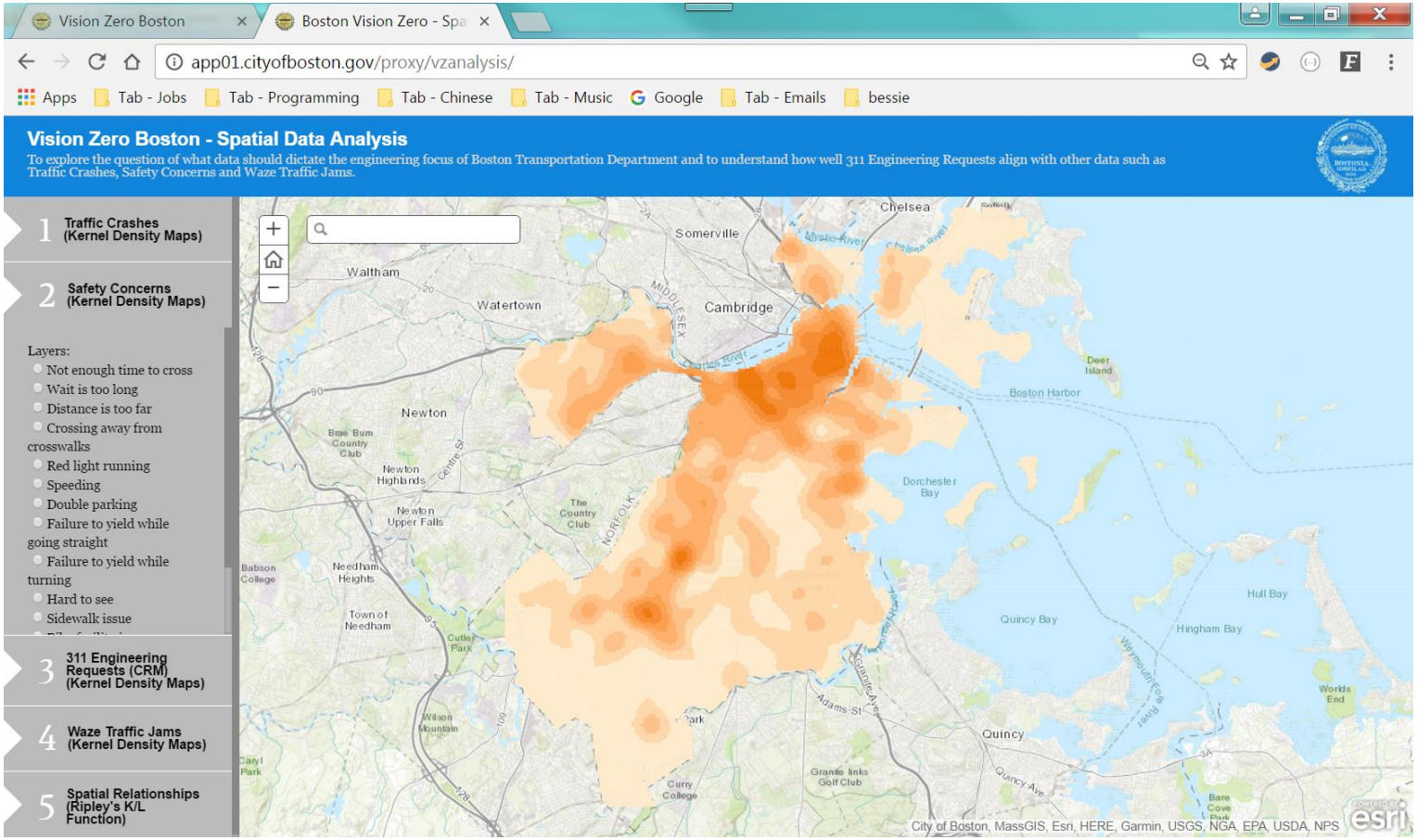
TRAFFIC CRASHES (KERNEL DENSITY MAP)

- http://app01.cityofboston.gov/proxy/vzanalysis/



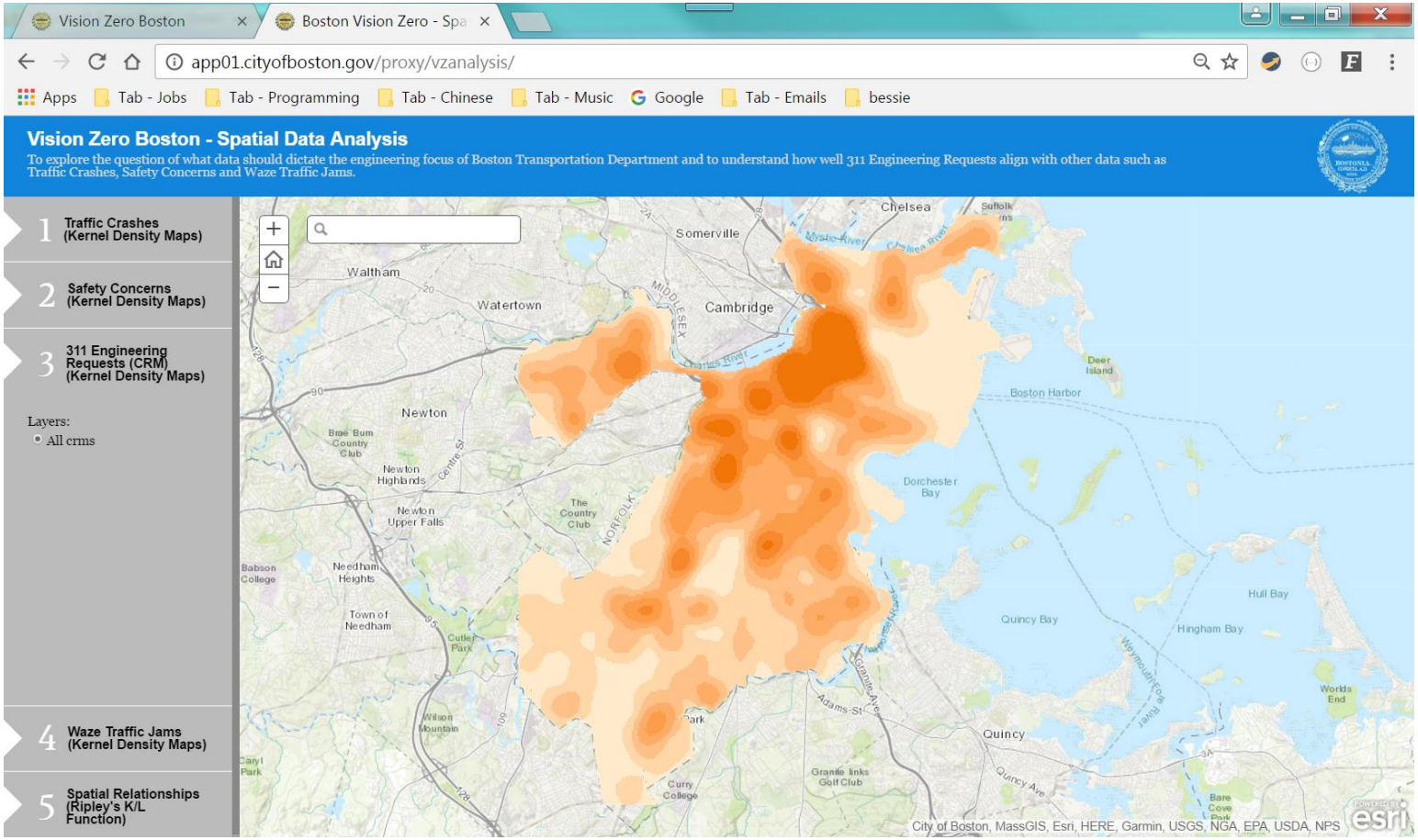
SAFETY CONCERNS (KERNEL DENSITY MAP)

- http://app01.cityofboston.gov/proxy/vzanalysis/



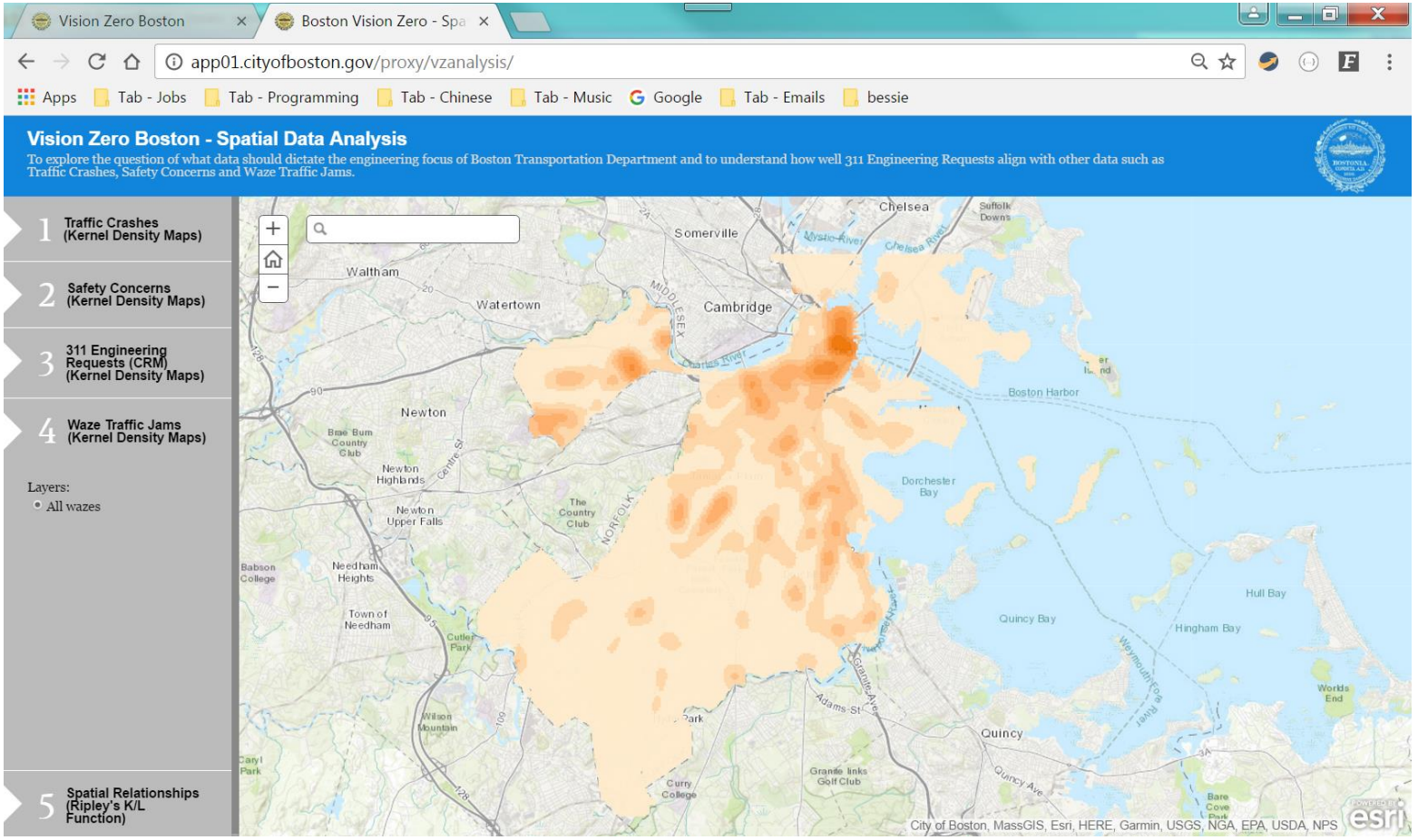
311 ENGINEERING REQUESTS (KERNEL DENSITY MAP)

- http://app01.cityofboston.gov/proxy/vzanalysis/



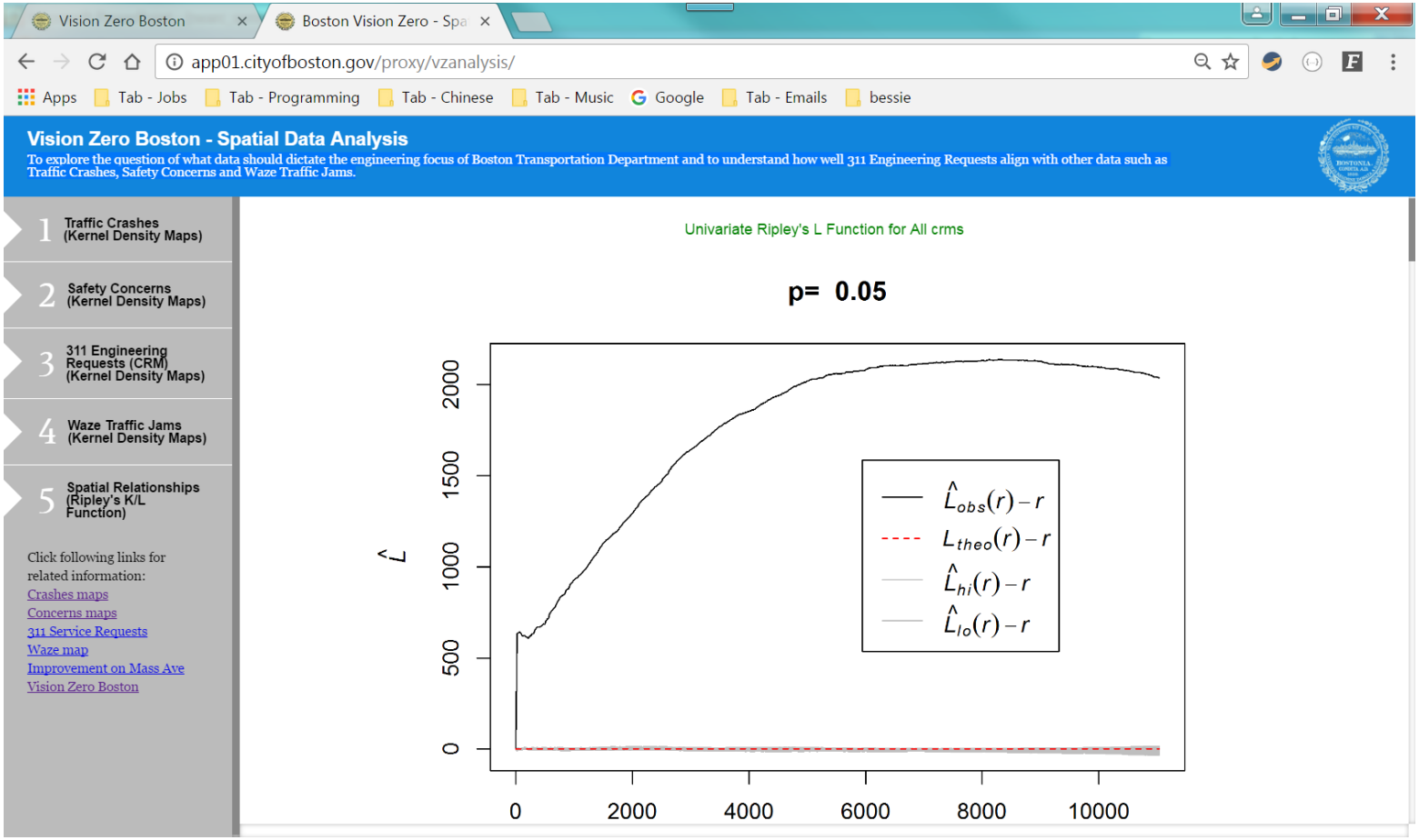
WAZE
(KERNEL DENSITY MAP)

- http://app01.cityofboston.gov/proxy/vzanalysis/



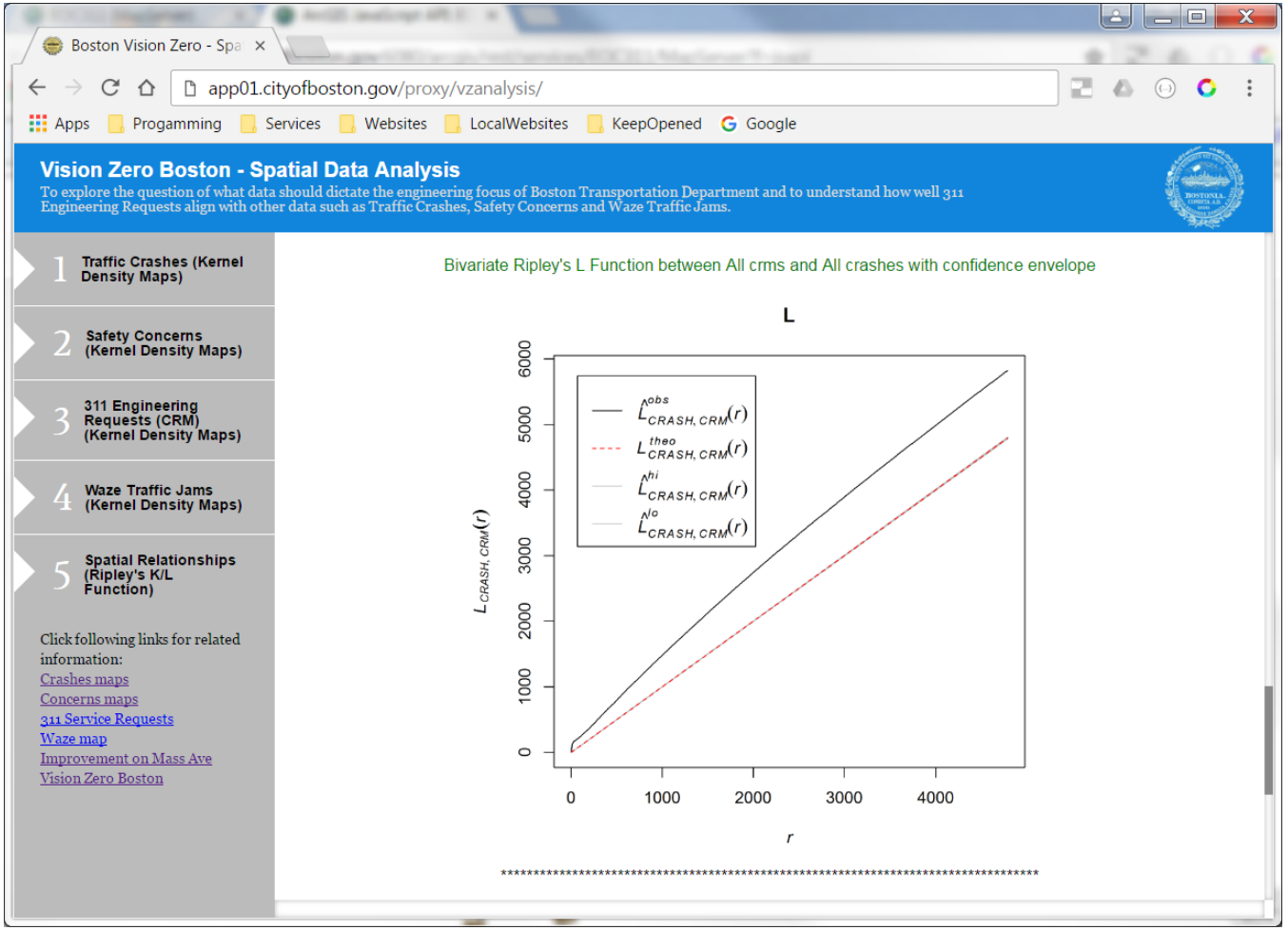
UNIVARIATE RIPLEY'S K/L FUNCTION

- http://app01.cityofboston.gov/proxy/vzanalysis/



BIVARIATE RIPLEY'S K/L FUNCTION

- http://app01.cityofboston.gov/proxy/vzanalysis



KERNEL DENSITY ESTIMATION (KDE)

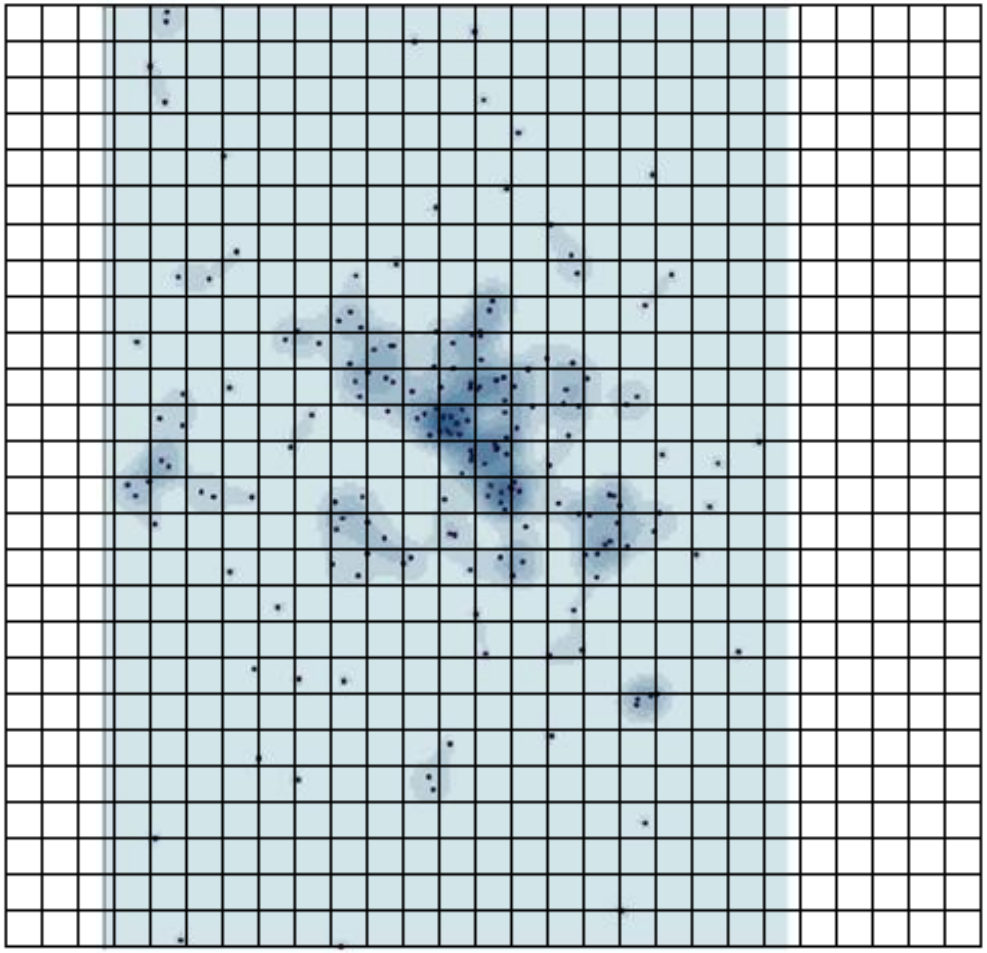
- In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable.

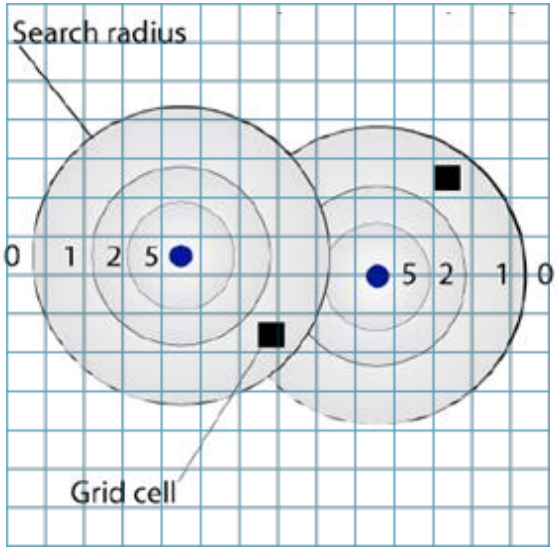
- Let (x1, x2, …, xn) be an independent and identically distributed sample drawn from some distribution with an unknown density f. We are interested in estimating the shape of this function f. Its kernel density estimator is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big),$$

  where K(•) is the kernel — a non-negative function that integrates to one and has mean zero — and h > 0 is a smoothing parameter called the bandwidth.

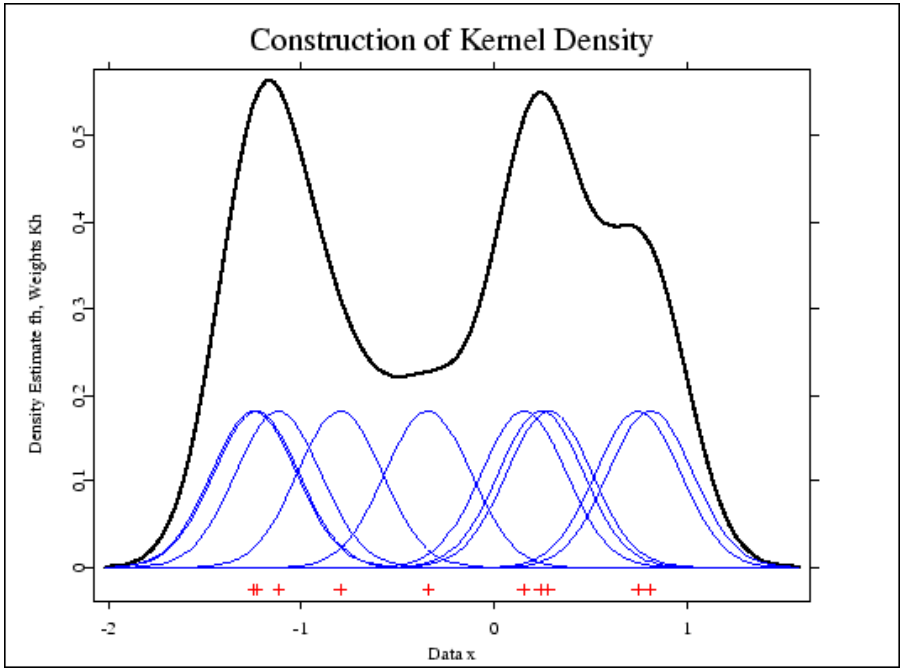
KERNEL DENSITY ESTIMATION (KDE)

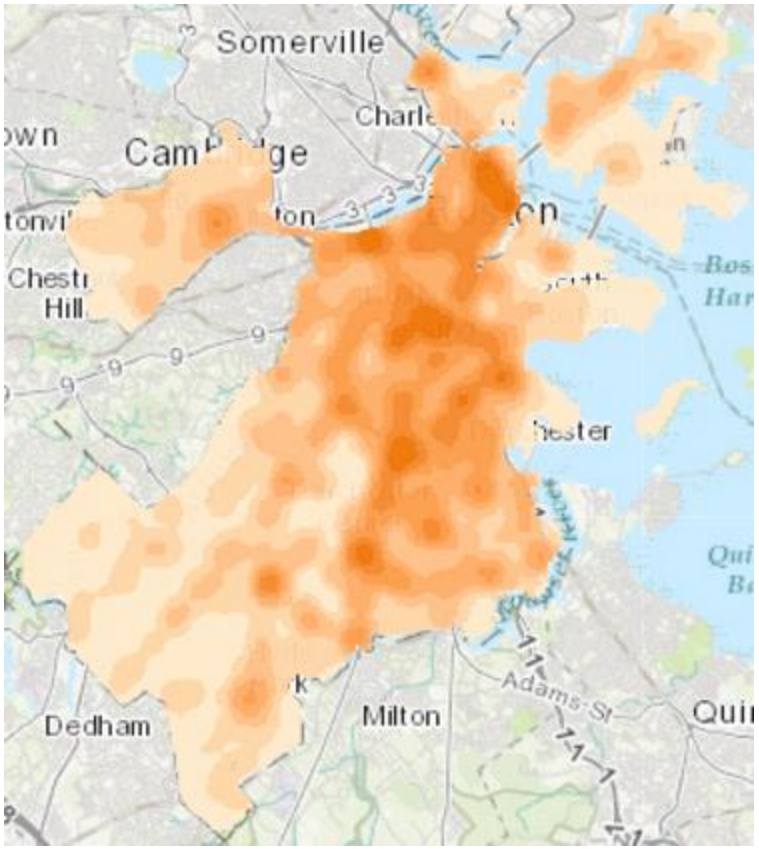




KERNEL DENSITY ESTIMATION (KDE)

- Kernel Function: Epanechnikov, Gaussian and Tri−cube etc.
- Bandwidth: crucial





RIPLEY'S K/L FUNCTION

- Ripley's *K* and *L* functions are closely related descriptive statistics for detecting deviations from spatial homogeneity. The *K* function (technically its sample-based estimate) is defined as

$$\hat{K}(t) = \lambda^{-1} \sum_{i \neq j} I(d_{ij} < t)/n,$$

- where d_{ij} is the Euclidean distance between the i^{th} and j^{th} points in a data set of *n* points, t is the search radius, λ is the average density of points (generally estimated as *n*/*A*, where *A* is the area of the region containing all points) and *I* is the indicator function (1 if its operand is true, 0 otherwise). If the points are approximately homogeneous, hat K(t) should be approximately equal to πt^2.

RIPLEY'S K/L FUNCTION

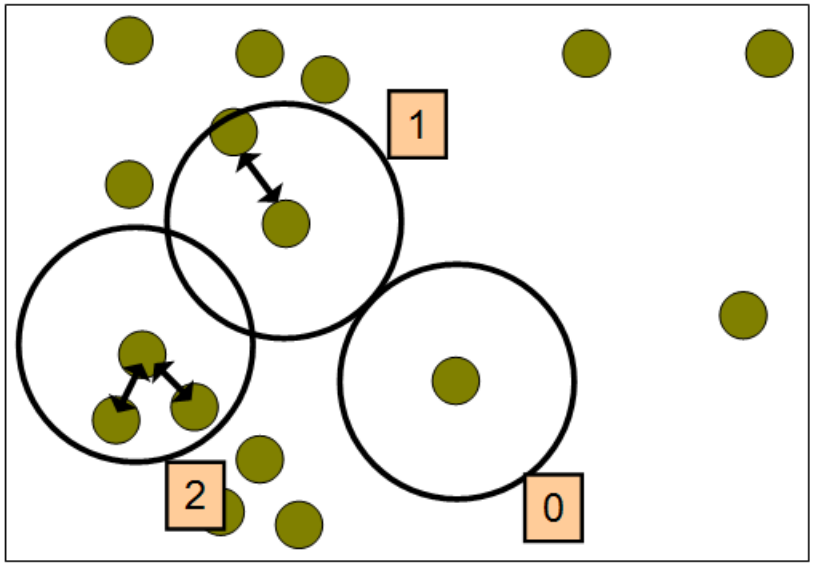
- For data analysis, the variance stabilized Ripley *K* function called the *L* function is generally used. The sample version of the *L* function is defined as

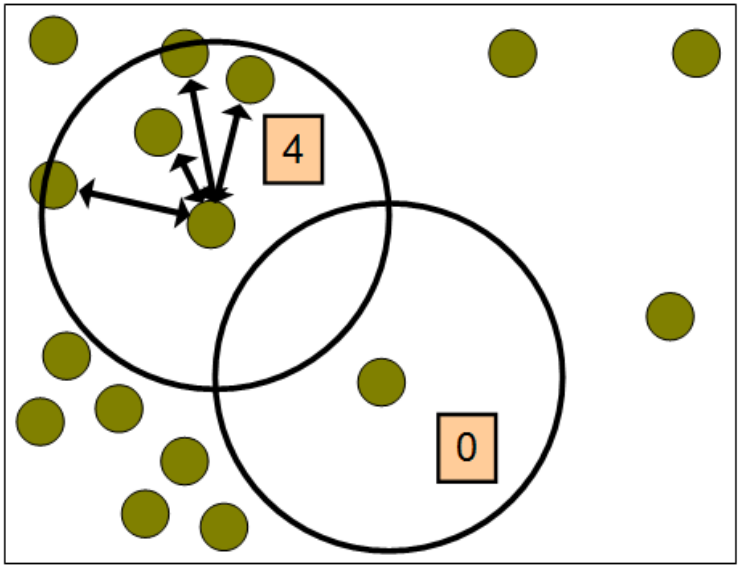
$$\hat{L}(t) = \left(\hat{K}(t)/\pi\right)^{1/2}.$$

- For approximately homogeneous data, the *L* function has expected value *t* and its variance is approximately constant in *t*. A common plot is a graph of hat L(t) against *t*, which will approximately follow the horizontal zero-axis with constant dispersion if the data follow a homogeneous Poisson process.

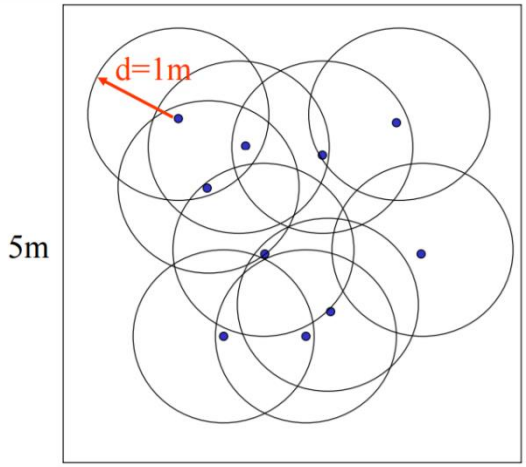
- Using Ripley's K function you can determine whether points have a random, dispersed or clustered distribution pattern at a certain scale.

RIPLEY'S K/L FUNCTION



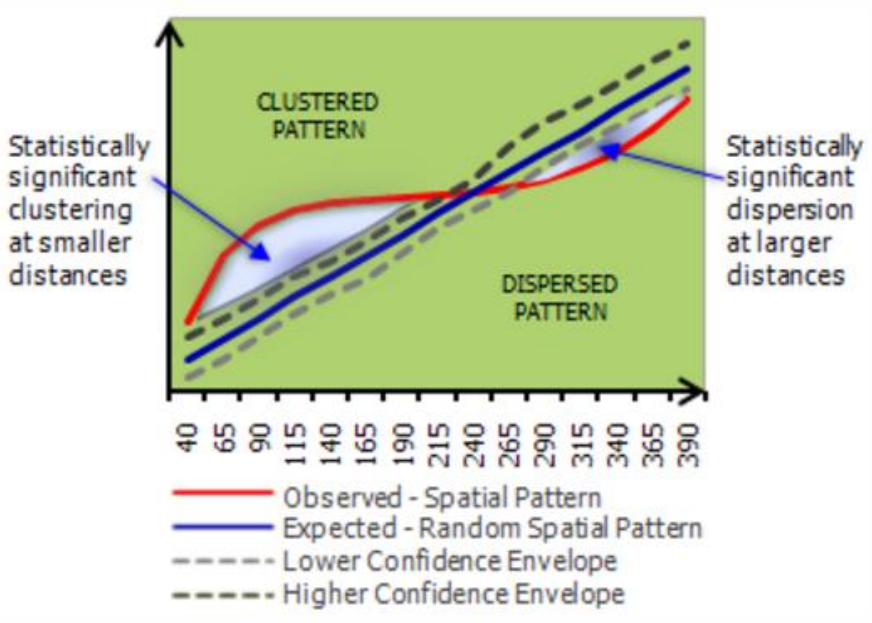


RIPLEY'S K/L FUNCTION

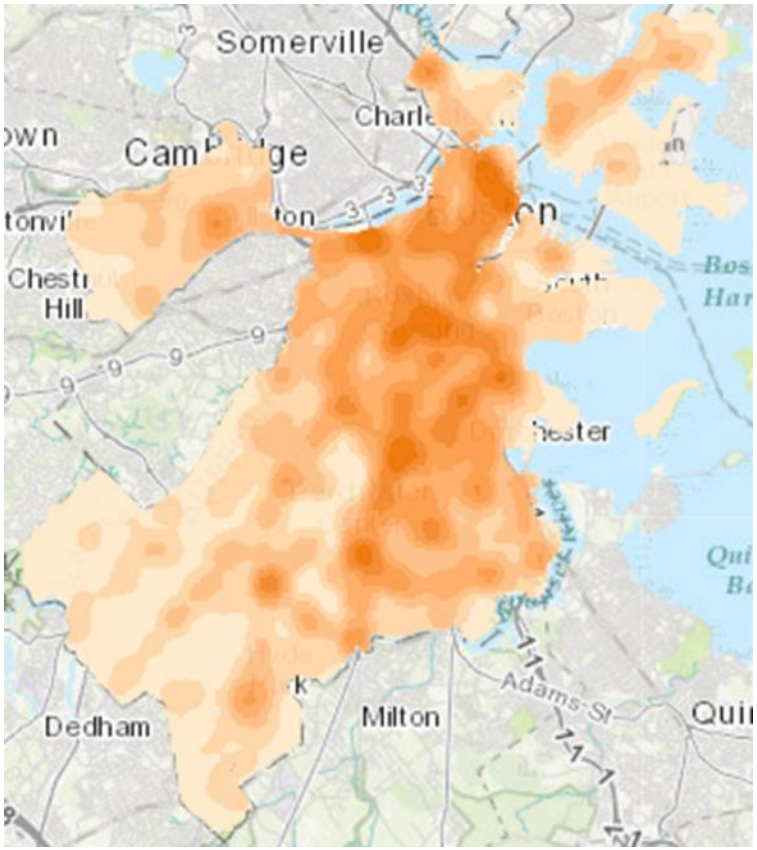


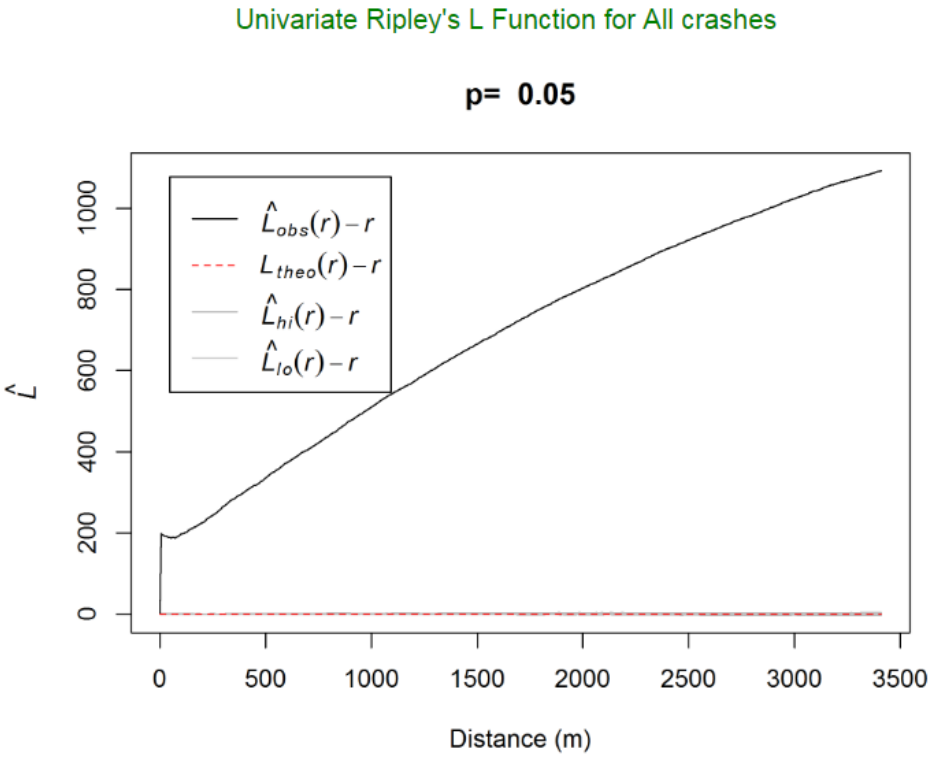
If uniform $K(1) = \pi = 3.14$

N(1)
2
3
2
1
2
1
0
2
1
2
E[N(1)] = 16/10
λ=10/25=0.4
K(1) = 1.6/0.4 = 4

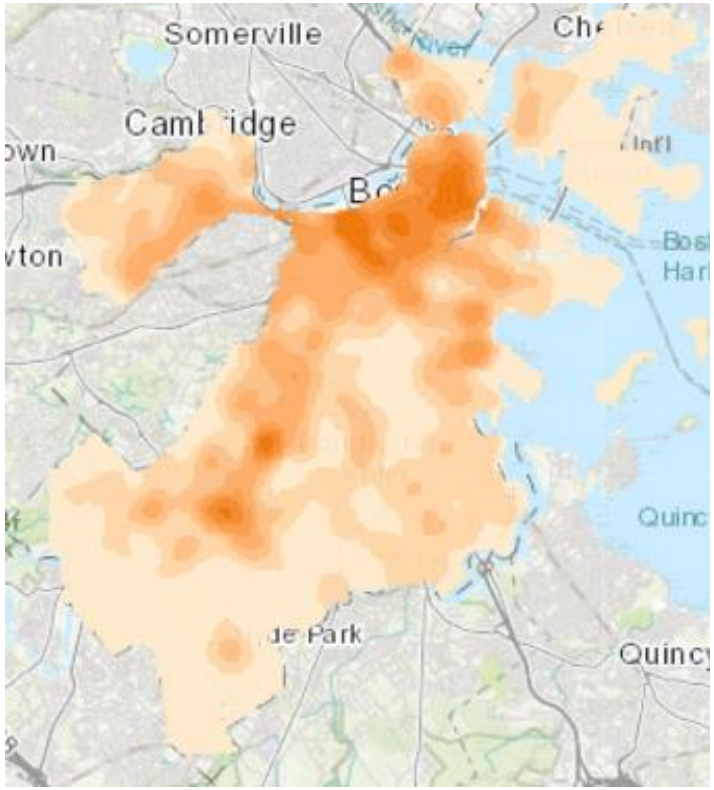


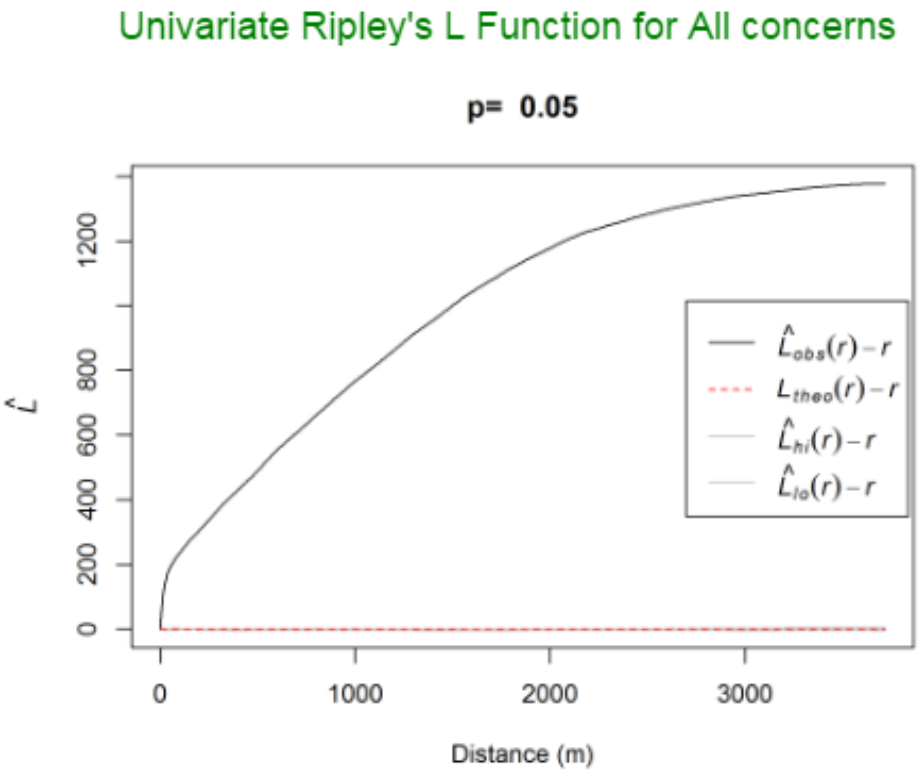
SPATIAL PATTERN OF TRAFFIC CRASHES



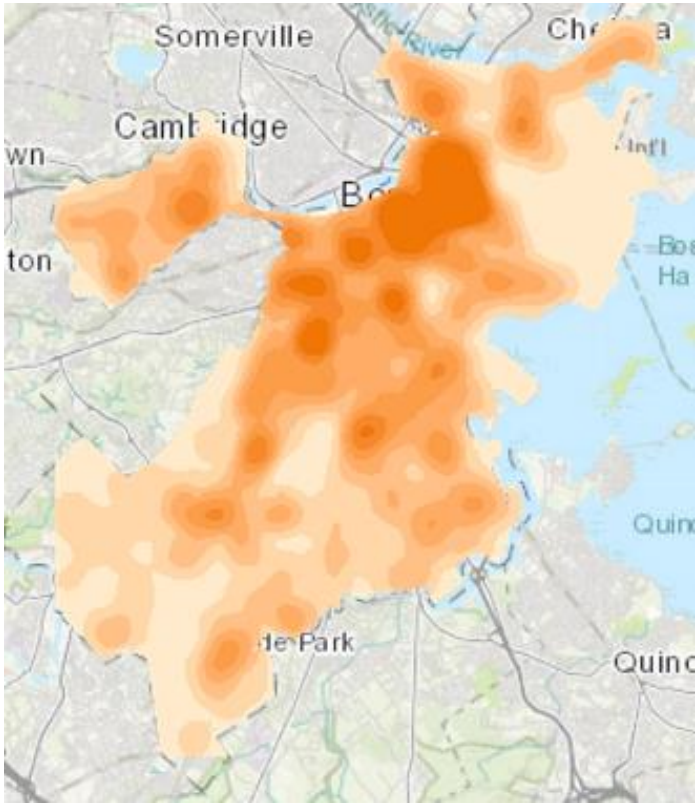


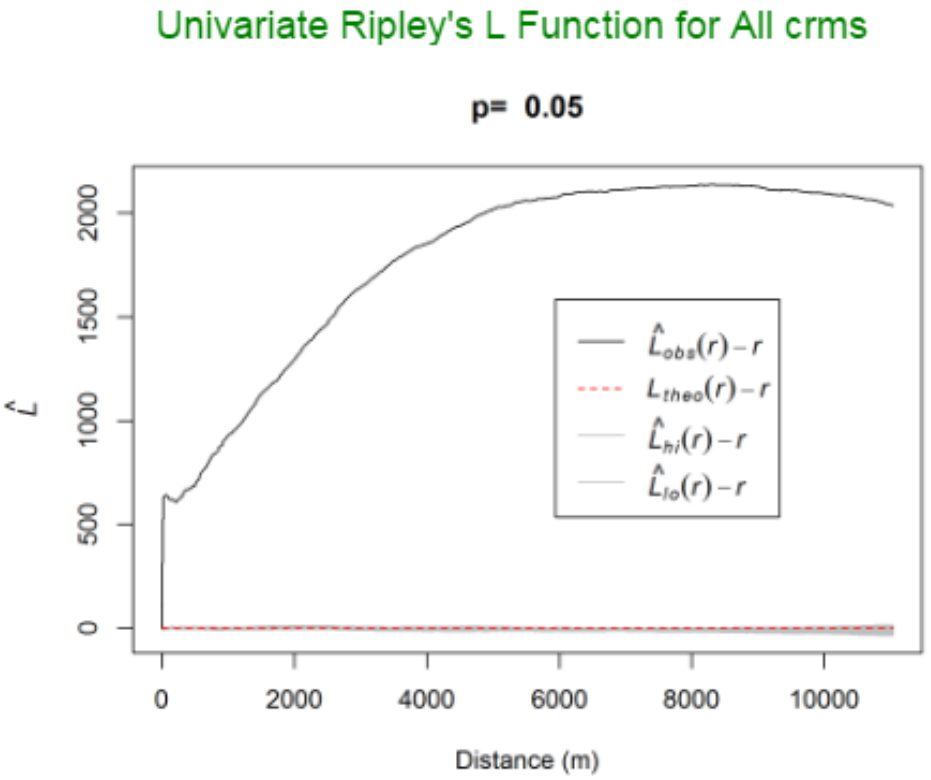
SPATIAL PATTERN OF SAFETY CONCERNS



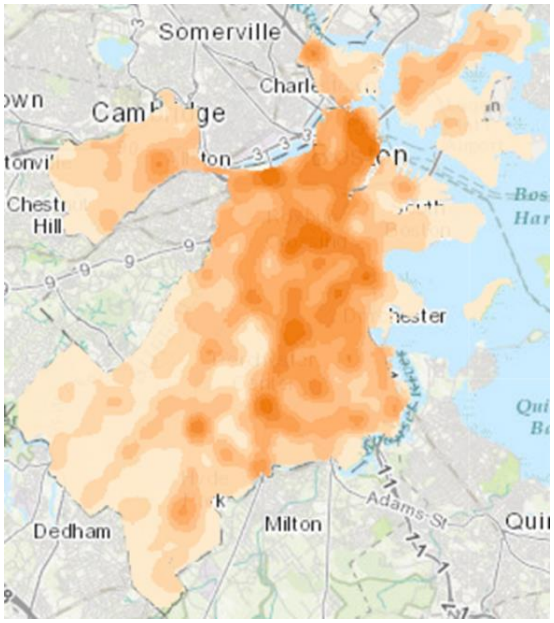


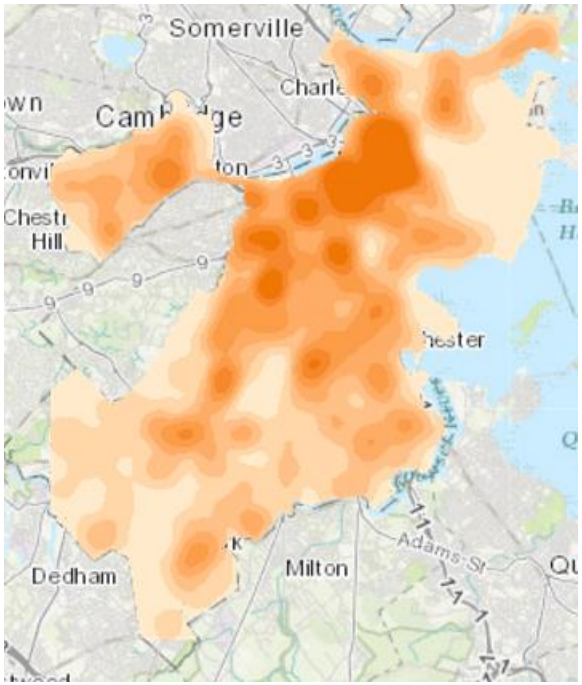
SPATIAL PATTERN OF 311 ENGINEERING REQUESTS

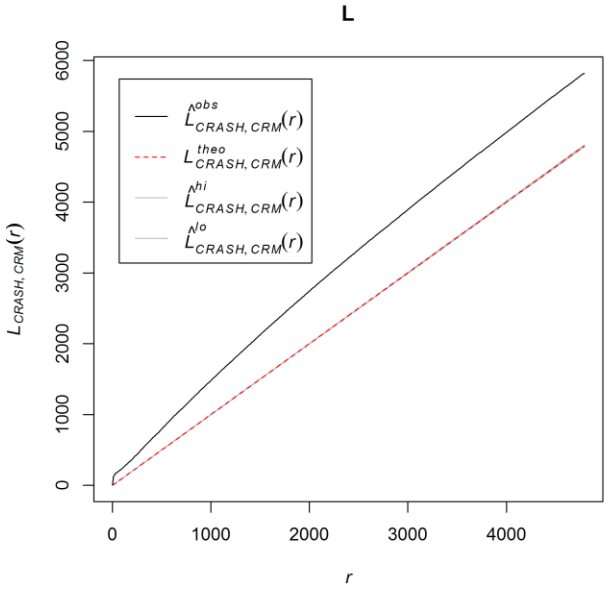




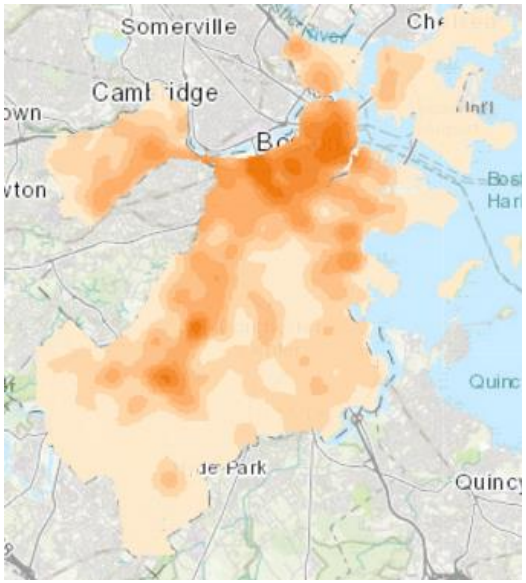
SPATIAL RELATIONSHIP BETWEEN TRAFFIC CRASHES AND 311 ENGINEERING REQUESTS

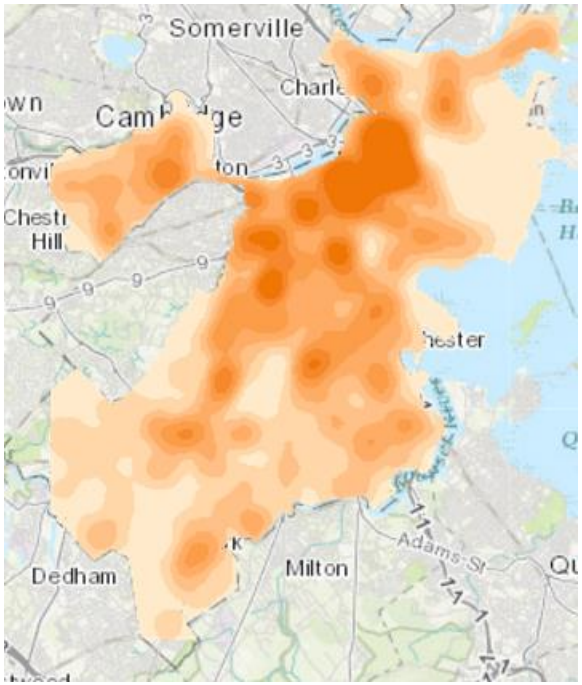


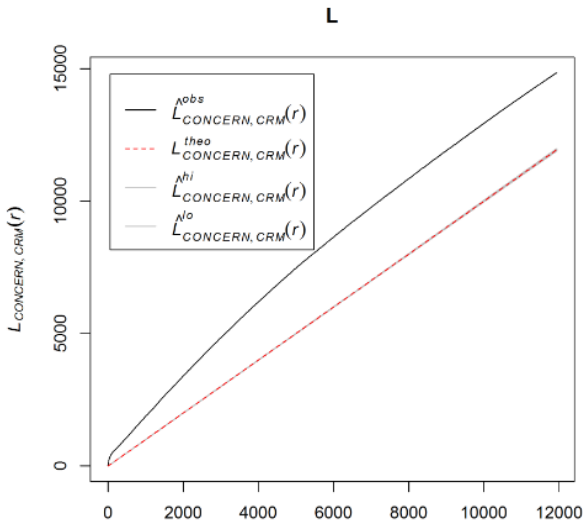




SPATIAL RELATIONSHIP BETWEEN SAFETY CONCERNS AND 311 ENGINEERING REQUESTS







CONCLUSION AND FURTHER STUDY

- The Kernel Density Estimation (KDE) tells us where we have clusters. The Ripley's K/L function does not tell us where we have clusters, but tells us we do have clusters and the clusters are statistically significant.

- The Kernel Density Estimation (KDE) and Ripley's K/L function together indicate we would (very likely) continue to have more Traffic Crashes, more 311 Engineering Requests and more Safety Concerns in these clustered areas in the future.

- See http://app01.cityofboston.gov/proxy/vzanalysispart2 for further study: *Detecting Traffic Crash Patterns and Identifying the most Risky Street Segments and Sections for Vision Zero Boston using Network Constrained Spatial Statistical Methods.*

ABSTRACT

Exploring Spatial Patterns and Relationships in GIS Datasets for Vision Zero Boston

Youshe Li

Department of Innovation and Technology, City of Boston

The spatial statistical methods: Kernel Density Estimation (KDE) and Ripley's K/L Function have been used to explore the question of what data should dictate the engineering focus of the Boston Transportation Department and to understand how well 311 Engineering Requests align with other data such as Traffic Crashes, Safety Concerns, and Waze Traffic Jams.

Statistically, there are clusters on the maps for 311 Engineering Requests, Traffic Crashes, and Safety Concerns at a 95% confidence level, and the clusters for 311 Engineering Requests are spatially correlated with those for Traffic Crashes and Safety Concerns also at a 95% confidence level.

The Kernel Density Estimation and Ripley’s K/L Function together indicate we would (very likely) continue to have more Traffic Crashes, more 311 Engineering Requests, and more Safety Concerns in these clustered areas in the future. Visit http://app01.cityofboston.gov/proxy/vzanalysis/ for more details.

Thank you!

Questions?