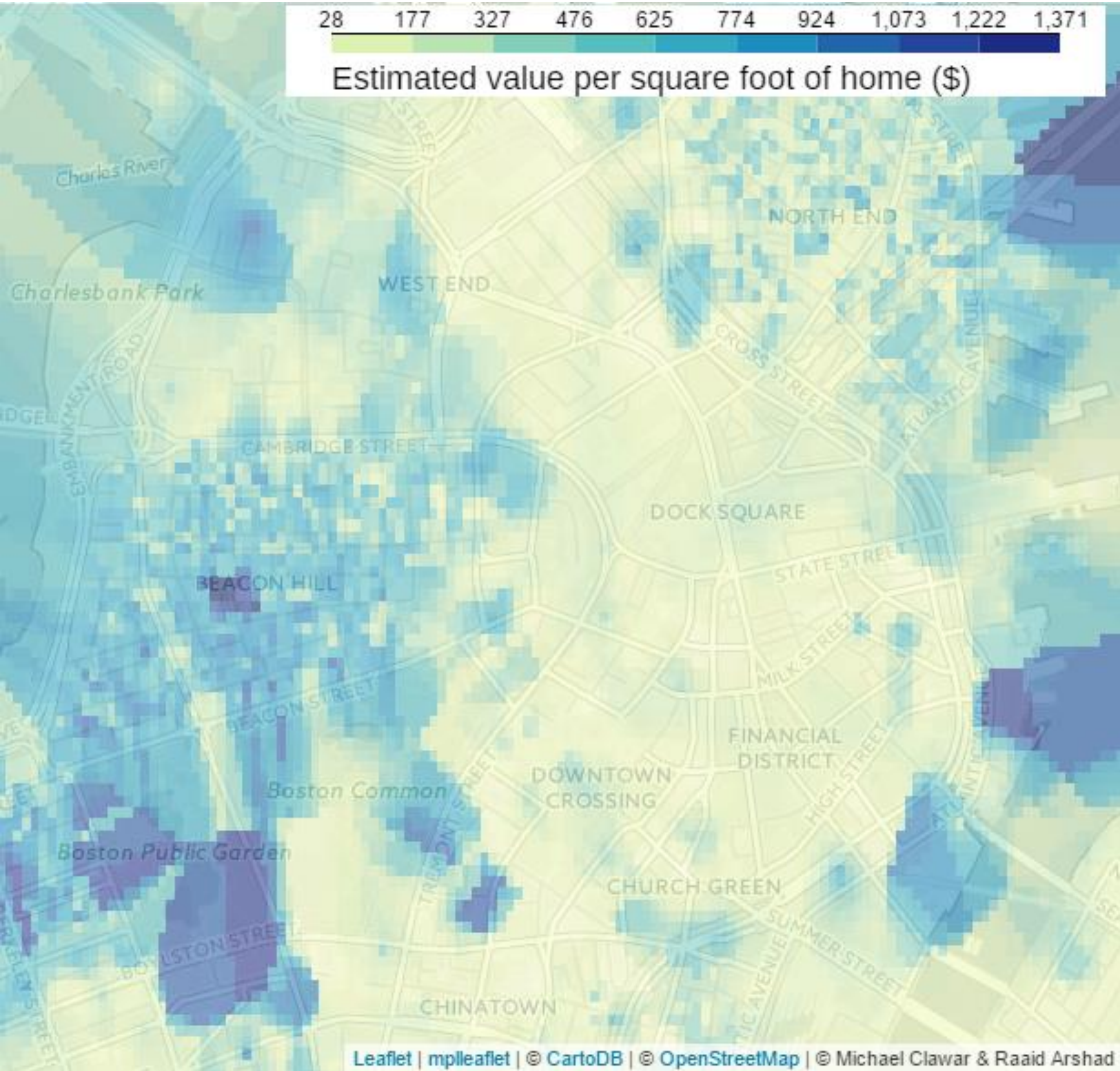


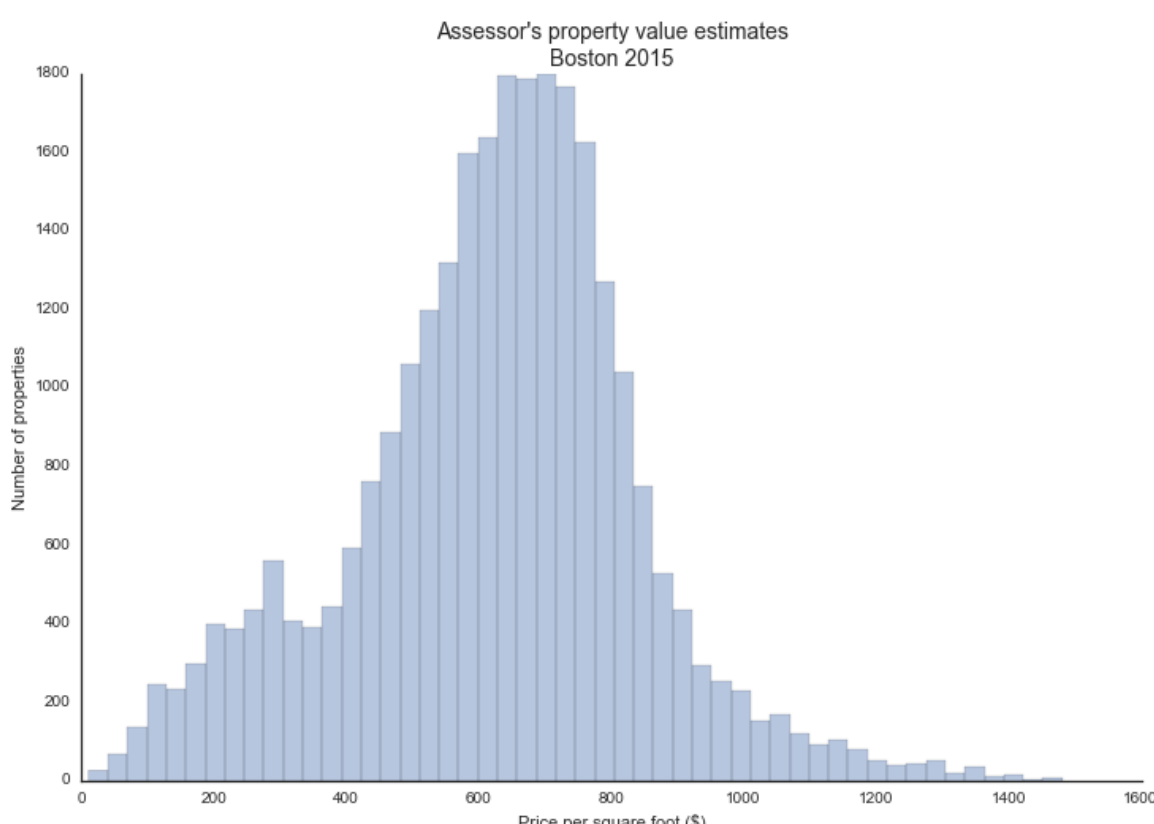
# Neighborhood clustering around Boston

## 2015 crimes committed and property values

Raaid Arshad and Michael Clawar



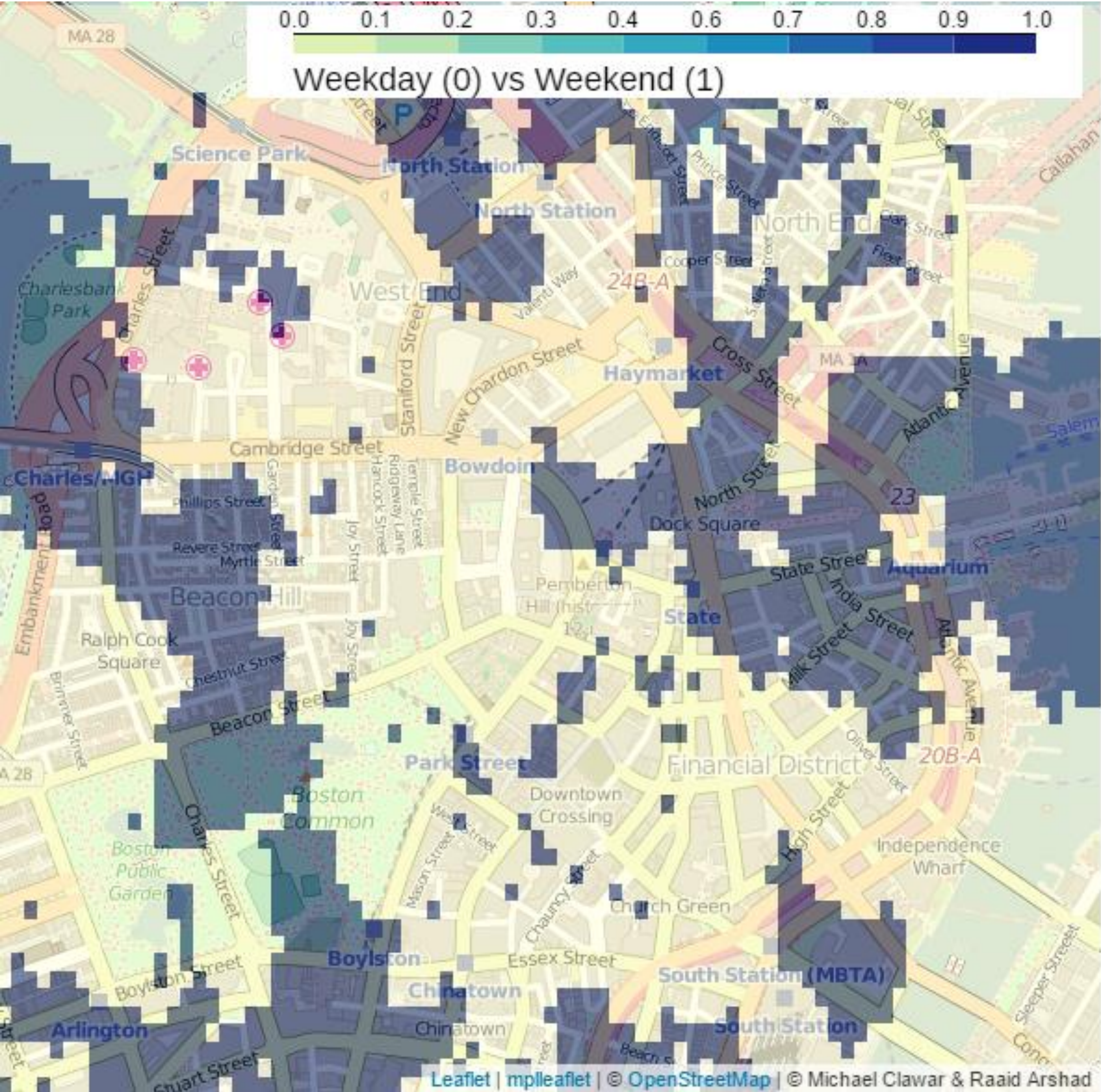
**Methodology:** Using 5-nearest neighbors regression, we fit a model over the area of downtown Boston with latitude and longitude as predictors and home values from 2015 assessors data to create a spatial model of home prices across the city.



Beacon Hill is most obviously expensive area, while Chinatown and East Boston are relatively less expensive. The center of the downtown area, composed of mostly commercial buildings, appears to be less expensive per square foot than the residential areas surrounding it. In addition, the waterfront properties in East Boston are relatively more expensive, as are pockets in the North End.

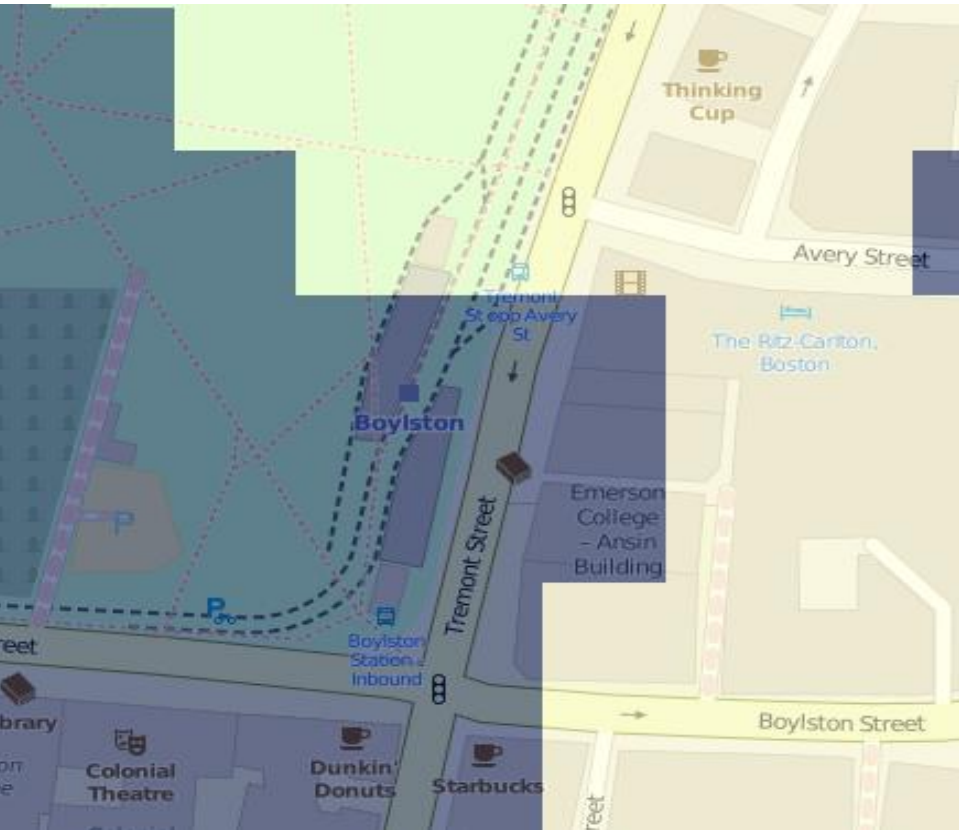
Metric	Property Value
25 <sup>th</sup> percentile	\$499
50 <sup>th</sup> percentile	\$642
75 <sup>th</sup> percentile	\$755
Mean	\$623

Our model also predicts the Boston Public Garden to be relatively expensive, despite having no property in the area. This approach may be a helpful path to pursue for developers or government in valuing public or undeveloped land.

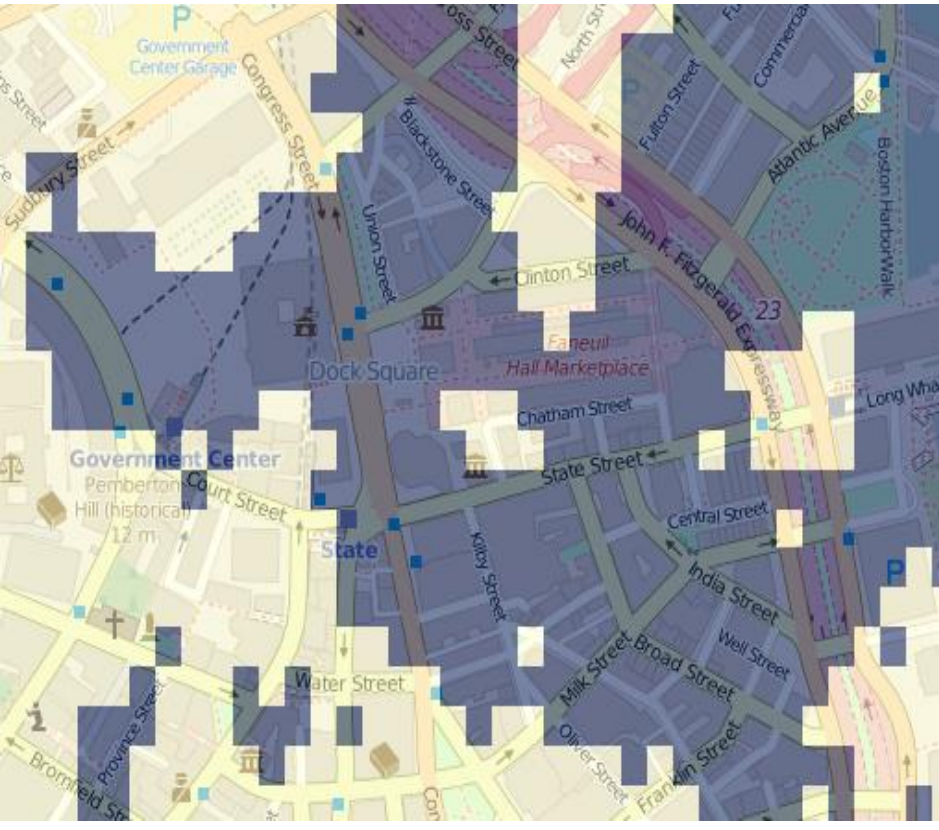


**Methodology:** Using 10-nearest neighbors classification, we fit a model over the area of downtown Boston with latitude and longitude as predictors and all crimes from January to August 2015, flagged with day of the week they were committed.

- From this model, we note three key findings:
- Tourist areas tend to have more crime on weekends
  - T-stops and the surrounding areas appear to also have relatively more crime on weekends than on weekdays
  - The Financial District has relatively more crime on weekdays when it has significantly more possibility for crime given the presence of workers.

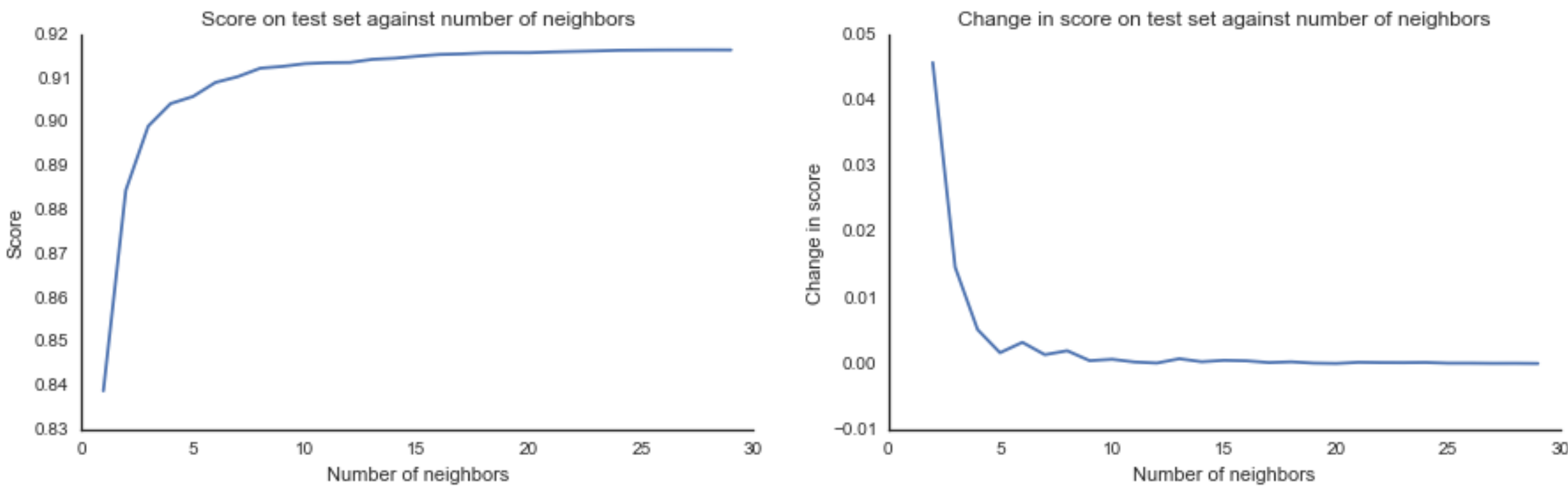


Weekend crime predicted around the Boylston T-stop

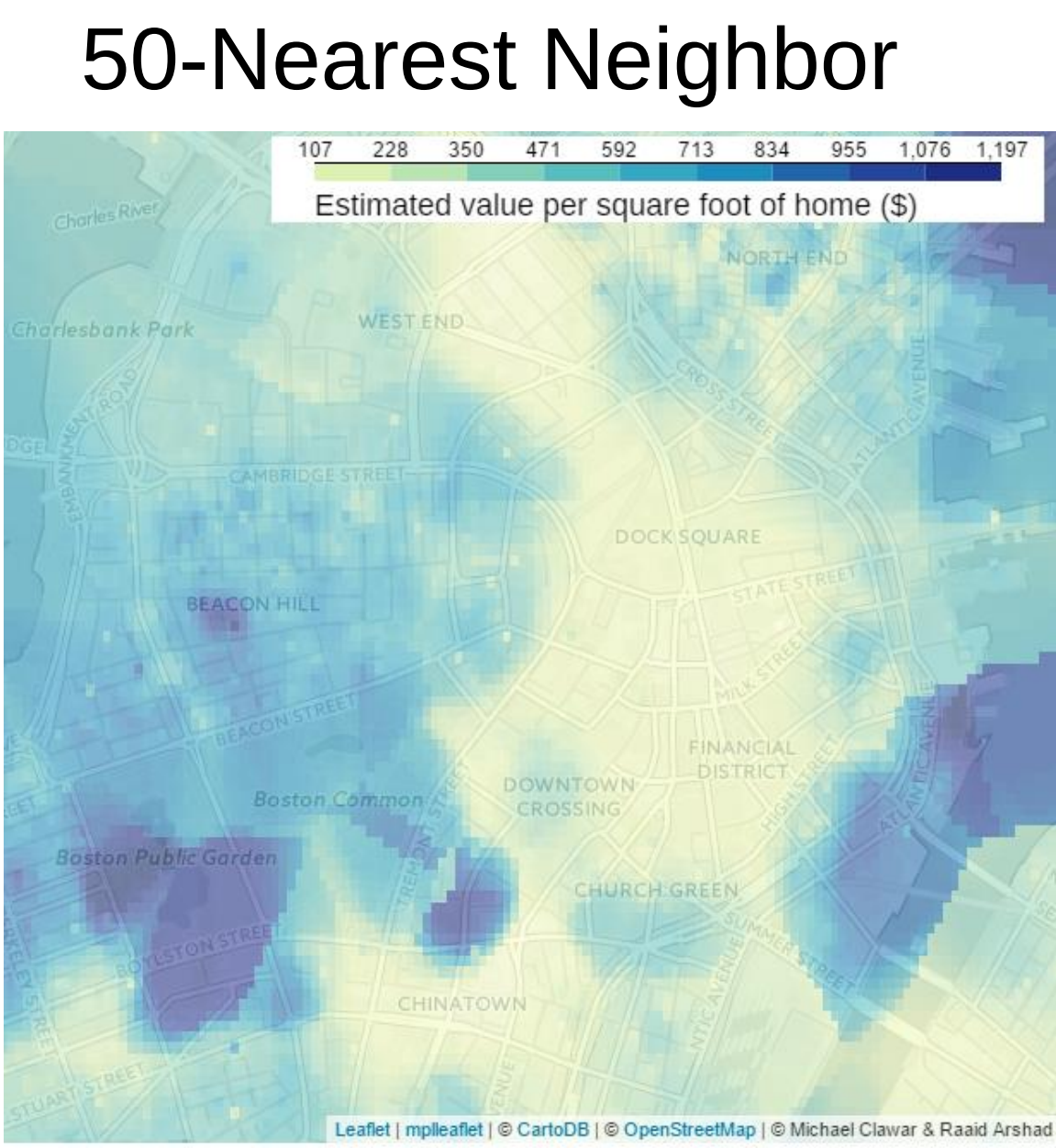
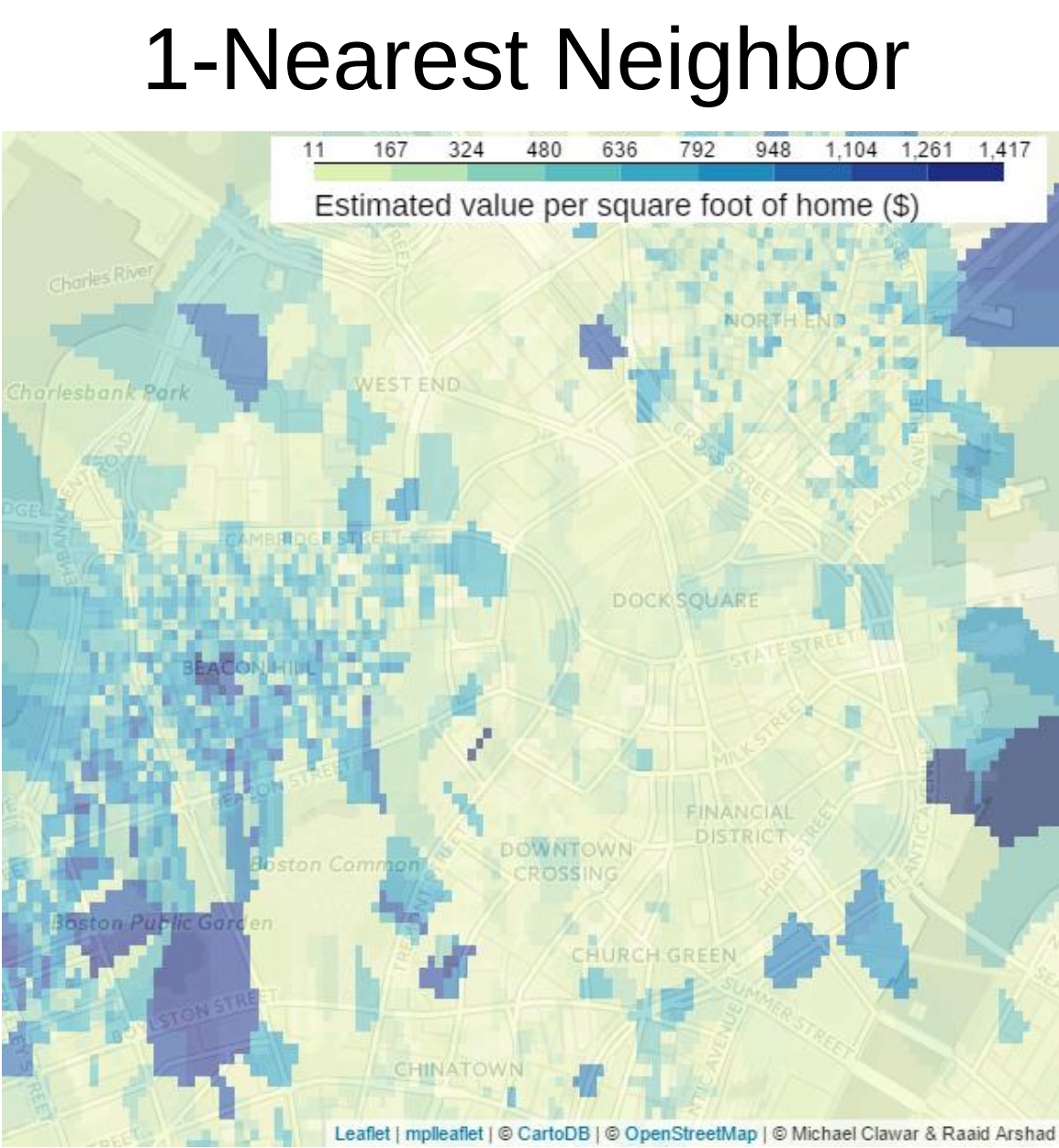


Weekend crime predicted around Faneuil Hall

Using a training set of roughly 23,000 properties, we fit a model for varying numbers of neighbors from 1 to 30. To choose the number of neighbors for the final model, we look at the point where the change in the  $R^2$  (score) for a test set of 5,000 observations levels off. In this case, we chose 5 neighbors for the final model. For crime classification, we use the same methodology.



Below, we display two maps: one with predictions using 1-nearest neighbor and the other using 50-nearest neighbors. The 1-nearest neighbor overfits, creating the pixelated view of the city, while the 30-nearest neighbor model oversmooths, which may not capture property variability from block to block.



Our data reveals clusters of valuable properties in the downtown Boston area, concentrated in the Beacon Hill and waterfront neighborhoods. We also find that tourist and commuter hubs are more likely to experience crime on weekends.