
An Analysis of Pedestrian Safety in Cincinnati, Ohio

Patrick McDevitt

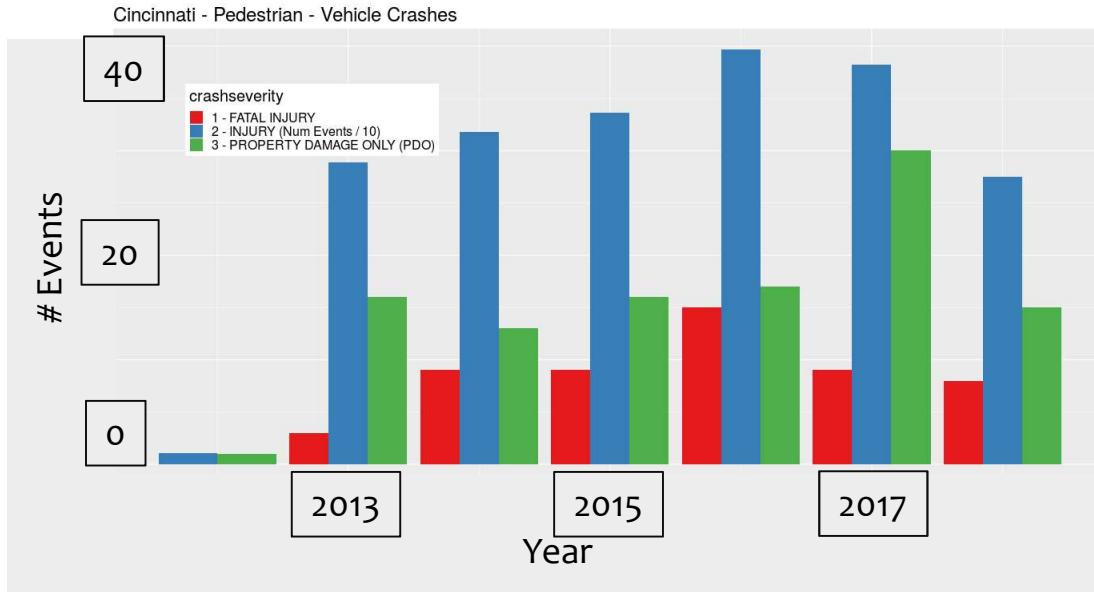
Preeti Swaminathan

Joshua Herrera

Raghuram Srinivas

01-feb-2019

Cincinnati – Pedestrian Safety Experience



Property damage

Injuries (/10)

Fatalities

- Fatalities : ~12 / year
- Accidents w/ Injuries : 350 / year

Model Concept – Grid Cell Aggregation

Pedestrian Accidents

- Cost factor = Sum(accident * severity)
- Event factor = Sum(number events)

Independent Estimators

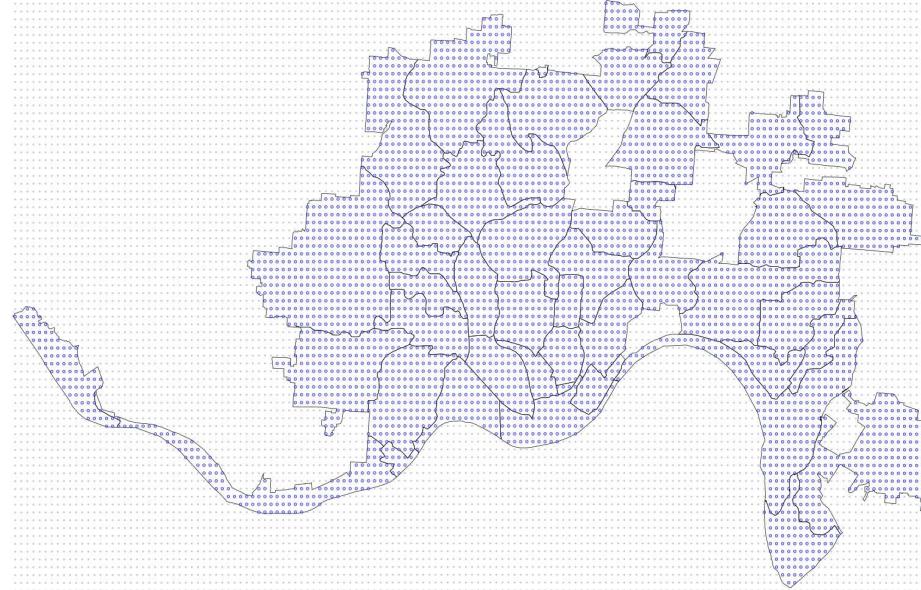
- Non-pedestrian accidents
- Pedestrian survey responses
- Reported near-misses
- Street surface areas
- Walk Scores
- Fire incidents
- Non-emergency service requests
- Property values & uses
- Public transportation accessibility

cell_id	530045	630027	530038	
Pedestrian Event Costs	0.749	10.69	24.993	Traffic Accidents
sum_cost_y	973.74	1724.687	1796.4	
num_events_y	117	196	210	
num_fire_incd	0	0	0	
med_sale_res_y	50,000	989,016	13,356	
med_sale_res_n	25,600,000	12,815,000	33,618	
med_sale_com_y	27,413	351,822	8,303,951	
med_sale_com_n	25,600,000	25,600,000	25,600,000	
med_sale_ind_y	0	444,023	0	
med_sale_ind_n	0	0	0	
med_sale_pbo_y	0	0	0	
med_sale_pbo_n	0	0	0	
mean_walk_score	46	46	68	Walk Score
min_walk_score	44	44	67	
max_walk_score	48	48	69	
num_walk_scores	0	0	3	
sum_lane_cnt	206	89	517	Street Surfaces
sum_width	2024	907	5209	
sum_area	601532	122758	2745087	
num_streets	43	43	97	
dist	0.000416949	0.000330175	0.000343675	Bus Dist
n_object	3	12	6	
animals_insects	36	2	4	
building_related	203	38	52	
construction	0	0	0	
food	0	6	1	
others	124	155	156	
police.property	4	4	3	
service.complaint	2	1	1	
street_sidewalk	11	19	40	
traffic_sigol	0	3	1	
trash	91	45	64	
trees_plants	8	3	29	
water.leak	2	1	1	
zoning_parking	0	0	1	
n_request	481	277	353	
access	0	0	0	
dblprk	0	0	0	
dntyld	0	0	0	
jywalk	0	0	0	
lvisib	0	0	0	
lwfwfs	0	0	0	
nobikef	0	0	0	
noswlk	0	0	0	
other	0	0	0	
prkint	0	0	0	
prkswlk	0	0	0	
speed	0	0	0	
num_near_misses	0	0	0	Misc

- All events occurring within same geographic sector aggregated as a common response – predictor data set

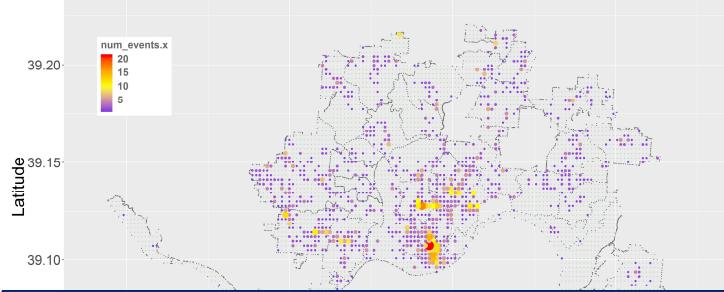
Analysis Method Concept

- Grid cell approach
- Accumulate experience of candidate features in each cell
- Grid cell size : 250m x 250m
 - uniform distribution, equidistant edge distances
- 4196 grid cells
- (~80 sq-miles)



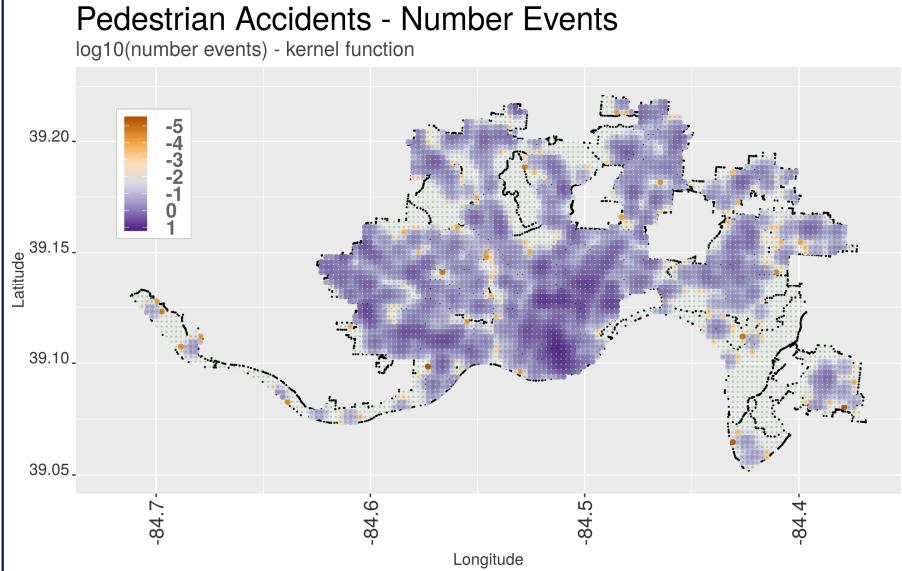
Pedestrian Accidents – Locations / Severity

Pedestrian Accidents - Number Events

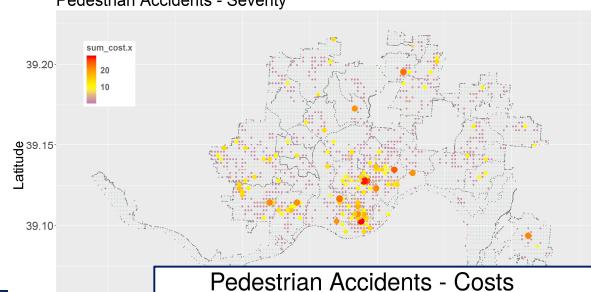


Pedestrian Accidents - Number Events

$\log_{10}(\text{number events})$ - kernel function

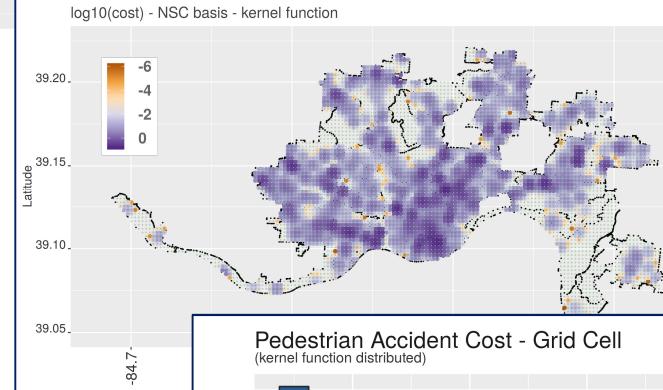


Pedestrian Accidents - Severity

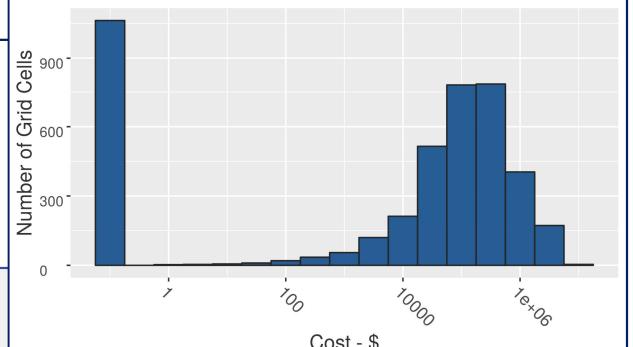


Pedestrian Accidents - Costs

$\log_{10}(\text{cost})$ - NSC basis - kernel function



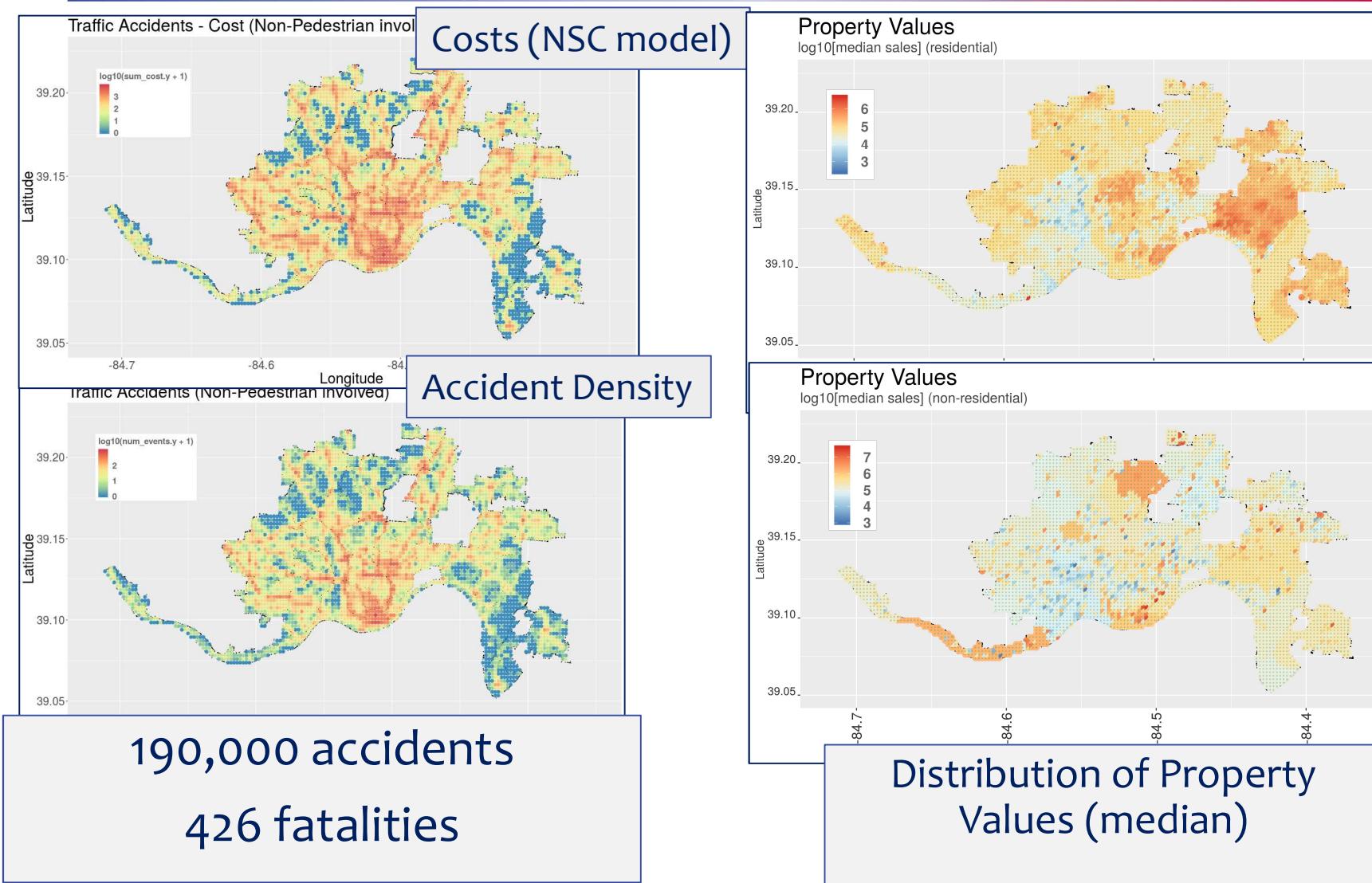
Pedestrian Accident Cost - Grid Cell
(kernel function distributed)



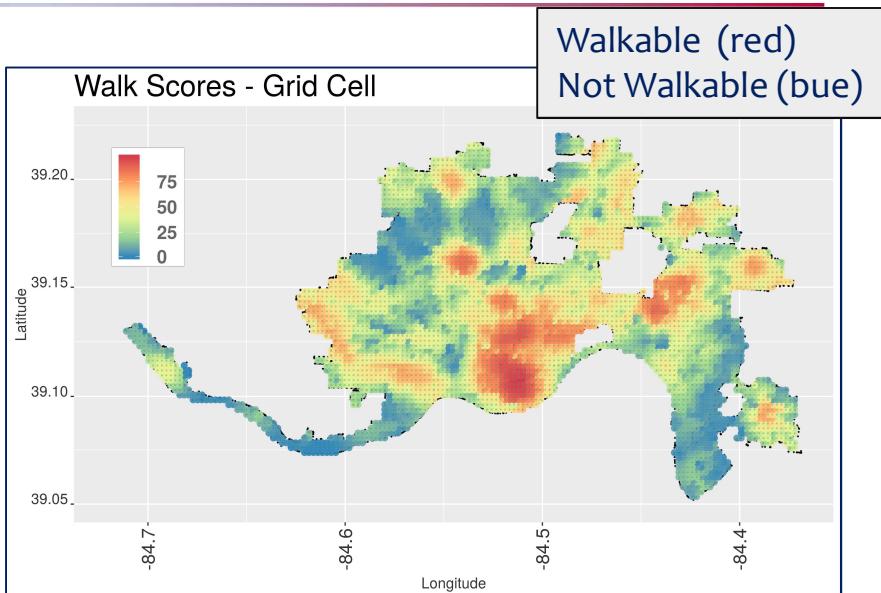
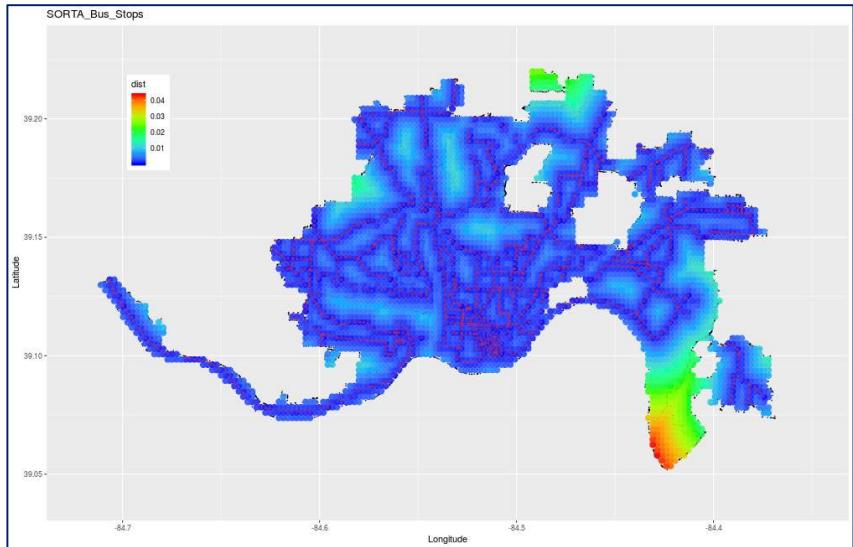
- Characteristics

- widely dispersed across the city
- heaviest concentration at city center & near east neighborhoods
- National Safety Council estimated cost : \$500M / year (all-in)

Traffic Accidents (Non-Pedestrian) / Property Values



Transportation / Walk-ability



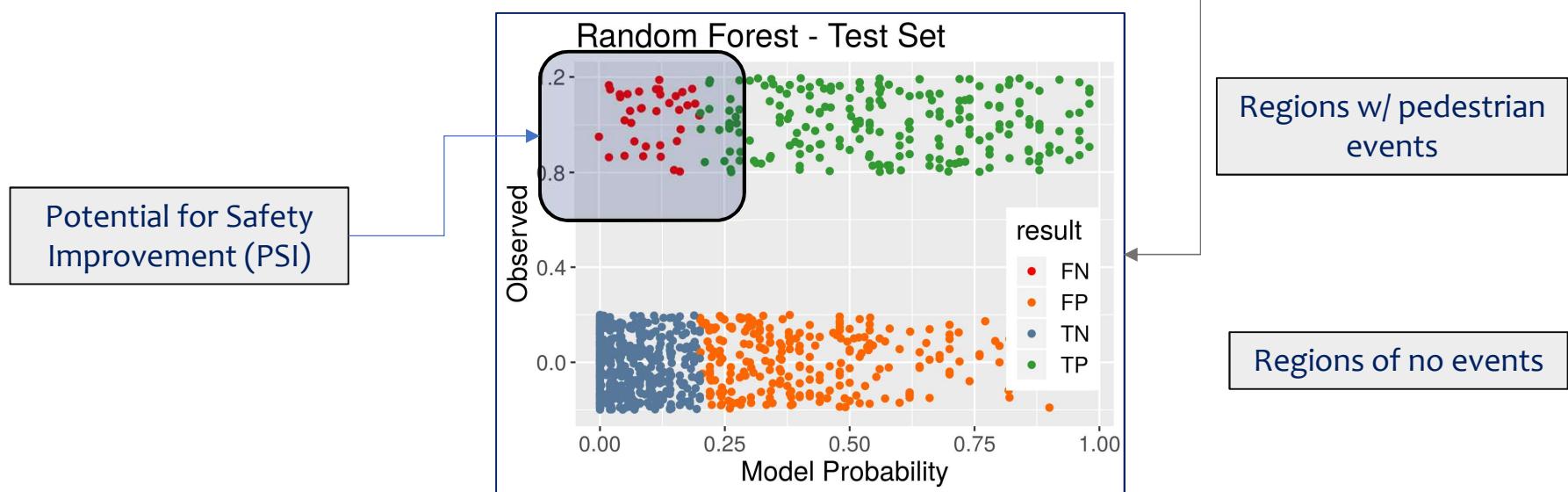
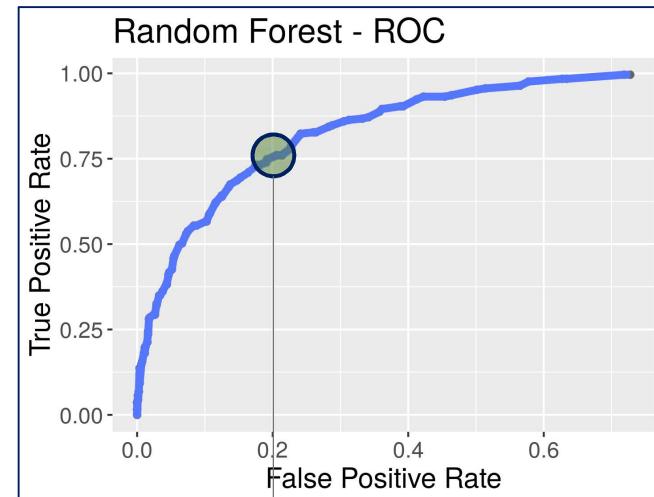
Public transportation accessibility
distance to nearest bus stop

Walk Score – scale from 0 – 100

- walking distances, block length, intersection density
- calculated every city block
- weighted by population density

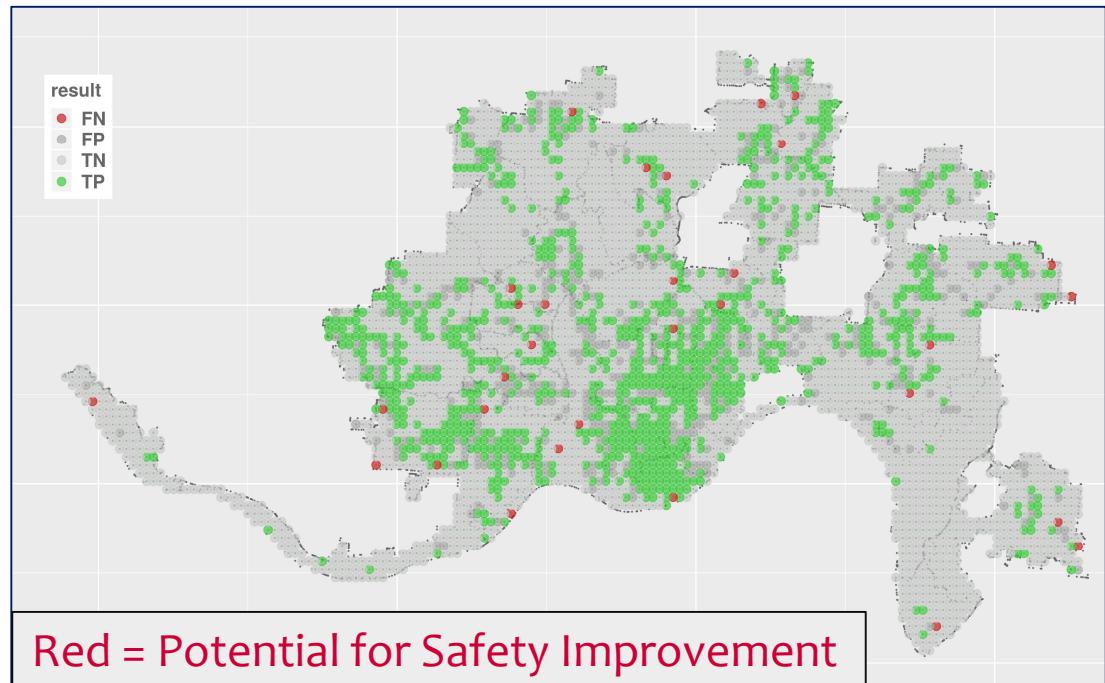
Random Forest Analysis

- Binary & Cost Models implemented
- Binary model provides useful results
 - ~75% TPR / 20% FPR
- Influential factors
 - Non-pedestrian involved accidents
 - Public transportation accessibility
 - Number of service requests for buildings, sidewalk, and vegetation

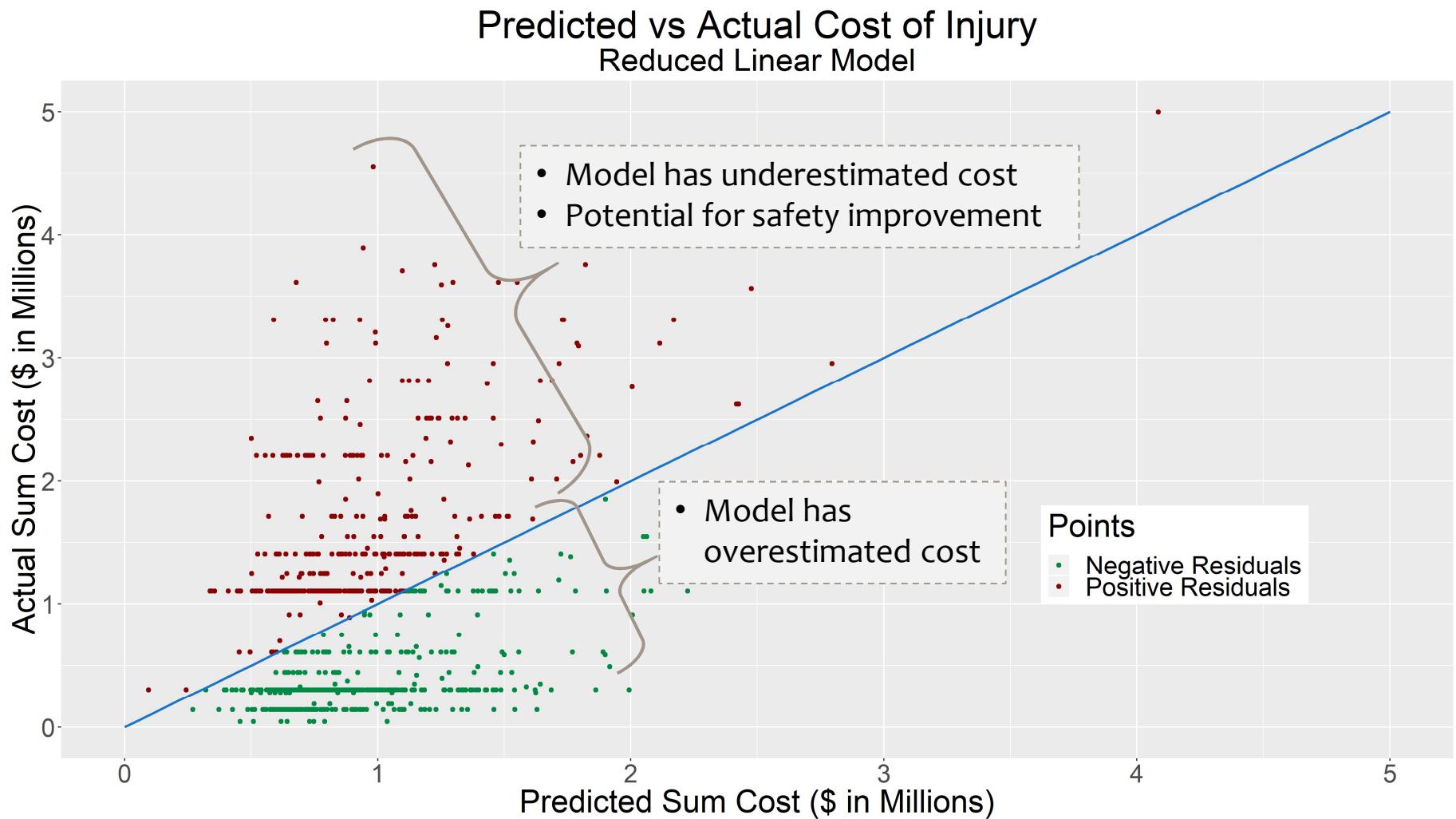


Random Forest – Geographic View

- Strong contiguity for TP, FP, FN regions
- Large surface areas of TN – high model accuracy
- PSI identification (FN) perimeter of TP regions
 - Widely distributed (consistent with experience of accidents)
 - Some at perimeter of city incorporation boundary
 - Reasonable number of specific regions for further investigation



Regression Analysis

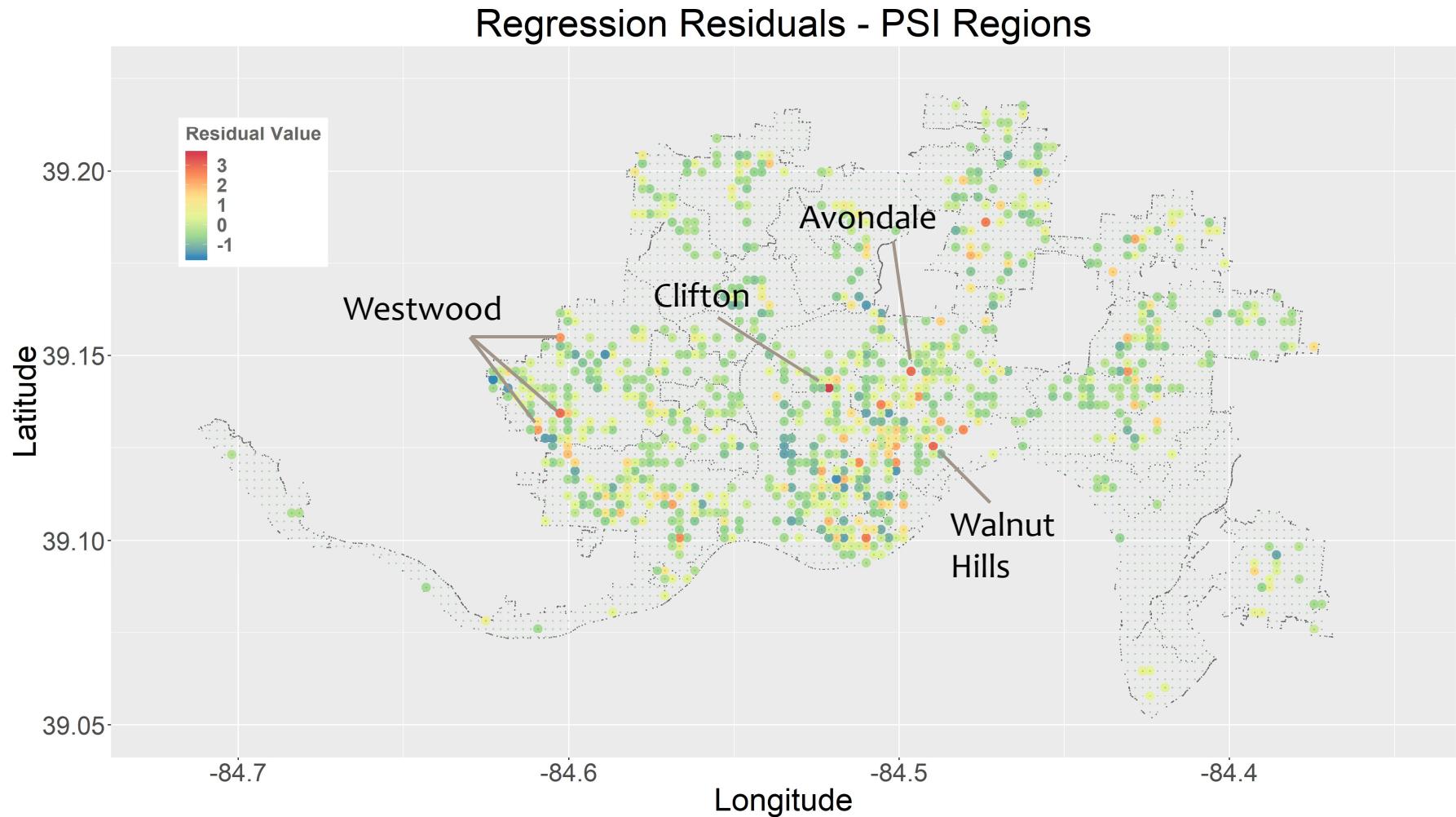


Variable Analysis

Coefficients Sorted by Significance	
	Pr(> t)
Sum Cost of Damage to People in Non-Pedestrian Accidents	< 2e-16
Unsanitary Food Operations	1.40E-06
Maximum Walk Score	0.0014
Water Leaks and Breaks	0.0037
Crosswalk Needed	0.0105
Number of Police Properties	0.0138
Vehicles Running Red Lights / Stop Signs	0.0386
Unkempt Flora	0.0643
Speeding	0.0666
Parking Close to Intersection	0.0778
Parking on Sidewalk	0.0812

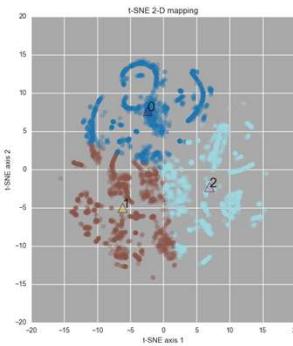
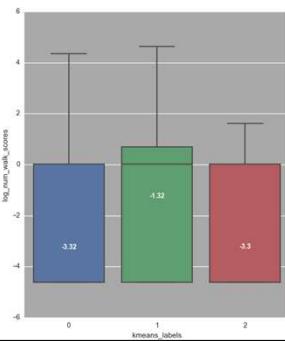
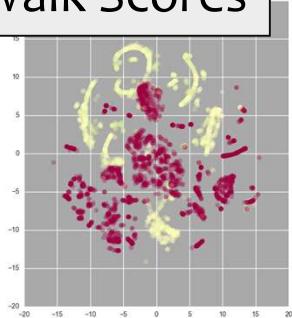
Coefficients below 0.05 are not considered statistically, but are suggestive of being relevant to the model's predictive capabilities

Visualization of Results



Non-Supervised Learning – K-means Clustering

Walk Scores

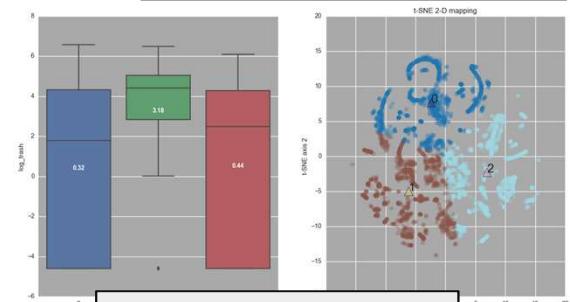


Magnitude distribution of feature

Box plot distribution of property values for each cluster

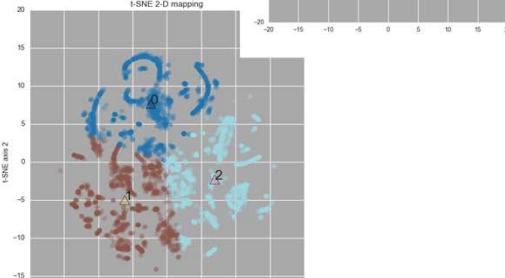
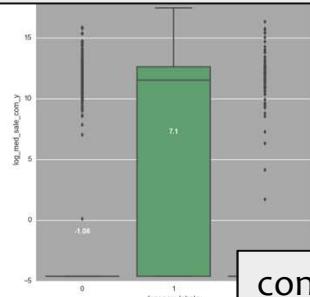
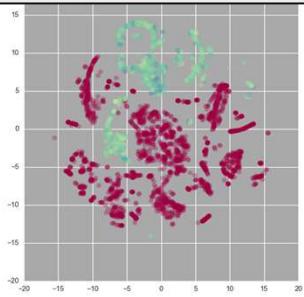
k-means cluster identification in t-SNE 2D mapping

Service Requests



highest in 2nd cluster

Commercial Property Values



concentrated in 2nd cluster

Non-Supervised Learning

Cluster 0: Blue

Primarily Residential

Cluster 1: Green

Primarily Commercial

Cluster 2 : Red

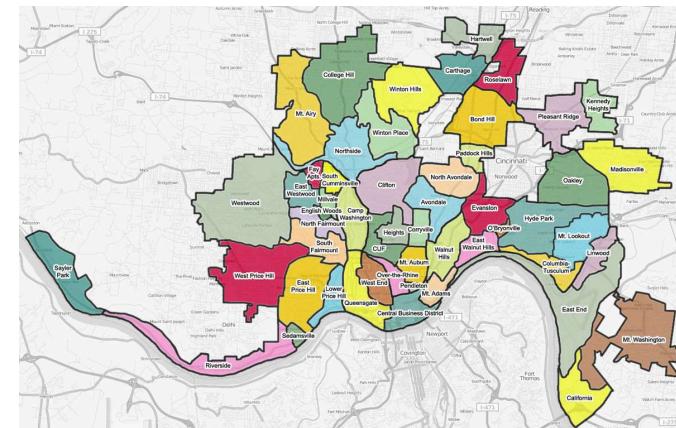
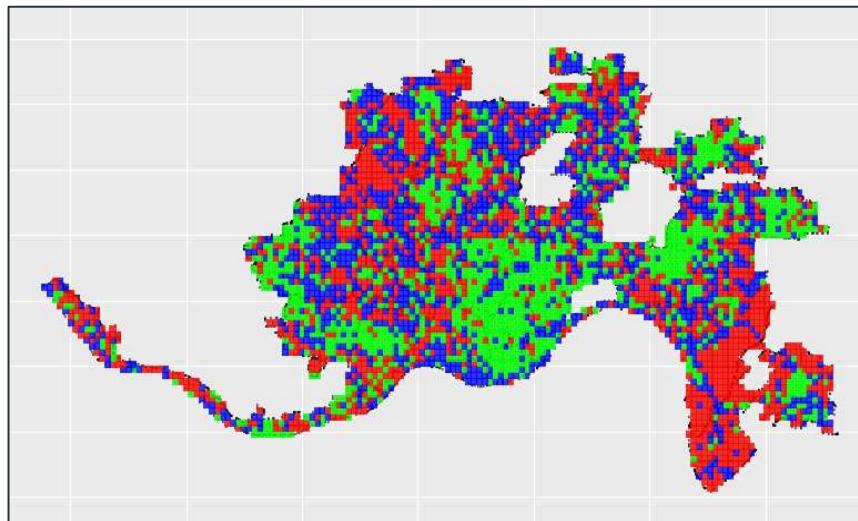
Industrial

Identifying characteristics

- Animals and Insects
- Unkempt Vegetation
- Vehicles Running Red Lights and Stop Signs

- Speeding
- Water Leaks
- Heavy Traffic
- Vehicles not Yielding

- Metal Trash
- Cars Unable to Find Parking
- Lack of Sidewalks
- Potholes



Conclusions

- PSI Regions Identified
 - ~30 grid cells targeted from random forest binary model
 - RF Binary model – 75% TP / 20% FP Rate
 - Regression – R^2 consistent with published results from other cities
- Influential Factors
 - Non-pedestrian involved accidents
 - Public transportation accessibility
 - Number of service requests for buildings, sidewalk, and vegetation
- Non-Supervised Learning
 - Westwood, Mt. Washington and East End
 - Speeding vehicles and non-yielding to pedestrians
 - Clifton, Madisonville, Walnut Hills and Mt. Washington
 - Construct additional sidewalks and crosswalks
 - Over-the-Rhine, Downtown and Clifton
 - Increase bike-ability

appendix

Table 3. Data sets supporting model

Data Set	Description
Pedestrian accidents ⁷	Traffic Crash Reports (CPD) - Contains date, time, weather, location, road information, severity, demographic information
Traffic accidents	Same data source as for pedestrian accidents
Streets Infrastructure ⁸	Street Centerlines (w/ PCI rating) - Street segment locations, length, width, area, material
Bus transportation ⁹	SORTA Bus Stops - Bus stops along SORTA bus routes in the city of Cincinnati, locations, line name
Non-emergency requests ¹⁰	Cincinnati 311 (Non-Emergency) Service - dates, request type, location, agency responsible, status
Cincinnati pedestrian safety survey data ¹¹	Survey input data from citizens; location, date, concern type, comments
Pedestrian near miss data ¹²	Locations, dates of near-miss incidents reported directly to the DOTE
Property valuation ¹³	Each property transfer, date of sale, purchase amount, property, buyer, and seller information
Walk Score ²⁰	A scoring scale rating the walkability of every address in the city
Google Maps	API used to identify latitude and longitude coordinates from local addresses
Zillow ^{14,15}	Zillow Home Value Index provides time-series of data for median market values, used to support property valuations. Zillow neighborhood provides shapefiles used to delineate neighborhood boundaries in the city.

where $Cost_g$ is the crash cost assigned to grid cell g , d_{ig} is the distance from the identified pedestrian accident site to the local grid cell centroids, r is the (constant) search radius, C_i is the cost severity of the accident i , and ρ_i is

⁷<https://data.cincinnati-oh.gov/Safer-Streets/Traffic-Crash-Reports-CPD-/rvmt-pkmq>

⁸<https://data.cincinnati-oh.gov/Fiscal-Sustainability-Strategic-Investment/Street-Centerlines-w-PCI-rating-/574p-8utc>

⁹<https://data.cagisportal.opendata.arcgis.com/datasets/sorta-bus-stops>

¹⁰<https://data.cincinnati-oh.gov/Thriving-Healthy-Neighborhoods/Cincinnati-311-Non-Emergency-Service-Requests/4cjh-bm8b>

¹¹direct communication from Cincinnati DOTE

¹²direct communication from Cincinnati DOTE

¹³https://www.hamiltoncountyauditor.org/transfer_policies.asp

¹⁴https://www.quandl.com/data/ZILLOW/M26_NFS-Zillow-Home-Value-Index-Metro-NF-Sales-Cincinnati-OH

¹⁵<https://www.zillow.com/howto/api/neighborhood-boundaries.htm>