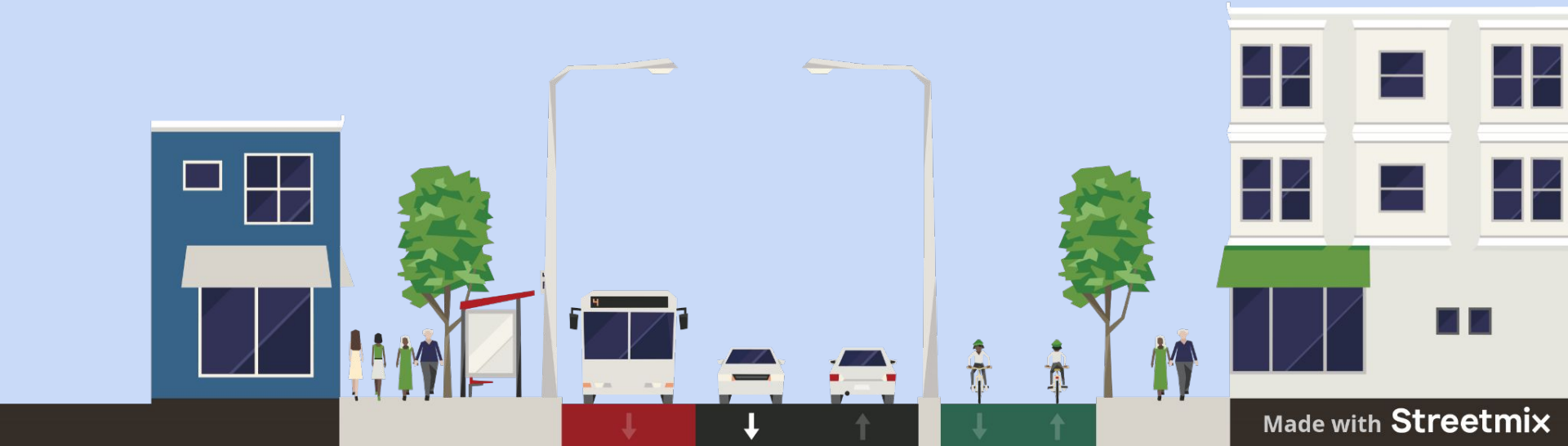
Analysing Road Safety 

Erdős Institute project
Assaf Bar-Natan, Jesse Frohlich, Pedro Lemos, Jacob van Hook

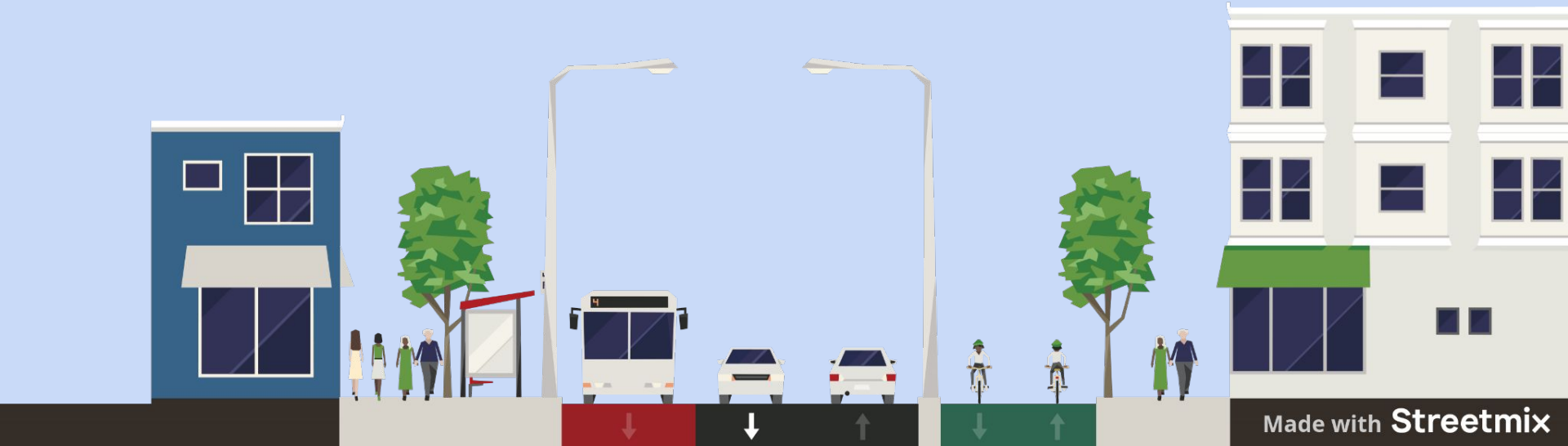
Overview

Increasingly, cities across North America are adopting postures that injuries and deaths on roadways are not inevitable, but *preventable*, and that safety is a property of a street's design, and less a consequence of enforcement. This is commonly referred to as "Vision Zero".

Goal: Which changes make streets safer?



Goal: Predict collision rate from physical design

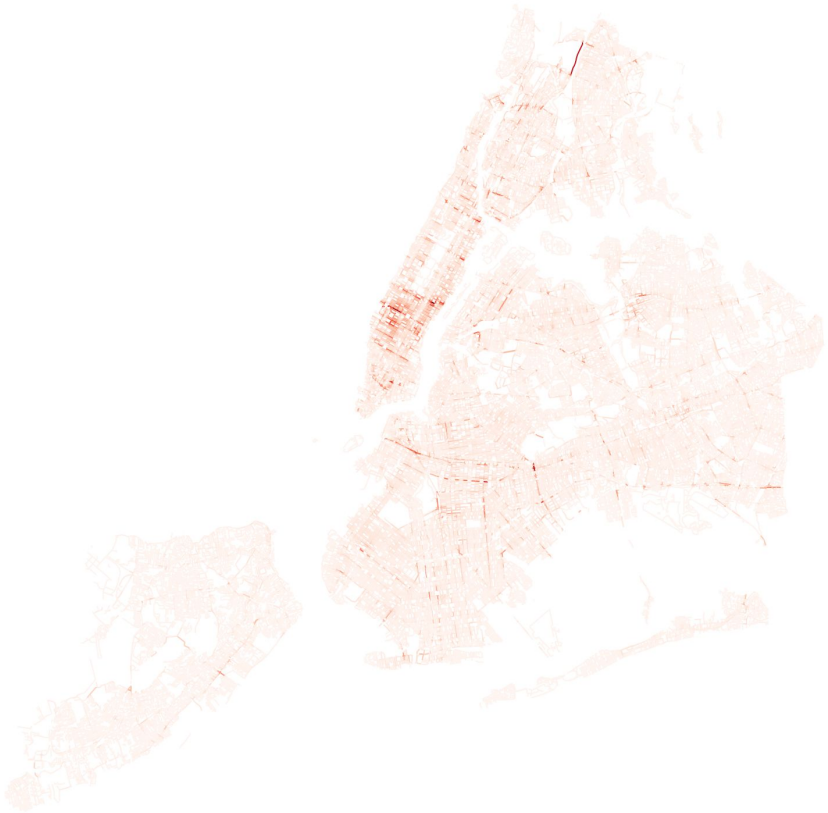


New York City

This city has both a Vision Zero policy and many publicly available data.

Stakeholders:

- **Residents**
- **City Planners**
- **Neighbourhood advocacy groups**



(Higher collision rates in darker red)

Datasets

NYC Open Data

- Street width
- Weekly collision rate
- Presence of bike lanes
- Number of trees planted
- Traffic volume
- Presence of speed humps
- Presence of parking meters
- Speed limit
- ... and more

Data Processing

Selecting streets whose traffic volume is measured

Street segments with traffic volume data shown in blue.

Data Processing

Temporal and spatial transformations

- Streets which underwent a recent change (e.g. installing speed humps) were split into multiple entries.
- Geographic data was assigned to the street it lay on.

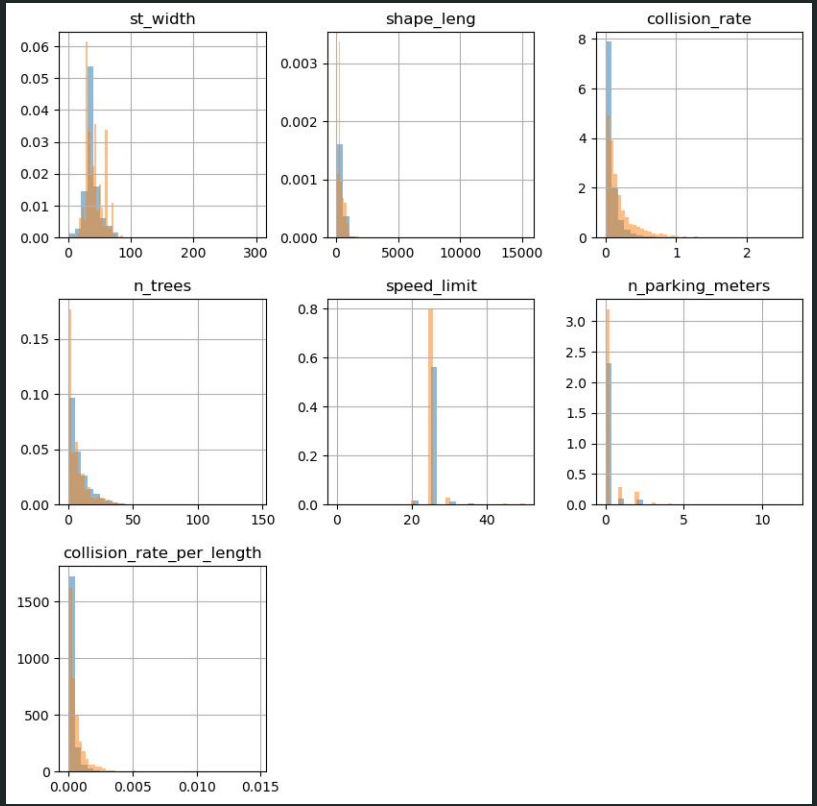
Initial Observations

Observations on collisions:

- Relatively rare
- Happen in city centers
- Happen along certain corridors
- Concentrated on a few streets

Initial Observations

Streets with volume data are similar to ones without

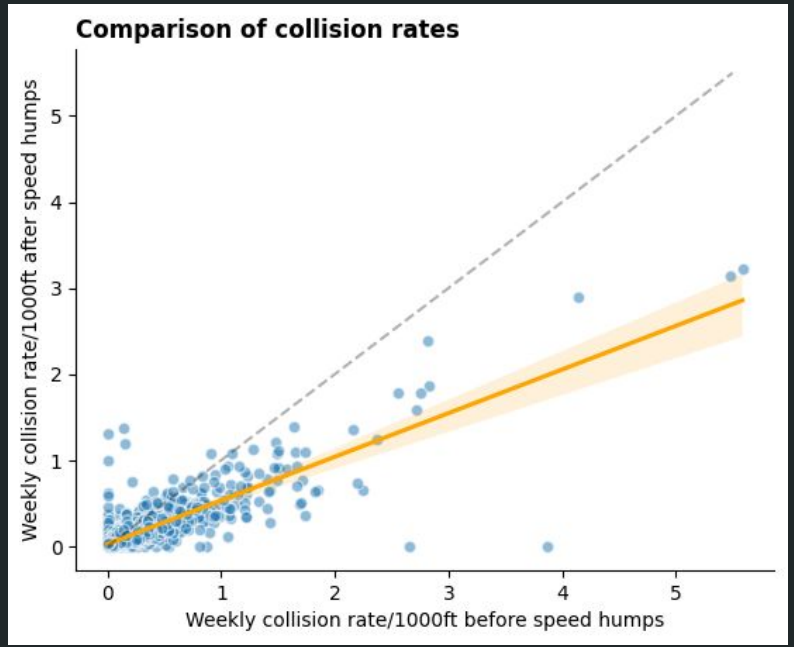


Distributions of volume vs no volume data

Initial Observations

Speed hump analysis

For roads with recent speed hump installations, we observed a trend consistent with other research:



Models

- Baseline constant model
- *k*-nearest neighbours
- Linear regression
- Random forest with XGBoost

Model Performance

Best model:
K-Nearest Neighbours

Model	Mean Square Error ($\times 10^{-7}$)
Baseline	4.85
Linear regression	3.83
Random forest	3.56
K-Nearest Neighbours	3.48

Feature Importance

Using random forest to determine feature importance

Top 5 important features:

- Traffic Volume
- Street Width
- Length of road segment
- Number of Trees
- Type of street (ave or st)

Less impactful features:

- Speed limit
- Has a bike lane
- Parking meters
- Speed humps

Final thoughts

We are thankful to the Erdős Institute for this opportunity and the continued support!

Special thanks go to Alec Clott, Steven Gubkin, Matt Osborne, and Roman Holowinsky, with special credits to Michael Darcy.