

# Mini-Project: Analysis of NYC Vehicle Collisions

CS-GY 6323

Andrew Liang, Caroline Barker, Omar Hammami

4/20/2020

# Overview

- Background information
- Goals
- Questions
  - Question 1 Analysis
  - Question 2 Analysis
  - Question 3 Analysis
- Summary
- Implications

# Background Information

- “Motor Vehicle Collisions – Crashes” dataset provided by NYPD
- Contains
  - Location
  - Time
  - Contributing Factor
  - # of casualties
  - Vehicle Type
- Date Created: April 28, 2014
- Date Updated: April 18, 2020, Daily
- 1.67M Rows

GOAL:

Determine which areas should be  
targeted for redesign efforts

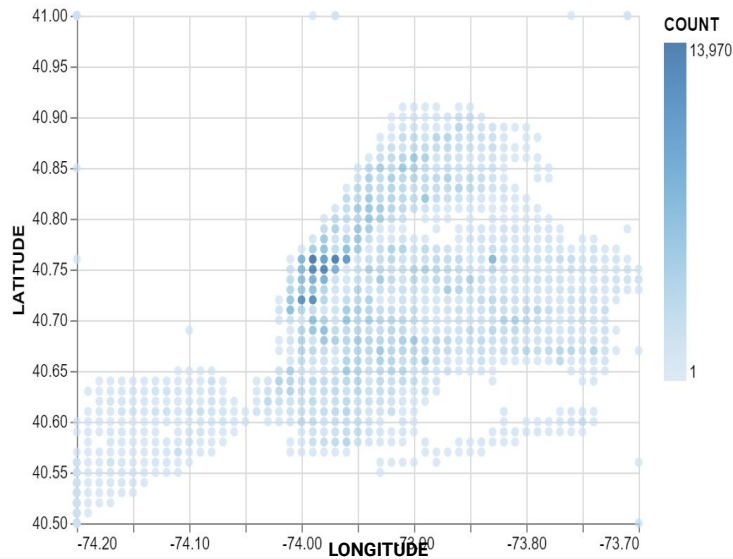
# Questions

1. Which places are the most collision prone?
2. How do casualties compare between types of commuters?
3. At what times are collisions more likely to occur?

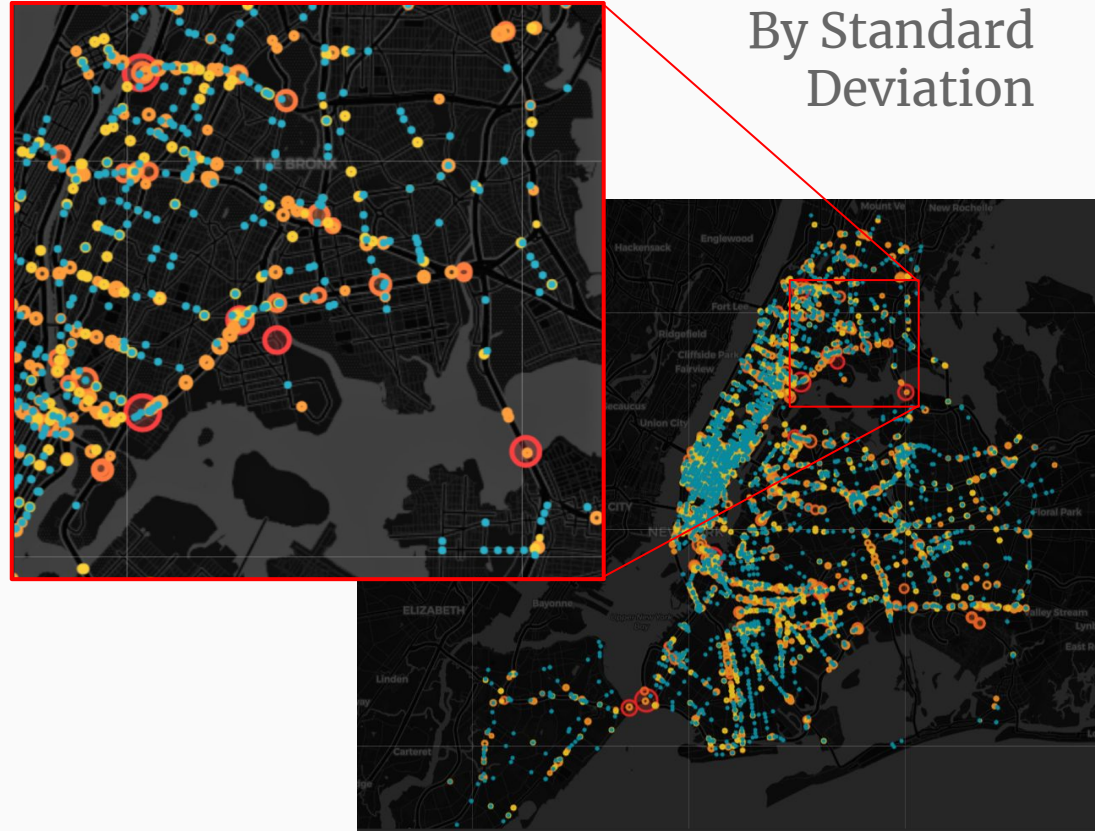
Question 1:  
Which places are the most  
collision prone?

# Question 1: Which locations are most collision prone?

## By Amount of Collisions



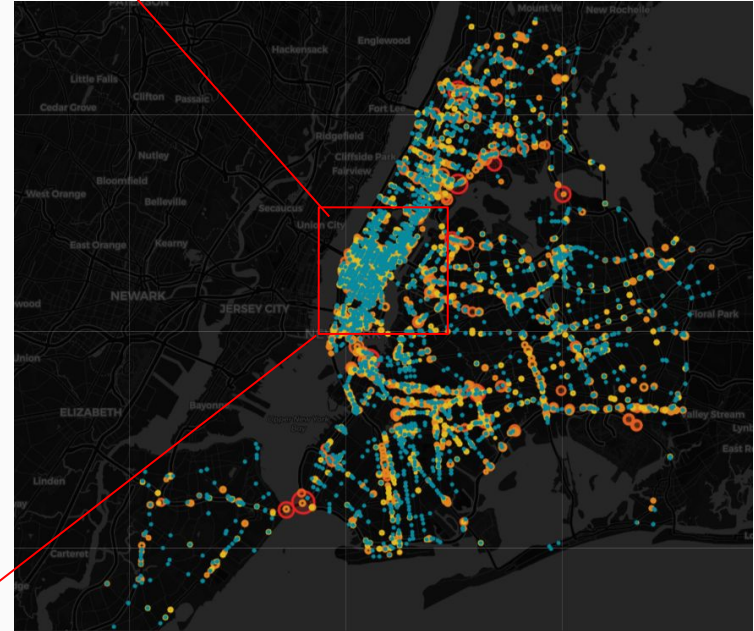
## By Standard Deviation



# Question 1: Which locations are most collision prone?



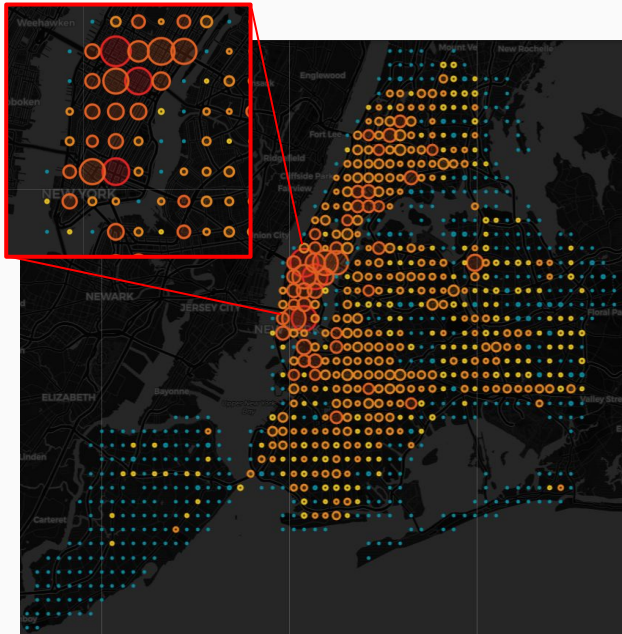
By Standard  
Deviation



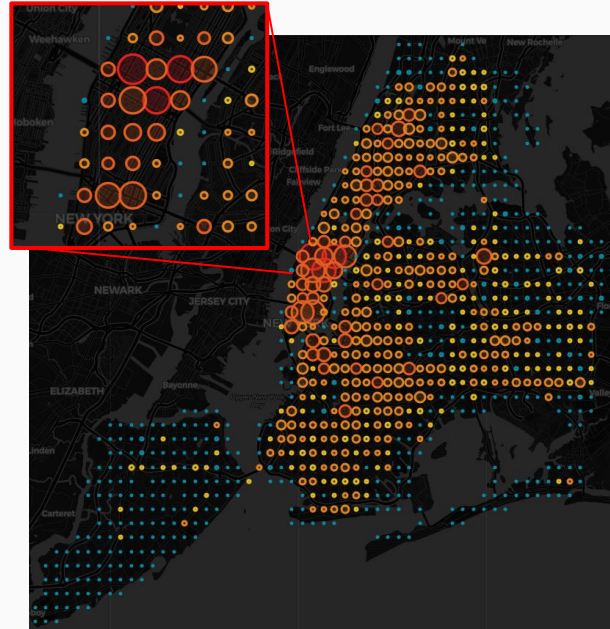


## Question 1: Which locations are most collision prone?

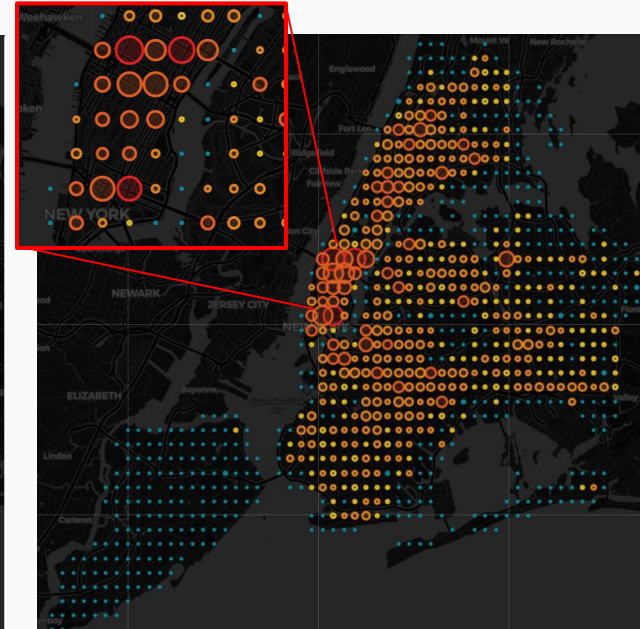
## Rounding Coordinates and separating by year



2017



2018



2019

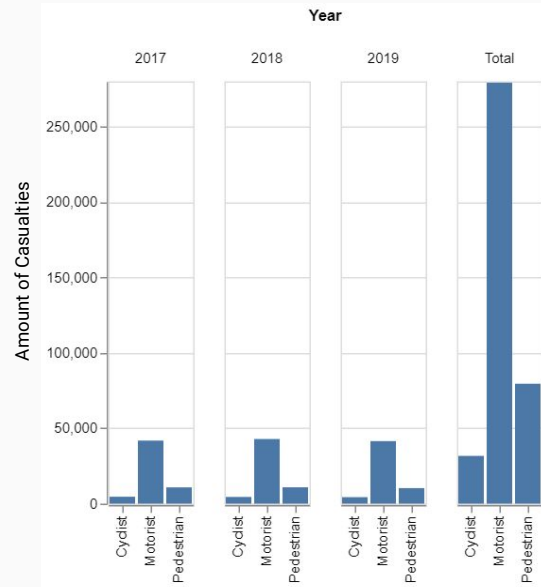
## Question 2:

How do casualties compare  
between type of commuters?

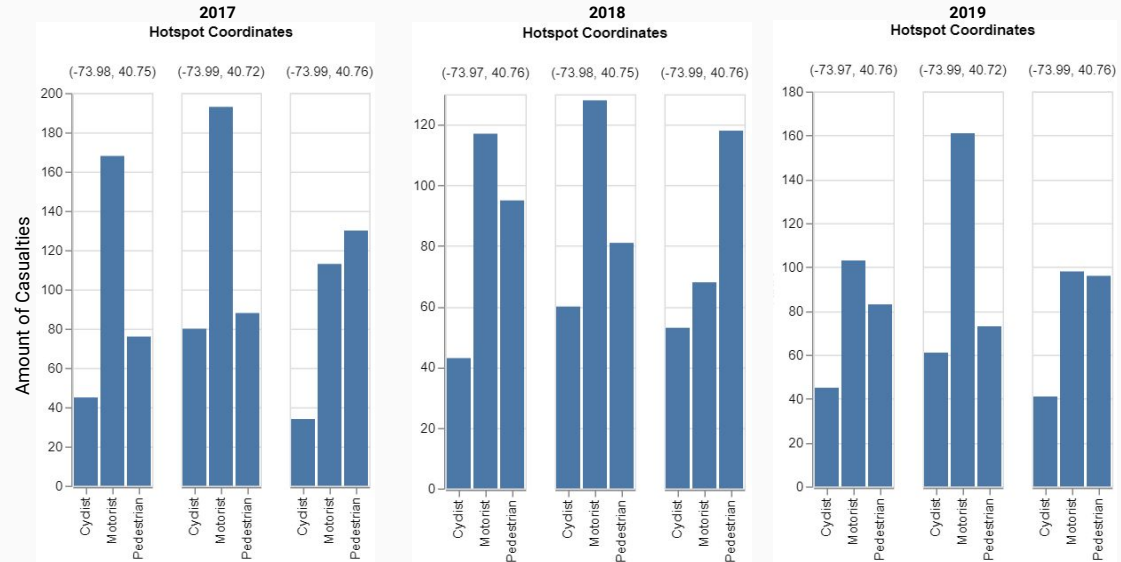
# Question 2: How do casualties compare between type of commuter?

## Total casualties between pedestrians, cyclists, and motorists

For all locations for all years



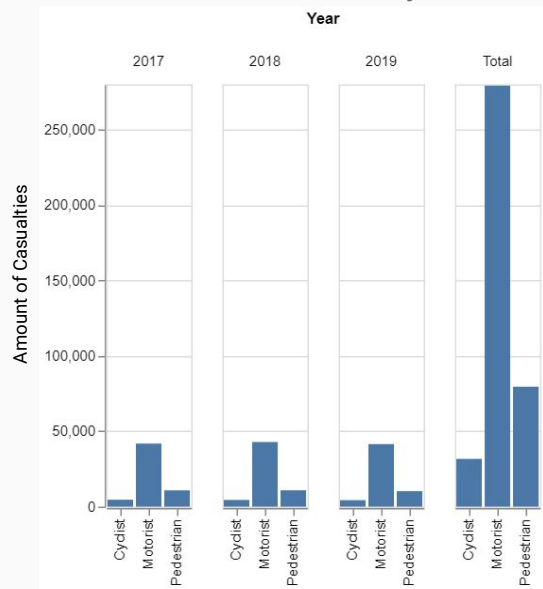
Hotspot Locations for each Year



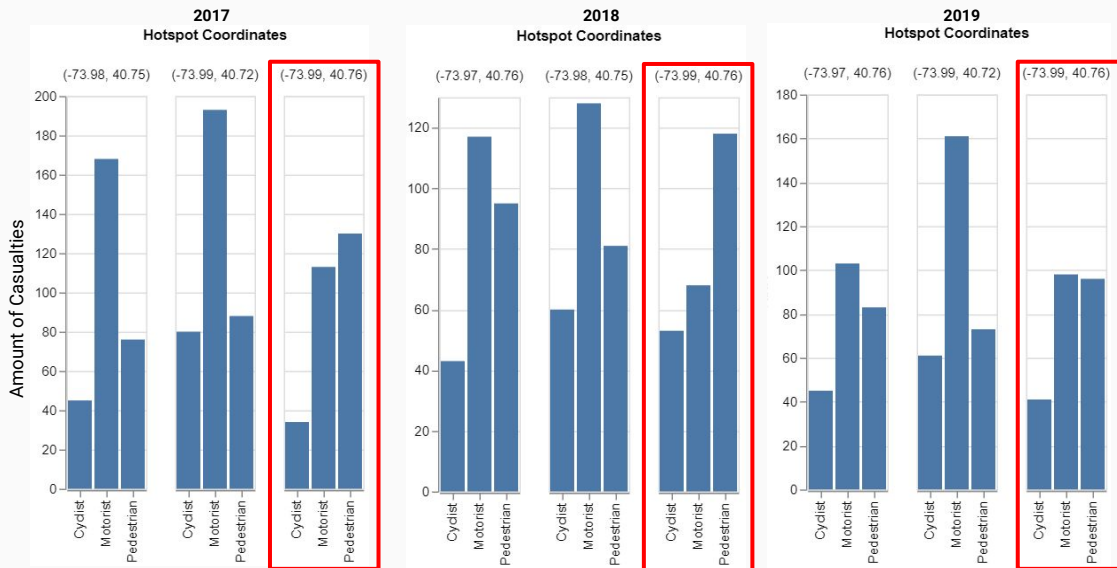
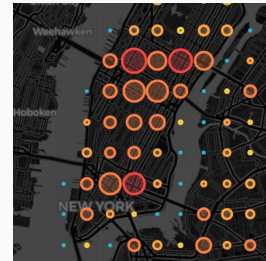
# Question 2: How do casualties compare between type of commuter?

## Total casualties between pedestrians, cyclists, and motorists

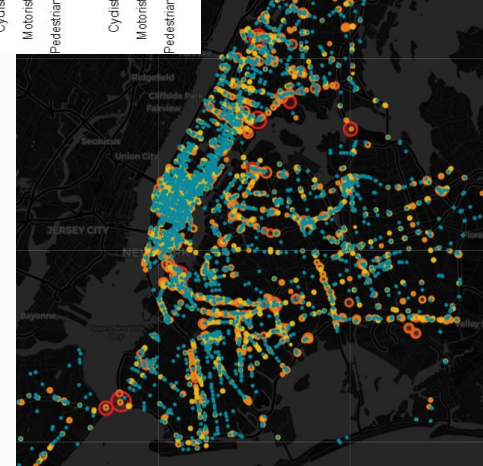
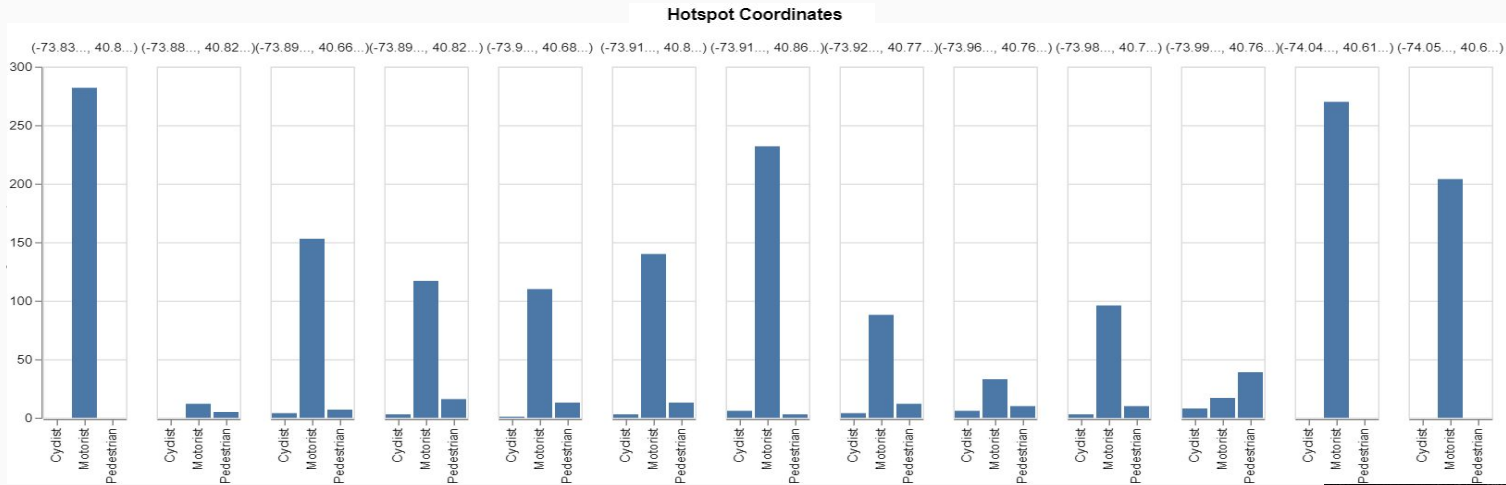
For all locations for all years



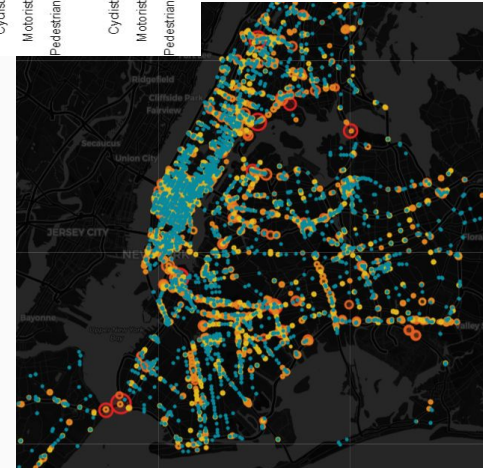
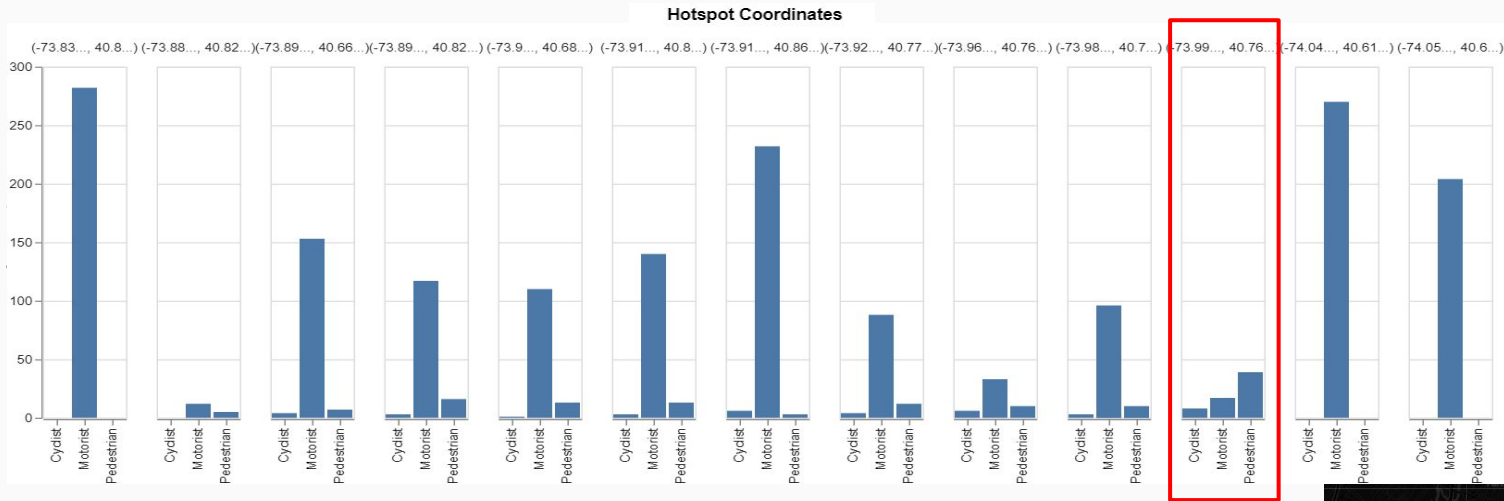
Hotspot Locations for each Year



# Question 2: How do casualties compare between type of commuter?



## Question 2: How do casualties compare between type of commuter?

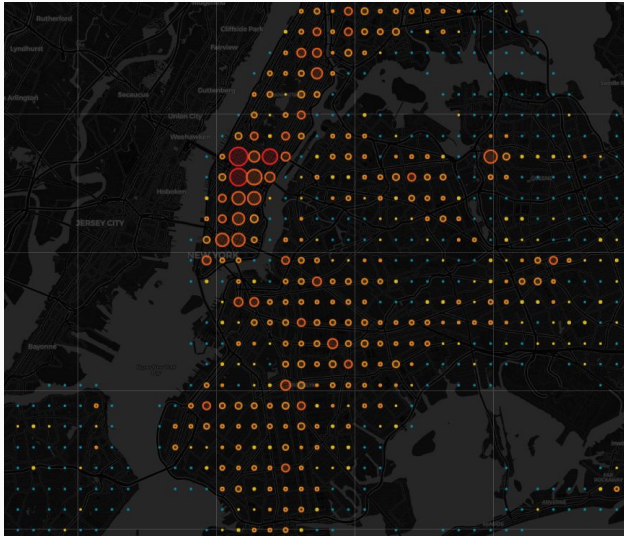




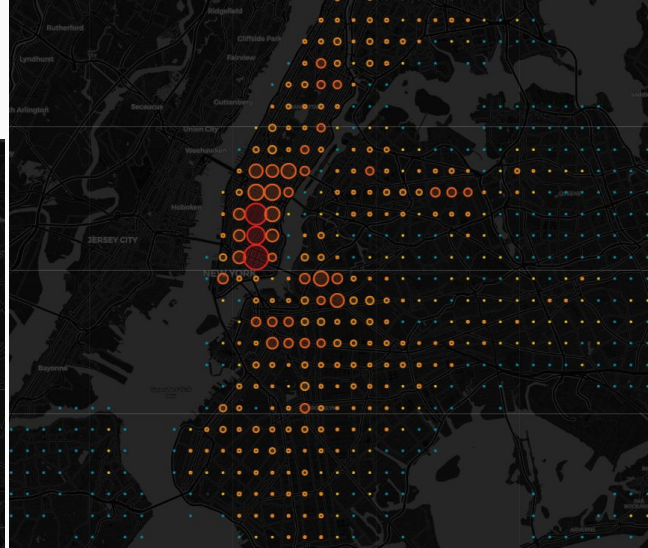
# Question 2: How do casualties compare between type of commuter?

## Casualties by Victim Category (for all years)

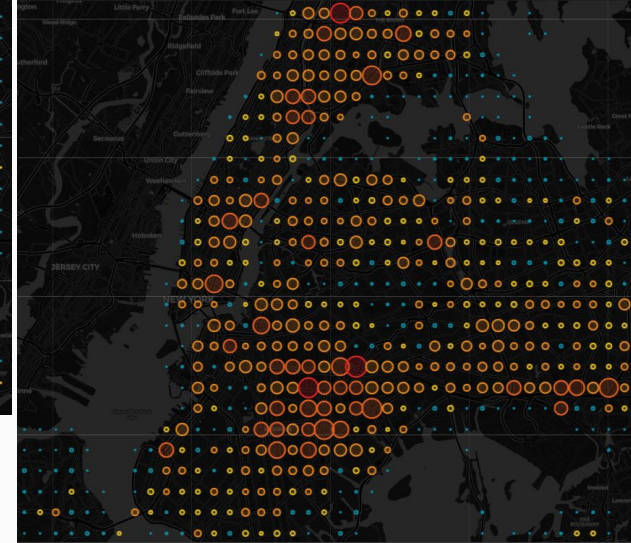
Pedestrians



Cyclists



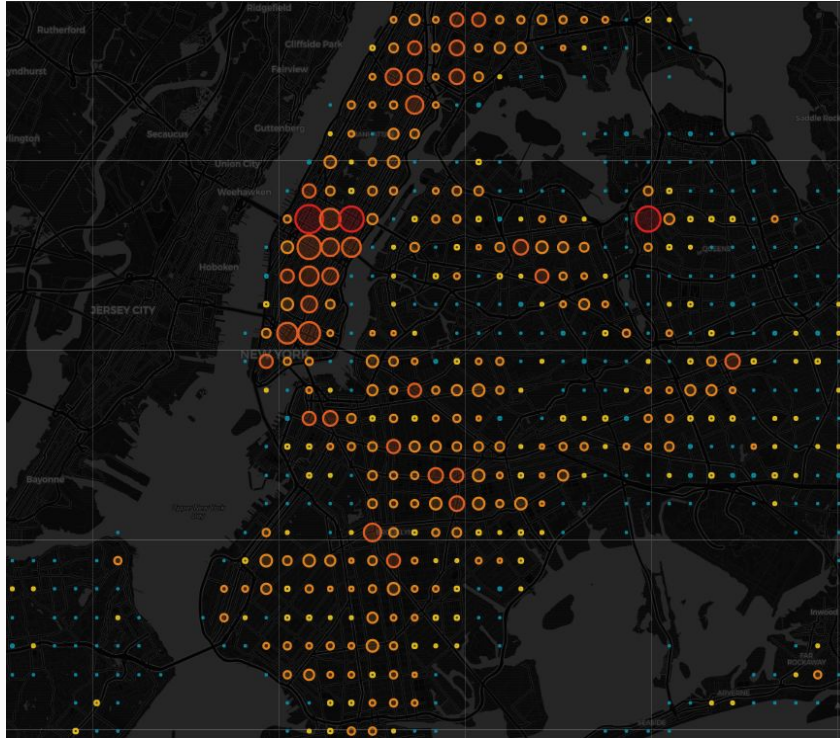
Motorists



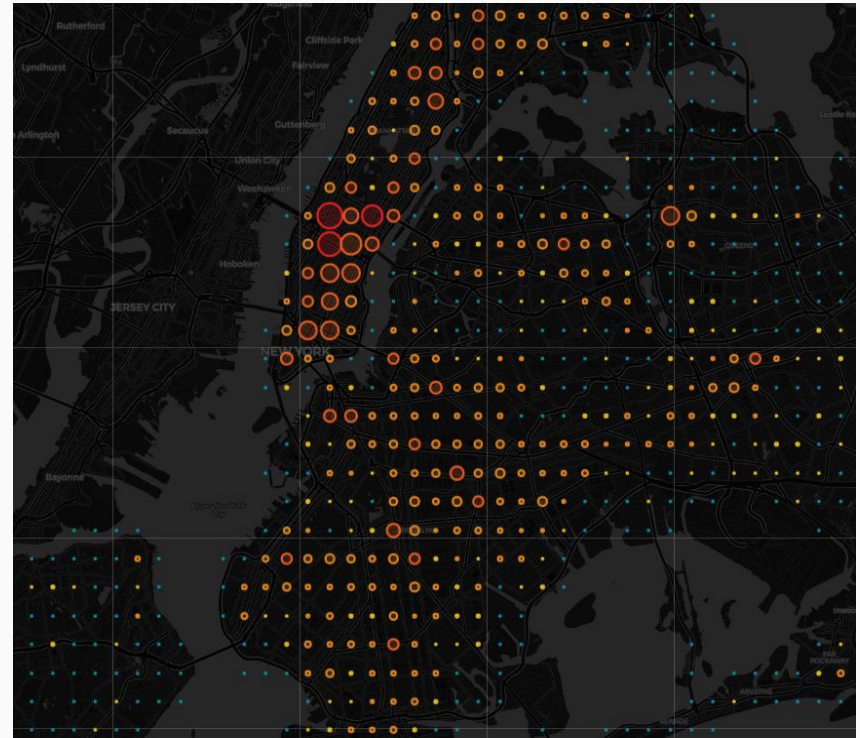
# Question 2: How do casualties compare between type of commuter?

## Casualties by Commuter Category: Pedestrians

2019



All Years

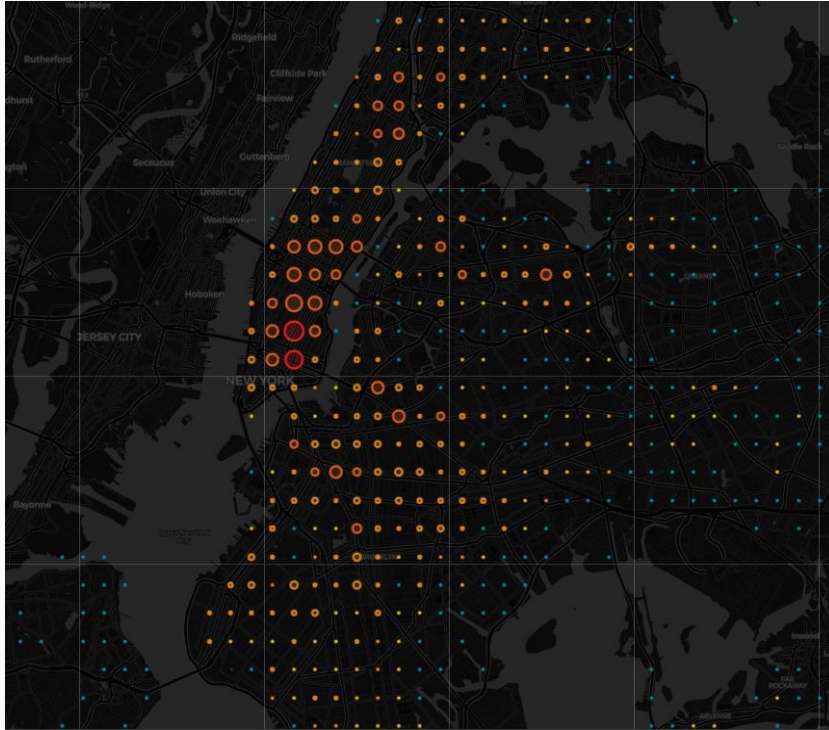




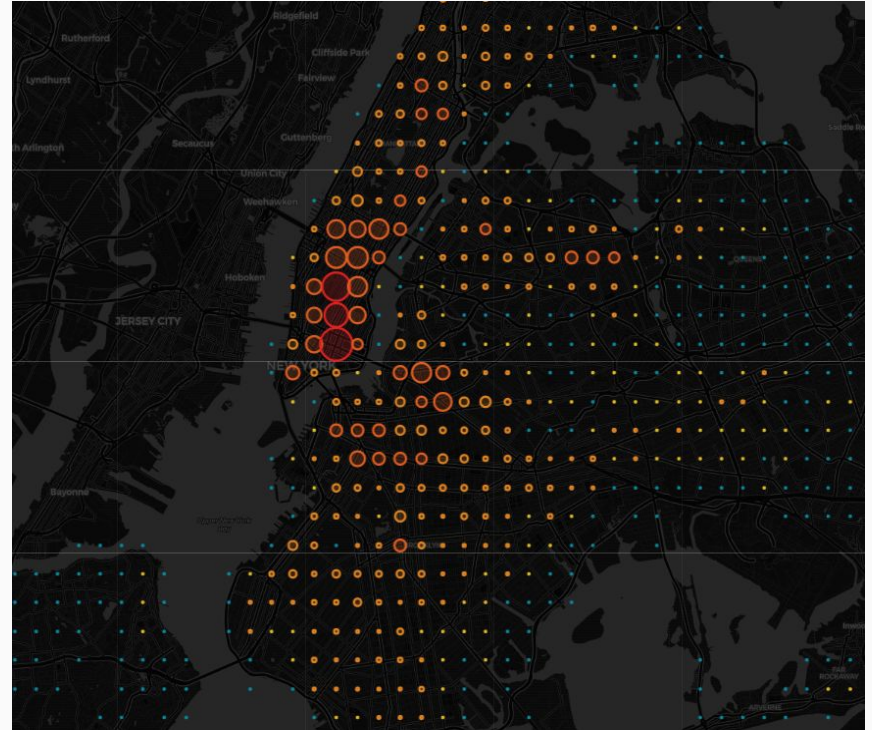
# Question 2: How do casualties compare between type of commuter?

## Casualties by Commuter Category: Cyclists

2019



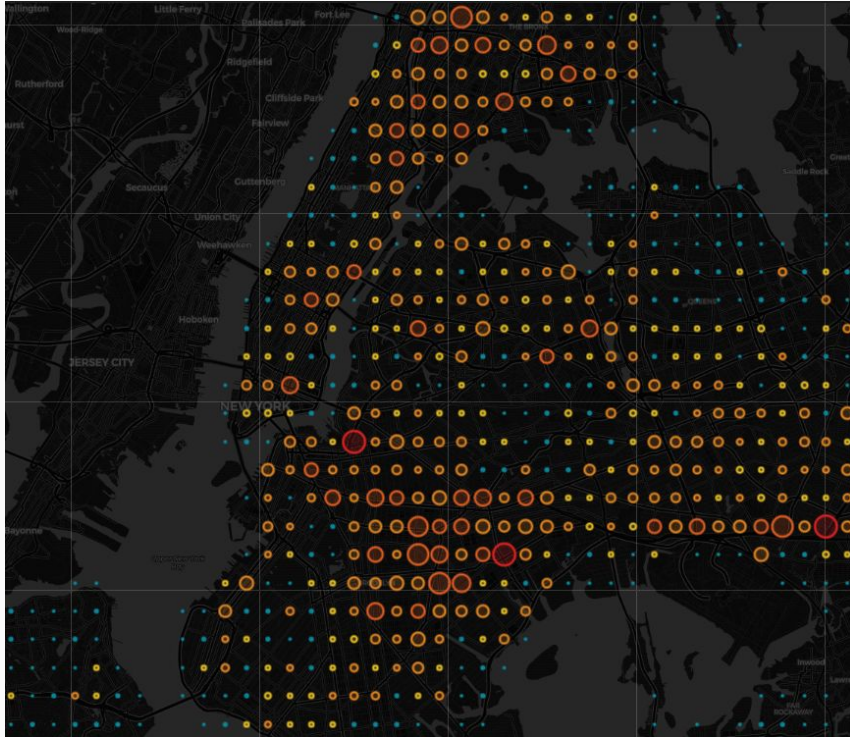
All Years



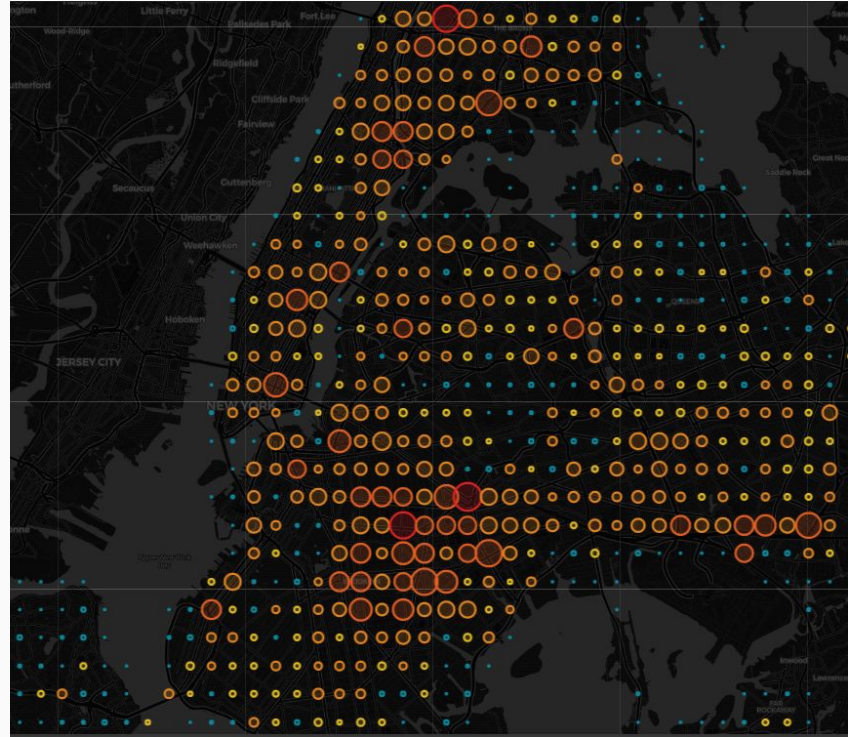
# Question 2: How do casualties compare between type of commuter?

## Casualties by Commuter Category: Motorists

2019



All Years



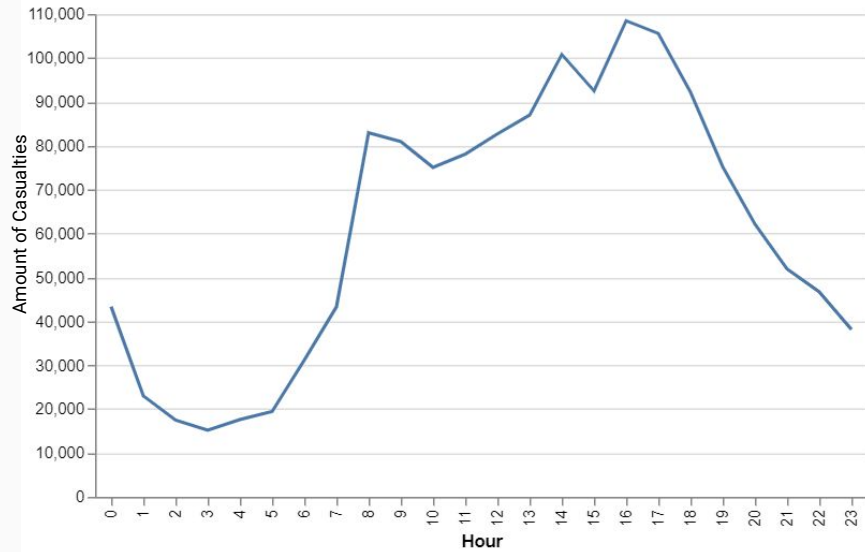
# Question 3:

At what times are collisions more likely to occur?

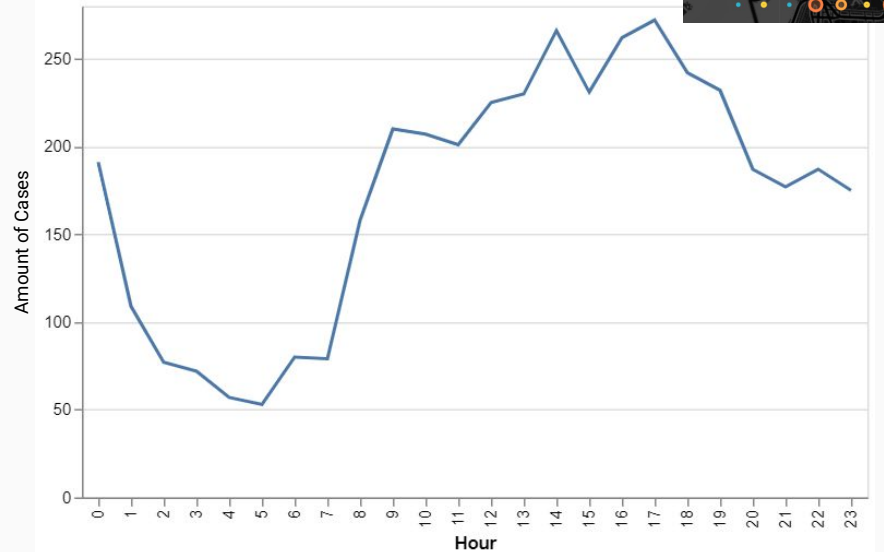
# Question 3: At what times are collisions more likely to occur?

## 2019 Collisions by Hour:

For all Locations



Across all Hotspots



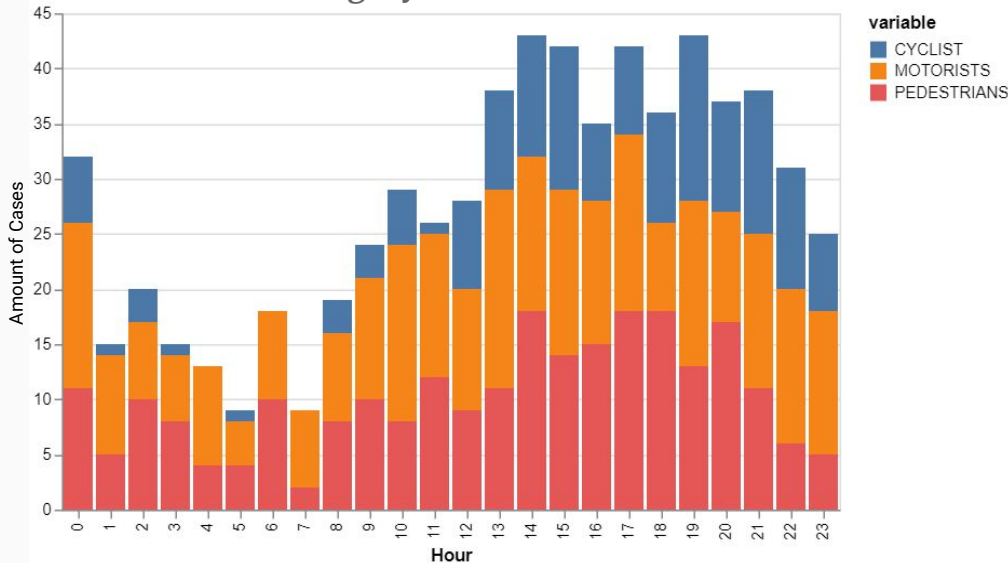


# Question 3: At what times are collisions more likely to occur?

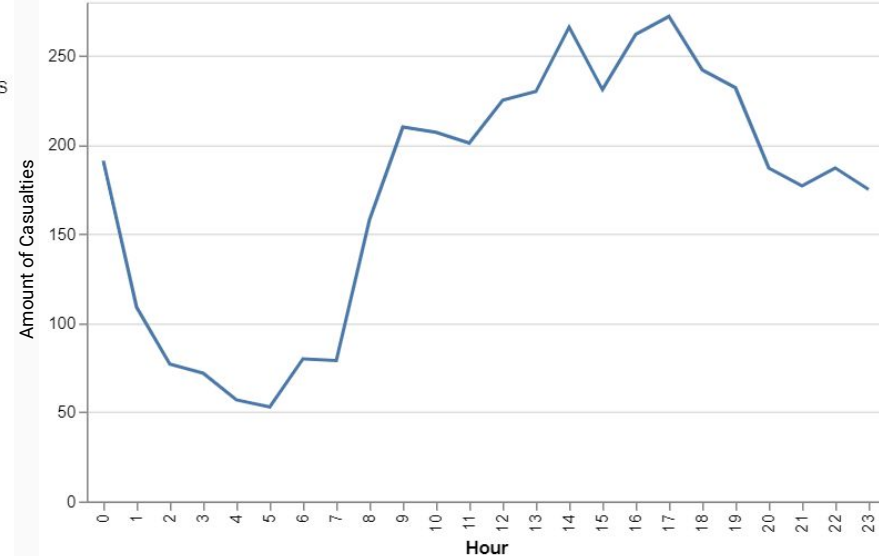
Hourly collisions for all hotspots:



Category Distributions

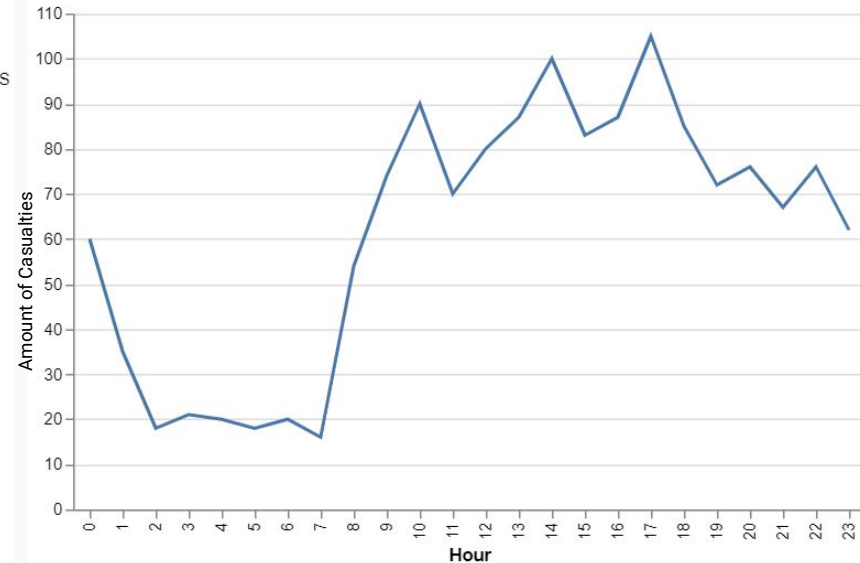
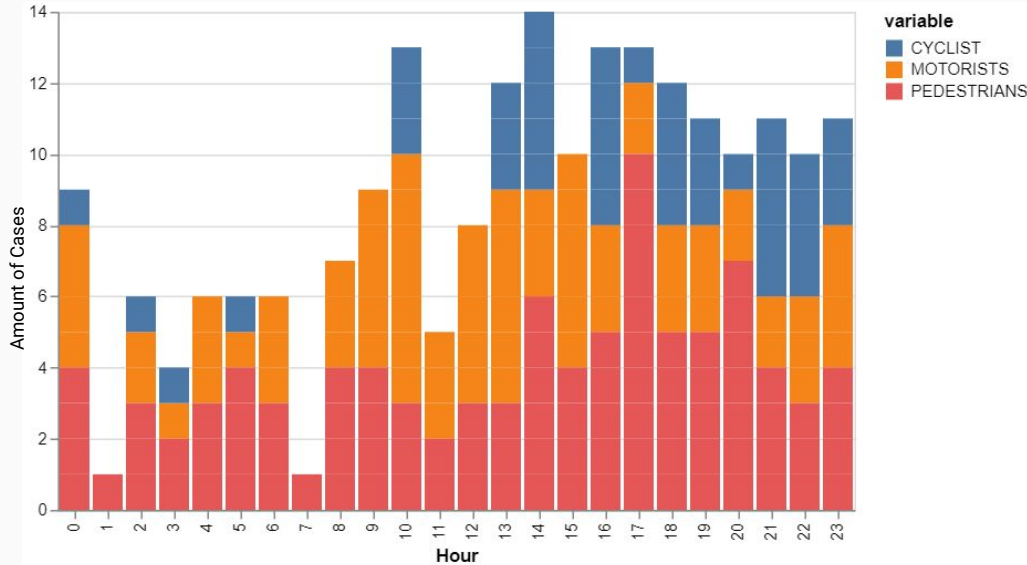
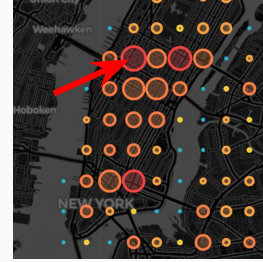


Aggregate Trend



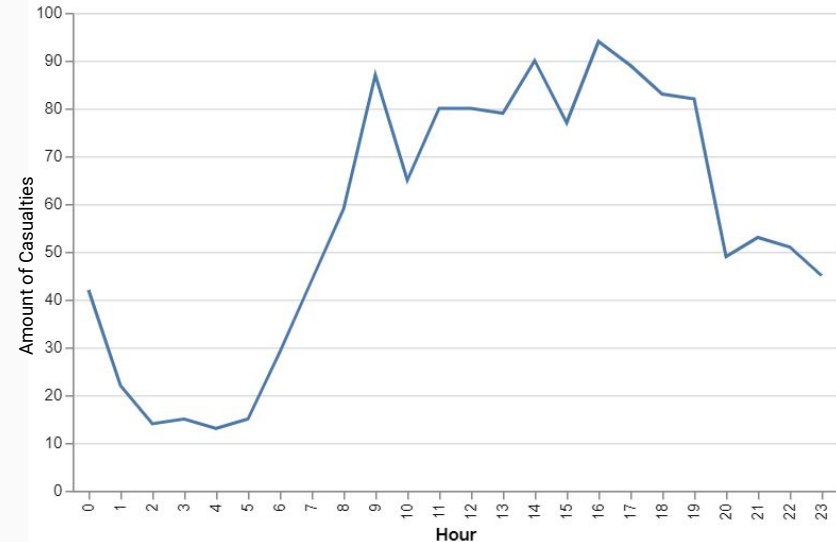
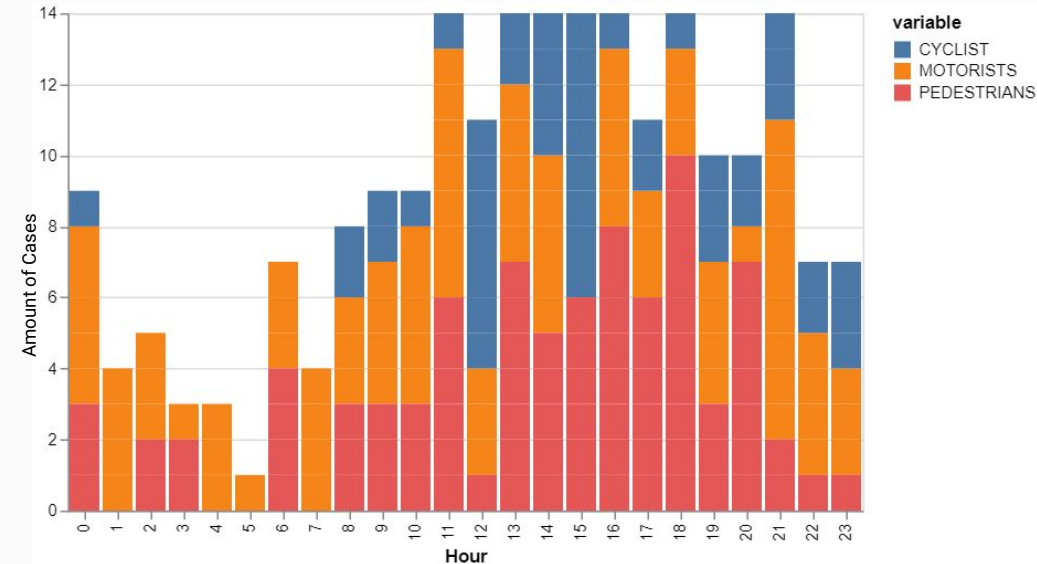
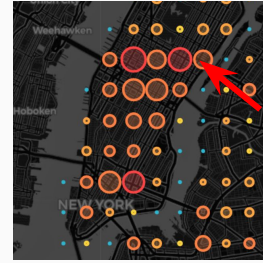
# Question 3: At what times are collisions more likely to occur?

Hourly collisions for Hotspot 1 (-73.99, 40.76):



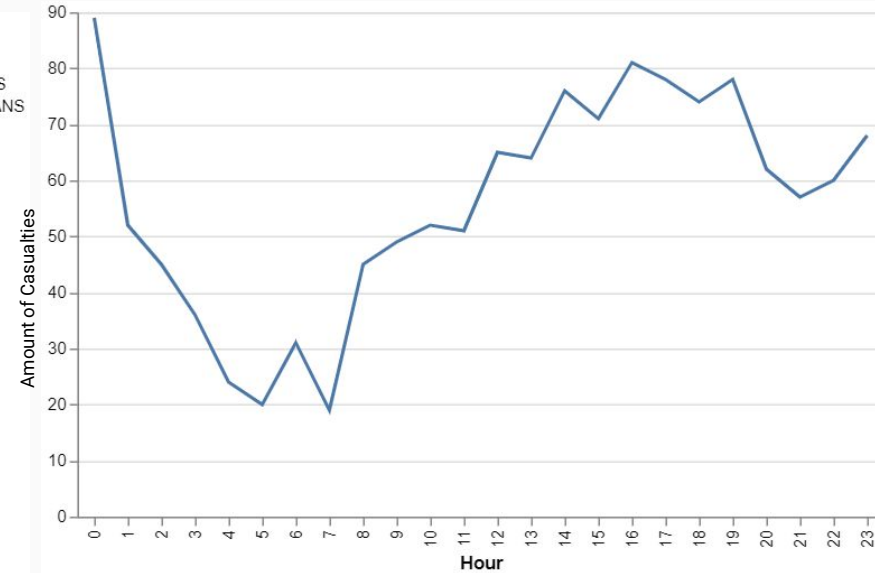
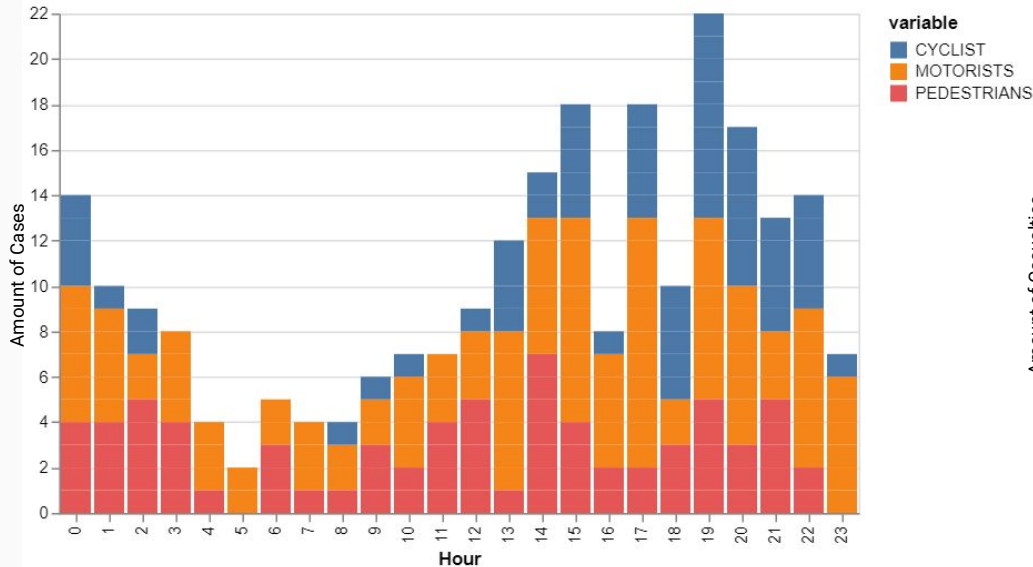
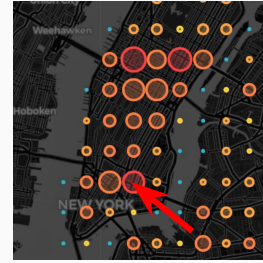
# Question 3: At what times are collisions more likely to occur?

Hourly collisions for Hotspot 2 (-73.97, 40.76):



# Question 3: At what times are collisions more likely to occur?

Hourly collisions for Hotspot 3 (-73.99, 40.72):

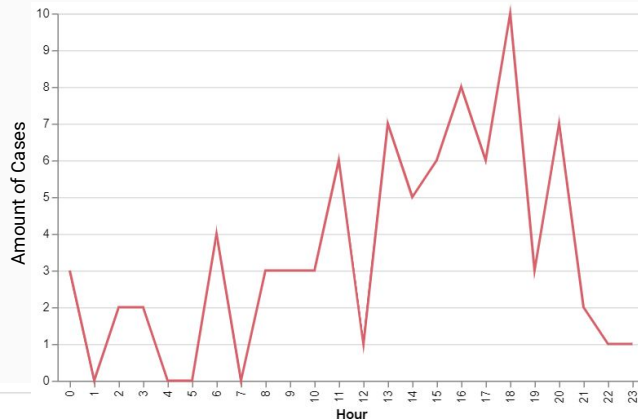
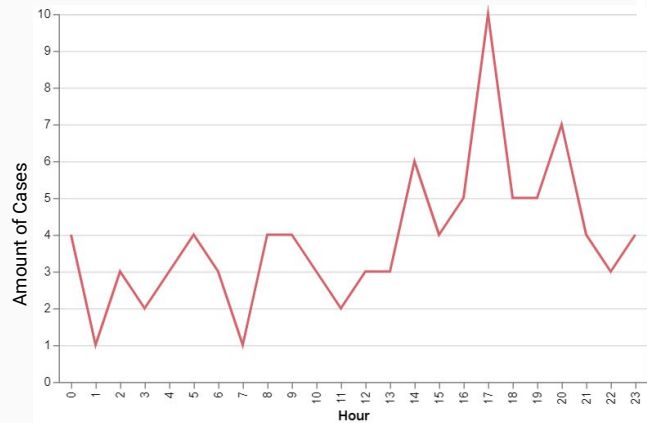




# Question 3: At what times are collisions more likely to occur?

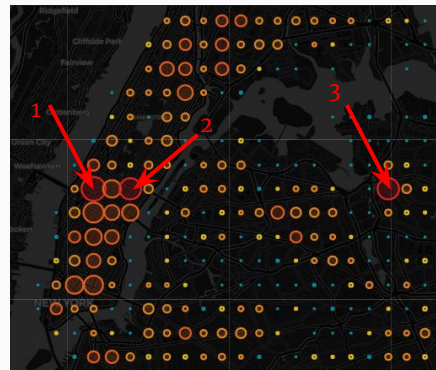
Hourly collisions for **pedestrians** at each hotspot:

Hotspot 1 (-73.99, 40.67)

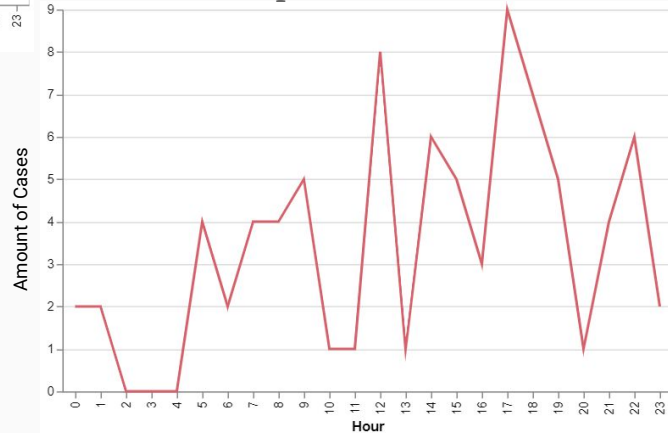


Hotspot 2 (-73.97, 40.67)

Pedestrian Casualties (2019)



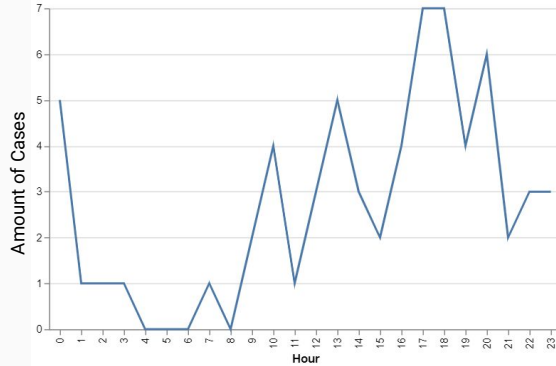
Hotspot 3 (-73.83, 40.67)



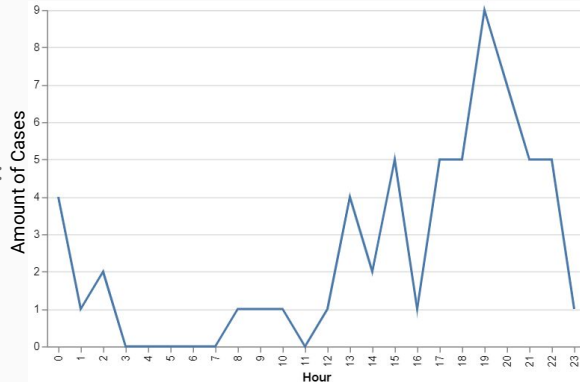
# Question 3: At what times are collisions more likely to occur?

Hourly collisions for **cyclists** at each hotspot:

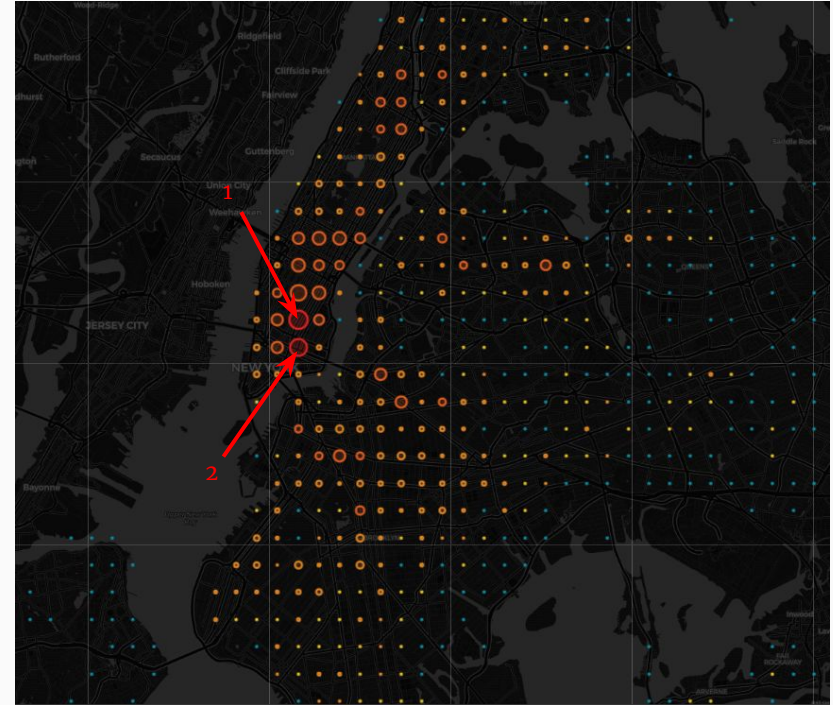
Hotspot 1 (-73.99, 40.73):



Hotspot 2 (-73.99, 40.72):



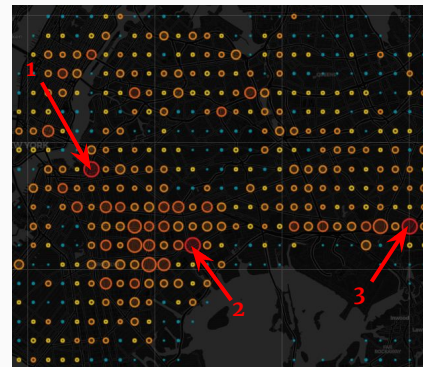
Cyclist Casualties (2019)



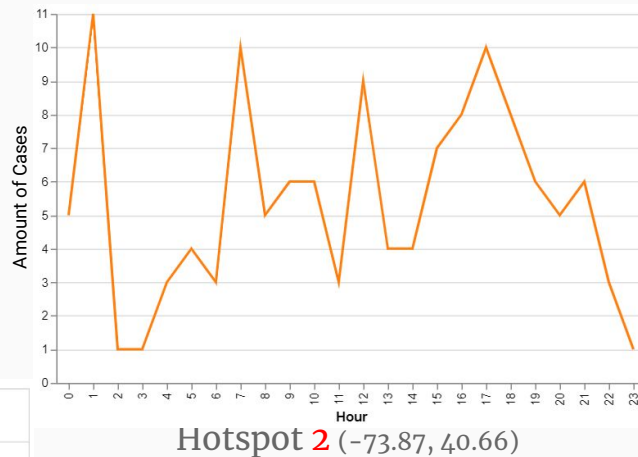
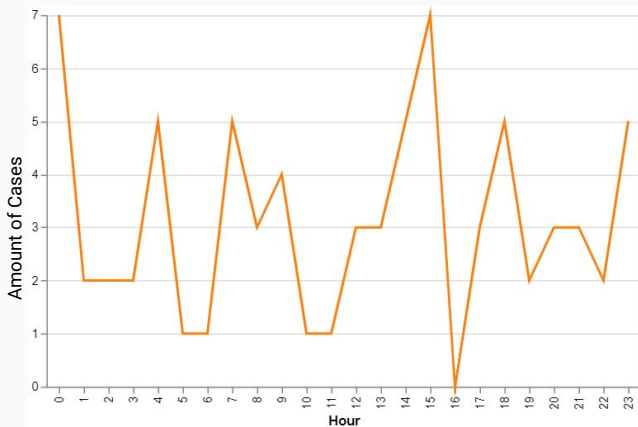
# Question 3: At what times are collisions more likely to occur?

Hourly collisions for **motorists** at each hotspot:

Motorist Casualties (2019)

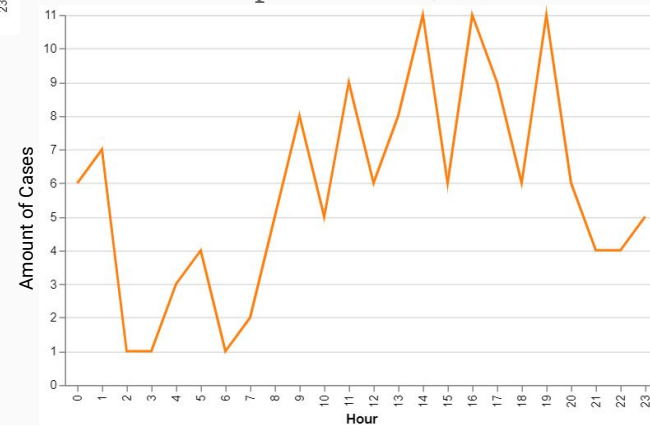


Hotspot 1 (-73.79, 40.67)



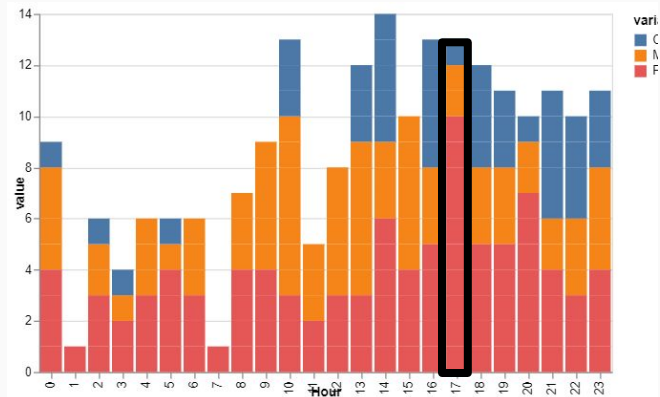
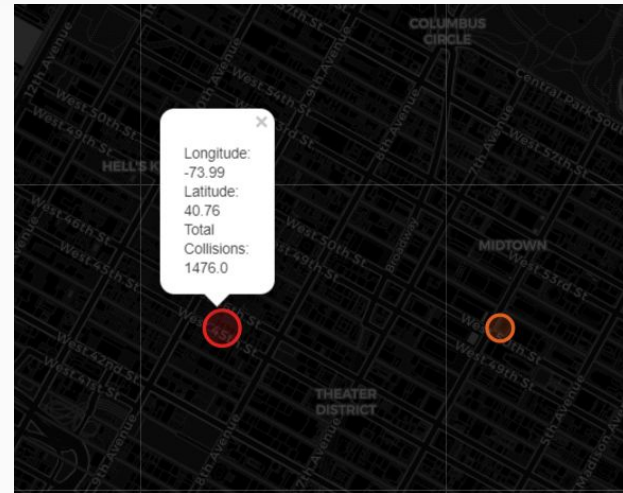
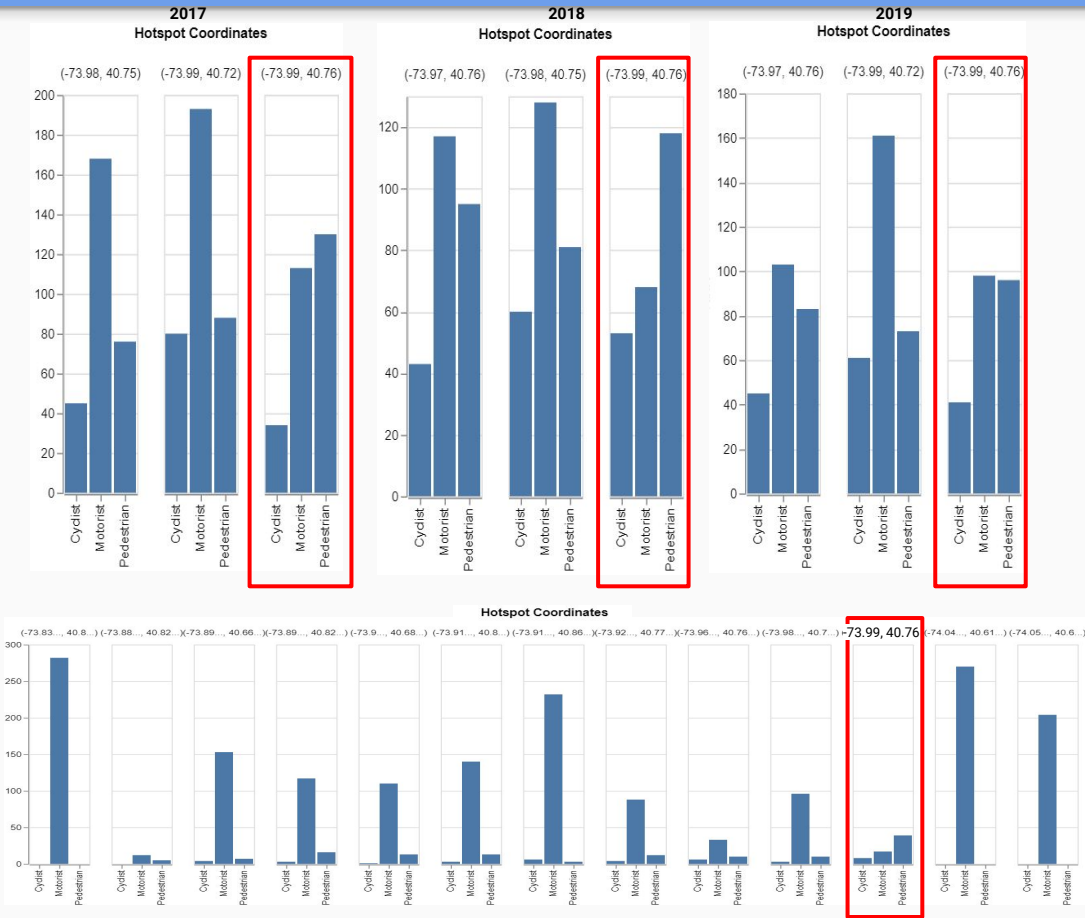
Hotspot 2 (-73.87, 40.66)

Hotspot 3 (-73.96, 40.7)



# Findings and Implications

## Observing unusually many pedestrian collisions at $(-73.99, 40.76)$



# Findings and Implications

- Certain collision hotspot coordinate regions have significant casualty abnormalities for specific commuter categories
  - Pedestrians at pedestrian collision hotspot and overall hotspot 1
    - $(-73.99, 40.76)$  at 5:00 PM EST
  - Cyclists at overall hotspot 2
    - $(-73.97, 40.76)$  at 3:00 PM EST
- **Identified areas to investigate:** Something causes pedestrians and cyclists to be especially vulnerable at those locations / times
  - Future work: Weather, transportation density, crosswalk and lane conditions

# Future Work

- Investigate cause of problems meriting redesign of roads
  - Impact of weather
  - Transportation density
  - Crosswalk and lane conditions (width, etc)
- Additional visualizations
  - Choropleth Pie Chart
  - Interactive Temporal Map