Running Head - Program Evaluation

# Assessing Vision Zero Impact Through Data Science

**Rafael Garcia Cano Da Costa[1], Jirou Xu[1], Lucien Banks Wilson[2]**

**[1]NYU Center for Data Science, [2]KPF Urban Interface**

**New York University**

**December 2016**

# CONTENTS

# ABSTRACT

This study evaluates the Vision Zero program in New York City (NYC) through speed, street redesign, data exploration, collision, and correlation discovery analysis in an effort to offer a feedback to the city that can be used for changing, improving, or creating action plans to decrease the number of deaths and serious injuries on traffic. The program have success so far in an integrated approach, but the main contributor seems to be people's education as a driver and pedestrian. For improving the results, the city should create initiatives addressing high density areas, and enhancing its tools of law enforcement. This work can be used as a guideline for evaluating policies in other cities, and it can be improved by using new datasets, softwares, algorithms, information, and more granular data analysis.

# CHAPTER 1

# INTRODUCTION

## Motivation

Vision Zero was first introduced in Sweden by 1997. After that, different countries through some of their cities adopted the same model for bringing the number of deaths and serious injuries on traffic to zero. Today, states such as Canada, United Kingdom, Netherlands, United States, Australia, and others implemented the program characterizing it as an multinational approach to traffic safety (Wikipedia, 2016).

In NYC, serious injury and deaths caused by the daily traffic receive a special attention from the Mayor's office. Starting in 2014, Mayor Bill De Blasio launched a program called Vision Zero. The undertaking requires the involvement of different agencies and people in the city on an effort to cover all possible contributors when it comes to traffic accidents having people as victims.

The city started a myriad of initiatives to propel the decrease of serious injury and deaths. First, in partnership with the Limousine and Taxi Commission, registered drivers have to go through training every 3 years in order to keep their work concession. The training comprises orientations on yielding when pedestrians are crossing streets, on respecting speed limits and red lights, and so on. Next, streets were redesigned as a way to keep pedestrians safer. New crossing points based on pedestrians needs are being created, and intermediate areas called islands where people can wait while crossing streets in case of a light turning red before reaching the next curb are another reengineering example. Finally, new laws were enacted reducing the citywide speed limit from 30 to 25 miles per hour, while also strengthening sanctions to whom hit someone on the traffic and run, and requiring changing automobile design for minimizing drivers' blind spots. Although other approaches are being taken as a means to keep people away from serious injuries and deaths on the traffic, they are not the focus of this analysis.

Resorting to data analysis techniques, this project seeks to quantify the impact of Vision Zero, and to find out options to potentiate its effectiveness.

## Contributions

This work aims to offer some feedback to the city of New York on the Vision Zero program initiative, and a guideline for conducting data analysis to evaluate cities' policies in order to reassure or improve action plans in place. Once this study is completed, the city will have a good idea if the action plans implemented so far should continue or be improved, and if new ones should be adopted.

# CHAPTER 2

# RELATED WORK

Papers and reports were studied in an effort to better understand the goals and approaches taken via a Vision Zero program, and how this kind of initiative can be evaluated.

On the goals aimed by Vision Zero, it was stated that everyone using the road system carries a piece of responsibilities when it comes to traffic accident (Claes Tingvall and Narelle Haworth, 1999). This calls attention to the fact that an integrated work is fundamental for the success of the referred program. The fact that not only drivers and pedestrians, but also laws and urban engineering are important players guided the mindset throughout this work.

As a means to have a bigger picture on all the moves taken once Vision Zero took off, the reports released by the City of New York on the action plans were studied in order to be aware of each step towards achieving the desired outcome of decreasing the deaths and serious injuries on traffic. One of the measures adopted by the program was the establishment of training before and after taxi and for-hire drivers receive their license, which addresses drivers practices on traffic for raising awareness and minimizing the occurrence of accidents. Also, a joint work involving both private and public sectors is imperative for having an impactful action plan (Vision Zero Action Plan, 2014). Good examples of the level of integration of the program are the workshops reaching out to the people, building a communication channel hearing the citizens concerns and suggestions to improve safety on the streets (Pedestrian Safety Action Plan - Manhattan, 2015). Another instance is enforcement and enactment of new laws like the new city wide limit of twenty five miles per hour that can lower by fifty percent the chance of death when someone is hit by a car at this speed when compared to one at thirty miles per hour (Vision Zero 1 Year Report, 2015). On the industry side, the NYC Taxi and Limousine Commission (TLC) is spreading training on safe driving among professional drivers in the city believing that education is a big partner in the program undertaking (Vision Zero 2, 2016). Not less important, those responsible for urban engineering are contributing as well through street reengineering which deliveries safer roads to the users (Vision Zero 1 Year Report, 2015).

Trying to resort to recent insights, a recent research released that the number of trips among taxis and for-hire vehicles have been stable (FiveThirtyEight, 2015). This information helps this work to be more confident of some conclusions when analysing the available data.

Finally, to find different alternatives for evaluating the data on hand, ArcGis seemed a resourceful software for analysing spatial data. This tool could offer a good way to compare changes on the spatial data before and after an action plan was implemented, which could add value in evaluating the percentual change of data points from one heat map to another (Katharine D. Bennett, 2010).

# CHAPTER 3

# METHODOLOGY, DATA, AND ANALYSIS

Once a policy is implemented in a city, finding out all possible available data on the impacted areas is one important aspect. The City of New York makes resourceful datasets available to the public through the internet (NYC Open Data, 2016).

After having the data on hand, it is necessary to explore it for identifying if the content of a file or database is relevant to the needs faced by the team conducting an analysis. The exploration will also point to features (columns) that should be kept during the data cleaning process. This is important for eliminating unnecessary data for mitigating memory handling issues when manipulating the content on a computer or a cluster of machines.

The plotting of the data as a line, bar, geolocation map, or heat map graph can lead to useful insights on the data. It is a good way to have the data conveying some good information. Generally, trends can be identified when count of occurrences or average of metrics are observed over time. In sequence, the data can be validated through hypothesis testing.

In the end, predictive techniques like machine learning can identify new factors that can be addressed for amplifying the impact of a policy. Many times, this approach will require finding a common factor among disparate datasets for integrating them in order to gauge the interaction and contribution of each feature towards the desired outcome of a policy.

As aforementioned, those steps can be summarized as data collection, data cleaning, data exploration, data engineering, and data analysis.

## Data Source

Different datasets were used, which can be mainly divided into 4 categories. First, Taxi data, made available by the TLC, was used to evaluate if the law passed on October 2014 reducing the citywide speed limit had any impact. In addition to this dataset, Citi Bike data supplemented some findings on the traffic condition of streets. Second, NYPD Crime data and NYPD collision data provided records on the traffic injuries and fatalities along with other occurrences related to street safety before and after the enacted program. Third, Subway Entrance Locations, Bus Stop Locations, and Public Schools Locations were used to integrate the collision dataset as a means to identify features contributing to cases of collision. Lastly, the remaining datasets offered additional features to the aforementioned intents. Most of the supporting datasets are available on NYC Open Data. The detailed information of each of the datasets are described below:

Fig. 1

- Green Cab: ~53436857 records, 21 columns, ~7.92 GB
- FHV Cab: 178469955 records, 18 columns, ~5.49 GB
- Yellow Cab: ~761273804 records, 19 columns, ~119.41 GB
- Citi Bike Data: ~28548005 records, 15 columns, 921.82 MB
- NYPD Crime Data : 1123465 records, 20 columns, 194.1 MB

- NYPD Collision Data : 892927 records, 29 columns, 171.3 MB
- Subway Entrance Locations: 1905 records, 4 columns, 234 KB
- Parking Violation Issued: 39454354 records, 43 columns, ~7.68 GB
- NYC WI-FI Hotspot Locations: 3359 records, 16 columns, 373 MB
- NYC Pluto: 859469 records (split into 5 .csv files, one per borough) 83 columns, 432.9 M
- Bus Stop Locations: 13280 records, 11 columns, 1.2 MB

## Data Exploration

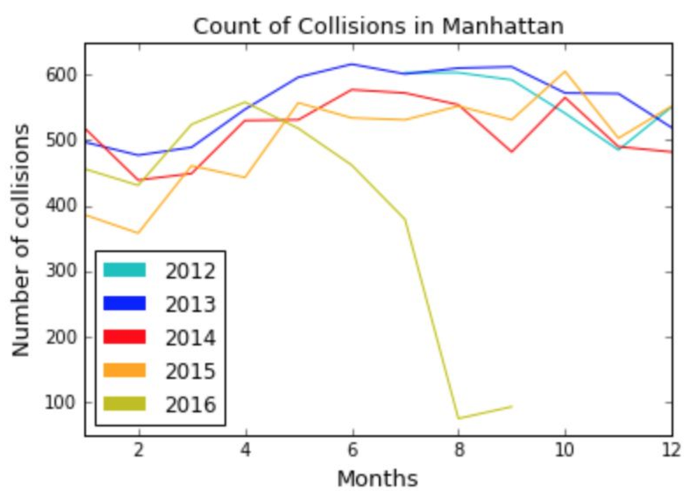
The initial approach was to narrow down the data available on the five boroughs for mitigating the computational cost of working on the whole data at once. First, the choice was to work on the data available on Manhattan. Once this was done, the idea was to work on a sampled data from Manhattan for deriving some statistics and conclusions that could be generalized to the whole island. However, this approach was not offering the expected efficiency. Summarizing the initial approach, Manhattan was divided into thirty seven polygons using some current main streets as delimiters. After this, a definition of a block was written down, and an algorithm on how to identify blocks were designed. Having these established, each polygon had a number assigned to its blocks within, and the *random* python library picked one block from each polygon. After having random blocks from each polygon, polygon queries through Hadoop would collect data based on pick up/drop off latitude and longitude. This approach was not being successful because the length of a street next to a selected block was taking too long to be defined since the goal was to have the shortest length possible, but being able to have the polygon query retrieving a resourceful amount of data. As a result, another reasoning was applied, and the whole data on Manhattan was used.

After filtering the data on Manhattan using polygon query through Hadoop, the dataset was again partitioned by year. Also, the data from 2014 and 2015 had their month of September filtered as well. These steps were implemented on the Taxi, Citi Bike, NYPD Collision, and NYPD Crime datasets.

## Collision Analysis

The data on collisions was plotted using Matplotlib and Ipython Notebook after preprocessing on Hadoop that counted the number of collisions per month between 2011 and 2016. The expectation was to verify a decrease on the occurrences of collisions just after 2014. However, the graph showed an interesting change on the trend just on June 2016. This fact can be observed on the picture below:

Fig. 2




This fact was a surprise and gave the idea that a deeper research and study would be needed to understand the impact of Vision Zero.

Enhancing the previous picture contribution, it was thought that a heat map showing the density of occurrences of collisions in Manhattan could lead to new ways of mitigating the number of future cases through initiatives focusing on the higher density areas. This plot displays areas with lower/higher density of collision cases. The darker the area, the higher the density of cases.

Fig. 3

**The picture on the left depicts 2014 while the other one 2015**

Both pictures display the highest concentration in midtown, and on specific intersections in lower Manhattan and in Harlem. This kind of map can point to areas where new actions can be introduced by the city to lower even more the number of collisions on the traffic.

## Speed Analysis

The data on taxis and for hire vehicle was used for evaluating the trend of speed on traffic after the new law reducing the citywide limit to twenty miles per hour. The expectation was to see a decreasing trend after October 2014. The data was plotted using Seaborn and Ipython Notebook. The distribution of the data was plotted on September 2014 and 2015, and the speeds over seventy miles per hour were not considered for computing some statistics and conducting a hypothesis testing.

The distributions and statistics on yellow and green cabs are as follow:

Fig. 4

**Graphs on 09-2014**

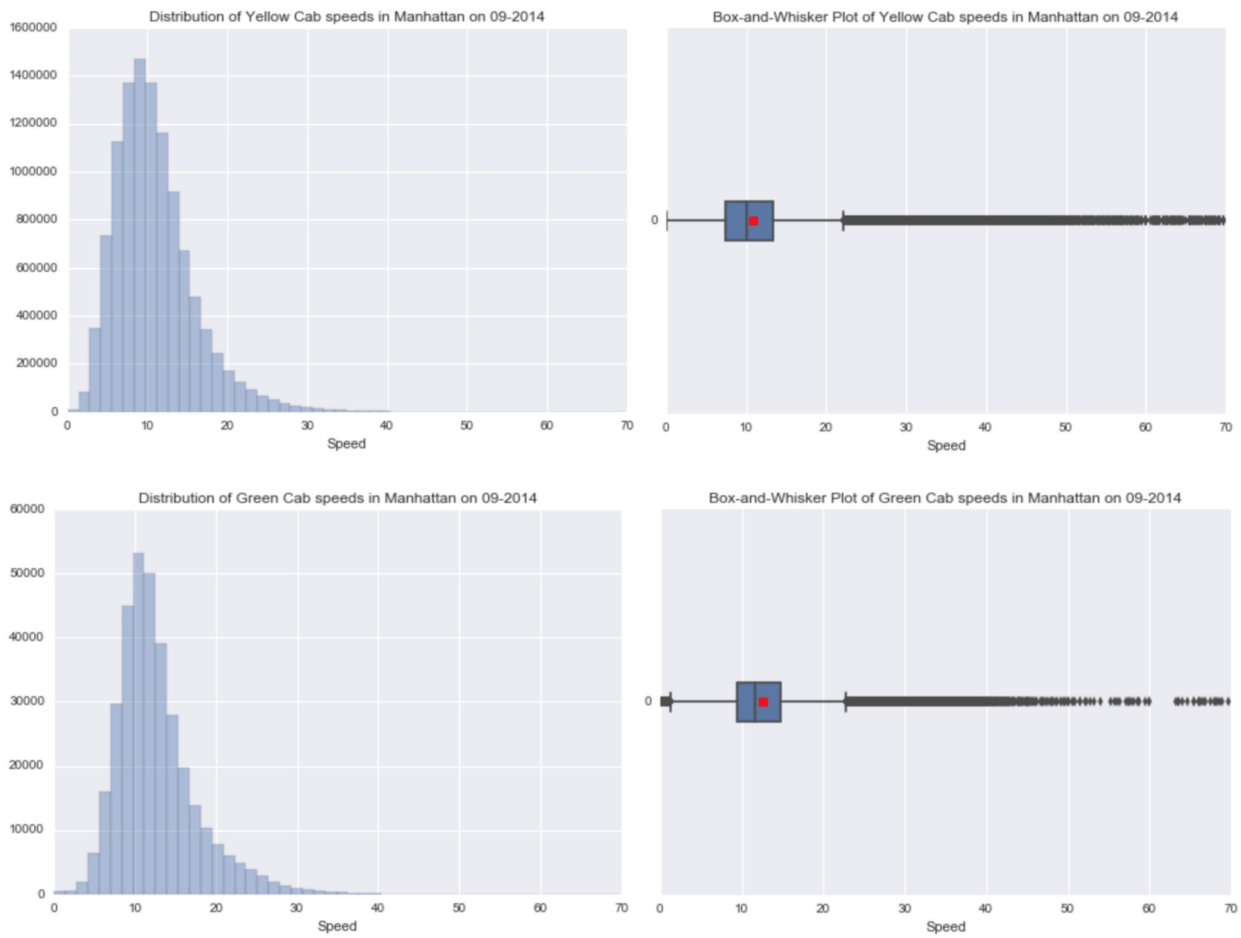



Fig. 5

**Descriptive Statistics on 09-2014**

| | Measure of Central Tendency | | | | Measure of Spread | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Resistant | | Nonresistant | | Resistant | Nonresistant | | | | |
| | Median | Midhinge | Mean | Midrange | IQR | Range | SD | Coef V. | Skew | Skewed? |
| yellow | 10.10 | 10.4 | 10.90 | 34.82 | 5.9 | 69.65 | 4.99 | 0.457 | 1.30 | yes |
| green | 11.68 | 12.05 | 12.63 | 34.83 | 5.35 | 69.67 | 5.00 | 0.39 | 1.41 | yes |

Fig. 6

**Graphs on 09-2015**

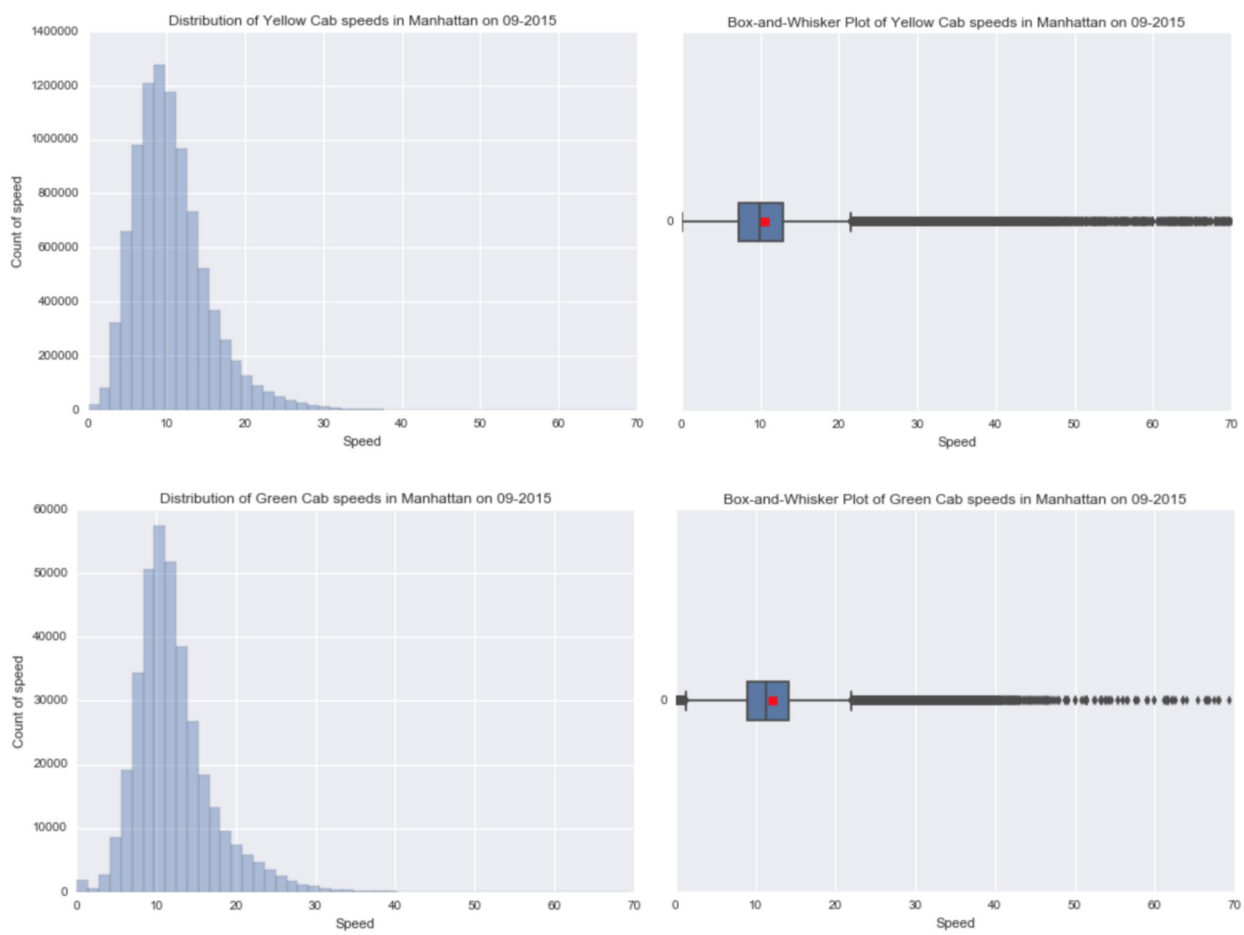


Fig. 7

**Descriptive Statistics on 09-2015**

| | Measure of Central Tendency | | | | Measure of Spread | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Resistant | | Nonresistant | | Resistant | Nonresistant | | | | |
| | Median | Midhinge | Mean | Midrange | IQR | Range | SD | Coef V. | Skew | Skewed? |
| yellow | 9.84 | 10.1 | 10.57 | 34.94 | 5.66 | 69.89 | 4.78 | 0.452 | 1.21 | yes |
| green | 11.26 | 11.61 | 12.14 | 34.69 | 5.16 | 69.39 | 4.91 | 0.40 | 1.37 | yes |

The descriptive statistics give some hint that there was a shift of the data points to the left of the x-axis. For instance, the median is a good resistant measure to observe this fact. It has dropped from 10.10 to 9.84 on yellow cabs, and from 11.68 to 11.26 on green cabs. Also, the IQR (Interquartile Range) shows that fifty percent of the data is more concentrated since the distance between its extreme values decreased from 5.9 to 5.66 on yellow cabs, and from 5.35 to 5.16 on green cabs. Another good indicator is the skewness that went from 1.30 to 1.21 on yellow cabs, and from 1.41 to 1.37 on green cabs, meaning that the data is less spread to the right tail. Therefore, it is expected that a hypothesis testing the difference of means among the data points from 09-2014 and 09-2015 will indicate a drop on the average speed of taxi trips in Manhattan after the change on the citywide speed limit.

The chosen approach was a paired *t* test for the mean difference in related populations. The details are as follows:

Fig. 8

**Null hypothesis:**

Test 1: $H_o$: $(\mu_{09\text{-}2014} - \mu_{09\text{-}2015}) = 0$

**Alternative hypothesis:**

Test 1: $H_A$: $(\mu_{09\text{-}2014} - \mu_{09\text{-}2015}) > 0$

**Level of Significance:**

5% = 0.05

**Sample size:**

$\text{Yellow}_{\text{cab data}}$: 9205643

$\text{Green}_{\text{cab data}}$: 346120

**Statistical technique & corresponding test statistic to use:**

Paired *t* test for the mean difference in related populations.

**Rules for rejecting or not rejecting the null hypothesis:**

Test 1: Reject the null hypothesis if the $t_{\text{statistic}}$ > upper critical value, otherwise do not reject.

Fig. 9

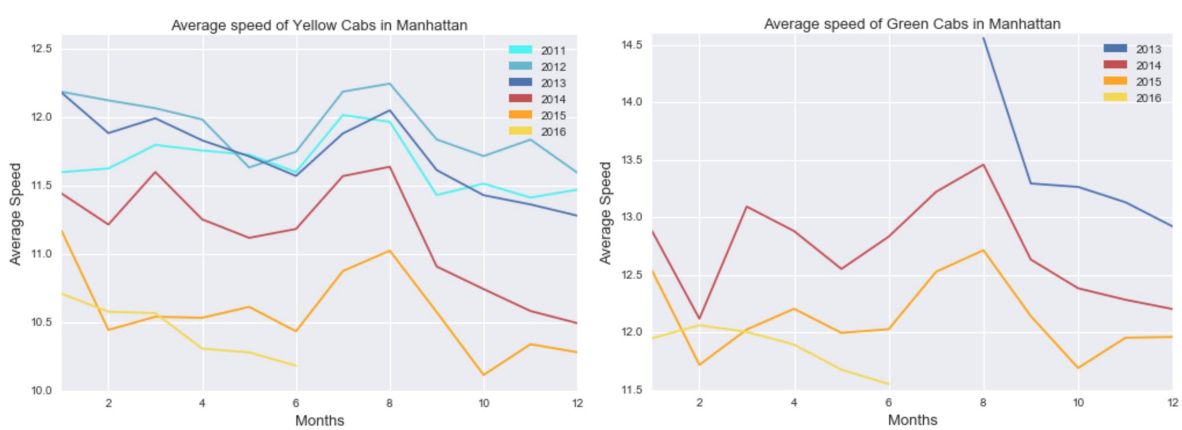
**Table containing involved values**

| | Critical Value | $t_{\text{statistic}}$ |
|---|---|---|
| Test 1 - yellow | 1.64 | 145.75 |
| Test 1 - green | 1.64 | 41.05 |

The conclusion was that the null hypothesis should be rejected since both $t_{statistic}$ are bigger than 1.64. Therefore, it can be said with level of significance of five percent that the average speed from 09-2014 and 09-2015 are not the same, and in fact there was a probable decrease on the average speed in 09-2015. This result provides an indicator that the new citywide limit is having an impact. Moreover, the ratio of average speeds above twenty five miles per hour and those below or equal was 0.017 on September, 2014, and 0.013 on September, 2015 for yellow cabs. Respectively, 0.028 and 0.023 for green cabs, which shows that the number of average speeds above the new limit retracted.

In addition, a graph of the average speed trend over months per year was plotted to verify if the trend of decrease on the average speed is actually happening.

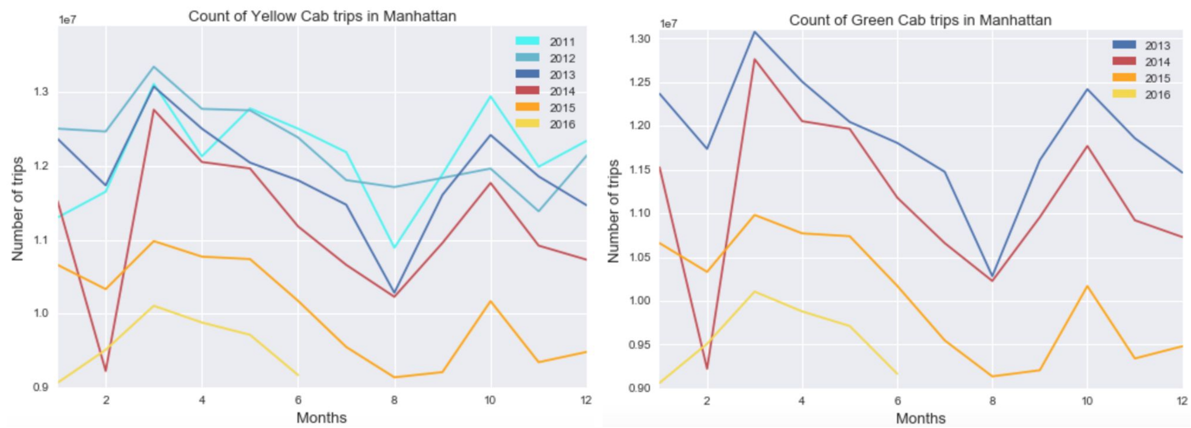
Fig. 10




It is possible the verification that over the years the average speed is in fact dropping after 2014. Moreover, 2016 presents the lowest trend.

Although the previous testing and visualization are helpful to derive some insight, it does not seem enough. The fact that the city is growing can increase the number of cars on the streets, and this could lead to a drop on the average speed. In order to clarify this point, a graph plotting the count of trips in Manhattan was developed.

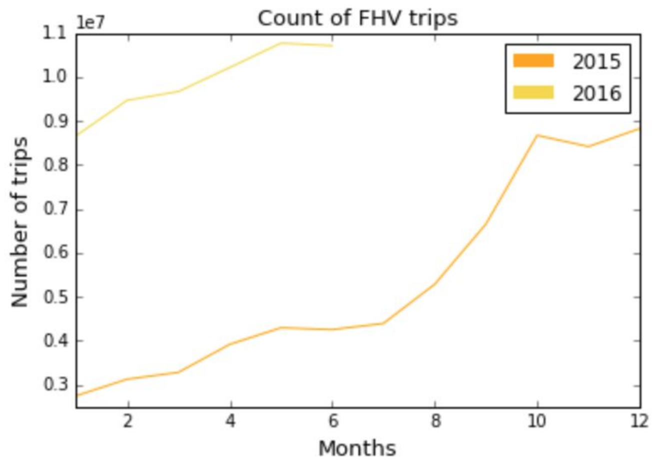
Fig. 11




One interesting fact is the deep valley on February, 2014. Based on research, it was found that February 2014 had a severe winter with deadly snowing storms (New York Post, 2014) which probably one of the aftermaths was the atypical decline on the number of trips in Manhattan. Also, the regular annual trend of decline around August seems related to the school vacation period in New York City, which ranges between June and September (schools.nyc.gov, 2016).

The above graph shows that the number of trips in Manhattan is decreasing for yellow and green cabs. This raises some questions because the decrease on the number of trips is not very compatible with the drop on the average speed. As a way to have some answers, a graph was plotted using the for hire dataset which includes trips from Lyft and Uber among others.

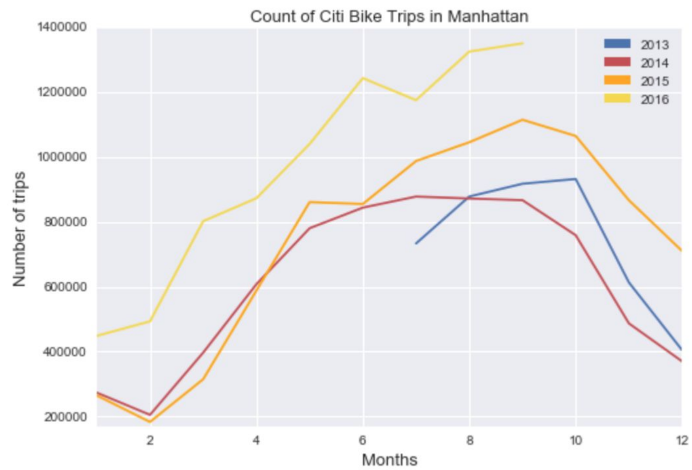
Fig. 12




The FHV(For Hire Vehicle) graph indicates an increase on the number of trips among the FHV category. Even though there was no data prior to 2015, the increase was significant from 3,000,000 to 11,000,000 trips. This can explain the drop on the number of trips on yellow and green cabs. As a result, there must be happening a shift of customers to the FHV service. Also, to corroborate this finding, a research done by FiveThirtyEight concluded on October 2015 that the number of regular cabs and for-hire trips is pretty much stable over previous years. Consequently, it seems reasonable to conclude that the average speed is really decreasing over the past years. However, to have more certainty, it is useful to know the number of licensed cars among cabs and for-hire vehicles over years as a means to evaluate if there is an increase on the number even though the total amount of pickups has been being stable.

Furthermore, the number of people using the Citi Bike service is also increasing, which can contribute to the decrease on the number of cab trips. This fact can be observed on the graph below:

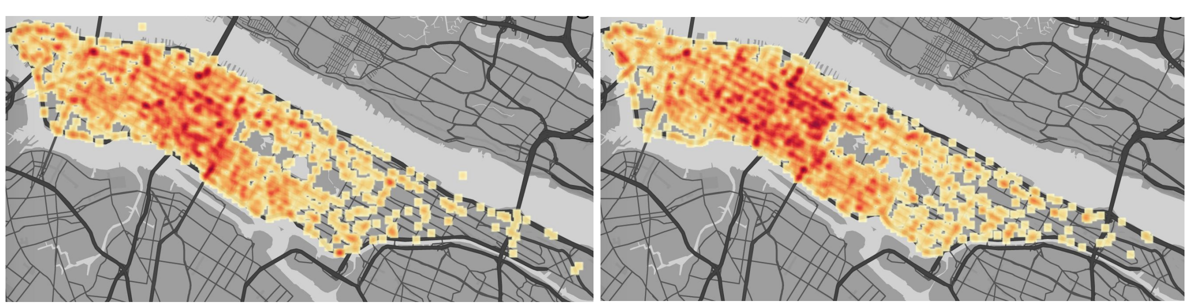
Fig. 13




Envisioning future implementations for improving the Vision Zero Program, a heat map graph was created to display occurrences of overspeed(above 25mph) among yellow cabs on trips shorter than or equal to .1 of the mile. This plot displays areas with lower/higher density of overspeeding cases. The darker the area, the higher the density of cases.

Fig. 14

**The picture on the left depicts 2014 while the other one 2015**



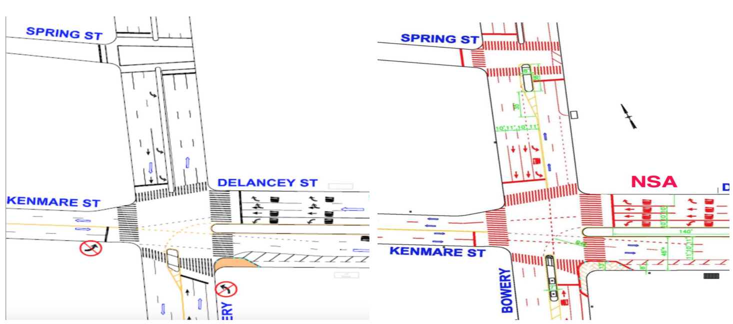
In both graphs, it is possible to see that the higher density of overspeed occurs in midtown. This information could be exploited by choosing areas where devices can be installed to identify drivers incurring in a wrong behaviour which could lead to a later punishment through fees for example. Another aspect is the possibility to allocate NYPD personal to the higher density areas to enforce the law. Also, it raises questions on the possible relationship between the higher number of collisions and overspeed in Midtown, because Midtown is an area with high density of people walking around. Among the people, tourists account for a big group that can be vulnerable to traffic accidents for being distracted while crossing roads. A further study on this subject can lead to ways of keeping tourists away from accidents on traffic.

## Street Redesign Analysis

The trend on the number of collisions before and after a specific street is redesigned was picked as a way to evaluate if the Vision Zero Program is being effective on its intent. Among different implemented redesign, the number of collisions per year was plotted, and the average was computed in order to measure the impact of this action plan on the program.

The first redesign to be analysed is the one that was completed on February 2014 on the intersections comprising Bowery, Kenmare, Delancey, and Spring St. The picture on the left depicts the streets before the changes.

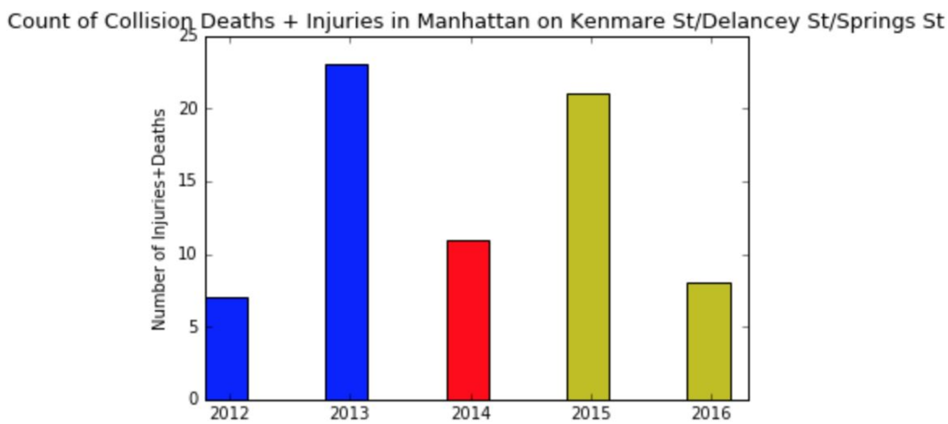
Fig. 15



Some changes are the redesigned central islands on Bowery close to Kenmare St for allowing easier turn and to Spring St for improving pedestrian safety.

The available collision data on the intersections from 07-2012 to 09-2016 was plotted as follows:

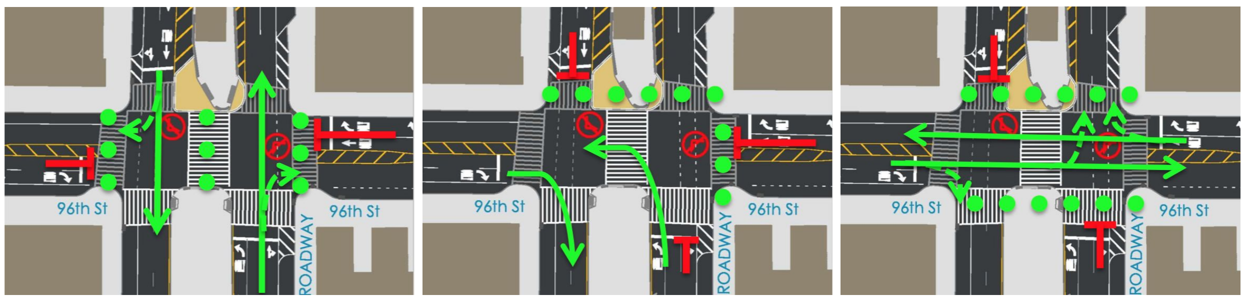
Fig. 16




The focus here is on the years following 2013. In 2014 there was a great decrease on the number of collisions. However, there was an increase on the number of collisions in 2015. It is possible that people were getting used to the new changes, and for this reason they were more cautious when driving or crossing the street in 2014. In sequence, people were already familiar with the redesign in 2015, and the lack of care brought the occurrences to levels observed in 2013. Finally in 2016, there was a huge drop on that number. This was an interesting fact because it matched the trend observed on the first collision line plot. Therefore, something is driving the positive impact of Vision Zero with more power than just the decreasing on the average speed or the street redesign.

The second redesign to be analysed was completed on January 2014 on the intersection comprising Broadway and 96 ST. The pictures depict the streets after the changes.

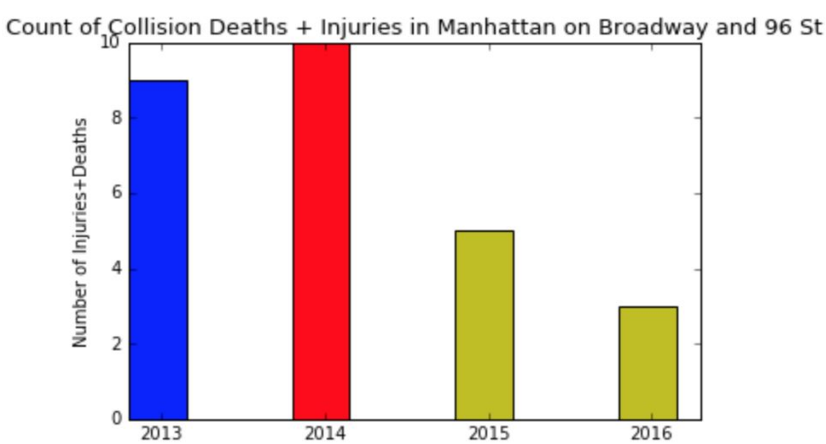
Fig. 17




The main changes are two left turn that were banned, new phasing of signals, expanded pedestrian space, and others .

The available collision data on the intersections from 01-2013 to 09-2016 was plotted as follows:

Fig. 18




Contrasting with the previous redesign, the modifications on Broadway and 96 St seem to have less impact for causing no decrease on the number of deaths and injuries in 2014 in relation to 2013. This could happen because drivers were not aware of the new phasing and not right to left turn. They could speedup before the light turned green for expecting a faster phasing as before, or they could turn left when not supposed to do so. However, there was an improvement in 2015 as opposed to the previous redesign. Maybe the drop in 2015 was due to familiarity with the new redesign. Also, in 2016 the drop was significant, and it worth mention that there was no occurrence on the dataset from March to September in 2016. It is getting more and more clear that the occurrence of traffic accidents is closely related to pedestrians and drivers behavior, and educations seems to have an important role as a result.

To verify what is behind this huge drop on the number of collisions in 2016, a research was done on the annual reports that were released by the New York City for the Vision Zero Program. The reports mention that training on defensive driving and yielding in favor of pedestrians would be made available to professional drivers as a requirement to have their licenses renewed. The renewal happens each three years, and maybe most of the drivers would have to renew their licenses by the end of 2015 or beginning of 2016. The number of drivers receiving this kind of training over the years should be verified as a next step. Moreover, a massive amount of initiatives took place in 2015 related to raising drivers and pedestrians awareness to traffic safety through training. The city reached out to people in schools as a way to educate them on how to behave on streets. The TLC(Taxi and Limousine Commission) awarded and recognized drivers with outstanding driving practices, and issued 283% more summonses for overspeeding in 2015 than in 2014 (Vision Zero 2, 2016). Additionally, Another possible explanation to this surprising change in 2016 could be that there are plenty of street redesign updates and adjustments done at the beginning of 2016. Besides, compared to the the street redesign projects in 2016, the number of street redesign plans in 2014 is relatively small.

Summing up, the Vision Zero Program is comprised by different action plans that work in an integrated manner. Somehow, each action plan is making a contribution, but it seems that the education of the people through training and awareness adds a substantial improvement to the effectiveness of the program.

## Correlation Discovery

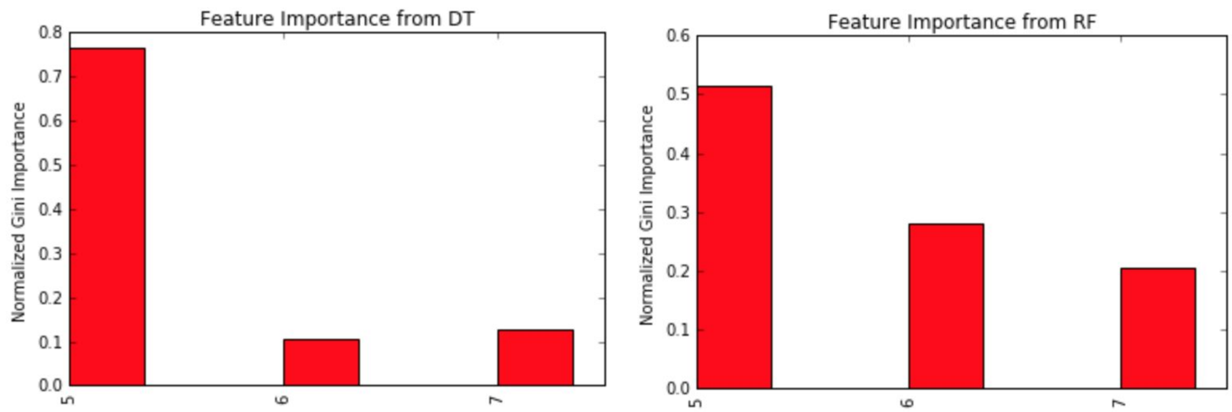
Another goal of this work is to offer ways to Vision Zero be more impactful. This is possible if hidden correlation are brought to light. One way to do it is engineering a dataset based on the collisions dataset. This dataset has latitude and longitude where the occurrence happened, and this can be used to build new features to integrate the dataset. Applying this reasoning, the Bus Stop Locations dataset was integrated into the collision dataset. If a bus stop were placed up to twenty five meters from a collision point, the new feature would present one or zero otherwise. This same procedure was implemented for the Subway Entrance Locations dataset and the NYC WI-FI Hotspot Locations.

After having the dataset built, a Decision Tree and Random Forest model were applied to the data in order to extract the feature importance for predicting the occurrence or not (target variable) of any death on each collision. The feature importance can be visited below where each feature is represented by a number on the x-axis.

**Features:**

5 - Bus Stop Presence
6 - Subway Entrance Presence
7 - WI-FI Hotspot Presence

Fig. 19



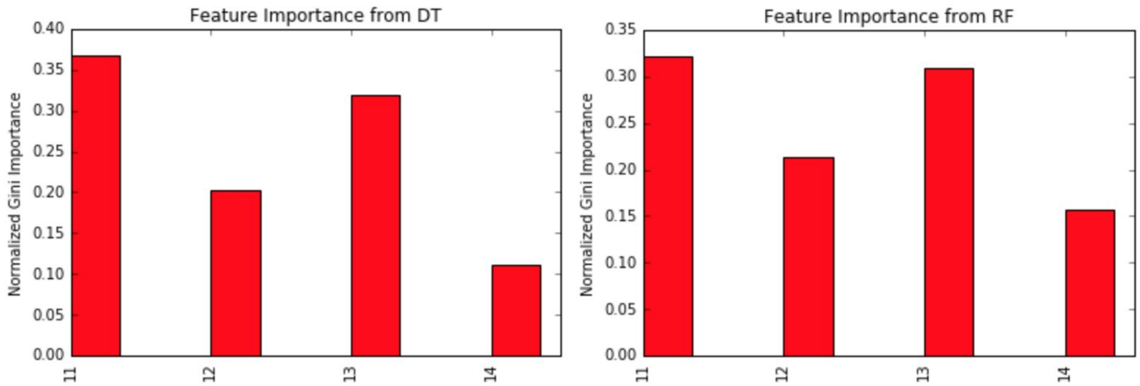

From both graphs, it is clear that the presence of a bus stop is highly correlated with deaths on collisions. Also, subway entrance presence has some influence as well. This information can help the city to create new ways to minimize collision leading to death around bus stop sites.

Trying to optimize the model, a new dataset was engineered. Now, the presence of public schools was introduced. Another change was in the distance between the place where the collision took place and the other features. If the bus stop, subway entrance, and wifi spot were up to fifty meters from a collision point, the new feature would present one or zero otherwise. For the schools, the range was a hundred meters for schools being big building with latitude and longitude centered in the middle of the construction.

**Features:**

11 - Bus Stop Presence
12 - Subway Entrance Presence
13 - WI-FI Hotspot Presence
14 - Schools Presence

Fig. 20




It is curious how the wifi hotspot got importance although it started to be available by the end of 2015. This can call attention to the fact that they are located in areas with high density of people like Midtown (LinkNYC, 2016). Therefore, the city can start new initiatives addressing high density areas for minimizing deaths on traffic.

# CHAPTER 4

# CONCLUSIONS

An important aspect noticed is that Vision Zero is transforming the lives of pedestrians and drivers in New York City. This change is positive, and is being successful given the decrease on the number of collisions on traffic.

The action plan resorting to law enactment and enforcement is being effective on its intent of reducing the citywide speed limit to twenty five miles per hour. The great outcome of this approach is going to be the fifty percent higher chance of survival if someone is hit by a car at the new speed when compared to the old one. It is important that the city keeps intensifying its work through law enforcement and education for achieving even lower rates of death on traffic.

The street redesign analysis gave a mix of information, but in the end it came to be very resourceful. It made possible to understand that a key factor is the people's behaviour. As a result, no other better way to straight a bad behaviour exists than education. The city started a widespread education approach reaching out to the public, and this needs to be enhanced and endured, which will probably bring the best results to the Vision Zero Program.

The correlation discovery pointed out to new aspects. The city should create policies towards high density areas like Midtown. Also, bus stops should receive special attention from drivers. One way of mitigating accidents close to this locations would be through the enactment of severe laws for drivers involved in cases close to bus stops or a set of supposed high density areas.

# CHAPTER 5

# DISCUSSION AND FUTURE WORK

This work has some limitations. Most of them are linked to available datasets, chosen algorithms, how datasets were split, collected information, area of coverage, and software limitation.

The are datasets that can strengthen this paper such as one on the drivers who went through training for driving safer, and another on the number of cabs and for-hire cars in NYC. These new sources would bring more certainty on the insights derived so far.

The algorithm used for calculating the distance between collision sites and other features was based on *Haversine*, but a google map api could be more precise leading to a better modeling.

Another detail is the way some data were split. For instance, the data analysed under the street redesign approach had the street changed in February 2014, but the data on 2014 was altogether while it should be split between before and after that date. This detail could lead to variation on the findings. Moreover, the other redesigns need analysis to have more consistent findings, and this paper should understand the reason of each design change. In addition, some data like the average speed one should be split into different levels of time such as day or night and seasons of the year which would provide more insights.

Some calls reaching out to the DOT (Department of Transportation) were not successful on its intent. The idea was to be aware of different information like the frequency of repaving the streets, which could be used as a new feature on building a dataset since drivers can get distracted by avoiding holes on the street and hit a person or another car as a consequence. However, the people on the phone did not have further information other than the ones available on their websites.

This work should be developed on the datasets for the other NYC boroughs in order to derive broader insights about the effectiveness of the program.

One software limitation is that the heat map visualizer did not have any statistical tool attached to it. This could be overcome by using ArcGIS. ArcGIS can draw heat maps deriving conclusions on the percentual change in density from one point in time to another, and it would be useful for studying the impact of Vision Zero in a different level. Also, the models used to discover correlations are not resistant to imbalanced datasets, which do not have the number of occurrences around fifty percent of the whole dataset. Decision Tree and Random Forest can oscillate on their results at each iteration if the datasets are imbalanced. This problem could be minimized with ArcGis. Moreover, a cluster analysis should be done for reinforcing the current insights or coming up with new ones.

Despite the referenced limitations, this work can help NYC to evaluate the Vision Zero program and find ways to improve it. Moreover, other cities can follow this paper as a guideline for verifying the level of success of their implemented policies.

## REFERENCES

"Holidays and Key Dates." *Holidays and Key Dates*. New York City Department of Education, 2016. Web. 07 Dec. 2016.

"Uber Is Taking Millions Of Manhattan Rides Away From Taxis." *Uber Is Taking Millions Of Manhattan Rides Away From Taxis*. Ed. Reuben Fischer-Baum and Carl Bialik. FiveThirtyEight, 08 Mar. 2016. Web. 07 Dec. 2016.

"Mid-February 2014 North American Winter Storm." *Wikipedia*. Wikimedia Foundation, n.d. Web. 07 Dec. 2016.

Service, National Weather. "Major Winter Storms." *US Department of Commerce, NOAA, National Weather Service*. NOAA's National Weather Service, n.d. Web. 07 Dec. 2016

Pavelski, Joel. "Deadly Winter Storm Hits Northeast: The Latest." *New York Post*. N.p., 13 Feb. 2014. Web. 07 Dec. 2016.

"Vision Zero - An Ethical Approach to Safety and Mobility." *Vision Zero - An Ethical Approach to Safety and Mobility - Accident Research Centre*. N.p., n.d. Web. 07 Dec. 2016.

"Library." *Library - Vision Zero*. NYC.gov, n.d. Web. 07 Dec. 2016.

Benne , Katharine D., "Spatial Analysis of Motor Vehicle Accidents in Johnson City, Tennessee, as Reported to Washington County Emergency Communications District (911)." (2010). *Electronic Theses and Dissertations*. Paper 1778. h p://dc.etsu.edu/etd/1778.

*Vision Zero Action Plan*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016.

*Pedestrian Safety Action Plan Manhattan*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016.

*Vision Zero 1 Year Report*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016.

*Vision Zero 2 Year Report*. Rep. NYC.gov, n.d. Web. 07 Dec. 2016.

"Vision Zero." *Wikipedia*. Wikimedia Foundation, n.d. Web. 07 Dec. 2016.

"NYC Open Data." *NYC Open Data*. NYC.gov, n.d. Web. 07 Dec. 2016.

Intersection. "CityBridge | LinkNYC." *CityBridge | LinkNYC*. N.p., n.d. Web. 11 Dec. 2016.