Pengzi Li, Po-Yang Kang, Sam Burns, Asilayi Bahetibieke

Senior Pedestrians in NYC: A Diff-in-Diff Approach to Evaluating Safe Streets for Seniors

Abstract

We provide a condensed version of our study in order to cover the information that is vital to understanding our project, recent progress in our research, and expected goals and challenges. We recently decided to reframe our analysis at a larger geographic scale in order to compensate for the relatively narrow range of possible predicted values for our previous modelling methods. The revised scale presented unforeseen technical challenges but we successfully overcame them and now have updated results from new regression models. These models support our hypothesis that NYC Vision Zero's Safe Streets for Seniors program is reducing pedestrian accidents involving seniors. In the next phase of our research, we will iterate on our previous regression models as a "sanity" check on our methodology, per the advice of Professor Hochfellner. In addition, we have researched various options for displaying our work online. We have developed a website framework using GitHub Pages and experimented with Tableau for creating interactive data visualizations. We are still deciding on the format and preferred hosting service for our final, publicly-accessible web page for displaying our work.

Introduction

Motivation

Nations around the world are continuing to see more and more people flock to cities. The year 2008 marked the first time in human history when more than half of the global population could be found in cities or towns rather than rural environments (Netherland et al. 2011). And that number is only expected to grow: by 2030, two thirds of the global population will reside in cities (Buffel et al. 2016). At the same time, many countries are aging. In member nations of the Organization for Economic Co-operation and Development, the proportion of citizens aged 65 and older has grown from 7.7% in 1950 to 17.8% in 2010 and will reach 25% by 2050 (Buffel et al. 2016).

New York City is projected to see similar increases in its senior population along with new demands for urban infrastructure and policy geared towards seniors. Creating an age-friendly NYC, however, is no easy task. As Buffel and Phillipson illustrated in their study of cities across the globe, significant obstacles inhibit the realization of age-friendly policies and urban design. Austerity measures have been taken in many cities including NYC while other pressures like housing redevelopment and gentrification add additional resistance to the realization of age-friendly policies (Buffel et al. 2016).

Pedestrian Safety as a Key Aspect of Senior Health

It should be clear at this point that major challenges exist for NYC to accommodate the impending growth of its senior population. And these challenges will come from many different angles, whether it be access to affordable housing, medical care, or adequate transportation options, to name a few. One critical aspect of the age-friendliness of NYC will be the walkability of its streets. Many New Yorkers spend a significant part of each day walking to a variety of destinations. Seniors, too, require the same level of accessibility. Whether it is for work, grocery shopping, visiting friends, or just for exercise, seniors need adequate access to NYC sidewalks.

Studies show that seniors face unique challenges when it comes to navigating NYC on foot. Senior pedestrians suffer higher fatality rates in pedestrian accidents in NYC than any other part of the US (Nicaj et al. 2006) - an outcome that leads to the highest highest hospitalization and fatality rates among all injured pedestrians accident (Miller et al. 2004).

NYC Policy and Relevant Literature

To combat this problem, NYC DOT came up with its Safe Streets for Seniors program (SSfS), which is designed to reduce traffic accidents that result in serious injuries to or the death of senior pedestrians. Since 2008, the DOT has implemented changes in 41 "Senior Pedestrian Focus Areas" (SPFAs), which the agency identified as areas with high densities of vulnerable senior pedestrians based on the density of crashes involving senior pedestrians, senior pedestrian traffic, and the locations of senior centers and senior housing.

Once SPFAs had been identified, the DOT added new structural changes (called "Street Improvement Projects" aka SIPs) to these SPFAs, such as "extending pedestrian crossing times at crosswalks to accommodate slower walking speeds, constructing pedestrian safety islands, widening curbs and medians, narrowing roadways, and installing new stop controls and signals" ("NYC DOT - Safe Streets for Seniors", 2019).

Researchers have been hesitant to declare the program a success. One study found parts of the program to be effective while other implementations (midblock pedestrian fencing) were found to have no effect on senior pedestrian accidents (Chen et al. 2013). A separate study found 10 candidate community districts in NYC with relatively high concentrations of senior residents and pedestrian accidents that had not been targeted as SPFAs ("Aging with Dignity: A Blueprint for Serving NYC's Growing Senior Population").

Hypothesis

Ultimately, our hypothesis is that SPFAs reduce the rate at which senior pedestrians in NYC are killed by motor vehicles. Should our hypothesis prove to be true, the city's policy could be promoted and adopted by other urban areas. If our hypothesis proves to be false, we could conduct a further analysis of the city's selection of SPFAs in order to find potential improvements in their site selection process.

Data

Our diff-in-diff analysis relies primarily on two data sets - traffic fatality data collected from the National Highway Traffic Safety Administration (NHTSA) and SPFA spatial data (see Figure 1 for an overview of the NHTSA and SPFA datasets).

NHTSA Data

The NHTSA publishes nationwide motor vehicle crash data going back decades as part of its Fatality Analysis Reporting System (FARS). Crashes reported in FARS "must involve a motor vehicle traveling on a trafficway customarily open to the public and must result in the death of at least one person (occupant of a vehicle or a non-motorist) within 30 days of the crash" ("Fatality Analysis Reporting System", 2016). The NHTSA publishes data on each accident (date, time, and location of the crash, for example) along with other, related data (number of motorists and pedestrians involved, the age and health outcome of each person involved in the crash, etc.).

We collected crash data from 2001 to 2017 along with data on individuals involved in each crash. The data related to some of these variables was kept strictly for experimental control purposes. These control variables include: whether or not drugs or alcohol were involved in the accident, the total number of pedestrians involved, age and gender, and the severity of the injuries suffered.

SPFAs

As previously mentioned, SPFAs are specific areas in NYC that have been identified by NYC DOT as especially dangerous for senior pedestrians living in those areas. DOT has identified 41 SPFAs in total and has undertaken various infrastructure projects to improve the safety of these areas. Spatial data that describes SPFAs is published online by NYC Open Data. We have collected this data and used it for visualization and initial descriptive statistics

It should be noted that there is some limitation to the data available to the public. While the DOT does disclose the year in which SPFAs have been implemented (25 in 2008, 12 in 2012-2013, and 4 in 2017), more specific information regarding the timing of these projects is not available. We made repeated attempts to request this information from NYC DOT, however, the agency declined to provide it. We are, however, currently in touch with Cordell Schachter, the Chief Technology Officer at NYC DOT, who may be able to provide further details on the timing of SPFA implementation.

Methodology

The difference-in-difference (diff-in-diff) method is a statistical technique used in econometrics and quantitative research in the social sciences that attempts to mimic an experimental research design using observational study data by analyzing the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment. It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group（"How Much Should We Trust Differences-In-Differences Estimates?").

Previous research has neglected to apply diff-in-diff methodology specifically to the analysis of SPFA effectiveness despite the apparent applicability of the method to this subject area. We therefore decided to analyze traffic accidents in NYC involving senior pedestrians using a diff-in-diff methodology, where we would compare accident rates over time between neighborhoods that received the SPFA "treatment" and those that did not.

Results

Regression

Since our last presentation, we have turned our focus toward regression analysis of the data. Our regression models include an interaction term between time and the treatment as well as covariance terms for the control variables previously described (e.g. presence/absence of drugs/alcohol, total number of pedestrians involved in accident, etc.):

$$Y = \beta 0 + \beta 1 \times Time + \beta 2 \times Treatment + \beta 3 \times (Time \times Treatment) + \beta 4 \times Covariates + \varepsilon$$

Up until this point, we have worked with the data primarily on an accident-by-accident level. That is, our data set included one row per accident. This data structure was suboptimal since our regression model would have a very narrow range of predicted values (the number of senior pedestrians killed in an accident). More than 88% of accidents involved either 0 or 1 senior pedestrians and the maximum number of seniors involved in a single accident was 6. Based on these statistics it is clear that the power of a regression model for the accident-by-accident data would be very limited.

We therefore decided to aggregate the data at the level of the census tract. Despite some initial challenges, we successfully aggregated the data (see Figure 2) and have applied various linear regression models to understand how different combinations of exogenous variables affect our analysis. We tested models for three separate endogenous variables: the number of seniors involved in accidents that involved pedestrians, the number of seniors killed in accidents that involved pedestrians, and the number of senior pedestrians killed in vehicular accidents. For each of these endogenous variables, we tested four separate regression models, making a total of 12 separate models.

For each endogenous variable, we created the following models: model 1 included exogenous variables for when the accident occurred and whether or not the first SPFAs had been implemented; model 2 added exogenous variables for the presence of alcohol or drugs, inclement weather conditions, and time of day; model 3 added exogenous variables for whether or not the accident took place in round 2 or 3 SPFAs (implemented in 2012 and 2017, respectively);

finally, model 4 added an exogenous variable for the gender of senior pedestrians involved in accidents.

The simplest model, model 1, had low R-squared values for each endogenous variable. Models 2, 3, and 4, however, produced much more promising R-squared values between .43 and .68. The models also showed that there was a statistically significant negative trend for our interaction variable for the year when the accident occurred and whether or not the accident occurred in an SPFA. For a summary of our modelling results, see Figure 3 and Figure 4.

Website

In addition to data aggregation and regression modelling, we have also researched and experimented with different options for making our work publicly-accessible via the internet. We considered a number of potential hosts - including popular website building services like Squarespace and Wix - but found that GitHub Pages and/or Tableau would provide the service and functionality we need for our project.

We have already created a template site (https://asilayi.github.io/Senior-Pedestrians-in-NYC-A-Diff-in-Diff-Approach-to-Evaluating-Safe-Streets-for-Seniors/) with which we could easily add text, images, and other content in order to finalize our end product. Our only concern at this time is whether and how to display information interactively. GitHub Pages is not well suited for interactive data visualization, however, Tableau makes this relatively simple. We have therefore experimented with adding data to Tableau and creating interactive plots of our data that include spatial and temporal filters (see Figure 5). We are currently testing whether or not our interactive visualization can be embedded in our GitHub

Pages site. If it cannot be embedded, we are considering creating a dashboard in Tableau that could be publicly displayed via Tableau Public.

We expect to finalize our decision-making regarding the website by July 1st. As previously described, we have already created sample website versions in GitHub Pages and Tableau. The final implementation of our site will therefore simply build off one or both of these templates and is expected to have a complete first draft by July 15th. We will then schedule a short phase of user testing and feedback so that we can refine the site, which will be completed by July 22nd. As we complete the site and our final regression results, we will also continue to update our final report and create our final presentation, which we plan to have completed by July 23rd.

Conclusions

Our regression models showed that there is a negative correlation between accidents that occurred within SPFAs and the number of seniors killed in total and, specifically, in the number of senior pedestrians killed. In other words, the models suggest that NYC's SSfS program is helping to reduce the number of seniors killed each year by motor vehicles. Given that our models show statistically significant improvement, other cities should consider adopting a program similar to SSfS in order to protect their own senior populations.

In the coming weeks, we will finalize our website and create additional regression models as a "sanity" check for the results we have acquired thus far. In these models, we will experiment with different combinations of exogenous variables in order to ensure that we have not achieved spurious results. Finally, should time allow, we will create an additional model predicting where senior pedestrian accidents are likely to occur based on patterns in where seniors live now and where they are expected to live in the future. DOT could use the results of our analysis to determine where to expand the program to.

The following provides a final summary of our expected project timeline (see corresponding Gantt chart):

- Final selection and planning for website design elements - July 1, 2019
- Regression modelling “sanity” check and finalization - July 8, 2019
- Final conclusions drawn from all regression modelling results - July 15, 2019
- Completed first draft of website - July 15, 2019
- Implementation of user feedback and completion of website - July 22, 2019
- Completion of final report and presentation - July 23, 2019

Team collaboration statement

The members of our team have contributed in the following ways:

- Pengzi Li - Data wrangling, feature engineering, linear regression model construction, and statistical analysis.
- Po-Yang Kang - Researching policies on the city, state, and country levels, and cleaning datasets corresponding to those policies.
- Sam Burns - Report writing, communications with outside organizations, project coordination, analysis.
- Asilayi Bahetibieke - Data searching, data cleaning, unify data format, report writing.

Figures and Tables

Type of Data	*Datasets*	*Relevant Fields*
FARS Data	**Accident Dataset**	**ST_CASE, Date, Latitude, Longitude**
FARS Data	**Person Dataset**	**Age, Sex, Person Type, Date of Death, Drinking involvement, Drug involvement, Related Factors, Dead at scene or not, Injury severity**
SPFA Data	**SPFA shapefile**	**SPFA Name, SPFA Implementation Round, Shape area, Longitude, Latitude**

Figure 1: Table describing each of the two primary data sets used in our analysis of NYC's SPFAs.

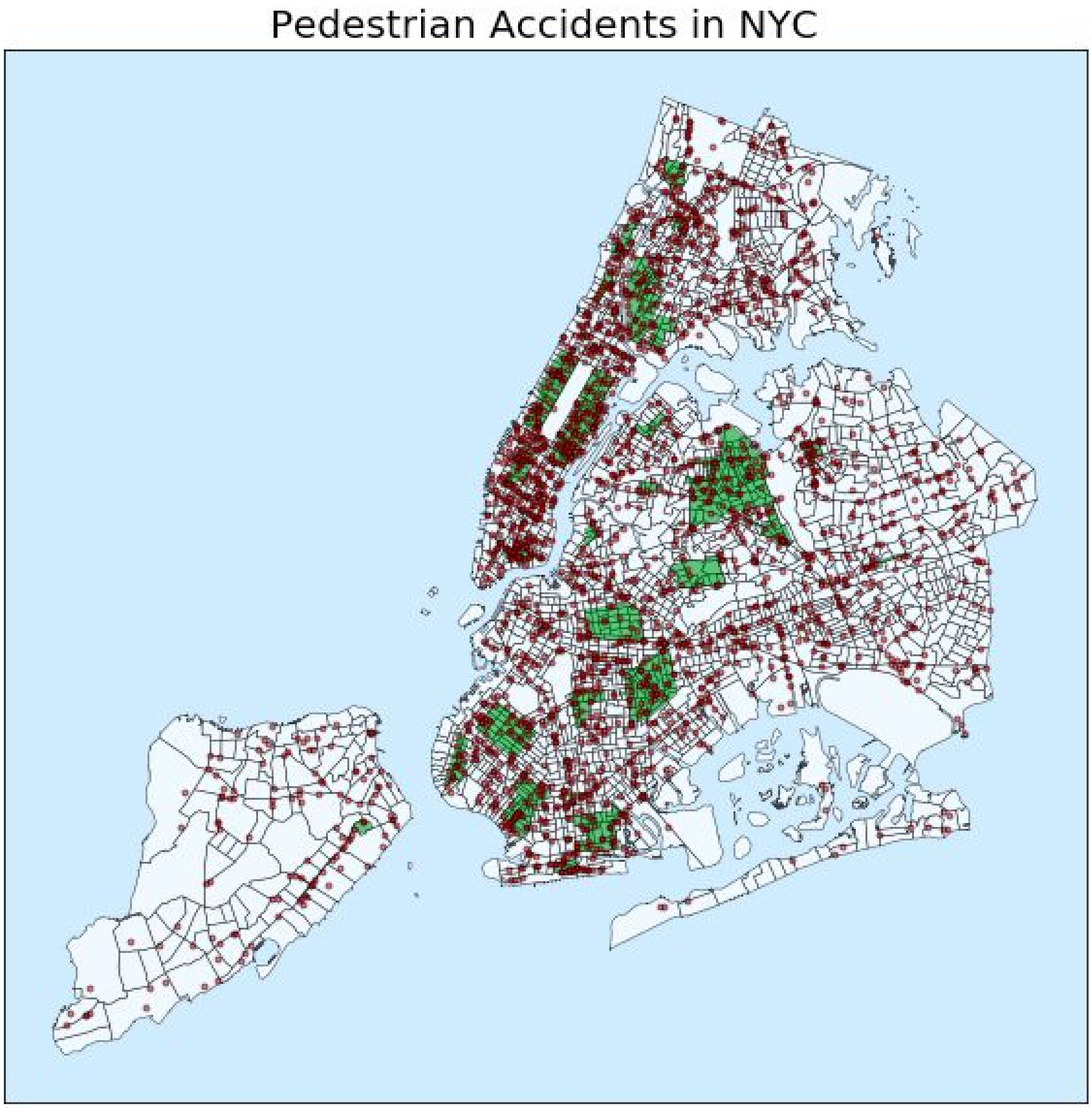

Figure 2: Plot of pedestrian accidents in NYC in red with SPFAs plotted in green. Note that the base layer for NYC is now plotted with census tracts, rather than PUMAs.

Endogenous Variable	Model 1	Model 2	Model 3	Model 4
Accident involved a senior	0.067	0.475	0.477	0.686
Number of senior pedestrians killed in accidents	0.068	0.437	0.446	0.572
Number of seniors killed in accidents	0.072	0.434	0.439	0.601

Figure 3: R-squared values for each model and each endogenous variable. Notice that R-squared values improve dramatically from model 1 to model 2 due to the addition of other exogenous variables, such as weather conditions and the presence of drugs and/or alcohol.

Endogenous Variable	Model 1	Model 2	Model 3	Model 4
Accident involved a senior	Not significant	Not significant	Not significant	Not significant
Number of senior pedestrians killed in accidents	Significant	Significant	Significant	Significant
Number of seniors killed in accidents	Significant	Significant	Significant	Significant

Figure 4: Significance results for our interaction term between the year in which the accident occurred and whether or not it occurred within an SPFA. Coefficients for this variable were negative across all models, however, only the models predicting on the number of senior pedestrians killed or overall seniors killed in accidents had significant results for our interaction term. Still, the models suggest that NYC's SSfS program is helping to reduce fatalities for seniors across the city and, in particular, senior pedestrians.

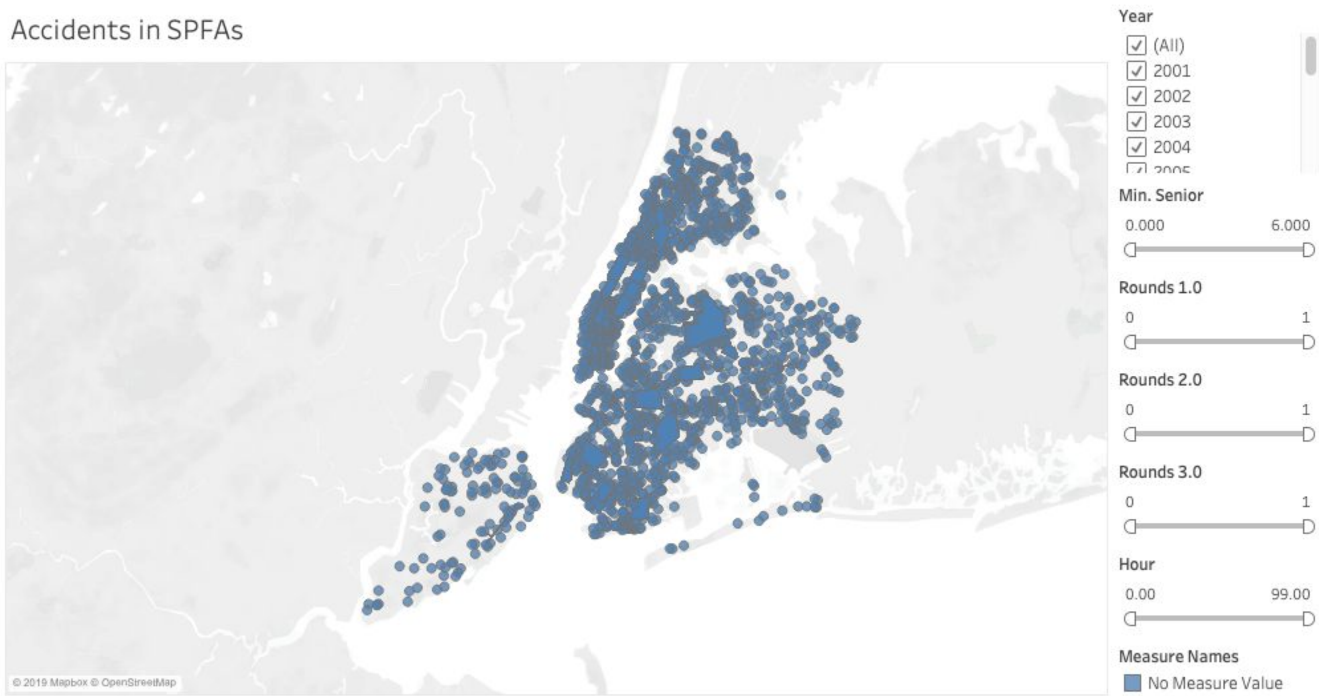

Figure 5: Screenshot from Tableau Desktop dashboard of data analysis. The plot can be adjusted temporally and in terms of the points and polygons that are displayed.

References

"About the Global Network for Age-Friendly Cities and Communities." *Age-Friendly World* (blog). Accessed April 19, 2019. https://extranet.who.int/agefriendlyworld/who-network/.

"Aging with Dignity: A Blueprint for Serving NYC's Growing Senior Population : Office of the New York City Comptroller Scott M. Stringer." Accessed April 27, 2019. https://comptroller.nyc.gov/reports/aging-with-dignity-a-blueprint-for-serving-nycs-growing-senior-population/.

Angrist, Joshua D. and Pischke, Jörn-Steffen. "Mostly Harmless Econometrics." Princeton University Press. Accessed April 27, 2019. https://press.princeton.edu/titles/8769.html.

Marianne Bertrand, Esther Duflo, Sendhil Mullainathan. "How Much Should We Trust Differences-In-Differences Estimates?" The Quarterly Journal of Economics, (February, 2004): 249–275. https://doi.org/10.1162/003355304772839588.

Buffel, Tine, and Chris Phillipson. "Can Global Cities Be 'Age-Friendly Cities'? Urban Development and Ageing Populations." *Cities* 55 (June 1, 2016): 94–100. https://doi.org/10.1016/j.cities.2016.03.016.

Chen, Li, Cynthia Chen, Reid Ewing, Claire E. McKnight, Raghavan Srinivasan, and Matthew Roe. "Safety Countermeasures and Crash Reduction in New York City--Experience and Lessons Learned." *Accident; Analysis and Prevention* 50 (January 2013): 312–22. https://doi.org/10.1016/j.aap.2012.05.009.

"Difference-in-Difference Estimation | Columbia University Mailman School of Public Health." Accessed April 27, 2019. https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation.

"Fatality Analysis Reporting System (FARS)." Text. NHTSA, November 14, 2016. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars.

Miller, Ted R., Eduard Zaloshnja, Bruce A. Lawrence, Jeff Crandall, Johan Ivarsson, and A. Eric Finkelstein. "Pedestrian and Pedalcyclist Injury Costs in the United States by Age and Injury Severity." *Annual Proceedings / Association for the Advancement of Automotive Medicine* 48 (2004): 265–84.

"Motor Vehicle Crash Pedestrian Deaths in New York City: The Plight of the Older Pedestrian. - PubMed - NCBI." Accessed April 27, 2019. https://www.ncbi.nlm.nih.gov/pubmed/17170193.

Netherland, Julie, Ruth Finkelstein, and Paula Gardner. "The Age-Friendly New York City Project: An Environmental Intervention to Increase Aging Resilience." In *Resilience in Aging: Concepts, Research, and Outcomes*, edited by Barbara Resnick, Lisa P. Gwyther, and Karen A. Roberto, 273–87. New York, NY: Springer New York, 2011. https://doi.org/10.1007/978-1-4419-0232-0_18.

Nicaj, L., S. Wilt, and K. Henning. "Motor Vehicle Crash Pedestrian Deaths in New York City: The Plight of the Older Pedestrian." Injury Prevention 12, no. 6 (December 1, 2006): 414–16. https://doi.org/10.1136/ip.2005.010082.

"NYC DOT - Safe Streets for Seniors." Accessed April 15, 2019. https://www1.nyc.gov/html/dot/html/pedestrians/safeseniors.shtml.