

Testing for similarity in area-based spatial patterns: Alternative methods to Andresen's spatial point pattern test

Andrew P. Wheeler^{1,**}, Wouter Steenbeek², and Martin A. Andresen³

*Data and code to replicate the findings can be obtained from
https://www.dropbox.com/s/zx5db9qashfufrw/Analysis_AltMethods_SPPT.zip?dl=0

**Corresponding author, email: apwheele@gmail.com

1. School of Economic, Political, and Policy Sciences -- Criminology Program

The University of Texas at Dallas

800 West Campbell Road, Mail Station GR 31

Richardson, Texas 75080-3021

2. Netherlands Institute for the Study of Crime and Law Enforcement

3. Simon Fraser University, School of Criminology and Institute for Canadian Urban Research Studies

Andrew P. Wheeler is an Assistant Professor of criminology at the University of Texas at Dallas in the School of Economic, Political, and Policy Sciences. His research focuses on the spatial analysis of crime at micro places and practical problems faced by crime analysts. His recent work has been published in the *Homicide Studies*, *Journal of Quantitative Criminology*, *Crime & Delinquency*, *Security Journal*, *Cartography and Geographic Information Sciences*, the *International Journal of Police Science and Management*, and the *Journal of Investigative Psychology and Offender Profiling*.

Wouter Steenbeek is a researcher at the Netherlands Institute for the Study of Crime and Law Enforcement. His research focuses on spatio-temporal crime patterns, crime concentration, social capital, and quantitative research methods. His recent research has been published in the *Journal of Quantitative Criminology*, *Applied Geography*, *The Review of Economics and Statistics*, *Health Services Research*, and *Social Forces*.

Martin A. Andresen is a Professor of criminology at Simon Fraser University, School of Criminology and Institute for Canadian Urban Research Studies. His research focuses on spatial crime analysis, crime and place, geography of crime, environmental criminology, applied spatial statistics and geographical information analysis. His recent research has been published in *Applied Geography*, *Canadian Geographer*, *Crime Science*, *Crime Prevention & Community Safety*, *Policing and Society*, *Security Journal*, *Journal of Quantitative Criminology*, and *Urban Geography*.

Testing for similarity in area-based spatial patterns: Alternative methods to Andresen's spatial point pattern test

Abstract: Andresen's spatial point pattern test (SPPT) compares two spatial point patterns on defined areal units: it identifies areas where the spatial point patterns diverge and aggregates these local (dis)similarities to one global measure. We discuss the limitations of the SPPT and provide two alternative methods to calculate differences in the point patterns. In the first approach we use differences in proportions tests corrected for multiple comparisons. We show how the *size* of differences matter, as with large point patterns many areas will be identified by SPPT as statistically different, even if those differences are substantively trivial. The second approach uses multinomial logistic regression, which can be extended to identify differences in proportions over continuous time. We demonstrate these methods on identifying areas where pedestrian stops by the New York City Police Department are different from violent crimes from 2006 through 2016.

Keywords: hot spots, point pattern, pedestrian stops, violent crime, crime analysis

Data and anonymous code to replicate the analysis can be downloaded from

https://www.dropbox.com/s/j8bhyo5y994yyy0/Analysis_AltMethods_SPPT_Anony.zip?dl=0

Introduction

Andresen's point pattern test (SPPT) has been widely applied to geographic analyses of crime and police behavior (Andresen, 2009; 2016). SPPT examines the relative difference in the proportions of two point patterns, either within administrative areas or a regular lattice. For example, one may wish to examine the proportion of traffic stops for speeding conducted in patrol zone A (out of all traffic stops) relative to the proportion of accidents in patrol zone A (out of all accidents in that jurisdiction). If the proportion of traffic stops are lower than accidents, a police department may wish to allocate more resources to that patrol zone. Thus, one looks at the *relative* distribution of two point patterns, and not at the absolute density of the two patterns.

We discuss two alternative methods that have the same purpose but improve upon the existing SPPT: a direct test of the differences in the proportions and multinomial logistic regression. This paper makes two specific contributions. First, we emphasize the absolute size of the differences in proportions in contrast to SPPT's reliance on a dichotomous decision of whether an area is different or not. This is especially important when considering large sample sizes, in which case many statistically significant but substantively trivial differences may be detected using traditional null hypothesis testing. Second, we show how to go beyond a simple comparison of two samples via the inclusion of covariates. In particular, we illustrate how one can assess changes in the proportions over continuous time.

We illustrate these methods via a case study examining pedestrian stops and violent crimes in New York City occurring from 2006 through 2016. Pedestrian stops, instances in which police officers stop individuals on the street due to reasonable suspicion that a crime may occur (MacDonald et al., 2016), were a critical component of the New York City Police Departments (NYPD) broken windows policing strategy, but were criticized as

disproportionately targeting minority communities (Fagan & Davies, 2000). Subsequent civil litigation resulted in dramatic reductions of the total number of stops, from over half a million per year to under 20,000. Our analyses compare the spatial pattern of pedestrian stops to the pattern of violent crimes in grid cells of 500 by 500 meters over the entire city, as well as at the precinct level over time.

Andresen's Point Pattern Test

Andresen's procedure to identify statistically different locations can be described briefly as follows: 1) identify a point pattern as a base set and set the other point pattern as the test set, 2) use a resampling procedure to estimate a confidence interval around the proportion in each area of the test set, and then 3) determine whether the base set proportion falls inside or outside of the confidence interval for the test set in each area. Continuing with the traffic stops and accidents example, say patrol zone A had 5% of stops, and 10% of accidents. With the accidents point pattern as the test set, the resampling procedure may construct an interval of 4% to 16%. In that case, one would not conclude that the proportion of traffic stops is statistically different than the proportion of traffic accidents, as the 5% of stops is within the confidence interval for accidents (see Steenbeek et al. 2017 for an R implementation of the test and further examples).

Example applications in criminology have examined temporal changes in crime patterns over time, different seasons, or different days of the week (Andresen & Malleson, 2011; 2013; 2015; Andresen et al., 2017; Linning, 2015; Vandeviver & Steenbeek, 2017; Pereira et al., 2016), differences in the spatial distribution of different crime types (Andresen & Linning, 2012; Melo et al., 2015; Vaughan et al., 2016), the change in crime distribution after an intervention (Andresen & Malleson, 2014; Hodgkinson et al., 2016), differences in drug activity as measured via police calls versus EMS data (Hibdon et al., 2017), estimate the amount of error introduced via aggregating police incidents (Tompson et al., 2015), and the overlap of proactive policing and reported crime locations (Wu and Lum, 2017).

SPPT's original procedure has two limitations. First, one must arbitrarily choose which point pattern is the base and test dataset, but results are not necessarily invariant to this choice. For example, if the test set has a much larger sample size, it will result in smaller confidence intervals and, subsequently, more areas may be marked as statistically different. If you switched the analysis so that the smaller sample size was the test set, the intervals around the proportions would be much larger and, thus, the researcher may conclude the patterns are more similar.

The second problem is that areas with zero points in the test set will always have a confidence interval from 0% to 0% (Wang, 2013).¹ For example, imagine a point pattern with 100 traffic accidents, but one area did not receive any accidents. In the resampling approach, that area will never result in being chosen, and so the confidence interval for the proportion will always be zero. Instead, the confidence interval should be (small but) non-zero (assuming it can viably be the location of a traffic accident, e.g. it is not a lake with no roads): e.g. a 95% exact

¹ There is actually no need to construct confidence intervals via a resampling approach as they can be calculated exactly. When resampling with replacement, the probability that a particular point will be drawn is defined by the binomial distribution. Resampling without replacement is defined by the hypergeometric distribution. We illustrate how to construct these 'exact' resampling intervals via R code in the supplementary data files.

binomial confidence interval given an observed proportion of 0/100 is 0% to 3% (Agresti, 2002). So, SPPTs bootstrapping approach falsely assumes the proportion will always be zero, and will subsequently conclude the patterns are different, even if in the base set the proportion is quite small (but non-zero).

To address these limitations we suggest two approaches. When the number of areas to assess changes in proportions is large, and some areas can have zero observations, we formulate a test for the difference in proportions in each area (Campbell, 2009). Because this involves a large number of hypothesis tests, we use multiple comparison corrections to the p-values to limit the number of false positives (Brunsdon and Charlton, 2011). This approach does not require one to select a particular point pattern as a base and test set. Additionally, one can construct a continuous measure of the size of the differences in the two point patterns rather than a dichotomous measure. This is important because differences may be statistically significant but substantively trivial, which is not taken into account by SPPT.

The second approach we suggest when the number of areas is smaller² is multinomial logistic regression models (Agresti, 2002). We estimate the differences in relative proportions for the two point patterns using a multinomial model with the areal unit as the outcome. We also show how the multinomial model can be extended to more complicated research designs. In particular, we show how it can be used to model the relative proportions in continuous time, which prevents one from needing to bin the observations into arbitrary time periods, such as seasons or years.

Police Stops and Crime in New York City

To illustrate the use of our proposed methods we conduct a case study on the differences in the spatial allocation of pedestrian stops by the NYPD and violent crimes in New York City occurring from 2006 through 2016.

The motivation for such analysis is because pedestrian stops have been a main component of broken windows policing, i.e. targeted enforcement of minor offenses, by the NYPD (Kelling and Coles, 1997). The NYPD emphasized targeting hot spots of crime through COMPSTAT meetings in which the display of maps and recent crime statistics were a major component (Silverman, 1999). ‘Stop, question, and frisk’ became a particularly major crime fighting component of the NYPD during Police Commissioner Raymond Kelly’s second term, which saw the number of recorded stops rise to over half a million per year during the aughts (Levitt, 2018).

Prior research shows that pedestrian stops by the NYPD are targeted at hot spots of crime (Weisburd et al., 2014), and some have shown stops have slight deterrent effects on crime (MacDonald et al., 2016; Rosenfeld & Fornango, 2017; Weisburd et al., 2016; Wooditch & Weisburd, 2016), but others have not found similar deterrent effects (Greenberg, 2014; Rosenfeld & Fornango, 2014; Sullivan & O’Keeffe, 2017). In addition to hot spots identified by precinct commanders at monthly COMPSTAT meetings, the NYPD additionally identified

² While we have no hard and fast rules for what counts as large or small in these circumstances, generally a large number of areal units will be when analysts use micro spatial aggregations, like street segments or a detailed grid over the study area, leading to much more than 100-200 spatial units in total.

IMPACT zones, which were areas with high levels of violent crime that were additionally targeted for proactive policing (MacDonald et al., 2016; Smith & Purtell, 2007).

While these policies have (arguably) been attributed to the crime decline starting in the 1990s in NYC (Rosenfeld et al., 2007; Silverman, 1999), pedestrian stops by the NYPD have come under additional scrutiny, as they were disproportionately concentrated in minority communities (Gelman et al., 2007; Fagan & Davies, 2000; Ferrandino, 2015). This subsequently resulted in civil litigation against the NYPD, which resulted in a court appointed monitor and a large reduction of stops (Sweeten, 2016). In 2011 the NYPD recorded over 600,000 pedestrian stops, but in 2014 they only recorded a total of 45,000 stops. Such a decline did not occur on a specific date, but showed decreases starting April 2012.

These events provide the backdrop to two particular research questions. The first is to what extent pedestrian stops and violent crimes overlap. Prior analysis has accused the NYPD of not targeting enforcement at the places of highest crime, but at minority communities (Fagan & Davies, 2000). The maps and analysis provided here provide a graphical test to show where stops and violent crimes occur disproportionately. The second research question examines how the dramatic reduction of stops due to civil litigation change the spatial distribution of stops. After such reductions, did the NYPD allocate stops in a more equitable manner? Or did the relative spatial distribution of stops stay consistent, even though the total number of stops were greatly reduced?

Data Description

Our data come from publicly available sources. Crime data geocoded to the street segment level were obtained from New York City's open data portal, <https://opendata.cityofnewyork.us/>. Recorded pedestrian stops geocoded to the address level by the NYPD were obtained from <https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page>. We analyze stops and part 1 violent crimes (murder, assault, robbery, and rape) from 2006 through 2016. While publicly available pedestrian stop data are available going back to 2003, the publicly available crime data are only provided starting in 2006. After eliminating events that were not geocoded or were missing the date the event was reported in the original datasets³, there were a total of 4,067,570 pedestrian stops and 412,698 part 1 violent crimes over the study period.

Two geographic units are used in this study. For small area analyses, we construct a regular grid of 500 by 500 meters over the study area, and aggregate stops and part 1 violent crimes to those grid cells via a point-in-polygon check.⁴ There are in total 2,976 grid cells that had at least one violent crime or stop in the study area over the eleven-year period. We eliminate grid cells with zero observed stops *and* violent crimes, as many of these locations cannot have a recorded incident, such as water areas. Only analyzing locations with at least one event has been described as a *robust* measure for Andresen's global S index (Vandeviver & Steenbeek, 2017). While this may eliminate land area that could have a recorded stop or crime, because we use a

³ We additionally eliminated violent crimes that were recorded on Riker's Island, the local jail for all New York City boroughs. This is because no *pedestrian* stops could occur within the jail.

⁴ The data were projected into the 1983 North American Datum state plane system for New York and Long Island.

long sample period and a large number of incidents, it is unlikely that we eliminate many areas that can have a recorded incident.

The second geographic unit of analysis are the 77 NYPD precincts in New York City. The original database of crimes and stops both had a field recording the precinct of the stop. NYPD precincts (as well as borders for the New York City boroughs) were obtained from the Bytes of the Big Apple administrative boundaries, currently available at <https://www1.nyc.gov/site/planning/data-maps/open-data.page>.

Grid cells are used as the geocoding in the provided public data make it difficult to use other geographies, such as street segments (Weisburd et al., 2014; Vandeviver & Steenbeek, 2017), or census geographies like blocks or block groups (Bernasco & Block, 2010; MacDonald et al., 2016). Crimes are provided at the middle of street segments (Singleton & Brunsdon, 2014), which provides a degree of privacy, but prevents aggregating crimes to census geographies, whose borders are defined by streets. Pedestrian stops are geocoded to the address level, but many stops (over 50% in this sample) are recorded at intersections (Wooditch & Weisburd, 2016). This is a common procedure for police departments for events occurring outdoors where it may be difficult to associate an event at a specific address (Levine & Kim, 1998; Wheeler et al., 2016). This makes it difficult to use street segments as a unit of analysis though, as intersections are of course in-between multiple street segments.

Constructing a regular grid of 500 by 500 meters is ad hoc, but mostly eliminates the problems of inconsistency between the geocoding of the two point patterns and provides reasonably small areas, close in size to census block groups (Wheeler, 2017). It also provides enough observations to reasonably estimate the difference in percentages between the two point patterns in many areas. Smaller areas will result in smaller percentages of the total and, thus, have a larger standard error (Gelman & Price, 1999).

Methods

Proportion test

We propose two different methods to address some of the shortcomings of the Monte Carlo approach of SPPT. The first is to construct a direct test of the differences in the proportions, as opposed to constructing a confidence interval around one pattern, and then treating the other as fixed. While this direct test in the differences in proportions can be constructed in several ways (Agresti, 2003), here we use a simple Chi-Square approach with Yate's N-1 continuity correction. Other potential ways would be to estimate Fisher's exact test (Fisher, 1925), estimate bootstrap differences in proportions for both sets, or a shrunk estimate of the differences in proportions (Agresti & Caffo, 2000). We use the continuity corrected Chi-Square based on the recommendations of Campbell (2007), showing the continuity corrected Chi-Square has more power than Fisher's Exact test, and tends to work well even when cells have an expected number of observations as few as one. We then apply this test to 2,976 grid cell areas for the violent crime and pedestrian stop point patterns aggregated over the 11-year period, correcting p-values based on the false discovery rate, as suggested in Brunsdon & Charlton (2011).

Multinomial logistic regression

The second method we suggest is using multinomial logistic regression to predict the differences in probability between the proportion of stops and violent crime at the precinct level. If we consider the probability that a particular point i falls within precinct j to equal $\pi_j(i)$ conditional on covariates x , the multinomial logistic regression model can be written as (Agresti, 2002):

$$\log \frac{\pi_j(i)}{\pi_1(i)} = \alpha_j + \beta_j' x, \quad j = 1, \dots, J - 1 \quad (1)$$

In this formulation, one area must be chosen as the reference for the denominator, but that choice is immaterial as we are ultimately only interested in the estimated probability of a point being contained within a particular area. To obtain the predicted probabilities from this model for any specific area j , one sums over all of the coefficients for the $J-1$ different categories.

$$\pi_j(i) = \frac{\exp(a_j + \beta_j' x)}{1 + \sum_{h=1}^{J-1} \exp(a_h + \beta_h' x)} \quad (2)$$

To replicate the results in the basic SPPT test, a dummy variable is included to distinguish between the point pattern types, so one can create an estimate probability of an incident occurring in precinct j for the stops, $\hat{\pi}_j(S)$, and for the violent crimes, $\hat{\pi}_j(V)$. Subsequently, we estimate a difference in those proportions, $\Delta\hat{\pi}_j$ and the variance of that difference.⁵ Finally, we construct a confidence interval around that estimate of the differences in proportions (Reeve, 2016).

To illustrate the multinomial model on a simple example, one constructs a dataset (see Table 1), and uses a dummy variable to distinguish between the two point patterns. We select only three precincts (30, 73, and 106) to simplify the example. Given these three areas, we choose the reference area to be precinct 30, and estimate a total of four different coefficients with this multinomial model. Here we use superscripts to denote predicting the specific precinct.

$$\begin{aligned} \log \frac{\pi(\text{PCT} = 73)}{\pi(\text{PCT} = 30)} &= \text{Intercept}^{73} + \beta^{73}(\text{Violent}) \\ \log \frac{\pi(\text{PCT} = 106)}{\pi(\text{PCT} = 30)} &= \text{Intercept}^{106} + \beta^{106}(\text{Violent}) \end{aligned} \quad (3)$$

[INSERT TABLE 1 AND TABLE 2 HERE]

⁵ This can be constructed via the delta method and estimates of the standard error of the original probability estimates. The equations are relatively complex, and involve multivariate differentiation of the covariance matrix of the coefficients. We refer the reader to Fox & Anderson (2006) for a detailed treatment.

Table 2 shows the estimated coefficients from the multinomial logistic regression using the data from Table 1 and the total number of incidents as frequency weights.⁶ Substituting the estimated coefficients from Table 2 into equation 2, the estimated percentages for precinct 73 for stops and violent crimes, respectively, are:

$$\pi(\text{PCT} = 73 \mid \text{Violent} = 0) = \frac{\exp(1.348)}{1 + [\exp(1.348) + \exp(0.153)]} \approx 0.64$$

$$\pi(\text{PCT} = 73 \mid \text{Violent} = 1) = \frac{\exp(1.348 - 0.401)}{1 + [\exp(1.348 - 0.401) + \exp(0.153 - 0.019)]} \approx 0.55$$

These just reproduce the observed proportions for each set, but one can construct a standard error around those proportions and a standard error of that difference. Here the difference in the proportions is 0.094, and the standard error of that difference is 0.003. This procedure is repeated for each area under study.

There are no obvious improvements of using multinomial logistic regression over other approaches of calculating the difference in proportions in this simple example. Being a generalized linear model, the standard errors are based on asymptotic properties. Also the multinomial regression equation is not identified if there are areas with zero observations, as this will cause perfect separation. Neither of these are the case when using the simpler differences in proportion test.

However, an advantage of using multinomial logistic regression over a differences in proportion test is one can incorporate additional covariates into the model. For example, it may be the case that such differences in proportions are due to the fact that some areas are busier in terms of foot traffic and have more individuals at risk of being stopped. To account for this, a covariate capturing the ambient population can be included. Thus one can identify not only whether two areas are different in their proportions, but potentially attribute *why* those areas are different.

An additional benefit of this approach is that one is not limited to just comparing two point patterns. Prior uses of SPPT have binned crimes into years and made comparisons over time (Andresen et al., 2017; Vandeviver & Steenbeek, 2017), e.g. comparing crime in 2010 to crime in 2011. Incorporating time as a covariate allows one to not just make simple year over year comparisons, but allows one to estimate the changes in those proportions continuously over time. This is important if we do not expect the proportions to change abruptly with the coming of a new year, but to vary more subtly over time. In our example case, the NYPD did not decrease the number of pedestrian stops at the start of a new calendar year, but started decreasing stops a few months into 2012.

To capture this process, we incorporate the report date of the incident as a covariate via a set of splines (Durrleman & Simon, 1989), and allow the differences in proportions of stops and

⁶ One can use frequency weights and use the smaller weighted dataset, as opposed to building a larger dataset where each row is an individual incident. In this case it is the difference between having a dataset of six rows versus a dataset of over 290,000 rows.

violent crimes at precincts to vary smoothly over time. We can then assess the size of the differences in proportions between the two point patterns and the changes over time. One could similarly accomplish this via polynomial terms of time, but we prefer splines as they prevent estimates in the tail from greatly affecting the rest of the function (Harrell, 2001). To illustrate the method we use restricted cubic splines and set the locations of knots to every-other year in the sample (i.e. 2008, 2010, 2012, and 2016) which allows us to approximate a variety of functions over the limited temporal domain (Harrell, 2001).

Grid Cell Analysis

Figure 1 displays grid cells having either a higher proportion of stops or violent crime, after correcting for multiple comparisons using the false discovery rate (Brunsdon and Charlton, 2011). Without correcting for multiple comparisons, there were a total of 1,675 grid cells with a statistically significant difference in the proportions (for a global S index of 0.44) at the $p < 0.05$ level. After correcting for multiple comparisons, only 1,222 grid cells shown a statistically significant difference in the proportions (for a global S index of 0.59). Compared to the traditional SPPT Monte Carlo approach, there were many fewer areas identified as different. Using stops as the test set, there were 2,539 areas identified as statistically different, for an S index of 0.15). Using violent crimes as the test set there where 2,799 areas identified as statistically different, for an S index of 0.06.⁷

With such a large number of observations, many locations may still have a trivial difference in the percentages while having a statistically significant difference. Figure 1, therefore, not only filters out grid cells without a statistically significant difference (Mennis, 2006), but displays the grid cells as graduated circles to symbolize the absolute differences.⁸ Blue areas indicate locations with a higher proportional share of stops, and red areas indicate a higher proportional share of violent crimes.

[INSERT FIGURE 1 HERE]

One can identify several interesting spatial clusters in Figure 1. For example, there is a large cluster of locations with a high proportion of stops in upper Manhattan, but subsequently when one travels just slightly north to the Bronx there is a high proportion of violent crime. It, thus, appears more stops should be allocated to the Bronx. Within Brooklyn there are areas with high stops and high violent crime and, thus, it suggests that only slight shifts in the spatial distribution may make each pattern more spatially equitable. There are additionally several clusters within Queens and Staten Island of higher numbers of stops relative to crimes, but these are smaller in comparison to the clusters within Manhattan, the Bronx, and Brooklyn.

⁷ These statistics were computed using the default version of the SPPT, without replacement, with a sample size 85% of the original sample, and with a 95% confidence interval. They however were calculated using an exact interval, so are the intervals one would estimate if one conducted an infinite number of simulations.

⁸ Another potential way to map such results is via value-by-alpha maps (Roth et al., 2010), in which grid cells with higher p-values are shown with a higher level of transparency. We show this, along with the results of the original SPPT, in a supplementary appendix.

Precinct Level Analysis

We now illustrate multinomial logistic regression to estimate the differences in proportions of pedestrian stops and violent crimes within 77 precincts. Because of the large number of precincts, examining the multinomial output in a table is quite difficult. Here we draw graphs to simplify the presentation, laying out the time series of the difference in proportions according to a grid that approximates the geographic contiguity of the NYPD precincts (McNeill & Hale, 2017). Figure 2 displays those differences in proportions over the 11-year period, the percentage of stops minus the percentage of violent crime, using color to distinguish between the five boroughs in New York City. The colored lines display the estimated difference in proportions, which using splines can have non-linear trends over time. The actual differences in proportions at the monthly level are displayed as small dots, and one can see that the splines do a good job of approximating the trends for many precincts, but for some the fit of the splines appears to be inadequate. The grey guideline is set at January in 2012, just before stops started to dramatically decrease.

Given the large sample size, most differences are statistically significant at the 99% level. However, the majority of precincts do not show any substantively interesting differences but instead appear near zero and flat over the study period. Several precincts do show several large increases in the proportion of stops relative to crime, such as precinct 49 in the Bronx (Blue) and precinct 106 in Queens (Purple), although the increase did not appear immediately in 2012. Several precincts show decreases as well; in particular, many of the precincts in Brooklyn (Green) appear to have decreased post 2012. Precinct 121 in Staten Island was not formed until July, 2013 (Staten Island Advance Editorial, 2013), so its line is perfectly flat until that time period. Because it formed from parts of prior precincts 120 and 122, one might expect the relative share of stops in the precincts to decrease, but the opposite appeared to be the case, and precincts 120 and 122 saw their share increase, and precinct 123 stayed quite flat.

[INSERT FIGURE 2 HERE AND FIGURE 3 HERE]

Any increase or decrease could be the result of proportional changes in crimes, stops, or both simultaneously. The next step is to focus on the specific precincts where substantively interesting differences are found. For example, Figure 3 displays the proportion of violent crimes and stops in precinct 106 in Queens, which saw the most dramatic increase in the proportion of stops relative to crimes post 2012 and through 2016. The trend is entirely due to the increase in the proportion of stops: the violent crime share in the precinct has consistently been within one to two percent over the entire study period. The increase also did not occur until starting around 2013, and has continued to climb over time.

The increase in the *proportion* of stops in this precinct is actually not due to an increasing *count* of stops: in 2010 there were over 1,000 stops per month as compared to under 100 stops per month in 2016. Instead, the *relative share* of stops increased dramatically, suggesting the decrease in stops in Precinct 106 was not as dramatic as other places. Take for example Precinct 73 in Brooklyn, which has a larger share of violent crime than stops currently. In 2011 and 2012, Precinct 73 had over 2,000 stops per month, but in 2016 only had 25 or fewer stops per month.

Discussion

In this article we illustrate two new ways to compare point patterns for a chosen areal unit. First, we calculate intervals around the difference in percentage, and show how to appropriately map those results. This is an improvement over prior methods, as it allows one to focus on the magnitude of the differences. Using this technique, we show both areas where stops appear to be disproportionately targeted and areas where violent crimes have a higher share of the proportion. Large clusters of stops in north Manhattan, and a cluster of low stops relative to violent crime in the neighboring borough of the Bronx are one example of where NYPD resources may be reasonably redistributed.

Prior research focused on regression analysis identifying citywide metrics of whether stops were disproportionately targeted at minority neighborhoods (Fagan & Davies, 2000; Gelman et al., 2007), whereas our analyses identify particular spatial areas where stops are greater or lesser than one would expect under a model where stops are in a direct ratio to violent crime. Those areas identified are not uniformly in neighborhoods of a particular racial category. For example, there were areas of both high stops and low stops in Brooklyn spatially adjacent, which are among mixed racial neighborhoods of White, Hispanic, and Asian residents.⁹ Such an analysis provides a more in-depth look at the geography of stops and crime, without painting a simple picture of whether the NYPD disproportionately targeted stops in minority neighborhoods, as one can find examples across the city of where minority neighborhoods had too few of stops as well as too many relative to reported violent crime.

The second example uses multinomial logistic regression to examine the differences in stops and violent crimes over time. After identifying particular areas that deserve further scrutiny, additional graphs examine the changes in either a single point pattern or two point patterns over time. This analysis illustrates that before stops were dramatically reduced in 2012, precinct level stops and violent crime tended to be quite even and stable, although there were several precincts in Brooklyn that had a high share of stops relative to violent crime. After stops dramatically decreased, there appeared more volatile patterns in the percentage of stops relative to crime. Such volatility was mostly driven by changes in the relative distribution of stops. Although the dramatic reduction of stops likely reduced the absolute amount of interaction individuals in particular areas had with police, the dramatic reduction did not result in a more equitable spatial distribution of where stops took place at the precinct level.

We discuss a number of limitations of the proposed methods (that also apply to Andresen's original SPPT). In terms of evaluating whether the NYPD 'stops and questions' disproportionately in certain locations, one needs to keep in mind that different forces impact the distribution of stops. While we focus on reported Part 1 violent crimes in this analysis, the NYPD does not necessarily devote all of its resources to targeting only violent crime, but can focus on other serious crimes, like burglary (Kelling & Sousa, 2001).

⁹ To examine the racial distribution within New York City see an interactive dot density map of different races by the New York Times, *Mapping Segregation*, <https://www.nytimes.com/interactive/2015/07/08/us/census-race-map.html> (Bloch et al., 2015).

An important presumption behind the analysis is whether a direct ratio model, e.g. when a location has 1% of the crime it should have 1% of stops, is reasonable in and of itself. Prior analyses claiming the NYPD stopped individuals disproportionately in minority neighborhoods implicitly made this assumption, but it is at odds with hot spot policing, a NYPD tactic. It may be the case that one should not focus on the relative disparity in police resources, but the absolute number of police stops of minorities (Manski & Nagin, 2017).

For the differences in proportion test, while we used an adjusted Chi-Square version here, there are many potential options, including bootstrapping both point patterns (instead of just one, as in the original SPPT), or using other differences in proportions intervals. While we relied on past simulation studies of simpler two-by-two tables to choose the adjusted Chi-Square test, more detailed simulation studies are needed to show how well any of these techniques will perform in the case of doing many comparisons of proportions.

A second aspect that needs to be considered is the trade-off in power versus examining smaller areas. Prior analyses using the SPPT have examined very small areas, such as street segments, but it is not clear whether so many comparisons will produce a reasonable number of false positives and false negatives given the low number of observed points within each area. Future work should address how small a grid an analyst can reasonably assess and still maintain a particular power to detect differences in the two point patterns while still limiting false positives.

The multinomial model is presented as an extension to the SPPT to examine more complicated situations, such as changes over time or controlling for different area characteristics. However, the multinomial model is difficult to estimate for many areas, so it is currently limited to only analyzing larger administrative units for most study designs. Although, again, we cannot explicitly say what are too many areas to make the approach infeasible. Given that it is a generalized linear model, one needs to also rely on asymptotic behavior to estimate unbiased effects and standard errors around those effects. Finally, the multinomial model is only identified in the case that each area has more than zero incidents in either point pattern, although this assumption can be relaxed by using a regularized approach to estimating the multinomial regression (e.g. Kosmidis, 2017).

In both the differences in proportions test as well as the multinomial models (and the original SPPT), the spatial relationship between areas are not taken into account explicitly. That is, the tests do not take into account the proportions of the point patterns in neighboring areas, but treat each area as an island. In the case of the proportions test, one may wish to spatially smooth the differences, to limit the impact of areas with small baseline rates. Additionally one may be able to provide a spatially lagged version of the Chi-Square test in the differences in proportions (Rogerson, 2010).

Notwithstanding these limitations, all of which also apply to Andresen's SPPT, we believe we have made significant strides in the analysis and the presentation of spatial point pattern comparisons. Researchers should emphasize the size of the differences in any maps and graphs presented, and not rely simply on a dichotomous decision of different or similar. Using multinomial logistic regression presents opportunities to examine more complicated patterns over time or control for other characteristics, and so is a substantial improvement over the traditional SPPT that has been applied in past research.

References

- Agresti, A. (2002). *Categorical Data Analysis*. 2nd Ed. John Wiley & Sons: Hoboken, NJ.
- Agresti, A. (2003). Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research*, 12, 3-21.
- Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54, 280-288.
- Andresen, M.A. (2009). Testing for similarity in area-based spatial-patterns: A nonparametric Monte Carlo approach. *Applied Geography* 29, 333-345.
- Andresen, M.A. (2013). Crime seasonality and its variations across space. *Applied Geography* 43, 25-35.
- Andresen, M.A. (2016). An area-based nonparametric spatial point pattern test: the test, its applications, and the future. *Methodological Innovations*, 9, 1-11.
- Andresen, M.A., and Linning, S.J. (2012). The (in)appropriateness of aggregating across crime types. *Applied Geography*, 35, 275-282.
- Andresen, M.A., Linning, S.J., & Malleson, N. (2017). Crime at places and spatial concentrations: Exploring the spatial stability of property crime in Vancouver, B.C., 2003-2013. *Journal of Quantitative Criminology*, 33, 255-275.
- Andresen, M.A., & Malleson, N. (2011). Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime & Delinquency*, 48, 58-82.
- Andresen, M.A., & Malleson, N. (2013). Crime seasonality and its variations across space. *Applied Geography*, 43, 25-35.
- Andresen, M.A., & Malleson, N. (2014). Police foot patrol and crime displacement: A local analysis. *Journal of Contemporary Criminal Justice*, 30, 186-199.
- Andresen, M.A., & Malleson, N. (2015). Intra-week spatial-temporal patterns of crime. *Crime Science*, 4, 12.

- Bernasco, W., & Block, R.L. (2010). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime & Delinquency*, 48, 33-57.
- Bloch, M., Cox, A., & Giratikanon. (2015). Mapping segregation: New government rules will require all cities and towns receiving federal housing funds to assess patterns of segregation. *New York Times*. Available at <https://www.nytimes.com/interactive/2015/07/08/us/census-race-map.html>.
- Brunsdon, C., & Charlton, M. (2011). An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning B: Planning and Design*, 38, 216-230.
- Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661-3675.
- Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8, 551-561.
- Fagan, J., & Davies, G. (2000). Street stops and broken windows: Terry, race, and disorder in New York City. *Fordham Urban Law Journal*, 28, 457-504.
- Ferrandino, J. (2015). Minority threat hypothesis and NYPD stop and frisk policy. *Criminal Justice Review*, 40, 209-229.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Oliver and Boyd: Edinburgh
- Fox, J., & Andersen, R. (2006). Effect displays for multinomial and proportional-odds logit models. *Sociological Methodology*, 36, 225-255.
- Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the New York City Police Departments "Stop-and-Frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102, 813-823.
- Gelman, A., & Price, P.N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine*, 18, 3221-3234.
- Greenberg, D.F. (2014). Studying New York City's crime decline: Methodological issues. *Justice Quarterly*, 31, 154-188.
- Harrell, F. (2001). *Regression modelling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer-Verlag: New York, NY.

- Hibdon, J., Telep, C., & Groff, E.R. (2017). The concentration and stability of drug activity in Seattle, Washington using police and emergency medical services data. *Journal of Quantitative Criminology*, 33, 497-517.
- Hodgkinson, T., Andresen, M.A., & Farrell, G. (2016). The decline and locational shift of automotive theft: A local level analysis. *Journal of Criminal Justice*, 44, 49-57.
- Kelling, G.L., & Coles, C.M. (1997). *Fixing broken windows: Restoring order and reducing crime in our communities*. The Free Press: New York, NY.
- Kelling, G.L., & Sousa, W.H. (2001). *Do police matter? An analysis of the impact of New York City's police reforms*. *The Manhattan Institution*. Available at https://www.manhattan-institute.org/pdf/cr_22.pdf.
- Kosmidis, I. (2017). brglm2: Bias reduction in generalized linear models. *R package version 0.1.5*, <https://github.com/ikosmidis/brglm2>.
- Levine, N., & Kim, K.E. (1998). The location of motor vehicle crashes in Honolulu: A methodology for geocoding intersections. *Computers, Environment and Urban Systems*, 22, 557-576.
- Levitt, L. (2018). The mysteries of crime. *NYPD Confidential: An Inside Look at the New York Police Department*. Available at <http://nypdconfidential.com/columns/2018/180101.html>.
- Linning, S. (2015). Crime seasonality and the micro-spatial patterns of property crime in Vancouver, BC and Ottawa, ON. *Journal of Criminal Justice*, 43, 544-555.
- MacDonald, J, Fagan, J, & Geller, A. (2016). The effects of local police surges on crime and arrests in New York City. *PLoS ONE*, 11, e0157223.
- Manski, C.F., & Nagin, D.S. (2017). Assessing benefits, costs, and disparate racial impacts of confrontational proactive policing. *Proceedings of the National Academy of Sciences*, 114, 9308-9313.
- McNeill, G., & Hale, S.A. (2017). Generating tile maps. *Computer Graphics Forum*, 36, 435-445.
- Melo, S.N., Matias, L.F., & Andresen, M.A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography*, 62, 314-324.
- Mennis, J. (2006). Mapping the results of geographically weighted regression. *The Cartographic Journal*, 42, 171-179.

- Pereira, D.V.S., Mota, C.M.M., & Andresen, M.A. (2016). The homicide drop in Recife, Brazil: A study of crime concentrations and spatial patterns. *Homicide Studies*, 21, 21 - 38.
- Reeve, R. (2016). Confidence interval of difference of proportions in logistic regression in presence of covariates. *Statistical Methods in Medical Research*, Online First.
- Rogerson, P.A. (2010). The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic. *Geographical Analysis*, 31, 130-147.
- Rosenfeld, R., & Fornango, R. (2014). The impact of police stops on precinct robbery and burglary rates in New York City, 2003-2010. *Justice Quarterly*, 31, 96-122.
- Rosenfeld, R., & Fornango, R. (2017). The relationship between crime and stop, question, and frisk rates in New York City neighborhoods. *Justice Quarterly*, 34, 931-951.
- Rosenfeld, R., Fornango, R., & Rengifo, A.F. (2007). The impact of order-maintenance policing on New York City homicide and robbery rates: 1988-2001. *Criminology*, 45, 355-384.
- Roth, R.E., Woodruff, A.W., & Johnson, Z.F. (2010). Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal*, 47, 130-140.
- Silverman, E.B. (1999). *NYPD battles crime: Innovative strategies in policing*. Northeastern University Press: Boston, MA.
- Singleton, A., & Brunson, C. (2014). Escaping the pushpin paradigm in geographic information science: (re)presenting national crime data. *Area*, 46, 294-304.
- Smith, D.C., & Purtell, R. (2007). *An empirical assessment of NYPD's "Operation Impact": A targeted zone crime reduction strategy*. NYU Wagner Graduate School of Public Service Paper. Available at <https://wagner.nyu.edu/node/1396>.
- Staten Island Advance Editorial (2013). A welcome newcomer: NYPD's new 121st precinct on Staten Island is a change for the better. *Silive.com*. Available at http://www.silive.com/opinion/index.ssf/2013/11/a_welcome_newcomer_nypds_new_p.html.
- Steenbeek, W., Vandeviver, C., Andresen, M., & Malleson, N. (2017). *sppt: Spatial Point Pattern Test*. R package version 0.1.4. URL: <https://github.com/wsteenbeek/sppt>
- Sullivan, C.M., & O'Keeffe, Z.P. (2017). Evidence that curtailing proactive policing can reduce major crime. *Nature Human Behaviour*, 1, 730-737.

- Sweeten, Gary. (2016). What works, what doesn't, what's constitutional? *Criminology & Public Policy*, 15, 67-73.
- Thompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2015). UK open source crime data: Accuracy and possibilities for research. *Cartography and Geographic Information Science*, 42, 97-111.
- Vandeviver, C., & Steenbeek, W. (2017). The (in)stability of residential burglary patterns on street segments: The case of Antwerp, Belgium 2005-2016. *Journal of Quantitative Criminology*, Online First.
- Vaughan, A., Hewitt, A.N., Andresen, M.A., Brantingham, P.L. (2016). Exploring the role of the environmental context in the spatial distribution of calls-for-service associated with emotionally disturbed persons. *Policing: A Journal of Policy and Practice*, 10, 121-133.
- Wang, W. (2013). A note on the bootstrap confidence intervals for proportions. *Statistics & Probability Letters*, 83, 2699-2702.
- Weisburd, D., Telep, C.W., & Lawton, B.A. (2014). Could innovations in policing have contributed to the New York City crime drop even in a period of declining police strength?: The case of stop, question and frisk as a hot spots policing strategy. *Justice Quarterly*, 31, 129-153.
- Weisburd, D., Wooditch, A., Weisburd, S., & Yang, S.M. (2016). Do stop, question, and frisk practices deter crime? *Criminology & Public Policy*, 15, 31-56.
- Wheeler, A.P. (2017). The effect of 311 calls for service on crime in D.C. at microplaces. *Crime & Delinquency*, Online First.
- Wheeler, A.P., Worden, R.E., & McLean, S.J. (2016). Replicating group-based trajectory models of crime at micro-places in Albany, NY. *Journal of Quantitative Criminology*, 32, 589-612.
- Wickham, H., Hoffman, H., Wickham, C., & Cook, D. (2012). Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, 23, 382-393.
- Wooditch, A., & Weisburd, D. (2016). Using space-time analysis to evaluate criminal justice programs: An application to stop-question-frisk practices. *Journal of Quantitative Criminology*, 32, 191-213.
- Wu, Xiaoyun, & Lum, C. (2017). Measuring the spatial and temporal patterns of police proactivity. *Journal of Quantitative Criminology*, 33, 915-934.

Table 1: Example of Stacking the Dataset to Estimate Multinomial Model

| <i>Precinct</i> | <i>Violent</i> | <i>Total Incidents</i> |
|-----------------|----------------|------------------------|
| 30 | 0 | 44,687 |
| 73 | 0 | 171,983 |
| 106 | 0 | 52,058 |
| 30 | 1 | 4,699 |
| 73 | 1 | 12,113 |
| 106 | 1 | 5,369 |

Table 2: Multinomial Logistic Coefficients Predicting the Precinct of Stop or Violent Crime

| | PCT = 73 | | PCT = 106 | |
|------------------|--------------------|-----------------------|--------------------|-----------------------|
| | <i>Coefficient</i> | <i>Standard Error</i> | <i>Coefficient</i> | <i>Standard Error</i> |
| <i>Intercept</i> | 1.348 | 0.005 | 0.153 | 0.006 |
| <i>Violent</i> | -0.401 | 0.018 | -0.019 | 0.021 |

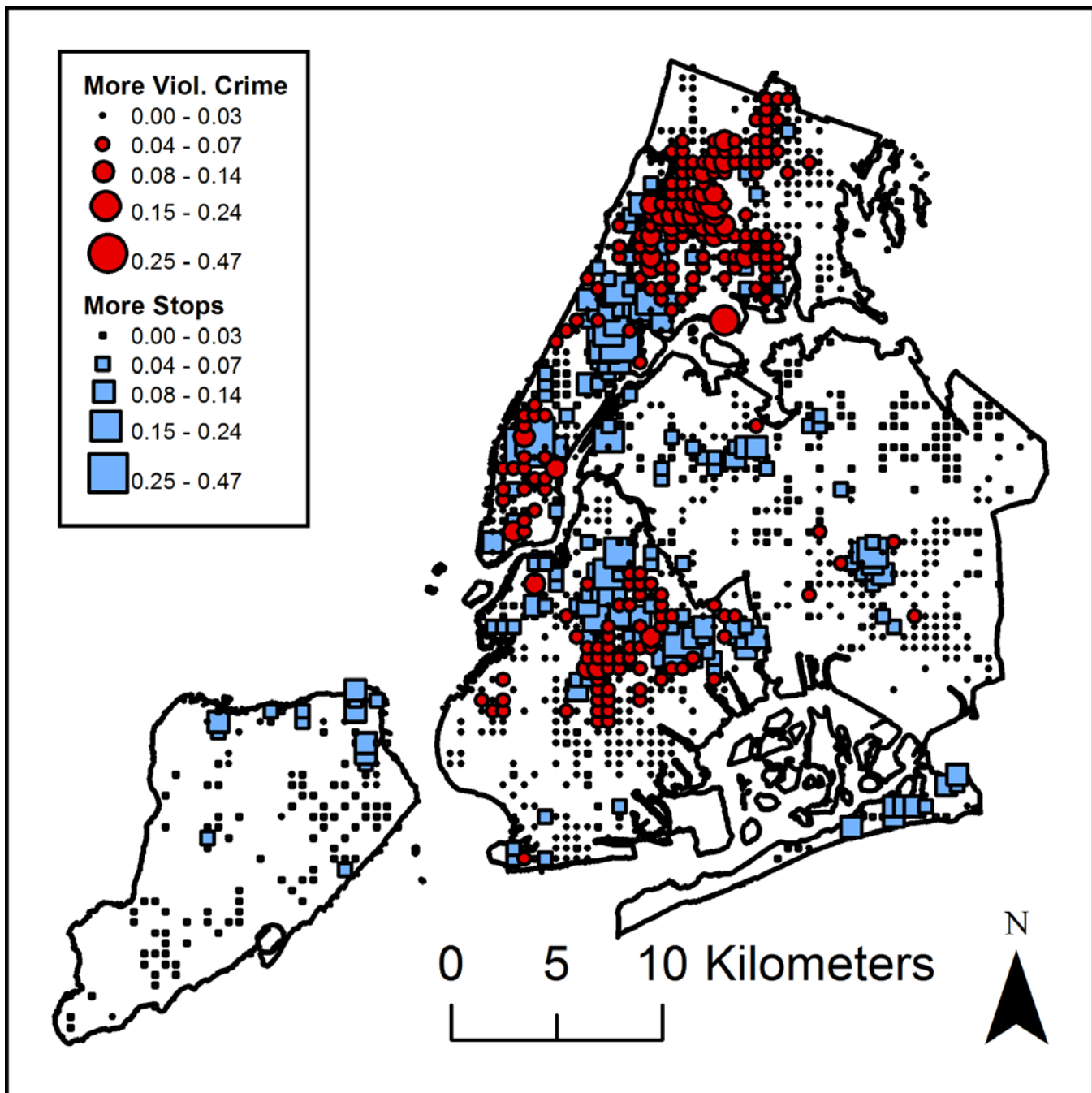


Figure 1: Grid cells that have a statistically significant difference in the proportion of stops and violent crimes at the $p < 0.05$ level after adjusting for multiple comparisons. Graduated symbols are displayed according to the difference in the percentage of stops and the percentage of violent crimes.

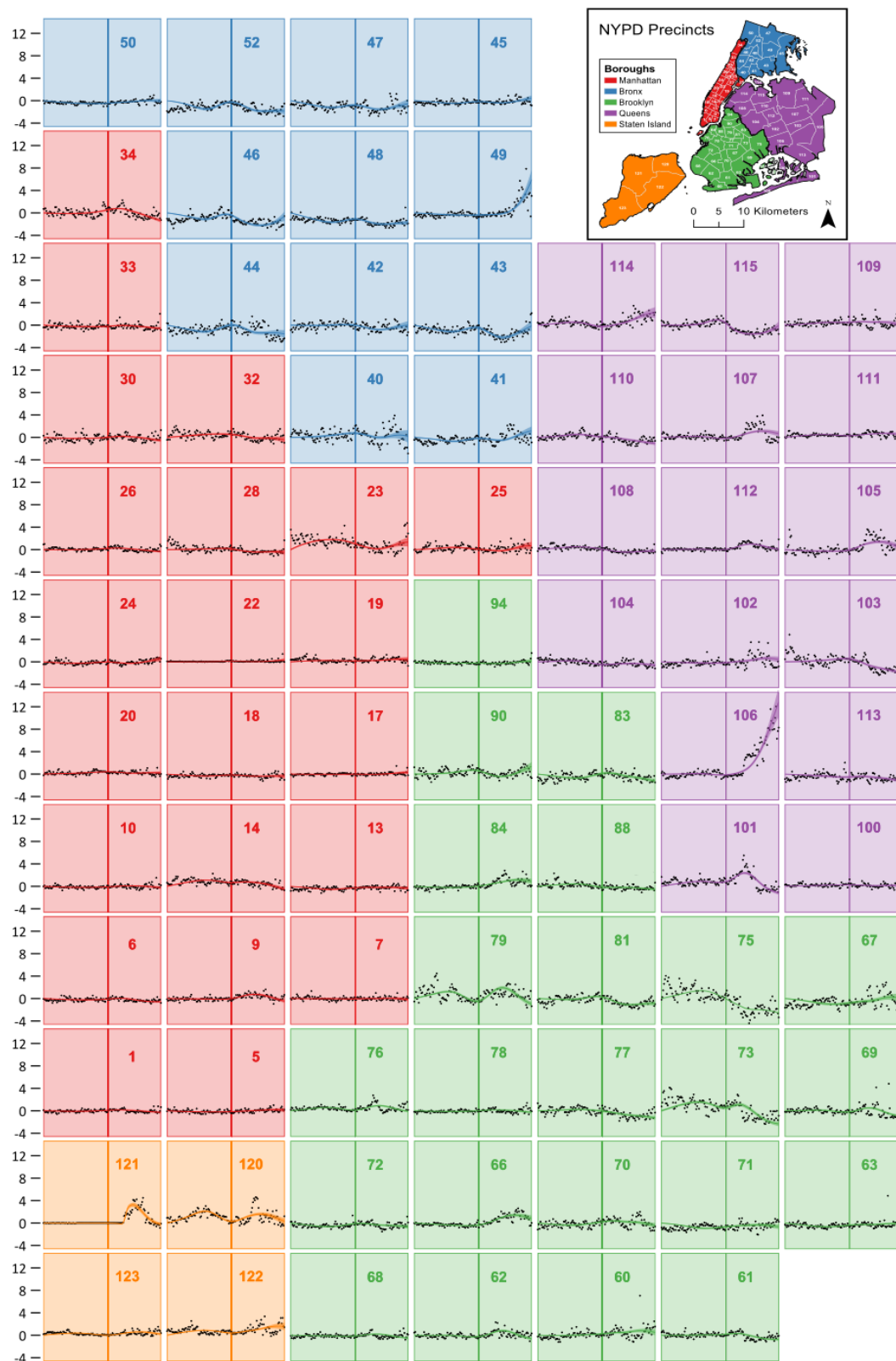


Figure 2: Estimates of the differences in the percentage of pedestrian stops minus the percentage of violent crimes from 2006 through 2016. Dots indicate the observed monthly differences in the percentage, and the colored line and error bars are the estimates and 99% confidence interval of the difference via multinomial logistic regression. The vertical grey bar represents the month of January, 2012. The precinct number is labelled in the upper portion of the graph, and the color represents the borough. A vector copy of the image can be seen at https://www.dropbox.com/s/4ow3p9lo41fpilc/TileMap_Vector.pdf?dl=0.

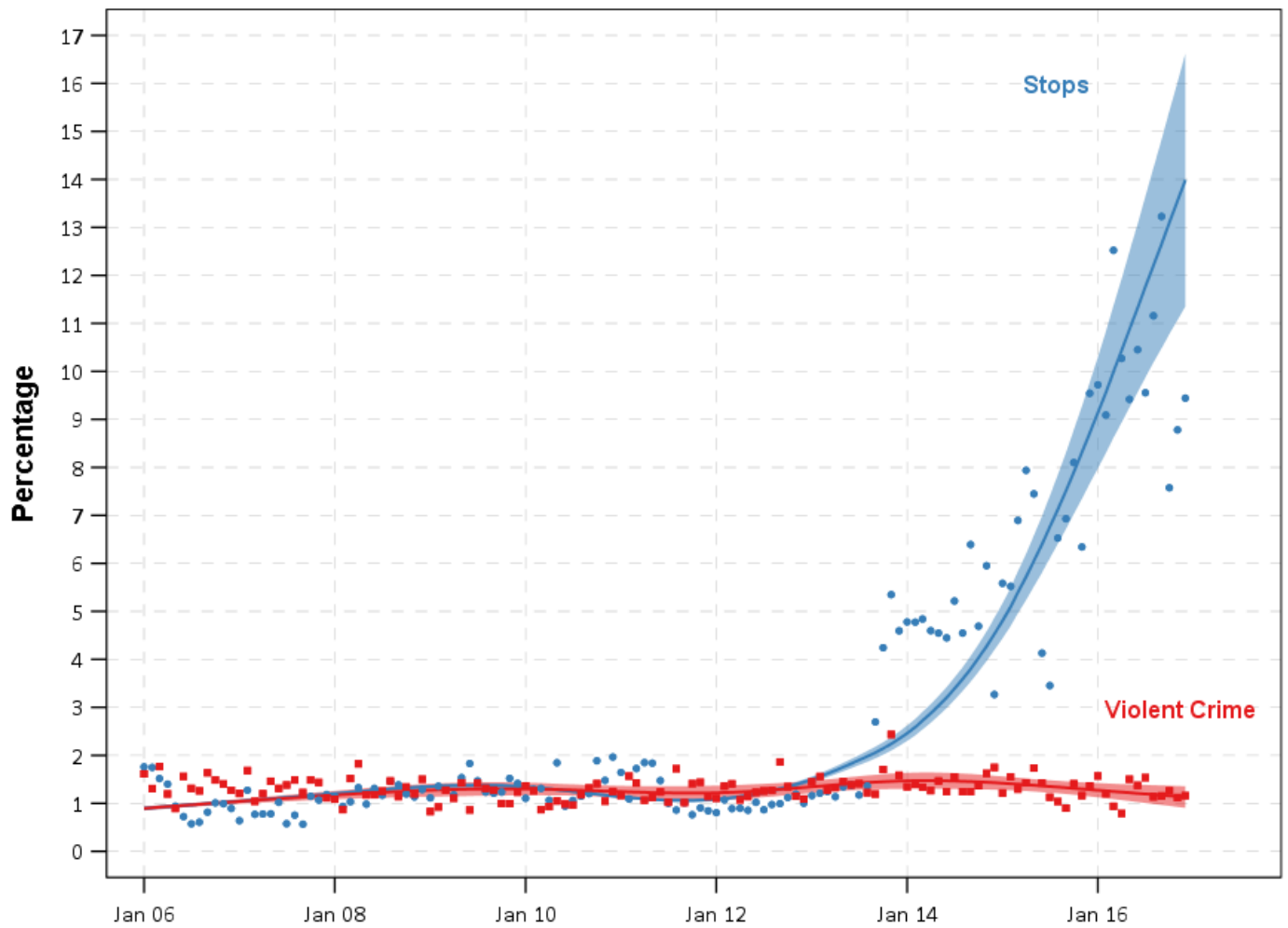


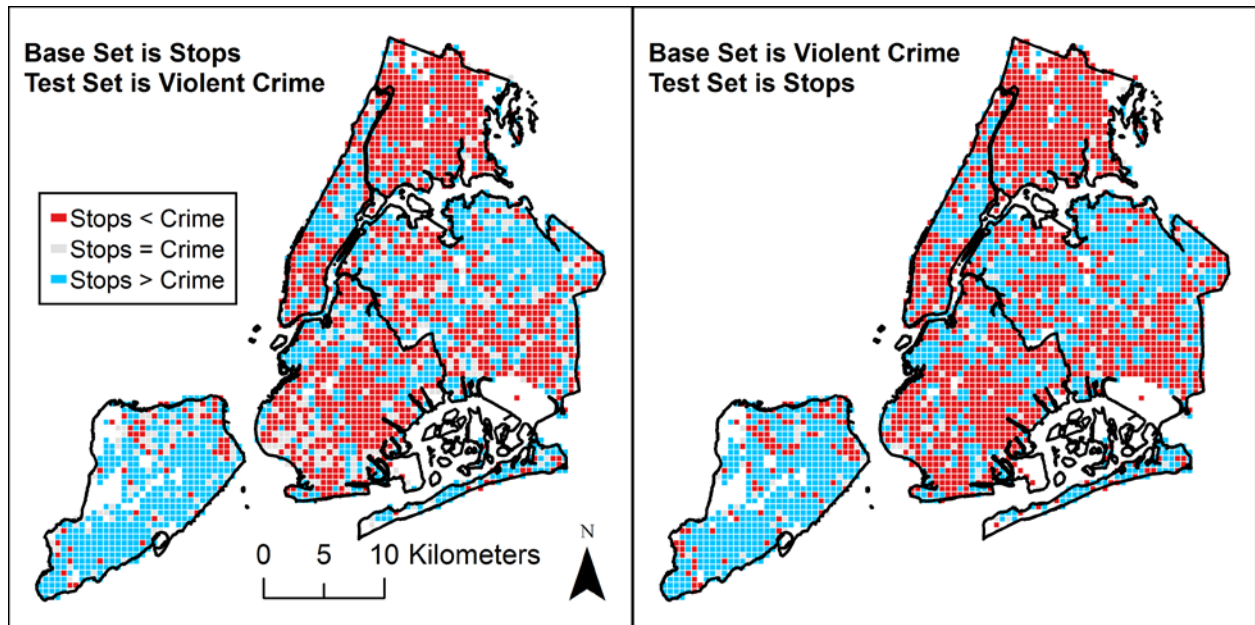
Figure 3: Change in the proportion of violent crime and the proportion of stops occurring in Precinct 106. Points are the observed proportions at the monthly level, the lines are the predicted proportions via the multinomial logistic regression model, and the area around the lines are the 99% confidence interval for the predicted percent.

Appendix A: Comparing our Results to the Default SPPT Procedure

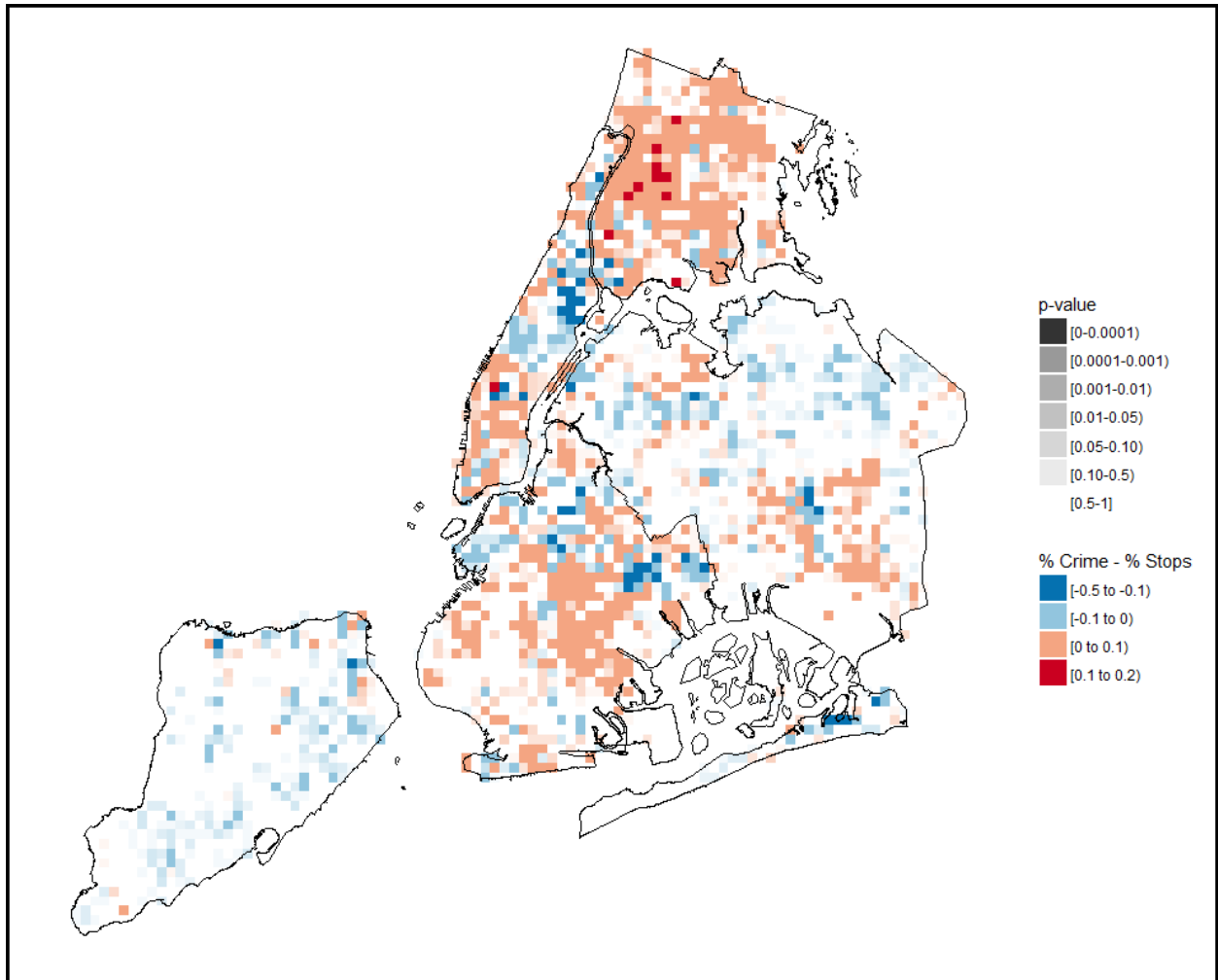
This appendix contains maps of the results of the original SPPT procedure (Appendix Figure 1), as well as an example value-by-alpha map of the results of our suggested difference in proportions Chi-square test (Appendix Figure 2). The original SPPT procedure used a 85% resampling rate, as well as a 95% confidence interval. We calculate the exact interval based on the hypergeometric distribution, so it is equivalent to doing an infinite number of Monte Carlo estimates.

In figure 1 you can see that each map is mostly filled in, whether one is using stops or violent crimes as the base set. Given that stops are more numerous, there are slightly more observations noted as statistically different, but in each map the overall S index is quite low, with 2,539 and 2,799 areas as different (out of 2,976 areas) for violent crime and stops as base sets respectively, for S indexes of 0.15 and 0.06. As one can see, with large sample sizes this provides very little discrimination in the map.

Compare this to the value-by-alpha map in Appendix Figure 2, that both highlights the size of the proportion difference, as well as mutes areas that have higher p-values (that have also been corrected for multiple comparisons). This map highlights similar areas compared to the varying glyph sized map shown in the paper (Figure 1), and one can identify the Bronx as having too few of stops and upper Manhattan as having too many stops. While there are pockets in Brooklyn that appear to have too many stops, the majority appears to have too few of stops. Queens and Staten Island are mostly muted in this particular map.



Appendix Figure 1: Output from traditional SPPT procedure.



Appendix Figure 2: Value by alpha map of the SPPT results using difference in proportions Chi-square test.