Walmart and Crime

Lander Basterra, Seth Green, Chris Patrick, Kevin Sun

Introduction

The impact of Walmart on the U.S. economy is difficult to overestimate. The company's first quarter U.S. revenue for 2016 was a reported $115.9 billion, and according to *Business Insider*, Walmart employed 1.4 million U.S. workers in 2010, representing 1% of the working population in the U.S. [1],[2]. Along with its influence in the marketplace, there are a host of societal ramifications of the company's culture, practices, and interactions with the communities containing its stores. For example, one study discovered a relationship between number of Walmart Supercenters with increased body mass index and obesity for a particular area [3]. Despite the fact that many of these factors have been extensively studied, there are still a multitude of unknown facets of what has been called the "Walmart effect" [4].

One aspect of Walmart's impact that should be better understood is whether it is associated with crime events in its communities; as a top employer and retail operation that brings many people together into a specific location, a relationship between Walmart and crime occurrences may exist. A recent high-profile article on Bloomberg Businessweek's site has called attention to what the authors describe as Walmart's "out-of-control" problem with crime, noting that in 2015, police were called to four Walmart locations in Tulsa, Oklahoma on almost 2,000 occasions, while being called to four Target locations only 300 times during that same period [5], and a 2014 study determined that the presence of Walmart in counties is associated with smaller reduction in crime rates for those counties as contrasted with counties with no Walmart [6]. However, formal analysis of crime rates at specific Walmart stores in comparison with crime rates of similar stores appears to be lacking and is needed to determine whether claims such as those made in the Bloomberg article are justified or hyperbolic.

Why should we care about this issue? The Bloomberg article suggests that crime at Walmart is becoming a drain on police departments around the country. The authors also make a strong claim that the problem of crime at Walmart is a direct consequence of heavy cost-cutting measures _including elimination of greeters at the front of the store_by leadership at the top levels of the organization; the retail space to worker ratio is now 524 square feet per employee, which represents "…a 19 percent increase in space per employee from a decade ago" [5]. Although Walmart contracts with a data analysis company named Cap Index to determine the relative risk of crime at each store using a number of variables, including "neighborhood demographics, local housing values, national and local crime statistics, and internal company records of theft," the company does not share this data, so local governments and police departments are missing important information that could help them determine how problematic specific stores are and perhaps redirect their limited resources accordingly [5]. Furthermore, there is some indication that when city leaders attempt to engage Walmart on the problem of crime at their stores, there can be recalcitrance on the part of the company to address the issue; the authors note the efforts of mayor Dennis Buckley of Beech Grove, Indiana, who used social and television media to put pressure on his local Walmart to follow through on changes they promised to make but were not delivering [5].

Normative Analysis

As noted above, our current understanding of crime associations for different retailers is primarily based on anecdotal reporting. The problem with these anecdotal reports is that they do not account for a number of factors that, if controlled for, might demonstrate that differences in crime rates between retailers are insignificant. For example, Walmart stores tend to be open much longer hours than other retailers; many of their stores are open 24 hours seven days a week. One possibility is that the amount of crime is simply a factor of how many hours each store is open. In addition, we might suspect that location of the store could be an important factor, perhaps even more important than whether the store is a Walmart or some other retailer. One of the key ideas from the groundbreaking work of Dr. John Snow in mapping the London cholera epidemic of 1854 is confirmation of a relatively simple assumption: people tend to go to the most convenient location to do what they need to do (in the case of Dr. Snow's analysis, this was the water pump closest to the person's house) [7]. People are most likely to shop at the stores most convenient for them, and this has implications for any analysis of retail stores and crime, because locations have a set of characteristics based on the people who live in those locations (e.g., income level, education level) and other features of the environment (e.g., access to public transit). Finally, each retail location has other important characteristics that may impact frequency of crime as well, such as number of parking spaces and square footage of retail space. Any analysis comparing crimes at retail locations should take as many of these factors into account as possible.

Stakeholders

The amount of effort already expended on improving the issue of crime in general is indicative of its importance to a number of different stakeholders. First and foremost, police departments are looking for the best guidance possible on how to allocate their limited resources, and a better understanding of whether crime is problematic for a particular retailer will be helpful to that end. In a similar vein, city government officials, such as the mayor mentioned above, need information on crime at locations in their jurisdictions in order to successfully partner with store owners on crime mitigation activities. Access to data on the relative risk of crime for specific retailers will also help city and county development and planning boards, developers, and shopping center owners make better informed strategic decisions related to higher-risk stores. The public at large is also a key stakeholder for several reasons: (1) citizens pay taxes to support the work of police and city government, and if a retailer is shifting the burden of crime management at its stores to police departments, the public is ultimately paying the bill; and (2) understanding whether crime is more likely to happen at one store versus another is useful in making decisions about where to shop. Last and certainly not least, retail owners themselves are key stakeholders. They should bear a great deal of the responsibility for ensuring that crimes are less likely to happen at their stores by budgeting appropriately and designing for crime prevention; the financial impact of crime on retailers is significant, and the possibility of negative publicity from articles such as the Bloomberg story are not helpful to retailers' bottom lines.

Impact

A successful analysis of crimes at various retailers should result in a reduction of the frequency of crimes, especially at problem retail locations. Informed analysis will provide city and police leadership with the information they need to impel owners of problematic stores to take appropriate action, such as increasing staffing for store security, and also to offer incentives to stores that demonstrate they are actively partnering with their communities on crime reduction. A reduction in crime at problematic stores will mean fewer calls to police and fewer taxpayer dollars being spent, and will allow police departments to direct their attention to other essential activities. Also, the number of people going through the judicial system and being incarcerated should be reduced, which will not only improve the lives of the people involved, but also reduce spending on public defense, trials, and prisons. Finally, having open access to the relevant data will allow the public to make more informed choices about where to shop.

Objectives

Our goals for this study are to:

- Unpack the relationship between Walmart and crime
- Compare crime at Walmart with crime at retail stores of similar type
- Understand how other variables such as income disparity, education level, and access to public transportation impact the frequency of crime at particular retail locations
- Determine if the amount of crime at various retail stores is primarily a feature of the hours that store is open
- Offer retail stores information on crime at their locations with the purpose of reducing crime
- Encourage retail stores to allocate appropriate resources for crime prevention
- Offer city governments and police departments information on how to better target their resources to crime at retail locations
- Reduce the financial impact to retailers, victims, police, and city governments
- Reduce time spent by police officers at problematic retail locations

- Prevent perpetrators from entering the justice system by helping to prevent crimes in the first place
- Reduce crime overall by increasing our understanding of related factors

Metrics

To determine the effectiveness of our approach, we propose measuring the following:

- Amount of crime at Walmart and other related retail store locations from month to month
- Number of calls to the police to retail store locations
- Monthly spending for police departments
- Numbers of security staff (measured as FTEs or full time equivalents) at problematic retail locations
- Store spending on crime prevention measures
- Number of theft and shoplifting cases entering the judicial system

Background

In order to investigate whether individual Walmart stores are associated with higher crime rates than their competitors, it is necessary to understand other factors that may be positively or negatively associated with crimes at a particular retail location. Every crime event has three features that will be important to consider: *who* was involved, *when* it occurred, and *where* it took place. In our literature review, we have broken down a sample of crime studies into these three broad categories. Although exploration of the timing of crime events will be outside the scope of our analysis of Walmart and crime, we include it in hopes that it might inform further analysis of the Walmart-crime relationship.

<u>Who</u>

Many factors contribute to who commits crime. Economists and crime researchers have tried to use economic and socioeconomic-demographic variables to try to capture this effect. Gumus found a strong correlation between cities that had a higher percentage of people under the poverty line and crime rate. His research confirmed previous empirical studies where income inequality variables were used [8]. In order to support our desired result of determining whether Walmart stores are associated with higher crime than their competitors, it is important to consider whether absolute income or relative income is a better predictor of criminal activity at store locations and then control for that variable when comparing crime rates at different stores.

Another popular predictor variable past researchers have analyzed was education level. Although Gumus found no statistical significant relationship between education level and crime [8], others have found not only a strong negative relationship between years of education and crime, but also evidence of a causal relationship between education level and crime [10]. Several theoretical reasons for expecting a relationship between education and crime are that schooling increases the economic return of legitimate work, increases cost of committing a crime, increases risk aversion, and increases the psychic cost of breaking the law [10].

In order to show that schooling had a causal relationship with crime, Moretti explored changes in state compulsory attendance laws; specifically, changes over time in the number of years of compulsory education that states mandated acted as an instrumental variable for education. OLS and 2SLS regressions were carried out and Moretti concluded that high school graduates had a lower probability of committing a crime compared to high school drop-outs. Interestingly, Moretti found that a one-year increase in schooling on average decreased murder and assault by 30%, arson by 13%, burglary and larceny by 6%, but had no effect on robberies and increased rape by 15%. Moretti also calculated that for every 1% increase in high school graduation rates, the total social savings would amount to $1.4 billion [10]. Based on these results, we will need to attempt to control for mean educational level of the area in which a store is based when comparing crime rates at different stores.

As mentioned before, there have been conflicting conclusions from various research literatures on crime. One possible explanation to the varying results may be due to the lack of consistent data used by the researchers. Gumus investigated 75 cities with population greater than 200,000 from the 12th edition of the *County and City Data Book*. However, bias might have been introduced due to the lack of consistency of data dates (due to difficulty of getting data), as all of his response variables were from 1991, but the socioeconomic-demographic variable data was from 1990 and economic data was from 1989 (except unemployment rate data which was from 1991) [8]. Furthermore, the research cited by both Gumus and Moretti was conducted from the mid to late 1900s, in which either the data quality or socioeconomic behavior might have changed since Gumus and Moretti [8],[10]. Moretti used data from the FBI Uniform Crime Report (UCR) as well as data from the National Longitudinal Survey of Youth (NLSY) [10]. Data from the NLSY is self-reported and thus may not capture all crimes that were committed. Thus, the reason the problem of identifying *who* will commit crimes is still unsolved may be due to the nature of how crime data is reported and the availability of the data throughout the decades. A representative sample of data sources, data mining methods, and some of the evaluation methods and their outcomes, for studies on *who* commits crimes, is summarized in the table below:

Data sources	
Gumus	Moretti
Urban Crime: *County and City Data Book*	Education level and crime: FBI Uniform Crime Report, National Longitudinal Survey of Youth
Data mining methods	
Ordinary Least Squares Regression	Ordinary Least Squares Regression, 2SLS Regression
Evaluation methods and outcomes	
Used both crime and crime rate, as well as their respective log forms as separate response variables to verify the power of statistically significant predictor variables. Result was that income equality, population, and per capita income had high power and percentage of black population had low power but was statistically significant in some of the models.	Used the IV variable *compulsory schooling* for education. Used 2SLS regression along with the IV variable to verify if the result of schooling was different than their OLS regression without the IV variable. 2SLS results showed that schooling was indeed statistically significant and signs matched the coefficients in the OLS regression.

<u>When</u>

There is general agreement in conducted research that climate time series is highly correlated to crime. Two of the most common findings are that property crimes such as burglary, robbery, and theft, the types of crimes we might expect to see more frequently in retail, are more predominant during fall and winter and violent crimes peak during summer and drop during winter [14]. One of the most researched hypotheses dealing with crime seasonality is the temperature aggression hypothesis, which states that violent crimes increase as temperature rises and drops when temperature gets around 90 degrees Fahrenheit [15].

However, surrounding events and other factors could be contributing to these findings. Hipp et al. also stated that seasonal crime opportunities are concentrated in time and space, mentioning three conditions that could contribute to violent crimes: (1) motivated offenders, (2) suitable targets, and (3) the absence of a capable guardian [14]. Some researchers have continued studying seasonal crime under the hypothesis that different offenses yield distinct seasonal patterns and cities with different climates would generate different seasonal crime trends [16]. In their research, which was focused on Vancouver and Ottawa, they found that seasonal crime patterns vary depending on the crime type, and hence, demonstrated the importance of disaggregating data when looking at crime in general. Also, Andresen and Malleson investigated days of the week as a key factor in seasonal crime patterns for each type of crime [17]. They found that criminal events are more common during the week than on the weekend with the beginning of the week having a greater proportion than the rest of the week, while assaults are more common during weekends. In general, Andresen and Malleson found that all crime types had distinctive temporal patterns as the week progressed [17]. The temperature aggression hypothesis has also continued to be a topic of interest among researchers since they want to study this correlation in different cities and see how the distribution of daily temperatures can affect this hypothesis. For example, Bushman et al. found that in certain cities such as Dallas, violent crimes do not drop when temperatures are greater than 90 degrees during nighttime hours; in Minneapolis the drop was also non-existent as temperatures got closer to 90 degrees, but as the authors note this information is less useful for Minneapolis because of the relative dearth of observations for the temperature reaching 90 degrees [18].

Seasonal crime could also be affected by calendar events. Instead of looking at temporal trends, a 2003 study looked at specific days of the year that correlate with more crime. In examining major and minor holidays, they found that “both violent and property crimes were significantly related to major (or legal) holidays, whereas neither type of crime was more likely to occur on minor holidays. Crimes of expressive violence were significantly more prevalent on major holidays, whereas property crimes were less frequent on those days” [19]. Their hypothesis is founded on routine activity (RA) theory, which postulates that “changes in routine or typical activities increase the probability that individuals will be vulnerable to certain types of criminal victimization” [19].

The study used was fairly thorough in its use of dummy variables to attempt to control for other factors such as temperature, day of week, and whether the day was close to the first of the month (typically payday). In all 67 dummy variables were included in the model for purposes of controlling confounding factors. Many of the effects they observed attained significance only when adding those controls (particularly for weather and time of year) into the model [19].

Overall, this study faulted its predecessors for *not* distinguishing between different types of holidays. That critique was ably dealt with, as Cohn and Rotton note that holidays where people typically gather in the home (Christmas, Thanksgiving, etc.) show very different correlations from holidays such as Independence Day where people typically gather in public [19]. As one might expect, Independence Day correlates with disorderly conduct charges, whereas Christmas correlates to an increase in domestic violence.

Cohn and Rotton claim to have confirmed their hypothesis though they acknowledge that “the social and cultural effects of holidays on crime were more complex than originally anticipated” [19]. One notable limitation of the study was that it used data from only one city (Minneapolis). They note that different holidays (St. Patrick's Day is used as an example) are celebrated differently, and to a far different scale, in different cities. The study also noted in many places that certain attributes of a holiday correlate to different kinds of crime. In terms of an analysis of Walmart and crime, or crime at retail locations in general, more research could be done on the impact of holidays on crimes at those stores. It would seem to be a given that crime would drop at stores if the stores are closed, but examining crime at stores that are open for specific holidays might offer some interesting insights. It would also be useful to pursue routine activity theory further as it relates to crime at Walmart and other similar stores. If the theory holds, it could be extremely useful to be able to predict situations where individuals have significant changes in their shopping routines and activities and attempt to set up some kind of safeguard to make them less susceptible as victims.

Although a large number of studies have been conducted to assess crime seasonality, predicting the time of a crime is still a hard problem, one that remains unsolved. One of the reasons could be that in some cases when offenses were evaluated, researchers often collapsed the crime types into an aggregate crime variable. We need to take into account that not all crimes, calendar events, and even cities_and particularly relevant for the purposes of this case study, not all stores_behave the same way. It would be better to look at each separately and use findings to arrive to a general consensus. As noted above, this line of inquiry is largely outside of the scope of this study. Among other reasons, this is because our data set encompasses only one calendar year, which makes it difficult to make any conclusions about seasonal or holiday-related effects. However, future inquiry into these issues could prove fruitful, if appropriate data could be collected.

Data sources		
Hipp et al.	Linning et al.	Cohn and Rotton
Crime data: The National Archive of Criminal Justice Data Weather data: The National Climate Data Center	Census records: Vancouver Census Metropolitan Area Crime data: Vancouver Police Department's Calls for Service Database	Data on calls for service in Minneapolis, Minnesota, in 1985, 1987, and 1988
Data mining methods		
Time Series: Latent Curve Model (LCM)	Regression Analysis: Ordinary Least Squares(OLS) and Negative Binomial	Multi-variable regression Dummy variable controls RA Theory
Evaluation methods and outcomes		
The (LCM) allows to model seasonal oscillations in crime for a large number of cities. Their approach implemented a nonlinear cosine function to capture the oscillatory patterns observed in crime rates over seasons. They chose this method because 1) employs a highly confirmatory factor for repeated measures 2) allows to predict that not all communities experience the same magnitude of seasonal crime oscillations.	Eleven distinct models were run on each of the independent variables (both aggregate and disaggregate) Initial tests checking for OLS assumptions were conducted and revealed that error terms in all models were not independently or identically distributed. Both autocorrelation and heteroscedasticity were present in the data. The use of robust standard errors most often lead to a smaller number of statistically significant variables in all of the models.	Examined correlation between specific holidays and categories of holiday and specific types of crimes. Found that major holidays correlate to an increase in many types of crime, especially when controls for temperature and time are considered.

Where

In examining the possible connection between Walmart and crime, we must understand both the current state of research on crime location as well as location-related attributes of crime events that we may need to control for in our analysis. Studies of geographical locations as they relate to crime events began as early as the 1800's, but police departments and researchers did not have a sustained focus on crime mapping until the late 1960's when computers could be engaged in automating the process [20]. The relatively recent advent of sophisticated modeling computing tools such as ArcGIS maximizes the opportunity to map crime to location, and as might be expected, there has been a shift in recent years from examination of the larger environment, such as the neighborhood, to a narrowed focus on crime events at places within that neighborhood, such as intersections, addresses, or individual businesses [21], [22]. For example, one recent study featured an analysis of impact on crime rates within 250 feet of individual foreclosed residential properties; the authors discovered a 19% increase in violent crimes for foreclosed properties that became vacant [23]. Although a number of different interesting analyses have been performed for crime frequency around different types of specific locations, such as liquor stores and drug treatment centers [22], [24], there are still many types of business properties, such as Walmart stores, and subsets of properties, such as "big box" stores in general, to be examined to enhance our ability to use location to predict and understand criminal activity. We also must seek a better understanding of the geographical context for crimes at specific locations; Deryol, et al. highlight the importance of other nearby destinations (nodes), travel routes (paths), and the broader environmental landscape in close proximity to crime hotspots (locations of high criminal activity) to provide contextual information for crime events [21].

Many factors contribute to whether or not a crime occurs at a specific location. Eck and Weisburd note the importance of controllers, or those whose presence and effectiveness can prevent a crime event. These controllers include: intimate handlers, or those who have influence over the offender, such as a parent or a teacher; place managers, responsible for oversight of a location, such as a building superintendent; and guardians, or those who protect the criminal targets, such as a friend or police officer [25]. As noted above, the greater geographical context for a location is also critical in influencing a crime event; for example, if one location is less proximal to a bus stop than another, it may be more difficult to reach by criminal offenders as well as targets and thus characterized by a lower frequency of crime. Critical infrastructure such as street lighting and public transportation and the socioeconomic status of the location's neighborhood will also be factors to consider when examining the likelihood of criminal activity at a particular location, and we will attempt to control for some of these factors when comparing crimes that occur at different store locations.

Data mining for geographic information related to crimes is aided by the number of cities that make data on criminal activity publicly available; many of these datasets are available in a single location at data.gov. These datasets typically include the category of offense, at least some type of location information, and the date of the event, as well as other markers. To protect the identity of victims, the exact address of the crime event is frequently masked and only the block-level information is provided. There do appear to be some cities that provide exact addresses at least for some crimes, and for this reason we will be working with crime data from Austin, Texas, which according to Ron MacKay of the Austin Police Department, only excludes addresses for sensitive crimes, such as rape and crimes against children [26]. Also of note, many cities use the Uniform Crime Report (UCR) format; one of the issues with this data is that if a single event involves multiple crimes, only the most serious crime will be listed. For example, if someone breaks into a house and murders an occupant, only the murder will be indicated [23]; this means that less serious crimes will tend to be underreported in the dataset. However, as we will be examining crimes that occur at retail locations, we expect this issue to be less problematic than might otherwise be the case if we were examining crime events more broadly.

A representative sample of data sources, data mining methods, and some of the evaluation methods and their outcomes for studies on crime places is summarized in the table below. Given the previous work on crime places, the problem remains unsolved because of the multitude of factors influencing whether a crime occurs at a particular location or not. Although we are better able to narrow our focus to the microenvironment with today's technology and statistical methods, a crime event is characterized by a confluence of factors and our models must attempt to represent this complexity. Because of this inherent complexity, it may be helpful to narrow in on a more specific set of locations, as is the intent of our investigation of crimes at Walmart and other related stores.

Data sources			
Cui and Walsh	Furr-Holden et al.	Deryol et al.	Han et al.
Crime data: Police Department of the City of Pittsburgh	Publicly funded outpatient drug treatment centers: Baltimore Substance Abuse Systems, Inc. (BSAS)	Commercial density: Cincinnati Area Geographic Information System's open website	Liquor stores: Pennsylvania Liquor Control Board (PLCB)
Data mining methods			
Classification (identifying different periods associated with foreclosures: pre-foreclosure, foreclosure, vacancy, reoccupation)	Negative binomial regression modeling	Poisson-based hierarchical regression analysis	Difference-in-difference-in-differences (triple difference) modeling
Evaluation methods and outcomes			
Tested for potentially treated controls by defining larger control areas. For violent crime, the resulting reduction in magnitude indicated that the effective treatment area is smaller than 500 feet, thus validating the original approach.	Reran regression models only using drug treatment centers, convenience stores, liquor stores, and corner stores that had no other businesses of a similar type within the buffer range. The results were similar to the original findings, thus validating the approach.	Assessed whether there was a relative improvement across different explanatory models by re-running the analysis within a linear model specification. The decrease in deviance statistics were only significant when moving toward the model looking at the additive effects of the variables under consideration; the decrease was not significant for a model looking at a product of the variable values.	Used a falsification model, arbitrarily examining a different day (Tuesday) than the target day (Sunday). The falsification model did not demonstrate a statistically significant increase in crime for the false variable value, thus validating the original model.

Hypothesis and Approach

Hypothesis

When other variables are controlled, being a Walmart is significantly correlated to a higher number of crimes committed at a given store, compared to other similar establishments.

Data

We focused on Austin to test our hypothesis. One of the main reasons was that the Austin Crime Database for 2015 includes addresses of crimes. This aspect was particularly important in our research since we wanted to get the specific location of crimes and not a radius around them. The latitude, longitude, store addresses and hours of operation used in our research were manually retrieved from Google Maps. We also included in our master dataset a count of the number of crimes at each address, as well as a measure for "normalized crime", which divided this count by the percentage of possible hours in a week that the store was open.

We used the United States Census Bureau to get demographic information about zip codes where these establishments are located, such as population, average income, median income, college graduation rate, high school graduation rate, unemployment, and percentage of families under poverty line. We also calculated a measure for income inequality by taking the difference between the income of the approximate 80th percentile of residents in a given zip code and the income of the approximate 20th percentile.

The dataset also includes the Kernel Density Estimation (KDE) density value that we calculated on the 2015 Austin Crime Database to predict crime density values across the city of Austin. For this KDE, we removed all crimes that occurred at the stores we were analyzing (2,675 crimes in total at 27 different stores) and then calculated the geographic crime density based on the remaining 35,306 in the 2015 database. This was done so that the KDE was an estimate of the density of *other* crimes in the area surrounding the stores, and was not biased by the crimes *at* the stores (since this is what we were attempting to predict). A KDE value for the location of each store was then considered as a predictor variable in our model. Fig. 1 shows a visualization of this KDE, as well as the locations of the stores in our dataset. The large teal circles indicate Walmarts, while the small diamonds indicate other similar stores.

Finally, we also included the distance from the closest bus stop to each store in our dataset. The inclusion of this measure was based on previous research that indicated infrastructural aspects having an impact on local crime rates [21]. Our objective was to find as much information related to the stores and location as possible in order to find a possible correlation which could prove or disprove our hypothesis.

The stores used in our research were Walmart, Target, Costco, Sam's Club, and Big Lots since all of them are comparable in size and service offered. In total, there were 27 of these stores within the geographical area covered by the Austin dataset. We selected these stores as being similar to Walmart by looking at the New York Stock Exchange Nasdaq Index page for Walmart's competitors [27] and then cross-referencing that list with all the stores included in Austin's Yellow Pages for the three categories in which Walmart is included ("Discount Stores," "Department Stores," and "General Merchandise") [28], [29], [30].

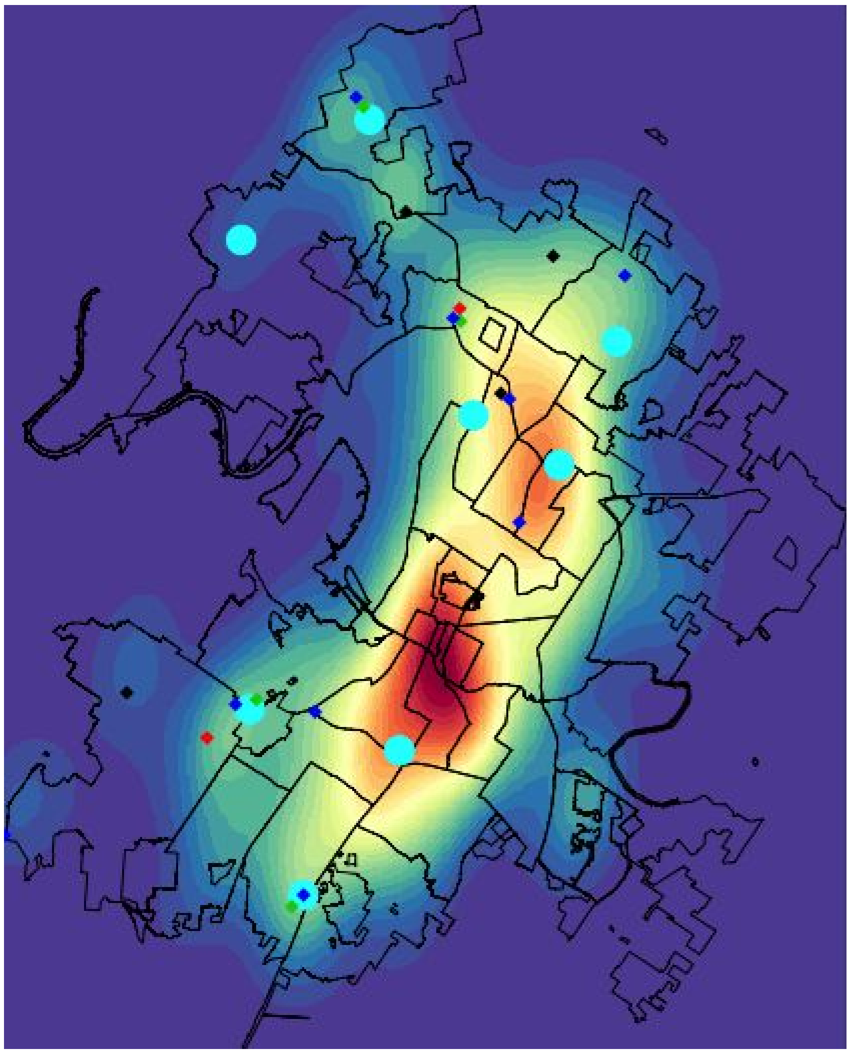

Fig.1

There is a probability that biases might exist in our data. For example, the Austin Police Department might have not recorded every single crime committed at these stores and stores might have not reported every crime to the police. Also, there is a possibility of human error at the time of computing the income inequality (20% to 80% range) for every single relevant zip code from the United States Census Bureau database or mistakes already present in this data.

Appendix A is a table summarizing all of the variables we considered in our analysis.

Methods and Evaluation

Our method was to attempt to isolate the effect of a store being a Walmart on the amount of crime at that store. To do so, we constructed linear models (using ordinary least squares) which predicted crime based on a number of factors. The purpose of our models was not explicitly to predict the number of crimes at a given store, but to give us an indication of which of the variables that we included in a given model had the most significant effect on the outcome. This method allowed us to control for a variety of different factors which had historically been correlated with crime rates. In all, we built over 30 different models with different combinations of these variables.

We used a number of methods to test the significance of different variables in our models, in order to better understand the explanatory effect that each was having on the amount of crime at a given store. First we applied rigorous Variance Inflation Factor tests to our model as a whole, as well as certain sub-groupings of variables, to be sure that the significance we were measuring was not being masked or distorted by multicollinearity.

In order to accurately select appropriate variables that may infer crime, two approaches were used: top-down and bottom-up approach. The top-down approach modeled every predictive variable against the response variable **crimes2015**. The first model contained every predictor variable and exhibited multicollinearity with **povertyline** having the highest VIF of 33. We removed **povertyline,** but the model still showed multicollinearity with **MedIncome** having the highest VIF of 7.34. By iterating this process of checking multicollinearity and removing the variable with the highest VIF of at least 5, we removed **CollegeGradPercent** as the last multicollinear variable. We arrived at a model (Appendix B) that contained the variables **Population, AvgIncome, IncomeInequality, Unemployment, KDEraw, closest_stops_in_meter,** and **wal** and had an adjusted R-square of 0.542. **Umemployment, IncomeInequality,** and **closest_stops_in_meter** had unexpected signs. **IncomeInequality** was calculated differently in this study compared to the method of using the Gini Index in other crime literary research, which might explain the unexpected sign. After taking **IncomeInequality** out, we arrived at a better model with an adjusted R-squared of 0.563, but still got unexpected signs for **unemployment** and **closest_stops_in_meter** (Appendix C). We found **wal** to have a positive coefficient and to be statistically significant at the $\alpha = 0.05$ level, and **population** to be slightly significant at the $\alpha = .05$ level.

Another way to select significant variables was from a bottom-up approach. Since there were many variables that captured the idea of income (**MedIncome, AverageIncome, IncomeInequality, Unemployment, PovertyLine),** we wanted to identify any multicollinearity among these variables and pick the best variables that had statistically significant associations with crime. After conducting separate models for each individual income-related variable to **crimes2015**, we found that **PovertyLine** was the most significant in predicting crime with a p-value = 0.009 and adjusted R-squared = 0.221. We also found that **unemployment** was not significant, and **IncomeInequality** had unexpected sign and was only significant at the $\alpha = 5\%$ level. The study then regressed all income variables to **crimes2015** and found multicollinearity among **povertyline, medIncome, and AvgIncome**. Because **medIncome** had the highest adjusted R-squared in the univariate model compared to **AvgIncome**, we decided to remove **AvgIncome** as a regressor. In addition, from basic statistical theory, **MedIncome** should be a more robust metric than **AvgIncome**. We also decided to remove **IncomeInequality** as a variable because of the way **IncomeInequality** was calculated as mentioned before.

We then modeled **MedIncome, Unemployment,** and **PovertyLine** to **crimes2015** and found **MedIncome** to have the highest VIF of 5.63480. After removing **MedIncome** from our model, we arrived at a model with no multicollinearity, and with an adjusted R-squared of 0.1817. This adjusted R-squared was lower than the univariate model that regressed **PovertyLine** with **crimes2015**, which had an adjusted R-squared of 0.2112. We then tested to see if adding another income variable to a model that already contained **PovertyLine** would result in a model with a higher Adjusted R-squared, which did not occur. Thus, we concluded to use only **PovertyLine** as the variable to represent **Income** in our model.

Moving forward, we regressed **PovertyLine** along with the other non-income variables on **crimes2015** to see if multicollinearity occurred. From the iterative analysis of removing the variable with the highest VIF, we removed **Percentage.hours.open.per.week, HighSchoolGradRate**, and **Hours.open.per.week**, resulting in our final model using the bottom-up approach (Appendix D). In this model, we found that **wal** had a positive coefficient and was statistically significant at the $\alpha = 0.05$ level and **population** to be weakly statistically significant at the $\alpha = .1$ level, which was similar to our findings using the top-down approach. Both of these systematic approaches show that a Walmart store is statistically significant in inferring crime when controlling for other factors.

Once we had eliminated variables with significant multicollinearity, we performed significance tests on a number of different candidate models. We looked for high adjusted R-squared values in potential models and then looked for variables with low multicollinearity and p-values, and coefficient signs matching our expectations. These variables could then safely said to be having a significant effect on the amount of crime at a given store, when controlling for the other variables included in that model.

We began by iteratively creating models including as many variables as we could without introducing multicollinearity. Next we built a model based on each variable of interest by itself or with only it and an indicator of whether or not the store was a Walmart. For each of these models, a raw count of crimes at that store in 2015 (**crimes2015**) was the response variable. A table of the statistics mentioned above for some of the more significant models is featured in Appendix E. We compared these statistics for each of the models and quickly began to notice some patterns.

In almost every model we ran, the variable indicating whether or not the store was a Walmart (**wal**) was the most significant, and had a fairly large positive correlation to the response. On models where three or more predictor variables were included, Walmart was almost always the *only* variable with a p-value less than 0.05.

We mentioned earlier how we had also calculated a measure for number of crimes divided by the percentage of hours per week that a store was open (**normalizedCrime**). This allowed us to control for the fact that most Walmarts are open 24 hours a day, whereas almost no competitors are. As previously mentioned, it was hypothesized that perhaps this was in some way explanatory of the higher incidence of crime at Walmart. However, when we repeated our analysis using our **normalizedCrime** count as the response variable, we noticed almost the exact same pattern. The most significant variable in every model we built was the indicator of whether or not the store was a Walmart (the **wal** variable). A sample of this is shown in Appendix E.

Based on these observations, and looking at the adjusted R-squared for several models, we picked three models as optimal models for further investigation:

Optimal Models

Response variable	**Predictor variables**	**Adj. R-squared**
crimes2015	KDEraw + wal	0.4706
crimes2015	Population + wal	0.5947
normalizedCrime	KDEraw + Population + wal	0.5565

In order to visualize the effect of **wal** in these models, we created a simulation data set by simply selecting one row from our original data (the entry for the Walmart at 2525 W Anderson Ln), copying it five times, and changing only the name of the store for each new row. This allowed us to compare predictions that the different models (trained on the original data) would make for these fictional stores. A good mental approximation of this simulation would be imagining a shopping center which featured all five candidate stores sharing a single parking lot. In that scenario, each predictor variable would have the exact same value (they would all be in the same zip code, and would have essentially equal KDE values), except for store name. When the predictions for each of the optimal models are plotted in Fig. 2, it is clear that each model predicts the Walmart to have significantly more crime than any of its competitors. In fact, it appears that there is a non-linear relationship, with a steep incline in crime for Walmarts as compared to its competitors.

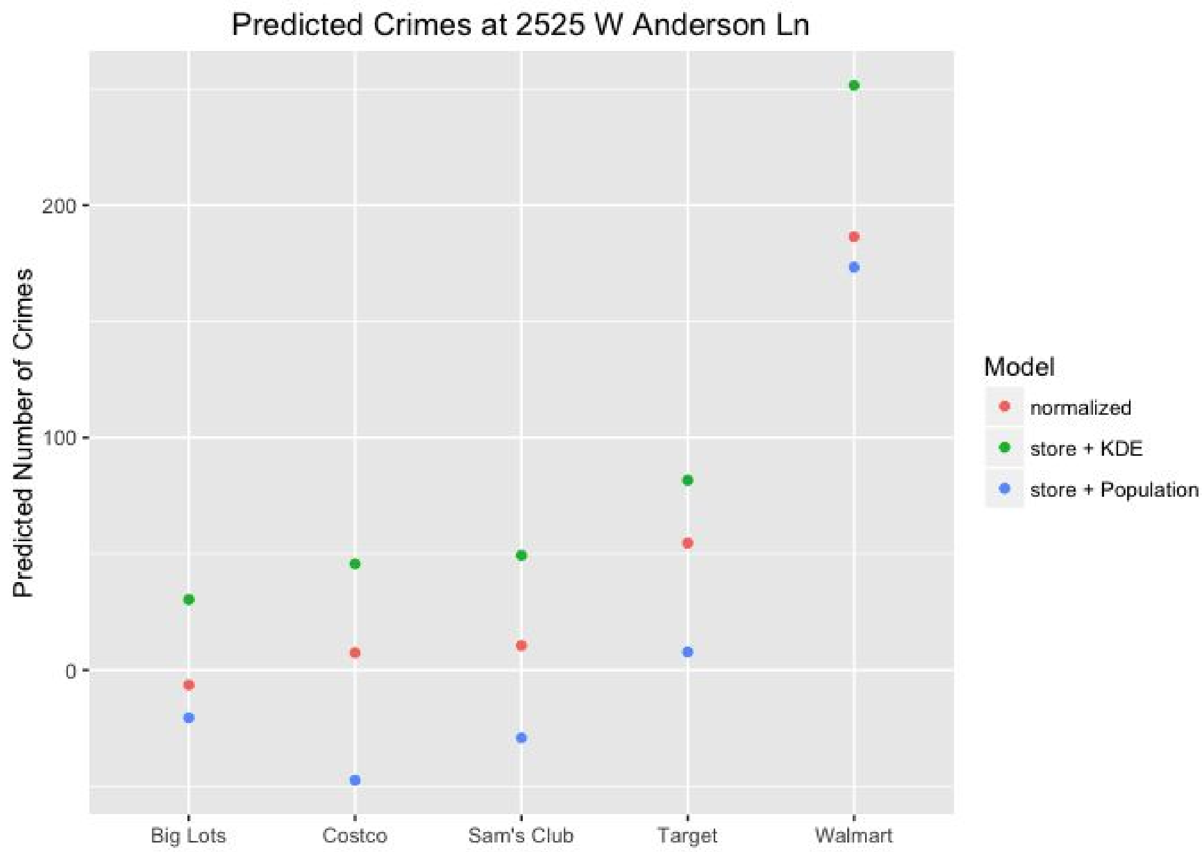

Fig. 2

Once these optimal models had been identified and examined, we performed residual analysis to further assess the validity of our modeling and the robustness of our conclusions. Plotting the R-student residuals for these models indicated issues with non-constant variance, confirming our previous intuition that the true relationship between regressors and response was nonlinear. Thus, we attempted several transformations of our response variable. Although there were improvements in the model by using the square root of the response (see Appendix E), there tended to still be a discernable funnel pattern. This indicates that different transformations might be needed, that we may be missing important variables, or perhaps that there is a quadratic relationship between crime and one or more of our regressors.

In the spirit of open data and reproducible research, we have also included a link to download all of our raw data and analysis code in Appendix F.

Conclusion

In summary, our optimal models indicate that indeed whether or not a store is a Walmart is highly correlated to the number of crimes at that location. In fact, throughout *all* of the models we tested, the only variable that was significant *every* time was whether or not the store was a Walmart. Furthermore, it was the *most* significant variable in almost every model. This is a strong indication that, even when controlling for a variety of socioeconomic and geographic factors, Walmarts have significantly more crime than similar stores.

We consider our research and methods to be novel because, while there have been anecdotal reports that a correlation exists between individual Walmart stores and crime rates across the United States [5], there have been no investigations that have applied data mining or robust statistical analysis to prove or disprove such a correlation. Furthermore, while a more formal study has compared county-level crime rates in counties with or without a Walmart, our analysis is the first to examine the amount of crimes that were reported *at the actual stores*. By looking at this, while controlling for the wide array of confounding factors which we considered, we have isolated the increase in crime that is directly attributable to Walmart, as opposed to crime that may be attributed to the areas in which Walmart chooses to build. This provides a stronger foundation upon which to pursue further study, and also to recommend practical action.

In regards to further research on this subject, there are a number of important lines of inquiry that should be pursued. First off, our analysis was limited to the city of Austin, Texas in 2015. A more robust treatment should incorporate data from other cities to make sure this effect is not specific to any one region. It would also benefit greatly from the inclusion of data from multiple years. Not only would this bolster the basic analysis, but it would allow future researchers to control for various temporal effects as well. Our literature review revealed many analyses which indicate that time of year, day of week, weather, and various holidays have a significant effect on crime. It would be useful to know whether the effect we are attributing to Walmart was particularly strong, for example, on holidays, or perhaps in the hot summer months. Additionally, we found some unexpected coefficient signs on a few of the variables we examined, as well as some troubling patterns in the residual analysis. If an attempt was made to make a more accurately predictive model, these issues should be investigated further.

Recommendations for action are always difficult to make, but it appears that some action is warranted in this case. Valuable public resources are being devoted to responding to a disproportionate amount of crimes at a single retail establishment. Further study is necessary to attempt to determine the *reasons* that Walmart is so strongly correlated with crimes, but some anecdotal evidence suggests that the amount of unmonitored parking they offer, as well as the amount of square footage which each employee is expected to cover, may be issues of concern. [5] These things deserve their own consideration in future research. Local governments, police departments, and the business community could also put pressure on Walmart to release some of its own relevant market research, in order help communities better cope with the addition of what appears to be a legitimate disruptive force. And finally, local officials could follow the lead of the Indiana mayor mentioned earlier and make efforts to affect change at specific stores in their own communities.

Practical considerations of these recommendations aside, we feel confident in again stating our result that, at least in Austin, Texas in 2015, Walmart stores contribute a significant amount to local crime rates as compared to their competitor stores. Considering the prevalence of Walmart in communities throughout America, we believe our findings merit further studies that should incorporate variables missing from our study, as well as data from other cities and multiple years, in order to have a more accurate view of Walmart's effect on crime.

Appendix A: Summary of Variables

Variable Name	Response/ Predictor	Description	Expected Sign
Crimes2015	Response	Number of crimes occurred at the store	n/a
NormalizedCrime	Response	Number of crimes occurred at the store, normalized to its store hours	n/a
MedIncome	Predictor	Median Income of residents in the store's zip code	Negative
AvgIncome	Predictor	Average Income of residents in the store's zip code	Negative
Population	Predictor	Population of the store's zip code	Positive
CollegeGradPercent	Predictor	Percent of residents in the zip code that holds a bachelor's degree	Negative
HighSchoolGradRate	Predictor	Percent of residents in the zip code that graduated high school	Negative
IncomeInequality	Predictor	The difference between the approximated 80^{th} income percentile and the approximated 20^{th} income percentile in the zip code	Positive
Unemployment	Predictor	The unemployment rate of the zip code	Positive
PovertyLine	Predictor	The % of residents under the poverty-line in the zip code	Positive
KDEraw	Predictor	The KDE value for the store's location	Positive
closest_stops_in_ meters	Predictor	Distance to closest bus stop in meters	Negative
wal	Predictor	Indicator variable if store is a Walmart or not	Positive

Appendix B: Model of crimes2015 with predictor variables using top-down approach

Variable	**Coefficient**	**P-value**	**Significance**
Intercept	11.48	0.95	
Population	4.98e-03	0.06	10% level
AvgIncome	-7.77e-04	0.487	
IncomeInequality	-2.46e-04	0.796	
Unemployment	-1.502e+01	0.5022	
KDE raw	1.838e+10	0.39	
Closest_stops_in_ meters	2.624e-02	0.43	
wal	1.786e+02	0.000414	0.05% level
Adjusted R-squared	0.542		

Appendix C: Model of crimes2015 with predictor variables using top-down approach without IncomeInequality

Variable	**Coefficient**	**P-value**	**Significance**
Intercept	-30.4	0.845	
Population	5.373e-03	0.0127	5% level
AvgIncome	-7.705e-04	0.481	
Unemployment	-1.485e+01	0.49637	
KDE raw	2.018e+10	0.3094	
Closest_stops_in_meters	2.710e-02	0.4055	
wal	1.791e+02	0.000281	0.05% level
Adjusted R-squared	0.563		

Appendix D: Model of crimes2015 with predictor variables using bottom-up approach without IncomeInequality

Variable	**Coefficient**	**P-value**	**Significance**
Intercept	-62.56	0.7467	
Population	4.198e-03	0.059237	10% level
CollegeGradPercent	-1.117	0.629145	
PovertyLine	-2.539e-01	0.954016	
KDE raw	1.536e10	0.468188	
Closest_stops_in_meters	1.409e-02	0.656871	
wal	1.812e02	0.000338	0.05% level
Adjusted R-squared	0.554		

Appendix E: Statistics for several models we tested

Response variable	**All predictor variables included in model**	**Significant variables (and associated p-values)**	**Adjusted R-squared**
crimes2015	CollegeGradPercent + PovertyLine + KDEraw + Population + wal	wal (0.0002)	0.5708
crimes2015	Population + KDEraw + CollegeGradPercent + wal	wal (0.0001)	0.5903
crimes2015	CollegeGradPercent + PovertyLine + wal	wal (0.0009)	0.5075
crimes2015	PovertyLine + wal	wal (0.0010)	0.4795
crimes2015 *(optimal model)*	KDEraw + wal	KDEraw (0.0223) wal (0.0001)	0.4706
crimes2015	KDEraw + wal + closest_stops_in_meters	KDEraw (0.0428) wal (0.0001)	0.4489
crimes2015 *(optimal model)*	Population + wal	Population (0.0006) wal (0.00004)	0.5947
normalizedCrime	KDEraw + wal + CollegeGradPercent	CollegeGradPercent (0.0280) wal (0.0055)	0.4704
normalizedCrime *(optimal model)*	KDEraw + wal + Population	Population (0.0029) wal (0.0004)	0.5565
normalizedCrime	KDEraw + wal	wal (0.3711)	0.0017
normalizedCrime	Population + wal	Population (0.0002) wal (0.0005)	0.5557
sqrt(crimes2015)	Population + wal	Population (0.000065) wal (0.000007)	0.6730 *(best overall)*

Appendix F

In the spirit of open data and reproducible research, we have provided access to our data sets and analysis code at the following link:

https://github.com/seth127/SYS-6018-cs1

References

[1] CNBC, "Wal-Mart profit, revenue beat expectations; US sales up more than expected," cnbc.com, para. 3, May 19, 2016. [Online]. Available: http://www.cnbc.com/2016/05/19/ walmart-q1-earnings.html. [Accessed Sept. 25, 2016].

[2] H. Blodget, "Walmart employs 1% of America. Should it be forced to pay its employees more?" *Business Insider*, para. 6, Sept. 20, 2010. [Online]. Available: http://www.businessinsider.com/walmart-employees-pay. [Accessed Sept. 25, 2016].

[3] C. Courtemanche and A. Carden, "Supersizing supercenters? The impact of Walmart Supercenters on body mass index and obesity," *Journal of Urban Economics*, vol. 69, no. 2, March 2011. [Online]. Available: http://ac.els-cdn.com/S0094119010000690/1-s2.0-S0094119010000690-main.pdf?_tid =ec8c48e0-8333-11e6-9404-00000aacb35d&acdnat= 1474817151_526a325d3939d205671962d4ed3b1a21. [Accessed Sept. 25, 2016].

[4] C. Fishman, *The Wal-Mart Effect: How the World's Most Powerful Company Really Works and How It's Transforming the American Economy*. New York, NY: Penguin Press, 2006, p. 9. [Online]. Available: https://books.google.com/books?id= oq2pgMztlzsC&printsec=frontcover&dq=%22wal-mart+effect%22#v=onepage&q= %22wal-mart%20effect%22&f=false. [Accessed Sept. 25, 2016].

[5] S. Pettypiece and D. Voreacos, "Walmart's out-of-control crime problem is driving police crazy," *Bloomberg Businessweek*, Aug. 17, 2016. [Online]. Available: http://www.bloomberg.com/features/2016-walmart-crime/. [Accessed Aug. 27, 2016].

[6] S.E. Wolfe and D.C. Pyrooz, "Rolling back prices and raising crime rates? The Walmart effect on crime in the United States," *British Journal of Criminology*, vol. 54, no. 2, pp. 199-221, Jan. 2014. [Online]. Available: https://bjc-oxfordjournals-org.proxy.its.virginia. edu/content/54/2/199.full.pdf+html?sid=2792c7d0-fda7-46cb-9f13-e11fbe85ba5c. [Accessed Sept. 25, 2016].

[7] S. Johnson, *The Ghost Map: The Story of London's Most Terrifying Epidemic--and How It Changed Science, Cities, and the Modern World*, New York, NY: Riverhead Books, 2006.

[8] E. Gumus, "Crime in urban areas: An empirical investigation," *Akdeniz I.I.B.F. Dergisi*, vol. 4, no. 7, 2004, pp. 98-109. [Online]. Available: http://poseidon01.ssrn.com/delivery. php?ID=539085124070071086005069074107115011059080066060051090009027114 124070030010073116007001050008026018126038029006076091119093018026063049093124066088084090026038055037091089124031016092024028078024109109090

50770911100241030230681080950910830900898&EXT=pdf. [Accessed Aug. 29, 2016].

[9] C. Jones, "Genetic and environmental influences on criminal behavior." [Online]. Available: http://www.personalityresearch.org/papers/jones.html [Accessed Aug. 29, 2016].

[10] E. Moretti, "Does education reduce participation in criminal activities?" in *Teacher's College, Columbia University Fall 2005 Symposium on the Social Costs of Inadequate Education, 24-25 October 2005, New York, NY* [Online]. Available: http://devweb.tc.columbia.edu/manager/symposium/Files/74_Moretti_Symp.pdf. [Accessed: Aug. 29, 2016].

[11] R. Kersley and M. Stierli, "Global wealth report 2015," *Credit Suisse AG Research Institute*, 2015. [Online]. Available: https://publications.credit-suisse.com/tasks/render/file/?fileID=F2425415-DCA7- 80B8-EAD989AF9341D47E. [Accessed Aug. 29, 2016].

[12] D. Hardoon, et al, "An Economy for the 1%: How privilege and power in the economy drive extreme inequality and how this can be stopped," *Oxfam International*, 210 Oxfam Briefing Paper, 2016. [Online]. Available: https://www.oxfam.org/sites/www.oxfam. org/files/file_attachments/bp210-economy-one-percent-tax-havens-180116-en_0.pdf. [Accessed Aug. 29, 2016].

[13] L. Stolzenberg, et al., "Race, economic inequality, and violent crime," *Journal of Criminal Justice*, vol. 34, no. 3, pp. 303–316, May-June 2006. [Online]. Available: http://ac.els-cdn.com/ S0047235206000250/1-s2.0-S0047235206000250-main.pdf?_tid=6649de34-8354-11e6-9438-00000aacb35f&acdnat=1474831099_ad0b1681f3e7117791717892 e0bc6d74. [Accessed Aug. 29, 2016].

[14] J.R. Hipp, et al., "Crimes of opportunity or crimes of emotion? Testing two explanations of seasonal change in crime," *Social Forces*, vol. 82, no. 4, pp. 1333-1372, June 2004. [Online]. Available: https://sf-oxfordjournals-org.proxy.its.virginia.edu/content/82/4/1333.full.pdf+html?sid=5ba868b6-be6e-465b-be69-cd5c6bb26ca3. [Accessed Aug. 28, 2016].

[15] C.A. Anderson and D.C. Anderson, "Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses." *Journal of Personality and Social Psychology*, vol. 46, no. 1, pp. 91-97, Jan. 1984. [Online]. Available: http://blog.afletcher.net/articles/AmbientTemperatureandViolentCrime.pdf. [Accessed Aug. 28, 2016].

[16] S.J. Linning, et al., "Crime seasonality: Examining the temporal fluctuations of property crime in cities with varying climates," *International Journal of Offender Therapy and Comparative Criminology*, pp. 1-26, Mar. 2016,

doi:10.1177/0306624X16632259. [Online]. Available: http://ijo.sagepub.com.proxy.its.virginia.edu/content/early/2016/03/16/0306624X16632259.full.pdf+html. [Accessed Aug. 29, 2016].

[17] M.A. Andresen and N. Malleson, "Intra-week spatial-temporal patterns of crime," *Crime Science*, vol. 4, no. 12, pp. 1-11, July 2015. [Online]. Available: http://download.springer.com/static/pdf/997/art%253A10.1186%252Fs40163-015-0024-7.pdf?originUrl=http%3A%2F%2Fcrimesciencejournal.springeropen.com%2Farticle%2F10.1186%2Fs40163-015-0024-7&token2=exp=1474846016~acl=%2Fstatic%2Fpdf%2F997%2Fart%25253A10.1186%25252Fs40163-015-0024-7.pdf*~hmac=6148dc91de804330f3e42af77270ccabe3a490e6a0a4363784afd6f54c3d7c40. [Accessed Aug. 27, 2016].

[18] B.J. Bushman, et al., "Is the curve relating temperature to aggression linear or curvilinear? Assaults and temperature in Minneapolis reexamined," *Journal of Personality and Social Psychology*, vol. 89, no. 1, pp. 62-66, July 2005. [Online]. Available: http://psycnet.apa.org.proxy.its.virginia.edu/journals/psp/89/1/62.pdf&productCode=pa. [Accessed Aug. 28, 2016].

[19] E.G. Cohn and J. Rotton, "Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays," *Journal of Criminal Justice*, vol. 31, no. 4, pp. 351-360, July-Aug. 2003. [Online]. Available: http://ac.els-cdn.com/S0047235203000291/1-s2.0-S0047235203000291-main.pdf?_tid=f83ce46c-8377-11e6-a5d8-00000aab0f27&acdnat=1474846377_5a2b1fcb174efaf889ea199385e3ec7d. [Accessed Aug. 28, 2016].

[20] D. Weisburd and T. McEwen, "Crime mapping and crime prevention," in *Crime Prevention Studies*, Vol. 8, R.V. Clarke, Ed. Monsey, NY: Criminal Justice Press, 1997, pp. 1-23. [Online]. Available: http://www.popcenter.org/library/crimeprevention/volume_08/title_and_copyrights_pages.pdf. [Accessed Aug. 29, 2016].

[21] R. Deryol, et al., "Crime places in context: An illustration of the multilevel nature of hot spot development," *Journal of Quantitative Criminology*, vol. 32, no. 2, pp. 305-325, June 2016. [Online]. Available: http://link.springer.com.proxy.its.virginia.edu/ article/10.1007/s10940-015-9278-1. [Accessed Aug. 29, 2016].

[22] S. Han, et al., "The effect of a Sunday liquor-sales ban repeal on crime: A triple-difference analysis," *Alcoholism: Clinical and Experimental Research*, vol. 40, no. 5, pp. 1111-1121, May 2016. [Online]. Available: http://onlinelibrary.wiley.com.proxy.its.virginia.edu/doi/10.1111/acer.13047/epdf. [Accessed Aug. 29, 2016].

[23] L. Cui and R. Walsh, "Foreclosure, vacancy and crime," *Journal of Urban Economics*, vol. 87, pp. 72-84, May 2015. [Online]. Available: http://ac.els-cdn.com/S0094119015000029/1-s2.0-S0094119015000029-main.pdf?_tid=37b1d5c2-8363-11e6-b1e0-00000

aab0f02&acdnat=1474837464_3b791a51c571b858a20d032738fad48e. [Accessed Aug. 29, 2016].

[24] C.D.M. Furr-Holden, et al., "Not in my back yard: A comparative analysis of crime around publicly funded drug treatment centers, liquor stores, convenience stores, and corner stores in one Mid-Atlantic city," *Journal of Studies on Alcohol and Drugs*, vol. 77, no. 1, pp. 17-24, Jan. 2016. [Online]. Available: http://www.jsad.com/doi/full/10.15288/jsad.2016.77.17. [Accessed Aug. 29, 2016].

[25] J.E. Eck and D. Weisburd, "Crime places in crime theory," in *Crime and Place*, J.E. Eck and D. Weisburd, Eds. Cincinnati, OH: Criminal Justice Press, 1995, pp. 1-33. [Online]. Available: http://www.popcenter.org/library/crimeprevention/volume_04/01-eckweisburd.pdf. [Accessed Aug. 29, 2016].

[26] R. MacKay, "Annual crime dataset 2015," *data.austintexas.gov*, July 2016. [Online] Available: https://data.austintexas.gov/Public-Safety/Annual-Crime-Dataset-2015/spbg-9v94, July 2016. [Accessed Aug. 29, 2016].

[27] Nasdaq, "Wal-Mart Stores, Inc. competitors," nasdaq.com, Sept. 25, 2016. [Online]. Available: http://www.nasdaq.com/symbol/wmt/competitors. [Accessed Sept. 25, 2016].

[28] Yellow Pages, "Austin discount stores," yellow.pages.com. [Online]. Available: http://www.yellowpages.com/austin-tx/discount-stores. [Accessed Sept. 25, 2016].

[29] Yellow Pages, "Austin department stores," yellow.pages.com. [Online]. Available: http://www.yellowpages.com/austin-tx/department-stores. [Accessed Sept. 25, 2016].

[30] Yellow Pages, "Austin general merchandise," yellow.pages.com. [Online]. Available: http://www.yellowpages.com/austin-tx/general-merchandise. [Accessed Sept. 25, 2016].