The Effects of Marijuana Legalization on Crime and Other Drug Usage in Seattle

Benjamin Gaiser and Jeremy Russell

16 Decembre 2016

Contents

Introduction (400 words?)

Lorem ipsum dolor sit amet, oportere pertinacia repudiandae ne mea, eam ne meis eleifend. Ferri splendide tincidunt cum ad, vis mandamus posidonium cu, eos no vitae sententiae honestatis. Tantas inciderint sed eu, aeque sonet efficiantur no sed. Porro debet pertinax per at, ea mea unum vide.

Literature Review and Policy Relevance (600 words?)

Lorem ipsum dolor sit amet, oportere pertinacia repudiandae ne mea, eam ne meis eleifend. Ferri splendide tincidunt cum ad, vis mandamus posidonium cu, eos no vitae sententiae honestatis. Tantas inciderint sed eu, aeque sonet efficiantur no sed. Porro debet pertinax per at, ea mea unum vide.

Data Gathering and Methodology (1000 words?)

Data Gathering (so far: 657 words)

Our project will be using three main data sets: i) cannabis data on Washington State, ii) crime data on Seattle, WA, and iii) socio-economic data on Seattle, WA. As the policy field of cannabis legalization is a rather nascent field of study, there has been limited availability of reliable data (previous data sources were clouted by the illegality of possession and consumption and thus approximated dark figures).

Our project has taken advantage of one of the provisions within the 'Washington Initiative 502', the initiative that brought about the legalization of cannabis in Washington State, that established disclosure of retailers, processors and producers as well as some general economic data online. The data has been web-scraped using the rvest package (Wickham 2016a). As there have been difficulties due to the 502data site being scripted in Java, we additionally needed the 'phantomjs' file to use the RSelenium package (Harrison 2016) to make the web-scraping possible.

As the ".exe"-file does not work on Macs and because scraping is time-consuming, we will provide an unaldutered ".txt"-file called "ProdProc" for all producers and processors and one called "Retailer" for retailers. This way we allow for as much reproducibility as possible. The data lists the names and locations of producers/processors in the state and of retailers including YTD sales and tax revenue. For the later analysis, we will constrain the data set to Seattle. The second set of data concerns data on crime. We scraped the data for Seattle with the jsonlite package (Ooms, Temple Lang, and Hilaiel 2016) and formatted it into R-readable format with the gdata package (Warnes et al. 2015). The initial data set is rather large (>80 MB). We have transformed it into a zipped-file called "CrimeSeattle.gz". For the analysis, we cleaned the data so to only contain information about the longitude/latitude, the type of crime and the time it has occurde. For the types of crime we have aggregated individual types into seven broad categories: i) alcohol-related crime (includes Liquor offenses and DUI offenses), ii) burglary/theft, iii) marijuana-related crime, iv) narcotics-related crime, v) other-drug related crimes, vi) property crime, and vii) violent crime (includes armed robberies, assaults, drive by shootings, strong arm robbery and homicides).

The third set of data comes from the U.S. Census data and provides us with socio-economic factors for each district. Our code to obtain these data follows the concept of the blog post by zevross. One limitation is that we, however, only have this data annually and only for 2014. The variables included for this analysis are the share of poverty for each district, the average age level, the prevalence of minority races, as well as the level of education. While these are classic socio-economic factors, the lack of variation means that they will only be useful for the regular OLS regression as well as for the propensity score matching, which means that this set of data will be of limited value overall.[1]

In order to be able to merge all three data sets, we needed to ensure that we had geospatial data in all of them. Only for the first data set did we not already have longitude and latitude data. To obtain these, we scraped the street locations for all available retailers from kushtourism.com. With this information, we were

[1] should find quote for stereotypical soes

able to use a simple Google API query in combination with the RJSONIO package (Temple Lang 2014) to determine latitude and longitude for each observation. This left us with 15 individual retailers for Seattle for which we have data.[2] For the merger of all three data sets, we calculated the centroid of each district of the U.S. Census Data and then calculated the minimum distance of crime incidents and retailers using the fields package (Nychka et al. 2016). The final data set is called "SeattleCrimeAnalysis.gz". As we are not looking at the individual perpetrator but at the crime incident of a district, we feel that this is an adequate way of merging the data set.

Methodology (so far 198 words)

This study will employ several methods in order to analyse whether there is a discernible effect between there being a dispensary in a given district of Seattle and the logarithmic crime incident per 1000 people in that district. We will first use descriptive statistics to show the development of the cannabis industry in Washington State and Seattle to give a feeling for the scope of the topic at hand. The descriptive statistic will be rounded up by a quick look at the differences in means between districts with and districts without a dispensary. Secondly, we will perform a standard ordinary least squares analysis (OLS) to analyse the main relationship of this paper. Unfortunately, normal OLS is unable to adequately identify causal effects when data is observational. It might be good for an initial understanding but not more. To account for this, we complement this analysis with a propensity score matching model. "Matching on the propensity score is esentially a weighting scheme, which determines what weights are placed on comparison units when computing the estimated treatment"[3] This means that far from looking at the average effect across all districts, we will try to identify how the establishment of a dispensary has changed crime incidents for similar districts (which should boost our confidence in causality). As propensity score matching has recently been under fire, we will also use another robustness checks: the robustness check is a simple fixed effects model, as it "avoids the ommited variable bias through controlling for [district] level heterogeneity by means of dummy variables.[4] To implement this, we will be using the plm package (Croissant, Millo, and Tappe 2016). A robustness check could be a difference-in-differences analysis that uses establishment as *treatment* and all other districts as *control* groups. Yet, because we have different shops opening at different times throughout the period 2014-2016, we will not adopt this additional model due to its complexity.

Analysis

Descritpive Statistics

Let us begin with with simple descriptive statistics. While the 502data is rather well documented, they do, however, run in the issue that their retailers are not uniquely identified by names. Thus, while we will be using the names for producers/processors to get their numbers, we will be using the URL extensions that we received from the scraping process to identify how many retailers there are.

As we can see from Figure 1, the nascent industry is already booming in Washington with both the numbers of actors and the sales/tax revenues being already higher in October 2016 than in all of 2015. The numbers are quite staggering:

1. Number of Producers and Processors: The number increased by 210% YoY in 2015 from 185 to 574 and by another 53% in the ten months leading up to October 2016 to 881.
2. Number of Retailers: The number increased by 130% YoY in 2015 from 82 to 189 and by another 69% in the ten months leading up to October 2016 to 319.
3. Increase of revenues in sales: The number increased by 960% YoY in 2015 from $29.21[5] Mill to $309.86 Mill and by another 77% in the ten months leading up to October 2016 to $549.93 Mill.

[2] The query was conducted on 10 November 2016

[3] @dehejia2002propensity: 153.

[4] Möhring (2012)

[5] when did the first one open

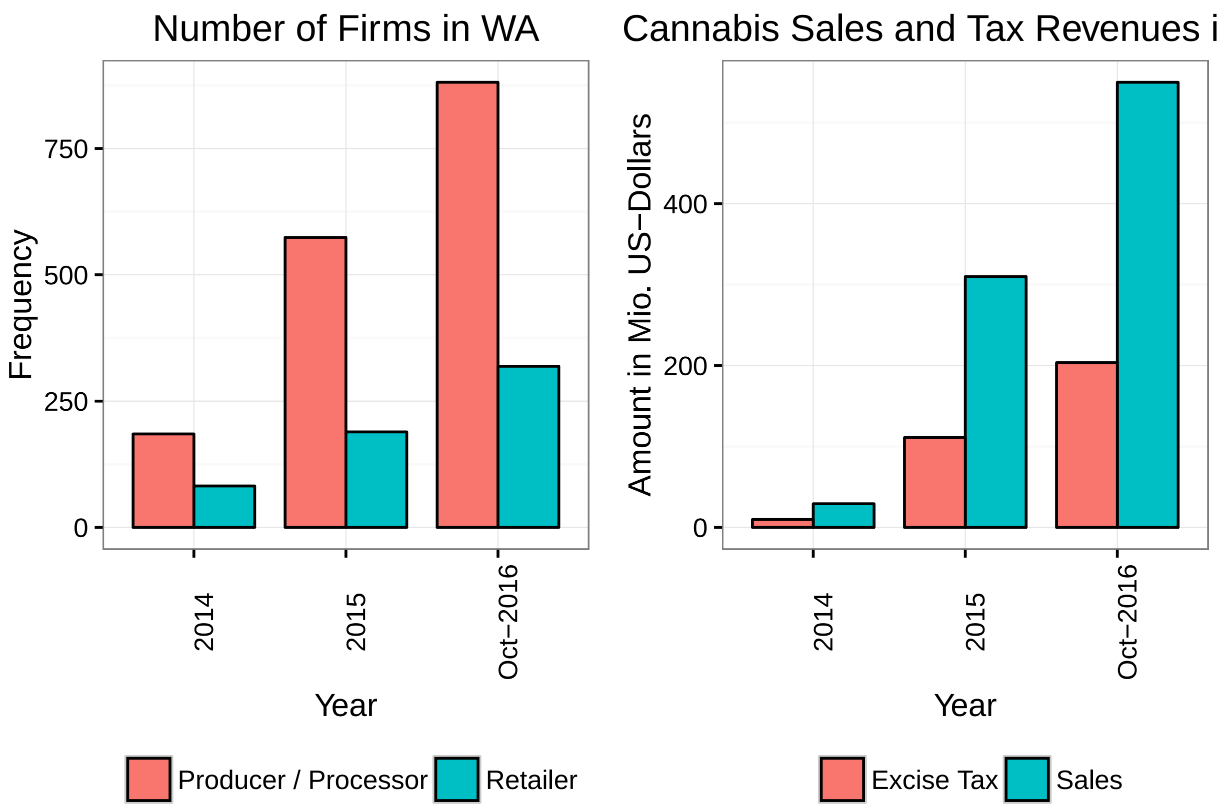

Figure 1: Business Development of Cannabis Production and Sales in Washington State

4. Increase of tax revenues: The number increased by 1040% YoY in 2015 from \$9.74 Mill to \$110.97 Mill and by another 83% in the ten months leading up to October 2016 to \$203.47 Mill.

It is a fast growing industry and shows increasing interest of buyers. In relation to Washington State's GDP it is, however, still relatively small with a share of 0.12 % of a GDP of \$445.4 Bill.[6] Unfortunately, we do not have data on the quantities that have been sold or reliable data on the price development of the good. In either of this scenario we could have otherwise also analysed the usage.

Table 1 shows the variables used in the next part of our analysis. While Seattle is divided into 132 districts by the U.S. Census, crime incidents only occured in 35 of them once accounting for other NA-related issues with our merged data set. Eight of these 35 districts have a dispensary, which is represented by the *treatment* variable "Established". The dependent variable is the natural logarithm of crime incidents per 1000 inhabitants.[7] The poverty rate, age, diversity and education are socio-economic control variables. Table 1 also indicates how these variables have been operationalised.

	Variables	Operationalisation
1	Producer/Processor	Count of Producers/Processors in WA
2	Retailer	Count of Retailers in WA
3	Excise Tax	Sum of taxes paid by retailers
4	Sales	Sum of revenue made by retailers
5	Crime Categories	Aggregates of Crime incidents from Seattle 911-Incident Database in log/1000 citizens
6	Established	Treatment variable: 0 if no dispensary, 1 otherwise
7	Poverty Rate	Share of People < \$15,000/year as percent of total people in district
8	Age	Shares of Adults in comparison to mean of Seattle
9	Diversity	Share of Whites in comparison to mean of Seattle
10	Education	Share of Degree holders in comparison to mean of Seattle
11	GEOID	U.S. Census District
12	Time	Months from Jan 2014 to Oct 2016

Table 1: List of dependent and explanatory variables used in analysis

Crime is definitely an issue in Seattle. If we look at Figure 2, we can see the distribution of our different crime categories for October 2016. We can clearly see that crime is concentrated in central Seattle and that there is less crime around the red dots that represent the existence of a dispensary. The stark concentration of crime follows the experience of other big cities: take Boston as an example where "fewer than 5 percent of [its] street corners and block faces generated 74 percent of fatal and non-fatal shootings between 1980 and 2008."[8]

If we look at crime incidents per 1,000 citizens in Figure 3 in the Appendix, then we can see that the development has been different between the different categories. As you can see from Figure 4 in the Appendix, however, only looking at the distribution, however, might be a problem due to the heavy right-tail skewdness. Thus, and in order to make interpretation easier, we will take the natural logarithm. Figure 4 in the Appendix now shows a slightly different picture where most of the crime follows a random walk, i.e. there is change in either direction without there being a trend. Only burglary/theft, property, and violent crime seems to have an upward trend, while other drug related crimes seem to have dropped recently.

We tested if the our analysis might be skewed due to the jumps from January 2014 to February 2014 that we can see in the Figures in the Appendix. The jumps might be due to coding changes/errors from one year to the next. We have not found any statistical significance. You can find the results in first table of the Appendix.

[6]Data from the Bureau of Economic Analysis was last retrieved 15/12/2016
[7]As a total, as well as for each of the seven different crime categories
[8]@ferguson2016policing: 20-21.

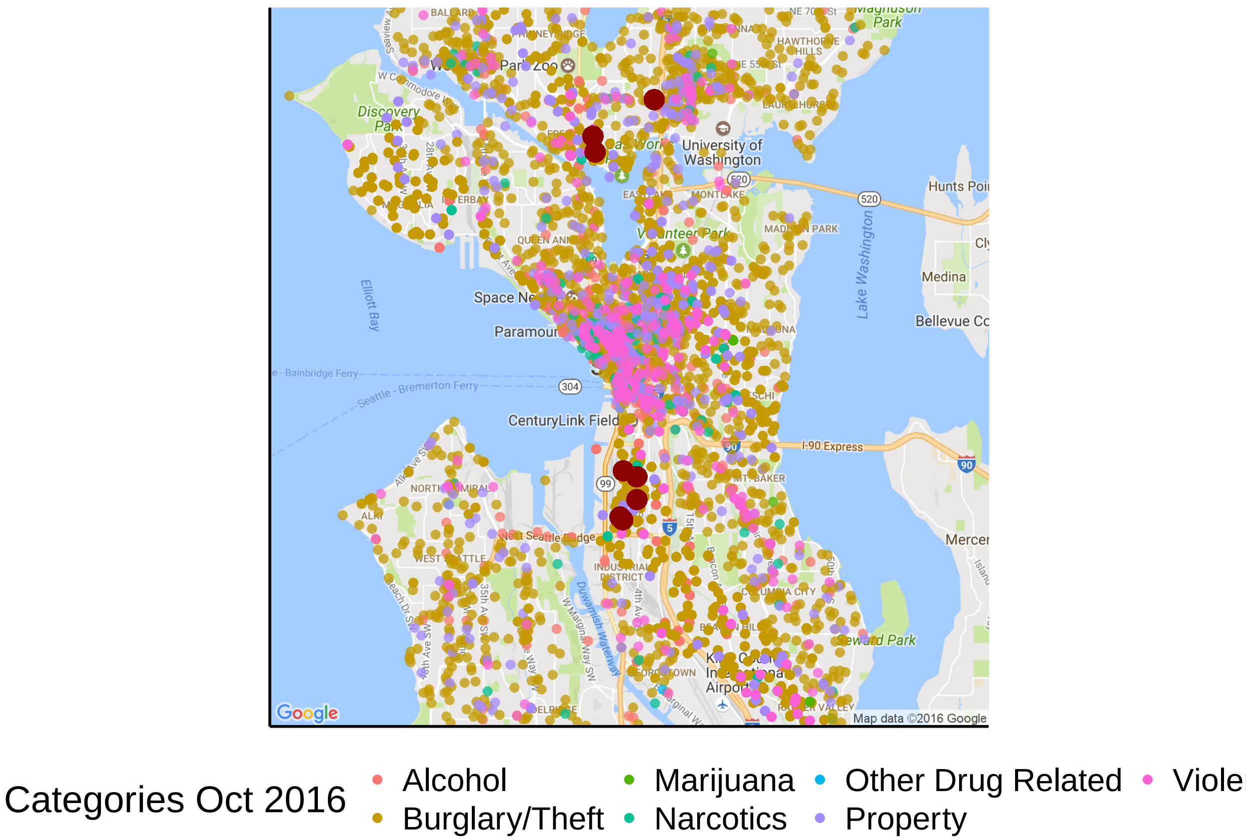

Figure 2: Map of Crime incidents in Seattle in October 2016

Inferential Statistics

Let us now look at the actual inferential statistics. We can see from Figure XXX that the impact of the *treatment* differs for each of the crime categories differently. It seems that alcohol-related crime is not statistically affected at all by the establishment of a dispensary. Burglary/Theft, property and violent crime, however all seem to be positively impacted by the establishment, whereas marijuana-related crime is substantially smaller and narcotics-related crime somewhat smaller. Other drug-related crime seem to have changed little.

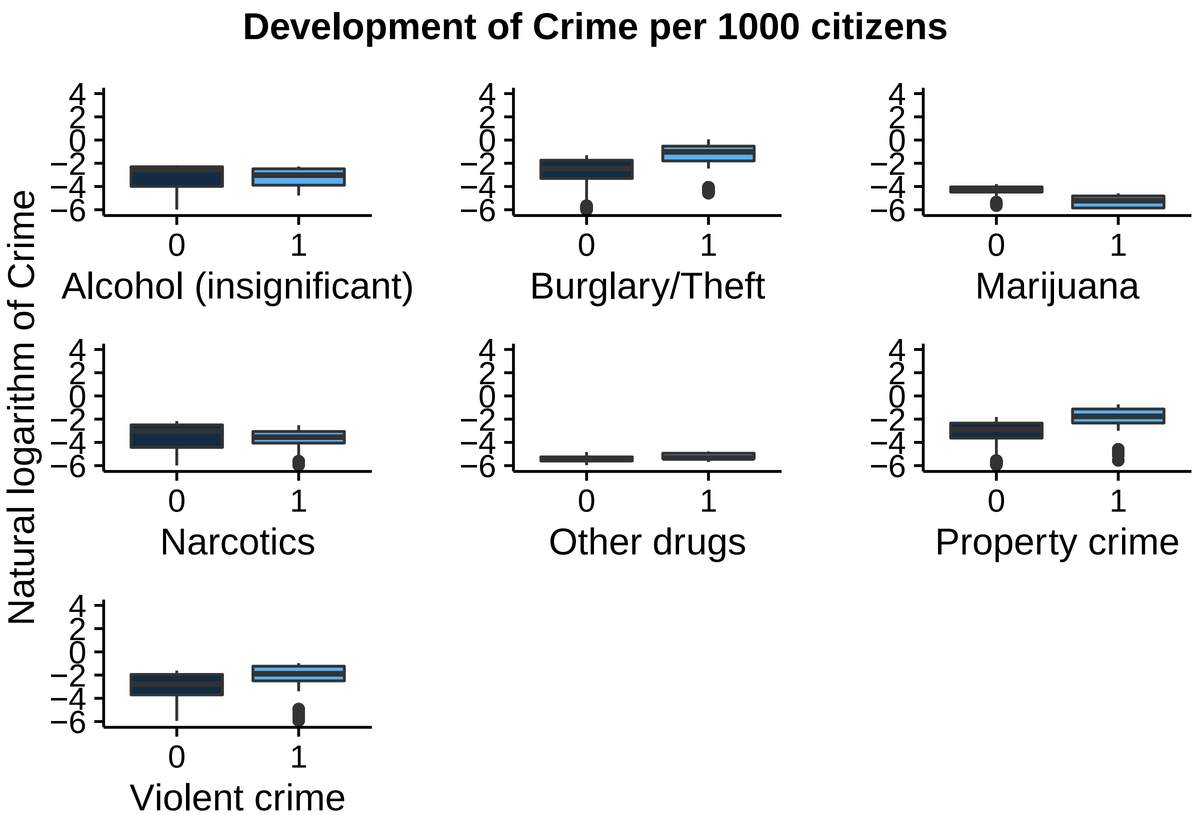

Figure 3: Impact of Dispensaries on Crime incident - Mean-difference

Yet, a simple differences in means analysis does not do our question justice. So we are looking at a regression.

A simple OLS with only our dependent and our treatment variable shows actually that there is a strong positive effect of a dispensary and the crime incident. Remember that our dependent variable is in log-form which means that this equates to a 1.34 percent increase in crime. The variation explained is 17% which is not too bad for a single variable. Yet, if we apply our more rational model of propensity score matching, looking at similar units, we see that the effect drops substantially to only 0.55 percent increase in crime. Yet, still a positive and highly statistically significant effect. If we look, however at the R^2-value now, we can see that the variation in crime is no longer explained at all by this variable, thus suggesting that there are other variables that are important.

Once we include our covariates, we are able to look at the relationship in more detail. The establishment of a dispensary is still positively related with the crime rate and is substantial in the Full OLS model and about half in the propensity score matching scenario. This time, the R^2-value does not drop but rather increases to 53%, suggesting that the socio-economic variables chosen do explain the variation in crime quite well. In particular, we have that below average adults, i.e. when there are more retirees that could be prone to

Table 2: Analysis of Crime

	Dependent variable:	
	ln(Crime Rate Per Thousand)	
	OLS Simple	Propensity Simple
	(1)	(2)
Established	1.34*** (1.33, 1.35)	0.55*** (0.54, 0.57)
Constant	−2.98*** (−2.98, −2.97)	−2.98*** (−2.98, −2.97)
Observations	228,574	208,172
R^2	0.17	0.02
Adjusted R^2	0.17	0.02
Residual Std. Error	1.20 (df = 228572)	1.16 (df = 208170)
F Statistic	46,214.14*** (df = 1; 228572)	5,178.47*** (df = 1; 208170)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3: Analysis of Crime

	Dependent variable:	
	ln(Crime Rate Per Thousand)	
	OLS Full	Propensity Full
	(1)	(2)
Established	1.40*** (1.39, 1.42)	0.66*** (0.65, 0.67)
share_poverty	0.04*** (0.04, 0.04)	0.05*** (0.05, 0.05)
Adulthood# Adults: Above average	−0.11*** (−0.13, −0.09)	−1.21*** (−1.24, −1.19)
Adulthood# Adults: Below average	1.10*** (1.08, 1.12)	0.35*** (0.34, 0.36)
Adulthood# Adults: Just above average	0.24*** (0.22, 0.26)	−0.52*** (−0.53, −0.51)
EduCat# Graduates: High	0.02 (−0.01, 0.05)	−1.33*** (−1.37, −1.29)
EduCat# Graduates: Very High	−0.79*** (−0.83, −0.76)	−2.42*** (−2.47, −2.38)
EduCat# Graduates: Very low	−1.29*** (−1.33, −1.26)	−2.42*** (−2.47, −2.38)
DiversityDiversity: High	0.17*** (0.15, 0.19)	0.11*** (0.10, 0.12)
DiversityDiversity: Very high	0.06*** (0.04, 0.08)	−0.23*** (−0.26, −0.21)
DiversityDiversity: Very Low	0.40*** (0.37, 0.43)	1.03*** (1.01, 1.05)
Constant	−3.51*** (−3.54, −3.48)	−1.22*** (−1.25, −1.18)
Observations	228,574	175,643
R^2	0.33	0.53
Adjusted R^2	0.33	0.53
Residual Std. Error	1.07 (df = 228562)	0.53 (df = 175631)
F Statistic	10,319.37*** (df = 11; 228562)	18,020.97*** (df = 11; 175631)
Note:		*p<0.1; **p<0.05; ***p<0.01

be victims or more adolescents that could be prone to become criminals, then we see more crime. A very high level of graduate degree holders in a district does also reduce the crime rate, yet so does a very low number of graduate holders, which might in this case be due to the concentration on only 35 districts.[9] For the Propensity model, it is interesting to note that almost every variable seems to have a negative effect on the crime rate, except of the share of poverty, there being less adults than on average, and there being either high or very low levels of diversity. Yet, looking at Figure 7: A.5, we must concede that our findings here might be driven for other reasons because we face the so-called common support problem, when the overlap between treated and untreated is so low that we lose confidence in our findings.

Table 4: Analysis of Crime

	Dependent variable:
	ln(Crime Rate Per Thousand) Fixed Effects
Established	0.59*** (0.58, 0.61)
Observations	228,574
R^2	0.02
Adjusted R^2	0.02
F Statistic	4,178.01*** (df = 1; 228538)
Note:	*p<0.1; **p<0.05; ***p<0.01

issue with data availability

Discussion

Burglary/Theft, property and violent crime either up because of other omitted variables or potentially also due to previous drug sellers out of job (but then why only so late in the data -> e.g. see trend is picking up recently; displacement takes a while? no rather ovb)

Conclusion

Hello

[9] in discussion

Appendix

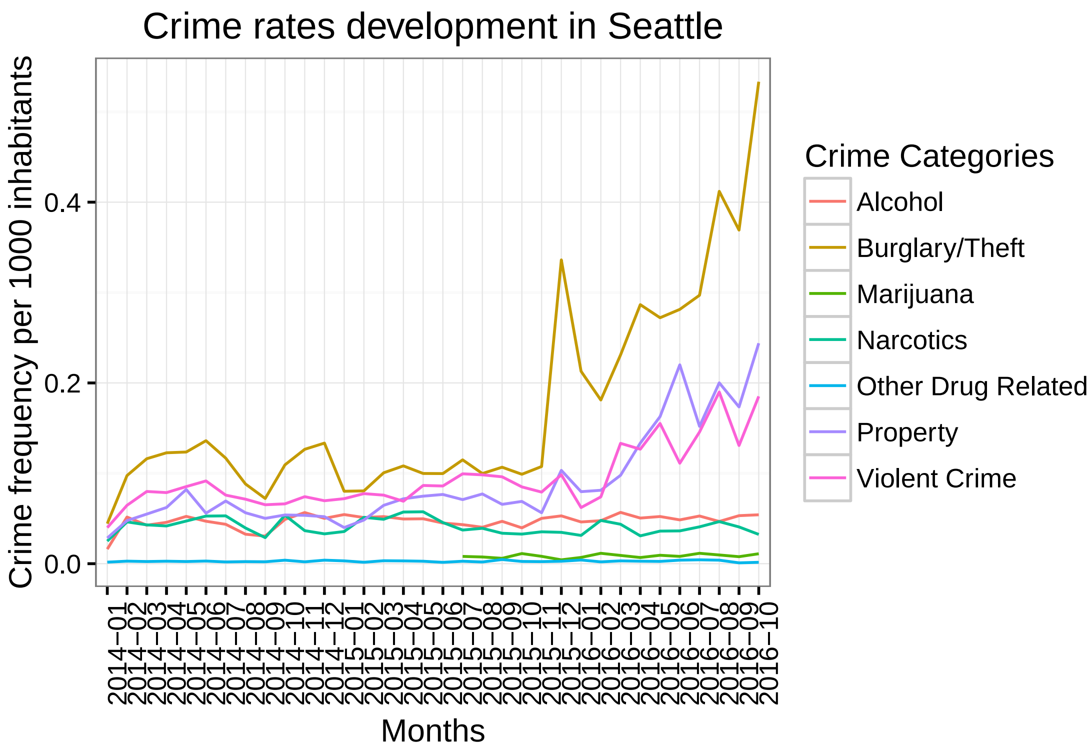

Figure 4: A.1 - Crime development in Seattle

Table 5: Analysis of Crime

	Dependent variable:	
	ln(Crime Rate Per Thousand)	
	Full Model	Without Jan 2014
	(1)	(2)
Established	1.34*** (1.33, 1.35)	1.33*** (1.32, 1.34)
Constant	−2.98*** (−2.98, −2.97)	−2.97*** (−2.97, −2.96)
Observations	228,574	226,147
R^2	0.17	0.17
Adjusted R^2	0.17	0.17
Note:		*p<0.1; **p<0.05; ***p<0.01

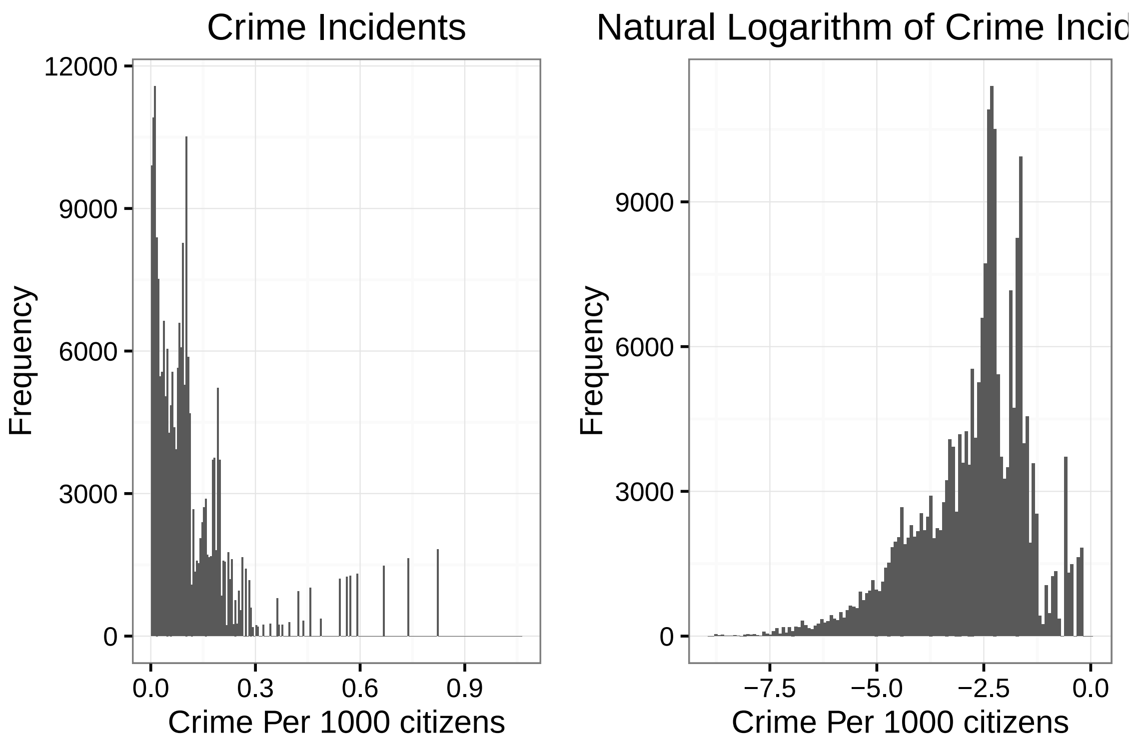

Figure 5: A.2 - Distribution of Crime across the Districts

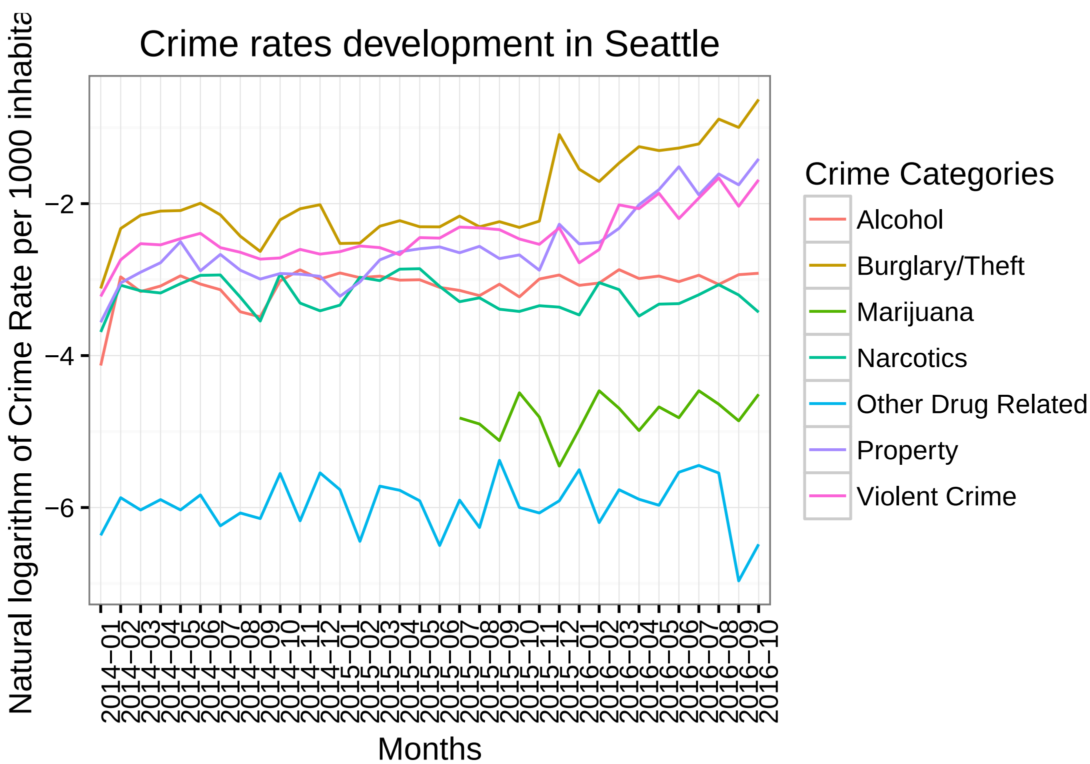

Figure 6: A.3 - Percentage change Month on Month in Seattle

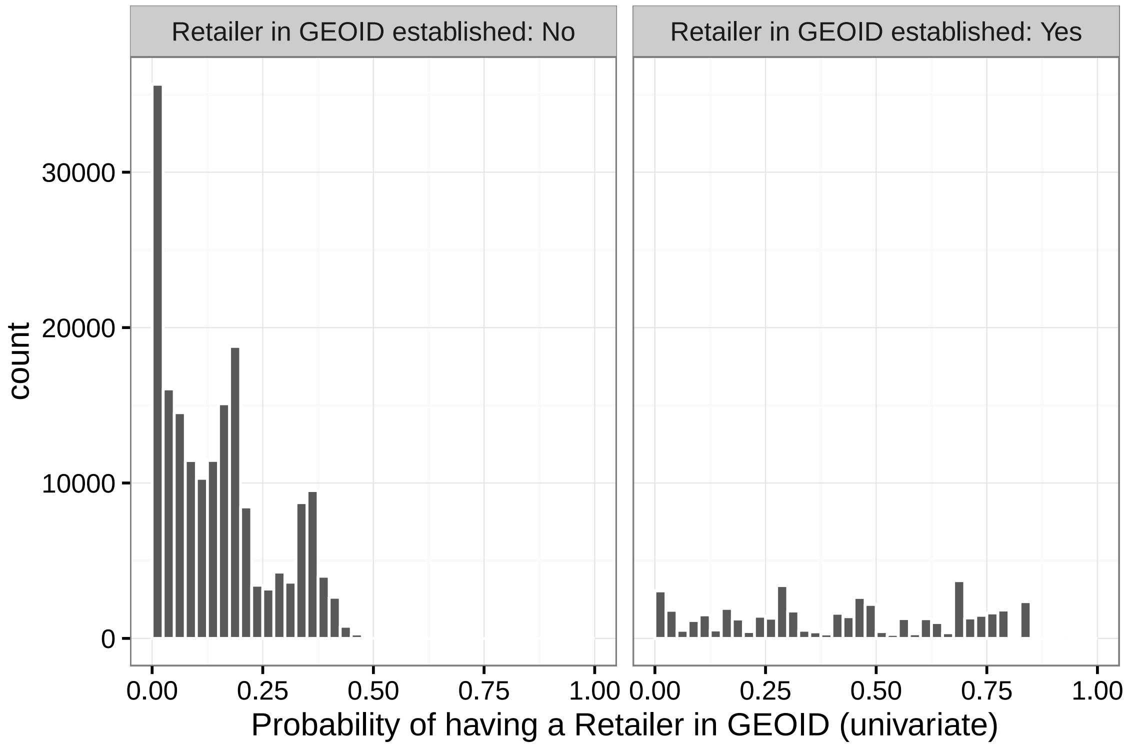

Figure 7: A.5 - Common Support Graph/Visualisation propensity score matching

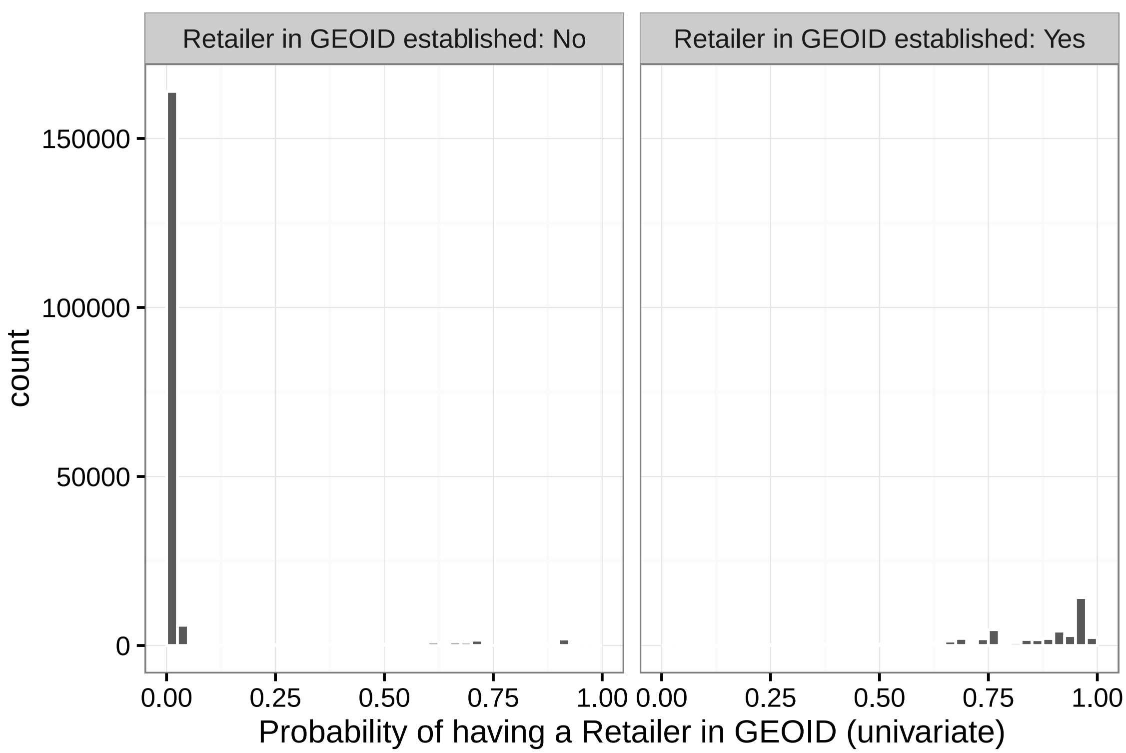

Figure 8: A.6 - Common Support Graph/Visualisation propensity score matching with covariates

References

Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr.

Bivand, Roger, and Nicholas Lewin-Koh. 2016. *Maptools: Tools for Reading and Handling Spatial Objects*. https://CRAN.R-project.org/package=maptools.

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2016. *Rgdal: Bindings for the Geospatial Data Abstraction Library*. https://CRAN.R-project.org/package=rgdal.

Croissant, Yves, Giovanni Millo, and Kevin Tappe. 2016. *Plm: Linear Models for Panel Data*. https://CRAN.R-project.org/package=plm.

Glenn, Ezra Haber. 2016. *Acs: Download, Manipulate, and Present American Community Survey and Decennial Data from the US Census*. https://CRAN.R-project.org/package=acs.

Harrison, John. 2016. *RSelenium: R Bindings for Selenium WebDriver*. https://CRAN.R-project.org/package=RSelenium.

Möhring, Katja. 2012. "The Fixed Effect as an Alternative to Multilevel Analysis for Cross-National Analyses." *Cologne: University of Cologne (GK Soclife Working Paper 16)*.

Nychka, Douglas, Reinhard Furrer, John Paige, and Stephan Sain. 2016. *Fields: Tools for Spatial Data*. https://CRAN.R-project.org/package=fields.

Ooms, Jeroen, Duncan Temple Lang, and Lloyd Hilaiel. 2016. *Jsonlite: A Robust, High Performance JSON Parser and Generator for R*. https://CRAN.R-project.org/package=jsonlite.

Pebesma, Edzer, and Roger Bivand. 2016. *Sp: Classes and Methods for Spatial Data*. https://CRAN.R-project.org/package=sp.

Temple Lang, Duncan. 2014. *RJSONIO: Serialize R Objects to JSON, JavaScript Object Notation*. https://CRAN.R-project.org/package=RJSONIO.

Temple Lang, Duncan, and the CRAN team. 2016. *RCurl: General Network (HTTP/FTP/.) Client Interface for R*. https://CRAN.R-project.org/package=RCurl.

Walker, Kyle. 2016. *Tigris: Load Census TIGER/Line Shapefiles into R*. https://CRAN.R-project.org/package=tigris.

Warnes, Gregory R., Ben Bolker, Gregor Gorjanc, Gabor Grothendieck, Ales Korosec, Thomas Lumley, Don MacQueen, Arni Magnusson, Jim Rogers, and others. 2015. *Gdata: Various R Programming Tools for Data Manipulation*. https://CRAN.R-project.org/package=gdata.

Wickham, Hadley. 2016a. *Rvest: Easily Harvest (Scrape) Web Pages*. https://CRAN.R-project.org/package=rvest.

———. 2016b. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Wickham, Hadley, and Romain Francois. 2016. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.