

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?)	1
Literature Review and Policy Relevance (600 words?)	1
Data Gathering and Methodology (1000 words?)	1
Analysis	3
References	5

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, aequae 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, aequae 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 clouded 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 unaltered “.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 occurred. 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 kush tourism.com. With this information, we were 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 essentially 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 omitted variable bias through controlling for [district] level heterogeneity by means of dummy variables.”⁴ To implement this, we will be using the lfe package (Gaure and Economic Research 2016), checking its robustness in turn via the plm package (Croissant, Milla, and Tappe 2016). An alternative robustness check could have been 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.

¹should find quote for stereotypical soes

²The query was conducted on 10 November 2016

³@dehejia2002propensity: 153.

⁴Möhring (2012)

Analysis

Descriptive Statistics

Let us begin 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.

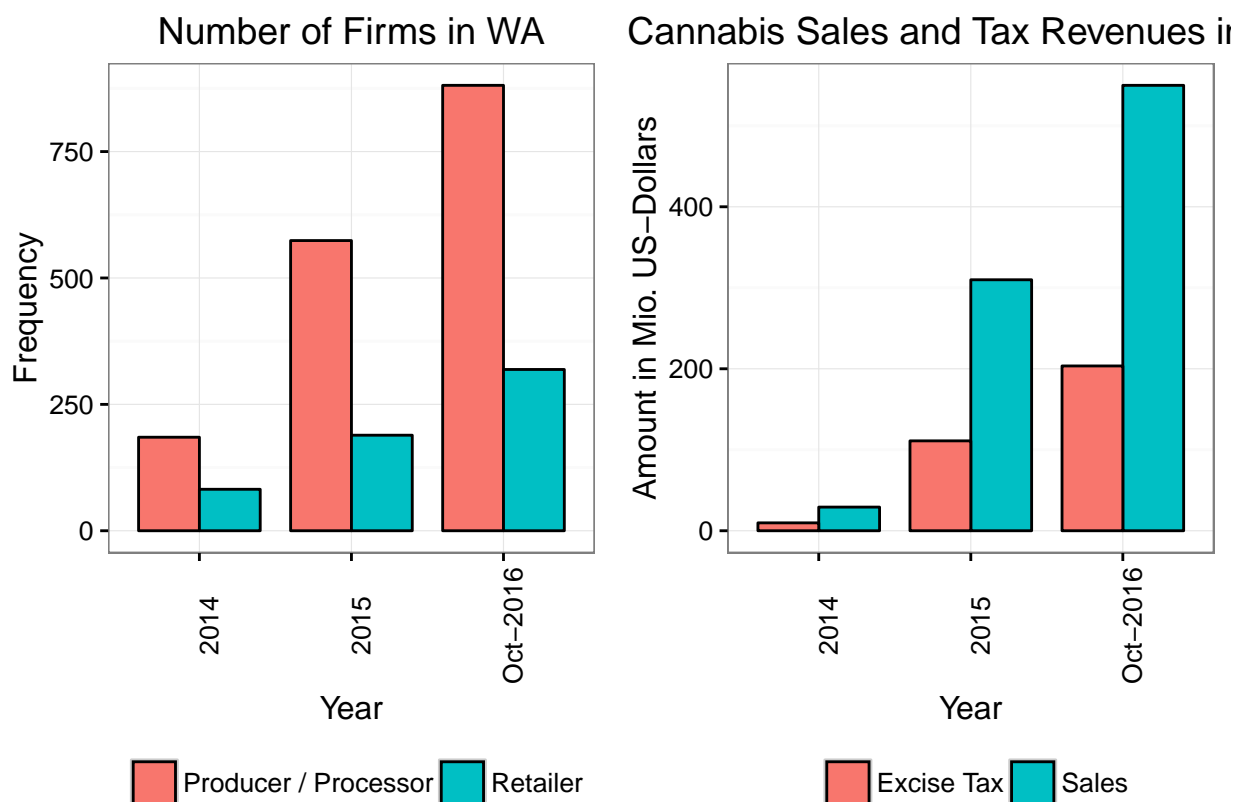


Figure 1: Business Development of Cannabis Production and Sales in Washington State

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 \$29.21⁵ Mill to \$309.86 Mill and by another 77% in the ten months leading up to October 2016 to \$549.93 Mill.
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.

⁵when did the first one open

Thus, we know that this is not just a niche industry.⁶ Unfortunately, we do not have data on how many quantities have been sold or reliable data on the price development because otherwise we could also analyse the actual use in quantities.

Let me now introduce to you the dependent and explanatory variables that we use in this analysis

	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
8	Age	Share of Kids, Adults and Retirees
9	Diversity	Share of Whites, Blacks and Multiple Races
10	GEOID	U.S. Census District

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

there was a problem with the names of retailers not uniquely identifying retail shops. To uniquely identify the length of our data set, we thus need to look at the unique URLs rather than the names. This is not yet a concern however, because the observations that are affected drop out in our later analysis of Seattle. While Seattle is divided into 132 districts by the U.S. Census, crime incidents have only occurred in 35 of them. Out of these 35 districts, we have eight that also have a dispensary.

⁶maybe add here the GDP of WA to make a % comparison

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>.
- Gaure, Simen, and Ragnar Frisch Centre for Economic Research. 2016. *Lfe: Linear Group Fixed Effects*. <https://CRAN.R-project.org/package=lfe>.
- 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>.