

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	2
References	3

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

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 kushtourism.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

other packages that have been used will be referred to below initially descriptive statistics then ols then matching then fel

Analysis

Descriptive Statistics

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.

¹should find quote for stereotypical soes

²The query was conducted on 10 November 2016

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>.
- 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>.
- 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>.