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

Benjamin Gaiser and Jeremy Russell

11 November 2016

OVERVIEW:

Building on our work since the second assignment and given time constraints, we decided to focus solely on the city of Seattle, Washington for this third assignment. A full comparative analysis of Seattle and Denver is to come in the final assignment. As mentioned, residents of nine other states in addition to Colorado, Washington, Oregon and Alaska recently voted on measures to legalize either recreational or medicinal cannabis use. California, Maine, Massachusetts and Nevada voted to legalize recreational use, while the outcome in Arizona is still forthcoming. Florida, Arkansas and North Dakota voted to legalize the medicinal use of cannabis, while the outcome in Montana is also forthcoming.

REPOSITORY:

This repository contains the R code for web scraping, cleaning and analyzing our data. Currently we have data for marijuana production and distribution in the state of Washington, and crime and marijuana statistics for the city of Seattle. We added U.S. census data to account for socio-economic factors.

DATA PROCESSES:

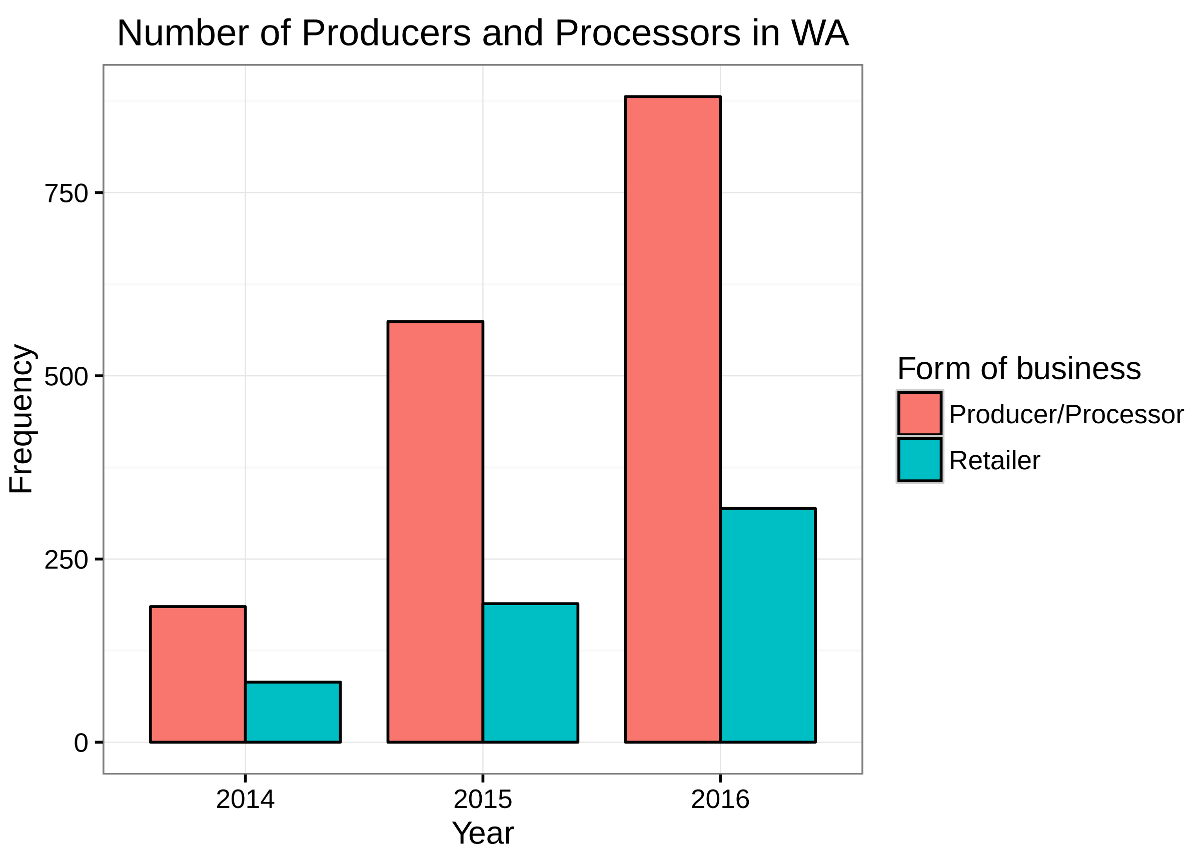
We use the R packages built by (Gandrud 2016) to construct this PDF. Further packages have been used along the data gathering, cleaning and analysis part and will be cited either there or in the references.

We began by scraping the data for cannabis retailers and producers separately, using the packages (Wickham 2016a) and (Harrison 2016). The latter builds on a phantomjs ".exe"-file, which coincidentally is not compatible with a Mac. So we ran the code on a Windows and made a simple ".txt" file of the data before cleaning it. After cleaning the data, we made variables for tax, sales and recorded month. This heavily relied on the packages by (Wickham and Francois 2016) and (Bache and Wickham 2014). 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. Next, we subset the data on retailers from Washington into King County; downloaded the Seattle crime data with (Ooms, Temple Lang, and Hilaiel 2016); subset this data into narcotics and/or alcohol related data, since we wanted to focus on drug-related crimes; and wrote this information into a zip file, as it has taken quite considerable processing time. We would like to note that we are considering adding data on the effects of legalization on more serious crimes (property and violent crimes) in addition to our current analysis of drug related crimes to more accurately specify the effects of legalization. But for now, this helps us to get a first understanding of the relationship between crime and marijuana legalization. Then we scraped U.S. census data, using an array of different packages cited below in the Reference section (2016), (2014), (2015), (2016), (2016b), (2016), (2016), (2016), (2016) and determining which variables to include in our analysis; we decided to include (1) high household incomes as a share of all (> \$200k), (2) age, (3) race and (4) education. We consulted the following blog post as a reference point for our code. In order to merge all of these data, we need to ensure that we had geospatial data for all of them. For the web-scraped data on retailers in Seattle, we did not have this information. So we have scraped information about the data address from another site. With this we sent a Google API query to determine the latitude and longitude for each using commands from the package by(Temple Lang 2014). This left us with 15 individual retailers for Seattle for which we have data. We then saved the geolocation data as a ".txt" file; created parameters to assess the minimum distance using the package by (Nychka et al. 2016) between latitude/longitude points between the Seattle.Retailer data frame and the Seattle.Crime.Narcotics data frame, creating the final data frame Seattle.Merged. We then made a binary variable representing from which month onwards a retailer was established. As for the merging of our U.S. census data with the final data frame, we had to make a strong assumption to link the socio-economic characteristics of a district to a geopoint of crime. Our assumption here for the U.S. census data on district-specific crimes is that a perpetrator of some crime is probably more likely to commit that crime in their own neighborhood than to travel far away to do so, but we recognize the limitations of this logic.

DESCRIPTIVE STATISTICS:

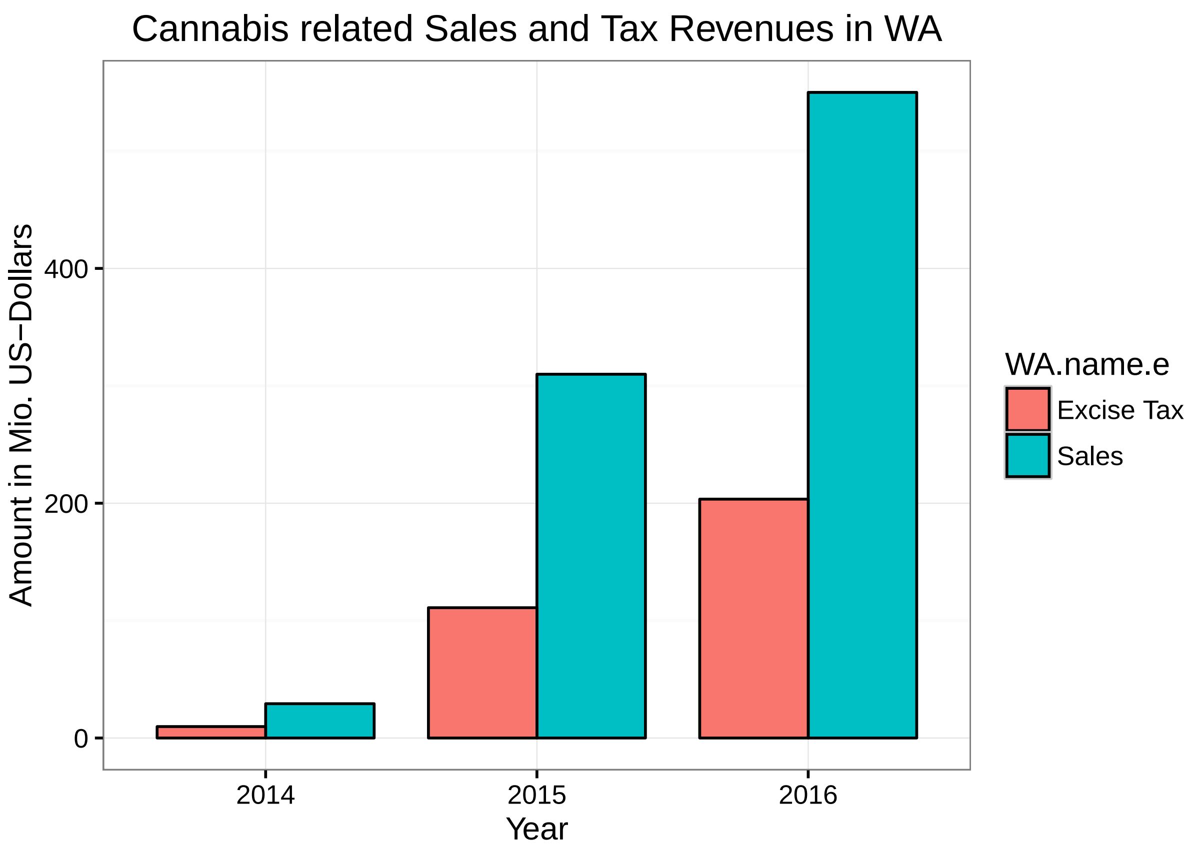
The graphs that follow are built using the package ggplot2 by (Wickham and Chang 2016). We used a simple bar chart to show the growth of producers/processors and retailers of cannabis in Seattle between the years 2014-16. Between 2014-15, producers grew rapidly. They surged by 201% (from 185 to 574); between 2015-16, they grew more slowly - by 53% (from 574 to 881). Retailers grew by 130% (from 82 to 189) and 69% (from 189 to 319) respectively between 2014-15 and 2015-16.

```
washington.barplot <- ggplot(WA.bar, aes(x=Year, y=WA.numbers, fill=WA.name)) +
  geom_bar(stat="identity", colour="black",width=.8, position=position_dodge()) +
  theme_bw() + xlab("Year") + ylab("Frequency") +
  scale_fill_hue(name="Form of business") +
  ggtitle("Number of Producers and Processors in WA")
washington.barplot
```



Data on sales and tax revenue also show a massive increase over these two years. By 2015, sales revenue leaped by 960% from $29.21 million to $309.86 million, and in 2016, by 77% from $309.86 million to $549.93 million. Tax revenue in 2015 grew by an astounding 1040% from $9.74 million to $110.97 million, but slowed down a bit in 2016, increasing by 83% from $110.97 million to $203.47 million.

```
washington.barplot.e <- ggplot(WA.bar.e, aes(x=Year, y=WA.numbers.e, fill=WA.name.e)) +
  geom_bar(stat="identity", colour="black",width=.8, position=position_dodge()) +
  theme_bw() + xlab("Year") + ylab("Amount in Mio. US-Dollars") +
  ggtitle("Cannabis related Sales and Tax Revenues in WA")
washington.barplot.e
```

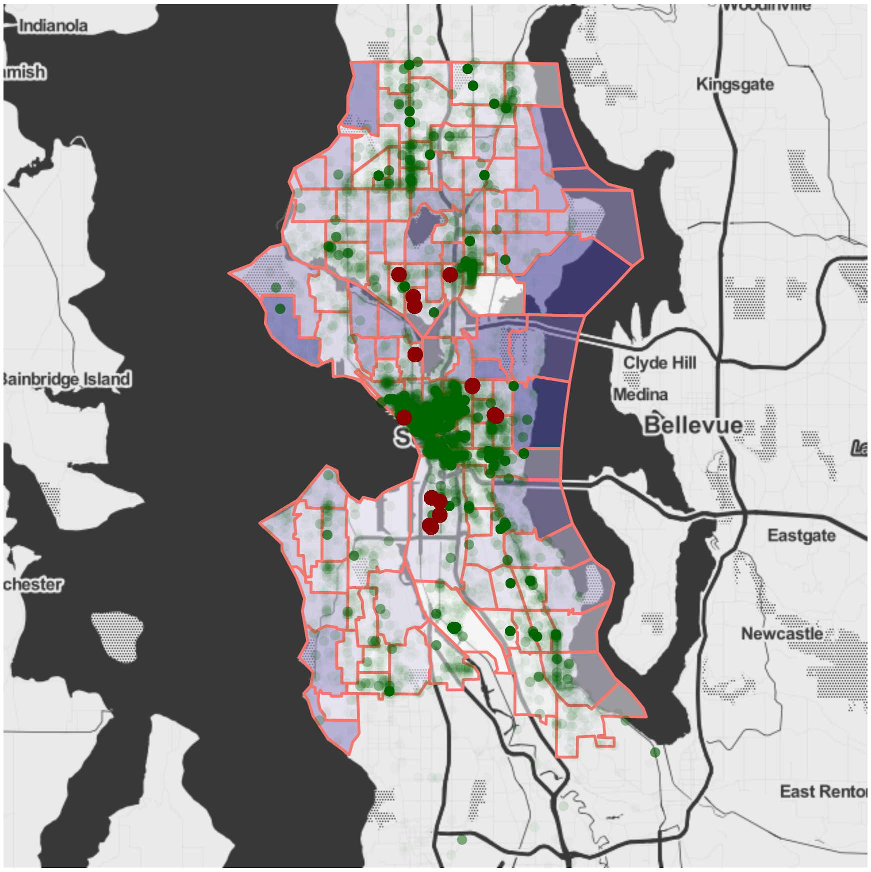


Next we created a map showing all narcotics-related crimes in green, all retailers in red, and wealth as a heat map where the share of household income was above $200,000+ per year (i.e. the more intensive the colour, the higher the share). As we can see, crime was concentrated in the central area of Seattle, and was not clustered around retailers. Furthermore, the type of crime in our data set does not seem to affect the wealthy areas of Seattle.

```
map.seattle_city <- qmap("seattle", zoom = 11, source="stamen", maptype="toner",
                         darken = c(.3,"#BBBBBB"))
map.seattle_city +
  geom_polygon(data=Seattle.SOE, aes(x=lon,y=lat, group=group, fill=percent,
                                     color=muted("blue"), alpha=0.9)) +
  theme_nothing(legend=FALSE) +
  scale_fill_gradient2(breaks=pretty_breaks(n=5),
                       guide=guide_legend(title="Percent living > $200k", keywidth = 2)) +
  geom_point(data=Only.Narcotics, aes(x=lon.Narc, y=lat.Narc),
             color="dark green", alpha=0.03, size=1.1) +
  geom_point(data=Seattle.Merged, aes(x=lon.Ret, y=lat.Ret),
             color="dark red", size=2) +
  labs(title = "Narcotic Crime in Seattle")
```

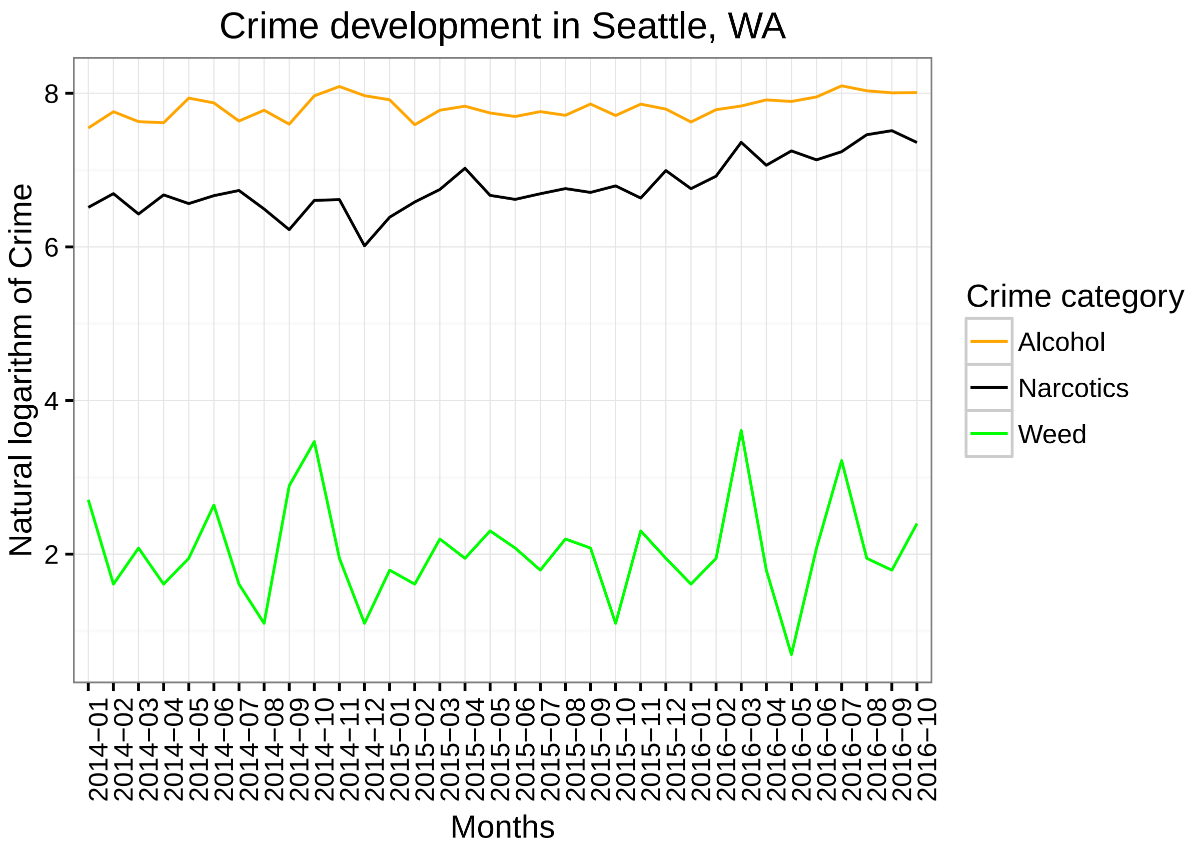
```
## Warning: Removed 5028 rows containing missing values (geom_point).
```

```
## Warning: Removed 428406 rows containing missing values (geom_point).
```



Lastly, we plot crime development in Seattle, divided into narcotics-related, alcohol-related and weed-related. As one can see from the line graph, they do not seem to show any trend, though weed has a higher variation than the other two variables.

```
crime.graph <- ggplot(Crime.lines, aes(Month)) +
  geom_line(aes(y = AlcoholCrimePerMonth, colour ="AlcoholCrimePerMonth",
              group=AlcGroup)) +
  geom_line(aes(y = Only.NarcoticsCrimePerMonth, colour ="Only.NarcoticsCrimePerMonth",
              group=NarcGroup)) +
  geom_line(aes(y = WeedCrimePerMonth, colour = "WeedCrimePerMonth",
              group=WeedGroup)) +
  scale_colour_manual(name="Crime category",
                    labels=c("Alcohol", "Narcotics", "Weed"),
                    breaks=c("AlcoholCrimePerMonth", "Only.NarcoticsCrimePerMonth",
                             "WeedCrimePerMonth"),
                    values=c("orange", "black", "green")) +
  xlab("Months") + ylab("Natural logarithm of Crime") +
  labs(title="Crime development in Seattle, WA") + theme_bw() +
  theme(axis.text.x = element_text(angle=90))
crime.graph
```



OUTCOMES: As retrieving, cleaning and merging the data took quite some time, our focus was more on descriptive statistics than on inferential. Nevertheless, we have created a variable to do a simple univariate regression upon which we will be able to expand on in our next assignment. WeedCrime is a binary variable and assumes 1 for all crimes related to marijuana and 0 otherwise. We can see that the fact that a retailer was established nearby was highly significant but on a negligible level (both in terms of strength of the relationship and the variance explained), thus tentatively supporting the idea that marijuana legalization does not open the gates of Sodom and Gomorrah concerning crime. Yet, lacking any further control variables, little can be said about this association at this point in time.

```
fit <- lm(WeedCrime ~ Est, data=Seattle.Merged)

stargazer::stargazer(fit,
                     title="Univariate Regression of Crime Incidence",
                     digits=2, type="latex", header=FALSE)
```

Table 1: Univariate Regression of Crime Incidence

	Dependent variable:
	WeedCrime
Est	0.03***
	(0.001)
Constant	0.01***
	(0.0001)
Observations	435,068
R^2	0.002
Adjusted R^2	0.002
Residual Std. Error	0.07 (df = 435066)
F Statistic	898.12*** (df = 1; 435066)
Note:	*p<0.1; **p<0.05; ***p<0.01

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.

Gandrud, Christopher. 2016. *Repmis: Miscellaneous Tools for Reproducible Research*. https://CRAN.R-project.org/package=repmis.

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 Winston Chang. 2016. *Ggplot2: An Implementation of the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

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