

Effect of Marijuana Legalization

Introduction/Motivation

In the 20th and 21st century, marijuana and its usage have globally disrupted society. Widespread usage impacts regulation and legalization in different parts of the world, health effects on specific demographics, and economic growth due to social media's influence on stocks related to marijuana. The goal of this study is to elucidate marijuana's impact on society in terms of security, privacy and public health.

Problem statement:

What are the long- and short-term effects of marijuana usage and legalization in terms of crime, health, social sentiment and related stocks?

Contributions

All team members contributed similar effort. Specific tasking is summarized in the Appendix.

Current State-of-the-Art

Current analysis and visualization are done in a static way. In addition, the collective impact of marijuana usage on society is unknown relative to the following four topics:

1. Impact on Crime

Historically, there have been reports on the effects on crime from medical marijuana usage [6]. With legalization of marijuana in certain states, there have been preliminary reports of reduced crime [5]. More data is needed comparing the reduction in crime in legalized and non-legalized states. Longitudinal studies have also shown significant correlations between marijuana use and criminal charges during the formative years into young adulthood [4]. These findings will be useful in quantifying how marijuana usage effects not only crime rate but also geography and social strata. Bringing these analyses together will provide more insights on crime and correlations to legalization.

Data for this effort can be scraped from the FBI crime repository. The existing analysis only looks at medical marijuana [16] and does not really incorporate the new issues with legalization.

2. Health Effects

The effects of cannabis use on health have been hotly debated and studied dating back to the India Hemp Commission of 1894 and prevalently studied in the modern era [21] [22]. There is general consensus about the acute negative effects of illicit drug use, but more research and analysis must be performed to fully understand its effects.

3. Social Impact

Marijuana has gained lots of popularity on social media platforms. Researchers have investigated popularity and trend [7-10], social psychology [10], based on Twitter content [12], locations [8], attitude [9, 12] and time [10]. More recently, Facebook enabled emoji functions which provides another dimension to characterize people's attitudes toward marijuana [11].

4. Economic Effect

Given commercialization efforts, as well as decades of legalization issues with marijuana, economic effects can be better understood through analytics and visualization [13]. As discussed earlier, social media has impacted the popularity of marijuana. One aspect of social media is the ability to mine data for public opinion using sentiment and stock price analyses, which are potentially extendable to the topic of marijuana [14].

Proposed Method

Intuition/ List of Innovation:

Novelty and enhancements from this study are listed. For this project, we attempt to collect data from public databases and utilize free/open source tools to minimize cost.

1. Analysis and visualization of groundbreaking new data from current events, including Canadian legalization of marijuana, which occurred in the Fall of 2018. Our tweet and stock analysis focus on how such events affect public opinion towards marijuana.
2. Combination of short and long time frame to determine effects of marijuana legalization in variety of time scales. Crime and health data are examined in a long-term scale, spanning 20 years from 1995 to 2014. Twitter and stock data focuses on the short-term, from Sept. 2018 to Nov 2018. In this study, we examine both short-term effects (social awareness and stock) and long-term effects (crime rate and public health), in order to provide a full picture of marijuana use.
3. Correlation analysis between variety of data source. For example, correlation was built between public sentiment of marijuana, and stocks related to marijuana. We also examined correlation between health/crime data and marijuana usage and legalization.
4. Interactive data visualization using D3, HTML/CSS, Python plotting libraries and Tableau. Techniques from this course were utilized to build an interactive platform visualizing information about marijuana in terms of crime rate, health impact, social sentiment and stock effect.

Description of approaches:

1. Impact on Crime

Difference in difference (DID) is incorporated to see the change in state crime rates relative to marijuana legalization. For the analysis, simple DID regression is used on the predictive variable to see the effect of the different violent activities.

$$Total\ Crime = \beta_0 + \beta_1 Recreational + \beta_2 Medical + \beta_3 Year$$

Where β_0 , β_1 , β_2 , β_3 are coefficients for the linear regression. With limited data recreational legalization, the analysis was restricted to the few states and looked at the overall trend and behaviour of the information.

2. Health Effects

Text analytics using tokenization and word count analyses are performed on unstructured data taken from the study Minnesota Medical Cannabis Program: Patient Experiences From the First Program Year [21] and the Report of the National Commission on Ganja [22]. These references provide perspectives from both sides of the marijuana debate -- from the healthcare providers of patients with severe health conditions to those recreational users (some of whom would cite religious purposes for their marijuana use).

3. Social Impact

Twitter data is crawled with three key words “marijuana”, “weed”, and “cannabis” with Twitter id, time created, user id, user location and text. The timeframe of the Twitter data spans from Oct 25, 2018 up to Nov 17, 2018 due to restrictions of Twitter’s search policy. Using crawled Twitter data, sentiment analysis is performed using Naive Bayes classification algorithm to create labels from most negative (0) to most positive (4). The whole pipeline of data processing is illustrated in the following graph, reproduced from ref [15]. More details can be found in this reference [15].

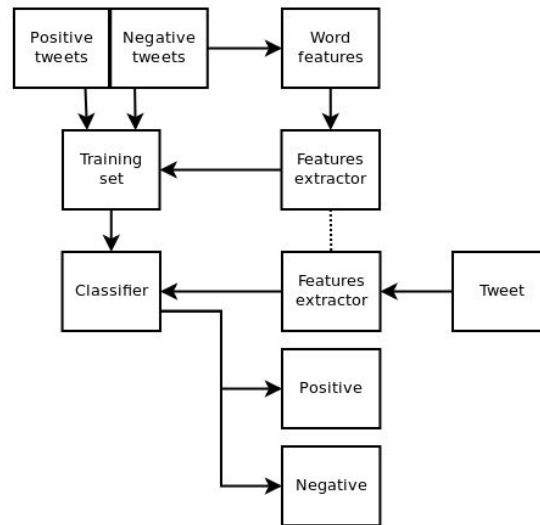


Figure 1: Pipeline of Data Processing

4. Economic Effects

Using similar techniques for understanding the social impact, understanding economics uses the same tweet sentiment classification tools. Approximately 2,500 tweets are collected from the Twitter API during Sep 5 2018 to Sep 7 2018 with query "Tesla marijuana", and Oct 15 2018 to Oct 18 2018 with query "Canopy Growth". These public companies are selected because these companies had current events related to marijuana, and have plenty of Twitter data. Additionally, Canopy Growth is a global leader in the sale of marijuana products, and Tesla, is in the automobile industry and energy sector, run by an influential billionaire. Stock data is crawled from the Investor's Exchange API (IEX) using Python and pandas.

To understand the relationship between Twitter sentiment and stock performance, the tweet sentiments, and the dividend yield of stocks (a variable of closing price) are aggregated with Python, numpy, and pandas. Then, the strength of correlation is calculated through the Pearson Correlation Coefficient (r), which is a value between -1 and 1. If r is -1, there is absolute negative linear dependence, and if r is 1, then there is absolute positive linear dependence. The data is visualized as scatter and matrix plots using matplotlib.

Experiments / Evaluation

List of Questions to Answer:

1. How does legalization affect the crime rate in the US?
2. What is the long-term effect of medical marijuana on the overall US crime?
3. What, if any, associations link health effects and marijuana use? What is the societal benefit?
4. Can news media or public events generate more public interests about marijuana?
5. Are social attitudes affected by marijuana legalization? Does legalization of medical usage and recreational usage have different effects?
6. Which area of people are most interested in discussion about marijuana? What is their general attitude towards marijuana?
7. Can social media influence be attributed to the increase or decrease in a publicly traded company's portfolio when marijuana is discussed? More specifically, is this the case if the company sells marijuana only, or is it also the case if the company is in a different industry and publicly endorses or condones marijuana usage and/or policy?

Experiment Detail:

1. Crime Analysis

To determine the impact on the recreational usage of marijuana in United States, all individual state's Federal Bureau of Investigation's Uniform Crime Reports (UCR) datasets were combined. The scope of this experiment was 1994 - 2014. Next, information on the legalization of cannabis in USA was used and variables were created which denote the year when the marijuana was legalized in the USA for medical and recreational purposes.

	Alaska	Colorado	DC	Oregon	Washington
Total Robbery	11848	68578	83733	57116	122289
MML Years	20	2	17	20	2
Recreational Years	1	2	1	1	2
Total Rape	10215	39747	4522	24447	53009
Total Murder	748	3362	4152	1804	3934
Total Assault	61904	210087	89134	141367	259622

Table 1: Crime Summary in States with Legalized Recreational Usage of Marijuana

	Last 3 years	Alaska	Colorado	DC	Oregon	Washington
Total Robbery	2012	630	3392	4037	2419	5801
	2013	623	3136	4082	2394	5829
	2014	629	3039	3497	2093	5640
Total Rape	2012	583	2122	236	1159	2250
	2013	657	2198	297	1000	2072
	2014	555	2121	352	1052	2171
Total Assault	2012	3169	10285	3505	5969	12285
	2013	3127	9826	3835	5596	11623
	2014	3243	10325	4125	5592	11627
Total Murder	2012	30	152	88	91	217
	2013	34	174	103	82	167
	2014	41	151	105	81	174

Table 2: 3-Year Crime Summary in States with Legalized Recreational Marijuana Usage

The data distribution of the states having recreational information gives us information on how things have changed since marijuana legalization. Using DID, medical marijuana legalization is shown to be highly influential on crime overall with p value 0.00847. The overall regression analysis on all the critical variables shows a high influence of the legalization.

Some of the insights observed by only looking at the recreational data is below. In the recent trends, total crime has increased overall due to many factors. Drilling down to the actual date of legalization, there is a spike of crime during the events.

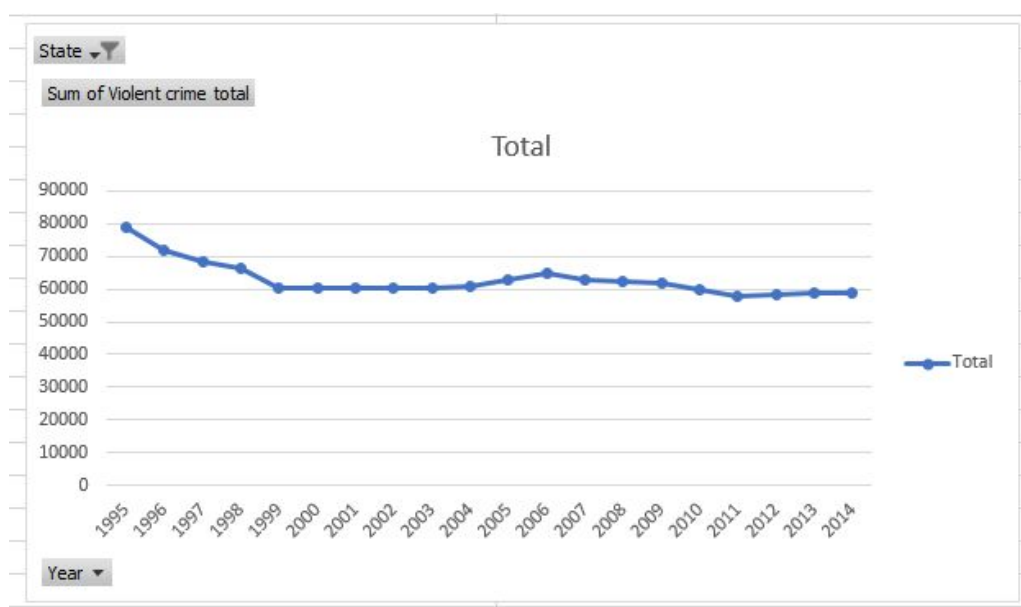


Figure 2: Summary of Violent Crime Totals

On the recreational legalization, there are five states with data. After visualizing the information based on the aggravated assault, the rate of crime has increased significantly from 1994 - 2014. There is a clear upward trend and strong influence of external factors like marijuana which might have driven up the aggravated assault.

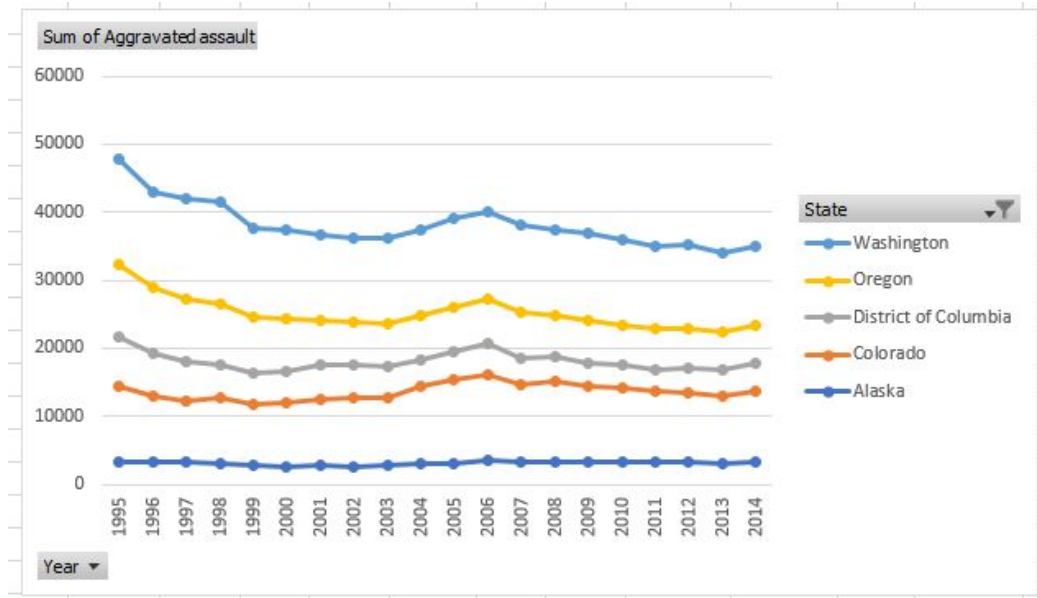


Figure 3: Summary of Aggravated Assault

On the other side, robbery has significantly gone down since legalization. There can be various factors influencing this including the theory that marijuana legalization could have helped reduce robbery given that once people are under the influence of marijuana it is difficult to rob but more easy to assault others.

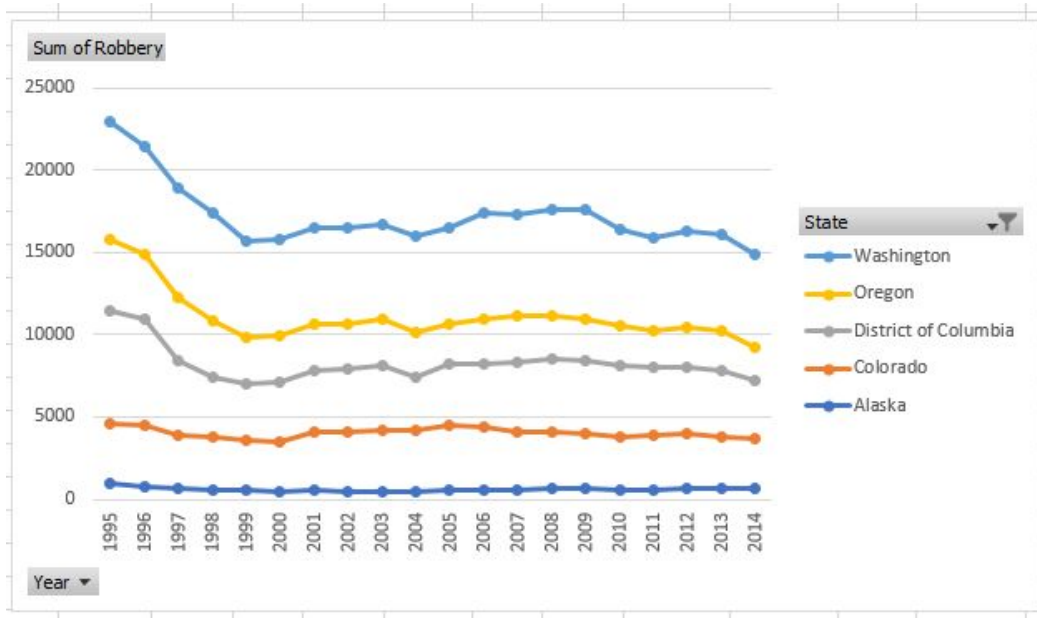


Figure 4: Summary of Robbery

A visual analysis was created to demonstrate the correlation of the data over a period of time using a slider to vary the year and view how the crime data has changed in the last 20 years, as in reference [23] to total crime.

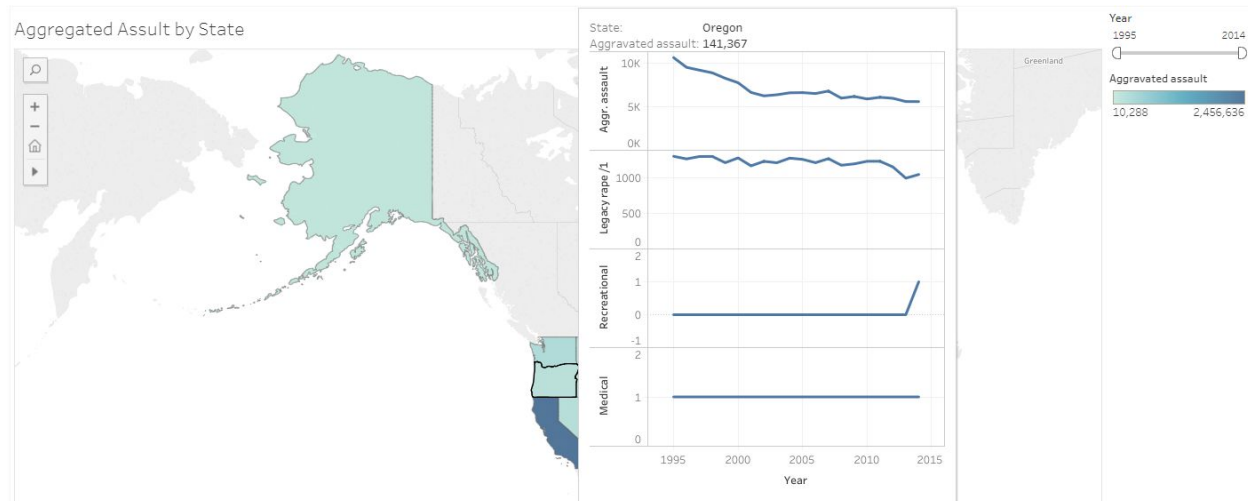


Figure 5: Correlation of Crime Data (last 20 years)

2. Health Analysis

The unstructured data from each referenced publication was input into an .xls file. Using Python libraries (i.e. nltk, xlrd, and xtutils) and regular expressions, the data was cleaned (i.e., removed stopwords, special characters, newline returns, and any numbers) and tokenized. Using Python, the cleaned text and keywords were re-written back to the original .xls file. From this point, the analysis branched further and resulted in: 1) a word cloud and 2) the interactive human body map which visualizes the extent of the key benefits and negative effects of marijuana use.

Word Cloud

OpenRefine was used to cluster and edit the keywords using the nearest-neighbor method. After merging and re-clustering the data, Tableau was used to create the word cloud. Provided below is a word cloud which visualizes the key marijuana benefits (to

patients) as observed by their healthcare practitioners as well as the negative effects of marijuana use.

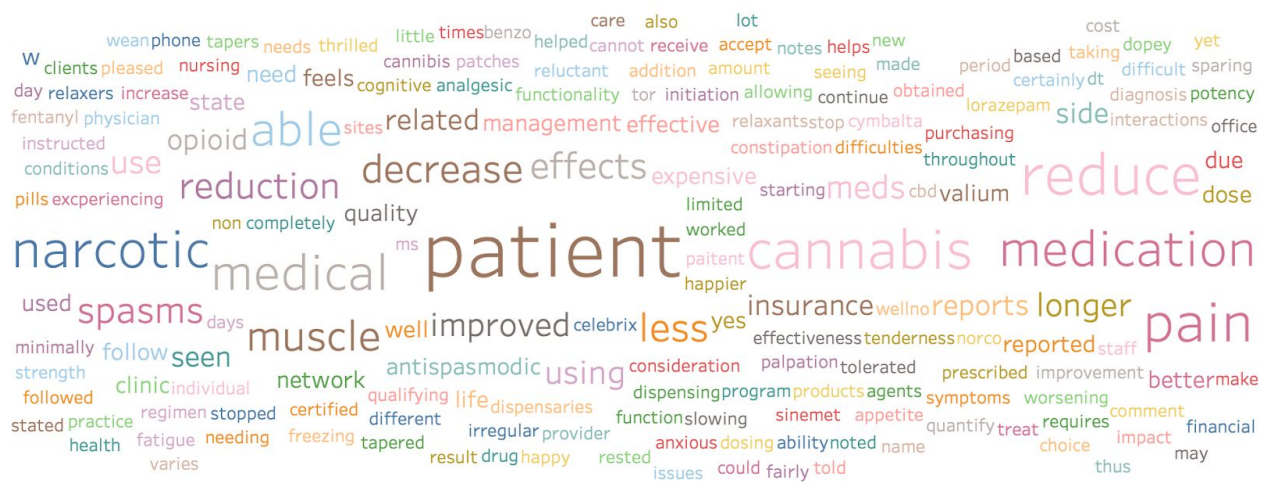


Figure 6: Marijuana Importance for Medicinal and Recreational Users

Several key insights arise out of this visualization:

- Fewer non-medical terms in the word cloud attributable to the lack of objective survey data (e.g., from health practitioner observations) of recreational users
- Of the medical terms, the keywords 'reduce' and 'pain' are among the highest in importance (i.e., high word counts)
- The keyword 'cannabis' is listed prominently among other prescription pain and symptom management drugs including Benzo, Fentanyl, Valium, Cymbalta, Celebrix, Sinemet, and opioids

Interactive Human Body Map

Tableau was used to annotate x- and y-coordinates on the image of the human body relative to its dimensions. The proof of concept was demonstrated using Tableau and fully developed using JavaScript/CSS/HTML from an interactive D3 human body map [25], as shown in ref [23] . Below is a static image of the map.

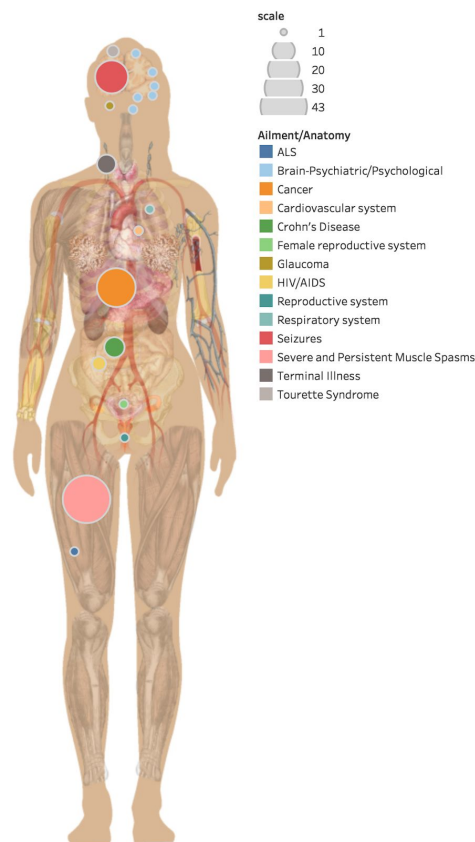


Figure 7: Health Effects of Marijuana (Medical Use vs. Drug Abuse)

Key insights of societal benefit can be gleaned from this interactive map:

- Similar to the insights from the word cloud above, healthcare practitioners are prescribing medical marijuana to treat pain and nausea associated with severe conditions.
- Analysis reveals the top three ailments for which Minnesota healthcare practitioners prescribed marijuana are: pain, nausea, and seizures. Patients now have viable natural remedies to address these common discomforts.

- The majority of the negative effects of recreational marijuana use are brain and psychiatric related. These effects create short-term impairment of cognitive function, psychic effects, signs of intoxication, and exacerbates existing mental health issues.
- This interactive map can be used as an educational tool for families considering traditional and non-traditional treatments to chronic diseases. With further research, an marijuana effectiveness scale could be established and incorporated into this interactive map to assist doctors in informing patients of their medical treatment options.

3. Twitter Analysis

Following graph shows number of tweets per day from October 25, 2018 and November 17, 2018.

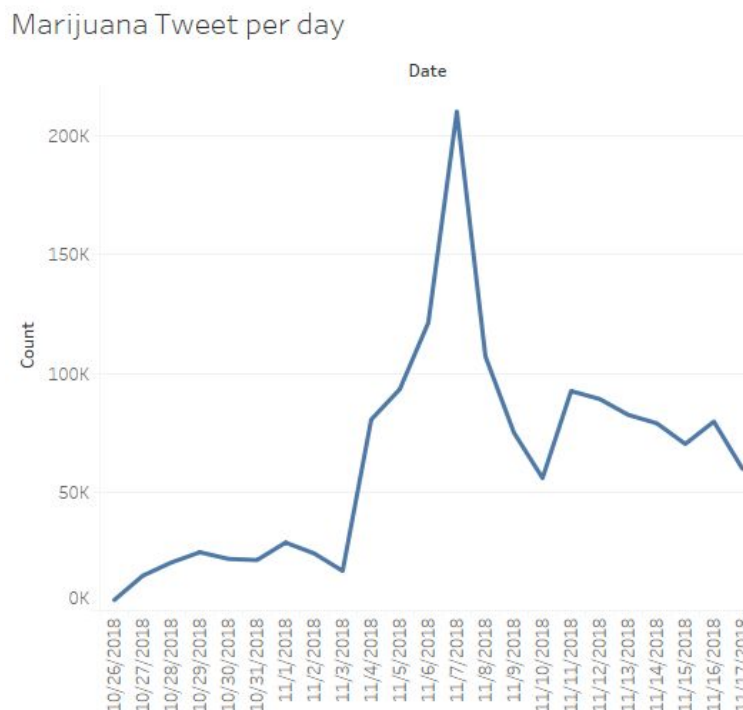


Figure 8: Tweets per day (10/25/2018 - 11/17/2018)

From this graph, it is safe to say that people are not talking about marijuana all the time. There are certain dates in which more discussion about marijuana are stimulated.

During the time period examined, a few peaks were observed: November 7, November 11 and November 17. Interestingly, these dates correspond to news coverage about marijuana. [18-20]. It is very clear that social events and media coverage stimulate people's interests in marijuana and generate more online discussion.

The following graph shows counts of daily marijuana tweets by hour. It is interesting to note that 9am observed least marijuana tweets while 5 pm shows most marijuana tweets, which is probably related to the use of marijuana after work.

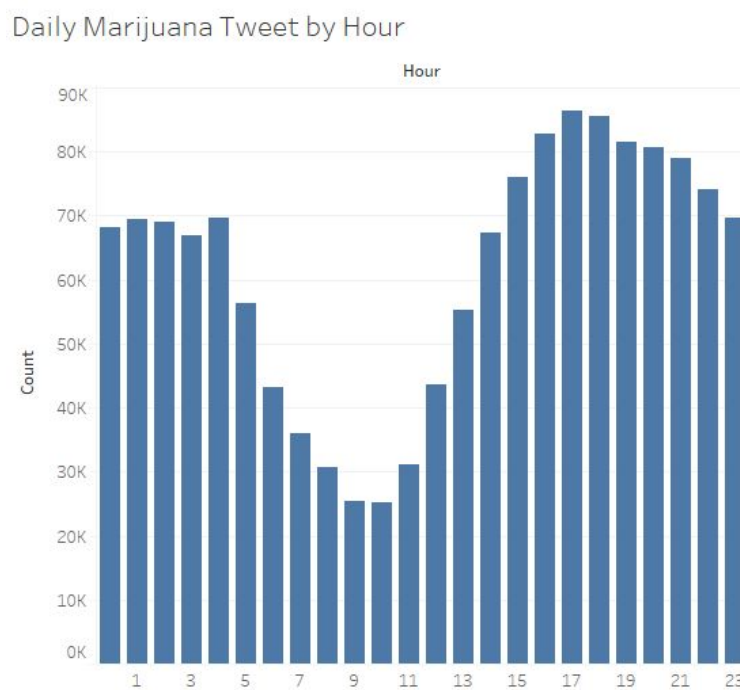


Figure 9: Daily Marijuana Tweets by Hour

In order to examine popularity of marijuana for different user, a plot number of users vs number of tweets by each user as shown below. The trend follows power law very well with following equation:

$$\ln(\text{No. of users in each bin}) = -2.38461 \cdot \ln(\text{No. of tweets by each user}) + 12.661$$

This power law relationship indicates that most people only talk about marijuana once and very a few tend to talk about it many times.

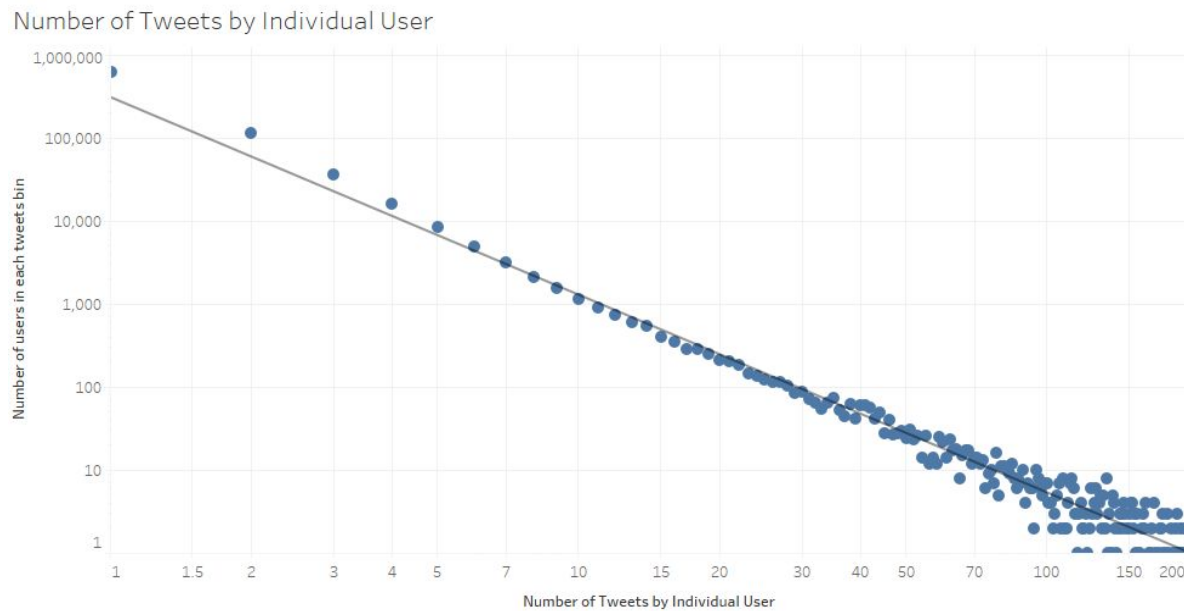


Figure 10: Number of Tweets by Individual User

Tweets also include geographic information. Tweets outside of the United States were screened out. Graphs below show average sentiment and count of tweets vs locations. As for average sentiment, Vermont and Colorado shows highest sentiment while Texas has lowest sentiment. This is in contrast with the count of tweets where California and Texas has most tweets.

Average Sentiment vs State

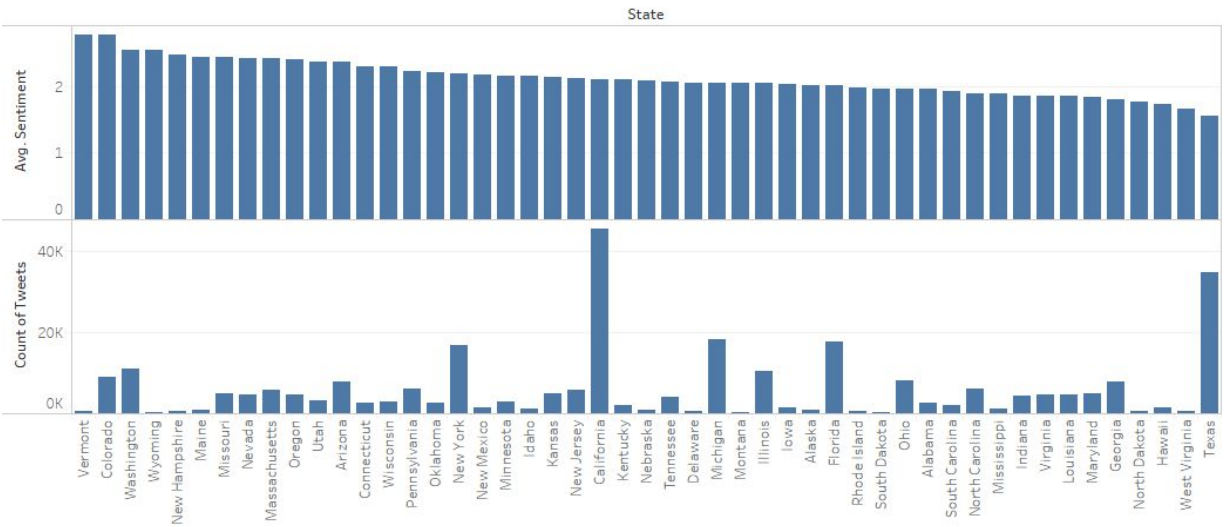


Figure 11: Average Sentiment vs. State

The average sentiment dataset was implemented on geographic map to visualize the data interactively in ref [23]. Following graph shows a screenshot of the visualization.

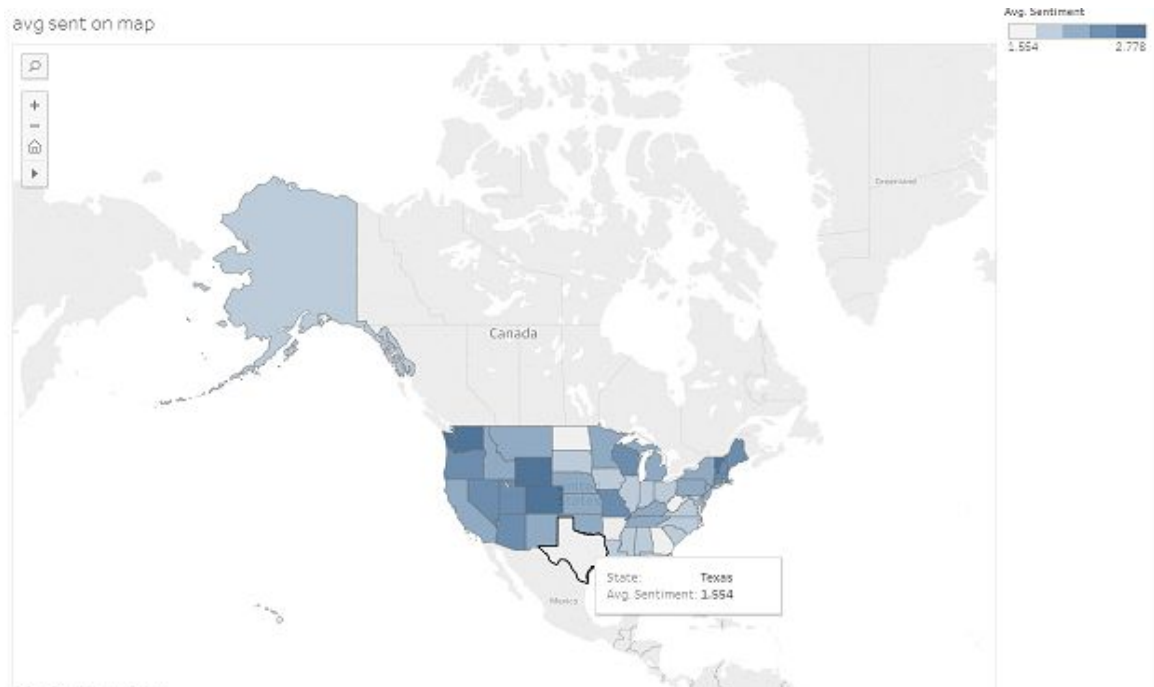


Figure 12: Average Sentiment Data on a Geographic Map

To check correlation between marijuana legalization and average sentiment, average sentiment was plotted vs how many years of legalization for medical use and recreational use extracted from ref [24].

Sentiment correlation

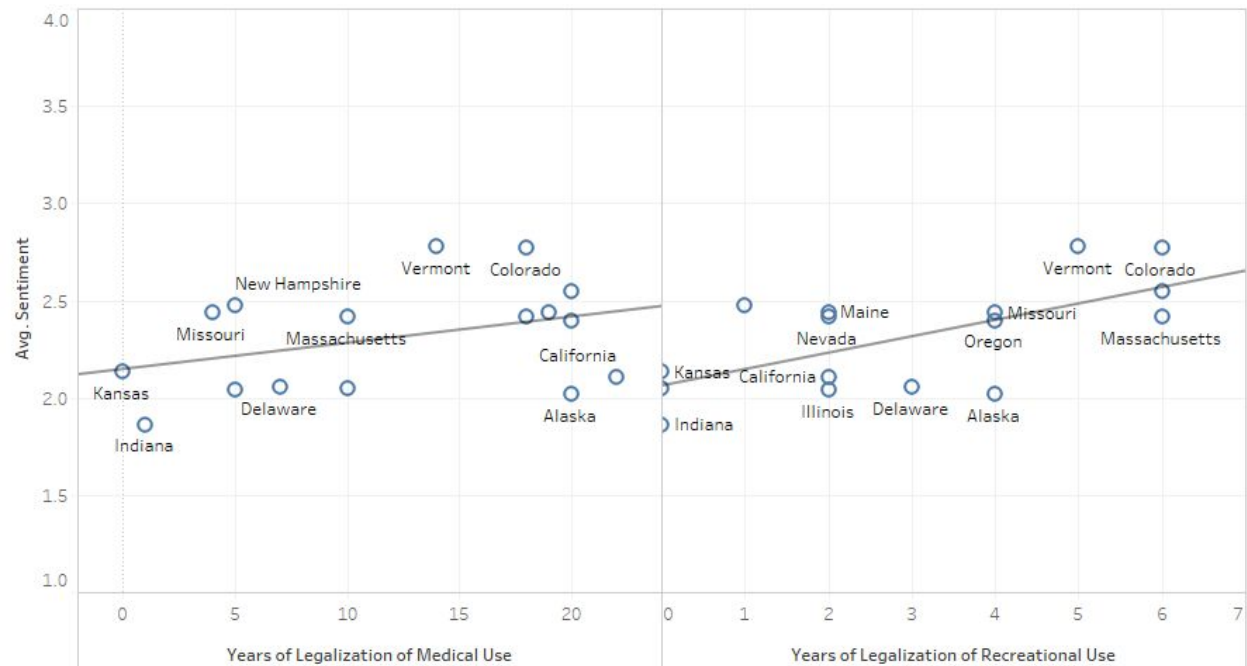


Figure 13: Sentiment Correlation

The left figure shows the effect of legalization of medical use, and the right figures shows that of recreational use. Following table shows the statistics from linear regression fitting for each graph.

	Medical use	Recreational use
slope	0.0134948	0.0843252
R squared	0.139986	0.42969
P value	0.153388	0.0058385

Table 3: Linear Regression Fitting

First, both slopes are positive, indicating that longer the legalization is, the more positive people treat marijuana. The R squared and p value comparison also show that legalization of medical use doesn't affect public sentiment much, whereas legalization of recreational use shows stronger correlation.

4. Stock Correlation

With the use of the Twitter sentiment classifier that was trained and the Pearson Correlation Algorithm, a value r of ± 0.67 was calculated. It is possible that alternate time windows of tweet and stock performance collection could have improved the magnitude of r , and observable correlation. The results still imply strong correlation between social media influence and stocks regardless of whether the company's primary product is marijuana. Figures 14 and 15 show scatter plots with a trend line to show the linearity of the data, and Figure 16 shows a matrix graph to show how much of the data is specifically correlated with a color map.

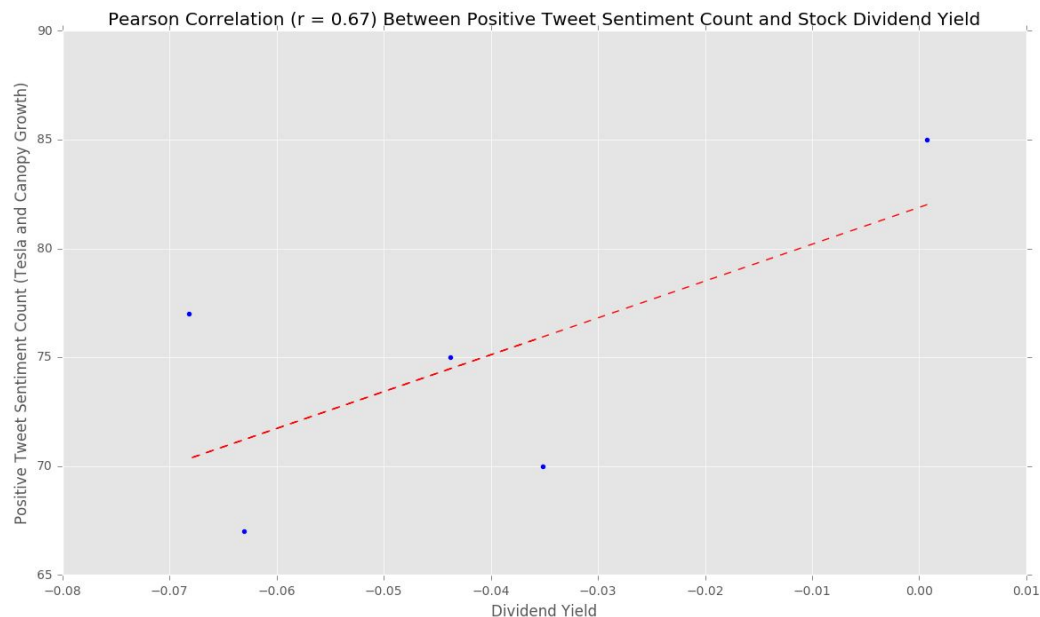


Figure 14: Scatter Plot: Positive Sentiment and Yield Correlation

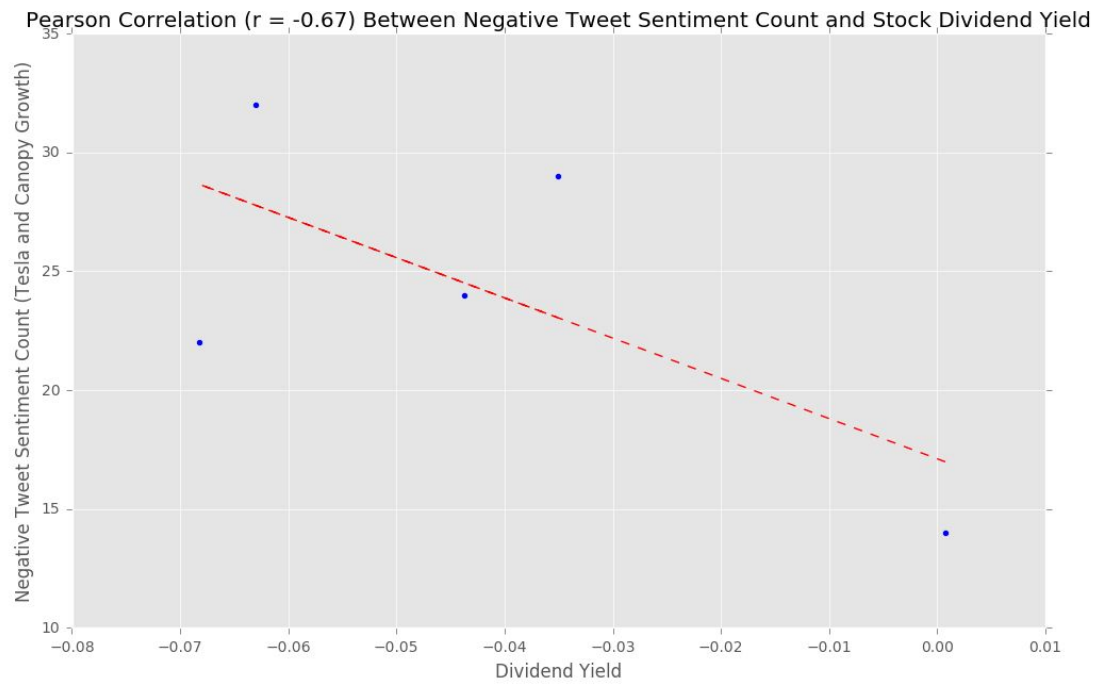


Figure 15: Scatter Plot: Negative Sentiment and Yield Correlation

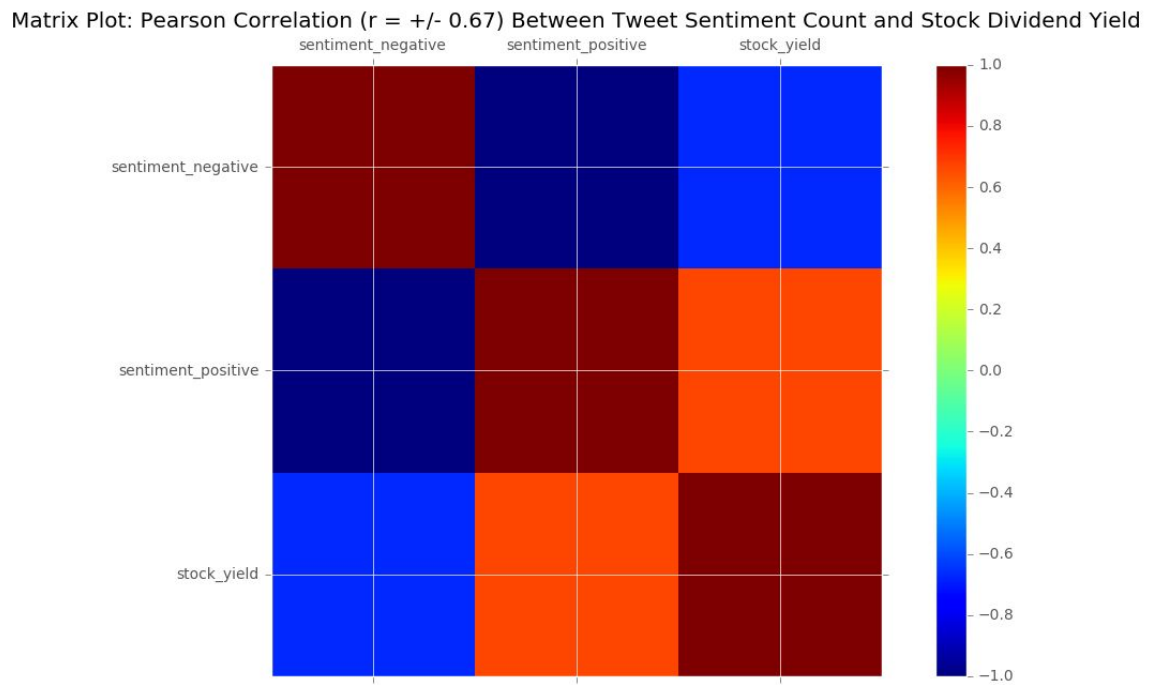


Figure 16: Matrix Plot -- Sentiment and Yield Correlation

Conclusion

The survey studies with the data before the legalization of marijuana on the recreational front concluded that they are not seeing any trend in the crime related to marijuana. But with the new laws and data we obtained we can definitely see a trend developing due to legalization. With many states coming forward to legalize in the future with a much larger dataset we can see the influence on crime increase.

Analysis of health effects of marijuana use provides effectiveness evidence of marijuana legalization on our society. As more states move towards legalization, a richer dataset will be available for further analysis. Future considerations can test how marijuana use correlates to the criminality of legalization by exploring: a) the prevalence of patients sharing or selling their prescription marijuana and b) whether patients are purchasing marijuana from dispensaries and pharmacies only.

In the short term, news media coverage and public events generate more public interests about marijuana. Marijuana legalization does affect social attitude, in which legalization of recreational use has stronger correlation than medical use. California and Texas has most tweets about marijuana, during times of interest. The state that treats marijuana most positively is Vermont, and the state that hates marijuana most is Texas.

There is an empirical correlation between Twitter sentiment of companies involving marijuana usage, or policy and stock performance with a Pearson Correlation Coefficient of ± 0.67 . The experiment design in terms of the time window size for data collection and analysis possibly led to error, and the algorithms could be optimized further to show a stronger correlation and dependence.

With these effects of marijuana use and legalization, it is important that policy makers have a comprehensive understanding of the impact of legalization with thoughtful

consideration of legalization for recreational use. Teenagers and parents should also be aware of social media exposure.

Work Cited

1. National Conference of State Legislatures. "State Medical Marijuana Laws." 2018, <http://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx>
2. Warf, Barney. "High Points: An Historical Geography of Cannabis." *Geographical Review*, vol. 104, no. 4, 2014, pp. 414–438.,
[doi:10.1111/j.1931-0846.2014.12038.x](https://doi.org/10.1111/j.1931-0846.2014.12038.x).
3. David Blake & Jack Finlaw, Marijuana Legalization in Colorado: Learned Lessons, 8 *Harv. L. & Pol'y Rev.* 359, 371 (2014)
4. Pedersen W, Skardhamar T (2010) Cannabis and crime: Findings from a longitudinal study. *Addict* 105: 109–118.
5. Doward, Jamie. "Legal Marijuana Cuts Violence Says US Study, as Medical-Use Laws See Crime Fall." *The Guardian*, Guardian News and Media, 14 Jan. 2018, www.theguardian.com/world/2018/jan/14/legal-marijuana-medical-use-crime-rate-plummets-us-study.
6. The Effect of Medical Marijuana Laws on Crime: Evidence from State Panel Data, 1990-2006,
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092816#>
7. Daniulaityte, Raminta, et al. "'Time for dabs': Analyzing Twitter data on marijuana concentrates across the US." *Drug and alcohol dependence* 155 (2015): 307-311.
8. Cavazos-Rehg, Patricia, et al. "Characterizing the followers and tweets of a marijuana-focused Twitter handle." *Journal of medical Internet research* 16.6 (2014).
9. Nguyen, Anh, et al. "Evaluating marijuana-related tweets on Twitter." *Computing and Communication Workshop and Conference (CCWC)*, 2017 IEEE 7th Annual. IEEE, 2017.
10. Thompson, Leah, Frederick P. Rivara, and Jennifer M. Whitehill. "Prevalence of marijuana-related traffic on Twitter, 2012–2013: a content analysis." *Cyberpsychology, Behavior, and Social Networking* 18.6 (2015): 311-319.
11. Tran, Tuan, et al. "Sentiment Analysis of Marijuana Content via Facebook Emoji-Based Reactions." *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018.
12. Cavazos-Rehg, Patricia A., et al. "Twitter chatter about marijuana." *Journal of Adolescent Health* 56.2 (2015): 139-145.
13. "OVERVIEW OF CURRENT MARIJUANA POLICIES." *Analysis of Marijuana Policy*, An, National Academies Press, 1982.
14. Pagolu, Venkata Sasank, et al. "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements." *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)*, 2016,
[doi:10.1109/scopes.2016.7955659](https://doi.org/10.1109/scopes.2016.7955659).
15. "Laurent Luce's Blog." *Twitter Sentiment Analysis Using Python and NLTK* | Laurent Luce's Blog,
www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/.
16. Berman, Mackenzie. "MEDICAL MARIJUANA LAWS AND CRIME: An Empirical Analysis of Market Design and Racial Implications." Princeton University, 2016.

17. United States, Congress, “The National Survey on Drug Use and Health (NSDUH).” The National Survey on Drug Use and Health (NSDUH), Office of Applied Studies, Substance Abuse and Mental Health Services Administration, Dept. Health & Human Services, 2003.
18. “Marijuana Embraced in Michigan, Utah and Missouri, but Rejected in North Dakota”, Christine Hauser, The New York Times, Nov. 7, 2018
<https://www.nytimes.com/2018/11/07/us/politics/michigan-marijuana-legalization.html>
19. “Canada’s Message to Teenagers: Marijuana Is Legal Now. Please Don’t Smoke It”, Catherine Porter, The New York Times, Nov. 11, 2018
<https://www.nytimes.com/2018/11/11/world/canada/marijuana-legalization-teenagers.html>
20. “Recreational marijuana sales will begin in Mass. on Tuesday”, Jessica Barlett, Boston Business Journal, Nov. 16, 2018
<https://www.bizjournals.com/boston/news/2018/11/16/recreational-marijuana-sales-cleared-to-begin-in.html>
21. Minnesota Medical Cannabis Program: Patient Experiences From the First Program Year, Minnesota Department of Health Office of Cannabis, 2017.
<https://www.leg.state.mn.us/docs/2017/other/170739/FirstYearReport.pdf>
22. A Report of the National Commission on Ganja to Rt. Hon. P.J. Patterson, Q.C., M.P. Prime Minister of Jamaica, August 2001.
<https://www.cannabis-med.org/science/Jamaica.htm>
23. Interactive Visualization: <https://public.tableau.com/profile/andrew.xu#!/>
24. “Timeline of cannabis laws in the United States”. Wikipedia.
https://en.wikipedia.org/wiki/Timeline_of_cannabis_laws_in_the_United_States
25. Female Shadow Anatomy image, Mikael Häggström, April 2013.
https://commons.wikimedia.org/wiki/File:Female_shadow_anatomy_without_labels.svg

APPENDIX: Plan of Activities

This project was completed in two months. Midterm and final key performance indicators include the interactive platform and presentation. A task lead has been established for each phase.			
<u>Teammate effort distribution</u>	<u>Task</u>	<u>Duration</u>	<u>Status</u>
Content	<u>Crime Effects:</u> Umashankar Kulasekaran <u>Health Effects:</u> Jennifer S. Johnson <u>Social Influence:</u> Andrew Xu <u>Economic Impact:</u> Jeremy Rosenfeld	9/25/2018 - 11/28/2018	Completed
Project Proposal & Presentation	All team members have contributed similar amount of effort.	9/25/2018 - 10/10/2018	Completed
Final Report, Poster and Presentation	All team members contribute similar amount of effort.	11/12/2018 - 11/28/2018	Completed

Table 4: Plan of Activities