Medha Verma
Mary Zhu
CS448B Final Project

**Marijuana in the United States**

**Website to visualization:** http://web.stanford.edu/~medhav/final_project_448B/

**Introduction**

Marijuana is the most commonly used psychotropic drug in the United States, after alcohol (NIH, 2019). Made from the dried stems and leaves of the *Cannabis sativa* or *Cannabis indica* plant, the drug's allure comes from its mind-altering primary chemical THC (NIH, 2019). It was first outlawed for any use in 1970 under the Controlled Substances Act, but was eventually legalized in some states for only medical purposes in 1990. Since 2010, multiple states have legalized the recreational use of marijuana, stimulating its rising popularity in the U.S. (Shapiro & Mettler, n.d.).

However, for many people aware of its effects, this is a major cause for concern. Studies have shown that marijuana hinders brain development in teenagers, while other marijuana users have reported experiencing breathing problems, increased heart rate, problems with child development in pregnancy, and poorer mental and physical health (NIH, 2019). If marijuana is widely known as a drug with such dangerous side effects, why then, has its usage been so steeply on the rise?

Our data analysis project sets out to explore this question by investigating three potential factors that may influence marijuana usage in the U.S.: the legality of recreational marijuana in different states, perceptions of the dangers of marijuana, and the demographics of the people who chose to use it. First, we obtained a comprehensive dataset from the Substance Abuse and Mental Health Services Administration (SAMHSA) of the U.S. Department of Health & Human Services, and performed preliminary data analyses and experiments with the recorded data variables. We choose legality as one factor to analyze, as our dataset uses 2017 and onward survey data, and the 2016 election doubled the number of states that fully legalized marijuana, starting a nationwide movement. We choose perception as a factor, as the impression of the dangers of marijuana may strongly influence its actual usage. We also choose age as another factor to investigate, in the hopes of determining age-related trends, such as whether marijunana use begins in younger age groups or starts in older ones. Finally, in order to obtain a more granular understanding of marijuana users, we built three of our own datasets (also based on SAMHSA, but augmented with additional data), which looked into their demographics (race, gender, education). We hope to measure how influential, if at all, each of these three factors are on marijuana usage.

**Methods**

Legality/Perception: This visualization aims to study how marijuana use laws in different states may impact the perception of the risk from marijuana. The input data we used counts the number of people in each U.S. state who perceives a great risk from smoking marijuana once a month. We first preprocessed the input data in the dataset by transforming the number of people per state to the percentage of each

state's population perceiving this risk. Like in the previous visualization, this decision was made in order to use percentage as a common metric between states of all population sizes. We designed this visualization as a play on the topic of perception - by creating a mini-quiz inquiring the user to compare his/her guess of percentage of perceivers of great risk from marijuana, the visualization itself challenges the user's perceptions. This You-Draw-It style of interactive visualization is inspired by the New York Times' interactive articles. Originally, our group had considered creating a classic You-Draw-It line graph (a partially drawn line graph that challenges the user to predict the trend of the hidden portion of the line by drawing onto the canvas, before revealing the true trend); however, a line graph implies that the variable on the x-axis of the graph can be displayed sequentially, which in this case it (the legality of states) cannot.

Age/Geography/Legality: This visualization aims to study U.S. marijuana use in 2017, by age group, state, and state legality. We first preprocessed the input data in the dataset by transforming the number of users per state to the percentage of each state's population that are users. This decision was made in order to use percentage as a common metric between states with differing sizes of populations. To create an effective visualization of the potential relationships between the variables involved, we chose to represent the geographic variable through a map of the U.S., the age variable as a filter, the usage variable through color, and state legality through an additional filter. The legality design decision was the most difficult to make - since the other variables had already been defined by the visualization's color and shape, we needed to make sure that highlighting legality would not drown out or impede upon the data illustrated by the other variables. Ultimately, the choice to represent legal states by hiding illegal states allows users to toggle the filter back and forth to observe whether all legal states have high percentages of users, or whether all illegal states have low percentages of users.

Gender/Ethnicity/Education: This final visualization aims to study first time marijuana use, while considering factors like gender, ethnicity, and education. The input data we used for the previous two visualizations were not informative enough to create a meaningful visualization studying demographics. Consequently, we took the entire SAMHSA NSDUH data set and processed it to include a weighted count of first time use at every age for every variable within a demographic. As a team, we hypothesized that gender, race, and education impacted first time use for people. Consequently, we chose to focus our dataset on these three areas. In order to create an effective visualization, we decided to design a multi line chart with age of first time use on the x-axis and the weighted count on the y-axis. The different lines represent attributes of a variable, and the buttons allow users to toggle between studying different variables (gender, education, race). One of the challenges was determining whether percentage or a weighted count was the best way to represent the data. While analyzing the data, it became evident that percentage made it more comprehensible to see similarities in trends for different attributes. However, we concluded that a count might be more insightful as it would allow users to compare the number of users varied from attribute to attribute. For instance, while analyzing our data we saw that the trend for Caucasians and Asians were nearly identical. However, when considering count, we not only got to see the same trend, but also that Asians have significantly fewer first time users over all ages. Because our visualization aimed to study how demographics impacted first time use, we believed that this would be a meaningful addition.

**Results**

## Question (1/3): Legal for Recreational Use

What % of residents in states where marijuana is legal for recreational use perceive a great risk from smoking marijuana once a month?



Figure 1. Visualization comparing legality of marjiuana in U.S. states with perception of risk associated with marjuana use.

Legality/Perception: The major trend covered in Figure 1 is the relationship between the legality of marijuana in different U.S. states (legal for recreational use, legal for only medical use, or illegal) and the perception of risk from smoking marijuana once a month.
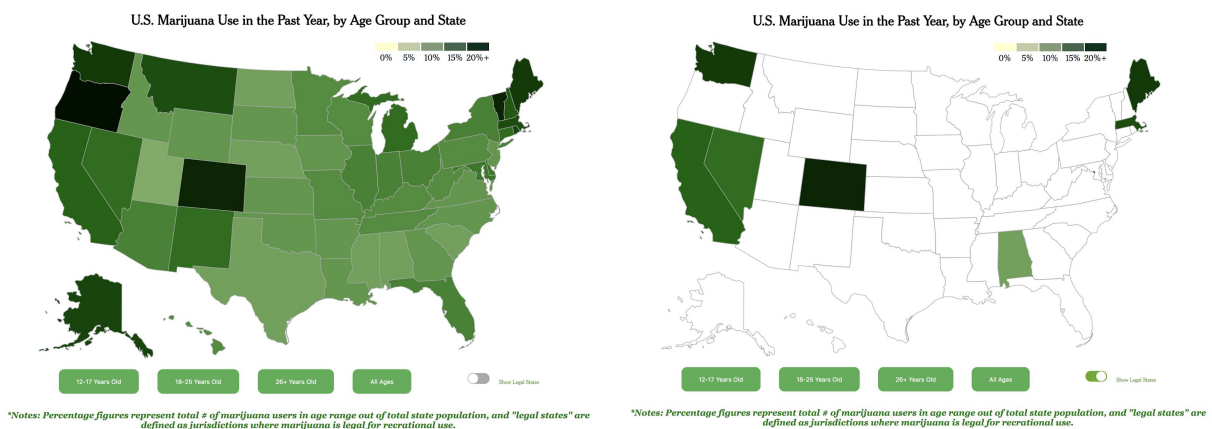


Figure 2. Visualization comparing age, geography, state legality, and marijuana use.

Age/Geography/Legality: The major trends that can be drawn from Figure 2 are the relationship between U.S. states and the percentage of state population that are marijuana users, the relationship between age and marijuana use, and the impact of legality for recreational marijuana use across states.
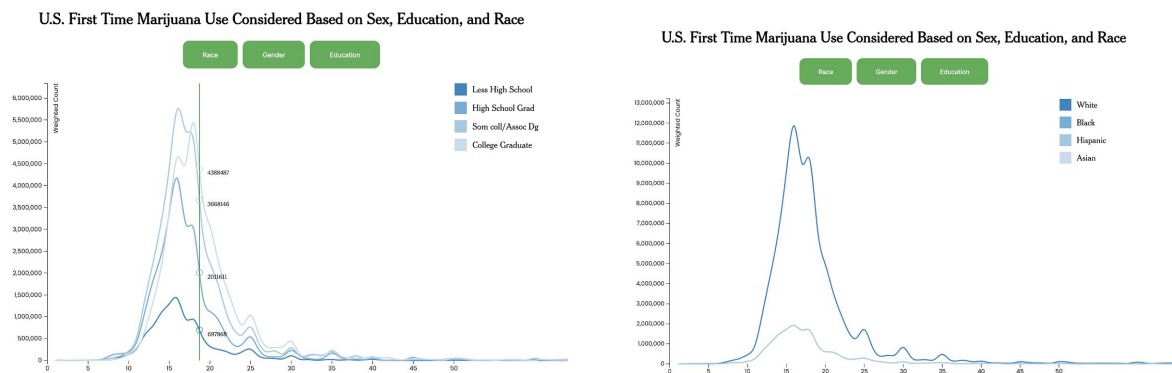


Figure 3. Visualization comparing gender, education, and race.

Race/Gender/Education: Major trends shown in Figure 3 are the relationships between the age of first time mariijuana use and race, gender, and education. It also studies the relationship between different attributes of a specific demographic across all ages.

**Discussion**

Legality/Perception: From this visualization, we see that the more harshly marijuana is restricted legally in a state, the higher the percentage of residents in that state who perceive a great risk from smoking marijuana once a month. Across all states, an average of 21% of residents believe that monthly use of marijuana is a concern of great risk. In states where the recreational use of marijuana is legal, only 18.2% of residents believe that there is a great risk. In states where only medical use is permitted, 20.4% of residents believe that there is great risk. Finally, in states where marijuana is illegal for any use, 22.4% of residents believe that there is great risk. As a result, we see that the intuition that the more stringent the policy against marijuana use, the greater the perception of risk associated with marijuana.

Age/Geography/Legality: In studying marijuana use in the U.S. during 2017, we immediately see a positive trend between usage and age range. We first notice that only 1% of every state's population are marijuana users in the youngest age range (12-17 years old). As we progress with older age ranges, we notice more marijunana users in the 18-25 age range, and the most marijuana users in the oldest age range (26+ years old) in every state. Although this may simply be a byproduct of the oldest age group being the largest and prevents us from drawing conclusions about how old specifically the typical user is, we do see that most users across America are adult age or older.

When looking at marijuana usage and states where marijuana was legal for recreational use in 2017, we find a positive relationship. Among all eight jurisdictions where recreational marijuana is legal (Alabama, California, Colorado, Washington, D.C., Maine, Massachusetts, Nevada, and Washington), the average percentage of marijuana users out of the state population is ~17%, while the average among states where recreational marijuana is illegal is ~12%. However, we do spot a few states where this relationship does not apply. Consider, for example, Alabama - although recreational marijuana is legal here, only 9% of its

total state population used marijuana in the past year. Contrast this with Oregon and Vermont, states where recreational marijuana is illegal, but where total usage is at 23% and 21%, respectively. Despite these outlier states, we do see that, generally, states where marijuana is legal for recreational usage do have a higher percentage of total users.

Demographics/First Time Use: In studying first time marijuana use in the U.S., it's evident that the first time use of marijuana is seen most during teenage years with a sharp steady decline afterwards. When studying gender, overall we see that there doesn't seem to be any significant difference between men and women for first time use. However, we notice that between the ages of 5 and 10, more than twice as many men than women have tried marijuana for the first time. As we move to the older ranges, we see that the peak for both genders is around 15-16 years old, being ~8.7 million for men and ~7.8 million for women. We also see smaller peaks at 18, 25, and 30 years old for both men and women.
When looking at education levels, we noticed a much more significant difference between attributes. Although the attributes less than a high school degree, a high school degree, and some type of associate degree all had similar peaks for first time use, those who graduated college didn't. College graduates had their peak around 18-19 years old versus all other education levels had their peak closer to 15-16 years old. We also see that an associate degree or some college has the highest count of people at every age until it's peak. However, after the associate degree hits its peak at 15-16, college graduates have a higher count of first time users at every age than associate degrees and every other attribute. This is especially interesting because there are nearly 10% more associate degrees than college degrees annually.
Finally, we also can see many learnings from looking at how race plays a role in first time marijuana use. We see that Caucasian people make up the majority of our first time users overall ages. We also see that all races follow a similar curve, with their peaks being around 16 and 18. However, Hispanic and Caucasian people have a steeper curve of first time users in between the ages of 5-10 years old, so they generally experiment with Maijuana at a younger age than other ethnicities.

**Future Work**

With additional time, future work could certainly improve our visual explainer. In regards to our visualization exploring legality and perceptions of danger, perhaps there could be a way to frame all three mini-quiz bar graphs so that they lie on the same graph and can serve as points of references to one another. This particular visualization could also be extended to include comparisons of perception with other factors, such as age, gender, education, and geography (as this data exists in our dataset), although a concise and clear design of these factors is yet to be explored. We also believe providing feedback based on how all users responded could provide an additional insight on perceptions. Regarding our visualization depicting the relationships between age, geography, and legality on marijuana use by state: in order to show a more clear representation of the impact of legality on use, we could figure out how to code up some method to indicate legality without completely hiding the other states (as they serve as points of reference) by perhaps overlaying different textures indicating legality on the colored states. By using this texture method, we could also illustrate varying degrees of legality (such as recreationally legal, medically legal, or illegal), instead of strictly either recreationally legal or not. Finally, with the last visualization it would be interesting to be able to visualize the demographics both through count and

percentage, making certain trends between two different attributes more evident. Furthermore, it would be interesting to explore perceptions at the time of first time use. Do most people who use for the first time have a positive or negative perception of marijuana and does that relationship stay the same with age? We would also have liked to put tooltips at points where there were interesting insights, so the user would have less cognitive load when looking at the visualization.

**Circumstances and Work Breakdown**

Explanation of team reduction and circumstances:
We started out our project as a team of three: Sean Lee, Medha Verma, and Mary Zhu. However, on the afternoon of Monday, March 23, Sean informed us that he could no longer finish the project given external circumstances. At this point, Medha and Mary were under the impression that submitting a partial team project meant that only ⅔ of the requirements were to be submitted (in reference to Piazza question @143). However, on Tuesday, March 24, after they posted a clarification question on Piazza, they discovered that all components of the project were to be submitted regardless of team size. Also, in light of having to move out of Stanford and the emotional distress over the last couple of weeks, we weren't able to commit as much time as we had hoped. The following lists the work breakdown between Medha and Mary:

**Medha Verma:**
Medha designed and implemented the visualization titled "U.S. First Time Marijuana Use Considered Based on Sex, Education, and Race." For her visualization, she started by researching and finding all the datasets for the project and then processing and cleaning the three datasets used for this visualization. She then went on to design and build the visualization, using multiple sources to create it. Medha also wrote her respective parts pertaining to the visualization and contributed to writing all other sections of the report. Finally, Medha compiled, formatted, and styled the interactive article to come together cohesively, and uploaded the code as a running website.

**Mary Zhu:**
Mary designed and implemented her original visualization titled "U.S. Marijuana Use in the Past Year, by Age Group and State" and the additional visualization (to cover for Sean) titled "How does the legality of marijuana impact perceptions of risk in different U.S. states?". For her original visualization, Mary adopted a mosaic approach of referencing, designing, and synthesizing several sources to create it. After realizing that the submission required three visualizations, she asked Sean to send over his code. However, Sean had not added additional functionality to his code since the Zoom project check-in (the beginnings of a You-Draw-It line graph). Further, she noticed that the type of visualization he was attempting to implement was not the most appropriate for the variables he was working with (as described in the Methods section). As a result, she decided to implement a new visualization to better fit the variables. For this visualization, Mary used existing files (youdrawit.js and style.css files), tweaking the frontend to fit the data. Finally, Mary wrote the sections of the report pertaining to both visualizations + some other parts, and minor stylistic editing of the interactive article.

**References** (Additional references to code commented in files)

National Institute on Drug Abuse (2019). What is marijuana? Retrieved from
https://www.drugabuse.gov/publications/drugfacts/marijuana.

Shapiro, L. & Mettler, K. (n.d.). U.S. Marijuana Laws: A History. Retrieved from
https://www.washingtonpost.com/graphics/health/marijuana-laws-timeline/.

Substance Abuse and Mental Health Services Administration (2017). Welcome to Data and
Dissemination. Retrieved from https://www.samhsa.gov/data/.

Substance Abuse and Mental Health Services Administration (2018). Welcome to Data and
Dissemination. Retrieved from https://www.samhsa.gov/data/.