MathWorks Math Modeling Challenge 2019

Wayzata High School–

Team # 12063 Plymouth, Minnesota

Coach: William Skerbitz

Students: Amanda Chan, George Lyu, Zachary Xiong, Caroline Zeng, Alisha Zhu

MathWorks Math Modeling Challenge Finalist

$5,000 Team Prize

MathWorks Math Modeling Challenge

***Note: This cover sheet has been added by SIAM to identify the winning team after judging was completed. Any identifying information other than team # on a MathWorks Math Modeling Challenge submission is a rules violation.

***Note: This paper underwent a light edit by SIAM staff prior to posting.

ONE IS TOO MANY AND A THOUSAND NOT ENOUGH: SUBSTANCE USE AND ABUSE

Team #12063

Table of Contents

Executive Summary

Substance abuse has plagued American society for hundreds of years. Recently, the opioid epidemic has caused a proliferation of prescription drug abuse and fatalities due to overdoses. Similarly, the rise of e-cigarettes has reignited the consumption of nicotine in the United States. These trends foreshadow many detrimental effects to both individuals and the American society as a whole. How will drug use continue to spread in the future, and how can we quantify the consequences of this epidemic?

Our team was asked to predict the spread of vaping in youth over the next 10 years and compare it to the past trend of cigarette smoking. First, we assumed that only data regarding high school students is relevant for increased e-cigarette usage. Additionally, we assumed that the growth of e-cigarettes began in 2003 and that the growth of traditional cigarettes started in 1900. Our model uses a logistic regression to predict the percentage of teens using e-cigarettes, which is justifiable by an upper limit of student drug use. This carrying capacity is found to be 22.8% of high schoolers, and our model also indicates that this limit will be reached in 10 years from this year. Compared to the past trend of cigarette consumption, electronic cigarette usage was growing much more rapidly initially, but will slow down.

We were also asked to create a model to predict the probability that a given individual will use a given substance. We assumed that race, sex, and academic grades are all independent of each other while drug usage is dependent on all three factors. We created a probability matrix for each drug type to relate an individual's demographic characteristics to the probability that an individual will use each drug. We were also tasked with applying this model to a class of 300 high school seniors and predicting how many of them will use nicotine, marijuana, alcohol, and un-prescribed opioids. We created a simulation that randomly constructed 300 individuals and counted the number of individuals who were expected to use each drug. The simulation found that in this class, 69 students would use alcohol, 36 students would use marijuana, 34 students would use nicotine, and 15 students would use un-prescribed opioids.

Finally, our team was tasked to create a model that assigns an index score to controlled substances based on their impact on society. Then, we were asked to use this model to rank nicotine, marijuana, alcohol, and un-prescribed opioids. Our model assumed that financial factors and non-financial factors are equal in weight. We normalized the monetary costs to society and aggression scores. Then, we calculated an average of their values and determined that alcohol is the most damaging to society, while marijuana is the least damaging of the given substances.

Drugs have the potential to adversely affect the livelihoods, careers, academics, and finances of both drug users and those close to them. This demands a thorough examination and continued monitoring of the spread of drugs and the impact of drugs on society.

Background

In 2014, the National Drug Survey on Abuse and Health found that one in every 10 Americans aged 12 and older deals with substance abuse disorder [1]. Drug abuse has been a prominent issue in society, and both financial and non-financial costs have taken their toll on the population. Every year, 78.5 billion dollars are lost to prescription opioids alone, which factors in the costs of healthcare, lost productivity, addiction treatment, and criminal justice involvement [2]. This value is only a minor part of the costs of all drug abuse, which is estimated to reach 820 billion dollars when combined with tobacco, alcohol, and other illicit drugs [2].

Now, focus is brought on the population of high schoolers who are at risk of substance abuse with alcohol, marijuana, prescription drugs, tobacco, and, most notably, e-cigarettes. Since 1983, implementations of drug education programs such as Drug Abuse Resistance Education (DARE) have attempted to limit the spread of illicit underage drug use [3]. However, recent surveys have found an exponential increase in the use of e-cigarettes among high schoolers, a surprising trend given the apparent success of drug education in slowly decreasing all other underage drug abuse [3].

These increases in substance use necessitate further investigation of the spread of drug consumption and the effects of these drugs on individuals and society.

Global Assumptions

G.1 The federal government does not implement any new policies to regulate the drugs analyzed in these models.
G.2 Drug users will continue their current usage trends, and drug dealers will continue their current marketing trends.

Part I: Darth Vapor

1.1 Restatement of Problem

We are asked to create a model that can predict the spread of vaping over the next 10 years and compare it to the past spread of smoking cigarettes. We quantify the spread of a drug as the percentage of the United States population that uses that drug.

1.2 Local Assumptions

1. Only data regarding high school students is relevant for e-cigarettes.
    a. **Justification:** Our model is looking at drug usage for the new generation, which specifies high school students. E-cigarettes are also the most prominent among high school students [4]. Thus, other age groups are negligible.
2. The growth in the number of people smoking cigarettes in the United States is comparable to the consumption of cigarettes over time.

   a. **Justification:** The number of people who smoke cigarettes is proportional to the number of cigarettes consumed.

3. The growth in the number of people who use e-cigarettes is comparable to the percentage of high schoolers who vape.
   a. **Justification:** E-cigarettes are targeted at the youth, specifically high school students [5].
4. E-cigarette usage can be accurately modeled with the percentage of high school students who used e-cigarettes in the past 30 days.
   a. **Justification:** If a student has used an e-cigarette in the last 30 days, he or she is likely to be a regular user.
5. The growth of usage of e-cigarettes over time began in 2003.
   a. **Justification:** The first commercially successful e-cigarette was patented in 2003 [6].
6. The growth of usage of cigarettes over time began in 1900.
   a. **Justification:** Cigarettes became popular at the beginning of the 20th century [7].

1.3 Variables

Symbol	Definition	Units
K	Carrying capacity	Percent of high schoolers who have used e-cigarettes in the past 30 days
E(t)	Percent of high schoolers who have used e-cigarettes in the past 30 days	Percent
t	Time since 2003	Years
D	Drug type	Unitless
Y	Years after drug's first significant usage	Years
U(D,Y)	Usage of drug D in year Y since first significant usage	Defined below in 1.4.2
GP	Relative growth proportion of drug's usage	Unitless
GPR	Relative growth proportion ratio of electronic cigarettes to combustible cigarettes	Unitless

1.4 Solution & Results

1.4.1 Predicted Electronic Cigarettes

Teen e-cigarette usage has accelerated over the past decade. Below is a plot showing high school e-cigarette usage since 2011 [8].

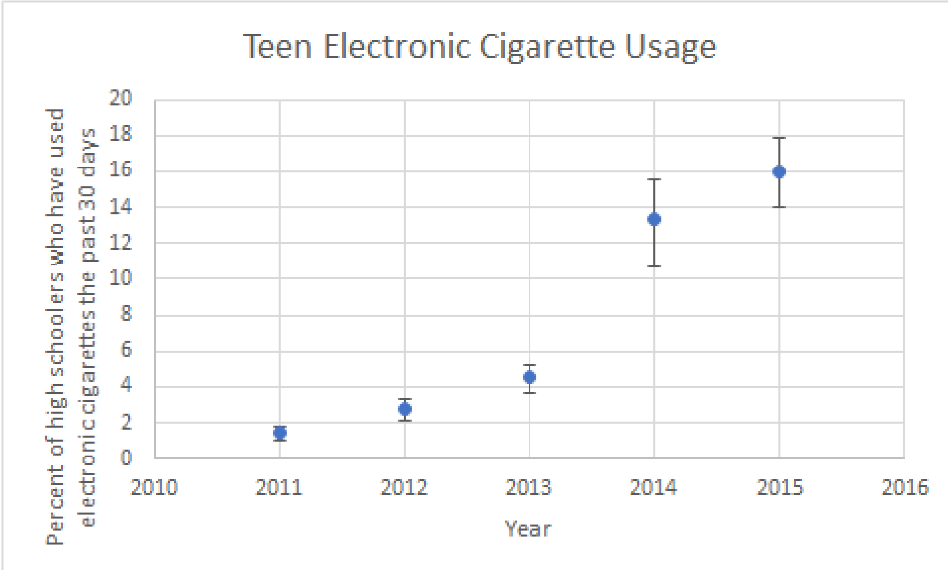

Figure 1: E-cigarette use among high school students

Table 1: Percentage of high schoolers who use e-cigarettes

Year	**Percent of high schoolers who have used e-cigarettes the past 30 days**
2011	1.5
2012	2.8
2013	4.5
2014	13.4
2015	16

We used a logistic regression to model e-cigarette usage for high schoolers because there is a lower bound (at least 0%) and upper bound (at most 100%) to the percentage of high schoolers who use e-cigarettes.

Logistic equations are in the form $E(t) = \frac{K * E_0 e^{rt}}{K + E_0(e^{rt} - 1)}$, where *E(t)* is the electronic cigarette usage at *t* years since 2003, and K is the maximum electronic cigarette usage (usage is defined as the percent of high schoolers who have used electronic cigarettes in the last 30 days). We let $E_0 = e^{\alpha}$ because E_0, which is a population value, has the same possible values as e^{α}. Through substitution, the equation may be rewritten as $E(t) = \frac{K * e^{rt+\alpha}}{K + e^{rt+\alpha} - e^{\alpha}}$.

We assume that $e^{\alpha} \ll K + e^{rt+\alpha}$ because the determined α value, shown later in the final logistic equation to be -7.17, is a negative number of high magnitude, so we may ignore e^{α} in the denominator. Therefore, we have the refined logistic model:

$$E(t) = \frac{K * e^{rt+\alpha}}{K + e^{rt+\alpha}}$$

Through algebraic manipulation, this equation may be rewritten as

$\ln\left(\frac{K*E(t)}{K-E(t)}\right) = rt + \alpha.$

There are three parameters in the equation that we must determine: *K*, *r*, and α. We do this by varying the carrying capacity *K* (starting at 100% and decreasing by 0.1% increments). For each *K* value, we perform a linear regression of $\ln\left(\frac{KE(t)}{K-E(t)}\right)$ over time *t* in order to determine the respective *r* and α values. Then, using the *K*, *r*, and α values, we determine our equation's predicted e-cigarette usage for each *K* value. We then find the percent error of these predicted e-cigarette usage values by comparing to the true e-cigarette usage from the data in Table 2. For clarity, let $y = \ln\left(\frac{K*E(t)}{K-E(t)}\right) = rt + \alpha$.

Below is a sample of carrying capacity values:

Table 2: Percentage of high schoolers who have used e-cigarettes in the past 30 days

Years since 2003	**True value**	**y, K**						
		23.1%	**23.0%**	**22.9%**	**22.8%**	**22.7%**	**22.6%**	**22.5%**
8	1.5	1.2763	1.2744	1.2724	1.2704	1.2684	1.2663	1.2642
9	2.8	2.9725	2.9737	2.9749	2.9762	2.9775	2.9788	2.9801
10	4.5	6.2748	6.2837	6.2927	6.3019	6.3113	6.3209	6.3306
11	13.4	11.2037	11.2147	11.2259	11.2373	11.2489	11.2607	11.2728
12	16	16.2622	16.2530	16.2438	16.2344	16.2251	16.2156	16.2061
Deter-mined constants	r	0.9264	0.9288	0.9313	0.9338	0.9364	0.9391	0.9418
	α	-7.1101	-7.1310	-7.1524	-7.1741	-7.1963	-7.2188	-7.2419

Using this method, we determine *K* = 22.8%, which corresponds with *r* = 0.934 and α = -7.17, to be the *K* value that yields the lowest average percent error.

Table 3: Percent error of our model

Years since 2003	**True value**	**y, K**						
		23.1%	**23.0%**	**22.9%**	**22.8%**	**22.7%**	**22.6%**	**22.5%**
8	1.5	0.824	0.547	1.326	0.042	29.209	0.957	0.322
9	2.8	12.733	0.767	2.881	0.033	89.490	0.967	2.083
10	4.5	782.744	0.992	5.344	0.179	34.193	0.815	6.766
11	13.4	130.669	0.914	11.280	0.004	2983.178	0.996	10.316
12	16	252.587	0.936	16.361	0.008	2098.822	0.992	15.332
Average		235.911	0.831	7.4384	0.053	1046.979	0.9454	6.964

Therefore, our model yields the following equation:

$$E(t) = \frac{22.8e^{.934t-7.17}}{22.8 + e^{.934t-7.17}}$$

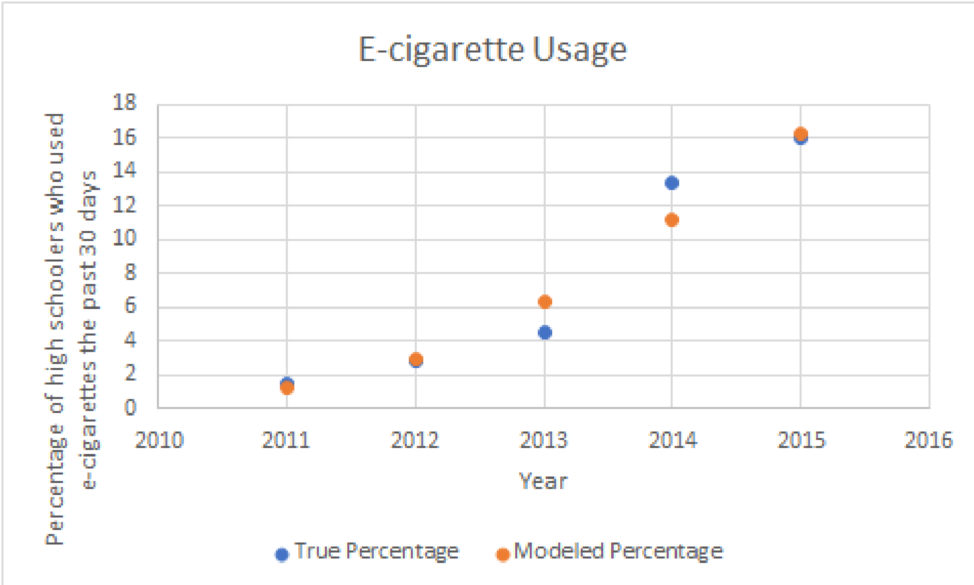

Figure 2: Comparison between our modeled growth of e-cigarette use and the true values

Using this equation, we substituted t = 26 years since 2003 to find the usage in 10 years from 2019. We obtained $E(26) = 22.8$, so our model predicts that 22.8% of high schoolers will regularly use e-cigarettes in 10 years from this year.

1.4.2 Growth of Electronic Cigarettes Compared to Growth of Combustible Cigarettes

We compared the electronic cigarettes to combustible cigarettes by comparing their relative rates of growth using data from the American Lung Association [9]. We assumed that the number of combustible cigarettes consumed is proportional to the number of combustible cigarette users. Thus, we may compare electronic cigarette user quantity with combustible cigarette product consumption.

We defined *U(D, Y)* as the usage of drug D in Y years after that drug had its first significant usage. We can use *U(D, Y)* to represent the percent of high schoolers who used electronic cigarettes in the past 30 days (units of %) and the combustible cigarette consumption (units of cigarettes) because the final result is a unitless ratio of relative growth in usage of electronic cigarettes (unitless) to the relative growth in usage of combustible cigarettes (unitless). We defined $\Delta U(D,Y) = U(D,Y) - U(D,Y-1)$ as the amount the usage of drug D grew between years Y-1 and Y. We defined $GP(D,Y) = \frac{\Delta U(D,Y)}{U(D,Y-1)}$

as the proportional growth of drug D between year Y-1 and Y. Finally, we defined $GPR(Y) = \frac{GP(Ecig,Y)}{GP(Cig,Y)}$ as the ratio of the proportional growth of electronic cigarette usage to the proportional growth of combustible cigarette usage.

A high GPR implies that electronic cigarette usage grew much more rapidly than combustible cigarette usage at Y years after the first significant usage of the drugs, and the converse is true as well.

Table 4: Comparing growth rates of e-cigarettes and cigarettes over time

Years since first significant usage	**E-cigarette**			**Cigarette**			**E-cig to cig growth proportion ratio (unitless)**
	Usage (%)	**Usage growth (%)**	**Growth proportion (unitless)**	**Usage (cigs)**	**Usage growth (cigs)**	**Growth Proportion (unitless)**	
Y	*U(E-cig, Y)*	*ΔU(E-cig,Y)*	*GP(E-cig,Y)*	*U(Cig, Y)*	*ΔU(Cig,Y)*	*GP(Cig,Y)*	*GPR(Y)*
0	0.001			2.5			
1	0.002	0.001	1.545	2.500	0.000	0.000	
2	0.005	0.003	1.544	2.800	0.300	0.120	12.869
3	0.013	0.008	1.544	3.100	0.300	0.107	14.409
4	0.032	0.020	1.542	3.300	0.200	0.065	23.909
5	0.082	0.050	1.539	3.600	0.300	0.091	16.930
6	0.207	0.125	1.531	4.500	0.900	0.250	6.123
7	0.519	0.312	1.509	5.300	0.800	0.178	8.491
8	1.277	0.757	1.458	5.700	0.400	0.075	19.321
9	2.990	1.714	1.342	7.000	1.300	0.228	5.884
10	6.328	3.337	1.116	8.600	1.600	0.229	4.882
11	11.270	4.943	0.781	10.100	1.500	0.174	4.478
12	16.262	4.992	0.443	13.200	3.100	0.307	1.443
13	19.689	3.427	0.211	15.800	2.600	0.197	1.070
14	21.467	1.778	0.090	16.500	0.700	0.044	2.038
15	22.257	0.790	0.037	17.900	1.400	0.085	0.434
16	22.583	0.327	0.015	25.200	7.300	0.408	0.036
17	22.714	0.131	0.006	35.700	10.500	0.417	0.014
18	22.766	0.052	0.002	45.600	9.900	0.277	0.008
19	22.787	0.020	0.001	48.000	2.400	0.053	0.017
20	22.795	0.008	0.000	44.600	-3.400	-0.071	-0.005
21	22.798	0.003	0.000	50.700	6.100	0.137	0.001
22	22.799	0.001	0.000	53.400	2.700	0.053	0.001
23	22.800	0.000	0.000	64.400	11.000	0.206	0.000

24	22.800	0.000	0.000	71.000	6.600	0.102	0.000
25	22.800	0.000	0.000	79.800	8.800	0.124	0.000
26	22.800	0.000	0.000	89.100	9.300	0.117	0.000

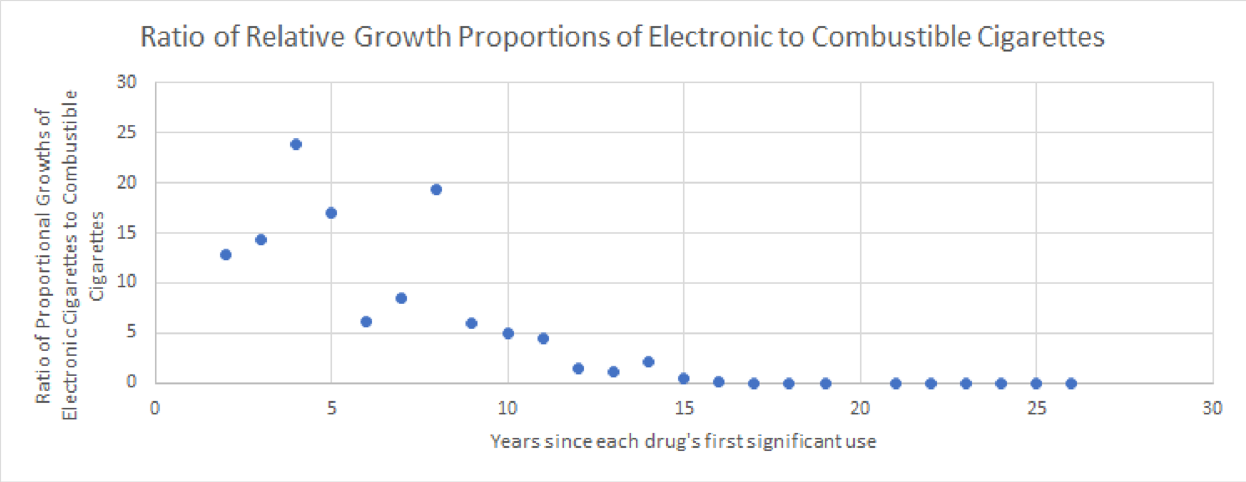

Figure 3: Ratio of growth of e-cigs to cigs since each drug's first significant use

The GPR is extremely high at years before year 9. After that, GPR approaches 0. This implies that electronic cigarette usage grew much more rapidly than combustible cigarettes in each drug's initial years, but electronic cigarette usage growth has slowed significantly relative to how quickly combustible cigarettes usage grew. This suggests that electronic cigarettes appear to be prevalent now, but the usage of electronic cigarettes has mostly saturated and will not grow much more.

This is reinforced by each drug's carrying capacities. The carrying capacity of electronic cigarettes, as determined by our model, is 22.8% percent of high schoolers, whereas the carrying capacity of combustible cigarettes is far greater: Combustible cigarette usage reached 44.5% of the population in 1938 [10].

This comparison strongly implies that electronic cigarettes are an exaggerated fad. The extremely rapid initial growth of electronic cigarettes (relative to empirical cigarette growth) gave the impression that electronic cigarettes would invade every high schooler's life, but the low carrying capacity of electronic cigarette usage and the slowing growth rate of electronic cigarette usage disprove fears of electronic cigarettes reaching the prevalence of cigarettes.

1.5 Validation

To validate our model, we found the most recent data in the National Youth Tobacco Survey reported 20.8% percent of high schoolers using e-cigarettes [11]. Our model, based on data from 2011 to 2015, found that 22.3% of high schoolers would be users. This yields a percent error of 7.2% for our prediction, which is relatively low considering the limited data and susceptibility to change for e-cigarette use in high school. The National Youth Tobacco Survey thus validates our model.

1.6 Strengths & Weaknesses

Strengths

- Our model maintains simplicity. We only created a logistic regression to extrapolate and determine the spread of vaping in youth in the next 10 years.
- We used a logistic regression model, which more accurately represents the spread of vaping because there is a cap at 100% for the percentage of the population that can vape.

Weaknesses

- We compared growth rates between the percentage of high schoolers who used e-cigarettes and the total consumption of cigarettes, which may be inconsistent.
- We did not use data that is representative of the entire United States population in our analysis. This may have skewed our data and results.

Part II: Above or Under the Influence?

2.1 Restatement of Problem

We are asked to create a mathematical model that will determine the probability that an individual with certain demographic characteristics has used each of four drugs: nicotine, marijuana, alcohol, and un-prescribed opioids. We are also asked to apply this model to determine the number of students in a 300-student class who have used each drug.

2.2 Local Assumptions

1. Race, sex, and academic grades are all independent from each other, even though drug use is dependent on all three.
    a. **Justification:** Race and sex are clearly independent from each other, as sex is randomly determined by genetics and does not depend on race. Grades sufficiently are independent from sex and race and because schools typically endeavor to provide equal educational opportunity to all groups of people. Any dependence between these factors is considered insignificant compared to the dependence of drugs on these factors.
2. We can compare data regarding 11th graders with data regarding 12th graders.
    a. **Justification:** 11th graders and 12th graders behave similarly because they are very similar in age.
3. Students who have previously used certain drugs can be considered as individuals who would use that specific drug again.

a. **Justification:** First time use of a drug is much harder to accomplish than returning use, increasing the likelihood that one with past use would use again.

2.3 Solution & Results

Our model considered three factors – race, sex, and academic grades – that would affect the likelihood that a given individual will use a given substance. We obtained race, sex, and academic grade data from the Centers for Disease Control [12]. We now derive the function used for our model:

Consider *n* demographic characteristics (such as having an A GPA or a male sex) $X_1, X_2, \ . \ . \ . \ , X_n$ that affect the likelihood an individual will use a given drug *D.* We assume the *n* demographic factors are independent of each other even though the likelihood of using drug *D* is dependent on each factor. This assumption justifies

$$\begin{aligned} P(D \cap X_1 \cap X_2 \cap \ . \ . \ . \cap X_n) &= P(D \cap X_1 \cap D \cap X_2 \cap \ . \ . \ . \cap D \cap X_n) \\ &= P((D \cap X_1) \cap (D \cap X_2) \cap \ . \ . \ . \cap (D \cap X_n)) \\ &= P(D \cap X_1) \cap P(D \cap X_2) \cap \ . \ . \ . \cap P(D \cap X_n) \\ &= P(D \cap X_1) * P(D \cap X_2) * \ . \ . \ . * P(D \cap X_n) \end{aligned}$$

Using the conditional probability rule,

$$\begin{aligned} P\big(D|(X_1 \cap X_2 \cap \ . \ . \ . \cap X_n)\big) &= \frac{P(D \cap X_1 \cap X_2 \cap \ . \ . \ . \cap X_n)}{P(X_1 \cap X_2 \cap \ . \ . \ . \cap X_n)} \\ &= \frac{P(D \cap X_1) * P(D \cap X_2) * \ . \ . \ . * P(D \cap X_n)}{P(X_1) * P(X_2) * \ . \ . \ . * P(X_n)} \\ &= \frac{P(D \cap X_1)}{P(X_1)} * \frac{P(D \cap X_2)}{P(X_2)} * \ . \ . \ . * \frac{P(D \cap X_n)}{P(X_n)} \\ &= P(D|X_1) * P(D|X_2) * \ . \ . \ . * P(D|X_n) \\ &= \prod_{i=1}^{n} P(D|X_i) \end{aligned}$$

For a drug *D*, we define the following probability matrix:

$$Q_D = \begin{bmatrix} \text{P(D|Native American)} & \text{P(D|Asian)} & \text{P(D|Black)} & \text{P(D|Hispanic)} & \text{P(D|White)} & \text{P(D|Multiple race)} \\ \text{P(D|Female)} & \text{P(D|Male)} & 0 & 0 & 0 & 0 \\ \text{P(D|A)} & \text{P(D|B)} & \text{P(D|C)} & \text{P(D|D/F)} & 0 & 0 \end{bmatrix} = \begin{bmatrix} q(D)_{1,1} & \cdots & q(D)_{1,n} \\ \vdots & \ddots & \vdots \\ q(D)_{n,1} & \cdots & q(D)_{n,n} \end{bmatrix}$$

Inputting our data into this probability matrix yields a probability matrix for each drug:

$$Q_{Cigarettes} = \begin{bmatrix} 0.4186 & 0.1483 & 0.2109 & 0.297 & 0.3095 & 0.315 \\ 0.2728 & 0.3067 & 0 & 0 & 0 & 0 \\ 0.31 & 0.46 & 0.6 & 0.74 & 0 & 0 \end{bmatrix}$$

$$Q_{Alcohol} = \begin{bmatrix} 0.5692 & 0.4239 & 0.5132 & 0.6467 & 0.6171 & 0.648 \\ 0.6258 & 0.5809 & 0 & 0 & 0 & 0 \\ 0.24 & 0.35 & 0.4 & 0.52 & 0 & 0 \end{bmatrix}$$

$$Q_{Marijuana} = \begin{bmatrix} 0.3741 & 0.1303 & 0.4282 & 0.4239 & 0.3195 & 0.4044 \\ 0.3589 & 0.3523 & 0 & 0 & 0 & 0 \\ 0.24 & 0.39 & 0.53 & 0.66 & 0 & 0 \end{bmatrix}$$

$$Q_{Opioid} = \begin{bmatrix} 0.268 & 0.0796 & 0.1228 & 0.1505 & 0.1352 & 0.1854 \\ 0.1442 & 0.1335 & 0 & 0 & 0 & 0 \\ 0.11 & 0.17 & 0.22 & 0.34 & 0 & 0 \end{bmatrix}$$

We also define the trait matrix of a single high schooler *k*, *where all non-1 elements are equal to 0*.

$$T_K = \begin{bmatrix} =1 \text{ if Native American} & =1 \text{ if Asian} & =1 \text{ if Black} & =1 \text{ if Hispanic} & =1 \text{ if White} & =1 \text{ if Multiple race} \\ =1 \text{ if Female} & =1 \text{ if Male} & 0 & 0 & 0 & 0 \\ =1 \text{ if GPA is A} & =1 \text{ if GPA is B} & =1 \text{ if GPA is C} & =1 \text{ if GPA is D/F} & 0 & 0 \end{bmatrix} = \begin{bmatrix} t(k)_{1,1} & \cdots & t(k)_{1,n} \\ \vdots & \ddots & \vdots \\ t(k)_{n,1} & \cdots & t(k)_{n,n} \end{bmatrix}$$

For example, a person who is Asian and female and earns B grades corresponds with the following trait matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, an individual *k*'s trait matrix represents their demographics.

Next, we determined the product of the elements of the Hadamard Product of matrices Q_d and T_k:

$$\text{prod}(Q_d \text{ o } T_k) = \prod_{j,l=1} q(D)_{j,l} * t(k)_{j,l}, \text{omitting } t(k)_{j,l} \text{ that are equal to zero}$$

$$= \prod_{i=1}^{n} P(D|X_i)$$

Thus, $\text{prod}(Q_d \text{ o } T_k)$ is the probability that individual *k* has used drug *D*. We will call this function the *drug use probability function*.

To find the number of students in a 300-student class who have used a drug, we conducted 10 trials of a simulation. For each trial, we constructed 300 students whose characteristics were randomly determined using the distribution of American high school student characteristics shown in the matrix below [13, 14, 15]. The elements of the matrix represent the proportion of high school students with the corresponding characteristic.

$$\begin{bmatrix} P(\text{Native American}) & P(\text{Asian}) & P(\text{Black}) & P(\text{Hispanic}) & P(\text{White}) & P(\text{Multiple race}) \\ P(\text{Female}) & P(\text{Male}) & 0 & 0 & 0 & 0 \\ P(A) & P(B) & P(C) & P(D/F) & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.01 & 0.058 & 0.149 & 0.264 & 0.494 & 0.025 \\ 0.508 & 0.492 & 0 & 0 & 0 & 0 \\ 0.06 & 0.31 & 0.35 & 0.28 & 0 & 0 \end{bmatrix}$$

For each of the 300 students, we use the drug use probability function to determine the likelihood that the student has used each drug. For each drug, we determine the expected count of students who have used that drug by summing each student's probability of having used that drug. This simulation yielded the following results:

Table 5: Simulation results of expected number of high schoolers who use each drug

Trial	**Expected number of high schoolers from a 300-student class who have used a drug**			
	Cigarettes	**Alcohol**	**Marijuana**	**Opioid**
1	37.64	67.43	40.00	14.28
2	35.51	71.25	38.08	10.18
3	37.34	69.35	36.13	11.73
4	34.25	69.02	41.98	14.63
5	34.14	70.95	37.69	14.95
6	38.91	63.80	45.65	13.14
7	41.05	66.41	35.04	9.05
8	33.42	62.18	38.18	11.45
9	33.89	63.53	35.38	13.90
10	34.42	69.43	36.13	15.04
Average	**36.06**	**67.34**	**38.43**	**12.84**
Standard deviation	**2.42**	**3.07**	**3.15**	**2.01**
Coefficient of variation	**0.067**	**0.046**	**0.082**	**0.157**

Therefore, the average expected number of students from a 300-student class who have used cigarettes is 36, who have used alcohol is 67, who have used marijuana is 38, and who have used opioids is 13. The low coefficients of variation validate the consistency of our simulation.

2.4 Validation

According to a statewide survey conducted by the Minnesota Department of Education, 24.5% of 11th grade students had an alcoholic beverage within the past 30 days in 2016 [16]. The following table compares the results of the Minnesota Education Survey with the results of our model. We then calculate the percent difference of our model.

Table 6: True percentages of students who use drugs and percent difference of simulation

Drug	**Actual Percent of Students**	**Predicted Percent of Students**	**Percent Difference**
Nicotine	13.0	12.0	8.0
Alcohol	24.5	22.4	9.0
Marijuana	16.0	12.8	22.2
Un-prescribed Opioids	5.0	4.3	15.1

As shown, the percent differences are in an acceptable range, which shows that our results are relatively consistent with those of the Minnesota Department of Education.

2.5 Strengths & Weaknesses

Strengths

- Our model is easily applicable to all students across the nation because the only input into the model is the demographic data of that student.

Weaknesses

- There are many important factors that we did not incorporate into our model: Specifically, we did not incorporate income, the price of each drug, the addictiveness of each drug, and the prevalence of each drug.

Part III: Ripples

3.1 Restatement of Problem

We are asked to create a mathematical model that incorporates financial and non-financial factors in order to give a rating score to an illicit substance based on its societal impact. Then, we are asked to use this model to rank the severity of nicotine, marijuana, alcohol, and un-prescribed opioids.

3.2 Local Assumptions

1. Financial factors are equal in importance to non-financial factors and should therefore be weighted equally.
    a. **Justification:** Drugs have severe consequences that have both financial and non-financial impacts. This assumption is necessary for the simplicity of the model.
2. Drug-related aggression incorporates all non-financial factors, such as divorce and domestic abuse.
    a. **Justification:** Aggression is a primary psychological consequence of drug-related body chemistry. Additionally, aggression is a cause of many non-financial factors, as aggression is one cause of domestic violence.

3.3 Variables

Symbol	Definition	Units
I	Societal impact index	Unitless
M	Monetary cost	US dollars
A	Aggression scores	Unitless

3.4 Solution & Results

We found the monetary cost to society of each drug from the National Institute on Drug Abuse [17], and we found how aggression relates to each drug from the National Academy of Sciences [18]. The monetary cost to society comprises costs from healthcare, the criminal justice system, and workplace productivity loss resulting from each drug. The standardized monetary cost M of each drug is found by dividing that drug's monetary costs by the total monetary costs of all four drugs. Similarly, the standardized aggression score A of each drug is equal to the specific drug's aggression score by the total aggression score of all four drugs. To determine the final index I for the given drugs, we used the following equation:

$$I = \frac{M + A}{2}$$

Table 7: Monetary cost, aggression scores, and societal impact indexes for four drugs

Drug	Monetary Cost to Society	M	Aggression Scores	A	I
Alcohol	$249 billion	0.387	4	0.4	0.394
Marijuana	$16.4 billion	0.025	1	0.1	0.063
Nicotine	$300 billion	0.466	3	0.3	0.383
Un-prescribed Opioids	$78.5 billion	0.122	2	0.2	0.161

According to Table 7, alcohol with an I value of 0.394 is the most damaging to society, closely followed by nicotine with an I value of 0.383. Un-prescribed opioids are the third most damaging to society with an I value of 0.161. Marijuana is the least damaging to society with an I value of 0.063.

3.5 Validation

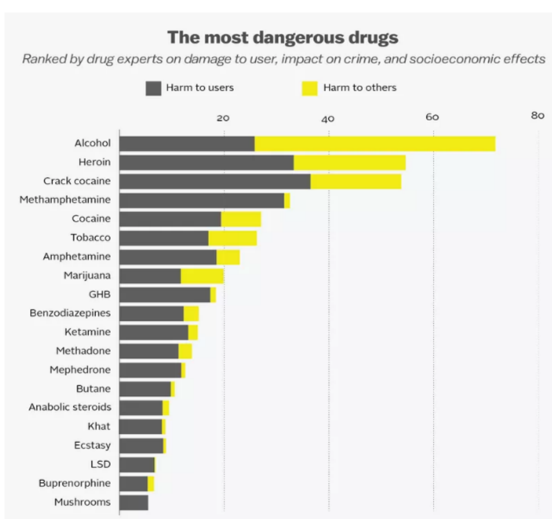

Figure 4: Danger ranking of various drugs

According to Figure 4 from Vox Media [19], alcohol has the most negative impact on society, followed by un-prescribed opioids (approximated by heroin), nicotine, and marijuana. This ranking validates our ranking of the given substances because they are very similar: Our model and Vox Media agree that alcohol is the most damaging and that marijuana is the least damaging.

3.6 Strengths & Weaknesses

Strengths

- Our model maintains simplicity. The model incorporates only monetary cost and aggression to determine the societal impact index.

Weaknesses

- The societal impact index only conveys information about whether a certain drug is more or less impactful than another drug. The societal impact index does not describe *how much more* impactful one drug is than another.

Part IV: Conclusion

First, we were asked to predict the growing trend of electronic cigarettes among high school students. We found that e-cigarettes, despite having exponential growth in usage in recent years, were fast approaching a carrying capacity limit, which was found by utilizing a logistic equation. Then, we compared the rise of e-cigarettes to the first rise in combustible cigarettes. Although it appeared e-cigarettes would follow the titanic growth in popularity that traditional cigarettes first garnered, e-cigarettes are actually soon coming to a limited percentage of student users, about 22.8%.

Next, we were tasked with developing a model that could give the probability that an individual has used alcohol, nicotine, marijuana, or un-prescribed opioids. We developed a trait matrix to quantify each individual's demographic characteristics, and we developed a probability matrix to relate an individual's demographic characteristics to the chance that the individual has used each drug. We created a simulation that applied this model to a class of 300 students to determine the expected number of students who have used each drug.

Finally, we focused on both the financial and non-financial costs of alcohol, marijuana, nicotine, and opioids. We assigned a societal impact index score to each of the four drugs based on relative economic costs to society (crime, healthcare, productivity loss) and likelihood of causing aggressive nature, a central element to drug-caused domestic abuse and divorce. The produced ranking of drugs found alcohol to be the most detrimental to people around the user, while marijuana was least likely to negatively impact society.

As our nation approaches an ever-growing number of at-risk teens, monitoring the spread and severity of drugs becomes of utmost importance in protecting and informing the youth. Drugs have devastating impacts on society and our livelihoods, and while there is no way to prevent our posterity from falling to the dark side, we must be vigilant in our pursuit of a cleaner, healthier society.

Bibliography

[1] https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf
[2] https://www.drugabuse.gov/related-topics/trends-statistics#supplemental-references-for-economic-costs
[3] https://www.scientificamerican.com/article/the-new-d-a-r-e-program-this-one-works/
[4] https://truthinitiative.org/news/new-study-reveals-teens-16-times-more-likely-use-juul-older-age-groups
[5] https://truthinitiative.org/news/4-marketing-tactics-e-cigarette-companies-use-target-youth
[6] http://www.casaa.org/historical-timeline-of-electronic-cigarettes/
[7] https://www.atlasobscura.com/articles/in-the-1900s-being-a-cigarette-fiend-was-a-legitimate-defense-for-murder
[8] https://www.cdc.gov/tobacco/data_statistics/sgr/e-cigarettes/
[9] https://www.lung.org/assets/documents/research/tobacco-trend-report.pdf
[10] https://news.gallup.com/poll/1717/tobacco-smoking.aspx
[11] https://www.cdc.gov/mmwr/volumes/67/wr/mm6745a5.htm?s_cid=mm6745a5_w
[12] https://nccd.cdc.gov/youthonline/App/Results.aspx?LID=XX
[13] https://www.cdc.gov/healthyyouth/health_and_academics/pdf/DASHFactSheetDrugUse.pdf
[14] https://www.cdc.gov/healthyschools/health_and_academics/pdf/factsheetAlcohol.pdf
[15] https://www.surgeongeneral.gov/library/reports/50-years-of-progress/sgr50-chap-13.pdf
[16] https://education.mn.gov/mdeprod/groups/communications/documents/basic/bwrl/mdu5/~edisp/mde059325.pdf
[17] https://www.drugabuse.gov/related-topics/trends-statistics#supplemental-references-for-economic-costs
[18] https://www.nap.edu/read/4421/chapter/5#396
[19] https://www.vox.com/2015/2/24/8094759/alcohol-marijuana