# Usage Patterns, Effects, and Subjective Experiences of Illicit Drugs: A Text Mining Approach

# INTRODUCTION

Drug abuse is prevalent in societies around the globe, yet the illegal status of many drugs drives the use and discussion of these illegal drugs underground. Nevertheless, discussion about illegal drug use is extremely public: from the idea of "gateway drugs"," to the historical prohibition of alcohol, to recent discussions about the legalization of marijuana, to cutting-edge mainstream medical research investigating the therapeutic properties of now-illegal drugs. Unfortunately, the illegality of illicit drugs makes them especially difficult to study, and a major gap exists between the large amount of drug abuse and the relatively few studies into the effects and properties of illicit drugs.

Despite the lack of formal research into illicit drug use, a significant body of informal discussion of illicit drugs is available publicly on the Internet. The basis of this report is to examine subjective drug experiences from one of these many sources.

Erowid.org is a long-running website that stores a tremendous amount of information about both legal and illegal drugs. Their tagline is "Documenting the Complex Relationship Between Humans & Pyschoactives," and part of the "documenting" is over 30,000 first-person subjective reports of drug use (called "Experiences") written by users of the site. These reports are from people all over the globe; cover a wide variety of drug combinations, dosages, and methods of use. The earliest reports date from the year 2000. The reports are from people of both genders, a wide range of ages, and many different countries. The reports themselves are written in multiple languages and cover a wide range of styles from short, general write-ups about the user's history with a drug, to extremely long and detailed reports with minute-by-minute updates of drug effects. Short selections from some of the reports can be found in Appendix A.

# ACQUIRING THE DATA

From 30 March 2014 to 1 April 2014, Python and the BeautifulSoup web scraper were used to download all available Experience reports from Erowid.org. 23,819 reports were downloaded, containing a title, an author, the name(s) of the drug(s) taken, the text of the report, when the report was published, and (when available) information about dosage, timing, gender, and age. We focused this study primarily on the drugs taken and the main text of the report.

First, the text of the 23,819 reports was stripped of editorial comments. The Erowid.org staff frequently adds comments to the reports commenting on inappropriate dosage, dangerous drugs, and dangerous combinations, and we needed to strip these comments so our text mining efforts captured only user-written comments. 1,780 comments were identified and removed from reports. Next, 128 non-English reports were identified and removed. Finally, 171 regular expression-based search-and-replacements were done to clean the drug names.

Spellings and capitalizations were corrected, different slang terms for single drugs were identified and consolidated, drug names with dosages were stripped down to merely the drug name, and slight variations of a drug (e.g. smoked marijuana and pharmaceutical THC) were combined together.

From this cleaned data, 34 drugs were identified as having more than 100 reports of that drug taken alone. We then added other drugs we felt should be included, despite having less than 100 reports of the drug taken alone. Our final list of target drugs can be viewed in Appendix B.

### ANALYSIS OF DRUG CO-USAGE

We then turned our focus to reports where more two or more of the 34 target drugs were used together. Using R, we determined relative co-usage between target drugs. We defined relative co-usage as how many times a target drug was taken with one other target drug divided by the total number of times a target drug was taken with all other target drugs. After putting this information into a matrix with both a row and a column for each drug, we had a dissimilarity matrix. This dissimilarity matrix was used to create a multi-dimensional scaling (using R's cmdscale function)—essentially a 2d map of the different drugs, where the distances between drugs most used together is small and the distance between drugs least used together is large.

MDS MAP	
Inheitaturge MelatoninDXM	Oxycodone Vicodin Codeine Morphine Ferail
Ayahuaseástanéoqi6fñifigG	mbien Slory Tobacco Paxil Cocaine Xanax
M DPT Salvia AMT twtot03028fivenW880	lethamphetamine
two	Mushrooms Cannabis
	Ketamine Nitrous LSD
fiveMeODMT	MDMA
DMT	

In the MDS map we can see some interesting relationships. The two extreme outliers, DMT and 5-MeO-DMT (fiveMeODMT), are closely related but notorious for being very strong—it is not surprising to find these drugs standing out from the rest. Alcohol and cannabis seem to be further apart than expected, but this could reflect the fact that alcohol and cannabis are often used in tandem with all other drugs. We can see that LSD and MDMA are close together, a combination which is popular enough to have its own name (candyflip). At the top of the MDS map is a cluster of opiates, suggesting their similar effects cause them to be used together regularly. Xanax and cocaine are close together, revealing an interesting relationship: people tend to use the effect of one drug to counteract the effect of the other; in this case, the downer Xanax provides a "comedown" from the upper cocaine. There are some more unexpected relationships that are revealed through this MDS map: ketamine and nitrous, where one tends to amplify the effect of the other; absinthe and crack; and heroin and Adderall.

We then lift by applying the following function:

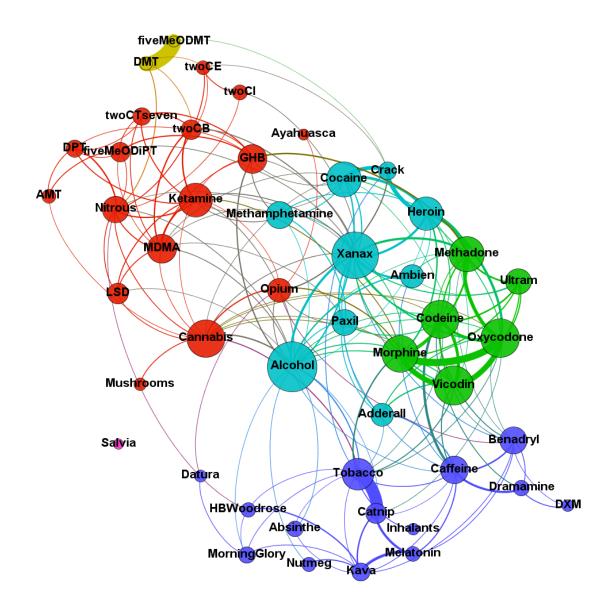
# $Lift_{A\&B} = count(drug A \& drug B used together in a report) / [count(all reports with drug A and at least one other drug) * count(all reports with drug B and at least one other drug)]$

The interpretation of this lift metric is that 1 equals what we expect the co-usage of two drugs to be by chance. Therefore, lifts greater than 1 correspond to drug combinations occurring more than chance would suggest, and lifts less than 1 correspond to drug combinations occurring less than chance would suggest.

A matrix of lifts (where cells in the matrix were the lift of the two drugs represented by the row and column) was then exported to Gephi to create a network diagram (next page). Only drug combinations with lifts above 1.0 were included in the network diagram.

In the network diagram the thickness of the edges corresponds to lift values. We see that some drug combinations have high lifts (thick lines), and some drugs combinations have lifts closer to 1 (thin lines). To get a better sense of how the drugs relate to one another, we grouped the drugs using a Gephi clustering algorithm and found 6 distinct clusters. Node colors indicate the cluster the drug belongs to. The sizes of the nodes depict each drug's eigenvector centrality score—essentially a measure of how popular that drug is across all combinations.

We see that Salvia is in a cluster by itself. Salvia (which is legal in some areas of the US) is known to be rather intense and has effects lasting only a few minutes, making it difficult to take other drugs after taking salvia. DMT and 5-MeO-DMT (fiveMeODMT) are two drugs with very similar effects; their lift values are so strong that the two are in a cluster by themselves. The green cluster corresponds to painkillers, or opiates (though opium is not part of this cluster, possibly given its low availability compared to other opiates, or its low efficacy in soothing opiate addiction). The opiates are a very tight cluster—the lift values between members of this group are very high. The red cluster contains many designer drugs and traditional psychedelics, indicating a strong relationship between the two groups of drugs. The most prominent drug in this group is cannabis, and it is interesting that cannabis does not have a higher centrality score (node size). Given the perception of cannabis it as a "gateway drug," we expected it to have the highest centrality. Perhaps this "gateway drug" notion is incorrect, or perhaps cannabis use is so routine to Erowid users they did not always note it in their reports (however, the high centrality of alcohol casts doubt on this hypothesis of routine use leading to non-reporting).



The teal cluster is dominated by stimulants and antidepressants, indicating the two groups of drugs are abused simultaneously. Research suggests stimulants can temporarily remedy depression, and our network diagram shows drug users know this. However, there are some outliers in this group--neither alcohol nor Ambien are clearly a stimulant or an antidepressant. However, many people self-medicate depression with alcohol, and it is possible that many people who suffer from depression might also have trouble sleeping and turn to Ambien.

Alcohol has the highest centrality (node size) of all of the drugs. It is not noteworthy to see alcohol's high centrality score, but Xanax has the second-highest centrality after alcohol higher than tobacco or cannabis. As mentioned earlier, one explanation is that drug users might not mention every time they smoked tobacco or cannabis, since these drugs are used so often. However, it is also possible that Xanax abuse is/was extremely common.

The purple cluster is an interesting mix of drugs that are easily acquired. Caffeine, Benadryl, DXM (Robitussin), Dramamine, Kava, Melatonin, inhalants (e.g. glue), catnip (note how high the lift between catnip and tobacco is), and nutmeg can be purchased legally by anyone, and absinthe and tobacco can be purchased by anyone of age. Datura, H. B. Woodrose seeds (HBWoodrose), and Morning Glory seeds (MorningGlory) are all from plants that are easy and legal to grow. The clustering of all these substances together suggest there is a class of drug user that tends to abuse only legal substances. Perhaps these users fear drug tests, do not want to break the law, or do not have access to illicit drugs. Regardless, the clustering of all these drugs together suggests a very specific pattern of drug abuse, and raises questions about the ability of laws to stop drug abuse.

## SENTIMENT ANALYSIS OF FULL REPORTS

Having gained an understanding of the co-usage within our set of target drugs, we turned toward sentiment analysis of full written reports as a means of insight into subjective drug experiences.

First, we stripped our data down to only reports of our target drugs taken alone—not in combination with any other drugs (we refer to this as "solo drug use"). This left us with 10,080 reports. Then we took 83 reports of solo drug use from drugs outside of our chosen set (out of a total of 6,594). These reports were read and manually assigned a rating of "positive" or "negative". Our general approach to rating was how much the report author enjoyed the drug, if the author would recommend the drug to other drug users, and if the author would use the drug again. However, we attempted to control for the addictive nature of some drugs, which was clear in some very negative reports where the author recounted highly negative experiences but then stated how much they loved the drug.

We then used the Lightside software package to automatically classify unread reports as either "positive" or "negative". Our classification method first removed all stop words (e.g. "a", "and", "the"), numbers, and punctuation, and then stemmed the remaining words (e.g. "waste", "wasted", "wasted", "wasting" all become "wast"). Then each individual report was converted into a feature vector (collection) of bigrams (one word followed by another), and part-of-speech bigrams (one part of speech followed by another, e.g. "transitive verb/indirect object"). Multimodal naive Bayes was used for classification, meaning a count of each feature in the feature vector was considered, and then a probability of being negative or positive was calculated by comparing the features of an unread report to the set of 83 reports that we read.

We then tested the algorithm on our 83 read reports. Using cross-validation, our accuracy was 66%.

Predicted> Actual Below	Positive	Negative	Total
Positive	37	8	45
Negative	20	18	38
Total	57	26	83

There was a greater tendency to classify negative reports as positive rather than positive reports as negative. This caused us to turn our analysis to the percentage of negative reports, since this would be a more telling metric—we could not be sure if reports classified as positive were truly positive (given our high rate of false positives), but we could be more assured reports classified as negative were truly negative (given our comparably low rate of false negatives).

Rank (by % of negative sentiment)	Drug Name	Negative Sentiment Percentage
1	Melatonin	68.52%
2	Paxil	63.49%
3	Ultram	57.58%
4	Catnip	53.66%
5	Tobacco	53.33%
6	Ambien	52.94%
7	Kava	49.02%
8	Opium	48.72%
9	Caffeine	48.25%
10	GHB	46.34%
11	Methadone	43.40%
12	Xanax	41.38%
13	Morphine	40.00%
14	Cocaine	39.13%
15	Codeine	38.67%
16	Absinthe	34.83%
17	Oxycodone	34.31%

18Acoha3.33%19Vicolin3.33%19Numeg3.33%20Numeg2.75%21Aderal3.18%22Aderal3.18%23Nitrous2.77%24Datura2.70%25Datura3.33%26Crack3.33%27MDMA3.40%28Menaryla3.40%29Methamphetami8.69%30Ketamine1.86%31Ketamine1.86%32Mondos1.93%34Mondos1.23%35Ketamine1.62%36MorrigGiory1.57%37Ketamine1.57%36MorrigGiory1.57%37MorrigGiory1.53%38MorrigGiory1.33%39MorrigGiory1.33%39MorrigGiory1.33%30MorrigGiory1.33%31MorrigGiory1.33%32MorrigGiory1.33%33MorrigGiory1.33%34MorrigGiory1.29%35MorrigGiory1.29%36MorrigGiory1.29%37MorrigGiory1.29%38MorrigGiory1.29%39MorrigGiory1.29%30MorrigGiory1.29%31MorrigGiory1.29%32MorrigGiory1.29%33MorrigGiory1.29%34 <th></th> <th>[</th> <th></th>		[	
20Nutmeg32.75%21Inhalats32.31%21AdderallS2.31%22AdderallS2.13%23Nitrous27.27%24DXM6.07%25Datura25.70%26Crack33.3%27MDMA3.40%28Benadryl21.98%29Methamphetamin18.69%31Ketamine18.69%32Cranabis17.93%34MorningGlory15.73%35Kutorose15.23%36NorningGlory15.73%37MordeSuery15.73%38NorningGlory15.23%39MordeSuery15.33%39MordeSuery15.33%39MordeSuery13.33%30Diramanine13.33%31Suery14.34%32NordeSuery10.43%34NordeSuery10.43%34Suery10.23%35Suery10.23%36Suery10.33%37Suery10.23%38Suery10.23%39Suery10.23%30Suery10.23%31Suery10.23%32Suery10.23%33Suery10.23%34Suery10.23%35Suery10.23%36Suery10.23%37Suery10.23%38Suery10.23% <td>18</td> <td>Alcohol</td> <td>33.33%</td>	18	Alcohol	33.33%
ProcessProcess21Inhalants32.31%22Adderall32.18%23Nitrous7.27%24DXM26.07%25Datura25.70%26Crack23.33%27MDMA23.40%28Benadryl21.98%29Methamphetamin18.69%30Heroin18.69%31Canabis17.93%32Canabis17.93%34MorringGlory15.23%35fiveMcODIPT15.76%36twoCTseven15.23%37AMT11.31%38DPT13.38%39AMT11.31%41twoCE10.48%42twoCE10.20%43Musrioms8.43%44twoCI8.53%45LDDKatyacout46LDDKatyacout	19	Vicodin	33.33%
22Adderall32.18%23Nirous27.27%24DXM6.07%25Datura25.70%26Crack3.53%27MDMA23.40%28Benadryl1.98%29Methamphetamin18.69%30Heroin18.69%31Canabis17.93%34MOMA15.25%35H8Woodrose17.23%36KoringGlory15.76%37KovCrseven15.23%38Dramanice15.37%39AMT1.31%40DPT1.31%41twoCE10.48%42twoCI8.53%44Kukrooms8.43%45LD8.43%46LD8.43%	20	Nutmeg	32.75%
11123Nitrous27.27%24DXM26.07%25Datura25.07%26Crack23.53%27MDMA23.40%28Benadryl19.8%29Methamphetamin18.69%30Heroin18.69%31Canabis17.93%32Canabis17.93%34MorningGlory15.76%35fiveMcODIPT15.76%36twoCTseven13.38%37Datura13.38%39AMT11.21%41twoCE10.48%42twoCI3.53%44twoCI8.43%45Kushrooms8.43%46LDD8.43%	21	Inhalants	32.31%
24DXM26.07%25Datura25.70%26Crack23.53%27MDMA23.40%28Benadryl21.98%29Methamphetamin8.69%30Heroin18.69%31Ketamine18.06%32Cannabis7.93%34MorningGlory16.72%35fiveMeODPT15.76%36twoCTseven15.33%37MorningCom15.33%38Damamine13.38%39AMT11.31%41twoCB11.29%42twoCE0.20%43Salvia9.41%44twoClame8.53%45Lixplen5.53%46Lixplen8.09%	22	Adderall	32.18%
25Datura25.70%25Datura25.70%26Crack35.3%27MDMA3.40%28Benadryl21.98%29Methamphetamin8.78%30Heroin18.69%31Ketamine18.06%32Cannabis17.93%34MorningGlory16.72%35twoCrseven5.23%36biveMcODPT15.07%37fiveMeODPT15.07%38Daramine13.3%39AMT1.31%41twoCB1.29%42kwoCl9.41%43Salvia8.53%44twoCl8.3%45Subian5.3%46SD8.41%	23	Nitrous	27.27%
26Crack23.53%27MDMA23.40%28Benadryl21.98%29Methamphetamine18.78%30Heroin18.69%31Ketamine18.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37AMT11.31%39AMT11.31%40DPT11.29%41twoCE10.20%42twoCI9.41%44twoCI8.33%45LSD8.43%46LSD8.41%	24	DXM	26.07%
27MDMA23.40%27MDMA23.40%28Benadryl21.98%29Methamphetamine18.78%30Heroin18.69%31Ketamine18.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODIPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramanine13.38%39AMT11.31%40DPT11.29%41twoCB10.20%43Salvia9.41%44twoCI8.53%45LSD8.43%46LSD8.09%	25	Datura	25.70%
28Benadryl21.98%29Methamphetamine18.78%30Heroin18.69%31Ketamine18.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramanine13.38%39AMT11.31%40DPT11.29%41twoCE10.48%42Salvia9.41%43Salvia8.33%44LSD8.43%45LSD8.41%	26	Crack	23.53%
29Methamphetamine18.78%30Heroin18.69%31Ketamine18.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODIPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramamine13.38%39AMT11.31%40DPT11.29%41twoCE10.48%42Salvia9.41%43Salvia8.43%45LSD8.41%47Ayahuasca8.09%	27	MDMA	23.40%
30Heroin18.69%31Ketamine18.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramamine13.38%39AMT11.31%40DPT11.29%41twoCB10.48%42twoCI8.53%43Salvia9.41%44LSD8.43%45LSD8.41%47Ayahuasca8.09%	28	Benadryl	21.98%
And 31Ketamine8.06%32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramanine13.38%39AMT11.31%41twoCB10.48%42twoCE10.20%43Salvia9.41%44twoCl8.53%45LSD8.43%47Kaphasca8.09%	29	Methamphetamine	18.78%
32Cannabis17.93%33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramamine13.38%39AMT11.31%40DPT11.29%41twoCE10.48%42twoCE10.20%43Salvia9.41%44LtwoCl8.53%45LSD8.41%47Ayahuasca8.09%	30	Heroin	18.69%
33HBWoodrose17.82%34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramamine13.38%39AMT11.31%40DPT11.29%41twoCB10.48%42twoCE10.20%43Salvia9.41%44twoCI8.53%45Mushrooms8.43%46LSD8.09%	31	Ketamine	18.06%
34MorningGlory16.72%35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramanine13.38%39AMT11.31%40DPT11.29%41twoCB10.48%42twoCE10.20%43Salvia9.41%44twoCl8.53%45LSD8.41%47Ayahuasca8.09%	32	Cannabis	17.93%
35fiveMeODiPT15.76%36twoCTseven15.23%37fiveMeODMT15.07%38Dramamine13.38%39AMT11.31%40DPT11.29%41twoCB10.48%42twoCE10.20%43Salvia9.41%44twoCl8.53%45Mushrooms8.43%46LSD8.09%	33	HBWoodrose	17.82%
36 twoCTseven 15.23%   37 fiveMeODMT 15.07%   38 Dramamine 13.38%   39 AMT 11.31%   40 DPT 11.29%   41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%	34	MorningGlory	16.72%
37 fiveMeODMT 15.07%   38 Dramamine 13.38%   39 AMT 11.31%   40 DPT 11.29%   41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%	35	fiveMeODiPT	15.76%
38 Dramamine 13.38%   39 AMT 11.31%   40 DPT 11.29%   41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%	36	twoCTseven	15.23%
39 AMT 11.31%   40 DPT 11.29%   41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%	37	fiveMeODMT	15.07%
40 DPT 11.29%   41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%	38	Dramamine	13.38%
41 twoCB 10.48%   42 twoCE 10.20%   43 Salvia 9.41%   44 twoCI 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%   47 Mushrooms 8.09%	39	AMT	11.31%
42 twoCE 10.20%   43 Salvia 9.41%   44 twoCl 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%   47 Ayahuasca 8.09%	40	DPT	11.29%
43 Salvia 9.41%   44 twoCl 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%   47 Ayahuasca 8.09%	41	twoCB	10.48%
44 twoCl 8.53%   45 Mushrooms 8.43%   46 LSD 8.41%   47 Ayahuasca 8.09%	42	twoCE	10.20%
45 Mushrooms 8.43%   46 LSD 8.41%   47 Ayahuasca 8.09%	43	Salvia	9.41%
46 LSD 8.41%   47 Ayahuasca 8.09%	44	twoCl	8.53%
47 Ayahuasca 8.09%	45	Mushrooms	8.43%
	46	LSD	8.41%
48 DMT 7.37%	47	Ayahuasca	8.09%
	48	DMT	7.37%

Given the imprecise nature of sentiment analysis, we are hesitant to make conclusions about individual drugs. However, we do notice some trends. Heavy psychedelics tend to have very positive sentiments (ranks 48-33 are all psychedelics, with the exception of rank 43 and rank 38). Among the more negatively experienced drugs, we see a lot of prescription drugs and legal drugs. It is possible the negative sentiments of these drugs are due to a lack of powerful effects, rather than truly negative experiences. We also see a clustering of most of the opiate based drugs in the middle in ranks 11-19 (with the exception of ranks 14, 16, and 19). This might reflect the mixed experience of opiates, with their pleasurable effects tainted by the high possibility of addiction. The only opiates outside of this area are the prescription drug Ultram in rank 3, Opium in row 8, and Heroin in row 30. The low negative sentiment of heroin is unsurprising—if heroin left users feeling negative, it would not be so addictive.

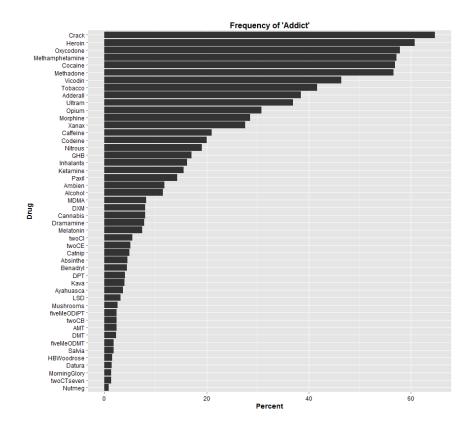
### ANALYSIS OF WORD FREQUENCY BY DRUG

Next, we analyzed what topics are being disused most with what drugs. To do this, we considered the frequency of different stemmed words in reports of different drugs. These words were: body, music, myself, sleep, think, and trip (note these are stems, so "trip", "tripping", and "tripped" would all be counted as "trip"). The percentage of reports of each target drug containing the searched word (counting only solo drug reports) is below:

twoCTseven	56.91	0	55.17	36.84	61.22	61.5
twoCl	61	68.24	50.33	43.67	57.6	56.63
twoCE	54.3	68.87	52.63	32.61	58.62	53.83
twoCB	61.7	68.33	41.33	0	51.52	53.88
Salvia	43.24	58.9	38.69	36.64	52.62	54.05
Nutmeg	41.11	65.71	38.27	32.89	39.86	51.79
Mushrooms	41.15	57.66	42.76	28.05	47.74	49.4
MorningGlory	40.27	54.76	41.62	34.27	44.23	49.16
MDMA-	44.93	56.84	46.93	31.61	48.33	47.83
LSD	43.94	59.87	40.14	26.91	44.85	47.67
HBWoodrose	51.85	46.88	44.71	25	53.25	49.64
fiveMeODiPT	66.86	66	47.31	29.07	50.62	59.56
DXM	41.86	63.87	37.93	30.51	49.77	50.71
Dramamine	38.03	0	30.14	38.3	40	44.19
DPT	54.44	50.98	36.78	0	53.25	56.93
DMT	38.98	61.11	50	0	46.9	55.19
Datura	24.59	0	41.67	41.3	32.8	46.63
Cannabis	37.83	50.67	39.14	29.39	38.51	52.22
Benadryl	37.63	0	28.28	35.23	39.5	51
Ayahuasca	47.14	75.44	45.21	0	53.51	58.68
AMT	53.51	53.45	40.43	23.58	43.06	51.4
	body	music	myself	sleep	think	trip

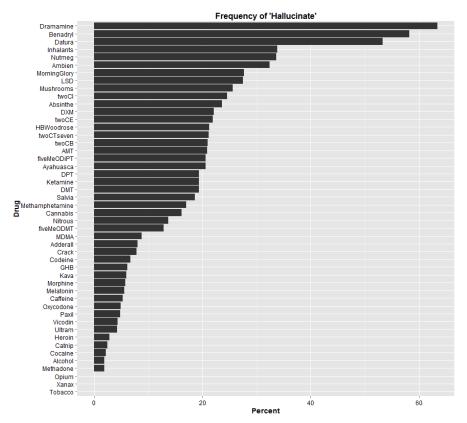
From this, we can pick out strong associations such as melatonin & sleep or LSD & music and infer that people taking melatonin are discussing sleep and people taking LSD are discussing music. It is difficult to see significant patterns here, and a better approach might have been to show which drugs have word occurrences above and below the average.

We then went a bit deeper with two more words, addict and hallucinate (both stemmed), and expanded the selection of drugs:

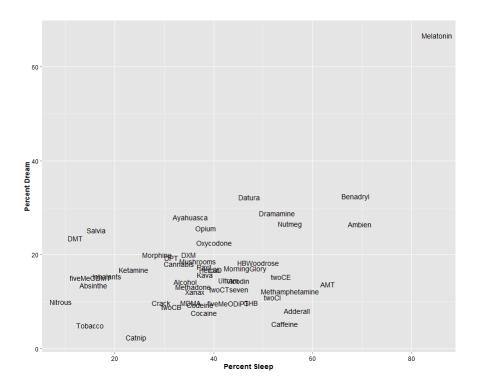


With addict, we unsurprisingly see that crack, heroin, methamphetamine, cocaine, and various opiates have reports where forms of the word "addict" appear most. More surprising is how many drugs are never discussed in an addition context. Addiction is a much greater concern to users of prescription drugs, tobacco, and caffeine than to users of powerful hallucinogens and cannabis.

A similar plot for forms of the word "hallucinate" also shows interesting findings (next page). We see that the two drugs garnering the most "hallucination" related discussion are Dramamine and Benadryl, two over-the-counter drugs. This tells us that these drugs are being abused with the goal of hallucinations. While this analysis does not tell us if these drugs actually produce hallucinations, we can infer that users of these drugs are at least hoping to hallucinate. Inhalants, nutmeg, and Ambien are also seen near the top of the plot. All of these drugs are above traditional hallucinogens like mushrooms and LSD, suggesting there is more discussed than hallucinations when people write about mushrooms and LSD.



Lastly, we plotted drugs on a scatterplot where the percentage of a drug's reports mentioning "sleep" is on the x-axis and the percentages a drug's reports mentioning "dream" is on the y-axis:



Melatonin, as mentioned before, is clearly a drug which lends itself to discussion of sleep and dreams. On the other hand, tobacco, catnip, and nitrous have little to no association with sleep or dreams. Obviously, we expect a positive relationship between occurrences of "dream" and occurrences of "sleep" in a report. However, some drugs are outside of this positive relationship. DMT and salvia are the subject of dream discussion much more than sleep discussion, and Adderall and caffeine are the subject of sleep discussion but very little dream discussion. This exposes some of the limitations of this analysis—DMT and salvia users might be discussing hallucinations with phrases like "dream-like" or words like "dreamy", and Adderall and caffeine users are likely discussing the sleep problems these stimulants cause.

### INDIVIDUAL WORD SENTIMENT

Finally, we turned our attention to the sentiment individual words, e.g. if a report mentions "hallucinations", are those hallucinations positive or negative? A python program was written to capture text snippets around a target word. These snippets were isolated and grouped by target word stem and drug. We then took 653 snippets from reports on individual drugs that were not part of our target drug list, and graded the sentiment of these snippets individually. Because many of the snippets had very little to say about the experience of the word in relation to the drug, we ended up with 112 positive, 90 negative, and 451 irrelevant snippets to use for training our classifier. We only classified a snippet as positive or negative if it was clearly positive or negative AND directly discussing the drug or the drug's effects in relation to the target word.

We then fed our snippets into Lightside, and built a model identical to the model we used for overall sentiment analysis, with the addition of unigrams (single words) to our model. Our final model had a test accuracy of 53.6%, but focusing only on "positive" vs. "negative" classification/misclassification, our model had a test accuracy of 67.8%.

Predicted> Actual Below	Positive	Negative	Irrelevant	Total
Positive	41	11	88	112
Negative	14	33	87	90
Irrelevant	57	46	276	451
Total	112	90	451	653

For our analysis, we only focused on "positive" vs. "negative", noting that misclassifications of either "positive" or "negative" to "irrelevant" were about equal in our test data. We also stripped out our target words to remove their impact from the sentiment classification. This gave us an overall idea of positive or negative experiences for each word in relation to the reported drug.

We set a threshold of 50 total mentions of a word with a drug to consider the sentiment. We then eliminated words that did not have 50 total mentions for at least 75% of the drugs, and then eliminated drugs that lacked 50 mentions for more than one word.

	body	music	myself	sleep	think	trip
Salvia	43.24%	58.90%	38.69%	36.64%	52.62%	54.05%
Mushroom	41.15%	57.66%	42.76%	28.05%	47.74%	49.40%
Cannabis	37.83%	50.67%	39.14%	29.39%	38.51%	52.22%
LSD	43.94%	59.87%	40.14%	26.91%	44.85%	47.67%
MDMA	44.93%	56.84%	46.93%	31.61%	48.33%	47.83%
DXM	41.86%	63.87%	37.93%	30.51%	49.77%	50.71%
twoCl	61.00%	68.24%	50.33%	43.67%	57.60%	56.63%
MorningGl	40.27%	54.76%	41.62%	34.27%	44.23%	49.16%
Nutmeg	41.11%	65.71%	38.27%	32.89%	39.86%	51.79%
DMT	38.98%	61.11%	50.00%		46.90%	55.19%
Datura	24.59%		41.67%	41.30%	32.80%	46.63%
HBWoodro	51.85%	46.88%	44.71%	25.00%	53.25%	49.64%
twoCE	54.30%	68.87%	52.63%	32.61%	58.62%	53.83%
Benadryl	37.63%		28.28%	35.23%	39.50%	51.00%
AMT	53.51%	53.45%	40.43%	23.58%	43.06%	51.40%
fiveMeODi	66.86%	66.00%	47.31%	29.07%	50.62%	59.56%
twoCTseve	56.91%		55.17%	36.84%	61.22%	61.50%
Dramamin	38.03%		30.14%	38.30%	40.00%	44.19%
Ayahuasca	47.14%	75.44%	45.21%		53.51%	58.68%
twoCB	61.70%	68.33%	41.33%		51.52%	53.88%
DPT	54.44%	50.98%	36.78%		53.25%	56.93%

There are a handful of interesting findings from this analysis, but we start with the less interesting findings. First, none of these drugs have a very positive association with "sleep", although since most of these drugs are hallucinogens that is not surprising. What is more interesting is that Benadryl has low sentiment for "sleep"—this suggests many Benadryl report writers did not desire this effect. Second, we see that most of these drugs had positive sentiments for "music", again not surprising given the sample of users and the goal of the drug use. H. B. Woodrose seeds (HBWoodrose) came in very low with "music", and ayahuasca came in very high. While we lack the domain knowledge to speculate why H. B. Woodrose seeds users did not find "music" positive, ayahuasca is often done in a ceremonial setting complete with music, which might explain why music perceived so positively, as it is an integral part of the experience. Lastly, the word "trip" shows relatively consistent results across all the drugs.

"Body", "myself", and "think" differ greatly across the drugs and are thus more interesting for analysis. "Body" has extreme low sentiment for Dramanine, Benadryl, and Datura. While we lack domain knowledge to speculate on the reasons Datura has such negative sentiment for "body", we believe that Benadryl's and Dramamine's sedative side effects lead to a negative "body" experience. 2-CI (twoCI), 5-MeO-DIPT (5MeoDIPT), and 2-CB (twoCB) have high positive sentiment with "body", possibly indicative of their designer drug origins. Designer drugs are especially targeted towards use in clubs and at raves, and the fact that these drugs have positive sentiment for "music" as well as "body" are suggestive of this designer drug origin and purpose. "Myself", which relates generally to the effect of a drug on one's sense of self, also shows a wide mix of results. We again see low values for Benadryl and Dramamine, although the reason for this is unknown to us. And again, we see positive sentiments for designer drugs.

Lastly, we turn our attention towards "think", a proxy for the effect of the drug on cognition. Datura comes in the most negative, which is not surprising. Datura blocks acetylcholine, an important neurotransmitter which has a key role in cognition. And again, we see low sentiment for Benadryl and Dramamine. What is most interesting here is the low sentiment for cannabis and "think," echoing the popular belief that marijuana impairs thinking. We also note that many classic drugs like LSD and mushrooms come in with lower "think" sentiment than many designer drugs. Overall, across the whole report, this is a general trend we see—newer designer drugs are more positively viewed than their traditional counterparts, which say a lot about their growing popularity.

As a final note, it is not hard to see some correlation between the sentiments of these words and the overall sentiments of drugs in the earlier section. However, we feel this individual word analysis could contribute much to a breakdown of how different factors are taken into account by Erowid's users when evaluating a drug, and could also help provide information about \*why\* a drug was perceived negatively or positively.

### **Appendix A: Sample Reports**

"I found a bottle of 30 lyrica capsules. i took them and felt relaxed. I was on Xanax, Klonopin, and Restoril awhile back, and was a big pharmaceutical abuser. I knew Lyrica was Schedule 5 for some reason. I took about 5-10 of the 50mg capsules and felt the sedation and euphoria from benzos and barbituates. be safe because this is a new drug. i am going to ask my doctor for a script."

"Wreathed by the silence of the woods around the yurt where this is taking place, I state my question and take a sip of the smoke. Then I try another. A sudden feeling of gravity surges down my body, startling me into quickly handing my friend the pipe. Then I close my eyes, steady my breath, sit up straight, get attentive to what I see, and revisit my question. I throw in some prayers and affirmations for good measure. I am present and in control rather than feeling required to endure something I can't handle. This is a smooth encounter, a conversation, rather than a journey. The visuals are extremely mild; muted dark blues and yellows form gently morphing arabesques."

"12:23 am - OK, I'm ready for this to be over. It has lessened immensely but I am still feeling the effects. I just can't stand being jittery at all. To me its the WORST feeling, and its the reason I was taking benzos to begin with. Still shaking my leg, and I want to puke again. i can still hear sounds that I haven't noticed before... the distinct whir of each & every propeller of my ceiling fan, cars passing that are blocks away from my house.12:34 am - This is definitely interesting stuff that I took too much of for my first (and last) time. Heart is still beating fast & uneven-like. Arm, neck, shoulder & back muscles are sore. I keep wanting to stretch. Still itchy, and I have a mild headache. My stomach is unsettled still. 1:02 am - Still exactly the same as an hour ago. Does this end??? 1:12 am - I feel horrible. SO nauseous. Jittery. Achy & sore. I'd love to go to sleep but my stomach is turning over & over. Not a fan anymore."

"Yep, a night in the hospital in the most awful psychic agony I have ever experienced while they poured ipecac (SP) and charcoal down me 'cause the little bastards refused to let me puke. I was pumped. I was then hooked to a heart monitor and had the distinct displeasure of seeing my heart stop a couple of times. After one of them and a couple of firm thumps to my sternum I asked the doctor if I was going to make it. He was rather preocupied with saving my life and sorta muttered, 'we don't know.' They would not treat me with anything until the mushroom was identified which they did by flying it to a poison control center in Denver, I think it was, by Navy jet from Moffet field. They then shot me with something that had me down in 15 minutes after about six hours of mental horror. The next morning every damn med student and intern came by to find out what it was like. I didn't have good things to report. Your mileage may vary."

"I chewed betel nut many times when I was in Papua New Guinea on a research project. Everybody (including kids of 3) chews it there in the areas that it grows naturally. It has a kind of sour/bitter taste that I liked a lot. I usually just chewed it green, but a few times I chewed it with lime and daka, or betel pepper. A couple times stand out in my memory. Once I asked for a really strong betel nut and the local buai (betel) fiend gave me one. I chewed it up with a bunch of lime and daka. It totally made my head spin and it was a lot like the first time you smoke a cigar and forget that you're not supposed to inhale. I sweated a lot and I was kind of zoning out. Somebody asked if I was OK and I said 'Oh, mi spak nau' (Tok Pisin for 'I'm buzzing') Everybody laughed. Most times when I chewed it with lime though I just got animated and chatty (not my usually speed, I'm normally a fairly quiet guy), sweated a lot, and generally felt up, like I'd just had a good strong unfiltered cigarette or cup of coffee. Overall I really enjoyed it. I think it's great stuff as long as you don't go nuts with it. There were people with black teeth in PNG from chewing it constantly, and one guy I knew was clearly addicted. I wish I could find some in the states. It would be great every now and again. Also, I would love to see people's reaction when I spit a big stream of orange red betel juice!"

> 100 reports				
Salvia	Mushrooms	Cannabis	LSD	MDMA
DXM	twoCl	MorningGlory	Cocaine	Nutmet
Methamphetamine	fiveMeODMT	DMT	Datura	HBWoodrose
twoCE	Benadryl	Adderall	AMT	fiveMeODiPT
Ultram	Ketamine	twoCTseven	Dramamine	Ayahuasca
Inhalants	twoCB	DPT	Caffeine	Nitrous
Heroin	Alcohol	Oxycodone	Kava	Ambien
< 100 reports				
Absinthe	GHB	Codeine	Vicodin	Paxil
Tobacco	Xanax	Melatonin	Methadone	Crack
< 50 reports				
Catnip	Opium	Morphine		

Appendix B: Drugs Used in	Analysis with Report C	ounts of the Drug Taken by Itself