Springboard—DSC Program

Capstone Project 1 Milestone Report

Determining the Likelihood and Type of Drug Abuse Visits to Emergency Departments in the US

By Laura Eshee

April, 2020

Background

According to The National Institute on Drug Abuse, “Addiction is a chronic disease characterized by drug seeking and use that is compulsive, or difficult to control, despite harmful consequences. The initial decision to take drugs is voluntary for most people, but repeated drug use can lead to brain changes that challenge an addicted person’s self-control and interfere with their ability to resist intense urges to take drugs. These brain changes can be persistent, which is why drug addiction is considered a "relapsing" disease—people in recovery from drug use disorders are at increased risk for returning to drug use even after years of not taking the drug.”[1]

Drug abuse occurs when a person takes a substance, whether illegal, prescribed or over the counter, for purposes other than those in which they are meant to be used, or when a person takes large quantities of the substance. Typically, the person is using the drug to alter his or her mood or feel better and not for a health reason.

Statistics for drug abuse include:

- Almost 21 million Americans have at least one addiction, yet only 10% of them receive treatment.[2]
- Drug overdose deaths have more than tripled since 1990.[2]
- From 1999 to 2017, more than 700,000 Americans died from overdosing on a drug.[2]
- More than 90% of people who have an addiction started to drink alcohol or use drugs before they were 18 years old.[2]
- Americans between the ages of 18 and 25 are most likely to use addictive drugs.[2]
- Alcohol and drug addiction cost the U.S. economy over $600 billion every year.[2]
- During 2008–2011, an average of 1.1 million emergency department (ED) visits were made each year for drug poisoning, with a visit rate of 35.4 per 10,000 persons.[3]

- About one-quarter (24.5%) of drug-poisoning ED visits resulted in hospital admission.[3]

Client and Problem Statement

Emergency Departments in hospitals nationwide must be prepared to accept patents that are suffering from drug abuse. They need to plan for enough staffing and supplies to handle the expected types and percentages of drug abuse related Emergency Department visits so that the departments' can be prepared with enough staffing, medications, procedures, etc. to properly deal with these cases and have good outcomes.

Dataset

The Drug Abuse Warning Network (DAWN) is a public health surveillance system that monitors drug abuse related visits to emergency departments in hospitals in large metro areas across the US. According to the Substance Abuse and Mental Health Services Administration (SAMHDA),

> "A DAWN case is any ED visit involving recent drug use that is implicated in the ED visit. DAWN captures both ED visits that are directly caused by drugs and those in which drugs are a contributing factor, but not the direct cause of the ED visit. Annually, DAWN produces estimates of drug-related visits to hospital EDs for the nation as a whole and for selected metropolitan areas.
>
> DAWN is used to monitor trends in drug misuse and abuse, identify the emergence of new substances and drug combinations, assess health hazards associated with drug abuse, and estimate the impact of drug misuse and abuse on the Nation's health care system. DAWN relies on a longitudinal probability sample of hospitals located throughout the United States.
>
> To be eligible for selection into the DAWN sample, a hospital must be a non-federal, short-stay, general surgical and medical hospital located in the United States, with at least one 24-hour ED. The dataset includes demographics, drugs involved in the ED visit (up to 16 drugs from 2004 through 2008 and up to 22 drugs from 2009 through 2011), toxicology confirmation, route of administration, type of case, and disposition of the patient following the visit.
>
> Prepared DAWN Emergency Department National and Metro data tables are available on the DAWN website. The DAWN website also provides access to DAWN reports."[4]

Data Wrangling

The initial step in the project was to 'wrangle' the data, which is the process of taking raw data and transforming it into a format that is suitable for analysis.

First, the DataFrame was examined to see its size and column names. It has 284 columns, and 229,221 rows. Since the DataFrame is very large and somewhat unwieldy, the decision was made to remove the 'sdled' and 'CATID' columns. These columns won't be used in the analysis because their inclusion is beyond the scope of this project. The modified DataFrame has 84 columns.

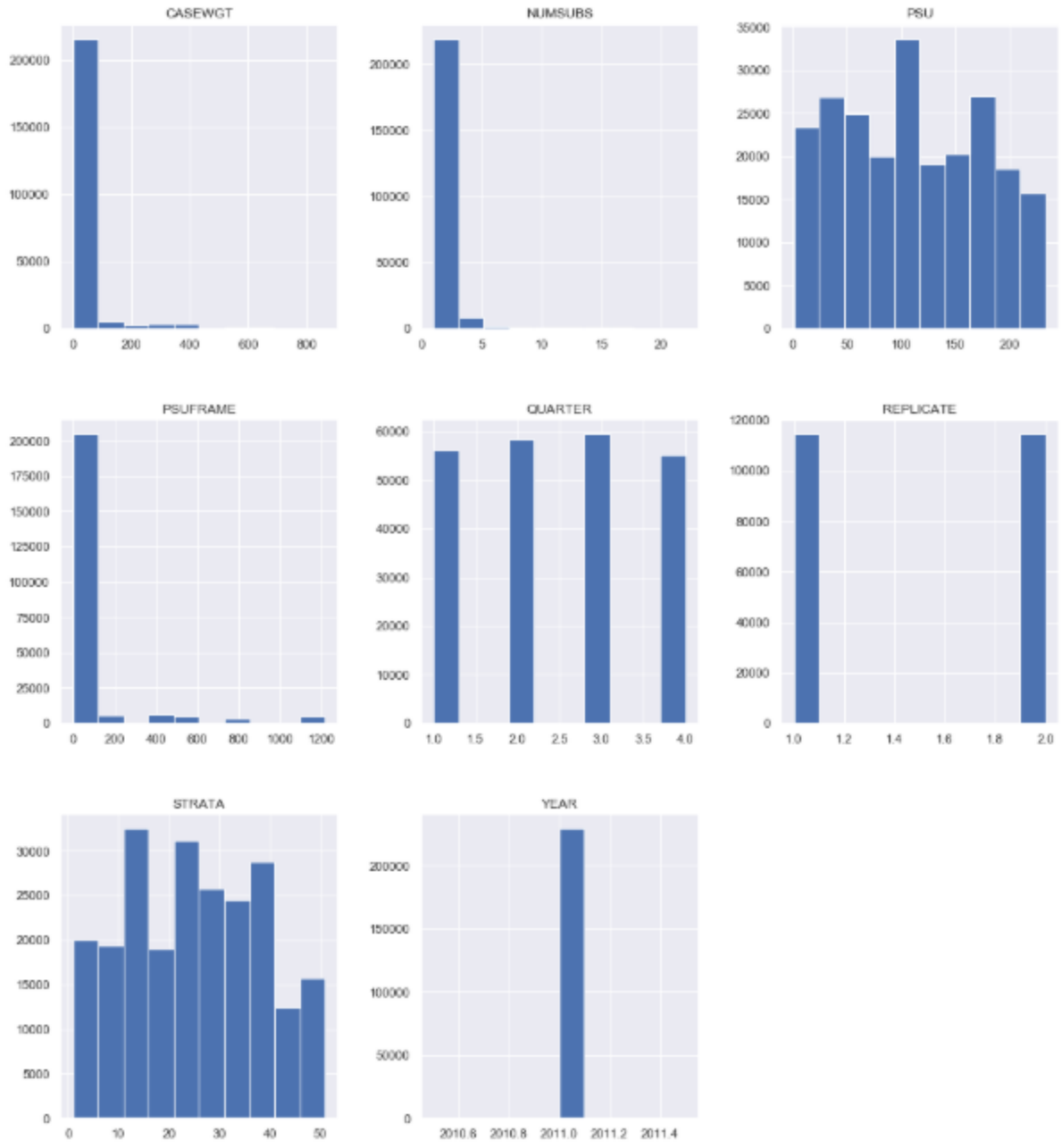
Second, summary data for all columns was obtained. Below is a snippet of the results:

	METRO	STRATA	PSU	REPLICATE	CASEWGT	PSUFRAME	AGECAT	SEX	RACE	YEAR
count	229211.000000	229211.000000	229211.000000	229211.000000	229211.000000	229211.000000	229211.000000	229211.000000	229211.000000	229211.0
mean	6.670090	24.681551	109.610839	1.500028	22.107901	73.348304	7.352370	1.477289	0.278656	2011.0
std	4.719905	13.331152	64.444097	0.500001	68.403862	211.852785	2.691575	0.526280	3.463027	0.0
min	1.000000	1.000000	1.000000	1.000000	0.938440	2.000000	-8.000000	-8.000000	-8.000000	2011.0
25%	2.000000	13.000000	53.000000	1.000000	2.714999	8.000000	5.000000	1.000000	1.000000	2011.0
50%	5.000000	25.000000	109.000000	2.000000	4.190787	10.000000	8.000000	1.000000	1.000000	2011.0
75%	12.000000	35.000000	165.000000	2.000000	7.148615	17.000000	9.000000	2.000000	2.000000	2011.0
max	14.000000	51.000000	233.000000	2.000000	862.824350	1215.000000	11.000000	2.000000	4.000000	2011.0

Next, the number of distinct values per column was counted. The column with the most distinct values was CASEWGT, with 2,931. DRUGID_1 followed that with 1,604 values. Next were DRUGID_2 with 1,071 values, DRUGID_3 with 787 values, DRUGID_4 with 634 values…ALLABUSE with 2 values and YEAR with 1 value.

Many of the variables in the DataFrame are categorical but are represented by numbers. In order to tell the data story better, it was decided to replace the numbers with the corresponding label. This replacement was done for all the variables including the 2600 drug names in the data glossary that was available with the dataset. Also, though there exists a shorter, cleaner way to perform the replacements than to do them one by one, due to time constraints, it was decided to use the method that is working rather than spend more time researching a more 'pythonic' method. Additionally, there were too many drug names to replace them using the same method as with the others. Therefore, it was decided to replace them by importing a dictionary and mapping the names to the numbers.

Then, subplot histograms of all quantitative variables were plotted. The resulting plots are below:

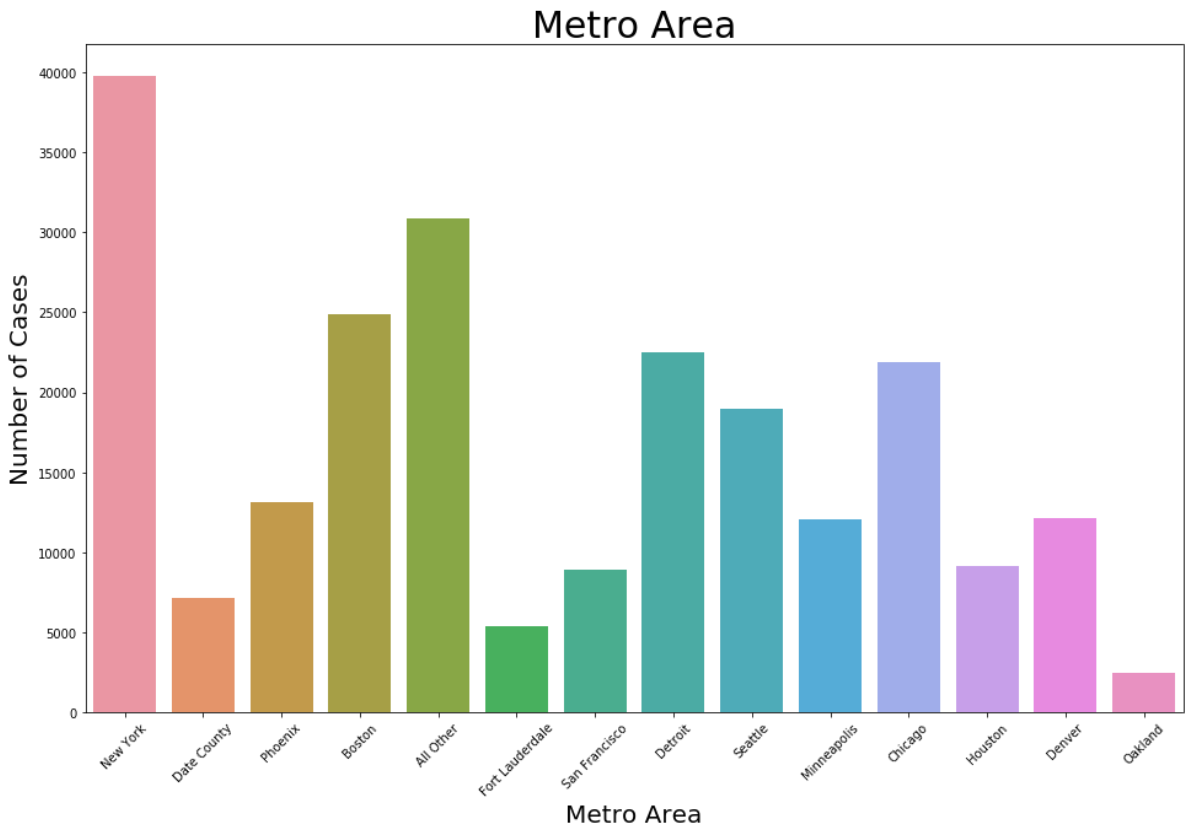


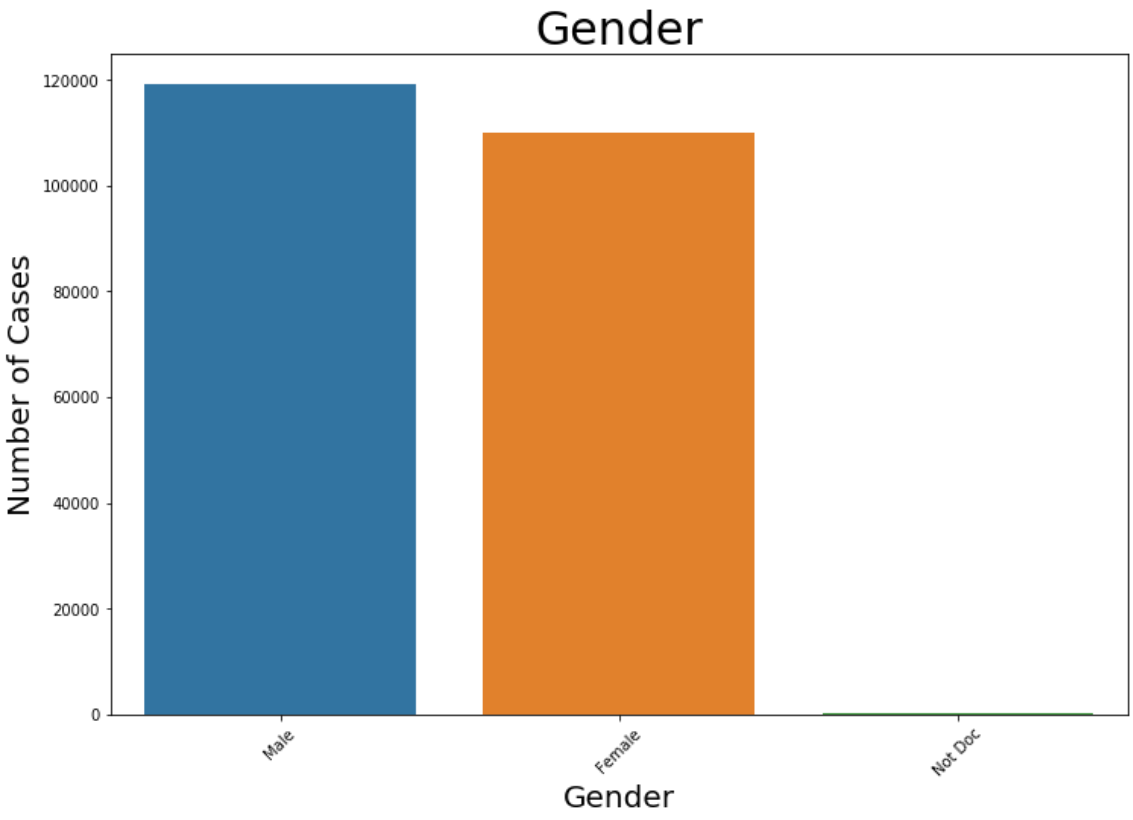
Finally, the DataFrame was checked for missing values. None were found.

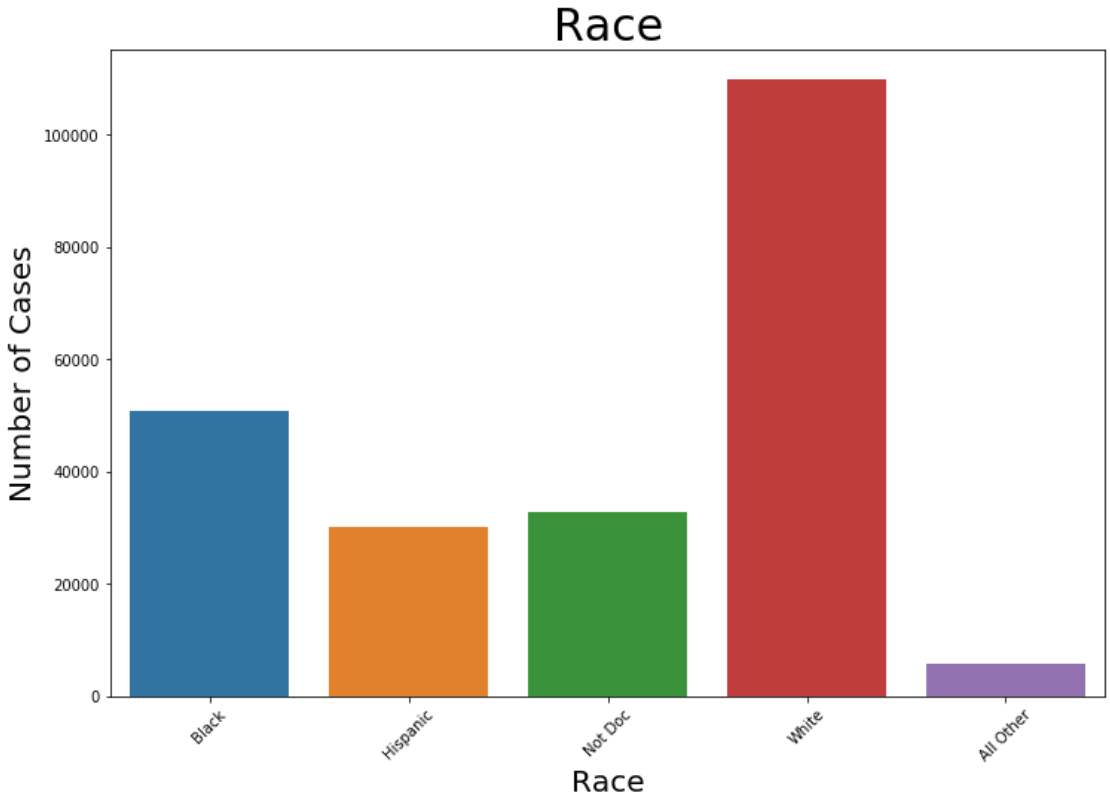
In summary, this data set has 86 columns and 229,211 rows. Most of the variables are categorical. In the original data set, the categorical variables were represented by numbers. Steps were taken to replace the numbers with their corresponding names. Since the DataFrame is very large and somewhat unwieldy the decision was made to remove the 'sdled' and 'CATID' columns. These columns won't be used in the analysis because their inclusion is beyond the scope of this project. Additionally, the number of distinct values for each variable was plotted, subplot histograms of the quantitative variables were plotted, and it was determined that the data set did not include any missing values.

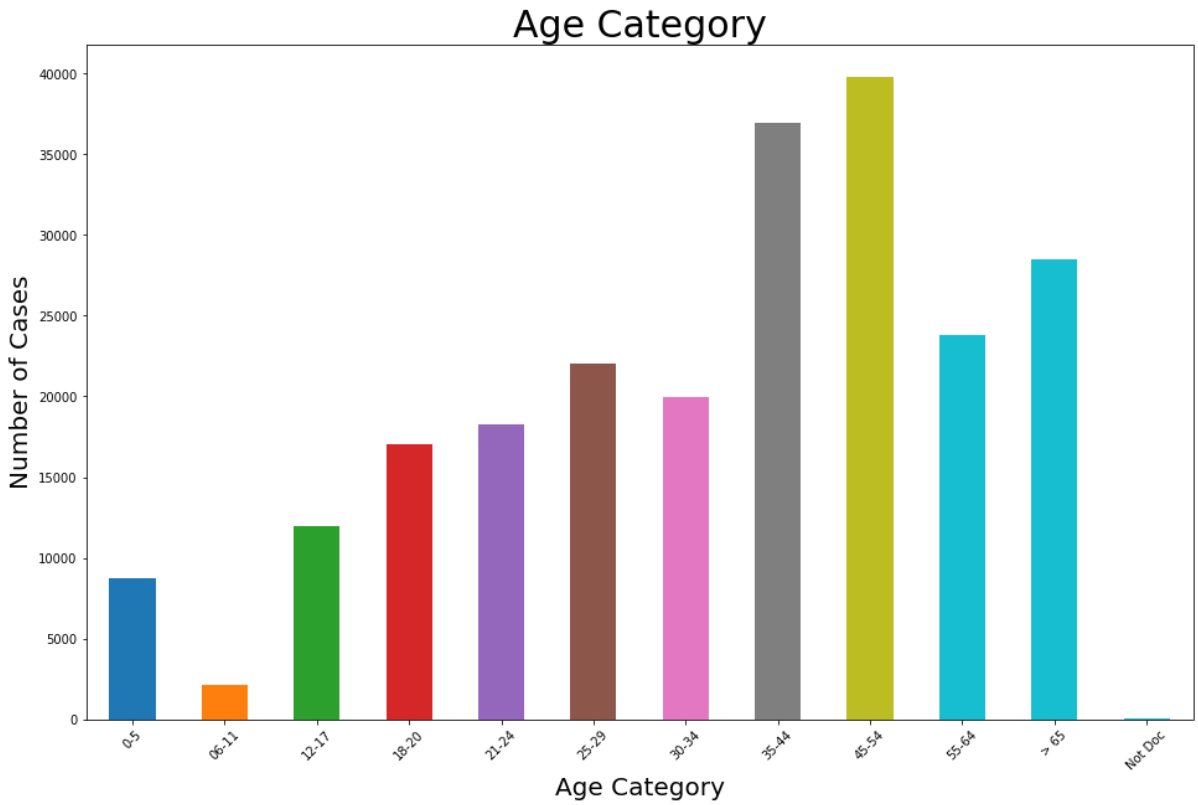
Data Story

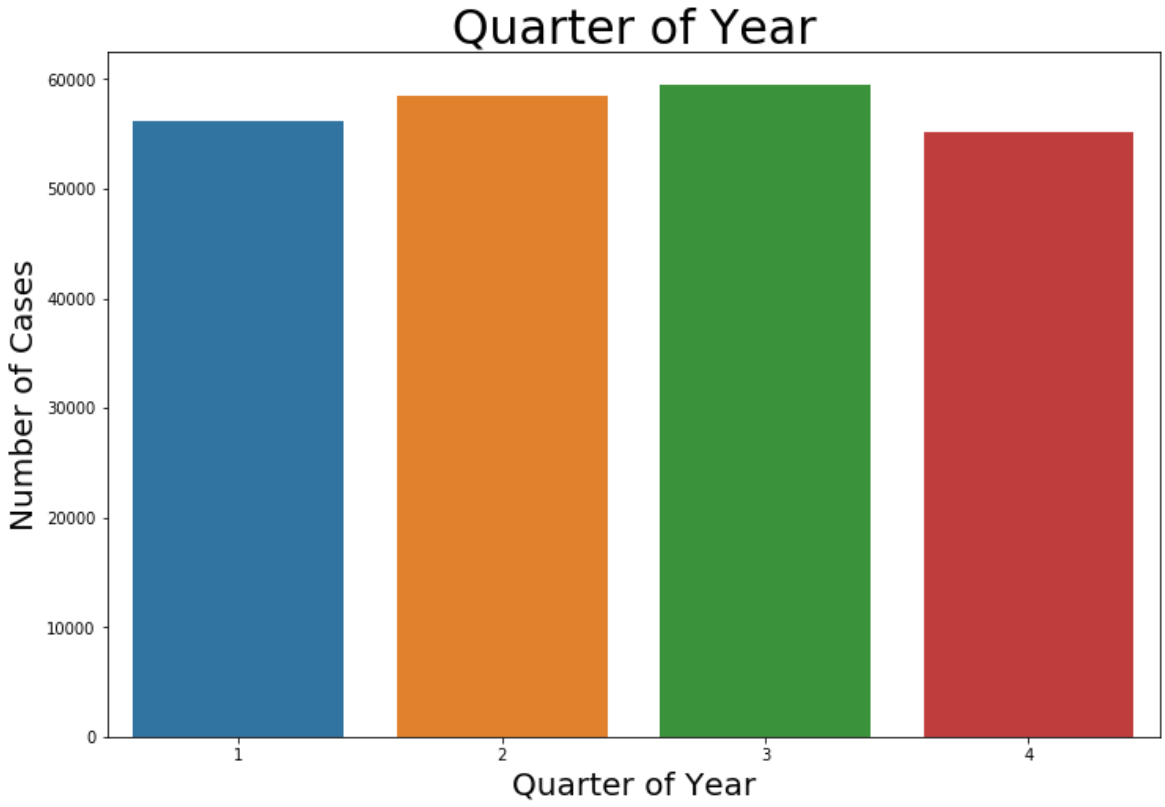
With the data wrangling done, it's time to see what story the data tells. In order to do this, all of the categorical variables were plotted. The plots are below:

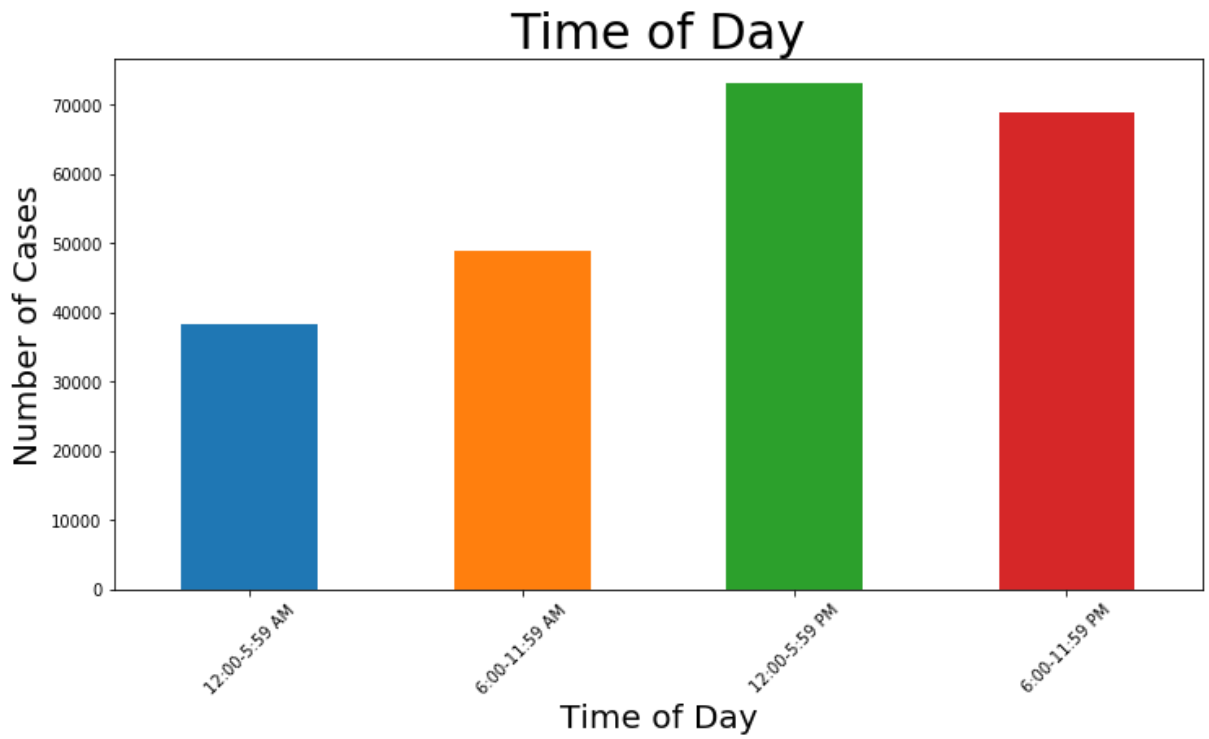



Gender
Number of Cases
120000
100000
80000
60000
40000
20000
0
Male
Female
Not Doc
Gender

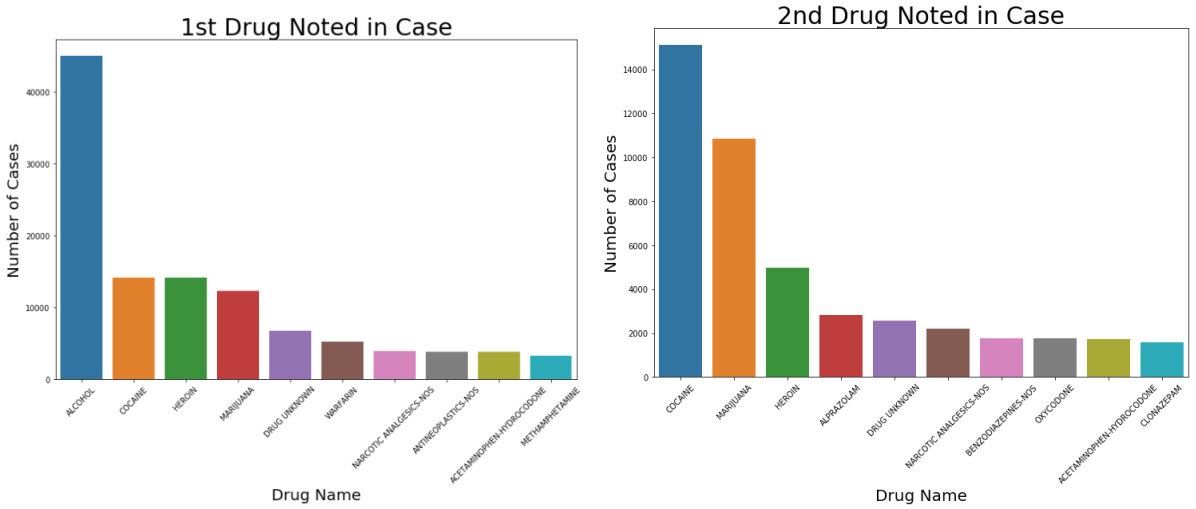

Race
Number of Cases
100000
80000
60000
40000
20000
0
Black
Hispanic
Not Doc
White
All Other
Race

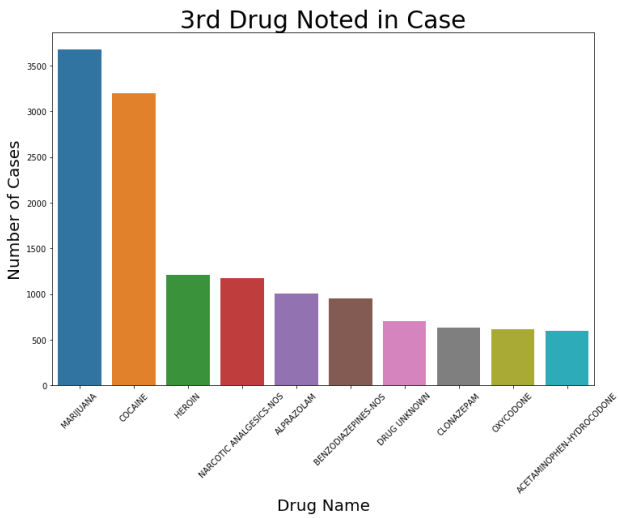

Age Category
Number of Cases
0
5000
10000
15000
20000
25000
30000
35000
40000
0-5
06-11
12-17
18-20
21-24
25-29
30-34
35-44
45-54
55-64
> 65
Not Doc
Age Category

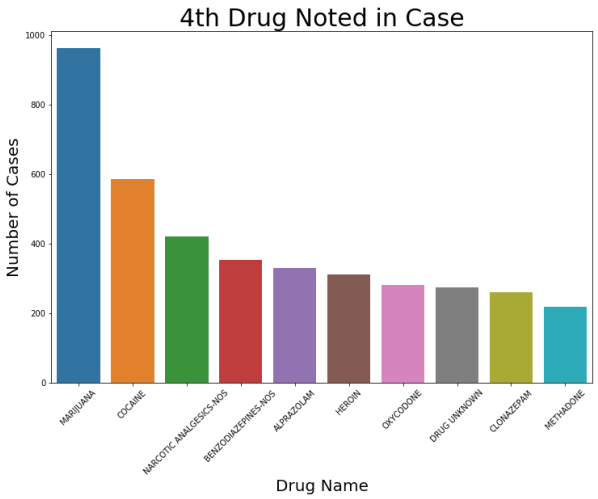

Quarter of Year
Number of Cases
0
10000
20000
30000
40000
50000
60000
1
2
3
4
Quarter of Year

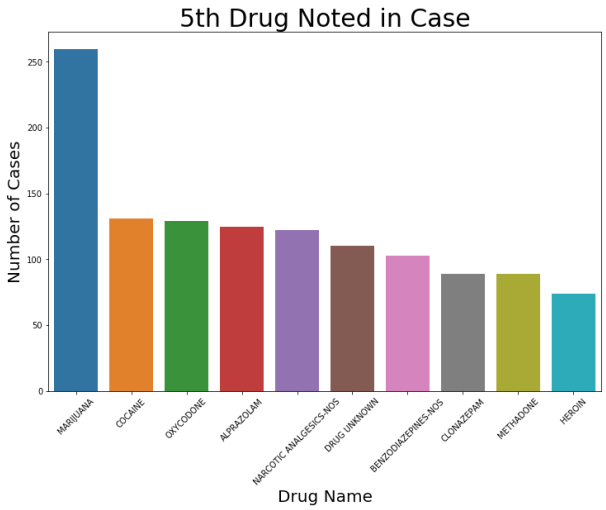


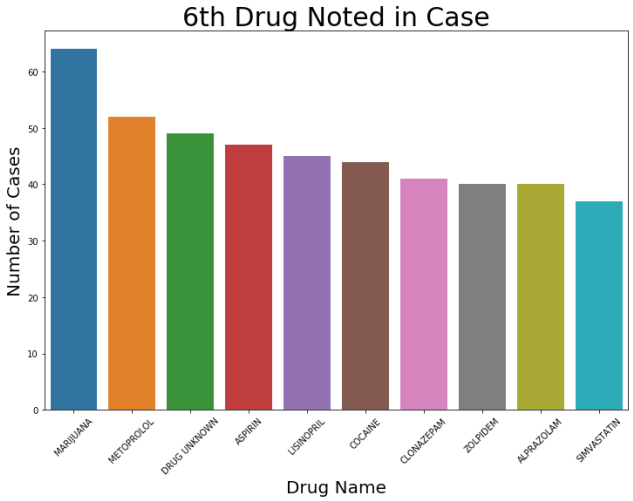
When plotting the distribution of the Drug(s) involved in the cases, it was decided to show only top 10 since there are > 2600 drugs possible. Otherwise, the graph would be too large and hard to read. Also, there were up to 22 substances involved in one case, and a plot was done for each instance.

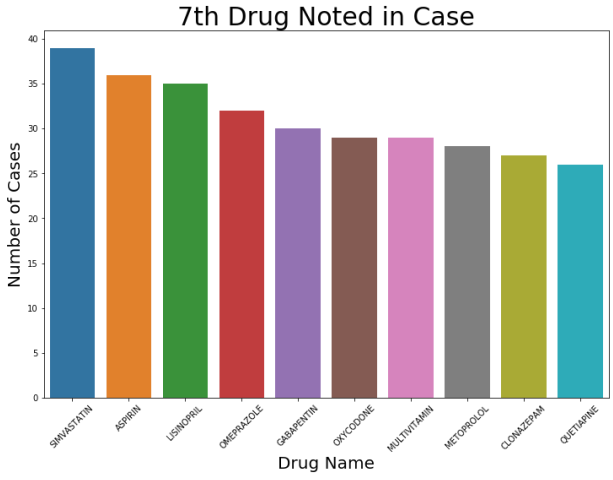


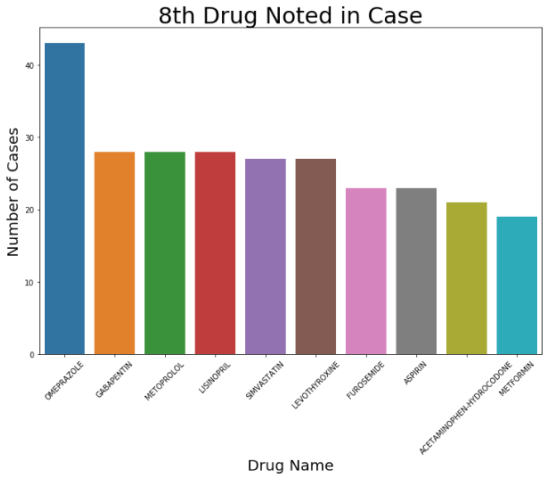

3rd Drug Noted in Case
Number of Cases
Drug Name
MARIJUANA
COCAINE
HEROIN
NARCOTIC ANALGESICS-NOS
ALPRAZOLAM
BENZODIAZEPINES-NOS
DRUG UNKNOWN
CLONAZEPAM
OXYCODONE
ACETAMINOPHEN-HYDROCODONE

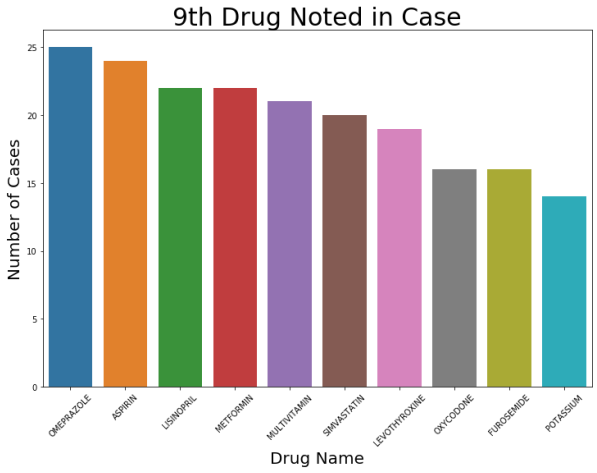

4th Drug Noted in Case
Number of Cases
Drug Name
MARIJUANA
COCAINE
NARCOTIC ANALGESICS-NOS
BENZODIAZEPINES-NOS
ALPRAZOLAM
HEROIN
OXYCODONE
DRUG UNKNOWN
CLONAZEPAM
METHADONE

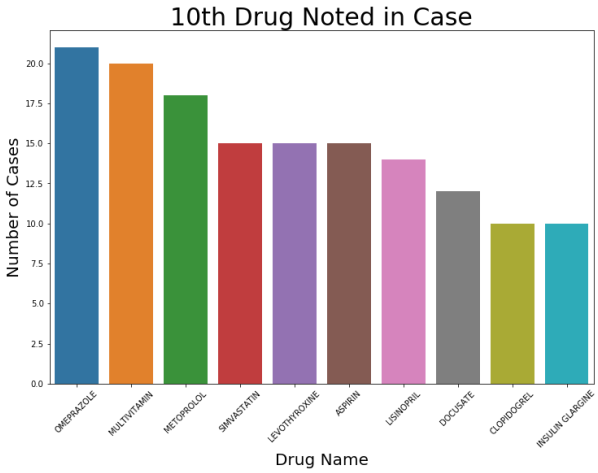

5th Drug Noted in Case
Number of Cases
Drug Name
MARIJUANA
COCAINE
OXYCODONE
ALPRAZOLAM
NARCOTIC ANALGESICS-NOS
DRUG UNKNOWN
BENZODIAZEPINES-NOS
CLONAZEPAM
METHADONE
HEROIN

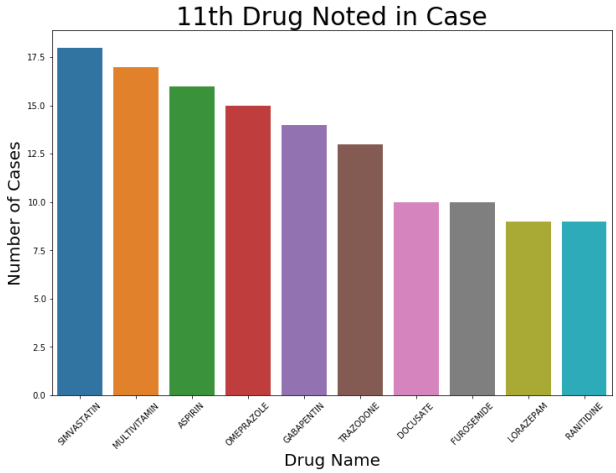

6th Drug Noted in Case
Number of Cases
Drug Name
MARIJUANA
METOPROLOL
DRUG UNKNOWN
ASPIRIN
LISINOPRIL
COCAINE
CLONAZEPAM
ZOLPIDEM
ALPRAZOLAM
SIMVASTATIN

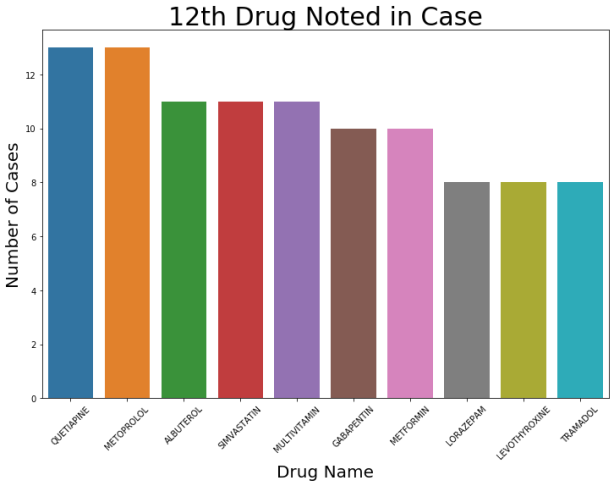

7th Drug Noted in Case
Number of Cases
Drug Name
SIMVASTATIN
ASPIRIN
LISINOPRIL
OMEPRAZOLE
GABAPENTIN
OXYCODONE
MULTIVITAMIN
METOPROLOL
CLONAZEPAM
QUETIAPINE

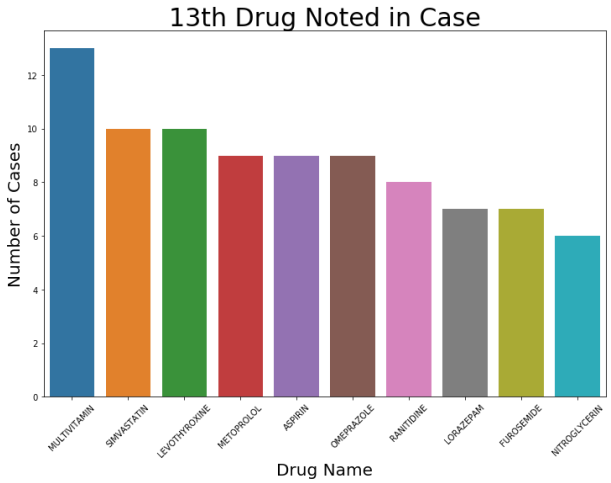

8th Drug Noted in Case
Number of Cases
Drug Name
OMEPRAZOLE
GABAPENTIN
METOPROLOL
LISINOPRIL
SIMVASTATIN
LEVOTHYROXINE
FUROSEMIDE
ASPIRIN
ACETAMINOPHEN-HYDROCODONE
METFORMIN

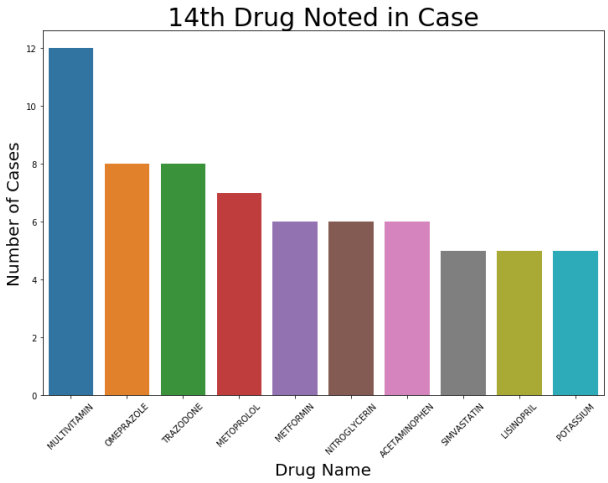

9th Drug Noted in Case
Number of Cases
Drug Name
OMEPRAZOLE
ASPIRIN
LISINOPRIL
METFORMIN
MULTIVITAMIN
SIMVASTATIN
LEVOTHYROXINE
OXYCODONE
FUROSEMIDE
POTASSIUM

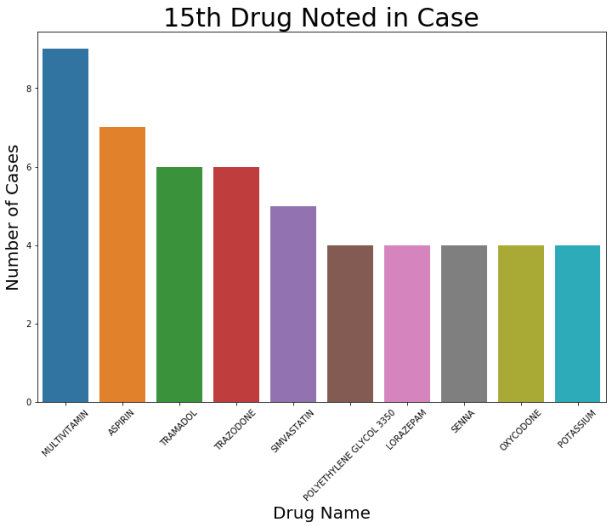

10th Drug Noted in Case
Number of Cases
Drug Name
OMEPRAZOLE
MULTIVITAMIN
METOPROLOL
SIMVASTATIN
LEVOTHYROXINE
ASPIRIN
LISINOPRIL
DOCUSATE
CLOPIDOGREL
INSULIN GLARGINE

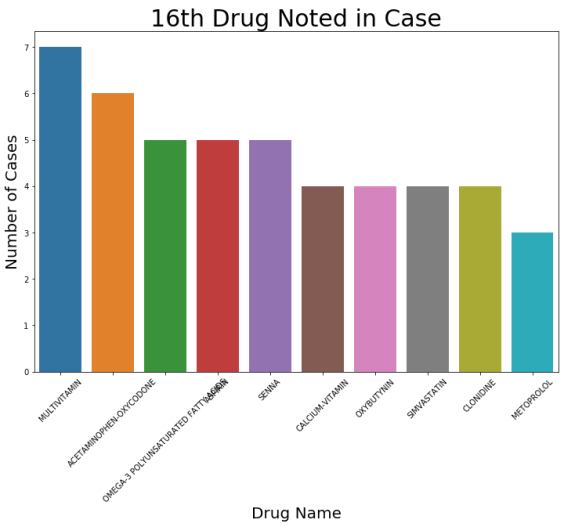

11th Drug Noted in Case
Number of Cases
Drug Name
SIMVASTATIN
MULTIVITAMIN
ASPIRIN
OMEPRAZOLE
GABAPENTIN
TRAZODONE
DOCUSATE
FUROSEMIDE
LORAZEPAM
RANITIDINE

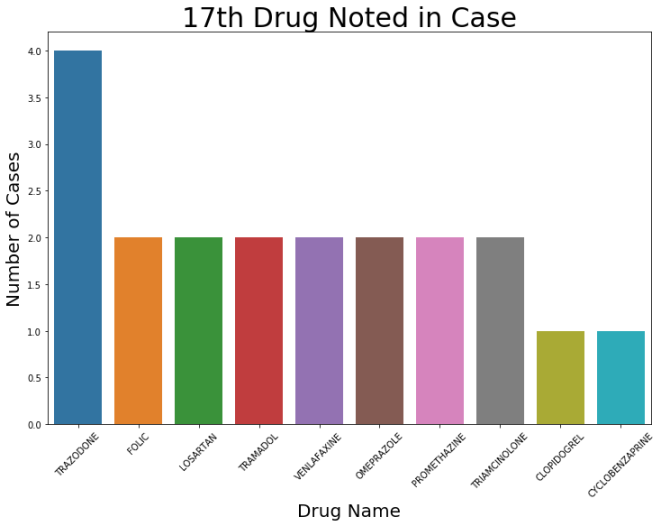

12th Drug Noted in Case
Number of Cases
Drug Name
QUETIAPINE
METOPROLOL
ALBUTEROL
SIMVASTATIN
MULTIVITAMIN
GABAPENTIN
METFORMIN
LORAZEPAM
LEVOTHYROXINE
TRAMADOL

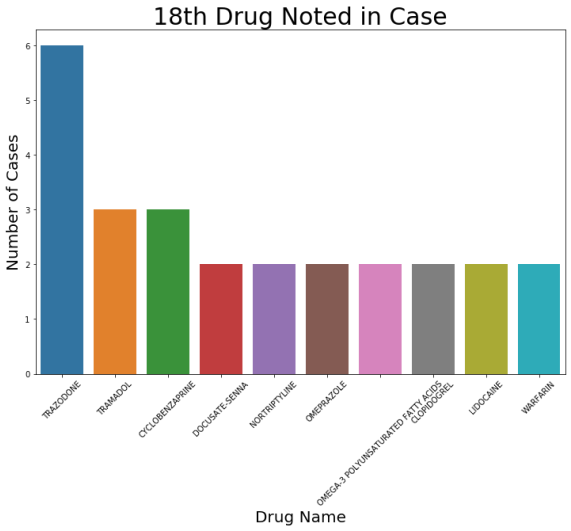

13th Drug Noted in Case
Number of Cases
Drug Name
MULTIVITAMIN
SIMVASTATIN
LEVOTHYROXINE
METOPROLOL
ASPIRIN
OMEPRAZOLE
RANITIDINE
LORAZEPAM
FUROSEMIDE
NITROGLYCERIN

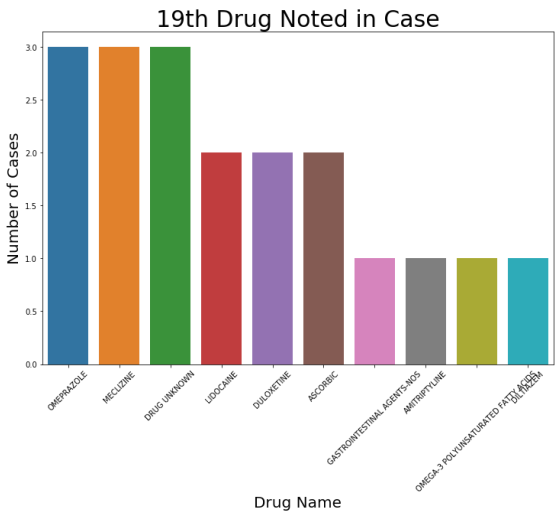

14th Drug Noted in Case
Number of Cases
Drug Name
MULTIVITAMIN
OMEPRAZOLE
TRAZODONE
METOPROLOL
METFORMIN
NITROGLYCERIN
ACETAMINOPHEN
SIMVASTATIN
LISINOPRIL
POTASSIUM

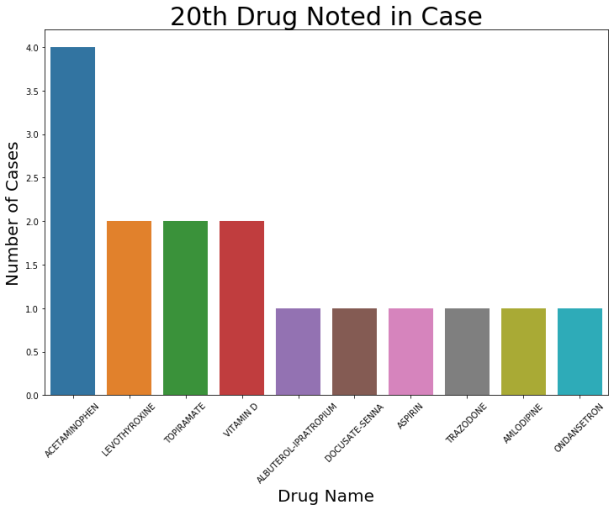

15th Drug Noted in Case
Number of Cases
Drug Name
MULTIVITAMIN
ASPIRIN
TRAMADOL
TRAZODONE
SIMVASTATIN
POLYETHYLENE GLYCOL 3350
LORAZEPAM
SENNA
OXYCODONE
POTASSIUM

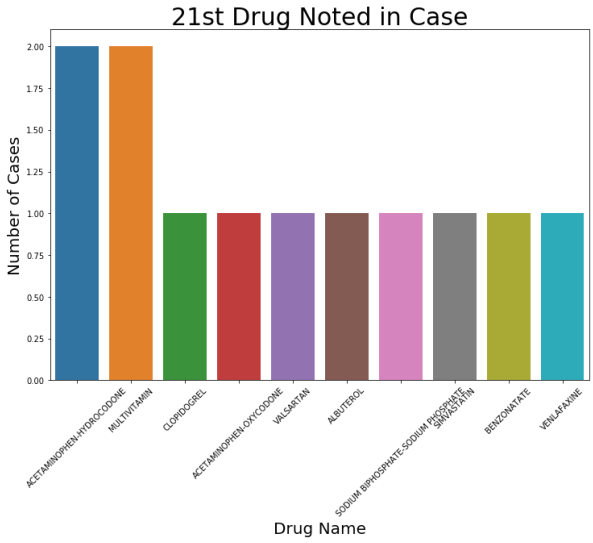

16th Drug Noted in Case
Number of Cases
Drug Name
MULTIVITAMIN
ACETAMINOPHEN-OXYCODONE
SENNA
CALCIUM-VITAMIN
OXYBUTYNIN
SIMVASTATIN
CLONIDINE
METOPROLOL

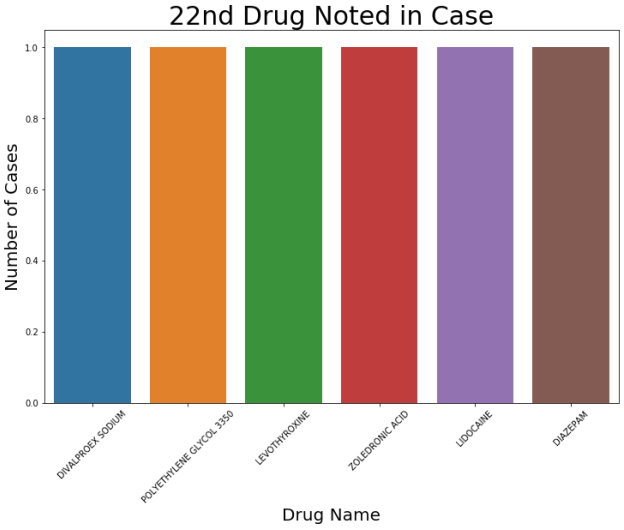

17th Drug Noted in Case
Number of Cases
Drug Name
TRAZODONE
FOLIC
LOSARTAN
TRAMADOL
VENLAFAXINE
OMEPRAZOLE
PROMETHAZINE
TRIAMCINOLONE
CLOPIDOGREL
CYCLOBENZAPRINE


18th Drug Noted in Case
Number of Cases
Drug Name
TRAZODONE
TRAMADOL
CYCLOBENZAPRINE
DOCUSATE-SENNA
NORTRIPTYLINE
OMEPRAZOLE
CLOPIDOGREL
LIDOCAINE
WARFARIN

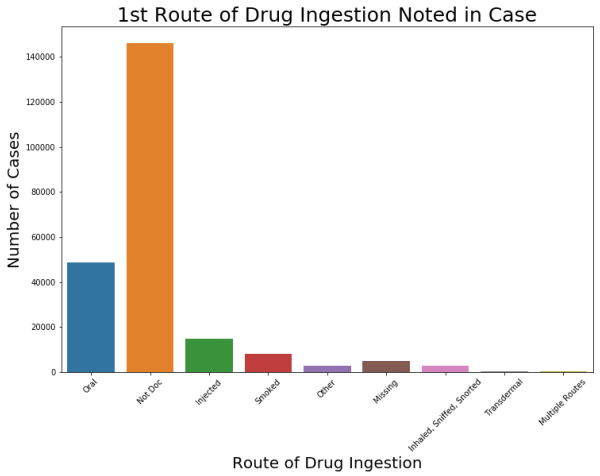

19th Drug Noted in Case
Number of Cases
Drug Name
OMEPRAZOLE
MECLIZINE
DRUG UNKNOWN
LIDOCAINE
DULOXETINE
ASCORBIC
GASTROINTESTINAL AGENTS-NOS
AMITRIPTYLINE
OMEGA-3 POLYUNSATURATED FATTY ACIDS

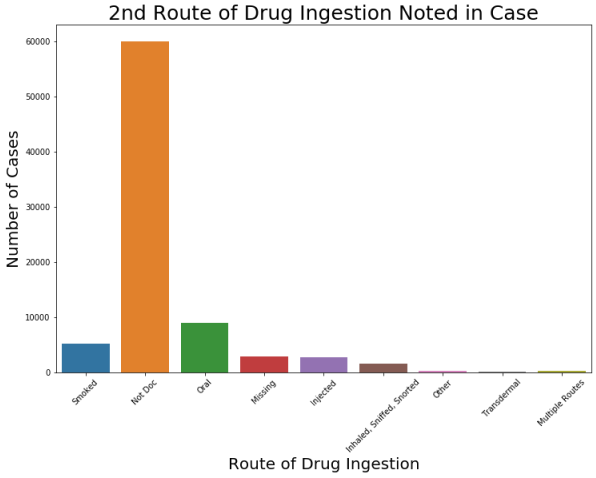

20th Drug Noted in Case
Number of Cases
Drug Name
ACETAMINOPHEN
LEVOTHYROXINE
TOPIRAMATE
VITAMIN D
ALBUTEROL-IPRATROPIUM
DOCUSATE-SENNA
ASPIRIN
TRAZODONE
AMLODIPINE
ONDANSETRON

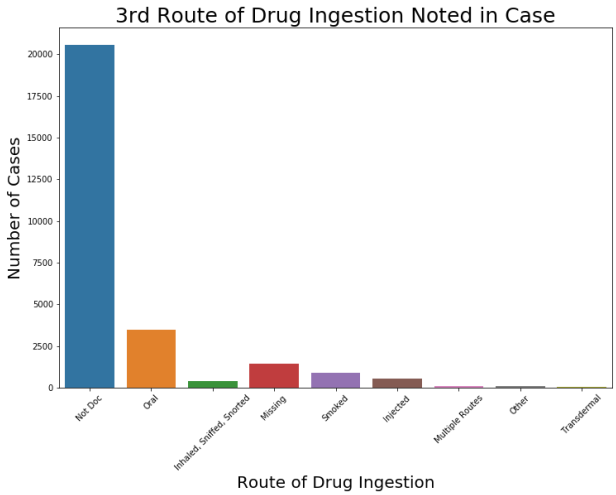

21st Drug Noted in Case
Number of Cases
Drug Name
ACETAMINOPHEN-HYDROCODONE
MULTIVITAMIN
CLOPIDOGREL
ACETAMINOPHEN-OXYCODONE
VALSARTAN
ALBUTEROL
SODIUM BIPHOSPHATE-SODIUM PHOSPHATE
SIMVASTATIN
BENZONATATE
VENLAFAXINE

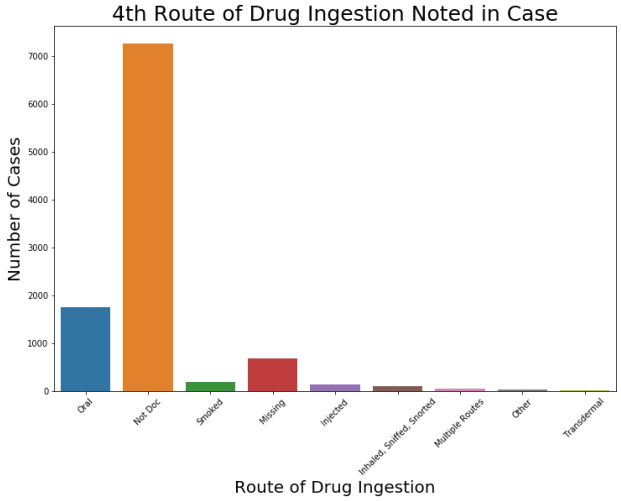

22nd Drug Noted in Case
Number of Cases
Drug Name
DIVALPROEX-SODIUM
POLYETHYLENE GLYCOL 3350
LEVOTHYROXINE
ZOLEDRONIC ACID
LIDOCAINE
DIAZEPAM

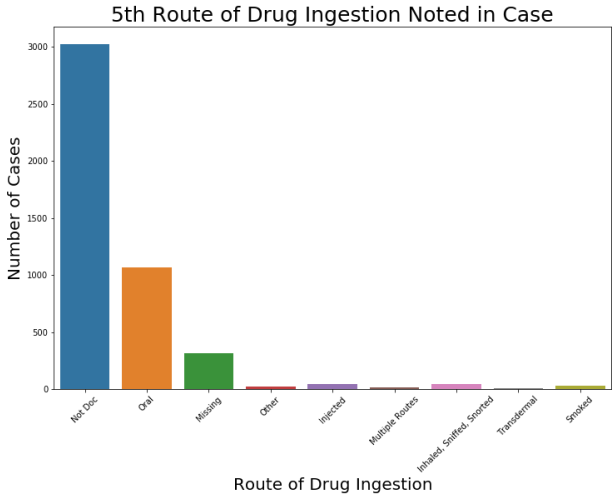
With each drug listed in a case, the route of administration was recorded. A plot of each route was done.

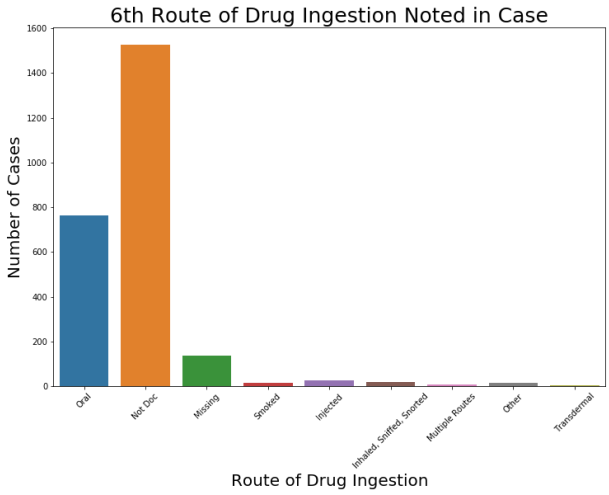


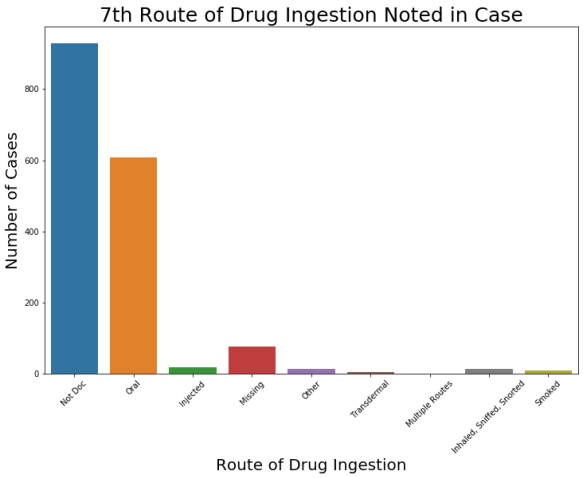


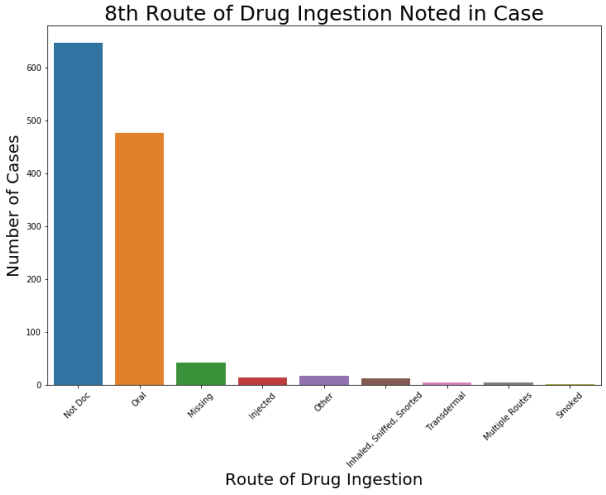


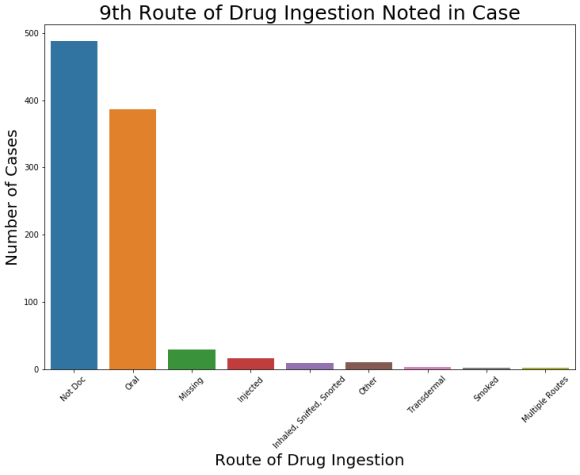


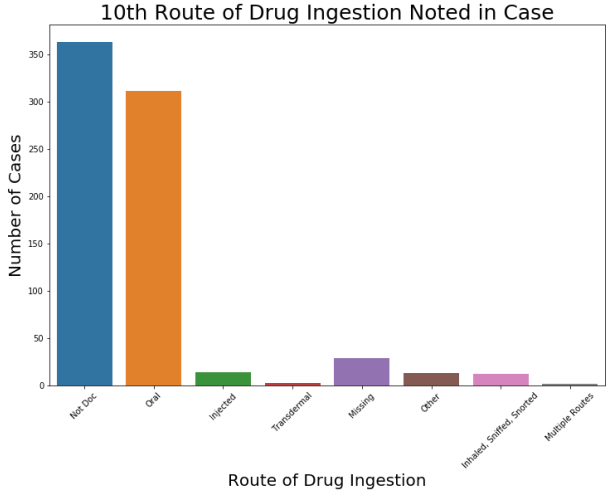


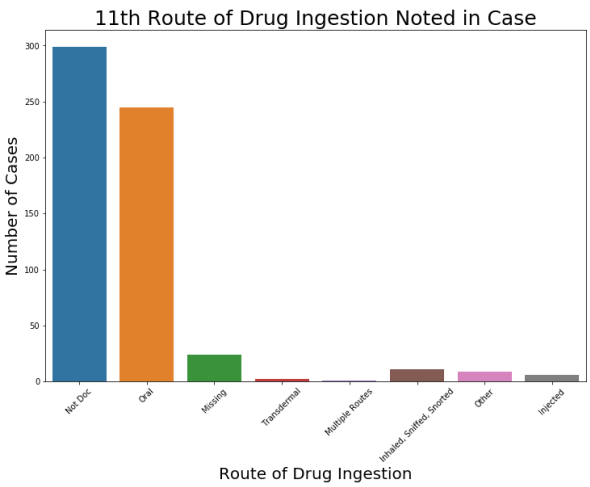


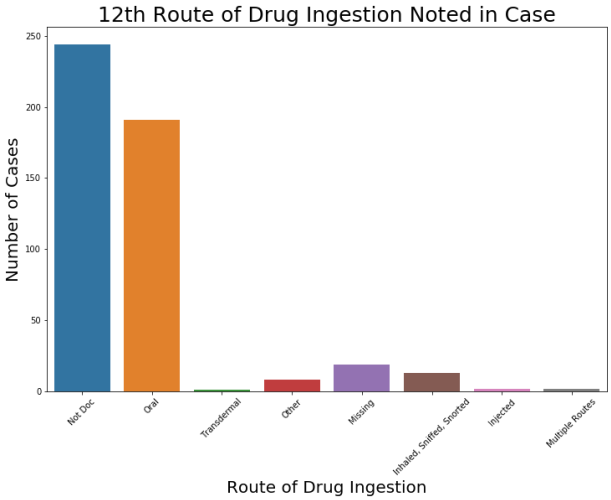

7th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Injected
Missing
Other
Transdermal
Multiple Routes
Inhaled, Sniffed, Snorted
Smoked

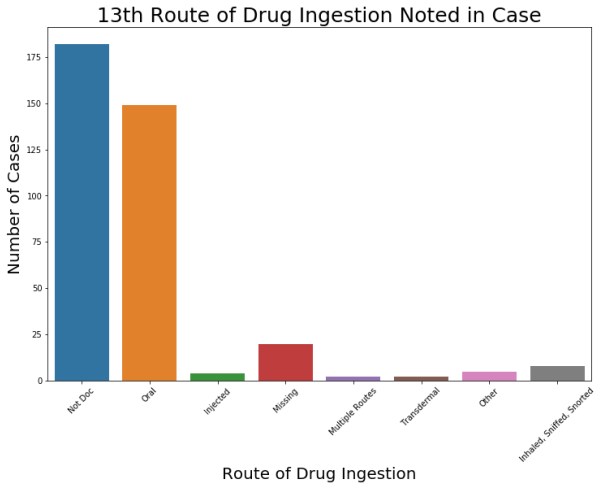

8th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Missing
Injected
Other
Inhaled, Sniffed, Snorted
Transdermal
Multiple Routes
Smoked

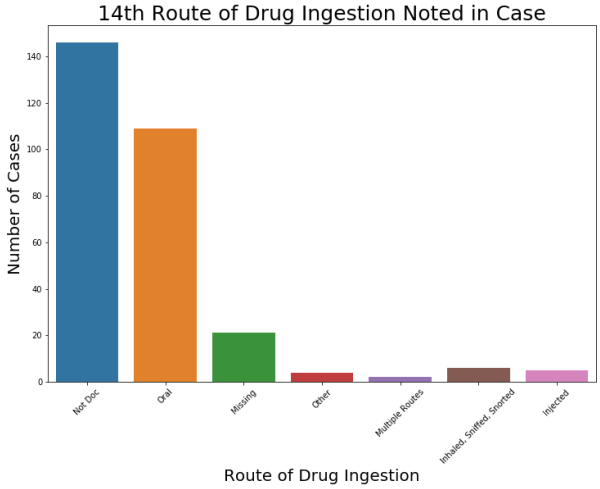

9th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Missing
Injected
Inhaled, Sniffed, Snorted
Other
Transdermal
Smoked
Multiple Routes

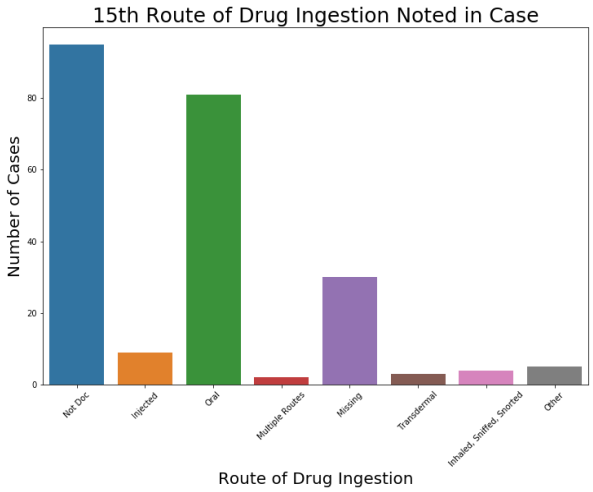

10th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Injected
Transdermal
Missing
Other
Inhaled, Sniffed, Snorted
Multiple Routes

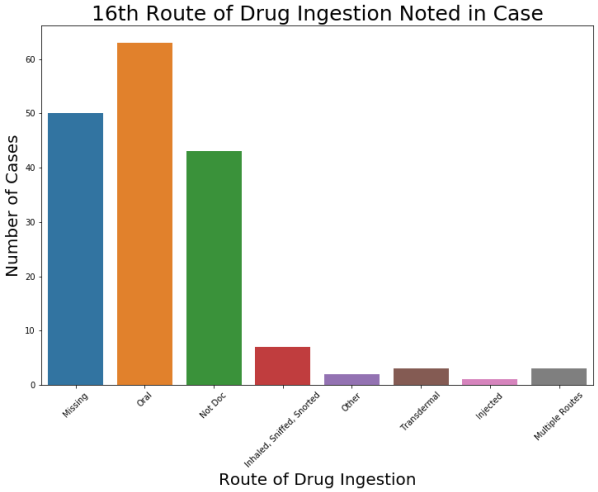

11th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Missing
Transdermal
Multiple Routes
Inhaled, Sniffed, Snorted
Other
Injected

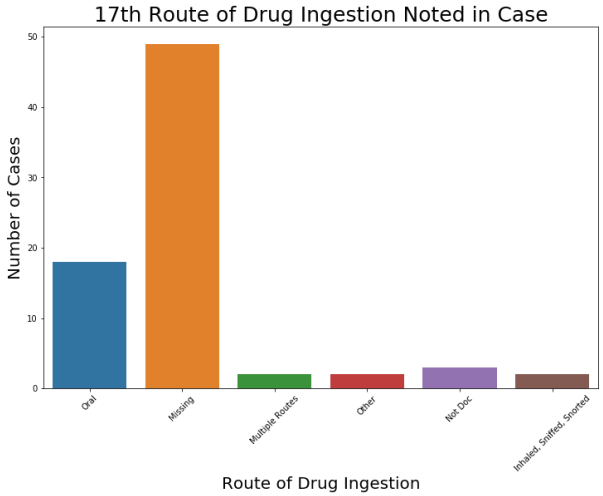

12th Route of Drug Ingestion Noted in Case
Number of Cases
Route of Drug Ingestion
Not Doc
Oral
Transdermal
Other
Missing
Inhaled, Sniffed, Snorted
Injected
Multiple Routes

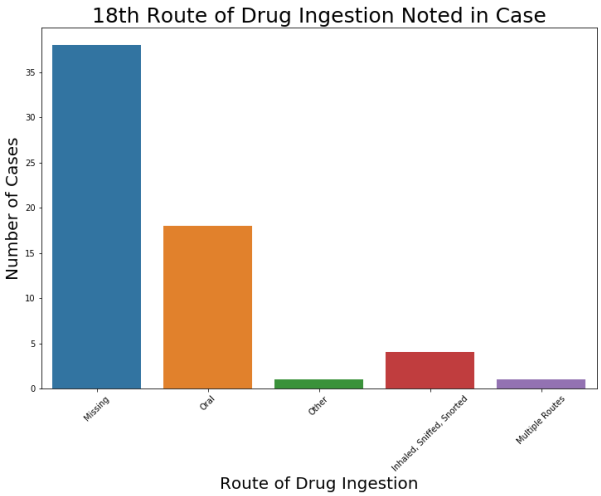

13th Route of Drug Ingestion Noted in Case
Number of Cases
Not Doc
Oral
Injected
Missing
Multiple Routes
Transdermal
Other
Inhaled, Sniffed, Snorted
Route of Drug Ingestion

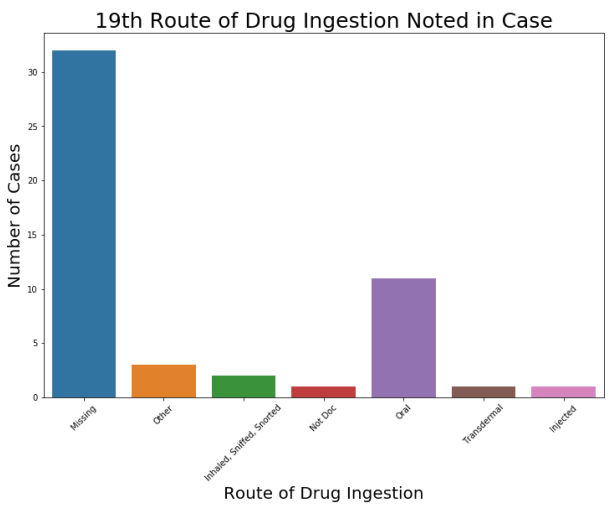

14th Route of Drug Ingestion Noted in Case
Number of Cases
Not Doc
Oral
Missing
Other
Multiple Routes
Inhaled, Sniffed, Snorted
Injected
Route of Drug Ingestion

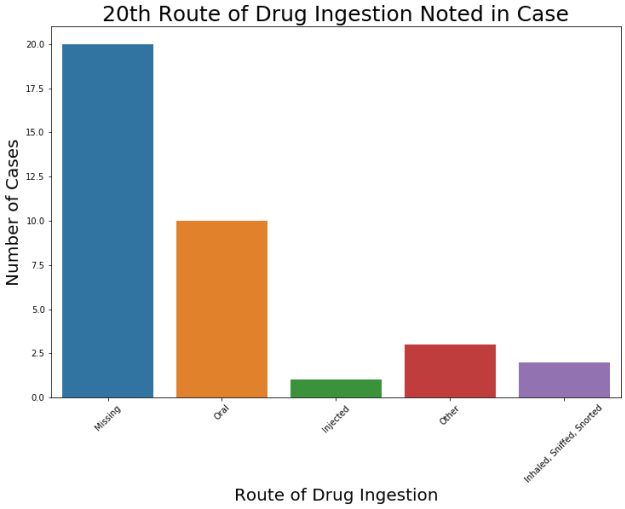

15th Route of Drug Ingestion Noted in Case
Number of Cases
Not Doc
Injected
Oral
Multiple Routes
Missing
Transdermal
Inhaled, Sniffed, Snorted
Other
Route of Drug Ingestion

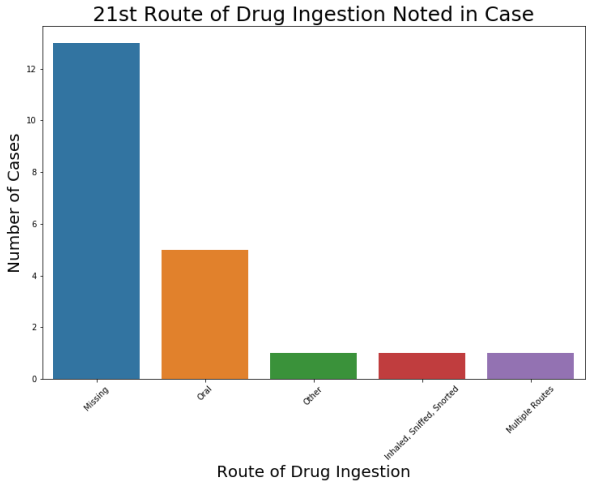

16th Route of Drug Ingestion Noted in Case
Number of Cases
Missing
Oral
Not Doc
Inhaled, Sniffed, Snorted
Other
Transdermal
Injected
Multiple Routes
Route of Drug Ingestion

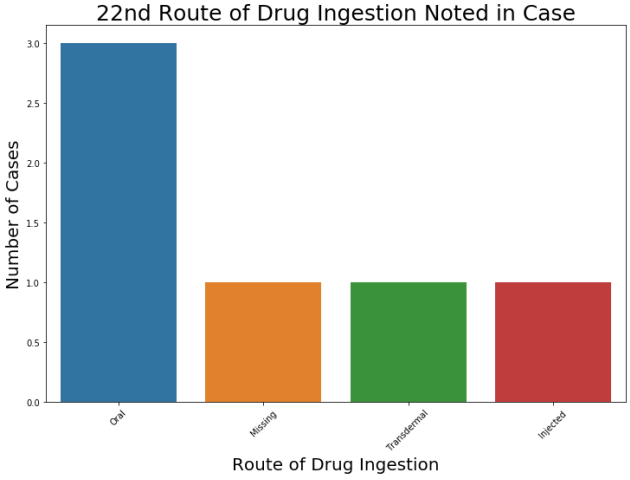

17th Route of Drug Ingestion Noted in Case
Number of Cases
Oral
Missing
Multiple Routes
Other
Not Doc
Inhaled, Sniffed, Snorted
Route of Drug Ingestion


18th Route of Drug Ingestion Noted in Case
Number of Cases
Missing
Oral
Other
Inhaled, Sniffed, Snorted
Multiple Routes
Route of Drug Ingestion

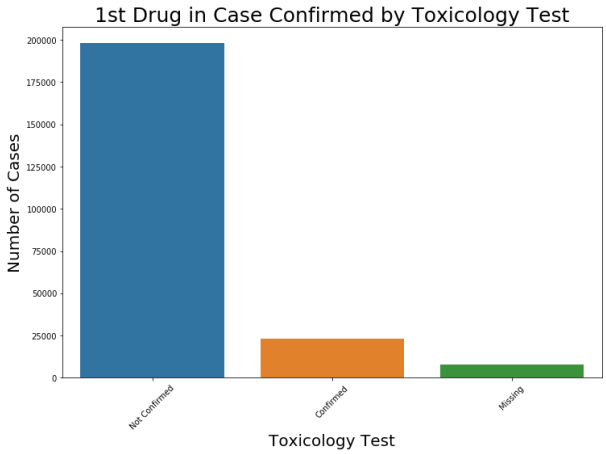

19th Route of Drug Ingestion Noted in Case
Number of Cases
Missing
Other
Inhaled, Sniffed, Snorted
Not Doc
Oral
Transdermal
Injected
Route of Drug Ingestion

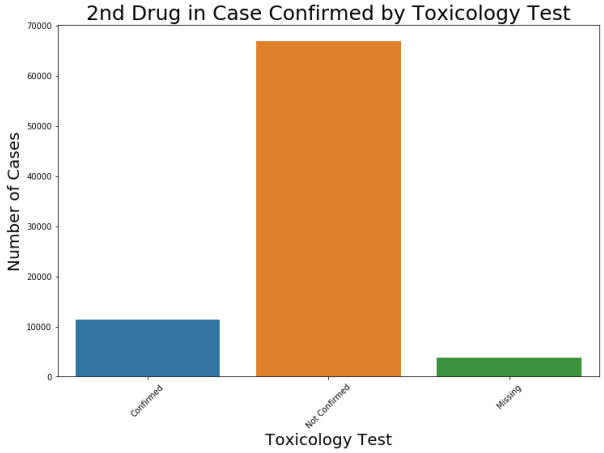

20th Route of Drug Ingestion Noted in Case
Number of Cases
Missing
Oral
Injected
Other
Inhaled, Sniffed, Snorted
Route of Drug Ingestion

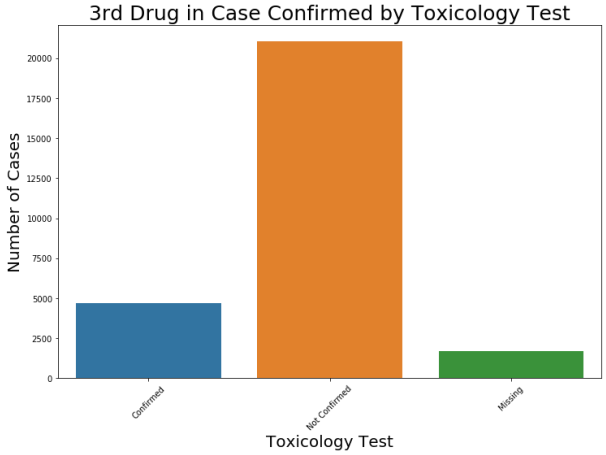

21st Route of Drug Ingestion Noted in Case
Number of Cases
Missing
Oral
Other
Inhaled, Sniffed, Snorted
Multiple Routes
Route of Drug Ingestion

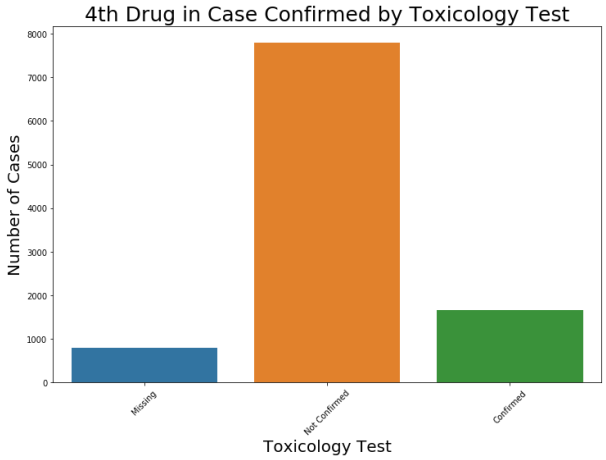

22nd Route of Drug Ingestion Noted in Case
Number of Cases
Oral
Missing
Transdermal
Injected
Route of Drug Ingestion

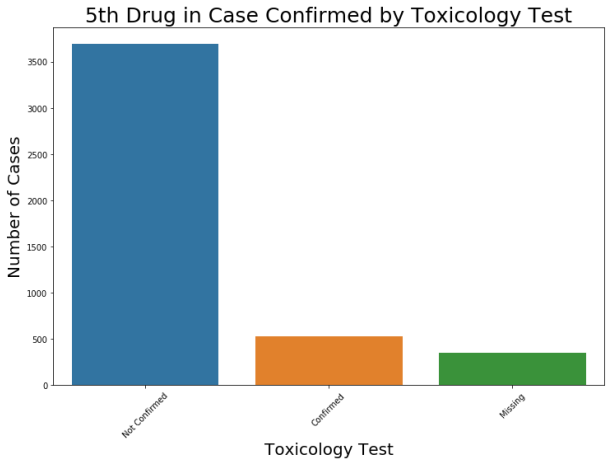
With each drug listed in a case, whether or not a toxicology test was done was recorded. Plots of the distribution of toxicology tests is below:

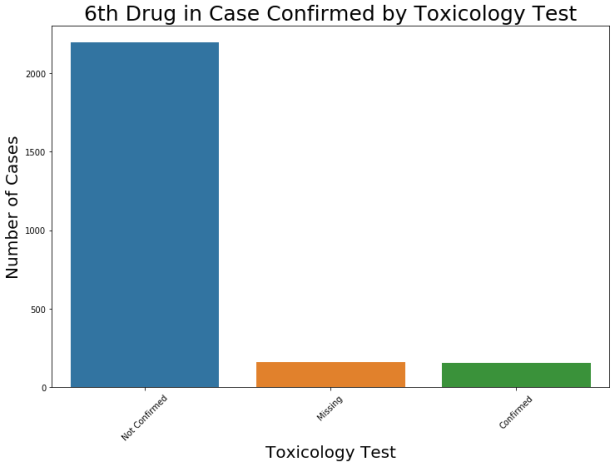


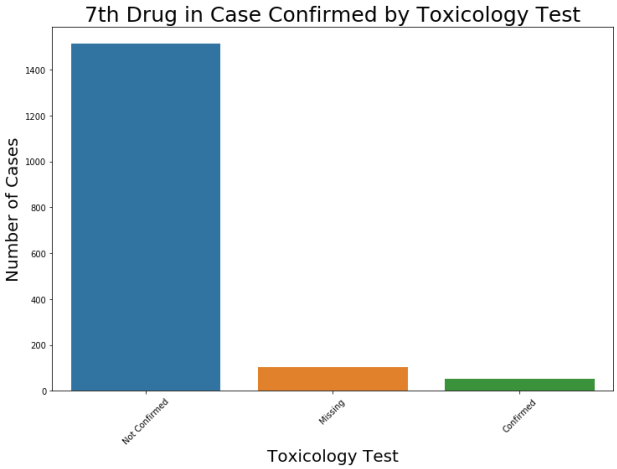


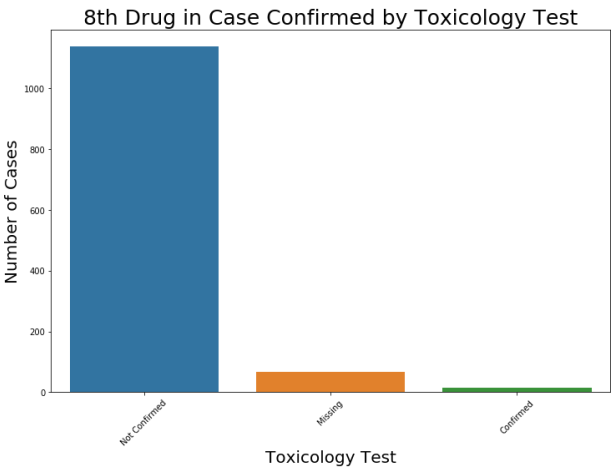


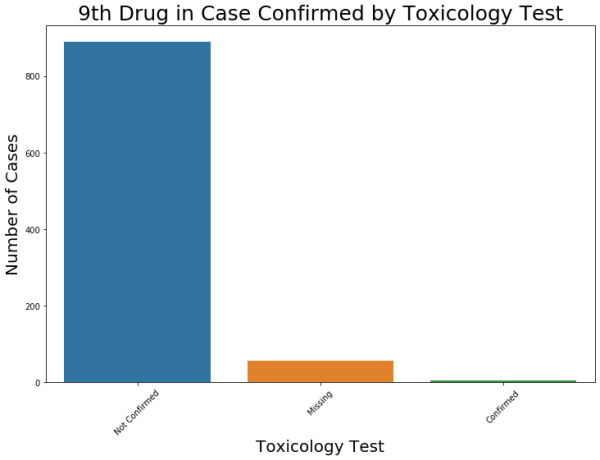


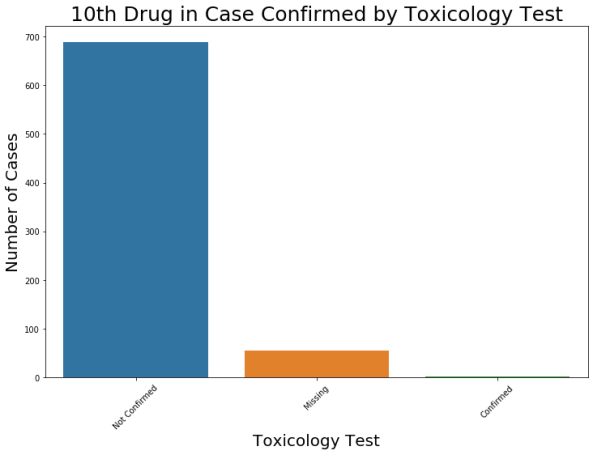


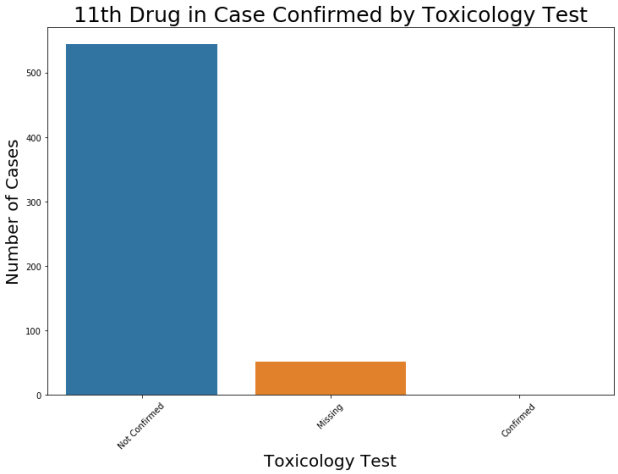


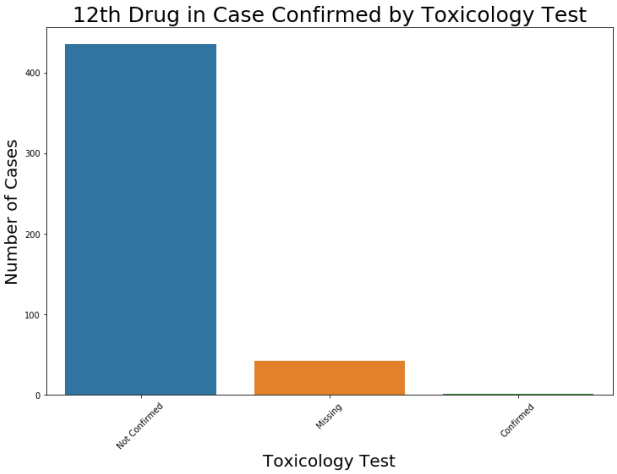

7th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test

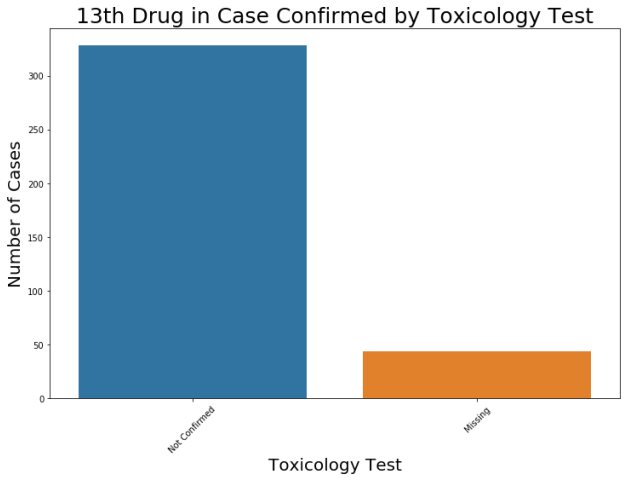

8th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test

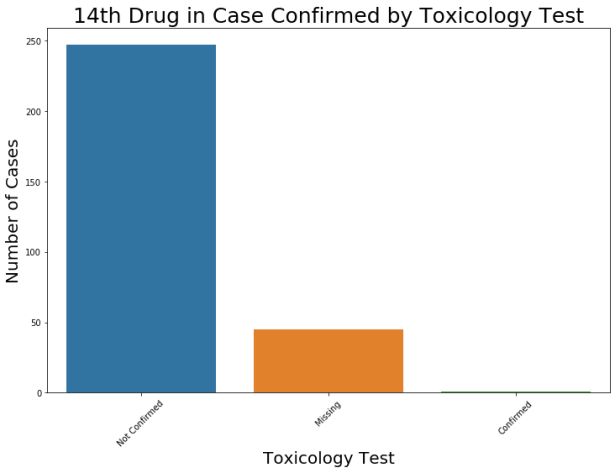

9th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test

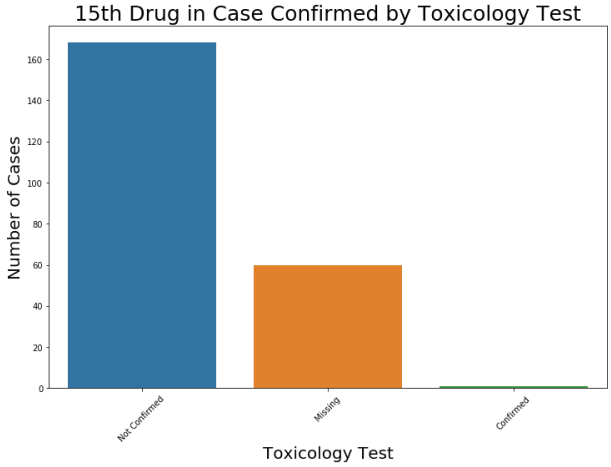

10th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test

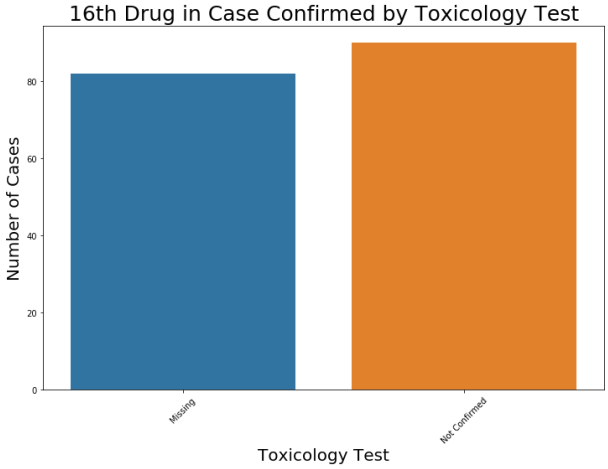

11th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test

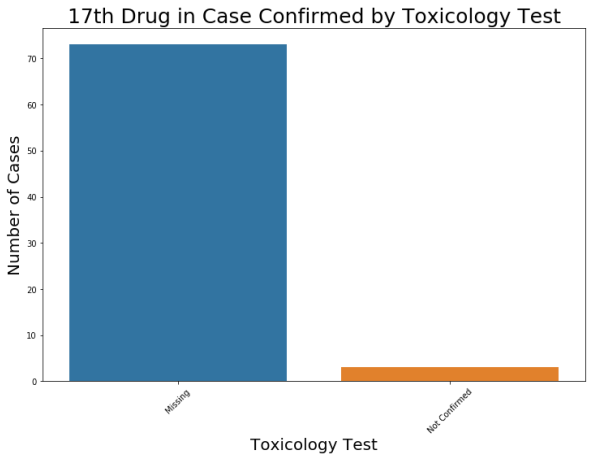

12th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test


13th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Toxicology Test

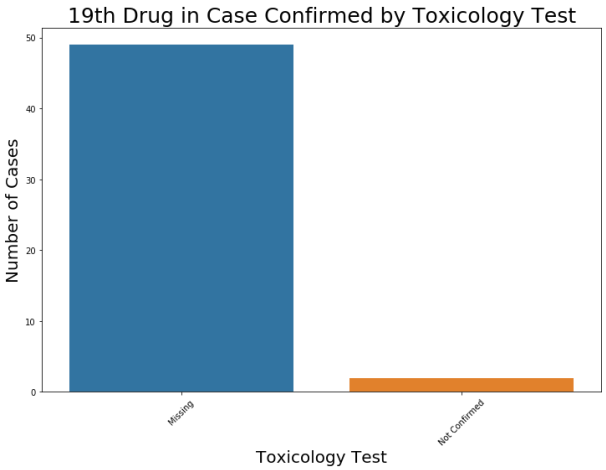

14th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test


15th Drug in Case Confirmed by Toxicology Test
Number of Cases
Not Confirmed
Missing
Confirmed
Toxicology Test


16th Drug in Case Confirmed by Toxicology Test
Number of Cases
Missing
Not Confirmed
Toxicology Test


17th Drug in Case Confirmed by Toxicology Test
Number of Cases
Missing
Not Confirmed
Toxicology Test

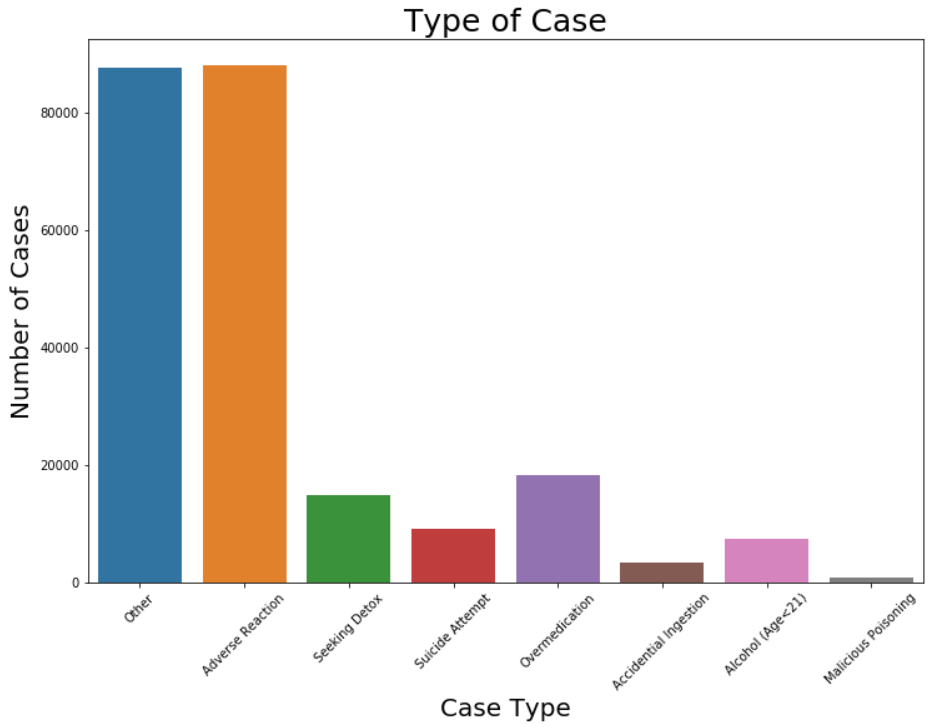

18th Drug in Case Confirmed by Toxicology Test
Number of Cases
Missing
Toxicology Test

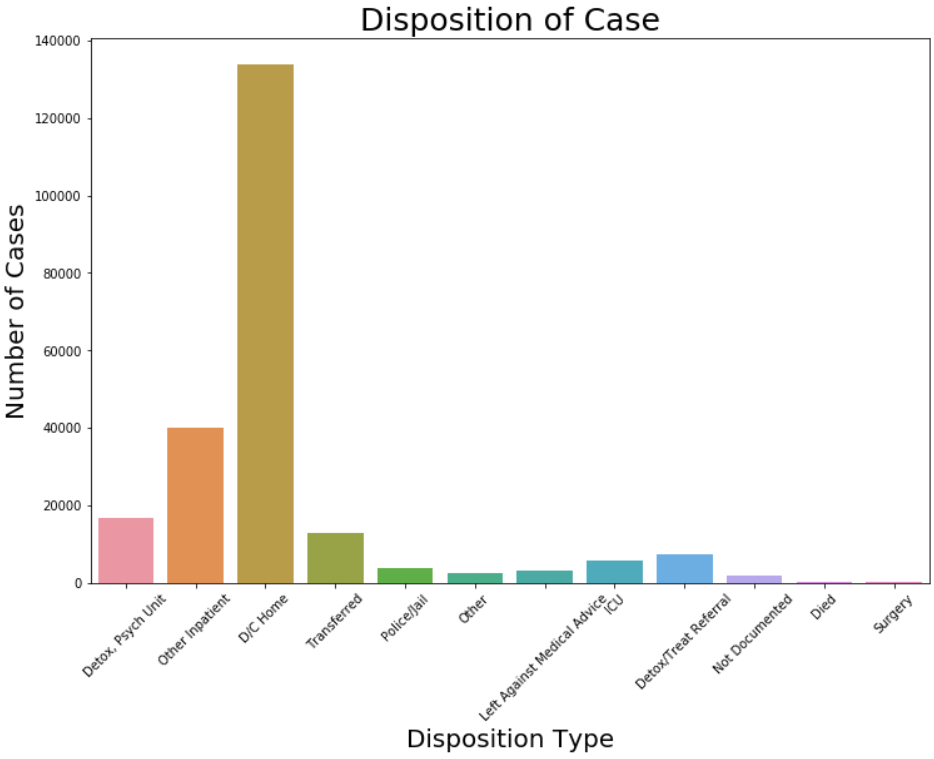

19th Drug in Case Confirmed by Toxicology Test
Number of Cases
Toxicology Test
Missing
Not Confirmed

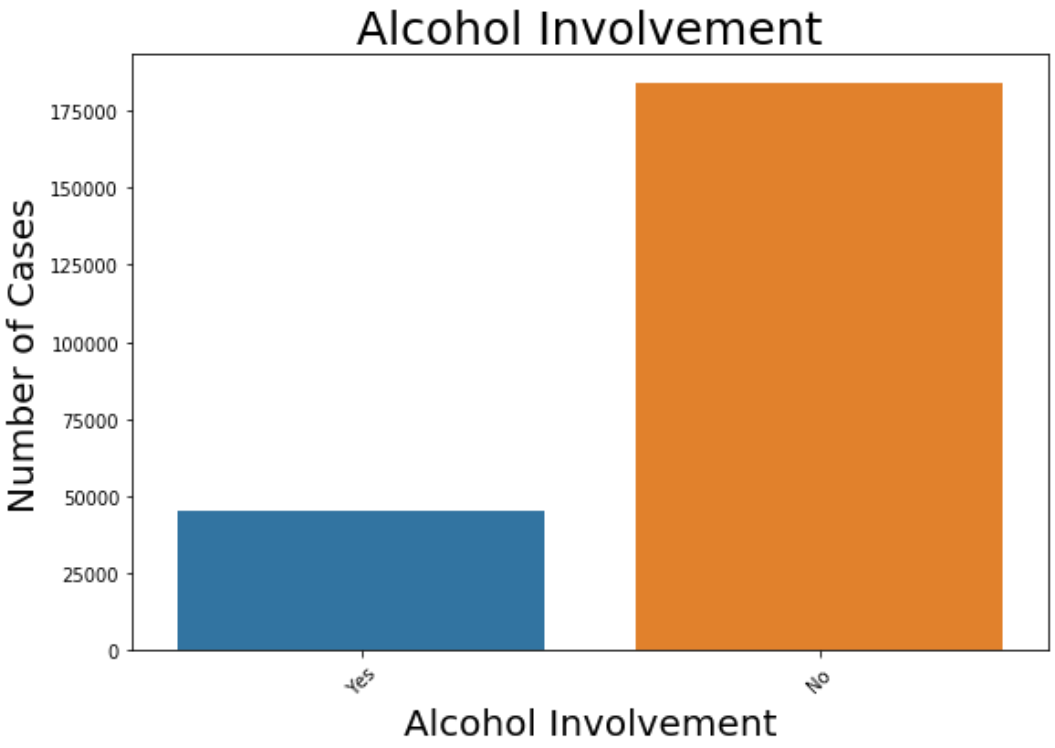

20th Drug in Case Confirmed by Toxicology Test
Number of Cases
Toxicology Test
Missing

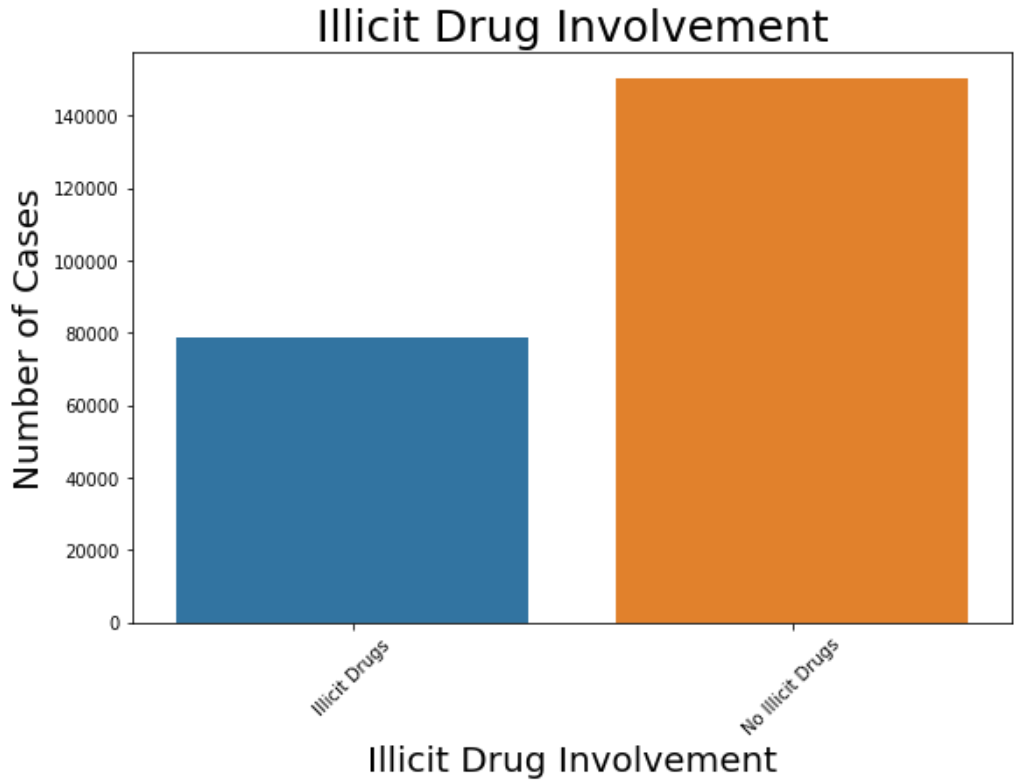

21st Drug in Case Confirmed by Toxicology Test
Number of Cases
Toxicology Test
Missing

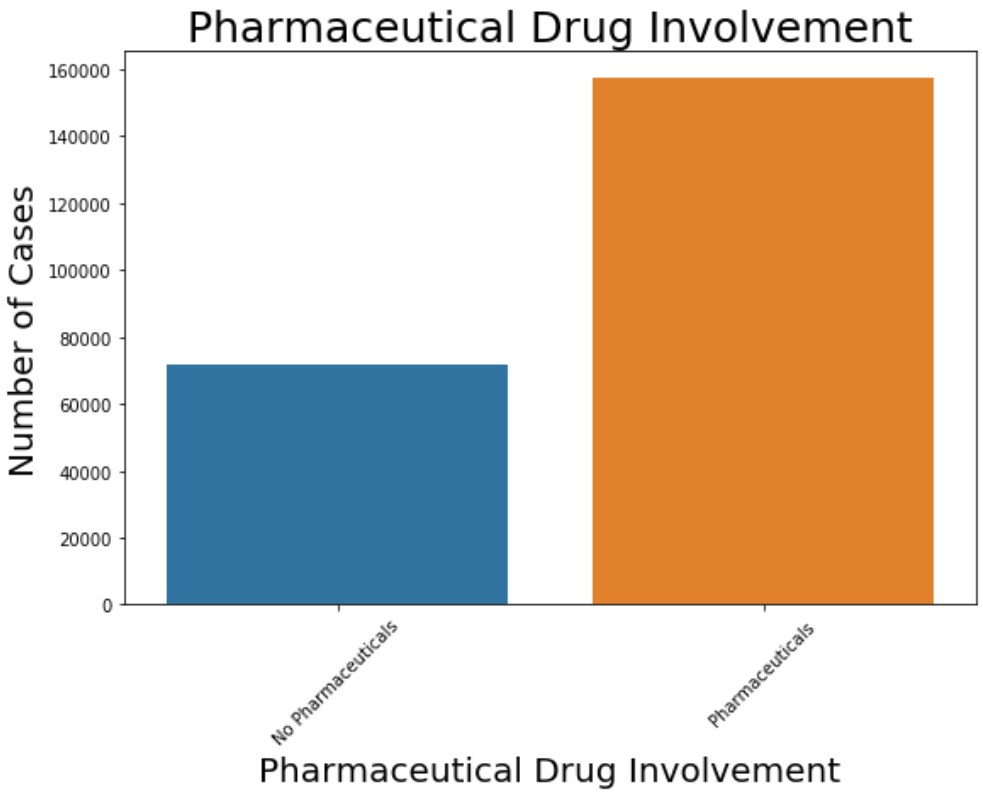

22nd Drug in Case Confirmed by Toxicology Test
Number of Cases
Toxicology Test
Missing

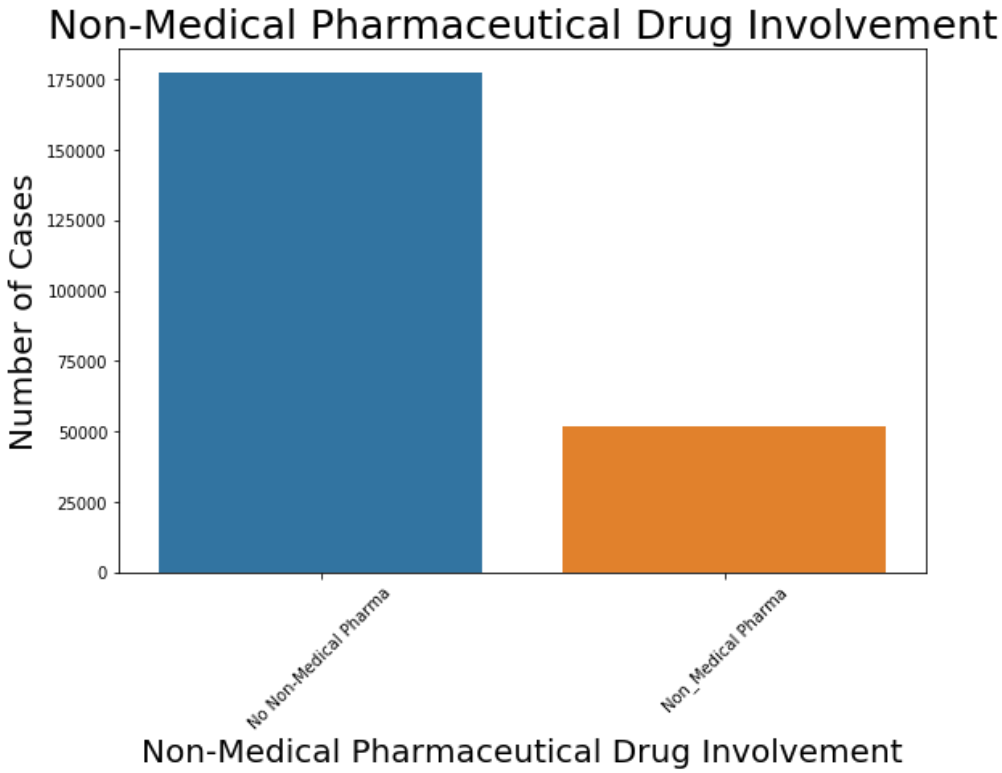

Type of Case
Number of Cases
0
20000
40000
60000
80000
Other
Adverse Reaction
Seeking Detox
Suicide Attempt
Overmedication
Accidential Ingestion
Alcohol (Age<21)
Malicious Poisoning
Case Type


Disposition of Case
Number of Cases
0
20000
40000
60000
80000
100000
120000
140000
Detox, Psych Unit
Other Inpatient
D/C Home
Transferred
Police/Jail
Other
Left Against Medical Advice
ICU
Detox/Treat Referral
Not Documented
Died
Surgery
Disposition Type


Alcohol Involvement
Number of Cases
175000
150000
125000
100000
75000
50000
25000
0
Yes
No
Alcohol Involvement


Illicit Drug Involvement
Number of Cases
140000
120000
100000
80000
60000
40000
20000
0
Illicit Drugs
No Illicit Drugs
Illicit Drug Involvement


Pharmaceutical Drug Involvement
Number of Cases
160000
140000
120000
100000
80000
60000
40000
20000
0
No Pharmaceuticals
Pharmaceuticals
Pharmaceutical Drug Involvement


Non-Medical Pharmaceutical Drug Involvement
Number of Cases
175000
150000
125000
100000
75000
50000
25000
0
No Non-Medical Pharma
Non_Medical Pharma
Non-Medical Pharmaceutical Drug Involvement

This data set has 85 variables. Eighty of them are categorical. These variables include demographic information on each case, such as metro area, age, sex and race. They also include the name of the drug(s) ingested, the route by which the drug(s) was ingested and whether a toxicology report was done. There are up to 22 drugs involved in a case. Also included in the data are the type of case, it's disposition, whether alcohol was involved, and whether the drugs were illicit, pharmaceutical or non-medical pharmaceutical.

The following observations were seen from the plots:

- The distribution of gender was fairly even.
- The predominate race involved in the cases was white.
- The majority of the cases were between the ages of 35-54.
- The time of year is evenly distributed.
- The time of day that most cases occur is between 12:00 pm and 11:59 pm.
- Alcohol, cocaine, heroin and marijuana are involved in many of the cases.
- Most of the drugs were taken orally.
- Most were not confirmed by a toxicology test.
- Most of the cases were due to adverse reactions to the drugs.
- Most cases were discharged home.
- Alcohol was not involved in many of the cases.

- Most of the drugs involved were not illicit.
- Most of the drugs involved were pharmaceutical.
- Most of the cases did not involve non-medical pharmaceutical drugs.
- Most of the cases qualified as all misuse and abuse episodes, which means that the drugs were used for purposes that is not consistent with legal or medical guidelines.

Inferential Statistics

The next step in the project was inferential statistics. Since this DataFrame has many variables, only a few were chosen for hypothesis testing. Since the variables are categorical, the Chi-Squared test for proportions and independence were used for the tests.

First, the locations of the cases were analyzed. Is there a significant difference in the frequency of cases in each of the cities?

- H_o: p1 = p2... = p15
- H_a: p1 != p2... != p15

The p-value was 0.0; therefore, H_o was rejected. It is reasonable to conclude that the distribution of the proportion of the location of the cases is not equal.

Second, the race of the cases will be analyzed. Does the race of the patient vary significantly?

- H_o: p1 = p2... = p15
- H_a: p1 != p2... != p15

The p-value was 0.0; therefore, H_o was rejected. It is reasonable to conclude that the distribution of the proportion of race is not equal.

Next, the independence of disposition vs. casetype was analyzed.

- H_o: the population frequencies are equal to the expected frequencies
- H_a: the population frequencies are not equal to the expected frequencies.

The p-value was 0.0; therefore, H_o was rejected. It is reasonable to conclude that the distribution of population frequencies are not equal to the expected frequencies.

Finally, the independence of sex vs. drugid_1 was analyzed.

- H_o: the population frequencies are equal to the expected frequencies
- H_a: the population frequencies are not equal to the expected frequencies.

The p-value was 0.0; therefore, H_o was rejected. It is reasonable to conclude that the distribution of population frequencies are not equal to the expected frequencies.

Conclusion

In summary, this milestone report shows the process for getting the data set ready for analysis, the plots done that show the story the data is telling, and the inference testing that was done. The project is now ready to move into the next phase, which focus on machine learning and regression algorithms and their evaluations.

Footnotes:

1. 'Understanding Drug Use and Addiction', *National Institute on Drug Abuse,* June, 2018, https://www.drugabuse.gov/publications/drugfacts/understanding-drug-use-addiction, (accessed December 14, 2019)
2. Yerby, Nathan, 'Statistics on Addiction in America', *Addiction Center,* December 5, 2019, https://www.addictioncenter.com/addiction/addiction-statistics/, (accessed December 14, 2019)
3. Albert, Michael, M.D., M.P.H.; McCaig, Linda F, M.P.H.; Uddin, Sayeedha, M.D., M.P.H., 'Emergency Department Visits for Drug Poisoning: United States, 2008-2011', *Centers for Disease Control and Prevention,* April, 2015, https://www.cdc.gov/nchs/products/databriefs/db196.htm, (accessed December 14, 2019)
4. 'Drug Abuse Warning Network 2011 (DAWN-2011-DS0001)', *Substance Abuse & Mental Health Services Administration,* 2011, https://www.datafiles.samhsa.gov/study-dataset/drug-abuse-warning-network-2011-dawn-2011-ds0001-nid13747, (accessed December 14, 2019)