Final Project: Analyzing Drug Overdose Deaths in Connecticut

Introduction

Over the past two decades, deaths from the abuse of opioids and other drugs have dramatically increased in the United States. Opiate deaths have risen from about 3 per 100,000 people in 1999 to more than 15 per 100,000 today.[1] Some scholars have attributed this rise to being part of a broader pattern of increased "deaths of despair", a category that also includes suicides and alcoholic liver disease, and link these deaths to hopelessness and malaise wrought by economic stagnation and inequality. Others have focused on the long-term ramifications of aggressive marketing tactics used by pharmaceutical companies to promote new prescription painkillers, arguing that these companies unethically dismissed the likelihood of addictive behavior in those given opioid-derived painkillers.

While a full determination of the causes of the crisis is beyond what can be gleaned from any single dataset, there are nonetheless patterns that can be discerned. We analyzed a dataset of all 5106 reported drug overdose deaths from Connecticut between 2012 and 2018. This dataset gives basic demographic information (age, race, sex) on all overdose decedents, and provides toxicology data on the drugs found in each person's body, as well as the location of death.

Business Questions

Since this data set only lists deaths, and does not provide overall demographic information or statistics on the relative frequency of the use of different drugs in different populations, we cannot directly answer questions about the factors that influence the probability of death (i.e., we can't meaningfully calculate P(death|X) for any X, as we lack any information about non-deaths). We can,

[1] See, e.g.,National Institute on Drug Abuse, "Overdose Death Rates", online at. https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates#:~:text=The%20figure%20above%20is%20a,in%202018%20to%2067%2C367%20deaths.

however, ask questions about the overall number of deaths, as well as the particular types of drugs responsible for the deaths.

Accordingly, our central business question is the extent to which demographic factors (and the year of death) can be used to predict the manner of death. There are three ways of classifying drugs that are of interest: opioids (such as heroin and oxycodone) versus non-opioids (such as cocaine and ethanol), prescription drugs (such as tramadol and fentanyl) versus non-prescription drugs (such as heroin and cocaine), and newer synthetic opioids (fentanyl and its analogues) versus all other drugs. Understanding these relationships can help us understand which demographics are particularly affected by particular forms of drug abuse. Looking at the temporal component also helps us understand the degree to which the drug epidemic has changed over time.

Data Acquisition, Cleansing, Transformation, Munging

The data set is downloadable in CSV and Excel form through the Connecticut government's open data initiative, and is found at https://data.ct.gov/Health-and-Human-Services/Accidental-Drug-Related-Deaths-2012-2018/rybz-nyjw . We used every row in the data set, as it is not nearly large enough to be difficult to load in memory. For the purpose of predictive analytics, we randomly chose 2/3rd of the data as a training set and 1/3rd as a test set.

The data is of medium quality. There are some telltale signs of hand-entered data with insufficient consistency checks – for example, one (but only one) row had "Chinese" instead of "Asian" in the Race column, while "morphine" was spelled incorrectly in one entry in the "OtherDrug" column. Basic demographic data was present in more than 99% of cases, so we felt justified in ignoring those rows where it was not.

We decided to focus on the year, on basic demographic information, as well as on the Boolean variables on whether particular drugs were found in a given body. We also looked at geocoded death

data for the sole purpose of preparing a map of overdoses. We discarded other geographic information, and also discarded several columns of presumably coroner-supplied information on cause of death, as these were not presented in a standardized machine-readable format.

Munging involved processing columns into a standardized format – for example, race and sex were converted from characters into factors, while columns on the presence or absence of specific drugs were coded as Booleans.

Data Dictionary

The raw data file had the following fields, with cleaning operations noted:

- ID – Ignored.
- Date – we only looked at the year.
- DateType – whether the date was the date of death, autopsy, or something else. Since these would not be more than a few days apart, we ignored this.
- Age – self-explanatory.
- Sex – self-explanatory. This was parsed as a string but converted into a factor.
- Race – we simplified into 4 factors: white/black/Hispanic/Asian and other
- Residence – several columns on location. Ignored.
- Death County and location – ignored.
- Injury county and location – ignored.
- COD.. Cause of death per coroner's report. Not in a standardized format, so ignored.
- 17 columns with Boolean values on whether specific drugs were found. We combined them (via logical OR-ing of the vectors) as follows:

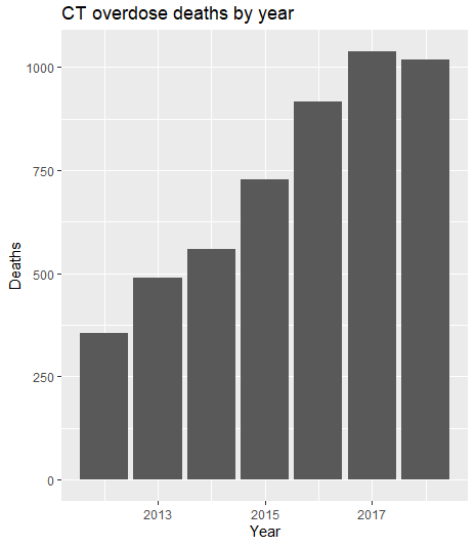
- IsOpiate – were any of Heroin, Fentayl, Fentanyl Analogue, Oxycodine, Oxymoprhone, Hydrocodone, Methadone, Tramad(ol), “Morphine_NotHeroin” OpiateNOS or “AnyOpioid” true?
- IsPrescription – is the drug commonly prescribed in the United States? True if any of Fentayl, Oxycodone, Oxymorphone, Hydrocodone, Benzodiazepine, Methdone or Tramadol were found.
- IsSynthetic – Were any newer synthetic opioids – often attributed for the spike in cases in recent years found? Combination of Fentanyln.

- Death GeoCode – the latitude and longitude of the death location, useful for mapping.

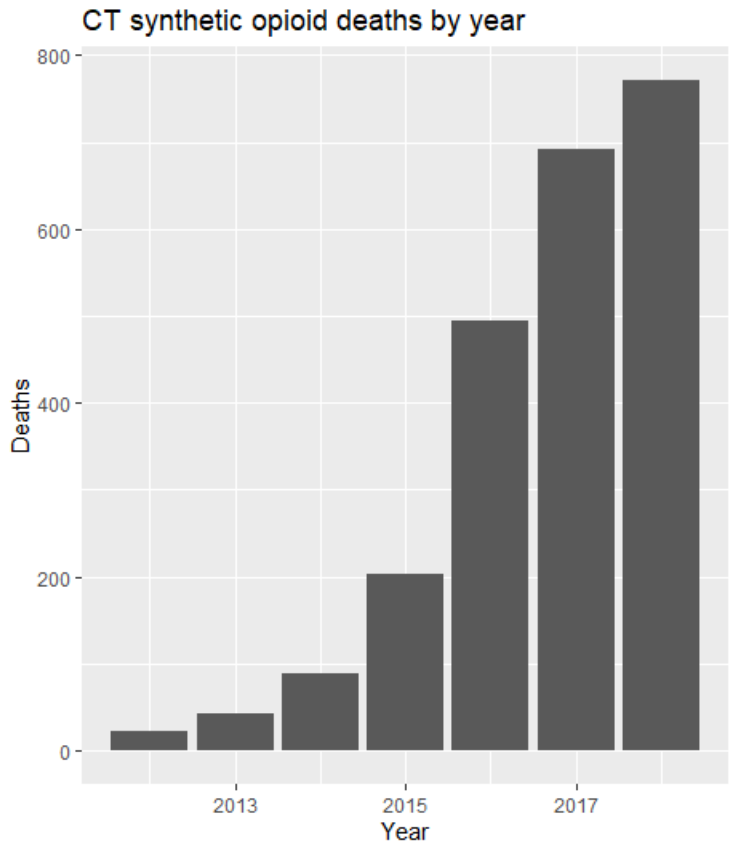
Descriptive Statistics

Some 89.9% of decedents had taken at least one opioid, 70.3% had taken a prescription drug, and 45.3% had taken a synthetic opioid.

One obvious trend was the year-on-year increase in total deaths through 2017:



Considering only deaths from synthetic opioids like fentanyl, the difference is even starker:



Predictive Analytics

We used two techniques to answer our central business question of the ability of demographics and year to predict which drugs were found. Since our business questions were all classification tasks, we used a logit regression (rather than a linear regression).

Looking at the proportion of deaths that were due to opioids, results were as follows:

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.726336   0.501059   3.445  0.00057 ***
RaceBlack    0.175882   0.466453   0.377  0.70613
RaceHispanic 1.300413   0.478190   2.719  0.00654 **
RaceWhite    1.416065   0.449855   3.148  0.00164 **
Age         -0.041266   0.005002  -8.250  < 2e-16 ***
SexMale      0.258647   0.128650   2.010  0.04438 *
YearCode     0.254366   0.030763   8.269  < 2e-16 ***
```

White and Hispanic race, younger age and male gender were all associated with a greater chance of deaths being due to opioids. Running this model on the test data set obtained 90.6% accuracy, which is relatively low since naively predicting all deaths to have been due to opiates is correct 89% of the time.

Looking at prescription drug deaths, the situation is more muddled:

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.585640   0.450588  -1.300 0.193695
RaceBlack    -0.631960   0.440400  -1.435 0.151297
RaceHispanic -0.177961   0.436246  -0.408 0.683320
RaceWhite     0.161004   0.422425   0.381 0.703098
Age           0.001465   0.003421   0.428 0.668452
SexMale      -0.337790   0.097499  -3.465 0.000531 ***
YearCode      0.480672   0.023122  20.788  < 2e-16 ***
```

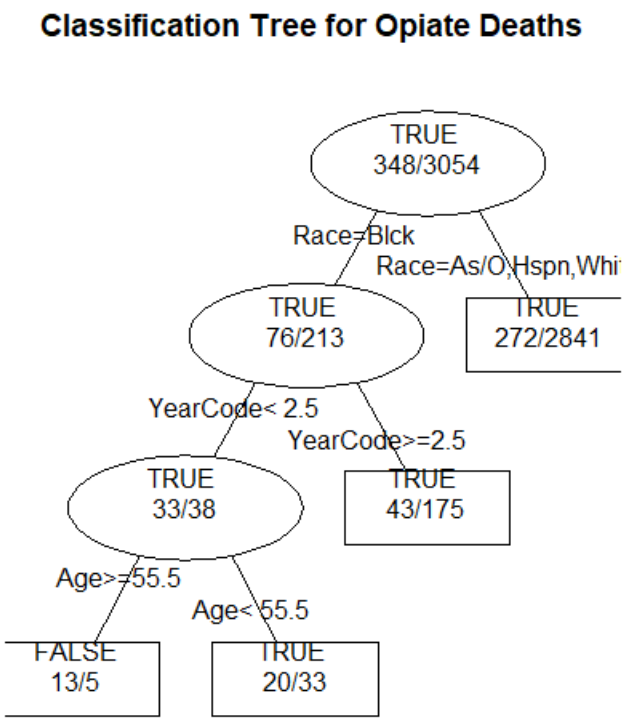
Women were more likely to use prescription drugs, and prescription drug abuse also increased over time. No other demographics had significant predictive effect. Overall, the predictive model was somewhat more effective (73% accuracy, against 70.3% naively).

Finally, deaths due to fentanyl and other synthetic opioids had the following regression:

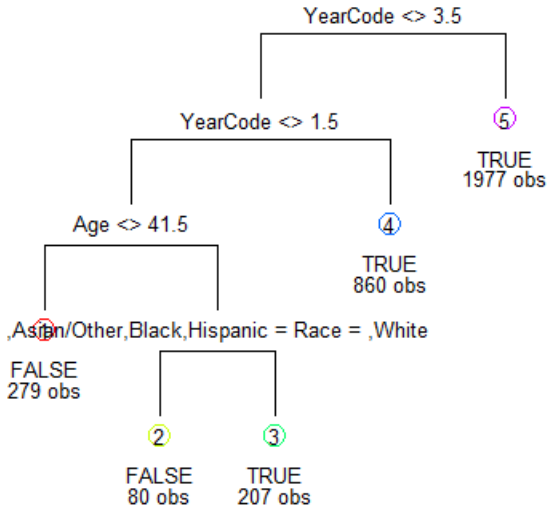
```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.20133    0.43694  -5.038 4.70e-07 ***
RaceBlack    -0.14586    0.42677  -0.342    0.733
RaceHispanic  0.04828    0.42030   0.115    0.909
RaceWhite    -0.12505    0.40462  -0.309    0.757
Age          -0.02625    0.00344  -7.632 2.31e-14 ***
SexMale       0.52535    0.09559   5.496 3.89e-08 ***
YearCode      0.73262    0.02764  26.509  < 2e-16 ***
```

Unsurprisingly, fentanyl deaths increased dramatically with time. Men were more likely to die of synthetic opioids, as were younger drug abusers, perhaps because older people find it easier to be prescribed prescription opioids. This suggests that strategies to fight overdoses should be targeted at the young. On the test data set, this was the most effective predictor, at 72.7% accuracy against 55% naively.

We also generated decision trees to answer our business question. In general, this approach was less successful, in some cases only generating trees when the cp hyperparameter was set to a low threshold. For overall opiate deaths, the decision tree only identified one cluster of non-opiate-overdoses, namely older black drug abusers – and even here, after 2015 this demographic was more likely to OD on opiates:

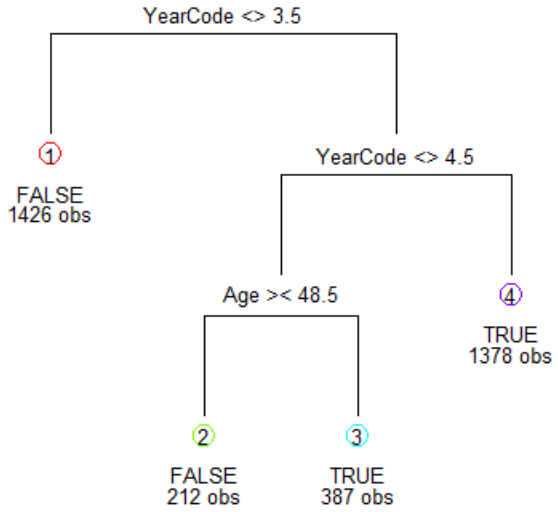


Looking at prescription drug deaths, the decision tree was more informative:



Before 2014, most overdoses, except among older whites, did not involve prescription drugs.

For fentanyl deaths, the increase over time is particularly stark:



Deaths before 2016 (Year Code 4) seldom involved synthetics like fentanyl. In 2016, as the epidemic picked up speed, deaths due to synthetics were common among the young. In 2017 and 2018, the epidemic was pervasive enough that most overdoses among all ages involved synthetics.

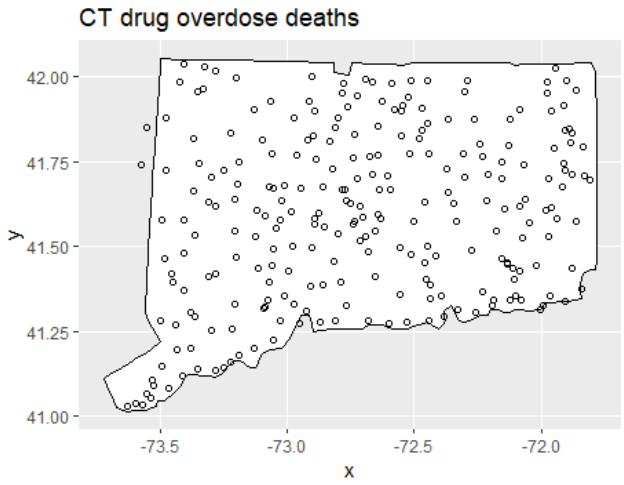
Data Mining

We ran an associative rules analysis (using the apriori() function) to discover combinations of drugs that frequently occurred together. Strikingly, of the 23 rules the algorithm found, almost all involve the combination of both an opiate and a stimulant (either amphetamine or cocaine). This is on the one hand not surprising – the potential lethality of mixing depressants and stimulants is well known – but at the same time, it runs quite opposite to the prevailing narrative that the epidemic is almost exclusively about opioids. Perhaps successful harm-reduction strategies could discourage the combined use of these drugs.

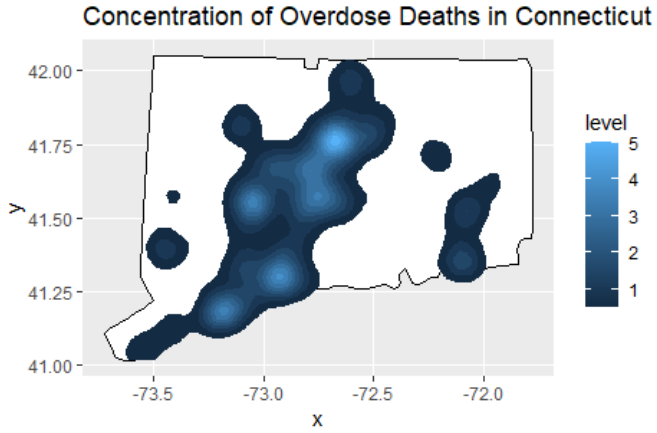
```
  lhs                                          rhs         support    confidence coverage   lift     count
[1]  {Oxymorphone}                             => {Oxycodone} 0.01743389 0.8240741  0.02115573 6.930639  89
[2]  {Amphet}                                  => {Fentanyl}  0.01567091 0.5031447  0.03114594 1.150786  80
[3]  {FentanylAnalogue}                        => {Fentanyl}  0.07600392 0.9974293  0.07619980 2.281307 388
[4]  {Fentanyl,Amphet}                         => {Heroin}    0.01018609 0.6500000  0.01567091 1.312080  52
[5]  {Heroin,Amphet}                           => {Fentanyl}  0.01018609 0.6753247  0.01508325 1.544593  52
[6]  {Fentanyl,Methadone}                      => {Heroin}    0.01527914 0.6341463  0.02409403 1.280078  78
[7]  {Heroin,Methadone}                        => {Fentanyl}  0.01527914 0.5492958  0.02781587 1.256342  78
[8]  {FentanylAnalogue,Ethanol}                => {Fentanyl}  0.01939275 1.0000000  0.01939275 2.287186  99
[9]  {FentanylAnalogue,Benzodiazepine}         => {Fentanyl}  0.02076396 0.9906542  0.02095984 2.265811 106
[10] {FentanylAnalogue,Benzodiazepine}         => {Heroin}    0.01116552 0.5327103  0.02095984 1.075321  57
[11] {Cocaine,FentanylAnalogue}                => {Fentanyl}  0.02742409 1.0000000  0.02742409 2.287186 140
[12] {Cocaine,FentanylAnalogue}                => {Heroin}    0.01371205 0.5000000  0.02742409 1.009292  70
[13] {Heroin,FentanylAnalogue}                 => {Fentanyl}  0.03643487 1.0000000  0.03643487 2.287186 186
[14] {Cocaine,Ethanol}                         => {Heroin}    0.03212537 0.5015291  0.06405485 1.012379 164
[15] {Cocaine,Benzodiazepine}                  => {Fentanyl}  0.03016650 0.5724907  0.05269344 1.309393 154
[16] {Cocaine,Benzodiazepine}                  => {Heroin}    0.02801175 0.5315985  0.05269344 1.073076 143
[17] {Fentanyl,Benzodiazepine}                 => {Heroin}    0.05269344 0.5046904  0.10440744 1.018760 269
[18] {Fentanyl,FentanylAnalogue,Benzodiazepine} => {Heroin}    0.01116552 0.5377358  0.02076396 1.085465  57
[19] {Heroin,FentanylAnalogue,Benzodiazepine}  => {Fentanyl}  0.01116552 1.0000000  0.01116552 2.287186  57
[20] {Cocaine,Fentanyl,FentanylAnalogue}       => {Heroin}    0.01371205 0.5000000  0.02742409 1.009292  70
[21] {Heroin,Cocaine,FentanylAnalogue}         => {Fentanyl}  0.01371205 1.0000000  0.01371205 2.287186  70
[22] {Cocaine,Fentanyl,Benzodiazepine}         => {Heroin}    0.01527914 0.5064935  0.03016650 1.022400  78
[23] {Heroin,Cocaine,Benzodiazepine}           => {Fentanyl}  0.01527914 0.5454545  0.02801175 1.247556  78

>
```

Geospatial Analytics

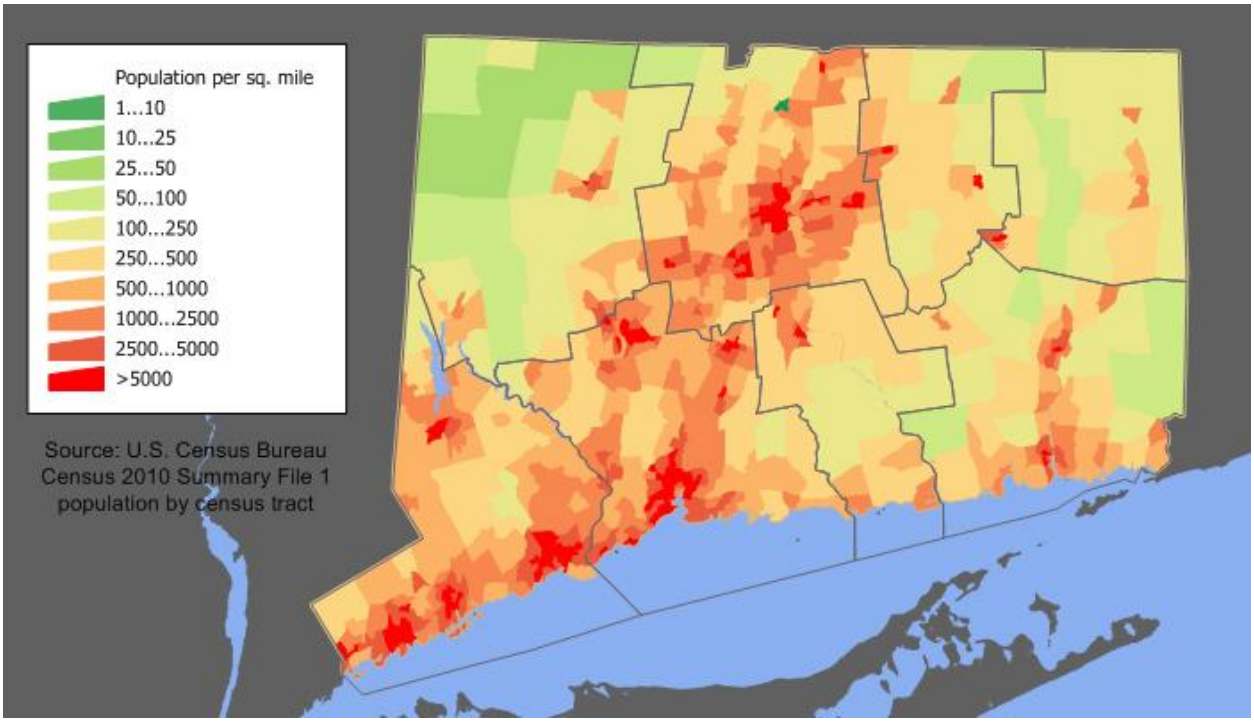
Unsurprisingly, overdose deaths occurred throughout the state:



However, deaths were not evenly concentrated in the state:



To some extent, these clusters just correspond to overal population density, as is clear when comparing the previous map to this density map from Wikipedia:



But some discrepancies jump out. Notably, the southwest corner of the state, though very populous, does not have as nearly a high number of overdose deaths. This likely is a function of wealth – southwestern Connecticut is home to Greenwich and similar towns, which are among the wealthiest in the country. Deaths are much more common in the poorer, more post-industrial center of the state.

Insights and Conclusion

Our analysis confirms the devasting growth in overdose deaths across Connecticut from 2012 to 2018. Two actionable insights stand out from the conventional narrative, however. First, our associative rules analysis shows while the specific drugs -and the number of people consuming them – are far more popular than in the past, the pharmacology of what combinations of drugs are particularly dangerous

has not varied as much. The conventional wisdom focuses primarily on depressants, especially synthetic opioids, but a remarkably common finding was the number of people who mixed opioids with stimulant drugs. Though stimulants have received far less attention than opioids (at least since the decline of the methamphetamine outbreak that was prominent about a decade ago), they greatly magnify the harms of opioids when taken in combination of them. Sensible anti-overdose policy, then, should focus particularly hard on discouraging the combination of these drugs.

Secondly, an unexpected gender component showed up in our logit regressions. Men were much more likely to die of synthetic opioids, while women were much more likely to die of prescription medications. Care should be taken to ascertain the extent to which this reflects differing degrees of access to prescription medications. But this analysis does suggest that efforts to fight the epidemic should take sex into account. Perhaps doctors are more likely to recognize (and prevent) drug-seeking behavior in men then women, and would benefit from being reminded that both sexes abuse opioids at staggering rates.

Appendix: R Code

```
#libraries


library(readr)

library(stringr)

library(ggplot2)

library(rpart)

library(dplyr)

library(tidyverse)

library(maptree)

library(arules)


###############################################


# loading the data


ctod <- read_csv("Accidental_Drug_Related_Deaths_2012-2018.csv")
```

```
####################

#Geospatial analysis

CityToLatLon <- function(dc)

{

  return(as.numeric(unlist(str_extract_all(dc, "-
?\\d+\\.*\\d*")),useNames = FALSE))

  }


tmp <- CityToLatLon(od$ResidenceCityGeo)


ctod$long <- tmp[seq(0,nrow(od)*2,2)]

ctod$lat <- tmp[seq(1,nrow(od)*2,2)]


ctod$state <- "connecticut"


dummyDF <- data.frame("connecticut", stringsAsFactors = FALSE)

dummyDF$state <- "connecticut"

#dummyDF$state <- tolower(dummyDF$state.name)

dummyDF
```

```
us <- map_data("state")

ct <- us[us$region == "connecticut",]

ct


map.simple <- ggplot(dummyDF, aes(map_id = state))


map.simple <- map.simple + geom_map(map = ct, fill = "white",

                                    color = "black")


map.simple <- map.simple + expand_limits(x = ct$long, y = ct$lat)


map.simple <- map.simple + coord_map() + ggtitle("CT drug overdose
deaths")

map.simple


#density map

map.simple + stat_density2d(data = ctod, aes(x = long, y = lat, fill=
after_stat(level)), geom="polygon") + ggtitle("Concentration of
Overdose Deaths in Connecticut")+ xlab("Longitude") + ylab("Latitude")


#point map
```

```
map.simple + geom_point(data = ctod, aes(x = long, y = lat), shape =
1) + xlab("Longitude") + ylab("Latitude")



###Main analysis


summary (ctod)

str(ctod)

##################

##process race into 4 factors (white/black/hispanic/asian or other)


unique(ctod$Race)

ctod$Race[str_detect(ctod$Race, "Hispanic")] <- "Hispanic"

ctod$Race[str_detect(ctod$Race, "Asian|Chinese|Native
American|Hawaiian|Other")] <- "Asian/Other"

ctod$Race[str_detect(ctod$Race, "Unknown")] <- NA

ctod$Race <- factor(ctod$Race)

unique(ctod$Race)


##Process sex into a factor
```

```
ctod$Sex[str_detect(ctod$Sex, "Unknown")] <- NA

ctod$Sex <- factor(ctod$Sex)


unique(ctod$Sex)







sapply(ctod, function(x) sum(is.na (x)))

#NAs are rare for demographic columns (age, sex, race), so safe to
ignore rows missing those.



# clean the data and make a subset with Date, Age and few drug
columns.


#updated_accidental_drug_deaths <-
subset(accidental_Drug_Related_Deaths_2012_2018,
accidental_Drug_Related_Deaths_2012_2018$date,
accidental_Drug_Related_Deaths_2012_2018$Age,
accidental_Drug_Related_Deaths_2012_2018$Heroin,
accidental_Drug_Related_Deaths_2012_2018$Cocaine,
```

```
accidental_Drug_Related_Deaths_2012_2018$Fentanyl,
accidental_Drug_Related_Deaths_2012_2018$FentanylAnalogue,
accidental_Drug_Related_Deaths_2012_2018$OpiateNOS,
accidental_Drug_Related_Deaths_2012_2018$AnyOpioid)

ctod_updated <- ctod[,c("Date", "Age", "Race", "Sex", "Heroin",
"Cocaine", "Fentanyl",
"FentanylAnalogue","Oxycodone","Oxymorphone","Ethanol","Hydrocodone","
Benzodiazepine","Methadone","Amphet","Tramad","Morphine_NotHeroin","Hy
dromorphone","Other", "OpiateNOS", "AnyOpioid")]



head(ctod_updated)



#Process drug columns into Booleans
ctod_updated$Heroin <- ifelse(is.na(ctod_updated$Heroin), FALSE,TRUE)
ctod_updated$Cocaine <- ifelse(is.na(ctod_updated$Cocaine),
FALSE,TRUE)
ctod_updated$Fentanyl <- ifelse(is.na(ctod_updated$Fentanyl),
FALSE,TRUE)
ctod_updated$FentanylAnalogue <-
ifelse(is.na(ctod_updated$FentanylAnalogue), FALSE,TRUE)
```

```
ctod_updated$Oxycodone <- ifelse(is.na(ctod_updated$Oxycodone),
FALSE,TRUE)

ctod_updated$Oxymorphone <- ifelse(is.na(ctod_updated$Oxymorphone),
FALSE,TRUE)

ctod_updated$Ethanol <- ifelse(is.na(ctod_updated$Ethanol),
FALSE,TRUE)

ctod_updated$Hydrocodone <- ifelse(is.na(ctod_updated$Hydrocodone),
FALSE,TRUE)

ctod_updated$Benzodiazepine <-
ifelse(is.na(ctod_updated$Benzodiazepine), FALSE,TRUE)

ctod_updated$Methadone <- ifelse(is.na(ctod_updated$Methadone),
FALSE,TRUE)

ctod_updated$Amphet <- ifelse(is.na(ctod_updated$Amphet), FALSE,TRUE)

ctod_updated$Tramad <- ifelse(is.na(ctod_updated$Tramad), FALSE,TRUE)

ctod_updated$Morphine_NotHeroin <-
ifelse(is.na(ctod_updated$Morphine_NotHeroin), FALSE,TRUE)

ctod_updated$Hydromorphone <-
ifelse(is.na(ctod_updated$Hydromorphone), FALSE,TRUE)

ctod_updated$Other <- ifelse(is.na(ctod_updated$Other), FALSE,TRUE)

ctod_updated$OpiateNOS <-
ifelse(is.na(ctod_updated$OpiateNOS),FALSE,TRUE)
```

```
ctod_updated$AnyOpioid <- ifelse(is.na(ctod_updated$AnyOpioid),
FALSE,TRUE)
ctod_updated


#fix date format
ctod_updated$Date <- as.Date(ctod_updated$Date,format = "%m/%d/%Y")
ctod_updated




ctod_updated$Year <- as.numeric(substr(ctod_updated$Date,1,4))
#Year code (starting at 0) makes more sense for regressions.
ctod_updated$YearCode <- ctod_updated$Year - 2012


#all of the following are opiates (per Wikipedia)
ctod_updated$IsOpiate <- (ctod_updated$Heroin | ctod_updated$Fentanyl
| ctod_updated$FentanylAnalogue | ctod_updated$Oxycodone |
                      ctod_updated$Oxymorphone | ctod_updated$Oxycodone
| ctod_updated$Hydrocodone | ctod_updated$Methadone
                      | ctod_updated$Tramad |
ctod_updated$Morphine_NotHeroin | ctod_updated$OpiateNOS |
ctod_updated$AnyOpioid)
```

```
#All of the following are (generally) prescription medications.

ctod_updated$IsPrescription <- (ctod_updated$Fentanyl |
ctod_updated$Oxycodone | ctod_updated$Oxycodone |
ctod_updated$Oxymorphone
                                | ctod_updated$Hydrocodone |
ctod_updated$Benzodiazepine | ctod_updated$Methadone |
ctod_updated$Tramad
                                | ctod_updated$Morphine_NotHeroin |
ctod_updated$Hydromorphone)


ctod_updated$IsSynthetic <- (ctod_updated$Fentanyl |
ctod_updated$FentanylAnalogue | ctod_updated$Tramad)



mean(ctod_updated$IsOpiate)

mean(ctod_updated$IsPrescription)

mean(ctod_updated$IsSynthetic)


#plot deaths from synthetic opiods
```

```
ggplot(ctod_updated[ctod_updated$IsSynthetic,], aes(x = Year)) +
geom_bar() + ggtitle("CT synthetic opioid deaths by year") +
ylab("Deaths")



##########################################



deathsbyyear <- ctod_updated %>% drop_na() %>% group_by(Year) %>%
summarize(Count = n())
#plot all deaths
ggplot(ctod_updated, aes(x = Year)) + geom_bar() + ggtitle("CT
overdose deaths by year") + ylab("Deaths")

# ggplot(updated_accidental_drug_deaths, aes(x = Age, y = Year)) +
geom_point() + stat_smooth()



######################
#Generate test and training subsets
```

```
randPermutation <-
sample(1:nrow(ctod_updated),nrow(ctod_updated),replace = FALSE)
randPermutation


cutoff <- floor(nrow(ctod_updated)*2/3)


ctTrain <- ctod_updated[randPermutation[1:cutoff],]
ctTest <- ctod_updated[randPermutation[(cutoff +
1):length(randPermutation)],]




# Multiple Linear Regression with Opiates (fit ) vs Non opiates(fit1)
#fit <- lm(Year ~ Heroin + Fentanyl + Oxycodone + Methadone + Tramad +
OpiateNOS + AnyOpioid, data=updated_accidental_drug_deaths)
logitModelOpiate <- glm(IsOpiate ~ Race + Age + + Sex + YearCode ,
data=ctTrain, family = binomial)
summary(logitModelOpiate)
predIsOpiate <- predict(logitModelOpiate, ctTest)


mean(predIsOpiate > 1, na.rm = TRUE)
```

```
predIsOpiate > 1

ctTest$IsOpiate

mean((predIsOpiate > .5) == ctTest$IsOpiate,na.rm = TRUE)




logitModelPrescription <- glm(IsPrescription ~ Race + Age + Sex +
YearCode , data=ctTrain, family = binomial)

summary(logitModelPrescription)

predIsPrescription <- predict(logitModelPrescription, ctTest)


mean((predIsPrescription > .5) == ctTest$IsPrescription,na.rm = TRUE)




##Synthetics

logitModelSynthetic <- glm(IsSynthetic ~ Race + Age + + Sex + YearCode
, data=ctTrain, family = binomial)

summary(logitModelSynthetic)

predIsSynthetic <- predict(logitModelSynthetic, ctTest)


mean((predIsSynthetic > .5) == ctTest$IsSynthetic,na.rm = TRUE)
```

```
##############################

#fit1 <- lm(Year ~  Cocaine +  FentanylAnalogue + Oxymorphone +
Ethanol + Hydrocodone + Benzodiazepine + Methadone + Amphet +
Morphine_NotHeroin + Hydromorphone + Other,
data=updated_accidental_drug_deaths)


dtModelOpiate <- rpart(IsOpiate ~ Race + Age + Sex + YearCode, data =
ctTrain, method = "class", cp = 0.005)

dtModelOpiate

draw.tree(dtModelOpiate,cex=0.8)

post(dtModelOpiate, file = "", title="Classification Tree for Opiate
Deaths")


dtModelPrescription <- rpart(IsPrescription ~ Race + Age + Sex +
YearCode, data = ctTrain, method = "class", cp = 0.01)

draw.tree(dtModelPrescription,cex=0.8)

post(dtModelPrescription, file = "", title="Classification Tree for
Prescription Drugs",cex = 0.3, tweak = .5)

dtModelPrescription
```

```
dtModelSynthetic <- rpart(IsSynthetic ~ Race + Age + Sex + YearCode,
data = ctTrain, method = "class", cp = 0.01)

draw.tree(dtModelSynthetic,cex=0.8)

post(dtModelSynthetic, file = "", title="Classification Tree for
Synthetic Opioids",cex = 0.5, tweak = .5)

dtModelPrescription

####################

# Associative rule mapping

drugs <- data.frame(ctod_updated[,c("Heroin","Cocaine", "Fentanyl",
"FentanylAnalogue",'Oxycodone',"Oxymorphone","Ethanol","Hydrocodone"
,
"Benzodiazepine","Methadone","Amphet","Tramad","Morphine_NotHeroin","H
ydromorphone","OpiateNOS")])



drugs <- as(data.matrix(drugs), "itemMatrix")

itemFrequencyPlot(drugs)

?itemFrequencyPlot

?apriori

ar <- apriori(drugs, parameter = list(support=0.01, confidence =
.5,maxlen = 10))

inspect(ar)
```