Statistical Analysis and Presentation on Drug Misuses

Presented to the Rocky Mountain Poison and Drug Safety center

ASA Datafest 2021

Team: Library (Purrr)

Team Members: Tina Wang, Rain Wu, George Huang, Ben Min

Guiding Research Questions

1. **Which type of people are more likely to misuse drugs?**
    a. Variables: demographic profiles, living habits, general health
    b. Predict whether one will misuse based on variables
    c. Find the 10 most significant variables for questionnaire
2. **Which type of drugs are more likely to be misused?**
    a. Proportion of misuse by drug
    b. Drugs that frequently get misused together
    c. Separate analysis by major drug types
3. **Do regional factors relate to drug misuse?**
    a. External Data: Healthcare CPI, Crime Severity, Access to Healthcare

1. Which type of people are more likely to misuse drugs?

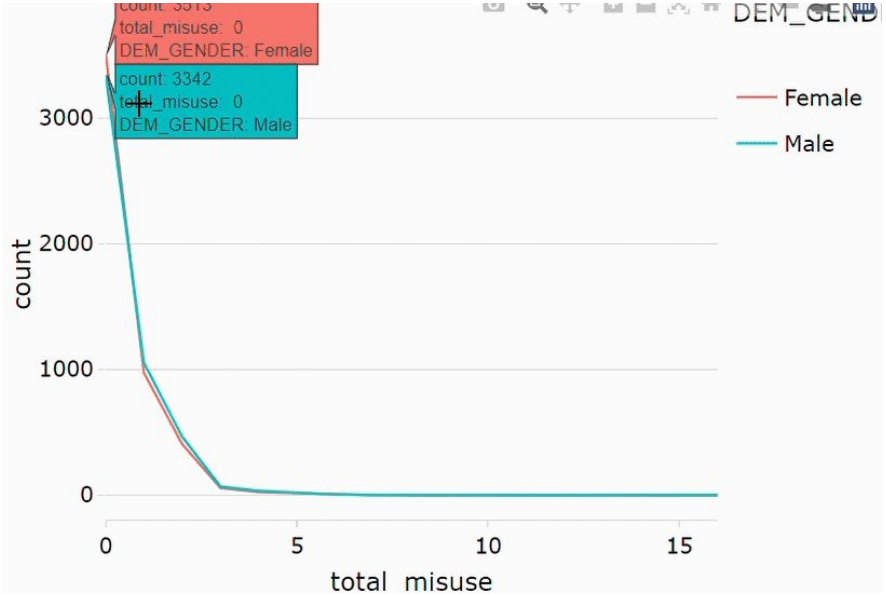
total_misuse (from 0 to 16) : counts the number of drugs one has misused
has_misuse (binary 0/1) : identifies whether one has ever misused any drug

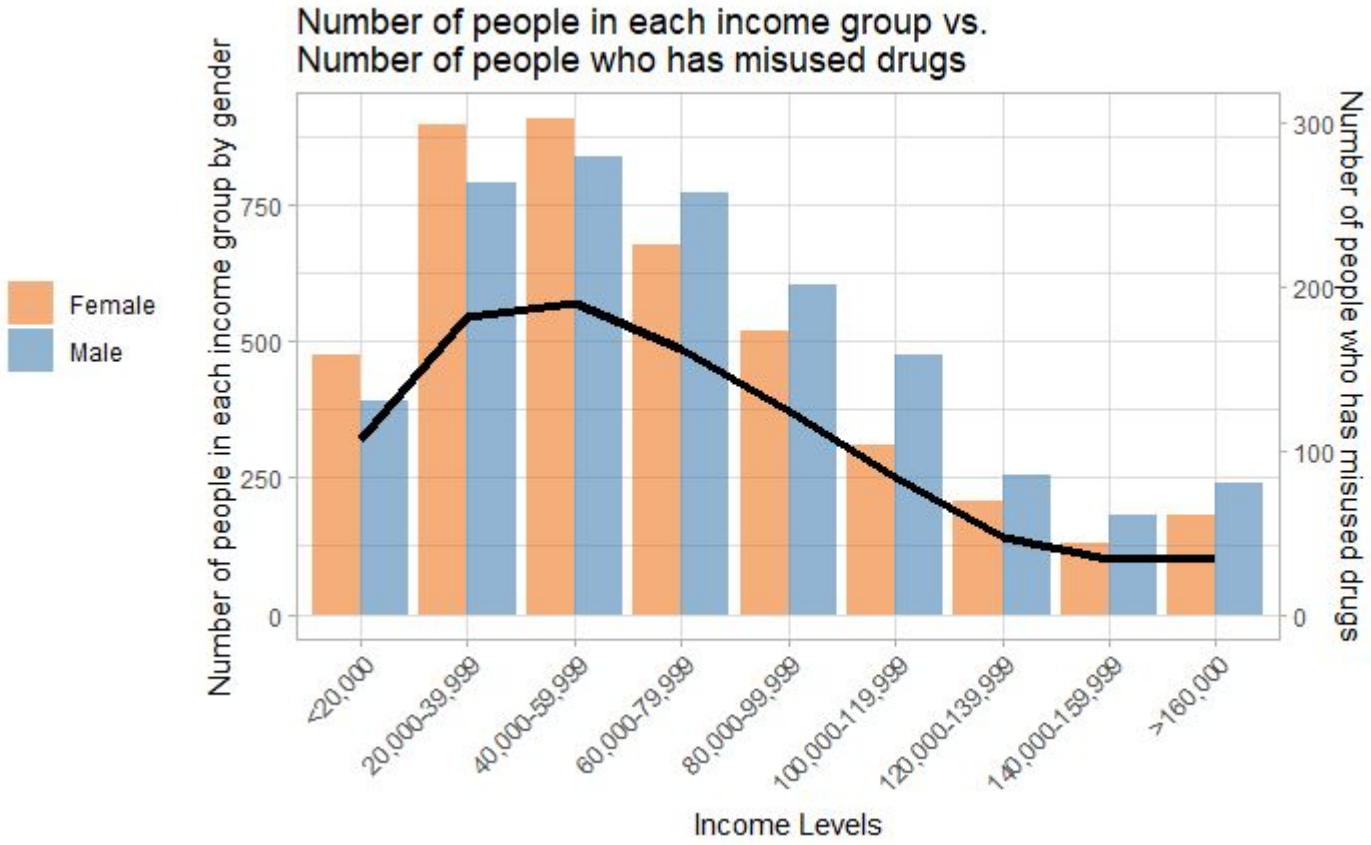
Variable	Female, N = 5,002[1]	Male, N = 5,005[1]
total_misuse		
0	3,513 (70%)	3,342 (67%)
1	974 (19%)	1,052 (21%)
2	407 (8.1%)	466 (9.3%)
3	57 (1.1%)	68 (1.4%)
4	25 (0.5%)	36 (0.7%)
5	15 (0.3%)	23 (0.5%)
6	8 (0.2%)	5 (<0.1%)
7	1 (<0.1%)	2 (<0.1%)
8	0 (0%)	3 (<0.1%)
9	0 (0%)	5 (<0.1%)
10	1 (<0.1%)	1 (<0.1%)
11	1 (<0.1%)	0 (0%)
16	0 (0%)	2 (<0.1%)

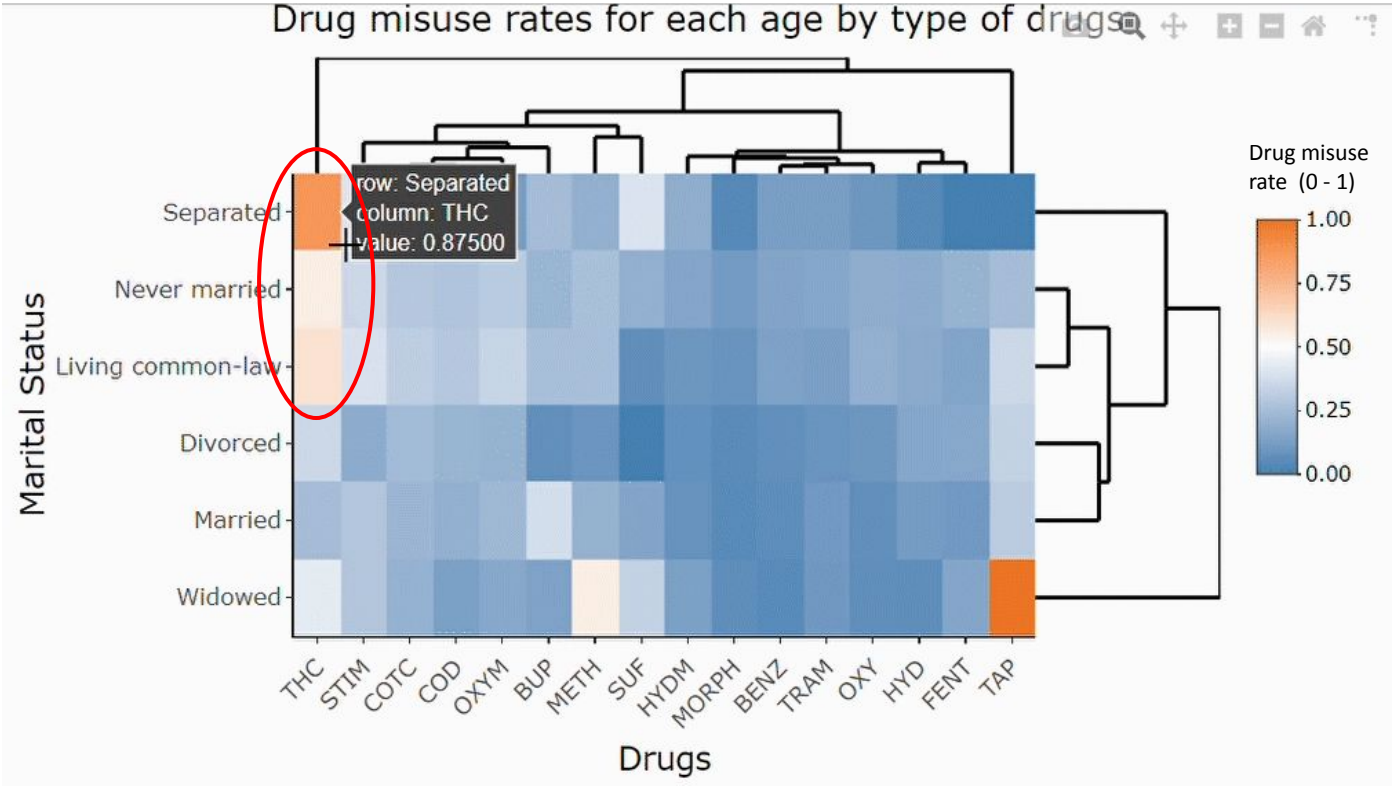
[1] n (%)

has_misuse = 0

has_misuse = 1




Number of people in each income group vs.
Number of people who has misused drugs
Female
Male
Number of people in each income group by gender
0
250
500
750
Number of people who has misused drugs
0
100
200
300
<20,000
20,000-39,999
40,000-59,999
60,000-79,999
80,000-99,999
100,000-119,999
120,000-139,999
140,000-159,999
>160,000
Income Levels


Drug misuse rates for each age by type of drugs
row: Separated
column: THC
value: 0.87500
Separated
Never married
Living common-law
Divorced
Married
Widowed
Marital Status
THC
STIM
COTC
COD
OXYM
BUP
METH
SUF
HYDM
MORPH
BENZ
TRAM
OXY
HYD
FENT
TAP
Drugs
Drug misuse rate (0 - 1)
1.00
0.75
0.50
0.25
0.00

Naive linear regressions for variable selection

Model 1 : demographic + other life-style habits

Model 2 : pregnancy + intentions to misuse or sell drugs

Model 3 : other pains + mental health disorders

Significant variables when regressed on total_misuse (Using Ca Data)

DEM_GENDER + DEM_AGE10 + DEM_ABOR + DEM_LOCATION+ DEM_MARITAL + DEM_INCOME + DEM_EDU + ALC_USE + TOB_USE + OTH_RX_DRUG_USE + HELP_SUB_USE + PAIN_CHRONIC + PAIN_ACUTE + MENT_NONE

Random Forest Classification for Misuse

Feature Selection:

Top 14 significant variables from previous regressions

Output:

Classification of whether one misuses or not (has_misuse)

Data:

80% training and 20% testing

Check Performance:

Prediction Confusion Matrix -> Accuracy, Sensitivity, Specificity

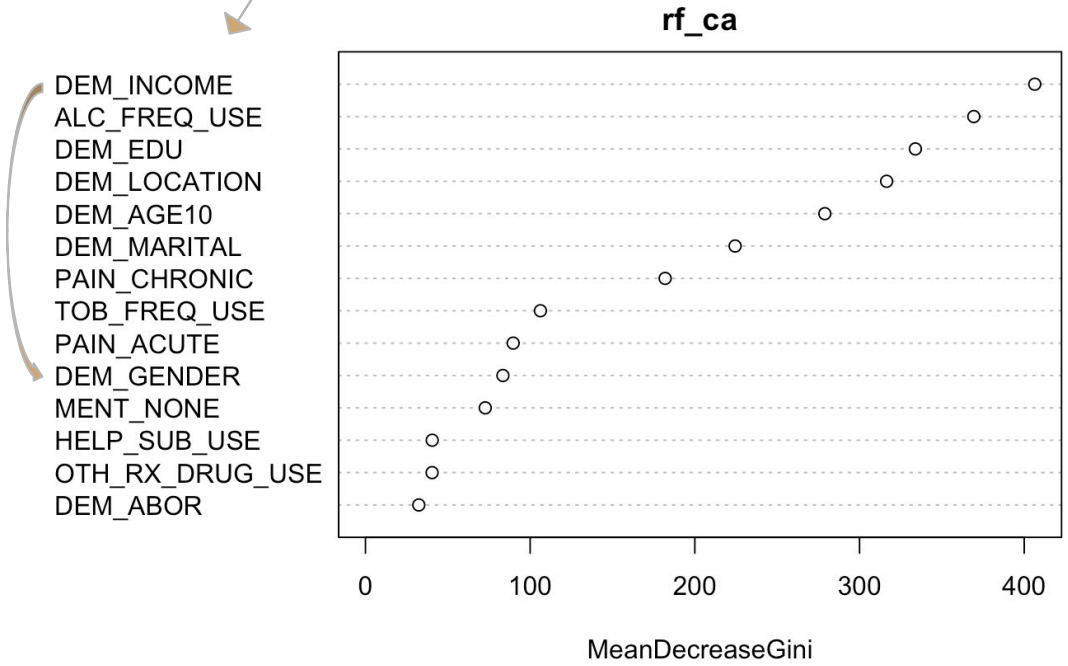
Technical:

Using randomForest(), predict() and library(ROCR) in R

```
                Type of random forest: classification
                      Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 30.18%
Confusion matrix:
     0   1 class.error
0 5109 387  0.07041485
1 2029 480  0.80868872
                MeanDecreaseGini
DEM_GENDER              83.48069
DEM_AGE10              278.92307
DEM_ABOR                32.40608
DEM_LOCATION           316.45244
DEM_MARITAL            224.46492
DEM_INCOME             406.39444
DEM_EDU                333.90860
ALC_FREQ_USE           369.38007
TOB_FREQ_USE           106.24833
OTH_RX_DRUG_USE         40.48211
HELP_SUB_USE            40.55028
PAIN_CHRONIC           181.97881
PAIN_ACUTE              89.70709
MENT_NONE               72.74968
```

Rank variables by the order of importance, where the top 10 are chosen for questionnaire



Survey of Non-Medical Use of Prescription Drugs Program

The Survey of Non-Medical Use of Prescription Drugs (NMURx) Program employs an online survey of the general adult population to understand non-medical use (NMU) of prescription drugs. Volunteers from the general population are queried about NMU of prescription drugs. This program collects demographic information and whether the respondent is a student, healthcare professional, or current/former member of the armed forces.

*Required

Are you *

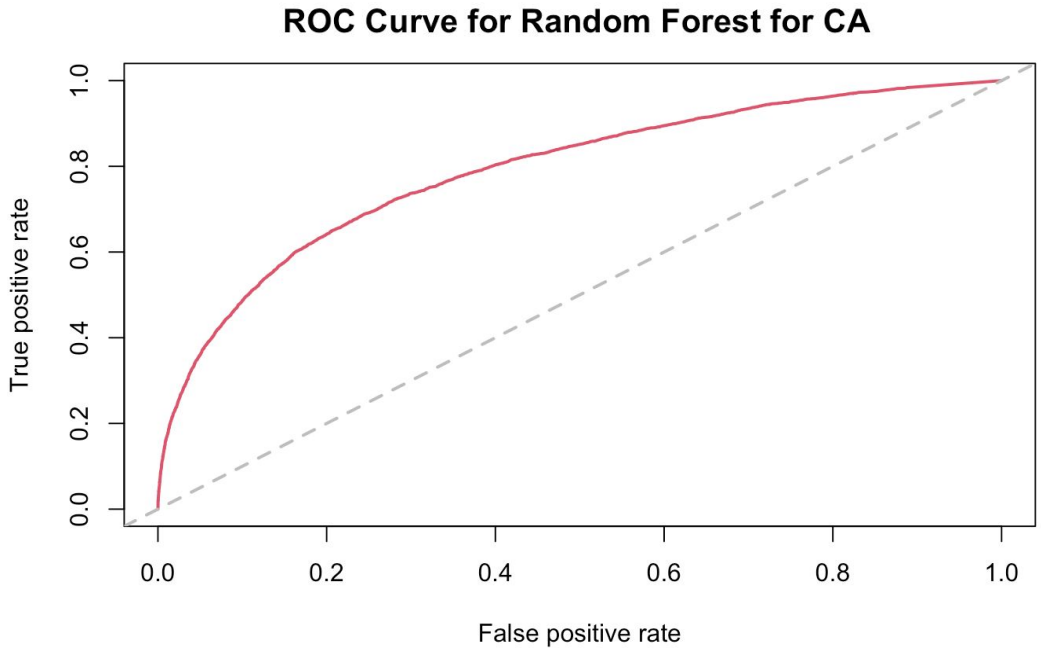
- ◯ Male
- ◯ Female

How old are you? *

- ◯ 35-44 years old
- ◯ 25-34 years old
- ◯ 15-24 years old
- ◯ 45-54 years old
- ◯ 65 or more years old
- ◯ 55-64 years old

Which province or territory in Canada do you currently live in? *

- ◯ Alberta
- ◯ Québec
- ◯ Ontario
- ◯ Saskatchewan




```
          Reference
Prediction    0    1
         0 1355  104
         1    4  539
```

High Prediction Accuracy

```
               Accuracy : 0.9461
                 95% CI : (0.9352, 0.9555)
    No Information Rate : 0.6788
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.871

 Mcnemar's Test P-Value : < 2.2e-16
```

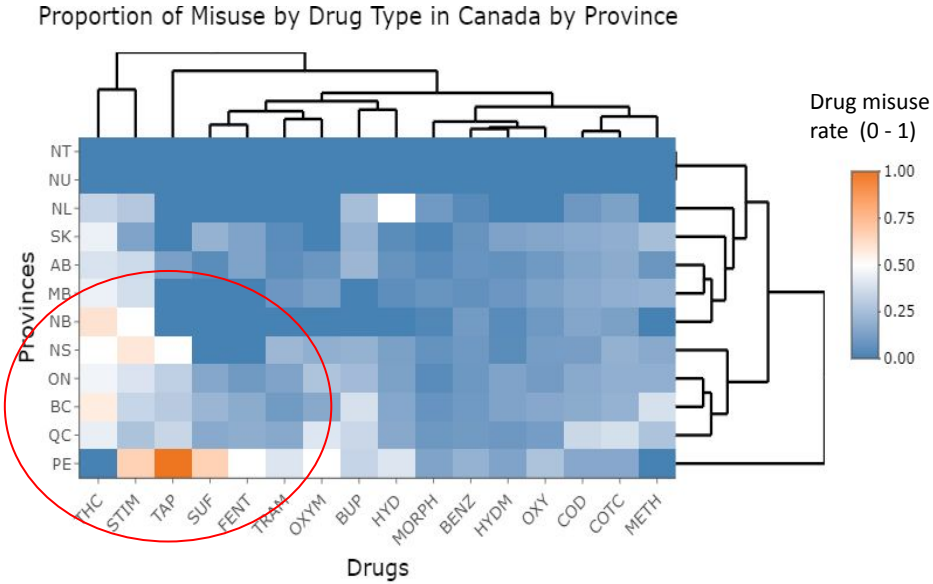
Accurately predicts misuse

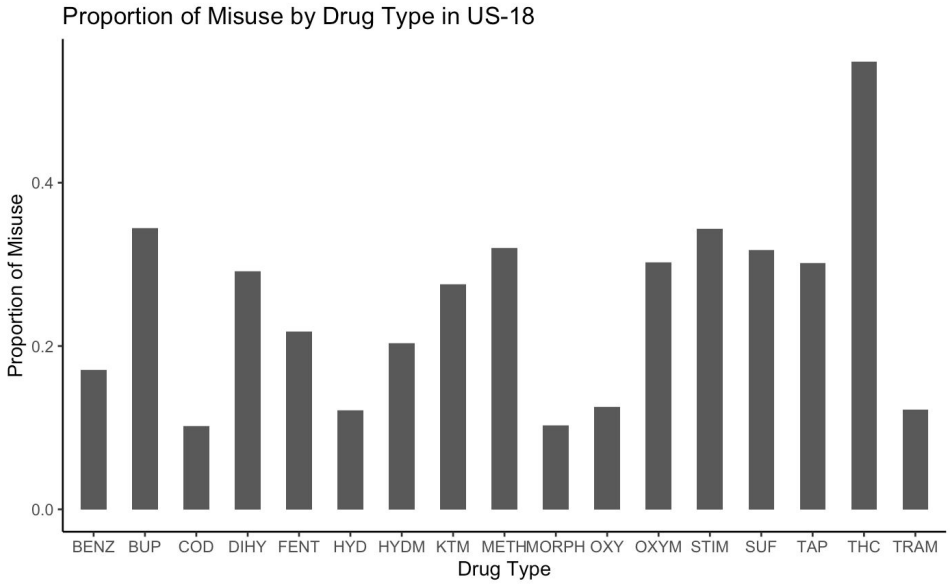
```
            Sensitivity : 0.9971
            Specificity : 0.8383
         Pos Pred Value : 0.9287
         Neg Pred Value : 0.9926
             Prevalence : 0.6788
         Detection Rate : 0.6768
   Detection Prevalence : 0.7288
      Balanced Accuracy : 0.9177

       'Positive' Class : 0
```

2. Which type of drugs are more likely to be misused?

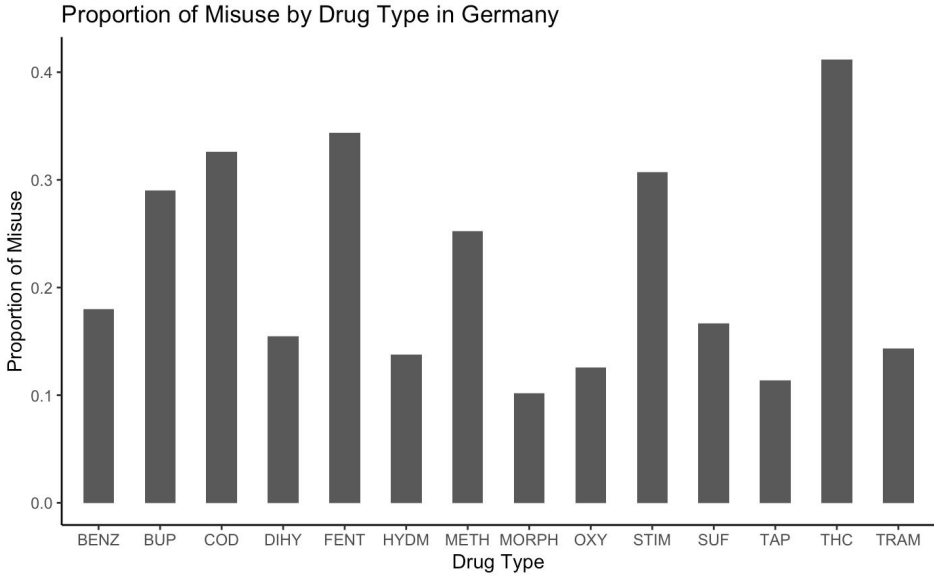
Proportion of drug misuses in each country

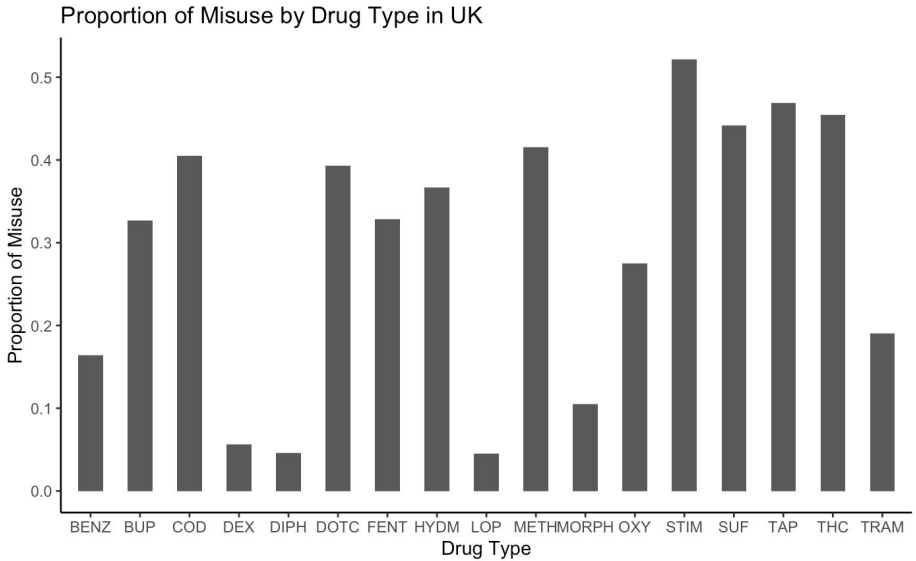




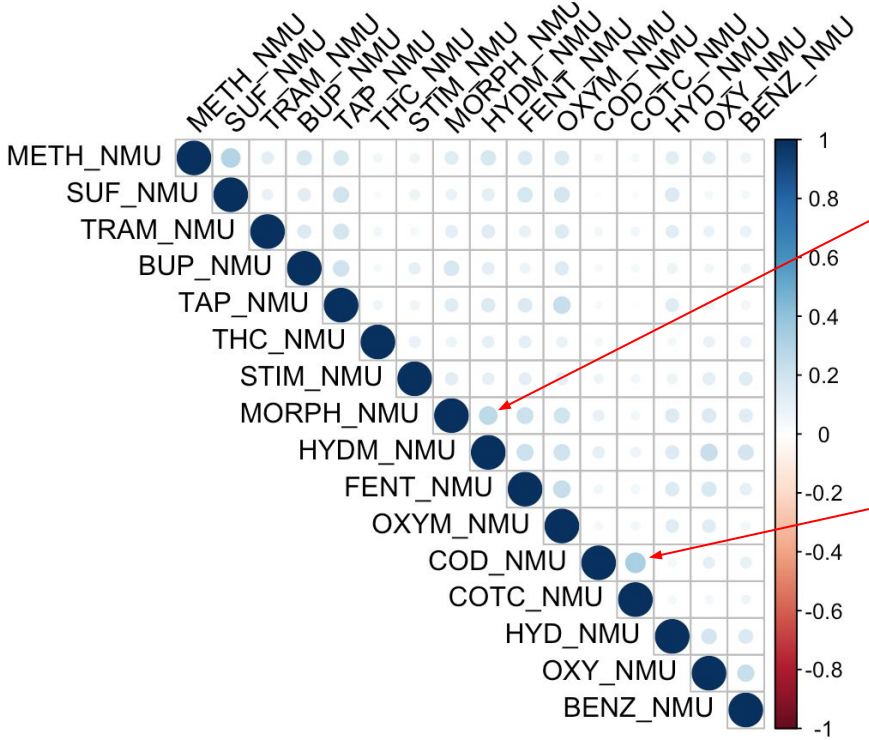
Top 3 most misused drugs by country:

Canada: 1. THC 2. STIM 3. TAP
US-18: 1. THC 2. BUP 3. STIM
Germany: 1. THC 2. FENT 3. COD
UK: 1. STIM 2. TAP 3. THC





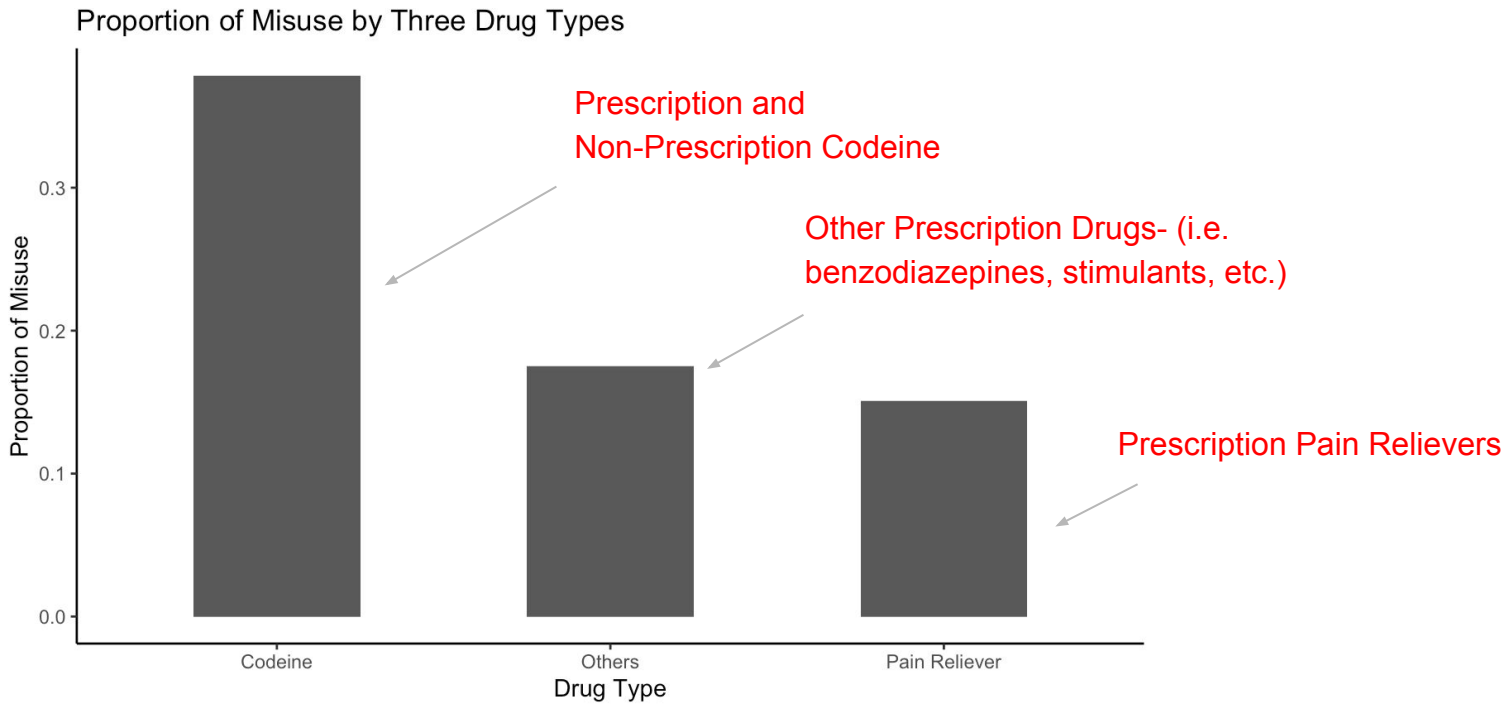
Correlation Matrix plotted for Drug Misuses (CA data)



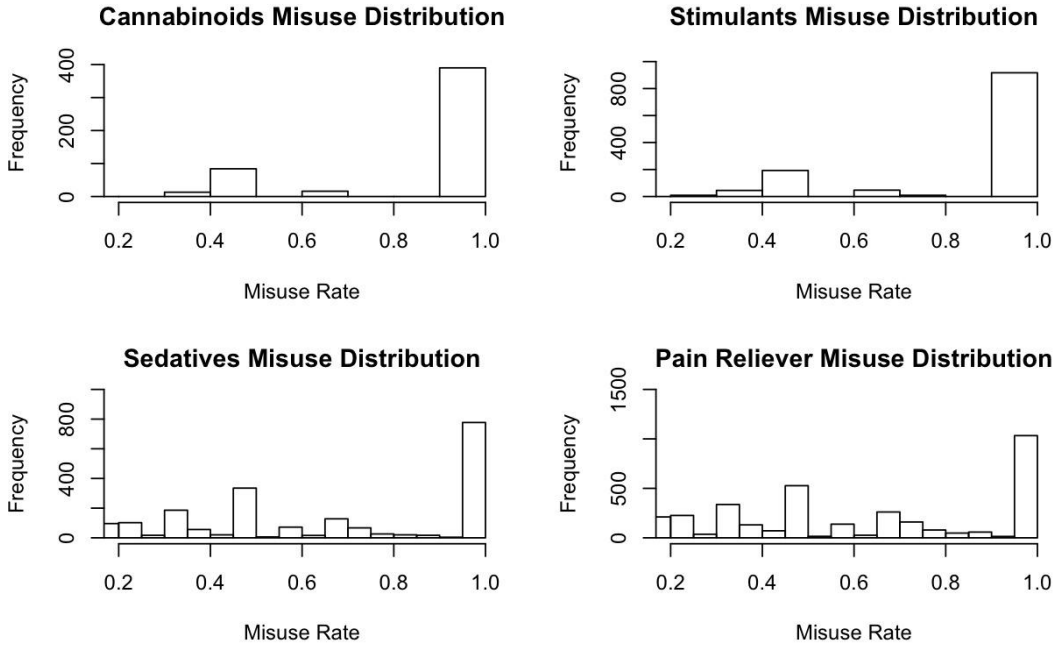
Eg. the correlation between **Morphine and HydroMorphone** is relatively large at 0.26

Eg. the correlation between **Codeine and Non-prescription Codeine** is relatively large at 0.33

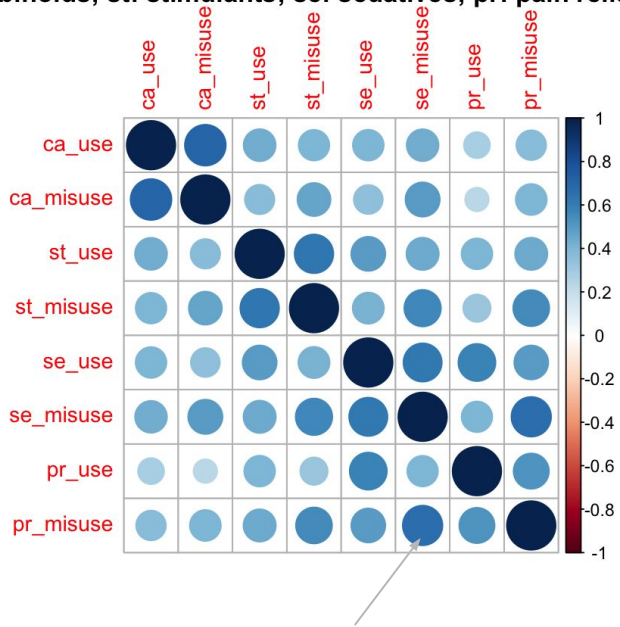
Logistic Regression based on Drug Types



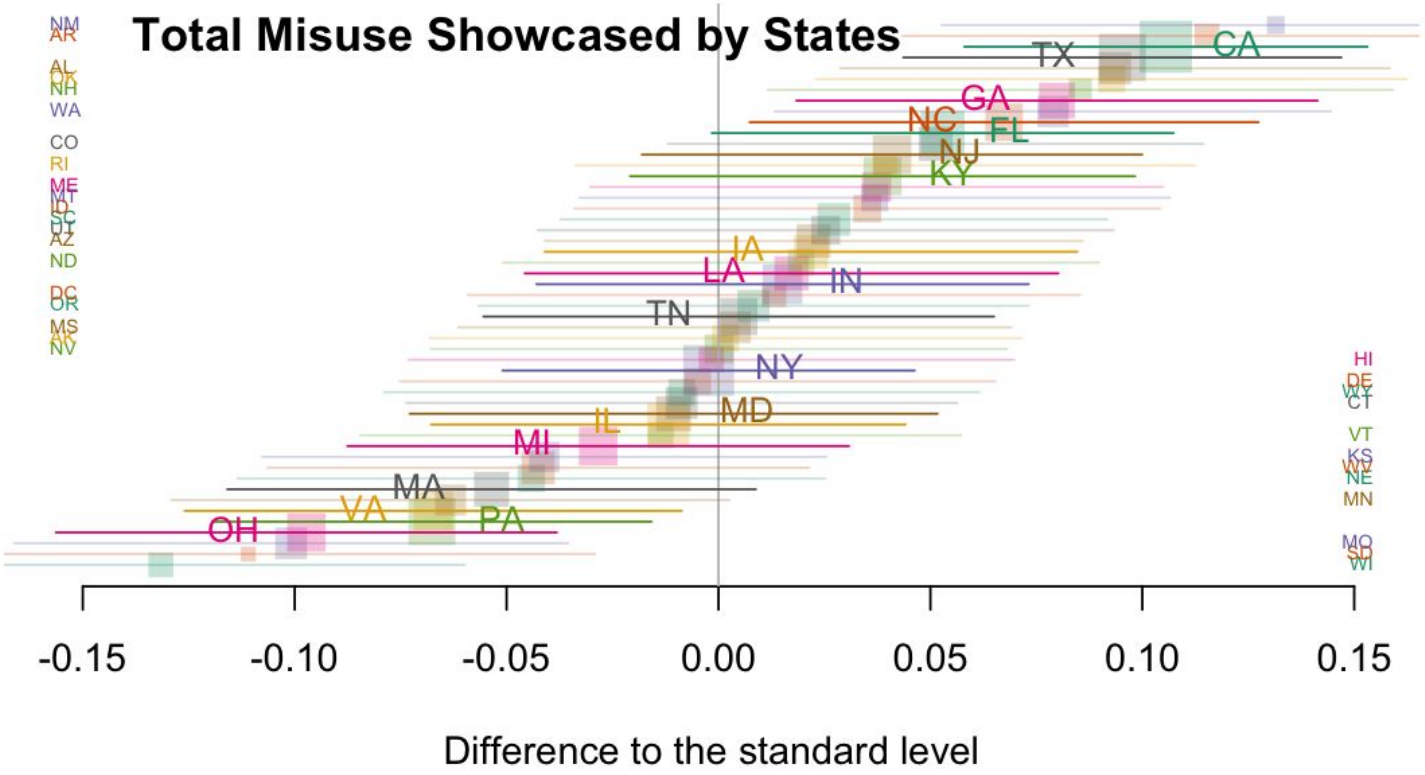
Analysis of US-19 Data - 4 drug types



ca: cannabinoids; st: stimulants; se: sedatives; pr: pain relievers



Eg. high corr between pain reliever and sedatives misuse


Total Misuse Showcased by States
NM
AR
AL
NH
WA
CO
RI
ME
ID
SC
AZ
ND
DC
OR
MS
NV
TX
CA
GA
NC
FL
NJ
KY
IA
LA
IN
TN
NY
MD
IL
MI
MA
VA
PA
OH
HI
DE
CT
VT
KS
NE
MN
MO
SD
WI
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
Difference to the standard level

3. Which regional factors relate to drug misuse?

External Data #1

-Health Care CPI measured in Dec 2016 by Province

External Data #2

-Crime Severity Index (CSI) measured in 2017 by Province

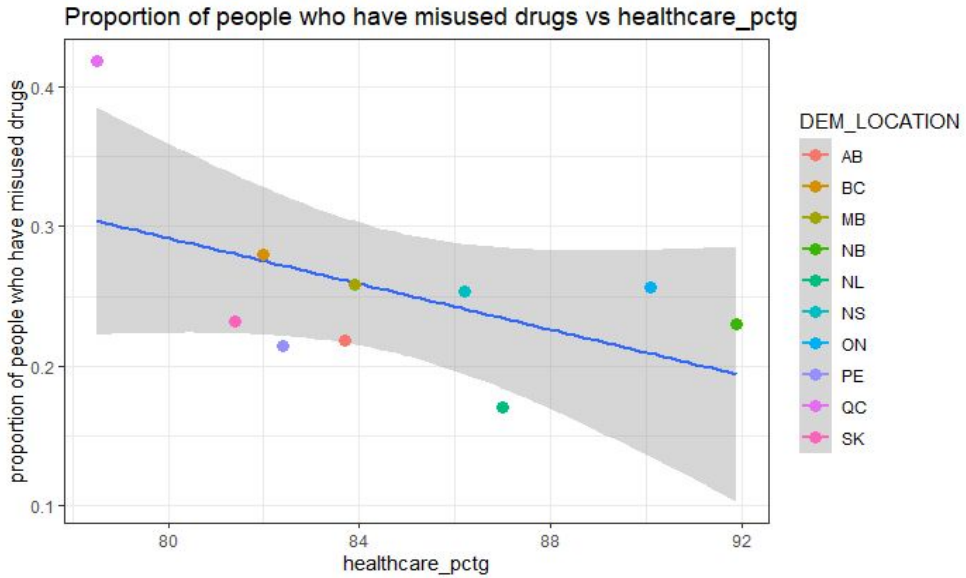
External Data #3

-Percentage of Pop. having access to healthcare measured in 2017 by Province

Source: Statistics Canada

Logistic Regression controlling for other demographic variables

```
healthcpi          -0.002342   0.003584  -0.653  0.51352
csi                -0.009678   0.001015  -9.533  < 2e-16 ***
healthcare_pctg -0.063032   0.004730 -13.325  < 2e-16 ***
```



A significant negative relationship:
higher % of pop with healthcare access, lower provincial drug misuse rate

Limitations and Future Improvements

- Time constraint on more detailed analysis or sophisticated models
- Communicating, editing and sharing of codes virtually

- Collect more data from under-represented groups
- Justify that extreme responses (16 misuses) are reliable
- Find more relevant external sources not limited to regional data
- Test model prediction accuracy with data collected from questionnaire
- A stronger model that predicts possible misuse in specific drugs

Conclusions

1. Constructed a questionnaire to predict misuse based on top 10 significant variables

2. Identified, for all datasets, the ranking of most misused individual drugs and most misused drug type

3. Visualized correlations amongst individual drugs and drug types (within & across categories)

4. External data reveals a significant negative relationship between percentage of population with access to healthcare and provincial proportion of drug misuse

Thank you

External Data References

Government of Canada, Statistics Canada. *Consumer Price Index, Monthly, Percentage Change, Not Seasonally Adjusted, Canada, Provinces, Whitehorse and Yellowknife - Health and Personal Care*, Government of Canada, Statistics Canada, 21 Apr. 2021, www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000408&pickMembers%5B0%5D=1.2&cubeTimeFrame.startMonth=12&cubeTimeFrame.startYear=2017&referencePeriods=20171201%2C20171201.

Government of Canada, Statistics Canada. *Crime Severity Index and Weighted Clearance Rates, Canada, Provinces, Territories and Census Metropolitan Areas*, Government of Canada, Statistics Canada, 29 Oct. 2020, www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510002601.

Government of Canada, Statistics Canada. “Health Characteristics.” *Health Characteristics, Two-Year Period Estimates*, Government of Canada, Statistics Canada, 22 Oct. 2019, www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1310011301.