Term Project Step 3

Burden Joshua

2022-06-04

Legalization of Marijuana

Introduction

With more and more states adding medical and recreational cannabis to their ballots, questions are now being asked more openly about the impact of cannabis on public health, crime statistics, and popularity of legalization. The data out there is sparse, but the data that has been collected could show trends and insights on the future of legislation and public sentiment of legalization, along with showing the potential for lowered crime and raised tax revenue in states and at the federal level.

Research questions

I want to know, based on the data that is out there, if legalization of marijuana:

- Reduces the number of arrests of all nonviolent offenses significantly
- Would increase the tax income for states and government entities
- Would drug usage go down by making it legal
- What is the state tax collections for legal rec states vs med only vs no rec/med
- What is the public support for legalization

Approach

I will look at the data that is out there for sales, crime, and public sentiment and see if there is anything that stands out to answer any of the questions above.

How your approach addresses (fully or partially) the problem.

The data will tell me whether or not the any or all of my questions can be answered and if there is discrepancies in the data, more research might be conducted to see if there is correlation.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

- https://data.world/denver/marijuana-related-crime
- https://www.kaggle.com/datasets/tunguz/drug-use-by-age
- https://www.kaggle.com/datasets/mykeysw/marijuana-sales-forecasting-in-tx
- https://data.world/sya/marijuana-laws-by-state
- https://data.world/denver/marijuana-gross-sales
- https://data.world/health/support-for-legal-marijuana
- https://www.kaggle.com/datasets/terenceshin/historical-prices-for-biggest-weed-stocks

Bad dataset

- https://data.world/opensavannah/cannabis-justice

Required Packages

- Gplot2

Plots and Table Needs

- Histograms
- Scatter plots
- CDF
- Linear Regressions

Questions for future steps

Is there more data out there that would help with this research, and how deep should these models go in terms of covering the questions?

Step 2

How to import and clean my data

Google is very good actually for making suggestions on cleaning datasets. That is my first step, by bringing these datasets into google sheets, I am effectively having google help me get the datasets clean. One other thing that I needed to do was to also parse what is important and what isn't based on the datasets. There were quite a number of datasets that were not helpful or would be considered a large reach for the dataset to fit into the story, so I removed those datasets from the list of datasets I had. I also made the decision early on to remove header column names that either didn't make sense or were not required for the dataset story.

What does the final data set look like?

Summaries of the dataset, as to not get overwelmed by the vast amounts of data.

```
setwd('/Users/joshua/Documents/PERSONAL_GITHUB_REPOS/dsc520/TermProject')
cosalesrev <- read.csv("data/cleaned_COMJSalesMonthReports2019.csv")
cotaxrev <- read.csv("data/cleaned_COMJTaxMonthReports2019.csv")
costatepop <- read.csv("data/cleaned_COCountyPop20142018.csv")
drug_use <- read.csv("data/cleaned_drug-use-by-age.csv")
crime_denver <- read.csv("data/cleaned_crime_marijuana.csv")
legal_support <-read.csv("data/cleaned_legal_marijuana_support.csv")
```

Summaries of datasets

```
# summaries of the datasets
summary(cosalesrev)
```

```
##     Month            MedicalSales       MSYearToDate        RetailSales
##  Length:65          Min.   :24082927   Min.   : 25680596   Min.   : 14022213
##  Class :character   1st Qu.:29238678   1st Qu.:100238704   1st Qu.: 44590853
##  Mode  :character   Median :32686869   Median :193094556   Median : 76018423
##                     Mean   :32736868   Mean   :204611449   Mean   : 70423116
##                     3rd Qu.:36010893   3rd Qu.:302926736   3rd Qu.: 96642441
##                     Max.   :41056948   Max.   :445616062   Max.   :114317739
##   RSYearToDate         TotalSales         TSYearToDate
##  Min.   :1.402e+07   Min.   : 45987045   Min.   :4.656e+07
##  1st Qu.:1.599e+08   1st Qu.: 78199969   1st Qu.:2.831e+08
```

```
##  Median :3.088e+08   Median :109222331   Median :5.399e+08
##  Mean   :4.086e+08   Mean   :103159984   Mean   :6.132e+08
##  3rd Qu.:5.775e+08   3rd Qu.:127706166   3rd Qu.:8.809e+08
##  Max.   :1.214e+09   Max.   :143107279   Max.   :1.546e+09
##   TotalToDate
##  Min.   :4.656e+07
##  1st Qu.:1.047e+09
##  Median :2.644e+09
##  Mean   :2.895e+09
##  3rd Qu.:4.612e+09
##  Max.   :6.705e+09
```

```
summary(cotaxrev)
```

```
##     Month              SalesTax        LicenseFees       SalesTaxFees
##  Length:65          Min.   : 628947   Min.   : 592661   Min.   :1570401
##  Class :character   1st Qu.:1014752   1st Qu.: 972122   1st Qu.:2137462
##  Mode  :character   Median :1808419   Median :1063563   Median :3067394
##                     Mean   :1970529   Mean   :1097654   Mean   :3068184
##                     3rd Qu.:2763721   3rd Qu.:1221521   3rd Qu.:4016545
##                     Max.   :3692930   Max.   :1663120   Max.   :4892115
##  RetailSalesTax     RetailExciseTax    TotalTaxFees        YearToDate
##  Min.   : 1401568   Min.   : 195318   Min.   : 3519756   Min.   :  3519756
##  1st Qu.: 4394550   1st Qu.:2796865   1st Qu.:10856584   1st Qu.: 37511919
##  Median : 7746575   Median :4683825   Median :17694953   Median : 76306924
##  Mean   : 8877802   Mean   :4114956   Mean   :16060942   Mean   : 93255239
##  3rd Qu.:14608085   3rd Qu.:5598581   3rd Qu.:21622509   3rd Qu.:134971077
##  Max.   :18698640   Max.   :7867853   Max.   :26841073   Max.   :266529637
##   TotalToDate
##  Min.   :3.520e+06
##  1st Qu.:1.260e+08
##  Median :3.555e+08
##  Mean   :4.161e+08
##  3rd Qu.:6.818e+08
##  Max.   :1.044e+09
```

```
summary(costatepop)
```

```
##     County             X2014Pop         X2015Pop         X2016Pop
##  Length:64          Min.   :   703   Min.   :   689   Min.   :   690
##  Class :character   1st Qu.:  5692   1st Qu.:  5736   1st Qu.:  5636
##  Mode  :character   Median : 14288   Median : 14358   Median : 14592
##                     Mean   : 83526   Mean   : 85076   Mean   : 86472
##                     3rd Qu.: 41946   3rd Qu.: 41975   3rd Qu.: 42592
##                     Max.   :664715   Max.   :683081   Max.   :696347
##     X2017Pop         X2018Pop
##  Min.   :   714   Min.   :   762
##  1st Qu.:  5837   1st Qu.:  5876
##  Median : 14747   Median : 15014
##  Mean   : 87648   Mean   : 88993
##  3rd Qu.: 43202   3rd Qu.: 43666
##  Max.   :705651   Max.   :716492
```

```
summary(drug_use)
```

```
##      age             sample_size   marijuana_use_by_percentage
```

```
##  Length:17          Min.   :2223   Min.   : 1.10
##  Class :character   1st Qu.:2469   1st Qu.: 8.70
##  Mode  :character   Median :2798   Median :20.80
##                     Mean   :3251   Mean   :18.92
##                     3rd Qu.:3058   3rd Qu.:28.40
##                     Max.   :7391   Max.   :34.00
##  marijuana_frequency_over_12_months
##  Min.   : 4.00
##  1st Qu.:30.00
##  Median :52.00
##  Mean   :42.94
##  3rd Qu.:52.00
##  Max.   :72.00
```

```
summary(crime_denver)
```

```
##   REPORTDATE         OFFENSE_CODE  OFFENSE_TYPE_ID    OFFENSE_CATEGORY_ID
##  Length:1254        Min.   :1006   Length:1254        Length:1254
##  Class :character   1st Qu.:2203   Class :character   Class :character
##  Mode  :character   Median :2203   Mode  :character   Mode  :character
##                     Mean   :2249
##                     3rd Qu.:2206
##                     Max.   :7399
##  NEIGHBORHOOD_ID
##  Length:1254
##  Class :character
##  Mode  :character
##
##
##
```

```
summary(legal_support)
```

```
##       Year         Month           Asked_half_sample    Yes_Legal
##  Min.   :1969   Length:20          Length:20          Min.   :12.00
##  1st Qu.:1980   Class :character   Class :character   1st Qu.:25.00
##  Median :2002   Mode  :character   Mode  :character   Median :34.00
##  Mean   :1997                                         Mean   :35.95
##  3rd Qu.:2011                                         3rd Qu.:48.50
##  Max.   :2016                                         Max.   :60.00
##    No_Illegal      No_Opinion   Percent_Yes         Percent_No
##  Min.   :39.00   Min.   :1.00   Length:20          Length:20
##  1st Qu.:49.25   1st Qu.:2.00   Class :character   Class :character
##  Median :63.00   Median :4.00   Mode  :character   Mode  :character
##  Mean   :60.50   Mean   :3.45
##  3rd Qu.:70.75   3rd Qu.:4.25
##  Max.   :84.00   Max.   :6.00
##  Percent_No_Opinion
##  Length:20
##  Class :character
##  Mode  :character
##
##
##
```

glimpse of datasets

```
glimpse(cosalesrev)
```

```
## Rows: 65
## Columns: 8
## $ Month        <chr> "Jan 2014", "Feb 2014", "Mar 2014", "Apr 2014", "May 2014~
## $ MedicalSales <int> 32541720, 31738572, 34821878, 32686869, 31355208, 2995030~
## $ MSYearToDate <int> 32541720, 64280292, 99102170, 131789039, 163144247, 19309~
## $ RetailSales  <int> 14022213, 14248473, 19881631, 20765986, 21375001, 2397808~
## $ RSYearToDate <int> 14022213, 28270686, 48152317, 68918303, 90293304, 1142713~
## $ TotalSales   <int> 46563933, 45987045, 54703509, 53452855, 52730209, 5392839~
## $ TSYearToDate <int> 46563933, 92550978, 147254487, 200707342, 253437551, 3073~
## $ TotalToDate  <dbl> 46563933, 92550978, 147254487, 200707342, 253437551, 3073~
```

```
glimpse(cotaxrev)
```

```
## Rows: 65
## Columns: 9
## $ Month           <chr> "Feb 2014", "Mar 2014", "Apr 2014", "May 2014", "Jun 2~
## $ SalesTax        <int> 1330209, 1460429, 1569405, 1559710, 1569454, 1530968, ~
## $ LicenseFees     <int> 592661, 857615, 902995, 761687, 940028, 1547853, 13795~
## $ SalesTaxFees    <int> 1922870, 2318044, 2472400, 2321397, 2509482, 3078821, ~
## $ RetailSalesTax  <int> 1401568, 1434916, 1898685, 2217607, 2070577, 2473627, ~
## $ RetailExciseTax <int> 195318, 339615, 609907, 734351, 1135648, 969637, 13979~
## $ TotalTaxFees    <int> 3519756, 4092575, 4980992, 5273355, 5715707, 6522085, ~
## $ YearToDate      <int> 3519756, 7612330, 12593322, 17866677, 23582384, 301044~
## $ TotalToDate     <int> 3519756, 7612330, 12593322, 17866677, 23582384, 301044~
```

```
glimpse(costatepop)
```

```
## Rows: 64
## Columns: 6
## $ County   <chr> "Adams", "Alamosa", "Arapahoe", "Archuleta", "Baca", "Bent", ~
## $ X2014Pop <int> 479477, 15758, 617498, 12240, 3576, 5777, 312588, 61617, 1845~
## $ X2015Pop <int> 489774, 15854, 628951, 12401, 3544, 5885, 318071, 64713, 1857~
## $ X2016Pop <int> 497419, 16006, 637266, 12839, 3522, 5664, 321363, 66399, 1907~
## $ X2017Pop <int> 503375, 16056, 643257, 13316, 3539, 5866, 322854, 68169, 1962~
## $ X2018Pop <int> 511868, 16683, 651215, 13765, 3585, 5882, 326078, 69267, 2002~
```

```
glimpse(drug_use)
```

```
## Rows: 17
## Columns: 4
## $ age                                <chr> "12", "13", "14", "15", "16", "17",~
## $ sample_size                        <int> 2798, 2757, 2792, 2956, 3058, 3038,~
## $ marijuana_use_by_percentage        <dbl> 1.1, 3.4, 8.7, 14.5, 22.5, 28.0, 33~
## $ marijuana_frequency_over_12_months <int> 4, 15, 24, 25, 30, 36, 52, 60, 60, ~
```

```
glimpse(crime_denver)
```

```
## Rows: 1,254
## Columns: 5
## $ REPORTDATE          <chr> "2/27/2012", "8/6/2012", "9/18/2012", "8/19/2012",~
## $ OFFENSE_CODE        <int> 2203, 2203, 2203, 5707, 2203, 2203, 2203, 2203, 22~
## $ OFFENSE_TYPE_ID     <chr> "BURGLARY - BUSINESS BY FORCE", "BURGLARY - BUSINE~
## $ OFFENSE_CATEGORY_ID <chr> "Burglary", "Burglary", "Burglary", "All Other Cri~
```

```
## $ NEIGHBORHOOD_ID     <chr> "montclair", "five-points", "hampden-south", "mont~
```

```
glimpse(legal_support)
```

```
## Rows: 20
## Columns: 9
## $ Year               <int> 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 200~
## $ Month              <chr> "Oct", "Oct", "Oct", "Oct", "Nov", "Oct", "Oct", "O~
## $ Asked_half_sample  <chr> "no", "no", "no", "no", "no", "no", "yes", "yes", "~
## $ Yes_Legal          <int> 60, 58, 51, 58, 48, 50, 46, 44, 36, 34, 34, 31, 25,~
## $ No_Illegal         <int> 39, 40, 47, 39, 50, 46, 50, 54, 60, 64, 62, 64, 73,~
## $ No_Opinion         <int> 1, 2, 2, 3, 1, 3, 4, 2, 4, 2, 4, 5, 2, 4, 5, 5, 6, ~
## $ Percent_Yes        <chr> "60%", "58%", "51%", "58%", "48%", "51%", "46%", "4~
## $ Percent_No         <chr> "39%", "40%", "47%", "39%", "51%", "46%", "50%", "5~
## $ Percent_No_Opinion <chr> "1%", "2%", "2%", "3%", "1%", "3%", "4%", "2%", "4%~
```

Questions for future steps.

I am in a situation where there is almost too much data, but not enough connections as there

What information is not self-evident?

I was going to add this datasets to other states, but for many states there is not an aggregate set of information out there for just marijuana related offenses, and without being specific, it becomes difficult to see patterns and trends when the net is cast to wide.

What are different ways you could look at this data?

Data like this is not very connected. Making a story out of the datasets requires looking at each dataset as a piece of a puzzle and not trying to force the datasets to work with each other but rather, answer a question and check if the answer relates to the next question.

How do you plan to slice and dice the data?

For one set of data, I plan to see the sales trend for Colorado based on sales, and tax data. I will also look at if crime went up specifically in Denver, and then see if there is another data set out there to see if there is a relationship with higher sales, with uprising crime and youth uses.

What types of plots and tables will help you to illustrate the findings to your questions?

barcharts and lineplots seem to make the most sense in these instances. I might also look at doing some data plotting with a scatterplot.

Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.

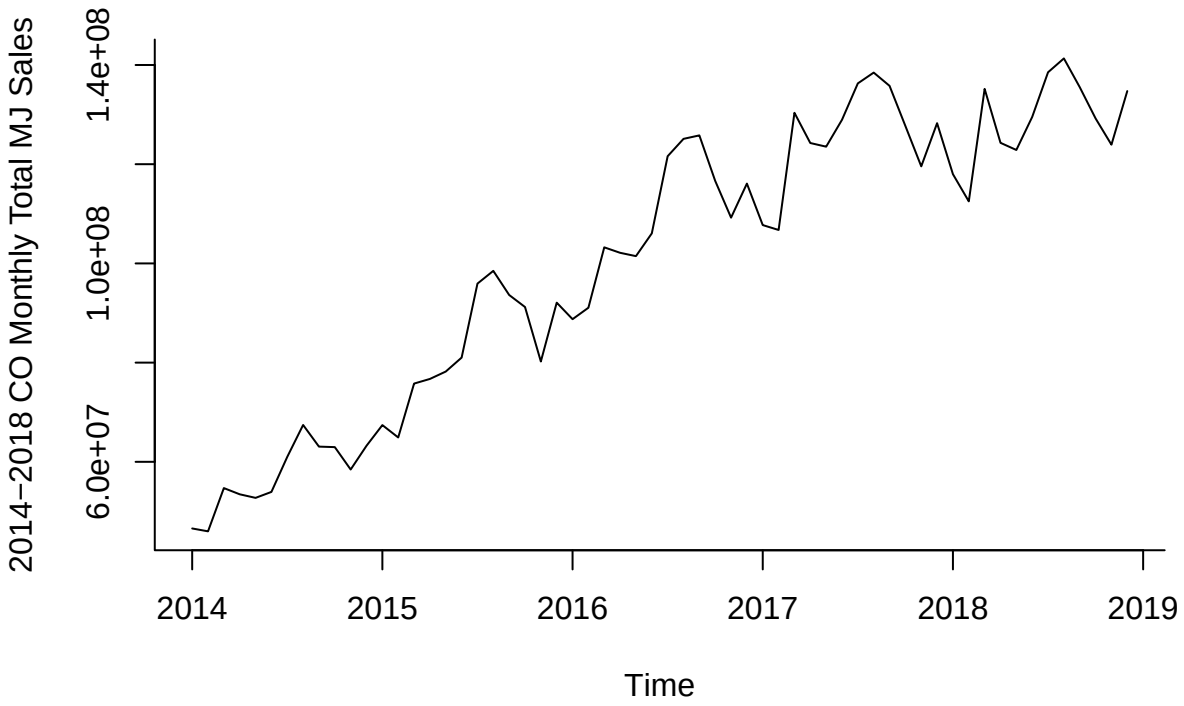
I am unsure at this time I will add a machine learning algorithm to these datasets. I would like to say yes, but I am still parsing data, and narrowing my search queries to something smaller than my original scope.

Questions for future steps.

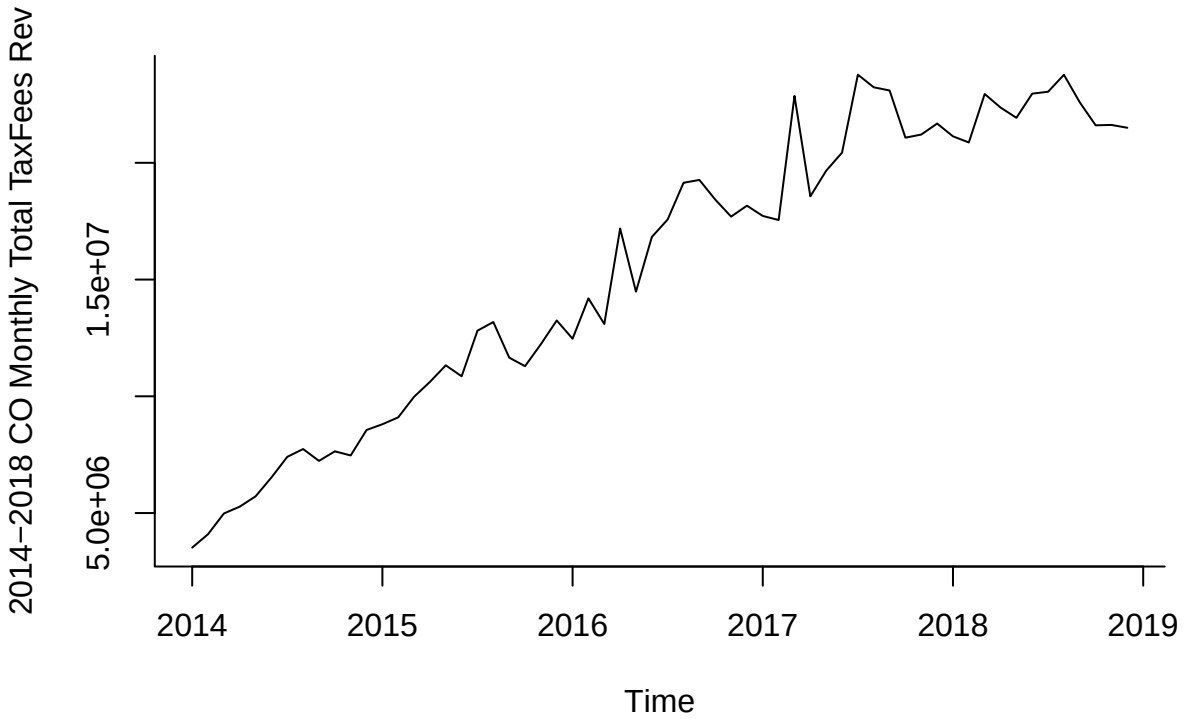
Is there more data out there that I just have not found yet that might give me a better and more up to date dataset for the questions I am posing?

Findings

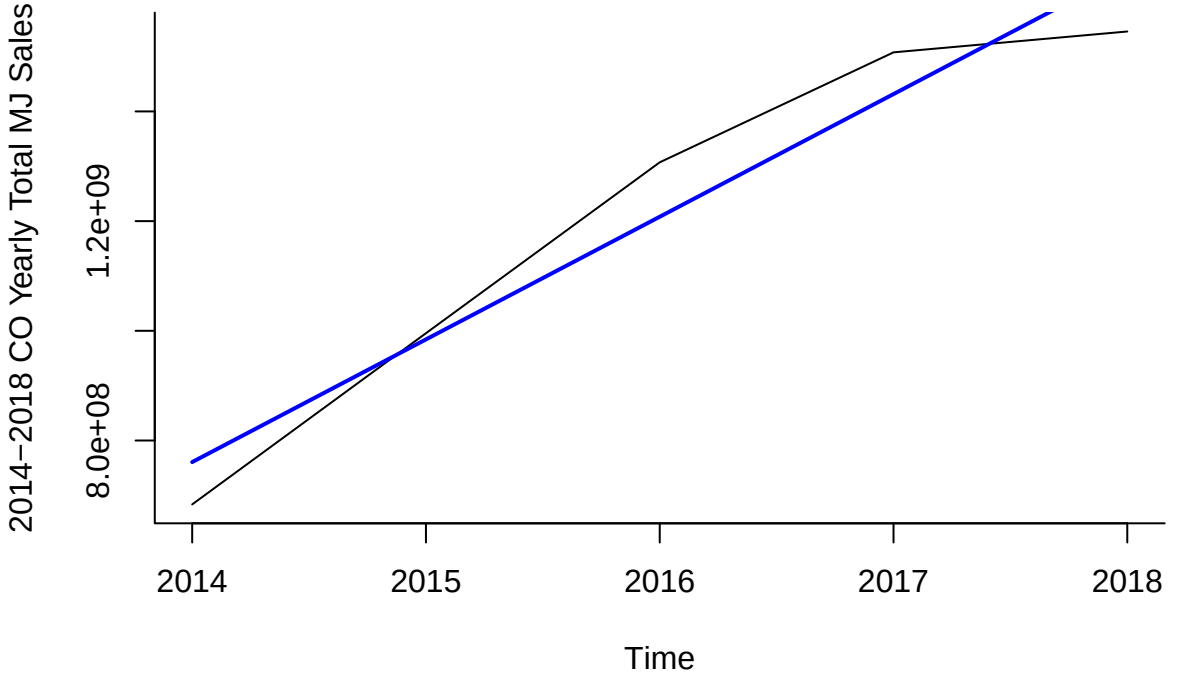
```
# Exploratory Analysis CO 2014-2019 marijuana sales & tax revenue
cosalesrev <- read.csv("data/cleaned_COMJSalesMonthReports2019.csv")
cotaxrev <- read.csv("data/cleaned_COMJTaxMonthReports2019.csv")
# Visualize CO 2014-2019 monthly total marijuana sales as a time series
cosalesrev.ts <- ts(cosalesrev$TotalSales, start = c(2014,1), end = c(2018,12), freq = 12)
plot(cosalesrev.ts, xlab = "Time", ylab = "2014-2018 CO Monthly Total MJ Sales", bty = "l")
```



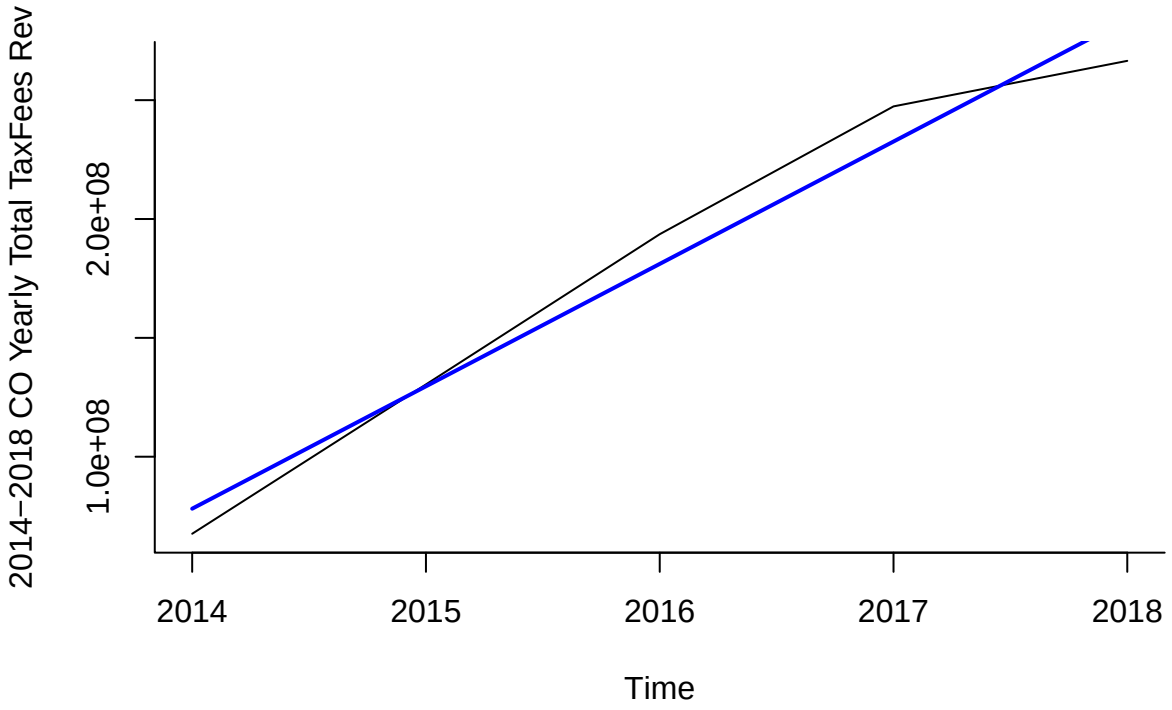

```
# Visualize CO 2014-2019 monthly total tax fees revenue as a time series
cotaxrev.ts <- ts(cotaxrev$TotalTaxFees, start = c(2014,1), end = c(2018,12), freq = 12)
plot(cotaxrev.ts, xlab = "Time", ylab = "2014-2018 CO Monthly Total TaxFees Rev", bty = "l")
```



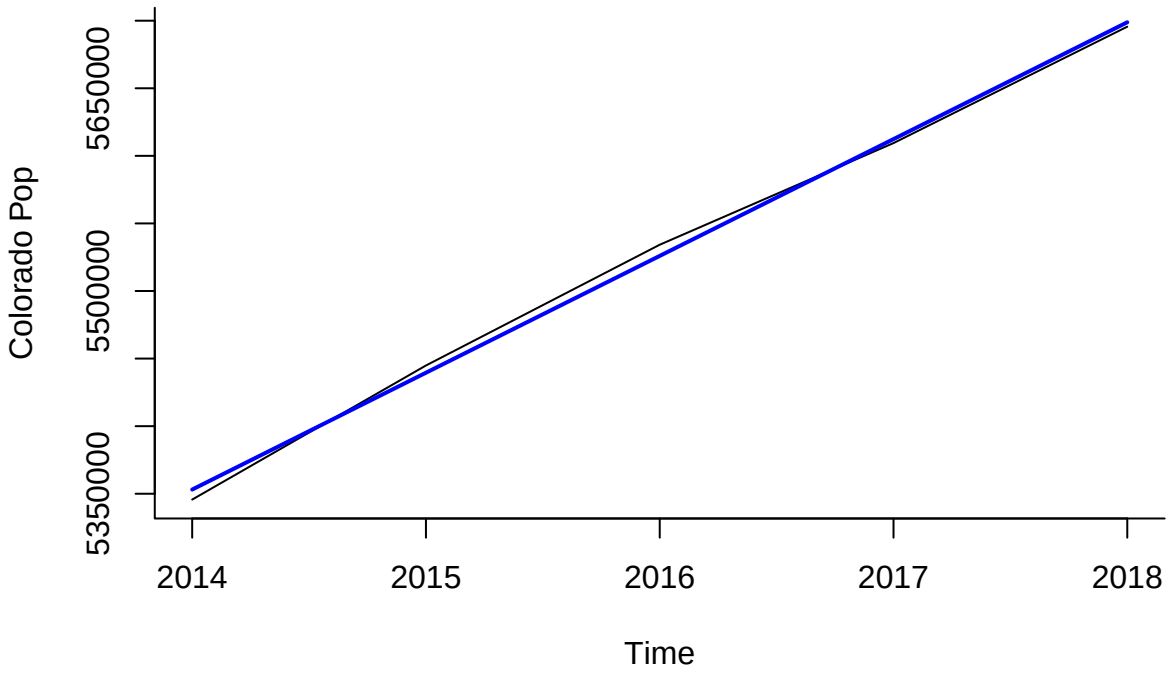

```
# Exploratory Analysis CO 2014-2018 marijuana sales & tax summaries
cosalessummary <- read.csv("data/cleaned_COMJSalesSummary2019.csv")
cotaxsummary <- read.csv("data/cleaned_COMJTaxSummary2019.csv")
# Visualize CO 2014-2018 total marijuana sales summary by years as time series
cosalessummary.ts <- ts(cosalessummary$TotalSales, start = c(2014), end = c(2018), freq = 1)
cosalessummary.lm <- tslm(cosalessummary.ts ~ trend)
plot(cosalessummary.ts, xlab = "Time", ylab = "2014-2018 CO Yearly Total MJ Sales", bty = "l")
lines(cosalessummary.lm$fitted, col="blue", lwd = 2)
```



```r
# Visualize CO 2014-2019 monthly tax summary by years as a time series
cotaxsummary.ts <- ts(cotaxsummary$TotalTaxFees, start = c(2014), end = c(2018), freq = 1)
cotaxsummary.lm <- tslm(cotaxsummary.ts ~ trend)
plot(cotaxsummary.ts, xlab = "Time", ylab = "2014-2018 CO Yearly Total TaxFees Rev", bty = "l")
lines(cotaxsummary.lm$fitted, col="blue", lwd = 2)
```



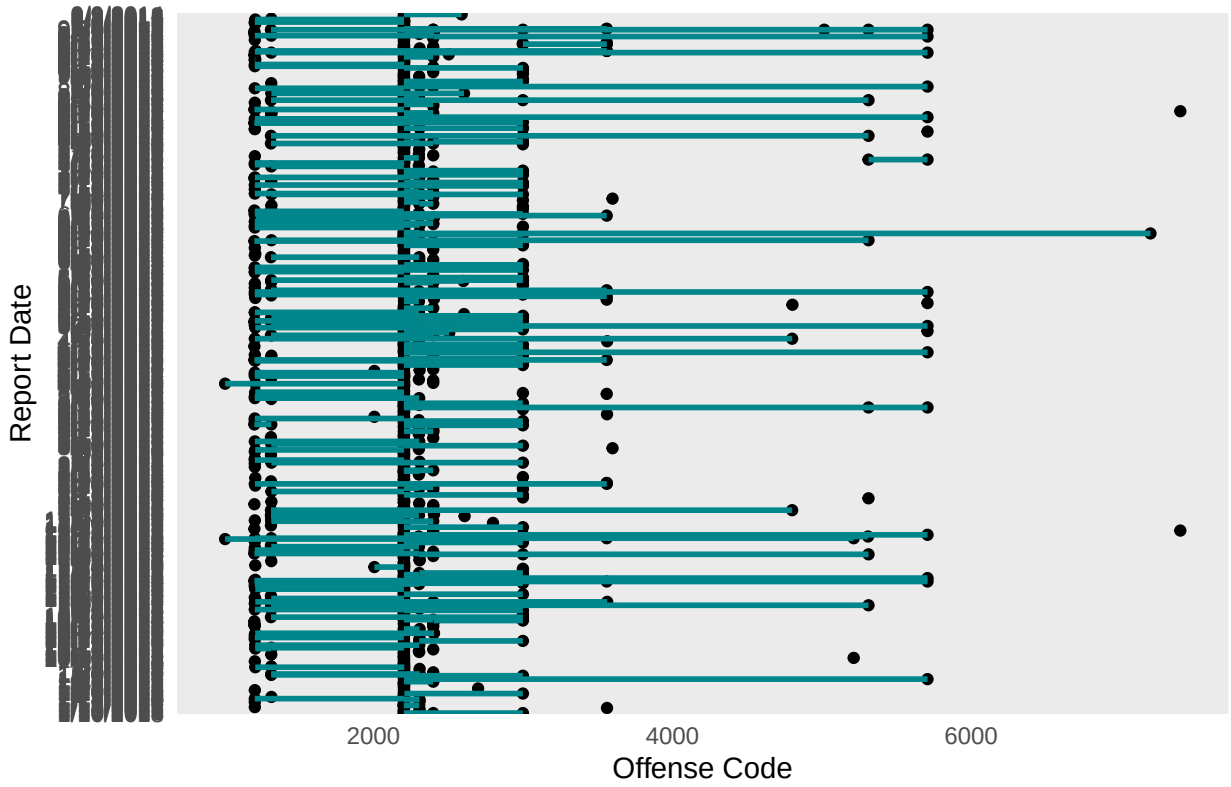

```r
# Exploratory Analysis CO 2014-2018 population estimates
costatepop <- read.csv("data/cleaned_COCountyPop20142018.csv")
# Calculate 2014-2017 CO state pop from all CO counties
Year <- c(2014,2015,2016,2017, 2018)
Colorado <- c((sum(costatepop$X2014Pop)),(sum(costatepop$X2015Pop)), (sum(costatepop$X2016Pop)), (sum(c
# Create a 2014-2017 Colorado state pop df
costatepop.df <- data.frame(Year,Colorado)
# Visualize CO pop as a time series
costatepop.ts <- ts(costatepop.df$Colorado, start = c(2014), end = c(2018), freq = 1)
# The Linear trend line on CO Pop time series
costatepop.lm <- tslm(costatepop.ts ~ trend)
plot(costatepop.ts, xlab = "Time", ylab = "Colorado Pop", bty = "l")
lines(costatepop.lm$fitted, col="blue", lwd = 2)
```




```
library(readxl)
library(ggplot2)

crime_denver <- read.csv("data/cleaned_crime_marijuana.csv")
crime_denver[is.na(crime_denver) | crime_denver=="Inf"] = NA
crime_stat = lm(OFFENSE_CODE~REPORTDATE, data = crime_denver)
ggplot(crime_stat, aes(OFFENSE_CODE,REPORTDATE))+
  geom_point() +
  geom_smooth(method='lm', se=FALSE, color='turquoise4') +
  theme_minimal() +
  labs(x='Offense Code', y = 'Report Date' , title = 'Linear Regression model of Crime in Denver') +
  theme(plot.title = element_text(hjust=0.5, size=20, face='bold'))
```

```
## `geom_smooth()` using formula 'y ~ x'
```

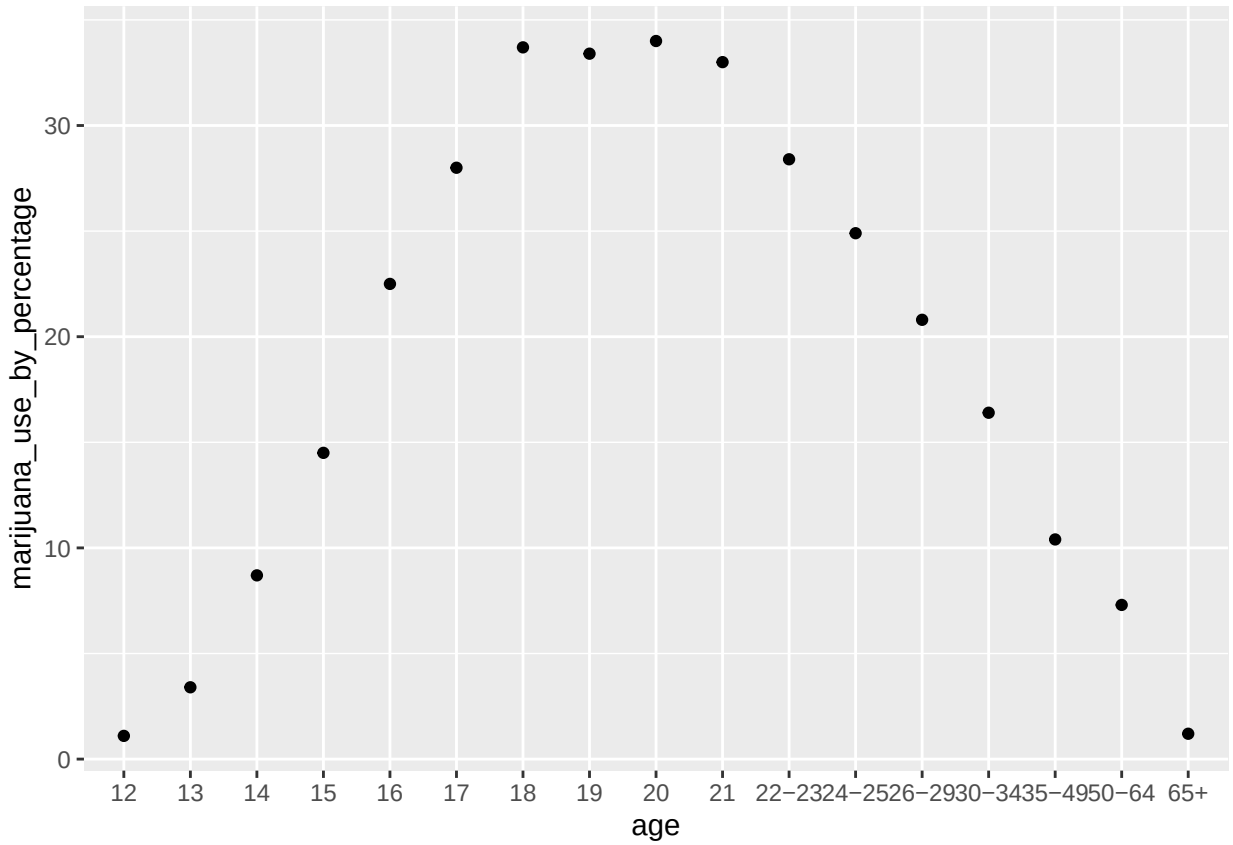



```
cleaned_drug_use <- read.csv("data/cleaned_drug-use-by-age.csv")
summary(cleaned_drug_use)
```

```
##      age             sample_size   marijuana_use_by_percentage
##  Length:17          Min.   :2223   Min.   : 1.10
##  Class :character   1st Qu.:2469   1st Qu.: 8.70
##  Mode  :character   Median :2798   Median :20.80
##                     Mean   :3251   Mean   :18.92
##                     3rd Qu.:3058   3rd Qu.:28.40
##                     Max.   :7391   Max.   :34.00
##  marijuana_frequency_over_12_months
##  Min.   : 4.00
##  1st Qu.:30.00
##  Median :52.00
##  Mean   :42.94
##  3rd Qu.:52.00
##  Max.   :72.00
```

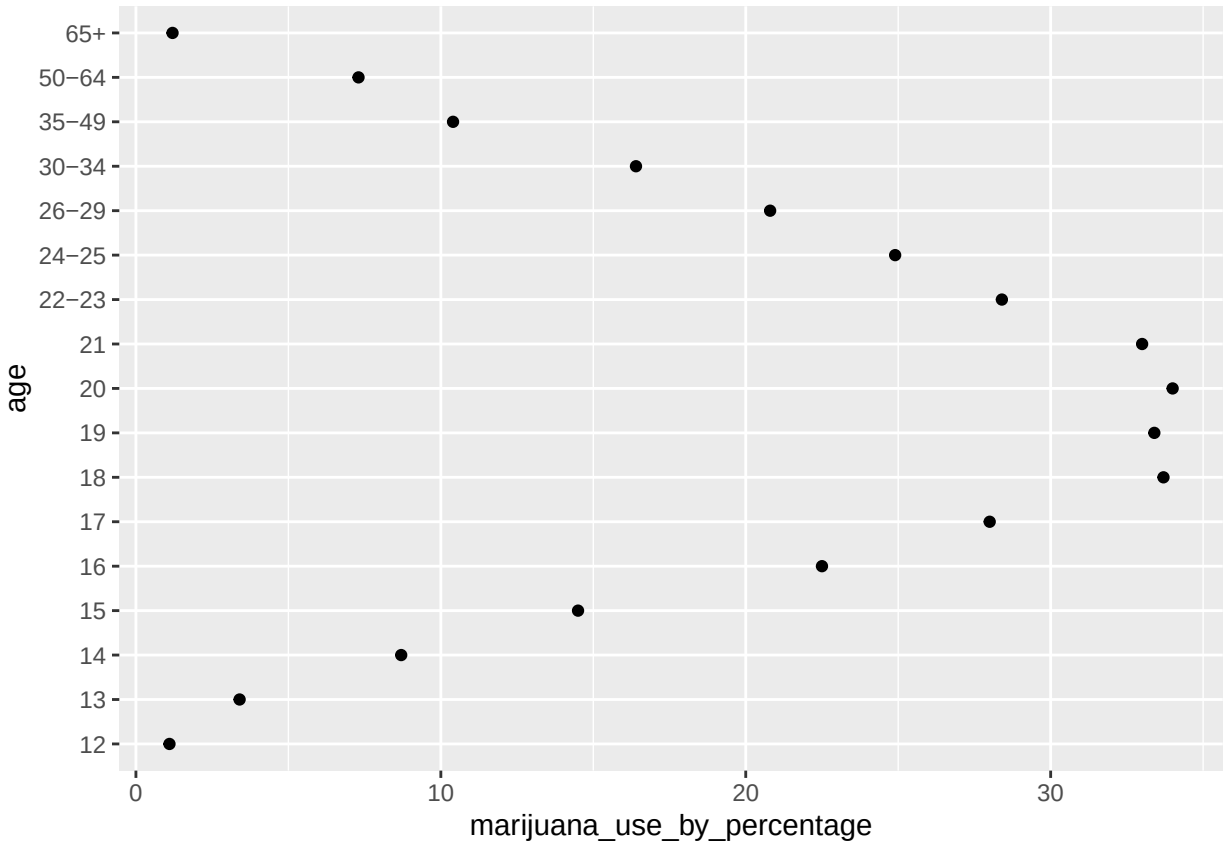
```
ggplot(cleaned_drug_use, aes(x=age, y=marijuana_use_by_percentage)) + geom_point() +  geom_smooth(metho
```

```
## `geom_smooth()` using formula 'y ~ x'
```



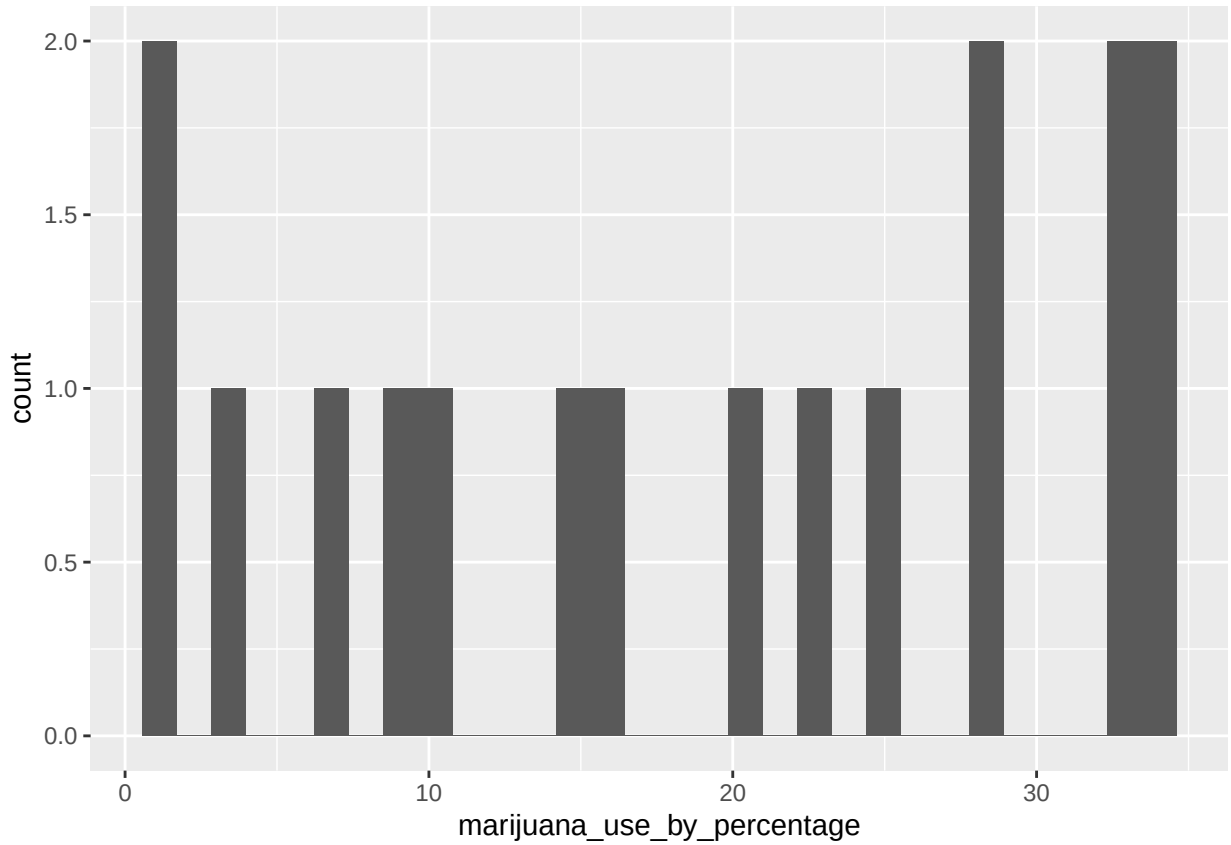

```
ggplot(cleaned_drug_use, aes(x=marijuana_use_by_percentage, y=age)) + geom_point() +  geom_smooth(metho
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(cleaned_drug_use, aes(x=marijuana_use_by_percentage)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
cleaned_support <- read.csv("data/cleaned_legal_marijuana_support.csv")
summary(cleaned_support)
```

```
##       Year         Month           Asked_half_sample    Yes_Legal
##  Min.   :1969   Length:20          Length:20          Min.   :12.00
##  1st Qu.:1980   Class :character   Class :character   1st Qu.:25.00
##  Median :2002   Mode  :character   Mode  :character   Median :34.00
##  Mean   :1997                                         Mean   :35.95
##  3rd Qu.:2011                                         3rd Qu.:48.50
##  Max.   :2016                                         Max.   :60.00
##    No_Illegal      No_Opinion   Percent_Yes         Percent_No
##  Min.   :39.00   Min.   :1.00   Length:20          Length:20
##  1st Qu.:49.25   1st Qu.:2.00   Class :character   Class :character
##  Median :63.00   Median :4.00   Mode  :character   Mode  :character
##  Mean   :60.50   Mean   :3.45
##  3rd Qu.:70.75   3rd Qu.:4.25
##  Max.   :84.00   Max.   :6.00
##  Percent_No_Opinion
##  Length:20
##  Class :character
##  Mode  :character
##
##
##
```

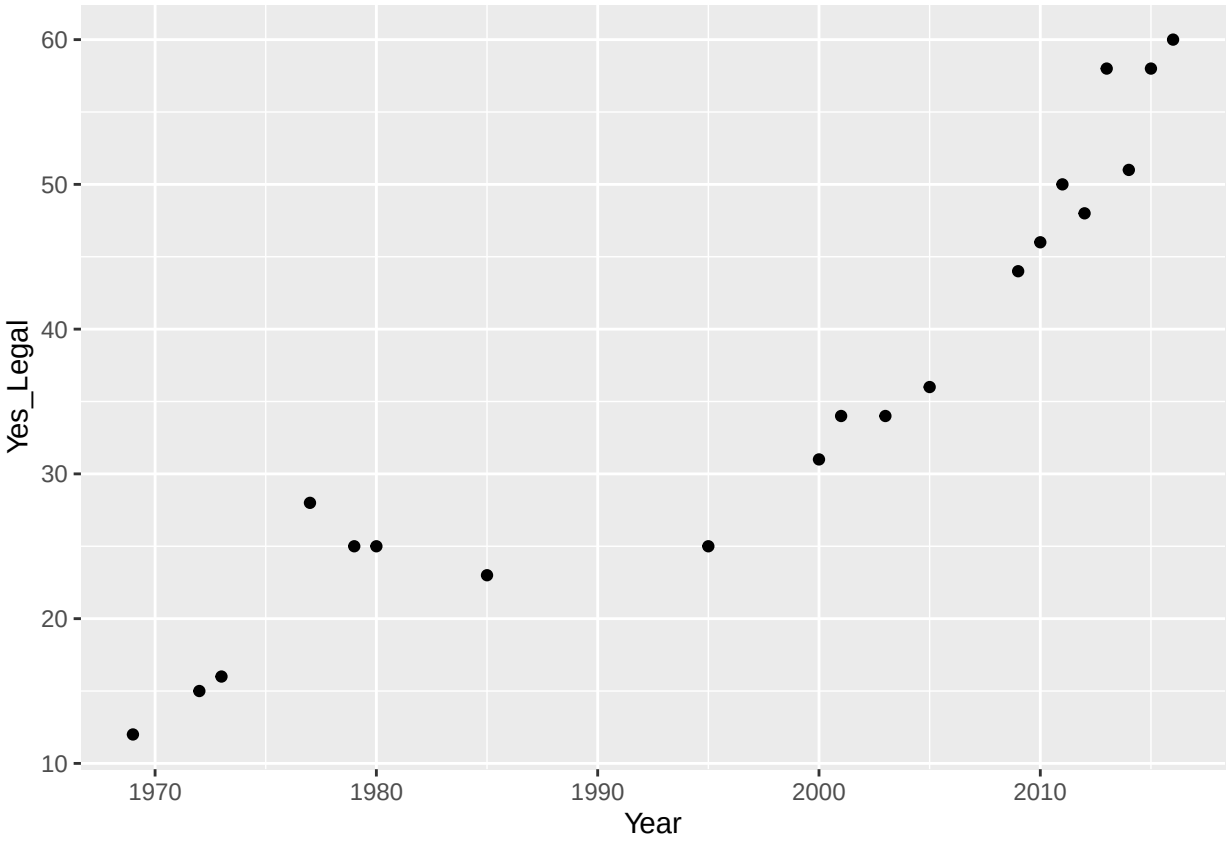
```
head(cleaned_support)
```

```
##   Year Month Asked_half_sample Yes_Legal No_Illegal No_Opinion Percent_Yes
```

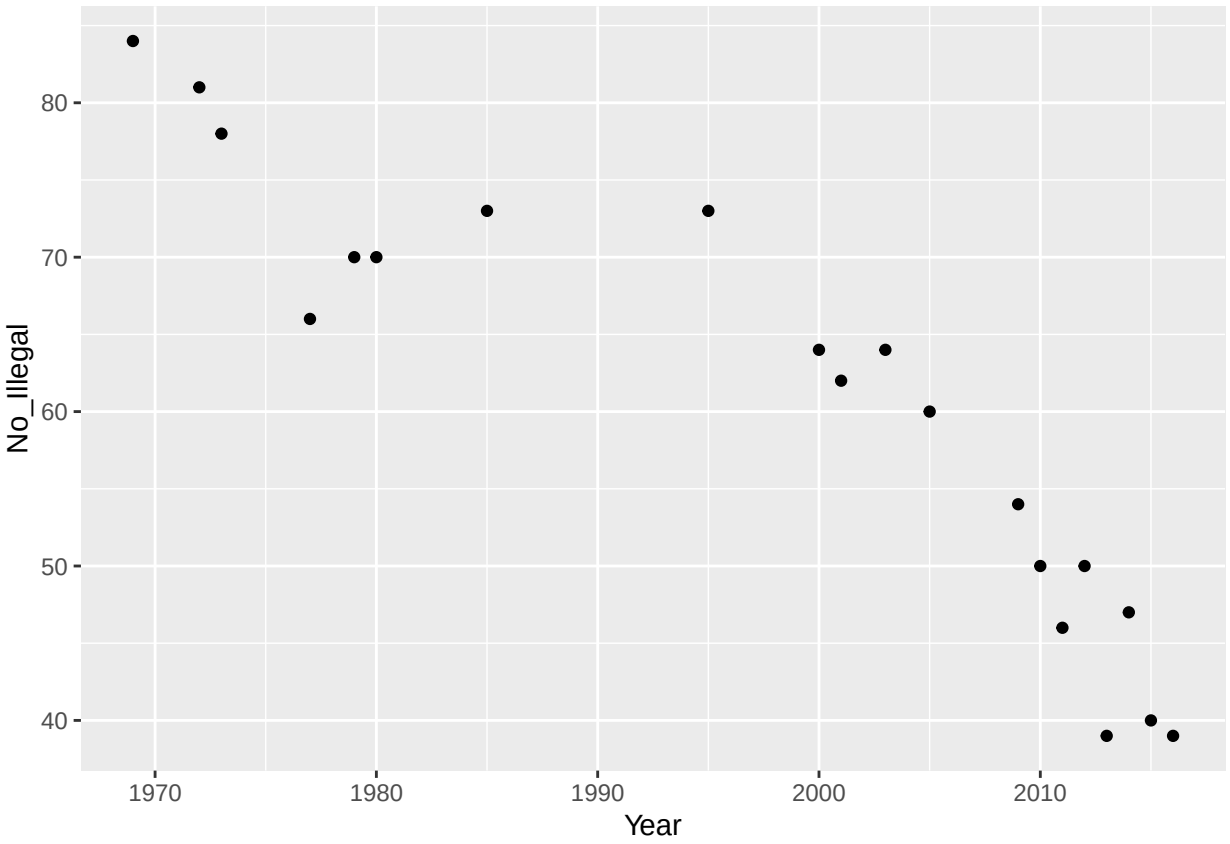
```
## 1 2016   Oct                no         60         39          1         60%
## 2 2015   Oct                no         58         40          2         58%
## 3 2014   Oct                no         51         47          2         51%
## 4 2013   Oct                no         58         39          3         58%
## 5 2012   Nov                no         48         50          1         48%
## 6 2011   Oct                no         50         46          3         51%
##   Percent_No Percent_No_Opinion
## 1        39%                 1%
## 2        40%                 2%
## 3        47%                 2%
## 4        39%                 3%
## 5        51%                 1%
## 6        46%                 3%
```

```
library(patchwork)
library(dplyr)

ggplot(cleaned_support, aes(y=Yes_Legal, x=Year)) + geom_point()
```




```
ggplot(cleaned_support, aes(y=No_Illegal, x=Year)) + geom_point()
```



Step 3

Introduction

With more and more states adding medical and recreational cannabis to their ballots, questions are now being asked more openly about the impact of cannabis on public health, crime statistics, and popularity of legalization. The data out there is sparse, but the data that has been collected could show trends and insights on the future of legislation and public sentiment of legalization, along with showing the potential for lowered crime and raised tax revenue in states and at the federal level.

The data collected is attempting to find out if the sale and legality of cannabis shows an increase in crime and use in teens. The data collected shows the sales of cannabis based on annual sales since its legalization recreationally in 2014 in the state of Colorado, and the statistics of crime in the city of Denver. Data that did not support the investigation was scrubbed from the datasets and a new aggregate dataset was formed using the data that best fits the model example we are trying to follow. Results are interesting after running the models and building out visualizations. Some models were not as clean as I had hoped to be, but the linear regresssion lines did often model predictions accurrately with only a few outliers.

The problem statement

The data required to answer the questions was not only sparse but also very difficult to parse through. The data in and of itself did not always correlate well with the questions being asked, along with the age old statement of causation does not equal coorelation often occuring. The issues that seemed to be the biggest is that the legalization of recreational marijuana in the state of Denver was the best kept up to date datasets but also tended to end the reporting at 2019. I

was not able to find existing data that continued on past a certian point, nor was I able to find out how well the data was kept up. The data that I was able to find also delved into topics and subjects that was outside of scope for my initial line of questioning.

How I addressed the Problem Statement

After cleaning the data as best as possible, piecing together ways to at least make a connection between datasets was the primary focus. It was difficult to parse through everything as the data was not initially related to each other and thus the data did not always play nice with the one another. Metrics also were an issue, but isolating and asking more specific questions made it easier to form the right philosophies and make it easier to understand the data.

Analysis

From looking at the data, there seems to be little difference in crime related to the legaliziation of marijuana except for 22 instances from 2012 - 2016 and nearly all offenses were Possession, cultivation, or Sales charges. The data suggests that the most of the crimes were sales. The law in Denver states taht the sale of more than 6 ounces – 2.5 pounds of marijuana concentrate is a level 3 drug felony punishable by 2 – 6 years imprisonment as well as a fine between $5,000-$500,000. The sale of more than 2.5 – 25 pounds is a level 2 drug felony punishable by a sentence of 4 – 16 years and a fine of $ 3,000 – $ 750,000. The stats found did not however indicate what statute was broken nor did it display what the offense was other than a category_id and a offense code.

Another interesting finding is that the data suggests that public opinion for legalization had a converse relationship sometime between 1990-2000 (most likely sometime around 1995). The only other highest point was sometime in the late 70's for a yes to legalization but was still nowhere near the amount of support in 1995. What happened at that time to change public perception is most likelly many factors that the data could not display nor suggest as to the reason why.

Also something to look at was that from a usage scenario by age, 18-21 year olds were the group most likely to use marijuana frequently. It tapers off the older one gets by average after the age of 21, while it tapers up at 12 years of age and use is as frequent as the 65 and older group.

Limitations

The largest limitation I found with these data sets was that the data was large for the crime statistics but also was in a small area (Denver Area) and only covered crime from 2014-2016, where it accounted for all the crime reported such as car theft, homicide, and vandalism, etc.. but crime after legalization for possession and distrubition was lower and also there was not enough detail to describe the instances of why the crime was a crime in the first place. When looking at the datasets for how much revenue was produced, that dataset was the most informative, but also did not show much reason as to why crime and usage either increased or decreased based on the number of sales. The limitations that really stood out where that the data was not directly related as reports, and that all the data only focused on a specific angle of the questions that I was asking.

Concluding Remarks

More research must be done with a broader net when gathering information on the impact of recreational sales of Cannabis and crime, usage by users. The research sets were interesting and the topic itself is an interesting question to see if it can be asked and solved. I believe that the data is out there currently, but aggregation of the data is sparce and other data sets seem to be focusing specifically on a broad range question analysis such as "How many people are using X Drug (cocaine, Methamphetamine, Marijuana, Alcohol, etc. . .)" than how many people are using Marijuana and commiting crimes that is more severe than a drug related offense.

I think though with the amount of research available that it would be hard to really say with confidence that the recreational sales of marijuana has done anything good or bad for the crime rate/ useage by people, as a model that could be used to predict the crime rate and usage in a newly legal state, but I am confident with the research available that there is at least an implication that with the legalization of Cannabis in a new state, crime related to Cannabis would lower, and usage might go up for those college age, meanwhile seeing a boom in financial resources collected by the state.