A Mapping of Denver Marijuana Businesses and Arrests

Author: David Martinez Institution: Regis University CLass: MSDS692 Data Science Practicum #1 Professor: Dr. Michael Busch

Introduction

On November 7th, 2000, voters in Colorado amended the state Constitution to allow the sale and use of Marijuana upon written consent by medical professional (Amendment 20, Colorado Constitution, 2000). Twelve years later, Colorado approved the sale and use of marijuana recreational use for adults over the age of twenty-one (Amendment 64, COlorado Constitution, 2012). Denver, the state capital and the largest population center in Colorado, has published data records since 2010 for medical marijuana and 2013 for recreational marijuana, including sales, government revenue, licensing information, and crime statistics.

This project will focus on identifying the types and locations of Marijuana businesses as well as the types and locations of arrests made.

The following datasets are used in this project: https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-marijuana-active-business-licenses https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime-marijuana

Other references: https://ballotpedia.org/Marijuana_on_the_ballot https://developers.google.com/maps/documentation/geocoding/usage-and-billing

citations: ggmap - D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

stringr - Hadley Wickham (2017). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.2.0. https://CRAN.R-project.org/package=stringr

ggplot2 - H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

dplyr - Hadley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.4. https://CRAN.R-project.org/package=dplyr

RgoogleMaps - Markus Loecher and Karl Ropkins (2015). RgoogleMaps and loa: Unleashing R Graphics Power on Map Tiles. Journal of Statistical Software 63(4), 1-18. URL http://www.jstatsoft.org/v63/i04/.

Retrieve Denver Marijuana Licenses

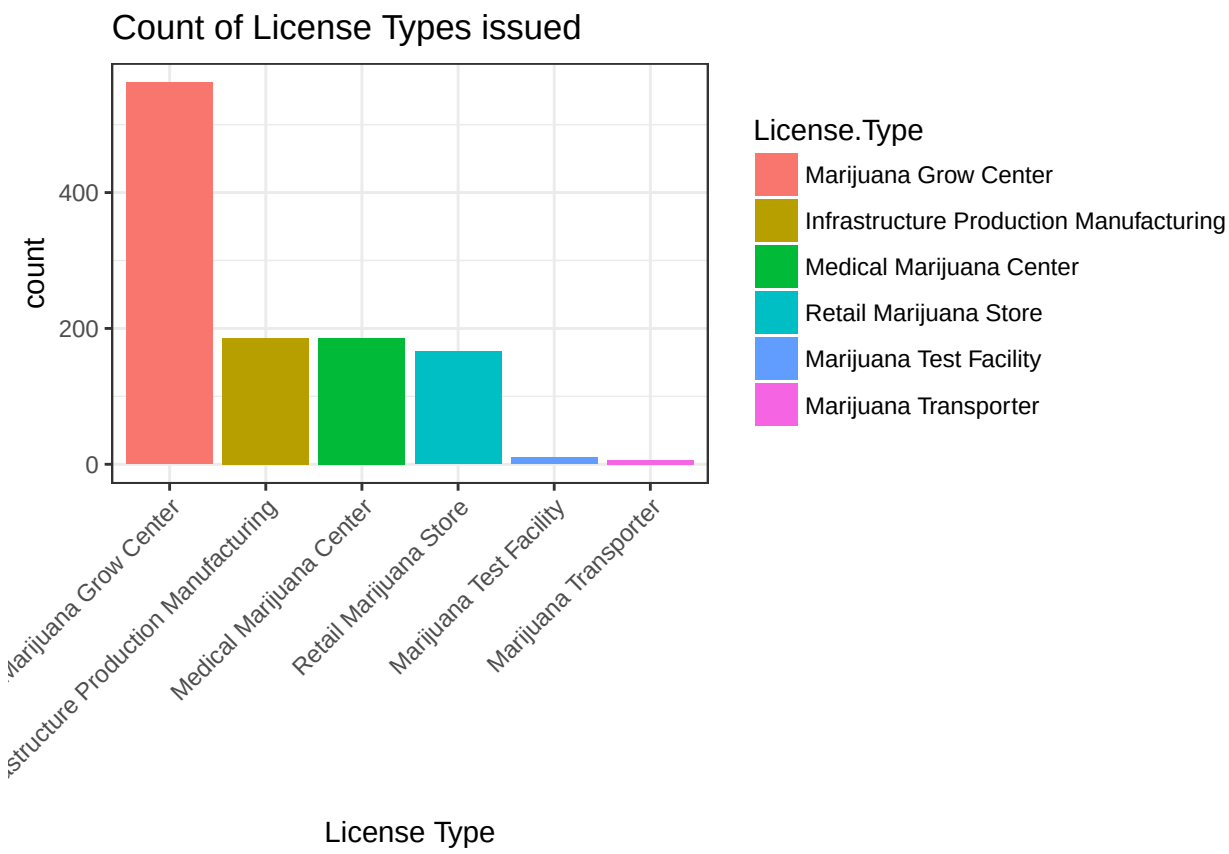
```
library(stringr)  #string operations
library(ggplot2)  #graphics
library(ggmap)    #used for geocoding
library(dplyr)    #dataframe manipulation

#retrieve dataset from denvergov.org
denver_mj_licenses <- read.csv("https://www.denvergov.org/media/gis/DataCatalog/marijuana_active_busines

#create factor for license type
denver_mj_licenses$License.Type <- factor(denver_mj_licenses$License.Type)

#combine levels and rename for easier reading
denver_mj_licenses$License.Type <- plyr::revalue(denver_mj_licenses$License.Type, c("Med Marijuana Inf

#create factor for zip code
denver_mj_licenses$Facility.Zip.Code <- factor(denver_mj_licenses$Facility.Zip.Code)
```

```r
#current license status and expiration date don't appear to be very interesting so they can go
denver_mj_licenses <- denver_mj_licenses[-c(5:6)]

#convert street address information to one field for geocoding
num <- paste(word(denver_mj_licenses$Facility.Street.Number))
dir <- paste(word(denver_mj_licenses$Facility.Pre.Direction))
street <- paste(word(denver_mj_licenses$Facility.Street.Name))
type <- paste(word(denver_mj_licenses$Facility.Street.Type))
denver_mj_licenses$ADDRESS <- paste(num, dir, street, type, ", DENVER, CO", sep=" ")

#remove old address information (except zip code)
denver_mj_licenses <- denver_mj_licenses[-c(5:10)]

#geocode for lat/long - Data Science Toolkit (dsk) is used here instead of to retrieve the lat/long
for(i in 1:nrow(denver_mj_licenses)) {
  result <- geocode(denver_mj_licenses$ADDRESS[i], output="latlona", source="dsk")
  denver_mj_licenses$LONGITUDE[i] <- as.numeric(result[1])
  denver_mj_licenses$LATITUDE[i] <- as.numeric(result[2])
}

#save for posterity
#write.csv(denver_mj_licenses, "mj_licenses_geocoded.csv", row.names=FALSE)

summary(denver_mj_licenses)
```

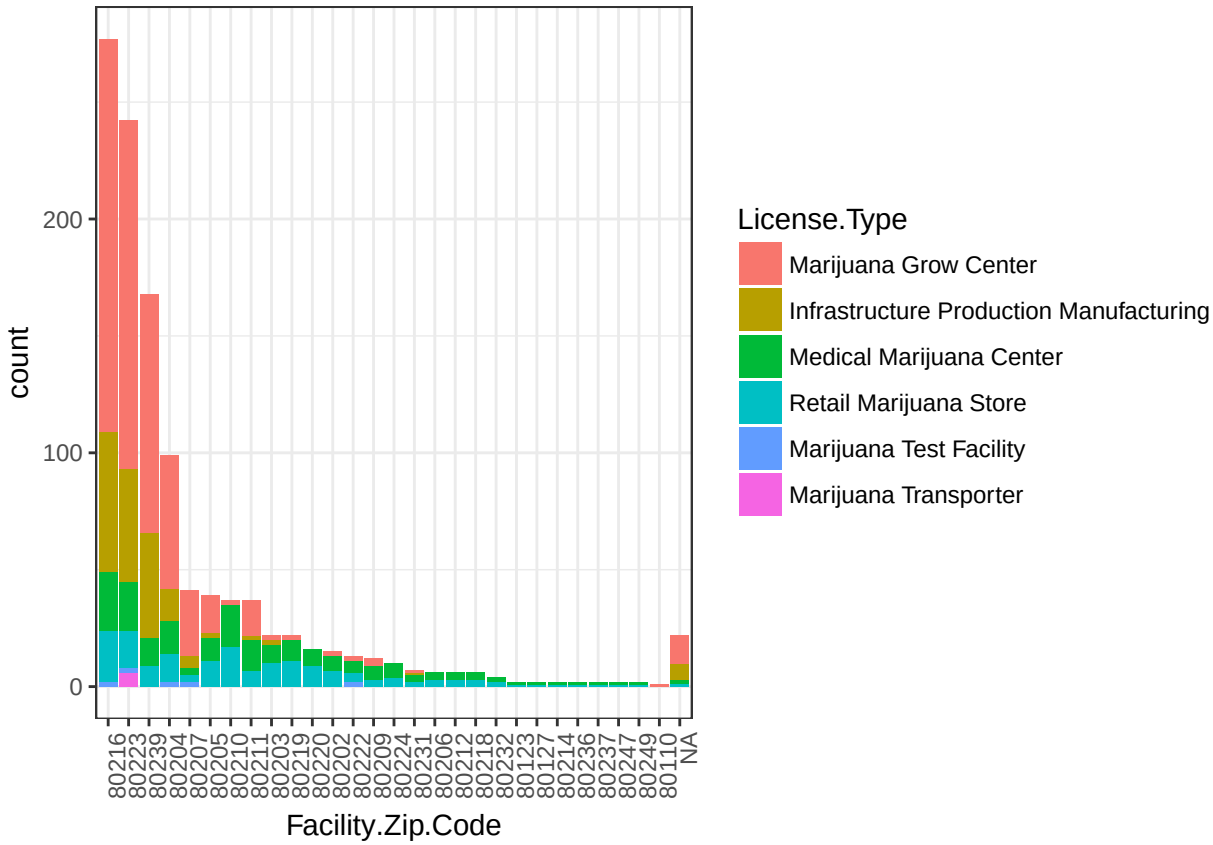
```
##  Business.File.Number                                    License.Type
##  Length:1116          Marijuana Transporter                    :  6
##  Class :character     Infrastructure Production Manufacturing:186
##  Mode  :character     Marijuana Grow Center                    :562
##                       Marijuana Test Facility                  : 10
##                       Medical Marijuana Center                 :186
##                       Retail Marijuana Store                   :166
##
##  Entity.Name         Trade.Name        Facility.Zip.Code
##  Length:1116        Length:1116        80216  :277
##  Class :character   Class :character   80223  :242
##  Mode  :character   Mode  :character   80239  :168
##                                        80204  : 99
##                                        80207  : 41
##                                        (Other):267
##                                        NA's   : 22
##    ADDRESS            LONGITUDE         LATITUDE
##  Length:1116        Min.   :-105.1   Min.   :39.63
##  Class :character   1st Qu.:-105.0   1st Qu.:39.68
##  Mode  :character   Median :-105.0   Median :39.71
##                     Mean   :-105.0   Mean   :39.72
##                     3rd Qu.:-104.9   3rd Qu.:39.77
##                     Max.   :-104.8   Max.   :39.92
##
```

```r
#cleanup again - darn OCD
rm(i, num, dir, street, type, result)

#sort by License.Type
```

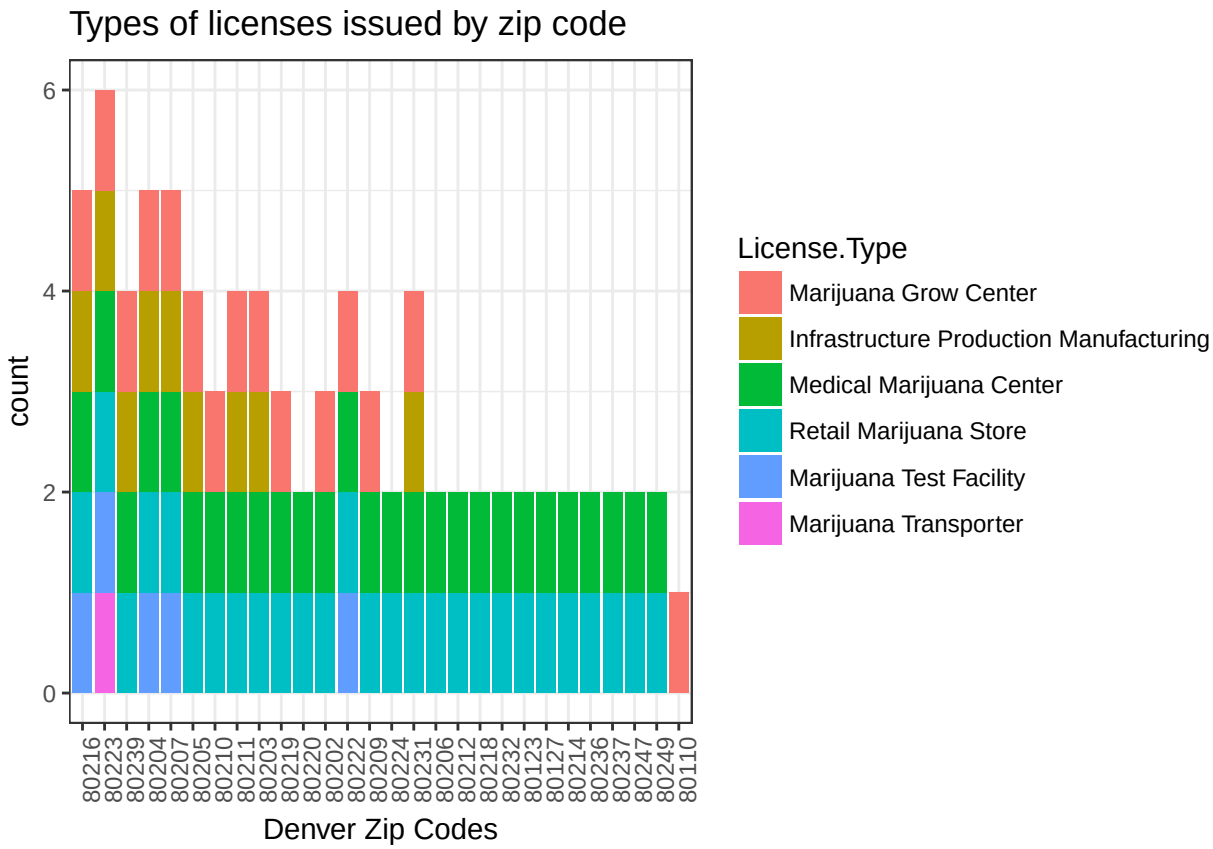
```
denver_mj_licenses <- within(denver_mj_licenses, License.Type <- factor(License.Type,
                                                                        levels=names(sort(table(License

#generate plot to show graphically the number of license types issued
ggplot(denver_mj_licenses,aes(x=License.Type, fill=License.Type))+
  geom_bar()+
  labs(title="Count of License Types issued", x= "License Type")+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#sort by Facility Zip Code
denver_mj_licenses <- within(denver_mj_licenses, Facility.Zip.Code <- factor(Facility.Zip.Code,
                                                                              levels=names(sort(table(Facility.

#generate plot to show where licenses are being issued
ggplot(denver_mj_licenses,aes(x=Facility.Zip.Code, fill=License.Type))+
  geom_bar()+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```
#create a dataframe of the types of licenses issued to each zip code
zip_tab <- as.data.frame(table(denver_mj_licenses$Facility.Zip.Code, denver_mj_licenses$License.Type))

#rename for more descriptive tags
zip_tab <- rename(zip_tab, ZipCode = Var1, License.Type = Var2)

#filter out the 0 frequency occurences
zip_tab <- filter(zip_tab, zip_tab$Freq > 0)

#generate a plot to show the types of licenses issued to each zip code
ggplot(zip_tab,aes(x=ZipCode, fill=License.Type))+
  geom_bar()+
  labs(title="Types of licenses issued by zip code", x= "Denver Zip Codes")+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Generating Maps

Clean and save marijuana crime dataset

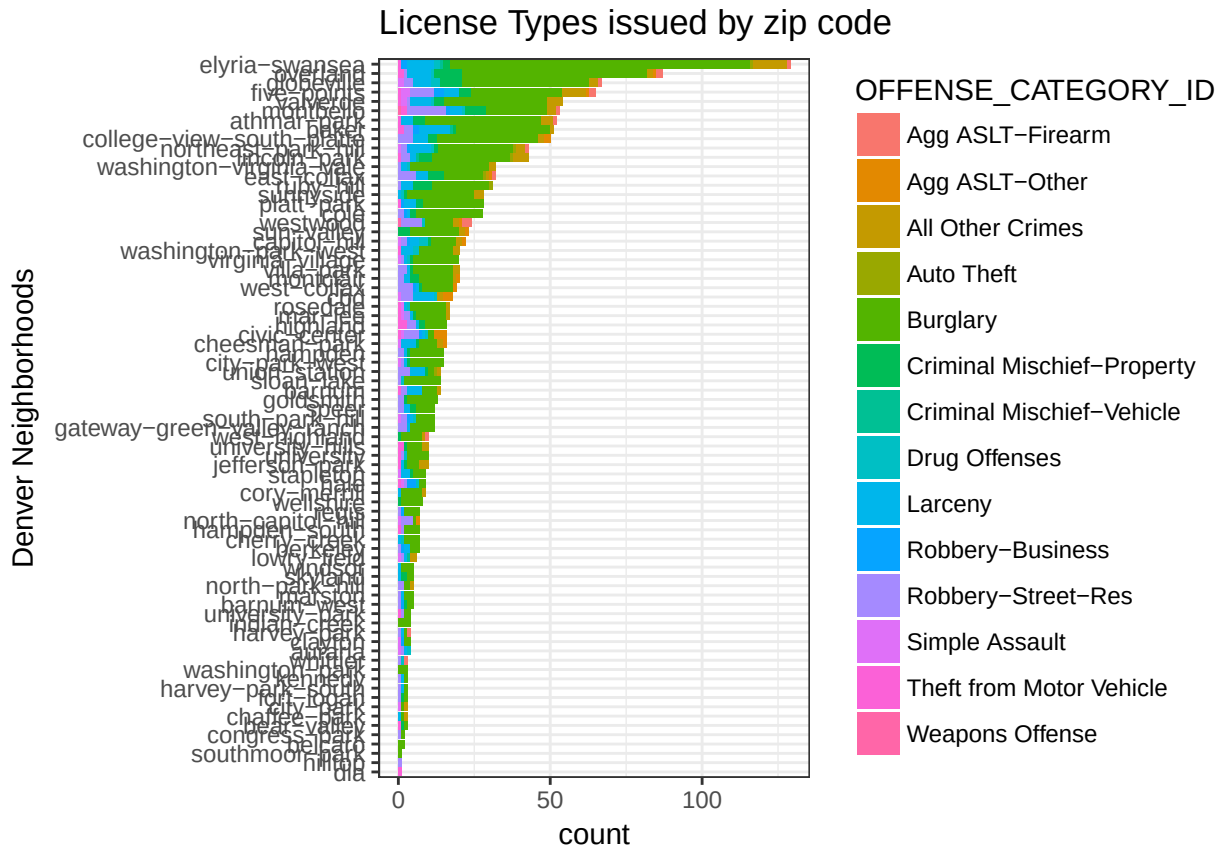
```
#retrieve the DPD marijuana crime file from Denver open data portal
denver_mj_crime <- read.csv("https://www.denvergov.org/media/gis/DataCatalog/crime_marijuana/csv/crime_

#write original file to disk for posterity
write.csv(denver_mj_crime, "crime_marijuana.csv", row.names=FALSE)

#incident ID to character
denver_mj_crime$INCIDENT_ID <- as.character(denver_mj_crime$INCIDENT_ID)

#Only one date field is needed
denver_mj_crime <- data.frame(denver_mj_crime[-c(2:3, 6:7, 10:11)])

#The Report date needs to be in an R compliant format
day <- paste(word(denver_mj_crime$REPORTDATE, 1, sep="-"))
month <- paste(word(denver_mj_crime$REPORTDATE, 2, sep="-"))
month <- plyr::revalue(month, c("JAN"="01", "FEB"="02", "MAR"="03", "APR"="04", "MAY"="05", "JUN"="06",
year <- as.integer(paste(word(denver_mj_crime$REPORTDATE, -1, sep="-")))
year <- paste(year+2000)

#denver_mj_crime$REPORTDATE <- paste(year, month, sep="-")
denver_mj_crime$REPORTDATE <- paste(year, month, day, sep="-")
denver_mj_crime$REPORTDATE <- as.Date(denver_mj_crime$REPORTDATE)
```

```
#append city and state information to address - necessary for geocoding
denver_mj_crime$INCIDENT_ADDRESS <- sapply(denver_mj_crime$INCIDENT_ADDRESS , paste, ", Denver, CO", sep

#geo_x, geo_y are not needed for this activity
#denver_mj_crime <- data.frame(denver_mj_crime[-c(4:5)])

#convert fields to factors
denver_mj_crime$DISTRICT_ID <- factor(denver_mj_crime$DISTRICT_ID)
denver_mj_crime$PRECINCT_ID <- factor(denver_mj_crime$PRECINCT_ID)
denver_mj_crime$OFFENSE_CATEGORY_ID <- factor(denver_mj_crime$OFFENSE_CATEGORY_ID)
denver_mj_crime$MJ_RELATION_TYPE <- factor(denver_mj_crime$MJ_RELATION_TYPE)
denver_mj_crime$NEIGHBORHOOD_ID <- factor(denver_mj_crime$NEIGHBORHOOD_ID)

#geocode to add Longitude/Latitude data
for(i in 1:nrow(denver_mj_crime)) {
  result <- geocode(denver_mj_crime$INCIDENT_ADDRESS[i], output="latlona", source="dsk")
  denver_mj_crime$LONGITUDE[i] <- as.numeric(result[1])
  denver_mj_crime$LATITUDE[i] <- as.numeric(result[2])
}

summary(denver_mj_crime)
```

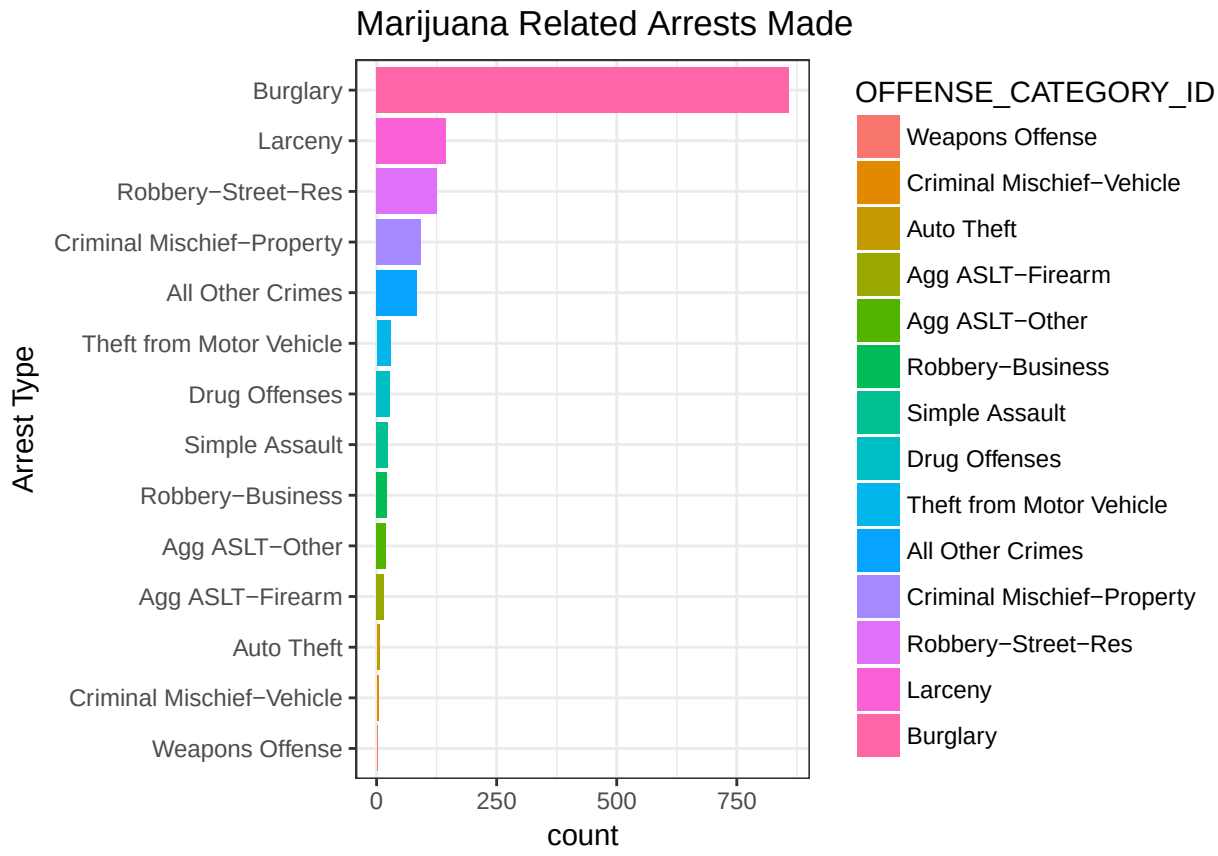
```
##  INCIDENT_ID          REPORTDATE         INCIDENT_ADDRESS    DISTRICT_ID
##  Length:1454        Min.   :2012-01-03   Length:1454        3      :375
##  Class :character   1st Qu.:2013-08-08   Class :character   2      :329
##  Mode  :character   Median :2014-12-18   Mode  :character   1      :272
##                     Mean   :2014-12-16                      4      :265
##                     3rd Qu.:2016-05-21                      6      :137
##                     Max.   :2017-12-17                      5      : 74
##                                                             (Other):  2
##   PRECINCT_ID                  OFFENSE_CATEGORY_ID      MJ_RELATION_TYPE
##  313    :145   Burglary                  :859      INDUSTRY\n    :1039
##  212    :130   Larceny                   :144      NON-INDUSTRY\n: 415
##  112    : 83   Robbery-Street-Res        :125
##  422    : 81   Criminal Mischief-Property: 92
##  412    : 76   All Other Crimes          : 84
##  411    : 73   Theft from Motor Vehicle  : 29
##  (Other):866   (Other)                   :121
##        NEIGHBORHOOD_ID   LONGITUDE         LATITUDE
##  elyria-swansea:129    Min.   :-105.1   Min.   :39.63
##  overland      : 87    1st Qu.:-105.0   1st Qu.:39.68
##  globeville    : 67    Median :-105.0   Median :39.71
##  five-points   : 65    Mean   :-105.0   Mean   :39.72
##  valverde      : 54    3rd Qu.:-104.9   3rd Qu.:39.76
##  montbello     : 53    Max.   :-104.7   Max.   :39.91
##  (Other)       :999
```

```
#remove intermediate variables and capture the clean file for posterity
rm(result, i, day, month, year)

#sort by neighborhood
denver_mj_crime <- within(denver_mj_crime, NEIGHBORHOOD_ID <- factor(NEIGHBORHOOD_ID,
                                                                     levels=names(sort(table(NEIGHBORH
```

```
#generate plot to show neighborhoods and arrests
ggplot(denver_mj_crime,aes(x=NEIGHBORHOOD_ID, fill=OFFENSE_CATEGORY_ID))+
  geom_bar()+
  coord_flip()+
  labs(title="License Types issued by zip code", x= "Denver Neighborhoods")+
  theme_bw()
```

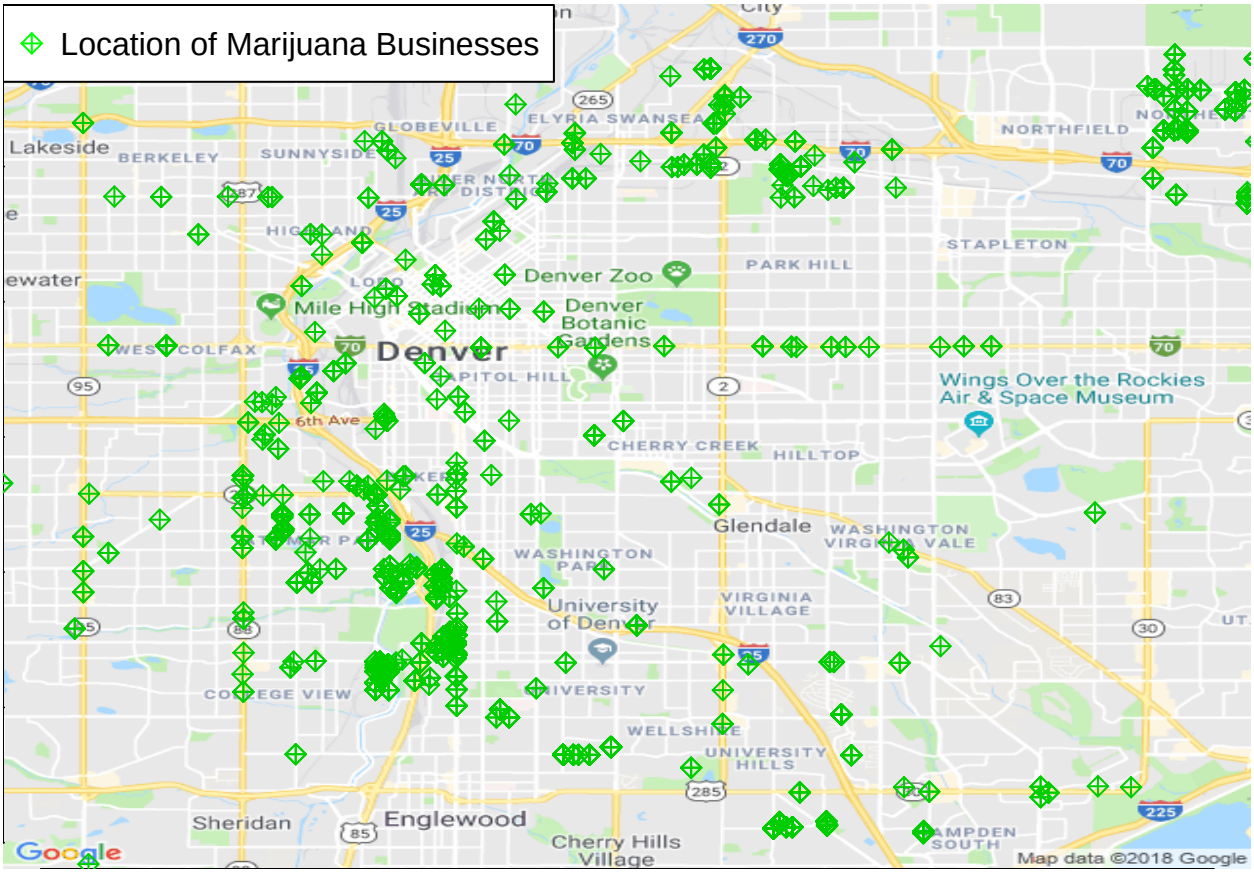



```
#sort by crime type
denver_mj_crime <- within(denver_mj_crime, OFFENSE_CATEGORY_ID <- factor(OFFENSE_CATEGORY_ID,
                                                                         levels=names(sort(table(OFFENSE_C

#generate plot to show where licenses are being issued
ggplot(denver_mj_crime,aes(x=OFFENSE_CATEGORY_ID, fill=OFFENSE_CATEGORY_ID))+
  geom_bar()+
  coord_flip()+
  labs(title="Marijuana Related Arrests Made", x= "Arrest Type")+
  theme_bw()
```

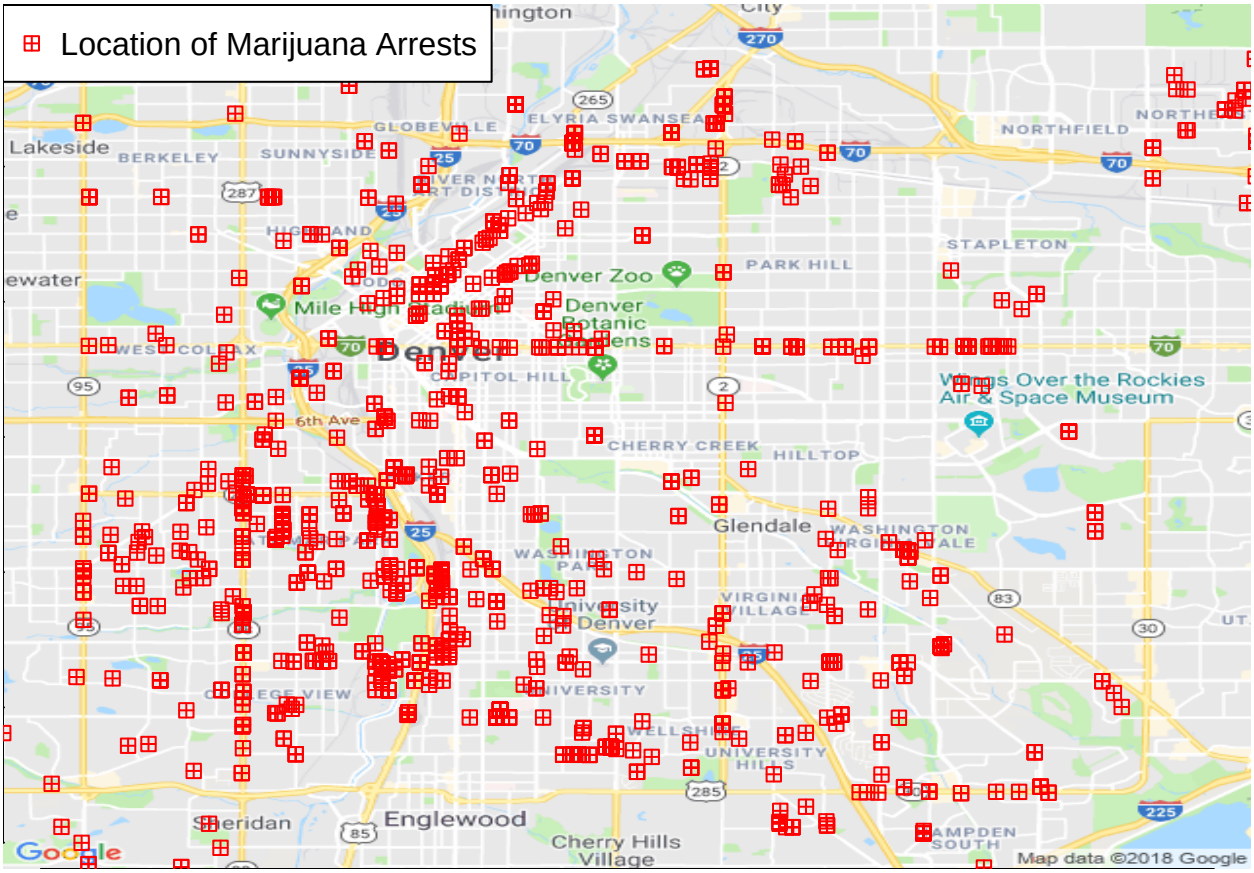


This section requires a Google developers API Key: (https://developers.google.com/maps/documentation/geocoding/usage-and-billing). For security purposes this section is masked.

```
library(RgoogleMaps) #interface to google maps

#API key
#apikey <- "your api key"

#generate denver map.
denver_map <- GetMap(center = c(lat = mean(denver_mj_licenses$LATITUDE), lon = mean(denver_mj_licenses$

p1 <- PlotOnStaticMap(denver_map, lat = denver_mj_licenses$LATITUDE, lon = denver_mj_licenses$LONGITUDE
legend("topleft", legend = "Location of Marijuana Businesses", col = "green", bg = "white", pch=9)
```




```
p2 <- PlotOnStaticMap(denver_map, lat = denver_mj_crime$LATITUDE, lon = denver_mj_crime$LONGITUDE, dest
legend("topleft", legend = "Location of Marijuana Arrests", col = "red", bg = "white", pch=12)
```


Location of Marijuana Arrests
Lakeside
BERKELEY
SUNNYSIDE
GLOBEVILLE
ELYRIA SWANSEA
NORTHFIELD
STAPLETON
PARK HILL
Denver Zoo
Denver Botanic Gardens
Mile High Stadium
Denver
CAPITOL HILL
6th Ave
CHERRY CREEK
HILLTOP
Glendale
WASHINGTON PARK
University of Denver
VIRGINIA VILLAGE
UNIVERSITY HILLS
COLLEGE VIEW
Sheridan
Englewood
Cherry Hills Village
Wings Over the Rockies Air & Space Museum
Google
Map data ©2018 Google