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>

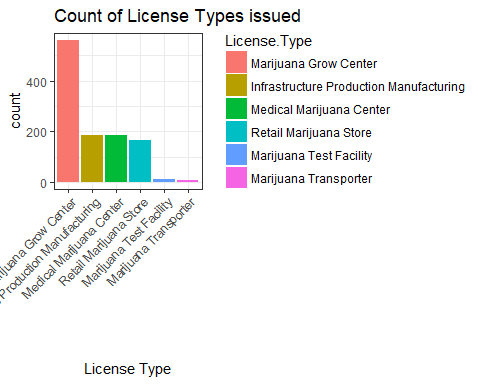
RgoogleMaps - Markus Loecher and Karl Ropkins (2015). RgoogleMaps and loa: Unleashing R Graphics Power on Map Tiles. Journal of Statistical Software 63(4), 1-18. URL <http://www.jstatsoft.org/v63/i04/>.

### Retrieve Denver Marijuana Licenses

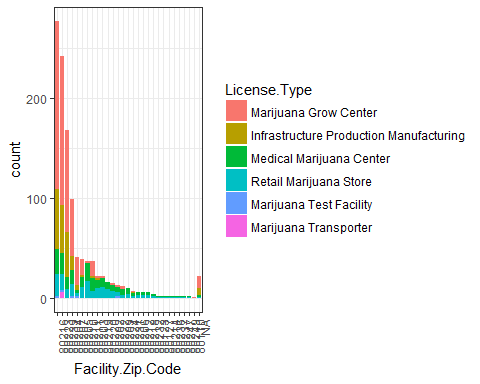
library(stringr) #string operations  
library(ggplot2) #graphics  
library(ggmap) #used for geocoding   
library(dplyr) #dataframe manipulation  
  
#retrieve dataset from denvergov.org  
denver\_mj\_licenses <- read.csv("https://www.denvergov.org/media/gis/DataCatalog/marijuana\_active\_business\_licenses/csv/marijuana\_active\_business\_licenses.csv", header=TRUE, stringsAsFactors = FALSE)  
  
#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 Prod Manuf" = "Infrastructure Production Manufacturing", "Retail Marijuana Inf Prod Mfg" = "Infrastructure Production Manufacturing", "Med Marijuana Test Facility" = "Marijuana Test Facility", "Retail Marijuana Test Facility" = "Marijuana Test Facility", "Med Marijuana Opt Prem Cultiv" = "Marijuana Grow Center", "Retail Marij Opt. Prem. Cultiv" = "Marijuana Grow Center"))  
  
#create factor for zip code  
denver\_mj\_licenses$Facility.Zip.Code <- factor(denver\_mj\_licenses$Facility.Zip.Code)  
  
#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   
##

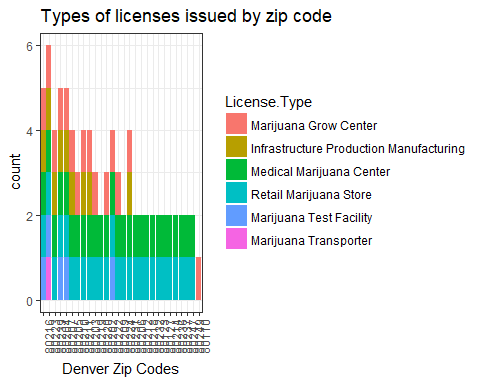
#cleanup again - darn OCD  
rm(i, num, dir, street, type, result)  
  
#sort by License.Type  
denver\_mj\_licenses <- within(denver\_mj\_licenses, License.Type <- factor(License.Type,   
 levels=names(sort(table(License.Type), decreasing=TRUE))))  
  
#generate plot to show graphically the number of 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.Zip.Code), decreasing=TRUE))))  
#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))

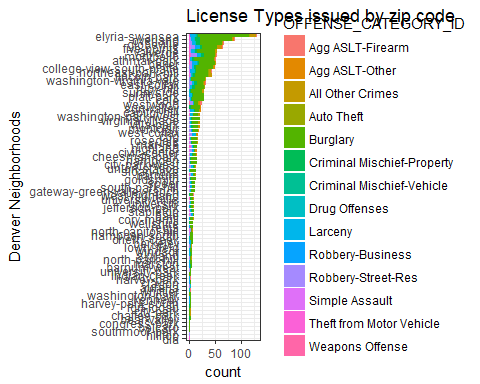
 ### Generating Maps

### Clean and save marijuana crime dataset

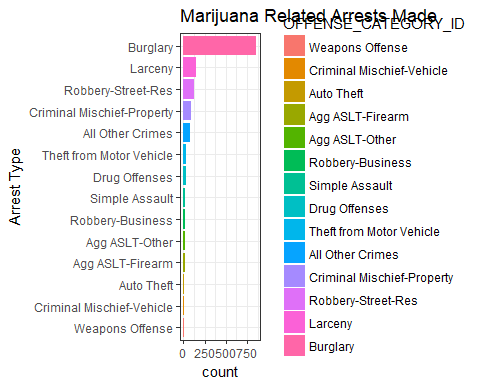
#retrieve the DPD marijuana crime file from Denver open data portal  
denver\_mj\_crime <- read.csv("https://www.denvergov.org/media/gis/DataCatalog/crime\_marijuana/csv/crime\_marijuana.csv", header=TRUE, strip.white = TRUE, stringsAsFactors = FALSE)  
  
#write original file to disk for posterity  
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", "JUL"="07", "AUG"="08", "SEP"="09", "OCT"="10", "NOV"="11", "DEC"="12"))  
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(NEIGHBORHOOD\_ID), decreasing=FALSE))))  
  
#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()

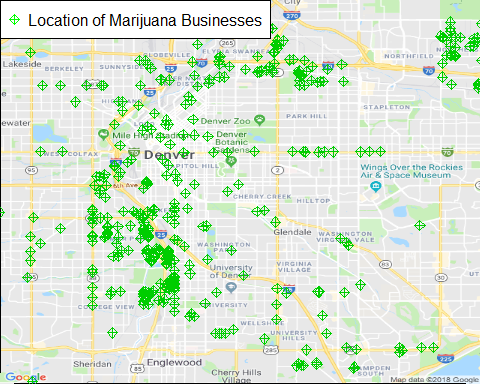


#sort by crime type  
denver\_mj\_crime <- within(denver\_mj\_crime, OFFENSE\_CATEGORY\_ID <- factor(OFFENSE\_CATEGORY\_ID,   
 levels=names(sort(table(OFFENSE\_CATEGORY\_ID), decreasing=FALSE))))  
  
#generate plot to show where licenses are being issued  
ggplot(denver\_mj\_crime,aes(x=OFFENSE\_CATEGORY\_ID, 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$LATITUDE), lon = mean(denver\_mj\_licenses$LONGITUDE)), destfile="denver\_map.png", size = c(640, 640), zoom = 12, sensor = "false", maptype = "roadmap", format = "png", RETURNIMAGE = TRUE, GRAYSCALE = FALSE, NEWMAP = TRUE, SCALE = 1, API\_console\_key = apikey, verbose = 0)  
  
p1 <- PlotOnStaticMap(denver\_map, lat = denver\_mj\_licenses$LATITUDE, lon = denver\_mj\_licenses$LONGITUDE, destfile="denver\_map.png", GRAYSCALE = FALSE, add = FALSE, FUN = points, NEWMAP = TRUE, TrueProj = TRUE, axes = TRUE, atX = NULL, atY = NULL, col = 3, pch=9)  
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, destfile="denver\_map.png", GRAYSCALE = FALSE, add = FALSE, FUN = points, NEWMAP = TRUE, TrueProj = TRUE, axes = TRUE, atX = NULL, atY = NULL, col = 2, pch=12)  
legend("topleft", legend = "Location of Marijuana Arrests", col = "red", bg = "white", pch=12)

