NOISE

Juan Nunez

April 4, 2019

Set working directory.

setwd("C:\\Users\\Juan Nunez\\Desktop\\DATA\_Capstone")

Open packages.

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readr)  
library(ggplot2)  
library(tidyr)

Open dataset.

noise <- read\_csv("DATASETS\\Noise\_Complaints.csv")

## Parsed with column specification:  
## cols(  
## `Case Number` = col\_double(),  
## `Case Open Date` = col\_character(),  
## `Case Close Date` = col\_character(),  
## `Case Year` = col\_double(),  
## `Case Sub-Type` = col\_character(),  
## `Address Zip Code` = col\_double()  
## )

## Warning: 1 parsing failure.  
## row col expected actual file  
## 2255 Address Zip Code no trailing characters 4023 'DATASETS\Noise\_Complaints.csv'

Now let’s look at the top dataset

dim(noise)

## [1] 6954 6

head(noise)

## # A tibble: 6 x 6  
## `Case Number` `Case Open Date` `Case Close Dat~ `Case Year`  
## <dbl> <chr> <chr> <dbl>  
## 1 19002 05/10/2007 01/08/2013 2007  
## 2 20161106 02/05/2016 04/07/2016 2016  
## 3 19847 12/28/2007 04/25/2008 2007  
## 4 15816 01/18/2005 01/27/2005 2005  
## 5 20161573 06/03/2016 06/20/2016 2016  
## 6 8280 04/09/1999 05/06/1999 1999  
## # ... with 2 more variables: `Case Sub-Type` <chr>, `Address Zip  
## # Code` <dbl>

Add underscores to column names.

colnames(noise) = gsub(" ", "\_",colnames(noise))

Let’s see if the zip codes need zeros.

table(noise$Address\_Zip\_Code)

##   
## 20083 20110 20646 20705 20707 20743 20770 20774 20782 20812 20813 20814   
## 4 1 1 2 3 1 1 4 1 7 2 665   
## 20815 20816 20817 20818 20822 20827 20832 20833 20837 20838 20839 20841   
## 368 200 274 15 1 2 131 40 25 2 5 28   
## 20842 20850 20851 20852 20853 20854 20855 20860 20861 20862 20866 20868   
## 24 384 41 349 111 304 91 12 19 2 52 8   
## 20870 20871 20872 20874 20876 20877 20878 20879 20880 20882 20884 20885   
## 1 83 50 267 109 84 175 125 4 74 1 1   
## 20886 20889 20891 20892 20895 20896 20899 20901 20902 20903 20904 20905   
## 147 3 4 3 162 6 1 145 257 38 165 51   
## 20906 20910 20912 20985 20987 21036 21061 21207 21666 21754 21771 21816   
## 237 584 57 1 1 1 1 1 1 1 4 1   
## 22101 35512 77380 88191 92024   
## 1 1 1 1 1

Create a date variable.

noise$Date <- as.Date(noise$Case\_Open\_Date , "%m/%d/%Y")  
class(noise$Date)

## [1] "Date"

Let’s see the Case Sub\_Type variable in detail.

table(noise$`Case\_Sub-Type`)

##   
## Industrial/commercial Electrical Amp Equip   
## 488   
## Industrial/commercial Equipment Noise   
## 1333   
## Industrial/commercial heat Pump/ac Noise   
## 96   
## Industrial/commercial Refuse Noise Collection   
## 187   
## Industrial/commercial construction/repair/demolit   
## 1330   
## Information Request   
## 38   
## Noise Industrial/commercial Leaf Blower Noise   
## 368   
## Project Community Noise Survey   
## 7   
## Residential Construction/repair/demolition   
## 604   
## Residential El Amplification Equip   
## 736   
## Residential Mechanical   
## 422   
## Residential Motor Vehicle Noise   
## 208   
## Residential House/yard   
## 436   
## Residential Muti Tenant Dwelling   
## 409   
## Waiver Request   
## 292

Create a three category variable for Case Type.

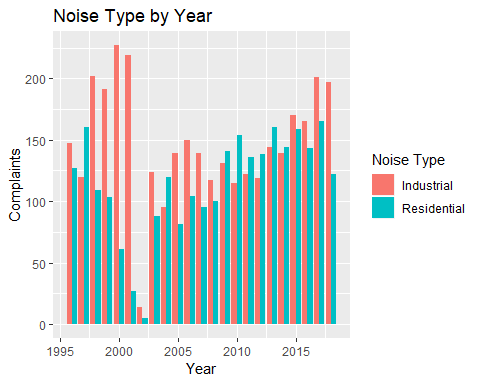
noise['NOISE\_TYPE']= "Industrial"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential Construction/repair/demolition"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential Mechanical"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential House/yard"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential El Amplification Equip"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential Motor Vehicle Noise"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Residential Muti Tenant Dwelling"] <- "Residential"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Waiver Request"] <- "Other"  
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Information Request"] <- "Other"   
noise$NOISE\_TYPE[noise$`Case\_Sub-Type` == "Project Community Noise Survey"] <- "Other"   
  
  
table(noise$NOISE\_TYPE)

##   
## Industrial Other Residential   
## 3802 337 2815

Save data as csv.

write.csv(noise, file = "noise.csv",row.names=FALSE, na="")

#filter(!is.na(AGE\_GROUP\_fixed)) %>%  
noise %>% filter(NOISE\_TYPE != "Other") %>% filter(Case\_Year > 1995 & Case\_Year < 2019) %>%  
ggplot(aes(x = Case\_Year, fill = NOISE\_TYPE)) +  
geom\_bar(position = 'dodge') +  
labs(title = "Noise Type by Year", x= "Year", y = "Complaints") +  
guides(fill=guide\_legend(title = "Noise Type"))



table(noise$Case\_Year)

##   
## 1989 1990 1991 1992 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004   
## 55 94 85 60 9 210 276 281 311 295 288 246 19 230 216   
## 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019   
## 228 266 243 230 279 287 275 283 330 297 353 341 406 370 91