San Francisco Crime Report - Jingbin Xu 24 Dec 2016

INTRODUCTION

The San Francisco crime dataset provides a special opportunity to play with data visually. As a hobbyist photographer, I am a big fan of pretty pictures. Since the dataset covers a wide variety of crimes, visualizing them all at once might wash out any patterns contained in smaller subgroups. I have elected to see if there are any patterns within violent crime that can be explored visually.

THE DATA

```
##Columns include logitude and latitude, the category the crime is classified as,and
date and time. When we import the data, we extract the Year, Month, Day, and Hour as
individual columns. This will allow us to break out our visualizations more easily in
to these dimensions. Lets look at all the crime categories to see which might fall un
der the umbrella of ???Violent Crime???
```

```
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
```

```r
library(readr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```r
coltypes <-list(Dates = col_datetime("%Y-%m-%d %H:%M:%S"))
train <-read_csv("train-2.csv",col_types=coltypes)
test <-read_csv("test-2.csv",col_types=coltypes)
train <-train %>%
    mutate(Year  = factor(year(Dates), levels=2003:2015),
           Month = factor(month(Dates), levels=1:12),
           Day   = day(Dates),
           Hour  = factor(hour(Dates), levels=0:23),
           dayDate = as.POSIXct(round(Dates, units = "days")),
           DayOfWeek = factor(DayOfWeek, levels=c("Monday",
                                                  "Tuesday",
                                                  "Wednesday",
                                                  "Thursday",
                                                  "Friday",
                                                  "Saturday",
                                                  "Sunday"))
           )
unique(train$Category)
```

```
##  [1] "WARRANTS"                   "OTHER OFFENSES"
##  [3] "LARCENY/THEFT"              "VEHICLE THEFT"
##  [5] "VANDALISM"                  "NON-CRIMINAL"
##  [7] "ROBBERY"                    "ASSAULT"
##  [9] "WEAPON LAWS"                "BURGLARY"
## [11] "SUSPICIOUS OCC"             "DRUNKENNESS"
## [13] "FORGERY/COUNTERFEITING"     "DRUG/NARCOTIC"
## [15] "STOLEN PROPERTY"            "SECONDARY CODES"
## [17] "TRESPASS"                   "MISSING PERSON"
## [19] "FRAUD"                      "KIDNAPPING"
## [21] "RUNAWAY"                    "DRIVING UNDER THE INFLUENCE"
## [23] "SEX OFFENSES FORCIBLE"      "PROSTITUTION"
## [25] "DISORDERLY CONDUCT"         "ARSON"
## [27] "FAMILY OFFENSES"            "LIQUOR LAWS"
## [29] "BRIBERY"                    "EMBEZZLEMENT"
## [31] "SUICIDE"                    "LOITERING"
## [33] "SEX OFFENSES NON FORCIBLE"  "EXTORTION"
## [35] "GAMBLING"                   "BAD CHECKS"
## [37] "TREA"                       "RECOVERED VEHICLE"
## [39] "PORNOGRAPHY/OBSCENE MAT"
```

Observation

```
## # A tibble: 3 <U+00D7> 2
##                 Category       n
##                    <chr> <int>
## 1              ASSAULT 76876
## 2              ROBBERY 23000
## 3 SEX OFFENSES FORCIBLE  4388
```

Occurences of assault far outnumber the other crimes, and as such will be heavily represented in any aggregate observations. It will be interesting to look at violent crime both in aggregate and by specific category.