# Predict the category of crimes that occurred in the San-Francisco

# Statistical Data Analysis Project

(Based on R)

BY,

Drashti Patel

Stevens Institute of Technology, Hoboken, NJ.

Spring 2016.

# Abstract

The project is statistical data analysis project. In which, we are going to analyse the Crime-Data about San-Francisco to predict the different level of crime rates over the city. The project will provide different types of analysis from the data like statistical analysis of each type of crime, individual Geographic presentation of each type of crime etc. This analysis will definitely help the police department with many different security aspects. The statistical analysis helps to know the crime rates. As well as, Geographic analysis will help them to develop and maintain security to make the city more secure.

# Data

- We have a data about the different type of crimes in San-Francisco. Which includes Date, Time, Type of Crime, Description of step taken by Police, District, and the Exact location of the Crime Place.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2003-01-07 07:52:00 | WARRANTS | WARRANT ARREST | Tuesday | SOUTHERN | ARREST, BOOKED | 5TH ST / SHIPLEY ST | -122.402843 | 37.779829 |
| 2003-01-07 04:49:00 | WARRANTS | ENROUTE TO OUTSIDE JURISDICTION | Tuesday | TENDERLOIN | ARREST, BOOKED | CYRIL MAGNIN STORTH ST / EDDY ST | -122.408495 | 37.784452 |
| 2003-01-07 03:52:00 | WARRANTS | WARRANT ARREST | Tuesday | NORTHERN | ARREST, BOOKED | OFARRELL ST / LARKIN ST | -122.417904 | 37.785167 |
| 2003-01-07 03:34:00 | WARRANTS | WARRANT ARREST | Tuesday | NORTHERN | ARREST, BOOKED | DIVISADERO ST / LOMBARD ST | -122.442650 | 37.796999 |
| 2003-01-07 01:22:00 | WARRANTS | WARRANT ARREST | Tuesday | SOUTHERN | ARREST, BOOKED | 900 Block of MARKET ST | -122.409537 | 37.782691 |
| 2003-01-06 23:30:00 | WARRANTS | ENROUTE TO OUTSIDE JURISDICTION | Monday | BAYVIEW | ARREST, BOOKED | REVERE AV / INGALLS ST | -122.384557 | 37.728487 |
| 2003-01-06 23:14:00 | WARRANTS | WARRANT ARREST | Monday | CENTRAL | ARREST, BOOKED | BUSH ST / HYDE ST | -122.417019 | 37.789110 |
| 2003-01-06 22:45:00 | WARRANTS | WARRANT ARREST | Monday | SOUTHERN | ARREST, BOOKED | 800 Block of BRYANT ST | -122.403405 | 37.775421 |
| 2003-01-06 22:45:00 | WARRANTS | ENROUTE TO OUTSIDE JURISDICTION | Monday | SOUTHERN | ARREST, BOOKED | 800 Block of BRYANT ST | -122.403405 | 37.775421 |
| 2003-01-06 22:19:00 | WARRANTS | ENROUTE TO OUTSIDE JURISDICTION | Monday | NORTHERN | ARREST, BOOKED | GEARY ST / POLK ST | -122.419740 | 37.785893 |
| 2003-01-06 21:54:00 | WARRANTS | ENROUTE TO OUTSIDE JURISDICTION | Monday | NORTHERN | ARREST, BOOKED | SUTTER ST / POLK ST | -122.420120 | 37.787757 |

Sample Data

**Data fields**

- **Dates** - timestamp of the crime incident
- **Category** - category of the crime incident (only in train.csv). **This is the target variable you are going to predict.**
- **Descript** - detailed description of the crime incident (only in train.csv)
- **DayOfWeek** - the day of the week
- **PdDistrict** - name of the Police Department District
- **Resolution** - how the crime incident was resolved (only in train.csv)
- **Address** - the approximate street address of the crime incident
- **X** - Longitude
- **Y** - Latitude

# Analysis Descriptions

- According to the definition of Crime *(As provided in Wikipedia),* the crime mainly taxonomies into two categories.

  o White Collar Crime
    It refers to financially motivated nonviolent crime committed by business and government professionals. Typical white-collar crimes include **fraud, bribery, Ponzi schemes, insider trading, embezzlement, cybercrime, copyright infringement, money laundering, identity theft, and forgery.**

- o Blue Collar Crime
  It is any crime committed by an individual from a lower social class as opposed to white-collar crime which is associated with crime committed by someone of a higher-level social class. **Typical blue-collar crimes include vandalism, Larceny/Theft, Stolen Property, Robbery, Driving Under the Influence, Disorderly Conduct, Liquor Laws, Vehicle Theft, Assault, Kidnapping, Trespass, Arson, Recovered Vehicle.**

- o Other Crime
  It includes **Missing Person, Runway, Family Offenses, Sex Offenses Non Forcible, Pornography/Obscene mat, Weapon laws, Drunkenness, Suicide, Trae, Drug/Narcotic, Sex Offenses Forcible, Loitering etc.**

Based on to the types of crime we are planning to provide different types of analysis as mentioned below.

## I. Statistical Analysis
- o In Statistical analysis we are planning to provide all three types of crime rates on yearly bases.
- o Top 10 crimes based on Max number of Crimes in Individual categories

| Category | count |
|---|---|
| LARCENY/THEFT | 174900 |
| OTHER OFFENSES | 126182 |
| NON-CRIMINAL | 92304 |
| ASSAULT | 76876 |
| DRUG/NARCOTIC | 53971 |
| VEHICLE THEFT | 53781 |
| VANDALISM | 44725 |
| WARRANTS | 42214 |
| BURGLARY | 36755 |
| SUSPICIOUS OCC | 31414 |
| MISSING PERSON | 25989 |
| ROBBERY | 23000 |
| FRAUD | 16679 |
| FORGERY/COUNTERFEITING | 10609 |
| SECONDARY CODES | 9985 |
| WEAPON LAWS | 8555 |
| PROSTITUTION | 7484 |
| TRESPASS | 7326 |
| STOLEN PROPERTY | 4540 |
| SEX OFFENSES FORCIBLE | 4388 |

Result

## II. Geographical Analysis
- o Here, we are planning to map the crime locations into the Geographic ears of the city. This is the important asset of the project. We will able to conclude many security pros and cons by analyzing the data on geographic manner.

## III. Crime Variation by Year
## IV. Total no of crime in Individual Categories

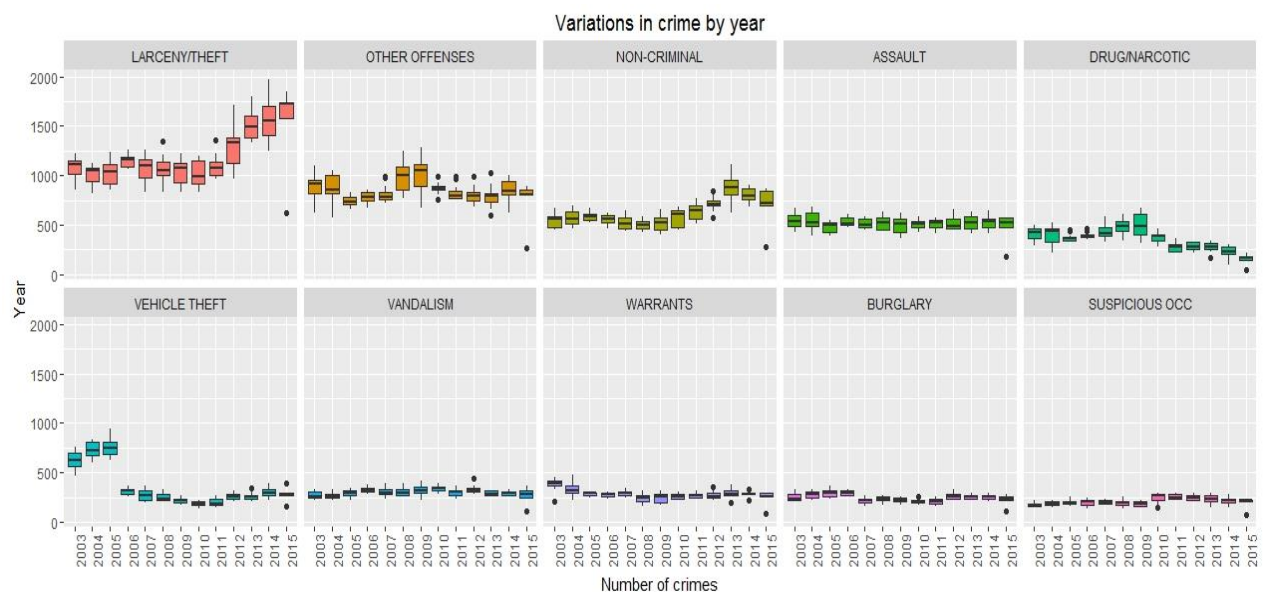## **Implementation Environment**

### **We are planning to perform the analysis in to the R language.**

We used R-studio and Different R-libraries to complete our analysis & have studied different R-Packages and decided to use which packages to perform our Project at the very moment. We have used some basic R- libraries like,

- library(dplyr): Tools for splitting, applying and combining data.
- library(ggplot2): An implementation of the Grammar of Graphics.
- library(data.table): Fast aggregation of large data (e.g. 100GB in RAM), fast ordered joins, fast add/modify/delete of columns by group using no copies at all, list columns and a fast file reader.
- library(reshape2): Flexibly reshape data: a reboot of the reshape package.
- library(readr): Provide a fast and friendly way to read tabular data into R.
- library(lubridate): fast and userfriendly parsing of date-time data, extraction and updating of components of a date-time (years, months, days, hours, minutes, and seconds).
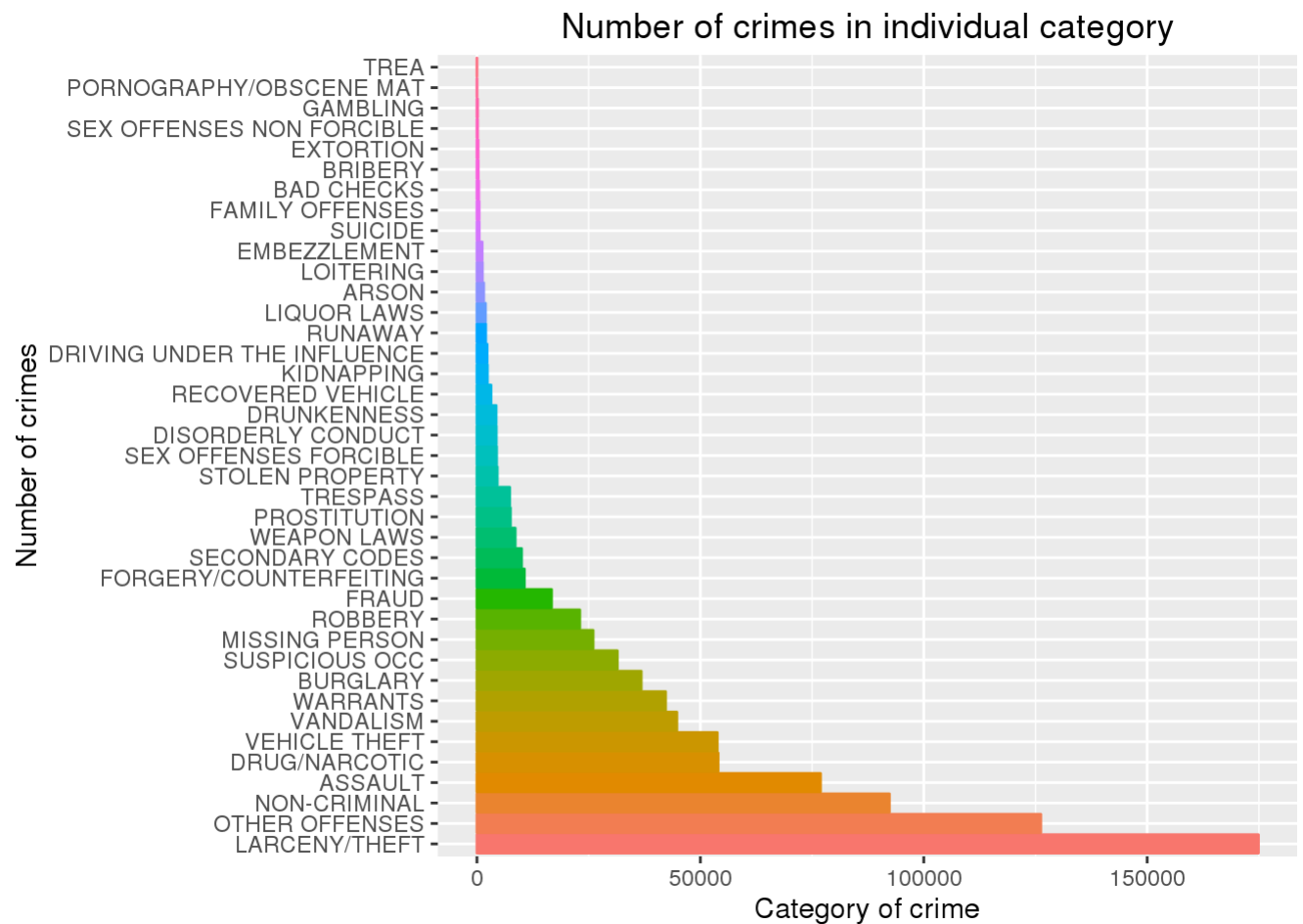
### **Analysis Results:**

The below plot shows median total crimes per month from 2003 to 2015. Plot indicates that larceny/theft rates are on rise. Most interesting trend is reduction in number of vehicle thefts from 2006 to 2007.
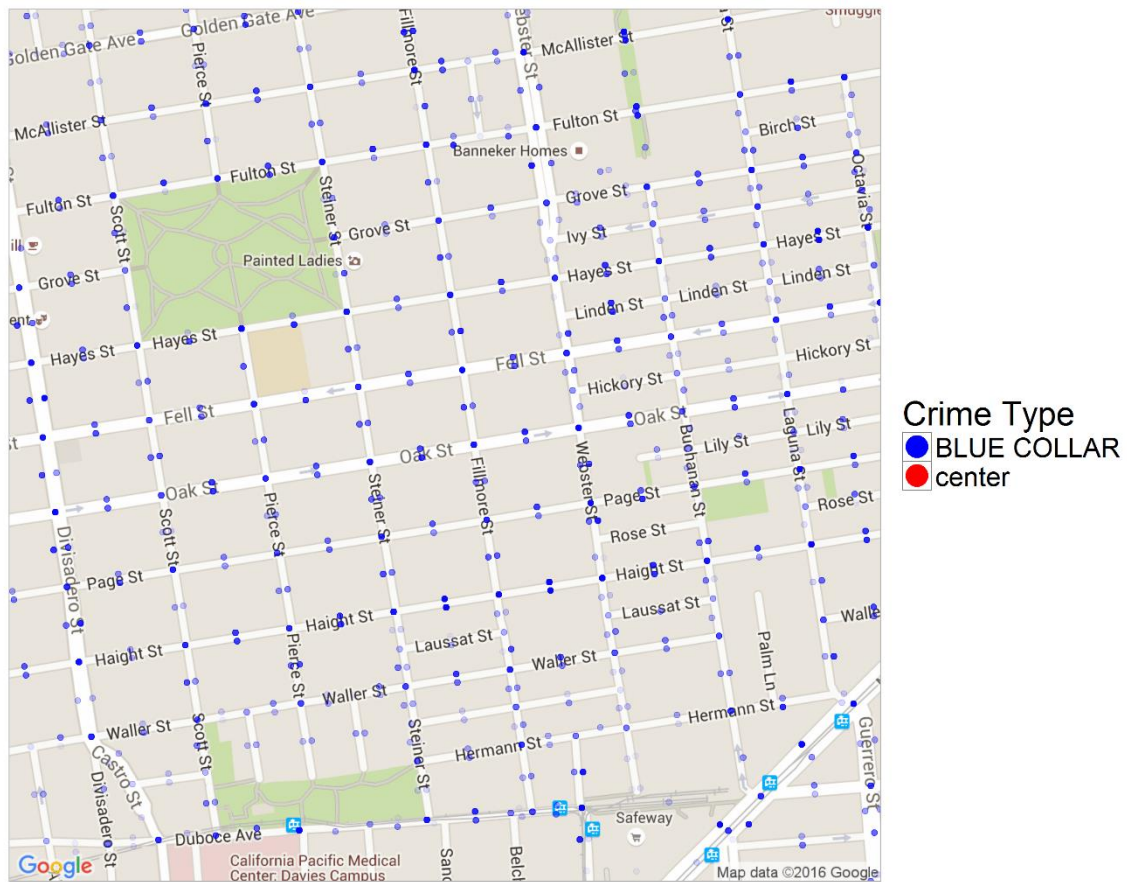


### **Types of crime.**

We next draw a bar plot to show number of crimes in each category. Bar plots indicate that the top crime category is Larceny/Theft. Further, top 20 crime types account for 97% of the crimes.

## Number of crimes in individual category



```
## [1] "Top 10 crimes"
##                  Category   count
## 17          LARCENY/THEFT  174900
## 22         OTHER OFFENSES  126182
## 21           NON-CRIMINAL   92304
## 2                 ASSAULT   76876
## 8           DRUG/NARCOTIC   53971
## 37          VEHICLE THEFT   53781
## 36              VANDALISM   44725
## 38               WARRANTS   42214
## 5                BURGLARY   36755
## 33         SUSPICIOUS OCC   31414
## 20         MISSING PERSON   25989
## 26                ROBBERY   23000
## 14                  FRAUD   16679
## 13 FORGERY/COUNTERFEITING   10609
## 28         SECONDARY CODES    9985
## 39            WEAPON LAWS    8555
## 24            PROSTITUTION    7484
## 35               TRESPASS    7326
## 31         STOLEN PROPERTY    4540
## 29    SEX OFFENSES FORCIBLE   4388
## [1] "Percentage of crimes in top 20 categories =  96.9965229730915"
```
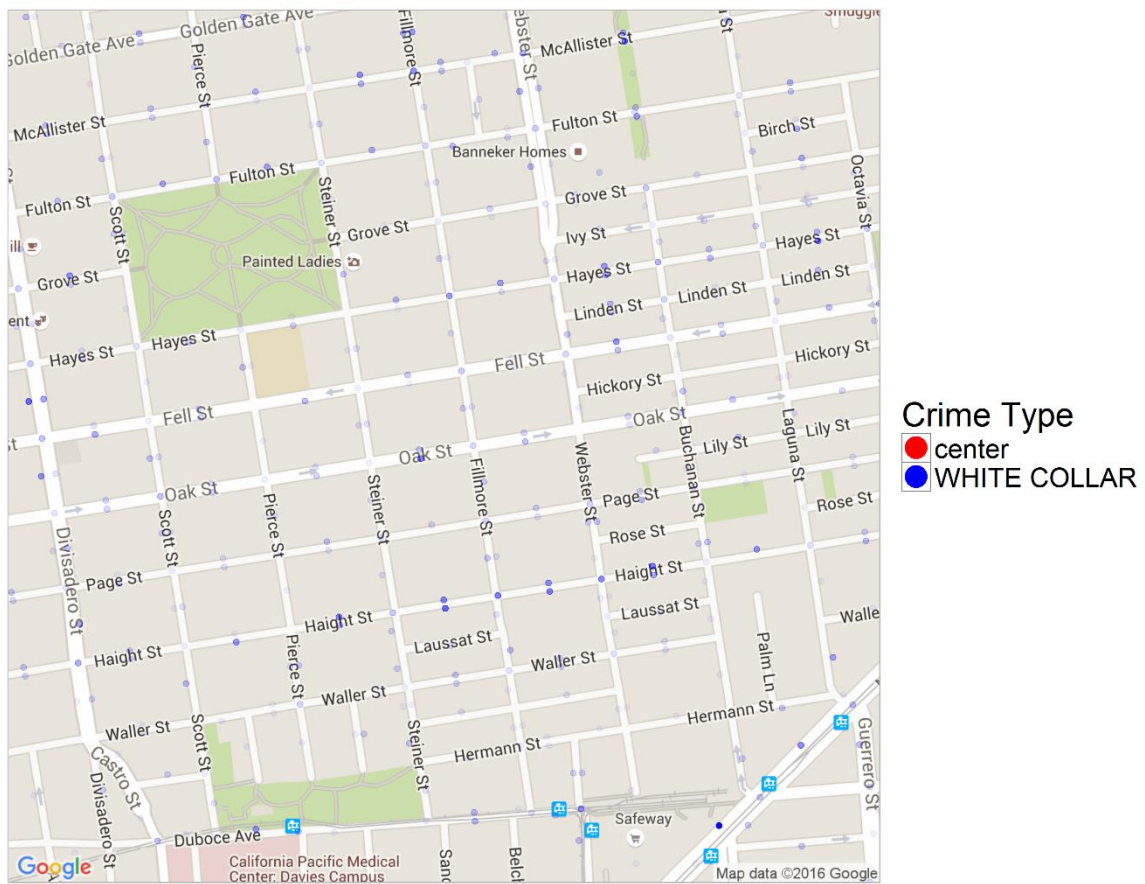
**Geometric Representations:**

Blue-Collar crimes in SF

# White-Collar crimes in SF



## Crime Type
● center
● WHITE COLLAR

## Total Crime Analysis:



White-Collar vs Blue-Collar Crimes in SF