Los Angeles Crime Data

(2010-Present)

Date: April 27, 2019

Vaibhav Rastogi

Agenda

1. Introduction & Problem Statement
2. The Data
3. Exploratory Analysis
4. Challenges
5. Insights & Next Steps

Introduction

The City of Los Angeles publishes Los Angeles Police District (LAPD) crime and arrest data from 2010 to present (updated weekly) on Kaggle link. We explored both crime (1.96M records) and arrest (1.25M records) datasets as at 19 April 2019.

We were interested in looking at what clustering methods could tell us about crime hot spots. Our problem statement is premised on one of the 5 questions data science can answer framed as *"Which areas in Los Angeles are Low Risk, Medium Risk or High Risk based on the frequency of crime incidents?"*

In our exploratory analysis we looked at crimes vs arrests by location, differences in the profile of offenders vs victims and offenders by arrest type. For the clustering analysis we adopted two approaches. The first approach focused on 2017, the year with the highest frequency of crimes using hierarchical clustering. The second approach focused on 2010 to present using clustering along with PCA.

The Data

Crimes

DR_Number	Date_Reported	Date_Occurred	Time_Occurred
131318519	2013-09-20T00:00:00	2010-09-10T00:00:00	2200
160106330	2016-01-29T00:00:00	2010-01-01T00:00:00	700
171508002	2017-03-16T00:00:00	2010-09-15T00:00:00	2045
171608778	2017-04-19T00:00:00	2010-01-01T00:00:00	700
171608779	2017-04-19T00:00:00	2010-01-01T00:00:00	700
170512370	2017-06-29T00:00:00	2010-11-03T00:00:00	1

Arrests

Report_ID	Arrest_Date	Time
101000690	2010-05-07T00:00:00	2100
101000692	2010-05-07T00:00:00	2100
101000990	2010-11-23T00:00:00	1130
101000991	2010-11-23T00:00:00	1130
101004315	2010-01-06T00:00:00	1600
101004316	2010-01-07T00:00:00	1600

Area_ID	Area_Name	Reporting_District	Address	Cross_Street	Location
10	West Valley	1027	17300 VANOWEN ST		{'needs_recoding': False, 'latitude': '34.1939', 'lo
10	West Valley	1027	17300 VANOWEN ST		{'needs_recoding': False, 'latitude': '34.1939', 'lo
10	West Valley	1027	16900 VOSE ST		{'needs_recoding': False, 'latitude': '34.1987', 'lo
10	West Valley	1027	16900 VOSE ST		{'needs_recoding': False, 'latitude': '34.1987', 'lo
10	West Valley	1011	19300 SATICOY ST		{'needs_recoding': False, 'latitude': '34.2084', 'lo
10	West Valley	1011	19300 SATICOY ST		{'needs_recoding': False, 'latitude': '34.2084', 'lo

Age	Sex_Code	Descent_Code
19	M	H
19	M	W
52	M	H
62	M	H
40	M	H
45	M	W

Arrest_Type_Code	Charge	Charge_Description
M	11357(B)HS	POSSESS 28.5 GRAMS OR LESS OF MARIJUANA
M	11357(B)HS	POSSESS 28.5 GRAMS OR LESS OF MARIJUANA
M	66.28I.AMC	TAMPERING WITH RECYCLING OR REFUSE
M	66.28I.AMC	TAMPERING WITH RECYCLING OR REFUSE
M	41.27CLAMC	DRINKING IN PUBLIC
M	41.27CLAMC	DRINKING IN PUBLIC

Area_ID	Area_Name	Reporting_District	Address	Cross_Street	Location
13	Newton	1313	700 E WASHINGTON BL		{'needs_recoding': False, 'latitude': '34.0277', 'longitu
1	Central	152	500 S GRAND AV		{'needs_recoding': False, 'latitude': '34.0502', 'longitu
19	Mission	1901	FOOTHILL BL	GLENOAKS AV	{'needs_recoding': False, 'latitude': '34.3229', 'longitu
16	Foothill	1611	13300 BROWNELL ST		{'needs_recoding': False, 'latitude': '34.2831', 'longitu
16	Foothill	1611	13300 BROWNELL ST		{'needs_recoding': False, 'latitude': '34.2831', 'longitu
5	Harbor	523	2100 JOHN S GIBSON BL		{'needs_recoding': False, 'latitude': '33.7577', 'longitu

Premise_Code	Premise_Description	Weapon_Used_Code	Weapon_Description
203	OTHER BUSINESS	NA	
503	HOTEL	NA	
501	SINGLE FAMILY DWELLING	301	BELT FLAILING INSTRUMENT/CHAIN
501	SINGLE FAMILY DWELLING	400	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
501	SINGLE FAMILY DWELLING	NA	
726	POLICE FACILITY	NA	

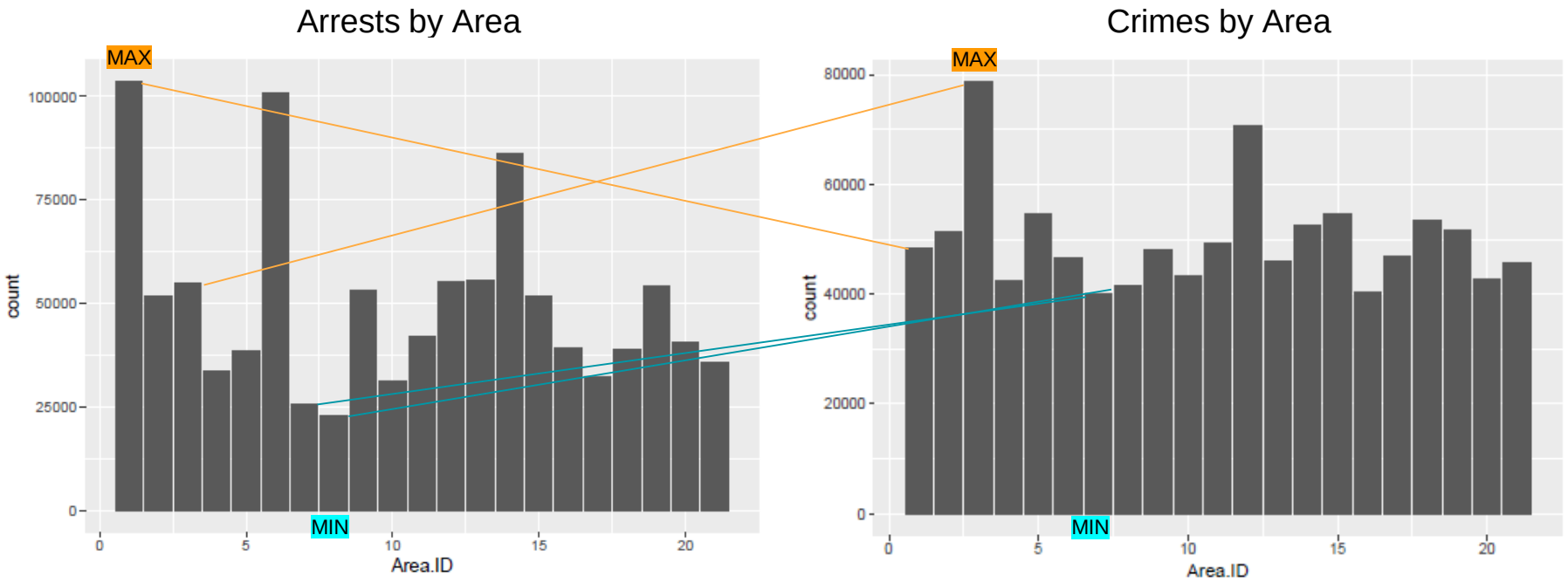
Victim_Age	Victim_Sex	Victim_Descent
62	M	W
50	F	A
20	F	H
2	F	H
5	F	H
-8	X	X

Crime_Code	Crime_Code_Description	Crime_Code_2	Crime_Code_3	Crime_Code_4
668	EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)	NA	NA	NA
668	EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)	NA	NA	NA
626	INTIMATE PARTNER - SIMPLE ASSAULT	NA	NA	NA
812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	860	NA	NA
810	SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W/ FRGN OBJ	812	NA	NA
946	OTHER MISCELLANEOUS CRIME	946	NA	NA

Status_Code	Status_Description
AA	Adult Arrest
IC	Invest Cont
AO	Adult Other
IC	Invest Cont
IC	Invest Cont
IC	Invest Cont

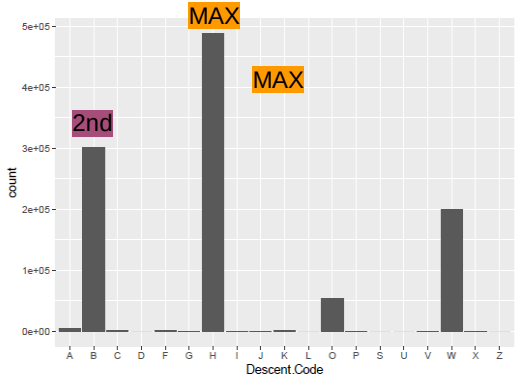
Exploratory Analysis

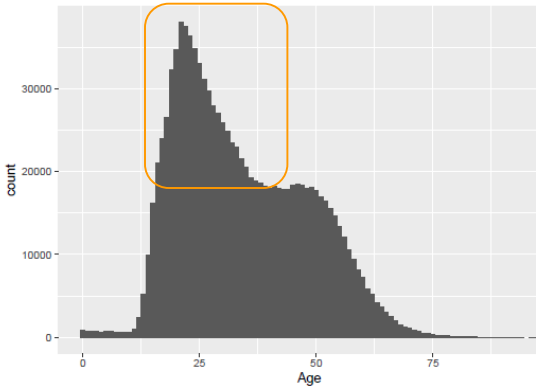
Arrests vs Crimes - by location

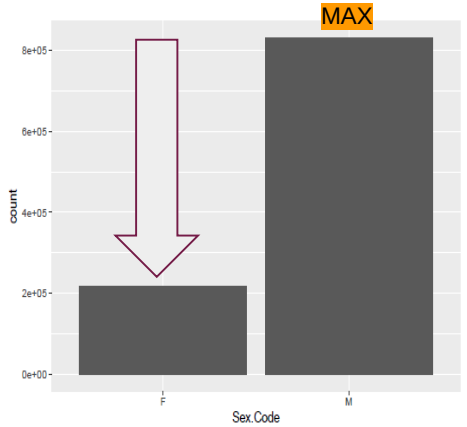


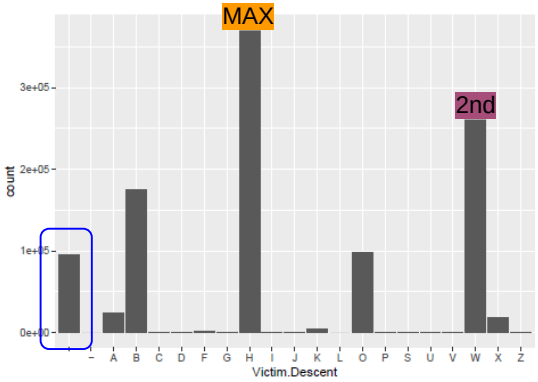
Exploratory Analysis

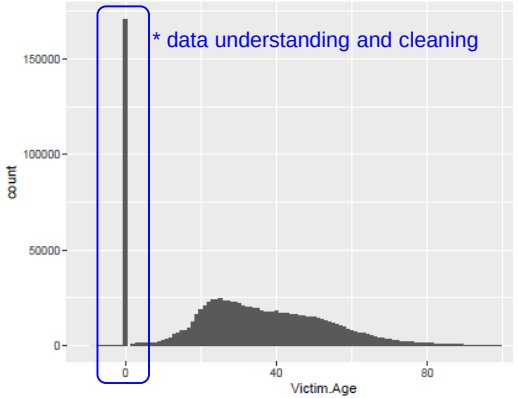
Offenders vs Victims

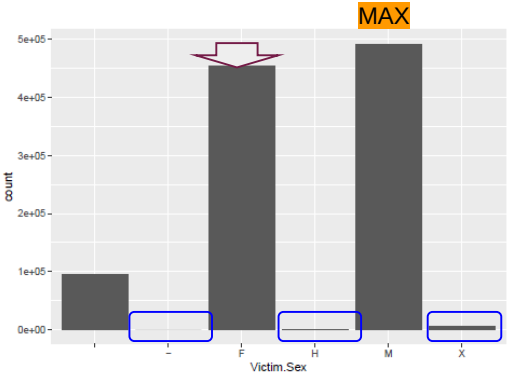








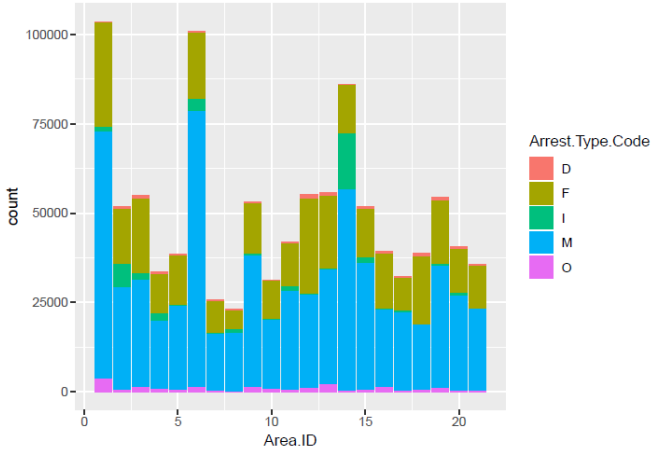


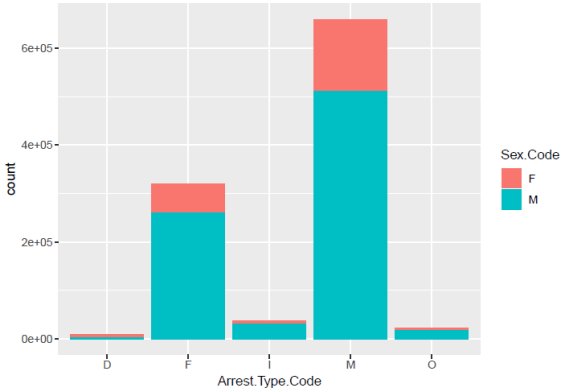


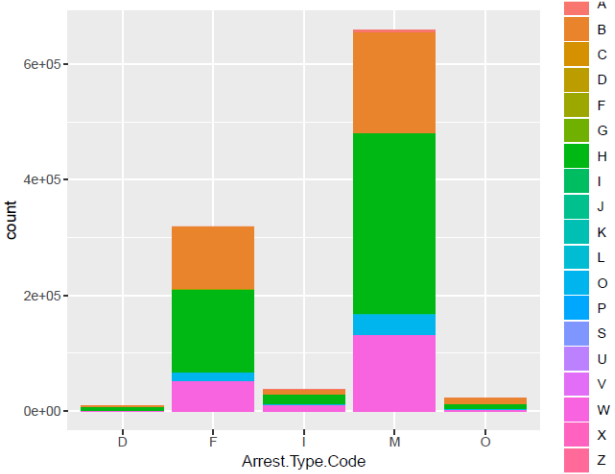
Exploratory Analysis

Offenders by Arrest type, Cross-tabs

D - Dependent
F - Felony
I - Infraction
M - Misdemeanor
O - Other



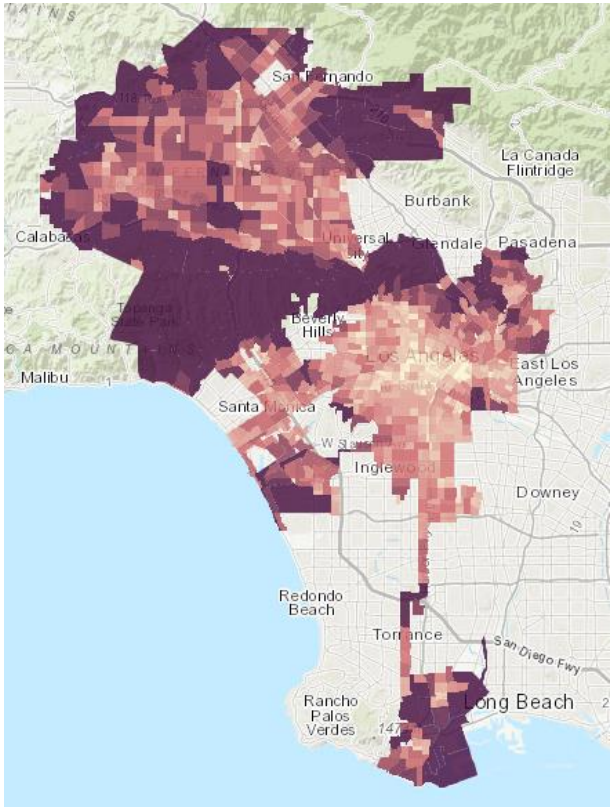




Challenge 1: Non-numerical data

```
 [1] ABORTION/ILLEGAL                                  ARSON
 [3] ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER      ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
 [5] ATTEMPTED ROBBERY                                 BATTERY - SIMPLE ASSAULT
 [7] BATTERY ON A FIREFIGHTER                          BATTERY POLICE (SIMPLE)
 [9] BATTERY WITH SEXUAL CONTACT                       BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM
[11] BIGAMY                                            BIKE - ATTEMPTED STOLEN
[13] BIKE - STOLEN                                     BOAT - STOLEN
[15] BOMB SCARE                                        BRANDISH WEAPON
[17] BRIBERY                                           BUNCO, ATTEMPT
[19] BUNCO, GRAND THEFT                                BUNCO, PETTY THEFT
[21] BURGLARY                                          BURGLARY FROM VEHICLE
[23] BURGLARY FROM VEHICLE, ATTEMPTED                  BURGLARY, ATTEMPTED
[25] CHILD ABANDONMENT                                 CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT
[27] CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT           CHILD ANNOYING (17YRS & UNDER)
[29] CHILD NEGLECT (SEE 300 W.I.C.)                    CHILD PORNOGRAPHY
[31] CHILD STEALING                                    CONSPIRACY
[33] CONTEMPT OF COURT                                 CONTRIBUTING
[35] COUNTERFEIT                                       CREDIT CARDS, FRAUD USE ($950 & UNDER
[37] CREDIT CARDS, FRAUD USE ($950.01 & OVER)          CRIMINAL HOMICIDE
[39] CRIMINAL THREATS - NO WEAPON DISPLAYED            CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)
[41] CRUELTY TO ANIMALS                                DEFRAUDING INNKEEPER/THEFT OF SERVICES, $400 & UNDER
[43] DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER $400 DISCHARGE FIREARMS/SHOTS FIRED
[45] DISHONEST EMPLOYEE - GRAND THEFT                  DISHONEST EMPLOYEE - PETTY THEFT
[47] DISHONEST EMPLOYEE ATTEMPTED THEFT                DISRUPT SCHOOL
[49] DISTURBING THE PEACE                              DOCUMENT FORGERY / STOLEN FELONY
[51] DOCUMENT WORTHLESS ($200 & UNDER)                 DOCUMENT WORTHLESS ($200.01 & OVER)
[53] DRIVING WITHOUT OWNER CONSENT (DWOC)              DRUGS, TO A MINOR
[55] DRUNK ROLL                                        DRUNK ROLL - ATTEMPT
[57] EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)        EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)
[59] EXTORTION                                         FAILURE TO DISPERSE
[61] FAILURE TO YIELD                                  FALSE IMPRISONMENT
[63] FALSE POLICE REPORT                               GRAND THEFT / AUTO REPAIR
[65] GRAND THEFT / INSURANCE FRAUD                     HUMAN TRAFFICKING - COMMERCIAL SEX ACTS
[67] HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE         ILLEGAL DUMPING
[69] INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)      INCITING A RIOT
[71] INDECENT EXPOSURE                                 INTIMATE PARTNER - AGGRAVATED ASSAULT
[73] INTIMATE PARTNER - SIMPLE ASSAULT                 KIDNAPPING
[75] KIDNAPPING - GRAND ATTEMPT                        LETTERS, LEWD  -  TELEPHONE CALLS, LEWD
[77] LEWD CONDUCT                                      LEWD/LASCIVIOUS ACTS WITH CHILD
[79] LYNCHING                                          LYNCHING - ATTEMPTED
[81] MANSLAUGHTER, NEGLIGENT                           ORAL COPULATION
[83] OTHER ASSAULT                                     OTHER MISCELLANEOUS CRIME
[85] PANDERING                                         PEEPING TOM
[87] PETTY THEFT - AUTO REPAIR                         PICKPOCKET
[89] PICKPOCKET, ATTEMPT                               PIMPING
```

Challenge 2: Visualizing Hierarchical data



Area

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21
```

This shows that there are 21 Area ID values for the target variable.

Now let's see if the Area.Name field has the same number of values, in which case they are likely related.

```
##  [1] 77th Street Central     Devonshire  Foothill    Harbor
##  [6] Hollenbeck  Hollywood   Mission     N Hollywood Newton
## [11] Northeast   Olympic     Pacific     Rampart     Southeast
## [16] Southwest   Topanga     Van Nuys    West LA     West Valley
## [21] Wilshire
## 21 Levels: 77th Street Central Devonshire Foothill Harbor ... Wilshire
```

District

Coordinates and Addresses

Location
{'needs_recoding': False, 'latitude': '34.1939', 'longitude': '-11
{'needs_recoding': False, 'latitude': '34.1939', 'longitude': '-11
{'needs_recoding': False, 'latitude': '34.1987', 'longitude': '-11
{'needs_recoding': False, 'latitude': '34.1987', 'longitude': '-11
{'needs_recoding': False, 'latitude': '34.2084', 'longitude': '-11
{'needs_recoding': False, 'latitude': '34.2084', 'longitude': '-11

Address
17300 VANOWEN ST
17300 VANOWEN ST
16900 VOSE ST
16900 VOSE ST
19300 SATICOY ST
19300 SATICOY ST

Challenge 3: What is our concept of distance?

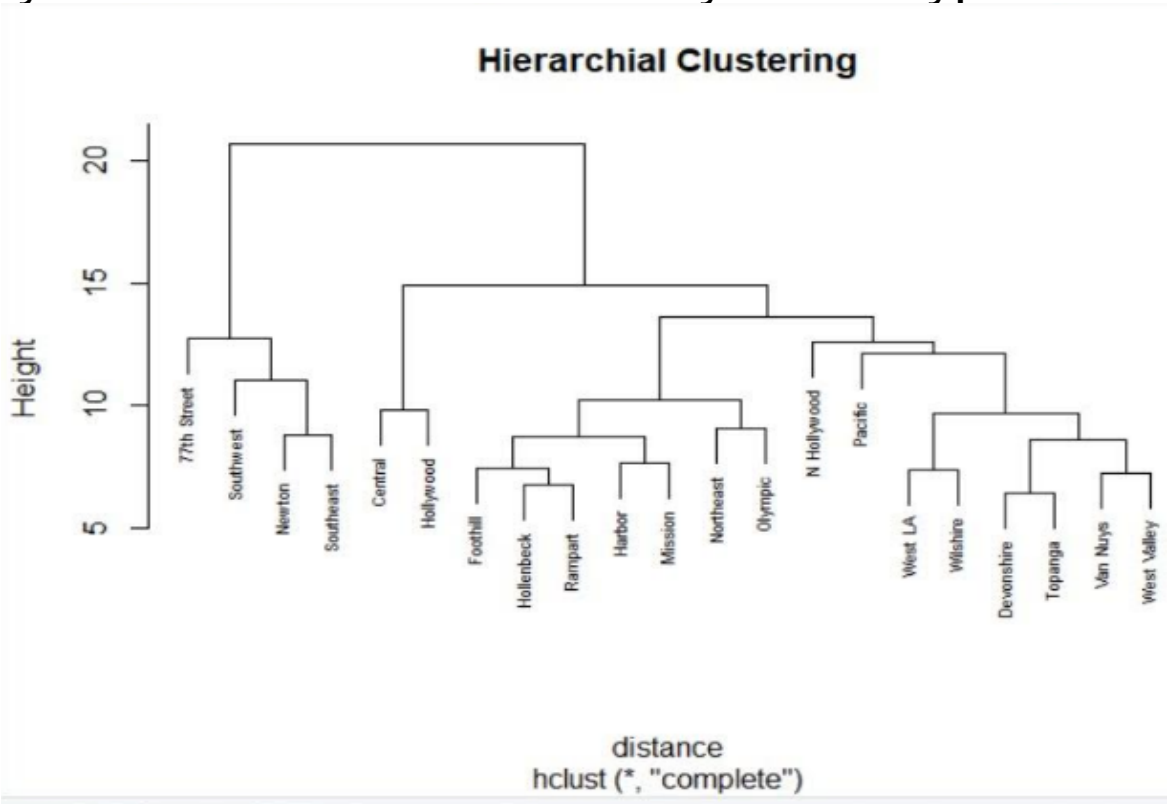
1. **Distance – Similarity , Compactness and Distinction**
2. **Average distance between points in the same cluster**
3. **Average distance between centroids / means**
4. **Average distance between points in different clusters**

Preparing for modelling

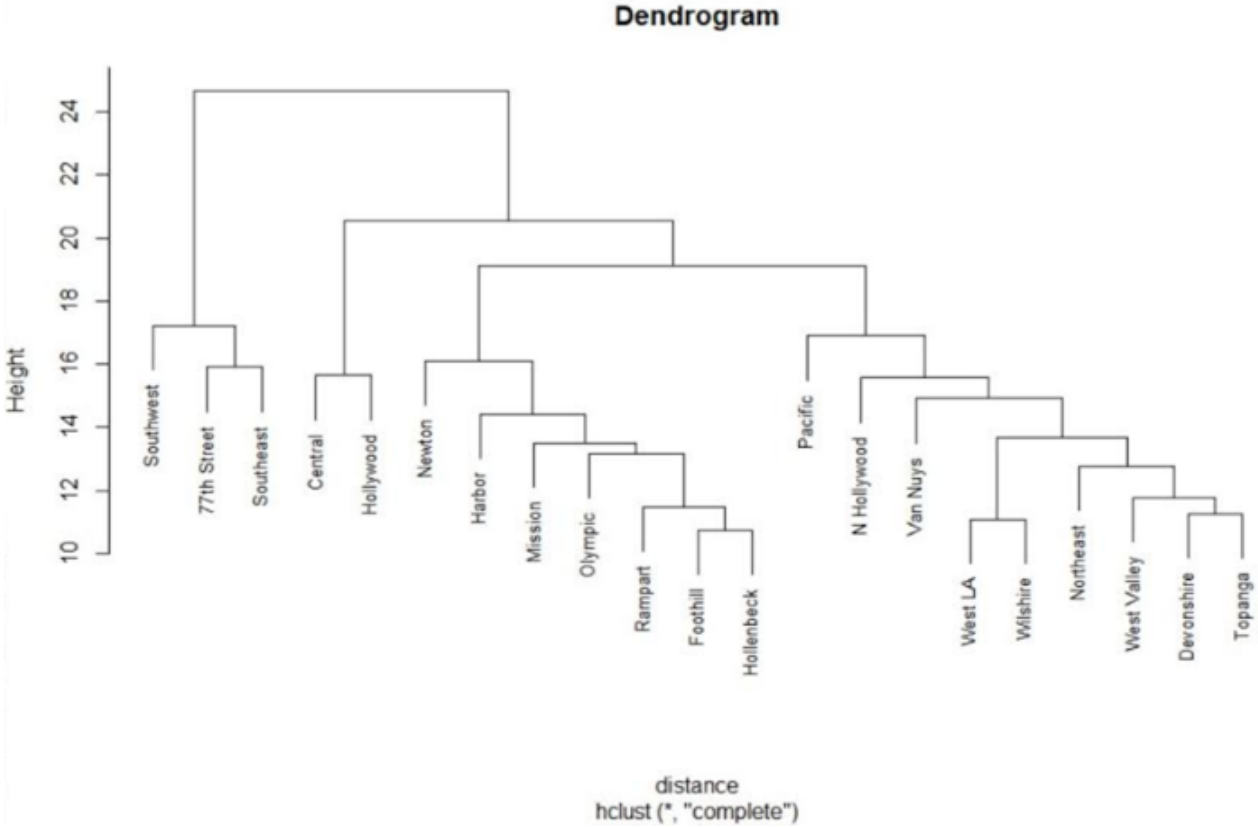
1. Select clustering method
2. Transform data into matrix
3. Normalize data, reduce dimensions
4. Determine # of clusters

Approach 1: Hierarchical, 2017 crimes

of crimes by Area.Name vs # of crimes by Crime.Type



Approach 1, Take 2: What's changed?

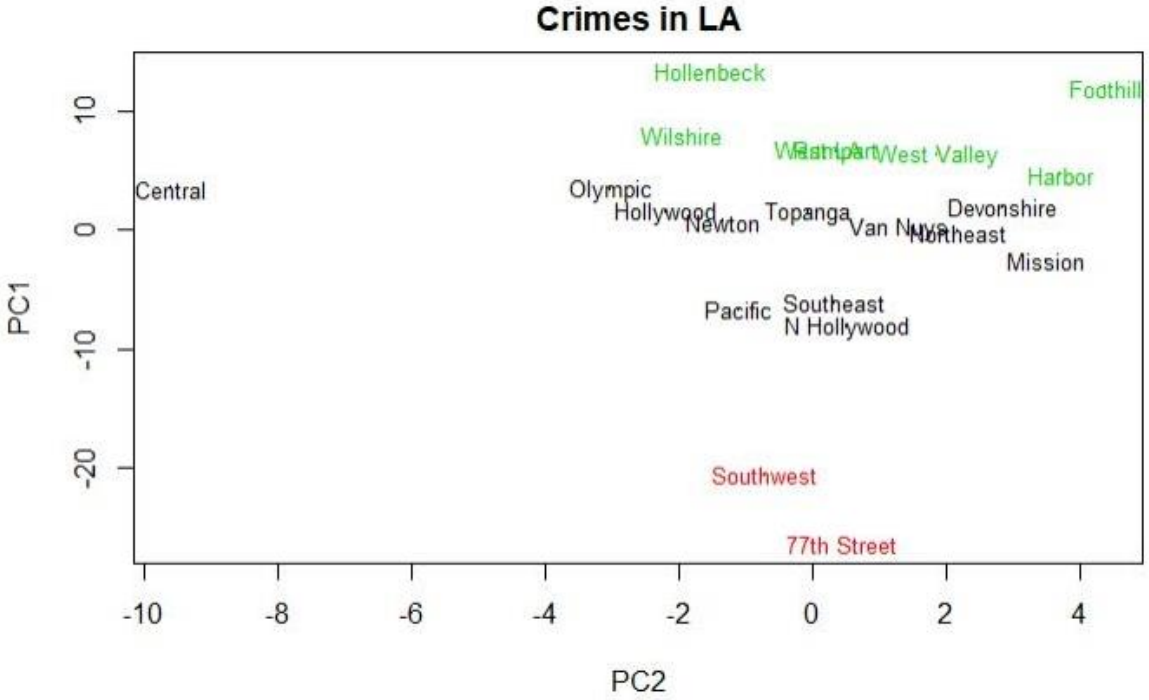


Approach 2: Clustering K-Means (K=3), 2010-18

Using first 2 PCs after running PCA on summary dataset

Area.Name <fctr>	Jan 2010 <int>	Feb 2010 <int>	Mar 2010 <int>	Apr 2010 <int>
77th Street	1278	1072	1218	1261
Central	611	473	612	561
Devonshire	971	855	976	906
Foothill	949	739	824	807
Harbor	882	729	847	772
Hollenbeck	728	558	622	628
Hollywood	853	696	778	778
Mission	1044	842	988	869
N Hollywood	1093	966	972	958
Newton	925	771	947	819

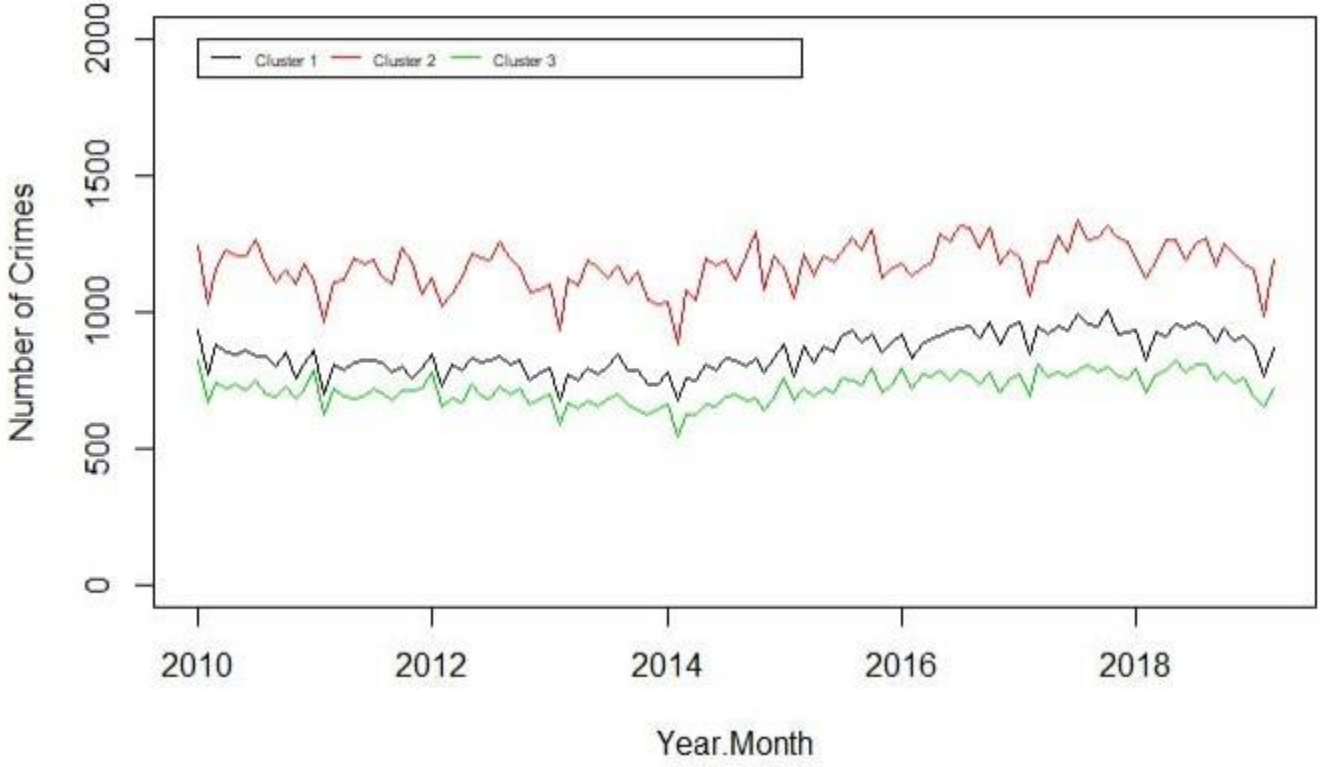
1-10 of 21 rows | 1-9 of 112 columns



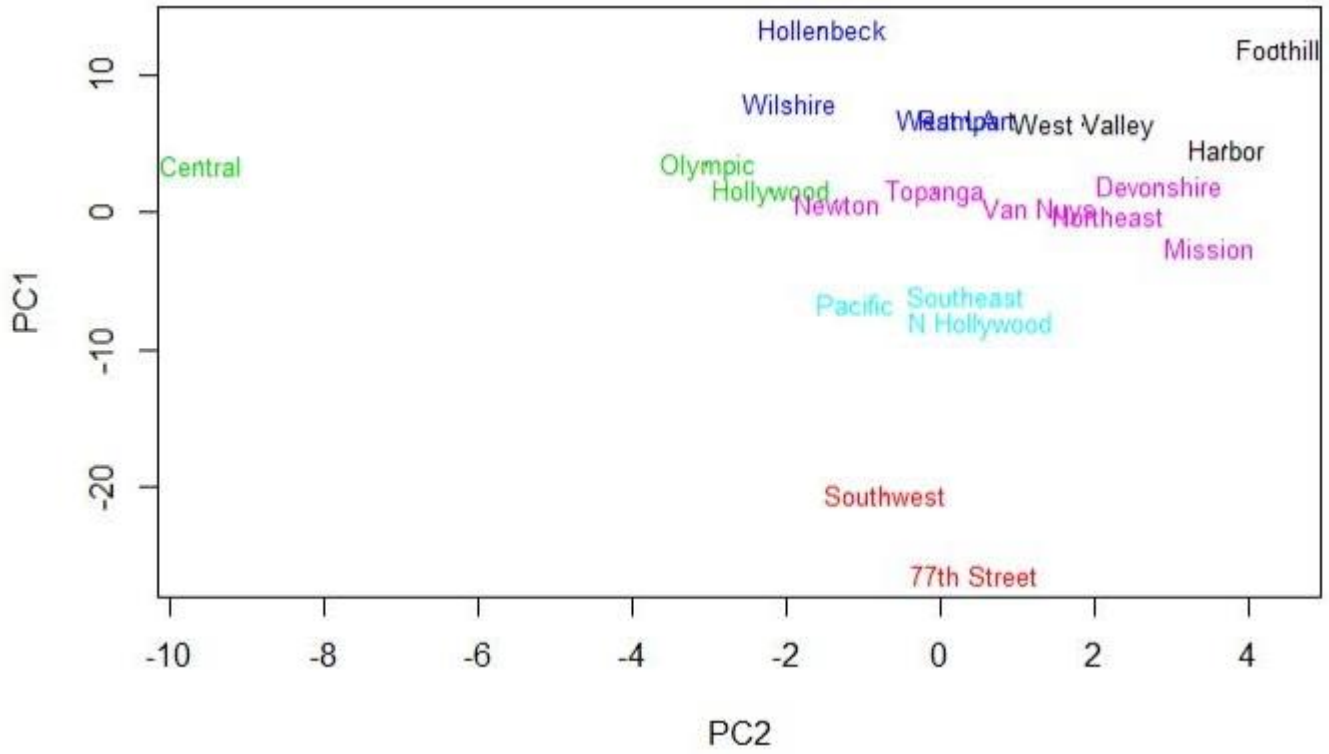
Insights:

Crime volume by cluster centroid, 2010 to present

Relative distance between high, medium, and low centroids fairly stable over time; seasonal patterns



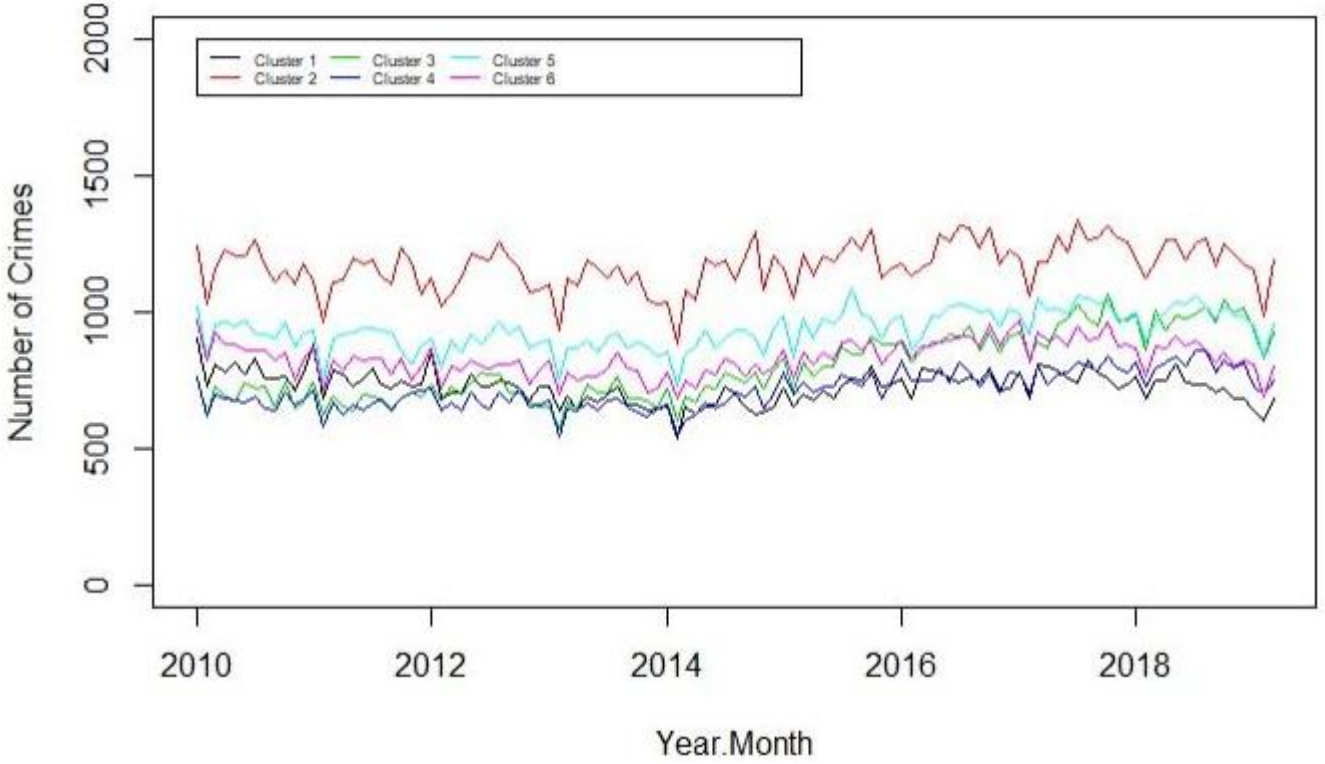
Approach 2, Take 2: K-Means (K=6)



Insights:

Crime volume by cluster centroid (K =6), 2010 to present

Using volume by month as a concept of distance is useful to differentiate between clusters based patterns that are sustained over time.



Next Steps

- Cluster validation - kmeans, kcentroid, neural gas. Adjusted rand index , CH index their meanings and comparisons
- Distance comparison and External validation
- Scaling data - relook

Questions & Answers

THANK YOU