case_7.2

May 29, 2020

0.1 Do the types of crimes committed in Chicago depend on location and time?

[1]:
```
import pandas as pd
import numpy as np
from scipy.stats import chi2_contingency
import matplotlib.pyplot as plt
from scipy.stats import chi2
```

0.2 Introduction (5 mts)

Business Context. Previously, you investigated crime data for the Chicago police department, and discovered many potential factors that could be associated with crime incidents. Now, the police department wants you to finalize your report to them so that they can start implementing some strategies based on your findings. However, because deploying a new strategy is resource intensive, they want you to confirm that the patterns you observed are not merely due to randomness.

Business Problem. The department wants you to determine: **"Are the crime patterns you observed in your prior analysis merely due to chance, or do they represent an actionable signal?"**

Analytical Context. In this case, we will learn how to perform hypothesis tests to find out if two discrete variables are independent of each other or if there are patterns between them not due to chance. This establishes if the observed interaction between such variables during exploratory data analysis are **statistically significant**. The testing procedure is usually referred to as the **Chi-square test** and it is performed on contigency tables.

The case is structured as follows: you will (1) set up the contigency table for crime type vs. location; (2) learn about the chi-square test and apply it to this pair to ascertain statistical significance of the pattern we observed during EDA; and (3) apply this test to a few other patterns we observed before.

[2]:
```
df = pd.read_csv('Chicago_crime_data.csv', dtype={'ID': object, 'beat_num':␣
↪object})
pd.options.display.max_rows = 200
```

0.3 Contingency tables (20 mts)

Recall that our original Chicago crime dataset consisted of records of individual incidents. For any given variable (e.g. primary type of crime), each incident has a particular value. For example, the first incident in the dataset is a burglary case and it happened in an apartment:

[3]:
```python
df.head()
```

[3]:
```
         ID Case Number            Date                    Block  IUCR  \
0  11192233    JB100016  12/31/17 23:58      046XX N ST LOUIS AVE   630
1  11196379    JB105867  12/31/17 23:50  024XX N LAKE SHORE DR NB   460
2  11192540    JB100551  12/31/17 23:48        001XX E SUPERIOR ST   890
3  11192239    JB100032  12/31/17 23:45           019XX S CANAL ST  1320
4  11192254    JB100003  12/31/17 23:45           115XX S STATE ST  041A

      Primary Type            Description Location Description  Arrest  \
0         BURGLARY  ATTEMPT FORCIBLE ENTRY            APARTMENT   False
1          BATTERY                  SIMPLE  MOVIE HOUSE/THEATER   False
2            THEFT           FROM BUILDING          HOTEL/MOTEL   False
3  CRIMINAL DAMAGE              TO VEHICLE               STREET   False
4          BATTERY     AGGRAVATED: HANDGUN            RESIDENCE   False

   Domestic  …  Ward  Community Area  FBI Code  X Coordinate Y Coordinate  \
0     False  …  33.0              14         5     1152214.0    1930694.0
1     False  …  43.0               7       08B     1175293.0    1916610.0
2     False  …  42.0               8         6     1177508.0    1905401.0
3      True  …  25.0              31        14     1173432.0    1891037.0
4      True  …  34.0              53       04B     1178329.0    1828012.0

   Year    Updated On   Latitude  Longitude                      Location
0  2017  5/4/18 15:51  41.965694 -87.715726  (41.965693651, -87.715726125)
1  2017  5/4/18 15:51  41.926559 -87.631294  (41.926558908, -87.631294073)
2  2017  5/4/18 15:51  41.895751 -87.623496  (41.895750913, -87.623495923)
3  2017  5/4/18 15:51  41.856427 -87.638893  (41.856426716, -87.638892854)
4  2017  5/4/18 15:51  41.683369 -87.622830  (41.683369303, -87.622829524)

[5 rows x 22 columns]
```

Recall also that we used contingency tables in order to investigate possible correlations and relationships among the different variables. The following table gives the full contingency table for `Primary Type` vs. `Location`:

[4]:
```python
type_loc_cross = pd.crosstab(df["Primary Type"], df["Location Description"])
type_loc_cross
```

[4]:
```
Location Description                ABANDONED BUILDING  AIRCRAFT  \
Primary Type
ARSON                                               10         0
```

ASSAULT	5	0
BATTERY	12	19
BURGLARY	54	0
CONCEALED CARRY LICENSE VIOLATION	0	0
CRIM SEXUAL ASSAULT	19	0
CRIMINAL DAMAGE	30	0
CRIMINAL TRESPASS	29	0
DECEPTIVE PRACTICE	2	2
GAMBLING	1	0
HOMICIDE	0	0
HUMAN TRAFFICKING	0	0
INTERFERENCE WITH PUBLIC OFFICER	6	0
INTIMIDATION	0	0
KIDNAPPING	0	0
LIQUOR LAW VIOLATION	0	0
MOTOR VEHICLE THEFT	0	0
NARCOTICS	88	1
NON-CRIMINAL	0	0
NON-CRIMINAL (SUBJECT SPECIFIED)	0	0
OBSCENITY	0	0
OFFENSE INVOLVING CHILDREN	0	0
OTHER NARCOTIC VIOLATION	0	0
OTHER OFFENSE	18	1
PROSTITUTION	0	0
PUBLIC INDECENCY	0	0
PUBLIC PEACE VIOLATION	5	15
ROBBERY	9	0
SEX OFFENSE	1	2
STALKING	0	0
THEFT	29	34
WEAPONS VIOLATION	12	0

```
Location Description              AIRPORT BUILDING NON-TERMINAL - NON-SECURE
AREA  \
Primary Type
ARSON
0
ASSAULT
6
BATTERY
5
BURGLARY
0
CONCEALED CARRY LICENSE VIOLATION
1
CRIM SEXUAL ASSAULT
0
```

CRIMINAL DAMAGE
5
CRIMINAL TRESPASS
5
DECEPTIVE PRACTICE
19
GAMBLING
0
HOMICIDE
0
HUMAN TRAFFICKING
0
INTERFERENCE WITH PUBLIC OFFICER
0
INTIMIDATION
0
KIDNAPPING
0
LIQUOR LAW VIOLATION
0
MOTOR VEHICLE THEFT
5
NARCOTICS
0
NON-CRIMINAL
0
NON-CRIMINAL (SUBJECT SPECIFIED)
0
OBSCENITY
0
OFFENSE INVOLVING CHILDREN
0
OTHER NARCOTIC VIOLATION
0
OTHER OFFENSE
5
PROSTITUTION
0
PUBLIC INDECENCY
0
PUBLIC PEACE VIOLATION
1
ROBBERY
0
SEX OFFENSE
0
STALKING

```
0
THEFT
47
WEAPONS VIOLATION
0

Location Description              AIRPORT BUILDING NON-TERMINAL - SECURE AREA
\
Primary Type
ARSON                                                                       0
ASSAULT                                                                     1
BATTERY                                                                     6
BURGLARY                                                                    0
CONCEALED CARRY LICENSE VIOLATION                                           2
CRIM SEXUAL ASSAULT                                                         1
CRIMINAL DAMAGE                                                             0
CRIMINAL TRESPASS                                                           0
DECEPTIVE PRACTICE                                                         11
GAMBLING                                                                    0
HOMICIDE                                                                    0
HUMAN TRAFFICKING                                                           0
INTERFERENCE WITH PUBLIC OFFICER                                            0
INTIMIDATION                                                                0
KIDNAPPING                                                                  0
LIQUOR LAW VIOLATION                                                        0
MOTOR VEHICLE THEFT                                                         0
NARCOTICS                                                                   2
NON-CRIMINAL                                                                1
NON-CRIMINAL (SUBJECT SPECIFIED)                                            0
OBSCENITY                                                                   0
OFFENSE INVOLVING CHILDREN                                                  0
OTHER NARCOTIC VIOLATION                                                    0
OTHER OFFENSE                                                               4
PROSTITUTION                                                                0
PUBLIC INDECENCY                                                            0
PUBLIC PEACE VIOLATION                                                      2
ROBBERY                                                                     0
SEX OFFENSE                                                                 1
STALKING                                                                    0
THEFT                                                                      49
WEAPONS VIOLATION                                                           0

Location Description              AIRPORT EXTERIOR - NON-SECURE AREA  \
Primary Type
ARSON                                                              0
ASSAULT                                                            9
BATTERY                                                           13
```

```
BURGLARY                                                             0
CONCEALED CARRY LICENSE VIOLATION                                    1
CRIM SEXUAL ASSAULT                                                  0
CRIMINAL DAMAGE                                                      2
CRIMINAL TRESPASS                                                    2
DECEPTIVE PRACTICE                                                  27
GAMBLING                                                             0
HOMICIDE                                                             0
HUMAN TRAFFICKING                                                    0
INTERFERENCE WITH PUBLIC OFFICER                                     1
INTIMIDATION                                                         0
KIDNAPPING                                                           0
LIQUOR LAW VIOLATION                                                 0
MOTOR VEHICLE THEFT                                                  9
NARCOTICS                                                            0
NON-CRIMINAL                                                         0
NON-CRIMINAL (SUBJECT SPECIFIED)                                     0
OBSCENITY                                                            0
OFFENSE INVOLVING CHILDREN                                           0
OTHER NARCOTIC VIOLATION                                             0
OTHER OFFENSE                                                        6
PROSTITUTION                                                         0
PUBLIC INDECENCY                                                     0
PUBLIC PEACE VIOLATION                                               1
ROBBERY                                                              0
SEX OFFENSE                                                          0
STALKING                                                             0
THEFT                                                               19
WEAPONS VIOLATION                                                    1

Location Description                AIRPORT EXTERIOR - SECURE AREA  \
Primary Type
ARSON                                                            0
ASSAULT                                                          2
BATTERY                                                          2
BURGLARY                                                         0
CONCEALED CARRY LICENSE VIOLATION                                0
CRIM SEXUAL ASSAULT                                              0
CRIMINAL DAMAGE                                                  2
CRIMINAL TRESPASS                                                1
DECEPTIVE PRACTICE                                               3
GAMBLING                                                         0
HOMICIDE                                                         0
HUMAN TRAFFICKING                                                0
INTERFERENCE WITH PUBLIC OFFICER                                 0
INTIMIDATION                                                     0
KIDNAPPING                                                       0
```

```
LIQUOR LAW VIOLATION                                              0
MOTOR VEHICLE THEFT                                               1
NARCOTICS                                                         0
NON-CRIMINAL                                                      0
NON-CRIMINAL (SUBJECT SPECIFIED)                                  0
OBSCENITY                                                         0
OFFENSE INVOLVING CHILDREN                                        0
OTHER NARCOTIC VIOLATION                                          0
OTHER OFFENSE                                                     2
PROSTITUTION                                                      0
PUBLIC INDECENCY                                                  0
PUBLIC PEACE VIOLATION                                            0
ROBBERY                                                           0
SEX OFFENSE                                                       1
STALKING                                                          0
THEFT                                                            11
WEAPONS VIOLATION                                                 0

Location Description                  AIRPORT PARKING LOT  \
Primary Type
ARSON                                                   0
ASSAULT                                                 8
BATTERY                                                 4
BURGLARY                                                0
CONCEALED CARRY LICENSE VIOLATION                       0
CRIM SEXUAL ASSAULT                                     0
CRIMINAL DAMAGE                                        14
CRIMINAL TRESPASS                                       1
DECEPTIVE PRACTICE                                     11
GAMBLING                                                0
HOMICIDE                                                0
HUMAN TRAFFICKING                                       0
INTERFERENCE WITH PUBLIC OFFICER                        0
INTIMIDATION                                            0
KIDNAPPING                                              0
LIQUOR LAW VIOLATION                                    0
MOTOR VEHICLE THEFT                                    12
NARCOTICS                                               1
NON-CRIMINAL                                            0
NON-CRIMINAL (SUBJECT SPECIFIED)                        0
OBSCENITY                                               0
OFFENSE INVOLVING CHILDREN                              0
OTHER NARCOTIC VIOLATION                                0
OTHER OFFENSE                                           4
PROSTITUTION                                            0
PUBLIC INDECENCY                                        0
PUBLIC PEACE VIOLATION                                  0
```

ROBBERY	0
SEX OFFENSE	0
STALKING	0
THEFT	31
WEAPONS VIOLATION	0

Location Description AIRPORT TERMINAL LOWER LEVEL - NON-SECURE AREA \

Primary Type
ARSON
0
ASSAULT
8
BATTERY
25
BURGLARY
1
CONCEALED CARRY LICENSE VIOLATION
0
CRIM SEXUAL ASSAULT
0
CRIMINAL DAMAGE
3
CRIMINAL TRESPASS
87
DECEPTIVE PRACTICE
8
GAMBLING
0
HOMICIDE
0
HUMAN TRAFFICKING
0
INTERFERENCE WITH PUBLIC OFFICER
0
INTIMIDATION
0
KIDNAPPING
0
LIQUOR LAW VIOLATION
0
MOTOR VEHICLE THEFT
0
NARCOTICS
0
NON-CRIMINAL
0

```
NON-CRIMINAL (SUBJECT SPECIFIED)
0
OBSCENITY
0
OFFENSE INVOLVING CHILDREN
1
OTHER NARCOTIC VIOLATION
0
OTHER OFFENSE
2
PROSTITUTION
0
PUBLIC INDECENCY
0
PUBLIC PEACE VIOLATION
0
ROBBERY
0
SEX OFFENSE
0
STALKING
0
THEFT
77
WEAPONS VIOLATION
0

Location Description                AIRPORT TERMINAL LOWER LEVEL - SECURE AREA  \
Primary Type
ARSON                                                                        0
ASSAULT                                                                      1
BATTERY                                                                      5
BURGLARY                                                                     0
CONCEALED CARRY LICENSE VIOLATION                                            1
CRIM SEXUAL ASSAULT                                                          0
CRIMINAL DAMAGE                                                              1
CRIMINAL TRESPASS                                                            2
DECEPTIVE PRACTICE                                                           4
GAMBLING                                                                     0
HOMICIDE                                                                     0
HUMAN TRAFFICKING                                                            0
INTERFERENCE WITH PUBLIC OFFICER                                             0
INTIMIDATION                                                                 0
KIDNAPPING                                                                   1
LIQUOR LAW VIOLATION                                                         0
MOTOR VEHICLE THEFT                                                          0
NARCOTICS                                                                    3
```

NON-CRIMINAL	1
NON-CRIMINAL (SUBJECT SPECIFIED)	0
OBSCENITY	0
OFFENSE INVOLVING CHILDREN	1
OTHER NARCOTIC VIOLATION	0
OTHER OFFENSE	0
PROSTITUTION	0
PUBLIC INDECENCY	0
PUBLIC PEACE VIOLATION	0
ROBBERY	0
SEX OFFENSE	0
STALKING	0
THEFT	37
WEAPONS VIOLATION	0

Location Description \ Primary Type	AIRPORT TERMINAL MEZZANINE - NON-SECURE AREA
ARSON	0
ASSAULT	0
BATTERY	1
BURGLARY	0
CONCEALED CARRY LICENSE VIOLATION	0
CRIM SEXUAL ASSAULT	0
CRIMINAL DAMAGE	1
CRIMINAL TRESPASS	0
DECEPTIVE PRACTICE	3
GAMBLING	0
HOMICIDE	0
HUMAN TRAFFICKING	0
INTERFERENCE WITH PUBLIC OFFICER	0
INTIMIDATION	0
KIDNAPPING	0
LIQUOR LAW VIOLATION	0
MOTOR VEHICLE THEFT	0
NARCOTICS	0
NON-CRIMINAL	0
NON-CRIMINAL (SUBJECT SPECIFIED)	0
OBSCENITY	0
OFFENSE INVOLVING CHILDREN	0
OTHER NARCOTIC VIOLATION	0
OTHER OFFENSE	0
PROSTITUTION	0
PUBLIC INDECENCY	0
PUBLIC PEACE VIOLATION	0
ROBBERY	0
SEX OFFENSE	0

STALKING	0
THEFT	6
WEAPONS VIOLATION	0

Location Description	…	VACANT LOT	VACANT LOT/LAND	\
Primary Type	…			
ARSON	…	0	7	
ASSAULT	…	0	15	
BATTERY	…	0	37	
BURGLARY	…	0	247	
CONCEALED CARRY LICENSE VIOLATION	…	0	0	
CRIM SEXUAL ASSAULT	…	0	4	
CRIMINAL DAMAGE	…	0	132	
CRIMINAL TRESPASS	…	0	21	
DECEPTIVE PRACTICE	…	0	1	
GAMBLING	…	0	1	
HOMICIDE	…	3	0	
HUMAN TRAFFICKING	…	0	0	
INTERFERENCE WITH PUBLIC OFFICER	…	0	4	
INTIMIDATION	…	0	0	
KIDNAPPING	…	0	0	
LIQUOR LAW VIOLATION	…	0	0	
MOTOR VEHICLE THEFT	…	0	35	
NARCOTICS	…	0	171	
NON-CRIMINAL	…	0	0	
NON-CRIMINAL (SUBJECT SPECIFIED)	…	0	0	
OBSCENITY	…	0	0	
OFFENSE INVOLVING CHILDREN	…	0	0	
OTHER NARCOTIC VIOLATION	…	0	0	
OTHER OFFENSE	…	0	13	
PROSTITUTION	…	0	0	
PUBLIC INDECENCY	…	0	0	
PUBLIC PEACE VIOLATION	…	0	7	
ROBBERY	…	0	19	
SEX OFFENSE	…	0	0	
STALKING	…	0	0	
THEFT	…	0	116	
WEAPONS VIOLATION	…	0	41	

Location Description	VEHICLE - DELIVERY TRUCK	\
Primary Type		
ARSON	0	
ASSAULT	1	
BATTERY	0	
BURGLARY	0	
CONCEALED CARRY LICENSE VIOLATION	0	
CRIM SEXUAL ASSAULT	0	

```
CRIMINAL DAMAGE                                             0
CRIMINAL TRESPASS                                           0
DECEPTIVE PRACTICE                                          2
GAMBLING                                                    0
HOMICIDE                                                    0
HUMAN TRAFFICKING                                           0
INTERFERENCE WITH PUBLIC OFFICER                            0
INTIMIDATION                                                0
KIDNAPPING                                                  0
LIQUOR LAW VIOLATION                                        0
MOTOR VEHICLE THEFT                                         1
NARCOTICS                                                   0
NON-CRIMINAL                                                0
NON-CRIMINAL (SUBJECT SPECIFIED)                            0
OBSCENITY                                                   0
OFFENSE INVOLVING CHILDREN                                  0
OTHER NARCOTIC VIOLATION                                    0
OTHER OFFENSE                                               0
PROSTITUTION                                                0
PUBLIC INDECENCY                                            0
PUBLIC PEACE VIOLATION                                      0
ROBBERY                                                     0
SEX OFFENSE                                                 0
STALKING                                                    0
THEFT                                                      16
WEAPONS VIOLATION                                           0

Location Description                VEHICLE - OTHER RIDE SERVICE  \
Primary Type
ARSON                                                          0
ASSAULT                                                       11
BATTERY                                                       42
BURGLARY                                                       0
CONCEALED CARRY LICENSE VIOLATION                              0
CRIM SEXUAL ASSAULT                                            2
CRIMINAL DAMAGE                                                6
CRIMINAL TRESPASS                                              0
DECEPTIVE PRACTICE                                            23
GAMBLING                                                       0
HOMICIDE                                                       0
HUMAN TRAFFICKING                                              0
INTERFERENCE WITH PUBLIC OFFICER                               1
INTIMIDATION                                                   0
KIDNAPPING                                                     0
LIQUOR LAW VIOLATION                                           0
MOTOR VEHICLE THEFT                                            0
NARCOTICS                                                      0
```

```
NON-CRIMINAL                                          1
NON-CRIMINAL (SUBJECT SPECIFIED)                      0
OBSCENITY                                             0
OFFENSE INVOLVING CHILDREN                            0
OTHER NARCOTIC VIOLATION                              0
OTHER OFFENSE                                         1
PROSTITUTION                                          0
PUBLIC INDECENCY                                      0
PUBLIC PEACE VIOLATION                                0
ROBBERY                                               9
SEX OFFENSE                                           4
STALKING                                              0
THEFT                                                47
WEAPONS VIOLATION                                     2

Location Description                   VEHICLE - OTHER RIDE SHARE SERVICE (E.G.,
UBER, LYFT)  \
Primary Type
ARSON
0
ASSAULT
0
BATTERY
4
BURGLARY
0
CONCEALED CARRY LICENSE VIOLATION
0
CRIM SEXUAL ASSAULT
2
CRIMINAL DAMAGE
0
CRIMINAL TRESPASS
0
DECEPTIVE PRACTICE
3
GAMBLING
0
HOMICIDE
0
HUMAN TRAFFICKING
0
INTERFERENCE WITH PUBLIC OFFICER
0
INTIMIDATION
0
KIDNAPPING
```

```
0
LIQUOR LAW VIOLATION
0
MOTOR VEHICLE THEFT
0
NARCOTICS
0
NON-CRIMINAL
0
NON-CRIMINAL (SUBJECT SPECIFIED)
0
OBSCENITY
0
OFFENSE INVOLVING CHILDREN
0
OTHER NARCOTIC VIOLATION
0
OTHER OFFENSE
0
PROSTITUTION
0
PUBLIC INDECENCY
0
PUBLIC PEACE VIOLATION
0
ROBBERY
3
SEX OFFENSE
2
STALKING
0
THEFT
1
WEAPONS VIOLATION
0

Location Description               VEHICLE NON-COMMERCIAL  VEHICLE-COMMERCIAL  \
Primary Type
ARSON                                                 100                   2
ASSAULT                                                98                   3
BATTERY                                               646                  21
BURGLARY                                               13                   3
CONCEALED CARRY LICENSE VIOLATION                       3                   0
CRIM SEXUAL ASSAULT                                    42                   3
CRIMINAL DAMAGE                                       543                  26
CRIMINAL TRESPASS                                      89                   4
DECEPTIVE PRACTICE                                     41                  32
```

GAMBLING	0	0
HOMICIDE	0	0
HUMAN TRAFFICKING	0	0
INTERFERENCE WITH PUBLIC OFFICER	29	2
INTIMIDATION	0	0
KIDNAPPING	6	1
LIQUOR LAW VIOLATION	1	0
MOTOR VEHICLE THEFT	174	7
NARCOTICS	621	6
NON-CRIMINAL	0	0
NON-CRIMINAL (SUBJECT SPECIFIED)	0	0
OBSCENITY	0	0
OFFENSE INVOLVING CHILDREN	18	0
OTHER NARCOTIC VIOLATION	0	0
OTHER OFFENSE	243	11
PROSTITUTION	14	0
PUBLIC INDECENCY	0	0
PUBLIC PEACE VIOLATION	6	0
ROBBERY	139	6
SEX OFFENSE	21	3
STALKING	1	0
THEFT	1675	149
WEAPONS VIOLATION	200	3

Location Description	VESTIBULE	WAREHOUSE	YARD
Primary Type			
ARSON	0	1	0
ASSAULT	0	20	0
BATTERY	0	28	0
BURGLARY	0	74	0
CONCEALED CARRY LICENSE VIOLATION	0	0	0
CRIM SEXUAL ASSAULT	0	2	0
CRIMINAL DAMAGE	0	27	0
CRIMINAL TRESPASS	0	14	0
DECEPTIVE PRACTICE	0	21	0
GAMBLING	0	0	0
HOMICIDE	1	0	22
HUMAN TRAFFICKING	0	0	0
INTERFERENCE WITH PUBLIC OFFICER	0	0	0
INTIMIDATION	0	0	0
KIDNAPPING	0	0	0
LIQUOR LAW VIOLATION	0	0	0
MOTOR VEHICLE THEFT	0	5	0
NARCOTICS	0	4	0
NON-CRIMINAL	0	0	0
NON-CRIMINAL (SUBJECT SPECIFIED)	0	0	0
OBSCENITY	0	0	0

```
OFFENSE INVOLVING CHILDREN                    0          0     0
OTHER NARCOTIC VIOLATION                      0          0     0
OTHER OFFENSE                                 0          8     0
PROSTITUTION                                  0          0     0
PUBLIC INDECENCY                              0          0     0
PUBLIC PEACE VIOLATION                        0          0     0
ROBBERY                                       0          3     0
SEX OFFENSE                                   0          1     0
STALKING                                      0          0     0
THEFT                                         0        107     0
WEAPONS VIOLATION                             0          0     0

[32 rows x 128 columns]
```

The resulting table is a bit too large; let's narrow this to focus on the most prevalent crime types and locations. The results are shown below:

[5]:
```
row_idx = df['Primary Type'].value_counts().index[:8]
col_idx = df['Location Description'].value_counts().index[:8]
type_loc_cross.loc[row_idx, col_idx]
```

[5]:
```
                    STREET  RESIDENCE  APARTMENT  SIDEWALK  OTHER  \
THEFT                15801       5048       3716      2152   3313
BATTERY               6732      10136      11706      6812   1021
CRIMINAL DAMAGE       9997       5524       3767       251    804
ASSAULT               3533       3189       2839      2189    753
DECEPTIVE PRACTICE     965       6225       1762       226   2261
OTHER OFFENSE         3451       6697       2473       593   1355
BURGLARY                49       4170       3956         6    494
ROBBERY               3548        234        235      3513    230

                    PARKING LOT/GARAGE(NON.RESID.)  RESTAURANT  \
THEFT                                         2931        3101
BATTERY                                        849         692
CRIMINAL DAMAGE                               1442         433
ASSAULT                                        499         541
DECEPTIVE PRACTICE                             201         925
OTHER OFFENSE                                  158         185
BURGLARY                                        40         312
ROBBERY                                        383         195

                    SMALL RETAIL STORE
THEFT                             4048
BATTERY                            321
CRIMINAL DAMAGE                    347
ASSAULT                            363
DECEPTIVE PRACTICE                 601
```

```
OTHER OFFENSE                     175
BURGLARY                          319
ROBBERY                           366
```

This is not the best table for determining if different values of `Primary Type` are strongly assoociated with certain values of `Location Description`, but still there is some glaring evidence. For example, out of more than 7,000 burglary cases, only six cases happened on sidewalks, in constrast to 4,170 cases that happened in residence. On the flip side, around 40% of narcotic cases happened on sidewalks.

We can extract this evidence by looking at the proportion of each location type represented for a specific crime type. This can be done by dividing each row by the row sum. For theft cases, the proportions of the top 10 locations are:

[6]:
```
round(type_loc_cross.loc["THEFT",col_idx]/type_loc_cross.loc["THEFT",:].
 ↪sum()*100,2)
```

[6]:
```
STREET                           24.55
RESIDENCE                         7.84
APARTMENT                         5.77
SIDEWALK                          3.34
OTHER                             5.15
PARKING LOT/GARAGE(NON.RESID.)    4.55
RESTAURANT                        4.82
SMALL RETAIL STORE                6.29
Name: THEFT, dtype: float64
```

Let's do this for all of the top 10 crime types:

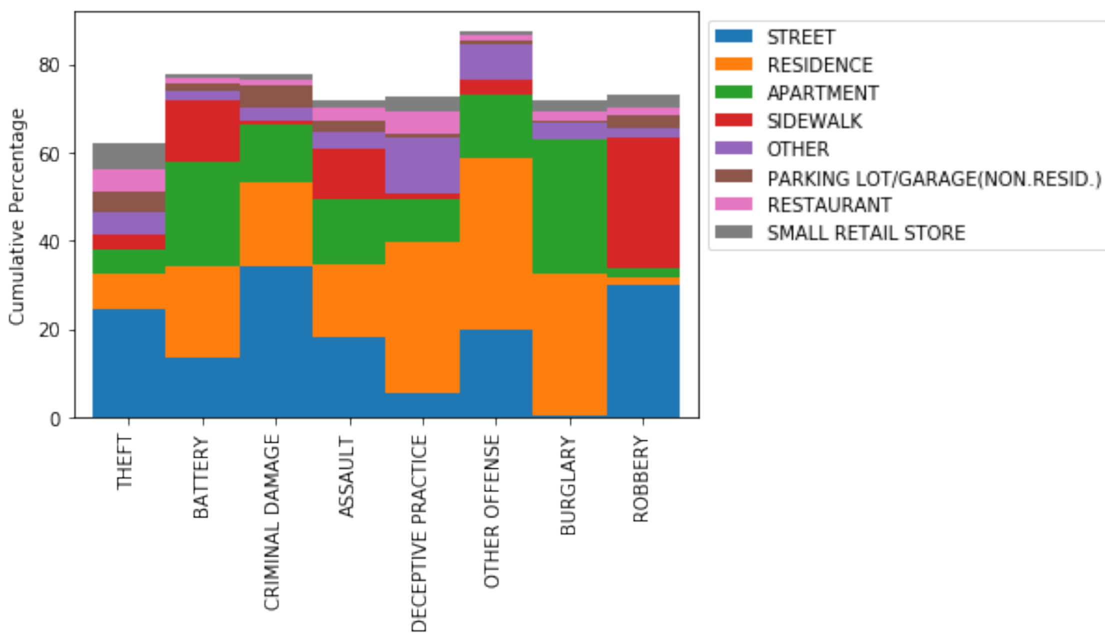
[7]:
```
type_loc_prop = round(type_loc_cross.div(type_loc_cross.sum(axis=1), axis=0).
 ↪loc[row_idx,col_idx]*100,2)
type_loc_prop
```

[7]:
```
                    STREET  RESIDENCE  APARTMENT  SIDEWALK  OTHER  \
THEFT                24.55       7.84       5.77      3.34   5.15
BATTERY              13.68      20.59      23.78     13.84   2.07
CRIMINAL DAMAGE      34.42      19.02      12.97      0.86   2.77
ASSAULT              18.30      16.52      14.71     11.34   3.90
DECEPTIVE PRACTICE    5.34      34.43       9.75      1.25  12.50
OTHER OFFENSE        20.02      38.86      14.35      3.44   7.86
BURGLARY              0.38      32.08      30.43      0.05   3.80
ROBBERY              29.87       1.97       1.98     29.57   1.94

                    PARKING LOT/GARAGE(NON.RESID.)  RESTAURANT  \
THEFT                                         4.55        4.82
BATTERY                                       1.72        1.41
CRIMINAL DAMAGE                               4.97        1.49
ASSAULT                                       2.58        2.80
```

```
DECEPTIVE PRACTICE                                     1.11         5.12
OTHER OFFENSE                                          0.92         1.07
BURGLARY                                               0.31         2.40
ROBBERY                                                3.22         1.64

                    SMALL RETAIL STORE
THEFT                             6.29
BATTERY                           0.65
CRIMINAL DAMAGE                   1.19
ASSAULT                           1.88
DECEPTIVE PRACTICE                3.32
OTHER OFFENSE                     1.02
BURGLARY                          2.45
ROBBERY                           3.08
```

We can easily spot that different types of crimes are distributed differently across locations. For example, theft cases are highly likely to happen on the street, but deceptive practices are more likely to happen in residence. We can visualize these proportions with a **stacked bar chart**, which illustrates the hotspot differences between different crime types more clearly:

[8]:
```
plt_prop = type_loc_prop.plot(kind='bar', stacked = True, width = 1)
plt_prop.legend(bbox_to_anchor=(1,1), loc='upper left', ncol = 1)
_ = plt.ylabel("Cumulative Percentage")
```



It is clear that the color compositions vary a lot across different types of crimes. We can interpret this variation as the result of crime location-type interaction. That is, types of crimes have influence on how crimes are distributed across different types of locations. If this interaction really exists, focusing on a specific location will only affect a subset of crimes, and if the target is a specific type of crime (e.g. theft), it is not enough to only focus on the most prevalent crime locations.

0.3.1 Exercise 1: (10 mts)

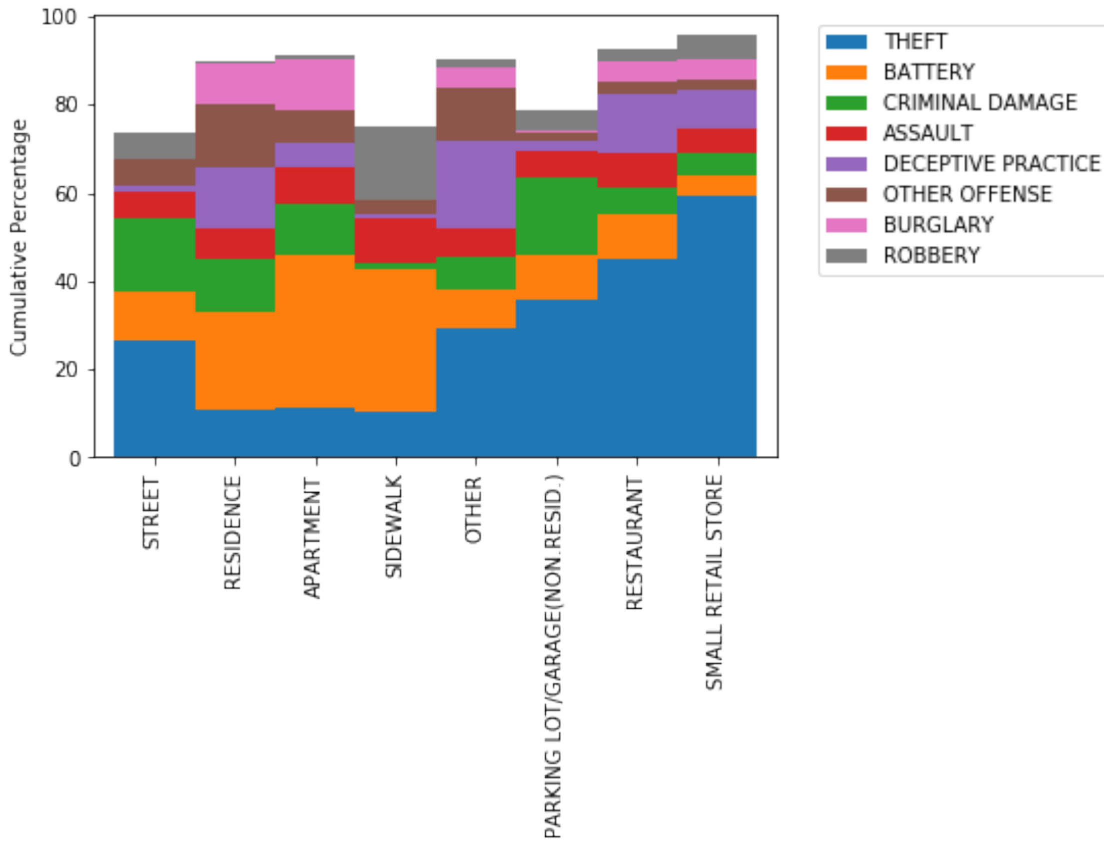
Let's flip the above script; use the above code and instead of constructing the table of proportions of crime locations for each crime type, construct the table of crime types for each crime location. Again, only include the top 10 prevalent types and locations. Plot the results with a barplot. Do your results still support crime location-type interaction?

[9]:
```
loc_type_cross = pd.crosstab(df["Location Description"], df["Primary Type"])
loc_type_prop = round(loc_type_cross.div(loc_type_cross.sum(axis=1),␣
 ↪axis=0)*100,2).loc[col_idx,row_idx]
loc_type_prop
```

[9]:
```
                              THEFT  BATTERY  CRIMINAL DAMAGE  ASSAULT  \
STREET                        26.35    11.22            16.67     5.89
RESIDENCE                     11.00    22.09            12.04     6.95
APARTMENT                     11.11    35.00            11.26     8.49
SIDEWALK                      10.24    32.43             1.19    10.42
OTHER                         29.25     9.01             7.10     6.65
PARKING LOT/GARAGE(NON.RESID.) 35.54    10.29            17.49     6.05
RESTAURANT                    44.99    10.04             6.28     7.85
SMALL RETAIL STORE            59.25     4.70             5.08     5.31

                              DECEPTIVE PRACTICE  OTHER OFFENSE  BURGLARY  \
STREET                                      1.61           5.75      0.08
RESIDENCE                                  13.57          14.60      9.09
APARTMENT                                   5.27           7.39     11.83
SIDEWALK                                    1.08           2.82      0.03
OTHER                                      19.96          11.96      4.36
PARKING LOT/GARAGE(NON.RESID.)              2.44           1.92      0.49
RESTAURANT                                 13.42           2.68      4.53
SMALL RETAIL STORE                          8.80           2.56      4.67

                              ROBBERY
STREET                           5.92
RESIDENCE                        0.51
APARTMENT                        0.70
SIDEWALK                        16.72
OTHER                            2.03
PARKING LOT/GARAGE(NON.RESID.)   4.64
RESTAURANT                       2.83
SMALL RETAIL STORE               5.36
```

[10]:
```
plt_prop_new = loc_type_prop.plot(kind='bar', stacked = True, width = 1)
plt_prop_new.legend(bbox_to_anchor=(1.5,1), loc='upper right', ncol = 1)
```

```python
_ = plt.ylabel("Cumulative Percentage")
```



We can find again the color compositions vary across different crime locations. So the results still support crime location-type interaction.

0.4 Chi-square test based on contingency tables (10 mts)

We have generated the contingency table of `Primary Type` vs. `Location Description` and observed that the crime type specific breakdowns of locations are not uniform. We conclude that there might be an interaction between these two variables. We can formally test if the variations we observed indeed reflect actual differences or if they are just a byproduct of randomness. There are many different ways to perform the test but we will focus on the most widely used test: the **Chi-square test**. The null hypothesis for the Chi-square test is:

$$H_0 : \text{Primary Type is independent of Location Description}$$

We do not need a formal definition for "independent". Intuitively, "independence between two variables" means that the distribution of values of one variable remains the same even as the value of the second variable changes (and vice versa). In our case, this means that the proportions of different crime types remains the same even as we look at different crime locations. The data seems to indicate otherwise, so let's discuss how to numerically summarize the data to formally examine the null hypothesis:

[11]:
```
type_prop = (df["Primary Type"].value_counts()/df["Primary Type"].count()).
↪sort_index()
type_prop
```

[11]:
```
ARSON                                0.001662
ASSAULT                              0.072251
BATTERY                              0.184214
BURGLARY                             0.048657
CONCEALED CARRY LICENSE VIOLATION    0.000258
CRIM SEXUAL ASSAULT                  0.006105
CRIMINAL DAMAGE                      0.108703
CRIMINAL TRESPASS                    0.025507
DECEPTIVE PRACTICE                   0.067674
GAMBLING                             0.000715
HOMICIDE                             0.002526
HUMAN TRAFFICKING                    0.000030
INTERFERENCE WITH PUBLIC OFFICER     0.004068
INTIMIDATION                         0.000565
KIDNAPPING                           0.000711
LIQUOR LAW VIOLATION                 0.000715
MOTOR VEHICLE THEFT                  0.042612
NARCOTICS                            0.043638
NON-CRIMINAL                         0.000138
NON-CRIMINAL (SUBJECT SPECIFIED)     0.000007
OBSCENITY                            0.000322
OFFENSE INVOLVING CHILDREN           0.008541
OTHER NARCOTIC VIOLATION             0.000041
OTHER OFFENSE                        0.064508
PROSTITUTION                         0.002751
PUBLIC INDECENCY                     0.000037
PUBLIC PEACE VIOLATION               0.005607
ROBBERY                              0.044461
SEX OFFENSE                          0.003859
STALKING                             0.000711
THEFT                                0.240866
WEAPONS VIOLATION                    0.017539
Name: Primary Type, dtype: float64
```

Similarly, the proportions of all distinct values of `Location Description` are:

[12]:
```
location_prop = (df["Location Description"].value_counts()/df["Location␣
↪Description"].count()).sort_index()
location_prop
```

[12]:
```
ABANDONED BUILDING                                   0.001235
AIRCRAFT                                             0.000277
AIRPORT BUILDING NON-TERMINAL - NON-SECURE AREA      0.000371
AIRPORT BUILDING NON-TERMINAL - SECURE AREA          0.000299
AIRPORT EXTERIOR - NON-SECURE AREA                   0.000341
AIRPORT EXTERIOR - SECURE AREA                       0.000094
AIRPORT PARKING LOT                                  0.000322
AIRPORT TERMINAL LOWER LEVEL - NON-SECURE AREA       0.000793
AIRPORT TERMINAL LOWER LEVEL - SECURE AREA           0.000213
AIRPORT TERMINAL MEZZANINE - NON-SECURE AREA         0.000041
AIRPORT TERMINAL UPPER LEVEL - NON-SECURE AREA       0.000296
AIRPORT TERMINAL UPPER LEVEL - SECURE AREA           0.000861
AIRPORT TRANSPORTATION SYSTEM (ATS)                  0.000034
AIRPORT VENDING ESTABLISHMENT                        0.000442
AIRPORT/AIRCRAFT                                     0.000277
ALLEY                                                0.019773
ANIMAL HOSPITAL                                      0.000180
APARTMENT                                            0.125190
APPLIANCE STORE                                      0.000344
ATHLETIC CLUB                                        0.001991
ATM (AUTOMATIC TELLER MACHINE)                       0.001995
AUTO                                                 0.000307
AUTO / BOAT / RV DEALERSHIP                          0.000052
BANK                                                 0.003859
BAR OR TAVERN                                        0.007362
BARBERSHOP                                           0.000868
BASEMENT                                             0.000004
BOAT/WATERCRAFT                                      0.000094
BOWLING ALLEY                                        0.000138
BRIDGE                                               0.000094
CAR WASH                                             0.000513
CEMETARY                                             0.000075
CHA APARTMENT                                        0.002796
CHA HALLWAY                                          0.000004
CHA HALLWAY/STAIRWELL/ELEVATOR                       0.000782
CHA PARKING LOT                                      0.000015
CHA PARKING LOT/GROUNDS                              0.002246
CHURCH                                               0.000007
CHURCH/SYNAGOGUE/PLACE OF WORSHIP                    0.002193
CLEANING STORE                                       0.000337
CLUB                                                 0.000004
COIN OPERATED MACHINE                                0.000180
COLLEGE/UNIVERSITY GROUNDS                           0.000603
COLLEGE/UNIVERSITY RESIDENCE HALL                    0.000138
COMMERCIAL / BUSINESS OFFICE                         0.006172
CONSTRUCTION SITE                                    0.001437
CONVENIENCE STORE                                    0.006816
```

CREDIT UNION	0.000060
CTA "L" PLATFORM	0.000004
CTA BUS	0.003234
CTA BUS STOP	0.002074
CTA GARAGE / OTHER PROPERTY	0.000838
CTA PLATFORM	0.003114
CTA PROPERTY	0.000004
CTA STATION	0.003462
CTA TRACKS - RIGHT OF WAY	0.000086
CTA TRAIN	0.006415
CURRENCY EXCHANGE	0.001437
DAY CARE CENTER	0.000700
DEPARTMENT STORE	0.016966
DRIVEWAY	0.000011
DRIVEWAY - RESIDENTIAL	0.002927
DRUG STORE	0.004753
FACTORY/MANUFACTURING BUILDING	0.000438
FEDERAL BUILDING	0.000124
FIRE STATION	0.000131
FOREST PRESERVE	0.000041
GANGWAY	0.000007
GARAGE	0.000011
GAS STATION	0.014537
GAS STATION DRIVE/PROP.	0.000004
GOVERNMENT BUILDING/PROPERTY	0.002126
GROCERY FOOD STORE	0.013126
HALLWAY	0.000007
HIGHWAY/EXPRESSWAY	0.000150
HOSPITAL BUILDING/GROUNDS	0.004327
HOTEL/MOTEL	0.005382
HOUSE	0.000082
JAIL / LOCK-UP FACILITY	0.000356
LAKEFRONT/WATERFRONT/RIVERBANK	0.000311
LIBRARY	0.000999
MEDICAL/DENTAL OFFICE	0.001142
MOVIE HOUSE/THEATER	0.000408
NEWSSTAND	0.000026
NURSING HOME	0.000004
NURSING HOME/RETIREMENT HOME	0.003020
OTHER	0.042395
OTHER COMMERCIAL TRANSPORTATION	0.000546
OTHER RAILROAD PROP / TRAIN DEPOT	0.000719
PARK PROPERTY	0.007793
PARKING LOT	0.000030
PARKING LOT/GARAGE(NON.RESID.)	0.030867
PAWN SHOP	0.000183
POLICE FACILITY/VEH PARKING LOT	0.003357

```
POOL ROOM                                                   0.000187
PORCH                                                       0.000090
RESIDENCE                                                   0.171721
RESIDENCE PORCH/HALLWAY                                     0.017318
RESIDENCE-GARAGE                                            0.017168
RESIDENTIAL YARD (FRONT/BACK)                               0.019878
RESTAURANT                                                  0.025799
RETAIL STORE                                                0.000007
RIVER BANK                                                  0.000004
ROOMING HOUSE                                               0.000004
SAVINGS AND LOAN                                            0.000026
SCHOOL YARD                                                 0.000007
SCHOOL, PRIVATE, BUILDING                                   0.001991
SCHOOL, PRIVATE, GROUNDS                                    0.000726
SCHOOL, PUBLIC, BUILDING                                    0.012771
SCHOOL, PUBLIC, GROUNDS                                     0.003937
SIDEWALK                                                    0.078622
SMALL RETAIL STORE                                          0.025571
SPORTS ARENA/STADIUM                                        0.000939
STAIRWELL                                                   0.000011
STREET                                                      0.224476
TAVERN                                                      0.000007
TAVERN/LIQUOR STORE                                         0.002006
TAXICAB                                                     0.001576
VACANT LOT                                                  0.000011
VACANT LOT/LAND                                             0.003260
VEHICLE - DELIVERY TRUCK                                    0.000075
VEHICLE - OTHER RIDE SERVICE                                0.000558
VEHICLE - OTHER RIDE SHARE SERVICE (E.G., UBER, LYFT)       0.000056
VEHICLE NON-COMMERCIAL                                      0.017677
VEHICLE-COMMERCIAL                                          0.001055
VESTIBULE                                                   0.000004
WAREHOUSE                                                   0.001179
YARD                                                        0.000082
Name: Location Description, dtype: float64
```

Under the null hypothesis, the expected numbers of occurrences for all pairwise combinations of values of `Primary Type` and `Location Description` should be the product of the corresponding values in the above two tables, times the total number of records in the dataset. For example, the total number of occurrences of battery crime in apartments should be approximately 0.184214 times 0.125190 times the total number of crime instances in our entire dataset.

[13]:
```python
primary_location_cross = pd.crosstab(df['Primary Type'], df['Location'])
g, p, dof, expctd = chi2_contingency(primary_location_cross)
print("p-value of Chi-square test for Primary Type vs. Location =", p)
```

```
p-value of Chi-square test for Primary Type vs. Location = 0.0
```

We can see the p - value is extremely small and thus reject the null hypothesis and conclude that `Primary Type` and `Location Description` are not independent. In other words, the proportions of distinct value of `Primary Type` do not remain the same across different values of `Location Description`, which is exactly what we observed in the data.

0.5 Chi-square test for primary type vs. day of week (20 mts)

Sometimes, when we perform the Chi-square test, one of the variables (or even both of them) is not naturally discrete (for example, crime time). However, we can discretize the variable and perform the Chi-square test on the discretized versions. We will now discretize the time variable into day-of-the-week buckets and test if the day of the week is independent of crime types. This test will inform us if we should vary police force deployment according to the day of week. Let's get started:

```
[ ]: # discretize time
df["date_py"] = pd.to_datetime(df.Date)
df["day_of_week"] = df.date_py.dt.dayofweek
type_dow_cross = pd.crosstab(df['Primary Type'], df['day_of_week'])
type_dow_cross
```

The following code gives the result of performing a Chi-square test:

```
[ ]: g, p, dof, expctd = chi2_contingency(type_dow_cross)
print("p-value of Chi-square test for Primary Type vs. Day of week =", p)
```

The results indicate that `Primary Type` and `day_of_week` are not independent. Let's visualize the distribution of the top 10 crimes for each day of the week with a stacked bar chart:

```
[ ]: type_dow_plt_dat = round(type_dow_cross.div(type_dow_cross.sum(axis=0), axis=1).
 ↪loc[row_idx,:]*100,2).T
plt_type_dow = type_dow_plt_dat.plot(kind='bar', stacked = True, rot = 0)
plt_type_dow.legend(bbox_to_anchor=(1.5,1), loc='upper right', ncol = 1)
_ = plt.ylabel("Cumulative Percentage")
```

From this, we can see that battery tends to be more prevalent on Fridays and Saturdays, while theft tends to decrease on Saturdays.

0.5.1 Exercise 2: (10 mts)

We suspect that throughout the course of a typical day, the distribution of crime locations may shift materially. Conduct a test to determine if this is the case. If this is the case, identify the potential shift by constructing a stacked bar chart that shows the proportion of crimes in each of the top 10 locations for each hour of the day.

Answer. We discretize the time into hours of the day and perform the Chi-square test:

[ ]:
```
df["hour_of_day"] = df.date_py.dt.hour
hod_loc_cross = pd.crosstab(df['hour_of_day'], df['Location Description'])
g, p, dof, expctd = chi2_contingency(hod_loc_cross)
print("Test for independence of crime locations and hour of the day: p-value␣
 ↪=", p)
```

We find that location and time of day are indeed dependent. Let's visualize this:

[ ]:
```
hod_loc_plt_dat = round(hod_loc_cross.div(hod_loc_cross.sum(axis=1), axis=0).
 ↪loc[:,col_idx]*100,2)
plt_hod_loc = hod_loc_plt_dat.plot(kind='bar', stacked = True, rot = 0)
plt_hod_loc.legend(bbox_to_anchor=(1.7,1), loc='upper right', ncol = 1)
_ = plt.ylabel("Cumulative Percentage")
```

We can see that in the morning (5AM - 12PM), crimes in residence (orange) tend to increase and after 6PM, crimes begin to flock towards the streets.

0.6 Conclusions (5 mts)

In this case, we performed the Chi-square test to validate various patterns and relationships that we observed between various features in our previous EDA of Chicago crime incidents. This test provided statistical evidence that the pattern we saw in the contingency tables previously was not just due to chance. This provides strong backing for the police department to take the big step of reorganizing their force in line with our observations.

0.7 Takeaways (5 mts)

In this case, we learned the concept of feature indepedence and learned how to perform Chi-square tests to examine if two discrete factors are independent. The Chi-square test helps statistically validate the patterns we observe from exploratory data analysis.