NYU TANDON

Crimes Across New York City

Final Presentation

Kewei Yu, Manas Johri, Sara Kou, Shihang Zhu
& Utkarsha Chourasia

Team Members	Roles
Kewei Yu	Organizer
Manas Johri	Agenda Maker
Sara Kou	Task Manager
Shihang Zhu	Analyst
Utkarsha Chourasia	Long-Term Strategy

Introduction & Background

- We aim to create a platform for **NYU Students** that can help them summarize **areas of interest while moving into the city.**

- We realized that the main concerns students have while moving here are related to **Crime, followed by Health Facilities, Transportation Services** and so on.

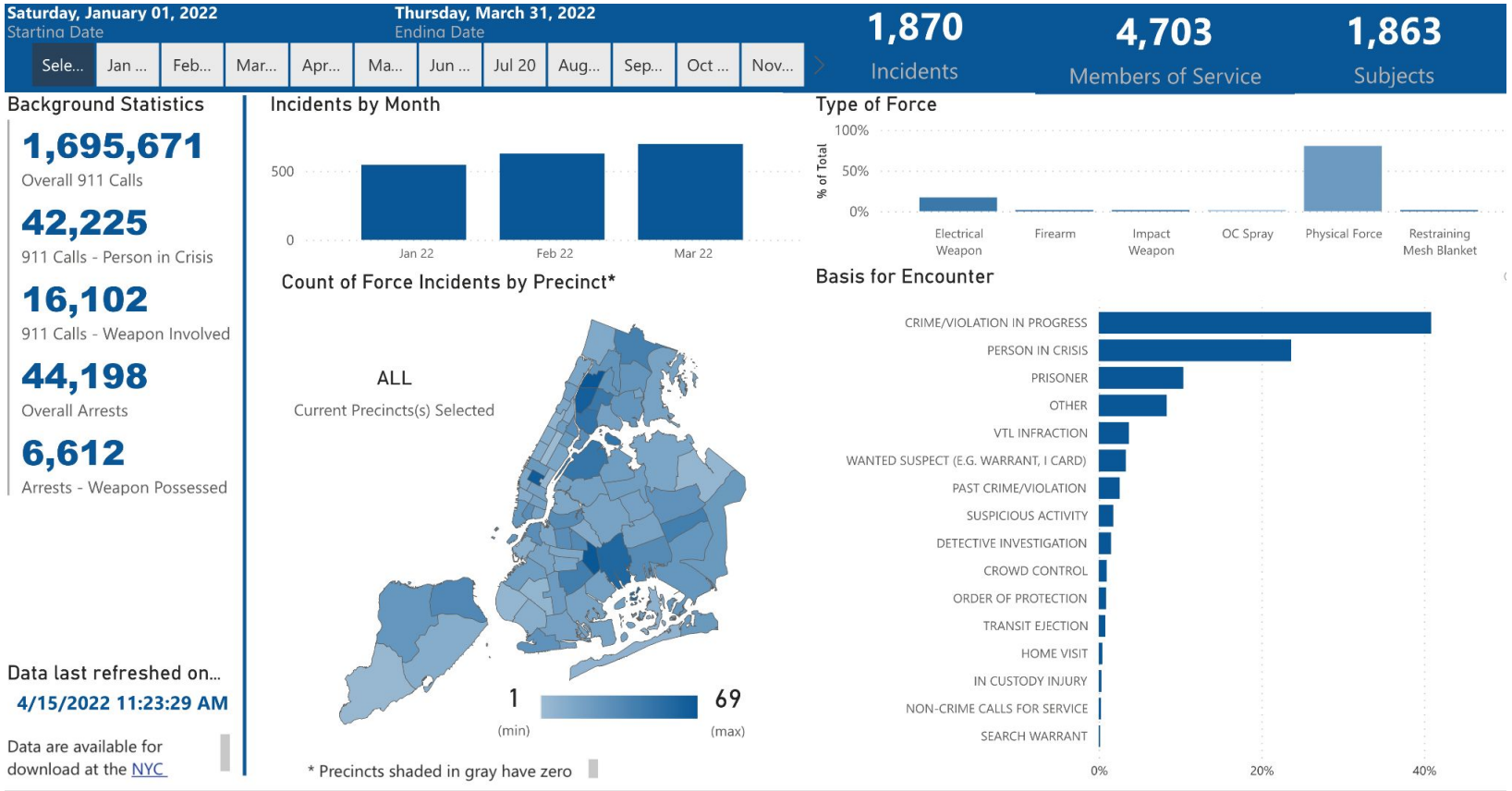
- Hence, we created an **Interactive Dashboard (Tableau Story)** that covers **'Crime Across NYC'**, using data from **Open NYC and NYPD Data Sources.**

- Additionally, we incorporated a few more facilities related to **Health and Transportation** into our Dashboard for a broader usage.

The Ultimate Goal: Help and benefit the population in NYC, especially NYU students.

Motivation

Rising Crime - According to NYPD statistics, “for the month of January 2022, New York City saw a 38.5% increase in overall index crime compared with January 2021 (9,566 v. 6,905)”.



Tools used throughout the project

- We have primarily used Tableau for our Project. Excel was used for Data Filtering, Sorting and Tinkering

1. Tableau - Data visualization

2. Microsoft Excel - Data filtering and sorting

3. R - Data processing and filtering

4. Power BI - Open for potential use

Data Sets Used

- **We have used 5 Data Sets extracted from NYC OpenData, namely -**

NYPD Arrests Data (Historic) Public Safety

List of every arrest in NYC going back to 2006 through the end of the previous calendar year. This is a breakdown of every arrest effected in NYC by the NYPD going back to 2006 through the end of the previous calendar year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. In addition, information related to suspect demographics is also included.
This data can be used by the public to explore the nature of police enforcement activity. Please refer to the attached data footnotes for additional information about this dataset.

NYPD Complaint Data Historic Public Safety

This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of last year (2019). For additional details, please see the attached data dictionary in the 'About' section.

Athletic Facilities

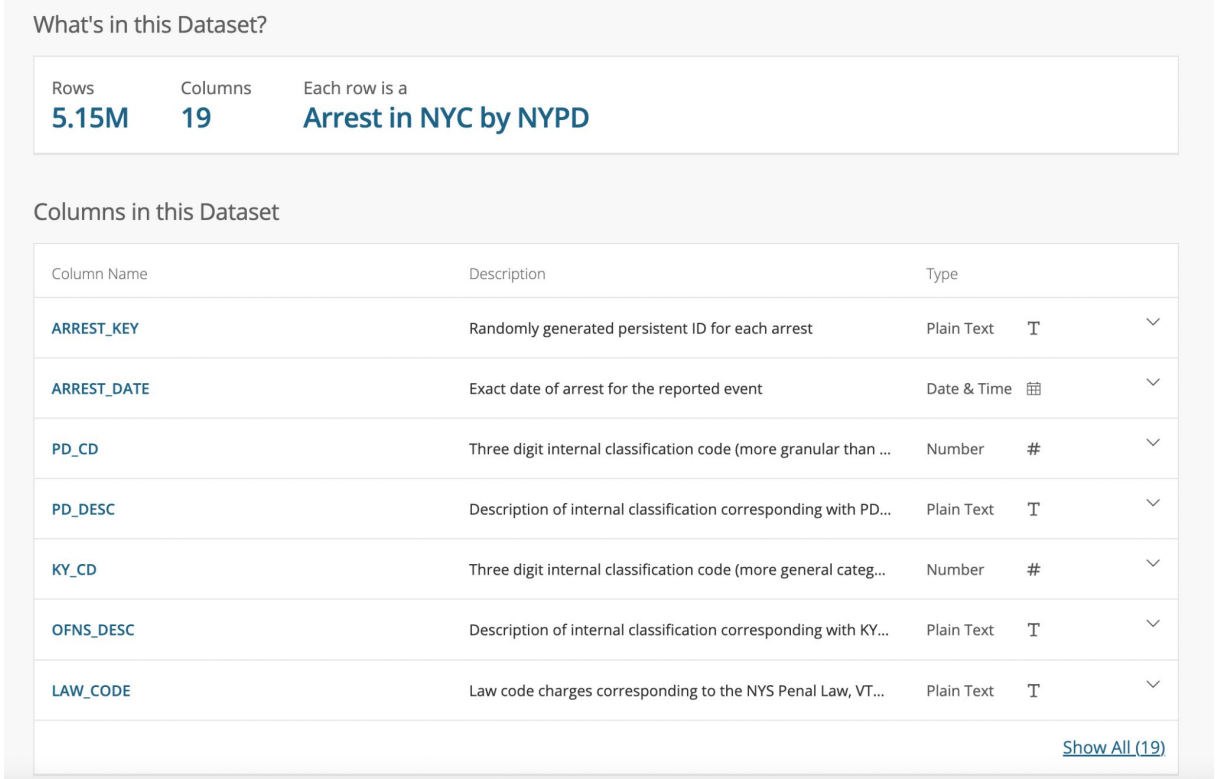
This dataset contains facilities that were designed specifically with sports in mind. These facilities are not the only places within parks where sports can be played. This layer contains all of the facilities that can be permitted for athletic activity and facilities that are specifically designated for sports but are not available for permitting.

NYC Health + Hospitals patient satisfaction scores – 2009 Health

These are patient satisfaction scores measured by a standardized survey known as the Hospital Consumer Assessment of Healthcare Providers & Systems (HCAHPS). The survey is conducted by the federal Centers for Medicare and Medicaid Services (CMS) as a standard assessment tool for all hospitals throughout the nation. This data shows the 2009 patient satisfaction scores in New York City's 11 public hospitals.

Subway Entrances

Map of NYC Subway Entrances



NYPD Arrest Data(Historical)

- 5.15 millions rows/19 columns
- Offence description/Arrest Boroughs/Arrest Date/Age Group

Example of datatable using R

Show 5 entries Search:

	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DESC	LAW_CODE	LAW_CAT_CD
1	32311380	06/18/2007	511	CONTROLLED SUBSTANCE, POSSESSION 7	235	DANGEROUS DRUGS	PL 2200300	M
2	192799737	01/26/2019	177	SEXUAL ABUSE	116	SEX CRIMES	PL 1306503	F
3	193260691	02/06/2019					PL 2203400	F
4	149117452	01/06/2016	153	RAPE 3	104	RAPE	PL 1302503	F
5	190049060	11/15/2018	157	RAPE 1	104	RAPE	PL 1303501	F

Showing 1 to 5 of 100 entries Previous 1 2 3 4 5 … 20 Next

NYU TANDON

Data Cleaning and Exploration

```
'data.frame':   5153369 obs. of  19 variables:
 $ ARREST_KEY        : int  32311380 192799737 193260691 149117452 190049060 24288194 189182271 196324211 19
197554056 ...
 $ ARREST_DATE       : chr  "06/18/2007" "01/26/2019" "02/06/2019" "01/06/2016" ...
 $ PD_CD             : int  511 177 NA 153 157 203 153 157 175 175 ...
 $ PD_DESC           : chr  "CONTROLLED SUBSTANCE, POSSESSION 7" "SEXUAL ABUSE" "" "RAPE 3" ...
 $ KY_CD             : int  235 116 NA 104 104 352 104 104 233 233 ...
 $ OFNS_DESC         : chr  "DANGEROUS DRUGS" "SEX CRIMES" "" "RAPE" ...
 $ LAW_CODE          : chr  "PL 2200300" "PL 1306503" "PL 2203400" "PL 1302503" ...
 $ LAW_CAT_CD        : chr  "M" "F" "F" "F" ...
 $ ARREST_BORO       : chr  "Q" "M" "M" "K" ...
 $ ARREST_PRECINCT   : int  27 25 14 67 77 77 5 77 50 26 ...
 $ JURISDICTION_CODE: int  1 0 0 0 0 2 0 0 0 0 ...
 $ AGE_GROUP         : chr  "18-24" "45-64" "25-44" "25-44" ...
 $ PERP_SEX          : chr  "M" "M" "M" "M" ...
 $ PERP_RACE         : chr  "BLACK" "BLACK" "UNKNOWN" "BLACK" ...
 $ X_COORD_CD        : num  NA 1000555 986685 998032 1003606 ...
 $ Y_COORD_CD        : num  NA 230994 215375 175598 185050 ...
 $ Latitude          : num  NA 40.8 40.8 40.6 40.7 ...
 $ Longitude         : num  NA -73.9 -74 -74 -73.9 ...
 $ Lon_Lat           : chr  "" "POINT (-73.94110928599997 40.800694331000045)" "POINT (-73.99121211099998
40.75783900300007)" "POINT (-73.95033556299995 40.648650085000035)" ...
```

NYU TANDON

Data Cleaning and Exploration

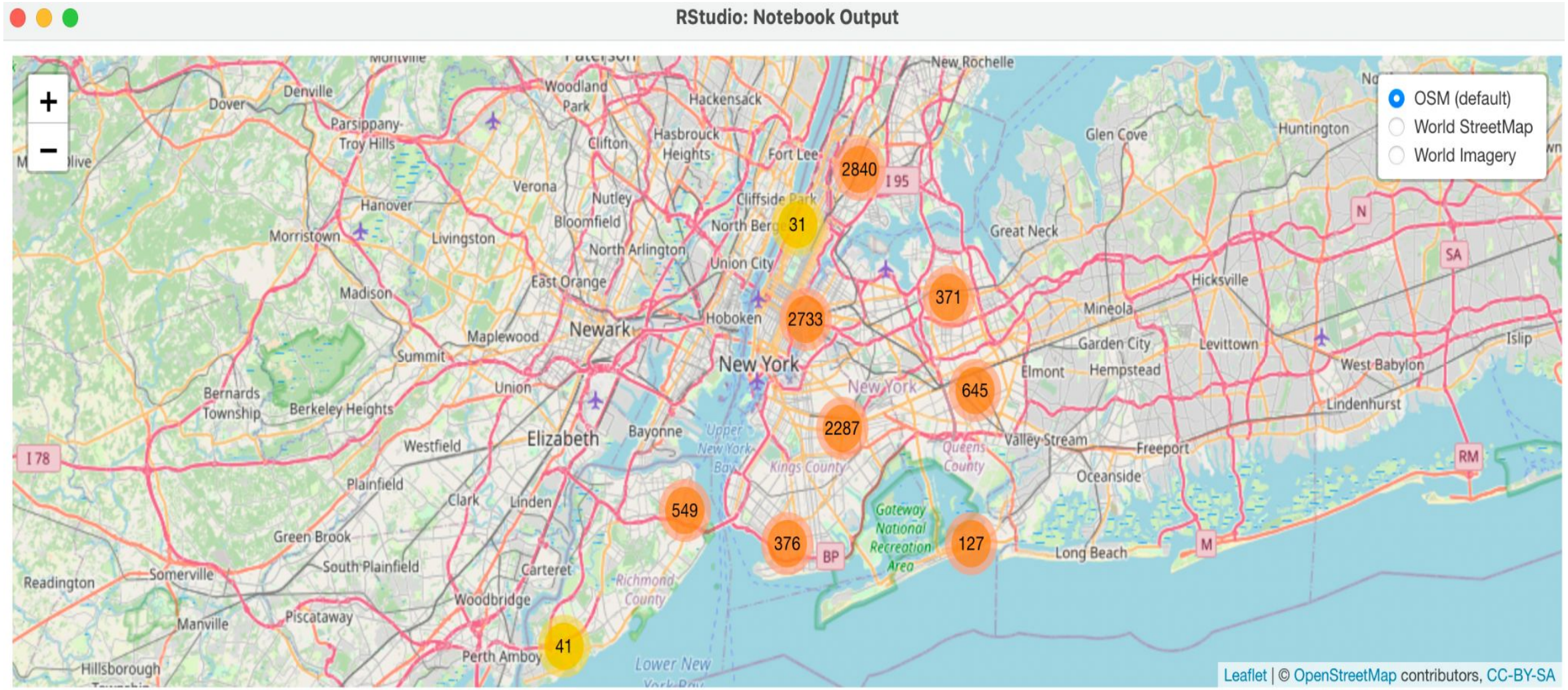
```r
summary(df2)
```

```
  ARREST_KEY        ARREST_DATE            PD_CD        PD_DESC              KY_CD       OFNS_DESC
Min.   :  9926901   Length:5153369     Min.   :  0.0   Length:5153369     Min.   :101    Length:5153369
1st Qu.: 60181681   Class :character   1st Qu.:269.0   Class :character   1st Qu.:126    Class :character
Median : 84499756   Mode  :character   Median :511.0   Mode  :character   Median :341    Mode  :character
Mean   : 99029881                      Mean   :508.8                      Mean   :300
3rd Qu.:146381373                      3rd Qu.:748.0                      3rd Qu.:348
Max.   :222500606                      Max.   :997.0                      Max.   :995
                                       NA's   :284                        NA's   :9066
  LAW_CODE          LAW_CAT_CD        ARREST_BORO        ARREST_PRECINCT JURISDICTION_CODE
Length:5153369     Length:5153369     Length:5153369     Min.   :  1.0   Min.   : 0.000
Class :character   Class :character   Class :character   1st Qu.: 33.0   1st Qu.: 0.000
Mode  :character   Mode  :character   Mode  :character   Median : 60.0   Median : 0.000
                                                         Mean   : 60.7   Mean   : 1.307
                                                         3rd Qu.: 84.0   3rd Qu.: 0.000
                                                         Max.   :123.0   Max.   :97.000
                                                                         NA's   :10
 AGE_GROUP           PERP_SEX          PERP_RACE          X_COORD_CD       Y_COORD_CD        Latitude
Length:5153369     Length:5153369     Length:5153369     Min.   : 913357   Min.   : 121131   Min.   :40.50
Class :character   Class :character   Class :character   1st Qu.: 993347   1st Qu.: 186857   1st Qu.:40.68
Mode  :character   Mode  :character   Mode  :character   Median :1004904   Median : 209391   Median :40.74
                                                         Mean   :1005366   Mean   : 214781   Mean   :40.76
                                                         3rd Qu.:1015896   3rd Qu.: 236614   3rd Qu.:40.82
                                                         Max.   :1067302   Max.   :8202360   Max.   :62.08
                                                         NA's   :1         NA's   :1         NA's   :1
  Longitude        Lon_Lat
Min.   :-74.25   Length:5153369
1st Qu.:-73.97   Class :character
Median :-73.93   Mode  :character
Mean   :-73.92
3rd Qu.:-73.89
Max.   :-73.68
NA's   :1
```

Remove null value and outlier in age group

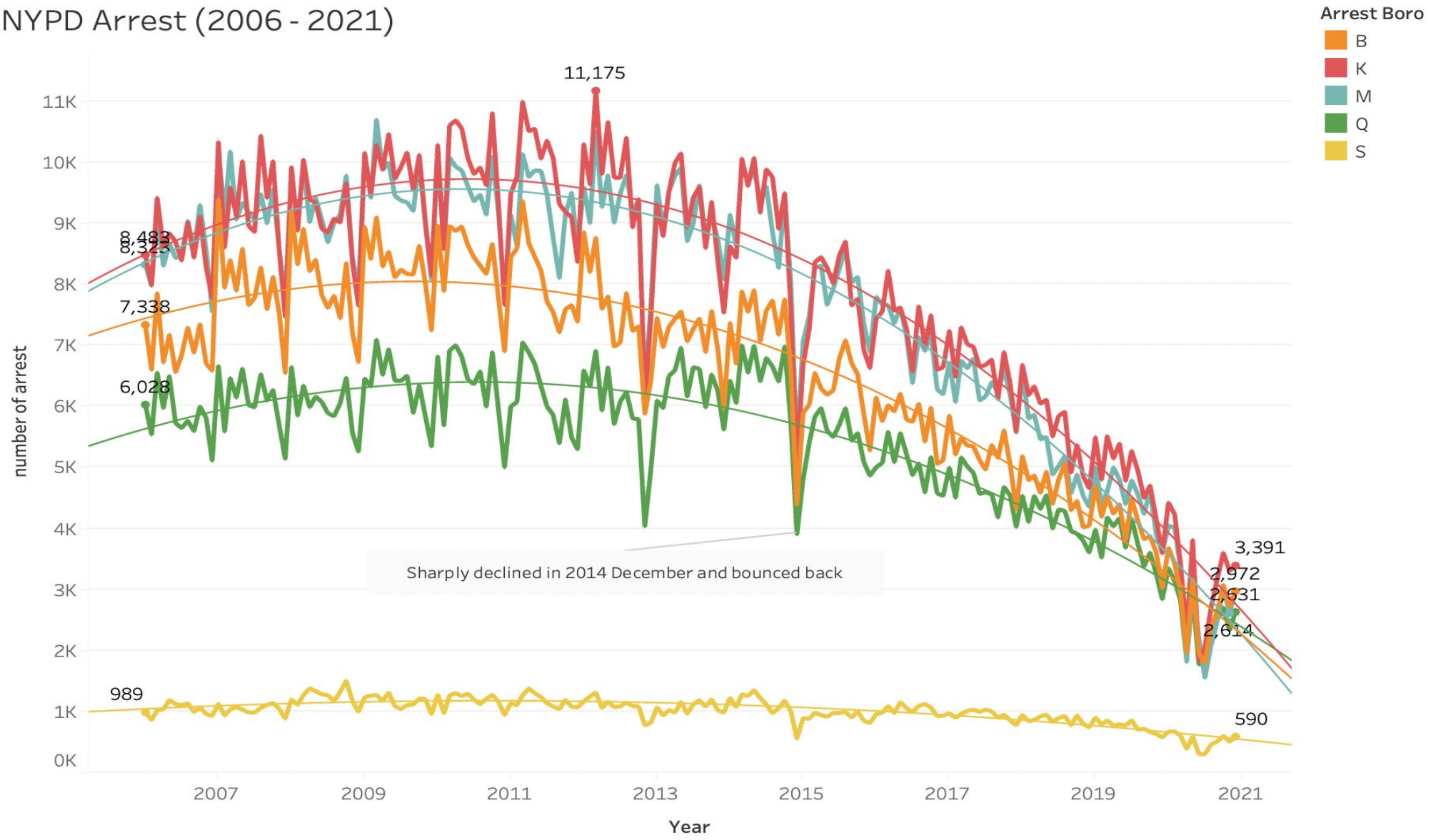
NYU TANDON

Using R for the same goal



NYU TANDON

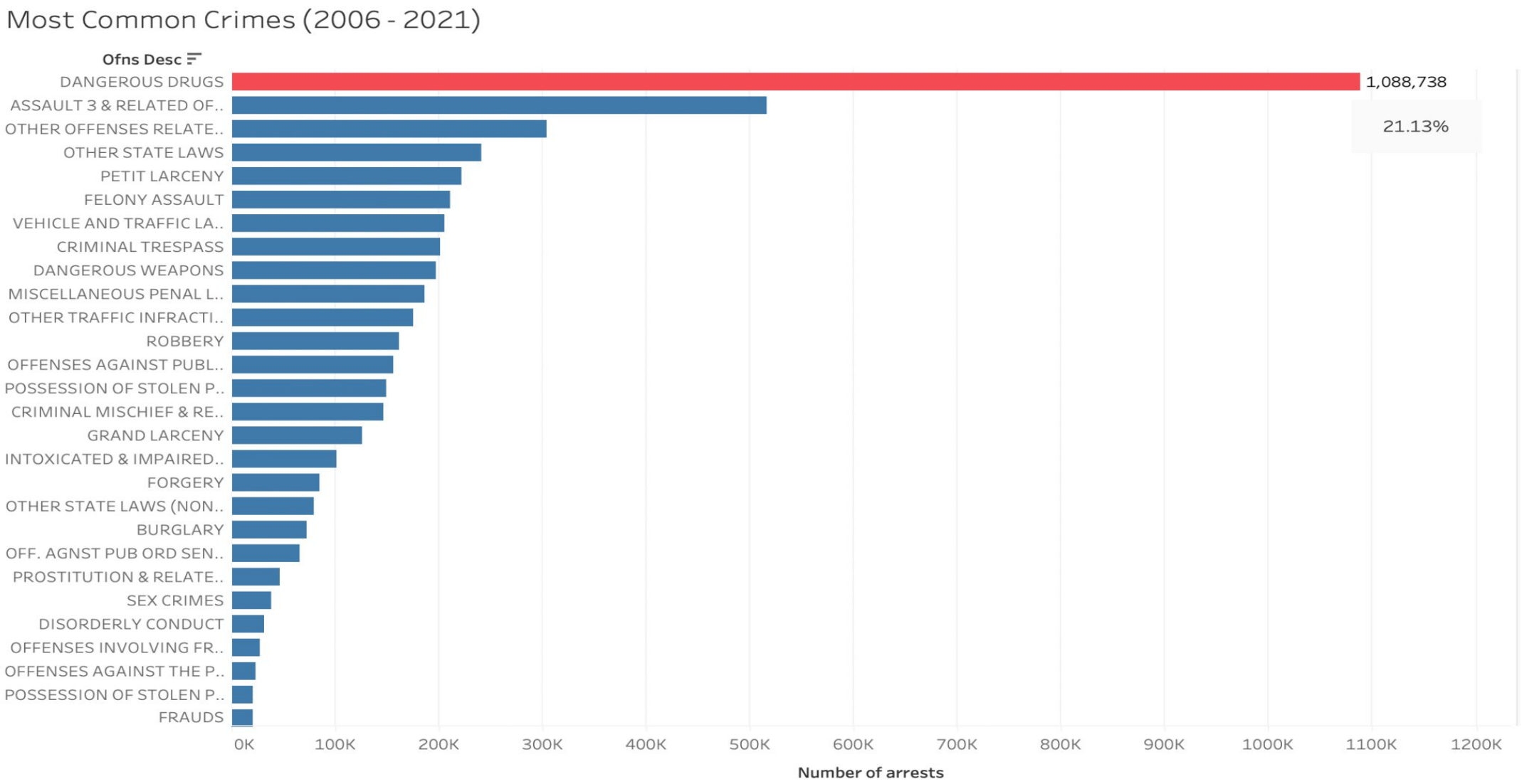
Map visualization in R



NYU TANDON

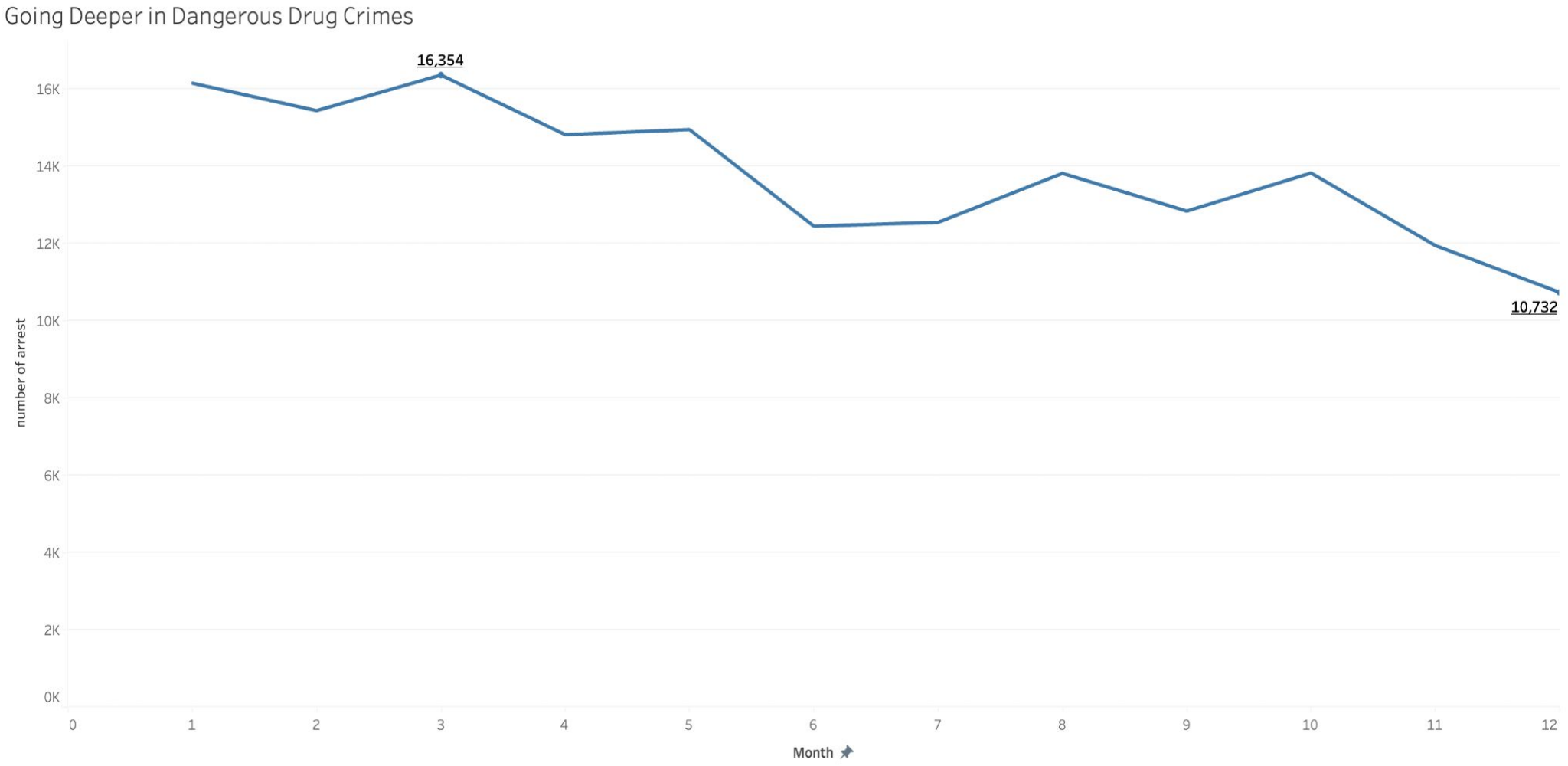
Citywide crimes continued to fall

Drugs related crimes were prevalent over last 15 years

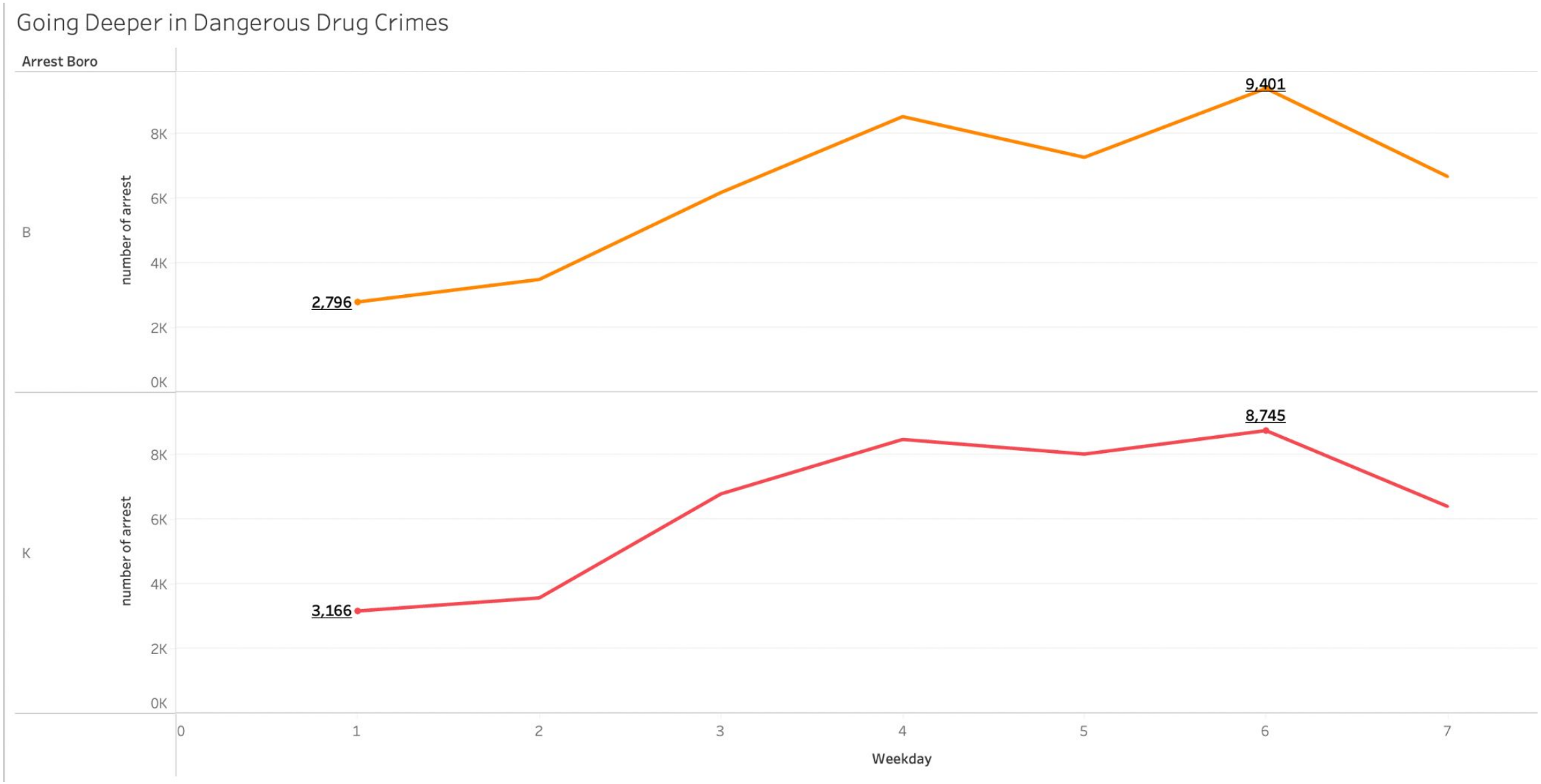


NYU | TANDON

Drugs crimes occur in peak period in March and fall gradually till December



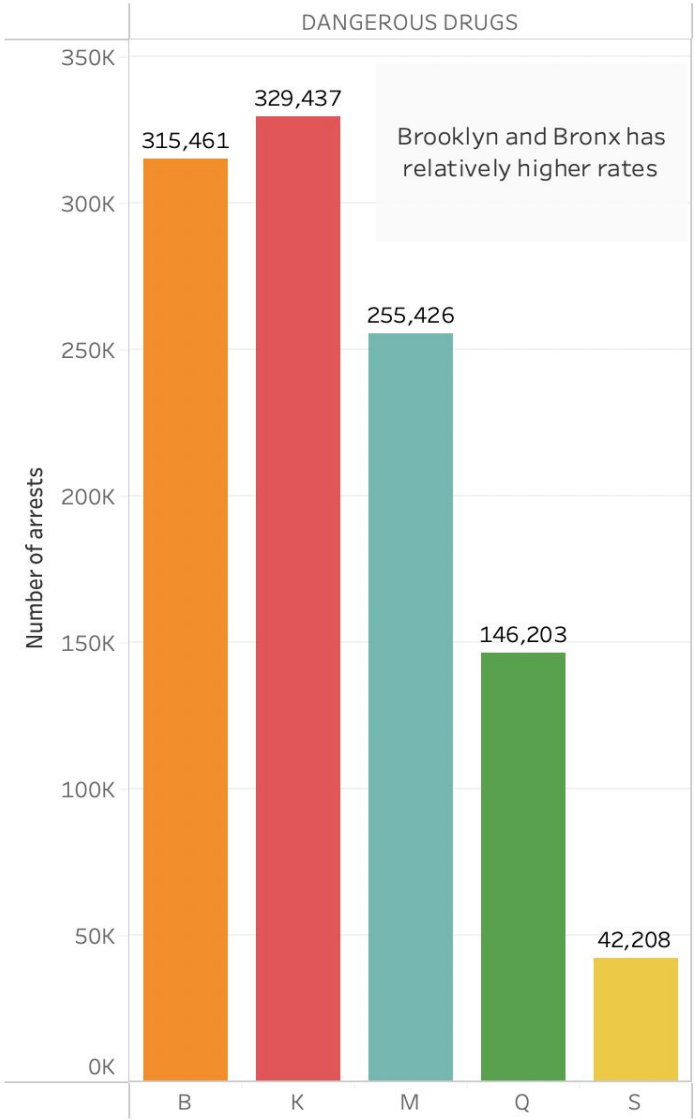
NYU TANDON



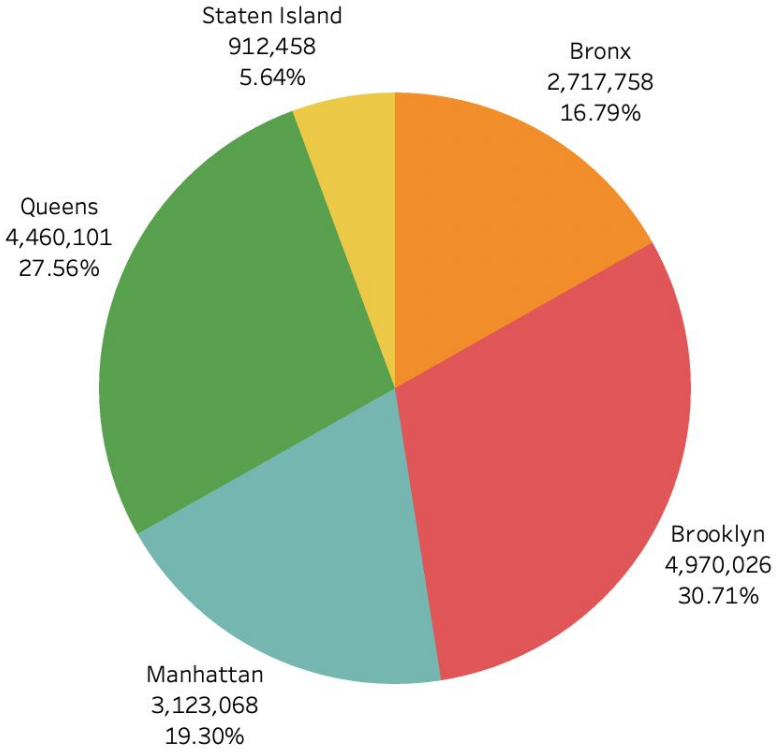
NYU TANDON

Drug crimes most likely happen on Saturday and least on Monday

Going Deeper in Dangerous Drug Crimes



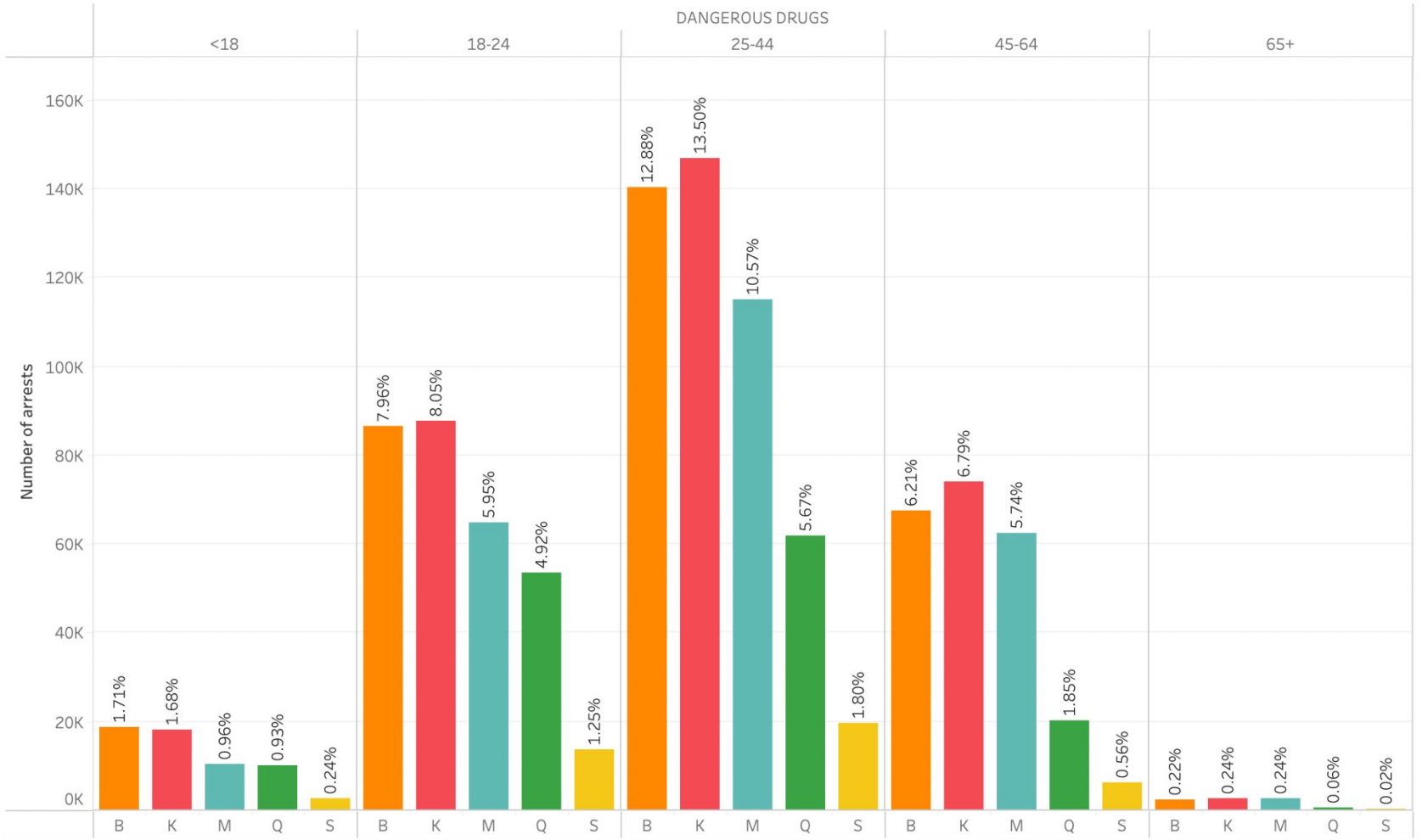
NYC Population in 2020



NYU TANDON

After normalization, Bronx has the highest crime rate (11.61%), Manhattan has the second (8.18%) and Brooklyn has the third (6.63%).

Going Deeper in Dangerous Drug Crimes

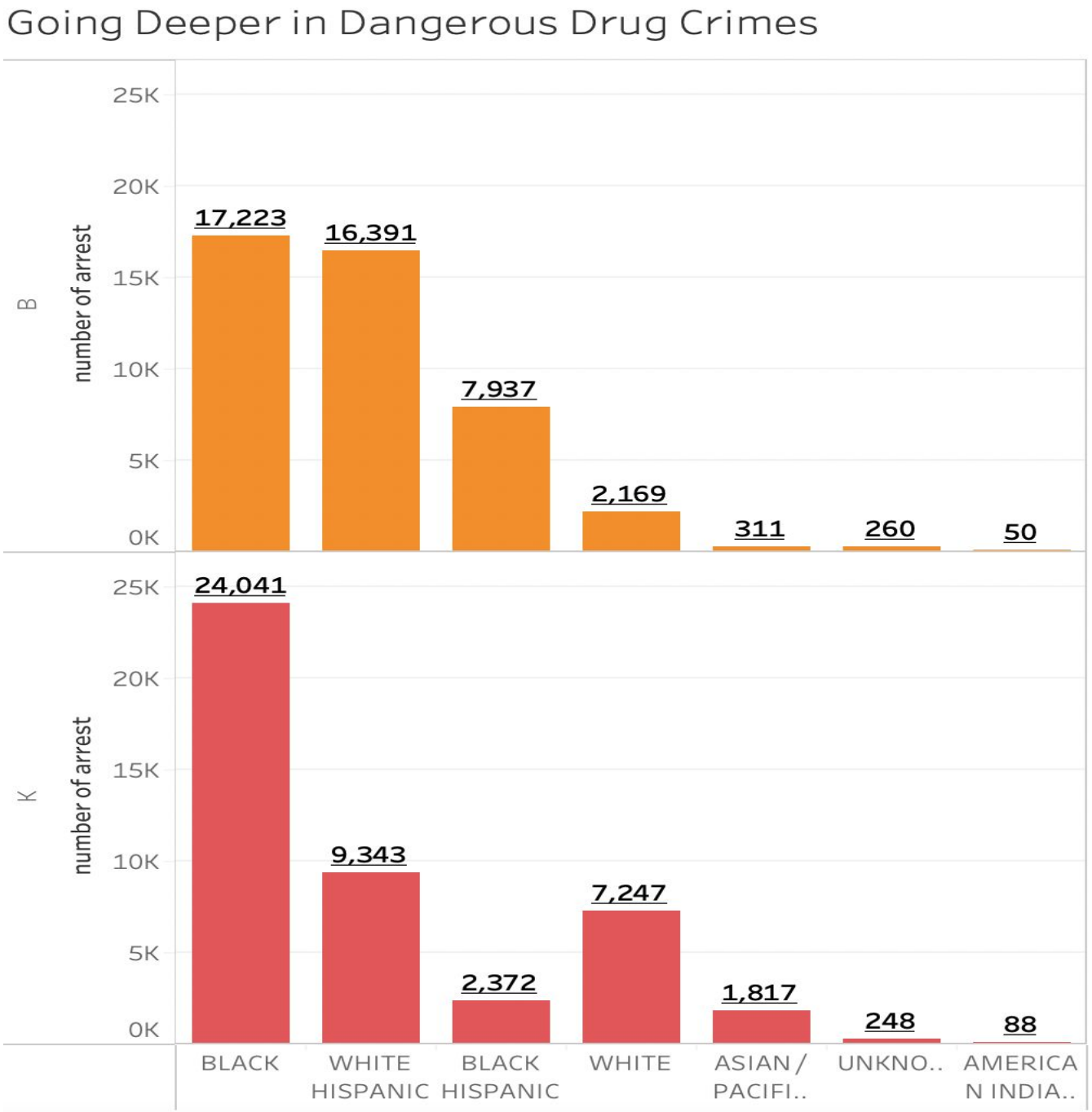


Most criminal's age fall in the range from 25 to 44

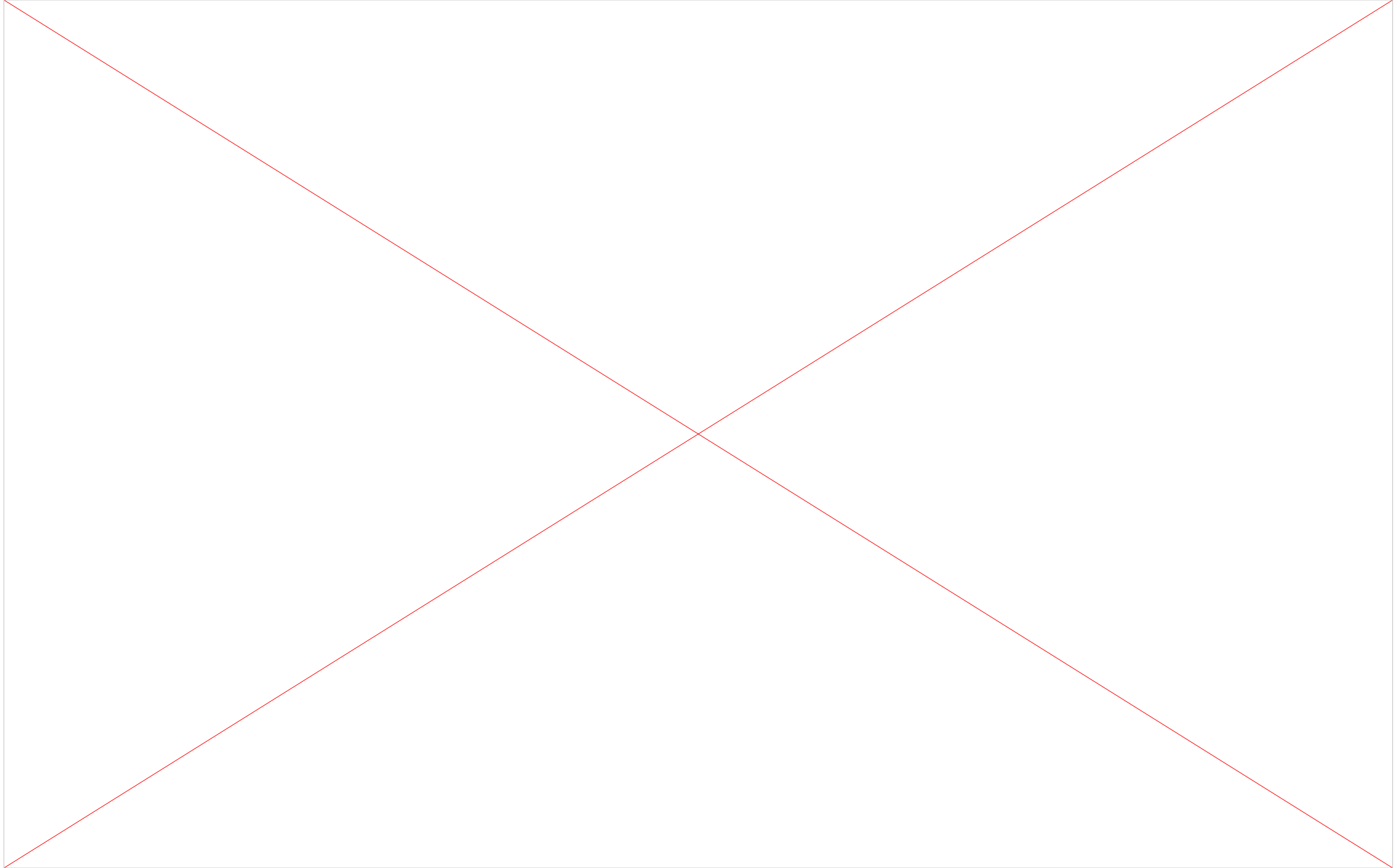
Drug crimes related to teenager and elderly are very little compared to the whole group

NYU TANDON

Black and White Hispanic are the majority in the drug crimes



NYU TANDON



Summary

- We have successfully created a **Tableau Dashboard** which covers all of the major aspects that students have while moving into NYC.
- We have made it as **interactive** and **minimal** to use as possible, such that anybody can summarize their areas of interest while hunting for apartments/dorms.
- Our project is **open to suggestions**, we believe it has **immense potential** and can be easily **integrated onto the NYU website**.
- By covering all of **Crime, Hospital Facilities, Athletic Facilities as well as Subway Entrances**, we have created a tool which (*if published on the NYU website*) can be immensely helpful for the upcoming cohorts.

References

- NYPD Arrest Data (Historical)
- NYPD Force Dashboard
- NYC Athletic Facilities Map

Thank you.

Crimes Across New York City

Kewei Yu, Manas Johri, Sara Kou, Shihang Zhu & Utkarsha Chourasia

NYU | TANDON

Determining Key Process Metrics

- **The overall process you will follow for the entire data visualization.**
  - Collection of data from reliable sources
  - Cleaning and sorting of relevant data from this collection
  - Correlation Analysis/ hypothesis testing using R
  - Create Dashboard with Tableau
  - Put necessary interactive slides and button for areas of focus
  - Presenting data with accordance with Melissa Anderson's Principles of Design to get best results.

- **A description of your data including how you obtained it.**
  - Define the purpose and know our audiences
  - Collect and clean data from NYC Open Data website(NYC Compliant Data in 2021, year to date) and NYPD Arrests Data(Historical)
  - Use Tableau and R to visualize data across space and time
  - Develop possible regression model and correlation analysis
  - Conclude our findings

Team Charter

Team Members	Email IDs	Roles
Kewei Yu	ky957@nyu.edu	Organizer
Manas Johri	maj8626@nyu.edu	Agenda Maker
Sara Kou	xxg200@nyu.edu	Task Manager
Shihang Zhu	sz3496@nyu.edu	Analyst
Utkarsha Chourasia	uc2012@nyu.edu	Long-Term Strategy