Final Paper - Open Data
What data on police conduct tells us about reforms needed in Chicago

#1 **Download datasets**

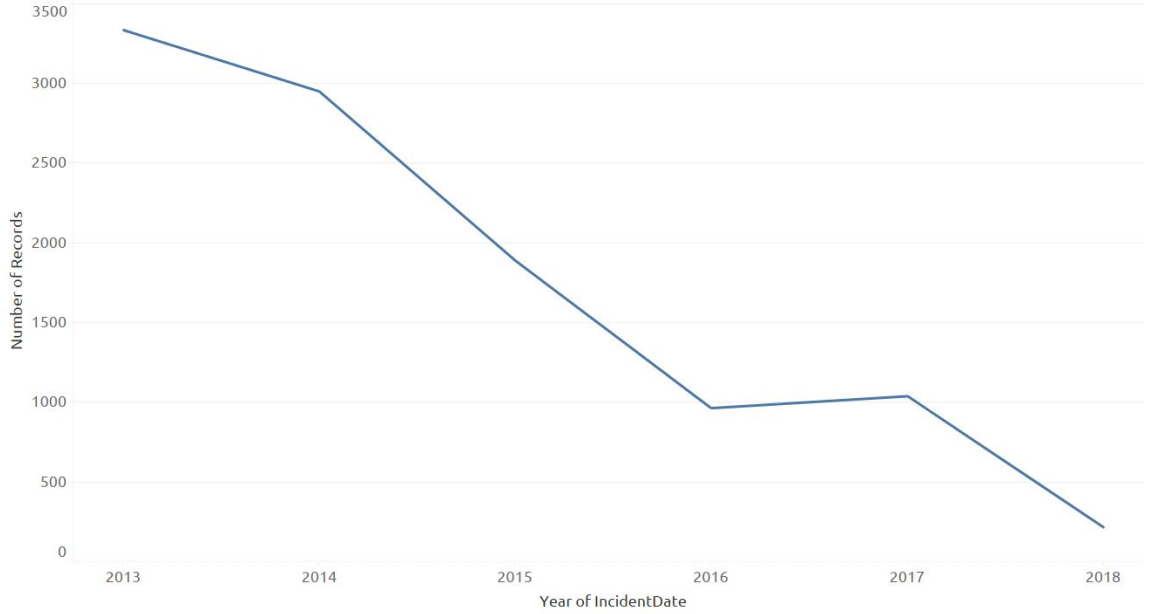
a. *Citizens Police Data Project*:
https://data.cpdp.co/data/DR8zlN/citizens-police-data-project
Download complaints for Category 'Use of Force'
Download complaints for Category 'Illegal Search'

b. *Boundaries - Census Tracts - 2010*:
https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik
Download Shapefile

c. *2013-2017 American Community Survey 5-Year Estimates*:
https://factfinder.census.gov
Search 'Census Tract for Cook County, Illinois'
Search 'Research/Ethnicity Individual'

A look at the Citizens Police Data Project dataset

The dataset that has been released by the city of Chicago since 2016 comprises over 100,000 formal complaints submitted to Chicago Police Department since 1997. However, it is a grassroot organization, the Citizens Police Data Project (CPDP), that has made these data easily accessible to the wider public. In terms of Completeness, it therefore appears that not all records have been disclosed by the city. With regards to timeliness, it is also unclear if the data is made available to the public as soon as possible.



As the graph shows, the n. of complaints stored in the database goes down year by year. We could assume that the city of Chicago prefers to release older records. On the positive side, the data score high in terms of granularity but some columns in the dataset , such as 'Beat', need some complementary information.
The CPDP website is easy to navigate but it is not possible to download the entire dataset at once. Only one category at the time can be downloaded. Moreover, each dataset

downloaded stores information in different sheets of the same Excel file. Connecting and crossing the information present in different sheets cannot easily be done by anyone. These details limit the reusability of the datasets.

#2 **Clean Citizens Police Data Project datasets**

a. Merge together 'Use of Force' and 'Illegal Search'
b. Delete sheets:
    'DISCLAIMER'
    'Police Witnesses'
c. In sheet 'Allegations' keep columns:
    CRID
    OfficerID
    Category
    Allegation
    Outcome
    Incident Date
    Latitude
    Longitude
d. Modify Timestamp in column 'Incident' Date to keep only the year
    Function Text-to-Columns
e. Make more uniform the Allegation categories
    Function Text-to-Columns
f. Keep only complaints from 2013 onwards to consider a 5-year range
g. Move information in 'Officer Profile' in 'Allegations':
    Using function VLOOKUP match OfficerID column in 'Allegations' with OfficerID in 'Officer Profile column and create 4 new columns:
    Officer Gender
    Officer Race
    Officer Rank
    Officer Age
h. Delete sheet 'Officer Profile'
i. Save information contained in 'Complaining Witnesses' sheet in a separate file
j. Delete sheet 'Complaining Witnesses'

#3 **Upload cleaned '*AGGREGATED_Complaints_Details*' file in Tableau Public**

a. Perform a left join in Tableau between the Complaints file and the previously saved '*AGGREGATED_Witnesses_Details*'

#4 **Visualize data**

a. Visualizations can be consulted on https://public.tableau.com/profile/maria.giorda#!/vizhome/OpenDataDashboard/Sheet4

#5 **Clean *2013-2017 American Community Survey 5-Year Estimates (ACS)***

a. Keep only columns:
    Geography
    Estimate Population - White
    Estimate Population - Black

Estimate Population - Native American
Estimate Population - Asian
Estimate Population - Latino

b. Isolate Census Tract Number in Geography column
Function Text-to-Columns
c. Create a Total Pop column
d. Create a Percent Population column for each race
e. Create a column called Predominant Ethnicity that returns the column header of the largest value amongst Percent Population by Race
Use together functions INDEX, MATCH, MAX

#6 **Visualise Geographical and Census data using QGIS**

a. Upload *Boundaries - Census Tracts - 2010* in QGIS
b. Upload *ACS* in QGIS
c. Make a joint between Census Tracts boundaries and Census data using Tract Number as Join Field
d. Colour each census tract based on the Predominant Ethnicity
e. Upload CSV file version of the *Complaints* file
f. Map complaints occurrences to identify neighbourhoods in which occurrences concentrate
g. Save Map