

# Crimes in Chicago city from 2001 to 2017

Arthur Le Calvez

NOVA FCT

Lisbon, Portugal

a.calvez@campus.fct.unl.pt



Figure 1: Chicago crime scene

## ABSTRACT

This project, by visualization of data, aims to bring answers to questions that attract the attention of the world when it comes to crimes. The chosen dataset is from the website kaggle, a referenced website when it comes to dataset and data analysis, and it is about crimes in Chicago from 2001 to 2017.

Permission to make digital or hard copies of all or part of this work for personal or internal use is granted without fee, provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Woodstock '18, April, 2021, Lisbon, PT

© 2018 Association for Computing Machinery.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00  
<https://doi.org/10.1145/1122445.1122456>

2021-11-24 10:02. Page 1 of 1-14.

This dataset is composed by multifarious data for each crime (reported by an unique ID) as the location, date, etc, extracted from the Chicago Police Department's CLEAT system.

It follows interesting questions to ask, as how evolved the crime in Chicago since 2001, which area are the most touched by crime, by what type of crime, what is the ratio arrest crime and many more. In this paper, we will introduce our database and do visualization of our data to answer to our questions.

## ACM Reference Format:

Arthur Le Calvez. 2018. Crimes in Chicago city from 2001 to 2017. In *Proceedings of World Health Statistics 2020 (Woodstock '18)*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/1122445.1122456>

## 117 1 INTRODUCTION

118 The subject of this database is the crimes in Chicago from 2001 to  
 119 2017. This dataset is really relevant to model because it is very com-  
 120 plete, with location and date for every crime reported and the type  
 121 of crime committed. Moreover, the scale of time and the number of  
 122 data are big so it is conceivable to expect a precise analysis of crime  
 123 in Chicago, and so we can expect to learn a lots of things about the  
 124 evolution of crime in Chicago through the years and even maybe  
 125 some realistic predictions. Furthermore, crime in Chicago is a real  
 126 issue in United States, and with the recent events in United States  
 127 against Police forces, it's more than never an actual subject.  
 128

129 The dataset is divided into 4 files, for crimes from 2001 to 2004,  
 130 from 2005 to 2007, from 2008 to 2011, and from 20012 to 2017.  
 131

132 Thereby, those of the most important information record in the  
 133 dataset are the Case Number, which is the Chicago Police Depart-  
 134 ment RD Number because it is unique, the Date of when the inci-  
 135 dent occurred, the primary Type which is the type of crime made  
 136 according to the IUCR code, the location description which is the  
 137 description of the location where the incident occurred, the distric  
 138 which indicates the police district where the incident occurred and  
 139 the longitude, latitude and location which are geographical infor-  
 140 mation.  
 141

142 Our models will have for objectives to answer to questions as how  
 143 evolved the crime in Chicago since 2001 by years and months, which  
 144 are the most touched area by crime, which type of crime are the  
 145 most recurrent, what are the most commons locations where crimes  
 146 happened, how is evolving the ration arrestation/crime through  
 147 different variable.

148 First of all, we will be looking at our research questions and dataset.  
 149 Then, we will move onto the state of the art. Thridly, we will discuss  
 150 about our proposal and finally we will finish with evaluation and  
 151 conclusions about our work.  
 152

## 154 2 RESEARCH QUESTIONS

- 156 (1) How evolved the crime in Chicago through the time since  
 157 2012 ?

158 With the features

159 *Date* : Date when the incident occurred

- 160 – *Year* : Year the incident occurred

161 We should be able to calculate the number of crime that  
 162 happened by year in Chicago and so answer to the ques-  
 163 tions. If the visualization shows tendencies for some months,  
 164 it should be interesting to compare it with database about  
 165 weather in Chicago. Maybe we could show correlation  
 166 between weather and crime rate.

- 167 (2) Which type of crime are the most recurrent according to  
 168 the area ?

169 With the features

170 *Primary Type* : The primary description of the IUCR  
 171 code

- 172 – *District* : Indicates the police district where the incident  
 173 occurred.

175 we should be able to visualize which crimes are the most  
 176 recurrent and where they are.

- 177 (3) How is evolving the ratio arrestation/crime made accord-  
 178 ing to the years, the location and the type of crime ?  
 179 With the future,

180 *Location Description* : Description of the location where  
 181 the incident occurred.

- 182 – *Year* : Year the incident occured.  
 183 – *Primary Type* : The primary description of the IUCR  
 184 code.  
 185 – *Arrest* : Indicates whether an arrest was made.  
 186 we should be able to answer to the questions.

- 187 (4) What are the most dangerous hours in Chicago according  
 188 to the area and the type of crime ?  
 189 With the future,

190 *Primary Type* : The primary description of the IUCR  
 191 code.

- 192 – *Date* : Date when the incident occurred. this is some-  
 193 times a best estimate.  
 194 – *District* : Indicates the police district where the incident  
 195 occurred.  
 196 – *Latitude/Longitude*

197 we should be able to answer to the question.

## 198 3 DATA SET

200 The dataset is divided into 4 csv files, for crimes from 2001 to 2004,  
 201 from 2005 to 2007, from 2008 to 2011, and from 20012 to 2017. Each  
 202 rows of this dataset correspond to a crime committed in Chicago's  
 203 area between 2001 and 2017. There is a total of 7 941 292 records.  
 204 Each records has its own temporal and spatial indications :

- 205 • the temporal indication is in date (when the incident occurred)  
 206 and year (year the incident occurred) between 2001 and 2017.
- 207 • the spatial indication is also indicated by differents ways but  
 208 the more important should be the location column because  
 209 it's the one in a format that allows for creation of maps and  
 210 other geographic. We should also notice that in order to  
 211 protect the privacy of crime victims, addresses are shown at  
 212 the block level only and specific locations are not identified.

213 We will describe each field in the annex.

214 The dataset's total size is 1.85 GB.

## 217 4 STATE OF THE ART

### 218 4.1 Related papers and works

- 219 *4.1.1 Exploring Chicago Crimes 2012-2016* : [1] This works  
 220 is exploring the crime in Chicago between 2012 and 2016. It uses  
 221 the same dataset as us and only the dataset that records crimes  
 222 between 2012 and 2017.

223 This works is exploring multiples research questions which join  
 224 ours, as : Did homicides rates really getting worse in Chicago in  
 225 2016 ? Why are the number of arrests so less compared to number  
 226 of crimes ? How the crimes are distributed by months ? Which are  
 227 the locations with most crimes ? How crime numbers vary in top  
 228 crime locations ? Which are the most common crime types ?

229 We can find the papers at this address <https://www.kaggle.com/umeshnarayanappa/exploring-chicago-crimes-2012-2016>.

233   **4.1.2 Homicide in Chicago: Data Storytelling:** [2] This works  
 234   is also exploring the crime in Chicago between 2001 and 2016. It  
 235   also uses the same dataset as us.

236   This works is exploring multiples research questions which join  
 237   ours, as : How crime is distributed accross the city between 2001  
 238   and 2018, and more particularly homicides ? How homicides are  
 239   distributed through the years, months, days and hours ? Which loca-  
 240   tions are the most touched by homicides ? Which is the relationship  
 241   between arrests and non-arrests for homicides ?

242   We can find the paper at this address <https://towardsdatascience.com/homicide-in-chicago-data-storytelling-part-one-e6fb77afc07>  
 243   for the part 1 and <https://towardsdatascience.com/homicide-in-chicago-data-storytelling-part-two-e8748602daca> for the part two.  
 244

245   **4.1.3 Analysis and Visualization of Crime in Chicago :** [3]  
 246   This works is also exploring the crime in Chicago between 2012  
 247   and 2016. It also uses the same dataset as us and only the dataset  
 248   that records crimes between 2012 and 2017, and removed data from  
 249   2017.

250   This works is exploring multiples research questions which join  
 251   ours, as : visualization of density of crime in different areas filter-  
 252   ing by charge (misdemeanor or felony), location, and year range),  
 253   visualization of crime by showing crime counts clustered by area,  
 254   the crime trends over years.

255   We can find the paper at this address <https://nycdatascience.com/blog/student-works/r-shiny/analysis-and-visualization-of-crime-in-chicago/>.

256   **4.1.4 Chicago crime data analysis :** [4] This works is also ex-  
 257   ploring the crime in Chicago between 2012 and 2016. It also uses  
 258   the same dataset as us and only the dataset that records crimes  
 259   between 2012 and 2017, and removed data from 2017.

260   This works is exploring multiples research questions which join  
 261   ours, as : Which are the most common crime types ? How are  
 262   they evolve through the years ? How evolves the number of crime  
 263   through the years, the quarters and the hours ? How evolves the  
 264   arrest and unarrest ? How crime are distributed through the location  
 265   and time ?

266   We can find the work at this address <https://github.com/khatrider/Chicago-Crime>.

## 275   **4.2 Evolution of the crime through the time**

276   It is interesting to know how evolve the crime in Chicago through  
 277   the time. To visualize the evolution of a data through the time, a  
 278   bar chart is an efficient way to do it, and that's what [2] have done.  
 279

280   In those bar-charts, thereafter, we have in lines the number of  
 281   homicide committed and in columns the years, the months, the days  
 282   and the hours respectively. Those bar-charts allows us to see im-  
 283   mediately which years, months, days or hours there is the more  
 284   homicides committed in Chicago.

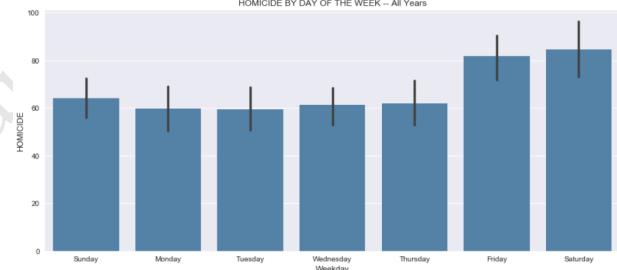
285   So it is very clear that the years with the most homicides are 2016  
 286   and 2017, that the months with the most homicides are summers  
 287   months, that the days with the most homicides are friday and sat-  
 288   urday and that the hours with the most homicides are night hours.  
 289



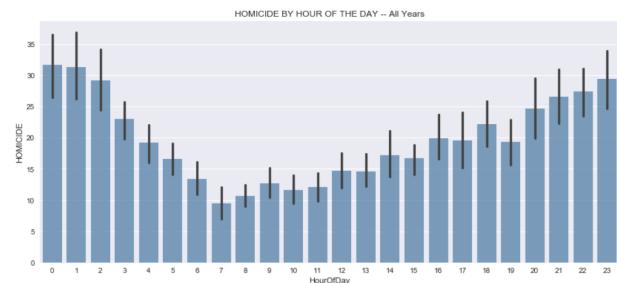
291   **Figure 2: Yearly Homicide Rates**



295   **Figure 3: Yearly Homicide Rates**

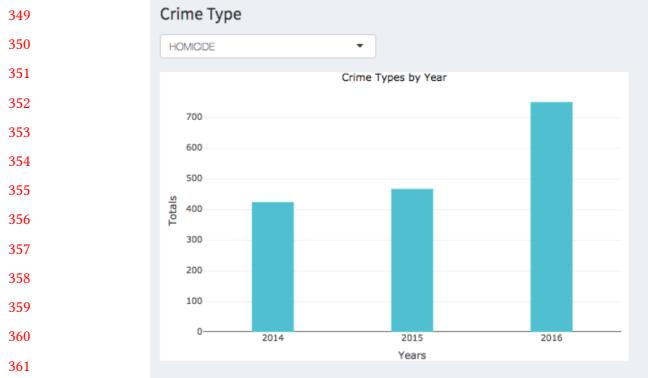


299   **Figure 4: Daily Homicide Rates**



343   **Figure 5: Hourly Homicide Rates**

344   In paper [3], they used also bar-chart. We can then see crime  
 345   types by year. The below shows the spike in homicide counts from  
 346   2015 - 2016 very clearly too.



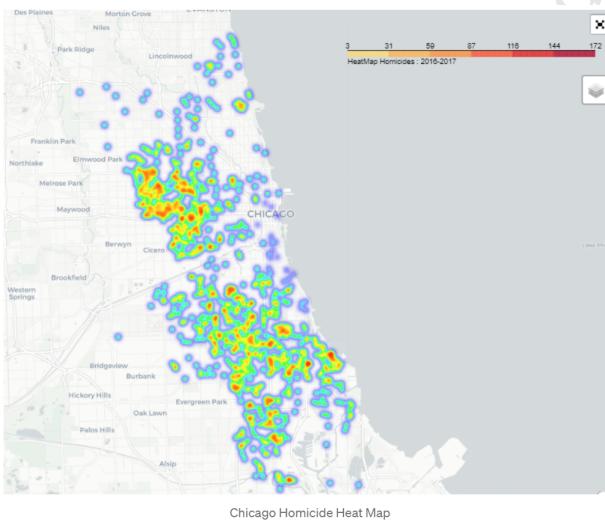
**Figure 6: Crime Types by Year**

In paper [4], they also used bar-charts, but they visualize at the same time the years and the quarters. We can see that most crimes happened in quarter 2 and 3, that is when weather is good in summer months. However, the bar-charts describing crimes by hour is not really readable, we see there is a spike every year but we don't know at which hour of date it corresponds.

### 4.3 Distribution of the crime accross the city

In our dataset, we got a lots of geolocalisation data. So, it is great to visualize our data on a map to see the distribution of the crime accross the city.

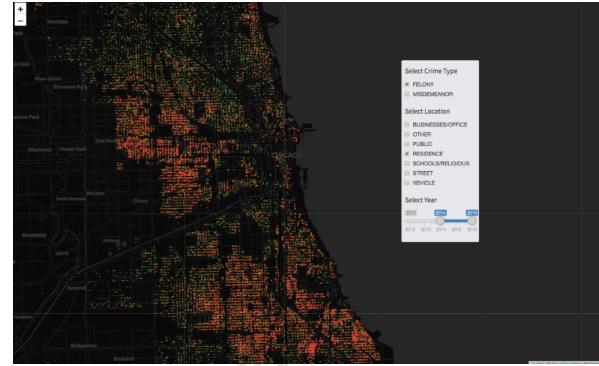
In [2], they choose to visualize the distribution of crime accross the city with an heatmap.



**Figure 7: Chicago Homicide Heat Map**

The heat map is for the 2016-2017 time frame. The individual crime scene locations are mapped on the city's canvas and represented as colors. This heatmap shows us immediatly locations on the map where the frequency of crimes is high or low. The more the color is hot, the more the crime frequency is high. We see that homicides are higher on west side and south of Chicago.

In [3], they also choose to visualize the distribution of crime accross the city with an heatmap.

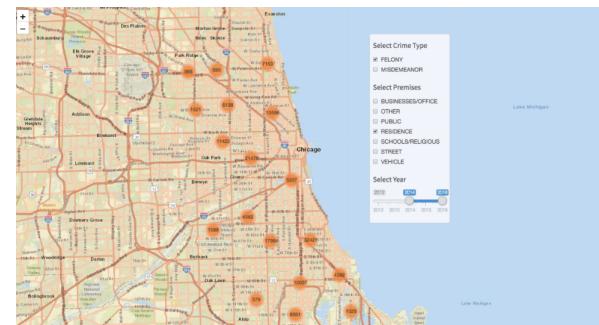


**Figure 8: Chicago crime Heat Map**

This visualization allow the user to see the density in crimes by crime type, location, and year. The density of crime is represented on the map by the concentration of red on area : the higher the concentration of red is, the higher the density in crimes is. For this one, we see for crime type "Felony" and location "Residence", for years 2014 to 2016, that there is a higher concentration of red in the southern half of the city.

This heatmap displays almost the same informations as the previous one, however the choose of red concentration to represented the density instead of a cold/hot color made it way more difficult to read.

In [3], they also choosed to visualize the distribution of crime accross the city with a cluster map.

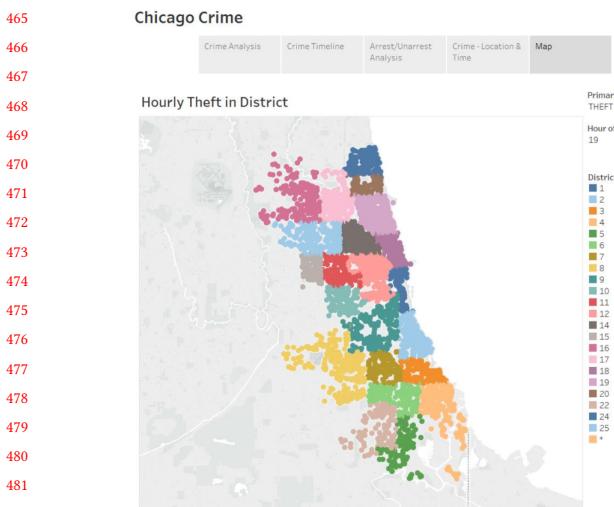


**Figure 9: Chicago crime Cluster map**

The cluster map displays number of crimes in the different area of the city filtered by crime type, location and year. We don't see it with this image, but when you zoom into a particular location you can see details for crime as the date, the arrest or not and the location.

This visualization is very effective because it shows us a lot of of informations but it stays really clear and simple to use and read.

In [4], we have another choice of representation of the distribution of crime in Chicago with a map.



**Figure 10: Chicago map Hourly Theft in District**

On the map, each color represent a district. The more there is point in a district, the more there is a selected type of crime that happened at the hour of date chosen.

For these example, the crime type chosen is "THEFT" and the Hour of Date is "19".

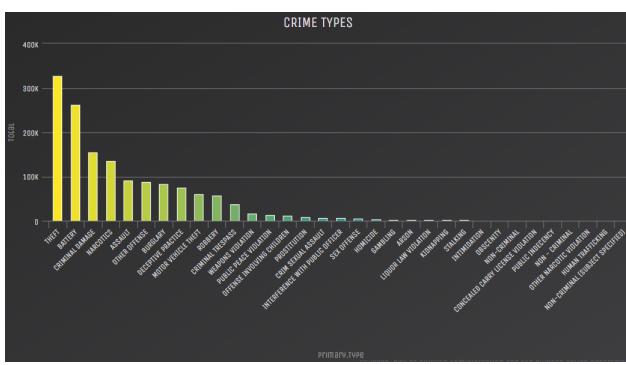
We see that a lots of theft where records every district of Chicago at the Hour of Date 19, but this is more dense in the north district of Chicago.

However, on these type of visualization, if there is too much point we will not be able to see the difference of count between the different district, even if it is big.

#### 4.4 Different type of crime in Chicago

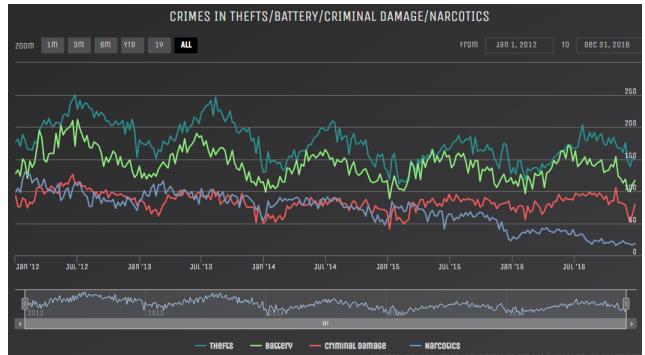
There is a lot of different type of crime that were records in Chicago, so it interesting to visualize for example which type are the most common.

In [1], they choose to represent it with bar-charts and graph.



**Figure 11: Crime type ordered by total number of crime.**

2021-11-24 10:02. Page 5 of 1-14.



**Figure 12: Crimes in thefts/battery/criminal damage/narcotics through the years**

In the bar-charts, the type of crime is in column and total numbers of crime in lines.

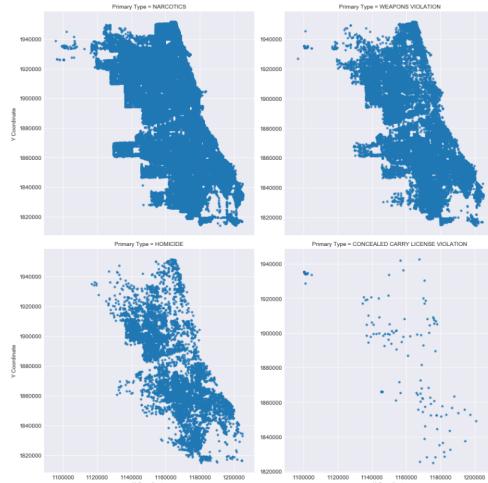
This is an effective way to visualize the number of crimes by type of crime because we immediately see that Theft, Battery, Criminal Damage and Narcotics are the most represented.

However, the choice of a color gradient for each type of crime according to his total of crime doesn't add any others information.

The graph displays the evolution of crimes in Theft, Battery, Criminal Damage and Narcotics through the time. We can choose the years span, and if we want for 1 years, 6 months, 3 months or 1 month. Each colors represent a type of crime between the four represented.

In this example, we see that number of narcotics crimes have reduced and number of thefts, battery and criminal damage remained the same through the years.

In [2], they choose to represent crime type on a map.



**Figure 13: Geographical distribution of a chosen type crime**

On the lines, it's represented the Y coordinate of the crime and on the columns its X coordinate.

It looks like Narcotics and Weapons Violation crimes are common

581 all over the city but Homicide and Concealed Carry Violation crimes  
 582 have more a specific geographic characteristic and are less represented.  
 583

584

585

586

587

588

589

## 590 4.5 Arrest and unarrest analysis

591 When we talk about record crime, it is natural to be interested by  
 592 the correlation between arrest and unarrest crime in those records.  
 593

594

595 In [4], arrest/unarrest trend are analysed through a graph and  
 596 bar-charts.  
 597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615



616 **Figure 14: Arrest/Unarrest Analysis**

617 At the top, with a graph is display the arrest/unarrest trend. On  
 618 lines there is the arrest count and unarrest count, and on column  
 619 the Month of date.  
 620 We see that there is way more unarrest crime than arrest crime.  
 621 However, the choice of not choose the same scale for arrest count  
 622 and unarrest count make it difficult to interpret by a simple look.  
 623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639 At the bottom, there is the arrest/unarrest count by carter of  
 640 date represented with bar-charts.  
 641 On the columns there is the total count of arrest/unarrest crime  
 642 and on the columns, there is each quarter between 2012 and 2017,  
 643 and for each quarter Arrest/Unarrest Count.  
 644 Here, at contrary of the top visualization, we immediately clearly  
 645 see that the non arrest count is higher than the arrest count.

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695



656 **Figure 15: Crimes and arrests heatmap by year and month (2012-2016)**

657 On the column there is the years, on the line the month. Then a  
 658 color scale represent for each month by year (represent by a square)  
 659 the total of crimes and arrest, which is also write into the square.  
 660 The heatmap shows how the number of arrests have decreased by  
 661 more than a half between 2012 and 2016 but the crimes have not  
 662 reduced at the same rate, still the arrests have gone done drastically.  
 663

664 However, the choice of the colour scale is not very pertinent and the  
 665 scale is not the same between crimes and arrests (8000 arrests will  
 666 have the same colour than 28000 crimes), so to see that the number  
 667 of arrests have gone done drastically compared to the number of  
 668 crimes, we need to read the number in each cases, which is not very  
 669 convenient.  
 670

## 697 5 PROPOSAL

### 698 5.1 How evolve the crime in Chicago through the time since 2012 ?

699 Our first dashboard visualize the crime through the time. There are four visualizations for each time resolution.

700 The one on the up-right is a bar-charts that visualize the number of crime by year between 2012 and 2016.

701 The one on the up-left is an heatmap that visualize the evolution of crime through the month for each year between 2012 and 2016 : the  
702 more the square is red, the more there was crime record on this month.

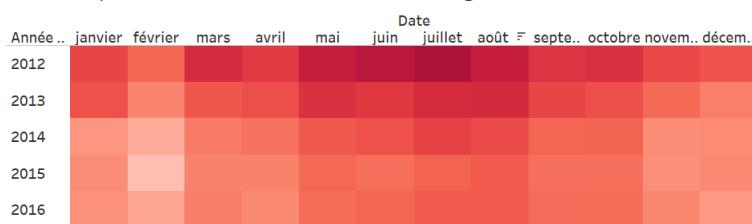
703 The one on the bottom-left is a bar-chart that shows evolution of total crime record by day week.

704 For the last one on the bottom-right, this is graph that represents the total crime record through the hour of the day.

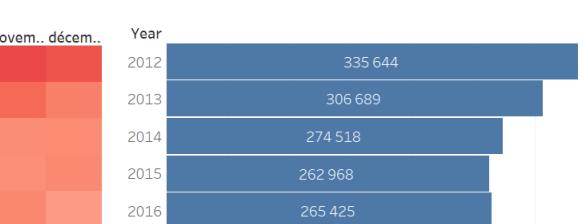
715 Total distinct de Case Number  
716 16 287  31 942

### 717 Visualization of the crime through the time

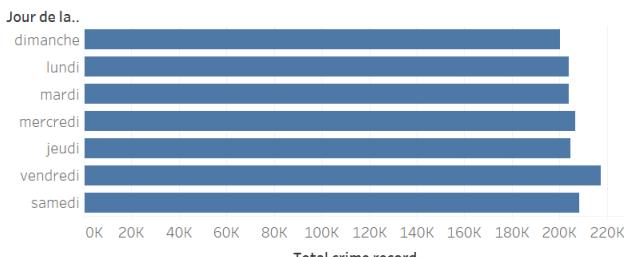
718 Heatmap of the evolution of crime through Year and Month



719 Number of crime per year



720 Crime per day in Tout in 2012, 2013, 2014 et 2 de plus



721 Crime per hour in Tout in Tout in Tout



722 **Figure 16: Dashboard : Visualization of the crime between 2012 and 2016.**

723 For the image above, we see with the bar-chart "Crime per year" that the number of crime decreased for each year since 2012, with a  
724 stagnation between 2015 and 2016. The result is validate with he heatmap and we also see that the most crime happened in summer months.  
725 Now, we can choose by clicking on the heatmap a year, a month or a month in a year : for example on the picture below we choose July 2012,  
726 and we see crime for Friday (vendredi). In consequence, on the visualization on the bottom-left, the bar chart are the crime per day in July  
727 2012, and for the one on the bottom-right, we see the evolution of crime per hour for Friday, in July 2012.

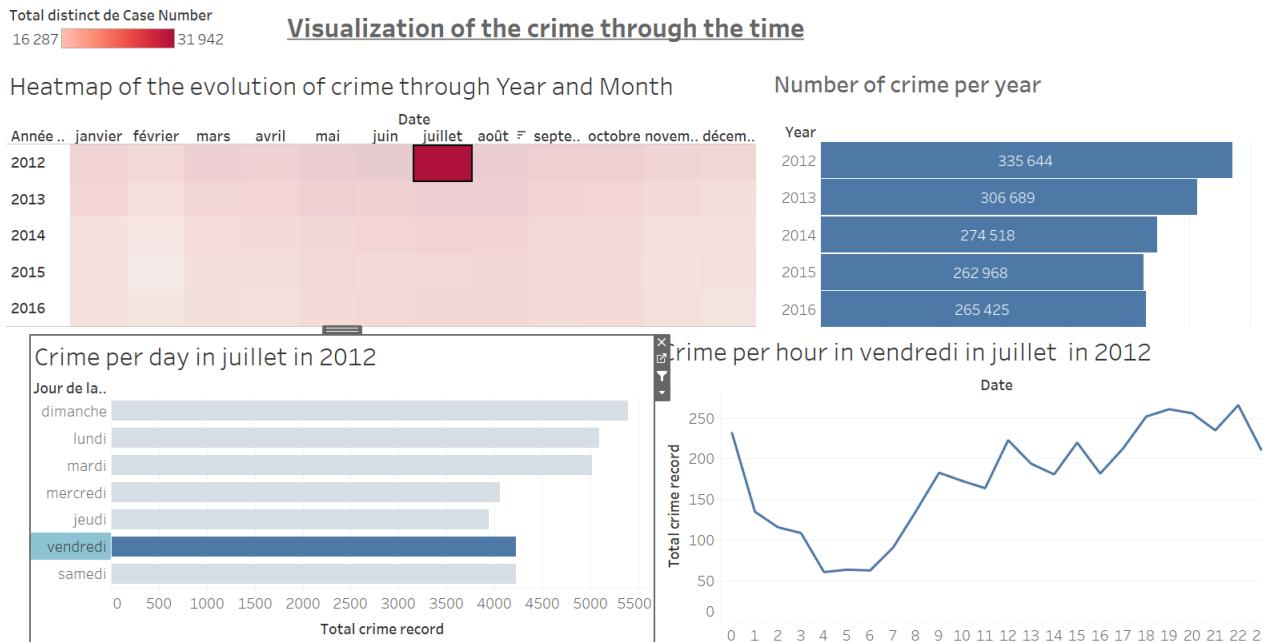
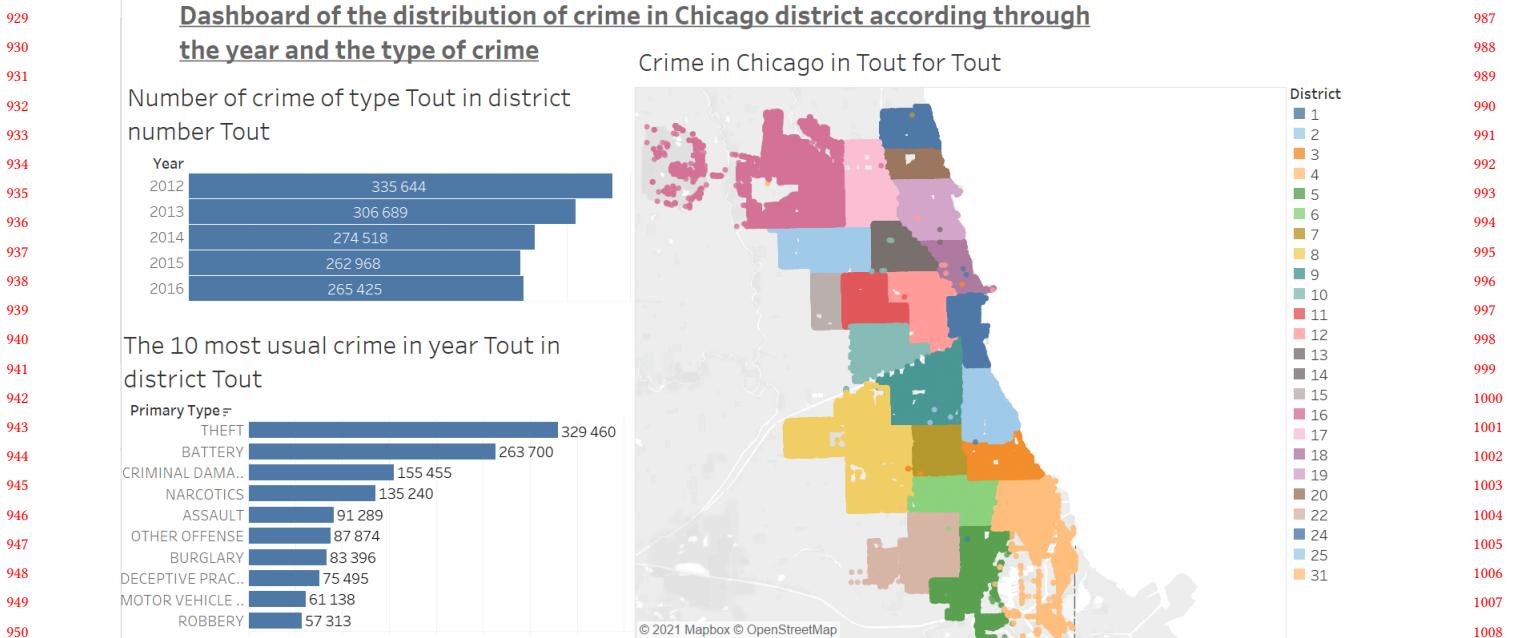


Figure 17: Dashboard : Visualization of the crime on Friday during July 2012.

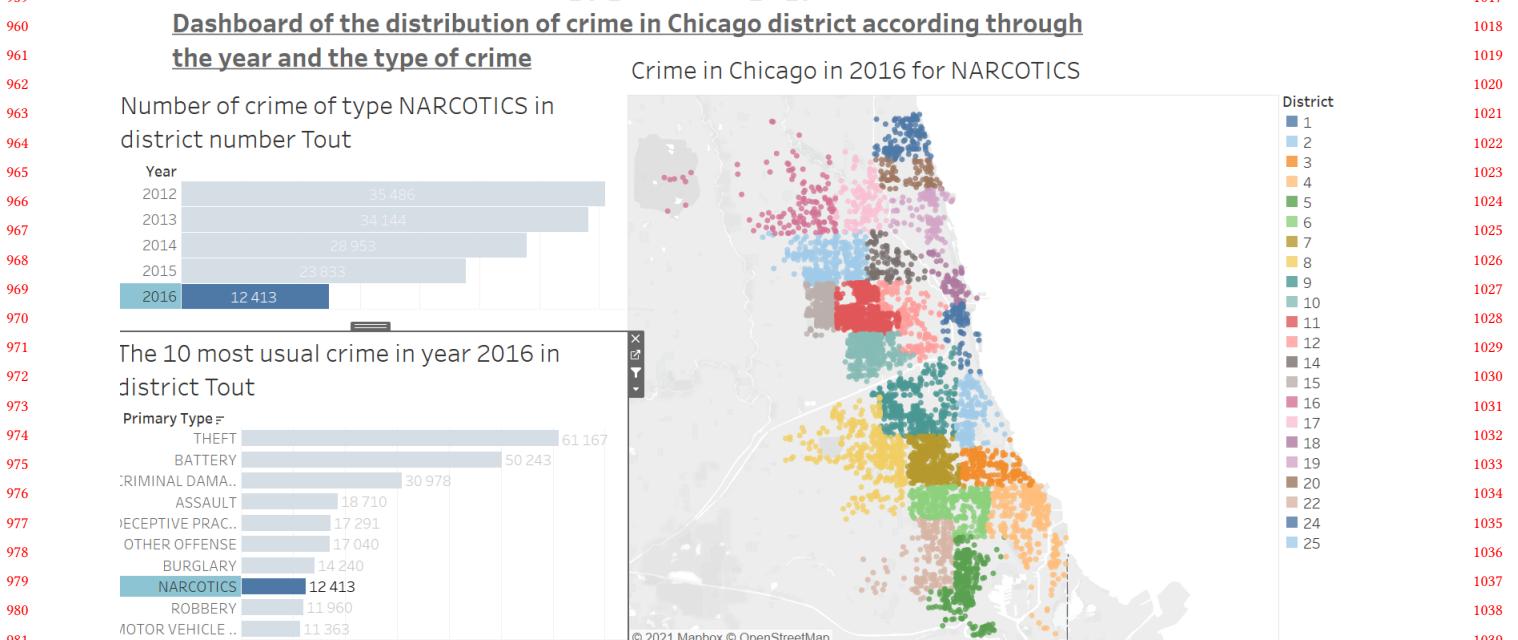
847 It's chosen the day Friday of July and year 2012. We see that that there is more record crime on the day Sunday, Monday and Tuesday and  
 848 that on Friday, most of the crime happened between 9:00am and 00:00pm.  
 849

## 5.2 Which type of crime are the most recurrent according to the area ?

863 On the second dashboard, we visualize the distribution of crime for each district in Chicago according to the year and the type of crime with  
 864 a map. // On the right, there is a map that show the distribution of crime in Chicago according to the year and type of crime chosen. We see  
 865 that in the caption, each district is represented by a colour, and when a crime occur it is represented on the map by a point at the location  
 866 which where it occurs and of the colour of the district in which it occurs. // The more there are points of a colour on a area, the more the  
 867 type of crime selectionned happened in this district for the year selectionned.// On the left, there is two horizontal bar chart, the first one  
 868 represent the number of crime through the years according to the primary type (type of crime) chosen in the district chosen , and the second  
 869 one represent the 10 most usual crime for the year chosen in the district chosen.//  
 870

**Figure 18: Dashboard : Visualization of the distribution of crime in Chicago district according to the year and the type of crime.**

To interact with the dashboard, we can choose the district, the year and the type of crime we want to see. We can choose the year by clicking on the one we want to see on the first bar chart. It will change the bar chart of the 10 most usual crime and the map of crime in Chicago for the year selected. We can also select the type of crime we want to see by clicking on the one we want on the second bar chart, it will change the bar chart of the number of crime and the map of crime in Chicago for the type of crime selected. And we can also chose the district we want to see by clicking on a point in the colour of the district we want. It will change the two bar chart and visualize the data only for the district selected.

**Figure 19: Dashboard : Visualization of the distribution of crime in Chicago district according to the year and the type of crime.**

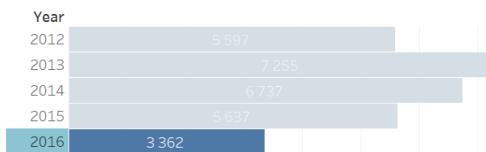
For the image above, we see that we select the year 2016, the type of crime "NARCOTICS" . In consequence, we see on the first bar chart the number of crime of type "NARCOTICS" in all district between 2012 and 2016, and on the seconnd bar chart the 10 most usual crime in 2021-11-24 10:02. Page 9 of 1-14.

1045 year 2016 for all district. We see on the map the distribution of crime of "NARCOTICS" type in Chicago in 2016, we see that the majority of  
 1046 "NARCOTICS" crime happened on the north west and south of Chicago in 2016.  
 1047  
 1048  
 1049  
 1050  
 1051

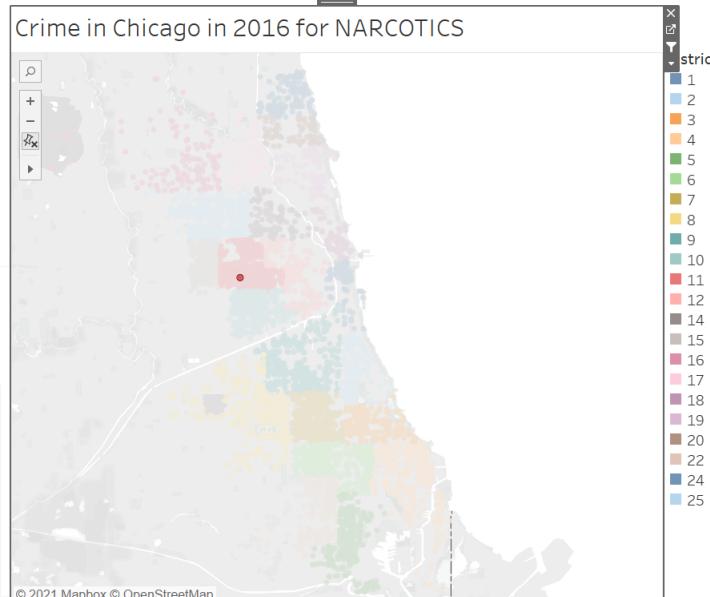
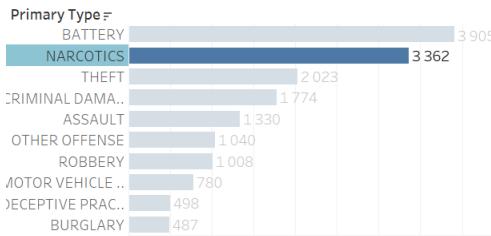
1103  
 1104  
 1105  
 1106  
 1107  
 1108  
 1109

### Dashboard of the distribution of crime in Chicago district according through the year and the type of crime

1055 Number of crime of type NARCOTICS in  
 1056 district number 11



1063 The 10 most usual crime in year 2016 in  
 1064 district 11



1075 **Figure 20: Dashboard : Visualization of the distribution of crime in Chicago district according to the year and the type of crime.**

1081 For the image above, we did the same but we select also the district 11 on the map, so the two bar chart are for the district 11.

1133  
 1134  
 1135  
 1136  
 1137  
 1138  
 1139  
 1140  
 1141  
 1142  
 1143  
 1144  
 1145  
 1146  
 1147  
 1148

### 5.3 How is evolving the ratio arrest/crime made according to the years, the location and the type of crime ?

1091 Our third dashboard visualize the evolution of crime and arrest according to the year, the location and the type of crime.

1149

1092 There are three visualization : at the top a graph, and on the bottom two bar-charts.

1150

1093 On the top, the graph represents the evolution of number of arrest and crime recorded for the year selected. There is two, in green the arrest  
 1094 and in blue the no arrest.

1151  
 1152

1095 On the bottom, there is two bar-chart. The one on the left represents the 10 first location where crime happened, ordering by descendant  
 1096 order for total crime record for each location. The one on the right represents the 10 first type of crime , ordering by descendant order for  
 1097 total crime record for each type of crime.

1153  
 1154  
 1155

1098 For both, bar are divided into two parts : in green the arrest and in blue the no arrest.

1156

1100 We can choose the period of time we want to visualize on the two dashboard by selecting on the graph the period we want.

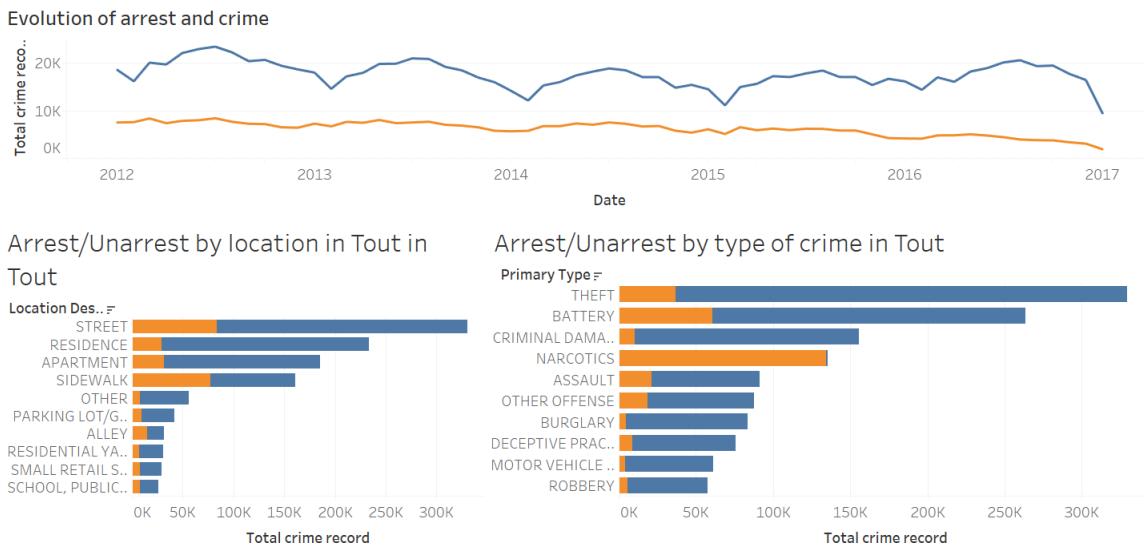
1158

1101

1159

1102

### 1161 Visualization of the crime and arrest



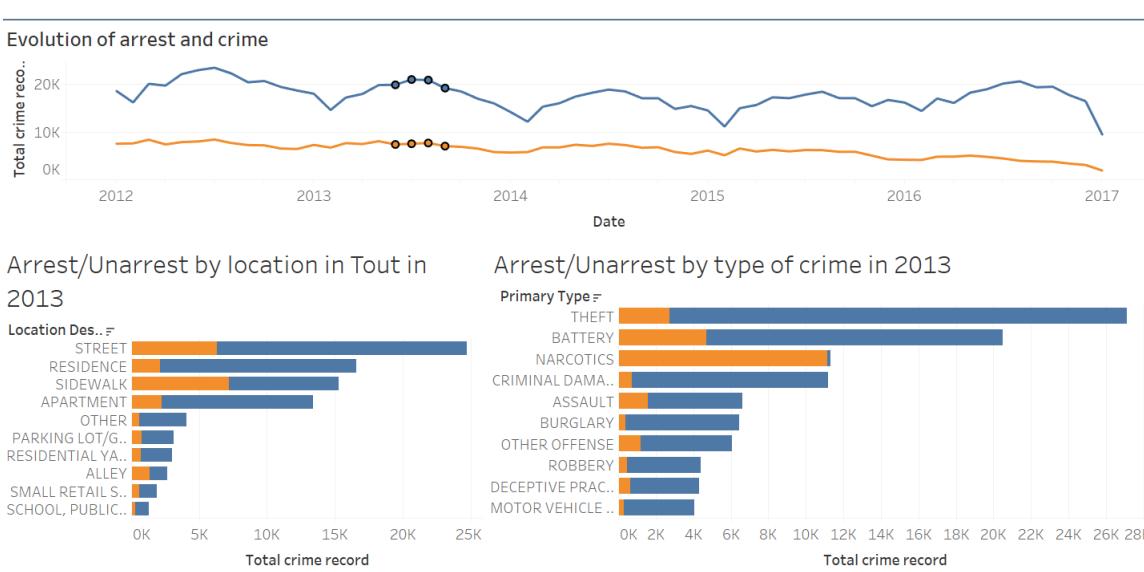
1167 **Figure 21: Visualization of the crime and arrest between 2012 and 2016.**

1168 For the image below, the chosen period is June 2013 to September 2013. We see that in this period, there is way more (more than the double) crime than arrest crime : the up line in blue represents the total crime records and the below line in green represents the total number of arrest crime during 2014 by months.

1169 We see on the two bar-chart that the ratio between arrest/unarrest crime corresponds to the the graph above. The four location where most of the crime happened in 2014 are in order Street, Residence, Sidewalk and Apartment. There is more arrests crime compared to the number of crime for sidewalk than the three others.

1170 The four crime type which happened the most are in order Theft, Battery, Narcotics and Criminal Damage. We see that for Narcotics, almost all crime where arrested.

### 1171 Visualization of the crime and arrest



1177 **Figure 22: Visualization of the crime and arrest between June 2013 and September 2013.**

#### 5.4 What are the most dangerous hours in Chicago according to the area and the type of crime ?

This last visualization represents the number of crimes in Chicago in function of the type of crime, the location and the daytime. We can find three main visualizations here.

First, we have the map of Chicago, we can observe here the location of all the crimes selected. These locations teach us two main information, the districts where the crime has been made (Color) for the hour selected and their numbers.

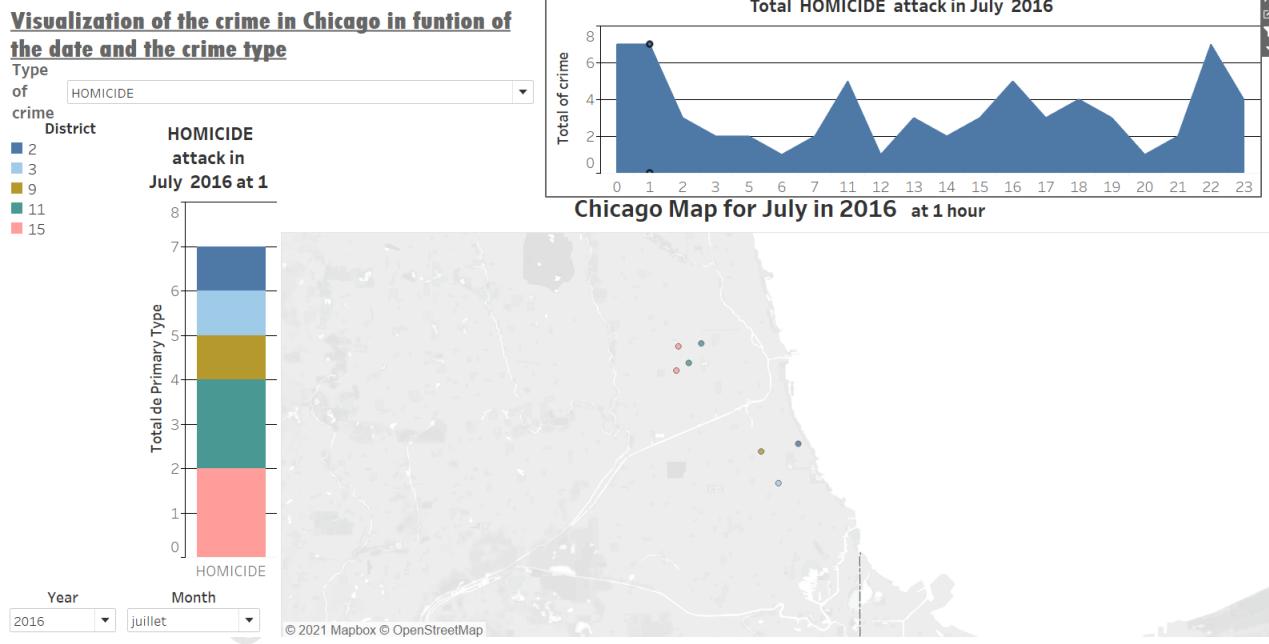
We can observe on the right corner a line visualization, it represents the total of crime (Previous selected) for each hour of the day for precis Year and Month. It allows the user to understand the dangerousness of the districts according to the different hours of the Month.

The last visualization is a histogram bar which gives us a fast computing information of the crimes made during the hour selected inside each districts.

This visualization contains five captions. The most important is the primary type which selects the type of crime wanted, it is linked with our three visualization.

The second caption is the District information on the left size, it gives the name of the districts according to the color. This caption is not dynamically, the user can't interact with it. It just gives the District information for the histogram bar and the map (Color).

The last captions can be gathered in one explication, indeed they are all responsible of the date wanted. We have the years and the month which can be selected with lists. Because of the large amount of information present in the database, we decided here to visualize our data with the hour. Indeed, we can choose the hour we want to see by selectionning on the histogram bar the one we want. It allows a clearer geographical visualization. The user can easily find the hours and the districts the most dangerous for a given month and year.



**Figure 23: Visualization of the crime in Chicago for homicide in July 2016, at 1pm.**

1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392

## 1393 6 EVALUATION AND CONCLUSIONS

1394 First, the visualization main goal is to answer at the best to the research questions that highlight the most important idea of the database. For  
 1395 that, we choose to visualize mainly the data link to the location (map), the time, the number of crime and arrest, the type of crime and the  
 1396 location of the map.

1397

1398 There is four dashboards to answer to the different research question, one for each.

1399

1400 On the first dashboard, we choose to use an heatmap to see the visualization through the months and year so we can see everything at the  
 1401 same time immediatly. We see that for all years, there is an increasing of number of crime recorded at the summer, and by playing a little bit  
 1402 with the dashboard we see with bar chart on the bottom left and the graph on the bottom right that the number of crime per day week  
 1403 and by hour of the day remains stable through the time. It would have been interesting to see if the number of crime is correlated with the  
 1404 temperature because there is some rumors about that in Chicago : the hotter the day is, the more there is crime.

1405

1406 On the second dashboard, we choose to use a map accompanied by bar chart. It is a little difficult to see correctly the density of the crime  
 1407 in the city when we choose options where there is too much crime, because the map is overlapped when there is too much point. However,  
 1408 we see that for the top 10 crime, there is a tendency to number of crime decrease between 2012 and 2015 with for some a little increase in  
 1409 2016. We also see that the top crime are happening almost everywhere in Chicago, in all district. It could have been a good idea to divide the  
 1410 city of Chicago otherwise than by District, for example we could have divide by communities area and compare both.

1411

1412 On the third dashboard, we choose to visualize the comparison of arrest and unarrest crime with a graph and two bar-chart. It could have  
 1413 been interesting to add a map to see where in Chicago, people who did crime are the most susceptible to be arrested. However, we see that  
 1414 the ration arrest/unarrest remains stable through the time, and for the location and type of crime : for almost all cases there is a lot of chance  
 1415 that the author of the crime will not be arrested, except of the crime of type Narcotics, where almost everyone is getting arrested.

1416 On the fourth dashboard, we choose to visualize with a map and an histogram the most dangerous hours in Chicago according to the area  
 1417 and the type of crime. The histogram shoes the most dangerous hours and the map the area where the crime happened at this hour. Our  
 1418 selection which display the crime on the map is very precise with the year, the month, the type of homicide and the hour so there is not so  
 1419 much data to show, and then we can clearly see where it happens.

1420 To conclude, it could have been a good idea to use others type of map than DotMap, as density map to see the distribution of crime in  
 1421 Chicago, and also to analyse the crime not only between 2012 and 2016 but between 2001 and 2016, to have a better idea of the evolution of  
 1422 crime in the time, but unfortunately it was too long to compute with our computer.

1423

1424

1425

1426

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

1437

1438

1439

1440

1441

1442

1443

1444

1445

1446

1447

1448

1449

1450 2021-11-24 10:02. Page 13 of 1-14.

1451

1452

1453

1454

1455

1456

1457

1458

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477

1478

1479

1480

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

1494

1495

1496

1497

1498

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

## 1509 REFERENCES

- 1510 [1] Umesh, "Exploring chicago crimes 2012-2016." <https://www.kaggle.com/umeshnarayananappa/exploring-chicago-crimes-2012-2016>, 2017. 1567
- 1511 [2] R. Wahome, "Homicide in chicago: Data storytelling." <https://towardsdatascience.com/homicide-in-chicago-data-storytelling-part-one-e6fb77afc07>, 2018. 1568
- 1512 [3] K. Sohn, "Analysis and visualization of crime in chicago." <https://nycdatascience.com/blog/student-works/r-shiny/analysis-and-visualization-of-crime-in-chicago/>, 2018. 1569
- 1513 [4] R. J. S. B. Deepthi Khatri, Preethi Kannan, "Chicago crime data analysis." <https://github.com/khatrdeepti/Chicago-Crime>, 2019. 1570

## 1514 7 ANNEX

1515 Column	1516 Explanation	1517
1518 ID	1519 Unique identifier for the record	1567
1520 Case Number	1521 The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.	1568
1522 Date	1523 Date when the incident occurred. this is sometimes a best estimate.	1569
1524 Block	1525 The partially redacted address where the incident occurred, placing it on the same block as the actual address.	1570
1526 IUCR	1527 The Illinois Unifrom Crime Reporting code. This is directly linked to the Primary Type and Description.	1571
1528 Primary Type	1529 The primary description of the IUCR code	1572
1530 Description	1531 The secondary description of the IUCR code, a subcategory of the primary description	1573
1532 Location Description	1533 Description of the location where the incident occurred.	1574
1534 Arrest	1535 Indicates whether an arrest was made.	1575
1536 Domestic	1537 Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.	1576
1538 Beat	1539 Indicates the beat where the incident occurred A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts	1577
1540 District	1541 Indicates the police district where the incident occurred	1578
1542 Ward	1543 The ward (City Council district) where the incident occurred	1579
1544 Community Area	1545 Indicates the community area where the incident occurred. Chicago has 77 community areas.	1580
1546 FBI Code	1547 Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).	1581
1548 X Coordinate	1549 The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	1582
1550 Y Coordinate	1551 The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	1583
1552 Year	1553 Year the incident occurred.	1584
1554 Updated On	1555 Date and time the record was last updated.	1585
1556 Latitude	1557 The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	1586
1558 Longitude	1559 The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	1587
1560 Location	1561 The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.	1588
1562		1589
1563		1590
1564		1591
1565		1592
1566		1593
		1594
		1595
		1596
		1597
		1598
		1599
		1600
		1601
		1602
		1603
		1604
		1605
		1606
		1607
		1608
		1609
		1610
		1611
		1612
		1613
		1614
		1615
		1616
		1617
		1618
		1619
		1620
		1621
		1622
		1623