

# Practical Assignment: Visualizing Data with Tableau

---

Davide Pedranz

d.pedranz@student.rug.nl

Student Number: S3543757

November 3, 2017

## ABSTRACT

The scope of this assignment is to answer a number of non trivial question using visualizations built with Tableau. The dataset contains incidents reported to the police of Seattle in the period 2009 to 2017 and is available online at <http://data.seattle.org> under the name of **Seattle Police Department 911 Incident Response**. We use a dump of the dataset of September 17.

## CONTENTS

1 Exploring the incidents' geographical distribution	2
2 Exploring the incidents' geographical and temporal distribution	7
3 Exploring the resolution speed	8
4 Exploring the incidents' classification	12
5 Exploring the incidents' temporal distribution	17

## TABLEAU VISUALIZATION

All visualization used to answer the questions in this report can be found in Tableau Public at the following address: [https://public.tableau.com/profile/davide.pedranz#/vizhome/seattle\\_police\\_department\\_911\\_incident\\_response/GeographicalDistribution](https://public.tableau.com/profile/davide.pedranz#/vizhome/seattle_police_department_911_incident_response/GeographicalDistribution).

# 1 EXPLORING THE INCIDENTS' GEOGRAPHICAL DISTRIBUTION

## QUESTION 1.1

*Are there high variations of incidents densities over Seattle? Which are the low-density zones? Which are the high density zones?*

The visualization uses a map of the city of Seattle (a scatter plot that uses the maps from OpenStreetMap for the background): each point in the dataset is visualized as a circle on the map. Columns and rows are used to encode respectively longitude and latitude. The size of the circles is set to the minimum, the opacity to the value 8%. Size and opacity do not encode any particular information. However, the combination of small size, opacity and zoom level allows to visualize an approximation of the density of incidents. The zoom level is small in order to get an overview of the entire city that fits the screen. Size and opacity are adjusted accordingly. We added some layer to the city map in background to visualize the main streets and the coast. To add or remove layers in a map in Tableau, we used the Map > Map Layers option in the top menu bar: Tableau opens a left panel with all available layers and marks the selected one. In our case, we added the Coastline and the Streets and Highways layers.

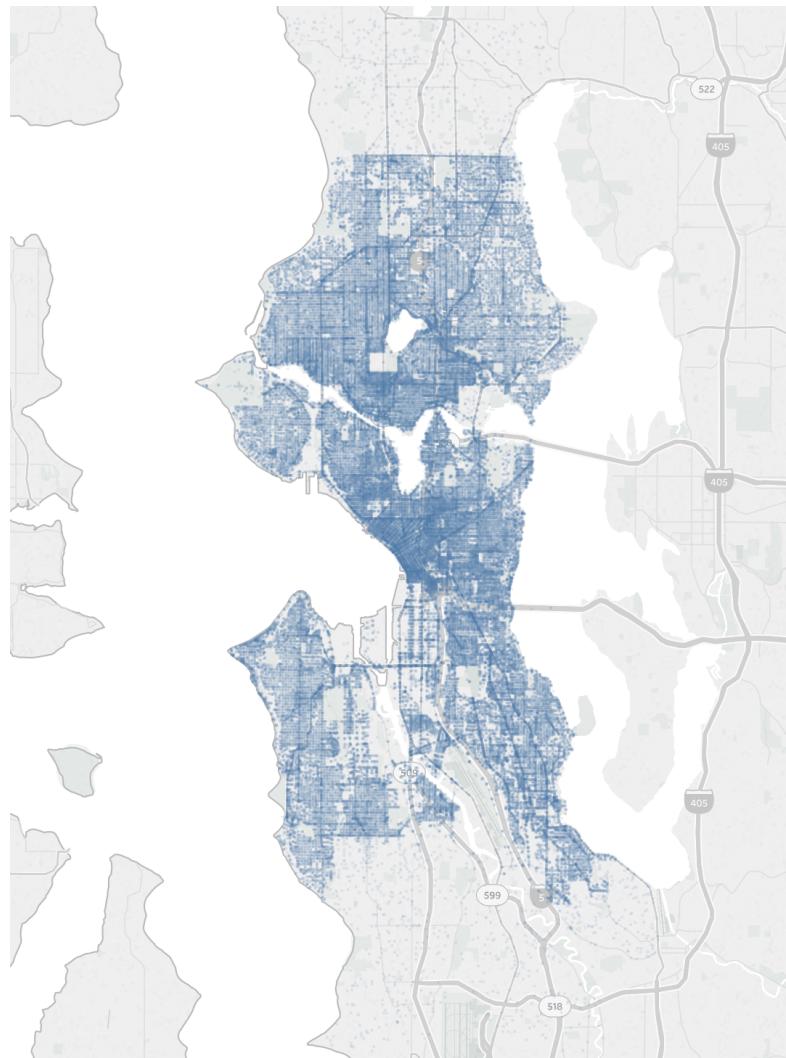


Figure 1.1: Geographical distribution of the incidents in Seattle. The screenshot is taken from the sheet called *Geographical Distribution of Incidents in Seattle* in Tableau.

Figure 1.1 shows that there are indeed variations in the density of incidents in Seattle:

- The regions in the very north and south of the city have a very small number of incidents. These areas are already outside the borders of Seattle. We suspect they are not usually covered by the police of Seattle, so the dataset may not contain all the incidents that occurred in those areas.
- Incidents seem to be more concentrated in the central area of the city, around the Elliott Bay, and immediately to its north.
- Outside the city center, the main roads have a higher number of incidents than the surrounding areas (in particular in the south part of the city).
- The density of incidents in the rest of the city is lower and seems to be approximately uniform.

The visualization answers the question, but it is not optimal. A better solution would be to use a density plot which is able to visualize better the differences of incidents' densities in the city. Unfortunately, Tableau does implement density plots: the proposed solution tries to approximate a density plot by finding the right level of opacity, size and zoom using Tableau.

#### QUESTION 1.2

*Are there high variations of incidents densities over Seattle for specific types of incidents? Which are these types and which are the variations you found?*

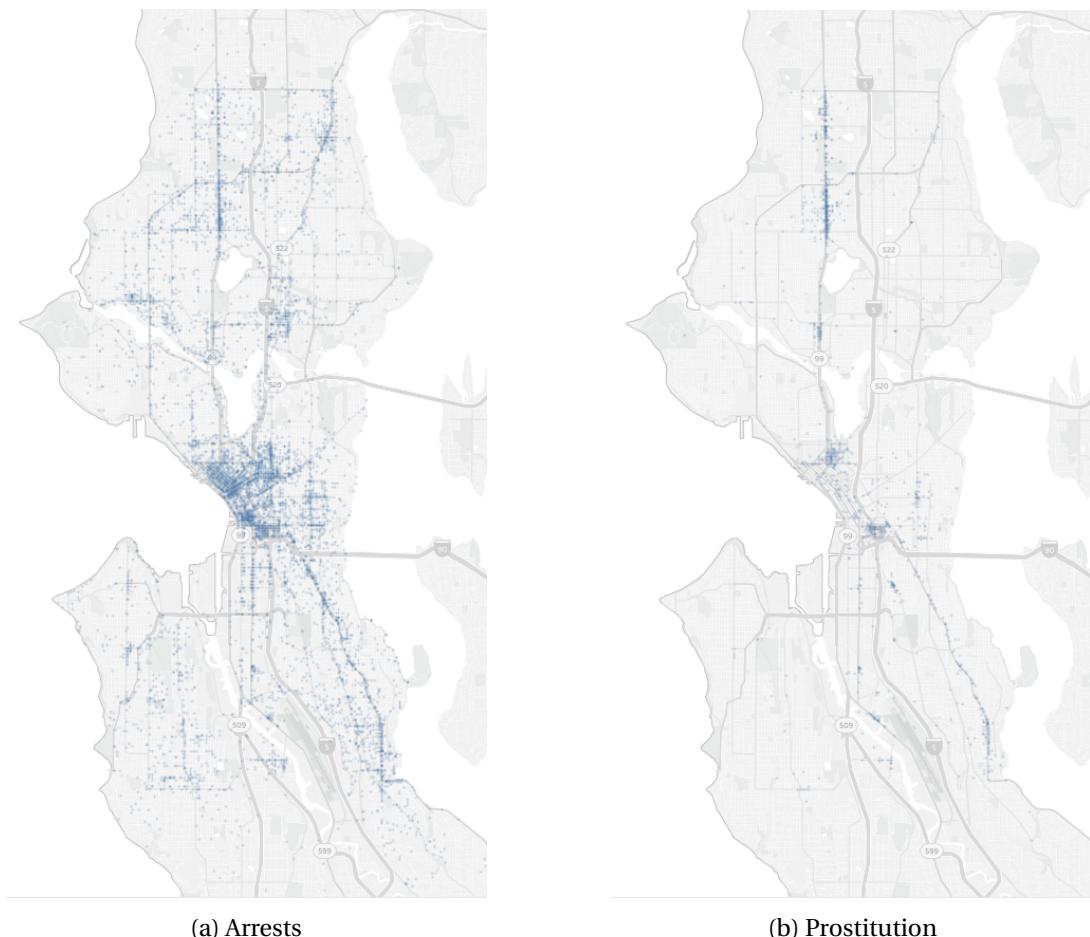


Figure 1.2: Geographical distribution of the incidents of type *arrest* (on the left) and *prostitution* (on the right) in Seattle. The screenshots are taken from the sheet called *Sheet 22* in Tableau.

The visualization uses an approach similar to the previous section, but exploits pagination to display the different types of incidents individually. This is done by using the “Pages” option in Tableau: it is possible to interactively change the type visualized by using the dedicated menu. The visualization uses a filter to exclude the observations without a type. Since there are less observations in each page, the opacity value is set to a higher value (30%).

Figure 1.2 shows 2 snapshots for the types *arrest* and *prostitution*. From the figure it is already possible to notice that different types of incidents have different distributions in the city. Table 1.1 contains the observations for every type of incident.

Incident Type	Description
Accident Investigation	Higher concentration in the city center. Outside the city center, main roads have a higher concentration than the surrounding areas.
Animal Complaints	Slightly higher concentration in the city center.
Arrest	Higher concentration in the city center and along main roads in the north and south of the city.
Assaults	Like <i>Arrest</i> .
Auto Thefts	No significant variations.
Behavioral Health	Higher concentration in the city center.
Bike	Higher concentration in the city center. The north has a higher concentration than the south.
Burglary	No significant variations.
Car Prowl	Slightly more concentrated in the city center.
Disturbances	Like <i>Car Prowl</i> .
Drive By (No Injury)	Higher concentration in south east, almost absent in the other sectors.
Failure to Register (Sex Offender)	Slightly more concentrated in the city center and south east.
False Alacad	No significant variations.
False Alarms	No significant variations.
Fraud Calls	Like <i>Car Prowl</i> .
Harbor Calls	Concentrated along the coast.
Hazards	Like <i>Arrest</i> .
Homicide	Slightly more concentrated in the city center.
Lewd Conduct	Higher concentration in the city center.
Liquor Violations	Like <i>Arrest</i> .
Miscellaneous Misdemeanors	Like <i>Arrest</i> .

Motor Vehicle Collision Investigation	More concentrated in the center and along the streets.
Narcotics Complaints	Like Arrest.
Nuisance, Mischief	Like Arrest.
Other Property	Like Arrest.
Person Down / Injury	Higher concentration in the city center.
Persons - Lost, Found, Missing	Slightly more concentrated in the city center.
Property - Missing, Found	Slightly more concentrated in the city center.
Property Damage	Like Arrest.
Prostitution	Concentrated along the main streets in the north and in the south. Present also in the city center.
Prowler	No significant variations.
Public Gatherings	Higher concentration in the city center.
Reckless Burning	No significant variations.
Robbery	Like Arrest.
Shoplifting	Like Arrest.
Suspicious Circumstances	No significant variations.
Threats, Harassment	Like Arrest.
Traffic Related Calls	No significant variations.
Trespass	Like Arrest.
Vice Calls	No significant variations.
Weapon Calls	Like Arrest.

Table 1.1: Distribution of incidents in Seattle by type.

The proposed visualization allows to analyze the distribution of incidents' types one by one. However, it makes really difficult to compare different types, since the user needs to change page each time.

To overcome this limitation, we have to display different types together. The first solution is to visualize all incidents together on the same map and use a categorical color map to distinguish them (see sheet *Incidents Distribution (all types together)* in Tableau). The user can interactively select subsets of types to display and filter out the others. This solution presents 2 main problems:

- A categorical color map should use color which appear very different for each other to be useful, in the ideal case a set of color for which each pair has the same distance in RGB from each other. Unfortunately, it is impossible to find such a set for more than 4 colors. We have about 40 different types, so such an encoding is not effective.
- There are too many points, so even if we could find a good set of colors for the encoding, the

point would overlap with each other. We can not use transparency to visualize also the underlying points since, since colors would mix together and create new ones.

One could use the following approaches to solve the problem:

- Select a subset of types to show. Unfortunately, there is not an easy choice. Types with a higher number of incidents do not in general have an interesting distribution (i.e. their geographical distribution is similar to the one for all incidents). On the other hand, less frequent types have often a very interesting behaviour (see *Prostitution*).
- Cluster similar types together. Unfortunately, it is not easy to build a types' hierarchy.

An alternative solution is to apply the small multiple design idea, i.e. create a visualization for a single type and replicate it multiple times. This idea is implemented in sheet *Incidents Distribution (small multiple design)* in Tableau: each type is displayed on its own map, and maps for different types are shown in columns adjacent to each other (we use columns because the shape of the city on the chart is stretched out along the y-axis). The type of incident is encoded both in space (different charts) and in color: this overloading makes easier to distinguish different types. Types are sorted in alphabetical order by default, but the user can interactively re-order them in order to compare 2 or 3 types at a time. Figure 1.3 shows an example of such a comparison: types *Prostitution*, *Drive by* and *Failure to register (sex offender)* seem to be somehow correlated in the south-east part of the city but not in the north. The size of each point is chosen to be small, but the opacity is set larger than for Question 1.1 since there are fewer point in each map.



Figure 1.3: Geographical distribution of incidents of types *Prostitution*, *Drive by* and *Failure to register (sex offender)* in Seattle. The screenshot is taken from sheet *Incidents Distribution (small multiple design)* in Tableau.

## 2 EXPLORING THE INCIDENTS' GEOGRAPHICAL AND TEMPORAL DISTRIBUTION

### QUESTION 2.1

*Do you see a different spatial distribution of incidents over the different years? If so, which are the differences you found?*

Before trying to answer this question, it is interesting to take a closer look to the data. The data span a temporal period of 9 years, from 2009 to 2017. The dataset contains 1.441.208 records, of which 1.027.546 (about 70%) have no information about the time of the incident. Also, there are only 7 incidents for year 2009 and 84 for year 2010 (we suspect that many incidents without the time information have happened in these years, but we have no evidence of that). Our analysis concentrate thus on years 2011 to 2017.

Similar to Question 1.2, we use the small multiple design technique. The visualization is composed of multiple maps: each map shows the incidents in Seattle for a single year. The maps are sorted in chronological order, i.e. 2011 is the first on the left, 2017 the last on the right. The user is free to change the order to make accurate comparison between pairs of years, if needed. Since the visualization only shows data for 7 years, it fits in a single screen.

We use a small size for each point on the map and a medium level of transparency to be able to distinguish areas with different densities. We use color to overload the encoding of the year: this allows to combine the visualization with the histogram of incidents' frequency by year, as shown in Figure 2.1.

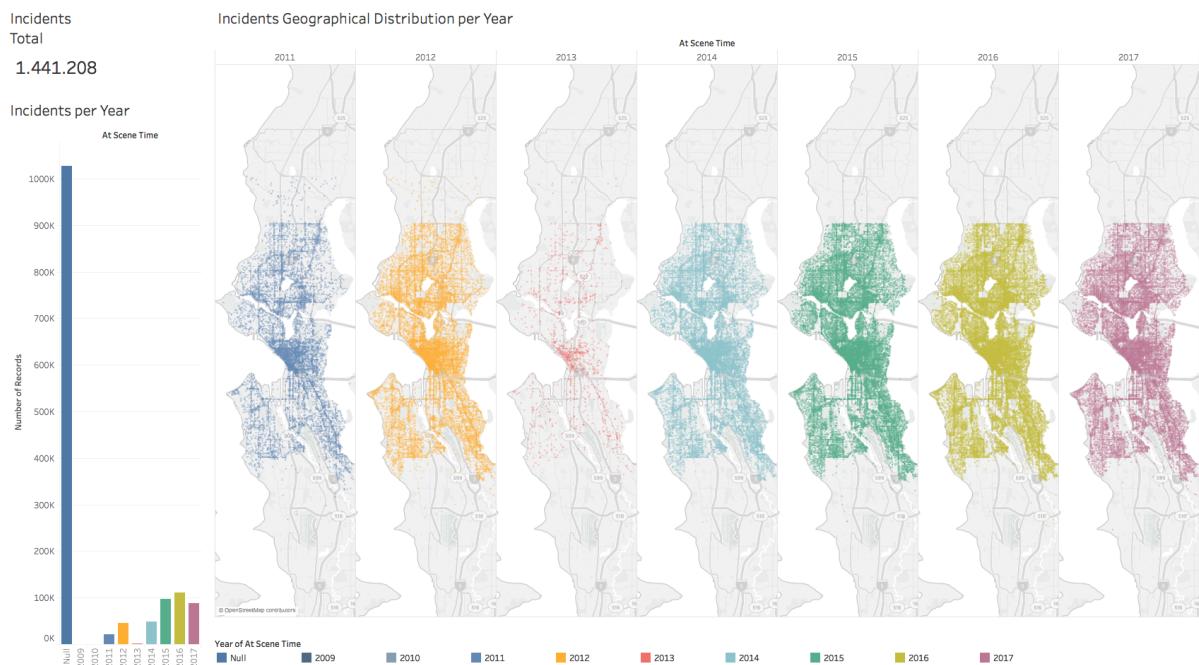


Figure 2.1: Incidents spatial distribution over different years in Seattle. The dashboard is called *Incidents geographical and temporal distribution* in Tableau.

From the visualization, we can not notice any significant difference in the spatial distribution of incidents over the last 5 years. Year 2013 has significantly less incidents, but this is probably due to the missing temporal information on 70% of the entries. Anyway, their distribution is similar to the general one: incidents are more concentrated in the city center (see Question 1.1).

## QUESTION 2.2

*Are there zones with a consistent low incidents density over all years? Are there zones with a consistent high density over all years?*

We can use the visualization built for the previous question to answer. Over all years, the city center has a higher concentration of incidents. Outside the city center, incidents seems to be slightly more concentrated along the main streets. The distribution in each year is similar to the distribution over all years and is described in greater details in Section 1.

## 3 EXPLORING THE RESOLUTION SPEED

### QUESTION 3.1

*What is the average resolution speed to an incident? Resolution speed is defined as the difference between the Event Clearance Date value (moment when the police closed the file of an incident) and the At Scene Time value (moment when the police arrived at the scene of the incident).*

The dataset does not define the average resolution time explicitly. Thus, we need to add a new columns to the dataset with this information. Tableau is able to compute new dimensions using the existing ones. Since we compute a difference between 2 timestamp, we use the function datediff to get a proper result.

The average resolution time of the entire dataset is about 2 hours.

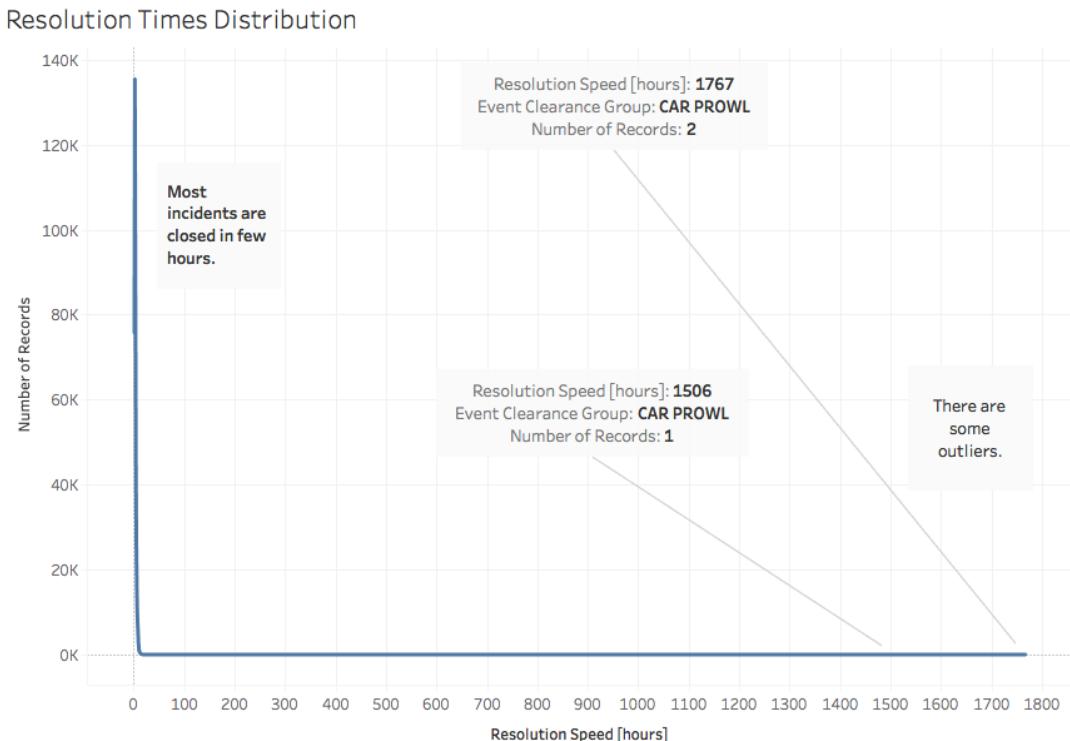


Figure 3.1: Distribution of the resolution times. The annotations on the plot show some outliers with very high resolution times. Most of the incidents are closed in few hours. This is a screenshot of sheet *Resolution Times Distribution* in Tableau.

Figure 3.1 shows the distribution of resolution times. We can notice that:

- Most incidents are closed in few hours.
- There are few outliers that have a resolution time of over 1500 hours (approximately 2 months). All outliers share the same type: *Car Prowl*.

The line chart gives already some interesting insights about incidents resolution times. Due to the described outliers, it is difficult to read the left part of the chart, where most of the distribution mass is.

To solve the problem, we add an histogram of the resolution times. Resolution time is a continuous measure, but we are not interested in the precise value for each incident, rather in the overall distribution. To visualize the distribution, we discretize the resolution time in buckets of 1 hour. Since most of the density mass in the first 15 buckets, we cut the histogram after the first 20 ones. Since we are losing the outliers, we add a table that contains such information.

Figure 3.2 shows the final dashboard. We can notice that:

- The average resolution time is about 2 hours.
- There are some outliers, but their number is really low in comparison to the amount of entries.
- Most incidents are closed in less than 10 hours.

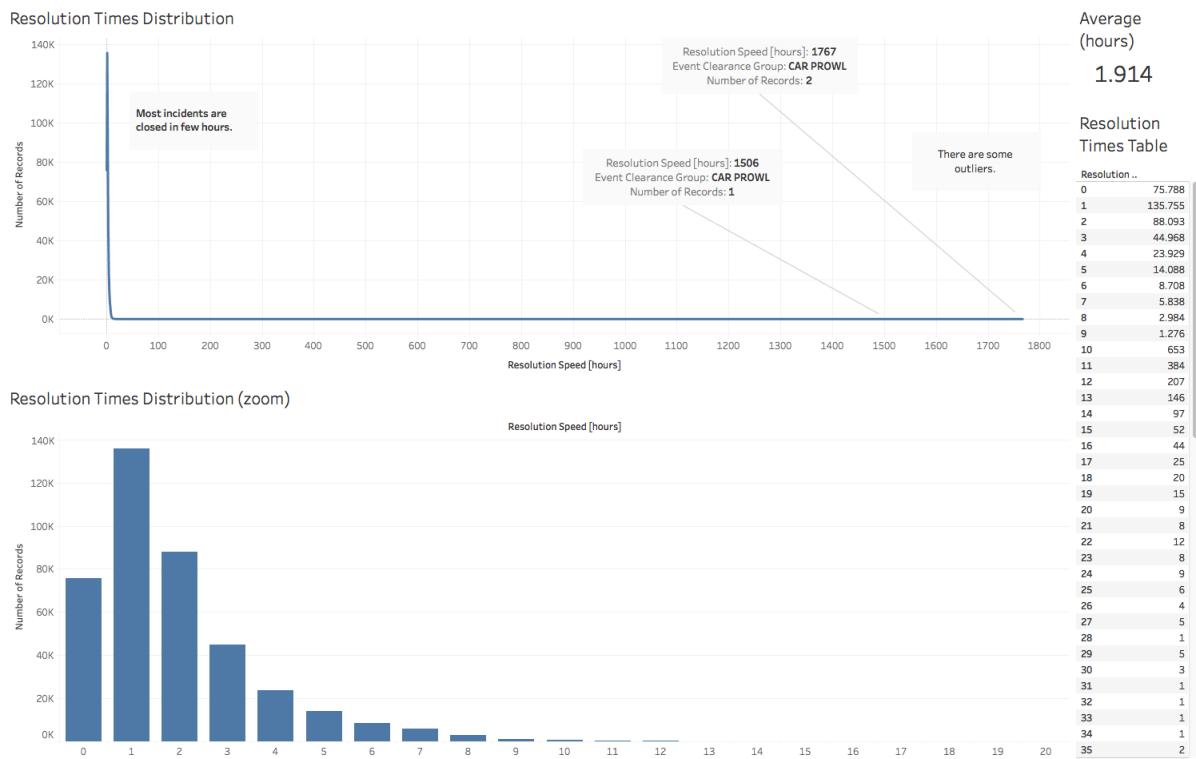


Figure 3.2: Distribution of the resolution times. The dashboard is called *Resolution Times* in Tableau.

### QUESTION 3.2

*Are there certain types of incidents having a much lower resolution speed than others? If so, which are these?*

The visualization uses a bar chart. Each type of incident is encoded in one bar and the height of the bar corresponds to the average resolution time for that type. The types are ordered by average resolution time decreasing, so it is easy to spot the outliers. Color is used to overload the encoding of the incident type. The bar chart shows the average line to make comparison and reasoning easier. We prefer to

use the x-axis for the type and the y-axis for the average resolution time since the visualization with swapped axes does not fit a single screen.

Figure 3.3 shows the visualization. We can notice that:

- *Homicide* is the type with the highest average resolution time. This is not a surprise, since such an incident requires very accurate investigations.
- *Public Gatherings* have the seconds highest average resolution time. This is probably due to the high number of people involved in the incident.
- *Vice Calls* have the fastest resolution time, on average only about 30 minutes.
- *False Alarms* are also resolved quite fast, less than 1 hour on average.
- All other types have resolution times between 1 and 3 hours on average.
- The distribution is skewed to towards types with higher resolution times.

Resolution Time by Type

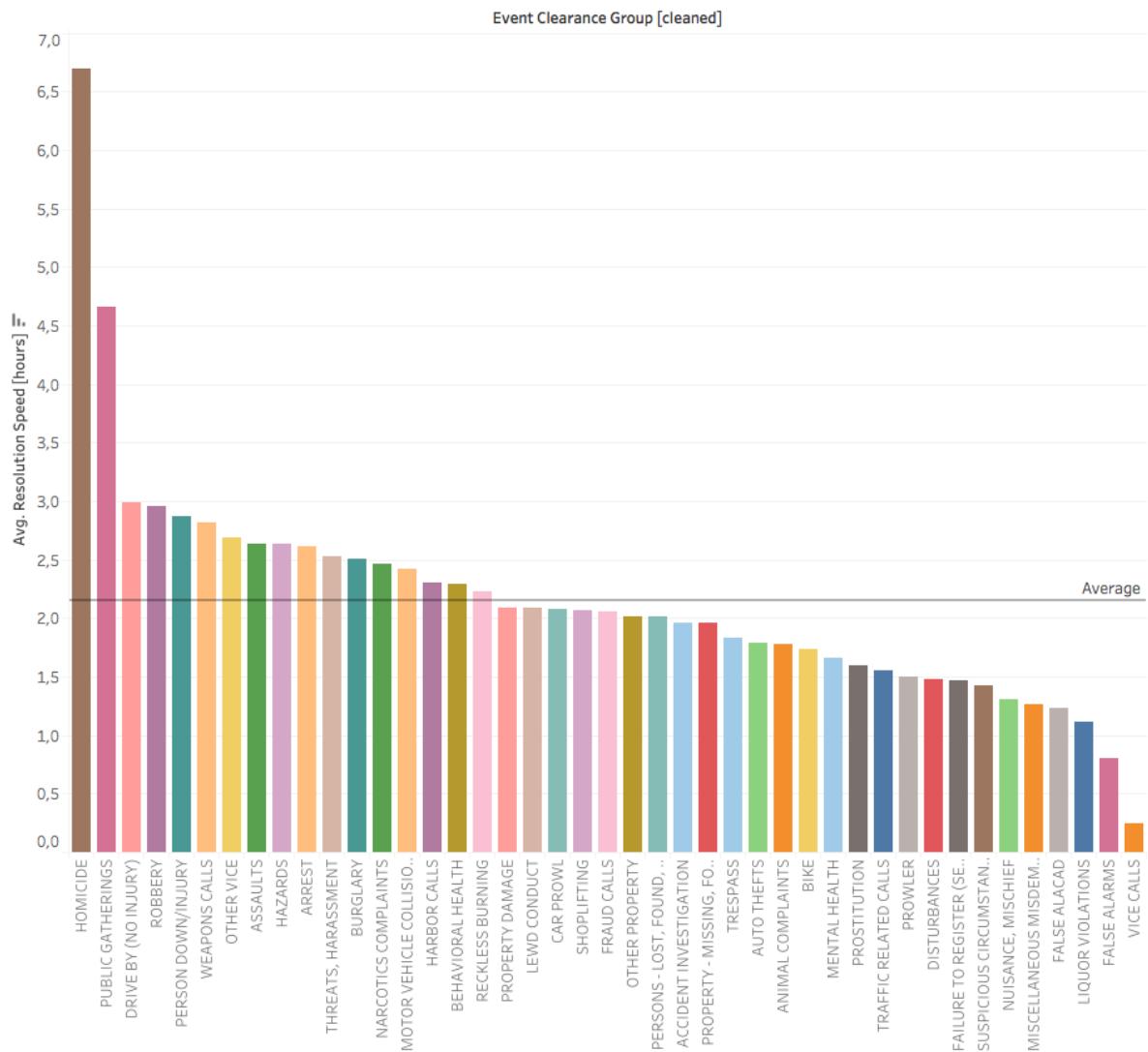


Figure 3.3: Incidents' resolution time by type. The sheet is called *Resolution Time by Type* in Tableau.

### QUESTION 3.3

*Does the resolution speed depend on the time period (e.g., year, season of the year)?*

The visualization uses a scatterplot to visualize the resolution speed of incidents over time. The x-axis encode the time at scene, the y-axis the resolution speed in hours. We use color to overload the encoding of the resolution speed. In particular, we use a continuos heat colormap. We show the average line in order to be able to compare time ranges with the average behaviour.

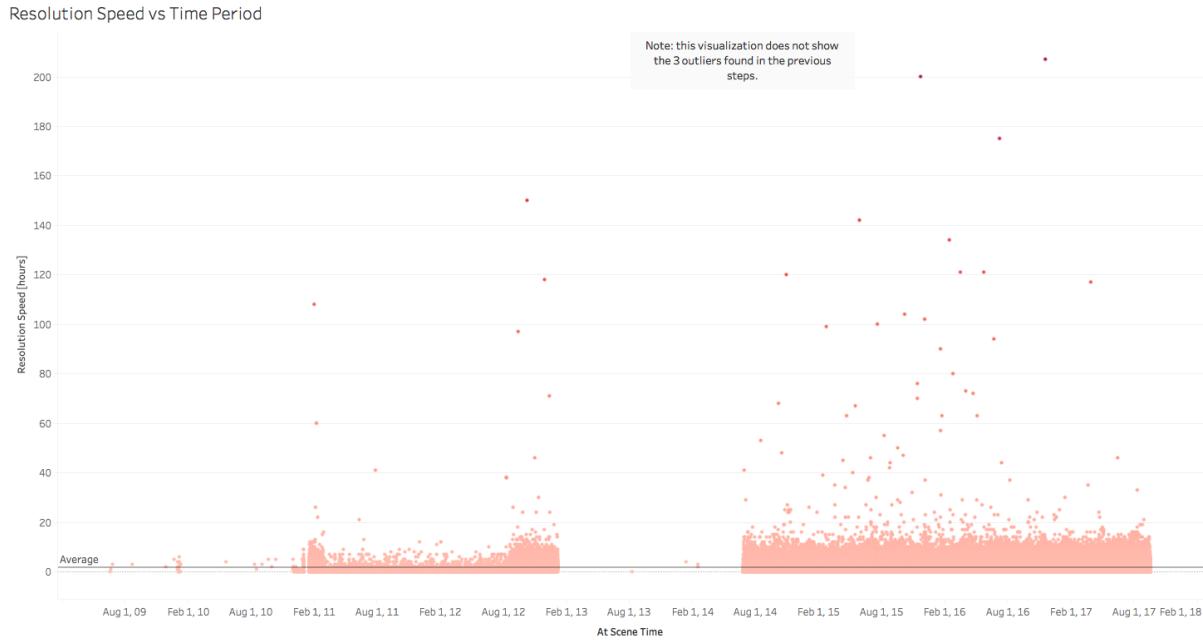
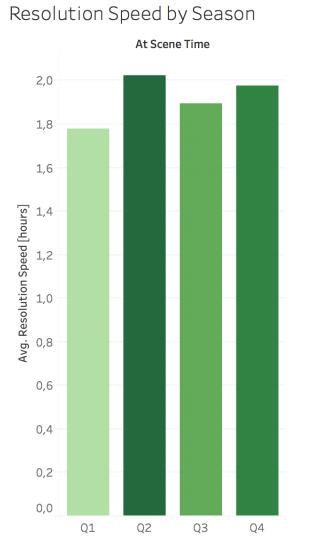


Figure 3.4: Incidents' resolution speed over time. Please note that the 3 outliers cited in the text are excluded from this visualization. The sheet is called *Resolution Speed vs Time* in Tableau.

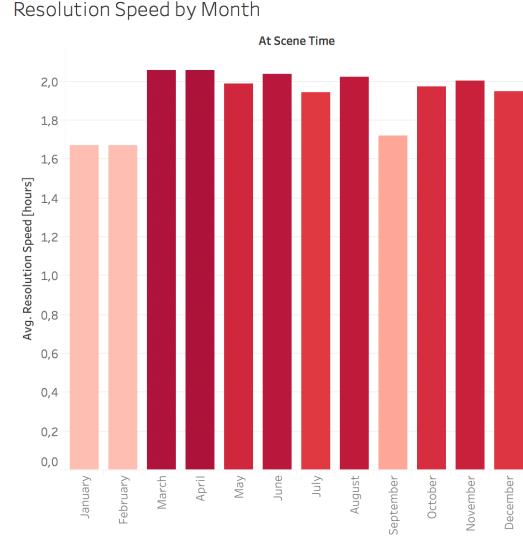
The visualization is difficult to read due to the presence of the 3 outliers described in Question 3.1. Since we have discussed them already, we remove them from this visualization. The result is shown in Figure 3.4:

- There are 4 slots of times that appears differently. The 1<sup>st</sup> (02/2009 - 01/2010) and 3<sup>rd</sup> (02/2013 - 07/2014) slots have almost no data. There is not much we can say about those periods.
- The 2<sup>nd</sup> slot (02/2011 - 01/2013) has a period where incidents are resolved faster, in particular from 03/2011 to 08/2012; in this period most incidents are closed in less than 2 hours.
- The 4<sup>th</sup> slot (08/2014 - 09/2017) seems to have a uniform distribution for the resolution speeds, with an average higher than the 2<sup>nd</sup> slot.

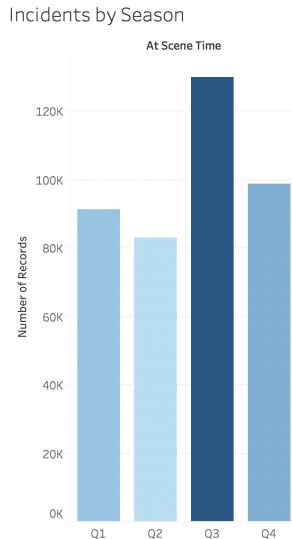
Overall, the resolution speed seems to be not correlated to the particular period of the year (i.e. month or season). To confirm this hypothesis, we create 2 new bar charts that show respectively the average resolution speed by season and by month. We use color to overload the encoding of the values showed by the bars to make the plots easier to read. We also create 2 bar charts to show the number of incidents by season and month, to make sure that the number of incidents does not vary significantly over the year (since this would probably affect the average resolution speed). From Figure 3.5 we can confirm that there is no significant correlation between the period of the year and the incidents' resolution speed. Also, we can notice that there is a higher number of incidents (about 20%) during the summer months (July, August and September); however, this does not influence the average resolution speed. The scatterplot and the bar charts are grouped in the *Resolution Speed* dashboard.



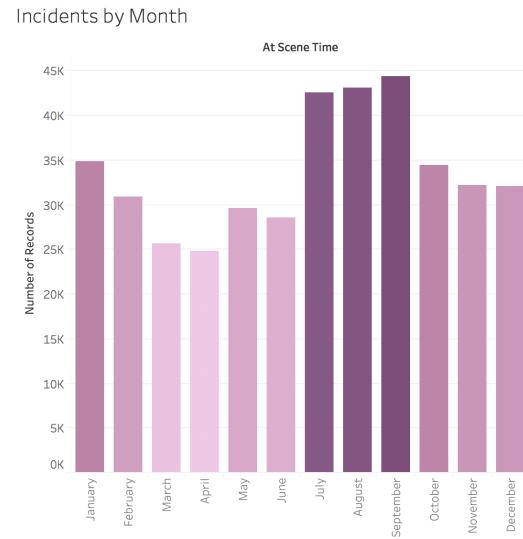
(a) Resolution speed by season.



(b) Resolution speed by month.



(c) Number of incidents by season.



(d) Number of incidents by month.

Figure 3.5: Overview of the resolution speed and number of incidents by season and month. The bar chart are grouped in the *Resolution Speed* dashboard in Tableau.

## 4 EXPLORING THE INCIDENTS' CLASSIFICATION

### QUESTION 4.1

*Incidents have two types: the type of the incident at the moment it was reported (Initial Type Group) and the type of the incident after it has been studied and classified (Event Clearance Group). Obviously, there should be some kind of correlation of the latter with the former (for instance, incidents reported as robberies are very likely to be also classified finally as robberies). How strong is this correlation?*

The visualization uses a matrix plot to show this correlation. The rows show the incident types at the moment it was reported, while the columns show the type it was classified as (we will refer to it as the “final” type). Each cell in the map contains a square, whose area and color are proportional to the

number of record for the particular row and column. Color uses a continuos heat colormap.

The order of rows and columns is initially alphabetical. However, this order does not have any particular meaning, since there is no 1-to-1 mapping between the categorical values of *Initial Type Group* and *Event Clearance Group*. Starting from such an order, we have manually moved some rows and columns to types which seems to be possibly correlated, in order to show this correlation along the main diagonal. For example, *Initial Type Group* has the value *Parking Violation*, but *Event Clearance Group* does not have it: we have it closed to *Traffic Related Calls*.

Figure 4.1 shows the visualization. We can see that:

- After some reordering of columns, most of the big squares are along the main diagonal. This means there is a high correlation between the initial and final types of incidents.
- In many cases there is a match between the initial and final type (e.g. *Disturbances*). Nevertheless, in some cases a single initial type matches multiple more specific final types (e.g. *Theft* matches *Car prowls*, *Other property* and *Shoplifting*).
- The first row of the table has a different behaviour and represents the incidents without an initial type. This group matches different final types, mostly *Disturbances*, *Suspicious circumstances* and *Traffic related calls*.
- There is quite a high number of incidents without an initial type, but only few without a final type. Incidents are better classified after their resolution.

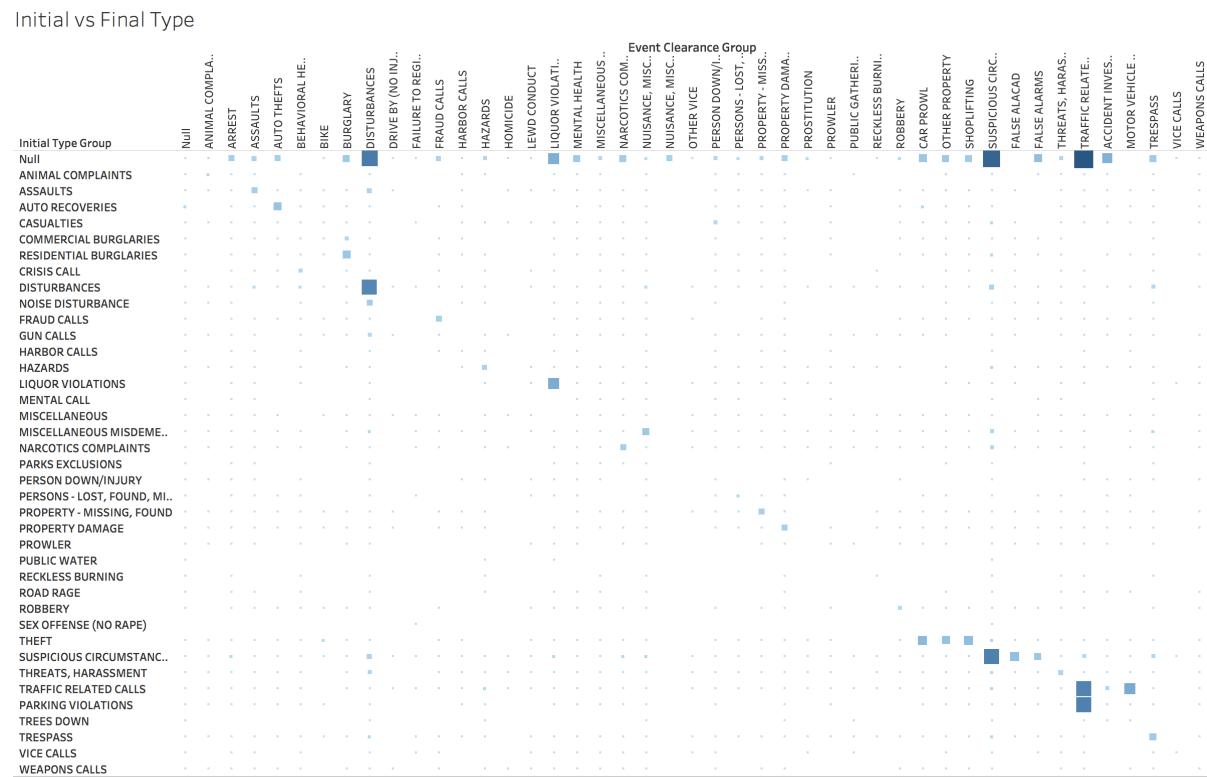


Figure 4.1: Correlation between the initial and final incident type. The sheet is called *Initial vs Final Type* in Tableau.

The visualization answers the question, but requires some manual work to find the matching between initial and final types. Moreover, it wastes a lot of space, since most initial types are highly correlated to only a few final ones. This makes the square dedicated to interesting correlations very small and difficult to compare with the others. The use of colors as an overloading helps with that, but does not really solve the problem.

To better visualize the interesting correlations and remove the manual reordering step, we use a treemap. We filter out values without an initial type since we have discussed this case previously. We use the initial type for the first split for the treemap. Each rectangle corresponds to a pair initial and final types. Both the color and the area of each rectangle encode the number of elements for the particular pair. Tooltips allow the user to interactively get additional information for each rectangle, namely initial type, final type and number of matching incidents.

Figure 4.2 shows the visualization:

- There are some very big rectangles, which indicates a high correlation between the initial and final types.
- There are many small rectangles, which indicates a low correlation. For most types the area occupied from small rectangles is negligible, but for some other (e.g. *Suspicious circumstances*) this area is quite relevant. This fact was very difficult to spot in the previous visualization.
- Mapping between the initial and final types do not need to be specified interactively by the user. This is the biggest improvement over the previous design.

To sum up, there is a quite strong correlation between initial and final types. However, a lot of incidents with some particular initial cases are often classified in a different way afterwards (e.g. *Suspicious circumstances*).

Initial vs Final type (exclude nulls)



Figure 4.2: Correlation between the initial and final incident type (excluding incidents without an initial type). The sheet is called *Initial vs Final Type (exclude nulls)* in Tableau.

Another interesting question is whether final event types emerged from very different initial types, i.e. if there are types of incidents that are particular difficult to classify by the initial incident report. To answer this question, we use the same treemap, but with a different color coding: color now represents the final events type. Since there are too many types for an effective categorical color map, we group the less frequent ones into the *Others* type. The user can interactively use the legend to select only some types of final incidents in order to compare and find types with similar patterns.

Initial vs Final Type (final types spread)



Figure 4.3: Correlation between the initial and final incident type, using color coding for the final events types (excluding incidents without an initial type). The sheet is called *Initial vs Final Type (final types spread)* in Tableau.

Figure 4.3 shows the visualization. We can notice the following:

- Most final groups contain incidents that are initially classified in many different group.
- About half of the incidents in the final groups *Suspicious Circumstances*, *Disturbances*, *Trespass* and *Nuisance, Mischief* are initially classified as a different type. A possible interpretation is that these types are quite generic and match a wide range of incidents.
- Most of the incidents in the other final groups are classified correctly from the beginning.
- An exception is the final group *Traffic Related Calls* that maps both the initial groups *Traffic Related Calls* and *Parking Violations*. For some reason, the latter is not a final group.

## QUESTION 4.2

*How do the total number of incidents break down per incident reported type (Initial Type Group) and incident resolution type (Event Clearance Group)?*

To answer this question we have created 2 treemaps very similar to Figure 4.2. The former is a shows first the initial and then the final types (including null values), the latter shows first the final and then the initial types.

We can notice the following:

- About 3/8 of the incidents have no information about the initial type. Most of them are than classified in the most frequent final types (*Traffic Related Calls*, *Suspicious Circumstances*, *Disturbances* etc.).
- There are 5 main initial types of incidents that include about another 3/8 of the total, namely

*Suspicious Circumstances, Traffic Related Calls, Disturbances, Theft and Parking Violations.*

- The remaining 2/8 of the incidents are spread among different initial types.
- There is no problem of missing data for the final type: less than 1% of the incidents miss it.
- There are 4 main final types of incidents that include about 1/2 of the entries, namely *Traffic Related Calls, Suspicious Circumstances, Disturbances* and *Liquor Violations*.
- The remaining 1/2 is distributed among the other types.
- In most cases, the majority of incidents in each final type group matches the initial type. There is thus a correlation between the initial and final incident type, as said for Question 4.1

### QUESTION 4.3

*How do the total number of incidents break down per incident resolution type (Event Clearance Group) and incident resolution subtype (Event Clearance Subgroup)?*

The visualization uses a treemap, since groups and subgroups can naturally be organized in a tree hierarchy. Each group is represented by a rectangle, which can be divided in sub-rectangles to represent the subgroups, if any. Both color and area of the rectangles encode the number of incidents with the given type / subtype.

Incidents Types Distribution (Group and Subgroup)



Figure 4.4: Incidents by Type and Subtype. The sheet is called *Incidents Types Distribution (Group and Subgroup)* in Tableau.

Figure 4.4 shows the visualization. We can see that:

- There are 4 types that include about 50% of all of incidents, namely *Traffic Related Calls, Suspicious Circumstances, Disturbances* and *Liquor Violations*.
- Only 10 out of 45 type groups have subtypes.

- Groups that are divided in subgroups have only 2 or 3 of them.
- In most cases there is a dominant subgroups that includes most of the cases; the only exception is the group *Traffic Related Calls*, where incidents are distributed almost equally in both subgroups *Traffic Related Calls* and *Parking Violations*.

## 5 EXPLORING THE INCIDENTS' TEMPORAL DISTRIBUTION

### QUESTION 5.1

*How do the different types of incidents (as represented by Event Clearance Group) vary in number over the different years? Do you see different types of incidents having the same temporal variation pattern over the same year(s)?*

The visualization uses bar charts and the small multiple design principle. The x-axis represent the time (in years), while the y-axis encodes the number of incidents for the given year. The bar chart is replicated for each incident type; bar charts for different types are organized in columns. Color is used to overload the encoding of the number of incidents (showed by the y-axis). Trend lines are drawn for each small bar chart to show the trend of the number of incidents over year for each different type. The bar charts are ordered for the total number of incidents per type (decreasing). The types with less incidents are aggregated together in the *Others* group to spare space on the screen and thus better fit the visualization. Since there are very few data for 2009 and 2010, those years are removed from the visualization. We filter also null values for the date since we are interested in the trend of incidents over years.

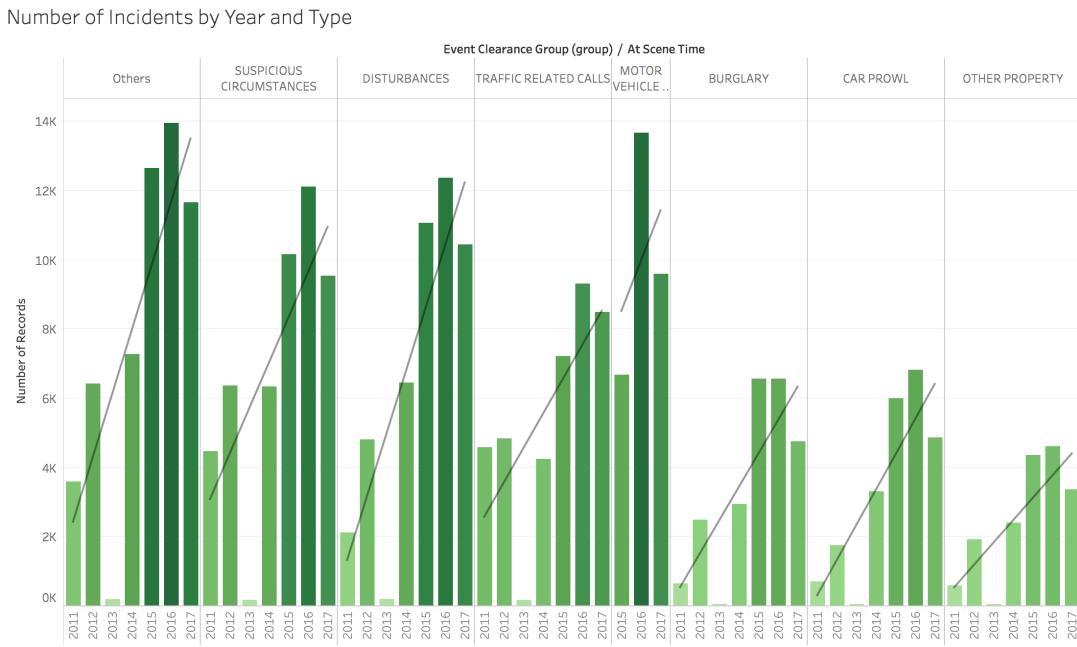


Figure 5.1: Number of incidents by year and type. The sheet is called *Number of Incidents by Year and Type* in Tableau.

Figure 5.1 shows the visualization. We can observe the following:

- Only 2 incidents' types decreased over time, namely *Accident Investigation* and *Mental Health*. For both types there is no single incident reported in years 2016 and 2017. We believe that this might be due to a change in the classification of incidents in types between 2015 and 2016.

- All other types of incidents showed an increase over time.
- Most incidents show a common pattern: there is an increase from 2011 to 2012 and a clear drop in 2013; 2014 has about the same number of incidents as 2013; years 2015 and 2016 have a fast increase, with the peak in 2016; 2017 has slightly less incidents than 2016 (this is probably due to the missing data for the last 3 months of 2017).

The visualization allows to answer the question. However, it does not show the general trend of the incidents number over years. To solve this problem, we created a second visualization using another bar chart. The y-axis represents the incidents' types, the x-axis and color encode the number of incidents for the given year. We use a continuous heatmap for the color, but a different color for the maximum value in comparison with the visualization in Figure 5.1.

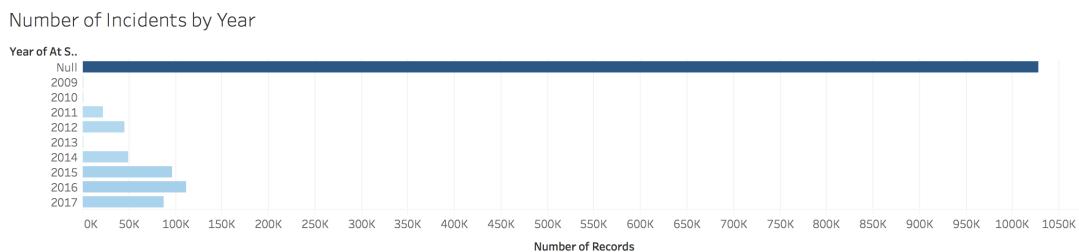


Figure 5.2: Number of incidents by year. The sheet is called *Number of Incidents by Year* in Tableau.

The visualization is showed in Figure 5.2. We can notice that:

- The majority of incidents in the dataset do not have the date information. The confident of the answer to the question is thus low.
- The general temporal distribution of incidents has the same trend of most incidents' types.

The visualizations are grouped together in the *Incidents Temporal Distribution by Type and Year* dashboard.

## QUESTION 5.2

*How do the different types of reported incidents (as represented by Initial Type Group) vary over the 24 hours of a day, for the entire data collection? Are there certain hours having a higher rate of reported incidents?*

The visualization uses bar charts and the small multiple design principle. For each chart, the y-axis represent the hour of the day, while the x-axis and the color encode the number of incidents for the given hour. The chart is replicated for each incident type, organized in columns. The bar charts are ordered for the total number of incidents per type (decreasing, from left to right). The types with less incidents are aggregated together in the *Others* group to spare space on the screen thus better and fit the visualization and null values are filtered out. We add the average line for each type in order to easier spot hours with more or less incidents than the average.

From Figure 5.3 we can notice a common pattern among different types:

- The number of incidents is generally smaller in the night / morning (23 - 11) and higher in the other hours.
- There are 2 peaks, namely at 12 and 20: after these hours, incidents are still higher than the average but tend to gradually reduce in number.
- For most incidents the peaks have a similar number of incidents; there are a couple of exceptions, namely *Fraud Calls* and *Commercial Burglaries*, where the peak at 12 is much higher

than the one at 20.

Incidents Distribution over the Day

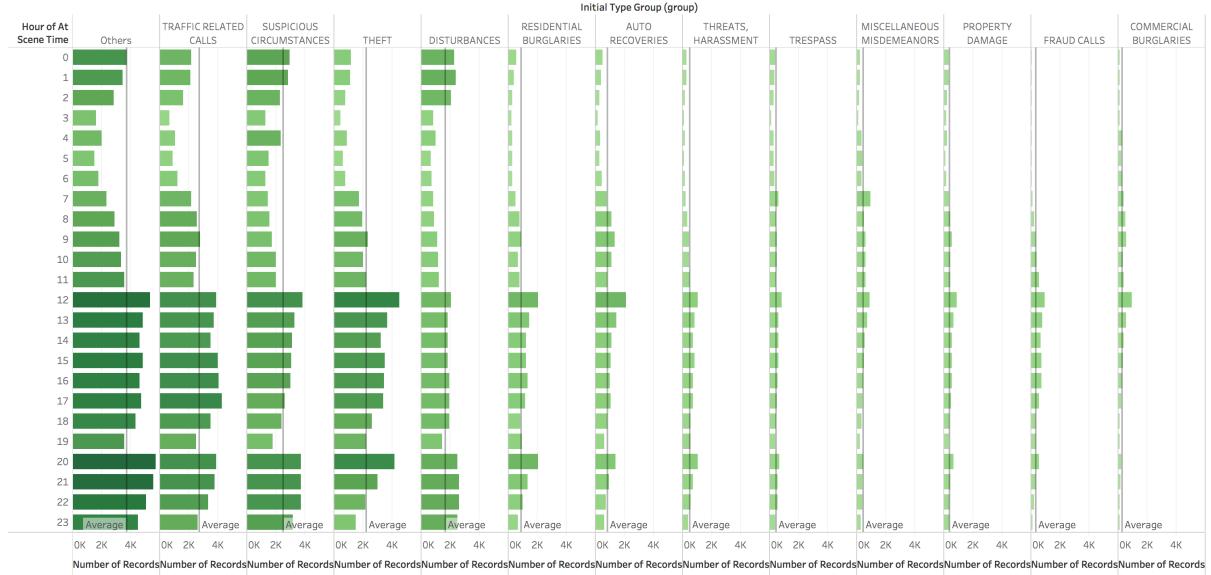


Figure 5.3: Number of incidents by type over the different hours of the day. The sheet is called *Incidents Distribution over the Day* in Tableau.

### QUESTION 5.3

*How long does it take to clear incidents, as a function of the hour when they were reported? For example, do incidents reported at noon get cleared faster than incidents reported in the middle of the night?*

The visualization uses a bar chart. The x-axis represents the hour of the day, while the y-axis shows the average resolution time for incidents reported in the given hour. Color overloads the value showed in the y-axis. The average line helps to compare particular time slots with the average behaviour.

Best practices for visualization suggests to use line charts instead of bar charts for continuous values like time. In this case, however, we have discretized time into buckets of one hour and aggregated them together to compute the average. In other words, we treat time as a discrete ordinal value.

Figure 5.4 shows the visualization. We can notice that:

- The average time required to close incidents is not constant over the different hours of the day.
- Incidents that happen in the slots 11 – 14 and 19 – 20 take significant longer to be closed.
- Incidents that happens in the early morning (5 – 7) require on average more time than the incidents in the night and late morning (9 – 10).
- The time slots between 5 – 7, 11 – 14 and 19 – 20 have a faster resolution speed.

To sum up, there is a correlation between the hour of the day and the time required to close incidents. The visualization fully answers the question.

We have created a second visualization to check if this behaviour is common to all years. The visualization uses again bar charts, one for each year. The last column shows the average over all years as a reference for comparisons. The axes are swapped in comparison to Figure 5.4 to better fit the screen.

From Figure 5.5 we can notice that:

Resolution Speed vs Day Hour

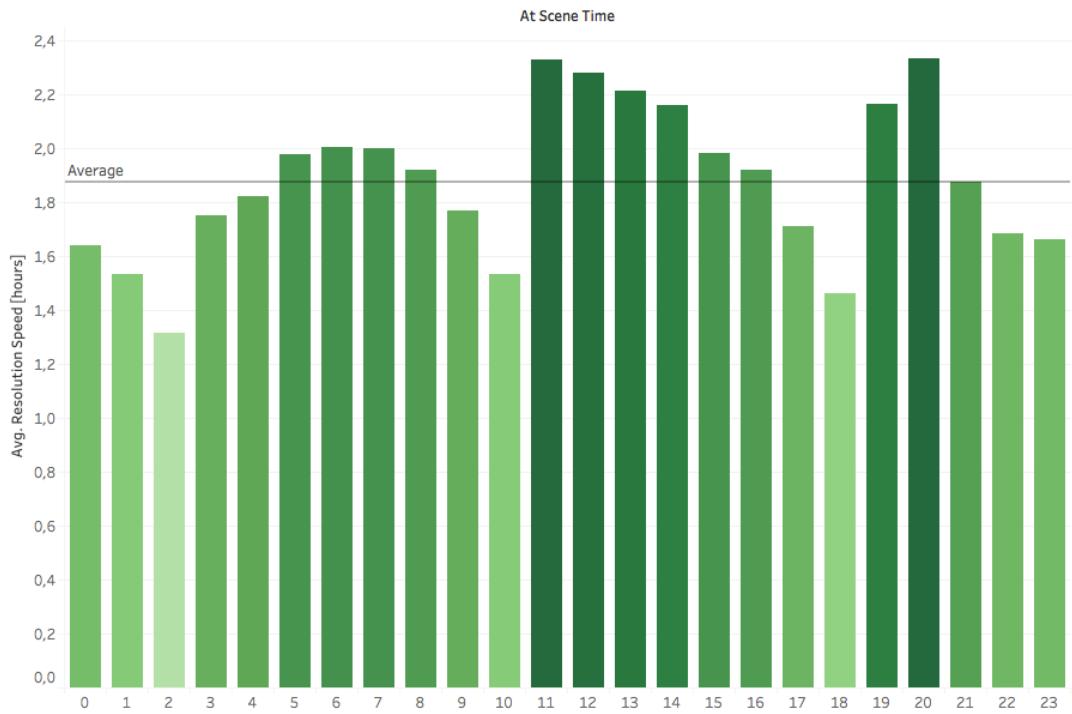


Figure 5.4: Resolution speed of incidents reported in different hour. The sheet is called *Resolution Speed vs Day Hour* in Tableau.

- In 2011 incidents are closed much faster than the average. There are some hours with a slightly higher resolution speed (9, 14, 19), but the difference with the rest of the day is small.
- Incidents that happens at 20 in year 2012 seems to take much longer to close than the average. This is probably due to the outliers discussed in Section 3, since both have very long resolution times and are in that slot of time.
- Years 2013 have some picks in the slots 12–13 and 18–20. However, there are probably missing data for this year (see Section 1, so the confidence for this observation is low).
- Years 2014, 2015, 2016 and 2017 have a behaviour identical to the average over all years.

To sum up, the resolution speed for the different hours of the day seems to be stable over years 2014 to 2017. The previous years have slightly different behaviour, but still share some patterns. Incidents that occurs in slots 5–7, 11–14 and 19–20 take more time to be closed, while incidents between this slots are closed much faster.



Figure 5.5: Resolution speed of incidents reported in different hour by year. The sheet is called *Resolution Speed vs Day Hour (by year)* in Tableau.