Practical Assignment: Constructing Visualizations

1. Introduction

This document describes the partial solution to the practical assignment coming with the Practical Information Visualization course. To understand the material in here, you should have read the assignment description (**Practical Assignment – Description**). Also, to make best use of the material here, you should have tried to build a few Tableau visualizations trying to answer the questions described in the assignment description.

The aim of this document is to provide details on how to build visualizations which can be used to answer the assignment's questions. The document can thus be used in two ways:

- If you get stuck with one of the tasks in the assignment, read here how to construct visualizations which cover that task
- If you have already constructed visualizations for solving a task, compare them with the ones presented here. Note that there is no unique way to construct a visualization that covers a given task. As such, your solutions may be different than the ones presented here, and still be good solutions.

Notes:

- Read this document step-by-step: Study the solution presented here only after you have tried yourself to solve the respective step.
- This document does not provide (full) answers to the question in the assignment description. It only shows you how to build visualizations that you should next use and interpret to get the actual answers.
  The actual implementations of the Tableau visualizations described here are provided online at https://public.tableau.com/profile/alex.telea#!/vizhome/SeattlePoliceDeptIncidents
  You should download and use these only when you have not been able to construct the visualizations following the textual description in this document.

Task 1: Exploring the incidents' geographical distribution

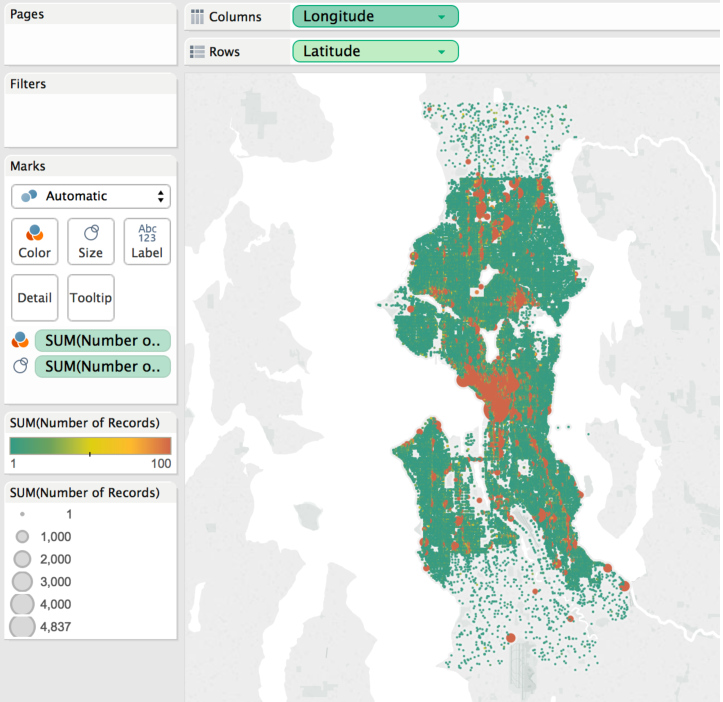
For this task, the best way is to use a map visualization. We can create this by dragging the Longitude and Latitude fields (from the Measures panel) to the Columns and Rows fields of the current sheet. However, if Longitude and Latitude are seen as measures, their values will be aggregated (averaged), thereby resulting in the display of a single dot. To get dots corresponding to all different locations, right-click Longitude and Latitude in the Columns and Rows fields, and change their type from measure to dimension.

Q1.1

To answer this question, we can color the incidents' locations by the number of incidents happening at that location. For this, drag the Number of records field (from the measures panel) to the Color field in the Marks panel. Next, select some suggestive colormap – for example, a green-yellow-red one, which suggests criticality (green=few incidents, red=many incidents). Next, examine the range of the number of incidents per location. You will see that it is quite high (thousands). However, the map appears largely green. That means that there are just a few locations with a large number of incidents. To get a better view of the variation of number of incidents over the map, we can reduce the range to, say [0..100]. Do this by right-clicking the color legend, select 'Edit colors', and change the range using the advanced options.

Still, since the incidents are quite dense over the map, high-frequency locations do not show up very well. We can show them better by using larger dots for high-frequency locations. For this, drag Number of recods to the Size field in the Marks panel, and optionally tune the scale of the dots.

The resulting visualization should resemble the one below:



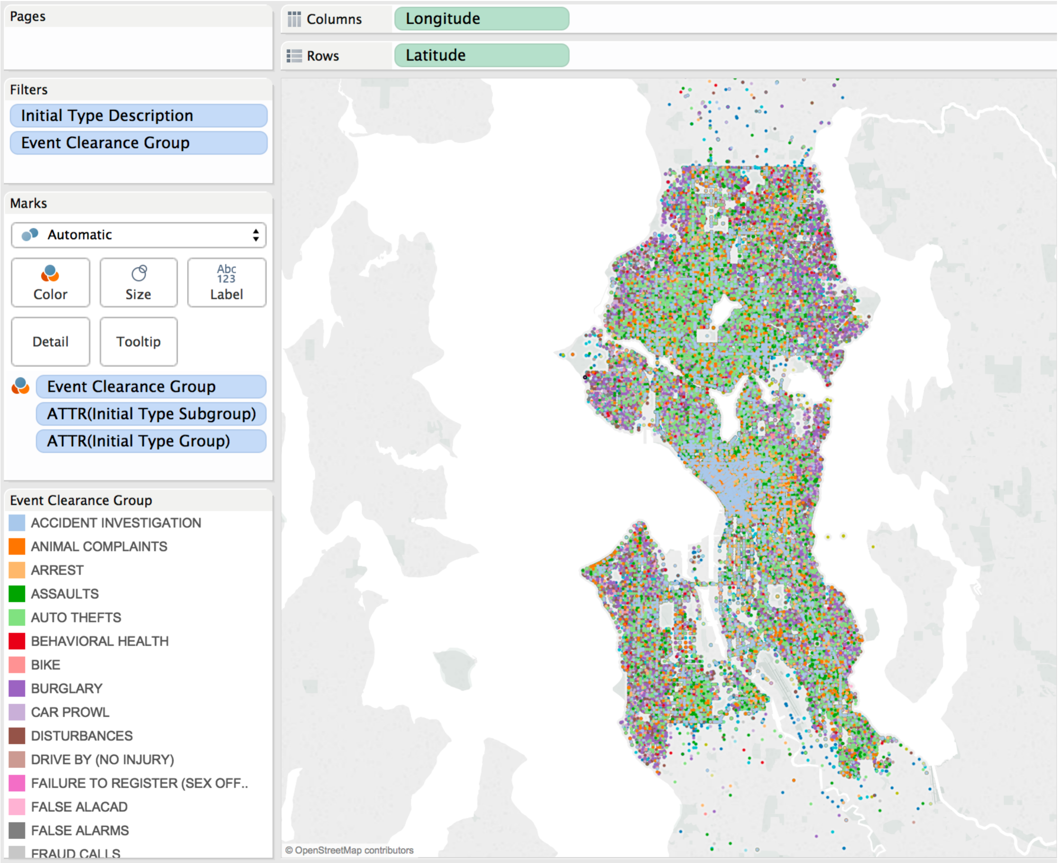
Columns: Longitude
Rows: Latitude
Marks: Automatic
Color: SUM(Number o..
Size: SUM(Number o..

Q1.2

To answer this question, we need to show somehow the type of incident. Let us select, for type, the field Event Clearance Group (you can also use Initial Type Group). We construct a map just as described for Q1.1. However, since we do not want to show incident density, we do not use the dots' sizes. Also, to show event types, we map Event Clearance Group to the Color field.

When doing this, you will notice a very large number of dark blue dots. Upon closer inspection, we see that these correspond to incidents with no Event Clearance Group data (this is indicated by the Null value corresponding to dark blue in the color legend). Showing these events does not help us much. To remove them from the view, we use a filter: Drag the Event Clearance Group field from the Dimensions panel to the Filters panel. Next, double-click this filter and set it to exclude all Null values.

The resulting visualization should resemble the one below:



Columns: Longitude

Rows: Latitude

Filters: Initial Type Description, Event Clearance Group

Marks: Automatic

Color: Event Clearance Group
ATTR(Initial Type Subgroup)
ATTR(Initial Type Group)

Event Clearance Group
- ■ ACCIDENT INVESTIGATION
- ■ ANIMAL COMPLAINTS
- ■ ARREST
- ■ ASSAULTS
- ■ AUTO THEFTS
- ■ BEHAVIORAL HEALTH
- ■ BIKE
- ■ BURGLARY
- ■ CAR PROWL
- ■ DISTURBANCES
- ■ DRIVE BY (NO INJURY)
- ■ FAILURE TO REGISTER (SEX OFF..
- ■ FALSE ALACAD
- ■ FALSE ALARMS
- ■ FRAUD CALLS

To see the specific spatial distribution of a given event type, we can now select that event type in the color legend. When doing so, only locations corresponding to that type of event are shown on the map.

Task 2: Exploring the incidents' geographical and temporal distribution

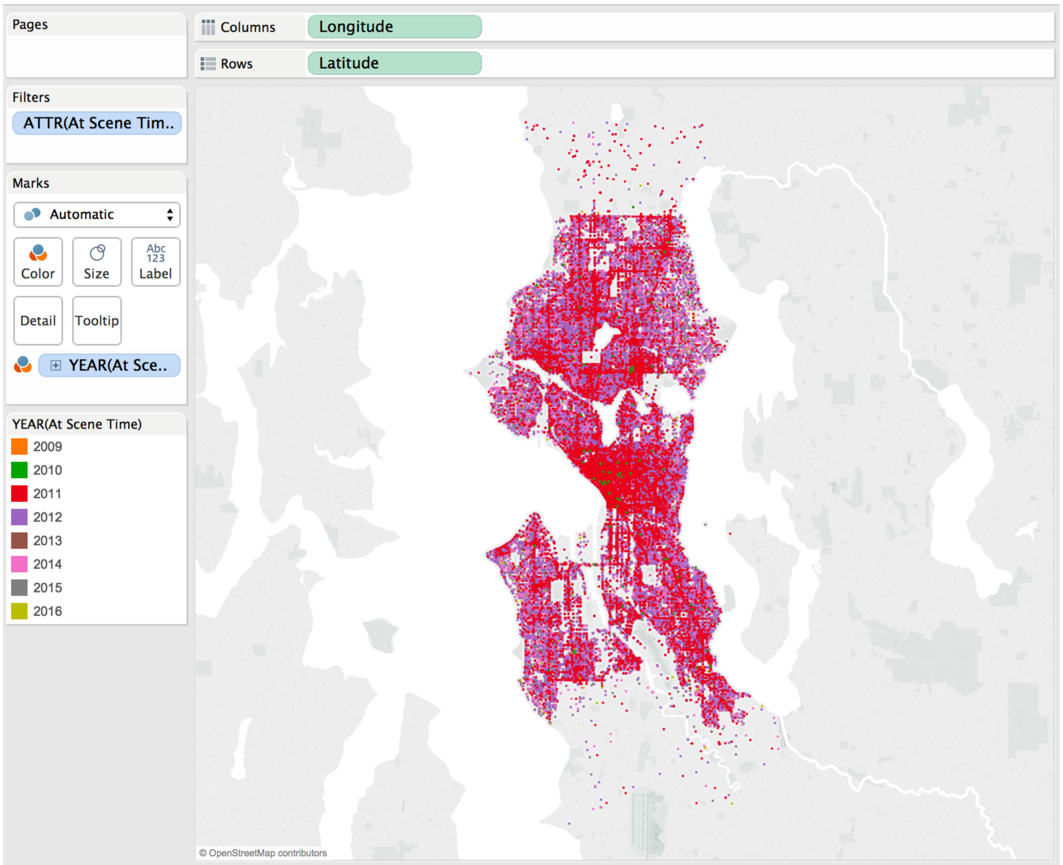
Q2.1, Q2.2

For these questions, we can refine the same type of map used for Task 1. Specifically, we map Latitude and Longitude to Rows and Columns, and map the year of the event to color. Note that the year of an incident is present in various fields – both At Scene Time and Event Clearance Date could be used for that, if we assume that the time difference between reporting an incident and solving it is far smaller than one year. Let us next assume we select the At Scene Time field for this.

To obtain a color mapping which corresponds to years, we have to bin the At Scene Time into values corresponding to different years. To do this, right-click the field At Scene Time in the Color slot of the Marks panel, and select the 'Year' option in the menu. This ignores the months/day/hour/... details of the At Scene Time field. Next, change the variable's type from continuous to discrete. This reduces the At Scene Time variable to discrete values corresponding to the different year values present in the dataset.

Looking at the resulting visualization, we find a lot of dark blue points. These correspond to incidents having Null for At Scene Time. We would like to filter these out. For this, drag At Scene Time to the Filters panel, and right-click it to exclude the Null values.

The resulting visualization should resemble the image below:



Columns: Longitude
Rows: Latitude
Filters: ATTR(At Scene Tim..
Marks: Automatic
YEAR(At Sce..

YEAR(At Scene Time)
🟧 2009
🟩 2010
🟥 2011
🟪 2012
🟫 2013
2014
⬛ 2015
🟨 2016

To further see the spatial distribution of incidents over a single year, select the respective year in the color legend.

Task 3: Exploring the resolution speed

Q3.1, Q3.2, Q3.3

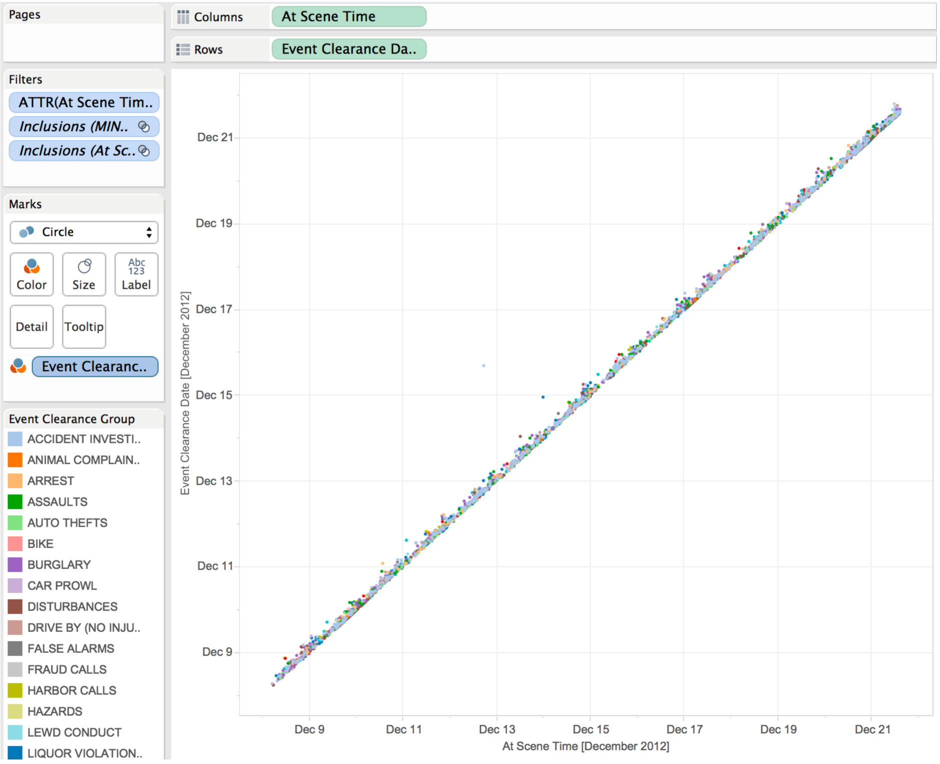
These questions are a bit more subtle to address. Thinking it over, we basically have two variables: At Scene Time and Event Clearance Date. Obviously, the second one is a function of the first one (the clearance date can only take place after the police has been at the event scene and analyzed the event). As such, a good way to answer this question is to construct a scatterplot of the two variables.

For this, drag At Scene Time and Event Clearance Date to the Columns and Rows of the current sheet. This creates a scatterplot with At Scene Time on the *x* axis and Event Clearance Date on the *y* axis. Next, make sure that the two fields are classified as continuous in the Columns and Rows slots, and that the option 'exact date' is selected for both (since we really want to look at the correlation of the precise, exact, dates). For the above, use the right-button menu on the At Scene Time and Event Clearance Date fields in Columns and Rows.

You will notice that the obtained scatterplot shows an almost perfect linear correlation of the two variables. However, the time range (six years) makes it very hard to see minute details of the correlation of the two fields. Also, interacting with the scatterplot is quite slow (no wonder, since we're visualizing over one million records). To simplify the scatterplot, first exclude all records having Null values for the At Scene Time and/or Event Clearance Date. Next, use the mouse to select a small rectangular area capturing a part of the dot-line appearing in the plot. A good idea is to select a time-range of about a week.

Finally, to understand which events we are looking at, drag Event Clearance Group to the Color field of the Marks panel.

The resulting visualization should resemble the image below:

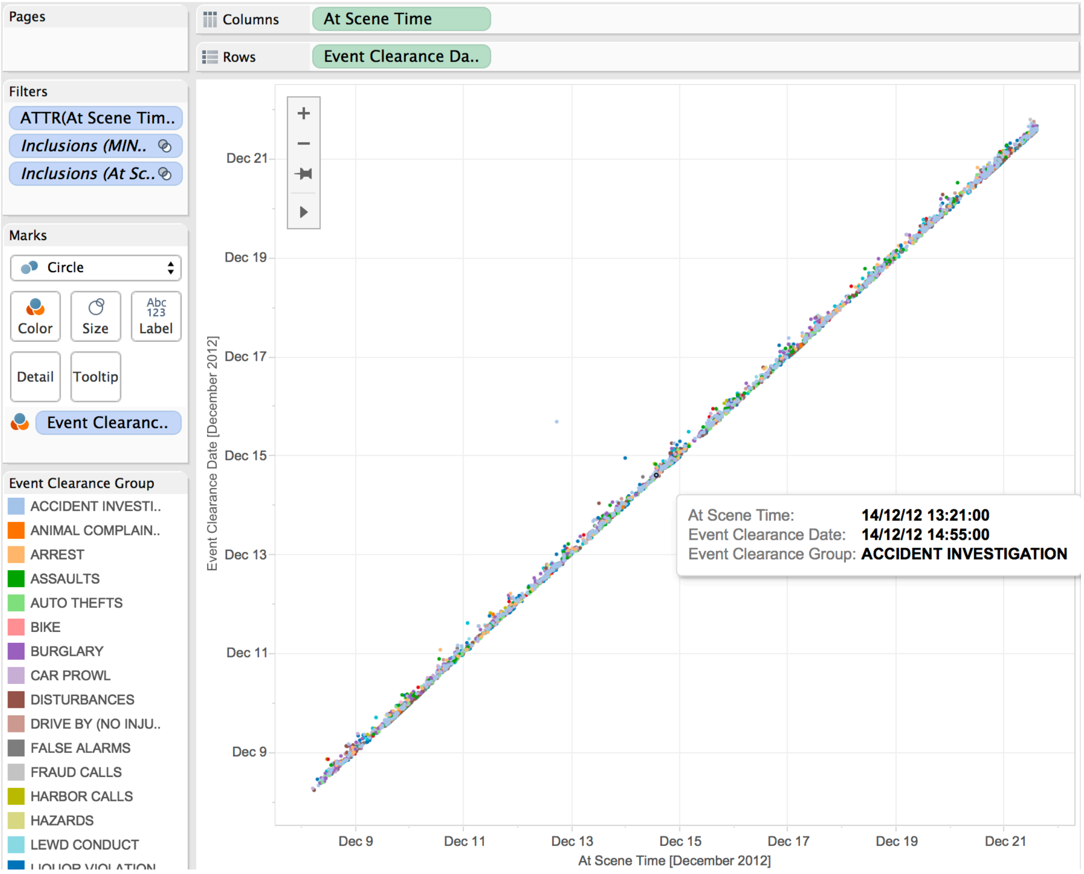


We can now see the linear correlation much better. As visible, all points (events) are located above the diagonal line. This is normal – an event is always cleared *after* it happened. Points very close to the diagonal have very short resolution times. Points located higher above the diagonal have longer resolution times.

To discover possible correlations between the event type and resolution speed, select the different event types in the color legend. Next, compare the distribution of dots (for each event type) with respect to the diagonal. Again, dots higher than the diagonal indicate longer response times.

To find what is the average response time, you can you can simply brush several points close to the middle of the local distribution-band for a given At Scene Time value. Use the tooltip to investigate the actual difference between Event Clearance Date and At Scene Time for the brushed point. To have the tooltip show these values, drag the Event Clearance Date and At Scene Time fields into the Tooltip slot of the Marks panel. Repeat the operation for a few points, and note down the time differences.

The figure below shows the effect of brushing a point to investigate its actual time values.



At Scene Time: 14/12/12 13:21:00
Event Clearance Date: 14/12/12 14:55:00
Event Clearance Group: ACCIDENT INVESTIGATION

Task 4: Exploring the incidents' classification

Q4.1

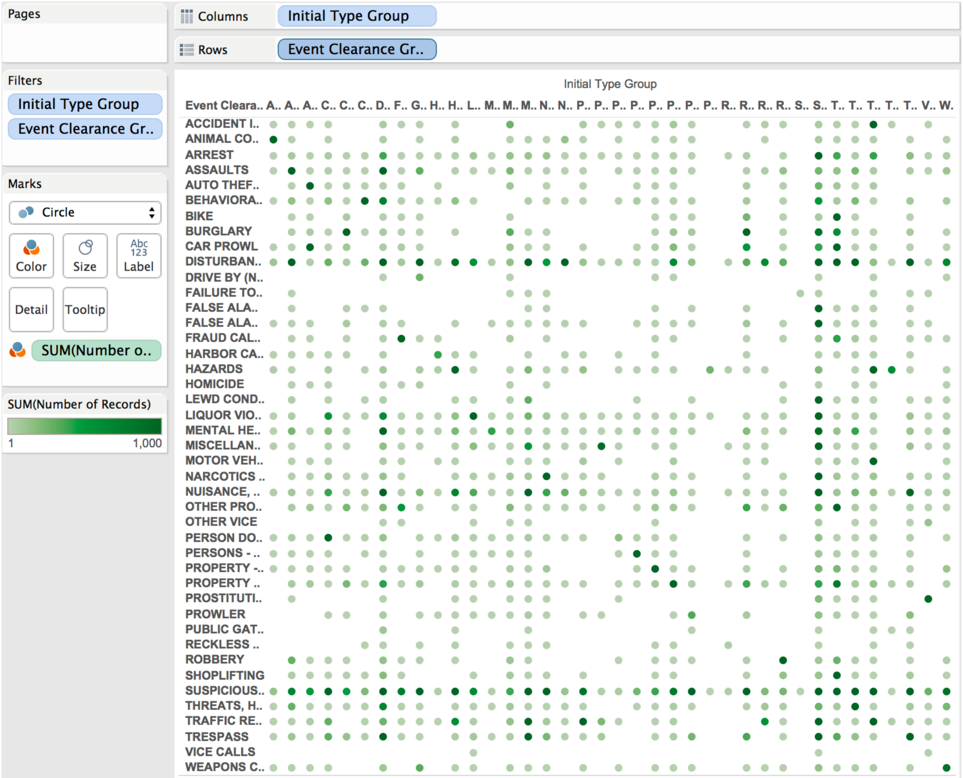
This question also requires a bit of thinking. Consider that we have N possible values for Initial Type Group and M possible values for Event Clearance Group. In theory, any of the Initial Type Group values can lead to any of the Event Clearance Group values, if the police is allowed to reclassify any event (reported as one type) to any other type. So, to study this correlation, we nee to show an N*M matrix of possibilities.

Creating such a matrix visualization is simple: Drag the Initial Type Group and Event Clearance Group fields to the Columns and Rows of a new sheet. Next, we want to show which of the N*M classification possibilities is actually taking place. That is, we want to show which cells in the matrix are not empty. To do this, drag the Number of records to the Color field of the Marks panel. Finally, to make the visualization more suggestive, select Circle in the pull-down menu in the Marks panel.

This creates a matrix-like visualization where white spots indicate classification possibilities which never occur in the data; light-green spots indicate classification possibilities which do not occur often; and dark-green spots indicate classification possibilities which occur very often. To get an even better view of the variation of the data, use the pull-down menu of the color legend to change the range of the data to [1..1000] or something similar. This makes low values more visible in the plot.

Similarly to previous examples, we'd like to exclude records having Null fields for Initial Type Group and/or Event Clearance Group. Do this by constructing two filters for the respective values (see earlier explanations on filters in this document).

The obtained visualization should resemble the image below:



Columns: Initial Type Group

Rows: Event Clearance Gr..

Filters: Initial Type Group, Event Clearance Gr..

Marks: Circle

Color: SUM(Number o..

SUM(Number of Records): 1 – 1,000

Note that you can improve the readability of this visualization by rotating the labels of the Initial Type Group axis – for this, use the right-click menu on any of those labels.

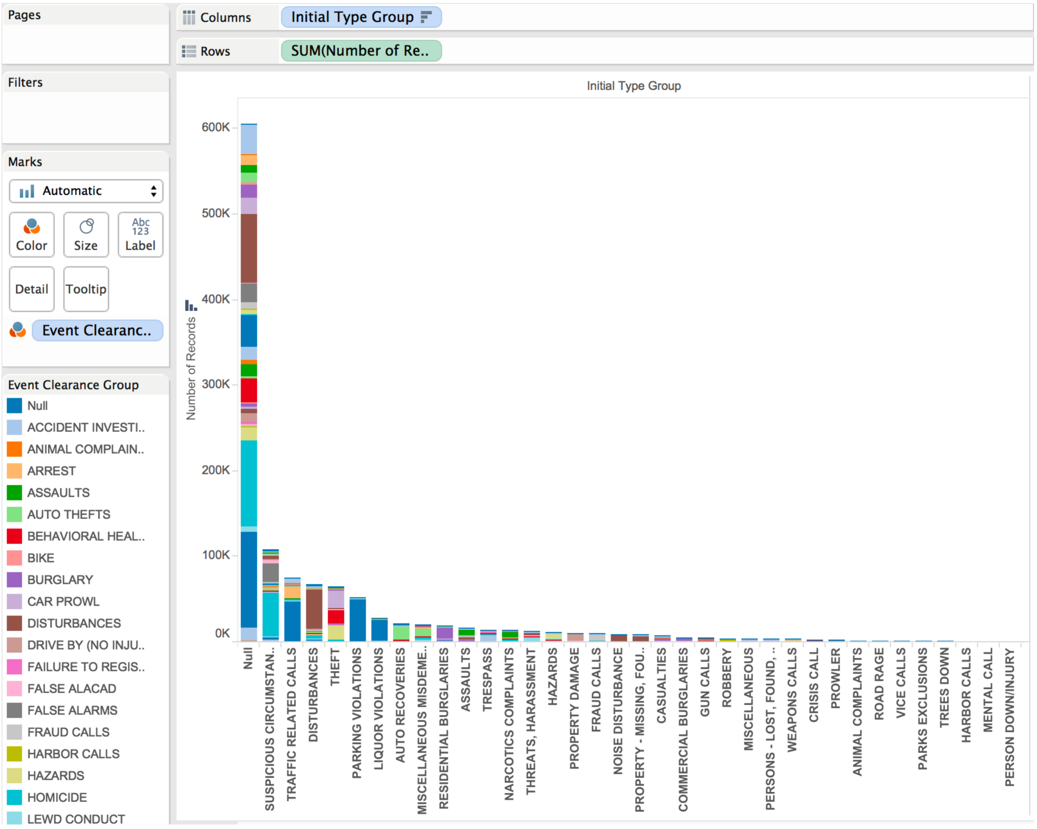
The insights provided by this visualization are quite interesting. For any row or column, we see many dots (and actually quite a number of dark dots). This means that there is no 1-to-1 mapping between a reported incident type and the type to which it was finally classified (!)

Q4.2

This is an easier question to answer. For this, we essentially can create a stacked bar chart. First, drag Initial Type Group to the Columns slot in a new sheet. Next, drag Number of Records in the Rows slot. This creates basically a simple bar chart showing the breakdown of number of incidents per Initial Type Group value. To further refine the view, we can break down each bar in the chart in the number of events of a certain Event Clearance Group. For this, drag the Event Clearance Group field to the Color field of the Marks panel.

Finally, you can use the sort buttons in the top toolbar to sort the bar chart e.g. decreasingly on the total number of incidents per Initial Type Group.

This should result in a visualization similar to the image below:



Columns: Initial Type Group

Rows: SUM(Number of Re..)

Marks: Automatic

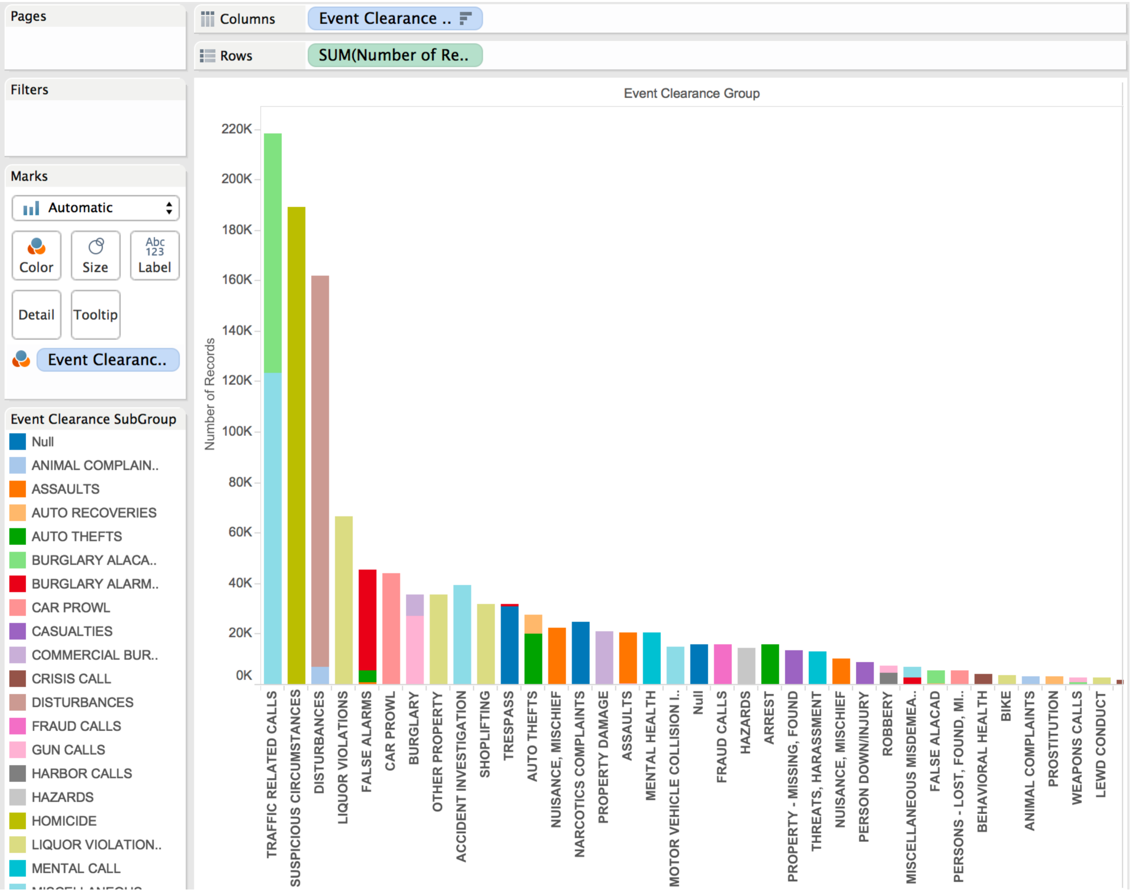
Color: Event Clearanc..

An interesting insight here is that the leftmost bar is much longer than the others. This bar indicates all events which have been reported without a specific type (that is, have Null for Initial Type Group value). Most events in the dataset are of this kind. Even more interestingly, we see that this bar is broken down in many different colors, which indicate how these events have been finally classified (value of Event Clearance Group). Looking at the color legend, we see that blue maps the Null value for Event Clearance Group (meaning, events which haven't been classified in any way). If we select this value, we see that the dark blue bar-fragments that get highlighted in the visualization are tiny. This is good news for the Seattle police department: Overall, it means that even if most the incidents were reported without a specific type, they managed to sort out and classify nearly all these events in the end (!)

Q4.3

This question is very similar to the previous one and admits a similar visualization. For example, we can construct a stacked bar chart, as follows. First, we drag Event Clearance Group and Number of records to the Columns and Rows fields of a new sheet, respectively. This creates a bar chart showing the break down of events per Event Clearance Group value. Next, we drag Event Clearance Subgroup to the Color field of the Marks panel. This splits the bars into ranges corresponding to the different types of event subgroups. Finally, we can sort the bar chart e.g. decreasingly on the number of records per Event Clearance Group value.

The resulting visualization should resemble the image below:



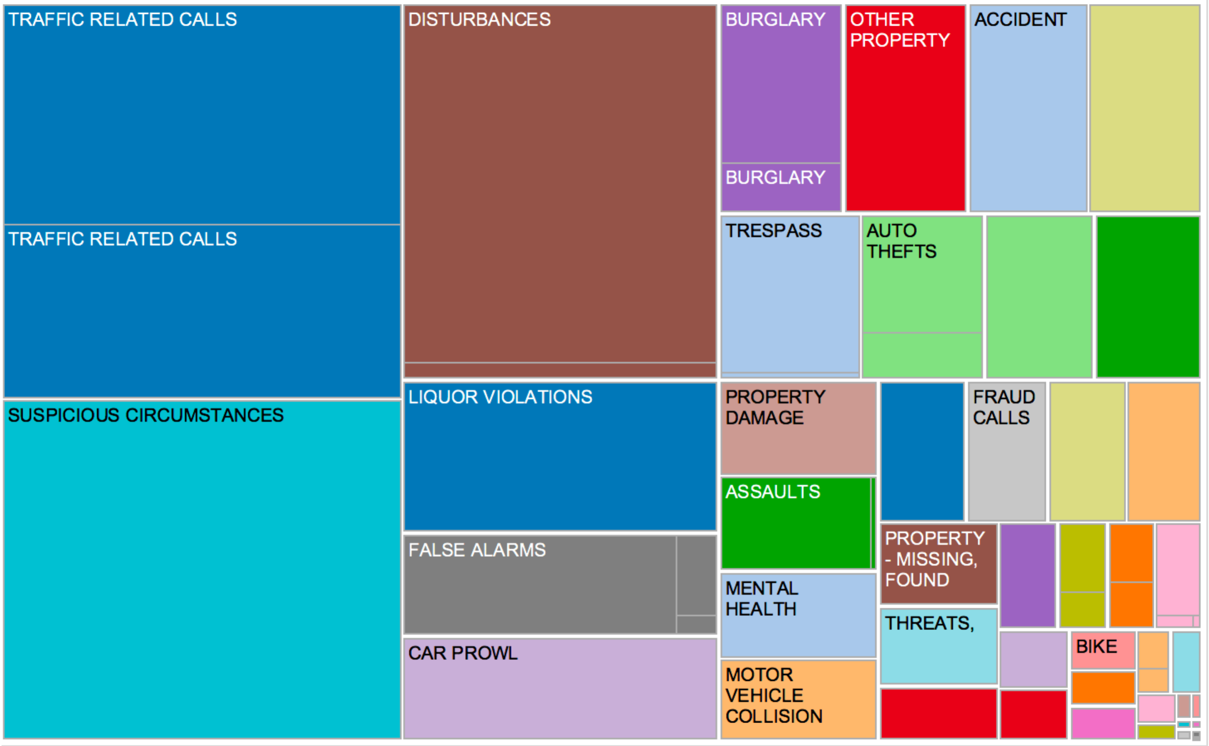
It is interesting to see how Null values correlate here. We see that there are Null values for both Event Clearance Group (this is the small bar labeled 'Null' roughly in the middle of the bar chart), and also for the Event Clearance Subgroup (this is the dark-blue value labeled 'Null' in the color legend). If we select Event Clearance Subgroup in the color legend, we see that only the above-mentioned small bar gets highlighted in the visualization. Thus, Event Clearance Subgroup takes Null values only for Null values of Event Clearance Group. Conversely: If we brush the bar marked 'Null' in the bar chart, or zoom in to better see its extent, we see it is entirely dark blue. This means that all records having a Null value for Event Clearance Group also have a Null value for Event Clearance Subgroup. In other words, we just found out that the two variables are Null *precisely* in the same time.

We can answer Q4.3 also using a treemap. For this, drag Event Clearance Group to the Columns field and Number of records to the Columns field of a new sheet. You get a bar chart. Next, use the Show Me panel to select the treemap option. The data is now displayed as a treemap, where Event Clearance Group is used to create different cells, and the sizes of the cells are determined by the Number of records (number of incidents) having the same Event Clearance Group value.

To make the treemap more readable, you can drag Event Clearance Group also to the Color field of the Marks panel. Now the type of events in a treemap cell is marked not only by the label of the cell, but also its color.

Further on, you can split the treemap cells based on Event Clearance Subgroup. For this, drag this fields in the Marks panel, *just under* the bottom-most Event Clearance Group field. This creates a two-level treemap: the first level groups incidents on same-values of Event Clearance Group; the second level groups incidents on same-values of Event Clearance Subgroup.

The resulting visualization should resemble the image below:



Task 5: Exploring the incidents' temporal distribution

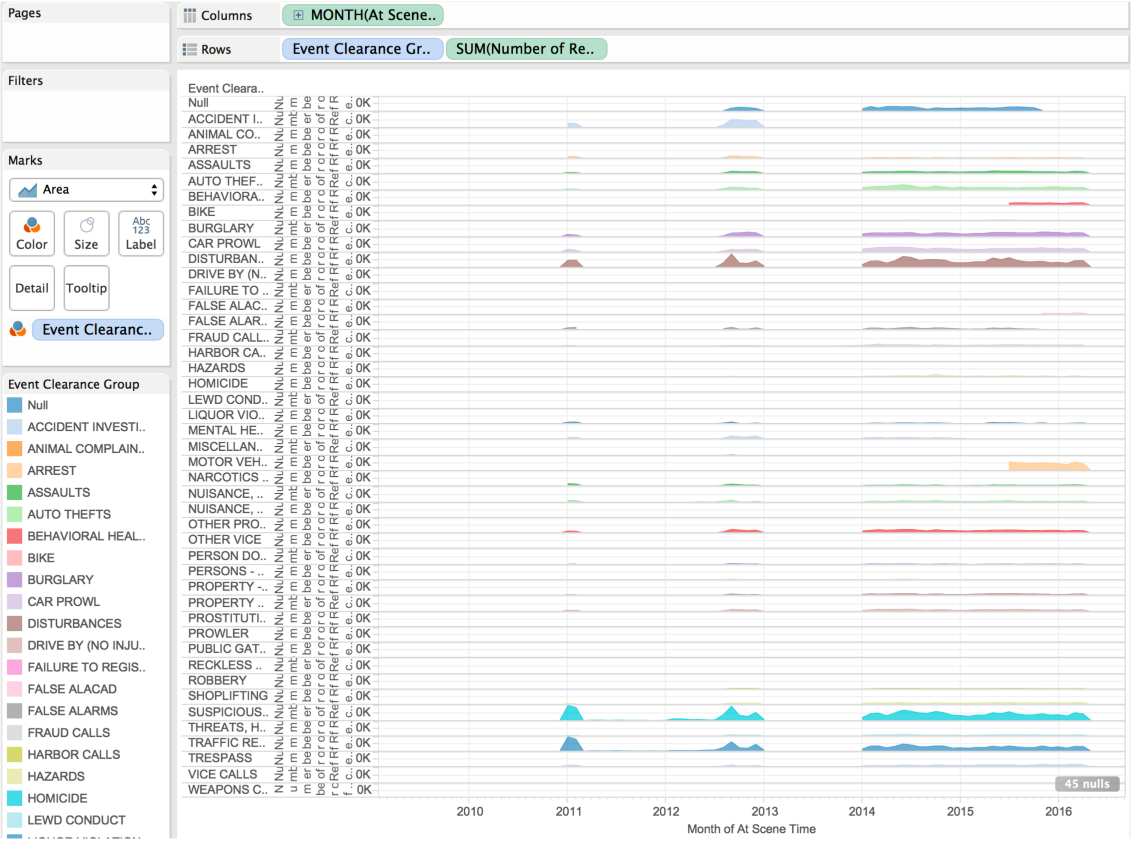
Q5.1

To answer this question, we can use a small-multiples line-chart visualization. The idea is to draw, per incident type (Event Clearance Group value) a line-chart, or timeline, showing the number of incidents of that type over the entire time range. Next, we can stack such timelines for all Event Clearance Group values atop of each other, to compare them.

To obtain this, drag At Scene Time to the Columns field of a new sheet, and select next the *second* 'Month' option in the pull-down menu over At Scene Time. This tells us that the resolution of the x axis is years-and-months. Note that if we selected the *first* 'Month' option in the menu, months of different years would be considered the same (so the resolution of the x axis would be just months). Next, drag Event Clearance Group to the Rows field. This creates a set of stacked charts, one per Event Clearance Group value. Finally, to show the number of incidents in each such chart, drag Number of records to the same Rows field. Note that Number of records gets placed to the *right* of Event Clearance Group. Indeed, data can only be shown by grouping *first* by same-values of Event Clearance Group, and then showing the number of records for each such group. Finally, drag Event Clearance Group to the Color field of the Marks panel. This shows each horizontal time-line with a different color.

Note that you can choose different styles of graphs (line, area, bar, etc) by selecting the desired option in the top pull-down menu of the Marks panel.

The resulting visualization should look similar to the image below:



Columns: MONTH(At Scene..

Rows: Event Clearance Gr.. SUM(Number of Re..

Marks: Area

Color: Event Clearanc..

Interpreting this visualization is now easy: Look horizontally along a timeline to see how the number of incidents of that type changes over the years. Look vertically along a year or month to see which is the most numerous number of incidents of all types. Look over the entire visualization to see if you find similar patterns in terms of variation of number of incidents of different types and/or at different moments in time.

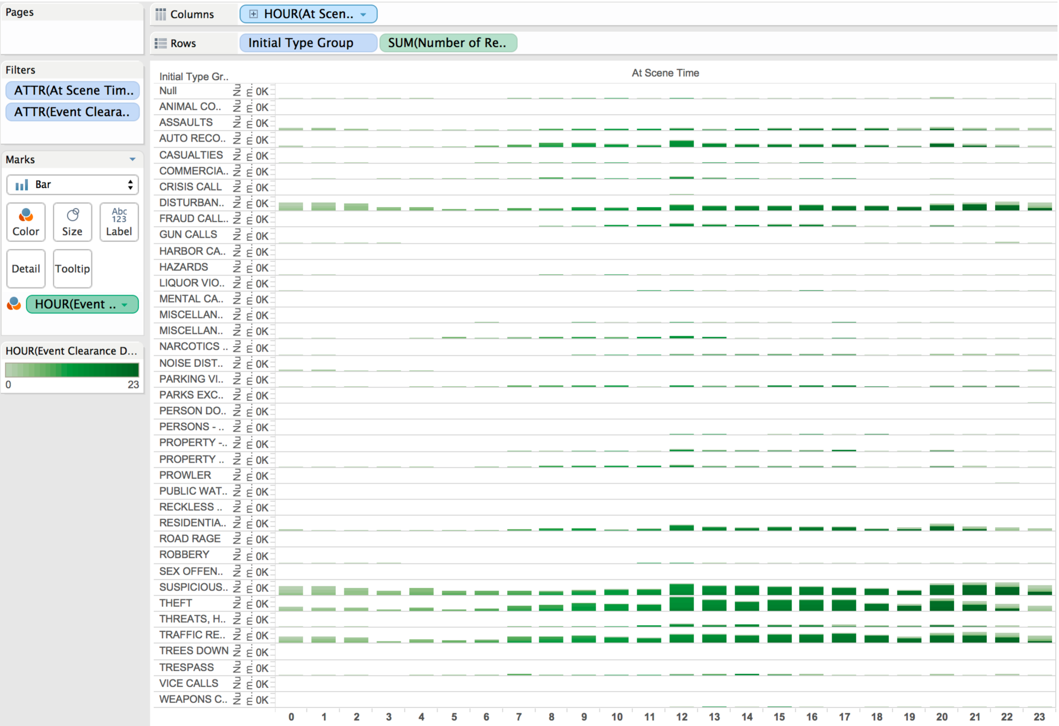
Q5.2, Q5.3

To answer this, we can also use a stacked bar chart, similar (but not identical) to toe one constructed for the previous question. First, drag At Scene Time to the Columns field of a new sheet. Then select the first 'hours' entry in the right-click menu for At Scene Time. This tells that we want to bin the data per hours only, that is disregarding years, months, and days. Next, change the data type of At Scene Time to discrete. This tells that we want to see each range of one hour (during a 24-hour day) separately from all other ranges of one hour.

Next drag Initial Type Group to the Rows field. This creates a set of stacked charts, one for each value of Initial Type Group. Next, drag Number of records to the right of Initial Type Group in the Rows field. This tells that we want to show the number of events for each 'cell' of our chart. Also, select the chart type as 'bar' in the top pull-down menu of the Marks panel.

Finally, drag Event Clearance Date to the Color field of the Marks panel, to color the chart elements by the time when incidents were cleared. Since we want to show hours over a 24-hour day, and we don't care about the year/month/day, select the first 'hours' item in the right-click menu over the Events Clearance Date field in the Marks panel. Finally, to make the color mapping more intuitive, select the type 'continuous' for Events Clearance Date. This will force the using of a continuous colormap to indicate the hours of the day (this is good, since we can determine if an hour value is early or late just by looking at the color).

The resulting visualization should resemble the image below:



Interpreting this image is relatively easy: Different horizontal bar sequences indicate the number of incidents of a certain type over the duration of a day. Vertical columns indicate a certain hour of the day. The colors of bars indicate when the respective incidents (reported at the value shown by the x axis of the visualization) have been solved. Note that a single bar can be split in multiple colors – indeed, an incident of a given type, reported at a given moment, can be solved at multiple times in the future.

Looking at the color gradient from left to right, we see a light-green to dark-green pattern. This tells us that there's a good correlation between when an incident was reported and when it was solved: Incidents reported early in the night (left bars) tend to be solved quite early in the morning; and incidents reported late in the evening tend to be solved later in the night. There's also an interesting pattern in the rightmost bars: We see that parts of them are light green. These indicate that some incidents reported late in the evening have been actually solved early in the morning *of the next day*.

To explore more, you can add the desired fields to the Tooltip slot of the Marks panel, and brush the bars for which you want more information.