Visualization
For Seattle Crime Dataset

Team: XFactor
Behrooz Shahriari
Sreenidhi Pundi Muralidharan
Parineetha Gandhi

What is it for?
Help the Seattle Police Department (SPD) analyze the data.

How does it help?
With such a visualization, it helps the SPD gain a better insight on the data.

And hence...
Helps SPD in maintaining a crime-free city.

Why did we choose this dataset?
This dataset (2010-2015) had more than 1 million records with 38 types of crimes that occurred in 19 districts in 54 different zones.

Purpose:
Seattle Police department authorities would find this visualization project useful to analyze data and plan for faster responses to incidents and hence reduce the crime-rate. This analysis hence provides better INSIGHT into the data. Effective visualization hence catches attention.

Target Users:
This application was developed keeping in mind about the potential users who can take an action to reduce the occurrence of the crimes and those users could be Seattle police department or related authorities. This application helps the authorities to know the crimes that have occurred over the time.

Goals:
The main motive of this application is to help the Seattle Police department make the City of Seattle, a safer place to live in for its public. We have talked about other primary goals in the following sections.
The following are the fields in a particular data set:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Clearance Group</td>
<td>The type of incidents after authorities got to scene.</td>
</tr>
<tr>
<td>Event Clearance SubGroup</td>
<td>The sub-type of incidents after authorities got to scene.</td>
</tr>
<tr>
<td>Event Clearance Description</td>
<td>A complete description of incidents after authorities got to scene.</td>
</tr>
<tr>
<td>Event Clearance Date</td>
<td>Date and time when authorities got to incident’s scene.</td>
</tr>
<tr>
<td>Hundred Block Location</td>
<td>Address of incident.</td>
</tr>
<tr>
<td>District/Sector</td>
<td>Name of the district for incident.</td>
</tr>
<tr>
<td>Zone/Beat</td>
<td>Every district is divided into zones. And this field is the name of the zone.</td>
</tr>
<tr>
<td>Census Tract</td>
<td>NA</td>
</tr>
<tr>
<td>Longitude</td>
<td>Longitude of the location where the incident occurred.</td>
</tr>
<tr>
<td>Latitude</td>
<td>Latitude of the location where the incident occurred.</td>
</tr>
<tr>
<td>Initial Type Description</td>
<td>Description of the event reported to 911.</td>
</tr>
<tr>
<td>Initial Type Subgroup</td>
<td>Description of the subgroup of the event reported to 911.</td>
</tr>
<tr>
<td>Initial Type Group</td>
<td>The initial type of incident that has been reported to 911.</td>
</tr>
<tr>
<td>At Scene Time</td>
<td>The time that authorities got to scene.</td>
</tr>
<tr>
<td>Process Time</td>
<td>Total time from when the incident has been reported to 911 to the time authorities got to scene.</td>
</tr>
<tr>
<td>District Process Time</td>
<td>Average processing time for each district.</td>
</tr>
<tr>
<td>District Zone Process Time</td>
<td>Average processing time for each zone in each districts.</td>
</tr>
<tr>
<td>District Long</td>
<td>Longitude of the district’s center based on events.</td>
</tr>
<tr>
<td>District Lat</td>
<td>Latitude of the district’s center based on events.</td>
</tr>
<tr>
<td>District Zone Long</td>
<td>Longitude of the zone’s center based on events.</td>
</tr>
<tr>
<td>District Zone Lat</td>
<td>Latitude of the zone’s center based on events.</td>
</tr>
</tbody>
</table>

**Dataset Used:**

We used Seattle city’s crime dataset. We decided to analyze the crimes from years 2010-2015 (last 5 years). This dataset is in .csv format and has been acquired from:

Goal 1: Visualizing the data

Description: Here, the SPD authorities can use the visualized data and hence know where crimes have occurred in the past, as well as currently occurring crimes. The dashboard provides effective analysis of data by showing past as well as real-time feed.

Visualization Pattern Used: Dashboard. This dashboard shows a collection of data, which can be changed according to the filters set.

How did we do it? We used tableau to design a simple dashboard for normal users and a complex one for advance users for complicated analysis.

Tools Used:

<table>
<thead>
<tr>
<th>Highcharts</th>
<th>Tableau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map Reduce</td>
<td>Infogr.am</td>
</tr>
</tbody>
</table>

Thus, data is presented visually to the user rather than raw data.

Primary Goals:
1. The police department will use visualized data and hence analyze the crimes. This will pave way for a crime-free city.
2. Comparative study of crimes from 2010 to 2015; comparative study between each crime type.
3. Hourly and monthly based visualization of raw data.
4. Comparison between processing times for each type of crimes.
5. Comparison between average processing times and number of records.
Goal 2: Comparative study of crimes from 2010 to 2015; comparative study between each crime type.

Description: Comparative study of crimes happening in Seattle with the comparison being done over the years, from 2010 to 2015. Using this comparison, it is easier to check if these crimes have been thwarted; or if they still persist, how quickly they can be attended to.

Visualization Pattern Used: Bar Graphs.

How did we do it? We used bar graphs because they are very good for comparisons between each crime type. Using such bar graphs for sorted data, it is easier to find which crime occurred the maximum number of times and which ones occurred the least.
For example, here, suspicious circumstances and traffic related crimes are high for all years; Narcotics crimes have not been occurring in April and August, in any year.

Goal 3: Hourly and monthly based visualization of raw data.

Description: Data that is represented for each hour and each month, for all five years (as shown above) is useful in comparing which months/times are better than the rest.

Visualization Pattern Used: Line and Bar Graphs with Geo Maps.

How did we do it? We chose bar graphs because they are good for comparisons between the crime types for each hour/month. Apart from that, we also chose to show line graphs as they make peak/critical points clearly stand out in any unsorted data.
Goal 4: Comparison between processing times for each type of crimes.

Description: Processing time is the time between when a crime is actually reported to when the SPD actually gets to the scene of incident. Some crimes like homicides may have faster processing times while others like false burglary alarms may have slower attention rate.

Visual Design Pattern: Pyramid Chart

How did we do it? This kind of chart shows the comparison between processing times for each crime type. Amongst the ones with higher attention rate, homicides, narcotics and harbor vessel recovery top the charts.

Be careful when going out at 12 noon and at 8 PM.

Safest to go out between 3 to 5 AM!
”Crime statistics represent only police services where a report was made and do not include any other type of calls.”

Areas with a high volume of foot traffic or that are more densely populated may have more reported crime. This does not necessarily mean more crime occurs there, but that more crime is reported there.

Data will sometimes reflect where the crime was reported versus where the crime actually occurred.

The crime index of Seattle is 4. This means that it is safer than 4% of the cities in the United States.

Goal 5: Comparison between average processing times and number of records.

Description: With this kind of visualization, it is easier to compare two different values. This may assist the SPD in taking quick action against the crimes that need much attention.

Visualization Pattern Used: Overlaid line graph on bar graph

How did we do it? We used an overlaid line graph, on top of a bar graph to show the differences between the processing times and number of records. Traffic related crimes were the highest in number while false alarms received the least processing time.
Creating and working with the Visualizations and Data:

We used MapReduce to query and process the data. The data set did not contain all the necessary information beforehand; we used certain tools to compute the information we needed. For example, we calculated the processing time for each crime from the “event clearance time” and “at scene time”.

The following visualizations depict the purpose they satisfy. Graphs were plotted for visualizing data analyzed as well as computed.

1. Major crimes in Seattle in the past 5 years: Traffic related crimes, Suspicious Circumstances and other Disturbances have been topping the charts from 2010-2015. Even in the current year, these kinds of crimes have not reduced and the SPD apparently seems to have done nothing about it.
The above picture is a WordCloud and depicts the major crimes that have been going on in Seattle.

2. Compare number of incidences; Real-time data with history of over 6 months for better comparison: As we have discussed above, we used bar charts for effective comparison through visualization. We found out that every month, traffic-related crimes and suspicious circumstances have taken the toll, while vice crimes and homicides occur rarely.

3. Data tips on maps: The data tips can be revealed when clicking on markers. They give details about the place, its location, crime rate, processing times for each crime, etc.
4. **Data spotlight**: Clicking on an entity highlights the entity and dims the remaining. We used this so that it catches the immediate attention of the police personnel using this application.

5. **Dynamic Querying**: The dashboard’s dynamic querying can be used to set the filters dynamically and the visualization changes accordingly.

6. **Local zooming**: Zooming on particular areas helps the user (SPD) to look into the data more deeply and hence find out more information about it.
7. Sortable Data: Raw data can be either sorted or unsorted. While unsorted data representation as bar graphs is difficult to comprehend, the data can hence be sorted and represented as bar graphs. This can also be used to represent highest peak/critical points.

8. Multi-Y graphs: Multi-Y graphs are used when the number of layers equals to the columns plotted. Instead of plotting in different columns in the y-axis, using such a multi-Y graph makes it easier for comparison amongst different components that are plotted.

9. Geo Maps and Heat maps: Geographical maps can be used to show the exact map of the places with markers on top of districts/places, which says the rate of crime (red marker – highest crime rate area; cyan marker- lowest crime rate area). Heat maps on the other hand can be visually used to represent the density of crime in specific area; higher crime areas have denser colors.
To make this application look like a real time application, we designed a homepage with two links. One link directs to the Dashboard where a user can see the visualization of the overview of the incidents. The second link "Maps/Dashboard" in the homepage directs to the page where a user can view the visualization related to maps exclusively.

1. **Overview Data Page:** This page shows the visualizations related to the entire data. The left panel has all the filters where user can filter out and view the visualizations accordingly. This page consists of visualizations according to the month, according to the hour, according to the type of incident. There is also live streaming visualization that can show the real-time data feed.

2. **Maps Visualization Page:** The motive behind keeping the maps in separate page is to keep the application simple and easy to use by the user. This page shows up two different maps each of which serves different purpose. The map on the left shows all the districts and the heat map of number of incidents in Seattle; color of each marker and its region demonstrates the ratio of incidents in the related regions to the total number of incidents in Seattle.

   The map on the right shows all the zones in districts and heat map of process times for incidents in Seattle. Color of each marker demonstrates the ratio of process time for incidents, in the related zone, to the overall process time in Seattle.

   There are three sets of button groups, one button group is to view the visualization for a particular year and the other two button groups are to customize the heat map.

   (a) Toggle heat map button is used to toggle between a heat map or a geo map.

   (b) Change gradient button is used to change the color gradient button in the heat map to cyan, blue and red.
(c) Change radius button is used to change the radius of the focus of the point on the heat map.

(d) Change opacity button is used to increase or decrease the opaqueness of the heat map.

Districts map (left): On hovering over the map user can identify different districts of Seattle. On clicking on a district the data related to that district is shown in the tooltip. This tooltip has a link to view the visualizations related to that particular district. This link directs to the dashboard of that particular district in a new tab, which has visualization for monthly report, hourly report, average process time report, number of incidents report.

Most importantly, it has visualization for live streaming of the data according to the process time of history of incidents and number of incidents. This streaming is adjustable according to weeks, months and also can be viewed for the overall data with a tailored range.

**UI Design Patterns followed:**

1. **Organizational Design Patterns:**

(a) **Dashboard:** As shown and explained previously.

(b) **Settings editor:** Has various options to customize the map as explained before.

(c) **Multiply Workspace:** Two different workspaces for maps- efficient for comparisons between years, areas, crimes, processing times, etc.

2. **Layout Design Pattern:**

(a) **Visual Layout:** A common framework for all application pages but which allows flexibility to handle varying page content. We used it in maps.

(b) **Tiled Section:** Each graph in the dashboard is represented as a tile.

(c) **Collapsible Panel:** (In maps) hide one to focus on the other.

(d) **Self-adjusting layout:** The entire application is responsive to changes in size.
3. Navigational Design Patterns:

(a) Hub and Spoke: Transition from all other pages to home page.

(b) Sequential: Transition from complex dashboard to map.

(c) Pyramid: Used for changing years in map page.

4. Action Design Patterns:

(a) Button Groups: Buttons that perform the similar kind of action are grouped together. All these buttons are used to change the appearance of the heat map and so are grouped together.

(b) Hover tools: Hover tools are used to highlight the information related to the area where the mouse hovers. We have used this tool in various locations. In the following screenshot the district gets highlighted when the mouse hovers on that location. This is helpful in identifying the boundaries of that district.

5. Choice Control Design Patterns:

(a) Slider: As the check boxes consume lot of space and it takes extra effort from the user to check and uncheck the boxes to select a month we thought slider is the better option.

(b) Dropdown List: Used for listing down all crimes.

(c) Checkbox List: As there are multiple number of crimes and to visualize the data of one or more number of crimes the user can check or uncheck the crime to view the respective visualization.
Screenshots:

Home page of Dashboard
Overview of all incidents
Real time feed of data
Maps Dashboard
How to run this application?

1. Download the project’s zip file in a location.
2. Unzip the file’s contents into a location that is convenient.

3. Get into the Seattle Visualization Site folder. Double click on home.html. This is the main page that leads you to the home page.
• Please find in the attached zip folder, along with this report:

  1. Our presentation slides
  2. The source code needed for running the visualization tool.

• The demo of our project is put up as a video on YouTube. The link is at: https://youtu.be/EOFQfd3F6Wc