The Crime Landscape of Salt Lake City

# Basic Information

Members:

* Archit Rathore – u1144416 (archit@cs.utah.edu)
* Yash Gangrade – u1143811 (yashgangrade09@gmail.com)
* Rebeka Mukherjee – u1141112 (rebeka.mukherjee@utah.edu)

Github Repository: <https://github.com/yashgangrade09/dataviscourse-pr-crime-landscape.git>

# Motivation

The general consensus between us team members was to work with data that pertained to some societal factors. Another thing that we wanted out of our project was to be able to derive insights on a much finer geospatial resolution (state/city). Finally, we also wanted our visualization to be relatable and not dealing with data of a technical nature. This led us to explore the Utah Open Data catalog (<https://opendata.utah.gov/>) and we finally zeroed in on the police cases dataset that lists all reported crimes in the Salt Lake county.

This dataset checks off all our boxes, namely:

* Pertains to societal factors
* Is localized
* Is relatable and may be of equal interest to both a layman and a specialist
* Has the potential to inform policy decisions

# Data

We are using the [Police cases dataset](https://opendata.utah.gov/Public-Safety/SALT-LAKE-CITY-POLICE-CASES-2016/trgz-4r9d) from 2008-2016 (tentatively) from Utah Open Data catalog. Each record in the data corresponds to an instance of reported crime at one of the police stations in Salt Lake City. It contains information about the type, time and date of occurrence and reporting, and location of the crime. There are about 50K+ records per year. The data is open access.

The geospatial data will be pulled from either Google Maps or OpenStreetMaps (more inclined to use OpenStreetMaps). Showing crime statistics in context of local information such as businesses presents would be a ‘nice-to-have' addition.

# Data Processing

The Open Data catalog provides an export API to easily download data in a variety of formats. We will be exporting the data as CSV through the web pages for each year. The data is mostly clean with very few missing values (<0.1%) for which we intend to drop the row itself.

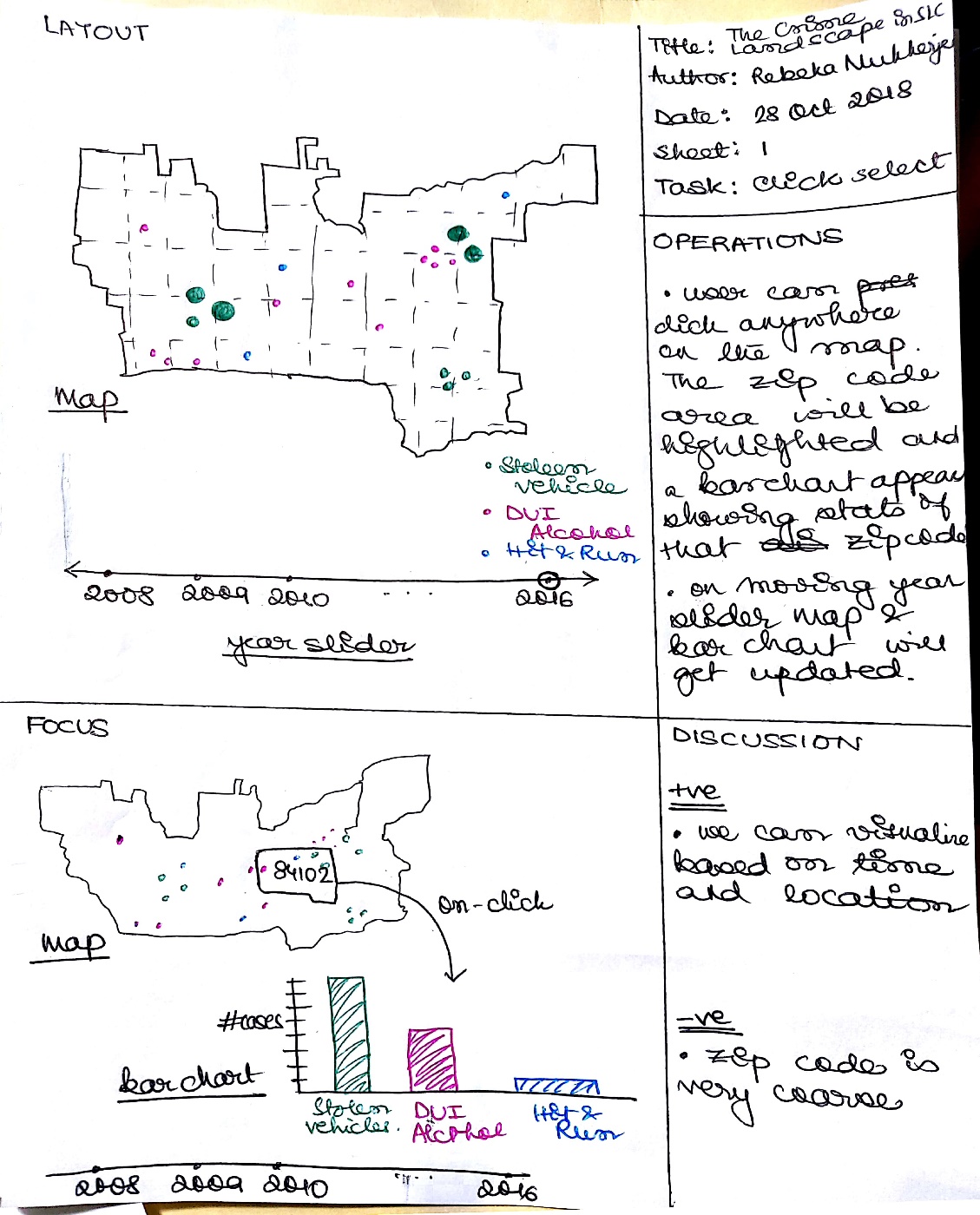
We will perform data aggregation based on time of occurrence, location, type of crime etc. This will be done using python/Excel and stored as separate files that are loaded on demand (instead of aggregation in browser).

The maps data should not require much processing as it will be fetched as is from the API.

# Visualization Design

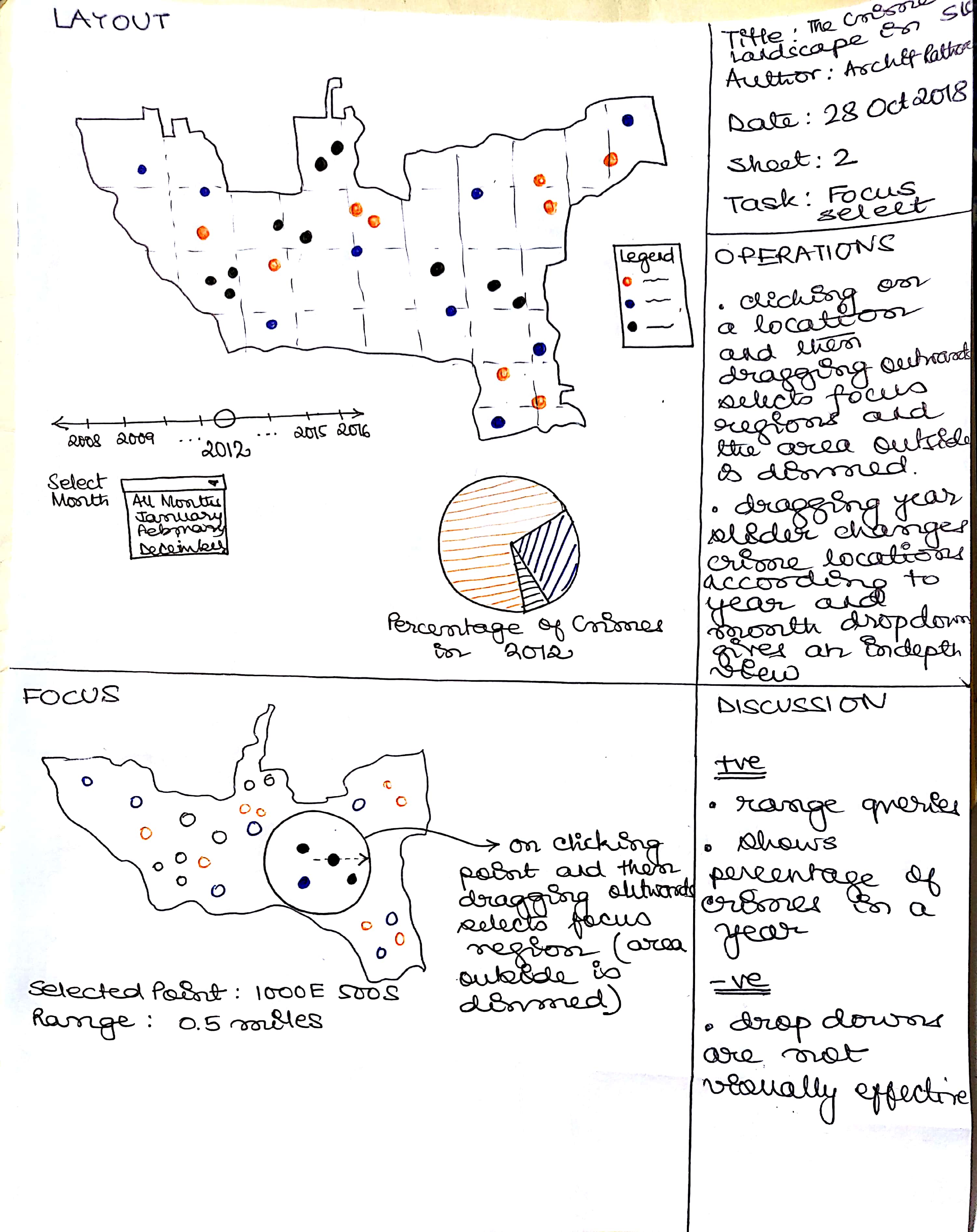
**Prototype 1**

In this prototype design, we are using a map view to show a map of Salt Lake county. This map will have a year slider and a month slider corresponding to the selected year. It will have certain points highlighted with different colors. These points represent different types of crimes happened at different locations in the county. A user can select a zip code and that entire area will be highlighted and a bar chart of different crimes will be created for that zip area. For every year and month selected, there will be a single bar chart.



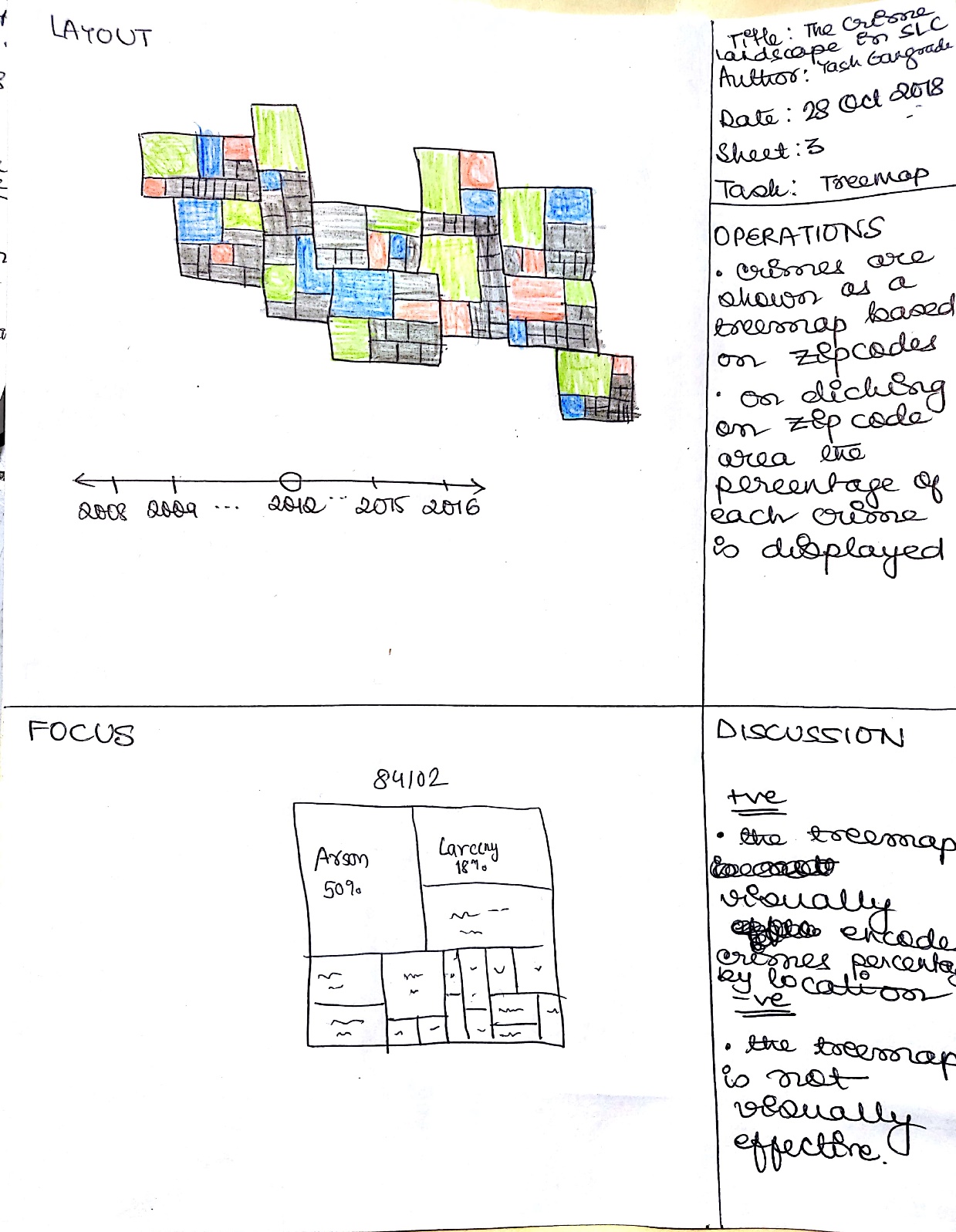
**Prototype 2**

We have a similar map view as prototype 1. A pie chart is added to represent distribution of crimes in the county for a selected year and month. However, here instead of clicking the user can click on a point and drag outwards to trace a circle which will then populate the information about that region below the map.

****

**Prototype 3**

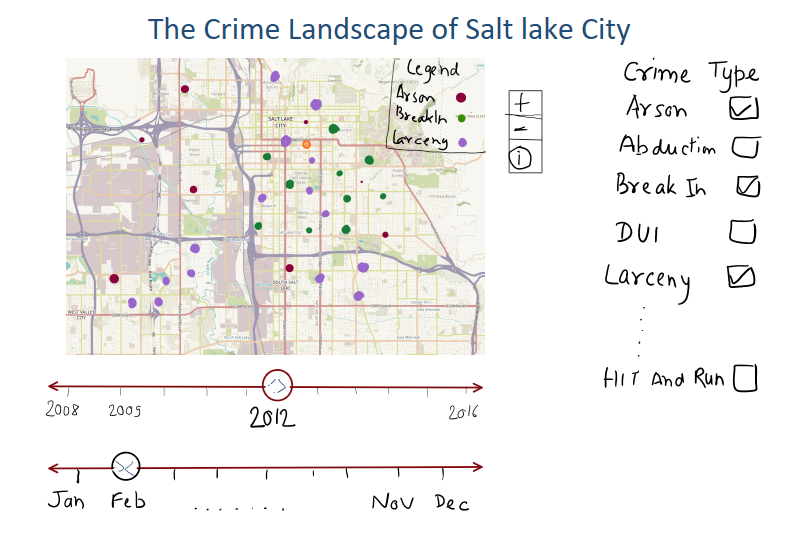
We replaced the geological map with a tile map where each tile corresponds to one zip code. Each tile houses a treemap visualization of crimes for that zip. Clicking on any of the tiles expands the treemap in the tile and shows a detailed view of the distribution of crime.

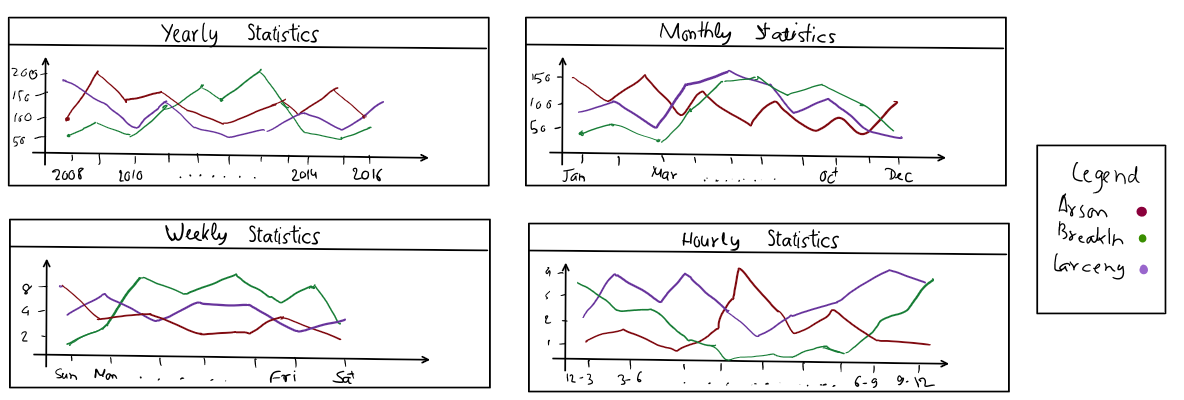
****

**Final Design**

In the final design, we intend to show a map view of Salt Lake County, and a statistics view to show the yearly, monthly, weekly and hourly statistics.

* The map view shows the location points of crimes selected in selection view and time selected by the time slider. The mark used is points and channels are color (type of crime) and position (location of crime).
* A time slider enables the user to see the trend of crimes over the years. On clicking the year slider it expands to show the months for the selected year. This enables the user to visualize the crime trends on a finer scale. It collapses to show only the years when clicked again.
* A selection view enables users to select the crimes they wish to visualize through checkboxes.
* The map view has a semantic zooming feature with zip code level aggregation at the topmost level and blocks at the lowest level.
* The statistics view shows yearly, monthly, weekly and hourly statistics of aggregated data about selected crimes as line charts.





# Must-Have Features

* Functional map of Salt Lake City with an interactive overlay of the crime data.
* Interacting with individual data points shows details through tooltip.
* Show data over a span of years (through a year slider).
* “Semantic zooming” - when zoomed out the interface shows aggregated statistics, zooming in progressively refines the scale of aggregation until only individual datapoints remain. This would allow user to control the level of aggregation.

# Optional Features

* Incorporate other data from the Map API such as number of businesses in the region, distance from the nearest police station etc.
* Toggle between detailed and summary view.
* Radial range selector.

# Project Schedule

**Week 1: October 29 – November 5**

* Oct 30 – Peer Feedback
* Work to do: Pre-process Datasets

**Week 2: November 6 – November 13**

* Nov 9 – Project Milestone due
* Work to do: Create and populate the map view

**Week 3: November 14 – November 21**

* Meeting with Mentor to finalize things
* Work to do: Implement Statistics view

**Week 4: November 22 – November 30**

* Nov 30 - Final Report due
* Work to do: Wrapping up things, attempt optional features, creating screencast.

------------------------------------------<Project Proposal Ends>------------------------------------------

------------------------<Project Milestone Starts on Next Page>------------------------------------------

# Motivation

The general consensus between us team members was to work with data that pertained to some societal factors. Another thing that we wanted out of our project was to be able to derive insights on a much finer geospatial resolution (state/city). Finally, we also wanted our visualization to be relatable and not dealing with data of a technical nature. This led us to explore the Utah Open Data catalog (<https://opendata.utah.gov/>) and we finally zeroed in on the police cases dataset that lists all reported crimes in the Salt Lake county.

This dataset checks off all our boxes, namely:

* Pertains to societal factors
* Is localized
* Is relatable and may be of equal interest to both a layman and a specialist
* Has the potential to inform policy decisions

# Questions

There are many questions that we want to answer through our visualization. Many questions arise when we were creating the visualizations and processing the data. Some of them are listed below;

* What’s the general crime landscape in Salt Lake City and how it has changed over the years? Essentially, an overview about the living in Salt Lake City.
* Are there any crimes which have been constantly increasing over time?
* If a person wants to buy a house, he/she can query what kind of crimes are prevalent near the location.
* Is there any correlation between the crime types and the businesses present in the area?
* How’s one neighborhood different from other neighborhoods in terms of crime types?
* What hour of the day/night is the most dangerous in terms of crimes in a neighborhood? Ex. what’s the number of stealing incidents on Friday or Saturday night etc.

# Data Collection and Processing

As we discussed before, we are using the [Police cases dataset](https://opendata.utah.gov/Public-Safety/SALT-LAKE-CITY-POLICE-CASES-2016/trgz-4r9d) from 2008-2016 from Utah Open Data catalog. Each record in the data corresponds to an instance of reported crime at one of the police stations in Salt Lake City. It contains information about the type, time and date of occurrence and reporting, and location of the crime. There are about 50K+ records per year. The data is open access.

For data clean up, we drop the rows where any piece of vital information like X-coordinate, Y-coordinate, Location, Time etc. is not available. We also need the Latitude and Longitude information of the location of crime using the given coordinates information. In the data, we are given X and Y coordinates which needs to be converted to Latitude and Longitude. As soon as we get the conversion formula, we will process our actual data and put it in place with the views.

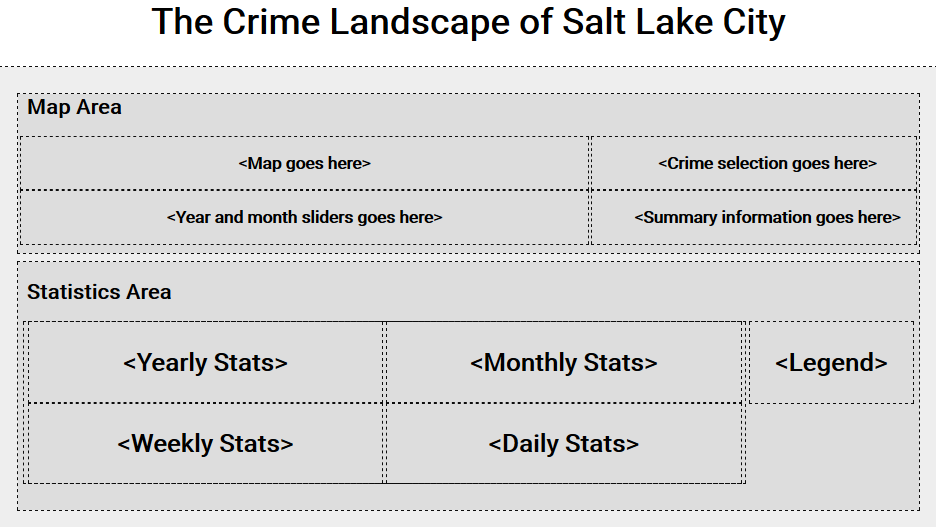
For now, to create and work with different views in visualization, we are using small subset of dataset for three years with a few data points in each file. We took ten data points for each year from 2008 – 2010 and then we got the latitude and longitude manually from the google maps by entering the street address.

For geospatial data, we tried a few different ways to fetch and display the data namely Google Maps, OpenStreetMap, OpenLayers etc. and we are finally using Leaflet library for displaying the map. This is also a mobile friendly and responsive design of the map, so it will be beneficial and easy to use for everyone.

# Implementation

## Page Structure

We first created the basic structure of the directory and created the views structure of the visualization. First, we created a container which holds all the views and structure in several nested divs. Each div represents a view like map-view, summary view etc. and we are assigning classes to each of the divs. This assists us in arranging the divs according to the structure required. Also, we can work on a view separately without disturbing the entire structure. Below is the initial starting structure of the visualization.



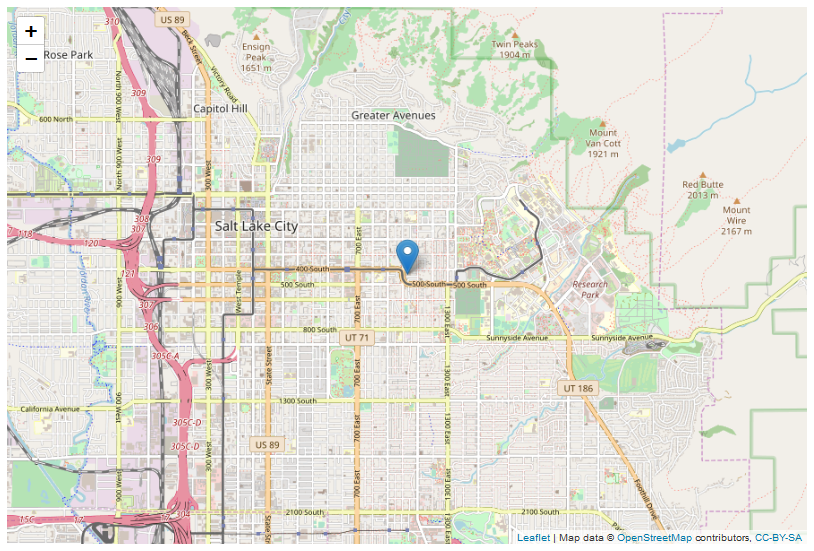
## Year Slider

The visualization needed a year slider to update the map when the year is changed. We directly used the year slider structure from the Gap Minder HW here. Similar slider will be attached for the month data as well. An image of year slider is attached below.

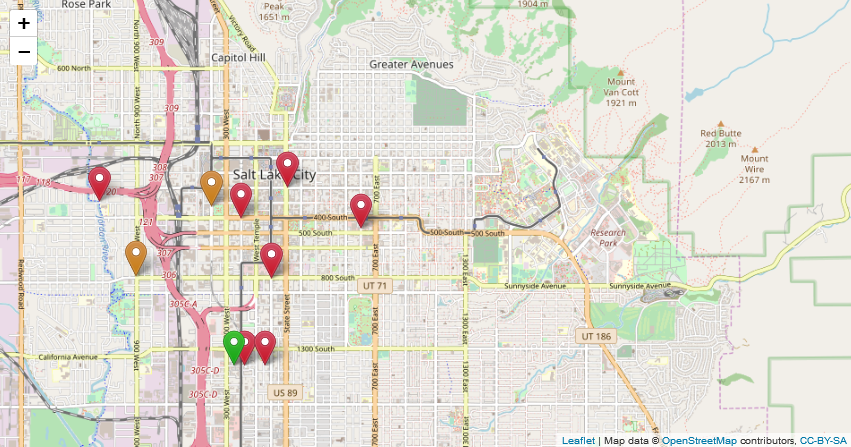


## Map View

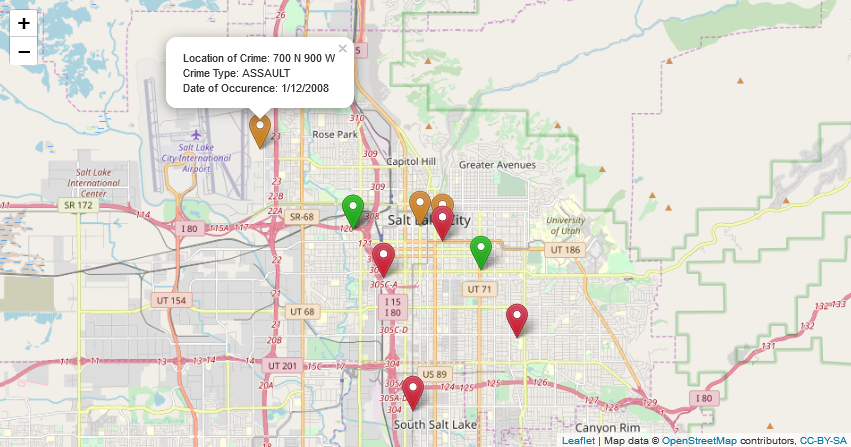
One of the key elements in our visualization is the map view of Salt Lake City county. We want to fetch the data from the server then display it along with different markers for the crimes. For this, we evaluated three options namely, Google Maps API, MapBox, and OpenStreetMap for fetching the map data of Salt Lake City. We finally selected OpenStreetMap to do this because it’s free and it doesn’t restrict the number of requests, modifications to the maps etc. Then, to display the data we again explored multiple options like OpenLayers, GoogleMaps, Leaflet etc. and we finally selected Leaflet to display the map (data from OpenStreetMap) for our visualization. We centered the map at Salt Lake City and kept a marker at our home address for testing. Initial Map is shown below.



Next step was to create markers for different crime types at different locations. For this, we converted the CSV data to a JSON object and then we use this JSON object to put different markers and assign different icons to them according to the crime type. The map and the markers get updated if the year is changed. Clicking on any marker will open a small popup showing information about the location and crime. An example of map with markers is shown below for the years 2008 and 2009.



Crimes in year 2008 (each color represents a different type of crime)

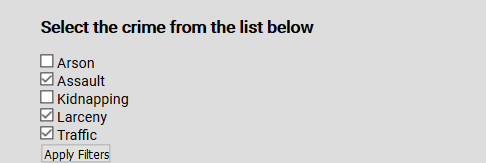


*Crimes in year 2009 (popup shows more information about the crime)*

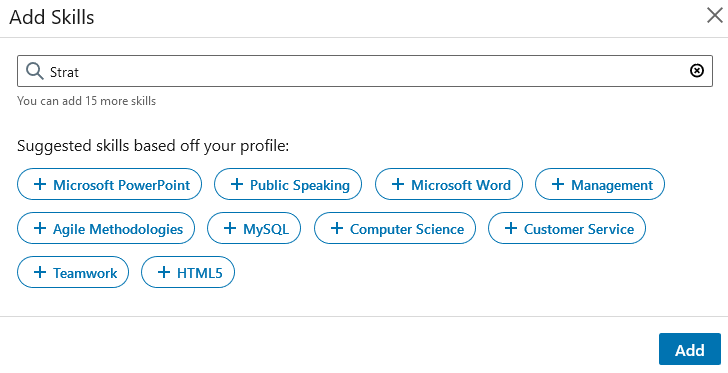
When we were visualizing the crimes using markers for our small subset of dataset, we realized that these markers are not the best way to represent the crimes in Salt Lake because with large dataset, markers will cover most of the map and thus it will become hard to infer use information from the map. We will change the markers to heat maps or we will use some form of clustering/aggregating the markers by crime type etc. We also can try to use hue and saturation to represent different types of crimes in the clustered view.

## Selection View

In the selection view, we currently have a small subset of crimes to select from. They are also shown as checkboxes which will be later changed to multiple select checkboxes in textbox (something like LinkedIn Skills section) to accommodate all the type of crimes. Currently, our selection view is not functional as in we haven’t appended the update function to the “Apply Filters” option.



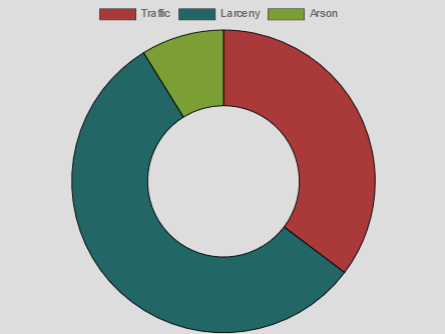
Selection View (will change for the final design)



Example of desired Selection View

## Summary View

We are using Doughnut chart from Chart.js and it’s demonstrating the ratio of incidents of selected crimes against total crimes. We understand that this might not be optimal, and we may have to opt for some other form of visualizations.



## Statistics View

This view uses the data from different years, months, days and even hours. We are creating four different line charts for Yearly, Monthly, Weekly, and Hourly statistics. Here, we will be using different colored lines to show the trend over different temporal quantities. We can compare multiple different crime and understand their correlation with each other and with different times. For example, we can see that traffic accidents increases in the evening time because it’s usually high traffic that time. We will explore more visualizations like bar chart, pie chart etc. and see which visualization conveys the information in the most efficient and informative way.

