POT TRACKER

Process Book: CS -5630 / CS – 6630 Data Visualization

HTTP://WWW.INDIANEXPRESS.COM

CONTENTS

PROJECT INFORMATION

TITLE:

PotTracker

PROJECT DESCRIPTION:

An application that gives ward wise analysis of potholes in Mumbai, India.

PROJECT MEMBERS:

a) Sunny Hardasani - u1000118, sunny.hardasani@gmail.com
b) Mohan Krishna Manchala – u0942454, manchalamohan@gmail.com
c) Yogesh Mishra – u1010866, yogeshmishra.india@gmail.com

REPOSITORY:

https://github.com/PotTracker/web-application

OVERVIEW AND MOTIVATION

Accidents due to potholes are one of the major problem in developing countries like India. Poor construction quality leads to a large number of potholes which cause accidents and traffic issues. According to the Road Accident Report (2014) published by the road transport and highways ministry of India, 6,672 people died in accidents caused due to potholes and speed breakers. In this project, we try to present a visualization of the potholes in a city and help the government in taking corrective measures.

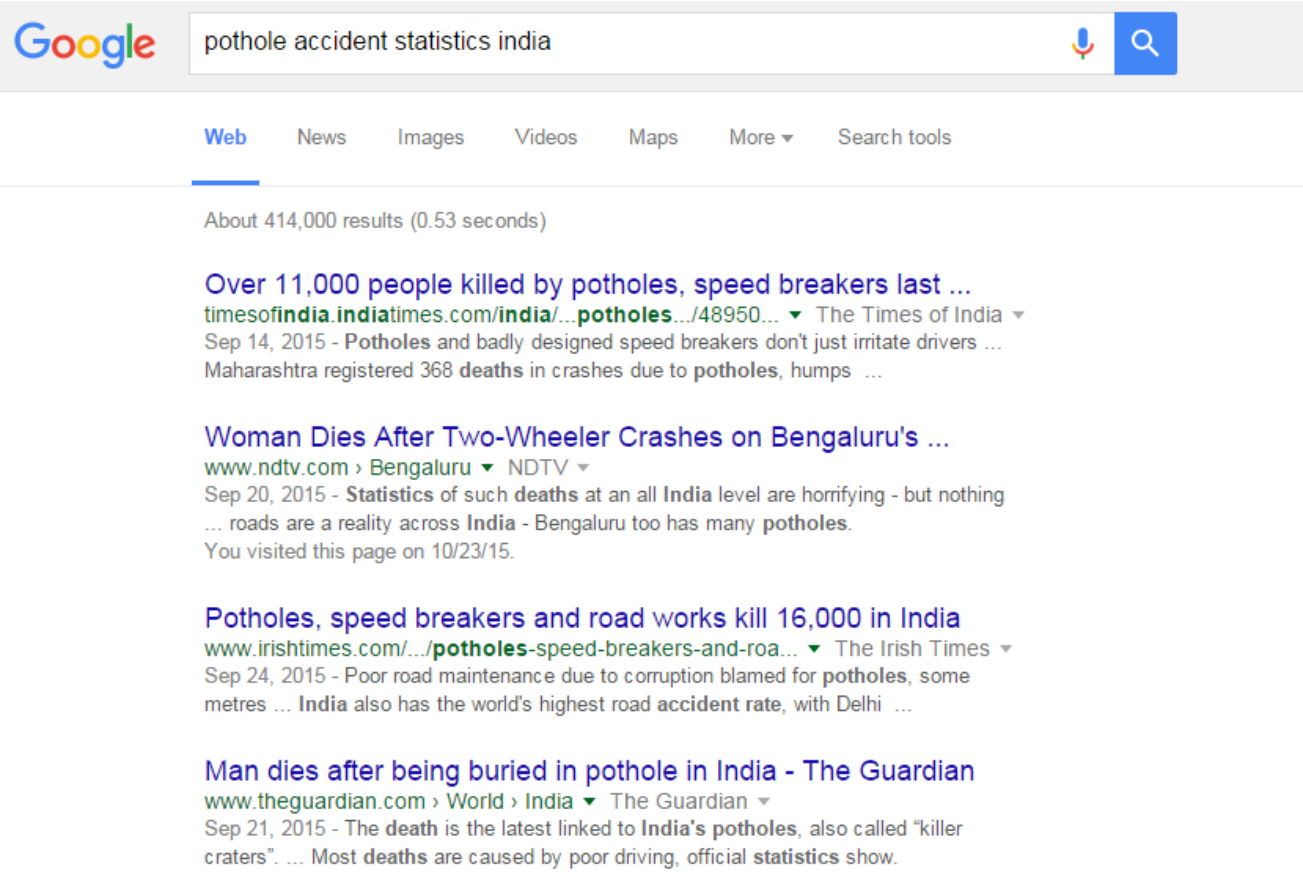


FIGURE -GOOGLE SEARCH SNAPSHOT OF ACCIDENT DUE TO POTHOLES IN INDIA

Basically, our visualization aims to keep track of the area wise data of open potholes, closed potholes and the fresh potholes in Mumbai city. Both the government and people of the city are our audience. Government can use our visualization to effectively track potholes and create a plan to take appropriate measures. People can use it to compare the area-wise increase and decrease in number of potholes and gauge the efficiency of the roads in the respective area.

Our project is a part of a major project that has two components, Mobile application and Visualization application. We are assuming that there is a mobile application which will run as a background service on the mobile phones and tracks the location of the potholes whenever a vehicle running the mobile application crosses it. This application will sense the twitch and send the latitude and longitude of that (pothole) location to the cloud server. Our project focuses on the visualization application. This application will read the data from the server and show the stats of the potholes on page. Visualization will change in real time as the new data gets added on the server by the vehicles.

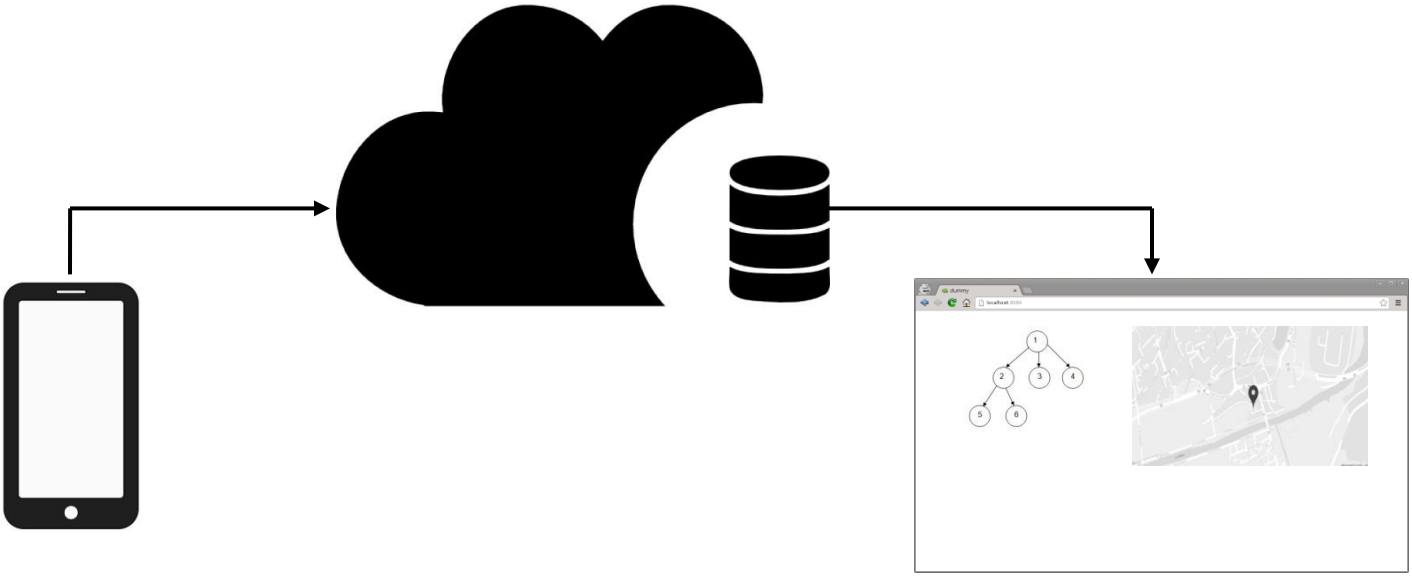


FIGURE- SCHEMATIC OF POTTRACKER

By this application we intent to help the user by focusing on providing the following information.

1. The location of the current potholes in the Mumbai city.
2. The potholes that effect the commuters most.
3. Potholes containing key institutions such as hospitals and schools near them.
4. Area wise information previously opened pothole count with respect to a particular timeline.
5. The most vulnerable roads in an area.

In this project, we convey the above information by means of various visualization techniques learned in the class.

RELATED WORK

In this technology age, it is very important to develop applications that actually solve ground level problems of the society to help the common man rather than creating something fancy. Our group always wanted to do something like this and we got the opportunity in this visualization project.

As mentioned in the motivation section, people in India face a hard time in commuting because of the potholes and poor quality of roads. To tackle this issue with technology, one of our project team members, previously worked on a personal Mobile application that gets pothole location using the GPS whenever there is a bump while commuting. This information is stored in the database and displayed whenever a user opens the application. This kind of application cautions the commuters and helps them in avoiding accidents and other troubles involved.

We were inspired by this application and felt we could take it to a different level if we could incorporate the visualization of the data. We were particularly drawn towards this project because it actually deals with and tries to alleviate the problems of a common commuter

FIGURE - A CAR STUCK INSIDE A POTHOLE, INTERUPTING THE TRAFFIC.

Along with the pothole locations, we also wanted to associate a lot of other key information such as the location of hospitals and schools near the potholes, the most problematic potholes based on the vehicle hit count, area wise comparison of the number of potholes and its variation with respect to a particular timeline. We feel that this information helps the commuters and the government immensely. As we can observe, this is a lot of data and it gives us good scope to use various attractive visualization techniques learned in class.

QUESTION

This visualization project tries to analyze certain statistics of the potholes in a city and help give live details to the public and government. In particular, we would like to answer the following Questions.

1) Which roads in Mumbai have been affected by potholes?
2) What is the frequency of occurrence and closure of potholes on Mumbai roads on yearly and monthly basis?
3) Number of open potholes in a particular month and its variation over a period of time?
4) Comparative analysis of all areas based on the number of potholes?
5) What are the Top potholes that have been causing problems to the public and their number of hits (Vehicle encounters)? These potholes would need immediate attention of the government.
6) Area wise comparative analysis of the potholes based on the hit count?
7) Which area in Mumbai has been most effected by potholes?
8) How the potholes are varying according to the climatic seasons? How are the roads being affected in rainy season, etc?
9) Which potholes have key public institution such as hospitals and schools near them?

Apart from the above questions, over the course of working on this project, we thought we should give some sort of performance assessment of road contractors through our statistics and visualization. This would help the public and government to decide the efficiency of the contractors work. Therefore, if we are successful in answering the above mentioned question we would also provide information about the area wise quality of construction of the roads and the corresponding road contractors. In future, we would also like to forecast the per month increase in the potholes for the next year.

DATA

Two types of data are required to build this application:

ADMINISTRATIVE BOUNDARIES (WARD) SHAPE FILES.

The Shapefile format is a common standard for representing geospatial vector data. It is in fact a grouping of several files formatted to represent different aspects of geodata:

- .shp — shape format; the feature geometry itself.
- .shx — shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly.
- .dbf — attribute format; columnar attributes for each shape, in dBase IV format.

We downloaded the shapefile of India from the following source:

Link: http://www.diva-gis.org/gdata

FIGURE 4: WARD WISE SHAPE FILE VISUALIZATION FOR MUMBAI CITY

Processing: The shape file of Mumbai city was manually extracted from the India shapefile.

We have used this shapefile to display the boundaries of wards of Mumbai city in the Google maps. To overlay the shapefile on google map, we had to convert the shapefile into GeoJSON format. We have used the Qgis software for the conversion.

LOCATION DATA READ FROM THE PRIVATE MOBILE DATABASE

Assuming that we have the mobile app is up and ready, it gives the location of the potholes in the Mumbai city. Our visualization reads in the locations and generates the associated extra data which will be used for the visualization. We have written a script to read in the location of the pothole and output the data required for visualization. Our data primarily has following fields:

- Address
- Latitude, Longitude of the pothole
- Start and End timestamp

- Vehicle hit count.
- Status(active/close)
- Key institutions near the potholes
- Area of pothole

We have created a simulation to automatically generate this data for the past 3 years by adding new potholes and deleting some of the existing ones.

We have used Google maps API to get some of the associated fields of a pothole. The following explains briefly about the APIs used.

1) Geocoding

It is a process of converting geographic coordinates into a human-readable address. The Google Maps Geocoding API's reverse geocoding service also lets you find the address for a given place ID. This API is used to get the address location of the potholes given its latitude and longitude coordinates.

2) Nearby places

A Nearby Search lets you search for places within a specified area by keyword or type. A Nearby Search must always include a location, which can be specified in one of two ways:

- a LatLngBounds.
- a circular area defined as the combination of the location property — specifying the center of the circle as a LatLng object — and a radius, measured in meters.

This API is used to get the key public institutions such as hospitals and schools near to the potholes.

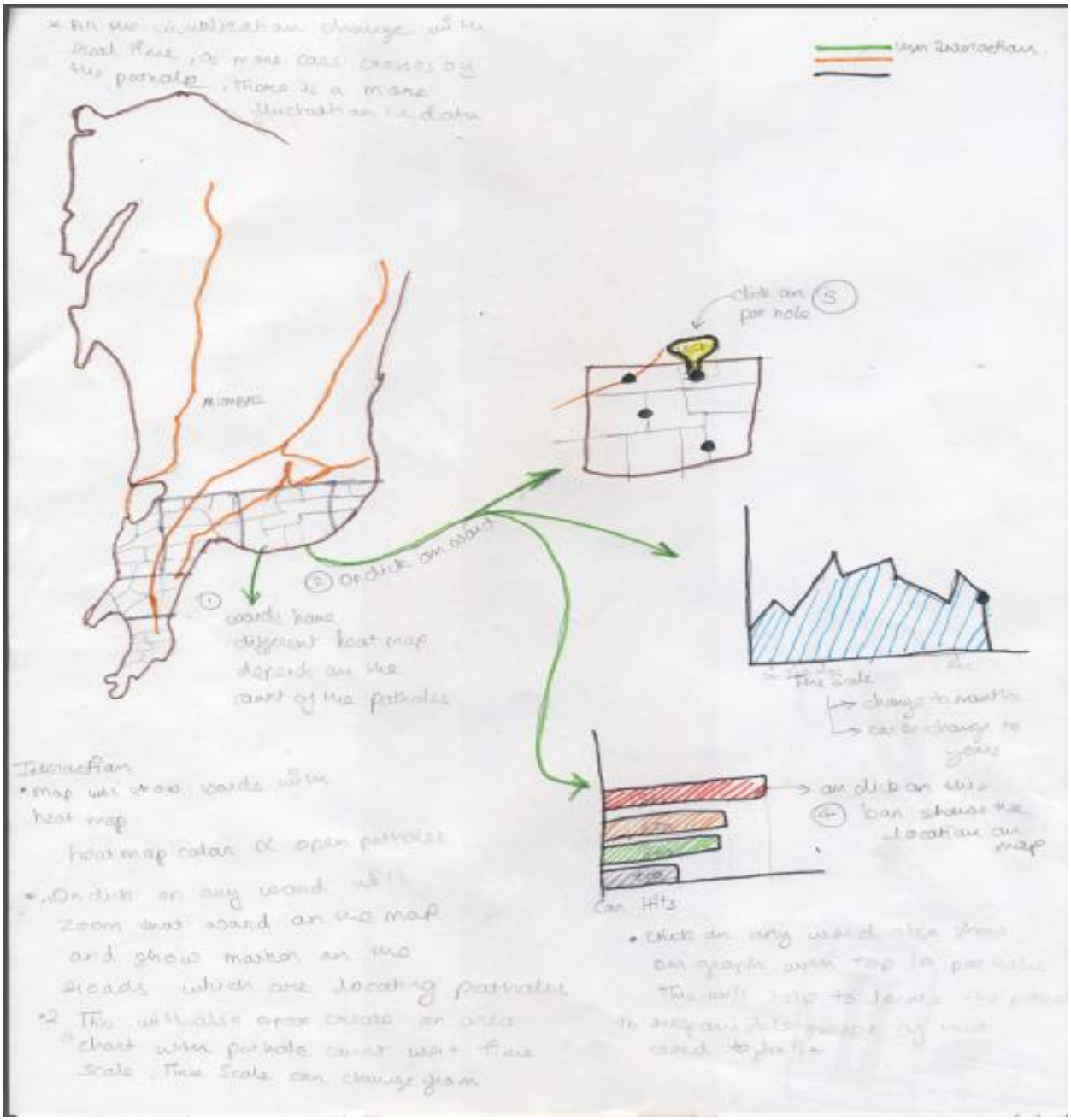
3) Contains location

Each area of the Mumbai city is defined by a polygon with the help of the shapefile. We use the Contains location API to find out the area in which the pothole falls under.

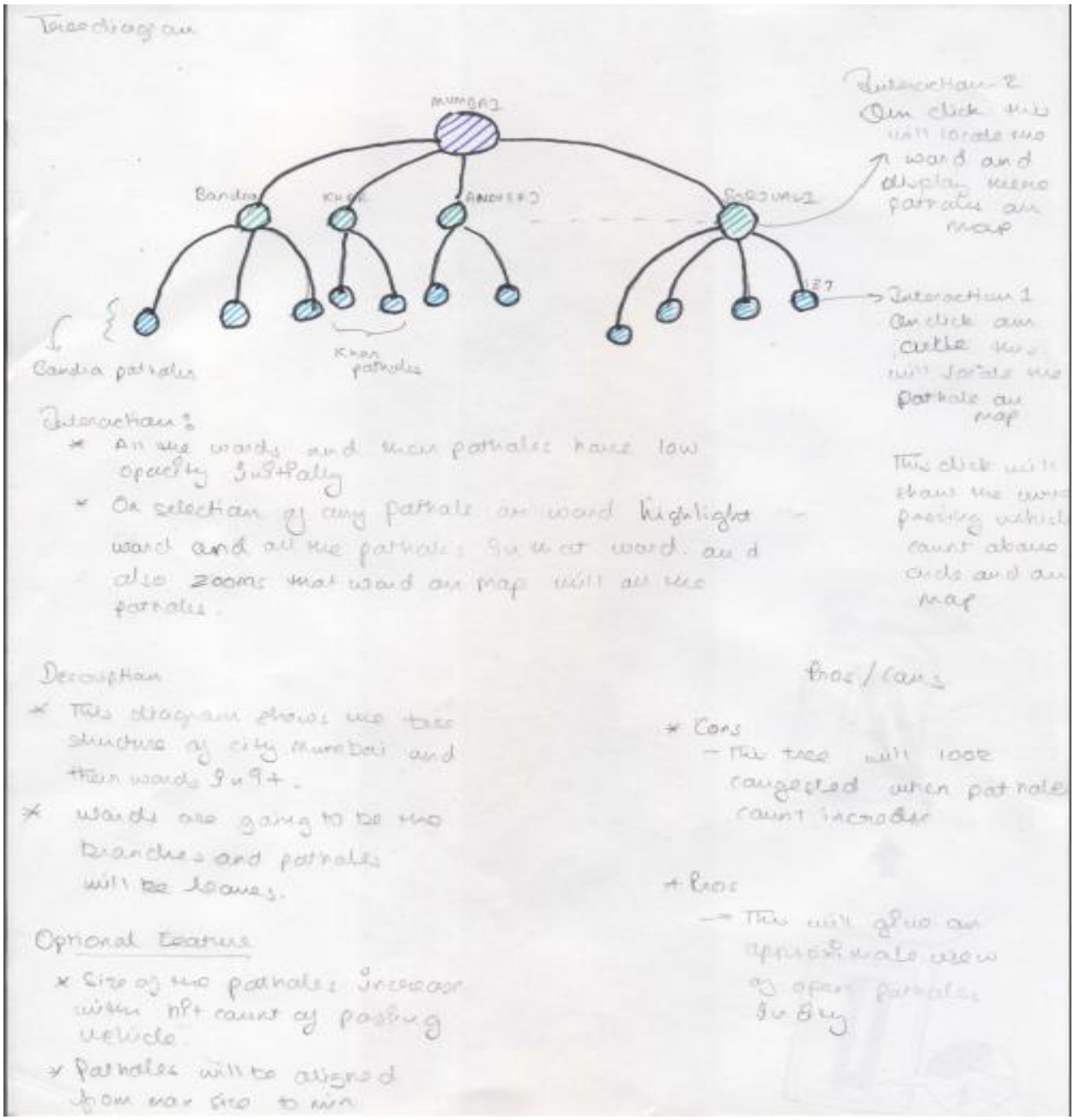
DESIGN EVOLUTION

This section discusses briefly about our initial designs, the problems that popped up because of our initial designs, modifications to the design and the final design.

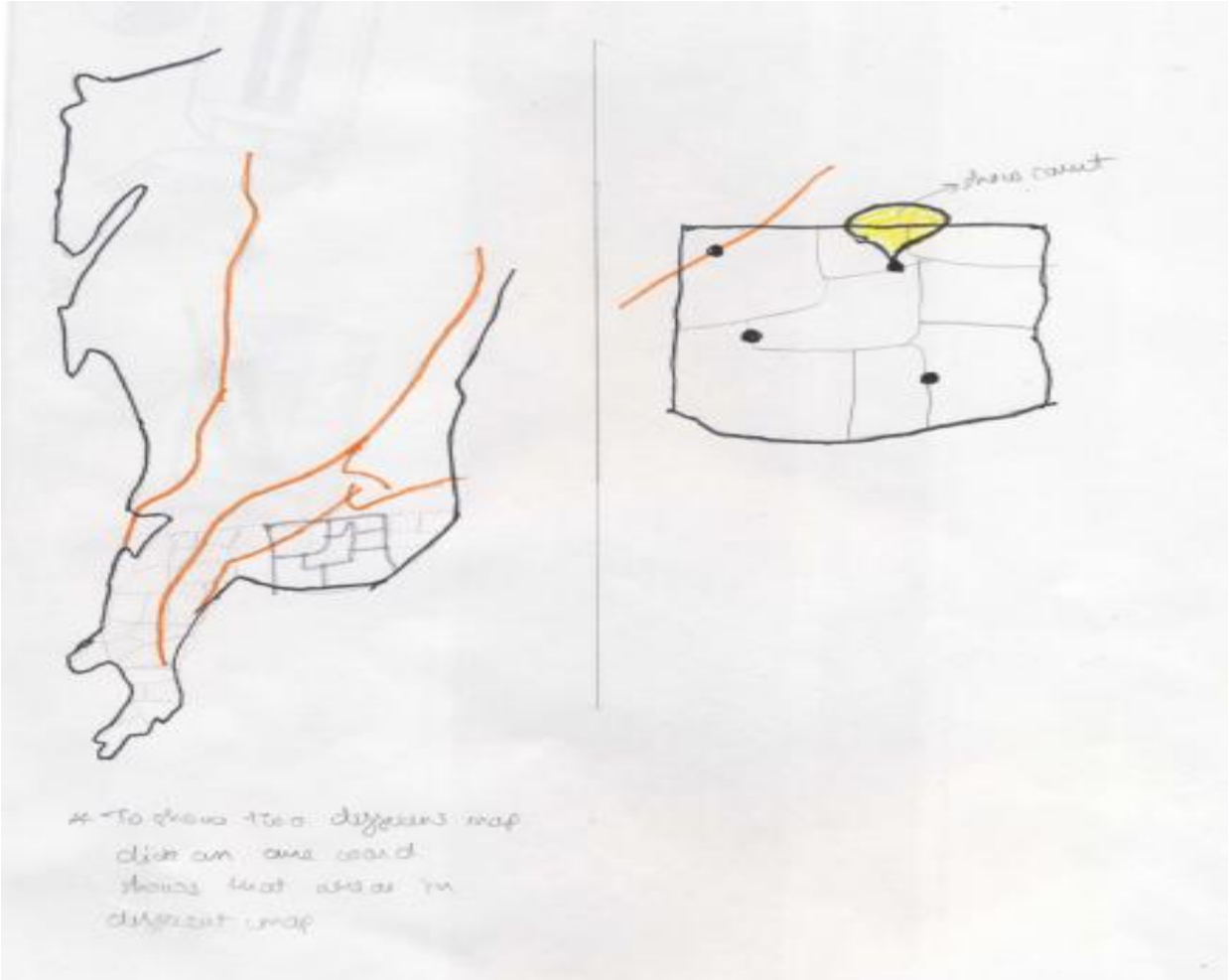
Preliminary Design Abstract 1:



Preliminary Design Abstract 2:



Preliminary Design Abstract 3:



The above figures show our preliminary design abstracts.

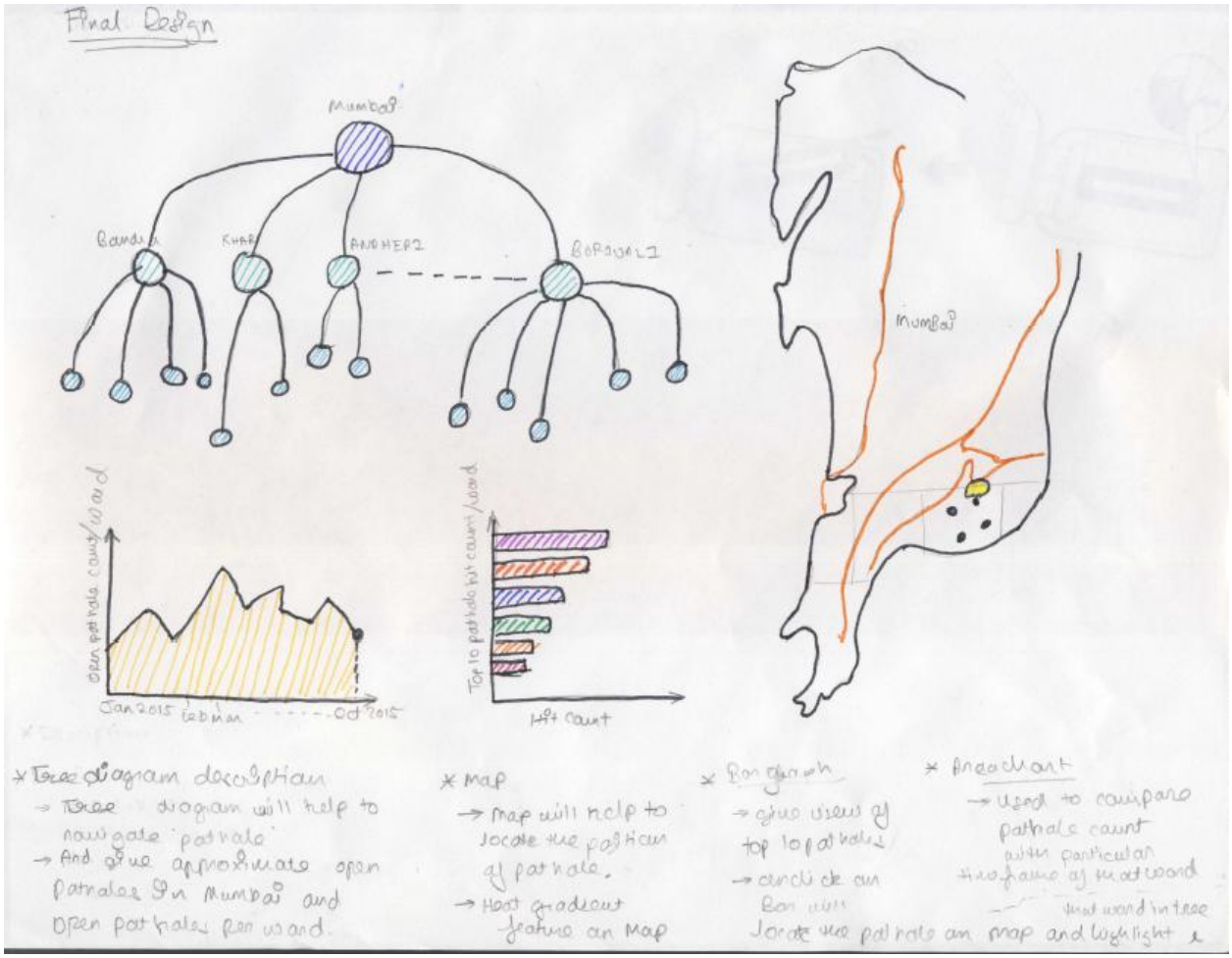
- Abstract 1 consists of the Mumbai city google map to visualize the areas and potholes, Area chart to visualize the pothole count of a selected area with respect to a particular timeline and Bar chart to visualize the hit counts of the top ten problematic potholes in a selected area. But in this design the user does not see all the potholes in the map unless he zooms into area where the pothole is present. This is because the potholes are very small when compared to the Mumbai city.

- Abstract 2 uses the Tree diagram with gives the areas and potholes as the nodes. So we can select each area and potholes by clicking on the corresponding nodes and get the associated Area and Bar chart.

- Abstract 3 divide's the map into two parts, one to visualize and select the Mumbai city areas and the other to visualize the selected area.

Initial Design:

We combined ideas of the abstracts to form the initial design. As shown in above figure, in our initial design, we used the following visualization components.



1) Google Maps

Purpose:

- To visualize and navigate the area wise Mumbai city.
- To visualize the locations of the potholes and their corresponding roads.

2) Tree Diagram

Purpose:

- To see all the potholes and directly navigate to the individual pothole location in the map by selection of the node.

3) Area chart

Purpose:

- To show the count of the potholes in a selected area with respect to particular timeline.

4) Bar chart

Purpose:

- Gives the top ten potholes with maximum count of hits (vehicle encounters).
- Gives the area-wise pothole hit counts.

But we faced many visualization issues with this design. In this design, if the number of potholes is high, user does not see all the pot holes as the tree will be extremely clustered. Hence, the user cannot directly navigate to the pothole in the map. Also the normal tree diagram occupies a lot of space as the number of nodes (areas) increase. In Area and Bar charts, the user can only see the respective data for the selected area. This is very ineffective as there is no means of doing inter-area comparison of the data.

Changes to the initial Design

We tackled the above problems by making the following modifications to our design.

a) Replacing Normal Tree with Reingold–Tilford Tree

We removed the normal tree and replaced it with a Reingold–Tilford Tree into our design. This tree is a radially wrapped version of normal tree. This saves and lot of space. The inner layer of the Reingold–Tilford Tree ford tree consists of area nodes and the outer layer consists of the pothole node. If the number of potholes is very high, the nodes on the tree get very congested and it would be difficult for the user to select each pothole. Hence we made another modification to the tree. When the user clicks the area node of the tree, all the pothole nodes get opened, making them more visible and easy to select as shown in the figure below. By this design, the user can see all the potholes and directly navigate to them in the map by clicking the corresponding node in the tree.

b) Removing Bar Chart

To make the design more effective, we decided to remove the bar chart as we can modify the tree to show the problematic potholes in an area. We show the number of hits of a pothole by the diameter size the node. That is, the more the size of the node, the more the hit count and the corresponding pothole is more problematic. The Reingold–Tilford Tree can be seen in the figure below.

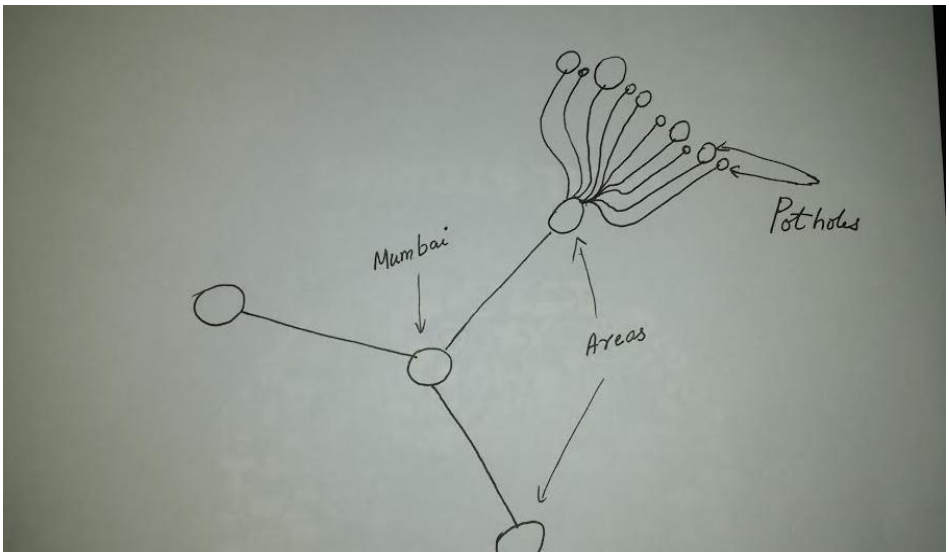


FIGURE - ROUGH SKETCH OF REINGOLD–TILFORD TREE

c) Changes to Area Chart

We also wanted to improve our area chart and make it more effective for comparing and analyzing the data. Hence we decided to modify it to stacked area chart. In this design, we can dynamically add and remove the areas that we want to analyze using buttons as shown in the figure below. This would be very effective for comparing number of potholes in the respective areas for a given time. We also decided to add a tooltip to show the data on hovering as shown in the diagram below.

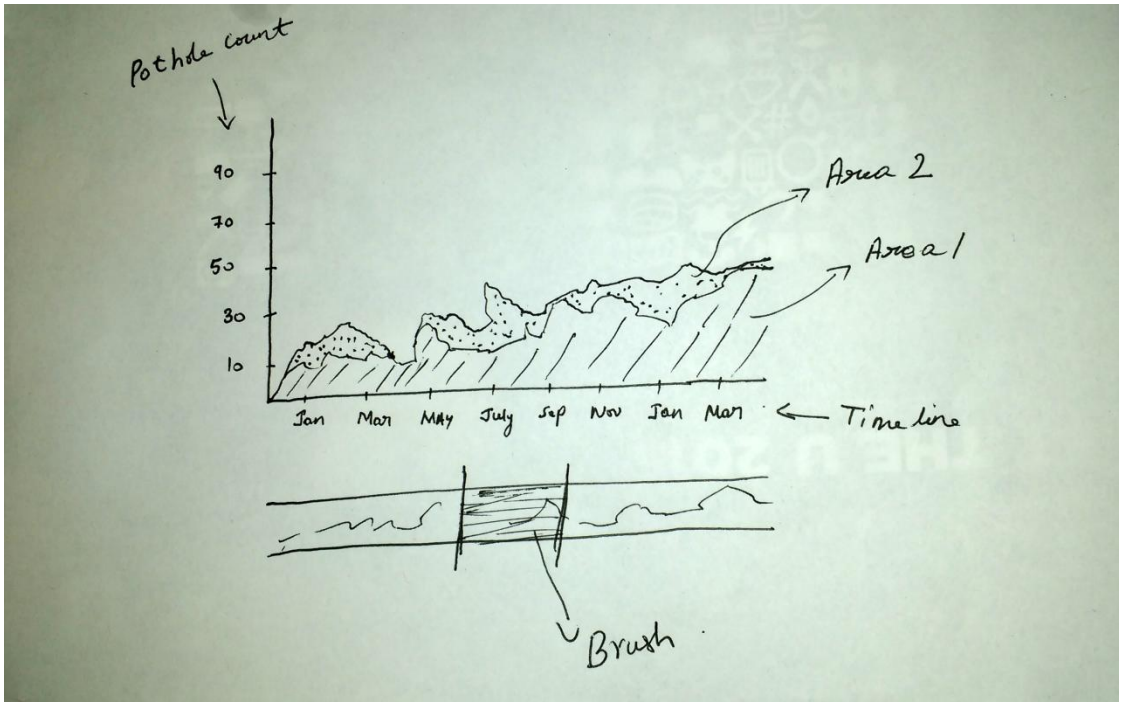


FIGURE - ROUGH SKETCH OF THE STACKED AREA CHART

For interactively selecting the time range, we have added the brush to our stacked area chart. Using this brush, the user can select the time range for which he wants to see the pothole counts.

Final Design

The following is the diagram of our final design.

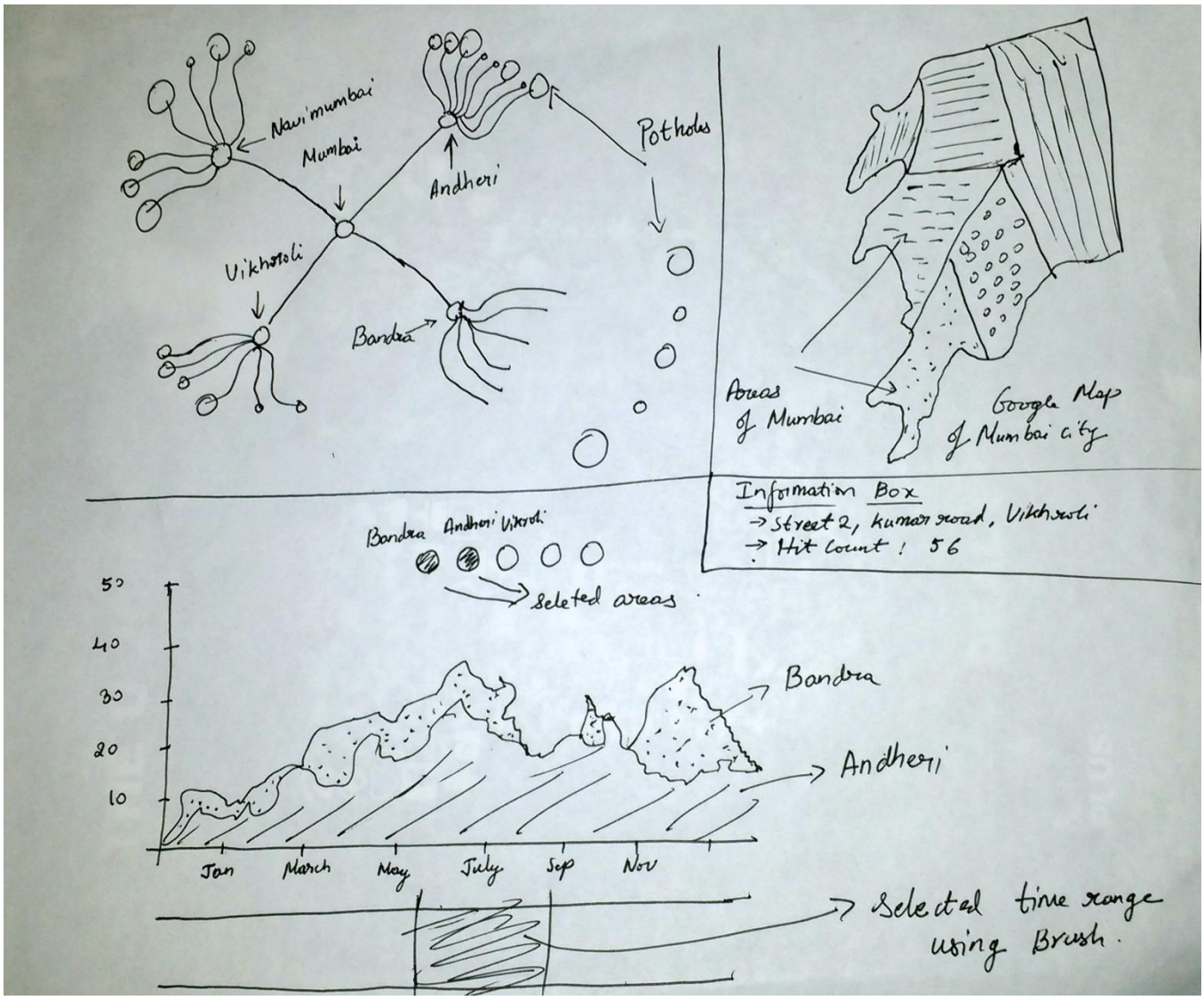


FIGURE - ROUGH SKETCH OF OUR FINAL DESIGN

Finally after many modifications and improvements our design consists of 3 visualization components. The final design shows all the visualization in a single webpage. Initially, when the user opens our visualization webpage, he sees the Tilford tree and Google map. If he scrolls down, he would see the stacked area chart.

The following explains the role of each chart in our visualization.

1) Tilford Tree

- The Tilford tree is mainly used for navigating to specific Areas and potholes in the Map by clicking on the nodes.
- It informs the user about the pothole hit counts.
- It also does Area wise visual comparison of the pothole hit count using the size of the nodes.
- On hovering the pothole nodes, the total information about the pothole including address, key institutions nearby and hit count.
- On clicking an area node, its pole hole count area chart emerges.

2) Google Map

- The Google map is mainly used for visualizing the area and potholes of Mumbai city.
- It shows each area with a chloropath intensity based on the number of potholes on that area.

3) Stacked Area chart

- Compare the pothole counts of selected areas using the stacked area chart.
- On hovering the chart at particular time/date, shows the pothole count of each area at that time.
- We can dynamically add and remove different areas by selection.

4) Information Box

- It gives the complete information of a pothole including Address, hit count and nearby key places.

FUNCTIONAL PROTOTYPE FOR MILESTONE 1

FIGURE - SHAPE FILE OF ADMINISTRATIVE AREAS

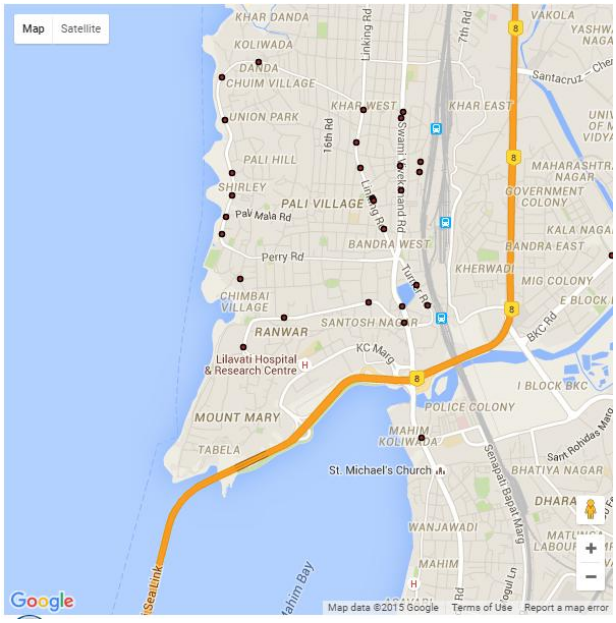


FIGURE – POTHOLE IN THE ANDHERI EAST AREA

We have used Google maps to display the potholes in Mumbai city. Initially the map is centered at centre of Mumbai city at a zoom level of 11. We are doing our pothole analysis per each administrative area of Mumbai city. Hence, we have overlaid the shape files of administrative areas of Mumbai using D3 overlay. These administrative areas can be seen in the figure - 5.

For the milestone 1, we have implemented the following.

Currently, we are only doing for Andheri East area (figure - 9), but the final design would be for all areas.

- On clicking any of the administrative area, the map would zoom in to centre of that administrative block.
- The potholes currently open in that administrative area will be displayed.

For the final design implementation, we will implement few more features like setting the chloropleth (opacity and color) of the administrative area on based on the number of open potholes and establishing the linking with other plots.

REINGOLD–TILFORD TREE

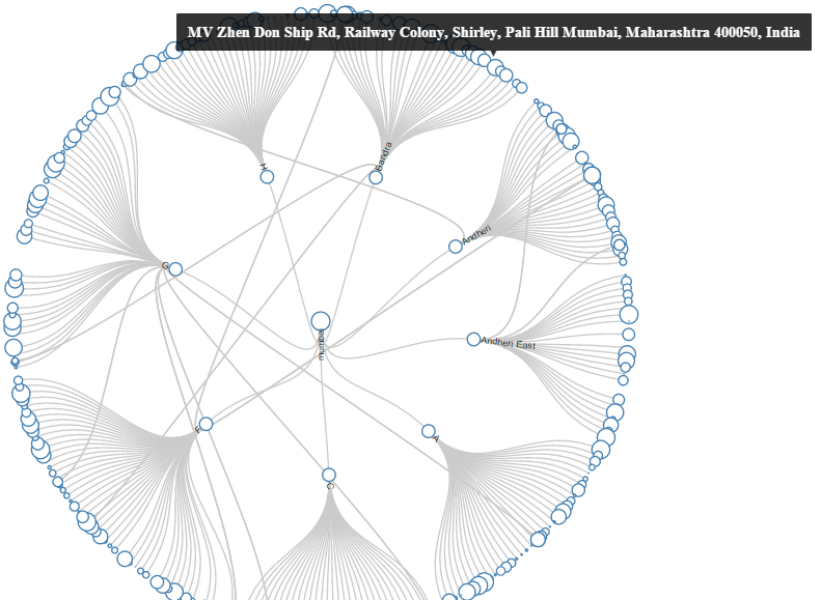


FIGURE 10 – REINGOLD–TILFORD TREE

In this intermediate version, we are using Tilford tree to map the potholes from Mumbai to Area (Bandra, Andheri, etc.) and Area to Potholes. In this pothole size (circle at depth-2) changes according to the hit count. It helps the government to identify the priority to fix the potholes. In this we are reading data from the csv file.

In final delivery, this tree contains different color for city, area and potholes and all the potholes circle comes in the sorted order and on clicking on any pothole will raise the event which is read by the other visualization to change their fields.

AREA CHART

In the Area Chart, we plot the months in x-axis and number of potholes in the y-axis. As of now, we have just plotted the area chart of only one administrative boundary. In this chart, we can know the number of open potholes in an area at a particular time of year. As the final work, we would plot a different area chart of different administrative boundary and invoke the corresponding plot by clicking on the administrative boundary in the map.

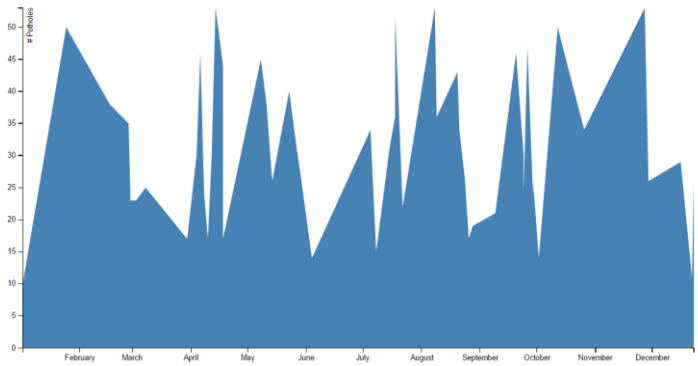


FIGURE - AREA CHART OF BANDRA

BAR GRAPH

This bar graph show the top 10 potholes of the selected region with their hit count, this will help government to take action for the most serious pothole.

In final delivery, links will get implemented which will point to that particular pothole in the map and tree.

FIGURE - BAR GRAPH

IMPLEMENTATION

This section discusses the intent and content of each functionality of the interactive visualization components with clear and well-referenced images showing the key design and interaction elements.

REINGOLD–TILFORD TREE

The following are the various functionalities of the Reingold–Tilford tree in our visualization.

- Basic layout of the tree

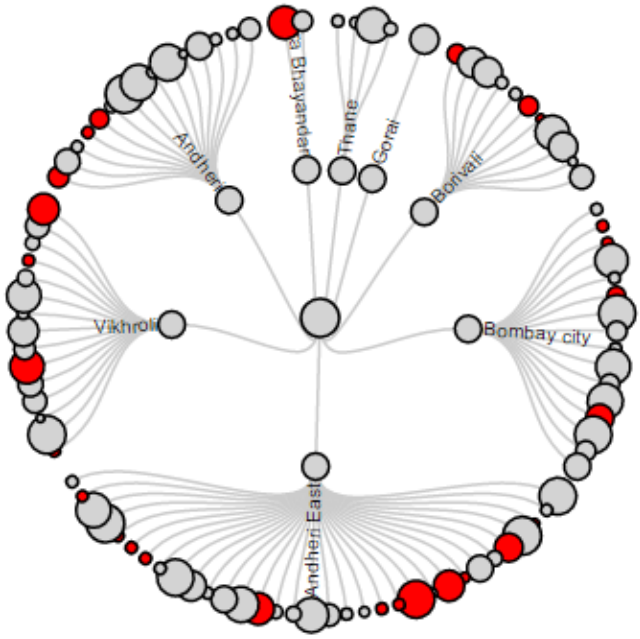


FIGURE - BASIC LAYOUT OF THE TILFORD TREE

  - As shown in the above figure, this layout gives the user an overview of all the areas and potholes in the Mumbai city as the nodes of the tree.

- On clicking the Area node

- If the number of potholes is very high, they would be very close to each other as a cluster in the basic layout.

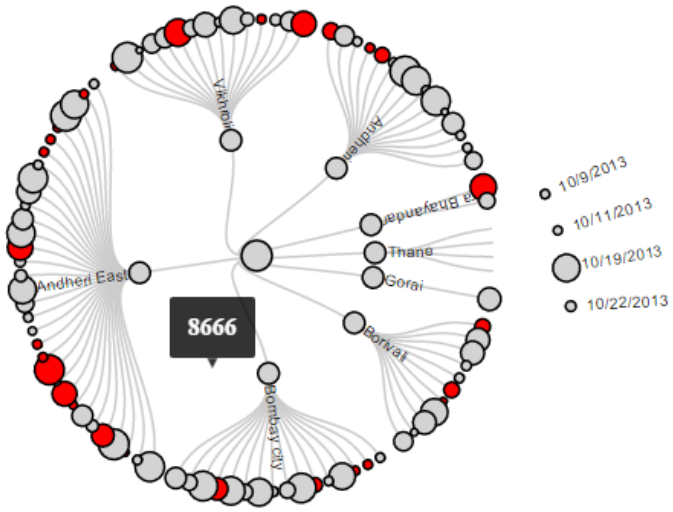


FIGURE - POTHOLE NODES OF THANE AREA DEPECTING THE ON CLICK FEATURE

- The user would find it very difficult to select individual nodes. Hence we added this functionality.
- On clicking a particular area, the corresponding potholes open up in the Tilford tree as shown in the above figure.
- The pothole nodes are sorted according to the date as shown in figure 14.
- On clicking a particular area node in the Tilford tree will also trigger the Google map to zoom into corresponding area as shown in the figure 15.

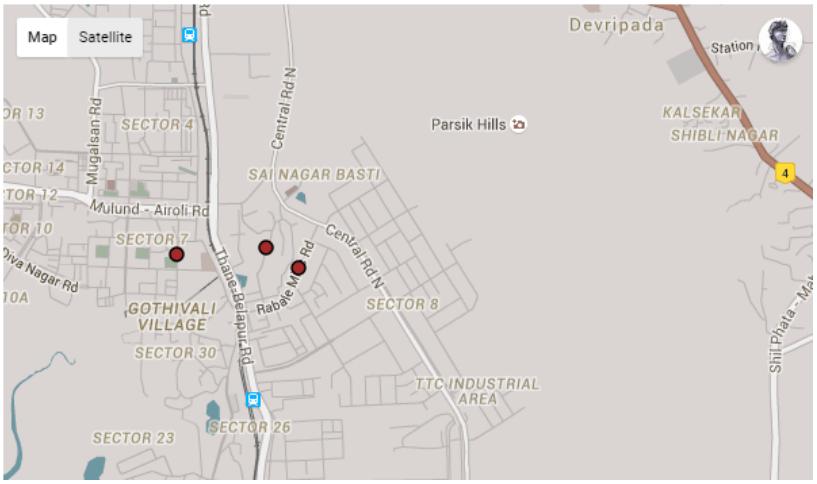


FIGURE - GOOGLE MAP SHOWING THE THANE AREA, ZOOMING TRIGGERED BY ONCLICKING THE THANE NODE IN THE TREE

- Clicking the Pothole node
    - On clicking the pothole node will trigger the map to zoom further inside, showing the corresponding pothole.

    - It will also trigger the text box to show the complete information of the pothole including the address, vehicle hit count and key places nearby.

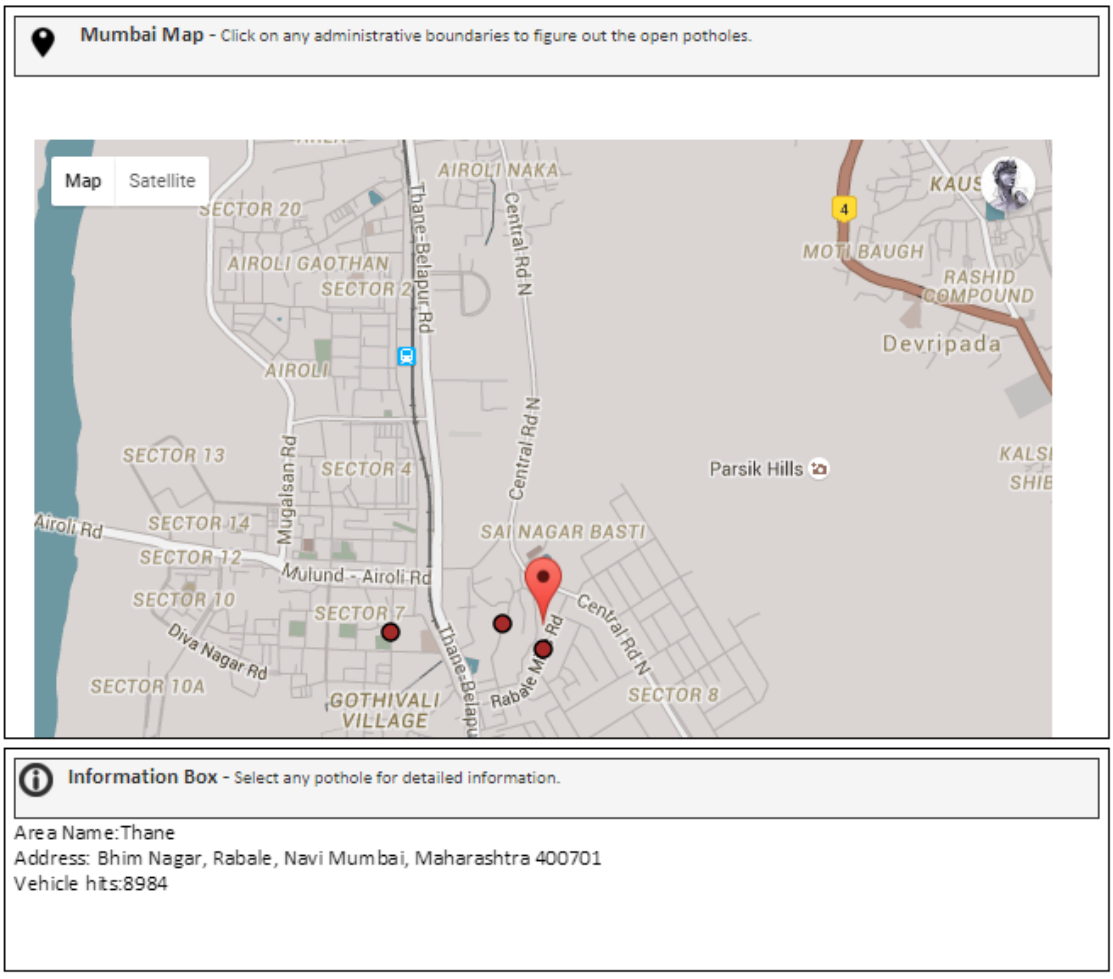


FIGURE - GOOGLE MAP SHOWING THE SELECTED POTHOLE AND THE INFORMATION BOX CONTAINING INFORMATION OF THAT POTHOLE

- Size of the pothole nodes.

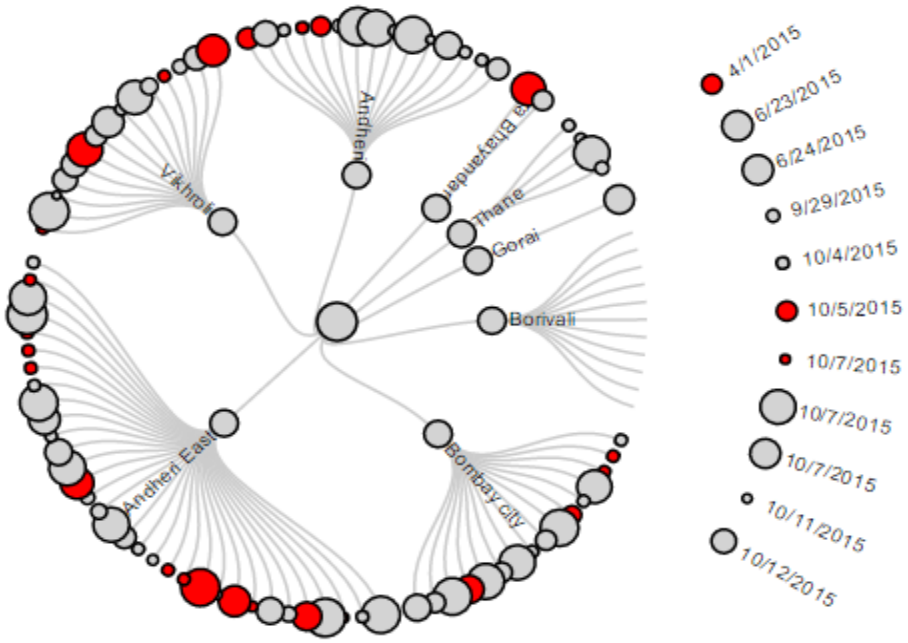


FIGURE - TREE DIAGRAM SHOWING THE POTHOLE NODES OF VARYING SIZES AND COLORS

  - The size of the pothole nodes informs the user about the vehicle hit count of the potholes.
  - This functionality enables the user to do a visual comparative analysis with respect to the vehicle hit count.
  - The pothole nodes with larger sizes have higher hit counts and are more problematic.

- Color of the pothole nodes.

  - With respect to the color there are two types of potholes.
  - The grey colored nodes indicate normal potholes and the red colored nodes indicate the potholes that have key places such as hospitals and schools.
  - Hence the red colored potholes pose more problems.

- On clicking Mumbai node
  - This triggers the Google map to zoom out to the level of Mumbai city.

GOOGLE MAP

The following are the various functionalities of the Google maps in our visualizations.

- Basic layout

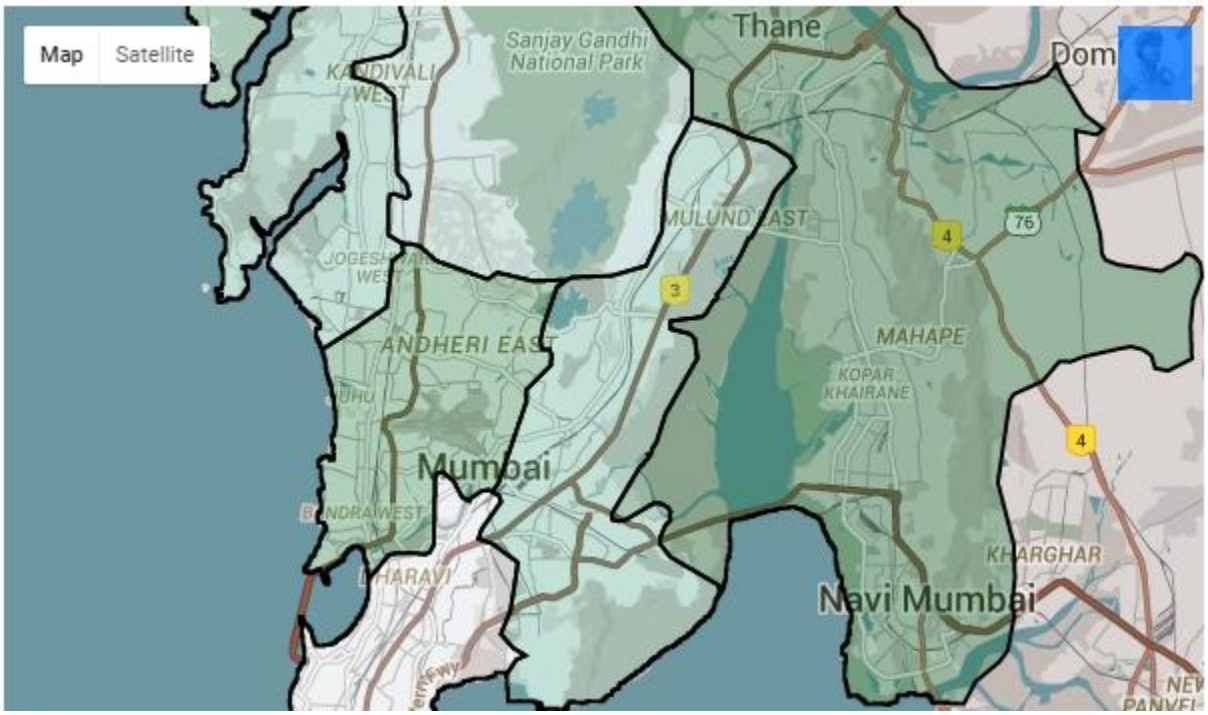


FIGURE- BASIC LAYOUT OF THE GOOGLE MAP SHOWING THE MUMBAI CITY AND ITS AREAS

The basic layout shows the map of the Mumbai city and its areas.

- Chloropath of the Areas
  - The intensity of the area chloropath indicates the number of potholes in the area.
  - The Area with higher intensity of the chloropath has more potholes.
- On clicking the areas
  - On clicking an area, the map will zoom to the center of that area. The potholes in that area will also be visible as shown in the figure 19.

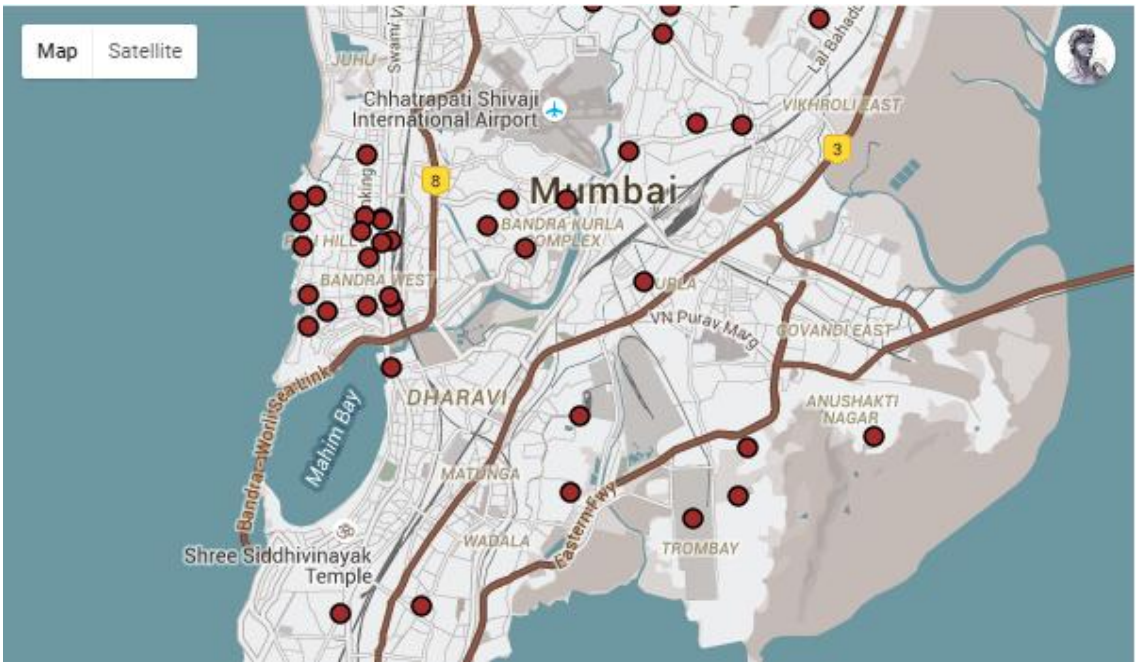


FIGURE - DIFFERENT POTHOLED OF THE SELECTED AREA

- On clicking the potholes.

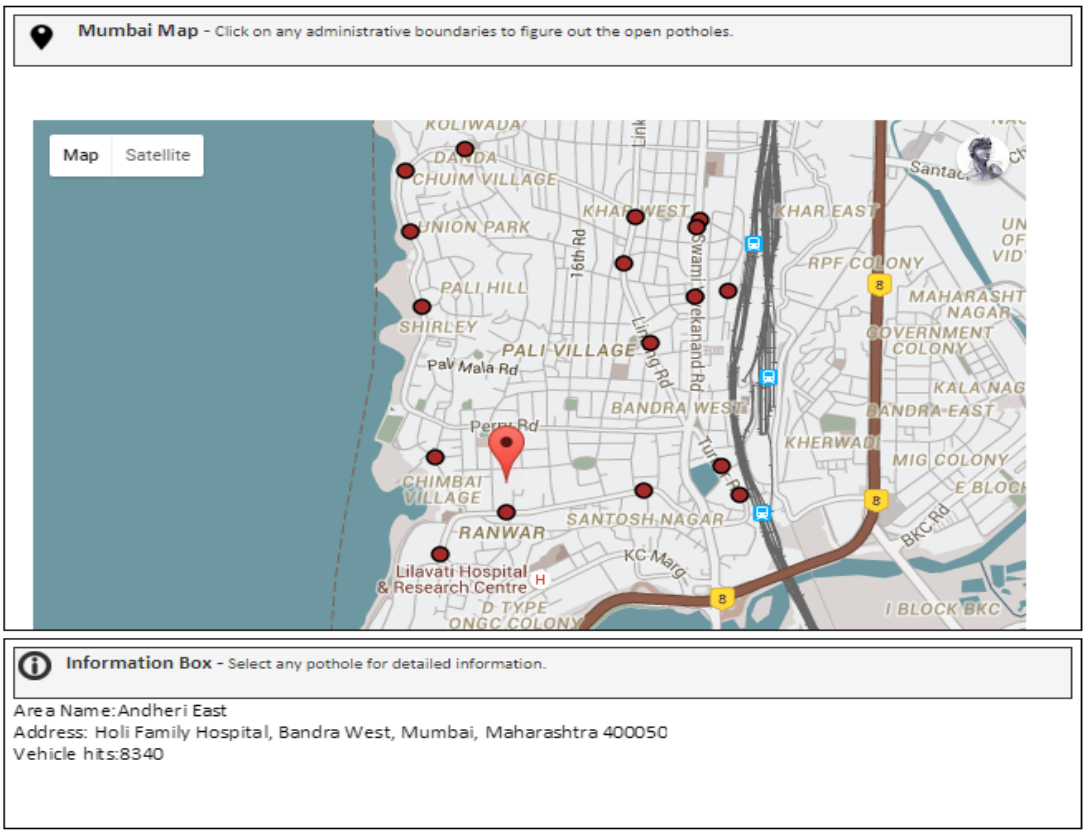


FIGURE - CLICKED POTHOLE WITH A MARKER ON IT AND THE ASSOCIATED DATA IN THE INFORMATION BOX.

- On clicking a pothole in the map, a marker appears on it as shown in the above figure
- It will also trigger the text box to show the complete information of the pothole including the address, vehicle hit count and key places nearby.

Stacked Area Chart

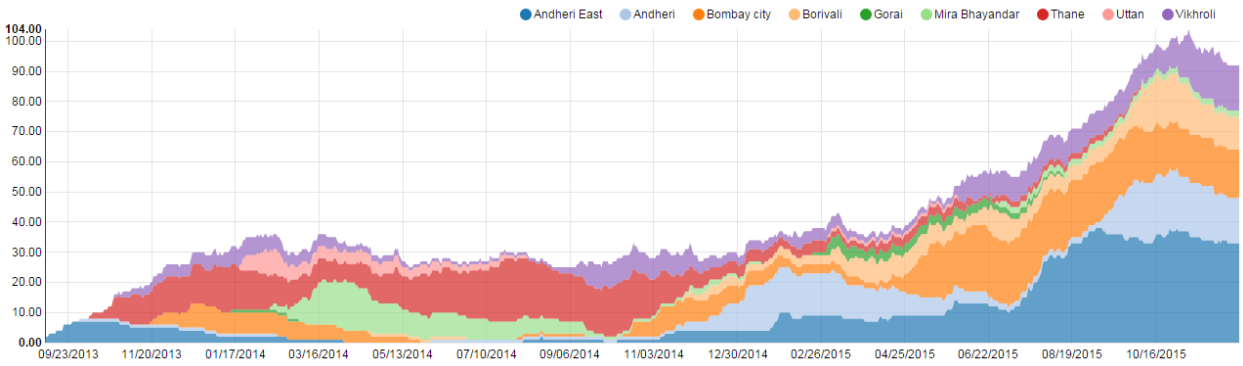


FIGURE - STACKED AREA CHART

The following are the various functionalities of the stacked area chart in our visualization:

1) Tooltip on hover

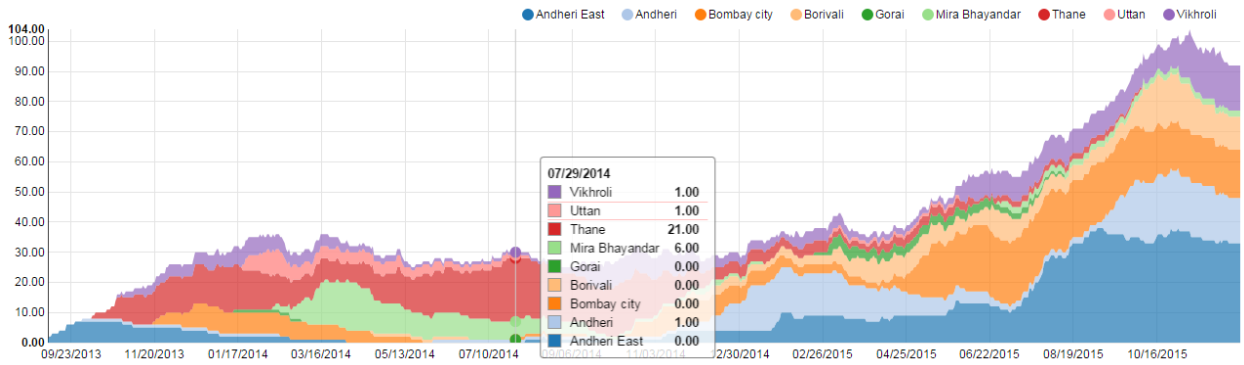


FIGURE - TOOLTIP HOVER GIVE DETAILED INFORMATION OF THE SELECTED AREAS

- On hover on the stacked area chart for a particular time, show the pothole count data of all the areas in the chart for that time.

2) Area selection

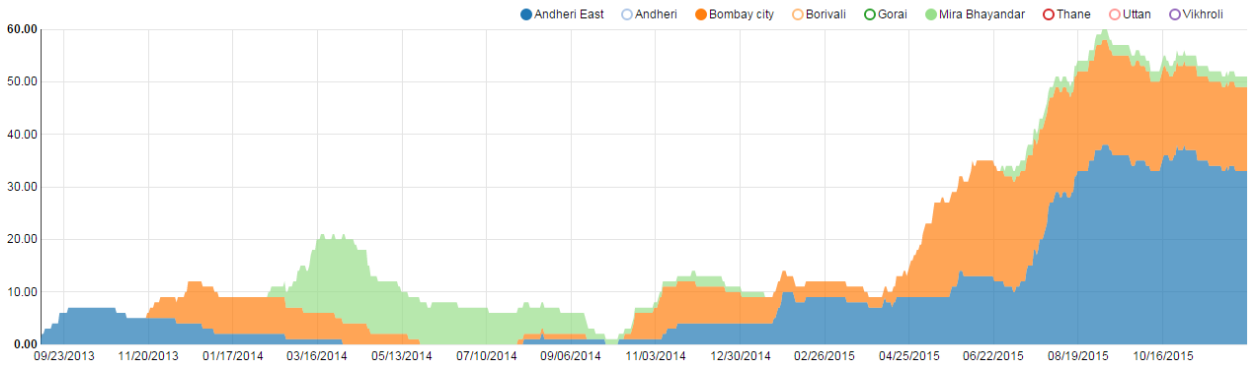


FIGURE - COMPARSION OF SELECTED AREA

- This feature enables the user to dynamically add and delete the areas in the stacked chart.
- Using this feature, the user can select only the required areas and compare there pothole counts.

3) Brush

Optional Feature: This feature enables the user to select time range for which he wants to analyse the stacked area chart. We have implemented this feature but we are not added it because of the performance issue.

ANALYSIS OF DATA AND CONCLUSION

The main take away from this visualization project is the interactivity and visualization of the associated pothole data. A typical user can derive interesting analysis and results from the project.

The following is the screen shot of the website of our visualization project.

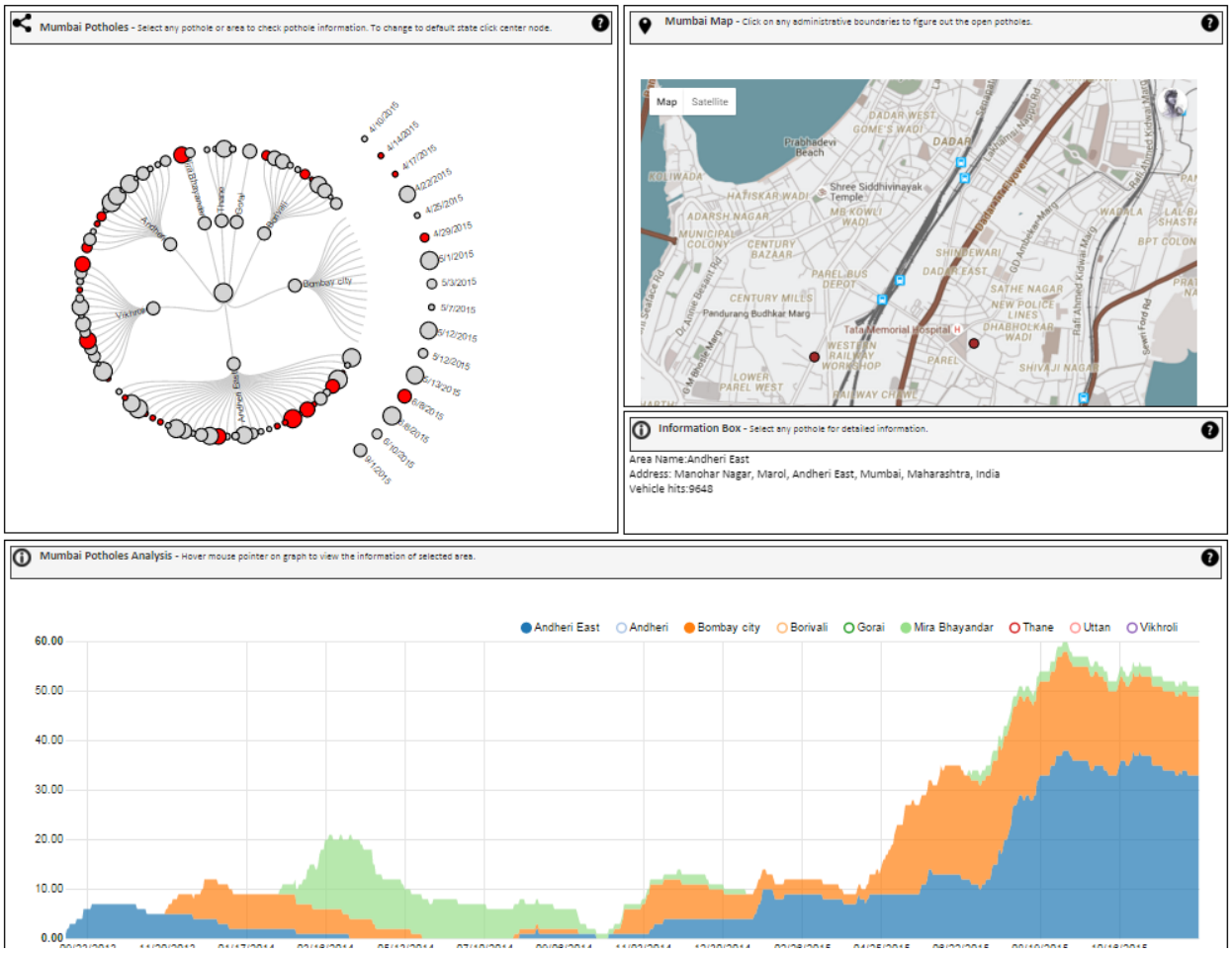


FIGURE - IMAGE OF THE COMPLETE WEBSITE

As mentioned earlier, we manually selected few lat long locations of the Mumbai city roads and generated the associated pothole data with a script. Given that it is highly impossible, we have tried our best to generate the data that closely reflects the real world.

As the stacked area chart gives the information of the pothole count of many areas in a particular, the user do comparative analysis of area wise occurrence of the new potholes based on the seasons of the year using the stacked area chart.

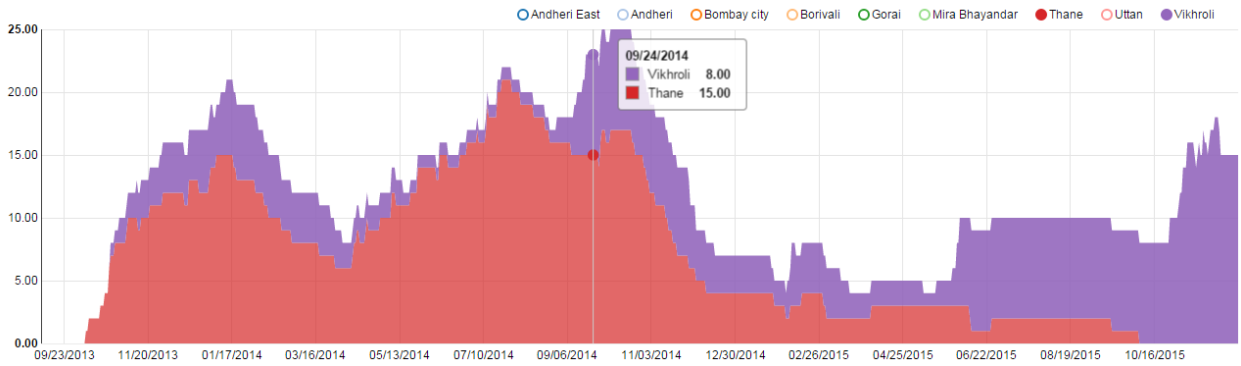


FIGURE - COMPARISON OF TWO AREA'S DURING RAINY SEASON

The above area chart shows two spikes in the number of potholes at the months of September and October. These months typically fall in the monsoon season in India. Hence the user can infer that the occurrence of the potholes is high in the monsoon season. From the area chart, the user can also do a comparison of different areas and gauge the roads of those areas.

From the REINGOLD–TILFORD tree, the user can interactively get a lot of data about the potholes. The size of the pothole nodes give the vehicle hit count of the nodes. Hence the user can know the most dangerous potholes and their corresponding roads from the Google maps for a particular area. Information about the key places like hospitals and schools nearby the pothole would be handy.

From our Visualization project, we feel that we are successful in answering Questions mentioned elsewhere in the process book. There is scope for adding more components into this visualization. In the future, we would like to forecast the area wise occurrence of the potholes based on the previous data. We believe that this could be done using Machine learning algorithm. Though this kind of analysis would require a lot more data, we would also like to gauge the effectiveness of the local contractors and government with respect to road constructions.

Working on this project was a good opportunity for our group. We got to do some cool visualizations that would actually alleviate a ground level problem of a common commuter. We have also learned a lot new things like the REINGOLD–TILFORD tree, working with Google map APIs and developing a very interactive stacked area chart.

RESOURCES

Tree Implementation - Radial Reingold–Tilford Tree - http://bl.ocks.org/mbostock/4063550

Area Chart Implementation – NV D3 - http://NVD3.org

Map Implementation - Google Map and Geocoder API - https://developers.google.com/maps/