VISUALIZING VEHICLE CRASH INCIDENTS IN MASSACHUSETTS

CS 573 Final Project

SACHIN SUDARSHANA (*Email*: ssudarshana@wpi.edu, *Github*: sachin019)
SOUNTHAR MANICKAVASAGAM (*Email*: smanickavasagam@wpi.edu, *Github*: msounthar)
SNEHASISH BARMAN (*Email*: sbarman@wpi.edu, *Github*: sbarman-mi9)

Table of Contents

Background & Motivation

We had worked on the New York Motor Vehicle Collisions for a previous assignment in a Data Science class. That provided us the opportunity to realize the importance of the dataset and how good visualizations can effectively convey deep insights and broad perspectives.

Using this dataset gives us a scope to come up many different types of visualizations, all of us which will be closely connected, So designing and implementing them will be both be interesting and a good learning experience.

Project Objectives

In this project, we hope to uncover any hidden patterns in the crash incidents in Massachusetts and also try to answer the following questions

Are there locations that are more prone to these incidents?
Are there any factors that influence these collisions?
Which type of vehicles are involved in an accident and is there a pattern?
When do these collisions happen the most?

Data

Data Source

The dataset we used is the vehicles crash data in various towns of the Commonwealth of Massachusetts. The source of the dataset is the Massachusetts Registry of Motor Vehicles (RMV).

1	Crash_Number	Town_Name	Crash_Date	Crash_Time	Crash_Severity	Number_of_Vehicles	Total_Nor	Total_Fata	Manner_c	Vehicle_A	Vehicle_T	Most_Hari	Vehicle_C	Road_Surf	Ambient_	Weather_	At_Roadw	Distance_	Distance_	Distance_	I
2	3420645	WORCESTER	1-Jan-13	1:36 AM	Unknown	3	0	0	Angle	V1: Backin	V1:Not re	V1: Collisi	V1: Passer	Snow	Dark - ligh	Clear/Clear		114 STERLING STREET			
3	3379220	WORCESTER	1-Jan-13	2:37 AM	Property damage only (none i	2	0	0	Angle	V1: Other	V1:Eastbo	V1: Collisi	V1: Passer	Wet	Dark - ligh	Clear	HARDING STREET / VERNON STREET Rte 1:				I
4	3323604	WORCESTER	1-Jan-13	3:14 AM	Non-fatal injury	1	1	0	Single veh	V1: Travel	V1:Eastbo	V1: Collisi	V1: Passer	Dry	Dark - ligh	Not Reported		Rte 290 E		Exit 13 on F	
5	3404812	WORCESTER	1-Jan-13	9:00 AM	Unknown	1	0	0	Single veh	V1: Unknc	V1:Northb	V1: Collisi	V1: Passer	Unknown	Dark - unk	Unknown		273 STAFFORD STREET			
6	3378019	WORCESTER	1-Jan-13	9:53 AM	Non-fatal injury	2	1	0	Rear-end	V1: Travel	V1:Not re	V1: Collisi	V1: Passer	Dry	Daylight	Cloudy	MAYWOOD STREET / MAIN STREET				
7	3378012	WORCESTER	1-Jan-13	1:00 PM	Property damage only (none i	1	0	0	Single veh	V1: Travel	V1:Not re	V1: Collisi	V1: Passer	Wet	Daylight	Clear/Cloudy		117 SHREWSBURY STREET			
8	3378136	WORCESTER	1-Jan-13	1:58 PM	Non-fatal injury	1	1	0	Single veh	V1: Travel	V1:Southb	V1: Collisi	V1: Light t	Dry	Dark - ligh	Clear/Clear		317 MASSASOIT ROAD			
9	3378026	WORCESTER	1-Jan-13	2:05 PM	Property damage only (none i	2	0	0	Head-on	V1: Travel	V1:Southb	V1: Collisi	V1: Passer	Ice	Dark - roa	Snow/Sleet, hail (fre		19 WALL STREET / CHROME STRE			
10	3378023	WORCESTER	1-Jan-13	2:07 PM	Non-fatal injury	1	1	0	Single veh	V1: Travel	V1:Eastbo	V1: Collisi	V1: Passer	Dry	Dark - ligh	Cloudy		192 PLEASANT STREET			
11	3378107	WORCESTER	1-Jan-13	2:45 PM	Property damage only (none i	2	0	0	Rear-end	V1: Backin	V1:Not re	V1: Collisi	V1: Passer	Snow	Dark - ligh	Snow/Snc	MAY STREET / MASON STREET				
12	3378016	WORCESTER	1-Jan-13	2:55 PM	Non-fatal injury	2	1	0	Rear-end	V1: Slowir	V1:Southb	V1: Collisi	V1: Passer	Unknown	Dark - unk	Unknown/Unknown		215 MILL STREET			I
13	3378011	WORCESTER	1-Jan-13	5:07 PM	Property damage only (none i	2	0	0	Sideswipe	V1: Enteri	V1:Not re	V1: Collisi	V1: Passer	Dry	Dark - ligh	Clear		932 MAIN STREET			
14	3363381	WORCESTER	1-Jan-13	5:30 PM	Non-fatal injury	2	1	0	Rear-end	V1: Travel	V1:Northb	V1: Collisi	V1: Passer	Dry	Dark - ligh	Clear		499 PARK AVENUE			
15	3378010	WORCESTER	1-Jan-13	5:40 PM	Property damage only (none i	2	0	0	Angle	V1: Turnir	V1:Not re	V1: Collisi	V1: Passer	Dry	Dark - ligh	Clear/Cloudy		548 LINCOLN STREET			I
16	3378169	WORCESTER	1-Jan-13	9:34 PM	Property damage only (none i	2	0	0	Angle	V1: Travel	V1:Southb	V1: Collisi	V1: Passer	Dry	Dark - ligh	Clear	VERNON STREET / Rte 290 W				
17	3420647	WORCESTER	1-Jan-13	10:31 PM	Property damage only (none i	2	0	0	Angle	V1: Travel	V1:Northb	V1: Collisi	V1: Passer	Wet	Dark - ligh	Clear/Clear		160 GOLD STAR BOULEVARD			
18	3380022	WORCESTER	1-Jan-13	11:00 PM	Not Reported	2	0	0	Head-on	V1: Travel	V1:Southb	V1: Not re	V1: Passer	Dry	Dark - ligh	Clear		5658 OLGA AVENUE			
19	3420646	WORCESTER	2-Jan-13	6:50 AM	Property damage only (none i	2	0	0	Angle	V1: Travel	V1:Southb	V1: Collisi	V1: Light t	Dry	Daylight	Clear/Cle	MAJOR TAYLOR BOULEVARD / SCHOOL STR				
20	3388837	WORCESTER	2-Jan-13	7:30 AM	Not Reported	2	0	0	Angle	V1: Parke	V1:Not re	V1: Collisi	V1: Passer	Unknown	Daylight	Unknown/Unknown		68 ALBANY STREET			
21	3403165	WORCESTER	2-Jan-13	8:59 AM	Property damage only (none i	2	0	0	Angle	V1: Travel	V1:Northb	V1: Collisi	V1: Passer	Dry	Daylight	Clear/Cle	HIGHLAND STREET / WACHUSETT STREET				
22	3388839	WORCESTER	2-Jan-13	11:07 AM	Non-fatal injury	2	1	0	Angle	V1: Travel	V1:Westb	V1: Collisi	V1: Passer	Dry	Daylight	Clear/Cle	HIGHLAND STREET / WACHUSETT STREET				
23	3388840	WORCESTER	2-Jan-13	12:00 PM	Property damage only (none i	1	0	0	Head-on	V1: Travel	V1:Northb	V1: Collisi	V1: Passer	Dry	Daylight	Clear	CLARK STREET / BURNCOAT STREET				
24	3388841	WORCESTER	2-Jan-13	12:20 PM	Property damage only (none i	2	0	0	Rear-end	V1: Slowir	V1:Northb	V1: Collisi	V1: Passer	Dry	Daylight	Cloudy/Cl	LAFAYETTE STREET / SOUTHBRIDGE STREET				
25	3388842	WORCESTER	2-Jan-13	12:49 PM	Property damage only (none i	2	0	0	Angle	V1: Travel	V1:Eastbo	V1: Collisi	V1: Passer	Dry	Daylight	Clear/Clear		199 HIGHLAND STREET Rte 9 W			
26	3388843	WORCESTER	2-Jan-13	12:49 PM	Property damage only (none i	2	0	0	Rear-end	V1: Chang	V1:Westb	V1: Collisi	V1: Light t	Dry	Daylight	Cloudy	MADISON STREET / BEACON STREET				

Figure 1 Data Snapshot

The crash data for every town in Massachusetts can be downloaded by completing the request form at: http://services.massdot.state.ma.us/crashportal/DataRequest.aspx

We collected the data for the year 2013 of the top 20 towns and cities by population in Massachusetts.

The individual csv files of all the 20 towns and cities were then merged into a Master csv file. This csv file was used for further data processing.

Top 20 cities/towns of MA
Boston
Worcester
Springfield
Lowell city
Cambridge
New Bedford
Brockton
Quincy
Lynn
Fall River
Newton
Lawrence
Somerville
Framingham
Haverhill
Waltham
Malden
Brookline
Plymouth
Medford
Taunton

Figure 2 Top 20 towns in Massachusetts

Data Description

The data contains about 24 different attributes for every crash incident reported.

No.	Attribute Name	Description
1	Crash Number	A unique identifier for the crash
2	City/Town Name	Place where the crash occurred
3	Crash date	Date of occurrence
4	Crash time	Time of occurrence
5	Crash severity	Severity of the incident
6	Number of Vehicles	Number of vehicles involved in the incident
7	Total Non-fatal injuries	Total non-fatal injuries reported
8	Total Fatal injuries	Total fatal injuries reported
9	Manner of Collision	Description of the crash
10	Vehicle Action prior to crash	Description of the actions of the vehicles involved prior to the crash
11	Vehicle travel directions	Directions of travel of all the vehicles involved

12	Most harmful events	Events caused by the vehicles as a result of the crash
13	Vehicle configuration	Type of the vehicles involved
14	Road surface condition	Condition of the road
15	Ambient Light	Lighting at the scene of crash based on time of day
16	Weather condition	Weather conditions at the time of crash
17	At Roadway intersection	Street names if the crash occurred at an intersection
18	Distance from nearest roadway intersection	Address of the nearest interesection
19	Distance from nearest mile marker	Nearest Mile marker if incident occurred on a freeway
20	Distance from nearest exit	Nearest exit number if incident occurred on a freeway
21	Distance from nearest landmark	Nearest landmark from site of crash
22	Non Motorist type	Type of non-motorists involved
23	X Coordinate	Location encoded by GIS software
24	Y Coordinate	Location encoded by GIS software

Figure 3 Data Description

Data Processing

Modifying the Master Dataset

The data was processed primarily in MS Excel. There were several inconsistencies in the data and to render our visualizations we had to create different subsets of the data.

The following steps were performed in excel to prepare the data

1. Location of the Incident:

The dataset contained multiple columns that indicates the location of the incident. "At Roadway intersection", "Distance from nearest roadway intersection", "Distance from nearest mile marker", "Distance from nearest exit", "Distance from nearest landmark". There was a value present in only 1 or 2 columns for every row. All these values were merged / reconstructed into a single column called "LOCATION". These columns were then deleted.

2. Vehicles involved in the incident:

The "Vehicle configuration" column contained data about the vehicles involved in the crash. This was a composite column which had information about all the vehicles in a single cell.

Ex: *V1: Passenger car / V2:Light truck(van, mini-van, panel, pickup, sport utility) with only four tires*

This column was decompoased into 2 separate columns titled "Vehicle 1" and "Vehicle 2" and unwanted details about the vehicles were deleted.

Vehicle 1	*Vehicle 2*
Passenger Car	*Light Truck*

There were many vehicle types that were reported in the dataset. We considered only the following types for our visualization.

Passenger Car, Light Truck, Bus, Motorcycle, MOPED, Other, Not Reported, Unknown Vehicle Configuration, Single-Unit Truck, Motor home recreational vehicle, Truck/Trailer.

All data points that had any other type, were purged.

Also, to help our visualization, we performed discretization of the Vehicle type data.

Ex.

V1	*V2*	*Pass. Car*	*L Truck*	*Bus*	*M.cycle*	*Moped*	*Other*	*Nt. Rptd.*	*Unknown*	*Single Truck*	*Recreational*
Pass. Car	*Pass. Car*	*2*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Pass. Car	*L Truck*	*1*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

*Abbreviated due to lack of space

3. Vehicle Directions:

The "Vehicle travel directions" column also was a composite column. Similar decomposition was also performed on this column

Ex:

Vehicle travel directions
V1:Southbound/V2:Northbound
V1:Eastbound/V2:Eastbound

Vehicle1 travel directions	**Vehicle2 travel directions**
Southbound	**Northbound**
Eastbound	**Eastbound**

4. Addition of new columns/ Merging of Columns:

We added a few new columns in our dataset that we required for our visualization.
From the date column, we created new "Month" column.
From this month column, we created a "Season" column.
The value for the season column was generated as follows:

spring - March to June,
summer - June to September,
fall (autumn) - September to December, and
winter - December to March.

The "Total Non-fatal injuries" and "Total Fatal injuries" columns were merged into a single column titled "Total Injuries"

After the above processing steps, our final processed dataset had **35809** data points with **30** Attributes worth **15 MB**.

Creating subsets of data for visualization

To aid our visualizations, we created subsets of data from this master dataset. By this, we could focus more on the visualization code in JavaScript than worrying about data processing/filtering.

The following subsets of data were created:

1. **Street & Time:** Dataset containing No. of Accidents for every city for the Top 10 locations during various times of the day. The times were grouped into 1 hour intervals. Ex: 12AM- 1AM, 1AM – 2AM etc.
2. **City & Time:** Dataset containing total number of accidents per city grouped into 1 hour intervals.
3. **Type & Time:** Dataset containing total number of vehicles by type involved in accidents in every city grouped into 1 hour intervals
4. **City & Type:** Dataset containing total number of vehicles by vehicle type involved in accidents per city grouped into 1 hour intervals
5. **Season & Type:** Dataset containing total number of vehicles by vehicle type involved in accidents per city grouped into the 4 different seasons
6. **Street & Season: :** Dataset containing No. of Accidents for every city in each of the 4 seasons grouped into 1 hour intervals

Design Evolution

We initially planned to do analysis on a few interesting attributes: time, factor/causes of crash, conditions, vehicle types, injuries etc. Our initial ideas in implementing the code for this project are described in the following section.

Initial Ideas

We came up with 3 initial ideas.

Idea 1: Single Page Visualization

Here, each visualization will be displayed on separate pages. On the first page, we show a map of all the towns of Massachusetts, the color of which can be changed by moving the "range slider" to different ranges of time in a day. On selection of a town, we drill down to the details of individual towns and provides visualizations such as streamgraph (visualize the number of crash incidents by vehicle types and time), and parallel coordinate plot of all the locations in a town (compare the crash incidents by different hours in a day).

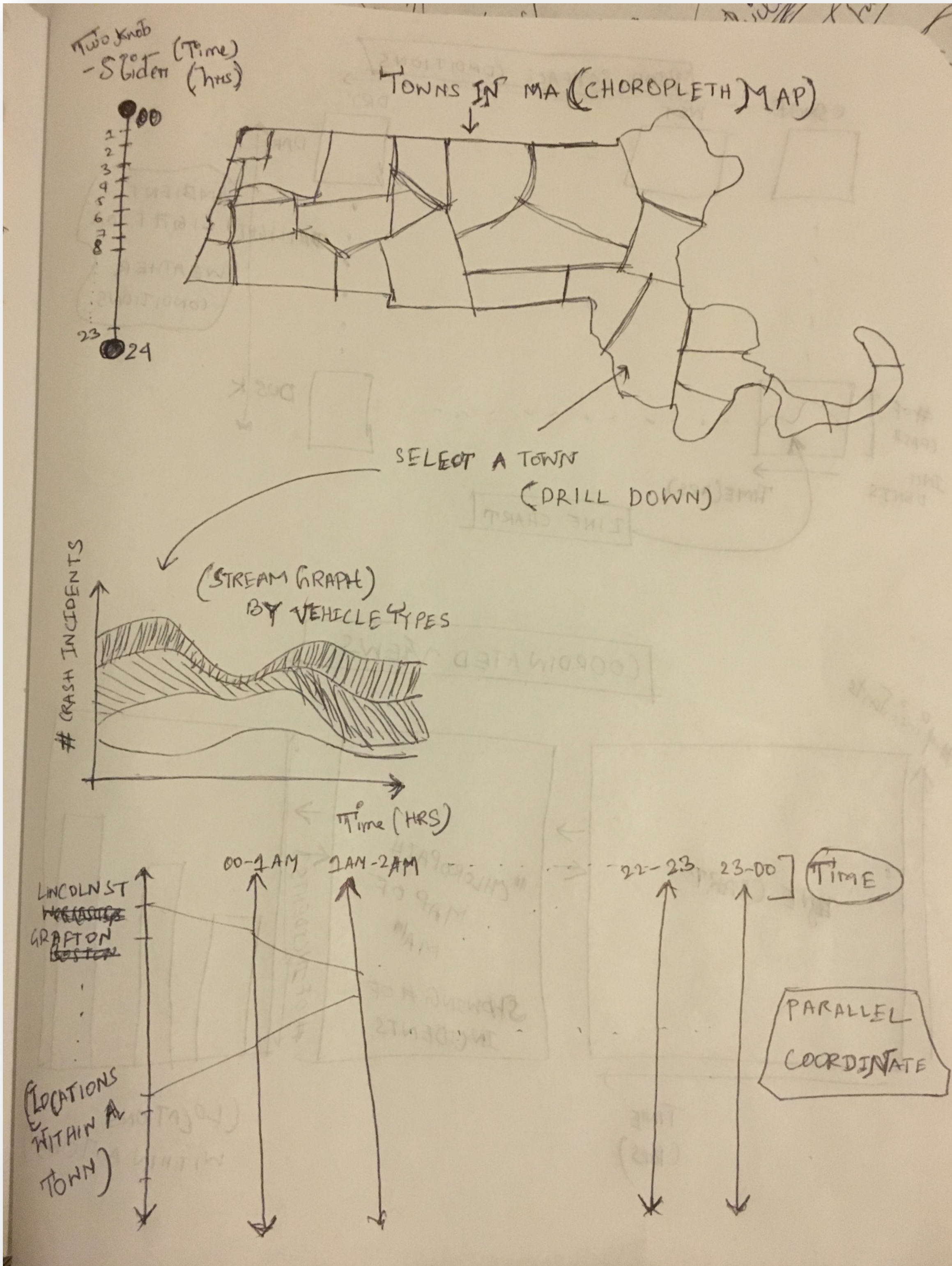

Figure 4 Idea #1 –Map, Steam graph and Parallel plots

Idea 2: Matrix Views

Here, we the first page remains the same as in Idea #1, but the drill down pages to individual towns, and the associated variables are displayed as matrices, In the example shown above, each individual cell is a line chart displaying no. of crash incidents vs. time (hrs) and they are bifurcated by different variables such as weather conditions, road conditions, etc.

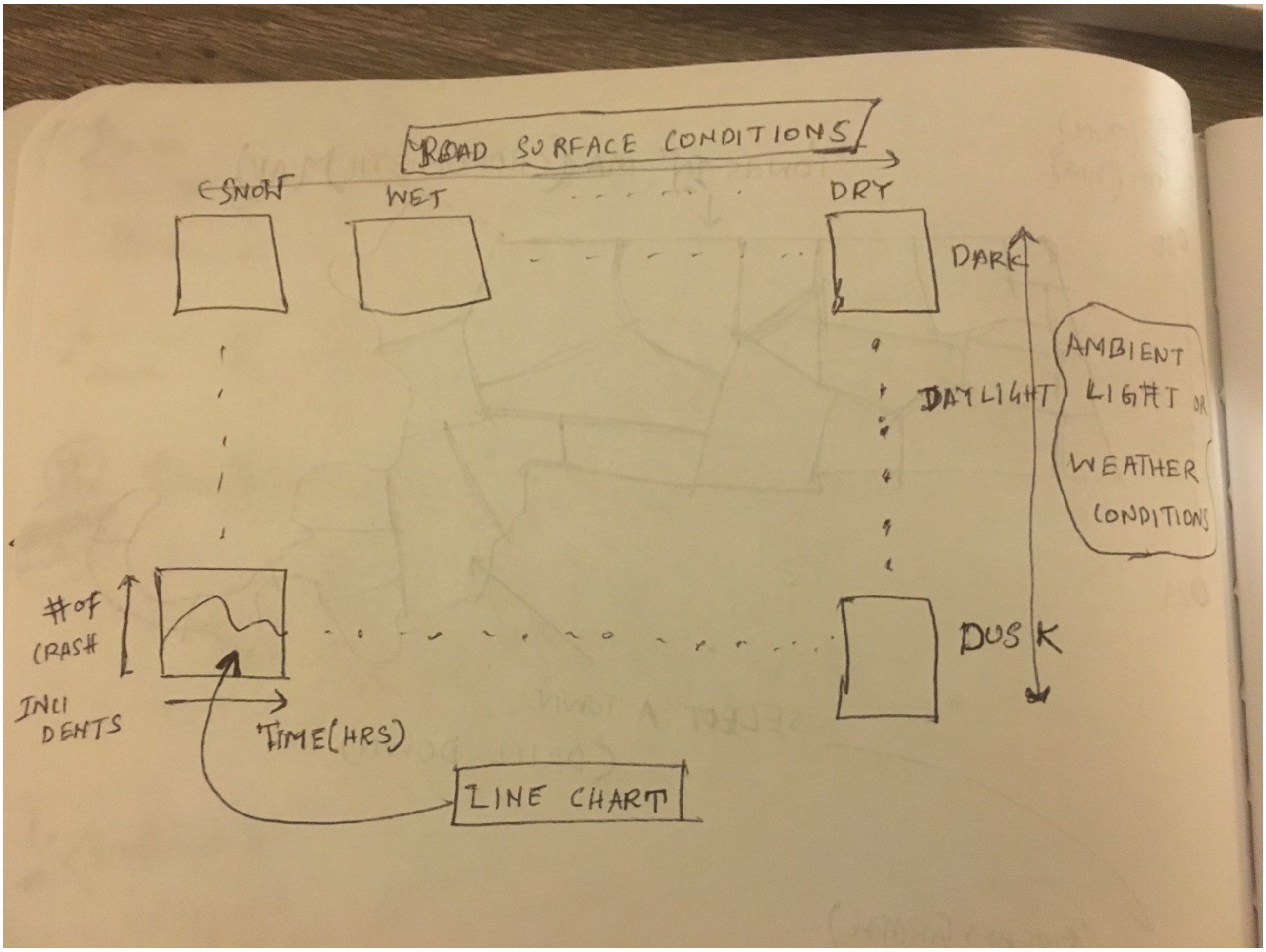

Figure 5 Idea #2 - Matrix views

Idea 3: Co-ordinated Views

Our third idea uses coordinated views, where we show different views. Interaction on any of these views changes the other.

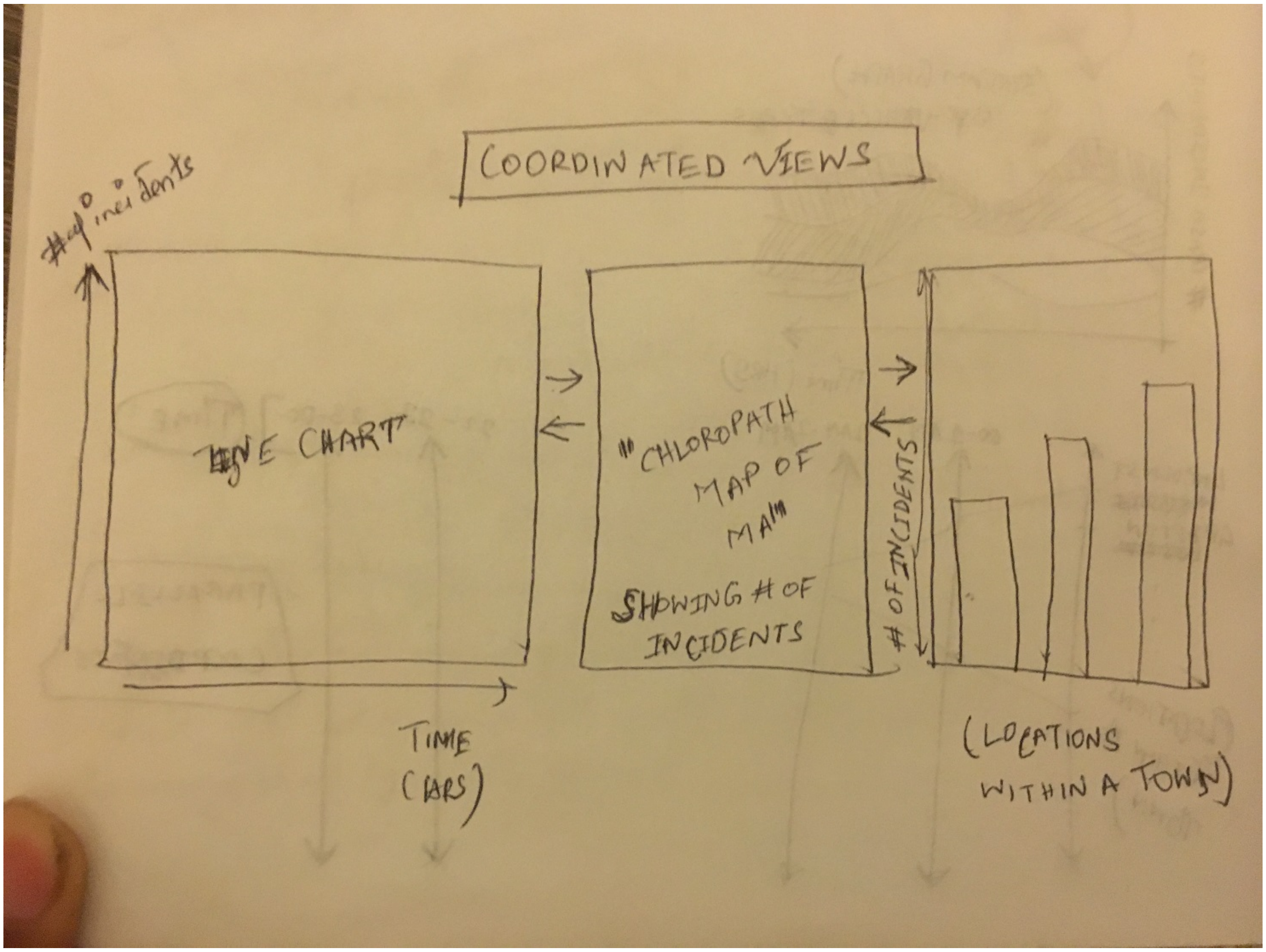

Figure 6 Idea #3 - Coordinated Views

Prototyping

For developing the prototype we initially developed the charts in the Single Page Visualization idea #1.

The following are the screenshots for Maps, ThemeRiver and Parallel Plot.

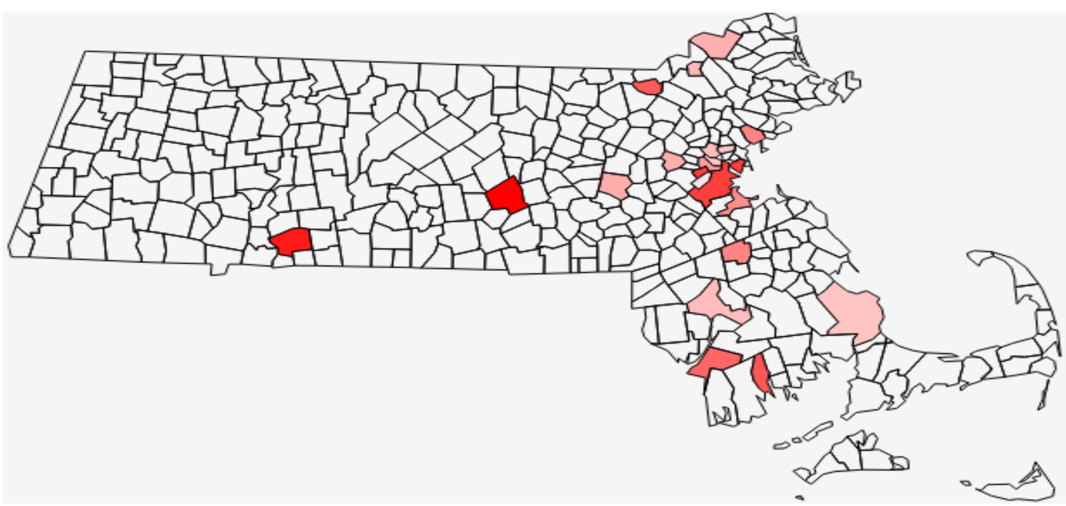

Figure 7 Massachusetts map - color represents number of accidents in that city/town

In the above figure, we can see the top 20 towns that we have considered for our project. A brighter shade of red indicates more number of crashes.

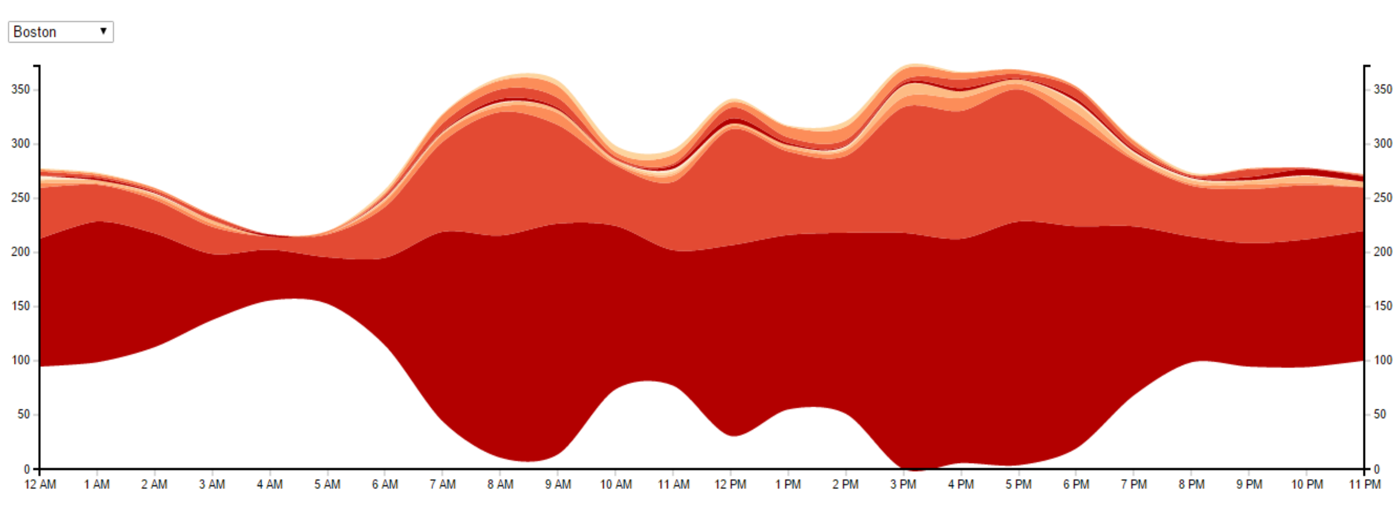

Figure 8 Theme-River

The above figure depicts a Theme-River for the different types of vehicles and the total number of crashes they have been involved in. The different shades of red denote the various types of vehicles and the thickness is the cumulative frequency of the crashes.

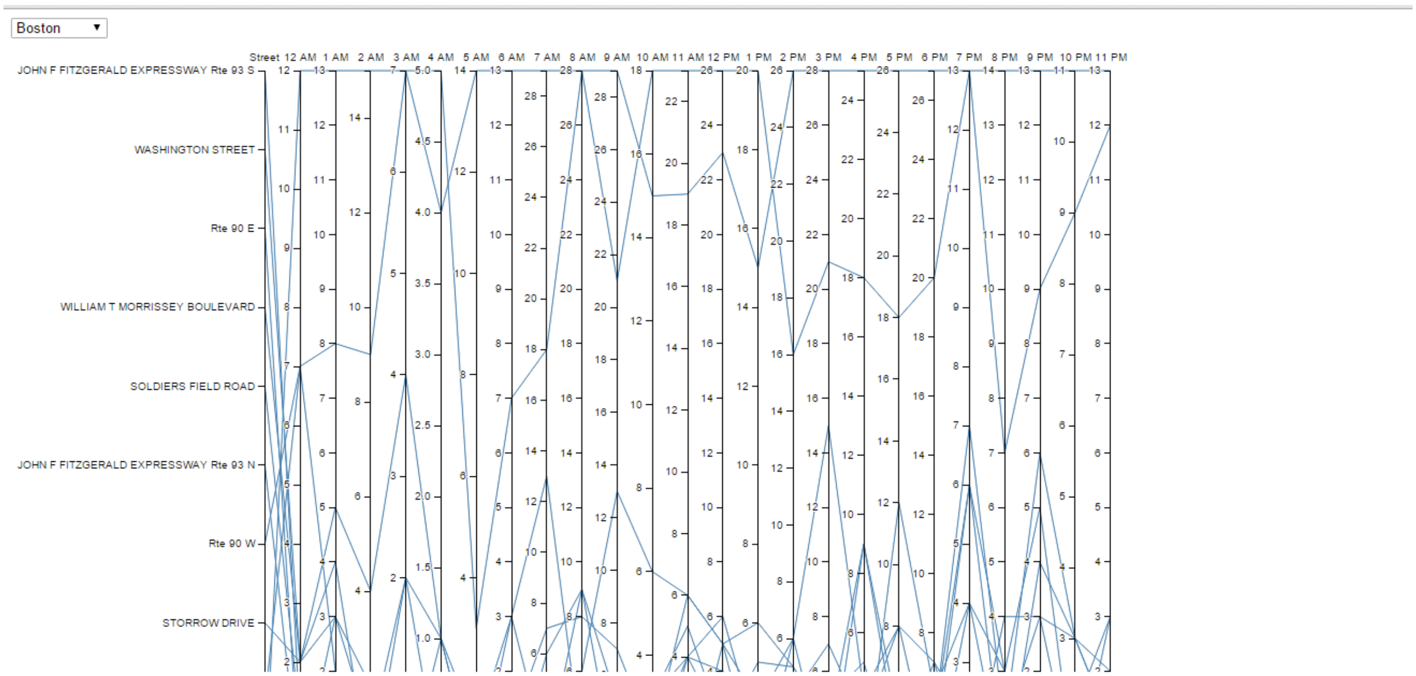

Figure 9 Parallel Plot

The above parallel plot shows the total number of crashes reported in the top 10 locations of Boston grouped into time intervals of 1 hour each. The vertical axes represent time intervals and the horizontal lines denote the locations. The distance from the bottom to a point on the vertical axis represents the number of crashes at that time interval for a given location.

We received some feedback based on the visualizations that we submitted as a part of the initial prototype.

Based on the feedback received, we decided to kill the Map and do a simple bar chart instead. Since we were considering only the top 20 towns, showing them on the map wouldn't make much of a difference than a bar chart. Also, geographical location is not as important as the name. The bar chart would do a better job in this case.

We also decided to eliminate the parallel plot. For starters, it looked a bit confusing and also was very crowded on the screen. We tried making it better by making the lines larger, changing colors to make patterns more apparent. This was a bit redundant to the Theme-river, hence we decided to take it off completely.

In place of a parallel plot, we came up with a Radar Map which was much better aesthetically and also did not create any confusion in the mind of the viewer.

So, the final designs we decided upon were: **Bar Chart**, **Heat Map (Matrix), Theme-River** and **Radar Map**

Final Design

For our final designs, we zeroed in on the 4 charts mentioned above

Bar Chart

The Bar chart was an introduction to the entire visualization project. Below is a screenshot of our Bar chart.

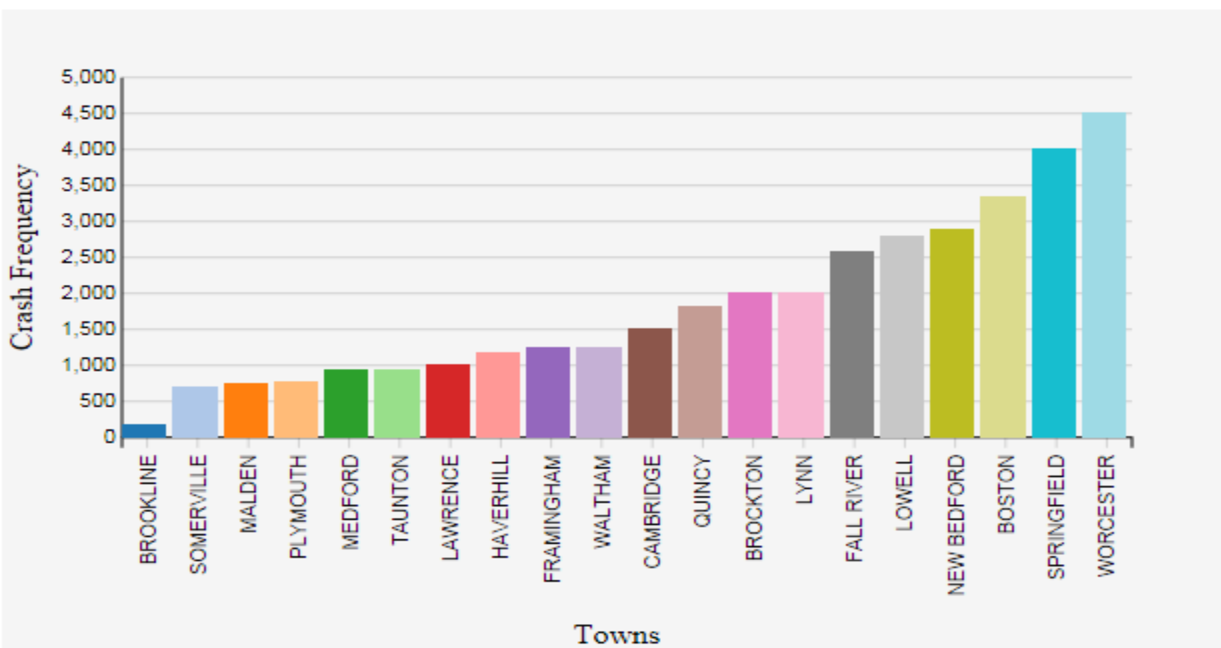

Figure 10 Bar chart

The above given bar chart is a peek into the entire dataset. It basically summarizes all the crashes that have taken place in all the towns in the year 2013.

We then decided to add more functionality to this chart. We wanted to view this information in a temporal way. We added the time and the season attribute to this visualization. We thought that the best way to view this temporally is by adding a slider for viewing it by time. The slider could be adjusted to select a time range. We added radio buttons to view the same according to the seasons.

Below is the screenshot of the improved bar chart

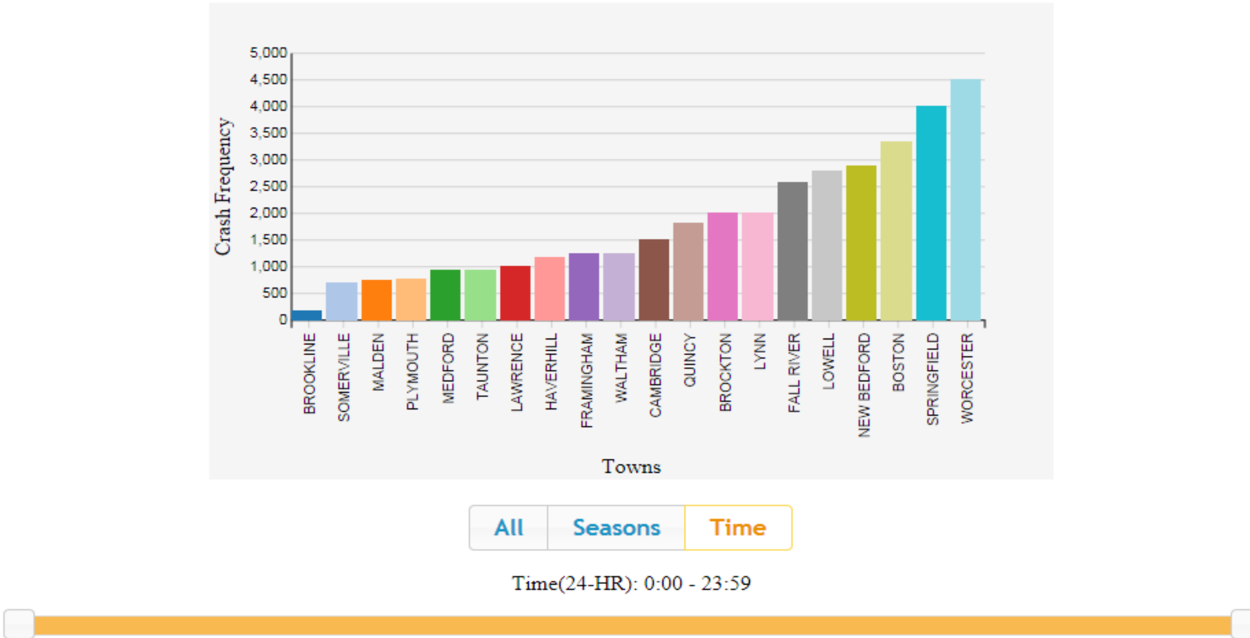

Figure 11 Bar chart with slider

We can see a slider in the above graph that allows to choose a time range by adjusting the slider. Upon clicking the Seasons tab, we are presented with a choice of 4 radio buttons that correspond to the 4 seasons: Fall, winter, spring and summer.

Heat map (Matrix)

Our next design was the heat map that denotes the Road and Light conditions at the time of crash. We can view this chart for every city and gives an understanding of how many crashes have occurred in the given scenario.

Below is a screenshot of the Heat map.

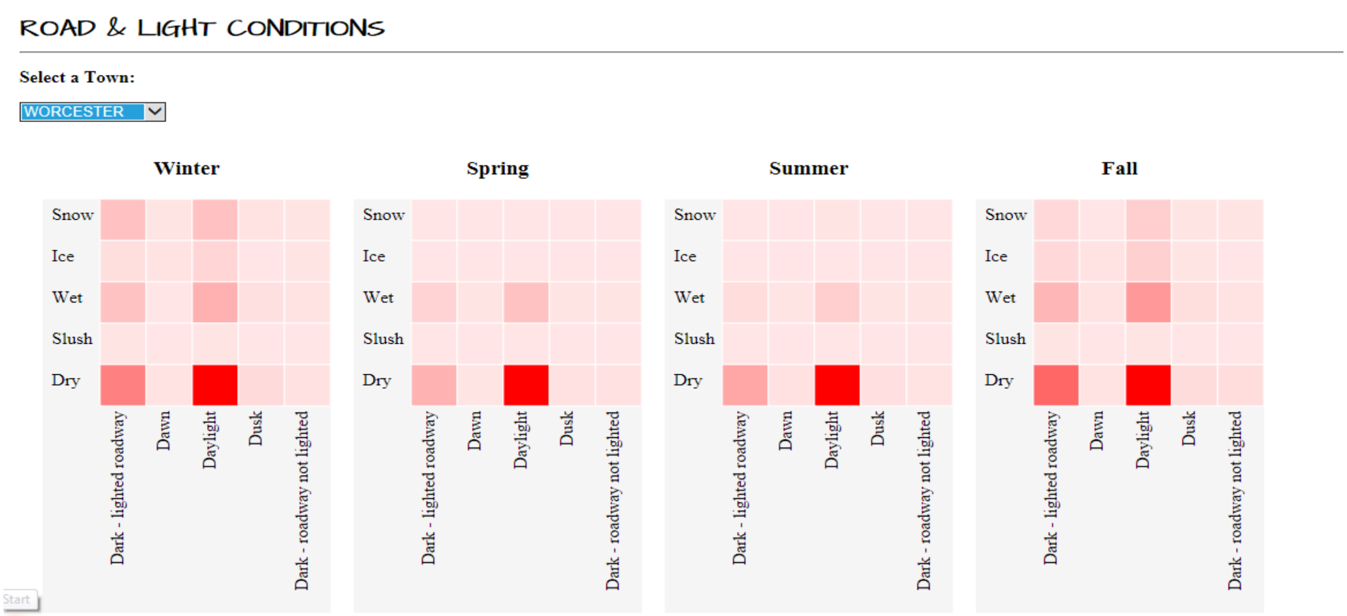

Figure 12 Heat Map (Matrix)

The above screenshot depicts the road and light conditions for the city of Worcester. The darker the square for a given road and light condition, the more number of crashes.

Theme-River

Theme-River was one of the designs we had developed ompletely at the prototype stage itself. The only thing that was left to add was the data label when we mouse over on the graph.

The below figure shows the improved Theme-River

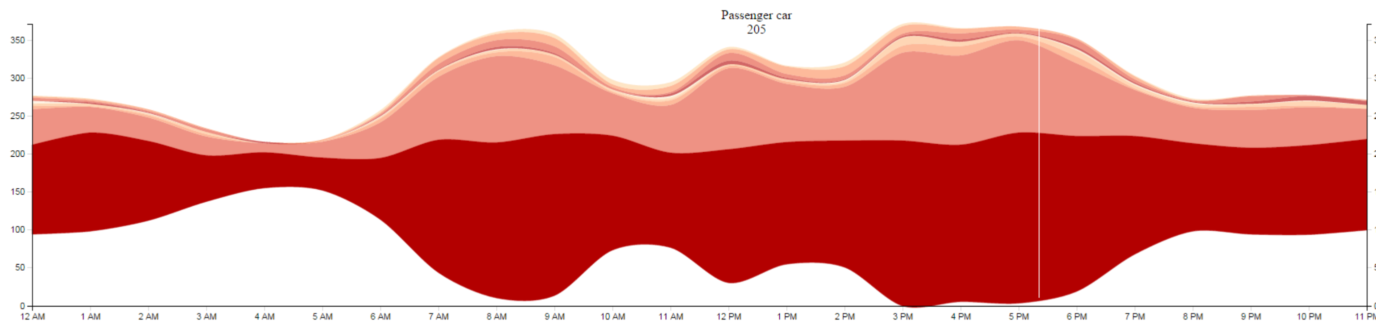

Figure 13 Theme-River

As seen in the above figure, hovering upon the chart will show the data label along with the value. The Example shows "Passenger Car, 205"

Radar Map

Our new design that we implemented as a substitute to the parallel plot was the Radar Map. Prior to finalizing the radar map, we tried to depict something different rather than the number of accidents per street in the given time range. We thought showing the direction of the vehicle at the site of the crash would be a good piece of information to show.

However, our data had many missing values for the direction of travel of the second vehicle involved in the crash. In some cases the direction of travel was the same and depicting this in a non-confusing way on a parallel plot was a challenging task. We came up with many different designs (shown below), but weren't able to show something that wouldn't cause confusion in the minds of the viewer.

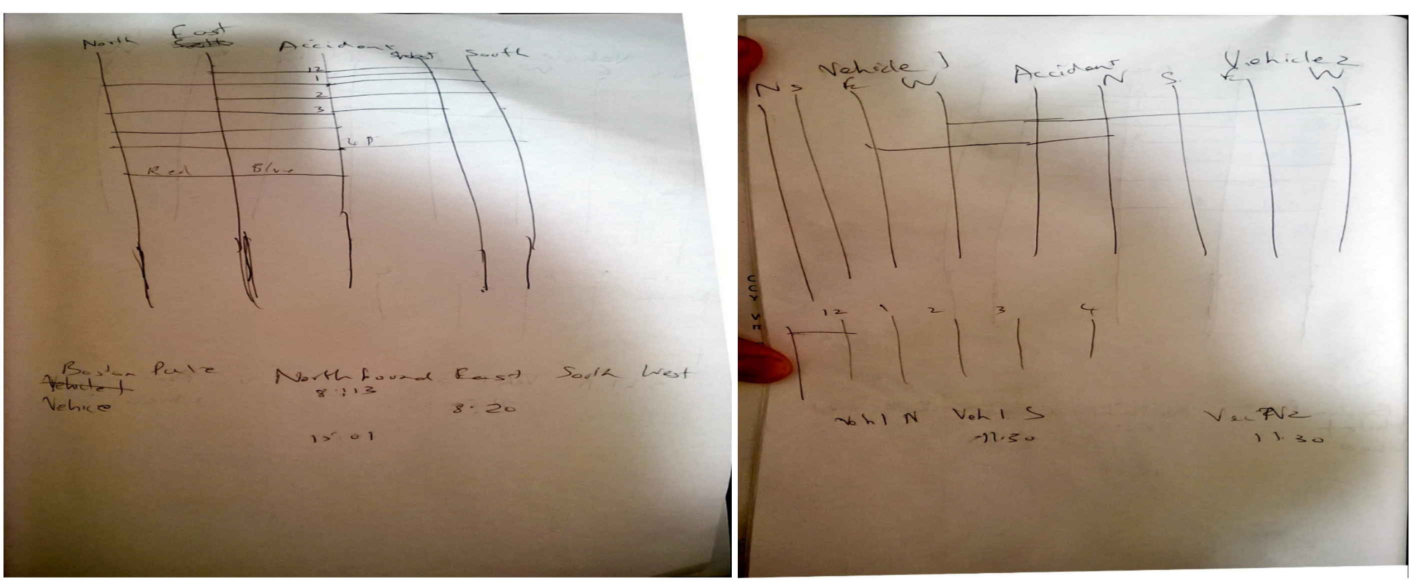

Figure 14 Modified parallel plot

So, we shunned this idea and decided on implementing a Radar Map showing the same earlier information itself, i.e. the number of accidents per street in all the time intervals.

Below is a screenshot of the Radar Map we implemented

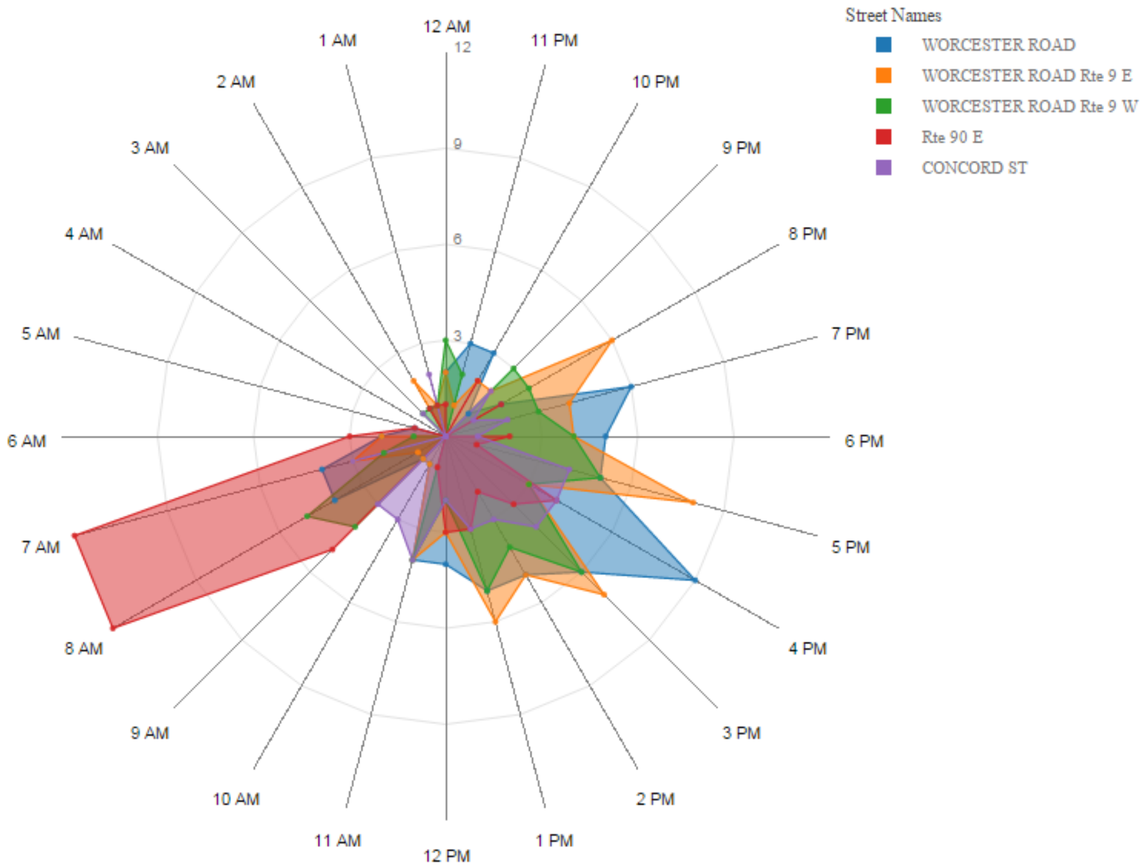

Figure 15 Radar Map

The above figure shows the Radar Map for the town of Framingham. As you can see, this visual is simpler that a confusing parallel plot.

Integration of the Designs

Once we finished developing the individual designs for all the charts, we next focused on integrating all the components together and defining interactivity between them.

This was a challenging task for us because we had developed the components individually and while integrating them, we had to ensure code consistency and correctness.

Process Flow & Interactivity

After considering a few options for the process flow, we decided on using the following design to integrate our visualizations together.

The initial screen shows the bar chart that has information on all the towns/cities – summary of the entire dataset. The control for the Season and Time range is located above the bar chart on the top. Upon changing the controls here, the bar chart gets updated.

Below is the screenshot for the same.

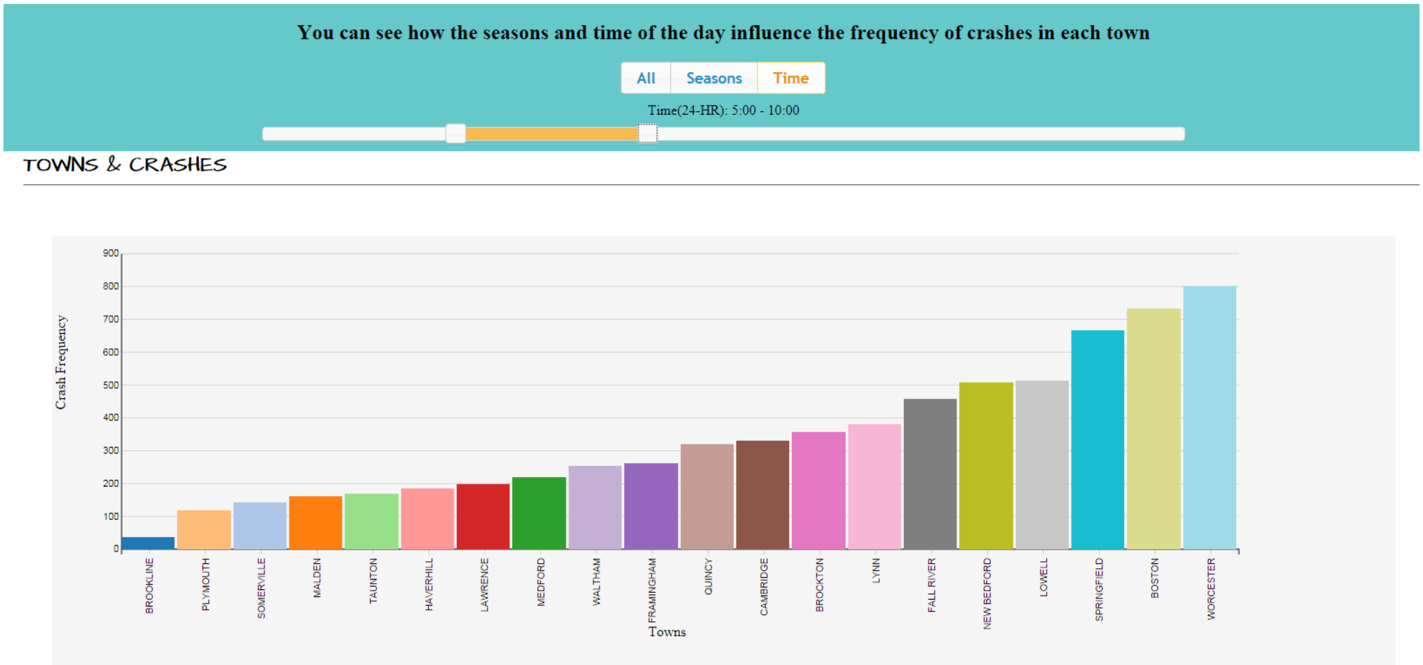

Figure 16 Bar Chart with controls

In the above chart we have shown the crash frequency in all the towns from 5:00 – 10:00. Worcester has the highest frequency.

The next chart is the Heat Map which can be viewed once the user scrolls down. Upon scrolling, the controls at the top of the screen are updated to display a dropdown menu for the city/town. The heat map is updated to show the road and light conditions according to the city/town selected

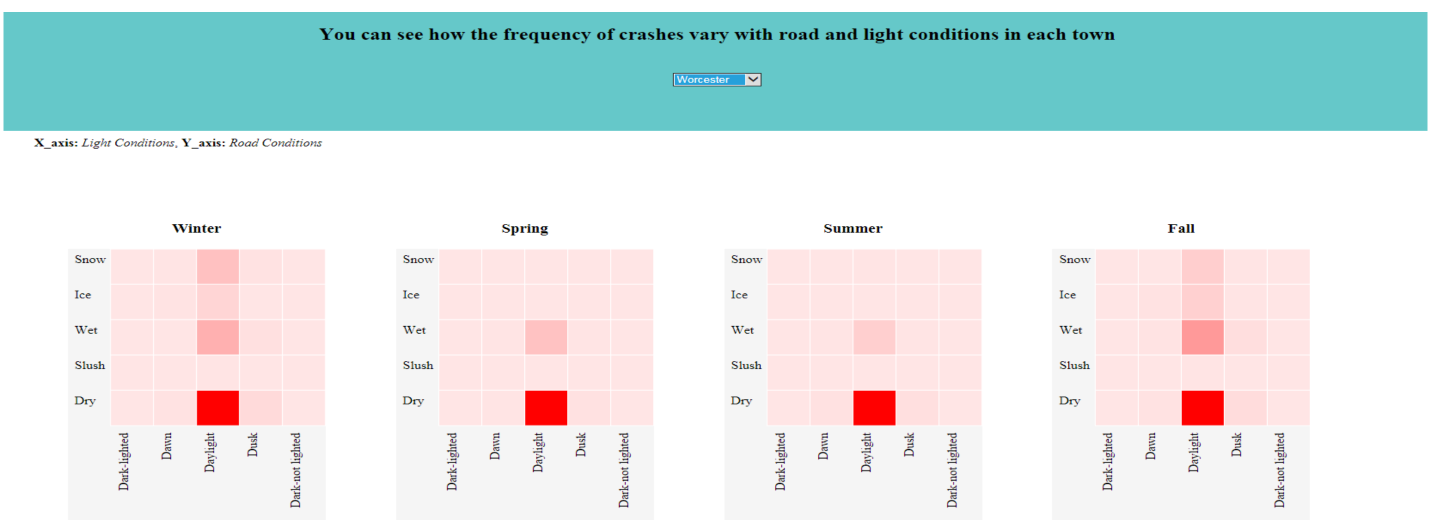

Figure 17 Heat Map for Worcester

The above chart shows information on crashes in Worcester under various road and light conditions.

Going one screen further down shows the Theme-River chart. The controls section is automatically updated once again and displays the seasons tab and the time tab.

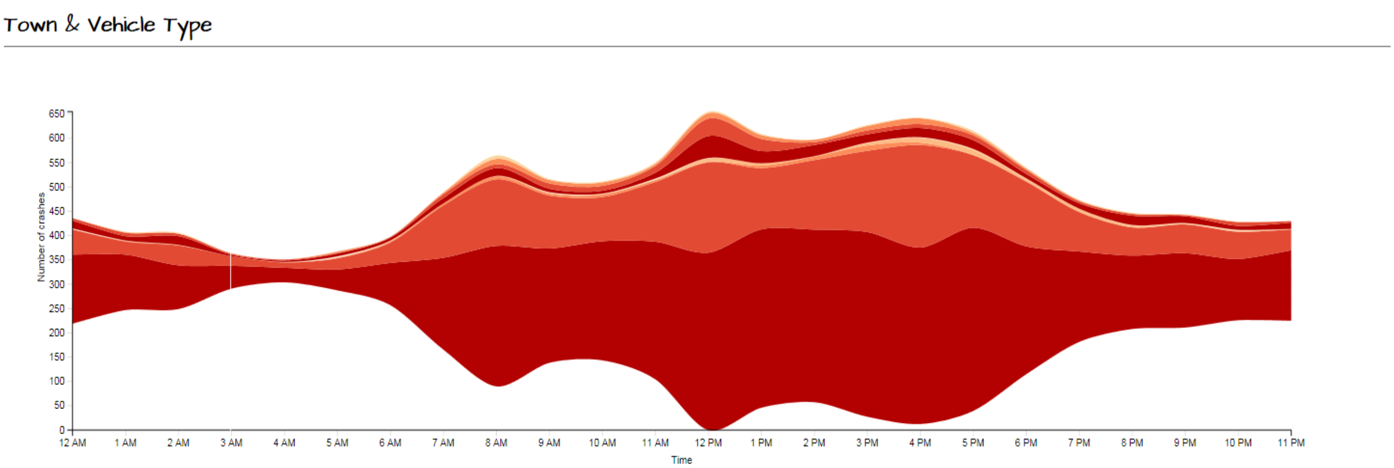

Figure 18 Theme River for Worcester

The above figure shows the theme-river chart for Worcester depicting the various type of vehicles involved in crashes throughout the day.

The last screen displays the Radar Map. The controls section remains the same as Theme-River chart for this.

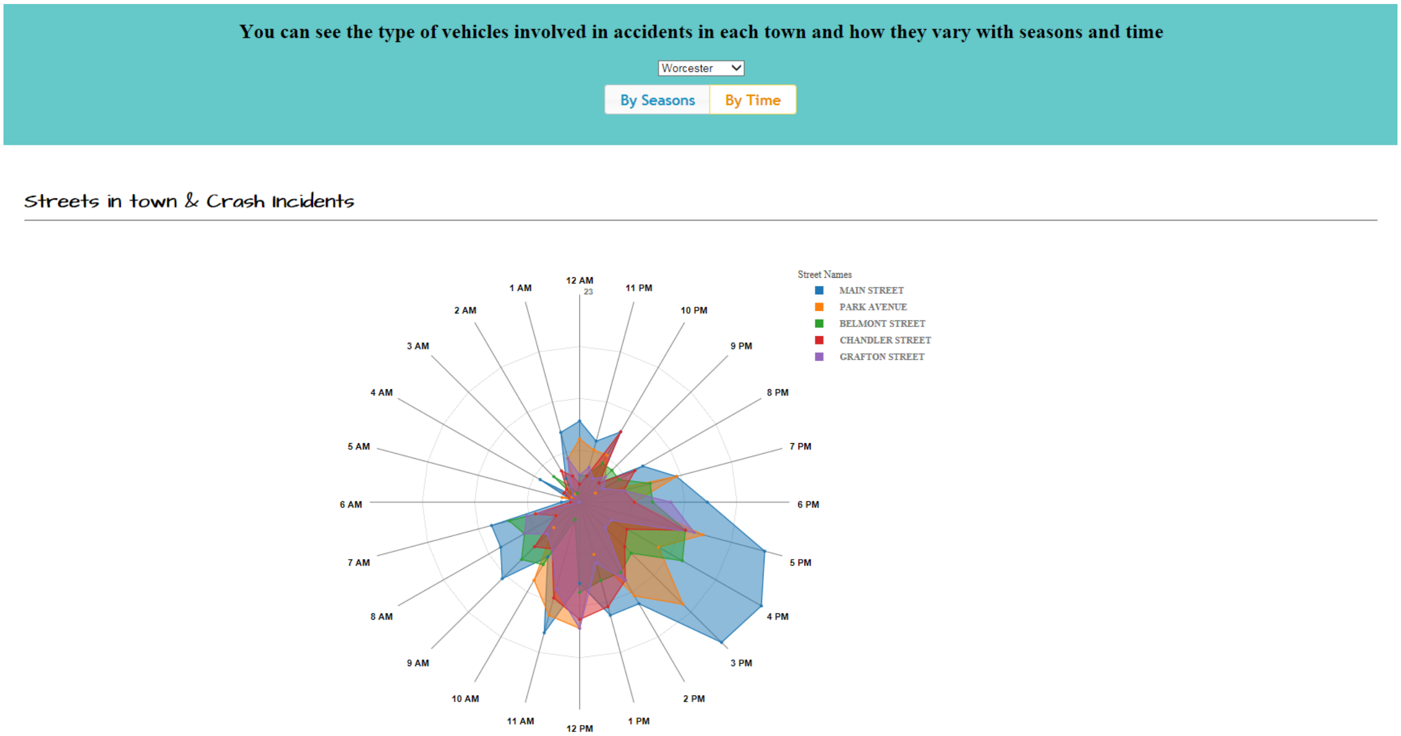

Figure 19 Radar Map for Worcester

The above chart shows the radar map for the city of Worcester depicting the crashes in the top 5 streets throughout the day.

Implementation

We have used D3.js, jQuery, and Underscore.js libraries to create our individual visualizations. For the bar chart, we choose to visually encode each bar with Hue to make them distinct from each other. JQuery UI controls were used to implement segmented button and range slider.

We have used a sequence of four heat maps to let the user explore the crash incidents under different road and lighting conditions, for every town and by seasons. The cells are visually encoded using color saturation. With it, the user can figure out trends in the crash incidents from winter to fall (assuming winter is the beginning of the year).

While developing the theme-river and radar map, we had to convert the csv data to json format so that it works correctly and the visuals are properly rendered.

For the controls section, we used Jquery.viewport to handle and detect the changes on scroll. All views were created with dynamic width that will be determined from the resolution of the display screen so that visualizations do not appear pixelated or out of place.

We also ran into multiple css issues while trying to integrate the individual components and spent quite some time to resolve them.

Experimental Observations

We were trying a few things and looking to uncover some hidden patterns in the data. Upon initial exploration, we observed that for almost all the season and for most of the times of the day, the crash frequency in **Worcester** was the highest. We decided to investigate it further and rendered all the remaining visualizations for Worcester.

As shown in Figure 17, we can see that for almost all the seasons, the highest crash frequencies were recorded in daylight and in dry conditions. This is an interesting observation. It is sort of counter-intuitive because one would think most accidents happen when the conditions are not favorable for driving.

Also, in Figure 18, we can see the Theme River for Worcester during the time of day. Passenger cars seem to have the highest frequency of crashes especially during 4 PM – 6 PM. This seems to be normal since this is the time most commuters usually get back home and the traffic density would be higher.

Figure 19, shows the Radar Map for the top 5 locations of Worcester. Main Street seems to be the worst during 3PM – 5PM during the afternoon with major portion of crashes reported there.

We also took a look at the same graph for Boston

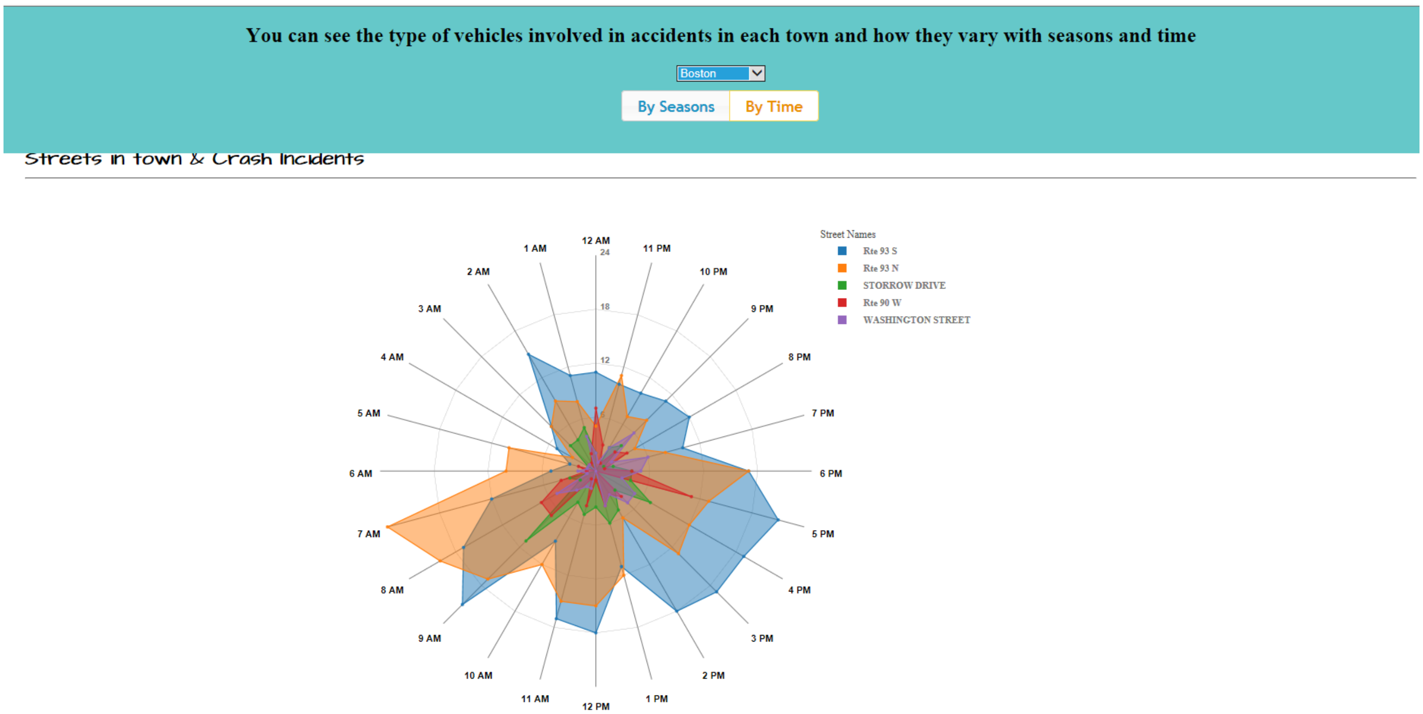

Figure 20 Radar Map for Boston

We can see an interesting pattern in the above graph for Boston. The I-93 N reports more accidents in the morning hours and the I-93 S reports more accidents in the evening hours. This suggests that a major chunk of the traffic heads north from the city for work etc. And hence more traffic density. This could additionally help the city administration to mobilize more resources and move towards reducing accidents in the area.

Conclusion

Our set of 4 different visuals depict a nice hierarchical flow from high level view to a lower level. The Bar chart gives the summary of the dataset that can be categorized by season as well as time. The next visualizations were helpful in drilling down the data by every city against the important set of data such as road and light conditions, location and the different types of vehicles involved in all the incidents. This visualization charts helped us uncover some hidden patterns that were counter-intuitive (such as more accidents in daylight and dry conditions) and also validate some notions such as more traffic density can be a cause to more number of crashes.