Title:

Visualization vehicle crash incidents in Massachusetts

Team:

Snehasish Barman (*Email*: sbarman@wpi.edu, *Github*: sbarman-mi9)
Sounthar M (*Email*: smanickavasagam@wpi.edu, *Github*: msounthar)
Sachin Sudarshana (*Email*: ssudarshana@wpi.edu, *Github*: sachin019)

Project repo:

https://github.com/sbarman-mi9/CS573FinalViz.git

Background & Motivation:

We had worked on the New York Motor Vehicle Collisions for a previous assignment. 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 are going to answer the following questions,

What are the factors which influence these collisions?
Which type of vehicles are involved in an accident and is there a pattern?
When do these collisions happen the most?
Where do the vehicles were travelling and at location the crash happened?

Data:

The dataset we plan to work with 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).
The crash data for every town can be requested using this form:
http://services.massdot.state.ma.us/crashportal/DataRequest.aspx

We collected the data for the year 2013 of all the towns in Massachusetts. The data is the form of separate excel files for all the different towns and cities.

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

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. Ex: Fatal, Non-Fatal etc.
6. **No. of Vehicles:** Number of vehicles involved in the crash
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. Ex: Single vehicle, Angle, Head-on etc.
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. Ex: Passenger car, Truck etc.
14. **Road surface condition:** Condition of the road
15. **Ambient Light:** Lighting at the scene of crash based on time of day. Ex: Daylight, dark - lighted driveway etc.
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. Ex: Cyclist, Pedestrian etc.
23. **X Coordinate:** Location encoded by GIS software.
24. **Y Coordinate:** Location encoded by GIS software.

Data Processing:

For our visualization we primarily need to perform 2 steps for Data processing:

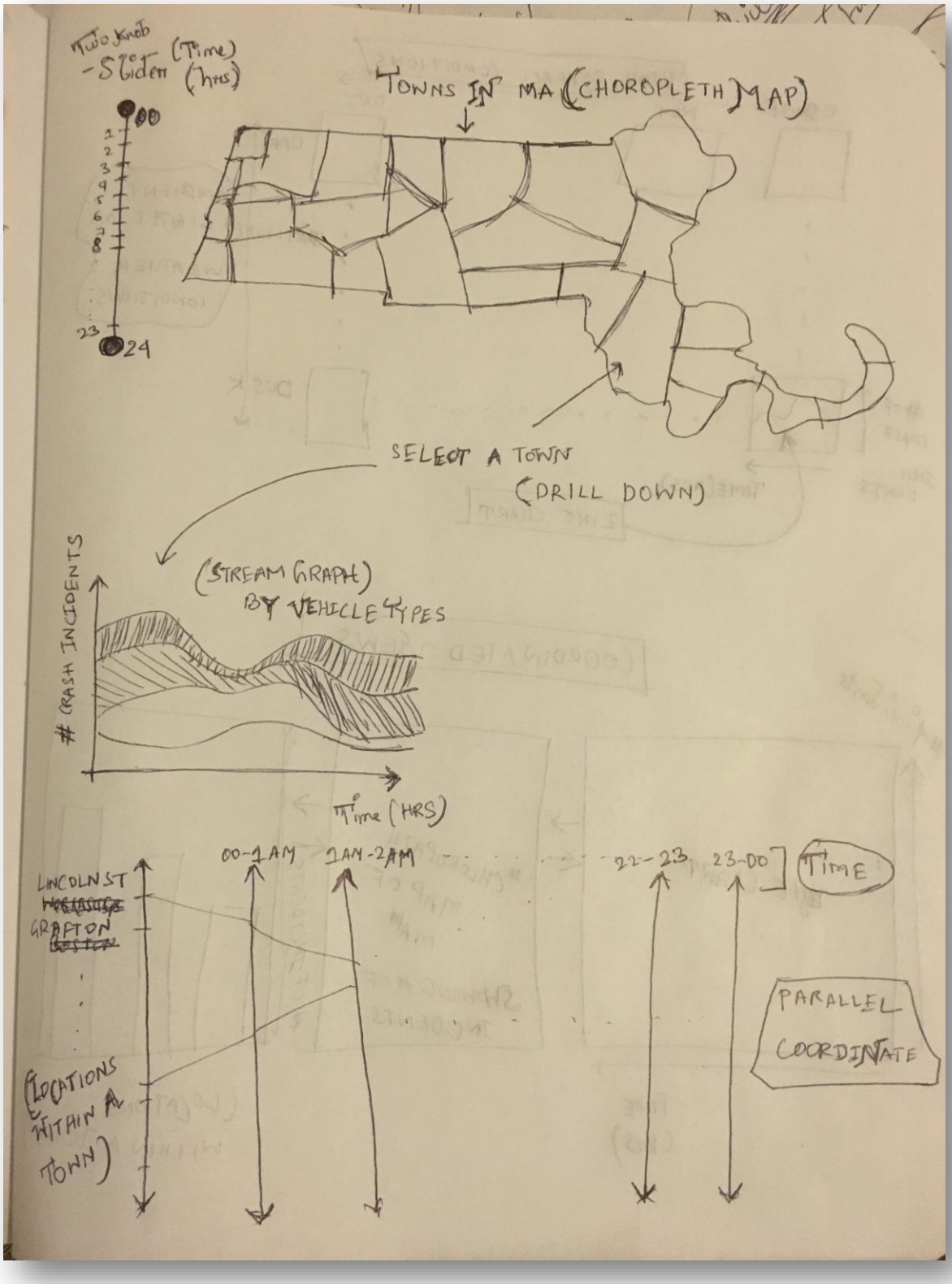
1. Since the datasets are obtained individually for all the towns and cities, we need to merge all the files into a single master dataset.
2. All the non-required columns in the master dataset has to be deleted. Attributes such as 'At roadway intersection', 'distance from nearest mile marker' etc. is purged from the dataset.

In addition to these steps, we will also perform treatment of missing values in columns that are required for generating our visuals.

Visualization Design:

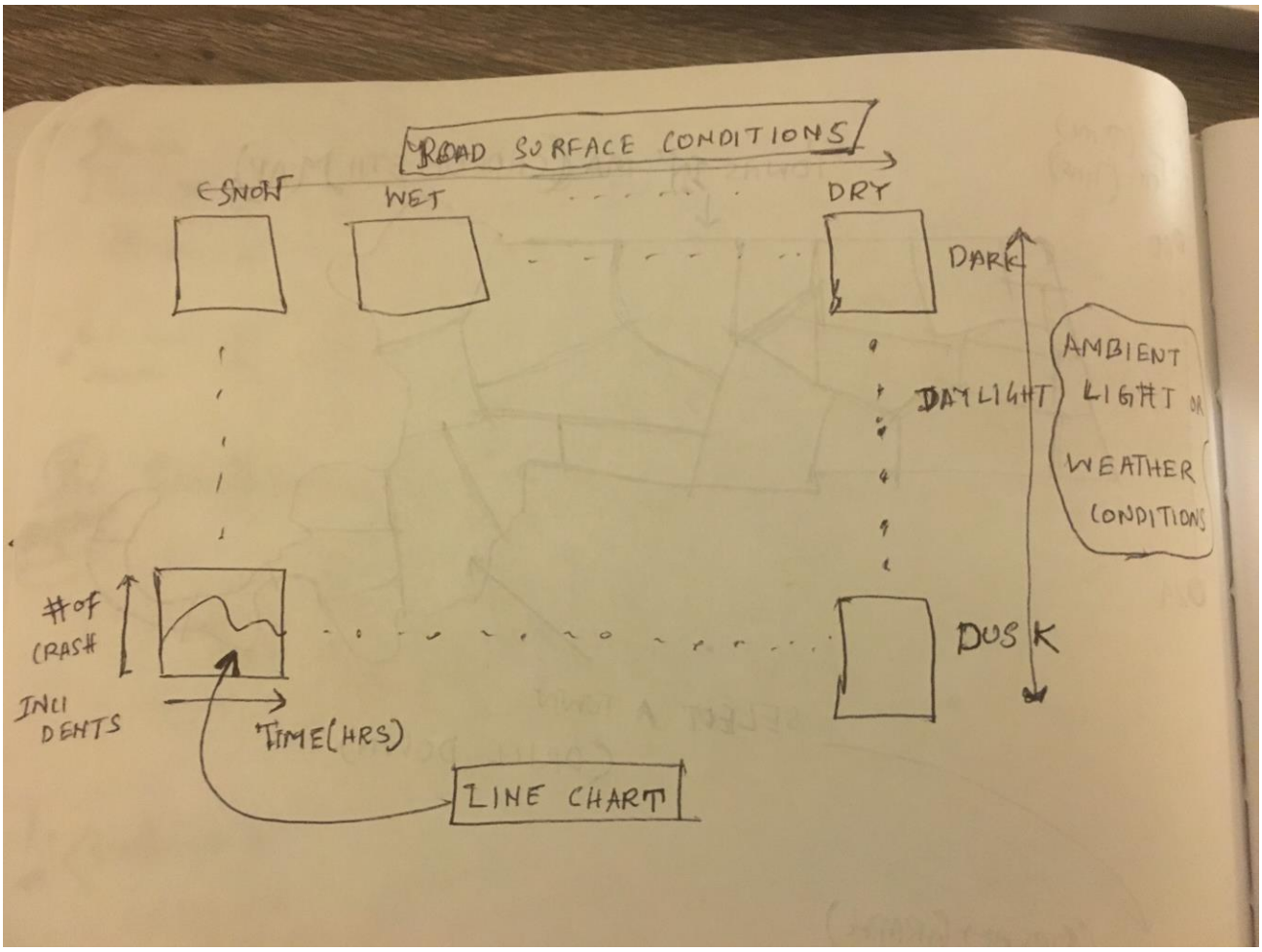
We plan to do analysis on few variables: time, factor/causes of crash, conditions, vehicle types, injuries, distance from and to, and others as per time availability.

Idea #1: Single Page Visualizations


Two knob
- Slider (Time)
(hrs)
00
1
2
3
4
5
6
7
8
23
24
TOWNS IN MA (CHOROPLETH) MAP
SELECT A TOWN
(DRILL DOWN)
(STREAM GRAPH)
BY VEHICLE TYPES
CRASH INCIDENTS
Time (HRS)
00-1AM
1AM-2AM
22-23
23-00
Time
LINCOLN ST
GRAFTON
LOCATIONS WITHIN A TOWN)
PARALLEL
COORDINATE

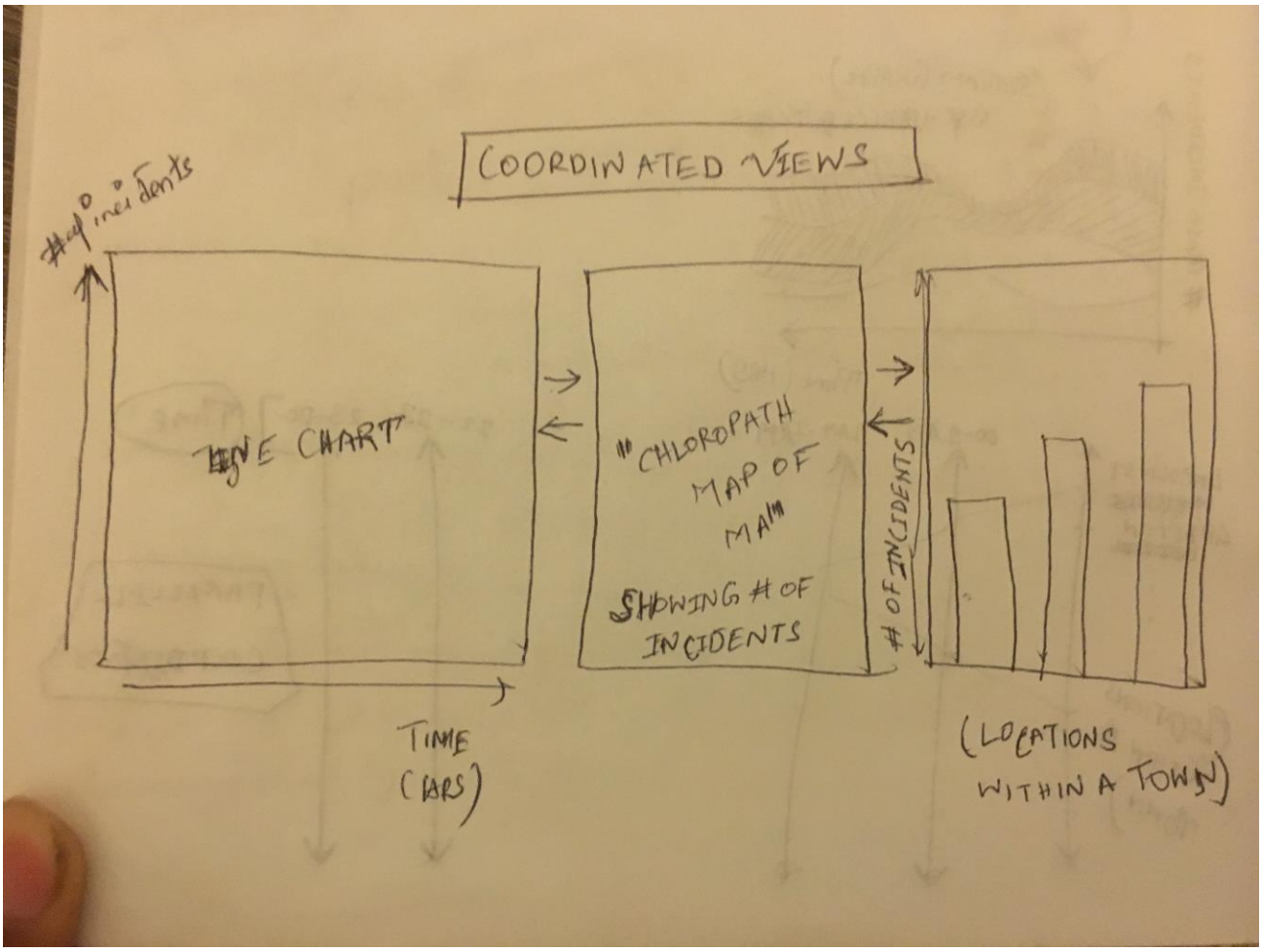
Here, each visualization will be displayed on separate pages. On the first page, we show a map of all the towns of Massachusetts, the color of which can be changed by moving the "range slider" to different ranges of time in a day. On selection of a town, we drill down to the details of individual towns 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).

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...

Idea #3: Co-ordinated Views



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

Final Idea: Our final idea might be a combination of all three above.

Must-Have Features:

Map interactivity, matrix analysis, analysis by different factors using appropriate visualization paradigm. We plan to set up a server and database for this application.

Optional Features:

Given the time permits, we may integrate NYC motor incidents dataset to show correlations between the two datasets among different conditions/causes of crashes and among different times of the day.

Project Schedule:

Week No.	Dates	Task
1.	Nov. 18 - Nov. 24	Data exploration and processing
2.	Nov. 24 - Dec. 1	Design and Development
3.	Dec. 1 - Dec. 8	Development (contd)
4.	Dec. 8 - Dec. 15	Testing and Documentation