

CS 171 - Visualization

Crashcaster

Staying Safe in Cambridge

Project Plan

Team HarvardViz:

Kyle Maguire

Rick Farmer

Kartik Trasi

Adam DiChiara

Contents

1. Overview and Motivation
2. Second iteration of sketches
3. Description of team roles
4. Feature list
5. Story board for interaction
6. Webpage layout/storytelling
7. Project timeline and Milestones
8. Complete dataset

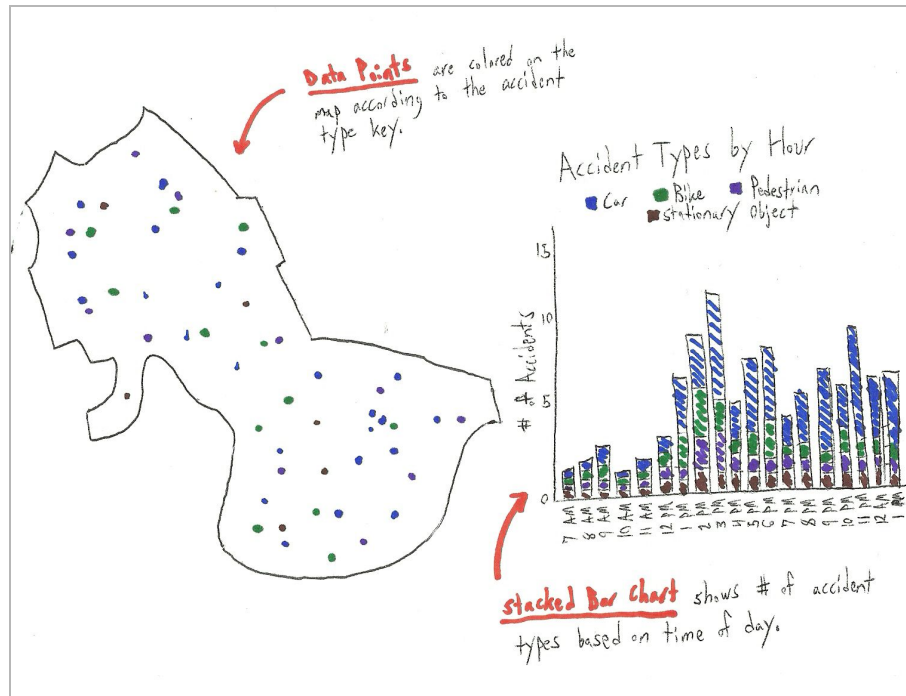
Overview and Motivation

Public safety is an important concern for residents and commuters in the Boston metro area, but the information they rely on to avoid safety risks may be lacking. In particular, traffic safety issues are common and frequent, yet with the exception of severe accidents the information the general public has regarding traffic safety is mostly based on personal judgements from anecdotal or personal experiences. In general, any information provided fails to inform the public in a way that represents systemic problems or complex inter-related traffic dynamics at play. Our project aims to develop insightful visualizations representing public traffic safety data in a way which reveals probable "cause and effect" relationships between influencing factors and incident occurrences. Our goal is to employ good data visualization design principles and effective storytelling to create an end product that is both informative and visually appealing.

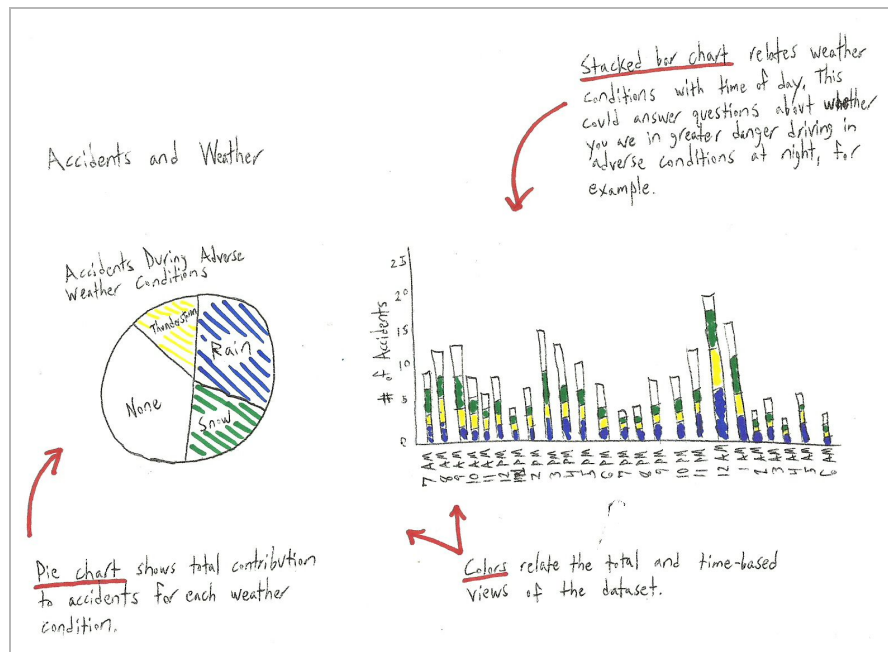
We plan on exploring datasets on city traffic data provided by the cities of Cambridge and/or Boston, MA. Our identified datasets include geographic location, incident type, date, time of day, and other related details which allow filtering, faceting, and sorting with various methods. This data can also be merged with other available datasets on traffic density and potentially weather. Additionally, we plan to classify and represent geo-areas into safe and potentially unsafe zones based on risk factors, which would give residents a guide to reroute their travel via safer paths. We were attracted to these datasets due to their completeness and readiness for analysis, which will allow us to focus our efforts on developing effective visualizations. Also, the cities of Cambridge and Boston provide a significant number of other related datasets with information that could potentially enrich our visualizations in meaningful ways. We think there's a lot to be explored in the data.

Second iteration of Sketches

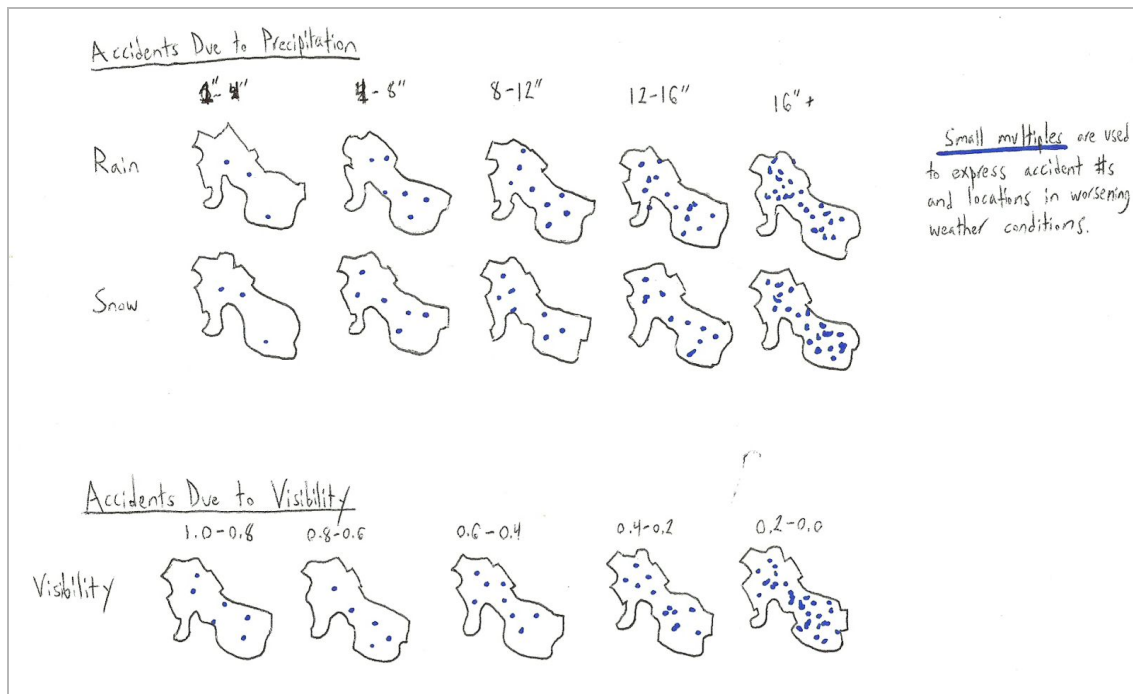
Here are the second iteration of the sketches. There is a common theme of using geo-mapping and small multiples or trends with crossfiltering.



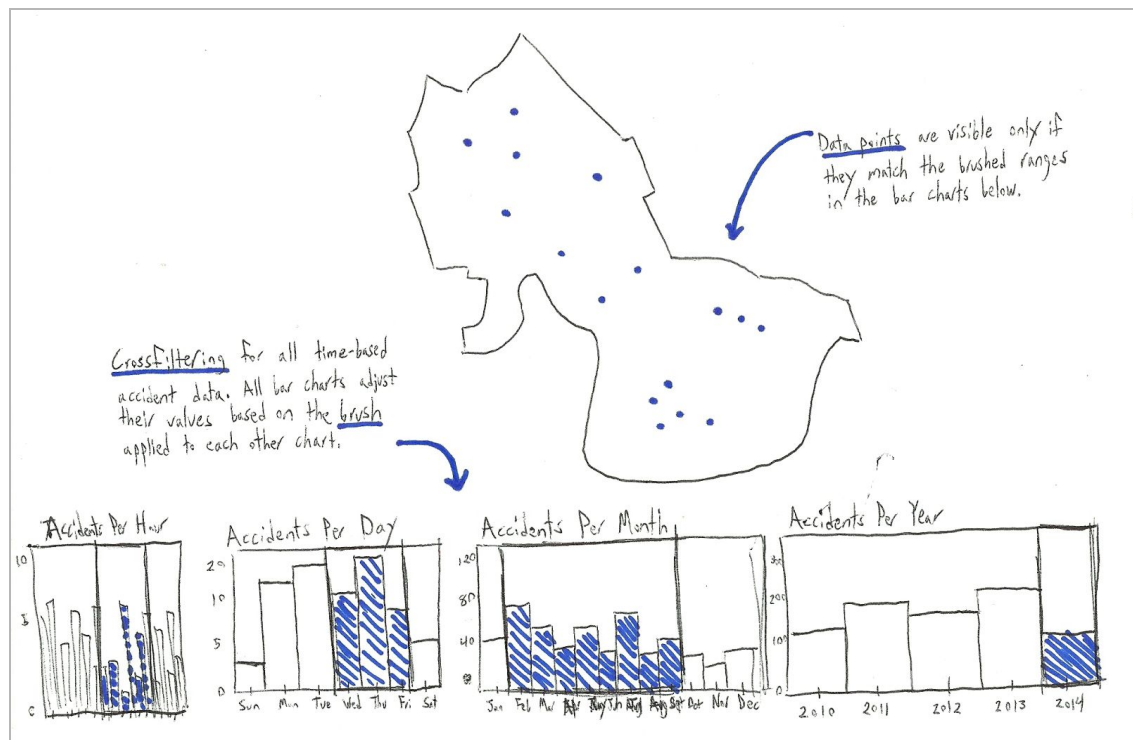
Sketch 1: Accidents expressed by accident type, and by time of day.



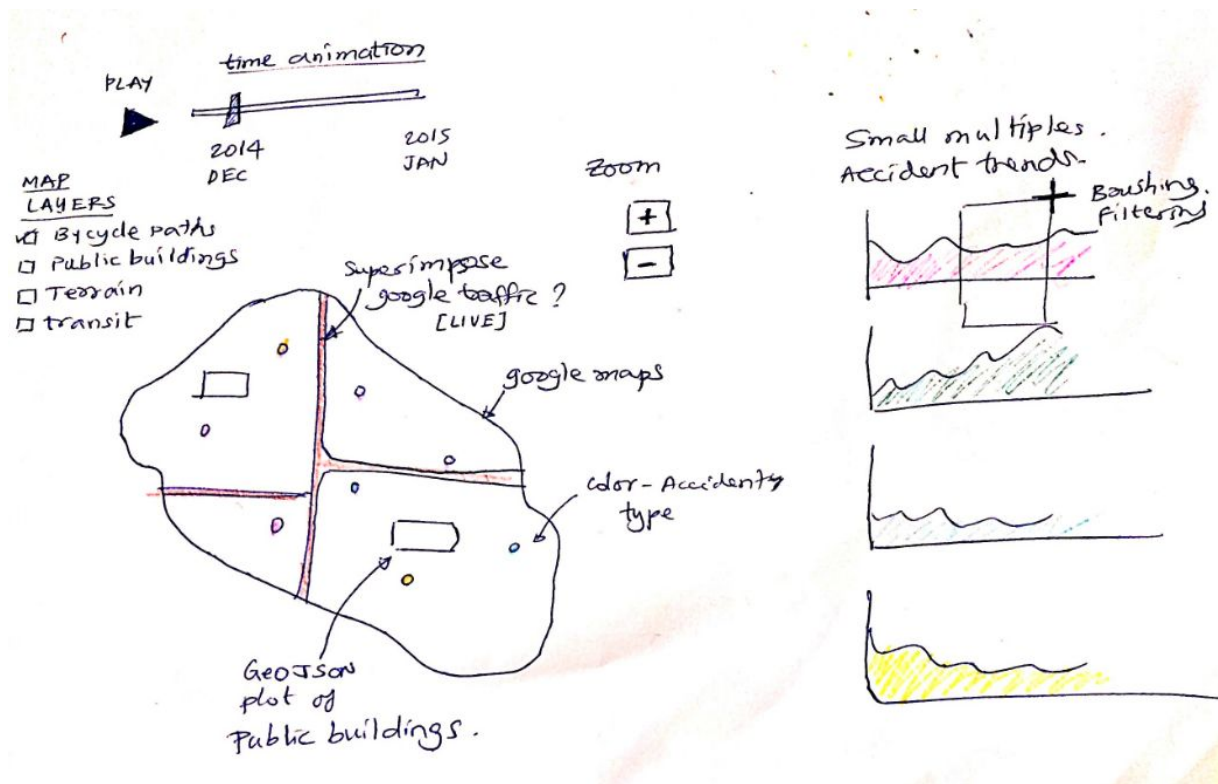
Sketch 2: Weather contribution to accidents, both overall and based on time of day.



Sketch 3. Small multiples to express adverse weather conditions and their contribution to accidents at increasing levels of severity.

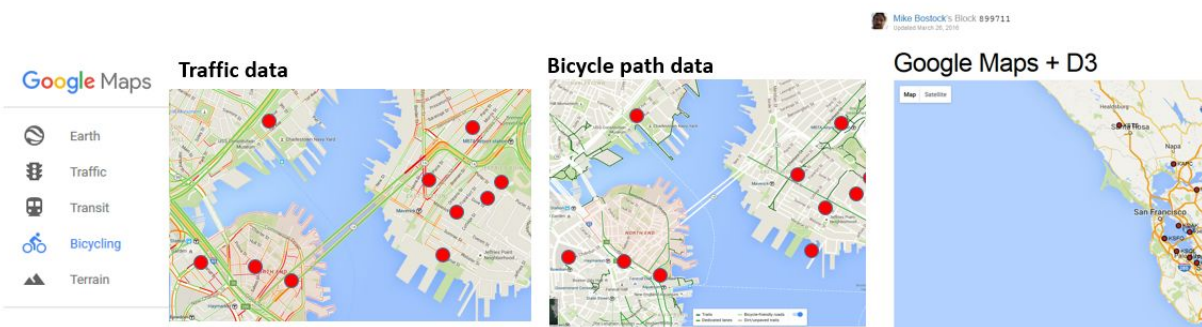


Sketch 4. Extensive time-based cross-filtering to display accidents in Cambridge.



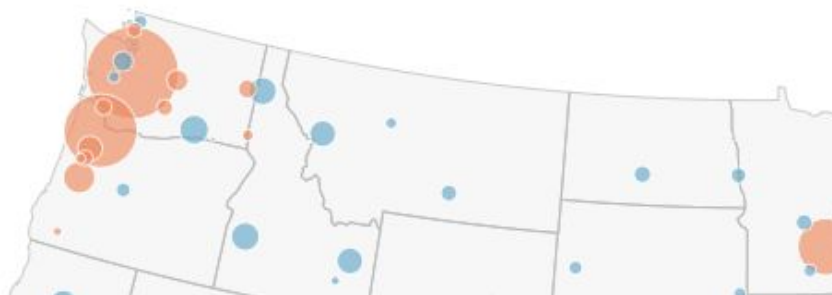
Sketch 5 - Geomaps, cross filtering, brushing, google maps / or equivalent (layers: traffic, terrain, transit, bicycle paths), public buildings

1. Google map integration will allow flexibility to use additional google map features
2. Live traffic data superimposed with our data will be interesting viz
3. Adding bicycle paths, terrain, transit... could be interesting as well
4. <http://nyc.crashmapper.com/11/8/14/2/standard/collisions/2/11/40.711/-73.772>



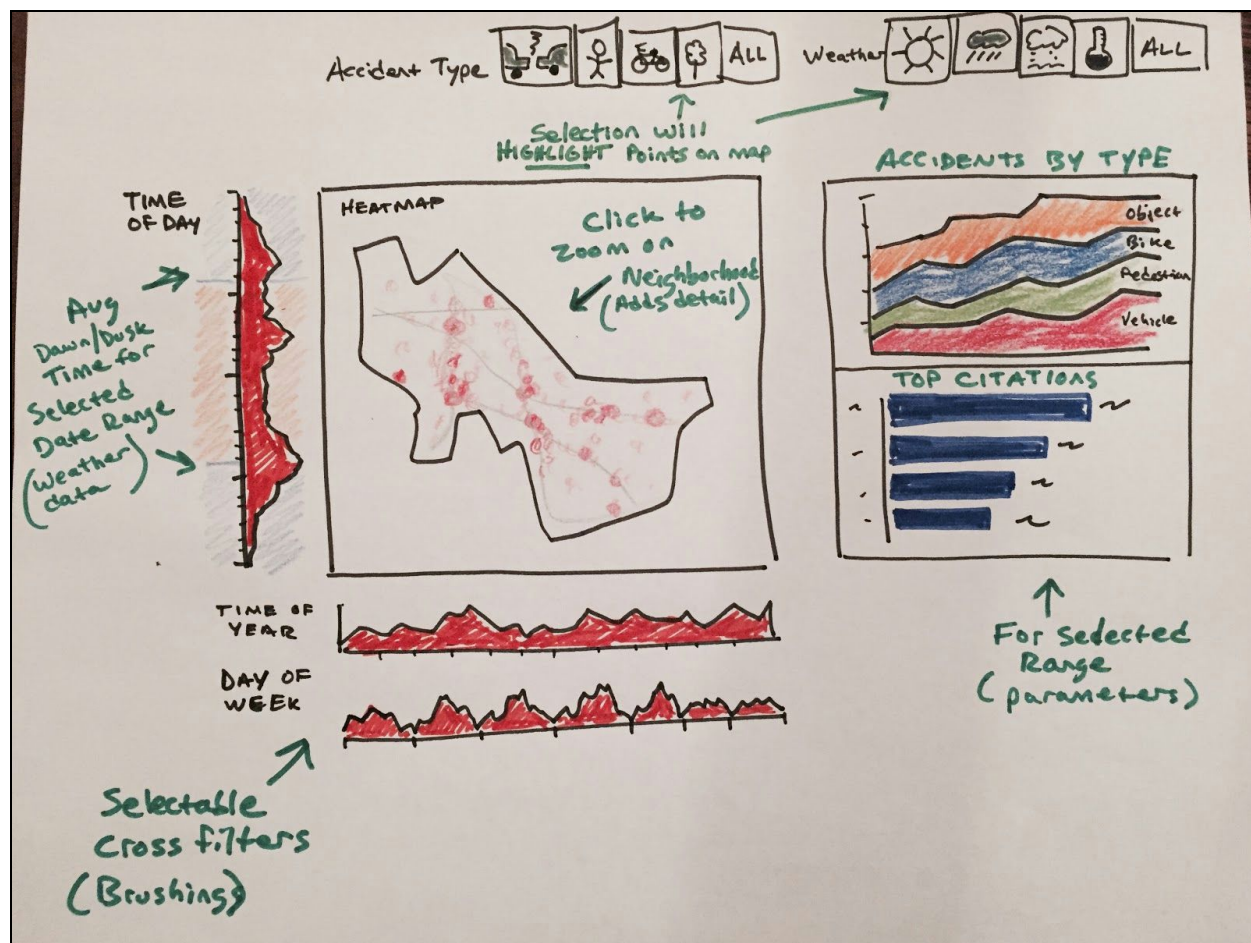
Sketch 5A: - Layers showing traffic, terrain, transit, bicycle paths...

Link to D3 sample code: <http://bl.ocks.org/mbostock/899711>

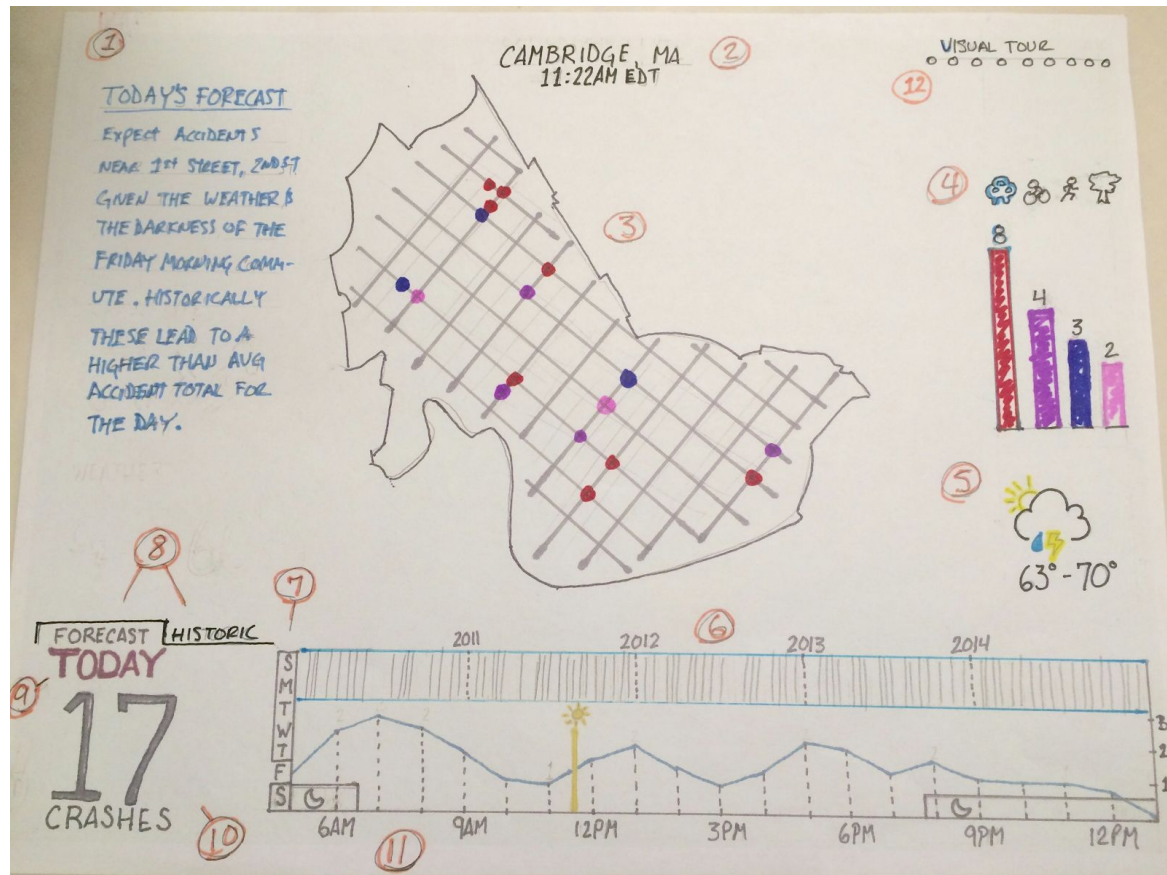


Sketch 5B: Animation through time will give the appearance of the data being dynamic

Link to example: <http://tipstrategies.com/geography-of-jobs/>



Sketch 6: main display multiple linked visualizations with multiple crossfilters.



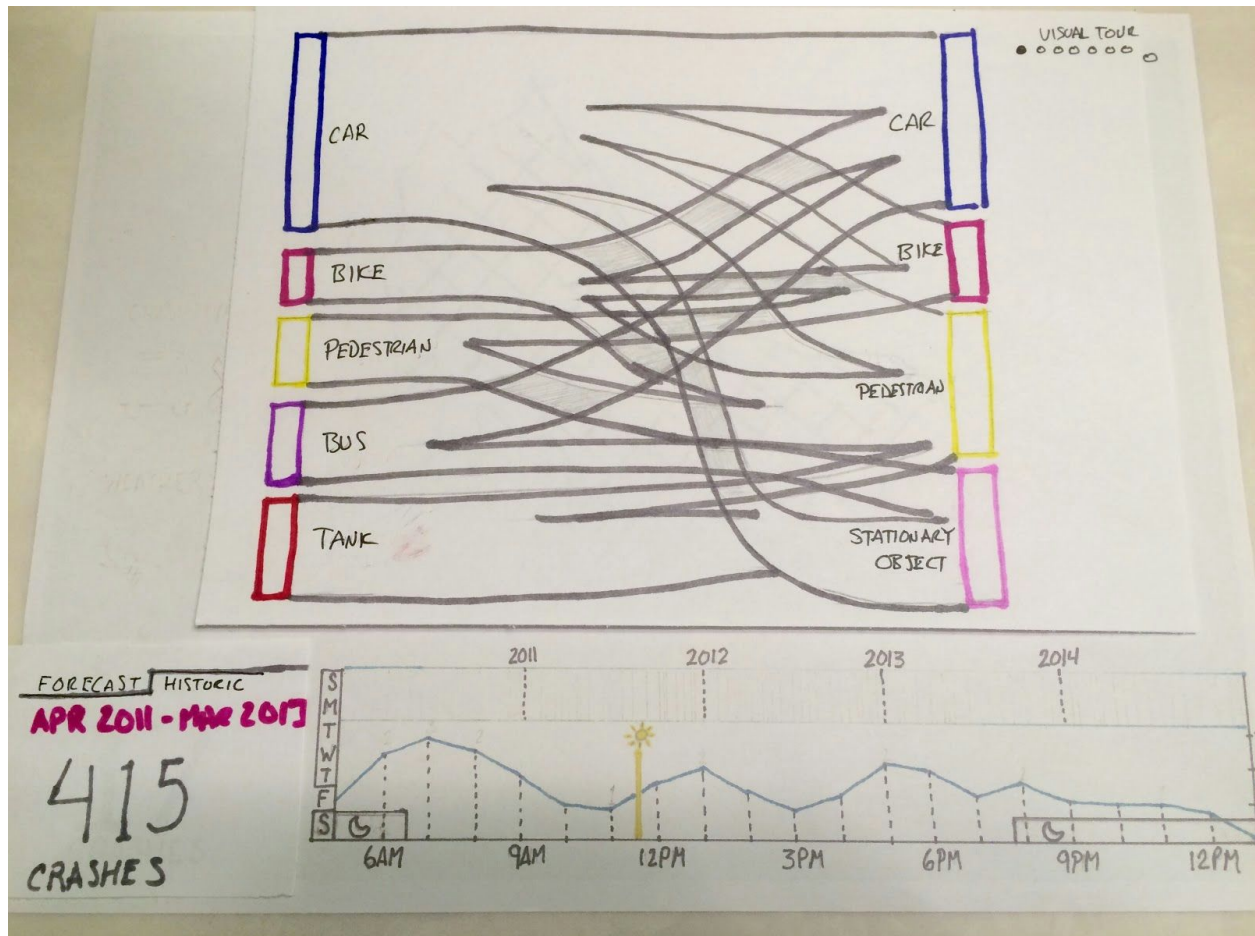
Sketch 6A: Crash Forecast (Crashcast?) landing page. This would make the project useful everyday for a commuter in Cambridge. Additionally, the user may explore historic crash data and could take a visual tour of the story-based data about crashes in Cambridge using the excellent visual design the rest of the team has suggested above.

Using the circled numbers above:

1. Today's crash forecast. Auto-generated forecast text that takes into account the various dimensions of Cambridge data (week of year, day of week, weather conditions, etc.) to make some generalized prediction about intersections to avoid based on the average expected total of accidents for a similar day.
2. City, State, current time
3. Map that shows the various types of accidents at the intersections that have previously had them. Clicking on a colored dot will give a hover tip that describes the previous accidents that the forecast is based upon.
4. The bars and icons are clickable filters that are linked to the map. Clicking either adds or removes the dots from the map based on crash types, for example cars, bicycles, pedestrians, stationary objects, etc.
5. Today's weather forecast that is used to help make the forecast. Clicking the icon, expands the weather condition icons out to the left and allows the selection of different weather conditions for

historic filtering. The temperature below is clickable and will allow the selection of a range of temperatures to filter by for analyzing historic data on the map.

- 6. A brush filter that allows the selection of a range of years to display in the visualization above (the map may not be present, if the user selected visual tour, there may be some other viz present that can be manipulated as well). The vertical lines show the days with crashes on them.*
- 7. A filter to allow the selection of "day of week" that works with items (6) and (11). For instance the day of week shown is Friday.*
- 8. Selection of "Forecast" or "Historic" mode. Forecast mode is shown which predicts the number of crashes today or for the week, e.g. 17 as shown in item (10). In "Historic" mode "Today's Forecast" (item 1) is hidden.*
- 9. The selected date. In "Forecast" mode this shows "Today". In "Historic" mode it shows the date range the brush filter (item 6) is set to, it changes as the brush ends are moved.*
- 10. The count of either "Forecast" crashes or in "Historic" mode the total count of crashes within the year brush range (item 6). Updates as the brush ends are moved.*
- 11. Time of day brush. In "Forecast" mode this shows the count of accident per hour throughout the day, along with the current time (see the sun icon and line at 11:22am), darkness based on the time of sunrise and sunset (see the moon icon and dark band at the bottom). In "Historic" mode the brush end allow the user to select a specific time of day range, the sun and moon icons are hidden.*
- 12. "Visual Tour" allows the selection of additional visualizations that the team has chosen for storytelling about Cambridge via the dots or a left swipe.*



Sketch 6B: Crash Forecast (Crashcast?) “Visual Tour” page. This shows what happens once a user has selected a dot under the “Visual Tour” or swiped to the left to move to a curated selection of additional visualizations that the team has chosen for storytelling about Cambridge. Everything above the the bottom filters (year range, day of week, time of day range, historic/forecast) may be changed based on the various visualization the team has determined help tell the story of traffic safety in Cambridge.

Here I’ve used a Sankey diagram to map the size of the relationships between crasher and crashee. For example code that look much closer to the final viz, please see the experimental code at <http://tributary.io/inlet/1217283ecbd2093073d4>

Description of team roles

Our project roles are determined by the work we are currently doing. Each of us contributes to assigned tasks that we agree to take on throughout the project. As a result, we all get to learn a variety of skills and engage in the process. We pair up to tackle complex tasks as needed to ensure that one person does not get stuck or overloaded. We each take responsibility to contribute to the success of our project in any role needed.

Administrative Roles are specifically assigned but Functional Roles and Domain Anchors will change from week to week, depending on what is required for current tasks.

Administrative Roles:

- Communications Coordinator (**Rick Farmer**): Responsible for managing communications and sharing platforms (Slack, GitHub, Pivotal Tracker, conference calls).
- Agenda Coordinator (**Adam DiChiara**): Proposes order of weekly tasks and initiates calls for group action. Responsible for keeping forward progress towards weekly goals.
- Documents Coordinator (**Kyle Maguire**): Responsible for coordinating assignment documents, ensuring completeness, finalizing, and submitting.
- External Liaison (**Kartik Trasi**): Responsible for coordinating with TF and/or studio on any team issues, questions, or feedback. (via Piazza)

Functional Roles:

- Presenter: Presents team assignments via Piazza, screencasts, YouTube, etc.
- Writer: Consolidates team input into initial drafts for assignments.
- Sketch artist: Consolidates team sketch designs / ideas into final presentable sketches.
- Researcher: Conducts online research on assigned topics to inform team decisions / alternatives.

Domain Anchors:

- Targeting: Choose domain, define question(s), explore existing solutions, formulate data analysis tasks.
- Data Wrangling: Find and clean data, EDA (exploratory data analysis), transform and summarize data.

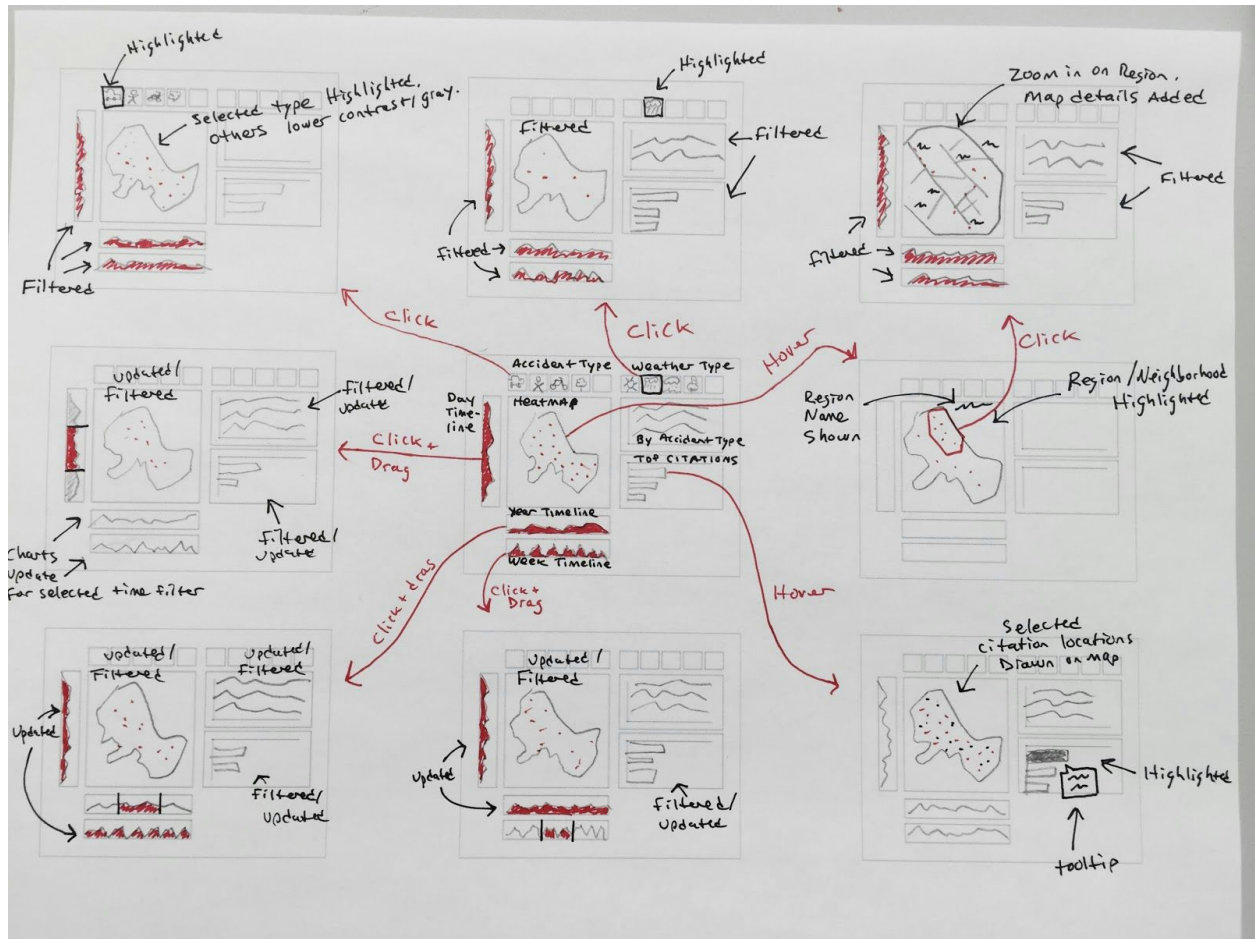
- Design: Design visual encoding, design interactions, design layout and storytelling.
- Implementation: Rapid prototypes, define data structure, explore efficient algorithms, design system architecture.
- Evaluation: Is the abstract right? Do encoding and interaction support the task? And, provide (new) insights?

3. Feature list

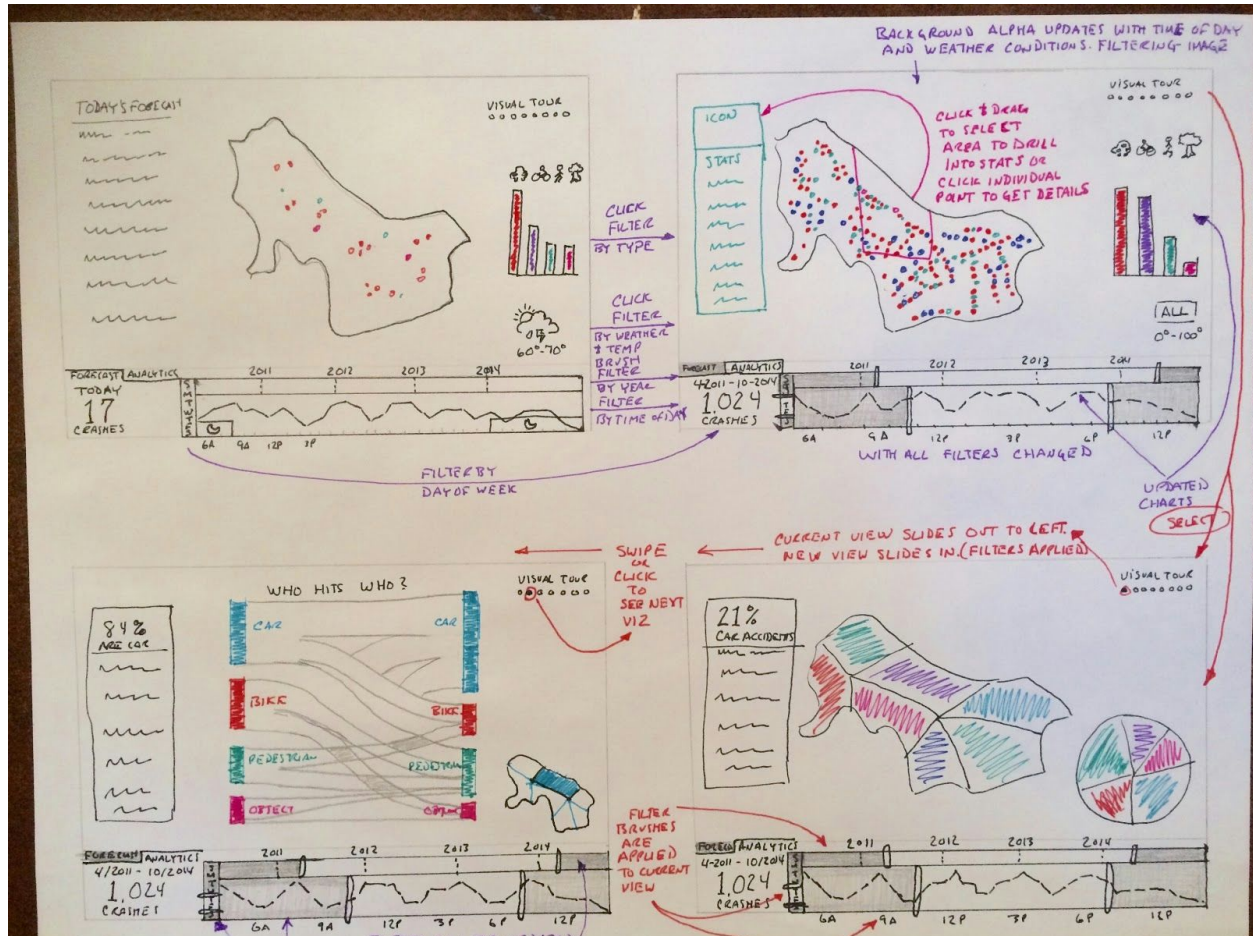
We plan to implement the following features:

1. Geo-mapping of data points
2. Forecast data plots
3. Trends by accident types
4. Cross-filtering
5. Plot of citations
6. Correlation with weather data
7. Additional map layers (transit, bicycle maps, public buildings)
8. Sankey diagram to map the size of the relationships between crasher and crashee

Interaction Storyboard

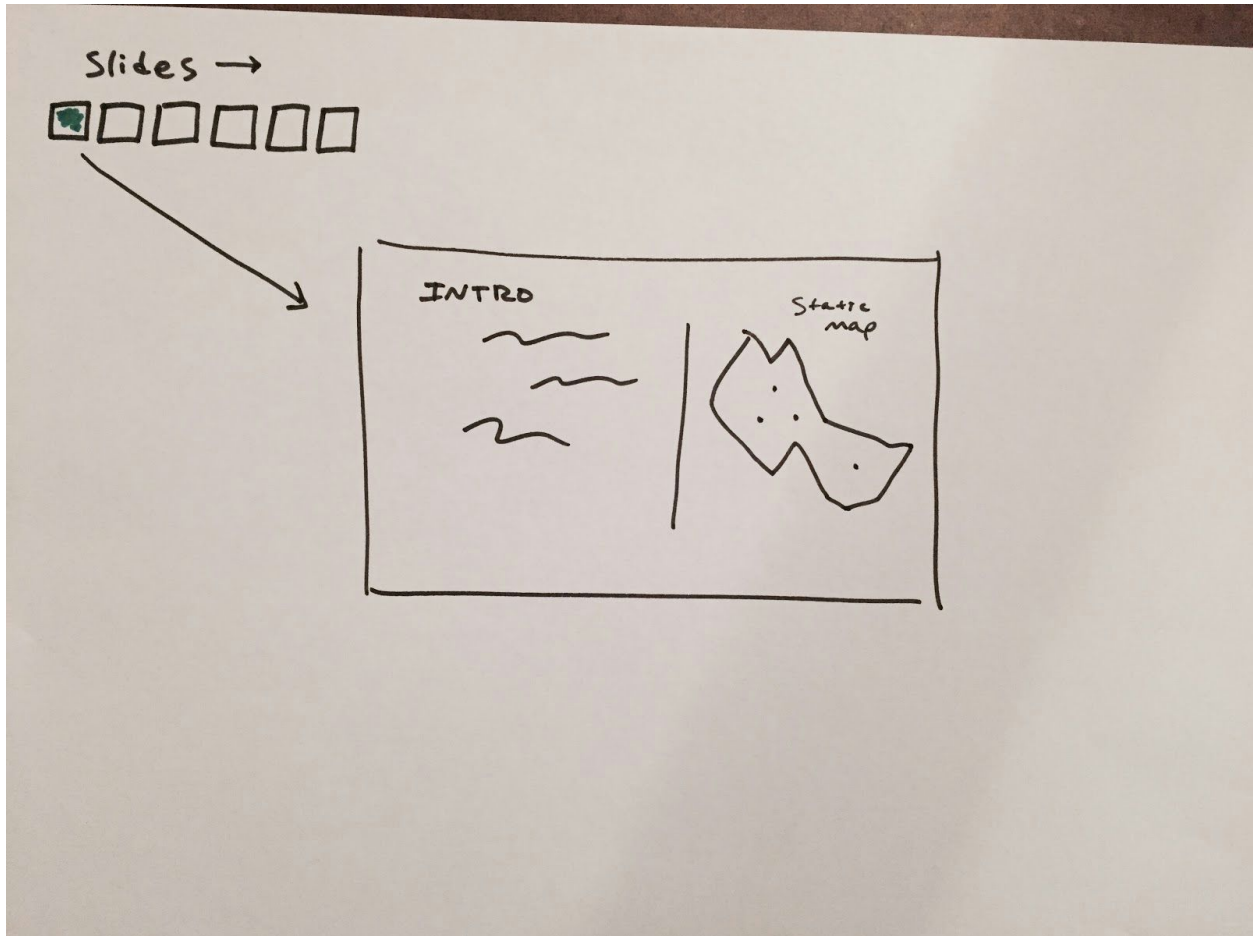


Single page, multiple views - Initial page is a forecast of today's accident hotspots to avoid. The user may switch from "forecast mode" into "analytics mode" by changing any of the filters (day of week, brush range over dd/mm/yyyy, brush range over time of day, type of weather, temperature, type of accident).



Webpage Layout / Storytelling

Our story telling approach will use a “martini glass approach” where we progress through highlight slides telling story and then lets user explore data in free form



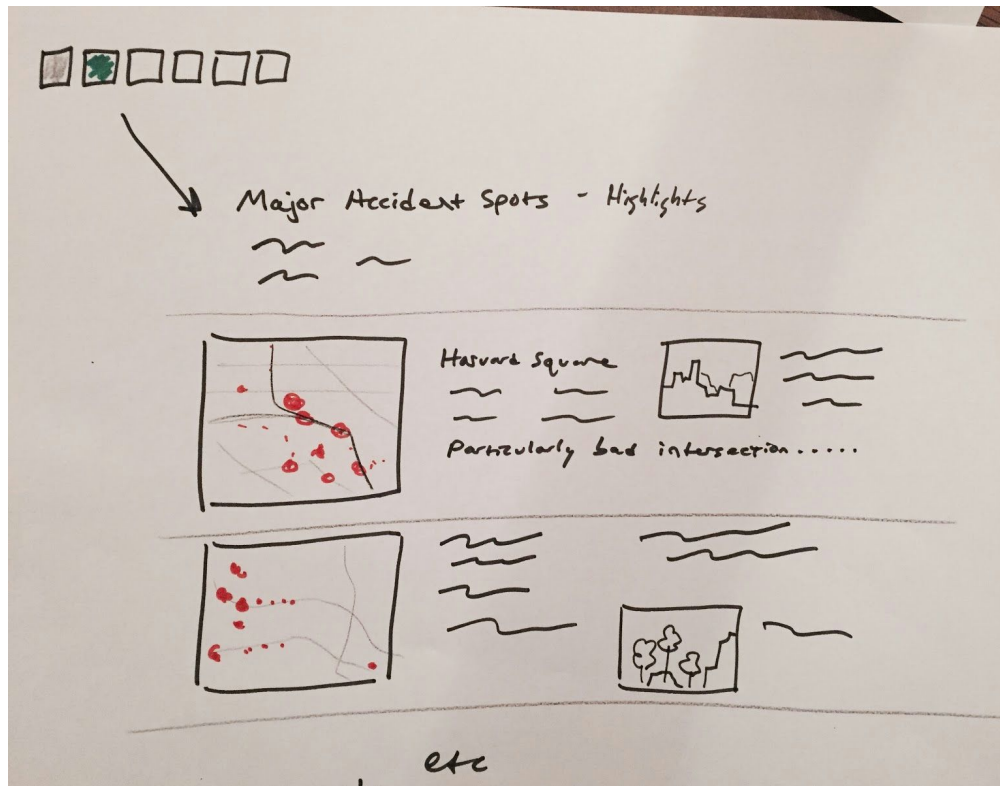


Figure: The first visualization will show our first highlighted story, navigation bar allows user to navigate through multiple stories

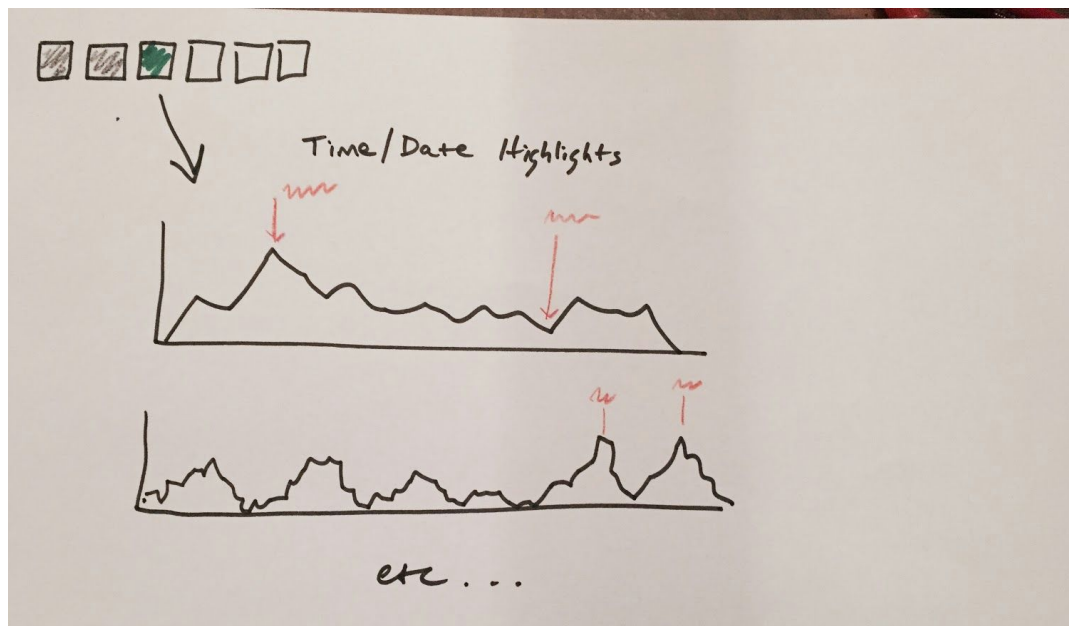
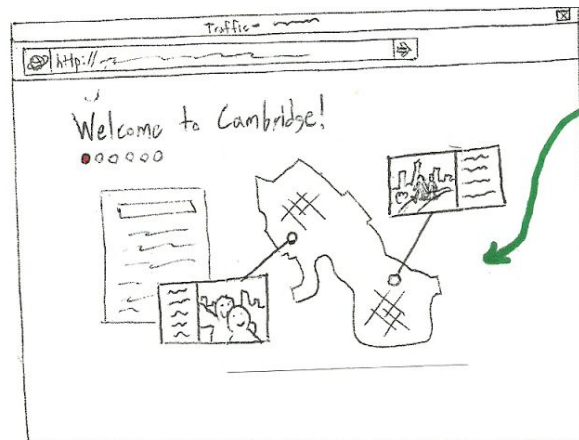


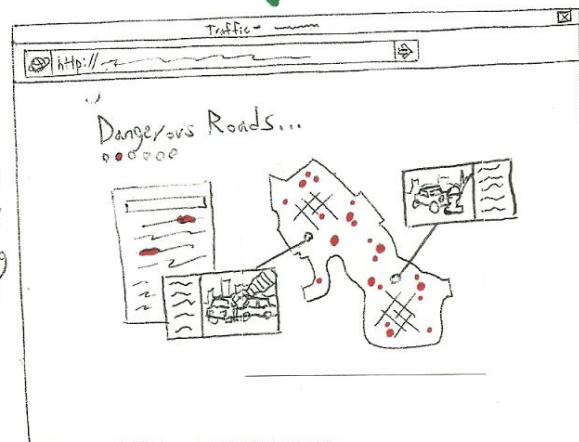
Figure: The first visualization will show our second highlighted story, navigation bar allows user to navigate through multiple stories

- ① Set the stage with some human elements. Engage with the viewer emotionally.

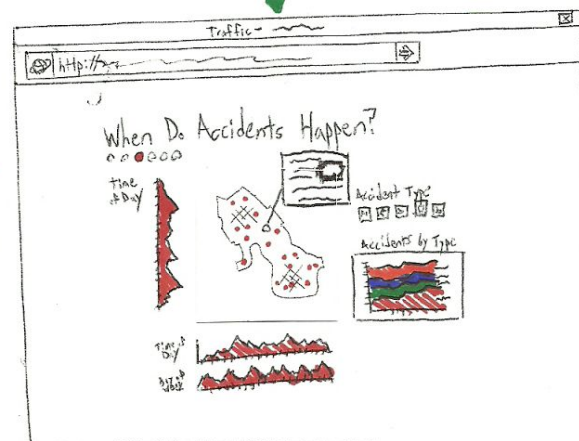


A common thread. The map of Cambridge acts as a constant touchstone for the storytelling and presentation.

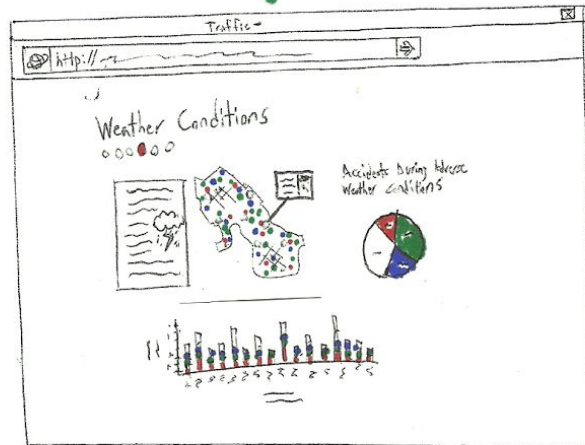
- ② Introduce the danger and show some basic, introductory data points without overwhelming the viewer. Highlight areas of interest.



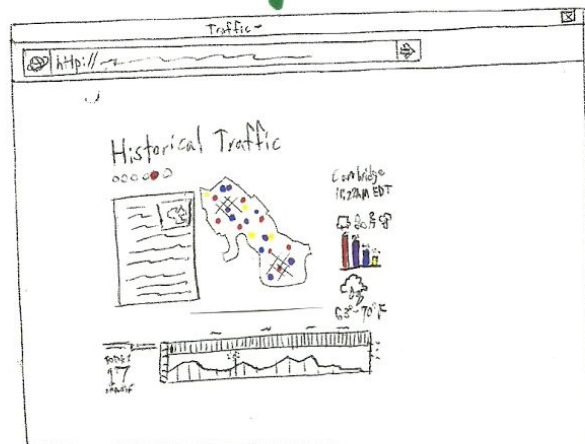
- ③ Discuss more specifics about when accidents happen and the contributing factors. Allow for some exploration on the part of the viewer.



- ④ Continue to elaborate on the various factors that may contribute to the accidents. Bring up all parts of the contributing sources.



- ⑤ Present an extremely flexible and dynamic view, allowing for unhindered exploration of the data by the viewer.



- ⑥ Summarize, and connect the story to the present. Present other sources of information and an infographic.



Complete Dataset

We are trying to answer the following questions at a high level with our vizualization project:

- Where are the most dangerous intersections/road locations?
- Where are most police citations for dangerous driving issued?
- How are traffic accidents/citations affected by various factors (time/weather/traffic-volume/road safety-features)?
- Measurable Outcomes – “Effects” (Dependent Variables)
 - Traffic Accidents
 - Location:
 - Coordinates (for mapping)
 - Street / Cross-street info (for referencing)
 - Filterable by accident type (auto/auto, auto/pedestrian, etc)
 - Sources
 - 2010-2013:
<https://data.cambridgema.gov/Public-Safety/ACCIDENT-2010-2013/ybny-g9cv>
 - 2014:
<https://data.cambridgema.gov/Public-Safety/ACCIDENT-2014/7fai-h9wk>
 - Visualization ideas:
 - Plot locations of accidents on map.
 - Color code by type?
 - Filterable by type and cross-linked factors?
 - Police Citations
 - Location:
 - Coordinates (for mapping)
 - Street / Cross-street info (for referencing)
 - Filterable by citation type (Speeding, run light, drunk driving, etc.) (some non-relevant types (e.g., revoked license, no insurance) can be filtered out of dataset.
 - Sources: 2010-2014
 - <https://data.cambridgema.gov/Public-Safety/Police-Citations-2010-2014/gmq6-8ver>
 - Visualization ideas:
 - Plot locations of citations?
 - Heat-map of citations (Overlaid with accident data as points)?
- Measurable Factors – “Causes” (Independent Variables)
 - Date/Time of Accident/Citation

- *Year (2010 – 2014)*
 - Potential factor: Increase in accidents due to population growth, commuter rise.
- *Time of Year (Date/time string, parseable)*
 - Potential factor: More incidents during school year, etc. More during winter.
- *Day of Week (Day value present)*
 - Potential factor: More commuter accidents during work week. More nighttime accidents on Friday/Saturday.
- *Time of Day ((Date/time string, parseable)*
 - Potential factor: More accidents during rush hour.
- Visualization ideas:
 - Filter mapped data by timeline brushing?
 - Summary charts showing distribution of total accidents/citations by date/time
- Weather
 - *Weather condition*
 - Potential factor: More accidents during adverse weather conditions
 - *Daily Temperature?*
 - Potential factor: More road accidents during freezing temperatures. More pedestrian/bike accidents during warmer temperatures.
 - Visualization ideas:
 - Filter mapped data by selectable weather condition icons / filter options?
 - Filter mapped data by brush-able temperature range?
- Traffic Volume
 - Data for certain intersections in Cambridge showing average & peak daily traffic counts.
 - Potential factor: Higher traffic volume = higher accident rate
 - Mapped by location (Same as above)
 - *Source: 1972-2014 (can be filtered to our applicable range)*
 - <https://data.cambridgema.gov/Traffic-Parking-and-Transportation/Average-Daily-Traffic-Counts-1972-to-2014/v43b-kqeq>
 - Visualization ideas:
 - Summary chart(s) showing distribution of volume levels by date/time?
- Presence/Non-Presence of Traffic Signal
 - GeoJSON locations of traffic lights.
 - Potential factor: Safer intersections, less accidents.
 - *Source:*
https://github.com/codeforboston/open_data_cambridge/blob/master/Traffic/TrafficLights.geojson

- Visualization ideas:
 - Enable/Disable viewable layer showing locations of traffic signals?
- Presence/Non-Presence of Walk Signal
 - GeoJSON locations of pedestrian crosswalk signals.
 - Potential factor: Safer road crossings for pedestrians, less pedestrian accidents.
 - Source:

https://github.com/codeforboston/open_data_cambridge/blob/master/Traffic/WalkSignals.geojson
 - Visualization ideas:
 - Enable/Disable viewable layer showing locations of walk signals?
- Presence/Non-Presence of Crosswalk
 - GeoJSON layer of marked crosswalks on roads.
 - Potential factor: Safer intersections for pedestrians
 - Source:

https://raw.githubusercontent.com/codeforboston/open_data_cambridge/master/Traffic/PavementMarkings.geojson
 - Visualization ideas:
 - Enable/Disable viewable layer showing locations of bike routes on roads?
- Presence/Non-Presence of Bike Route
 - GeoJSON layer of marked bike routes on roads.
 - Potential factor: Safer roads for bike riders
 - Source:

<https://data.cambridgema.gov/Geographic-Information-GIS-/Bike-Facilities/tdg4-6twm>
 - https://github.com/codeforboston/open_data_cambridge/blob/master/Recreation/BikeFacilities.geojson
 - Visualization ideas:
 - Enable/Disable viewable layer showing locations of bike routes on roads?
- Supplemental Data
 - GIS data layers (Basemap, Roads, Sidewalks, Buildings, place names, locations of interest)

Project Timeline and Milestones

Milestones

Milestone	Submit Project Proposal	Final Project Plan	Project re-design	Project prototype V1	Project prototype V2	Final project V2	Demo
Progress							
Deadline	3/27	4/3	4/10	4/17	4/24	5/1	5/8

Goals and Tasks

Our preliminary goals and tasks for this project are based on visualization process steps outlined in class. These goals and tasks will be refined as we make progress on the project.

Screencast

A screencast presentation of the Initial Project Plan can be found at <https://youtu.be/L33ZCXyVEVo>.

1. Target

a) Choose domain. We've generally decided upon studying traffic data from Cambridge, MA. We must decide what aspect(s) we will specifically examine. Some initial ideas are traffic density, traffic crimes, and street layouts.

b) Define question. We need to decide what questions we're trying to answer using the datasets we have access to. There are the obvious high level questions like "Where and when is traffic the worst / best?" and "Where are you most likely to get a ticket?". There are also the more interesting questions that involve correlation of multiple datasets like "How is traffic density related to traffic crime?" or even more interestingly "How do traffic crime and safety relate to poverty levels?". These are the deeper questions we need to better define moving forward.

c) Explore existing solutions. We will look at existing visualizations and data analysis that relate to our subject. Luckily, traffic data is a common subject of analysis. Additionally, any effective display of location-based data could be relevant to our project. We will compile our findings to be used as reference when we decide how we want to present our data.

d) Formulate data analysis tasks. Using the questions we've decided we want to answer, we must define how we will go about analyzing the data. This essentially means taking those questions and making specific decisions about what data will be involved and how to actually go about finding an answer.

2. Data Wrangling

a) Find and clean data. Once we know what questions we want to answer, we will have a better idea of what format we need the data in and what parts of the data are important to our project. We need to define the high-level concept of how we want to manipulate and format the data, then actually implement code to export the data in this format.

b) EDA (exploratory data analysis). This is the first step in drawing conclusions from our datasets. We will need to do some more data manipulation and try to find very basic answers to some of our questions. This might mean computing some simple averages or counts for data points in a given time period or physical area. This will be our first look into what the data can tell us and will set the stage for all of our analysis and presentation of the data moving forward.

c) Transform and summarize data. Having explored what the data can tell us, we will transform the data into a format that gets even closer to the final data we will use for creating our visualizations. After making a decision on the ideal format(s), we will once again implement code to export the data in this format.

3. Design

a) Design visual encoding. Having decided upon interesting questions and formatted the data, we will decide the best way to present that data in an intuitive and understandable way. We will likely reference resources like Bertin's Taxonomy and the existing solutions that we explored for guidance on what will best serve our needs.

b) Design interactions. Going a step further, we will expand the visual encoding design with useful interactions that can provide even more insights into the data. Our data will probably lend itself well to interactions as it generally spans both time and geographic location, thus it has many levels of depth (high-dimensional data). We will drive our visual interactions using the "Modalities for Interaction", User Intent, and the Heer and Shneiderman taxonomy for interaction.

c) Design layout and storytelling. We now must take all of the analytical work and frame it in a design layout to drive a compelling story. We will use Gestalt principles not only in our visualizations but in our design layout which will adhere to CRAP and Tufte's principles.

4. Implement

a) Rapid prototypes. At this point we'll have a great idea of what our final product might look like and we can begin implementing some basic visualizations that reflect all of our planning work. During this step we can further explore what is effective and what is possible / feasible.

b) Define data structure. We have worked with the datasets quite a bit at this point and during this step we will potentially perform the final transformations of the data. This involves implementing code and exporting to a suitable format. The final format of the data will be informed by all of our exploratory work up to this point, and what we have decided on for our final implementation targets.

c) Explore efficient algorithms. With our implementation goals in mind, we must consider what may be computationally demanding and how we can best approach the problem. We will avoid premature optimization, though, and simply try to approach problems in a logical and methodical way so as not to create unnecessary problems for ourselves in the future.

d) Design system architecture. We will implement the final program structure, having explored data formats and efficient algorithms. This implementation will require a full knowledge of exactly what we want the final product to be.

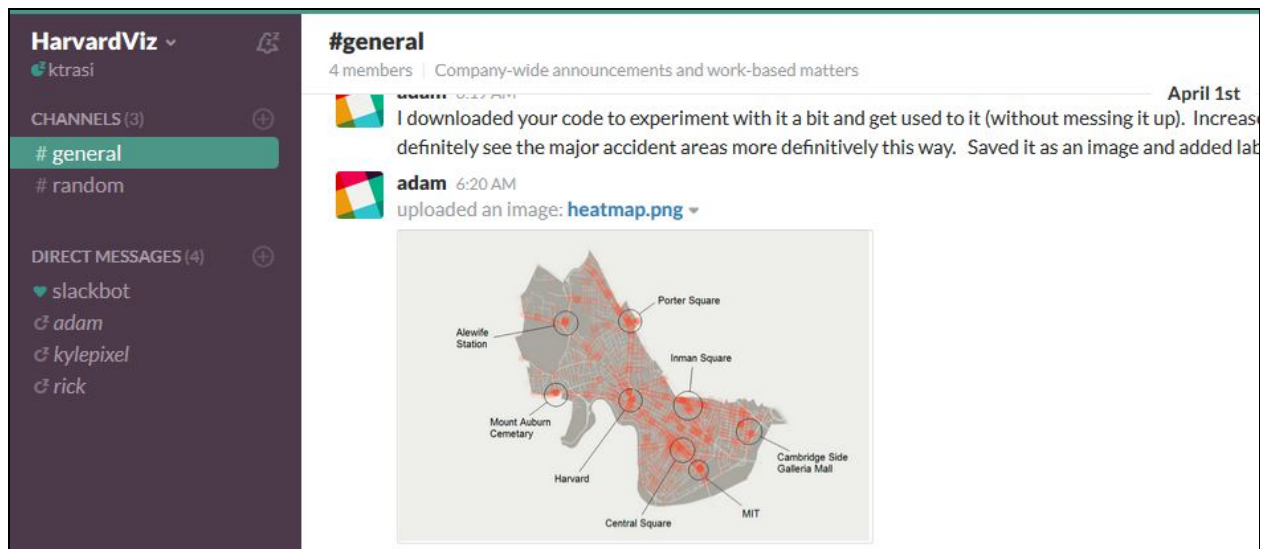
5. Evaluate

a) Is the abstract right? We will evaluate whether we accomplished what we set out to do. Does our implementation properly explore the questions we set out to answer? Can any small adjustments be made to get us closer to our original goal?

b) Do encoding and interaction support the task and/or provide (new) insights? We must look back at all of the planning we have done, even as far back as the existing solutions we explored, and ask ourselves if we chose the best possible encoding for the data. Does the visualization convey the information we wanted to convey? Do the interactions support this, as well? Can the user explore the data in a way that is useful for answering the questions that interest us? Can the interaction lead to interesting insights that are not necessarily present in the static visualization?

Communication Method(s) / Expectations:

Slack will be our primary method of communication. The expectation is that the team members will check slack throughout the project duration and respond as needed to notifications for @everyone or @username.



On Slack, to notify all members of the team please use the @everyone or @channel keywords. To notify individual members of the team please use the form @jamesbond

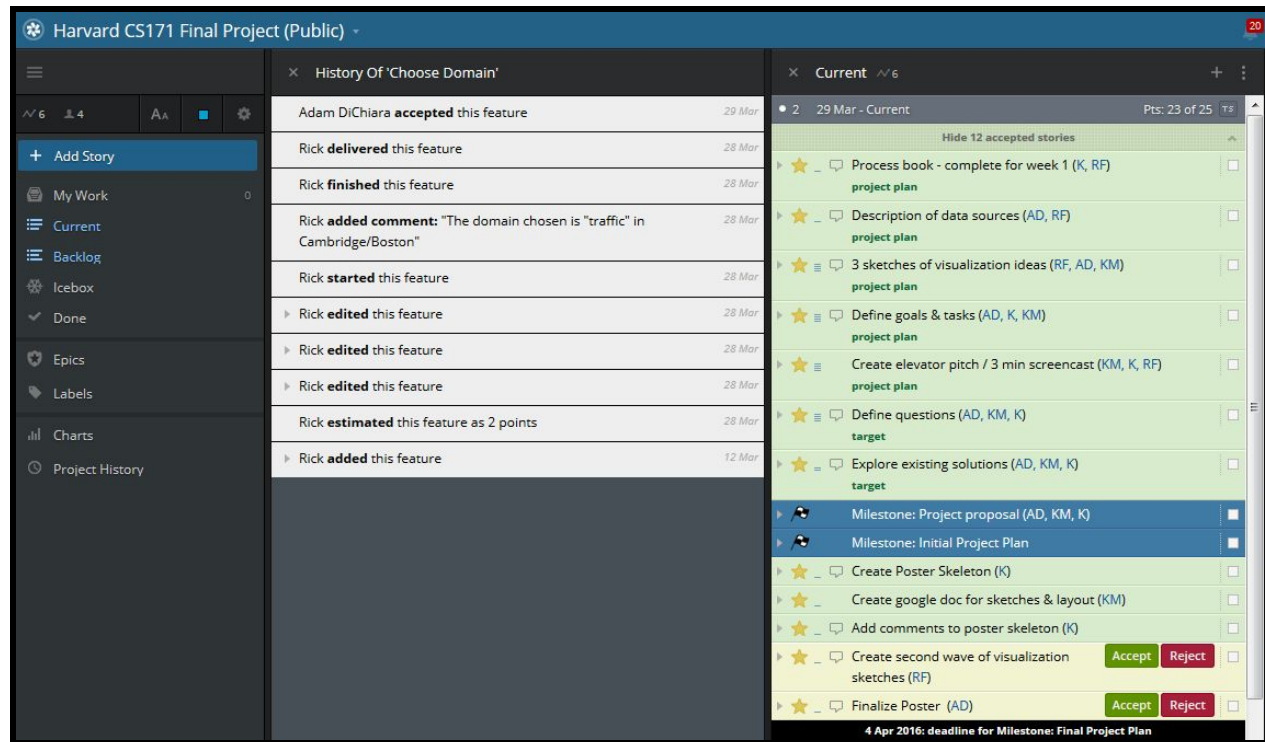
Meeting Schedule / Method(s):

A dial-in conference call is used to ensure that the team has assigned current work to individuals, is on track for milestones, and to demo individual work to the other team members as we progress, as well as discuss planning and resolve outstanding questions.

We have a weekly meeting at 8pm EST each Wednesday. Our dial-in number is 1 (888) 643-3084,,,96848588#. Additional ad hoc meetings may be scheduled (most likely on the weekend) for required collaboration on tasks and mid-week updates. We will use Screenhero for screen sharing (<https://screenhero.com>).

Implementation / Collaboration Plan:

Pivotal Tracker will contain the living implementation and collaboration plan (<https://www.pivotaltracker.com/n/projects/1554113>). As the project evolves our tracker will be updated with the details of the assignments for each member of our team and associated delivery horizon.



As we begin to implement our project, when possible the team will pair on items in tracker. Pair-programming (aka XP) helps to make complex items less difficult, promote efficiency, and share knowledge among our team members while allowing us to deliver quality work more quickly. Screenhero is a tool that enables remote pairing.