Utah Avalanches Process Book

Proposal

Basic info

Group Information

Names	Emails	UIDs
Evan Colt	evanc1229@gmail.com	u1046766
Christopher de Freitas	christopherdef@gmail.com	u1076216
Tanner Benedict	u0993730@utah.edu	u0993730

Project Title
Utah Avalanches

Github Repo
https://github.com/evanc1229/utah_avalanches

Background and Motivation
In the United States, there are on average 27 avalanche related deaths per year and countless more avalanche related injuries. This death count is concentrated in about 8 states, Utah being in the top 4 of those states for the highest death count. While this number may seem small, so is the winter backcountry recreation community. As this community grows, so does the number of backcountry travel-related deaths. Because this is a growing issue, it is increasingly important to analyze our existing avalanche data to give people the ability to make more informed, and thus life-saving decisions, when traveling in the backcountry. Our group was interested in this for the above-stated reasons and because member(s) of our group are frequently in backcountry avalanche terrain and would like a better tool for visually understanding this dataset that is widely used by the backcountry community.

Project Objectives

The primary goal for our visualization is to make the complex risk analysis done by avalanche analysts more accessible and readable for everyday backcountry travelers. Tabular avalanche data alone is great, but the ability to assess risk at a glance is critical for ensuring the safety of outdoor enthusiasts everywhere.

Our next goal is to take and improve upon the visualizations already developed by the Utah Avalanche Center. As a non-profit, there's only so much money they can sink into making their website and interfaces slick and modern (outreach and research already take so much money to maintain). Consequently, the visualizations on the site are lacking in a few key areas we'd like to address: only a small set of features collected are able to be visualized and interactivity is limited.

Finally, we want to enable travelers to better understand the risks in their region over time. It's expected that a project like this may be used to forecast risk before upcoming trips. But building a longer-term set of expectations in a region is also extremely helpful. Visualizations that aggregate data over time could reveal important regional trends that will help travelers make more informed decisions.

Data

This dataset is available for download from the Utah Avalanche Center website [https://utahavalanchecenter.org/avalanches] and is composed of crowdsourced and official reports regarding avalanches that have occurred in Utah. These reports are generated in one of three ways:

1) A SAR mission was required because of an avalanche that killed or seriously injured a person in the backcountry.
2) An employee or forecaster from Utah Avalanche Center witnessed an avalanche in the backcountry and filed an avalanche report.
3) A good samaritan witnessed or was affected by an avalanche in the Utah backcountry, collected data on the avalanche, and submitted a report to the Utah Avalanche Center. This report is then reviewed and edited by forecasters from UAC to ensure its authenticity and objectivity.

These reports are then compiled into a master data sheet showing every recorded avalanche in Utah on a given day. The data set contains information on the physical features of the avalanche, where and when it was, and why it might have occurred. This data is used by forecasters from the Utah Avalanche Center to create a daily avalanche report and is actively read by enthusiasts looking to collect more information when making their plan to head into the backcountry.

Data Processing

Data Features

The visualizations provided by the Utah Avalanche Center only leverage a small subset of physical descriptors of avalanche events (i.e. Date, Depth, Region), but the data collected is much richer. Asterisks (*) next to names indicate a column is a free-text entry. Many of which we do not intend to use, but here are ones we've decided to make visualizations for:

Column	Description	Example
Date	Date of the incident	2/11/2020
Region	Approximate region of incident	"Uintas"
Place	The location that the event occurred chosen from a list of all ski areas in Utah	"West Desolation Ridge"
Trigger	The cause of the avalanche	"Skier" or "natural"
Weak Layer	The snow conditions that made the avalanche possible	"New Snow"
Depth	The depth of the slab of snow	2ft
Width	How wide the slab of snow was across the mountain	100ft
Vertical	How long the slab of snow was from top to bottom	400ft
Aspect	The direction the face that the avalanche came from was facing	"North-East"
Elevation	The elevation of the location of the avalanche	10,000ft

Coordinates	Coordinates of the avalanche derived from the place column	40.622507000000, -111.660829000000
Caught	The avalanche engulfed someone but did not bury or move them	True/False
Carried	The avalanche caught someone and moved them	True/False
Buried	The avalanche fully buried partner rescue required or partially buried someone	True/False
Injured	The avalanche injured someone	True/False
Killed	One or more people were killed by the avalanche	True/False
*Accident and Rescue Summary	Only created when a Search and Rescue operation was necessary and is created by Utah Avalanche Center	Free Text Entry
*Terrain Summary	Description of the terrain features and how they contributed to the avalanche risk factor	Free Text Entry
*Comments	Comments left by person who filed report	Free Text Entry

Data Cleaning

For generating our text based visualizations, we'll be using the columns "Accident and Rescue Summary", "Terrain Summary", and "Comments". As free text entries, they will need to be cleaned first if we want to extract meaningful information from them. We'll follow a typical NLP pipeline by case folding, stemming, and removing stopwords and punctuation. After that, we'll be able to effectively identify patterns in language usage, and write simpler regex patterns for finding certain information.

Data Processing Pipeline

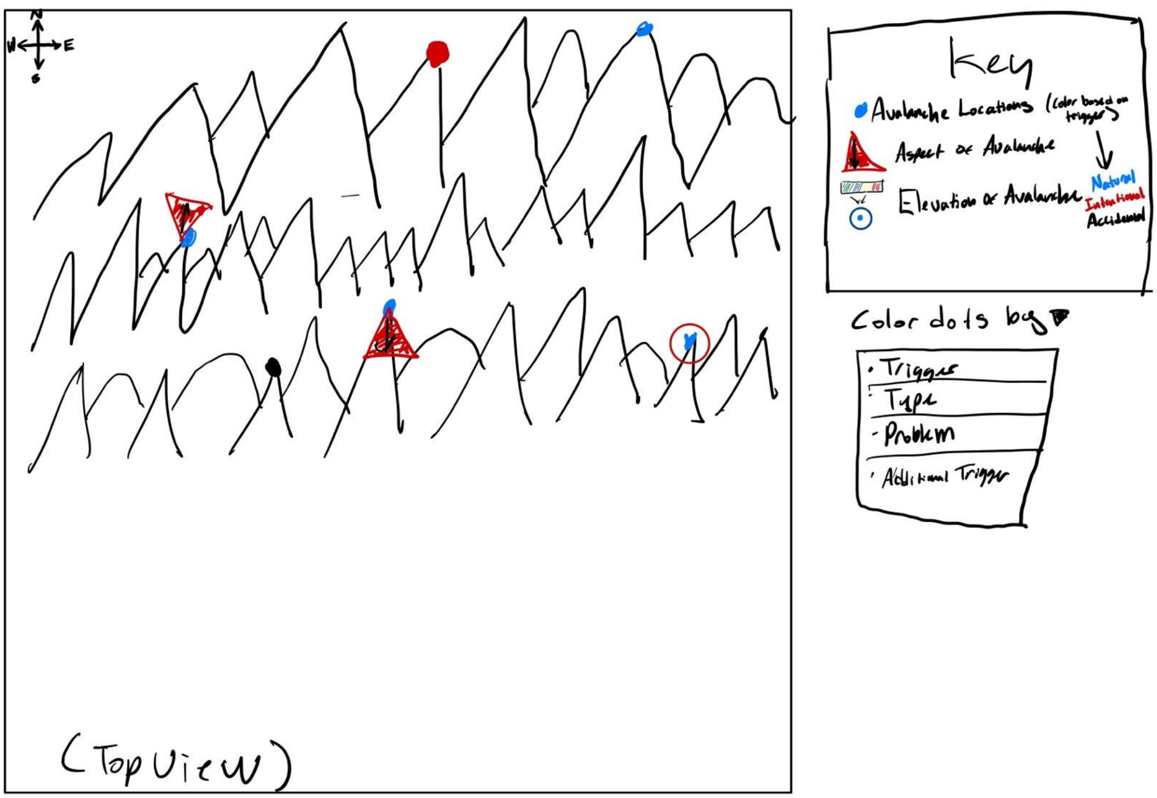
With the static dataset we're using, and our team's experience with Python, we plan to perform most of our data cleaning and transformation in Python before serving it to D3. However, given the amount of interactivity and flexibility we plan to include in our project, a lot of data processing will still need to be done on the fly in D3.

Initial Visualization Design Sketches

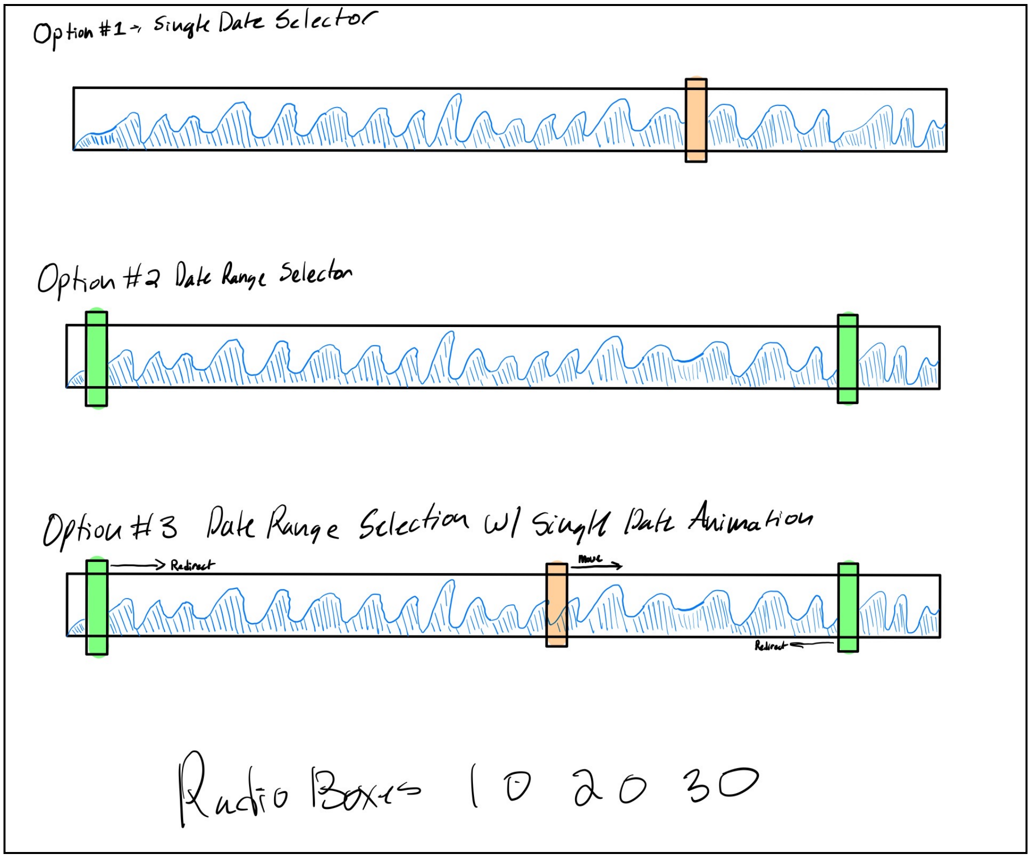
The following is a list of our individual visualizations and thought processes linking them together. To start, we'll go through individual components we thought up and our reasonings for each. In the end, the [Final Layout] section will show what our final vision for this project looks like.

Component Ideas and Prototype Sketches

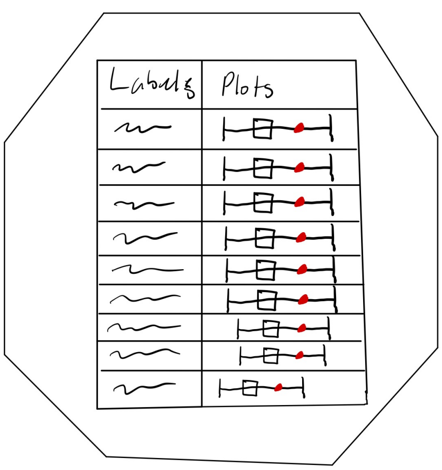
At a Glance View



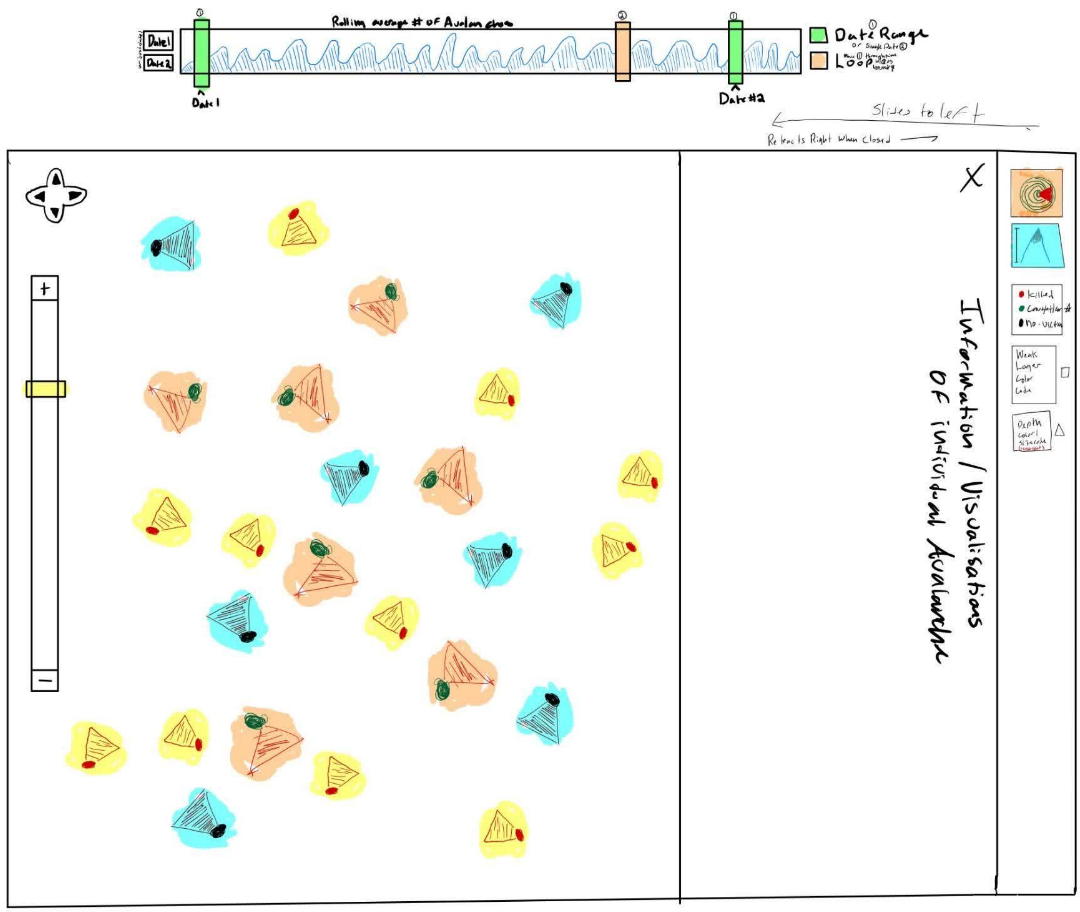
First Draft of this view. Has all of the main features but is less polished.



Component for next draft of the At a Glance view. This is a data selection visualization that is composed of an area chart depicting (# avalanches / Time). This can be used to select a range of dates, a single date or to iterate through a range of dates.



This is another component for the At a Glance view. This component would appear when an avalanche is clicked and will contain a tabular version of the data on the avalanche along with a chart showing how its stats match up to other avalanches during the same time period.



This is the third draft of this visualization and is defined by 3 main features:

1) A dynamic time axis that also serves as an area chart showing (# of avalanches/time). This time axis can be used in 3 different ways. Time can be selected for individual data, a range of dates, or as an animation looping between two selected dates. This is important because this visualization should be useful for looking at the past in addition to looking at the avalanche status in Utah on a given day. Making the time scale dynamic like this will allow for this flexibility.

2) The locations of avalanches will be plotted on the map and can be coded by aspect, elevation, human impact, weak layer, and depth. These will be represented through several visual channels such as color, direction, position, and length. Because we cannot use all of these channels at once we will disable some of these filters as necessary based on previous selections by the user. Showing this data on the map is important because it will provide the user with up-to-date information on a given location at a glance.

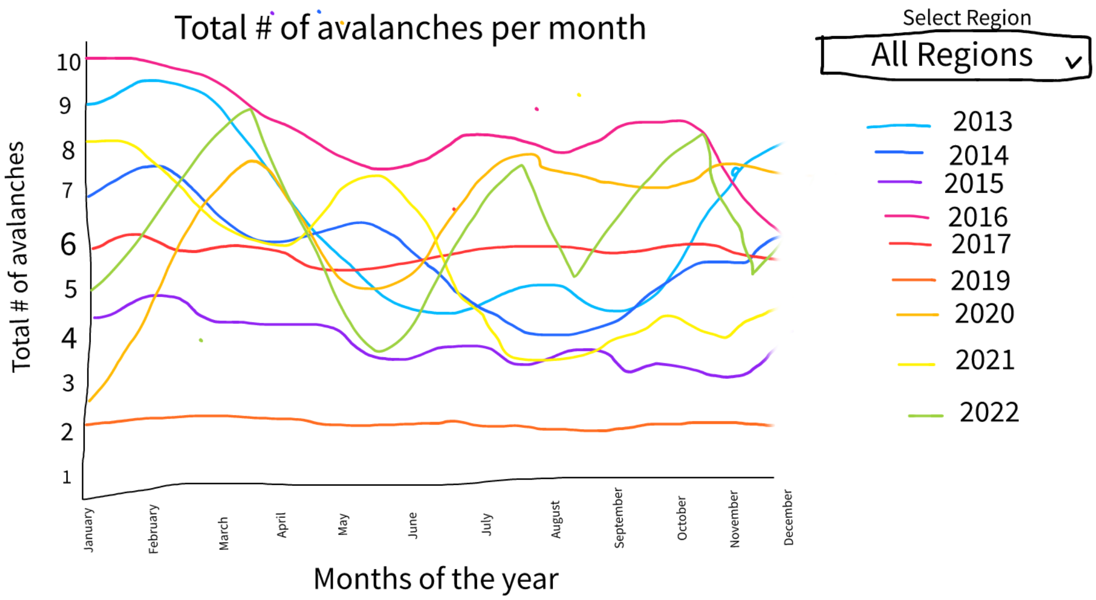
3) When a given avalanche on the map is clicked on, there will be an individual visualization for that avalanche including tabular data and plots (possibly utilizing general avalanche forecast data for that day). This is important because it will allow the user to dig deeper into what is going on in a given region.

This visualization is a component of the above prototype and uses a map to show the density of avalanches and where they occur. The size and color of the circle determine how many avalanches of that type occurred in that location. Zooming in will divide the circles into smaller circles in different sections until you zoom in to the exact location of the avalanche. Using the drop down you can select what year you would like to look at.

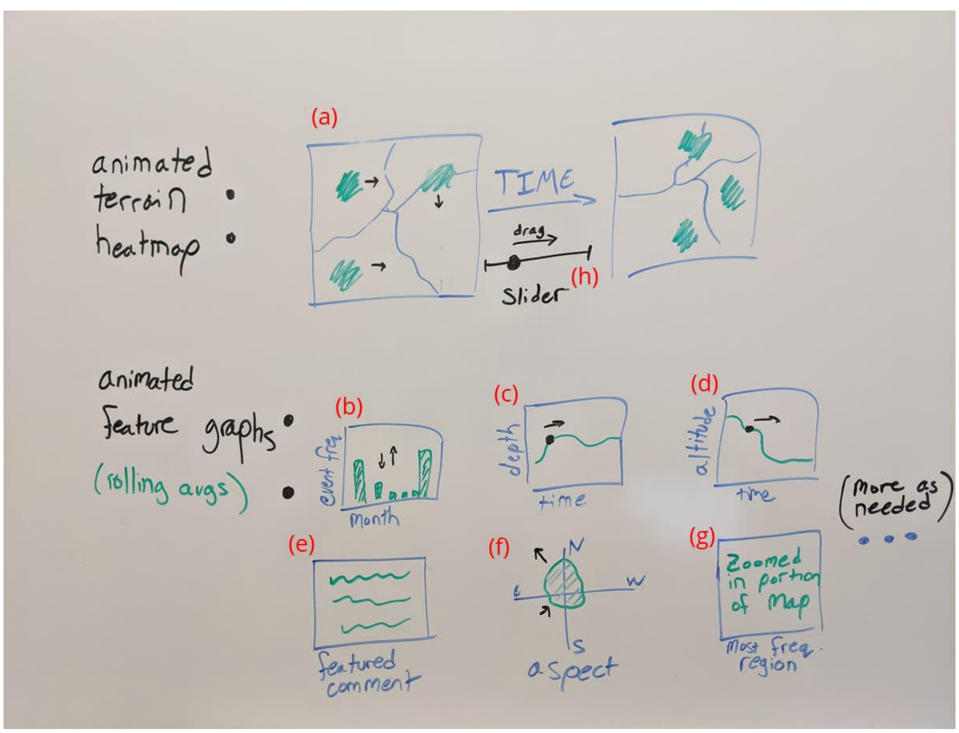
Text Insights View

A large source of unstructured, and underutilized, data in the Utah Avalanche Center dataset is the report comments. By looking at word clouds, readers could get an intuition on which words to focus on when reading individual reports about incidents in their area. (component)

Historical View



This is a representation of what a line chart could look like to represent the data per month so that patterns can be recognized about when the majority of avalanches happen. Each year will be represented by a colored line on the graph. The user can use the drop-down box to select what region they would like to view, or view the data for all regions. The y-axis will change depending on the region selected and the max number of avalanches for that region.



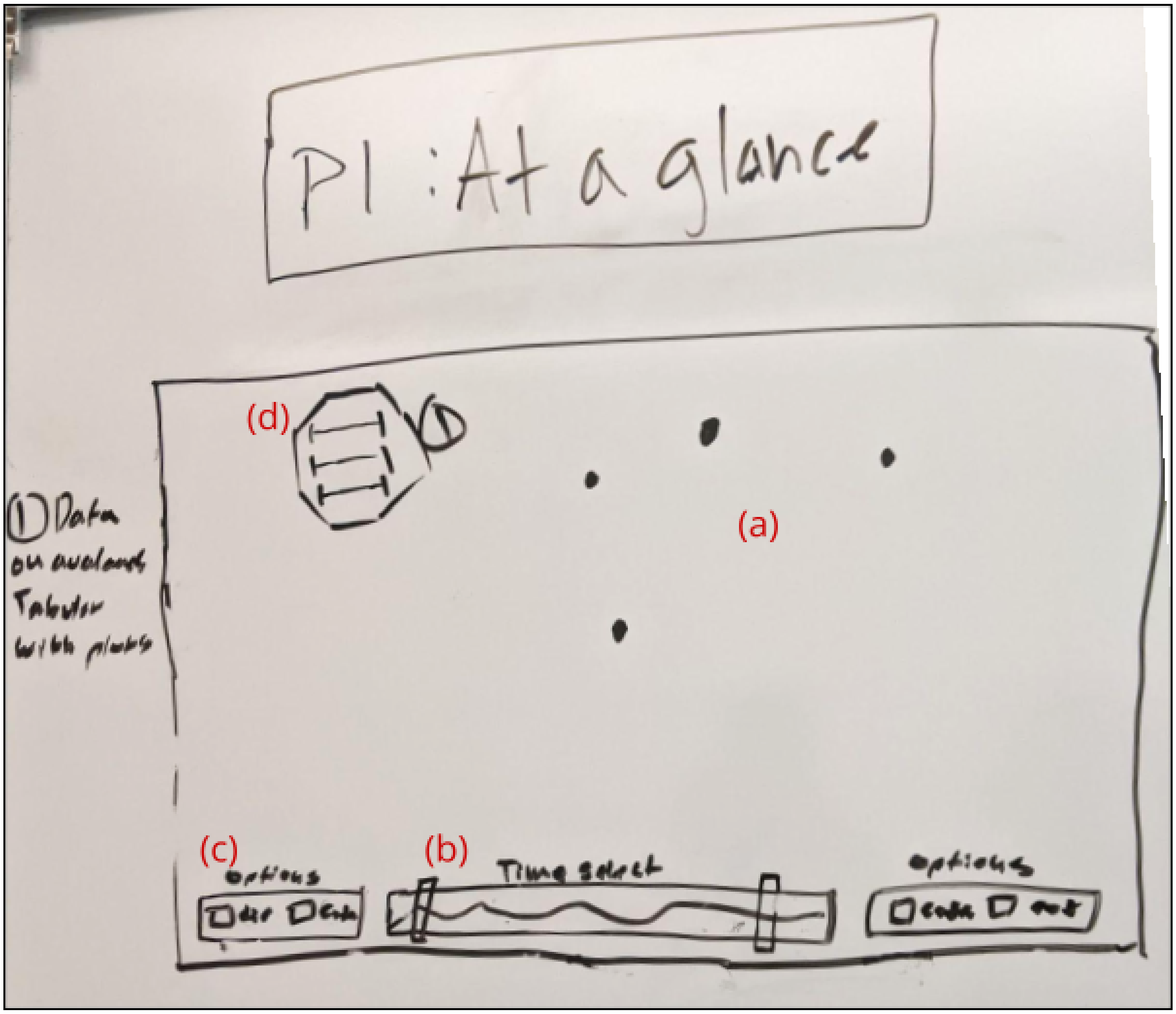
Two major parts of this visualization, both are linked temporally by a slider (h) that can be clicked and dragged. The slider represents position in time. As the slider is dragged, each visualization smoothly transitions to represent the same information at different points in time. For the charts, blue is the figure structure, green is the data we'll plot, and black shows the directionality of changes over time.

1) On top, there's a heatmap (a) over terrain that represents the rolling average of incident counts over time (likely averaging at the resolution of a month/season).
2) On the bottom, there's a panel holding various graphs (b)-(g) providing more context about that particular moment in time. The following section will describe each element, and how it will change with time.

> (a): a map over the entire region will have a "heatmap layer" overlaid on top to indicate the proportions of incidents in a region; the heatmap will smoothly transition between states with time
> (b): running average frequency of incidents per month; the bars raise and lower with time
> (c)-(d): simple line graphs showing the rolling average depth/altitude over *all* time; a black dot will show the user's position in time and move left or right along the graph with time
> (e): a featured comment will be selected using heuristics we plan to establish (i.e. picking the longest or earliest comment); the comment displayed will change with time
> (f): rose plot to show the proportions of avalanches at certain aspects; the center blob will smoothly morph its shape with time
> (g): a "hotspot region" will be selected based off heuristics we plan to establish (i.e. hiking location closest to the largest number of incidents), a zoomed in portion of the map will be shown with a red bounding box and text to describe the reason it's being considered dangerous (i.e. the number of nearby incidents); the map will zoom and pan as needed to shift to new locations with time

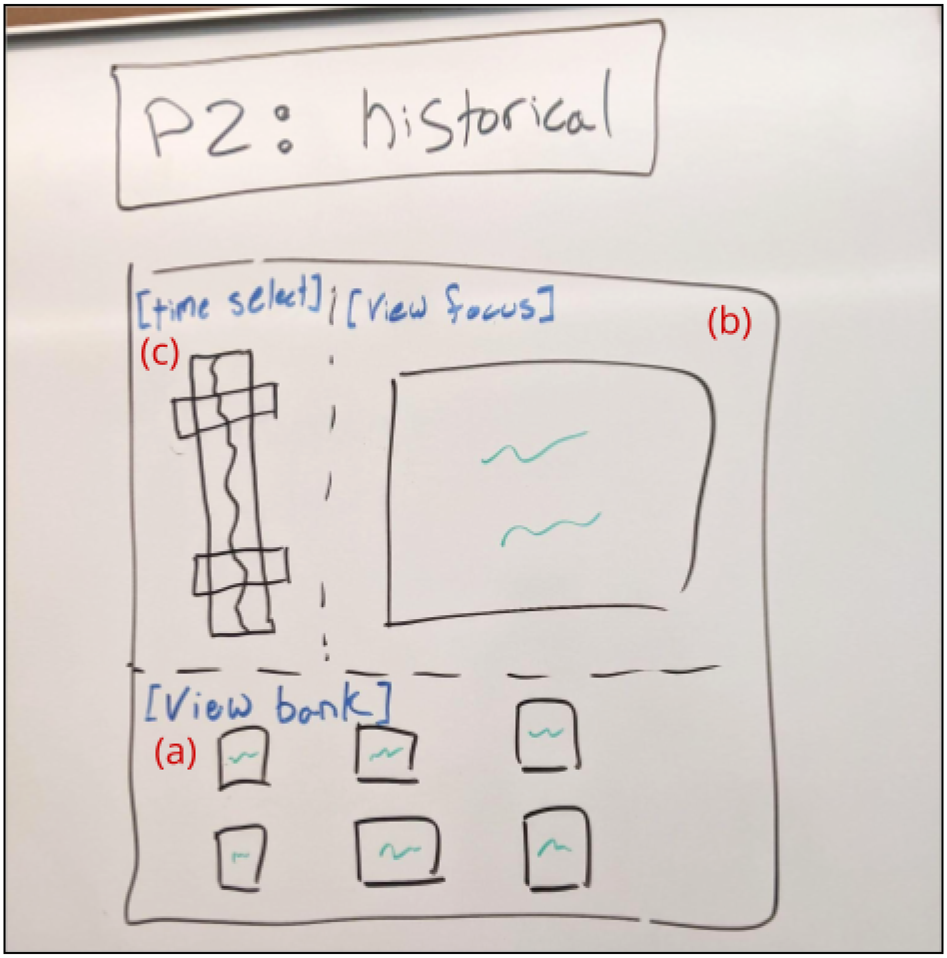
Final Layout Sketches

At a Glance View



Final Prototype of view that incorporates the designs and description illustrated in the above Ideas and Prototypes section. This view utilizes the timeline, tabular data set and mapping components created above.

Historical View



Is the Final Prototype of the view that incorporates the designs and description illustrated in the above Ideas and Prototypes section. This view utilizes several bars and line charts, including the line chart outlined above, and mimics the prototype in the Historical view prototypes section, but with a different layout.

Text Insights View

Final view sketch that will be several word clouds based on the various free text columns available in the dataset. This view will also incorporate the same date selection visualization that is used by other views.

Milestone 1

<u>First Prototype of ‘At a Glance’ / Milestone</u>

The following are prototype subviews for our ‘at a glance’ main view. This view will serve as the initial landing page for our visualization and will be the primary page used by backcountry enthusiasts on a day-to-day basis.

The three major components for the first page are independently complete, for the most part. By organizing each component into its own modular class, we’ve made it easy to combine

<u>TimeSelect</u>

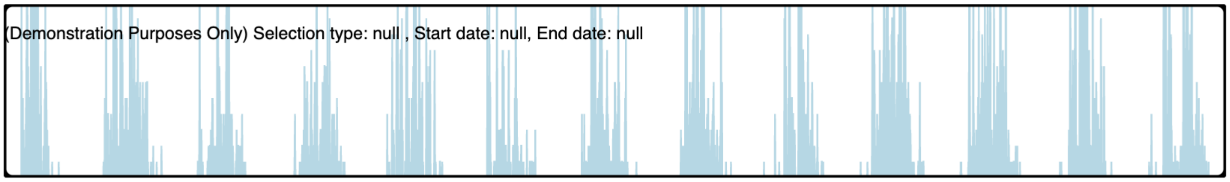
Date Processing

In order to produce the following sub-visualization our main csv was grouped by number of recorded avalanches per date using d3’s rollup method. After grouping, all data for avalanches prior to 2012 was removed as the number of avalanches prior to this time was extremely sparse. Missing dates on which avalanches had not occurred were then added to the dataset with an avalanche count value of 0. This smoothed out the area graph and made it more readable.

Interaction and Functionality

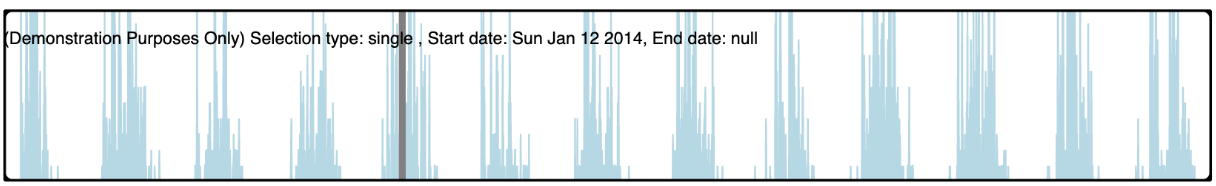
This sub-view is used to select a given time period to display on the map view described below. The view uses an area plot to give the user an idea of the volume of avalanches in Utah over time period and give them greater context for the date/dates that they are selecting. We chose an area plot because it is more readable than a line plot and effectively conveys when there is a high or low volume of avalanches in the Utah mountains. This sub-view can be interacted with through a series of clicks.

Default) The view displays an area plot of the volume of avalanches in the utah mountains. The overlaid text is only visible for demonstration purposes.



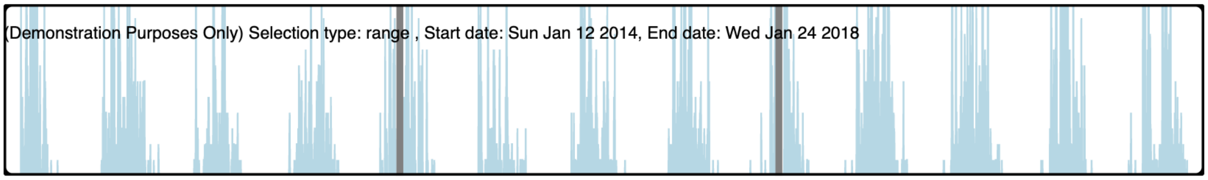
Click1) The chart displays a selection bar at the scaledX location of the mouse click. The click location is used to caluate the selected date and the view enters selection mode “single”. When querying which date has been selected via a getDates() method. The returned value will be the following JS object:

{
mode: “single”,
start: <SELECTED DATE1>,
end: null
}

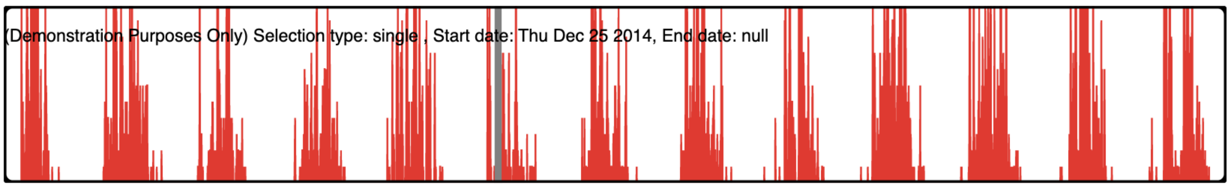


Click 3) The chart displays a selection bar at the scaledX location of the first mouse click and a selection bar at the scaledX location of the second mouse click. The click locations are used to caluate the selected date and the view enters selection mode “range”. When querying which dates have been selected via a getDates() method. The returned value will be the following JS object:

{
mode: “range”,
start: <SELECTED DATE1>,
end: <SELECTED DATE2>
}



Click 3 Error) If the mouseX location for click 3 is less than the mouseX location for click 2 then the view will flash red and will revert back to the state of click two. This was designed to prevent a backwards date range an provide the user with a warning that they had made and error.



Click 4)
The bars are cleared and the sub-view is returned to the default state.

ToolTip

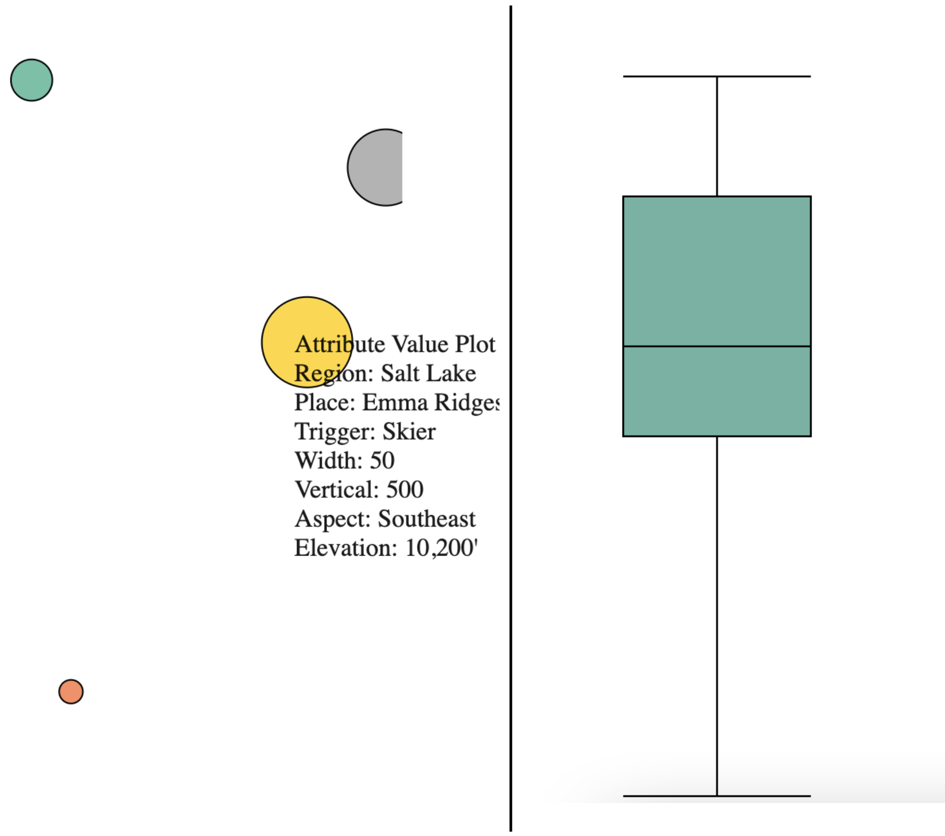
This view is used to see information about a specific avalanche that is being hovered over on the map view. This view will show a number of box plots that demonstrate how the selected avalanche measures up to avalanches in the given time period. If a single date is selected and not a time period. This tooltip will show how the selected avalanche compares to all other avalanches in the dataset

Data Preprocessing

There was little need for data preprocessing for this tooltip other than converting string integers to proper ints.

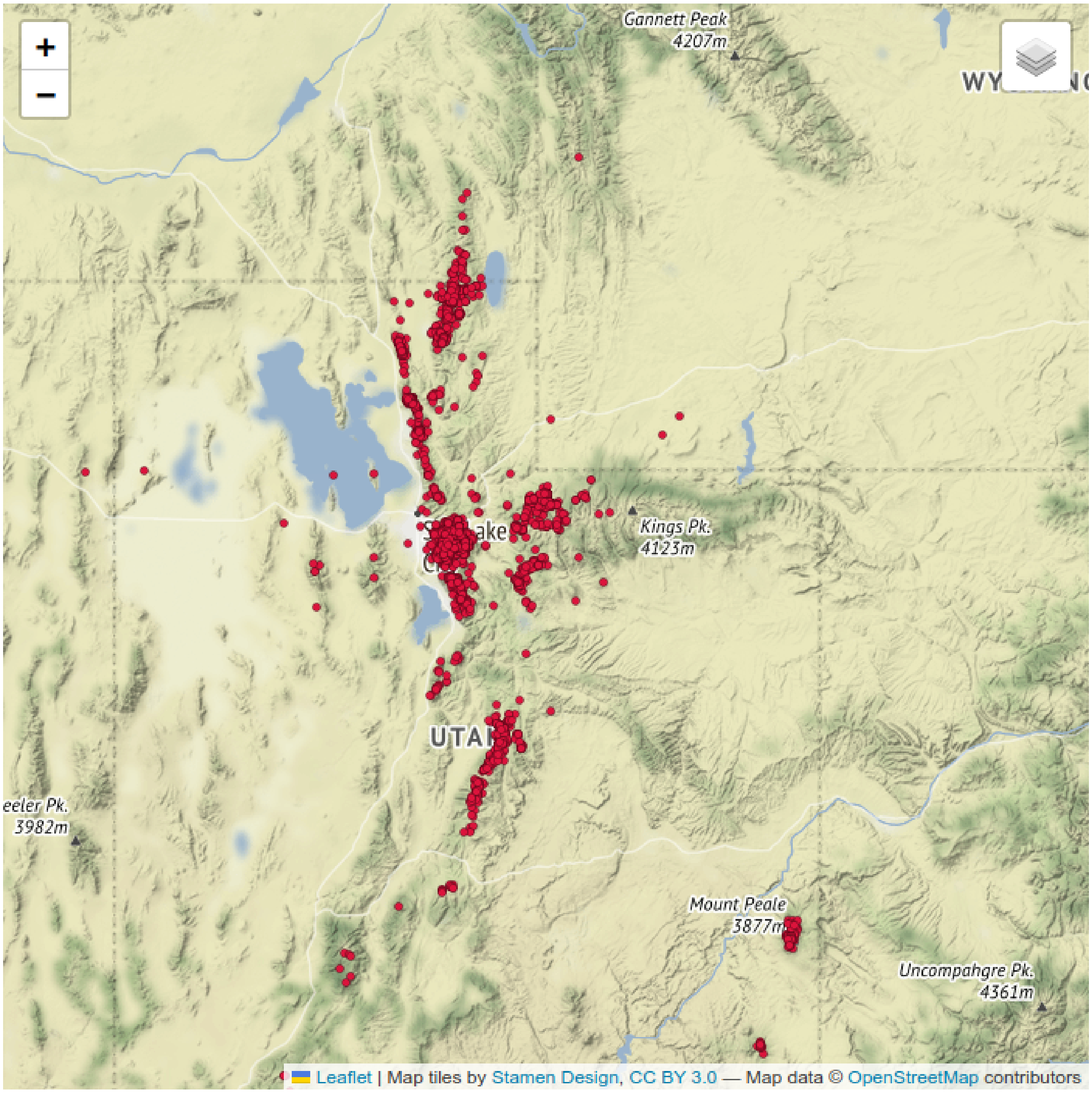
Interaction and Functionality

This view is static in the sense that the tooltip itself is not interactive. However, this tooltip will change relative to the data bound to the selected avalanche. This can be seen in the following screenshot where data about the avalanche is displayed on hover.



Region	Place	Trigger	Depth	Width	Vertical	Aspect	Elevation
Salt Lake	Sunset Peak	Snowboarder	14"	80'	30'	North	10,400'
Salt Lake	Pioneer Ridge	Skier	2'	20'	125'	North	9,900'
Salt Lake	Pioneer Bowl	Natural				North	9,700'
Salt Lake	Patsy Marly	Skier	2.5'	100'	100'	North	9,700'
Salt Lake	Two Dogs	Skier	3'	70'	450'	North	10,200'
Salt Lake	Emma Ridges	Skier	18"	50'	500'	Southeast	10,200'
Salt Lake	Sunset Peak	Skier	3.5'	1,500'	800'	North	10,400'
Salt Lake	Grizzly Gulch	Skier	2'	125'	150'	North	10,000'
Salt Lake	Silver Fork Headwall	Skier	2.5'	60'	400'	Northeast	9,900'
Salt Lake	Snake Creek	Skier	6"	15'	200'	Northeast	9,700'

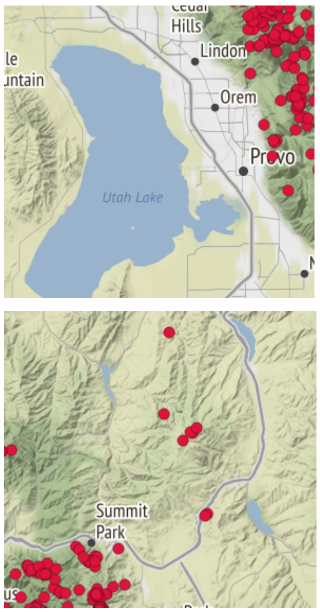
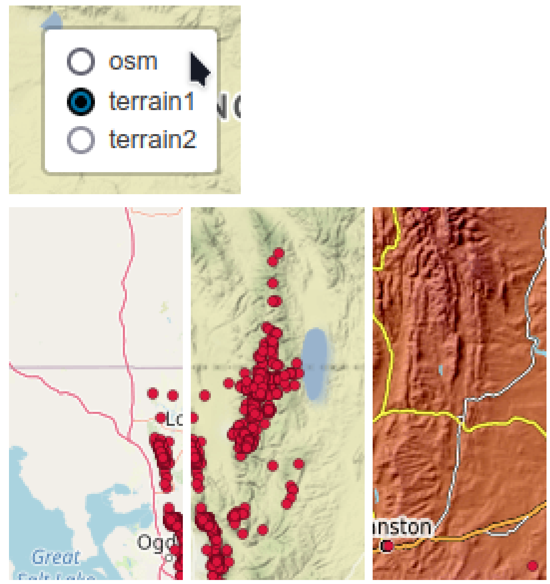
Map



Data Preprocessing

Plotting coordinates on the map required preprocessing of the coordinates provided, which are initially just strings (i.e. “(100.2543, 91.4345)”). The strings needed to be parsed, then coordinate conversion functions between the display space and latitude-longitude coordinates needed to be established.

Interaction and Functionality

Free Panning and Zooming	Basic Interactivity	Three Different Layers
		

The map currently can be panned and zoomed, provides basic reactivity to *"mouseover"* events (to be used leveraged by the ToolTip component later), and can switch between three different layers of the map.