Characterizing Utah Avalanche Observations Over Time

Project: Characterizing Utah Avalanche Observations Over Time
Repository: https://github.com/jmaxmarno/avy-obs
Team members:

Jadie Adams

u0930409

jadieraeadams@gmail.com
u0930409@utah.edu

Max Marno

u0656379

jmaxmarno@gmail.com
u0656379@utah.edu

Background and Motivation

As winter recreationists we have spent time pouring avalanche forecasts and observations to better understand the environment around us. Making informed decisions is a critical component for winter backcountry travelers and while published advisories are the first step in assessing risk, past observations are also relevant. The Utah Avalanche Center collects and makes available a dataset of avalanche observations from trained forecasters, and the public. These observations represent a sample of the real occurrences, but can be useful for exploratory analysis. The dataset contains observations from as far back as 60 years ago, but starting around 2009 there is a more consistent number of observations. We are interested in visualizing how the distribution of these observations has changed over the last decade. Our project will enable users to explore how reported rates of avalanches with different characteristics have changed, potentially drawing attention to the increase in backcountry traffic and avalanche caused fatalities. We hope to bring this design to the Utah Avalanche Center and if they agree it is effective in helping users understand and explore the observations, we plan to collaborate with them to make it accessible through their website.

Project Objectives

Our objective is to provide interested parties with an interactive visualization to explore the distribution of avalanche observation characteristics over time. We are aware of the presence of reporting bias, and will focus on faceting the observations into attribute sets, and displaying the normalized distributions of these attributes. Currently, interested users can export this data as CSV or explore it in a table format. We hope to make the data more accessible and useful by implementing mechanisms for exploration, while reinforcing the notion that

observations are not a representative sample of all Utah avalanche phenomenon and absolute inference should not be made about the population distribution.

Our visualization will have both user-driven interactive components as well as guided story-telling aspects to highlight points of interest. It will allow the user to explore the following characteristics of avalanche observations:

- Trigger / cause of avalanche
- Discretized elevation of avalanche
- Discretized size of avalanche
- Aspect or cardinal facing of avalanche
- Absolute number of people carried, caught, and buried as well as number of injuries and fatalities and the proportion of avalanches that involved these events.

Data

Data of recorded avalanche observations is provided on the Utah Avalanche Center website: https://utahavalanchecenter.org/avalanches They have made the data available to download in a csv file.

Below is a histogram of the observations from 01/01/2010 to now, it is binned by month. It can be seen that in the winter months there are around 100 observations on average and in the summer months there are none.
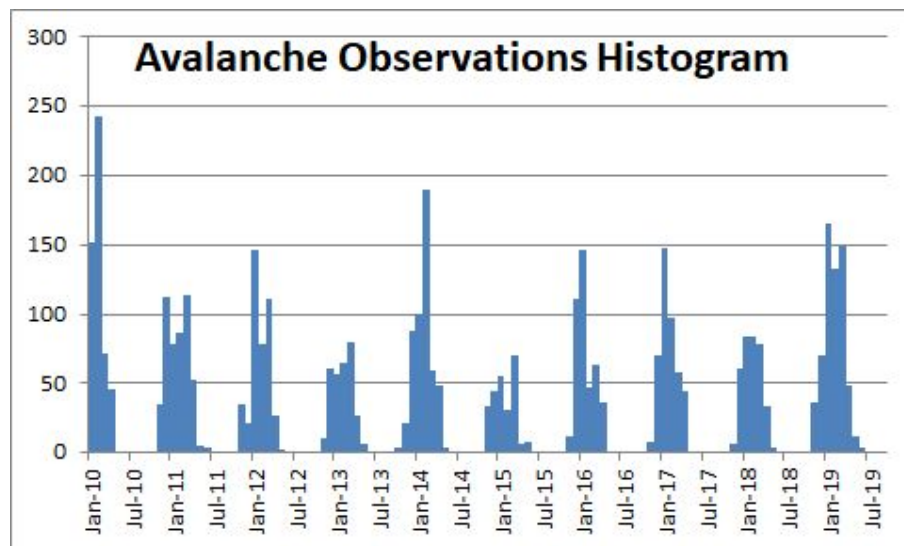
The data recorded for each observation includes:

Date, Region, Place, Trigger, Depth, Width, Vertical, Aspect, Elevation, Coordinates, Weak Layer, Caught, Carried, Buried - Partly, Buried - Fully, Injured, Killed, Accident and Rescue Summary, Terrain Summary, Weather Conditions and History, and Comments.

Often some of these fields are left blank but they all include at least a date and nearly all have the first half of the characteristic accounted for.

Here is the distribution of observations among the different regions and shows why we decided to exclude a cartographic visualization:
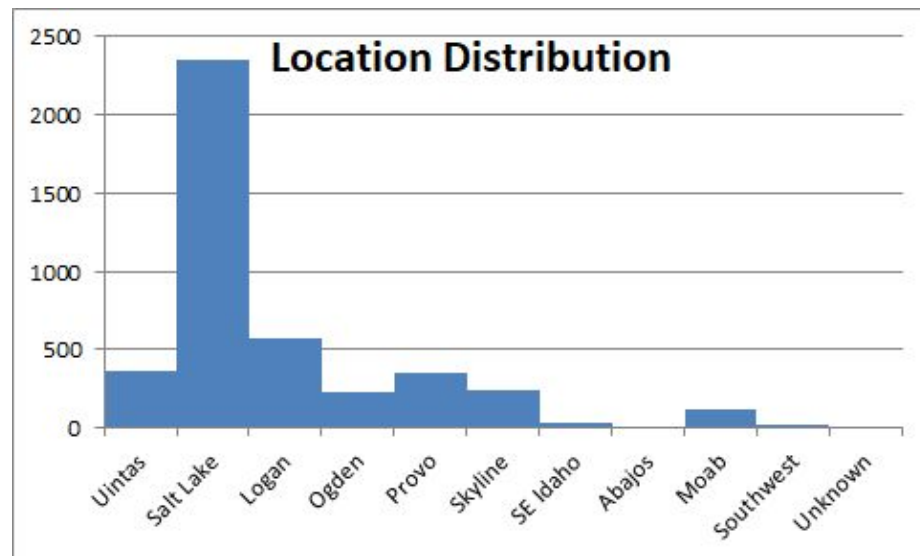
Data Processing

Our dataset is already relatively 'clean' however we do expect to deal with some erroneos values as the dataset incorporates crowdsourced observations.We will also need to decide how to handle missing values for the characteristics of interest, these appear to be sparse so the best tactic may be to omit these instances but we will verify that. We will be working with categorical attribute sets, as well as binned quantitative attributes. We will pre-process our data using R or Python to minimize the overhead needed for our visualization. This process will be as modular as possible so that future incorporation with an updated dataset will be possible.
- Cleaning/enforcing categorical values (if needed)
- Binning
- Normalization

Visualization Design

Our visualization focuses on exploring the change over time of the avalanche characteristics. We will provide the ability to select a desired attribute set (aspect, trigger, elevation, size, etc...) and visualize the distribution of observations within this selected category over time. This steers the user away from making inferences about the real distribution of all avalanches, and provides the benefit of describing how the characteristics of the selected avalanche observations has changed (over the selected time frame). The user will specify:
- Time frame (based on month and year)
    - The grid can be brushed to select multiple cells
- Category selection (aspect, trigger, size, elevation)

- We may choose to incorporate some category sub-selections

Must-Have Features:

- Stream/Area chart
- Attribute drop down to select what to display
- Heatmap or gid month/year selector
- Preset 'stories'

Optional Features:

- Year clock chart
- Scatter or graph view of how attribute frequency has changed over time
- Derived attributes
  - Weather
 - Forecasted avalanche danger (from archives)

Project Schedule

- Oct. 18
  - Team announcement completed
- Oct. 25
  - **Project Proposal** completed
- Nov. 1
  - Data cleaning completed (final csv) and Javascript data structures in place
  - Overall HTML layout in place
  - Process book: Overview and motivation, related work, questions
- Nov. 8
  - Area chart that appears on load
  - Heat map
  - All buttons, labels, and text on screen (even if not all functional)
  - Process book: Data, exploratory data analysis, design evolution
  - **Project Milestone** completed
- Nov 15
  - Interactive functionality and linked view functionality
  - Storytelling
  - Process Book: Design evolution, implementation
- Nov. 22
  - Finalize code and design aesthetics
  - Make project screen-cast video
  - Process Book: Analysis
- Nov. 27

- ○ Host on website
- ○ Finalize process book
- ○ **Project Completed**

Design Sketches

Sketch number 1: This sketch included an area chart of the observation distribution over time. This design incorporated a normalized and unnormalized view. We decided the unnormalized view may be misleading because of the reporting bias, so we will not include this in the final design. The user could select a display category as well as filter by location (by region and then more granularly place). Ultimately we decided that the location filtering wouldn't be as useful to visualize because such a large proportion of the observations are in the Salt Lake region and most of the observation "places" listed tend to only tend to have one or two observations associated with them.
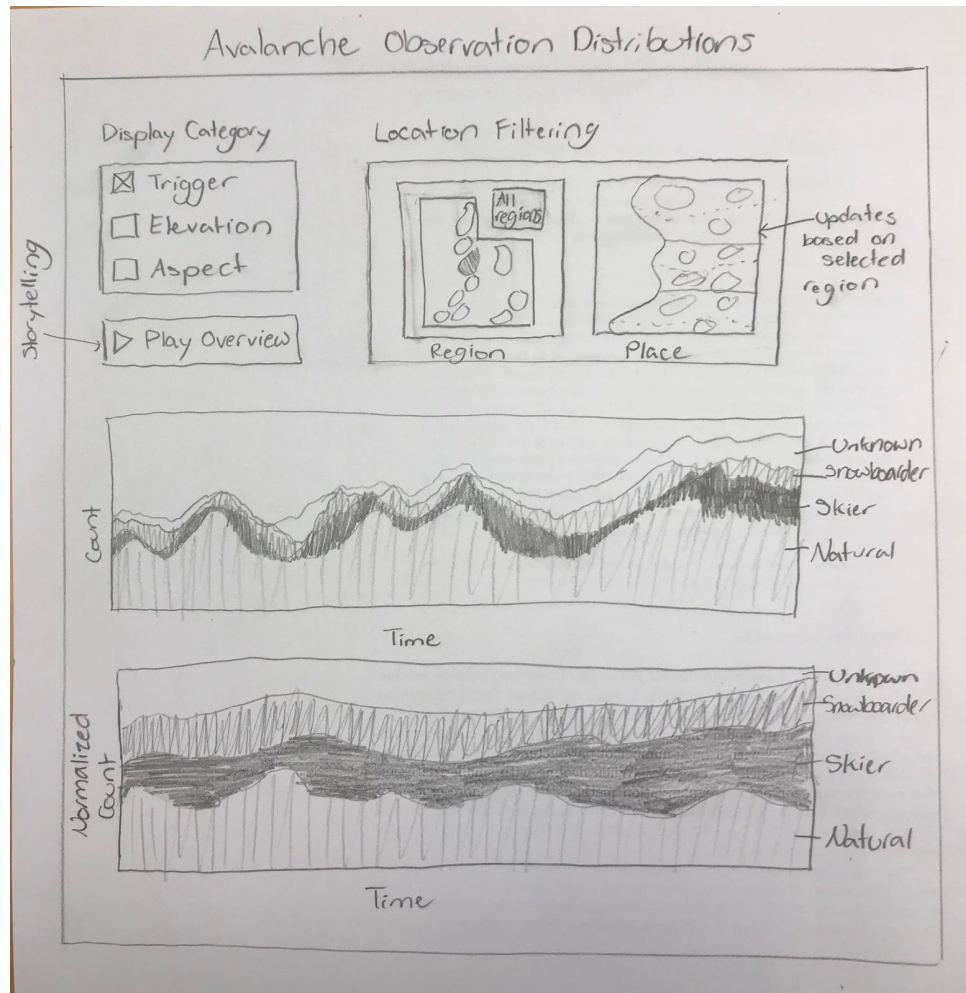
Sketch number 2: This sketch encorporated a similar area chart as the previous one but with the added feature that the user can zoom in on a specific month to see more detail. This design also includes a scatter-plot as another way to visualize the data shown in the area chart. This design also has a heat map of the months and years. We liked this idea because it provides a way to visualize

how the frequency of all observations has changed over time and provides a convenient way for the user to select which year and month to display in the area chart.
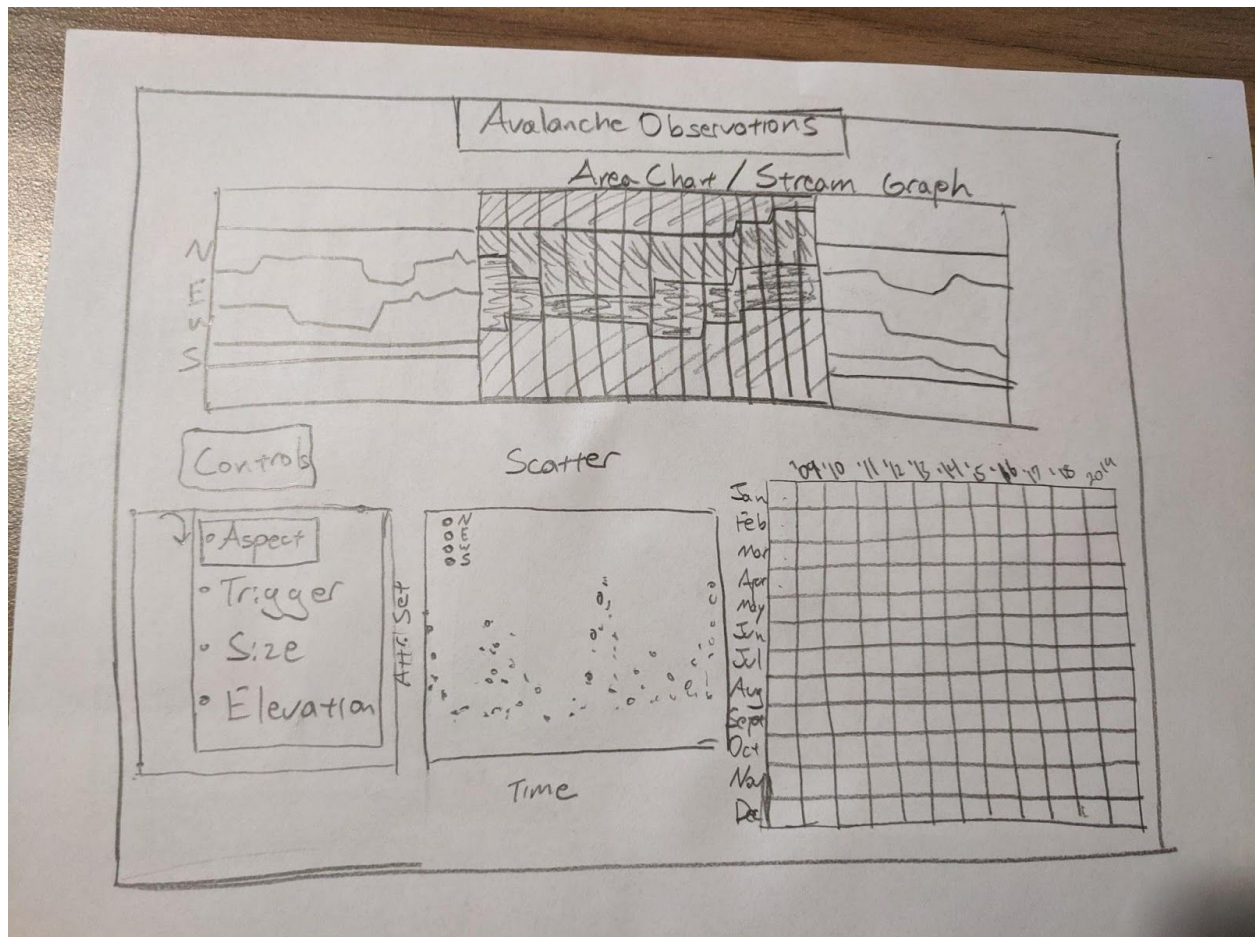
Sketch number 3 In this sketch we kept the area chart as well as the heat map, and we added a year clock inspired by Mortiz Stefaner's 'Rhythm of Food. The year clock would provide a way to see how the frequency of observation characteristics change in a given year without the constraints posed by a classic stream map or area chart. Given that this data is cyclical with the seasons, it makes sense to encode the frequencies in a cyclical way. The additional attribute filter was intended to provide the option to select a single attribute from selected attribute set, and isolate it's change over time. This ended up not being part of our final design.
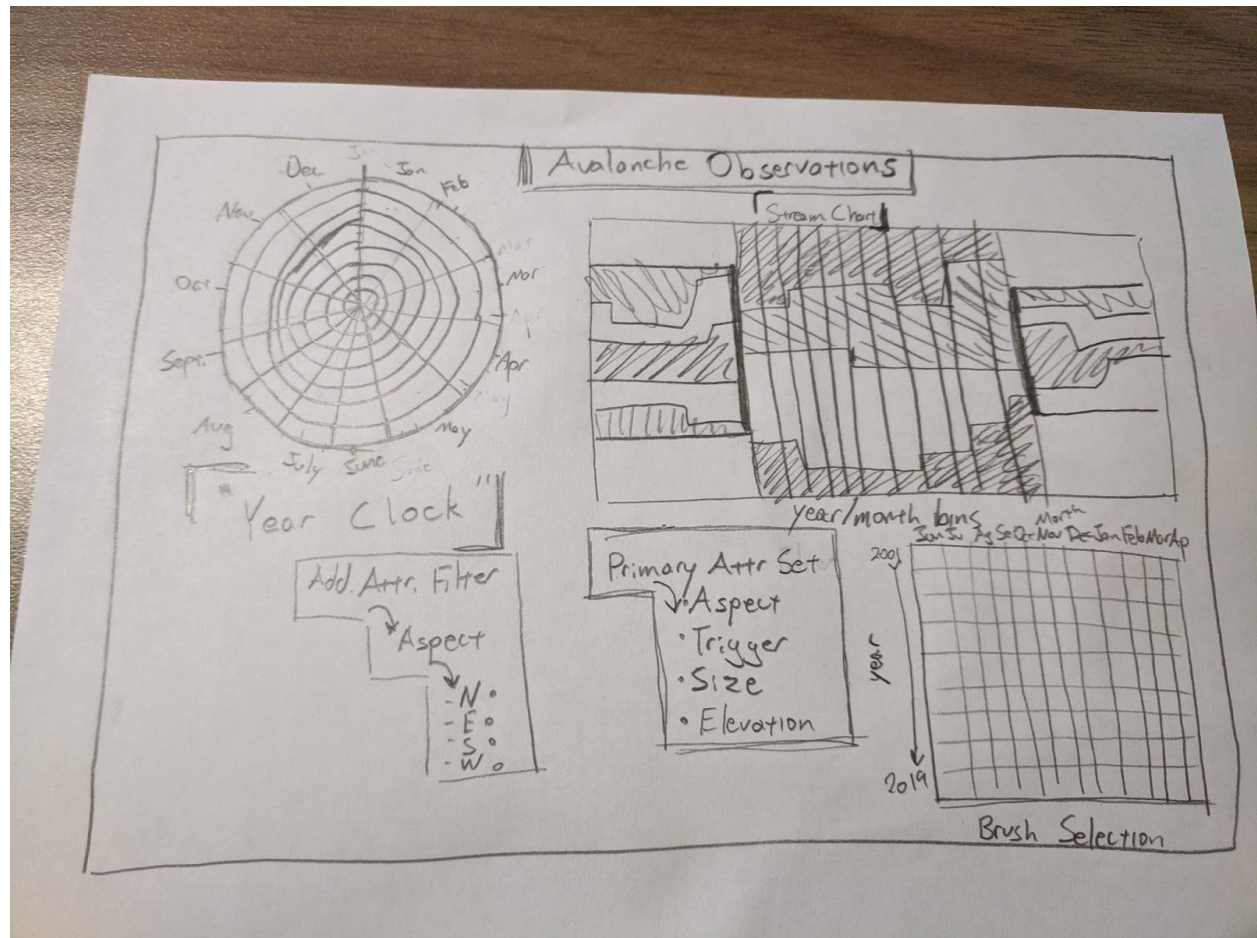
Final Design: Our final design combines the area chart with the heat map view. The user will be able to select which attributes they want to visualize using the drop down. They will also be able to select the year, and month that are displayed using the heatmap which will also indicate the number of recorded observations within each cell encoded by saturation. An area chart/stacked bar chart is an appropriate form of visualization because we are interested in showing parts of a

whole. This will enable the user to see which categories are most often reported and how those distributions change over time. The heat map is an appropriate way of encoding the total number of observations because it will allow the user to visualize the cyclical properties of the data and get an idea of how these counts have changed throughout the years. The story and provenance section will include both a data-driven written description of what attributes are being displayed and how the data is filtered, as well as an overview of interesting points and storytelling.

Avalanche Observation Characteristics

Stacked bar (distribution of whole and steam graph on outsides)

Attribute Dropdown
Aspect
Trigger
Elevation
Size

Storytelling
And presets

Year vs. Month grid (heatmap)