Utah Avalanche Observations - Characterized

Process Book

Jadie Adams and Max Marno

Table of Contents

Project Repository: https://github.com/jmaxmarno/avy-obs

Overview and Motivation

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

As winter recreationists we have spent time pouring over 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.

Our visualization illustrates how the distribution of these observations has changed over the last decade. Our project enables 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. As we are aware of the presence of reporting bias, our design focuses on faceting the observations into attribute sets, and displaying the normalized distributions of these attributes. Previously, interested users could only export this data as CSV or explore it in a table format. We 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 has both user-driven interactive components as well as guided story-telling aspects to highlight points of interest. It allows 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.

Say something about collaborating with UAC if that happens

Related Work

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?
We want to give users the opportunity to explore a rich dataset of crowdsourced avalanche observations that is otherwise difficult to access. Considering our dataset and the presence of reporting bias, we chose to visualize the attributes as proportions of the whole. By normalizing the attribute sets in this way it is less likely that someone may incorrectly infer a trend that does not exist. Ex; 'Look, the number of avalanches triggered by skiers went way up!'.
We emphasize the fact that these observations only represent a sample of avalanches, and provide as much insight as to where people are traveling, as much as where avalanches are occurring.
However, the visualization allows one to answer questions such as:

- How does the distribution of avalanche observations change over aspect throughout the winter?
- Were observed low elevation avalanches more prevalent in one year compared to another?
- Has the proportion of observed skier triggered avalanches changed throughout the year?

Data

Source

Source, scraping method, cleanup, etc.

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.

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.
For our stacked bar visualization we needed to bin/discretize the continuous variables; width and elevation. For width we used the following scale:

- Large (width > 600)
- Medium (300<width<600
- Small (300>width)

For elevation we followed a common tier commonly used in avalanche forecasting

- Above 9,500ft
- 8,000ft - 9,500ft
- Below 8,000ft

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

A pre-processing script developed in python takes the raw .csv file and converts it to an appropriate .json file with the appropriate nesting.
The structure looks like this:

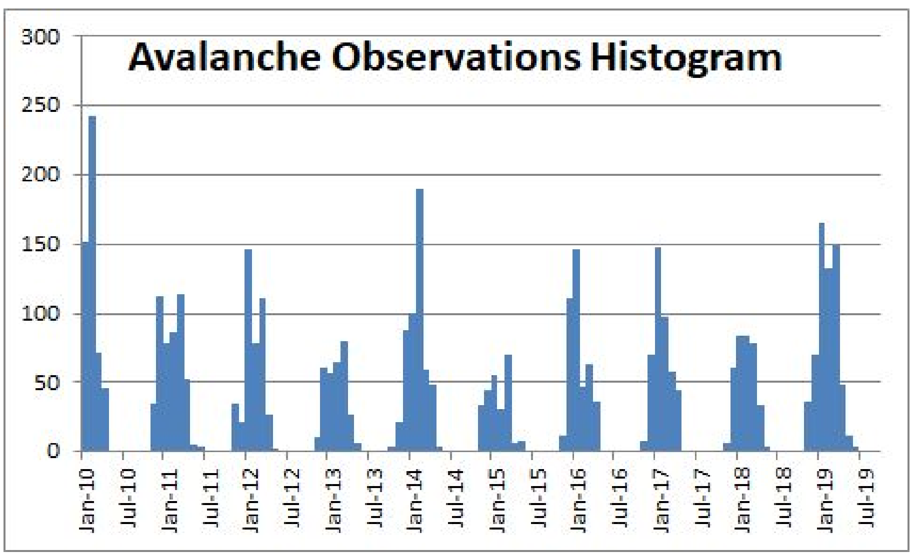
```
'''
JSON structure:
{
year{
  month{
    total_count
    region {}
    trigger {}
    weak_layer {}
    size {}
    aspect {}
    human_stats{
      caught
      carried
```

```
            burried_partly
            buried_fully
            injured
            killed
            }
        }
    month ...
    }
year ...
}
```
Missing categories are marked "unknown"
Missing human stats are assumed to be 0

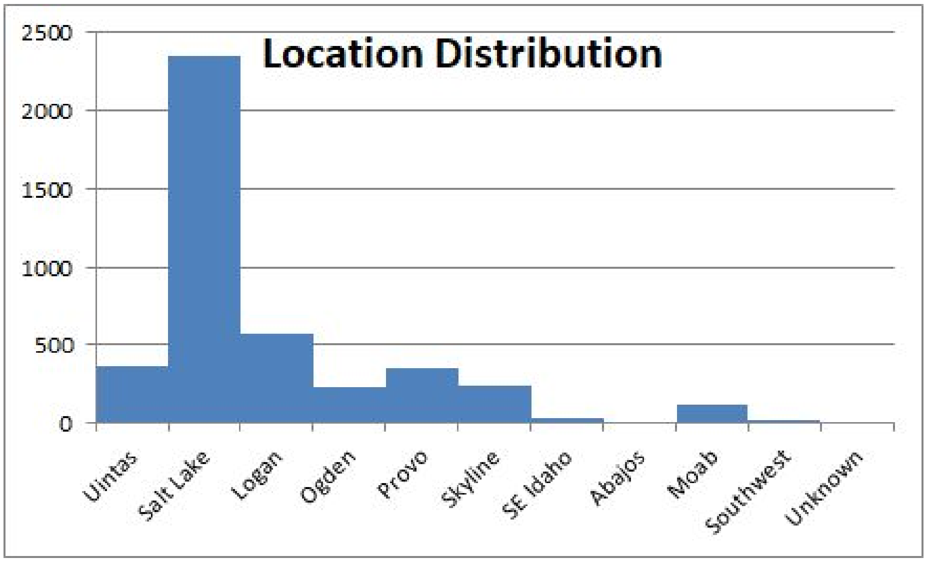
Exploratory Data Analysis

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

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.



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



Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

Proposal

As our visualization aims to allow users to explore the change over time of the avalanche characteristics, we wanted to 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). We proposed a design in which the user will specify:

- Time frame (based on month and year) selected via brushing grid
- Category selection (aspect, trigger, size, elevation and maybe some sub-categories)

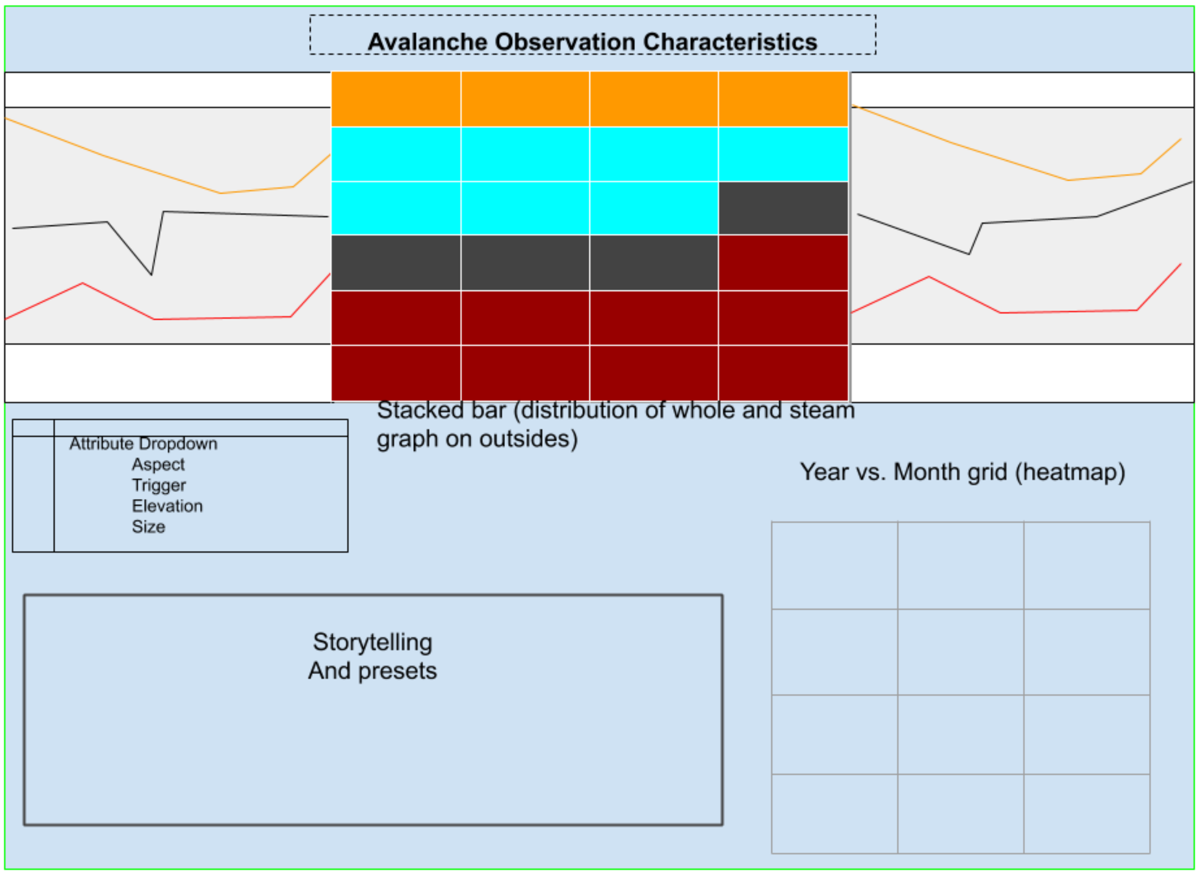
The must have features we defined in the proposal were:

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

The optional features we defined were:

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

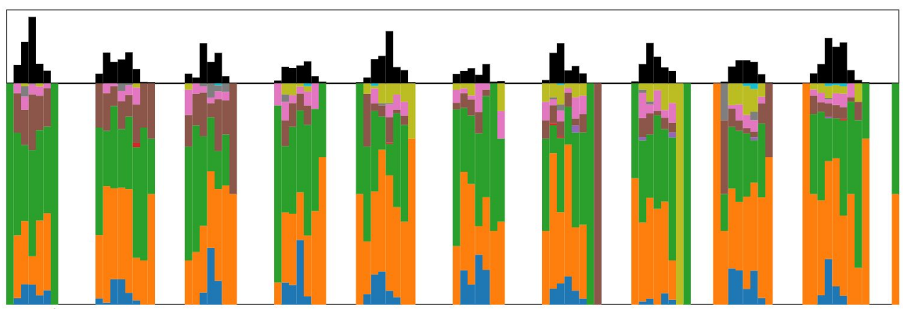
The following is the design sketch we decided on in our proposal.



Our proposal design combined an area chart with the heat map view. The idea was user will be able to select which attributes they want to visualize using the drop down and select the year and month displayed using the heatmap, which will also indicate the number of encode observations within each cell by saturation. We decided that an area chart/stacked bar chart is an appropriate form of visualization because we are interested in showing parts of a whole and it enables the user to see which categories are most often reported and how those distributions change over time. The heat map is also 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. We also chose to include a story and provenance section with 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.

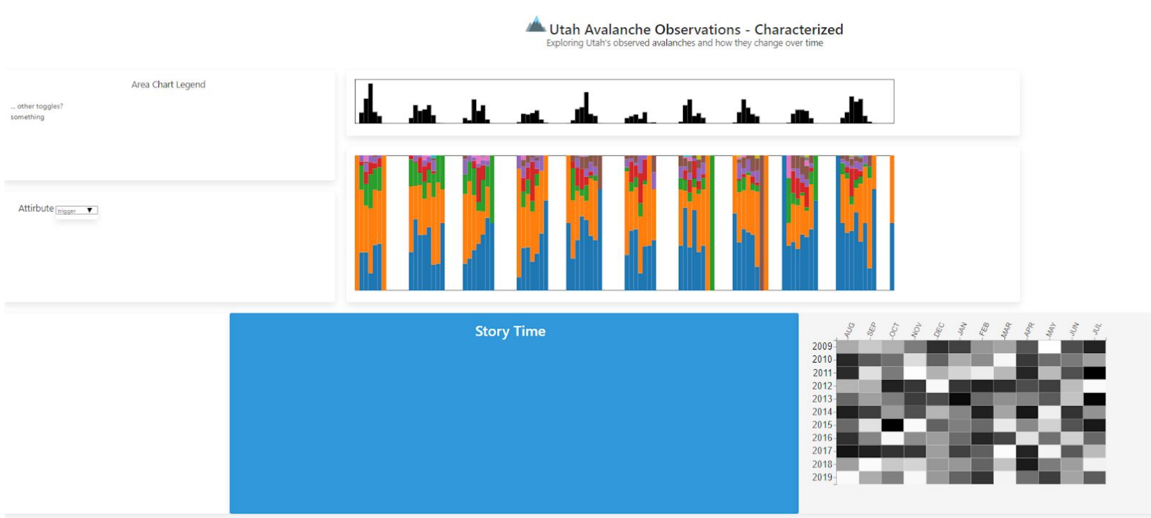
Milestone

One of the first design changes we decided to make in implementing the milestone was to add a histogram of observations above our area chart as seen below.



We decided that the histogram gave necessary context for the normalized area chart about the total observation count. It helps illustrate, for example, why there are blank bands throughout the chart (summer months with no observations). It can also help a user understand that a proportion which seems extremely high may be an artifact of a total low observation count.

We also decided that the stacked bar design was a better choice than a stream graph for the zoomed out data because it shows finer detail without looking to crowded. Our milestone design look pretty close to our proposed visualization in the end.



Final Design

Implementation

Views

Area Chart

Describe area chart and functionality

Year Grid

Describe year grid and functionality

StoryBoard

Describe story board and functionality
Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?