Exploratory Data Analysis of Colorado Avalanches

Jackson Baker and Yusef Haswarey

The goal of this project is to perform exploratory data analysis on observed avalanche records obtained from the Colorado Avalanche Information Center (CAIC) as well as all recorded fatal avalanches in the United States and determine if there are any clear patterns in the data that will allow us to identify potentially dangerous situations. We will try to identify patterns to determine if certain conditions, areas, or activities are more dangerous than others. We will analyze two datasets:

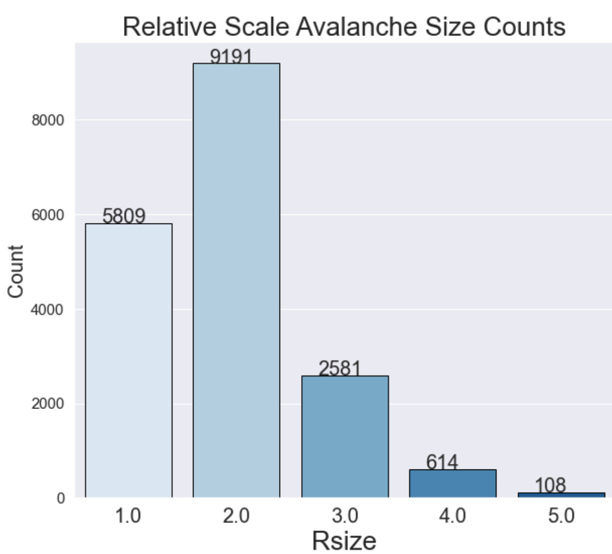
- The primary dataset includes all observed avalanches from 2010 to present day in Colorado.
- The secondary dataset includes all recorded fatal avalanches from 1951 to present day across the entire United States.

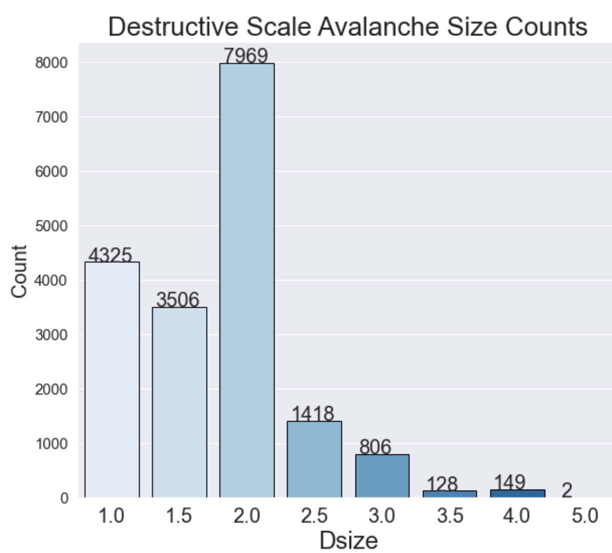
Literature Review

- Similar statistical analysis has been performed by the CAIC. Their public statistics can be found on their website: avalanche.state.co.us

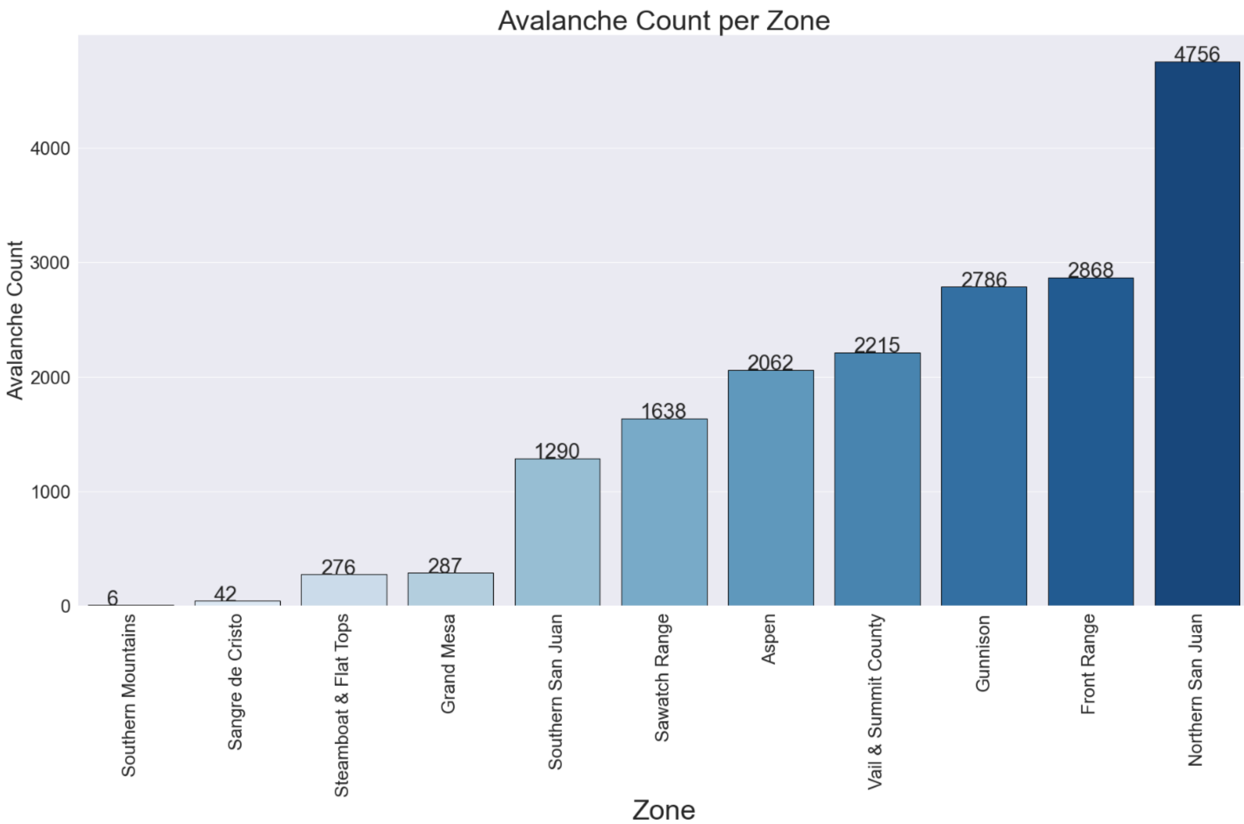
Primary Dataset

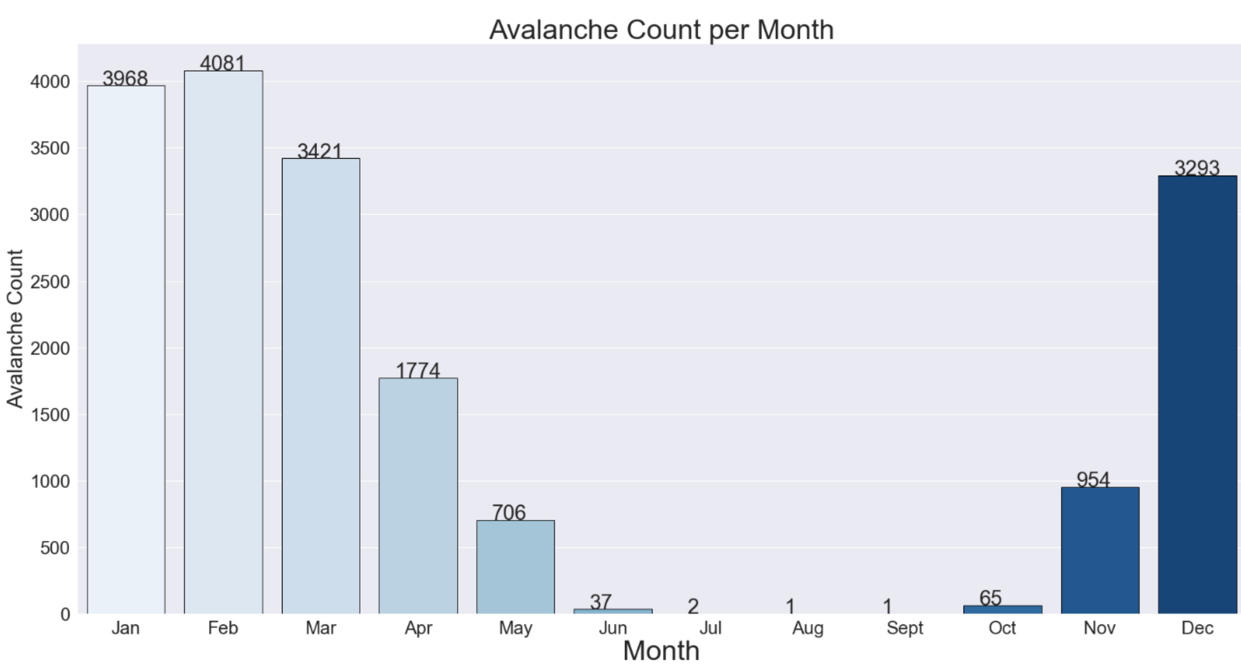
This dataset includes all observed avalanches in Colorado from 2010 to present day. The size of the avalanche is denoted in two different formats, relative scale and destructive scale. Relative scale started as the standard classification system for the United States and describes the size of the avalanche relative to the avalanche path. 1 is very small relative to the path and 5 is the avalanche covered its path entirely. The destructive scale was the standard classification system for Canada but has since been adopted in the US as well. It describes the destructive potential of an avalanche with 1 being relatively harmless to people to 5 being capable of destroying a huge amount of landscape or even entire towns. We graphed the counts of both scales to determine the distribution.



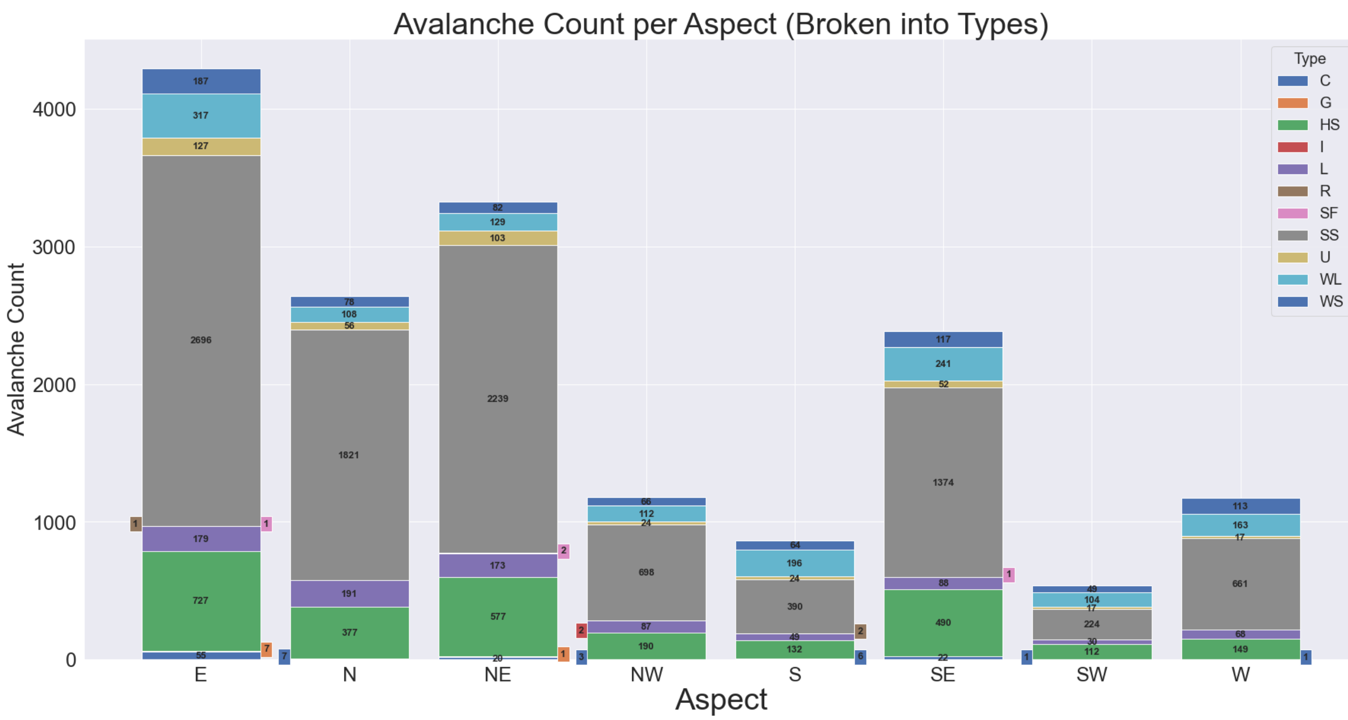


Both scales are right skewed with a size 2 avalanche as the mode. It is to be expected that smaller avalanches occur more frequently than absolutely massive ones, so no red flags in the data thus far.





Next, we looked at the most dangerous Colorado zones, or more specifically the zones that see the most avalanches. Zone size is variable and may contribute to higher avalanche counts in those areas. The Northern San Juan mountains have by far the most avalanches reported, and the more south-eastern mountains such as the Southern and Sangre de Cristo's have very few avalanches reported. As expected, the areas with high reporting are the areas that contain some of the best skiable terrain, and the areas with low reporting do not contain nearly as much ski country. We also considered avalanche count per month, although unsurprisingly the distribution shows more avalanches in the deepest winter months, falling off in spring as the snowpack becomes more stable, and dropping off completely in the summer as it becomes non-existent.

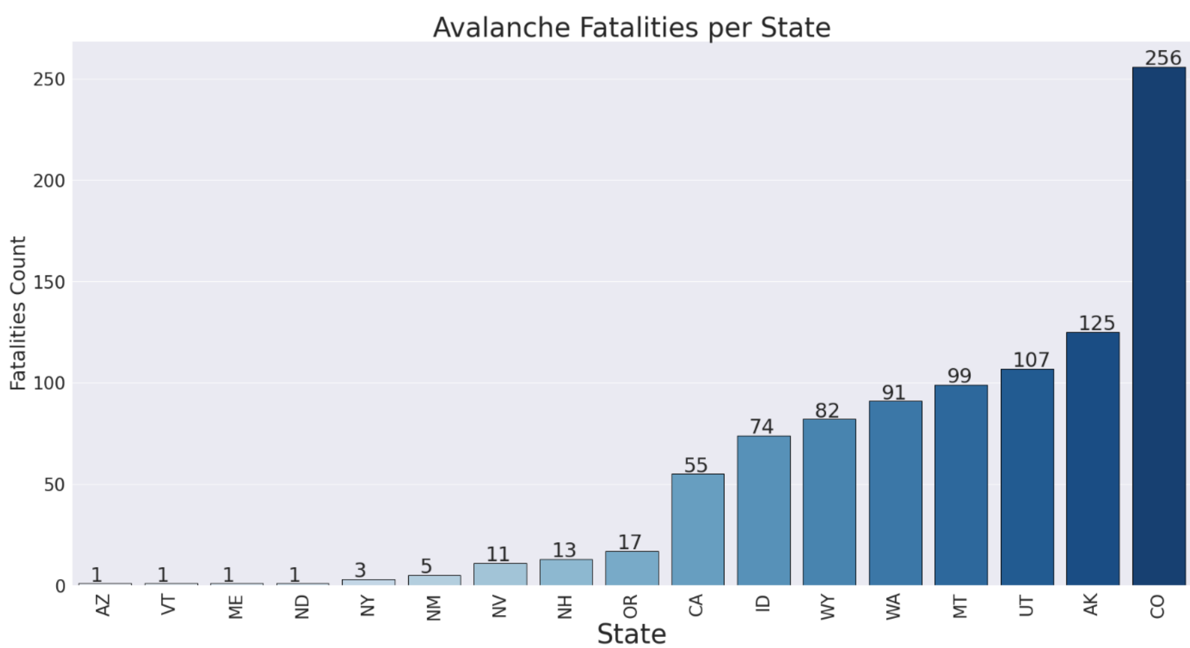


Another concern we investigated was if certain aspects yielded more dangerous types of avalanches more frequently (Storm Slab, Persistent Slab, etc..). This does not appear to be the case as the distribution of avalanche types appears almost completely even across all aspects. This is surprising considering the eastern and southern aspects in Colorado see more warming in the winter, which one would think would lead to first a more cohesive and safe snow layer, but then more loose wet and slough avalanches in the spring.

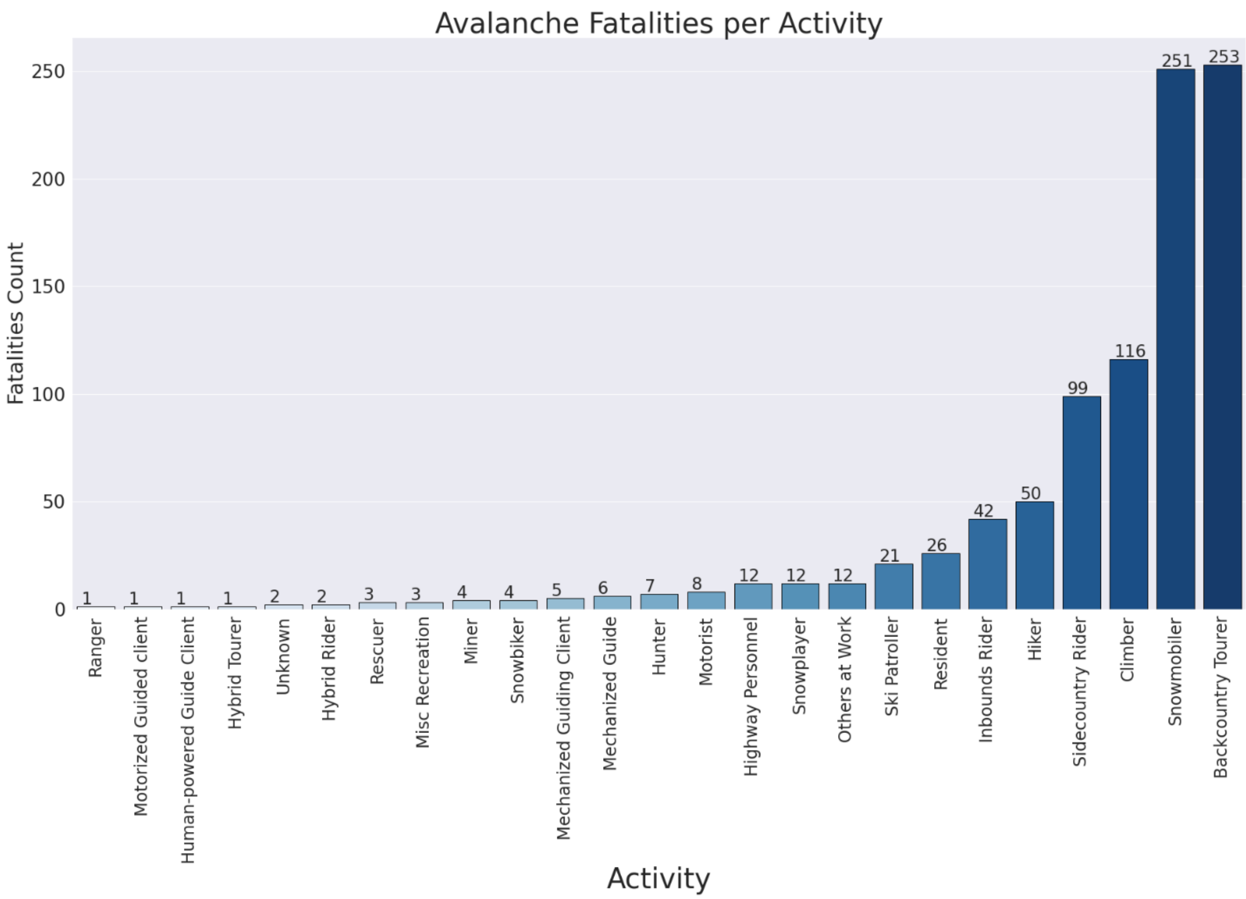
Finally, we looked to the description section of the avalanche report fields and created a lemmatized wordcloud to see which words people used most frequently in their reports. Unsurprisingly, words like avalanche, skier, and aspect were used frequently. Also many descriptors such as crown (the top of an avalanche where it breaks from the snowpack), bowl, chute, gully, and ridge (terrain descriptors) were also used.

Secondary Dataset

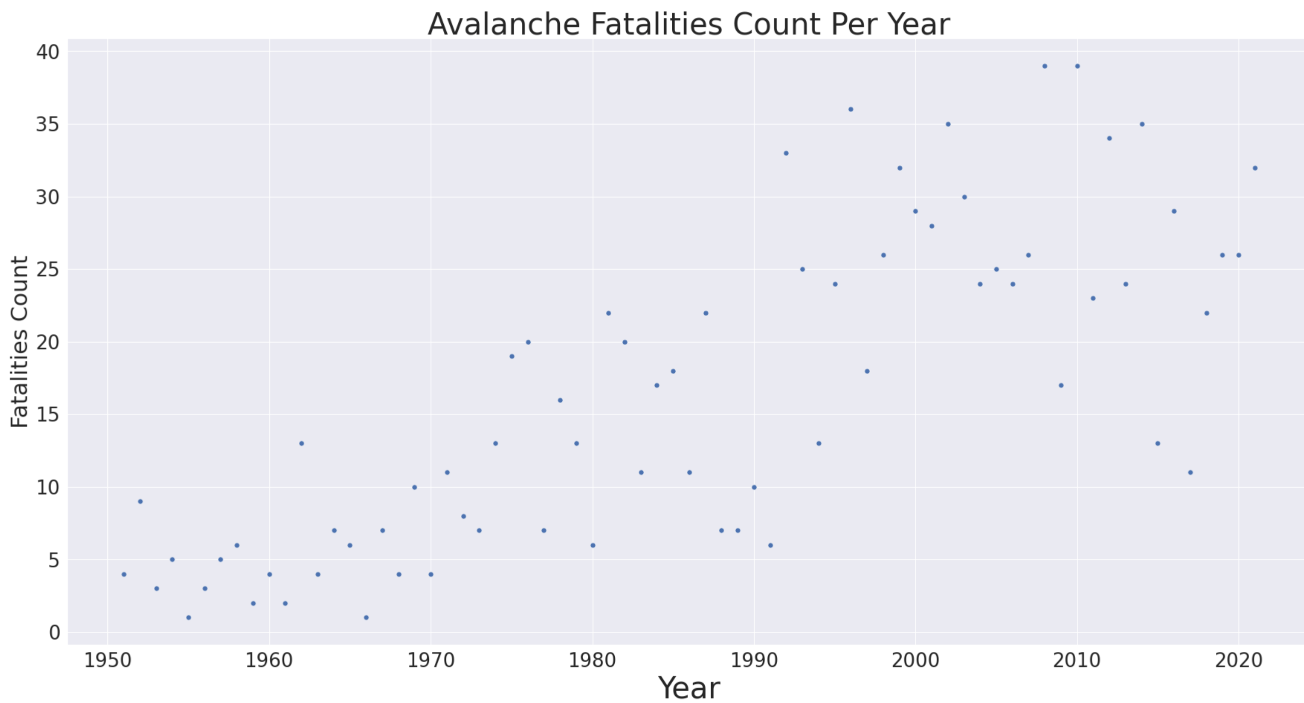
This dataset is a collection of all reported avalanche fatalities across the entire United States from 1951 to present day. First, we looked at the most dangerous State's and the most dangerous activities. Usurpingly for those who know a little bit about avalanche sciences, Colorado is the deadliest state. This is not just because Colorado is a hotbed for winter sports, but also because of our infamous snowpack. Colorado frequently gets very early season snowstorms, but the weather doesn't stay cold enough and continue snowing. This means that early season snow doesn't warm enough to melt, but it does warm enough to create facets and form a very dangerous layer. Later in the season when things cool down further and more snow starts to fall, it all rests on the rotten early season layer, leaving very dangerous conditions for persistent slab avalanches.

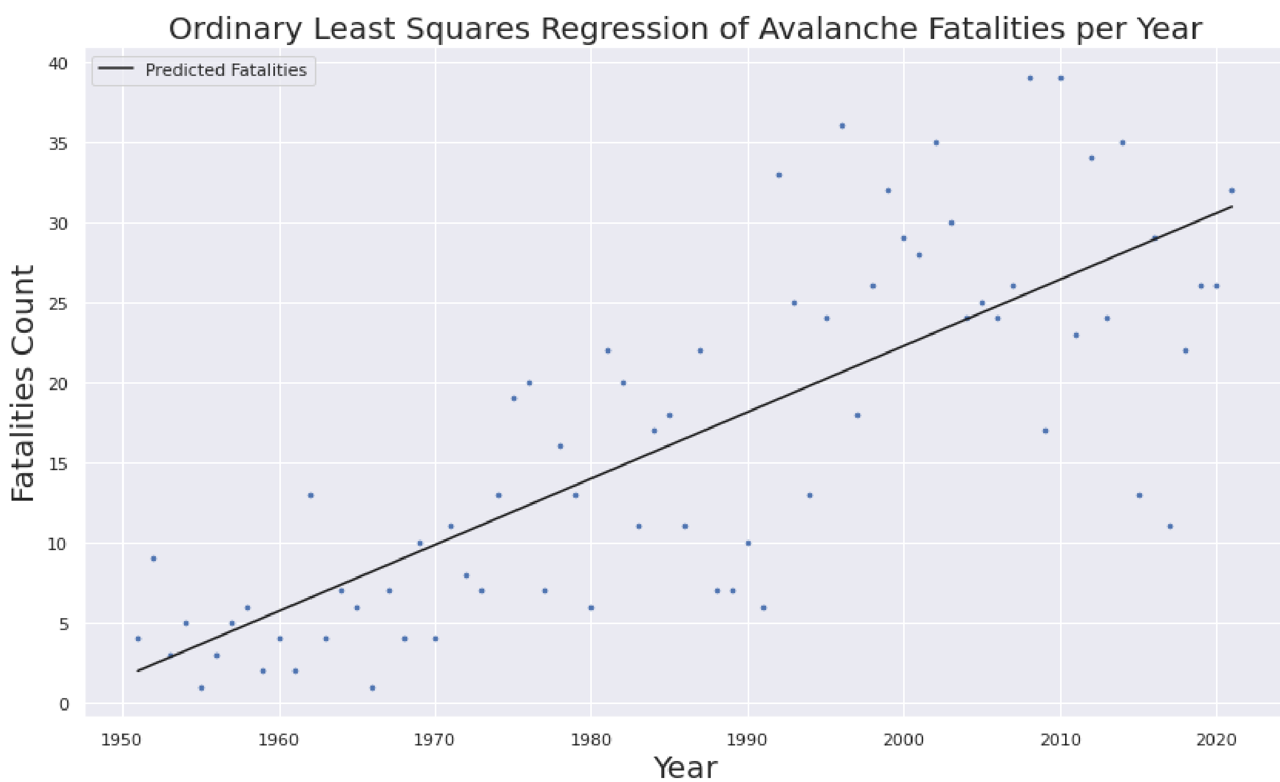


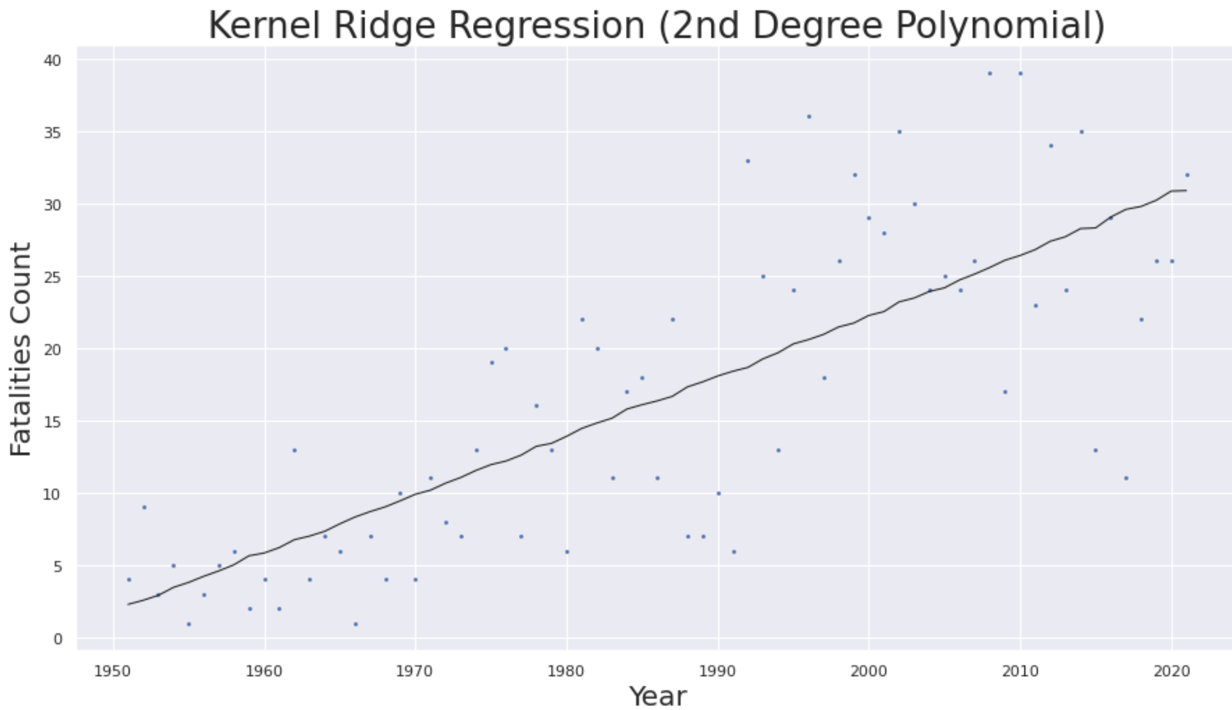
As far as activities go, backcountry touring and snowmobiling are the two most dangerous. This is likely because these two activities naturally involve taking the highest risks in avalanche terrain. Climbing and hiking also see their fair share of deaths, which some speculate is because while climbers and hikers don't venture into as dangerous of terrain, they also don't typically know how to spot the warning signs that backcountry skiers do.



We then looked to avalanche fatalities per year and discovered a positive linear trend. We tried OLS regression and a variety of kernel smoothing methods to try and predict future avalanche fatality trends. The Ordinary Least Squares regression had an R squared value of 0.6119 and a Root Mean Squared Error value of 6.756. Out of all the kernel smoothing methods (Linear, RBF, Polynomial, Laplacian, and Cosine), 2nd degree polynomial had the best RMSE of 6.767, which is slightly worse than the basic OLS regression. Overall, none of the methods provided a very accurate model for predicting future death toll.


Avalanche Fatalities Count Per Year
Fatalities Count
Year
0
5
10
15
20
25
30
35
40
1950
1960
1970
1980
1990
2000
2010
2020


Ordinary Least Squares Regression of Avalanche Fatalities per Year
Predicted Fatalities
Fatalities Count
Year
0
5
10
15
20
25
30
35
40
1950
1960
1970
1980
1990
2000
2010
2020



Overall, a few trends were identified in avalanche activity. Colorado is the most dangerous state (by number of fatalities), with the Front Range and the Northern San Juans being the two most avalanche ridden zones in Colorado, and January and February the two most avalanche ridden months. Backcountry touring and snowmobiling are the two most dangerous activities by overall fatality count. Types of avalanches are relatively evenly distributed across aspects, although overall avalanche count is not. Eastern and northern aspects see far more avalanches than their southern and western counterparts. Avalanche fatalities have consistently increased over the years since 1951, however ordinary least squares regression and kernel smoothing methods failed to produce a very accurate model for future prediction. Avalanches are extremely unpredictable by nature and trying to predict when and where they will occur is very difficult.