Ross Brancati, Patrick Kelly, Brent Kohl
Data Visualization Project Deliverable 1
March 3, 2022

Part 1 - Description of the Dataset

For our project, we will be using a dataset that was found on github from Ewiseman. This dataset, ski_resort_stats, is publicly available and contains data about ski resorts around the United States. The dataset contains 364 items, where each item represents a ski mountain. There are 15 fields which describe the location (state, latitude, and longitude) and statistics of the mountain (base height, summit height, total vertical height, lift count, number of runs, total acreage, and number of trails by rating of difficulty). These descriptive statistics characterize mountains by their size, location, and difficulty, and are informative for making decisions about where someone may want to take a ski trip with their family or even move to if they are looking to live in a ski town.

In addition to this dataset, we can also merge annual snowfall data to inform skiers and snowboarders of potential destinations. We could also include data on proximity to airports to give an idea of ease of access. Another field of data that we could incorporate would be a rating of how the living conditions are in a particular area (cost of living, restaurants, summer entertainment, etc.).

Part 2 - User Identification

User 1 - User planning a long distance trip to a mountain
This application is targeted at users who are looking to plan a long distance ski trip to a single mountain or multiple mountains (either alone, with friends, or with family) in the United States. To define long distance, we will estimate either traveling via airplane or driving for more than 10-12 hours. Many skiers and snowboarders go on trips to destinations far from their homes every year, but the task of finding a good destination that meets their wants and needs is challenging. It involves individually looking up the mountain, cost of lift tickets or passes, trying to find lodging, booking airline tickets, and deciding which time of the year is best to ride at that particular mountain. Since skiers/snowboarders range in skill level, knowledge of mountains, and other factors related to the sport, it can be challenging to find a mountain that is right for the individual or group. Some people have already been on trips before, so they have their favorite places. However, others may have very minimal knowledge of where they should plan their trip to. Given that the mountains get bigger as you move out West, this user group is targeted at individuals who want to ride big mountains (for example, individuals who live in Colorado probably do not want to come to Massachusetts or Vermont on a trip like this).

These users have many goals that they want to achieve through planning the trip. The first decision they want to make is to choose the mountain that is going to provide the best riding experience. Fields that would come into play here would be terrain difficulty and annual snowfall. For example, a novice skier or snowboarder probably wouldn't want to travel to a mountain that gets 20 feet of snow a year and is 75% expert terrain. Other goals they would want to achieve would be ease of access (via airport, for example), or proximity to other mountains if they want to travel to multiple mountains in the trip.

Since there are many options of mountains to choose from, they would need to survey a large number of websites or maps to see which mountains are available to them, how close they are to airports, and the available accomodations. Without this visualization, it would be time consuming and tedious to do. However, our visualizations will allow them to select the important factors of their destination and make decisions without having to sift through multiple websites or tables of data. We imagine our visualizations will portray size, proximity to airports, terrain difficulty, and annual snowfall so that the user can select their own important criteria and make their decisions by looking at them.

There could be potential difficulties with our visualizations. There are many mountains in the dataset, so there will also be a large number of datapoints in the visualization. Weather is also a chaotic system (meaning it can change quickly and is not always

consistent), so annual snowfall may not be reliable. I also think that the user will have to have some sort of background information about mountains. For example, if a user is looking to travel to a big mountain, they should be looking out West somewhere because that is where the biggest mountains are. Without knowing that beforehand, they could be trying to scan the northeast for large mountains which they won't find. We could potentially solve this problem by including a brief text guide on where to look depending on the user's wants and/or needs.

User 2 - Family looking to buy a home in a ski town
The application can be used by a family looking for more information about a ski town they are considering purchasing a home in. Many well-off families have vacation homes, some families like to buy houses in a tropical climate and others like to purchase vacation homes in ski resorts. The task of finding the perfect town to purchase a winter vacation house that meets a whole family's needs is challenging. Each member of the family can have different levels of skill when it comes to skiing/snowboarding and they would want to find a mountain/resort that can satisfy every family member's skill level. In addition, the family may want to find a location that is not too far away from their main residence so as to avoid having to spend copious amounts of money on flights for each member of the family every time they want to visit their vacation home. Instead they could find a location that meets their needs and is only a few hours drive. This application's user group is families that want to buy a house in a ski town/resort that has a good distribution of green, blue and black level trails. A family would need enough lower level trails for the less experienced skier and snowboards to learn but also enough higher level trails for the more experienced.

The main goal of these users would be to find a resort/mountain that has enough higher and lower level trails for their family and is also not too far away from their main residence. Some fields that would come into play to determine which mountain would be the best fit is green/blue/black percent and the specific location of the mountain relative to a location that is provided by the user. In addition, a family may want to buy a house on a mountain that is on the larger side so they are not skiing the same couple of trails when they go on vacation so they would want to look at total acres of the mountains as well. Finally, the user could look at historical averages for snowfall by month for each mountain. If the user's family usually goes on vacation the most in the month of February they may want to look at mountains that have a higher historical average snowfall in that month.

Without this visualization these families would need to search state by state resort by resort to find a mountain that meets their needs. The user may have to go digging deep on each mountain's website to find specific figures on the percentage of each level trail.

This visualization would be able to provide that data to the user immediately. This visualization will portray mountain size, proximity to airport, proximity to a user provided location, terrain difficulty distribution, and annual snowfall.

There could be potential difficulties with our visualization. If a family is looking to buy a winter vacation home the only factor in their decision is not the quality of the ski resort. These families will be looking at cost, size, style, and many other house specific factors when deciding where to purchase their home. Our visualization can be an aid to these to narrow down their options to a couple locations that fit their needs and then they can research house specific requirements based on the locations they were given.

User 3 - Individual looking to recommend regions of the country to ski at multiple mountains
This application can be used by a person that is looking for regions of the country where they can go on a ski trip to multiple different mountains. Many skiers will plan trips and want to ski more than one mountain while in the area to get the whole experience. The task of choosing where to go on a ski trip is a tough one as you want the best experience possible and the ability to ski in more than one location. As many people go skiing with their family it is also important to determine if there is skiing for the whole family and not just the best skiers.

The main goal of this user is to find the best regions in the country for them to ski in to be able to ski multiple mountains that fit their level of skiing as well as their families. Some of the fields that would be useful for this are the latitude and longitude of each of the resorts and the percentage color of the trails. When the user wants to ski multiple mountains on a single trip, the density of the mountains can also play a factor. They will also want to look at the average annual snowfall so that they can plan their trip to a place that has snow.

Without a visualization of this data, they would have to look at mountains by state, and even then it wouldn't include surrounding states that might have mountains they are interested in going to. Having a visualization to help with this, would make it easy to see the density of mountains in each region and then you can narrow down the mountains based on trail types and snowfall.

Some potential difficulties that we could run into with this visualization is that there are other factors that are included in a ski trip such as the price they are willing to spend on it as well as how far they want to travel. Just because they can find the high density areas for skiing, they may not be near them or willing to spend the money to get to them. This

visualization can help the users gain the knowledge in order to find the regions that best fit their needs and determine where the next trip may be to.

User 4 - Weatherman who wants to analyze which areas of the country receive the most snowfall

This application can be used by weathermen to analyze the snowfall across different areas of the country. This can be useful as they can see the changes over time and compare different areas to their average snowfall. This is especially useful for skiers as they look to weathermen to tell them about the snowfall of the mountains that they look to ski at. Having information about the highest snowfall based on the areas may also correlate with the ski areas and mountains, giving more information to the skiers.

The main goal of this user is to correlate the snowfall of each area and be able to compare the areas. Having this ability can help them give more detailed information to their viewers with an easy to read visualization. The fields that will come into play are the snowfall, the latitude, longitude and altitude. Having altitude can help correlate if the snow is only falling on the tops of the mountains or if it will hit all the surrounding area.The more information that a weatherman has at their disposal is better as they can better inform their viewers.

Without the visualization, the weatherman would have to sift through tons of data to gain this information. They would have to look state by state and see if they have this information readily available and then collect it. The visualization would give them easy access to this data in a quick and easy manner. This visualization will provide the snowfall, location, and altitude.

The potential difficulties with this visualization is that the snowfall recorded might not look at the different altitudes when looking at the snowfall and may not be as specific as a weatherman might like. That being said, out visualization can make it easy to see this information at a glance when looking for this information.

3 – Task identification

User 1 Tasks - User planning a long distance trip to a mountain

Task 1:

A high level task of this user is simply picking a destination to travel to before planning a ski trip. The ski trip planner will have to choose a mountain that has the desired quantity of trails, that receives a lot of snowfall each year, and that is relatively easy to get to. To achieve this goal, they could look at the skiable acres of a mountain, the snowfall received each year, and the location of airports. We could potentially use a three dimensional chart to visualize this correlation, or label data points with text in a two-dimensional scatter plot. Another approach would be to include all of this data in a single datapoint on the visualization, such as distance to closest airport and the mountain statistics.

Task 2:
Another high level task of this user is to understand which regions of the country offer the most challenging terrain and how the terrain difficulty correlates with vertical elevation. To achieve this goal, we could create a visualization that separates the mountains by region or mountain range, and then correlate the regions with the amount of expert terrain and total vertical drop. We could also extend this to include the beginner and intermediate terrain. This would involve some data transformations or feature engineering, but overall the user would have an idea of which states they should target depending on the difficulty of terrain they would like to ski.

Task 3:
A third high level task of this user would be to plan a trip to a state or region where they can visit multiple resorts. To achieve this task, they would need to look at the density of ski mountains in a given state or region, along with calculating the distance to the nearest mountains. For example, if they wanted to visit 4 mountains on one trip, they should choose a mountain that has at least three other accessible mountains. They also may want to consider total skiable acres in their analysis, so that data could also be included. For example, within 100 miles of Denver, CO there are *x* number of skiable acres. To visualize this, we could plot the region or state with the proximity of the mountains to a centralized location in the state, such as an airport, and also include the total number of mountains or skiable acres in that particular area.

User 2 Tasks - Family looking to buy a home in a ski town

Task 1:
This user wants to choose a mountain that is large and one that has a good distribution of different terrain trail levels. To achieve this goal we could look at total acres of each resort and the percent of each trail difficulty on the mountain. We could eliminate all

resorts that have a certain percentage of one trail type. For example, if a resort has 50% or more of a given trail this may not be the trail difficulty distribution that a family is looking for.

Task 2:
This user wants to find a mountain that is within a certain distance from a location they provide and get a certain amount of snowfall per year. For this task we would look at the location of each resort and compare it to the location the user provides. Also, we would color these locations based on the amount of snowfall they have per year. To visualize this we could plot each resort on a map of the United States and show its relative distance to the location that the user provided. We would then color each resort on the map based on the annual snowfall.

Task 3:
The user wants to find a breakdown of which states have the most acres of black level trails. The user's family may all be expert skiers and snowboarders and they want to find the most advanced mountain to purchase a home near. To do this we could sum all the black level trail acres for each state. To show this to the user we could use a bar chart where the x axis is each state and the y axis is the number of black trail acres. We could also take a map of the US and use a different color saturation depending on the number of black level trails there are.

User 3 - Individual looking to recommend regions of the country to ski at multiple mountains

Task 1:
A high level task of this user is to find the most densely populated areas of mountains so that it is easy to travel from one mountain to another. The user looking or recommending regions to go skiing in will be able to look at the visualization and see what regions are the best place for this type of trip. We could potentially use a map of the United States as a chart, showing the density of the mountains using a heatmap and give information about the trails via hovering over the area on the map.

Task 2:
Another high level task is to correlate the average area of a mountain in each given region so that the users can see the amount of skiable area in each region. Having this information will tell the user how large the skiing areas are helping them decide where to go. We could use a heatmap for the area of the mountains using position and color to show the area of skiable areas in each region.

Task 3:

A third high level task of this user is to find Multiple mountains in a region that have the most mountains with the highest level of the type of trail that they like to ski. To show this, we can use a visualization that can use size and color to show the number of mountains in the area and the number of each type of run. This would make it easy to see both the number of a certain type of trail and the number of mountains in a certain region.

User 4 - Weatherman who wants to analyze which areas of the country receive the most snowfall

Task 1:
One high level task for this user would be to analyze the snowfall data in order to find the areas of the country that have the highest snowfall. To do this we can create a visualization of each snowfall in each region and use a heat map model to use color to denote the amount of snowfall in each location. Using this method it would be easy to see the different snowfall levels across the country.

Task 2:
Another high level task for this user would be to track the difference in snowfall between different altitudes and regions. To do this we could correlate the location and snowfall mixed with altitude to see if the altitude has an impact on the snowfall. To do this we can use a topographical map with a snowfall heatmap overlay and give the user the ability to compare the altitude to snowfall and different regions.

Task 3:
A third high level task for this user would be to see if the snowfall on skiing areas is higher than other areas. To do this the user would need to be able to see where the location of the snowfall is and where the location of each of the skiing mountains. To do this, we would create a visualization with points for each of the ski mountains and a heatmap of snowfall over the country. This would give the user easy access to see if the ski mountains have snow and give them the ability to report on it.

4 – Data and Task Abstraction

User 1 Tasks - User planning a long distance trip to a mountain

Task 1:
Attribute 1:
- Name: acres
- Meaning: total skiable acres at the mountain

- Type: quantitative
- Range: 7295 acres, mean: 598 acres
- Further notes: describes the total available surface that one can ski on, which is indicative of the size of the mountain

Attribute 2:
- Name: annual_snowfall
- Meaning: average amount of snow at the mountain per year in inches
- Type: quantitative
- Range: 596 inches, mean: 247.5 inches
- Further notes: data for annual snowfall is not available for every mountain, so we will not be able to provide this datapoint for some destinations

Attribute 3:
- Name: airport_proximity
- Meaning: distance from nearest airport to mountain
- Type: quantitative
- Still trying to merge databases to figure out closest airports
- Further notes: we may need to generalize this to airports in the state because the location of the nearest airport will be challenging to find for this attribute

Attribute 4:
- Name: state
- Meaning: the state that the mountain is in
- Type: categorical
- Cardinality: 38
- Further notes: not every state in the country has a skiable mountain which is why the cardinality is only 38 for this attribute

For this task, we will need a chart that marks the position of each ski mountain. Each mark can be labeled with the attributes listed above including acres, annual snowfall, and proximity to an airport. We may attempt to overlay these places on a map, which will give the user a sense of what state the mountain is in. To calculate the distance to an airport, we will need to transform the data from latitude and longitude to a distance in miles. We will need additional data that is not provided in the dataset such as annual snowfall and airport location. Using Munzner's taxonomy, we can classify this task as exploring features because the user will be looking at many potential locations for their trip.

Task 2:
Attribute 1:

- Name: state
- Meaning: the state that the mountain is in
- Type: categorical
- Cardinality: 38
- Further notes: not every state in the country has a skiable mountain which is why the cardinality is only 38 for this attribute

Attribute 2:
- Name: vertical
- Meaning: total vertical elevation gain of the mountain in feet
- Type: quantitative
- Range: 4510 feet, mean: 1177.5 feet
- Further notes: it would also be interesting to plot a histogram of elevation

Attribute 3:
- Name: black_acres
- Meaning: total amount of acres of expert rated terrain
- Type: quantitative
- Range: 3480 acres, mean: 226.18 acres
- Further notes: some mountains don't offer expert rated terrain and other offer mostly expert terrain, which explains why the range is so high

For this task, we can create a scatter plot with the vertical attribute on one axis and the black_acres on the other axis. We may also label the points with the mountain name in case a skier wants to know which mountain is where in this plot. We can also add a line of best fit and generate a regression equation to quantify the correlation between these two variables. Another plot could be a bar plot of the state and its respective number of black_acres. This would provide a sense of which states offer the most challenging terrain, and the user can then derive the regions of the country that offer this. For Munzner's taxonomy, this task is about exploring correlations between vertical drop and challenging terrain. Another taxonomy is discovering a distribution of which states/regions offer the most challenging terrain.

Task 3:
Attribute 1:
- Names: lat and lon
- Meaning: latitude and longitude of the mountain
- Type: quantitative
- lat range: 64.92, lat mean: 58.97 | lon range: 81.67, lon mean: -93.62

- Further notes: the US lies in negative longitude ranges, so we will have to be careful when plotting these values and add an annotation to ensure the user understands where 0 longitude is located

Attribute 2:
- Name: acres
- Meaning: total skiable acres at the mountain
- Type: quantitative
- Range: 7295 acres, mean: 598 acres
- Further notes: describes the total available surface that one can ski on, which is indicative of the size of the mountain

For this task, we could generate a new attribute that represents the total skiable acreage within a certain distance of the mountain that the user originally chose. For some mountains, this number may just be the total number of acres at that mountain because there are not any mountains within a given distance. However, in areas where the density of ski mountains is greater, this attribute would be representative of how many acres of terrain are within a given distance of the chosen mountain. This would allow the user to make a decision on if they want to ski more acres or different mountains on their trip. For Munzner's taxonomy, this is considered deriving features because we are creating a new attribute that is a feature of every mountain.

User 2 - Family looking to buy a home in a ski town

Task 1-3:
Attribute 1:
- Name: Lat and Lon
- Meaning: The latitude and longitude of the mountains
- Type: quantitative
- Range: 64.92, lat mean: 58.97 | lon range: 81.67, lon mean: -93.62
- Further Notes: As there are negative longiitudes in the US, we will make sure to denote this so that the users are well aware of the where each mountain is located.

Attribute 2:
- Name: annual_snowfall
- Meaning: The amount of snow that falls on each mountain a year
- Type: quantitative
- Range: 596 inches, mean: 247.5 inches

- Further Notes: As there is not snowfall data for every mountain, there will be some mountains with missing data points.

Attribute 3:
- Name: acres
- Meaning: total skiable acres at the mountain
- Type: quantitative
- Range: 7295 acres, mean: 598 acres
- Further notes: describes the total available surface that one can ski on, which is indicative of the size of the mountain

Attribute 4:
- Name: black_acres
- Meaning: total amount of acres of expert rated terrain
- Type: quantitative
- Range: 3480 acres, mean: 226.18 acres
- Further notes: some mountains don't offer expert rated terrain and other offer mostly expert terrain, which explains why the range is so high

Attribute 4:
- Name: green_acres
- Meaning: total amount of acres of beginner rated terrain
- Type: quantitative
- Range: 3480 acres, mean: 120.31 acres
- Further notes: some mountains don't offer beginner rated terrain and others offer mostly beginner terrain, which explains why the range is so high

Attribute 5:
- Name: blue_acres
- Meaning: total amount of acres of intermediate rated terrain
- Type: quantitative
- Range: 3480 acres, mean: 251.56 acres
- Further notes: some mountains don't offer intermediate rated terrain and others offer mostly intermediate terrain, which explains why the range is so high

For this user we could look at the latitude and longitude of the ski mountain/resort and determine its distance from a user provided location. Since this user is a family we would only want to look at mountains that are above a certain number of acres. Using these two sets of data will narrow down the number of valid ski resorts that the user will see. To further narrow down this search we could then introduce the number of acres of each

different trail level. The user could then see which trail difficulty has the highest number of acres per mountain and further narrow down their results. Finally, they could use the annual snowfall attribute to see which mountain gets the most snow during the time when they vacation the most. For Munzners taxonomy, we can classify this as exploration since the user will be searching for multiple suitable locations to purchase a vacation home.

User 3 - Individual looking to recommend regions of the country to ski at multiple mountains

Task 1-3:
Attribute 1:
- Name: acres
- Meaning: total skiable acres at the mountain
- Type: quantitative
- Range: 7295 acres, mean: 598 acres
- Further notes: describes the total available surface that one can ski on, which is indicative of the size of the mountain

Attribute 2:
- Name: annual_snowfall
- Meaning: average amount of snow at the mountain per year in inches
- Type: quantitative
- Range: 596 inches, mean: 247.5 inches
- Further notes: data for annual snowfall is not available for every mountain, so we will not be able to provide this datapoint for some destinations

Attribute 3:
- Name: state
- Meaning: the state that the mountain is in
- Type: categorical
- Cardinality: 38
- Further notes: not every state in the country has a skiable mountain which is why the cardinality is only 38 for this attribute

Attribute 4:
- Name: Lat and Lon
- Meaning: Tha latitude and longitude of the mountains
- Type: quantitative
- Range: 64.92, lat mean: 58.97 | lon range: 81.67, lon mean: -93.62

- Further Notes: As there are negative longiitudes in the US, we will make sure to denote this so that the users are well aware of the where each mountain is located.

For this user, we can use each of these attributes in order to find the location and therefore the region that a person would want to ski in. We would derive the density of the ski resorts so that the user is able to see how many resorts are in a certain region that they plan to visit. We can also use a similar derivation to find the density of resorts based on the acres because there are many areas that have plenty of ski resorts but they are all very small. If there are a lot of resorts but they are small the user may not want to travel to ski there. We can also overlay the snowfall over the regions in a separate visualization to show the amount of snowfall in each region as the snowfall can directly impact the quality of skiing. For Munzer's Taxonomy, these tasks would mainly be deriving features as we need to find the density based off of different attributes for many of the tasks.

User 4 - Weatherman who wants to analyze which areas of the country receive the most snowfall

Tasks 1-3:
Attribute 1:
- Name: Lat and Lon
- Meaning: Tha latitude and longitude of the mountains
- Type: quantitative
- Range: 64.92, lat mean: 58.97 | lon range: 81.67, lon mean: -93.62
- Further Notes: As there are negative longiitudes in the US, we will make sure to denote this so that the users are well aware of the where each mountain is located.

Attribute 2:
- Name: annual_snowfall
- Meaning: The amount of snow that falls on each mountain a year
- Type: quantitative
- Range: 596 inches, mean: 247.5 inches
- Further Notes: As there is not snowfall data for every mountain, there will be some mountains with missing data points.

Attribute 3:
- Name: vertical
- Meaning: total vertical elevation gain of the mountain in feet
- Type: quantitative
- Range: 4510 feet, mean: 1177.5 feet

- Further notes: Using Elevation can be interesting to see if there is a correlation between altitude and snowfall.

For this user, we can use the latitude and longitude to find the exact locations of the mountains and what regions they are located in. By using the location data and the annual snowfall of the mountains we can see the data and compare the snowfall across different regions of the United States. We can also create a new attribute of snowfall density by comparing the snowfall and how close it is together in the region. This can be good to show where in each region has the highest density of snow. As there are some mountains that aren't near others, the density will be much lower as there is less overall snowfall in the area. For Munzer's Taxonomy, this would fall under the derivation of features because it is creating a new attribute to compare different regions.

References:

[1] Eric Wiseman. *Ski_resort_stats.* Created: 2017. Accessed: February 2022.
https://gist.github.com/Ewiseman/b251e5eaf70ca52a4b9b10dce9e635a4

[2] Z Rankings (Christopher Steiner). *Annual_snowfall*. Updated: 2022. Accessed:
February 2022. https://www.zrankings.com/ski-resorts/snow