Rock Climbing Data Analysis

Michelle Gottlieb Chernov, Aaron Armbruster, Jasper Wu, Gokul Deep, Nicole Gonzaga

Abstract

Rock climbers require a lot of skill and time in order to master difficult climbs; however, the quality of a climb can determine whether a climber will be able to successfully traverse a climb. A good quality climb allows for a comfortable climbing experience which helps a climber to advance their skill set. Since many climbers fall into the category of weekend warriors having predetermined good quality climbs reduces the allocated time a climber searches for an adequate climbing location. By exploring the relationship between quality and other factors we can determine the most beneficial climbs a climber should travel towards. Through our findings we found that there is minimal correlation between the quality of a climb and other factors such as location, length, and type of climbing style. Quality of a climb is most likely related to the skills a climber has in order for them to rate each climb.

Introduction

1. Background

This data comes from the Mountain Project Website which is a website that hosts information on over 100,000 rock climbing routes. The purpose of the site is to allow climbers to find routes in their area and from all over the world. The routes are posted by users along with as much information as they can furnish to help other climbers find the routes and allow them to bring the proper equipment. Each route has a location, name, type, grade, length, and rating. Both the rating and grade are chosen through user votes. The csv file is from kaggle; where the kaggle poster wrote a python script from the scraped mountain project for the routes.

2. Aims

Many climbers fall into what are known as weekend warriors, meaning the only time they can get out to climb is the weekend. Oftentimes climbers wonder where and what grades they should aim for to get the most out of their weekend. To be able to climb the upper grade levels, it takes a lot of time and devotion. For example, going from climbing 5.12 to 5.13 might take multiple years worth of effort and specific training. With many jobs offering full work from home options, many climbers are leaving their expensive metropolitan areas with very little climbing to cheaper areas with tons of climbing. Using data visualization, we hope to explore the relationship between the quality of climbs and other factors in order to analyze which climbs would be most beneficial for weekend warrior climbers.

Questions and Goals

Difficulty and Quality
Question 1: How does the voting for quality vary with respect to the difficulty of the climb?
Approach: Plot a histogram of quality against votes faceted by the difficulty range

Climbing Type and Quality
Question 2: How many votes were observed relative to the quality and the route types traditional (trad) and sport?
Approach: Plot a histogram grouped by the type to see if there is a difference in voting across Quality.

Length and Quality
Question 3: Does the length of the route affect the quality ?
Approach: Plot a histogram of the length of the route versus its quality.

Location and Quality
Question 4: Where are the highest quality routes located in the U.S.?
Approach: Create a map to locate the climbs and differentiate their quality using a color scale.

Materials and methods

1. Datasets

Dataset 1: Three columns describing the variables we used by their Name, Description, and Type.

Name	Variable description	Type
Name	Unique name of each climb trail	String
Location	Pathway of climbing trail and surronding area	String
Quality	Rating of each specific climb	Numeric
Type	Identifier of sport or traditional climbing route	String
Pitches	Amount of pitches per climb	Numeric
Length	Length of the route in meters	Numeric
Latitude	The latitudinal coordinates of the climb	Numeric
Longitude	The longitudinal coordinates of the climb	Numeric
Protection	Gear required for the climb	String
Votes	Amount of people rating climb	Numeric
Grade	Difficulty of each unique climb	Float

Figure 1: Head of the dataset with each variable and its corresponding value

[19]:

	Name	Location	Quality	Type	Pitches	Length	Latitude	Longitude	Protection	Votes	Grade
0	Access Denied	El Mirador > El Potrero Chico > Nuevo Leon > N...	2.9	Sport	4	350.0	25.95044	-100.47755	12 draws + 60m Rope Take 22 draws if you wan...	22	5.10
1	Agave Nectar	Sugar Shack > Cougar Canyon (Creek) - CONSTRUC...	2.0	Sport	1	NaN	51.09642	-115.31767	4 bolts to anchor	1	5.10
2	Ant & Bee do Yoga	The Hen House > Kamloops > British Columbia > ...	2.7	Trad	1	NaN	50.57212	-120.13874	mixed~ gear to 4"	3	5.10
3	Besame Fuerte	Pilon De Lolita > Loreto Area > Baja Californi...	2.0	Sport	1	80.0	26.01097	-111.34166	bolts	1	5.10

2. Methods

Our exploratory analysis aimed at visualizing and exploring different relationships between the quality of climbs vs. other variables (number of votes, Location, length, and climbing technique), in order to determine the best quality climb. We visualized these relationships by creating distinct scatter plots, histograms, faceted histograms and faceted scatter plots. After conducting exploratory analyses, we decided to determine whether quality could be predicted from the variables Length, Location, Grade, and Type. Linear regression was performed to predict the quality of a climb. We consider two models, ignoring the variable Length in model 2 as most of the observations lacked values:

$$\text{Quality} = \text{Length} + \text{Grade} + \text{Location} + \text{Type}$$

```
Quality = Location + Type + Grade
```

Results

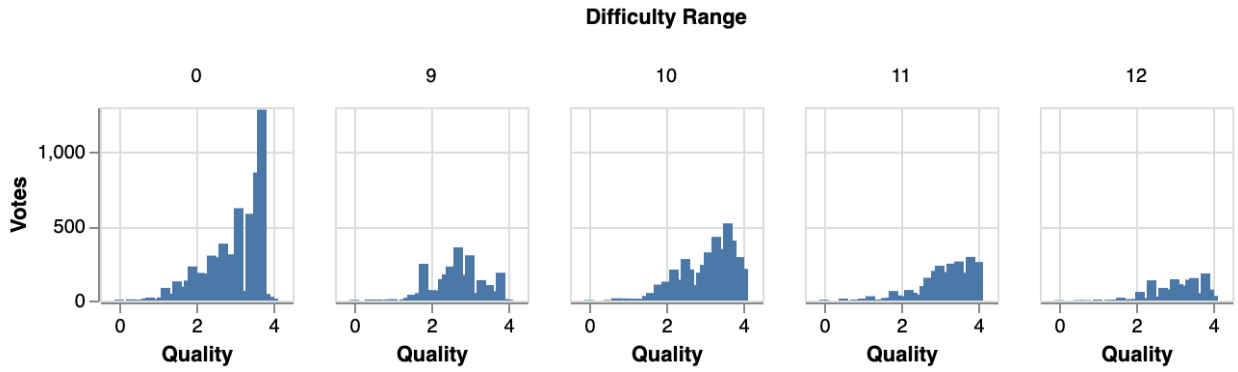

Figure 2: Quality of a Climb VS Number of Votes split into level of difficulty

Figure 2 compares the quality of a climb compared to the number of votes it received, primarily focusing on the higher difficulty level climbs ranging from 9 to 12. We can see that the more difficult a climb is the less number of votes, which means that less people have the skillset to complete these climbs. Therefore, we can infer that their quality rating is mostly based on their climbing skills.

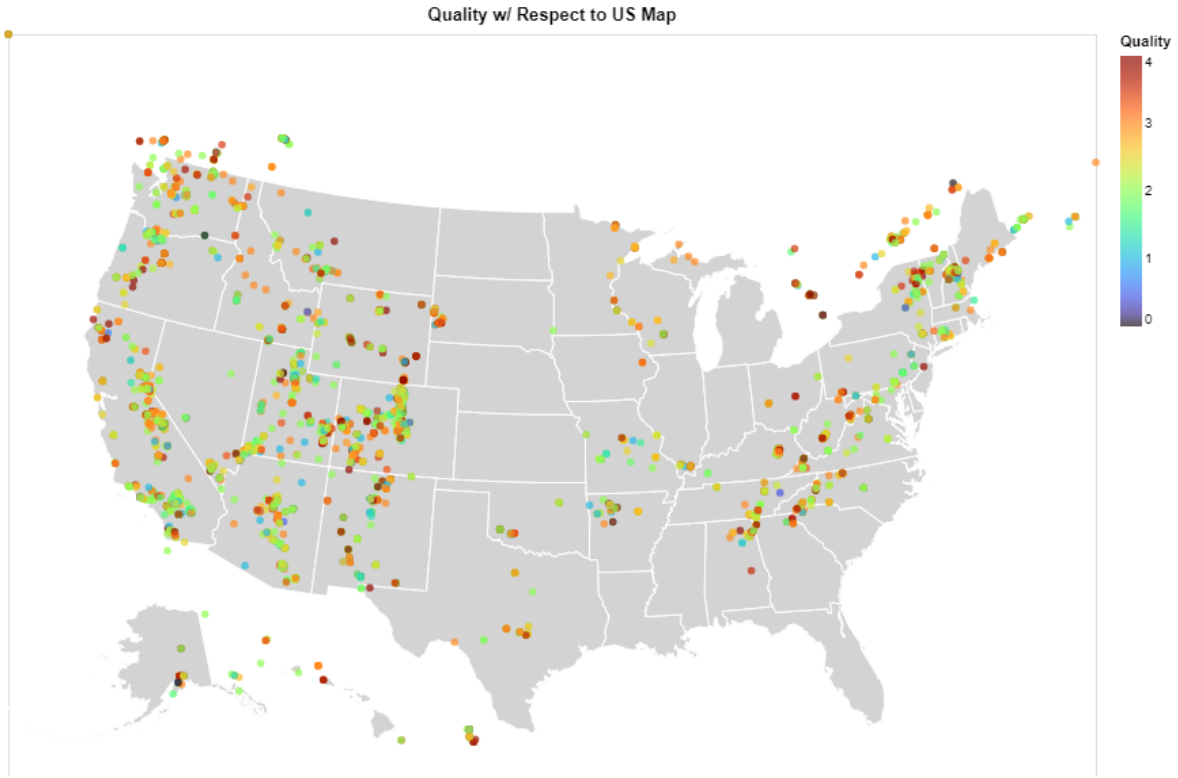

Figure 3: Location of Best Quality Climbs

Figure 3 represents the best quality climbs throughout the United States, with the maroon dots representing the best climbs. We can see geographically that the best quality climbs are in areas with large mountain ranges, and national parks. This would represent that most climbers prefer climbing in well traveled areas versus more unique climbing areas.

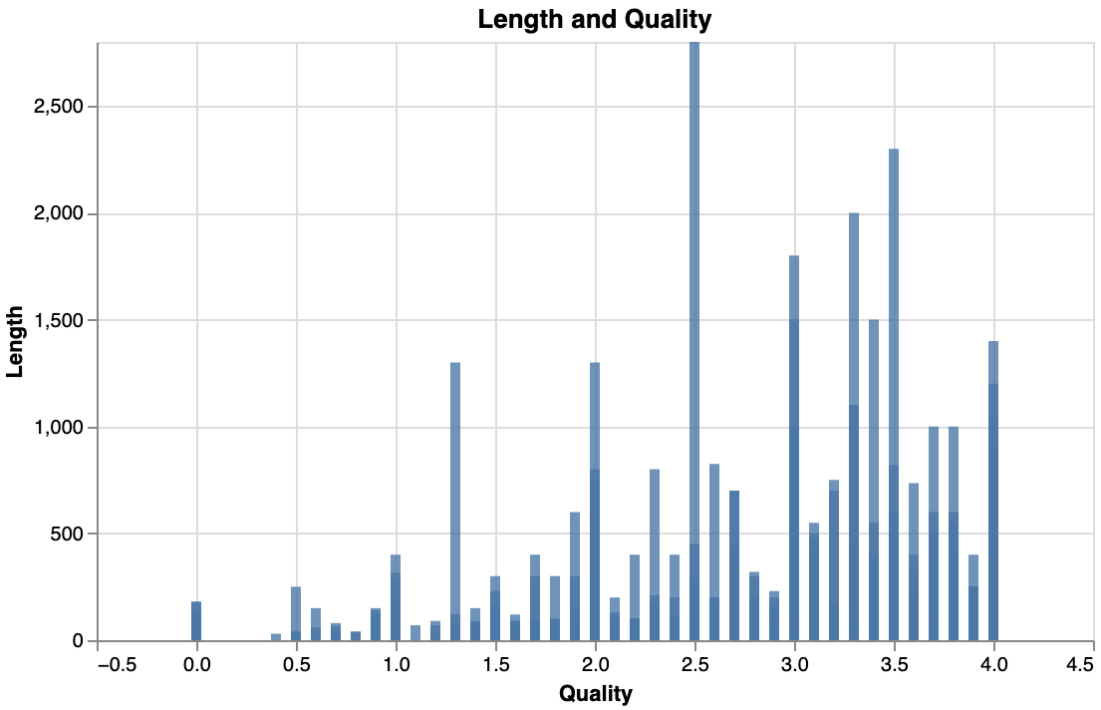

Figure 4: Length of a climb compared to the quality

Figure 4 represents whether a climb's quality is affected by its length. From the figure, we can see that the longer length of a climb is relatively higher quality. This implies that climbers prefer longer length climbs rather than shorter climbs.

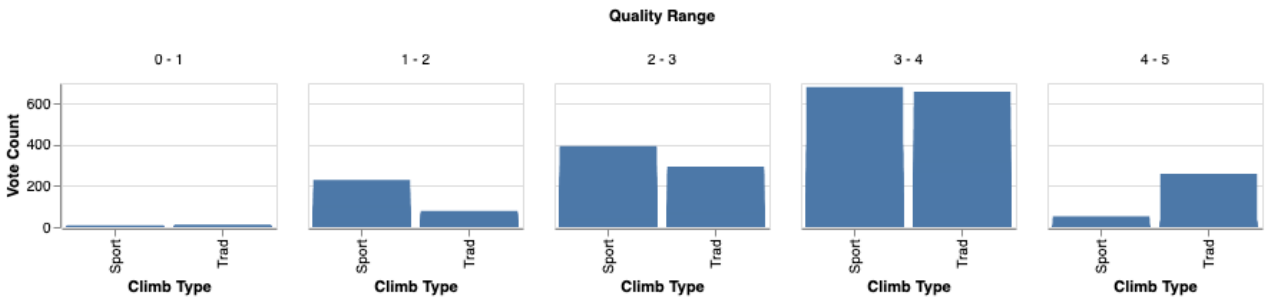

Figure 5: Relationship between Type of climbing technique and quality of climb on location

Figure 5 represents the two different climbing styles and plots the vote counts relative to the quality. We can interpret from figure 5 that the

higher quality climbs are traditional while the lower quality climbs are for sport.

Linear Regression Models for predicting Quality of Climb:

After conducting exploratory analyses, we decided to determine whether quality could be predicted from the variables Length, Location, Grade, and Type. In one of the linear regression models (Model 1), we ignored Length because most of the observations lacked a value for this variable; hence, the model was Quality = Location + Type + Grade. The Locations were converted from coordinates of latitude and longitude to regions of the United States: Southwest, Midwest,Northeast, and South. The "type" variable remained Sport and Trad, and Grade/Difficulty was converted to levels from 1-13. The computed coefficient estimates for the models were:

Model 1:

	coefficient estimate	standard error
Type_Trad	2.067083	0.322270
Difficulty_1	0.140964	0.025602
Difficulty_2	-0.339571	0.482663
Difficulty_3	-0.461304	0.378458
Difficulty_4	-0.486137	0.363799
Difficulty_5	-0.381237	0.341402
Difficulty_6	-0.380249	0.331930
Difficulty_7	-0.307065	0.327533
Difficulty_8	-0.182330	0.324672
Difficulty_9	-0.105705	0.323843
Difficulty_10	-0.006915	0.323787
Difficulty_11	0.284049	0.322462
Difficulty_12	0.486058	0.322682
Difficulty_13	0.791818	0.323341
Difficulty_14	1.069958	0.327515
Region_Midwest	0.982065	0.368709
Region_Southwest	0.309585	0.052730
Region_South	-0.052769	0.035345
Region_Northeast	0.196046	0.034967
intercept	0.134004	0.045047
error_variance	0.515895	NaN

The R-squared value for Model 1 was:

0.2055627405864468

Model 2:

	estimate	standard error
intercept	2.054105	0.046519
Length	0.001132	0.000137
Region_Midwest	0.362628	0.068918
Region_Southwest	0.027548	0.045533
Region_Southeast	0.162287	0.044999
Region_Northeast	0.087530	0.060981
Type_Trad	0.044433	0.032820
Difficulty_10	0.220616	0.047001
Difficulty_11	0.447551	0.048492
Difficulty_12	0.768690	0.053024
error variance	0.517058	NaN

The R-squared value for Model 2 was:

0.14174298245142924

Discussion

Firstly, we generated some visualizations including scatterplots, barplots, and a map. Since we are interested in how our variables affect the quality reports, we started off by plotting the quality of climb against the number of votes faceted by difficulty; we gathered that higher difficulties correlate to a reduction in the vote counts. From this information, we can assume that the quality ratings are associated with the climbers' degree of experience. For our next plot, we mapped out all the locations of the climbs across the US with respect to their quality. We consequently gathered that quality is positively correlated to large mountain ranges as well as national parks. Our next plot was a horizontal bar plot with quality against length; this demonstrated that lengthier climbs tend to be associated with higher quality. Our final visualization was a comparison between quality for climbing types: sport and traditional. We gathered that traditional climbs tend to have more high quality votes.

After conducting exploratory analyses with Length, Location, Grade, and Type, we decided to determine whether quality could be predicted from these variables. The models proved that predicting quality was difficult because R^2 was very low. This could imply that quality is subjective, and climbers have significantly different factors that they consider when providing a quality rating.