**Project Report**

**A Soil Property Analysis of America’s Rock Climbing Areas**

**1. Introduction**

Rock climbing routes found outdoors are often thought to be exclusive to mountainous regions with rocky soil, but is that really the case? Do you really need to travel to the Rocky Mountains or the Appalachians to climb quality routes? Quality climbing areas exist elsewhere and may not even be discovered yet. Being an avid climber, I am also curious if the types of climbing in regions throughout the United States differs. Is trad climbing more prevalent in the West while sport climbing is more popular in the Central United States?

In this project, I plan to use Mountain Project’s1 outdoor climbing route database to gather the locations of rock-climbing hubs across America. I will compare that data to soil property data obtained through UC Davis’2 soil properties map to discover which soil properties are common throughout U.S climbing destinations. Based on the common soil properties I can determine if there are any new potential rock climbing areas to be discovered in the U.S. Based on the ability to scrape Mountain Project, my analysis questions have changed slightly from my proposal. This is addressed further in the Analysis and Conclusion sections.

**2. Data**

For this project I used two sources of data: Mountain Project’s1 climbing areas database and UC Davis’s2 soil property database.

*2.1 Climbing Areas*

I collected data from Mountain Project, which allowed me to find information about the top 10 climbing areas in the United States.

Mountain Project’s data was found on ten different webpages. After sorting through each climbing area in the United States and determining the top ten most popular areas, I wrote a web crawling script to obtain data about these ten climbing areas. I collected the *area name, area page views, state, GPS coordinates, and number of routes for each climbing type.* After scraping I created a csv file which is named *0\_mp\_df.csv* in the project folder*.* The web crawling script I wrote is contained in the file *0\_mp\_scraping.R* in the project folder.

1 <https://mountainproject.com>

2 <https://casoilresource.lawr.ucdavis.edu/soil-properties/>

The Mountain Project data required a heavy amount of cleaning. The GPS coordinates needed to be split into longitude and latitude to be used in conjunction with the UC Davis soil properties database. The area page views and total number of climbs per area also needed to be cleaned and reformatted as integers. The cleaning was done in the file named *2\_cleaning\_area\_properties.R* in the project folder.

*2.2 Soil Properties*

The UC Davis Soil Properties website provides user soil property data for any location in the mainland United States.

The information is contained on a single webpage which includes an interactive map that users can utilize to find soil properties of any location in the continental United States. The data was obtained by aggregating USDA-NCSS soil survey data. The website allows users to click on the map or enter coordinates.

Due to the point and click nature of the map website I had to manually find and enter soil property information using each of the ten coordinates for climbing areas. I collected the *available water holder capacity, drainage class, rock fragments, amount of sand, amount of silt, amount of clay, soil depth, and soil order* for each climbing area. I entered the information using the file named *0\_soil\_properties.R* in the project folder. The data was exported to a csv named *0\_soil\_df.csv* in the project folder.

The data didn’t require any cleaning since it was manually entered. However, I did need to add each climbing area name so that the data frame could be merged with the climbing area csv in the future. I did this by referencing the GPS coordinates scraped in the first script. I did this using the tidyr3 package.

*2.3 Combining Climbing Areas and Soil Properties*

Since both datasets contained the same climbing area name, I was able to merge the datasets to create a longer single dataset. I used the **merge** function to combine the datasets horizontally. All this work is completed in the file named *1\_combining\_area\_properties.R* in the project folder.

Once the datasets were merged, I saved them to a csv titled *1\_combined\_df.csv.* I imported that csv into the *2\_cleaning\_area\_properties.R* file to complete the final cleaning steps mentioned in section 1.1. After all cleaning was done, the final csv was saved to the file named *2\_cleaned\_area\_properties.csv* in the project folder. The final dataset had 10 observations of climbing areas and 22 variables of information about each climbing area. A description of each variable can be found below in Table 1.

3 <https://cran.r-project.org/web/packages/tidyr/index.html>

*Table 1 Data Dictionary*

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Type | Source | Description |
| area\_name | Text | Mountain Project | The name of the climbing area |
| state | Text | Mountain Project | The state in which the route/area is located |
| area\_page\_views | Numeric | Mountain Project | The number of page views for a climbing area |
| latitude | Text | Mountain Project | The latitude of the climbing area |
| longitude | Text | Mountain Project | The longitude of the climbing area |
| total\_climbs | Numeric | Mountain Project | The total number of climbs in each area |
| trad\_climbs | Numeric | Mountain Project | The total number of trad climbs in each area |
| sport\_climbs | Numeric | Mountain Project | The total number of sport climbs in each area |
| toprope\_climbs | Numeric | Mountain Project | The total number of toprope climbs in each area |
| boulder\_climbs | Numeric | Mountain Project | The total number of boulders in each area |
| ice\_climbs | Numeric | Mountain Project | The total number of ice climbs in each area |
| aid\_climbs | Numeric | Mountain Project | The total number of aid climbs in each area |
| mixed\_climbs | Numeric | Mountain Project | The total number of mixed climbs in each area |
| alpine\_climbs | Numeric | Mountain Project | The total number of alpine climbs in each area |
| avail\_water\_holding\_capacity | Number | UC Davis | The amount of water that a soil can store that is available for use by plants (cm) |
| drainage\_class | Text | UC Davis | The moisture condition of the soil in its natural condition throughout the year. |
| rock\_fragments | Number | UC Davis | The percent by volume of the soil that is occupied by rock fragments |
| sand | Number | UC Davis | The weight percentage of the sand particles less than 2 mm and greater than or equal to 0.05 mm |
| silt | Number | UC Davis | The weight percentage of the silt particles greater than or equal to 0.002 mm but less than 0.05 mm |
| clay | Number | UC Davis | The weight percentage of the clay particles less than 0.002 mm in equivalent diameter in the less than 2 mm |
| soil\_depth | Text | UC Davis | Depth of soil (cm) |
| soil\_order | Text | UC Davis | Soil taxonomy is a system for classifying soils. At the highest level, soils are placed in one of 12 categories known as orders |

**3. Analysis**

All analysis and further description can be found in the file named *3\_area\_properties\_analysis.Rmd* in the project folder.

*3.1 Soil Properties*

The first aspect of the data I wanted to investigate was the soil properties at each climbing area. I was curious if there were common soil properties between climbing areas. To complete this analysis, I started by creating bins for each numerical soil property. This included binning *available water capacity, rock fragments, sand, silt, clay, and depth.* For simplicity I used the bin criteria found on the UC Davis soil property website, the criteria can be found below in Table 2.

*Table 2 Bin Criteria*

|  |  |
| --- | --- |
| AWC | **A screen shot of a chart  Description automatically generated** |
| drainage\_class | **A chart of different colors  Description automatically generated** |
| rock\_fragments | **A chart with numbers and text  Description automatically generated with medium confidence** |
| sand | **A chart of different colors  Description automatically generated** |
| silt | **A chart of different colors  Description automatically generated with medium confidence** |
| clay | **A chart of different colors  Description automatically generated** |
| soil\_depth | **A white background with black text  Description automatically generated** |
| soil\_order | **A chart of different colors  Description automatically generated with medium confidence** |

After each variable was essentially turning into a categorical variable, I utilized ggplot4 and Rcolorbrewer5 to plot each soil property in a colorblind friendly way. Each plot is specific to a single soil property and displays a count of each category of the soil property for the ten climbing areas.

The plots showed that five out of the eight soil properties measured were unique to climbing areas. Those soil properties were *available water holding capacity, drainage class, weight percentage of clay, soil depth, and soil order.* This means that specific categories of these soil properties occur much more often in climbing areas than the rest of the United States.

Starting with *available water holding capacity (AWC),* I observed a high frequency of climbing areas with low water holding capacity soil. Climbing areas were rarely found in areas with high water holding capacity soil and are not found in this dataset. This observation makes sense for climbing areas found in the desert but what is more interesting is even with areas not in deserts such as The Gunks and Smith Rock, the soil at these areas still had low *AWC* while soil in the surrounding area did not. Similarly, climbing areas also had well drained or excessively drained soil. This makes sense because *AWC* is directly related to *drainage class*.

*Weight percentage of clay* was also found to be a significantly different soil property for climbing areas. Climbing areas often had low amounts of clay in the soil (between 10-35%), while the rest of the United States often has up to 55% of clay in the soil. This indicates that climbing areas are often found in areas with low amounts of clay in the soil.

Climbing areas often had deep *soil depth.* This observation makes sense for climbing areas in mountainous regions such as Bolder or The Wasatch Range but even climbing areas in deserts such as Joshua Tree had deep soil where the rest of the surrounding area did not.

Lastly, out of the 11 *soil orders* found in the United States only 3 soil orders were found in the climbing areas that I studied. Those *soil orders* being entisols, inceptisols, and mollisols. This observation is quite significant because there are many soil orders found throughout the U.S but great rock climbing is typically found in areas with one of the three soils mentioned above.

Surprisingly, *weight percentage of clay, weight percentage of silt,* and *the amount of rock fragments* in the soil were not consistent across climbing areas. Therefore, these soil properties are not significant in identifying common soil properties across climbing areas.

For a more in-depth look at the soil property frequency plots built using ggplot and Rcolorbrewer please refer to Figure 1 and 2.

4 <https://cran.r-project.org/web/packages/ggplot2/index.html>

5 <https://cran.r-project.org/web/packages/RColorBrewer/index.html>

A collage of different colored bars

Description automatically generated

Figure 1 Plots for AWC, Soil Drainage, Clay, and Sand

A collage of different colored bars

Description automatically generated

Figure 2 Plots for Rock Fragments, Silt, Soil Depth, and Soil Order

*3.2 Soil Properties Elsewhere in the U.S*

After analyzing the soil properties for the climbing areas in my dataset, I was curious to see if these common soil properties existed at less popular, Midwest climbing areas that I frequent. After searching the UC Davis soil database again I came to the conclusion that the common soil properties found at uber-popular climbing areas also exist at the climbing areas that I love. This proves that these common soil properties can be used to search for potential new climbing areas in the U.S. Figure 3 shows the soil properties at my favorite climbing area in the state of Iowa: Pictured Rocks. I observed that all five of the common soil properties are also found at Pictured Rocks.

A screenshot of a computer screen

Description automatically generated

Figure 3 Soil Properties for Pictured Rocks, Monticello, IA

*3.3 Predicting Potential Climbing Areas*

Predictive analytics is outside the scope of this project but I think the finding could be used for such. Using the five common soil properties discussed in section 3.1, you could utilize the UC Davis soil property map to search for the common soil properties and check Mountain Project to see if rock climbing has been established in that area. If climbing hasn’t been established, there may be untapped potential rock climbing in that area.

*3.4 Dominant Climbing Types in Different Regions*

Once I had various climbing areas across the U.S all in to one database I was curious if the types of climbing done in the Western United States is the same as the Central or Eastern United States. To answer this question I classified each climbing area into a region, either West, Central, or East. Then I totaled the number of routes for each climbing type in each region. The result was the three plots shown in Figure 4. From the plots, I observed that Trad climbing is very popular at the Eastern and Western United State’s most popular climbing areas. Sport and trad climbing is the most popular in the Central Region. Obscure forms of climbing like ice, alpine, and aid are mainly available in the Western United States.

A group of pie charts

Description automatically generated

Figure 4 Types of Climbing by Region

**4. Conclusion**

Throughout this project I analyzed the soil properties found at the top ten climbing areas in the United States. I also analyzed climbing areas broken down by region. After adjusting some questions from my project proposal, I decided to answer these four analysis questions:

1. *What are the most common soil properties among U.S climbing areas?*

Out of the eight soil properties analyzed for each climbing area, there were five categories that were consistent across all ten climbing areas. Those soil properties were: Low available water holding capacity, excessively or well-drained soil, a low weight percentage of clay, deep or very deep soil depth, and the soil order being either entisols, mollisols, or inceptisols.

1. *Do these common soil properties exist elsewhere in the United States where rock climbing hasn’t been developed?*

After checking the common soil properties with climbing areas not included in my dataset such as Pictured Rocks I confirmed that these soil properties do exist at various other climbing areas. There is also an abundance of areas that match the common soil property list where there is not climbing developed. Anyone can access the UC Davis soil property map and find these areas.

1. *Are there any soil properties that are varied across climbing areas?*

I found that weight percentage of clay, weight percentage of silt, and the amount of rock fragments in soil had no consistent measure across climbing areas. These soil properties are insignificant when trying to find new climbing areas and do not predict quality climbing.

1. *Does the type of climbing differ between West, Central, and Eastern regions?*

The dominant type of climbing does vary between regions. The West has very diverse climbing types with trad climbing being most popular and sport and boulder climbing also having lots of routes. In the Central United States, sport and alpine climbing is dominant and you cannot find ice, mixed, or aid climbing in abundance like you could in the West. In the East, trad is the most popular type of climbing and in The Gunks I observed no sport climbing whatsoever.

My project has a few limitations, including the number of climbing areas scraped from mountain project, the inability to quickly pull soil property data from UC Davis’s database, and there being no specific climbing route information (information was limited to climbing areas). For the future, I could expand this project to include more climbing areas and routes. This expansion would allow me to improve my analysis of common soil properties. It would also improve the dominant climbing type analysis by region, especially for the Central and Eastern United States. Adding more soil property data would allow me to lengthen my list of soil properties to look for when searching for new climbing areas.