A First Analysis of 8A User Data

Tristan Blackburn, Anthony Hayes
& Oliver Winston

April 13, 2018

This document serves as an introduction to research performed by Tristan Blackburn, Anthony Hayes & Oliver Winston on the 8a.nu database. The three of us are currently concluding our PhD's in physics at the University of Sussex and, through our mutual love of climbing, have turned our attention to analyzing the relatively young field of climbing sports-science. Our hope is to collaborate with the good folk at Lattice to produce both academic and public information.

Key Findings

1. A BMI analysis of 8A users on a per country, per gender, basis. EU data is used for comparison with the wider population.
2. A height analysis of boulders in Fontainebleau. Categorizing the most popular boulders by the height disparity between the average ascensionist of a given boulder and the average height of all Font climbers.
3. A 'softness' analysis of boulders in Fontainebleau. Categorizing the most popular boulders in Font into the hardest and easiest climbs relative to their established grade, given the distributions in performance of the listed ascensionists.

1 Introduction

1.1 Data and Software

The research was largely performed in the python interpretive programming language. C++ was used for some pre-processing of the data using the data analysis package ROOT, developed at CERN.

A complete list of python packages utilized:

- pandas
- Levenshtein-python
- matplotlib
- numpy
- tables

A software framework that allows portable analysis and quick replicable results was developed in both C++ and Python. This is maintained, alongside all code, data and output, through the GitHub repository: `https://github.com/ollywinston/ClimbingStats`

1.2 Mapping Users to Ascents

8a data is divided into two databases - users and ascents. Each entry in the ascents database has a user ID associated with it. Ascents were mapped to each user and stored as vectors containing the user ID and all matching ascents. From this, a distribution of each users ascents for any metric in the ascents database was cross comparable with the user information provided in the user table.

1.3 Creating Climb ID's

The current 8a data is messy. User inputs have zero automated corection. The boulders 'Le Toit de Cul de Chien' and 'Le Toit du Cul de Chien' are both unique entries, with different numbers of ascents, grades and meta-info.

For the height and grade analyses this was remedied by fuzzy clustering using a string comparator called the Levenshtein distance. This returns the number of iterations required to changed one set of characters to another. In the previous paragraphs example the score returned would be 1: '...du...' to "...de...'. A ratio of the number of character changes to the total number of characters was required to be less than 0.80 to identify unique boulders. Boulders which failed this criteria were concatenated into a single boulder ID, where ascents, grades and meta-info were added in summation and averaged where appropriate. Further data cleaning took place by removing extraneous special characters, null entries in name (such as 'n/a' or 'unknown') and boulders with low numbers of ascents.

Both analyses are for Fontainebleau. The nature of the data framework allows fast replication of these analyses for any other crag, or routes instead of boulders. 8a does not have a strict flag for distinguishing between trad and sport climbing. The processing time for the fuzzy clustering and concatenation process is approximately twenty minutes.

2 BMI Analysis

Figure 1 shows fits of the BMI distributions for men and women in the 10 most populated countries in the 8A database. Male data is shown in blue, female in yellow. Black lines represent the quantile boundaries, purple lines show the BMI classification boundaries. Data was cleaned to remove extraneous outliers resulting from unit errors at user input.

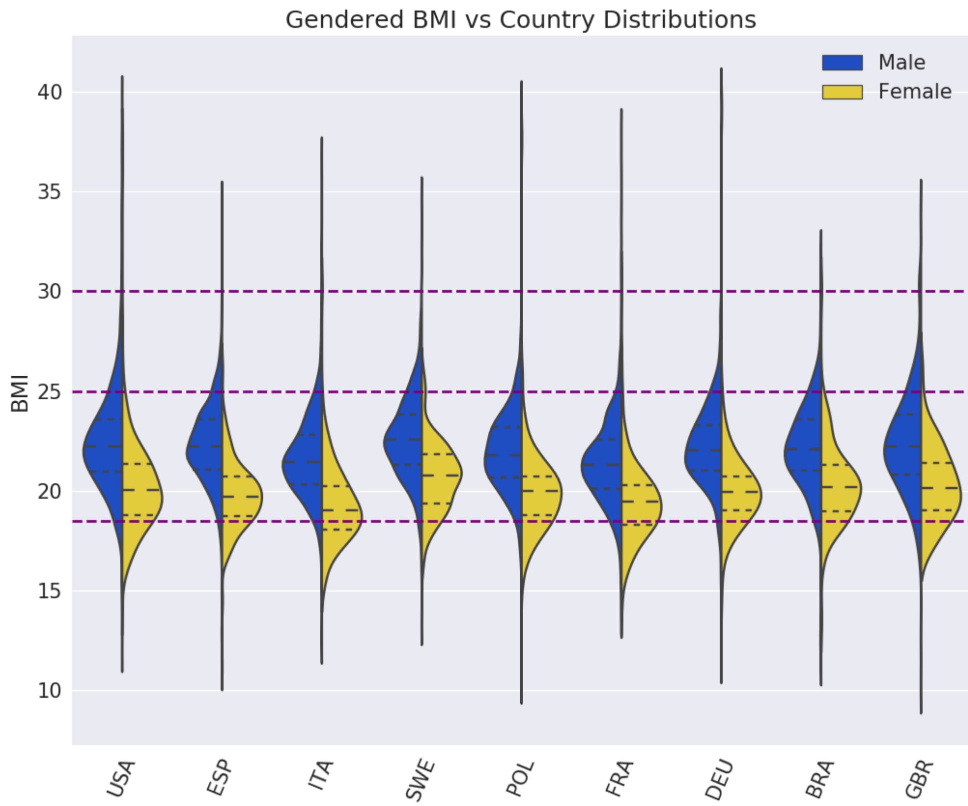

Figure 1: BMI distributions of women and men for the most populated countries in the 8A.nu database. Fits are kernel density equations. Quantile boundaries are shown in black. Purple lines show BMI classification boundaries.

Figure 2 & Figure 3 show the male and female percentage breakdowns of BMI categories with respect to the total populations of the relative sub-plot respectively. The top left plot shows the percentage of each BMI classification for the 10 most populated countries in the 8a database, for 8a users. The bottom-left plot shows the percentage of each BMI classification for the country populations as taken from the 2014 EU census. The countries selected are matched to the most populated countries in the 8a database. The right hand plot shows the ratio between the 8a percentage BMI classification and the EU census data.

The results show that both female and male 8a users are significantly more likely to be in the ‘healthy’ BMI category with respect to the general population. As follows, the probability of an 8a user being in either the ‘overweight’ or ‘obese’ category is significantly reduced with respect to the general population. In both the female and male case, an 8a user has a greatly increased probability of being in the underweight category when compared with the general population. This result is particularly apparent in the female distributions. Underweight German female 8a users are over 10 times more frequent than in the general population. For men, the ratio remains consistent across countries at approximately twice that of the general population.

8a users exhibit healthier BMI distributions than the general populations of their respective countries in both genders. The percentage of unhealthy 8a.nu climbers (comprising 3 BMI categories) is primarily populated by underweight individuals - the general population considered unhealthy is weighted towards the overweight and obese BMI categories. In the case of female 8a climbers some countries exhibit greater than 25% populations of underweight individuals. This is a drastic increase with respect to the EU census. For both genders, in all country distributions, the percentage of the 8a population in the healthy range greatly exceeds that of the general populations. 8a users are arguably significantly healthier than the average individual.

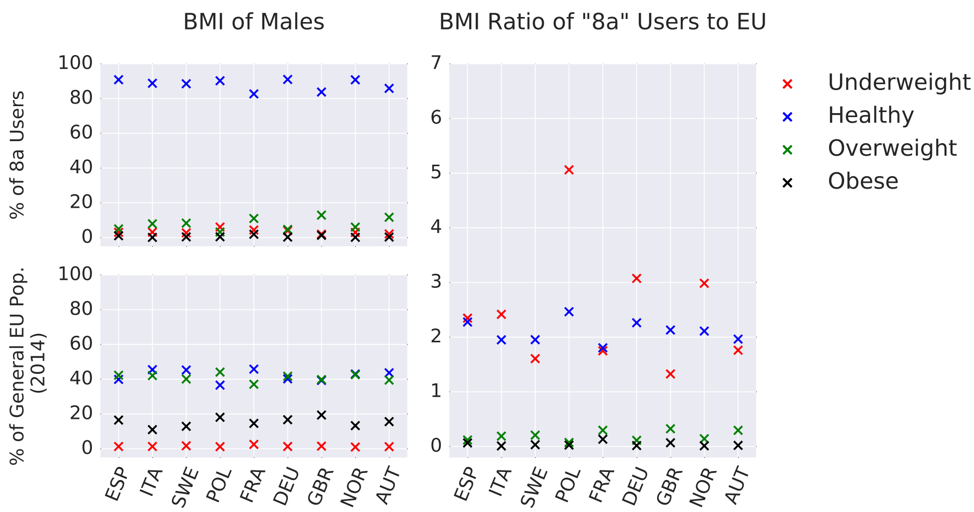

Figure 2: Percentage breakdowns of male BMI classifications in the 8A.nu database, EU 2014 census data and the respective ratio between the two

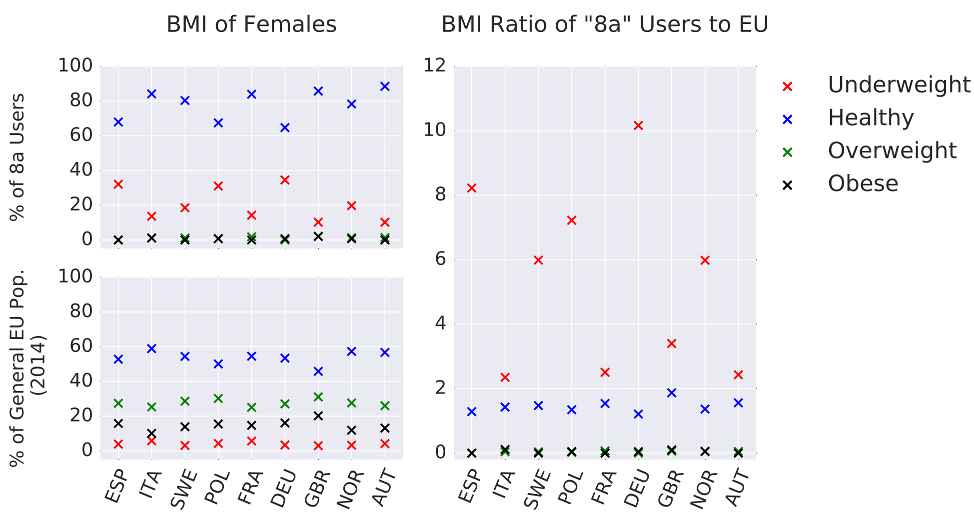

Figure 3: Percentage breakdowns of female BMI classifications in the 8A.nu database, EU 2014 census data and the respective ratio between the two

3 Height Analysis

Height advantage is a hard topic to avoid in the climbing community. Few visits to a climbing centre are free of some reference to height. We investigated the correlation of height with ascents for boulders in Fontainebleau. The 200 most popular, by ascents, boulders in Fontainebleau, were mapped to heights of each user that had non-zero heights in the database. The height of all users with logged ascents in Fontainebleau were then averaged to provide a benchmark for comparison. The disparity between the average height of a given boulders ascensionists and the average Font climber was then used to score each boulder. Figures 4 show the results.

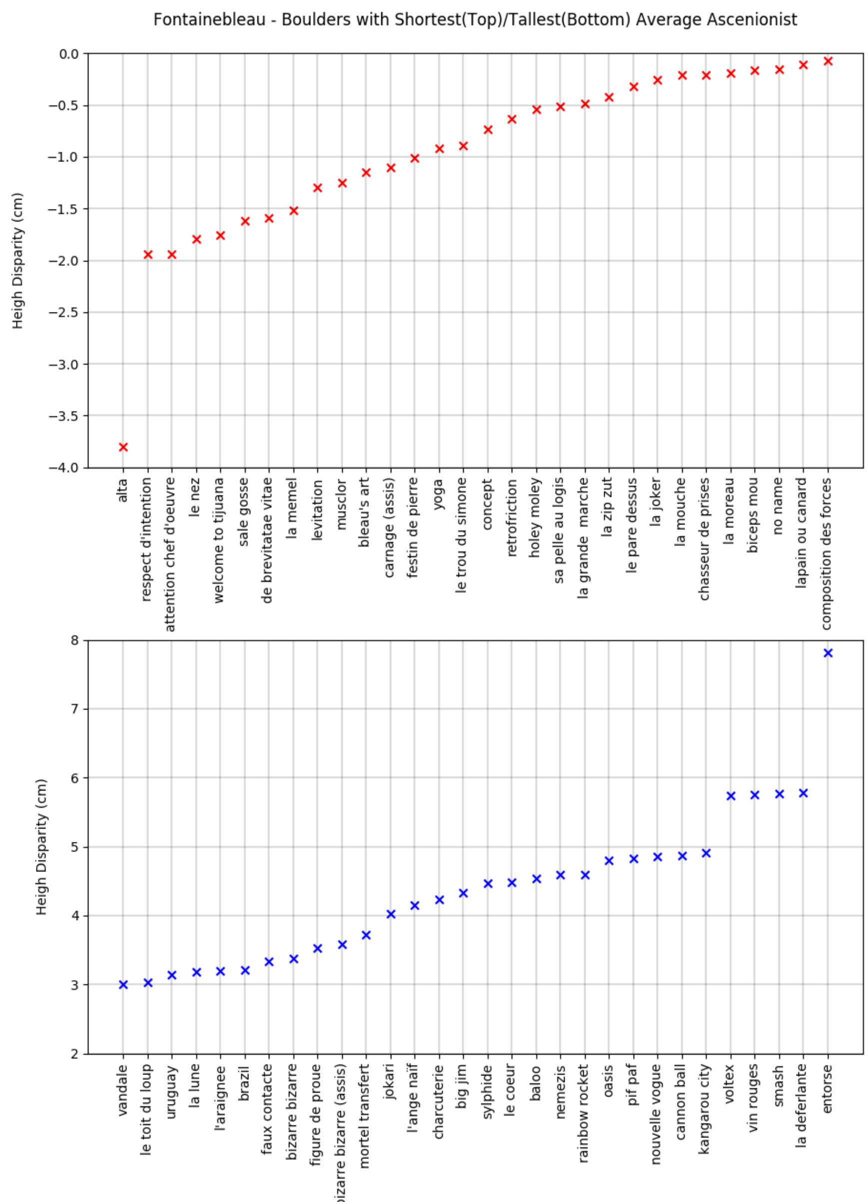

Figure 4: Boulders with the tallest and shortest average ascenionist with respect to the average height of a Font climber (176.4cm)

The climbs which most favour tall climbers, are perhaps unsurprisingly, almost all dynos. Famous climbs like Rainbow Rocket(8A) and Entorse(7A+) are close to impossible for short climbers and are almost never climbed by a user with a height at or below the Font-average. Boulders favouring short climbers are often less publicized. Alta(7C), La Joker(7A) and Sale Gosse(7C) all involve moves with small boxes and cramped positions, disfavouring the gangly.

4 Softness Analysis

Climbers love few things more than arguining over subjective levels of difficulty. We took a shoot at quantifying it. Our dataset only looks the most popular climbs in Font. This has a double implication. Firstly, Font is renowned for it's grade stiffness, in spite of being the birthplace of the grading system... Even the softest climb may still be solid at its grade in a more global context. Secondly, the measure of softness has a systematic error. A highly popular but hard climb, targetted frequently by climbers for which it is their maximum grade, would score as softer. This may also be the reason the boulder is targetted by people looking to push their grade...

The scores were calculated as follows:

$$S = 1 - \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{t_g}{T}$$

where:

- n & N are user and total number of users
- t & T are user boulder and total number of boulders done by user
- t_g is the number of boulders done by this user at or above the consensus grade of the relevant boulder.

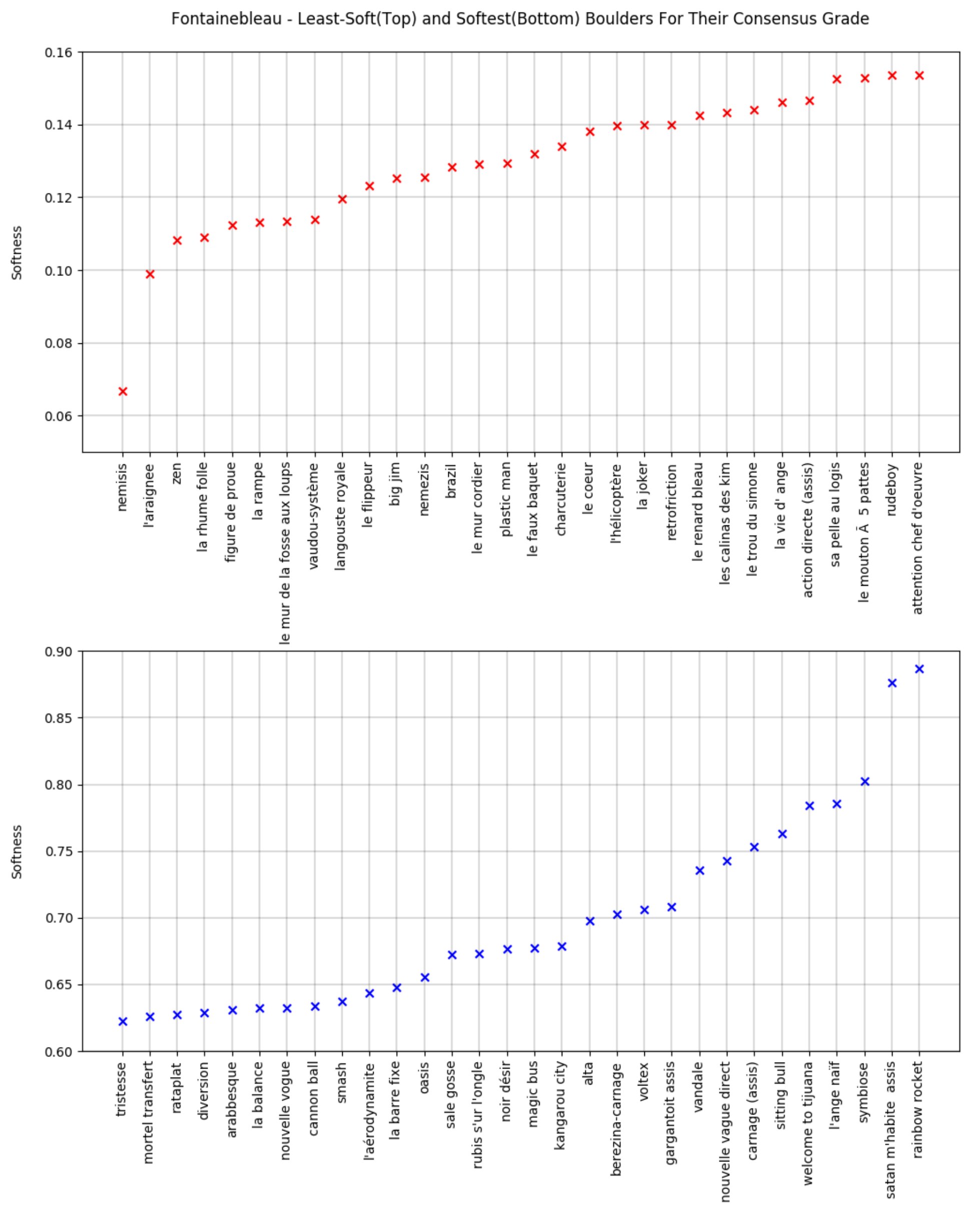

Figure 5: Boulders with the highest and lowest 'softness' rating as measured by % likelihood of a given ascent occuring for a users grade distribution, average over all user ascents of a given boulder. Scores run from 0 to 1.