

# Wanderlust: Discovering New Hiking Trails

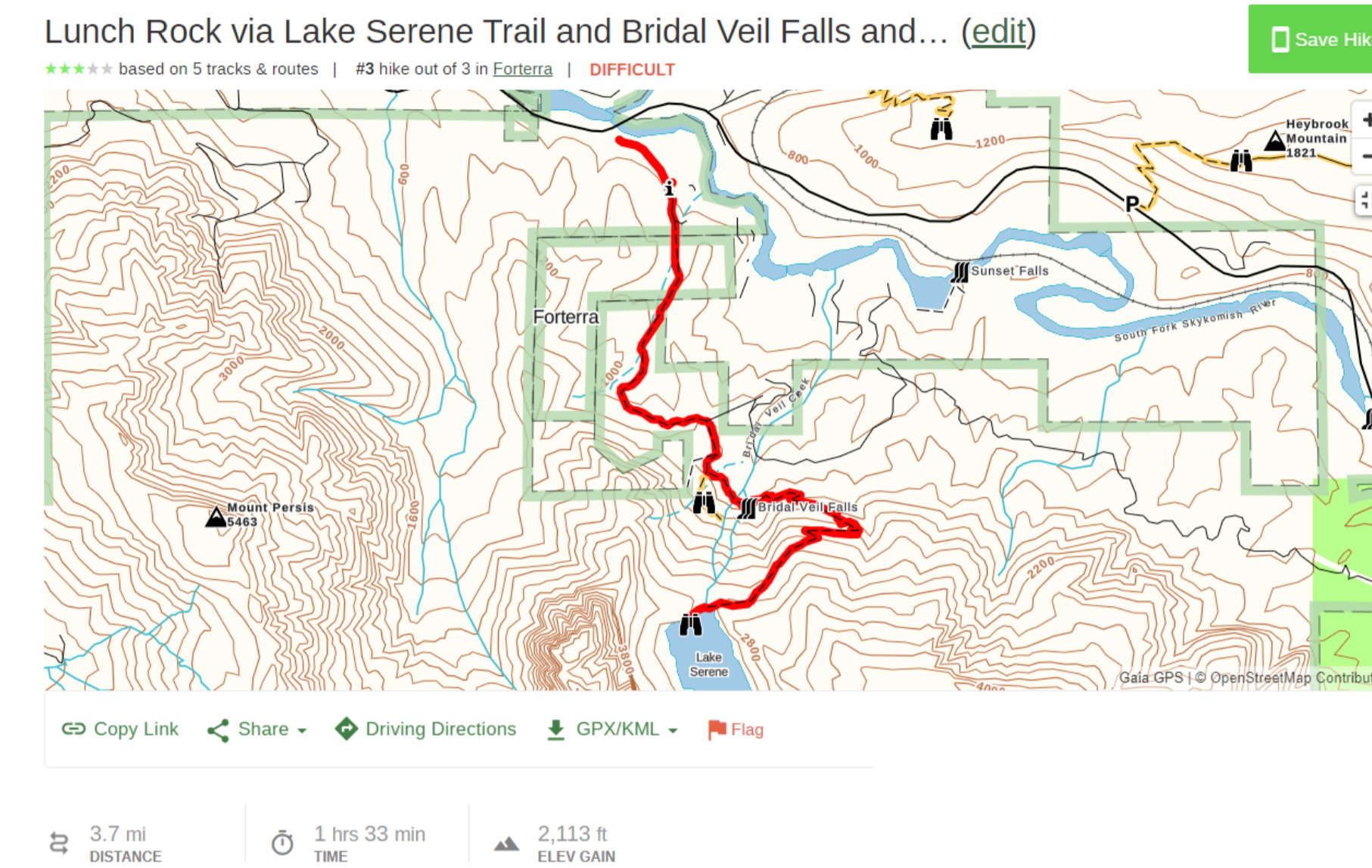
Nick Kaashoek

## Motivation

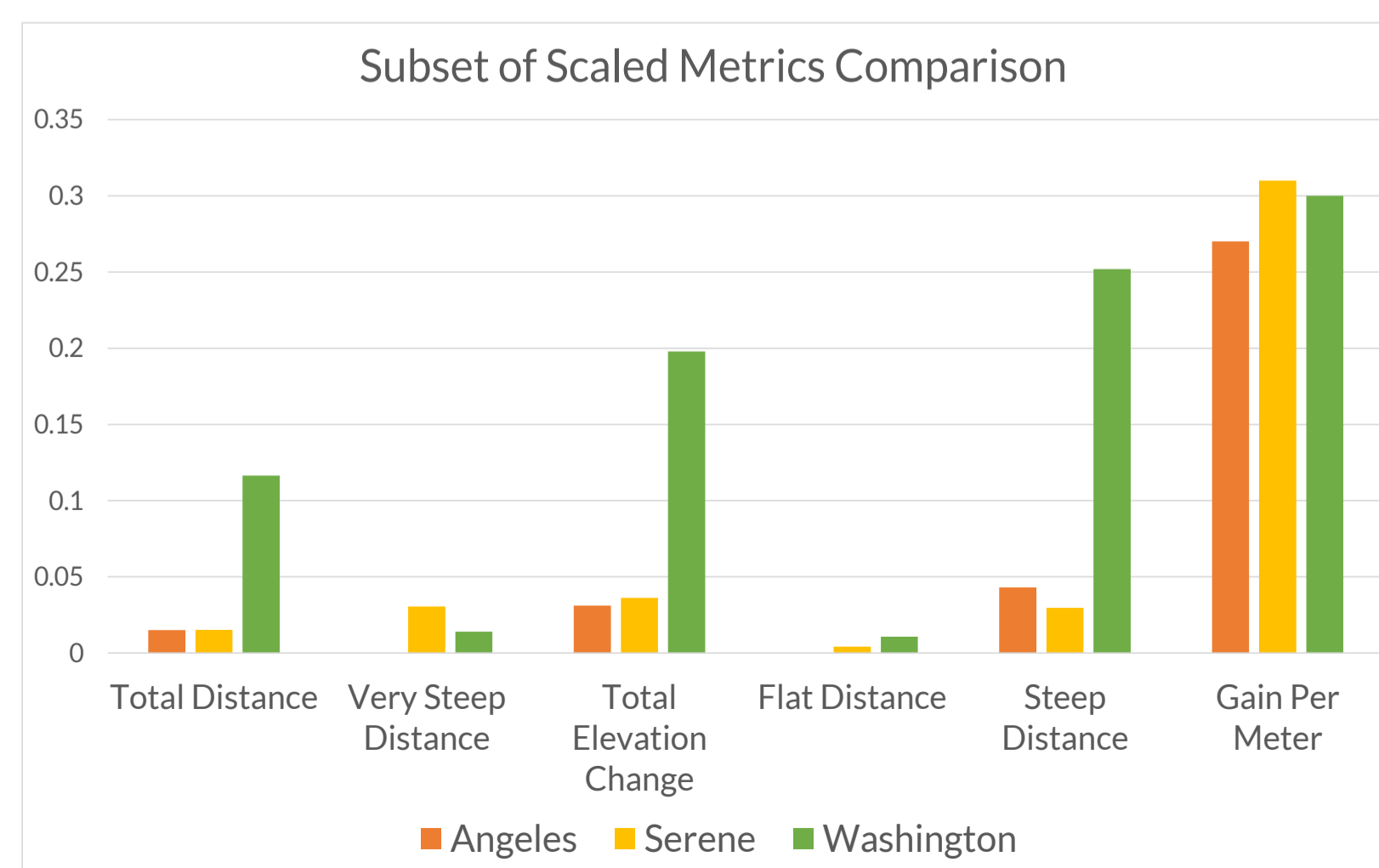
- Websites do a poor job in exposing less frequently hiked trails
- Websites have bloated difficulty categories
- Very little personalization; every hiker different
- Use machine learning techniques combined with more detailed analysis to increase personalization

## Data

- Data started as GPX files scraped from GaiaGPS and AllTrails
- GPX Files consist of an array of points containing elevation and positional data
- Parsed each GPX file into data structure
- Ran computations on the data structures to produce the metrics seen to the right
- Total of 16,723 trails parsed



| Metric                        | S    | A    | W     |
|-------------------------------|------|------|-------|
| Distance on Very Steep        | 0    | 172  | 19    |
| Total Elevation Change        | 846  | 985  | 5375  |
| Number of Steep Sections      | 22   | 15   | 97    |
| Average Altitude              | 951  | 437  | 1434  |
| Elevation Change per Meter    | 0.15 | 0.18 | 0.17  |
| Number of Flat Sections       | 0    | 4    | 9     |
| Switchbackiness               | 0    | 0    | 0.01  |
| Distance on Switchbacks       | 0    | 0    | 132   |
| Number of Switchback Sections | 0    | 0    | 1     |
| Total Elevation Climbed       | 794  | 792  | 2775  |
| Total Distance                | 5558 | 5576 | 32274 |
| Distance on Flat              | 0    | 241  | 625   |
| Extremely Steep Sections      | 0    | 1    | 0     |
| Total Elevation Descended     | 52   | 193  | 2600  |
| Peak Altitude                 | 1316 | 805  | 1910  |
| Distance on Steep             | 2165 | 1494 | 12678 |
| Very Steep Sections           | 0    | 5    | 2     |
| Distance on Extremely Steep   | 0    | 23   | 0     |



## Estimating Difficulty

- Use Nearest Neighbors Regressor to estimate difficulty
- Combine metrics from all hiked trails for user
- Weighted Minkowski distance to give precedence to some metrics
- More similar trails = closer difficulty

## Recommendations

- Collaborative filtering on ratings
- Test both model based and memory based
- Settled on SVD++, lowest error
- Users query for all unhiked trails, select best 10 predictions and recommend to the user
- See Testing for details on tests

## Trail Similarity

- Based on Nearest Neighbors
- Obtain feature vector from metrics table above
- Run Nearest neighbors with all trails as training set
- Query for 5 Nearest neighbors and display those on webpage
- Weighted distance

## Challenges

- Hard to evaluate the system, can't hike all the trails
- Tested many different ML algorithms to decide on one; no prior experience with machine learning
- What kind of features are most important?
- How to weight the different metrics?

## Testing

- Generate fake data to run realistic tests
- 1) 10 groups of 100 users each
- 2) 15 groups of ~40 trails each
- 3) Different groups rate different subsets of trails consistently
- 4) Generates expected sparse matrix to test algorithms on
- 5) Generate set of test users and manually check where they're recommendations and estimates fall

## Results

- Web application successfully recommends new trails
- Adapts difficulty as users hike more trails
- Manual inspection confirms that trails marked similar appear to be similar
- Using testing strategy above, RMSE/MAE for recommendations is 0.61/0.58
- Difficulty is 0.43/0.38