

A Non-Profit's Effectiveness: What Influences It?

SCA Program Data Analysis



Matthew Terrible
mtteribile@gmail.com
September 2017

Agenda



- Goals
- Executive Summary
- Data and Techniques
- Key Findings
- Summary and Next Steps

Background – SCA



- Formerly The Student Conservation Association (thesca.org)
- Arlington, VA based charitable organization
- Deploys young people to work in US parks, public lands, and urban green spaces
- Teaches them how to plan, enact, and lead, all while making a tangible impact in conservation

Mission (excerpt):

Building ‘the next generation of conservation leaders and inspiring lifelong stewardship of the environment and communities by engaging young people in hands-on service to the land.’

Goal



Analyze survey data collected from participants to help SCA better understand what contributes to a positive experience, and thus improve their programs



Executive Summary



- SCA participant and project data, combined with publicly available data, was statistically analyzed and modeled to gain insights on what contributed to a higher likelihood for participants to recommend the SCA to others.
- Significant differences between programs were found in the recommendation scores.
- Answers to a few other survey questions correlated significantly with the recommendation score. The questions varied by program.
- For each program, some participant and project characteristics (such as age of participant) correlated with the recommendation score, while other characteristics (such as gender of participant) did not.

Data



1. Participant survey data from late 2014 to mid - 2017 (xls file, 4912 rows x 79 columns)*
2. Participant demographic data and project time and place data (xls file, 7923 x 23 columns)*
3. 'Selected Economic Characteristics' and 'Household Income' by zip code from the American Community Survey (csv files downloaded from factfinder.census.gov)
4. Weather data for project dates and locations pulled via API from Weather Underground (<https://www.wunderground.com/weather/api/d/docs>)
5. For data unavailable from Weather Underground (Kalispell), csv file was requested from NOAA (<https://www.ncdc.noaa.gov/cdo-web/>)

*Personally identifying information was removed by SCA

Data Preparation



1. Incomplete data
 - Observations that lacked fields to merge data and lacked most other fields were eliminated (very small number)
 - Observations that lacked a Recommendation score were eliminated (about 2%)
 - Observations that lacked a key piece of data (for example, age), but had other data were used except when analysis was done using that piece of data (for example, age)
2. Duplicate data
 - About 1% of the surveys were duplicates. The survey with the latest time stamp was used.
3. Merging data sets
 - SCA data sets were merged on the 'opening ID' field, a unique identifier for a participant in a role.
 - Participant home demographic data was merged based on home zip code
 - Weather data was merged in based on location (city, state) and date
4. Data cleanup and formatting

Statistical Techniques and Modelling



1. Kolmogorov – Smirnov Test for difference in distributions
2. Correlation matrices to find other questions with high correlation
3. Logistic regression with Lasso regularization (Sklearn) to eliminate features
4. Linear regression (StatsModel) with a few variables at a time for inference
5. Train / Test split, Cross validation (Sklearn) to guard against overfitting
6. Logistic regression and Random Forest models (Sklearn) to confirm inferences about variables that gave the best models

Summary of Data Fields



Dependent variable:

- Recommendation Score – 1 to 10 scale ‘How likely are you to recommend an SCA experience?’

Independent variables:

1. Other survey questions – 1 to 5 scale – example ‘I did something I was really excited about’
2. Participant demographic data (gender, age, ethnicity, home town, etc.)
3. Project information (location, dates, managing organization)
4. Project category
5. Demographic information of participants home zip code
6. Weather information during project

Need
something
visual

Findings - notes



- All findings are statistically significant at a 95% confidence level unless otherwise noted.
- Correlation does not imply causation. No finding in this report is proof that one thing caused another. There may be other underlying root causes.
- Because of the variations in the data (uniqueness of each participant), no model that I built could predict a participant's recommendation score any better than with 46% accuracy, which is equivalent to the percentage of '10' recommendations scores
- Findings with less than 5 data points are not reported

Program Differences

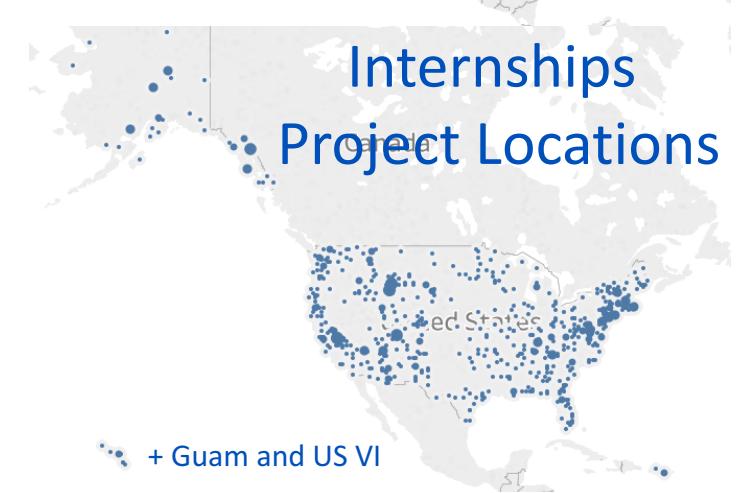
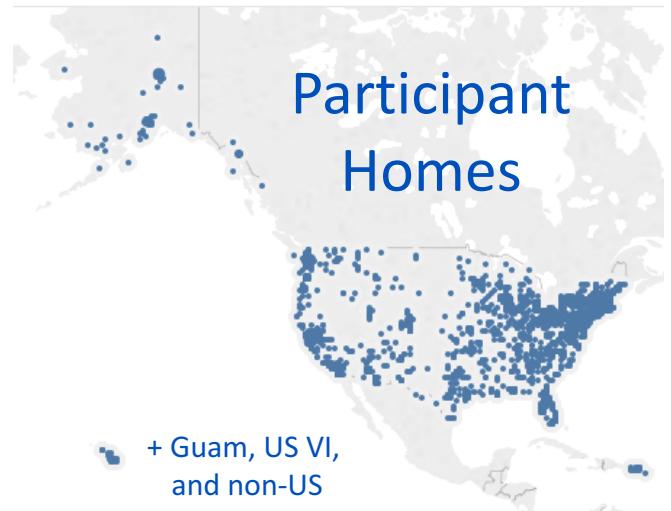
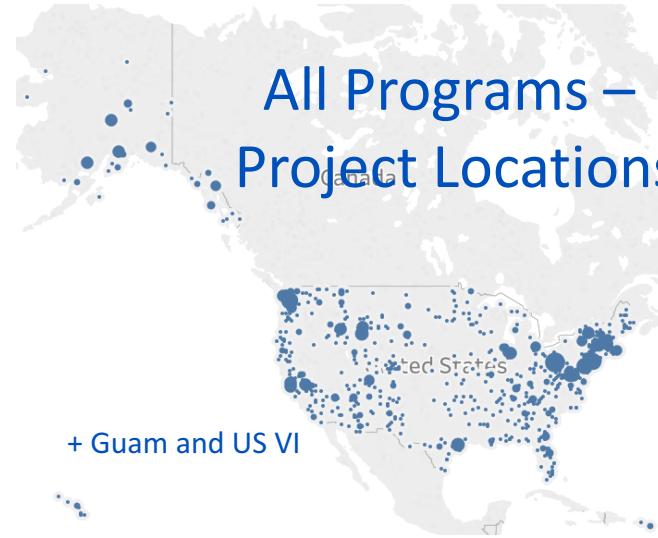


	Community	Crew	Corps	Internships	Leader Crew
Age (target)	Under 18	Under 18	18+	18+	18+
Location	Local	Away	Away	Away	Away
Team vs. Individual	Team	Team	Team	Individual	Team
# of Surveys	1161	434	496	2474	23
Average Recommendation Score	8.67	9.27	8.45	8.52	9.52
Net Promoter Score*	52.4	76.5	46.8	48.5	82.6

Crews and Leader Crews have significantly higher recommendation scores than other programs

* Net Promotor Score = (% of 9 or 10 scores) - (% 1 through 6 scores)

Where?



Who?

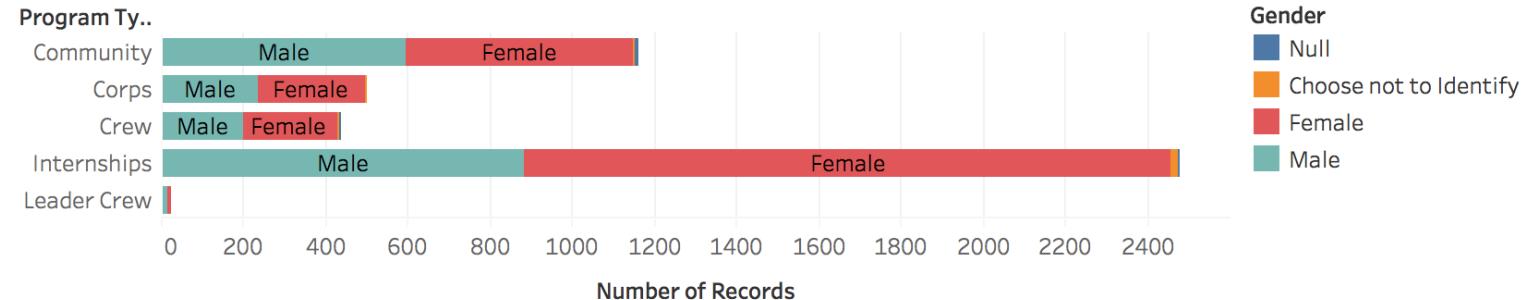


Crew Participants come from wealthier zip codes, Community participants from less wealthy*

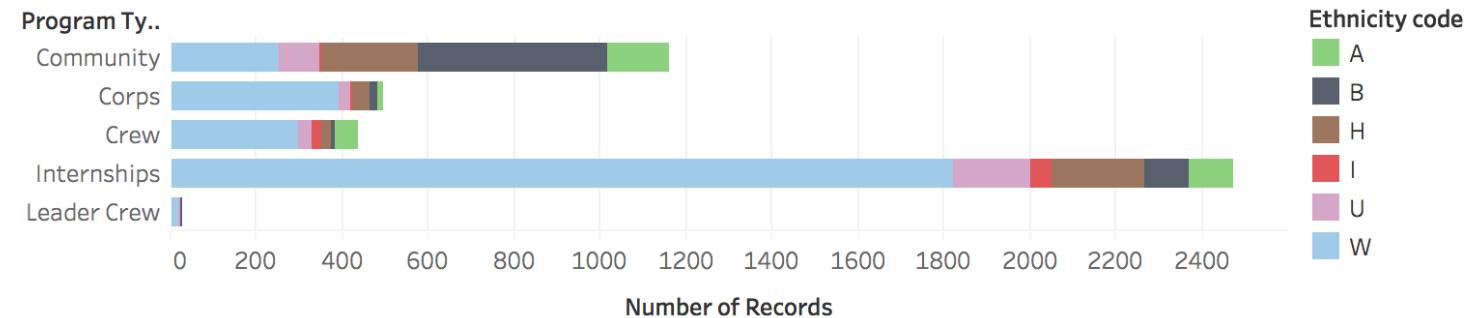


* bars of different colors are statistically different

Except for Internships, Most Programs Evenly Split Male / Female



Community Programs have More Non-Whites than Other Programs



Community Recommendation Scores



- Projects in Pittsburgh (320 surveys), Baltimore (20 surveys), and Lebanon NJ (6 surveys) had lower scores than others
- 25 projects had lower scores than others, including some in the cities
- 14 and under participants give lower scores than other age groups
- Programs greater than 8 weeks (weekend?) have higher recommendation scores (average of 8.9 vs. 8.5) than programs less than 8 weeks (summer?)
- Higher Median Income of home zip code and lower % below poverty line (92% significance) correlated with higher scores
- Not significant: Gender, job title, budget code

Chart?

Community Correlated Questions



Higher scores on these questions correlated to higher recommendation scores (with confidence level):

- Because of my recent SCA experience, I am more interested in a conservation career (96%)
- I plan to spend more time outdoors or in nature (90%)
- The partner (park, etc) where I did the work was supportive (87%)
- I felt like an important part of the larger SCA organization and its mission (87%)

Crew Recommendation Scores



- Two locations had lower recommendation scores
 - Estes Park, CO (22 surveys)
 - Franklin, NC (5 surveys)
- Older or more educated participants gave higher scores
- Whites and Hispanics gave higher scores
- Not significant: Gender, economics of home zip code, weather during project, # of weeks for project, budget code

Add chart

Crew Correlated Questions



Higher scores on these questions correlated to higher recommendation scores (with confidence level):

- I did something I was really excited about (99.9%)
- I felt like an important part of the larger SCA organization and its mission (99.9%)
- I learned how to solve problems that emerged in my team (99%)
- When you think about life after this SCA experience, how likely is it that you will do volunteer work to help people in need (97%)
- Because of my recent SCA experience, I am more interested in a conservation career (91%)
- I learned skills that I can use in a future job (88%)

Corps Recommendation Scores



- Three managing organizations had lower recommendation scores
 - Allegheny County, PA (17 surveys)
 - Bighorn National Forest, WY (7 surveys)
 - SCA Oakland Office, CA (5 surveys)
- One location had lower recommendation scores
 - Staatsburg, NY (6 surveys)
- Not significant: Gender, ethnicity, age, education level, length in weeks, budget code, Americorp participation, economic indicators

Corps Correlated Questions



Higher scores on these questions correlated to higher recommendation scores (with confidence level):

- Did your leader or supervisor work with you to develop goals for your SCA experience? (99.9%)
- I felt like an important part of the larger SCA organization and its mission (99.9%)
- Because of my recent SCA experience, I am more certain about the career I want to have (95%)
- I did something I was really excited about (94%)
- I learned how to cooperate with others (93%)
- The projects I worked on were challenging in a good way (92%)
- I learned new ways to navigate large agencies to get things done (88%)

Lower scores on these questions correlated to higher recommendation scores (with confidence level):

- Would you have participated in SCA if the work had been unpaid? (99.9%)
- I had genuine opportunities to lead the group (96%)

Internships Recommendation Scores



- These budget codes had lower scores as compared to 205 with these confidence levels:
 - 206 (99%) – 696 surveys
 - 330 (93%) – 17 surveys
 - 257 (86%) – 460 surveys
- Those not in Americorps (about 1/3 of participants) gave higher scores
 - Look at education more
- Older or more educated participants gave lower scores
 - Ethnicity – most were whites. Any way to get whites as ref vs. Asians
- Whites and Undefined ethnicity gave lower scores
 - Look at weeks more
- More weeks gave lower scores
 - Look at weeks more
- Not significant: Gender, Veteran status, Location (or Organization Managing), economics of home zip code

Internships Correlated Questions



Higher scores on these questions correlated to higher recommendation scores (with confidence level):

- Because of my recent SCA experience, I am more certain about the career I want to have (99.9%)
- The stipend I received was enough to buy essentials (99.9%)
- As part of my recent SCA experience, I learned new ways to navigate large agencies to get things done (99%)
- My supervisor/leader always prioritized safety when making decisions (98%)
- The work that I performed matched the position that was described to me before I started (98%)

Summary of Findings



- There are differences between programs in:
 - Demographic makeup
 - Recommendation scores
 - Survey questions correlated with higher (or lower) recommendation scores
 - Location or participant attributes that correlate with higher scores
- Factors that generally don't correlate with higher recommendation scores:
 - Gender
 - Economic characteristics of participants home zip code (except for community programs)
 - Weather (for crew programs)

Economic conclusion doesn't fully align with exec summary

Follow up



- SCA evaluate the findings to improve program effectiveness
- Matt is available to
 - Provide further information on any point
 - Delve deeper into any areas of interest