

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

## SCA Program Data Analysis



Matthew Terrible  
[mtteribile@gmail.com](mailto:mtteribile@gmail.com)  
September 2017

# Agenda



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

# Background – SCA



- The Student Conservation Association ([thesca.org](http://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

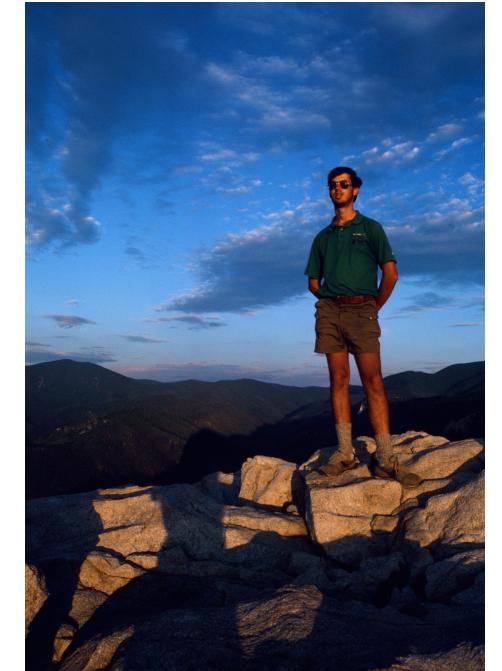
Mission (excerpt):

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 that participants would recommend the SCA to others.
- Significant differences between programs were found in the recommendation scores.
- 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.
- Answers to a few other survey questions correlated significantly with the recommendation score. The correlated questions varied by program.

# 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 rows 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, .csv files were requested from NOAA (<https://www.ncdc.noaa.gov/cdo-web/>)

\*Personally identifying information was removed by SCA

# Summary of Data Fields



## Dependent variable:

- Recommendation Score (1 to 10 scale)  
'How likely are you to recommend an SCA experience?'



## Independent variables:

- Program category (Community, Crew, Corps, Internships, Leader Crew)
- Project information (location, dates, managing organization, budget code)
- Participant demographic data (gender, age, ethnicity, home town, etc.)
- Demographic information of participant's home zip code (median income, etc.)
- Weather information during project
- Other survey questions (1 to 5 scale) – for example, 'I did something I was really excited about'

# Data Preparation



## 1. Incomplete data

- Surveys that lacked fields to merge data and lacked most other fields were eliminated (very small number).
- Surveys that lacked a Recommendation score were eliminated (about 2%).
- Surveys that did not match a member and project profile were eliminated (about 1%)
- 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.

## 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



# Data Science Tools and Statistical Techniques



1. Correlation matrices to find questions with high correlation
2. Logistic regression with Lasso regularization ([scikit-learn](#)) to eliminate features
3. Linear regression ([StatsModel](#)) with a few variables at a time for inference
4. Logistic regression and Random Forest models ([scikit-learn](#)) to confirm inferences about variables that gave the best models
5. Train / Test split, Cross validation ([scikit-learn](#)) to guard against overfitting
6. Matplotlib, Seaborn, and Tableau visualizations to better understand data
7. Kolmogorov – Smirnov test ([SciPy](#)) to confirm difference in non-parametric distributions

# 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.
- No model that I built could predict a participant's recommendation score any better than the percentage of '10' recommendation scores.
- Findings with less than 5 data points are not reported

# Program Differences

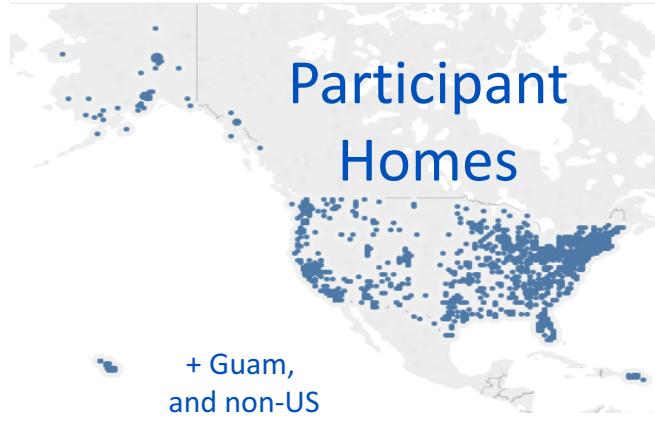


	Community	Crew	Corps	Internships	Leader Crew
Age (target)	Under 18	Under 18	18+	18+	18+
Location	Local	Away	Away	Away	Away
Team vs. Individual	Team	Team	Team	Individual	Team
# of Surveys	1161	434	496	2474	23
Average Recommendation Score	8.67	9.27	8.45	8.52	9.52
Net Promoter Score*	52.4	76.5	46.8	48.5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

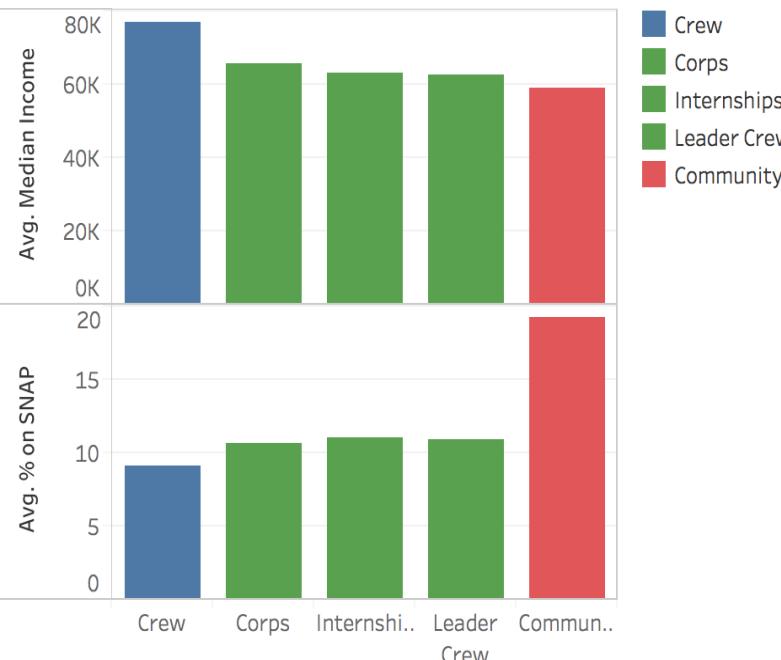


Participants come from all over the US and beyond. Programs are located throughout the US, although team based programs, especially Community programs, are concentrated in certain areas.

# Who?

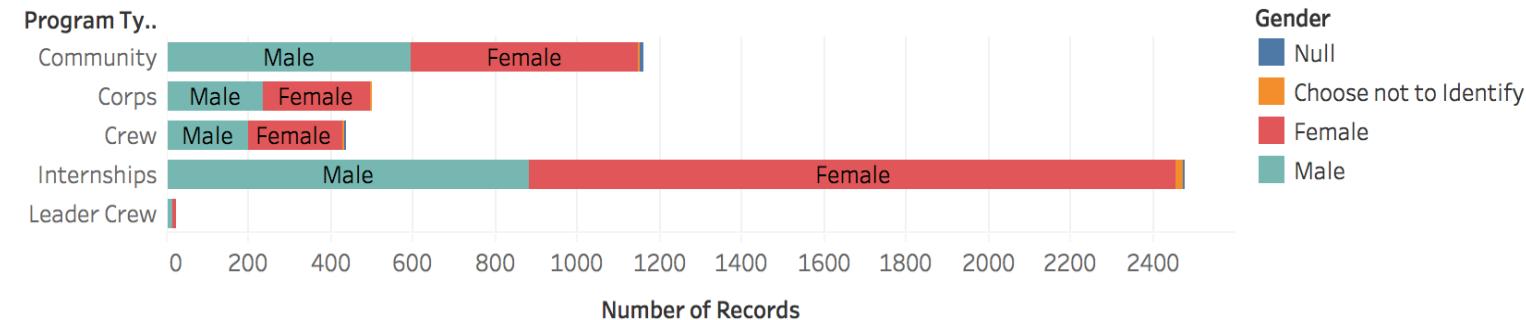


Crew Participants come from wealthier zip codes, Community participants from less wealthy\*

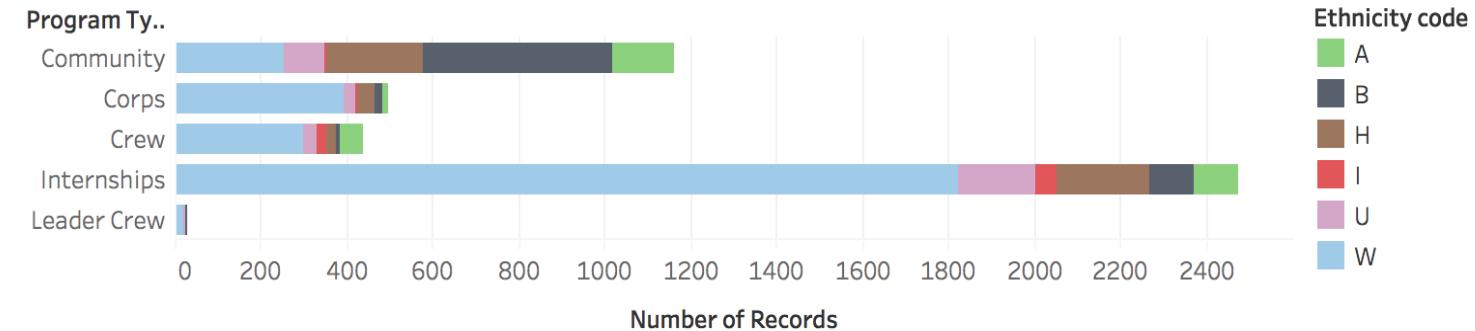


\* bars of different colors are statistically different

Except for Internships, Most Programs Evenly Split Male / Female



Community Programs have More Non-Whites than Other Programs



# Findings for each Program



Internships

Community

Crews

Corps

(Leader Crews not analyzed due to small # of participants)

# Internships Recommendation Scores



## Higher scores from:

- Participants not in AmeriCorps (about 1/3 of participants)
- Asian Americans, African Americans, Hispanics, and Native Americans
- Younger participants\*

### \*Impact of age:

Age	15-17	18-19	20-21	21+
# data pts	11	163	331	1,969
Average score	9.00	8.78	8.77	8.46

## Lower scores from:

- Participants in these budget codes (confidence levels):
  - 206 (99%) – 696 surveys
  - 330 (93%) – 17 surveys
  - 257 (86%) – 460 surveys
- Participants in longer assignments\*\*

### \*\*Impact of assignment length:

# Weeks	13	16	26	52
# data pts	149	295	122	160
Average score	8.89	8.73	8.21	7.99

Not significant: Gender, veteran status, location (or organization managing), economics of home zip code

# Internships Correlated Questions



Higher scores on these questions correlate to higher recommendation scores (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%)

# Community Recommendation Scores



## Higher scores from:

- Programs greater than 8 weeks long vs. programs less than 8 weeks long (average of 8.9 vs. 8.5). This roughly corresponds to weekend vs. summer programs.
- Participants who live in zip codes with higher median incomes and fewer people below poverty line.\*

## Lower scores from:

- Participants in
  - Pittsburgh (320 surveys)
  - Baltimore (20 surveys)
  - Lebanon NJ (6 surveys)
- 25 projects, including some in the above cities. (See appendix for list)
- Age 14 and under participants

Not significant: Gender, job title, budget code

\*99% confidence on median income, 92% on poverty statistic

# Community Correlated Questions



Higher scores on these questions correlate to higher recommendation scores (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



## Higher scores from:

- Older or more educated participants
- Whites and Hispanics

## Lower scores from:

- Participants in
  - Estes Park, CO (22 surveys)
  - Franklin, NC (5 surveys)

Not significant: Gender, economics of home zip code, weather during project, # of weeks for project, budget code

# Crew Correlated Questions



Higher scores on these questions correlate to higher recommendation scores (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



## Lower scores from:

- Participants with these Managing Organizations:
  - Allegheny County, PA (17 surveys)
  - Bighorn National Forest, WY (7 surveys)
  - SCA Oakland Office, CA (5 surveys)
- Participants in this location:
  - Staatsburg, NY (6 surveys)

Not significant: Gender, ethnicity, age, education level, length in weeks, budget code, AmeriCorps participation, economics of home zip code

# Corps Correlated Questions



Higher scores on these questions correlate to higher recommendation scores (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 correlate to higher recommendation scores (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%)

# Summary of Findings



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



# Follow up



- SCA to evaluate the findings
  - Goal is to improve program effectiveness
- Matt is available to
  - Provide further information on any point
  - Delve deeper into areas of interest
- Matt considering doing similar work for other non-profits
  - Missions he believes in
  - Small to medium size organizations are likely target



## Appendix - Community Projects with Lower Recommendation Scores



- Baltimore Community Crew - Storm Water Management
- 2015 NYC High School Community Fall Intern Team
- 2016 Baltimore Community Crew - Farm Team
- 2017 SMC Summer Crew South
- FY16 EBRPD Summer Crew 1 Richmond
- FY16 EBRPD Summer Crew 2 Oakland
- Fall-Winter 2015-16 YAIP Crew
- Friends of the Wissahickon (FOW) 2015
- John Heinz (Kingsessing/Eastwick) School Year Community Crew 1-Eastwick
- NAC Summer Crew 2016
- NJ Community Roving Crew
- NYC SYEP Crew 2015
- Northwest CP- Seattle Crew 3 - FY17
- Peninsula Summer Crew GGNRA
- Pittsburgh Summer: Emerald View/McKinley
- Pittsburgh Summer: Frick
- Pittsburgh Summer: Highland
- Pittsburgh Summer: Point Park & Emerald View Park
- Pittsburgh Summer: Riverview
- Pittsburgh Summer: Roving Crew
- Pittsburgh Summer: Schenley
- Pittsburgh Summer: Southside
- SMC School Year Crew 3 2016\_Team Green
- South Side Park Spring Crew
- Spring 2017 YAIP Crew