Statistical Volunteering With SAS - Experiences and Opportunities

David Corliss, Peace-Work, Plymouth, MI

ABSTRACT

This presentation brings together experiences from SAS® professionals working as volunteers for organizations, charities, and in academic research. Pro bono work, much like that done by physicians, attorneys and professionals in other areas is rapidly growing in statistical practice as an important part of a statistical career, offering the opportunity to utilize your skills in a places where they are so needed but cannot be supported in a for-pay position. Statistical volunteers also gain important learning experiences, mentoring, networking, and other opportunities for professional development. The presenter will share experiences from volunteering for local charities, NGO's and other organizations and causes, both in the US and around the world. The mission, methods and focus of some organizations are presented, including DataKind, Statistics Without Borders, Peace-Work and others.

INTRODUCTION

In many professions, pro bono work – literally, 'For Good" – is an ordinary, even expected part of every career. Physicians may take certain cases without charging for their services and volunteer at a free clinic. Accountants will volunteer for charitable organizations. Attorneys have a long history of pro bono activity, with the American Bar Association effectively requiring no less than 50 volunteer hours each year to remain a member in good standing. Other professional organizations such as the American Medical Association and American Institute of Architects encourage volunteer work while setting no hard and fast expectations, preferring instead to encourage participation and listing benefits of pro bono work in their standards of ethical practice while leaving the details to individuals.

In recent years, volunteer work in statistics and data science have become much more common. Often called "Data for Good", this growing movement has been embraced by individuals, companies, and professional associations. A number of organizations have sprung up to design and manage statistical volunteering projects, recruiting volunteers and connecting them with projects and organizations that will benefit from analytic support. This addresses a prominent need in organizations pursuing good causes: they often have data but lack statistical skills and resources to perform analytics and cannot afford to hire highly skilled consultants. In these situations, statistical volunteers have the opportunity to make a real difference if the world doing the things we love the most.

Statistical volunteering offers many advantages. In addition the primary importance of using our technical skills to support causes and organizations we care about deeply, pro bono also work offers many benefits for career development. While volunteering certainly looks very good on a resume – especially to desirable employers who encourage contribution to the larger community as an ordinary part of a career – there is much more. People early in their career gain important practice and learn new analytic techniques. Volunteer consultants are able to broaden their experiences and gain practical experience in new techniques – something that can be very limited in professional situations, which often focus on the same handful of methods used over and over again for a commercial product. Statistical volunteering is often the best way to try out and enjoy a wide variety analytic tools and emerging technology.

Statistical volunteering often pairs more experienced data scientists and analysts with those having less experience. This naturally leads to mentoring relationships than benefit both parties. All participants benefit from networking with people from different backgrounds, skill sets and work situations. More junior volunteers gain practical experience, the chance to work directly with the people needing the analysis – a quality often very restricted at work, where newer team members can rarely see anyone who directly uses their analysis. Statistical volunteers gain presentation skills, practical working experience and confidence in their work. As experienced is gained, taking a larger role on volunteer projects creates opportunities for leadership, furthering career development.

The most important benefits of statistical volunteering, however, will never appear on a resume but only in the lives impacted by the projects we perform. As statistical and data science professionals, the need for what we do every day on the job is so great but volunteers are still far too few. The Data for Good movement has grown so rapidly because it makes so much more than a good resume: it makes a career by making a real and tangible difference in our profession, our society and our world.

VOLUNTEER ORGANIZATIONS AND EXPERIENCES

The growing Data for Good movement has attracted a number of volunteer organizations, each with a distinctive character. Many involved in statistical volunteering remain unaffiliated, preferring the flexibility and close personal relationships of individual consulting. Also, many companies now offer employees volunteer opportunities through work – although statistical projects are still few and far between, with more advocacy, volunteers, and projects needed in this area. This section describes some of the different organizations and avenues for statistical volunteering, along with experiences of the people involved.

DATAKIND

DataKind is one of the most prominent organizations in in the Data for Good volunteer movement. Founded in 2011 by data scientist Jake Porway, DataKind has grown rapidly to become a large, global organization making a great impact for social change, with thousands of volunteers working on dozens of projects. Headquartered in New York, DataKind has chapters in San Francisco, Washington DC, London, Dublin, Bangalore and Singapore and hosts events around the world. The organization combines the resources of a large number of volunteers with a smaller cadre of professional staff. With substantial financial support from foundations, grants and other donations supporting a multitude of projects, DataKind is a powerful engine for using data and analytic science to make a difference in communities and around the world.

Most commonly, DataKind partners directly with other organizations – not-for-profits, NGOs and government agencies – engaged in addressing problems and creating solutions with data science. DataKind also collaborates with corporate data science efforts, partnering with Informatica, Cloudera, Pivotal, Microsoft, and others. Volunteers often work closely with a project team centered on a particular issue or organization. DataKind also offers opportunities for short-term projects, including engagements with a four-hour stint, and one-day events called Data Dives. While this paper focuses on volunteering, it is noteworthy that, unlike many organizations in the field, DataKind offers some opportunities for direct employment through grants and paid staff positions. For those interested in combining career advancement with statistical altruism, DataKind is well known and well respected in professional circles.

DataKind has a long history of collaboration with the SAS community. DataKind founder Jake Porway was a keynote speaker at SAS Global Forum in 2015. Employees at the SAS Institute have donated to DonorsChoose, an organization support classroom projects by public school teachers, since 2006. DataKind has gotten involved, with volunteer data scientists mining DonorChoose data and providing the results to schools districts, government agencies and other organizations supporting public schools.

One large project now underway is a partnership between DataKind and SAS to optimize bus routes in the Boston Public Schools. This large school district has 125 schools, with 33,000 students riding buses a total of 26,000 every day. This vast optimization problem is being tackled by DataKind, the SAS Advanced Analytics and Optimization Services Team and SAS Educational Practice to develop and analyze multiple scenarios to improve cost and efficiency. In addition to traditional route optimization methodology, the team from DataKind and SAS will examine school start times, having students attending different schools ride the same bus and other strategies to improve a complex and inter-related set of performance metrics. As a result, students will get to and from school faster and the money saved can strengthen educational programs for all the students.

This project displays the most important characteristics of the Data for Good movement – an organization or activity engaged in the public good has the data but lacks the analytic resources to leverage it. Volunteer data scientists provide the needed expertise to turn raw data into action to benefit others.

Boston Public School bus routes: a complex optimization problem

Boston Public Schools
SY 2016 - 2017
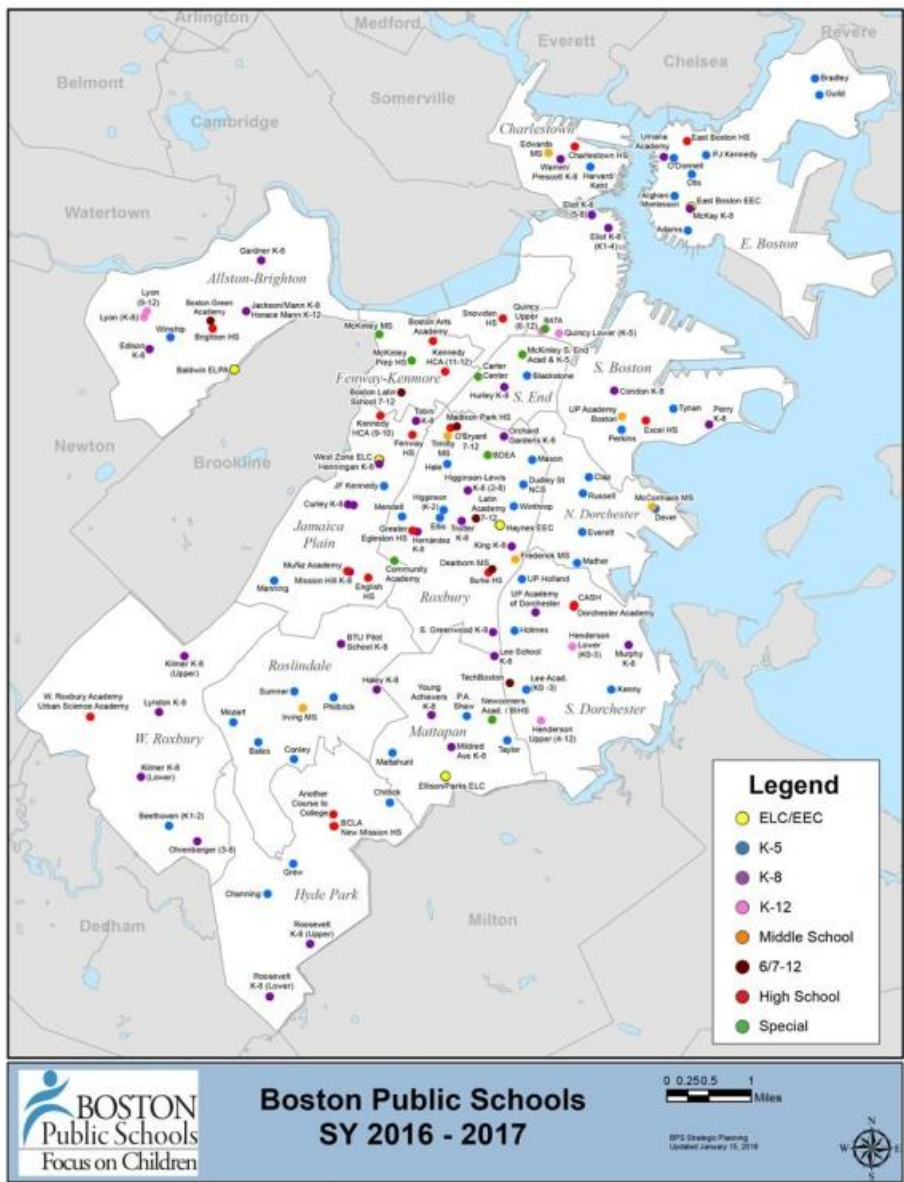
STATISTICS WITHOUT BORDERS

Statistics Without Borders (SWB) is an outreach group from the American Statistical Association. This all-volunteer organization was established in 2008, making it one of the first Data for Good organizations. SWB offers free statistical consulting to governments and Non-Governmental Organizations (NGO). Inspired by Doctors Without Borders and other similar organizations, SWB is non-partisan and secular with much of their work supporting initiatives in the developing world. As with other Data for Good organizations, the nations and organizations SWB supports have data but lack advanced analytic personnel and resources leverage it.

SWB has more than a thousand volunteers, including the professional sector working on a pro bono basis, academia, students and retirees. While this is an outreach of the American Statistical Association, membership in that organization is not required to participate in projects. As an international organization with projects around the world, SWB is especially in need of volunteers fluent in languages other than English and cultural experience and understanding beyond developed nations. In the United States, the

SWB is active at local and national meetings and conferences, which serve as a focal point for networking, reaching out to new volunteers and presenting project results.

As the leading professional organization for statisticians in the United States, the leaders of the American Statistical Association are often approached with ideas for projects. When the need for pro bono statistical consulting occurs around the world, especially in developing countries, SWB volunteers can get involved. Many SWB projects are short-term projects, such as partnering with the Millennium Villages project on a cluster analysis of medical and development programs in Africa and with the AES World Languages and Cultures Institute to set up a research group studying health care services in the presence of language barriers. SWB also has a number of continuing programs, such as technical support and analytic assessment of the Rotary Oceania Medical Aid for Children program and developing a strategy for use of statistical software, including SAS, in developing nations. Since the volunteers tend to work in the analytic platform with which they are most familiar, SAS often plays an important role.

One prominent SWB project addressed the needs for data collection is the wake of the 2010 Haitian earthquake. Dr. Jean Orelien, a public health researcher and a native of Haiti, noted the lack of organization in the early days of the response to the disaster. Realizing the impact data collection could have on improving outcomes, he contacted the American Statistical Association and Statisticians Without Borders became involved. A team of SWB volunteers led by Orelian travelled to Haiti and conducted a survey. While most public infrastructure was very badly damaged, cell phone service was still operational. A cell phone survey was used to conduct a survey of living conditions, especially sanitation and potential disease vectors. Returning to the United States to compile and process the data, the SWB team provided the database to governments and NGOs without cost. The anonymized data is publicly available for download in SAS from Statisticians Without Borders.

Summary Statistics on the impact of the Haitian earthquake from SWB data in SAS
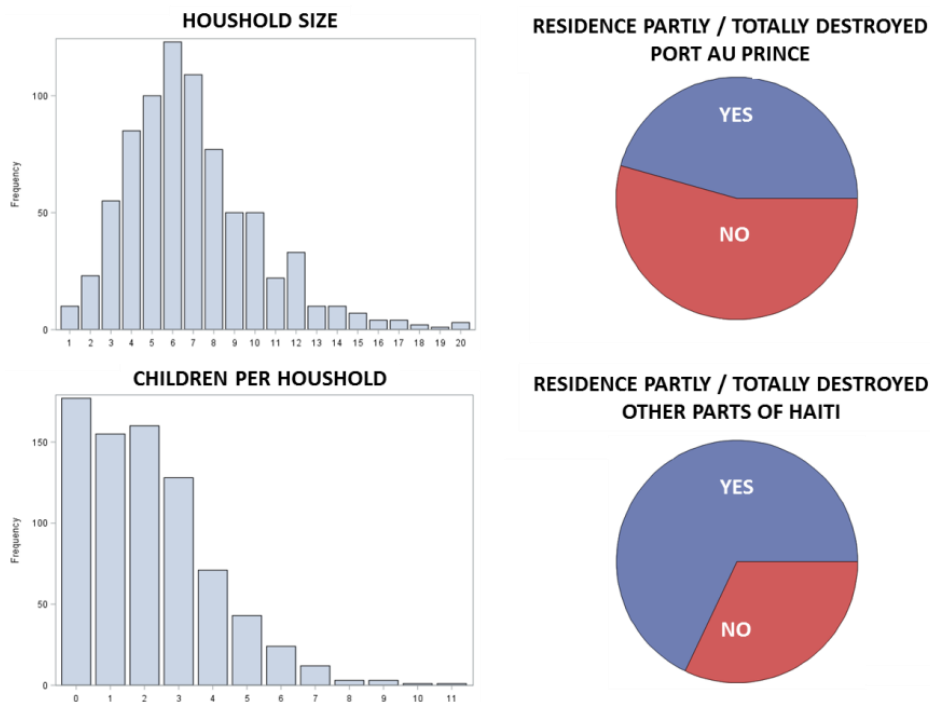
This project shows how an individual statistical volunteer can initiate a project at a Data for Good organization and work with a team of volunteers to make a real difference affecting many people.
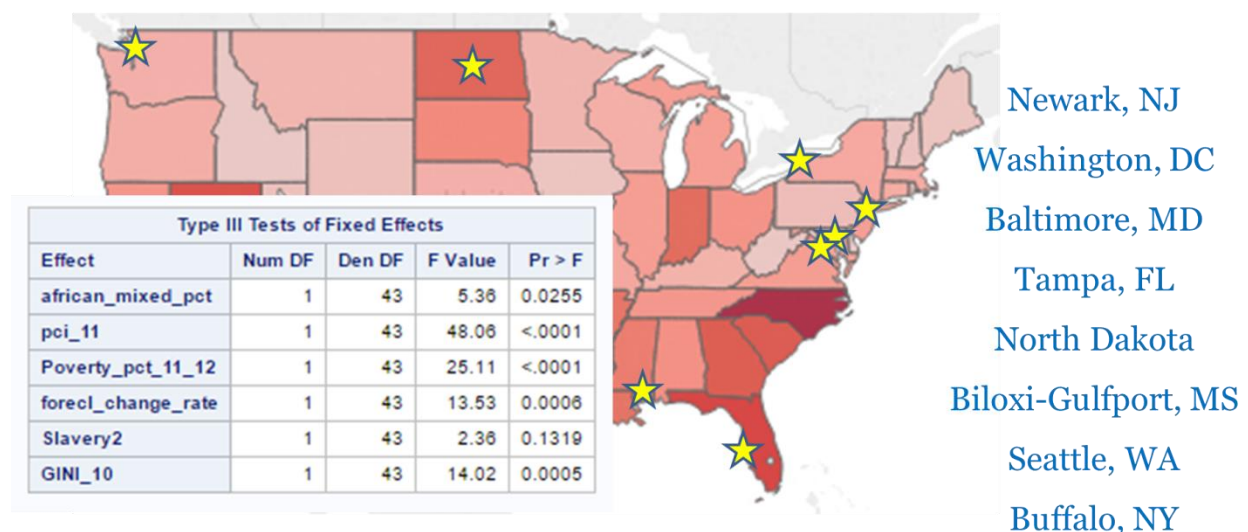
PEACE-WORK

Peace-Work is a volunteer cooperative of statisticians, data scientists and other researchers applying analytics to issues in poverty, education and social justice. Although Peace-Work is the new kid on the block in terms of organizations, currently filing for official Not-For-Profit status, its founding members have a long history in individual statistical advocacy. An all-volunteer organization, Peace-Work projects are often in the area of academic and policy research, with volunteers as likely to be found working with government economic data to write a position paper for publication as to work hand-in-hand with a social justice organization. With a practice focusing on issue-driven advocacy, projects have included education performance metrics, root cause analysis of homelessness, descriptive statistics of privilege and the impact of racial bias, among others.

Peace-Work operates by connecting volunteers with organizations and datasets, often from governments and other official sources, to address issues volunteers care about deeply. With no full-time staff, larger statistical project are broken into small pieces to accommodate volunteers busy schedules. Peace-Work has recently begun sponsoring events at analytic conferences, beginning at the Midwest SAS User Group conference in 2016. Called "Distributed Justice Events", these activities invite conference participants to contribute volunteer hours during the conference on projects targeted local issues.

While volunteers rely on the software tools of their own choosing, SAS is used in many projects. Peace-Work actively promotes the use of SAS University Edition as a powerful analytic tool with extensive statistical, machine learning and big data capabilities available to their researchers for free.

One current Peace-Work project focuses on the problem of human trafficking, which has attracted considerable interest in recent years. The National Human Trafficking Resource Center (NHTRC) operates a national hotline for victims. Peace-Work volunteers combined state-level summary statistics published on the NHTRC website with demographics and socio-economic data. SAS University Edition was used to perform a meta-analysis, identifying factors predicting high level of human trafficking while accounting for variations between states. The fixed effects found by the state-level meta-analysis were then applied to data on large cities to create a decision tree model to identify potential unidentified centers of human trafficking activity.

Peace-Work model indicating previously undetected centers of human trafficking

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
african_mixed_pct	1	43	5.36	0.0255
pci_11	1	43	48.06	<.0001
Poverty_pct_11_12	1	43	25.11	<.0001
forecl_change_rate	1	43	13.53	0.0006
Slavery2	1	43	2.36	0.1319
GINI_10	1	43	14.02	0.0005

Newark, NJ
Washington, DC
Baltimore, MD
Tampa, FL
North Dakota
Biloxi-Gulfport, MS
Seattle, WA
Buffalo, NY

Peace-Work has begun to partner with local agencies to repeat meta-analysis of human trafficking at the level of metropolitan areas. As variations between states appears to be driven, in part, by differences in state laws, work has begun to advocate for changes to implement across the country best practices for finding traffickers and supporting victims.

VOLUNTEERING THROUGH WORK

Companies often encourage employees to volunteer in the community where they work, with many offering time off for teams or individuals volunteering for local organizations. Companies engaged in data science and technology have sponsored projects where groups of employees apply the same skills they use on the job to support important causes on a pro bono basis.

Well-known employers supporting statistical volunteering include Cloudera, which recently sponsored a hackathon to compile data on the spread of the Zika virus. AS described above, the SAS DataKind has partnered with Teradata, Pivotal, Informatica and others on projects. The number of companies supporting the Data for Good movement with volunteer opportunities continues to grow. In addition to large corporations with substantial HR departments to encourage and coordinate statistical volunteering, many small consulting companies and Analytic Services Providers are involved and often have the most flexibility in setting up Data for Good volunteering programs at work.

Company volunteer programs usually have a particular set of ground rules. Projects and the organizations receiving support should be approved through the HR department or, in the case of smaller companies, the Operations lead. Projects must not interfere with or delay regular work for company customers. Companies will often choose the cause they want to support and then seek employees to volunteer, generally preferring activities that support important needs in the community. As the company will want to let others know about the good they are doing in the community, communication with HR or operations, an employee newsletter and even the outside press might be a requirement. Employee volunteer projects often serve as team building activities, so it's a good chance several of your co-workers will be involved. Many companies will give a limited, specified amount of time off for volunteering through work, often one day a year or sometimes two. As the company is performing the work as a charitable activity, careful documentation is needed for tax purposes. This is usually done by a Team Captain who submits a request for an employee volunteer activity, helps recruit volunteers, records attendance in hours for each participating employee, and reports by to HR or Operations about how everything went.

One advantage of volunteering through work is the company often makes available analytic resources and software the volunteers use every day. The familiar SAS environment at work usually becomes the platform for pro bono projects as well, maximizing the efficiency and effectiveness of statistical volunteering using SAS.

Microsoft's MySkills4Africa is an excellent example of a Data for Good program created by large tech corporation relying on employee volunteers. Established in 2013, MySkills4Africa offers Microsoft employees from around the world the opportunity to develop practical and affordable technology in developing African nations. The Microsoft program partners with governments, NGOs, commercial enterprises and schools and universities. Characteristic of MySkills4Africa projects is the opportunity to use advanced skills normally applied only at work and use them in service to others as part of the Data for Good movement. Many MySkills4Africa projects involve longer-term consulting projects working virtually with teams in Africa. Microsoft also sponsors 1-2 week hands-on projects in Africa. One project developed a team of Subject Matter Experts to partner with the Rwanda Ministry of Education. Months of preparation - all done as volunteers, in addition to work responsibilities – culminated in a two week trip to Rwanda to train professionals and small business owners in emerging technologies. At the same time, the employee team worked with Rwanda education officials to establish an on-going program to continue the good work.

Employee volunteering is supported by many companies but not all of them have started projects using analytics and data science. If there are not volunteer opportunities on the Data for Good projects where you work, a contact with your HR department might be a good place to start.
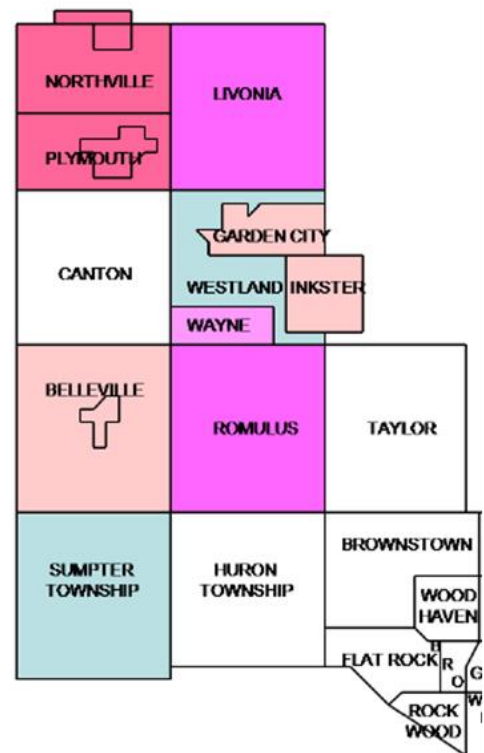
INDIVIDUAL VOLUNTEERING

Of course, not all or even most statistical volunteering goes through an organization. Individuals working on their own make up a large portion of the Data for Good movement. Often, a personal connection to a group needing statistical help is the most important factor in matching people and projects.

Working with a volunteer organization offers many advantages, connecting people to volunteer projects and groups needing assistance, providing mentoring, software support, project planning and opportunities

to work on larger projects. Working through an established organization can do more career development through mentoring, working with a larger team and name recognition for your resume. However, working as an individual or in a small group of colleagues and friends offers substantial advantages of its own, including complete creative control of methodology and analytic practices, greater flexibility in which analytic tools are used, working with a circle of friends in familiar surroundings. Individual volunteer projects can be less formal, without a definitely timeline, and often can be put on a shelf for a time and re-started later when more time is available.

The nature of individual volunteering and the opportunities it offers can be seen in a project by the author for a local Habitat for Humanity chapter in 2005. After volunteering on a number of construction projects and getting to know the chapter's Board of Directors, the question was raised of how data science can be used to support the organization. The chapter is located in western Wayne County, Michigan, in the western suburbs of Detroit. Analysis of an anonymized list of donors and construction volunteers revealed most of the support for the chapter came from a small portion of their geographic area, with many cities largely unreached. A cluster analysis binned all of the communities in the chapter's area into one of three groups, based on level of charitable need in the community and their ability to meet it. Each of the three clusters, classifying all of the communities in the area served by Western Wayne County Habitat for Humanity, were matched with the zip code counts – no information that could identify an individual! Looking at the whole area, instead one small part where most of the board members lived, produced a list of communities to target where the organization had very little presence. This Data for Good project had an unintended consequence – for the good! In addition to finding places to search for more donors and volunteers, it documented that the one place where the local Habitat group was building houses was very poor and weak infrastructure and especially poor schools. This produced a recommendation to build house in a different, neighboring community where the families would receive more support and the children attend better schools – doing more good for the families in new homes even if the land prices there were somewhat higher.

Habitat for Humanity donor and volunteer analysis - Western Wayne

This project tells the story of how most statistical volunteering by unaffiliated individuals happens: a person is already helping an organization in a non-statistical way – working at the public library, walking dogs at the local animal shelter, volunteering at a school, house of worship or community center – any of a hundred things. Individual statistical volunteering happens when people with analytic skills know how those skills can be used to help the organizations and causes they already support. It's such a rare skill, so needed by groups across our communities that it can make a huge impact right where we live every day.

HOW TO GET INVOLVED

There are many excellent organizations in the Data for Good movement, including

DataKind: http://www.datakind.org/

Statistics Without Borders: http://community.amstat.org/statisticswithoutborders/home

Peace-Work: davidjcorliss@peace-work.org

DataDriven – "Data science competitions to save the world": https://www.drivendata.org/

Data for Good researchers in the SAS community will want to look at SAS University Edition. More information is available at: http://www.sas.com/en_us/software/university-edition.html

SAS blog by Becky Gaebe on becoming a part of the Data for Good movement: http://blogs.sas.com/content/sascom/2015/05/01/data-for-good/

CONCLUSION

The writer Frederick Buechner observed a person's calling is found "where your deep gladness and the world's deep hunger meet." Participation in the growing Data for Good movement provides opportunities to use our technical skills to support the organizations, projects and causes we care about the most. As the movement continues to grow, we can envision a day when pro bono activity is considered to be more than an added benefit, becoming a normal, ordinary part of a career in data science, statistics and analytics. Participation is encouraged from everyone from students through the most experienced and entry-level analytic workers through top executives. Finding your place of deep gladness and sharing your skills for the greater good can be a part of everyone's career journey. Where is your place?

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David J Corliss
Peace-Work
davidjcorliss@peace-work.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.