DataKind

An Exploratory Study on the Data for Good Movement and Volunteer Motivations

Benjamin Kinsella, PhD
benjamin.kinsella@datakind.org

Introduction

1. Overview and Data for Good volunteering
2. Methods
3. Findings
4. Conclusion/Future Research

DataKind

DataKind

We help social organizations identify their data and AI opportunities.

We recruit and manage a team of pro bono experts to build their solutions.

We stick around to ensure the solutions are used and impact is achieved.

DataKind
by the Numbers

From evening or weekend events to multi-month projects, our programs are designed to provide social organizations with the pro bono data science innovation team they need to tackle critical humanitarian issues.

20,000+
community members worldwide in **98** countries, representing the largest global data science for social good network

5 global chapters

250+ events around the world

Volunteer sign-ups from **174** countries

300+ projects completed, providing the most comprehensive library of data science for social good projects

150+ organizations helped

200,000+ hours donated

$35M+ pro bono services delivered

Overview

- Volunteerism and motivation:
  - Functional approaches (Clary et al., 1998)
  - Behavioral and psychological (Stukas, Snyder, & Clary, 2015)
  - External and extrinsic (e.g., Snyder et al., 2000)

- Data for Good (D4G) volunteering:
  - A growing global movement characterized by the use of data science and AI techniques that solve important societal challenges vis-á-vis a prosocial volunteer framework.

- Value on skills-based volunteerism, thoughtfulness of the acknowledged risks of AI and data science and UN's 17 Sustainable Development Goals (see Chui et al., 2018, Varshney and Mojsilovic, 2019)

Overview

- Several models of engagements:
  - Short term data science competitions/hackathons
  - Corporate philanthropy (e.g., TwoSigma Data Clinic).
  - Specialized nongovernmental organizations (e.g., Bayes Impact) and innovation units within large development organizations (United Nations Global Pulse).
  - Longer term volunteer engagements (e.g., DataKind's DataCorps model)

Data for Good Volunteering

Questions:

- What are the **characteristics** of D4G volunteers (e.g., geography, gender, age, technical skills, etc.)?
- What are the **underlying motivations** of volunteers?
- What **opportunities/limitations** exist across D4G volunteers?

Methods

- Survey:
  - Multiple choice, Likert scale, open ended text

- Three sections:
  - General background (e.g., background, geography, etc.)
  - Functional categories of motivation (Clary et al., 1998)
  - Open ended text (e.g., obstacles)

- Sample collected through subset of DataKind's listserv (approximately N = 8,000)

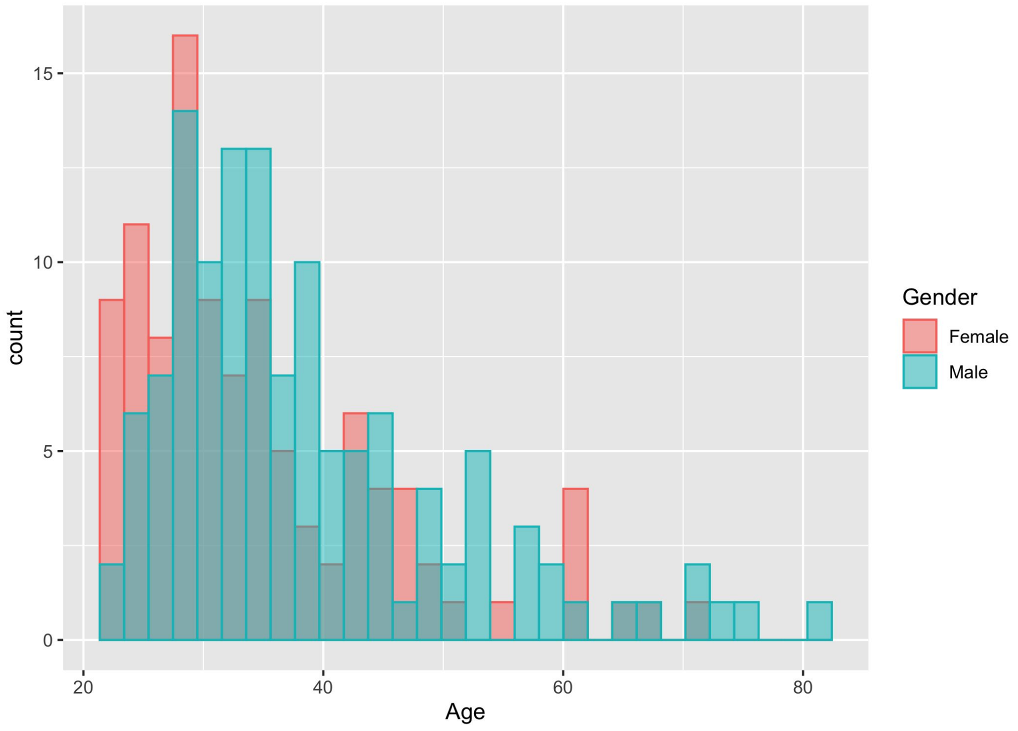
DataKind

Findings

Background and Characteristics

General Background

- N = 238
- Age:
  - Range: 22-81, Median: 34
- Gender:
  - Male: 53%
  - Female: 45%
  - Non-Binary 1.3%
- Education
  - Masters: 58%
  - PhD: 21%
  - Undergraduate: 20%



DataKind

Respondents

DataKind

Motivations and Event Interest

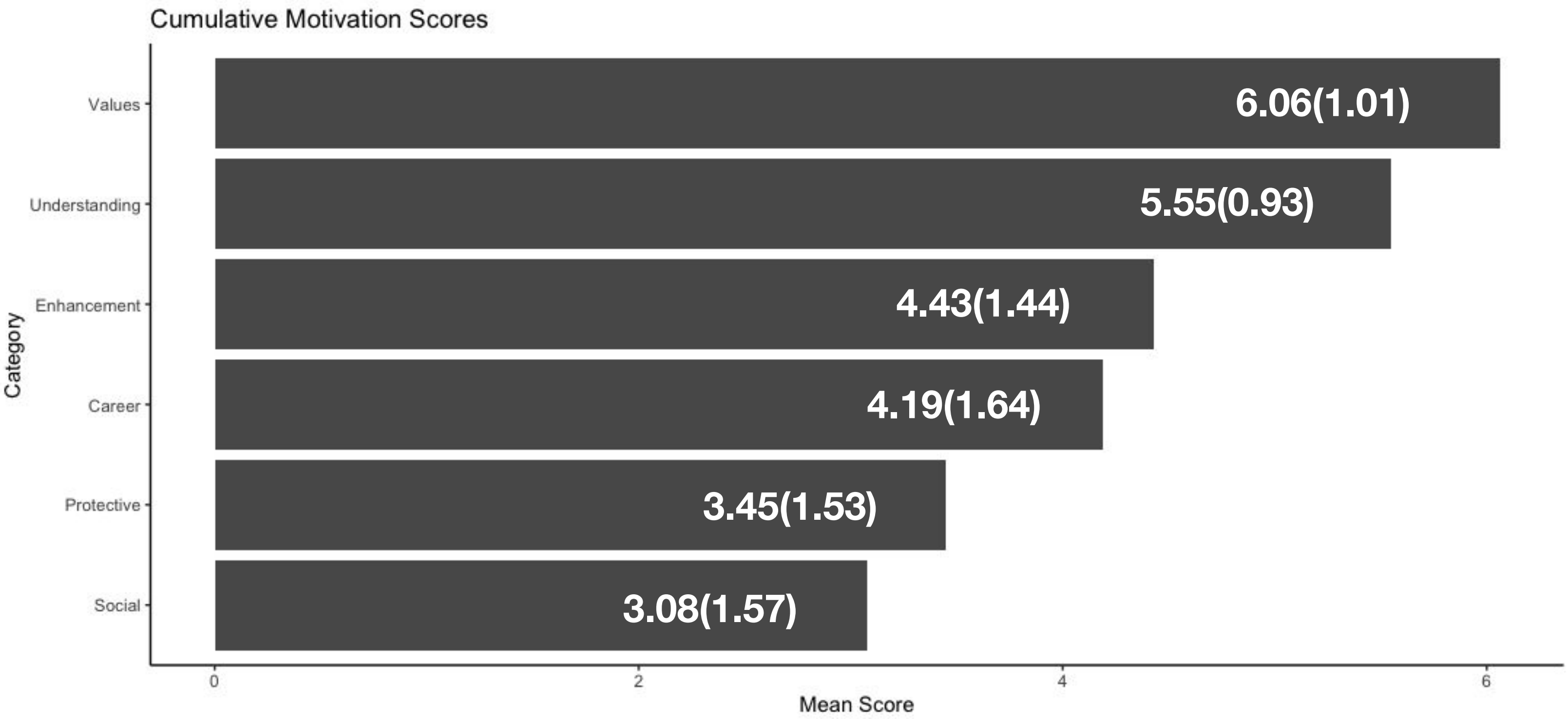
Functional Motivational Categories to Volunteering

Table 1: Clary, Snyder and Ridge (1998) functional categories

Motivation	**Question example**
Protective	"Volunteering on a D4G project helps me put my personal problems into perspective"
Values	"I feel it's important to help others."
Career	"Volunteering on a D4G project can help me to get my foot in the door where I'd like to work."
Social	"People I'm close to *want* me to volunteer."
Understanding	"Volunteering on D4G projects lets me learn through direct, hands on experience."
Enhancement	"Volunteering on D4G projects makes me feel useful."

DataKind

Cumulative Mean Scores of Functional Categories



DataKind

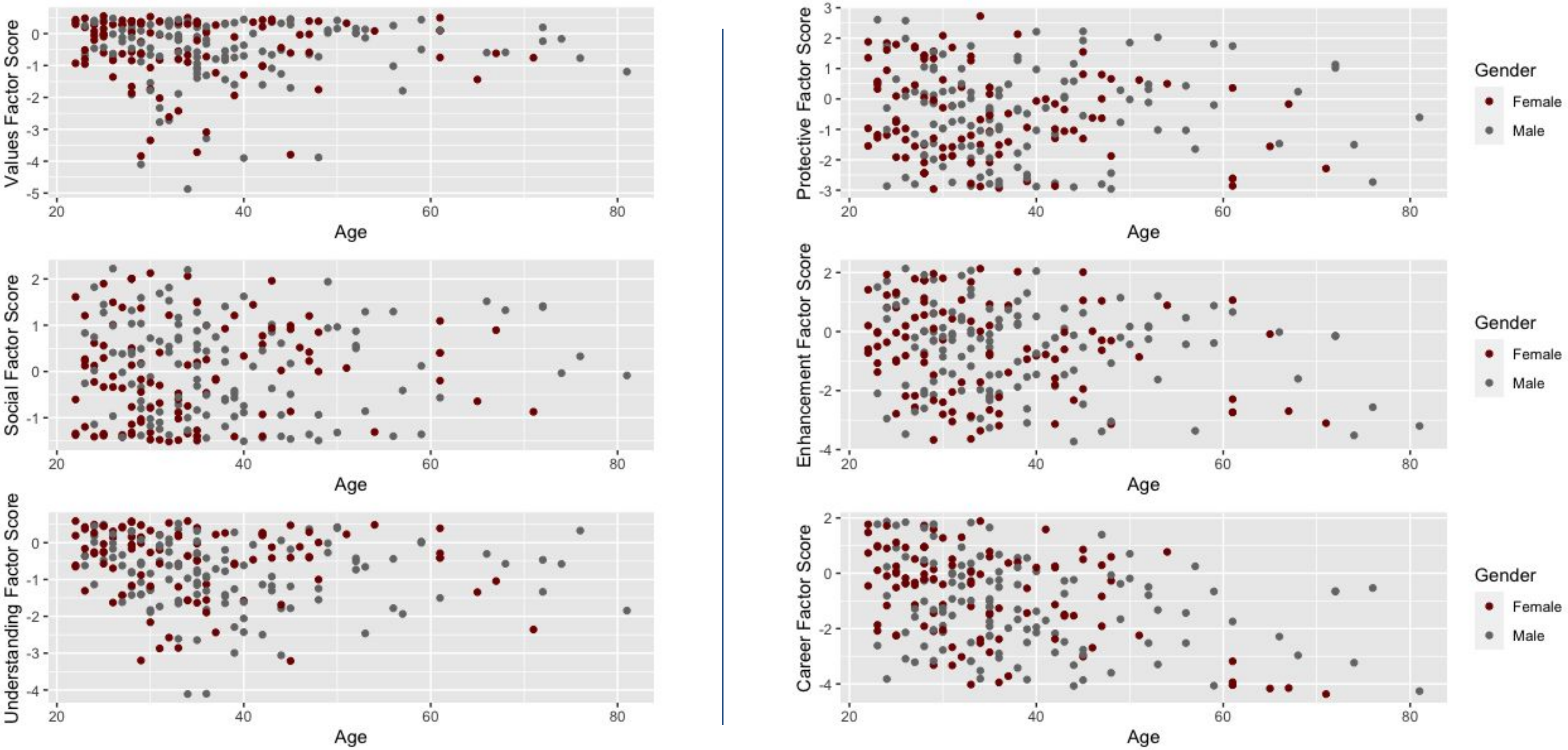
Confirmatory Factor Analysis

- CFA was conducted on latent variables for each category using Lavaan package in R (Roseel et al., 2020)

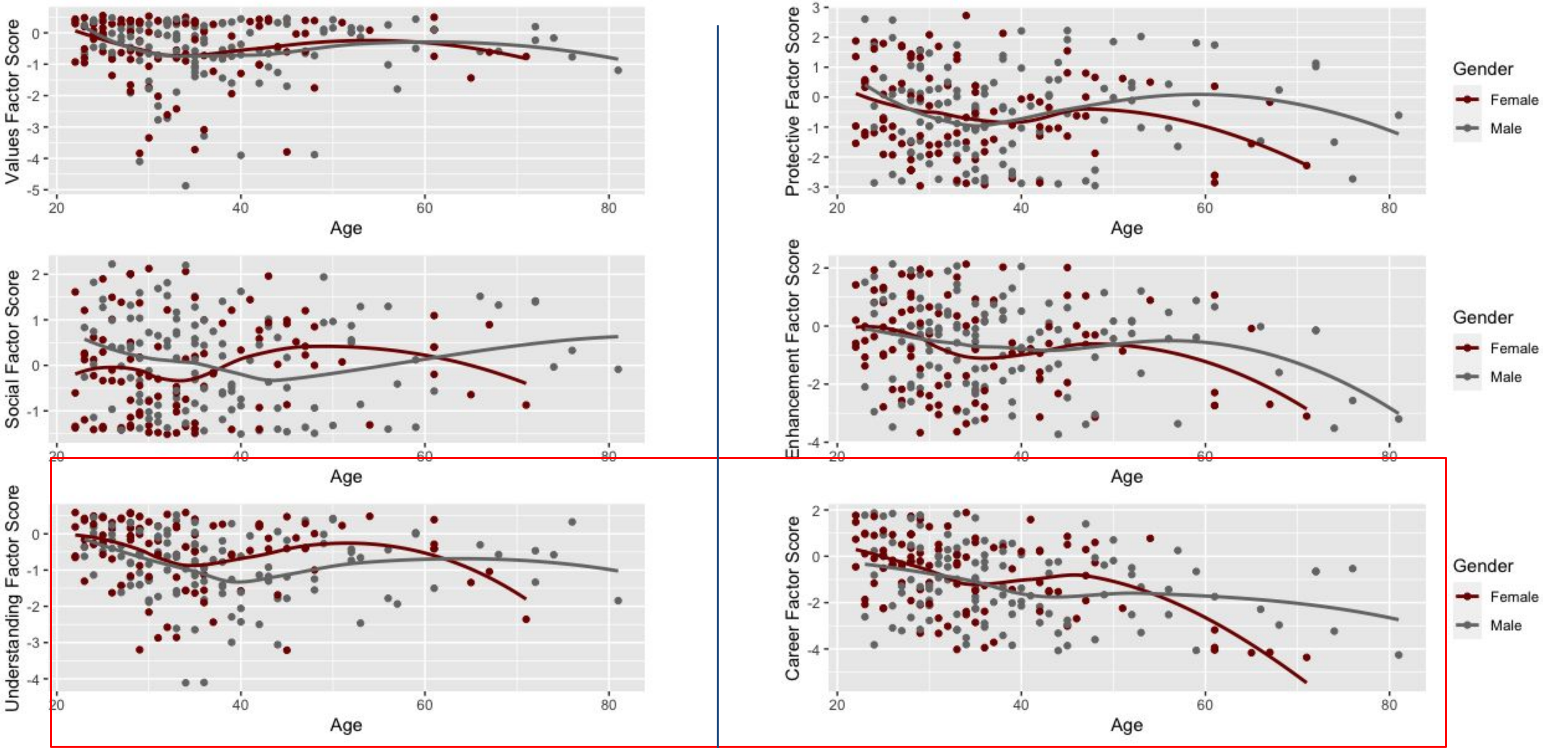
- Fit indices:
  - CFI (0.87)
  - RMSEA (0.071)
  - 90% confidence interval (0.066 and 0.077).
  - The standardized root mean square residual was 0.071
  - Chi-square and log of likelihood test significant to the 1% level.

- Model appears to be a decent fit to the data.

Regression - Factor Scores for Each Category



Regression - Factor Scores for Each Category



Category	R^2
Values	0.003 3
Protective	0.021
Social	0.017
Enhanceme nt	0.05
Understand- ing	0.102
Career	0.147

DataKind

Regression - Career Factor

$R^2 = 0.147$

DataKind

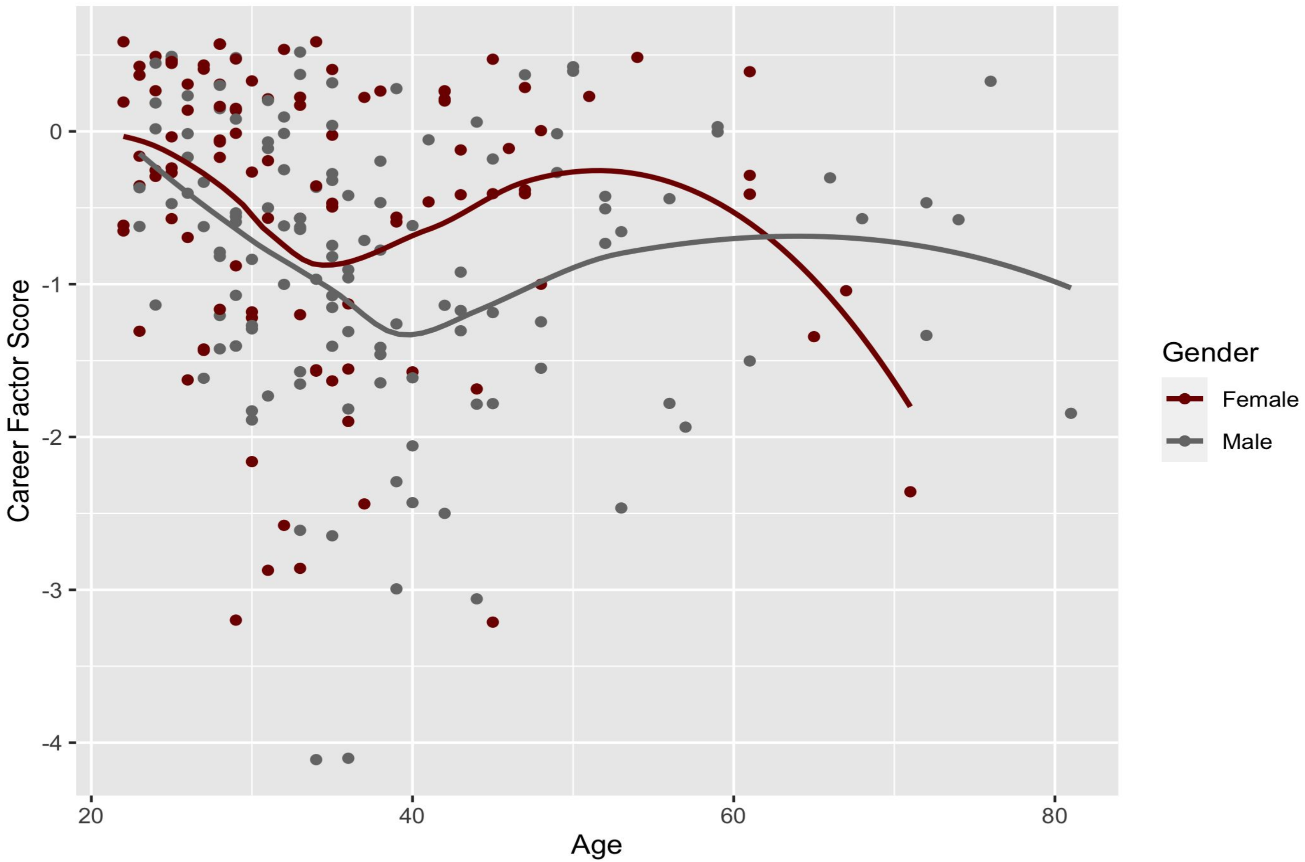
Regression - Career Factor

Group	Estimate	std.all	Std. Err	z-value	p-value
25-34	-0.621	-0.121	0.351	-1.769	0.077
35-44	-0.656	-0.191	0.277	-2.368	0.018
45-54	0.382	0.087	0.418	0.915	0.360
55-64	-1.692	0.280	0.613	-2.763	0.006
65+	0.159	0.018	0.743	0.213	0.831
Gender (Male)	-0.178	-0.053	0.223	-0.801	0.423

DataKind

Regression - Understanding Factor

$R^2 = 0.102$



DataKind

Regression - Understanding Factor

Group	Estimate	std.all	Std. Err	z-value	p-value
25-34	-0.474	-0.153	0.225	-2.106	0.035
35-44	-0.415	-0.200	0.178	-2.330	0.020
45-54	0.511	0.192	0.268	1.905	0.057
55-64	-0.157	-0.043	0.388	-0.404	0.686
65+	-0.139	-0.026	0.473	-0.294	0.769
Gender (Male)	-0.301	-0.147	0.143	-2.102	0.036

DataKind

Obstacles that prevent volunteering

D4G Volunteering Obstacles

Theme	Description
Family obligations	Time constraints given family obligations and childcare.
Work commitments	Balancing work commitments, as well as legal restrictions imposed by employers
Long-term planning	Coordination in advance to prepare for long-term engagements.
Range of required tech skills	Discovery of new projects that are aligned with skill sets outside traditional data science.

D4G Volunteering Obstacles

“I understand why there's a **professionalization** driven by DataKind, but I find it puts up barriers if you're being asked to commit quite a lot of time…you're **excluding people** with families, children, caring responsibilities etc.” *Female, 35, London, UK.*

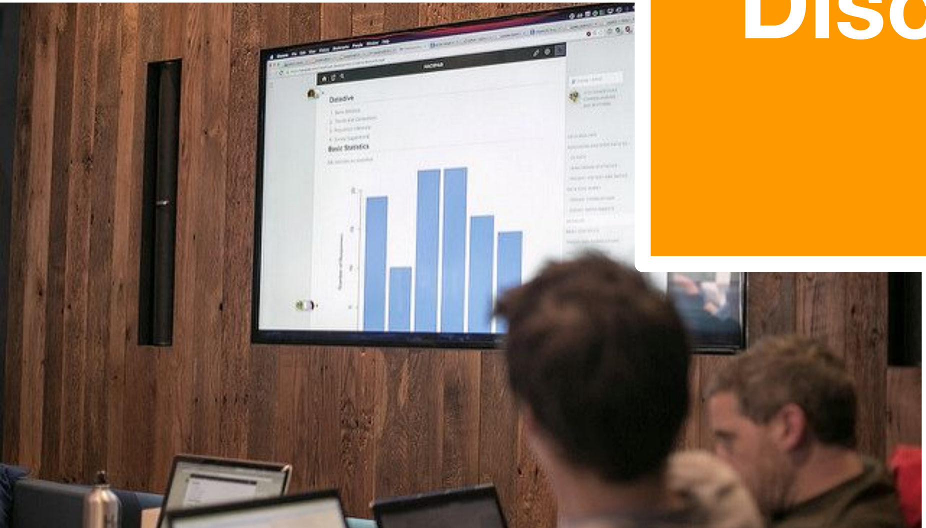
“I'm the **primary caregiver** for my infant daughter, so any work needs to fit in early morning, evenings, and weekends. A **flexible schedule** with minimal **"synchronous"** components is important for me right now.” *Female, 35, Austin, TX.*

D4G Volunteering Obstacles

“**Legal restraints**: My current employer (Google) doesn't really allow side projects using the same skills we use in our day-to-day job. I'd need to get legal approval to participate.” *Male, 45, New York, NY.*

“I'm hoping to volunteer as much as my schedule allows. I may be limited in that **my technical background in data science isn't very cutting edge** - I'm more comfortable in the data exploration, data viz, or even project development areas. I've recently been volunteering with a local food pantry where my biggest value is as someone who's familiar with Google Sheets! It's been so fulfilling”, *Female, 35, Austin, TX.*

Discussion



Implications

- Align activities with volunteer motivations
  - Offer learning and skill building activities, as well as those related to career enhancement; also consider motives for adults.

- Expand new pathways for engagement
  - Incorporate new technologies and virtual work
  - Consider other skills, in addition to core data science approaches

- Attract and retain diverse volunteers through partnerships and program design
  - Build and deploy ethical solutions with end users and ensure long-term sustainability of projects through diverse representation on projects

Conclusions and Future Research

- Sample included several limitations:
  - D4G volunteering is broad (e.g., different data science skills, career trajectories, and geographies)
    - Cannot be generalized to other types of skill-based volunteering (Open Source Software for Good) and different cultural contexts
  - Respondents had opted in to receive emails from DataKind's listserv. So it may be that some of the study's findings do not generalize to individuals who are outside the DataKind community
  - Estimated response rate volunteers was low, only 5% of total reach.

Acknowledgments

We would like to thank our contributors who volunteered to help develop the survey and analyze findings:

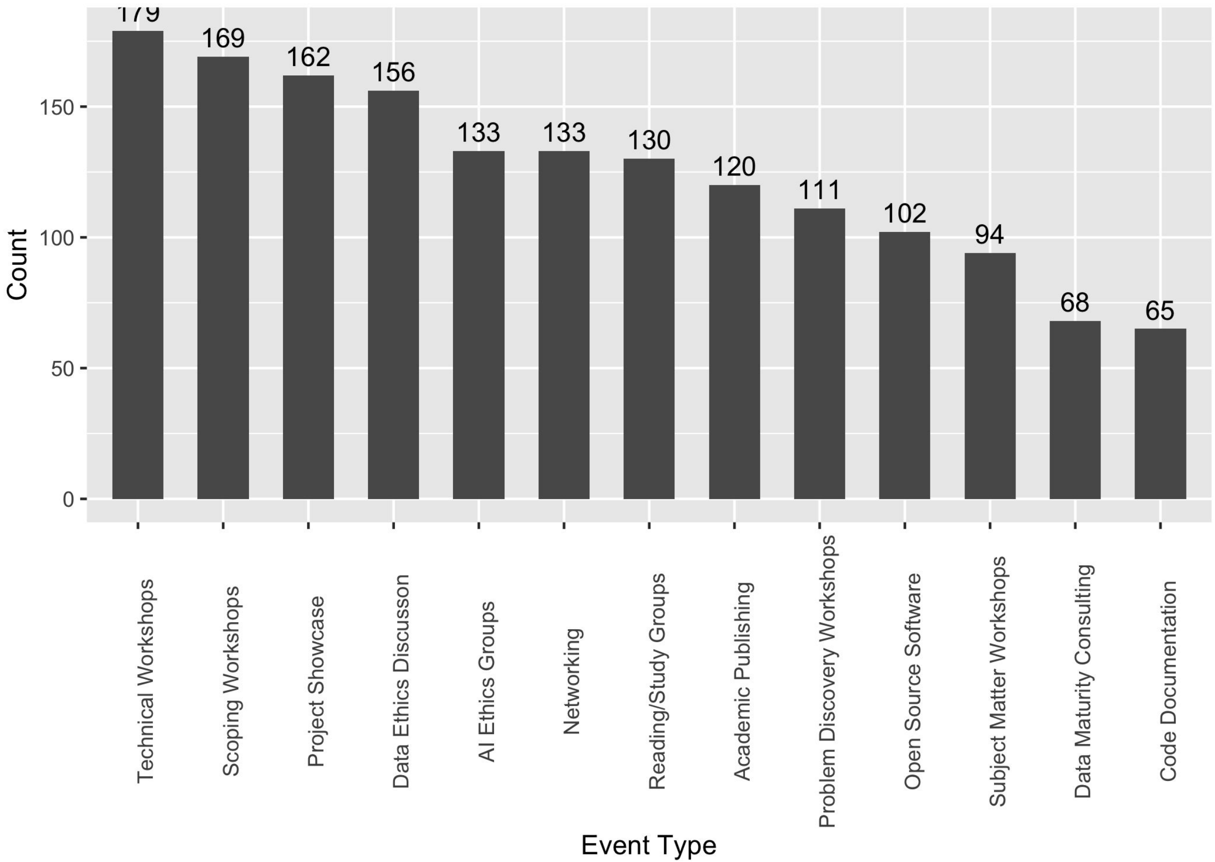
- Daniel Nissani
- Manojit Nandi
- Neal Fultz
- Jack Craft
- Camille Metzinger
- Paula Hwang
- Smit Mehta

Your volunteer insight and technical expertise were critical in interpreting these findings!

DataKind

WE DATA

Event and Activity Interest



DataKind