Breakout Session 7:

ScHARe - Science collaborative for Health disparities and Artificial intelligence bias Reduction

Dr. Deborah Duran (Moderator)
Senior Advisor, Office of the Director, NIH/NIMHD

ScHARe

Science **c**ollaborative for
Health disparities and
Artificial intelligence bias **Re**duction

STRIDES Performance

Deborah Duran, PhD · NIMHD

January 18, 2024

National Institutes of Health

NIH National Institute on Minority Health and Health Disparities + NIH Office of Data Science Strategy + NIH National Institute of Nursing Research

Thank you

NIMHD

Dr. Eliseo Perez-Stable

ODSS

Dr. Susan Gregurick

NIH/OD

Dr. Larry Tabak

NINR

Dr. Shannon Zenk

NINR

Rebecca Hawes
Micheal Steele
John Grason

ORWH

OMH

NIMHD OCPL

Kelli Carrington
Thoko Kachipande
Corinne Baker

BioTeam

STRIDES

Terra

SIDEM

RLA

Broad Institute

CDE Working Group

Deborah Duran
Luca Calzoni
Rebecca Hawes
Micheal Steele
Kelvin Choi
Paula Strassle
Deborah Linares
Crystal Barksdale
Gneisha Dinwiddie
Jennifer Alvidrez
Matthew McAuliffe
Carolina Mendoza-Puccini
Simrann Sidhu
Tu Le

ScHARe

ScHARe is a cloud-based population science data platform designed to accelerate research in health disparities, health and healthcare delivery outcomes, and artificial intelligence (AI) bias mitigation strategies

ScHARe aims to fill three critical gaps:

- Increase participation of **women & underrepresented populations with health disparities** in data science through data science skills training, cross-discipline mentoring, and multi-career level collaborating on research
- Leverage population science, SDoH, and behavioral Big Data and cloud computing tools to foster a **paradigm shift** in healthy disparity, and health and healthcare delivery outcomes research
- **Advance AI bias mitigation and ethical inquiry** by developing innovative strategies and securing diverse perspectives

nimhd.nih.gov/schare

ScHARe

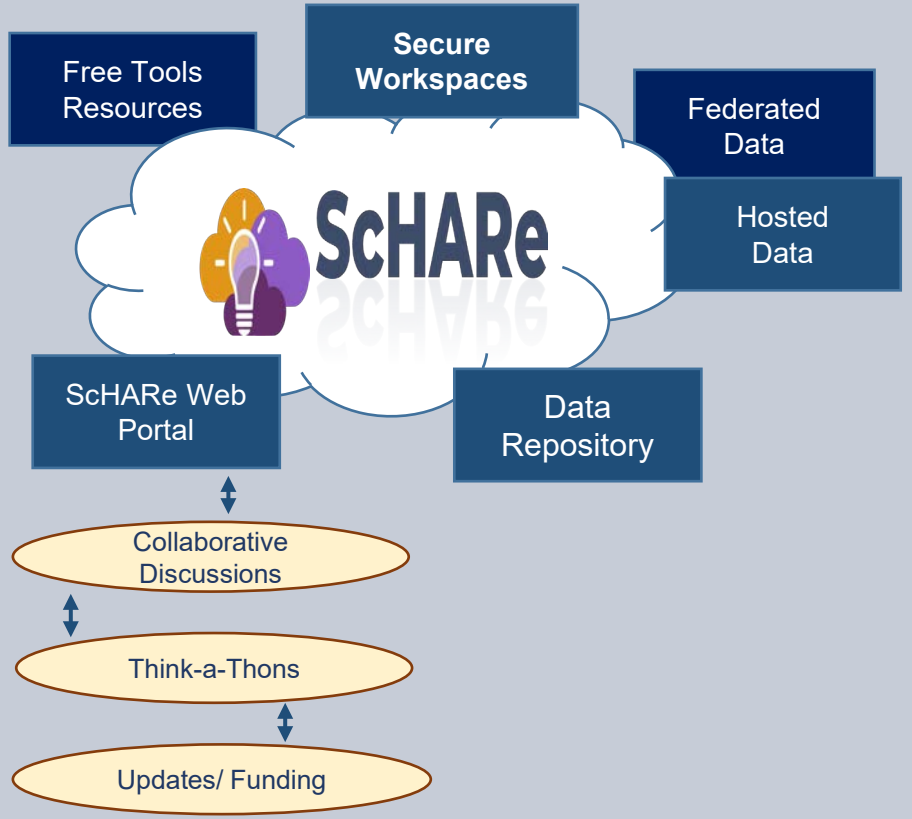
ScHARe co-localizes within the cloud:

- **Datasets** (including social determinants of health and social science data) relevant to minority health, health disparities, and health care outcomes research
- **Data repository** to comply with the required hosting, managing, and sharing of data from NIMHD- and NINR-funded research programs
- **Computational capabilities** analytic tools, resources, code
- **Secure, collaborative workspaces** for students and all career level researchers
- **Tools for collaboratively evaluating and mitigating biases** associated with datasets and algorithms utilized to inform healthcare and policy decisions

Frameworks: Google Platform, Terra, GitHub, NIMHD Web ScHARe Portal

Components

Intramural & Extramural Resource



nimhd.nih.gov/schare

ScHARe Data Ecosystem

OVER 240 DATA SETS CENTRALIZED

Researchers can access, link, analyze, and export **a wealth of datasets** within and across platforms relevant to research about health disparities, health care outcomes and bias mitigation, including:

- **Google Cloud Public Datasets:** publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program

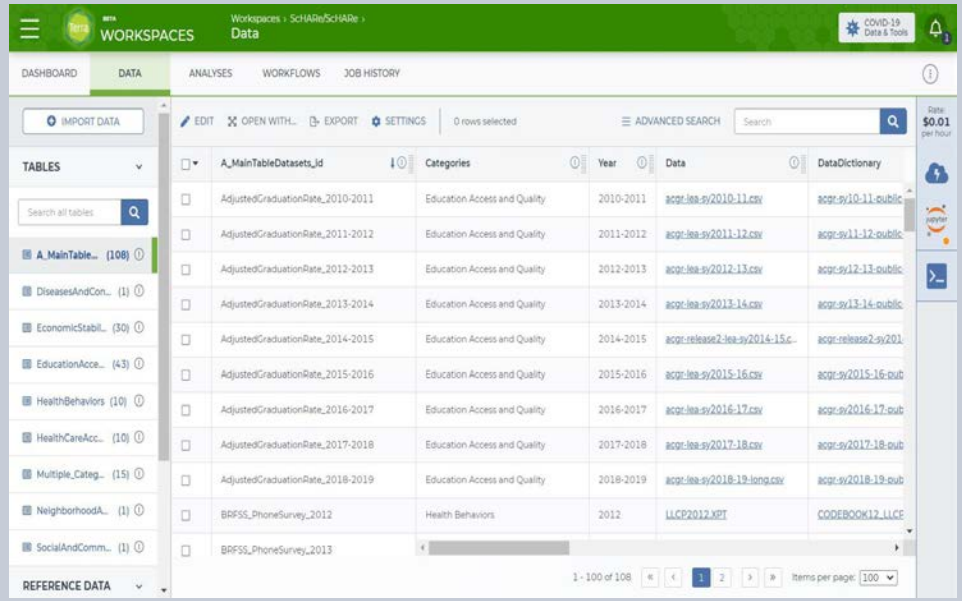
  Example: *American Community Survey (ACS)*

- **ScHARe Hosted Public Datasets:** publicly accessible, de-identified datasets hosted by ScHARe

  Example: *Behavioral Risk Factor Surveillance System (BRFSS)*

- **Funded Datasets on ScHARe:** publicly accessible and controlled-access, funded program/project datasets using Core Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy

  Examples: *Jackson Heart Study (JHS); Extramural Grant Data; Intramural Project Data*



Datasets are categorized by content based on the CDC **Social Determinants of Health categories**:

1. Economic Stability
2. Education Access and Quality
3. Health Care Access and Quality
4. Neighborhood and Built Environment
5. Social and Community Context

with the addition of:

- **Health Behaviors**
- **Diseases and Conditions**

Users will be able to **map and link** across datasets

Data Ecosystem Structure

FEDERATED PUBLIC DATA +240

(Hosted by Google & ScHARe)

REPOSITORY

CDE FOCUSED

CDEs enhances Data Interoperability (Aggregation) by using semantic standards and concept codes

What is a CDE?

A common data element (CDE) is a standardized, precisely defined question that is paired with a set of specific allowable responses, that is then used systematically across different sites, studies, or clinical trials to ensure consistent data collection

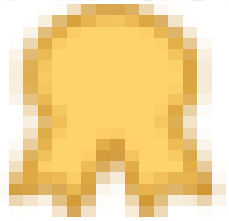
Innovative Approach: CDE Concept Codes Uniform Resource Identifier (URI)

Ecosystem: ScHARe hosted PROJECT DATA

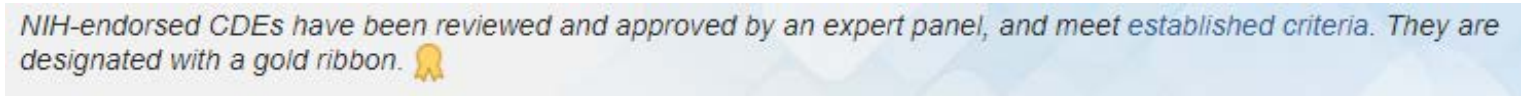
- Age
- Birthplace
- Zip Code
- Race and Ethnicity
- Sex
- Gender
- Sexual Orientation
- Marital Status
- Education
- Annual Household Income
- Household Size

- English Proficiency
- Disabilities
- Health Insurance
- Employment Status
- Usual Place of Health Care
- Financial Security / Social Needs
- Self Reported Health
- Health Conditions (Associated Medications/Txs)
- NIMHD Framework
- Health Disparity Outcomes

NIH Endorsed



CDE Repository: https://cde.nlm.nih.gov/home



Cross-walked with PhenX SDoH

ScHARe

COMMON DATA ELEMENTS

NLM CDE Repository

Coded NIMHDCommon Data Elements

- Labels
- Questions
- Permissible Values

DATA UPLOAD

Acquired **Google and ScHARe Hosted Datasets**

- Overview
- Data Dictionaries
- Data Updates

REPOSITORY

Project and Key Acquired Datasets

Overview

Description and Links to Overview Material

4-Privacy Levels

COMMON DATA ELEMENTS

Data

Metadata

Data Dictionaries

Analysis Ready

RAS Single Sign-on

ATO

Common Data Elements + **Data**

Data Access

Based On PII Levels and User Needs:

- Public
- Data Use Agreement
- Private

DATA MAPPING, DOWNLOAD AND EXPORT

Other Cloud Platforms
AnVil, BDC, All of Us

DATA MAPPING

ACROSS DATASETS AND PLATFORMS BASED ON CDES

EXAMPLE: CDE linked

ACS NIMHD Project BioData Catalyst

Aggregated Data Set

CDE Linked Project Data

Data Download in a Variety of Formats

CSV, TSV, XLSX

Data Export to Terra for Analysis

Workspaces

Visualizations Tools

Shiny

Project & federated dataset mapping

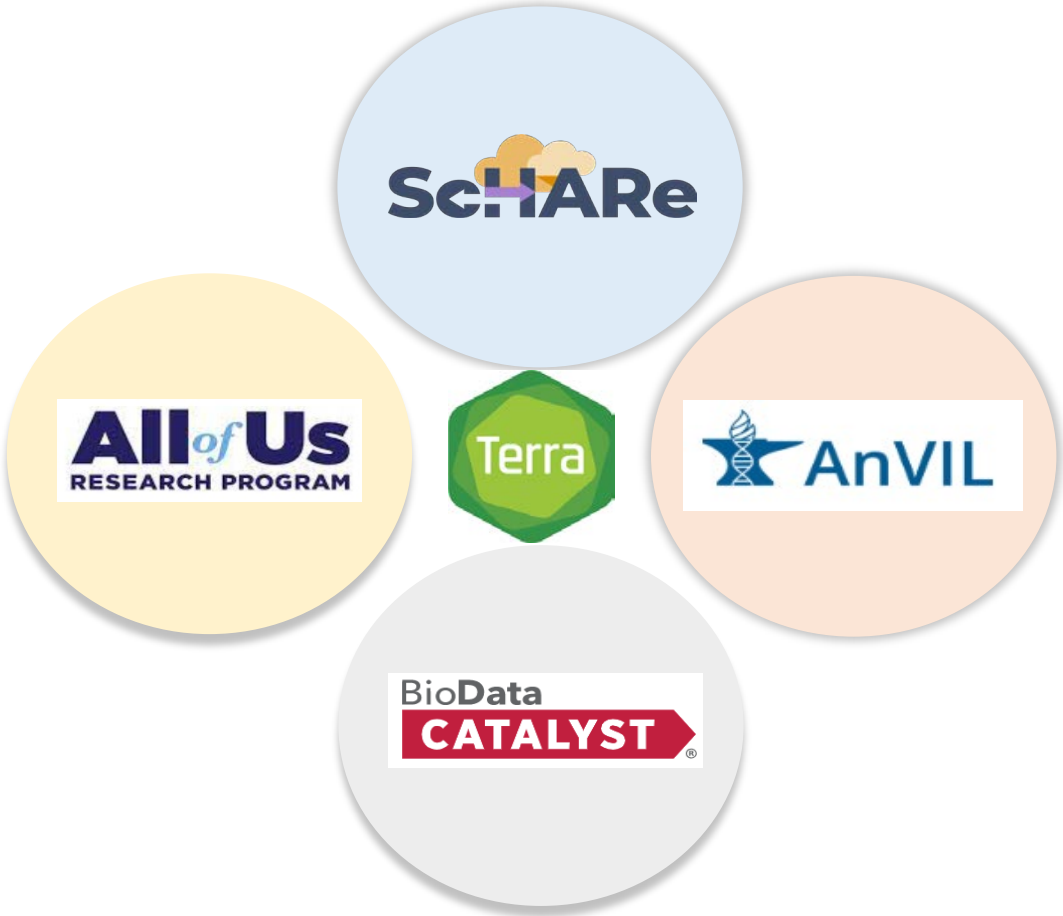
- Project Title
- Project Description
- Core Common Data Elements
- Other Project Data
- Data Dictionary

\+ AMERICAN COMMUNITY SURVEY

\+ Medical Expenditure Survey (MEPS)

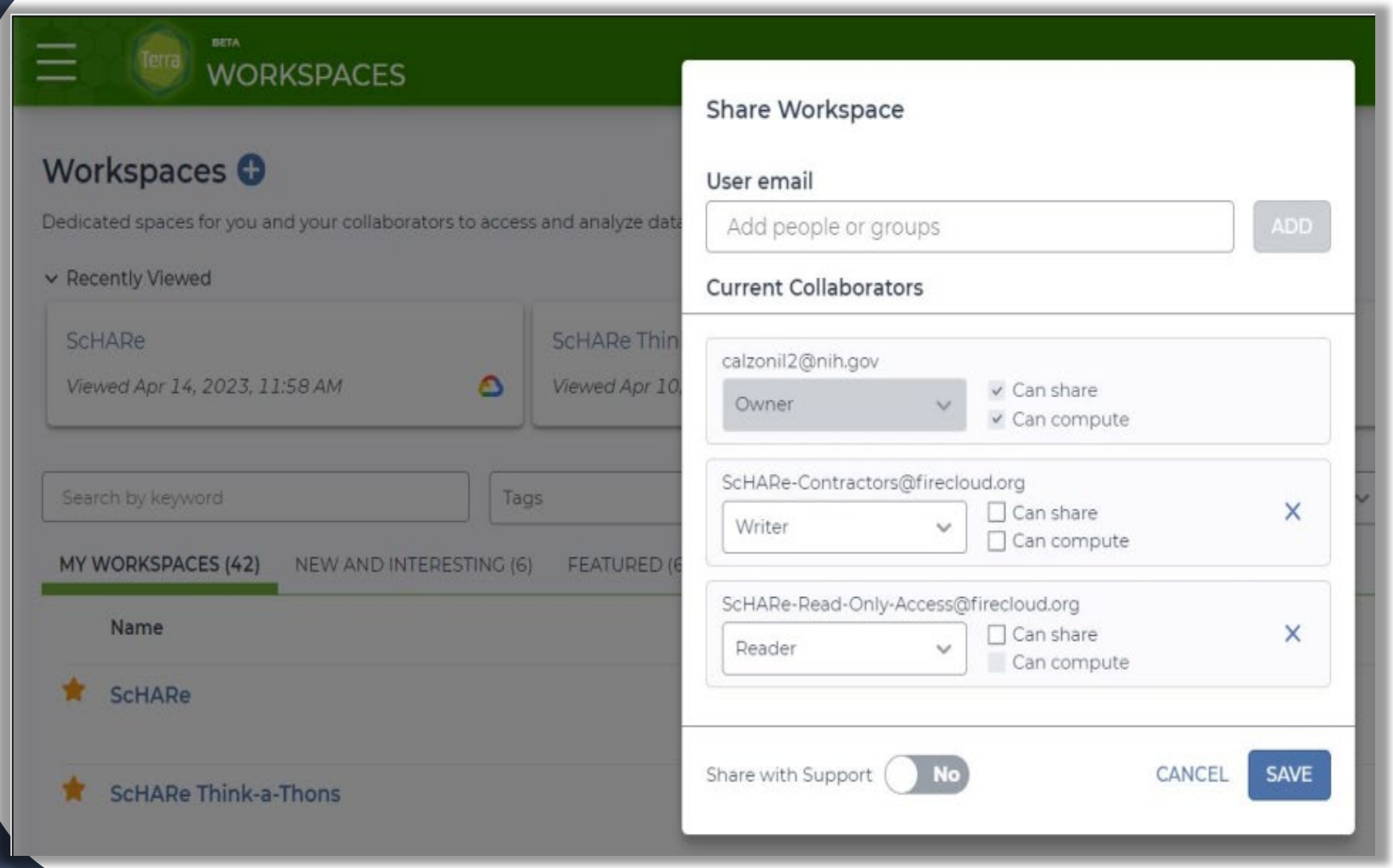
\+ Pharmacy and health insurance databases

Mapping across cloud platforms



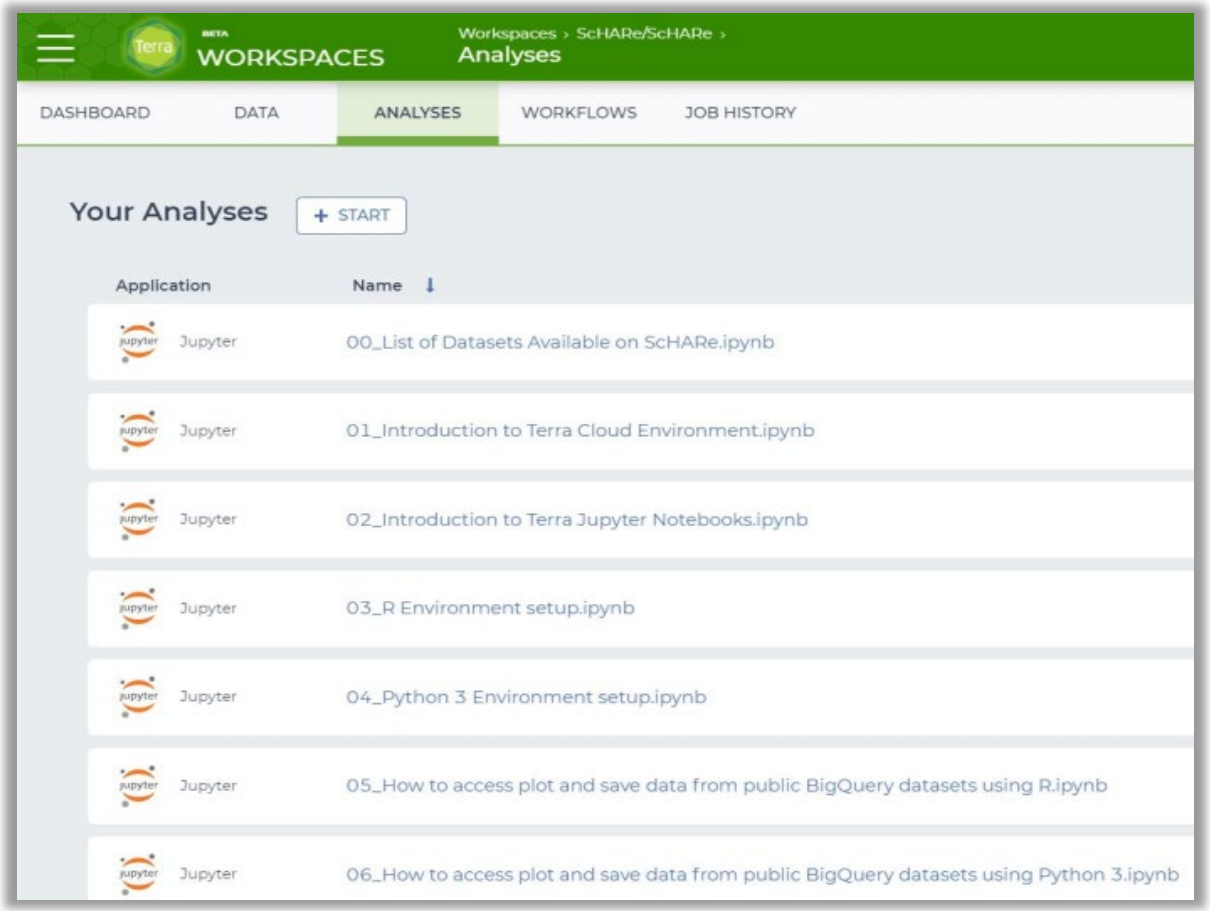
UPCOMING

Secure workspace



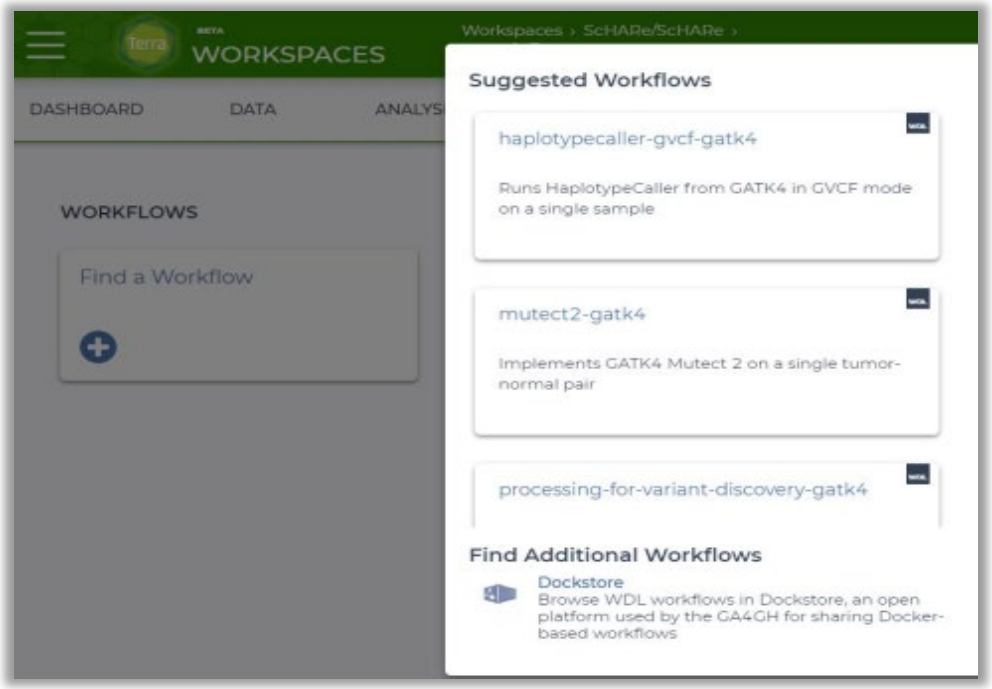
- Secure workspace **for self or collaborative research**
- **Assign roles**: review or admin
- **Host own data and code**

Notebooks analytics



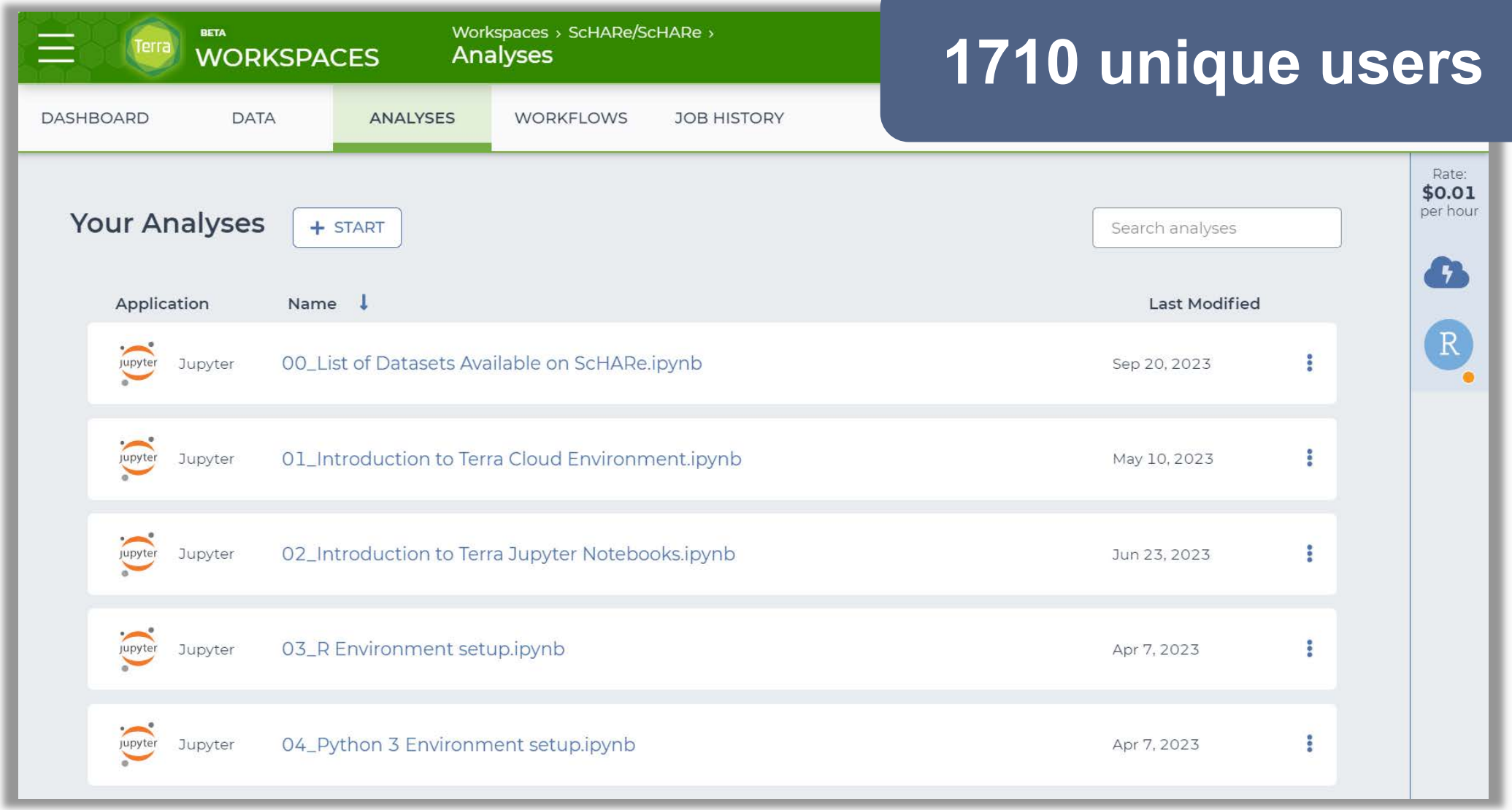
Workflows - Modular codes

- **Cut and paste analytics**



- Modular codes developed for reuse
- **Adding SAS**

ScHARe Registrations



1710 unique users

ScHARe

Think-a-Thons

National Institutes of Health

Think-a-Thon (TaT) Topics

February	**Artificial Intelligence and Cloud Computing 101**
March	**ScHARe 1 – Accounts and Workspaces**
April	**ScHARe 2 – Terra Datasets**
May	**ScHARe 3 – Terra Google-hosted Datasets**
	ScHARe for Educators (Community Colleges & Low Resource MSIs)
June	**ScHARe 4 – Terra ScHARe-hosted Datasets**
July	**An Introduction to Python for Data Science – Part 1**
August	**An Introduction to Python for Data Science – Part 2**
	ScHARe for American Indian / Alaska Native Researchers
September	**ScHARe 5: A Review of the ScHARe Platform and Data Ecosystem**
October	**Preparing for AI 1: Common Data Elements and Data Aggregation**
November	**Preparing for AI 2: An Introduction to FAIR Data and AI-ready Datasets**
January	**Preparing for AI 3: Computational Data Science Strategies 101**

Think-a-Thon (TaT) Topics

February	**Artificial Intelligence and Cloud Computing 101**
March	**ScHARe 1 – Accounts and Workspaces**
April	**ScHARe 2 – Terra Datasets**
May	**ScHARe 3 – Terra Google-hosted Datasets**
	ScHARe for Educators (Community Colleges & Low Resource MSIs)
June	**ScHARe 4 – Terra ScHARe-hosted Datasets**
July	**An Introduction to Python for Data Science – Part 1**
August	**An Introduction to Python for Data Science – Part 2**
	ScHARe for American Indian / Alaska Native Researchers
September	**ScHARe 5: A Review of the ScHARe Platform and Data Ecosystem**
October	**Preparing for AI 1: Common Data Elements and Data Aggregation**
November	**Preparing for AI 2: An Introduction to FAIR Data and AI-ready Datasets**
January	**Preparing for AI 3: Computational Data Science Strategies 101**

Upcoming

ScHARe

Think-a-Thons (TaT)

Research Teams

Title: Data Science Projects 1 – Health Disparities and Individual SDoH **Description:** Exploring the impact of individual Social Determinants of Health on health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.
Title: Data Science Projects 2 - Health Disparities and Structural SDoH **Description:** Assessing the impact of structural Social Determinants of Health on health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.
Title: Data Science Projects 3 – Health Outcomes **Description:** Investigating the influence of non-clinical factors on disparities in health care delivery: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.

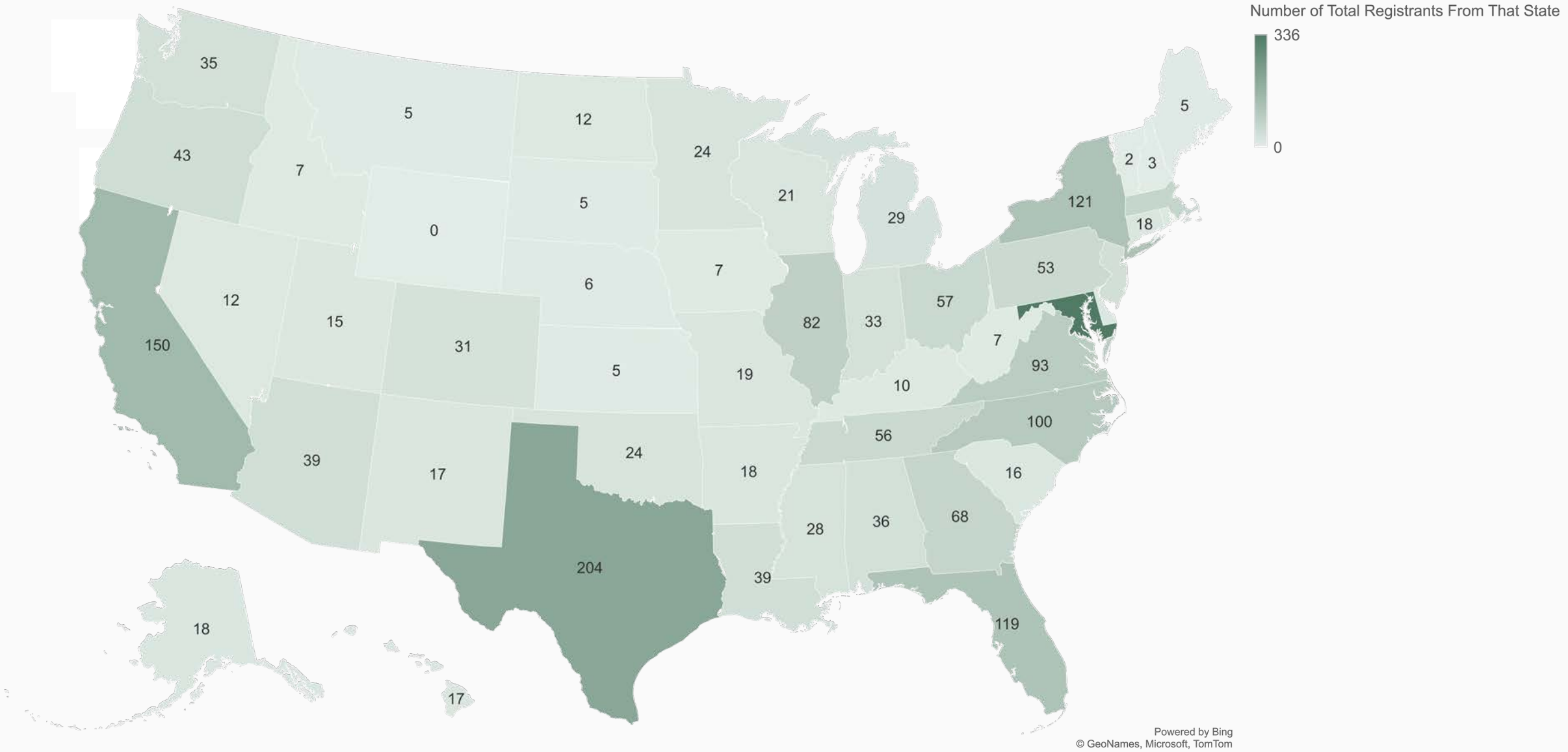
- Foster a research paradigm shift to use Big Data
- Promote use of Dark Data

- Multi-career (students to sr. investigators)
- Multi-discipline (data scientist & researchers)
- Feature Datasets with Guest Expert Leads
- Secure experts in topic area, analytics, data sources etc. to provide guidance
- Generate research idea - decide potential design, datasets & analytics
- Select co-leads to coordinate completion outside of TaT
- Publications

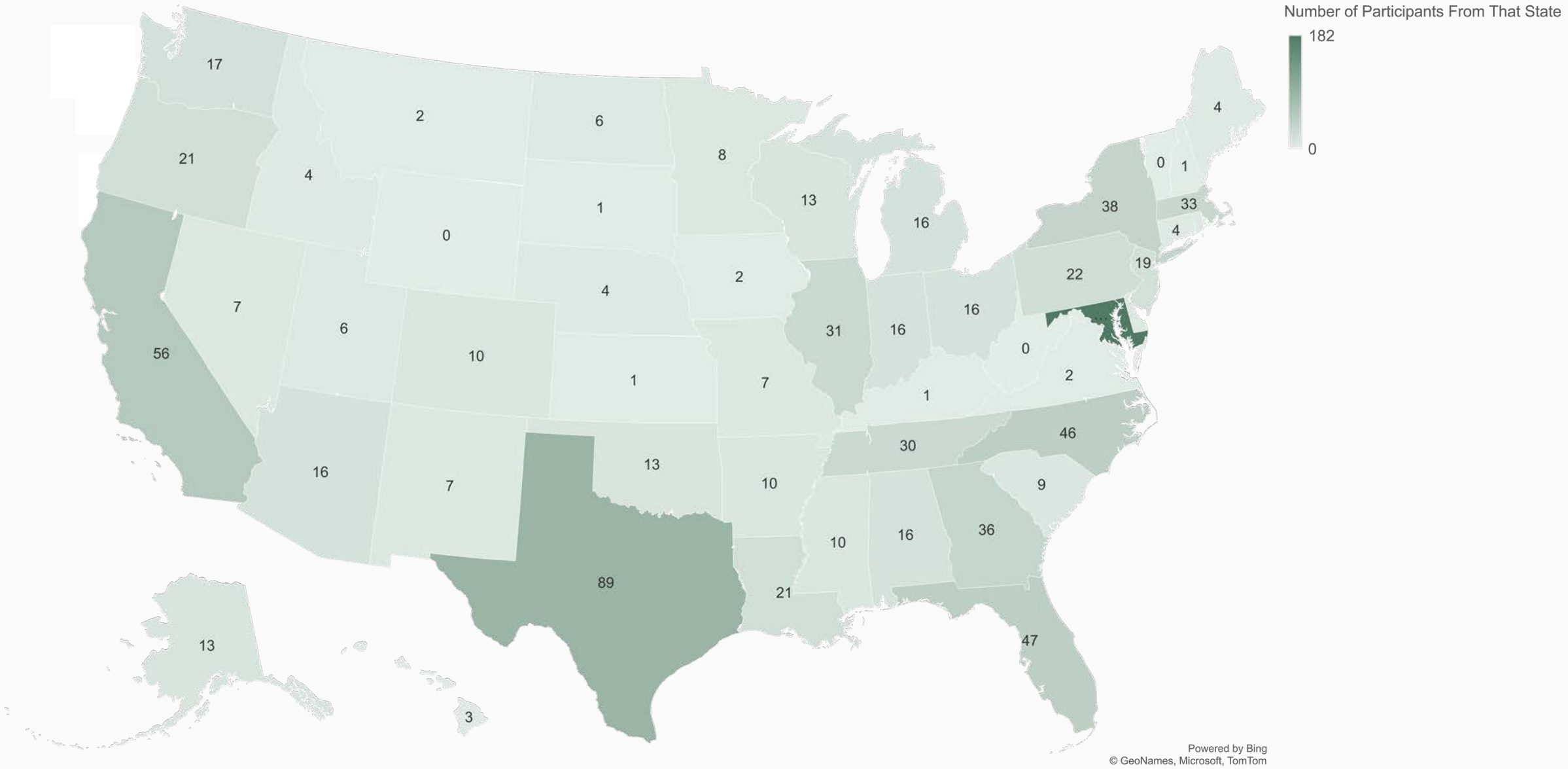
Register:

bit.ly/think-a-thons

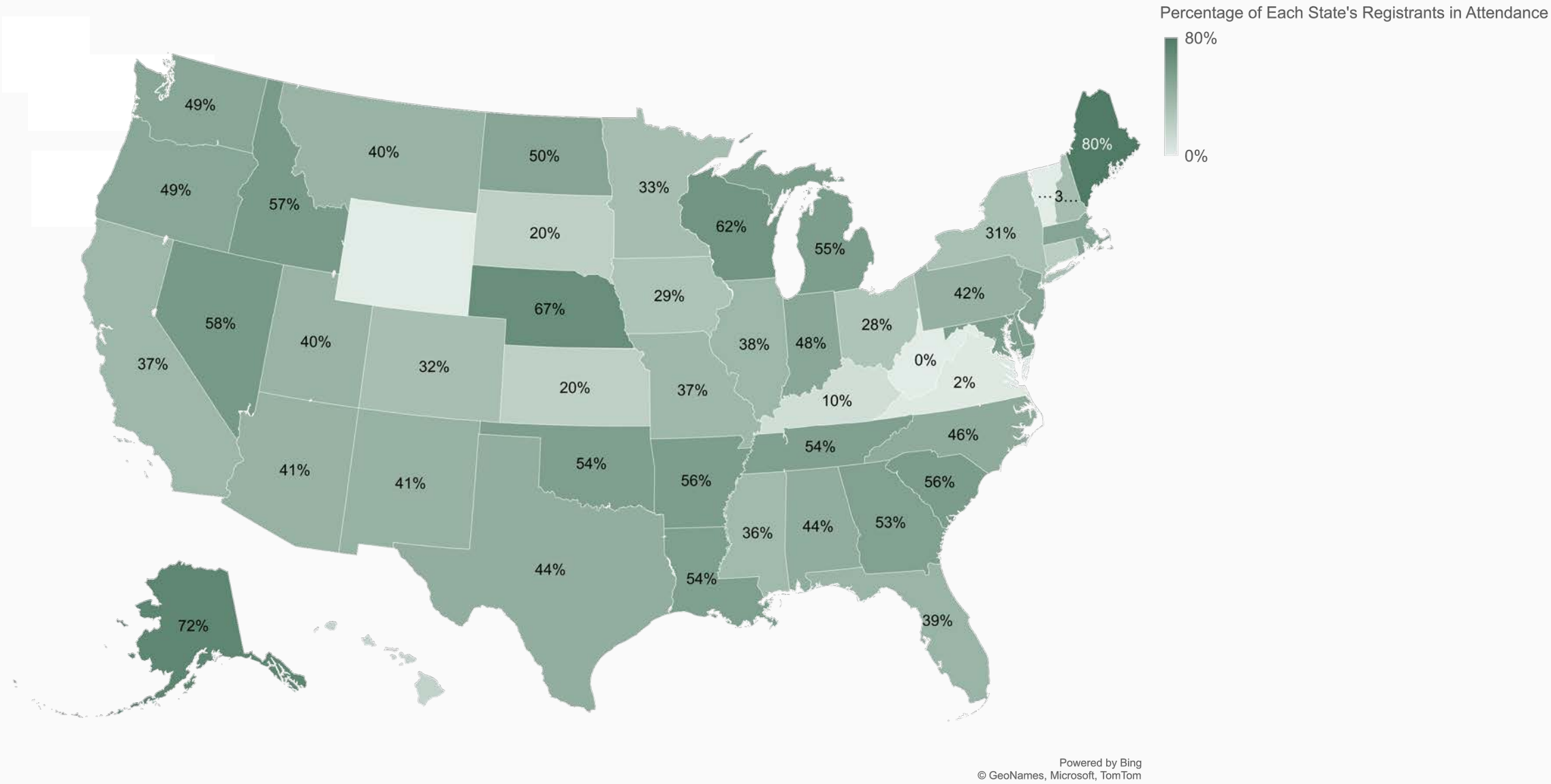
Number of Registrants by State



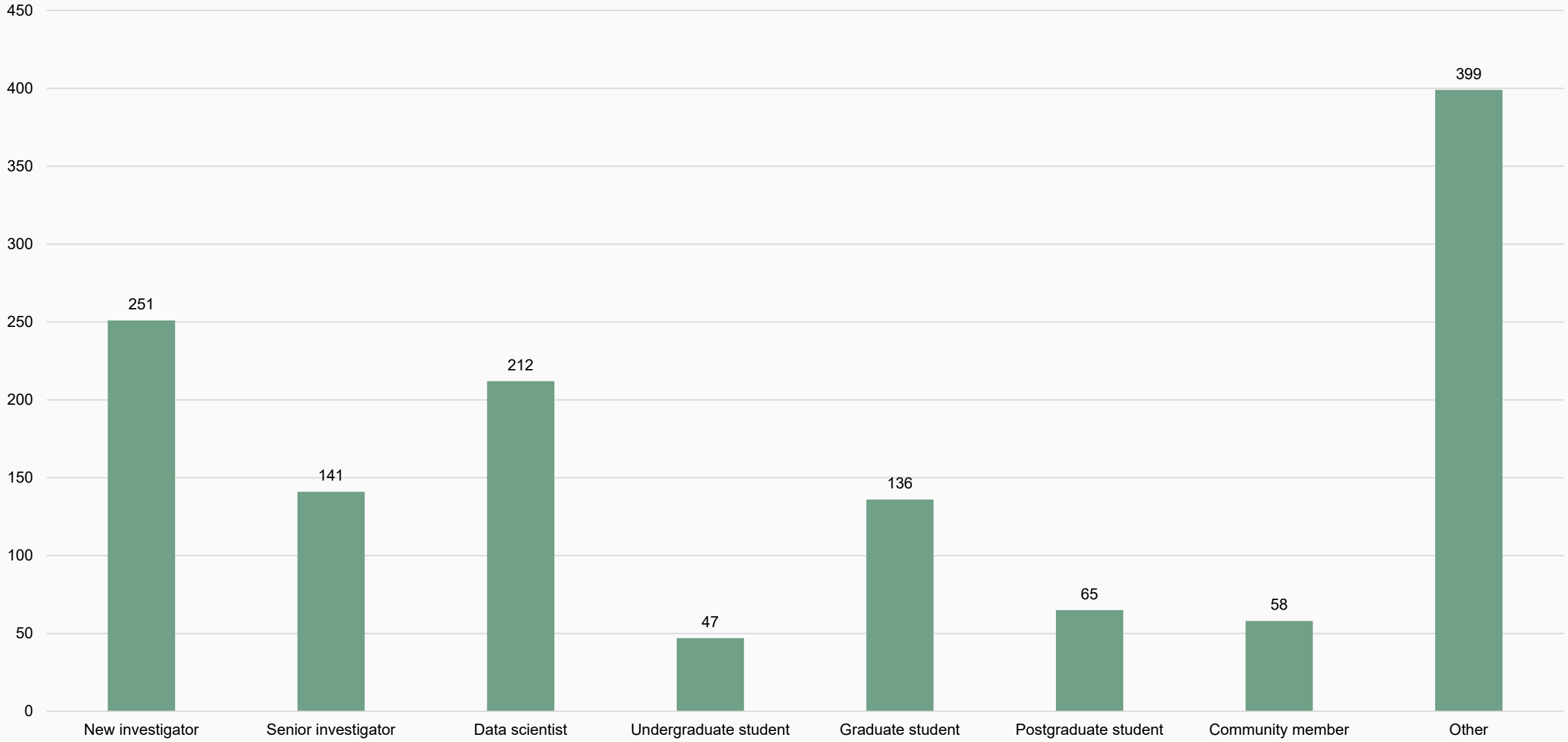
Number of Participants by State



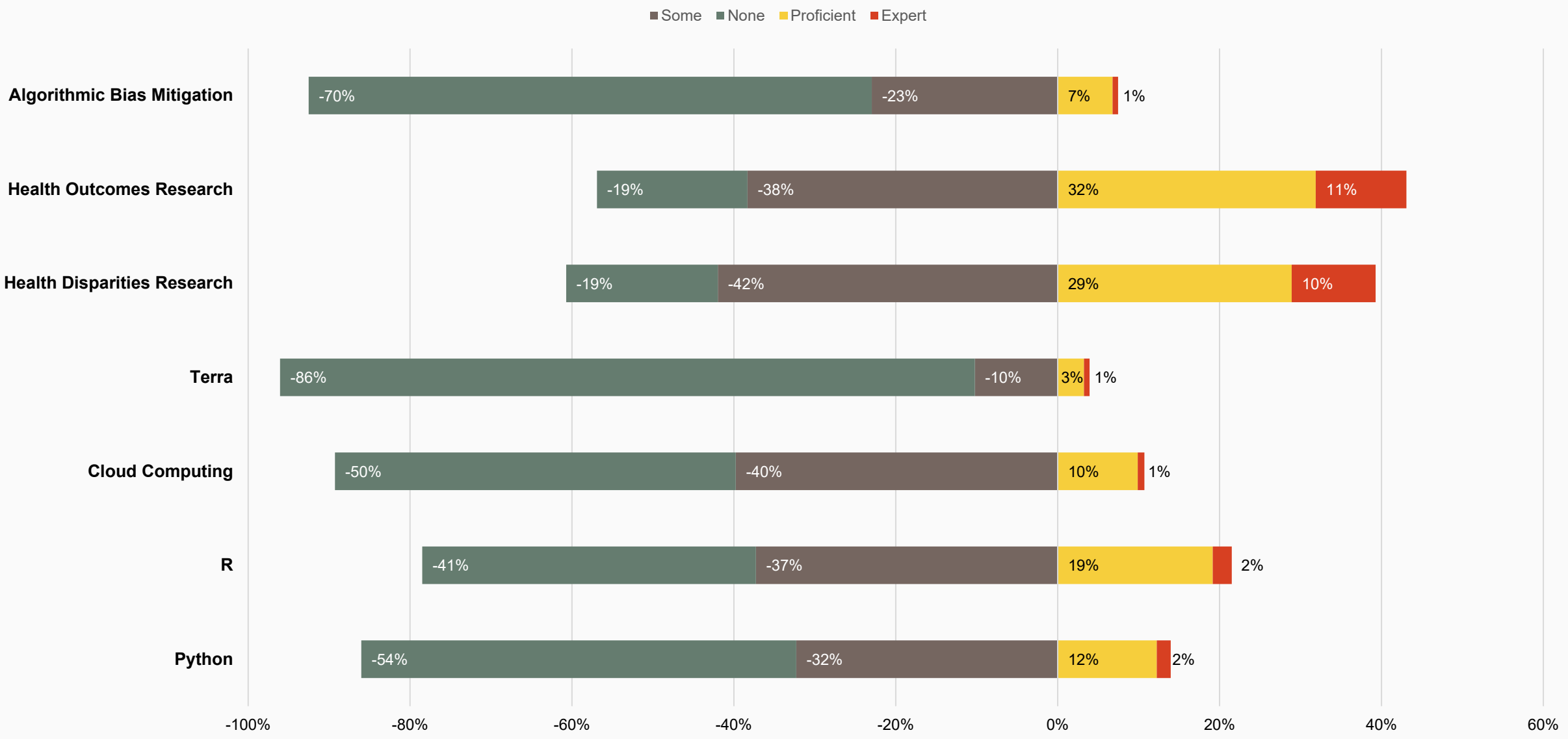
Percentage of Each State's Registrants in Attendance



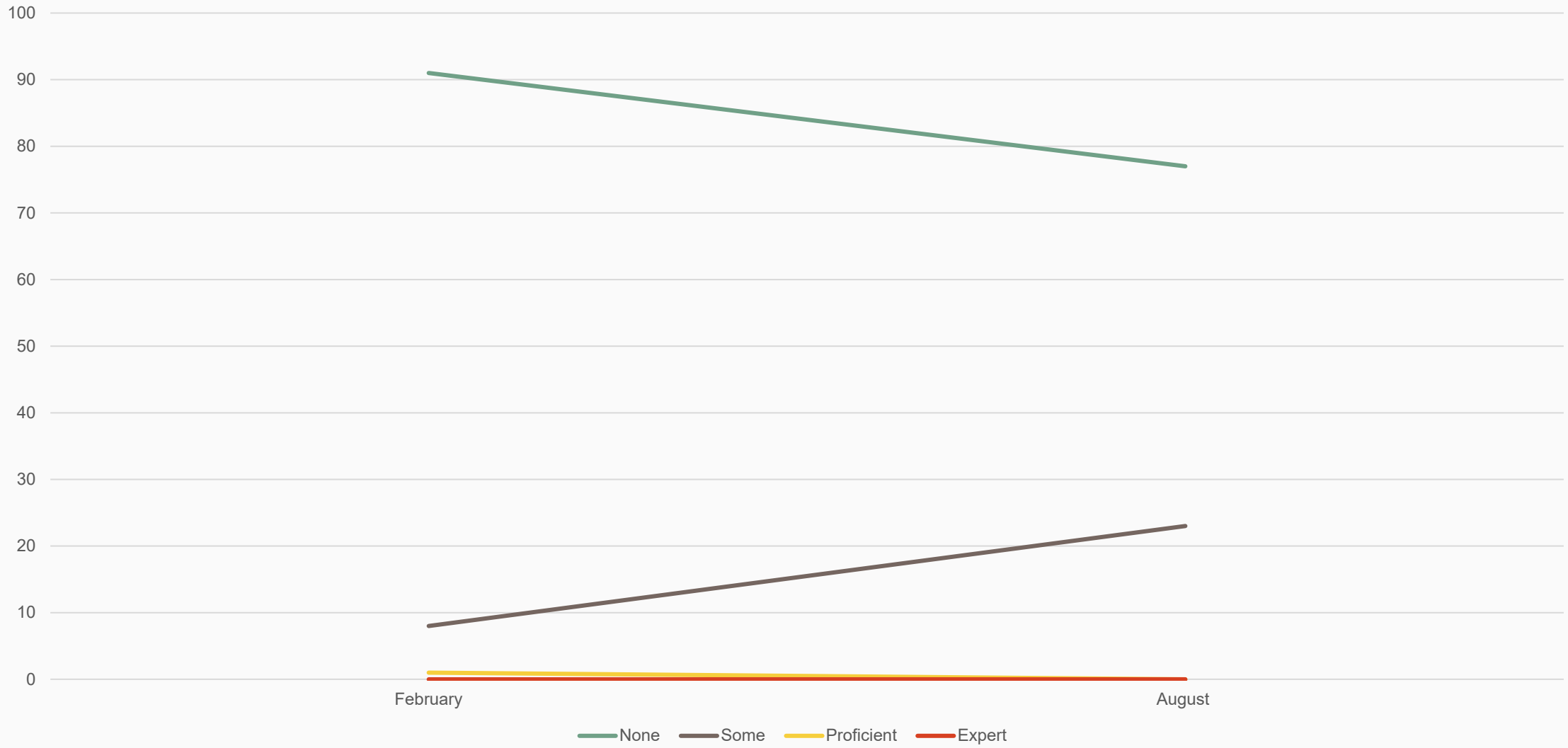
Role of Participants



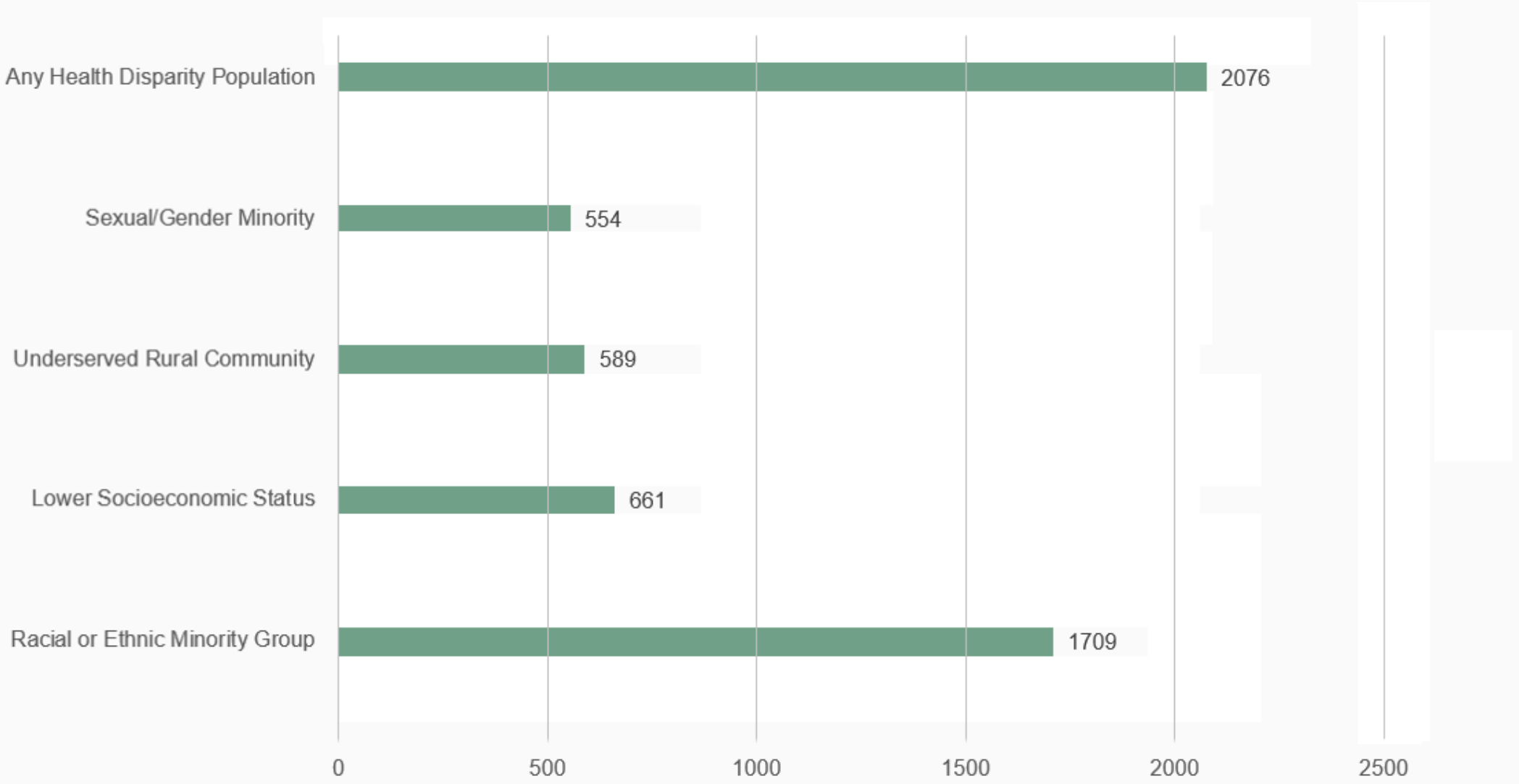
Level of Experience (participants only)



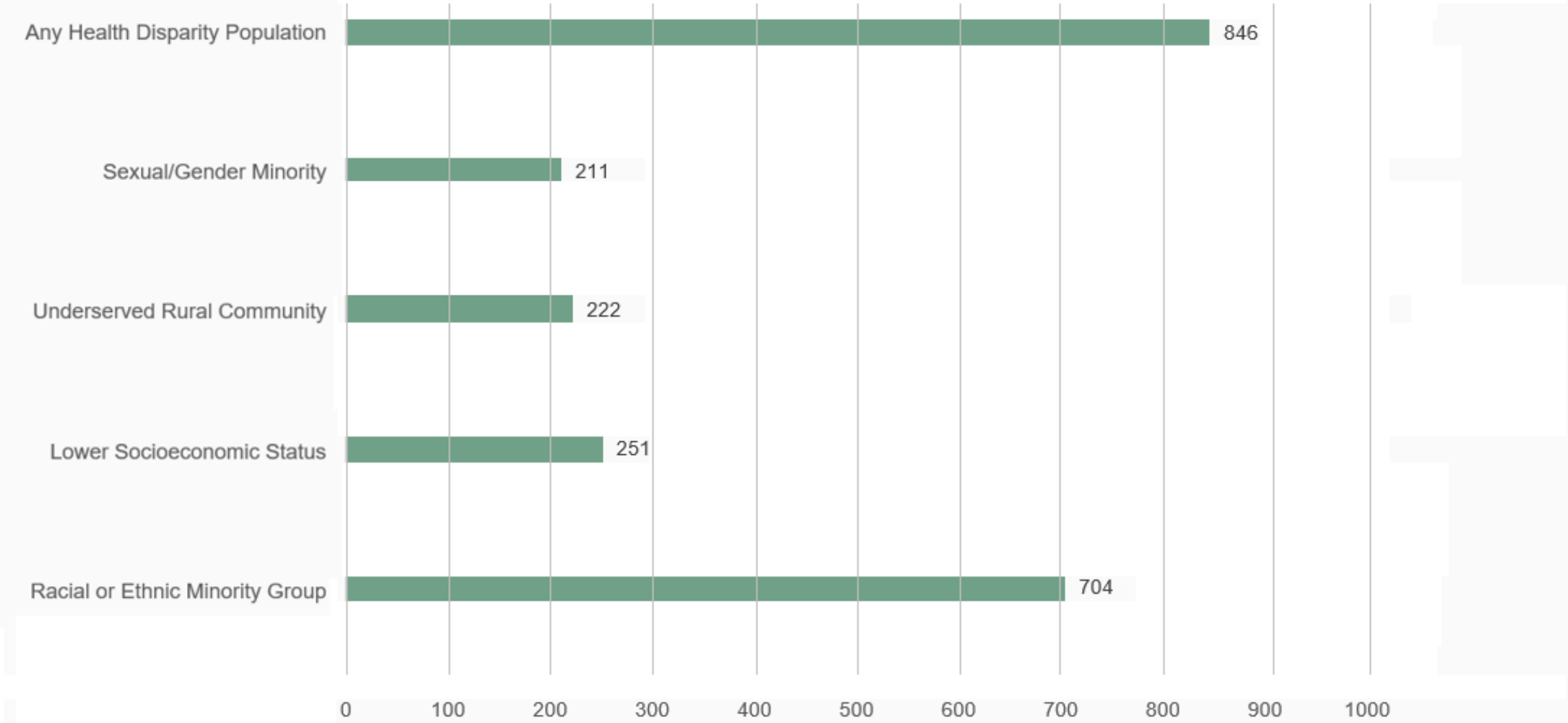
Level of Experience with the Terra Platform Over Time



Health Disparity Population Representation – All Registrants



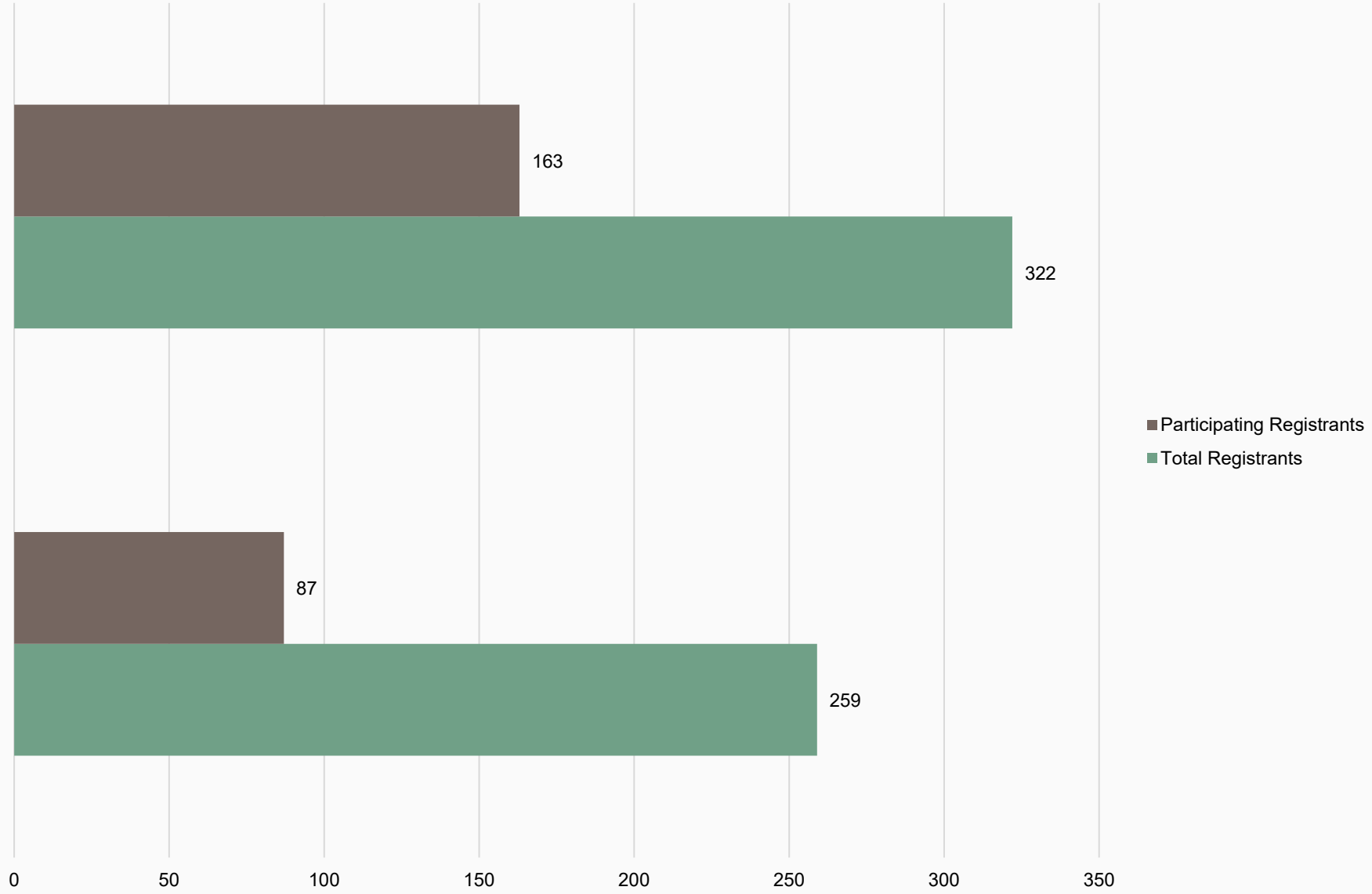
Health Disparity Population Representation - Participants Only



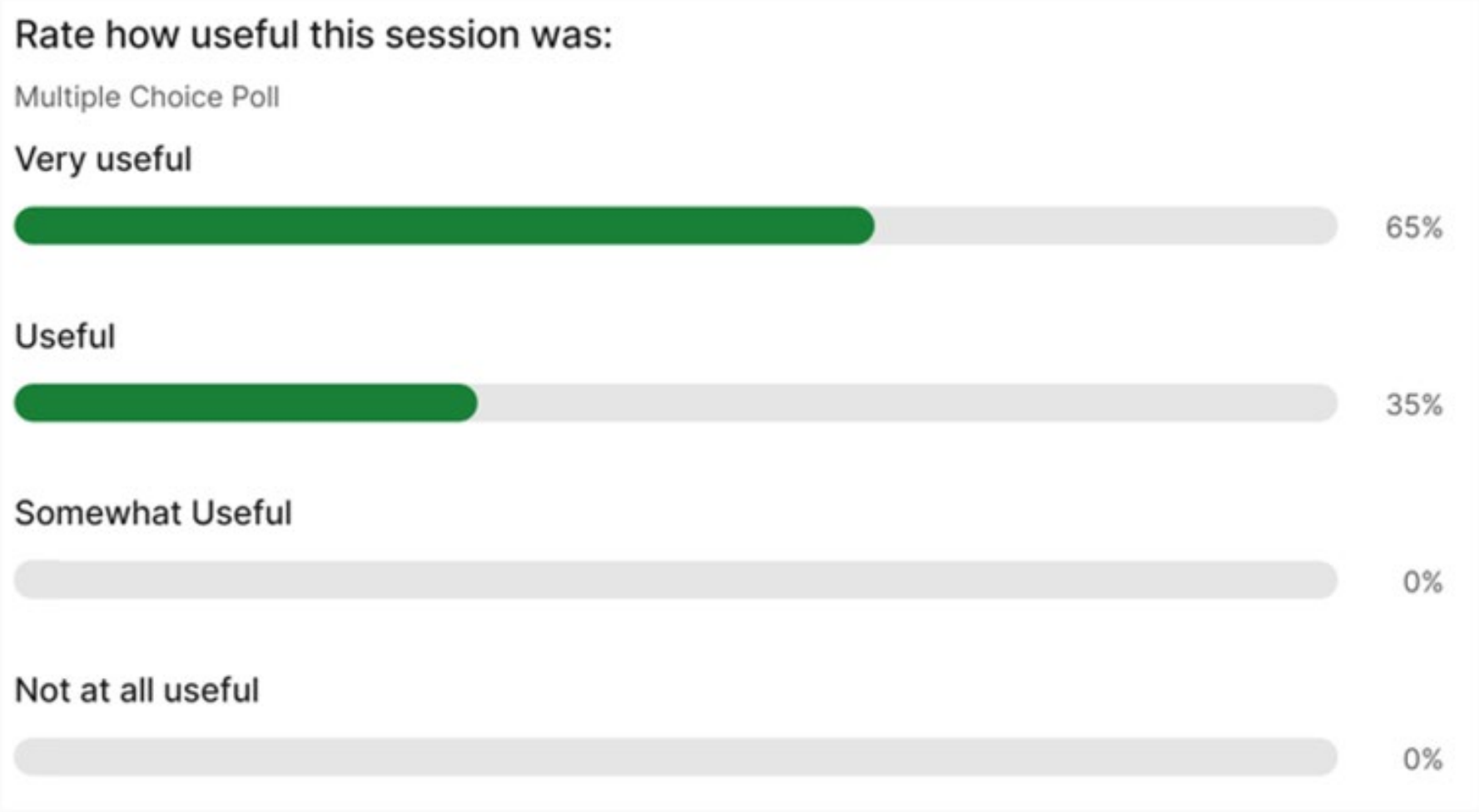
Special Think-a-Thons Registration and Participation

Special Think-a-Thon
ScHARe for AI/AN Researchers

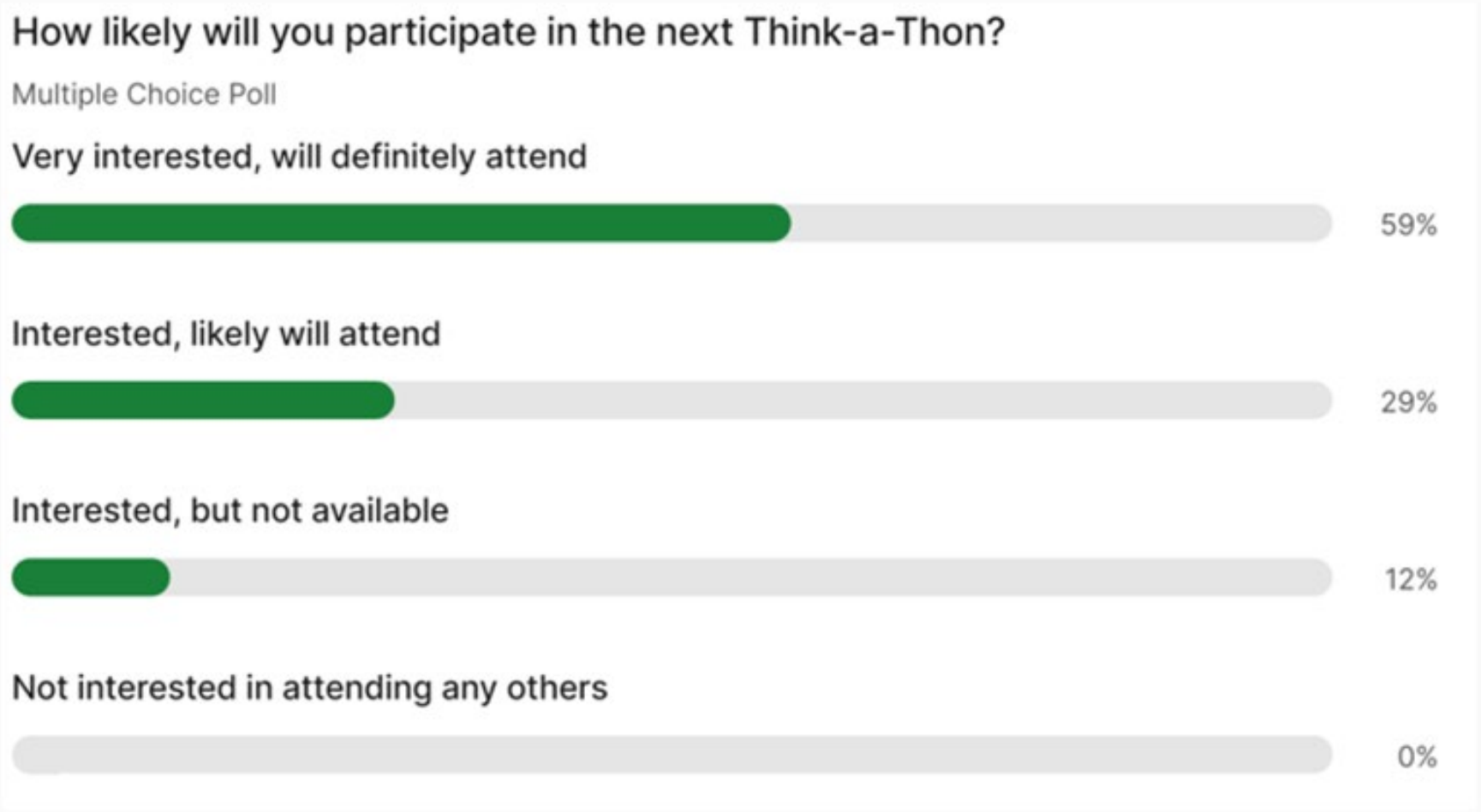
Special Think-a-Thon
ScHARe for Educators



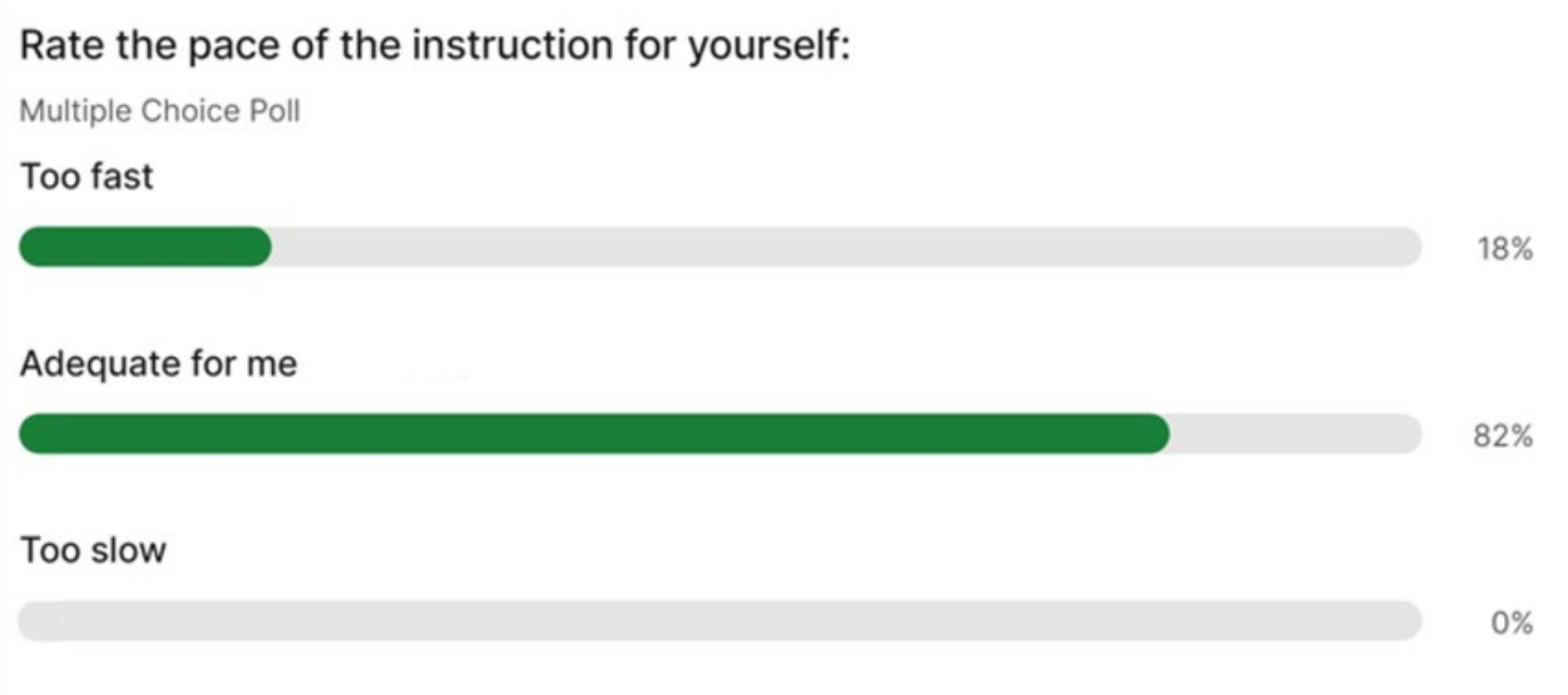
Evaluations (September Think-a-Thon)



Evaluations (September Think-a-Thon)



Evaluations (September Think-a-Thon)



Feedback Received

TJ

It feels so good to see how far we've come from onboarding and getting familiar with using the platform to here 🤗

To: schare <schare@mail.nih.gov>
Subject: [EXTERNAL] Fwd: Welcome to Terra

Today's webinar has been AMAZING.

Thank you for everything,

To: schare <schare@mail.nih.gov>
Subject: [EXTERNAL] ScHARe login

Good afternoon,

Thank you for the great intro to the ScHARe platform yesterday!

Thanks again, and I'll be looking forward to the demo next Wednesday!

Think-a-Thon Video Views

Think-a-Thon	Video Views as of 9/27/2023
February 2023	303
March 2023	171
April 2023	134
May 2023	82
June 2023	95
July 2023	365
August 2023	69

Partnerships and Collaborative Discussions

Partners

NINR - National Institute of Nursing Research

ORWH - Office of Research on Women's Health

OMH – Office of Minority Health

Collaborative Discussions

NIDCR - National Institute of Dental and Craniofacial Research

NIEHS - National Institute of Environmental Health Sciences

NHLBI - National Heart, Lung, and Blood Institute (BioData Catalyst)

NHGRI - National Human Genome Research Institute (AnVIL and NCPI Initiative)

All of Us

NLM – Collaborates with TaTs and CDE RepositoryNIH CHORD (Climate and Health Outcomes Research Data Systems)

EPA – Environmental Protection Agency

FAS - Federation of American Scientists and University of Maryland (Advancing equity in medical devices - Pulse oximeters)

NAIRR Pilot Project

Health Equity Action Network (HEAN) AI-Ready Data Project for ScHARe

The HEAN data warehouse will be hosted on ScHARe, with the ultimate goal of facilitating its integration into the open NAIRR infrastructure

The HEAN data warehouse:

- *contains data on multiple chronic diseases associated with health disparities*
- *will be a test case for AI-ready data that includes race and ethnicity to:*
  - *test mapping with ScHARe Core Common Data Elements*
  - *test bias mitigation strategies*
  - *use AI to advance health disparities research in the areas of chronic disease and health equity, especially as pertaining to SDoH*

NIMHD Listserve – Outreach (6 months)

Total of 30,038 subscribers to ScHARe training, tools, and resources list.

- Training, Tools & Resources list – 29,839
- Research Collaboration list – 4,723
- Datasets list – 4,676
- AI Bias Mitigation list – 4,666