****

# **Instructor Lesson plan**

## **Day 4: Non-Probability-Based Surveys in Practice**

## **Updated by Naniette H. Coleman; Prepared by Lai Wei and Matt Salganik**

### **Instructor preparation**

To prepare for this activity you should:

* Carefully read the [lesson plan for participants](https://docs.google.com/document/d/15rvHiu1r0hhruYoBtrOD36kNiIZ0Ev_1G3YWsWYM06o/edit?usp=sharing) and the reading assignments included in it.
* This activity has three main elements: 1) creating the questionnaire with at least one module on an anti-Black racism or inequity topic; 2) collecting some data; and 3) analyzing a larger dataset that we have pre-collected for you. From the instructor perspective, step 2 can be the most complicated. SICSS-Howard-Mathematica, will have participants deploy their survey to MTurk, and we will provide them with funds (say $50 per account). If you don't have funds to support this you are welcome to skip the data collection step and just focus on steps (1) and (3). Also, if you have funds and you want to use an alternative platform, such as [Prolific](https://www.prolific.co/), that is OK; this activity is not specific to MTurk.
* If you want to provide your participants pre-loaded accounts, both MTurk and Prolific have tightened up verification and account security. Therefore, we recommend that you create these accounts a few days before the activity and we expect that your participants may run into problems accessing them.
* This activity will involve each group doing some programming. You could explore setting up some kind of collaborative coding environment, but don't have plans to do that at SICSS-Howard/Mathematica.

### **Group formation**

Form groups of about 3-4 people. We think small groups are better for remote collaboration (when doing this activity in person we often have groups of 4-5). If possible, each group should have one person with experience with MTurk and one person with experience in surveys.

If you are going to fund data collection for your participants, please note that the more groups you have, the higher the costs of data collection.

### **Key Tasks for instructors and TA if you do use MTurk for data collection**

* The organizer should double-check that groups have paid Turk workers by 2PM. Sometimes there is a question about whether the worker should be paid or not. When in doubt, we encourage you to just pay the worker.

### **Potential Discussion Questions**

When you come back for the group discussion, here are some possible questions:

* How close were the estimates for this activity to the estimates from Pew? Were there any important differences observed in the estimates of minoritized populations? What do you think might be causing some of these differences? How could you test thi s speculation?
* What issues have you noticed while you’re implementing the survey in Google Form that might affect how respondents answered the questions?
* Did you randomize the order of your questions? If not, should you have? Note that Pew did randomize the order in each survey.
* How important is it to open-source data that you have collected with other researchers?
* What are things we should keep in mind when open-sourcing data?
* Does this activity change how you think about your own research or the project you might want to do in week 2?

### **Notes**

* This activity is modular. You could easily do the afternoon part without doing the morning part.
* If you are collecting data when many people in the United States might be asleep (or not on MTurk), the data collection might be slower. You can relax the location requirement if you wish. The data we collected was from people in the United States, and your participants can do their analysis on that pre-collected dataset.
* Due to constraint in time and change towards remote collaboration, the survey activity will not ask participants to archive and release their data. You can review those materials from previous years [here](https://github.com/compsocialscience/summer-institute/blob/master/2019/materials/day4-surveys/06-intro-to-open-sourcing-data.pdf). However, this is a good change to remind participants that they should think about open and reproducible research and they will be expected to use those practices for the projects they start in week 2.
* Please help participants have realistic expectations of what can be completed during this activity. It is likely that participants do not have time to complete model-based post-stratification.

### **More information about Pew Research Center methodology**

The survey on policy priorities was collected using an online panel called the American Trends Panel (ATP), which was created by Pew. Here's more about the [methodology for the survey on policy priorities](https://www.pewresearch.org/politics/2021/01/28/policy-priorities-methodology/)

The survey on social media use was done using a telephone sample (landline and cell phones). Here's more information about the [methodology for the survey about social media use](https://www.pewresearch.org/internet/2021/04/07/social-media-use-methodology/)

### **Acknowledgments**

The updated version of this activity was conceptualized and created at the SICSS-Howard/Mathematica site in 2022 with the goal of our participants and all SICSS participants having the option to do exercises that allow them to directly engage the topics of anti-Black racism and inequity alongside learning about Computational Social Science. This exercise would not have been possible without support from Chris Bail, Matt Salganik, Akira Bell, Jesse Chandler, Joshua Dyck, and Amber Du, Alina Gu, Amanda Lee, Ava Wu, and the entire Coleman Research Lab at the University of California Berkeley. The original activity was designed with input from participants and TAs from SICSS 2017-2020, especially Yo-Yo Chen, Janet Xu, Simone Zhang, Robin Lee, and Ian Lundberg.