Gordon Kamer and Max Snyder

May 15, 2019

Math 23c

Paul Bamberg

A one-page document that discusses ethical issues related to collection of the dataset.

For our project, we worked with data collected by the American National Election Survey, which conducts huge surveys before and after federal elections. We used their 2016 data.

The first issue that arises is the anonymity of such data. While respondents’ names are not distributed as a part of the dataset, the number of questions asked makes anonymity almost impossible. I can tell what kind of area a respondent lives in, what TV shows the respondent watches, how old the respondent is, how long the respondent has lived at his/her address, and more. If I had an idea about someone who took the survey, I would surely be able to determine which ID corresponded with that person and find all of that person’s political beliefs and that person’s votes for every single political office for the past 2 election cycles. Some of the answers are personal, too, such as answers to religious and moral questions. Because the data are effectively not anonymous – even though the survey is purported to be anonymous – I can find private information on the people in the dataset.

Second, there is an ethical impetus on the part of the group collecting the data to not influence or mess with the data in any way. If sample-takers were to deliberately not take a representative sample in order to bias the data one way or the other, that would be a serious ethical violation, as would be nudging respondents in certain ways when they answer questions.