Increasing Understanding of Survey Re-Weighting with Visualization

Yufei Zhang*
University of North Carolina at Chapel Hill

David Borland†
University of North Carolina at Chapel Hill

David Gotz‡
University of North Carolina at Chapel Hill

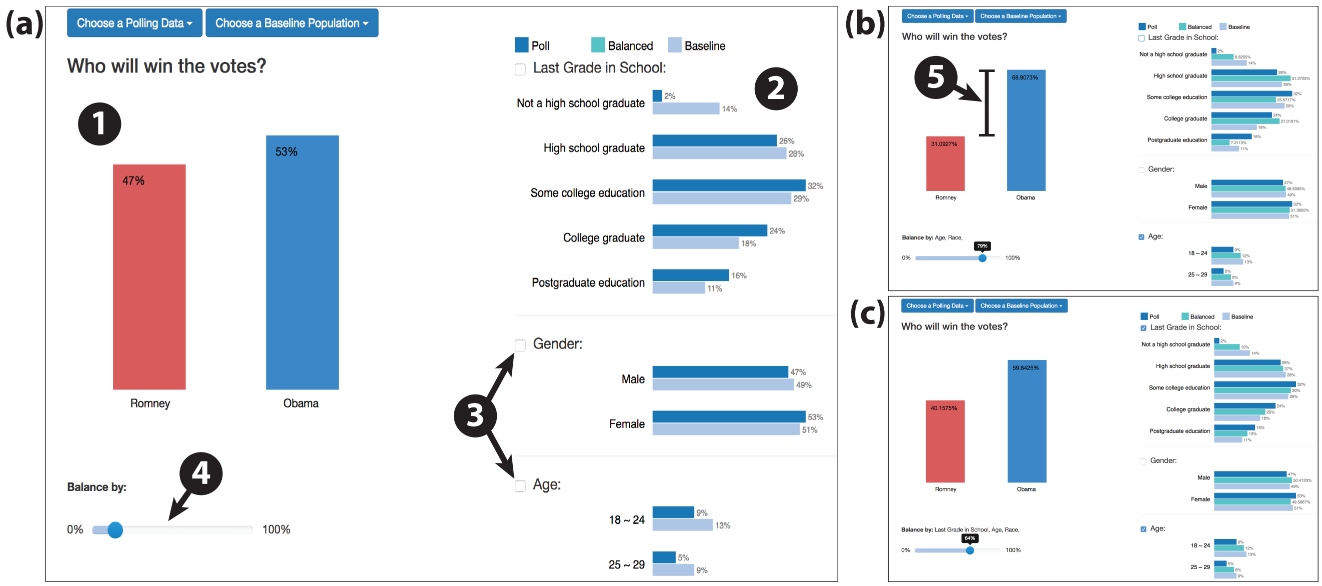

Figure 1: (a) Before balancing the visualization shows the raw survey results (1) and histograms (2) comparing the demographics of the survey sample population to the intended baseline population for the survey. Users can click checkboxes (3) for each demographic to re-weight the survey sample to be more representative for the corresponding dimension, and adjust a slider (4) for the amount of correction toward the baseline to be performed. (b) After re-weighting by age and race, the lead for Obama is larger. (c) However, after including education level the margin tightens. Study results indicate that this interactive approach helps communicate the re-weighting process.

Abstract

Surveys are a widely used tool for inferring information about a target population by collecting data from a smaller sample. However information is always lost, as the sample cannot in general fully represent the target population. Thus predictions made from surveys may contain biases. To mitigate these biases, sample subgroups can be re-weighted to match their known distributions in the target population. We introduce a web-based interactive tool to visualize the re-weighting process in surveys, with specific application to presidential election polls. A detailed description of the system's user interface and re-weighting algorithm are provided. In addition, the results of a twenty-person user study evaluating the system are presented and discussed.

Keywords: Visualization, Surveys, Polling, Selection Bias

1 Introduction

Survey polls are widely used by media outlets and political organizations to analyze public opinion and make predictions in the run-up to elections. These polls are particularly common in the USA for presidential campaigns. Nationwide telephone polls are often used for collecting data for such polls, usually conducted monthly by collecting demographic information, and opinions towards presidential candidates and policies from a sample of the nationwide population. We use the term *poll population* to describe this sample population. Polls often track the same group of people over time in order to control for unexpected variables.

Information obtained from polls are analysed to discover trends in the campaign and to predict the likelihood of each candidate winning the vote. However, biases are unavoidable when the survey sample population is not the same as the target population. As with other surveys, the poll population needs to be adjusted to simulate the larger population of interest (i.e., the *baseline population*), by applying different weights to subgroups within the sample. Subgroups are usually divided by demographic information such as age, gender, race, region, education level, income, party, and political views [1]. We introduce a visualization system designed to help users understand this re-weighting process.

2 Design

Our web-based *Balance* application displays the bias introduced in political polls, and elucidates the process of eliminating bias.

Visual Design. An interactive interface was designed to visualize the re-weighting process. Figure 1a shows four panels of the interface before balancing. Figures 1b and 1c show the interface after balancing.

Drop-down lists in the upper left are used to choose polling data and baseline data from available datasets. The result panel shows the predicted share of votes for each candidate from the selected data. Initially, predictions are raw results from the poll, without any re-weighting. Predictions may change according to different weighting methods.

* e-mail: yuffie@live.unc.edu † e-mail: borland@renci.org
‡ email: gotz@unc.edu

The bottom left of the interface contains a list of the current balancing dimensions, with a slider to control the balancing extent. The amount of balancing ranges from 0% (no balancing of sample) to 100% (balancing of sample to fully match the baseline distribution).

The right half of the interface contains a list of bar charts showing distributions for the available demographic information. Three types of percentage values are shown in three different colors to describe the distribution in (1) the poll population, (2) after re-weighting, and (3) in the baseline population. The percentage after re-weighting will range between the poll (0% balancing) and baseline (100% balancing) percentages. A check box is provided for each chart to add or remove this dimension from the list of dimensions to use for balancing.

Re-Weighting Algorithm. One common approach to correct for sampling bias with respect to the target population of interest, assuming that the subgroup distribution in the population is known or can be estimated, is to give different weight to subgroups in the sample such that subgroups that are under-represented in the sample obtain a higher weight, whereas subgroups that are over-represented are down-weighted [2]. After choosing a poll dataset and a baseline dataset, a user can choose a set of dimensions which the user wishes to use in re-weighting. We refer to these as totem dimensions, T.

For example, Jane may choose to re-weight race and gender because she thinks they're the most influential factors. The user can then decide to what degree they want to correct the polling data, by selecting the *re-weighting extent*, E. Jane may choose to fully re-weight the poll data to ensure her findings reflect the overall population. However, Jack may re-weight by 50% because he wants to maintain some of the features in the poll data. Once T and E are decided, a balancing algorithm is applied to the dataset to generate weights for each subgroup. Subgroups are defined by combinations of values in T. For example, given that Jane specified race and gender in T, six subgroups would be created based on the available data: (1) Female Asian, (2) Female Black, (3) Female White, (4) Male Asian, (5) Male Black, and (6) Male White. If any subgroup has zero members, it will be omitted from the algorithm because it is impossible to re-weight.

A weight (w_i) is then computed for each of the n subgroups as follows:

$$\boldsymbol{w_i = \frac{p_i + E(b_i - p_i)}{p_i}} \quad (1)$$

where p_i is the proportion of the subgroup i in the poll population, b_i is the proportion of the subgroup i in the baseline population, and E is the re-weighting extent.

After computing the weight of each subgroup, each respondent's vote will be re-weighted based on which subgroup the respondent belongs to. The new result is then derived by accumulating re-weighted votes. At this point users can have a sense of how the predicted share of votes changes for each candidate after any change in weights.

3 Study

A user study was conducted to evaluate how the *Balance* system helped users understand the impact of re-weighting.

Dataset and Participants. The study used two datasets: (1) the poll population, which included demographic information and voting preferences for each participant in the September 2012 Monmouth University Poll of nationwide voters in the 2012 presidential race between Mitt Romney and Barack Obama, and (2) the baseline population statistics for voting-age adults in the USA retrieved from Wikipedia.

Study participants comprised 20 graduate students at UNC-Chapel Hill recruited via email. None of the participants had prior experience with the system.

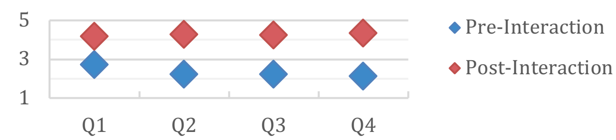

Figure 1: Figure 2: Mean responses in pre- and post-questionnaires showed benefits for interaction across all questions.

Protocol and Results. Participants interacted with the prototype by completing 15 tasks (7 practice, 8 experimental). Tasks were designed to test participants' understanding of the system and the impact of re-weighting. In addition, pre- and post-questionnaires were used to collect data from participants. In each questionnaire, participants were required to answer the same 4 questions, in which they reported how much they understood a specific aspect of re-weighting in surveys. A 5-point Likert scale was used for each question, where 1 represents "Not at all" and 5 represents "Very Well". The four questions probed the following concepts:

Q1. How well do you understand the overall re-weighting process in presidential surveys?

Q2. How well do you understand the sensitivity of the survey result to different biases in the poll population?

Q3. How well do you understand the impact of weighting extent on the survey results?

Q4. How well do you understand the impact of weighting overall?

A total of 160 experimental tasks were performed (8x20), of which 156 were completed correctly by the participants. This equates to a high task accuracy rate of 97.5%. Moreover, the questionnaires provided qualitative feedback with respect to the benefits of interacting with our system. Figure 2 shows the mean value of the answers to each question on the questionnaires. Comparing post-interaction (red) to pre-interaction (blue) shows a statistically significant improvement ($p < 0.0001$) for all questions. This suggests that participants felt that they understood bias and the re-weighting process better after interacting with the software. Together, these results suggest that *Balance* provides a highly usable interface that has a positive effect for individuals in terms of how they understand the impact of survey re-weighting.

4 Conclusion

We have reported on a visualization tool designed to visualize the bias introducing by survey sampling and the re-weighting process of eliminating bias. We introduced the *Balance* interface and functionalities, including the re-weighting algorithm, and provided the results of a user study evaluating the effectiveness of the tool. Task accuracies as well as opinions collected from pre- and post-questionnaires indicate that the software is easy to use and was able to improve participants' understanding of survey re-weighting.

The initial design and study are promising, but several challenges remain. Future work includes (1) using a quiz instead of self-reporting to evaluate participants' understanding of re-weighting in surveys, and (2) adding customized analysis that enable users to upload their own survey and baseline data.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1704018.

6 References

[1] S. Keeter, "The Impact of Cell Phone Noncoverage Bias on Polling in the 2004 Presidential Election," *Public Opin Q*, vol. 70, no. 1, pp. 88–98, Jan. 2006.

[2] A. Gelman, "Struggles with Survey Weighting and Regression Modeling," *Statistical Science*, vol. 22, no. 2, pp. 153–164, 2007.