Increasing Understanding of Survey Re-Weighting with Visualization

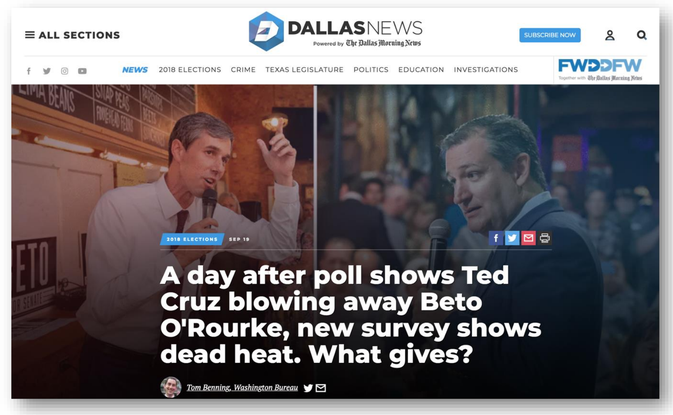
Yufei Zhang, David Borland, and David Gotz, *University of North Carolina at Chapel Hill*

Introduction

- **Surveys widely used to report public opinion**
- **Extremely valuable in wide range of applications**, e.g.:
  - *Politics*
  - *Marketing*
  - *Customer Relationship Mgmt.*
- **Often reported as "facts,"** as in *"Gillum leads by 9"*

The Challenge

- **Surveys are based on responses from a population sample**
- **The sample is rarely representative of the target baseline population**
- **Accurate results require re-weighting of samples**
  - *How to weight?*
  - *What is impact of re-weighting?*



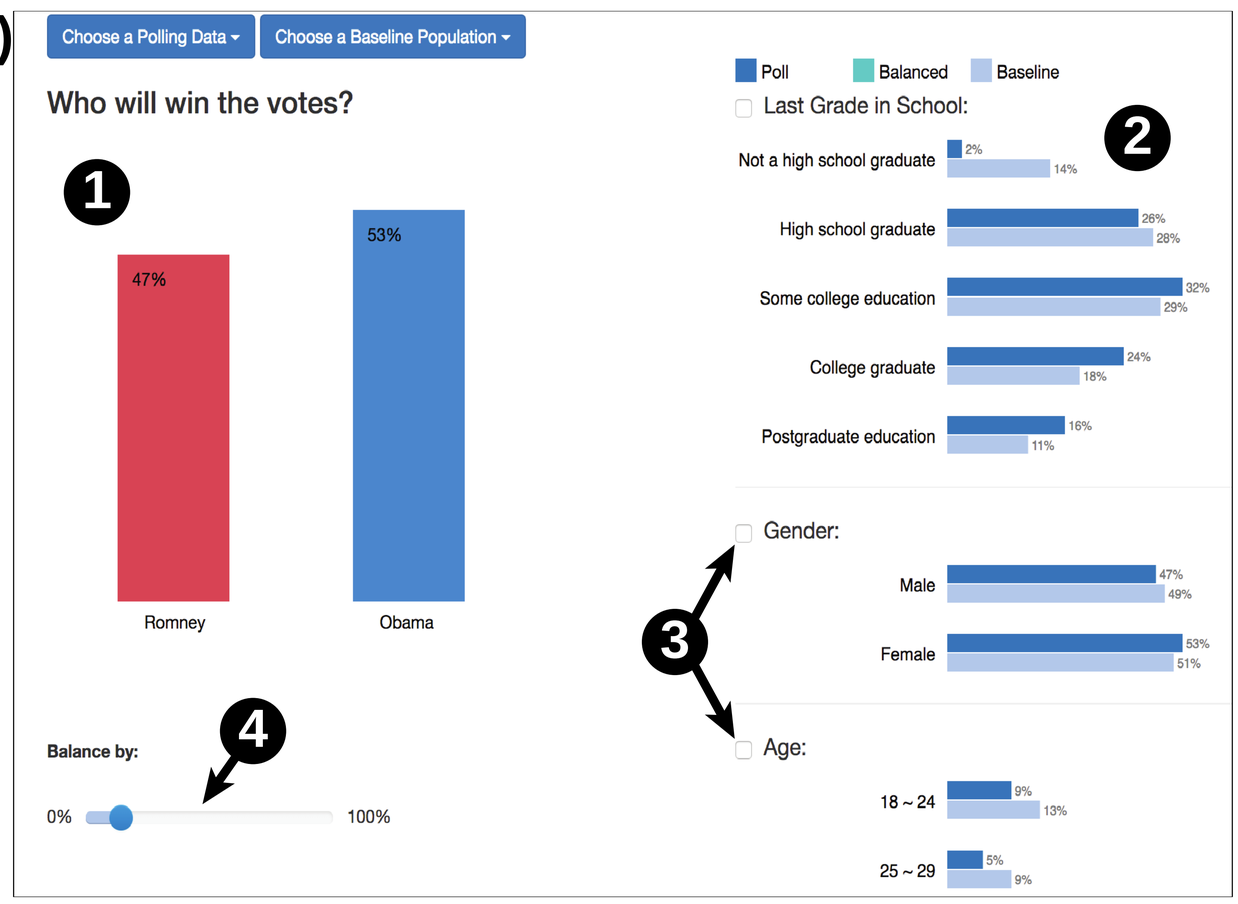
Visualization of Survey Re-Weighting

1. Raw survey response data is shown by default (a). In this case, 2012 presidential election survey.

2. Histograms comparing the sample population with the baseline.

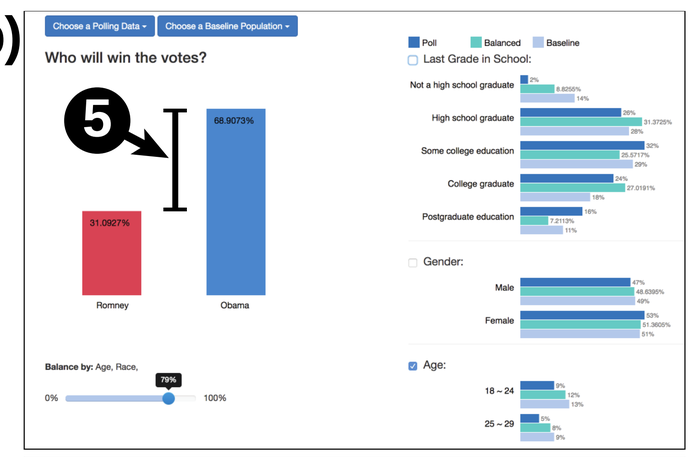
3. Checkboxes allow users to choose which dimensions are included when re-weighting.

(a)

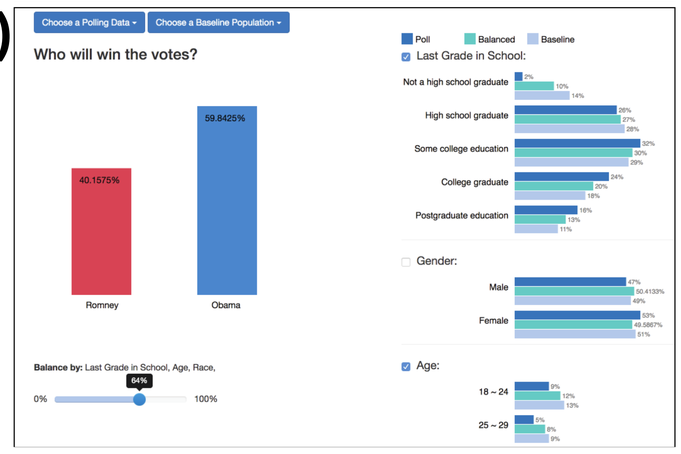


4. A slider allows users to interactively adjust the amount of re-weighting. A value of 0% produces the raw data, 100% will produce a "completely balanced" sample.

(b)



(c)



5. Correcting for age and race gives Obama a larger lead (b). However, adding education level as a dimension for re-weighting results in a narrower survey margin (c).

User Study

- **Does visualization of re-weighting help users understand the impact of the process on survey results?**
- **Study design**
  - *20 participants*
    - Grad students @ UNC-Chapel Hill
    - No prior experience with system
  - *2 datasets*

Poll Data
Monmouth University Poll of National 2012 Presidential Race Barack Obama vs. Mitt Romney

Baseline Publication
Voting age adults in the USA (statistics gathered from Wikipedia)

- *Protocol and Results*
  - 15 tasks: 7 for practice, followed by 8 experimental tasks
    - 160 experimental tasks in total (20x8)
    - 156 answered correctly
    - 96.5% accuracy *shows that users could utilize the system effectively*
  - Pre-post questionnaires with 5-point scale asked how well users understand:
    Q1. overall re-weighting process in surveys? *Q3. impact of the scale of re-weighting?*
    Q2. sensitivity of results to sample bias? *Q4. the overall impact of re-weighting?*