1. Non-probability sampling phone survey

(a) See the filled out version of the PhoneSurvey.xlsx spreadsheet in the repository.

(b) I called 200 numbers, among which only one person responded and 199 people did not respond according to my Response variable. The response rate is 0.5%. Actually there are 7 people who picked up my call, yet 6 of them refused to respond or hanged up without saying a word after I introduced myself and explained my intent of calling them. The rest 193 numbers include quite a large portion of invalid numbers and numbers that are disconnected or no longer in service. The only respondent of my phone survey is a female working in sales industry. I guess the nature of her work probably has familiarized her with the terrible feeling of being refused, so that she could sympathize with my situation and agreed to take my survey.

(c) 100% of those for whom Response = 1 answered the voting question. 0% of those for whom Response = 1 answered the age question. What happened was that when I asked my only respondent the second question, she suddenly got a bit mad, saying that she did not vote and hated talking about it and hanged up right away.

(d) I called them at around 8:30 am - 12am on a working day. I guess my choice of calling time may have posed a negative effect on my respondent rate, because my potential respondents might be working and hence, unable to answer calls when I called them. Quite a few of my calls were not picked up or directed into the voice mailbox.

(e) Unfortunately, I did not manage to collect any age data from my respondents. But judging from voice and the way of speaking, all 7 people who picked up my phone seem to be young (at least below 45). According to the 2012-2016 American Community Survey 5-Year Estimates, the median age in California is 36. Since I did not really get a single response to the age question, I could not infer the reason for the age data matching question.

(f) 0% of my respondents voted Republican (Trump) or Democrat (Clinton) in the 2016 U.S. Presidential election. My only respondent did not vote at all. The actual voting percentages for Republican and Democrat in California in the 2016 election are 33.2% and 61.5%, respectively. I might call a same group of people on different days

to ask them the same survey question with different orders in which I say the candidates or categories. Then I will compare the results of different days to examine if the order influences the results. Another method is to call a group of people that are identified with same attributes (age, gender, education level, etc.) using different candidates/categories orders in the survey question and compare the results.

2. Read the paper Wang et al. (2015), and write a one-to-two-page responding to the following questions.

According to figure 1 in Wang et al. (2015), the three least representative variables are age, sex and education. The three most representative variables are state, race and ideology. Since younger male with lower education level has higher probability to play computer games frequently and involve in gaming platforms than their counterparts regarding each of the above three characteristics, the Xbox population is identified with a much higher portion of young men who are not college graduates than the broader voting population.

The authors use the state and national exit polls data from the 2008 presidential election and Xbox poll data to re-weight the respondents in the Xbox poll data.

According to figure 2 in Wang et al (2015), the Xbox raw data would have predicted that Romney wins, as the two-party Obama support remained below 50% in the last three weeks; Pollster.com forecast data would have predicted an uncertain outcome, as the two-party Obama support fluctuated around 50% in the last three weeks. According to figure 3 in Wang et al (2015), the Xbox post-stratified data would have predicted that Obama wins, as the two-party Obama support remained above 50% in the last three weeks.