1. Non-probability sampling phone survey

(a) The spreadsheet is attached.

(b) I called 200 numbers in total.

According to my response variable, 11 people responded. The other 189 calls don't get any respond, and some of them are not in service. The respond rate of this survey is 5.5%.

(c) No one of those 11 people answer the both questions, age and vote. They just hang up during or after my introduction.

(d) I called them at 3 different time, which are Friday afternoon, Saturday noon and Sunday night. From this phone survey experience, I don't think there are some differences across time. The respond rates are all very low. This may be because those phone numbers are randomly generated. My guess is the response rate will be higher during workdays. Nowadays, people are all contacting their friends and family members using social media. Phone calls are more often used in work place (both email and phone call). I conclude this from my own experience. When I was doing an internship in a company, I sometimes got a phone call from my boss. However, when I am in school, I seldomly get phone calls.

(e) According to U.S. Census Bureau Factfinder table, the median age in Texas is 34.2. However, since no one answers the age question in my phone survey. I cannot compare that. If I could have a larger data set, maybe I will have a conclusion. However, I think this is so difficult to get the similar result with the actual state median age. The first reason is the response rate is too low to give us a solid result. The second reason is there existing selection bias on the collected result. For example, the people who answer the phone survey are usually old people, because they have more free time and don't care talking with you for several minutes.

(f) The 2016 election result of Texas is Donald J. Trump. Since I don't have any valid response, I cannot answer the first three questions. For the last question, I think it will be milder asking did you vote at first. After this question, we can ask the year they were born in. This sequence will make respondents feel that asking age is not infringing on their privacy, but just research needs. Also, after the respondents answered the question about whether they voted, we can just ask: who did you vote? Instead of asking Hillary or Trump, this may be a little offensive to those fanatic supporters.

2. Predicting elections survey.

(a) From figure 1 in this paper(Wang, Rothschild, Goel, and Gelman, 2015), I can clearly see that these three variables are significantly different from 2012 exit poll, which are sex, age and education. From my perspective, the most representative variables are race, state and 2008 vote. These variables will not change or have minor difference between the specific crowds of X-box and all voters.

There are far more males in X-box dataset. This is because man prefer to play games than women in general. My guess is women will spend more time on other things like shopping, makeup and dancing, etc. Also, age and education pattern are quite different. X-box respondents are more inclined to be young people. With the age increasing, the number of people who use X-box will drop substantially. This is not the same age pattern

with the actual voters. The age distribution of voters and the overall population of the United States should be very close. Also, the education distribution is very different. There is a significant drop in the number of college graduates playing X-box. As before, this is dissimilar from the actual voters.

(b) I think the two data sources are X-box data and 2008 exit poll data. The author says: "Having detailed the multilevel regression step, we now turn to poststratification, where the cell-level estimates are weighted by the proportion of the electorate in each cell and aggregated to the appropriate level (i.e., state or national). To compute cell weights, we require cross-tabulated population data. (Wang, Rothschild, Goel, and Gelman, 2015)" Then, the author mentions the data of Current Population Survey(CPS), which is a commonly used data source for adjusting population pattern. However, it do have a shortcoming here. "the CPS is missing some key poststratification variables, such as party identification. (Wang, Rothschild, Goel, and Gelman, 2015)" Therefore, the author uses 2008 exit poll data instead. The advantage of using 2008 exit poll data, not 2012, is this could help the author to predict the 2012 election results. The disadvantage is "the author cannot capture the demographic shifts of the intervening four years." Also, the author mentions why using only 2008 exit poll data instead of combination of 2008 exit poll and CPS data. "While combining exit poll and CPS data could arguably yield improved results, we limit ourselves to the 2008 exit poll summaries for our poststratification, for the sake of simplicity and transparency. (Wang, Rothschild, Goel, and Gelman, 2015)"

(c) According to this paper, in the last three weeks of election, X-box row data would have predicted that Romney would win; Pollster.com would have predicted the uncertain result because the line is around 50% (Figure 2, Wang, Rothschild, Goel, and Gelman, 2015); and X-box post-stratified would have predicted the correct result: Obama would win (Figure 3, Wang, Rothschild, Goel, and Gelman, 2015).