Question 1.

Part a.
Please check PhoneSurvey.xlsx, uploaded on github repository.

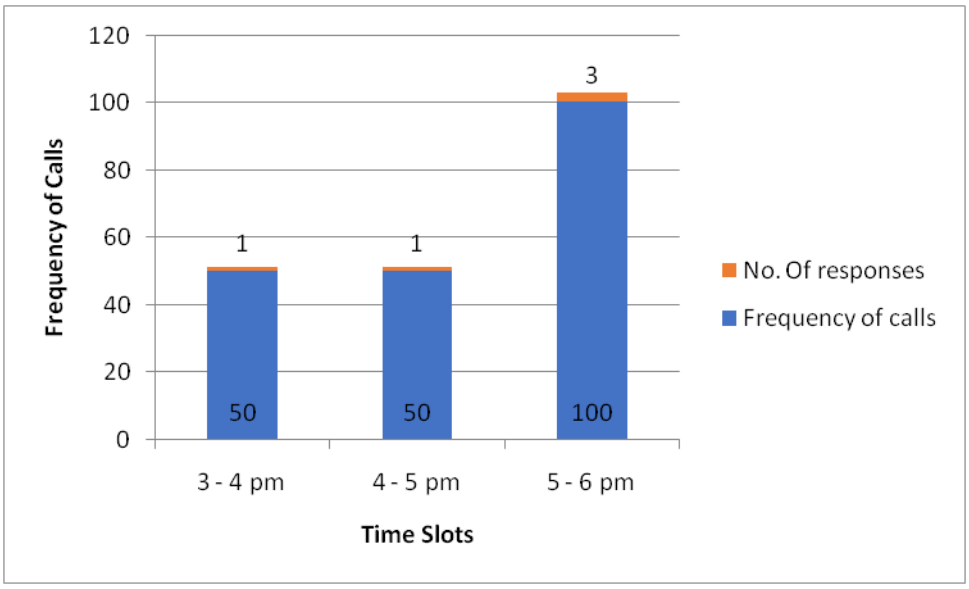
Part b.
I called up the 200 numbers provided in the PhoneSurvey excel file. According to the definition of the response variable, 5 people responded while 195 didn't respond to the call at all or dropped before I could ask them the voting question, "In the 2016 U.S. Presidential Election, did you vote Democrat (Clinton), Republican (Trump), Other, or Did Not Vote?"

The response rate was 2.5% (5 responses out of 200).

Part c.
Out of the five responses (Response equals to 1), 4 people answered the voting question. Surprisingly, all of the four responses to the voting question were different and mutually covered the four distinct answer choices to the voting question.

Out of the five responses, 3 people answered the age question. Out of the three answers, two were exact numbers while one answered in the form of range, "age over 40".

Part d.
Overall, all the calls were made between 3 pm and 6 pm. The majority of the calls (100) were made between 5 – 6 pm while the time slots 3 – 4 pm and 4 – 5 pm had 50 calls each.



I think time of the call played a little role in the response rate. Number of responses observed during 5 – 6 pm are triple the number of responses observed in the other two time slots, even though the frequency of calls is just double of the other timings. It could be because 5-6 pm slot is when the work shift has ended and people are comparatively free. An important point to be noted over here is that many of these phone numbers weren't operating, so a definite conclusion cannot be based on the reported frequency of calls.

Part e.
Out of the five responses, only three people answered the age question. Even amongst those three responses, one of the individuals replied with "Over 40" to the question, "What is your age?" The other ages were reported to be 35 and 53.

Hence, the median age is above 40 and less than or equal to 56 in the observed data. The median age as seen in U.S. Census Bureau data (State: Florida, Year: 2016) is 42.1[1], which definitely lies in the range of sample median. If we drop the "Over 40" value and re-calculate median using only two values, we obtain 43 which is close to 42.1 years of age. One of the possible reasons for a close match between the two medians could be that the younger generation don't tend to pick up unknown numbers and even if they do, they have a higher tendency to not to take a survey. The relatively older generation are more patient with phone calls and tend to accept calls from unknown numbers and talk for some duration.

Part f.
According to my survey data, 20% of the respondents voted Republican (Trump) and 20% of the respondents voted Democrat (Clinton) in the 2016 U.S. Presidential election. Based on the election results, Republican (Trump) had received 49.1% of the votes while Democrat (Clinton) received 47.8% of the votes[2]. A difference of 1.3% was observed in the actual voting results while there were no differences in our findings from the survey data.

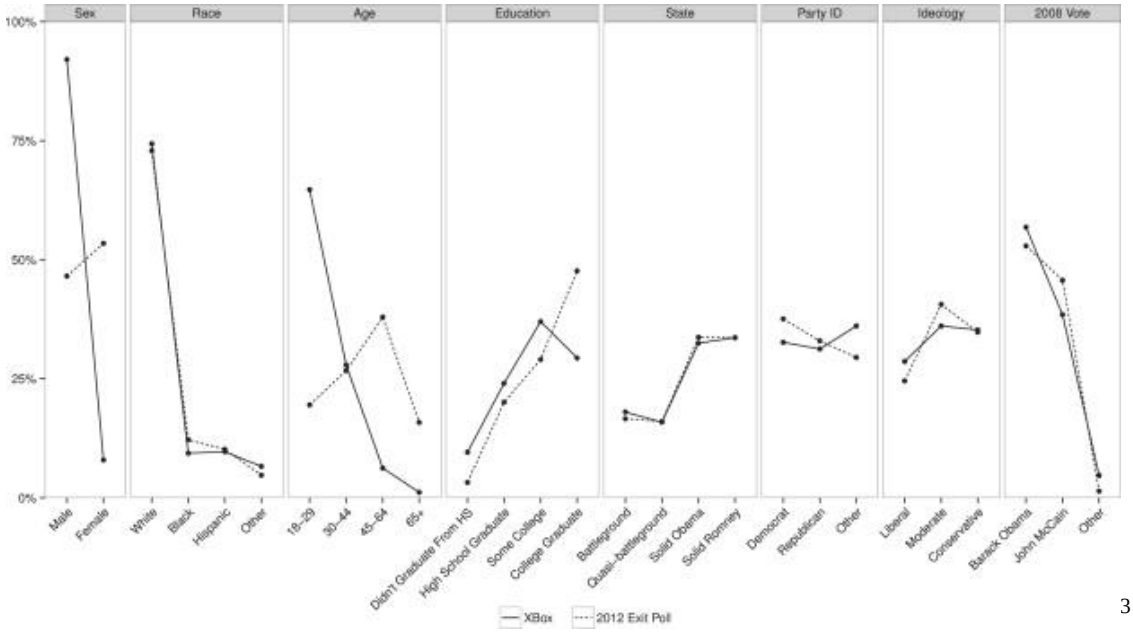
To test if the order in which the candidates or categories are said in the survey question has any influence on the result, I would randomly assign the possible different orderings (4! = 24, in our case) to the different phone numbers equally. And then we could easily observe if the ordering of the categories had any influence on the survey results based on the category wise results.

[1] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk
[2] https://www.politico.com/2016-election/results/map/president/florida

Question 2.

Part a.

The eight variables reported from the respondents are: Sex, Race, Age, Education, State, Party ID, Ideology and 2008 Vote.



[3]

The three most representative variables of the data are:

- Race: White and Other races are very minutely over-represented while Black is under-represented, but even that has a very small deviation from representative 2012 Exit Poll data. Hispanics have the (almost) same representation in XBox and 2012 Exit Poll data.

- State: Quasi-battleground and Solid Romney states have same representation in XBox and 2012 Exit Poll data. Battleground and Solid Obama states are over-represented and under-represented, respectively. But the deviations in both the cases are too small to be considered significant.

- Ideology: XBox data represents the Conservatives very well while it over-represents and under-represents Liberals and Moderates, respectively. But the deviations are smaller when compared with the deviations for the other variables except Race and State.

[3] Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980-991.

The three least representative variables of the data are:

- Sex: XBox data representation does a poor job representing sex variable. The deviations are off by a large magnitude in case of both, males and females.

- Age: XBox data gets representation of 30-44 age group correct and also informs us about the decline in population from 45-64 age group to 65+ age group. Though the magnitude of decline is not same as observed in representative 2012 Exit Poll data, it captures the decline nevertheless. The other age groups (18-29 and 45-64) have very large deviations, contributing to least representativeness.

- Education: Though XBox has a similar trend like representative 2012 Exit Poll data for the initial education categories till college education (included), the data is always over-represented. For college graduates, XBox data over-represents with large differences.

A possible explanation for three least representative variables of XBox data to be different from the broader voting population:
A larger population of males play videogames than females and also, most of the video-gamers are in 10-25 age group. In addition, to avoid cyber-threat, a lot of females register themselves as males while playing online[4]. It could add to the already significant gender gap between people who would play XBox. The older generation isn't as tech-savvy and play very less of XBox and have a reasonable interest in politics. Hence, 45-64 age group is under-represented in XBox data. Generally, most of the XBox players are young students in High Schools or College and hence these groups are well represented through XBox data while more education (in terms of college degrees) is positively correlated with age, which in turn has a bad representation through XBox data (in comparison to 2012 Exit Poll data).

Part b.
The authors have used exit poll data from presidential elections conducted in 2008 to perform a post-stratification re-weighting of the respondents. In order to be able to make predictions for 2012 they don't use more recent 2012 Exit Poll data.

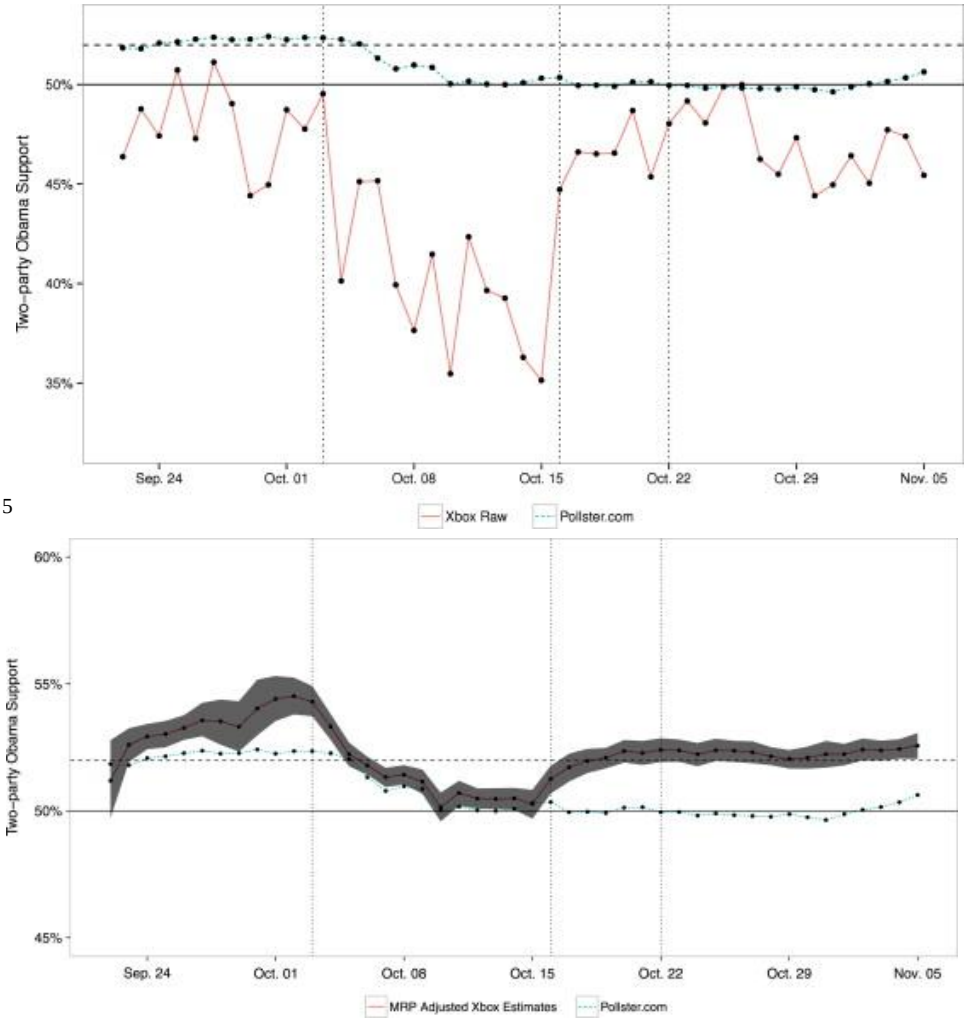
Part c.
In the following two figures, Prediction of 2012 U.S. Presidential election is mapped on the Y-axis in the form of Two-Party Obama Support. On the y-axis, a value above 50% corresponds to Obama winning the elections and a value less than 50% corresponds to Romney winning the elections. A value of 50% implies uncertainty regarding who will win the elections.

[4] https://www.washingtonpost.com/news/the-switch/wp/2014/10/17/more-women-play-video-games-than-boys-and-other-surprising-facts-lost-in-the-mess-of-gamergate/?noredirect=on&utm_term=.76366f168043

Consider, X-Box raw data which is plotted in the next figure. In the last three weeks of the election, the support for Obama starts building up from 35% to around 48% at the end of the last third week. In the second last week, it is mostly between 45-50 % but then it successively drops to 45%. In the last week, there is high volatility throughout and at the end the data predicts around 46% Obama support, which implies, win for Romney.

Pollster.com has a representative data unline XBox data and hence, have a very much stable trend during the last three weeks. Around the starting of the last week, there is a decline in Obama support but it ends up predicting win for Obama (greater than 50% for Obama support).

In the last three weeks of the election, XBox post-stratified started deviating from the Pollster data in support of Obama. After the end of last third week, there was an approximate increase of 3% in Obama support. During the last two weeks of the election, the predictions are more or less stable and in favour of Obama winning the Presidential Elections, the original outcome.



[5] Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980-991.