Part 1 If we're trying to predict the results of the Clinton vs. Trump presidential race, what is the population of interest?

The eligible voting population within the United States.

Part 2 What is the sampling frame?

Correspondents reached through random digit dialing or other communication methods who answer the poll.

0.0.1 Question 5

Why can't we assess the impact of the other two biases (voters changing preference and voters hiding their preference)?

Note: You might find it easier to complete this question after you've completed the rest of the homework including the simulation study.

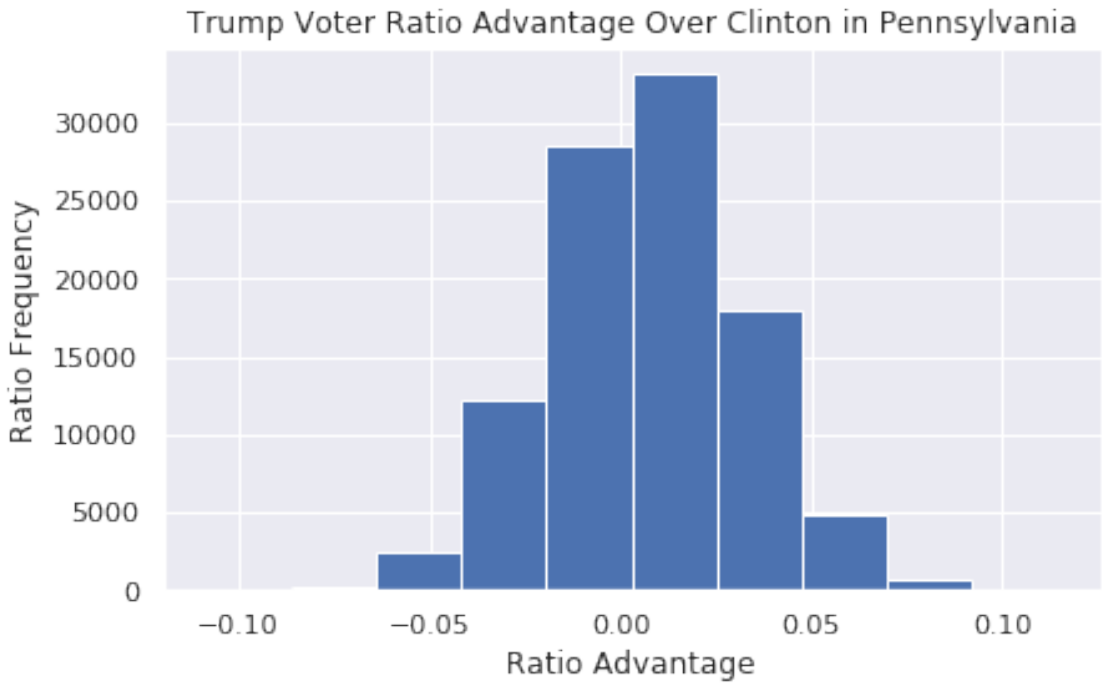
There is no way for us to determine how many people hid their support for by hiding from pollsters, so this will be an unknown amount of voters. Also, when people vote on polls, there is no systematic way to determine which people lied on the polls and changed their votes during the actual election. Therefore, we cannot analyze the effect of voters changing and hiding their preference.

Part 4 Make a histogram of the sampling distribution of Trump's proportion advantage in Pennsylvania. Make sure to give your plot a title and add labels where appropriate. Hint: You should use the `plt.hist` function in your code.

Make sure to include a title as well as axis labels. You can do this using `plt.title`, `plt.xlabel`, and `plt.ylabel`.

```
In [38]: plt.hist(simulations)
         plt.title('Trump Voter Ratio Advantage Over Clinton in Pennsylvania')
         plt.xlabel('Ratio Advantage')
         plt.ylabel('Ratio Frequency')
```

```
Out[38]: Text(0, 0.5, 'Ratio Frequency')
```

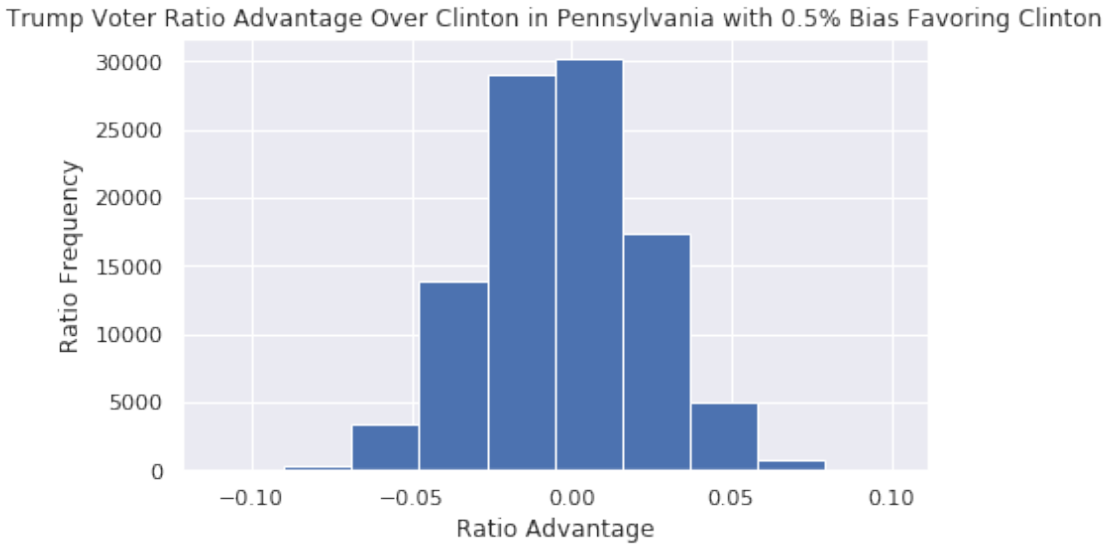


Part 2 Make a histogram of the new sampling distribution of Trump's proportion advantage now using these biased samples. That is, your histogram should be the same as in Q6.4, but now using the biased samples.

Make sure to give your plot a title and add labels where appropriate.

```
In [45]: plt.hist(biased_simulations)
         plt.title('Trump Voter Ratio Advantage Over Clinton in Pennsylvania with 0.5% Bias Favoring Cl
         plt.xlabel('Ratio Advantage')
         plt.ylabel('Ratio Frequency')
```

```
Out[45]: Text(0, 0.5, 'Ratio Frequency')
```



Part 3 Compare the histogram you created in Q7.2 to that in Q6.4.

In the biased histogram, Trump on average has much less of a voter ratio advantage and has more negative and zero ratio advatanges compared with the unbiased histogram, where he has more positive ratios.

Write your answer in the cell below.

A larger sample size affects both sampling error and bias. Larger sample sizes reduce sampling errors because our sampling frame is much more representative of the population. However, larger sample sizes further increase error and amplify the bias towards a specific group or candidate in biased samples.

0.0.2 Question 9

According to FiveThirtyEight: "… Polls of the November 2016 presidential election were about as accurate as polls of presidential elections have been on average since 1972."

When the margin of victory may be relatively small as it was in 2016, why don't polling agencies simply gather significantly larger samples to bring this error close to zero?

As seen with question 8, larger sample sizes do not necessarily always help. Once the sample sizes have become large enough, the winning ratio for Trump began to converge. So, the error would not have been reduced signifcantly by obtaining larger samples.