Assignment 3

Several factors complicate the analysis of this particular experiment. First, the probability of being randomly assigned to treatment was different for urban and non-urban areas. Second, some people assigned to treatment did not receive the treatment. And third, we are unable to observe turnout for some of the subjects. See the README file for more details.

a) Calculate the mean value of turnout for people who did and did not receive the treatment, and interpret the implied effect of get-out-the-vote interventions on turnout. Think about the likely biases that arise from the three complications listed above. If you had to guess, would you say that you are likely over- or under-estimating the average effect with this analysis? Explain your answer.

Below I find the naive implied effect of get-out-the-vote interventions on turnout by finding the average difference in turnout between those who were and were not successfully treated. The resulting estimate, importantly, does not account for any of the complications enumerated above, except I drop those subjects who attrited to remove their missing values from calculations.

```
# find average turnout of those who received the treatment
mean_turnout_t1 <- data %>%
  filter(successfultreatment == 1) %>%
  filter(is.na(turnout) != 1) %>%
  summarise(mean = mean(turnout))
mean_turnout_t1
```

```
##        mean
## 1 0.5935141
```

```
# find average turnout of those who did not receive the treatment
mean_turnout_t0 <- data %>%
  filter(successfultreatment == 0) %>%
  filter(is.na(turnout) != 1) %>%
  summarise(mean = mean(turnout))
mean_turnout_t0
```

```
##        mean
## 1 0.4742721
```

```
# estimate the implied effect of treatment
implied_effect <- mean_turnout_t1 - mean_turnout_t0
implied_effect
```

```
##        mean
## 1 0.1192421
```

The average difference in turnout for people who did and did not receive the treatment was .119. This can be naitvley interpreted as the treatment, a get-out-the-vote intervention, having caused an 11.9% increase in turnout.

This number is likely an overestimate of the average effect of the treatment. Two of the three complications to the experiment–non-random assignment and presence of noncompliance–create bias and baseline differences between our treatment and control groups. The third complication–attrition–does not create bias in either direction.

The first complication–non-random assignment–makes living in an urban area a confounder because residential location influences turnout and likelihood of treatment above and beyond random assignment. Urbanites are, on average, less likely to vote and are also disproportionately assigned to the control group. The combination of these two factors results in a positive bias that overstates the effect of the treatment.

```
# find proprotion of urbanites in the sample
prop_urban <- data %>%
  filter(urban == 1) %>%
  count()/50000
prop_urban
```

```
##        n
## 1 0.4975
```

```
# find proportion of non-urbanites in the sample
prop_nonurban <- data %>%
  filter(urban == 0) %>%
  count()/50000
prop_nonurban
```

```
##        n
## 1 0.5025
```

```
# find proportion of urbanites who were successfully treated
prop_urban_treat <- data %>%
  filter(urban == 1 & successfultreatment == 1) %>%
  count()/24875
prop_urban_treat
```

```
##           n
## 1 0.1507136
```

```
# find proportion of urbanites who were not successfully treated
prop_urban_control <- data %>%
  filter(urban == 1 & successfultreatment == 0) %>%
  count()/24875
prop_urban_control
```

```
##           n
## 1 0.8492864
```

```
# find proportion of non-urbanites who were successfully treated
prop_nonurban_treat <- data %>%
  filter(urban == 0 & successfultreatment == 1) %>%
  count()/25125
prop_nonurban_treat
```

```
##           n
## 1 0.6188657
```

```r
# find proportion of non-urbanites who were not successfully treated
prop_nonurban_control <- data %>%
  filter(urban == 0 & successfultreatment == 0) %>%
  count()/25125
prop_nonurban_control
```

```
##           n
## 1 0.3811343
```

Urbanites and non-urbanites represent roughly half of the sample, yet as the numbers above show, they both groups not equally likely to receive treatment. This non-random assignment creates baseline differences between those were and were not successfully treated because urbanites are, on average, less likely to vote (shown below).

```r
# find proportion of urbanites in the control group who voted
control_urban_turnout <- data %>%
  filter(urban == 1 & treatmentattempt == 0 & turnout == 1) %>%
  count()/19868
control_urban_turnout
```

```
##           n
## 1 0.4695994
```

```r
# find proportion of non-urbanites in control who voted
control_nonurban_turnout <- data %>%
  filter(urban == 0 & treatmentattempt == 0 & turnout == 1) %>%
  count()/5033
control_nonurban_turnout
```

```
##           n
## 1 0.5306974
```

```r
# find average difference in likelihood to turnout
turnout_difference <- control_urban_turnout - control_nonurban_turnout
turnout_difference
```

```
##             n
## 1 -0.06109804
```

This means that those who were successfully treated (disproportionately non-urban) were more likely to turnout absent the intervention; the experiment is not comparing apples-to-oranges. As a result, urban residence is a confounder that creates positive bias.

The second complication is the presence of noncompliers. Not all people assigned to treatment were successfully treated. The bias created by compliance is also likely positive. We can reason that noncompliers are less likely to turnout than compliers. Perhaps noncompliers, on average, are less friendly or community-oriented and so are less likely to answer their door when they hear a political canvasser knocking. I would wager that these characteristics are correlated with lower probabilities of turnout.

As such, baseline differences would be present, and noncompliance is a confounder. Those who who successfully received the treatment would be, before treatment, more likley to vote which results in an overestimation of the effect of get-out-the-vote interventions.

The third complication is attrition. Many of the observations in both the control and the treatment groups attrited from the sample, that is, whether they voted could not be ascertained perhaps because they moved out of state, died, or mysteriously vanished. Crucially, whether or not this attrition biases the implied effect of get-out-the-vote interventions depends on the nature of the attrition. If attrition occurred at random (i.e., is unrelated to assigmnment), then attrition does not bias our estimate (only reduces our sample size).

I do not think there are strong reasons to believe attrition systematically more likely to occur because of assignment to be treated. The hypothetical reasons enumerated above are rather idiosyncratic; assignment to the treatment group is hardlly likely to make a participant move out of state. As such, I would wager that attrition is random in nature and therefore does not bias our the implied effect of get-out-the-vote interventions.

b) Using the lessons from Chapter 10, try to account for the fact that probability of treatment between urban and non-urban areas. How did you estimate your change? Why?

```r
# regress turnout on treatment
regression <- lm(data = data, turnout ~ successfultreatment)
summary(regression)
```

```
##
## Call:
## lm(formula = turnout ~ successfultreatment, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5935 -0.4743  0.4065  0.5257  0.5257
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.47427    0.00284  167.01   <2e-16 ***
## successfultreatment  0.11924    0.00457   26.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4962 on 49740 degrees of freedom
##   (258 observations deleted due to missingness)
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.01348
## F-statistic: 680.9 on 1 and 49740 DF,  p-value: < 2.2e-16
```

```r
# regress turnout on treatment and urban
regression_test <- lm(data = data, turnout ~ successfultreatment + urban)
summary(regression_test)
```

```
##
## Call:
## lm(formula = turnout ~ successfultreatment + urban, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5994 -0.4951  0.4006  0.5049  0.5352
```

```
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.495063   0.004499 110.044  < 2e-16 ***
## successfultreatment  0.104319   0.005210  20.024  < 2e-16 ***
## urban               -0.030223   0.005073  -5.957 2.58e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.496 on 49739 degrees of freedom
##   (258 observations deleted due to missingness)
## Multiple R-squared:  0.01421,    Adjusted R-squared:  0.01417
## F-statistic: 358.4 on 2 and 49739 DF,  p-value: < 2.2e-16
```

I tried accounting for the bias introduced by the unequal probability of being assigned to treatment between urban and non-urban areas by regressing turnout on both treatment and the "urban" variable.

This procedure controls for the bias introduced by the confounder because including an urban variable in the regression by incorporating information about the effect of this confounder upon the outcome. This improves the accuracy of the estimated relationship between the treatment and the outcome by holding the urban variable constant.

My estimate decreased by from .119 to .104, for a decrease of .015. This likely occurred because urban residence makes an individual both less likely to turnout and be successfully treated. Making this effect explicit, reduces the reported relationship between treatment and turnout.

c) Using the lessons from the chapter, let's try to account for noncompliance. First, try to estimate the intent-to-treat effect (reduced form) and the compliance rate (first stage). Now divide the former by the latter to estimate the compiler average treatment effect.

I will estimate a separate complier average treatment effect (CATE) for urbanites and non-urbanites in the sample. Finding the CATE uses instrumental variables to account for noncompliance and estimating separate CATEs mitigates the bias produced by non-random assignment.

In practice, I will find the CATE by first identifying the "intent-to-treat" (ITT) effect, the average difference in turnout between those assigned to treatment and control, for both groups. Then I will find the proportion of compliers and divide the ITT effect by that proportion to get an unbiased estimate of the CATE. This implements the "Wald Estimator" for both urban and non-urban groups. I choose to drop those observations for whom we do not know their turnout.

```r
# find the intent-to-treat effect for both groups
# urban first - estimate number of non-attrited urbanites in treatment group
n_urban_treatment <- data %>%
  filter(urban == 1 & treatmentattempt == 1 & is.na(turnout) != 1) %>%
  count()

# estimate number of non-attrited urbanites in control group
n_urban_control <- data %>%
  filter(urban == 1 & treatmentattempt == 0 & is.na(turnout) != 1) %>%
  count()

# find average turnout for control and treatment groups
urban_turnout_treatment <- data %>%
  filter(urban == 1 & treatmentattempt == 1 & turnout == 1) %>%
  count()/n_urban_treatment
```

```r
urban_turnout_control <- data %>%
  filter(urban == 1 & treatmentattempt == 0 & turnout == 1) %>%
  count()/n_urban_control

# find the average difference in turnout for treatment and control groups
ITT_urban <- urban_turnout_treatment - urban_turnout_control
ITT_urban
```

```
##            n
## 1 0.04104669
```

```r
# non-urban second - estimate number of non-attrited non-urban residents in treatment group
n_nonurban_treatment <- data %>%
  filter(urban == 0 & treatmentattempt == 1 & is.na(turnout) != 1) %>%
  count()

# estimate number of non-attrited non-urban residents in treatment group
n_nonurban_control <- data %>%
  filter(urban == 0 & treatmentattempt == 0 & is.na(turnout) != 1) %>%
  count()

# find average turnout for control and treatment groups
nonurban_turnout_treatment <- data %>%
  filter(urban == 0 & treatmentattempt == 1 & turnout == 1) %>%
  count()/n_nonurban_treatment

nonurban_turnout_control <- data %>%
  filter(urban == 0 & treatmentattempt == 0 & turnout == 1) %>%
  count()/n_nonurban_control

# find the average difference in turnout for treatment and control groups
ITT_nonurban <- nonurban_turnout_treatment - nonurban_turnout_control
ITT_nonurban
```

```
##            n
## 1 0.03274022
```

The estimated intent-to-treat effects are .041 and .033 for urban and non-areas, respectively, which describe the average difference in turnout between treatment and control groups for urban and non-urban areas. Next I estimate the proportion of compliers in each group.

```r
# find the proportion of compliers for both groups
# urban first - estimate proportion of compliers and always-takers in treatment group
urban_compliers <- data %>%
  filter(urban == 1 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/n_urban_treatment
urban_compliers
```

```
##           n
## 1 0.7529624
```

```r
# non-urban second - estimate proportion of compliers and always-takers in treatment group
nonurban_compliers <- data %>%
  filter(urban == 0 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/n_nonurban_treatment
nonurban_compliers
```

```
##           n
## 1 0.7773723
```

The numbers above imply that the proportion of compliers for urban and non-urban areas are .753 .777, respectively. Now we implement the Wald Estimator for both groups to find an unbiased estimate of their CATEs.

```r
# implement the Wald Estimator by dividing the ITT by proportion of compliers
# urban first - divide the ITT by proportion of urban compliers
urban_CATE <- ITT_urban/urban_compliers
urban_CATE
```

```
##           n
## 1 0.0545136
```

```r
# non-urban second - divide the ITT by proportion of non-urban compliers
nonurban_CATE <- ITT_nonurban/nonurban_compliers
nonurban_CATE
```

```
##            n
## 1 0.04211652
```

After implementing the Wald Estimator, we find unbiased estimates of the CATE for both urban and non-urban areas, which are .055 and .042, respectively. We now find an overall CATE for get-out-the-vote interventions by taking an average, weighted by both group's proportion in the overall sample. I calculate these proportions and take a weighted average to find a final CATE.

```r
# find proportion of urbanites in the sample
prop_urban <- data %>%
  filter(urban == 1) %>%
  count()/50000

# find proportion of non-urbanites in the sample
prop_nonurban <- data %>%
  filter(urban == 0) %>%
  count()/50000

# weight average the two unbiased estimates of CATE
final_CATE <- prop_urban * urban_CATE + prop_nonurban * nonurban_CATE
final_CATE
```

```
##            n
## 1 0.04828407
```

The final, unbiased estimate of the CATE for get-out-the-vote interventions is 0.0483, that is, compliers were 4.83% more likely to turnout after being assigned to the treatment group.

d) Think about the attrition problem. What are you implicitly assuming if you just drop the subjects for whom we don't observe their turnout? Let's see how their estimates change under different assumptions. Estimate the complier average treatment effect assuming that none of the subjects who attrited would have voted. What would your estimate be under the worst-case scenario for the effectiveness of GOTV? What about the best-case scenario?

If I were to drop the subjects for whom we do not observe their turnout, I would be implicitly assuming that the group that attrited was not any different, on average, from the group overall. In other words, I would be assuming that attrition was random. This is not always the case. If attrition was systematically related to assignment in some unobserved way, then dropping the former from the experiment might bias the calculated estimates of the treatment'seffect.

Just in case, I will estimate the treatment effect assuming that none of the subject who attrited would have voted (the base-case). I will also identify and interpret the worst- and base-case scenario. I start by imputing the missing values from the data set for each of these scenarios: in the base-case, all attrited are assumed to not vote; in the worst-case, the attrited in the control are assumed to have voted and in treatment are assumed to not; in the best-case, the attrited in the control are assumed to not have voted and in treatment are assumed to have.

```
# impute no voting for all attrited (the base case)
base_case_data <- data %>%
  mutate(turnout = replace_na(turnout, 0))

# impute all voting for attrited in control and no voting in treatment (the worst case)
# replace all missing values of turnout in control with 1
control_worst_case <- data %>%
  filter(treatmentattempt == 0) %>%
  mutate(turnout = replace_na(turnout, 1))

# replace all missing values of turnout in control with 0
treatment_worst_case <- data %>%
  filter(treatmentattempt == 1) %>%
  mutate(turnout = replace_na(turnout, 0))

# rejoin the two data sets
worst_case_data <- control_worst_case %>%
  full_join(treatment_worst_case)
```

```
## Joining, by = c("female", "age", "white", "black", "employed", "urban", "treatmentattempt", "success
```

```
# impute no voting for attrited in control and all voting in treatment (the best case)
# replace all missing values of turnout in control with 0
control_best_case <- data %>%
  filter(treatmentattempt == 0) %>%
  mutate(turnout = replace_na(turnout, 0))

# replace all missing values of turnout in control with 1
treatment_best_case <- data %>%
  filter(treatmentattempt == 1) %>%
  mutate(turnout = replace_na(turnout, 1))

# rejoin the two data sets
```

```r
best_case_data <- control_best_case %>%
  full_join(treatment_best_case)
```

```
## Joining, by = c("female", "age", "white", "black", "employed", "urban", "treatmentattempt", "success
```

Now, with values imputed for the base, worst, and best case of attrition, we can calculate and compare the complier average treatment across these scenarios. This will give us upper and lower bounds on the bias from attrition. For each group, I follow the same procedure as in part and report the resulting CATEs and then interpret their meaning as upper and lower bounds on the bias introduced by attrition.

```r
# find the intent-to-treat effect for urban and rural groups in the base case
# urban first - estimate number of urbanites in treatment group
urban_treatment_base <- base_case_data %>%
  filter(urban == 1 & treatmentattempt == 1) %>%
  count()
urban_treatment_base
```

```
##      n
## 1 5007
```

```r
# estimate number of urbanites in control group
urban_control_base <- base_case_data %>%
  filter(urban == 1 & treatmentattempt == 0) %>%
  count()
urban_control_base
```

```
##       n
## 1 19868
```

```r
# find average turnout for control and treatment groups
urban_turnout_treatment_base <- base_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & turnout == 1) %>%
  count()/urban_treatment_base

urban_turnout_control_base <- base_case_data %>%
  filter(urban == 1 & treatmentattempt == 0 & turnout == 1) %>%
  count()/urban_control_base

# find the average difference in turnout for treatment and control groups
ITT_urban_base <- urban_turnout_treatment_base - urban_turnout_control_base
ITT_urban_base
```

```
##            n
## 1 0.04088596
```

```r
# non-urban second - estimate number of non-urban residents in treatment group
nonurban_treatment_base <- base_case_data %>%
  filter(urban == 0 & treatmentattempt == 1) %>%
  count()
nonurban_treatment_base
```

```
##       n
## 1 20092
```

```r
# estimate number of non-urban residents in treatment group
nonurban_control_base <- base_case_data %>%
  filter(urban == 0 & treatmentattempt == 0) %>%
  count()
nonurban_control_base
```

```
##      n
## 1 5033
```

```r
# find average turnout for control and treatment groups
nonurban_turnout_treatment_base <- base_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & turnout == 1) %>%
  count()/nonurban_treatment_base

nonurban_turnout_control_base <- base_case_data %>%
  filter(urban == 0 & treatmentattempt == 0 & turnout == 1) %>%
  count()/nonurban_control_base

# find the average difference in turnout for treatment and control groups
ITT_nonurban_base <- nonurban_turnout_treatment_base - nonurban_turnout_control_base
ITT_nonurban_base
```

```
##            n
## 1 0.03295978
```

```r
# find the proportion of compliers for both groups
# urban first - estimate proportion of compliers and always-takers in treatment group
urban_compliers_base <- base_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/urban_treatment_base
urban_compliers_base
```

```
##           n
## 1 0.7487517
```

```r
# non-urban second - estimate proportion of compliers and always-takers in treatment group
nonurban_compliers_base <- base_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/nonurban_treatment_base
nonurban_compliers_base
```

```
##           n
## 1 0.7738901
```

```r
# implement the Wald Estimator by dividing the ITT by proportion of compliers
# urban first - divide the ITT by proportion of urban compliers
urban_CATE_base <- ITT_urban_base/urban_compliers_base
urban_CATE_base
```

```
##           n
## 1 0.0546055
```

```
# non-urban second - divide the ITT by proportion of non-urban compliers
nonurban_CATE_base <- ITT_nonurban_base/nonurban_compliers_base
nonurban_CATE_base
```

```
##            n
## 1 0.04258974
```

```
# weight average the two unbiased estimates of CATE
final_CATE_base <- prop_urban * urban_CATE_base + prop_nonurban * nonurban_CATE_base
final_CATE_base
```

```
##            n
## 1 0.04856758
```

The resulting overall CATE is .0486. Assuming that all attrited subjects did not vote, then compliers were 4.86% more likely to vote after being assigned to treatment. This result is very similar to the estimate I found in part c and so does not help us get a sense of the extent of bias that could occur from attrition. I now estimate the lower bound of CATE.

```
# find the intent-to-treat effect for urban and rural groups in the worst case
# urban first - estimate number of urbanites in treatment group
urban_treatment_worst <- worst_case_data %>%
  filter(urban == 1 & treatmentattempt == 1) %>%
  count()
urban_treatment_worst
```

```
##      n
## 1 5007
```

```
# estimate number of urbanites in control group
urban_control_worst <- worst_case_data %>%
  filter(urban == 1 & treatmentattempt == 0) %>%
  count()
urban_control_worst
```

```
##       n
## 1 19868
```

```
# find average turnout for control and treatment groups
urban_turnout_treatment_worst <- worst_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & turnout == 1) %>%
  count()/urban_treatment_worst

urban_turnout_control_worst <- worst_case_data %>%
  filter(urban == 1 & treatmentattempt == 0 & turnout == 1) %>%
  count()/urban_control_worst

# find the average difference in turnout for treatment and control groups
ITT_urban_worst <- urban_turnout_treatment_worst - urban_turnout_control_worst
ITT_urban_worst
```

```
##            n
## 1 0.03514809
```

```r
# non-urban second - estimate number of non-urban residents in treatment group
nonurban_treatment_worst <- worst_case_data %>%
  filter(urban == 0 & treatmentattempt == 1) %>%
  count()
nonurban_treatment_worst
```

```
##       n
## 1 20092
```

```r
# estimate number of non-urban residents in treatment group
nonurban_control_worst <- worst_case_data %>%
  filter(urban == 0 & treatmentattempt == 0) %>%
  count()
nonurban_control_worst
```

```
##      n
## 1 5033
```

```r
# find average turnout for control and treatment groups
nonurban_turnout_treatment_worst <- worst_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & turnout == 1) %>%
  count()/nonurban_treatment_worst

nonurban_turnout_control_worst <- worst_case_data %>%
  filter(urban == 0 & treatmentattempt == 0 & turnout == 1) %>%
  count()/nonurban_control_worst

# find the average difference in turnout for treatment and control groups
ITT_nonurban_worst <- nonurban_turnout_treatment_worst - nonurban_turnout_control_worst
ITT_nonurban_worst
```

```
##            n
## 1 0.02779387
```

```r
# find the proportion of compliers for both groups
# urban first - estimate proportion of compliers and always-takers in treatment group
urban_compliers_worst <- worst_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/urban_treatment_worst
urban_compliers_worst
```

```
##           n
## 1 0.7487517
```

```r
# non-urban second - estimate proportion of compliers and always-takers in treatment group
nonurban_compliers_worst <- worst_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/nonurban_treatment_worst
nonurban_compliers_worst
```

```
##           n
## 1 0.7738901
```

```r
# implement the Wald Estimator by dividing the ITT by proportion of compliers
# urban first - divide the ITT by proportion of urban compliers
urban_CATE_worst <- ITT_urban_worst/urban_compliers_worst
urban_CATE_worst
```

```
##            n
## 1 0.04694225
```

```r
# non-urban second - divide the ITT by proportion of non-urban compliers
nonurban_CATE_worst <- ITT_nonurban_worst/nonurban_compliers_worst
nonurban_CATE_worst
```

```
##           n
## 1 0.0359145
```

```r
# weight average the two unbiased estimates of CATE
final_CATE_worst <- prop_urban * urban_CATE_worst + prop_nonurban * nonurban_CATE_worst
final_CATE_worst
```

```
##            n
## 1 0.04140081
```

The resulting overall CATE is .0414. Assuming that all attrited subjects in the control voted and treatment did not, then compliers were 4.14% more likely to vote after being assigned to treatment. Notably, this estimate is not very far off from the estimate found in part c, which suggests the negative bias potentially created by attrition would not be huge. This bounding procedures gives us a useful sense of how bad bias could be in the negative direction for our CATE. I now estimate the upper bound of positive bias by finding the CATE for the best case scenario.

```r
# find the intent-to-treat effect for urban and rural groups in the best case
# urban first - estimate number of urbanites in treatment group
urban_treatment_best <- best_case_data %>%
  filter(urban == 1 & treatmentattempt == 1) %>%
  count()
urban_treatment_best
```

```
##      n
## 1 5007
```

```r
# estimate number of urbanites in control group
urban_control_best <- best_case_data %>%
  filter(urban == 1 & treatmentattempt == 0) %>%
  count()
urban_control_best
```

```
##       n
## 1 19868
```

```r
# find average turnout for control and treatment groups
urban_turnout_treatment_best <- best_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & turnout == 1) %>%
  count()/urban_treatment_best

urban_turnout_control_best <- best_case_data %>%
  filter(urban == 1 & treatmentattempt == 0 & turnout == 1) %>%
  count()/urban_control_best

# find the average difference in turnout for treatment and control groups
ITT_urban_best <- urban_turnout_treatment_best - urban_turnout_control_best
ITT_urban_best
```

```
##            n
## 1 0.04647814
```

```r
# non-urban second - estimate number of non-urban residents in treatment group
nonurban_treatment_best <- best_case_data %>%
  filter(urban == 0 & treatmentattempt == 1) %>%
  count()
nonurban_treatment_best
```

```
##       n
## 1 20092
```

```r
# estimate number of non-urban residents in treatment group
nonurban_control_best <- best_case_data %>%
  filter(urban == 0 & treatmentattempt == 0) %>%
  count()
nonurban_control_best
```

```
##      n
## 1 5033
```

```r
# find average turnout for control and treatment groups
nonurban_turnout_treatment_best <- best_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & turnout == 1) %>%
  count()/nonurban_treatment_best

nonurban_turnout_control_best <- best_case_data %>%
  filter(urban == 0 & treatmentattempt == 0 & turnout == 1) %>%
  count()/nonurban_control_best

# find the average difference in turnout for treatment and control groups
ITT_nonurban_best <- nonurban_turnout_treatment_best - nonurban_turnout_control_best
ITT_nonurban_best
```

```
##            n
## 1 0.03743917
```

```r
# find the proportion of compliers for both groups
# urban first - estimate proportion of compliers and always-takers in treatment group
urban_compliers_best <- best_case_data %>%
  filter(urban == 1 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/urban_treatment_best
urban_compliers_best
```

```
##           n
## 1 0.7487517
```

```r
# non-urban second - estimate proportion of compliers and always-takers in treatment group
nonurban_compliers_best <- best_case_data %>%
  filter(urban == 0 & treatmentattempt == 1 & successfultreatment == 1) %>%
  count()/nonurban_treatment_best
nonurban_compliers_best
```

```
##           n
## 1 0.7738901
```

```r
# implement the Wald Estimator by dividing the ITT by proportion of compliers
# urban first - divide the ITT by proportion of urban compliers
urban_CATE_best <- ITT_urban_best/urban_compliers_best
urban_CATE_best
```

```
##            n
## 1 0.06207416
```

```r
# non-urban second - divide the ITT by proportion of non-urban compliers
nonurban_CATE_best <- ITT_nonurban_best/nonurban_compliers_best
nonurban_CATE_best
```

```
##           n
## 1 0.0483779
```

```r
# weight average the two unbiased estimates of CATE
final_CATE_best <- prop_urban * urban_CATE_best + prop_nonurban * nonurban_CATE_best
final_CATE_best
```

```
##            n
## 1 0.05519179
```

The resulting overall CATE is .0552. Assuming that all attrited subjects in the control did not vote and in the treatment did, then compliers were 5.52% more likely to vote after being assigned to treatment. Notably, this estimate is also not very far off from the estimate found in part c, which suggests the positive bias potentially created by attrition would not be huge. This bounding procedures gives us a useful sense of how bad bias could be in the positive direction for our CATE.