What is the Effect of the Death of the Leader on the Level of Democracy? Part III: Controlling for Confounders

Let's continue to work with the data on assassinations attempts against political leaders from 1875 to 2004. The dataset is in a file called "leaders.csv". Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is assassination attempts.

variable	description
year	year of the assassination attempt
country	name of the country where the assassination attempt took place
leadername	name of the leader whose life was at risk in the assassination attempt
died	whether the leader died as a result of the assassination attempt: 1=yes, 0=no
politybefore	polity scores of the country where the assassination attempt took place before the assassination attempt (in points, in a scale from -10 to 10)
polityafter	polity scores of the country where the assassination attempt took place after the assassination attempt (in points, in a scale from -10 to 10)

Table 1: Variables in "leaders.csv"

In this problem set, we will revisit the analysis we performed in *What is the Effect of the Death of the Leader on the Level of Democracy? Part II* in view of what we learned in *Part I*.

In this problem set, we practice estimating causal effects with observational data by fitting a multiple linear regression model where Y is the outcome variable of interest, X_1 is the treatment variable, and X_2 is the confounding variable we worry about.

As always, we start by loading and looking at the data:

```
## load and look at the data
leaders <- read.csv("leaders.csv") # reads and stores data
head(leaders) # shows first observations
##   year      country       leadername died politybefore polityafter
## 1 1929 Afghanistan Habibullah Ghazi    0           -6   -6.000000
## 2 1933 Afghanistan       Nadir Shah    1           -6   -7.333333
## 3 1934 Afghanistan      Hashim Khan    0           -6   -8.000000
## 4 1924      Albania             Zogu    0            0   -9.000000
## 5 1931      Albania             Zogu    0           -9   -9.000000
## 6 1968      Algeria       Boumedienne    0           -9   -9.000000
```

1. Let's start by quickly replicating the results from Part II: Use the lm() function to fit a line to the data and summarize the relationship between X and Y and answer the question: What is the average causal effect of the death of a leader on the polity scores of a country? (Please write a full sentence answering the question, including the assumption, why the assumption might be reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (5 points)
2. Note that, as stated in our conclusion statement above, the validity of the estimate we arrived at in Part II depends on whether the assumption we made is correct (that is, on whether the assassination attempts where the leader ended up dying are comparable to the assassination attempts where the leader ended up surviving). Given your answers to questions 4 and 5 in Part I, do you think the assumption is correct? (5 points)
3. Given what you learned in Part I about (1) the relationship between *politybefore* and *died* (questions 4 and 5) and (2) the relationship between *politybefore* and *polityafter* (questions 6 and 7), do you think *politybefore* is a confounder? Please provide your reasoning. (10 points)
4. Now, let's estimate the average causal effect of *died* on *polityafter*, while controlling for *politybefore*. Start by fitting the following linear model with R:

$$\widehat{polityafter} = \widehat{\alpha} + \widehat{\beta}_1\ died + \widehat{\beta}_2\ politybefore$$

   (Hint: To fit a multiple linear regression model in R, we can use the function lm() and specify as the main required argument a formula of the type $Y \sim X_1 + X_2 + ... + X_p$.)

   What is the fitted model? (5 points)
5. Assuming *politybefore* is the only confounding variable present, please provide a full substantive interpretation of $\widehat{\beta}_2$ (including the unit of measurement). (10 points)
6. Assuming *politybefore* is the only confounding variable present, what is the average causal effect of the death of a leader on the polity scores of a country? (Please write a full sentence answering the question, including the assumption, why the assumption might be reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (10 points)
7. Does the size of the average treatment effect change significantly as a result of controlling for the confounder *politybefore*? A yes/no answer will suffice. (5 points)