What is the Effect of the Death of the Leader on the Level of Democracy? Part I: Visualizations and Correlations (with Solutions)

(Based on Benjamin F. Jones and Benjamin A. Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics*, 1 (2): 55-87.)

There is a longstanding debate in the study of international relations on whether individual political leaders make a difference. To explore this issue, *What is the Effect of the Death of the Leader on the Level of Democracy? Part II*, we will estimate the causal effect of the death of the leader on the level of democracy of a country. For this purpose, we will analyze data on assassination attempts against political leaders from 1875 to 2004.

To measure the level of democracy of the country, we will use polity scores. Polity scores categorize the regime of a country on a 21-point scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy). The Polity Project has produced polity scores for all countries from 1800 and on. For example, here are the 2018 polity scores.

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 practice computing the number of observations in a dataset, creating table of frequencies, computing and interpreting means, creating density histograms, subsetting variables, creating scatter plots, and computing correlation coefficients.

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. How many assassination attempts are recorded in this dataset? (5 points)

   R code:

   ```
   dim(leaders) # provides dimensions of dataframe: rows, columns
   ## [1] 250   6
   ```

   (Recall: In R, the function dim() provides (1) the number of observations and (2) the number of variables in a dataset (in this order). This function requires that you specify inside the parentheses the name of the object where the dataset is stored, *leaders* in this case.)

   Answer: There are 250 assassination attempts recorded in this dataset.

2. How many of the assassination attempts in the dataset ended up with the leader dead? (Hint: The function table() might be helpful here.) (5 points)

   R code:

   ```
   table(leaders$died) # creates table of frequencies
   ##
   ##   0   1
   ## 196  54
   ```

   (Recall: In R, the function table() creates the table of frequencies of a variable, displaying the values the variable takes as well as their frequency. This function requires that you specify inside the parentheses the name of the variable, *leaders$died*, in this case.)

   Answer: Out of the 250 assassination attempts in the dataset, 54 ended up with the leader dead. (Note: The variable *died* is a binary variable that identifies the assassination attempts that were successful, that is, that ended with the leader dead.)

3. What is the success rate of assassination attempts? In other words, what proportion of assassination attempts ended up with the leader dead? Please answer this question by using the function mean(). (5 points)

   R code:

   ```
   mean(leaders$died) # calculates the mean of died
   ## [1] 0.216
   ```

(Recall: In R, the function mean() computes the mean of a variable. This function requires that you specify inside the parentheses the name of the variable, *leaders$died*, in this case.)

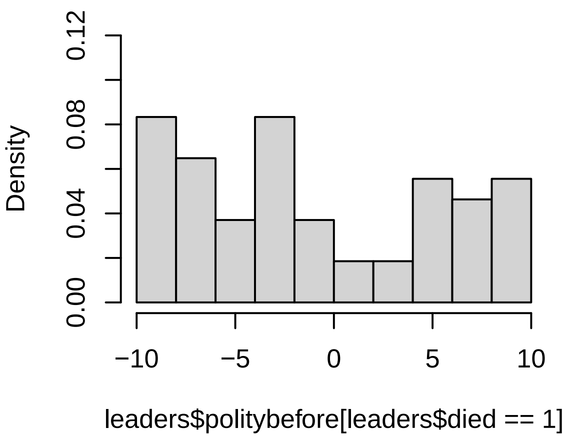
Answer: About 22% of the assassination attempts are successful, that is, end up with the leader dead. (Note: Since *died* is binary, the average of *died* can be interpreted as the proportion of the observations that have the characteristic identified by the variable (after multiplying the output by 100). In this case, then, the average of *died* can be interpreted as the proportion of assassination attempts that were successful (after multiplying the output by 100). There are many other ways we could have arrived to the same conclusion. For example, we could just have divided 54 by 250 or run prop.table(table(leaders$died)) to create the table of proportions of *died*.)

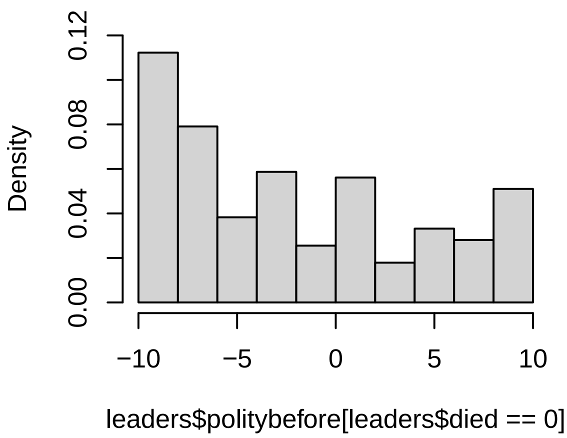
4. Do we observe any differences in the level of democracy *before* the assassination attempts took place between successful and unsuccessful assassination attempts? To answer this question, let's start by creating (1) the density histogram of *politybefore* for successful assassination attempts and (2) the density histogram of *politybefore* for unsuccessful assassination attempts. (Hint: The [] and == operators might be helpful when subsetting the variables here.) Do the two distributions look identical to each other? For example, do we find in each group the same proportion of *politybefore* equal to -10? (10 points)

   (Recall: When comparing distributions with different number of observations, it is best to create density histograms. To ask R to create a density histogram, we use the function hist() and set the optional argument freq to equal FALSE. Optional: To be able to compare the two histograms more easily, try specifying the following two optional arguments inside the hist() functions: breaks=10, and ylim=c(0,0.12). The first optional argument asks R to use 10 bins in the histogram. The second optional argument asks R to set the Y-axis to go from 0 to 0.12. Recall: In R functions, if we want to specify multiple arguments inside the parentheses, we should separate them with a comma.)

   R code:

```
## create histograms of polity scores before the assassination attempt took place
hist(leaders$politybefore[leaders$died==1],
     freq=FALSE, breaks=10, ylim=c(0,0.12)) # for successful assassination attempts
hist(leaders$politybefore[leaders$died==0],
     freq=FALSE, breaks=10, ylim=c(0,0.12)) # for unsuccessful assassination attempts
```





(Recall: The only required argument of the function hist() is the code identifying the variable we want to create the histogram of, which is *leaders$politybefore* in this case. Also recall that we use [] to subset a variable; inside the square brackets, we specify the criterion of selection. For example, we can use the relational operator == to set a logical test; only the observations for which the logical test are true will be extracted. In the first piece of code above, we extract from *politybefore* the observations that belong to successful assassination attempts (that is, for which *died* equals 1) and then create the histogram of those observations. In the second piece of code above, we extract from *politybefore* the observations that belong to unsuccessful assassination attempts (that is, for which *died* equals 0) and then create the histogram of those observations.)

Answer: No, the two distributions do not look identical to each other. For example, a smaller proportion of successful assassination attempts have a value of *politybefore* of -10. (Recall: Whenever the bins are of the same width as is the case here, the relative height of the bins in density histograms reflect the relative proportion of observations that fall in the bins.)

5. To further compare the two distributions above, compute the mean of both. Before the assassination attempt took place, were countries where the assassination attempt ended up being successful, on average, slightly more democratic/slightly less democratic/had the exact same level of democracy as compared to countries where the assassination attempt ended up not being successful? (5 points)

   R code:

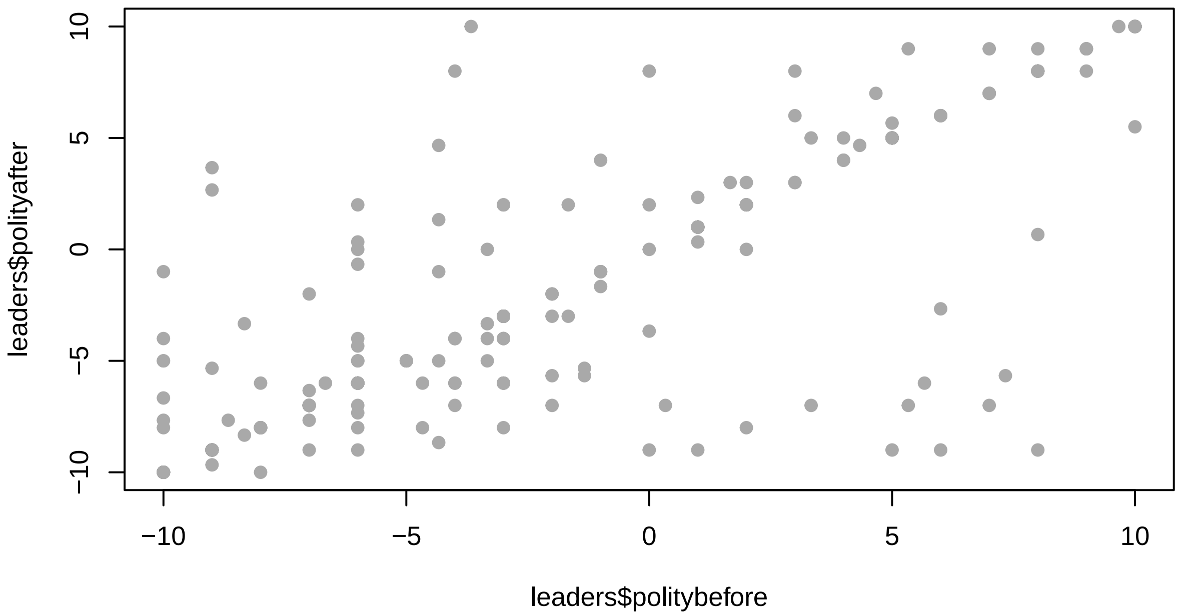
```
## compute means of politybefore
mean(leaders$politybefore[leaders$died==1]) # for successful assassination attempts
## [1] -0.7037037
mean(leaders$politybefore[leaders$died==0]) # for unsuccessful assassination attempts
## [1] -1.743197
```

   Answers: Countries where the assassination attempt ended up being successful were, on average, slightly more democratic to begin with (that is, before the assassination took place) than countries where the assassination attempt ended up not being successful. (Note: The average level of democracy before the assassination attempts is less negative for successful assassination attempts than for unsuccessful assassination attempts.)

6. Now, let's explore the relationship between *politybefore* and *polityafter* by creating the scatter plot between them. Make sure to place *politybefore* in the X axis and *polityafter* in the Y axis. Does the relationship look somewhat linear? (10 points)

   R code:

```
## create scatter plot
plot(leaders$politybefore, leaders$polityafter) # or
plot(x=leaders$politybefore, y=leaders$polityafter) # or
plot(y=leaders$polityafter, x=leaders$politybefore)
```



(Recall: the scatter plot of two variables shows how the two variables relate to each other. The function in R to create a scatter plot is plot(). The only two required arguments are the code identifying the two variables (first the X variable and then the Y variable, unless you use the name of the arguments x and y, in which case the order no longer matters).)

Answer: Yes, the relationship between *politybefore* and *polityafter* looks somewhat linear. (Note: There are a fair amount of observations exactly at the 45 degree angle line, which is where the observations would be when the value of *polityafter* is the exact same value as *politybefore*. This indicates that there are a fair amount of countries that had the exact same level of democracy before and after the assassination attempt, making the relationship between these two variables quite linear.)

7. To further explore the strength of the linear association between *politybefore* and *polityafter*, compute the correlation coefficient between them. Are these two variables moderately to highly correlated with each other? A yes/no answer will suffice. (10 points)

R code:

```
## compute correlation
cor(leaders$politybefore, leaders$polityafter)
## [1] 0.8283237
```

(Recall: The function in R to compute a correlation coefficient is cor(). The only two required arguments are the code identifying the two variables. The order of the variables does not matter since the correlation between X and Y is the same as the correlation between Y and X.)

Answer: Yes, these variables are moderately to highly correlated with each other. (Note: The correlation coefficient is 0.83, which is way closer to 1 than to 0.)