This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

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

(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
```

This material was produced for instructors using Llaudet, Elena and Kosuke Imai.
Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press)
and should not be shared beyond those who are enrolled in this class.

```
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)
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)
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)
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.)
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)
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)
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)