

## Do Women Promote Different Policies than Men?

### Part IV: Visualizations and Correlations

Let's continue working with the data from the experiment in India. As a reminder, Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is villages.

variable	description
<i>village</i>	village identifier ("Gram Panchayat number _ village number")
<i>female</i>	whether village was assigned a female politician: 1=yes, 0=no
<i>water</i>	number of new (or repaired) drinking water facilities in the village since random assignment
<i>irrigation</i>	number of new (or repaired) irrigation facilities in the village since random assignment

Table 1: Variables in "india.csv"

In this problem set, we practice (1) how to create and make sense of visualizations and (2) how to compute and interpret the correlation between two variables, among other things.

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

```
## load and look at the data
india <- read.csv("india.csv") # reads and stores data
head(india) # shows first observations
##      village female water  irrigation
## 1 GP1_village2      1    10           0
## 2 GP1_village1      1     0           5
## 3 GP2_village2      1     2           2
## 4 GP2_village1      1    31           4
## 5 GP3_village2      0     0           0
## 6 GP3_village1      0     0           0
```

1. Create a visualization of the distribution of the variable *water*.
  - a. Does this variable look normally distributed? (5 points)
  - b. Approximately how many villages in this experiment had about 250 new (or repaired) drinking water facilities since the randomization of politicians? (5 points)
2. Create a visualization of the relationship between *water* and *irrigation*.
  - a. Does the linear relationship between these two variables look positive or negative? A positive/negative answer will suffice. (5 points)
  - b. Does the relationship between these two variables look strongly linear? A yes/no answer will suffice. (5 points)
3. Compute the correlation between *water* and *irrigation*.

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.

- a. Are you surprised by the sign of the correlation? Provide your reason. (5 points)
- b. And are you surprised by the absolute value of the correlation? Provide your reason. (5 points)
4. If we wanted to use the sample of villages in this dataset to infer the characteristics of all villages in India, we would have to make sure that the sample is \_\_\_\_\_ of the population. (Please provide the missing word). (10 points)
5. What would have been the best way of selecting the villages for the sample to ensure that the statement above was true? (10 points)