Do Women Promote Different Policies than Men? Part I: Loading and Making Sense of Data

(Based on Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409–43.)

In a few problem sets, we will estimate the average causal effect of having a female politician on two different policy outcomes. For this purpose, we will analyze data from an experiment conducted in India, where villages were randomly assigned to have a female council head. The dataset we will use is in a file called "india.csv". Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is villages.

variable	description
village	village identifier ("Gram Panchayat number _ village number")
female	whether village was assigned a female politician: 1=yes, 0=no
water	number of new (or repaired) drinking water facilities in the village since random assignment
irrigation	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 how to load and make sense of data.

1. Use the function read.csv() to read the CSV file "india.csv" and use the assignment operator <- to store the data in an object called *india*. (Do not forget to set the working directory first.) Provide the R code you used (without the output). (10 points)
2. Use the function head() to view the first few observations of the dataset. Provide the R code you used (without the output). (10 points)
3. What does each observation in this dataset represent? (5 points)
4. Please substantively interpret the first observation in the dataset. (5 points)
5. For each variable in the dataset, please identify the type of variable (character vs. numeric binary vs. numeric non-binary) (10 points)
6. How many observations are in the dataset? In other words, how many villages were part of this experiment? (Hint: the function dim() might be helpful here.) Provide the R code you used (without the output) and provide the substantive answer. (10 points)