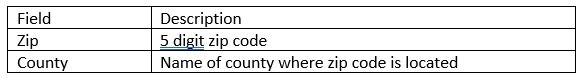
QUESTION 1

Write an R script that transforms the raw data file described and attached below into a tidy data set, following the instructions provided. Steps 8 & 9 also ask you to create a second data frame summarizing the data. When you are finished, save your R script and upload using the 'Browse Local Files' button below. I should be able to run your script start to finish and produce the two output tables shown.

Raw data file as csv

Raw data description: Number of responses to survey questions by question, survey and demographic codes. Demographic groups are broken down by gender (m, f) and zip code.

Instructions:



1. Load survey\_data.csv into R.

2. Subset data to include only surveys in 2018.

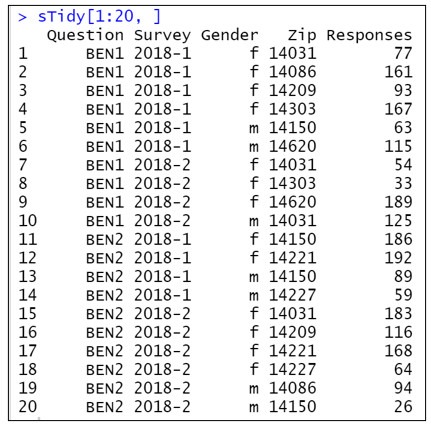
3. Remove column new\_sp from data.

4. Melt data to a long data set, using “Question” and “Survey” for id variables and removing NA values from data.

5. Create “Gender” column from the demographic variable (m or f).

6. Create “Zip” column from the demographic variable (five digit number)

7. Order the columns: Question, Survey, Gender, Zip, Responses. Sort data by Question, Survey, Gender, Zip.



Data description: Cross reference table for zip codes (5 digit number) and county name

8. Load survey\_state\_lookup.csv into R. Merge survey\_state\_lookup with your tidy data set, using zip code as the key field.

9. Create new summary data frame called s\_summary. s\_summary should have one row for each question and county and 3 data columns: male responses, female responses, and total responses (total across all surveys and genders).

