GH 811 PROBLEM SET 2

Background

Your analyses of the Ghana Demographic and Health Survey generated important insights into cooking patterns in that country. Soon after publishing your findings in the prestigious medical journal *The Lancet*, you received a telephone call from the Malawian Minister of Health expressing interest in your work and inviting you to collaborate with her team on a follow-up investigation in Malawi. The Minister expresses dismay that uptake of improved wood stoves remains so low despite efforts made by the Ministry and various NGOs to disseminate improved stoves. She tells you her team is about to launch a new initiative aimed at expanding access to and use of improved wood stoves in Malawi. For purposes of planning the initiative she asks you to investigate factors associated with current use of improved woods stoves. She says evidence on the characteristics of 'early adopters' will aid her team in designing a more effective and targeted intervention strategy.

Source of Data

Malawi Demographic and Health Survey 2010 (the dataset, final report and codebook are posted on Blackboard).

Data Processing

Your analytic dataset should consist of households that meet the following inclusion criteria:

a. Primary fuel used for cooking is wood
b. Household cooking is performed either in the main residence or a separate building, but not outdoors

Output

1.How many households were eliminated as a result of applying the stated inclusion criteria and how many remain in the final analytic dataset?

2.Using a nested ifelse statement, generate a new variable that records whether each household has an improved wood stove.* Drop from the dataset households for which it is not possible to determine whether there is an improved stove. How many households were eliminated as a result of this deletion? How many remain in the dataset? What is the percentage of households with an improved wood stove?

*Consider the stove an improved wood stove if any of the following criteria are met:

GH 811 PROBLEM SET 2

a. Food is cooked on an open stove
b. Food is cooked on a closed stove with chimney
c. Household has a chimney
d. Household has a hood

*Consider it not an improved stove if all the following criteria are met:

a. Food is cooked over an open fire
b. Household has neither chimney or hood

3.Compare characteristics of households that have an improved stove to those who don't have an improved stove. Include all the following characteristics in your table and three other relevant variables of your choosing. You may create the table in MS Excel if you prefer; however, the values of the variables must be estimated using R Studio.

a. Mean value of wealth index
b. Percent of households with wealth index of 'poorest' or 'poorer'
c. Percent of households from the southern region of Malawi
d. Percent of households that are urban
e. Percent of households that have a bednet for sleeping

4.Describe the pattern of findings in the table.

5.Which district(var_name: shdist) of Malawi has the highest prevalence of improved wood stoves (note: districts are a separate and smaller administrative unit than regions)? What is the percentage in that district? Why might this district have a higher prevalence of improved wood stoves than other districts in Malawi?

6.Restrict the districts to those that have a prevalence of improved wood stoves greater than the median and represent the prevalence in these districts with a barchart. Make sure to include appropriate labels, a title, and appropriate y-axis limit.

7.Use a boxplot to show the distribution of education (in units of single years) of the household head comparing households that have an improved wood stove to those who do not. Explain. Hints: (1) Household head is assigned line number 1 in the survey. Thus, the relevant variable will be of the form hvYYY_01. (2) Exclude outliers prior to making your plot.

8.Using a for loop, generate the proportion of improved stoves within each wealth status category. Create a barchart to display this information. Include a title, labels for both axes, correct names for the bars, and an appropriate y-axis limit.

*Submit answers to the above questions along with your commented R code attached to the end of your word document by **2 PM October 31, 2017.***

GH 811 PROBLEM SET 2

Hints:

**The use of the dplyr package is highly suggested for this problem set

. 1) Try performing an operation using the dim() function.

. 2) Keep in mind the definition for improved stove involves values across multiple variables. A nested ifelse should be used when generating the improved wood stove variable. (look at codebook for missing value)

. 3) *do regular way then try dplyr (look forward to module 5)

   a. Wealth index is coded as text and must be recoded with numeric values to proceed (remember this is ordinal when recoding)
   b. May help to create an indicator variable
   c. May help to create a dummy variable
   d. If using dplyr may help to a dummy variable
   e. No hints

. 4) .

. 5) You can do this several ways, but the easiest might be using dplyr

. 6) Looking at ?plot or ?barplot in r can help with arguments and options for visuals (use ggplot if you like)

. 7) Make sure to deal with missing values before creating boxplots. Use ?boxplot in r for information on syntax. x~y ...

. 8) Using a for loop here is required.