1. Read in all three data sets.

2. Merge the three data sets, keeping only those observations that were in all three data sets.

3. Subset the data to include only those people 45 years and older at time of screening.

4. Omit any variables not included in the data dictionary.

5. Rename three variables to make it easier for you to know what the variables represent.

6. Clean the data.

7. How many missing values do each of the variables have?

8. Combine Mexican Americans and Other Hispanics into one category.

9. Categorize age into 3 categories – 45 to 60, 61 to 75, and 75 and over.

10. Find the mean age by sex.

11. Create a cross table of age (as categorized above) and new ethnicity.

**demo**

seqn identification number

riagendr sex 1 = male 2 = female

ridageyr age in years at screening 0-79, 80 = 80 years and older

ridreth1 ethnic identity 1 = Mexican American

2 = Other Hispanic

3 = Non-Hispanic White

4 = Non-Hispanic Black

5 = Other (including Multi)

**medcond**

seqn identification number

mcq010 ever told you have asthma 1 = yes, 2 = no

mcq082 ever told have celiac disease 1 = yes, 2 = no

mcq220 ever told you have cancer or malignancy 1 = yes, 2 = no

**phyfunc**

seqn identification number

pfq059 physical, mental, emotional limitations 1 = yes, 2 = no

pfq049 limitations keeping you from working 1 = yes, 2 = no

pfq061c walking up ten steps difficult 1 = no difficulty

2 = some difficulty

3 = much difficulty

4 = unable to do

5 = do not do this activity