For this assignment, you'll be working with the county level dataset to predict a home ownership rates using conditional means. You'll need to select the county-level characteristics that you think might be related to home ownership rates. Please complete the following steps:

Hint: Watch ALL of the video lectures prior to attempting the HW. Will covers examples very similar to those you are asked to complete here.

1. Calculate the mean of the outcome.

Translation "outcome": calculate the mean of your dependent variable, <u>home ownership rates.</u>

Review the RMD file for week 2 and step through and execute the code. Identify the correct syntax for computing the mean.

2. Use your mean as a prediction: Create a new variable that consists of the mean of the outcome.

Translation "outcome": calculate the mean of your dependent variable, <u>home ownership rates.</u>

Hint: Use the ← operator.

3. Calculate a summary measure of the errors for each observation—the difference between your prediction and the outcome.

Hint: Find the use of the RMSE method in the week 2 RMD file.

4. Calculate the mean of the outcome at levels of a predictor variable.

The predictor variable here refers to the conditioned variable used for prediction. We can compute the mean of a variable conditioned on levels of another (conditioned) variable.

You should ask yourself:
- *What variable is being predicted? Hint: We will compute the mean of this variable*
- *What variable is the conditioned (predictor) variable? We will group or compute levels of this variable.*
- *What are the levels? These may be chosen by you the investigator, e.g. quartiles; or they may be implicit given the data, e.g. county name.*

See the following example in the RMD file:
" …
##Conditional Means With One Predictor Variable

To incorporate additional information into the mean, we need to calculate averages at levels of other predictors. Let's calculate the average per capita income at different levels of college education. The code below will calculate average income across counties at four different levels of college education-- the four quantiles of college education in the dataset.
…"

5. Use these conditional means as a prediction: for every county, use the conditional mean to provide a ''best guess" as to that county's level of the outcome.

6. Calculate a summary measure of the error in your predictions.
I expect that the .Rmd file you submit will run cleanly, and that there shouldn't be any errors. Use LOTS of comments to tell me what you are doing.