

You can download the data-set from this link (<https://goo.gl/KKzQKx>). The description of the columns are provided on this link (<https://goo.gl/cj1WW9>). You will need the **county name**, **state name** and the following columns:

**LFE305213**: Mean travel time to work (minutes)  
**HSG445213**: Homeownership rate, 2009-2013  
**INC110213**: Median household income, 2009-2013  
**POP010210**: Population, 2010  
**EDU685213**: Bachelor's degree or higher

**Step 0**: Create a class representing a County containing the properties we will need for each county to perform the necessary analysis (**20**). Read the csv file and create a list (or dictionary) of county objects (**20**). While reading the file, make sure to skip the rows where State (state\_abbreviation) column is empty. All the features listed below are to be performed on this list (or dictionary).

**Step 1(20)**: Display the least and most populous counties.

**Step 2(30)**: Display the states with least and most population.

**Step 3(40)**: We want to check if there is a correlation between happiness and the rate of higher education in counties.

A county's happiness rate is computed by multiplying the median household income and homeownership rate, and then dividing by the mean travel times to work -- (income \* homeownership) / commute. Make sure to write a method as part of the class you created in Step 0 that returns the happiness rating of the county. Whenever you need the happiness rating, you will call this method of the class to get it.

Now we can check for correlation between happiness and higher education. We want to do this simple. Find the average happiness rating among all the counties in US and then group the counties into two:

- G1- Those counties whose happiness rating is **above** the average
- G2- Those counties whose happiness rating is **below** the average

Then find the average higher education rate among counties in G1 (*avgHappy*). Do the same for counties in G2 (*avgUnhappy*). If the difference between *avgHappy* and *avgUnhappy* is more than 20, then the correlation is significant. If it is more than 5, there is slight correlation, otherwise we could not find any significant correlation. Your program will output **one** of the following appropriate messages below:

- a) Happy counties have significantly more (**or less**) college graduates.
- b) Happy counties have slightly more (**or less**) college graduates.
- c) Could not find any significant correlation.

Taking the words **more** or **less** into account, notice that your program may output 5 different messages.

**Step 4(40)**: Find and report all counties that are similar. Two counties are similar if all the information on them are within 2% of each other. So find all counties, whose properties (except county name and state) are within only %2 different.

**Step 5(30)**: Create a graph displaying correlation between population and income. As the population grows in a county, does the median income grow? Find the minimum and maximum populations (you found this in Step 1), divide all counties into 3 equal segments. For example, if the population of the least populous county is 100 and max populous is 1000, you will have 3 groups ((1000-100)/3=300 spacing): 100~400, 400~700, 700~1000. Within each group, find and plot the average income.

Q1: Counties with least and most population

=====

Loving County(TX): 82.0

Los Angeles County(CA): 9818605.0

Q2: States with least and most population

=====

WY: 563626.0

CA: 37253956.0

Q3: Happiness vs Higher Education Rate

=====

Happy Counties College Edu Rate : 23.50780559646543%

Unhappy Counties College Edu Rate: 16.912829131652664%

Difference is 6.594976464812767%.

Happy counties have slightly more college graduates.

Q4: Similar Counties (2%)

=====

Jackson County(MN) and Yellow Medicine County(MN) are similar.

Q4: Population vs Income Rate

=====

