

Project 7.2 Does education influence unemployment?

Conventional wisdom dictates that those with more education have an easier time finding a job. In this project, we will investigate this claim by analyzing 2018 data from the U.S. Census Bureau.²

Part 1: Get the data

Download the data file from the book website and look at it in a text editor. You will notice that it is a tab-separated file with one line per U.S. metropolitan area. In each line are 60 columns containing population data broken down by educational attainment and employment status. There are four educational attainment categories: less than high school graduate (i.e., no high school diploma), high school graduate, some college or associate's degree, and college graduate.

Economists typically define unemployment in terms of the “labor force,” people who are available for employment at any given time, whether or not they are actually employed. So we will define the unemployment rate to be the fraction of the (civilian) labor force that is unemployed. To compute the unemployment rate for each category of educational attainment, we will need data from the following ten columns of the file (column numbers start at 1):

Column	Contents
2	Name of metropolitan area
3	Total population of metropolitan area
11	No high school diploma, total in civilian labor force
15	No high school diploma, unemployed in civilian labor force
25	High school graduate, total in civilian labor force
29	High school graduate, unemployed in civilian labor force
39	Some college, total in civilian labor force
43	Some college, unemployed in civilian labor force
53	College graduate, total in civilian labor force
57	College graduate, unemployed in civilian labor force

Read this data from the file and store it in six lists that contain the names of the metropolitan areas, the total populations, and the unemployment rates for each of the four educational attainment categories. Each unemployment rate should be stored as a floating point value between 0 and 100, rounded to one decimal point. For example, 0.123456 should be stored as 12.3.

Next, use four calls to the `matplotlib` `plot` function to plot the unemployment rate data in a single figure with the metropolitan areas on the x -axis, and the unemployment rates for each educational attainment category on the y -axis. Be sure to label each plot and display a legend. You can place the names of the metropolitan areas on the x -axis with the `xticks` function:

²*Educational Attainment by Employment Status for the Population 25 to 64 Years* (Table B23006), U.S. Census Bureau, 2018.

P7.2-2 ■ Discovering Computer Science, Second Edition

```
plt.xticks(range(len(names)), names, rotation = 270, fontsize = 'small')
```

The first argument is the locations of the ticks on the x -axis, the second argument is a list of labels to place at those ticks, and the third and fourth arguments optionally rotate the text and change its size.

Part 2: Filter the data

Your plot should show data for 519 metropolitan areas. To make trends easier to discern, let's narrow the data down to the most populous areas. Create six new lists that contain the same data as the original six lists, but for only the thirty most populous metropolitan areas. Be sure to maintain the correct correspondence between values in the six lists. Generate the same plot as above, but for your six shorter lists.

Part 3: Analysis

Write a program to answer each of the following questions.

Question 7.2.1 *In which of the 30 metropolitan areas is the unemployment rate higher for HS graduates than for those without a HS diploma?*

Question 7.2.2 *Which of the 30 metropolitan areas have the highest and lowest unemployment rates for each of the four categories of educational attainment? Use a single loop to compute all of the answers, and do not use the built-in `min` and `max` functions.*

Question 7.2.3 *Which of the 30 metropolitan areas has the largest difference between the unemployment rates of college graduates and those with only a high school diploma? What is this difference?*

Question 7.2.4 *Print a formatted table that ranks the 30 metropolitan areas by the unemployment rates of their college graduates. (Hint: use a dictionary.)*