

## Dataset

The dataset `scp-1205.csv` contains information on Medicare Advantage (MA) enrollment by contract-county pair. The variables in the dataset are:

<code>countyname</code>	name of the county
<code>state</code>	state postal code
<code>healthplanname</code>	name of the health plan
<code>typeofplan</code>	type of health plan
<code>countysa</code>	Social Security Administration county code
<code>eligibles</code>	number of individuals in the county that are Medicare eligible
<code>enrollees</code>	number of individuals enrolled in the specific health plan
<code>penetration</code>	percent of individuals in the county enrolled in the plan, defined as $100 \times \text{enrollees}/\text{eligibles}$
<code>ABrate</code>	Medicare's monthly payments to the health plan

A row in the data corresponds to a contract-county pair. For example the row that begins

```
countyname  state  contracthealthplanname
AUTAUGA    AL    H0150          HEALTHSPRING OF ALABAMA INC.
```

indicates that there are 6883 eligibles in Autauga, AL and that 313 of them are enrollees in the Healthspring of Alabama plan.

County-plan pairs that have missing values for eligibles, enrollees, and penetration should be considered as having zero values for these variables.

## Objective

The goal of the exercise is to write a short script that produces a county-level dataset that identifies the number of plans and total enrollment in each county (you may exclude territories such as Puerto Rico and Guam). In particular, we would like you to produce a dataset that has following variables:

<code>countyname</code>	name of the county
<code>state</code>	state postal code
<code>numberofplans1</code>	number of health plans with <i>more than 10 enrollees</i>
<code>numberofplans2</code>	number of health plans with <i>penetration &gt; 0.5</i>
<code>countysa</code>	Social Security Administration county code
<code>eligibles</code>	number of individuals in the county that are Medicare eligible
<code>totalenrollees</code>	number of individuals in the county with a MA health plan
<code>totalpenetration</code>	percent of individuals in the county enrolled in a MA plan, defined as $100 \times \text{totalenrollees}/\text{eligibles}$

The dataset should include one observation for each county and should be sorted by state and then by county. If you need to make judgment calls about how to process the data, please write us a *short* note describing the decisions you made.

In evaluating the project we will not only look at whether the output file is correct, but also at: (i) whether the algorithm is efficient and concise, (ii) whether the code is clean, easy-to-read, and well-commented, and (iii) how quickly the code was produced.

So we can evaluate (iii), we ask that you *accurately record the total time spent on the project beginning when you first read through the instructions.*