
The following problems should be done in julia. Type up the code you used to find your results and answer all questions. Submit your homework via Blackboard by using any program you want and submitting the PDF version of it or an ipynb file. This assignment should be submitted no later than 11:59pm on November 29, 2016.

1. This problem uses the data at <http://www.census.gov/geo/maps-data/data/gazetteer2010.html> and download the data under the header *ZIP Code Tabulation Areas*. You should download the data, unzip it and it should be called: *Gaz_zcta-national.txt*. Note: this is a different file then we loaded in class, but it is tab delimited.
 - (a) Load the data into a DataFrame in julia and make sure that the data types are all correct.
 - (b) Find the total population of the U.S. according to this dataset.
 - (c) Find the top 5 zip code areas by population. List the zip codes and the population. Use the internet to look up the cities where the zip codes are.
 - (d) Create a new DataFrame with 2 columns, the first is the zip code and the 2nd is the population density, which is the population divided by the area (in square miles) within the zip code. List both the top 5 zip codes by density and the cities of the zip codes.
 - (e) Find the mean longitude and latitude and describe where that lies in the U.S. (Note: most mapping websites will map a longitude and latitude for you.) This is a way of measuring the center of the U.S. There are two alternative ways of measuring this.
 - (f) Before finding other centers of the U.S., we need to define a weighted mean. In short, a weighted mean is a mean where not every point has the same weight. For example, if we have the numbers 1, 4, 5, 8, 12, then the mean is 6. However if we have a vector of weights [1, 3, 3, 1, 4], then we consider find the mean of the numbers 1, 4, 4, 4, 5, 5, 5, 8, 12, 12, 12, 12, and taking the mean of these is 7. In mathematical notation, a weighted mean of the numbers in x with weights w , is $\bar{x}_w = \frac{\sum_i x_i w_i}{\sum w_i}$ or using julia notation if the vectors \mathbf{x} and \mathbf{w} have the same length then the weighted mean is


```
sum(x.*w)/sum(w)
```

 Find both the regular mean and the weighted mean of [1, 3, 5, 7, 9, 11, 13] with weights [1, 2, 1, 2, 1, 2, 3]
 - (g) Find the weighted mean of the longitude and latitude in the dataset taking the total area (land and water) as the weight. Find the town its nearest. This is called the geographical center of the U.S. Note: you can use either the area in meter or square miles. The result will be the same.
 - (h) Find the weighted mean of the longitude and latitude in the dataset taking the population as the weight. Find the town its nearest. This is called the population center of the U.S.
 - (i) Find the largest and smallest area codes by total land and water area. (Hint: you may need to make a new DataFrame with a column of total area.)
2. Use the Olympic Data from class to answer all of these questions. Load the data and put each column in a proper form. Note: as we saw in class, there are athletes in the data without names or ages. You should not include them in the following calculations. You may either throw out their results or not load them in the dataset to begin with.

Also, for the top 5 listed, if there is a tie for 5th, list all athletes with that value.

- (a) Find the 5 oldest and 5 youngest olympians in the data set and list each participant's name, age, sport and country. (Note: if there is a tie, list all participant's in the tie.)
- (b) Find the mean, median and standard deviation of the ages of the olympians.
- (c) Plot a histogram of the ages of all olympians.

- (d) Which athlete won the most gold medals in 2000? Most Silver medals in 2004, most bronze medals in 2002?
- (e) For this problem, combine the data over all years as we discussed in class. Find the top 5 gold, silver and bronze medal winners over all years in the dataset.
- (f) Combine the data by both Country Name and Year and then sum the total number of medals. Find the top 5 countries by total medals in each of the years 2000 through 2012? Again, if there are ties, list all all countries.