Finding a Convenient Conference Location

It's time for your organization to plan for a national conference. One of the most important decisions to make is where it's going to be held. Since your members live all over the United States, you have decided to find a location that will minimize everyone's aggregate distance to travel to the conference.

However, since you do not know in advance who is coming to the conference, how would you decide on a convenient location? Let's assume that the attendees are drawn at large from the U.S. population. I have a file that you can download that gives a summary of the U.S. population from the 2010 census.

http://cs.furman.edu/~chealy/census/block2010.txt.gz

After you uncompress this plain text file, you will notice that it's really big! Millions of lines, where each line describes the population and location of one block. A typical line of the input file looks like this:

```
AL Autauga County       021100 2001     2     0     2     0 32555 086804 192
```

The first 24 characters on the line give the state abbreviation and the name of a county. After the 24th character, the rest of the line consists of tokens, separated by spaces.

- Census tract number
- Census block number
- Total population
- 3 more tokens indicating something about the racial distribution of the population
- Latitude, in thousandths of a degree
- Longitude, in thousandths of a degree
- Area of the block in acres

The only information you need from each line is its total population, latitude and longitude.

You need to write a program that reads the entire file and grabs the population and location information for each block. But let's think for a moment on how we are going to use this massive information.

The ultimate objective of your program is to select a convenient location for a conference. Let's begin by creating a set of candidate locations. For each candidate location, we need to scan our list of blocks (either by having the information stored inside a data structure in our program or by reading the file again), in order to compute the weighted average distance that everyone in the United States would have to travel to get to your candidate location. And then this distance will be stored with this candidate location. Once you are done, you can print out all of these candidate locations along with their distances, as well as identifying the one location with the lowest average distance.

What are the candidate locations? Here is a suggestion: consider every whole degree of latitude and longitude across the United States. Let's assume that the conference will not be in Alaska or Hawaii. It doesn't take too much imagination to see why they would not be suitable. So, your program would go something like this:

For latitude = 25 to 49
  For longitude = 70 to 122
    avgDistance = calculate the aggregate average distance to (latitude, longitude)

And at the end of your program, you can print the results.

When you are done, you will have identified a point located somewhere in the interior of the United States. But it could be the middle of a cornfield. Look on a map (e.g. Google Earth) to find the nearest major city, one that is convenient for road and airline travel.

Based on your program's output, does your output make intuitive sense? Where do you recommend having the conference?