**CSE564 Final Project Prelim Report**

Xinhe Huang 108390641

Silu Ou 110563639

**Preparation**

In order to get a better understanding in the *gps.csv* file, we use the app ArcGIS to open the ESRI shapefile *Abila* and translate the map into a file named *location.csv*. This file records the gps of all the locations shown in the map. Then we can translate the records of latitude and longitude in the *gps.csv* file into corresponding locations. The ESRI shapefile *Abila* and the map are shown as below.



Abila



map

We compare these two files carefully and record the range of each location’s gps. Then we use the mean value of the range as the gps of one location.

**Data Analysis**

First we take a look at each csv file roughly. We find that *gps.csv* file records each car’s gps at every second as long as the car is moving. Thus we need to handle the data in the *gps.csv* file to get the start location and end location of each journey. Then for each car we calculate the frequency of its visit to each place. If one place has a frequency far below others, it means this car seldom visit this place, which is an unusual behavior. We also need to calculate the time period of each car’s journey since for the user of one car he/she should have a similar schedule for each day. If one day there is journey having a time period that is far from other days, it means that on this day the user has an unusual behavior.

For *cc\_data.csv* file we do the similar analysis work. But before we analyze the data, we first need to filter them. We find that some people only go to one or two place and have tractions there. It is not necessary to do analysis on these people since we can’t get a lot of useful information from their monotonous life. Thus we need delete these people’s records from the original data. Then we calculate the frequency of each person’s transaction at each location to find their unusual behavior. In addition, we calculate the total price of each person’s transaction at each location. Then we make comparison between each person to find if there is some difference.

**Charts/Graphs**