**Data Set Information:**

The dataset is composed by two tables. The first table go\_track\_tracks presents general attributes and each instance has one trajectory that is represented by the table   
go\_track\_trackspoints.

**Attribute Information:**

1. go\_track\_tracks.csv: a list of trajectories

id: trajectory的关键字  
id\_android - it represents the device used to capture the instance;   
speed - it represents the average speed (Km/H)   
distance - it represent the total distance (Km)   
rating - it is an evaluation parameter. Evaluation the traffic is a way to verify the volunteers perception about the traffic during the travel, in other words,   
if volunteers move to some place and face traffic jam, maybe they will evaluate 'bad'. (3- good, 2- normal, 1-bad).   
rating\_bus - it is other evaluation parameter. (1 - The amount of people inside the bus is little, 2 - The bus is not crowded, 3- The bus is crowded.   
rating\_weather - it is another evaluation parameter. ( 2- sunny, 1- raining).   
car\_or\_bus - (1 - car, 2-bus)   
linha - information about the bus that does the pathway   
  
  
  
(2) go\_track\_trackspoints.csv: localization points of each trajectory   
id: unique key to identify each point   
latitude: latitude from where the point is   
longitude: longitude from where the point is   
track\_id: identify the trajectory which the point belong ，与go\_track\_tracks.csv中id对应  
time: datetime when the point was collected (GMT-3)

题目要求：

1 通过对trajectory或point进行聚类，划分出样本为若干活动热区

聚类的数据点选取方法有两种方案，可选择任意一种：

1. 以单个trajectory为计算单位(以go\_track\_trackspoints.csv 中track\_id为关键字)，该方法需要计算trajectory的代表节点（计算方法自行决定）
2. 以trajectory中的每个point为计算单位

2 聚类方法采用密度聚类

3 把聚类结果以图形化方式展示出来

4 从空间和时间上分析运行效率

5 分析算法的优缺点，提出改进的方向

提交报告要求：

1 对题目要求1中算法

1. 写出伪代码和程序代码（标注主要过程及方法）
2. 如果数据点选取方法采用1），请写出计算方法，并提供部分选取的结果
3. 展示出20个数据对象的算法执行过程

2 写出数据的分类结果

3 数据划分的图形化展示结果

4 运行效率的分析过程

5 算法优缺点的分析，及改进意见

6 运行的交互界面

7 程序源码，并对主要过程和方法进行注释