**交通雾霾相关性分析小组工作进展报告(20160328)**

胡煜成，姚嘉豪

**目标：**探索交通与雾霾的相关性

**已完成工作：**

* AQI数据整理

为了和交通数据的匹配，我们使用的也是2014-01到2014-06一共185天的雾霾的数据，其中数据的来源来自青悦，总共35个采样点，其中每一个小时记录一次数据。

数据中有其他指标，比如SO2、NO2等等，但是我们初步只使用AQI的值。

数据之中有大量的缺失，比如说某一天的某一个小时没有记录，2014-4月的数据有大量存在缺失等等。但是，这些问题最后，和数据组商量之后，得到了比较好的解决。

雾霾的数据尤其困难，因为每个数据都是在一个cvx文件中的，关键在于如何从已有的文件中扣出自己所要的数据。

对于每一小时AQI指标的平均，我们是将35个检查点进行加合取平均的办法，同时，我们还做了这一小时AQI增量的数据。

* 交通数据整理

初步整理了从2014-01到2014-06一共185天的北京市路段微波数据。

数据是按每个月31天来算的，这是为了数据记录的方便，所以很多的NA,但是我们作图的时候使用办法消除了这样的问题。

数据一共包含北京市区约1500条路段。

数据包含每个路段的车流量和平均车速随时间变化的信息。

数据记录的时间步长为2分钟。

为了和雾霾数据进行匹配，因为交通的时间是以两分钟为界限的和雾霾是以1小时为界限的，所以我们将交通数据转换成一个小时的交通量，这样更加有利于数据的对比。

数据中含有缺失和错误记录。

数据中并没有我们需要的路段坐标信息。

姚嘉豪和张日月明将路段的坐标信息生成了出来。

为了得到GPS坐标的信息，从百度API上批量得到其坐标，但过程中使用了很多时间去钻研如何去批量操作。

* GPS坐标与地图坐标的制式与转换

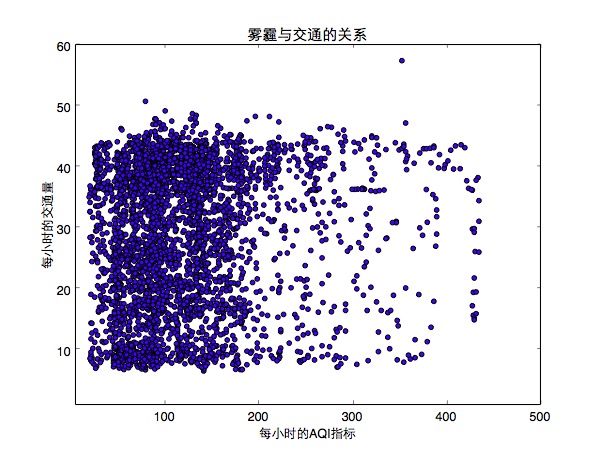
见张日月明报告

* 宏观数据分析

这只是一个宏观的分析，因为我们很多数据仅仅是全局的北京城的平均，比如说将交通进行的平均，以及将雾霾指数的平均，（尤其是雾霾的平均，没有考虑到地理位置的面积，仅仅是一个简单的平均，即忽略了权重的影响。），但是，我们只是希望在这简单的宏观之中，窥出一二端倪。故，作图如下：

平均车辆密度与平均AQI值之间的关系。

【图】

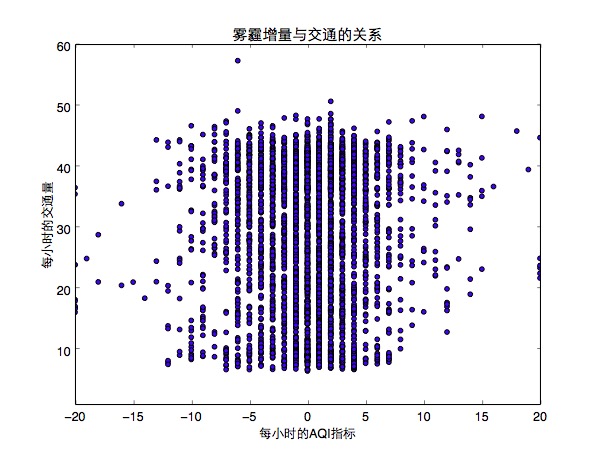


可以看出，这里北京的天气还是很好的，大量的集中在AQI指标比较小的地方，这也是比较合理的地方吧，但是也似乎可以看出当交通量越密集，我们的点越集中，于是交通确实对于空气是有一定的影响。

平均车辆密度与平均AQI增量之间的关系。

【图】

交通对于AQI指标的影响似乎，也就只是在20-30左右的样子，没有交通是对空气有影响，但是影响的程度没有那么的大吧。



结论：

**待完成工作：**

* 进一步分析北京市区交通总量与AQI平均水平的关系。
* 得到带有空间信息的交通数据。
* 结合时空性分析二者的相关性。

存在的问题：

我们对于数据的清理上花费了大量的时间，交通地点的GPS坐标的获取、AQI数据的处理、交通量数据的整合，都花上了很大的时间。这上面的效率有待进一步的提高。