# 

# **数据挖掘中期实验报告**

# 

|  |  |
| --- | --- |
| **组长** | **周晖** |
| **班级** | **计算机技术1班** |
| **组员** | **陈思宇、李鑫、王宏宇** |
| **指导教师** | **郑皎凌** |
| **所在学院** | **计算机** |
| **提交日期** | **2019年10月23日** |

2019 年 10 月

成都信息工程大学 计算机学院

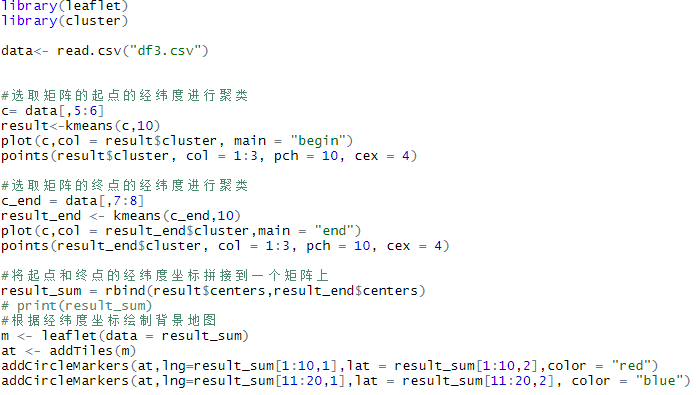
# 1.起点聚类与终点聚类

## 1.1 说明

起点和终点的聚类过程概述：

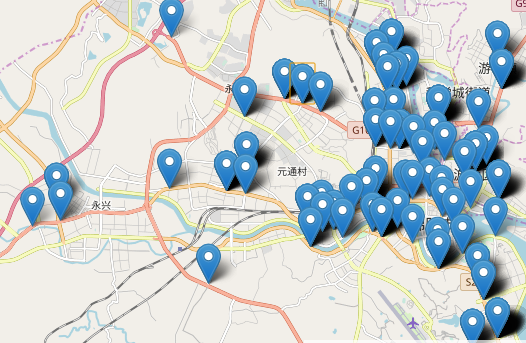
* 加载编程所需要的包，leaflet、cluster包
* 读取excel数据，并截取数据的第5、6列为起点数据，7、8列为终点数据
* 将起点数据和终点的数据放入系统自带的kmeans函数来进行聚类
* 将聚类后的结果通过散点图的方式可视化出来
* 通过leaflet包将聚类的中心点在地图上进行显示

## 1.2 代码展示

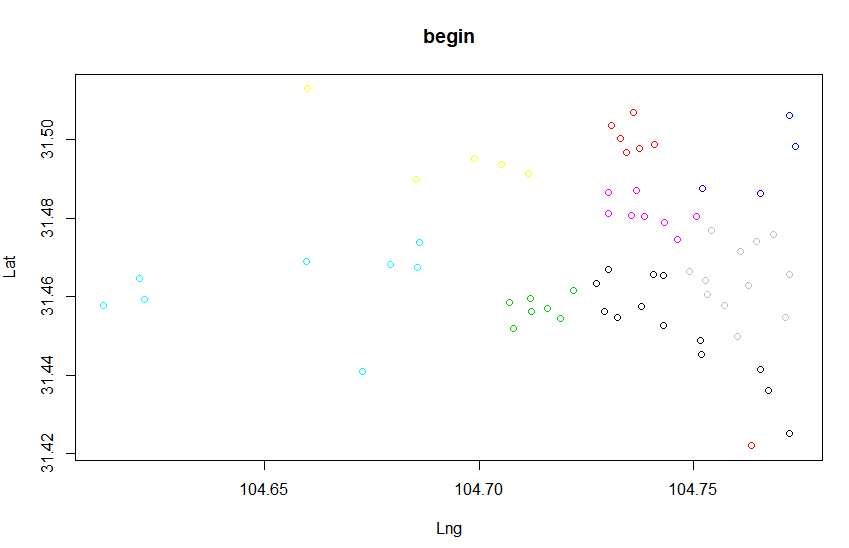


## 1.3效果图展示

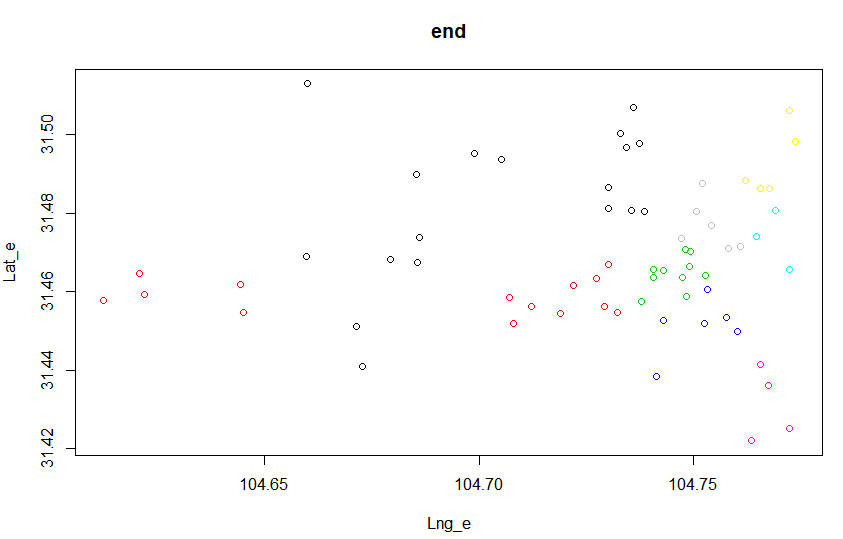
原始数据在地图上显示图：



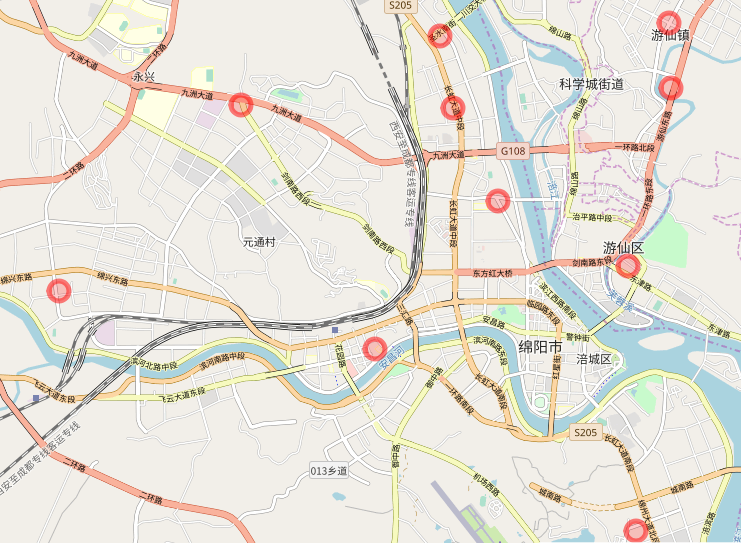
起点聚类散点图结果



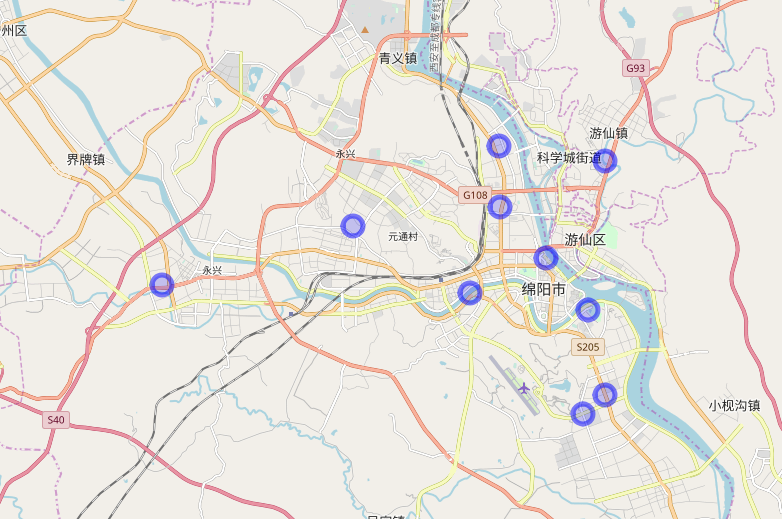
终点聚类散点图结果



根据起点的聚类中心点的实景图：



根据终点的聚类中心点绘制的实景图



# 2.OD线绘制

## 2.1 说明

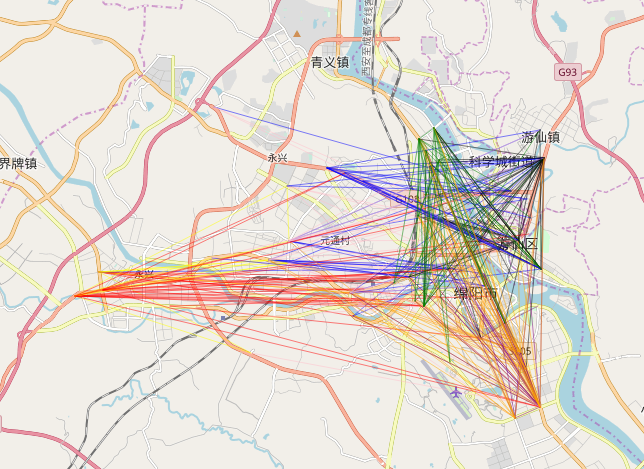
OD线绘制过程概述：

* 加载编程所需要的包，leaflet、cluster包
* 读取excel数据，并提取数据的第5、6列为起点数据，7、8列为终点数据
* 根据起点和终点的数据，用kmeans方法进行聚类
* 通过leaflet包绘制连线图并在地图上进行显示

## 2.2 代码展示



## 2.3 效果图展示



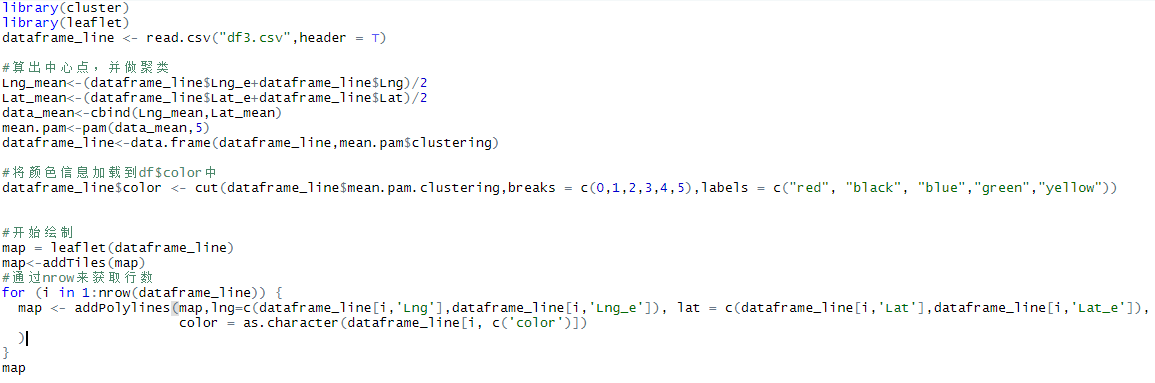
# 3.OD线聚类

## 3.1 说明

OD线聚类过程概述：

* 计算出连线的中心点的值。
* 将中心点的值用R语言自带的pam算法进行聚类
* 将颜色信息加入data\_frame中
* 进行绘图

## 3.2 代码展示



## 3.3 效果图展示

# 4.总结

起点聚类与终点聚类设置了10个簇，并且将聚类后的中心点结果在地图上进行显示，能直观看到各个点所在的区域。

OD线聚类使用线的中心点来做聚类的方法，该方法比较简洁快速，但是总体对线的聚类的效果不是太好，因此后续需要继续改进。

在本次实验中学习到了Rstudio工具的使用，了解了R语言强大的统计学特性，便于后续做自己的研究课题的时候可以借助R语言来处理数据，实现更好的效果。