# 数据挖掘第一次实验报告

# 

# 班级 软工学硕1班

# 小组成员 杨蕾、何瑞、宋琼成

# 靳含、卢晓晓、Meltem Kaplan

**任帅龙**

# 1.起点聚类

### 1.1代码

library(leaflet)

dataAll<- read.csv("df3.csv")

lng<-dataAll[,5]

lat<-dataAll[,6]

start\_lat\_lng<-dataAll[,5:6]

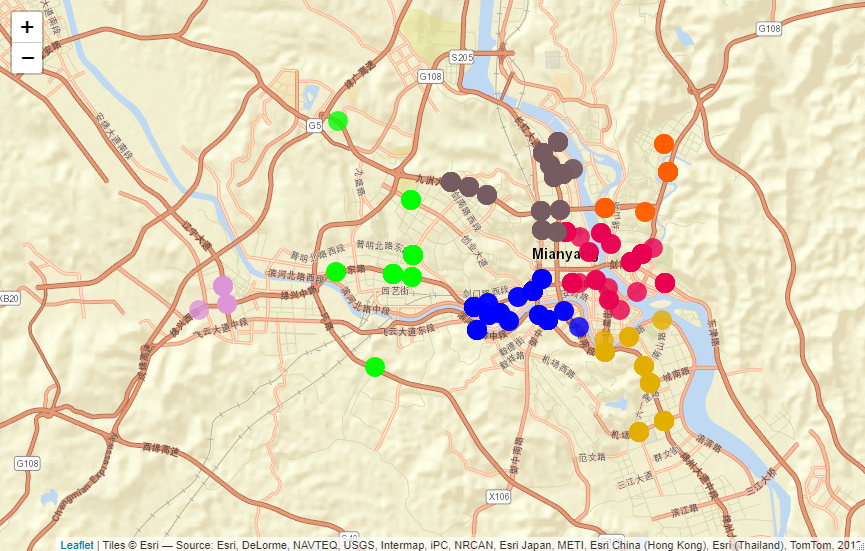
k\_start<-kmeans(start\_lat\_lng,centers = 7)

pal <- colorFactor(domain = k\_start$cluster)

leaflet(start\_lat\_lng)%>%addProviderTiles("Esri.WorldStreetMap")%>%

addCircleMarkers(fillColor = ~pal(k\_start$cluster),stroke = FALSE,fillOpacity = 0.8,popup=~as.character(k\_start$cluster))

### 1.2效果图



# 2.终点聚类

### 2.1代码

end\_lat\_lng<-dataAll[,7:8]

k\_end<-kmeans(end\_lat\_lng,centers = 6)

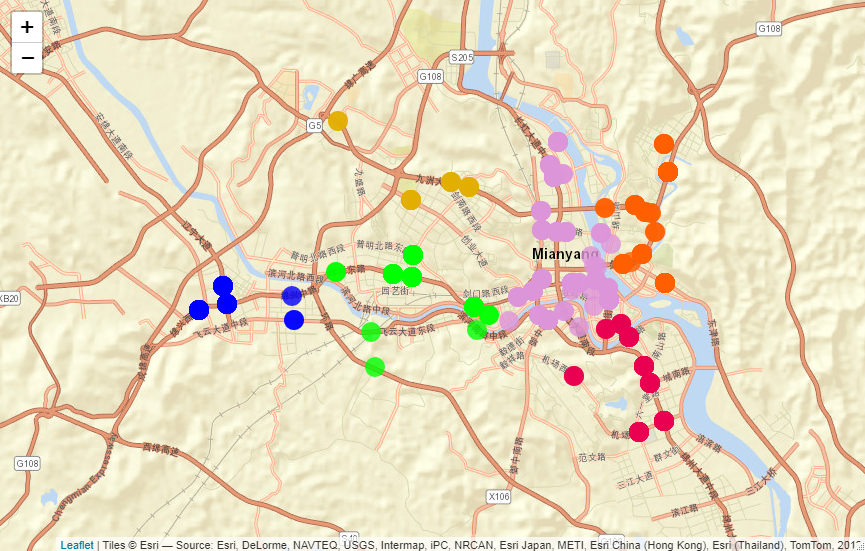
names(end\_lat\_lng)<-c("Lng","Lat")

pal <- colorFactor(domain = k\_end$cluster)

leaflet(end\_lat\_lng)%>%addProviderTiles("Esri.WorldStreetMap")%>%

addCircleMarkers(fillColor = ~pal(k\_end$cluster),stroke = FALSE,fillOpacity = 0.8,popup=~as.character(k\_end$cluster))

### 2.2效果图

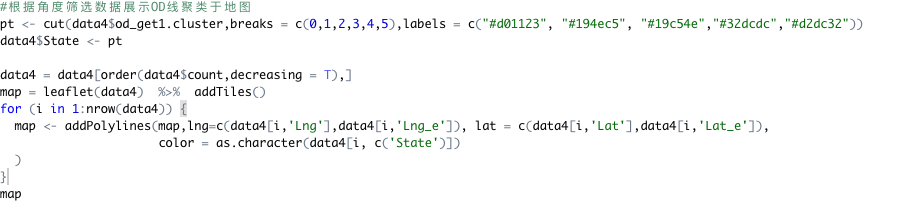


## 3.OD线聚类

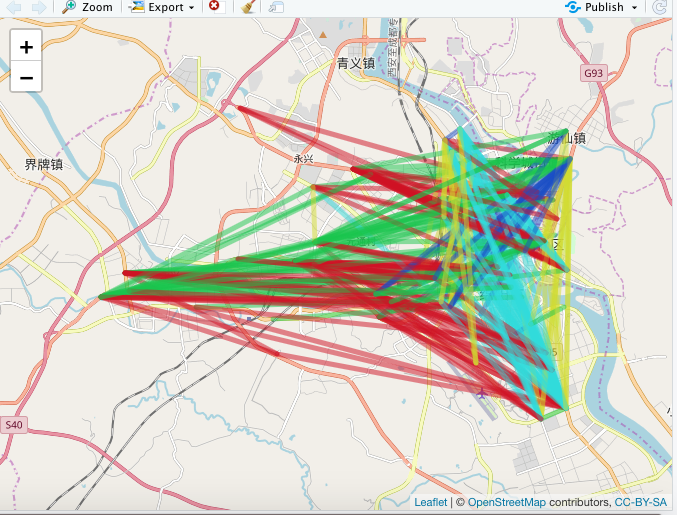
### 3.1角度

3.1.1 代码



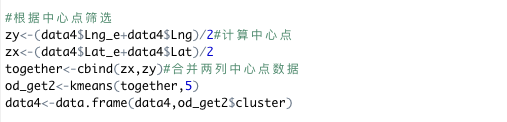


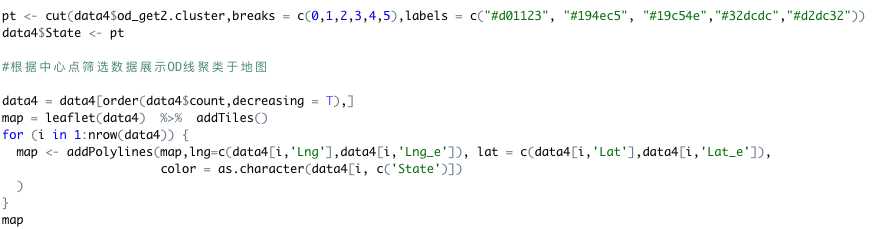
3.1.2效果图



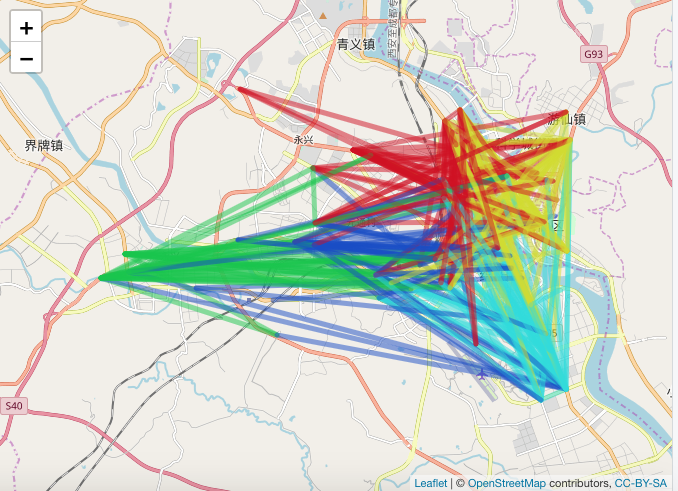
### 3.2 中心点

3.2.1 代码





3.2.2 效果图



# 4.总结

起点聚类我们设置了7个簇，根据绵阳市的片区我们设立了7个K点，从聚类结果的可视化图中我们可以明显的看出每个点所在的片区和分布。

OD线聚类最开始是从起始点经纬度和终点经纬度的连线与起始点的纬线的角度作为线的一个特征值来聚类。我们将其划分为5族。其效果图见3.1.2.其中可以看出大面积都是蓝色和红色的线段，聚类效果不是很理想。所以我们考虑了另外一种算法。我们利用起始点和终点的经纬度求出两点连线的中心点，将中心点作为线的又一个特征值。由此一来我们的一条线有了角度与中心点两个特征值。再根据这两个特征值对线使用K-means算法，进行聚类。效果图见3.2.2.可以由图中看出，其聚类结果明显好于第一种只用角度作为特征值的聚类结果。