Branch: master ▾ Find file Copy path

SJWJQZZY / 第三组 / REAMDME.md

ly1102 fix bug

01c8a5b 8 minutes ago

1 contributor

Raw Blame History

192 lines (171 sloc) 6.3 KB

组员介绍

学院：网络空间安全学院

- 李源　3190802008
- 彭乔　3190802007
- 刘国航　3190802009
- 杨璐　3190809001
- 张玥靓　3190809002
- 谭鹏飞　3190802020
- 邓杭杭　3190802015
- 张运理　3190802017
- 刘盈江　3190802013

作业完成简介

小组分为三个小队，分别使用了自己的方法完成了作业。

第一小分队

1. 安装写代码需要的包

```
# install.packages("cluster")
library(cluster)
# install.packages("leaflet")
```

```
library(leaflet)
# install.packages("geosphere")
library(geosphere)
# install.packages("ggplot2")
library(ggplot2)
# install.packages("ggfortify")
library(ggfortify)
# install.packages("fpc")
library(fpc)
```

2. 数据读取与初步处理

```
# 读取数据集
data_read <- read.csv("C:/Users/c401/Desktop/df3.CSV",header = T)
# 读取起点
data_start <- data_read[,5:6]
# 使用fpc包里的pamk函数来估计类的个数
pamk_start.best <- pamk(data_start)
# 输出合理的类个数
cat("number of start clusters estimated by optimum average silhouette width:", pa
plot(pam(data_start, pamk_start.best$nc))
data_start.pam <- pam(data_start, pamk_start.best$nc)
data_read$clustering <- data_start.pam$clustering
pt1 <- cut(as.numeric(data_read$clustering),breaks = c(0,1,2,3,4,5,6,7,8,9,10),la
data_read$color_set1 <- pt1

# 读取终点
data_end <- data_read[,7:8]
pamk_end.best <- pamk(data_end)
cat("number of end clusters estimated by optimum average silhouette width:", pamk
plot(pam(data_end, pamk_end.best$nc))
data_end.pam <- pam(data_end, pamk_end.best$nc)
data_read$clustering <- data_end.pam$clustering
pt2 <- cut(as.numeric(data_read$clustering),breaks = c(0,1,2,3,4,5,6,7,8,9,10),la
data_read$color_set2 <- pt2
```

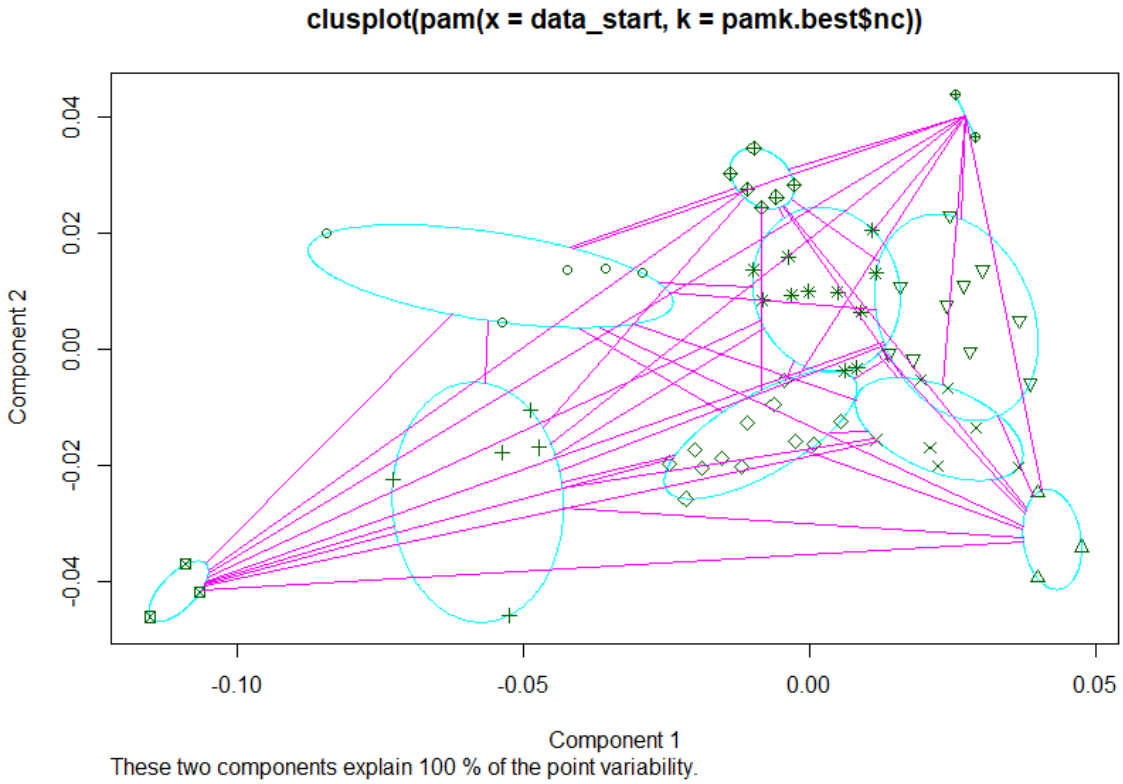
3. 分类回归与绘图

```
#求距离
for (i in 1:nrow(data_read)){
  x1 = c(data_read[i,"Lng"],data_read[i,'Lat'])
  y1 = c(data_read[i,"Lng_e"],data_read[i,'Lat_e'])
  dist1 = distm(x1,y1)
  dist1 = dist1[1,1]
  data_read$dist[i] = dist1
}


#添加起点，终点颜色
color_start = data.frame(color_start=c("#000000"))
```

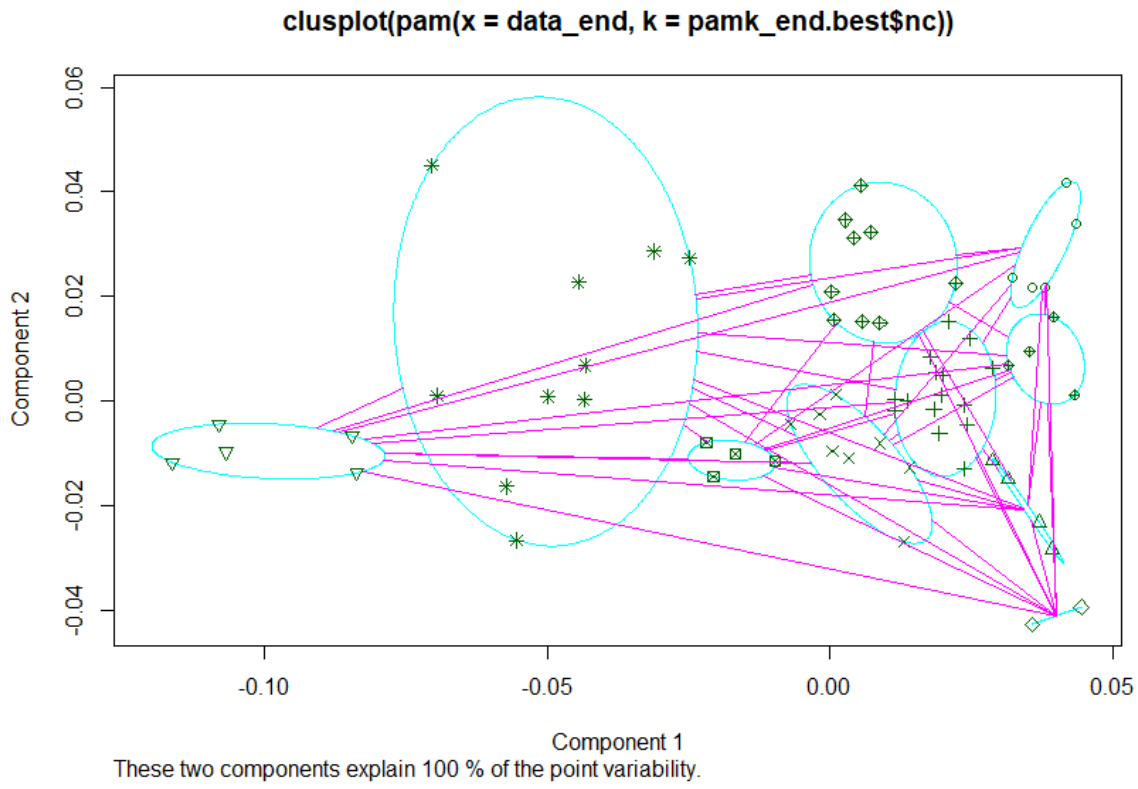
```
color_end = data.frame(color_end=c("#fffafa"))
data_read = merge(data_read,color_start,all = T)
data_read = merge(data_read,color_end,all = T)

#画地图
map = leaflet(data_read)  %>%  addTiles()
for (i in 1:nrow(data_read)) {
  map <- addCircles(map,lng=c(data_read[i,'Lng']), lat = c(data_read[i,'Lat']),
                    color = as.character(data_read[i, c('color_set1')]))
}
map
map = leaflet(data_read)  %>%  addTiles()
for (i in 1:nrow(data_read)) {
  map <- addCircles(map,lng=c(data_read[i,'Lng_e']), lat = c(data_read[i,'Lat_e']
                    color = as.character(data_read[i, c('color_set2')]))
}
map

# 绘线
y1<-data.frame(data_read$Lng_e)
y0<-data.frame(data_read$Lng)
z1<-data.frame(data_read$Lat_e)
z0<-data.frame(data_read$Lat)
jd<-(y1-y0)/(z1-z0)
df4<-data.frame(data_read,jd)
df4<-df4[complete.cases(df4[,10]),]
#pamk.best <- pamk(jd)
#x4<-pam(jd, pamk.best$nc)
jd<-na.omit(jd)
x4<-kmeans(jd,10)
df4 <- na.omit(df4)
df4<-data.frame(df4,x4$cluster)
zy<-(df4$Lng_e+df4$Lng)/2
zx<-(df4$Lat_e+df4$Lat)/2
c<-cbind(zx,zy)
#pamk.best <- pamk(c)
#x5<-pam(c, pamk.best$nc)
x5<-kmeans(c,10)
df4<-data.frame(df4,x5$cluster)
pt <- cut(as.numeric(df4$x4.cluster),breaks = c(0,1,2,3,4,5,6,7,8,9,10),labels =
df4$State <- pt

df4 = df4[order(df4$count,decreasing = T),]
map = leaflet(df4)  %>%  addTiles()
for (i in 1:nrow(df4)) {
  map <- addPolylines(map,lng=c(df4[i,'Lng'],df4[i,'Lng_e']), lat = c(df4[i,'Lat'
                      color = as.character(df4[i, c('State')])
  )
}
map

pt <- cut(df4$x5.cluster,breaks = c(0,1,2,3,4,5,6,7,8,9,10),labels = c("#000000",
df4$State2 <- pt
map = leaflet(df4)  %>%  addTiles()
for (i in 1:nrow(df4)) {
```

```
  map <- addPolylines(map,lng=c(df4[i,'Lng'],df4[i,'Lng_e']), lat = c(df4[i,'Lat'
                      color = as.character(df4[i, c('State2')])
  )
}
map
```

- 起点

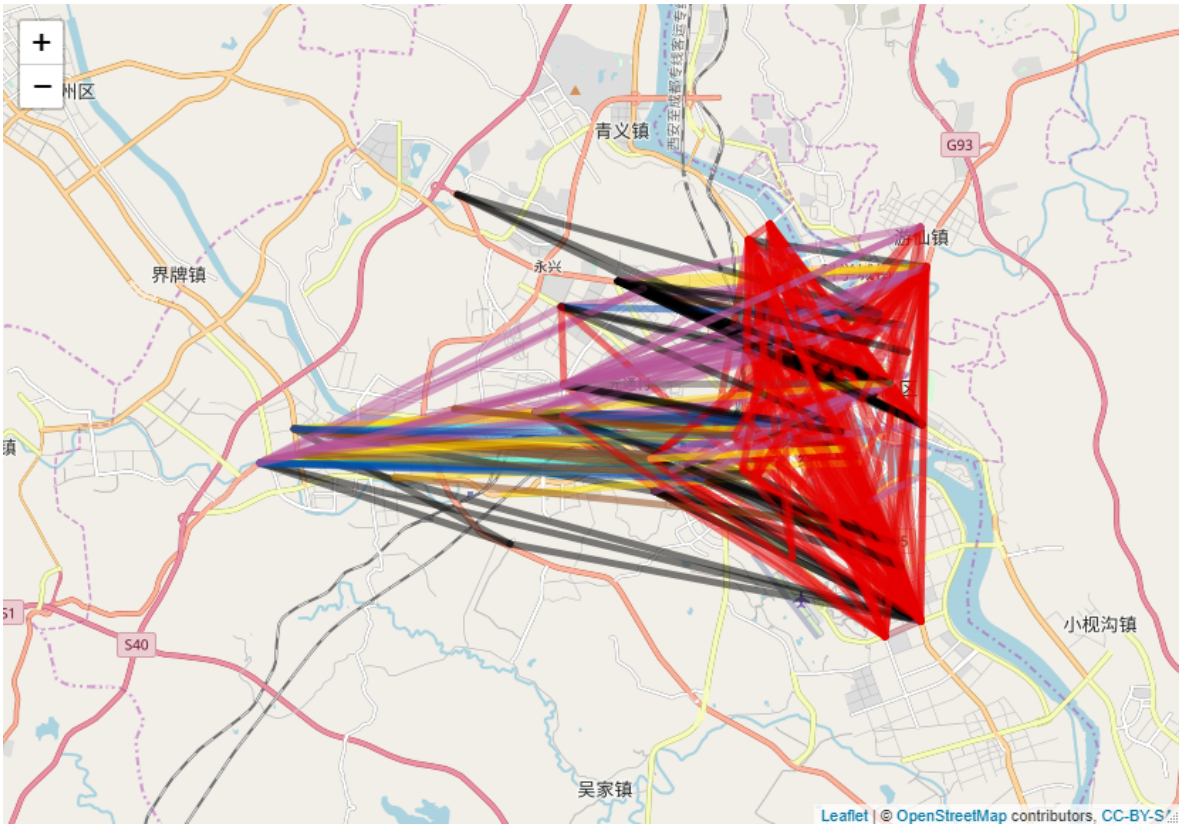


- 终点


clusplot(pam(x = data_end, k = pamk_end.best$nc))
Component 2
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.10
-0.05
0.00
0.05
Component 1
These two components explain 100 % of the point variability.


+
−
州区
青义镇
G93
游仙镇
界牌镇
永兴
科学城街道
G10
元通村
游仙区
永兴
绵阳市
S40
小枧沟镇
吴家镇
Leaflet | © OpenStreetMap contributors, CC-BY-S

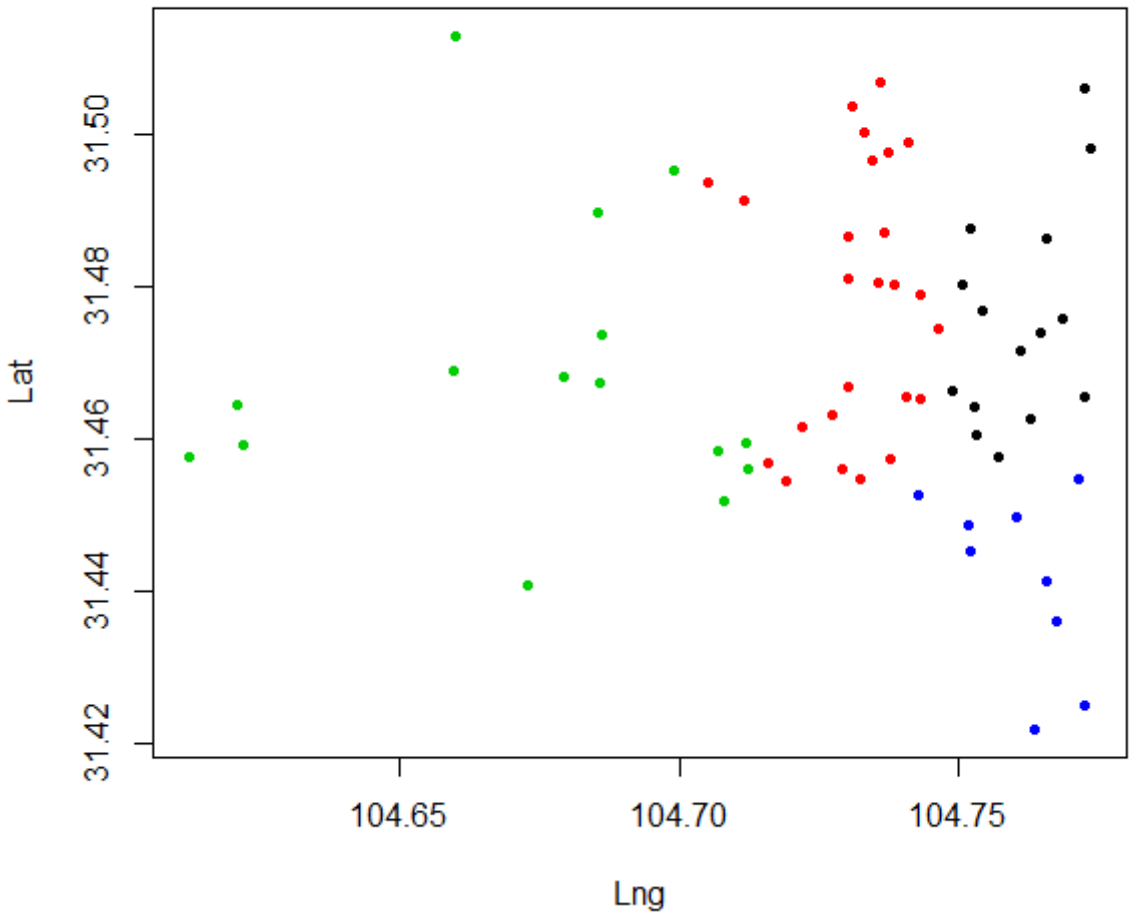
- OD



第二小分队

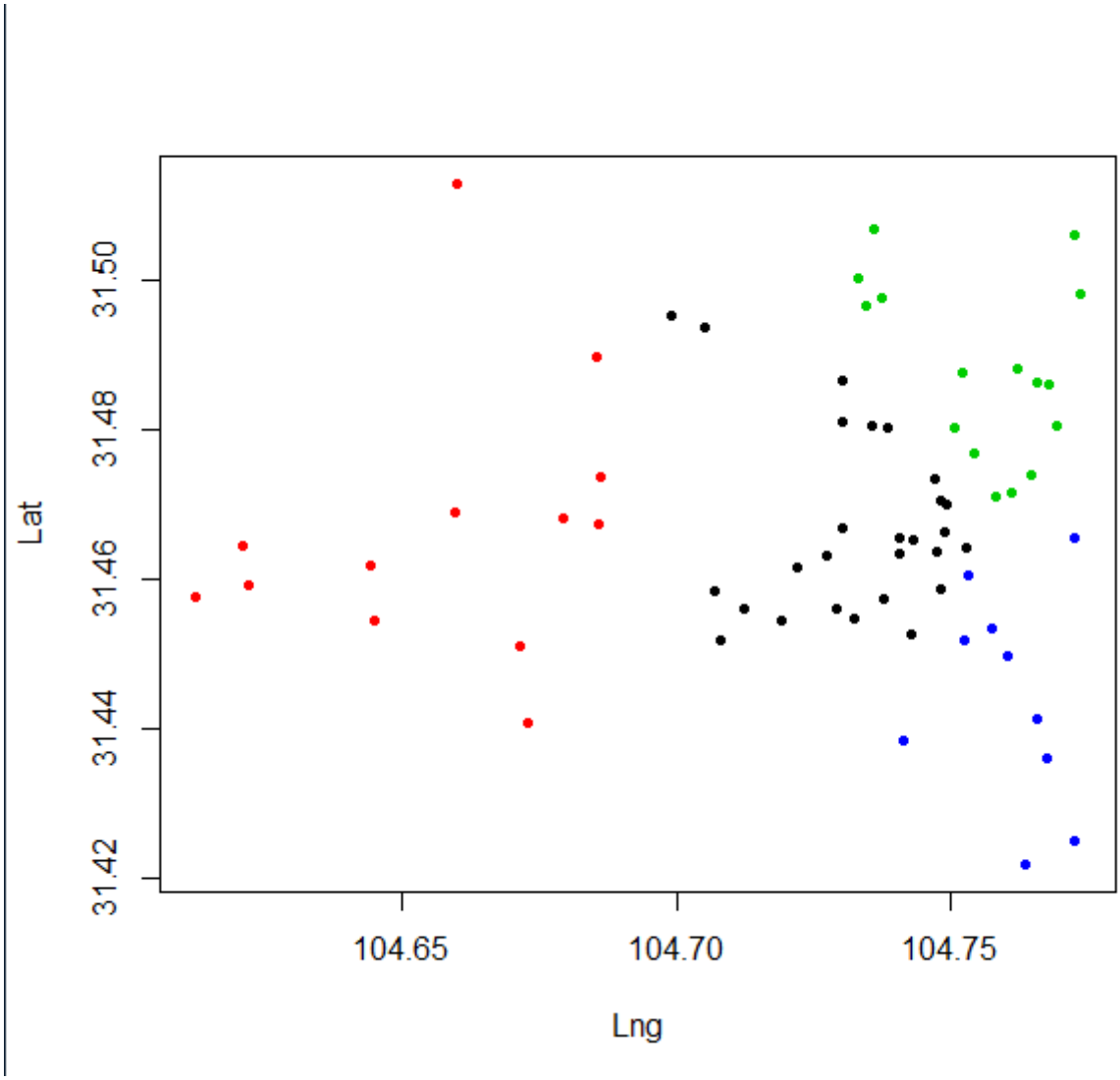
- 起点

```
library(ggplot2)
oddataset <- read.csv("./df3.csv")
kc_means_set <- cbind(oddataset[5], oddataset[6])
kc_set <- kmeans(kc_means_set, 4)
Lng = as.matrix(oddataset[5])
Lat = as.matrix(oddataset[6])
plot(Lng,Lat,col=kc_set$cluster, pch=20)
```



- 终点

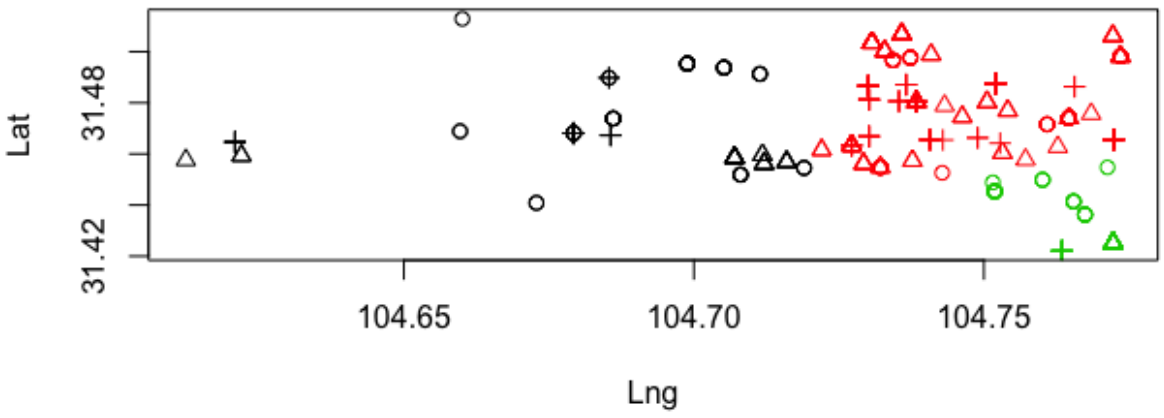
```
library(ggplot2)
oddataset <- read.csv("./df3.csv")
kc_means_set <- cbind(oddataset[5], oddataset[6])
kc_set <- kmeans(kc_means_set, 4)
Lng = as.matrix(oddataset[7])
Lat = as.matrix(oddataset[8])
plot(Lng,Lat,col=kc_set$cluster, pch=20)
```



第三小分队

- 起点

```
read.csv("df3.csv", header = FALSE)
km <- kmeans(df3[,5:6], 3,iter.max = 30)
plot(df3[c("Lng", "Lat")], col = km$cluster,pch = as.integer(iris$Species))
```



- 终点

```
read.csv("df3.csv", header = FALSE)
km <- kmeans(df3[,7:8], 3)
plot(df3[c("Lng_e", "Lat_e")], col = km$cluster,pch = as.integer(iris$Species))
```