1. 餐饮企业的部分客户的消费行为

1.1 读入数据

In [ ]:

```
# 读入数据
Data <- read.csv("C:/Users/bff/Desktop/me-ppt/consumption_data.csv", header = TRUE)[, 2:4]
km <- kmeans(Data, center = 3) #使用k均值方法聚类
print(km)
km$size / sum(km$size)
```

In [ ]:

```
km$centers# R:最近一次消费时间间隔；F，消费频率；M消费总金额
```

In [ ]:

```
## 数据分组
aaa <- data.frame(Data, km$cluster)
Data1 <- Data[which(aaa$km.cluster == 1), ]
Data2 <- Data[which(aaa$km.cluster == 2), ]
Data3 <- Data[which(aaa$km.cluster == 3), ]
```

In [ ]:

```
# 客户分群“1”的概率密度函数图
par(mfrow = c(1,3))
plot(density(Data1[, 1]), col = "red", main = "R")
plot(density(Data1[, 2]), col = "red", main = "F")
plot(density(Data1[, 3]), col = "red", main = "M")
```

In [ ]:

```
# 客户分群“2”的概率密度函数图
par(mfrow = c(1, 3))
plot(density(Data2[, 1]), col="red", main = "R")
plot(density(Data2[, 2]), col="red", main = "F")
plot(density(Data2[, 3]), col="red", main = "M")
```

In [ ]:

```
# 客户分群“3”的概率密度函数图
par(mfrow = c(1, 3))
plot(density(Data3[, 1]), col="red", main = "R")
plot(density(Data3[, 2]), col="red", main = "F")
plot(density(Data3[, 3]), col="red", main = "M")

```

2. 基于基站定位数据的商圈分析

2.1 读取数据及树图

In [ ]:

```r
#读取数据并得到谱系聚类图
Data <- read.csv("C:/Users/bff/Desktop/me-ppt/standardized.csv", header = FALSE)
# colnames(Data) <- c("x1", "x2", "x3","x4")
attach(Data)
dist <- dist(Data, method = 'euclidean')
hc1 <- hclust(dist, "ward.D2")
plot(hc1, hang = -1)
detach(Data)
```

从上述结果可看出分为三类较合适

2.2 层次聚类算法

In [ ]:

```r
Data <- read.csv("C:/Users/bff/Desktop/me-ppt/standardized.csv", header = FALSE)
attach(Data)
dist <- dist(Data, method = 'euclidean')
hc1 <- hclust(dist, "ward.D2")
plot(hc1)
# 分成三类
re1 <- rect.hclust(hc1, k = 3)

a <- re1[[2]]
b <- re1[[3]]
c <- re1[[1]]
```

In [ ]:

```r
library(ggplot2)
# 商圈类别1
matrix <- Data[a,]
d <- dim(matrix)
y <- as.numeric(t(matrix))
row <- factor(rep(1:d[1], each = d[2]))
x <- rep(1:d[2], times = d[1])
data <- data.frame(y = y, x = x, row = row)
ggplot(data = data, aes(x = x, y = y, group = row)) + geom_line() +
  scale_x_continuous(breaks = c(1, 2, 3, 4),
                     labels = c("工作日人均停留时间", "凌晨人均停留时间",
                                "周末人均停留时间", "日均人流量")) +
  labs(title = "商圈类别1", x = "", y = "")
```

图形语法中至少包括了如下几个图形部件：

数据(data) 映射(mapping) 几何对象(geom) 统计变换(stats) 标度(scale) 坐标系(coord) 分面(facet) 这些组件之间是通过“+”, 以图层(layer)的方式来粘合构图的, 所以图层是ggplot2中一个重要的概念。

In [ ]:

```r
# 商圈类别2
matrix <- Data[b, ]
d <- dim(matrix)
y <- as.numeric(t(matrix))
row <- factor(rep(1:d[1], each = d[2]))
x <- rep(1:d[2], times = d[1])
data <- data.frame(y = y, x = x, row = row)
ggplot(data = data, aes(x = x, y = y, group = row)) + geom_line() +
  scale_x_continuous(breaks = c(1, 2, 3, 4),
                     labels = c("工作日人均停留时间", "凌晨人均停留时间",
                                "周末人均停留时间", "日均人流量")) +
  labs(title="商圈类别2", x="", y="")
```

In [ ]:

```r
# 商圈类别3
matrix <- Data[c,]
d <- dim(matrix)
y <- as.numeric(t(matrix))
row <- factor(rep(1:d[1], each = d[2]))
x <- rep(1:d[2], times = d[1])
data <- data.frame(y = y, x = x, row = row)
ggplot(data = data, aes(x = x, y = y, group = row)) + geom_line() +
  scale_x_continuous(breaks = c(1, 2, 3, 4),
                     labels = c("工作日人均停留时间", "凌晨人均停留时间",
                                "周末人均停留时间", "日均人流量")) +
  labs(title = "商圈类别3", x = "", y = "")
```

In [ ]:

```r

```