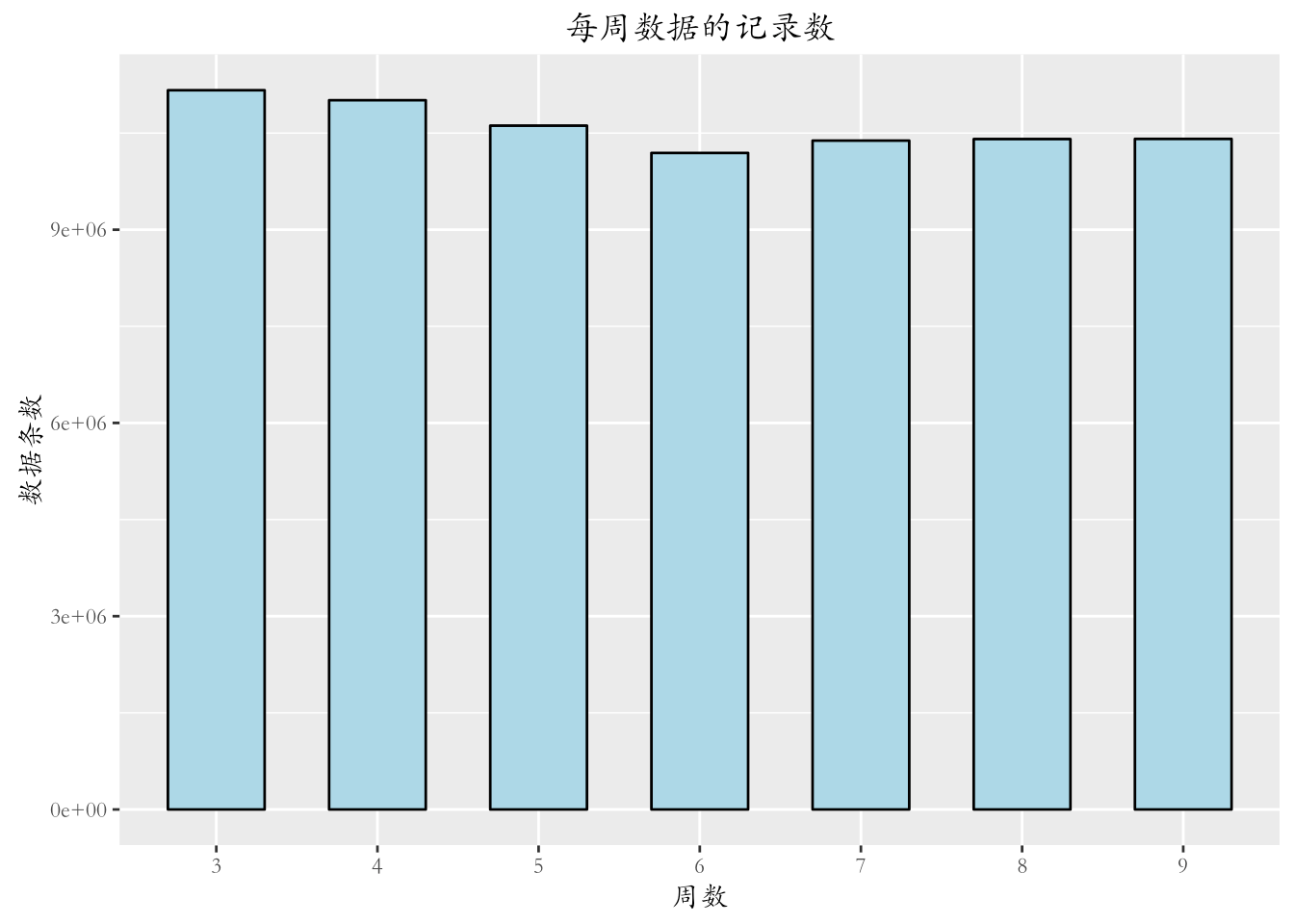
## 2016年暑期课程设计  
## 问题：Grupo Bimbo Inventory Demand  
## 宾堡集团的库存需求  
## 最大限度地提高销售和最大限度地减少烘焙食品的退回  
## Daitu  
## start:2016.06.21  
## 参考借鉴kaggle上的公开程序

**数据分析：**

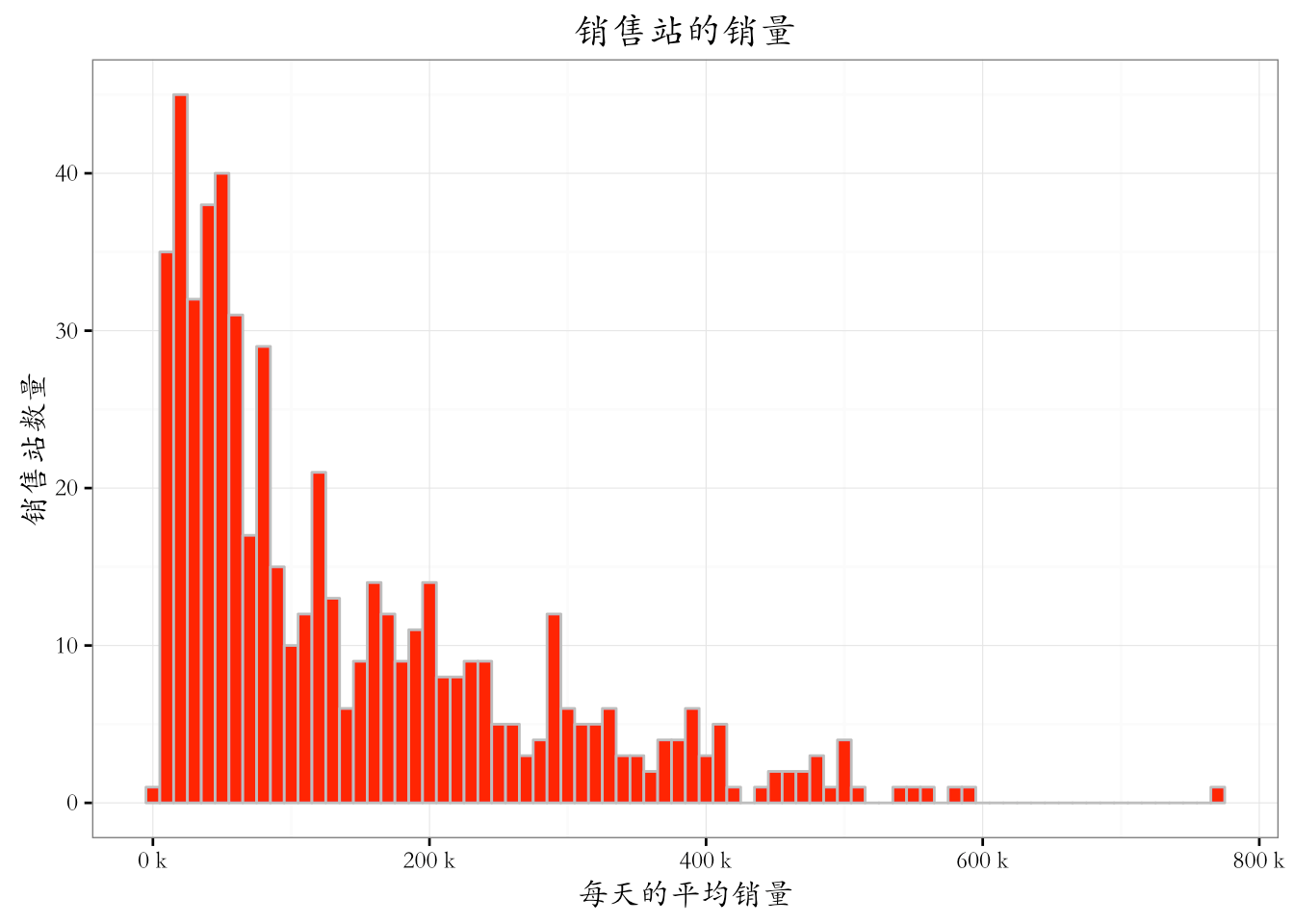
**描述性统计**

1、分析数据的周数：Semana  
## 条形图

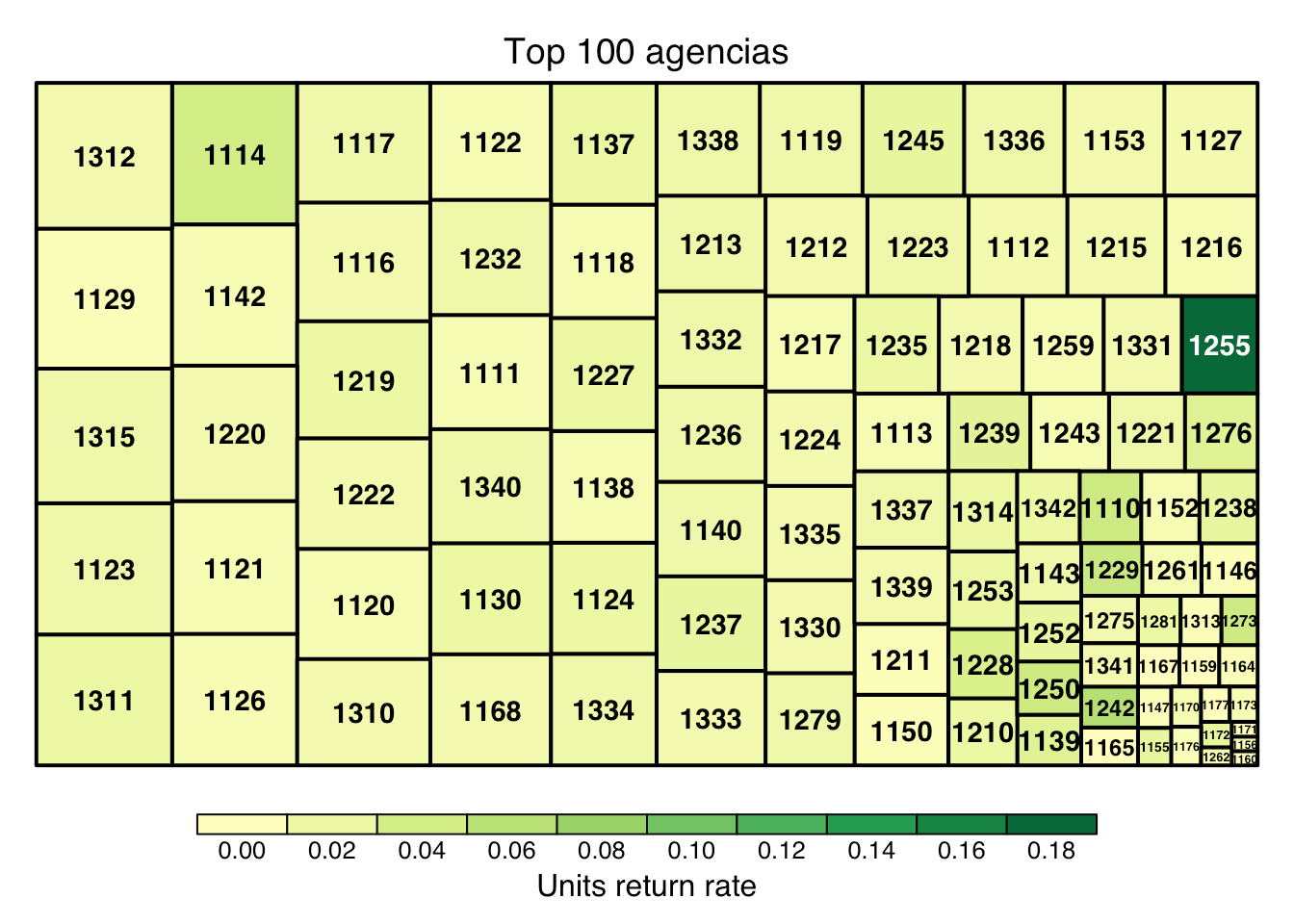


分析:该图表示一星期中每天数据数量，星期六相对较少，星期三最多，但大体分布比较均匀，即每天的情况相差不大。

2、销售站的数据分析  
销售站Agencia 和 州State

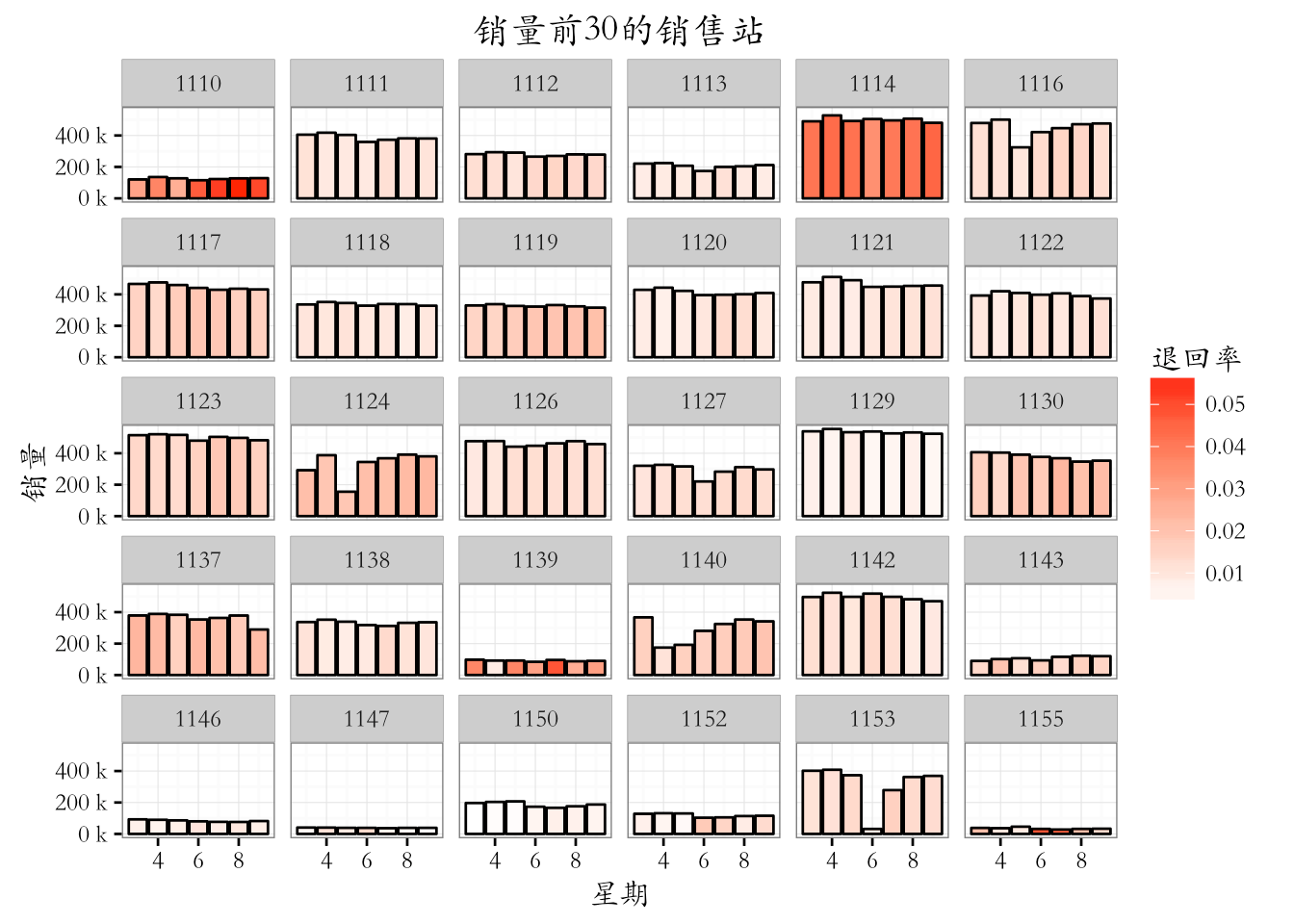


前100的销售站的销量树图



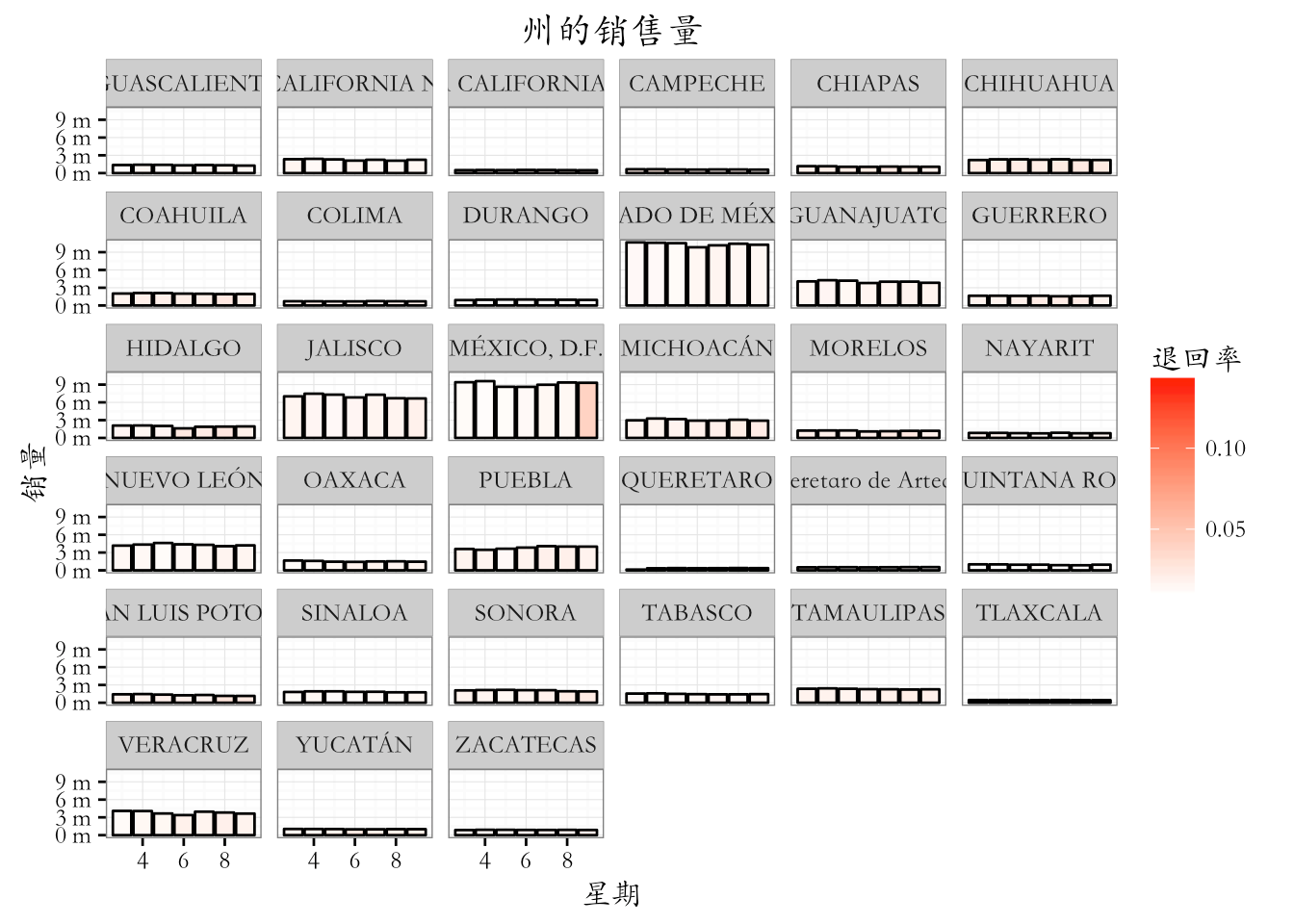
分析：由图可知，每个销售站每天的平均销量分布相对较集中，大多数都分布在0k~200k期间，说明多数的销售站销量还是一般，只有少数的销量较高，达到400k的更是很少。

销量前30的销售站每周的销量和退货率图



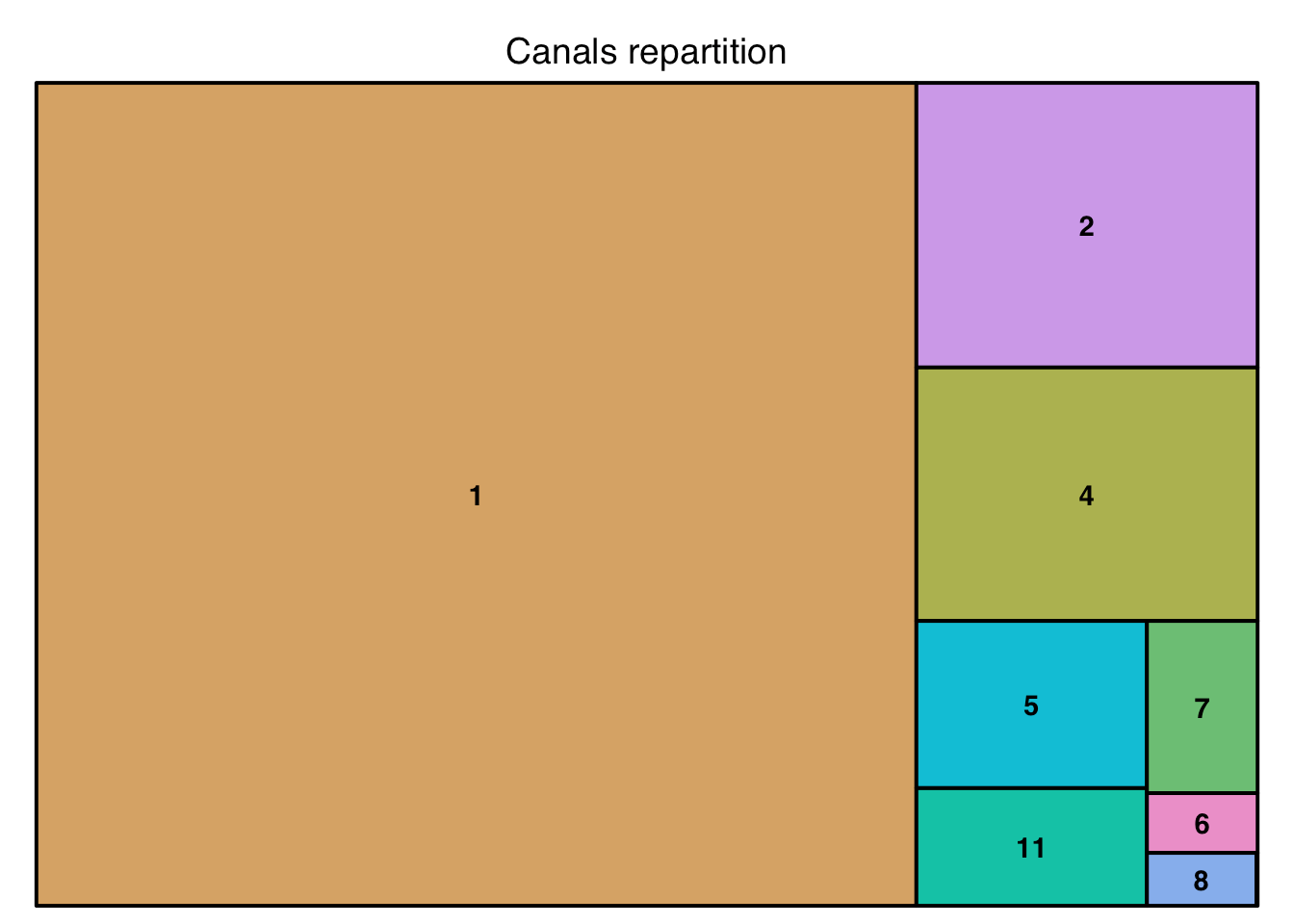
分析：在销量为前30的销售站中，在一星期中，其各自销量都较平均，退回率也相差不多。例如，销售站1114每天销量都较靠前，但是每天的退回率也都较高；销售站1129销量很高，退回率很低，还有1110销量低，退回率高，1146销量低，退回率低。

3、每个州的销售数据的分析



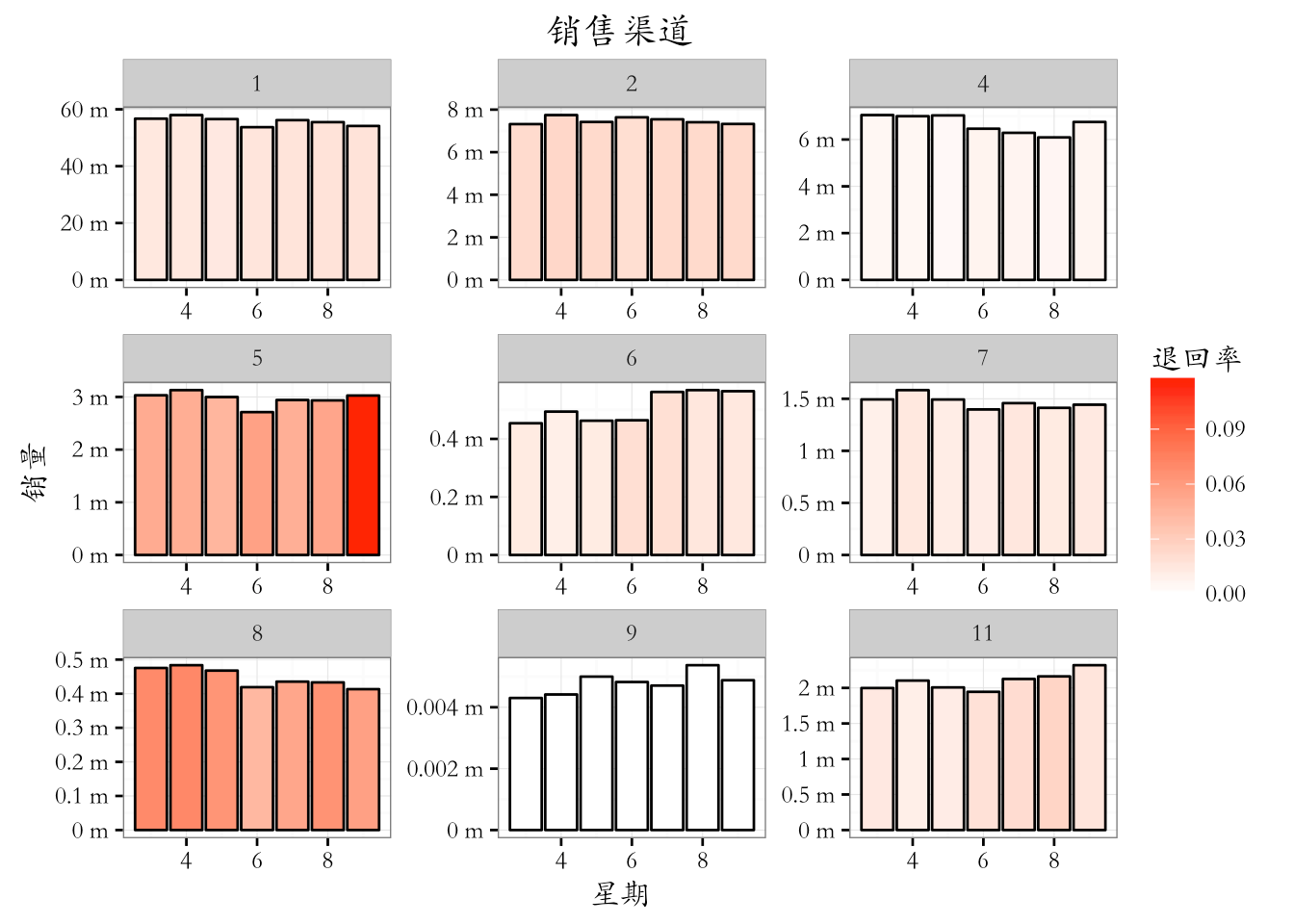
分析：该图按照州的分布进行一星期的销售量统计，并对每个州的退回率也进行统计。只有几个州的销量较高，大体销量销量都较低。每个州的每天销量分布相对较平均。

4、销售渠道的分析  
# 销售渠道1占据主要的销量



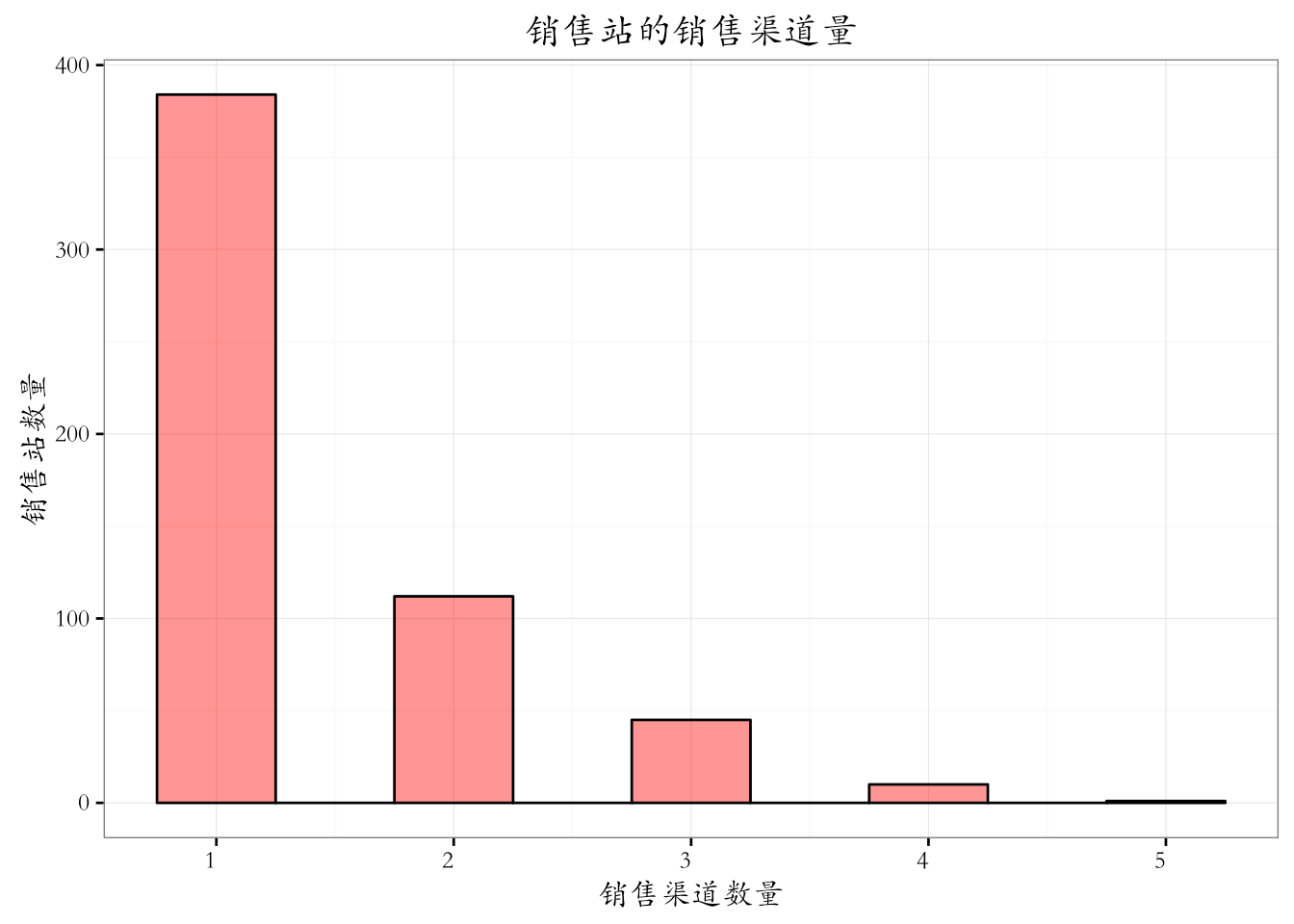
分析：主要有8种销售渠道，其中销售渠道1占销售量的绝大部分，其次为销售渠道2和4，所占比重最少的为销售渠道6和8.由此可知，大部分销售商都更愿意选择销售渠道1，并且从销量看来，选择销售渠道1是个不错的选择。

5、销售渠道的销量和星期和退货率



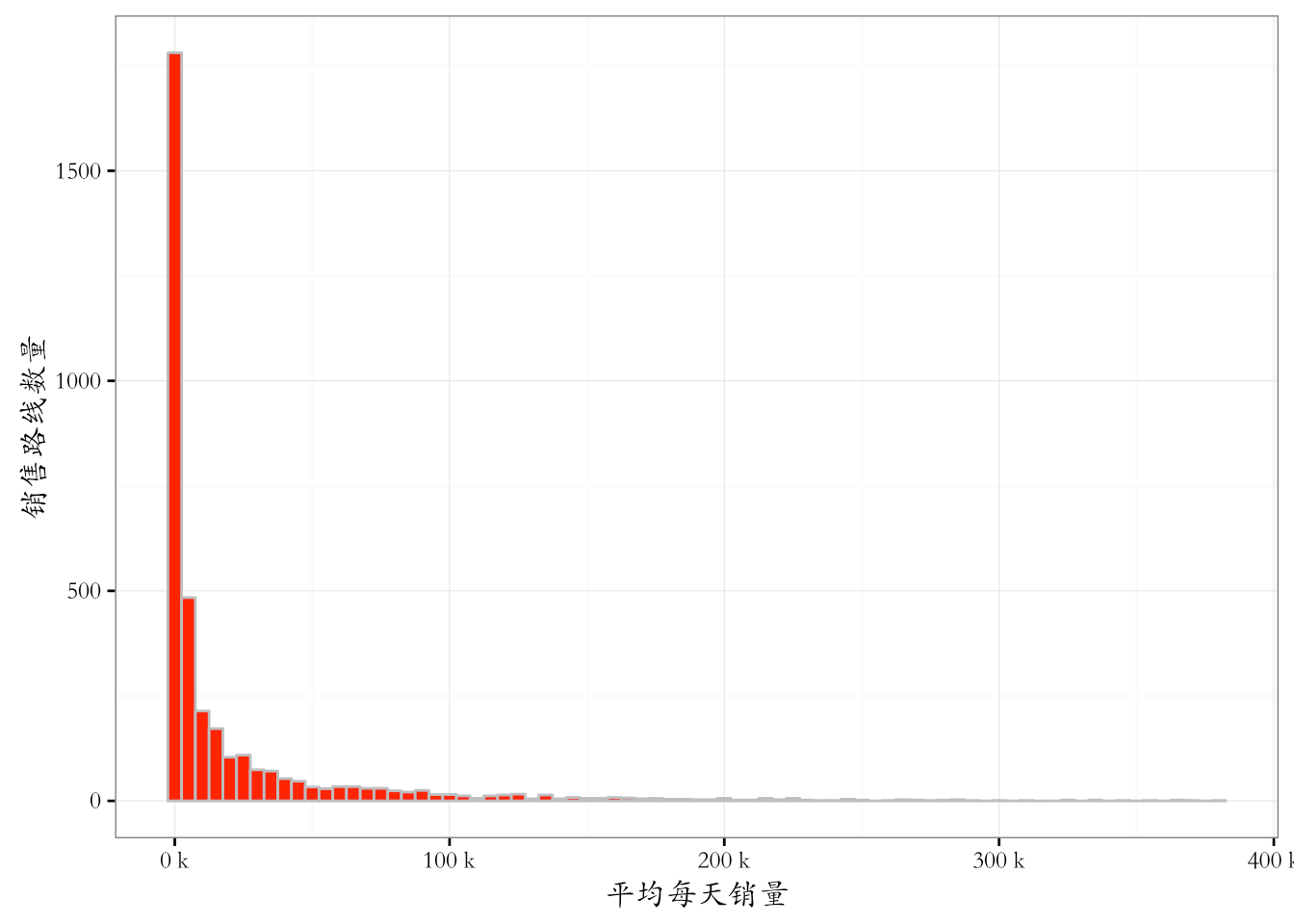
分析：该图按星期分布，呈现了不同的销售渠道所对应的销量，可见每种销售渠道对应的销量都相对较均匀，但是销售渠道5、8和2，退回率都较高，特别是销售渠道5的星期二，退回率是最高的。

6、销售渠道和销售站分析  
销售渠道有多少销售站可视化



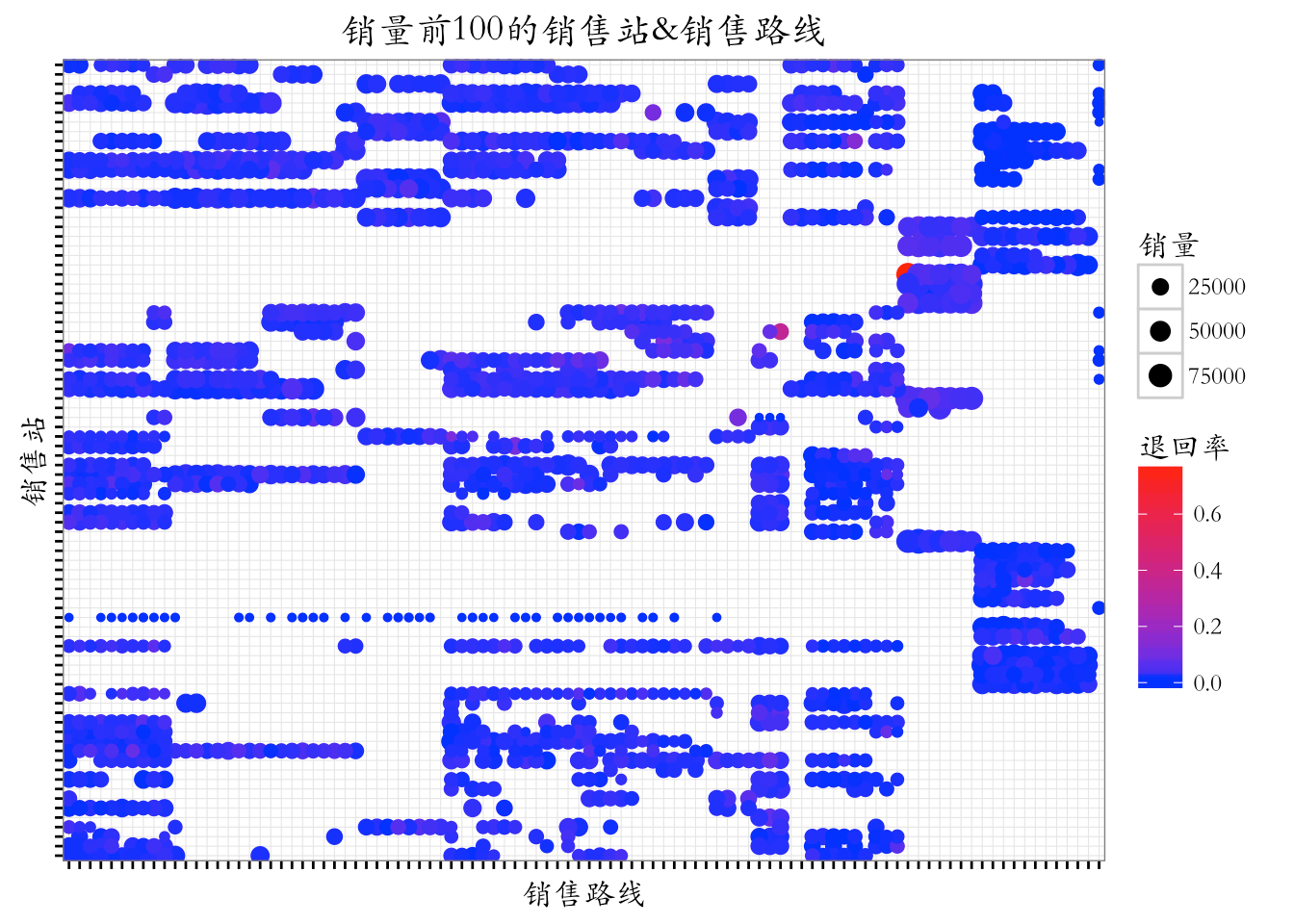
分析：由该图可知，销售渠道1所含的销售站远远高于其他，这与前面图6所呈现的结果相一致。因为销售渠道1所占销售站数量较大，所以通过渠道1的销售量最多，并占总销售量的比例最大。

# 大部分销售站有1条销售渠道，只有很少的销售站有超过三条的销售渠道  
7、销售路线的分析



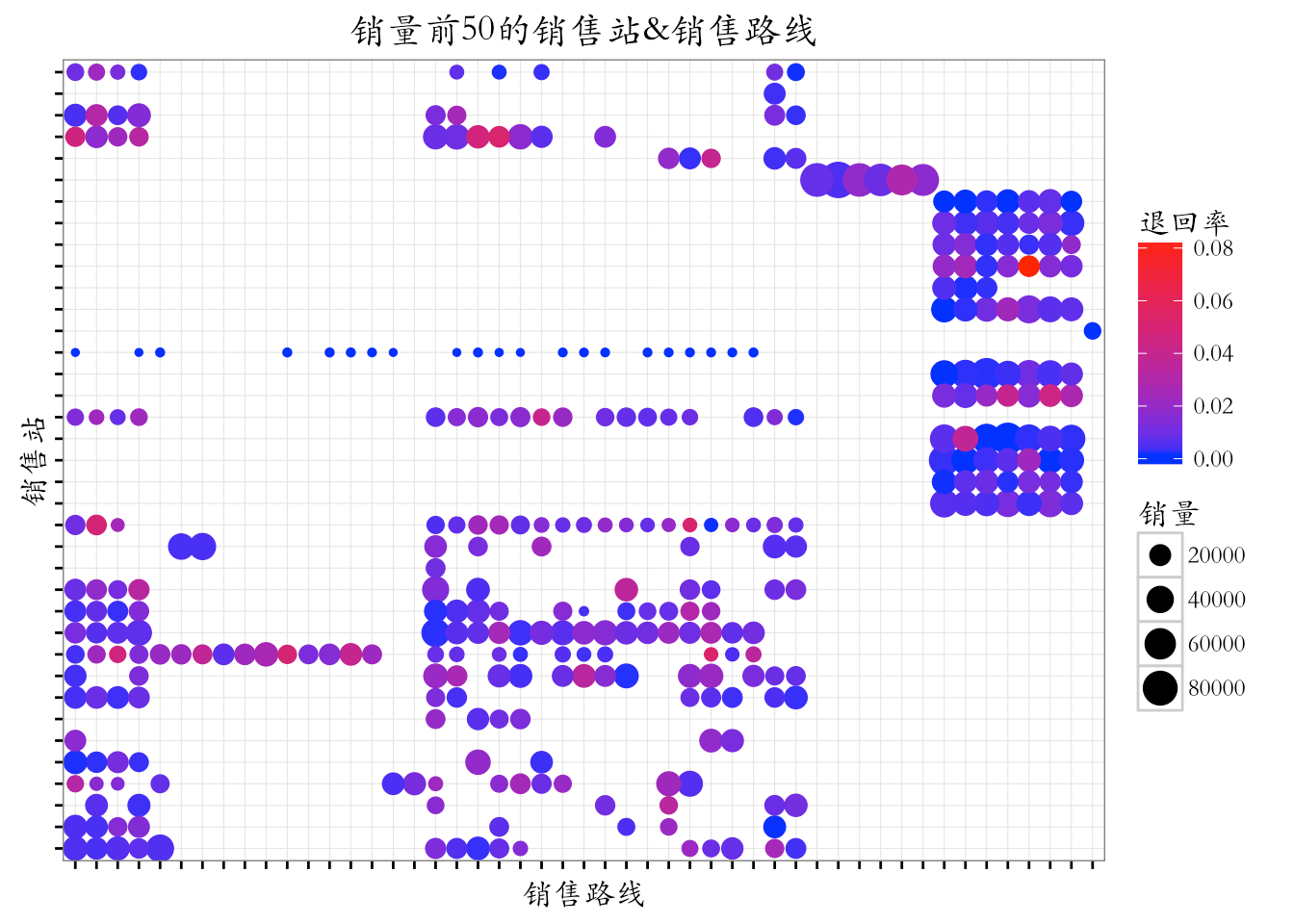
分析：从平均每天销量和销售路线数量的条形图可以看出，大部分销售路线的平均销售量并不多，超过2/3的销售路线每天的销售量不超过10000，且随着销量的增多销售路线的数量有递减的趋势。  
8、销售路线和销售站

销量前100的销售站&销售路线：



分析：图中每一个点都是对应的销售路线与销售站的销量，其中点的大小表示销量的多少，点分布越多的地方销量越高；点的颜色表示退回的百分比，由红到蓝退回百分比递减。销量前100的点的分布较为均匀，无法得出体现其中内在规律的结论。

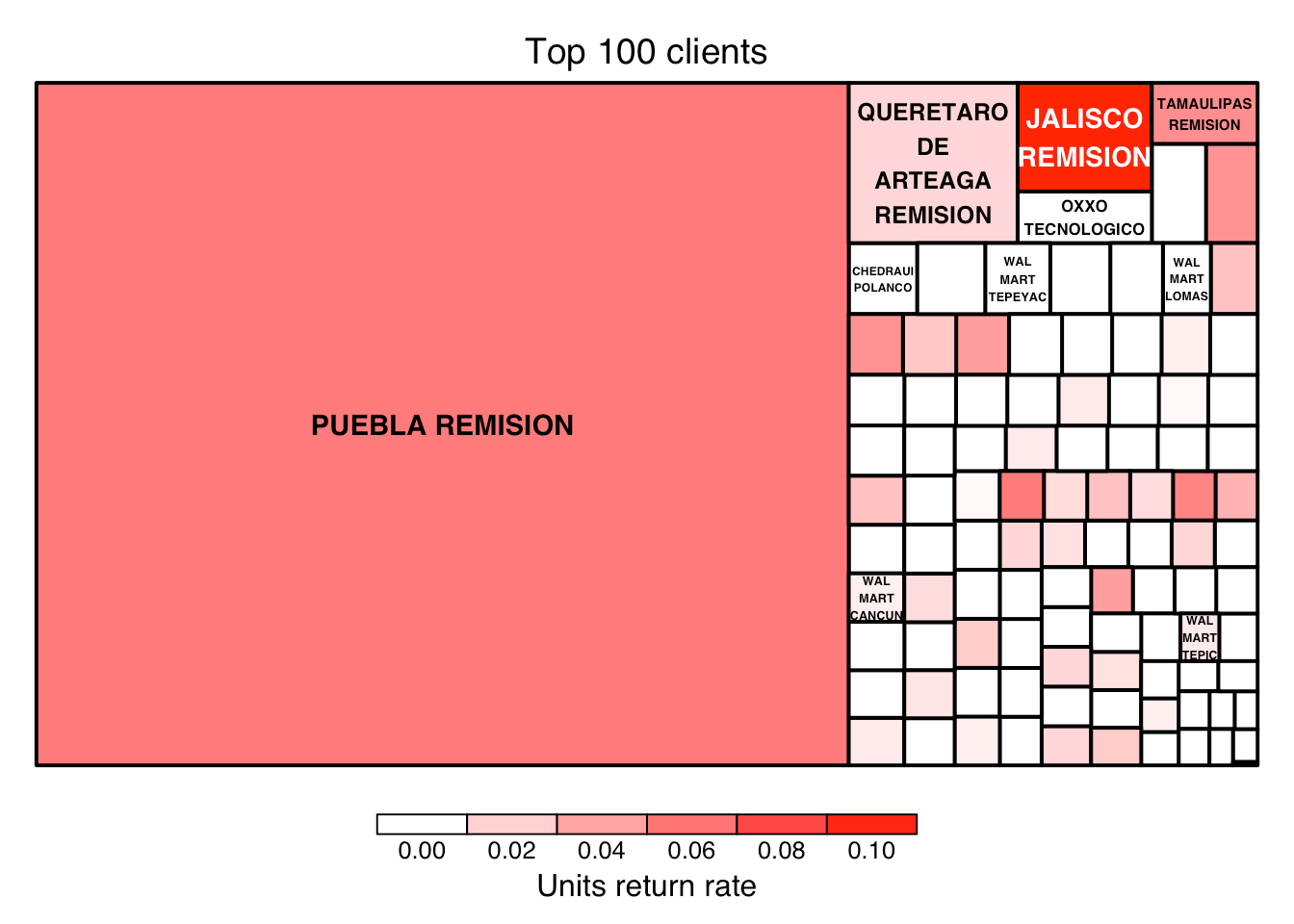
销量前50的销售站&销售路线：



分析：从图中可以看出，在图的右上侧点的 分布较为密集，点较大，颜色偏蓝，说明销售量大相应的退回百分比较小。

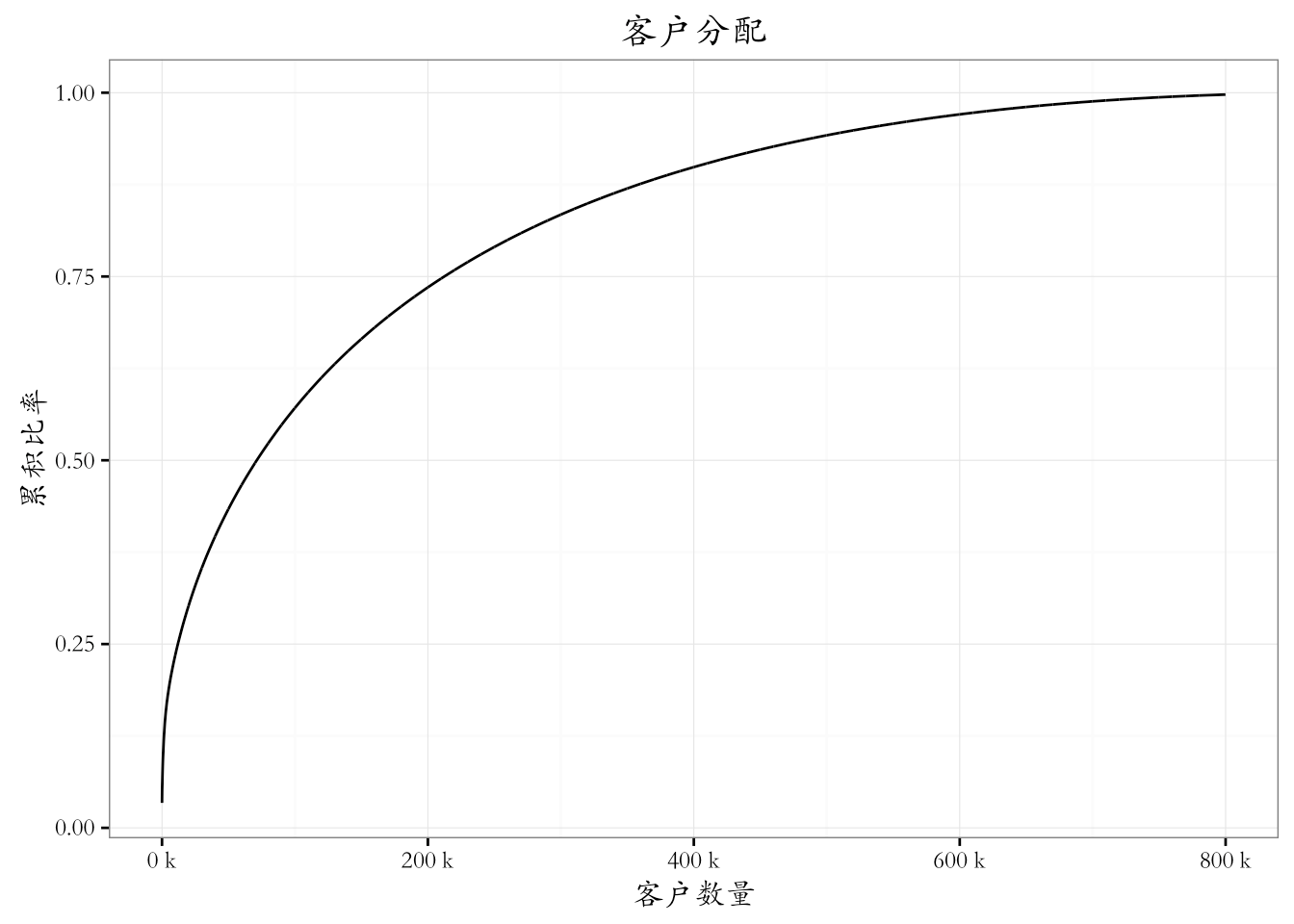
9、对客户数据进行分析

# 花费量前100个客户的树形图



分析：在树形图中面积最大的是Puebla Remision，可见他是一个大客户，但是退回的率也偏高。

10、客户的累积消耗量



分析：曲线随着客户数量的增多逐渐平缓，销量的累积百分比逐渐趋于1，可以看出前20万客户约贡献了75％的销售量。  
11、客户和销售站分析

## Cliente\_ID n\_agencias NombreCliente

## 1 26 2 BODEGA COMERCIAL MEXICANA TOLUCA

## 2 60 2 SAMS CLUB TOLUCA

## 3 65 2 WAL MART METEPEC

## 4 101 1 WAL MART TOLUCA

## 5 105 1 SUPER KOMPRAS SAN BUENAVENTURA

## 6 106 1 ISSSTE 21

分析：客户ID信息其对应的销售站数据集的前6条有3个客户通过2个销售站购买产品，另外3个客户则通过1个销售站购买产品。

dim(agencias\_by\_client)

## [1] 885416 3

分析：一共有885416个销售站信息。

单个客户使用销售站的数量

table(agencias\_by\_client$n\_agencias)

##

## 1 2 3 4 5 9 62

## 844113 37510 3771 19 1 1 1

分析：大部分的客户只从一个销售站购买产品，只有几个客户通过5个以上的销售站购买产品。

使用销售站多的客户信息

agencias\_by\_client %>% filter(n\_agencias %in% c(5, 9, 62)) #返回符合条件的行

## Source: local data table [3 x 3]

##

## Cliente\_ID n\_agencias NombreCliente

## (int) (int) (chr)

## 1 188391 9 DESAYUNOS ESCOLARES

## 2 653378 62 PUEBLA REMISION

## 3 1274327 5 COMERCIALIZADORA LA PUERTA DEL SOL

分析：客户DESAYUNOS ESCOLARES、PUEBLA REMISION、COMERCIALIZADORA LA PUERTA DEL SOL分别通过9、62、5个销售站购买产品。

12、客户和购买渠道分析  
（1）table(clients\_canals$n\_canals)

# 1 2 3 4   
# 874022 6516 65 1   
分析：大多数客户只有一个购买渠道。不同的销售渠道可以为一个客户提供服务。  
（2）table(clients\_agencies\_canals$n\_canals)  
## 1 2 3   
## 922108 3271 3

分析：很少有销售站有同一个客户通过多个渠道。

13、客户和路线分析

## Cliente\_ID n\_routes

## 1 15766 1

## 2 22926 2

## 3 24080 1

## 4 24695 1

## 5 50379 1

## 6 50395 1

分析：客户ID信息其对应的销售路线数据集的前6条有5个客户通过1个路线购买产品，只有1个客户通过2条路线购买产品。

dim(clients\_routes)

## [1] 880604 2

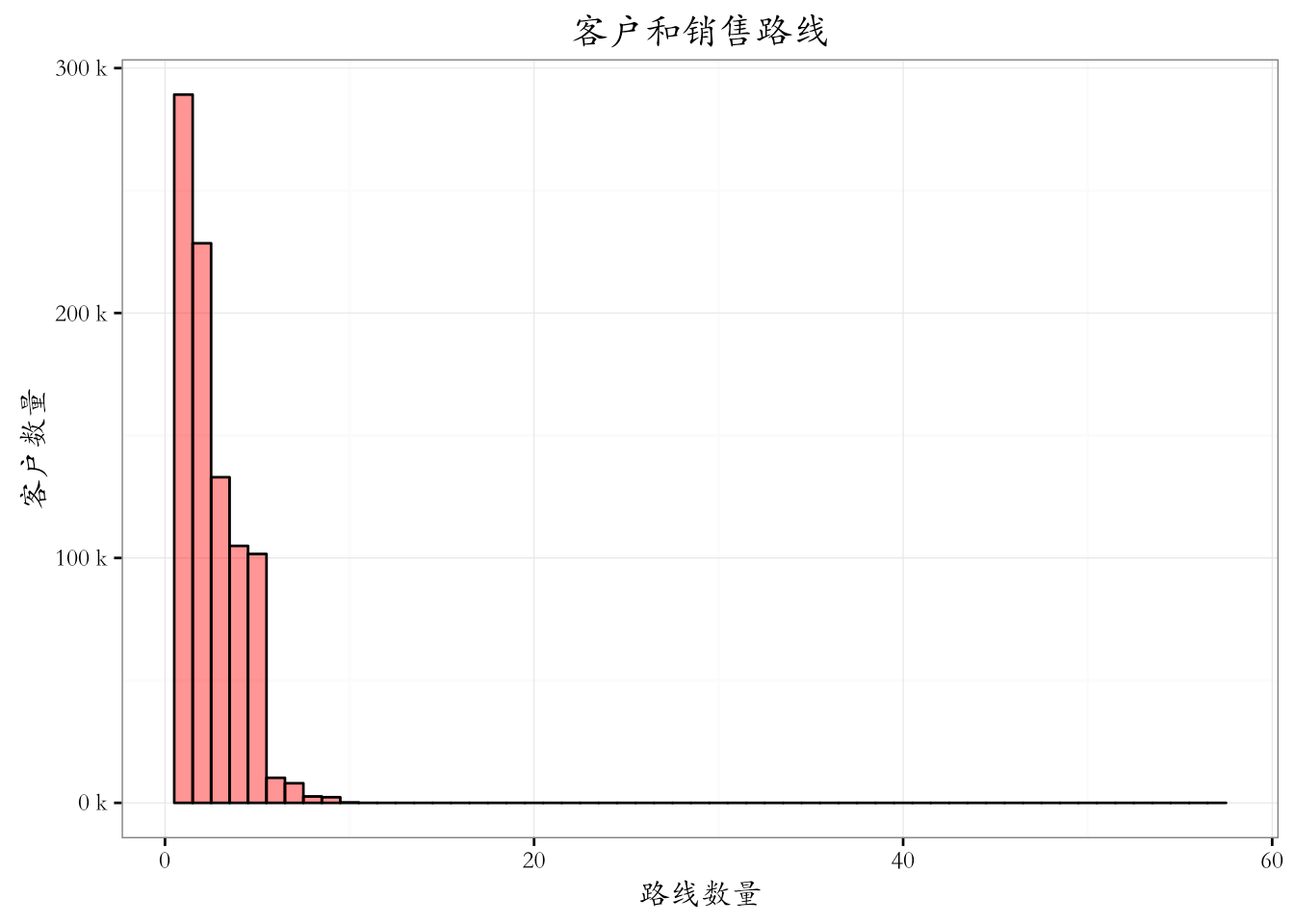
分析：一共有880604条购买产品的路线。

## 大多数客户只有不到5个仓库的交货，但超过240个客户的工作与10个仓库或更多。

sum(clients\_routes$n\_routes >= 10)

## [1] 242

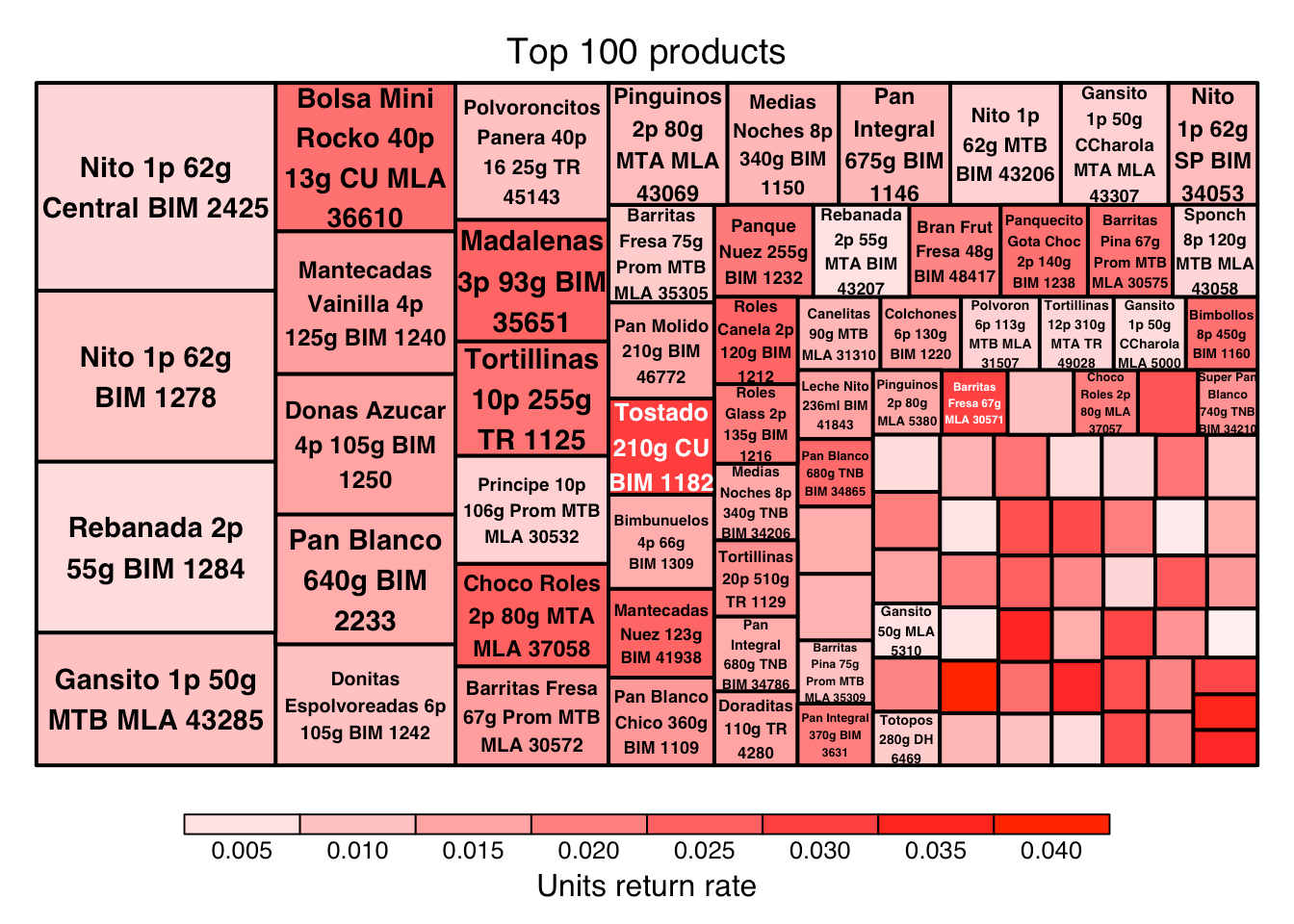
sum(clients\_routes$n\_routes >= 10)

## [1] 242  
分析：大多数客户只有不到5个仓库的交货，但超过240个客户的工作与10个仓库或更多。  


分析：大部分客户选择的销售路线数量不超过10条，5条路线以下的客户数量较多，随着路线的增多客户数量递减。

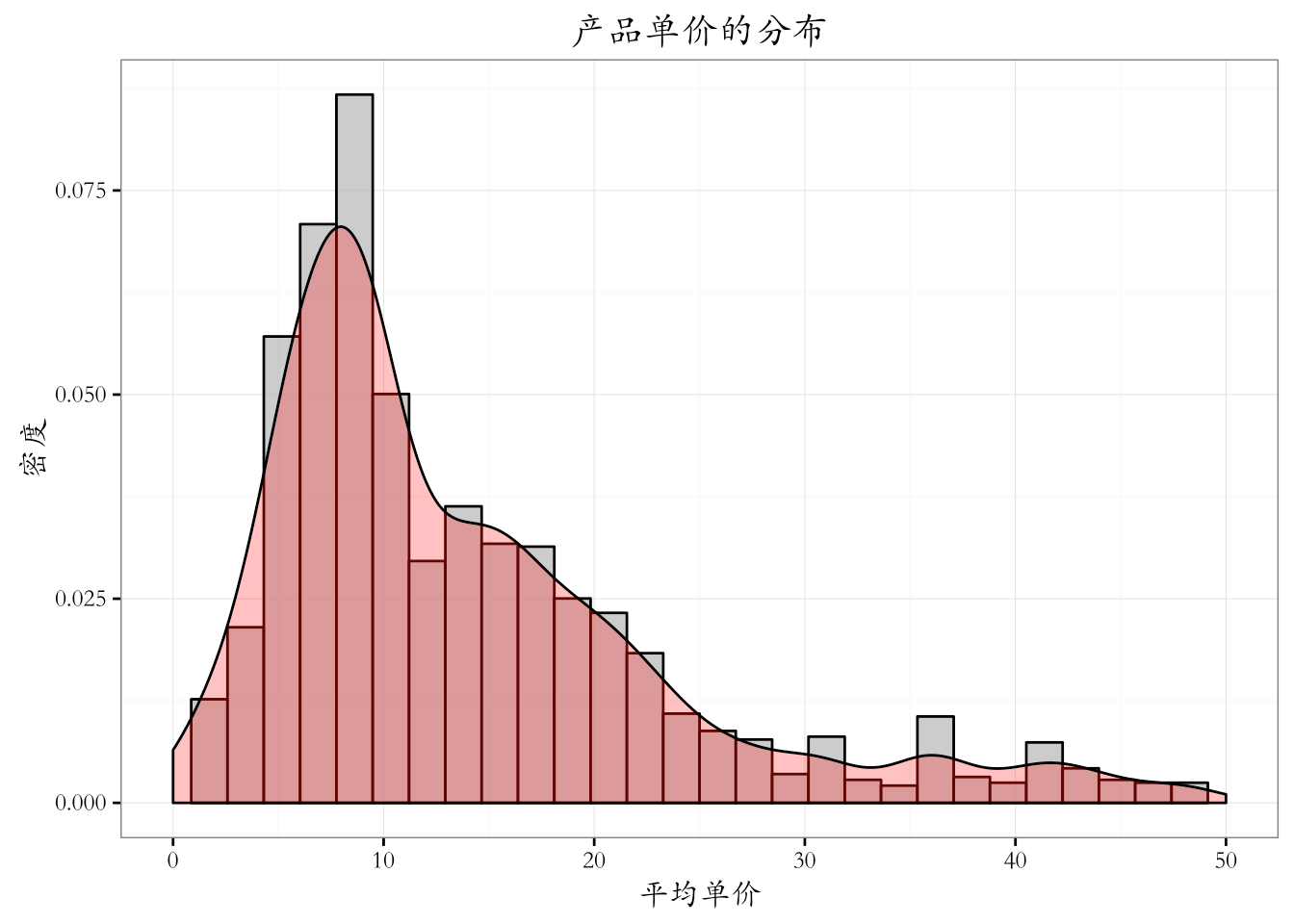
1. 对集团销售的产品进行分析

销量前100的产品树形图

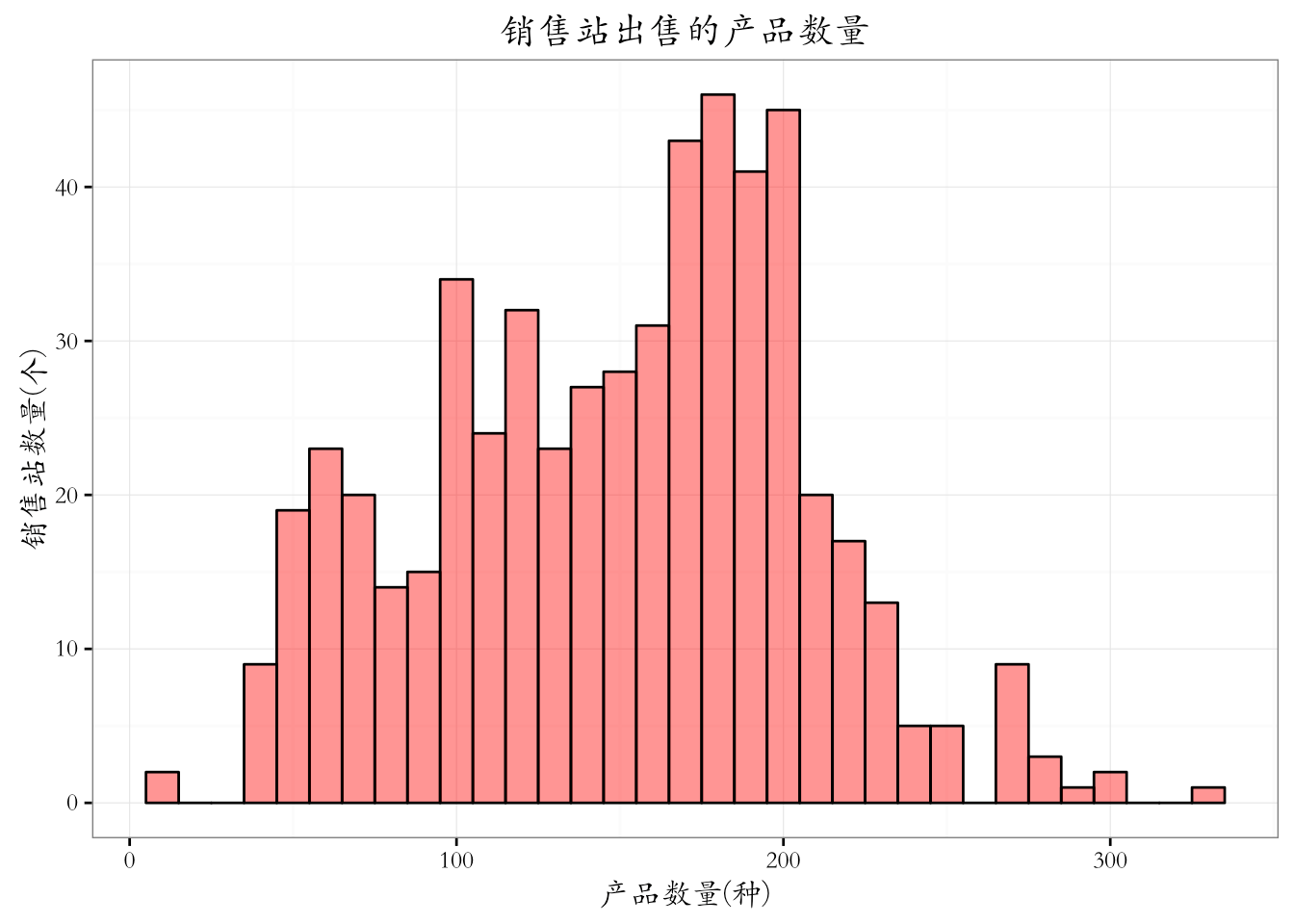


分析：销量较多的产品前三名是Nito 1p 62g Central BIM 2425，Rebanada 2p 55g BIM 1284，Nito 1p 62g BIM 1278。

15、产品单价的密度分布

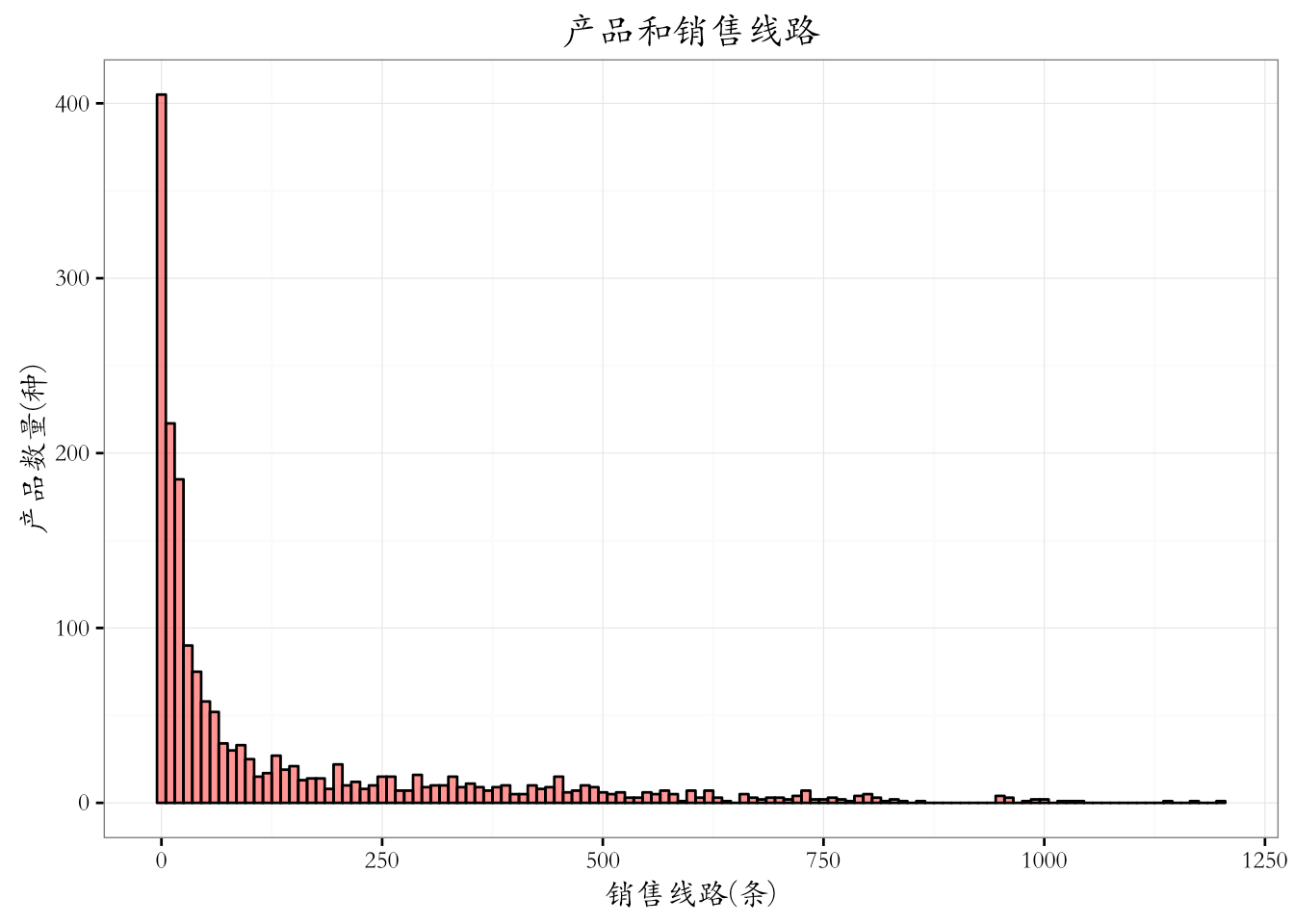


分析：每种产品的单价=该产品的营业额/该产品的销量，条形图和和密度曲线图可以看出平均单价主要集中在[0,20]，单价为8的产品销量最多。

16、产品和销售站  


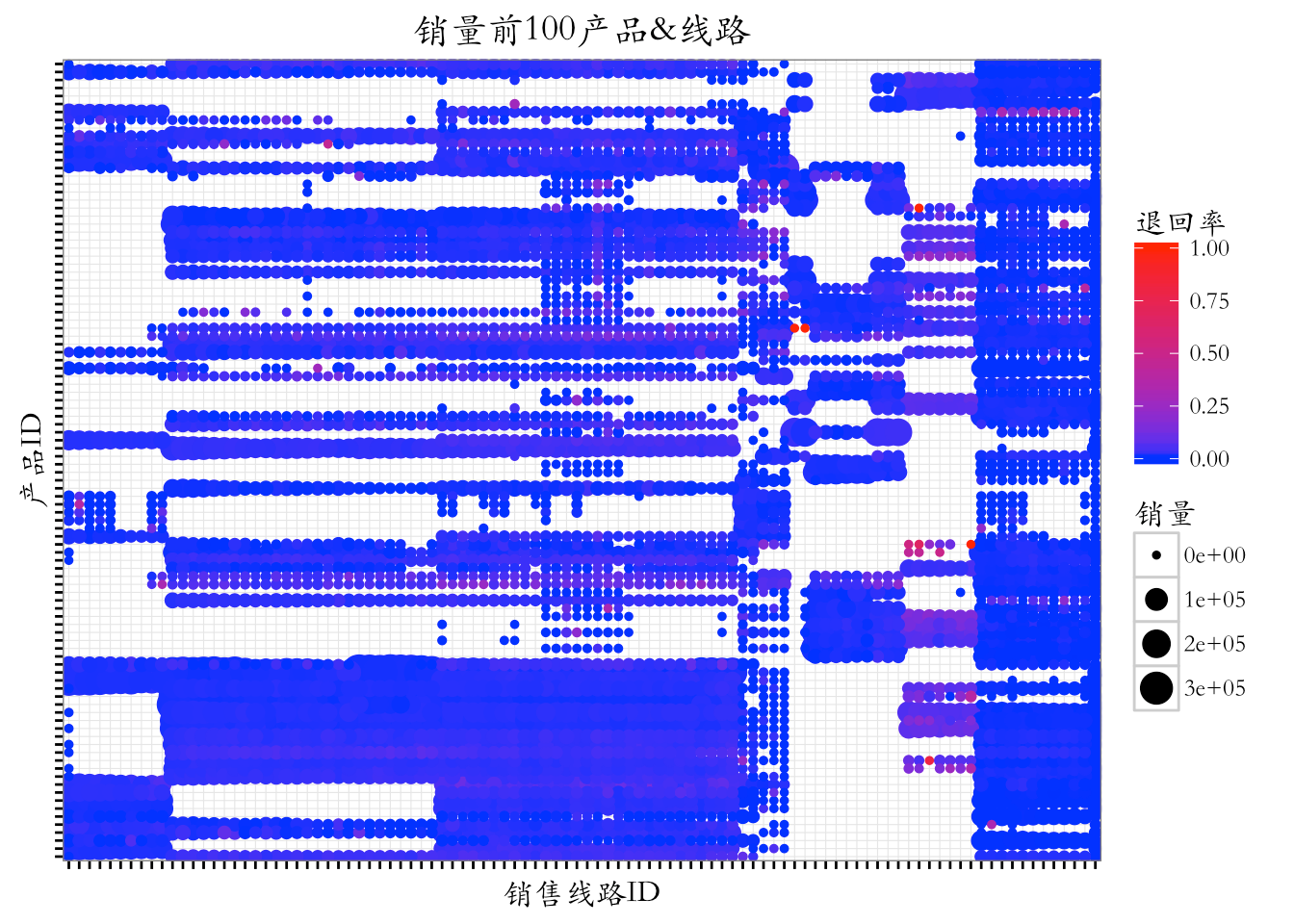
分析：大部分的销售站产品数量都在100~200之内。

17、产品和销售路线



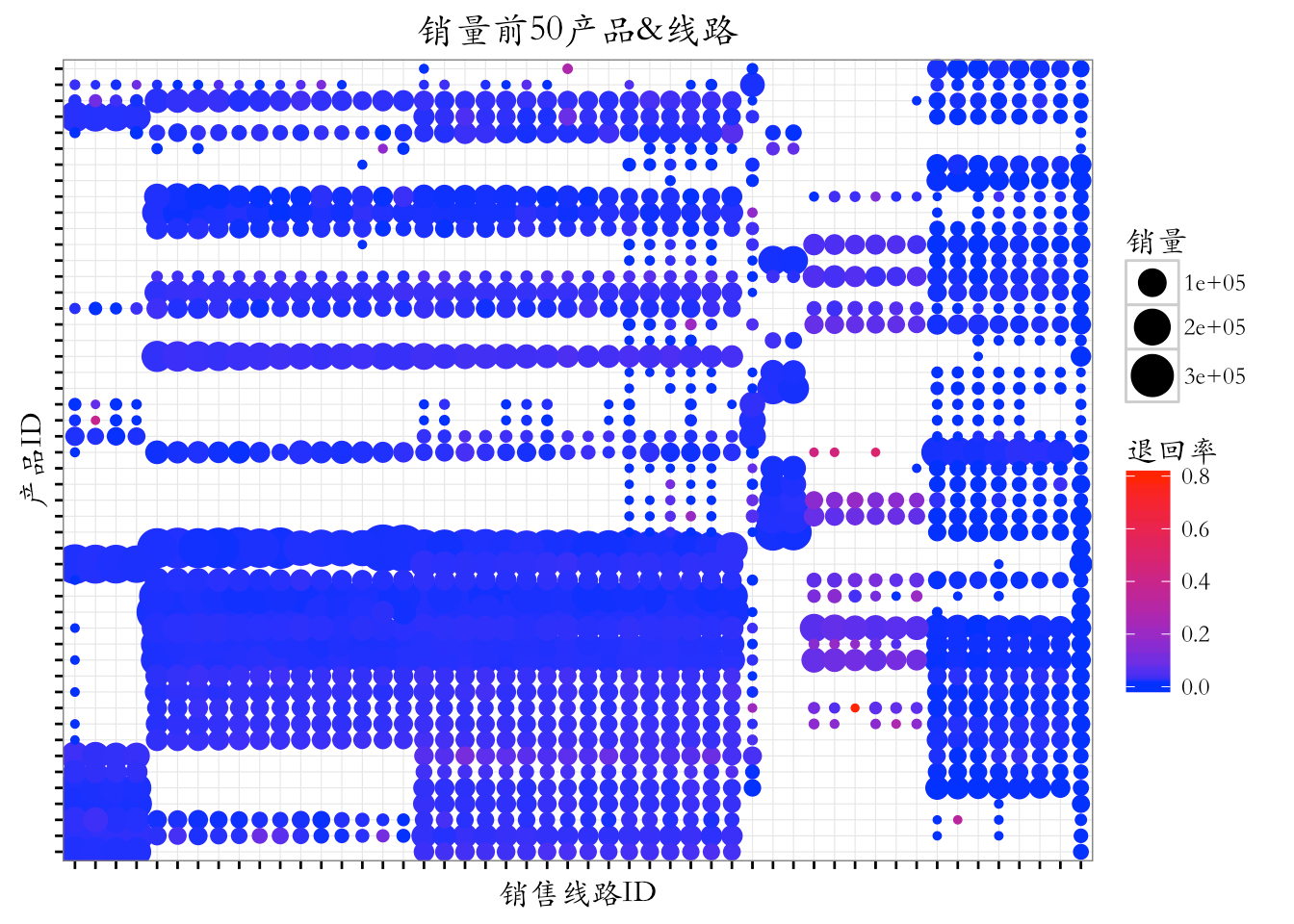
分析：大部分产品只有几条销售路线，只有几种产品的销售路线很多

产品和销售路线 top100



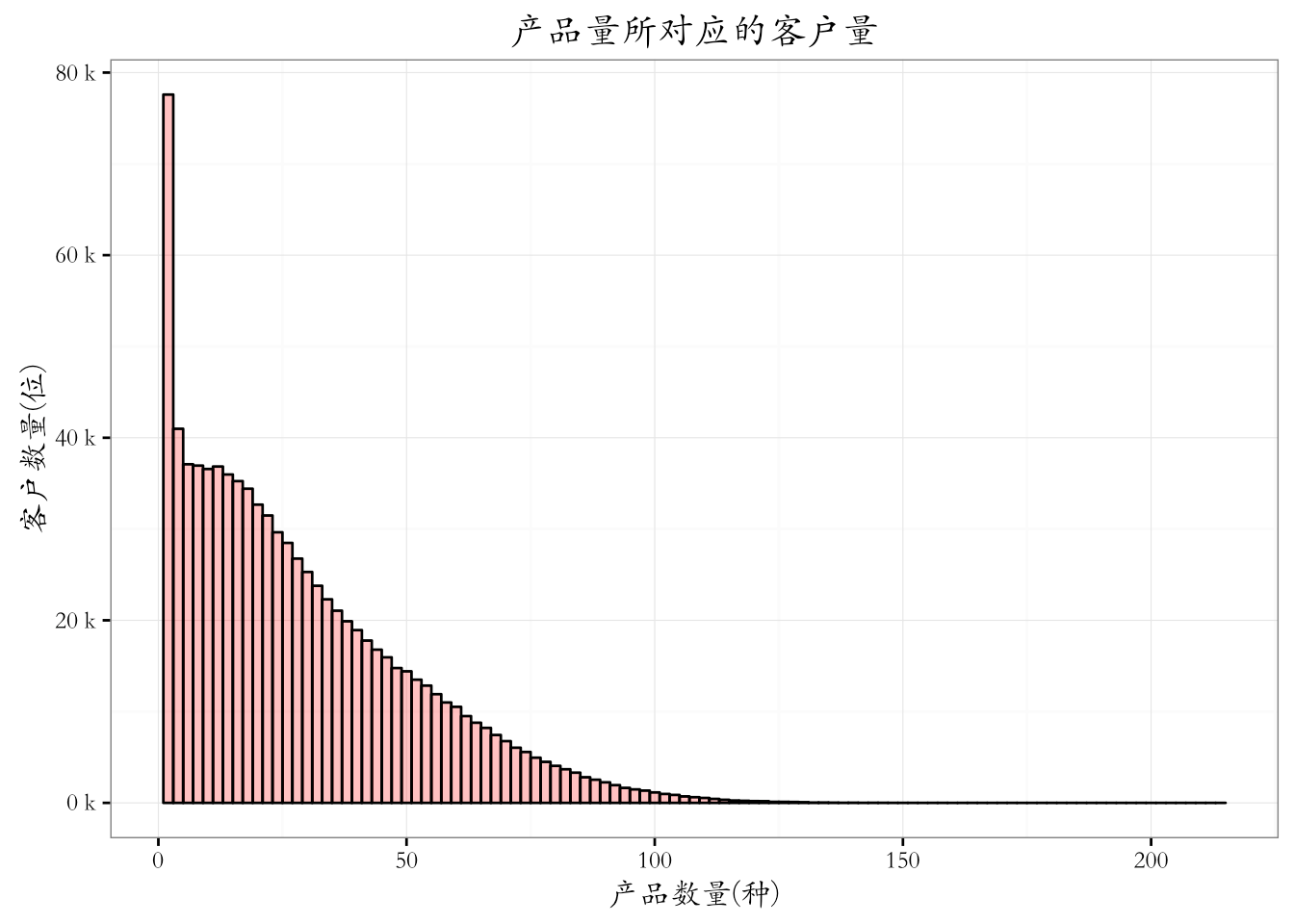
分析:图中每一个点都是对应的销售路线与产品ID的销量，其中点的大小表示销量的多少，点分布越多的地方销量越高；点的颜色表示退回的百分比，由红到蓝退回百分比递减。销量前100的点的分布较为均匀，无法得出体现其中内在规律的结论。继续分析产品和销售路线前50的销量及返回率。

产品和销售路线 top50



分析：从图中可以看出，在图的左下侧点的分布较为密集，点较大，颜色偏蓝，说明销售量大相应的退回百分比较小。与销量前50的销售站&销售路线分析的结果一致。

18、产品和客户



分析：从图中可以看出，随着产品数量的增多，客户量逐渐减少。产品数量小于10时客户数量最多。汪嫂得出结论，物以稀为贵。