

# 飞车玩家分类探索

飞车项目组

**Fisher**

# 研究思路

- ◆ 为什么做——分类初衷
- ◆ 拿什么做——分类依据
- ◆ 该怎么做——如何分类
- ◆ 做成了吗——分类结果
- ◆ 做有啥用——结果解释
- ◆ 还能做啥——改进空间
- ◆ 附录

# 为什么做？——精细化理念

认知

- 更加清楚哪些玩家具备哪些特征

运营

- 适当的**时间**适当的**场合**对适当的玩家推送适当的**活动**。

营销

- 适当的**时间**适当的**场合**对适当的玩家推送适当的**商品**。

# 拿什么做？——分类指标的寻找

- 原则：
1. 分类的指标能够与游戏内容相契合
  2. 尽可能少的指标，反映尽可能多的特征
  3. 不同的指标能够全面、独立地反映特征



# 拿什么做？——分类指标的寻找

## 游戏元素

- 开局数；游戏时长；炫点数；经验；酷币；冠亚季军次数；合作值；荣誉值等。

## 社交元素

- 好友数；招募好友数；缘分天空情况等。

## 消费元素

- 魅力值；消费积分；点券；消费券；幸运币等。

## 情侣元素

- 舞力值；心动值等。

## 休闲元素

- 休闲时长；偷猪总经验；养鱼等。

# 拿什么做？——分类指标的寻找

将原始指标标准化并赋予不同权重生成分类新指标。

原始指标

新生指标

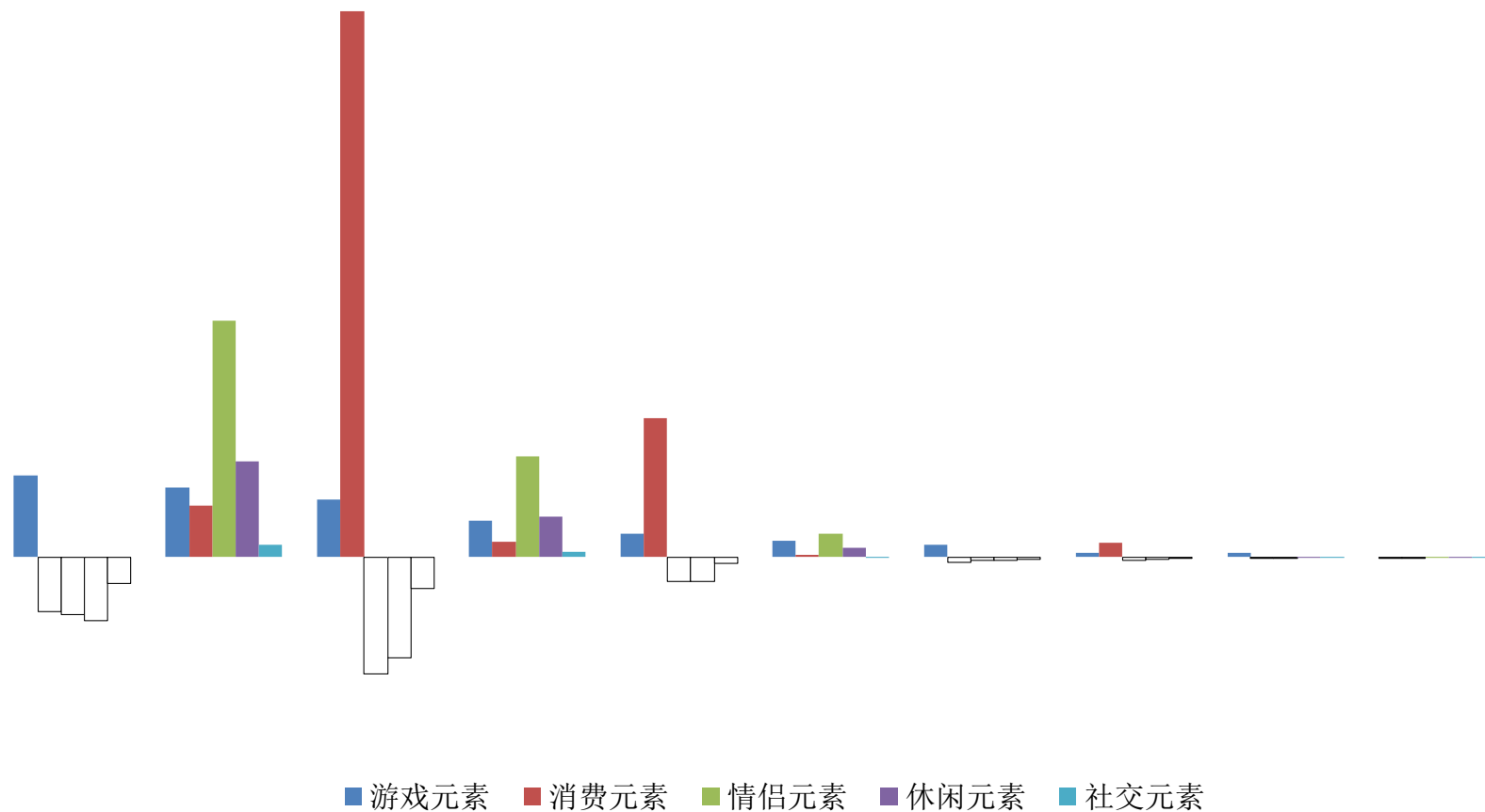
	游戏	消费	情侣	休闲	社交
经验	.704	-.203	.201	-.246	-.095
酷币	.847	-.210	.195	-.173	-.061
开局数	.908	-.117	.159	-.060	.023
已休闲时长	.116	.035	.340	.674	-.111
总在线时长	.055	-.007	.255	.624	-.380
总游戏时长	.957	-.085	.003	.016	.029
魅力值	.428	.771	.105	-.113	-.016
舞力值	.552	.113	-.700	.279	.033
心动值	.651	.082	-.625	.244	.020
合作值	.804	-.195	.074	-.134	-.086
炫点数	.768	.005	.123	.091	.040
好友数	.047	-.031	.077	.083	.705
累计消费积分	.248	.846	.174	-.129	-.039
招募集合大小	.157	.033	.258	.309	.594

新生指标相当于原始指标的线性组合；  
新生指标之间无关联，每一指标独立反映某一特征。

# 该怎么做？——分类结果数据展示

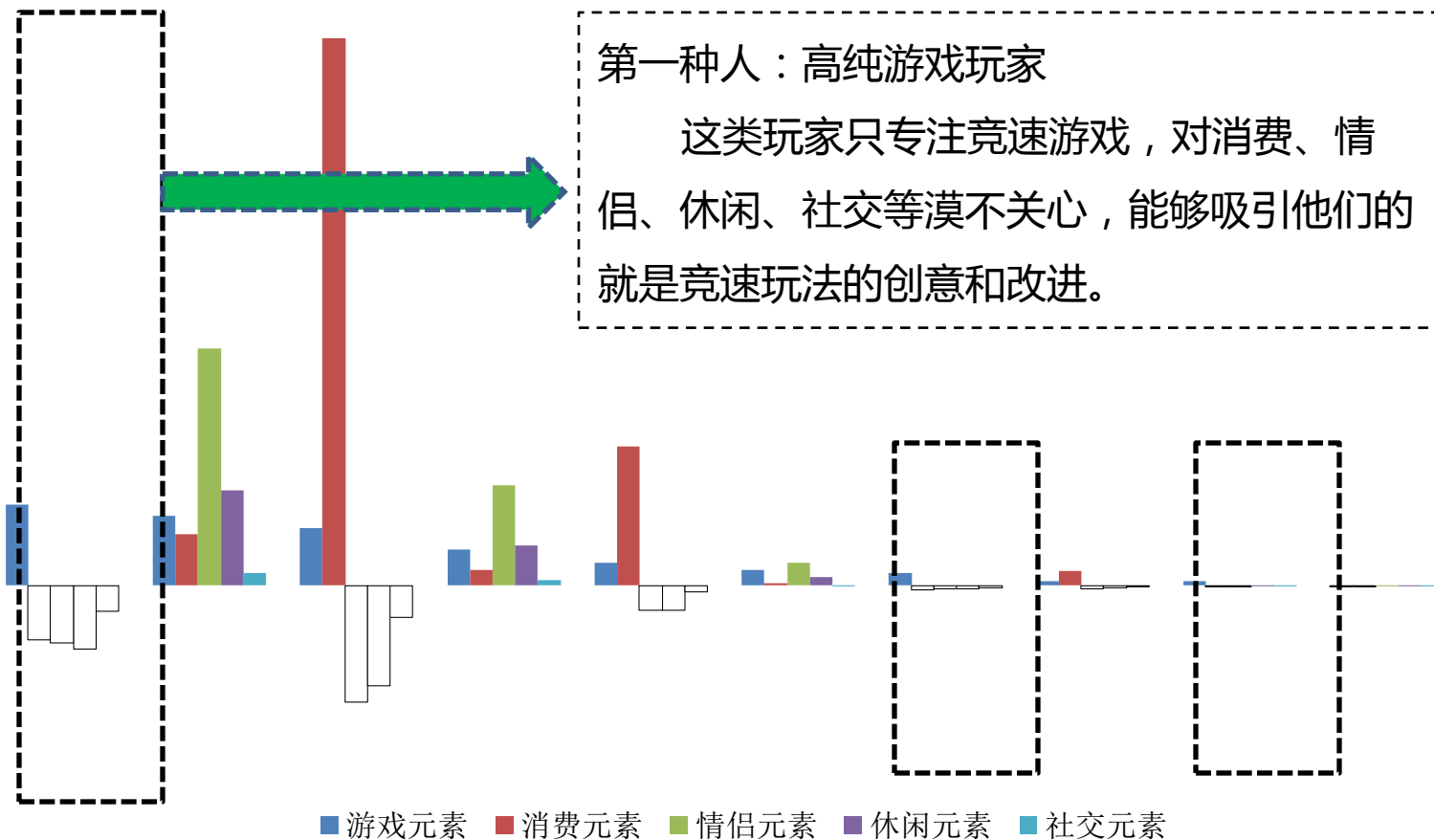
利用前述五个指标对随机样本数据（样本量：**678469**）进行聚类，其结果如下：

类别	1	2	3	4	5	6	7	8	9	10
数量	2	20	1	25	10	484	1709	1561	21151	653506



# 做有啥用？——分类结果数据解释

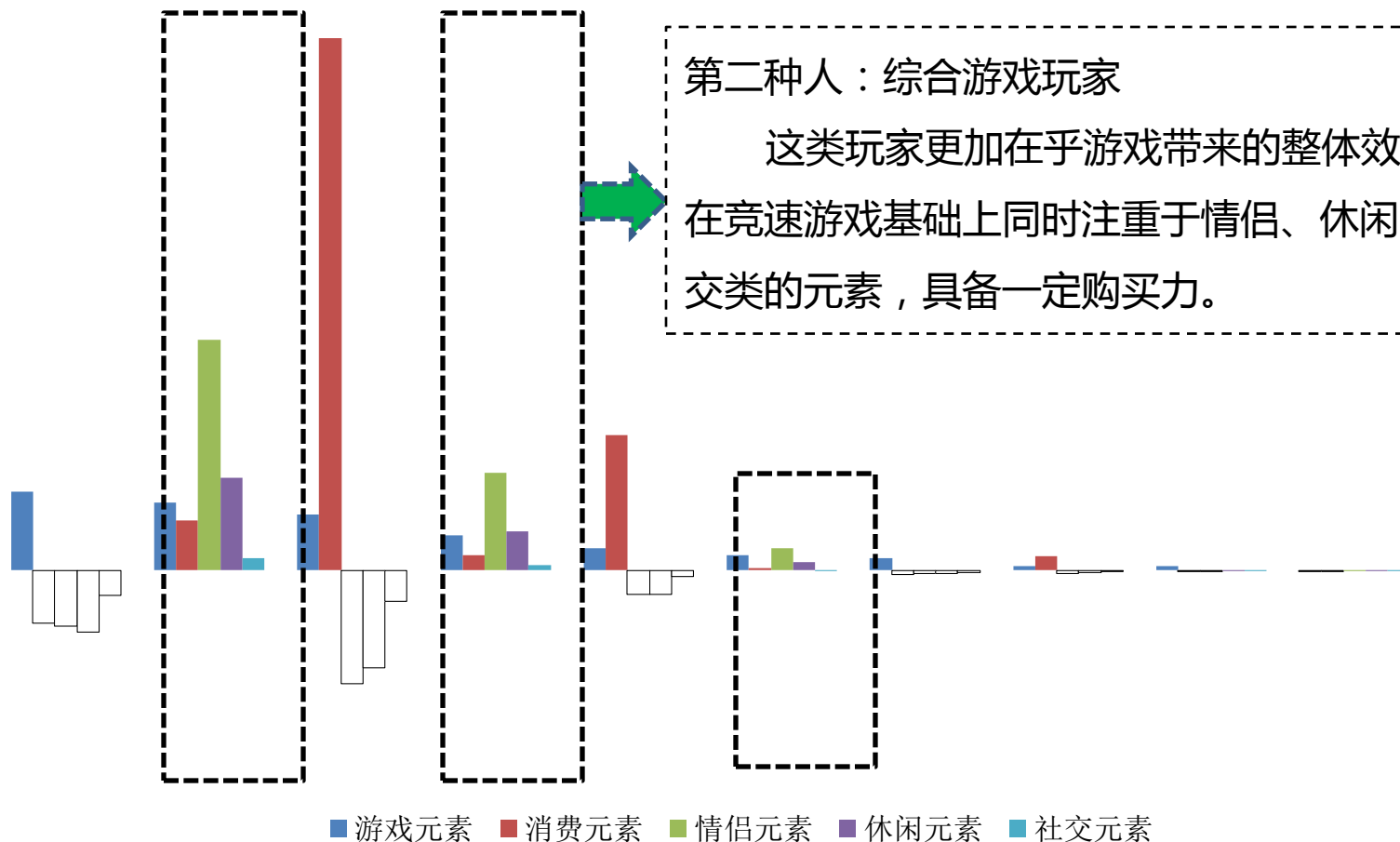
类别	1	2	3	4	5	6	7	8	9	10
数量	2	20	1	25	10	484	1709	1561	21151	653506





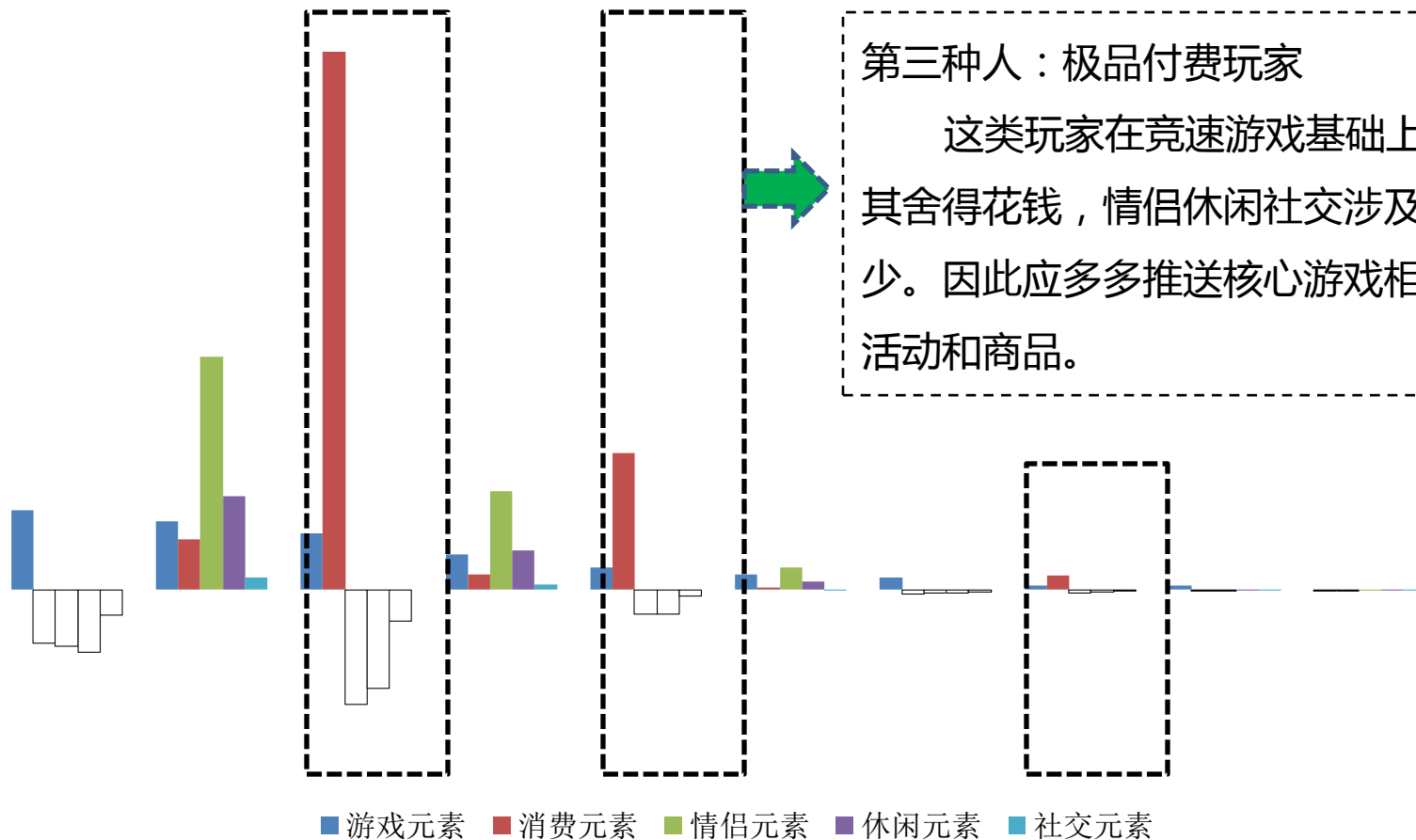
# 做有啥用？——分类结果数据解释

类别	1	2	3	4	5	6	7	8	9	10
数量	2	20	1	25	10	484	1709	1561	21151	653506



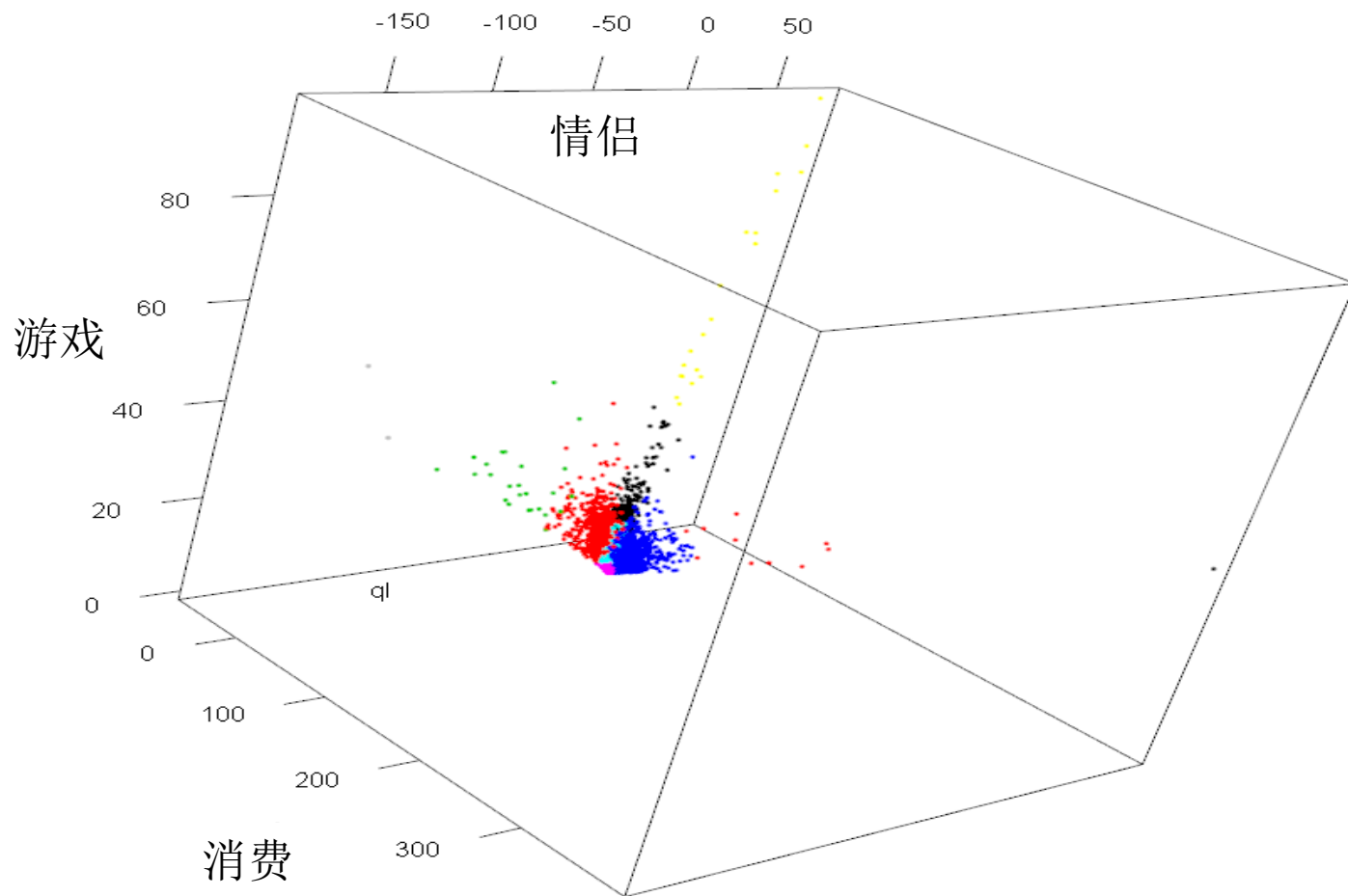
# 做有啥用？——分类结果数据解释

类别	1	2	3	4	5	6	7	8	9	10
数量	2	20	1	25	10	484	1709	1561	21151	653506



# 做成了啥？——分类结果空间展示

选择游戏、消费、情侣三指标，其在空间中的结构为：

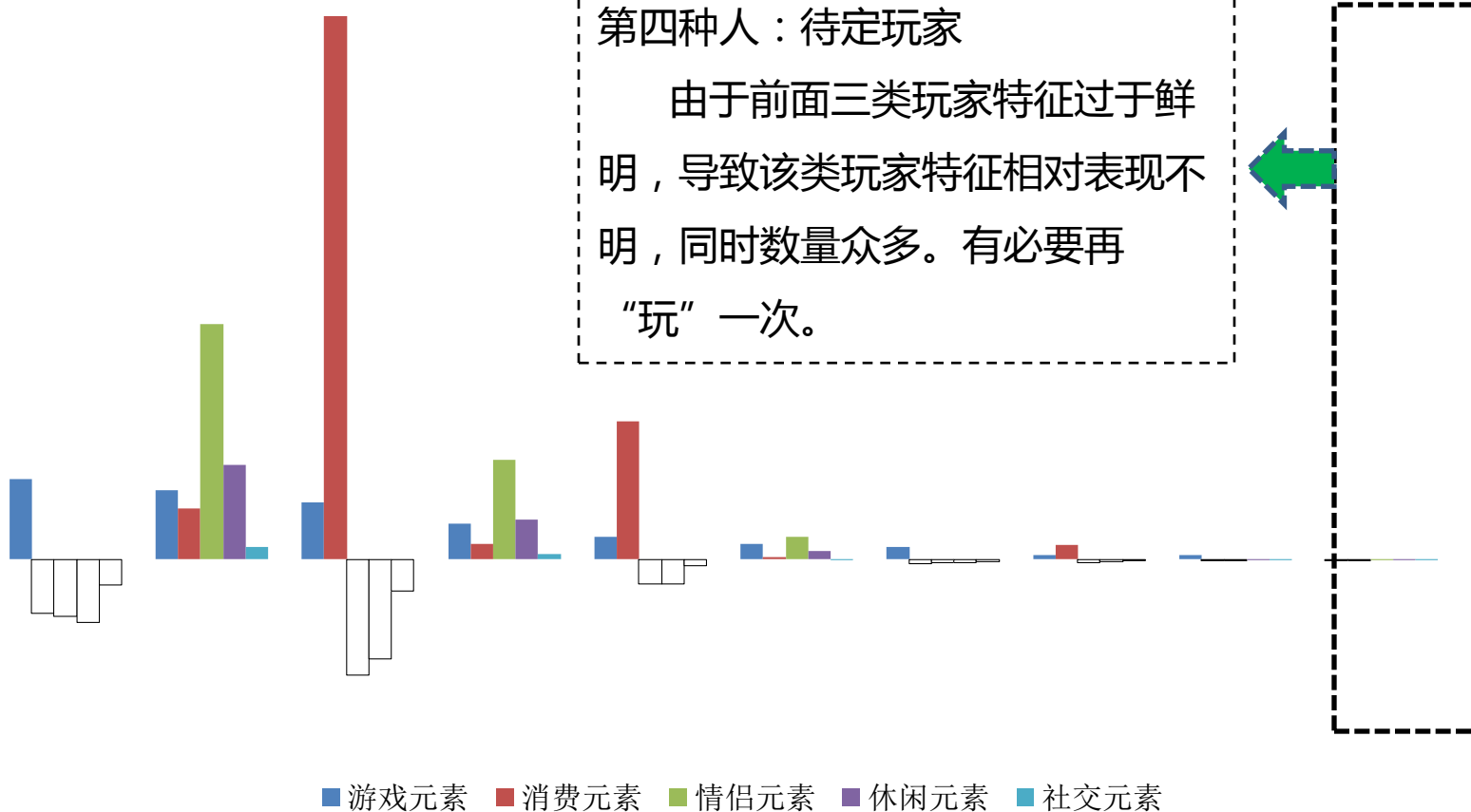


# 做有啥用？——分类结果数据解释

类别	1	2	3	4	5	6	7	8	9	10
数量	2	20	1	25	10	484	1709	1561	21151	653506

## 第四种人：待定玩家

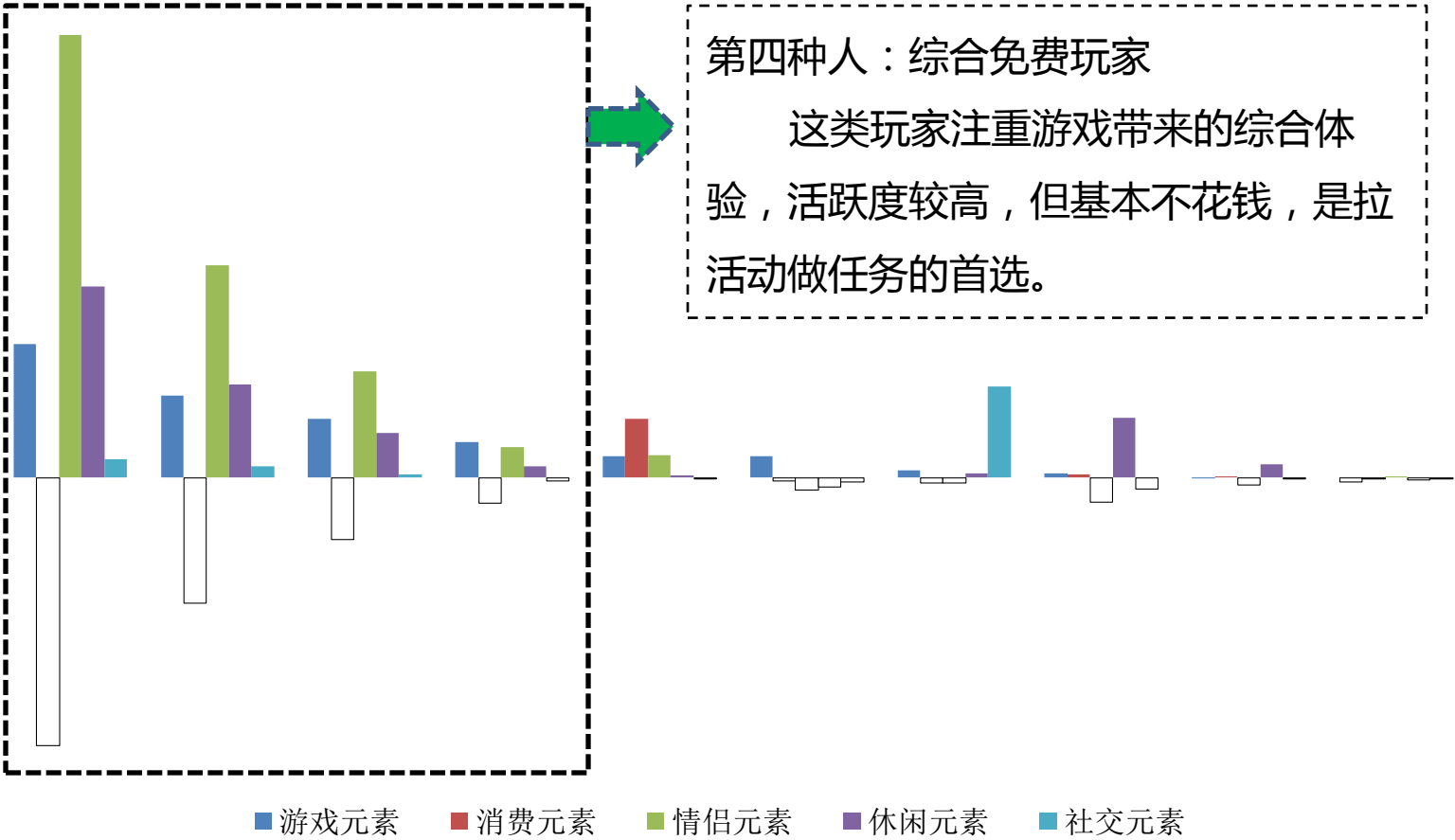
由于前面三类玩家特征过于鲜明，导致该类玩家特征相对表现不明，同时数量众多。有必要再“玩”一次。



# 做有啥用？——二次分类结果数据展示

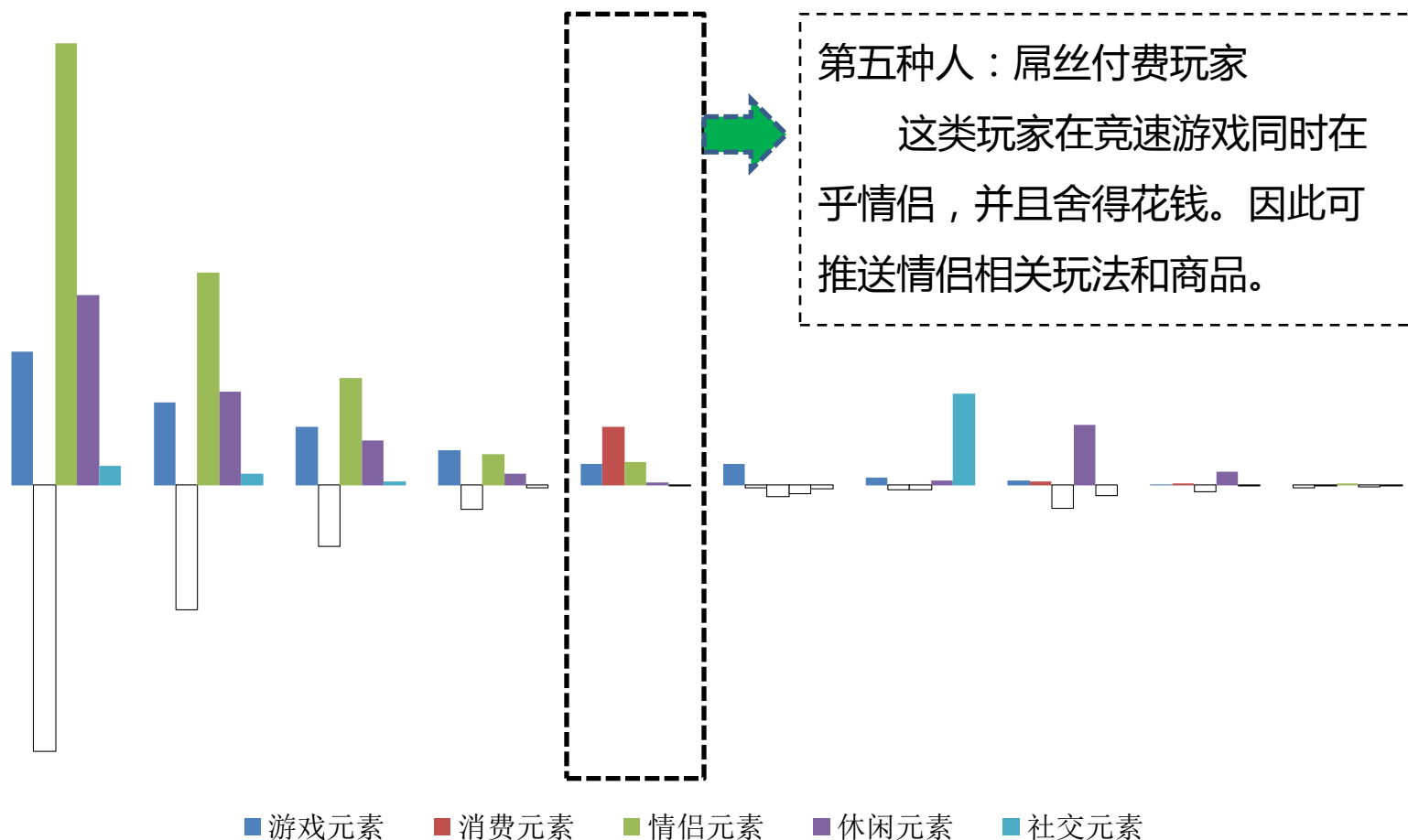
利用五个指标对上述**第10类**样本数据再次进行聚类，其结果如下：

类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782



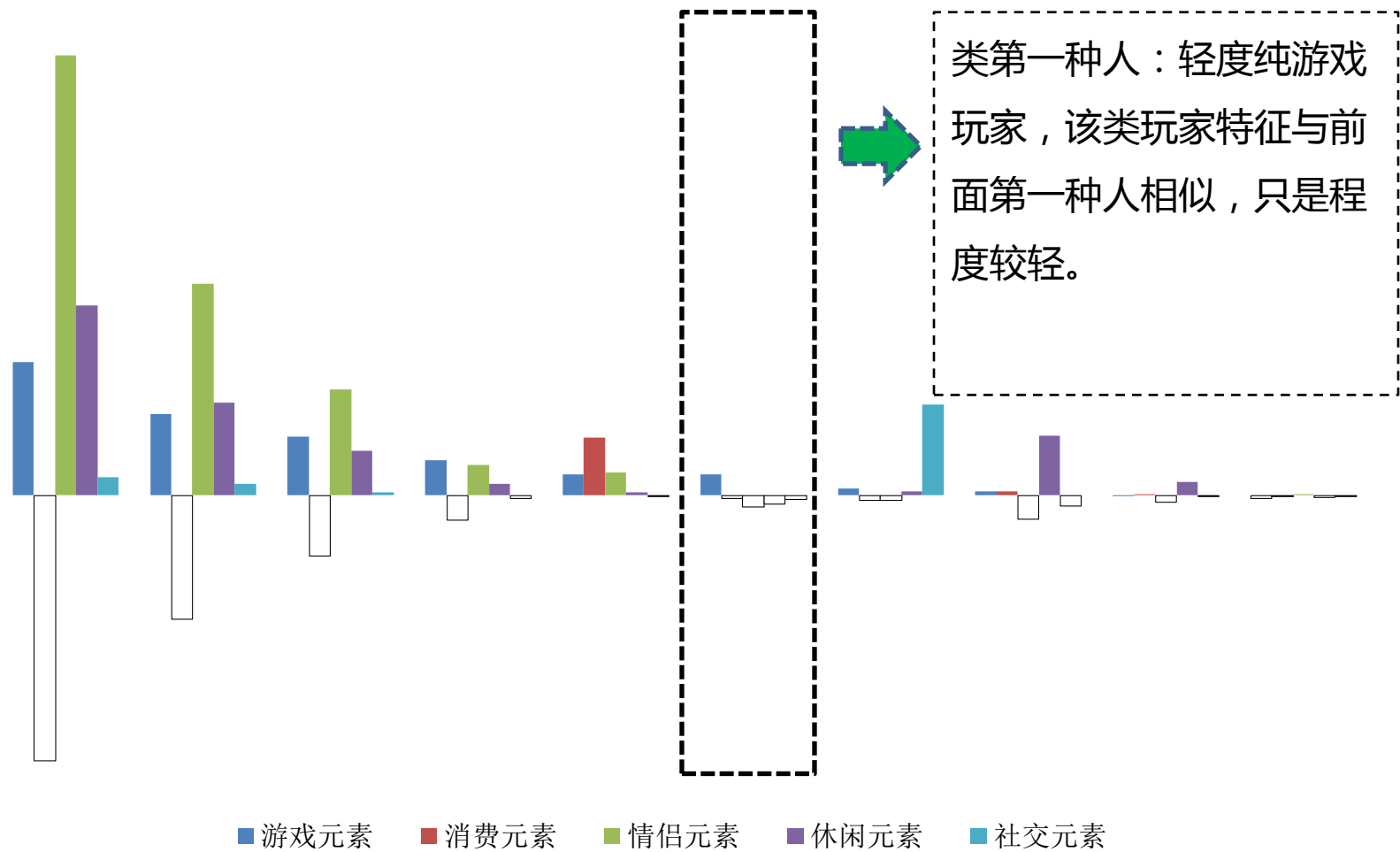
# 做有啥用？——二次分类结果数据展示

类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782



# 做有啥用？——二次分类结果数据展示

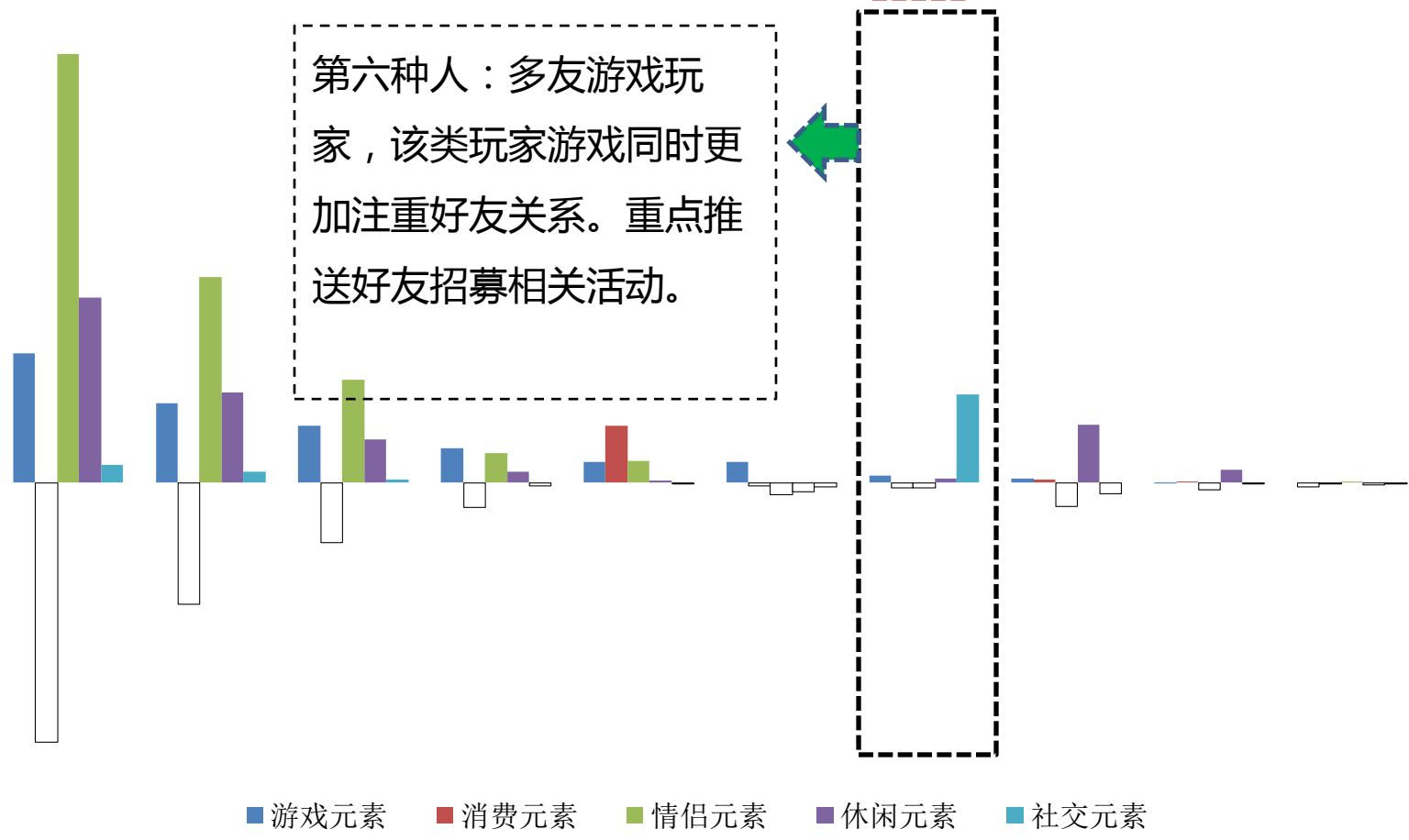
类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782



# 做有啥用？——二次分类结果数据展示

利用五个指标对上述第10类样本数据再次进行聚类，其结果如下：

类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782

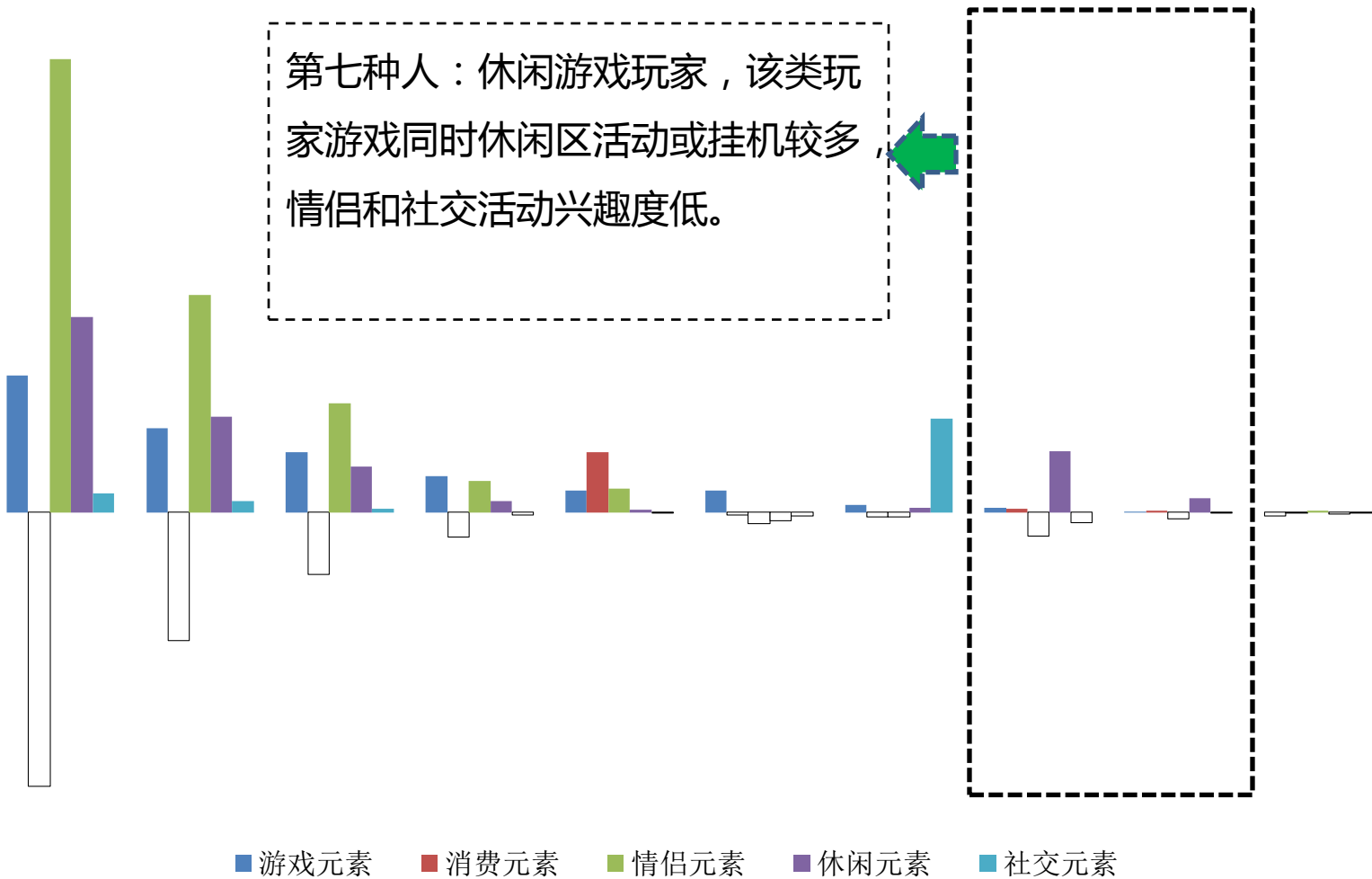




# 做有啥用？——二次分类结果数据展示

类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782

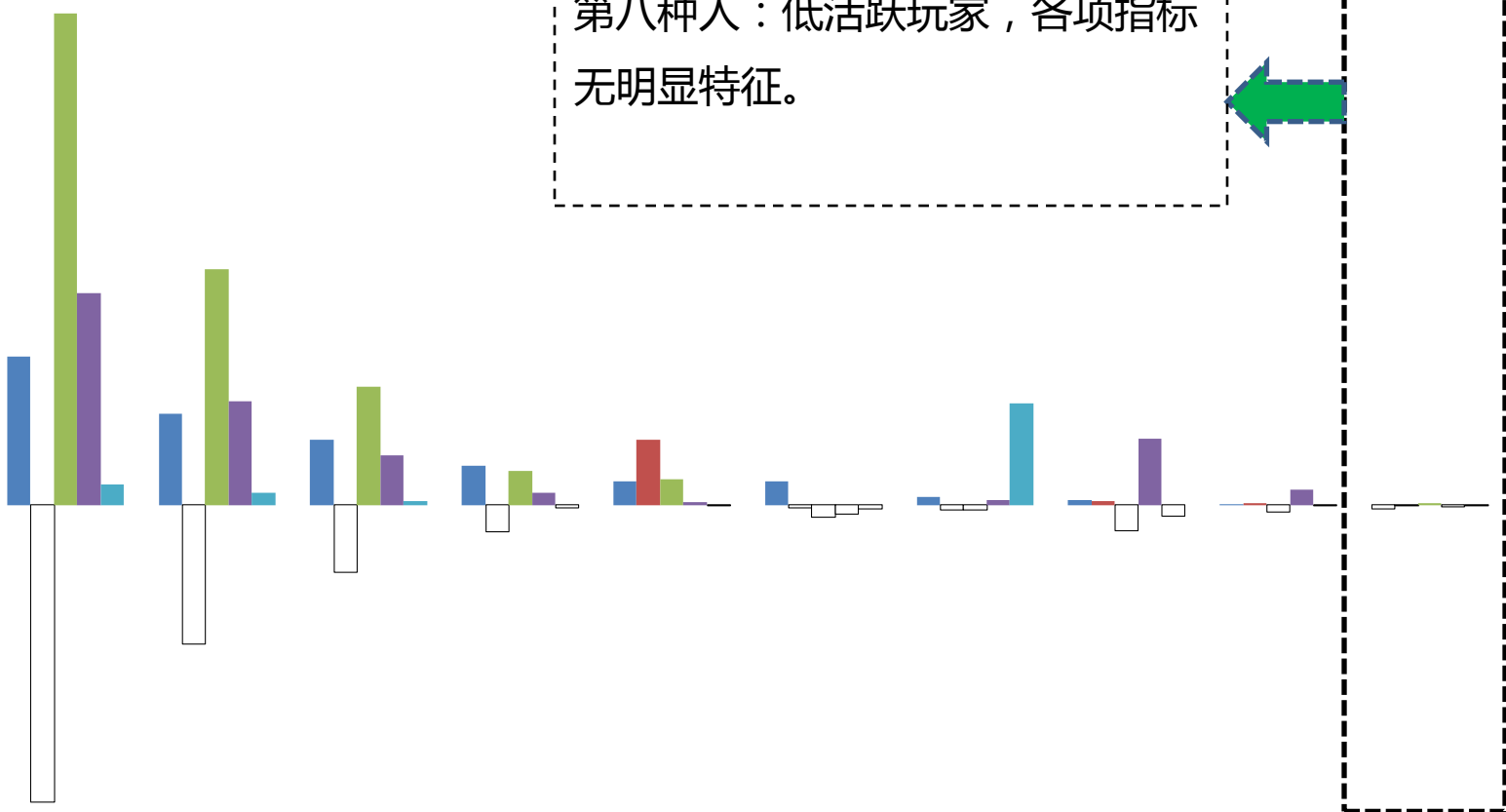
第七种人：休闲游戏玩家，该类玩家游戏同时休闲区活动或挂机较多，情侣和社交活动兴趣度低。



# 做有啥用？——二次分类结果数据展示

类别	1	2	3	4	5	6	7	8	9	10
数量	11	239	963	5217	9498	55721	6506	6201	71368	497782

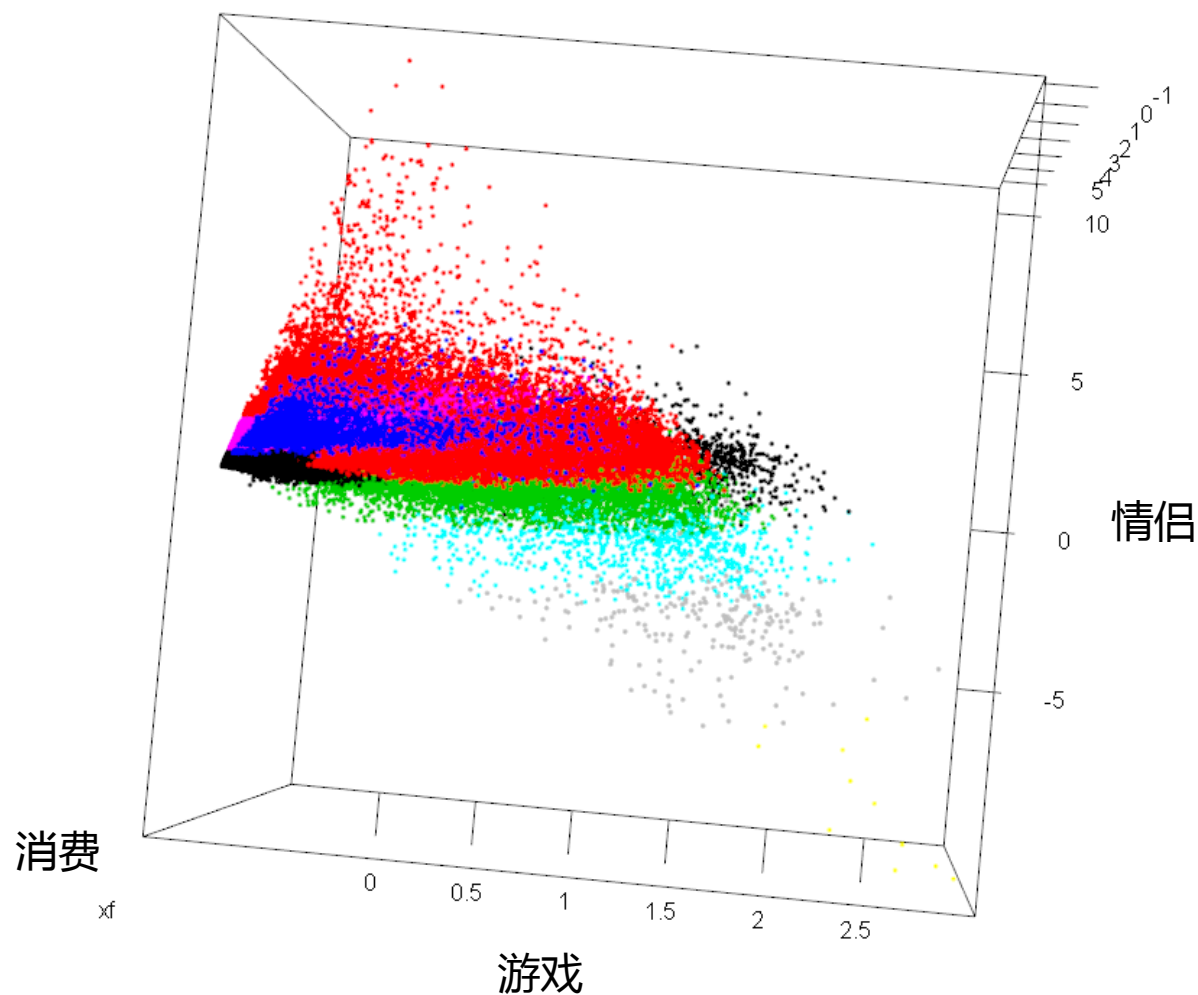
第八种人：低活跃玩家，各项指标无明显特征。



■ 游戏元素   ■ 消费元素   ■ 情侣元素   ■ 休闲元素   ■ 社交元素

# 做的结果？——二次分类结果空间展示

选择游戏、消费、情侣三指标，其在空间中的结构为：

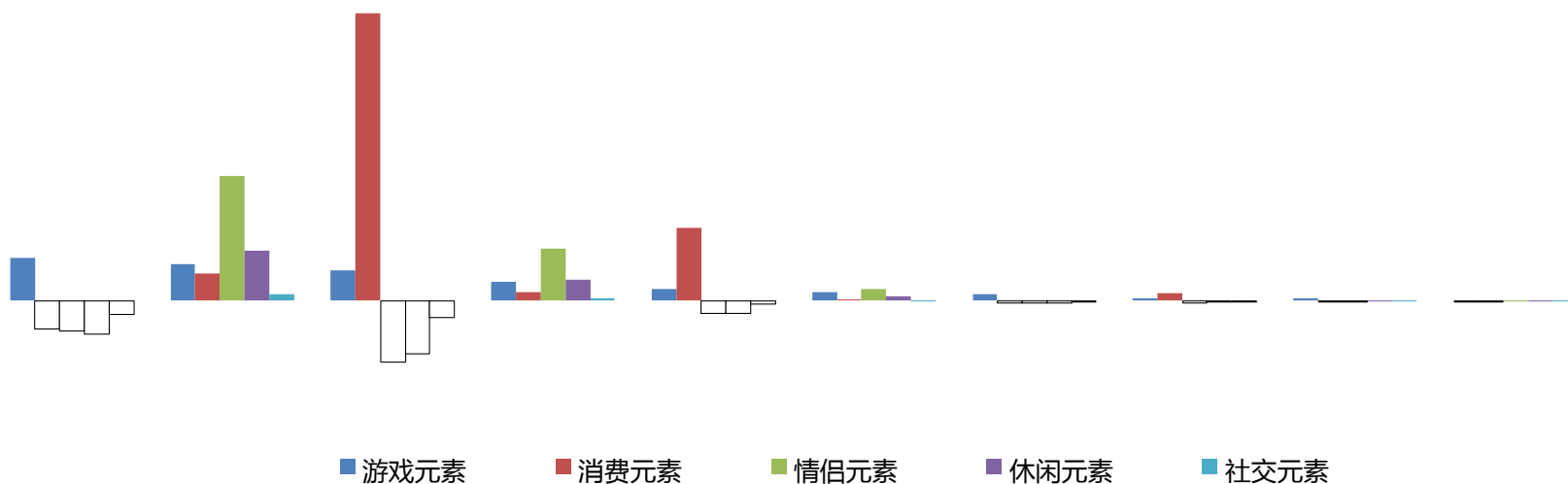


# 都做了啥？——总结

类别	竞速	消费	情侣	休闲	社交
第一种人	★★★★★	★	★	★	★
第二种人	★★★★	★★★★	★★★★★	★★★★★	★★★★
第三种人	★★★★	★★★★★	★	★	★
第四种人	★★★	★	★★★	★★★	★★★
第五种人	★★★	★★★	★★★	★★	★
第六种人	★★	★	★	★★	★★★★
第七种人	★★	★★	★	★★★	★
第八种人	★	★	★	★	★

# 分类准确性验证

为了进一步验证上述分类效果，现选取一段时间各类玩家消费金额数据，计算每类玩家平均消费金额，与分类结果中各类消费元素数值基本保持一致。



不同类别玩家平均每人消费情况

# 后续做啥？——改进空间

## ◆ 指标选择上

**精确性**：目前选取的原始指标是否能够评价玩家某一特征？

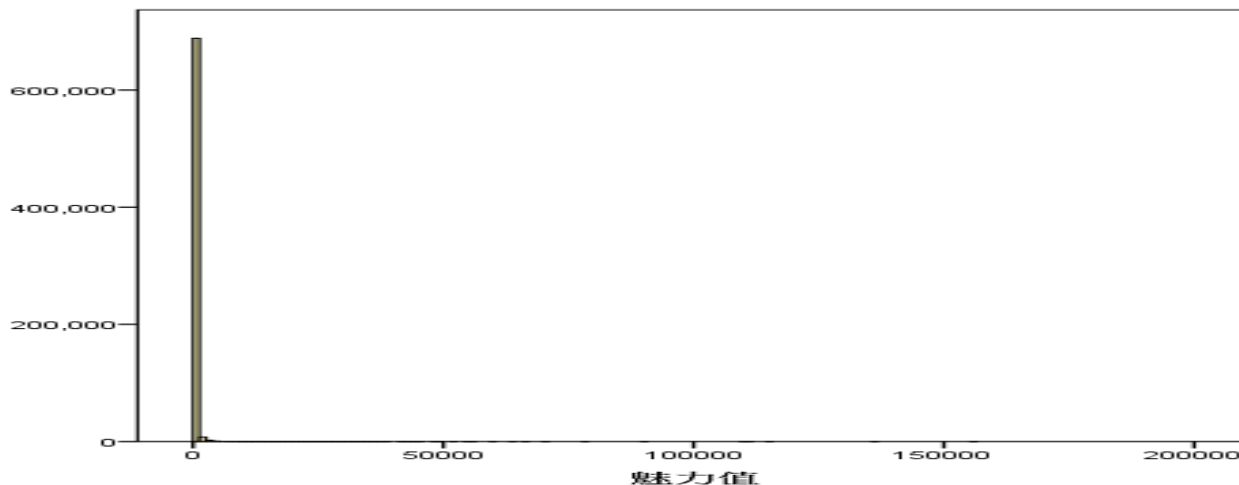
如消费指标，消费券、点券、幸运币由于月末清空机制，不能全面反映用户消费积累。

如反映休闲活动的养鱼、偷猪，只对部分用户起到评价作用，但不能作为全部用户评价指标。

**全面性**：目前选取这些原始指标是否全面，是否遗漏某些重要指标？

## ◆ 聚类方式上

由于游戏本身特征，几乎所有变量呈典型的右偏分布（二八定律），因此分类结果只能将少量极具特征的玩家挑选出来，即使采用二次分类，大部分普通玩家仍然归为一类。



# 附： R程序

```
aa=read.csv("D:\\专题-用户分群\\变量.csv",header=T)
yx=aa[,1]
xf=aa[,2]
ql=aa[,3]
xx=aa[,4]
sj=aa[,5]
lb=aa[,6]
lb2=aa[,7]
library( rgl )
plot3d(yx,xf,ql,col=lb) #第一次分类结果展示#
```

```
aa=read.csv("D:\\专题-用户分群\\变量2.csv",header=T)
yx=aa[,1]
xf=aa[,2]
ql=aa[,3]
xx=aa[,4]
sj=aa[,5]
lb=aa[,6]
lb2=aa[,7]
library( rgl )
plot3d(yx,xf,ql,col=lb2) #第二次分类结果展示#
```