

将影片的类型/语言/制片国家第一个取值提取出来，制作词云图，通过词云图我们可以得出如下结论：

影片量超过1000的影片类型分别为：**剧情/喜剧/动作/科幻/动画/悬疑/恐怖/音乐**；

影片量超过1000的影片语言分别为：**英语/汉语(普通话/粤语等)/日语**；

影片量超过1000的影片制片国家分别为：**美国/中国(大陆/台湾/香港等)/日本/英国**；

影片第一类型词云分布概况



影片第一语言词云分布图



电影第一制片国家词云分布图



通过柱状图可以得到如下结论：

影片评分>0的影片总共**36558**部，优秀和良共占比**14.5%**；影片评价人数1万以上的影片占比**13%**；影片片长**53.8%**落在了1.5-2小时分组区间内。

通过联动可知影片优分组中88.2%为五星人数百分比50%以上，1-3星人数100%落在20%以下，

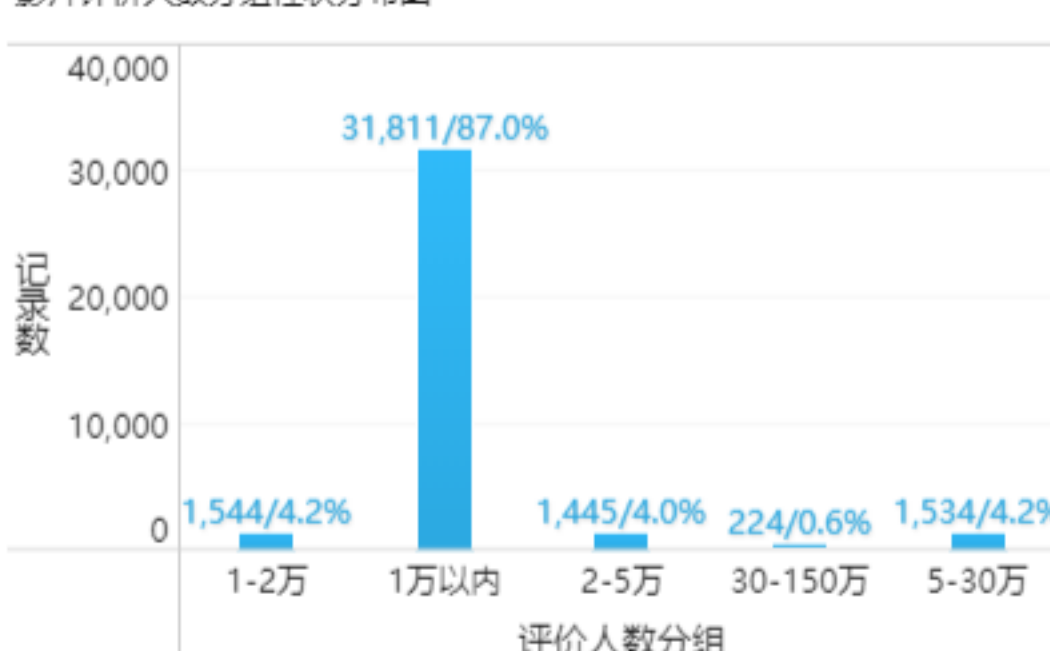
说明**5星人数占比**越高，越有可能成为优秀影片。因此**5星人数占比**是影响影片评分的重要因素之一。

通过联动可发现，随着评价人数的增多，成为优分组影片的概率越大，说明评价人数也是影响应聘评分因素之一。

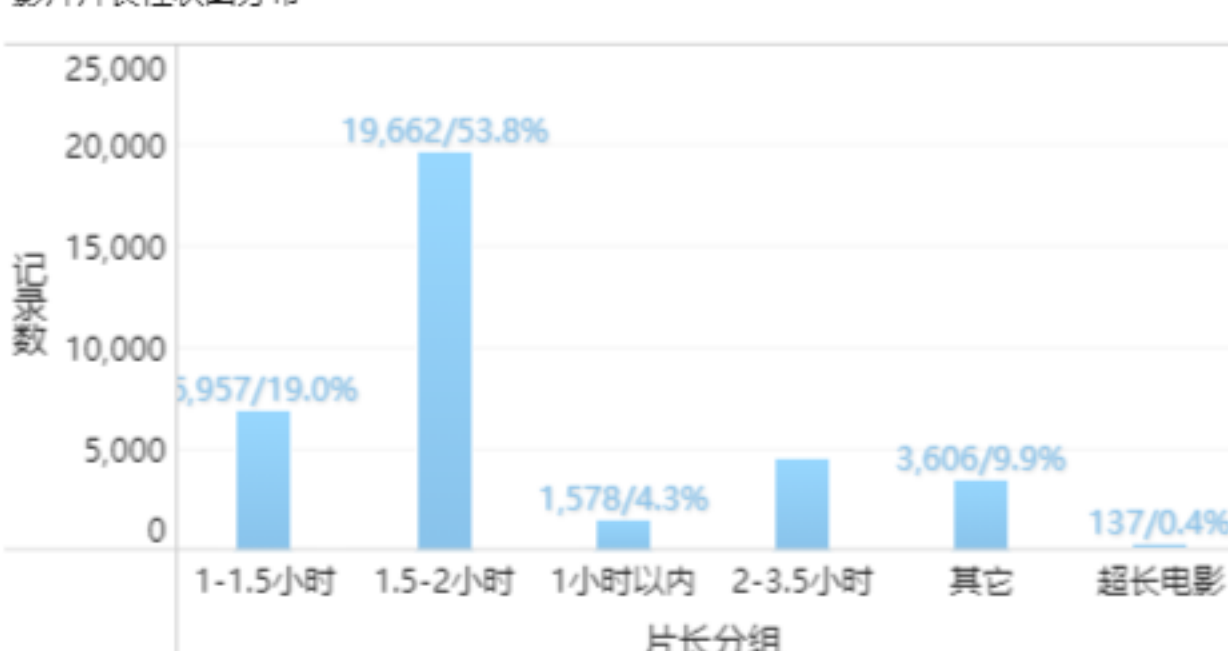
通过联动可发现，随着短评数量的增多，成为优分组影片的概率越大，说明短评数量也是影响应聘评分因素之一。

影片评分取值区间  < 值 <=

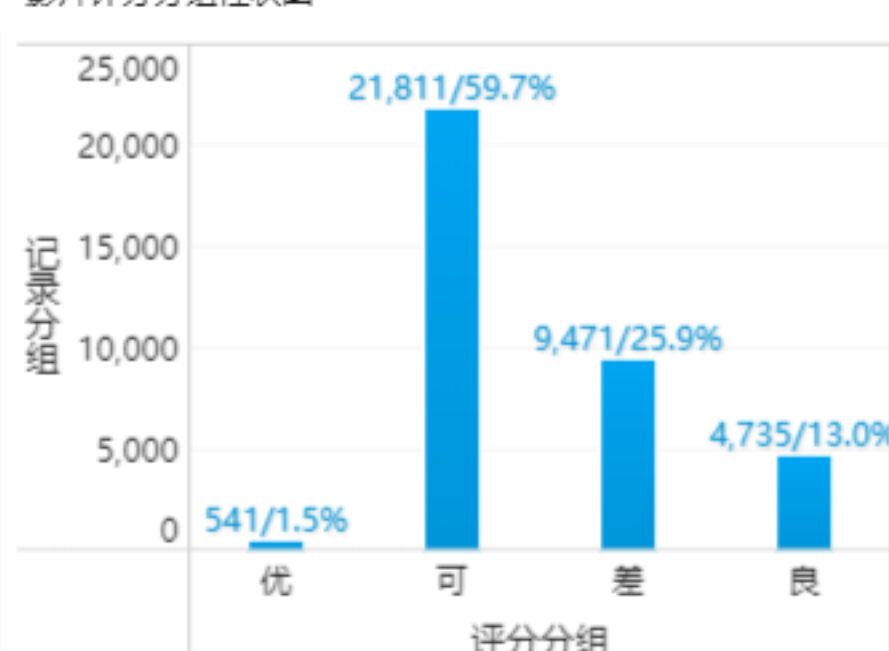
影片评价人数分组柱状图分布



影片片长柱状图分布



影片评分分组柱状图



因为分析目的为需求决定评分的因素，因此以下柱状图都会包含过滤条件(评分>0)。

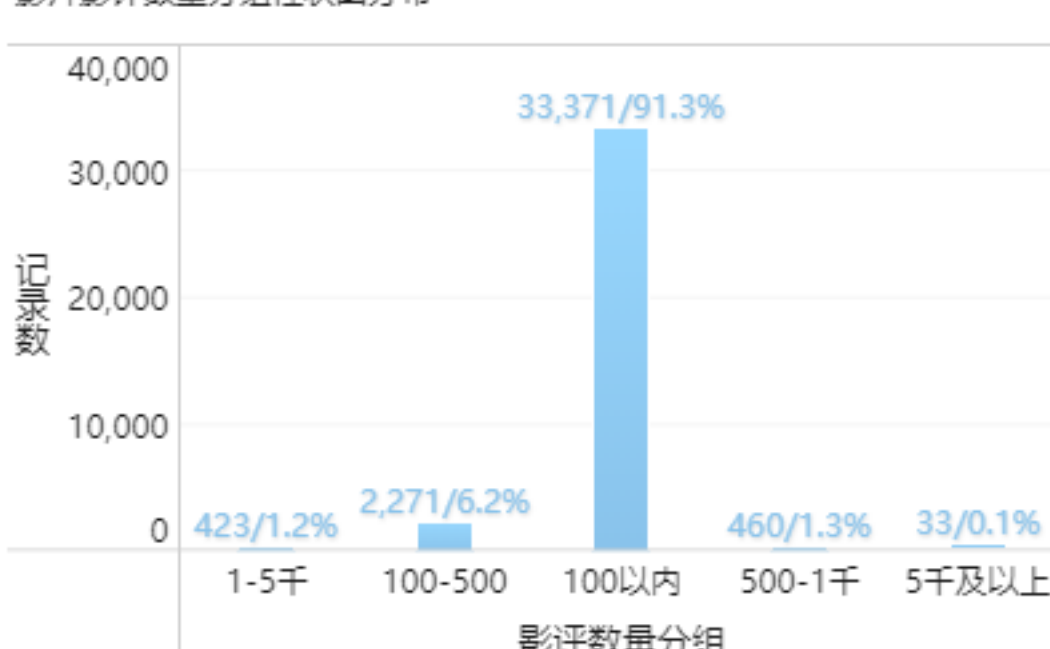
通过柱状图可以得到如下结论：

影片评分>0的影片总共**36558**部，影评数量100及以上的占**8.7%**；

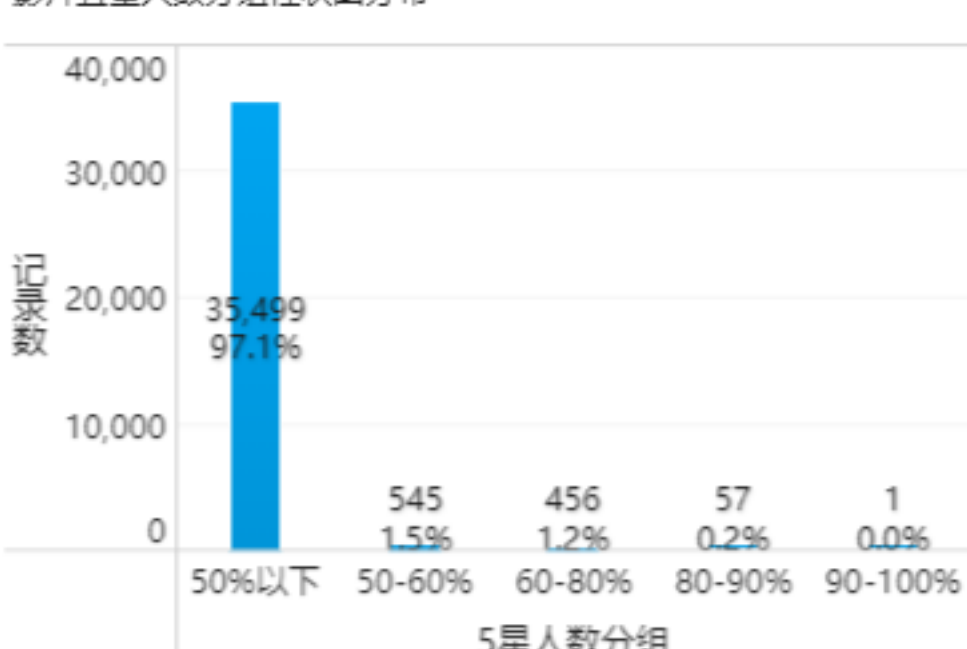
影片五星人数占总评人数50%及以上的占总影片**2.9%**；影片四星人数占总评人数50%及以上的占总影片**7.8%**。

电影评分取值区间  < 值 <=

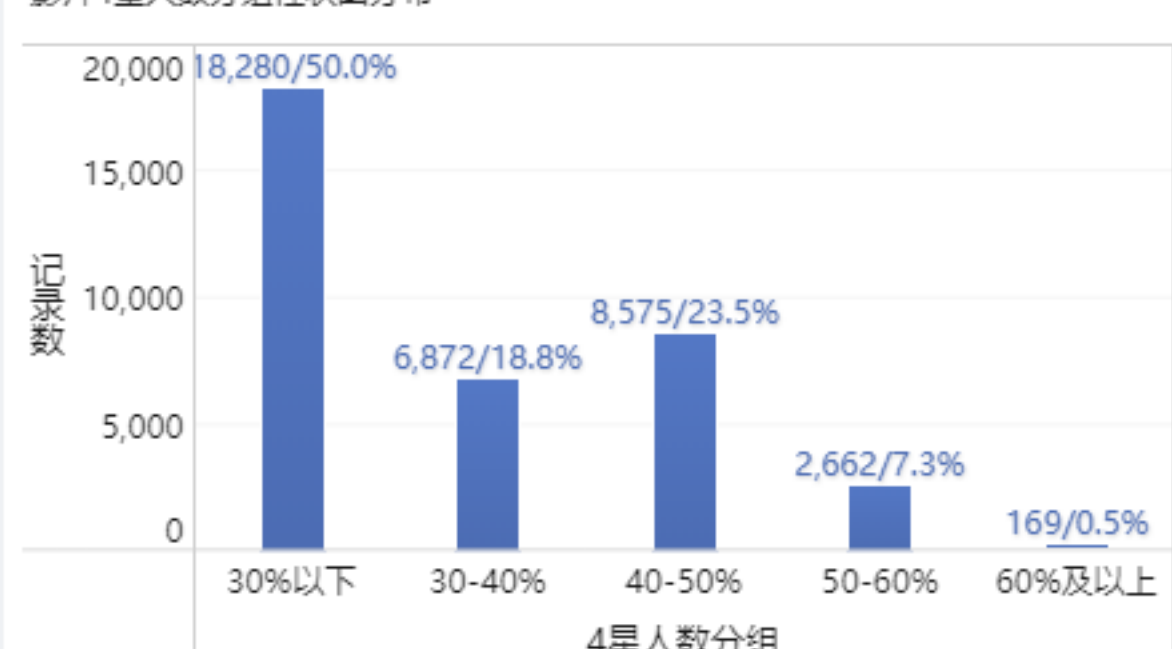
影片影评数量分组柱状图分布



影片五星人数分组柱状图分布



影片四星人数分组柱状图分布



因为分析目的为需求决定评分的因素，因此以下柱状图都会包含过滤条件(评分>0)。

通过柱状图可以得到如下结论：

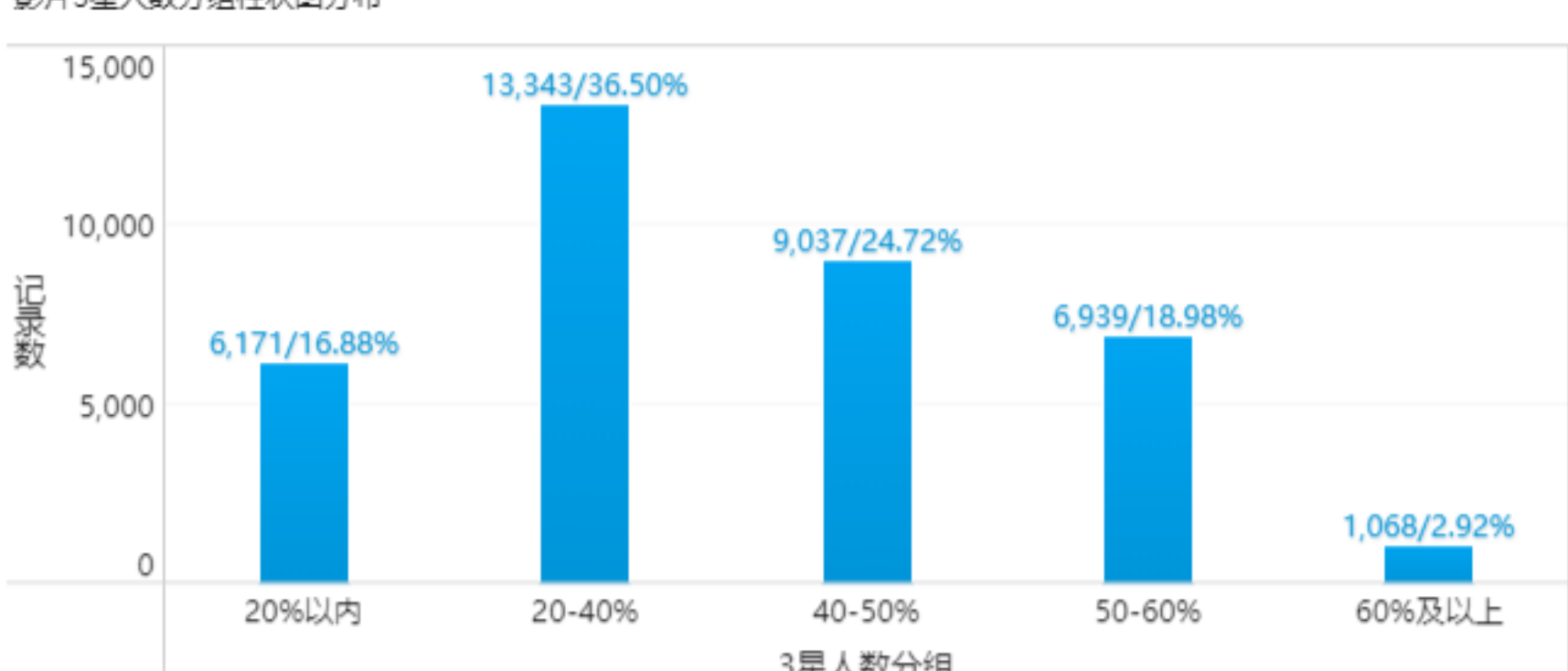
影片评分>0的影片总共**36558**部，影片三星人数占总评人数40%以内的超过总影片**50%**；

影片二星人数占总评人数40%及以上的占总影片**2.21%**；影片一星人数占总评人数20%及以上的占总影片**10.5%**；

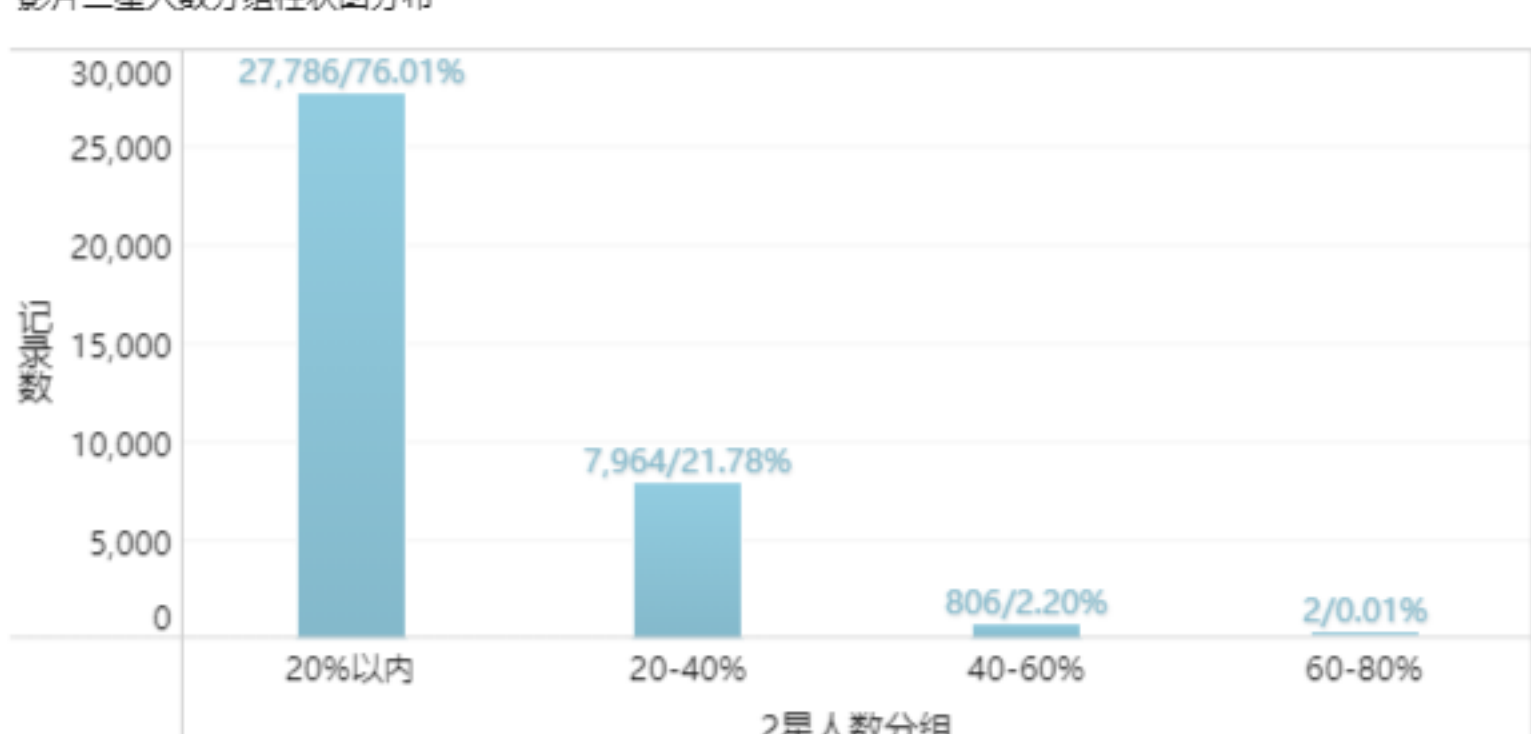
通过发布年份-影片数折线图可以看出2000年以前，电影发布数量逐年攀升，2000-2016年增速较快，且2016年达到高峰，21

评分取值区间  < 值 <=

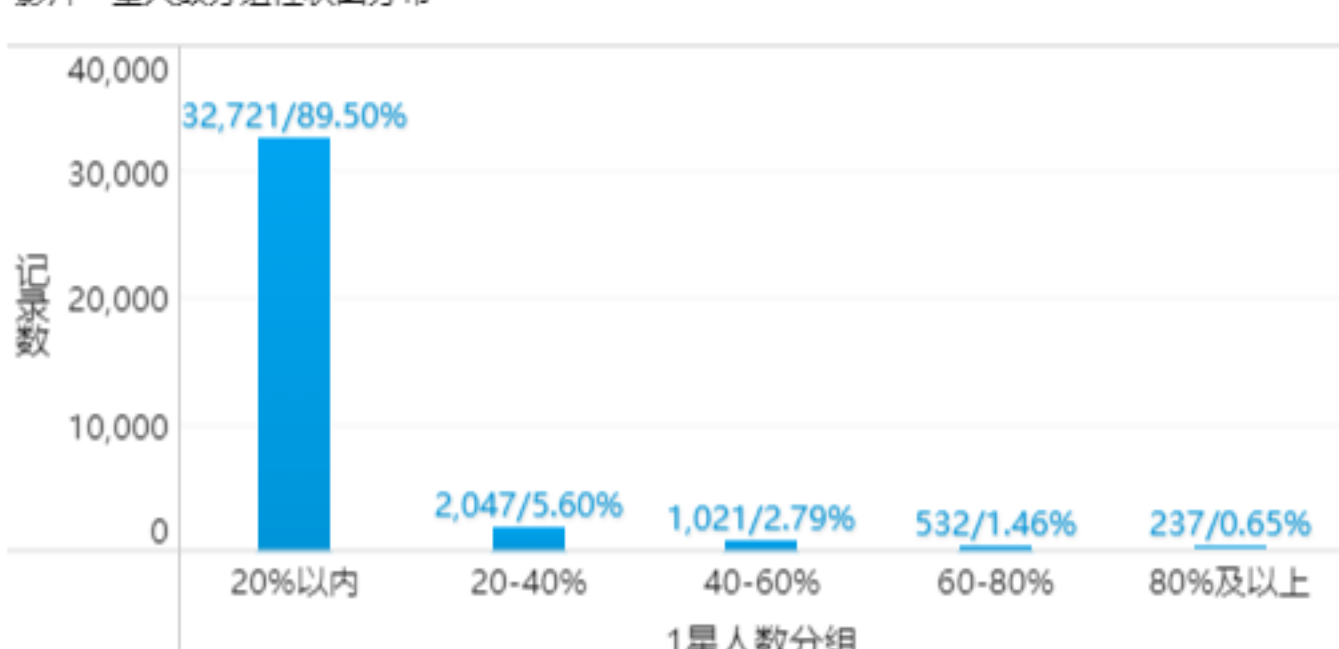
影片3星人数分组柱状图分布



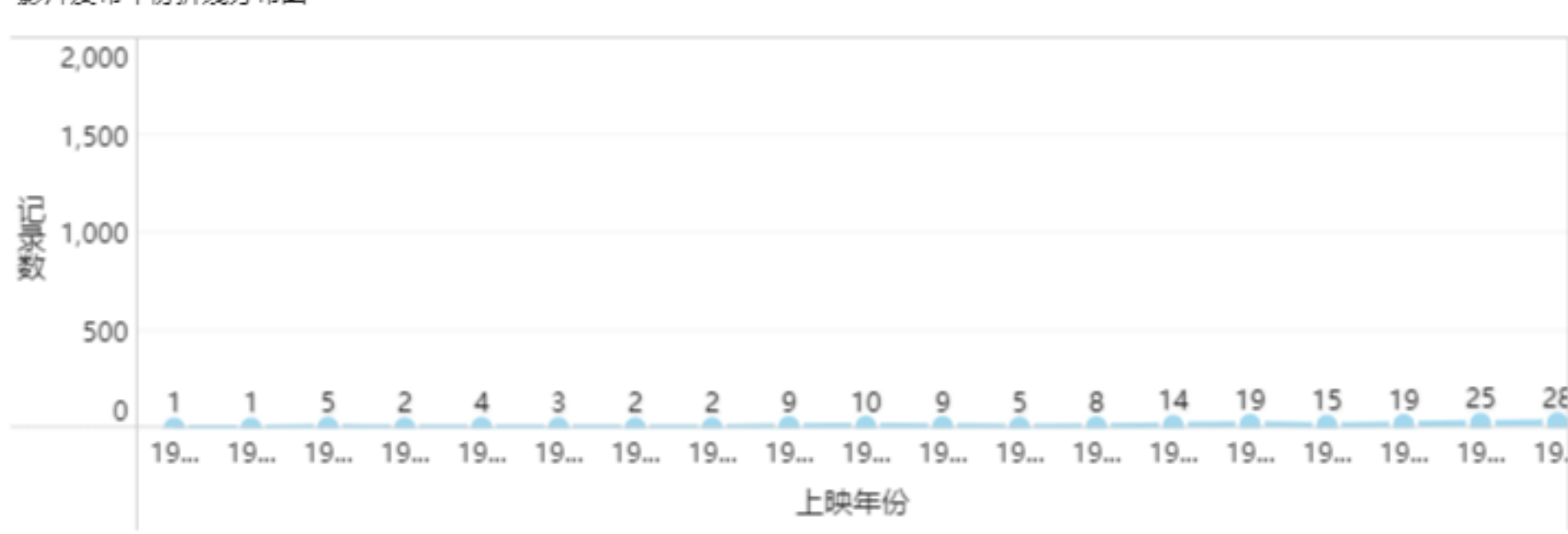
影片二星人数分组柱状图分布



影片一星人数分组柱状图分布



影片发布年份折线图分布



通过下面三张图可知：优秀影片从数量上看，前3的类型为**剧情/音乐/喜剧**；

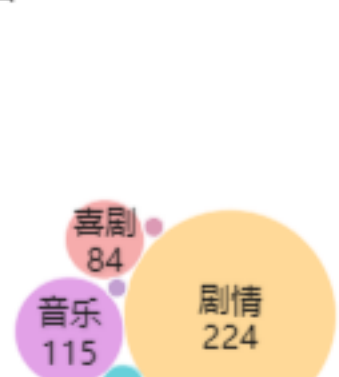
前三的语言为**英语/日语/汉语**；

前三的制片国家为**美国/日本/英国**。

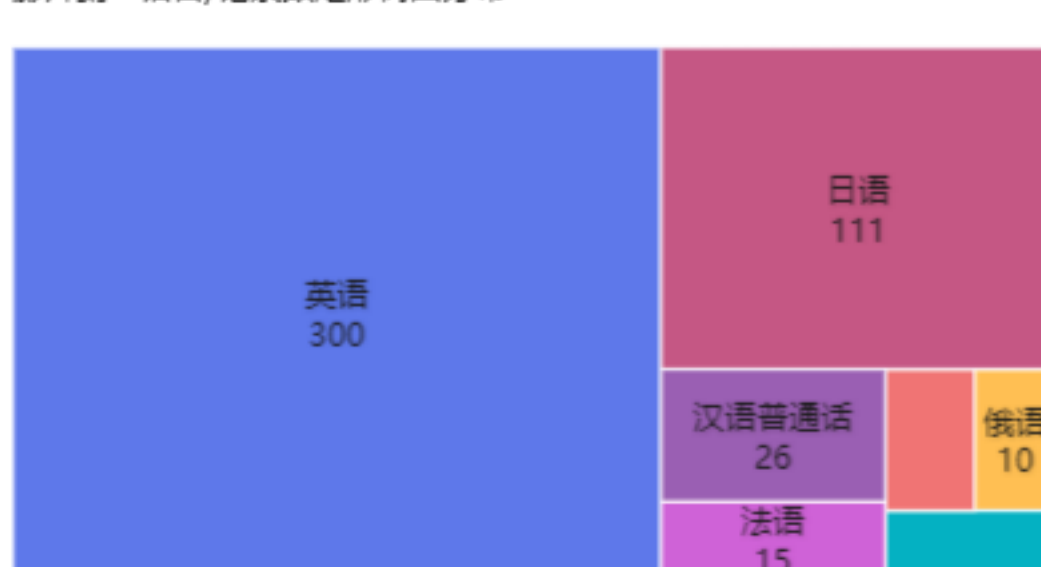
但并不意味着这些类型/语言/制片国家更有概率制作优秀影片。

评分分组

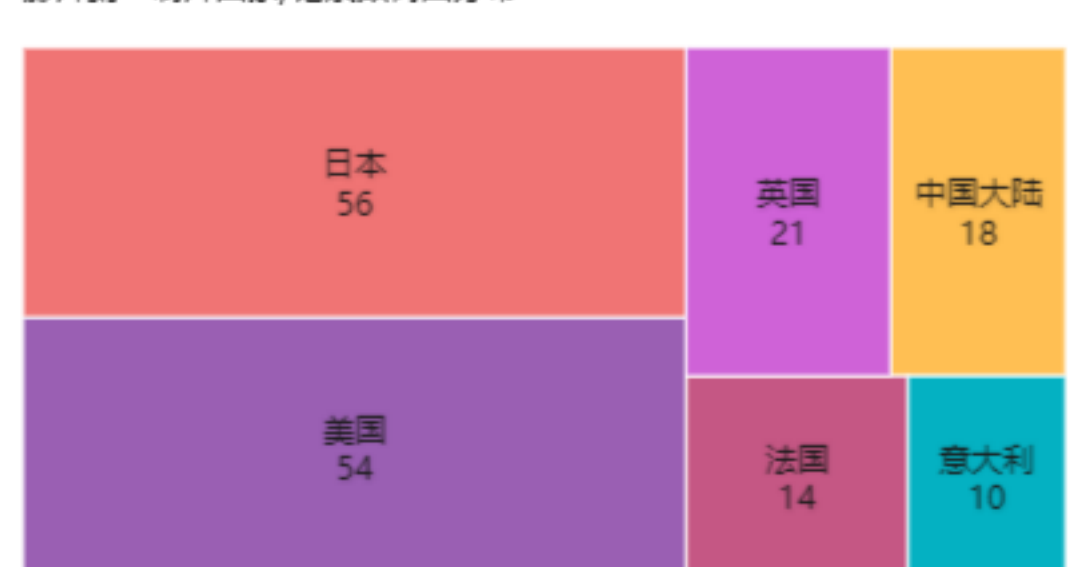
影片类型数量气泡分布图



影片第一语言/记录数矩形图分布



影片第一制片国家/记录数树图分布



明细数据表格

Table with columns: 电影名称, 5星人数分组, 短评数量分组, 第一类型, 第一语言, 第一制片国家, 评分分组, 评价人数. Rows include The Shawshank Redemption, Léon, Forrest Gump, Inception, etc.

共 541 条数据

1 / 6 < >

总结：

1、5星评价人数占比越高，越有可能是优质影片；

2、评价人数越多，越有可能为优质影片；

3、短评条数越多，越有可能成为优质影片；

说明：**5星人数占比、评价人数、短评条数**是影响评分高低的重要因素。

4、优秀影片产量较多的影片类型为**喜剧/剧情/音乐**；

5、优秀影片产量较多的第一语言为**英语/日语/汉语**；

6、优秀影片产量较多的第一国家为**日本/英国/美国**；

说明产量高，但不能证明生产优秀影片的概率高。