1. 看单一语料库对bcc整体的排名影响。目前做了微博和无微博排名，计算一下排名升降的绝对值，再用MT的leave-one-out方法，看看其他语料库对bcc的排名影响。
2. 微博（或其他分支语料库）对总排名的影响相差多少为合理？

*#两组数据的相关性（某一个值对应的两个数据，相差多大是正常？参考testing，两个人对同一个人的打分，相差百分之几不能用？*）

*#如果某一句话在广告中大量出现，要剔除吗？这算语言中的异常现象吗？*

1. 初部排了两人打分和语料库频率的图，可能有三个原因导致图不好看：长恨歌本身不够常用；长恨歌只有一百二十句，语料太少；只有两个人评分，评得分太少。