* 关于豆瓣top250 电影的数据分析
  + 这个实验的主要目的是对于豆瓣 top250电影的数据进行整理和发掘 让我们对于一部 能够进入豆瓣top250 的“好电影”电影的特质有更清晰直观的认识
  + 在能够爬取的数据中 有片名…… 所以我选取了留个角度进行分析
  + 首先是类型分析
    - 由于几乎素有的影片都有剧情这一项 但是这种类型不能反映电影的真实特质 所以做了两版分析，考虑到剧情的，和排除剧情的
    - 统计结果……这里用到了条形图和饼图分析 ，原来matplotlib的配色比较丑，就发现原来python有好多特别好看的颜色，还蛮开心的。
  + 导演分析 ……
    - ……最多 还有……也是榜上有名的
  + 产地分析
    - 通过条形图和饼图可以看出……
  + 人气
    - 按照打分人数和短评数量再根据年份处理给出人气值 人气-排名的关系
    - 在我的人气值评分体系下 评分1-100 显然是没有什么关系的
    - 但是豆瓣top250 应当还是将评分人数考虑在内了 这里可以看到 有一定的相关关系 线性回归的相关系数相关系数为0.448
    - 折线图和散点图
  + 年代分析
    - ……
    - 这里用了直方图更直观地现实频数分布
    - ……
  + 短评
    - ……
    - 大家能看出来吗 这是一只皮卡丘
    - 从词云中可以看出……

总而言之 豆瓣top250应当反映了大部分中国文艺青年的电影洗好 爱情和喜剧在top250 中所占最多， 宫崎骏 诺兰 王家卫 斯皮尔伯格 这些导演 当直观地看到这些作品的时候 才能真实地感受到 这些导演被大家推崇尊敬的愿意 在产地方面 美国最多 中国显然乏力

而人气和 在豆瓣top250的排名与评分都没有直接的关系