

Hello 大家好，我是一个小舞见，也是计算机专业的学生。

2009--2019，宅舞区经历了十年的沉淀，正站在崭新的起点上努力触碰更高更远的天空。

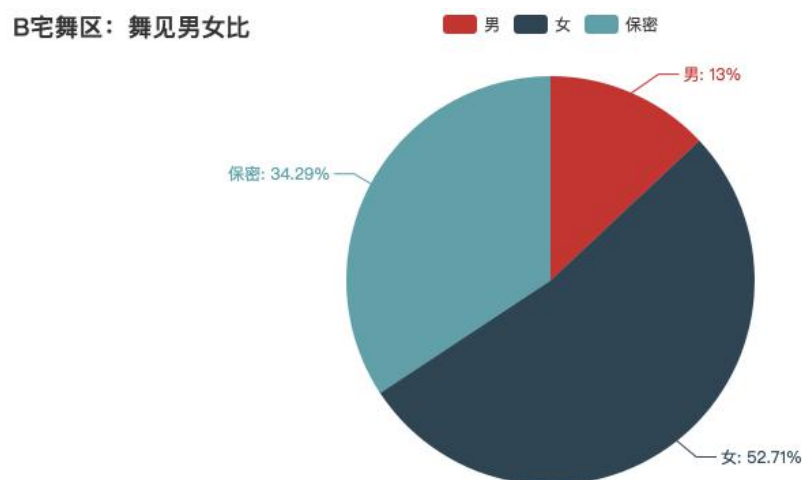
值此情 wang 人 wang 节 wang 的大好时节，我将用可视化方法，从多角度直观展示宅舞区十年发展的各项数据。相信以下总有一个是屏幕前的你有兴趣了解的~[多图预警]

（由于条件有限，样本爬取不到 2w 条，所以结果似乎娱乐成分更大一点 hhh 对代码项目感兴趣的童鞋，评论区我会说明~）



## Question1: 舞见男女比?

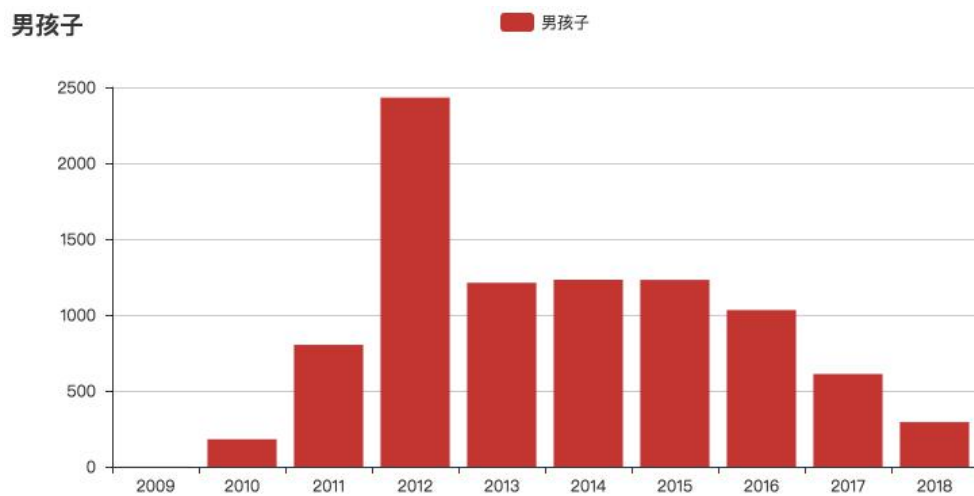
在宅舞区应该叫做女女....女男比:) 如下图:



宅舞区舞见男女比例

手动看了一下，其实性别为“保密”的一项有相当多数量的小姐姐（比如咬人猫）。所以算起来比例应该还不止这些。

突然想看看"稀有物种"男性出现在弹幕里数量的十年变化曲线了，如图：

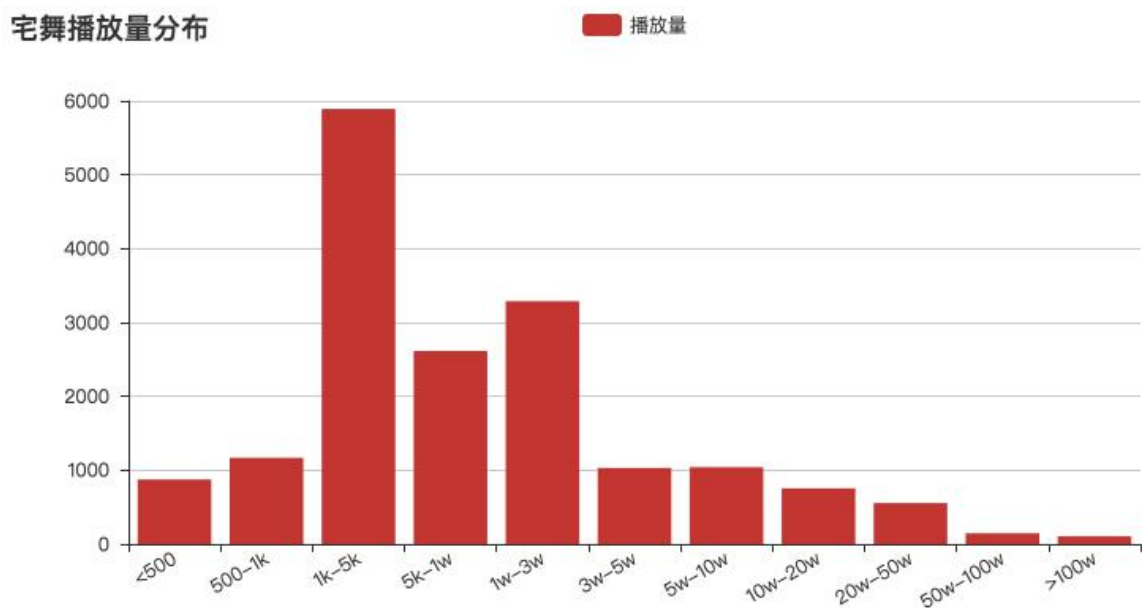


2009--2018 弹幕关键词"男"变化曲线

我做统计的时候，把"汉子","小哥哥"以及包含"男"的词语都算进去了。遗憾的是，尽管我尽量去抓关键词，但是弹幕里包含男性的内容自峰值年以来越来越少，这是否能反应男舞见数量正在下滑呢？



Question2: 播放量分布图



2009--2018 宅舞播放量分布图

两个播放量峰值就是舞见需要努力爬的“坡”，即 1k 和 1w+。(不遵循正态分布果然是因为样本少嘛...哪天拟合一下会不会是哪个指数分布族？)

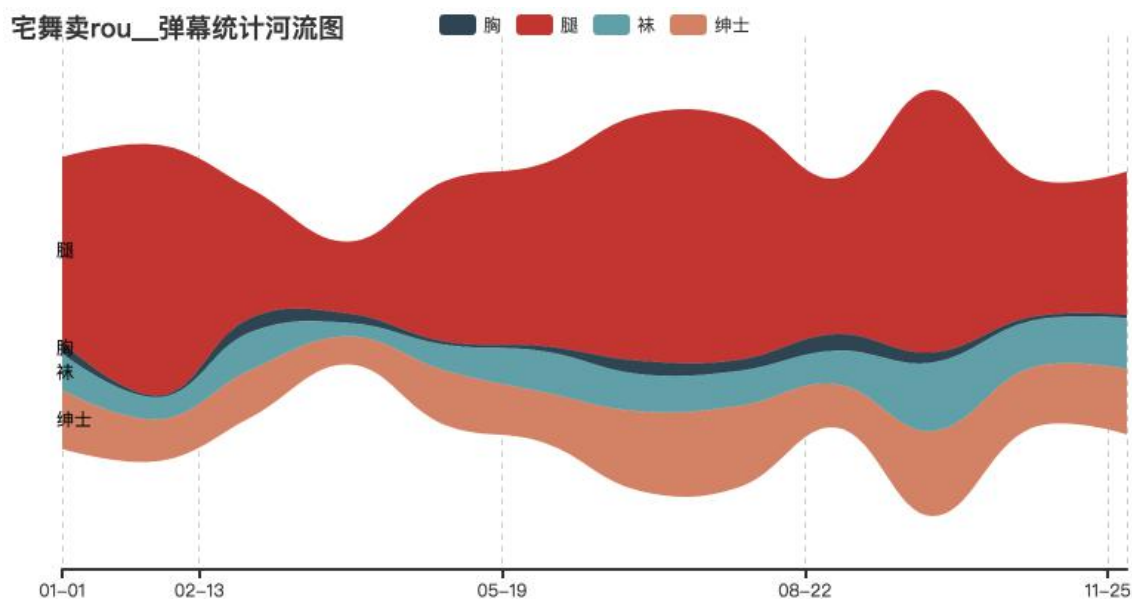
顺便一说我是最左边那个...OTZ



### Question3: 关于舞区风气的一点分析

不论是一直没有停息的“舞区卖腿”的讨论，还是前一阵子“讨论舞区卖腿”的潮流来袭，都说明大家对宅舞区的要求越来越严格，这是一件好事。那么数据会告诉我们什么呢？

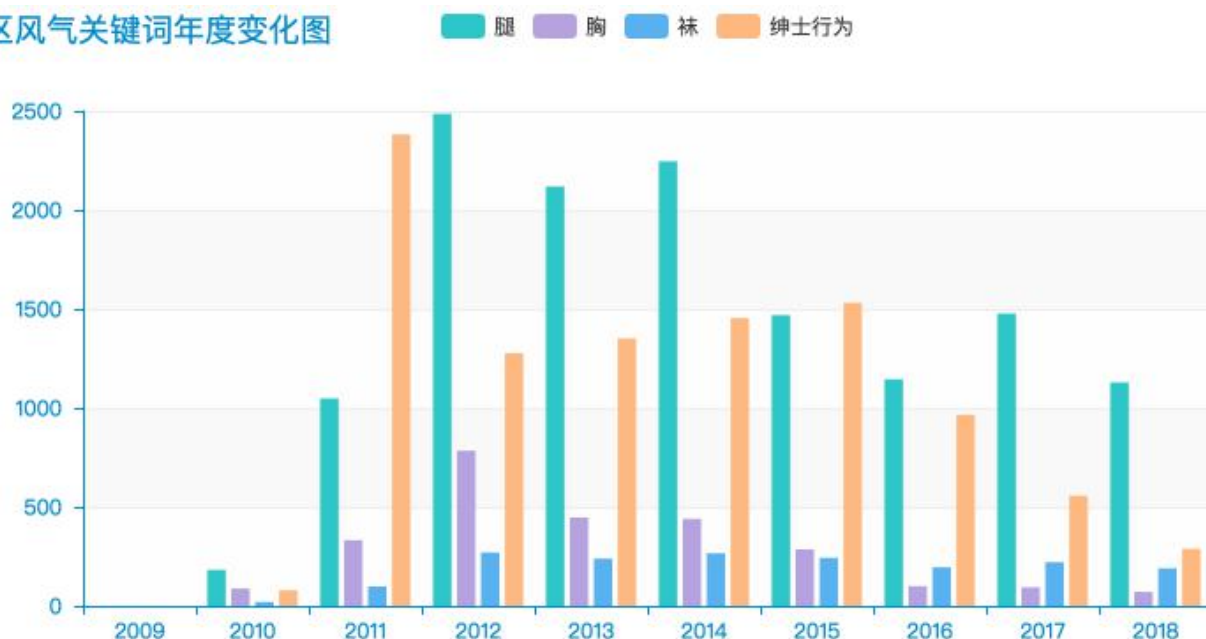
下图是从弹幕总结出来的去年“卖 rou”随月份变化图，四种颜色表示四种代表卖 rou 行为的特征，其中“绅士”包含一些不雅用语：



高峰集中在寒暑假和国庆假期左右（因为有空投稿了嘛），而除了特征“腿”之外，其他三个特征的变化趋势是一致的，与腿刚好相反。也许这说明了最大的亮点集中在”腿”上时，观众对其他地方的想法就削弱一些；而当“胸”和腿上的装饰物本身比较突出时，就会有很多很多的想法(w)

下图是 2009--2018 以上四种特征出现频次的变化图（求不要问为什么不用折线 QAQ）

舞区风气关键词年度变化图



2009-2018 舞区风气关键词年度变化图

果然观众的注意力一直集中在”腿“和”绅士行为“两方面，而”绅士行为“数量的减少证明了舞区弹幕环境在逐年改善。从图中也能发现，其实去年可能并不是卖腿的高发年份，而在远古时期大家可能更加愿意在弹幕里刷腿子（当时宅舞还是非常新鲜的玩意儿）。不过 12 年左右的视频还是靠搬运 N 站为主吧？

私下里采访了一些童鞋，搜集了一些可能存在“卖腿”的视频弹幕并生成词云图供大家参考（可以与接下来 Question5 的年度弹幕云图进行对比，发现这里弹幕的不同点 www）



## Question4: 宅舞区第一投稿, 第一个国人稿件, 以及十年投稿数量变化

宅舞区第一稿:

av152, 爬出来后看了一下投稿发现被尊称为 152 神教...不过是 N 站的搬运

aid	cid	mid	video_name	up_name	view	danmaku	reply	favorite	coin	share	like	history_rank	pubtime
152	3642989	1878	萌娘新作 - 不知道新不新	猥琐过度	13186	432	139	154	22	364	41	0	2009-07-15 20:34:50

1 row in set (0.00 sec)

```
mysql> SELECT * FROM up WHERE up_name="猥琐过度";
```

mid	up_name	sex	avatar	sign
1878	猥琐过度	男	http://il.hdslb.com/bfs/face/16e9a87305165eb1f6ae4da811777986a5fed26e.jpg	

av152 信息图



## 第一个国人稿件：

在我的数据库里第一个投稿的国人宅舞稿件是 av38103, 不过这个不是宅舞，不知道为什么放在了宅舞分区。

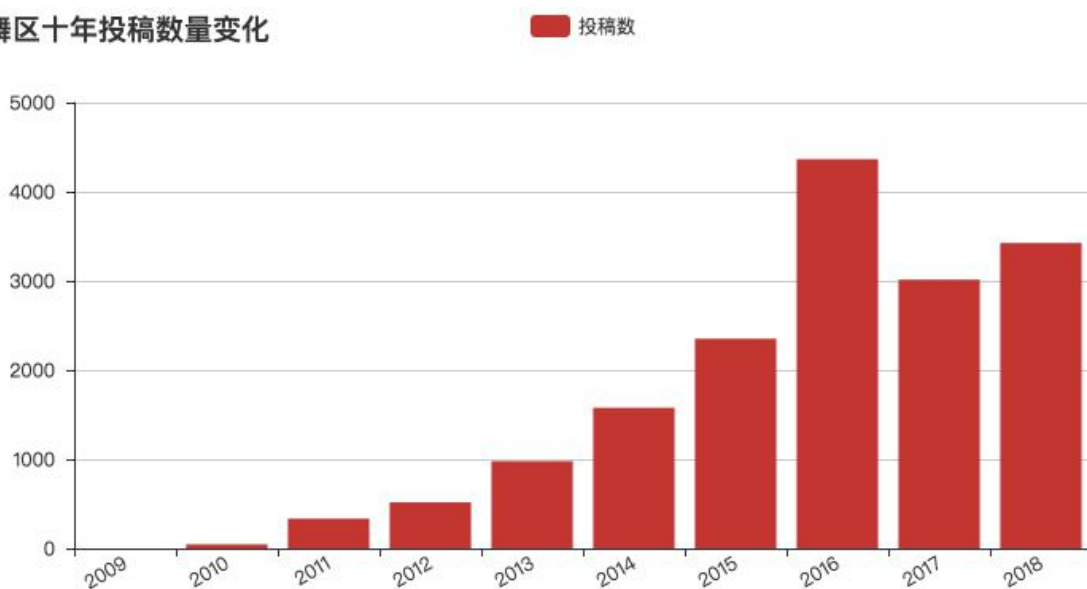
所以真正的第一稿应该是 av39292 的华师大七草祭投稿视频，跳的是“恋爱循环”。2010 年就在大学跳了宅舞，真的好早啊！



由于数据库不全所以不继续查找第一个单人舞见投稿了~以后继续完善啦~

## 十年投稿数量变化：

宅舞区十年投稿数量变化



2009-2018t 投稿数量变化

我觉得应该是样本太少的缘故，不然我都纳闷为什么会显示 17 年有投稿量的滑坡？怎么想都不明白啊？17 年是 b 站“百人宅舞”BDF 企划元年（不是第一次 BDF），会带动很多稿件。假如真的有滑坡的话，那原因到底是什么呢？homework~



## Question5: 十年弹幕氛围变化

通过 2009--2018 [年度弹幕词云分析](#)，最直观的结论是这样的：

1. 一直是热点的用词：好可爱 啊啊啊 爱爱爱 承包 妹子 小姐姐 腿 喜欢 加油 表白 ...
2. 2017 年及往后，“小姐姐”的数量超越了“妹子”，成为最常见的舞见代词
3. 早期弹幕不乏粗俗用语和乱七八糟的词，但是往后弹幕质量逐渐提高，2015 年往后逐渐形成了以“虚”的称赞居多的弹幕氛围。舞区的整体气氛应该是越来越好的。  
  
注释：“虚”的意思就是泛泛的夸奖，而不是指出动作，着装，表情什么的很细致的分析（概率意义上）
4. ‘生日快乐’近两年成为热词(暂)，预计‘awsl’或者‘啊我死了’能成为今年新进热词
5. 好像大家一直很喜欢马尾？（狗头）



2009-2017 的词云图依次展示如下：（封面图是 2018 的）



## 2009 弹幕词云



## 2010 弹幕词云



2011 弹幕词云



2012 弹幕词云









## 2016 弹幕词云



## 2017 弹幕词云





## 2018 弹幕词云

(为什么不拿评论分析呢? 太菜了爬不到评论数据 QAQ 有路过的大腿请提点一下)

相信大家可以从词云里得到更多有趣的信息~~~



报告到此结束啦~就像我在开头说的那样，当成一次娱乐向的阅读就好~或许可以成为舞见小姐姐练舞休息时不错的放松材料呢~

年也过啦，春天马上就要到咯，宅舞区的大家一起撸起袖子加油肝起来！夜空划过璀璨的流星雨，星空下是心意交织的我和你；要和你一直，一直在一起，向更高更远的天空，出发！

bilibili 宅舞区, []~(¯▽¯)~\*干杯