

世界大学大乱斗

论文草稿，英文版待更新，先靠母语打份草稿(黄瓚 北京理工大学 dreaming_hz@hotmail.com) (第七组 黄瓚 乔伟)

最近卡在数学模型的建立与进一步的数值化分析处理工作上

摘要

世界大学排名受到越来越多的关注，是求学与求职时的重要参考。现已有若干有影响力的榜单逐年更新，每个榜单都有自己的一套评价方法。大学排名仍然是一个模糊而充满争议的话题。本文中，我们参照几个公开的大学排名数据集，试图一窥创建一流大学的秘诀。

内容介绍

本文中，我们对世界大学排名数据集进行探索并试图通过数据分析找寻使一所大学誉满全球的关键因素。

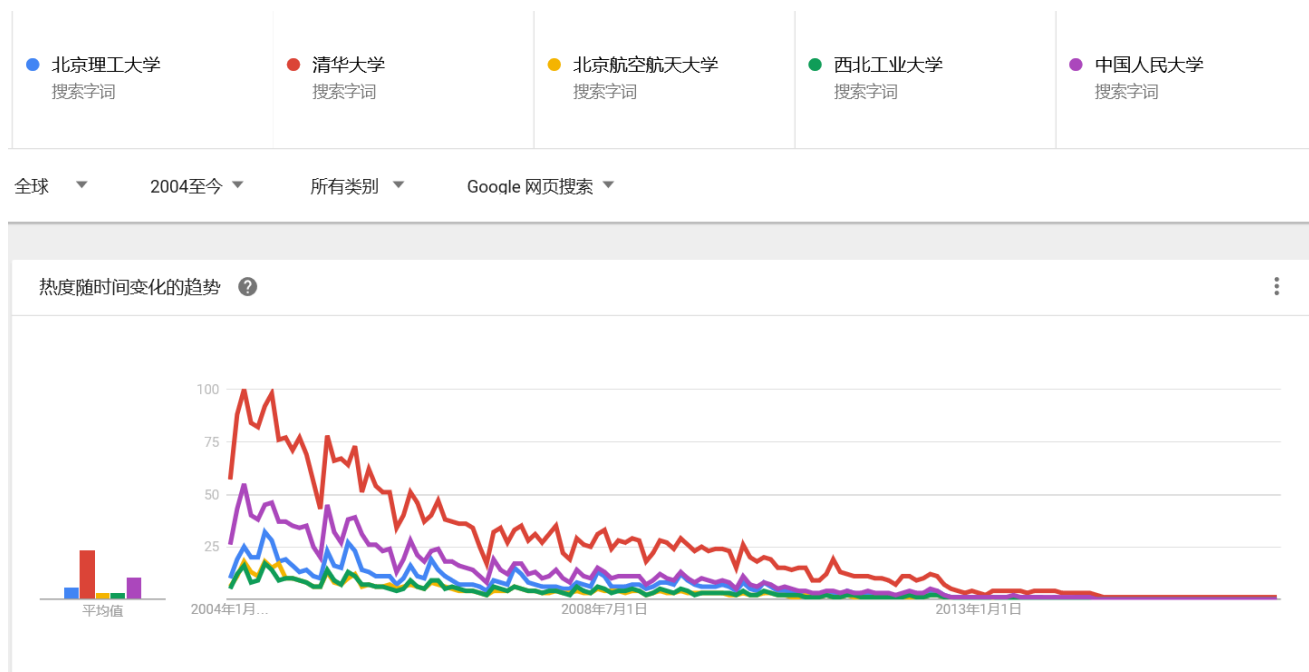
全球各地分布着成千上万所大学，很难断言究竟孰优孰劣。但另一方面大学排名对我们的生活其实有很深刻的影响，尤其在东亚地区。不论我们对其好恶程度如何，大学排名信息都应被更深入地分析探讨。

我们依照数据挖掘的通用流程展开工作，提出问题是首要的一步，围绕着对应问题进行数据的收集与清洗，数据分析，数据可视化，根据结果进行总结报告。

换个视角重新审视大学排名问题，用数据搭建起通向智慧的桥梁，在生活的混沌中寻求真知。

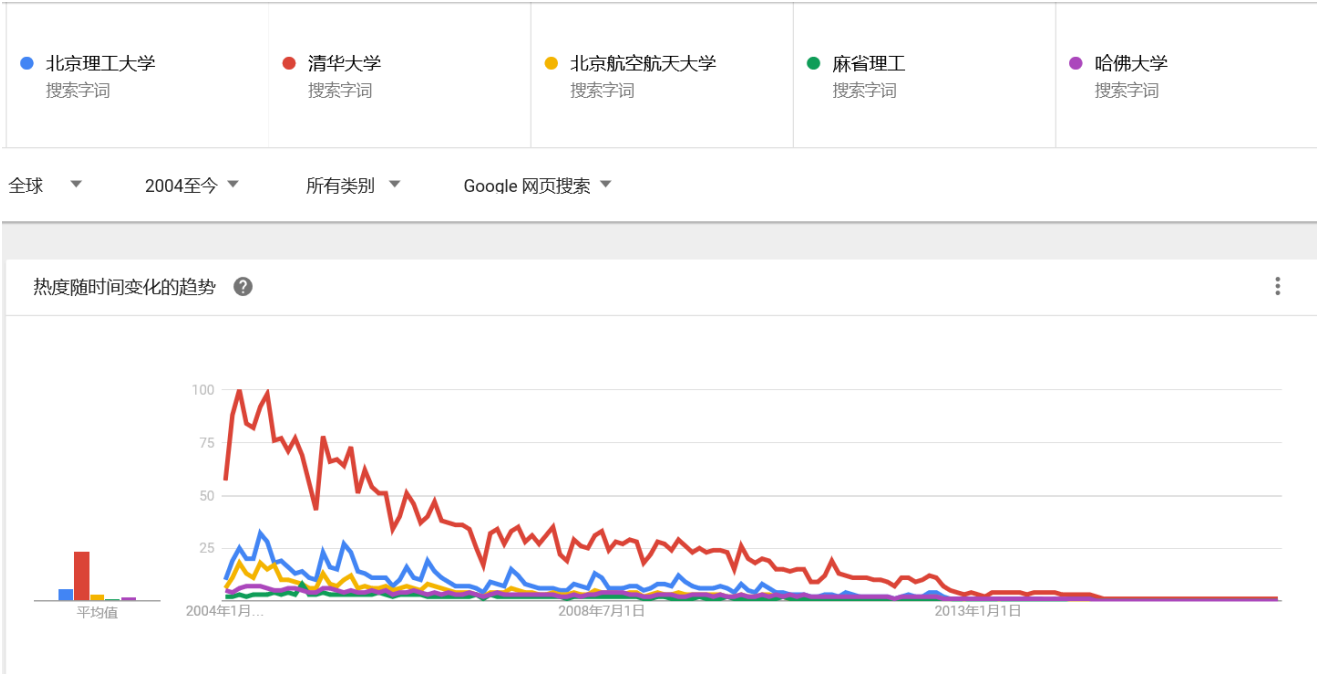
预调查

在百度百科上有一个著名的撕逼词条叫作中国十大名校，由于各方的编辑大战最后给关停了，在大陆，大学排名是个很敏感的问题，我们顺手使用 Google Trends 通过关键词作了对比分析，结果如下



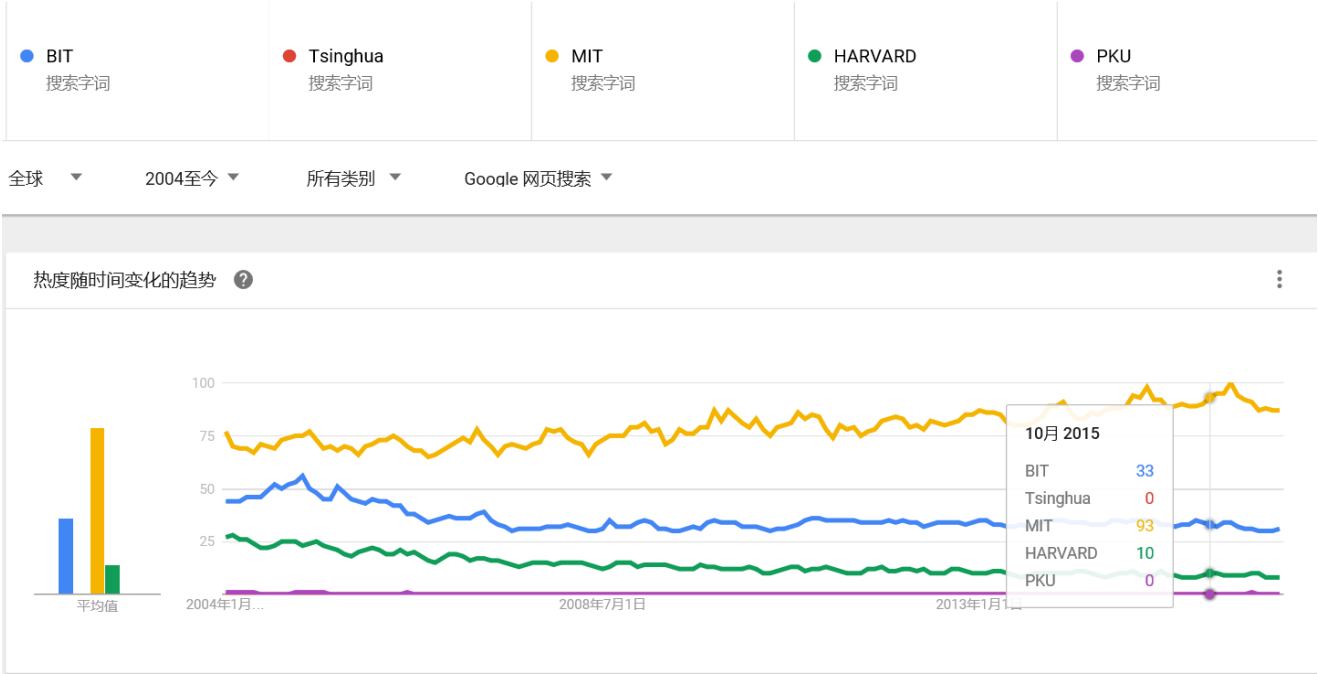
名校往往会受到更多关注，拿北理、清华、北航、西工大、人大五所学校作为关键词对比其 Google 流行趋势，以某关键词最高搜索量为100分，按比例计算得到各校校名在时间轴上热度对比可以看出，五道口热度最高，舒克贝塔相爱相杀中间夹杂着西工大，人大搜索热度要高于三所工科院校；由于谷歌撤离中国大陆，在13年以后的数据不再具有参考价值，同时我们也可以看到中文谷歌总体用户数的下降。

换掉两所学校，我们再来看一下对比结果



趋势喜人，麻省和哈佛一同垫底人生赢家感爆棚

我们换个玩法，使用学校缩略词回归英文世界看看哪方在国际舞台更受关注



理工这是脚踏哈佛直追麻省的节奏，还能甩另两位好几条街，看来名字起的好确实很重要Orz

有兴趣的朋友可以转战百度做类似工作，从曲线中能够明显发现每年六月末七月初填报志愿期间是校名搜索的高峰期



至此，我们的关键词游戏告一段落，从中可以看到获取可靠数据进行分析工作的难度。鉴于两人小组时间有限，我们主要在 Kaggle 的 World University Ranking 数据集基础上继续工作，在此向该社区致以诚挚谢意。

基本假设

一切数据分析工作都要以针对特定目的所提问题为中心，我们在此提出的问题是：哪些因素让一所学校比其他学校更好？

这是一个相对模糊的问题，首先对于好的定义，是录取难度、经济实力还是最简单的对于特定的人给出所有学校作为选项，他所选择的那所学校即为最好的一个？

我们这里的“好”，是一个待挖掘的概念，我们会首先关注于更加底层的方便数值化的概念，比如：

- 资金-我们认为钱越多越好
- 人员-人多入少说不定，但是师生比越低越好
- 科研成果-发表论文数及影响力
- 准入门槛-成绩要求及申请难度

其它内容也纳入考虑范畴之中：

- 人员流动-人们倾向到自己心目中更好的学校就读，到资源更优越的学校工作，加之学校人员聘用时默认规则的影响；人员流动包含着我们所需信息，但数据的获取分析难度更大
- 语言-国民以英语为母语的国家在若干榜单上排位优势明显，部分学校被低估

待解问题

- 在中国 90 年代高校合并潮中，大而全则好的理念被一再贯彻，体量会带来的论文发表量等优势，那么大校一定利于提升排位么？
- 很多高校是依靠最突出的一小部分人通过展示学术成果等方式提升学校声望，但对于校内的大多数来讲却并没有太大关系；学校是否能帮助学生真正实现自我发展而非招好学生筛选出精英小队进行军备竞赛，这一因素是否列出到大学排名考察当中？
- 排位高一定好？

数据集

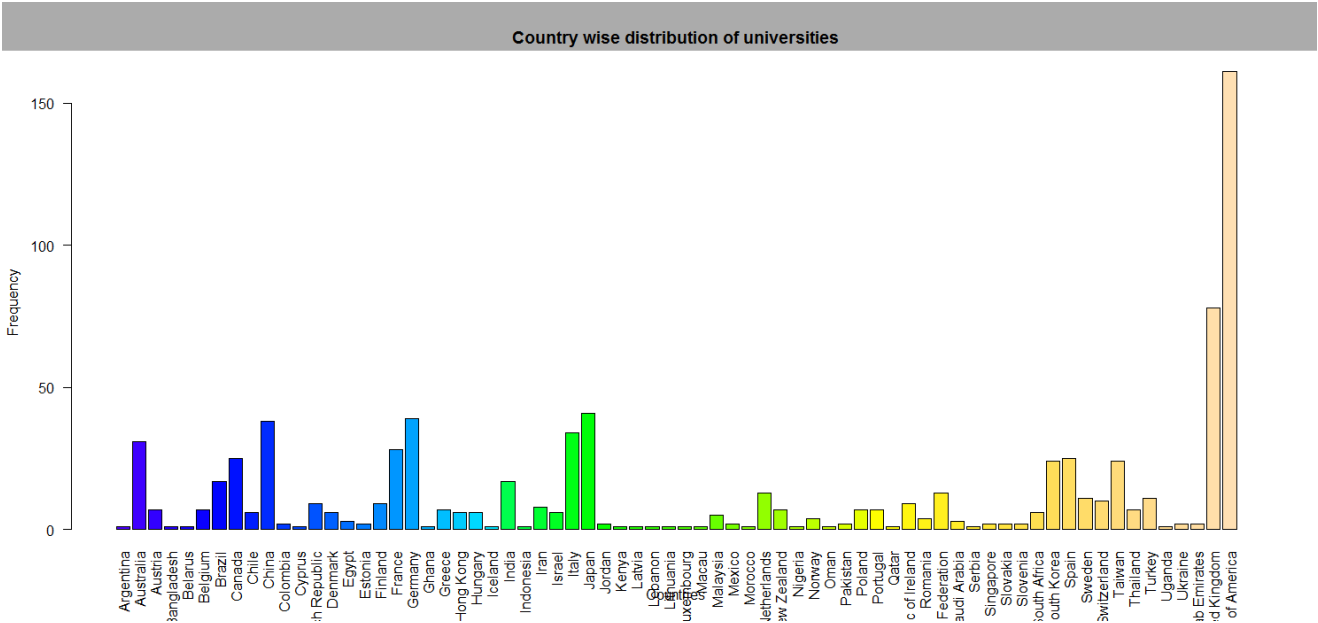
我们所用数据集构成以及具体的数据获取和清洗方法将在该部分进行说明。

Kaggle World University Ranking Dataset

部分数据集从 Kaggle 下载获得，其中包含 2005 年到 2015 年的上交世界大学学术排名，Times 排名，CWUR 排名等多项内容，具体可以登录 Kaggle 进行查看。

地理分布

选取十年内各榜单 500 名内出现过的学校按国家进行划分，作出对应柱状图：



可以看到，这些学校多集中于北美、西欧各国，尤其是美国拥有超过百所上榜学校，在该排行榜上遥遥领先。

我们对数据作进一步处理，从按国家名字排序的方式转到依大学数进行排序，取 50% 为界看有多少国家的大学就够占到此时考察对象的一半。

个人角度

大学选择因人而异，合适才是最好的，贵不一定好吃。