

## Print（“hello world!”）有多难？

### ——中国各省份少儿编程发展状况及相关因素探析

编程是人工智能发展的基础，编程能力也是未来互联网经济时代的核心竞争力。早在1984年邓小平就提出计算机要从娃娃抓起。中国于2017年《国务院关于印发新一代人工智能发展规划的通知》（国发〔2017〕35号）提出：“实施全民智能教育项目，在中小学阶段设置人工智能相关课程，逐步推广编程教育，鼓励社会力量参与寓教于乐的编程教学软件、游戏的开发和推广。”

尽管少儿编程教育已经成为国家教育规划的一部分，中国各省市地区编程教育发展状况大相径庭。各省市地区的经济发展、人口规模、教育市场的竞争状况，以及互联网发展均有参差，这些差异影响了编程教育的发展。

因此，本文将探讨中国少儿编程教育在各省市地区的差异，以及探究相关的经济、社会、科技水平等因素对少儿编程教育发展状况的影响程度。

### 数据与方法

我们运用 OSEM pipeline 来进行数据处理与分析，即

- O —— 获取（Obtaining）数据
- S —— 清洗（Scrubbing）数据
- E —— 探索（Exploring）数据，通过数据可视化等方式，找到潜藏的模式和趋势
- M —— 建模（Modeling），对数据进行建模，会给我们相关预测
- N —— 解释（Interpreting）数据

#### 一、获取数据

我们将关注的变量、指标、以及数据来源列在表一

表一： 变量、测量指标、与数据来源

变量	指标	数据来源
编程教育的发展	少儿编程教育公司注册数量	企查查
	线下培训机构数量	大众点评搜索
编程教育的受关注程度	媒体关注度	百度搜索指数
经济因素	2019 年各省人均 GDP	国民经济和社会发展报告
社会因素	中国中小学人数	国家统计局
科技因素	中国互联网普及率	中国互联网发展报告

\*所有数据来源均不包括香港，澳门和中国台湾。

\*文章中的少儿编程针对用户年龄群体为 4-16 岁。

我们将各省市的少儿编程企业的数量作为衡量该地区少儿编程教育发展状况的标准，以大众点评中的线下少儿编程机构数量作为补充数据，以此衡量中国编程教育发展的客观状况。同时关注经济、社会以及科技等影响因素，用以上数据作为探讨中国少儿编程教育的地区差异数据来源。

由于一些政府网站并未提供一些公开数据，导致直接爬取数据较为困难。我们分别从国家统计局、2018 年国民经济和社会发展报告以及近三年的《互联网发展报告》中获得相关的二手数据。

在百度搜索指数中，我们统一的用“省份+少儿编程”来进行搜索。同时在大众点评上，我们尝试着用少儿编程进行搜索，但涵盖的搜索错误相对较大，故最终只采取用“编程”进行关键词搜索。

由于 2019 年互联网普及率尚未公布，我们使用了 2018 年的数据。近几年各地区互联网普及率趋势大致相同

## 二、清洗数据

1. 少儿编程教育企业数量：筛选出被取消或关闭的企业，计算出各省市区的公司数量。
2. GDP 数据：分别按照总 GDP 和人均 GDP 对各省市区别进行排名。
3. 人口：选择小学，初中和高中的学生数据，然后计算各省市区的学生总数。

## 三、数据探索、建模与解释

### 问题一：中国少儿编程教育在各省市地区的发展现状

我们采用两个指标来考察少儿编程教育的发展：少儿编程教育企业数量，大众点评上搜索数量结果，结果如图 1. 。

#### 1.1 中国大部分少儿编程机构和企业分布于少数几个省份

图 1. 显示，广东、浙江、江苏的少儿编程企业数量是全国的前三名，前五名的省份少儿编程企业注册数量占到了总量的一半以上。广州拥有 1336 家线下培训机构；虽然在北京注册的企业数量不超过 10 家，但线下培训机构的总数却高达 1215 家。少数几个省份涵盖了中国大部分少儿编程企业和机构。

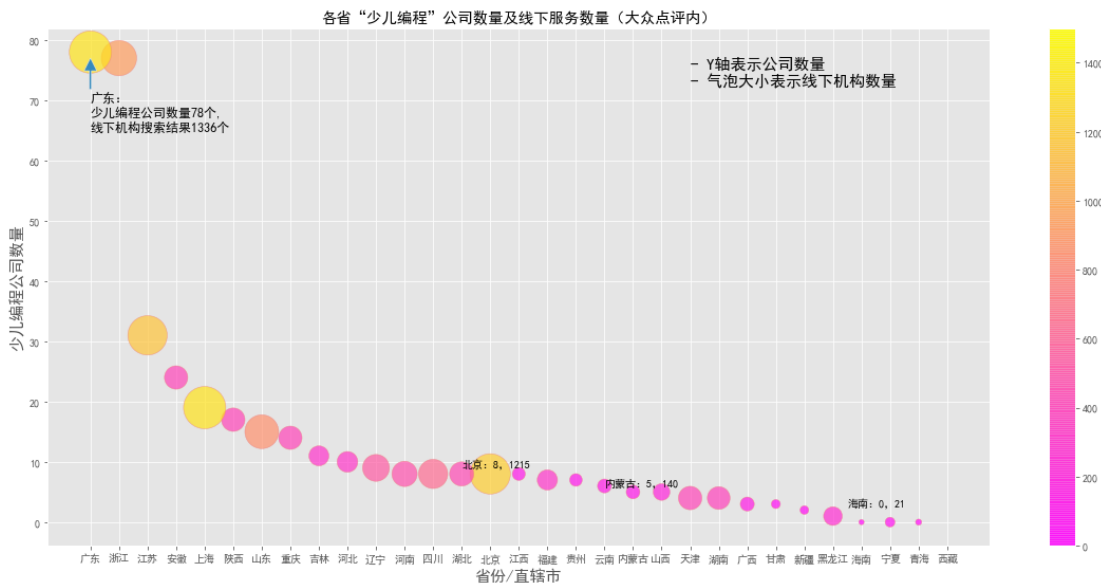


图 1.

（各省份少儿编程企业和机构的发展状况，包括企业数量和线下培训机构数量）

问题二： 影响编程教育发展的因素探析

2.1 编程教育与经济发展的关系

图 2 显示，上海、北京两大直辖市人均 GDP 抢眼，但少儿编程企业注册数量远少于头部分省份，其余省市的少儿编程企业数量和人均 GDP 的排名则具有较强的一致性。与少儿编程企业数量相比，图 3 显示了线下培训机构和各省人均 GDP 有着更紧一点的相关性。值得注意的是，福建省存在 GDP 排名高但是企业数量和线下机构均较少的现象。

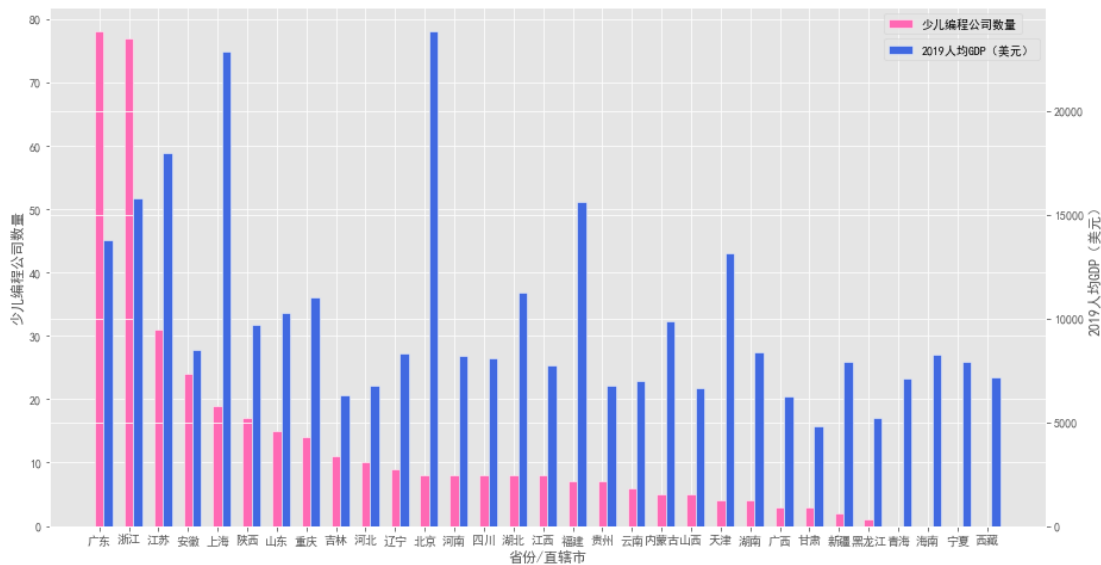


图 2.

(少儿编程企业数量与各省人均 GDP 的对比)

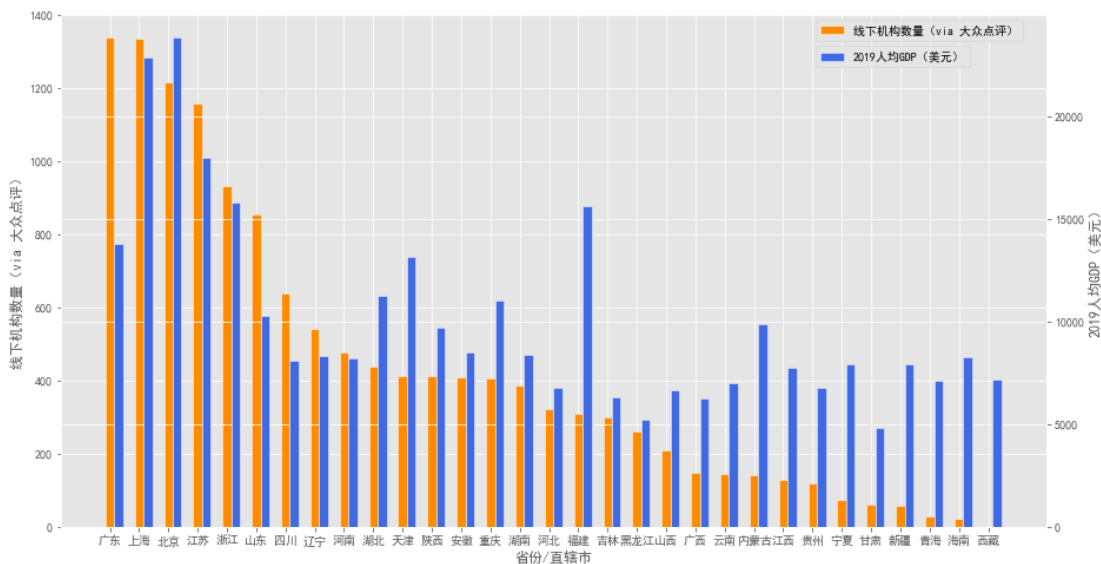


图 3.

(少儿编程线下机构数量和各省人均 GDP 的对比)

## 2.2 编程教育与社会因素的关系

我们通过各省市地区的青少年人口来反应社会因素即所对应的受众面。图4与图5显示，通过少儿编程企业数量、线下培训机构数量和中小学人口数据的对比可以看出，**各省市地区的少儿编程教育分布与青少年人口数量存在较弱的相关性**。上海和北京中小学人口数量虽少，但线下少儿编程机构数量居全国前列。与之相反的是，河南省拥有最多的青少年人口数量，但其线下教育机构却不足上海的一半。**少儿编程线下培训机构能服务到的青少年受众在全国分布并不均匀。**

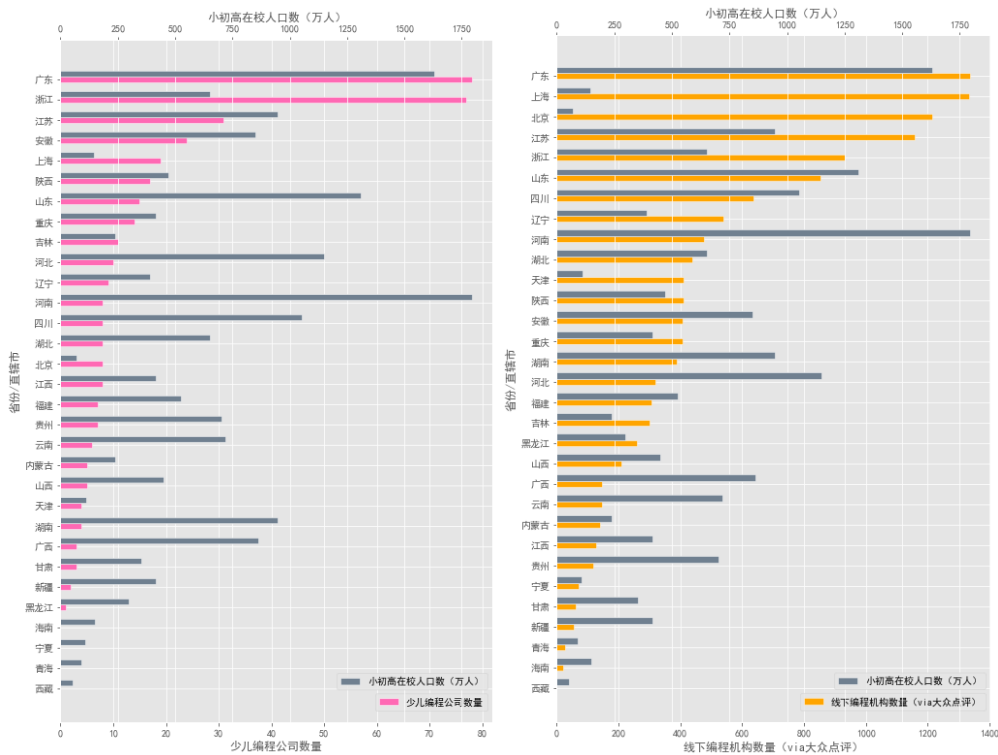
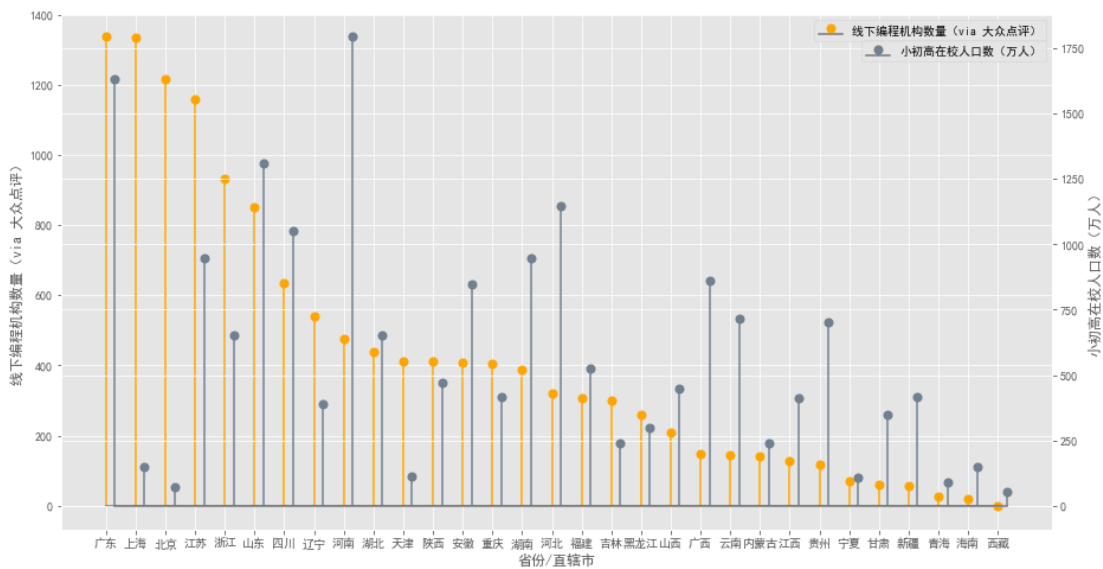
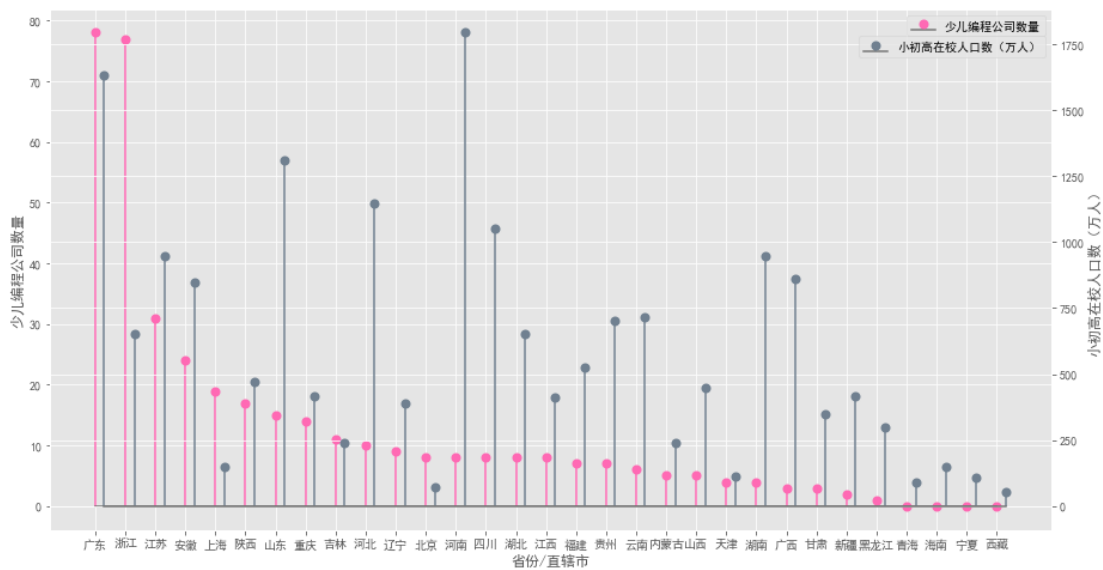


图4 & 图5

(少儿编程企业和线下培训机构与中小学人口数量对比)



#备用图，请老师参考

## 2.3 少儿编程发展与互联网普及率

近些年来，随着我国互联网产业的飞速发展，互联网触及越来越多的人民，各省份的互联网普及率大部分已超过 50%。在图 6. 和图 7. 中对比我们可以发现，网络普及率与各省少儿编程发展关联性较弱。

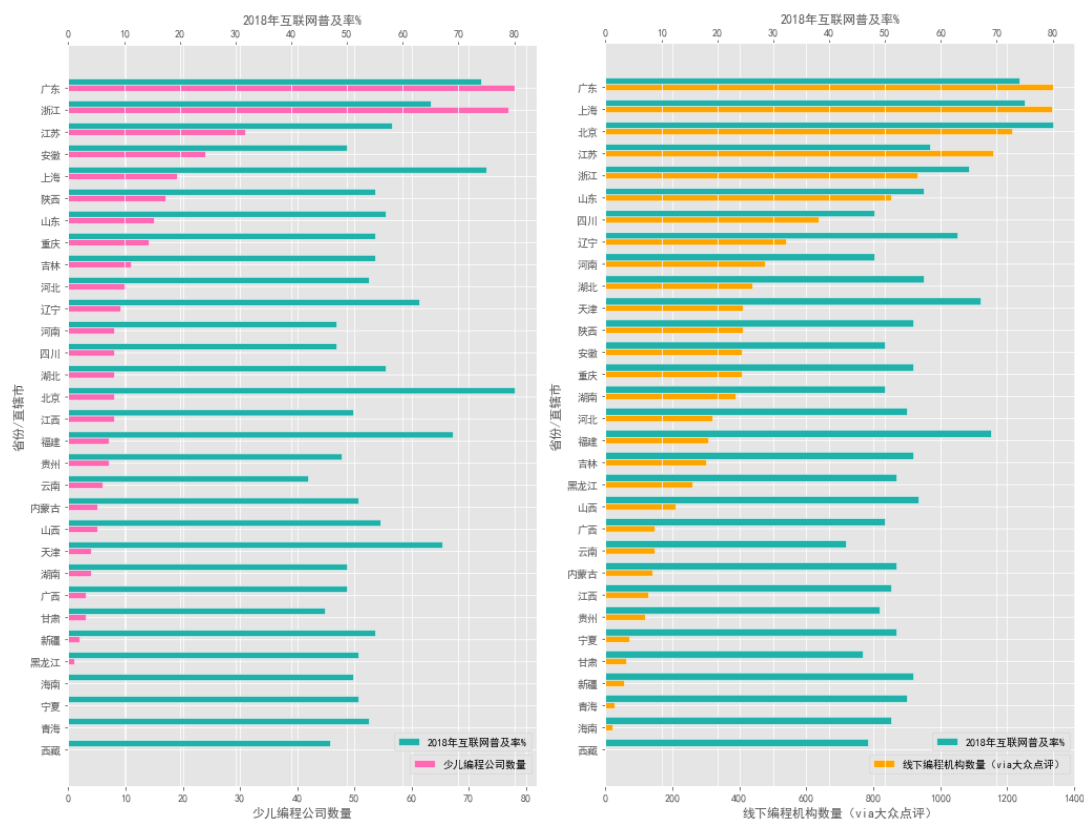


图 6 & 图 7

(少儿编程企业数量及线下培训机构数量与互联网普及率的对比)

## 2.4 少儿编程发展与媒体关注程度

我们使用百度搜索指数来衡量互联网用户对于少儿编程教育的关注度。相对少儿编程企业数量分布来说，各省的百度搜索指数更为均衡。其中企业数量遥遥领先的广东和浙江，互联网搜索指数也高居前列。北京作为首都，互联网用户对于少儿编程的关注度尤其高，但注册在当地的少儿编程企业数量却很少。

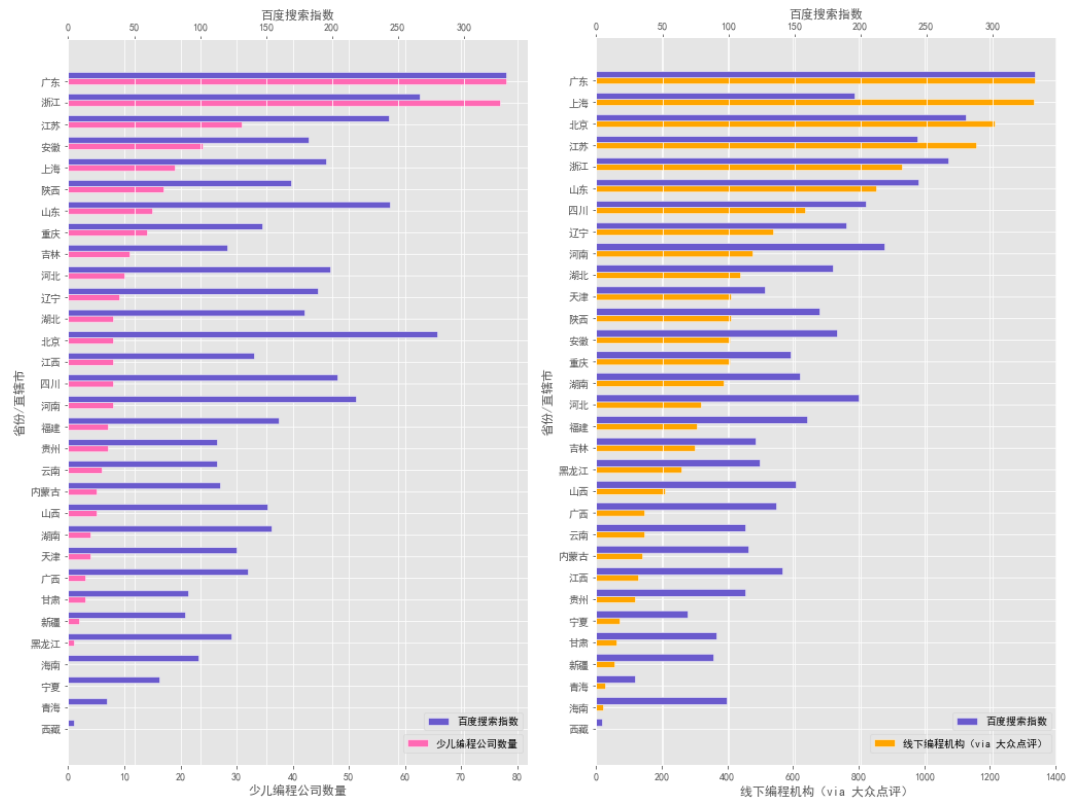
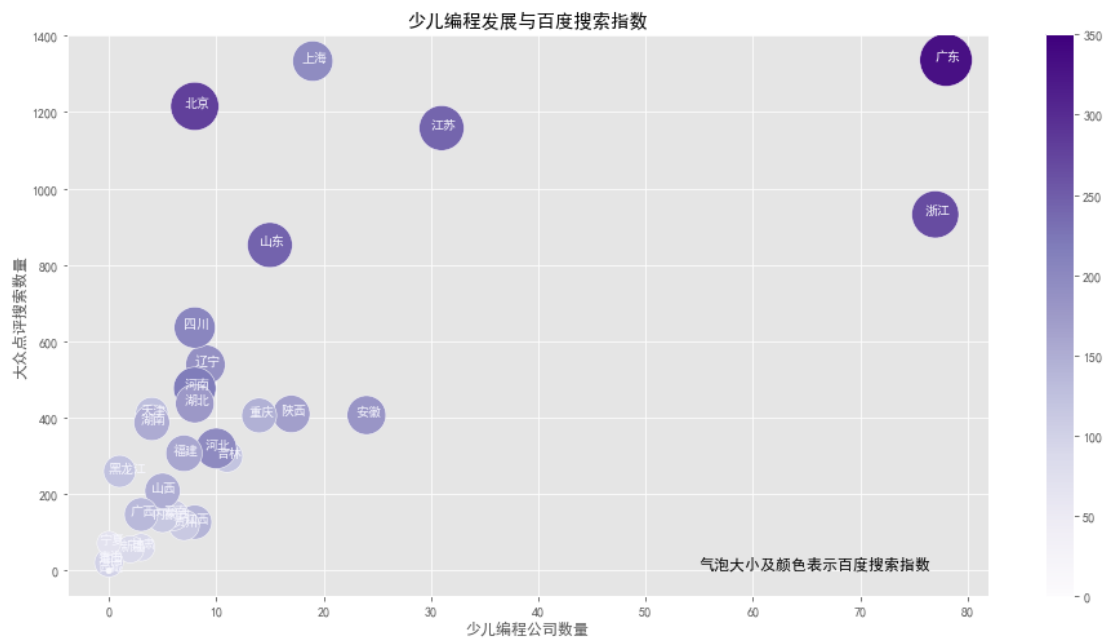


图 8.& 图 9.

(少儿编程企业数量及线下培训机构数量与媒体关注度的对比)



# (少儿编程企业数量及线下培训机构数量与媒体关注度气泡图)

#备用图，请老师们参考

总结

我们将所有可能影响的因素归纳到了图 10. 中。在各因素柱状图的比对分析中，我们发现有许多省份互联网用户对少儿编程关注度较高，并且青少年人口基数大，但少儿编程企业数量和线下培训机构较少，例如山东、河南与四川。这样的现状在未来也许会催生一定的市场契机。





## 使用工具

Python

## 数据处理代码及参考文献

[https://github.com/superhq221/hkbu\\_comm7330\\_2019/blob/7330group/7330group/Reference.ipynb](https://github.com/superhq221/hkbu_comm7330_2019/blob/7330group/7330group/Reference.ipynb)

## 作者简介



乔 玥

自我介绍：2019 AIDMer，性格开朗，置身于数据新闻的不断学习和探索之中。

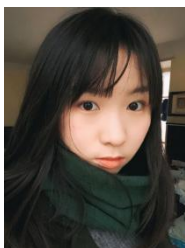
任务分配：数据搜集、文案撰写



施 意

自我介绍：可可爱爱，没有烦恼。

任务分配：数据分析、内容呈现



何 青

自我介绍：满头大汗的 python 入门玩家

任务分配：数据搜索、清理以及可视化



赵冰洁

自我介绍：不断努力奋进 er，希望得到 Jupyter 加持的文转理极度困难症患者。

任务分配：数据搜集，文案撰写