

# 中国性别问题研究

小组成员：高宇菲，林以任，姚修齐，李睿恩

# 工作摘要

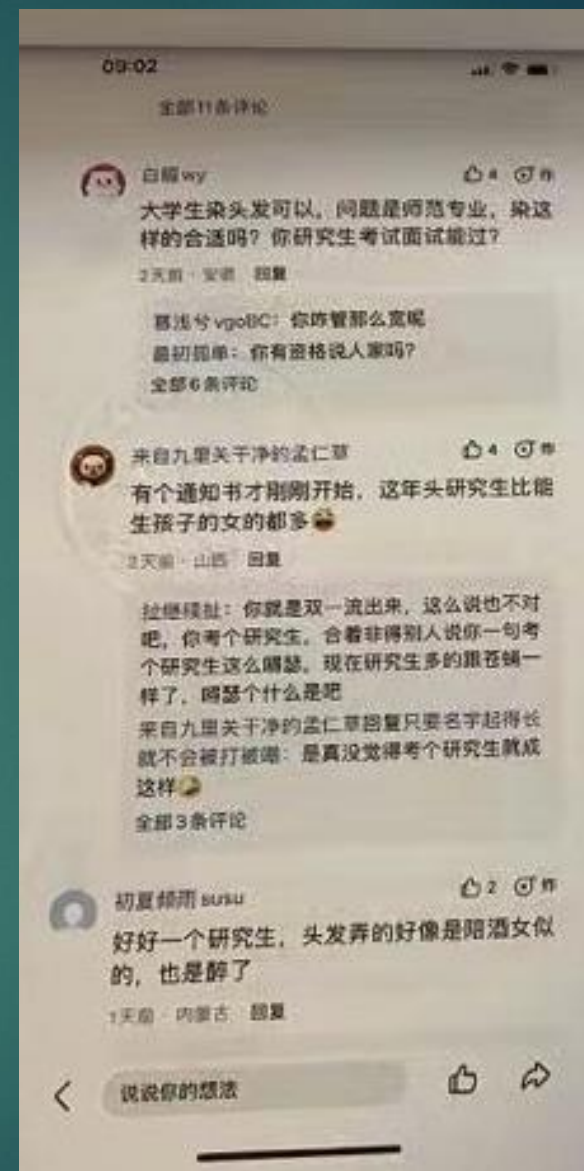
- ▶ 现状分析：伴随着互联网应用的兴起，我们得以听到他人的心声。然而，我们观察到在中文互联网上（微博、知乎、豆瓣、百度贴吧……），与性别有关的话题总是容易获得关注，基于性别展开的攻讦屡见不鲜。我们希望通过查找文献了解这一现象的具体情况，并希望通过自行爬取近期的互联网数据，基于自然语言处理的方法分析目前中国互联网的性别对立现象。
- ▶ 偏见检测：相比于非常明显的性别对立，一种更隐蔽的性别歧视实际上是语言中暗藏的基于性别的偏见。我们希望通过基于自然语言处理的分类模型来检测一句话中是否存在性别偏见。

# 现状分析

- ▶ 性别对立的话题在近几年频繁出现。



2022年河北邯郸“毒驾”事件微博评论区



2022年一位华东师范大学女研究生因染粉色头发被网暴, 最终抑郁逝世

# 现状分析

- ▶ 我们希望以更严谨的角度了解中国互联网的性别对立现象。
- ▶ 文献法
- ▶ 百度指数
- ▶ 基于自然语言处理的热点话题微博评论分析

# 现状分析——文献法

- ▶ 黑龙江大学的张新苗，以语言学的角度撰写了硕士学位论文《微博评论性别对立评价研究》。
- ▶ “微博中的性别对立容易造成刻板印象，将男性塑造成“好色”“暴力”“不负责任”的形象，将女性塑造成“贪财”“能力差”“不理性”的形象，将与性别相关的制度、机构塑造成“偏心男性”“无用”的形象，同时传播了对于异性和婚姻的厌恶和恐慌情绪。大多信息呈现引导性、批判性，夸大性的特点。”
- ▶ “微博的性别对立舆论战中女性处于优势地位。”
- ▶ “虽然性别对立程度更加激烈，热度逐渐增长，性别对立逐渐向着娱乐化的方向发展，但从评价对象的变化来看，增加了对不分男女的性别对立的批判，可以看出微博中的性别对立还是在向着好的趋势发展的。”

---

[1] 张新苗. 微博评论性别对立评价研究[D]. 黑龙江大学, 2022. DOI:10.27123/d.cnki.ghlju.2022.000634.



# 现状分析——百度指数

- ▶ 以百度海量网民行为数据为基础的数据分析平台，是当前互联网乃至整个数据时代最重要的统计分析平台之一，自发布之日便成为众多企业营销决策的重要依据。
- ▶ 通过检索特定关键词，可以得到该关键词在不同时间下在百度中被搜索的频次，并且可以得到一些异常峰值出现时的日期里的新闻头条事件。同时，我们还可以获得搜索该词语的用户的画像，如搜索者所在地区、年龄、性别。



百度指数 官方

## 体验大数据之美 AI引导数据革命

趋势研究，近十年关键词变化趋势免费查 >

需求图谱，呈现用户隐藏的关注焦点 >

人群画像，呈现关注用户的群体画像 >

# 现状分析——百度指数

- ▶ 检索词语“女拳”、“小仙女”、“下头男”、“普信男”的数据。
- ▶ 女拳：“女权”的谐音，常被部分男性用于批评部分他们眼中的极端或不理性的女权主义者。
- ▶ 小仙女：数年前常被部分女性用以形容自己精致美丽，现在常被部分男性用于批评部分他们眼中极度自恋、浮夸、矫情的女性。
- ▶ 下头男：常被部分女性用于批评部分他们眼中令人感到扫兴、无语、恶心，并且特别希望远离的男性。
- ▶ 普信男：常被部分女性用于批评部分男性“普通且自信”，明明没有什么特点却总是自以为是。

# 现状分析——百度指数

- ▶ 检索词语“女拳”、“小仙女”、“下头男”、“普信男”的数据。
- ▶ 尽管个别时候，这些词语的确可以精准地描绘出个别人的形象，但目前这些词语已经有明显的滥用现象，可以在尚不明确该词定义或尚不明确对方行为的情况下，直接为一众人“贴标签”、“扣帽子”，成为了引发性别对立攻讦的最佳武器。



# 现状分析——时序数据

关键词

女拳

⌵

+ 添加对比

确定

搜索指数 ?

对比时间段

2017-12-02 ~ 2024-06-25

自定义

PC+移动

全国



■ 女拳

一部名为《女拳》  
的励志爱情剧

是什么让“女拳”肆无忌惮？  
如何分辨女权与田园女拳？

各类性别相关话题

☒ 新闻头条 ☐ 平均值



# 现状分析——时序数据

关键词

小仙女

+ 添加对比

确定

搜索指数 小仙女陪你逛西博车展·春季展!

五一长假 小仙女陪你逛西博车展春季展  
多多与妹妹穿同款连衣裙 好像两个小仙女

对比时间段

2018-12-04 ~ 2024-06-25

自定义

PC+移动

全国



☒ 新闻头条 ☐ 平均值



# 现状分析——时序数据

关键词

下头男

+ 添加对比

确定

搜索指数 ?

对比时间段

2021-12-04 ~ 2024-06-21

自定义

PC+移动

全国



下头男

☒ 新闻头条 ☐ 平均值



# 现状分析——时序数据

关键词

普信男

+ 添加对比

确定

搜索指数 ?

对比时间段 | 2020-12-02 ~ 2024-06-25 | 自定义 | PC+移动 | 全国 |

普信男

☒ 新闻头条 ☐ 平均值

12,000

10,000

8,000

6,000 无数据

4,000

2,000

PDD与Blackpink开黑吃鸡,回看自己操作后直呼:终于理解普信男了  
杨倩直播被吐槽颜值太低?遭普信男抨击:还是把口罩戴上  
甩掉“普信男”后,他开始对我狂舔

2021-02-15 2021-05-03 2021-07-19 2021-10-04 2021-12-20 2022-03-07 2022-05-23 2022-08-08 2022-10-24 2023-01-09 2023-03-27 2023-06-12 2023-08-28 2023-11-13 2024-01-29 2024-04-15 2024-06-24

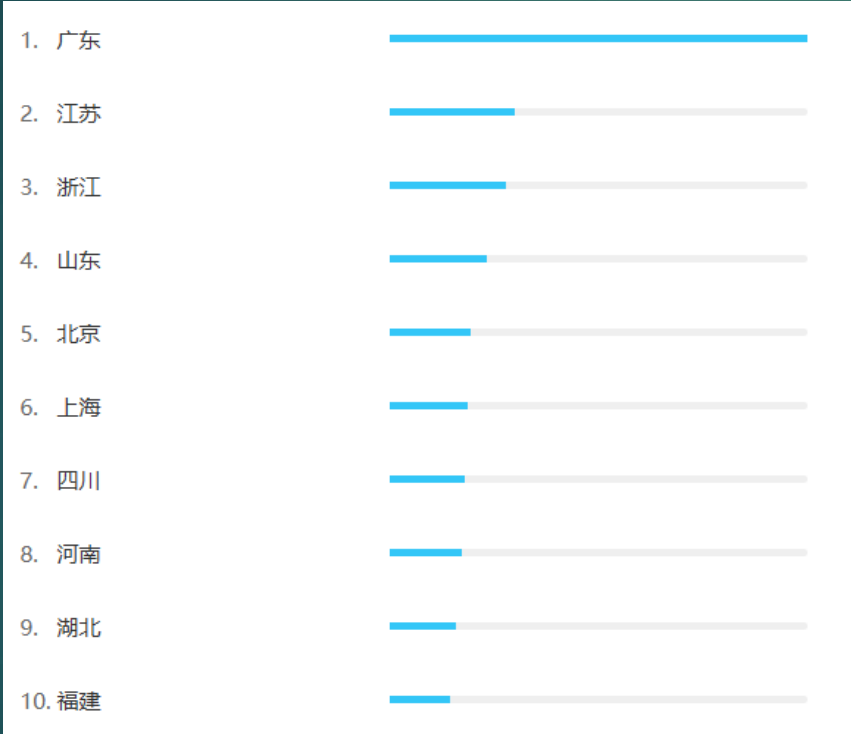
@百度指数

# 现状分析——时序数据

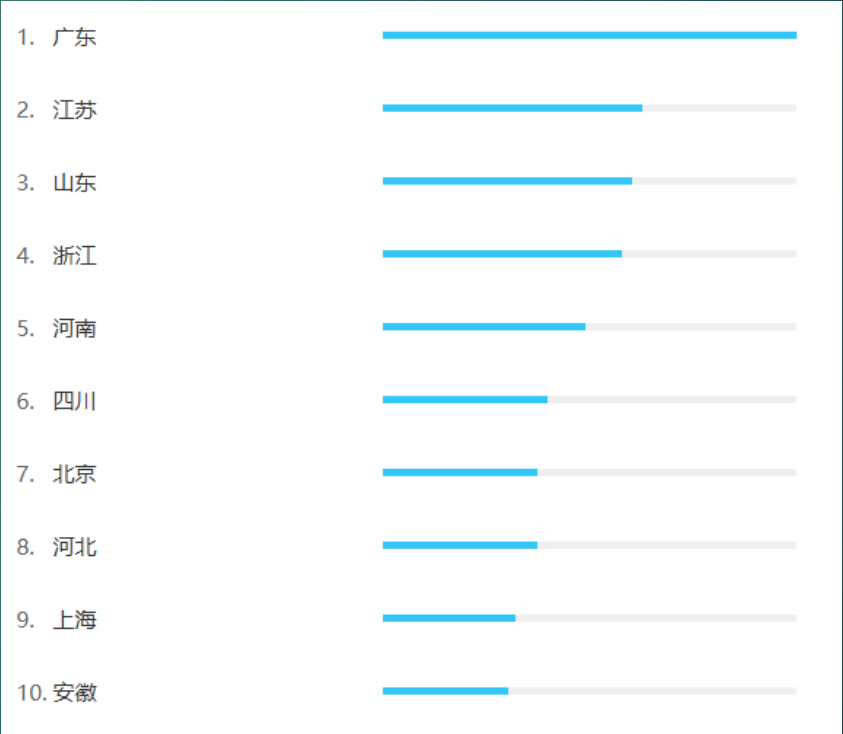
- ▶ 自2019年以来出现的与性别对立相关的词汇，自诞生之后便一直在被人使用，至今热度没有明显下降，性别对立的词语有“日常化”趋势。
- ▶ 这些词语在诞生后，会在个别时间点出现突然的搜索量陡增。这极有可能是在当时发生了与性别高度相关的事件，引发了公众的强烈讨论，故这些具有攻击性的词汇再次被人们高强度的反复使用。
- ▶ “普信男”的最高搜索指数可超过 10,000，是其他词语的最高搜索指数的四倍以上。这体现出“普信男”是一个非常可以引发大众共鸣与讨论的词语。

# 现状分析——地域特性

女拳



小仙女





# 现状分析——地域特性

下头男



普信男

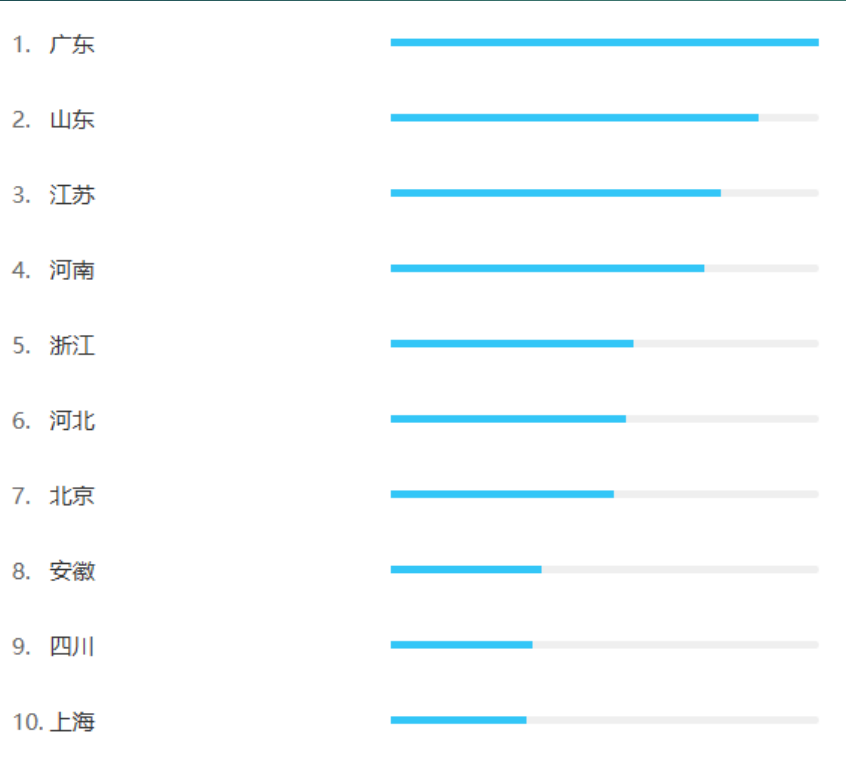


# 现状分析——地域特性

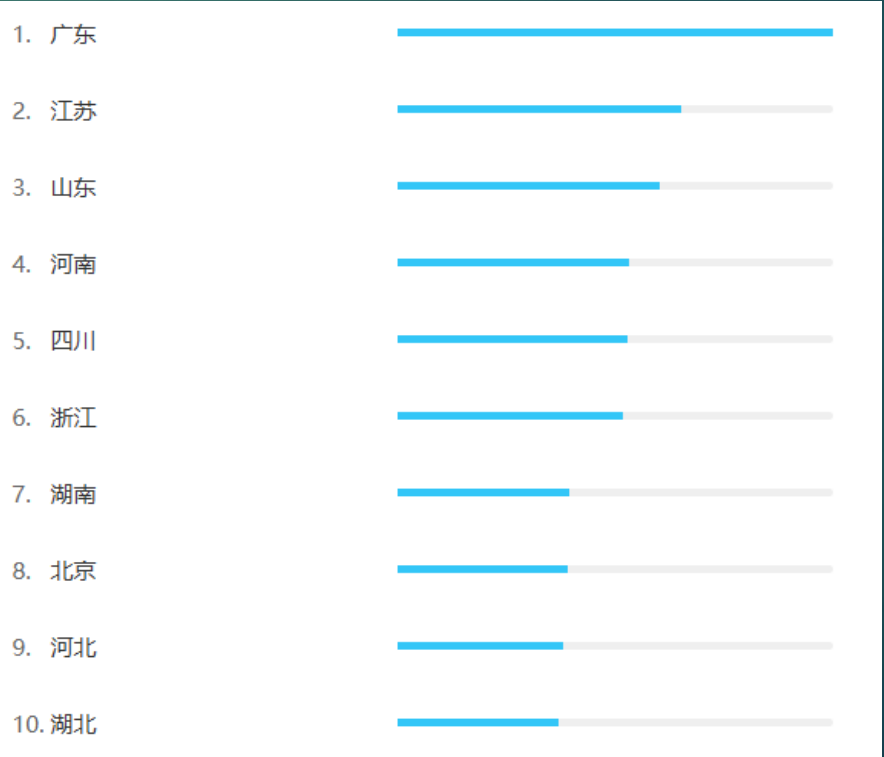
- ▶ 按照先前所展示的图表，似乎可以得到一个结论：
- ▶ 广东人非常在意性别对立！
- ▶ 但实际上，如果我们检索其他与性别相关度低的词语，我们会发现，广东人其实一直在高强度搜索所有话题。

# 现状分析——地域特性

小麦



新世纪福音战士（日本动画作品）



# 现状分析——地域特性

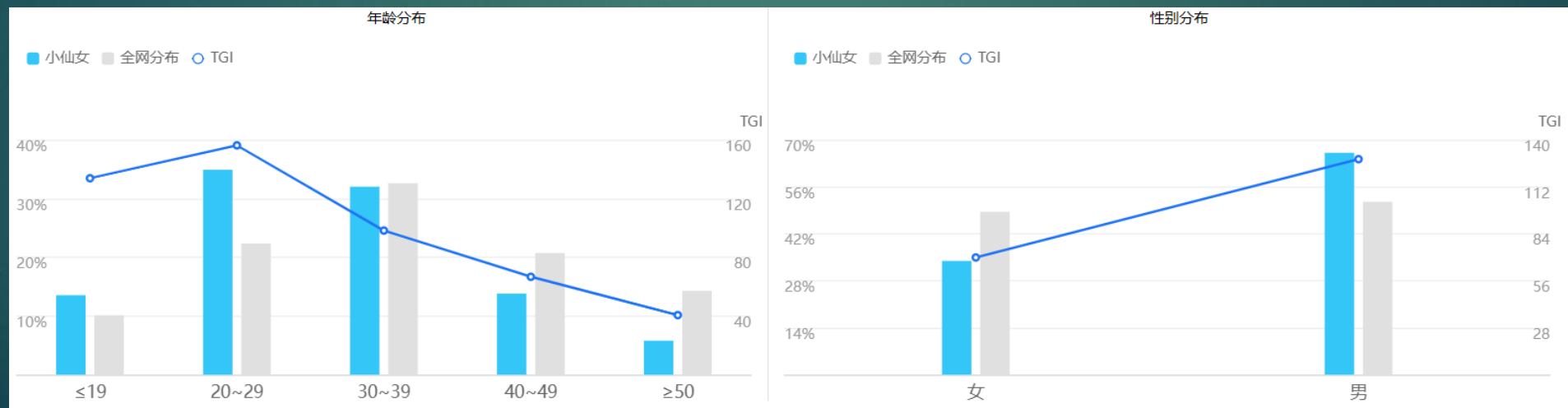
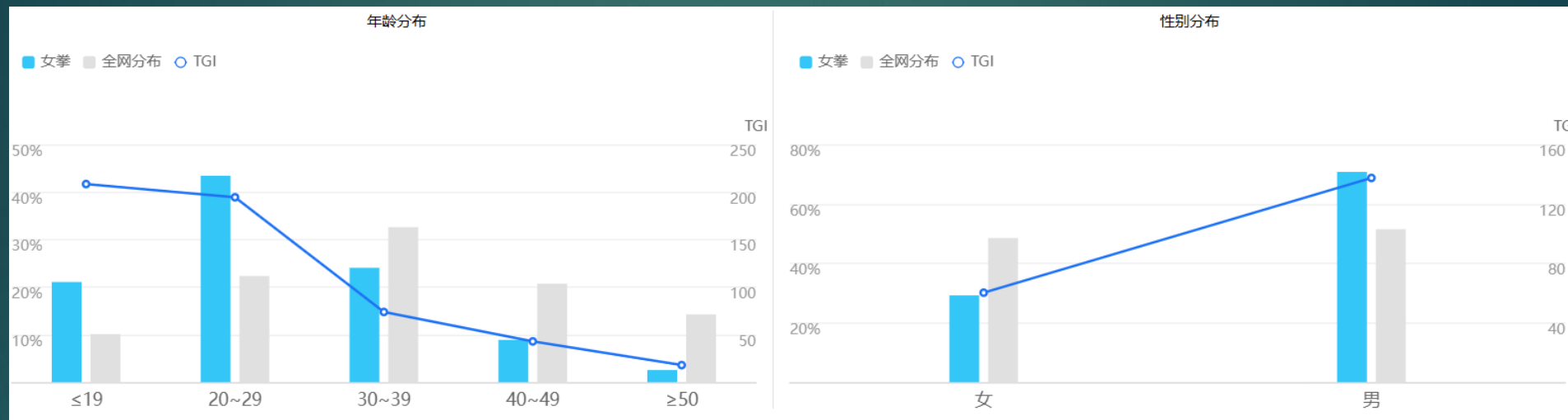
- ▶ 通过“百度指数”，我们只能得到“广东省的网民很多”这一结论。这个结论也非常自然，因为根据第七次全国人口普查数据<sup>[2]</sup>，广东省是全国人口最多的省，人口高达 1.26 亿。
- ▶ 因此，由“百度指数”的人群画像中的地域特性来判断哪一个地区的人对性别对立更感兴趣，效果不佳。
- ▶ 考虑后续对微博数据进行基于 IP 地址的地域特性分析。

---

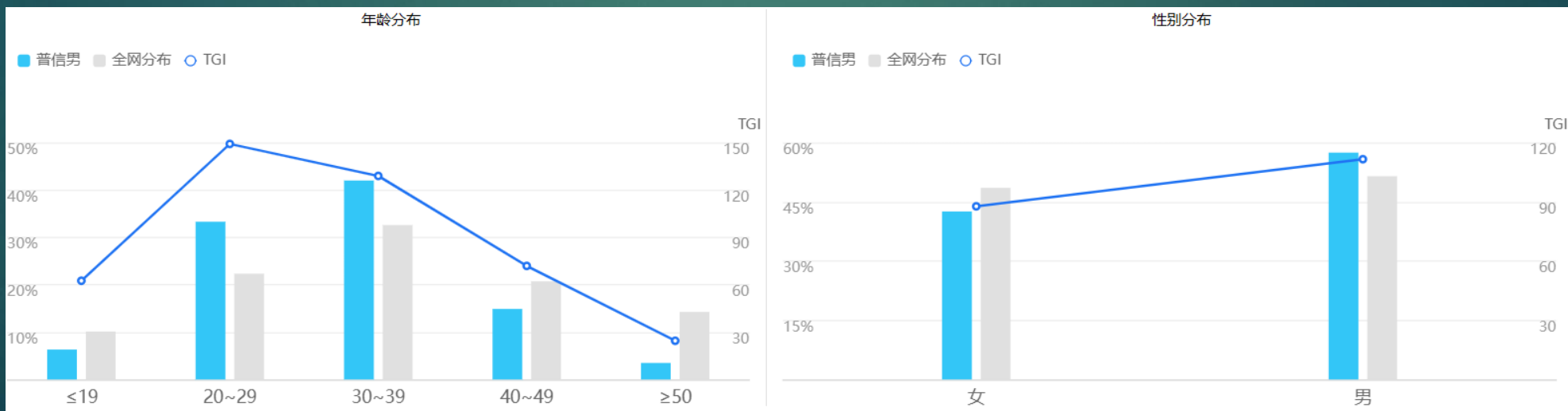
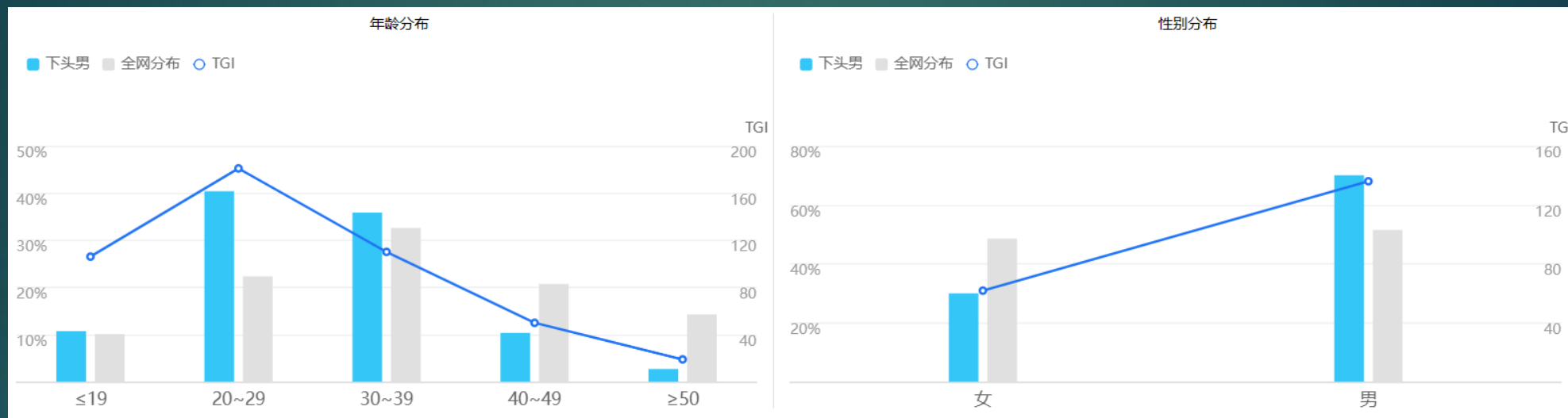
[2] 第七次全国人口普查公报（第三号）

[https://www.stats.gov.cn/sj/zxfb/202302/t20230203\\_1901083.html](https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901083.html)

# 现状分析——人群属性



# 现状分析——人群属性





# 现状分析——人群属性

- ▶ 在年龄角度上，对这些性别对立词语比较在意的人以 20~39 岁居多。这说明，青年人是最在意性别对立话题的。
- ▶ 在性别角度上，对这些性别对立词语的一般均以男性居多。在这四个词上，男性占据了所有搜索用户的55%以上。
- ▶ 男性对这些容易引发性别对立的词汇的关注度均比女性高，这是一个非常有趣的现象。我们在此处追加搜索“性别歧视”这一词语。

# 现状分析——人群属性

- ▶ 检索“性别歧视”的女性要比男性明显更多。
- ▶ 假设大部分百度账号持有者的性别是真实的，这体现出，男性往往对性别对立的词语有较高关注度，女性则对于性别歧视现象有更高的关注度。

